ANTHONY SEQUEIRA

Cert Guide

Learn, prepare, and practice for exam success

AWS Certified Solutions Architect - Associate (SAA-C01)

PEARSON IT CERTIFICATION

AWS Certified Solutions Architect - Associate (SAA-C01) Cert Guide

Anthony Sequeira, CCIE No. 15626

AWS Certified Solutions Architect - Associate (SAA-C01) Cert Guide

ISBN-13: 978-0-7897-6049-4

ISBN-10: 0-7897-6049-5

Library of Congress Control Number: 2018963110

01 19

Editor-in-Chief
Mark Taub

Product Line Manager
Brett Bartow

Acquisitions Editor
Paul Carlstroem

Development Editor
Christopher Cleveland

Managing Editor
Sandra Schroeder

Senior Project Editor
Tonya Simpson

Copy Editor
Chuck Hutchinson

Indexer
Erika Millen

Proofreader
Abigail Manheim

Technical Editor
Ryan Dymek

Publishing Coordinator
Cindy J. Teeters

Cover Designer
Chuti Prasertsith

Compositor
codemantra

Contents at a Glance

Part VII: Appendixes

Table of Contents

Part VII: Appendixes

About the Author

Anthony Sequeira, CCIE No. 15626, is a seasoned trainer and author regarding various levels and tracks of Cisco, Microsoft, and AWS certifications. In 1994, Anthony formally began his career in the information technology industry with IBM in Tampa, Florida. He quickly formed his own computer consultancy, Computer Solutions, and then discovered his true passion—teaching and writing about information technologies.

Anthony joined Mastering Computers in 1996 and lectured to massive audiences around the world about the latest in computer technologies. Mastering Computers became the revolutionary online training company KnowledgeNet, and Anthony trained there for many years.

Anthony is currently pursuing his second CCIE in the area of Cisco Data Center! He is a full-time instructor at CBT Nuggets.

Dedication

I dedicate this book to my incredible daughter, Bella Sequeira. While she may never read it to master AWS, she can at least use it as a rather expensive coaster in her beautiful new Oregon home.

Acknowledgments

This manuscript was made truly great by the incredible technical review of Ryan Dymek. Sometimes I think he might have invented AWS.

I would also like to express my gratitude to Chris Cleveland, development editor of this book. I was so incredibly lucky to work with him again on this text. Like Ryan, he made this book several cuts above the rest.

About the Technical Reviewer

Ryan Dymek has been working with Amazon Web Services (AWS) for more than 9 years and holds all nine AWS certifications as well as various Google Cloud Platform (GCP) certifications. Ryan trains and advises some of the largest companies in the world on sound architectural practices in cloud strategy and DevOps principles. While working with business leaders, developers, and engineers, Ryan bridges the gap between business and technology, maintaining the understanding and skills required to be able to perform at a deep technical level. Ryan runs his own cloud consulting practice advising more than 20 companies on the Fortune 500 list and has helped many startups find their way in the cloud.

In addition to cloud and technical acumen, Ryan is a certified business coach personally trained by John Maxwell. He uses these professional skills not only to advise companies on best cloud practices but also on how to align with a business's needs and culture, making confident business and technical decisions and cultivating a transformation into DevOps.

We Want to Hear from You!

As the reader of this book, *you* are our most important critic and commentator. We value your opinion and want to know what we're doing right, what we could do better, what areas you'd like to see us publish in, and any other words of wisdom you're willing to pass our way.

We welcome your comments. You can email to let us know what you did or didn't like about this book—as well as what we can do to make our books better.

Please note that we cannot help you with technical problems related to the topic of this book.

When you write, please be sure to include this book's title and author as well as your name and email address. We will carefully review your comments and share them with the author and editors who worked on the book.

Email: feedback@pearsonitcertification.com

Introduction

The AWS Certified Solutions Architect - Associate is a cloud-related certification that tests a candidate's ability to architect effective solutions calling upon the most popular aspects of Amazon Web Services. The solutions architect candidates must demonstrate their skills on how to plan and implement a sophisticated design that saves costs, is secure, and perhaps most importantly, operates with excellence. Candidates are also required to know the most important facts regarding various services and their capabilities.

The AWS Certified Solutions Architect is an Associate-level cloud career certification. This certification is an excellent second step after the achievement of the AWS Certified Cloud Practitioner certification, although seasoned AWS users might choose to skip that entry-level exam. Following this certification, AWS offers a Professional level of certification for the solutions architect.

AWS also offers certifications in which you might be interested in different tracks. For example, a Developer track for AWS also includes Associate and Professional levels. Amazon also uses specialty certifications to deep dive into many different areas such as security and advanced networking.

NOTE The AWS Certified Solutions Architect - Associate certification is globally recognized and does an excellent job of demonstrating that the holder has knowledge and skills across a broad range of AWS topics.

The Goals of the AWS Certified Solutions Architect - Associate Certification

The AWS Certified Solutions Architect - Associate certification is intended for individuals who perform a solutions architect role. The certification seeks to validate your ability to effectively demonstrate knowledge of how to architect and deploy secure and robust applications on AWS technologies.

You should be able to define a solution using architectural design principles based on customer requirements. You should also be able to provide implementation guidance based on best practices to the organization throughout the lifecycle of the project.

Ideal Candidates

Although this text provides you with the information required to pass this exam, Amazon considers ideal candidates to be those who possess the following:

- One year of hands-on experience designing available, cost-efficient, fault-tolerant, and scalable distributed systems on AWS

- Hands-on experience using computing, networking, storage, and database AWS services
- Hands-on experience with AWS deployment and management services
- The ability to identify and define technical requirements for an AWS-based application
- The ability to identify which AWS services meet a given technical requirement
- Knowledge of recommended best practices for building secure and reliable applications on the AWS platform
- An understanding of the basic architectural principles of building on the AWS cloud
- An understanding of the AWS global infrastructure
- An understanding of network technologies as they relate to AWS
- An understanding of security features and tools that AWS provides and how they relate to traditional services

The Exam Objectives (Domains)

The AWS Certified Solutions Architect - Associate (SAA-C01) exam is broken down into five major domains. The contents of this book cover each of the domains and the subtopics included in them as illustrated in the following descriptions.

The following table breaks down each domain represented in the exam.

Domain	Percentage of Representation in Exam
1: Design Resilient Architectures	34%
2: Define Performant Architectures	24%
3: Specify Secure Applications and Architectures	26%
4: Design Cost-Optimized Architectures	10%
5: Define Operationally Excellent Architectures	6%
	Total 100%

1.0 Design Resilient Architectures

The Design Resilient Architectures domain is covered in Chapters 1 through 5. It covers critical information for designing and deploying highly available services and resources in the cloud. It comprises 34 percent of the exam. Topics include

- 1.1 Choose reliable/resilient storage.
- 1.2 Determine how to design decoupling mechanisms using AWS services.
- 1.3 Determine how to design a multitier architecture solution.
- 1.4 Determine how to design high availability and/or fault-tolerant architectures.

2.0 Define Performant Architectures

The Define Performant Architectures domain is covered in Chapters 6 through 9. This domain ensures that you consider and properly implement solutions that perform per the client requirements. This makes up 24 percent of the exam. Topics include

- 2.1 Choose performant storage and databases.
- 2.2 Apply caching to improve performance.
- 2.3 Design solutions for elasticity and scalability.

3.0 Specify Secure Applications and Architectures

The Specify Secure Applications and Architectures domain is covered in Chapters 10 through 12. This domain is critical, especially in today's landscape, because it assists you in an AWS implementation that features security at many layers. This builds a Defense in Depth type of strategy that increases your chances of warding off would-be attackers. It encompasses 26 percent of the exam. Topics include

- 3.1 Determine how to secure application tiers.
- 3.2 Determine how to secure data.
- 3.3 Define the networking infrastructure for a single VPC application.

4.0 Design Cost-Optimized Architectures

The Design Cost-Optimized Architectures domain is covered in Chapters 13 and 14. Here you learn about saving on costs where these costs could be the most substantial—in the areas of compute and storage. This domain embodies 10 percent of the exam. The topics include

- 4.1 Determine how to design cost-optimized storage.
- 4.2 Determine how to design cost-optimized compute.

5.0 Define Operationally Excellent Architectures

The Define Operationally Excellent Architectures domain is covered in Chapter 15. This chapter ensures you understand how to use AWS best practices around the three main steps of Prepare, Operate, and Evolve with AWS solutions. This domain makes up 6 percent of the exam. Topics include

- 5.1 Choose design features in solutions that enable operational excellence.

Steps to Becoming an AWS Certified Solutions Architect – Associate

To become an AWS Certified Solutions Architect - Associate, a test candidate should meet certain prerequisites (none of these are formal prerequisites) and follow specific procedures. Test candidates must qualify for the exam and sign up for the exam.

Recommended Experience

There are no prerequisites for the Solutions Architect - Associate certification. However, Amazon recommends that candidates possess the Certified Cloud Practitioner certification or equivalent knowledge.

NOTE Other certifications you might possess in related areas, such as Microsoft Azure, can also prove beneficial.

Signing Up for the Exam

The steps required to sign up for the Solutions Architect - Associate exam are as follows:

Step 1. Create an AWS Certification account at https://www.aws.training/Certification and schedule your exam.

Step 2. Complete the Examination Agreement, attesting to the truth of your assertions regarding professional experience and legally committing to the adherence of the testing policies.

Step 3. Submit the examination fee.

Facts About the Exam

The exam is a computer-based test. The exam consists of multiple-choice questions only. You must bring a government-issued identification card. No other forms of ID will be accepted.

TIP Refer to the AWS Certification site at https://aws.amazon.com/certification/ for more information regarding this, and other, AWS Certifications. I am also in the process of building a simple hub site for everything AWS Certification related at awscerthub.com. This site is made up of 100 percent AWS solutions. Of course!

About the Solutions Architect – Associate Cert Guide

This book maps directly to the topic areas of the exam and uses a number of features to help you understand the topics and prepare for the exam.

Objectives and Methods

This book uses several key methodologies to help you discover the exam topics on which you need more review, to help you fully understand and remember those details, and to help you prove to yourself that you have retained your knowledge of those topics. This book does not try to help you pass the exam only by memorization; it seeks to help you to truly learn and understand the topics. This book is designed to help you pass the AWS Certified Solutions Architect - Associate exam by using the following methods:

- Helping you discover which exam topics you have not mastered
- Providing explanations and information to fill in your knowledge gaps
- Supplying exercises that enhance your ability to recall and deduce the answers to test questions
- Providing practice exercises on the topics and the testing process via test questions on the companion website

Book Features

To help you customize your study time using this book, the core chapters have several features that help you make the best use of your time:

- **Foundation Topics:** These are the core sections of each chapter. They explain the concepts for the topics in that chapter.
- **Exam Preparation Tasks:** After the "Foundation Topics" section of each chapter, the "Exam Preparation Tasks" section lists a series of study activities that you should do at the end of the chapter:
  - **Review All Key Topics:** The Key Topic icon appears next to the most important items in the "Foundation Topics" section of the chapter. The "Review All Key Topics" activity lists the key topics from the chapter,

along with their page numbers. Although the contents of the entire chapter could be on the exam, you should definitely know the information listed in each key topic, so you should review these.

- **Define Key Terms:** Although the Solutions Architect exam may be unlikely to ask a question such as "Define this term," the exam does require that you learn and know a lot of new terminology. This section lists the most important terms from the chapter, asking you to write a short definition and compare your answer to the glossary at the end of the book.
- **Review Questions:** Confirm that you understand the content that you just covered by answering these questions and reading the answer explanations.

- **Web-based Practice Exam:** The companion website includes the Pearson Cert Practice test engine that allows you to take practice exam questions. Use it to prepare with a sample exam and to pinpoint topics where you need more study.

Figure Credits

Chapter 1, Figures 1-1 through 1-7, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 2, Figures 2-1 through 2-4, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 3, Figures 3-1 through 3-3, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 4, Figure 4-2, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 5, Figures 5-1 and 5-2, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 6, quote from Amazon in the note courtesy of Amazon Web Services, Inc.

Chapter 6, Figure 6-1, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 7, Figures 7-1 and 7-2, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 8, Figures 8-1 through 8-4, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 9, Figures 9-1 and 9-2, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 10, Figures 10-1 through 10-4, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 11, reference to NIST 800-88 courtesy of NIST 800-88 "Guidelines for Media Sanitization," Recommendations of the National Institute of Standards and Technology, Richard Kissel September, 2006.

Chapter 12, Figures 12-1 through 12-3, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 13, Figures 13-1 through 13-3, screenshot of AWS © 2018, Amazon Web Services, Inc.

Chapter 14, Figures 14-1 through 14-3, screenshot of AWS © 2018, Amazon Web Services, Inc.

Designing for Elasticity

We discussed elasticity briefly when we presented the various advantages that cloud technologies provide IT solutions in Chapter 1, "The Fundamentals of AWS." Remember, *elasticity* refers to the capability of your resources to scale up and down as well as in and out when resource requirements change due to changes in demand.

"Do I Know This Already?" Quiz

The "Do I Know This Already?" quiz allows you to assess whether you should read the entire chapter. Table 9-1 lists the major headings in this chapter and the "Do I Know This Already?" quiz questions covering the material in those headings so you can assess your knowledge of these specific areas. The answers to the "Do I Know This Already?" quiz appear in Appendix A.

Table 9-1 "Do I Know This Already?" Foundation Topics Section-to-Question Mapping

Foundation Topics Section	Question
Elastic Load Balancing	1–2
Auto Scaling	3–4

CAUTION The goal of self-assessment is to gauge your mastery of the topics in this chapter. If you do not know the answer to a question or are only partially sure of the answer, you should mark that question as wrong for purposes of the self-assessment. Giving yourself credit for an answer you correctly guess skews your self-assessment results and might provide you with a false sense of security.

1. Which is not a common service used with ELB?
   a. EC2
   b. CloudWatch
   c. EBS
   d. Route 53

2. Which of the following is not a valid type of Elastic Load Balancer in AWS?
   a. Classic
   b. Service-based
   c. Network
   d. Application

3. How many Auto Scaling policies can you set against an AWS resource?
   a. 1
   b. 10
   c. 50
   d. 100

4. What is the cooldown period used with Auto Scaling?
   a. It is the mandatory time that a service must run before sampling can begin.
   b. It is the mandatory sample time of metrics for policy-based Auto Scaling.
   c. It is the mandatory time that Auto Scaling must wait after a service reboots before scale-out operations.
   d. It is the mandatory time that Auto Scaling must wait before it takes additional scaling actions.

Foundation Topics

Elastic Load Balancing

Elastic Load Balancing (ELB) distributes incoming application or network traffic across multiple targets. These targets are AWS resources such as Amazon EC2 instances, containers, and IP addresses. To foster high availability, ensure the resources are in multiple Availability Zones.

Elastic Load Balancing scales your load balancer as traffic to your application changes over time and can scale to a majority of workloads automatically. You can add and remove compute resources from your load balancer as your needs change, without disrupting the overall flow of requests to your applications.

You can configure health checks, which are used to monitor the health of the computing resources so that the load balancer can send requests only to the healthy ones. You can also offload the work of SSL/TLS encryption and decryption to your load balancer so that your compute resources can focus on their main work.

Elastic Load Balancing supports three types of load balancers: Application Load Balancers, Network Load Balancers, and Classic Load Balancers. You can select a load balancer based on your application needs. Figure 9-1 shows the configuration of a Network Load Balancer in AWS.

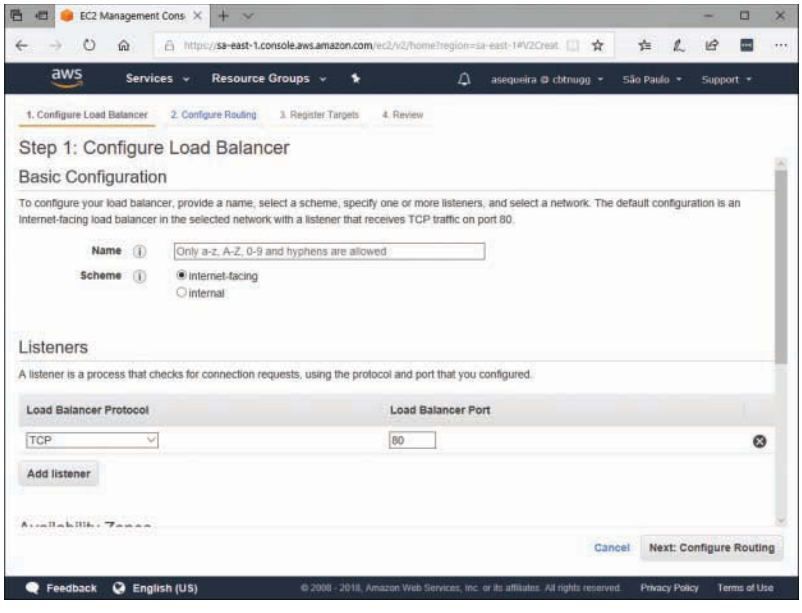

FIGURE 9-1 Configuring a Network Load Balancer in AWS

Although you can monitor and configure your Elastic Load Balancer with the traditional options of AWS (such as the Management Console or SDKs), you can also use the Query API. This interface provides low-level API actions that you call using HTTPS requests.

Elastic Load Balancing works with the following services to improve the availability and scalability of your applications:

Key Topic

- **EC2:** These virtual servers run your applications in the cloud. You can configure your load balancer to route traffic to your EC2 instances.
- **ECS:** This service enables you to run, stop, and manage Docker containers on a cluster of EC2 instances. You can configure your load balancer to route traffic to your containers.
- **Auto Scaling:** This service ensures that you are running your desired number of instances, even if an instance fails, and enables you to automatically increase or decrease the number of instances as the demand on your instances changes. If you enable Auto Scaling with Elastic Load Balancing, instances that are launched by Auto Scaling are automatically registered with the load balancer, and instances that are terminated by Auto Scaling are automatically deregistered from the load balancer.
- **CloudWatch:** This service enables you to monitor your load balancer and take action as needed.
- **Route 53:** This service provides a reliable and cost-effective way to route visitors to websites by translating domain names (such as www.ajsnetworking.com) into the numeric IP addresses (such as 69.163.163.123) that computers use to connect to each other. AWS assigns URLs to your resources, such as load balancers. However, you might want a URL that is easy for users to remember. For example, you can map your domain name to a load balancer.

Auto Scaling

Auto Scaling enables you to quickly discover the scalable AWS resources that are part of your application and configure dynamic scaling in a matter of minutes. The Auto Scaling console provides a single user interface to use the automatic scaling features of multiple AWS services. It also offers recommendations to configure scaling for the scalable resources in your application.

Use Auto Scaling to automatically scale the following resources that support your application:

Key Topic

- EC2 Auto Scaling groups
- Aurora DB clusters

- DynamoDB global secondary indexes
- DynamoDB tables
- ECS services
- Spot Fleet requests

With Auto Scaling, you create a scaling plan with a set of instructions used to configure dynamic scaling for the scalable resources in your application. Auto Scaling creates target tracking scaling policies for the scalable resources in your scaling plan. Target tracking scaling policies adjust the capacity of your scalable resource as required to maintain resource utilization at the target value that you specified.

You can create one scaling plan per application source (an AWS CloudFormation stack or a set of tags). You can add each scalable resource to one scaling plan. If you have already configured scaling policies for a scalable resource in your application, Auto Scaling keeps the existing scaling policies instead of creating additional scaling policies for the resource.

Auto Scaling involves the creation of a Launch Configuration and an Auto Scaling group. Figure 9-2 shows the configuration of Auto Scaling in AWS.

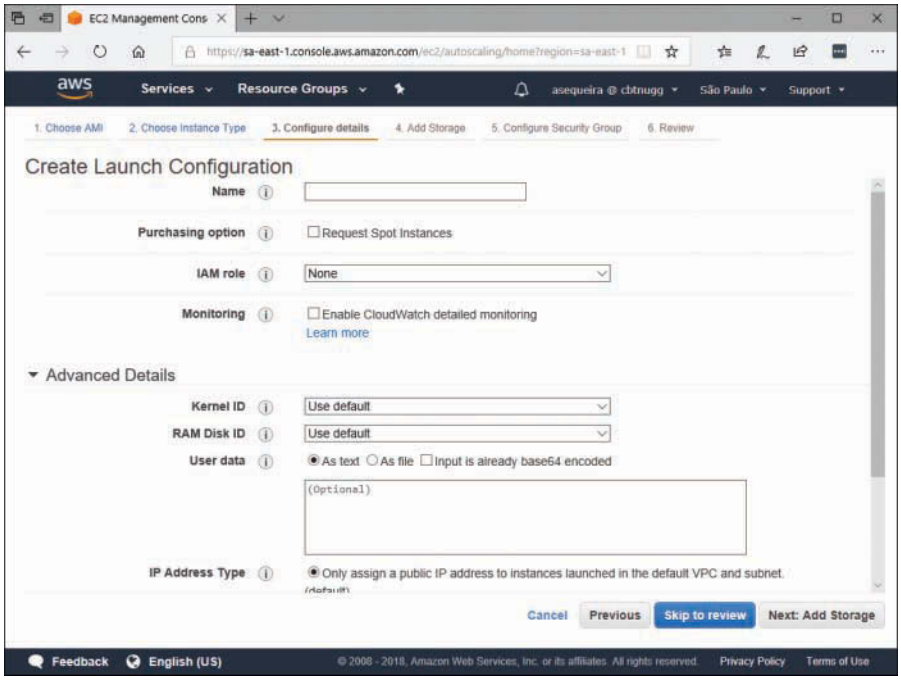

FIGURE 9-2 Configuring Auto Scaling in AWS

Target Tracking Scaling Policies

With target tracking scaling policies, you select a predefined metric or configure a customized metric and set a target value. Application Auto Scaling creates and manages the CloudWatch alarms that trigger the scaling policy and calculates the scaling adjustment based on the metric and the target value. The scaling policy adds or removes capacity as required to keep the metric at, or close to, the specified target value. In addition to keeping the metric close to the target value, a target tracking scaling policy also adjusts to changes in the metric due to a changing load pattern and minimizes changes to the capacity of the scalable target.

When specifying a customized metric, be aware that not all metrics work for target tracking. The metric must be a valid utilization metric and describe how busy a scalable target is. The metric value must increase or decrease proportionally to the capacity of the scalable target so that the metric data can be used to proportionally scale the scalable target.

You can have multiple target tracking scaling policies for a scalable target, provided that each of them uses a different metric. Application Auto Scaling scales based on the policy that provides the largest capacity for both scale-in and scale-out. This provides greater flexibility to cover multiple scenarios and ensures that there is always enough capacity to process your application workloads.

NOTE When discussing scaling the resources of a service, we are scaling those resources horizontally (out and in with elasticity), while the service made up of those resources is being scaled up and down (vertically because the single service is getting bigger or smaller). A single service scales both up and down and out and in, depending on the context.

You can also optionally disable the scale-in portion of a target tracking scaling policy. This feature provides the flexibility to use a different method for scale-in than you use for scale-out.

Keep the following in mind for Auto Scaling:

- You cannot create target tracking scaling policies for Amazon EMR clusters or AppStream 2.0 fleets.
- You can create 50 scaling policies per scalable target. This includes both step scaling policies and target tracking policies.
- A target tracking scaling policy assumes that it should perform scale-out when the specified metric is above the target value. You cannot use a target tracking scaling policy to scale out when the specified metric is below the target value.

- A target tracking scaling policy does not perform scaling when the specified metric has insufficient data. It does not perform scale-in because it does not interpret insufficient data as low utilization. To scale in when a metric has insufficient data, create a step scaling policy and have an alarm invoke the scaling policy when it changes to the INSUFFICIENT_DATA state.
- You may see gaps between the target value and the actual metric data points. The reason is that Application Auto Scaling always acts conservatively by rounding up or down when it determines how much capacity to add or remove. This prevents it from adding insufficient capacity or removing too much capacity. However, for a scalable target with small capacity, the actual metric data points might seem far from the target value. For a scalable target with larger capacity, adding or removing capacity causes less of a gap between the target value and the actual metric data points.
- We recommend that you scale based on metrics with a 1-minute frequency because that ensures a faster response to utilization changes. Scaling on metrics with a 5-minute frequency can result in slower response time and scaling on stale metric data.
- To ensure application availability, Application Auto Scaling scales out proportionally to the metric as fast as it can but scales in more gradually.
- Do not edit or delete the CloudWatch alarms that Application Auto Scaling manages for a target tracking scaling policy. Application Auto Scaling deletes the alarms automatically when you delete the Auto Scaling policy.

The Cooldown Period

The scale-*out* cooldown period is the amount of time, in seconds, after a scale-out activity completes before another scale-out activity can start. While this cooldown period is in effect, the capacity that has been added by the previous scale-out event that initiated the cooldown is calculated as part of the desired capacity for the next scale-out event. The intention is to continuously scale out.

The scale-*in* cooldown period is the amount of time, in seconds, after a scale-in activity completes before another scale-in activity can start. This cooldown period is used to block subsequent scale-in events until it has expired. The intention is to scale in conservatively to protect your application's availability. However, if another alarm triggers a scale-out policy during the cooldown period after a scale-in event, Application Auto Scaling scales out your scalable target immediately.

Exam Preparation Tasks

As mentioned in the section "How to Use This Book" in the Introduction, you have a couple of choices for exam preparation: the exercises here, Chapter 16, "Final Preparation," and the exam simulation questions in the Pearson Test Prep practice test software.

Review All Key Topics

Review the most important topics in this chapter, noted with the Key Topics icon in the margin of the page. Table 9-2 lists a reference of these key topics and the page numbers on which each is found.

Key Topic

Table 9-2 Key Topics for Chapter 9

Key Topic Element	Description	Page Number
List	Resources you can use with Elastic Load Balancing	160
List	Resources you can use with Auto Scaling	160

Complete Tables and Lists from Memory

There are no memory tables in this chapter.

Define Key Terms

Define the following key term from this chapter and check your answer in the glossary:

Cooldown period

Q&A

The answers to these questions appear in Appendix A. For more practice with exam format questions, use the Pearson Test Prep practice test software.

1. What are the three types of load balancers in AWS?
2. Do you need to delete the individual CloudWatch alarms that you use for Auto Scaling when you delete the Auto Scaling policy ?

Index

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