

CompTIA Server+ Certification Guide

A comprehensive, end-to-end study guide for the SK0-004 certification,
along with mock exams



Packt>

www.packt.com

Ron Price

CompTIA Server+ Certification Guide

A comprehensive, end-to-end study guide for the SK0-004 certification, along with mock exams

Ron Price



BIRMINGHAM - MUMBAI

CompTIA Server+ Certification Guide

Copyright © 2019 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing or its dealers and distributors, will be held liable for any damages caused or alleged to have been caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

Commissioning Editor: Gebin George
Acquisition Editor: Rohit Rajkumar
Content Development Editor: Ronn Kurien
Technical Editor: Swathy Mohan
Copy Editor: Safis Editing
Project Coordinator: Jagdish Prabhu
Proofreader: Safis Editing
Indexer: Tejal Daruwale Soni
Graphics: Tom Scaria
Production Coordinator: Arvindkumar Gupta

First published: February 2019

Production reference: 1250219

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham
B3 2PB, UK.

ISBN 978-1-78953-481-8

www.packtpub.com



`mapt.io`

Mapt is an online digital library that gives you full access to over 5,000 books and videos, as well as industry leading tools to help you plan your personal development and advance your career. For more information, please visit our website.

Why subscribe?

- Spend less time learning and more time coding with practical eBooks and Videos from over 4,000 industry professionals
- Improve your learning with Skill Plans built especially for you
- Get a free eBook or video every month
- Mapt is fully searchable
- Copy and paste, print, and bookmark content

Packt.com

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.packt.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at customercare@packtpub.com for more details.

At www.packt.com, you can also read a collection of free technical articles, sign up for a range of free newsletters, and receive exclusive discounts and offers on Packt books and eBooks.

Contributors

About the author

Ron Price (Server+, A+, Network+, Security+, CCNA, MBA, AAGG) began his experience in computing as a programmer on a mainframe operating system project. He has experience in system design, database systems, operational administration and senior management. In addition to his writing, Ron is an instructor of information systems at Spokane Falls Community College.

About the reviewer

Christopher Rees is a lifelong learner, an IT technology leader, an author at Pluralsight, and a former law enforcement officer who focused on computer crime investigations. For the past 20 years, he has been working in the enterprise IT space and has trained over 100,000 people from around the world via the online training courses he's developed in the areas of networking, cybersecurity, and business continuity management.

Married for more than 20 years with 3 beautiful children, Chris enjoys keeping fit, maximizing his time with his family and friends, and, of course, keeping up with the latest tech and industry trends.

Packt is searching for authors like you

If you're interested in becoming an author for Packt, please visit authors.packtpub.com and apply today. We have worked with thousands of developers and tech professionals, just like you, to help them share their insight with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Table of Contents

Preface	1
Section 1: Section 1: System Architecture	
Chapter 1: Server Hardware	7
Server roles	7
Application servers	8
Database servers	9
Directory servers	9
File servers	10
Mail servers	11
Messaging servers	11
Network services servers	11
Print servers	12
Proxy server	13
Routing and Remote Access Service (RRAS)	14
Virtual server	14
Form factors	15
Tower servers	16
Rack mounts	17
Blade technology	18
Server power systems	19
Electrical power	20
AC versus DC / 110V versus 230V	20
Wye and delta	21
Negative 48V	21
One phase versus three phases	21
PSU	23
Wattage	23
The 80-plus certification	24
Selecting the right PSU	24
Redundancy	26
System heat	28
Cooling systems	29
Air cooling and air flow	29
Summary	30
Questions	31
Chapter 2: Server Internals	34
CPUs	34
Multiprocessors	35

Symmetrical Multiprocessing (SMP) versus Asymmetrical Multiprocessing (ASMP)	35
SIMD, MISD, and MIMD	36
Multiple core processing	36
CPU packages and sockets	38
Cache memory	41
CPU cache memory	41
CPU cache memory levels	42
Write-back/write-through cache	42
Advanced RISC Machine (ARM) servers	43
CPU multiplier	43
CPU stepping	43
Main memory	44
RAM	44
Double Data Rate (DDR) RAM	44
RAM packaging	45
Memory timing	47
Error-correction code (ECC) versus non-ECC	48
Dual channel memory	48
Color-coded RAM slots	49
Buses, channels, and expansion slots	50
Bus width	50
Peripheral Component Interconnect (PCI) bus	51
PCI size and fit standards	52
PCI conventional	52
PCI-e	54
Expansion cards	55
Network interface controller (NIC)	55
Host Bus Adapter (HBA)	55
Redundant Array of Independent Disks (RAID) controller	55
Riser cards	56
USB interface and port	56
Configuration	57
BIOS	58
UEFI	58
Summary	58
Questions	59
Chapter 3: Data Storage	62
Data storage devices and their specifications	62
Hard drive specifications	63
Form factors	63
Small form factor (SFF)	63
Large form factor (LFF)	64
HDD specification and configuration	65
Disk capacity – decimal versus binary	66
Hard disk drive (HDD) versus solid-state drive (SSD)	67

SSD specification and configuration	69
Hard disk interfaces	69
Data storage systems	70
Direct-attached storage (DAS)	71
Network-attached storage (NAS)	71
Storage area network (SAN)	72
SAN fabric	72
SAN communications	73
Logical Unit Number (LUN) zoning and masking	73
Filesystem	74
Operating systems and filesystems	75
File sharing	75
RAID	76
Striping and mirroring	76
RAID levels	77
RAID implementation	79
Disk quotas	80
Disk compression	81
High availability (HA)	82
The nines	83
Fault tolerance	83
Replacing failed components	84
Disk storage capacity planning	84
Other storage devices	85
Magnetic tape	86
Optical storage	86
Summary	87
Questions	88
Chapter 4: Server Operating Systems	91
The network server	92
Server functions	92
Network server operating systems	92
Operating system (OS) functions	93
User/computer communications	93
Memory management	94
Dynamic loading and linking	94
Memory allocation	94
Control and coordination of hardware	95
The use of system resources	95
Internal and network file management	100
User, data, application, and resource security	101
Hardware configuration	101
The primary parts of an OS	102
The OS and hardware	103
Boot sequence	104
Firmware	104

Preparing a disk for the OS	106
Filesystems	109
Formatting	109
Filesystems by OS	109
Journaling	110
Special function filesystems	110
Network configuration	110
Configuring the hostname	111
Configuring a hostname on Windows Server	111
Configuring a hostname on a Linux server	112
User accounts	114
Creating a local user account	114
Creating a domain user account	117
Adding a workstation to a domain	118
Connecting to a network	119
Connecting a PC to a network	119
Adding server roles and features	120
Unattended and remote installations	123
NOS optimization	124
Summary	124
Questions	125
Chapter 5: Addressing	128
IP addressing	129
IP version 4	129
The IPv4 address structure	129
Classful IP addressing	130
LAN addressing	130
Private IP addresses	131
Network and host IDs	132
Network Address Translation (NAT)	133
Collision domains	134
Broadcast domains	135
Classless Interdomain Routing (CIDR)	136
Subnetting	137
Subnets and hosts	137
Subnet masks	139
Network and broadcast addresses	140
Internet Protocol version 6 (IPv6)	140
The IPv6 address structure	141
Reserved prefixes	142
IPv6 address compression	143
IPv6 leading zero compression	143
IPv6 network ID	143
Address categories	144
MAC addressing	144
Address resolution	145
ARP	146
DNS	146

DNS search	146
Domain suffix	147
The Windows Internet Name Service (WINS)	147
Ports and protocols	148
Well-known ports	148
Registered ports	149
Summary	150
Questions	151
Chapter 6: Cabling	154
Copper cabling	154
Twisted-pair cabling	155
Coaxial cabling	158
Network connectors	159
EIA/TIA 568 facility standards	160
Category cabling	161
Ethernet cable standards	162
Fiber-optic cabling	163
Fiber-optic cable modes	165
SM fiber-optic cable	165
MM fiber-optic cable	165
Fiber-optic cable connectors	166
Network cable installation	168
Summary	171
Questions	172
Section 2: Section 2: Administration	
<hr/>	
Chapter 7: Server Administration	176
Hardware administration	176
Network administration	177
Configuring, updating, and maintaining network hardware	177
KVM interfaces	178
Serial interfaces	180
Network-based hardware administration	181
Network-based operating system administration	182
Asset management	184
Information Technology Asset Management (ITAM)	184
IT life cycle asset management	185
Additional ITAM terms	185
System documentation	186
Service manuals	186
System and network documentation	187
System diagrams	188
System documentation	189
Other documents and documentation	190
Storing sensitive documentation	190

Summary	191
Questions	191
Chapter 8: Server Maintenance	194
Change and patch management	195
Change control process	195
Patch management	196
OS updates	198
Device driver updates	200
Firmware updates	201
Hardware maintenance	202
Server monitoring systems	202
Light Emitting Diodes (LED) server status indicator	205
Liquid Crystal Display (LCD) messages	205
Beep codes	206
Replace failed components	207
Preventive maintenance	208
Fault tolerance and high availability	209
Clustering	210
Active/active versus active/passive clusters	211
Load balancing	212
Heartbeat	214
Hot and not hot	214
Hot swap	214
Non-hot swap	215
Service level agreements (SLA)	215
Summary	216
Questions	217
Chapter 9: Virtualization	220
Virtual networking	221
Virtual network components	221
Virtual devices	222
Virtual servers	222
Hypervisors	223
Hosts and guests	224
Virtual machine (VM)	224
Hardware configuration for a virtual environment	225
Virtual resource allocation	226
Network connectivity	227
Virtual internetworking devices	227
Summary	228
Questions	229
Chapter 10: Disaster Recovery	231
Business continuity plan (BCP)	231

BIA	232
Risk assessment	233
Continuity of operations	233
DRP	234
Recovery plans	234
Recovery sites	235
Replication and backup	236
Data replication	236
Synchronous and asynchronous	236
Replication methods	238
Data backup	239
Archive bit	239
Backup methods	239
Data versus OS restore	240
Backup media	241
Media storage	241
Backup media integrity	243
Backup media retention	243
Summary	244
Questions	245
Section 3: Section 3: Security	
<hr/>	
Chapter 11: Security Systems and Protocols	248
Security zones	249
Firewall zones	249
Demilitarized zone (DMZ)	249
Browser zones	250
Security devices	251
Authentication protocols	251
Authentication methods	252
Point-to-point authentication protocols	253
AAA authentication protocols	254
Secure Sockets Layer (SSL)/Transport Layer Security (TLS)	255
Internet Protocol Security (IPSec)	256
IPSec policies	256
IPSec modes	258
Port security	259
Port-based security	260
IEEE 802.1x	260
Access control list (ACL)	261
Router ACLs	261
Access list content	262
ACL types	263
Standard ACLs	264
Extended ACLs	264

Other ACL types	265
ACE types	266
Wildcard masks	267
Public key infrastructure (PKI)	268
PKI features	268
Encryption and authentication	268
Virtual private network (VPN)	269
Virtual LAN (VLAN)	269
Summary	270
Questions	271
Chapter 12: Physical Security and Environmental Controls	273
MFA	273
Passwords	274
Authentication factors	275
General physical security concepts	276
Threats to physical security	276
Environmental threats	277
Man-made threats	277
Site-specific threats	278
Technical threats	279
Physical security devices	280
Environmental controls	281
Environmental monitoring	281
Electrical power	282
Uninterruptable Power Supplies (UPS)	283
UPS ratings	284
Automated shutdown of attached devices	284
Power distribution	285
PDU types	286
PDU ratings	287
Physical safety issues	288
Summary	289
Questions	290
Chapter 13: Logical Security	293
Access control	293
Access control criteria	294
Access control levels	295
Filesystem access control	295
Access control to peripherals	297
Administration access control	298
Security and distribution groups	298
Network access control (NAC)	298
Data encryption	299
Storage encryption	299
Data retention and disposal	300
Erasing a disk	300

Formatting	301
Physically destroying a disk drive	302
Hardening	303
OS hardening	303
System hardening	304
Application hardening	304
Hardware hardening	305
Host hardware hardening	305
Network device hardening	305
Endpoint security	306
Summary	307
Questions	308
<hr/> Section 4: Section 4: Troubleshooting <hr/>	
Chapter 14: Troubleshooting Methods	312
Troubleshooting steps	312
Identifying the problem	313
Hardware or software?	314
Hardware problems	314
Software problems	315
Establishing a probable cause	315
Define a plan of action	316
Verifying functionality	316
Documenting findings, actions, and outcomes	317
Summary	318
Questions	319
Chapter 15: Common Hardware Issues	321
Hardware problems	321
Identifying a hardware problem	322
Common problems	323
POST failure	323
Overheating	324
Processor failure	325
Memory failure	327
Motherboard and component issues	328
Capacitor issues	328
Burns	330
USB not recognized	330
Expansion bus	331
PSUs	332
Hard Disk Drives (HDDs)	333
Video display	335
Other common problems	337
Environmental issues	338

Summary	339
Questions	340
Chapter 16: Common Software Issues	343
Software problems	343
Hardware-related software problems	344
Common operating systems problems	345
Common problem causes	348
User Account Control (UAC)	348
Windows UAC	349
Access control	350
Corrupted files	352
Windows file recovery	352
Linux file recovery	353
Hard disk space problems	354
Lack of system resources	355
Virtual memory problems	356
Fragmentation	357
Printing issues	358
Log files	360
Operating system monitoring tools	361
Summary	362
Questions	363
Chapter 17: Common Network Issues	365
Common network problems	365
Internet connectivity	366
Configurations	368
Dynamic Host Configuration Protocol (DHCP) server	368
APIPA	368
DHCP addresses	369
Other misconfigured devices	369
Email problems	370
Hosts file configuration	372
Misconfigured NIC	372
Routing and switching issues	374
VLAN configuration errors	374
Default gateway not available	376
Firewall failure	377
Miscellaneous common problems	378
Troubleshooting tools	379
ping	379
tracert/traceroute	379
ipconfig/ifconfig	380
nslookup	381
net use/mount	382
nbtstat and netstat	382
Summary	382

Questions	383
Chapter 18: Common Storage Issues	385
Data storage device problems	385
Common HDD problems	387
Causes of common problems	390
Media failures	391
Hard disk media	391
SSD media	391
Magnetic tape media	392
Optical drives	393
Common storage problems causes	393
Drive and connector failures	393
HDD problems	394
Cable and connector problems	395
Storage system issues	395
Software-related failures	396
Hardware-related issues	396
Storage array issues	397
Administrative tools	399
Disk management	399
Disk partitioning tools	400
Map, mount, and net use	401
Disk arrays	401
RAID arrays	402
Storage monitoring tools	404
Summary	407
Questions	408
Chapter 19: Common Security Issues	410
Common data security problems	410
Causes of common security problems	414
Security tools	419
Summary	422
Questions	422
Appendix A: CompTIA Server+ Examination	425
The exam	425
Registering for the Exam	427
Preparing for the exam	428
The certification	429
Appendix B: Glossary	430
0-9	430
A	430
B	431

C	432
D	434
E	436
F	436
G	437
H	437
I	438
J	439
K	439
L	439
M	440
N	441
O	442
P	442
Q	444
R	444
S	445
T	447
U	448
V	449
W	449
Z	449
Assessment	450
Other Books You May Enjoy	458
Index	461

Preface

The CompTIA Server+ certification is one of the top five IT certifications that is vendor neutral. System administrators opt for the CompTIA Server + certification to gain advanced knowledge on concepts such as troubleshooting and networking.

This book will start with the configuration of a basic network server and the configuration of each of its myriad roles. The next set of chapters will provide an overview of the responsibilities of and the tasks performed by a system administrator to manage and maintain a network server. Going ahead, you will learn about the basic security technologies, methods, and procedures that can be applied to a server and its network.

Next, you will cover troubleshooting procedures and methods in general, and specifically for hardware, software, networks, storage devices, and security applications. Towards the end of this book, you will cover a few troubleshooting and security mitigation concepts for running admin servers with ease. This guide is packed with test questions and mock papers, which will help you pass the exam.

By the end of this book, you will be in a position to pass the CompTIA Server+ certification with ease.

Who this book is for

This book is targeted toward professionals seeking to gain the CompTIA Server+ certification. People from a Microsoft background with basic operating system and networking skills will also find this book useful. Basic experience of working with system administration is mandatory.

What this book covers

Chapter 1, *Server Hardware*, provides a review of the components that are likely to be found in a network server and gives enough detail to help an inexperienced reader understand the what and the why of server hardware. In the discussion of each of the functional server modes, its protocols, services, and purpose are also discussed. The chapter also looks at server power, cooling, and form factors.

Chapter 2, *Server Internals*, examines the components and systems inside the server's computer case to provide a brief overview of the purpose and function of each of them, as well as how they interact with other components.

Chapter 3, *Data Storage*, examines the devices and components that make up data storage systems that are common on networks. This chapter also discusses the various interfaces, technologies, and configurations of magnetic storage devices.

Chapter 4, *Server Operating Systems*, discusses the installation, configuration, and management of a network server operating system. Both Windows Server and Linux are covered. The chapter also takes a look at creating performance baselines and the configuration and administration of unattended or remote server installations.

Chapter 5, *Addressing*, provides a detailed look at IPv4 and IPv6 addressing, including discussions on CIDR, subnetting, DNS, MAC, and FQDN. This chapter also includes information on network interfaces and TCP/UDP protocols and ports.

Chapter 6, *Cabling*, provides information on copper and fiber-optic cabling systems, including their connectors, configurations, designations, and installation.

Chapter 7, *Server Administration*, covers the tools, components, tasks, processes and management responsibilities used or performed to administer and maintain a server.

Chapter 8, *Server Maintenance*, covers the duties and activities involved in maintaining a server. This includes change and patch application and management, performance monitoring, and preventive maintenance.

Chapter 9, *Virtualization*, covers the concepts, configuration, and operation of virtualization technology, including hypervisors, hardware compatibility, allocation of resources, and virtual devices.

Chapter 10, *Disaster Recovery*, reviews the definitions, methods, products, and applications involved in disaster recovery and business continuity planning and execution.

Chapter 11, *Security Systems and Protocols*, covers the systems, protocols, and encryption key methods applied to secure a server. This includes firewalls, authentications, PKI, and security zones.

Chapter 12, *Physical Security and Environmental Controls*, covers the concepts, technologies, and methods applied in physical security programs, including MFA, security devices, and practices. This chapter also includes a discussion of the various electrical power concepts and applications, safety procedures, and the elements of environmental control.

Chapter 13, *Logical Security*, covers the concepts, technologies, and applications used to define and apply security procedures through system administration. This chapter also discusses data encryption, data storage security, hardening, and endpoint security.

Chapter 14, *Troubleshooting Methods*, discusses the procedures that should be used in any troubleshooting activity.

Chapter 15, *Common Hardware Issues*, identifies common hardware issues and the processes or methods used to isolate hardware issues and their causes.

Chapter 16, *Common Software Issues*, identifies common software issues on a server, their causes, and the tools used to detect, prevent, and resolve them.

Chapter 17, *Common Network Issues*, identifies common network issues on networks, their causes, and the tools used to detect, prevent, and resolve them.

Chapter 18, *Common Storage Issues*, identifies common hardware and software issues associated with disk drive storage attached to a server or network, their causes, and the tools used to detect, prevent, and resolve them.

Chapter 19, *Common Security Issues*, identifies common hardware and software issues associated with server and network security, their causes, and the tools used to detect, prevent, and resolve them.

Appendix A, *CompTIA Server+ Examination*, in this section, this section will go through the basic pre-requisites to clear the exam.

Appendix B, *Glossary*, this section will walk-through the basic term and definitions that are used throughout the book.

Appendix C, *Server+ Practice Exam*, you can test your knowledge of concepts required for CompTIA's Server+ exam by visiting the following link: https://www.packtpub.com/sites/default/files/downloads/Server_plus_Practice_Exams.pdf.

To get the most out of this book

In this book, you need the following:

- A PC with a working internet connection
- Windows system, preferably Windows Server, but Windows 10 is okay. You need Administrator permissions as well.
- A Linux system, an emulator running on Windows works too.

Download the color images

We also provide a PDF file that has color images of the screenshots/diagrams used in this book. You can download it here: https://www.packtpub.com/sites/default/files/downloads/9781789534818_ColorImages.pdf.

Conventions used

There are a number of text conventions used throughout this book.

CodeInText: Indicates code words in text, database table names, folder names, filenames, file extensions, pathnames, dummy URLs, user input, and Twitter handles. Here is an example: "In this entry, 162.29.5.12 is the source IP address and 0.0.0.0 is the wildcard mask."

A block of code is set as follows:

```
11111111.11111111.11111111.11111111 (decimal 255.255.255.255)
```

Any command-line input or output is written as follows:

```
$ nslookup packt.com
```

Bold: Indicates a new term, an important word, or words that you see onscreen. For example, words in menus or dialog boxes appear in the text like this. Here is an example: "The message **The service cannot be started, either because it is disabled or because it has no enabled devices associated with it** indicates that one or more services, programs, or scripts has failed to start."



Warnings or important notes appear like this.



Tips and tricks appear like this.

Get in touch

Feedback from our readers is always welcome.

General feedback: If you have questions about any aspect of this book, mention the book title in the subject of your message and email us at customercare@packtpub.com.

Errata: Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you have found a mistake in this book, we would be grateful if you would report this to us. Please visit www.packt.com/submit-errata, selecting your book, clicking on the Errata Submission Form link, and entering the details.

Piracy: If you come across any illegal copies of our works in any form on the Internet, we would be grateful if you would provide us with the location address or website name. Please contact us at copyright@packt.com with a link to the material.

If you are interested in becoming an author: If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, please visit authors.packtpub.com.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions, we at Packt can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about Packt, please visit packt.com.

1

Section 1: System Architecture

This part of the book covers the configuration of a basic network server and the configuration appropriate to each of its myriad roles. The chapters in this part discuss hardware, operating systems, data storage, network addressing, and cabling.

The following chapters are included in this section:

- Chapter 1, Server Hardware
- Chapter 2, Server Internals
- Chapter 3, Data Storage
- Chapter 4, Server Operating Systems
- Chapter 5, Addressing
- Chapter 6, Cabling

1

Server Hardware

It's generally assumed that a computer network server, at least in the way we talk about it, is hardware first and software second. While it's easier to envision a computer as a network server, in fact, the server is a piece of software running on the computer. In its most strict definition, a server is anything that provides services to fulfill requests made to it. Therefore, someone who takes our order and brings us our meal in a restaurant is a server and, in the same way, software running on a computer that processes an SQL request on a database and returns the data to the requester is also a server. Regardless of the way you envision a server, for the sake of learning about servers, let's agree that a centralized computer running server software that provides services to a network is a server.

With that understanding, let's look at the various roles fulfilled by a computer network server and the hardware of a typical computer in the role of a server.

In this chapter, we will cover the following topics:

- Server roles
- Form factors
- Server power systems
- System heat

Server roles

The software running on a server defines the role of that server. In fact, a server can have two or more different roles at times; it depends on the software. The list of the different roles a server can fulfill is long, but for the purposes of the Server+ exam, you should know the role and function of each of the following server types:

- Application server
- Database server
- Directory server

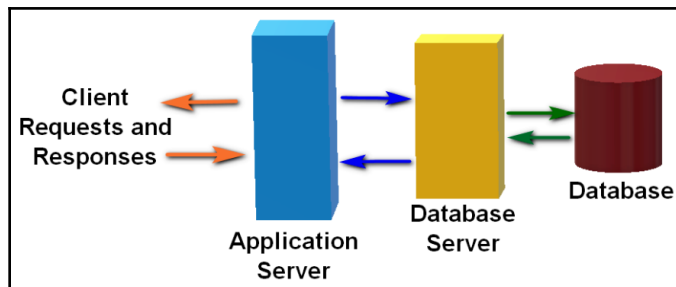
- File server
- Mail server
- Messaging server
- Network services server
- Print server
- Routing and remote access server
- Web server

The following sections explain each of these server roles.

Application servers

In the current environment of web-enabled or **Software-as-a-Service (SaaS)** applications in the cloud, an application server functions much like the generic description given above. An application server often provides services for one or more applications and serves as a mid-level service between user requests and other server- or network-based functions, such as a database system. There are three basic types of application servers. Their differences lie in what they do and where they fit into a process. The three types of application servers are as follows:

- **LAN application servers:** This type of application server can exist internally within an organization's local network and provide data-processing support to network users on one or more applications. They may host an entire application's processing or share the processing with a user's computer. A common implementation of this type of application server is a three-tier client/server environment in which the application server is middle-ware between a network user and a database management system. The following diagram illustrates a three-tier client/server system:



In a three-tier client/server system, an application server provides services to both the user and a database management system or other function-specific servers

- **Query-based application servers:** This type of application server hosts one or more scripting or programming language services used to request data from a database. A user's computer may have an active dashboard, a status board, or a specific scripting or service request system, such as **Active Server Page (ASP)**, **JavaServer Pages (JSP)**, Django, or Ruby on Rails. The application server accesses a database and returns current or real-time data back to the client software.
- **Application/web servers:** In many cases, application servers are becoming web servers and vice versa. Either type of server can support **Hypertext Transfer Protocol (HTTP)** request-and-response traffic and interact with client browsers. A stand alone web server (also called an HTTP server) typically includes several specialized scripts and database query services in addition to performing basic web server duties. A web-enabled application server includes the capability to deliver web content to a client's browser. Examples of web/application servers are IBM WebSphere, Oracle iPlanet, and Apache Tomcat, and Microsoft **Internet Information Services (IIS)**.

Database servers

As shown in the preceding diagram, in the *Application servers* section, a database server provides an interface between client requests, either directly or through an application server, and a database management system and its database. In most cases, an application server passes data requests to the database server for the processing and retrieval of the requested data. The database server then returns the data back to the requesting node. In a database client/server environment, the database management system, which performs the input/output operations on the database, is the backend. The software running on a host computer or an application server is the frontend. Requests for data flow from the frontend to the backend and back again.

Directory servers

A directory server supports directory services. Okay, *but what are directory services? Have you ever entered the lobby of a very tall building and used the directory board to locate where in the building the person or organization was that you needed to find?* Typically, you'd find the name, which has the location on the same line. *Sound familiar?* Directory services cross-reference or map the names, designations, or locations of computer or network resources to their respective local or network addresses.

The resources identified and addressed typically include disk volumes, directories, folders, files, input devices, output devices, and any other devices attached or installed on a system. This service is essential in a network. With this information, a resource is located, used, and administered. Without directory services, network resource addressing would be like a town in which the houses don't have street addresses. Efficient network operations would be impossible. Directory services are also known as name services because they manage a namespace. A **namespace** is a data abstraction that holds a list of names or identities of system resources, in this case, and their network addresses or locations. The namespace allows users, applications, and other services to access resources without the need to know their locations in advance. A directory server, or name server, is a server application that provides the organization, management, and security of the directory or name services, for example, Microsoft Active Directory, Red Hat Directory Server, Lotus Domino.

File servers

A file server is just what its name suggests—a server for files. There are several different types of file servers, but in general, a file server provides data resources to other nodes on a network. The configuration of a file server is a combination of several factors, including storage capacity, access time, security, fault tolerance, and, of course, budget. To best serve the data needs of an organization, a file server must be set up with the right blend of these factors. File servers can serve one of two roles:

- **Dedicated file servers:** This type of file server expressly provides file or database content to clients. A dedicated file server serves in that capacity only.
- **Non-dedicated file servers:** This type of file server supports two or more server services or functions.

What defines each of these roles is the method used for data sharing. File servers can be a **File Transfer Protocol (FTP)** server, a **Service Message Block/Common Internet File System (SMB/CIFS)** protocol server, an HTTP server, or a **Network File System (NFS)** server. Another form of file server arrangement is a **network-attached storage (NAS)** system.

Mail servers

Mail servers, which are also known as **email servers** or **mail transport agent (MTA)**, process and transport electronic mail messages for a network, up to and including the internet. A mail server emulates the functions of human postal workers in that it receives incoming mail and forwards it on to its destination, typically another mail server. The two primary protocols involved with mail servers and the delivery of emails are the **Simple Message Transport Protocol (SMTP)** and the **Post Office Protocol 3 (POP3)**. SMTP transports messages between mail servers. POP3 is a client-based protocol that interacts with a mail server to send and receive messages addressed to a particular user.

Messaging servers

A messaging server is a middleware service that receives, forwards, or holds messages between client applications and services. These messages communicate requests, responses, and status updates between client processes running on a network. There are two primary types of messaging servers:

- **Point-to-point messaging servers:** This type of messaging is a communication between one client, through a messaging server, and a single addressee client. Although other clients may be monitoring the messaging channel, only the single client to which the message is addressed will receive the message. An example of a point-to-point messaging service is the Java message service.
- **Publish-subscribe messaging servers:** This type of messaging communicates a message from a client (the publisher), through the messaging server, to a messaging category that includes multiple subscribed clients. The subscribers indicate which message categories they wish to receive. The clients then receive messages from only the categories to which they have subscribed. An example of publish-subscribe messaging services are Faye, NATS, and Redis.

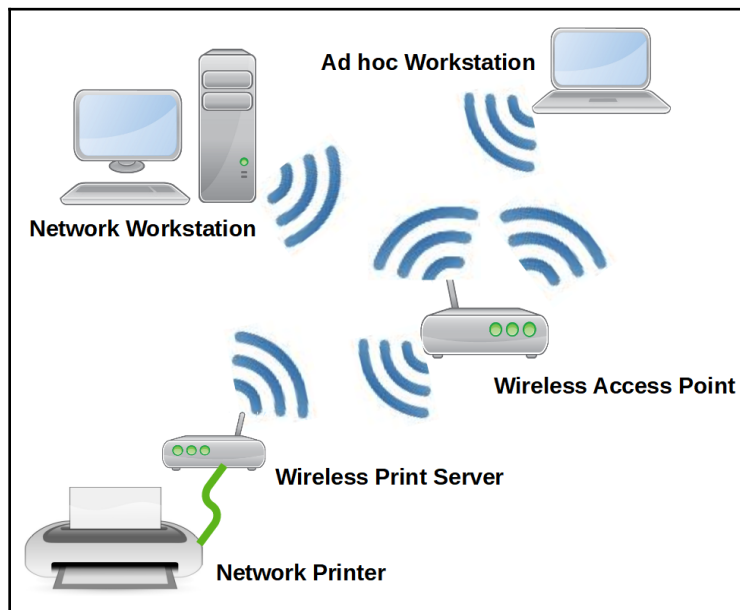
Network services servers

Network services are services provided by the network server to the network clients to provide core services, such as data storage **input/output (I/O)** operations, information display, peer-to-peer communication, and many others. A network service operates on the OSI application layer.

Although the **network operating system (NOS)** provides most network services, protocols, and services such as the **Domain Name System (DNS)**, the **Dynamic Host Configuration Protocol (DHCP)**, instant messaging, **Voice over Internet Protocol (VoIP)**, **Network Time Protocol (NTP)**, and email can run from a centralized network services server.

Print servers

A print server is a device (a computer, appliance, or software) that accepts print requests from clients and provides the sequencing and management of a network-attached printer, plotter, or other imaging device. A printer attached to a desktop computer directly can manage the print function through a print queue, typically on a first-come-first-served basis. On a network, with any number of clients requesting print services, access to a printer can be contentious at times. In addition to managing a network's print queue, a print server can also manage or enforce print policies, such as volume, color printing, and others. Today's print servers are stand-alone network devices dedicated to the single function of printing. The following diagram illustrates a wireless network that includes a print server:



A wireless LAN that includes a print server

Proxy server

Proxy servers are intermediate network services that accept network client requests for resources from remote servers. A proxy server examines a client's request and determines the most efficient way of providing the requested resource. Client requests can be for a service, a file, or a web page, among other network-based resources. In today's networks, proxy servers are web proxies that provide several functions, such as reducing network traffic, concealing a requester's identity, and, getting past IP address blocking. A proxy server doesn't necessarily require a centralized network computer to operate. A proxy server may be on one or more users' workstations, one server on a network, or at several points in between. The location of the proxy server isn't nearly as important as its capability to connect a user's workstation to the sought-after servers on the internet. There are several types of proxy servers, each of which provides a primary service. The most common types of proxy servers are:

- **Gateway proxy servers:** This type of proxy server, also known as **application-level gateways** or **tunneling proxy servers**, serve as portals between a local network and the internet, sending and receiving unchanged client requests and the resulting responses.
- **Internet-facing (forward) proxy servers:** This type of proxy server facilitate requests from their internal networks for resources from the internet.
- **Open proxy servers:** These are forward proxy servers that will send request-and-response messages to or from anywhere on an inter-network.
- **Internal-facing proxy servers:** This type of proxy server provides several ways to protect and service their internal networks. Reverse proxies can perform authentication, authorization, caching, decryption, and load balancing.
- **Reverse proxy servers:** A common use for internal-facing proxy servers is as reverse proxy servers. This type of proxy server accepts requests from the internet, such as HTTP requests, and passes them to the appropriate internal network server for processing.

Routing and Remote Access Service (RRAS)

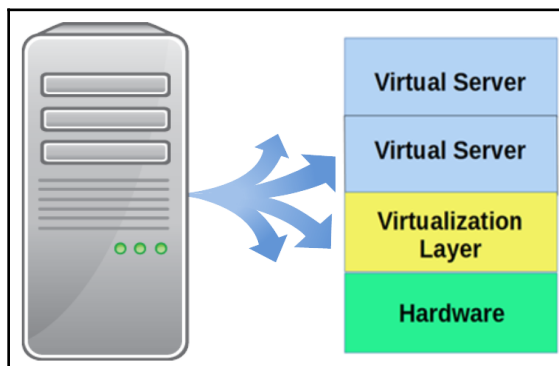
RRAS is a Microsoft suite of protocols configured to provide three basic functions:

- **Firewall:** Windows Firewall in Windows Server 2008 replaced the basic firewall function in RRAS
- **Router:** The server configured to run RRAS can perform multi-protocol routing, including the routing of IP, IPX, AppleTalk, **Routing Information Protocol (RIP)**, **Open Shortest Path First (OSPF)**, and **Internet Group Management Protocol (IGMP)** messages
- **Remote access:** Provides remote access connectivity for dial-up and **virtual private network (VPN)** clients using AppleTalk, IP, or IPX

RRAS incorporates the use of **Point-to-Point Protocol (PPP)** as its transport protocol. This allows RRAS to combine the router and the remote access functions.

Virtual server

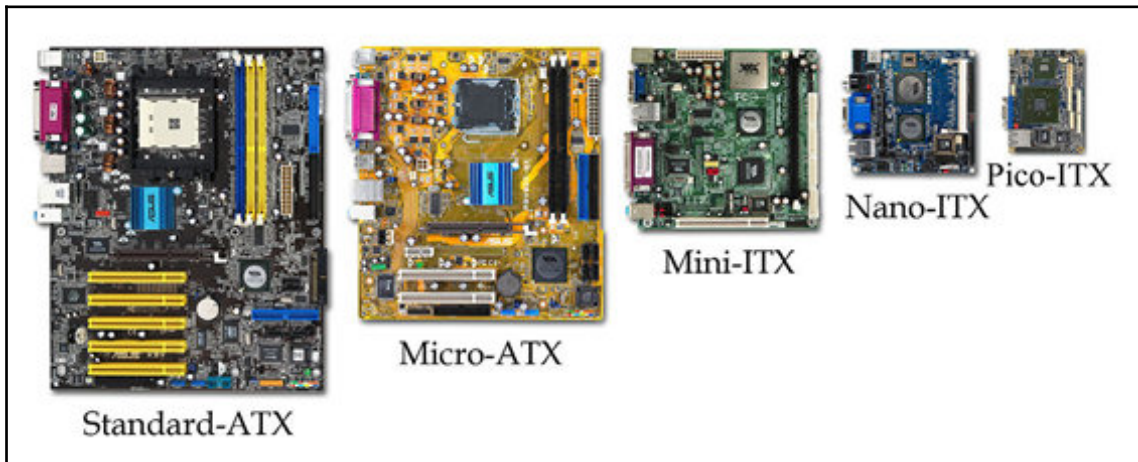
Anything virtual is like something, but it's not really it. So, a **virtual server** is like a server, without being one. Well, almost. A virtual server is a software-enabled logic object operating in the memory of a physical computer. A single physical computer can support several virtual servers, provided it has the hardware resources, primarily memory, to do so. As illustrated in the following diagram, a physical computer can support one, two, or even more virtual servers. In addition to the hardware and the appropriate device drivers, the virtualization layer, known as a **hypervisor**, provides direct support to the virtual servers, each of which occupies a shell in memory. Each virtual server can support numerous virtual machines, installed on the same host hardware or on other network computers:



A single physical computer can host one or more virtual servers

Form factors

For computing hardware, a **form factor** designates the dimensions, shape, and other physical characteristics of a computer case and its contents, including the power supply, mountings for internal storage devices, the motherboard and its mountings, RAM, expansion cards, the socket for the microprocessor, and other slots and mountings. The image that follows shows a variety of motherboard form factors, each of which have been made to fit inside a computer case of the same form factor. An ATX motherboard mounts in an ATX computer case, for example:



Motherboards in five different form factors
Image courtesy: VIA Technologies, Inc

Tower servers

It's common, especially in smaller networks and some home networks, to use a single tower computer as the network server. Tower computers, like the one shown in the following image, are inside a standing case or cabinet. Towers are commonly network servers. This means that a tower computer tends to have more different components and connectors than a **Small Office/Home Office (SOHO)** computer, even one in a tower case:

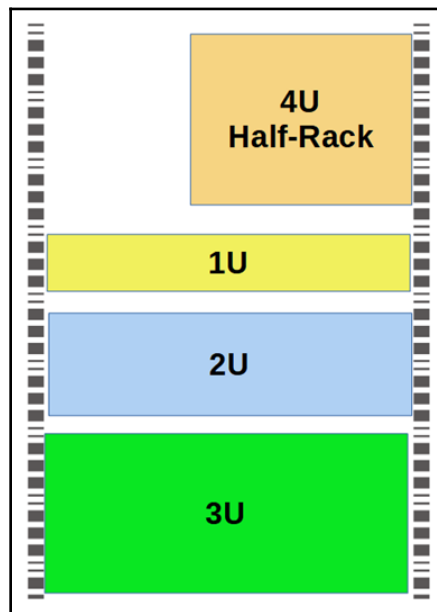


A tower network server

The upright and tall design of the tower case provides better cooling of the internal components. However, when tower computers are clustered, they take up more space and can create a complicated cabling arrangement. Plus, towers aren't the quietest computers around.

Rack mounts

The computer hardware on which a server runs fits into a slim chassis mounted in a rack system. The rack itself is typically either a two- or four-rail vertical structure. A server, or other rack-mounted device, attaches to the vertical rails using a rail kit, which consists of horizontally-mounted rails on which the device sits, and the fasteners to attach both to the vertical struts. The cabling that attaches to the rear of the rack-mounted device may install on a cable management arm, which helps to 1) organize the cabling on the device, and 2) allow the cabling to be out of the way when servicing or performing an upgrade on the unit. The height of a rack-mountable device is in rack units (U's). A rack unit is 1.75 inches (44.45 millimeters) tall. The size of a rack-mountable device is in the number of U's it will occupy in a vertical rack. The following diagram illustrates the relative sizes of a **1U**, **2U**, **3U**, and a **4U** half-rack mount. Servers are most commonly **1U** or **2U** in size:



Rack-mounted servers are sized in rack units (U's)

The **Electronics Industries Alliance (EIA)** has established a standard for rack systems of 42U in height and either 19-inches or 23-inches wide (48.3 cm to 58.4 cm). The depth of the rack can vary with the size of the overall structure or cabinet, as illustrated in the following image:



1U rack servers installed in a rack system
Image courtesy: 2018 FatCow Web Hosting

Blade technology

A blade server enclosure houses server blades, each of which is a scaled-down computer that fits into a slot in the rack-mountable blade enclosure chassis. The aim of the blade's design is to reduce the physical size, the number of direct interfaces, and the overall power usage of the server system. To do this, each blade has the components required to perform its internal processing. The cooling, power, networking, cabling, and management systems are a part of the blade enclosure or supplied by other devices in the rack mount or cabinet. As shown in the following image, a blade server enclosure supports several server blades. Each of the server blades installed in the blade server is, in fact, a discrete server that has a processor, memory, network adapter, and a **host bus adapter (HBA)**. It's common for a server blade to only support one application or service:



Multiple server blades in the rack-mounted chassis of a blade server
Image courtesy: Super Micro Computer, Inc

Server power systems

The power requirements of a server system, regardless of its form, are higher than needed by a desktop or laptop computer. Gilster's law (of everything computing) says:

"You never can tell, and it all depends."

This pretty much sums up the power and cooling systems for servers. The amount of power required by a server, measured in watts, is determined by the components installed and the devices attached to it. The same goes for cooling. The amount and kind of cooling required is a function of the heat generated by the components under power. However, whether the server is a stand alone computer or a blade server in a data center, the device power and cooling systems must provide a sufficient level of service to power and ventilate its components. The challenge in choosing and installing the right equipment for these tasks is anticipating growth in the systems or increased demand for these services. The de facto form factor standard for network servers is the standard-ATX (shown earlier, in the image published in the *Form factors* section). The ATX standard sets the form, fit, and function of a server's major components, primarily the motherboard, **power supply unit (PSU)**, and case. This ensures that these components are compatible and interoperable.

Electrical power

Before we get too deep into electrical systems and electricity, let's establish the meanings of a few terms you'll find in the discussion:

- **Current:** The flow or movement of an electrical charge
- **Resistance:** The properties of a wire that oppose the current flow
- **Amperes/Amps:** The rate of flow of an electrical current
- **Voltage:** The standard measure for the electrical force of a current
- **Watts:** The output rate of energy radiated, absorbed, or dissipated
- **Ground:** The protective measure with a conductive connection to the earth

These terms and their meanings are very important to a discussion on power supply and cooling systems. The following sections look at the different properties and applications of electrical power for a network server.

AC versus DC / 110V versus 230V

The primary function of a PSU is to convert an **alternating current (AC)** or a **direct current (DC)** into the low-voltage DC that powers the server's internal components. In North America, the predominant domestic electrical service is 120V AC (referred to as 110V), which has an actual range of 115V to 127V. The rest of the world (and some commercial data centers elsewhere, including the US) have a DC mains power standard of 230V, +/- 10%. In the US, the output voltage from a PSU conforms to the ATX standard of +3.3VDC, +5VDC, and +/- 12VDC, regardless of the electrical input. Because voltages can and do vary within a range, some systems use different numbers, although they designate similar systems. For example, the standard household voltage in the US is 120V AC. Notice that, in the following table, a 120V in the **Wye voltage** column can also supply 208V or 240V, depending on the circuit's connections:

Wye voltage	Delta voltage
120	208
120	240
230	400
240	415
277	480
347	600

Wye and Delta voltages common to the US

Wye and delta

There are two standard configurations used in electrical circuit diagrams—**wye** and **delta**. These names describe the approximate shape each has in a circuit diagram. A wye configuration connects a current-bearing line to a neutral in a sort of Y pattern. A delta configuration connects two current-bearing lines together to create a triangular shape. The easy way to remember these is that a wye circuit uses a neutral and a delta circuit doesn't.

Negative 48V

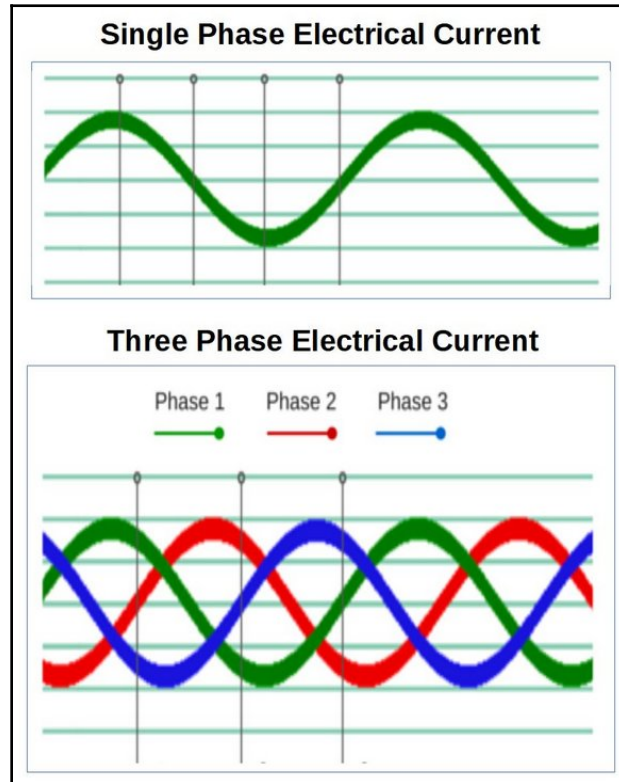
The voltage standard in telecommunications signaling, including wireless networking, is negative 48V power. All electrical circuits operate with plus (positive) and minus (negative) polarities, which yields one *live* side and one *ground* side. The 110V and 230V systems connect the grounding connection to the negative (minus) side. Negative 48V power connects its ground to the positive (plus) side. In case you're wondering why telecom systems use a negative voltage standard, it's because this voltage is safer for humans, especially those climbing up telephone poles.

One phase versus three phases

Electrical circuits transmit in one of three configurations—single-phase, split-phase, or triple-phase. Here are basic definitions for these terms:

- **Single-phase power:** A two-wire distribution system for AC, in which one wire carries the electrical current and the other wire is the neutral. The following diagram illustrates the wave form of a single-phase line.
- **Split-phase power:** A three-wire single-phase distribution system for AC, in which two wires carry electrical current and the third wire is the neutral. Split-phase distribution is common to homes and small business buildings.

- **Three-phase power:** A four-wire system, in which three overlapping wires carry an AC current. Each wire, and its current, are offset from the other wires, as shown in the following diagram. The fourth wire serves as the neutral. Three-phase is the transmission standard for larger electrical grids, industrial use, and data centers:



Single- and three-phase currents

PSU

Let's look at server power systems from the inside out, starting with the PSU mounted inside the server case. PSU modules are usually installed on the system case during manufacturing. However, because PSUs are the number one failure point in a computer system of any size, even those PSUs attached during manufacturing, are replaceable. The power needs of a server are dependent on its size and its installed or attached components. As the number of disk drives, network adapters, and RAM grows, the amount of electrical power a server needs also increases. A server's PSU needs to be more robust and operationally efficient than a standard PSU included in a typical desktop or portable computer. Because servers are essential to networks, they must be available, meaning that their power supply unit must continuously and consistently provide the voltage needed to power the server.

Wattage

Most of the power supplies available on the market, except those from the more *reputable* manufacturers, can have inferior components and claim overstated performance numbers. Although it is only one measure of a power supply's capabilities, many computer users rely on only the wattage rating of a PSU as the deciding factor for choosing one. Manufacturers understand that the wattage of a PSU is important to consumers, so they make sure it is very visible on the packaging and on the unit itself, as shown in the following image:



The product label on a 600-watt Cooler Master PSU
Image courtesy: Cooler Master Technology, Inc.

The 80-plus certification

The 80-plus program is a voluntary certification of computer PSUs based on their electrical efficiency. The *80* represents the quality threshold set as a minimum standard for operating efficiency at various load levels. Six levels of certification are available, based on the level of the unit's performance. Products that meet the requirements of each of the certification levels can include a badge on their packaging, marketing, and product labels, as shown in the subsequent image. The first level is the basic 80-plus certification (called *White*), which verifies that a PSU is 80 percent efficient under 20, 50, and 100 percent loads. The mid-level is *Gold*, which certifies units as having at least 87 percent efficiency at the three loads. The highest level is *Titanium* which verifies a unit with efficiencies above 90 percent under all load levels. This following image shows the various certifications of the 80-plus program:



The PSU certification levels of the 80-plus program

Selecting the right PSU

There are several factors that you should consider when selecting a power supply for a server, beyond how much wattage you need. The following lists the characteristics and capabilities that you should consider when selecting the PSU that's right for your server:

- **Wattage:** This number represents the power demands the server will make of the PSU as measured in watts. There are several wattage calculators on the web (see the following screenshot) that you can use to determine the total wattage your server needs. It's recommended that you add the wattage for any planned server or network expansions in your initial calculations:

OuterVision® Power Supply Calculator

Real Power Consumption

OuterVision Power Supply Calculator is the most accurate PC power consumption calculator available and is trusted by computer enthusiasts, PC hardware and power supply manufacturers across the Globe. Are you building a modern gaming PC, low power HTPC media server, or maybe you need to figure out power requirements for a rack in a data center? We've got you covered - OuterVision PSU Calculator will help you to select a suitable power supply unit and even Uninterruptible Power Supply (UPS) for your system. Building cryptocurrency mining rig? Check our Mining Rig Builder tool.

Video Card Overclock

Basic version of the OuterVision Power Supply Calculator allows users to quickly estimate power consumption with minimal selection of PC parts. On the other hand, our Expert, more advanced version of the **PSU Calculator** greatly extends the ability to select various PC parts and components, adds CPU and Graphics card overclocking, and allows consumers to calculate PC energy consumption, compare PSU efficiencies, and ultimately project energy cost.

Expert

Basic

Motherboard
Desktop

CPU
0 x Instant Search CPU
CPU count: by default it's 1 physical chip. Multicore CPU still counts as 1 physical CPU.
CPU Speed ☐ MHz
CPU Vcore ☐ V
CPU Utilization

Other Devices (USB, LED, Controllers, etc.)
0 x - Select
0 x - Select
0 x - Select
0 x - Select
Fans
0 x - Select
0 x - Select

A web-based interactive power supply calculator
Image courtesy: eXtreme Outer Vision, LLC

- **Connectors and modularity:** Make sure that the connectors provided with the PSU are compatible with the components that are to be attached and interconnected with it. A modular PSU has no built-in cables, only receptacles. This minimizes the amount of cabling on the PSU to only those that are necessary, which reduces clutter. Non-modular PSUs have different numbers and types of standard connectors. The connectors common to most current PSUs are:
 - ATX 24-pin or ATX 20+4-pin main power cable connector
 - 8-pin **entry-level power supply (EPS)** +12 volt
 - 4+4-pin +12 volt power cable connector
 - 6-pin **PCI express (PCIe)**
 - 8-pin PCIe

- 6+2-pin PCIe power cable connector
- 4-pin peripheral power cable connector
- SATA power cable connector
- **High-efficiency rating:** The 80-plus certification is an excellent guideline, but you should verify a PSU's efficiency rating as having been set from testing under actual load simulation. An 80% rating means that 20% of a PSU's energy (wattage) escapes as heat.
- **Rails:** In the context of a PSU, a rail is an output current of a single voltage. For example, an ATX PSU has one 3.3V rail, two for 5V (one each for +/- 5V), two for 12V (one each for +/- 12V), and a 5V standby rail.
- **External connection:** It may sound trivial, but without an external power cable with the appropriate connectors for your location (country), all your careful planning and selecting will have been for nothing. If you are in the US, use a **National Electrical Manufacturers Association (NEMA)** standards power cord and plugs. NEMA 5-15P connectors are the most commonplace in the US. In situations such as data centers with higher power ratings, a more robust connection can be wise. In these situations, a twist-lock connector such as the NEMA 5-30R locks the plug head into the electrical outlet.
- **Voltage switching:** Many PSUs include a voltage sensor that automatically detects the electrical current and switches to its voltage and mode. However, not all PSUs have this capability—some have a manual switch, and some have no switch at all and support only a single electrical service.

Redundancy

One way to ensure that a server or server cluster is fault-tolerant and provides high availability is to incorporate a redundant power supply system. Redundant power systems provide a safety net should the active power supply fail. In its most basic form, a redundant power supply has two separate PSUs that can provide power to the server together, alternatively, or with one PSU active and the other in standby mode. The image in this section shows a four-unit redundant power supply. The transition between the redundant units uses one of three configurations:

- **OR:** OR is a mathematical process that chooses between two (or more) options, as in *either or*. In this configuration, two PSUs can either share the power-load duties or one of the PSUs can be in standby mode. In either case, when a **metal-oxide-semiconductor field-effect transistor (MOSFET)** senses a drop in the power output of a unit, it switches to the standby PSU.

- **N+1:** The N+1 switch-over method is common for redundant systems that have three or more power supplies. In this arrangement, the +1 PSU is the standby unit and the N units share the power conversion operation.
- **OR of N+1:** This method is common to PSU blade systems. Each blade is a part of an N+1 grouping and interconnected to two or more power buses. Like the other redundancy configurations, each N+1 grouping can share power conversion or be in standby mode:



A four-unit redundant power supply
Image courtesy: Zippy Technology Corp.

System heat

All electronic devices produce heat. Some do so more than others, and cooling must reduce the heat effect to avoid failure or intermittent problems. The electronic components found inside a server (mostly on the motherboard) that produce significant heat include microprocessors, **graphics processing units (GPUs)**, chipsets, RAM, and **voltage regulator modules (VRMs)**. Of these, microprocessors (CPUs) and GPUs produce most of the heat inside a server. High heat conditions or a condition called **thermal stress** can affect the service life or operations of electronic components. Physics tells us that when things get hot, they expand, and when they cool off, they contract. Any electronic component that continuously goes through heat and cool cycles, meaning expansion and contraction, is stressed, which can lead to performance issues. The bigger the difference between how hot is and how cool is relates directly to the severity of the damage done. Most newer computer systems and processors now carry a rating called **thermal design power (TDP)**, which represents the amount of heat produced by the system or unit. The following table lists a few examples of the maximum TDP rating of several processors. This value indicates the amount of heat the cooling system must dissipate to keep the system running as it should. Although there are no standards for interpreting TDP, a lower value indicates the power usage and the heat produced is lower. TDP is only a general indicator of a system's cooling needs.

Processor	Maximum TDP
Intel Atom Z3740	4W
AMD A10 Micro-6700T	5W
Intel Core i3-5020U	15W
Intel Xeon E5-2630L v4	55W
AMD Ryzen 5 PRO 2600	65W
Intel Core i5-7600K	91W
AMD Ryzen 7 2700X	105W
Intel Core i9-7980XE	165W

A sampling of microprocessors and their TDP ratings

Cooling systems

Computers, regardless of shape, form, size, or application, require a cooling system. In desktops, towers, and some laptops, the cooling system is inside the case. For a blade server, the cooling may be in the blade cabinet, the rack cabinet, or in the computer room overall. Several different methods are available to cool the internal components of a computer system. Some of these systems are legacy and some are new. For the Server+ exam, you should understand the cooling systems described in the following sections.

Air cooling and air flow

A basic air-cooling system is typically a default system built into a computer's case and internal components. In its simplest form, an air-cooling system consists of a **heat sink**, **thermal paste**, and the computer's **case fan**. A heat sink attaches directly to a CPU with a small amount of thermal paste between the two to provide a thermal conductor. Air gaps between the heat sink and the CPU can act as thermal insulators, so the thermal paste eliminates this possibility. The heat sink is a ribbed metal extrusion that extends the surface of the CPU to allow more air to dissipate the heat. Air flow from one or more case fans moves across the fins of the heat sink to carry the heat away. This type of heat dissipation is known as **passive cooling**. Adding baffles, or air flow defectors, to passive cooling systems to specifically direct the air flow can enhance the effectiveness of the air-cooling system.

Some upper-end cases include not only multiple fans (two or more case fans and a graphics card fan), but a baffle system that directs the airflow to the CPU and other hot spots in the case. The alternative to cooling a computer with air flow is **liquid cooling**, which uses a coolant to pull the heat away from the CPU. Liquid cooling applies the thermodynamic principle that heat from a warm object will move to a cooler object. A CPU liquid cooling system works like the cooling system in an automobile. A liquid coolant, in this case distilled water, is pumped through an attachment on the CPU. The coolness of the water draws the heat of the CPU away and dissipates in the air flow.

The following image shows the radiator (on the left) and the CPU attachment (on the right):



A CPU liquid cooling system
Image courtesy: Asetek

Summary

In this chapter, you learned that the software running on a computer establishes its role as the server. A server can provide a variety of services to a network's clients, including: application servers, file servers, mail servers, messaging servers, network servers, print servers, RRAS, and web servers. The different types of application servers are LAN application servers: query-based application servers, and application/web servers. A file server is either dedicated or non-dedicated. Network services servers provide core services on the OSI application layer. Although the NOS provides many network services and protocols, services such as the DNS, DHCP, IM, VoIP, and NTP may come from a network services server. Proxy servers are intermediate network servers that accept, fulfill, filter, and forward requests from remote clients. The types of proxy servers are: gateway, internet-facing, open, internal-facing, and reverse. Two other server types that you may encounter on the Server+ exam are RRAS, which is a type of network services server that provides firewall, router, and remote access services, and a virtual server that is a software-enabled logic object operating in the memory of a physical computer. Servers and computers, in general, conform to a particular form factor standard. A form factor designates the dimensions, shape, physical characteristics, and power and cooling performance of a computer case along with the power supply, motherboard, RAM, and other components installed inside the system case.

The form factor standard for servers is ATX. Tower computers have a standing case or cabinet, as do desktop and notebook designs. Rack-mountable servers are commonly 1U or 2U in height with their width matching the width of the rack or cabinet. A blade server enclosure is a cabinet of sorts that houses server blades, each of which is a server itself. A PSU converts domestic AC or DC (110V or 230V) into +3.3VDC, +5VDC, and +/- 12VDC and telecommunication devices use -48V power. Factors for consideration when selecting a PSU for a server include wattage, connectors and modularity, efficiency rating, rails, external connection, and voltage switching. Redundant power supplies have two or more PSUs that can provide power in combination, alternatively, or with actives and standbys. The components inside a server, especially CPUs, GPUs, chipsets, RAM, and VRMs, produce heat that can cause thermal stress on some components. Thermal stress can, and usually does, affect electronic components.

Two types of cooling systems are common in servers: passive-cooling (air-cooling) and liquid-cooling systems. A passive system includes a heat sink, thermal paste, and a case fan. Liquid-cooling systems pump distilled water through piping and a radiator, where the heat absorbed by the water cools before being recycled. In the next chapter, we'll take a look at the key internal components of a server, the ones that use power and need cooling. This includes the CPU, RAM, the various bus structures, BIOS/UEFI and CMOS. These internal server components are extremely important to the overall operations of a server, which is why they are emphasized in the Server+ exam.

Questions

1. Which of the following is not a common type of application server?
 1. LAN application server
 2. Web server
 3. DHCP server
 4. Query-based application server
2. Which of the following is true about file servers?
 1. A file server can only be a dedicated server
 2. A file server can be either dedicated or non-dedicated
 3. A file server cannot be virtualized
 4. A file server and a database server are essentially the same

3. What type of server hosts protocols such as DNS, DHCP, VoIP, and NTP?
 1. Proxy server
 2. File server
 3. Messaging server
 4. Network services server
4. You have the task of installing and configuring a proxy server on your in-house network to reduce traffic from the LAN to the WAN. Which of the following general configurations should you use?
 1. Gateway
 2. Internet-facing
 3. Internal-facing
 4. Open
 5. Reverse
5. True or false: A virtual server is any server that hosts virtual private networks.
 1. True
 2. False
6. Which of the following is the de facto form factor standard for network servers?
 1. LTX
 2. AT
 3. ATX
 4. Mini-ATX
7. What is the standard height dimension of a rack unit or "U"?
 1. 1.75 inches
 2. 2.75 inches
 3. 3.50 inches
 4. 50 millimeters

8. What are the voltages produced by an ATX PSU for the internal components of server hardware? Choose all that apply?
1. +3.3VDC
 2. 110VAC
 3. +5VDC
 4. +/- 12VDC
 5. -48VDC
 6. 230VDC
9. A server's air-cooling system is what type of system?
1. Active
 2. Passive
 3. Liquid
 4. Baffled

2

Server Internals

In the previous chapter, we looked at the power and cooling systems of a server. Now, let's look at the key internal components of a server. You know, those powered and cooled. This includes the CPU, RAM, the various bus structures, and the **Basic Input Output (BIOS)/Unified Extensible Firmware Interface (UEFI)** and **complementary metal-oxide-semiconductor (CMOS)**. We'll also look at BIOS/UEFI server firmware. In this chapter, we will cover the following topics:

- **Central processing unit (CPU)**
- Main memory
- Buses, channels, and expansion slots
- BIOS/UEFI configuration

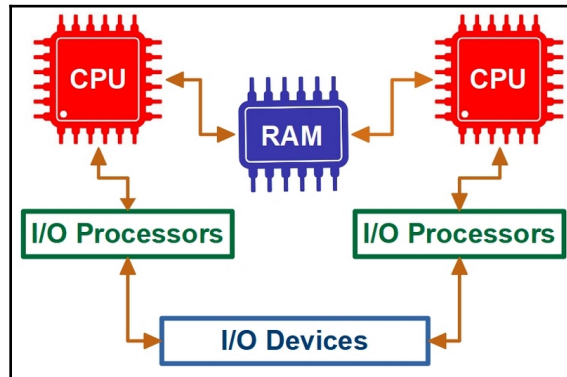
CPUs

Known by many names, the CPU (also known as the microprocessor, processor, brains, and other names), is the electronic component that runs program instructions, performs arithmetic functions, and controls the movement of data and the input and output functions of peripheral devices attached to or installed on the computer. But then, you knew all of that.

For the Server+ exam, what you need to know is less about how a CPU functions and more about its characteristics, including its mounting socket, clock speeds, cores, stepping, and more. So, let's get on with it.

Multiprocessors

A multiprocessor environment is a single computer system that has two or more integrated CPUs. The CPUs share the computer's memory, bus, and other resources. The CPUs work cooperatively to execute program instructions in series, with one CPU performing an instruction and the other CPUs simultaneously performing another. The purpose and result of this process is that the computer runs faster than a single processor (uniprocessor):



A simplified view of a two-processor multiprocessing system

Symmetrical Multiprocessing (SMP) versus Asymmetrical Multiprocessing (ASMP)

The CPUs in a multiprocessor system can be set up for either symmetrical or asymmetrical multiprocessing. In SMP, the CPUs equally share the operating system, main memory, bus, and input/output drivers and devices. The goal of an SMP system is to balance the computing loads between the processors and speed up processing. However, it's recommended that an SMP includes no more than 16 processors.

In ASMP, one CPU is the master and all other CPUs are slaves. The master handles the operating system tasks and assigns process requests to the slaves. Slaves can be general-purpose or dedicated to a specific processing responsibility. The master processor controls the functions of the system with the assistance, when necessary, of the slaves.

SIMD, MISD, and MIMD

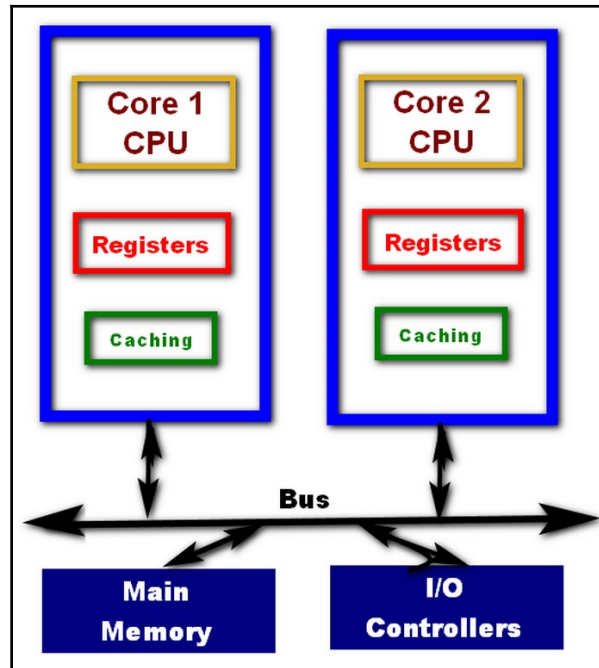
Multiple processor systems can perform *parallel processing*, in which each of its processors executes the same instruction or a unique set of instructions on either the same dataset or on several datasets. Today's multiprocessing computers support different flavors of parallel processing, with the three primary forms being:

- **Single-instruction, multiple-data (SIMD):** Multiple processors execute the same instruction on different blocks of a data source. SIMD speeds up multimedia processing.
- **Multiple-instruction, single-data (MISD):** Multiple processors execute different instructions on a single data source. MISD computing is not common because this mode of parallel processing is usually very specific to a problem.
- **Multiple-instruction, multiple-data (MIMD):** Multiple processors execute different instructions on different blocks of a data source. MIMD is what most people think of as parallel computing.

Multiple core processing

Where multiprocessing involves multiple microprocessors as parts of a single unit, multicore processing is one microprocessor that contains multiple processors or cores. Each core is a microprocessor and can process a different stream of instructions than any of the other cores on the same **integrated circuit (IC)** chip.

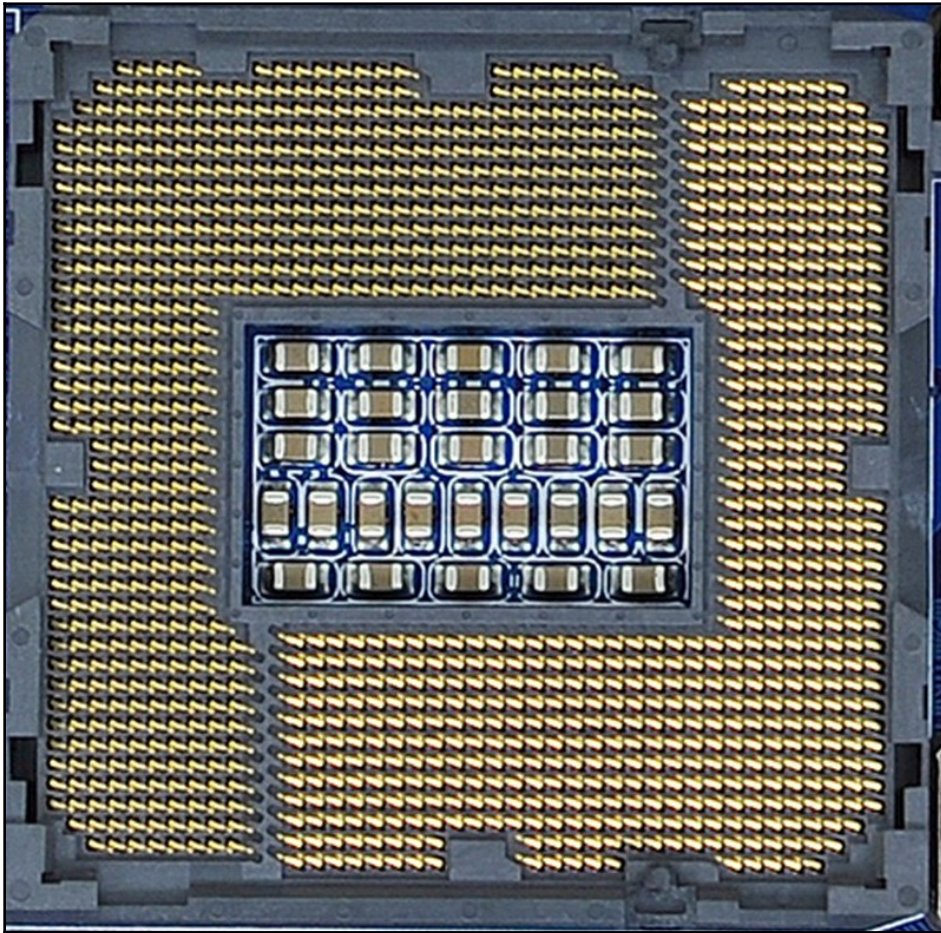
If you have a quad-core processor, you have four separate processors built into your CPU. This means that you can be checking your email, watching a video, working on your budget in a spreadsheet, and listening to a music stream all at the same time. Each of these actions runs separately on its own core:



A simplified view of the structure of a dual-core microprocessor

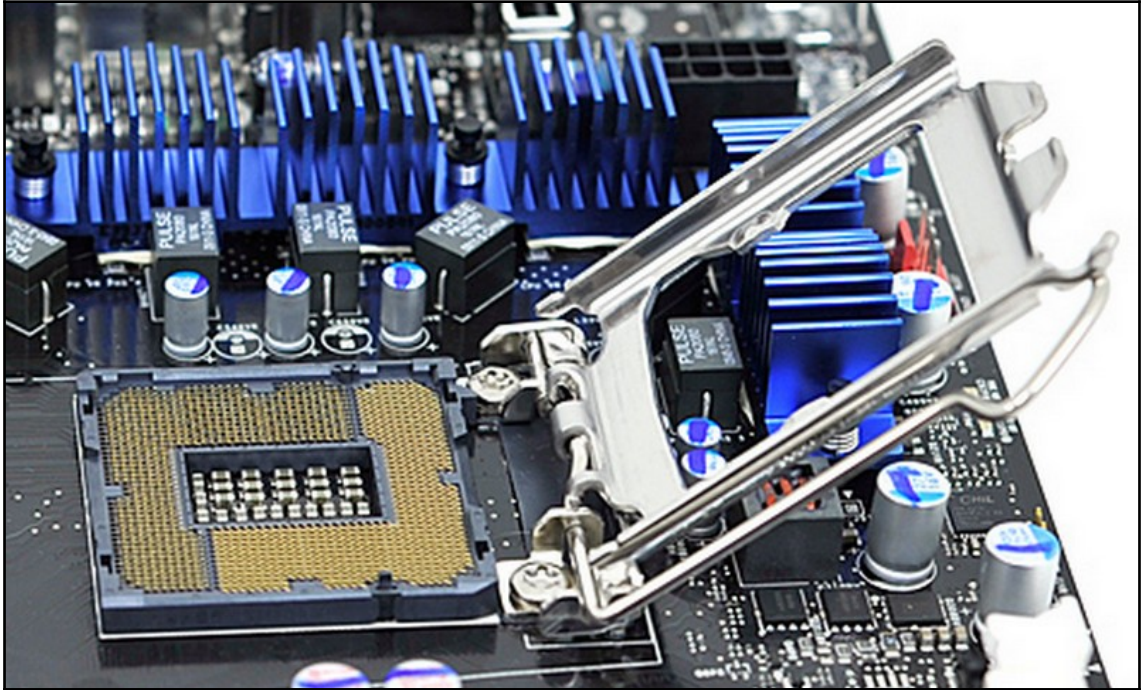
CPU packages and sockets

The shape and construction of a CPU is its packaging. Virtually all server CPUs are in **Land Grid Array (LGA)** packaging. In this packaging, the mounting pins are a part of the socket and the CPU has receiving holes (ports) to fit over each pin. The following image shows the ports on the underside of a CPU in an LGA package:



The underside of a CPU showing the mounting ports of LGA packaging
Image courtesy: AnandTech

There are nearly as many different types of CPU sockets as there are types of CPUs. The following image shows an LGA socket on a motherboard. Notice the locking arm, which secures the CPU in the socket:



An LGA socket mounting on a motherboard
Image courtesy: AnandTech

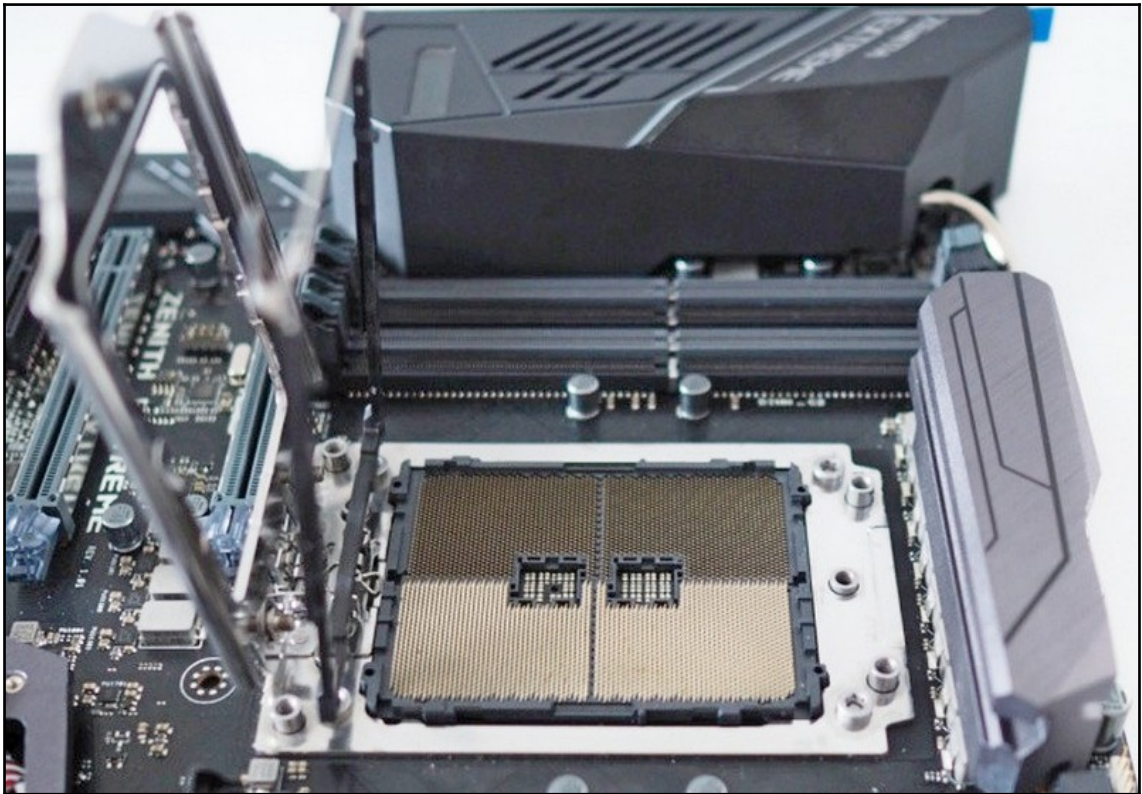
The following table lists a sampling of the sockets designed for use with a variety of processors:

Socket	Introduced	Compatible CPU	Package	Pins	Maximum speed
LGA 771/Socket J	2006	Intel Xeon	LGA	771	1600 MHz
Socket 1207FX	2006	AMD Athlon 64 FX	LGA	1207	2000 MHz
Socket G34	2010	AMD Opteron (6000 series)	LGA	1974	3200 MHz
LGA 1248	2010	Intel Itanium 9300-series	LGA	1248	1.47 GHz
LGA 1567/Socket LS	2010	Intel Xeon 6500/7500-series	LGA	1567	2.66 GHz

Socket SP3	2017	AMD Epyc	LGA	4094	2.9 GHz
LGA 2066/Socket R4	2017	Intel Skylake	LGA	2066	4.2 GHz

Common server sockets and the CPU they support

For instance, the LGA 1564 or Socket LS socket, introduced in 2010, was compatible with the Intel Xeon 6500 series CPUs. The Socket SP3, shown in the following image, holds the AMD EPYC processor. It's common for a paired new processor and new socket to be released together. However, a new CPU model or version may fit into an existing socket standard:



An AMD SP3 socket
Image courtesy: AnandTech

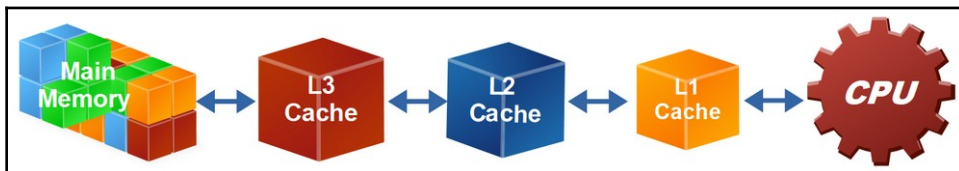
Cache memory

A cache (pronounced *cash*), according to dictionary folks, is anything you *store away in hiding for future use*. In a computer system, there are several types of cache:

- **Browser cache:** When you download a web page, the browser stores some or all of the page's content, especially the parts that aren't likely to change, such as images, headlines, text, scripts, and so on. This eliminates the need to download this content again as you page through the website.
- **Disk cache:** Many larger hard disk drives include a small amount of RAM, which functions as a cache. For example, a 1 TB hard disk drive has 32 MB of disk cache to improve the disk's I/O performance.
- **Memory cache:** Not to be confused with the processor cache, application software that requires a large amount of data, such as a graphics editor, will create a cache in RAM to reduce I/O operations and speed up processing.
- **Processor cache:** In the context of CPUs and servers, this type of cache (also known as CPU memory or cache memory) provides data and instructions to the CPU, eliminating the need to access the slower main memory. There is more on this in the following section.

CPU cache memory

A CPU has a small amount of memory reserved for its use. This memory, which is known as CPU memory or cache memory, consists of multiple levels of **static RAM (SRAM)**, most typically three levels, as shown in the following diagram in this section. The purpose of this cache of memory is to provide data and instructions to the CPU at a much faster rate than if these inputs were to come from main memory DRAM. The CPU's caching system also tries to guess ahead as to what the CPU will request next and, based on current processing, is right more often than not, which speeds things up more:



Caching provides a fast, multilevel buffer for the CPU

Cache memories have a couple of important characteristics:

- **Temporal (time) locality:** Cache memory holds images, data, and instructions that are unchanging, which eliminates the need to fetch them from their source or main memory.
- **Spatial (sequence) locality:** Commonly, the next CPU request for an instruction or data block is already in a cache. If the CPU requests something not in the cache, a *cache miss* occurs. If the CPU requests a cached item, it's a *cache hit*, which is a more common event.

CPU cache memory levels

As illustrated in the *CPU cache memory* section, most CPU systems include three levels of cache memory. **Level 3 (L3)** cache is slow, but faster than main memory, and is the larger in size of the three levels. **Level 2 (L2)** is faster than L3, but smaller in size. And **Level 1 (L1)** is the fastest of all three levels and the smallest in size.

As shown previously, in the *CPU cache memory* section, instructions and data pass from main memory to L3, which is large enough to contain enough of the active program or data to predictably include the next request from L2 and above. Multi-core processors typically share L3 cache memory. L2 applies spatial locality to predict what L1 will ask for and requests a block from L3, which contains the predicted items. L1 continues this process by predicting more precisely the item it anticipates the CPU to request next and asks L2 for that item.

Write-back/write-through cache

Cache memory systems take advantage of the time when the CPU may be idle—we're talking about milliseconds here—to write data passed down to it from the CPU or a higher cache level. The data is passed to main memory or directly to a secondary storage device. This action is called **write-back**.

In some instances, the CPU writes data directly to main memory or a storage device. When this happens, the CPU also passes the data to cache memory. This action is called **write-through**. The extra write step slows the process, but if the next CPU request results in a cache hit, the system realizes a time benefit.

Advanced RISC Machine (ARM) servers

It seems like the Intel x86 family of processors has been the heart and brains of networking from the beginning. That's not a bad thing. It's just that there may be a new kid on the block—an ARM. A **Reduced Instruction Set Computer (RISC)** is a specialized technology that involves the use of a more powerful instruction set with fewer commands. Cell phones and other mobile devices rely on ARM processors for many of their memory-and storage-related functions.

Because of its reduced instruction set, an ARM processor, which tends to be smaller in size than conventional processors, is readily adapted to a processor core. Servers that now run on processors with up to 12 cores can run equally as well on an ARM, which may involve dozens of less sophisticated processors that can share the computing task. While 12 processors result in substantial server, in x86 server, the processors could still become a bottleneck, depending on its loads.

CPU multiplier

CPUs are very fast in comparison to their support systems. Their speed or frequency (measured in MHz) is a function of the computer's **front-side bus (FSB)**, which is the bus that connects the CPU with the northbridge (memory controller hub) of the chipset. To set its internal frequency, the CPU applies a multiplier to the actual frequency of the FSB.

The multiplier is a ratio, referred to as the CPU multiplier, clock multiplier, or the clock ratio, that's applied to the frequency of the FSB to determine and set the internal frequency of the CPU. For a 100 MHz CPU with a 40 times (40X) multiplier defined in its BIOS, the internal clock rate of the CPU is 4.0 GHz.

The CPU multiplier is the key to overclocking or under-clocking a computer. To overclock a computer, common in gaming, raise the value of the multiplier. To under-clock a computer, which means to slow it down, lower the multiplier's value. Overclocking typically requires additional cooling capacity and under-clocking saves the battery in a portable device.

CPU stepping

CPU stepping refers to revisions and revision numbers applied to a CPU to fix bugs or improve functionality. When a manufacturer releases a new CPU, its revision level is generally zero or a variation of zero, such as **A0**. For example, Intel, which calls revision releases, specification updates, and raises the revision level or step. AMD releases *revision numbers*, each of which has a higher revision number than the last.

Main memory

Main memory, primary memory, primary storage, or RAM, known by all of these names, it is like the heart of a computer. If the CPU is the brain, RAM provides the circulation of instructions, data, addresses, and the status of everything flowing through the computer.

RAM

Random-access memory or RAM allows data, instructions, addresses, or state information to be located and accessed directly. Each byte in RAM is addressable and accessed by itself or in a block of data. Its random-access capability contributes to the overall speed of the computer. The types of RAM are as follows:

- **Dynamic RAM (DRAM)** is electrically volatile and must receive a refresher electrical charge periodically in order to hold its status (positive or negative). RAM consists of millions of single electronic components (transistors), each of which is able to store only a single value, which we represent with one of the two binary values—0 or 1. RAM consists of millions of transistors that can hold either a positive or negative charge.
- **Static RAM (SRAM)** is non-volatile, which means it doesn't require refreshing, and is able to hold its electrical charge as long as the computer has power. After the power is off, SRAM loses its stored charges along with the data it represents.

Double Data Rate (DDR) RAM

Like processors, RAM has moved through some evolutionary stages as well. Here's a bit of the timeline for RAM:

- **Dynamic RAM (DRAM):** The one characteristic of DRAM that led to its demise was that it operated independently of the processor, which could introduce latency when either component was waiting on the other.
- **Synchronous DRAM (SDRAM):** SDRAM coordinated its operations with control signals on the system bus, which allowed it to stay one step ahead. However, SDRAM was **single data-rate (SDR)**, meaning in a system clock cycle, it could only read or write once, on either the start or end of a clock cycle. With processors becoming more complicated and especially faster, SDRAM proved to be too slow.

- **DDR-SDRAM:** DDR memory improved on the SDR technology by allowing data I/O on both the start and end of the clock cycle, doubling the data rate.
- **DDR2/DDR3:** These two technologies incorporated internal clocks that operated at one half (DDR2) and one fourth (DDR3) of the original DDR memory. DDR3 improved memory size to as much as 8 GB, increased the data rate to as much as 2,133 Mbps, and reduced power consumption to 1.5V.
- **DDR4-SDRAM:** DDR, 4th generation (DDR4) SDRAM is the current evolution of the DDR technology. The main improvements of DDR4 over DDR3 are double the memory upper limit (up to 16 GB), a data rate of up to 3,200 Mbps, and uses less power (1.2V).

RAM packaging

The packaging for the RAM modules installed on home and desktop computers is in the form of a slot-mounted IC board with an edge connector with varying amounts of memory (in megabytes) and numbers of contacts (pins). Since the development of the DDR specification, memory boards have been in the form of **dual in-line memory modules (DIMMs)**, **small outline dual in-line memory modules (SO-DIMMs)**, and MicroDIMMs:



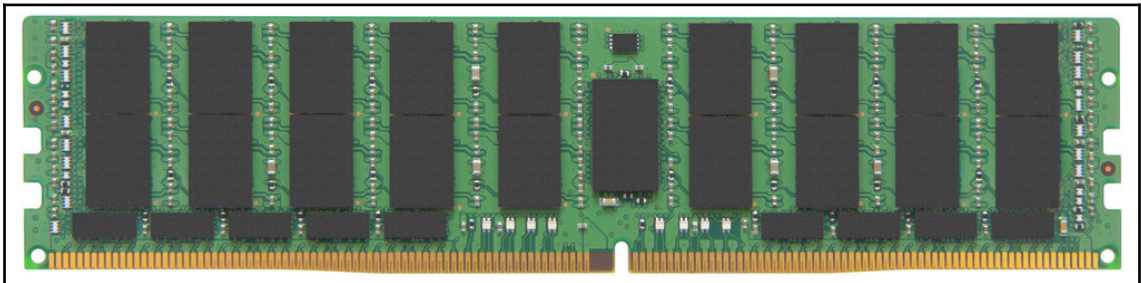
DDR4 SO-DIMM memory card
Image courtesy: Kingston Technologies, Inc.

The following table lists the number of pins in the edge connector of the various memory modules:

DIMM module	SDR	DDR	DDR2	DDR3	DDR4
Dual In-line Memory Module (DIMM)	168	184	240	240	288
SO-DIMM	100	200	200	204	256
MicroDIMM	172	214	214	214	-
Registered DIMM (RDIMM)	-	184	240	240	288
Load-reduced DIMM (LRDIMM)	-	184	240	240	288

Pin counts on different DIMM boards

However, in addition to the latest processors, manufacturers have also released new versions of DIMMs. Two of the newer DIMM versions are the RDIMM and the LRDIMM. An RDIMM register buffers between the memory controller and the DRAM on the DIMM. This allows a server to host more RDIMMs, but it can also increase power consumption and additional latency. An LRDIMM uses a memory buffer to combine the electrical signals of its DRAM into a single signal, which allows the DIMM to host up to eight DRAM units. However, the increase in power usage and latency is even higher than that of the RDIMM:



A 1.32 GB LRDIMM
Image courtesy: Kingston Technologies, Inc.

Memory timing

The general rule of thumb in RAM is *lower is better*, but *lower what*? Memory doesn't exactly operate in the way we often describe it. What we believe is that an operation involves a request for data from the CPU, the memory controller locating it, and sending the requested data upward. A simple three-step process, *right*? Well, no. The process involved for the CPU to request data or instructions from memory is more involved than that. In any case, the less time it takes for a memory I/O request to take place, the faster the overall speed of the computer will be.

When you shop for server memory, you may notice a set of numeric values that looks something like this: 9-10-11-24. In the acronyms for these measurements, the lowercase **t** in **tRCD**, **tRP**, and **tRAS** stands for *time*. This string of numbers represents the four primary measurements in the memory timings:

- **CAS latency (CL)**: The first number in the memory timings represents the **column address strobe (CAS)** latency or **CL**, which is the amount of time, in **nanoseconds (ns)**, it takes to receive and fulfill a request for data. The following table shows examples of CL for different memory technologies. Notice that the actual latency is a function of clock cycles multiplied by the CL.

Memory technology	Clock cycles (ns)	CL	Actual latency (ns)
SDR	7.50	3	22.50
DDR	5.00	3	15.00
DDR2	2.50	6	15.00
DDR3	1.25	11	13.75
DDR4	0.75	18	13.50

Examples of CL in various memory technologies

- **RAS to CAS delay (tRCD)**: The arrangement of data in memory is something like a spreadsheet with rows and columns. The second number in the memory timings is the **row address strobe (RAS)**, which is the time required to locate the row on which requested data is located. The **column address strobe (CAS)** then represents the time to move to the corresponding column. In other words, if the requested data is in location C15, RAS is the time to move to row 15 and CAS is the time required to move to column C and the data at C15.

- **RAS precharge (tRP):** Although it may sound like almost the opposite of its function, the RAS precharge releases the active row in memory and the tRP, the third value in the memory timings, is the time required to do so.
- **Row active time (tRAS):** Also known as **Active to Precharge Delay**, the fourth value, tRAS, is the time required to close an active row and to open a new row. You may also see tRAS described as the time required to complete an instruction and request and receive the next one.

Error-correction code (ECC) versus non-ECC

ECC memory is common in computers that process high-value or confidential data, such as computers supporting one or more servers. What differentiates ECC memory from non-ECC memory is that ECC memory modules include a dedicated memory unit that provides parity and error-correction to the other memory units of the module. ECC memory provides assurance of data integrity. Before ECC technology, the method used for error-correction was parity, either even or odd. Non-ECC memory has no error-correcting capability and acts only to fulfill requests from the processor.

Dual channel memory

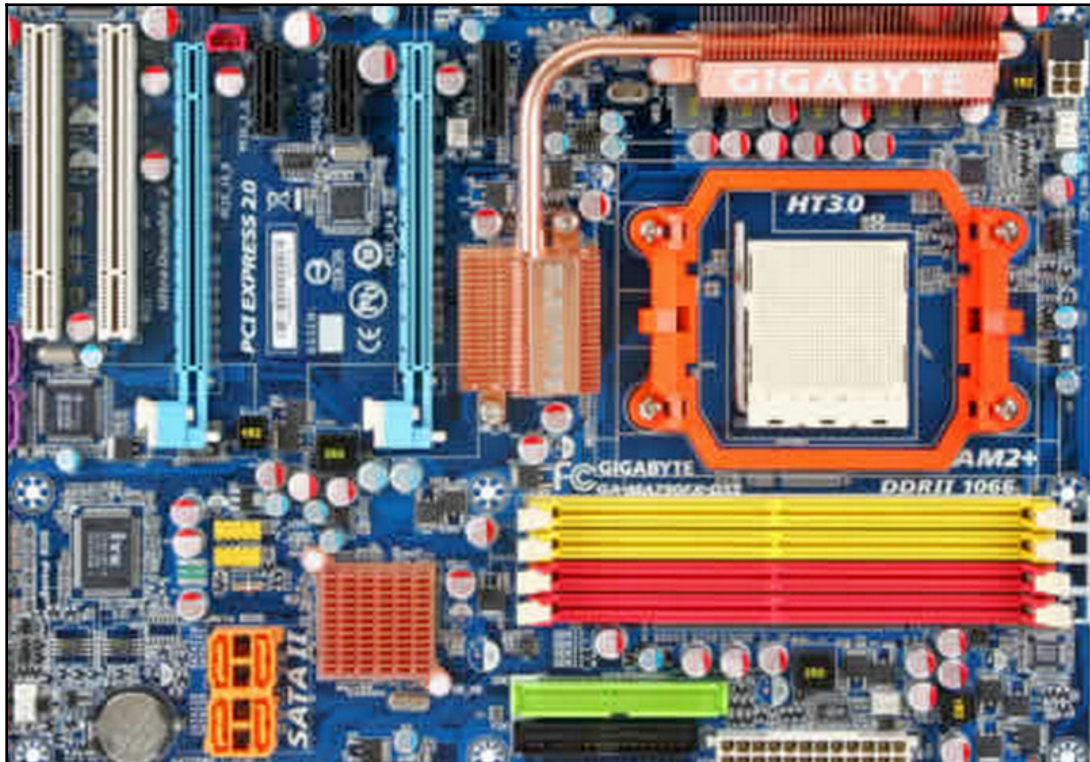
You may have heard about the need for memory pairing. You've heard about color coding memory slots. It seems important, but *what does either or both have to do with upgrading memory in a server?*

Many newer motherboards have color-coded dual channel RAM slots. When these slots contain matching memory modules, the system is able to move data to and from both memory boards simultaneously, which reduces the access time. Realistically though, matched memory that facilitates dual-channel mode does improve access time. The improvement isn't that significant over single-channel mode and unmatched memory pairs. However, performance is faster when you install matched memory pairs.

Color-coded RAM slots

Although there may be differences in the specific colors of the memory slots a manufacturer places on a motherboard, the basic rule of thumb is that memory slots are color-matched in pairs, threes, fours, or whatever the future holds, based on the memory channel technology of the motherboard. For example, in dual channel mode, paired memory slots will both be yellow, orange, or red to indicate the slots for a matching RAM pair.

The same holds true for the threes and fours, and so on:



Color-coded memory slots on a motherboard
Image courtesy: Gigabyte Technology Co., Ltd

Buses, channels, and expansion slots

The components on the motherboard of a computer, as well as the devices connected to it, pass data and instructions to each other constantly. Just like any other form of communication, there must be a medium connecting them. This medium, in this case a *bus channel*, provides a pathway on which data, addresses, and commands travel.

In a computer, there are two general categories of bus channels: internal bus and external bus. The *internal bus* is solely on the motherboard and it's used by the motherboard's components to pass data and instructions. The *external bus* provides a means for peripherals and expansion components to communicate with the components on the motherboard.

The term *bus* comes from the Latin word *omnibus*, which translates to *for everyone*. The connecting links in a computer carry *everything*, so they became buses. Bus shouldn't be confused with *buss*, which is another way to say kiss.

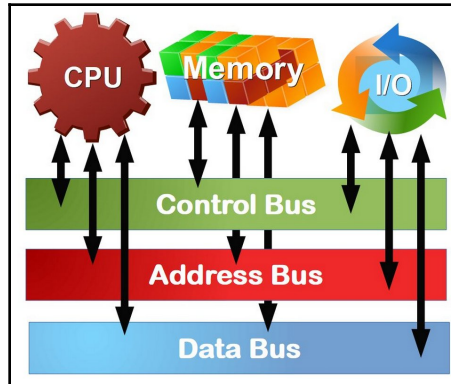
A bus structure consists of three coordinating bus lines (see the diagram in the following *Bus width* section):

- **Control bus:** The CPU transmits commands and instructions to a component or device to which it needs to send or from which it needs to receive, data or perform a command. This is a dedicated one-way bus that carries the command needed on the contents located at the address on the address bus.
- **Address bus:** The CPU transmits the address of data requiring a service to a component or device. The address bus is also a one-way bus.
- **Data bus:** The CPU sends or receives data from memory or a device controller.

Bus width

A bus channel consists of a number of wire traces, each of which carries a signal (bit) to a device pin. Bus channel's traces are typically in multiples of eight, as in 8, 16, 24, 32, 64, and so on, with the number of traces controlling the amount of memory possible on the system. The number of bits or traces in a bus channel limits the high-end address value. For example, an 8-bit address bus can only address memory cells up to 28 or 256. If the bus has 32 bits, the highest address in memory that it can reference is 232 or 4,294,967,296 (a bit over 4 GB or 500 MB). Obviously, the wider a bus channel is (that is, the more bits wide), the higher the number that it can represent in binary.

The descriptions and specifications of a microprocessor include the bus width with which it's compatible. For example, the Intel Xeon and the AMD Athlon are popular server CPUs, and both are 64-bit processors:



The components of the system bus structure

Peripheral Component Interconnect (PCI) bus

Not every motherboard comes complete with your favorite peripherals, controllers, or interfaces, in fact, few ever will. To help you get over this, motherboards include a few expansion slots into which you can insert the interface cards for the devices and functions you need. Over the years of computing and servers, a number of expansion card standards have come and gone, with new standards or extended standards taking their place. The *Color-coded RAM slots* section, earlier in the chapter, showed a motherboard that had two sets of expansion card slots (blue and white).

For the Server+ exam, the list of bus channel and expansion slot technologies you need to know in detail is short and from the same family: PCI—also known as **PCI conventional**, **PCI-extended (PCI-X)**, and **PCI-Express (PCI-e)**. The PCI slots and expansion cards of these standards differ in a number of areas, but the most significant differences are in their slot and card heights and lengths, bus width, and signal voltage.

PCI size and fit standards

All PCI and their variations apply the same physical board format standard. The standards for PCI boards specify four size and fit formats: full-height, low-profile, full-length, and half-length. There are a few exceptions though:

- A full-height expansion card won't fit a low-profile slot
- A half-height expansion card will fit a full-length slot
- A half-height expansion card won't fit a full-height slot
- A full-length expansion card won't fit a half-length slot

Okay, so the last one's a no-brainer! The best practice is to match PCI expansion cards to their corresponding PCI slots.

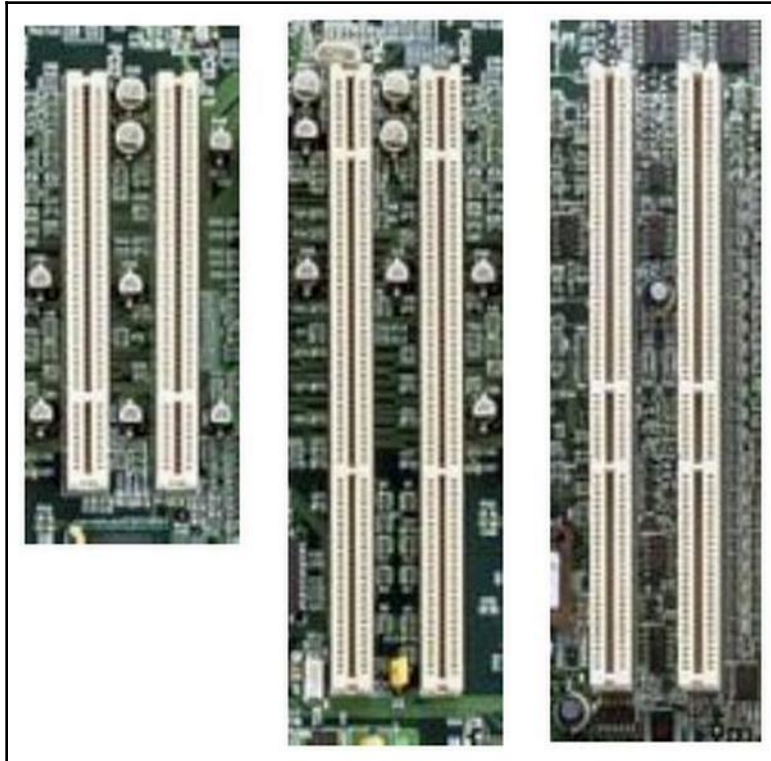
PCI conventional

The *bus width* of a bus channel is the number of traces in the channel. In general, the PCI bus and its variations, the bus width is either 32 or 64 traces. At the risk of being repetitive, a trace in a bus channel carries one bit of the data transmitted on the bus. The *signal voltage* of the PCI standard is one of either 5V or 3.3V.

The more common variations of the PCI conventional standard (see the following diagram) are:

- **32-bit PCI—5V signal voltage:** This is a common PCI slot on desktop computer motherboards. Its clock speed is 33 MHz and its DTR is a maximum of 1 Gbps.
- **64-bit PCI—5V signal voltage:** This is typically a server and dual processor motherboard expansion slot with a clock speed of 33 MHz and a maximum DTR of 2.1 Gbps.

- **64-bit PCI—3.3V signal voltage:** Also known as PCI-X, this PCI expansion slot is common to server motherboards. Its maximum clock speed ranges from 66 to 533 MHz and its maximum DTR is 4.3 Gbps (at 533 MHz):



Left to right: PCI 32-bit 33 MHz slots, PCI 64-bit 66 MHz slots, and 64-bit 33 MHz slots

The following table lists the maximum **data transfer rate (DTR)** for each of the PCI bus standards:

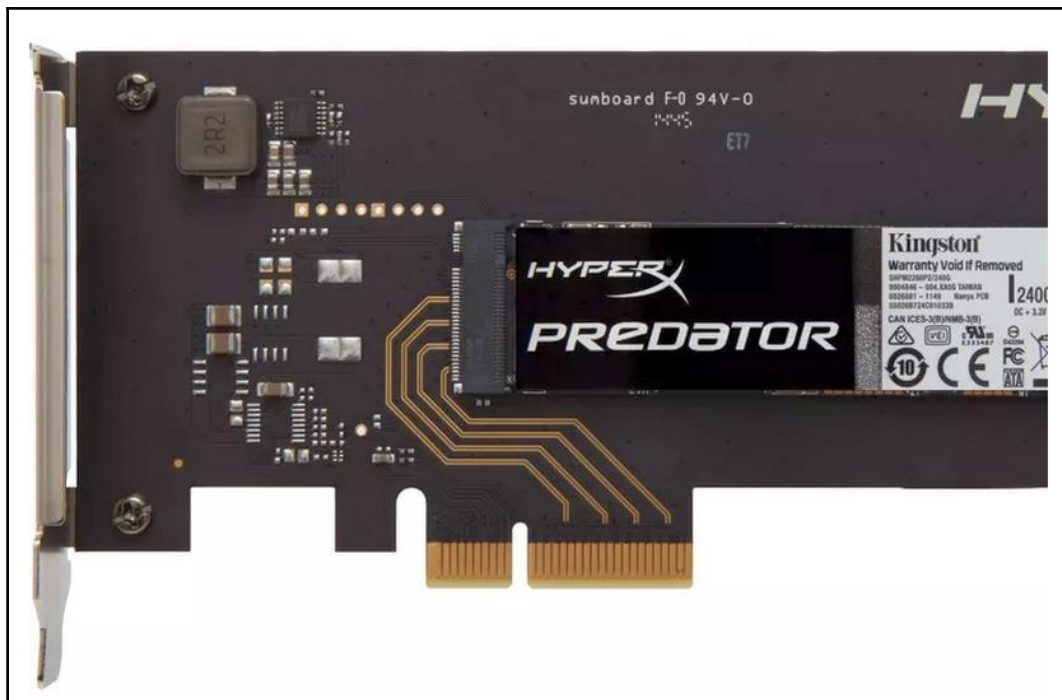
Technology	Maximum DTR (Mbps)	
Parallel	Serial (Half-duplex)	Serial (Full-duplex)
PCI (Conventional)	132	
PCI Express 2x	500	1000
PCI Express 4x	1000	2000

PCI Express 8x	2000	4000
PCI Express 16x	4000	8000
PCI Express 32x	8000	16000

Data transfer rates (DTRs) for PCI and PCI-e bus technologies

PCI-e

The PCI-e bus standard replaces the conventional PCI and the PCI-X bus standards. The main difference between PCI-e and PCI is bus topology (signal format). The PCI-e bus uses a point-to-point serial communication channel and PCI uses a shared parallel bus. Another difference is that PCI's bus is only as fast as the slowest device attached and PCI-e supports full-duplex communications between any end point devices:



An example of the PCIe 4X standard expansion card and interface
Image courtesy: Kingston Technologies, Inc.

Expansion cards

So, *what type of cards fit into a motherboard expansion slot?* The following sections describe the functions of the expansion cards you should know about for the Server+ exam.

Network interface controller (NIC)

Most newer motherboards incorporate all or some of the functions of NICs, also known as **network adapters**, into their chipsets. This is especially true for wireless communication. Many active network servers connect to network or communication links through one or more network adapters installed as expansion cards.

The primary function of an NIC is to provide a connection and interface between its host computer and a network. The NIC converts the data sent by the computer into a format compatible with the network's protocols and standards.

Host Bus Adapter (HBA)

An HBA provides a connecting point for peripheral devices to a computer. An HBA is usually an expansion card inserted into a slot on a motherboard. The HBA card serves as a conduit through which an external device and a computer can communicate. Perhaps the most common HBA installed in servers is a hard disk controller card, which includes a disk controller and provides an interface to one or more hard disk drives. The hard disk interfaces provided by an HBA include Ethernet, **Parallel Advanced Technology Attachment (PATA)/Integrated Drive Electronics (IDE)**, **Serial Advanced Technology Attachment (SATA)**, and **Small Computer System Interface (SCSI)** devices, among others.

Redundant Array of Independent Disks (RAID) controller

Another type of HBA is a RAID controller, also known as a *disk array controller*. A disk array controller provides management of multiple hard disk drives, which it presents to a CPU as logical, rather than physical, units. The RAID technology in use must be the same for the RAID controller and the disk drives. For example, a **RAID O** controller won't work with any RAID technology involving fault tolerance.

Riser cards

A riser card is a board that plugs into the system board and provides additional slots for adapter cards. Because it rises above the system board, it enables you to connect additional adapters to the system in an orientation that is parallel to the system board and saves space within the system case. Riser boards are common to rack-mounted servers. One rack-unit, 1U systems can have a single-slot riser integrated into the motherboard, which allows an expansion card to fit inside the case. A 2U computer may have an integrated riser like a 1U or an expansion slot for a 3.5-inch riser card with three slots. The following image shows an example of a 2U riser card with two PCI-e 8x ports:



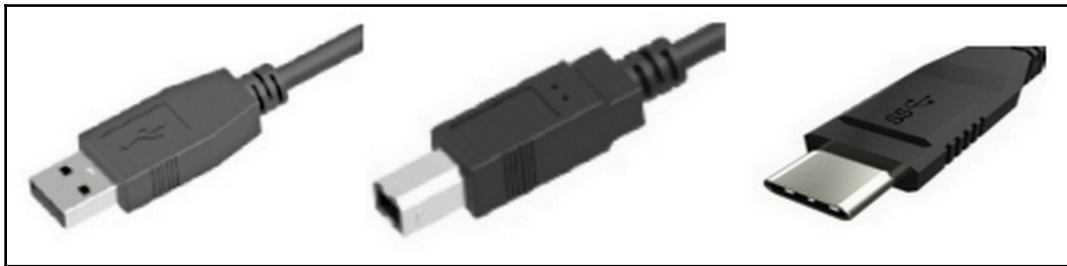
A 2U riser card with 2 PCI-e slots
Image courtesy: Super Micro Computer, Inc.

USB interface and port

There are several USB standards, each with its own size, shape, and purpose. All USB standards define a communication protocol, a physical cable, and one or more sets of connectors and ports. The shape and size of each USB connector and port, regardless of its version, is one of several connector types. Some of the standards are functionally compatible as long as the connector and port used are of the same type.

The different USB connector types, at least those still in use, are:

- **Type A:** Type A connectors and ports are what most people think of as a USB. This connector type is on virtually every computer, regardless of its portability. It's also found on many game consoles, television sets, audio devices, and so on. A-type connections provide a down-link between a peripheral device and a host with the host supplying 5V DC power to the connected device.
- **Type B:** Type B connectors and ports are common to the device end of the cables for printers, external disk drives, and other peripherals.
- **Type C:** USB Type C connectors and ports haven't found a purpose on a network server, yet. Type C has become common for cellular telephones, notebooks, and other portable devices. Its benefits are that a Type C connector isn't oriented (there is no up or down) and either end of a cable can connect to a host system:



Left to right: Type A, Type B, and Type C USB connectors
Images courtesy: Newnex Technology Corp.

There are other USB connector and port types, such as mini, micro, and internal. Each has its specific purpose, but don't fret about seeing them in the exam.

Configuration

An extremely important part of the administration of any server is its configuration settings. Remember that a computer is just a bunch of electrical and electronic components interconnected to accomplish a task repetitively, over and over. The configuration settings for a server are initially set in its firmware when the motherboard and its **read-only memory (ROM)** are manufactured. However, at some point in the life of any server, something in its configuration settings will change. New peripherals, storage devices, memory size, or other additions or adjustments are likely to keep the server supporting the needs of the network.

BIOS

BIOS has been with us since 1975 and basically hasn't changed much functionally over the years. BIOS resides on non-volatile memory (ROM) on a computer's motherboard and is 1) the first instruction set executed when a computer powers on and 2) provides operational support to the operating system for input and output operations.

The primary purpose of the BIOS is to provide the information needed to initialize the system, load the device drives for all connected storage devices and peripherals, and to load and initiate the operating system. The configuration data used by the boot process is in a small amount of memory on a CMOS chip. To ensure that the configuration settings are always available for the system startup process, a coin-style flat cell battery, called the CMOS battery, which has a service life of about 10 years, is used.

BIOS reads the first sector on the primary hard disk drive to access the address of the boot device (typically a hard disk drive) or the address of the initialization instructions (code). It then initializes the boot device and starts the operating system. BIOS has one big limitation, though—it only supports 16-bit data transfer, which restricts the amount of data read from ROM.

UEFI

Many newer computers used as servers now have a UEFI in place of BIOS and some systems even have both. UEFI is a system configuration technology developed to replace BIOS. While UEFI and BIOS essentially perform the same basic functions, UEFI stores its configuration data in a `.EFI` file on a hard disk drive in a special area called the **EFI System Partition (ESP)**. The files needed to load and initialize the operating system are also in the ESP.

Summary

A CPU is the electronic component that runs programs, performs arithmetic functions, and manages data and the input and output functions of a computer. A multiprocessor is a single computer with two or more CPUs, either symmetrical or asymmetrical. In SMP, CPUs are equal and share resources, but in ASMP, one CPU is a master and all other CPUs are slaves. Multiprocessing computers support SIMD, MISD, and MIMD. Multi-core processing involves a single microprocessor that contains multiple processors or cores.

Computer systems use several types of cache memory, including browser cache, disk cache, memory cache, and processor cache. Cache memory provides data to a CPU faster than from main memory. CPU systems include three levels of cache memory: L3, L2, and L1.

CPU speed, measured in MHz, is a function of FSB. The CPU's internal frequency applies the CPU multiplier to the FSB's frequency. Stepping is the application of revisions to a CPU. DRAM is volatile; SRAM is non-volatile but loses its stored contents when no system power is available. Memory timing measurements include CL, tRCD, tRP, and tRAS. ECC memory has a dedicated memory unit that provides parity and error-correction.

The bus channels are the internal bus and external bus. The internal bus is on the motherboard and passes data and instructions between its components. The external bus provides peripherals and expansion cards with a communications link. Bus structures include the control bus, address bus, and data bus. The bus channels and expansion slots you should know about are PCI, PCI-X, and PCI-e, and their significant differences.

BIOS is permanently on ROM and contains the first instructions executed when a computer powers on. BIOS boots the system, loads device drives, and initiates the OS. Configuration data for the boot process is on CMOS. Newer computers have UEFI in place of BIOS.

Questions

1. What is an electronic component that runs programs, performs arithmetic functions, and manages data and the I/O functions in a computer?
 1. GPU
 2. Control unit
 3. CPU
 4. Memory
2. Which of the following statements describes symmetrical multiprocessing?
 1. CPUs are unequal and prorate system resources
 2. CPUs are equal and share system resources
 3. One CPU is a master and all others are slaves
 4. Each of multiple CPUs processes the same instructions

3. Which of the following statements describes asymmetrical multiprocessing?
 1. CPUs are unequal and prorate system resources
 2. CPUs are equal and share system resources
 3. One CPU is a master and all others are slaves
 4. Each of multiple CPUs processes the same instructions
4. A single computing device with two or more CPUs that are either symmetrical or asymmetrical is a:
 1. Microprocessor
 2. Multiprocessor
 3. Uniprocessor
 4. None of the above
5. In this microprocessor type, multiple processors execute different instructions on a single data source:
 1. SIMD
 2. MISD
 3. MIMD
 4. UEFI
6. Which of the following is not a level of cache memory in a CPU system?
 1. Level 3
 2. Level 2
 3. Level 1
 4. Level 0
7. CPU speed is a function of which system feature?
 1. CPU multiplier and the frequency of the FSB
 2. ECC and memory parity
 3. BIOS and DRAM
 4. Memory timing and CL
8. Which of the following is not a PCI bus channel?
 1. PCI
 2. PCI-X
 3. PCI-e
 4. PCIC

9. Which two of the following contain the configuration data for a PC and is activated when a computer is powered up?
 1. BIOS
 2. CMOS
 3. BOOTP
 4. UEFI

10. What is the amount of time it takes to receive and fulfill a request for data from memory?
 1. CAS latency (CL)
 2. RAS to CAS Delay (tRCD)
 3. RAS precharge (tRP)
 4. Row active time (tRAS)

3

Data Storage

A crucial part of any network are the data storage devices that allow data to be organized, cataloged, and mapped in permanent storage for the purposes of retention and retrieval. In our day-to-day computer use, we almost take the hard disk drive in our desktop or portable computer for granted. It's just there, after all. Data storage on a network, which uses basically the same storage technologies as the disk drive in the desktop computer, is more complicated and does require more attention.

In this chapter, we'll focus on the technology, devices, protocols, medium, and applications of data storage on a network. In doing so, we'll look at the following topics:

- The hardware specifications of data storage devices
- The primary disk drive interface protocols
- Data storage systems
- File systems
- Redundant Array of Independent Disks (RAID)
- Capacity planning for data storage needs

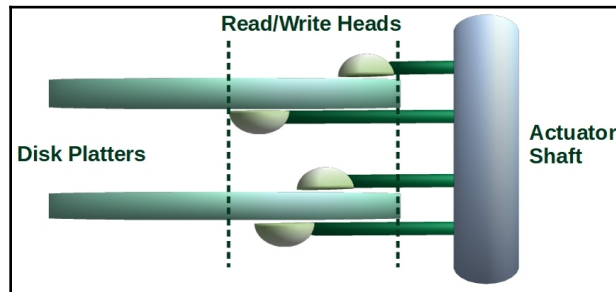
Data storage devices and their specifications

Data has become one of the most valuable assets a person, business, company, enterprise, corporation, or organization has and must protect. Computing and storage devices that process and store data must be reliable and available, and above all else, secure.

In this section, we'll look at disk drive hardware and configuration, data storage and transfer protocols and technologies, and we'll also take a hard look at RAID.

Hard drive specifications

On the most basic functioning levels, technology in hard disk drives, of all shapes and sizes is virtually the same. The following diagram shows the essential internal components of a basic hard disk drive. As this diagram illustrates, an actuator arm, attached to and positioned by the actuator shaft, extends, and retracts the read/write heads to place them over the appropriate track/cylinder and sector:



The primary components of a hard disk drive

As indicated earlier, hard disk drives come in a wide variety of sizes. In this case, size means different things: height, circumference, and capacity. A disk drive's form factor sets the first two characteristics. The capacity of a hard disk drive relates to the disk's form factor, but it's more a function of the technology in use.

Form factors

A form factor, as we discussed in an earlier chapter, sets the form and fit of a computer's case, motherboard, power supply, and more. However, for a hard disk drive, the form factor sets the height, width, length, and the type and placement of its connector to the host computer. The following sections discuss the more popular hard disk form factors found in servers.

Small form factor (SFF)

SFF is smaller than most other common forms. The diameter of the internal disk platters average 2.5 inches in diameter. The SFF standard specifies a drive width of 2.7 inches, an overall length of 100 mm, or 3.93 inches, and a height ranging from 5 mm (0.20 inches) to 15 mm (0.59 inches). Because of their compact size, SFF hard disk drives are common in notebooks and other portable computers. SFF drives can be used in desktop computers as well but require mounting brackets to fill up a 3.5-inch bay.

Because their storage capacity can be as much as 5 TB, SFF drives are in use in some server environments. In these systems, interactive and transaction-processing systems use SFF drives, and backup, archival, and larger volume backups use larger drives or technologies.

Large form factor (LFF)

The height dimension of a **large form factor (LFF)** hard disk drive can range from 19.9 mm (0.78 inches) to 26.1 mm (1.03-inches). While this may not seem *large*, the latest LFF drives are able to store as much as 100 TB, depending on their technology (more on this later). The length dimension of an LFF drive is 146 mm (5.75 inches) and its width is 101.6 mm (4.0 inches). The following image illustrates the difference in the sizes of the SFF and the LFF:



A comparison of a 3.5-inch LFF (left) and a 2.5-inch SFF (right)
Image courtesy: Seagate Technology, LLC

HDD specification and configuration

The technical specifications of an HDD describe its technology, interface, capacity, and operational speeds. For any HDD, there are two sets of specifications: the device specifications and the configured specifications. **Device specifications** indicate the *raw* or unformatted measurements of the disk drive's components and operations. The **configuration specifications**, which are also called **logical configurations**, reflect any changes to the device specifications after formatting or partitioning the disk media.

The device specifications important to either internal or external disk drives are:

- **Revolutions per minute (RPM):** The two dominant RPM rates in HDDs are 7,200 and 5,400, which indicate the number of times the disk platters in a disk drive rotate in a minute. A higher number means a faster disk drive. For example, a 7,200 RPM drive is roughly one-third faster than a 5,400 RPM drive. In this case, faster means that the stored bits on the disk platter move under the read/write head at a faster rate, which speeds up the data transfer process.
- **Interface:** The type of communication interface a drive supports can indicate the capacity of a drive, the bus width used to transfer data to or from the HDD, and its maximum (raw) data storage capacity. The following table shows the transfer rates and maximum capacity for each of the currently popular hard disk bus interfaces (see the *Hard disk interfaces* section later in this chapter for more information):

HDD Interface	Maximum transfer rate (Mbps)	Maximum capacity (TB)
PATA	133	1
SCSI	320	1
SATA	600	12
SAS	750	12

Characteristics of commonly used HDD interfaces

- **Access time:** This is the aggregate of each of the timing elements that occur after the disk controller initiates an I/O action and immediately before data can be read from or written to the disk medium. Access time includes the following:
 - **Seek time:** The time it takes for the actuator arm to position the read/write head over the appropriate disk platter and then over the disk track on that platter that contains the targeted data sector.
 - **Rotational latency:** Also known as rotational delay or latency, this is the time required for the rotating disk platter to move the targeted data sector under the read/write head.

- **Command execution time and settle time:** These measurements are generally very small and may not be included in a manufacturer's hard disk drive performance specifications. **Command execution time** is the time required to establish the link between the various elements involved in the transfer of data to or from the memory or the disk drive. **Settle time** is the time required to lower the read/write head into position to write data to the disk medium or read data from the disk medium. Most HDD specifications include settle time in seek time.
- **Throughput:** Also known as **data transfer rate (DTR)**, this is the time required to move data from one location to another in a specific operation. In the context of an HDD, throughput is measured as *read throughput* and *write throughput*. In either case, this is what is required to move data from the read/write head to memory (read throughput) or from memory to the read/write head (write throughput).
- **I/O operations per second (IOPS):** This is pronounced *eye-ops*, IOPS measures the number of read and write operations from and to random, non-contiguous addresses that an HDD can perform in one second. The structure of this measurement may vary from different manufacturers, so it may not be a reliable metric for comparing HDDs.

Disk capacity – decimal versus binary

The storage capacity of a storage device, such as an HDD, is expressed as both a decimal value and a binary value, which can confuse the amount of actual space available on the device. For example, the decimal value of 1000^{10} or 1 KB is often used interchangeably as the equivalent of 2^{10} or 1024^{10} which is also referred to as 1 KB.

A hard disk drive and a few other storage devices often appear to lose around seven percent of their stated or nominal capacity after they are formatted. For example, a new 1 GB HDD ends up having only approximately 938 MB after being formatted. This difference represents the average of seven percent of the total storage that seems to disappear.

The confusion comes from the fact that a computer, because of its design, uses binary representations of numeric values, such as 2^{20} being considered to be the same as 10^6 , or 1,048,576 being interchangeable with 1,000,000, or essentially 1 MB equaling 1 MB.

The following table lists the more commonly used prefix notations and their equivalent values:

Prefix	Value	Decimal power	Decimal value	Binary power	Binary value
Kilo	Thousand	10 ³	1,000	2 ¹⁰	1,024
Mega	Million	10 ⁶	1,000,000	2 ²⁰	1,048,576
Giga	Billion	10 ⁹	1,000,000,000	2 ³⁰	1,073,741,824
Tera	Trillion	10 ¹²	1,000,000,000,000	2 ⁴⁰	1,099,511,627,776
Peta	Quadrillion	10 ¹⁵	1,000,000,000,000,000	2 ⁵⁰	1,125,899,906,842,624
Exa	Quintillion	10 ¹⁸	1,000,000,000,000,000,000	2 ⁶⁰	1,152,921,504,606,846,976
Zetta	Sextillion	10 ²¹	1,000,000,000,000,000,000,000	2 ⁷⁰	1,180,591,620,714,113,034,240
Yotta	Septillion	10 ²⁴	1,000,000,000,000,000,000,000,000	2 ⁸⁰	1,208,925,819,614,629,174,706,176

Numeric notation prefixes

Hard disk drive (HDD) versus solid-state drive (SSD)

A mechanical, or common, HDD has at least eight major moving parts, some of which are shown in *The primary components of a hard disk drive* image in the *Hard drive specifications* section earlier in the chapter. Just the term *moving parts* may send up a red flag, because anything that moves can break or wear out. Okay, so not many hard drives fail due to a seized-up actuator arm or an intermittent read/write gap error, but they could. A hard disk drive that has no moving parts would be ideal and there'd be no threat of mechanical breakdown. An SSD is just such a device.

An SSD works along the lines of a USB flash memory drive in that both store data in semiconductor chips (see the following image). Solid-state is a term that has been around since the days of transistor radios. It refers to a circuit that is made up solely of semiconductors. Solid-state also means that there are no moving parts. SSDs are mostly designed for internal installation in a portable or compact computer, but they can also be external devices as well, sort of like very large flash drives. SSDs typically use either an SAS or SATA interface:



An exploded view of an SSD.
Image courtesy: Intel Corporation

The benefits of using an SSD over an HDD are characterized mainly by three things: dependability, speed, and reduced power consumption. As shown in the following table, an SSD has an edge over an HDD, but only in the extreme. An SSD is nearly four times faster than an HDD on average and uses about half as much power. On the other hand, the capacity of an SSD hasn't caught up with HDD yet and its cost per GB is much higher. What this all means is that an SSD storage device is more dependable, faster, and operationally less costly to use. However, the upfront costs are higher:

Characteristic	HDD	SSD
Failure rate	1.5 million hours MTBF*	2.0 million hours MTBF*
Read / write speed	50 – 120 MBps	200 – 550 MBps
Power usage	6 – 7 watts	2 – 3 watts
Maximum capacity	10 TB	4 TB
Cost	\$ 0.03/GB	\$ 0.20/GB

A comparison of the physical characteristics of an HDD and an SSD

* MTBF = mean time between failures

SSD specification and configuration

Comparing an HDD to an SSD on just their speed ratings is simple—the SSD is faster. The answer to the question, *What is the difference between SSD and HDD?* is *pretty much everything*. Well, maybe not everything. After all, an SSD doesn't have all of the moving parts. In fact, it has none of those of an HDD. This removes some performance specifications from the comparison, such as RPM, rotational delay, settle time, and a few others.

Whereas an HDD is a serially accessed device, an SSD is a random-access device. You can randomly access an HDD, but there is latency in getting to the data. On an SSD, you go directly to the data. None of the latency issues associated with the physics of an HDD effect the performance of an SSD. Measures such as IOPS, access time, and throughput are many times higher or lower, as the case may be. Remember too that the SSD is more expensive, at least for the time being.

Hard disk interfaces

Although it may be inside the system case, an HDD is a peripheral device to the motherboard and CPU. An HDD and a computer communicate, that is, pass data to each other, through a drive interface. A drive interface is the bus structure that the device driver, the hard disk device, RAM, and controllers use to pass data requests and the data requested.

HDD interfaces are either word-serial or bit-serial. Word-oriented interfaces transfer data in 8-bit, 16-bit, or 32-bit words using parallel bit signaling. Examples of word-oriented interfaces are the **Small Computer Serial Interface (SCSI)** and the **Parallel Advanced Technology Attachment (PATA)**. Bit-oriented interfaces interconnect through a **host bus adapter (HBA)**. Examples of bit-oriented interfaces are **fibre channel (FC)**, **serial ATA (SATA)**, and **serial attached SCSI (SAS)**.

The HDD drive interfaces you should know about for the Security+ exam are as follows:

- **PATA:** PATA uses parallel bit signaling to transmit word-length data between an HDD and controllers and drivers. It connects to the system through an 80-pin cable that carries a 5V signal. Not every conductor carries data. In fact, every second wire is a ground. Because of the size of its connecting cable, a PATA HDD is an internally mounted device. PATA devices can transfer data with DTRs as high as 80 MBps.

- **SATA:** SATA uses serial bit signaling to transmit data between an HDD and the motherboard, controllers, and drivers. SATA transmits data over a 7-conductor cable at higher speeds. SATA came about primarily because the PATA standard was unable to support the higher **data transfer rates (DTRs)** of larger capacity disk drives. SATA devices can transfer data with DTRs as high as 600 MBps.
- **SCSI:** SCSI (*scuzzy*) provides a multiple-connection interface for a variety of devices, including HDD, optical drives, printers, scanners, and other peripherals. As many as 15 devices can be attached to an SCSI interface that occupies only a single HBA slot in a host computer. SCSI uses parallel bit signaling and can transfer data with DTRs as high as 80 MBps.
- **SAS:** SAS improves on its base interface SCSI by expanding several of its capabilities and capacities. SAS controllers are connected directly to SAS HDD. SAS can connect up to 128 different devices on a single interface. It is a full-duplex interface that can transfer data with DTRs as high as 3.0 GBps. Another important characteristic of SAS is that it is a hot-plug technology.
- **FC:** The FC technology is primarily a communication protocol commonly used in data centers and server farms. Although its name suggests that it is based on fibre optic cabling, because it's based on a protocol, the transmission medium it uses can be coaxial, copper twisted-pair, and fibre optic cabling. The type of cable used determines the length of separation for FC-connected devices. Copper-connected nodes should be within 100 feet, but multimodal fibre optic cable-connected devices can be as far as 10 kilometers (6.67 miles) apart. The FC technology is replacing SCSI in data centers because of its reliability, flexibility, speed, and distances. Currently, FC can transmit data with a 128 GBps DTR.

Data storage systems

At this point, we need to squeeze in some basic disk storage terminology and concepts. For the Server+ exam, your knowledge of these terms and concepts is generally assumed, so this is something we need to cover. In this and the next few sections, we'll look at the technical organization of data storage and retrieval methods, technologies, and systems.

Direct-attached storage (DAS)

The name DAS should be self-explanatory. DAS is any storage device, HDD, SSD, USB, tape, optical disc, flash drive, and so on connected directly to or attached to a computer. An HDD installed inside a computer's case is DAS; a USB-connected external HDD is DAS; and an a USB tape drive is DAS. In other words, if a data storage device is directly attached to a computer and its bus, it's DAS.

Network-attached storage (NAS)

NAS, as shown in the next image, is one or more data storage devices attached to a network that network clients can share. NAS appears to network users much like DAS, especially when the **Network File System (NFS)** is in use. NFS is a network file-sharing protocol that allows authenticated and authorized users to access network file systems and resources, such as NAS. Although NFS is a common element of NAS systems, it can also work with unstructured clusters of HDDs, which are generally referred to as **just a bunch of disks (JBOD)**. In a JBOD arrangement, the disks operate as independent devices:



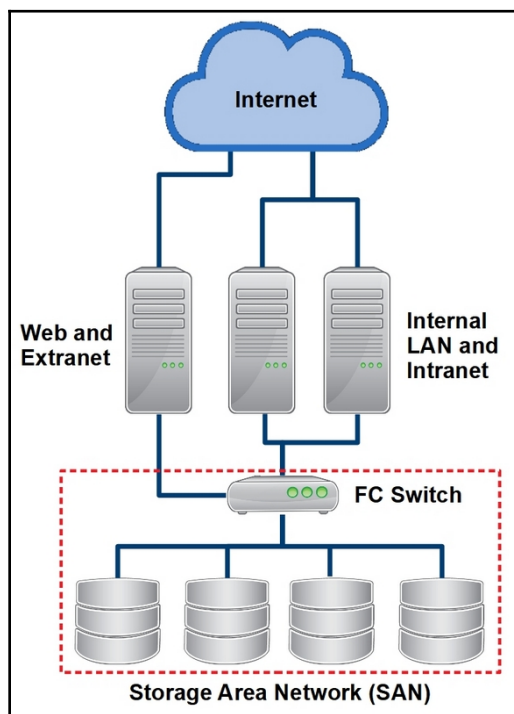
A 16-bay NAS unit
Image courtesy: Infortrend Technology, Inc

Storage area network (SAN)

One commonly used application for FC communication, especially in large networks and data centers, is a SAN. A SAN is a self-contained network of storage devices and specialized switches that provides high-speed data access to connected network nodes. To a network server, portions of a SAN appears to be directly connected to only that server. Every server connected to the SAN has this same view. A SAN provides network servers with the capability to move data between storage devices and perform backup and restore activities, in addition to providing multiple servers with data access.

SAN fabric

As the following diagram illustrates, a SAN, in its simplest form, is made up of one or more network servers that are connected to a data storage array of two or more hard disk drives through one or more dedicated switches. In this diagram, the red, dashed box encloses what is called the **fabric**, which is the switch-controlled disk drive network of the SAN:



The components of a SAN

SAN communications

Most SANs use one of two communication technologies: FC and **Internet SCSI (iSCSI)**:

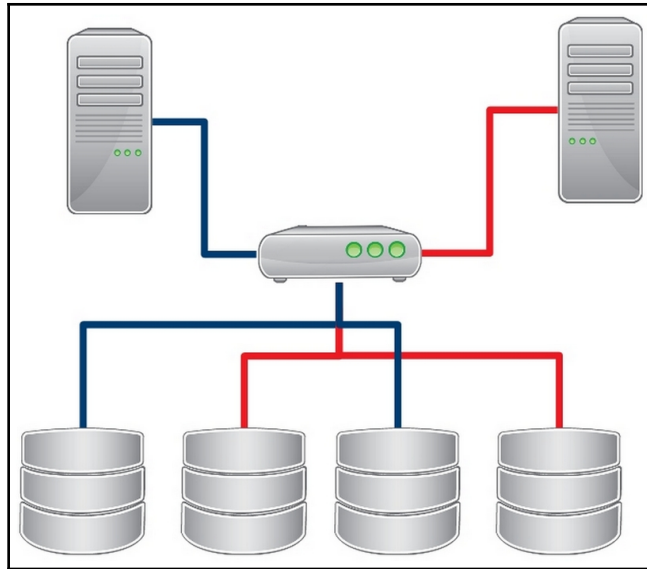
- **FC SAN:** FC is the most commonly used SAN communication technology. An FC-based SAN incorporates FC switches that interconnect storage devices and FC HBAs that connect FC switches to network servers.
- **iSCSI SAN:** The iSCSI alternative doesn't require the specialized devices of the FC-based SAN. An iSCSI SAN can operate on Ethernet switches and physical connections to storage devices and network servers. The result is a lower-cost, although lower-performing, SAN solution.

Another SAN option is **Fibre Channel over Ethernet (FCoE)**. FCoE merges Ethernet and FC signals on a common network medium. However, not all Ethernet networking devices support the FCoE standard.

Logical Unit Number (LUN) zoning and masking

One critical element in a SAN, based on any of the communication technologies, is the LUN. A LUN is an identity assigned to a unit of SAN storage, which can be a portion of an HDD, an entire HDD, or even more than one HDD, especially in RAID configurations. A LUN represents a mapping between the storage media and network clients that indicates which network hosts can access which SAN-managed data.

On an FC SAN, setting up LUN zoning defines specifically which network servers/hosts have access to certain LUNs and their contents. As illustrated in the following diagram, the server on the left has zone access to the first and third LUNs and the server on the right can access the second and the fourth LUNs. LUN zoning is specific to FC SANs. iSCSI and FCoE, because they use Ethernet technology, are set up much like a **virtual LAN (VLAN)**:



LUN zoning determines which hosts have access to which LUNs

Once LUN zoning is in place, LUN masking applies additional restrictions that can hide a LUN from specific servers and hosts. While LUN masking can be part of a security policy in many data centers, its primary usage is typically to restrict access to only the data a network server or host (and their users) requires.

Filesystem

Any computer, stand alone or network-attached, organizes the data it stores on secondary storage using a **filesystem**. A filesystem isn't just a pattern for storing data files. It also includes the processes, methods, and structures used to name, store, locate, and retrieve data. Although, it is also used by some to refer to a disk partition.

Operating systems and filesystems

Network servers can and do use a wide assortment of filesystems, some proprietary to an operating system and some standard or more generic. For example, as shown in the following table, several Linux versions from different providers all used the **Extended File System version 3 (ext3)** and later **ext4**. Windows systems relied on the **New Technology File System (NTFS)** for several years and it is still in use today, but newer filesystems are emerging:

Year released	Operating system	File system
2000	Windows Server 2000	NTFS
2005	Fedora/Ubuntu/Debian Linux	ext3
2006	Windows Vista	NTFS
2009	Windows 7	NTFS
2012	Windows Server 2012	Resilient File System (ReFS)
2013	Fedora/Ubuntu/Debian Linux	Extended File System version 4 (ext4)
2015	OpenSUSE 42.1	Better FS (Btrfs)/Extents File System (XFS)
2017	macOS	Apple File System (APFS)

Examples of operating systems and their associated file systems over the years

File sharing

The **Server Message Block (SMB)** is a file-sharing protocol that give applications the capability to access files and other network resources on a network. A network client is able to create, open, read, write, or move data files on any network server configured to respond to SMB client requests. The SMB protocol has a standard format for SMB request messages, but different *dialects*, format variations, developed for a variety of specific operating systems or system requirements. One such dialect, actually one of the original dialects, is the **Common Internet File System (CIFS)**, Microsoft's version of SMB for accessing file and print services on a network. Microsoft replaced CIFS with its own version of SMB.

Another commonly known dialect of SMB is **Samba**, which enables a Windows computer to share files with a computer running the UNIX or Linux operating system. Samba combines a number of different protocols and services, including SMB and **NetBIOS over TCP/IP (NBT)**.

RAID

An essential element of data center and server administration is **availability** as, in a network, data stores, and communication links are always available, meaning 24 hours a day, 7 days a week. In an electrically powered world, keeping electric devices always available goes beyond backup power units, **uninterruptible power systems (UPSs)**, and generators. Even when electrical power is guaranteed to be always available, there are several other threats to the availability of a system or network. However, in your preparation for the Server+ exam, we should focus on the availability of data storage devices and systems, and particularly their reliability.

RAID is a technical methodology used to improve the availability, reliability, and performance of data storage systems, especially in larger data centers, cloud systems, and high-volume **transaction processing systems (TPSs)**. In general, a RAID system is built on two or more HDD or SSD volumes that operate synchronously. There are no RAID standards and each company offering a RAID solution can offer its own specifications and levels. However, there are some practices and configurations that are generally implemented.

Striping and mirroring

Striping and mirroring are the methods that a RAID system can use to create data redundancy:

- **Striping** separates data into chunks that are stored on two or more disk drives or volumes. The size of these chunks varies by provider and ranges from single bytes and fixed-length blocks to entire partitions. For example, a RAID system with 5 HDDs may stripe 128 KB data blocks onto each drive and, if needed, would continue by repeating the sequence. A RAID system with 10 disks could stripe a 10 MB file to each of the 10 disks in 1 MB stripes.
- **Mirroring** is essentially duplicating or copying data onto two or more drives. The primary purpose of mirroring is to provide an exact copy of a data file, for instance, so that it could be a failover should something happen to the original file or the drive on which it's stored. Although a mirror is like a backup, it's online and immediately available, where a backup requires a few more steps. Mirrored files are generally synchronized, meaning updates to the original are also made to the mirror. This may also speed up data access operations because parts of a data file can be read from both copies.

RAID levels

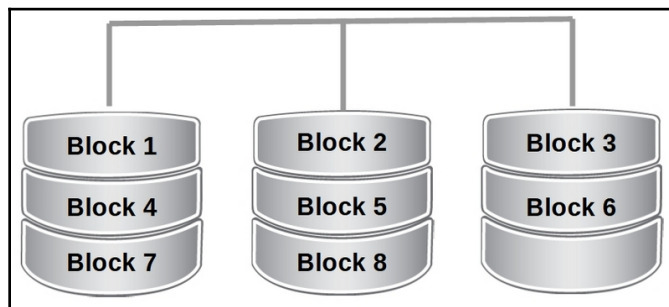
A RAID level is a functional methodology that incorporates striping or mirroring, or both, and may apply parity or not. Over the years, vendors and providers have defined 10 or more RAID levels, some of which are still in use as quasi-standards, and some have become obsolete.

The following table lists the various RAID levels that are commonly used and may show up on the Server+ exam:

RAID level	Description	Minimum number of disks
0	Block striping; no parity	2
1	Mirroring; no parity	2
5	Block striping; with parity	3
6	Block striping; double parity	4
1+0 (10)	Block striping with mirroring; no parity	4

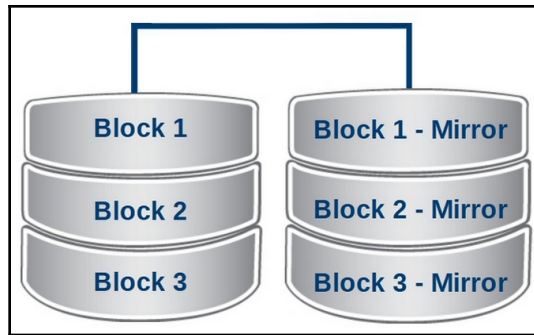
Active RAID levels

- **RAID 0:** RAID level 0 isn't a redundancy or data recovery method; it's more of a performance enhancer. By striping data to multiple disk drives, data I/O processes speed up. The next diagram illustrates striping across three disks:



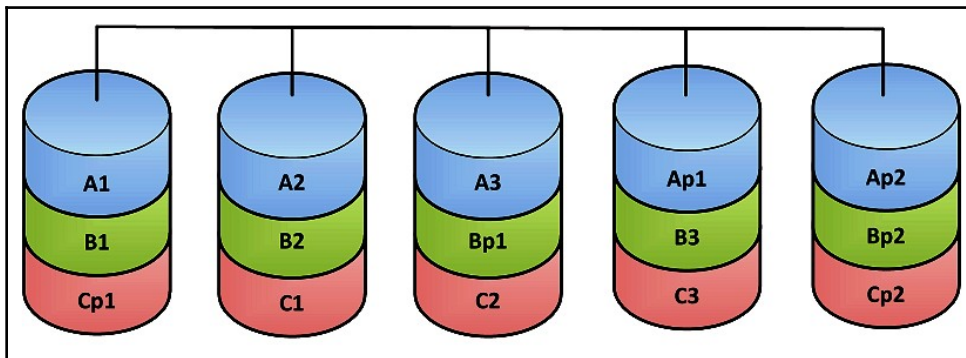
RAID 0 with data blocks striped across multiple disks

- **RAID 1:** RAID level 1 provides a mirror image or *replica* of a data block that is available for processing should a disk drive fail, or the primary data copy become corrupted. This diagram illustrates RAID 1 mirroring:



RAID 1 with data mirrors saved to another disk

- **RAID 5:** RAID level 5 looks like RAID 0 with parity added in (see the following diagram). The similarity goes only that far because RAID 5, which is the most common RAID implementation in enterprise networks and data centers, uses the striped data and parity information to recreate a data block from a failing drive so processing may continue uninterrupted. The downside to a RAID 5 implementation is that in I/O heavy environments, especially those with a lot of write actions, latency could be an issue:

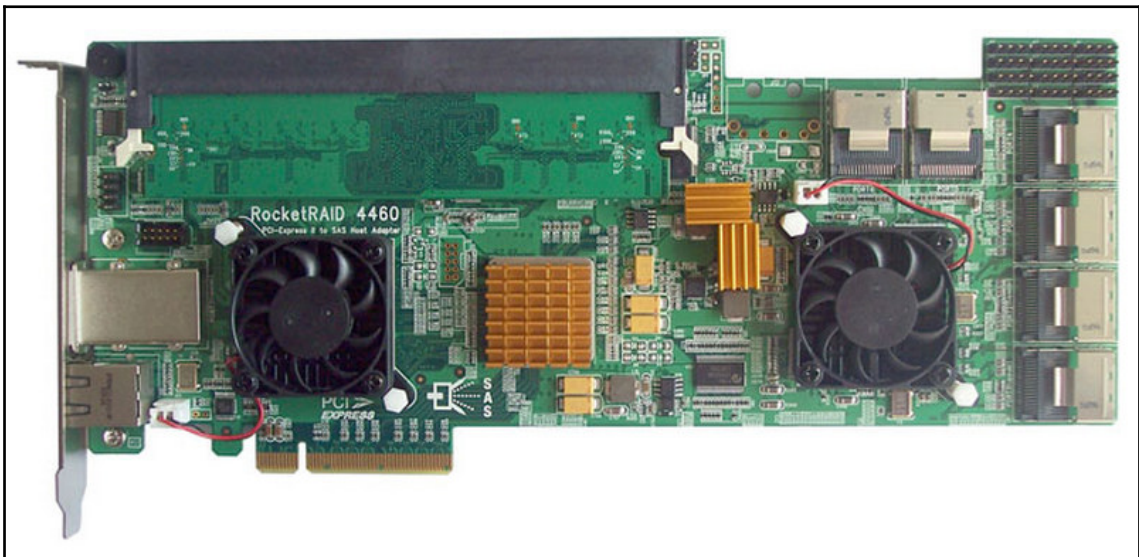


RAID 5 with data and parity information striped to multiple drives

- **RAID 6:** RAID level 6 is essentially RAID 5 with the parity information doubled, which means that two or more disk drives could fail and processing would continue. That is, providing that more than two drives are in the array. RAID 6 is also common to enterprise networks and data centers.
- **RAID 10:** RAID level 1+0 or 10 combines the mirroring of level 1 with the striping of level 0, but without parity. This RAID level is popular on smaller networks and servers and is applied through software RAID applications.

RAID implementation

A RAID system can be implemented on a server, SAN, or NAS as hardware RAID or software RAID. Operationally, there isn't much difference between the two RAID types, although many experts favor hardware RAID. If there are differences, they are a part of where the RAID processing takes place. In a software RAID implementation, RAID processing is performed by a server's (typically the RAID host server's) CPU. Hardware RAID is implemented in a variety of ways: a RAID controller expansion card installed in a host server (as shown in the following image); a stand-alone RAID controller appliance; or disk drives with embedded RAID controllers. Performance is often given as the primary difference between the two, with hardware RAID having an advantage. However, the differences are dependent on the number of disk drives and their formatting and the RAID level in use:



A hardware RAID controller expansion card
Image courtesy: HighPoint, Inc

An array controller, also known as a disk array controller, RAID controller, and **storage processor (SP)**, manages and controls an array of disk drives, which it presents to servers and hosts as one or more logical units. The disk array can be one big JBOD or configured as NAS or SAN. An array controller typically has three main components: a processor, RAM, and I/O interfaces. The processor interprets incoming requests for data to be provided or to be stored. The instructions and addressing required for data movement in or out is in RAM. The I/O component typically has at least one frontend port that communicates with the host computer's host adapter, and a backend port that links to the disk drives.

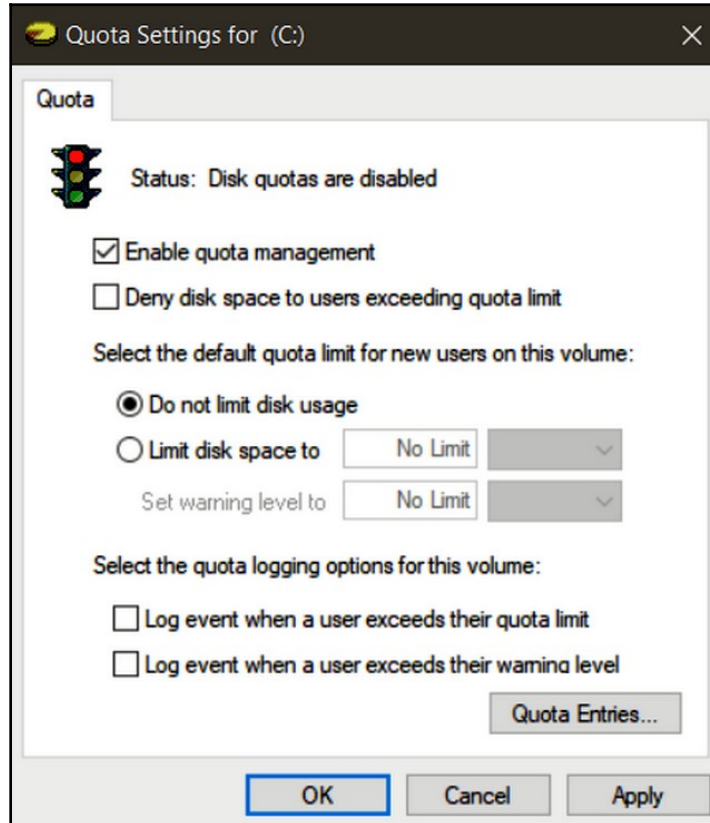
RAID systems commonly include **battery-backed cache memory**. This cache provides a buffer that enables the RAID controller to process data in either direction of I/O operations faster than its able to write it to a disk. The battery power enables the data to be retained in the event of a power failure. In larger RAID implementations, a **redundant disk array controller (RDAC)** that shares a dual-port connection with the primary disk array controller provides an immediate backup and failover should the primary fail. Without this redundancy, a failed disk array controller renders the disk array unreachable.

Disk quotas

System and network administrators not only have the job of ensuring that a system or network keeps running and remains available to its users. Included in their duties are the tasks of managing, controlling, and allocating system resources. Hard disk space on a multi-user system may require usage limits to provide adequate storage space for all users. One way of accomplishing this is to set disk quotas.

Disk quotas, which are set at user or group levels, limit the maximum size of single files or data blocks written to an HDD. Typically, a disk quota is specific to individual users or groups as one of two types of quota settings: user, or block, quotas or file, or inode (UNIX), quotas. A user quota limits an individual or group account to a specific maximum disk usage limit. File quotas limit the number and/or size of files created by a user or group.

In addition, disk quotas of either type are specific to a file system. Disk quotas can be set in most operating systems, including Windows (see the following screenshot), UNIX, and Linux:



Setting disk quotas on a Windows system

Disk compression

Many administrators see disk compression as a means to reduce the amount of space data files take up on an HDD. Some even claim that disk compression may speed up the I/O process since the compression/decompression actions take place in RAM. Well, this is another place where the following Gilster's Law applies:

"You never can tell, and it all depends."

Disk compression, in general, may add a small amount of time to the I/O process, in either direction. Depending on the nature of the data stored in compressed form, the increase in the amount of data stored on a disk may mitigate the latency of the compression. On a high-volume I/O system, a compressed disk may not be the best solution. However, a data archive or warehouse, where I/O activity is low, is a much better candidate.

Disk compression reduces the space that stored data uses on a disk medium. Disk compression, which is different from file compression, which compacts individual files, compresses all data written to the disk. When disk compression is enabled, a data compression utility that is between the operating system and the HDD controller intercepts any data written to or read from the affected HDD. The data is compressed or decompressed by the utility, as needed, and passed on.

High availability (HA)

An extremely important characteristic of a network server and all other network-connected systems is HA. A system that exhibits HA is operating and available a predominant percentage of the time. HA is measured in the percentage of uptime a system should or does realize. You may have heard the expression *four nines* (9999). What this refers to is the number of digits to the right of the decimal when expressing the uptime percentage (as shown in the following table):

Availability	Common name	Downtime/year	Downtime/month	Downtime/week	Downtime/day
99.9999999%	9-nines	31.56 ms	2.63 ms	604.80 μs	86.40 μs
99.999%	5-nines	5.26 min	26.30 sec	6.05 sec	864.00 ms
99.99%	4-nines	52.60 min	4.38 min	1.01 min	8.64 sec
99%	2-nines	3.65 days	7.31 hrs	1.68 hrs	14.40 min
55.5555555%	9-fives	162.33 days	13.53 days	74.92 hrs	10.67 hrs

Some of the more commonly used HA measures with their time equivalents

- * ms: milliseconds
- * μs: microseconds
- * sec: seconds
- * min: minutes
- * hrs: hours

The nines

As shown in the previous table, what is meant by HA can vary with the purpose and use of a network server and the network-connected devices. Each supporting device on a network, such as RAID, NAS, SAN, **Routing and Remote Access Service (RRAS)**, and so on, typically has an HA expectation to both the data center administrators and its users. A goal of *9-nines* may be a bit too aggressive for most server operations, allowing only 31.56 milliseconds of downtime per year. However, many **service level agreements (SLAs)**, or customer or subscriber support commitments, are typically for four or five-nines, committing to no more than 5-52 minutes of downtime a year. These commitments are either an average of the actual results experienced from a server or data center, or the aspirations of a network configuration built to provide a desired availability.

Fault tolerance

Where the goal of high availability is to keep a system available and accessible, the goal of **fault tolerance** is to retain all in-process data and operations in the event of a component or system failure. Tolerating a fault or failure is essential for any server or server-connected storage array. The level to which fault tolerance is incorporated into a system is dependent upon the nature of the services provided and the sensitivity of the data involved. A high-demand, around-the-clock system that must always be available combines high availability and fault tolerance principles. At the other end of the spectrum, the fault tolerance of a system may be more of a *soft landing* that provides the time needed to store all in-process data.

Any level of fault tolerance has three general characteristics:

- **There is no single point of failure:** Any element of a system that could possibly fail and take the system down is backed up by a failover or redundant element
- **A component failure shouldn't require the system to stop:** If a component fails, applying a replacement doesn't require any or all systems to be powered down
- **Any fault or failure is easily identified and isolated:** A failed component won't cause data or processing to be lost or interrupted

These principles can be applied to any of the three major failure areas of any system: hardware, software, and power.

Replacing failed components

The capability to replace or take a failed component offline without impacting users or operations is the primary goal of both high availability and fault tolerance. Avoiding the need to power down the entire system to replace a failed component is the ultimate goal. However, it isn't necessary to shut down the whole system or just the failed component in all cases. Some components have the capability of being replaced *on the fly*.

In any case, there are three general types of failed hardware and power component replacement methods:

- **Hot swapping:** Hot swapping involves the immediate switchover or physical replacement of a failed component while a system is running and fully operational. Examples of hot swappable devices, referred to as **hot spares**, are audio devices, displays, USB devices, SATA drives, and network jacks.
- **Warm swapping:** Warm swapping requires the suspension of a system's, or failed component's, operations while a replacement or switchover takes place. In general, warm swapping involves media changes, rather than entire components.
- **Cold swapping:** Cold swapping involves completely shutting down a system to affect upgrades, replacements, or repairs by applying **cold spares**.

Disk storage capacity planning

Forecasting how much data storage space a computer, system, or data center will need in the future, near or far, can be tricky. On the one hand, you don't want to bring in too much storage capacity unnecessarily and tie up money that you could use elsewhere. On the other hand, running out of storage space could mean system interruptions and higher last-minute prices. The trick is to find a happy medium where you will have the right amount of disk space over a specific time frame and a clear plan of when storage space should increase or decrease.

Most HDD manufacturers have disk capacity planning tools you can use to forecast your storage needs in the future. There are also several system management software packages that include a capacity planning function (see the following screenshot). However, before you begin forecasting future needs, be sure you understand where the disk storage demands will come from and what other technologies or approaches may address the capacity requirements. Disk storage technologies and methods such as tiered storage, data compression, virtualization, SAN, and NAS, may be a better and more economical way to go:

Disk Space Summary

Summarized Data

All Reports

PDF Version

Data shown for most recent scans

Summarized Data

Computer	Drive	Last Check...	Free (...)	% Free	Predicted Full ^	Total (GB)	Used (GB)	% Used
TEST-1	C:	8/1/2014 8:13:50 AM	1.5 GB	18 %	8/28/2014 10:24:39 AM	8 GB	6.5 GB	82 %
Archive [192.168.7.2]	C:	8/1/2014 10:16:50 AM	356 GB	38 %	7/18/2015 10:24:39 AM	931.5 GB	575.5 GB	62 %
192.168.7.101	C:	8/1/2014 8:13:51 AM	2.6 GB	31 %		8 GB	5.4 GB	69 %
BIGDELL	C:	8/1/2014 10:11:23 AM	1757 GB	94 %		1862.9 GB	106 GB	6 %
D2	C:	8/1/2014 9:37:15 AM	122.6 GB	26 %		465.8 GB	343.2 GB	74 %
D2	D:	8/1/2014 9:37:15 AM	13.5 GB	12 %		107.1 GB	93.7 GB	88 %
DOMAIN	C:	8/1/2014 10:11:23 AM	1.1 GB	13 %		8 GB	6.9 GB	87 %
LAB	C:	8/1/2014 8:13:51 AM	2.6 GB	31 %		8 GB	5.4 GB	69 %

A disk space capacity planning report
Image courtesy: Power Admin, LLC

Other storage devices

Other forms of data storage devices that just don't work for efficient data retrieval are appropriate for use in data backup, file archives, and disaster recovery. Of these, magnetic tape is perhaps the most commonly used medium for two reasons: portability and cost.

Magnetic tape

Magnetic tape is a plastic or mylar strip that is coated with an iron oxide or chromium layer that can be magnetized or demagnetized to represent data bits, which are written to the tape in either a parallel or helical pattern of up to 128 tracks. Originally, magnetic tape was stored on reels, but today, it's generally packaged inside plastic cases (cartridges).

Data is written to or read from tape serially, which means that to retrieve data from the end of the tape, all the preceding data must be skipped. Historically, magnetic tape wasn't a medium for random access. However, newer technologies, such as the **Linear Tape Open (LTO)**, perform like random access by storing data objects separately from their metadata. LTFS organization allows magnetic tape to store as much as 220 TB of data.

An enterprise or any other organization, typically stores its recorded tapes, whether for regular system backups, data archives, or disaster recovery, in a tape library.

A **tape library** can be an organized storage facility or cabinet, storing tapes in chronological, purpose, or content order. However, magnetic tape devices able to hold several tape cartridges that rotate sequentially on a set schedule are also known as tape libraries.

Optical storage

Another technology used for data backup, storing large files, transferring data or programming from one computer to another, or commercially packaged application software is optical data storage. An **optical disc drive (ODD)** reads from or writes to an optical disc, but, for the most part, optical discs are inputs.

Common examples of optical discs are **compact discs (CDs)**, **digital versatile/video discs (DVDs)**, and Blu-ray discs. Optical discs were at one time considered a better option than most portable magnetic media. A CD can store as much as 700 MB of data; a DVD can hold up to 8.4 GB; and a Blu-ray disk may contain as much as 50 GB. However, with USB flash memory drives (thumb drives/flash drives) now storing as much as 256 GB, optical drives may have a limited capability as backup media.

Summary

Form factors set the form and fit of a computer's case, motherboard, power supply, and disk storage drive, the form factor sets the height, width, length, and the type and placement of its connector to the host computer. HDD device specifications include RPM, interface, access time, throughput, and IOPS. An HDD is a serial device and an SSD is a random-access device.

An SSD stores data in solid-state semiconductor chips and has no moving parts. SSDs use a SAS or SATA interface. Disk drive interfaces are word-serial or bit-serial and bit-oriented interfaces that connect through HBA, FC, SATA, and SAS. PATA uses parallel bit signaling. SATA uses serial bit signaling. SCSI provides a multiple-connection interface for several devices. SAS expands SCSI. NAS is clustered storage devices that appear as a network share. SAN is a network of storage devices and switches providing access to connected network nodes. A LUN identifies a unit of SAN storage. Data is organized on secondary storage in a filesystem. SMB assists applications to access files and network resources on a network. SMB has different versions or dialects that are unique to specific OSes, such as CIFS.

RAID systems are built on two or more HDD or SSD volumes that operate synchronously. Striping and mirroring are the methods a RAID system uses to create data redundancy. Striping separates data into stripes stored on two or more disk drives. Mirroring duplicates data onto two or more drives. RAID levels incorporate striping or mirroring, or both, and may also apply parity. RAID 0 uses striping but isn't a redundancy method. RAID 1 provides mirroring that's available if a disk drive fails or data becomes corrupted. RAID 5 is like RAID 0 with parity added and uses striped data and parity information to recreate a data block on a failing drive. RAID 6 is the equivalent RAID 5 with the parity information doubled and is common in larger networks. RAID Level 1+0 or 10 combines the mirroring of Level 1 with the striping of Level 0, but without parity. Hardware RAID is implemented by a controller card in a host server, a standalone RAID appliance, or through HDDs with embedded controllers. Disk compression reduces the space data requires on a disk medium. Disk compression is different from file compression. A data compression utility between the OS and HDD intercepts data being passed and compresses or decompresses it as needed and passes it on.

HA is the uptime condition of a system that is available a significant percentage of the time. Fault-tolerant systems are able to continue operating in the event of a component or system failure. Hot swapping is the immediate switchover or replacement of a failed component completed while a system remains fully operational. Warm swapping requires the suspension of operations, although still powered, to affect the replacement of a failed component. Cold swapping requires the powering down of a system to affect replacements or repairs. Other media and storage devices may be used to store data, including magnetic tape and optical storage.

Questions

1. Which of the following are differences between an HDD and an SSD? Choose all that apply:
 1. Rotational delay
 2. Access time
 3. Storage capacity
 4. All of the above
2. An SSD device uses one of which two of the following interfaces? Choose two.
 1. PATA
 2. SATA
 3. SCSI
 4. SAS
3. Which of the following storage technologies describes a number of independent disks not configured into an array?
 1. RAID
 2. DAS
 3. SAN
 4. JBOD
4. Which of the following statements best describes a SAN?
 1. One or more storage devices directly connected to a computer
 2. A cluster of data storage devices that appear to be direct-attached storage
 3. A self-contained storage device network and switches that provide high-speed access to data
 4. Zoning that identifies the access permissions of servers and hosts

-
5. Which two of the following identify the access to data resources in a SAN for servers and hosts?
 1. File permissions
 2. LUN zoning
 3. LUN masking
 4. Filesystem
 6. The filesystems common to Linux and Windows respectively are:
 1. Linux and UNIX
 2. DOS and Windows
 3. ext3 and NTFS
 4. EFS and RAID
 7. SAMBA is an implementation of which application layer protocols?
 1. HTTP/FTP
 2. IPSec/WPA
 3. SBA/TCP
 4. SMB/CIFS
 8. What are two redundancy methods applied in the different RAID levels? Choose two.
 1. Mirroring
 2. Compression
 3. Encryption
 4. Striping
 9. Which of the following is not a data redundancy level of RAID?
 1. RAID 0
 2. RAID 1
 3. RAID 5
 4. RAID 10
 10. A system that is fully operational and accessible by users for a significant percentage of time is:
 1. Fault tolerant
 2. Fault averse
 3. Highly available
 4. Extremely available

11. Match the name on the left to the description on the right that best describes it.

a. Hot swapping

b. Cold swapping

c. Warm swapping

1. Requires the powering down of a system to affect replacements or repairs

2. The immediate switchover or replacement of a failed component completed while a system remains fully operational

3. The suspension of operations, although still powered, to affect the replacement of a failed component

4 Server Operating Systems

A computer system has five major components: hardware, software, people, documentation, and data. Without any one of these components, there would be little or no purpose of the system. Of course, there is another version that identifies the **input-process-output (IPOS)** model, which consists of inputs, processing, outputs, and storage (referring to main memory, not disk drives). In either of these models, software and processing, while equally important to the other components of their respective models, represent the catalyst that allows us to accomplish something on a computer.

In the first three chapters of this book, we looked at the typical hardware of a network server, including its external and internal hardware and the hardware that's used for storing data. In this chapter, we will move on and look at the primary software of all computer systems—the operating system and, in particular, **network operating systems (NOSes)**. In this chapter, we will focus on the three most popular (and the ones you'll encounter in the Server+ exam)—Windows Server, Linux Enterprise, and macOS Server.

In this chapter, we will cover the following topics:

- The network server
- OS and hardware
- Boot sequence
- Filesystems
- Network configuration
- User accounts
- NOS optimization

The network server

First of all, let's agree on what a network server is and what its functions may be. As we discussed in *Chapter 1, Server Hardware*, a server is technically a software package running on a network computer that processes, responds to requests from hosts, and provides resources to fulfill requests from network clients. There are many different types of servers, each with its own purpose and function. In the following sections, we will look at the various servers you may encounter in the Server+ exam.

Server functions

The general function of a network server is to provide services to network clients. At its most basic level, a server, just like those in a restaurant, serves the needs of its clients. And just like the servers in a restaurant, there are specialized servers for different responsibilities, such as the wine steward, the cook, and the bus person, focused on one purpose or group of related purposes.

However, if we expand the definition of a network server to also include the hardware, operating system, management utilities, protocols, data stores, and communications, its function will include the provisions of administration, security, resources, and other services to network clients. A network server may support a single role, function, or application. A network-connected computer may support several server roles and applications.

Regardless of the server software running on the server hardware and the roles the server fulfills on a network, one thing is constant on every server—the NOS. Not every network node that supports one type of server software or another requires an NOS, but those servers that administer user accounts, access, and security certainly do.

Network server operating systems

There are several differences between what we call an *operating system* and a *network operating system*. The differences go beyond, but are directly related to, the inclusion of *network* in the title. Like nearly all computer operating systems, such as Windows, Linux, and macOS, an NOS provides system control and management functions, but to multiple workstations.

Operating system (OS) functions

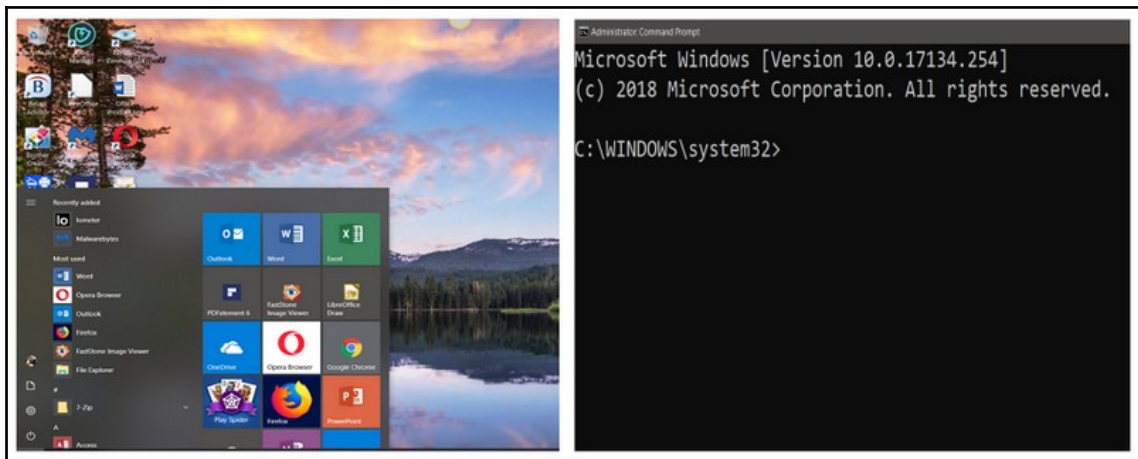
A computer's OS is the system software that provides the means for a user to make use of the computer hardware. An OS manages, controls, secures, and administers the physical electronics of a computer. For the most part, we tend to take the OS on our computers for granted, with little thought to how it works or just what it does. After all, it only makes everything we do possible!

An OS has five primary functions:

- User/computer communication
- Memory management
- Control and coordination of hardware
- Internal and network file management
- User, data, application, and resource security

User/computer communications

Users communicate with an OS or an application through a **graphical user interface (GUI)** or a **command-line interface (CLI)**. The facilitation of this communication is a primary function of any OS. The following screenshot illustrates both a GUI and a CLI through which users select or enter requests or commands:



Examples of a GUI (left) and a CLI (right)

Memory management

An OS, network or otherwise, has the responsibility of allocating memory to user or system-initiated programs. When a user starts a process or program that's not a part of the OS kernel, which means it's already in memory (more than likely), it's copied from secondary storage into an allocated space in main memory. The allocated memory space serves as a holding area for the program's instructions and data, before and after passing this to cache memory and the CPU.

Dynamic loading and linking

The OS typically loads smaller programs, those consisting of only one module, and data blocks often load to memory completely. However, larger programs, especially those with several modules, can't fit into the available memory to be allocated, so the OS uses a process called dynamic loading. **Dynamic loading** loads the first module of a program, the one with the first instruction, and then loads other modules into the same space when needed.

Another approach to memory management is **dynamic linking**. Many programs, especially those developed as **object-oriented programming (OOP)**, have an associated library of definitions, methods, and functions. As the program executes, the OS creates a link to any object invoked in a module to its definition in the library for execution. This avoids the need to load the entire library into memory.

Memory allocation

Memory allocation is the process that's used by the OS to allocate and assign memory space to a program. Generally, there are four approaches to the allocation process:

- **First fit:** The first memory space block in the memory available table that is large enough to fit the needs of the program or module
- **Next fit:** The next memory space block immediately after the last memory block allocated that fits the needs of the program or module
- **Best fit:** The smallest available block of memory large enough to meet the needs of the program or module
- **Worst fit:** Any available block of memory that is larger than the needs of the program or module

The memory management function of an OS may use any of the preceding memory allocation approaches. In any case, allocations are either static or a dynamic. Compilers often determine the **static allocations** of a program and its modules. **Dynamic allocations** are fluid. Dynamic loading and linking are two types of dynamic memory allocation.

Control and coordination of hardware

The OS is able to communicate with some devices, such as the keyboard, through the computer's BIOS, since there are standard devices that use a simple, standard command set. However, many devices and components have unique sets of commands and instructions that are typically not compatible with other devices. It's virtually impossible for an OS to include the commands needed to communicate with every possible make and model of internal and external devices installed on a computer. For this reason, device driver software, which is unique to a particular device make or model, acts as a go-between for the OS and the device controller. Device drivers make it possible for the OS to coordinate the actions of hardware components and devices.

Some hardware device drivers are in the OS, while some are in the **Basic Input Output System (BIOS)** and **Complementary Metal Oxide Semiconductor (CMOS)** and others are on the motherboard chipset. However, for the most part, device driver installation occurs either before or after the installation of the associated device on a computer. There are device software programs that assist the configuration of a device and the installation of the device driver.

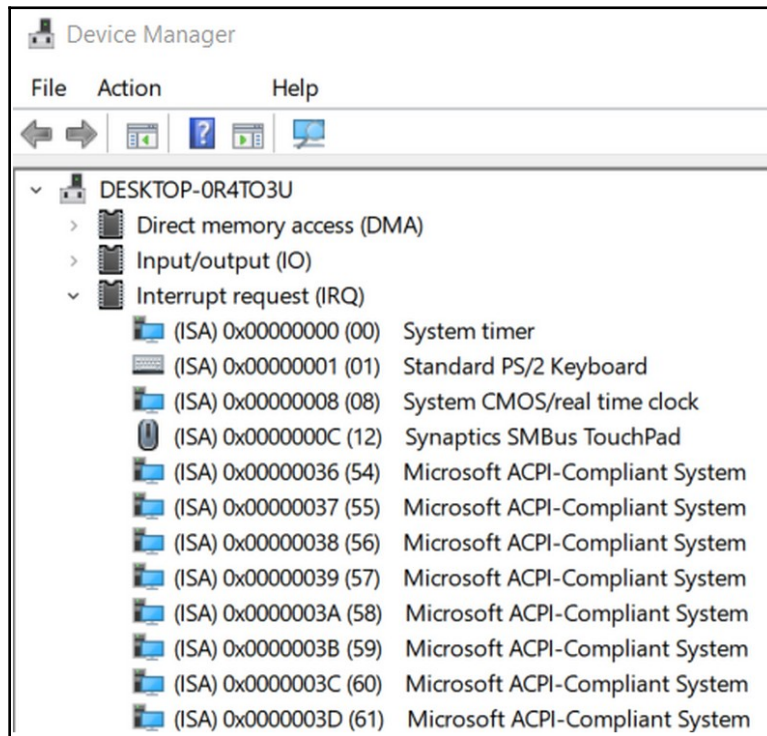
The use of system resources

The OS's communication with a device driver is a bit more involved than just passing data back and forth. *How does the device driver or the OS know that an action is needed by either? How do they communicate to each other to say that a requested action is complete and if data is involved, where is it?*

The most common answer to these questions involves **system resources**. A system resource is actually any addressable element of a computer. System resources are a portion of main memory (RAM) that is set aside for OS/device interactions. System resources can also be provided by the CPU, motherboard, chipset, and other system components. However, the system resources involved with device control and communication are those that are found in RAM.

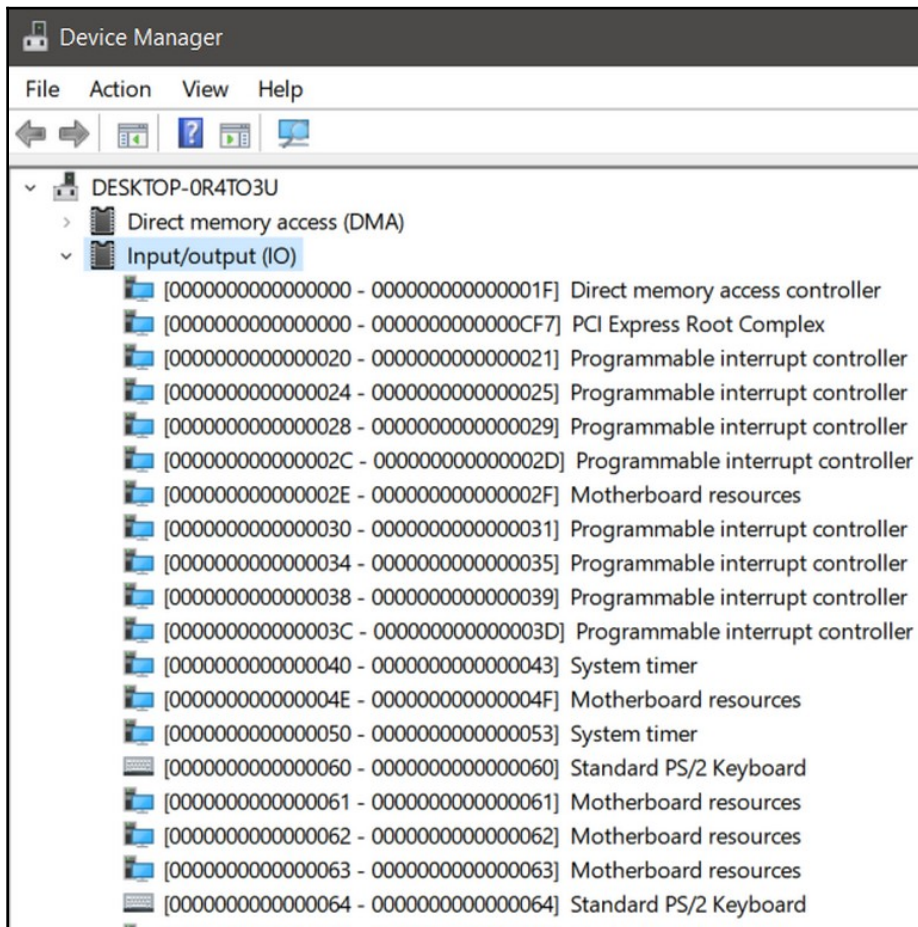
There are four categories of system resources:

- **Interrupt requests (IRQs):** In a classroom or a group meeting, if you wish to ask a question, you might raise your hand above your head and wait for recognition. An IRQ works essentially the same way—when a device or the OS needs the CPU to perform a task, it *raises its hand* by setting an assigned *IRQ on*. The CPU frequently checks the IRQs and if one is set to on, the CPU *interrupts* what it's doing to take care of the request, hence the name. After completing the task, the IRQ is set to *off*. The following screenshot shows the IRQs on a Windows system:



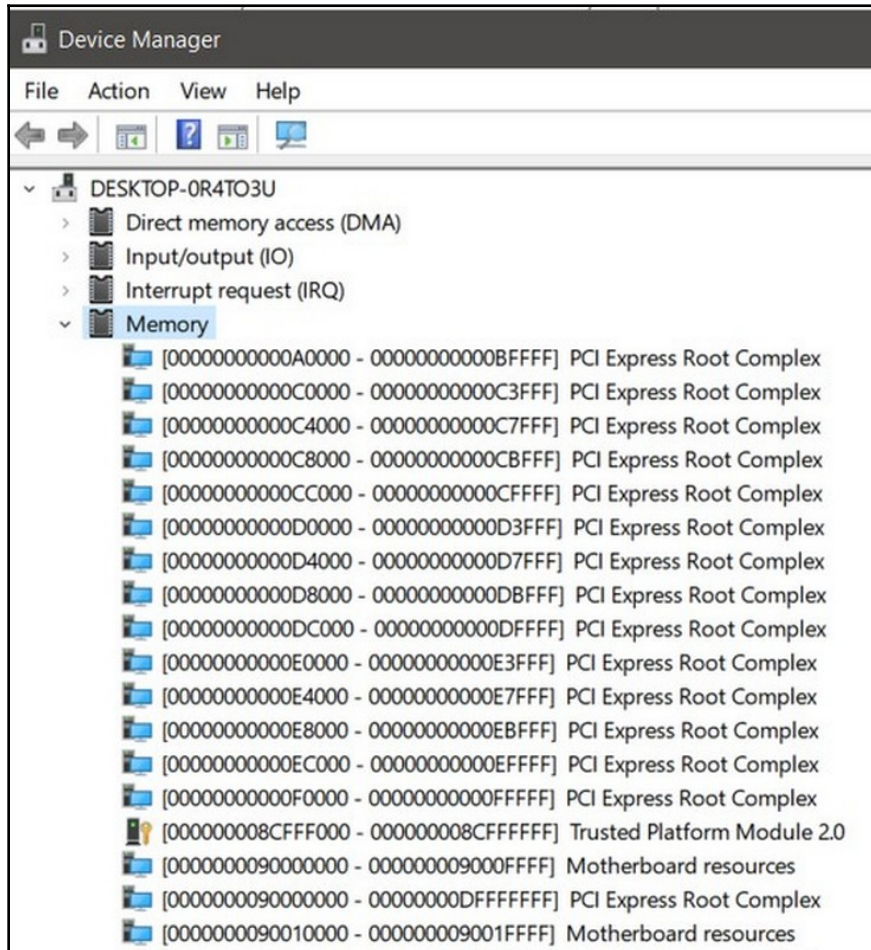
The IRQ assignments of a Windows system

- **I/O addresses:** Each installed I/O device on a computer has one or more addresses assigned to it. This address, which has several names, including I/O port, port address, or simply *port*, designates a device specifically and is not an address in memory. This address is like the street address on a house. The motherboard's address bus toggles between I/O addresses and memory addresses. When the bus is set to I/O addresses, the hardware device controllers monitor the bus for their individual addresses. If a device sees its I/O address, it responds to the request as appropriate. The following screenshot shows the I/O address assignments of a Windows system:



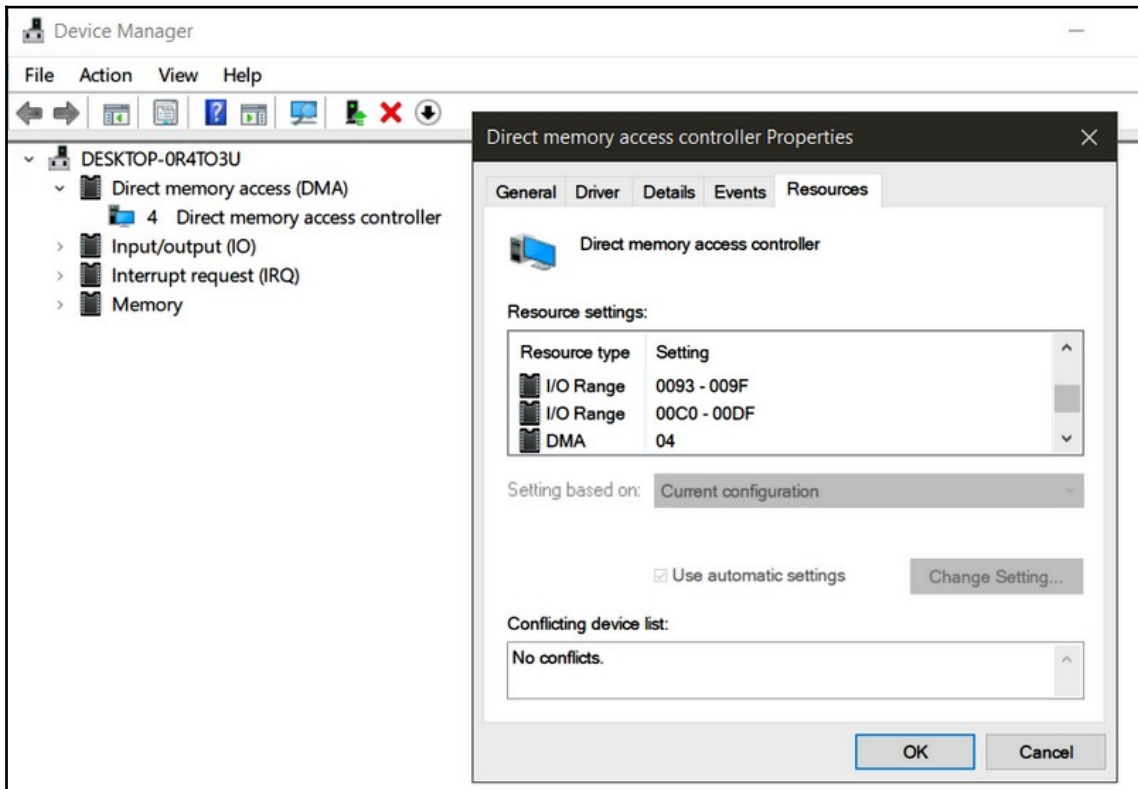
The I/O address assignments of a Windows system

- **Memory addresses:** Along with its I/O address, an I/O device may also be assigned a block of memory to use as a data buffer or as a scratchpad space. Not every device gets a memory address, and the space that is assigned is not all that large. As you can see in the following screenshot, some devices, both physical and logical, have more than one memory address assignment:



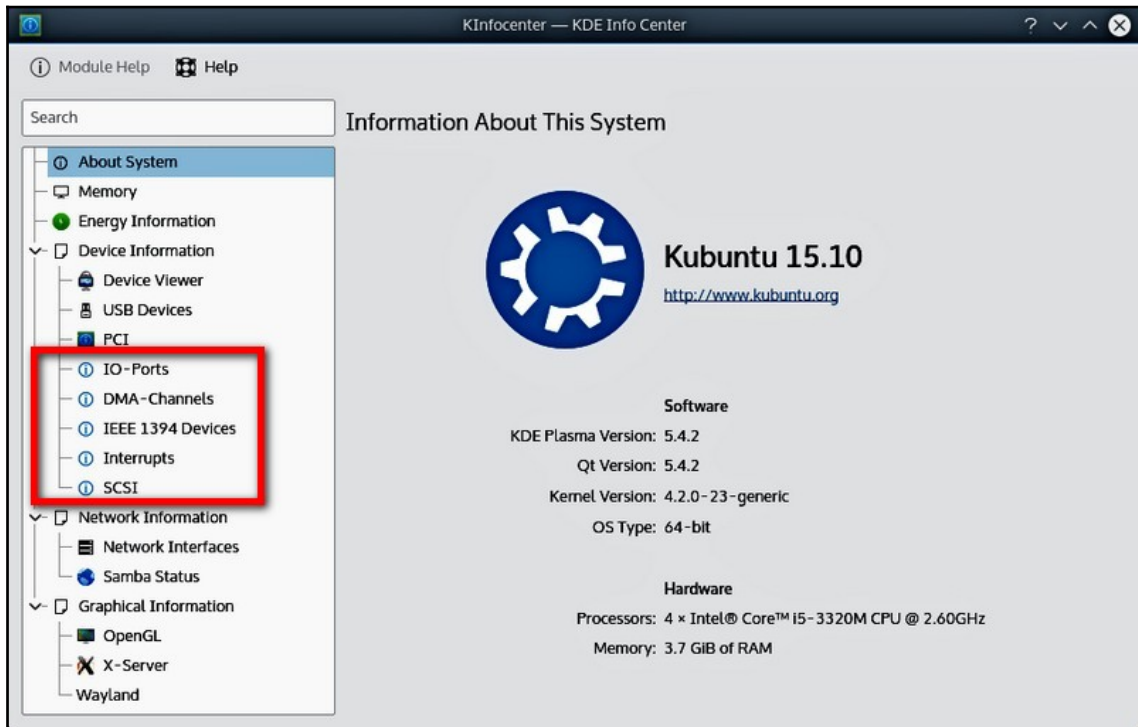
Memory address assignments of a Windows system

- **Direct memory access (DMA) addresses:** A DMA device is a computer I/O or storage device that is able to read and write data directly from or to main memory. DMA transfers don't involve the CPU, which can then take care of other tasks:



The Properties dialog box for a DMA device on a Windows system

The following screenshot shows an open source utility program that can be used to examine the systems information, including the system resources, of a Linux system:



KInfoCenter provides information about a Linux system, including the system resources

Internal and network file management

The management of data files, whether on an internal storage device or a NAS or SAN, is a primary function of an OS. File management involves the creation, modification, transfer, and removal of data units that are stored as a complete block. In performing file management, an OS interfaces with hardware controllers, driver software, and perhaps a data management system, such as a DBMS.

The tasks performed by the file management functions of an OS include the following:

- Creating new data files and recording their placement on the storage medium
- Providing for the modification of a data file and, if necessary, its relocation
- Performing the removal of a data file and any references to it
- Organizing data files into a filesystem, directory, or folder to facilitate its accessibility
- Facilitating access to a data file by multiple users

User, data, application, and resource security

The security provisions of an OS encompass the rules, functions, processes, and settings that the OS applies to implement and maintain the **confidentiality, integrity, and availability (CIA)** of its computer system. Organizational security policies must include the protection of the OS and its physical and environmental safety, including theft, damage, or destruction. However, the primary focus of OS security must be to enforce the rules and perform the tasks that prevent unauthorized intrusion or interference.

The security of an OS requires that certain activities are a part of normal operations, including the following:

- Performing patch and update management
- Applying antivirus and malware updates
- Examining all network traffic, inbound or outbound, through a firewall and/or other security appliances
- Regularly reviewing and managing user and group account permissions and rights

Hardware configuration

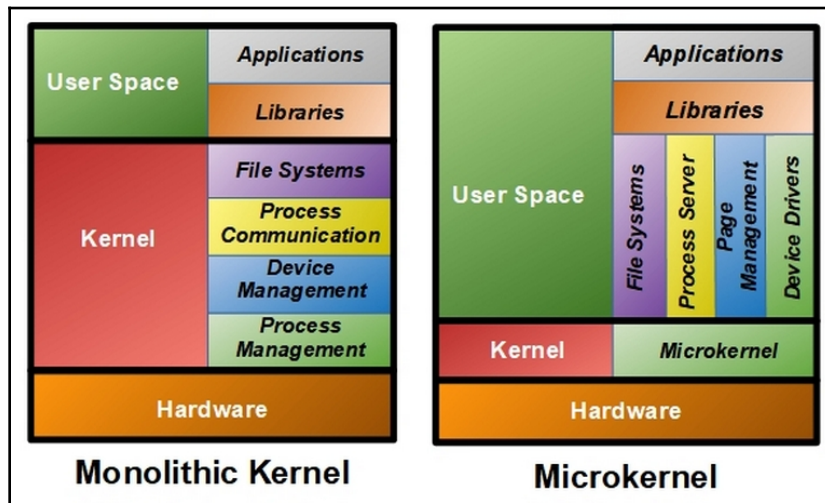
To understand the relationship between an OS and the configuration settings and specifications of the hardware, you need to understand that, by nature, an OS is sleek, and hardware is klutzy. A large part of an OS job is to make the hardware appear sleeker than it may actually be.

In the sections that follow, we look at the OS, its parts, and a look at its relationship with hardware.

The primary parts of an OS

An operating system has three primary parts:

- **Kernel:** Modern operating systems are modular, meaning that most essential services, such as memory management, I/O handling, and CPU interactions, are in the module that loads at startup and remains in memory—the **kernel**. Other functions and utilities load to memory as needed. As shown in the following diagram, there are two types of OS kernels—**monolithic** and **microkernel**. A monolithic OS, such as Windows, macOS, and Linux, includes the functions that are needed for the basic functions of the computer and user-initiated software. As you can see in the following diagram, a microkernel OS pushes support for user-initiated programs into the user space. At the present time, there are no microkernel systems being widely used:



The two types of OS kernels

- **Shell:** An OS shell is a program that the OS starts to provide a user interface, commonly as a CLI, as shown in the following screenshot. In UNIX/Linux systems, there are several shells, each with their own command set and functions, such as the **Bourne-Again Shell (Bash)**, the C shell, and the Korn shell. On a Windows system, the GUI displayed on the desktop represents a shell. The Windows Command Prompt also represents a shell:


```
[root@localhost init.d]# ls -l
    14 S01logging      13 S40network      12 rcS
    16 S10udev        15 S50dropbear
    17 S20urandom     11 rcK
[root@localhost init.d]# ls -l
total 28
    14 -rwxr-xr-x    1 root    root          649 May 25  2017 S01logging
    16 -rwxr-xr-x    1 root    root         1630 Jun 24  2017 S10udev
    17 -rwxr-xr-x    1 root    root         1321 May 24  2017 S20urandom
    13 -rwxr-xr-x    1 root    root          359 May 24  2017 S40network
    15 -rwxr-xr-x    1 root    root         1354 May 24  2017 S50dropbear
    11 -rwxr-xr-x    1 root    root          423 May 24  2017 rcK
    12 -rwxr-xr-x    1 root    root          408 May 24  2017 rcS
[root@localhost init.d]#
```

Commands entered at the command prompt of a Linux shell

- **Filesystem:** When you create a file and store it on a hard disk, you assume that you'll be able to find it in the future. A filesystem maintains the physical placement of a file or data block on a disk drive and keeps a cross-reference of the location and the filename in an index. Filesystems organize data storage in directories, folders, files, and objects. Commonly used filesystems are NTFS on Windows and ext3 and ext4 on Linux systems.

The OS and hardware

The configuration of a computer's hardware, software, and firmware specifies the parameters and settings for its functions and operations. The hardware configuration settings in a computer's BIOS or **Unified Extensible Firmware Interface (UEFI)**, OS, and support systems, define its startup, the installed devices, and several performance and operation parameters.

Not every program or application uses every hardware component available on a computer. The hardware any program uses will depend on its function and purpose. If a program only accepts two numbers from the user, adds them together, and prints the result to a display device without storing the result, the requirements of the program and the actions of the user affected the hardware used. In this particular case, the hardware involved was as follows:

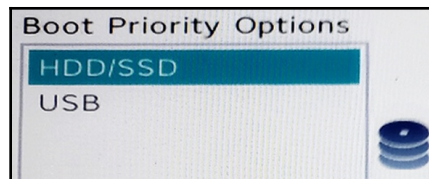
- **HDD:** The program loaded to RAM
- **RAM:** The program was assigned an allocated space and instructions were passed to the CPU

- **CPU:** The instructions were executed
- **Display:** The results were passed to the controller and displayed

Of course, the OS was the catalyst that guided this sequence, but in doing so, the OS involved only the hardware required to complete the task. So, *how does the OS know what hardware it has at its disposal?* That's what the BIOS/UEFI stores in CMOS. This information provides the OS with a list of the hardware devices it can use, and also how and where it can access each.

Boot sequence

A major part of the system startup of a computer, that is, its boot up, is loading the OS kernel into RAM so that the CPU can turn control over to the OS. To perform this process, the boot program must know where to look for the OS, which could be an HDD, a CD, a flash drive, or any bootable device. In the BIOS/UEFI settings, the sequence of devices that the boot process should look onto locate the **master boot record (MBR)** can be set. The boot utility will look at each of the devices in the order of priority and boot the computer from the first one it encounters with the boot information. The following screenshot shows a simple UEFI boot sequence list:



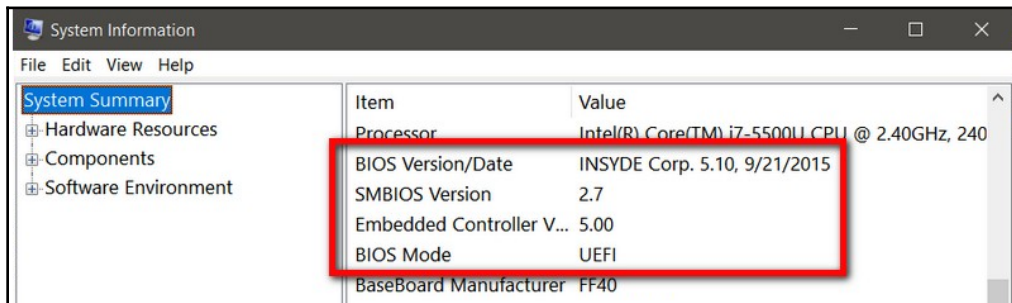
The boot priorities dialog box of a UEFI configuration

Firmware

Since the beginning of computer time, we have classified computer instructions and components by where they reside or their general characteristics. The components that we can touch and feel are *hardware*. Those we can't actually touch, feel, or hold, but can change, we call *software*. However, in computer systems, there is a component we can't touch, feel, or change, especially in older systems. This component exists somewhere between hardware and software, and we call it *firmware*.

What we refer to as firmware today can be one of two general technologies:

- **BIOS:** A legacy technology that uses data and instructions that are permanently loaded (*burned*) into a semiconductor during manufacturing. These instructions initiate the BIOS to begin the startup process and load the bootloader to complete the boot process. This low-level form of firmware is a part of the motherboard as a **read-only memory (ROM)/programmable read-only memory (PROM), one-time programmable (OTP), and programmable logic array (PLA)** chips.
- **UEFI:** This newer technology is replacing most BIOS on computers, but performs essentially the same functions. In fact, UEFI relies on BIOS for the **power-on self-test (POST)** function and the configuration specifications (commonly referred to as CMOS). Virtually all computers produced after 2010 have UEFI, with some also including BIOS. The following screenshot shows **System Information** with the BIOS/UEFI settings:



The BIOS/UEFI settings on the System Information dialog box

In spite of its *permanent* nature, the firmware on most computer systems is upgradeable. The exact procedure varies with the motherboard manufacturer, operating system, or the age of the system, but the following steps are generally the process that's used to update system firmware:

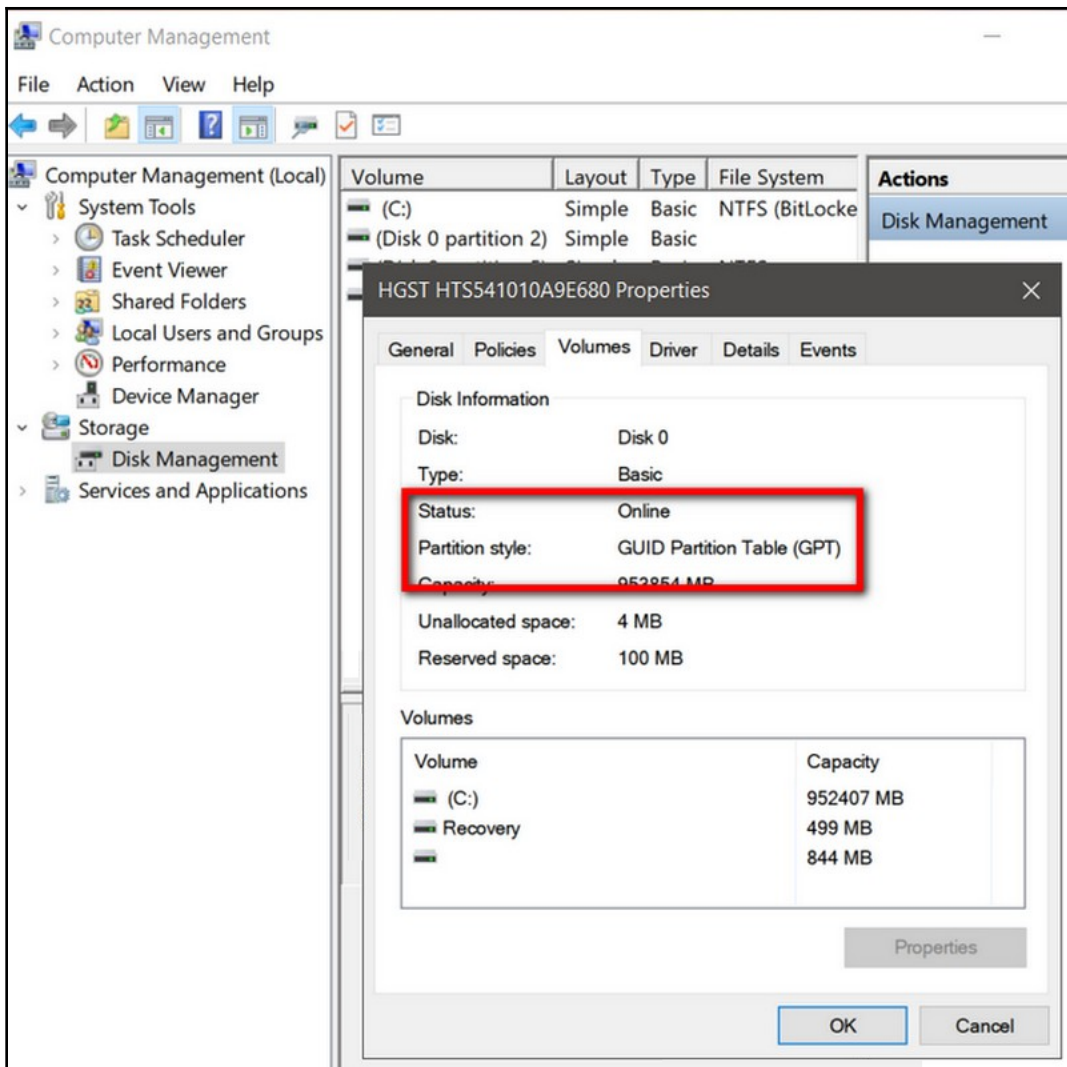
1. Verify that you are actually installing a newer version of the UEFI/BIOS. On a Windows system, run `MSINFO32`. On a Linux system, if the `/sys/firmware/efi` directory exists, the system is using UEFI. If that directory doesn't exist, the system is using BIOS. Record the version number and date and compare it to the information on the motherboard manufacturer's website.
2. Boot the computer and when the display informs you of which key to press to access the UEFI/BIOS settings, press it. Some systems include a firmware update function in the UEFI/BIOS utility.

3. If no update utility is available, download and decompress the update file and store it on a USB flash drive or external disk drive.
4. Restart the system and access the firmware update or flashing utility on the UEFI/BIOS settings page. Back up the existing firmware to an external drive, just in case something goes wrong in the update.
5. Using the firmware update utility, select the newer version image and start the update. The upgrade should only take a few minutes, but, in any case, absolutely don't restart, shut down, or power off the computer until the upgrade process completes.

Preparing a disk for the OS

There are two approaches to installing a new version of your existing OS or replacing it with another OS altogether. If you are staying with the same OS and just installing a newer version, such as upgrading Windows 10 to Windows Whatever, you have the choice of performing an update, or what's called a **clean install**. Essentially, a clean install removes any obsolete or out-of-date elements of the old version and replaces them with the newer versions. A clean install can also mean that the HDD is new and that the installation is on a *clean* disk drive. An upgrade from one version to another means that a clean install isn't necessary and that the installation will use an installer or setup program from the publisher. On the other hand, if you are replacing one OS with another, such as replacing Windows with Linux, there are things you should know about and a few steps you should take before installing the new OS.

The method that's used to prepare an HDD depends on the system configuration and boot process in use. BIOS creates a MBR and UEFI creates a **GUID partition table (GPT)**. An MBR creates legacy BIOS partition tables and a GPT creates UEFI partition tables. GUID is a Microsoft term for **Globally Unique Identifier** and is applicable only to its systems. The remainder of the OS world uses the term **Universally Unique Identifier (UUID)** for essentially the same structures. A GPT (UEFI) system is able to define and create more than four partitions on an HDD and is a must if the disk drive is 4 TB or more. Below those thresholds, the MBR (BIOS) system will work, if present. The following screenshot shows the partition table type for a Windows system HDD:



The Properties dialog box showing the partition style of an HDD

The following steps are for the general process of preparing a disk for the installation of an OS:

1. Verify that the computer on which the installation will occur meets or exceeds the system requirements published by the OS publisher.
2. If the HDD has data on it that you wish to preserve, take a total system backup to an external medium, such as an external HDD or a cloud service.

3. If you wish to remove all previous data and content from the HDD on a Windows system, use the **Disk Management** format option. On a Linux system, use the `fdisk` command.
4. Create the disk partitions needed for the installation using either the **Disk Management** utility or the `DISKPART.EXE` command at the Windows Command Prompt. The following screenshot shows an example of the `DISKPART` commands. The `fdisk` command is used on a Linux system.
5. After creating the partitions, format them to the appropriate partition table standard:

```
Microsoft DiskPart version 10.0.17134.1

Copyright (C) Microsoft Corporation.
On computer: DESKTOP-0R4T03U

DISKPART> LIST DISK

   Disk ###  Status       Size       Free      Dyn  Gpt
   -----  -
   Disk 0      Online         931 GB     1024 KB             *

DISKPART> SELECT DISK 0

Disk 0 is now the selected disk.

DISKPART> LIST PARTITION

   Partition ###  Type              Size       Offset
   -----  -
   Partition 1     Recovery          499 MB     1024 KB
   Partition 2     System            100 MB       500 MB
   Partition 3     Reserved           16 MB       600 MB
   Partition 4     Primary           930 GB       616 MB
   Partition 5     Recovery           844 MB       930 GB

DISKPART>
```

The display of the DISKPART.EXE command

Filesystems

When you store data on an HDD, USB-connected storage, or any other data storage device, it needs to have an organizational scheme so that individual files can be located in the future. A filesystem organizes and structures a storage medium and tracks the files stored on it. Along with location, a filesystem also catalogs identifying data for each file, including filename, size, status, creation and modification dates, access permissions, ownership, file type, and more.

Formatting

The placement of a filesystem on a data storage device, which may have several partitions or just one big one, happens through the process of formatting. Each partition may have a different filesystem. Formatting a drive generally does three things: recognizes a partition as a bounded structure; removes (erases) all existing data and indexing, if any, from the partition; and initiates a filesystem and its indexing in the partition. Each of the major operating systems has a utility to format a disk drive partition: **Windows Disk Management**, **Linux's GParted**, and **Mac OS Disk Utility**.

Filesystems by OS

In the Server+ exam, you will encounter questions and references about filesystems of Windows, Linux, and macOS. The filesystems that are included are as follows:

- **Windows:**
 - **File Allocation Table 32 (FAT32):** At one time, this was the default standard for Windows systems. It was used primarily for the format of flash memory drives.
 - **New Technology Filesystem (NTFS):** The default filesystem for Windows systems. The Windows system partition must be NTFS.
- **Linux/Unix:**
 - **Better Filesystem (Btrfs):** Adds pooling, snapshots, checksums, and other features to Linux.
 - **Extended Filesystem versions 2, 3, and 4 (ext2, ext3, and ext4):** A filesystem based on the **Unix Filesystem (UFS)** that tracks individual files.

- **Reiser Filesystem (ReiserFS):** A journaling filesystem for Linux.
- **Z File System (ZFS):** Originally developed for the Solaris OS, it is common for Linux systems to support NAS.
- **Mac OS:**
 - **Apple File System (APFS):** Replaces the HFS+ operating system on later Mac OS X systems.
 - **Hierarchical File System Plus (HFS+):** The de facto standard filesystem of older Mac OS systems.

Journaling

There are filesystems that perform journaling and some that don't, though most of those are legacy filesystems. A journaling filesystem records filesystem changes before applying them to the medium of a *journal*, which is essentially a filesystem activity log. Typically, a journal is on a separate device to the filesystem. The journal file provides recovery data should the filesystem be damaged. Examples of journaling filesystems are NTFS, ReiserFS, ext3, and ext4.

Special function filesystems

Sometimes, a filesystem really isn't a filesystem, but may perform some filesystem functions. The **Virtual Machine File System (VMFS)** works with a virtualized system's hypervisor to store and manage virtual machine snapshots and images. Swap, also known as **swap space**, is a space in a dedicated partition of a secondary storage device that the OS uses to store inactive memory pages to free up memory resources, as needed. The **Flash-Friendly File System (F2FS)** is an open source flash drive filesystem.

Network configuration

An essential part of installing and activating any NOS is configuring it for the network it's to support. In the following sections and steps, we will look at the processes that are used to configure a Windows Server system and a Linux system for some important settings and values.

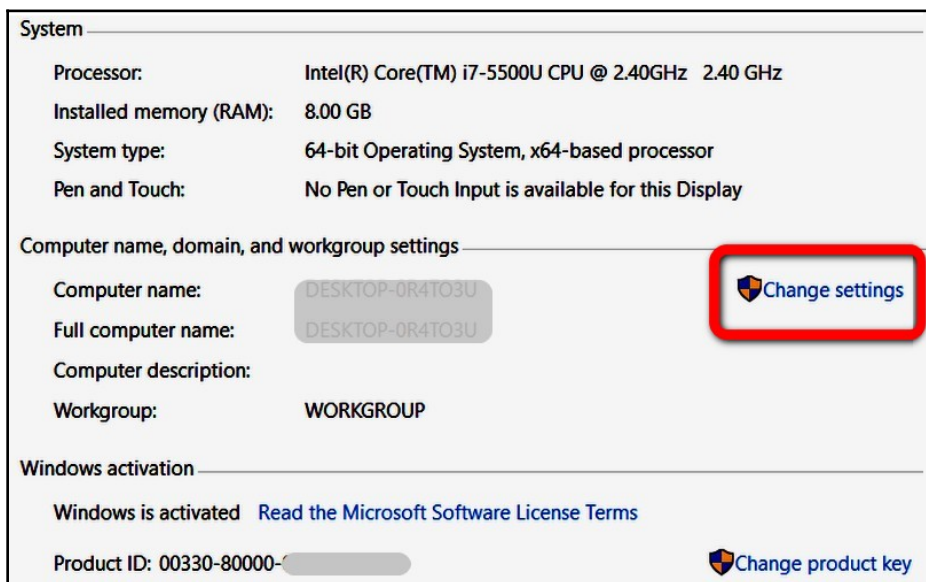
Configuring the hostname

The hostname of a server (or any node on a network) identifies the node to a computer network in communication and access to other nodes and off-network devices and services. In effect, the hostname of a device is its nickname.

Configuring a hostname on Windows Server

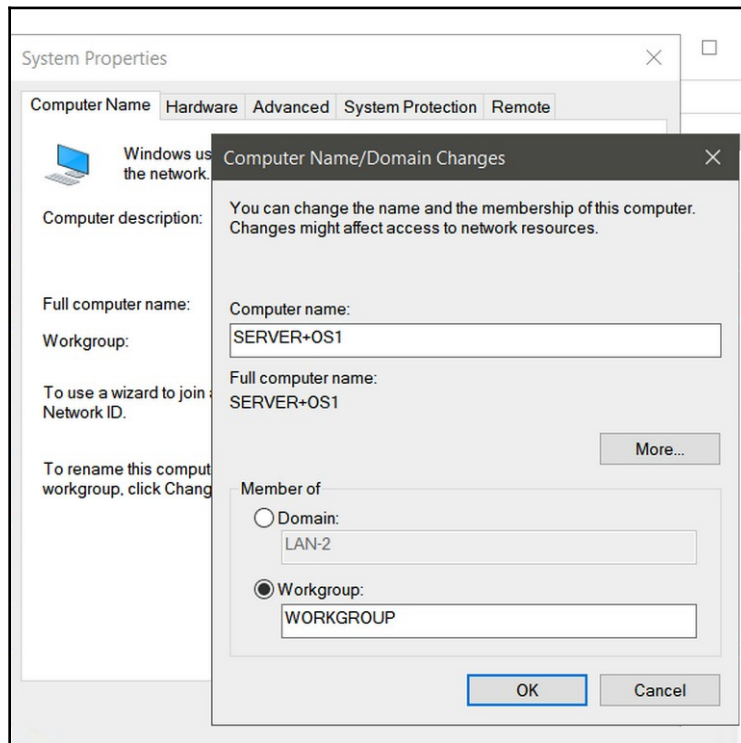
To set the hostname of a Windows Server installation, follow these steps:

1. Click on the Windows icon to display the Start menu.
2. Right-click the **This PC** option to display the **System** window.
3. On the right-hand side of the **System** window, click on the **Change settings** button to display the **System Properties** dialog box:



The Properties window of the This PC selection

4. Enter the name you wish to assign to the server in the **Computer name** textbox. The only hard rule is there can be no spaces in the name:



Assigning a hostname to a system

Configuring a hostname on a Linux server

On a Linux server, there are two hostnames to configure: a network-related hostname and a local hostname. The steps used to set the hostnames are as follows:

1. Log on to the system as the root administrator (the superuser).
2. Move to the `/etc/sysconfig` directory to locate the network file and display its contents, which should look something like the following:

```
# cat network
NETWORKING=yes
NETWORKING_IPV6=no
HOSTNAME=localhost.localdomain
```


The `cat` command displays the contents of the file `network`. Notice that the last line of the file contains the hostname setting.

3. Use a Linux editor (`vi` or `edit`) to change the `HOSTNAME` value to the hostname you wish to use. After you've completed the edit, save the file.
4. The edited file should now contain the revised hostname value. For example, the network file may now contain something like this:

```
# cat network
NETWORKING=yes
NETWORKING_IPV6=no
HOSTNAME=server+os1
```

5. The hostname in the `localhosts` file also needs to be set. Move to the `/etc` directory to find the `hosts` file. Its contents should be along the lines of the following:

```
# cat hosts
127.0.0.1 localhost.localdomain localhost
::1 localhost6.localdomain6 localhost6
```

6. Using an editor, replace the `localhost.localdomain localhost` value with the new hostname and save the file. After the change, the `hosts` file should contain something like this:

```
# cat hosts
127.0.0.1 server+os1
::1 localhost6.localdomain6 localhost6
```

7. You may want to also set the Terminal identity to the hostname of the server. The local network uses the Terminal name to address the server. To change the Terminal's hostname, use the `hostname` command:

```
# hostname server+os1
```

To verify the change, use the `hostname` command with no parameters.

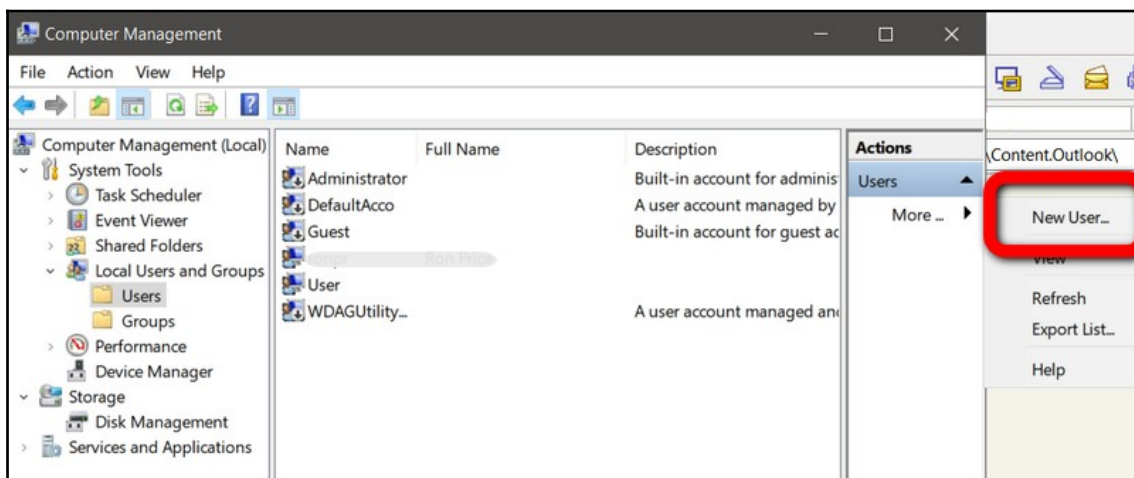
User accounts

On a Windows Server system, a user may have one or two account types: a local account or a domain user account. A **local user account** limits the user to only the resources for which file and folder permissions grant access to the individual or to any group the individual is a member on a single network node. A **domain user account** can access either local or network (or both) resources, according to the access permissions assigned to the individual or to the groups the individual is a member of.

Creating a local user account

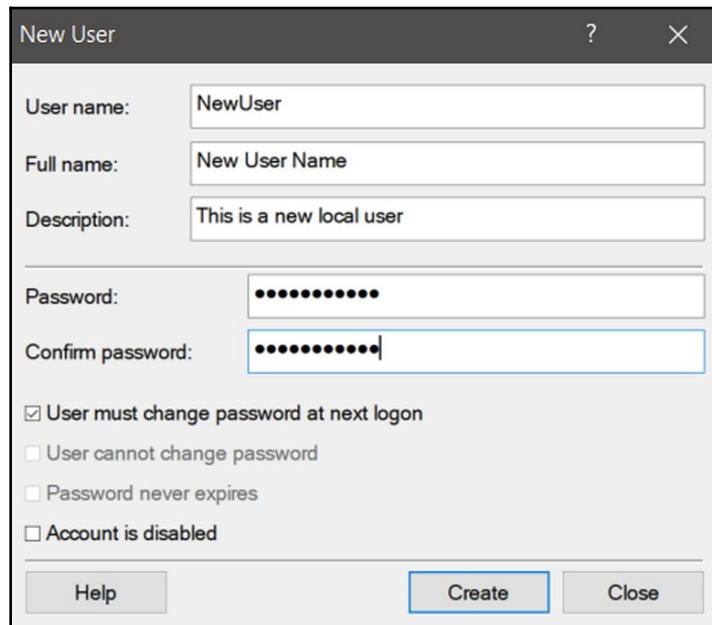
To set up a local user account on a Windows Server, follow these steps:

1. Right-click the Start icon or press the Windows key + x to open the Start right-click menu.
2. Select the **Computer Management** option to display its window.
3. In the left-hand navigation pane, click on **Users** under the **Local Users and Groups** option. The view shown in the following screenshot will appear:



Adding a new user through the Computer Management window

4. In the right-hand **Actions** pane, click on **More...** to open an options list. From that list, click on **New User...** to open the **New User** dialog box:

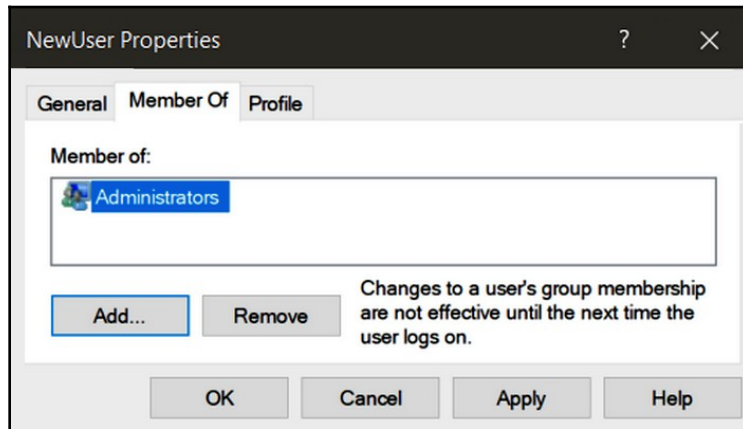


The New User dialog box

5. Fill in the new user's information and click on the **Create** button. The dialog box will clear and be ready to use so that you can add additional users. If finished, click on **Close**. Notice that the new user is now in the list of users in the Computer Management window. Remember that there's no way to recover a password on a local user account, so you may want to warn the user about using a password they will remember.

If the new user is to have administrator privileges or belong to any other group, use the following steps to add the user to the applicable groups:

1. Right-click the name of the new user in the center pane of the **Computer Management** window and choose **Properties** from the option list that appears to display the **Properties** dialog box for the user:



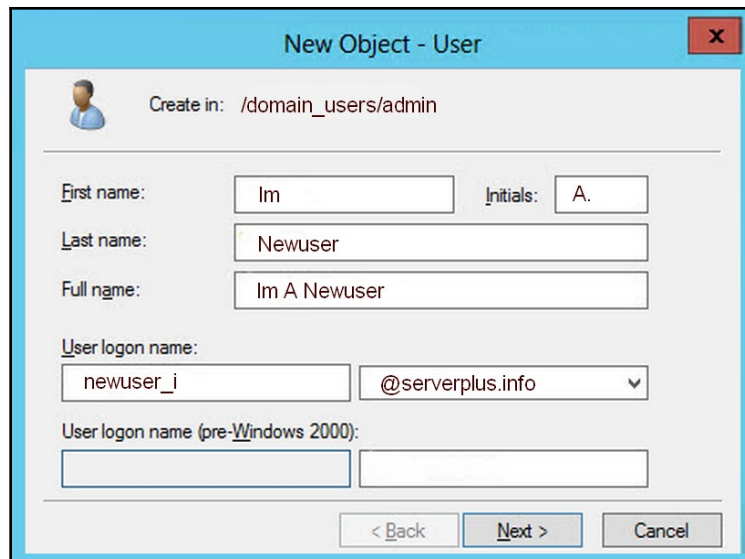
Adding a user to a group on the user's Properties dialog box

2. Click on a group name, in this case, **Administrators**, from the **Member of:** list and click **Add**. You can repeat this action to add the user to more groups.

Creating a domain user account

The creation of a domain user account assumes that at least one domain exists on a network. In the Windows world, a **domain** is a network where a database on a central domain controller contains information on all domain user accounts, security configuration, and the hardware devices on the network. The network's **Active Directory (AD)** service manages domain user accounts. To create a new domain user account on a Windows Server system, follow these steps:

1. From the Start menu, open **Server Manager** and pull down the **Tools** menu.
2. On the **Tools** menu, select the **Active Directory Users and Computer** option to display its dialog box.
3. In the left-hand navigation pane, click on the **Users** folder. On the option list that appears, click on **New**. On its option list, click on **User** to display the **New Object - User** dialog box, as shown in the following screenshot:



The screenshot shows the 'New Object - User' dialog box. The title bar is blue with a red close button. Below the title bar is a user icon and the text 'Create in: /domain_users/admin'. The main area contains several text boxes and a dropdown menu. The 'First name' box contains 'Im', the 'Initials' box contains 'A.', the 'Last name' box contains 'Newuser', and the 'Full name' box contains 'Im A Newuser'. The 'User login name' box contains 'newuser_i' and the dropdown menu shows '@serverplus.info'. The 'User login name (pre-Windows 2000)' box is empty. At the bottom are three buttons: '< Back', 'Next >', and 'Cancel'.

The New Object – User dialog box

4. After filling in the new user's information, click the **Next** button to move to another **New Object – User** dialog box and enter the user's password. Click **Next** to move to the **New Object – User** summary dialog box.
5. If all of the information that's shown is correct, click on the **Finish** button. If you need to make corrections, use the **Back** button.

Adding a workstation to a domain

To communicate on any network, a PC must join the network, which means that the network interface in the PC must connect to and join a network server's administrative systems. On a Windows system, the PC joins a domain. The process to add a PC to a domain is as follows:

1. On the **Control Panel**, click on the **System** option and choose **About** from the options in the left-hand side pane.
2. Clicking on the **Join a domain** button opens the first of the domain identification dialog boxes.
3. You will see three or four **Join a domain** dialog boxes in the next step. The first dialog box, as shown in the following screenshot, asks you for the name of the domain to which you wish to add the PC. If the domain name is valid, enter the username and password that you (personally) use to log on to the domain. While the system is authenticating the information you've entered, a blank copy of the first screen may be displayed:



The domain name dialog box

4. If the information that you've entered is valid, the next dialog box asks for the username of the person that will use this PC and the account type for the user, such as administrator, user, standard, and so on.

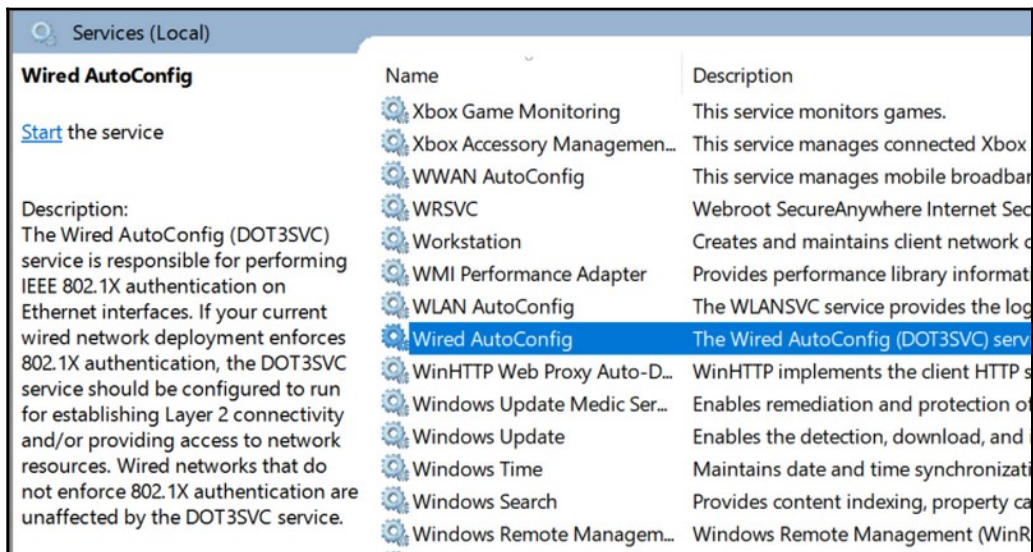
Connecting to a network

It's one thing to add a PC to a network domain and quite another to actually connect a PC to the network for active communication. Once a PC is in a domain, the next step is to verify that the electronics all synchronize and interoperate. In other words, *can the PC talk to the network?*

Of course, the medium used to communicate over a network does have an impact on the setup of network communications, as you might expect. However, you'd be surprised at how little of a difference this actually makes. Both physical cable and **radio frequency (RF)** communications are predominantly Ethernet networks.

Connecting a PC to a network

To ensure that a PC will connect to a network, there are a few steps to verify that the necessary elements are active before testing its communications. The first element to check is its connection service. A cable-connected network needs the **Wired AutoConfig** service's startup type, as shown in the following screenshot, set to automatic. If the workstation is connecting to a wireless local network, the **WLAN AutoConfig** service is set, or if the workstation is connecting to a wireless wide area network, configure the **WWAN AutoConfig** service. Whichever one of these three network services is activated, the other two should be deactivated:



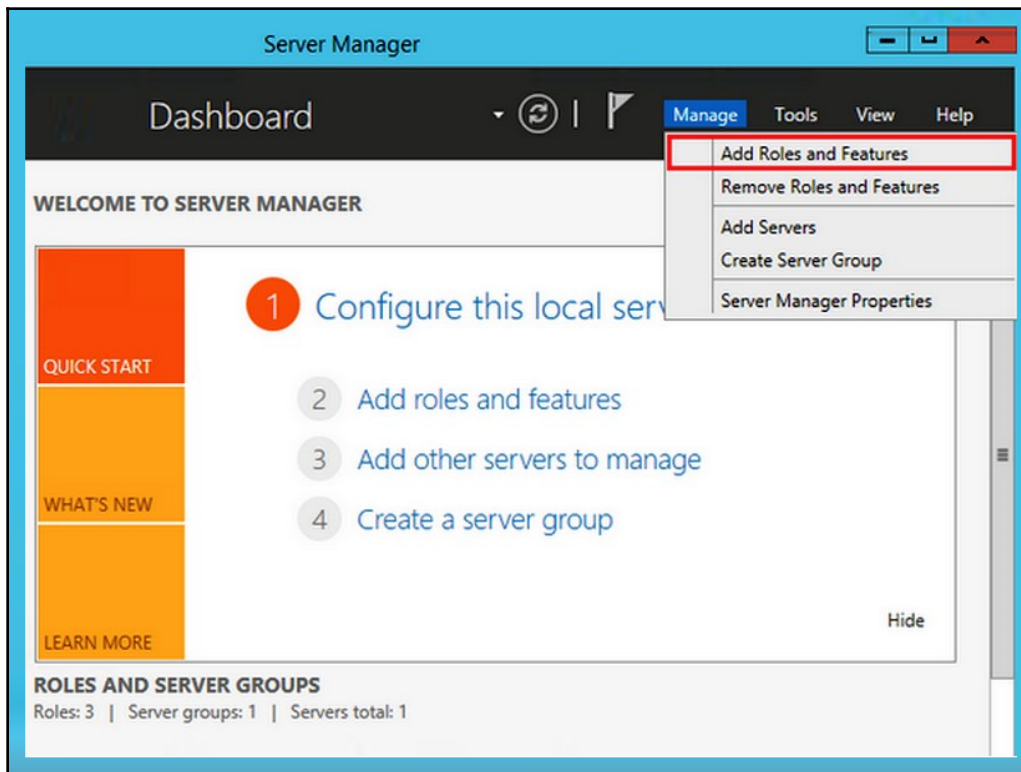
The Windows Services maintenance window

Adding server roles and features

A server role is one or more programs or sets of utilities that a server can use to provide a specific feature or function to a user, group of computers, or an application. Server roles and features set the purpose or role of a server. For example, a Windows server may have only a single role or be configured as an AD domain server, or it may have several different roles.

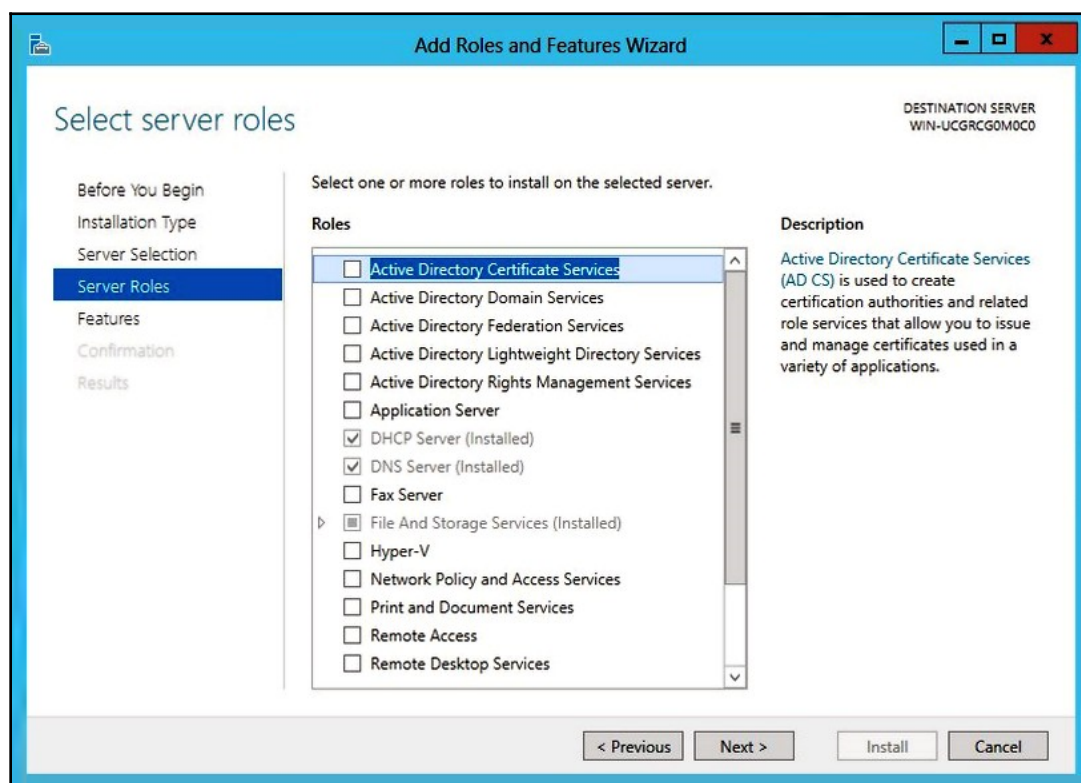
To add or remove roles and features from a Windows Server system, follow these steps:

1. On a Windows Server system, open **Server Manager**.
2. Click on the **Manage** menu option and select **Add Roles and Features**, as shown in the following screenshot, to open the **Before you begin** page of the **Add Roles and Features Wizard**. This page lists a series of tasks you should consider doing before adding, changing, or removing server roles and features:



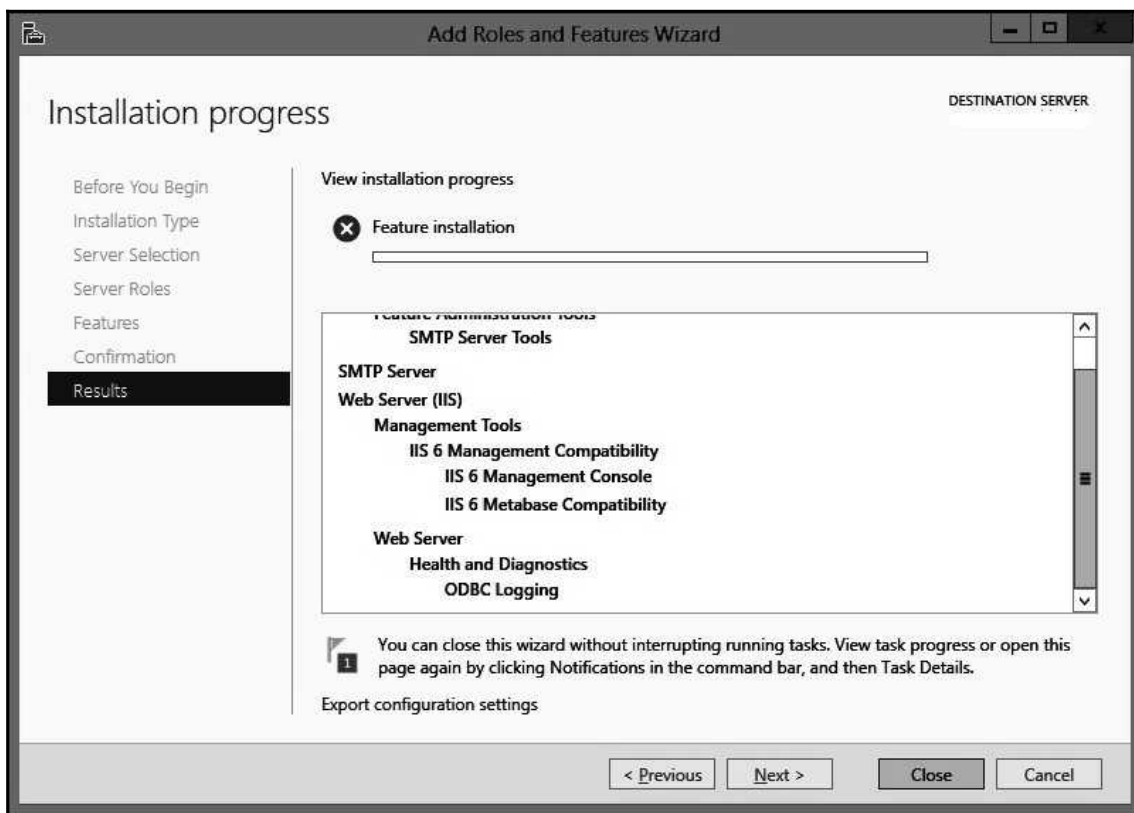
The Server Manager page of a Windows 2012 R2 system

3. Click **Next** to move to the **Select installation type** page. Select the **Role-based or feature-based installation** option. Click the **Next** button to advance.
4. On the **Select destination server** page, click on the **Server Selection** option in the left-hand navigation pane and then select either a server or a virtual hard disk from the network.
5. Click **Next** to advance to the **Select server roles** page, as shown in the following screenshot. In the main area of this page is a list of the roles and features to assign to the server or virtual disk chosen in the previous step. After you've completed your selections, click **Next** to move to the **Select features** page. Depending on the roles selected, a **Roles services** page may display first:



The Select server roles page of the Add Roles and Features Wizard

6. On the **Select features** page, select any features you wish to add to the target server. When you've finished, click the **Next** button to display the **Installation progress** page, which includes a summary of the roles and features selected, as shown in the following screenshot. If the list is correct, click the **Next** button to move to the **Install** page and begin the installation process:



The Installation progress page of the Add Roles and Features Wizard

7. When the installation completes, click on **Close** to return to the **Server Manager**.

Unattended and remote installations

Network administrators often need to access a remote computer for a variety of reasons. One of the more common tasks facing network administrators is the upgrade of software, including the OS, or running diagnostic utilities. In the past, these tasks required what the communication industry calls a *truck roll*, or a physical visit to the remote location. If remote means only across the building, there isn't usually too much of a problem. However, if remote means in another state, country, or continent, the time and cost required to affect changes can really grow. A number of different standards, software, and technologies are available to allow the network administrator to make these changes without the need to travel any further than from his or her desk to a network workstation.

The first of these tools is Intel's **Wired for Management (WfM)** specification, which allows a network administrator to create an automatic procedure to accomplish client maintenance and management over a network. You can also work with other tools, such as the **LANDesk Client Manager (LDCM)**, a software tool that monitors network clients for hardware issues, and the **Desktop Management Interface (DMI)**, which is an **application programming interface (API)** that allows software to inventory a network client, such as the details of which motherboard, expansion bus cards, and application software it has.

Another tool that WfM works with is the **Preboot Execution Environment (PXE)**, pronounced *pixi*. PXE enables a network administrator to boot a network client from a server by interfacing only with the client's PXE-enabled **network interface card (NIC)** or network adapter. PXE clients connect through their NICs to a network, even when the PC's power is off. The PXE client need not have an operating system, nor any other software for this remote boot to take place, even if it's on the client. The network administrator can then use a tool such as **Trivial File Transport Protocol (TFTP)** to transfer software or data to the client.

Windows systems and Linux systems, for the most part, provide a PXE **network boot program (NBP)**. Windows systems provide the **Remote Installation Service (RIS)**, which incorporates the PXE standard. Linux systems support the **PXELINUX** and the **gPXE** utilities.

NOS optimization

Network users expect the network to meet their individual needs in terms of bandwidth, speed, and responsiveness. These three expectations are parts of the same whole to the user—a network that performs consistently to provide the service that's expected. A network administrator can use a variety of processes, techniques, and methods to promote efficiency and effectiveness on a network at any level.

Some of the ways that the network administrator can ensure a network always meets its users' needs include the following:

- **Bandwidth:** Network administrators fully understand that a 10 Gbps network doesn't mean that users, individually or collectively, will realize that much bandwidth. Every network must consider the bandwidth loss from just providing a working network. To users, the network administrator's primary job is to keep bandwidth high for everyone.
- **High availability and fault tolerance:** As we discussed earlier in this chapter, HA and fault tolerance can be the key characteristics of a network that users depend upon. How much of the 525,600 minutes of uptime possible in a year a system is designed to achieve is an important design goal.
- **Load balancing:** On larger networks and data centers in which servers are clustered or distributed, load balancing can accept multiple input sources and distribute them to multiple processors, evening out the processing of the load and the outbound traffic.
- **Quality of Service (QoS):** Although its name reflects how users measure a network's performance, QoS is the ability of a network to provide a variety of services over a variety of communication methods at a level that meets the needs of the organization. A network that supports data, voice, and video consistently has QoS.

Summary

An NOS provides system control and management functions to network clients and resources and provide administration, security, resources, and other services to network clients. An OS has five primary functions: user/computer communications, memory management; control and coordinate hardware; internal and network file management; and user, data, application, and resource security. There are four categories of system resources; IRQs, I/O addresses, memory addresses, and **direct memory access (DMA)** addresses.

The file management functions of an OS include creating new files, providing for file modification, performing file removal, organizing files for accessibility, and facilitating access for multiple users. The security of an OS requires patch and update management, antivirus and malware updates, the examination of all traffic by a security appliance, and the regular review of user and group account permissions and rights. An operating system has three primary parts: the kernel, the shell, and the filesystem. A filesystem organizes a storage medium and tracks the files stored on it. It also catalogs identifying file data for each file. Formatting a drive recognizes a partition as a bounded structure, removes existing data and indexing, and initiates a filesystem and its indexing in the partition.

The hardware configuration settings in BIOS or UEFI, OS, and support systems, define startup, installed devices, and performance and operation parameters. Firmware is one of two general technologies: BIOS and UEFI. WfM allows an automatic procedure to do client maintenance and management over a network. DMI allows software to make an inventory of the hardware and software of a network client. PXE enables a network administrator to boot a network client remotely. TFTP transfers software or data to a remote client.

Questions

1. What is the basic function of a network server?
 1. Communicating with the internet
 2. Providing services to network clients
 3. Intrusion detection and prevention
 4. Providing proxy server functions
2. Which of the following is not a function of a network server?
 1. Administration
 2. Security
 3. Caching
 4. Resource allocation
3. Which of the following is not a system resource that's managed by an operating system?
 1. IRQ
 2. I/O address
 3. Device driver
 4. DMA address

4. The file management functions of an OS include the creation, modification, and removal of data files and what else?
 1. File accessibility
 2. Limiting access to only a single user
 3. Applying encryption to all files
 4. Converting filenames
5. An operating system has three primary parts. Which of the following is not one of those parts?
 1. Device drivers
 2. Kernel
 3. Shell
 4. Filesystem
6. What are the two most common firmware system configuration systems?
(Choose two)
 1. CMOS
 2. BIOS
 3. FAT32
 4. UEFI
 5. APFS
7. A filesystem configured on a Linux system could be which of the following?
 1. NTFS
 2. APFS
 3. BTRFS
 4. HFS+
8. Which identification identifies a server for the purposes of communication and access to network resources?
 1. User account name
 2. Hostname
 3. Server name
 4. Group account name

9. On a Windows Server network, the network administrator may grant a domain user which access level? (Choose all that apply)
 1. Local computer resources
 2. Server-based resources
 3. Network-attached resources
 4. WAN resources
10. A network administrator wishes to power up and boot a remote PC attached to an organization's network. Which of the following services will enable this action?
 1. QoS
 2. DMI
 3. PXE
 4. TFTP

5 Addressing

You'd have a difficult time locating a particular address in a large city using only a general description of the house in question. In the same way, a computer's **operating system (OS)** would have problems finding a data file for you if all you know is what's in it.

Back in the day, when secondary storage was small, it was relatively easy to find files on a floppy disk or hard disk. Connecting to another PC over a peer-to-peer network was straightforward and didn't require much in the way of an address—just the internal address of a parallel port.

However, on today's network and storage technologies, with their immensity, addressing on all levels is an essential part of computing.

In this chapter, we look at the variety of addressing schemes used in computing and networking—their structures, representations, and purposes. We also look at the interface between a computer and a network, and the protocols and services that provide their connection, transmission, and security.

We take a outside-in approach to addressing, beginning with **Internet Protocol (IP)** addressing and ending with **Media Access Control (MAC)** and port addressing.

We will cover the following topics in this chapter:

- IP addressing
- IP version 4
- IP version 6
- MAC addressing
- Address resolution
- Ports and protocols

IP addressing

For the purposes of the Server+ exam, we really don't need to cover the entire history of network addressing, from the development of packet switching 50 or so years ago to IP version 6 addressing today.

When TCP replaced the earlier **Network Control Protocol (NCP)** in 1974, it performed both data transmission and message routing. These two functions were later split into what we know today as TCP and IP, as in the TCP/IP protocol suite.

IP version 4

The primary purpose of an IP is to provide a path through a network of computing and routing devices interconnected by communication links in an unstructured manner.

Early on, it became clear that one of the underlying design features of the internet was that it provided no single point of failure. This led to its purposefully haphazard construction, and the need for a flexible addressing scheme.

The TCP/IP protocols developed over the years into the standard that is still largely in use today. Given that there was pressure on the developers from large communication corporations, government agencies, and small and medium-sized businesses, an address class system divided up the available address range to provide ample network addresses to all parties—at least at first.

The IPv4 address structure

What we have come to know as an IP address in general, the IPv4 address, consists of 32 bits divided into four eight-bit octets (groups of eight). For the sake of humans, a format called **dot-decimal** formats the binary address into four decimal numbers separated by dots, or periods.

For example, the following shows the binary number version of an IP address and its decimal number equivalent:

```
Binary value: 10101001.1110011.00011001.11011010  
Decimal value: 169. 115.25.218
```

Each set of four octets represents a single addressable network location.

An octet may hold binary values in the range from 00000000 to 11111111, or 0 to 255 in decimal, respectively. Therefore, the total range of IPv4 addresses starts at the following:

00000000.00000000.00000000.00000000 (decimal 0.0.0.0)

The range goes up to the following:

11111111.11111111.11111111.11111111 (decimal 255.255.255.255)

A significant number of the addresses in this range are set aside for special purposes. This means that not all of the addresses in the range are available for assignment to network nodes. We will talk more about this later in this section.

Classful IP addressing

As shown in the following table, the IPv4 addressing scheme includes five address classes, although only three have been in use extensively. Address classes **D** and **E** were set aside for multicasting under IPv4 and research and experimentation, respectively:

Address class	From	To	Network addresses	Hosts per network
A	1.0.0.0	126.255.255.255	126	16,777,214
B	128.0.0.0	191.255.255.255	16,382	65,534
C	192.0.0.0	223.255.255.255	2,097,150	254
D	224.0.0.0	239.255.255.255	Reserved for multicasting	
E	240.0.0.0	254.255.255.255	Experimental and research	

IPv4 address classes

The address classes shown in the table simplified the assignment of addresses to the various communication, manufacturing, services, government agencies, and companies wanting to be **internet service providers (ISPs)**. If an organization could demonstrate a certain level of need, based on the size of its network or business model, it received a block of addresses from an appropriate address class.

LAN addressing

However, like all good things, IPv4 addresses (all 4 billion of them) soon began running out. IPv6 was in development, but neither the technology nor the network managers were ready for the upgrade.

So, to avoid a complete exhaustion of IPv4 addresses, a variety of techniques emerged to maximize the usage of an IP address, and to minimize the need for additional addresses. These developments included private IP addresses, subnetting, CIDR, NAT, PAT, and others.

Private IP addresses

Somewhere along the line, it occurred to someone that, of the network-attached devices behind an organization's gateway router (edge router), those on a **local area network (LAN)** really don't need assigned public IP addresses. The majority of LAN traffic is between its nodes, and that occurs on the Data Link Layer (Layer 2) and makes use of the **Media Access Control (MAC)** sublayer address (physical address).

But *what if an organization with several LANs doesn't have a sufficient number of public IP addresses to configure every network node with its own?* Finding enough IPv4 addresses is the least of its problems.

Although it reduced the number of IPv4 addresses in the assignable pool, three blocks of addresses, one from each of the classes A, B, and C, were set aside for use as private addresses. The **Internet Assigned Numbers Authority (IANA)** designated private IP addresses for use on any network behind a gateway router or **network address translation (NAT)** device (more on which later in this section). Private IP addresses are, by definition, non-routable.

As shown in the following table, each of the assignable address classes (**A**, **B**, and **C**) have a range of private IP addresses:

IPv4 address class	Private address range	Private addresses available
A	10.0.0.0-10.255.255.255	16,777,216
B	172.16.0.0-172.31.255.254	1,048,576
C	192.168.0.0-192.168.255.254	65,536

IPv4 private addresses

As you can see, each block of addresses provides an adequate number of assignable private addresses for nearly all LANs.

An organization may also assign the same set of private addresses to two or more of its networks, as long as each network passes through a router or a NAT device to access outside the LAN.

Network and host IDs

IP addresses contain two significant pieces of information—a network identifier, and a host identifier. Each of the IPv4 address classes uses a different number of bits for each of the identifiers.

The following table shows the number of bits used for the **network (n)** and **host (h)** in each class:

Address class	High-order bits in first octet	Address mask pattern	Network ID bits	Host ID bits	Default address mask
A	0	n.h.h.h	7	24	255.0.0.0
B	10	n.n.h.h	14	16	255.255.0.0
C	110	n.n.n.h	21	8	255.255.255.0

Address masks and class IDs extract the network ID from an IPv4 address

The high-order (leftmost) bit or bits in the first (leftmost) octet indicate the address class of an IPv4 address. *Why?* The short answer is that extracting the network ID from the IPv4 address requires the use of the appropriate address mask.

The process of extracting the network ID from an IPv4 address applies a bit-wise AND operation to the address and the mask corresponding to the address class of the address. The following table shows the steps taken:

IPv4 address (decimal)	10.25.115.88
IPv4 address (binary)	00001010.00011001.01110011.01011000
Class A default mask	11111111.00000000.00000000.00000000
Network ID (binary)	00001010.00000000.00000000.00000000
Network ID (decimal)	10.0.0.0

The steps to extract the network ID from a Class A IPv4 address

First, the IPv4 dot.decimal address converts to binary. Then the binary form of the default Class A (0 in the first bit) address mask is ANDed to the binary IPv4 address.

The bit-wise ANDing process compares two bits, and if both have the value of 1, the result is set to a 1. All other combinations of the two bits result in zero. For example, in the preceding table, the IPv4 address has the binary value of 00001010, and the default Class A mask has the binary value of 11111111. The AND operation compares the two values; only when the two bits (upper and lower, address and mask) both have a 1 is the result set to a 1.

The operation looks something like the following:

10 = 0000	1010
255 = 1111	1111
10 = 0000	1010

The bitwise AND operation of an IPv4 address and address mask

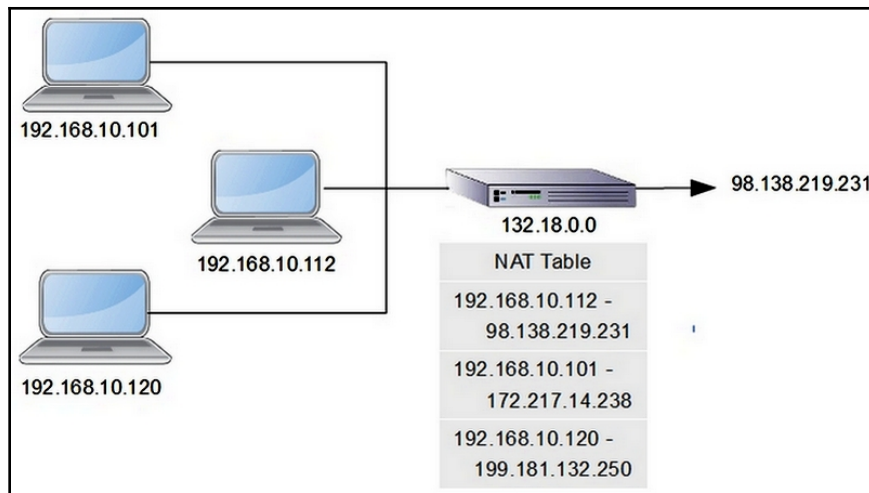
Network Address Translation (NAT)

Because private IPv4 addresses aren't routable, LAN nodes with private addresses can't request content from the internet. Any content request made to a server on the internet must contain the IP address of the requester.

If the source address is a private address, there is no way for the server or any intermediate devices between the server and the requester to know which 192.168.0.0 private network the requesting node is in.

To remedy this problem, the NAT protocol, configured on the gateway router, applies a public IP address as an alias for the requester's private address. NAT records the pairing of the assigned public address and the private address to which it's matched in a table. When the response arrives, NAT uses its source address (the sender) to look up the private address of the requester/destination, and forwards the information.

The following diagram illustrates the elements of this process:



The functional elements of the NAT protocol

Although the diagram shows the NAT operation with multiple alias addresses, many NAT implementations use only a single public address (typically, the address of the router), and tracks requests in the same manner as shown.

The primary mechanism that enables the use of a single public address is **Port Address Translation (PAT)**, otherwise known as NAT overloading. To differentiate between outbound requests, PAT applies a unique port number to the private address of the requesting LAN node.

Collision domains

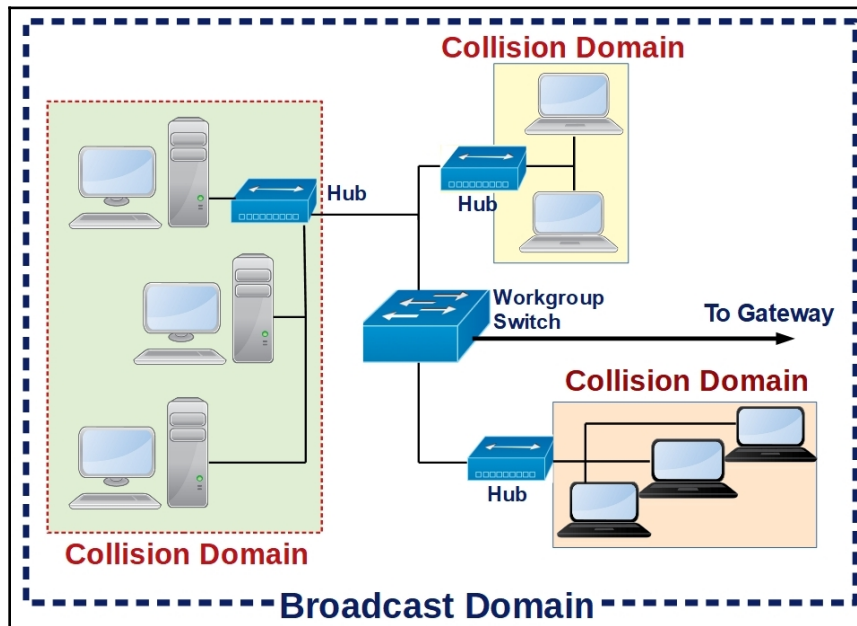
Two common problems inherent to Ethernet LANs are collisions and broadcast messages. Collisions occur when two (or more) nodes attempt to transmit on the network medium simultaneously, effectively destroying both messages.

Broadcast messages are a normal part of an Ethernet network's operations, but reducing the size of the network also reduces the impact these messages have on network performance.

Ethernet networks use a procedure called **Collision Sense Multiple Access/Collision Detection (CSMA/CD)** to detect when a collision has occurred on the network medium. CSMA/CD uses a back-off timer to stagger the retransmissions from the colliding nodes. On a wireless network, the procedure **CSMA/Collision Avoidance (CSMA/CA)** attempts to avoid collisions before they happen.

To regardless of whether the network medium is a cable or radio frequency waves, network nodes must share the common medium, often at the same time.

A grouping of network nodes connected through a hub or repeater is a collision domain. As shown in the following diagram, network segments behind a network hub are collision domains and, in this case, three collision domains exist behind a network switch:



Collision and broadcast domains on a LAN

In fact, all the nodes behind the switch are in one large collision domain. This would be especially true if these workstations were on a wireless network.

Broadcast domains

On the other hand, a broadcast domain includes all the network nodes that are able to communicate with each other. In virtually all networks, a broadcast domain encompasses all of the collision domains, as illustrated in the preceding diagram.

On a computer network, broadcasting means essentially the same as it does in radio or television. A network node transmits a message, typically a request, to the entire network, meaning every active connected device.

In most cases, a broadcast message on a LAN results in each receiving node sending out its own broadcast messages. This is the reason why network administrators try to limit the overall size of broadcast domains.

The two most common uses of broadcast messages are as follows:

- **Startup:** When you boot or restart a computer, its operating system sends a broadcast message on the network requesting its IP address configuration. It sends the broadcast message because, at that point, it doesn't have the address of the **Dynamic Host Configuration Protocol (DHCP)** server. If all is well, the DHCP server responds only to the requesting node with its configuration data. Any computers on a network that have a very specialized function or setup may not or should not use a DHCP configuration. In these cases, the node should be configured with static configuration settings.
- **Address resolution:** The **Address Resolution Protocol (ARP)** and its mirror-image, the **Reverse Address Resolution Protocol (RARP)**, use broadcast messages to ask network nodes which of them has a certain MAC address (more on this later), and to send back its IP address. The request could also be in reverse, with RARP asking for the MAC address that corresponds to a certain IP address.

Broadcast messages carry a unique IP address, one reserved specifically for this purpose: 255.255.255.255. When network devices see a message with this destination address, they know that it's a broadcast message.

Classless Interdomain Routing (CIDR)

Another way to express an IPv4 address that eliminates most of the need for an address mask is the CIDR address notation.

CIDR (pronounced variably as *cedar*, *cider*, or *kidder*, and also known as **supernetting**) denotes the number of bits in an IPv4 address that designate the network ID, without any consideration as to which address class it is.

For example, a standard Class A address has the following pattern:

nnnnnnnn.hhhhhhhh.hhhhhhhh.hhhhhhhh

Here, n denotes the network ID, and h denotes the host ID.

Therefore, in the address 101.15.105.10, the network ID is 101.0.0.0. Its address mask is 255.0.0.0. This same address expressed in CIDR notation is 101.15.105.10/8, which indicates that the leftmost eight bits are the network ID.

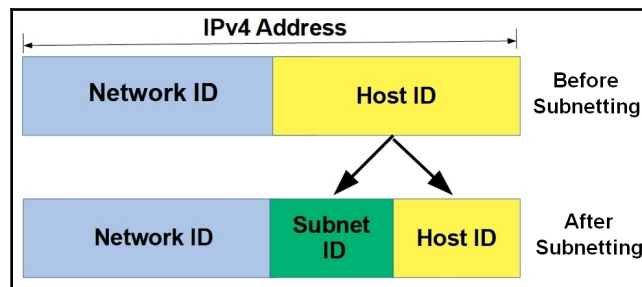
CIDR provides greater flexibility for extending networks without requiring additional IPv4 addresses.

Subnetting

A **subnetwork (subnet)** is a logical segment of a larger network that provides a variety of benefits to the network, including the following:

- **Limiting broadcast messages:** Subnetting a network creates smaller broadcast and collision domains
- **Expanding a network's size:** Subnetting facilitates the expansion of a network without purchasing additional IP addresses
- **Security:** Subnetting allows for the isolation of protected departments or functions, such as accounting or research

Subnetting is a logical addressing technique that creates smaller subnetworks using existing (in place) IPv4 addresses. The basic principle behind subnetting is the reassignment of one or more bits in the address mask, which in this case becomes a **subnet mask**, to provide for additional networks (subnetworks), as shown in the following diagram:



One or more bits borrowed from the host ID create a subnet ID in an IPv4 address

Subnets and hosts

First, we determine the number of subnets and the number of hosts available on each subnet. Typically, this part of the calculation can be a bit of trial and error.

The following table lists the departments of the **ServersRUs** company, and the number of workstations (hosts) required in each department:

Department	Number of hosts
Accounting	7
Customer support	18
IT	8
Management	6
HR	4

ServersRUs' subnet and host requirements

If we wish to set up a subnet for each department, we need to be able to address five subnets. In a general approach to subnets, each will have the same number of hosts, which in this case must be at least 18 hosts.

Remember, when determining how many bits to take from the host ID and apply to the subnet ID, that 2 raised to the power of the number of borrowed bits is the number of subnets that it creates. For example, ServersRUs needs five subnets, and 5 isn't a power of 2. At the low end of the powers of 2, we have the values of 1 (2^0), 2 (2^1), 4 (2^2), and 8 (2^3).

Of these, only 2^3 or 8, provides for the five subnets needed, with the remaining three subnets available if or when needed.

We have determined that three bits are to move from the host ID and to be used for the subnet ID. This leaves five bits for the host ID within each subnet. *Are five bits enough to provide each subnet with the number of hosts required?* The customer support department requires 18 hosts, which must be our target number. So, if we raise 2 to the power of 5 (number of bits), we are able to address 32 hosts on each subnet.

There is a formula in subnetting that works for calculating the number of subnets available and the number of hosts available on a subnet: $2^n - 2$ = addressable subnets or hosts. In this formula, n represents the number of bits borrowed (subnets) or remaining (hosts). We'll discuss the -2 in a minute.

The IPv4 Class C subnet addressing is shown in the following table:

Bits borrowed	Subnets	Hosts/subnets	Subnet mask	CIDR
0	1	254	255.255.255.0	/24
1	2	126	255.255.255.128	/25
2	4	62	255.255.255.192	/26
3	8	30	255.255.255.224	/27
4	16	14	255.255.255.240	/28
5	32	6	255.255.255.248	/29
6	64	2	255.255.255.252	/30

IPv4 Class C subnet addressing

Subnet masks

As shown in the preceding table, borrowing bits changes the values of the address mask/subnet mask. If we take zero bits, we have a standard Class C network with one network and 254 hosts. However, if we reassign four bits, we are able to address 16 subnets with 14 usable hosts per subnet.

Because we added four bits to the 24 bits already in a Class C address mask, we now have 28 (/28) bits in the network ID of our network. The following table shows how this impacts the value of the subnet mask:

Decimal value	255.255.255.0	Class C address mask
Binary value	11111111.11111111.11111111.00000000	Class C address mask
Borrowed bits	.11100000	23 = 8 subnets
New binary value	11111111.11111111.11111111.11100000	
New decimal value	255.255.255.224	128+61+32 (27+26+25)

Constructing a subnet mask

Network and broadcast addresses

On any network, the address of the entire network is always the first address of its assigned address range. For example, the IPv4 address 10.0.14.210 is the Class A address of a network host, and its network address, typically assigned to the gateway router, is 10.0.0.0.

In the ServersRUs example, each of the subnets has an assigned address range. The first address in the range is its network address. The broadcast address of a network or subnet is the last address in its address range.

In the 10.0.14.210 example, the broadcast address is 10.255.255.255, absolutely the last address in the 10.0.0.0 address range of one subnet with 16,777,214 hosts.

The following table shows the full range of addresses in the ServersRUs network:

Subnet network ID	Subnet hosts addresses	Subnet broadcast ID
192.168.32.0	192.168.32.1 to 192.168.32.30	192.168.32.31
192.168.32.32	192.168.32.33 to 192.168.32.62	192.168.32.63
192.168.32.64	192.168.32.65 to 192.168.32.94	192.168.32.95
192.168.32.96	192.168.32.97 to 192.168.32.126	192.168.32.127
192.168.32.128	192.168.32.129 to 192.168.32.158	192.168.32.159
192.168.32.160	192.168.32.161 to 192.168.32.190	192.168.32.191
192.168.32.192	192.168.32.193 to 192.168.32.222	192.168.32.223
192.168.32.224	192.168.32.225 to 192.168.32.254	192.168.32.255

Subnet addressing for the ServerRUs network

Note that the first subnet's network ID is the first address of the whole range (192.168.32.0 – 192.168.32.255). The network ID for the other seven subnets is the first address in its range, and the broadcast address is the last address in its range.

Each of the subnets shown in the preceding table has 30 addressable hosts, a network ID, and a broadcast address. The subnet mask (255.255.255.224) is the same for all subnets.

Internet Protocol version 6 (IPv6)

IPv6, first introduced in 1998, has slowly replaced or supplemented IPv4 addressing, particularly gaining momentum over the past few years. The reason for the development of IPv6 was to forestall the rapid depletion of IPv4 addresses.

The IPv6 structure extended the number of available network and host addresses to what should be a sufficient number for the foreseeable future. The IPv6 address space provides for 2,128 addresses, which is around 340,282,366,920,938,463,463,374,607,431,768,211,456 (340+ undecillion) individual addresses.

At the present time, only about 20 percent of these addresses are available. Don't worry, though, we won't be running out any time soon, because the current pool has enough addresses for everyone on the planet to be issued more than 3,000 addresses each.

The IPv6 address structure

IPv6 increases the IP address length to 128 bits, divided into eight 16-bit blocks. Each 16-bit block represents four hexadecimal values, called **hextets**. The following table shows the hexadecimal values represented in four bits:

Hexadecimal value	Binary value
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
A	1010
B	1011
C	1100
D	1101
E	1110
F	1111

Hexadecimal and binary value equivalents

The address `FE80:0000:0000:0000:0202:FFFA:C4DD:7435` is an example of the IPv6 format.



Note that, whereas IPv4 uses dots/periods to separate its sections, IPv6 uses colons (:).

The first three bits of an IPv6 address are set to 0012, which results in all public IPv6 addresses having just four hexadecimal values in the first hextet and beginning with either a 2 or a 3. For example, `1234:5678:98A4::ABCD` is not a valid public IPv6 address, because it begins with a 1, and `234:5678:98A4::ABCD` is not a valid address, because its first hextet has only three digits.

IPv6 addresses range from `2000::` to `3FFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF`, reducing the size of the pool to 2125.

Reserved prefixes

The first hextet of an IPv6 address may contain certain values to indicate a designated type of transaction or message. The reserved prefixes commonly used in IPv6 are as follows:

- `2002::/16`—6to4 routing. Provides for routing an IPv6 message over IPv4 tunneling.
- `fe80::/10`—link-local address. Primarily for configuration and discovery within a local network; routers do not forward.
- `ff00::/8`—multicast address. See the following table for some of the reserved multicast addresses:

Address	Description
<code>ff02::1</code>	All nodes on the local network segment
<code>ff02::2</code>	All routers on the local network segment
<code>ff02::5</code>	OSPFv3 All SPF routers
<code>ff02::9</code>	RIP routers
<code>ff05::1:3</code>	All DHCP servers on the local network site
<code>ff0x::fb</code>	Multicast DNS
<code>ff0x::114</code>	Used for experiments

A sample of the reserved prefixes for IPv6 multicast messages

IPv6 address compression

IPv6 allows for the compression of sections with only zeros, replacing two or more contiguous hexets with a double colon (: :). For example, the preceding address becomes `FE80::0202:FFFA:C4DD:7435` in compressed form.

However, an IPv6 address may contain only one double colon. This means that, if an address has two or more hexet groups of all zeros, compression can only apply to one of them. For example, the address `FE80:0000:0000:0000:0202:0000:0000:7435` compresses to `FE80::0202:0000:0000:7435`.

IPv6 leading zero compression

A hexet with one or more zeros in its most significant positions (leftmost) is compressible, removing the zeros. For example, a hexet of `00A7` becomes just `A7`, or a hexet of `0000` compresses to `0`. The address `FE80::0202:0000:0000:7435` further compresses to `FE80::202:0:0:7435`, which shows the application of both compression methods.

IPv6 network ID

Because IPv6 doesn't define address classes, a fixed segment of the address represents the network identifier. By standard, the first 64 bits of an IPv6 address identify the network address of the source or destination of a message, but this can vary.

A network of a single host has a network ID of 128 bits. IPv6 uses the CIDR notation to indicate the network ID in an address. An address with a /64 has the first 64 bits as its network address.

IPv6 also reserves certain bit-lengths for special purposes. The following table lists just a few of these:

Prefix	CIDR notation	Purpose
::	/96	An address compatible with IPv4
::	/128	An unspecified address used for internal addressing by software
:::1	/128	Loopback address, referring to the localhost equivalent to the IPv4 localhost 127.0.0.1
fc00::	/7	A Unique Local Address (ULA) , routed only within coordinated sites
fe80::	/10	A link-local address, valid on a local physical link only

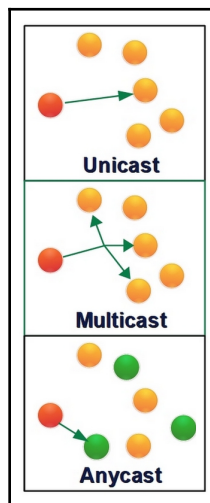
Reserved IPv6 addresses and CIDR designations

Address categories

IPv6 addresses generally fall into one of three address categories:

- **Unicast:** An IPv6 unicast address identifies a single destination, and packets with that unicast address go to that address.
- **Multicast:** An IPv6 multicast address identifies a group of nodes perhaps scattered across several networks. Each of the nodes included in the multicast address receives the transmission.
- **Anycast:** An IPv6 packet with an anycast address goes to only one of an identified set of nodes. The receiving node is typically the closest, in terms of distance and availability.

The following diagram illustrates the three categories:



IPv6 address categories

MAC addressing

IP addresses will work for the addressing within a LAN, but they are really not necessary for internal communications. Ethernet networks, for example, use an addressing scheme that operates on Layer 2 of the OSI model—the MAC address, otherwise known as the **Data Link Control (DLC)** address or physical address. While this number is more of an identification number, its universal uniqueness makes it ideal for addressing across a local network.

Every manufacturer of network or communication devices, such as NICs, switches, modems, routers, and so on, permanently embeds a unique identifying number into each device they produce. The manufacturer embeds this number into a ROM or firmware, so that it's permanently a part of the device.

A MAC/DLC address consists of 48 bits whose value in hexadecimal identifies both the manufacturer and the device. The first 24 bits of the MAC/DLC number identify the manufacturer with an **Organizationally Unique Identifier (OUI)** code, assigned to the producer by the **Institute for Electrical and Electronics Engineers (IEEE)**.

The remaining 24 bits contain what is in effect a serial number unique to the OUI. The combination of the OUI and the serial number creates an identifying number that is unique.

Here's an example—a network adapter in a notebook computer has the MAC/DLC address 5C-E0-C5-B6-B3-9A. The first three sections of this number (5C-E0-C5) indicate that its manufacturer is the Intel Corporation (Malaysia). The conversion of the serial number from hexadecimal to decimal gives the value 11973530, whatever that may mean. Regardless, this unique combination provides an unduplicated value that Layer 2 technologies use as an address.

Address resolution

When you have one type of address, but need another type, you use an address resolution protocol or service. An example might be when you have a person's name and address, but you need his or her telephone number. In this case, the address resolution method is to look up the name and address in a telephone directory to find the number you need.

On a computer network, there are device or configuration addresses that are appropriate for some purposes, but not for others. On an Ethernet network, the MAC/DLC address of a network adapter is sufficient for communications within the network. However, a message intended for a destination beyond the local network requires another level of address.

Using an address from one OSI layer to learn the corresponding address on another OSI layer is called address resolution. Commonly, this means resolving a MAC address to an IP address (or vice versa), or resolving a web domain name to an IP address.

The primary protocol and network service performing address resolution are, on local networks, the **Address Resolution Protocol (ARP)**, and, on public networks, the **Domain Name System (DNS)**.

ARP

ARP converts an IPv4 address (logical address) to its corresponding MAC/DLC address (physical address). This enables the forwarding of messages arriving from outside a local network to their destination. ARP is an OSI Layer 2 protocol which is typically part of the device driver for a network adapter.

A message packet that requires routing cannot use a MAC address for its source address, and must use an IP address. ARP performs this lookup, and provides the corresponding IP address.

At one time, a reverse action, **Reverse ARP (RARP)**, resolved IP to MAC addresses. However, this service is now part of **Dynamic Host Configuration Protocol (DHCP)**.

DNS

The internet would be a lot more complicated to use if all sites had only IP addresses. It would be difficult to remember all of the numerical IP addresses, especially IPv6 addresses. To resolve this potential problem, we use domain names and top-level domain designators in combination, such as `packt.com`, where `packt` is the domain name and `.com` is the **top-level domain (TLD)** designation.

DNS search

A domain name identifies a specific authority or autonomous realm on the web. Domain names are unique and issued to only a single entity, although some entities try to get close in spelling or sound to the domain names of others. Examples of familiar domain names are `google`, `amazon`, `euronews`, and `baidu`. The most popular TLDs include `.com`, `.net`, `.info`, `.gov`, and `.edu`.

Outside the United States, many domain names also include a country code as the TLD, such as `.uk`, `.ca`, `.cn`, or `.de`.

DNS search generally works as follows:

1. A user enters the **fully-qualified domain name (FQDN)** of a website into the location bar of a browser, for example, `www.packt.com`, in which `www` is the hostname assigned to the web server for `packt.com`.
2. The browser detects that it must resolve the FQDN to an IP address, and sends a query to its designated DNS server.

3. The DNS server searches in its `.com` entries for `packt`, and returns the corresponding IP address (`52.216.233.42`) to the browser.
4. The browser then issues an HTTP request to IP address `52.216.233.42` for the page requested.

Domain suffix

A default domain suffix search list allows for a one or more unqualified single name or identifying words to initiate a DNS search. This list contains the domains that should be a part of a DNS search for an IP address.

For example, if `generic.mysite.info` is the FQDN for resolution, and the `mysite.info` search domain includes other hostnames such as `normal`, `usual`, or `occasional`, a domain search using a search domain suffix is easier for the user. By including `mysite.info` in a domain search list, the user only needs to enter the hostname. Likewise, just entering `generic` in the address bar causes the browser to append the search domain/domain suffix to the hostname before sending a request to a DNS server.

The domain suffix search list may contain several search domains, each of which generates a separate request to a DNS server.

The Windows Internet Name Service (WINS)

On Windows systems, both DNS and WINS can resolve device and network names. However, these two services are quite different. WINS is a Microsoft utility that runs only on Windows systems, whereas DNS is system- and platform-independent.

In addition, there are several other differences, as listed in the following table:

Feature	DNS	WINS
IP addressing	Static IP addresses	Dynamic IP addresses
Name resolution	Host names to IP addresses	NetBIOS names to IP addresses
Database modifications	Copies entire database	Incremental modification
TCP/IP application services	Support all services	No support for TCP/IP services

DNS versus WINS basic features

In later releases of the Windows operating systems, improvements to WINS has led to WINS over TCP/IP, which makes it more compatible with DNS (and DNS with it). Legacy WINS is still in use, but is beginning to disappear.

Ports and protocols

Another type of network addressing uses transport layer ports to designate the endpoint to which a particular network packet is to be sent. The term *port* can refer to either of the following:

- A hardware connection point, such as an RJ-45 connection jack on a switch or router
- A software-defined construct that, along with an IP address, establishes a *socket*, or the total address of the software that is to process the incoming packet

It is the second definition that we are using here.

Data is transmitted on a network in **protocol data units (PDUs)**. On different levels of the OSI model, PDUs are given different names—for example, a PDU on the Network Layer is known as a **packet**, and on the Data Link Layer it's called a **frame**.

On the Transport Layer, a PDU is known by two names—segments (TCP protocol) and datagrams (UDP protocol). The point of this is that ports are associated with TCP and UDP, and the segments and datagrams of each.

Well-known ports

The port numbers available for assignment by the IANA range from 0 to 65535. Of these, ports 0 to 1023 are well-known ports. A **well-known port** is a service or protocol that resides on a private network or the public internet, and is a common server-based application.

As shown in the following table, each protocol/port combination has a numeric port associated with it:

Protocol	TCP port number	UDP port number
Domain Name System (DNS)	53	53
Dynamic Host Configuration Protocol (DHCP)		67 and 68
File Transfer Protocol (FTP)		20 and 21
FTP over TLS/SSL (FTPS)	989/990	989/990
HTTP over TLS/SSL (HTTPS)	443	
Hypertext Transfer Protocol (HTTP)	80	
Internet Message Access Protocol v4 (IMAP4)	143	143

Lightweight Data Access Protocol (LDAP)	389	389
Network Time Protocol (NTP)		123
Post Office Protocol 3 (POP3)	110	
Secure FTP (SFTP)		
Secure Shell (SSH) / Secure Copy (SCP) / Secure FTP (SFTP)	22	22
Simple Mail Transport Protocol (SMTP)	25	25
Simple Network Management Protocol (SNMP)	161	161
Telnet	23	23

Well-known TCP/UDP ports

The port number added to the end of an address designates the processing software, protocol, or service. For example, addressing a segment to 10.0.0.20:80 indicates that it's an HTTP message. The combination of the IP address and the port number creates a socket.

Registered ports

The next set of port numbers, ranging from 1024 to 49151, are registered ports, which are port numbers that individuals and companies can register for association with a particular software package or application.

You may encounter two registered ports on the Server+ exam, as follows:

Protocol	TCP port number	UDP port number
LDAP—AD	3268	3268
Remote Desktop Protocol (RDP)	3389	3389

Registered TCP/UDP for Server+ Exam

The last group of port numbers, those from 49152 to 65535, are dynamic (or private) ports. The IANA doesn't register these port numbers; they are available for general use by the public, primarily for internal purposes.

Summary

In this chapter, we discussed the fact that an IPv4 address is 32 bits divided into four octets in a dot-decimal format. An octet may hold values from 0 to 255.

We discussed how IPv4 addressing has address classes A, B, C, D, and E. Private addresses are for LANs. Each IPv4 address class has a standard subnet mask to identify its network ID: class A—eight bits; class B—16 bits, and class C—24 bits. NAT devices mask private addresses with public addresses, whereas PAT applies a port number to a private address.

Next, we covered how collisions occur when nodes attempt to transmit simultaneously. The impact of broadcast messages reduces on smaller networks. CSMA/CD detects collisions, and staggers the retransmissions of colliding nodes. Wireless networks use CSMA/CA to avoid collisions. Common broadcasts are startup and address resolution. CIDR denotes the number of bits in the network ID of an IPv4 address.

We then looked at subnets, which are logical segments of a larger network. Subnetting alters the subnet mask to provide additional subnet IDs. The first address in the range of addresses of a subnet is its network address. The broadcast address of a subnet is the last address in its address range.

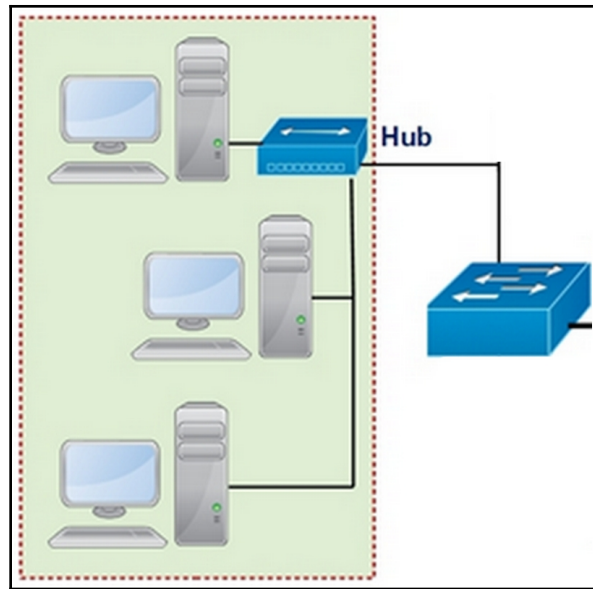
Next, we discussed IPv6. This increases the address length to 128 bits, divided into eight 16-bit blocks called hextets. The first 64 bits identify the network ID. IPv6 addresses can be unicast, multicast, or anycast. Ethernet networks use MAC/DLC addressing to identify the manufacturer and device. The first 24 bits is the OUI; the remaining 24 bits is a serial number.

We went on to look at the protocol and network services that perform address resolution, namely ARP, WINS, and DNS. ARP converts IP addresses into a corresponding MAC address, WINS converts NetBIOS names to MAC addresses, and DNS converts FQDNs to IP addresses.

Finally, we covered ports—software-defined identifiers that combine with an IP address to identify a socket, which designates the software to process the incoming packet. Port numbers range from 0 to 65535. Ports 0 to 1023 are well-known ports.

Questions

1. Which of the following best describes an IPv4 address?
 1. Classless
 2. Five classes of 32-bits in four octets
 3. 128-bits in eight sections of 16 bits
 4. The network identifier has a fixed size
2. Which IPv4 address class uses 16 bits to identify the network portion of an address?
 1. Class A
 2. Class B
 3. Class C
 4. Class D
 5. IPv4 is classless
3. Which network service provides an alias public address to internal network nodes with private LAN addresses?
 1. DHCP
 2. DNS
 3. WINS
 4. NAT
4. In the subnet shown in the following diagram, the red dotted line encompasses which of the following?
 1. Broadcast domain
 2. IPv4
 3. Collision domain
 4. IPv6



5. Wire-based Ethernet networks use which technology to manage message collisions?
 1. CSMA/CA
 2. TCP/IP
 3. CSMA/CD
 4. NAT
6. Which of the following is likely to be the broadcast address for a subnet?
 1. 10.0.0.0
 2. 192.168.32.10
 3. 201.255.255.255
 4. 168.92.15.0
7. Which of the following is an example of CIDR notation?
 1. 2020::15AD:0:25FF
 2. 201.110.25.16/24
 3. 10.0.0.0
 4. 198.168.32.10:80

8. What is the significance of a double colon inserted in an IPv6 address?
 1. It masks one or more sections not used for routing
 2. It masks one or more sections containing all zeros
 3. It masks one or more sections containing FFFF16
 4. It indicates a TCP/UDP port number
9. Which of the following is not an IPv6 address category?
 1. Multicast
 2. Unicast
 3. Anycast
 4. Broadcast
10. Which of the following is the port number range for well-known TCP/UDP ports?
 1. 0 to 65535
 2. 1024 to 49151
 3. 0 to 1023
 4. 49152 to 65535

6 Cabling

In order for networked devices to communicate with each other, some form of communication medium must be in place. This medium, regardless of type or technology, provides the conduit through which the transmitted signal travels from one node to another.

The technology used for wired and wireless network communication, beyond the physical medium involved, is not all that different.

The Server+ certification exam includes questions on the different connection media, their connectors, and proper installation procedures.

In this chapter, we'll look at the physical media that connects wired networks, and the **radio frequency (RF)** signals that connect wireless networks, local or wide in scope. We'll review the materials, construction, connections, installation, and technology that provide the communication connections linking a network to a network server.

We will cover the following topics in this chapter:

- The construction, connectors, and use of copper twisted-pair and coaxial cabling
- The construction, connectors, and use of fiber-optic cabling
- Cable installation methods and management devices
- Network cabling labeling

Copper cabling

Copper cables, in their various forms, have been the de facto standard for interconnecting network devices since the beginning of business-related computing (perhaps even longer, if we count the cabling systems of telephone service providers).

Copper cabling was an almost intuitive choice in early networks, which were mostly peer-to-peer topologies. It's relatively inexpensive, readily available, easily installed, and has low resistance to transmitted electrical signals. On the other hand, it has a limited transmission range and is susceptible to external electrical interference or noise.

When properly installed, copper cabling is both reliable and easy to maintain.

Twisted-pair cabling

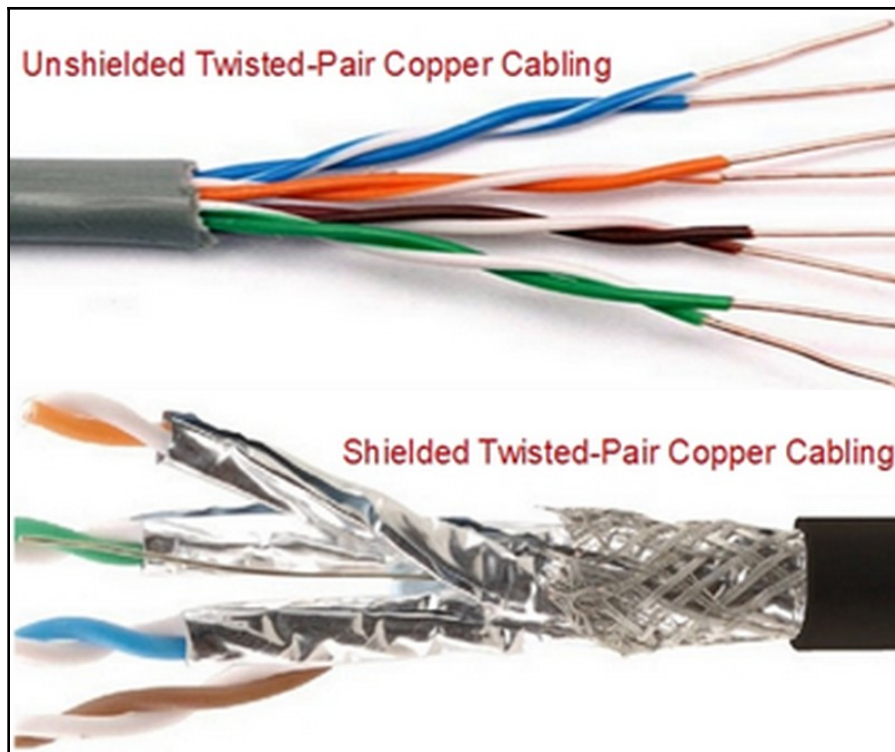
The copper cabling used in **local area networks (LANs)** is predominantly twisted-pair wire bundles.

One important point to know is that, even with insulation coating it, **twisted-pair (TP)** wiring is an **electro-magnetic interference (EMI)** sponge. Back in the late 1880s, the telephone people discovered that wrapping or twisting two wire strands together helped to reduce the amount of noise the wires absorbed from outside the cable. However, twisting wire pairs only reduces how susceptible TP is to electrical noise.

There are two types (and myriad variations) of twisted-pair copper cabling:

- **Unshielded twisted-pair (UTP):** Unshielded doesn't mean bare wire; what it does mean is that beyond the insulation around each wire strand, there is no other shielding in the cable to block EMI. UTP is inexpensive compared to other options, and in typically safe environments such as residential properties and small offices, its low cost more than compensates for its faults.
- **Shielded twisted-pair (STP):** Shielded means that, in addition to the insulation of each wire, shielding such as metal foil or braided mesh surrounds the internal cable construction to deflect or absorb interference.

The following image shows the physical differences between UTP and STP. Notice the lack of shielding on the UTP, and the plethora of shielding on the STP:



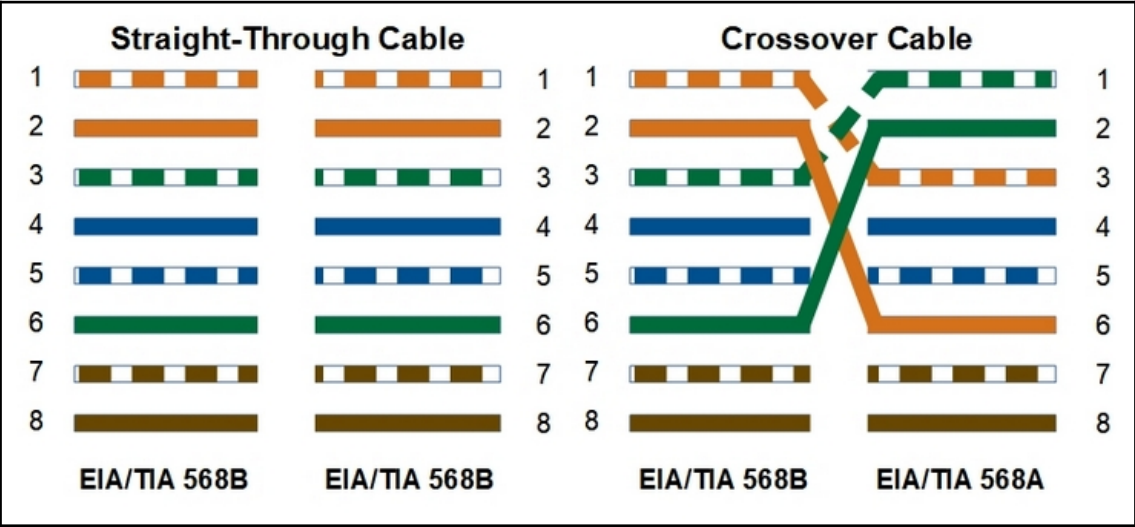
UTP versus STP copper cabling

Each segment of cable installed on a network has one or more specifications, standards, or configurations. In addition, a number of the various cable configurations have a special purpose that not only dictates the pattern and position of the individual wires, but also how they connect to a jack.

The termination and pinouts of twisted-pair cabling are governed by **Electronics Industry Association (EIA)/Telecommunications Industry Association (TIA)** standards 568A and 568B. Often, people refer to the two standards as being commercial and residential, respectively.

In any wired network, several different types of cables, meaning cables with different connector configurations, connect the various elements and devices of the network. The cable types you may encounter in the Server+ exam are as follows:

- **Crossover cables:** These combine the two 568 standards, with a connection of each standard on either end. This cable type typically connects two communication devices of the same type, such as two routers or two switches. A crossover cable has wires 1 and 3 switched with one another, and wires 2 and 6 switched with one another, at one end of the cable. The following diagram illustrates this configuration:



The pinouts of straight-through and crossover cables

- **Patch cables:** This is a generic term that describes any cable type, including crossover, rollover, and straight-through cables, used to establish a connection between two electronic devices.
- **Rollover cables:** These are also known as **Yost cables**, these have reversed or rolled-over pinouts, and aren't used for data transfer but rather for creating an interface between two devices. The pinouts for this cable type are listed in the following table:

Cable end 1	Cable end 2
Pin 1	Pin 8
Pin 2	Pin 7

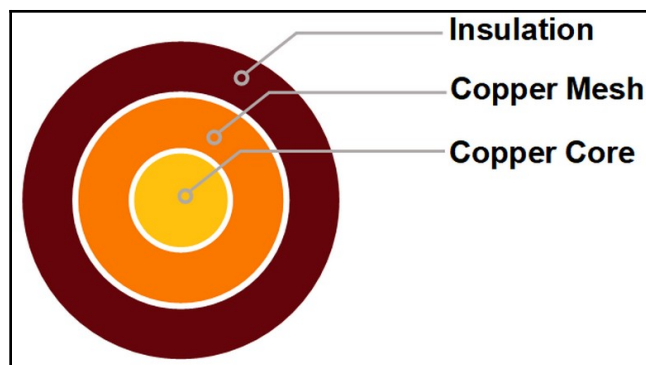
Pin 3	Pin 6
Pin 4	Pin 5
Pin 5	Pin 4
Pin 6	Pin 3
Pin 7	Pin 2
Pin 8	Pin 1

- **Straight-through cables:** These have connections of the same standard, 568a or 568b, on both ends, with the connector pins matched—that is to say, pin 1 to pin 1, pin 2 to pin 2, and so on. This cable type connects a network adapter to an internetworking device such as a router.

Coaxial cabling

Coaxial cables have two channels—the core wire and the metallic, usually copper, mesh layer, which each carry the same signal. The electrically-charged carriers each generate an electromagnetic field that cancels the other channel's emission, creating a barrier that external interference is unable to penetrate to alter the transmitted signal.

The following diagram shows the primary layers of a coaxial cable:



A cross-section view of a coaxial cable

Network connectors

Along with wire patterns and pinouts, the connectors that terminate network cables are standards-defined. For use with twisted-pair cabling, the EIA/TIA 568 standards govern the **registered jack-45 (RJ-45)** connector, which is an **eight position/eight contact (8P8C)** connector, as shown in the following image:



An 8P8C connector (RJ-45)

This connector is the standard for network cable connections to wall jacks, patch panels, and other networking devices. A smaller version of the RJ-45 is the RJ-11, which is a four-wire connector primarily used for telephone connections.

Network coaxial cables use one of two different styles of connectors. The most common coaxial connector is the F-type. This is familiar to most people, as it is the one that connects their television service to a cable provider. It also connects coaxial network cabling to a network adapter.

Another commonly used coaxial connector on data networks is the **Bayonet Neill-Concelman (BNC)** connector. Which of the two is a better connector for computer networks is a matter of choice—both have advantages and disadvantages.

Both connector types are shown in the following image:



Coaxial cable connectors—BNC and F-type

EIA/TIA 568 facility standards

The EIA/TIA 568 cabling standards also define standards for placement, distance, and use, related to where the cables are installed. These standards include the following:

- **Backbone cable:** This is the primary communication cabling that interconnects the primary, main, and intermediate distribution facilities (telecommunication and equipment rooms) in a facility or campus. Twisted-pair cables installed between network connections can be no more than 90 meters in length (about 300 feet).
- **Entrance facility:** Otherwise known as the demarcation point, or *demark*. This is the location where a service provider's network and the subscriber's backbone interconnect. The demark point should be at least 12 inches from where the service line enters the facility.
- **Cross-connect:** The 568 standard requires an equipment room that provides for the termination of the network backbone, as well as a cross-connection point for a facility's entire network.

- **Horizontal cable:** This connects the networking devices on a single level of a facility, such as a floor or story, to the facility's backbone. This cable terminates at an **intermediate distribution facility (IDF)** or telecommunication closet a level. The standard sets a maximum distance of 90 meters between the IDF and a connection point. However, a patch cable of no more than six meters (approximately 20 feet) can connect a network device to the connection point at each end of the horizontal cabling. The interconnecting patch cables cannot exceed a total of 10 meters together.
- **Telecommunication rooms/IDF:** Telecommunication rooms, or *closets*, provide an unlimited number of interconnections between horizontal cabling segments on one level of a facility.
- **Work areas:** The 568 standard requires that each workstation or network node location has access to two network connections, one for voice and one for data—each directly connected by the horizontal wiring to an IDF or telecommunication closet.

Category cabling

Cable standards also classify network twisted-pair cables into **categories (CATs)**. This mostly relates to Ethernet cabling, but has expanded to other types of network applications. Each CAT defines an evolutionary step in network cable capabilities, based on its frequency range, bandwidth, and **data transfer rates (DTRs)**.

Currently, the standards define seven categories, plus a few sub-categories. However, CATs 1 and 2 aren't network cable categories, CAT 4 is specific to ring topologies, and CAT 7 is still in development. Therefore, the only Ethernet cable categories you need to know about are as follows:

- **Category 3 (CAT 3) cable:** An eight-wire (four pairs) UTP cable, capable of 10 Mbps DTR and 16 MHz bandwidth. It is otherwise known as station wire. CAT 3 is still usable for older Ethernet networks, but is now obsolete, replaced by CAT 5 and later categories.
- **Category 5 (CAT 5) cable:** An eight-wire (four pairs) UTP cable that supports both 10 Mbps and 100 Mbps data speeds, and bandwidth of 100 MHz on an Ethernet network. It also supports voice and video transmissions. CAT 5 is still in use, but the CAT 5e standard superseded it.
- **Category 5 enhanced (CAT 5e) cable:** This upgrade to the CAT 5 standard reduced channel crosstalk, and extended its data speed to 1 Gbps with 100 MHz bandwidth. CATs 5 and 5e both use 24–26 AWG copper wire. CAT 5e cabling is backward-compatible with CAT 5.

- **Category 6 (CAT 6) cable:** This CAT raised bandwidth to 250 MHz with GB Ethernet speeds. It also provides for a better outer jacket and insulation, has thinner core copper wires (22–24 AWG), improved **signal-to-noise ratio (SNR)**, and is less susceptible to EMI. CAT 6 specifies both UTP and STP, and is backward-compatible with CATs 5 and 5e.
- **Category 6 augmented (CAT 6a) cable:** This upgrade to the CAT 6 standard raises the DTR to 10 Gbps and the bandwidth to 500 MHz. CAT 6a cables are STP, and require special connectors that provide grounding to the cable.

The following table provides a comparison of the Ethernet category cable standards:

Category	Cable	Maximum DTR	Maximum bandwidth
CAT 3	UTP	10 Mbps	16 MHz
CAT 5	UTP	10/100 Mbps	100 MHz
CAT 5e	UTP	1000 Mbps	100 MHz
CAT 6	UTP/STP	1000 Mbps	250 MHz
CAT 6a	STP	10,000 Mbps	500 MHz

Ethernet cable standards

As well as the category cable standards discussed in the previous section, IEEE 802.3 specifies additional cable standards. These overlap the category cable standards, but also provide a self-describing acronym to each cable type that makes it easier to know its capabilities.

The naming convention for cabling in the IEEE Ethernet standard consists of three primary parts: **DTR**, **signaling mode**, and **cable identification**. For example, 10BaseT represents a 10 Mbps, **baseband (Base)**, twisted-pair (**T**) cable.

The IEEE Ethernet cable standards specify the following cable types:

- **10Base2:** A legacy standard that is still in use. It represents a 10 Mbps, baseband coaxial cable with an attenuation distance of 185 meters (about 607 feet), or a close approximation of its effective range as a multiple of 100 meters. 10Base2 is also known as **Thinnet**.
- **10Base5:** Another coaxial cable with a speed of 10 Mbps on baseband signaling over a distance of 500 meters. Like 10Base2, 5 represents the range as a multiple of 100 meters. 10Base5 is also known as **Thicknet**.

- **10BaseT**: A 10 Mbps baseband UTP or STP cable. CATs 3 and higher support 10BaseT. All twisted-pair cabling has a standard range or attenuation distance of 90 meters.
- **10Base-FL**: A 10 Mbps baseband standard that runs on a **fiber-optic link (FL)**. Because of the fiber-optic medium, its range is 2 km.
- **100BaseTX**: A fast Ethernet specification with 100 Mbps baseband signaling on UTP cable, typically CAT 5 or above.
- **100BaseT4**: A specification that is essentially the same as 100BaseTX with STP cabling added.
- **100BaseFX**: A fiber-optic cable specification implemented on either **single-mode** or **multi-mode** cable. Its range varies with the specific medium. Single-mode fiber-optic cable has a range of 10,000 meters (a little over six miles), and multi-mode cable has a range of 412 meters (a bit less than a quarter of a mile).
- **1000BaseT**: A GB Ethernet standard, along with 1000BaseTX. Both run on CAT 5 or above UTP cable. However, with the increased speed, its range reduces to 75 meters (about 246 feet).
- **1000BaseCX**: Another GB Ethernet standard that runs on STP cabling with a range of only 25 meters (82 feet).
- **1000BaseSX**: A fiber-optic standard that transmits a short-wave laser stream on either 50 or 62.5-micron cable. On 50-micron filaments, its range is up to 550 meters (1850 feet); on 62.5-micron filaments, its range reduces to half that of the 50-micron cable. Both of these distances are full-duplex.
- **1000BaseLX**: Similar to the 1000BaseSX standard, but carrying long-wave laser streams to a distance of 5,000 meters (3.1 miles).
- **10GbE**: A standard that transmits at a speed of 10 Gbps or the equivalent of 10,000 Mbps. It uses the same medium specification as 1000BaseLX, but with technology that increases its range up to 40 km (a bit less than 25 miles).

Fiber-optic cabling

Networks that require longer distances and higher bandwidth than copper cabling can provide are candidates for fiber-optic cabling. Fiber-optic cables have two advantages over copper twisted-pair cables: range and bandwidth.

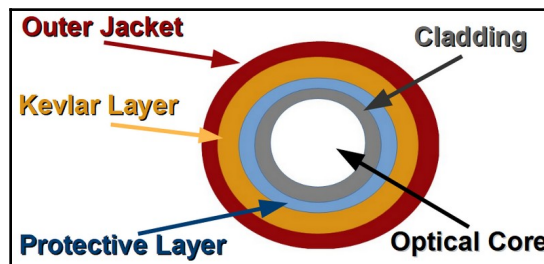
The maximum range on fiber-optic lines is around 20 km, whereas the maximum range for copper twisted-pair cables is, by standard, 100 meters or less. Bandwidth is another area in which fiber-optic cable surpasses copper. There are fiber-optic cables with bandwidth ratings of up to 60 Tbps, but fiber-optic network cabling is generally in the 10 Gbps to 100 Gbps range.

The following table summarizes the capability differences between fiber-optic and copper twisted-pair cabling:

Characteristic	Fiber-optic cable	Twisted-pair cable
Maximum bandwidth	60 Tbps	10 Gbps
Maximum range	20 Km	90 m
Interference	N/A	EMI/RFI, crosstalk, voltage fluctuations
Service life	30-50 years	Five years

Comparison of fiber-optic cables and copper twisted-pair cables

A fiber-optic cable has five layers, as illustrated in the following diagram:



The layers of a fiber-optic cable filament

Starting from the center and moving outward, the layers are as follows:

- **Optical Core:** A transparent glass or plastic filament through which light streams are transmitted.
- **Cladding:** This is typically a highly reflective material that coats the core filament. The effect of the cladding is very much like that of the backing material on a mirror, which reflects the light back toward its source. The purpose of the cladding is to keep the light streams within the core.

- **Protective Layer (buffer):** Many fiber-optic strands and single filament lines have a protective coating around the cladding and core to provide protection against bending, which may break the core.
- **Kevlar Layer:** A strengthening Kevlar fabric layer encases the inner layer (core, cladding, and possibly buffer) to provide additional stiffening and strength to the cable, as a precaution against installation mishaps.
- **Outer Jacket:** An outer jacket or sheath protects the fiber-optic cable assembly from any nicks, cuts, or abrasions it may encounter during installation.

Fiber-optic cable modes

Fiber-optic cabling is available in two modes: **single mode (SM)** and **multi-mode (MM)**.

SM fiber-optic cable

An SM cable has a single fiber filament at its center, which carries a single transmission mode or light stream. Because its core is relatively small in diameter, only one signal mode is able to pass through it. This small core and the single light wave it carries combine to eliminate problems with the transmitted light pulses.

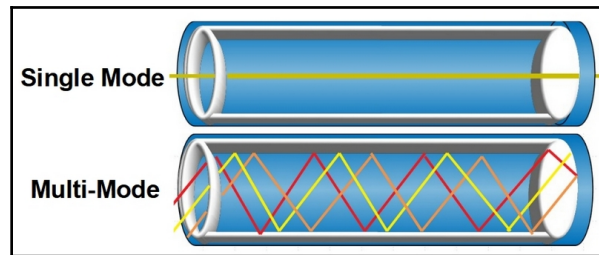
SM has a higher bandwidth than multi-mode cable, and has an attenuation point that is up to 50 times farther. However, you do have to pay more for this speed and distance. SM fiber-optics is also known as SM optical waveguide.

MM fiber-optic cable

As its name suggests, MM fiber is capable of transmitting several light streams at once. MM fiber provides higher bandwidth and higher speeds, but only medium-length distances.

In contrast to single mode, MM disperses light waves into multiple modes, or paths, for transmission. Whereas SM is able to carry a single light wave over longer distances, MM light streams can distort over longer distances, which limits its distance.

The two cable modes are shown in the following diagram:



Single mode transmits a single light stream; multi-mode carries multiple light streams

Which cable is best to use for any particular network depends on, first of all, budget, followed closely by transmission distance, as follows:

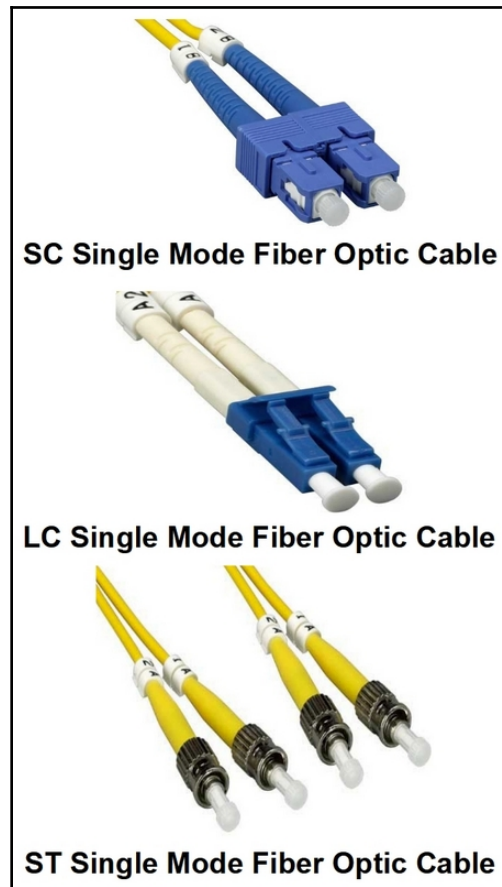
- If the transmission distance is less than 100 meters and is indoors, use copper cable
- For distances longer than 100 meters, but shorter than a couple of miles, MM is the sensible choice if you wish to avoid repeaters and such
- If the distance is two miles or more, SM is best

Fiber-optic cable connectors

Many different types of fiber-optic connector are available, each specialized to a particular use with a certain transmission mode. The connectors used in the deployment of a fiber-optic network will depend on cost, experience, and availability, among other factors.

Each connector has its own advantages and disadvantages. If the connectors used in a network are less than optimal, the performance of the network could degrade over time. The TIA defines fiber-optic cabling specifications in the **Fiber Optic Cable Intermateability Standard (FOCIS)**, which defines both the plugs and sockets of the various connector standards.

Common fiber-optic connectors are shown in the following diagram:



Common fiber-optic connectors
Image courtesy: L-com Global Connectivity

The fiber-optic connectors you may encounter in the Server+ exam are as follows:

- **Lucent Connector (LC):** The LC fiber-optic connector is smaller than most other connectors, but still includes a standard ceramic ferrule connector (the ferrule is the nozzle that protects the fiber in the connector body). FOCIS-10 defines standards for the LC connector.
- **Standard Connector (SC):** This connects with a push-pull locking mechanism that features a spring-loaded ferrule. SC has been a dominant fiber-optic connector type over the past few years, along with the ST connector, because of its high performance. FOCIS-3 defines standards for the SC connector.
- **Straight Tip (ST):** The ST connector is the most popular connector for MM fiber-optic cable. It's common in commercial fiber networks. FOCIS-2 defines standards for the ST connector.
- **Small Form-Factor Pluggable (SFP):** The SFP interface specification defines a transceiver for connecting **Fiber Channel (FC)** and **Gigabit Ethernet (GbE)** cabling to network switches and other internetworking devices using an LC connector. The **Quad-SFP (QSFP)** standard is a popular version of the SFP standard.

Network cable installation

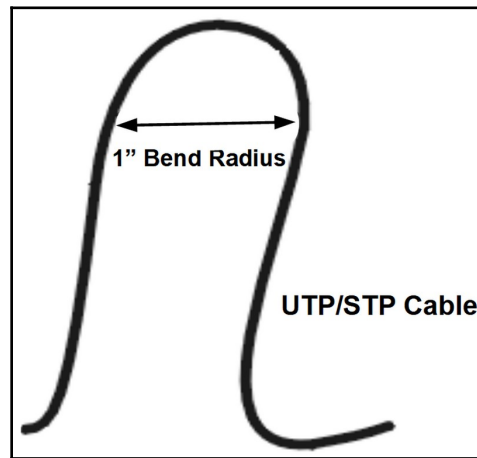
When installing physical network cable, whether twisted-pair, coaxial, or fiber-optic, there are considerations and practices that, if followed, can help to ensure an intact and fully functioning cable plant.

Each type of cable medium has its own special handling requirements and guidelines for proper installation. These address cabling issues that can occur because of improper installation, such as interference, cable damage, and signal loss. The following sections identify the considerations, practices, and guidelines for successfully installing each cable type.

There are certain installation guidelines specific to the copper twisted-pair and coaxial cabling, regarding cable placement, cable bundling, and cable suspension. Many of these same guidelines and recommendations also work for fiber-optic cable.

The following steps, processes, procedures, and practices can help to avoid transmission problems on twisted-pair cabling:

1. Install cable at a 90-degree angle (enter and exit perpendicularly) to any material it passes through, over, or under. If it's necessary to bend the cable, it's important not to exceed the cable's bend radius specification, which is 1 inch according to the EIA/TIA 568b standard (shown in the following diagram). Many certified installers use a 2-inch bend radius as best practice:



The standard bend radius for UTP/STP cable

2. Use predictable cable pathways, such as following hallways, major walls, and other substantial structural components. Avoid crossing over hallways and large office or living spaces, if possible.
3. Install cable in cable supports across open areas, such as basements, attics, and crawlspaces. Place hangers, hooks, and other cable supports every four to five feet across the open space. Tighten the cable run to remove any sag in the cable. A variety of cable supports are available, including the following:
 - **Cable ducts:** Typically a wire raceway or trough that facilitates installation (pulling), organization, and redirection of multiple cables. Most cable duct systems have slots on the side-through which cables are rerouted to drops along the cable run. Caps and solid ducts are available for additional protection.

- **Cable hangers:** These enable the running of cable across an open space, or support cable in spaces where ducts or trays really won't work. Different styles of cable hangers include loops, saddles, bridle rings, and J-Hooks, all of which will support bare cable or conduits. Cable hangers also provide the flexibility to run cable vertically, horizontally, at the top of walls, from ceilings, and along virtually any other surface.
 - **Cable raceways:** Any solid box-like rigid channel that is fully or partially closed and hides, secludes, or protects cabling is essentially a raceway. The different types of raceways include latching, J-Channel, J-Hook, corner duct, and power raceways.
 - **Cable sleeves:** Many refer to cable sleeves and conduits interchangeably, but there are several other types of this support, including braided sleeving, spiral wrap, and wire loom. Wire loom and conduits surround cables and wires to protect against damage and weather. Braided sleeving, which can be metallic or made of carbon fiber, fiber glass, Kevlar, or several other materials, forms itself to its contents to provide protection.
 - **Cable ties:** Otherwise known as **zip ties**, this cable support is actually more like a cable bundler as it wraps a group of cables reasonably tightly to hold the bundle as what amounts to one cable. Cable ties consist of a length of nylon, metal, or other materials, including fabric, and have a locking head that is able to hold in position by locking onto the ridges along the strip. Velcro cable ties are better in situations where they need to be reusable.
 - **Cable trays:** This type of cable support is best for supporting large numbers of cables from the ceiling or below the floor. Their name comes from their open-channel shape, but they are also baskets, cable ladders, and trunking lines.
3. Keep cables relatively slack at each end, and as much as possible throughout the cable run. Avoid or eliminate any taps, splices, or repairs in a cable between the distribution facility and the wall plate to which a station will connect.
 4. Keep cables at the following minimum safe distances:
 - 3 inches from any electrical power line of 2 kVA or less
 - 12 inches from other higher voltage sources, especially fluorescent lighting fixtures
 - 3 feet from electrical lines of 5 kVA or more
 - 3.5 feet away from any transformers or electrical motors, such as soft-drink machines

6. Label all cable runs at least at each end. Preferably, there should also be labeling on each cable run one or two places toward the middle of the cable pull.
7. Any cable installed in a duct or other space that is a part of a building's **heating, ventilation, and air conditioning (HVAC)** system, commonly in the ceiling or walls of a building, must be plenum-rated. Plenum-rated cabling has a fire-retardant coating, typically Teflon, which prevents the cable from giving off toxic gases and smoke as it burns.

Summary

In this chapter, we began by looking at LAN cabling, which is predominantly TP wire. There are two types of TP: UTP and STP. The installation and configuration of network cables are in the EIA/TIA 568A/B standards. Cables installed between network connections can be no more than 90 meters in length (about 300 feet).

We then looked at different types of cables. A crossover cable connects devices of the same type. A patch cable is a generic term or any cable type used to connect two devices. A rollover cable has reversed pinouts and creates an interface between two dissimilar devices. Straight-through cables have the same pinout standard on their ends and connect a network adapter to a device, such as a router.

Twisted-pair cables use RJ-45 connectors, an 8P8C, for cable connections, wall jacks, patch panels, and others.

Coaxial cable has two channels: the core wire carries the transmitted signal and the metallic mesh carries the same signal. Each generates an electromagnetic field that cancels the other's emissions. Coaxial cables use F-type and BNC connectors.

Next, we discussed the EIA/TIA 568 cabling standards. A backbone cable connects distribution facilities. An entrance facility is where a service provider's network and the subscriber's backbone interconnect. A horizontal cable connects network devices on a single level. IDF/telecommunication rooms provide for interconnections between horizontal cabling segments.

We then looked at category cabling, which defines Ethernet twisted-pair cabling. CAT 3 is a four-pair UTP cable capable of 10 Mbps and 16 MHz bandwidth. CAT 5 is a four-pair UTP cable supporting 10 Mbps or 100 Mbps and 100 MHz. CAT 5e reduces crosstalk, and supports 1 Gbps with 100 MHz. CAT 6 supports 250 MHz bandwidth and Gigabit speed. CAT 6a is an STP cable with specialized connectors and supports 10 Gbps and 500 MHz.

We went on to discuss the IEEE 802.3 standard, which specifies a cable identification convention with three primary parts: DTR, signaling mode, and cable identification. 10BaseT represents a 10 Mbps, **baseband (Base)**, twisted-pair (T) cable. Other standards include 10Base2, 10Base5, 10Base-FL, 100BaseTX, 100BaseT4, 100BaseFX, 1000BaseT, 1000BaseCX, and 10GbE.

Next, we turned to fiber-optic cabling. This has two advantages over copper twisted-pair cables: range and bandwidth. The maximum range on fiber-optic lines is around 20 km, and they generally support bandwidth in the 10 Gbps to 100 Gbps range. Fiber-optic cables have five layers: core, cladding, protective layer (buffer), Kevlar layer, and an outer jacket.

We looked at how fiber-optic cabling operates in two modes: SM and MM. SM has one fiber filament and carries a single transmission. SM has higher bandwidth than MM, and an attenuation point up to 50 times farther. MM transmits several light streams at once.

Next, we looked at fiber-optic cable connectors. FOCIS specifies standards for the plugs and sockets of connectors. LC is smaller than other connectors, but includes a standard ceramic ferrule connector. SC and ST have been the dominant fiber-optic connectors because of their high performance. SFP is a transceiver for connecting FC and GbE to switches and other devices through an LC connector.

Finally, we discussed how each physical cable medium has its own handling and installation guidelines that prevent interference, cable damage, and signal loss due to improper installation. Install cable at a 90-degree angle in accordance with the cable's bend radius specification, which is not less than 1 inch. Follow cable pathways and avoid crossing hallways and large open areas.

You should install cable supports in open areas. Common cable supports include cable ducts, hangers, raceways, sleeves, ties, and trays. Cables should be at least 6 inches from low voltage electrical lines and 12 inches from high-voltage lines and devices. Label all cable runs at each end. Cable installed in HVAC ducts must be plenum-rated.

Questions

1. Which two of the following are general types of TP cabling?
 1. Coaxial
 2. Fiber-optic
 3. UTP
 4. STP
 5. FDDI

-
2. What is the EIA/TIA 568 cabling standard for the maximum cable length between network stations?
 1. 500 meters
 2. 185 meters
 3. 100 meters
 4. 90 meters
 5. 10 meters
 3. Which of the following cable configurations commonly connects two dissimilar networking devices?
 1. Straight-through cable
 2. Patch cable
 3. Crossover cable
 4. Rollover cable
 5. Coaxial cable
 4. Which registered jack connector is the standard for Ethernet networks?
 1. RJ-11
 2. RJ-31
 3. RJ-45
 4. RS-232
 5. Which two of the following are connectors commonly used on thinnet coaxial cabling?
 1. RJ-11
 2. F-type
 3. USB
 4. BNC
 5. RJ-45
 6. Horizontal cabling run on the same level commonly terminates in which cabling structure?
 1. Backbone
 2. MDF
 3. IDF
 4. Entrance facility

7. Which two TP cable categories are considered to be the current de facto standards?
1. CAT 3
 2. CAT 5
 3. CAT 5e
 4. CAT 6
 5. CAT 7
8. Which three of the following IEEE 802.3 cable designator components represent Gigabit Ethernet UTP cabling?

Speed	Signaling	Medium
(1) 10	(2) BASE	(3) T
(4) 100	(5) BROAD	(6) F
(7) 1000	(8) NARROW	(9) L

1. (1)(2)(3)
 2. (1)(5)(6)
 3. (4)(2)(9)
 4. (7)(5)(3)
 5. (4)(5)(3)
 6. (7)(2)(3)
9. Which of the following transmission modes of fiber-optic cabling has the longest attenuation rating?
1. SM
 2. MM
 3. SC
 4. SFP
10. When installing network cabling, which restriction is in the EIA/TIA 568 cable standards regarding the need to bend a cable around an object or a corner?
1. A 12-inch bend radius
 2. Splice a cable in the bend
 3. A 1-inch bend radius
 4. No bends in the cable
 5. Install a hub

2

Section 2: Administration

The second part of the book provides an overview of the responsibilities of and the tasks performed by a system administrator to manage and maintain a network server. The chapters in this part of the book cover general administration, maintenance, performance monitoring, fault tolerance and availability, network virtualization, and recovery from a catastrophe.

The following chapters are included in this section:

- Chapter 7, Server Administration
- Chapter 8, Server Maintenance
- Chapter 9, Virtualization
- Chapter 10, Disaster Recovery

7

Server Administration

A large part of a network administrator's job is the administration of the network server or servers. This chapter covers the tools, components, tasks, processes, and management responsibilities that are used or performed to administer and maintain a server.

The duties and responsibilities of a server administrator (that is, the system administrator or network administrator) can vary from one organization to another, but, for the most part, there is a core group of tasks they perform or oversee that are essential to virtually any environment. These tasks and responsibilities generally include monitoring hardware and software performance, user administration, managing backup and recovery procedures, and the application of fixes and patches to name just a few. This part of the Server+ certification exam is process-oriented, meaning that it is less about what you know (although that is still important) and more about what you do. Let's get to it!

In this chapter, we will cover the following topics:

- Network hardware administration
- Server and network asset management
- Server and system documentation

Hardware administration

The responsibilities of a network administrator on a day-to-day basis boil down to monitoring, reviewing, reacting, and installing, when needed. You need to understand that a network administrator and a system administrator aren't exactly the same thing, but the workplace commonly uses these titles interchangeably. A system administrator's responsibilities often include those of the network administrator, but most organizations limit the scope of a network administrator's responsibilities to the configuration, operation, and installation of network hardware. So, since we're talking about hardware administration, we'll use the network administration title.

Network administration

The primary responsibility of a network administrator is to ensure that an organization's computer network conforms to the **Confidentiality, Integrity, and Availability (CIA)** model. To accomplish this, the administrator must keep the components of the network current, functional, and performing as required. Whether the duties of the administrator are a separate job description or embedded in the description of the **system administrator (sysadmin)** job, network hardware administration is an absolutely necessary activity. Regardless of where an organization defines the duties of its administrator, the definition should include these responsibilities:

- Configuring, updating, and maintaining network hardware
- Troubleshooting network issues
- Designing network models
- Implementing and monitoring hardware security policies
- Configuring, managing, and monitoring data storage systems

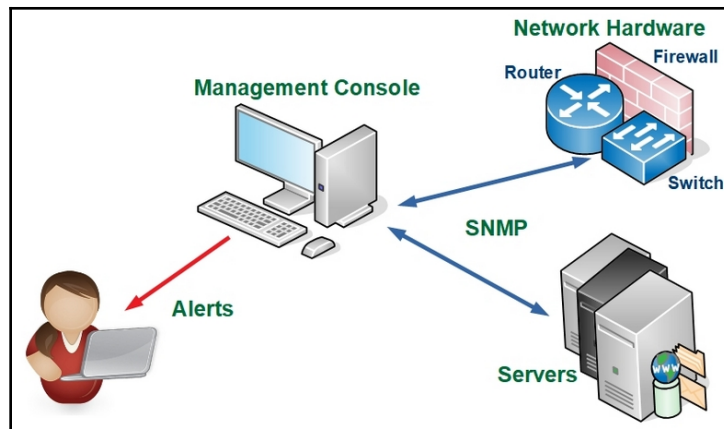
Other system and network areas may also fall under the responsibilities of the administrator or sysadmin. These include the following:

- Testing and installing network software
- Managing and maintaining the **network operating system (NOS)**
- Implementing and monitoring NOS and software security policies
- Administering and coordinating cloud computing services

Let's take a brief look at each of these areas and the responsibilities of the administrator.

Configuring, updating, and maintaining network hardware

An administrator's areas of responsibilities generally surround the network hardware in one form or another. For the most part, network hardware, ranging from the computer running the server software to the gateway router that connects to the outside world, have built-in performance and management utilities that monitor their operations. In some cases, protocols such as the **Simple Network Management Protocol (SNMP)** allow network devices to share information about status and preset conditions, as shown in the following diagram:



SNMP allows network devices to communicate status and issues
Image source: Ron Price

To perform the tasks that are required for network hardware administration, the administrator can interact with monitoring software from virtually any network workstation. However, in larger networks, such as a server farm or a data center, the administrator typically needs access to individual servers through a single console station. This interaction is capable through a few different technologies—**Keyboard, Video, Mouse (KVM)** switches, serial connections, and virtual administration consoles.

KVM interfaces

A KVM switch, such as the one shown in the following image, allows a centrally located administrator to control multiple computers individually through a single keyboard, video display, and mouse. The result is that the administrator's local devices replace those of the remote computer and function as if connected directly to the remote unit:



KVM devices: Front (top) and rear (bottom) views of a 16-port KVM switch
Image source: Raloy, Inc.

A KVM switch provides point-to-point administrative access to individual computers of a connected group through a single set of input/output devices, consisting of a keyboard, video display, and a mouse, from a central administrative site. A KVM, in effect, substitutes its devices for those of a selected computer. For example, the KVM switch shown in the preceding image is able to connect up to 16 computers. The administrator is able to select which of the connected computers is active and interact with it as if they were sitting in front of it.

There are two primary categories of KVM devices—access and control and application and technology. The categories overlap, and KVM switches typically fall into more than one:

- **Access and control:** There are two types of access and control KVMs, based on the number of operators requiring access:
 - A single-user KVM, the most common, connects a single administrator at a central point to multiple remote computers
 - Multi-user KVMs, generally used in larger data centers where multiple administrators need access to network computers, essentially function to provide each user with what appears to be a single-user KVM
- **Application and technology:** In some networks, all of the computers connected to the KVM are local and on others, the connected devices are remote. The primary two types of KVMs in this category are single-user analog devices and multi-user digital devices:
 - Analog KVMs directly connect to a computer through an **unshielded twisted-pair (UTP)** or coaxial cable. The KVM switch and the connected computers are usually on the same LAN.
 - Digital KVMs connect to local and remote computers by an **Internet Protocol (IP)** address over an internet connection. A digital KVM is able to interact with a computer in the next room or half-way around the world.

Another type of KVM switch is the **USB enumerated KVM switch**, that is, a hub-based KVM. When the active port on a multiport USB devices changes, an enumeration process takes place. This process, which also occurs when you insert or remove a USB device, initiates and activates the port. There is a small delay while this takes place, but this type of KVM works well in **small office/home office (SOHO)** situations.

The range of the three devices that connect to a KVM have relatively short communication ranges. The PS/2 (mini-DIN) and USB of legacy and current keyboards and mouse units and the digital video of most current displays start to weaken at about 5 meters (10 feet). Some are capable of maintaining their signals for up to 10 meters (33 feet), but that can depend on several factors. In situations in which you want to extend the effective range of these devices, a **KVM extender**, which works much like an Ethernet network repeater, can lengthen the range of the device's signals up to 150 meters (just under 500 feet) on UTP and much farther on fiber optic cable.

Serial interfaces

Virtually all server-level computers have at least one serial port, that is, a COM port. Typically, the connector for this port is a **D-subminiature-9 (DB-9)** male connector, as shown in the following image. In its name, the **D** references its shape, the **subminiature** represents its smaller shape, and the **9** is for the number of pins in the connector. COM ports are typically DB-9 connectors on the newest computers. Another, although less common, is the DB-25 serial port, which is a D-shaped connector with 25 pins.

DB connectors are either male or female, meaning that they have pins, like the top connector in the following image, or they have pin receptacles, just like the bottom connector in the following image, which is commonly known a VGA connector on a PC:



DB-9 connectors, male (top) and female (bottom)

System and network administrators often connect to networking and internetworking devices through a serial interface to access system and network resources for configuring, monitoring, troubleshooting, and maintenance using a system console, that is, the root console or admin console. In general, a system console is a text entry and display device, such as a PC or Terminal, with data entry capabilities. Windows and Linux operating systems provide Terminal emulation software, such as **Cmder** and **Console2** for Windows or Terminator and **Guake** for Linux, which enable a PC to communicate with network devices over a serial connection.

Network-based hardware administration

System and network administration, in the context of hardware, deals mostly with the remote access by an administrator through software or hardware. Three of the network-based hardware/software administration tools you may encounter in the Server+ exam are as follows:

- **KVM over IP switch:** This provides an administrator with the capability to access local or remote systems over an IP network through a web browser. Using a KVM over IP switch, an administrator can monitor and manage remote hardware and software, power management, and internetworking devices. Some models can also record the video display, thus showing the operations performed.
- **HP integrated Lights-Out (iLO):** A network hardware and system management tool that is proprietary to HP that provides the capability for systems management over the internet.
- **Integrated Dell Remote Access Controller (iDRAC):** Another proprietary network management package that is one part of the Dell EMC enterprise infrastructure management system. Like iLO, administrators have the complete capability to manage remote network resources.

Network-based operating system administration

On a widely distributed network, a centrally-located administrator is able to remotely manage, configure, and monitor the operating systems running remote servers and workstations. Some examples of the tools an administrator can use to manage the OS on remote systems are as follows:

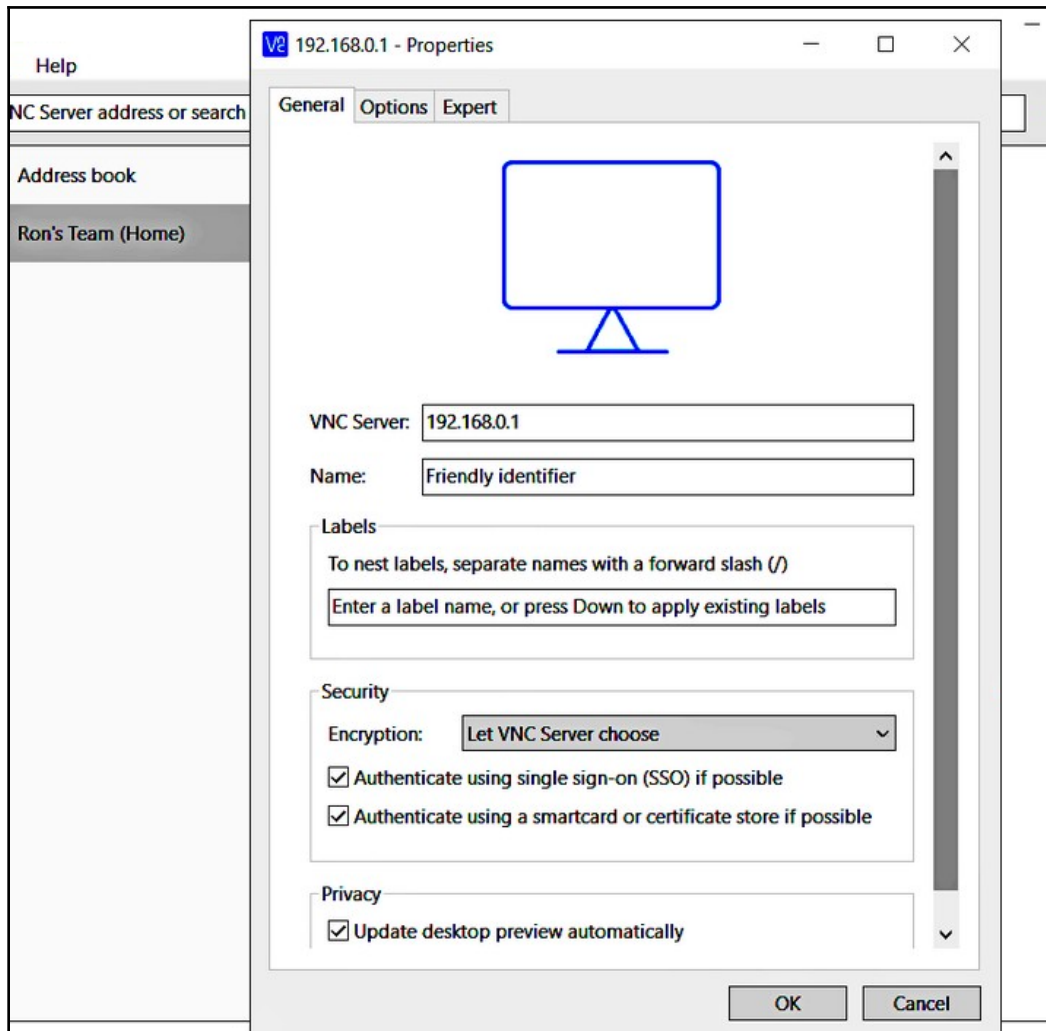
- **Remote Server Administration Tools (RSAT):** This Microsoft package enables an administrator to manage the configuration and features of Windows Server releases after version 2008 R2 from a Windows OS Vista and later system.
- **Remote Desktop Protocol (RDP):** This is another Microsoft system that provides remote access to network-connected systems that provide a GUI interface. The following screenshot shows the opening dialog box of RDP:



The RDP dialog box
Image source: Ron Price

- **Secure Shell (SSH):** A software utility that facilitates OS administration and file transfers over a secure remote connection.

- **Virtual Network Computing (VNC):** This provides a GUI desktop through the **Remote Frame Buffer (RFB)** protocol that enables an administrator to control and manage a remote system over a network. VNC transmits all keyboard and mouse actions to the remote system and returns all generated video back to the controlling system. The following screenshot shows the connection dialog box for the VNC Server:



The connection dialog box of a VNC server
Image source: Ron Price

- **Command line/shell commands:** Both Windows and Linux have the capability to invoke control of a remote system from the command line or via shell scripts. On a Windows system, PowerShell facilitates remote access and control. On a Linux system, the command-line feature over SSH provides remote system administration.

Asset management

In any organization, the assets of the computing and networking functions are among the most expensive and typically most valuable to its mission. Of course, this is a generalization, but data and the systems that store, protect, and process it into information require special handling, security, and maintenance.

Information Technology Asset Management (ITAM)

An ITAM program should include the financial, acquisition, and management activities that are applied to any major asset group of an organization. The primary purpose of an ITAM is to support both the short-term and long-term (tactical and strategic, respectively) life and application of the IT and networking functions of an organization.

The core of any successful ITAM program is the detailed information regarding the hardware and software and its inventory. This information provides the basis for the use, acquisition, replacement, and retirement of IT hardware and software components. Most organizations manage hardware assets independently of software assets. Typically, the IT hardware asset management activities include acquisition, application, and retirement and disposal. Depending on the asset type, an IT asset may also be depreciable. In contrast, not all software assets fall under the guidelines of managed asset properties. Software assets, which are generally software licenses, upgrades, and installations, are most definitely assets, depreciable or not.

An IT asset management program is likely to be different in each organization. There are several factors, components, and elements that potentially include in an ITAM. Some of the differences are large, such as what's included and what's not, or small, such as serial numbers, tags, and other identifying items. However, ITAM programs are about managing the IT asset life cycle.

IT life cycle asset management

An **IT life cycle asset management (LCAM)** program provides valuable initial and ongoing information to an ITAM concerning when existing equipment is in its life cycle and when the organization should acquire replacement equipment or new technology. A LCAM has five general phases:

- **Purchase:** When an organization acquires new equipment, it's added to the ITAM inventory and assigned an asset identity.
- **Implement:** Any necessary building changes or employee training takes place, along with the installation of the newly acquired equipment.
- **Maintain:** A program of periodic and preventive maintenance keeps the asset running as it should.
- **Support:** A periodic review of the value and contribution of an asset determines whether the asset is still contributing to the overall mission of the organization. If not, its life cycle may be shortened.
- **Dispose:** Not all assets at their end-of-life are ready for the trash heap. Especially with IT assets, many have some life or usefulness left in them and recycling may be a better choice than trashing them. Some IT assets have rare metals or other compounds that recyclers can extract from motherboards and other devices.

Of course, once a product has been useful between its purchase and its disposal, a replacement asset or a completely new technology may take its place, starting the cycle all over again.

Additional ITAM terms

Here are a few additional items you should about know for the Server+ exam, relating to asset management:

- **Asset inventory:** Another term for asset management is asset inventory, which means the periodic collection of detailed data of existing and new IT assets.
- **Asset tags:** A relatively permanent sticker or tag with a unique identification number marks any physical asset as an accountable asset.

- **Disposal/recycling:** An **IT asset disposition (ITAD)** program should include both the options of recycling serviceable IT assets and the proper disposal of each item type. So, if donating the equipment to a local charitable organization or school or trading it in isn't an option, check with your local government to learn how to recycle electronic equipment or dispose of it in your location.
- **End-of-life (EOL):** This is primarily a marketing term that refers to a product or service that is at the end of its life cycle and won't be available after a certain date. An organization with EOL devices installed knows that a new model, version, or edition of the device is to replace the EOL device. In most cases, the service and support for an EOL product may also be ending.

System documentation

Documentation is both a boon and a bane to a system or network administrator. It's extremely valuable when something is going wrong. It can also be a nuisance when it's time to create or update it. Regardless of how you look at it, love it or hate it, documentation is a valuable resource that requires active support. This section looks at the myriad of types and content of documentation that should exist for a network server, inter-networking devices, workstations, and, really, just about anything that attaches to a network.

The Server+ exam's objectives identify nine separate types of system documentation. The following sections outline each of the documentation types identified in the Server+ objectives.

Service manuals

Each hardware component added to a system should come with a service manual, commonly also called a user manual or owner manual. In some cases, the service manual is more of an installation guide that also includes information on minor problem troubleshooting methods with the bulk of the service, repair, or configuration information available through the manufacturer's website.

A service manual commonly contains some or all of the following contents:

- **Front matter:** This is the identification information on the first few pages of the manual that typically includes a cover page, introductory comments, product model and serial numbers, and a table of contents.
- **Installation/configuration guidelines:** Some service manuals have installation guidelines and configuration options and processes. Depending on the device or component, the installation and configuration processes are a combined process.
- **Troubleshooting:** In general, this portion of a service manual focuses on identifying common errors or problems with a device and perhaps the steps used to fix or remedy the issue. This section could contain detailed information on identifying a fault and its solution. It may just list the location of service centers or the customer support telephone numbers to call before attempting to repair the part.
- **Frequently asked questions:** Commonly asked questions related to the installation, configuration, and operations of the device, typically with short answers or explanations.
- **Glossary and index:** On occasion, a service manual may have a glossary to define and explain any technical or product-specific terms included in the manual and an alphabetical page index of its key terms and processes.

System and network documentation

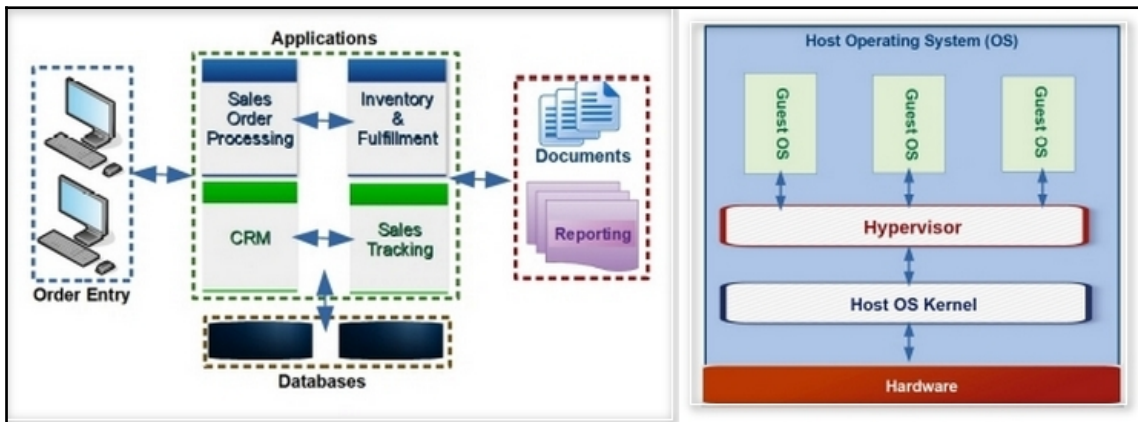
In far too many cases, the system administrator has an extremely detailed knowledge of the configurations of the various servers, a virtualized network environment, the storage attached network, and the inter-networking devices that connect to the internet. On the job, there is nothing the administrator doesn't know and can't fix, replace, or troubleshoot using only acquired knowledge. This administrator firmly believes that documentation really isn't necessary, because the administrator's knowledge would cover anything that could happen. Plus, keeping the documentation up to date is a major nuisance and a waste of time. Typically, in situations like this, management leaves the system and network administration to the administrator to do what is necessary, at least until an injury prevents the administrator from performing their duties.

While this scenario may seem like a scare tactic to convince you about the value of system and network documentation (only because it was), it's real focus is on developing a formalized process for creating, updating, and managing system and network documentation.

System diagrams

It's especially true in networking hardware that a picture is worth a long explanation. A diagram of a system's architecture, the topology of a network, or the data flow of a database system can quickly explain the structure, configuration, and functionality of an IT system. For the Server+ exam, you should be familiar with the following diagrams:

- **Architecture diagram:** The system architecture is shown in the following diagram. These diagrams depict the major components of an infrastructure, system, or even applications. Prior to system development, an architecture diagram models the intended development.
- **Data flow diagram (DFD):** A graphical depiction of how data flows through a system, network, or database. Typically, designers create DFDs in the early stages of a system's design. In the context of hardware, a DFD shows how data flows from device to device.
- **Network diagram:** A graphical representation of the devices and services in a local, wide, or other network. Network diagrams may show operating systems, routing protocols, and other services that are included:



Examples of a system architecture diagrams
Image source: Ron Price

System documentation

The documentation that's generated and maintained by an organization is directly related to the size and complexity of its data center, network, and applications. Smaller organizations require the same core documentation as larger ones, but bigger networks typically have additional areas that require documentation. At a minimum, the documentation of a computer system or a network should have two primary components: baseline and recovery.

Baseline documentation formalizes the equipment, devices, software, cabling, and all other components that have been installed in the system and documented in major subsystems. The baseline also documents the performance levels of the newly configured system or subsystems that will provide a comparison base for future performance measurements.

Recovery documentation, such as a disaster recovery plan or a business continuity plan, contains the details of the processes, locations, staffing, and procedures that are used to restore operations in the event of different levels of interruptions.

Effective documentation has four primary qualities:

- **Clear language:** The language and style of the documentation allows all affected parties to fully understand its purpose, specifications, requirements, and instructions.
- **Relative information:** Rather than include additional content, reference any other documents that relate to or are associated with the subject matter of a specific documentation.
- **Completeness:** Each specific documentation should contain all of the pertinent information so that readers are aware of all aspects of operations, good and bad.
- **Timely and accurate:** Update all affected documentation within a reasonable time frame after applying modifications, expansions, upgrades, or patching to the system. Also include the results of troubleshooting reported problems.

In addition, system documentation should exhibit four general characteristics: purpose, effectiveness, intended audience, and completeness. The goal is to achieve all four, if possible. Let's take a look at each of these characteristics:

- **Purpose:** The purpose or objective of the documentation is to provide the information required to detail the makeup and configuration of the system, create training for the system stakeholders, outline potential component upgrades or replacements, and to designate team member responsibilities.

- **Effectiveness:** In addition to the primary qualities that we've just listed, documentation must contain appropriate identification and replication procedures, as well as effective troubleshooting and resolution steps.
- **Audience:** The language should be free of technical jargon and terminology so that the intended audience can understand what you're talking about.
- **Completeness:** The documentation of an information system or network must contain the appropriate information for all aspects, components, operations, and uses of the system.

Someone must have said at one point, *you can never have enough documentation!* This statement may be true, but only if the documentation is up to date.

Other documents and documentation

There are several other types of documentation that relate to specific activities, such as the details of a system's configuration, the setup of an operating system, cabling specifications, service level agreements, change policies, personnel policies, authority structures, and the like. These documents tend to require less updating than, perhaps, main system documentation, but like it, they must be up to date.

Storing sensitive documentation

Some documentation may be of a sensitive or confidential nature or direct to the processing or handling of classified or categorized documents, data, and outputs. In these cases, the government, its agencies, the military, and businesses of all sizes, must have a separate, specific, and documented policy on its storage, safeguarding, and transport.

The US military defines a secured facility called a **sensitive compartmented information facility (SCIF)** in which documents are in open storage, closed storage, or continuous operations. Many non-governmental organizations store sensitive documentation and other documents in locked, fire-proof cabinets, or small versions of bank vaults. In any case, sensitive documentation requires the same level of protection as the sensitive data or materials it references.

Summary

A server administrator has several duties and responsibilities, including maintaining network and storage hardware, troubleshooting network issues, and implementing and monitoring security policies. The administrator's duties may also include installing software, maintaining the NOS, and implementing software security policies.

A KVM switch provides access and control for individual local or remote computers. We looked at two types of KVM devices—access and control and application and technology. Examples of network-based hardware/software administration tools include KVM over IP, iLO, and iDRAC. The tools that are used to manage an OS on a remote system are RSAT, RDP, and SSH. VNC provides a GUI desktop. Command-line/shell commands can also control a remote system. Windows PowerShell facilitates remote access and control and on a Linux system, command line over SSH provides remote system administration. An ITAM program manages major asset groups and the IT asset life cycle. An LCAM provides information on life cycle and replacement equipment using purchase, implement, maintain, support, and dispose.

Documentation should include the network server, inter-networking devices, and workstations. Types of documentation include service manuals, system and network documentation, architecture diagrams, data flow diagrams, network diagrams, baseline documentation, and recovery documentation. Effective documentation has clear language, relative information, completeness, and timely and accurate, as well as purpose, effectiveness, intended audience, and completeness. Documentation may include configuration, cable specifications, SLAs, and policies. Sensitive or classified documents, data, and outputs require special handling.

Questions

1. Which of the following are typically in the duties and responsibilities of a system or server administrator?
 1. Configuration
 2. Monitoring
 3. Implementing
 4. Construction
 5. 4 only
 6. 1 through 4
 7. 1, 2, and 3 only

2. What device allows an administrator to use local input and output devices to control a remote system?
 1. COM
 2. ITAM
 3. KVM
 4. LCAM
3. A DB-9 or DB-25 male connector on the back of a PC is typically what general port type?
 1. USB
 2. COM
 3. PS/2
 4. Mini-DIN
4. KVM over IP, iLO, and iDRAC are examples of what type of device or service?
 1. Localhost
 2. Network-based administration
 3. GUI
 4. SSH
5. A system administrator can use which of the following to manage the OS of a remote workstation?
 1. RSAT
 2. CLI over SSH
 3. VNC
 4. RDP
 5. All of the above
 6. None of the above
6. The management activities of information technology assets are known as:
 1. LCAM
 2. ITAM
 3. RSAT
 4. iDRAC

7. What is the acronym that refers to the purchase, implementation, maintenance, support, and disposal of an IT asset?
 1. LCAM
 2. ITAM
 3. RSAT
 4. iDRAC

8. Which of the following is not a common form of system documentation?
 1. Service manuals
 2. Architecture diagrams
 3. Recovery plans
 4. Typing tutor

9. Effective documentation has four qualities: clear language, relative information, _____, and timeliness.
 1. Policies
 2. Completeness
 3. Appropriateness
 4. Lengthy

10. System documentation has four characteristics:
 1. Purpose
 2. Effectiveness
 3. Intended audience
 4. Completeness
 5. All of the above
 6. None of the above

8

Server Maintenance

Remember that computers, no matter how sophisticated or expensive, are still just machines, although electronic. Electronic devices have very few moving parts. In fact, outside of the read/write heads in a secondary storage device, the open and close drawer on optical storage devices, and maybe a couple of more moveable items, the overall design of computers doesn't include many moving parts. We could say that the electrical signals that move about the computer are *moving parts*, but bits and their transmission media don't really cause many failure issues and server maintenance is all about preventing failure and resolving problems as they arise.

This chapter covers the duties and activities of the system administrator relating to the maintenance of a network server. This includes the definition and execution of a change and patch application and management program, monitoring the performance and health of a server's outward-facing components, the process used to troubleshoot and resolve issues on a network server, and the application of fault tolerance and high-availability technology.

We will cover the following topics in this chapter:

- Change and patch management
- Performance monitoring
- Hardware maintenance
- Fault tolerance

Change and patch management

Within the context of system and server administration, change management programs, commonly known as change control programs, have one purpose above all others—the assurance that the application of the appropriate and necessary patches, updates, and configuration changes follows a formalized controlled, coordinated, and consistent methodology. The two terms—**change management** and **change control**—commonly refer to service maintenance and software maintenance, respectively. However, for the Server+ exam, change management refers to server and hardware maintenance in general and OS maintenance in particular.

A formalized change management or control program is a written, distributed, reviewed, approved, and implemented document. Depending on the size of the organization, the people involved at different phases of the program include the IT staff, a change review board or committee, and possibly the stakeholders who the change impacts. Of course, in a smaller organization, the process typically involves only two or three IT people, due to the limited scope and effect of changes. In a larger organization, the number of participants grows with the potential scope and effect of a change. In either case, the process structure directs the reviewing, testing, and application of any system change.

The purpose of the change management program in a server management environment should include the following steps:

- Deciding the proposed change is both needed or applicable
- Identifying the impact of the proposed change and who or what it will affect
- Testing the change in a non-production environment
- Reviewing the results of the test
- Gaining approval to apply the change in the live environment
- Applying the change
- Benchmarking the post-change environment

Change control process

The process that should be a part of a change management and control program has six phases. These phases are as follows:

- **Purpose:** Understand what the vendor says about what the proposed change is to correct, patch, fix, or improve and decide whether this is a change that is necessary.

- **Scope:** Any change has primary and secondary effects on a data center's current operations. Therefore, knowing the full impact of a proposed change is essential before proceeding. This knowledge should include who, as well as what.
- **Approval:** Someone other than the system administrator should review the analysis of the preceding steps and grant approval to proceed or identify areas for further analysis.
- **Test:** Test proposed changes on a non-production system before implementing them on the live systems.
- **Implement:** If the results of the non-production environment testing are as expected, apply the change to the production environment. After applying the change, benchmark the system's performance.
- **Review:** For a set time period, check the results of the applied change to ensure its continued success.

Patch management

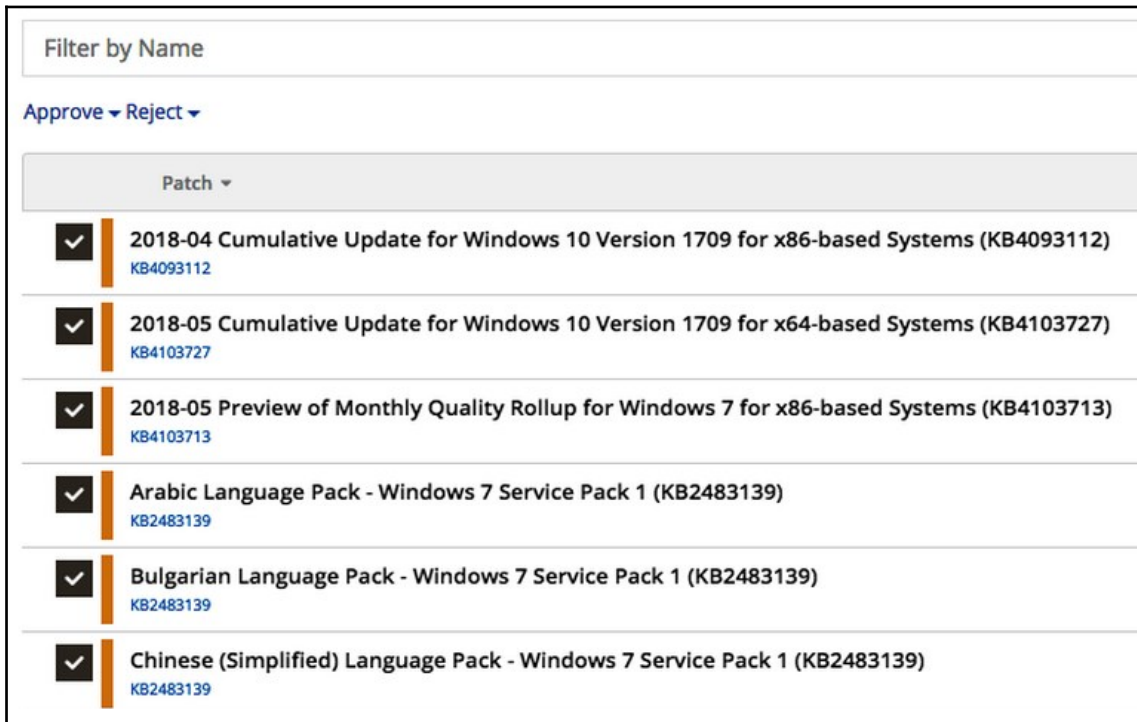
Patch management is a part of change management and control, but its importance is less for the operational system tasks and capabilities than it is for the security of the system. The majority of patches (vendor-supplied updates and fixes) address vulnerabilities identified in a software product, such as an OS or major application package. The application of patches to a system should generally follow the approaches discussed in the previous two sections. However, patch management does have a few requirements of its own:

- Patch management should be a priority
- Maintain an accurate inventory of IT assets
- Develop and apply a thorough testing process
- Assign responsibility
- Document the process, actions, and results

In today's environment of malware, hackers, worms, Trojans, and other evil-doers, the time frame between the identification of a vulnerability (a weakness or flaw in a system that a hacker could exploit) and the issuance of a patch to close it, may be only a few hours. In system environments that must exhibit **high availability (HA)**, the application of a patch could cause the system to be unavailable for an undetermined amount of time. The system administrator must analyze a patch for its appropriateness to the system and then decide on the best course for its implementation.

In larger, distributed data centers with remote locations managed from a central site, patch management (and change management for that matter) is a bit more complicated. Because of this, many use software tools, such as Microsoft's **Security Configuration and Analysis (SCA)** tool, **Windows Server Update Services (WSUS)**, or **System Center Configuration Manager (SCCM)**. Third-party patch management systems are available as either on-premise or cloud-based systems, such as SolarWinds Patch Manager (on-premise), ManageEngine Patch Manager Plus, or NinjaRMM (cloud-based).

The following screenshot shows the patch management display of the NinjaRMM system:



A patch management scheduling display. Image courtesy of NinjaRMM

OS updates

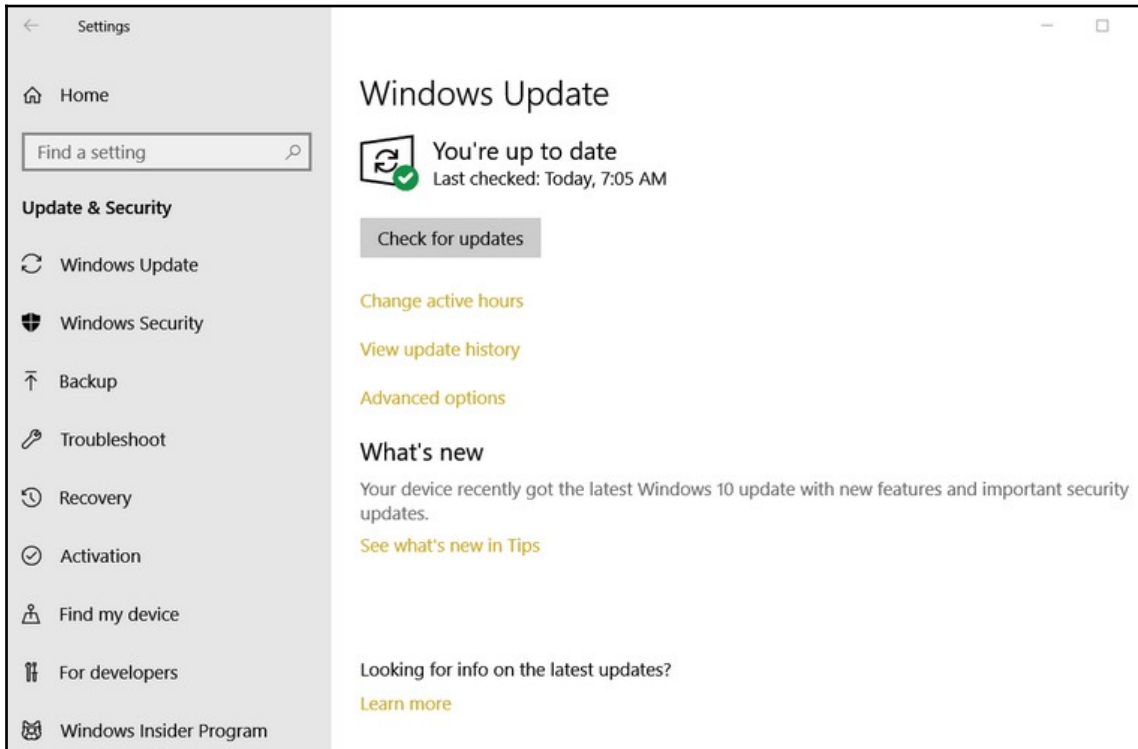
In addition to knowing if the OS is up-to-date and free of any known errors or bugs, applying fixes, patches, and updates is key to the security of the server and, most likely, its network. OS updates correct reported errors in the code and close vulnerabilities before or because of exploitation.

Most OS, including Windows, macOS, and many of the Linux distributions, include an automatic update feature. However, the update policy of the organization should direct whether this feature is enabled or disabled. Disabling automatic updates doesn't eliminate the administrator's ability to apply the update; it just won't happen automatically. Some OS updates can conflict with corrections or instructions inserted as workarounds for a flaw the patch is correcting. Prior to applying an update or patch, a thorough analysis of its effects and before and after testing should be a part of any change management procedure.

To enable or disable the automatic update features in Windows, macOS, and Linux , follow these guidelines:

- **Windows Server:** On an **Active Directory (AD)** system, control of the automatic update function may be at the group level. The **WSUS** and the **Windows Update** and **Maintenance Scheduler** settings in Group Policy set the parameters for when, what, and how OS updates occur. Use these settings to disable automatic updates as well. In addition to OS updates, WSUS also updates Microsoft applications.
- **Windows 7 and 8:** The feature used to enable and disable automatic updates on a Windows 7 or earlier version is on the Control Panel under the **System and Security** header and titled **Windows Update**. This feature provides three options:
 - **Turn automatic updating on and off:** If turned off, manual update is available through the Windows Update option on the Start Menu.
 - **Check for updates:** This searches for unapplied updates that are available for the Windows OS, Microsoft Office, and other Microsoft services for manual application.
 - **View installed updates:** This lists the installed updates of the OS, including their coverage and when posted.

- **Windows 10:** Control of automatic updates is enabled or disabled through the **Settings | Windows Update** functions (see the following screenshot):



Windows 10: Windows Update page

- **macOS:** Control of automatic or manual OS updates is effected through a page in the Apple Store. As shown in the following screenshot, there are three levels of control—automatically checking for updates; downloading and possibly installing the update; and downloading and installing previously purchased software installed on another macOS:



The Apple macOS automatic update configuration

- **Linux:** Each Linux distribution has an automatic update utility. For example, Ubuntu and Debian Linux releases have the *unattended-upgrades* utility that downloads and applies security updates and system patches. Another Linux utility for automatic updates is the `yum-cron` command and parameters in Red Hat Enterprise Linux and associated distributions.

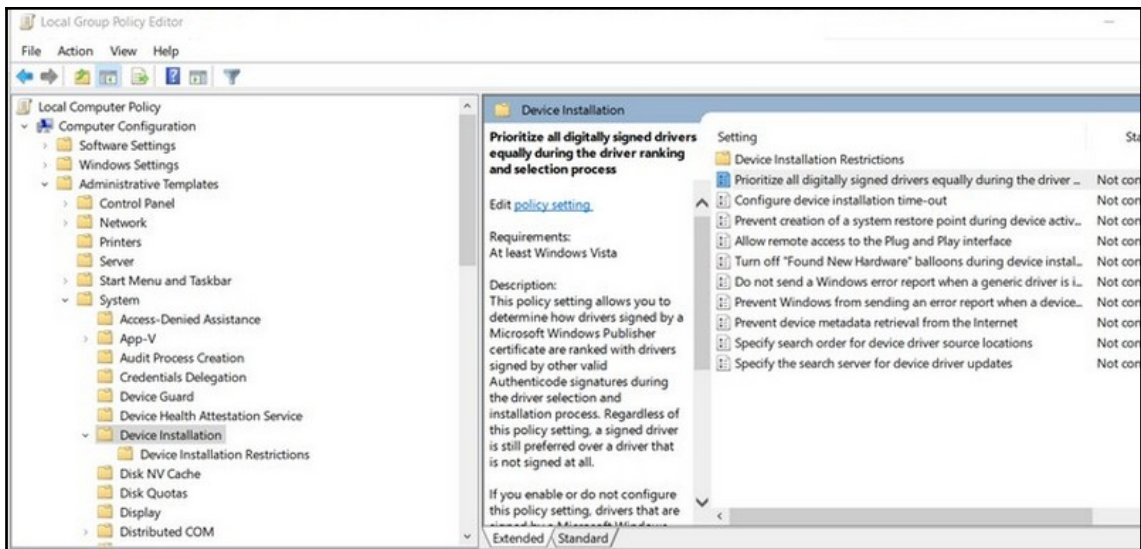
Device driver updates

On a Windows system, updates to device drivers can also be automated or turned off, whatever the case may be.

To set automatic device driver updates on (or off), apply the following steps:

1. Run the `gpedit.msc` utility to open the **Group Policy Editor** (see the following screenshot).
2. Navigate through **Computer Configuration** | **Administrative Templates** | **System** | **Device Installation** (and **Device Installation Restrictions**, if you are turning off this setting).

3. In the **Setting** window, select the policy settings you wish to activate:



Use the Group Policy Editor to configure automatic device driver update settings

Firmware updates

Many system administrators follow the, *if it isn't broke, don't fix it*, approach to firmware updates, and only consider updates that are improving or adding security functions. However, there are reasons to upgrade the firmware, even when security isn't the issue. For example, when you add a newer hardware component to a relatively older computer, the computer most likely needs a BIOS/UEFI update to enable a compatible interface to the component. Consider firmware updates in the same way as an OS or application update. A firmware update may not affect a particular server or network and may not be necessary. It's also important to remember that a firmware update intended to fix a bug may have bugs of its own.

It's also a good practice to update the firmware of a computer when building it up as a server, or when the manufacturer of the computer or motherboard says that you should update the firmware. In either case, how you go about updating the firmware varies among manufacturers of motherboards and computers. Check the manufacturer's website for the process and application instructions.

Hardware maintenance

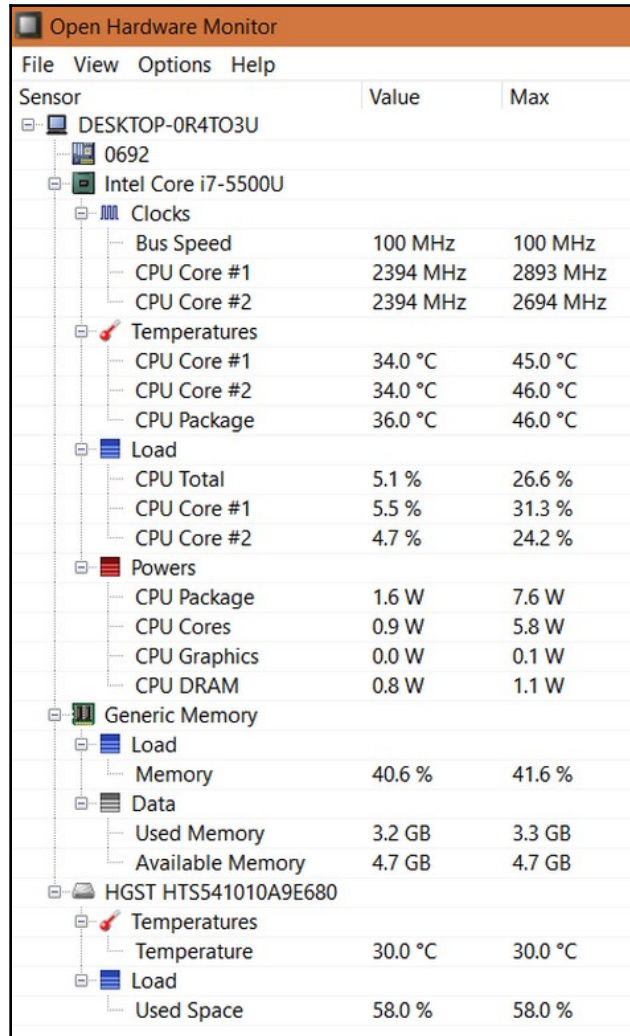
In networks that require high availability, network and system administrators must rely on a variety of monitoring, error notification, and troubleshooting guides to detect and alert them when problems occur. On high-availability systems, downtime is not an option. All support and maintenance activities focus on ensuring uptime and availability. In the following sections, we look at some of the tools available to assist administrators in achieving these goals.

Server monitoring systems

Server monitors provide system administrators with automated reporting, scheduled device checking, and warnings that a device or system may be nearing or has reached a preset threshold or ceiling setting and requires preemptive troubleshooting. The majority of server monitoring systems include checks and measurements and reporting, if necessary, in relation to the major systems, subsystems, and components of a network server, including the following:

- **CPU usage:** High CPU utilization can impact response times and productivity.
- **Hard disk space:** Low disk space can lead to slower performance, missing updates, high fragmentation, and slow indexed searches.
- **Disk input/output operations per second (IOPS):** Disk IOPS is a benchmark measurement that indicates the operating efficiency of a secondary storage device.
- **RAID health:** RAID systems can have a variety of issues, such as controller failings, errors in partitioning, and, of course, one or more physical disk drive failures.
- **RAM utilization:** High RAM utilization can cause slower runtimes, throughput, and perhaps system or application crashes.
- **Hardware status:** A poorly performing hardware component can slow or halt a network server. Proactively checking hardware health and processing capacity may avoid system downtime.

- **System temperatures:** Several components in a server (or network node) are sensitive to high heat, including the microprocessor, which is why system temperatures need monitoring (see the following screenshot):

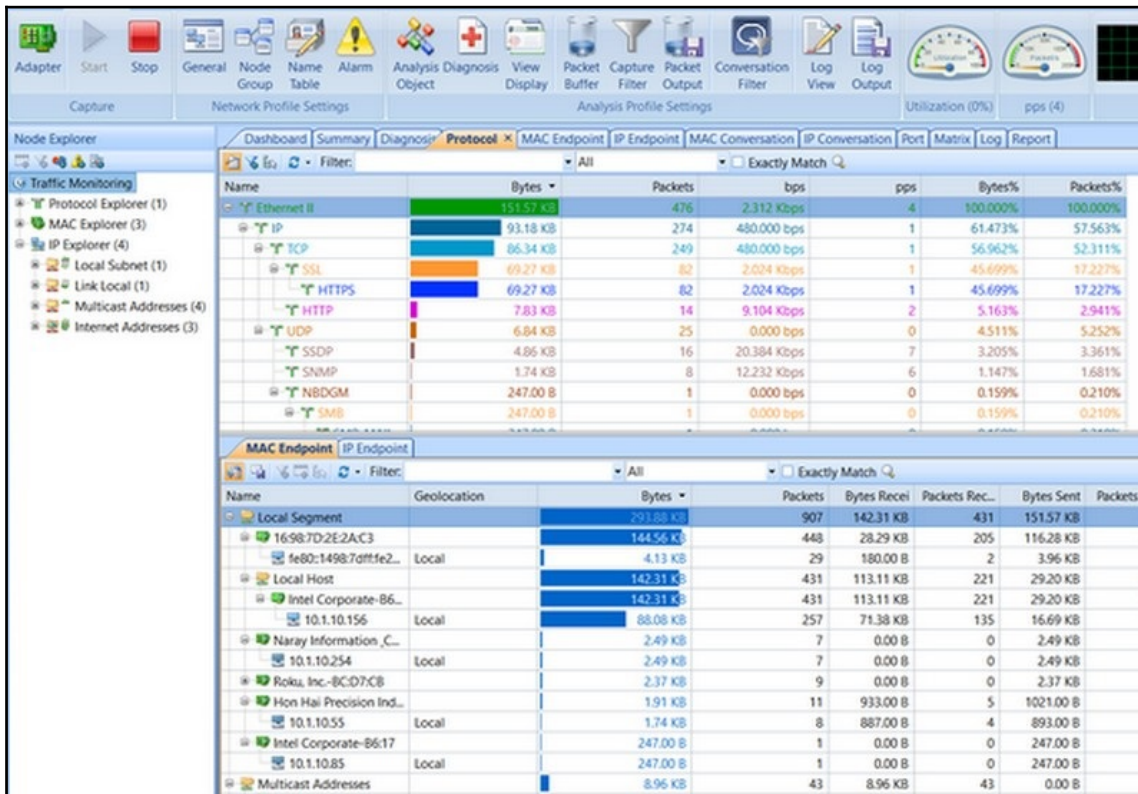


The screenshot shows the Open Hardware Monitor application window. The interface includes a menu bar (File, View, Options, Help) and a tree view on the left listing hardware components. The main area displays a table of sensor data for the selected component, DESKTOP-0R4TO3U. The table has columns for Sensor, Value, and Max. The components listed are Intel Core i7-5500U, Generic Memory, and HGST HTS541010A9E680. The Intel Core i7-5500U section shows sub-entries for Clocks, Temperatures, Load, and Powers. The Generic Memory section shows Load, Memory, and Data. The HGST HTS541010A9E680 section shows Temperatures and Load.

Sensor	Value	Max
DESKTOP-0R4TO3U		
0692		
Intel Core i7-5500U		
Clocks		
Bus Speed	100 MHz	100 MHz
CPU Core #1	2394 MHz	2893 MHz
CPU Core #2	2394 MHz	2694 MHz
Temperatures		
CPU Core #1	34.0 °C	45.0 °C
CPU Core #2	34.0 °C	46.0 °C
CPU Package	36.0 °C	46.0 °C
Load		
CPU Total	5.1 %	26.6 %
CPU Core #1	5.5 %	31.3 %
CPU Core #2	4.7 %	24.2 %
Powers		
CPU Package	1.6 W	7.6 W
CPU Cores	0.9 W	5.8 W
CPU Graphics	0.0 W	0.1 W
CPU DRAM	0.8 W	1.1 W
Generic Memory		
Load		
Memory	40.6 %	41.6 %
Data		
Used Memory	3.2 GB	3.3 GB
Available Memory	4.7 GB	4.7 GB
HGST HTS541010A9E680		
Temperatures		
Temperature	30.0 °C	30.0 °C
Load		
Used Space	58.0 %	58.0 %

An example of a hardware status monitor

- **Network utilization:** This metric is the ratio of the current traffic load on a network to the network medium's maximum throughput to indicate how busy the network medium truly is at any given moment. The following screenshot shows an example of a dashboard display from a network utilization program:



A screen capture of the Dashboard display of the ColaSoft Capsa software

- **Virtual machine performance:** An active virtual machine can have counters enabled to track its activity and performance levels, such as, input/output operations to a virtual disk, the amount of memory in use, and its network traffic volume.

Light Emitting Diodes (LED) server status indicator

Computer systems designed for use as a network server have one or more sets of LEDs on the motherboard, rear panel, or front panel, or some combination of these locations. The LEDs display color and/or blinking patterns to indicate the status of the system, which can range from *All is well* to *Power off now*. Server manufacturers differ on where the LEDs are located and what their combinations of colors and blinking means.

The following table shows an example of the LED server status indicators for an Intel server board.

Color	Display	Meaning	Action
Green	Steady	System normal	None
Green	Slow blink	System degraded	Memory error Baseboard Management Controller (BMC) detected error
Amber	Slow blink	Non-fatal flaw	Memory error—error threshold exceeded
Amber	Steady	Fatal	System failure—CPU configuration error
None	N/A	System not ready	AC power is off

Sampling of server board LED status indicator lights

Liquid Crystal Display (LCD) messages

In place of or in addition to LED indicators, several server systems display server status information on a small LCD that typically has one line of text limited to between 25 and 65 characters. Should an issue occur during the POST or a subsequent configuration, a short message displays on the LCD screen, like the examples given in the following table. This display is often exclusive to the server display and not displayed on the administrator console. Some servers, such as the Dell PowerEdge Server R-and Tx20 series, color the background in the display to indicate the severity of the issue, similar to the illustration in the following image:



An example of a server LCD display

Examples of server LCD error codes and messages are shown in the following table:

Error code	Message	Meaning
E1114	Temp Ambient	Ambient system temperature is out of acceptable range
E1210	CMOS Batt	CMOS battery is missing, or the voltage is out of acceptable range
E1410	CPU # IERR	Microprocessor # is reporting an internal error
E1714	Unknown Err	There has been an error, but BIOS is unable to determine its origin
E1810	HDD ## Fault	HDD ## has experienced a fault
E2014	CMOS Fail	CMOS RAM not functioning properly
E2019	Parity Error	Main memory parity error
E201E	POST Mem Test	BIOS POST memory test failure

Examples of server LCD error codes and messages
Source: <https://www.dell.com/support/>

Beep codes

When you power on a PC, a firmware utility, called the **power-on self-test (POST)**, checks the internal hardware components included in the BIOS or UEFI configuration. The POST checks each of the components for its presence (connection), compatibility, and function (response). If all is well, the POST issues an all-clear signal and continues the start up procedure. The signal given is generally a single *beep* tone, but some systems may beep twice.

However, should the POST encounter a problem, meaning the PC fails the POST, one of two things may happen—a beep code pattern is emitted that indicates the nature of the problem detected, or there is a power-off without sounding a beep. Beep codes, also known as POST error codes, when used, provide a general indication of the component causing the POST fail. It could be a bad connection, a removed device, or, perhaps, a failed device. The purpose of the beep codes is to provide a starting point for troubleshooting the issue.

Beep codes are a part of the BIOS/UEFI module stored in a ROM or NVRAM chip. Unfortunately, there is no standard, and each manufacturer can use a different beep code pattern scheme in its computers. Even different BIOS products of a single company may have completely different code meanings. As shown in the following table, the three manufacturers shown, while using the same beep code patterns, each assign a different issue or condition to their codes. A short tone or beep is a quick beep that is about one-tenth of a second in length. A long tone is about twice as long:

Failure /issue	AMIBIOS (American Megatrends)	Award BIOS (Phoenix Technologies)	IBM corporation	Dell computers
DRAM	1 beep	Continuous tone	-	2 shorts
CPU	5 beeps	Repeating low and high tones	1 long, 1 short	7 beeps
Keyboard	6 beeps	1 short, 2 shorts, 2 shorts, 1 short	3 longs	-
CMOS	10 beeps	1 long, 4 shorts	2 shorts	1 short
Display	8 beeps	1 long, 3 shorts	1 long, 3 shorts	6 shorts

Each vendor may have completely different beep tones

Replace failed components

The components that make up the hardware of a network server are in two major groups, each with its own handling, installation procedures, and complexity. The following table shows the categories of several server components.

- **Customer-Replaceable Units (CRUs):** These are the components that a user/customer is able to remove and replace. CRUs fall into levels of difficulty, such as a CRU 1 being relatively simple to replace and a CRU 3 being more difficult. In some systems, a CRU N or CRU X indicates a component that a user should not attempt to replace. Examples of CRUs are monitors, keyboards, batteries, and all external devices.

- **Field-Replaceable Units (FRUs):** These are the components that only a qualified field service representative should remove and replace. Examples of FRUs include hard disk drives, motherboards, internal control units, fans, backplanes, and memory:

Component	CRU/FRU
CMOS battery	FRU
DIMMs	FRU
HDDs	CRU
Internal cables	CRU
Memory card	FRU
Microprocessor	FRU
Power supply unit	CRU
RAID controller	FRU
SSDs	CRU

Examples of CRU/FRU designations for server components

Preventive maintenance

Unfortunately, computers, and especially servers, aren't *set 'em and forget 'em* devices. They require constant monitoring and administration. A very large part of these efforts is to organize a schedule of preventive maintenance activities, followed and recorded assiduously. They are necessary for just about every scheduling frequency: daily, weekly, monthly, quarterly, yearly, and perhaps between them.

The primary purpose of preventive maintenance is to avoid device or component failure, which creates corrective or restorative maintenance. Just like taking a flu shot helps to prevent getting sick with the flu, a solid **preventive maintenance (PM)** program can prevent serious problems down the road. Of course, the PM program is typically different for every system, data center, and network, but there is a list of tasks that any PM program should include.

The following table lists a sampling of these actions:

Frequency	Task
Daily	Check server error and usage logs to identify potential problems
Weekly	<ul style="list-style-type: none">• Check disk space on servers• Clean paper dust out of printers• Check for appropriate air flow• Ensure that antivirus software is up to date• Audit system users to ensure adherence to virus policy• Verify that OS patches and critical fixes are up to date and installed• Check event log for errors• Check system resources, such as HDD, RAM, and CPU, for availability and performance
Monthly	<ul style="list-style-type: none">• Check batteries on laptops and mobile devices• Monitor and collect data on trends for memory, CPU, and disk utilization
Quarterly	<ul style="list-style-type: none">• Clean dust from cooling system and fans• Clean keyboards, mice, and other moving parts
Annually	<ul style="list-style-type: none">• Test uninterruptible power supplies• Check network wiring• Review the effectiveness of the preventive maintenance program

A sample server preventive maintenance schedule

Fault tolerance and high availability

One of the terms used to describe server systems that are able to resist failures is *hardened*. While hardening is most associated with closing unused ports and other vulnerabilities, to achieve a truly hardened system usually requires the application of fault-tolerance techniques.

When one component fails, it may affect the function of another component, which then affects the functions of yet another component. This chain reaction of failing components creates a condition of cascading failure, which is not a good thing. There are several methods through which a server can be fault (fail) tolerant, meaning the server is able to withstand a component failure and stay available, achieving high availability.

Clustering

Where a server is generally a single computer running server software, a server cluster consists of servers arranged in an interactive group. The general purpose of a server cluster is to improve process throughput and responsiveness to users. A server cluster has two or more interconnected servers that operate as a single unit under the control of a primary or load-balancing server or controller.

The primary benefits provided by server clustering are as follows:

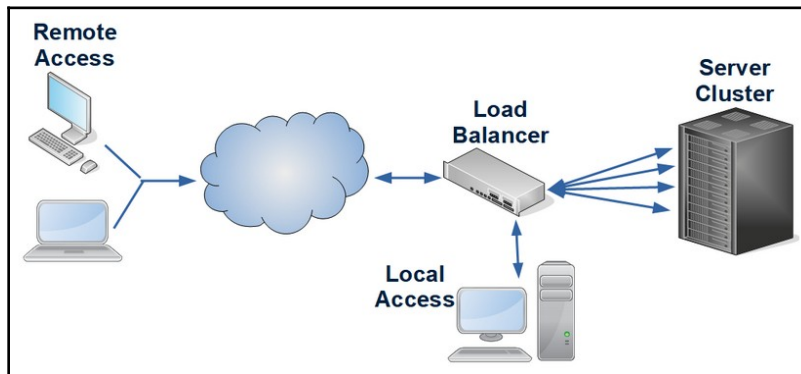
- **Scalability:** In situations where network or user demands exceed the capabilities of an existing cluster, it's relatively easy to incorporate additional servers into the cluster to handle the increased volume.
- **Reliability:** A single server is both a single point of failure and, at times, a bottleneck. A server cluster spreads the processing load across several servers, which is likely to withstand the failure of a single server in the cluster.
- **Manageability:** Maintenance on a single server generally requires downtime. However, a cluster of servers provides continuity of operations, while one or more of the clustered devices receive maintenance.

However, there are some disadvantages to server clustering, including the following:

- **Infrastructure:** A server cluster requires more computers and support infrastructure, which adds to the expense of maintaining the system.
- **Compatibility:** Not all computers designed as servers are cluster-friendly. Clustering is a fairly rigid structure and beyond add-on scalability is not very flexible. Some software applications aren't cluster-compatible either.
- **Cost:** Clustering can be expensive and its design must be carefully set to work efficiently, increasing the cost of administration and management.

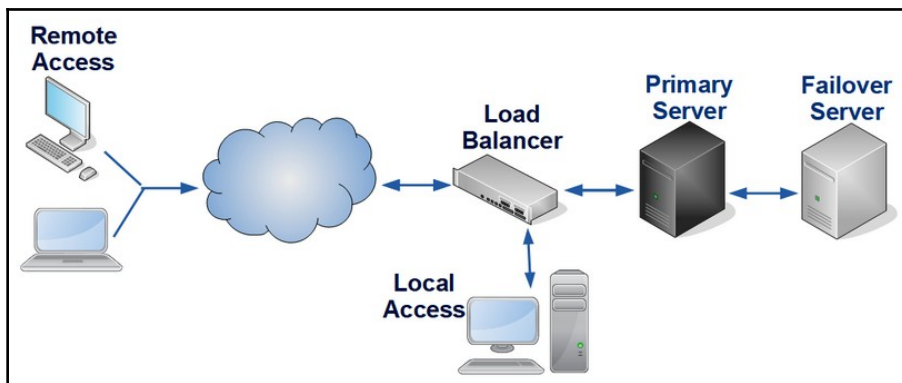
Active/active versus active/passive clusters

The two primary configurations for server clusters are **active/active** and **active/passive**. Each of these configurations can be a strategy for high availability. An active/active server cluster is commonly a load-balancing solution that interconnects two or more computers that perform the same processing steps. As depicted in the following diagram, a blade server chassis holds several active servers. A load-balancing appliance or server manages the process assignments to the servers optimizing the response time of the network:



An example of an active/active server cluster

By contrast, an active/passive server cluster configuration, as shown in the following diagram, also has two or more interconnected servers, and may have one or more active servers sharing the processing load and one or more failover, backup, or standby servers ready to replace a failed server or one taken out of service for maintenance:



An example of an active/passive server cluster

Load balancing

As discussed above, load balancing is a function that distributes incoming network traffic to two or more servers arranged in a pool, farm, or cluster. Load balancing can be performed by specific software, a hardware appliance, or a dedicated computer. A load-balancing server or appliance enables a client/server application or a content provider to even out high volumes by spreading the incoming service requests over the available capacity of the server group. In effect, a load balancer is a form of router. Where a router directs traffic based on the best path to a destination, the load balancer directs traffic to a server with the appropriate software or data or, like a router, the load balancer directs the traffic to the best available server.

The primary functions of a load-balancing application or appliance is to distribute incoming client requests for service to the server cluster to achieve efficiency, provide high availability by balancing its traffic only to active servers, and to provide flexibility by permitting administrators to add or remove servers to and from the cluster as needed.

Load balancers use a variety of algorithms to determine the server in a cluster or group to which an incoming message goes for processing. The most common of the load-balancing algorithms used are the following:

- **Agent-based adaptive load balancing:** This is a supplemental algorithm used in conjunction with a *weighted* algorithm. Each server in the cluster has an agent that interacts with the load balancer to provide a real-time update on its server's status. The load balancer uses this data to determine which server will process the request.
- **Chained failover:** The servers in the processing group or cluster form a serial chain. Incoming messages go to the first server in the chain. If that server isn't able to service the message, it passes to the next server for processing. If that server is unable to provide processing, the message passes along until a server is able to process it.
- **Least connections:** Neither of the round robin algorithms consider the current load of the server group when deciding where to send a message. However, the least connections algorithm looks at the load on each of the servers and assigns a message to the server with the least number of active processes.

- **Round robin:** In an active/active cluster, every server is able to process the message traffic equally. While all of the load balancers and each of the servers all share the same domain name, each has a unique IP address assigned. The primary DNS server for this domain associates the domain name with each of the IP addresses in the cluster. When a request asks for the IP address for the cluster domain, it's provided with a server IP. The IP address provided for the domain name rotates after each DNS request.
- **Software-defined networking (SDN) adaptive:** This load-balancing technique combines data about the network from the Presentation and Transport Layers with data from the Data Link and Network Layers to learn the status of the servers and their active applications, the status of the network itself, and whether there is any congestion or blockage on the network medium to make its load-balancing decision.
- **Source IP hash:** This algorithm converts the source and destination IP addresses of the client and server and hashes them together to create a key that it uses to forward the message to a particular server. The hash key is recalculable should processing interrupt and restart.
- **Weighted least connection:** The servers in the cluster or group receive a weighted value based on their resources and capabilities. Servers with more resources receive higher values. This value combines with the connections count of each server to determine a server prioritization. Servers with more resources get a bigger share of the load.
- **Weighted response time:** An intermittent server response time check sets a ranking for servers based on their response times. A server under a heavier processing load responds more slowly than a server with a lighter load, resulting in the servers with the fastest response times receiving more messages—at least until the response time check repeats.
- **Weighted round robin:** In addition to the basic round robin algorithm, each server receives a weighting factor. The servers with higher weightings receive a larger share of the incoming messages.

Heartbeat

A heartbeat mechanism, also known as a heartbeat network or heartbeat protocol, is a common distribution method in a clustered server arrangement. Each of the clustered servers communicates with a synchronizing server, or *sync*, that it is up and operating by sharing that they have a heartbeat.

Each of the active servers in the cluster periodically sends a heartbeat message to the *sync*, indicating its health. The sync adds the heartbeat messages to the bottom of a **first-in/first-out (FIFO)** push-up stack. The top entry in the FIFO stack identifies to the load balancer or cluster manager which of the active servers is next to receive a message request. However, if a server fails to provide its heartbeat message in the allotted time, the sync assumes that the server doesn't have a heartbeat and is therefore unavailable.

Hot and not hot

A **high-availability (HA)** strategy aims to keep a server, network, or cloud service running and available to users and subscribers without the need to take the server down to change out a failed or idled component. HA programs can address several aspects of server availability, including data restoration, server failover, component replacement, and disaster recovery. On the Server+ exam, you should expect to see a question or two concerning component replacement and the differences between hot-swappable and non-hot-swappable devices and procedures.

Hot swap

In spite of its common usage, hot plugging is not hot swapping. Typically, a hot plug device will connect to the system through plug and play, but often there is still another step required to complete the installation. On the other hand, a hot swap device installs on a running computer and is immediately usable.

Perhaps the most common hot swap device is a **Universal Serial Bus (USB)** flash memory drive. While USB flash drives aren't typically secondary storage on a network server, other USB, FireWire, Thunderbird, and eSATA devices, such as external storage devices, network adapters, and other peripheral devices, are generally hot swappable. Components usually thought of as internal devices, such as power supplies and hard disk drives, can also be hot swappable. However, not all power supplies and hard disk drives are hot swappable, and not all computers support hot swapping.

Non-hot swap

The primary difference between a hot swappable device and a non-hot swappable device is that the replacement of a non-hot swappable device typically requires a reboot of the computer. Regardless of how fast a system reboot completes, the system was not available to users during that time. This method of device replacement is a warm swap.

Another not-hot device replacement method is a cold swap that requires a system shutdown (powered down).

Service level agreements (SLA)

An SLA is a common instrument provided by a service provider to the service subscriber. The essential elements of an information system SLA are as follows:

- **Parties:** The specific identification of all of the parties to the agreement or their agents listing their authorities and abilities under the agreement
- **Services:** The services specifically covered by the agreement, defined by function, procedure, actions, time, and pertinent metrics
- **Performance:** A specification of the time, volumes, duration, acceptable performance, unacceptable performance, uptime commitment, downtime limits, and agreed-to metrics and thresholds
- **Implementation:** If the implementation of the covered services is on a schedule, the agreement must specify the timelines, tasks, objectives, and acceptance or completion criteria

Other areas typical to an SLA are as follows:

- **Scheduled downtime:** A delineation of any agreed-upon scheduled system downtime, during which access is unavailable
- **Unscheduled downtime:** *What are the rights of the service subscriber in the event the system suffers unscheduled, extended, or catastrophic downtime, and what are the agreed-upon recovery requirements of the service provider?*
- **Client notification:** The SLA should detail the methods the service provider will use to notify the subscriber of changes in system status and define the amount of time for a notification of a scheduled downtime event
- **Mean time to repair/restoration (MTTR):** A committed downtime duration, either historical, an estimate, or industry data, in the event of unanticipated downtime

Summary

Change management programs ensure that the necessary patches, updates, and changes follow a controlled and consistent process. Change management in a server environment should follow six phases—purpose, scope, approval, testing, implementation, and review. Patch management is a part of change management and control. Patch management should be a priority. Firmware updates generally fix a bug.

Server monitors provide automated reporting, scheduled device checking, and threshold and troubleshooting warnings for CPU usage, hard disk space, and utilization, disk IOPS, RAM utilization, and network traffic. The components of a network server are either CRUs or FRUs. Preventive maintenance avoids device or component failure.

The purpose of a server cluster is to improve process throughput and responsiveness to users through two or more interconnected servers operating as a single unit under the control of a primary or load-balancing server. The two configurations for server clusters are active/active and active/passive. An active/active server cluster is a load-balancing solution that interconnects two or more computers that perform the same processing steps. An active/passive server cluster has interconnected servers and one standby server to replace a failed server.

Load balancing distributes incoming network traffic to two or more servers to spread incoming service requests over the capacity of the server group. An HA strategy keeps a server available to users during the change-out of a failed component. A hot swappable device installs on a running computer and is immediately usable. A non-hot swappable device requires a reboot of the computer.

An SLA is a common instrument provided by a service provider to the service subscriber. The essential elements of an information system SLA are parties, services, performance, and implementation. Other areas typical to an SLA include scheduled downtime, unscheduled downtime, client notification, and MTTR.

Questions

1. Which of the following is not a phase of a change control process?
 1. Approval
 2. Immediate application
 3. Implement
 4. Purpose
 5. Scope
 6. Test
2. One difference between a change management program and a patch management program is as follows:
 1. Patch management should be a priority
 2. Patches should use the same change management process as major OS updates
 3. Patch management requires no testing
 4. A software patch only affects the user interface
3. What type of software provides for scheduled device checks, including checks against preset thresholds, hard disk utilization, disk IOPS, and other system health metrics on a server?
 1. Asset management systems
 2. SNMP-MIBs
 3. Server monitors
 4. Packet sniffers
4. What is the significance of the IOPS metric?
 1. It is an estimate of the maximum channel/bus bandwidth
 2. It represents the maximum number of reads and writes (input/output operations) to and from non-contiguous storage locations on secondary storage devices
 3. It reflects the number of independent operations a CPU performs in a second
 4. It measures the number of I/O operations per second made by the CPU

5. If the BIOS/UEFI start up process detects an error or possible error, it signals a code to identify the issue and its source. What are these audible signals commonly called?
 1. Error indicators
 2. Failure alarms
 3. Beep codes
 4. POST alerts
6. A server component or peripheral that can be installed, configured, or removed by the server owner's staff is categorized as what type of device?
 1. **Field replaceable unit (FRU)**
 2. **User replaceable unit (URU)**
 3. **Remote replaceable unit (RRU)**
 4. **Customer replaceable unit (CRU)**
7. The objective of a preventive maintenance program is:
 1. To extend the service life of a server component
 2. To avoid component failures
 3. To maintain server uptime commitments
 4. All of the above
8. Improving process throughput and responsiveness by interconnecting two or more servers in a cluster to operate as a single device describes what server configuration?
 1. Active/active
 2. Active/passive
 3. Passive/passive
 4. Hot swappable
9. The server cluster configuration in which a standby or failover server is available to replace a failed server, if required, is a(n):
 1. Active/active
 2. Active/passive
 3. Passive/passive
 4. Hot swappable

10. A failed server device or component that an administrator may replace while a system remains running and productive is a:
 1. Cold swap
 2. Warm swap
 3. Hot swap
 4. Just right swap

9 Virtualization

When something is virtual, it's real, but not really. Virtual computers exist inside an actual physical, reach-out-and-touch-it computer. Okay, it's not really real, but it's inside a computer? Yes, you've got it! However, just to clarify for our mutual understanding, Merriam Webster's dictionary defines virtual (in the context of computing) as *being on or simulated on a computer or computer network*. This is a very good definition in that it describes exactly how a virtual device comes to be and where it exists.

A virtualized network environment provides an organization with the capability to create multiple **virtual machines (VMs)**, each running a different operating system and applications. This allows the organization to maximize the capacity of its IT resources by extending the capabilities of a single computer to provide direct computing capacity to several users simultaneously.

In this chapter, we will look at the concepts and applications of virtualization in a computer network environment. This includes the terminology, topology, components (virtual and physical), and their configurations and purposes. With the understanding of these elements of a virtualized environment, we can then see the purpose and operation of a virtualized network. The specific areas we will discuss are as follows:

- Virtual environments
- Hypervisors and virtual network managers
- Virtual hosts and guests
- Hardware configurations for a virtual environment
- Resource allocations in a virtual network

Virtual networking

There are three types of virtualization that are used in networks, each with its own specific purpose and operation:

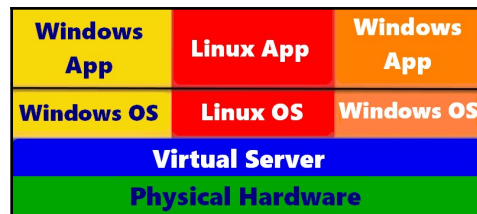
- **Virtual private network (VPN):** A secure connection through the internet using tunneling protocols
- **Virtual local area network (VLAN):** A grouping of network nodes in a logical domain
- **Virtual network:** Not to be confused with a VLAN, a virtual network employs virtual, software-induced components to extend the capabilities of physical computing equipment

A VPN creates a virtualized direct connection over the internetwork, between a remote device and an internal network. The purpose of a VPN is to provide the same safeguards and security that a directly wired connection should provide. A VLAN creates a logical arrangement of network nodes, regardless of their proximity, into a single domain structure. A VLAN is a virtual subset of a network that is a collision domain. Don't confuse a VLAN with a subnetwork, though.

Virtual network components

The essential parts of a virtualized network environment, as illustrated in the following diagram, are the physical hardware, a virtual network server, and one or more virtual machines, each running a guest operating system, which, in turn, supports applications running in the virtual space.

Underneath any virtual environment is a physical infrastructure that must be able to support the resource needs of the virtualized environment you envision. As we discuss the components of a virtual environment, bear in mind the importance of the physical hardware:



The structure of a virtualized environment

Virtual devices

Regardless of whether a network is physical or virtual, it requires the functionality of interconnecting network devices, such as switches, routers, and network interfaces. A virtual network differs from a physical network because many, if not all, of its connectivity devices are themselves virtualized versions of physical devices with the same functionality.

In a virtual network, the essential virtual devices are virtual servers, virtual machines, virtual network interfaces, virtual switches, and virtual routers. Any physical component required in a physical network has a virtual equivalent use in a virtual environment.

Virtual servers

Networks grow, adding nodes, applications, and scope. In the past, managing network growth meant adding one or more physical servers to handle the additional demand on existing servers. Adding a physical server can be costly in terms of hardware acquisition, configuration, and downtime. On the other hand, adding a virtual server requires only the time to configure it, resulting in a much lower implementation cost and a lot less service interruption.

Server virtualization is the partitioning of a physical server into two or more virtual servers. Server virtualization masks the physical hardware and users see only the resources that have been allocated to the virtual machine in use. There are three primary types of server virtualization:

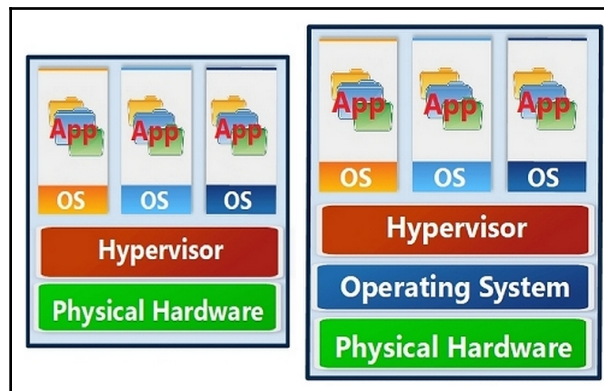
- **Full virtualization:** This form of server virtualization replaces the physical server's operating system with a specialized form of operations, either through a supervisor or a hypervisor. The hypervisor acts as a go-between for the virtual servers and virtual machines, which are unaware of each other, running in its environment and the physical computer and its resources. Examples of full virtualization systems include Adeos, Mac-on-Linux, Parallel Desktop for Mac, VMware ESXi, VirtualBox, Win4BSD, and **Quick Emulator (QEMU)** or Microsoft Hyper-V.
- **Para-virtualization:** In this form of server virtualization, a hypervisor or **virtual machine manager (VMM)** supports virtual servers that have been modified to run in this environment. This virtualization approach allows two or more different operating systems to share a physical computer and its resources. A primary example of this form of virtualization is the Xen Project Hypervisor.

- **Operating system level virtualization:** This form of server virtualization creates an environment in which several processing spaces, known as containers, zones, virtualization engines, or jails, which are independent processing spaces, are all running on the same operating system. In place of a hypervisor, the native operating system provides support for the virtualized environment. Examples of this type of server virtualization include Linux V-Server, FreeBSD Jail, AIX Workload Partitions, and Solaris Containers.

Hypervisors

A key part of virtualization is the hypervisor, or VMM, which is low-level software that facilitates multiple virtual machines running on a single physical computer. There are two types of hypervisors, plus a hybrid that blends the two. The two types of hypervisors are as follows:

- **Type I:** These are the bare-metal, embedded, or native hypervisors. This type of hypervisor installs itself directly onto the physical hardware or on the *bare metal* of the computer, as illustrated in the following diagram.
- **Type II:** These are the hosted hypervisors. As shown in the following diagram, a Type II hypervisor runs on the host computer's operating system:



A Type I hypervisor configuration (left) and a Type II hypervisor configuration (right)

As illustrated in the preceding diagram, the primary difference between a Type I and a Type II hypervisor is that one runs without an underlying host operating system (Type I) while the other does (Type II).

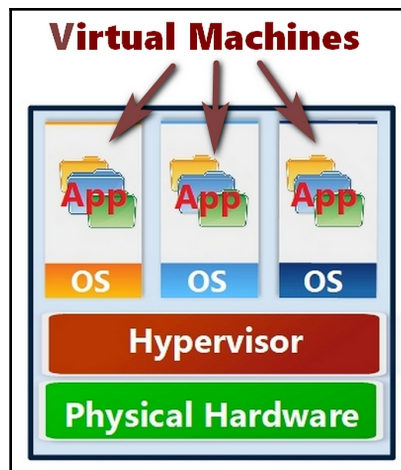
Although not technically a separate category of hypervisors, some companies are finding that fitting different hypervisors to different workloads creates a more efficient computing solution for them. Third-party software is available to manage multi-hypervisor environments, creating a hybrid hypervisor deployment.

Hosts and guests

Three virtualization terms that are often confused or used interchangeably are virtual machine, host, and guest. A virtual machine and a guest are terms for the same thing. Each refers to a software-created and managed workspace on the host computer that emulates a single virtual computer. The term *host* can refer to the host hardware, or the computer on which the virtual system is running, or to the host OS in a Type II environment. The host system is typically able to support multiple guest systems.

Virtual machine (VM)

A VM is a complete computer system, including its devices, operating system, application software, and so on, that are created through software to run on a physical computer, possibly along with other VMs. A VM is able to run programs or applications in the same way they would on a physical computer that's not been configured virtually:



Virtual machines on a host system

A **virtual machine management interface (VMMI)**, that is, a VM control panel, allows an administrator to monitor, create, delete, pause, start, and stop virtual machines in a virtualized environment. Some of these tools also provide the capability to reconfigure the allocation of the host system's physical resources for one or more virtual machines. Each of the major virtualized system developers (Citrix, Microsoft, and VMWare) offer VMMI software as part of their server virtualization packages. There are also third-party and open source VMMI packages that are compatible with many of the virtualization systems, including some that are based on the **Simple Network Management Protocol (SNMP)**.

Hardware configuration for a virtual environment

The hardware configuration required to support a virtual server, virtual machine, or other virtual devices is a function of each network environment, which is, in turn, a function of each individual organization. The major virtualization software providers have websites that enable users to determine the best hardware configuration for their needs, hardware, and plans. However, there are general guidelines, although not actually standards, for configuring a computer to support a virtual environment. The primary components for consideration are the CPU, memory, the BIOS/UEFI, and the physical network capacity.

Like just about everything else in computing, what you need in terms of computing power and memory really depends on what you plan to do. Your choices may depend on the answers to the questions that define the needs of your organization and its near-term requirements. Some of the considerations to address include the following:

- **CPU:** Along with primary memory, one or more processors sufficient in size to support the size and volume of the virtualized system are required:
 - If multiple physical computers are in the virtualized network, you should use the same CPU platform (Intel versus AMD) in every case. Virtual machines can move from one physical computer to another easily, but only if the computers have the same CPU platform.
 - Choose only those processors that are optimized for hardware virtualization support and include Intel-VT or AMD-V.
 - Multi-core processors can provide additional processing capabilities, especially in larger virtualizations.
- **Memory:** The number of virtual machines possible on a host computer is a function of the amount of memory available on the host. Memory is the most limiting resource in a virtualized environment.

- **BIOS/UEFI:** On AMD processors with AMD-V, this feature is automatically enabled. However, on Intel processors, the Intel-VT feature has an initial state disabled and must be enabled using either BIOS or UEFI.
- **Physical network:** Regardless of the structure of a virtualized of **software-defined network (SDN)** that overlays the physical network hardware, the physical network carries the network traffic. The physical network requires sufficient capacity to support all of the traffic generated by the virtual servers and machines, as if it were all directly part of the physical network.

Virtual resource allocation

In a virtualized environment, resources, such as the CPU, memory, and data storage, are allocated based on the settings of the shares, reservations, and limits of each resource for each VM. A virtual data center manages the resource pool and provides each VM with its initial resource allotment and on-demand resources, as needed, until the allocations reach an upper limit. The resource pool consists of the physical resources of the host computer minus the resource needs of the operating system, if any, and the hypervisor.

When the physical resources of the host no longer satisfy the resource demands of its virtual machines, the resources may require allocation or limits. The actions an administrator can take to assign resources to one or all VMs include the following:

- Create a reservation of one or more physical resources of the host computer or server cluster to one or more VMs
- Allocate a fixed share of a physical resource to a VM
- Assign a priority ranking to a VM that assures it a larger share of a physical resource
- Set upper limits on the allocation of resources to a VM

The primary allocation methods that are used to balance the assignment and use of the physical system resources to virtual machines are as follows:

- **Resource allocation shares:** A share indicates the priority or importance of a virtual machine with respect to a particular resource. Shares are in three levels—**High**, **Normal**, or **Low**, which translates to 4 shares for High, 2 shares for Normal, and 1 share for Low. An administrator may also set a custom level and set the number of shares to a specific number that set a proportional share ratio with the shares of other VMs.

- **Resource allocation reservation:** A resource reservation sets a minimum amount (in MBs or MHz) of a physical resource allocated to a VM. If the amount of a reserved resource is not available, the VM will start with a lower share, until such a time as the rest of the reserved resources become available.
- **Resource allocation limit:** An allocation limit sets an upper limit on the amount of a physical resource that may be allocated to a virtual machine. A virtual server can allocate more than one resource allocation reservation amount, but with an allocation limit specified, it cannot allocate the resource above the limit (expressed as MBs, MHz, or IOPS). By default, the limits for CPU, memory, and data storage are unlimited.

Network connectivity

Network connectivity and bandwidth are host resources that may be allocated to the VMs running on it. The amount of allocation can be preset, just like the preceding resources, or it can be allocated to fit the communication need of a particular transmission.

A network connection on a VM can have one of three configurations:

- **Direct access (bridged):** A VM, which has an assigned IP address, has direct access to an external network. The VM is able to communicate with the physical network through the host computer's network adapter and directly interact with network nodes.
- **Network Address Translation (NAT):** In this configuration, the VM communicates to the network using the IP address of the host computer. This option is best if the VM doesn't have an IP address of its own.
- **Host-only:** This configuration creates a connection between the host computer's network adapter and that of the VM. In other words, the virtual network adapter is visible to the host system. In this configuration, the VM is able to communicate only with the host and any other VMs running on the host.

Virtual internetworking devices

A virtual network uses essentially the same interconnection devices as a physical network—a **network interface controller (NIC)**, a network switch, and a network router. On a virtual network, these devices (and a few others) are software-defined and virtual. These are **virtual NICs (vNICs)**, **virtual switches (vSwitches)**, and **virtual routers (vRouters)**. The VMs connect to a virtual network through their vNICs, which logically connected to the **physical NICs (pNICs)** of the host.

On many virtual networks, vNICs connect to a vSwitch, which, in turn, connect to a vRouter. The functions of these virtual connectivity devices mirror that of their physical counterparts. However, like all virtual devices, at some point, they interconnect with the physical network.

Summary

There are three types of virtualization that are used in networks, each with its own specific purpose and operation—VPNs, VLANs, and virtual networks. The parts of a virtualized network environment are the physical hardware, a virtual network server, and virtual machines. In a virtual network, the devices are virtual servers, VMs, virtual network interfaces, virtual switches, and virtual routers. Server virtualization partitions a physical server into virtual servers. Three types of server virtualization are full virtualization, para-virtualization, and operating system level virtualization.

A hypervisor is a piece of software that facilitates multiple virtual machines running on a single physical computer. There are two types of hypervisors—Type I, or bare-metal, and Type II, or a hosted hypervisor. The difference between Type I and Type II hypervisors is that a Type I runs without an underlying host operating system and a Type II hypervisor runs on top of a host operating system.

A VM is a software-created and managed workspace on a host computer that emulates a single virtual computer. A host may refer to the physical hardware of a computer or a host OS in a Type II environment. A host system typically supports multiple guest (virtual) systems. A VM is a complete computer system that runs on a host computer, typically with other VMs.

The hardware configuration required for the configuration of a virtual environment is the CPU, with hardware virtualization, memory, the BIOS/UEFI, and the physical network. Resources are allocated through shares, reservations, and limits. Network connectivity and bandwidth are host resources that may be allocated as direct-access, NAT, and host-only.

Questions

1. Which of the following is not a virtualization technology that's used in computer networks?
 1. Virtual private network
 2. Virtual reality
 3. Virtual local area network
 4. Virtual server
2. Which of the following is not a type of server virtualization?
 1. Para-virtualization
 2. Full virtualization
 3. Quasi-virtualization
 4. Operating system level virtualization
3. The software that facilitates the creation and system services for a virtual machine is which of the following?
 1. Kernel
 2. Device driver
 3. Emulator
 4. Hypervisor
4. A bare-metal virtual machine manager is which of the following?
 1. Type I
 2. Type II
 3. Hybrid
 4. UEFI
5. Which type of virtualization runs on a host operating system?
 1. Type I
 2. Type II
 3. Hybrid
 4. UEFI

-
6. The term used for physical hardware, an operating system on a hosted system, or both is which of the following?
 1. Virtual
 2. Guest
 3. Host
 4. Ghost
 7. Which of the following is not a method for allocating physical or virtual hardware resources to VMs?
 1. Shares
 2. Assigned
 3. Limits
 4. Reservations
 8. Direct-access, NAT, and host-only are methods for allocating what resources? (Choose all that apply)
 1. Network connectivity
 2. Duplexity
 3. Bandwidth
 4. Disk storage
 9. Technologies such as Intel-VT and AMD-V optimize a processor for what adaptation?
 1. Remote communications
 2. Clustering
 3. Hardware virtualization
 4. Software virtualization
 10. What hardware resource influences the number of VMs possible on a host computer more than any other resource?
 1. Processor
 2. Hard disk drive
 3. Memory
 4. NIC

10

Disaster Recovery

For all that goes into the processes surrounding disaster recovery, it's hoped that it will never be necessary. The complexity of a disaster recovery program rises and falls with the size, locale, diversity, and application of the systems involved. Smaller system environments can protect themselves against a disaster merely by regularly taking a backup of their systems. Larger data center environments that support distributed systems, large networks, virtualization, and perhaps a cloud service, have a much more complicated task in identifying the required elements of continuity and their sequence for recovery—not to mention, exactly how soon and where the recovery is to happen. A medium-sized operation falls somewhere in between these two extremes.

In this chapter, we will review the definitions, methods, products, and applications involved in disaster recovery and business continuity planning and execution. We will cover the following topics:

- Continuity of operations
- Recovery
- Backups and replication

Business continuity plan (BCP)

Catastrophic events rarely provide a warning in advance. This is why any organization should have a formal, written plan to continue its operations during or after a catastrophic event. Having such a plan can help an organization remain viable and stay in business. A BCP outlines the objectives, procedures, and step-by-step actions required to restart or continue an organization's as-normal-as-possible operations after a disruptive event. A BCP is different from a **disaster recovery plan (DRP)**. A DRP focuses on the restoration of the computing infrastructure and its associated services. On the other hand, a BCP, which encompasses the DRP, takes a broader look at the business and the restoration or continuity of the functions of the entire organization, including its operational processes, equipment, staffing, and inventory.

The following diagram illustrates the components of a BCP:



Business impact analysis feeds into a DRP and both become a part of the BCP

Several examples, templates, and best practices are available on the web for preparing a BCP, but there isn't a universally accepted standard format with a list of required contents. However, in any form, the BCP focuses on getting the business up and running again. In most cases, the first phase of the development of a BCP is the performance of a **business impact analysis (BIA)**.

BIA

BIA takes a close look at an organization and projects the potential financial impact of any interruption in its operations from an extreme event. A BIA should identify how disruptive and destructive events could impact an organization. A BIA identifies what is likely to be affected by different events and severity levels. The following table shows one example of a BIA (with fictitious data, of course):

Event	Affected assets	Operational loss	Financial loss	Time to recover
Fire in server farm/water damage	All assets in data center	Loss of computing processes	\$1,000 loss per day	12 – 24 hours
Class 3+ hurricane	Building structure, window glass, roofing, signage	All business operations	\$10,000 loss per day	1 – 2 weeks
Power outage	Computing, electrical elements, lighting, HVAC	All business operations	\$800 per hour	0 – 12 hours

In general, a BIA project involves four phases:

- **Gathering:** This phase collects information from all sources concerning vulnerabilities, threats, loss, and recovery
- **Evaluating:** This phase evaluates the collected information for priority, degree of loss, and importance to business operations
- **Documenting:** A summary report of the information collected, the analysis process, and the conclusions of the impact analysis
- **Presenting:** The report is presented to senior management and stakeholders

Risk assessment

Another step in the development of a business continuity plan is risk assessment, which is generally a more detailed version of a BIA. With the approval of senior management and the stakeholders of the affected areas included in the BIA report and the events and results included, a risk assessment is the next step in this process. In smaller organizations, the BIA may include the results that a risk assessment would develop. However, in larger organizations, the events and threats of the BIA are expanded to identify what may be the specific assets or services that are lost in each case.

As its name implies, the purpose of risk assessment is to project which assets (equipment, people, tools, inventory, and so on) are at risk of loss or damage in an extreme event. Using the results of the risk assessment, a mitigation plan may be developed to remove or reduce the exposure of any identified vulnerabilities.

Continuity of operations

The immediate focus of virtually all organizations that suffer some form of catastrophic damage that causes an interruption to its operations is to restore its infrastructure to the point that it can resume providing its products or services. There is no standard or *cookie-cutter* way for any organization to follow in order to restore its operations. Each organization is relatively unique in its structure, operational procedures, and the services required for a desired level of functionality.

Effective and efficient recovery from a disaster, whether it be natural or human-caused, requires pre-planning. In the aftermath of a destructive catastrophe, there will be very little in the way of *clear thinking* or *rational actions*. Having a plan to follow, even one that's slightly out of date, is certainly better than no plan at all. In the following sections, we will discuss the plans that any business (yes, even those with only one person who performs all its duties) should consider developing.

DRP

While some forms of catastrophe that could cause destruction to computing resources may or may not happen, others might. If an organization chooses to disregard the potential damage that could occur, they do so at their own peril, to turn a phrase. So, *what kind of damage could be so bad that an organization may lose its capability to continue its business operations?* During a heavy rainstorm, a roof leaks a steady drip directly on a business' only computer, which is powered up and unattended. Eventually, its power supply, motherboard, and internal components short out, effectively destroying the business' data and the capability to recover it.

Alternatively, hurricane his/her-name completely wipes out the data center for a large metropolitan hospital. The data and systems are available on backups, both physical and in the cloud, but the recovery requires computer hardware to provide continuity of the hospital's operations. The availability and location of a computer system that is capable of adequately supporting the hospital's systems at a moment's notice will definitely solve the hospital's immediate problem, provided such a system exists nearby.

Recovery plans

There's more to what goes into a fully formed disaster recovery plan than just nailing down recovery hardware, loading data, and getting back online. A recovery plan should contain, as a minimum, the following elements:

- **Scope:** Identifies the events, causes, or situations covered
- **Evaluation:** The process to use for determining the extent of the damage and the appropriate actions for recovery
- **Staging:** The materials, equipment, staffing, and contingencies for the recovery actions appropriate to the damage
- **Restoration:** The detailed steps to restore, re-install, and recover systems, applications, and data
- **Re-evaluate:** A plan for the review and evaluation of the recovery and identification of any adjustments and improvements to the plan

The first and primary goal in disaster recovery planning of any organization is to have a plan. The plan should be specific, but it should also be realistic and address only the disaster events that are possible in that particular location. For example, a plan for a company in Kansas probably doesn't need to include volcanic eruptions in its disaster recovery plan, while an organization in Hawaii probably doesn't need to cover a blizzard.

A DRP may actually be several plans that share some or all of its stages. For example, a section of a disaster recovery plan on hurricanes will probably share some sections with heavy rainstorms or high winds. Whatever weather, war, or worse events the plan includes, each must be specific regarding the processes that are used to recover from it.

Recovery sites

One of the first orders of business in disaster recovery is choosing a recovery site. For some organizations, disaster recovery may require more than one site, depending on the severity of the damage or priorities of the recovery plan. While there are some overlaps, there are essentially three levels of recovery sites, each characterized by its requirements and responsiveness. Recovery sites are either hot, warm, or cold sites:

- **Hot site:** This type of recovery site is essentially a copy of a production system and its environment, in some cases right down to the office furniture. The purpose of a hot site is to provide a failover safety net for a destroyed or incapacitated computing environment. The systems of a hot site operate concurrently with the production systems so that minimal downtime happens in the switch-over. The location of a hot site must consider the threats that may make it necessary. Disaster events may require the hot site to be in a geographically distant location.
- **Warm site:** This type of recovery site contains the necessary equipment and environments to support the essential components of a production system. However, the installation of up-to-date data and systems must occur before processing can begin. A warm site is appropriate when the loss of computing operations is not an emergency, such as administrative systems.
- **Cold site:** This type of recovery site is essentially an office space with the necessary environmental and power systems to support the restoration of computing services. Obviously, a cold site supports recovery actions that aren't emergencies on any level. In many cases, a cold site becomes a temporary operations center until the restoration or relocation of the main facilities.

The strategy of the type of recovery site an organization includes in its disaster recovery plan depends on its business continuity needs, which are dependent on the mission, objectives, products, and services of the organization.

Replication and backup

Backup and replication are two data integrity and recovery methods that are frequently used interchangeably. However, they are actually quite different. A backup copy of data is just that, a copy. A backup is written to removable media and stored in an offsite location, physically or in the cloud. A replication is also a data copy process. However, replicated data has been copied for the purpose of storing copies of the data on distributed locations within an organization or on a subscribed cloud service.

A backup is generally created on a periodic basis, such as daily, weekly, or over longer periods. Replication is typically done online, in real-time or near-real-time. Restoring a backup can take some time if it must be retrieved from a remote storage facility, downloaded from the cloud, and then restored to a system. Restoring replicated data is immediately available, and the time involved amounts to not much more than its transmission time.

As of a disaster recovery plan, or as the core of a data integrity strategy, a data replication procedure ensures that the loss of one backup or copy of an organization's electronic data doesn't mean all is lost.

Data replication

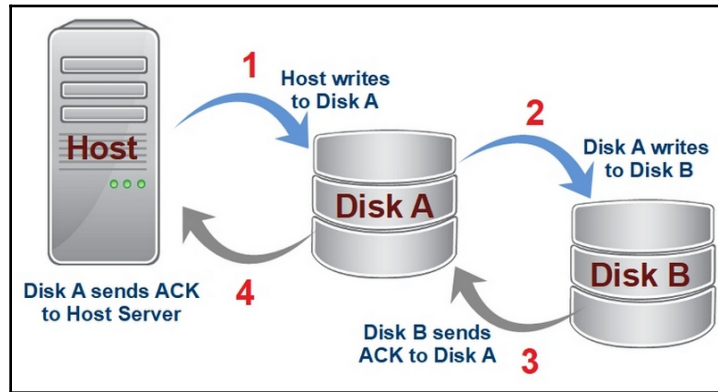
Data replication has several definitions and objectives, depending on how it's applied and its purpose. In one usage, data replication refers to a database being continuously updated so that two or more copies of the database are available in separate locations for processing. The purpose of data replication is to provide for coherent processing at distributed locations or to keep the data on the current hot site.

Synchronous and asynchronous

The two primary types of data replication are **synchronous** and **asynchronous**. Synchronous replication simultaneously writes to two storage devices, which guarantees a real-time distributed data source or a hot failover backup. Typically, the synchronization is a one-way process, as shown in the following diagram. This illustration shows the following steps in this process:

1. The host server writes to a primary storage (**hard disk drive (HDD)**, **network-attached storage (NAS)**, **storage area network (SAN)**).
2. The primary storage writes to the secondary storage.

3. The secondary storage acknowledges the receipt of the record to the primary storage device.
4. The primary storage sends an acknowledgement back to the host server. The storage devices in a synchronous replication system can be separated over a distance, but the distance between the storage devices may be limited by the latency that the distance introduces:

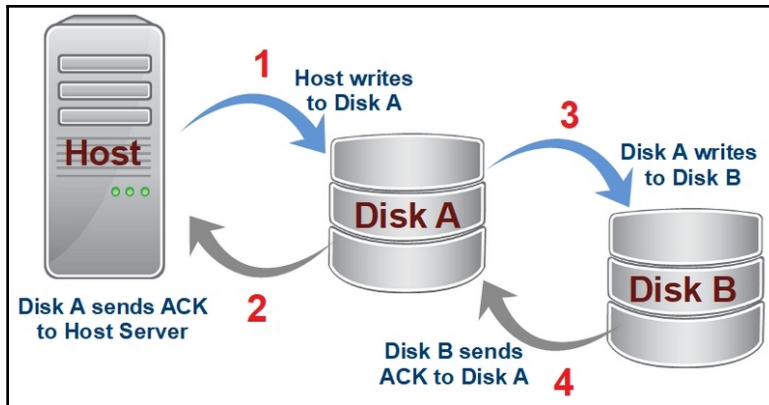


The process steps of a synchronous replication system

Asynchronous replication is a more commonly used approach for replication over a distance and is a primary part of a disaster recovery strategy. Where latency is an issue in the synchronization of the two data storage devices, asynchronous replication ignores the latency by buffering disk storage data actions. Typically, the secondary storage operates a few actions behind the primary storage. Operationally, the difference between synchronous and asynchronous replications is the sequence of the replication events.

The following diagram illustrates the process steps of asynchronous replication. This illustration shows the following steps in this process:

1. The host server writes to the primary storage device
2. The primary storage device acknowledges the action
3. The primary storage unit writes the record to the secondary storage device
4. The secondary storage device then acknowledges the action:



The process steps of an asynchronous replication system

An emerging replication method is near-synchronous replication. Near-synchronous replication only copies data changes to a secondary storage unit, typically with a latency of only a few seconds. Because near-synchronous replication writes each data change directly to the redundant site, the primary system can continue its processing without waiting for an acknowledgement from the secondary storage unit.

Replication methods

Several source-to-target variations exist for data replication, but the primary methods are as follows:

- **Disk-to-disk:** This replication method copies data from one data storage device to another. In some ways, this form of replication can essentially be a backup strategy, but data replication between two storage devices is a real-time or near-real-time process.
- **Server-based:** This form of replication is generally an element of a high availability or disaster recovery strategy. Server-based replication systems can be any of the following:
 - **Server-to-self:** One disk volume replicates to another volume on the same server.
 - **Cluster-to-cluster:** One disk cluster replicates to another disk cluster in a failover setup.
 - **Server-to-server:** This form of replication operates as synchronous or asynchronous, as discussed previously. This form of replication is also referred to as site-to-site.

Data backup

As we discussed earlier, a data backup is a copy of some or all of the data on a secondary storage that is written to removable media and securely stored in an offsite physical or logical location. However, in practice, there are several methodologies to follow, each of which can be a part of the backup strategy of virtually any organization.

Archive bit

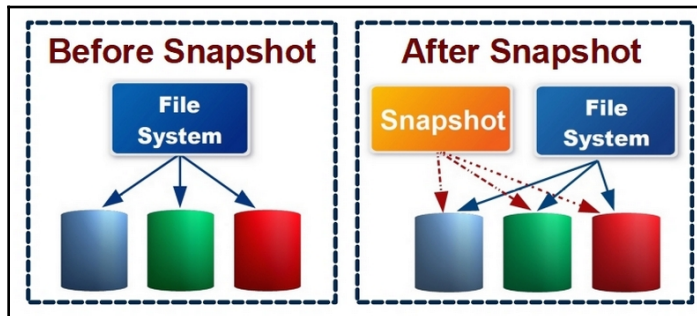
Backup software and utilities, for the most part, use the archive bit of a file to determine whether the file is to be written to the backup medium. While an archive bit set to *on* indicates that a file should be archived, not all backup methods set the archive bit to *off* after copying a file to the backup medium. The following section includes information on how each backup method deals with the archive bit.

Backup methods

Data backups may be created in a variety of scopes and frequencies. The backup methods you will encounter on the Server+ exam are as follows:

- **Full backup (normal backup):** This backup method copies all of the content stored on a secondary storage device and writes it—in compressed form, typically, to a removable medium for storage and archival purposes. Full backups are automatically taken on preset intervals, usually weekly, monthly, or at the end of a financial reporting period. Full backups turn off the archive bit.
- **Copy backup:** This backup method is used to create an archival copy of all or part of the data that's stored on secondary storage. A copy backup is just a copy of the data and does not affect the archive bit in any way.
- **Incremental backup:** This backup method copies only those files that have been modified or created since the last full or incremental backup. This means that only the files flagged with an *on* archive bit are captured. Typically, incremental backups are taken daily between full backups. However, to restore a system, you require the last full backup and each of the incremental backups since the last full backup must be reloaded. An incremental backup resets the archive bits to *off*.
- **Differential backup:** This backup method copies the files that have been modified or created since the last full backup. Each of the differential backups between full backups captures an accumulation of the data changes flagged by an *on* archive bit. However, a differential backup doesn't reset the archive bit, leaving that to the next full backup. To recover a system, only the last full backup and the last differential backup need be restored.

- **Snapshot:** A snapshot (an image) is a capture of the state of a system at one specific point in time. On a server, disk array, or a virtual machine, a storage snapshot is a variant backup that contains the information required to restore data back to the recovery point (the time at which the snapshot was taken). A snapshot file, like the one in the **After Snapshot** image in the following diagram, doesn't contain a copy of the data. Rather, it contains the locations of the data and its organization:



These two images illustrate the before and after structure of a data filesystem snapshot

- **Selective backup (partial backup):** This backup method requires the administrator to pre-select, list, or mark the files, folders, and so on that are to be captured in the backup. Only the files and folders listed or marked are captured to the backup medium.
- **Bare metal backup:** Although we have discussed full backups and other methods for storage backups, this method of backup literally captures everything on a system, including the OS, stored application, and system software and data. The purpose for and the source of its name, a bare-metal backup is used to load the entire system to a bare metal computer.
- **Open file backup:** When a system backup is taken, it's ideal to have no running jobs (other than the backup system) on the computer. While this isn't always possible, there should be no open files on the system. In some cases, neither is possible. There is a variety of add-in software products that include the capability to capture open files to the backup medium. Nearly all of the software products for Windows require the **Volume Shadow Copy Service (VSS)** to be enabled. Several backup software packages for Linux will also capture open files.

Data versus OS restore

Should it ever be necessary to restore some or all of a system from a backup medium, there are actually two types of restore processes—data restore, and OS or system restore. As their names imply, a data restore reloads all or some of the data on a backup medium and an OS restore reloads or resets the operating system and its settings back to their original states.

You should use a data restore procedure when some or all of the data on a secondary storage device is corrupted, erased, or compromised to reload specific files, an entire volume, or the entire data contents. An OS restore can restore a feature that's been removed erroneously or reset the system software back to a certain release.

Backup media

Just about any magnetic medium may be used as a backup storage medium, but for the most part, the three general medium categories are as follows:

- **Linear access:** There is really only one type of linear access storage medium – magnetic tape. Linear access refers to the fact that accessing a magnetic tape is always beginning to end, front to back, serial access. Magnetic tape or **linear tape-open (LTO)** is perhaps the most commonly used medium for storing backup data. The current standard, LTO-6, provides 2.5 TB data storage, but with a higher data compression, an LTO-6 tape cartridge can hold as much as 6.25 TB. Tape is the less costly of the standard backup media and is also the easiest to store.
- **Random access:** In many cases, archiving data and backing up data are two different functions, although both create copies of data and files. Tape backup is better suited for archiving data, which requires longer term storage and is generally restored on the whole. Magnetic disks or **solid-state drive (SSD)**, which are typically external peripherals, are much better for backing up data that may need to be restored in the near future, especially if only some of the backup's contents are to be accessed.
- **Removable media:** Unless an NAS or SAN is set aside for storing replicated or backed-up data, the medium that's used for storing the data is a form of removable media. Magnetic tape, external hard disk drives, external solid stage drives, optical media (CD-ROM, DVD, Blue-ray, **write once, read many (WORM)**, and more), and USB-connected flash memory devices are all forms of removable media that can hold archived, backed up, or replicated data.

Media storage

Regardless of the type of storage media used to back up a system on any backup type, there are two calculations that an organization must consider in choosing a storage location for the backup media:

- **Recovery time objective (RTO):** RTO indicates the desired length of time it takes to recover a failed system. If the backup or replicated data is to be restored in the event of a catastrophic system failure, regardless of the cause, the off-site location of the archive media is a major consideration. If the site is located far from the recovery site, the RTO will need to be longer. However, if the storage site is close to the recovery site, a shorter RTO is possible.
- **Recovery point objective (RPO):** RPO is the target point-in-time to be reestablished through data recovery for a failed system, for example, if a disaster recovery calls for the RPO to be no more than two days before the event that caused the need for a recovery.

With RTO and RPO guiding the choice of a storage location for the backup or replication media, the other issues, perhaps just as important, are as follows:

- **Access:** The storage location, regardless of being physical or logical, must be accessible when it is necessary to restore the system. If the location is only open during the day, it could pose a problem if the system crashes during the night. The same considerations apply to a cloud storage location, but who has access, how access is effected, and what is accessible, should be consistent with the needs of the DRP.
- **Security:** How well-secured the storage location is should be a major concern when choosing a site. A physical storage location must have the security devices and procedures that are in line with the criticality of the data being stored. A cloud storage service should provide information on its physical and logical security, which includes their procedures, should a security event occur.
- **Environmental:** A storage site may be accessible and secure, but if it is dusty, extremely hot or cold, and physically or structurally unsafe or dangerous, it's most likely not a good choice for storing your backups. The site (physical or logical) should have adequate air conditioning, air filtration, and monitoring to prevent damage to the storage media and its magnetic contents.

Backup media integrity

A backup medium unit needs two actions to verify its integrity and restorability. The first element is identity. The label on a backup storage unit should include the following:

- The data stored on the medium, listed as specifically as needed, to locate particular data for restoration. The unit label must be more than *daily backup*.
- The day, date, and time the data was written to the medium. If multiple units are involved, a sequence number of the set should be included using a *Unit n of x* units format.
- The identity of the person who created the backup or replication.
- Any other information pertinent to identifying the contents, security, or retention, as per the organization's standards.

The second verification and integrity procedures should include procedures to test the integrity of the data on the backup and its restorability. Many data backup or replication software packages include the capability to calculate and store a **cyclical redundancy check (CRC)** on the tape that can be verified at a later date to test the integrity of the data on the medium. In addition, on a rotating basis, test restores should be performed on a non-production system or the recovery site system.

Backup media retention

There are several backup data storage media retention and rotation strategies, most of which are based on the confidentiality or sensitivity of the data and its RTO/RPO metrics. However, the **3-2-1 backup strategy** is perhaps the basis for most backup plans. A 3-2-1 backup strategy requires that three copies of backup or replication data be created. Two of the copies, which are written to different mediums (such as one on tape and one in the cloud), are retained locally and one or more copies are kept offsite.

Many backup strategies also include a rotation schedule that recycles the backup medium after a certain period of time. For example, the medium of a backup created on Monday may be used again in three days, or perhaps after two weeks. A rotating medium helps to reduce cost and storage requirements.

Summary

An organization should have a formal, written plan that details the actions required to restart or continue operations after a disruptive event. A BCP details the restoration or continuity of the organization's processes, equipment, staffing, inventory, and so on. There's more to what goes into a fully formed DRP than just nailing down recovery hardware, loading data, and getting back online. A DRP focuses on the restoration of the computing infrastructure and its associated services and should include its scope statement, evaluation of the damage, a staging plan for each damage level, a restoration procedure, and a review of the plan's effectiveness. An important part of a DRP is the designation of a recovery site, which are classified as hot, warm, and cold sites.

A BIA projects the potential impact of an operational interruption from an extreme event by identifying the effect of different events and severity levels. A BIA project has four phases: gathering, evaluating, documenting, and presenting. A risk assessment identifies the events and threats, as well as the specific assets or services that would be lost in each case.

A backup is written to removable media and stored in an offsite location, physically or in the cloud. A replication stores a copy of data on a distributed location separately from the system. A backup is created on a periodic basis. Replication is an online, real-time duplicate of a primary data store. Data replication is synchronous or asynchronous. Synchronous simultaneously writes to two storage devices. Asynchronous is an approach that's used for replication over a distance. Latency is an issue regarding synchronous replication, but not with asynchronous replication. Some replication methods include disk-to-disk, server-based, server-to-self, cluster-to-cluster, and server-to-server, while some data backup methods include full, copy, incremental, differential, and snapshot. The general backup mediums are linear access, random access, and removable media.

The calculations that can determine the location of a data storage medium are RTO and RPO. The backup medium needs to verify its integrity and restorability. Its identification label should include the data on the medium, the date and time of the backup, and who created the backup. Most backup retention strategies are based on the 3-2-1 strategy.

Questions

1. Which of the following details the actions to be performed to restore or continue operations following a disruptive event?
 1. BIA
 2. DRP
 3. BCP
 4. RPO
2. What plan details the restoration of the computing infrastructure following a catastrophic event?
 1. BIA
 2. DRP
 3. BCP
 4. RPO
3. Which of the following is not a type of disaster recovery site?
 1. Cold
 2. Warm
 3. Hot
 4. Real-time
4. A BIA identifies the potential impact of the discontinuance of services in the event of a catastrophic event. What study identifies the specific threats and the impact an event would have on specific assets?
 1. Business impact assessment
 2. Disaster recovery plan
 3. Risk assessment
 4. Recovery time objective
5. Between a data backup and a data replication, which of these options stores a duplication of data that is real-time or near-real-time?
 1. Data backup
 2. Data replication
 3. Both
 4. Neither

6. Which replication method can be impacted by latency?
 1. Synchronous
 2. Asynchronous
 3. Near-synchronous
 4. Near-asynchronous
7. Which of the following are replication methodologies? (Choose all that apply)
 1. Disk-to-disk
 2. Cluster-to-cluster
 3. Server-to-server
 4. All of the above
 5. None of the above
8. Which data backup method resets an archive bit that was turned on after the last full back up?
 1. Full
 2. Incremental
 3. Snapshot
 4. Differential
 5. Copy
9. Which of the following is not a category of data backup media?
 1. Linear access
 2. Random access
 3. Layered access
 4. Removable media
10. What is the metric that is the desired completion time of a disaster recovery plan?
 1. 3-2-1
 2. RTO
 3. RPO
 4. R2D2

3

Section 3: Security

This part of the book looks at the basic security technologies, methods, and procedures that can be applied to secure a server and its network.

The following chapters are included in this section:

- Chapter 11, Security Systems and Protocols
- Chapter 12, Physical Security and Environmental Controls
- Chapter 13, Logical Security

11

Security Systems and Protocols

This part of this book covers the physical and logical security concepts, procedures, and devices you may encounter on the Server+ certification exam. The security and safety that's necessary to protect a network server and its immediate and remote connections involves systems that are dedicated to security; security protocols, security devices, and practices to physically protect data and equipment; and, of course, the various data protection technologies, such as encryption, which are used to secure data in use, in storage, and in transit.

There is certainly no part or configuration of a server that's more important than its security. Any security settings, configuration, or procedures must protect against external and internal intrusion, data corruption, and damage that's caused by malware. The objective of any computing network's security program is the CIA triad, which consists of confidentiality, integrity, and availability.

In this chapter, we'll look at the devices, software, methods, and protocols that can be combined to create a security system to protect a network server, its data, and its major systems. The following topics will be covered:

- Security zones
- Authentication protocols
- Port security

Security zones

Gateway security devices, such as firewalls, define security zones to apply security policies to incoming and outgoing network traffic. To put this in a different way, a **security zone** is a logical structure that's created from one or more device ports/interfaces that apply the same security policies. A security zone can be just one interface, or it can include several interfaces if the interfaces apply the same policies. Each interface may also be a security zone, and a security zone may also include two or more interfaces. However, an interface can belong to only one security zone.

Firewall zones

Many firewalls (hardware or software) predefine a set of security zones to facilitate initial configuration. Most installations require additional security zones, but for some the presets, just one security zone may be enough. The most common predefined security zones include the following:

Security zone	Description
LAN	Highest level of trust; includes VLANs
VPN/SSLVPN	Highest external level of trust; encrypted VPN traffic
Multicast	IP multicasting
DMZ/public	Publicly accessible servers
WAN	Connections to WANs; effects incoming and outgoing traffic
Untrusted	Lowest level of trust

The security zones that are listed in the previous table can be either private or public security zones. The first two zones (LAN and VPN) are private security zones and the remaining zones are public zones.

The configuration of a security zone applies two primary policies:

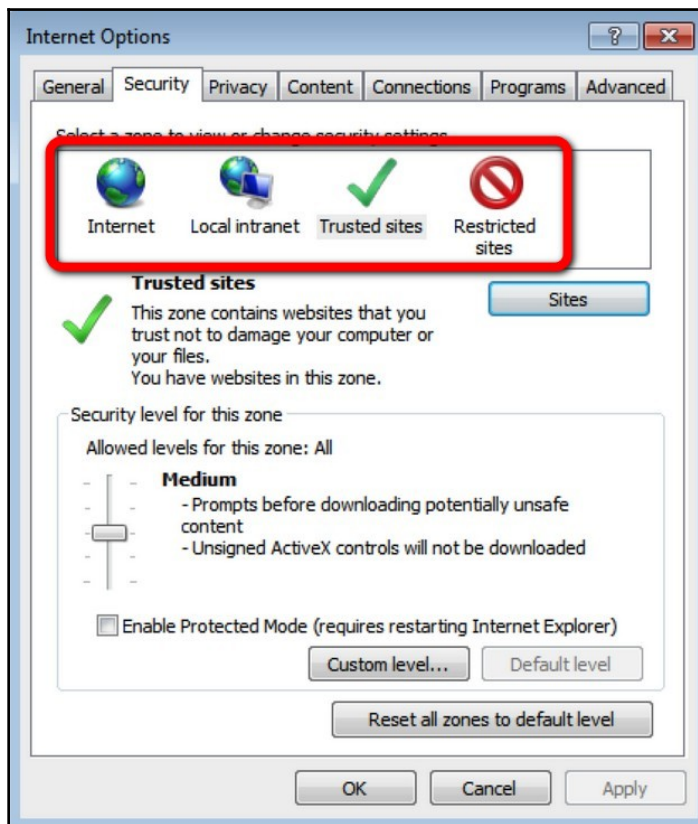
- **Security policies:** These are rules that control the permit or deny inbound and outbound traffic based on the source and destination addresses, protocols, port numbers, and content
- **Access control policies:** These define who or what has permission to gain access to the resources behind a firewall

Demilitarized zone (DMZ)

A DMZ or perimeter network can be either a physical or logical network segment or subnet that's used as a default landing space for external WAN traffic that's seeking organizational information from the organization's website, which is open to the untrusted WAN security zone.

Browser zones

On a local network, as well as on individual workstations, security policies can be set through a web browser. For example, the Microsoft **Internet Explorer (IE)** browser provides settings to restrict access to one or more predefined security zones, as shown in the following screenshot:



The security zones defined in IE

As shown in the preceding screenshot, IE defines four security zones—**Internet**, **Local intranet**, **Trusted sites**, and **Restricted sites**. Each zone may have a security level applied to control the level and type of content that's accessed and downloaded from sources or networks within each zone.

Security devices

Firewalls and other devices that perform intrusion detection and prevention are key security devices. In addition, each of these devices can be either a hardware appliance or a standalone or add-in software. In addition, a firewall can also be either **host-based** or **network-based**. Each has its advantages, but their applications are very different:

- **Host-based:** This type of firewall or security device is most common, as either software or hardware, on a network-connected host. The host can be a network node or a server, including a proxy server, that permits or denies message traffic that is either inbound or outbound. Host-based devices are more customized to the needs of smaller networks.
- **Network-based:** This is the type of firewall or intrusion security device most people envision: a security filtering device that permits or denies inbound traffic. A network firewall, an intrusion detection system, or an intrusion prevention system are examples of network-based security devices. Network-based devices, because they provide protection over a broader system, are, as the name implies, better suited to network-wide protection.

Firewalls, regardless of their location in a network, are essentially just firewalls. However, intrusion detection and prevention systems commonly carry their placement in a network as part of their device names. A host-based intrusion detection system is an HIDS and a network-based system intrusion detection system is an NIDS.

Authentication protocols

The first part of the AAA procedure is authentication (followed by authorization and accounting), and attempts to verify that the user who's attempting to gain access has provided verifiable credentials or biometric evidence that they are who they say they are. In other words, all is well, if the password and username provided is in the list of good identifying data.

This may not be the most sophisticated level of security for all situations, but it's the most commonly used method for authenticating a user. A variety of protocols that use similar methods perform authentication. In the following sections, we'll look at the most common authentication protocols.

Authentication methods

Authentication uses input data or images that are provided by a user to verify that the data or image is a match to the data or images stored in advance for that specific user. The most common of these input values include the following:

- **Password:** This is the most common form of authentication input that's combined with a username, word, phrase, token, card, or other identification to verify the person providing the data as an authenticated user.
- **One-time password:** As its name states, you use a one-time password just once. A one-time password can be either a challenge-response password or a password from a predefined password list. A challenge-response password generates a value that the user calculates into a response value or looks up a response in a table. A password list contains values that are entered once for authentication.
- **Public-key:** Two pass keys, one private and one public, encrypt, decrypt, and verify the data that's been provided for authentication.
- **Zero-knowledge:** Users receive a question or mathematical problem to answer or resolve that's unique each time. The CAPTCHA authentication method is an example of this.

As stated previously, an authentication protocol verifies that the identification data provided by a user is valid. More importantly, this procedure is the first line of security defense for an open and accessible network. Authentication protocols can be separated into two groups—**point-to-point protocol (PPP)** and **AAA**.

The following sections describe the most commonly used authentication protocols, with no order implied.

Point-to-point authentication protocols

This category of authentication protocols is a common form of identifying a user requesting access to a system or network. The client is in direct communication with the authentication server and enters an identity phrase, word, or value and a password. If the server can verify this information, the user gains access. The authentication protocols that are included in this category are as follows:

- **Password Authentication Protocol (PAP):** This is a legacy protocol that performs the absolute most basic authentication steps—a user's username and password go to an authentication server. If the server finds these values in its password tables, it sends an acceptance (permission to enter) message back. Because the transmissions between the client and server are in the clear, PAP isn't a secure choice.
- **Challenge-Handshake Authentication Protocol (CHAP):** After receiving a connection signal, the CHAP server (authenticator) transmits its hostname and a randomly generated data string (challenge) to the requesting client. The client then, using the challenge value, determines a corresponding secret value. The server then receives a bundle of data, including the secret value, the original challenge value—both encrypted with a one-way hash—and the client's hostname. The server duplicates the process that's performed by the client. If the result is the same as what was provided by the client, the client gains access. In addition, the CHAP handshake process may repeat periodically to reestablish its secure connection.
- **Extensible Authentication Protocol (EAP):** EAP is a general protocol that combines with different authentication methods, including Kerberos, one-time passwords, digital certifications, smart cards, PKI, and others. EAP versions support both wired and wireless LANs. The difference between EAP and the other APs is that EAP essentially acts as middleware between the client and an authentication server, such as RADIUS. The various versions of EAP include EAP-TLS, EAP-FAST, PEAP, and LEAP.

AAA authentication protocols

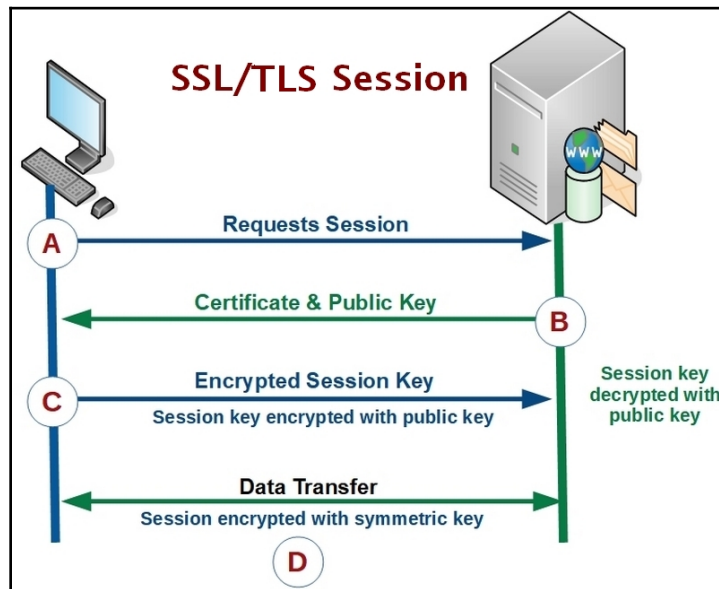
The *triple-A* (AAA) protocols provide the mechanisms to enforce, regulate, and monitor user access to a system or network. Authentication enforces who has access. Authorization regulates what a user may do, while accounting keeps track of what each authenticated and authorized user did. The most common of the AAA protocols include the following:

- **Kerberos:** Although not nearly as fierce as its namesake, Cerberus, the three-headed dog that guards the gates of Hades, Kerberos is a secure authentication protocol. Kerberos uses what it calls a *ticket*, which is an encrypted proof of identity to identify a user or a local network node. Kerberos authenticates a user to multiple systems rather than to a single system.
- **Lightweight Directory Access Protocol (LDAP):** An open standard for directory services, LDAP is common in authentication processes for storing and verifying user accounts. LDAP maintains a database (directory tree) of user credentials. Users attempting to log into a system or network enter their credentials and LDAP searches its tree. If LDAP finds the credentials, the user gains access.
- **Remote Authentication Dial-In User Service (RADIUS):** This authentication protocol provides centralized network access control, although, in most cases, not through a dial-in service. When a user attempts to log into a system, his or her credentials go to the RADIUS server in an access-request message. The RADIUS server, depending on whether the user's credentials are valid, responds with an **Access-Accept**, **Access-Reject**, or **Access-Challenge** message.
- **Terminal Access Controller Access Control System (TACACS):** This is another legacy AAA protocol that, like RADIUS, forwards user credentials to an authentication server for verification and permission to gain access to a system or network. A newer version, **Extended TACACS (XTACACS)**, includes authorization and accounting. Neither version should be confused with TACACS+, which is a totally different protocol.
- **TACACS+:** This is a version of TACACS that was developed by Cisco systems and encrypts its packets entirely before forwarding them to an authentication server. TACACS+ transmits on TCP, while TACACS/XTACACS use UDP.

Secure Sockets Layer (SSL)/Transport Layer Security (TLS)

The SSL and the TLS, which is an updated version of the SSL, secures TCP communications, especially for HTTP messages.

The following diagram illustrates the handshake process between a client and server using the SSL/TLS protocol:



The steps in the SSL/TLS handshake

Let's go through the steps of this handshake process:

- **Step A:** The client sends a *Hello* message to the target server. This message includes the version of SSL or TLS the client is using and the client's preferences for encryption algorithms (cipher suite), a compression method, and a random string value for use in computations.
- **Step B:** If all goes well, the server responds with its *Hello* message, which contains the server's choice for the cipher suite (from the client's list), a session ID, and its own random string value. The server then provides a digital certificate to the client. The server may request a certificate from the client.

- **Step C:** The client verifies the server's certificate. If verified, it sends the random string value, encrypted with the server's public key, to the server. This value will generate the encryption key for any subsequent messages.
- **Step D:** After exchanging encrypted *finished* messages to signal the end of the handshake's creation, the client and server transmit data and messages that have been encrypted with the shared key.

Internet Protocol Security (IPSec)

The IPSec defines a series of standards for both the encryption and transmission integrity of OSI network layer packets that are transmitted with transport layer protocols. IPSec uses tunneling protocols to securely transmit data on a network and to prevent corruption of the data while in transit.

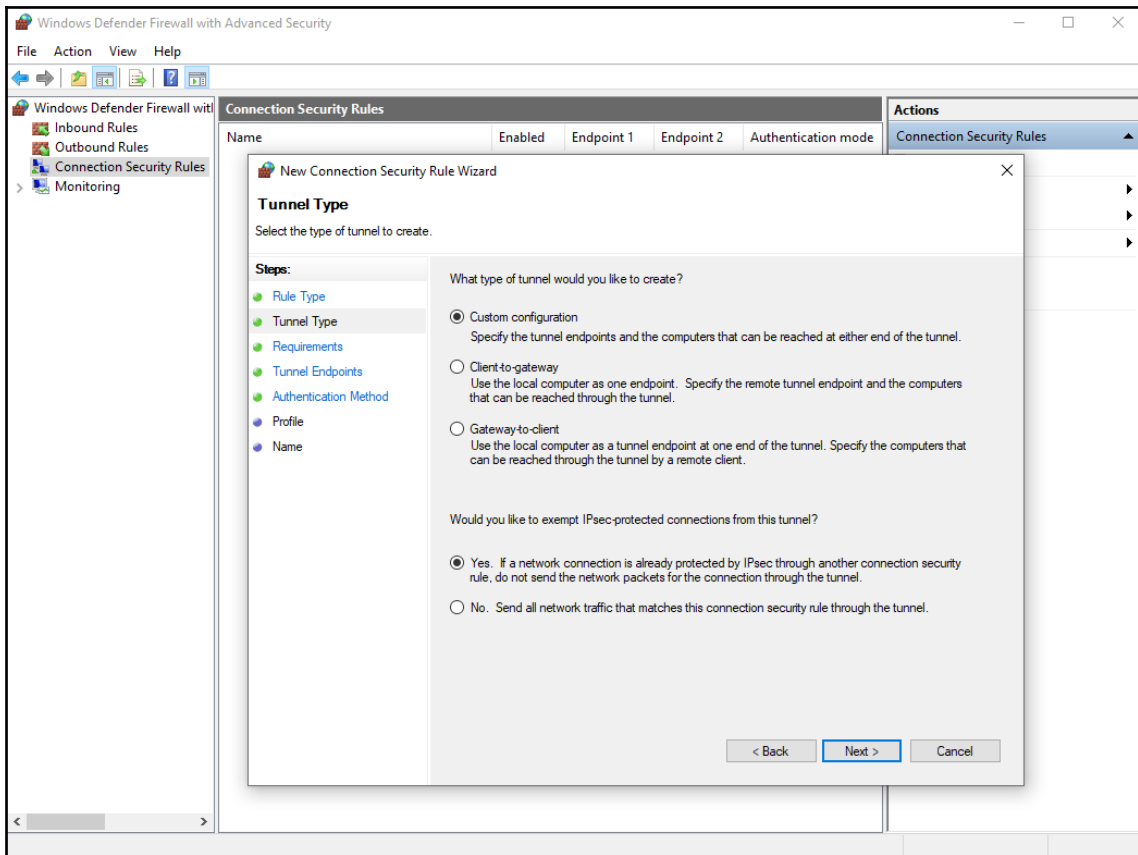
Security policies defining the security requirements of the network define IPSec functions. Each security policy defines a filter. These filters relate to the sources or destination IP addresses, transport layer port numbers, or the encapsulating protocol of a packet. When the data in a packet matches that of a filter, the filter's action applies.

IPSec policies

An IPSec policy contains rules and filters, which define the what, when, and how of securing transmitted network traffic. Rules and filters relate to specific types of network traffic and the level of security is required.

The functions and actions of an IPSec system are set through the rules that it's been configured to follow. Windows operating systems contain a set of example security rules, but these rules are strictly examples. They really don't provide much in the way of actual security. IPSec rules specifically define the domain to which it applies and whether it's in transport or tunnel mode (more on that later).

The following screenshot shows the dialog box being used to define a new security rule:



Defining a new IPsec policy rule on a Windows OS

IPsec policy rules may consist of one or more filters (so-called because they effectively perform packet filtering). Each filter may include the following information:

- **Filter list:** This is a list of the incoming or outgoing network messages on which the filter is to apply its filter actions.
- **Filter action:** A filter has three basic functions it can apply to a message—**permit**, **block**, or **negotiate** security. The actions of permit and block should be self-explanatory. The negotiate security filter action applies IKE to determine the security mode (**Authentication Header (AH)** or **Encapsulating Security Payload (ESP)**), the encryption method, and other session security data.

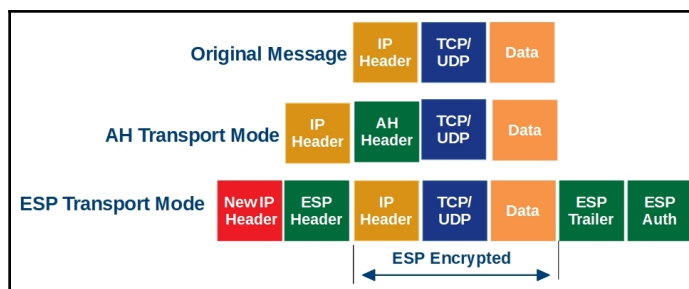
- **Filter authentication methods:** Kerberos, a CA certificate, or a pre-shared security key authenticate each endpoint in a security negotiation.
- **Tunnel endpoints:** If this value is present, it serves two purposes: an indication that tunneling protocols are in use and providing the IP address of one of the tunnel endpoints. Tunneling mode uses two filters, one for each end (and direction) of the tunnel.
- **Connection type:** This is an indicator of the type of connection to which a filter applies, which can be a LAN, a remote connection (VPN), or both.
- **Default response rule:** If one of the endpoints in an IPSec session doesn't have a security rule defined, the default response rule provides the missing filter.

The filter action applies when the filter list, and possibly other rules, matches the corresponding data in a packet, such as its IP addresses, port number, and protocol.

IPSec modes

IPSec can operate in two primary modes—**transport** mode or **tunnel** mode. In transport mode, which is IPSec's default mode, message security is end-to-end between a network client and a network server through two methods—**payload encryption** or **secure communications**. IPSec may operate in either of these configurations using one of the following modes:

- **AH transport mode:** As its name suggests, AH mode inserts a header into each packet, which contains a keyed hash total to ensure the packet's integrity. AH mode doesn't encrypt packets by default, but local administration may provide custom data checking and encryption rules.
- **ESP transport mode:** ESP mode essentially includes the message actions of the AH mode with the addition of the encryption of the payload (data) portion of the message only:

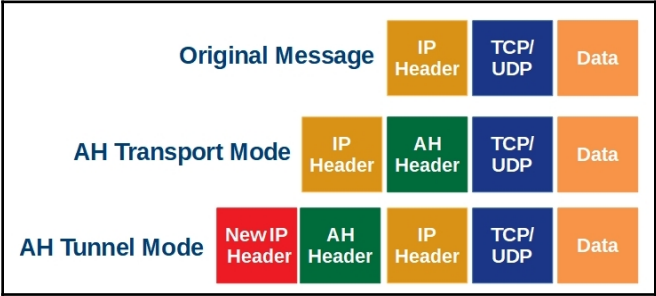


A comparison of the AH and ESP message formats

The preceding diagram shows the packet formats of the IPSec transport modes. As we can see, the primary differences are that ESP encrypts the original message and appends additional information to the packet for authentication and integrity checking.

Each transport mode has an associated tunnel mode. In either format, IPSec applies encryption to the entire message after it's formatted into either an AH or ESP transport mode packet. After encapsulating the encrypted message, an IP header with the IP addresses the source and destination endpoints of the IPSec tunnel.

As illustrated in the following diagram, AH tunnel mode places the AH header before the contents of the original packet. After encrypting this header and the packet's contents, AH adds an IP header with the associated IP addresses:



AH protocol transport and tunnel mode message formats

Another operational mode of IPSec is the **Internet Key Exchange (IKE)**, which can supervise the authentication, application of security policies and rules, and key exchange activities of each side of an IPSec interaction.

Port security

Many switches and routers have all of their interfaces enabled by default, right out of the box. As we've discussed, hardening—that is, closing unused ports—helps to secure a network by closing possible entry points. There are several approaches, standards, and methods used to secure internetworking devices, especially routers and switches, most of which are **port-based network access control (PNAC)** methods.

Port-based security

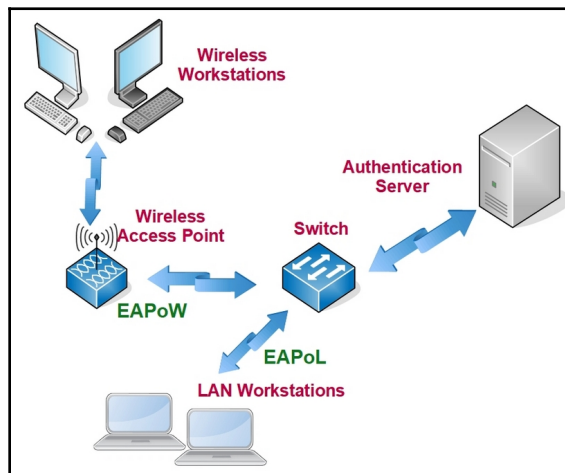
In general terms, port-based security secures the interfaces (ports) of a switch by limiting the number or specific devices that may forward packets or frames to one or more of its ports. Port security has two approaches:

- **Dynamic locking:** This sets the maximum number of MAC addresses the device can *learn*. After reaching the limit, the device ignores any additional unknown MAC addresses and any messages from those devices.
- **Static locking:** The device forwards only the MAC addresses included in a manually configured static address list.

IEEE 802.1x

Also known as *dot1x*, IEEE 802.1x defines a PNAC standard that performs an authentication process before allowing access to a device interface port. IEEE 802.1X provides for authentication and encryption key management on IEEE 802 networks, both wired and wireless. 802.1x is based on the EAP and is applied through **EAP over LAN (EAPoL)** and the **EAP over Wireless (EAPoW)** protocols.

In the following diagram, in the 802.1x authentication process, the client software on the wireless workstations and the LAN workstations is the supplicant, the wireless AP and the LAN switch serve as authenticators, and the authentication server is, well, the authentication server:



The flow of the IEEE 802.1x authentication process

The IEEE 802.1x authentication process goes like this:

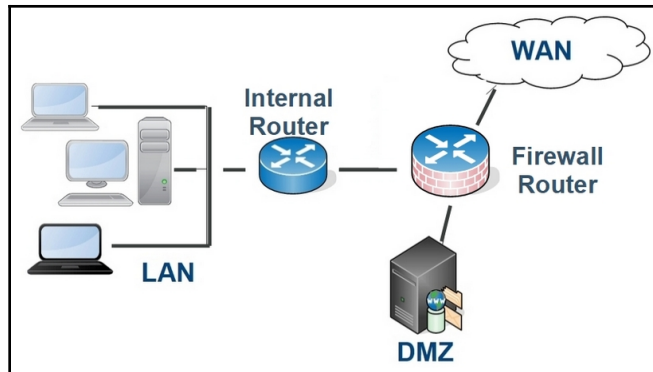
1. The client software (supplicant) on a LAN workstation (wireless or wired) sends an EAPoL start packet to an interface port on an **access point (AP)** or switch.
2. The AP or switch (authenticator) replies with an EAPoL identity request.
3. The supplicant sends back its identity, which is typically a username or an identity code.
4. The authenticator forwards the supplicant's information to the authentication server (commonly a RADIUS or DIAMETER server).
5. The authentication server, after verifying the identity of the supplicant, sends the authentication method it prefers to the authenticator, which forwards it to the supplicant. The authentication methods that are supported in this standard include **Tunneled Transport Layer Security (TTLS)**, **Transport Layer Security (TLS)**, **Message Digest version 5 (MD5)**, **Protected EAP (PEAP)**, and others.
6. The supplicant replies to the authentication server, through the authenticator, with its access credentials, such as a username and password or a digital certificate.
7. The authentication server verifies the information that's provided by the supplicants and either permits or denies access.

Access control list (ACL)

An ACL is a filter that's applied by internetworking devices, especially routers, firewalls, and switches, to identify and permit or deny access to an internal network to incoming message traffic. For the most part, ACLs are associated with an interface port on a device, and is under the assumption that the message traffic entering that interface is definable by source, destination, protocol, or port number.

Router ACLs

The objectives of the Server+ certification exam identifies router ACLs as something you will encounter on the exam. An ACL on a router performs the same basic function that we described in the previous section. However, router ACLs provide specific or targeted access to permit or deny filtering. A very common application for a router ACL is as part of an inbound DMZ, as illustrated in the following diagram:



An ACL can be the controlling mechanism in a DMZ setup

A router with one or more strictly configured ACLs can, in many respects, serve as a firewall first and a router second. The firewall/router, based on the filters in its ACL, can direct any incoming messages with unknown or blocked data.

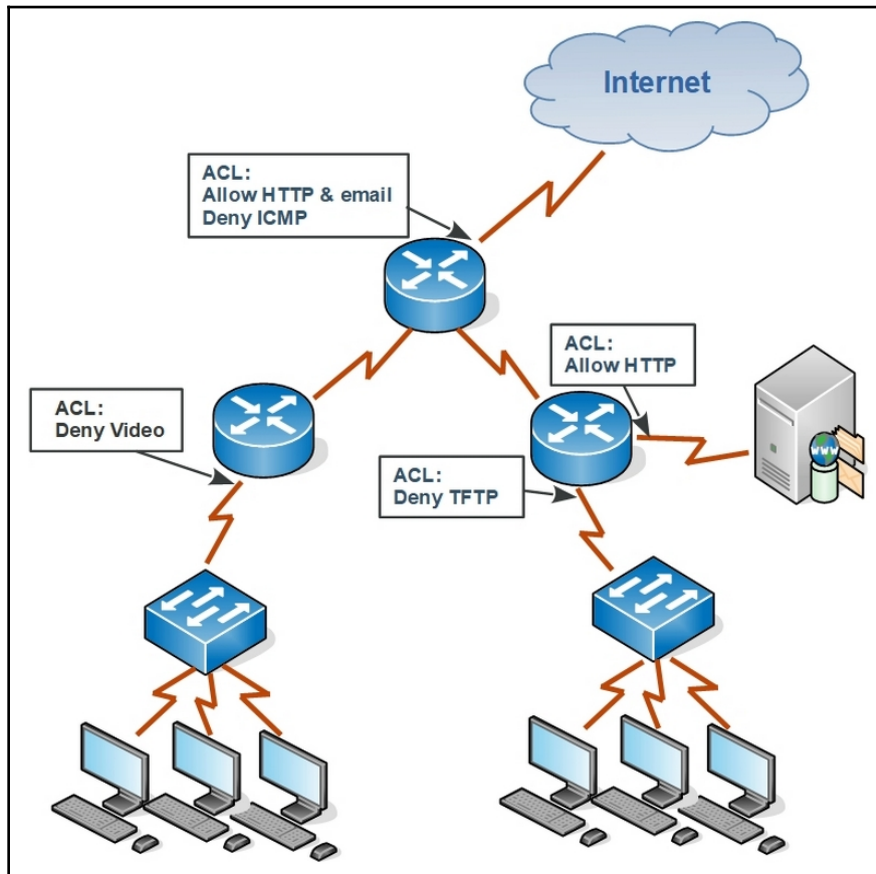
A network DMZ doesn't have to be a dead-end destination for denied messages. It's common for networks to route unknown source addresses and specific protocols to a DMZ. This may include a web server, remote access servers, and more.

Access list content

Generally, ACLs from different router providers can have features or controls that vary from model to model. However, nearly all routers support the following ACL features:

- **ACL identification:** In different routers, the identity of an ACL could be a name, a combination of alphanumeric characters, or a number, which could have significance in its value.
- **Access Control Entry (ACE):** An ACL may have one or more ACEs, each of which identifies a specific set of conditions that determine whether the content of an incoming (or perhaps outgoing) message permits it to be passed, denied or redirected.
- **Source and destination addresses:** ACLs and ACEs determine Permit or Deny using a message's source or destination IP address. The ACL compares the message's addresses to those defined in an ACE, which may be a range of addresses.

The following diagram illustrates how different parts of a network structure may place restrictions on different message types, originators, destinations, and protocols:



ACLs applied throughout a network infrastructure

ACL types

As with other features of the routers from various providers, the types or categories of ACLs may also vary. However, the most common ACL types are either standard or extended.

Standard ACLs

A standard ACL filters messages using only its source (originating) addresses. Standard ACLs are perhaps the simplest to use and create, but their simplicity also limits their effectiveness. An ACE in a standard ACL contains only the following content:

- **Access control list number:** A standard ACL must have an ACL number of 1 through 99 (inclusive) or 1,300 through 1,999 (inclusive). These number ranges identify it as a standard ACL.
- **Source IP address:** This is the specific IP address or address range compared to the source IP address of an incoming message. If there's a match, the corresponding Permit or Deny action applies.
- **Wildcard mask:** ACL entries may include a wildcard mask that's used similarly to a subnet mask. More on this in the *Wildcard masks* section later.
- **Permit or deny:** This is the action that's taken if the comparison of the source IP address against the address or address range is true.

Extended ACLs

An extended ACL provides the ability to filter message traffic using either or both the source and destination addresses, as well as a TCP/UDP port number. In other words, it extends the capabilities of the standard ACL. An ACE in an extended ACL may contain any of the following:

- **Access control list number:** Extended ACLs must have an ACL number of 100 through 199 (inclusive) or 2,000 through 2,699 (inclusive). These number ranges identify it as an extended ACL.
- **Permit or Deny:** This action is taken if the comparison of the message content against the ACE content is true.
- **Protocol:** These protocols are **Internet Protocol (IP)**, **Transmission Control Protocol (TCP)**, **User Datagram Protocol (UDP)**, **Internet Control Message Protocol (ICMP)**, **Generic Routing Encapsulation (GRE)**, and **Internet Gateway Routing Protocol (IGRP)**.
- **Source IP address:** This is the public IP address from which the message originated.
- **Source address mask:** This is the subnet mask associated with the source IP address' network or subnet.

- **Operator source port:** This indicates the condition of the comparison test involving the port number and contains both the conditional operator and the port number. The conditional operator may be **less than (lt)**, **greater than (gt)**, **equal (eq)**, or **not equal (neg)** and contains the TCP/UDP port number.
- **Destination IP address:** This is the public IP address of the receiver of the message.
- **Destination address mask:** This is the subnet mask associated with the network or subnet of the destination IP address.
- **Operator destination port:** This is the same as for the operator source port.

Other ACL types

In addition to standard and extended ACLs, some routers also implement two other ACL types:

- **Ethertype:** This permits or denies Layer 2 (Ethernet) frames.
- **Webtype:** There are two different webtype ACLs—the URL-based ACL filters specific protocol and the URL combinations. The TCP-based ACL filters permit or deny specific IP addresses and TCP port numbers.
- **Reflexive:** This ACL type, that is, an IP session, creates a permitting inbound ACL for responses to outbound messages.
- **Dynamic:** This type of ACL, that is, lock-and-key, allows a user to access a specific source or destination IP address.

In addition, ACLs also fall into one of two general types:

- **Discretionary ACL (DACL):** This type of ACL specifically identifies the IP addresses, user or group accounts, port numbers, and protocols with permission to access a resource
- **System ACL (SACL):** This type of ACL controls the router feature that generates log or audit entries and details attempts to access a resource

As an example of a numbering system for ACLs, Cisco Systems uses the designations that are shown in the following table:

Protocol	Range
Standard IP	1–99 and 1,300–1,999
Extended IP	100–199 and 2,000–2,699
Ethernet type code	200–299
AppleTalk	600–699
Ethernet address	700–799
Internetwork Packet Exchange (IPX)	800–899

ACE types

As we've already discussed, an ACL consists of one or more ACEs. Each ACE has one of three general purposes, which are as follows:

- **Access-denied ACE:** An entry in a DACL that specifically denies access based on a requestor's information
- **Access-allowed ACE:** An entry in a DACL that specifically permits access based on a requestor's information
- **System-audit ACE:** An entry in a SACL that generates a tracking entry when a requestor attempts to access to a resource

The testing and its resulting action of each ACE occur one entry at a time. It's almost as if no other entries are in the ACL. Each ACE has no relationship to any other ACE before or after it, other than where it is in the sequence of the ACL.

The sequence of the ACEs in an ACL is extremely important. The test conditions and their results should be logical and not contradictory. For example, if an ACE denies access to all messages from a certain IP address and later in the ACL another ACE permits only ICMP messages from that address, the deny ACE makes the permit ACE moot. After denying everything, there's nothing to permit. I bring this up not only to make a sequencing point, but to lead into the purposes of implicit and explicit deny actions.

An implicit anything is a result that occurs by default. An explicit anything is specific. Implicit is an assumption and explicit is a stated fact. For example, if a tour guide has a list of participants that can board the tour bus, not being on the list implicitly denies anyone else boarding. However, a competing tour company has a list of people who are absolutely barred from boarding their bus. In this case, being on the list explicitly denies participation.

Nearly all routers, firewalls, active directory servers, and other access control devices automatically include an implicit deny at the end of an ACL. This implicit deny functions to block any access requests that aren't included in the ACL, meaning that requests that have fallen through all of the ACEs in the ACL. An implicit deny as the last entry in an ACL says it denies any request that got that far.

An explicit deny can be anywhere in an ACL, depending on how specific it is. An explicit deny for a specific IP address should precede any general permit entry. An explicit deny at the end of an ACL is typically a *deny any* entry that functions exactly like the catchall implicit deny.

Wildcard masks

Also known as inverted masks, ACL IP address entries may include and apply a wildcard mask, which work much differently compared to the wildcard characters of the Windows world. In fact, ACL wildcard masks work essentially in the opposite way to subnetting masks.

In a subnet mask, the masking process only extracts the binary value positions that contain one value from an IP address. However, in an ACL wildcard mask, a binary zero indicates the address positions that must match. In other words, a zero bit in the mask means the corresponding bit in the message address must match (to create a true condition) and a one bit indicates an ignored bit position in the address.

For example, in a standard ACL entry, the source IP address and its wildcard mask appear (one after the other) as `162.29.5.12 0.0.0.0`. In this entry, `162.29.5.12` is the source IP address and `0.0.0.0` is the wildcard mask. The zeroes in the wildcard mask represent that each octet contains all zeroes. Otherwise, the wildcard mask would be `00000000.00000000.00000000.00000000`. Another way to represent the same setting is to use the `host2` term before the IP address: `host 162.29.5.12`.

The more common use of wildcard masks is in extended ACLs, where they identify entire or major subnets of a network. For example, if an ACL entry is to permit all addresses on a certain network, the one bits indicate the portion of the address that's permitted and the zeroes indicate the portion of the address that's ignored. If we wish to permit all of the nodes on the `210.20.158.0` network, the wildcard mask will be `0.0.0.255`. The 255 portion of the mask means that whatever host ID is in the fourth octet, permit it.

Public key infrastructure (PKI)

Essentially, the PKI is sort of like one of your close friends introducing you to some of their close friends. Officially, PKI is a suite of rules, guidelines, and policies that define, issue, manage, apply, store, and revoke digital certificates and the use of public key encryption. In simpler terms, PKI is an encryption and security methodology that provides better security for the identity of the correspondents and the content of the data than simpler methods, such as a password.

Practically, PKI is the coordinated interaction of its four parts, which are as follows:

- **Certificate authority (CA):** A trusted organization that provides unique digital certificates to subscribers and manages public keys and identity credentials for the data encryption of stored data, websites, and email
- **Registration authority (RA):** A network service that approves and forwards requests for identity verification (digital certificates) and the certificate authority that issues it
- **Certificate request database:** Stores requests for digital certificates
- **Certificate store:** Stores digital certificates

PKI features

PKI describes the following processes and procedures to provide a secure infrastructure:

- **Access control:** Using public and private key pairs, PKI assures that only identified parties have access to a document
- **Authentication:** PKI provides identity verification through digital certificates
- **Confidentiality:** PKI secures transmitted documents against unauthorized access through encryption
- **Integrity:** PKI assures transmitted data retains its integrity through message hashing
- **Non-repudiation:** A document's digital certificate permanently identifies its ownership

Encryption and authentication

These are security protocols that provide data encryption and authentication for messages that are transmitted over a network. Their differences aren't large, but they are not interoperable. However, since TLS is the replacement for SSL, you will often see them listed as SSL/TLS.

Digital certificates aren't dependent on specific protocols. The protocols that are enabled on a system or network are the property of the system or network configuration and not the security certificates in use. For example, an SSL digital certificate doesn't require only the SSL protocol, and TLS digital certificates don't require only TLS. In fact, for the most part, certificate authorities refer to these certificates as SSL/TLS certificates.

Virtual private network (VPN)

A VPN creates an encrypted end-to-end connection over an insecure network, primarily over the internet. VPNs are common in companies for their remotely authorized users, such as remote branch office workers, to access central office applications and other resources.

A VPN connects two locations over a network using an encrypted tunneling protocol. In fact, a VPN encrypts the elements of both the private network and the public network that's used to complete the connection. The three primary protocol types that are used with VPNs are as follows:

- **IP Security (IPSec):** A set of security and encryption protocols for protecting IP packets that are transmitted over a network using either transport mode or tunneling mode
- **Point-to-Point Tunneling Protocol (PPTP):** A protocol for creating VPNs
- **Layer 2 Tunneling Protocol (L2TP):** A protocol that allows **Internet Service Providers (ISPs)** to provide VPNs to subscribers

Virtual LAN (VLAN)

Where a VPN allows secured connections over a public network, a VLAN is a logical internal network configuration that can provide several benefits to users and administrators. A VLAN creates a group of network nodes into a logical network. This is a virtual network in that the included nodes may be on separate physical networks.

A VLAN can be one of two types:

- **Static VLAN:** Also known as a port-based VLAN, it consists of one or more network switch interface ports that have been configured to a VLAN. Any device that's connected to a port that's been configured to a static VLAN is automatically on that VLAN.
- **Dynamic VLAN:** The identification of a node in a dynamic VLAN is done by its physical address (MAC address) or network usernames. A **VLAN Member Policy Server (VMPS)** provides the list of nodes belonging to the dynamic VLANs and provide the configuration data to the effected network switches.

Summary

A security zone is a logical structure that's created from one or more interfaces that apply the same security policies. Many firewalls predefine security zones, with the most common being LAN, VPN, DMZ, and WAN. A security zone applies security policies and access control policies.

Security devices, such as firewalls and intrusion detection and prevention devices, are either hardware or software. Security devices can be host-based or network-based. A host-based intrusion detection system is a HIDS and a network-based system is a NIDS.

An authentication protocol verifies the credentials that have been provided by a user are valid and serves as the first line of security defense for an open network. Authentication protocols are either PPP or AAA. The most common of the AAA protocols include Kerberos and IPSec. An IPSec policy defines rules and filters for specific types of network traffic and security levels. IPSec operates in transport mode or tunnel mode and within each in AH or ESP modes.

Port-based security has two approaches: dynamic locking and static locking. IEEE 802.1x PNAC performs authentication before allowing access to an interface. 802.1x is based on EAP.

An ACL permits or denies access to the network via incoming messages. ACLs are associated with interfaces and apply to messages using source and/or destination addresses, protocols, or port numbers. Router ACLs include ACL ID, ACE, and source and destination addresses. A standard ACL filters messages using source addresses. An extended ACL filters messages using source and destination addresses and TCP/UDP port numbers. The two general types are DACL and SACL.

Routers and other access control devices automatically include an implicit deny at the end of an ACL. An explicit deny can be anywhere in an ACL.

PKI is an encryption and security method that secures the source and receiver of a message or document. PKI has four parts—CA, RA, certificate request DB, and certificate store. PKI provides access control, authentication, confidentiality, integrity, and non-repudiation.

A VPN creates an encrypted end-to-end connection over the internet using a tunneling protocol. The protocols that are used with VPNs are IPSec, PPTP, and L2TP. A VLAN is a logical network configuration that is either static or dynamic.

Questions

1. What is a logical structure of one or more interfaces with the same security policies?
 1. HIDS
 2. NIDS
 3. Control zone
 4. Security zone
2. Which two of the following are classifications for security devices?
 1. Host-based
 2. Host-attached
 3. Network-based
 4. Network-attached
3. A protocol that verifies user credentials performs which part of the AAA function?
 1. Authorization
 2. Accounting
 3. Authentication
 4. Association
4. The two most common of the IPSec protocols are Kerberos and what?
 1. L2TP
 2. EAP
 3. IPSec
 4. PNAC
5. Which of the following modes are associated with IPSec?
 1. AH
 2. Transport
 3. ESP
 4. Tunnel
 5. All of the above
 6. None of the above

6. What protocol standard is IEEE 802.1x based on?
 1. IPSec
 2. L2TP
 3. EAP
 4. DACL
7. What are the two approaches of port-based security?
 1. Dynamic locking
 2. Static locking
 3. Certificate locking
 4. Interface locking
8. Which of the following is the data that a standard ACL uses to filter messages?
 1. Source address
 2. Port number
 3. ACL identification number
 4. Protocol
9. What ACL entry is automatically added at the end of an ACL to deny any message that's not matched by its other entries?
 1. Explicit deny
 2. Explicit permit
 3. Implicit deny
 4. Implicit permit
10. Which of the following isn't one of the four parts of the PKI model?
 1. CA
 2. RA
 3. Certificate store
 4. PPTP

12

Physical Security and Environmental Controls

Not all network and server security devices and methods installed on, attached to, or integrate with computer hardware and software. The first line of defense for any server-based system, data center, or facility is its physical security. Physical security is all about keeping the computing assets safe, functioning, secure, and away from all intruders. Although we'll treat it as a separate topic, the environmental systems and controls for the physical environment should be a part of an organization's security policies.

In this chapter, we'll look at the physical security and environmental controls that protect computing systems. The topic areas covered are as follows:

- **Multi-factor authentication (MFA)**
- General physical security concepts
- Environmental controls
- Physical safety issues

MFA

Over the years in which users could access computers, systems, and networks, user authentication systems evolved from nearly nothing to combining multiple identity factors. This started with requiring user identity codes, then a secret password entered with the username, and now combining two or more identity factors. All of these levels had one primary purpose—to verify the identity of a user seeking access, as certainly as possible.

Passwords

Perhaps the first password, at least the first one written about, was the voice command used by Ali Baba to access a treasure—*open sesame*. At one time, a user account or identity code served the purpose of identifying who was accessing a system, which was typically a mainframe or a minicomputer. Eventually, because of sharing among co-workers and their simplistic content, these codes ceased to provide any level of secure access control. User account codes, also known as usernames, commonly consisted of only a first initial and a last name of each user with access to the system, such as `rprice`. This a standard practice that's still in use today. Usernames haven't really changed much over their 50-year history, nor has the security they provide, which is essentially none.



You must remember that the primary data storage media for early application systems was primarily removable media, such as hard disk platters, spindle packs, magnetic tape, and floppy disks. Physical security that blocked access to the physical media was a large part of access control.

As application systems grew in sophistication, data storage was on attached or internal disk storage devices. This meant that security had to become an absolute necessity. Maintaining data integrity and security required stronger authentication methods. Adding a *secret* password to the login process helped to verify that a user requesting access to a system was in fact who he or she purported themselves to be. The underlying assumption was that if the user knew the password, then the user was who the username identified.

The use of a username is merely one step (the identification step) in an authentication process. Pairing a password with the username creates a data bundle that serves as a **single-factor authentication (SFA)** value. If the password entered is the value paired with the username in the system's security files, the system grants access to the user.

Eventually, the use of the combined username and password no longer provided the access control and security it once did. Password crackers, social engineering, malware, and other shady means were able to obtain user passwords that lead to accessing the user's personal data. This is where MFA came in.

Authentication factors

MFA combines two or more identification factors to authenticate a user (supplicant) attempting to log on to a system or network. The MFA process involves the combination of two or more factors that uniquely identify the supplicant. So far, MFA can process, in various combinations, five general categories of authentication factors, which are as follows:

- **Something you know:** This category includes data, phrases, numbers, or any combination thereof that you've memorized and are able to recall when needed. This could be a password, a PIN code, the answer to a question, how many dogs George Washington owned (50), and so on. Usernames and email addresses, although they're something you know, are identifying codes, which is what we're trying to authenticate.
- **Something you have:** The factors in this category are physical items that contain information that assures that the holder/user of the item is, in fact, who they claim to be. This could be a small plastic token, such as an RSA SecurID, or a number generated from a bank's app on your mobile phone, a **Personal Identity Verification (PIV)** smart card or employee badge, or perhaps an RFID proximity card.
- **Something you are:** Many people confuse this category with the something you have or are categories, but essentially this category is biometrics. Biometrics uses parts of your body to uniquely identify you. Your body features, which can include fingerprints, handprints, retina and iris scans, voice prints, and facial scans, establish a pattern in binary coding, compared to when you attempt to log in.
- **Somewhere you are:** This type of authentication most commonly uses IP addresses to identify the location or source of a user request. A geolocation service verifies that the location from where a request originates is consistent with the location predefined for the user making the request.
- **Something you do:** This category incorporates touchpad and touchscreen movements (gestures), something such as a secret hand sign. Windows 8 introduced a feature called **Picture Password** that allowed users to record gestures for use in authentication.

Using only one factor, such as just a password, is SFA. Using just two factors is **Two-Factor Authentication (2FA)**. As you've probably figured out by now, more than two factors in an authentication is MFA. The one thing you should know regarding authentication factors is that the more you use, the safer your data will be.

General physical security concepts

The physical security of any organization must define the events, causes, actors, prevention, recovery, mitigation, and other relevant procedures regarding the security, safety, and operations of achieving an organization's mission.

A physical security program for any IT organization is based on the overall organization's security policies and plans. The IT security policy must recognize the need and purpose of physical security for the computing assets and functions. The policy that defines the IT physical security program should include the following objectives, and most likely others, depending on the nature of the organization:

- The physical safeguards that are in place, such as fire safety, intrusion prevention, business continuity, and disaster recovery
- The sources and types of threats to the organization
- The components of physical security monitoring systems
- The impact and recovery from utility service interruptions (electricity, natural gas, water, and so on)
- The countermeasures defined or in place in the event of unauthorized entry, vandalism, or the theft of computing resources

A physical security plan typically has two areas of focus—**deterrence** and **detection**. Deterrence procedures, methods, and equipment serve to discourage or prevent threats becoming security events. Detection, of course, involves the equipment, procedures, and methods used to discover events that threaten or defeat the physical security measures.

Threats to physical security

Threats to the physical security of any organization fall into four general categories—**environmental**, **man-made**, **site-specific**, and **technical**. The following sections give some examples of the threats in each of these categories.

Environmental threats

Environmental threats are weather, natural disaster, catastrophic event, and other events from the natural world. They include the following:

- **Dust:** Airborne dust can contribute to electrical problems in a computer, not to mention clogging up the airflow and cooling of the motherboard and CPU.
- **Earthquake:** The damage threat of an earthquake is usually structural, but with that may come electrical, plumbing, and certainly ventilation issues.
- **Extreme weather:** This threat could include all sorts of catastrophic weather events, including high winds, heavy rain, tornadoes, hurricanes/typhoons, ice, snow, hail, and so on.
- **Fire/explosion:** In addition to the damage a fire could do to computing equipment, damage could also come from the chemicals in the retardants, suppressors, and extinguishers, not to mention water.
- **Flood:** Floods are usually a result caused by other weather events, such as extreme weather, earthquakes, or your neighborhood volcano erupting.
- **Lightning:** Electrical storms are a real and serious threat. A direct lightning strike to the building housing a data center could send an extreme electric spike through even protected equipment.
- **Pests:** Unfortunately, a mouse chewing through a communication cable or a power cable can introduce hard-to-detect issues for a server and its nodes.

Man-made threats

As its name implies, these physical security threats are either a result of omission, negligence, or ignorance, or they come from the evil minds of evil-doers. They include the following:

- **Accidents:** Accidents may not be totally preventable, but there are measures that can help to minimize their occurrences: safety tread on laminated floor surfaces, proper spacing between equipment and cabinets, safety awareness training for all staff, and so on.
- **Malware:** Computer viruses, trojan horses, rootkits, bots, and other malicious software are a serious threat to any networked device, but especially to network servers.

- **Theft:** This isn't just the realm of external agents who gain entry to a secured site. Employees, vendors, contractors, and guests, among others, are responsible for a fair amount of equipment, document, and resource theft.
- **Unauthorized access:** Hackers gaining access to programming, data, and other resources are a threat that the security policies and programs must address. However, no less a threat is unauthorized physical access, which means that an intruder has physically gained entry and access to equipment, data archives, documents and forms, and other computing assets.
- **Vandalism:** Although typically thought of as defacement or damage to exterior surfaces, vandalism can also be window and fence breakage and damage, fire, and random destruction of equipment and buildings.

Site-specific threats

The scope of this group of physical security threats is usually a single site, building, or vicinity. For the most part, they involve purchased or contracted systems or services. Some examples are as follows:

- **Communication system failure:** Because communication services provide for a variety of communication technologies, which, in turn, can defeat or remove the performance and reliability of voice, data transfer, and feedback systems.
- **Computing equipment failure:** Obviously, one of the primary risk factors on any network is device and component failures. Physical security policies should address any possible vulnerabilities that could emerge due to certain equipment failures.
- **Fire suppression system failure:** Should a fire start in a secured area and the fire suppression system were to fail, procedures should exist to minimize the overall damage, if possible. It may be possible for technicians to only power-off all equipment and exit the facility.
- **Heating, Ventilating, and Air Conditioning (HVAC) failure:** The severity of this event depends on the location of the facility. In areas where temperatures and humidity are typically high, an HVAC failure could result in serious heat damage to computing equipment. On the other hand, in an extremely cold situation, the loss of heat could degrade the performance of critical components.
- **Key personnel resignation/departure:** Many organizations can't afford to have back-up personnel in place for every key position. When a valuable and hard-to-replace employee leaves the organization, especially if that employee was the one authority for certain activities, recovery from a security event may not be possible without additional help.

- **Power outage/brown-out:** We all know the threat of losing electrical power service to a building, wing, room, or circuit and the operational interruptions that go with it. Frequent power outages (black-outs), which typically involve a power sag followed by a power surge, and brown-outs or periods of low power voltage can degrade the performance and efficiency of a computer's power supply and eventually cause the device to fail.

Technical threats

Physical security threats tend to involve a human in one form or another, but typically because the human takes advantage of a technical vulnerability or opportunity created from a procedural gap. Some examples of technical threats are as follows:

- **Access beyond authorization:** *It happens!* An unsuspecting network user logs into the system and finds that he or she suddenly has access to confidential data that was unavailable to them before. Should the user inform system administration, an investigation and corrective action can take place. However, if it's not reported, an audit and review of user and group account permissions and rights should be immediate, as well as periodical.
- **Intruder attacks:** The attack surface of a computer network are the external points of entry or the internal access points that external and internal threat agents can exploit. An **intrusion detection system (IDS)**, an **intrusion prevention system (IPS)**, or a hybrid device that combines detection and prevention can detect or prevent, respectively.
- **Hardware failure:** Depending on the hardware component that fails, the failure can impact production, security, or both. For example, the failure of an authentication server or the malfunction of a biometric device could create an exploitable vulnerability.
- **Missing or inadequate operating procedures:** Regardless of the event, whether a physical access event or a logical security event, if a clear up-to-date procedure doesn't exist to handle the situation, the action taken could cause additional damage or make the issue unrecoverable.
- **Unauthorized modifications to software or hardware:** Perhaps this should be a man-made threat, but, either way, unauthorized and undocumented changes to production software or hardware could create a variety of physical or logical security vulnerabilities.

Physical security devices

The Server+ certification exam references a variety of physical security concepts and countermeasures. The following lists those items or concepts specifically listed in the exam's objectives with a brief explanation or definition of each and its use in a physical security application:

- **Badges and cards:** As we discussed earlier in this chapter, employee badges can include **Radio-Frequency Identification (RFID)** chips or other embedded circuitry that can be either a microcontroller with a small amount of memory or just a memory circuit that holds identifying data. Smartcards and PIVs are examples of this technology.
- **Biometric devices:** The most common PC-based biometric is a fingerprint scanner that's either a peripheral or built-in device. However, there are several physical security biometric devices that range from fingerprints and handprints patterns, hand and finger geometry, retinal and iris scanners, voice recognition, and vascular pattern scans that scan a hand to match the pattern of the internal blood veins.
- **Lockable cabinets and safes:** In a data center, certain removable media, forms (such as checks or order forms), hot swap components, and other essential items should be in a locked, and preferably fireproof cabinet or safe.
- **Locks and keys:** Perhaps the oldest physical security devices are door and other locks and the keys that open them. An extension of this technology are keypads on which a person seeking to enter an area keys in a security pass code.
- **Mantrap:** This is also known as an air lock or a sally port and provides an entry to a building or secured area with a small space, some call a room or vestibule, with interlocking doors on two sides that serve as an entry and an exit. The interlocking is such that only one door (or set of doors) can be open at any one time. Mantraps are common in areas where the external temperatures are far below or above the temperature inside a building or where some environmental or security procedure occurs between the two doors.
- **Rack mount cabinet:** In larger data centers, the computing and internetworking devices are rack-mounted frames and cabinets. See *Chapter 1, Server Hardware*, for more information on rack mount cabinets.

- **Security cameras and sensors:** Security cameras extend the field of vision for a security personnel (such as a security guard) who can monitor multiple camera images from a single location. A security camera may also provide an image stream for storage on magnetic medium for later review. Security sensors include motion detectors, which can activate other security devices, door and window opening sensors, glass breakage sensors, and a variety of other intrusion detection devices.

Environmental controls

The temperature, humidity, and airflow in a data center or computing facility can be as much a part of a physical security program as door locks. One of the principles of system security is availability and the environment in which the computing devices and servers operate must support the efficient operation of all equipment. Environmental controls implement and maintain the best operating environment for the equipment and devices contained in each space. The various components of environmental controls include the following:

- **HVAC:** In addition to heating and cooling, an HVAC system must maintain a constant even temperature and humidity range for the entire data center. The HVAC system should also interconnect to the fire control system so that the airflow system doesn't continue to feed oxygen to a fire.
- **Room, row, and rack temperatures:** As discussed earlier in this book, large data centers commonly have rows of rack mount servers, power supplies, and network-attached storage in open racks or enclosed rack-mount cabinets. Because different devices have different cooling requirements, a row of racks may be a *cold* row or a *hot* row. The HVAC system can direct cooling to one row and warmer air to another using airflow baffles and rack-blanking panels that create open spaces between rack-mounted devices.

Environmental monitoring

A variety of measurable conditions and ongoing functions contribute to the environment of a computer server room or data center. This environment is a moving target because of changes in the external weather and environment, the housing facility's HVAC system, and those dedicated to the data center.

The conditions that an environmental monitoring system should detect and report include the following:

- **Temperature:** Most environmental guidelines for computer rooms and data centers set the temperature range at 68 to 77 °F or 20 to 25 °C. Keeping the ambient temperature below or above this range for extended periods of time can damage the electronic components.
- **Humidity:** Depending on where a data center is located, humidity can be a problem of being too high or too low. An environment with high humidity may see oxidation and corrosion of electronic circuits and components. At the other extreme, low humidity may generate problems from static electricity. The standard recommendation for the relative humidity in a data center is 45 percent to 55 percent.
- **Voltage and power:** Voltage sensors and power-monitoring sensors monitoring electrical power voltage levels and power consumption against predefined high and low limits, brown outs, and power outages. Events measured to be too high or too low, and for too long, trigger a variety of alerts.
- **Physical intrusion:** Data center doors that remain locked during normal operating times should have open or ajar alarms installed. These alarms let the data center staff know that a door is open and a breach may be in progress.
- **Smoke:** As the saying goes, *Where there's smoke, there's fire*. Smoke and carbon monoxide fumes can be dangerous to people working in the data center. Smoke alarms and chemical detectors that interconnect to the power system and the environmental monitoring and alerting system can shut down the power and trigger alerts in the event it detects smoke or fumes.

Electrical power

Without electrical power, at least to this point in time, there would be no computers. Maybe someday, computers may run on air or banana skins, such as a Mr. Fusion Home Energy Reactor flux capacitor, but, for now, they're electrical devices powered by a connection to the power grid or by a battery. In this chapter, since we are looking at physical security, we'll focus on the power sources to computers and how it's best managed.

Uninterruptible Power Supplies (UPS)

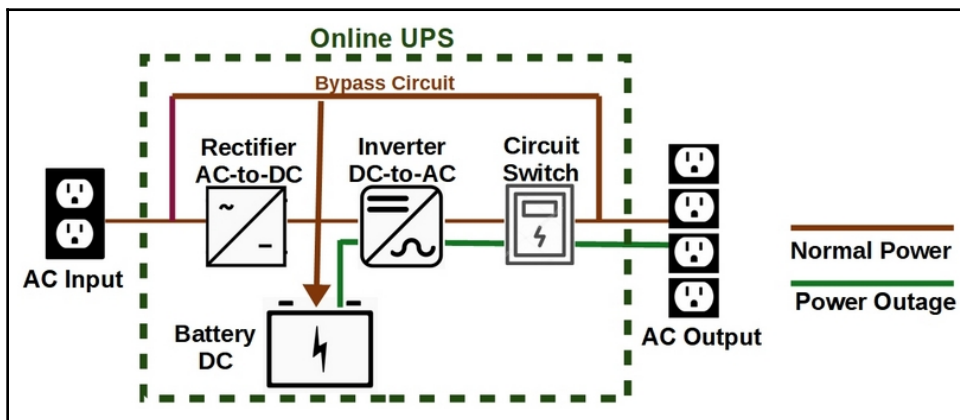
Safeguarding the electrical supply to a data center facility is another important consideration for a physical security program. However, a physical security plan should recognize that typically an UPS protects individual computers and peripherals adequately. The basic operating function of a UPS system is to pass an AC mains power source through to AC-powered devices, while charging a battery with DC power.

The primary purpose of a UPS, whether a standalone or a multi-faceted data center UPS, is to level out power surges and slumps and provide electrical power in the event of a loss of electrical service. Constant power fluctuations and failures can damage or destroy electrical components and may even cause the data center to fail its SLA. When a UPS senses an electrical power event is occurring, it initiates the mitigating action required. This could be to knock down an electrical spike, enhance a power slump, or switch to battery power to counter an outage.

In general, a UPS has multiple internal circuits and a battery:

- **Rectifier:** Converts the AC power source into DC power, which charges the battery
- **Inverter:** Converts DC power into AC power to the output electrical outlets on the UPS
- **Bypass:** Passes AC power directly to the electrical outlets of the UPS
- **Switch:** The circuit that switches between the inverter circuit output and the bypass circuit output

The following diagram shows the circuitry of a typical UPS:



The circuitry of an online UPS

There are two primary types of UPS—**offline** and **online**. An offline UPS, also known as a standby UPS, passes the AC mains power to its own AC outlets. When the AC power source is constant, the UPS passes it through to its output electrical outlets, drawing off only enough to charge its battery. Should the AC mains source drop, the offline UPS uses an inverter to convert the stored DC into AC.

An online UPS, illustrated in the preceding diagram uses a rectifier to convert the AC mains power into DC, which stabilizes the current and charges its batteries. The UPS then converts the DC power into clean and stable AC power. Online UPS systems provide constant AC power without breaks in the power stream, even when the power source drops.

UPS ratings

Typically, UPS ratings are in either maximum **volt-ampere (VA)**, maximum wattage, or both. These values indicate the upper limit of the capabilities of a UPS's, and the power demand on it shouldn't exceed these ratings. The rule of thumb for sizing a UPS for smaller power demands, such as a single computer, is that the wattage rating shouldn't exceed roughly 60 percent of the VA rating. However, on larger systems, it's common that the wattage and the VA ratings are equal.

Two other rating values that indicate the performance of a UPS are capacity and runtime. Capacity is the maximum amount of VA a UPS can output. Runtime, which is dependent on the size of the UPS battery, is the length of time the UPS can provide sufficient power to connected devices.

Automated shutdown of attached devices

Some UPS units can issue shutdown commands to connected devices, which can prevent data or processing losses; a panic-mode activity to shutdown servers and storage devices safely; or a lengthy system restoration. There are two ways to implement this capability—connecting a UPS to a network by adding a network adapter to the UPS or connecting a UPS to each of the devices should mains power drop.

Power distribution

The power supply unit is perhaps the most important component of a PC! This statement is certainly debatable, but consider that, without it, a PC is a bunch of plastic and metal that could be a boat's anchor or a very large paperweight. The power supply unit converts mains AC 120 volts or 240 volts electricity into a DC current typically 3.3 and 5 volts for internal circuits and 12 volts for internally mounted peripherals, such as disk drives, and cooling fans.

Power supplies are essential in desktop, tower, and portable computers. However, rack-or-cabinet mounted systems, such as servers, disk drives, and inter-networking devices, such as routers, switches, and firewalls, maximize computing power while minimizing the floor space required. Most of these devices require an external power source, which is either dedicated to a specific device or shared by several devices. In either case, the external power distribution device is a **power distribution unit (PDU)**:



A power strip (left) and a 16-outlet PDU (right)
Image courtesy: Tripp Lite

Power strips or surge suppressors are common to home and office systems. However, while some PDUs may look like a power strip, there are differences. A PDU, like a power strip, has multiple electric outlets and a mains power cord that connects the unit to an AC wall outlet, UPS, or another electric service distribution device. PDUs are common power units for devices installed in a rack or cabinet.

A PDU distributes electricity through multiple outlets on two primary device styles:

- **Floor-mounted:** This PDU converts raw electrical power sources (unconditioned AC power) into multiple lower-voltage electrical outlets. Floor-mounted PDUs are typically standalone units that power multiple devices, racks, or cabinets.
- **Rack-mounted:** This PDU provides mains electrical power to multiple outlets. Some rack-mounted PDUs include monitoring and control capabilities that an administrator can remotely access to track the power supply.

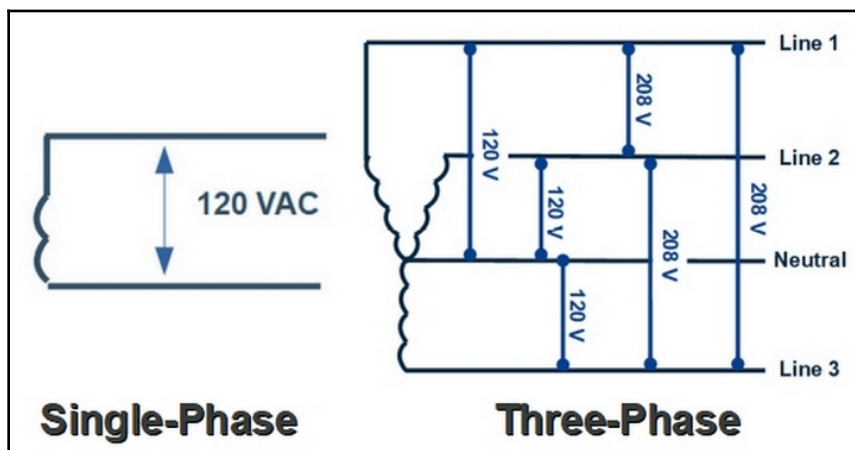
PDU types

PDUs are available in seven configurations, some of which are floor-mounted or rack-mounted units:

- **Standard (basic) PDU:** This PDU type is commonly a power outlet strip placed between a power source, such as a UPS, power generator, or a mains outlet, and a wiring closet, data center, or individual racks or cabinets. It typically has multiple low-amp outlets and is used to provide power to multiple rack-mounted devices.
- **Metered PDU:** This rack-mounted PDU type includes a digital current meter that monitors the electrical load to avoid power surge and overloads.
- **Monitored PDU:** This is another rack-mounted PDU that adds the capability to monitor electrical sources through a network connection to the functions of a metered PDU.
- **Switched PDU:** Also known as a switched-rack PDU, this rack-mounted PDU combines the functionality of a monitored PDU with the capability to turn on or turn off individual outlets to reboot devices over a network.
- **Automatic Transfer Switch (ATS) PDU:** This PDU type can have two (or more) independent power sources. In the event of a loss of power on one source, an ATS PDU can switch automatically to the back-up power source.
- **Hot-swap PDU:** Like an ATS PDU, this PDU type connects to two different power sources so that, when needed, a manual switch can change between the input power sources without a break in power.
- **Dual circuit PDU:** In effect, this type of PDU is equivalent to having two PDUs in one unit, which saves on space and provides redundancy, and supports devices with dual power supplies.

PDU ratings

In the US, the input power source to a PDU is generally either a single or a three-phase 208 V. Without going into too much detail, the difference between a three-phase and a single-phase electrical service is the number of cable wires supplying power. A single-phase circuit uses one of its two *hot* 120V wires and a neutral to provide a single 120 V charge, as shown in the following diagram. A three-phase cable has three hot 120V wires and a neutral wire. As shown in the following diagram, any single wire and the neutral can provide the equivalent of a single-phase circuit. More commonly, two of the hot wires combine to provide 208V:



A comparison of the structures of single-and three-phase circuits

In larger data centers and server rooms, three-phase power is the most commonly used. Single-phase power is typical in a home or small office environment.

The capabilities of a PDU fall into two general classifications—power ratings and load capacity:

- **Power rating:** Power ratings are in amperes, volts, and watts. These ratings indicate the characteristics of the electrical power provided by the PDU. Amperes indicate the strength of the electric current available at each of the PDU's outlets. Voltage measures the pressure required to move the amperes through a circuit, cable, or wire. In the US, voltage ratings are either 110V or 120V. The watts rating specifies the electrical power available to meet the demands of the device connected to the PDU.

- **Load capacity:** A PDU should match the power requirements of the devices it is to service. The load capacity of a PDU is the maximum amount of power a PDU can support. For example, a single rack-mounted server may require an operating range of 600 to 900 watts of power, which translates into a range of 5 to 7.5A. It's recommended that the demand on a PDU outlet not exceed more than 80 percent of its capacity. This means that the load on a 20-amp outlet shouldn't exceed 16 amps. See the following table for an illustration of these calculations:

Device	Nominal power draw	Minimum draw (- 20%)	Maximum draw (+ 20%)	Amperage (W/V)
1U server	750W	600W	900W	5A-7.5A
4U server	800W	640W	960W	5.3A-8A
Router	660W	530W	800W	4.5A-6.6A
Firewall	275W	250W	300W	2.1A-2.5A

Examples of power requirements for rack-mounted devices

Physical safety issues

Data centers and server rooms must be secure and safe environments and workspaces. There are a variety of guidelines and standards that list the requirements for a creating a functional, safe, and secure area. The following are space and safety requirements for server rooms and data centers:

- **Space:** The area or room that houses the servers, internetworking devices, UPS, PDU, and possibly cooling equipment, distribution panels, and cabling, should provide clearances between or among the equipment to allow for maintenance, servicing, installation or removal, and general passage. The aisles between devices, racks, or cabinets should be a minimum of 3.2 feet (1 meter), but racks and cabinets should have 4 feet to the front and 3 feet to the back of minimum clearance. Where one row of racks or cabinets is behind another, the front and back clearances should all be 4 feet. Entrance and exit access should be in accordance with local ordinances.
- **Lighting:** In a server room or data center of any size, all electronic equipment should be well lighted for safety and to provide a good working area. Lighting should illuminate both the fronts and backs of racks and cabinets, as well as the devices inside the unit. Emergency lighting should be in place to provide for a safe exit in the event of a power outage.

- **Electrical safety:** The electrical density of a server room or data center shouldn't exceed 300 watts per square foot of floorspace, on average. PDU and UPS units in the server room shouldn't exceed a rating of 100 kVA and, if larger, should be in a separate space.
- **Fire safety:** An enclosed server room or data center should have a fire suppression system and all cabling in the space should have a fire-rating or plenum. Fire extinguishers should be readily available.
- **Security:** Only authorized employees or vendors should have access to the server room or data center. Log file entries record entries into and exits from the space. As an emergency precaution, this space should also have at least one communication device for making contact outside the space.

Larger server rooms and data centers typically install raised flooring if for no other reason than to run the cabling for the installed devices under the floor surface. However, a raised floor, installed on a grid frame and filled in with flooring tiles, can't support the same weight as a concrete slab floor. Raised flooring, for all intents and purposes, is just as permanent as the concrete slab, which means that, as the enterprise grows and the computing equipment expands, the weight on the floor tends to increase.

There are three primary load types (weight measures) for raised flooring:

- **Point load:** This is the weight of an object on any one of its supports. For example, if a rack cabinet sits on four caster wheels, the point load for that cabinet is the weight on any one of the wheels. The flooring must be able to support the point load of the heaviest device or cabinet that will occupy the space.
- **Static load:** This is the combined total of all of the point loads on a specific tile. If two (or more) cabinets have a leg or wheel on one floor tiles, the static load is the total of the point loads for those units.
- **Rolling load:** Typically, this factor applies to perforated floor tiles and represents the weight of a piece of equipment or a rack cabinet moving over a tile. The rolling load on a tile is the total of two point loads of the device.

Summary

MFA combines two or more factors to authenticate a user. MFA uses five authentication factors: something you know, something you have, something you are, somewhere you are, and something you do.

Physical security is based on an organization's security policies and plans. A physical security program should address the following—physical safeguards, vulnerabilities and threats, monitoring systems, service and utility interruptions, and unauthorized entry, vandalism, or theft. Physical security should focus on deterrence to prevent security events and detection measures to discover events threatening the organization. Threats have four general categories—environmental, man-made, site-specific, and technical.

Environmental controls are a part of physical security. Server and networking devices can have a variety of cooling requirements, such as cold rows or hot rows. Environmental monitoring system should detect and report: temperature, humidity, voltage and power, physical intrusion, and smoke and chemical fumes.

A physical security plan includes UPS units to protect computers and peripherals. A UPS passes utility power to AC outlets and charges an internal DC battery. An offline UPS passes AC mains power through a bypass circuit to its AC outlets, while also charging its battery. An online UPS converts AC power into DC then converts the DC into AC power. The ratings of UPS units are in VA, watts, capacity, and runtime.

Many rack-mounted devices require an external power source, which is commonly a PDU. A PDU has multiple AC outlets and a power cord that connects to an AC outlet. PDUs are either floor-mounted or rack-mounted. PDUs are available in seven configurations: standard, metered, monitored, switched, automatic transfer switch, hot-swap, and dual circuit. PDU ratings also address power ratings and load capacity.

Data centers and server rooms must be secure and safe environments and workspaces. The guidelines and standards for creating a functional, safe, and secure area are as follows: space, lighting, electrical safety, fire safety, and security. Server rooms commonly have raised flooring, which has three load types—point load, static load, and rolling load.

Questions

1. What is the method that uses multiple authentication factors to verify a supplicant?
 1. SFA
 2. CHAP
 3. DHCP
 4. MFA

2. Which of the following is not a type of factor used in authentication?
 1. Something you know
 2. Something you have
 3. Something about you
 4. Something you do
3. What is the security area that prevents unauthorized entry, access, and malicious actions to a private area?
 1. Logical security
 2. Physical security
 3. Intrusion detection
 4. Mantrap
4. Which of the following is not a general threat category to a private area?
 1. Cyber
 2. Environmental
 3. Technical
 4. Man-made
5. The CompuGood Company has a large server room with several rows of rack-mounted servers, power units, and network attached storage. It has arranged the devices in rows so that the HVAC system provides the right amount of cooling to each row. What is the environmental control approach in use in its server room?
 1. Zone conditioning
 2. Compartmentalization
 3. Cold row/hot row
 4. On row/off row
6. Which of the following is not typically scanned or reported included in an environmental monitoring system?
 1. Temperature
 2. Logical intrusion
 3. Smoke and chemical fumes
 4. Humidity

7. What device can electrically protect computers and peripheral devices and provide electrical power in the event of a power outage?
 1. PDU
 2. HVAC
 3. UPS
 4. PSU
8. What device provides electrical power distribution to rack-mounted devices through multiple AC outlets?
 1. PDU
 2. HVAC
 3. UPS
 4. PSU
9. Which of the following is not a type of configuration for a PDU?
 1. Standard
 2. Cold-swap
 3. Dual circuit
 4. Switched
10. Raised flooring in a server room should meet three primary load types. Which of the following is not a load type rating for a raised floor?
 1. Broad load
 2. Point load
 3. Static load
 4. Rolling load

13

Logical Security

Physical security (see Chapter 12, *Physical Security and Environmental Controls*) protects an organization and its computing assets and resources from damage, destruction, and theft. Logical security is not the opposite of physical security. Rather, it's another part of security as a whole. Logical security has the same overall objectives as physical security—to protect, prevent, detect, and deter intrusions against a very valuable resource for any organization.

In this chapter, we will look at the various methods, procedures, and technology used to implement logical security policies to safeguard the data, software, and computing assets of an organization. We will cover the following topics:

- Access control
- Data encryption
- Data retention and disposal
- Physically destroying a disk drive
- OS, system, application, and hardware hardening

Access control

Earlier, we talked about access control on routers (see Chapter 11, *Security Systems and Protocols*) and the use of **access control lists (ACLs)**. However, the access control we talk about in the context of logical security, while different in structure and procedure, has the same overall objective—preventing unauthorized access to the software and data assets of a system.

Access control is a method of logical security that identifies what resources an individual user may access, modify, create, or remove. Using access controls, a system administrator can allow or restrict access to certain system assets and resources on an individual or group basis. Each authenticated and authorized user or group is assigned a **security identifier (SID)**, which is the identity of a system record that details the access permissions and rights of a user or group; that is, their security principals.

When a user, as an individual or as a member of a group, requests access to a system resource or asset, the SID record informs the **operating system (OS)** of the access permissions and rights that are granted to that user. Typically, the requested access comes in the form of an attempt to read or write the resource. In response, the OS scans the corresponding SID to determine whether it can permit the requested access.

Access control criteria

Granting access to system resources may be based on several different criteria. While access permissions and rights may be solely based on who the user is, the administration of granting access to individual users—a process called **discretionary access control (DAC)**—may be okay in a small organization. However, in a larger enterprise, especially one spread out geographically, the criteria that's used to assign permissions could come from a variety of identification factors, such as the following:

- **Group access control:** Group access defines a common set of rights and permissions for individual users that have been assigned to the group by the system administrators. Group accounts are typically generic in nature, such as a group account for general users or the accounting department. Group accounts and group access control allow administrators to implement an access control policy to groups of users that require the same permission sets.
- **Role-based access control (RBAC):** What a user does—that is, what his or her title, job, task, duties, or responsibilities may be—can define the permissions and rights the individual needs to perform or fulfill that assignment. Often, RBAC defines groups of users who perform essentially the same tasks and require the same resources.

- **Specific criteria access control:** Especially on geographically dispersed networks, certain criteria may define the access criteria of a group account. Some of the more common criteria are as follows:
 - **Location:** A location access control could restrict access to users based on their location, such as a branch office or remote manufacturing site.
 - **Time:** Here, access can occur only on certain days and only at certain times of the day. For example, a company policy may restrict access to financial data to only Monday through Thursday and only between 10:00 AM and 2:00 PM each day, and perhaps only to the users in the finance group.
 - **Transaction type:** On an online transaction processing system, access to the supporting database elements is specific to each of the various transaction types. Permissions may also be associated with the user's identity or a group account.
 - **Application:** Access control can also define permissions and rights based on online or offline application types and purposes. Typically, the rights and permissions of the logged in user extend to any application the user executes. This control, the **user account control (UAC)**, if disabled, extends the permissions of applications beyond a user's rights.

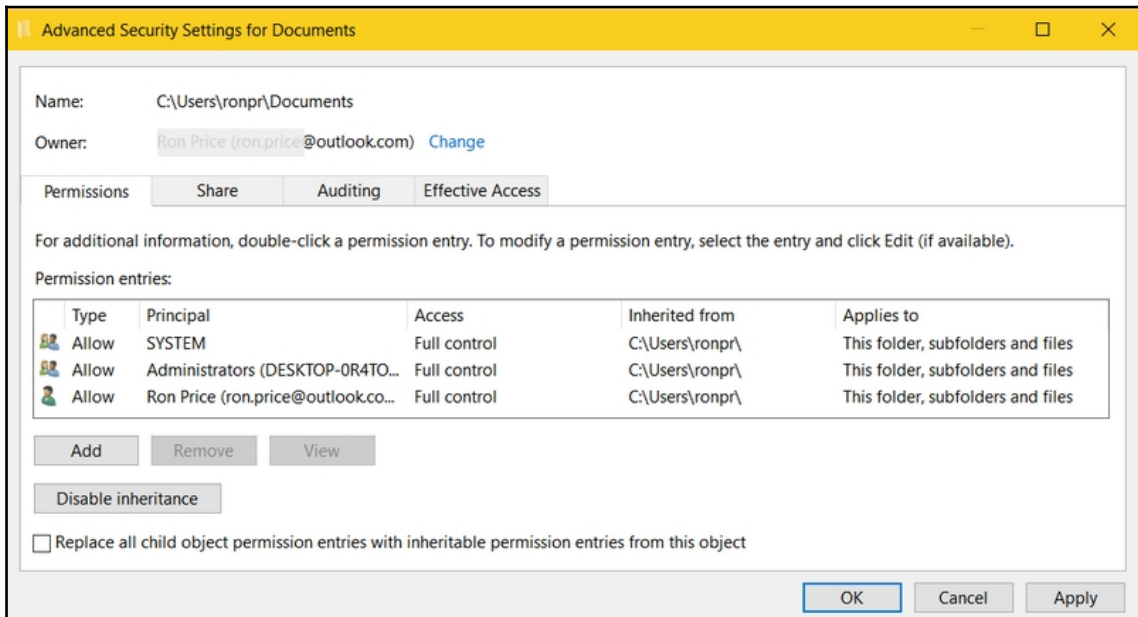
Access control levels

In addition to applying access control to users in various contexts, access control can also be applied to a system's individual hardware and software components and for the same purposes. Access control mechanisms can restrict access and the use of software and hardware components.

Filesystem access control

Beyond the permissions that are assigned to individual directories, folders, and files, filesystems, such as Windows NTFS, Unix POSIX, and Linux NFS, have an ACL that specifies user and group rights and permissions to the programs and files within the filesystem. Each item in the filesystem has an **access control entry (ACE)** in the filesystem ACL that defines the permissions associated with each individual item.

The permissions in an ACE are very much like those of individual filesystem elements, in that they specify who has access rights and what each group or user may do with an element. The following screenshot shows the Windows **Advanced Security Settings for Documents** dialog box on which permissions and rights can be assigned to groups and users on folders and files:



The Windows Advanced Security Settings is used to set permissions and rights for folders and files

The following screenshot shows the contents of the ACL for the `/usr` directory on a Linux system:

```
# getfacl /usr
getfacl: Removing leading '/' from absolute path names
# file: usr
# owner: root
# group: root
user::rwx
group::r-x
other::r-x
..
```

The `getfacl` command displays the permissions and rights for a directory or file on a Linux system

As illustrated in the preceding screenshot, the permissions and rights of this directory belong to three distinct levels of access—file owner, group owner (as in, the group of the file owner), and all other users, which in this case are `root`, `root`, and `other`. Characters indicating read (`r`), write (`w`), and execute (`x`) represent the permissions for each level. The user/owner of the `/usr` directory in the preceding screenshot has read, write, and execute permissions. Write permission includes both modifying and deleting. The `root` group and all other users and groups only have read and execute permissions, which means that they can open, scan, display, and, if the file is a program or script, run it.

Access control to peripherals

Controlling access to internal or peripheral devices of servers, and other networking devices, such as a router or switch, must be a vital part of any network security policy. Modern peripheral devices are essentially computers themselves, which makes them potential targets for attacks on a network. Some ways to limit unauthorized access to these devices are as follows:

- **Administrator passwords:** All peripheral devices include a configuration process that sets their performance and security settings, including an administrator-level password. Setting this password helps to prevent unauthorized access to a device's configuration.
- **Apply patches and firmware upgrades:** Most internal and peripheral devices have associated software that provides an interface to the computer or operating system. Keeping this software up-to-date can ensure that the security features are up-to-date as well.
- **Disable features not in use:** Just like any system, computer devices and peripherals can include features that are not in use. Disabling these features hardens the devices.
- **Limit access control:** These include any access capabilities that allow users to login to modify settings and configuration. Only administrators should have access to these, as well as print servers, network nodes that print directly to a device, and administrator workstations that manage the device and its network.
- **Isolate devices:** Peripheral devices should be a part of an internal network to which external access is not available.

Administration access control

The definition and specification of who has access to an organization's servers and networks and what they can do should be in a security or access control policy. This policy should detail the access rights and privileges of the various user and administrator groups. Administrator access control should restrict access to only those technicians who install, configure, manage, and monitor the software or hardware that performs access control on system resources. Even within the general group of system administrators, access can vary according to responsibility, position, skills, and knowledge.

Security and distribution groups

The Windows OS provides for two types of group accounts in its **Active Directory (AD)** services. A new group must have a group type, either a security group or a distribution group. A security group allows you to assign access control and file and folder rights and permissions. A distribution group is only to be used with distributed email messages. Security groups typically reflect an organization's group policies and allow for flexibility in both user and hardware security settings. On the other hand, a distribution group is strictly for distributing messages and not for security purposes.

Network access control (NAC)

NAC is a software or integrated device that does just what its name implies—it controls access to a network. NAC implements an organization's security policy, enforces logical compliance, and the administration of a network's access control rules. NAC applies to wired and wireless networks, as well as cloud computing.

NAC basically allows only authenticated devices and notes to gain access to a network's resources and monitors any access to the protected devices. Because NAC is made up of policies, procedures, utility, and application software, it's able to restrict user and device access and activity to the network-attached devices and resources. NAC covers all attached devices of a network.

Data encryption

Data or file encryption is the application of a cryptographic algorithm to the data or file so that when it's stored or transmitted, only those users or addressees with the appropriate access can open, execute, or apply its contents. Encryption and cryptography are expansive terms that cover a variety of approaches, applications, and methods to securing data and file assets. For the remainder of this discussion, we'll use the term encryption to cover data conversion through the application of cryptography.

At its most basic levels, encryption has three primary objectives—**confidentiality, integrity, and availability (CIA)**. In addition, the CIA triad is usually extended to include authentication, authorization, and non-repudiation. Encryption is a data integrity and confidentiality protection mechanism.

Storage encryption

On a computer system or network, data, including digital media, documents, and programming, can be in one of three states—in use, in transit, and at rest. Data in use has been decrypted and typically copied to memory for a logical operation. Data in transit needs protection from a variety of interception techniques (all of which are bad) as it moves across transmission lines. The same—if not more robust—protection is also a must for sensitive or confidential data that's stored on some form of data storage media, such as a hard disk, a flash drive, or perhaps even magnetic tape.

Encryption that's applied to data at rest on a storage device may have different levels of secured data. Encryption has applications on virtually all levels of a storage device, including an entire hard disk drive and, in some cases, removable media. **Full-disk encryption (FDE)** essentially creates a passageway between the storage medium and main memory. As data passes to the disk for storage, it is encrypted. As data is transferred to be placed in memory, it's decrypted. Virtually, the same encryption/decryption process may protect a disk partition, a filesystem, an individual file, or an individual data element, such as a column, row, or cell in a spreadsheet or database.

Data retention and disposal

Sometimes, securing data means that the data should be removed. When data is no longer required, it typically depends on the type of data, its operational context, its level of security, and any laws or regulations that require certain handling or retention. The following table shows examples of some generic data retention policy items and their retention periods:

Category	Item	Retention period	Retirement date	Disposal method
Organization records	IRS form 1023	Permanent	N/A	N/A
Financial records	Financial statements	Permanent	N/A	N/A
Financial records	Bank deposit	7 years	1/1/2024	Shred
Tax records	Payroll tax withholdings—2015	7 years	4/30/2022	Burn
Personnel records	I-9 forms	5 years after termination	Varies	Shred
Computer-based correspondence	Internal E-mail	12 years	Monthly EOM	Electronic soft wipe

Erasing a disk

Depending on your objective, there are several levels to removing data from a hard disk drive. At the lowest level (the least destructive) is *delete* and at the highest level (the most destructive) is *wipe*. Between these extremes is *erase* and *shred*. Users typically consider these terms (and their actions) to be different names for removing something semi-permanently or forever:

- **Delete:** Typically, a delete action applies to removing one or more files, directories, or folders. However, in comparison to a hard wipe, the delete command is essentially a soft wipe. After deleting objects from the disk, the electrical impulses (bits) that make up the data remain in place. The delete function removes the index pointer to the file's location on the disk and its space is set to available. Before new data overwrites the space, an *undelete* action will restore the file by replacing its index pointer.

- **Erase:** Although users may think that delete and erase are essentially the same thing, erase removes files and folders permanently, so they are gone forever. Erasing data from a disk involves the use of special software tools, such as **Darik's Boot and Nuke (DBAN)** 2.3.0, Eraser, and CBL Data Shredder. These tools overwrite a disk with random characters or binary zeroes (nulls), which permanently obliterates the data that was on the disk.
- **Wipe:** Wiping a disk drive is a bit more destructive than erasing it. When you wipe a storage device, whether it be secondary storage in a computer, on a cell phone, or a flash drive, you not only remove the active data, but any data that may still be recoverable. The software tools that perform an erase will also do a wipe. Some of the disk wipe software utilities include the ability to remotely format, erase, or wipe a hard disk.
- **Shred:** The shred action is basically a wipe, but limited to a single file or folder. Like the action in an erase or wipe, data is written over with random characters or zeroes.

Formatting

Many users believe that formatting removes any data or files on storage media. In fact, formatting a disk accomplishes just about the same as deleting a file—all the effected files and data appear to be gone. Like the deletion of an individual file, the data is still there. However, the formatting command removes all index entries, effectively removing all references to the existing data and marking all storage locations as free. As with a single file deletion, the data on a formatted disk is recoverable.

There are two distinct levels of formatting for a data storage device:

- **Low-level formatting (LLF):** Typically, a hard disk's manufacturer performs this process before installing the storage device in a computer case. LLF places digital sector markers on the disk to map the storage medium into cylinders, tracks, and sectors.
- **High-level formatting (HLF):** After low-level formatting, a disk still needs to be prepared for use by a OS. HLF adds the structures that the OS uses for partitions or logical volumes. Typically, HLF occurs during an OS installation or after installing a new unused hard disk.

So, does formatting a disk delete, or erase the data and files? Well, it depends. The Windows OS provides a utility called **Quick Format**, which is essentially a *soft* wipe of a whole partition or disk. A quick format is just a big delete, which only gets rid of the indexing, which *is* quick. Windows also provides a format function that does what some call a hard wipe because it makes a single pass over the entire disk, thus writing zeroes to all bits.

Physically destroying a disk drive

After wiping a disk at least once, if you want to be absolutely, positively, without-a-doubt sure that the disk and its now forgotten data are gone forever, you should physically destroy the disk drive. There are a number of ways to do this. The following are only a few of the ways that you can use to be confident that the disk will never be read:

- **Remove and destroy the disk platters:** Assuming that the hard disk drive we are looking at stores its data on ferrous oxide coated platters, we can use a star screwdriver to open the drive's case and the collar that holds the disk platters on the spindle. Remove the platters and break them into pieces, burn them, or use whatever method you wish to otherwise destroy them. If you wish, you could also just hammer the platters without removing them. In either case, take the drive case, any loose electronics, and the platter pieces to an electronics recycler.
- **Penetrate the disk drive:** Piercing a drive by drilling holes through it (using a power drill, of course), driving large screws or nails through the case, or penetrating the case with anything that renders the drive unusable, does not necessarily insure that the data is unreadable.
- **Shred the disk drive:** This method works for all types of data storage devices, including internal and external disk drives (HDD or SDD), flash drives, CD-ROM, and DVD-ROM. Of course, for CD-ROM and DVD-ROM, only the recorded media needs to be destroyed. Industrial shredding services will shred these devices into small bits for a small fee, which provides the highest assurance that any data is beyond accessibility.

Hardening

In general, hardening is the action that's taken to increase the security by reducing the vulnerability of a workstation, server, and network from exploitation. There are several levels of hardening, but the ones that you may encounter in the Server+ exam are **OS hardening**, **system hardening**, **application hardening**, and **hardware hardening**.

OS hardening

Securing an OS goes a long way in the process of securing its computer. The process of hardening an OS is slightly different for a client OS than for a server OS. Most of the steps are the same in either case, but let's focus on the hardening of a network server OS. Essentially, hardening an OS involves four major steps: disabling the unnecessary services, disabling unused TCP/UDP ports, installing only software that's needed for the tasks that are supported, and keeping the OS up to date by applying patches and updates.

Each of the major OSes have their own specific hardening practices, but there are some things that are appropriate for any OS. These include the following:

- **Unused software:** That freeware application you downloaded out of curiosity or to solve a one-time purpose may, like just about all software, be a possible way in for a hacker. Remove any unused or unnecessary software and ensure that your policy restricts the ability of users to install non-standard software on their workstations.
- **Group and user accounts:** A common source of harm to a server is caused by users who, through an administrative error, discover that they have access to confidential or private information. All user passwords should adhere to a common password policy. Review the permissions and rights of all groups and user accounts on a regular basis.
- **Patch management:** While it's a good practice to keep an OS up to date with the latest patches, error fixes, service packs, and updates, be sure that a fix is appropriate to your system. This may involve testing and auditing before and after applying the patch.
- **Establish and monitor baselines:** A baseline establishes a standard level of performance or capacity on a server (or client) for its OS and components. An operational baseline that's generated at installation and updated after each modification to the system serves as the norm for comparison of the current performance level. A large variance to the baseline could indicate a possible security event.

System hardening

System hardening, also known as workstation or PC hardening, is the process that improves the security of a network node. There are hardening actions that apply to just about every part of a workstation, including its major hardware and software components and ranging to its overall security. In addition to the hardening steps that are applied to a server, there are some steps that are specific to a network workstation, including the following:

- **User account policies:** Although technically a server-level policy, user account policies and permissions do impact a user's ability to access a network, as well as the software and data on a local host. The password policy should require periodic password change, a specified length, reuse policies, and login-failure lockout criteria. User rights and permissions should also be in line with the security policies and an access control model.
- **Physical security:** Users should know and understand the physical security policies that are applicable to their workstations. This should include the policies on the control of entrances and exits to a building, a work room, or an office or cubicle. A *clean desk* policy enforces the security of physical documents and access to a computer. Users should also be aware of their responsibilities in the event of a catastrophic event.
- **Logging:** Stopping short of a key-logger, record all system and network interactions by a user, including login and logout times, applications and data accessed, and any security-related program or data access attempts, successes, and failures.
- **Automatic patch application:** Fixes and patches that are performed on workstation OSes and applications can be posted automatically.
- **Disable sharing:** Disable file and device sharing to prevent network-based access through a shared system.

Application hardening

Application hardening is also known as application shielding because this process changes the binary code of the application or utility to make it more resistant to tampering.

Application software can be subject to reverse-engineering, intrusion, and in some cases, code insertion. Not only does the hardening process protect the application code itself, but in many cases, it protects the data from being accessed by the application. Common threats to application software include the discovery of a vulnerability, virus and other malware attacks, and possibly theft of the application and data.

Hardware hardening

In addition to operating system and application software hardening, the security of a system or network can be enhanced by applying hardening to certain hardware devices on workstations and the network.

Host hardware hardening

Most of the hardware hardening process on a host system is in the BIOS/UEFI configuration settings. Setting a password to access the configuration settings, disabling **Wake-on-LAN (WOL)**, if available, and setting the boot device sequence to eliminate a remote boot initiation are just three of the settings that can help prevent intrusion of the host system.

Network device hardening

An essential part of securing a network is to secure the network's infrastructure by hardening the network devices. However, network devices from different manufacturers may require slightly different steps to harden their devices. Generally, hardening a router, firewall, switch, bridge, and other devices accomplishes the same goals using roughly the same process. Some of the common steps that are used to harden a network device include the following:

- **Backup configuration:** This stores copies of the current configuration and settings of the device OS in a secure location
- **Disable admin protocols:** Disables any administrative protocols not in use, such as Telnet and **File Transfer Protocol (FTP)**
- **Disable unused or unnecessary services and protocols:** Disables device services and protocols that are not in use, such as any discovery protocols, HTTP, SNMP, and the **Bootstrap Protocol (BOOTP)**
- **Encrypt configuration files:** When transmitting configuration files, apply encryption to safeguard their contents
- **Secure access:** Use a strong password policy for access to the device console and its auxiliary and virtual Terminal interfaces
- **Define security standards:** Define security standards and procedures specifically for network devices and their operation

Endpoint security

Even after hardening a server and network, additional measures can enhance their overall security. When a client connects to a host, the connection creates two terminating points: an entry point and an endpoint. An entry point opens at the client end of the connection, and the server (its interface port) becomes the endpoint. Endpoint security has rapidly become an issue due to the increase of **bring your own device (BYOD)** policies in a growing number of companies. In these environments, as well as several forms of remote access, the security on the endpoint must be focused on the detection and prevention of intrusion and malware.

An **intrusion detection system (IDS)** can be either hardware or software. Regardless, an IDS scans network traffic (inbound or outbound) to identify malware or, a pattern of suspicious behavior, and monitoring system configuration settings for inadvertent changes. An endpoint IDS is a **security information and event management (SIEM)** application that monitors, scans, and gathers information on the endpoint's traffic. Some IDSes allow for performance and event thresholds above which an alert can go to the administrator. On a network, the endpoint and entry point IDS devices are designated as **network-based intrusion detection system (NIDS)** or **host-based intrusion detection system (HIDS)**. A NIDS is an independent device that monitors network traffic before it's passed to a network device. A HIDS can be either hardware or software that's attached to, or running on, a host computer.

There are essentially two types of IDS—**signature** and **heuristic**:

- A signature-based IDS extracts a segment of code from the suspected message, performs an algorithm on it, and then compares it to a database of known malware. If there is a match, the message is deleted, relevant log entries are created, and the sender is typically barred from the network.
- A heuristic IDS, that is, an anomaly-based IDS, compares the actions of the suspected program to a model of acceptable actions. If there is a match, there is the possibility of a false positive (a match, yet the message is valid and not malware) or a false negative (not a match, yet the message is invalid and possibly malware), but otherwise the IDS may block the source address from entering the network.

Summary

Access controls identify the resources an individual user may access, modify, create, or remove. Access controls allow or restrict access to system assets and resources. In larger organizations, the criteria to assign permissions may come from different identification factors, such as a group, a role, a location, a time, day or date, or a transaction type, or the application in use.

Access controls can apply to hardware and software. Filesystem ACLs specify rights and permissions. Each file or object has an ACE that defines its permissions for its owner, group, and other users. Each level may have read, write, and execute permissions. Access control for devices is done through administrator passwords, the application of patches and upgrades, disabling unused features, limiting access, and by isolating devices.

Encryption applies a cryptographic algorithm to a data or file to restrict its contents to those with appropriate access. Encryption and security have three primary objectives: confidentiality, integrity, and availability. CIA extends to include authentication, authorization, and non-repudiation. Data can be in use, in transit, or at rest.

There are several ways to remove data from a disk drive—delete, erase, wipe, and shred.

Formatting removes all index entries, but the data itself is still on the disk. Data on a formatted disk is recoverable. Low-level formatting places digital sector markers on the disk to map the storage medium into cylinders, tracks, and sectors. High-level formatting adds the structures that an OS uses for partitions or logical volumes. There are several ways to destroy a disk drive: removing and destroying the platters, penetrating the drive, or shredding the drive, among others.

Hardening increases the security of a system or device by reducing its vulnerability. The different types of hardening are OS hardening, system hardening, application hardening, and hardware hardening. Hardening an OS involves disabling unneeded services and unused TCP/UDP ports, installing only the required software, and applying patches and updates. In addition, hardening may include removing unused software, auditing group and user permissions, exercising patch management, and establishing and monitoring baselines.

System hardening improves the security of a network node. This can include enforcing user account policies, implementing a strong password policy, adhering to physical security policies, activating system logging, and disabling file and device sharing. Application hardening changes the binary code of the application to make it resistant to tampering, reverse-engineering, intrusion, and code insertion. Hardware hardening on a host system is done through BIOS/UEFI configuration settings, including setting a BIOS/UEFI password, disabling WoL, and setting the boot sequence to eliminate a remote boot. Hardening network devices includes backing up device configurations, disabling unused administrative protocols and services, encrypting configuration files, and securing access.

IDS is a piece of hardware or software that scans network traffic for malware or a pattern of suspicious behavior. An endpoint IDS monitors, scans, and gathers information on the endpoint's traffic. A NIDS monitors network traffic before a network device. A HIDS is a piece of hardware or software on a host computer. The two types of IDS are signature and heuristic.

Questions

1. Which of the following is not a commonly used identification factor of access control?
 1. Group
 2. Role or assignment
 3. Application type
 4. Computer name
2. The permissions and rights assigned to a specific user or group are defined in what element?
 1. ACL
 2. ACX
 3. ACE
 4. ACR
3. The application of a cryptographic algorithm to data for the purpose of restricting access to its content is what?
 1. Confidentiality
 2. Authentication
 3. Authorization
 4. Encryption

-
4. Which of the following is a method for removing data from a disk drive?
 1. Delete
 2. Wipe
 3. Erase
 4. Shred
 5. All the above
 6. None of the above
 5. Which of the following describes the status of any data on a disk drive after formatting?
 1. Unrecoverable
 2. Destroyed
 3. Recoverable
 4. Unreadable
 6. What is the objective of hardening a system or device?
 1. Extending service life
 2. Strengthening its case
 3. Reducing vulnerability
 4. Taking it offline
 7. Application hardening modifies the binary code of an application:
 1. True
 2. False
 8. Setting a BIOS/UEFI password and disabling WoL are examples of actions of which of the following?
 1. Application hardening
 2. OS hardening
 3. Network hardening
 4. Hardware hardening
 9. When an IDS compares content in a message to a database of identity codes, it is performing what action?
 1. Heuristic scanning
 2. Signature matching
 3. Binary matching
 4. Data scanning
-

10. Which software or hardware device can control access on a network by implementing an organization's security policy?
1. NAC
 2. SAC
 3. UPS
 4. IPSec

4

Section 4: Troubleshooting

This part of the book covers troubleshooting procedures and methods in general and specifically for hardware, software, networks, storage devices, and security applications.

The following chapters are included in this section:

- Chapter 14, Troubleshooting Methods
- Chapter 15, Common Hardware Issues
- Chapter 16, Common Software Issues
- Chapter 17, Common Network Issues
- Chapter 18, Common Storage Issues
- Chapter 19, Common Security Issues

14

Troubleshooting Methods

In this chapter, we'll review standard troubleshooting methods and procedures and common server issues that are caused by hardware, software, network, storage devices, and security systems.

In this chapter, we will review the general procedures and methods for identifying, isolating, recreating, applying, and testing a failure or fault in any general hardware, software, network, storage device, and security policies and devices. The following five chapters in this part of this book cover the specific troubleshooting processes for each of these subsystems, respectively. The topics in this chapter are a review of the troubleshooting procedures that apply to any troubleshooting situation, which are as follows:

- Troubleshooting steps
- Identifying the problem
- Establishing and testing a theory of cause
- Developing a plan of action to resolve a problem
- Documenting findings and outcomes

Troubleshooting steps

Troubleshooting, in the context of computers, servers, and networking, is a structured process that's used to identify, isolate, and resolve a performance issue or a malfunctioning component. Regardless of the size, purpose, use, or location of the computer, its problems are important to its users, operators, and stakeholders. The computer repair technician's responsibilities are simple, actually—identify the problem, resolve the problem, and document both actions. Yes, there are a few more steps to this process, but they're mostly to verify assumptions and test the solution.

While there's no real standard for the steps that are using in a troubleshooting process, the troubleshooting steps that are detailed in the Server+ exam's objectives are as follows:

1. Identify the problem and determine the scope
2. Establish a theory of probable cause
3. Test the theory to determine cause
4. Establish a plan of action to resolve the problem and notify impacted users
5. Implement the solution or escalate as appropriate
6. Verify full system functionality and, if applicable, implement preventive measures
7. Document findings, actions, and outcomes throughout the process

Understand that the application of these steps may or may not be necessary in every troubleshooting situation. For example, the failure of a common component is easy to identify, replace, test, and document. In other less obvious instances, each troubleshooting step is necessary and important.

The technician's knowledge of computer systems is an important tool for troubleshooting a computer problem. However, the technician's interpersonal skills can be as important, if not more important. The ability to interact with a frustrated or flustered end user, a supervisor, or a general stakeholder to learn more about the nature of the problem is an essential skill for a computer repair technician. In addition, the technician's work habits and ethics are on display throughout the troubleshooting and repair process. The work that's performed should be efficient, effective, and correct.

Identifying the problem

Identifying the actual cause of a problem is the first step of a troubleshooting process. However, the first step of that first step should be verifying that there's a problem. Commonly, problems occur not long, if not immediately, after changes are made to the hardware or software. Or the user or stakeholder thinks that something happened, but can't recreate what they think they saw, heard, or smelled. To verify that a problem is real, recreate it—if possible—or take the time to question the user with open questions (ones that cannot have just yes or no as answers). Remember that a user, supervisor, or stakeholders don't usually have your knowledge or understanding of computer systems.

After you're sure that there is a problem, regardless of whether it's the one reported or another cause altogether, start the process of identifying the source, cause, and location of the problem. Most PC hardware problems are relatively obvious, but what can appear to be a hardware problem could very well be a software issue. Or the symptoms may relate to problems that are possible on several components, or one problem is showing up as several different problems.

Regardless of the source of a problem, if possible, you should back up the hard disk and document an inventory of the installed components, if for no other reason than you will be able to restore a problem-free system back to where you started.

Hardware or software?

While this may seem like a simple question, some problems may be difficult to pin down based only on its symptoms. The troubleshooting approach and possibly the tools you use are different for these two areas. If you clearly can't make a decision on the source of the problem, draw on your knowledge and experience to begin troubleshooting what you believe is the issue.

Hardware problems

To determine whether a reported problem is a hardware issue or what appears to be a software issue that's really a hardware issue (for example, the **Blue Screen of Death (BSOD)**), a systematic approach is best. In fact, hardware issues are often related to loose connectors or jumpers, dirty fans and components, or a hardware component malfunctioning due to a software error. When troubleshooting a hardware issue, take care of any obvious issues first, regardless of whether you believe they have anything to do with the issue. A PC that isn't clean is much easier to work with when it's clean; dust can conduct electricity.

Common areas of hardware issues include the following:

- Main memory
- Power supply
- Cooling fan
- Hard Disk Drive (HDD)
- Video or graphics cards
- Expansion cards, slots, and ports
- Cable connections

Software problems

Software problems are much more frequent than issues with hardware components. The software on a computer is one of three general types: system software (for example, an operating system), application software (for example, LibreOffice), and utility software (for example, device drivers). Each software type can have its own set of problems. Operating system problems are typically the result of improper or erroneous configuration, settings, or compatibility issues. An operating system problem may require a minor adjustment or a complete re-installation. Application software issues may have some of the same problems, but generally a serious problem can be resolved through an update, patch, or correcting the registry. Utility software issues are primarily version issues and can be easily solved with a replacement installation.

Some causes of common software issues are as follows:

- Missing or corrupted system or configuration files
- Malware attacks
- Improperly installed or removed software
- Corrupted filesystems or registries

Establishing a probable cause

Identifying a problem is one thing, often obvious, but sometimes identifying the underlying cause can be more problematic. Establishing a probable cause is a process that identifies the possible causes of a problem, analyzes their probability of being a cause, and, in the end, identifying one or more causes to investigate further. The overall objective of this process is to identify the probable root causes of the problem. List the probable causes with the simpler or more obvious issues at the top and the more complex and less likely at the bottom.

Working from the prioritized list, top to bottom, begin testing each one to either prove or disprove the probable causes you identified. The possible causes at the top of the list should be the simplest to test. If one of the probabilities proves to be the cause of the problem, the next step is to develop a plan to resolve the issue. However, if after testing each of the probable causes in the list of possibilities none proved to be the exact cause of the problem, you should reconsider the problem and, if possible, develop another list of probable causes. If you exhaust all possible probable causes—at least the ones you can come up with—it may be necessary to escalate the investigation to a more skilled technician or perhaps the manufacturer.

Define a plan of action

In a vast majority of cases, the plan to resolve the problem is simply one of removing a hardware component, applying a patch to an application, or resetting or re-configuring system settings or parameters. The basic process that's involved is to acquire the replacement part or software (if the resolution requires the replacement of a hardware component), make the workspace free of static electricity to prevent electrostatic damage to the internal components, gather the appropriate tools, make the required fix, and thoroughly test the revised system completely.

If you're resolving more than one issue, affect the repair and test each one separately. As you proceed, each subsequent test step retests each of the earlier fixes, so you want to consider the sequence of your work to be accumulatively compatible. If, after making the change or one of a series of changes, the problem remains or a new problem appears, you should reverse the changes that were made to that point and reconsider your solution. After completing the repair and successfully testing it, notify its user or stakeholder and brief him or her on the cause of the problem, how you identified the cause, what you did to resolve the problem, and how you tested the solution.

Verifying functionality

At this point, you've tested the repair or resolution that was made to alleviate the problem and it's time to fully test the functionality of the full system and, if appropriate, the interaction of the system with any devices with which it interacts. Preferably, the assigned user, a supervisor, or a knowledgeable stakeholder should do this testing. Should this test fail at any level, document the point of failure and restart the entire troubleshooting process, focusing on the new failure.

If the testing is successful and the system functions as it should, determine whether there are any possible preventive measures that could help to avoid the same or similar failures in the future. These preventive measures may be as simple as improvements to the backup procedures or adjusting the **Heating, Ventilating, and Air Conditioning (HVAC)** system, or as severe as planning to replace the entire system as soon as possible.

Documenting findings, actions, and outcomes

Beginning with the initial installation of a network server or host, comprehensive documentation should be as much a part of a system as any of its hardware and software. The documentation of individual computers and network and system-attached devices should record its repair and maintenance history. This record must detail any troubleshooting activities, successful or not, any repairs, upgrades, patches, fixes, additions to, and even removals from, each system. Its importance as a repair and maintenance record comes from the fact that, typically, only one technician performs nearly all computer or system troubleshooting, repair, and maintenance actions.

Complete documentation of the troubleshooting, repairs, and maintenance is more important to the next technician who is going to be working on the system than it may be to you or anyone else. Since the document contains a written record of any reported problems and their resolution, it reduces any guesswork for future technicians. The information that a complete documentation of repair and maintenance actions should include the date, location, and contact at the trouble site; the reported problem/reason for the technician's visit; a description of the troubleshooting performed; a description of the repair made and the identity of any components that were replaced, added, or removed; a description or the identity of any software changes, including updates, configurations, installation, or removals; a detailed description of the testing performed and the results from each test; and any suspected issues not covered in this visit or potential maintenance or repair issues for later attention.

Each of the chapters that follow from Chapter 15, *Common Hardware Issues*, to Chapter 19, *Common Security Issues* focus on a single major subsystem or component category of a computer, its common repair and maintenance issues, and the steps of the troubleshooting process.

Summary

Troubleshooting is a structured process that's used to identify, isolate, and resolve a performance issue or a malfunctioning component. The troubleshooting steps in the Server+ objectives are first, identify the problem and its effects; identify one or more probable causes for the problem; choose the most likely cause of the problem and verify it as the cause; plan out the steps needed to solve the problem; explain the steps to be taken to any effected users; perform the resolution plan; test the corrections and the whole system; and, of course, document every step of this process and your actions. Knowledge of computer systems is important for troubleshooting, but interpersonal skills are just as important. A technician's good work habits and ethics, as well as his or her work should be efficient, effective, and correct.

Identifying the cause of a problem is the first step of troubleshooting, and the first step of that first step is to verify that there's a problem. Frequently, problems show up after changes to hardware or software. The user reports a problem but cannot recreate it. You can verify a problem by recreating it. Most PC hardware problems are obvious. However, what may appear to be a hardware problem may be a software issue. Back up the hard disk and document the installed components.

Common areas of hardware problems include main memory (RAM); power supply; cooling fans; hard disk drive; video or graphics cards; expansion cards, slots, and ports; and cables and connections. Computer software is system software, application software, and utility software. Common software issues include missing or corrupted system or configuration files; malware attacks; improperly installed software; and corrupted filesystems or registry entries.

Establishing a probable cause identifies the possible causes of a problem and their probability. Work from a prioritized list and test each possibility identified. Resolve the problem and test the revised system completely. Fully test the full system and involve the user. Documentation should include a complete record of the troubleshooting and work that was performed.

Questions

1. What diagnostic activity is a structured process that's used to identify and isolate a computer problem?
 1. Maintenance
 2. Analysis
 3. Troubleshooting
 4. System testing
2. What questioning technique is best for gaining information from a user that reported a problem about a computer?
 1. True/false
 2. Closed questioning
 3. Multiple choice
 4. Open questioning
3. In addition to technical knowledge about computer systems, what skill should a computer repair technician possess?
 1. Use of repair tools
 2. Interpersonal skills
 3. Interpersonal skills
 4. Use of diagnostic equipment
4. What's the first step in the process of determining the source or cause of a computer's performance issue?
 1. Identify the problem
 2. Reboot the computer
 3. Shut down the computer
 4. Test a possible solution
5. The best way to verify that a reported problem exists is to do which of the following?
 1. Determine whether the problem is hardware or software
 2. Remove the suspected hardware or software
 3. Recreate the problem
 4. Reboot the system

6. Before starting any work on a computer to isolate and identify a problem, what cautionary action should you take?
 1. Open a Command Prompt window
 2. Back up the entire hard disk
 3. Remove the hard disk drive
 4. Boot the computer from a floppy disk
7. Which of the following is not a general category of computer software?
 1. System software
 2. Entertainment software
 3. Utility software
 4. Application software
8. Which of the following component categories are common hardware problem areas? Choose all that apply.
 1. Main memory
 2. System case
 3. Power supply
 4. External connectors
 5. Cooling system
 6. Graphics card
9. Which of the following are common software issues? Choose all that apply.
 1. Corrupted system files
 2. Improperly installed software
 3. Malware attacks
 4. All of the above
 5. None of the above
10. When performing a full system test after affecting a hardware or software change to a system, which of the following steps should you include?
 1. Develop a test plan
 2. Involve the user
 3. Document the results
 4. All of the above
 5. None of the above

15

Common Hardware Issues

Although most computer problems are software-related and easy to resolve, hardware components can also have issues, which are usually at a minimum disruptive or, at the extreme, catastrophic. Your ability to troubleshoot, resolve, and document hardware issues directly impacts just how disruptive a hardware issue may be. In this chapter, we will look at PC hardware problems that occur relatively frequently. These problems are those that you are likely to see in your role as a certified server administrator.

In this chapter, we will cover the following topics:

- Common hardware problems and their causes
- Environmental issues

Hardware problems

Hardware problems can be a fault in or related to an electrical or electronic component, or possibly the setup, installation, or configuration of a component, which leads to the malfunction or failure of an entire computer. However, what may appear to be a hardware problem can be a software-caused issue. At the risk of being obvious, a major difference between a hardware problem and a software problem is that a hardware problem typically requires you to open or remove the system case, as opposed to simply applying a patch.

Identifying a hardware problem

Some computer problems are difficult to pin down as either software or hardware issues. While there are some that are just obvious, such as smoke coming out of the case, when it comes to others, you just can't tell. Some conditions that can help you identify a problem as an issue with hardware include the following:

- **Intermittent failure:** A problem that occurs irregularly, but in the same way each time, can be either hardware or software-related. If this problem begins immediately (or soon thereafter) after the installation of new software or a patch or upgrade to an existing operating system or application, it's most likely a software issue. However, if the problem begins to occur in a period with no software changes, then you've guessed it—it's probably hardware.
- **Access failure:** If a problem occurs after any use of a component or an I/O operation on a peripheral device, it's very likely that the problem is with the failing hardware. However, there is a chance that the problem could be the result of a software flaw. You should eliminate this issue as a hardware problem before testing for a software problem.
- **Random stopping or rebooting:** If you're working away on what must be a very valuable document, thesis, dissertation, or game and the computer suddenly freezes up and shuts down or arbitrarily reboots itself, then you have a hardware problem.
- **POST failure:** A part of the startup (boot) process is the **power-on self-test (POST)** that verifies the presence and functionality of the devices that were identified in the BIOS/UEFI configuration settings. If the boot process fails, the POST displays error message or codes on the display or sounds a pattern of beeps (representing a specific error condition) on the system speaker. Obviously, this is a hardware issue and typically one that occurs right after the installation of new hardware.

Other symptoms that usually indicate a hardware issue are as follows:

- **BSoD/RSoD:** The blue or red screen of death indicates an error between the operating system and the hardware configuration
- **I/O:** Reads and writes to secondary storage are very slow
- **Data integrity:** Files are inexplicably damaged, corrupted, or missing
- **Muddled display:** The display is missing parts of an image; the display is chaotic or not the expected image or content
- **Unusual noises:** Scraping, grinding, banging, and perhaps beeping noises emanating from the system case

Common problems

Most, but certainly not all, computer hardware problems are relatively easy to diagnose and, in many cases, simple to resolve. Otherwise, *why would a certification exam focused on network servers include troubleshooting and repair information?* As a system administrator, your skill set should include the ability to identify and resolve common hardware issues. The following sections identify failures for several hardware categories or components. Just so you know, each of these issues are very likely to appear on the certification exam in one form or another.

POST failure

As we discussed earlier, a failure during the POST phase of the boot process can cause the startup process to stop. Although POST communicates any problems it detects, the form of that communication varies among BIOS/UEFI versions. Both Windows PCs and macOS computers run a POST process as an initial part of the startup or boot process. The POST verifies the presence and functionality of essential components and devices, such as the **power supply unit (PSU)**, main memory (RAM), and the system bus structure, among others. Should any of these devices not respond to the POST, a pattern of audible beep tones will sound out to alert the user.

While the POST is running, the only output device that's available is the system speaker. For this reason, POST uses a pattern of short or long beep sounds to indicate any problem it has found. These codes are generally unique to the manufacturer of the motherboard or computer. There is no standard set of beep codes and each manufacturer has its own unique set. So, a technician working in an environment with PCs from several manufacturers may need to be multi-beep savvy.

The following table is an example of the use of beep codes with some illustrations of their meanings. As indicated, beep tones are short and long. A short beep is typically one second long and a long beep is two seconds long:

Number and length	Meaning
1 short	Memory refresh fault
2 shorts	POST fault
4 shorts	System timer fault
5 shorts	CPU fault
8 shorts	Video adapter fault
10 shorts	CMOS fault
1 long, 3 shorts	Memory fault
1 long	POST successful

Overheating

A computer overheats for a variety of reasons, but generally, it's caused by neglected preventive maintenance. Dust building up on the air vents and grills and inside the system case can block or restrict the airflow that was designed to cool the heat-emitting components. Heat inside the system case that slowly rises above its safe operating limits can also slowly cause the electronic components on the motherboard and installed devices to deteriorate. The damage comes from a heat buildup caused by the dust creating an insulating layer and trapping heat that can't escape.

Another factor that can contribute to a computer's overheating problem is a combination of its location and the ventilation and airflow it has. A well-ventilated space provides a base environment that provides for the underlying conditions that are assumed in the engineering of the computer's cooling system. Other factors include direct sunlight, closeness to exterior windows, or furniture or objects obstructing the airflow.

The resolution for a computer with an overheating problem, either intermittent or persistent, is to basically reverse the conditions causing the computer to overheat. Clean the inside of the system case with compressed air to blow any layers of dust away from any cooling system vents, fans, heat sinks, and air passages. You can vacuum the dust, but make sure you're using a vacuum cleaner that doesn't create or emit static electricity. In other words, don't use your household vacuum!

If cleaning doesn't resolve the problem, check the location and its access to the ambient ventilation system. Moving the computer out of direct sunlight and away from anything blocking the airflow is your next step. And, if this doesn't solve the problem, you may have a malfunctioning case fan, the heat sink or CPU fan is wrong, or the thermal paste is too much or too little. You'll need to work through these issues one at a time to eliminate them as possible causes.

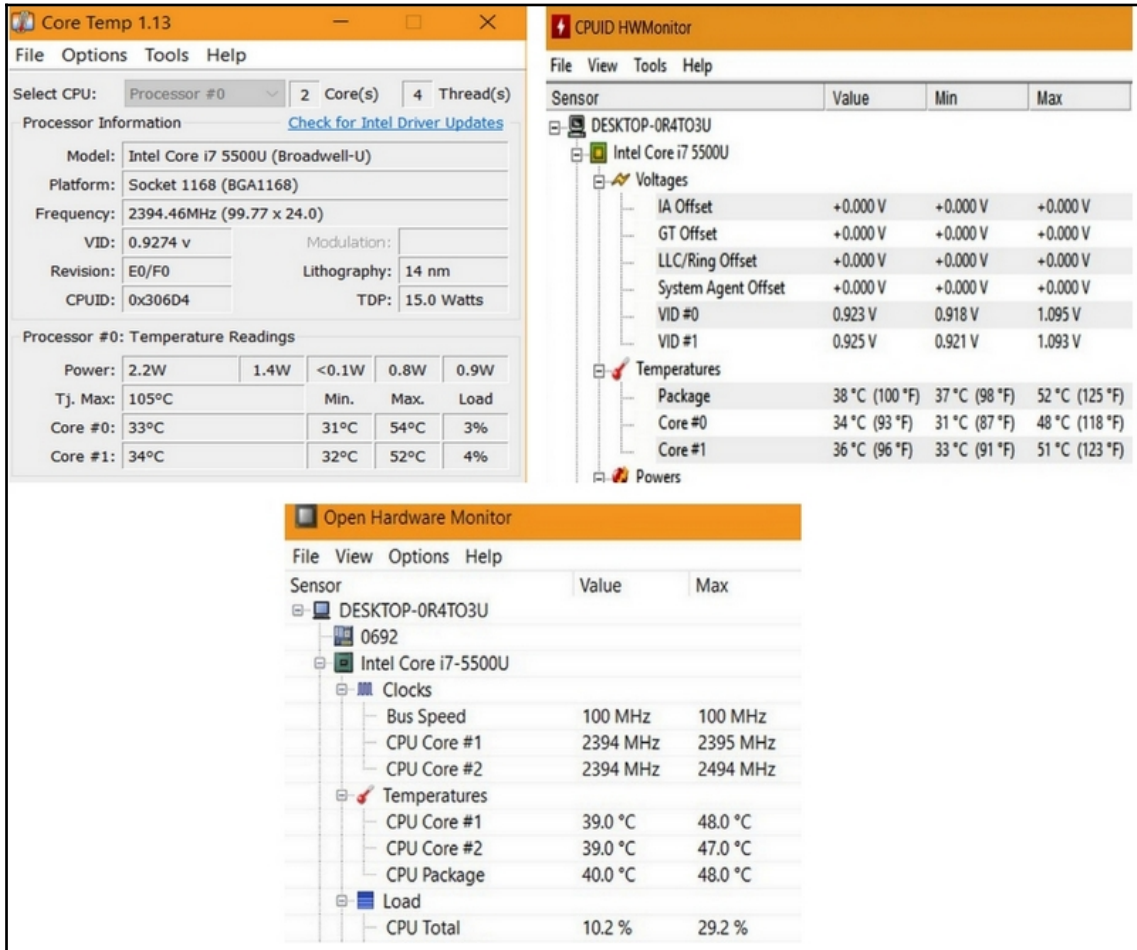
Processor failure

If the microprocessor in a computer fails, it's a very good indication of some other problems—not that it matters at that point. There are only a few reasons why a processor would fail, such as an extreme power surge or lightning strike, bungled attempts to overclock the CPU, or the accumulated damage of overheating. Most processors now include a protection mechanism that slows down or halts it completely if its operating temperature rises too high. While this feature does protect the processor against sudden and significant temperature increases, over the long term, a processor, especially one that's heavily used, can slowly begin to fail.

Monitoring the temperature of the CPU on a new commercially assembled computer can set a reliable baseline. However, if you were to install a new CPU or cooling devices, any unnoticed misalignment or improper application could create a false baseline. Checking the system regularly against this baseline is essentially keeping track of the **time to failure (TTF)**.

So, how do you take the temperature of a CPU? And what should the temperature be? It does vary a bit from processor to processor, but the rule of thumb is that a processor should normally operate in the range of 45° to 50° **Celsius (C)**, which is 113° to 122° **Fahrenheit (F)**. Under a full load, such as playing a game, the CPU should not exceed 75° C (167° F). Some processors can run at a higher temperature, but not for very long before damaging the CPU.

The following screenshot shows captures of three different software utilities—**Core Temp**, **CPUID HWMonitor**, and **Open Hardware Monitor**, clockwise from the upper-left-hand corner. These tools, and others, report several operating measurements for monitoring:



Examples of utility software for monitoring processor and core temperatures

Memory failure

A computer, much like a human, cannot function very well after its memory starts to go. Memory-related errors and failures on a computer aren't necessarily obvious and, in most cases, seem to indicate that other components may be the problem. Common memory problems and issues include the following:

- **Gradual slowdown:** When you first power up the computer, everything zips right along. However, as the day wears on, the computer gets slower and slower to the point that downloading a website or opening a document can take minutes, not seconds.
- **Random restarts:** As you are working, the computer restarts itself intermittently with a random amount of time in between or the computer locks up and the keyboard and mouse no longer seem to work.
- **Corrupt files:** While this often seems like a disk drive issue (and it could be), it could also be a memory issue. Commonly, memory failures can corrupt a single frequently-used file or filesystem to the point of rendering the disk drive unusable.
- **Failed installations:** When you are installing or reinstalling software, including an operating system or major application, the installation process halts, freezes, or displays a random error message that may not have anything to do with the actual problem, which is likely memory.
- **BSoD/RSoD:** The **Blue Screen of Death (BSoD)** or its cousin, red, flash up on the display quickly, telling you there's possibly a problem (usually memory), before continuing with the boot process.

For the most part, the causes of memory problems are essentially the same as those that create problems for the CPU – heat, power surges, electrostatic damage, improper installation, overclocking the CPU, a damaged memory module slot, or a fault in a memory module.

Motherboard and component issues

The bad news is that you think the problem with the primary network server is a motherboard problem. The good news is that the problem may not actually be the motherboard; it just appears to be. Remember that there are components that are permanent to the motherboard and there are other components installed on the motherboard. The add-on components, processors, RAM, cooling, and more, could be the source of the problem, but unfortunately, it just might be the motherboard after all.

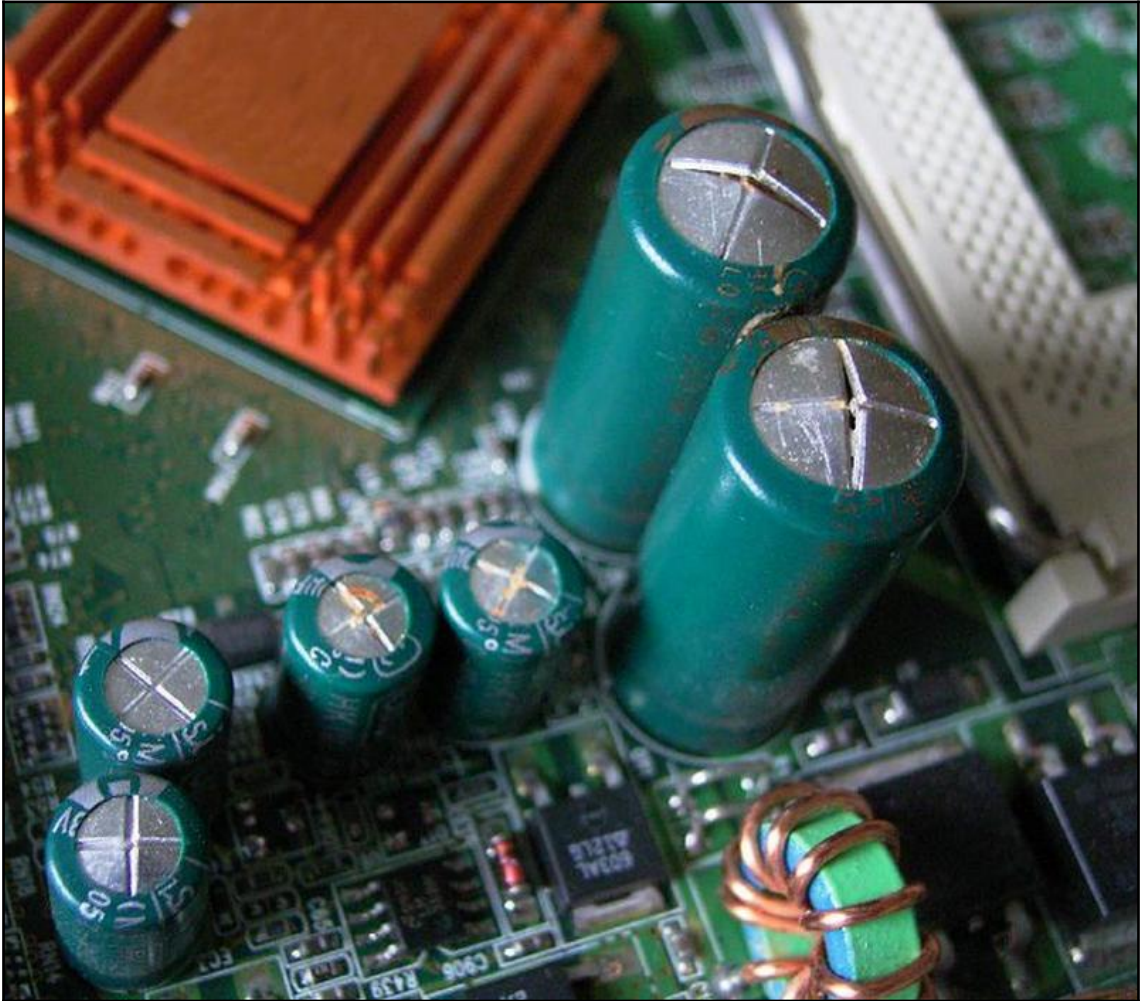
Never mind the fact that the motherboard may be the priciest component in the computer—it is likely the most difficult component to replace. Once you've convinced yourself that there is no doubt that the problem is absolutely, positively the motherboard, proceed with caution. Before removing the old board, label every wire and cable with what it connects to, where it connects to, and its place in the reassembly sequence. You'll be glad you did, trust me.

Rarely will a major component of a computer completely fail suddenly. Generally, it fails over time, giving signs that something may be wrong, and you should probably investigate. The following sections identify a few signs that something may be going wrong with a motherboard or its components, in no particular order of frequency or importance.

Capacitor issues

The three major electronic components on a motherboard are transformers, rectifiers, and electrolytic capacitors. A transformer in a computer's power supply decreases the voltage of the incoming electrical power. The *stepped down* power then passes through a rectifier that converts it from AC to DC. Since AC power fluctuates, the DC power that's produced by the rectifier also fluctuates. A capacitor, also known as an electrolytic capacitor or condenser, stores a static electrical charge. As the DC power on the motherboard fluctuates, a capacitor with **Equivalent Series Resistance (ESR)** provides enough current to *step up* the current.

A possible motherboard problem is the failure of a capacitor. As shown in the following image, a failing capacitor can begin to bulge or push up its top, leak out its electrolytic materials, or even catch fire. A failing capacitor isn't necessarily an emergency, but unregulated DC power can start to damage other components on the motherboard, including the processor:



Capacitors on a motherboard showing bulging caps

Image courtesy: Wikipedia.com (https://commons.wikimedia.org/wiki/File:Defekte_Kondensatoren.jpg)

Burns

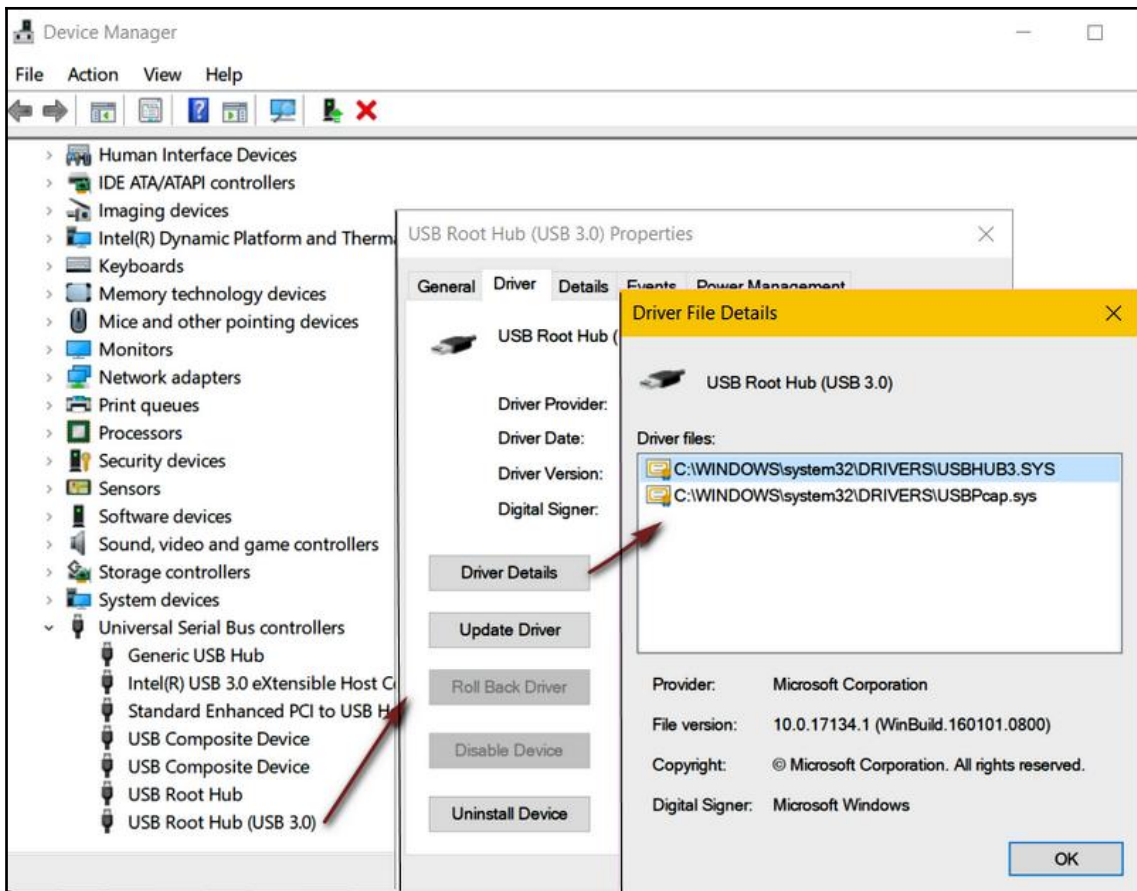
If an electrical component, such as a capacitor, erupts or burns, the result is typically obvious: there is an electrical burn odor (ozone) or signs of burned components or circuits on the motherboard. An electronic component doesn't have to be on fire to make it look like it has burned. An electrical short can cause the same effect.

USB not recognized

You're working on a network server and need to install a file. You plug in a USB flash drive, but you get an error stating **USB Device not recognized: One of the devices attached to this computer has malfunctioned and Windows does not recognize it** appears on the screen. Several issues could be the cause of this error. However, it's best to begin your troubleshooting with the most likely. One process goes like this:

1. Verify that the flash drive works on another port or computer. If it doesn't work there, you know what the problem is.
2. Now that you know that the flash drive is good, check the USB port for visual damage, especially if it's an external device, and replace it, if needed.
3. Verify the device drivers for the USB drive and the USB root hub (refer to the following screenshot) for validity and version (don't worry if there are more root hubs than actual USB ports). If needed, reinstall or update the device drivers.
4. Check for loose cables or connectors or wires that are not seated in a connector. Replace or repair any problems you find.

If any of these conditions appear to be causing the problem, take the appropriate action. However, if the problem doesn't appear to be hardware, it's likely that the problem lies with the operating system, device controller, firmware, or the root hub settings. Apply any operating system updates or patches pending installation. Update the system firmware, especially if the update references the USB or expansion bus. Check if the root hub has placed the USB device in a selective suspend state:



Checking the USB root hub device driver

Expansion bus

The expansion bus, also known as the external bus, provides a connection between an expansion slot (and an expansion card inserted into it) and the system bus. It's common for a motherboard to include multiple expansion bus standards to provide backward compatibility and component flexibility. The expansion buses on most current motherboards are the **Accelerated Graphics Port (AGP)**, **Peripheral Component Interconnect (PCI)**, **PCI Expanded (PCI-X)**, and **PCI Express (PCIe)**. The PCI bus is the most popular and is found on PC and Macintosh motherboards.

A common problem associated with expansion cards and buses is a form of *chip creep*. This condition, also known as power creep, is caused by the temperature cycles in the metal connectors of an expansion slot. Over time, an expansion card can begin to *wiggle* its way out of the expansion slot. To resolve this issue, firmly seat the card into the expansion slot (with the power off, of course).

If an expansion card was without problems when it was installed correctly, a failure or performance problem is likely to show up in the area associated with the card. Expansion cards are specific to one, maybe two, functions, components, or peripheral devices, such as video cards, sound cards, and network interface cards. So, if there is a problem with the card, it generally shows up on the supported device.

Should an expansion card that has been operating as it's meant to, suddenly show signs of problems, the most common causes are heat and power. Overheating can damage components on the card; a bad power supply can wreak havoc; or a major power surge may have affected all the internal electronics. Another cause for a condition that may appear to be an issue with an expansion card can occur when the motherboard and chipset include the same function as an installed expansion card. Removing the card may eliminate the problem.

Moving the card to a different expansion slot of the same type can verify whether the fault is with the card, with the expansion slot, or related to the bus, device driver, or chipset. However, the most common resolution to an expansion card problem is to replace the card with a known-good one.

PSUs

Without a fully functional PSU, there is no computer. The electronics inside or attached to a computer run on a completely different type and level of electricity than any of your other electrical devices. Yes, the computer does plug into the wall outlet that provides it with 110 volts or 220 volts of AC electrical service, but its internal components operate on +/- 3.3 volts, +/- 5 volts, and +/- 12 volts of DC power. The conversion between the voltages of AC from the wall and the DC power for the computer is what the PSU is all about.

Even though the PSU is a major component of a computer and has the highest failure rate, it's a big deal when a PSU starts to fail and an even bigger deal when it fails completely. A power supply problem that's detected early can prevent serious problems with other internal components, which can be much harder to diagnose. The hardest part of troubleshooting a problem that could be related to a faulty power supply is that it could be the cause of just about any electrical problem in the computer, and often is. Some common signs that the PSU may be failing or have failed include the following:

- Unusual noises coming from the back of the system case near the electrical cord's connection
- Turning on the power switch doesn't start the computer, but it does start a flashing light on the front or the back of the computer
- After starting up okay, the computer powers off after a very short time
- While playing a game or using an application, the computer powers off suddenly and without warning, and perhaps displays the BSoD
- The video display for a new game is distorted and the PSU's wattage and amperage are too low for the video card
- The computer starts up, but the hard disk drive and the cooling system do not start up, causing the system to lock up
- Touching the metal system case causes an electrical shock

The remedy for a failing power supply is to replace it, even if only to determine if the PSU is the source of the detected problems.

Hard Disk Drives (HDDs)

Mechanical failure is the most frequent cause of problems in HDDs and even **Solid State Drives (SSDs)**, to an extent. The mechanical internal components of an HDD are also its moving parts and, over time, moving parts can wear out and fail. On an HDD, failure comes with little, if any, warning. The failure horizon of an SSD is much longer. The common warranty on an SSD is ten years and manufacturers claim that an SSD will remain functional for over 300 years. So, let's focus on the HDD and its failure causes.

The frequent causes of HDD problems include the following:

- **Heat:** By now, you've figured out that excessive heat inside a computer is a bad thing and that long periods of high temperatures can damage or destroy key components of the system.
- **Moisture:** It seems logical that an HDD that has been immersed in water may have some problems. However, the problems aren't necessarily inside the hard case around the spindles, heads, and platters, but water and electronic components don't mix well. Powering up a system that may not be completely dry can short out the electronic components, and not just those on the HDD.
- **Electrostatic discharge (ESD):** On the other extreme, extremely dry conditions can produce almost the same damage as very humid or wet conditions. Static electricity can build up and cause components or circuits to fail, including the hard disk.
- **Power surge:** Routine power surges on the AC electrical service can degrade the overall performance of a computer over time, but severe electrical surges, such as a lightning strike, can easily destroy an HDD, along with all the electronics of a computer.
- **Physical damage:** The internal parts of an HDD operate with close tolerances that aren't designed for sudden jolts, drops, or impacts. If a computer, or even just an HDD, is dropped, hit, or impacted with enough force, the drive is likely to fail.

If you believe that a computer problem may be caused by a hard disk drive, you can use the **Self-Monitoring, Analysis, and Reporting Technology (SMART)** utility to display its status. This utility is available on nearly all operating systems, including Windows, Linux, and macOS X. The following screenshot shows the results of a SMART hard disk error 301, which indicates that the HDD is failing and should be immediately backed up:

```
SMART Hard Disk Error

The SMART hard disk check has detected an imminent failure.
To ensure no data loss, please backup the content immediately
and run the Hard Disk Test in System Diagnostics.

Hard Disk 1 (301)

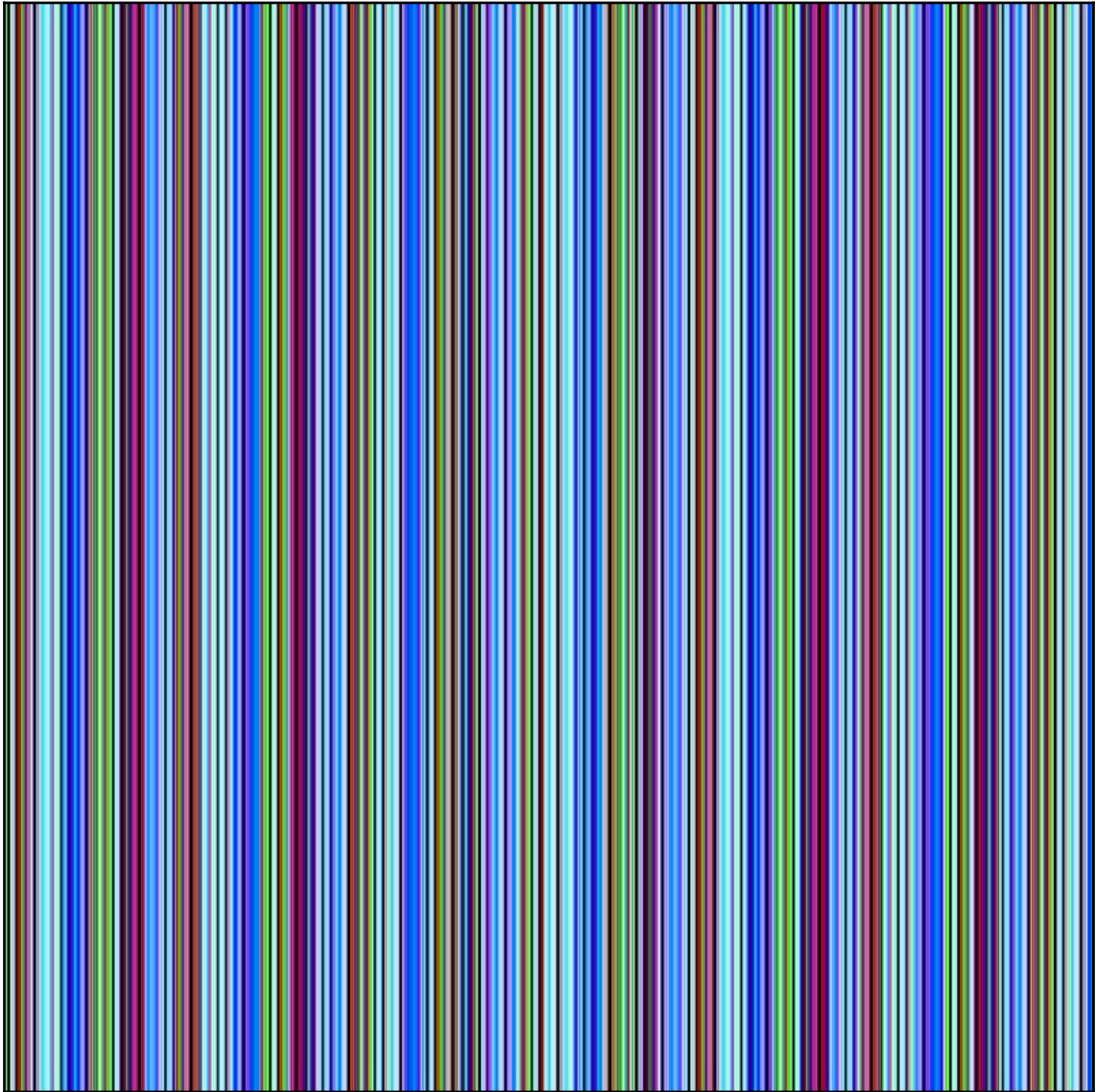
F2 System Diagnostics

ENTER - Continue Startup
```

The display of a SMART hard disk error

Video display

Obviously, if there are vertical or horizontal lines or bars covering a computer's video display, as shown in the following screenshot, there is something wrong. This condition is caused by either hardware or software. Choose one or the other to begin troubleshooting:



Vertical lines and bars on a video display, indicating an issue

If you begin troubleshooting the display as a software problem, use these steps:

1. Visit the website of the display's manufacturer, navigate to the support page, and access the graphics device drivers for your model
2. If there is more than one driver, install each one at a time and after completing them all, restart the computer

If these steps resolved the problem, it was a software issue, most likely caused by a conflict or incompatibility between the operating system and the previous versions of the device drivers.

If you begin by troubleshooting the problem as a hardware issue, or if the software troubleshooting steps didn't resolve the problem, use these steps on systems with detached displays:

1. Reboot the computer and open the BIOS/UEFI configuration system. If the problem is hardware related, booting into the BIOS/UEFI settings isolates the video display from the video drivers and chipset functions that are associated with the operating system and supports the display with only basic drivers.
2. If the display problem persists in the BIOS/UEFI system, it's a hardware problem. Correcting the problem is different for each manufacturer, and sometimes even between a manufacturer's different models. Go to the manufacturer's website to find the procedure to use to resolve the issue.

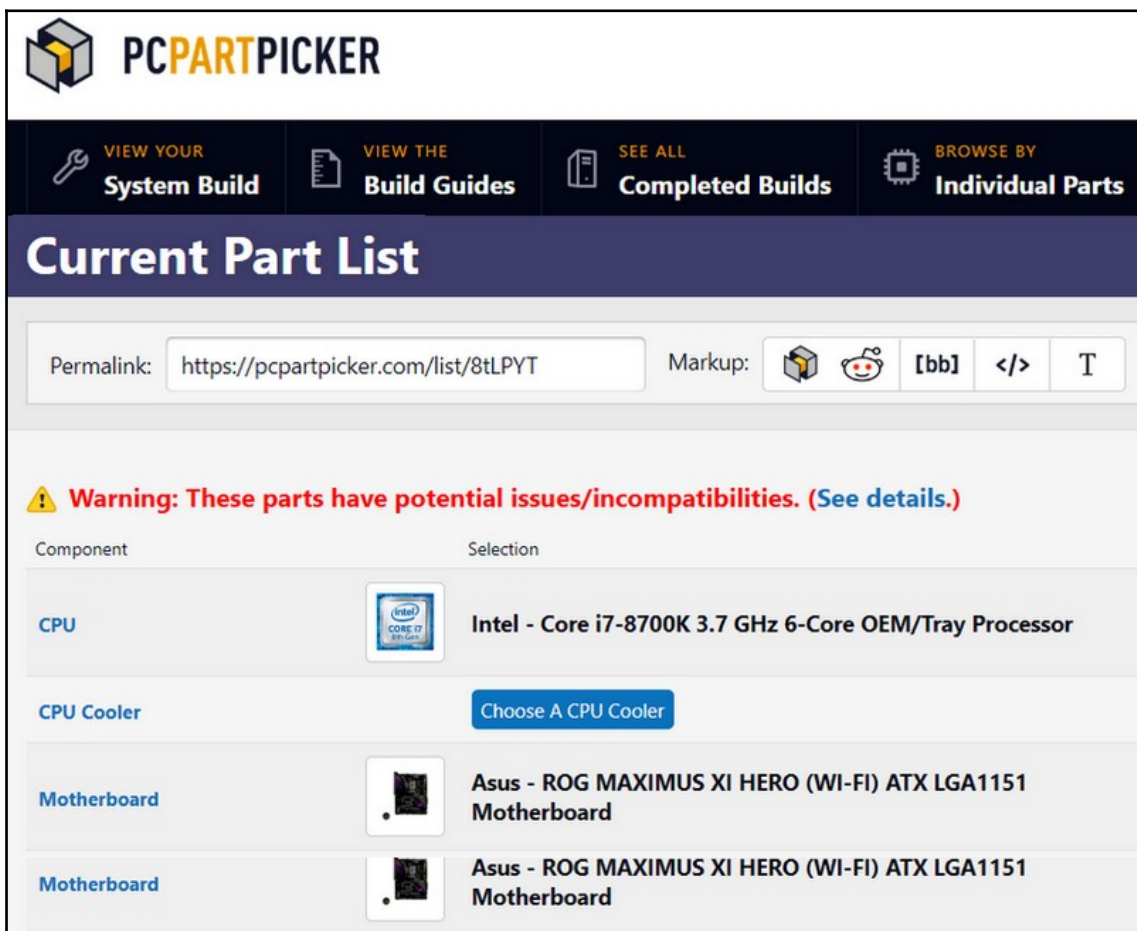
There are a few generic actions you can take, regardless of the make or model of the display. Check the cable to make sure it's connected properly at each end. Verify that the display is connected to the appropriate plug and video standard on the computer. Check the power cord and video connection cable for sharp bends, physical damage, and loose wire connections.

One last thing—if the video problem is on a laptop, notebook, or tablet computer, that is, any computer with an integrated display, it should be taken to a professional repair technician, preferably one representing the manufacturer.

Other common problems

Beyond the issues and causes that we've already discussed, there are other less specific problems that can show up on a computer system. These include the following:

- **Compatible components:** If it becomes necessary to replace a major component in a computer, the replacement's compatibility with the existing or other new or replacement parts is very important for the improved operation of the computer. There are several websites available that can assist you in verifying that components are compatible. The following screenshot shows a sample of one such site, <https://pcpartpicker.com/>:



The screenshot shows the PCPartPicker website interface. At the top, there's a navigation bar with links: "VIEW YOUR System Build", "VIEW THE Build Guides", "SEE ALL Completed Builds", and "BROWSE BY Individual Parts". Below this is a section titled "Current Part List".

Under "Current Part List", there's a permalink field showing "https://pcpartpicker.com/list/8tLPYT" and a markup section with icons for social media and code.

A warning message is displayed: **Warning: These parts have potential issues/incompatibilities. (See details.)**

The part list is organized into two columns: "Component" and "Selection".

Component	Selection
CPU	Intel - Core i7-8700K 3.7 GHz 6-Core OEM/Tray Processor
CPU Cooler	Choose A CPU Cooler
Motherboard	Asus - ROG MAXIMUS XI HERO (WI-FI) ATX LGA1151 Motherboard
Motherboard	Asus - ROG MAXIMUS XI HERO (WI-FI) ATX LGA1151 Motherboard

An example of the compatibility checking on PCPartPicker.com

- **Incompatible BIOS:** When installing or upgrading a Windows system, if the upgraded software stops and displays that Windows and the BIOS on the computer are incompatible, the problem is likely a CPU setting called the **non-execute (NX)** bit. This is a switch setting that controls the use of memory for various functions and blocks a part of the installation from running. To enable the NX bit, open a Command Prompt (with administrator privileges), enter the `exe /set {current} nx AlwaysOn` command, and press the *Enter* key. Shut down (not restart) the computer and then restart it. All should be well now.
- **Server backplane failure:** A backplane is a circuit board with several expansion slots that allows several components to share common bus structures and power sources. A backplane is only as good as its slot connectors and if there is a failure, a slot is left empty so that the backplane can be replaced.

Environmental issues

The environment of a server room or data center should be essentially the same, regardless of its location. A problem exists where the external environment may present one or more extremes in temperature, air quality, humidity, and severe weather conditions. A server room or data center must provide an environment that is conducive to the operations of electronic equipment, but to human administrators as well.

The primary environmental threats to a server room are temperature, humidity, contaminants in the air, and extreme fluctuations in the electrical power source. Here's a bit more on each of these elements:

- **Temperature:** The ambient temperature in a server room should always be between 50° F (10° C) and 82° F (28° C), and optimally between 68° F (20° C) and 71° F (22° C). Computers can operate in an environment with consistent temperatures, regardless of whether they are cold or hot. Problems, such as condensation inside the system case, crop up when the temperature rises or falls rapidly or often.
- **Air quality:** Airborne contaminants, such as dust, insulation, and other particles carried into a server room can, if not controlled or cleaned, accumulate in the cooling systems, main boards, and electronics of computers, storage devices, power supplies, and other devices. Water from humidity is another airborne contaminant that can slowly saturate internal components. Problems such as overheating, electrical shorts, and corrosion can result from this contamination.

- **Electrical power:** Variations in the current of the electrical power service can damage and possibly destroy the equipment in a server room. Power surges, dips, brown-outs, and black-outs can stress equipment and possibly burn out or short out motors, controllers, and other electronic components. Power conditioners and **uninterruptible power supply (UPS)** units, should be employed to level out the electrical service.

Summary

A computer hardware fault in an electronic component is a result of an issue in setup, installation, or configuration that leads to its malfunction or failure. Hardware issues include intermittent failures, access failures, random rebooting, and POST failures. System administrators should have the ability to identify and resolve common hardware issues.

The POST process verifies the presence and function of essential components and devices. Should the devices not respond, a pattern of audible beep tones sound to alert the user. POST uses a pattern of short or long beep sounds to indicate a problem. There is no standard set of beep codes.

Computers overheat commonly because of neglected preventive maintenance. Dust inside the case can block the airflow that's designed to cool components. Heat inside the system case can slowly deteriorate electronic components. The resolution of an overheating problem is to clean vents, fans, heat sinks, and air passages with compressed air.

There are a few reasons a processor fails: an extreme power surge, overclocking, or overheating. Measuring the CPU temperature sets a baseline. Any misalignment could create a false baseline. A processor should operate between 45° to 50° Celsius, which is 113° to 122° Fahrenheit, and not exceed 75° C (167° F).

Common memory problems and issues include gradual slowdowns, random restarts, corrupt files, failed installations, and BSoD. The causes of memory problems are the same as those for the CPU: heat, power surges, electrostatic damage, improper installation, overclocking, damaged memory slot, or faulty memory modules.

The three electronic components of a motherboard are transformers, rectifiers, and electrolytic capacitors. A transformer decreases the voltage of the incoming electrical power, which a rectifier then converts from AC to DC. A capacitor stores a static electrical charge, which the ESR uses to *step up* dips in the current. A failing capacitor bulges its top, leaks out its electrolytic material, or catches fire.

The expansion buses on most current motherboards are AGP, PCI, PCI-X, and PCIe. The PCI bus is the most popular and is found on PC and Macintosh motherboards.

The PSU converts AC of 110 volts or 220 volts of electrical service to DC power of 3.3 volts, 5 volts, and 12 volts of DC power. Common signs a PSU may be failing include strange noises, powering on not starting the computer, the computer powering off, and the HDD and cooling system not start. Mechanical failure is the most frequent cause of problems in HDDs. The causes of HDD problems include heat, moisture, ESD, power surge, and physical damage.

A server room or data center must provide an environment that is conducive to the operations of electronic equipment, but to human administrators as well. The primary environmental threats to a server room are temperature, humidity, contaminants in the air, and extreme fluctuations in the electrical power source.

Questions

1. An essential skill for a system administrator who's responsible for managing and maintaining a group of servers is which of the following?:
 1. Application programming
 2. Identifying common hardware issues
 3. Repairing HVAC system failures
 4. Monitoring environmental conditions
2. Which of the following is not a common computer hardware failure?
 1. Buffer overflow
 2. POST failure
 3. USB device unrecognizable
 4. Access failure
3. The process that runs immediately after powering up and verifies the presence of hardware components and devices is which of the following?
 1. BOOTP
 2. DHCP
 3. POST
 4. AGP

-
4. The audible sounds emitted by the POST process to indicate the source of an error are called which of the following?
 1. Tings
 2. Diffie-Helman
 3. Condition alerts
 4. Beep codes
 5. Which of the following is a common cause of a computer overheating?
 1. Blocked airflow vents
 2. A broken CPU fan
 3. Neglected preventive maintenance
 4. All of the above
 6. What is the temperature range inside a computer's case in which a processor should operate?
 1. 10° C to 45° C
 2. 45° F to 50° F
 3. 113° F to 122° F
 4. 75° C to 85° C
 7. Which of the following is not a common cause of a memory problem?
 1. ESD
 2. Overclocking
 3. A faulty memory module
 4. A bad PCI slot
 8. Which of the following is not an electronic component that's commonly found in a computer?
 1. A transformer
 2. A capacitor
 3. A CMOS
 4. The rectifier
 5. All the above
 6. None of the above

9. The acronym PCI-X stands for which of the following?
 1. Peripheral Component Interface Express
 2. Peripheral Clustering Interconnect Extended
 3. Peripheral Component Interconnect Expanded
 4. Printer Command Instruction Electronic

10. Which of the following is not typically considered a potential environmental hazard to a server room or data center?
 1. Dust
 2. Humidity
 3. Sunlight
 4. Power surges

16

Common Software Issues

Many software problems result from hardware issues. However, software can create, cause, or generate its own problems, faults, and errors related to, or completely independent of, hardware issues. In this chapter, we review common software problems, their causes, and some of the tools and utilities used to troubleshoot them.

The specific topic areas we look at in this chapter are as follows:

- Common software problems
- Common problem causes

Software problems

Typically, by the time an application or some system software is running on a server, it has been well tested in a test environment, run in a simulated production environment, and released to the production environment under close monitoring. Well, okay. Perhaps not all software goes through such a rigorous test and trial process as this, even though they should. Besides errors in logic, programming structure, and interfacing, what looks like a software error could be a hardware fault reflecting through a piece of software. Or, the software may be incompatible with the platform on which it's installed.

Software problems aren't the sole domain of application software. Problems and errors can be the result of misconfigured, improperly installed, or incompatible system software, device drivers, utility software, and even diagnostic and analysis software, and, of course, application software. In the sections that follow, common software problems are separated into two groups—hardware-related software problems and operating system problems.

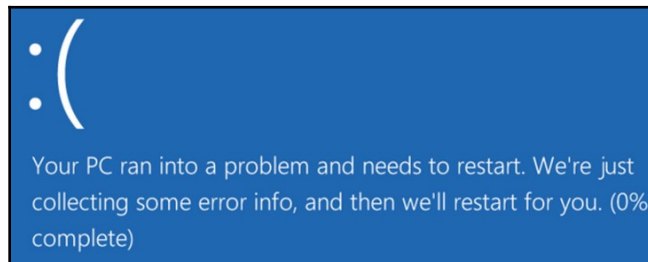
Hardware-related software problems

First off, some software problems are actually hardware problems or even malware problems. In some instances, it may be difficult to immediately determine the underlying cause of a problem and you must go with your best guess. However, next time, you'll have a better idea on what to do. What your first troubleshooting focus should be is fully explained by Gilster's law, as follows:

"You never can tell; and it all depends."

Common hardware-related problems that can appear to be software issues include the following:

- **Blue Screen of Death (BSoD):** A BSoD is a Windows error condition report screen, (see the following screenshot), that notifies the computer's operator of a *stop error* condition, which halts the operating system. This condition signals the occurrence of an operational error involving a specific hardware component or that hardware's device driver. Windows gathers data related to the failure and restarts. A BSoD can also be displayed by components of the operating system's kernel. Application software typically cannot cause a BSoD to display, unless it causes a hardware or operating system error. However, when an application fails, it rarely takes the operating system or hardware with it:



The BSoD display of a Windows 10 system

- **Disk boot failure:** This is also known as OS boot failure. Perhaps the most common cause of this error is that the boot sequence in BIOS/UEFI has a CD or DVD drive as the primary boot device and that drive is empty or has something other than a boot disk in it, in which case, the user should remove the CD or change the BIOS boot sequence and restart the computer. However, the problem could be a new unformatted hard disk, or a hard disk not connected to power.

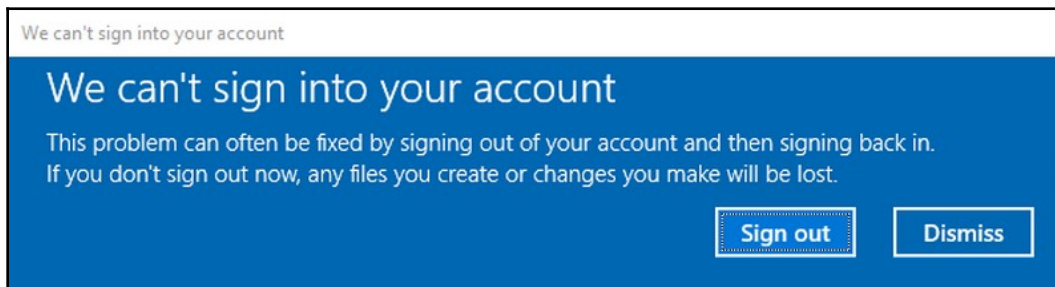
- **Cannot mount drive:** The cause of this error depends on which operating system you are running. On a Windows system, this problem is likely a hardware issue with an external, probably USB, hard drive and the problem is specifically with the USB port, the connecting cable, or the external drive. However, on a Linux system, this problem could be a corrupted filesystem or an issue in the filesystem table file (`fstab`).

Common operating systems problems

As much as we'd like to think that operating systems are rock solid and error-free, reality and experience tell us that this is definitely not the case. An operating system, like all software, can have problems and produce errors when it comes together with different, and often incompatible, resources.

The following are common operating system problems and errors you may encounter on the Server+ exam:

- **User is unable to log in:** Login problems for users are essentially the same on all of the big three operating systems—Windows, macOS, and Linux:
 - **Windows:** See the following table for common operating system-related login problems. The error message box shown in the following screenshot is an example of a common Windows login failure:



A Windows login error message box

- **macOS:** Like other operating systems, macOS doesn't require a password for users, but highly recommends its use. However, if the password or username is forgotten, either or both can be reset in admin mode, in single user mode, or through the Apple ID account.
- **Linux:** Most Linux releases and flavors incorporate the **Pluggable Authentication Module (PAM)** for authenticating username and password credentials. If a user is unable to log in with valid credentials, the PAM table may need to be reconstructed using the `pam-auth-update -package` command.

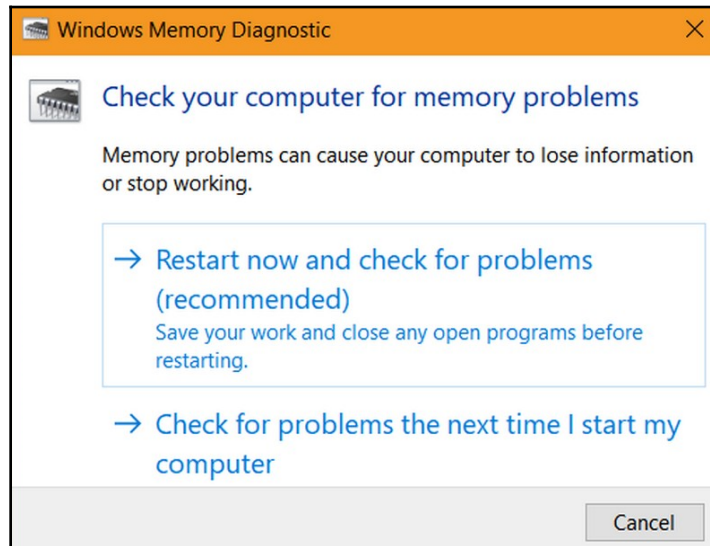
Common Windows user login errors and possible causes include the following:

Message	Problem
Windows can't sign into your account.	Password changed or upgrade installed.
Windows can't login with password.	Password on file is different from the one entered.
We can't sign into your account.	User profile is missing or corrupted. Windows upgrades can cause user login problems.
The User Profile Service service failed the sign-in. User profile cannot be loaded.	User profile is missing or corrupted.
No login screen is displayed.	This is likely a startup (boot) problem.

- **User cannot access resources:** If this error indicates a denied access to resources on a single computer or on a shared network resource, it is very likely a permissions problem for an individual user account or a group account of which the user is a member. However, if the error is about denied access to resources on a peer-to-peer network, the problem may be a TCP/IP configuration issue or a browser error on one of the peer stations.
- **Users cannot print:** In most cases, this condition may be something as simple as an incorrect printer driver on the affected computer, or it could be a group policy problem. Isolate the problem by logging in with a well-known account to determine if the issue is local or global.

- **Driver issues:** A new device trying to run with the driver of a previous model may not have the best results. Driver issues are typically compatibility problems with its associated device driver. These problems can also be caused by errors in the BIOS/UEFI configuration.
- **Cannot write to system log:** Users must have authorization to write to system log files. On a Windows system, the registry entry for the log file you wish to grant write permission to a user, edit the log file's entry in the registry's `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services` key. On a Linux system, a log file directory that is owned by the user should be created.
- **Slow OS performance:** All operating systems slow down when one or more programs dominate the system resources, primarily the CPU, memory, and secondary storage. Using the appropriate tool, such as Windows' Task Manager, macOS's Activity Monitor, and Linux's `systemctl` command, end the resource hogs that aren't crucial and perhaps review the programs and services in the startup file.
- **Runaway process:** Any program or service that gets into an infinite loop and perhaps launches other jobs is runaway. On a Windows system, open the Task Manager and end the process. On a macOS system, use the Activity Monitor to end the process. On a Linux system, use the `ps` command to *kill* the process.
- **Patch update failure:** If, after applying a patch to the operating system, one or more components, services, or applications begins having problems and you're sure the problem is the update or patch, use the system's backup utility to revert to your previous state. In Windows, this is the **Go back to a previous version** function on the **Settings** menu. On a macOS system, run the **Reinstall macOS** utility from recovery mode. On a Linux system, use the **Last-Known Good** option on the **GRUB** menu.
- **Service failure:** The message **The service cannot be started, either because it is disabled or because it has no enabled devices associated with it** indicates that one or more services, programs, or scripts have failed to start. You need to isolate the service and either reload it or a new version.
- **Hangs on shutdown:** If an operating system fails to shut down, when directed to do so, the problem likely is one or more of three basic issues—a process or service won't stop, a device driver is hung up or is holding a device open, or malware is causing the operating system to hang.

- **Memory leak:** This condition happens when a program fails to release some or all of its allocated memory when it's no longer needed. Memory leaks also happen when software loads an object to memory that cannot be accessed by its programming code. If multiple instances of the program causing the problem are executed, the system could eventually run out of available memory. Tools such as the **Windows Memory Diagnostic** utility (see the following screenshot) can help to resolve these errors:



The Windows Memory Diagnostic tool can be used to resolve memory leaks

Common problem causes

Many of the problems discussed in the preceding section have distinct causes, and some problems can cause other problems. In the sections that follow, we discuss the cause or causes of problems common to server hardware, and especially those you may encounter on the Server+ exam.

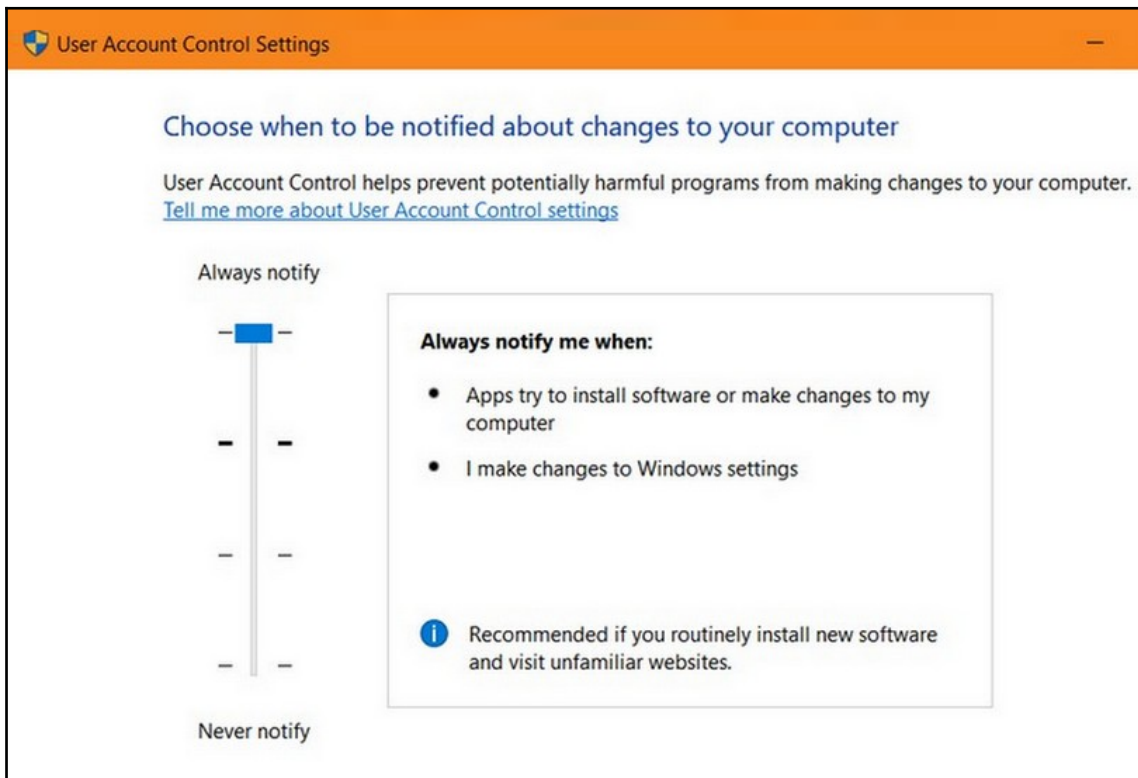
User Account Control (UAC)

On a Windows computer, this control feature is the UAC. On a Linux computer, the equivalent, which has been around much longer than the Windows UAC, is the `sudo` command.

Windows UAC

The purpose of the UAC (and `sudo`) is to restrict changes to the system to only those authorized to do so. In most situations, this would be the responsibility of the administrator using an administrator-level account. The problems associated with having the UAC enabled is that each time new or modified software installs, the user must enter an administrative password. UAC can be set at one of four levels, which range between **Always notify** and **Never notify**, as shown in the following screenshot.

Depending on where the system administrator sets the UAC level, the user must enter an administrator account password (**Always notify**) or is never asked for a password (**Never notify**). Too much user account control can lead to installation mistakes, malware, and serious application or system errors:

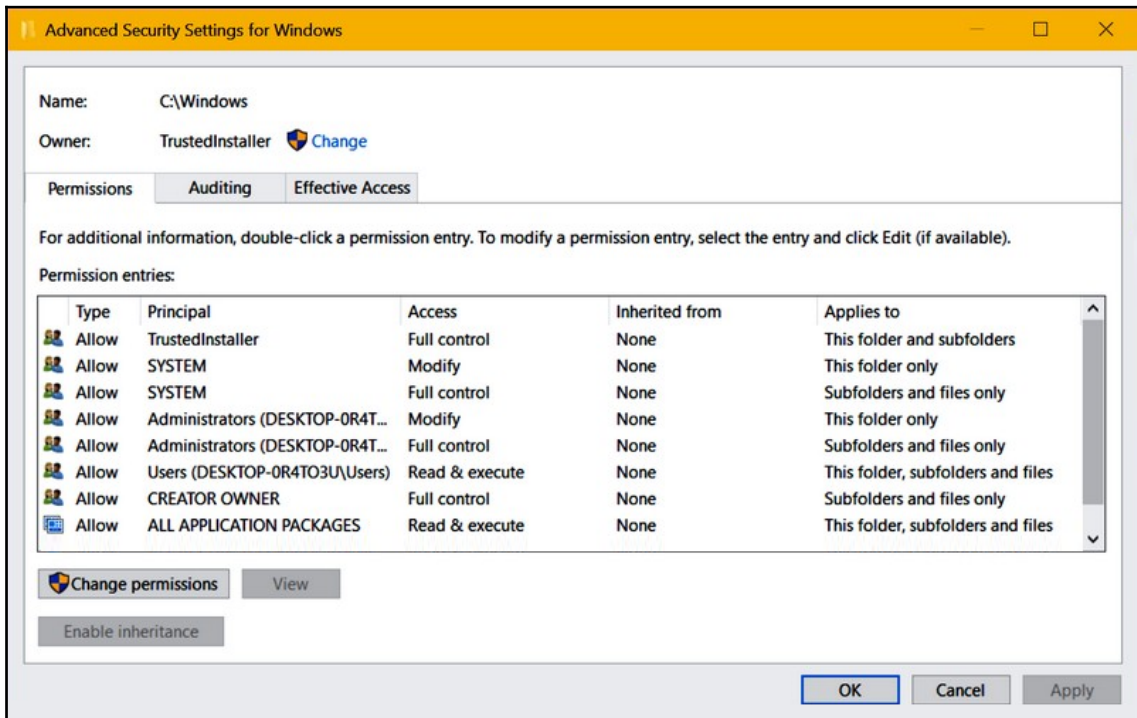


The Windows UAC sets the level of control for changes to the system

Access control

Although they use slightly different formats, Windows, Linux/Unix, and macOS use essentially the same access permissions structure. Access rights and permissions are specified on three levels—the owner (creator), the owner's group, and everyone else. Linux and macOS, which are built on a UNIX-like foundation, use the same permissions scheme. While Windows is similar, it presents it very differently.

On a Windows system, the equivalent of an access control list that defines what a user or group can access, open, modify, write, or delete is defined in the user's or group's permissions. The following screenshot shows the **Advanced Security Settings for Windows** dialog box which provides access to the existing permissions of all users or groups and access to administrators to modify the permissions:



The Windows Advanced Security Settings dialog box

Because the macOS operating system is built on top of **Portable Operating System Interface (POSIX)**, which is a UNIX/Linux workalike, macOS and Linux use essentially the same commands and permission sets to allow or restrict user access to resources. System administration can designate not only the access levels for the file owner, the owner's group, and everyone else, but can also control who is able to execute certain commands that may alter or manipulate a file, its directory, or its contents.

The file permissions of a Linux/UNIX/POSIX file or director is in the format shown in the following screenshot. In the leftmost section of this display, there are four sections. The first is a single character that is either a `d` for directory or a `-` for file. The next nine characters are three sets of three permission indicators, which represent, from left to right, the permissions of the file owner, the group of the owner, and all other users, respectively. For example, the `apps` directory is indicated with a `d` in the first position, and each of the three permission groups have read (`r`) and execute (`x`) rights. The `game1` file is not a directory (`-`) and the owner and group have read, write (`w`), and execute permissions. The permissions of others is only read and execute:

-rw-rw-r--	1	rprice	rprice	40	Feb	2	13:00	all
dr-xr-xr-x	1	rprice	rprice	4096	Feb	2	11:13	apps
-rwxrwxr-x	1	rprice	rprice	40	Feb	2	13:00	game1
-rwxrwxr-x	1	rprice	rprice	40	Feb	2	13:00	game2
drwxr--r--	1	rprice	rprice	4096	Feb	2	11:32	games
drwxr-xr-x	1	rprice	rprice	4096	Feb	2	11:13	misc
-r-xr-xr-x	1	rprice	rprice	40	Feb	2	13:00	newgame
-r--r--r--	1	rprice	rprice	40	Feb	2	13:00	oldgame
-rwxrwxr-x	1	rprice	rprice	40	Feb	2	13:00	puzzle
drwxr--r--	1	rprice	root	4096	Feb	2	11:13	usr
drwxrwxrwx	1	rprice	rprice	4096	Feb	2	11:13	usr

A Linux directory file list showing type, permissions, owner, group, size, and creation information

In either operating system, if a user tries to access or execute a file or directory for which they don't have permission, the user is asked for a password or at least challenged on whether or not this is the action they intended.

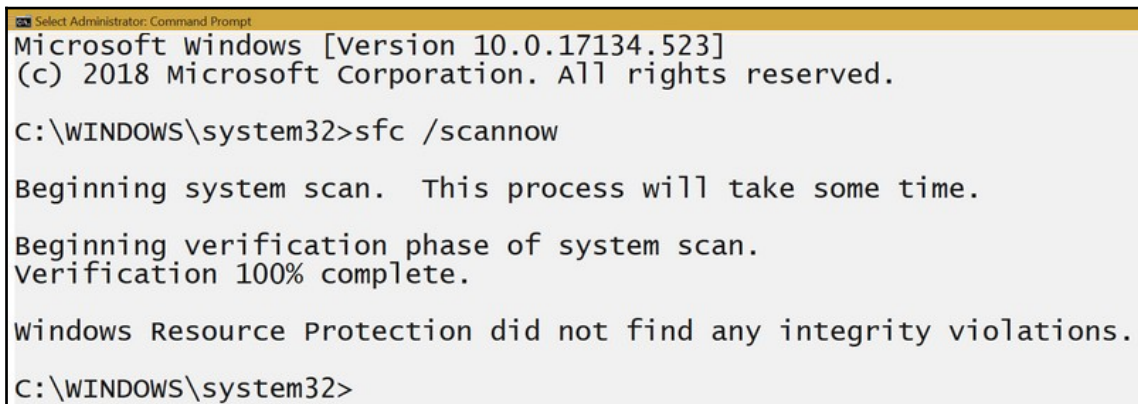
Corrupted files

The reasons as to why a file, filesystem, or even, heaven forbid, a hard disk partition, becomes corrupted are many and varied. Unfortunately, you typically discover this problem when the data or program is needed at that exact moment. While not all corrupted files are recoverable, the majority are, depending on the extent of the damage and, in some cases, what caused it. However, there are cases where data has been recovered from hard disk with broken platters, heavy water or fire damage, and other catastrophes. For the most part, a server administrator such as yourself may only encounter a file or two that the system can't find or read.

Windows file recovery

There are four types of file recovery on a Windows system:

- **System File Checker (SFC):** The `sfc.exe` file performs a scan of the hard disk looking for corrupted system files and folders. SFC works with **Windows Resource Protection (WRP)** which protects the registry files and other important system files. SFC can also be executed in Windows Safe Mode to eliminate any contention with third-party software. The following screenshot shows a typical result from `sfc.exe`:

A screenshot of a Windows Command Prompt window titled "Select Administrator: Command Prompt". The text inside shows the execution of the System File Checker (SFC) utility. It starts with the Windows version and copyright information, followed by the command `C:\WINDOWS\system32>sfc /scannow`. The output indicates the start of a system scan, followed by a verification phase that is 100% complete. The final message states that Windows Resource Protection did not find any integrity violations. The prompt ends at `C:\WINDOWS\system32>`.

```
Select Administrator: Command Prompt
Microsoft Windows [Version 10.0.17134.523]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>sfc /scannow

Beginning system scan. This process will take some time.

Beginning verification phase of system scan.
Verification 100% complete.

Windows Resource Protection did not find any integrity violations.

C:\WINDOWS\system32>
```

The output of the SFC utility

- **Deployment Image Servicing and Management (DISM):** This command-line utility is basically a multi-tool that can scan, check, restore, and repair corrupted files on a hard disk or stored as a part of an image file. The following screenshot shows an example of DISM:

```
C:\WINDOWS\system32>dism /Online /Cleanup-Image /CheckHealth  
  
Deployment Image Servicing and Management tool  
Version: 10.0.17134.1  
  
Image Version: 10.0.17134.523  
  
No component store corruption detected.  
The operation completed successfully.
```

The DISM.exe utility

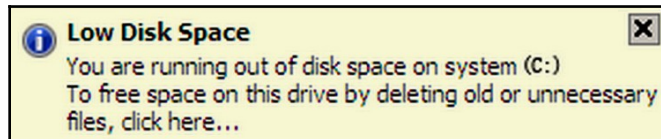
- **Replace the corrupted file:** Provided that there is a backup of the last-known good copy of a corrupted file, you can simply overwrite the bad file with a copy of the good file. Remember that you may need to take ownership (`takeown`) and gain full administrative rights (`icacls`) of the corrupted file first.
- **Restore system:** If all other processes fail to recover a corrupted file, or if the number of corrupted files, the filesystem, or the partition is too large, the best approach is most likely restoring the system to its last-known good point.

Linux file recovery

There are several data and file recovery utilities available for Linux systems. These tools range from those that find and copy retrievable data and files from a corrupted partition or volume to a good storage medium, such as **Ddrescue**. Others attempt to repair the structure and retrieve the data of corrupted files and images, such as **TestDisk**. Many of the same utilities that run on Linux are also available for macOS.

Hard disk space problems

You're working on installing new mission-critical software, when an error notice, such as the one shown in the following screenshot, pops up. The knee-jerk reaction, and this may be the correct reaction, is to remove anything on the hard disk that is no longer used or has simply been forgotten. However, there are other options to be considered:



A low disk space pop-up in Windows Server

In Windows, to free up disk space, you can make use of the following options:

- **Compression:** Compressing all or a large part of the files on the hard disk can free up space. In some instances (it really depends on the data being compressed), as much as 20 percent of the disk's capacity is made available. Remember that this could slow down I/O operations, though.
- **Disk management utility:** The Windows **Disk Management** utility allows you to reallocate disk space that is *healthy*, but doesn't have a drive letter assigned to it. Be sure that you have the entire disk backed up, just in case you forgot something about that space. Regardless, you can add this space to the system disk partition.
- **Temp folder:** The problem of low disk space could be the result of a Temp folder that hasn't been emptied recently, if ever.
- **Storage sense:** This is another Windows utility that can identify marginal files that are taking up a disproportionate amount of disk space and provide you with the opportunity to delete them.
- **Disk cleanup:** This is yet another Windows utility that reports the amount of space that several temporary data stores are using, with the option of clearing all or some of this space.

On Linux and macOS, the following options are available:

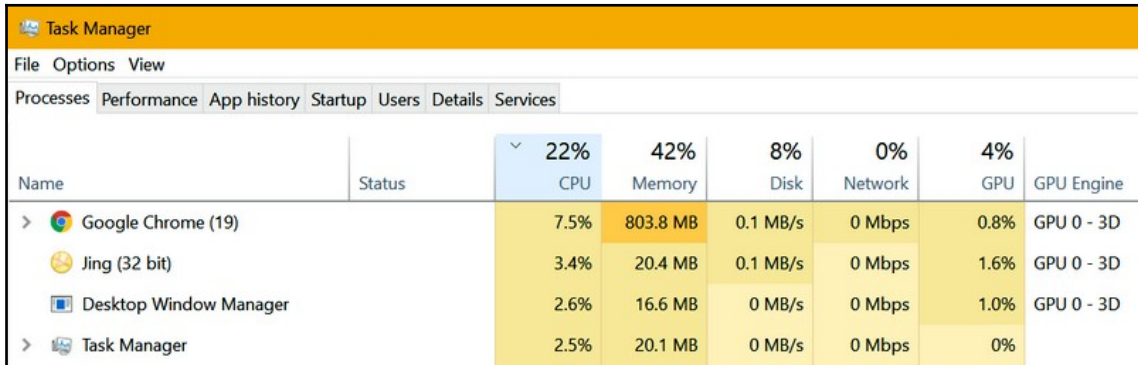
- **Disk usage/disk free:** These two commands (`du` and `df`) list the amount of disk space that is in use and is free, respectively, on the current mounted filesystem.
- **List open files:** Another command that can help you to find free space, or space that can be freed up, is the `lsof` command. This command has more than a dozen options that list files that are currently in use and are opened by running jobs at some point. This list identifies files that probably shouldn't be removed.

Lack of system resources

Unfortunately, the term *system resources* is non-specific and, when included in an error message, as in *memory error* or a *lack of system resources*, it may not be telling you anything specific. For example, a memory error doesn't mean you've used up all of the available main memory. In most cases, it actually means that the computer has run out of, or used up, an area in memory set aside by the system and called a **heap**. A heap is the area of memory that a running job has allocated dynamically over and above its allocation from the system. When the portion of memory reserved for heaps runs out, a memory error is generated. A heap is also considered to be a system resource.

In most cases, the immediate fix for an error resulting from a shortage of system resources is to restart the computer. However, to track down the problem and fix it so it doesn't keep happening, you need to isolate what is causing the problem. If this problem occurs each time you're running one particular application, take note of what else is running. You may have a compatibility or a resources contention issue.

In any case, you should use the Event logs and the Task Manager, (see the following screenshot) to investigate what was running at the time the problem happened and what jobs are dominating the CPU and main memory. You know, the problem could also be that you just need more RAM:



The screenshot shows the Windows Task Manager window with the 'Performance' tab selected. The window title is 'Task Manager'. The menu bar includes 'File', 'Options', and 'View'. The tabs at the bottom are 'Processes', 'Performance', 'App history', 'Startup', 'Users', 'Details', and 'Services'. The 'Performance' tab displays a table of system resource usage. The table has columns for Name, Status, CPU, Memory, Disk, Network, GPU, and GPU Engine. The data is as follows:

Name	Status	CPU	Memory	Disk	Network	GPU	GPU Engine
> Google Chrome (19)		7.5%	803.8 MB	0.1 MB/s	0 Mbps	0.8%	GPU 0 - 3D
Jing (32 bit)		3.4%	20.4 MB	0.1 MB/s	0 Mbps	1.6%	GPU 0 - 3D
Desktop Window Manager		2.6%	16.6 MB	0 MB/s	0 Mbps	1.0%	GPU 0 - 3D
> Task Manager		2.5%	20.1 MB	0 MB/s	0 Mbps	0%	

The Windows Task Manager can help to identify resource utilization

Virtual memory problems

Virtual memory, like all things virtual, is not memory at all. Rather, it is a portion of secondary storage that the operating system treats as an extension of main memory. Virtual memory, also known as **swap** file, allows more jobs to run on the computer. The operating system moves portions of idle jobs out of the main memory into virtual memory to allow an active job to use the now-available memory space. The drawback to virtual memory is that most secondary storage devices, such as a **hard disk drive (HDD)**, are much slower than RAM, which can add latency to a process.

One of the more common issues with virtual memory is evidenced when the **Out of Virtual Memory** alert displays, as shown in the following screenshot. What this error means is that there are too many jobs running in main memory and, as a result, there isn't enough RAM to go around. If this issue happens frequently, the best solution is to increase the RAM capacity. The next best solution is to increase the amount of virtual memory. There are no hard and fast rules concerning how much virtual memory you should configure, but its size should never be more than 150 percent of the amount of RAM installed on the computer. Getting the amount of virtual memory exactly right is a challenge. Too much or too little can work against you and slow down the computer:



Windows alert for low virtual memory

Another problem that can occur when virtual memory is enabled is caused by a corrupted pagefile. Virtual memory is actually a pagefile (`pagefile.sys` to be exact), which is an allocated space on a hard disk where the operating system can temporarily store idle programs and data. When you minimize a window on the desktop or pause an application, some or all of that job becomes idle. By moving it out of RAM and into a pagefile, the other jobs competing for RAM are able to run a bit faster. So, *what happens if the pagefile becomes corrupted?* The good news is that virtual memory is only used when there is not enough RAM to support the jobs running on the computer. If you never run out of RAM, the pagefile is idle anyway. Otherwise, run the `sfc /scannow` command that we discussed earlier to repair the pagefile.

Fragmentation

When a very small file is written to a hard disk, it's placed in a small disk area called a **sector**, which is typically 512 bytes in length. If a larger file, say one that is 1 MB, is written to the hard disk, obviously it won't fit into one sector. Therefore, the disk controller will break the file into about 2,000 pieces and store each into a sector. Not all of the available sectors are necessarily contiguous, so the file pieces end up being store all over the medium. As other files, large and small, take up sectors throughout the disk and the pieces of our first file reduce, enlarge, or disappear, sectors become open and new sectors are created. As sectors come and go, the original placement of the file pieces becomes more distributed and randomized. This is **fragmentation**.

Hard disk fragmentation can cause reliability problems, such as boot failures, file corruption, system hang-ups, and data errors or loss. It can also cause performance problems such as slowing I/O operations, lengthening the time for disk scans, and increasing the condition of disk thrashing. Defragmentation utilities shuffle the file pieces around so that the pieces of a file are as close to being on sequential sectors and sequential tracks as logically and physically possible.

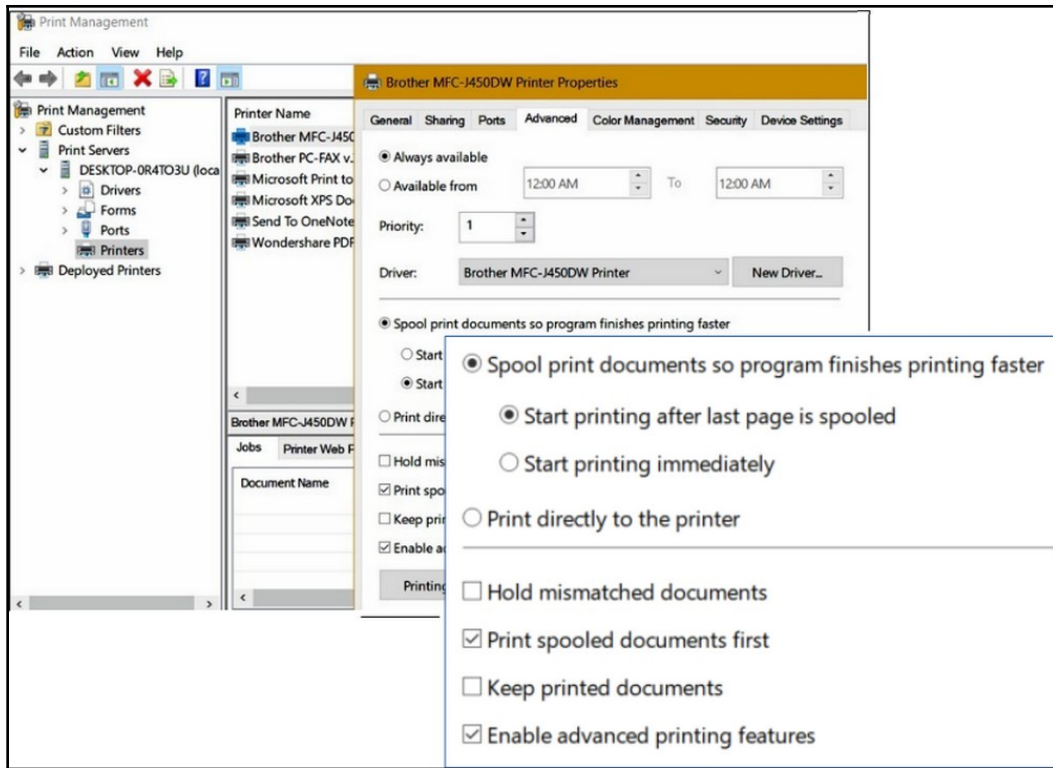
Printing issues

Printers of all types have remained an essential part of any network, large or small. However, the technology has allowed fewer printers to serve larger user groups. This is possible because of a variety of related technologies, but primarily due to the evolution of the print server.

A **print server** is either software or a dedicated device that provides an interface between network clients and one or more network printers. It is also a **single point of failure (SPOF)** in most cases. As with software and hardware, there is no one standard protocol used for sending a print job to a print server and then on to the printer. Adding a print server to a network can solve a number of printing issues, not only for users, but for the administrators as well.

A common problem with print servers is slowing down. There are several causes for a print server to run slow, including some unique to a particular installation, but a few others are as follows:

- **Bottlenecks:** If the print server seems to lock up with several print jobs in its queue, it's possible that there is a problem with memory or a service that may be consuming its resources. The best way to learn what may be causing this problem is to monitor the Windows Task Manager when no print jobs are running and then again with one print job and on up to everybody printing. Somewhere along this line, you should identify the culprit.
- **Print management:** You can access the settings for a network's print server(s) on the Windows **Print Management** page as shown in the following screenshot. Setting the options on each printer to print after spooling, print spooled jobs first, and assigning it a relative priority establishes a ranked queue that eliminates contention:



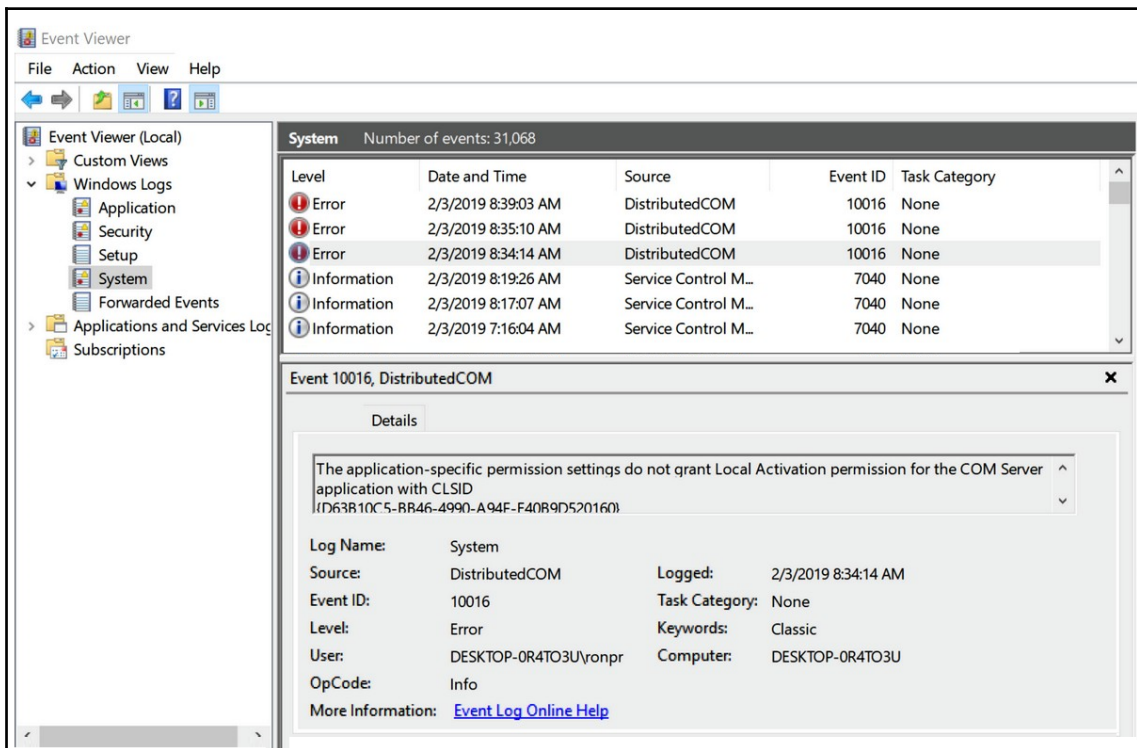
The Windows Print Management page showing the properties of a printer supported by the print server

- **Other remedies:** Every network is somewhat unique, especially concerning printers. To speed up network-based printing, there are a variety of actions that may or may not be appropriate to any particular network. A few of these actions are as follows:
 - **Print spooling on dedicated storage:** Spooling print jobs on a computer that supports active or large-volume applications can slow the I/O operations involved. Dedicating an isolated hard drive specifically to the spooler can speed things up.
 - **Add storage to the print server:** Something most users don't consider, if they know about it at all, is that print jobs grow in size when they are rendered for the printer. If several print jobs are vying for storage space on the print server, the logjam that results could slow the overall performance of the print server. Adding additional secondary storage to the print server can help to avoid this issue.

- **Clear the queue:** On occasion, a print job gets hung up in the print queue and blocks all other print jobs behind it. There are a couple of ways to force the print queue to empty—end the print spooler service or run `net stop spooler` from Command Prompt and delete the queue from the spooler.

Log files

Operating systems keep log files, and this is a good thing. Log files provide a running history of the events, actions, errors, and actions deserving a warning, and some good things too. Troubleshooting an operating system or an application problem should typically start with a look at, and some analysis of, one or more of the system's log files. The following screenshot shows the Windows **Event Viewer**, which provides a look at the contents of a variety of system log files:



The Windows Event Viewer can be used to scan log files for issues

On a Linux or macOS (POSIX) system, the system logs are centralized into the `/var/log` directory. The following screenshot shows the Command Prompt entries to list and then display one of the log files on a Linux system. There are also third-party software packages for viewing and analyzing log files on all three operating systems:

```
rprice@DESKTOP-0R4TO3U:/var/log/apt$ ls -l
total 24
-rw-r--r-- 1 root root 19176 Jul 25 2018 ejpp.log.xz
-rw-r--r-- 1 root root 883 Jul 25 2018 history.log
-rw-r----- 1 root adm 134 Jul 25 2018 term.log
rprice@DESKTOP-0R4TO3U:/var/log/apt$ sudo more history.log

Start-Date: 2018-07-25 15:56:27
Commandline: apt-get --purge remove --assume-yes ^linux-.* linux-
Purge: linux-headers-generic:amd64 (4.15.0-29.31), linux-image-4.
ders-4.15.0-29:amd64 (4.15.0-29.31), linux-virtual:amd64 (4.15.0.
31), linux-modules-4.15.0-29-generic:amd64 (4.15.0-29.31), linux-
linux-image-virtual:amd64 (4.15.0-29.31)
End-Date: 2018-07-25 15:56:30

Start-Date: 2018-07-25 15:56:33
Commandline: apt-get --purge remove --assume-yes ^grub-.*
Purge: os-prober:amd64 (1.74ubuntu1), grub-common:amd64 (2.02-2ut
grub-legacy-ec2:amd64 (1:1), grub-pc:amd64 (2.02-2ubuntu8.2), gr
oad-lists:amd64 (0.7)
```

A Linux log file displayed from the Command Prompt

Operating system monitoring tools

Windows, Linux, and macOS all include a set of commands or windows on which you can monitor the performance of all or part of a server. There are third-party tools that may be a bit more robust and intuitive, but if you need to check on one or more activities or conditions of the operating system and computer, the utilities included should do the job. The following table shows just a few of the commands or tools each one uses to monitor different performance areas:

Component/feature	Windows	Linux	macOS
Active processes	Task Manager	ps	Activity Monitor
Disk free space/utilization	File Explorer	df/du	About this Mac/Storage
Memory available/usage	Resource monitor	free	Activity Monitor
System status	Task Manager	top	About this Mac/More Info

The tools or commands available in Windows, Linux, and macOS for monitoring performance areas

Summary

System software problems can be from misconfigured, improperly installed, or incompatible device drivers, utility software, diagnostic software, and application software. Common hardware-related problems that can appear as software issues include BSoD, disk boot failure, and drive mounting issues. Common operating system problems include the user being unable to log in, the user being unable to access resources, users unable to print, device driver issues, system log errors, slow performance, runaway processes, service failure, and memory leaks.

Windows defines access permissions to a user or group. macOS and Linux allow or restrict user access and control the user's ability to manipulate a file. Windows file recovery options include SFC, DISM, replacing the corrupted file, and restoring the system to an earlier checkpoint. Linux systems use tools that find and copy retrievable data from the corrupted medium and tools that retrieve and repair data from corrupted files. Many Linux utilities also run on macOS. Low disk space warnings and errors on Windows can be resolved using compression, the Windows Disk Management utility, and unneeded data in temporary or search work files. On Linux and macOS, the `du` and `df` commands show the amount of disk space in use and the amount free, and the `ls -l` command shows the files that shouldn't be removed.

The fix for a shortage of system resources is to restart the computer. Use the Event logs and the Task Manager to see what's running when the problem happened and what's dominating the CPU and main memory. A common issue with virtual memory is the **Out of Virtual Memory** alert, which means there isn't enough RAM to support current demands. Hard disk fragmentation can cause boot failures, file corruption, and data errors. Defragmentation utilities move file pieces as close to being on sequential sectors and sequential tracks as possible.

A print server is a SPOF that can solve a number of printing issues. A common problem with print servers is slowing down, which may be caused by bottlenecks, the lack of printer management, or insufficient dedicated storage for print spooling, and can be solved by adding storage to the print server and clearing the print queue. Log files provide a record of the events, actions, errors, and warnings of the jobs running on the computer. Troubleshooting any software problems should start with a look at system log files. On Linux and macOS, system logs are in the `/var/log` directory.

Questions

1. You are training a new network server operator and technician. Unfortunately, he is unable to log in to the server. You are able to log in successfully. What could possibly be the problem?
 1. The server is powered off.
 2. He is using incorrect username and password credentials.
 3. He doesn't have access permissions to the server.
 4. Only one administrator account can be active at a time.
2. If you are able to access a folder and its files on a remote network server, but you are unable to delete out-of-date files, what folder permission are you lacking?
 1. Execute
 2. Read
 3. Write
 4. Delete
3. What notification display is caused by a variety of issues, including bad device drivers, faulty hardware, or components exceeding their operating limits?
 1. BSoD
 2. POST beep codes
 3. Flashing console lights
 4. Desktop freeze
4. The condition created when a program fails to release allocated memory when it's no longer in use is which of the following?
 1. Memory overflow
 2. Memory parity error
 3. Memory crash
 4. Memory leak
5. Which of the following can be used to identify corrupted files on a hard disk partition?
 1. `systemctl`
 2. GRUB
 3. SFC
 4. UAC

-
6. Of the following methods, which can be used to free up space on a HDD?
 1. Windows Disk Management
 2. Disk compression
 3. Virtual memory
 4. `fdisk`
 7. On a Linux system, what command-line command displays the location and amount of free disk drive space available on a filesystem?
 1. `du`
 2. `df`
 3. `lsdf`
 4. `fs`
 8. What is the feature that reserves a portion of the hard disk to extend the size and capacity of main memory?
 1. Virtual machine
 2. Virtual memory
 3. Virtual disk
 4. Virtual LAN
 9. Which system utility attempts to move file sectors into sequential locations on a hard disk drive?
 1. Degaussing
 2. Decompression
 3. Decryption
 4. Defragmentation
 10. Operating systems create and maintain a record of events, actions, errors, and warnings of the applications and services executing on a system. These records are recorded in what type of a file?
 1. Spool
 2. Queue
 3. Log
 4. Index
 5. Registry

17

Common Network Issues

Virtually every business and organization has come to rely on a computer network for its information, analysis, application, and media, as well as email, web access, scheduling, and personal productivity applications. In most cases, the network has become integrated into the fabric of these organizations with its availability taken for granted. However, when a technical problem causes the network to go down, network administrators must identify the problem, resolve it, and assure its integrity—all in as little time as possible.

In this chapter, we look at a number of common network problems, their causes, and some of the **Network Operating System (NOS)** utilities and tools that are commonly used as part of the troubleshooting process. The specific topics that will be introduced are as follows:

- Common network problems and their causes
- NOS resources used in troubleshooting network issues

Common network problems

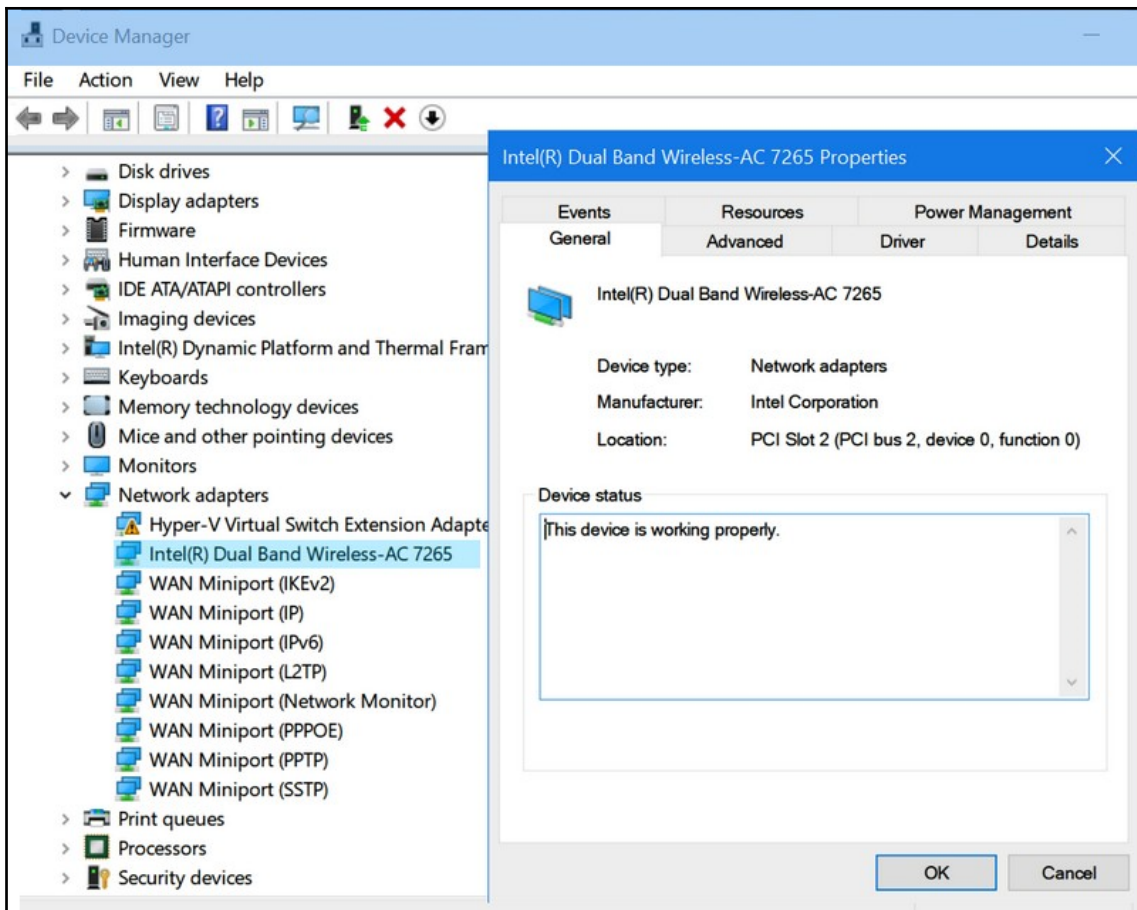
The list of common network problems is likely to vary from organization to organization. The common issues of any organization are a function of several variables, such as budget, staffing, training, and environment. The common network problems listed in the following sections are those identified in the Server+ objectives as being on the certification exam.

Internet connectivity

In some working environments, the loss of an internet connection may not be much of a problem, particularly if the activity on the network is local and doesn't need to go out onto the **Wide Area Network (WAN)**. However, at some point, access to the internetwork beyond the organization's gateway router becomes necessary, and users expect the service to be available.

The communication link between a **Local Area Network (LAN)** and a WAN includes several components, each of which could cause the link to fail. The following list includes the major pieces of this link and the potential for it to be the cause of a lost connection to the internet. Of course, you'd check these potential causes after you've checked the electrical cord and plug:

- **Cables:** On a wired network, a loose connection or a faulty connector will typically cause connectivity problems.
- **Connections:** On a wireless network, the host may not be able to connect to the access point or router and, if it can, it should be able to see the internet.
- **Network Interface Controller (NIC) failure:** If the host has an **Automatic Private IP Addressing (APIPA)** address, the next thing to check is the NIC or network adapter to see whether it's working. Perhaps the easiest way to check its status is in the Windows Device Manager. As shown in the following screenshot, opening **Properties** of the network adapter provides its functional status. If it is not working, the next step is to determine why:



Device Manager and the Properties box of a network adapter

- **Gateway configuration:** If the host can't see the gateway device, but you know the gateway has power and should be working, there is a possibility that the gateway configuration has become altered, corrupted, or hacked.
- **Service availability:** Regardless of the type of internet service you subscribe to, if the host is able to connect to the gateway device, but cannot connect to the internet, the problem could be that the device is faulty or the problem lies with the service provider. If the subscriber side of a connection is working without any faults, this could mean that the ISP's service is down. We'll go over more of the network status-checking commands later in this chapter.

Configurations

A common problem with network performance and connectivity is the configuration of the various components, both hardware and software. The configuration of the NOS, the host and its operating system, the network adapters in all the networking devices, and the interconnecting switching and routing devices must be compatible and configured to be in line with the purpose and function of the network.

The following sections are the common configuration issues that can be found on a network.

Dynamic Host Configuration Protocol (DHCP) server

One common cause for internet, and even some intranet, addressing problems is an improperly configured DHCP server. Another common related problem is the **Domain Name System (DNS)**, but more on this later. It's absolutely possible to boot up a network host without it receiving its IP configuration from the DHCP server. If the host's user never tries to access any resources outside of the host, no problems will typically appear. However, without a valid IP configuration, the host is very limited in terms of what it can reach.

There are two common reasons for a host failing to receive its IP configuration: no network connectivity and DHCP server issues. *Chapter 15, Common Hardware Issues*, dealt with the issues of the NIC, but let's take a look at DHCP server problems.

APIPA

One problem can be if a Windows host has a network connection, but, after booting up, it's configured with an unusual APIPA default address. This address is from a reserved Class B address group of 169.254.0.0/16. The purpose of the APIPA configuration is to allow the booting process and IP configuration to complete, so the host can operate on the local network. Routers don't forward APIPA addresses, but the resources of the local network should be available. However, this access is dependent on the configuration of any switches on the LAN.

After assigning the default IP address, the APIPA service checks for the presence of a DHCP server approximately every five minutes. When the DHCP server is online once again, it replaces the APIPA configuration with the DHCP configuration.

DHCP addresses

The most common issues with the configuration of a DHCP server are the issuance of duplicated IP address configurations, an exhausted address pool, and misconfigured static address assignments.

The issue of the same IP address being given to two or more hosts can happen when multiple devices (such as the router, switch, or network server) have an enabled DHCP server function using an overlapping range of addresses (called the **scope**), which duplicate static addresses that have been previously assigned to hosts.

An exhausted (or emptied) address pool means that the next host requesting a configuration will configure itself with an APIPA address.

DHCP servers assign an IP address and its related configuration to a network host in one of three ways:

- **Automatic:** This form of allocation designates an IP address to a host permanently, meaning an *unlimited* lease period.
- **Dynamic:** In this allocation option, the DHCP server assigns an IP address from a predefined address scope for a preset lease period. This option allows for renewals.
- **Static:** A specific IP address assigned to a host. A static IP address is a reserved address and is permanently linked to the host's MAC address.

Other misconfigured devices

If one or more components of a server or a network are misconfigured, it can have no effect at all. Alternatively, it can bring the system down and, as is commonly the case, without any real clue as to why. On the Server+ certification exam, you may encounter questions concerning the potential impact of several misconfigured devices, services, and protocols.

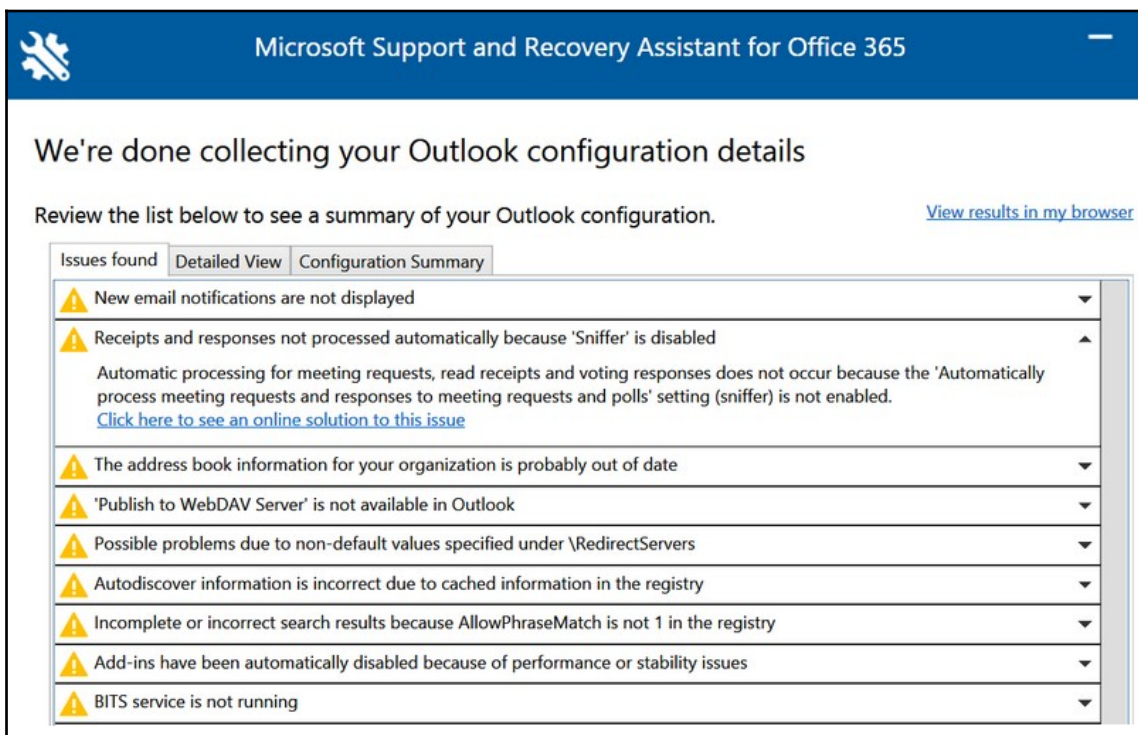
The following sections cover the issues you may see.

Email problems

For the most part, email problems typically result from misconfiguration. An email system typically involves a server, a transport mechanism (or protocol), and a client, all of which must work together. Some common problems that an email system may have include the following:

- **The client is unable to send or receive messages:** If your localhost is able to connect to the internet, the problem is likely to be with the mail system and commonly in the data entered to identify a mailbox identity (such as the email address and password).

In Microsoft Office, the account information of the Outlook client is common for all installed Office software. However, if Outlook is unable to download mail from the server, the **Microsoft Support and Recovery Assistant for Office 365** tool may be able to identify the problem, as shown in the following screenshot:



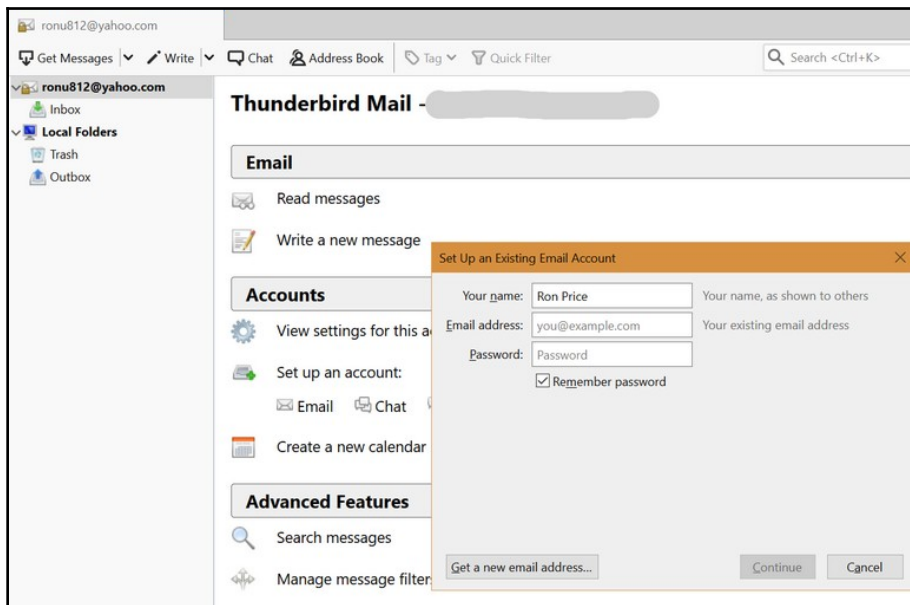
The screenshot displays the Microsoft Support and Recovery Assistant for Office 365 interface. The title bar is blue with a white gear icon on the left and the text "Microsoft Support and Recovery Assistant for Office 365" in the center. Below the title bar, the main content area has a white background. It starts with the heading "We're done collecting your Outlook configuration details" in bold. Below this, it says "Review the list below to see a summary of your Outlook configuration." with a link "View results in my browser" on the right. There are three tabs: "Issues found" (selected), "Detailed View", and "Configuration Summary". Below the tabs is a list of issues, each with a yellow warning icon and a dropdown arrow. The issues are:

- New email notifications are not displayed
- Receipts and responses not processed automatically because 'Sniffer' is disabled
Automatic processing for meeting requests, read receipts and voting responses does not occur because the 'Automatically process meeting requests and responses to meeting requests and polls' setting (sniffer) is not enabled.
[Click here to see an online solution to this issue](#)
- The address book information for your organization is probably out of date
- 'Publish to WebDAV Server' is not available in Outlook
- Possible problems due to non-default values specified under \RedirectServers
- Autodiscover information is incorrect due to cached information in the registry
- Incomplete or incorrect search results because AllowPhraseMatch is not 1 in the registry
- Add-ins have been automatically disabled because of performance or stability issues
- BITS service is not running

The Microsoft Support and Recovery Assistant can scan Outlook for email errors

Linux and Windows systems may also use a shareware or freeware mail client, such as Thunderbird (Mozilla), KMail, Evolution, or Geary. The configuration is straightforward and typically requires only an email address and a password, as shown in the upcoming screenshot.

- **The client can receive emails but is unable to send emails:** This situation normally means that the client may be trying to forward an email to TCP/UDP port 25. This well-known port was, at one time, the default port in nearly all email clients for outbound email. However, since spammers access port 25 routinely, most mail servers now block that port and use port 587 instead. Check the port assigned to the send mail function.
- **The client can send emails but is unable to receive emails:** This problem results from one or both of two conditions. Mail servers, especially those that support the **Internet Message Access Protocol (IMAP)**, operate within a set amount of allocated storage space for storing copies of forwarded messages. If the storage allocation fills up, email forwarding stops. The remedy for this is to access the mail server and delete any unneeded messages. The other condition is the configuration of the inbound account. Make sure that the incoming mail server, port number, email account, and password are all correct. Send yourself an email to check it:



The Thunderbird email client is a freeware tool for email

Hosts file configuration

On a Windows system, the `hosts` file provides a localhost DNS-type of lookup to provide the IP address of a **Fully Qualified Domain Name (FQDN)**. As illustrated in the following screenshot, the `hosts` file shows a domain name and its corresponding IP address. This entry eliminates the need to communicate with a DNS server to obtain the IP address. If the entry for a FQDN/IP pair is incorrect, or if the `hosts` file is missing, or if it has been saved with a file extension, the system bypasses the `hosts` file and sends out a DNS request:

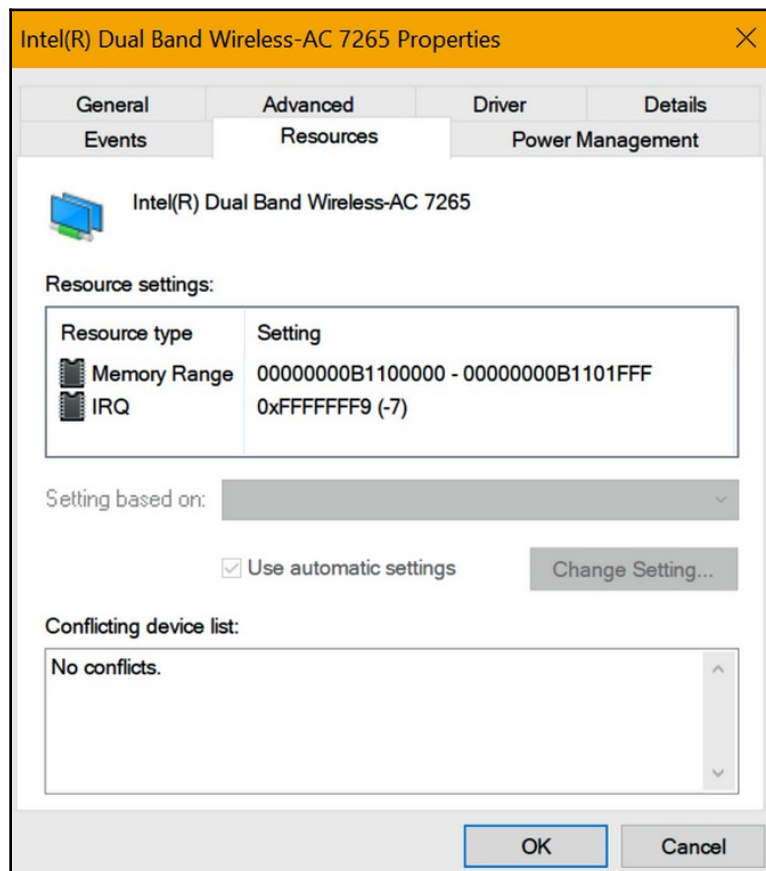
```
# Copyright (c) 1993-2019 Microsoft Corp.
#
# This is a sample HOSTS file used by Microsoft TCP/IP for Windows.
#
# This file contains the mappings of IP addresses to host names. Each
# entry should be kept on an individual line. The IP address should
# be placed in the first column followed by the corresponding host name.
# The IP address and the host name should be separated by at least one
# space.
#
# Additionally, comments (such as these) may be inserted on individual
# lines or following the machine name denoted by a '#' symbol.
#
# For example:
#
#       102.54.94.97       rhino.acme.com       # source server
#       38.25.63.10       x.acme.com           # x client host
#
# localhost name resolution is handled within DNS itself.
#   127.0.0.1             localhost
#   ::1                   localhost
#   52.216.238.66         www.packtpub.com
```

The hosts file on a Windows system

Misconfigured NIC

Although NIC cards, USB drives, and built-in network adapters rarely do so, they can fail. However, if you suspect the network card or adapter of interrupting communications to and from the network, typically the issue is one of two problems—either the NIC card or adapter is bad, or the NIC card or adapter is not getting the system resources it needs to do its job.

In most cases, at least with NICs, there are LED to help you identify whether it's working properly. When the system is powered up and connected to the network, if you don't see any green lights, you've found your problem. Still, the NIC may not be so bad that it requires replacing. It may be that it was reinstalled incorrectly, or a jumper block or switch setting may be incorrect. Check your documentation for these possibilities. Another issue involves the NIC card or the adapter with the same system resources as another I/O device. On a Windows system, Device Manager can help to identify any resource assignment issues, as you can see in the following screenshot. Device Manager can also check for a network adapter on the motherboard. Make sure in either case that the correct and up-to-date device drivers are in use:



Device Manager Component Properties dialog box

Routing and switching issues

Above all else, misconfigured routers and switches (as well as firewalls) are security issues. These devices can start out correctly configured to service and protect its network and interact with external networks. However, changes to a network, misinformation from a service provider, typing errors, and the ever-present *while I'm in here, let's see what else I can do* syndrome of network administrators can result in misconfiguration.

In addition to security, errors in the configuration of internetworking devices can also cause problems for the devices behind them, such as content servers, proxy servers, and LAN switches. Troubleshooting configuration problems on these devices can be a bit more complicated than on a network server.

VLAN configuration errors

Before we get into some of the common issues and problems associated with a **Virtual Local Area Network (VLAN)**, let's recall some of its terminology. Essentially, a VLAN is a logically created broadcast domain configured on a network switch. By designating one (or more) ports to a VLAN, besides VLAN1, which all ports belong to by default, the stations configured to that port are now on that VLAN.

As shown in the following diagram, dispersed hosts on different physical LAN segments can also be on a single VLAN. Also shown in this diagram, are the links that directly connect the switches and router to one another. These links are known as trunks, and the interface port to which the trunk connects is called a trunk port. Trunk ports run a special protocol, such as the IEEE 802.1Q, or the Cisco **Inter-Switch Link (ISL)**.

Two common problems of a VLAN typically have to do with IP addresses, such as the following:

- Network hosts configured with incorrect IP addresses, such as an APIPA, or an erroneous subnet address or mask, and not being able to access the network, let alone a VLAN
- Failing to configure interface ports to support each VLAN as a network segment or subnet

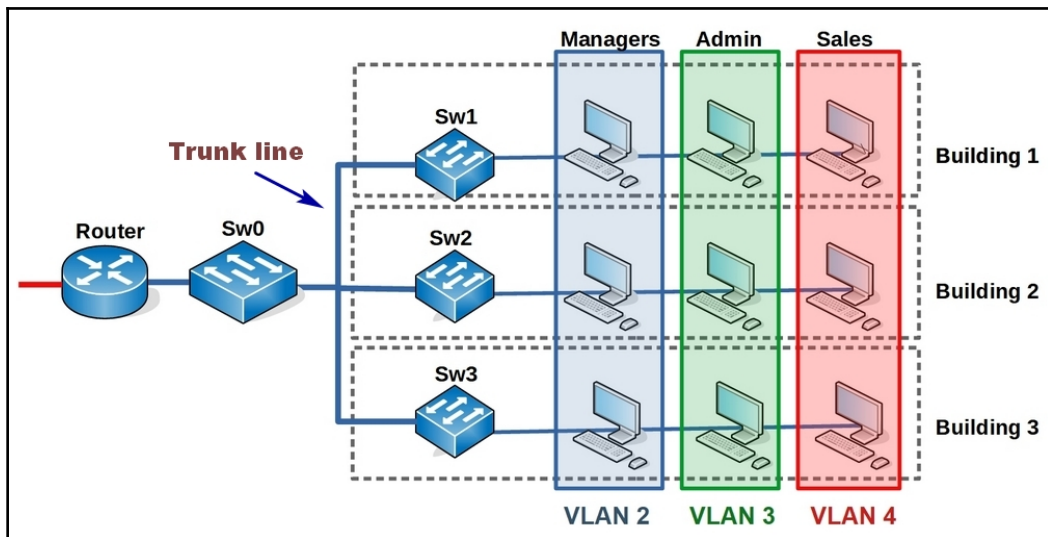
Other VLAN issues, beyond IP addresses, include the following:

- A VLAN is configured in a *down* state
- The VLAN's terminating trunk ports must be running the same trunking protocol and a common configuration as either a *tagged* or *untagged* VLAN

Some VLAN issues are the result of problems with a trunk port, such as the following:

- If two trunk ports each has a different trunk mode, they are unable to establish a link.
- A VLAN won't function if it's not on the list of authorized VLANs that can access a particular trunk port.
- A trunk port begins *flapping*. This condition is commonly associated with router ports, but switch ports can flap as well. **Flapping** is a hardware failure that causes the port to power on and off alternatively.

The following diagram shows you the VLAN configuration:

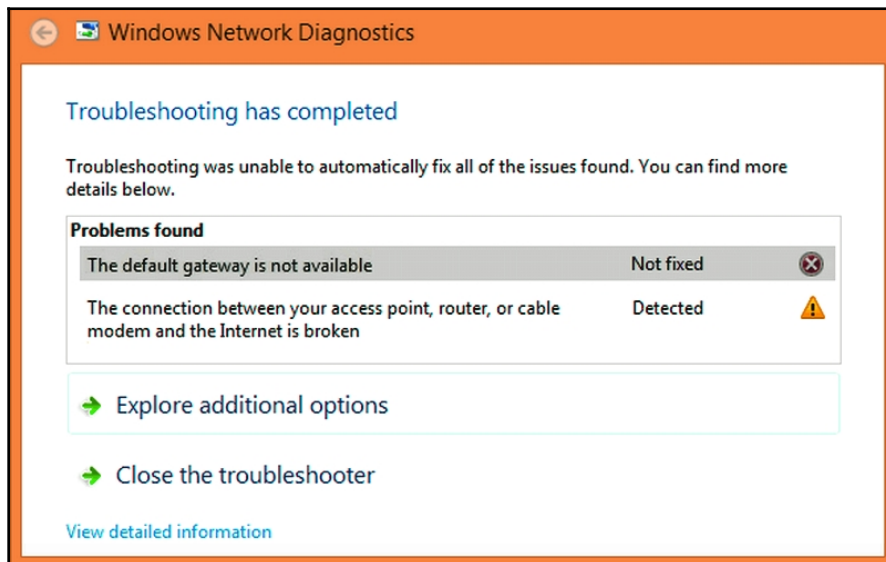


VLAN configuration

Default gateway not available

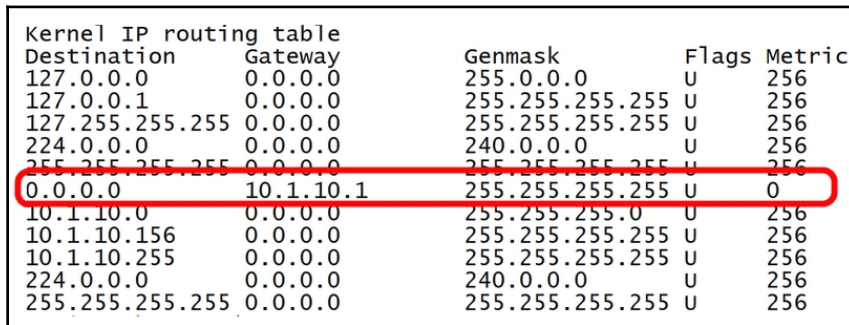
This is perhaps the most common network connection problem. Its causes are less than obvious and none of them is a problem of the gateway router itself. All you know is that you are suddenly unable to access the internet or any site beyond the gateway, and a message says that the default gateway is not available (see the following screenshot). Here are the common causes of this issue:

- **Antivirus or anti-malware software:** This has been known to change the network configuration of a host, including the default gateway's IP address. Remove this program and reset the default gateway configuration by rebooting the computer. If this solves the issue (you may have to wait a while before deciding), there you go.
- **Outdated network adapter drivers:** This may interfere with communications to the network connection on a computer. Check the default gateway connection after each of the following fixes (stopping after the one that seems to resolve the problem): reinstall the drivers, update the drivers, and replace or add in a network adapter.
- **Windows auto-login:** This has been known to mess up the IP configuration of a host at startup, including the IP address of the default gateway. There is no way to actually disable the auto-login feature, but you can set the login to require a username and password, which seems to solve the issue:



The Windows Network Diagnostics dialog box showing the default gateway issue

One result from a gateway error, such as the gateway address disappearing from the IP configuration on any operating system, is the message, *resource not available or unreachable*. The `route` command, which is available on virtually all operating systems, displays the content of the route table, which should include the address of the default gateway, as you can see in the following screenshot. The Windows `ipconfig` command or Linux/macOS `ifconfig` command also displays the address of the default gateway, if present:



Destination	Gateway	Genmask	Flags	Metric
127.0.0.0	0.0.0.0	255.0.0.0	U	256
127.0.0.1	0.0.0.0	255.255.255.255	U	256
127.255.255.255	0.0.0.0	255.255.255.255	U	256
224.0.0.0	0.0.0.0	240.0.0.0	U	256
255.255.255.255	0.0.0.0	255.255.255.255	U	256
0.0.0.0	10.1.10.1	255.255.255.255	U	0
10.1.10.0	0.0.0.0	255.255.255.0	U	256
10.1.10.156	0.0.0.0	255.255.255.255	U	256
10.1.10.255	0.0.0.0	255.255.255.255	U	256
224.0.0.0	0.0.0.0	240.0.0.0	U	256
255.255.255.255	0.0.0.0	255.255.255.255	U	256

The Linux `route` command results with the default gateway address highlighted

Firewall failure

A network firewall can be a standalone hardware device, a service provided by a router, or an independent software system. In any case, the purpose and performance of a firewall remains the same: to protect a network, server, or host from intrusion or attack. Like everything else in computing, firewalls can have problems. Here are a few of the firewall problems you might face:

- **Configuration issues:** If the configuration process of a firewall depends heavily on the default configuration, it's very likely that the resultant configuration will be too broad in scope and not specific enough in terms of detail. It's best to limit the configuration process to the core functions of the organization's security policies as a foundation.
- **Currency issues:** All too often, a firewall becomes a *set it and forget it* device, when, in fact, constant monitoring and maintenance to keep it current with its settings, rules, and updates should be normal procedure.

- **Processing power:** Software firewalls can become a bottleneck if installed on a computer with insufficient resources to handle the demand during peak operating hours. One way to lessen the load on the firewall is to limit or remove those features that aren't really necessary for the primary mission of the device or software.
- **IP address validity:** A common hack is to send a message into a local network's firewall that contains a bogus IP address in either the source or destination addresses (or both). If the pass/deny rules defined on the firewall don't consider phony or fictitious IP addresses outside the limits of IP standards, the incoming traffic of messages with dubious addresses could pass through to the internal network.

Miscellaneous common problems

In addition to the preceding, the Server+ objectives list a few other, more generic, hardware or software issues of which you should know the cause and cure. You may not encounter a question dealing specifically with one of these issues, but don't be surprised if they show up as a part of a question's scenario, or as one of its answer choices. Here they are, in no specific order:

- **Resource unavailable:** This error is very contextual. If it results from a mail client, then it's an email problem; if it results from a virtual machine, then it's a virtual machine problem; if it results from a network access issue, it could be a server or a network device problem. Most users associate this error with HTTP and either the 404 or 503 error conditions. The common meaning behind this message in any situation is that a requested file, service, or device is unavailable because of a name change, deletion of the file, or data corruption.
- **Destination host unreachable:** This error is associated with the **Packet Internet Groper** (the `ping` command) on a Linux or macOS system. On a Windows system, the message is likely to be **Request timed out**. The issue is that either there is no clear path to the destination address, or it doesn't exist.
- **Unknown host:** This error message results from an unreachable destination address. This could be a DNS problem, a configuration problem (such as the DNS server address), or just an incorrect destination address.

- **Failure of the service provider:** An ISP is actually a single source of failure and, if for some reason the ISP's routing services are no longer available, access to the internet also stops. One way around this potential error is by subscribing to more than one ISP for service.
- **Cannot resolve hostname/FQDN:** This error message indicates a problem with the information in the DNS server or the addressing to a remote site with either the hostname or the corresponding FQDN.

Troubleshooting tools

Although we have discussed each of the tools in the following list, let's take one more look at them in the context of their use in troubleshooting the types of problems included in this chapter. You should expect to see the tools in the following sections on the Server+ exam.

ping

`ping` is a command line utility program used to verify a connection between a source host and a destination host. Using the FQDN, hostname, domain name, or the IP address entered, `ping` sends out a 64-byte ICMP echo request message to that destination. Anything other than the IP address passes through the DNS or the local `hosts` file. After sending out the ICMP message, `ping` waits for the echo response from the destination. When a response comes back, `ping` displays the metrics of the transmitted and received messages.

tracert/traceroute

This command's purpose is to test and display the communication link between a source host and a destination host or network. `tracert` is the Windows version, and `traceroute` is the Linux/macOS version. This command uses an iterative process to incrementally test the path and connection to each of the routers on the path to a destination using the **Time-to-Live (TTL)** value. Initially, it sends out an ICMP echo request message with a TTL value of 1. This means that the router discards the message after reaching one router (hop) and only the first router on the path responds with the echo message. `tracert` then sends out another ICMP echo request, this time with a TTL value of 2. This process repeats until the TTL runs out (meaning the number of hops equals the default hop count), or until no incremental hops respond.

The following screenshot shows the results of a Windows `tracert` command:

```
C:\WINDOWS\system32>tracert comptia.org

Tracing route to comptia.org [198.134.5.6]
over a maximum of 30 hops:

  1    2 ms    9 ms    9 ms  10.1.10.1
  2   24 ms   19 ms   18 ms  96.120.102.249
  3   25 ms   27 ms   20 ms  po-107-rur01.spokane.wa.seattle.comcast.net [96.108.165.245]
  4   16 ms   20 ms   17 ms  po-2-rur02.spokane.wa.seattle.comcast.net [69.139.160.126]
  5   28 ms   28 ms   30 ms  be-37-ar01.seattle.wa.seattle.comcast.net [68.86.96.5]
  6   28 ms   29 ms   28 ms  be-33650-cr01.seattle.wa.ibone.comcast.net [68.86.93.165]
  7   30 ms   28 ms   28 ms  be-10847-pe02.seattle.wa.ibone.comcast.net [68.86.86.226]
  8   30 ms   28 ms   27 ms  66.208.232.182
  9   90 ms   88 ms   82 ms  216.156.16.80.ptr.us.xo.net [216.156.16.80]
 10   95 ms   98 ms   86 ms  207.88.12.228.ptr.us.xo.net [207.88.12.228]
 11   83 ms   88 ms   88 ms  207.88.12.144.ptr.us.xo.net [207.88.12.144]
 12   84 ms   98 ms   88 ms  207.88.12.190.ptr.us.xo.net [207.88.12.190]
 13   84 ms   86 ms   89 ms  207.88.12.189.ptr.us.xo.net [207.88.12.189]
 14   89 ms   88 ms   88 ms  207.88.12.164.ptr.us.xo.net [207.88.12.164]
 15   95 ms   88 ms   88 ms  216.156.16.199.ptr.us.xo.net [216.156.16.199]
 16   92 ms   86 ms   88 ms  216.55.11.62
 17  105 ms   90 ms  108 ms  198.134.5.6
 18  124 ms   98 ms   89 ms  198.134.5.6

Trace complete.
```

The results of a `tracert` command

ipconfig/ifconfig

While the Windows **Internet Protocol Configuration** (`ipconfig`) command, or the workalike **Interface Configuration** (`ifconfig`) command, are available on virtually all other operating systems, it is helpful to learn the configuration elements of a host and its network interface(s). The following screenshot shows an excerpt from its `/all` option. This command also allows for releasing or renewing the DHCP configuration and several other configuration adjustments:

```

Connection-specific DNS Suffix . : hsd1.wa.comcast.net
Description . . . . . : Intel(R) Dual Band Wireless-AC 7265
Physical Address. . . . . : 5C-E0-C5-B6-B3-9A
DHCP Enabled. . . . . : Yes
Autoconfiguration Enabled . . . . : Yes
IPv6 Address. . . . . : :e300::492f(Preferred)
Lease Obtained. . . . . : Saturday, February 2, 2019 2:30:29 PM
Lease Expires . . . . . : Saturday, February 16, 2019 9:47:35 AM
IPv6 Address. . . . . : :e300:8db7:a325:a3b3:74df(Preferred)
Link-local IPv6 Address . . . . . : 25:a3b3:74df%9(Preferred)
IPv4 Address. . . . . : 10.1.10.156(Preferred)
Subnet Mask . . . . . : 255.255.255.0
Lease Obtained. . . . . : Saturday, February 2, 2019 2:30:27 PM
Lease Expires . . . . . : Sunday, February 17, 2019 7:33:39 AM
Default Gateway . . . . . : ff:fe2e:2ac3%9
                          10.1.10.1
DHCP Server . . . . . : 10.1.10.1
DHCPv6 IAID . . . . . : 56418501
DHCPv6 Client DUID. . . . . : 00-01-00-01-23-50-8F-78-5C-E0-C5-B6-B3-9A
DNS Servers . . . . . : 2001:558:feed::1
                          75.75.75.75
                          75.75.76.76
NetBIOS over Tcpi. . . . . : Enabled

```

A sample of the results displayed by the `ipconfig` command

nslookup

The `nslookup` command-line utility, in virtually all operating systems, looks up names and IP addresses in a name server (hence its name). Using `nslookup`, you can find the IP address or domain of a host on the network. The following screenshot shows an example of `nslookup` on a Linux system, which provides the IP address for a domain. As shown, the response is **Non-authoritative answer**, which means this is the information in the DNS server at the present time:

```

rprice@ubuntu:~$ nslookup packt.com
Server:          75.75.75.75
Address:         75.75.75.75#53

Non-authoritative answer:
Name:   packt.com
Address: 52.216.225.242

```

The results of a `nslookup` command on a Linux system

net use/mount

The `net` command is actually more like a family of commands used to create a link to or to disconnect from a network shared resource, display all current connections on a host, share a resource with other hosts, manage passwords, control the print spooler, and more.

The `mount` command attaches (mounts) a filesystem or a secondary storage device. Mounting a device adds it to the active directory structure and makes its contents available for access. To dismount a mounted device or filesystem, the `umount` command notifies the operating system to complete any I/O operations to that unit and then remove it from the directory structure. Both the `net` and `mount` commands are available on Windows, Linux, and macOS.

nbtstat and netstat

NetBIOS over TCP/IP status (nbtstat) is a command-line tool used primarily for diagnosing or troubleshooting NetBIOS name issues. It can also display the NetBIOS services running on another host, or display whether there is a logged-in user on a particular host.

Network Status (netstat) displays information on the IP configuration of a host, but goes further to show its connections, including ports, protocols, and metrics on the communications.

Summary

The link between a LAN and a WAN includes several components that could cause a link to fail, including cables, connections, NICs, gateways, and the ISP. A common problem is the configuration of hardware and software. Areas where configuration issues are common include the DHCP server, email server and clients, the `hosts` file, NICs, routers and switches, VLANs, default gateways, and firewalls.

In many cases, the messages that are displayed characterize a problem, but not necessarily its cause. These include **resource unavailable**, **destination host unreachable**, **unknown host**, **failure of service provider**, and cannot resolve hostname/FQDN.

Some of the network problem troubleshooting tools available are `ping`, `tracert/traceroute`, `ipconfig`, `nslookup`, `net use`, `mount`, `nbtstat`, and `netstat`.

Questions

1. On a local network, a user is able to access resources on the LAN but is unable to access and download a web page. Which of the following areas would you suspect as the possible cause of this problem?
 1. Local network server
 2. Remote web server
 3. DHCP server
 4. Internet gateway
2. A local network host cannot complete its startup process without a successful interaction with the DHCP server. True or False?
 1. True
 2. False
3. What is the name given to the addresses in the Class B network 169.254.0.0/16 on a Windows system?
 1. EGRP
 2. FDDI
 3. APIPA
 4. Anycast
4. What TCP/UDP port has replaced port 25 for the SMTP electronic mail interface?
 1. Port 80
 2. Port 587
 3. Port 1024
 4. Port 20
5. What Windows file contains a list of hostnames and their associated IP addresses, which the operating system uses to look up the identity and location of a host?
 1. Hosts
 2. Pagefile
 3. WinSxS
 4. Registry

-
6. What is the formal name of a virtual structure that logically creates a broadcast domain through a switch port?
 1. VPN
 2. VM
 3. VLAN
 4. ISL
 7. A communication link that interconnects two switch interface ports and runs an IEEE 802.1Q protocol is known as a what?
 1. Frame relay
 2. Tagged port
 3. Flapping port
 4. Trunking port
 8. Which of the following could be a possible cause of a default gateway becoming unavailable?
 1. Antivirus software
 2. Network adapter device driver
 3. Auto-login service
 4. All the above
 5. None of the above
 9. A network technician installs a number of network hosts in a new branch office located in another state. What command-line utility should they use to verify that each of the new hosts is able to see the network server in the home office?
 1. nslookup
 2. ping
 3. tracert/traceroute
 4. ARP/RARP
 5. Either ping or tracert/traceroute, or both
 6. None of the above
 10. What is the Linux command-line utility used for displaying and modifying network interface configuration settings?
 1. ifconfig
 2. ipconfig
 3. net config
 4. net setup

18

Common Storage Issues

So far, a magnetic storage technology that's impregnable, and can withstand corruption and accidental or malicious erasure, doesn't exist—not yet, anyway. In the meantime, disk, tape, and optical storage devices continue to have their problems, issues, and faults. In this chapter, we look at the data storage issues and their respective causes that you should expect to encounter in the Server+ exam.

We will cover the following topics in this chapter:

- Data storage device problems
- Causes of common problems
- Administrative tools
- Storage monitoring tools

Data storage device problems

The ultimate safeguard against any data storage device failing is a formally defined and steadfastly followed data backup policy and procedure. A full system backup and incremental or differential backups are our insurance policy for any data storage device's failure. *Why do we need an insurance policy?* Well, we store one of our organization's most valuable assets (data) on what we believe is a reliable device. We then power it with one of the most unreliable and frequently failing devices in a PC—the power supply. If it avoids power issues, it's subject to viruses, physical damage, bad application software, and erroneous device drivers. On top of all that, there's the human. It's no wonder that data storage devices can develop problems.

In the sections that follow, we look at the common problems of three of the primary storage device types—**hard disk drives (HDDs)**, optical storage drives, and tape cartridge drives.

The following table lists some general causes of failure of data storage devices that you should know:

Problem	Cause	Chance of recovery
Deleted files, formatted drive	Human error	Good
Database or file corrupted or missing	Malware, software errors	Good
Filesystem corrupted	Media logical fault	Good
Can't access disk drive or garbled data	Media physical fault—heads, spindle motor, actuator	Fair to good
Device water-damaged	Natural disaster	Fair
Damaged device	External drive or laptop computer dropped	Fair
Flash media failed	Flash device overwrite limit exceeded*	Fair
Stored data garbled	Magnetic tape media degradation, failure, or damage*	Fair
Stored data unreadable	CD/DVD medium failure*	Poor

Common failure causes of data storage media

* The medium in some data storage units, such as flash, tape, and optical mediums, can and does wear out, and degrades over time to the point that anything stored on the medium becomes garbled, unreadable, or not there at all.

Common HDD problems

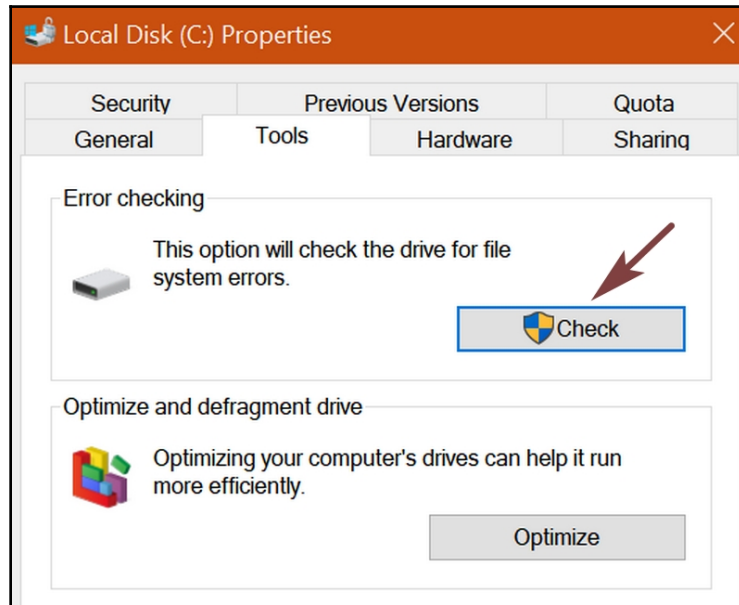
It can be difficult to identify certain HDD performance problems as such. The actions of a failing HDD can look very much like the symptoms of a malware attack, a memory problem, a failing power supply, or perhaps too many downloads running at the same time. The symptoms of a failing hard disk could be any of these issues and more. Some of the HDD problems you should know are as follows:

- **Operating system not found:** One of the first things the BIOS/UEFI does when a computer starts up is to access the OS on the system HDD, load its kernel to memory, and start up the system. This error says one of three things—there is a serious BIOS error, all or part of the HDD is faulty, or the **Master Boot Record (MBR)** is corrupted or missing.
- **Slow file access:** When opening a file for display, editing, printing, or any other use, if there seems to be a longer-than-usual pause or hesitation before the file shows up, the obvious problem is that the storage device has become slower. Actually, there could be several reasons why storage drives do take longer to access, read, and transfer data to memory and to the requesting program. The most common reasons for the I/O operations to slow down are as follows:
 - **Lack of sufficient space:** If an error message displays saying you're running out of disk space on a given drive as shown in the following screenshot, then the data storage drive indicated is almost filled to capacity. Many processes use secondary storage, and some are not able to execute properly if insufficient space is available for their needs:



Windows' low disk space warning

- **Corrupted data:** It's possible for a disk sector or cluster to become corrupted from several causes. Corrupted data on the hard disk could cause a slow data transfer speed. Use the **Error checking** tool, as shown in the following screenshot, on the HDD in question to search for corrupted data on that drive:



Windows Properties dialog box for a HDD showing the Error checking tool

- **File/data/object unavailable:** Although it's uncommon, files, data, and even a filesystem can suddenly disappear from a storage device—*apparently for no reason*. Well, for the sake of discussion, let's say that nothing happens in a computer by itself. Files, including documents, images, videos, programs, games, and so on, can become *unavailable, not found*, or some other way of saying, *it's gone*. There are several potential causes for this problem:
 - **Automatic operating system (OS) updates:** This is primarily a Windows issue, but automatically downloaded and applied upgrades, patches, or fixes may remove desktop files and installed applications, depending on their storage locations.
 - **Malware/virus:** Yes, these nefarious evil doers can and do delete or hide system, application, and personal files.

- **Login:** On occasion, users log in to their computers using different username credentials. An unavailable file may be there, but your other username may not have rights or permissions to see it.
- **HDD failing:** There is always the possibility that the HDD is logically or mechanically failing. Always assume this to be true and create a backup first thing.
- **HDD logical problems:** If a message appears, something like *Drive not available*, *Cannot access logical drive*, or *Unable to mount the device displays*, obviously there's a problem. There are several possible reasons why you may not be able to access a filesystem, file, or data on an HDD. Here are a few:
 - **Drive letter missing:** The drive designation, such as **C:** or **H:**, is missing from an HDD partition. This may be the result of a virus or an inadvertent action by a user.
 - **Hidden partition:** To protect personal or sensitive files, a user may decide to hide a partition. To access its contents, the filesystem must be unhidden.
 - **Partition inaccessible:** This condition indicates one of three possible causes—the access permissions for a particular partition are corrupted or gone altogether; the OS has detected an error in a value via the **cyclic redundancy check (CRC)**; or the **Master File Table (MFT)** or **File Allocation Table (FAT)** is corrupt and needs rebuilding.
 - **RAW partition:** A RAW partition is not in the format of a particular filesystem, such as NTFS, EXT4, or the like. A previously formatted filesystem can become RAW through a virus or result from an interruption during its formatting.
 - **Unallocated partition:** A power spike or surge could damage the logical partitioning of an HDD. A partition and its contents could change to unallocated.
- **Backup/restore problems:** A strictly followed backup program provides an insurance policy against catastrophic damage and loss. However, backup media can fail and become damaged. If the power fails or another event causes an interruption in the backup process, especially if the backup runs unattended, the resultant backup may be worthless. The discovery of a failed backup may not occur until it is needed to restore damaged or compromised files.

- **Cache failure:** In the context of data storage, caching refers to the disk cache, also known as the disk buffer or cache buffer, which is a small amount of disk space on an HDD. When the hard disk is busy, the disk cache receives and holds data for when the read/write mechanisms are available. A disk cache failure can occur from a power failure or any system error that halts the system, such as a kernel panic on a Linux or macOS system or the **Blue Screen of Death (BSOD)** on a Windows system. In any case, when the I/O operation interrupts, the cache empties and the data is lost.
- **Multiple HDD failure:** There are two ways you can look at this error condition—the failure of single HDDs installed in a computer one at a time, or an array of HDDs in a structure, such as **Redundant Array of Independent Disks (RAID)**, **Network-Attached Storage (NAS)**, or a **storage area network (SAN)**, all failing at once. In either situation, the first thing to suspect is the **power supply unit (PSU)**.
- **Status/error lights:** Most rackmount or tower servers have front-facing LEDs to indicate the activity, condition, and error status of their internal HDDs. Most HDD caddies and multiple-drive chassis devices, such as those for NAS or SANs, have lights for each bay or the ability to check each bay independently.

The following table shows a sampling of the meanings for an LED status light:

Pattern	Color	Meaning
Off	-	Empty slot
Steady	Green	Online
Flashing slow	Green	Drive administration in progress
Flashing fast	Yellow	Drive has failed
Flashing very slow	Alternating green/yellow	Drive failure imminent

Common HDD status light colors, patterns, and meanings

Causes of common problems

As you have most likely learned in your experience as a computing technician, the causes of most computer, software, and component failures, faults, and problems are relatively few in number. Identifying the source of a device or component problem typically involves looking for a familiar set of common system problems. In this section, we look at the list of common storage device problems you should expect to see on the Server+ certification exam.

Media failures

Storage media is something we take more or less for granted—but it is important. After all, that's where our data actually resides. However, since we cannot, and shouldn't, see, touch, or even smell, storage media for the most part, it's often a case of *out of sight, out of mind*.

Hard disk media

In spite of its rugged, metal-enclosed appearance, a hard disk drive, be it internal or external, is actually a relatively fragile device. Its internal components operate extremely close to each other, and at very high speeds. Any force or sudden motion to the system case or rackmount tray can, and usually does, cause damage to the disk and its storage medium. A crashed head that remains on the disk surface can scrape the medium material off the substrate, taking any data with it. One of the more common disk medium failures is the appearance of bad sectors on the disk. There are two types of bad sectors:

- **Physical bad sectors:** A portion of a hard disk medium damaged by a head crash or a contaminant inside the disk's case, such as dust, hair, and so on (typically from manufacturing). Neither the disk controller nor the OS is able to read or write to these areas. Physical bad sectors on an HDD are not generally repairable.
- **Logical bad sectors:** Unlike a physical bad sector, these areas of an HDD, also known as, soft bad sectors, are accessible but may cause I/O problems such as increased latency or read errors. The disk controller and OS can access a logical bad sector, and even write to it and attempt to read from it. Logical bad sectors are repairable.

SSD media

One of the advantages of an SSD over an HDD is the lack of moving parts. This means no head crashes or contaminant problems. However, SSD is an electronic device and, while its storage media is relatively error-free, the electronics around it aren't. Capacitors, circuits, and power supplies are known to fail, and if one or all of these that support the SSD do fail, then so does the SSD.

An SSD doesn't have bad sector problems, but it can have bad blocks. Here are a few of the symptoms to look for:

- The disk controller can't write to or read from the storage medium
- I/O-intensive applications may freeze up
- Copying or moving files causes system errors
- Accessing larger files takes longer than usual

Magnetic tape media

Magnetic tape, primarily in cartridges, has a limited, but very important, role in data integrity and system recovery, namely as a backup medium. A backup tape, regardless of its iteration, is necessary, but when something goes wrong with your main storage device, a backup becomes essential. Magnetic tape does have its problems, though. Here are the most common issues you may see on the exam:

- **Human errors:** Operators, technicians, and well-meaning, yet untrained, helpers are the causes of a majority of the problems of magnetic tape. Forgetting to change or load the tape, putting in the wrong tape for the current cycle, keeping tapes in service too long, and storing tapes in an improper location are just a few of the problems that humans create.
- **Script errors:** On some systems, rather than use a software utility, such as **Bacula** or **Duplicati** for macOS and Linux and **Windows Server Backup (WSB)** and **Acronis** for Windows, a script runs the system backups. Errors in the script, SAN or NAS addressing errors, or a folder not available can result in an incomplete backup that may not restore the system completely.
- **Hardware errors:** Tape drives and magnetic tape media can and do fail. In fact, one is often the cause of the other failing. A tape drive issue can affect the spooling of the tape medium, which can stretch, tear, or unravel the tape, usually outside the cartridge.

Optical drives

Optical drives, that is, CDs and DVDs, either work or they don't. Usually, there isn't any kind of a heads-up that the disc or the drive may be failing. It's typically a "*well, it worked yesterday*" kind of thing. The more common problems of an optical drive are as follows:

- **No drive at boot:** If the BIOS/UEFI doesn't *see* the optical drive during the boot process all of a sudden, chances are the drive's electronics are bad. Try rebooting, though. If that doesn't resurrect the drive, it's a goner.
- **Reads DVD, not CD:** If the drive reads a DVD with no problems, but won't read a CD at all, it's likely that one of the read lasers has failed. Try cleaning the disc and the drive and try again. If it's still a no-go, you'll need to replace the drive.
- **Read errors:** Typically, read errors indicate a dirty disc. Clean the disc and try again. If you still have problems, try another disc. If read errors continue, clean the drive.
- **Inconsistent operation:** If you have recently applied an upgrade or a patch to the OS, and since then the optical drive is operating erratically, then check with the manufacturer of the optical drive to see if a firmware update is available to fix the problem.

Common storage problems causes

Although we've discussed a few causes for storage device problems in the preceding section, let's focus now on specific causes of specific device or component failures. I've categorized the problem causes into the following groups—**drive and connector failures**, **controller and cache failures**, and **RAID and array failures**. The controller and cache failures and RAID and array failures are discussed in the *Hardware-related issues* section later in this chapter.

Drive and connector failures

The potential for storage device problems exists during almost every part of its installed life. In fact, just installing the device may introduce problems that may not be noticeable at first, but grow into serious issues. We discussed the problems a power supply can cause to storage devices and more, but even with general day-to-day levels of usage, there are several problems that can cause the storage device to be unavailable, perform erratically, or fail altogether.

HDD problems

Beyond human error, power supply issues, physical damage issues, and a few others, there are problems that can be the cause of problems with the OS, application software, and, okay, just about any program running on the computer that needs data from a storage device. The causes for the problems that fit this description are as follows:

- **Corrupt or missing MBR:** The MBR provides information to the boot process about the HDD on which it resides. When the boot process first accesses the HDD from which it will access the OS, it requests the MBR for that drive. The MBR includes information about the size of the disk, the number and type of logical partitions on it, and the like. However, if the MBR is corrupted or is missing, the boot process ignores the drive, which essentially says it doesn't exist. The three primary causes for an MBR to be missing or corrupt are a computer virus or malware, a program that has erroneously overwritten the MBR, that the drive has lost the clusters that include the MBR, or the drive has completely failed.
- **Corrupt filesystem table:** A part of the MBR is a master file table or a filesystem table that describes the logical partitions on the disk. The following table shows the contents of an MBR's filesystem table. Damage to the MBR will most likely also affect the filesystem table as well. A filesystem table looks something like the following:

Byte	Length (bytes)	Contents
0	1	Active boot indicator
1	3	Cylinder-Head-Sector (CHS) starting values
4	1	Partition type
5	3	CHS ending values
8	4	First sector
12	4	Partition size in sectors

The contents of the 16-byte filesystem table entry

- **Improper disk partition:** There are two types of partitions on a disk drive, namely, the system partition and one or more data partitions. The OS resides on the system partition, which is commonly designated as the **C:** drive. Data partitions store anything else and are designated as drives **D:**, **E:**, **F:**, and so on. Installing a lot of software, creating a large database, or saving data primarily to one partition can cause it to fill up. The message **Insufficient Disk Space**, which is common to all OSes, indicates that perhaps the partition sizing is incorrect for what the computer is being asked to support.

Cable and connector problems

On data center, rackmount, disks and disk arrays, a disk drive connects to the network through an Ethernet connector or **Serial Attached SCSI (SAS)** connection. The cables and connectors for these systems may have the same issues as standalone server connections, but typically on a much less frequent basis.

Standalone servers, whether free-standing (tower) or rackmount, generally have an onboard HDD of one form or another mounted inside the system case. These servers can also connect to SAN or NAS systems over the network. Storage drive failures do occur on systems of all sizes and technologies. The causes of these failures, regardless of the system, are essentially the same in all cases. These causes include the following:

- **Cable issues:** If the installation of the proper cabling for an internal storage drive is correct, the system should remain operational. If the system isn't recognizing the storage device, check the connectors and cabling for power or connectivity problems. If necessary, reseal the connectors or replace the cables. As a result of heating up and cooling down, a connector can wiggle out of a socket due to its metal contacts expanding and contracting. A faulty cable, even a new one, can cause intermittent problems or the boot process not seeing the device.
- **Improper termination:** This issue is one typically associated with SCSI and SAS devices and the *chain* that connects them. To prevent a signal from bouncing back onto the chain and creating *ghost devices*, properly terminate each end of a SCSI and SAS chain.
- **Serial Advanced Technology Attachment (SATA) connections:** Most of the time, SATA drives don't come with cabling. Match the cables to the specific drive manufacturer and model. The cabling should not be more than 1 m in length or intermittent data problems could occur. Make sure that all of the connectors are clean and snugly inserted.

Storage system issues

Data storage device failures fall into one of two general problem types—software-related failures and hardware-related failures.

Software-related failures

Data stored on a data storage device (that is, data at rest) can be like a proverbial *sitting duck*. Any number of factors can affect the data's quality, usability, and availability. Some common software-related or logical stored data issues are as follows:

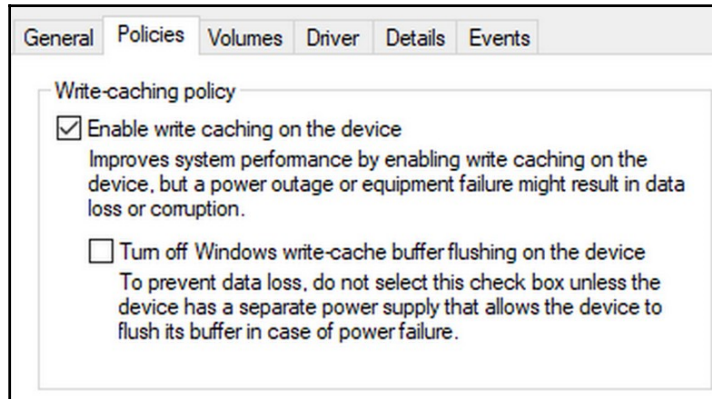
- **Formatting:** Formatting a data storage device, either purposely or accidentally, deletes the partition tables and the file location information. Any data stored on the device remains intact, but we can't access it because we've lost the map.
- **Deleted data:** When deleting a file, folder, or filesystem, regardless of whether you meant to or not, the data remains on the storage medium. In most cases, this data is recoverable, even though the OS can't find it. Deleting a filesystem removes the file indexing used to point to a folder, or file within it. However, in most cases, this data is recoverable.
- **Corrupted data:** Bad programming instructions, virus or malware infections, and hardware issues can be a cause of corrupted data.

Hardware-related issues

Because of its close operating distances, an HDD can be almost fragile, meaning it's relatively easy to damage it. Most data storage device problems occur during system startup. These problems are generally in the hard disk controller, but may also be in HDDs, as well. However, other forms of data storage devices can experience most of these problems. You should expect to see questions regarding HDD problems and their effect on stored data. The following are issues you may run into:

- **Controller failure:** On SATA and IDE HDDs, the disk controller, which may be in the disk housing or on the motherboard as a function of the chipset, could fail, which eliminates the capability of the OS to read and write from the disk medium. The HDD controller provides a go-between service for the OS and the HDD device. Commonly, although the message that's displayed indicates that the controller failed, it could very well be the HDD device itself.
- **Host bus adapter (HBA) failure:** An HBA is an electronic circuit board that relieves the CPU of the communications and actions with the data storage system, which may be a **fibre channel (FC)** SAN, NAS, RAID, or onboard data storage devices. A failure of an HBA controller can indicate installation issues, a failed component on the card, or a failure in the SAN controller.

- **Disk cache off:** By default, the disk-buffering or disk-caching function is enabled to assist an HDD to receive data from the system for writing to the disk while it's busy with other tasks, as shown in the following screenshot. If you disable this feature on an internal HDD, I/O speeds may be slow. On external drives, disabled disk caching is the default:



The write-caching policy settings in the Windows Device Manager

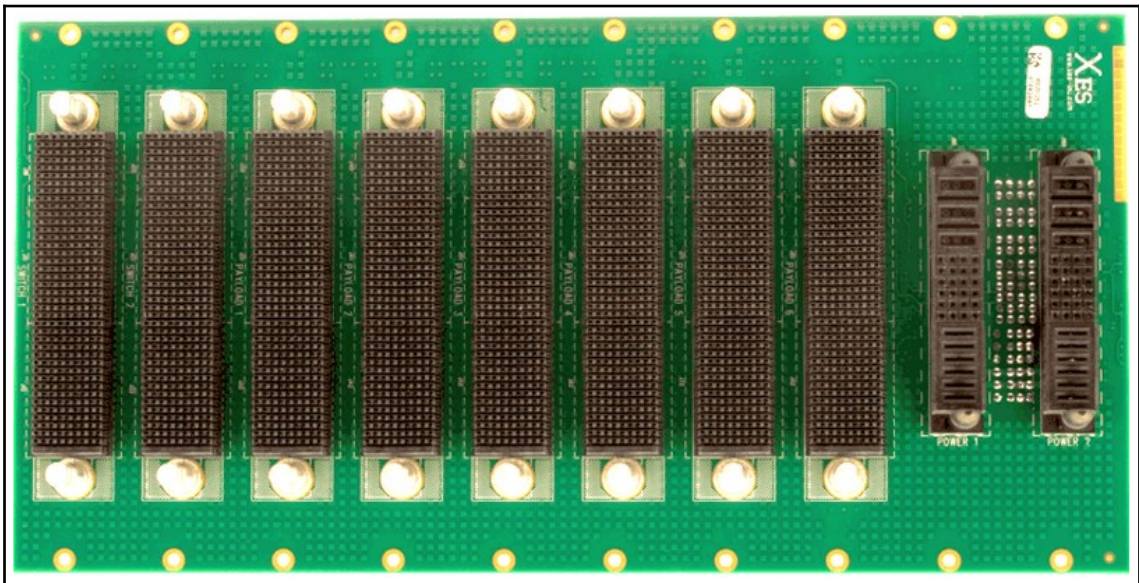
- **Cache battery failure:** RAID systems power their write cache memory with a battery so that, should the power fail, the system is able to clear the cache before shutting down. Any data in the cache when the cache battery's power weakens will be gone.

Storage array issues

The technologies and mechanisms available for implementation on a network to improve data I/O speed, reliability, redundancy, and availability for network users typically are relatively problem-free. However, improper configuration, incompatible devices, or hardware issues may defeat the purpose of the array. Specific issues that can arise are as follows:

- **Improper RAID configuration:** RAID is a disk storage technique that creates a network of HDDs that coordinate to prevent data loss while improving reliability, performance, and redundancy. However, if a RAID configuration is faulty, the system could work against itself. If performance is the priority, it could result in a data loss in the event of a failure. The reverse would be true as well. The configuration of the RAID arrangement should match the objectives of the system and the goals of the organization.

- **RAID rebuilds:** Should one of the storage devices in a RAID system fail and a replacement drive is available, the RAID array will reconstruct the piece that is now missing on the new drive based on its parity data. Unfortunately, while this is happening, users see longer-than-normal latency.
- **Mismatched drives:** In some data storage arrangements, such as NAS or SAN, differently sized drives work as well as drives that are identical. However, RAID needs the disk drives to all be the same size. If differently sized drives are in the RAID array, all drives, regardless of size, will only use the storage size of the smallest drive.
- **Backplane failure:** A common method used to implement a RAID system is through a backplane circuit or card, as shown in the following image. A backplane is often part of a card enclosure. Each disk drive unit connects into a slot or connector on the backplane, which eliminates the cabling clutter of multiple drives. Should a single backplane connector fail, its drive becomes inaccessible. Should the entire backplane fail, none of the drives connected to it are accessible. Typically, the drives themselves are okay though:



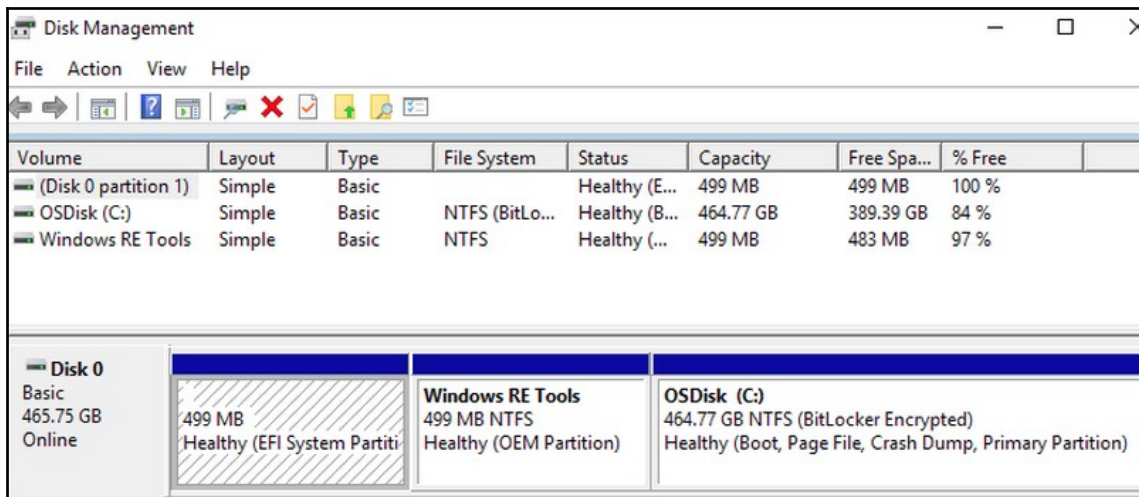
A data storage device backplane card
Image courtesy: kisspng.com

Administrative tools

The complexity of administering the data storage connected to a server depends on several factors, including size, technology, arrangement, and—let's not forget—security. Based on the complexity, the job of the storage administrator can range from the part-time duties of a single administrator to a full-time job for multiple technicians. There are a variety of disk or storage management software packages available that include the functions necessary to perform most of the tasks described in the following sections.

Disk management

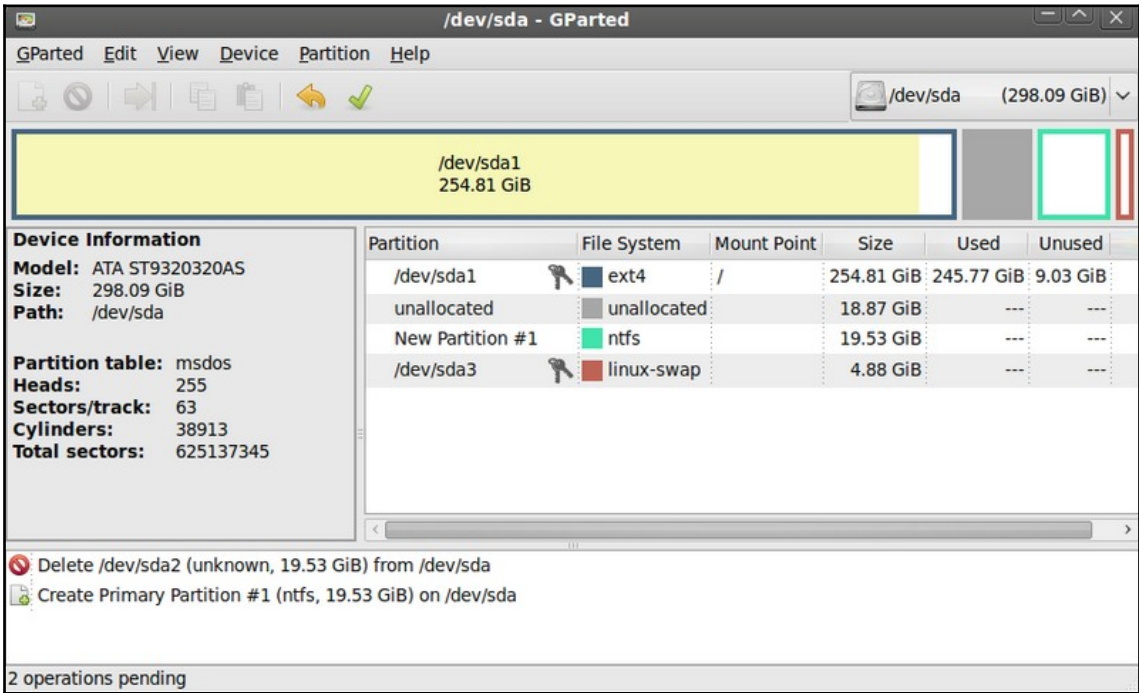
The term **disk management** has a few connotations, but it's most commonly used to describe a task or a utility feature of the Windows OS. Actually, the latter can help you with the former. The Windows **Disk Management** utility, as shown in the following screenshot, provides the capabilities to create, change, remove, extend, and rename disk volumes and partitions. For Linux systems, the `fdisk` and `disk` commands provide many of the same features:



The Windows Disk Management utility

Disk partitioning tools

The Windows **Disk Management** utility, as shown in the preceding screenshot, manages disk partitions. However, there are also many freeware and not-so-freeware packages available for this purpose, including the **MiniTool Partition Wizard** and the **AOMEI Partition Assistant**. On a Linux system, the `fdisk` command has been the standard, but **GParted**, as shown in the following screenshot, is an excellent open source tool for this purpose, as well:



GParted partitioning tool for Linux
Image courtesy: gparted.org

Map, mount, and net use

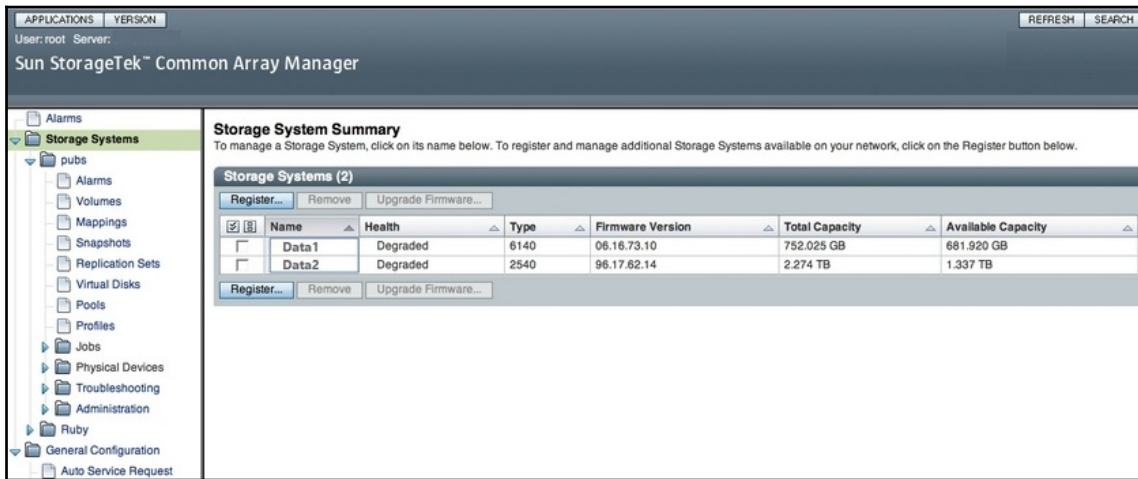
The process of mapping a network resource is a much different action from mounting a storage device or connecting to or creating a share on a network resource. But, these three functions are often confused. Let's look at each one separately:

- **Drive mapping:** Mapping a data storage resource involves identifying it by assigning it a drive identity, such as E:, and making it available to local network users. In effect, drive mapping creates a share, but on a broader scale than a resource on one node shared with a specific user. However, in most cases, adding a shortcut to the resource in the Windows File Explorer's **Network** folder may be a better option.
- **Mount command:** Both Windows and Linux use the `mount` command to make a storage device, a filesystem, or a group of files accessible to one or more users and connecting it to the active directory structure. However, the Linux `mount` command has a few more options that can make it the equivalent of the Windows `net use` command.
- **Net use:** The `use` option is just one of over 20 options that facilitate the management of just about any part of a network, including setting network shares, administering users and their permissions, and so on. The `net use` command can create, configure, or remove connections to shared resources on a network, including data storage drives and printers. The `net use` command produces essentially the same results as the **Map Network Drive** option in the Windows File Explorer.

Disk arrays

A disk array, also known as a storage array, combines several HDDs into a single data storage system that operates independently of network servers. Storage arrays also centralize storage devices for a single resource management function. Storage arrays provide the foundation for SANs and NAS.

Most disk array system manufacturers provide a form of storage management software. In addition, there are third-party storage management software packages:



An example of common disk array management software

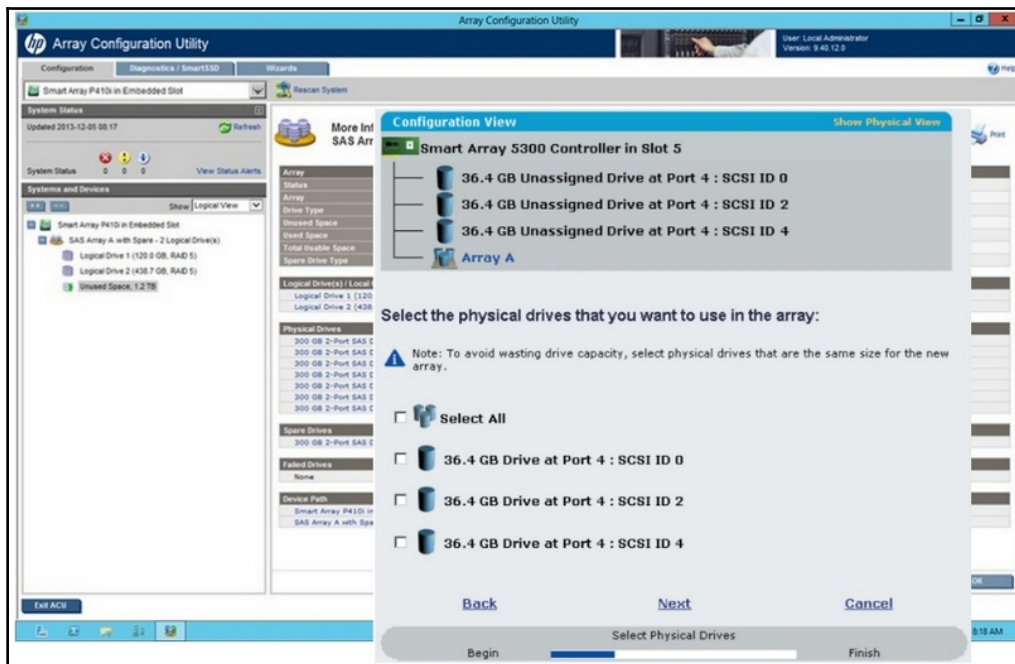
RAID arrays

Another type of storage device array is a RAID array, and, as with general disk arrays, there are specific software management tools for RAID implementations. On a Linux system, RAID management is through the `mdadm` command. The following screenshot shows the help display for this command:

```
mdadm is used for building, managing, and monitoring
Linux md devices (aka RAID arrays)
Usage: mdadm --create device options...
        Create a new array from unused devices.
mdadm --assemble device options...
        Assemble a previously created array.
mdadm --build device options...
        Create or assemble an array without metadata.
mdadm --manage device options...
        make changes to an existing array.
mdadm --misc options... devices
        report on or modify various md related devices.
mdadm --grow options device
        resize/reshape an active array
mdadm --incremental device
        add/remove a device to/from an array as appropriate
mdadm --monitor options...
        Monitor one or more array for significant changes.
mdadm device options...
        Shorthand for --manage.
Any parameter that does not start with '-' is treated as a device name
or, for --examine-bitmap, a file name.
The first such name is often the name of an md device. Subsequent
names are often names of component devices.
```

The help display for the mdadm Linux command

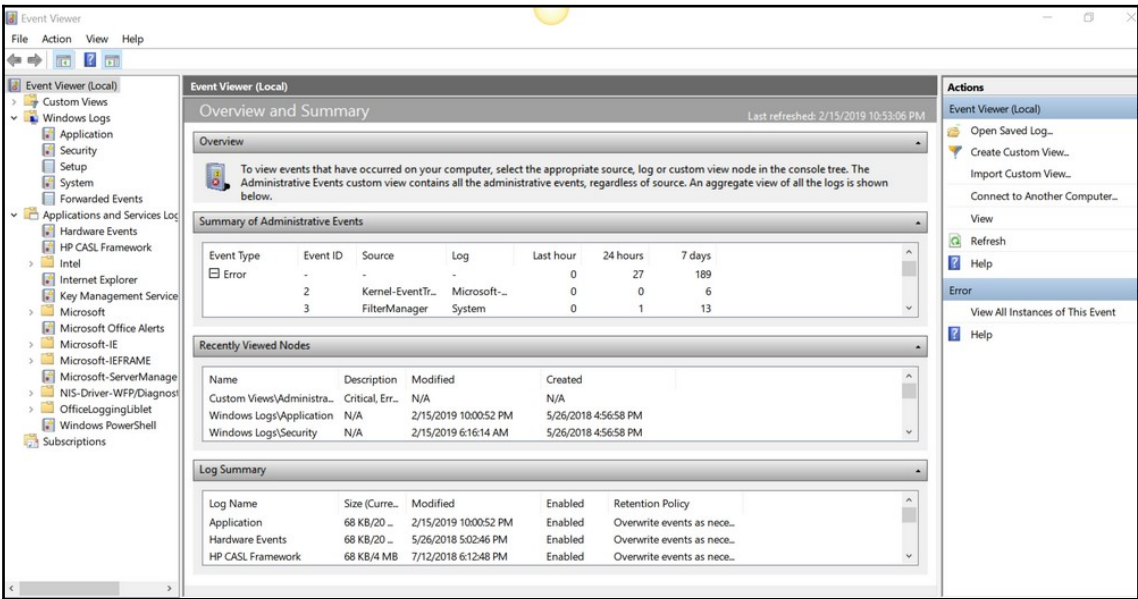
On Windows Server and other 64-bit Windows versions, utilities such as the **HP Array Configuration Utility**, shown in the following screenshot, support the creation, administration, and monitoring of RAID and other disk arrays:



The HP Array Configuration Utility is an example of RAID array management software

Storage monitoring tools

All OSeS record informational, cautionary, and error-alarm events in system log files. Windows has a robust log system. The Windows **Event Viewer** utility provides a tool to view, manage, archive, and customize the content and display of each of the log files. The following screenshot shows a screen capture of the Windows **Event Viewer**:



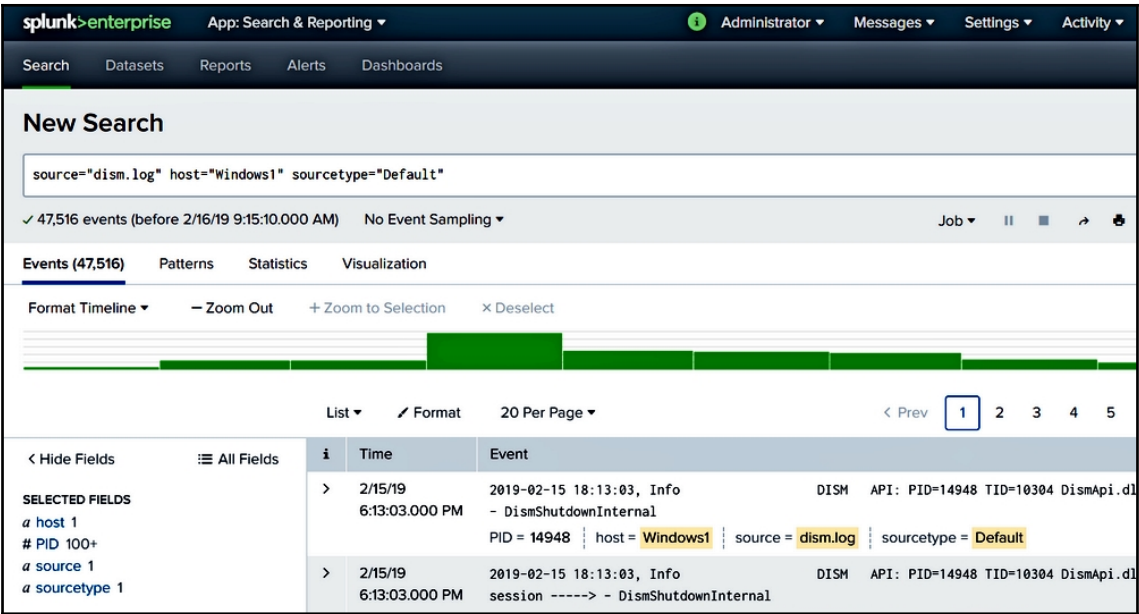
The Windows Event Viewer displays the contents of the system's log files

On a Linux system, the log files are in the `/var/log` directory. The following screenshot shows a display of the log entries in the `sysinfo.log` file, which contains entries for the general activities of the system:

```
2019-02-02 12:49:07,285 ERROR    Network plugin raised an exception.
Traceback (most recent call last):
  File "/usr/lib/python3/dist-packages/landscape/sysinfo/sysinfo.py", line 99, in run
    result = plugin.run()
  File "/usr/lib/python3/dist-packages/landscape/sysinfo/network.py", line 32, in run
    for info in self._get_device_info():
  File "/usr/lib/python3/dist-packages/landscape/lib/network.py", line 181, in get_active_device_info
    speed, duplex = get_network_interface_speed(sock, interface)
  File "/usr/lib/python3/dist-packages/landscape/lib/network.py", line 261, in get_network_interface_speed
    raise e
  File "/usr/lib/python3/dist-packages/landscape/lib/network.py", line 253, in get_network_interface_speed
    fcntl.ioctl(sock, SIOCEHTOOL, packed) # Status ioctl() call
OSError: [Errno 22] Invalid argument
2019-02-02 12:58:57,199 ERROR    Network plugin raised an exception.
Traceback (most recent call last):
  File "/usr/lib/python3/dist-packages/landscape/sysinfo/sysinfo.py", line 99, in run
    result = plugin.run()
  File "/usr/lib/python3/dist-packages/landscape/sysinfo/network.py", line 32, in run
    for info in self._get_device_info():
  File "/usr/lib/python3/dist-packages/landscape/lib/network.py", line 181, in get_active_device_info
    speed, duplex = get_network_interface_speed(sock, interface)
  File "/usr/lib/python3/dist-packages/landscape/lib/network.py", line 261, in get_network_interface_speed
    raise e
  File "/usr/lib/python3/dist-packages/landscape/lib/network.py", line 253, in get_network_interface_speed
    fcntl.ioctl(sock, SIOCEHTOOL, packed) # Status ioctl() call
OSError: [Errno 22] Invalid argument
2019-02-02 13:00:09,923 ERROR    Network plugin raised an exception.
Traceback (most recent call last):
  File "/usr/lib/python3/dist-packages/landscape/sysinfo/sysinfo.py", line 99, in run
    result = plugin.run()
  File "/usr/lib/python3/dist-packages/landscape/sysinfo/network.py", line 32, in run
    for info in self._get_device_info():
  File "/usr/lib/python3/dist-packages/landscape/lib/network.py", line 181, in get_active_device_info
    speed, duplex = get_network_interface_speed(sock, interface)
sysinfo.log
```

The contents of a Linux `sysinfo.log` file

In addition to the onboard log file utilities, several third-party systems are available for log file monitoring, analysis, and reporting. The following screenshot shows an event log analysis tool. Monitoring a server's log files is an important task. As the saying goes, *the devil is in the detail*, and it's certainly the case with log files:



The Splunk Enterprise log file analysis tool

Summary

The safeguard against storage device failures and data loss is a formal data backup procedure. System backups provide insurance against storage device failure. Some HDD problems you should know are *OS not found*, *slow file access*, *lack of sufficient space*, *corrupted data*, and *file/data/object unavailable*.

A common disk failure is physical or logical bad sectors appearing on a disk platter. Physical bad sectors are the result of a read/write head touching the disk medium surfaces. Logical bad sectors may cause I/O problems, including increased latency or read errors. An advantage of an SSD is its lack of moving parts, but its electronics can fail and it can have bad blocks. Magnetic tape cartridges are a common backup medium. Backup tapes are essential. Magnetic tape does have problems, including human errors, script errors, and hardware errors. The common problems of optical drives include no drive being detected at boot, reads a DVD—but not a CD, read errors, and inconsistent operation.

The causes for device or component failures fall into four groups—drive and connector failures, controller and cache failures, and RAID and array failures. The causes for HDD drive problems include a corrupt or missing MBR, a corrupt filesystem table, and an improper disk partition. Causes for cable and connector problems include cable issues, improper termination, and SATA connection issues.

The causes for storage system problems can be software-or hardware-related. The causes for software failures include formatting, deleted data, and corrupted data. The causes of hardware-related failures include controller failure, HBA failure, the disk cache being off, and a cache battery failure. The causes for storage array problems include improper RAID configuration, RAID rebuilds, mismatched drives, and backplane failures.

Storage management software provides the functionality to perform the following tasks of disk management: managing partitions, mapping a network resource, mounting filesystems, and adding, configuring, and removing partitions.

A storage array combines storage drives into a single data storage unit that operates independently of the network servers. Storage arrays are the foundation of SANs and NAS. Disk array manufacturers provide storage management software. There are specific software tools for RAID implementations. OSes record events in log files.

Questions

1. Which of the following is not a common problem of an HDD device?
 1. Slow access
 2. Corrupted data
 3. Too fast
 4. Bad sectors or clusters
2. Which of the following data storage devices has no storage-related moving parts?
 1. HDD
 2. SSD
 3. DVD
 4. CD
3. What does the acronym HBA stand for?
 1. Hierarchical Binary Array
 2. Host Base Algorithm
 3. Host Bus Adapter
 4. Hierarchical Bus Adapter
4. The most commonly accepted meaning of the acronym RAID is which of the following?
 1. Responsive arrays of independent disks
 2. Redundant arrays of independent disks
 3. Reproduced arrays of independent disks
 4. Readable arrays of independent disks
5. Which of the following is a common cause of backup failures? (Choose all that apply)
 1. Tape drive errors
 2. Human error
 3. OS errors
 4. Disk errors

-
6. Which of the following is not a type of device or component failure?
 1. Drive failures
 2. Connector failures
 3. Controller failures
 4. Cache failures
 5. RAID and array failures
 6. Cable termination failures
 7. All of the above
 8. None of the above
 7. The information needed by the boot process about a hard disk partition is found in which data block?
 1. Registry
 2. MBR
 3. Filesystem
 4. Disk buffer
 8. On a Windows system, which system utility can be used to view the contents of the system log files?
 1. Task Manager
 2. Device Manager
 3. Control Panel
 4. Event Viewer
 9. Which logical structure combines storage devices into a single unit that operates independently of network servers?
 1. Data array
 2. iSCSI chain
 3. Storage or disk array
 4. Database

19

Common Security Issues

Data security requires constant attention and scrutiny to ensure that stored data, data in transit, and data in use is under the **confidential, integrity, and availability (CIA)** model. In this chapter, we look at the security issues, problems, and consequences of gaps in or the absence of safeguards to protect a network's data, software, and hardware resources. The problems, causes, and tools discussed in this chapter are the topics and concepts you should expect to see on the Server+ exam. The general topics covered in this chapter are as follows:

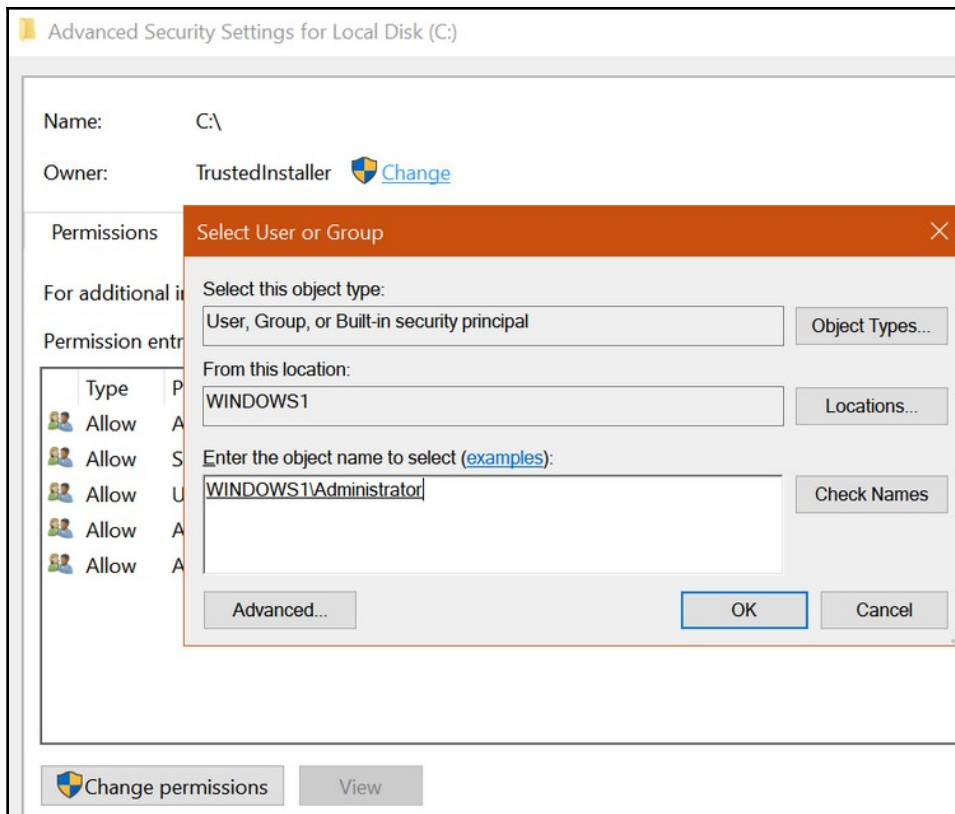
- Common data security problems
- Causes of common security problems
- Commonly used security tools and their use

Common data security problems

Network servers, regardless of their application, must deal with a variety of security issues, attacks, and, in some cases, harm. These problems aren't specifically related to hardware or software, but they are problems. The security problems you should understand include the following:

- **File integrity:** File integrity means that data and program files have not been the target of an attack and modified in any way by an unauthorized person or function. An attacker, either internal or external, can compromise the integrity of a file purposefully or accidentally. Monitoring the integrity of stored data and programs should be a part of a standard monitoring procedure. **File integrity monitor (FIM)** software uses checksums and hashing to detect changes to a file. The most common hashing algorithm used is **Message Digest 5 (MD5)**.

- **Privilege escalation:** This is an attack on the resources of a system or network in which the attacker is able to expand or elevate his or her rights and permissions to access resources with higher restrictions. There are two types of privilege escalation:
 - **Vertical:** This is the nominal privilege escalation, also known as, privilege elevation, in which a user (attacker) or application (malware) with low assigned privileges is able to access data, programs, or other content with higher privileges, rights, and permissions.
 - **Horizontal:** In this type of privilege escalation attack, a user increases his or her permissions and rights so that he or she is able to access resources reserved or assigned to another user or user group.
- **Applications will not load:** In the security settings of a Windows system, the ownership of a resource can affect the performance of an application from that resource. One common issue is that the Windows Update service isn't running. As suspicious as that may be, the problem could also be the ownership of a resource. The ownership designation of a partition, folder, or file controls who and what can read, write, or execute its contents. To change the ownership, access **Properties** on the hard disk from File Explorer (see the following screenshot). In addition, malware may have removed, renamed, or corrupted the disk, partition, or file. On a Linux system, the issue is likely the same and the remedy is too. Change the ownership of the filesystem, directory, or file as follows:



Changing the ownership of a disk drive

- **Cannot access network file/shares:** Somewhere in the chain of objects, elements, and devices between a network server and a host, the request to access a shared resource is failing. This means that the network adapter on either computer, the permissions of the remote user on the server, the permissions of the resource, and whether or not the resource is actually set as a share could be the issue.
- **Unable to open files:** This error and other similar messages are generally the result of the following:
 - Problems with the file, application, or service itself (see the following screenshot)
 - A third-party application that is not quite compatible with the operating system
 - A newly installed application or file that is in conflict with the anti-malware software

On a Windows system, there are a few more potential causes for this error:

- **User Account Control (UAC)** is not enabled.
- Certain applications will not open for the **Built-in Administrator (BIA)** account; log off and change the logged-in account.

On a Linux or a macOS system, the issues are likely a conflict of installed services:



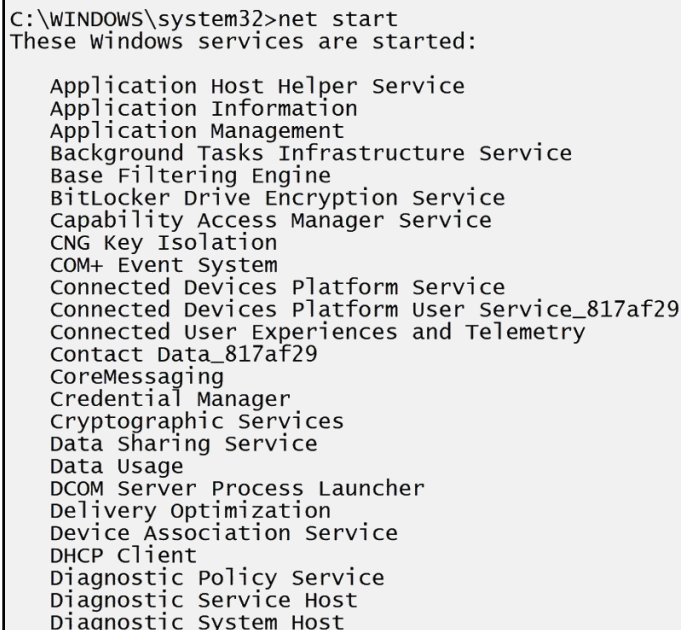
A file blocked for opening

- **Excessive access:** One of the core principles of resource access is the principle of least privilege, which says a user should only have access to the resources absolutely needed to complete or carry out their assigned task. If a particular user account, especially a remote user account, is accessing a specific data resource beyond reasonableness, it may be something to look into. The log files should provide the information needed to make a judgement on what may be excessive access. Another consideration is to verify that the access is, in fact, made by a person.
- **Excessive memory utilization:** High memory consumption can be an indication of a security issue, such as malware or an unauthorized user accessing data, or it could be excessive memory leaks. On a Windows system, use the Windows Memory Diagnostic utility to see what is taking up space in RAM. On Linux or macOS, the `syslog` log file may contain information on out of memory conditions and the `free -m` command enables you to see the allocation of memory usage.

Causes of common security problems

When you detect a security event that is underway, there is only so much you can do to stop it, block its damage, and mitigate the exploited vulnerability immediately. After recovering or rectifying the damage, your next task is to identify the vulnerability and the cause of the exploitation. The cause of common security problems is an area where the Server+ exam places some emphasis, so expect to see questions relating to the topics in the following list.

- **Active services:** Operating systems all start a group of services when they boot up and these services may start up other services (dependencies). A part of the security procedures on any server should be a periodical audit of the services actively running on it. The services audit should also note which TCP/UDP ports are in use and by which services. More active services can mean more vulnerabilities and threats. The information gained in the services audit can also help to structure the network's firewall and router rules. On a Windows system, list the active services using the `net start` command (see the following screenshot):



```
C:\WINDOWS\system32>net start
These Windows services are started:

Application Host Helper Service
Application Information
Application Management
Background Tasks Infrastructure Service
Base Filtering Engine
BitLocker Drive Encryption Service
Capability Access Manager Service
CNG Key Isolation
COM+ Event System
Connected Devices Platform Service
Connected Devices Platform User Service_817af29
Connected User Experiences and Telemetry
Contact Data_817af29
CoreMessaging
Credential Manager
Cryptographic Services
Data Sharing Service
Data Usage
DCOM Server Process Launcher
Delivery Optimization
Device Association Service
DHCP Client
Diagnostic Policy Service
Diagnostic Service Host
Diagnostic System Host
```

A portion of a list of the active services on a Windows system

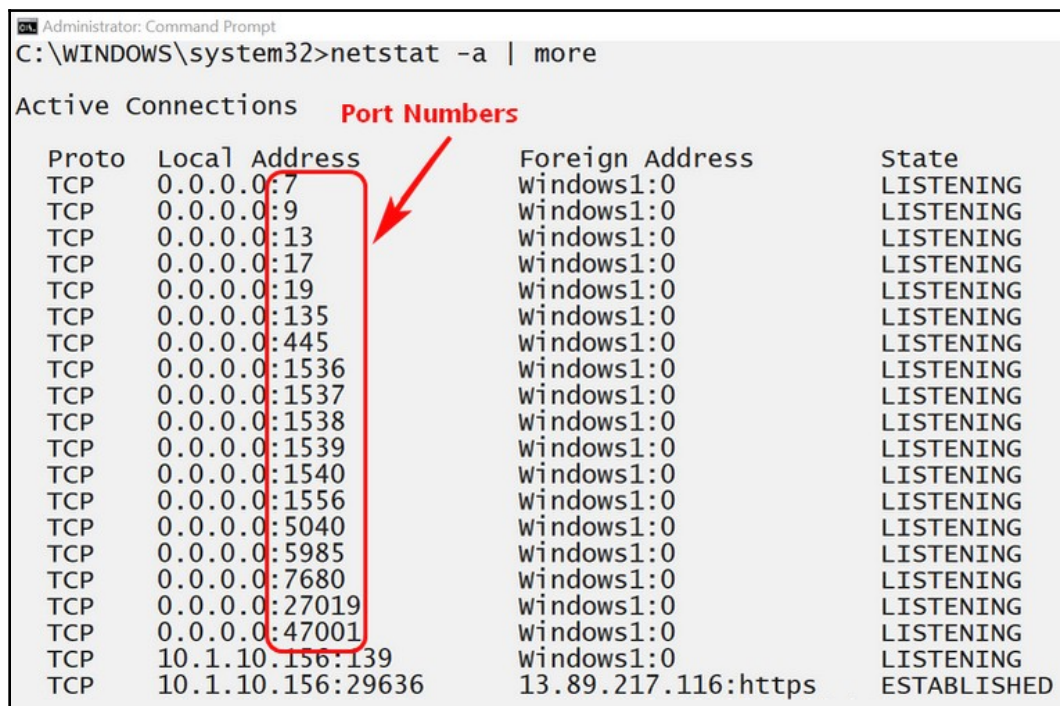
On a Linux system, the `service --status-all` command lists active services (see the following screenshot):

```
rprice@windows1:~$ sudo service --status-all | more
[ - ] acpid
[ - ] apparmor
[ ? ] apport
[ - ] atd
[ - ] console-setup.sh
[ - ] cron
[ ? ] cryptdisks
[ ? ] cryptdisks-early
[ - ] dbus
[ - ] ebttables
[ ? ] hwclock.sh
[ + ] irqbalance
[ - ] isc-dhcp-server
[ + ] iscsid
[ - ] keyboard-setup.sh
[ - ] kmod
[ - ] lvm2
[ + ] lvm2-lvmetad
[ + ] lvm2-lvmpolld
[ - ] lxcfs
[ - ] lxd
[ - ] mdadm
[ - ] mdadm-waitidle
[ + ] open-iscsi
[ - ] open-vm-tools
[ ? ] plymouth
[ ? ] plymouth-log
[ - ] procps
[ - ] rsync
[ - ] rsyslog
[ - ] screen-cleanup
--More--
```

The active services on a Linux system

- **Inactive accounts:** There is a big difference between an inactive account and a disabled account. An inactive account has not signed on or been active in a defined period of time. A disabled account is not available without being enabled. Administrators should monitor inactive accounts closely because they are usable accounts that could be a wide-open vulnerability door.
- **Anti-malware configurations:** Beyond choosing the best antivirus or anti-malware software for the organization's computing and networking needs, the configuration of that software is as important as the rules applied to a firewall and router, maybe more. In a networked environment, it is a server-based, anti-malware system that scans and protects all of the network's nodes automatically. But, the quality of that protection is a direct product of the anti-malware system's configuration. At minimum, the anti-malware's configuration should address the following issues:
 - **Schedule:** A full scan of all computers on the network every day. This includes the servers, of course, but also the network hosts. Schedule scans when the least number of humans are on the network.
 - **Updates:** Update malware databases or signature files automatically from the publisher's website as soon as the updates are available.
 - **Devices:** All allowed removable data storage devices should be as much a part of the anti-malware scan as possible.
 - **Review:** Review all anti-malware log files daily for alerts, false negatives, false positives, and other configuration-related errors.
- **Misconfigured permissions:** On larger networks, it's common for user account permissions to become conflicting with the permissions of the groups to which the user is a member. In either case, group or user, the application of the principle of least privilege can prevent the users of a group or any individual users having access to more than they need.

- **Open ports:** An open port is one that is enabled and *listening* for incoming IP addresses and port number combinations (sockets). Each open port is essentially an open door that an intruder could exploit. Hardening a server includes disabling all unnecessary open ports. The following screenshot shows `netstat` listing the active ports on a Windows system:

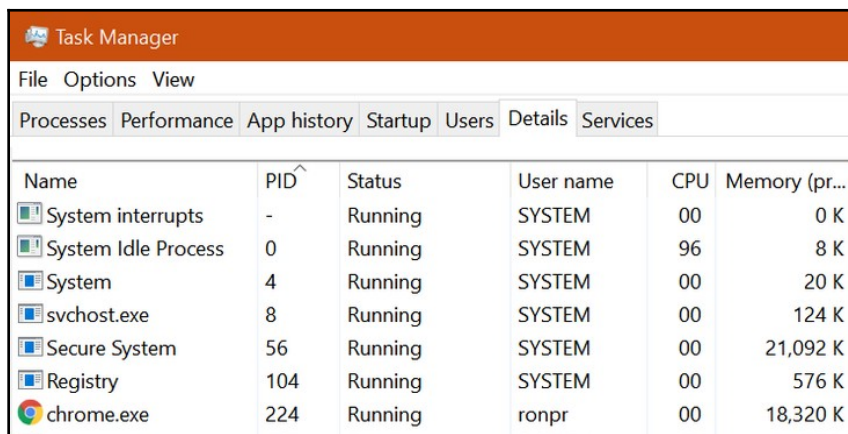


```
Administrator: Command Prompt
C:\WINDOWS\system32>netstat -a | more

Active Connections  Port Numbers
Proto Local Address Foreign Address State
TCP 0.0.0.0:7 Windows1:0 LISTENING
TCP 0.0.0.0:9 Windows1:0 LISTENING
TCP 0.0.0.0:13 Windows1:0 LISTENING
TCP 0.0.0.0:17 Windows1:0 LISTENING
TCP 0.0.0.0:19 Windows1:0 LISTENING
TCP 0.0.0.0:135 Windows1:0 LISTENING
TCP 0.0.0.0:445 Windows1:0 LISTENING
TCP 0.0.0.0:1536 Windows1:0 LISTENING
TCP 0.0.0.0:1537 Windows1:0 LISTENING
TCP 0.0.0.0:1538 Windows1:0 LISTENING
TCP 0.0.0.0:1539 Windows1:0 LISTENING
TCP 0.0.0.0:1540 Windows1:0 LISTENING
TCP 0.0.0.0:1556 Windows1:0 LISTENING
TCP 0.0.0.0:5040 Windows1:0 LISTENING
TCP 0.0.0.0:5985 Windows1:0 LISTENING
TCP 0.0.0.0:7680 Windows1:0 LISTENING
TCP 0.0.0.0:27019 Windows1:0 LISTENING
TCP 0.0.0.0:47001 Windows1:0 LISTENING
TCP 10.1.10.156:139 Windows1:0 LISTENING
TCP 10.1.10.156:29636 13.89.217.116:https ESTABLISHED
```

List the active ports using `netstat`

- Rogue processes/services:** A rogue process or service is one that is running on a computer that the user or operator did not initiate and is consuming resources, causing destruction, and performing mayhem. Regardless of what it's doing, kill it, but not in a physical, bloody way; in a humane, logical way. The key to killing it is to get its **process ID (PID)** first. In Windows, use the Task Manager (where you can also try to end it all) and in Linux or macOS, the `ps` command lists this information. The following screenshots show the PID numbers in the Windows Task Manager and a Linux `ps` command, respectively. With the PID, you can use the `taskkill /PID <PID>` command in Windows (at the command line), or the `kill` command in Linux/macOS:

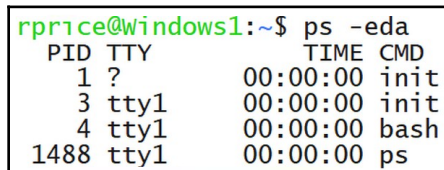


The screenshot shows the Windows Task Manager window with the 'Processes' tab selected. It displays a list of running processes with columns for Name, PID, Status, User name, CPU, and Memory (private).

Name	PID	Status	User name	CPU	Memory (private)
System interrupts	-	Running	SYSTEM	00	0 K
System Idle Process	0	Running	SYSTEM	96	8 K
System	4	Running	SYSTEM	00	20 K
svchost.exe	8	Running	SYSTEM	00	124 K
Secure System	56	Running	SYSTEM	00	21,092 K
Registry	104	Running	SYSTEM	00	576 K
chrome.exe	224	Running	ronpr	00	18,320 K

The Windows Task Manager

The following screenshot shows the output of the Linux `ps` command:



The screenshot shows a terminal window with the command `ps -eda` executed. The output displays process details including PID, TTY, TIME, and CMD for the current session and its parent processes.

```

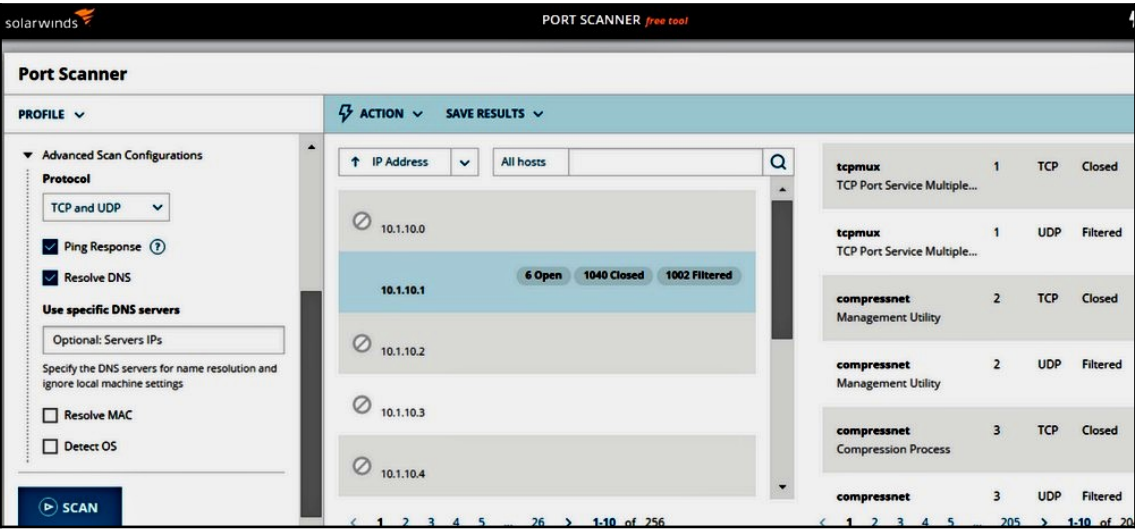
rprice@windows1:~$ ps -eda
  PID TTY          TIME CMD
    1 ?           00:00:00 init
    3 tty1        00:00:00 init
    4 tty1        00:00:00 bash
 1488 tty1        00:00:00 ps
  
```

The Linux `ps` command

Security tools

Like any skilled professional, such as a plumber, carpenter, or mechanic, a system administrator has a toolbox of tools specially made for the tasks and processes required to carry out his or her tasks. The following list shows the tools you need to be familiar with for the Server+ exam:

- **Port scanners:** This is a software application that scans a server (or host) to identify any open TCP/UDP ports. An open port is a possible entry point for an external attacker. Port scanners allow server administrators to ensure that the state of the server complies with the security policies of its organization. Not all port scanners scan for the full range of TCP/UDP port numbers. Some scan for only the well-known ports, others scan for the ports most commonly exploited, and yet others scan for the full range of port numbers (up to 65536). A port scan assigns one of three states to each port—**Open** (accepted), **Closed** (denied), or **Filtered** (blocked). The open ports pose the highest security vulnerabilities. The following screenshot shows an example of a port scanner utility:



The SolarWind port scanner

- **Sniffers:** Also known as packet sniffers, **sniffers** are utility software that examine the contents of network packet traffic. A sniffer's role depends on who is using it. System and network administrators use sniffers to monitor incoming network traffic, which could be part of an intrusion detection or prevention objective. Attackers use sniffers to obtain passwords, usernames, account numbers, and other **personally identifiable information (PII)**. Sniffers can be hardware appliances or software programs. A **snoop server** is a sniffer configured in promiscuous mode, which captures and examines all network traffic. Only the destination node examines a packet in non-promiscuous mode. Sniffing can also be active or passive.
- **Cipher:** In the context of server security, this term refers to an algorithm used to convert plain text into cipher text. In other words, the process used to encrypt or decrypt data. There are two basic types of ciphers—**transposition ciphers** and **substitution ciphers**. A transposition cipher shifts the characters of a string to hide their original order. For example, *ABCDE* may become *EACBD*. More sophisticated transposition ciphers may shift the bits of the string's characters.

A substitution cipher, at its simplest, assigns each character an alias value that replaces the original character in a string. The following screenshot shows an example of a substitution cipher:

Plaintext message: THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.																									
Substitution key:																									
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
G	A	B	S	L	Y	T	E	X	U	C	F	H	I	J	K	Z	M	N	O	P	Q	R	D	V	W
Ciphertext string: OEL ZPXBC AMJRI YJD UPHKN JQLM OEL FGWV SJT.																									

An example of the use of a substitution cipher

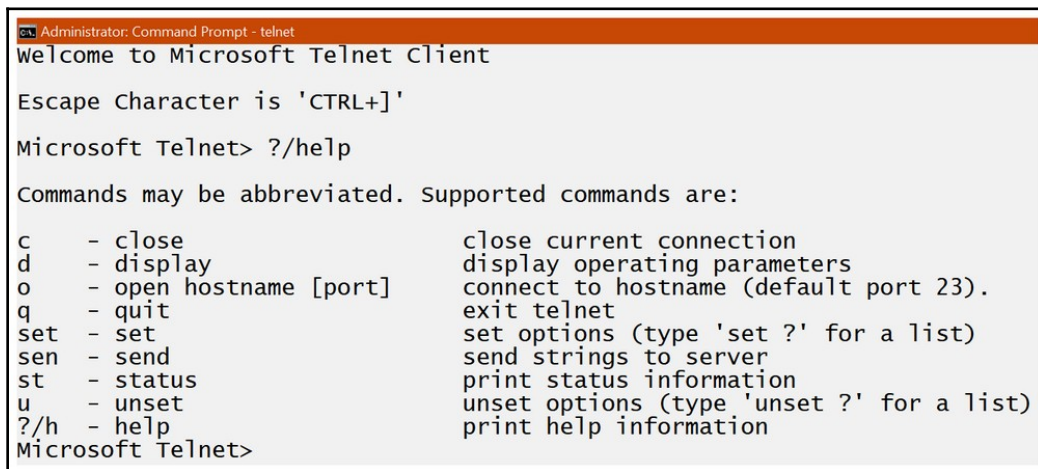
- **Checksums:** This is a value generated by an algorithm that consists of a sequence of characters that represents the contents of the original string. A checksum calculated on a copy of an original file, packet, string, or any other digital object should match the original checksum, indicating they are identical. The checksum algorithms in common use are the MD5, **Secure Hash Algorithm (SHA)**-1, SHA-256, and SHA-512. The numbers of these names refer to different things. The 5 in MD5 is a version number, as is the 1 in SHA-1. The numbers in SHA-256 and SHA-512 represent the numbers of bits in the signature (hash value) they generate.

The following table shows an example of the generated signatures from the checksum algorithms mentioned:

Checksum algorithm	Checksum
MD5	E4D909C290D0FB1CA068FFADDF22CBD0
SHA-1	22B759D30862CC7C7EB3CE9616A9D4E853B1E14D
SHA-256	EF537F25C895BFA782526529A9B63D97AA631564D5D789C2B765448C8635FB6C
SHA-512	91EA1245F20D46AE9A037A989F54F1F790F0A47607EEB8A14D12890CEA77A1BBC6C7ED9CF205E67B7F2B8FD4C7DFD3A7A8617E45F3C463D481C7E586C39AC1ED

The checksum results for "The quick brown fox jumps over the lazy dog"

- **Telnet:** Short for teletype network, Telnet has been around since the late 1960s and has been used to establish a command-line interface on a remote device, which is essentially how it's still used today. A Telnet client is available on virtually all operating systems, including Windows, Linux, and even those on access points, routers, firewalls, and other networking devices. Because Telnet preceded most internet protocol developments, it doesn't provide any form of encryption, which means its communications are all *in the clear*. Other tools that perform the same functions as Telnet are PuTTY, which is a Telnet work-alike that adds **Secure Shell (SSH)** encryption. Other alternatives are Microsoft PowerShell for Windows and `netcat` for macOS. The following screenshot shows the command set and options for the `telnet` command:



```

Administrator: Command Prompt - telnet
Welcome to Microsoft Telnet Client

Escape Character is 'CTRL+]'

Microsoft Telnet> ?/help

Commands may be abbreviated. Supported commands are:

c      - close           c\close current connection
d      - display         d\display operating parameters
o      - open hostname [port]  o\connect to hostname (default port 23).
q      - quit            q\exit telnet
set    - set             set\set options (type 'set ?' for a list)
sen    - send            sen\send strings to server
st     - status          st\print status information
u      - unset           u\unset options (type 'unset ?' for a list)
?/h   - help            ?/h\print help information
Microsoft Telnet>

```

The command set of the Windows Telnet

Summary

The server and network security problems you should understand are file integrity, privilege escalation, applications not loading, network file/share access, the inability to open files, excessive access, and excessive memory use. The causes of common security problems include active services, inactive accounts, anti-malware configuration, misconfigured permissions, open ports, and rogue processes.

A system administrator has a toolbox of tools that should include port scanners, sniffers, a Telnet client, and the application of cipher tools and checksums.

Questions

1. What assurance or security condition indicates that files have not changed through the actions of an unauthorized user or program?
 1. Confidentiality
 2. Availability
 3. Dependability
 4. Integrity
2. An attacker is able to modify the permissions of the user account he or she is using to provide access to resources with a higher level of permissions and rights. Of what is this an example?
 1. Horizontal privilege escalation
 2. Vertical privilege escalation
 3. Principle of least privilege
 4. Account spoofing
3. You attempt to open a Windows application, but a message displays saying that the application will not load. Which of the following could possibly be the issue?
 1. Windows Update service isn't running
 2. The ownership of the application is incorrect
 3. Malware may have deleted or renamed the application
 4. The hard disk location of the application is corrupted
 5. All of the above
 6. None of the above

-
4. Which of the following is not likely to be the problem when you are unable to open files from secondary storage?
 1. A problem with the file itself
 2. UAC is set incorrectly
 3. The file may be in conflict with the anti-malware software
 4. The file is in use by a remote connection
 5. A remote user seems to be accessing a particular data resource excessively. What system utility should you use to determine if your suspicions may be correct?
 1. Disk Management
 2. Task Manager
 3. Event Viewer
 4. Control Panel
 6. What Linux command displays a view of the amount of available main memory?
 1. `avlmain`
 2. `mem`
 3. `free`
 4. `dfsk`
 7. Which of the following resource vulnerabilities could an external hacker exploit to gain access to a server?
 1. Web browser
 2. TCP/UDP port
 3. Firewall
 4. Switch
 8. A user login account that has been idle for an extended period of time should be considered to be in what status?
 1. Disabled
 2. Blocked
 3. Idled
 4. Inactive

9. Which of the following is not an action that should be a part of the administration of an anti-malware system?
 1. Update signature database
 2. Establish scan schedule
 3. Review false positives
 4. Idle anti-malware software during peak hours
10. A device or a software utility that scans a system to identify any open ports that may create a vulnerability is what:
 1. Packet sniffer
 2. System monitor
 3. Port scanner
 4. Socket sniffer

CompTIA Server+ Examination

Taking the CompTIA Server+ certification examination and achieving a score of 750 points or more to pass validates you as an information technology specialist with the necessary knowledge and skills to perform as a network server administrator. This also validates your knowledge of server fundamentals, virtualization, data storage technologies, security, troubleshooting, and disaster recovery.

The Server+ examination is vendor-neutral and includes questions and challenges that cover the major operating systems, network services, internetworking devices, and intra-network communications. As a certified Server+ professional, you aren't limited to working with only a single manufacturer or provider.

The exam

The current version of the Server+ certification exam is SK0-004. It was released in July, 2015, and is scheduled to be updated in 2020. The exam has 100 questions that are prorated per the coverage percentage of each of the major topical areas. All of the questions are multiple choice. Once you begin the test, you have 90 minutes to complete it and at the end of this time period, the exam closes.

The examination is given online and one question at a time. You are able to mark questions you'd like to go back to for review. Before the time runs out, you can complete the exam by submitting it as completed. You are appraised of your score almost immediately after completing the exam.

The exam's 100 questions are distributed according to the weighting given to each of the exam domains. For the current exam version, the number of questions you should expect in each domain are as follows:

Domain	Percentage of exam	Number of questions
1.0 Server Architecture	12%	12
2.0 Server Administration	24%	24
3.0 Storage	12%	12
4.0 Security	13%	13
5.0 Networking	10%	10
6.0 Disaster Recovery	9%	9
7.0 Troubleshooting	20%	20

Before the actual exam starts, you must review the CompTIA Candidate Agreement and indicate that you understand it and agree to abide by it. You are given 28 minutes to review the agreement.

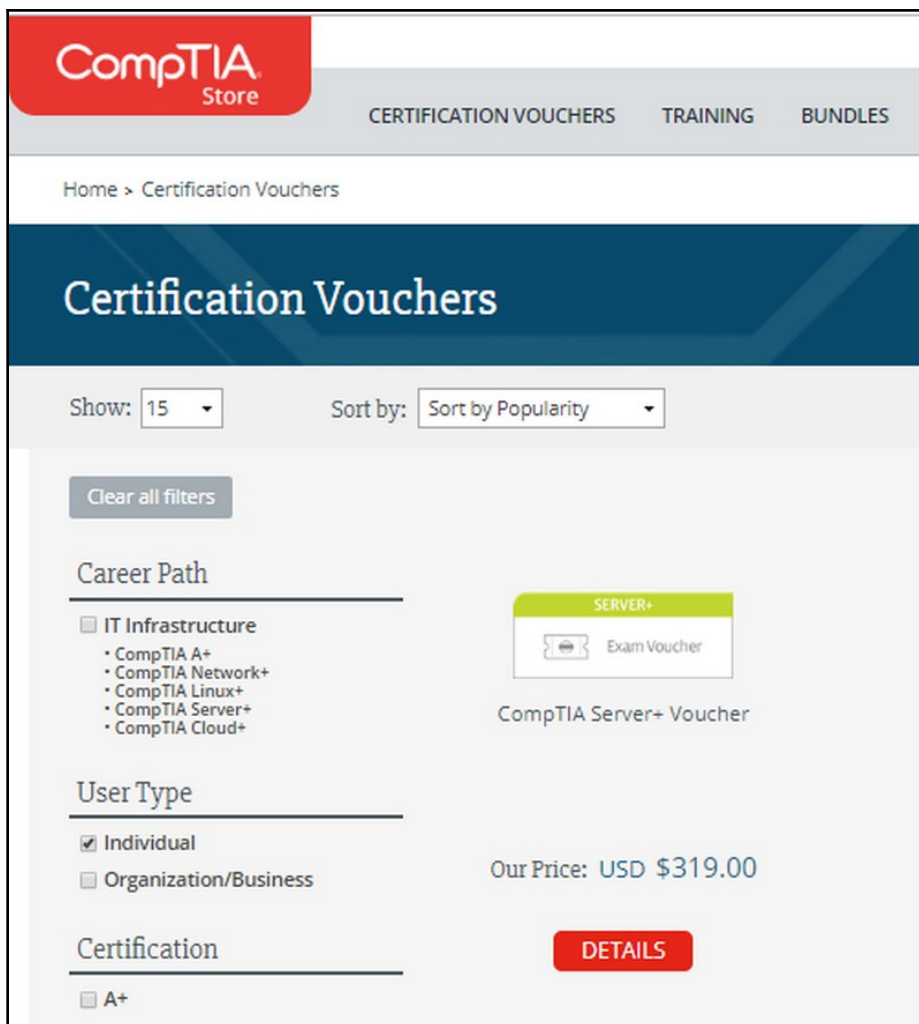
During the exam, you aren't allowed to use any notes or references. In fact, even though the exam is online, you don't have access to the Web to look something up. You are provided with some sheets of note paper to use during the exam, but you must turn it in when you complete the exam.

The Server+ exam is available in four languages—English, Japanese, and Simplified Chinese. When you register for the exam, you can indicate which language you prefer.

So, it boils down to just you and the exam, one-on-one.

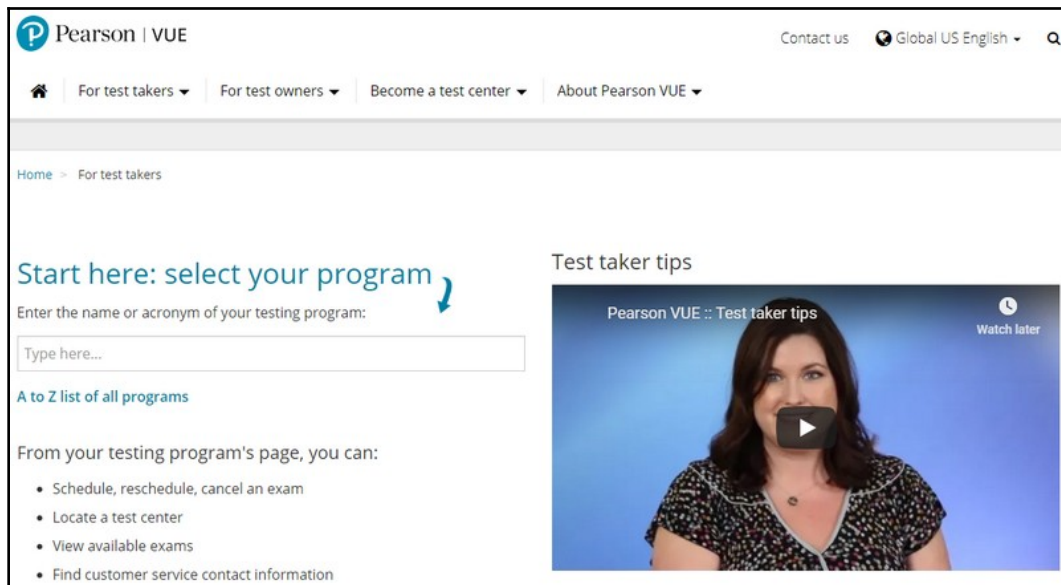
Registering for the Exam

You can register for the CompTIA Server+ exam, which means you can buy a test voucher for the exam on either the CompTIA (www.comptia.org), shown in the following screenshot:



The CompTIA Store web page

or the Pearson VUE (www.pearsonvue.com), shown in the following screenshot:



The Pearson VUE Schedule an Exam web page

On either site, you choose the in-person testing center you wish to use for the exam, schedule a date and time, from those available, and pay for the test voucher. However, you can purchase the test voucher at many of the testing centers as well. For the Server+ exam, the exam fee is \$319.00 USD.

Preparing for the exam

CompTIA recommends that to be prepared for taking the Server+ certification exam, you should have 18 to 24 months of information technology experience, preferably working with servers and networks. CompTIA also recommends that you hold the A+ certification prior to taking the Server+ exam. It's not a bad idea to also have the Network+ certification, as well. However, none of these recommendations are requirements. If you believe you have the knowledge to pass the exam, give it a go. But, remember that it's \$319.00 a pop.

In the CompTIA Candidate Agreement you agree to at the beginning of the testing period, it stipulates that you cannot use *brain dumps* and other unauthorized, and so-called **actual exam content** sources. You should read the Candidate Agreement prior to beginning your serious preparations and studies for the exam. It could save your certification.

The certification

Passing the Server+ exam certifies that you have the knowledge and skills commonly required of a server administrator, including:

- Perform all aspect of server administration: the installation, maintenance, troubleshooting, and security of server hardware, software, including virtualization
- Identify server types and their roles and interactions in a computing environment
- Specify and manage server-related environmental issues
- Execute and comply with business continuity, disaster recovery, and device failures

Holders of the Server+ certification, which I'm sure will include you, have more opportunities for employment. The jobs for which the Server+ exam qualifies you includes Server Support Technician or Administrator, Server Administrator, Data Storage Systems Administrator, and numerous others.



Test your knowledge of concepts required for CompTIA's Server+ exam by visiting the following link:

Glossary

0-9

- **1000BaseT**: Gigabit Ethernet standard with a range reduced to 75 meters.
- **100BaseFX**: A fiber optic cable specification implemented on either a single-mode or multi-mode cable. It has a range of 10,000 meters in single-mode and a range of 412 meters in multi-mode.
- **100BaseTX**: A Fast Ethernet specification with 100 Mbps baseband signaling on a UTP cable.
- **10Base2**: A 10 Mbps baseband coaxial cable with an attenuation distance of 185 meters.
- **10Base5**: Another coaxial baseband cable with a speed of 10 Mbps over 500 meters.
- **10Base-FL**: A 10 Mbps baseband standard that runs on a **fiber optic link (FL)**.
- **10BaseT**: Represents a 10 Mbps baseband UTP or STP cable.
- **10GbE**: This standard transmits at a speed of 10 Gbps at a range of up to 40 km.
- **80-plus**: This is a voluntary certification for a computer's PSU electrical efficiency.

A

- **AAA procedure**: Authentication, authorization, and accounting.
- **Access Control Entry (ACE)**: An entry in an ACL.
- **Access control list (ACL)**: Permits or denies access to incoming or outgoing network message traffic.
- **Access control policies**: Defines who or what has permission to gain access to the resources behind a firewall.
- **Access time**: The aggregate of the time after a disk controller initiates an I/O action and immediately before the data can be read from or written.
- **Active/active server cluster**: A load balancing solution interconnecting two or more computers performing the same processing steps.

- **Active/passive server cluster:** Two or more interconnected servers: one or more active servers and one or more failover servers.
- **Address bus:** The address of data requiring service.
- **Address Resolution Protocol (ARP):** A protocol used to translate IP addresses into MAC addresses.
- **Advanced RISC machine (ARM):** A specialized technology using an advanced instruction set.
- **Ampères/Amps:** The rate of flow of an electrical current.
- **Anycast:** An IPv6 packet with an anycast address goes to only one of an identified set of nodes. The receiving node is typically the "closest" in terms of distance and availability.
- **Application server:** Provides services for one or more applications and serves as a mid-level service between user requests and other server- or network-based functions.
- **Application/web servers:** A server that includes the capability of delivering web content to a client's browser as well as other applications.
- **Architecture diagram:** Depicts the major components of an infrastructure, system, or even applications.
- **Asymmetrical multiprocessing (ASMP):** One CPU handles the tasks of the OS and assigns process requests to the slaves and all other CPUs.
- **Asynchronous replication:** An approach for replication over a distance as a part of a disaster recovery strategy.
- **Authentication:** The verification of a user from the data or images provided in a login request.
- **Authentication Header (AH) transport mode:** IPSec AH mode inserts a header in each packet that contains a keyed hash total to ensure the packet's integrity.
- **Automatic Private IP Addressing (APIPA):** A default address from a reserved Class B address group of 169.254.0.0/16.

B

- **Backbone cable:** A primary communication cable that interconnects the primary, main, and intermediate distribution facilities.
- **Backend server:** A server that performs I/O operations on a database.
- **Baffles:** Air flow defectors that direct air flow to enhance the air-cooling system.
- **Bare metal backup:** A capture of everything on a system, including the OS, stored application, and system software and data.

- **Battery-backed cache memory:** Enables a RAID controller to process data in either direction of I/O operations faster than it's able to write it to a disk.
- **Bayonet Neill Concelman (BNC) connector:** Common connector for the coaxial cable.
- **Beep codes:** Tones sounded to indicate a failure during the bootup process.
- **Binary Input/Output System (BIOS):** Holds the first instruction, the hardware configuration, and support for I/O operations for startup.
- **Blade server:** A slim computer-on-a-card installed in a slot of a blade enclosure chassis.
- **Blue Screen of Death (BSoD):** Indicates an error between the operating system and the hardware configuration.
- **Broadcast:** A message type sent to all nodes, typically requesting information.
- **Broadcast address:** The highest (last) assignable address in a network range.
- **Broadcast domain:** The portion of a local network that *sees* a broadcast message on the medium.
- **Browser cache:** Stores a downloaded web page's content.
- **Bus channel:** Provides a pathway on which data, addresses, and commands travel.
- **Bus width:** The number of traces in the bus channel.
- **Business continuity plan (BCP):** Outlines the objectives, procedures, and step-by-step actions required to restart or continue an organizations' operations after a disruptive event.
- **Business impact analysis (BIA):** A study that projects the potential financial impact of any interruption in its operations from an extreme event.

C

- **Cable category (cat):** Defines network cable capabilities, based on its frequency range, bandwidth, and **data transfer rates (DTRs)**.
- **Cache memory:** A special purpose storage for a variety of uses, including staging instructions, preloading data, or buffering data or instructions before use.
- **Category 3 cable (CAT 3):** A an eight-wire (4 wire pairs) UTP cable capable of 10 Mbps DTR and 16Mhz bandwidth.
- **Category 5 cable (CAT 5):** A four-pair (8 wire) UTP cable that supports both 10 Mbps and 100 Mbps data speeds and a bandwidth of 100 MHz on an Ethernet network.

- **Category 5 enhanced cable (CAT 5e):** This enhancement of the CAT 5 standard reduced channel crosstalk and extended its data speed to 1 Gbps and with 100 MHz bandwidth on an Ethernet network.
- **Category 6 augmented cable (CAT 6a):** An enhancement of CAT 6 raises the DTR to 10 Gbps and the bandwidth to 500 MHz.
- **Category 6 cable (CAT 6):** Raises bandwidth to 250 MHz with Gigabit Ethernet speeds.
- **Central processing unit (CPU):** The electronic component that runs program instructions, performs arithmetic functions, and controls the movement of data and the input and output functions of peripheral devices attached to or installed in the computer.
- **Certificate authority (CA):** A trusted organization that provides unique digital certificates to subscribers and manages public keys and identity credentials for data encryption of stored data, websites, and email.
- **Challenge Handshake Authentication Protocol (CHAP):** After receiving a connection signal, the CHAP server transmits its hostname and a randomly generated challenge to the requesting client.
- **Checksums:** A value that's generated by an algorithm that consists of a sequence of characters that represents the contents of the original string.
- **Confidentiality, Integrity, and Availability (CIA) model:** The core model for network security.
- **Cipher:** An algorithm that's used to convert plaintext into ciphertext.
- **Cladding:** A reflective material coating the core filament of a fiber optic cable.
- **Classful IP addressing:** An IPv4 addressing scheme that defines five address classes: A, B, C, D and E.
- **Classless Interdomain Routing (CIDR):** An appendage to an IP address that indicates the number.
- **Clustering:** Consists of servers arranged in an interactive group.
- **Coaxial cable:** A copper core cable with two channels: solid core wire carries the transmitted signal and metal mesh layer also carries that same signal as a shield against EMI.
- **Cold site:** This type of recovery site has only the necessary environmental and power systems to support the restoration of computing services.
- **Cold swap:** Completely shutting down a system to affect upgrades, replacements, or repairs.
- **Collision domain:** Collisions occur when two nodes on the same network segment attempt to transmit on the network simultaneously.

- **Collision Sense Multiple Access/Collision Avoidance (CSMA/CA):** Used on a wireless network to avoid transmission collisions.
- **Collision Sense Multiple Access/Collision Detection (CSMA/CD):** A method used on Ethernet networks to clear the medium after a collision.
- **Column address strobe (CAS) Latency (CL):** The time in nanoseconds required to receive and fulfill a request for data.
- **COM port:** A serial interface, typically with a D-subminiature - 9 (DB-9) connector.
- **Configuration specifications:** Reflects any changes to the device specifications after formatting or partitioning the disk media.
- **Control bus:** A dedicated one-way bus that carries a command to use on the data at the address on the address bus.
- **Copy backup:** Creates an archival copy of the data on secondary storage.
- **Core:** A microprocessor that's able to process a separate and unique stream of instructions than other cores on the CPU.
- **CPU multiplier:** A factor that's applied to the frequency of the FSB to determine and set the internal frequency of the CPU.
- **CPU stepping:** Revisions applied to a CPU.
- **Cross-connect:** The point where the network backbone terminates and is a connection point for a facility's network.
- **Crossover cable:** This cable combines the two 568 standards with a connection of each standard on either end of a cable.
- **Current:** The flow or movement of an electrical charge.
- **Customer-replaceable unit (CRU):** A component that a user/customer can remove and replace.

D

- **Data bus:** The CPU sends or receives data from memory or a device controller. Data centers and server rooms must be secure and safe environments and workspaces.
- **Data encryption:** The application of a cryptographic algorithm to data so that when stored or transmitted only those with the appropriate access can open, execute, or apply its contents.
- **Data flow diagram (DFD):** A graphical depiction of how data flows through a system, network, or database.

- **Data transfer rate (DTR):** The time required to move data from one location to another in a specific operation.
- **Database server:** A server hosting a database management system and its database.
- **Dedicated file server:** Only provides file or database content to clients.
- **Delta:** Connects two current-bearing lines to create a triangular shape.
- **Deployment Image & Servicing Management (DISM):** A utility that can scan, check, restore, and repair corrupted files on a hard disk or stored as a part of an image file.
- **Device specifications:** Indicates the *raw* or unformatted measurements of a disk drive's components and operations.
- **Differential backup:** Copies the files that have been modified or created since the last full backup.
- **Direct memory access (DMA) addresses:** An I/O device that's able to read and write directly from or to main memory without assistance from the CPU.
- **Direct-Attached Storage (DAS):** A storage device connected directly to a computer.
- **Directory server:** Supports directory services that cross-reference or map the names, designations, or locations of computer or network resources to their respective local or network address.
- **Disaster recovery plan (DRP):** A plan for the restoration of the computing infrastructure and its associated services.
- **Discretionary ACL (DACL):** Identifies the IP addresses, user or group accounts, port numbers, and protocols with permission to access a resource.
- **Disk cache:** Hard disk drives have a small amount of RAM that's used as a buffer for I/O operations.
- **Disk compression:** Reduces the space that stored data uses on a disk medium.
- **Disk mirroring:** Duplicating data onto two or more drives.
- **Disk striping:** Separates data into chunks that are stored on two or more disk drives or volumes.
- **Disk usage/Disk free:** These two commands (`du` and `df`) list the amount of disk space that is in use and is free, respectively, on the current mounted filesystem.
- **Disk-to-disk replication:** Copies data from one data storage device to another.
- **DMZ:** A physical or logical network segment or subnet that's used as a default landing space for external traffic.
- **Domain Name Service (DNS), Domain Naming System (DNS):** A protocol that's used to convert FQDN into an IP address.

- **Drive Interface:** The communication interface a disk drive supports.
- **Dynamic Host Configuration Protocol (DHCP):** A protocol that provides a network configuration to hosts from a pool of available IP addresses.
- **Dynamic loading:** Loads the first module of a program and loads any other modules when needed.
- **Dynamic RAM (DRAM):** Electrically volatile memory.

E

- **EIA/TIA 568A/568B:** The standards for the pinouts of the RJ-45 connectors and twisted-pair cable for commercial and residential installations.
- **Encapsulating Security Payload (ESP) transport mode:** IPSec ESP mode includes the message actions of the AH mode and the encryption of the payload only.
- **Entrance facility:** The location where a service provider's network and the subscriber's backbone interconnect.
- **Environmental threats:** Weather, natural disaster, catastrophic event, and other events from the natural world.
- **Error-correction code (ECC) memory:** Includes a dedicated memory unit that provides parity and error-correction to the other memory units.
- **Extensible Authentication Protocol (EAP):** Combines with authentication methods and acts as middleware between the client and an authentication server.
- **External bus:** Provides a means for peripherals and expansion components to communicate with the components on a motherboard.

F

- **Fault tolerance:** Strategy to retain all in-process data and operations in the event of a component or system failure.
- **Fibre Channel (FC):** A communication protocol that's used in data centers and server farms.
- **Field-replaceable unit (FRU):** The components that only a qualified field service representative should remove and replace.
- **File integrity:** This means that data and program files have not been the target of an attack and modified in any way by an unauthorized person or function.
- **File server:** Provides data resources to other nodes on a network.

- **File system:** An organizing system for structuring data on secondary storage.
- **File Transfer Protocol (FTP) server**
- **Form factor:** Establishes the dimensions, shape, and other physical characteristics of a computer's hardware.
- **Frontend server:** Client host or application server requesting data from a database.
- **Front-side bus (FSB):** The bus that connects the CPU to the Northbridge of the chipset.
- **F-type connector:** Common connector for coaxial cable.
- **Full backup:** Copies all content on secondary storage device and writes it, in compressed form typically, to a removable medium.
- **Fully Qualified Domain Name (FQDN):** The full text name for a website, such as `www.packt.com`

G

- **Gateway proxy server:** Also known as the application-level gateway or a tunneling proxy server, it serves as a portal between a local network and the internet, sending and receiving client requests and the responses.
- **Gilster's Law (of everything computing):** *You never can tell, and it all depends.*
- **Ground:** The protective measure with a conductive connection to the earth.

H

- **Hardening:** Actions to increase security by reducing the vulnerability of a workstation, server, and network from exploitation.
- **Heat sink:** A device that attaches directly to a CPU to dissipate excess heat.
- **Hextet:** A grouping of 16-bits representing 4 hexadecimal value.
- **High availability (HA):** A system operating and available a predominant percentage of the time.
- **High-level formatting (HLF):** Adds the structures the OS uses for partitions or logical volumes.
- **Horizontal cable:** This is the cabling that connects the networking devices on a single level of a facility, such as a floor or story, to the facility's backbone.

- **Host bus adapter (HBA):** Provides a connecting point for peripheral devices to a computer.
- **Host-based firewall:** A firewall or security device on a network-connected host
- **HOSTS file:** On a Windows system, this provides a localhost DNS-type of lookup to provide the IP address of a FQDN.
- **Hot site:** This type of recovery site is essentially a copy of a production system and its environment.
- **Hot swap:** The immediate switchover or physical replacement of a failed component while a system is running and fully operational.
- **Hypervisor:** Provides direct support to virtual servers, each of which occupies a shell in memory.

I

- **I/O addresses:** Installed I/O devices have one or more addresses assigned for reference and addressing.
- **I/O operations per second (IOPS):** Measures the number of read and write operations from and to random, non-contiguous addresses an HDD can perform in one second.
- **IEEE 802.1x:** Defines **port-based network access control (PNAC)** that authenticates a process before allowing access to a device interface port.
- **Incremental backup:** Copies only those files that have been modified or created since the last full or incremental backup.
- **Intermediate distribution facility (IDF):** Provides an unlimited number of interconnections between horizontal cabling segments on a single level of a facility.
- **Internal bus:** Used by a motherboard's components to pass data and instructions.
- **Internal-facing proxy server:** Provides protection and services to its internal network.
- **Internet Group Management Protocol (IGMP) messages**
- **Internet Key Exchange (IKE):** Supervises the authentication, application of security policies and rules, and key exchange activities of each side of an IPsec interaction.
- **Internet Protocol version 6 (IPv6):** The replacement for IPv4 that uses 128-bits.
- **Internet security zones:** Windows defines four security zones: internet, local intranet, trusted sites, and restricted sites.

- **Internet-facing (forward) proxy server:** Facilitates requests from its internal network for resources from the internet.
- **Interrupt requests (IRQs):** A signaling device between a program and the OS asking for a service.
- **Intrusion detection system (IDS):** Scans network traffic for malware, tracks patterns of suspicious behavior, and monitors configuration settings for inadvertent changes.
- **ipconfig/ifconfig:** Displays the IP configuration of a host and its network interface(s).

J

- **Journaling:** A filesystem that records changes before applying them to the medium.

K

- **Kerberos:** A secure authentication protocol that uses an encrypted proof of identity to identify a user or a local network node.
- **Kernel:** The module of the OS that loads at startup and remains in memory.
- **Keyboard-Video-Mouse (KVM) switch:** Allows a centrally located administrator to control multiple computers individually through a single keyboard, video display, and mouse.

L

- **LAN application server:** Provides support to network nodes for one or more applications.
- **Land Grid Array (LGA) packaging:** The mounting pins are on the socket and the CPU has receiving ports for each pin.
- **Large form factor (LFF):** A hard disk drive sizing form for drives that are able to store as much as 100 TB.
- **Layer 2 Tunneling Protocol (L2TP):** A protocol that allows Internet Service Providers (ISPs) to provide VPNs to subscribers.

- **Level 1 (L1):** This is the fastest of all three levels and the smallest in size.
- **Level 2 (L2):** This is faster than L3, but smaller in size.
- **Level 3 (L3) cache:** This is slow, but faster than main memory, and is the largest in size of the three levels.
- **Lightweight Directory Access Protocol (LDAP):** Common in authentication processes for storing and verifying user accounts.
- **Linear access media:** Refers to magnetic tape and serial access.
- **Linear Tape – Open (LTO):** Stores data objects separately from their metadata so that data can be accessed randomly.
- **Liquid cooling:** Uses a coolant to pull the heat away from a CPU.
- **Load balancing:** A function that distributes incoming network traffic to two or more servers arranged in a pool, farm, or cluster.
- **logical unit number (LUN):** An identity assigned to a SAN unit.
- **Low-level formatting (LLF):** Places digital sector markers on the disk to map the storage medium into cylinders, tracks, and sectors.

M

- **Mail transport agents (MTAs):** Process and transport electronic mail messages for a network, up to and including the internet.
- **Memory addresses:** A block of memory for use as a data buffer that's assigned to an I/O device.
- **Memory allocation:** The process that's used to allocate memory space to a program.
- **Memory cache:** Application software requiring a large amount of data, such as a graphics editor, will create a cache in RAM to reduce I/O operations and speed up processing.
- **Memory leak:** This happens when a program fails to release some or all of its allocated memory when it's no longer needed.
- **Memory timings:** Given as four numerical values that represent the four memory timing measurements: CL, tRCD, tRP, and tRAS.
- **Messaging server:** Receives, forwards, or holds messages between client applications and services.
- **mount:** Attaches a filesystem and adds it to the active directory structure. This makes its contents available for access.

- **Multicast:** An IPv6 multicast address identifies a group of nodes perhaps scattered across several networks. Each of the nodes included in the multicast address receives the transmission.
- **Multi-core processing:** One microprocessor contains multiple cores.
- **Multifactor authentication:** A combination of two or more identification factors to authenticate a user requesting to log onto a system or network.
- **Multi-mode fiber optic cable:** A cable that's capable of transmitting several light streams at once.
- **Multiple-instruction, multiple-data (MIMD):** Multiple processors execute different instructions on different blocks of a data source. MIMD is what most people think of as parallel computing.
- **Multiple-instruction, single-data (MISD):** Multiple processors execute different instructions on a single data source. MISD computing is not common because this mode of parallel processing is usually very specific to a problem.
- **Multiprocessing:** A single computer system with two or more integrated CPUs.

N

- **Net use:** A family of commands used to create a link to or to disconnect from a network shared resource, display all current connections on a host, share a resource with other hosts, manage passwords, control the print spooler, and more.
- **Network adapter:** See **network interface controller (NIC)**
- **Network address:** The lowest assignable address in a network range.
- **Network Address Translation (NAT):** A protocol that applies a public IP address as an alias for a private address.
- **Network diagram:** A graphical representation of the devices and services in a local, wide, or other network.
- **Network File System (NFS) server**
- **Network interface controller (NIC):** Provides a connection and interface between a host computer and a network.
- **Network operating system (NOS):** Provides network services and protocol support from a centralized server.
- **Network services:** Core services provided to network clients on the OSI application layer.
- **Network-attached storage (NAS):** Data storage devices attached to a network that network clients can share.

- **Network-based firewall:** A security filtering device that permits or denies inbound traffic.
- **Non-dedicated file server:** Supports two or more applications and file or database services.
- **Northbridge:** (memory controller hub) of the chipset.
- **nslookup:** A command-line utility that's used to look up names and IP addresses in a name server.

O

- **Octet:** A grouping of 8 bits representing values from 0 to 255.
- **One-time password:** A challenge-response password or a password from a predefined password list.
- **Open proxy server:** Sends request and response messages to or from anywhere on an internetwork.
- **Organizationally Unique Identifier (OUI):** A 24-bit code assigned to a network device producer by the IEEE.

P

- **Packaging:** The shape and construction of a CPU. Virtually all server CPUs are in packaging.
- **Parallel Advanced Technology Attachment (PATA):** Uses parallel bit signaling to transmit word-length data between an HDD and controllers and drivers.
- **Parallel processing:** Multiple processors execute the same instruction or a unique set of instructions.
- **Passive cooling:** Cooling the CPU and computer through convection.
- **Password Authentication Protocol (PAP):** A legacy protocol that performs basic authentication steps.
- **Patch cables:** A generic cable type used to establish a connection between two electronic devices.
- **Patch management:** A part of change management and control that focuses on the security of a system.
- **Peripheral Component Interconnect (PCI):** An expansion card standard.

- **Physical security:** A program of the events, causes, actors, prevention, recovery, mitigation, and other relevant procedures regarding the security, safety, and operations of achieving an organization's mission.
- **Ping:** A command prompt utility that's used to verify a connection between a source host and a destination host.
- **Point-to-point messaging servers:** Communicates between a messaging server and a single addressee client.
- **Port Address Translation (PAT):** A protocol that applies a unique port number to the private address of the requesting LAN node.
- **Port scanners:** A software application that scans a server to identify any open TCP/UDP ports.
- **Post Office Protocol version 3 (POP3):** A client-based protocol that interacts with a mail server to send and receive messages that are addressed to a particular user.
- **Power distribution unit (PDU):** An appliance with multiple electric outlets and a mains AC connection that converts raw electrical power sources into multiple lower voltage electrical outlets.
- **Power-On Self-Test (POST):** As a part of the startup process, it checks the internal hardware components included in the BIOS or UEFI configuration.
- **Preventive maintenance:** A periodic schedule program of cleaning and testing to avoid device or component failure.
- **Print server:** Accepts print requests from clients and provides sequencing and management of a network-attached printer, plotter, or other imaging device.
- **Private IP addressing:** Three blocks of addresses from classes A, B, and C are set aside for use on any local network.
- **Privilege escalation:** An attacker is able to expand or elevate his or her rights and permissions to access resources with higher restrictions.
- **Processor cache:** Buffers data and instructions to the CPU, eliminating the need to access the main memory.
- **Protocol data unit (PDU):** The generic name for structured data transmitted on a network.
- **Proxy server:** An intermediate network services that accepts client requests for resources from remote servers.
- **Public key infrastructure (PKI):** A group of rules, policies, and procedures that *create, manage, distribute, use, store, and revoke digital certificates and manage public-key encryption.*

- **Public-key:** Two pass keys, one private and one public, encrypt, decrypt, and verify the data provided for authentication.
- **Publish-subscribe messaging servers:** Communicates a client message to multiple subscribed clients via a messaging server.

Q

- **Query-based application server:** Hosts one or more scripting or programming language services that are used to request data from a database.

R

- **Rack unit (U):** A rackmount measurement of 1.75 inches (44.45 millimeters) of height. The height of rackmount devices is in Us, such as 1U, 2U, 3U, or a 4U half-rack mount.
- **Rackmount:** A two- or four-rail vertical structure for servers and other rack-mount devices.
- **RAID 0:** Disk striping to multiple disk drives.
- **RAID 1:** Mirrors data to another disk volume.
- **RAID 10:** Combines the mirroring of RAID 1 with the striping of RAID 0, but without parity.
- **RAID 5 – RAID 0 with parity added.**
- **RAID 6 – RAID 5 with the parity doubled.**
- **Random access memory (RAM):** The volatile storage area that serves as a conduit for program instructions, data, and addressing going to or coming from the CPU.
- **RAS Precharge (tRP):** The time required to release the active row in memory.
- **Recovery point objective (RPO):** The target point-in-time to be reestablished through data recovery for a failed system.
- **Recovery time objective (RTO):** The desired length of time it takes to recover a failed system.
- **Registered jack 45 (RJ-45) connector:** EIA/TIA 568 standards specify an eight-position/eight-contact (8P8C) connector for twisted-pair cabling.
- **Registered port:** A TCP/UDP reference number to a servicing application or protocol in the range of 1024 to 49151.

- **Registration authority (RA):** A network service that approves and forwards requests for identity verification (digital certificate) and the certificate authority that issues it.
- **Remote Authentication Dial-In User Service (RADIUS):** An authentication protocol for centralized network access control.
- **Remote Desktop Protocol (RDP):** A Microsoft service that provides remote access to network-connected systems, providing a GUI interface.
- **Remote Server Administration Tools (RSAT):** A Microsoft package that's used to manage the configuration and features of Windows Server.
- **Resistance:** The properties of a wire that oppose the current flow.
- **Reverse Address Resolution Protocol (RARP):** A protocol that's used to translate MAC addresses into an IP addresses.
- **Reverse proxy server:** Performs authentication, authorization, caching, decryption, or load balancing.
- **RISC:** Reduced instruction set computer.
- **Risk assessment:** A study that projects which assets are at risk of loss or damage in an extreme event.
- **Role-based access control (RBAC):** Access control based on a user's duties.
- **Rollover cables:** This cable reverses some of the pinouts to create an interface between two devices.
- **Rotational latency:** Also known as rotational delay or latency, this is the time that's required for the rotating disk platter to move the targeted data sector under the read/write head.
- **Routing and Remote Access Service:** A suite of protocols that provides firewall, router, and remote access connectivity.
- **Row Active Time (tRAS):** The time required to close an active row and to open a new row.
- **Row address strobe (RAS) to CAS Delay (tRCD):** The time required to move to the row in which the requested data is located.

S

- **Secure Shell (SSH):** A software utility that facilitates OS administration and file transfers over a secure remote connection.
- **Secure Sockets Layer (SSL)/Transport Layer Security (TLS):** Protocols that secure TCP communications, especially **Hypertext Transfer Protocol (HTTP)** messages.

- **Security policies:** The rules that control the permit or deny of inbound and outbound traffic based on source and destination addresses, protocols, port numbers, and content.
- **Security zone:** A logical structure that's created from one or more device ports/interfaces that apply the same security policies.
- **Seek time:** The time it takes for the actuator arm of a hard disk drive to position the read/write head over the appropriate disk platter and then over the disk track on that platter that contains the targeted data sector.
- **Selective backup:** Copies only pre-selected or marked files and folders.
- **Serial ATA (SATA):** Uses serial bit signaling to transmit data between an HDD and the motherboard, controllers, and drivers.
- **Serial Attached SCSI (SAS):** Improves on SCSI by connecting directly to SAS HDD.
- **Server:** A centralized computer running server software that provides services to network clients.
- **Server Message Block (SMB):** A file-sharing protocol that give applications the capability to access files and other network resources on a network.
- **Server monitor:** Software that provides for automated reporting, scheduled device checking, and device failure warnings.
- **Server-based replication:** Part of a high availability or disaster recovery strategy. Server-based replication systems can be server-to-self, cluster-to-cluster, and server-to-server.
- **service level agreement (SLA):** A common instrument provided by a service provider to a service subscriber that details the scope of services and their level of performance.
- **Service Message Block/Common Internet File System (SMB/CIFS):** Protocol server.
- **Shielded twisted-pair (STP):** In addition to the insulation on each wire, a metal wrapping encases each or all of the internal wire pairs.
- **Simple Message Transport Protocol (SMTP):** Transports messages between mail servers.
- **Simple Network Management Protocol (SNMP):** A protocol that allows network devices to share information about status and measurements against preset conditions.
- **Single-instruction, multiple-data (SIMD):** Multiple processors execute the same instruction on different blocks of a data source. SIMD speeds up multimedia processing.

- **Single-mode fiber optic cable:** A cable with one fiber filament that carries a single transmission mode or light stream.
- **Single-phase power:** A two-wire distribution system for AC, in which one wire carries the electrical current and the other wire is neutral.
- **Small Computer Serial Interface (SCSI):** Provides a multiple-connection interface for as many as 15 devices on a SCSI interface in a single HBA slot.
- **Small form factor (SFF):** The smallest of the common forms, commonly used in portable computers.
- **Snapshot:** A capture of the state of a system at one specific point in time.
- **Sniffers:** Utility software that examines the content of network packet traffic.
- **Solid-state drive (SSD):** Stores data in semiconductor chips.
- **Split-phase power:** A three-wire single-phase distribution system for AC, in which two wires carry electrical current and the third wire is the neutral.
- **Static RAM (SRAM):** Non-volatile memory.
- **Storage area network (SAN):** A self-contained network of storage devices and specialized switches that provides high-speed data access to connected network nodes.
- **Straight-through cables:** The two ends of the cable terminate with the connector pins matched one-to-one.
- **Subnet mask:** A Boolean algebra masking value that's used to extract network addresses from an IPv4 address.
- **Subnetwork (subnet):** A logical segment of a larger network that creates a smaller broadcast domain and collision domain.
- **Symmetrical multiprocessing (SMP):** Two or more CPUs are equals and share the OS and system resources.
- **Synchronous replication:** Simultaneously writes to two storage devices to provide a real-time distributed data source or a hot failover backup.
- **System ACL (SACL):** Controls the router feature that generates log or audit entries detailing attempts to access a resource.
- **System File Checker (SFC):** Performs a scan of the hard disk looking for corrupted system files and folders.

T

- **TACACS+:** Encrypts its packets entirely before forwarding them on to an authentication server.

- **Terminal Access Controller Access Control System (TACACS):** Forwards user credentials to an authentication server for verification and permission to gain access to a system or network.
- **The Internet Protocol Security (IPSec):** A series of standards for both the encryption and the transmission integrity of transmitted packets.
- **Thermal Design Power (TDP):** The amount of heat produced by a computer system.
- **Thermal paste:** Seals a heat sink or fan to the CPU.
- **Three-phase power:** A four wire system in which three overlapping wires carry an AC current.
- **Three-tier client/server:** A network environment in which an application server is the middleware between a network user and a database management system.
- **Tower computer:** A computer in a vertical standing case, commonly used as a network server.
- **Tracert/traceroute:** A command-line utility that's used to test and display the router hops between a source host and a destination host or network.
- **Twisted-pair cable:** A network cable that has multiple pairs of copper wires with each pair of wires twisted around each other to minimize cross-talk.

U

- **Unicast:** An IPv6 unicast address identifies a single destination. Packets with that unicast address go to that address.
- **Unified Extensible Firmware Interface (UEFI):** A configuration technology replacing BIOS.
- **Uninterruptable Power Supply (UPS):** An appliance that levels power surges and slumps and provide electrical power in the event of a loss of electrical service.
- **Unshielded twisted-pair (UTP):** Twisted-pair cable that has no additional shielding in the cable beyond the sheathing on each wire.
- **User Account Control (UAC):** Restricts changes to the system to only those authorized to do so.

V

- **Virtual local area network (VLAN):** A logically created local network segment.
- **Virtual machine:** A software-created processing object operating under control of a virtual server.
- **Virtual Network Computing (VNC):** Provides a GUI desktop through the **Remote Frame Buffer (RFB)** protocol that enables an administrator to control and manage a remote system over a network.
- **Virtual private network (VPN):** An encrypted connection over an insecure network.
- **Virtual server:** A software-enabled logic object that supports virtual machines in the memory of a physical computer. A virtual server can support numerous virtual machines.
- **Voltage switching:** A voltage sensor that automatically detects the electrical current and switches to its voltage and mode.

W

- **Warm site:** This type of recovery site contains the necessary equipment and environment to support the essential components of a production system.
- **Warm swap:** Requires the suspension of a system's or failed component's operations while a replacement or switchover takes place.
- **Watts:** The output rate of energy radiated, absorbed, or dissipated.
- **Well-known port:** A TCP/UDP reference number to a servicing application or protocol in the range of 0 and 1023.
- **Windows Internet Name Service (WINS):** A Windows service that resolve devices and network names.
- **Write-through:** The CPU writes data directly to main memory or a storage device, bypassing cache memory.
- **Wye:** Connects a current-bearing line to a neutral in a Y pattern.

Z

- **Zero-knowledge authentication:** Users receive a question or arithmetic problem to answer or resolve that is unique each time.

Assessment

Chapter 1: Server Hardware

1. DHCP server
2. A file server can be either dedicated or non-dedicated
3. Network services server
4. Internal-facing
5. False
6. ATX
7. 1.75 inches
8. +3.3VDC, +5VDC, +/- 12VDC
9. Passive

Chapter 2: Server Internals

1. CPU
2. CPUs are equal and share system resources
3. One CPU is a master and all others are slaves
4. Multiprocessor
5. MISD
6. Level 0
7. CPU multiplier and the frequency of the FSB
8. PCIC
9. BIOS, UEFI
10. CAS latency (CL)

Chapter 3: Data Storage

1. Rotational delay, Access time
2. SATA, SCSI
3. JBOD
4. A self-contained storage device network and switches that provides high-speed access to data
5. LUN zoning, LUN masking
6. ext3 and NTFS
7. SMB/CIFS
8. Mirroring, Striping
9. RAID 0
10. Highly available
11. The matching is done as follows:
 - (a) Hot swapping-(2) The immediate switchover or replacement of a failed component completed while a system remains fully operational
 - (b) Cold swapping-(1) Requires the powering down of a system to affect replacements or repairs
 - (c) Warm swapping-(3) The suspension of operations, although still powered, to affect the replacement of a failed component

Chapter 4: Server Operating Systems

1. Providing services to network clients
2. Caching
3. Device driver
4. File accessibility
5. Device drivers
6. BIOS, UEFI
7. BTRFS
8. Hostname
9. Local computer resources, Network-attached resources
10. PXE

Chapter 5: Addressing

1. Five classes of 32-bits in four octets
2. Class B
3. NAT
4. Collision domain
5. CSMA/CD
6. 201.255.255.255
7. 201.110.25.16/24
8. It masks one or more sections containing all zeros
9. Broadcast
10. 0 to 1023

Chapter 6: Cabling

1. UTP, STP
2. 100 meters
3. Rollover cable
4. RJ-45
5. F-type, BNC
6. IDF
7. CAT 5e, CAT 6
8. (7)(2)(3)
9. SM
10. A 1-inch bend radius

Chapter 7: Server Administration

1. Configuration, Monitoring, Implementing
2. KVM
3. COM
4. Network-based administration
5. All of the above (RSAT, CLI over SSH, VNC, RDP)

6. ITAM
7. LCAM
8. Typing tutor
9. Completeness
10. All of the above (Purpose, Effectiveness, Intended audience, Completeness)

Chapter 8: Server Maintenance

1. Immediate application
2. Patch management should be a priority
3. Asset management systems
4. It represents the maximum number of reads and writes (input/output operations) to and from non-contiguous storage locations on secondary storage devices
5. Beep codes
6. **Customer replaceable unit (CRU)**
7. All of the above (Extend the service life of a server component, Avoid component failures, Maintain server uptime commitments)
8. Active/active
9. Active/passive
10. Hot swap

Chapter 9: Virtualization

1. Virtual reality
2. Quasi-virtualization
3. Hypervisor
4. Type I
5. Type II
6. Host
7. Assigned
8. Network connectivity, Bandwidth
9. Hardware virtualization
10. Memory

Chapter 10: Disaster Recovery

1. BCP
2. DRP
3. Real-time
4. Risk assessment
5. Data replication
6. Synchronous
7. All of the above (Disk-to-disk, Cluster-to-cluster, Server-to-server)
8. Differential
9. Layered access
10. RTO

Chapter 11: Security Systems and Protocols

1. Security zone
2. Host-based, Network-based
3. Authentication
4. IPSec
5. All of the above (AH, Transport, ESP, Tunnel)
6. EAP
7. Dynamic locking, Static locking
8. Source address
9. Implicit deny
10. PPTP

Chapter 12: Physical Security and Environmental Controls

1. MFA
2. Something about you
3. Physical security
4. Cyber

5. Cold row/hot row
6. Logical intrusion
7. UPS
8. PDU
9. Cold-swap
10. Broad load

Chapter 13: Logical Security

1. Computer name
2. ACE
3. Encryption
4. All the above (Delete, Wipe, Erase, Shred)
5. Recoverable
6. Reducing vulnerability
7. True
8. Hardware hardening
9. Signature matching
10. NAC

Chapter 14: Troubleshooting Methods

1. Troubleshooting
2. Open questioning
3. Interpersonal skills
4. Identify the problem
5. Recreate the problem
6. Back up the entire hard disk
7. Entertainment software
8. Main memory, System case, Cooling system, Graphics card
9. All of the above (Corrupted system files, Improperly installed software, Malware attacks)
10. All of the above (Develop a test plan, Involve the user, Document the results)

Chapter 15: Common Hardware Issues

1. Identify common hardware issues
2. Buffer overflow
3. POST
4. Beep codes
5. All of the above (Blocked airflow vents, Broken CPU fan, Neglected preventive maintenance)
6. 113° F to 122° F
7. Bad PCI slot
8. None of the above
9. Peripheral Clustering Interconnect Extended
10. Sunlight

Chapter 16: Common Software Issues

1. He doesn't have access permissions to the server.
2. Write
3. BSoD
4. Memory leak
5. SFC
6. Windows Disk Management, Disk compression
7. `df`
8. Virtual memory
9. Defragmentation
10. Log

Chapter 17: Common Network Issues

1. Internet gateway
2. False
3. APIPA
4. Port 587
5. Hosts

6. VLAN
7. Trunking port
8. All the above (Antivirus software, Network adapter device driver, Auto-login service)
9. Either ping **or** tracert/traceroute, or both
10. ifconfig

Chapter 18: Common Storage Issues

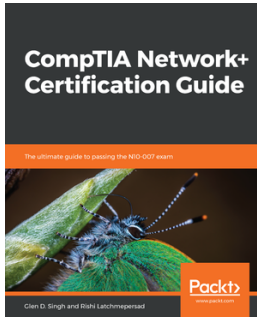
1. Too fast
2. SDD
3. Host Bus Adapter
4. Redundant arrays of independent disks
5. Tape drive errors, Human error
6. All of the above (Drive failures, Connector failures, Controller failures, Cache failures, RAID and array failures, Cable termination failures)
7. MBR
8. Event Viewer
9. Storage or disk array

Chapter 19: Common Security Issues

1. Integrity
2. Vertical privilege escalation
3. All of the above (Windows Update service isn't running, The ownership of the application is incorrect, Malware may have deleted or renamed the application, The hard disk location of the application is corrupted)
4. The file is in use by a remote connection
5. Event Viewer
6. free
7. TCP/UDP port
8. Inactive
9. Idle anti-malware software during peak hours
10. Port scanner

Other Books You May Enjoy

If you enjoyed this book, you may be interested in these other books by Packt:

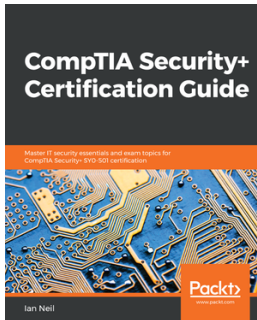


CompTIA Network+ Certification Guide

Glen D. Singh, Rishi Latchmepersad

ISBN: 978-1-78934-050-1

- Explain the purpose of a variety of networking concepts and implement them appropriately
- Understand physical security and common attacks while securing wired and wireless networks
- Understand the fundamentals of IPv4 and IPv6
- Determine and explain the appropriate cabling, device, and storage technologies
- Understand network troubleshooting methodology and appropriate tools to support connectivity and performance
- Use best practices to manage the network, determine policies, and ensure business continuity



CompTIA Security+ Certification Guide

Ian Neil

ISBN: 978-1-78934-801-9

- Get to grips with security fundamentals from Certificates and Encryption to Identity and Access Management
- Secure devices and applications that are used by your company
- Identify the different types of malware and virus and take appropriate actions to protect against them
- Protect your environment against social engineering and advanced attacks
- Implement PKI concepts
- Learn about secure coding techniques, quality control, and testing
- Troubleshoot common security issues

Leave a review - let other readers know what you think

Please share your thoughts on this book with others by leaving a review on the site that you bought it from. If you purchased the book from Amazon, please leave us an honest review on this book's Amazon page. This is vital so that other potential readers can see and use your unbiased opinion to make purchasing decisions, we can understand what our customers think about our products, and our authors can see your feedback on the title that they have worked with Packt to create. It will only take a few minutes of your time, but is valuable to other potential customers, our authors, and Packt. Thank you!

Index

3

3-2-1 backup strategy 243

8

80-plus certification 24

A

AAA authentication protocols

Kerberos 254

Lightweight Directory Access Protocol (LDAP)
254

Remote Authentication Dial-In User Service
(RADIUS) 254

TACACS+ 254

Terminal Access Controller Access Control
System (TACACS) 254

access control criteria

group access control 294

role-based access control (RBAC) 294

specific criteria access control 295

access control entry (ACE) 295

access control levels

administration access control 298

filesystem access control 295

network access control (NAC) 298

security and distribution groups 298

access control list (ACL)

about 261, 293

Discretionary ACL (DACL) 265

dynamic 265

ethertype 265

extended ACLs 264

reflexive 265

router ACLs 261

standard ACLs 264

System ACL (SACL) 265

types 263

webtype 265

access control

about 294

criteria 294

levels 295

access list content

access control entry (ACE) 262

ACL identification 262

source and destination address 262

access point 261

ACE types

access-allowed ACE 266

access-denied ACE 266

system-audit ACE 266

Acronis 392

Active Directory (AD) 117, 198

Active Server Page (ASP) 9

Active to Precharge Delay 48

active/active server clusters

versus active/passive server clusters 211

address bus 50

Address Resolution Protocol (ARP) 136, 145, 146

address resolution

about 145

ARP 146

DNS 146

Windows Internet Name Service (WINS) 147

administrative tools

about 399

disk management 399

disk partitioning tools 400

map 401

mount command 401

net use 401

RAID arrays 402

air conditioning (HVAC) 171

- air cooling 29
- air flow 29
- alternating current (AC) 20
- Amperes/Amps 20
- anti-malware configurations
 - devices 416
 - review 416
 - schedule 416
 - updates 416
- AOMEI Partition Assistant 400
- Apple File System (APFS) 110
- application programming interface (API) 123
- application servers
 - about 8, 9
 - LAN application servers 8
 - mail servers 11
 - messaging servers 11
 - network services servers 11
 - print servers 12
 - proxy server 13
 - query-based application servers 9
 - Routing and Remote Access Service (RRAS) 14
 - virtual server 14
- application-level gateways 13
- asset management
 - about 184
 - Information Technology Asset Management (ITAM) 184
 - IT life cycle asset management (LCAM) 185
- Asymmetrical Multiprocessing (ASMP)
 - versus Symmetrical Multiprocessing (SMP) 35
- asynchronous replication 236, 238
- authentication 268
- authentication methods
 - about 252
 - one-time password 252
 - password 252
 - public-key 252
 - zero-knowledge 252
- authentication protocols
 - AAA authentication protocols 254
 - about 251
 - Internet Protocol Security (IPSec) 256
 - point-to-point authentication protocols 253
 - Secure Sockets Layer (SSL)/Transport Layer

- Security (TLS) 255
- Automatic Private IP Addressing (APIPA) 366
- availability 76

B

- backup media 241
- backup media integrity 243
- backup media retention 243
- backup
 - about 236
 - data backup 239
 - media storage 242
- Bacula 392
- bad sectors
 - logical bad sectors 391
 - physical bad sectors 391
- baseband (Base) 162
- Baseboard Management Controller (BMC) 205
- Basic Input Output System (BIOS) 95
- battery-backed cache memory 80
- Bayonet Neill-Concelman (BNC) 159
- beep codes 207
- Better Filesystem (Btrfs) 109
- BIA project
 - documenting phase 233
 - evaluating phase 233
 - gathering phase 233
 - presenting phase 233
- binary
 - versus decimal 66
- BIOS 58
- blade technology 18
- Blue Screen of Death (BSOD) 344, 390
- boot sequence
 - about 104
 - disk, preparing for OS 106
 - firmware 104
- Bootstrap Protocol (BOOTP) 305
- Bourne-Again Shell (Bash) 102
- bring your own device (BYOD) 306
- broadcast addresses 140
- broadcast domains 135
- broadcast messages
 - uses 136
- browser cache 41

- browser zones 250
- bus channels
 - external bus 50
 - internal bus 50
- bus
 - address bus 50
 - control bus 50
 - data bus 50
 - Peripheral Component Interconnect (PCI) bus 51
 - width 50
- business continuity plan (BCP)
 - about 231
 - components 232
 - continuity of operations 233
 - risk assessment 233
- business impact analysis (BIA) 232

C

- cache memories
 - characteristics 42
- cache
 - about 41
 - browser cache 41
 - disk cache 41
 - memory cache 41
 - processor cache 41
- cannot mount drive issue 345
- case fan 29
- categories, backup media
 - linear access 241
 - random access 241
 - removable media 241
- category cabling 161
- Challenge-Handshake Authentication Protocol (CHAP) 253
- change control process 195
- change management
 - about 195
 - change control process 195
 - device driver updates 200
 - firmware updates 201
 - operating system updates 200
 - preventive maintenance (PM) 208
- checksum 420
- cipher 420
- Cisco Inter-Switch Link (ISL) 374
- classful IP addressing 130
- Classless Interdomain Routing (CIDR) 136
- clean install 106
- cluster-to-cluster replication 238
- coaxial cabling 158
- cold spares 84
- collision domains 134
- Collision Sense Multiple Access/Collision Detection (CSMA/CD) 134
- column address strobe (CAS) 47
- command execution time 66
- command-line interface (CLI) 93
- Common Internet File System (CIFS) 75
- compact discs (CDs) 86
- Complementary Metal Oxide Semiconductor (CMOS) 95
- concepts, Information Technology Asset Management (ITAM)
 - asset inventory 185
 - asset tags 185
 - end-of-life (EOL) 186
- conditions, detecting in environmental monitoring
 - humidity 282
 - physical intrusion 282
 - smoke 282
 - temperature 282
 - voltage and power 282
- conditions, for hardware issue identification
 - access failure 322
 - intermittent failure 322
 - POST failure 322
 - random stopping or rebooting 322
- confidentiality, integrity, and availability (CIA) 101, 177, 299
- configuration issues
 - default gateway unavailable 376
 - Dynamic Host Configuration Protocol (DHCP) server 368
 - firewall failure 377
 - miscellaneous issues 378
 - misconfigured devices 369
- configuration policies, security zones
 - access control policies 249
 - security policies 249

configuration specifications 65

configuration

about 57

BIOS 58

UEFI 58

continuity of operations 233

control bus 50

cooling systems 29

copper cabling

about 154

category cabling 161

coaxial cabling 158

EIA/TIA 568 facility standards 160

Ethernet cable standards 162

network connectors 159

twisted-pair cabling 155

CPU multiplier 43

CPU stepping 43

CPUs

about 34

Advanced RISC Machine (ARM) servers 43

cache memory 41

cache memory levels 42

common server sockets 40

microprocessors 35

multiple core processing 36

multiple-instruction, multiple-data (MIMD) 36

multiple-instruction, single-data (MISD) 36

packages 38

single-instruction, multiple-data (SIMD) 36

sockets 39

Symmetrical Multiprocessing (SMP), versus
Asymmetrical Multiprocessing (ASMP) 35

write-back/write-through cache 42

CSMA/Collision Avoidance (CSMA/CA) 134

current 20

Customer-Replaceable Units (CRUs) 207

cyclic redundancy check (CRC) 389

D

Darik's Boot and Nuke (DBAN) 2.3.0 301

data backup

about 239

archive bit 239

data restore, versus OS restore 241

methods 239

data bus 50

data encryption

about 299

storage encryption 299

Data Link Control (DLC) 144

data replication

about 236

asynchronous replication 236, 238

synchronous replication 236, 238

data security issues

about 410, 412, 413

reasons 414, 415, 416, 418

data storage device issues

about 385

causes 386

reasons 390

data storage systems

about 70

direct-attached storage (DAS) 71

network-attached storage (NAS) 71

storage area network (SAN) 72

data storage

devices 62

specifications 62

data transfer rate (DTR) 53, 66, 70, 161

data

deleting 300

disposal 300

erasing, from disk 300

formatting 301

retention 300

shred action 301

wiping 301

database servers 9

Ddrescue 353

decimal

versus binary 66

dedicated file servers 10

default gateway unavailable issue

reasons 376

delta 21

Demilitarized zone (DMZ) 250

Deployment Image Servicing and Management
(DISM) 353

- Desktop Management Interface (DMI) 123
- device specifications 65
- device specifications, disk drives
 - access time 65
 - I/O operations per second (IOPS) 66
 - interface 65
 - revolutions per minute (RPM) 65
 - throughput 66
- diagrams, system documentation
 - architecture diagram 188
 - data flow diagram (DFD) 188
 - network diagram 188
- digital versatile/video discs (DVDs) 86
- direct current (DC) 20
- direct-attached storage (DAS) 71
- directory servers 9
- disaster recovery plan (DRP)
 - about 231, 234
 - recovery plans 234
 - recovery sites 235
- discretionary access control (DAC) 294
- disk array 401
- disk boot failure 344
- disk cache 41
- disk compression 82
- disk drive
 - destroying physically 302
 - disk platters, removing 302
 - penetrating 302
 - shredding 302
- disk input/output operations per second (IOPS) 202
- disk management 399
- disk partitioning tools 400
- disk quotas 80
- disk storage capacity planning 84
- DNS search
 - about 146
 - working 146
- domain 117
- Domain Name System (DNS) 11, 145, 146, 368
- domain search 147
- domain suffix 147
- domain user account
 - about 114

- creating 117
- network, connecting 119
- remote installations 123
- unattended 123
- workstation, adding 118
- dot-decimal formats 129
- dot1x 260
- Double Data Rate (DDR) RAM 44
- drive and connector failures
 - cable issues 395
 - HDD issues 394
- dual in-line memory modules (DIMMs) 45
- Duplicati 392
- dynamic allocations 94
- Dynamic Host Configuration Protocol (DHCP) 12, 136, 146
- Dynamic Host Configuration Protocol (DHCP) server
 - addresses 369
 - APIPA 368
- dynamic linking 94
- dynamic loading 94
- dynamic RAM (DRAM) 44

E

- EAP over LAN (EAPoL) 260
- EAP over Wireless (EAPoW) 260
- EFI System Partition (ESP) 58
- EIA/TIA 568 facility standards
 - backbone cable 160
 - cross-connect 160
 - entrance facility 160
 - horizontal cable 161
 - telecommunication rooms/IDF 161
 - work areas 161
- eight position/eight contact (8P8C) connector 159
- electrical power
 - AC versus DC / 110V versus 230V 20
 - advantages 282
 - concepts 20
- electro-magnetic interference (EMI) 155
- Electronics Industries Alliance (EIA) 18
- Electronics Industry Association (EIA) 156
- email servers 11
- encryption 268

- endpoint security 306
- environmental controls
 - about 281
 - HVAC 281
 - room temperature 281
 - row and rack temperatures 281
- environmental threats
 - dust 277
 - earthquake 277
 - extreme weather 277
 - fire/explosion 277
 - flood 277
 - lightning 277
 - pests 277
- error-correction code (ECC)
 - versus non-ECC 48
- Ethernet cable standards 162
- Ethernet cable
 - categories 161
- examples, technical threat
 - access beyond authorization 279
 - hardware failure 279
 - inadequate operating procedures 279
 - intruder attacks 279
 - unauthorized modifications to software or hardware 279
- expansion cards
 - about 55
 - Host Bus Adapter (HBA) 55
 - network interface controller (NIC) 55
 - Redundant Array of Independent Disks (RAID)
 - controller 55
 - riser cards 56
- ext4 75
- extended ACLs 264
- Extended File System version 3 (ext3) 75
- Extended TACACS (XTACACS) 254
- Extensible Authentication Protocol (EAP) 253

F

- failed components, hardware maintenance
 - Customer-Replaceable Units (CRUs) 207
 - Field-Replaceable Units (FRUs) 208
- failed components
 - replacing 84

- failed hardware and power component replacement
 - methods
 - cold swap 84
 - hot swap 84
 - warm swap 84
- fault tolerance
 - about 83, 209
 - characteristics 83
- FC SAN 73
- fibre channel (FC) 69, 396
- Fibre Channel (FC) 168
- Fibre Channel over Ethernet (FCoE) 73
- Fibre Optic Cable Interchangeability Standard (FOCIS) 166
- fibre-optic cable
 - layers 164
- fibre-optic cabling
 - about 163
 - connectors 166
 - Lucent Connector (LC) 168
 - modes 165
 - multi-mode (MM) 165
 - single mode (SM) 165
 - Small Form-Factor Pluggable (SFP) 168
 - Standard Connector (SC) 168
 - Straight Tip (ST) 168
- fibre-optic link (FL) 163
- Field-Replaceable Units (FRUs) 208
- File Allocation Table (FAT) 389
- File Allocation Table 32 (FAT32) 109
- file integrity 410
- File integrity monitor (FIM) 410
- file servers
 - about 10
 - dedicated file servers 10
 - non-dedicated file servers 10
- file sharing 75
- File Transfer Protocol (FTP) 10, 305
- filesystems
 - about 74, 109
 - by OS 109
 - formatting 109
 - journaling 110
 - operating systems 75
 - special function 110

- firewall failure 377
- firewall zones 249
- firmware
 - BIOS 105
 - UEFI 105
- flapping 375
- Flash-Friendly File System (F2FS) 110
- form factors
 - about 15, 63
 - blade technology 18
 - large form factor (LFF) 64
 - rack mounts 17
 - small form factor (SFF) 63
 - tower servers 16
- formatting levels, data storage device
 - high-level formatting (HLF) 301
 - low-level formatting (LLF) 301
- fragmentation 357
- frame 148
- front-side bus (FSB) 43
- full-disk encryption (FDE) 299
- Fully Qualified Domain Name (FQDN) 372
- fully-qualified domain name (FQDN) 146

G

- gateway proxy servers 13
- Gigabit Ethernet (GbE) 168
- Globally Unique Identifier 106
- gPXE utilities 123
- graphical user interface (GUI) 93
- graphics processing units (GPUs) 28
- ground 20
- GUID partition table (GPT) 106

H

- hard disk drive (HDD)
 - about 236, 356, 386
 - disk capacity 66
 - specifications 65, 66
 - versus solid-state drive (SSD) 67, 68
- hardening
 - about 303
 - application hardening 304
 - endpoint security 306
 - hardware hardening 305

- OS hardening 303
 - system hardening 304
- hardware administration 176
- hardware configuration
 - for virtual environment 225
- hardware hardening
 - about 305
 - host hardware hardening 305
 - network device hardening 305
- hardware issues
 - about 321
 - identifying 322
 - memory failure 327
 - overheating 324
 - POST failure 323
 - processor failure 325
 - symptoms 323
- hardware maintenance
 - about 202
 - failed components, replacing 207
 - LED server status indicators 205
 - server monitoring systems 202, 204
- hardware-related issues, storage system
 - cache battery failure 397
 - controller failure 396
 - disk cache off 397
 - Host bus adapter (HBA) failure 396
 - storage array issues 397
- hardware-related software issues
 - about 344
 - Blue Screen of Death (BSOD) 344
 - cannot mount drive 345
 - disk boot failure 344
- HDD drive interfaces
 - FC technology 70
 - PATA 69
 - SAS 70
 - SATA 70
 - SCSI 70
- HDD interfaces
 - characteristics 65
- HDD issues 394
- HDD performance
 - issues 387, 388, 389, 390
- heap 355

- heartbeat network 214
- heat sink 29
- heating 171
- hexets 141
- Hierarchical File System Plus (HFS+) 110
- high availability 82, 196, 209, 214
- host bus adapter (HBA) 18, 69
- Host Bus Adapter (HBA) 55
- host-based firewall 251
- host-based intrusion detection system (HIDS) 306
- hostname
 - configuring 111
 - configuring, on Linux server 112
 - configuring, on Windows server 111
- hosts 138
- hot spares 84
- hot-swappable devices 214
- HP integrated Lights-Out (iLO) 181
- hypervisor 14, 223, 224

I

- I/O operations per second (IOPS) 66
- IDS
 - heuristic IDS 306
 - signature-based IDS 306
- IEEE 802.1x
 - about 260
 - authentication process 261
- IEEE Ethernet cable
 - types 162
- IEEE Ethernet standard
 - cable identification 162
 - DTR 162
 - signaling mode 162
- ifconfig 380
- Information Technology Asset Management (ITAM) 184
- input/output (I/O) operations 11
- input–process–output (IPOS) model 91
- Institute for Electrical and Electronics Engineers (IEEE) 145
- integrated circuit (IC) 37
- Integrated Dell Remote Access Controller (iDRAC) 181
- intermediate distribution facility (IDF) 161
- internal-facing proxy servers 13
- Internet Assigned Numbers Authority (IANA) 131
- Internet Group Management Protocol (IGMP) 14
- Internet Key Exchange (IKE) 259
- Internet Message Access Protocol (IMAP) 371
- Internet Protocol (IP) 179
- Internet Protocol Security (IPSec) 256
- Internet SCSI (iSCSI) 73
- Internet Service Providers (ISPs) 269
- internet service providers (ISPs) 130
- Internet-facing (forward) proxy servers 13
- intrusion detection system (IDS) 279, 306
- intrusion prevention system (IPS) 279
- IP addressing
 - about 129
 - address resolution 145
 - Media Access Control (MAC) addressing 144
 - subnetting 137
- IP Security (IPSec) 269
- ipconfig 380
- IPSec modes
 - about 258
 - AH transport mode 258
 - ESP transport mode 258
- IPSec policies
 - about 256
 - filters 257, 258
- IPv4 address
 - about 129
 - broadcast domains 135
 - classful IP addressing 130
 - Classless Interdomain Routing (CIDR) 136
 - collision domains 134
 - host IDs 132
 - LAN addressing 130
 - Network Address Translation (NAT) 133
 - network IDs 132
 - private IP addresses 131
 - structure 129
- IPv6 address
 - about 140
 - anycast 144
 - compression 143
 - leading zero compression 143
 - multicast 144

- network ID 143
- reserved prefixes 142
- structure 141
- unicast 144
- iSCSI SAN 73
- IT asset disposition (ITAD) 186
- IT life cycle asset management (LCAM) 185

J

- JavaServer Pages (JSP) 9
- just a bunch of disks (JBOD) 71

K

- Kerberos 254
- Keyboard, Video, Mouse (KVM) 178
- KVM extender 180

L

- LAN addressing 130
- LAN application servers 8
- Land Grid Array (LGA) 38
- LANDesk Client Manager (LDCM) 123
- large form factor (LFF) 64
- Layer 2 Tunneling Protocol (L2TP) 269
- LED server status indicators
 - about 205
 - beep codes 206
 - Liquid Crystal Display (LCD) messages 205
- Lightweight Directory Access Protocol (LDAP) 254
- Linear Tape Open (LTO) 86
- linear tape-open (LTO) 241
- Linux file recovery 353
- Linux's GParted 109
- liquid cooling 29
- Liquid Crystal Display (LCD) messages 205
- load balancing
 - about 212, 213
 - algorithms 212
- Local Area Network (LAN) 366
- local user account
 - about 114
 - creating 114
- logical configurations 65

M

- Mac OS Disk Utility 109
- magnetic tape 86
- mail servers 11
- mail transport agent (MTA) 11
- main memory
 - about 44
 - color-coded RAM slots 49
 - dual channel memory 48
 - error-correction code (ECC), versus non-ECC 48
 - memory timing 47
 - RAM 44
- man-made threats
 - accidents 277
 - malware 277
 - theft 278
 - unauthorized access 278
 - vandalism 278
- master boot record (MBR) 104
- Master Boot Record (MBR) 387
- Master File Table (MFT) 389
- mean time to repair/restoration (MTTR) 215
- Media Access Control (MAC) 131
- Media Access Control (MAC) addressing 144
- media failures
 - hard disk media 391
 - magnetic tape media 392
 - optical drives 393
 - SDD media 391
- media storage
 - about 242
 - recovery point objective (RPO) 242
 - recovery time objective (RTO) 242
- memory allocation
 - best fit 94
 - first fit 94
 - next fit 94
 - worst fit 94
- memory cache 41
- memory management
 - about 94
 - dynamic linking 94
 - dynamic loading 94
 - memory allocation 94

- memory timings
 - CAS latency (CL) 47
 - RAS precharge (tRP) 48
 - RAS to CAS delay (tRCD) 47
 - row active time (tRAS) 48
- Message Digest 5 (MD5) 261, 410
- messaging servers
 - about 11
 - point-to-point messaging servers 11
 - publish-subscribe messaging servers 11
- metal-oxide-semiconductor field-effect transistor (MOSFET) 26
- methods, data backup
 - bare metal backup 240
 - copy backup 239
 - differential backup 239
 - full backup 239
 - incremental backup 239
 - open file backup 240
 - selective backup 240
 - snapshot 240
- methods, replication
 - disk-to-disk replication 238
 - server-based replication 238
- microkernel 102
- MiniTool Partition Wizard 400
- mirroring 76
- misconfigured devices
 - email issues 370, 371
 - hosts file configuration 372
 - misconfigured NIC 372
 - routing issues 374
 - switching issues 374
 - VLAN configuration errors 374
- monolithic 102
- mount command 382
- multi-factor authentication (MFA)
 - about 273
 - authentication factors, categories 275
 - passwords 274
- multi-mode cable 163
- multiple core processing 36
- multiple-instruction, multiple-data (MIMD) 36
- multiple-instruction, single-data (MISD) 36
- multiprocessors 35

N

- namespace 10
- National Electrical Manufacturers Association (NEMA) 26
- negative 48V 21
- net command 382
- NetBIOS over TCP/IP (NBT) 75
- NetBIOS over TCP/IP status (nbtstat) 382
- network adapters 55
- Network Address Translation (NAT) 133
- network addresses 140
- network administration 177
- network boot program (NBP) 123
- network cable
 - cable ducts 169
 - cable hangers 170
 - cable raceways 170
 - cable sleeves 170
 - cable ties 170
 - Cable trays 170
 - installing 168
- network configuration
 - about 110
 - hostname, configuring 111
- network connection configurations, VM
 - direct access (bridged) 227
 - host-only 227
 - Network Address Translation (NAT) 227
- network connectors 159
- Network Control Protocol (NCP) 129
- network device
 - hardening, steps 305
- Network File System (NFS) 10, 71
- network hardware
 - configuring 177
 - KVM interfaces 178, 180
 - maintaining 177
 - serial interfaces 180, 181
 - updating 177
- network interface card (NIC) 123
- network interface controller (NIC) 55, 227
- network issues
 - about 365
 - configurations 368

- Internet connectivity 366, 367
- network operating system (NOS) 11, 91, 177
- network server
 - about 92
 - operating system (OS) functions 93
 - operating systems 92
 - server functions 92
- network services servers 11
- Network Status (netstat) 382
- Network Time Protocol (NTP) 12
- network-attached storage (NAS) 10, 71, 236
- network-based firewall 251
- network-based intrusion detection system (NIDS) 306
- network-based operating system administration 182
- network
 - connecting to 119
 - features, adding 120
 - PC, connecting to 119
 - server roles, adding 120
- New Technology Filesystem (NTFS) 109
- non-dedicated file servers 10
- non-ECC
 - versus error-correction code (ECC) 48
- non-hot-swappable devices 214, 215
- NOS optimization
 - about 124
 - bandwidth 124
 - high availability and fault tolerance 124
 - load balancing 124
 - Quality of Service (QoS) 124
- nslookup 381

O

- object-oriented programming (OOP) 94
- one-time programmable (OTP) 105
- open proxy servers 13
- Open Shortest Path First (OSPF) 14
- operating system (OS) functions
 - about 93
 - application 101
 - control hardware 95
 - coordinate hardware 95
 - data 101

- internal file management 100
- memory management 94
- network file management 100
- resource security 101
- user 101
- user/computer communications 93
- operating systems
 - filesystems 75
 - issues 345, 346, 347
- optical disc drive (ODD) 86
- optical storage 86
- Organizationally Unique Identifier (OUI) 145
- OS hardening
 - establish and monitor baselines 303
 - group and user accounts 303
 - patch management 303
 - unused software 303
- OS, primary parts
 - about 102
 - filesystem 103
 - kernel 102
 - shell 102
- OS
 - about 103
 - hardware configuration 103

P

- packet 148
- Packet Internet Groper 378
- Parallel Advanced Technology Attachment (PATA) 69
- passive cooling 29
- Password Authentication Protocol (PAP) 253
- patch management 195, 196, 197
- PCI conventional standard 51, 52
- PCI-e bus standard 54
- PCI-Express (PCI-e) 51
- PCI-extended (PCI-X) 51
- Peripheral Component Interconnect (PCI) bus
 - about 51
 - fit standards 52
 - size standards 52
- Personal Identity Verification (PIV) 275
- personally identifiable information (PII) 420
- physical NICs (pNICs) 228

- physical security devices
 - badges and cards 280
 - biometric devices 280
 - lockable cabinets and safes 280
 - locks and keys 280
 - mantrap 280
 - rack mount cabinet 280
 - security cameras and sensors 281
- physical security
 - about 276
 - detection 276
 - deterrence 276
 - devices 280
 - threats 276
- Picture Password 275
- ping 379
- Pluggable Authentication Module (PAM) 346
- point-to-point authentication protocols
 - about 253
 - Challenge-Handshake Authentication Protocol (CHAP) 253
 - Extensible Authentication Protocol (EAP) 253
 - Password Authentication Protocol (PAP) 253
- point-to-point messaging servers 11
- Point-to-Point Protocol (PPP) 14
- Point-to-Point Tunneling Protocol (PPTP) 269
- Port Address Translation (PAT) 134
- port scanners 419
- port security
 - about 259
 - access control list (ACL) 261
 - dynamic locking 260
 - IEEE 802.1x 260
 - static locking 260
 - wildcard masks 267
- port-based network access control (PNAC) 259
- Portable Operating System Interface (POSIX) 351
- ports
 - about 148
 - registered ports 149
 - well-known ports 148
- Post Office Protocol 3 (POP3) 11
- power distribution unit (PDU)
 - about 285
 - Automatic Transfer Switch (ATS) PDU 286
 - dual circuit PDU 286
 - floor-mounted 286
 - hot-swap PDU 286
 - load capacity 288
 - metered PDU 286
 - monitored PDU 286
 - power rating 287
 - rack-mounted 286
 - ratings 287
 - standard (basic) PDU 286
 - switched PDU 286
- power sources
 - automated shutdown of attached devices 284
 - Uninterruptable Power Supplies (UPS) 283
- power supply unit (PSU) 19, 390
- power-on self-test (POST) 105, 322
- Preboot Execution Environment (PXE) 123
- primary load types
 - point load 289
 - rolling load 289
 - static load 289
- print server 358
- printer servers 12
- private IP addresses 131
- privilege escalation
 - horizontal 411
 - vertical 411
- process ID (PID) 418
- processor cache 41
- programmable logic array (PLA) 105
- programmable read-only memory (PROM) 105
- Protected EAP (PEAP) 261
- protocol data units (PDUs) 148
- protocols 148
- proxy servers
 - about 13
 - gateway proxy servers 13
 - internal-facing proxy servers 13
 - Internet-facing (forward) proxy servers 13
 - open proxy servers 13
 - reverse proxy servers 13
- PSU
 - about 23
 - selecting 24, 25
- public key infrastructure (PKI)

- about 268
- certificate authority (CA) 268
- certificate request database 268
- certificate store 268
- features 268
- registration authority (RA) 268
- publish-subscribe messaging servers 11
- PXELINUX utilities 123

Q

- Quad-SFP (QSFP) 168
- query-based application servers 9
- Quick Emulator (QEMU) 222

R

- rack mounts 17
- radio frequency (RF) 119, 154
- RAID 0 77
- RAID 1 77
- RAID 10 78
- RAID 5 78
- RAID 6 78
- RAID arrays 402
- RAID
 - about 76
 - implementing 79, 80
 - levels 77, 78
 - mirroring 76
 - striping 76
- RAM packaging 45
- RAM
 - about 44
 - DDR-SDRAM 45
 - DDR2/DDR3 45
 - DDR4-SDRAM 45
 - dynamic RAM (DRAM) 44
 - static RAM (SRAM) 44
 - synchronous DRAM (SDRAM) 44
- read-only memory (ROM) 57, 105
- reasons, security issues
 - active services 414
 - anti-malware configurations 416
 - inactive accounts 416
 - misconfigured permissions 416
 - open ports 417

- rogue process 418
- reasons, storage device issues
 - drive and connector failures 393
 - media failure 391
- recovery point objective (RPO) 242
- recovery sites, disaster recovery plan (DRP)
 - cold site 235
 - hot site 235
 - warm site 235
- recovery time objective (RTO) 242
- Reduced Instruction Set Computer (RISC) 43
- Redundant Array of Independent Disks (RAID) 390
- Redundant Array of Independent Disks (RAID)
 - controller 55
- redundant disk array controller (RDAC) 80
- registered jack-45 (RJ-45) connector 159
- registered ports 149
- Reiser Filesystem (ReiserFS) 110
- Remote Authentication Dial-In User Service (RADIUS) 254
- Remote Desktop Protocol (RDP) 182
- Remote Frame Buffer (RFB) 183
- Remote Installation Service (RIS) 123
- Remote Server Administration Tools (RSAT) 182
- replication
 - about 236
 - data replication 236
 - methods 238
- resistance 20
- Reverse Address Resolution Protocol (RARP) 136
- Reverse ARP (RARP) 146
- reverse proxy servers 13
- riser cards 56
- risk assessment 233
- router ACLs 261
- Routing and Remote Access Service (RRAS)
 - about 14, 83
 - firewall 14
 - remote access 14
 - router 14
- Routing Information Protocol (RIP) 14
- row address strobe (RAS) 47

S

- Samba 75
- SAN communications
 - about 73
 - FC SAN 73
 - iSCSI SAN 73
- SAN fabric 72
- sector 357
- Secure Hash Algorithm (SHA)-1 420
- Secure Shell (SSH) 182, 421
- Secure Sockets Layer (SSL)/Transport Layer Security (TLS)
 - about 255
 - handshake process, steps 255
- Security Configuration and Analysis (SCA) 197
- security identifier (SID) 294
- security information and event management (SIEM) 306
- security tools
 - about 419
 - checksums 420
 - cipher 420
 - port scanners 419
 - sniffers 420
 - Telnet 421
- security zones
 - about 249
 - browser zones 250
 - configuration policies 249
 - Demilitarized zone (DMZ) 250
 - firewall zones 249
- sensitive compartmented information facility (SCIF) 190
- Serial Advanced Technology Attachment (SATA) 55
- Serial Advanced Technology Attachment (SATA) connections 395
- serial attached SCSI (SAS) 69
- Serial Attached SCSI (SAS) 395
- server clustering
 - about 210
 - benefits 210
 - disadvantages 210
- Server Manager 117
- Server Message Block (SMB) 75
- server power systems
 - 80-plus certification 24
 - about 19
 - delta 21
 - electrical power 20
 - PSU 23
 - redundancy 26
 - wattage 23
 - wye 21
- server roles
 - about 7
 - application servers 8
 - database servers 9
 - directory servers 9
 - file servers 10
- server virtualization
 - about 222
 - full virtualization 222
 - operating system level virtualization 223
 - para-virtualization 222
- Server+ exam
 - crossover cables 157
 - patch cables 157
 - rollover cables 157
 - straight-through cables 158
- server-based replication
 - about 238
 - cluster-to-cluster replication 238
 - server-to-self replication 238
 - server-to-server replication 238
- server-to-self replication 238
- server-to-server replication 238
- service level agreements (SLA) 83, 215
- service manuals 186
- Service Message Block/Common Internet File System (SMB/CIFS) 10
- settle time 66
- signal-to-noise ratio (SNR) 162
- Simple Message Transport Protocol (SMTP) 11
- Simple Network Management Protocol (SNMP) 177, 225
- single data-rate (SDR) 44
- single point of failure (SPOF) 358
- single-factor authentication (SFA) 274
- single-instruction, multiple-data (SIMD) 36

- single-mode cable 163
- single-phase power 21
- site-specific threat
 - communication system failure 278
 - computing equipment failure 278
 - fire suppression system failure 278
 - Heating, Ventilating, and Air Conditioning (HVAC) failure 278
 - key personnel resignation/departure 278
 - power outage/brown-out 279
- Small Computer Serial Interface (SCSI) 69
- Small Computer System Interface (SCSI) 55
- small form factor (SFF) 63
- small office/home office (SOHO) 179
- Small Office/Home Office (SOHO) computer 16
- small outline dual in-line memory modules (SO-DIMMs) 45
- sniffers 420
- snoop server 420
- software issue, causes
 - about 348
 - corrupted files 352
 - fragmentation 357
 - hard disk space issue 354
 - lack of system resources 355
 - log files 360, 361
 - operating system monitoring tools 361
 - printing issues 358, 359
 - User Account Control (UAC) 348
 - virtual memory issues 356
 - Windows UAC 349
- software issues
 - about 343
 - hardware-related software issues 344
 - operating system issues 346, 347
 - operating systems issues 345
- Software-as-a-Service (SaaS) 8
- software-defined network (SDN) 226
- software-related issues, storage system
 - formatting 396
- software-related issues
 - corrupted data 396
 - deleted data 396
- solid-state drive (SSD)
 - about 241
 - configuration 69
 - specifications 69
 - versus hard disk drive (HDD) 67, 68
- specific criteria access control
 - location 295
 - time 295
- specific criteria access
 - application 295
 - transaction type 295
- split-phase power 21
- standard ACLs 264
- standards, for creating secure area
 - electrical safety 289
 - fire security 289
 - lighting 288
 - security 289
 - space 288
- standby UPS 284
- static allocations 94
- static RAM (SRAM) 41, 44
- storage area network (SAN)
 - about 72, 237, 390
 - Logical Unit Number (LUN) masking 74
 - Logical Unit Number (LUN) zoning 73
- storage array issues
 - backplane failure 398
 - improper RAID configuration 397
 - mismatched drives 398
 - RAID rebuilds 398
- storage device issue
 - reasons 393
- storage devices
 - about 85
 - magnetic tape 86
 - optical storage 86
- storage monitoring tools 404, 405
- storage processor (SP) 80
- storage system issues
 - about 395
 - hardware-related issues 396
 - software-related issues 396
- striping 76
- subnet mask 137, 139
- subnets 138
- subnetwork (subnet) 137

- substitution cipher 420
- supernetting 136
- swap file 356
- swap space 110
- Symmetrical Multiprocessing (SMP)
 - versus Asymmetrical Multiprocessing (ASMP) 35
- symptoms, hardware issues
 - BSoD/RSoD 323
 - data integrity 323
 - I/O 323
 - muddled display 323
 - unusual noises 323
- synchronous replication 236, 238
- system administrator (sysadmin) 177
- system and network documentation 187
- System Center Configuration Manager (SCCM) 197
- system documentation
 - about 186, 189, 190
 - diagrams 188
 - qualities 189
 - service manuals 186
 - types 190
- System File Checker (SFC) 352
- system hardening
 - automatic patch application 304
 - logging 304
 - physical security 304
 - sharing, disabling 304
 - user account policies 304
- system heat 28
- system resources
 - about 95
 - direct memory access (DMA) addresses 99
 - I/O addresses 97
 - interrupt requests (IRQs) 96
 - memory addresses 98
 - using 95
- system sensitive documentation
 - storing 190

T

- TACACS+ 254
- tape library 86
- Telecommunications Industry Association (TIA)

- 157
- Telnet 421
- Terminal Access Controller Access Control System (TACACS) 254
- TestDisk 353
- thermal design power (TDP) 28
- thermal paste 29
- thermal stress 28
- Thinnet 162
- threats, physical security
 - about 276
 - environmental threats 277
 - man-made threats 277
 - site-specific threats 278
 - technical threats 279
- three-phase power 22
- Time-to-Live (TTL) 379
- tools, network-based hardware administration
 - about 181
 - HP integrated Lights-Out (iLO) 181
 - Integrated Dell Remote Access Controller (iDRAC) 181
 - KVM over IP switch 181
- tools, network-based operating system
 - administration
 - about 182
 - command line 184
 - Remote Desktop Protocol (RDP) 182
 - Remote Server Administration Tools (RSAT) 182
 - Secure Shell (SSH) 182
 - shell commands 184
 - Virtual Network Computing (VNC) 183
- top-level domain (TLD) 146
- tower servers 16
- traceroute 379
- tracert 379
- transaction processing system (TPS) 76
- Transport Layer Security (TLS) 261
- transposition cipher 420
- Trivial File Transport Protocol (TFTP) 123
- troubleshooting tools
 - about 379
 - ifconfig 380
 - ipconfig 380
 - mount 382

- nbtstat 382
- net 382
- netstat 382
- nslookup 381
- ping 379
- tracert 379
- tracert 379
- troubleshooting
 - steps 312
- Tunneled Transport Layer Security (TTLS) 261
- tunneling proxy servers 13
- twisted-pair (TP) 155
- twisted-pair cabling
 - about 155
 - shielded twisted pair (STP) 155
 - unshielded twisted pair (UTP) 155
- Two-Factor Authentication (2FA) 275

U

- UEFI 58
- unauthorized access, limiting
 - to devices 297
- Unified Extensible Firmware Interface (UEFI) 103
- Uninterruptable Power Supplies (UPS)
 - bypass 283
 - inverter 283
 - offline 284
 - online 284
 - rating 284
 - rectifier 283
 - switch 283
- uninterruptible power system (UPS) 76
- Unique Local Address (ULA) 143
- Universal Serial Bus (USB) 214
- Universally Unique Identifier (UUID) 106
- Unix Filesystem (UFS) 109
- unshielded twisted-pair (UTP) 179
- USB connector types
 - type A 57
 - type B 57
 - type C 57
- user account control (UAC) 295
- user accounts
 - about 114
 - domain user account, creating 117
 - local user account, creating 114
 - user/computer communications 93

V

- ventilation 171
- virtual devices 222
- virtual environment, hardware configuration
 - network connectivity 227
 - resource allocation 226
- virtual environment
 - hardware configuration 225
- virtual guests 224
- virtual hosts 224
- virtual internetworking devices
 - about 227
 - network interface controller (NIC) 227
 - physical NICs (pNICs) 228
 - virtual NICs (vNICs) 227
 - virtual routers (vRouters) 228
 - virtual switches (vSwitches) 228
- virtual LAN (VLAN)
 - about 73, 269
 - dynamic VLAN 269
 - static VLAN 269
- virtual local area network (VLAN) 221
- Virtual Local Area Network (VLAN) 374
- virtual machine (VM) 224
- Virtual Machine File System (VMFS) 110
- virtual machine management interface (VMMI) 225
- virtual machine manager (VMM) 222
- virtual memory 356
- Virtual Network Computing (VNC) 183
- virtual networking
 - about 221
 - components 221
 - hypervisors 223, 224
 - virtual devices 222
 - virtual guests 224
 - virtual hosts 224
 - virtual local area network (VLAN) 221
 - virtual machine (VM) 224
 - virtual private network (VPN) 221
 - virtual servers 222
- virtual NICs (vNICs) 227
- virtual private network (VPN) 14, 221, 269

- virtual routers (vRouters) 228
- virtual server 14, 222
- virtual switches (vSwitches) 228
- VLAN Member Policy Server (VMPS) 269
- Voice over Internet Protocol (VoIP) 12
- voltage 20
- voltage regulator modules (VRMs) 28
- Volume Shadow Copy Service (VSS) 240

W

- Wake-on-LAN (WOL) 305
- wattage 23
- watts 20
- web servers 9
- well-known port 148
- Wide Area Network (WAN) 366
- wildcard masks
 - about 267
 - public key infrastructure (PKI) 268
- Windows Disk Management utility 109, 399
- Windows file recovery
 - about 352

- corrupted file, replacing 353
- Deployment Image Servicing and Management (DISM) 353
- restore system 353
- System File Checker (SFC) 352
- Windows Internet Name Service (WINS) 147
- Windows Resource Protection (WRP) 352
- Windows Server Backup (WSB) 392
- Windows Server Update Services (WSUS) 197
- Windows UAC
 - access control 350, 351
- Wired for Management (WfM) 123
- write once, read many (WORM) 241
- wye 21

Y

- Yost cables 157

Z

- Z File System (ZFS) 110
- zip ties 170