Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Poll question

Which of the following best describes your familiarity with serverless architectures?

A. I have built solutions using serverless architectures.

B. I understand serverless architectures, but have not used them.

C. I know a little bit about serverless architectures.

D. I am not familiar with serverless architectures.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Module overview

- Business request
- What is serverless?
- Amazon API Gateway
- Amazon Simple Queue Service (Amazon SQS)
- Amazon Simple Notification Service (Amazon SNS)
- Amazon Kinesis
- AWS Step Functions
- Present solutions
- Knowledge check
- Lab 5: Build a Serverless Architecture

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Imagine you are meeting with an application development manager who is preparing to build on AWS. The manager is interested in the benefits of serverless architectures.

At the end of this module, you meet with the application development manager and present some solutions.

DO NOT COPY
javierartiga.espublico@gmail.com

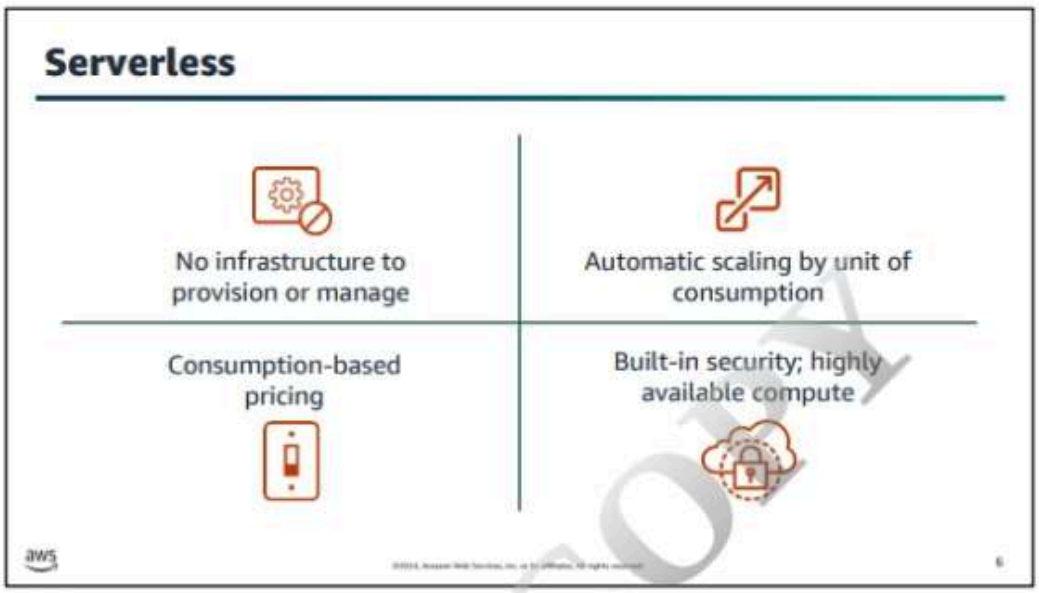
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The application development manager asks, "How can we reduce operational overhead and optimize our resource costs?" The company is looking for ways to shift resources towards developing new features and services. They are concerned about the additional infrastructure and cost that this could create. The company is asking you to research different ways to build in the cloud.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Serverless is a way to describe the services, practices, and strategies that you can use to build more agile applications. With serverless computing, you can innovate and respond to change faster. AWS handles infrastructure management tasks, such as capacity provisioning and patching, so you can focus on writing code that serves your customers.

Serverless computing provides the following advantages:

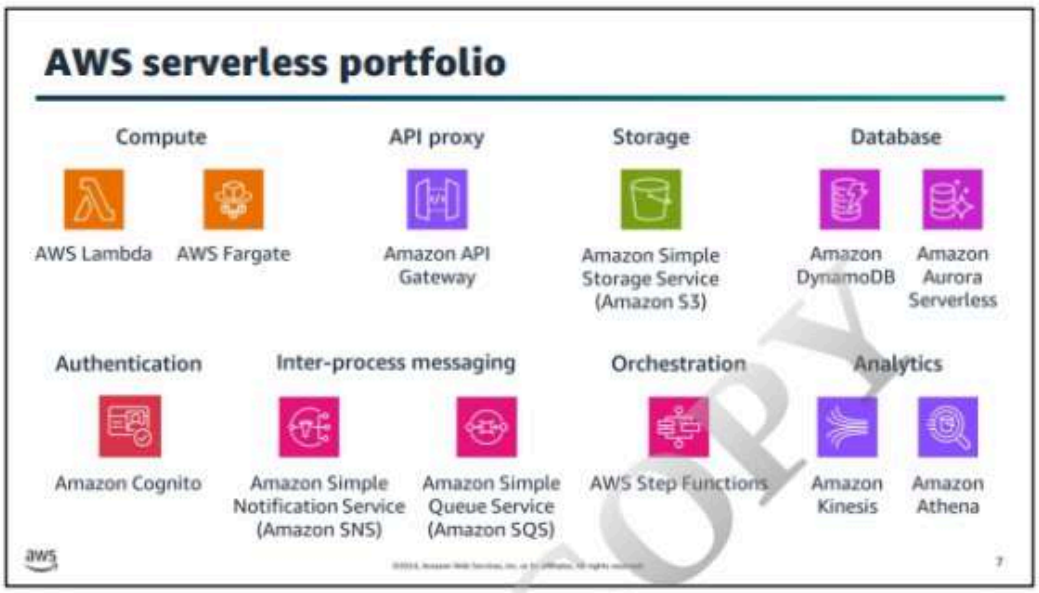
- No infrastructure to provision or manage
- No servers to provision, operate, or patch
- Automatic scaling by unit of consumption, rather than by server unit
- Pay-for-value billing model (pay for the unit, rather than by server unit)
- Built-in availability and fault tolerance
- No need to architect for availability because it is built into the service

For more information, see Serverless on AWS at https://aws.amazon.com/serverless/.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS provides a set of fully managed services that you can use to build and run serverless applications. Serverless applications don't require you to provision, maintain, and administer servers for backend components. Examples of components include compute, databases, storage, stream processing, messaging, and queueing. You also no longer need to worry about ensuring application fault tolerance and availability. AWS provides all these capabilities for you so you can focus on product innovation while enjoying faster time to market.

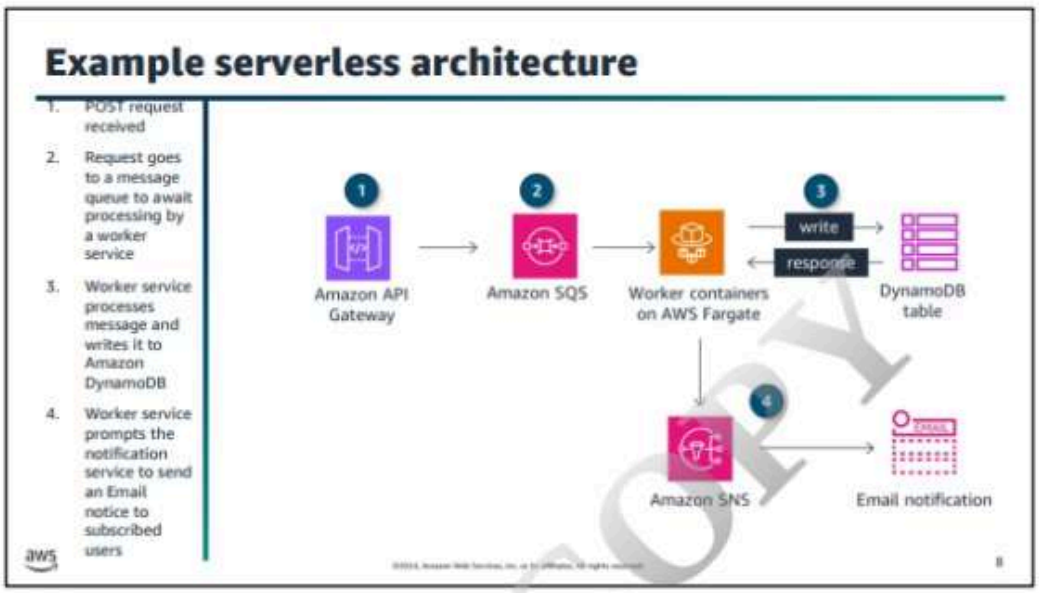
You learned about compute, storage, and database services in previous modules. In this module, you learn about the following:

- API Gateway
- Amazon SQS
- Amazon SNS
- Kinesis
- Step Functions

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



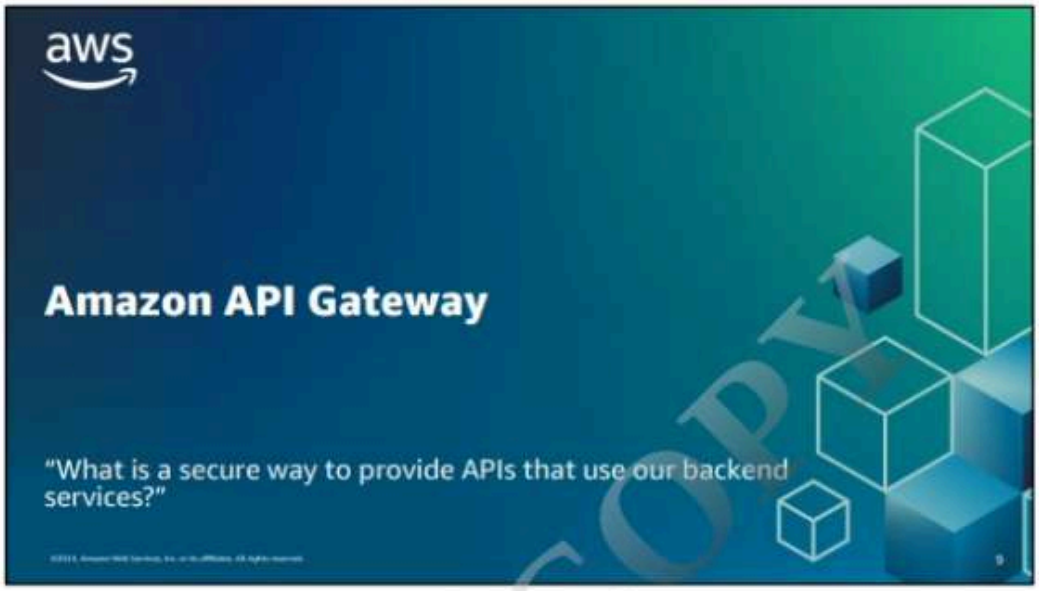
This example of a serverless architecture demonstrates a messaging application.

This example uses Amazon API Gateway to handle all of the tasks involved in accepting and processing API calls. API Gateway feeds messages to an application messaging queue built in Amazon SQS, where the messages are processed by a worker service. This queue provides a buffer for the worker service in the event of traffic spikes. The containerized worker service uses serverless hosting in AWS Fargate, which automatically scales hosting capacity with each newly launched container. The worker service writes the processed message to an Amazon DynamoDB table. Once the worker receives a success response, the worker notifies Amazon SNS, which pushes an Email notification to all subscribed users.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

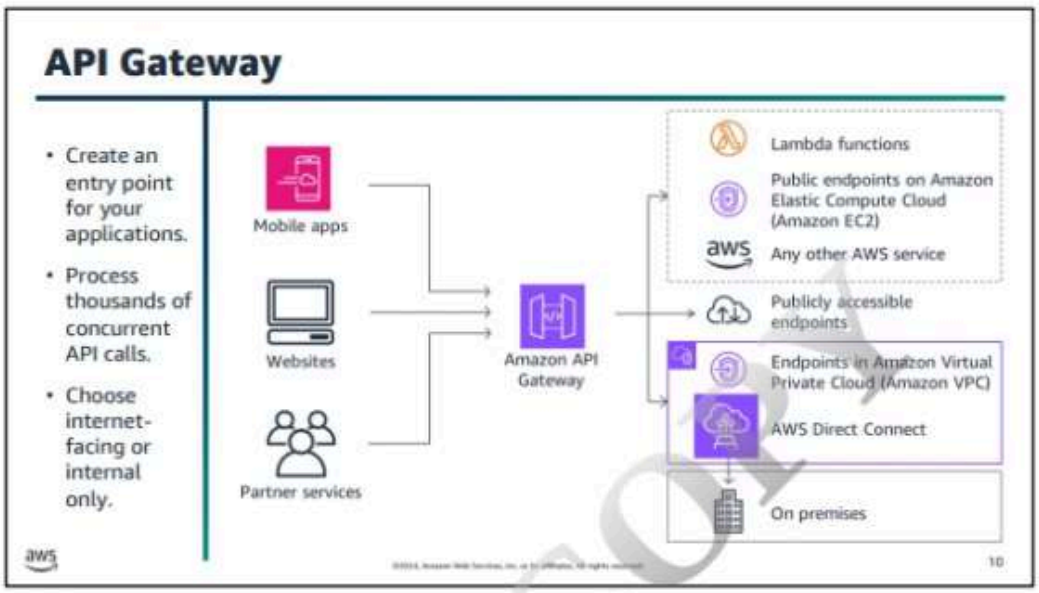


The application development manager asks, "What is a secure way to provide APIs that use our backend services?" The company has some services that they would like to expose for third-party developers using a REST API. They have concerns about how to do this in a secure way while also managing scale, throttling, and monitoring. They are asking you to recommend a service that minimizes operational overhead and can protect against security threats.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



With Amazon API Gateway, you can create, publish, maintain, monitor, and secure APIs. You can use API Gateway to connect your applications to AWS services and other public or private websites. It provides consistent RESTful and HTTP APIs for mobile and web applications to access AWS services and other resources hosted outside of AWS.

As a gateway, API Gateway handles all of the tasks involved in accepting and processing up to hundreds of thousands of concurrent API calls. These tasks include traffic management, authorization and access control, monitoring, and API version management.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

API Gateway features

- Creates a unified API frontend for multiple microservices
- Provides distributed denial of service (DDoS) protection and throttling for your backend
- Authenticates and authorizes requests to a backend
- Throttles, meters, and monetizes API usage by third-party developers

11

With API Gateway, you can do the following:

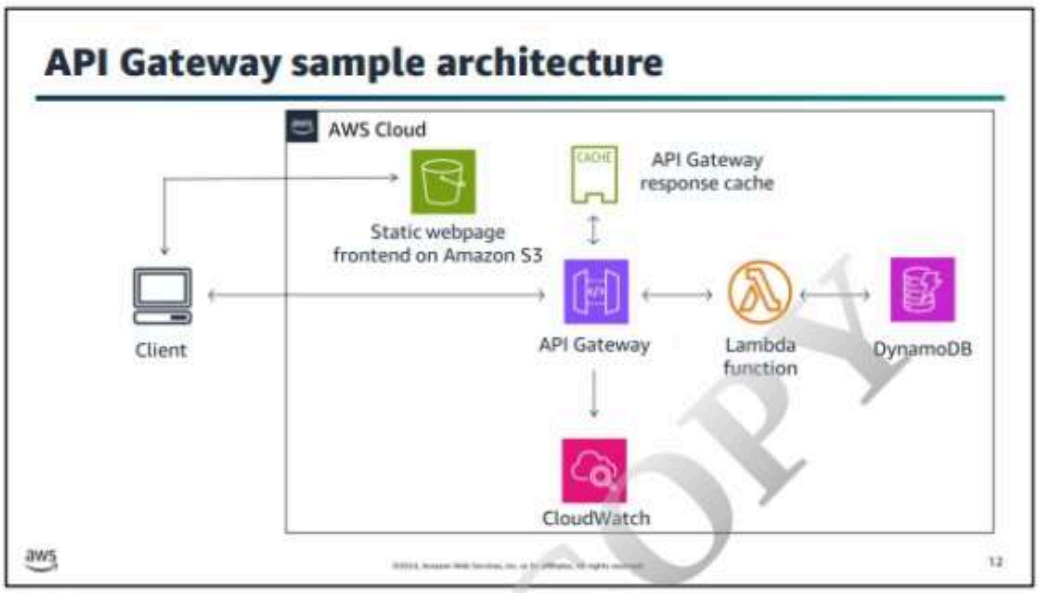
- Host and use multiple versions and stages of your APIs.
- Create and distribute API keys to developers.
- Use Signature Version 4 (SigV4) to authorize access to APIs.
- Use RESTful or WebSocket APIs.

For more information, see "AWS Signature Version 4 for API requests" in the *AWS General Reference* at https://docs.aws.amazon.com/IAM/latest/UserGuide/reference_sigv.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



In this example, the client browser requests a static webpage hosted in Amazon S3. Using this webpage, the client browser communicates with API Gateway using a REST API. API Gateway authenticates and authorizes the request, and invokes a Lambda function that communicates with DynamoDB. This example includes the optional API Gateway cache to reduce backend load and reduce latency when serving recurring requests.

API Gateway also sends logs to Amazon CloudWatch. API Gateway can send logs to CloudWatch for each stage in your API or for each method. You can set the verbosity of the logging (Error or Info), and whether full request and response data should be logged.

The detailed metrics that API Gateway can send to CloudWatch are the following:

- Number of API calls
- Latency
- Integration latency
- HTTP 400 and 500 errors

You can also activate access logging to log who has accessed your API and how they accessed it.

DO NOT COPY javierartiga.espublico@gmail.com

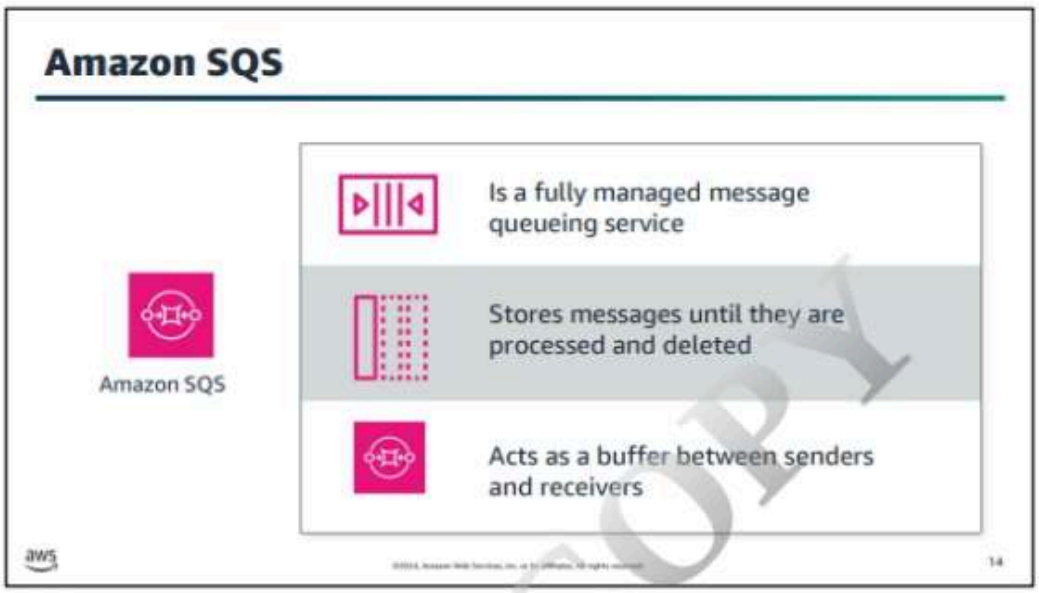
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The application development manager asks, "How do we create a message queue for reliable service-to-service communication?" The company is converting some of their monolithic applications into microservices. The development team has decided that these services can use asynchronous communication to communicate with one another. The company is asking you to recommend a solution for building a message queue.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



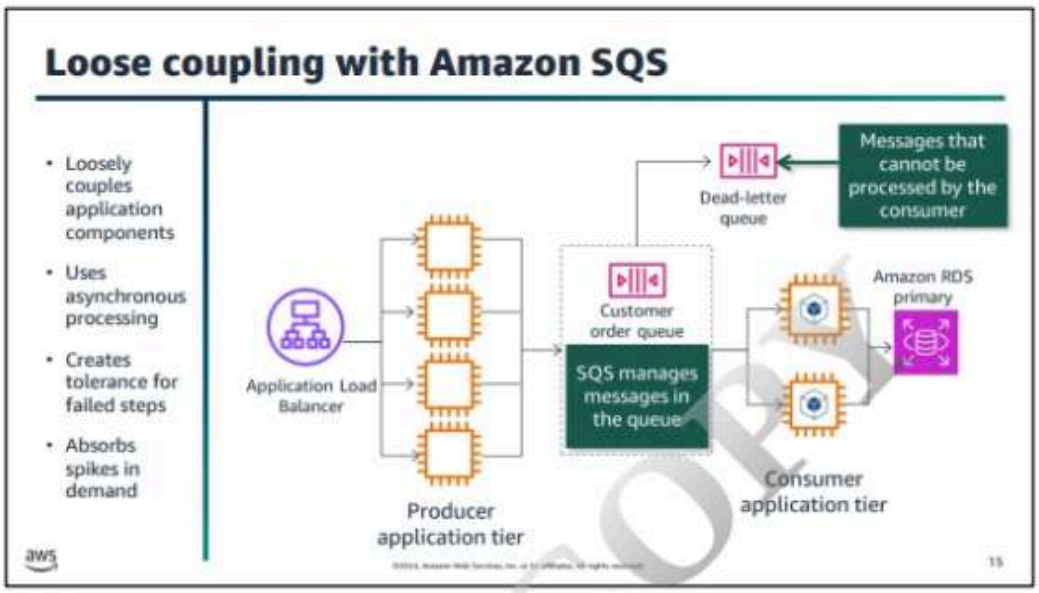
Amazon SQS is a fully managed service that requires no administrative overhead and little configuration. The service works on a massive scale, processing billions of messages per day. Amazon SQS stores all message queues and messages within a single, highly available AWS Region with multiple redundant Availability Zones. This prevents a single computer, network, or Availability Zone failure from making messages inaccessible. You can send and read messages simultaneously.

Developers can securely share SQS queues anonymously or with specific AWS accounts. You can also restrict queue sharing by IP address and time of day. Streaming Single Instruction Multiple Data (SIMD) Extensions (SSE) protects the contents of messages in SQS queues using keys managed in AWS Key Management Service (AWS KMS). SIMD Extensions encrypts messages as soon as Amazon SQS receives them. The messages are stored in encrypted form, and Amazon SQS decrypts messages only when they are sent to an authorized consumer.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



In this example, a producer application creates customer orders and sends them to an Amazon SQS queue. A consumer application processes orders from the producer application tier. The consumer application polls the queue and receives messages. It then records the messages in an Amazon Relational Database Service (Amazon RDS) database and deletes the processed messages from the SQS queue. Amazon SQS sends messages that cannot be processed to a dead-letter queue where they can be processed later.

Using an SQS queue provides the following benefits:

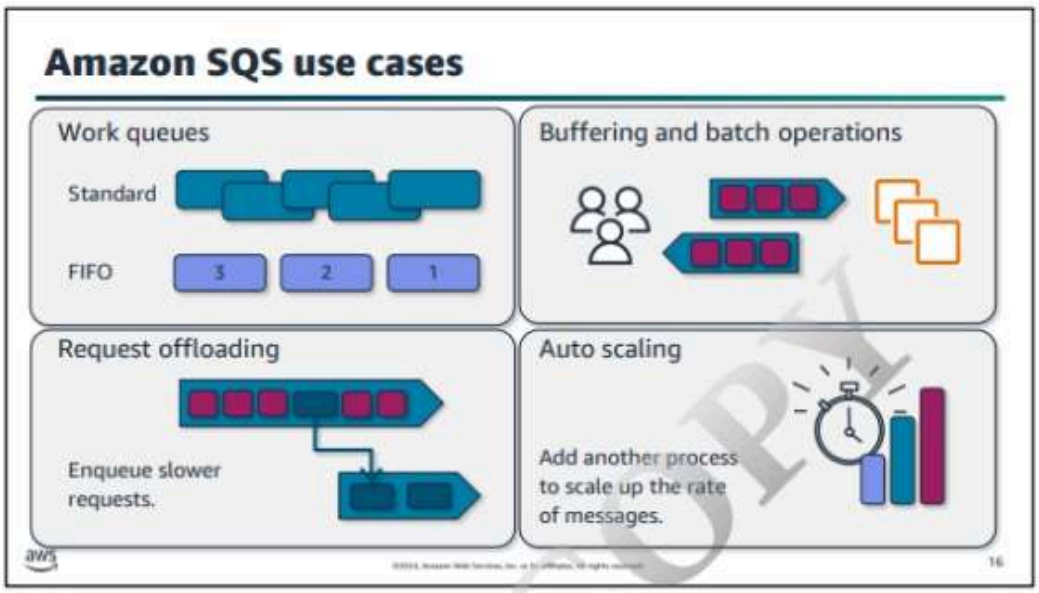
- **Loose coupling** – With Amazon SQS, you can decouple preprocessing steps from compute steps and postprocessing steps. Using asynchronous processing isolates the producer logic into its own component separate from the consumer logic.
- **Failure tolerance** – In the event of an application exception or transaction failure, the order processing can be retried. Once the maximum number of retries is reached, SQS can redirect the message to an Amazon SQS dead-letter queue where you can reprocess or debug it later. The loss of one node or job in a loosely coupled workload usually doesn't delay the entire calculation.
- **Absorbs spikes** – An Amazon SQS queue makes the system more resilient. The queue acts as a buffer to absorb spikes in traffic. This gives your application additional time to complete scale-out actions. It is also cost effective because you don't need to provision as much idle compute to absorb spikes.

For more information about this Amazon SQS use case, see "Building Loosely Coupled, Scalable, C# Applications with Amazon SQS and Amazon SNS" in the *AWS Compute Blog* at https://aws.amazon.com/blogs/compute/building-loosely-coupled-scalable-c-applications-with-amazon-sqs-and-amazon-sns/.

For more information about loosely coupled architectures, see "Loosely Coupled Scenarios" in the *High Performance Computing Lens: AWS Well-Architected Framework* guide at https://docs.aws.amazon.com/wellarchitected/latest/high-performance-computing-lens/loosely-coupled-scenarios.html.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



There are many different ways to use SQS queues.

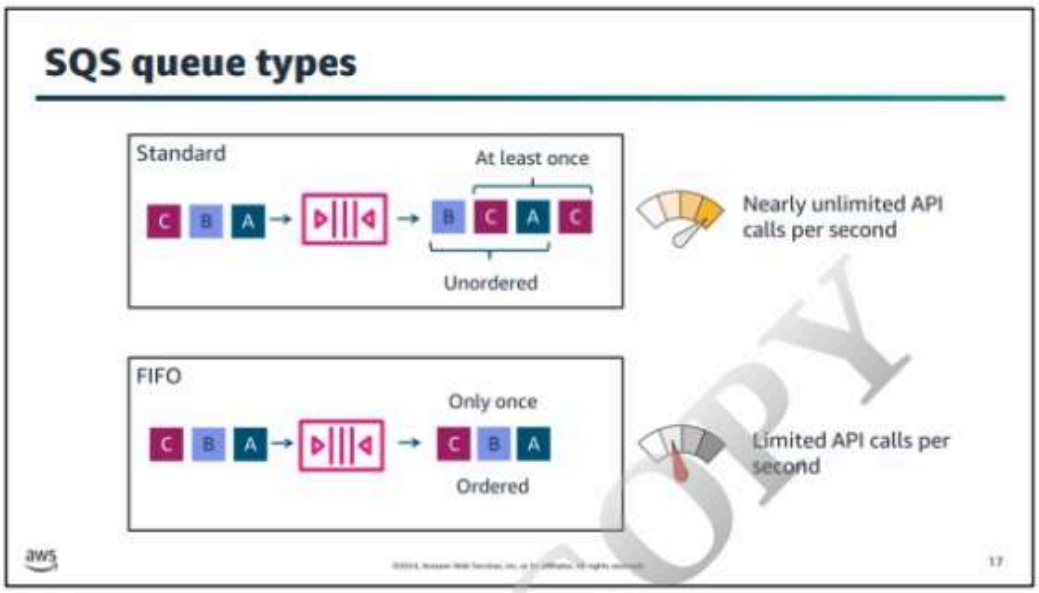
Review the following use cases:

- **Work queues** – Decouple components of a distributed application that might not all process the same amount of work simultaneously. You can choose a standard queue or a first-in-first-out (FIFO) queue depending on the requirements of your application.
- **Buffering and batch operations** – Add scalability and reliability to your architecture and smooth out temporary volume spikes without losing messages or increasing latency.
- **Request offloading** – Move slow operations off of interactive request paths by enqueueing the request.
- **Auto scaling instances** – Use SQS queues to help determine the load on an application. When combined with auto scaling, you can scale the number of Amazon EC2 instances in or out, depending on the volume of traffic.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon SQS offers two types of message queues.

Standard queues support at-least-once message delivery and provide best-effort ordering. Messages are generally delivered in the same order in which they are sent. However, because of the highly distributed architecture, more than one copy of a message might be delivered out of order. Standard queues can handle a nearly unlimited number of API calls per second. You can use standard message queues if your application can process messages that arrive more than once and out of order.

FIFO queues are designed to enhance messaging between applications when the order of operations and events is critical or where duplicates can't be tolerated. FIFO queues also provide exactly-once processing, but have a limited number of API calls per second.

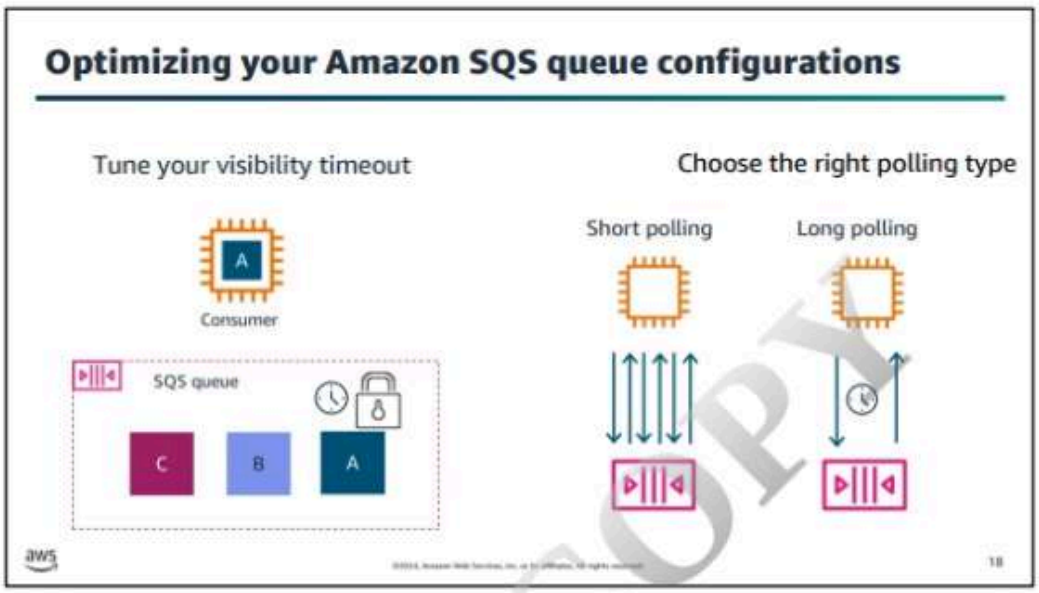
For more information about Amazon SQS standard queues, see "Amazon SQS standard queues" in the *Amazon Simple Queue Service Developer Guide* at http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/standard-queues.html.

For more information about Amazon SQS FIFO, see "Amazon SQS FIFO queues" in the *Amazon Simple Queue Service Developer Guide* at https://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-fifo-queues.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



When creating an Amazon SQS queue, you need to consider how your application interacts with the queue. This information will help you optimize the configuration of your queue to control costs and increase performance.

Visibility timeout

When a consumer receives an SQS message, that message remains in the queue until the consumer deletes it. You can configure the SQS queue's *visibility timeout* setting to make that message invisible to other consumers for a period of time. This helps to prevent another consumer from processing the same message. The default visibility timeout is 30 seconds. The consumer deletes the message once the consumer completes processing the message. If the consumer fails to delete the message before the visibility timeout expires, it becomes visible to other consumers and can be processed again.

Typically, you should set the visibility timeout to the maximum time that it takes your application to process and delete a message from the queue. Setting too short of a timeout increases the possibility of your application processing a message twice. Too long of a visibility timeout delays subsequent attempts at processing a message.

Polling type

You can configure an Amazon SQS queue to use either short polling or long polling.

An SQS queue with short polling does the following:

- Sends a response to the consumer immediately after receiving a request, providing a faster response
- Increases the number of responses and therefore costs

An SQS queue with long polling does the following:

- Does not return a response until at least one message arrives or the poll times out
- Provides less frequent responses but decreases costs

Depending on the frequency of messages arriving in your queue, many of the responses from a queue using short polling could be reporting an empty queue. Unless your application requires an immediate response to its poll requests, long polling is the preferable option.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

When to use message queues

✓ Service-to-service communication	✗ Selecting specific messages
✓ Asynchronous work items	✗ Large messages
✓ State change notifications	

19

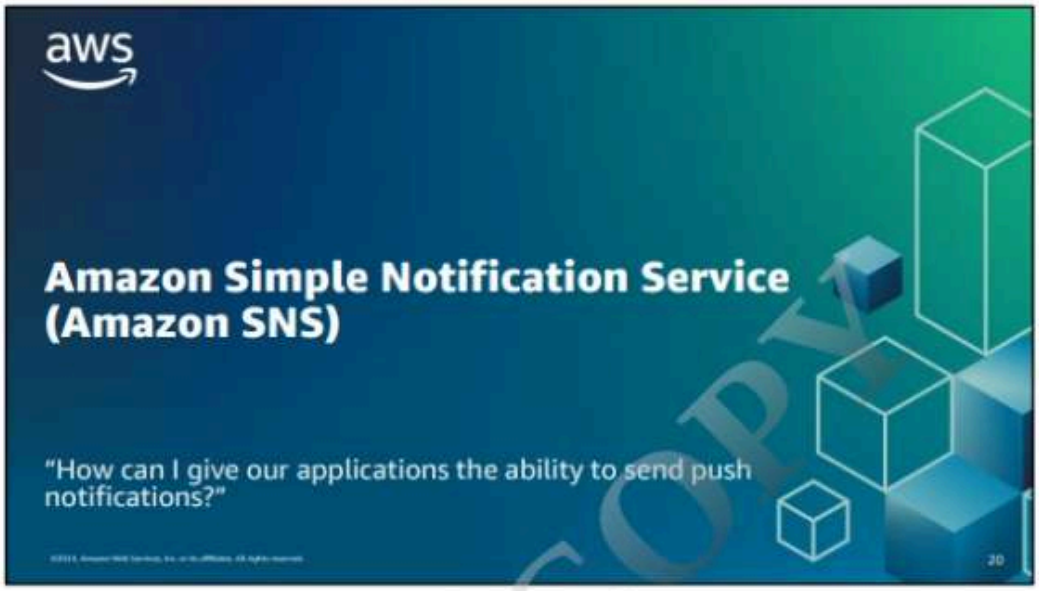
It's important to know when a particular technology won't fit well with your use case. Messaging has its own set of commonly encountered anti-patterns. It's tempting to retrieve messages selectively that match a particular set of attributes, or even match a one-time logical query. For example, a service requests a message with a particular attribute because it contains a response to another message that the service sent out. This can lead to a scenario where there are messages in the queue that no one is polling for and are never consumed.

Most messaging protocols and implementations work best with reasonably sized messages (in the tens or hundreds of kilobytes). As message sizes grow, it's best to use a dedicated storage system, such as Amazon S3, and pass a reference to an object in that store in the message itself.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

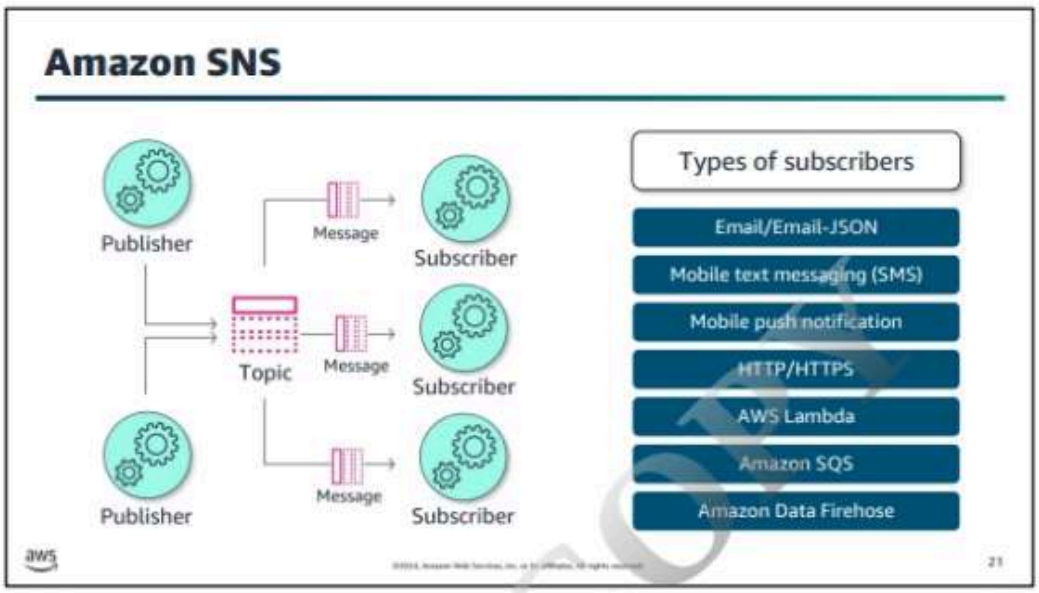


The application development manager asks, "How can we give our applications the ability to send push notifications?" The company would like to notify users when certain events occur, such as application updates or promotional messages. They would like to deliver these messages using email and text messaging. The company is asking you to research how they can most easily do this in the AWS Cloud.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon SNS is a web service that helps you to set up, operate, and send notifications from the cloud. The service follows the *publish-subscribe* (pub-sub) messaging paradigm, with notifications being delivered to clients using a *push* mechanism.

You create a topic and control access to it by defining policies that determine which publishers and subscribers can communicate with the topic. A publisher sends messages to topics they have created, or to topics that they have permission to publish to.

Instead of including a specific destination address in each message, a publisher sends a message to the topic. Amazon SNS matches the topic to a list of subscribers for that topic, and delivers the message to each subscriber.

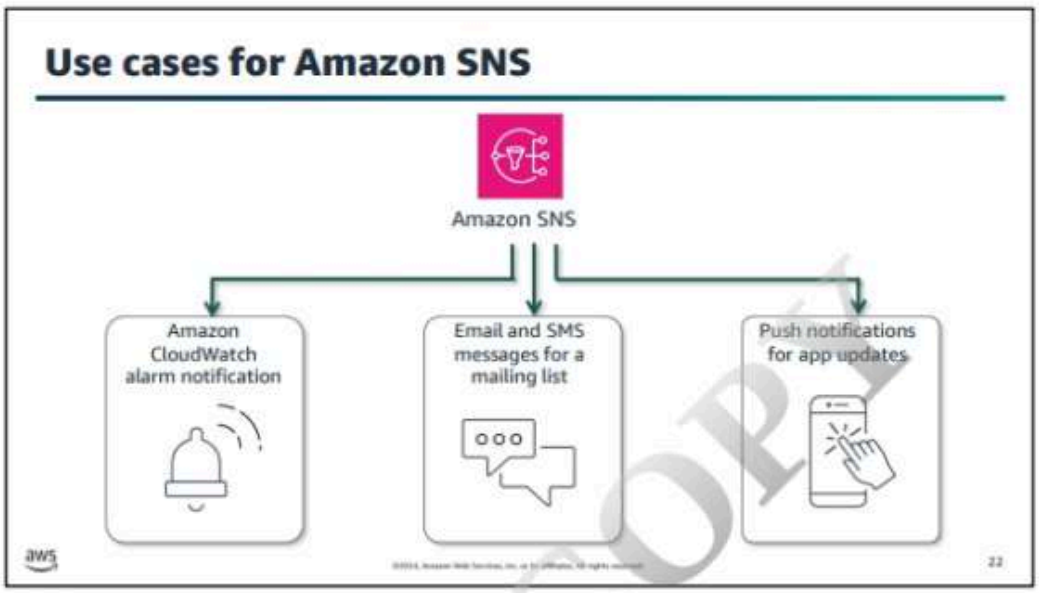
Each topic has a unique name that identifies the Amazon SNS endpoint where publishers post messages, and where subscribers register for notifications. Subscribers receive all messages published to their subscribed topics. All subscribers to a topic receive the same messages.

Amazon SNS supports encrypted topics. When you publish messages to encrypted topics, Amazon SNS uses AWS KMS keys to encrypt your messages.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

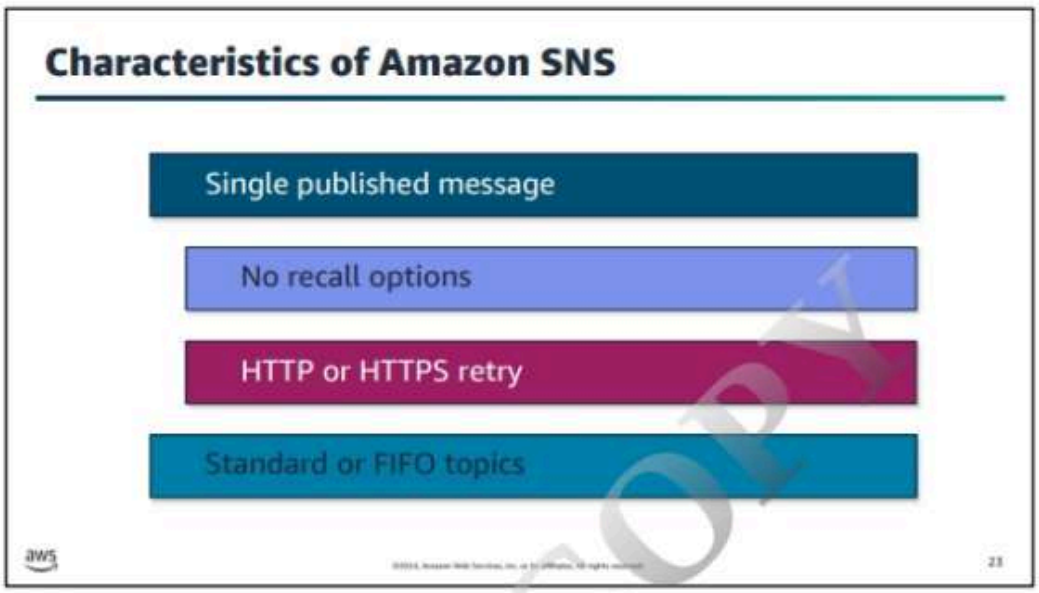


You can use Amazon SNS notifications in many ways—for example, the following:

- You can receive immediate notification when an event occurs, such as a specific change to your Auto Scaling group.
- You can push targeted news headlines to subscribers by email or SMS. Upon receiving the email or SMS text, interested readers can choose to learn more by visiting a website or launching an application.
- You can send notifications to an app, indicating that an update is available. The notification message can include a link to download and install the update.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



All notification messages contain a single published message. Amazon SNS will attempt to deliver messages from the publisher in the order in which they were published into the topic. However, network issues could potentially result in out-of-order messages at the subscriber end. Use Amazon SNS FIFO topics if you require strict message ordering and deduplicated message delivery to one or more subscribers.

Amazon SNS has the following characteristics:

- When a message is delivered successfully, there is no way to recall it.
- You can use an Amazon SNS Delivery Policy to control the retry pattern: linear, geometric, exponential backoff, maximum and minimum retry delays, and other patterns.
- To prevent messages from being lost, all messages are stored redundantly across multiple servers and data centers.
- Amazon SNS is designed to meet the needs of the largest and most demanding applications, allowing applications to publish a large number of messages at any time.
- Amazon SNS allows applications and end users on different devices to receive notifications by Mobile Push notification. This includes Apple, Google, and Kindle Fire Devices; HTTP or HTTPS; email or email-JSON; SMS or Amazon SQS queues; and Lambda functions.
- Amazon SNS provides access control mechanisms to protect topics and messages from unauthorized access. Topic owners can set policies for a topic that restrict who can publish or subscribe to a topic. Topic owners can encrypt notifications by specifying that the delivery mechanism must be HTTPS.

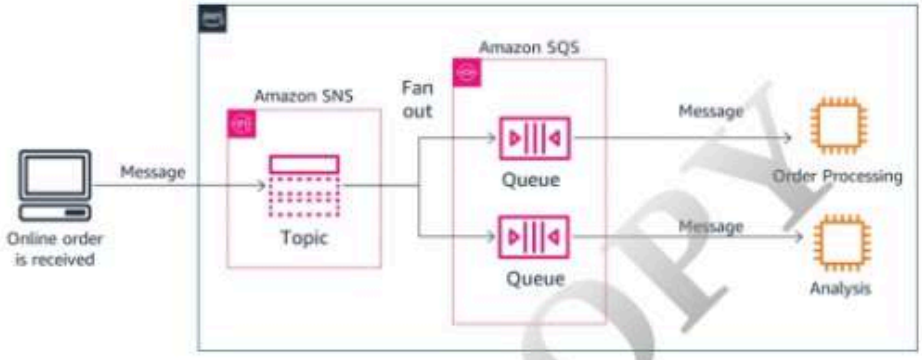
NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SNS publish to multiple SQS queues

Architecture example



24

Using highly available services such as Amazon SNS to perform basic message routing is an effective way of distributing messages to microservices. The two main forms of communications between microservices are *request-response* and *observer*.

This example uses the observer type and illustrates a *fan-out* scenario using Amazon SNS and Amazon SQS. In a fan-out scenario, a message is sent to an SNS topic and is then replicated and pushed to multiple SQS queues, HTTP endpoints, or email addresses. This allows for parallel asynchronous processing.

In this example, Amazon SNS fans out orders to two different SQS queues. Two Amazon EC2 instances each observe a queue. One of the instances handles the processing or fulfillment of the order, while the other is attached to a data warehouse for analysis of all orders received.

To deliver Amazon SNS notifications to an SQS queue, you subscribe to a topic, and specify Amazon SQS as the transport and a valid SQS queue as the endpoint. To permit the SQS queue to receive notifications from Amazon SNS, the SQS queue owner must subscribe the SQS queue to the topic for Amazon SNS. If the user owns the Amazon SNS topic being subscribed to and the SQS queue receiving the notifications, nothing else is required. Any message published to the topic will automatically be delivered to the specified SQS queue. If the owner of the SQS queue is not the owner of the topic, Amazon SNS requires an explicit confirmation to the subscription request.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Amazon SNS and Amazon SQS

Features	Amazon SNS	Amazon SQS
Message persistence	No	Yes
Delivery mechanism	Push (passive)	Poll (active)
Producer and consumer	Publisher and subscriber	Send or receive
Distribution model	One to many	One to one

25

Amazon SNS messages are not persistent. Amazon SNS defines a delivery policy for each delivery protocol. The delivery policy defines how Amazon SNS retries the delivery of messages when server-side errors occur. When the delivery policy is exhausted, Amazon SNS stops retrying the delivery and discards the message unless a dead-letter queue is attached to the subscription.

Amazon SNS permits applications to send time-critical messages to multiple subscribers through a push mechanism.

Amazon SQS exchanges messages through a polling model. Sending and receiving components are decoupled.

Amazon SQS provides flexibility for distributed components of applications to send and receive messages without requiring each component to be concurrently available.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

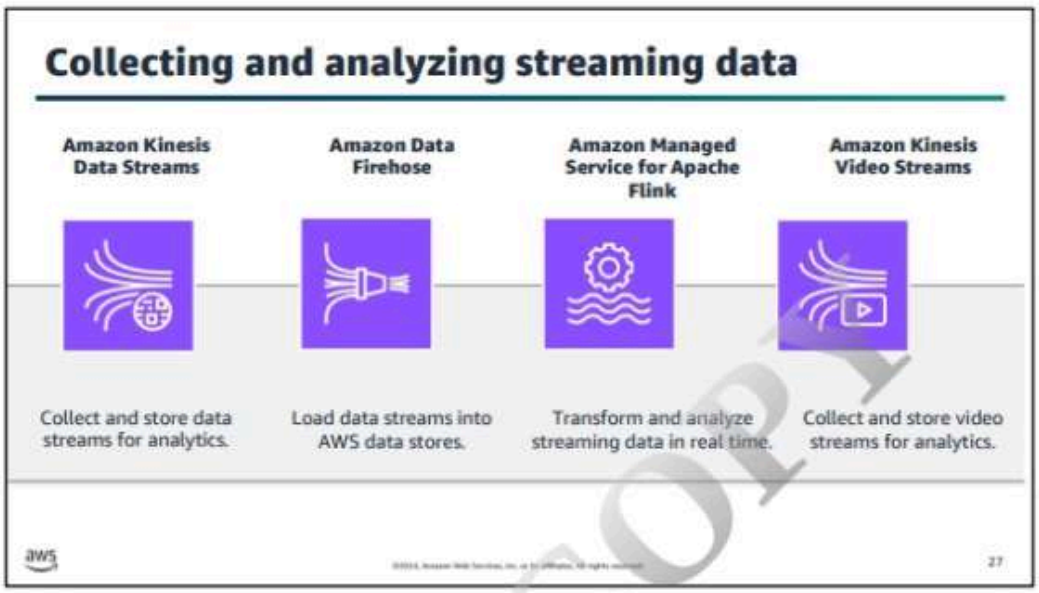
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The application development manager asks, "How do we ingest streaming data to power our real-time applications?" The company would like to capture clickstream data from their web applications and perform real-time analytics. The company is asking you identify a solution for ingesting streaming data.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



With Amazon Kinesis, you can do the following:

- Collect, process, and analyze data streams in real time. Kinesis has the capacity to process streaming data at any scale. You have the flexibility to choose the tools that best suit the requirements of your application in a cost-effective way.
- Ingest real-time data, such as video, audio, application logs, website clickstreams, and Internet of Things (IoT) telemetry data. The ingested data can be used for machine learning, analytics, and other applications.
- Query data streams in real time, and build and run stream processing applications using standard SQL, Python, and Scala. You don't have to wait until all data is collected before the processing begins.

This section describes Amazon Kinesis Data Streams and Amazon Data Firehose.

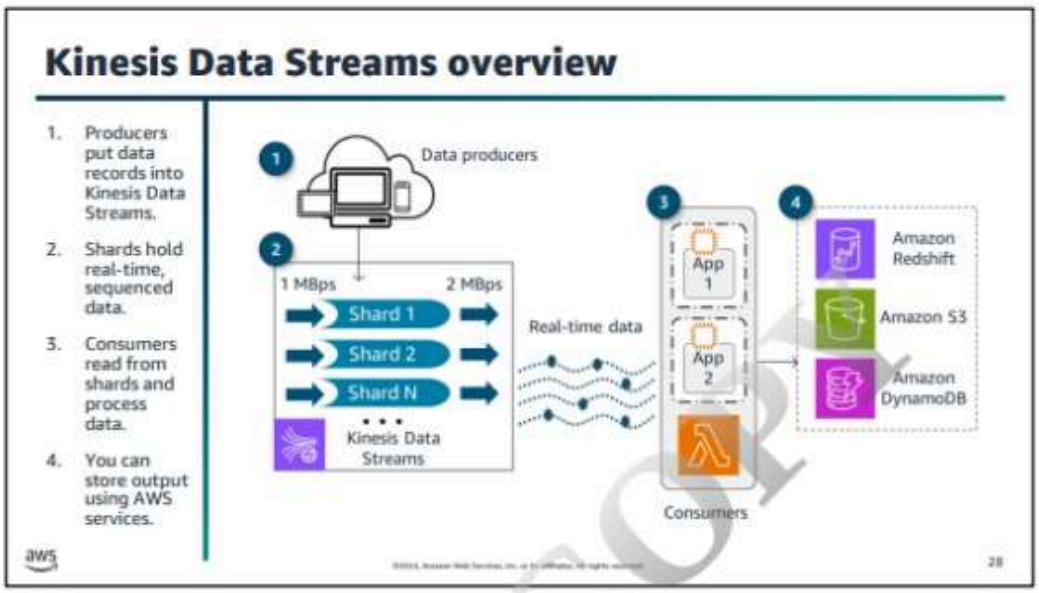
For more information, see Amazon Managed Service for Apache Flink at https://aws.amazon.com/managed-service-apache-flink/.

For more information about Amazon Kinesis Video Streams, see Amazon Kinesis at https://aws.amazon.com/kinesis/.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



To get started using Kinesis Data Streams, you create a stream and specify the number of shards. Each shard is a uniquely identified sequence of data records in a stream. Your stream can receive 1 megabyte per second (MBps) per shard. Each shard has a read limit of 2 MBps for your applications. The total capacity of a stream is the sum of the capacities of its shards. Use resharding to increase or decrease the number of shards in your stream as needed.

Producers write data into the stream. A producer might be an EC2 instance, a mobile client, an on-premises server, or an IoT device. You can send data such as infrastructure logs, application logs, market data feeds, and web clickstream data.

Consumers read the streaming data that the producers generate. A consumer might be an application running on an EC2 instance or AWS Lambda. An application on an EC2 instance will need to scale as the amount of streaming data increases. If this is the case, run it in an Auto Scaling group. Another way to write a consumer application is to use a Lambda function, which you can use to run code without having to provision or manage servers. More than one application can process the same data in the stream.

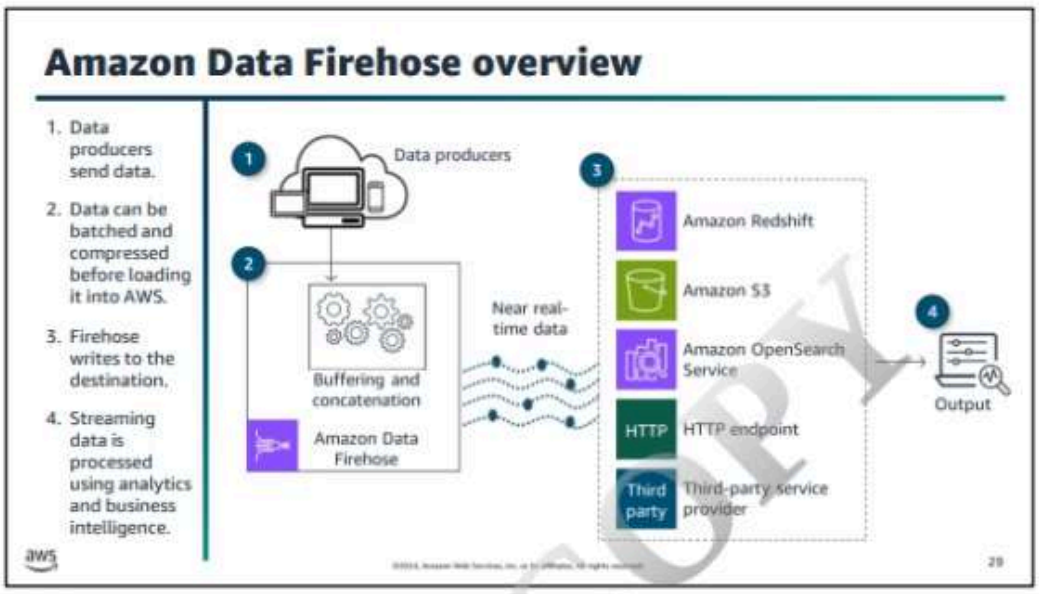
You can store the results of the consumer applications using AWS services such as Amazon S3, Amazon DynamoDB, and Amazon Redshift.

For more information about what you can do with Kinesis Data Streams and the benefits, see "What is Amazon Kinesis Data Streams?" in the *Amazon Kinesis Data Streams Developer Guide* at https://docs.aws.amazon.com/streams/latest/dev/introduction.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon Data Firehose starts to process data in near real time. Firehose can send records to Amazon S3, Amazon Redshift, Amazon OpenSearch Service, and any HTTP endpoint that you own. It can also send records to any of your third-party service providers, including Datadog, New Relic, and Splunk.

For data delivery to Amazon S3, Firehose concatenates multiple incoming records based on the buffering configuration of your delivery stream. It then delivers the concatenated record to Amazon S3 as an S3 object.

To deliver data to Amazon Redshift, Firehose does the following:

1. Delivers incoming data to your S3 bucket in the format described earlier
2. Issues an Amazon Redshift COPY command to load the data from your S3 bucket to your Amazon Redshift cluster

Amazon OpenSearch Service is a fully managed service that delivers OpenSearch APIs and real-time analytics capabilities. With the processed data, you can produce near real-time analytics using your existing business intelligence tools and dashboards. With Amazon OpenSearch Service, Firehose buffers incoming records based on the buffering configuration of your delivery stream. Then, Firehose generates a bulk request to index multiple records to your Amazon OpenSearch Service cluster.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

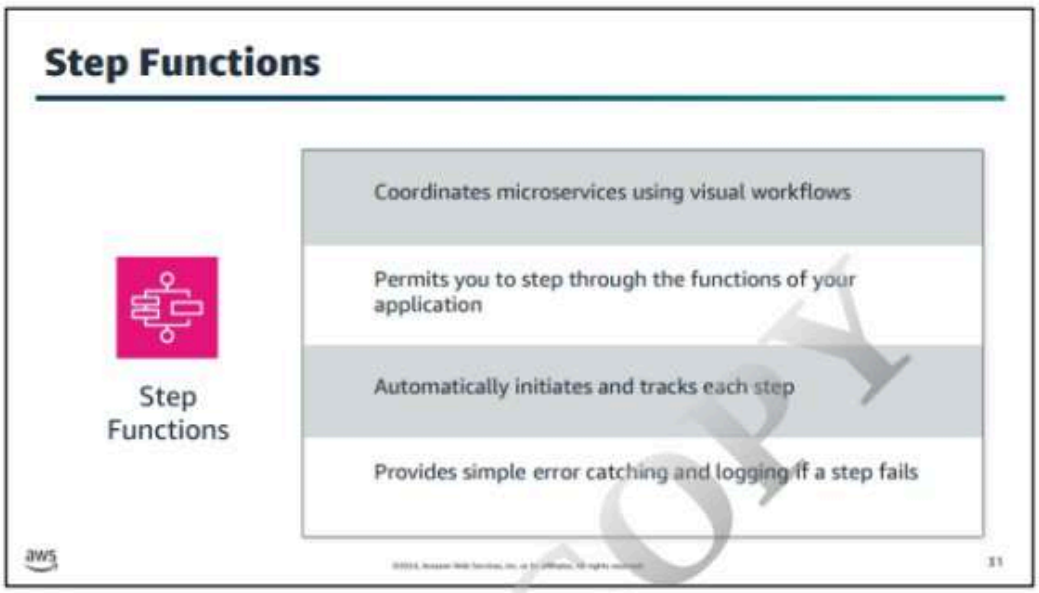
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The application development manager asks, "What is an easy way to orchestrate multi-step workflows?" The company is rearchitecting some of its monolithic applications and decoupling their logic into individual functions and services. They are looking for ways to coordinate workflows through these newly independent components. The company is asking you to identify a tool that facilitates workflows between functions and services.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



It's common for modern cloud applications to be composed of many services and components. As applications grow, you must write an increasing amount of code to coordinate the interaction of all components. With Step Functions, you can focus on defining the component interactions rather than writing all of the software to make the interactions work.

Step Functions integrates with the following AWS services. You can directly call API actions in Step Functions and pass parameters to the APIs of these services.

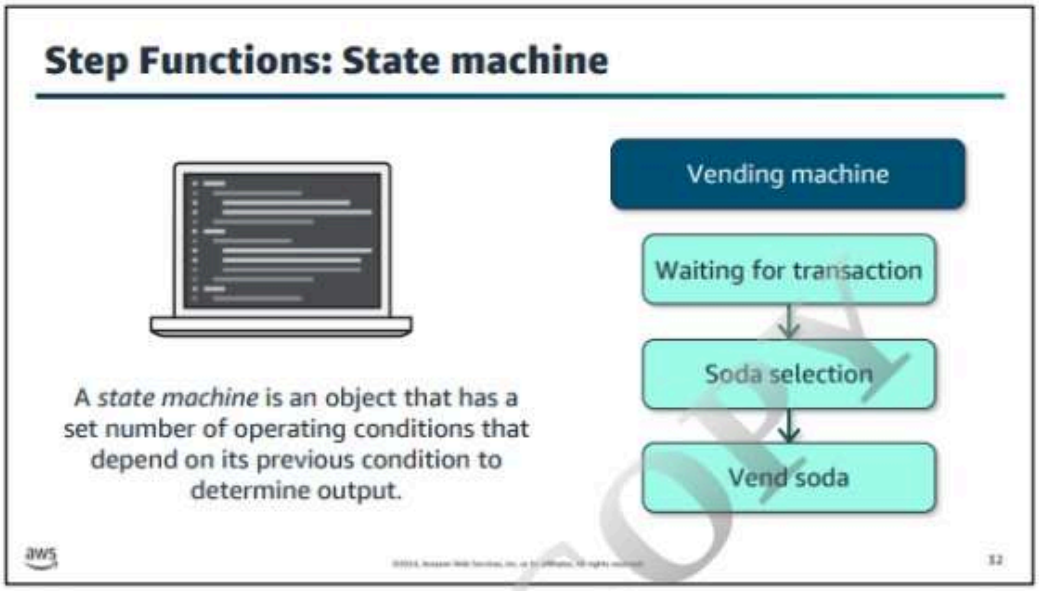
- Compute services: AWS Lambda, Amazon Elastic Container Service (Amazon ECS), Amazon Elastic Kubernetes Service (Amazon EKS), and AWS Fargate
- Database services: Amazon DynamoDB
- Messaging services: Amazon SNS and Amazon SQS
- Data processing and analytics services: Amazon Athena, AWS Batch, AWS Glue, Amazon EMR, and AWS Glue DataBrew
- Machine learning services: Amazon SageMaker AI
- APIs created by API Gateway

You can configure your Step Functions workflow to call other AWS services using Step Functions service tasks. For more information about AWS service integrations, see "Integrating services with Step Functions" in the *AWS Step Functions Developer Guide* at https://docs.aws.amazon.com/step-functions/latest/dg/integrate-services.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



A *state machine* is an object that has a set number of operating conditions that depend on its previous condition to determine output.

A common example of a state machine is the soda vending machine. The machine starts in the operating state (waiting for a transaction), and then moves to soda selection when someone adds money. After that, the machine enters a vending state, where it deploys the soda to the customer. After completion, the state returns to operating.

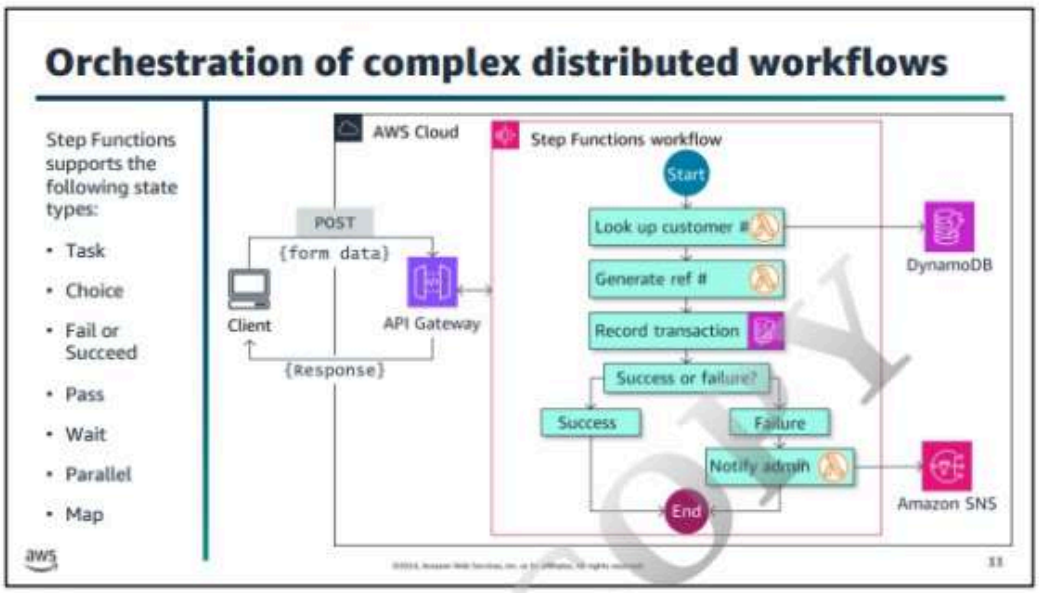
With Step Functions, you can create and automate your own state machines within the AWS environment. Step Functions use a JSON-based Amazon States Language, which contains a structure made of various states, tasks, choices, error handling, and more.

For more information about states, see "Discovering workflow states to use in Step Functions" in the *AWS Step Functions Developer Guide* at https://docs.aws.amazon.com/step-functions/latest/dg/workflow-states.html.

For more information about Amazon States Language, see "Amazon States Language" at https://states-language.net/spec.html.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



States are elements in your state machine. States can perform a variety of functions in your state machine. A state can do the following:

- Do some work in your state machine (a Task state).
- Make a choice between different branches to run (a Choice state).
- Stop with a failure or success (a Fail or Succeed state).
- Pass its input to its output or inject some fixed data (a Pass state).
- Provide a delay for a certain amount of time or until a specified time or date (a Wait state).
- Begin parallel branches (a Parallel state).
- Dynamically iterate steps (a Map state).

With Step Functions you can create two types of workflows, Standard and Express.

- You use the *Standard workflow* type for long-running, durable, and auditable workflows. These workflows can run for up to a year and you can access the full history of workflow activity for up to 90 days after a workflow completes. Standard workflows use an exactly-once model, where your tasks and states are never run more than once unless you specify a Retry behavior.
- You use the *Express workflow* type for high-volume, event-processing workloads such as IoT data ingestion, streaming data processing and transformation, and mobile application backends. They can run for up to five minutes. Express workflows use an at-least-once model, where there is a possibility that an execution might be run more than once.

In this example, a client completes a transaction, which sends a POST request to API Gateway. This event in API Gateway initiates a Step Functions workflow to record the transaction in DynamoDB. The workflow moves through each state until it completes the workflow and ends.

For more information about Step Functions workflows, see "Choosing workflow type in Step Functions" in the *AWS Step Functions Developer Guide* at https://docs.aws.amazon.com/step-functions/latest/dg/choosing-workflow-type.html.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Imagine you are now ready to talk to the application development manager and present solutions that meet their architectural needs. Think about how you would answer the questions from the beginning of the lesson.

Your answers should include the following solutions:

- One way to reduce operational overhead and costs is to adopt serverless architectures. With serverless computing, you do not have to manage or provision infrastructure. It can scale automatically and you can control costs by paying for only the resources you use.
- You can use API Gateway to provide both REST APIs and WebSocket APIs that interact with your backend services. API Gateway can accept and process up to hundreds of thousands of concurrent API calls. It also handles traffic management, authorization and access control, throttling, monitoring, and API version management.
- With Amazon SQS you can send, store, and receive messages between software components. Because Amazon SQS is a managed service, you do not have to manage and operate message-oriented middleware.
- You can use Amazon SNS for application-to-application and application-to-person communication. Amazon SNS supports messaging between distributed systems and event-driven architectures. It can also send messages to users at scale over SMS, mobile push, and email.
- You can use Amazon Kinesis to collect, process, and analyze real-time streaming data.
- With Step Functions, you can build stateful workflows that connect services and systems. Step Functions integrates with many AWS services.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Module review

In this module, you learned about the following:

- ✓ What is serverless?
- ✓ API Gateway
- ✓ Amazon SQS
- ✓ Amazon SNS
- ✓ Amazon Kinesis
- ✓ AWS Step Functions

Next, you review the following:

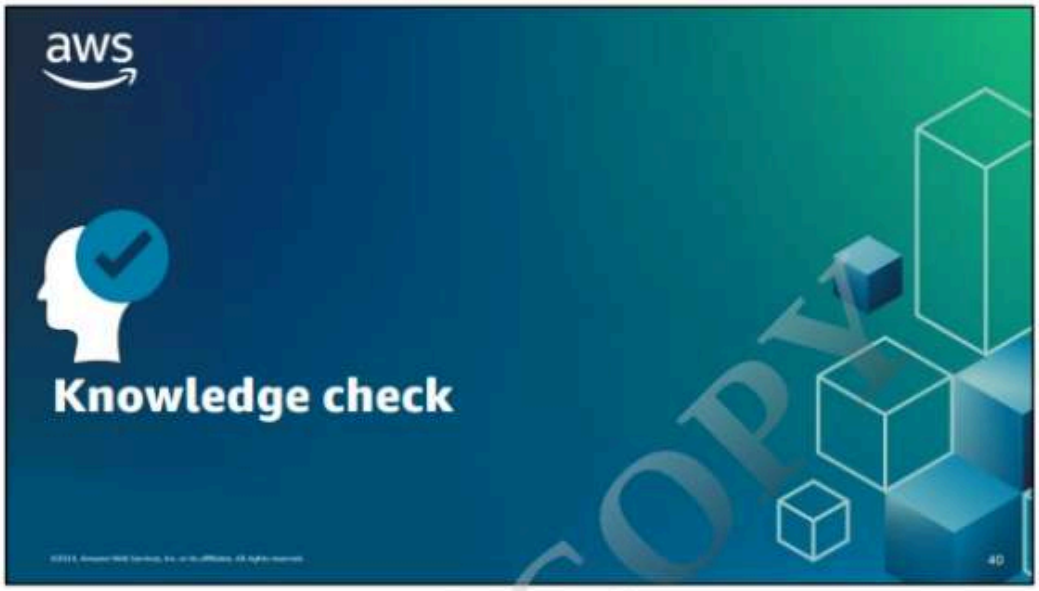
- Knowledge check
- Lab introduction

19

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 1

Which type of Amazon SQS queue provides at-least-once delivery?

Choice	Response
A	FIFO queue
B	Standard queue
C	Dead-letter queue
D	Long polling

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 1 and answer

Which type of Amazon SQS queue provides at-least-once delivery?

Choice	Response
A	FIFO queue
B correct	Standard queue
C	Dead-letter queue
D	Long polling

42

The correct answer is B, Standard queue.

Standard queues support at-least-once message delivery and provide best-effort ordering. Messages are generally delivered in the same order in which they are sent. However, because of the highly distributed architecture, more than one copy of a message might be delivered out of order.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 2

What is an advantage of long polling compared to short polling?

Choice	Response
A	Long polling provides an immediate response from a `ReceiveMessage` call.
B	Long polling is more stable when using a single thread to poll multiple queues.
C	Long polling reduces the cost of using Amazon SQS by reducing the number of empty responses and false empty responses.
D	Long polling reduces cost by only sampling a subset of Amazon SQS servers.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 2 and answer

What is an advantage of long polling compared to short polling?

Choice	Response
A	Long polling provides an immediate response from a `ReceiveMessage` call.
B	Long polling is more stable when using a single thread to poll multiple queues.
C correct	Long polling reduces the cost of using Amazon SQS by reducing the number of empty responses and false empty responses.
D	Long polling reduces cost by only sampling a subset of Amazon SQS servers.

The correct answer is C, Long polling reduces the cost of using Amazon SQS by reducing the number of empty responses and false empty responses.

Long polling doesn't return a response until a message arrives in the queue or the long poll times out. Long polling reduces the cost of using Amazon SQS by reducing the number of empty responses and false empty responses.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 3

What is a feature of Amazon SNS?

Choice	Response
A	Amazon SNS exchanges messages through a polling model.
B	Amazon SNS can send messages to decoupled components of a distributed application that do not process the same amount of work simultaneously.
C	Amazon SNS can push messages to multiple subscribers.
D	Amazon SNS keeps messages persistent.

45

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 3 and answer

What is a feature of Amazon SNS?

Choice	Response
A	Amazon SNS exchanges messages through a polling model.
B	Amazon SNS can send messages to decoupled components of a distributed application that do not process the same amount of work simultaneously.
C correct	Amazon SNS can push messages to multiple subscribers.
D	Amazon SNS keeps messages persistent.

46

The correct answer is C, Amazon SNS can push messages to multiple subscribers.

Amazon SNS can push messages to multiple subscribers.

- For Standard topics, each account can support 100,000 Standard topics, and each topic supports up to 12.5 million subscriptions.
- For FIFO topics, each account can support 1,000 FIFO topics, and each topic supports up to 100 subscriptions.

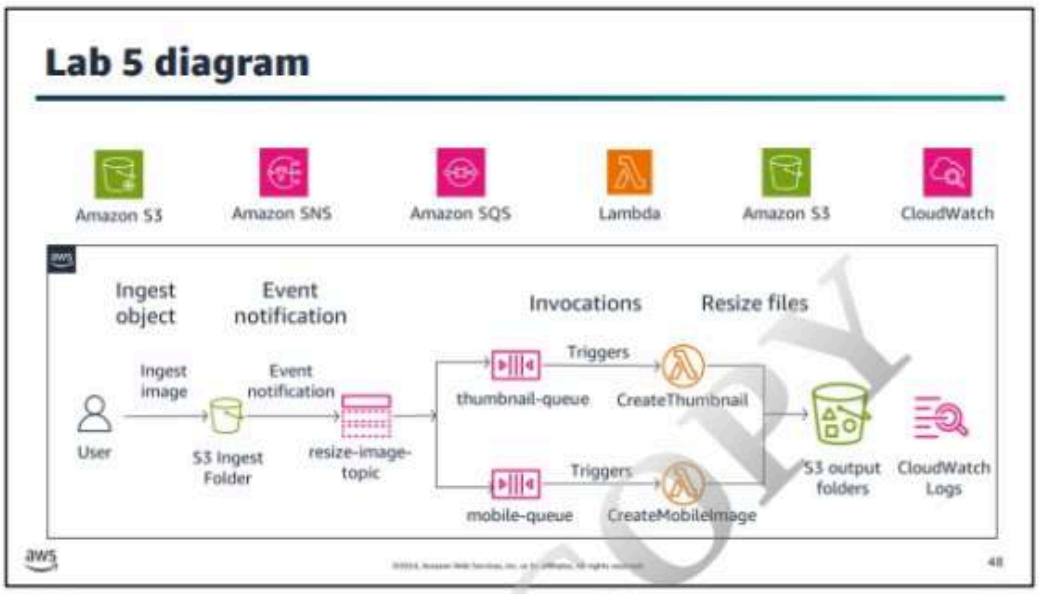
For more information, see Amazon SNS features at https://aws.amazon.com/sns/features/.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



In this lab, you create a serverless architecture. When a user adds an image to an Amazon S3 bucket, the event is published to an SNS topic. Three SQS queues are subscribed to the topic. Each queue has a Lambda function that is invoked when its SQS queue sends a message. Each Lambda function resizes the image and adds it to another S3 bucket. These events are logged in Amazon CloudWatch.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Lab tasks

Task 1: Create a standard Amazon SNS topic.

Task 2: Create two Amazon SQS queues.

Task 3: Create an Amazon S3 event notification.

Task 4: Create and configure two AWS Lambda functions.

Task 5: Upload an object to an Amazon S3 bucket.

Task 6: Validate the processed file.

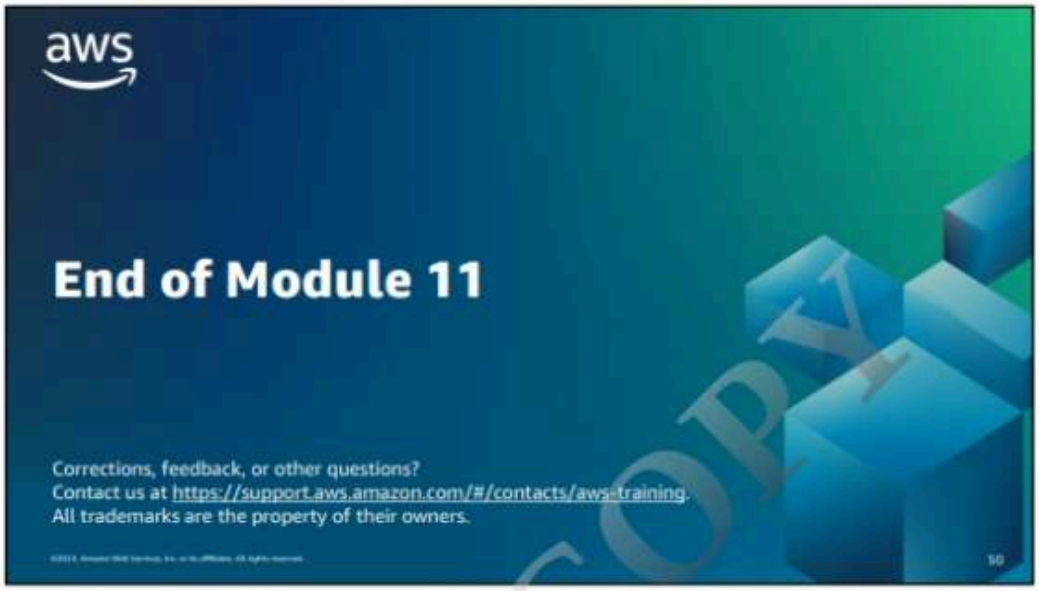
(Optional) Task 1: Create a lifecycle configuration to delete files in the ingest bucket after 30 days.

(Optional) Task 2: Add an SNS email notification to the existing SNS topic.

49

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Poll question

What percentage of your workloads requires ultra-low latency to devices and users?

A. 0%

B. 1–25%

C. 26–50%

D. 51–75%

E. 76–100%

DO NOT COPY
javierartiga.espublico@gmail.com

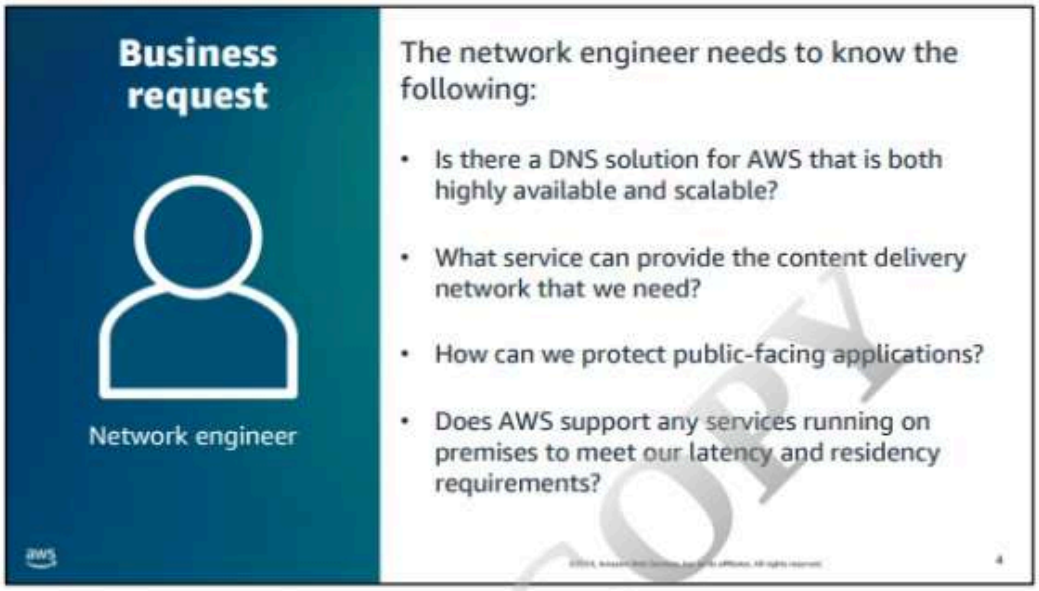
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Module overview

- Business request
- Edge fundamentals
- Amazon Route 53
- Amazon CloudFront
- Distributed denial of service (DDoS) protection
- AWS Outposts
- Present solutions
- Knowledge check
- Lab 6: Configure an Amazon CloudFront Distribution with an Amazon S3 Origin

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



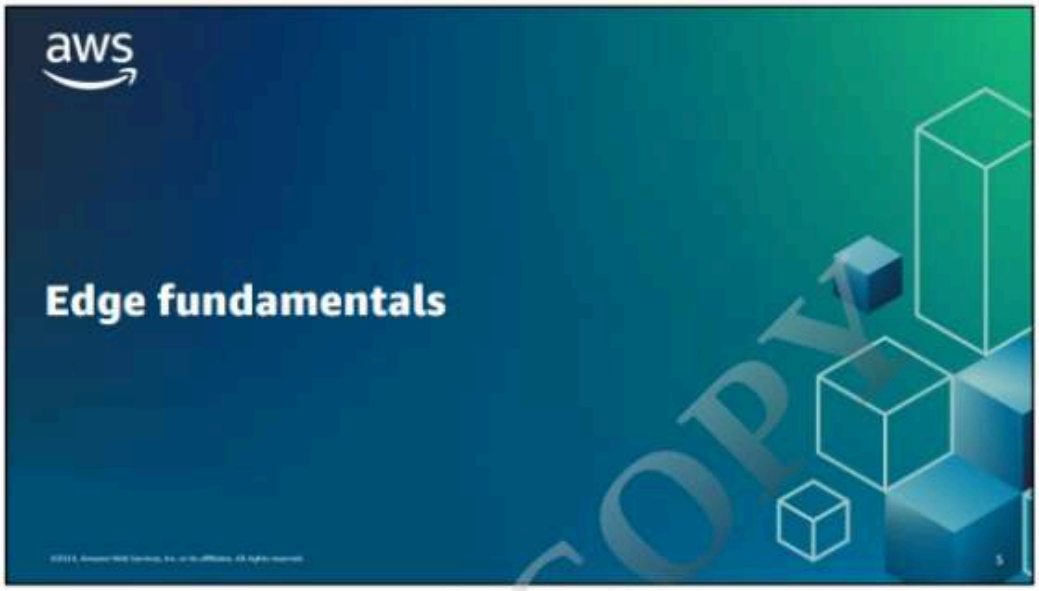
Imagine your network engineer meets with you to discuss AWS services that can improve availability, latency, and the security of your applications. Here are some questions they are asking about edge services.

At the end of this module, you meet with the network engineer and present some solutions.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

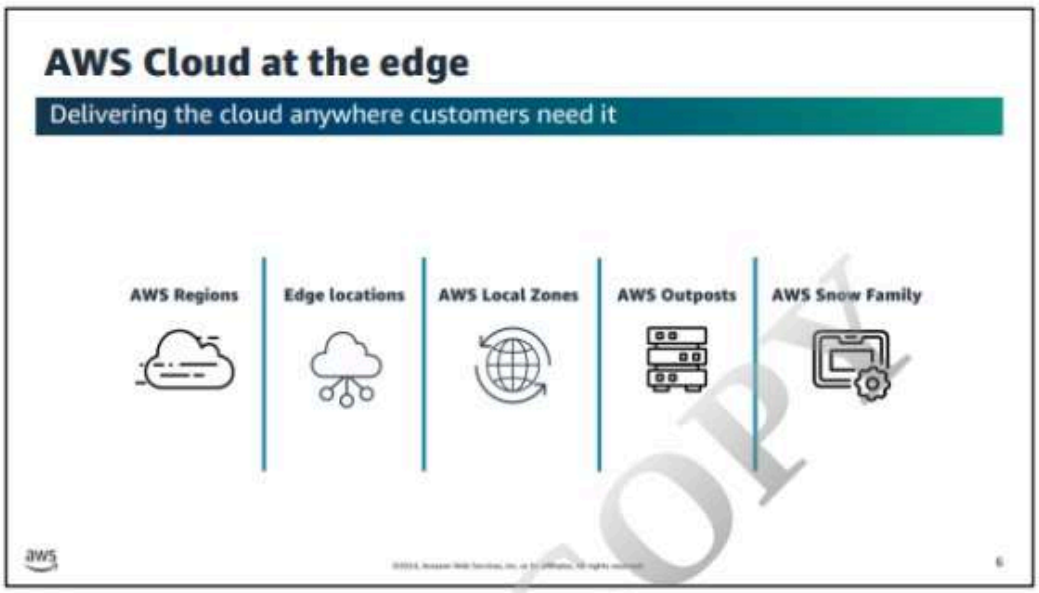


Start by reviewing some of the services that run on the edge of the AWS network.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS edge computing services provide infrastructure and software that move data processing and analysis as close to the endpoint as necessary. This includes deploying AWS managed hardware and software to locations outside AWS data centers, and even onto customer-owned devices.

You can extend the cloud for a consistent hybrid experience using these AWS edge services related to locations:

- **AWS edge locations** – Edge locations are connected to the AWS Regions through the AWS network backbone. Amazon CloudFront, AWS WAF, and AWS Shield are services you use here.
- **AWS Local Zones** – Local Zones are an extension of the AWS Cloud. They are located close to large population and industry centers. You learned about Local Zones in Module 1: Architecting Fundamentals.
- **AWS Outposts** – With AWS Outposts, you can run some AWS services on premises or at your own data center.
- **AWS Snow Family** – The Snow Family of products provides offline storage at the edge, which is used to deliver data back to AWS Regions.

You can use the same infrastructure, services, APIs, and tools for a consistent experience.

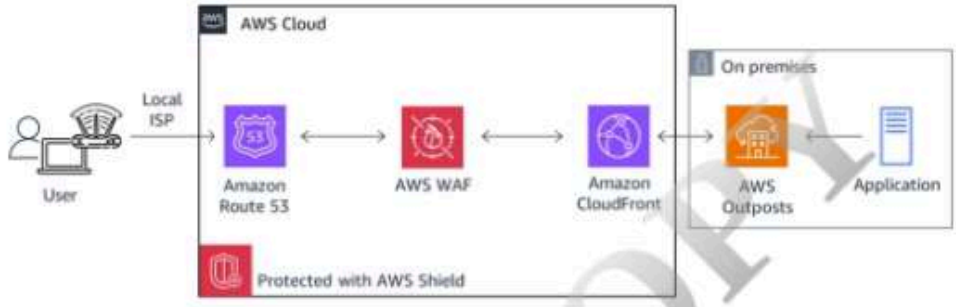
DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Edge services architecture

In this module, you learn about each AWS service involved in this example:



Review the edge services architecture. A user sends a request to an application partly hosted on premises. The user's request interacts with Amazon Route 53, AWS WAF, Amazon CloudFront, and AWS Outposts. The AWS services hosted in the cloud are protected with AWS Shield.

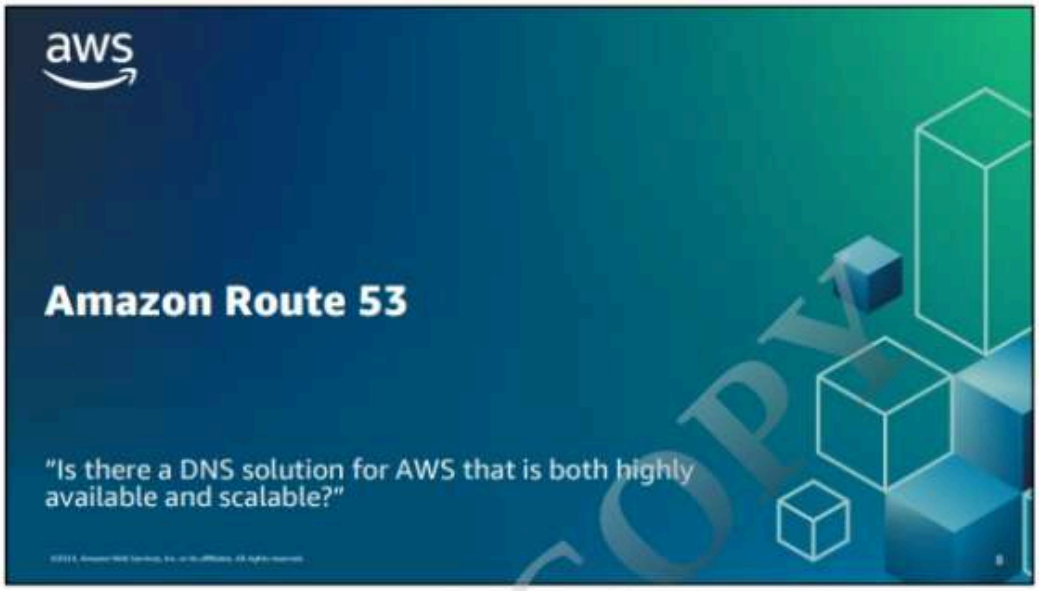
In this module, you learn about each of the services labeled in the example.

1. First, you learn about Amazon Route 53 and how you can use it to direct traffic to your resources. Route 53 operates at edge locations around the world.
2. Next, you learn about the AWS content delivery network known as Amazon CloudFront.
3. Following CloudFront, you learn about AWS WAF and AWS Shield and how they relate to distributed denial of service (DDoS) protection.
4. Last, you discover how AWS Outposts brings the power of the AWS Cloud to your on-premises network.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

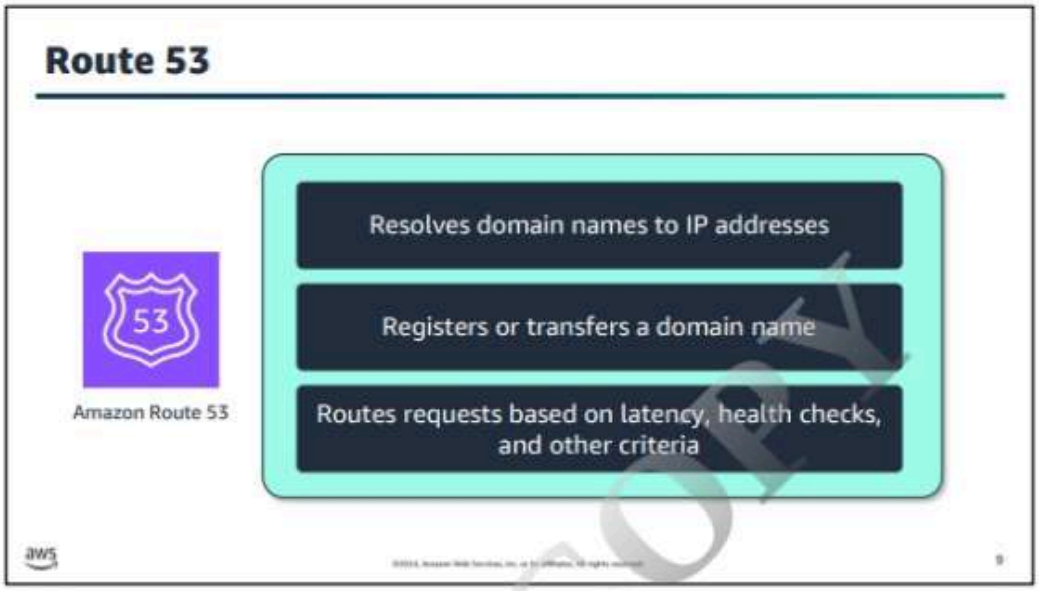


The network engineer asks, "Is there a DNS solution for AWS that is both highly available and scalable?" The networking team must choose whether to host their own DNS or use another service. The company wants your advice about how to integrate DNS services into the other AWS services that they are using.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Route 53 provides a DNS, domain name registration, and health checks. Route 53 was designed to give developers and businesses a reliable and cost-effective way to route end users to internet applications. It translates names like *example.com* into the numeric IP addresses that computers use to connect to each other.

With Route 53, you can purchase and manage domain names, such as *example.com*, and automatically configure DNS settings for your domains.

Route 53 effectively connects user requests to infrastructure running in AWS, such as Amazon Elastic Compute Cloud (Amazon EC2) instances, Elastic Load Balancing (ELB) load balancers, or Amazon Simple Storage Service (Amazon S3) buckets. You can also use it to route users to infrastructure outside of AWS.

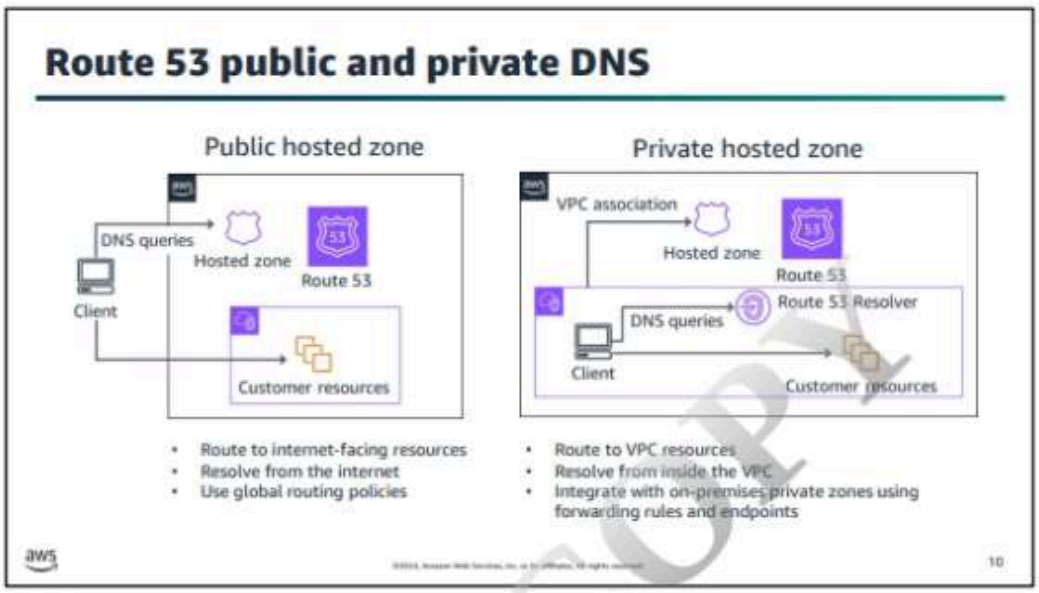
You can configure an Amazon CloudWatch alarm to check on the state of your endpoints. Combine your DNS with health check metrics to monitor and route traffic to healthy endpoints.

For more information about Route 53 health checks, see "Monitoring health checks using CloudWatch" in the *Amazon Route 53 Developer Guide* at https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/monitoring-health-checks.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



A *hosted zone* is a container for records. Records contain information about how you want to route traffic for a specific domain, such as example.com, and its subdomains, such as dev.example.com or mail.example.com. A hosted zone and the corresponding domain have the same name.

There are two types of hosted zones:

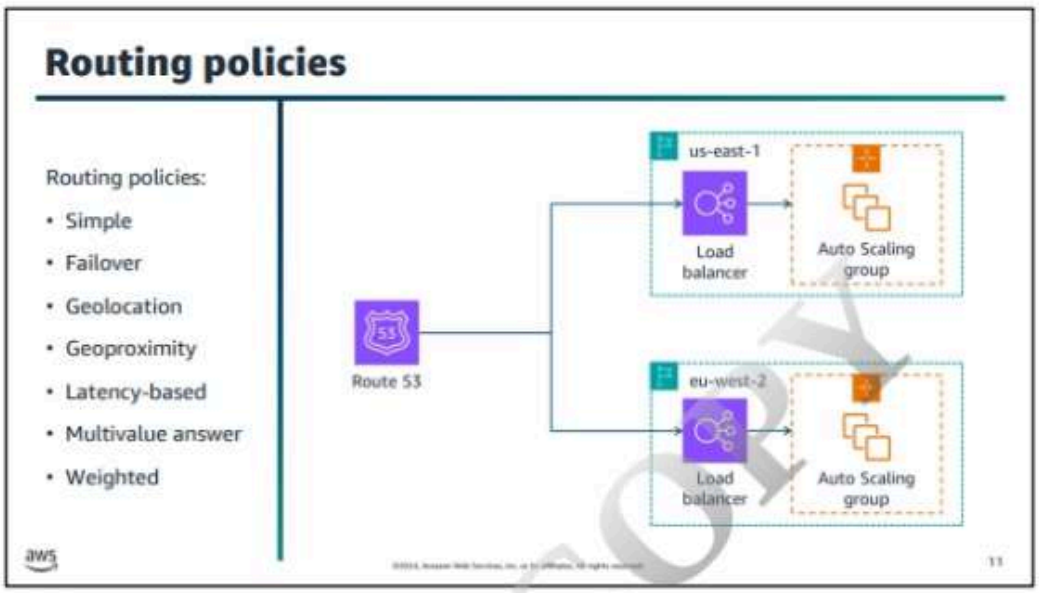
- *Public hosted zones* contain records that specify how you want to route traffic on the internet.
  - For internet name resolution
  - Delegation set, for authoritative name servers to be provided to the registrar or parent domain
- *Private hosted zones* contain records that specify how you want to route traffic in your Amazon Virtual Private Cloud (Amazon VPC).
  - For name resolution inside a VPC
  - Can be associated with multiple VPCs and across accounts

For more information, see "Working with hosted zones" in the *Amazon Route 53 Developer Guide* at https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/hosted-zones-working-with.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



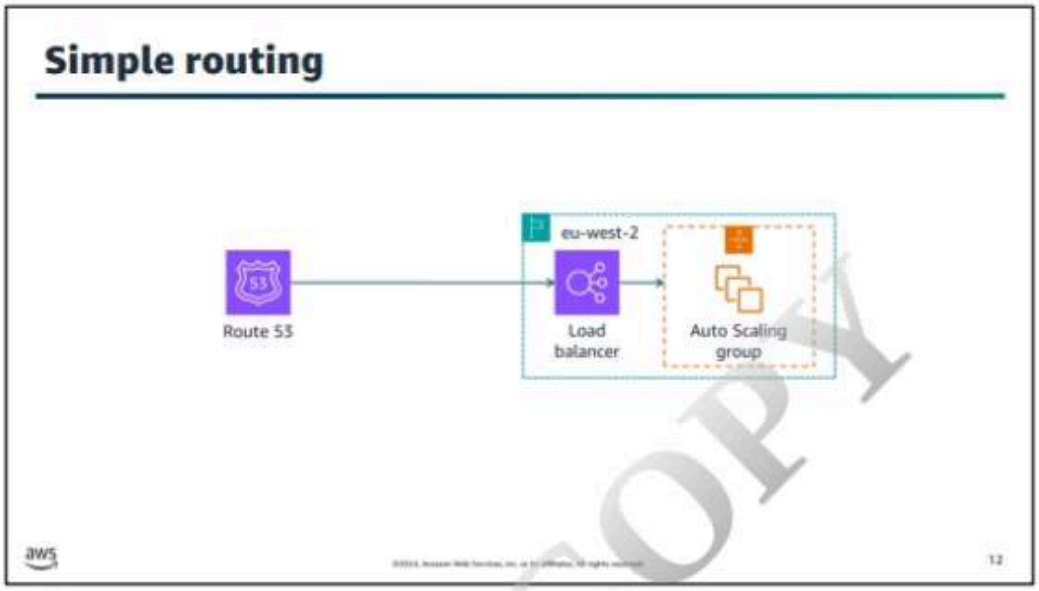
When you create a record, you choose a routing policy, which determines how Amazon Route 53 responds to queries:

- **Simple routing policy** – Use for a single resource that performs a given function for your domain. An example is a web server that serves content for the example.com website.
- **Failover routing policy** – Use when you want to configure active-passive failover.
- **Geolocation routing policy** – Use when you want to route traffic based on the location of your users.
- **Geoproximity routing policy** – Use when you want to route traffic based on the location of your resources and, optionally, shift traffic from resources in one location to resources in another.
- **Latency routing policy** – Use when you have resources in multiple AWS Regions and you want to route traffic to the Region that provides the lowest latency with less round-trip time.
- **Multivalue answer routing policy** – Use when you want Route 53 to respond to DNS queries with up to eight healthy records selected at random.
- **Weighted routing policy** – Use to route traffic to multiple resources in proportions that you specify.

For more information about routing policies, see "Choosing a routing policy" in the *Amazon Route 53 Developer Guide* at https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

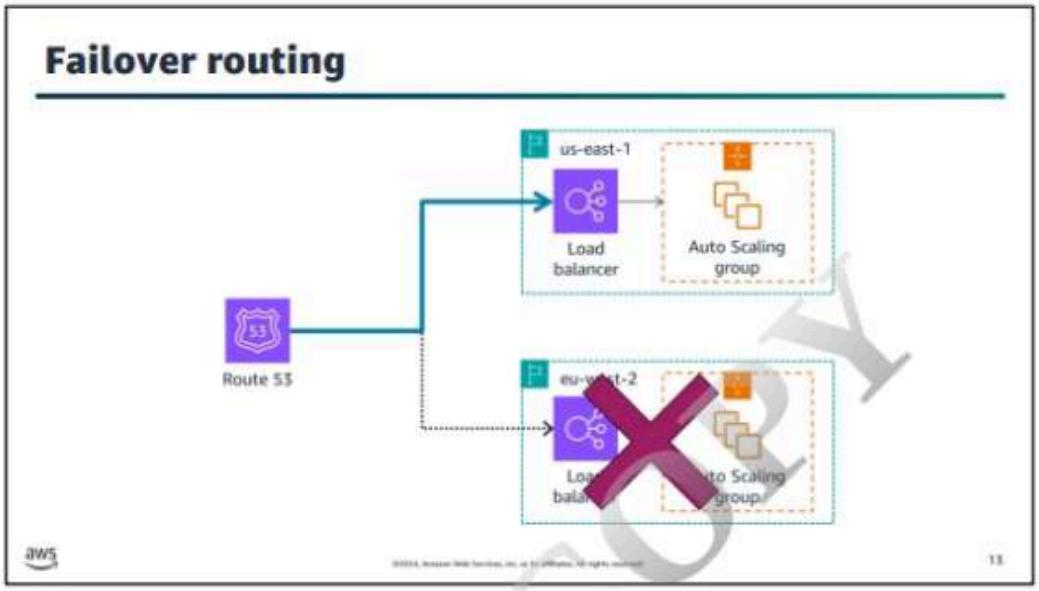


Using simple routing, you can configure standard DNS records, with no special Route 53 routing such as weighted or latency. With simple routing, you typically route traffic to a single resource, for example, to a web server for your website.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Route 53 health checks monitor the health and performance of your web applications, web servers, and other resources.

Each health check that you create can monitor one of the following:

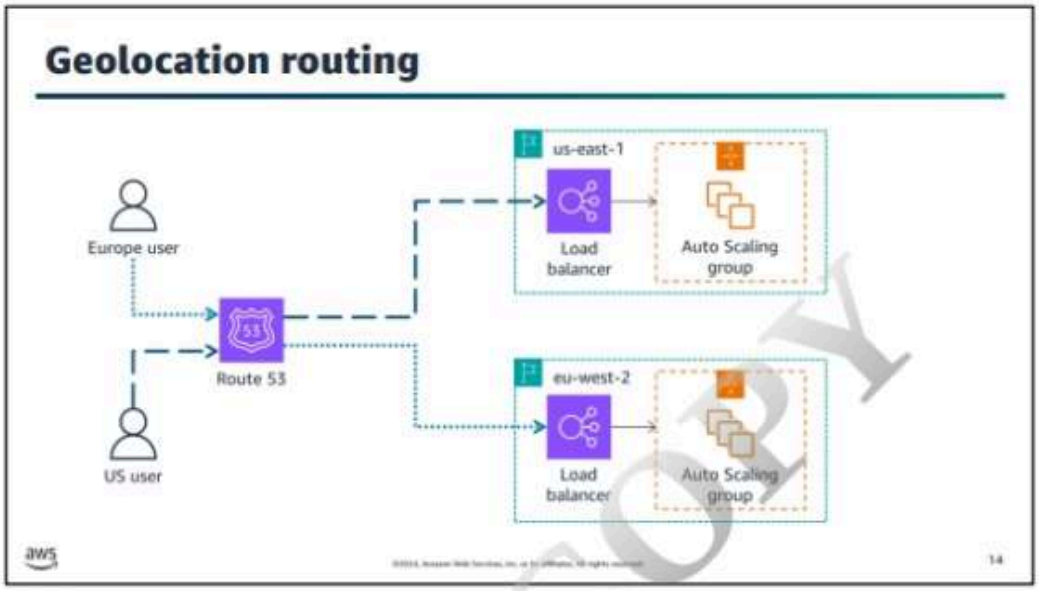
- The health of a specified resource, such as a web server
- The status of other health checks
- The status of a CloudWatch alarm

After you create a health check, you can get the status of the health check, get notifications when the status changes, and configure DNS failover.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

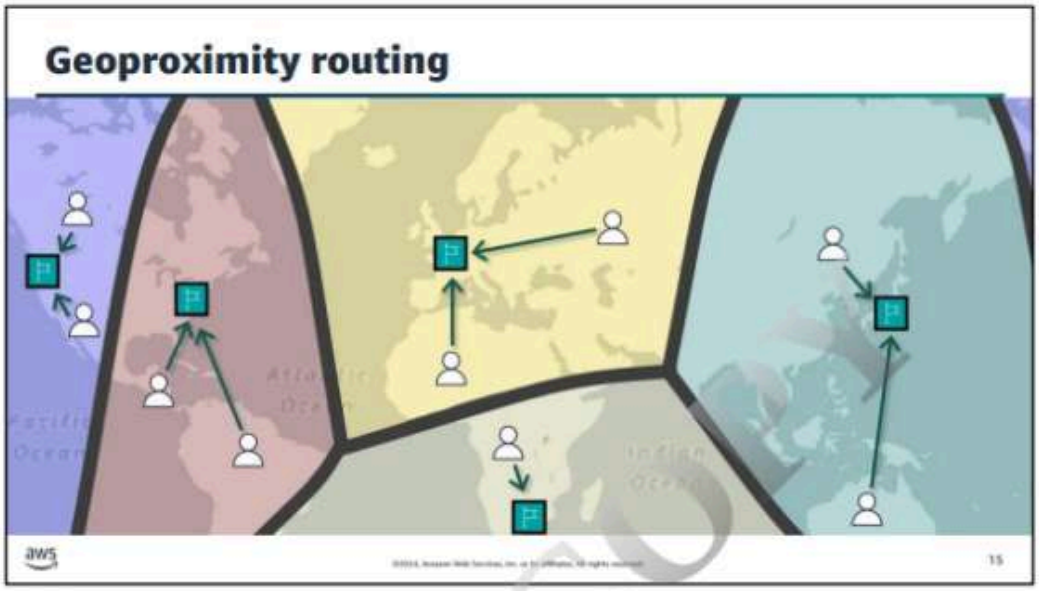


Using geolocation routing, you can choose the resources that serve your traffic based on the geographic location of your users, meaning the location that DNS queries originate from. For example, you might want all queries from Europe to be routed to an ELB load balancer in the London Region.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



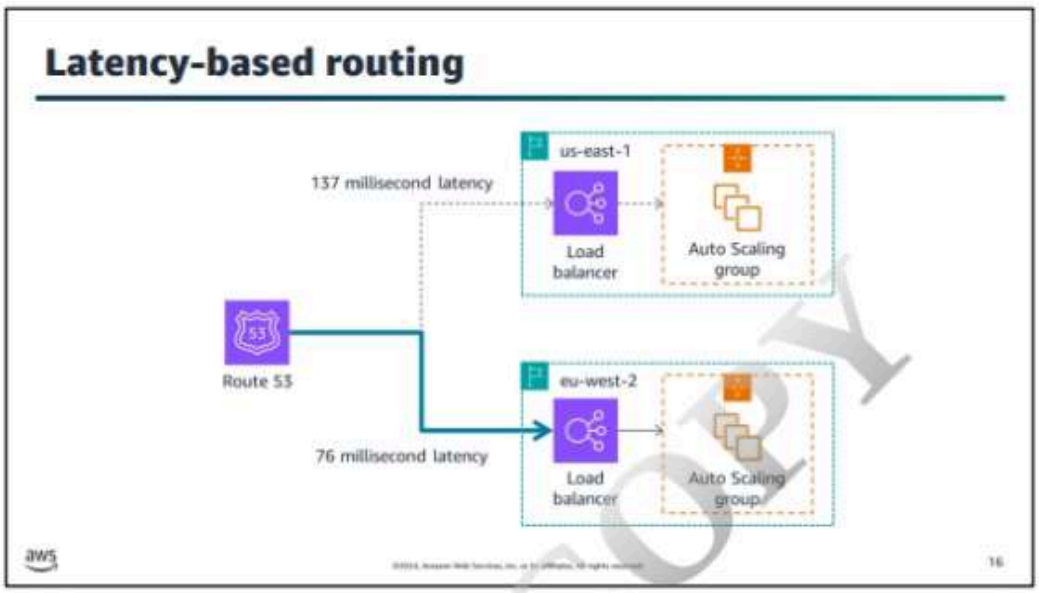
Using geoproximity routing, Route 53 can route traffic to your resources based on the geographic location of your users and your resources. You can optionally choose to route more, or less, traffic to a given resource by specifying a value, known as a *bias*. A bias expands or shrinks the size of the geographic Region from which traffic is routed to a resource.

In this example, if you increase the bias of one of the Regions, its geographic boundary will increase. This will consequently decrease the boundaries of the surrounding Regions.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

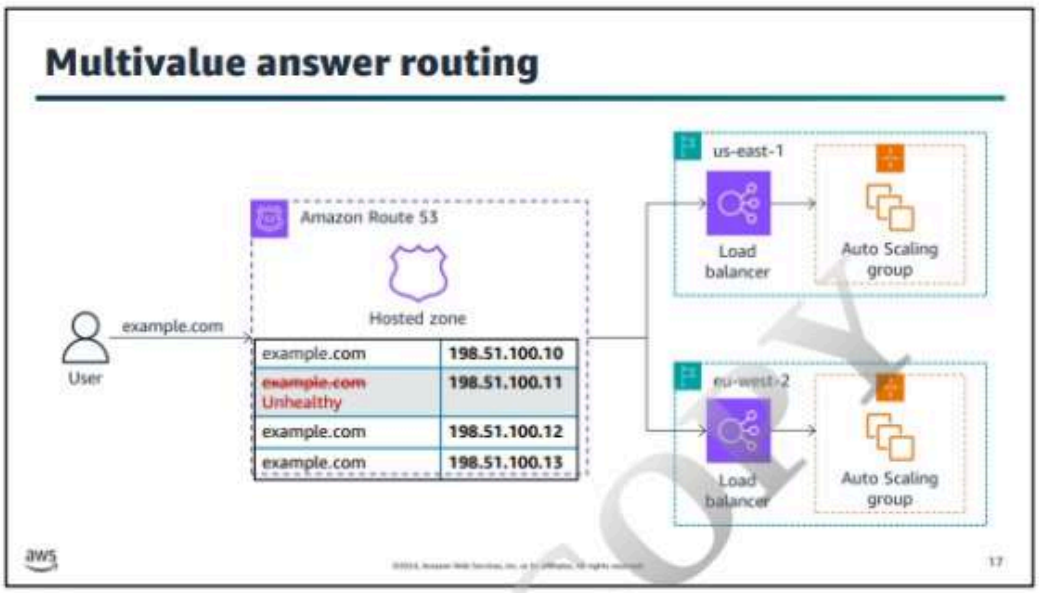


If your application is hosted in multiple AWS Regions, you can improve performance for your users by serving their requests from the AWS Region that provides the lowest latency.

Data about the latency between users and your resources is based entirely on traffic between users and AWS data centers. If you aren't using resources in an AWS Region, the actual latency between your users and your resources can vary significantly from AWS latency data. This is true even if your resources are located in the same city as an AWS Region.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



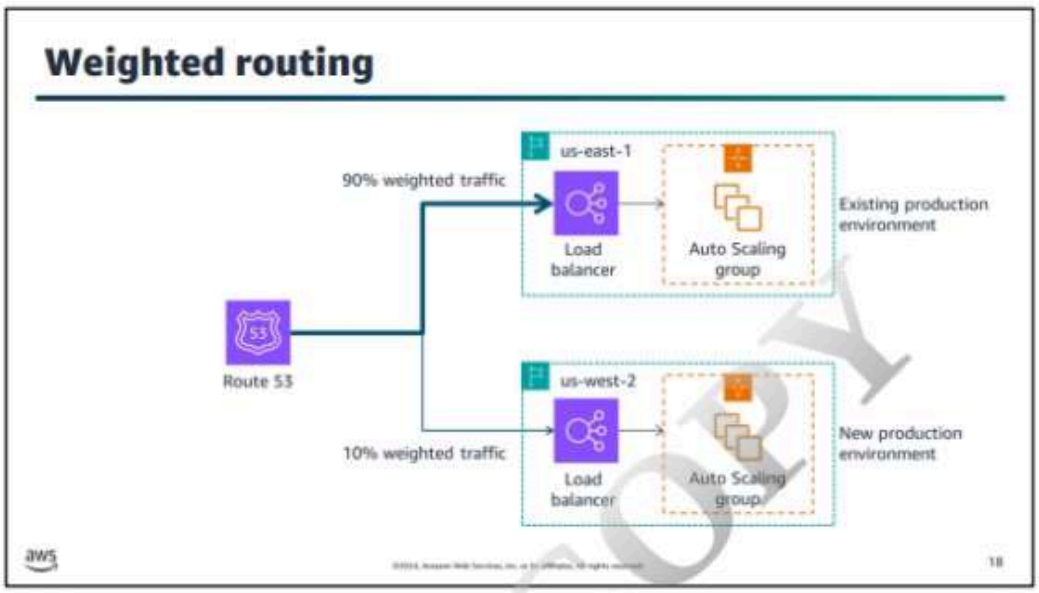
With multivalue answer routing, you can configure Route 53 to return multiple values, such as IP addresses for your web servers, in response to DNS queries. You can specify multiple values for almost any record, but with multivalue answer routing, you can also check the health of each resource. Route 53 returns values only for healthy resources.

The ability to return multiple IP addresses whose health can be checked is a way for you to use DNS to improve availability and load balancing. However, it is not a substitute for a load balancer.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



With weighted routing, you can assign weights to a resource record set to specify the frequency with which different responses are served.

In this example of a blue/green deployment, a weighted routing policy is used to send a small amount of traffic to a new production environment. If the new environment is operating as intended, the amount of weighted traffic can be increased to confirm that it can handle the increased load. If the test is successful, all traffic can be sent to the new environment.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

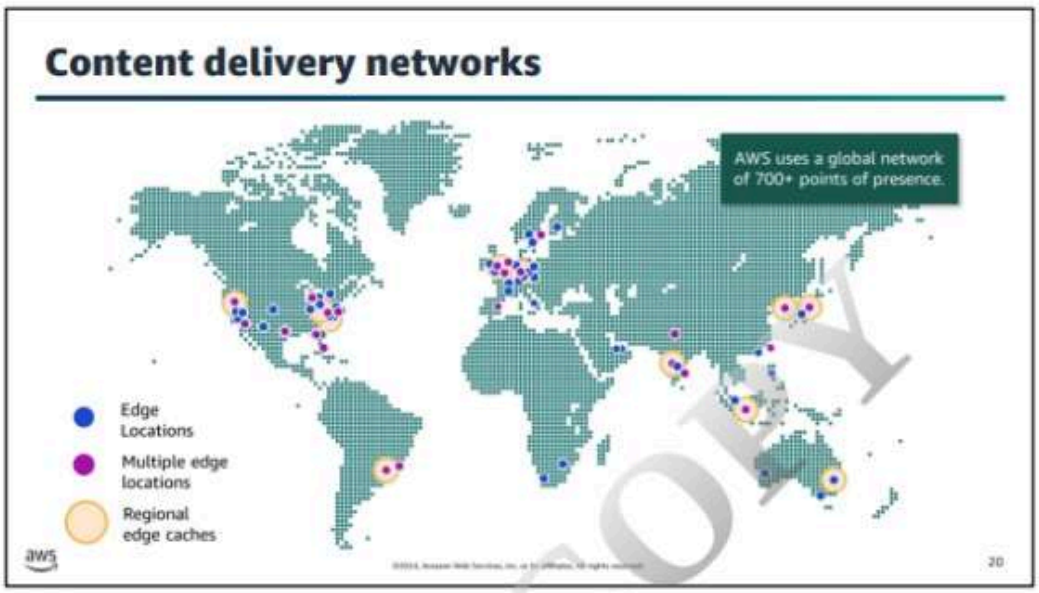
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The network engineer asks, "What service can provide the content delivery network that we need?" The engineering team needs to speed up content delivery for customers. You should explore how AWS resources can be served by edge locations on the AWS network.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



It's not always possible to replicate your entire infrastructure across the globe when your web traffic is geo-dispersed. It is also not cost effective. With a content delivery network (CDN), you can use its global network of edge locations to deliver a cached copy of your web content to your customers.

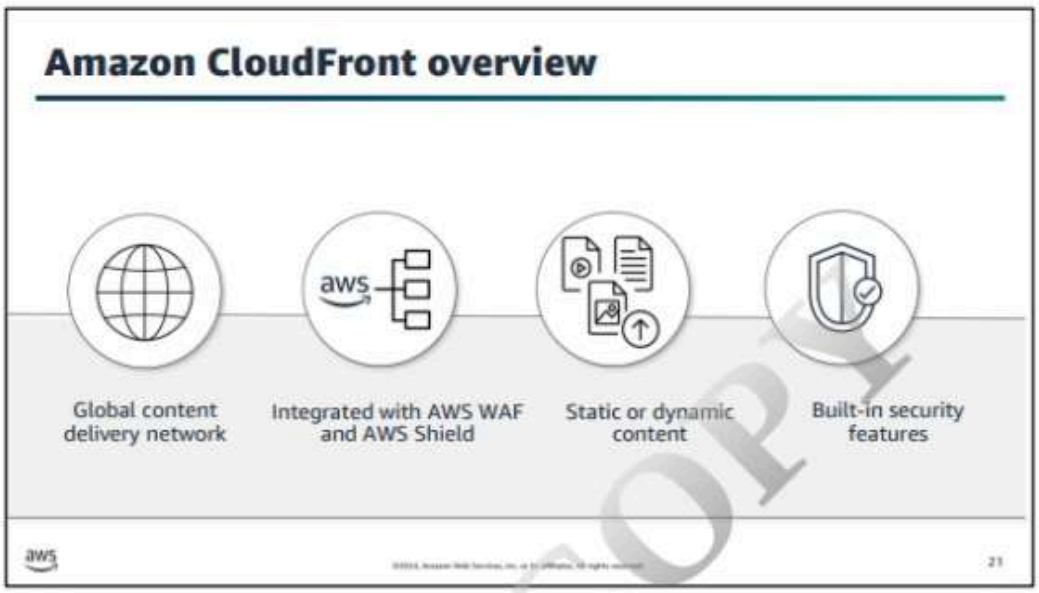
To reduce response time, the CDN uses the nearest edge location to the customer or the originating request location. Using the nearest edge location dramatically increases throughput because the web assets are delivered from cache. For dynamic data, you can configure many CDNs to retrieve data from the origin servers.

Use Regional edge caches when you have content that is not accessed frequently enough to remain in an edge location. Regional edge caches absorb this content and provide an alternative to having to retrieve that content from the origin server.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Amazon CloudFront is a global CDN service that accelerates delivery of your websites, APIs, video content, or other web assets. It integrates with other AWS products to give developers and businesses a straightforward way to accelerate content to users. There are no minimum usage commitments.

CloudFront provides extensive flexibility for optimizing cache behavior, coupled with network-layer optimizations for latency and throughput. The CDN offers a multi-tier cache by default. It has Regional edge caches that improve latency and lower the load on your origin servers when the object is not already cached at the edge.

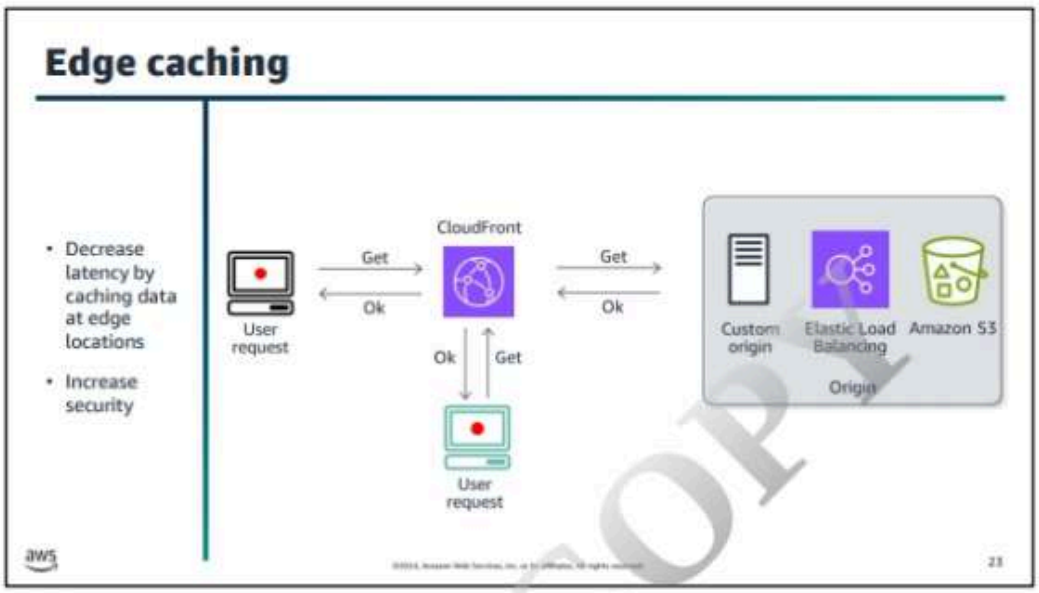
CloudFront supports real-time, bidirectional communication over the WebSocket protocol. With this persistent connection, clients and servers can send real-time data to one another without the overhead of repeatedly opening connections. This is especially useful for communications applications such as chat, collaboration, gaming, and financial trading.

With support for WebSockets in CloudFront, customers can manage WebSocket traffic through the same avenues as any other dynamic and static content. With CloudFront, you can take advantage of DDoS protection by using the built-in CloudFront integrations with AWS Shield and AWS WAF.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Imagine you are serving an image from a traditional web server, not from CloudFront. You might serve an image named sunsetphoto.png using the URL http://example.com/sunsetphoto.png. Your users can navigate to this URL and see the image.

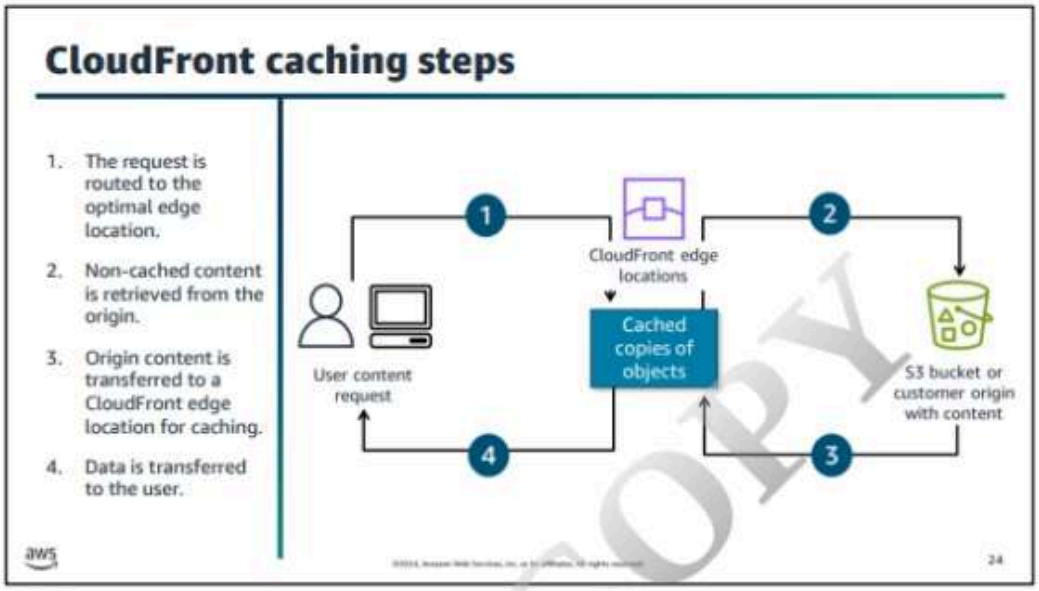
The users don't realize that their request was routed from one network to another (through the complex collection of interconnected networks that comprise the internet) until the image was found.

CloudFront speeds up the distribution of your content by routing each user request through the AWS backbone network to the edge location that can best serve your content. Typically, this is a CloudFront edge server that provides the fastest delivery to the viewer. Using the AWS network can dramatically reduce the number of networks your users' requests must pass through, which improves performance. Users get lower latency (the time it takes to load the first byte of the file) and higher data transfer rates. You also get increased reliability and availability because copies of your files (also called *objects)* are now held (or cached) in multiple edge locations around the world.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



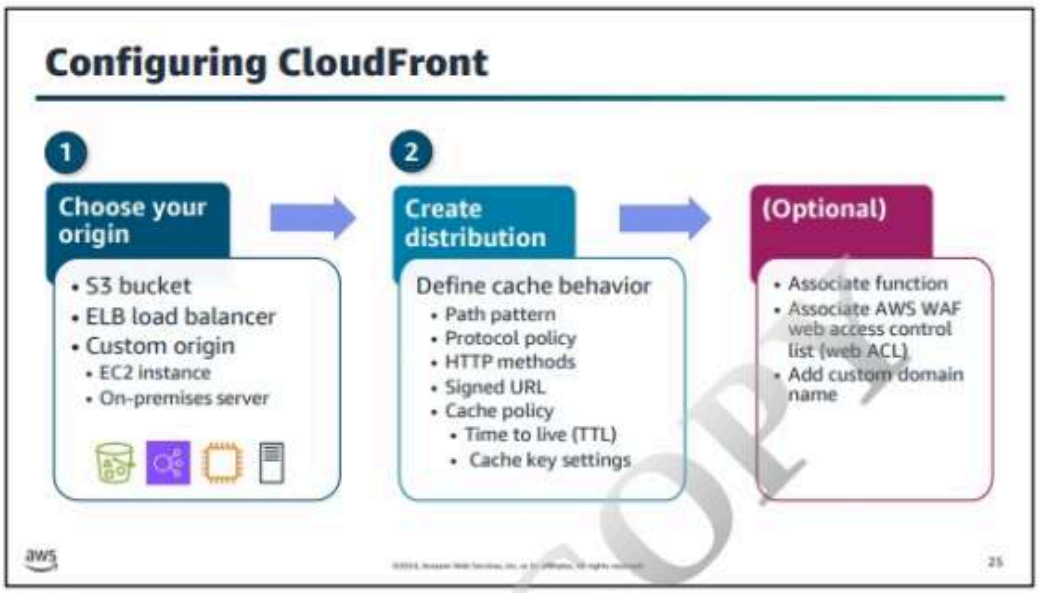
Review the steps to learn what happens when you request an object from an origin.

If content is already cached and its time-to-live (TTL) has not expired, then steps 2 and 3 are skipped, and the data is transferred from the edge location to the content requester.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



To configure CloudFront, you first specify an *origin server* from which CloudFront gets your files—for example, an S3 bucket or your HTTP server. These will be distributed from CloudFront edge locations all over the world.

- An origin server stores the original, definitive version of your objects.
- If you are serving content over HTTP, your origin server is either an S3 bucket or an HTTP server, such as a web server.
- Your HTTP server can run on an EC2 instance or on an on-premises server that you manage. These servers are also known as *custom origins*.

Next, you create a CloudFront distribution. This tells CloudFront which origin servers to get your files from when users request the files through your website or application. At the same time, you specify details such as whether you want CloudFront to log all requests and whether you want the distribution to be activated as soon as it's created.

CloudFront assigns a domain name to your new distribution. The service sends your distribution's configuration, but not the content, to all of its edge locations.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Improving performance

What AWS does:	What you can do:
• TCP optimization	• Choose your caching strategy
• TLS 1.3 support	• Improve your cache hit ratio
• Dynamic content placement	• Use CloudFront Origin Shield

26

Amazon CloudFront is a managed service. You should understand what services AWS provides to help you improve performance. You are also responsible for configuring CloudFront to optimize your application's performance.

AWS provides features that improve the performance of your content delivery, such as the following:

- **TCP optimization** – CloudFront uses TCP optimization to observe how fast a network is already delivering your traffic and the latency of your current round trips. It then uses that data as input to automatically improve performance.
- **TLS 1.3 support** – CloudFront supports TLS 1.3, which provides better performance with a simpler handshake process that requires fewer round trips. It also adds improved security features.
- **Dynamic content placement** – Use CloudFront to serve dynamic content, such as web applications or APIs from ELB load balancers or Amazon EC2 instances. You can improve the performance, availability, and security of your content.

You can also adjust the configuration of your CloudFront distribution in the following ways to accommodate for better performance:

- **Define your caching strategy** – Choosing an appropriate TTL is important. In addition, consider caching based on things like query string parameters, cookies, or request headers.
- **Improve your cache hit ratio** – You can view the percentage of viewer requests that are hits, misses, and errors in the CloudFront console. Make changes to your distribution based on statistics collected in the CloudFront cache statistics report.
- **Use CloudFront Origin Shield** – Get an additional layer of caching between the Regional edge caches and your origin. It is not always a best fit for your use case, but it can be beneficial for viewers that are spread across geographic regions or on-premises origins with capacity or bandwidth constraints.

For more information about dynamic content delivery, see Amazon CloudFront Dynamic Content Delivery at https://aws.amazon.com/cloudfront/dynamic-content/.

For more information about improving caching and availability in your distributions, see "Caching and availability" in the *Amazon CloudFront Developer Guide* https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/ConfiguringCaching.html.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

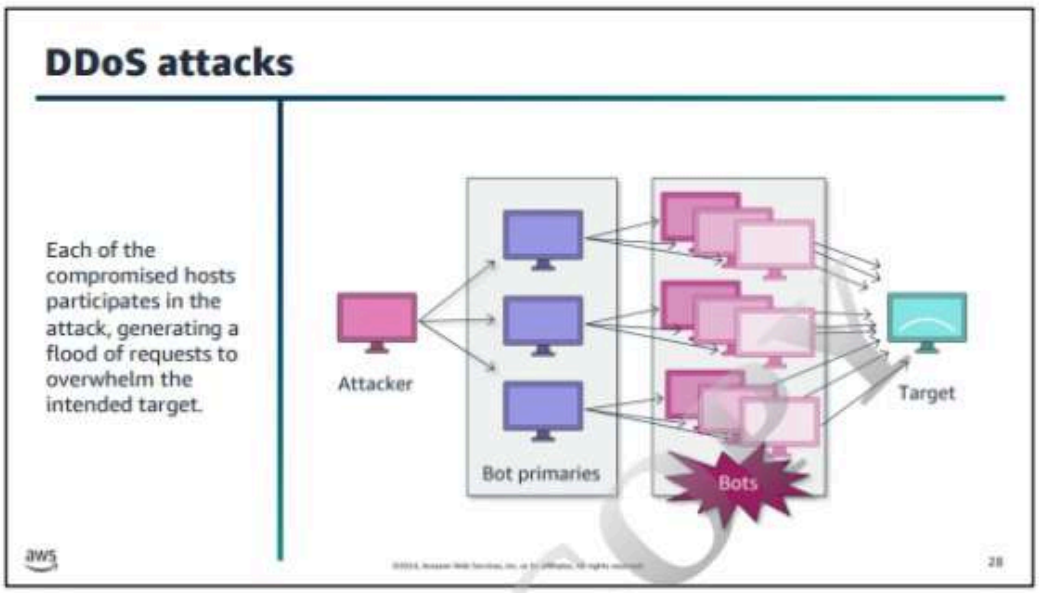
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The network engineer asks, "How can we protect public-facing applications?" The engineering team needs to prepare for distributed denial of service (DDoS) attacks. They want you to explain to them the best practices for protecting AWS services at the edge.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



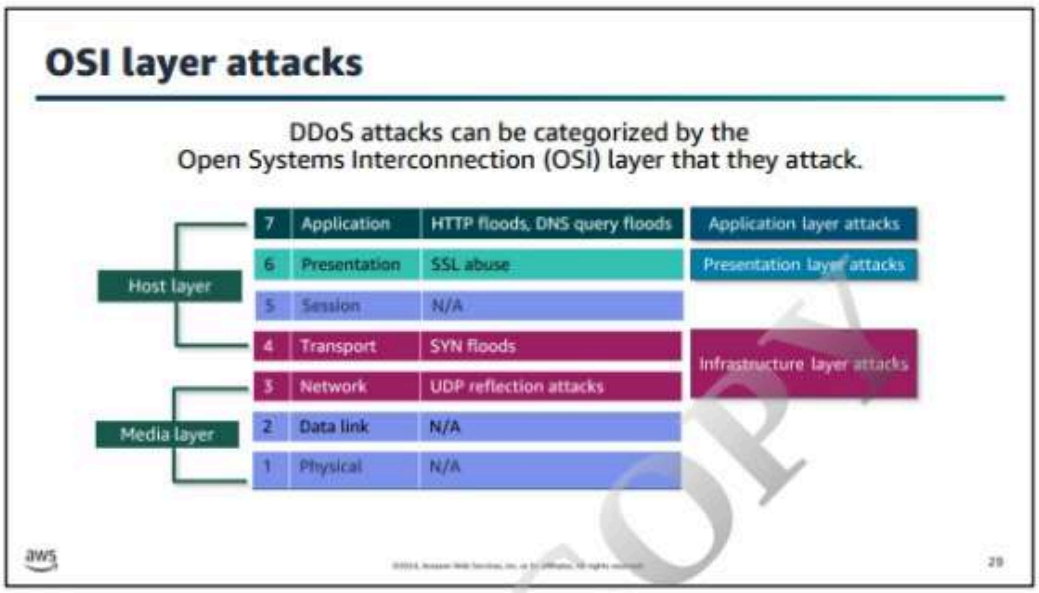
A DDoS attack is an attack in which multiple compromised systems attempt to flood a target, such as a network or web application, with traffic. A DDoS attack can prevent legitimate users from accessing a service and can cause the system to crash because of the overwhelming traffic volume.

The general concept of a DDoS attack is to use additional hosts to amplify the requests made to the target, rendering them at full capacity and unavailable. In the diagram, a single attacker uses three bot primaries. Each bot primary controls multiple bots that all attack the same target.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



In general, DDoS attacks can be isolated by the Open Systems Interconnection (OSI) model layer that they attack. Attacks are most common at the network layer (Layer 3), transport layer (Layer 4), presentation layer (Layer 6), and application layer (Layer 7).

Infrastructure layer attacks
Attacks at Layer 3 and Layer 4 are typically categorized as infrastructure layer attacks. These are also the most common type of DDoS attack and include vectors such as synchronized (SYN) floods and other reflection attacks, such as UDP packet floods. These attacks are usually large in volume and aim to overload the capacity of the network or the application servers. Fortunately, these are also the type of attacks that have clear signatures and are easier to detect.

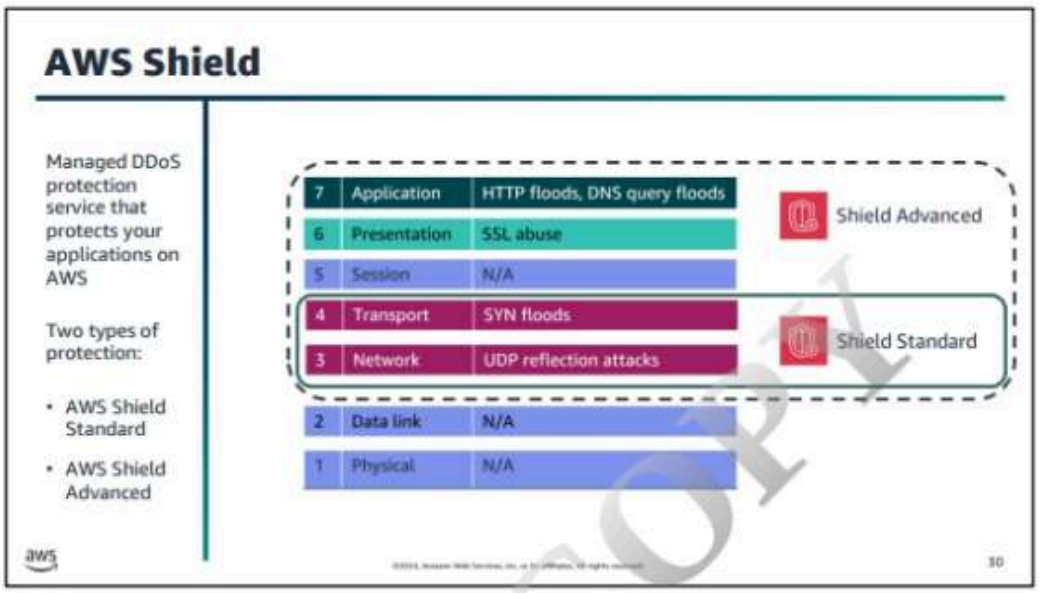
Application layer attacks
An attacker might target the application itself by using a Layer 7, or application layer, attack. In these attacks, similar to SYN flood infrastructure attacks, the attacker attempts to overload specific functions of an application to make the application unavailable or extremely unresponsive to legitimate users.

Presentation layer attacks
An attacker can use a Layer 6, or presentation layer, attack by exploiting vulnerabilities in the protocols responsible for data representation, encryption, and compression. Attackers can exploit vulnerabilities in SSL/TLS protocols to decrypt encrypted traffic or perform man-in-the-middle attacks, giving attackers the ability to intercept and manipulate sensitive data.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS Shield is a managed DDoS protection service that safeguards your applications running on AWS. It provides dynamic detection and automatic inline mitigations that minimize application downtime and latency. There are two tiers of AWS Shield: Shield Standard and Shield Advanced.

Shield Standard provides protection against some of the most common and frequently occurring infrastructure (Layer 3 and 4) attacks. This includes SYN/UDP floods and reflection attacks. Shield Standard improves availability of your applications on AWS. The service applies a combination of traffic signatures, anomaly algorithms, and other analysis techniques. Shield Standard detects malicious traffic and provides real-time issue mitigation. You are protected by Shield Standard at no additional charge.

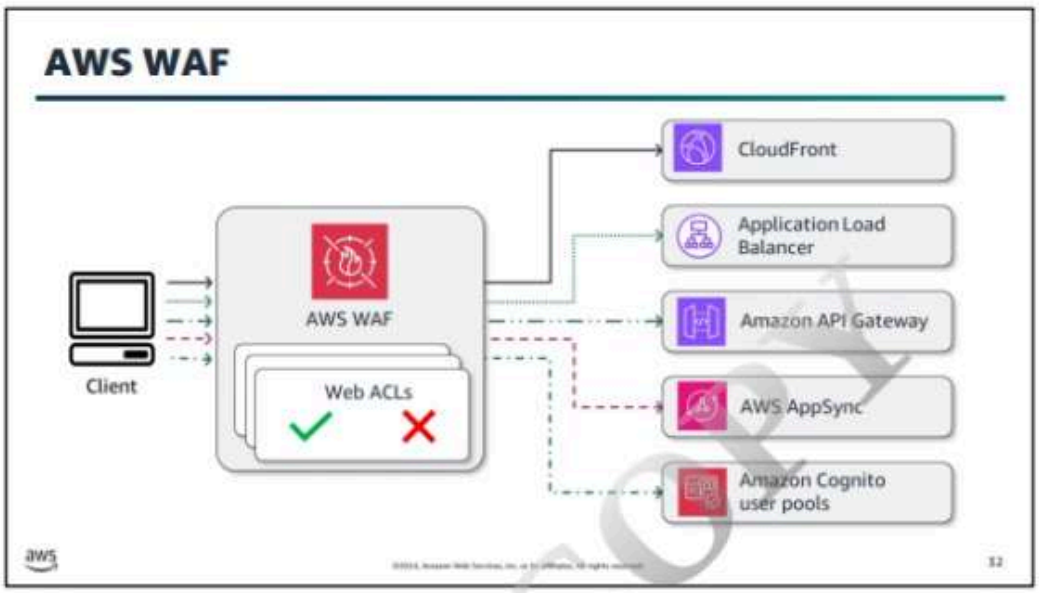
If you need even more protection from DDoS attacks on your applications, consider using Shield Advanced. You get additional detection and mitigation against large and sophisticated DDoS attacks, near real-time visibility, and integration with AWS WAF, a web application firewall.

For more information, see "How AWS Shield and Shield Advanced work" in the *AWS WAF, AWS Firewall Manager, and AWS Shield Advanced Developer Guide* at https://docs.aws.amazon.com/waf/latest/developerguide/ddos-overview.html.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS WAF is a web application firewall that helps protect your web applications or APIs against common web exploits and bots. AWS WAF gives you control over how traffic reaches your applications. Create security rules that control bot traffic and block common attack patterns, such as SQL injection (SQLi) or cross-site scripting (XSS). You can also monitor HTTP(S) requests that are forwarded to your compatible AWS services.

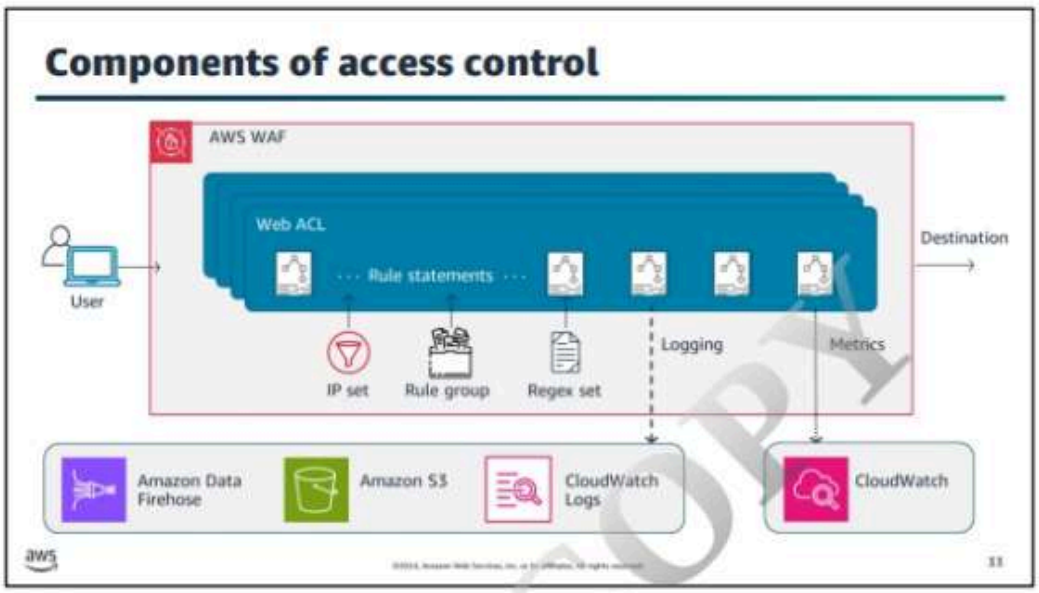
In the diagram, a client sends an inbound request to one of the five supported services: CloudFront, Application Load Balancer, Amazon API Gateway, AWS AppSync, or Amazon Cognito user pools. The traffic is evaluated against web ACLs before it reaches the destination. If the traffic passes all web ACLs without a deny, it is sent to the destination AWS service.

For more information about AWS WAF, see "Setting up AWS WAF and its components" at https://docs.aws.amazon.com/waf/latest/developerguide/getting-started.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Before configuring AWS WAF, you should understand the components used to control access to your AWS resources.

- **Web ACLs** – You use a web ACL to protect a set of AWS resources. You create a web ACL and define its protection strategy by adding rules.
- **Rules** – Rules define criteria for inspecting web requests and specify how to handle requests that match the criteria.
- **Rule groups** – You can use rules individually or in reusable rule groups. AWS Managed Rules for AWS AWF and AWS Marketplace sellers provide managed rule groups for your use. You can also define your own rule groups.
- **Rule statements** – This part of a rule tells AWS WAF how to inspect a web request. When AWS WAF finds the inspection criteria in a web request, we say that the web request matches the statement.
- **IP set** – This is a collection of IP addresses and IP address ranges that you want to use together in a rule statement. IP sets are AWS resources.
- **Regex pattern set** – This is a collection of regular expressions that you want to use together in a rule statement. Regex pattern sets are AWS resources.
- **Monitoring and logging** – You can monitor web requests, web ACLs, and rules using CloudWatch. You can also activate logging to get detailed information about traffic that is analyzed by your web ACL. You choose where to send your logs: CloudWatch Logs, Amazon S3, or Amazon Data Firehose.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Control traffic with ACL rule statements

Attack prevention
- SQL injection and cross-site scripting detection
- AWS Managed Rules for AWS WAF
- AWS Marketplace managed rule groups

Traffic filtering
- Rate limiting
- IP filtering with full Classless Inter-Domain Routing (CIDR) range support
- Geofencing by country

Pattern matching
- Regex match
- String match
- Size constraint match

Logical operation
- AND statement
- OR statement
- NOT statement

14

Rule statements are the part of a rule that tells AWS WAF how to inspect a web request. When AWS WAF finds the inspection criteria in a web request, we say that the web request matches the statement.

Every rule statement does the following:
- Specifies what to look for and how, according to the statement type
- Has a single top-level rule statement that can contain other statements

Rule statements can be very basic. For example, you could have a statement that provides a set of originating countries for which you should check web requests. Rule statements can also be very complex. Examples include statements that combine many other statements with logical AND, OR, and NOT statements.

Web ACLs can contain rule statements that reference rule groups. On the console, you don't see these represented as rule statements. Every web ACL has a JSON format representation. In the JSON, you can see these special types of rule statements. For rules of any complexity, manage your web ACL with the JSON editor.

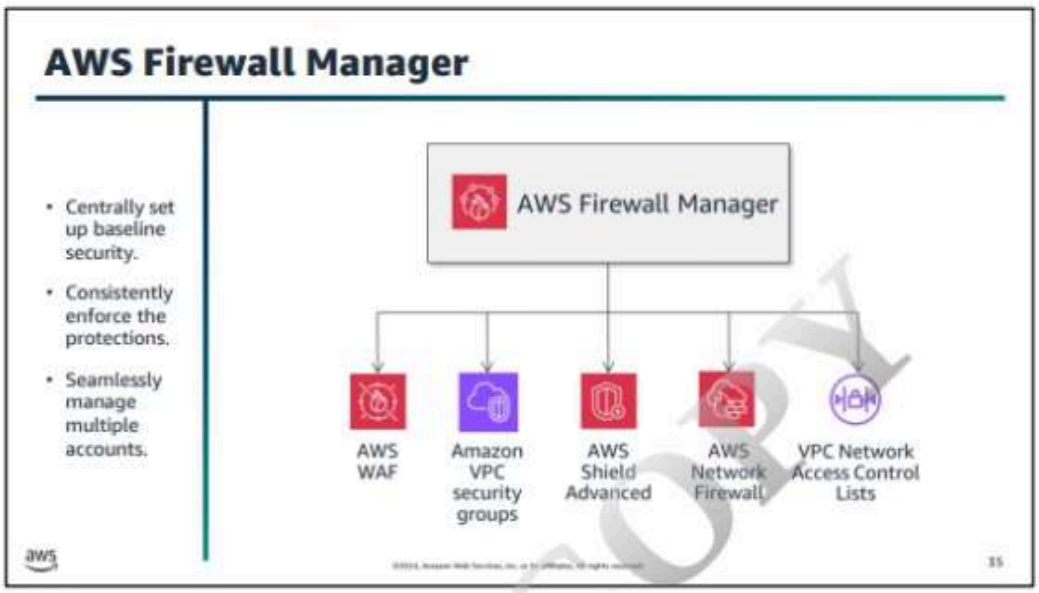
AWS WAF supports nesting for many rule statements. For example, use nesting to control the rate of requests coming from a specific geographic area. Use a rate-based rule and nest a geographic match rule inside it to narrow the scope.

For more information, see "Using rule statements in AWS WAF" in the *AWS WAF, AWS Firewall Manager, and AWS Shield Advanced Developer Guide* at https://docs.aws.amazon.com/waf/latest/developerguide/waf-rule-statements.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS Firewall Manager helps you with the administration and maintenance tasks of your AWS WAF and Amazon VPC security groups. You set up your AWS WAF firewall rules, Shield protections, and Amazon VPC security groups once. Firewall Manager automatically applies the rules and protections across your accounts and resources, even as you add new resources.

Firewall Manager helps you do the following:
- Simplify rule management across applications and accounts.
- Automatically discover new accounts and remediate noncompliant events.
- Deploy AWS WAF rules from AWS Marketplace.
- Activate rapid response to attacks across all accounts.

As new applications are created, Firewall Manager also facilitates bringing new applications and resources into compliance with a common set of security rules from day one. Now you have a single service to build firewall rules, create security policies, and enforce them in a consistent, hierarchical manner across your entire AWS infrastructure.

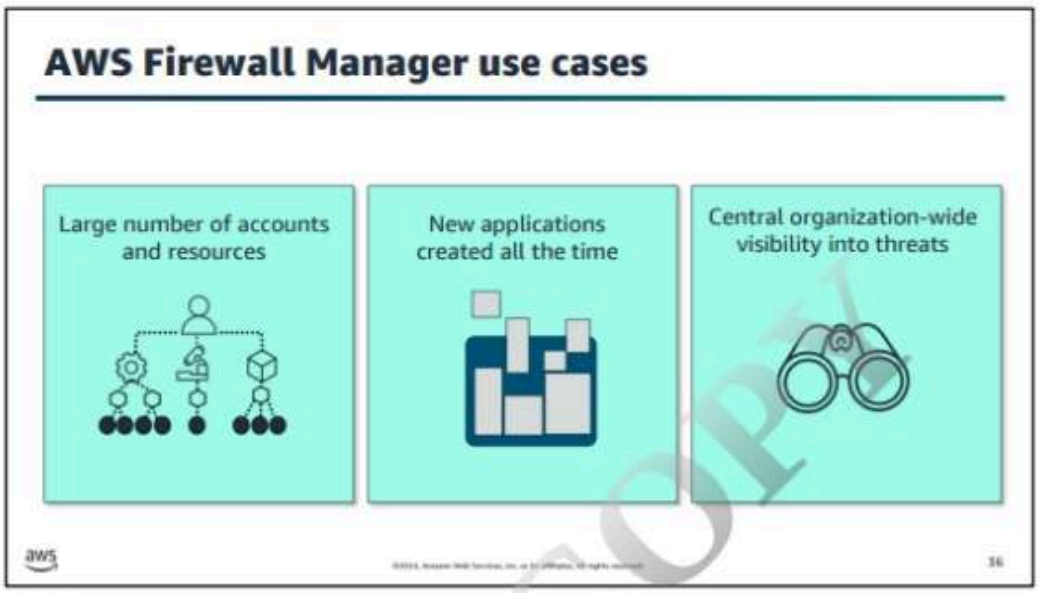
There are three prerequisites for using it: activate AWS Organizations with full features, use AWS Config, and have an assigned user as the Firewall Manager administrator.

For more information, see "Use AWS Firewall Manager to deploy protection at scale in AWS Organizations" in the *AWS Security Blog* at https://aws.amazon.com/blogs/security/use-aws-firewall-manager-to-deploy-protection-at-scale-in-aws-organizations/.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



As your number of applications in the AWS Cloud grows, you should be familiar with how to manage compliance at scale.

Some of the challenges you will face include the following:

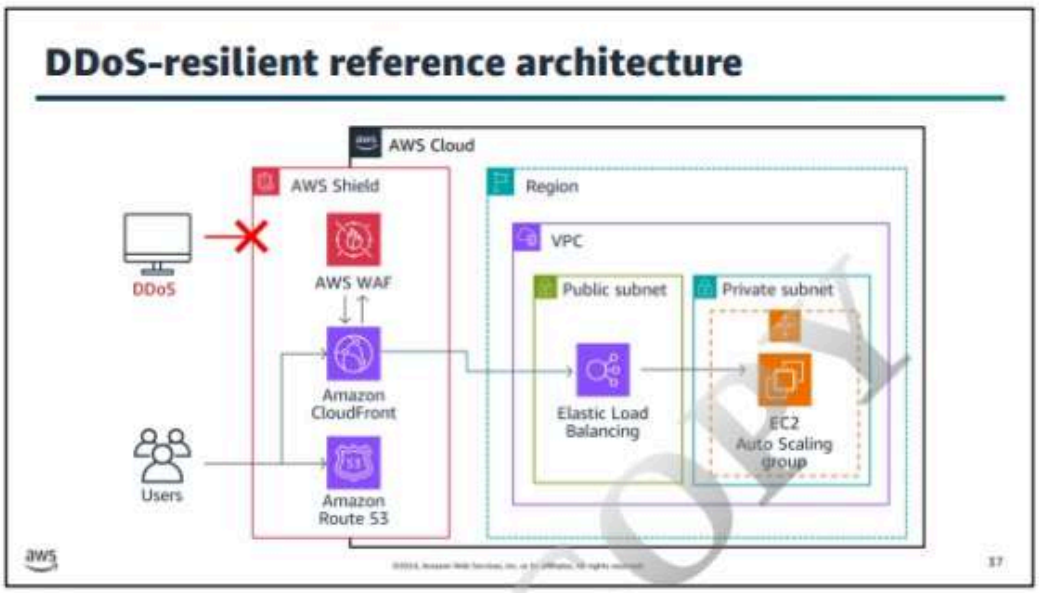
- **Large number of accounts and resources** – It is hard to manage security policies centrally across all accounts and resources.
- **New applications created all the time** – It is difficult to be sure all applications are consistently protected on day one.
- **Organization-wide visibility into threats** – There is no single place to monitor and respond to any threats across the organization.

Use AWS Firewall Manager to scale up cloud security and monitoring in your environments.

For more information, see "Centrally manage AWS WAF (API v2) and AWS Managed Rules at scale with Firewall Manager" in the *AWS Security Blog* at https://aws.amazon.com/blogs/security/centrally-manage-aws-waf-api-v2-and-aws-managed-rules-at-scale-with-firewall-manager/.

DO NOT COPY
javierartigaespublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



In the diagram, AWS Shield and WAF sit on the edge of the architecture and act as gatekeepers, allowing or denying traffic. Recall that Shield protects common Layer 3 and Layer 4 infrastructure attacks. AWS WAF protects Layer 7, the application layer.

You keep bad traffic at the edge by using services such as Route 53 and CloudFront with AWS WAF and Shield. In this architecture, auto scaling is a last line of defense. Each of these AWS services, depending on the attacks, catch attacks and block them before they reach your VPC.

Services that are available in AWS edge locations, such as CloudFront, AWS WAF, and Route 53, take advantage of a global network of edge locations. Multiple edge locations provide applications with greater fault tolerance and increased scale for managing larger volumes of traffic.

In a traditional data center environment, you can mitigate DDoS attacks of the infrastructure layer. You can use techniques such as overprovisioning capacity, deploying DDoS mitigation systems, or scrubbing traffic with the help of DDoS mitigation services. AWS automatically provides DDoS mitigation capabilities. You can optimize your application's DDoS resilience by making architecture choices that best use those capabilities and give you the ability to scale for excess traffic.

For more information, see the AWS Whitepaper, "AWS Best Practices for DDoS Resiliency" at https://docs.aws.amazon.com/whitepapers/latest/aws-best-practices-ddos-resiliency/welcome.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

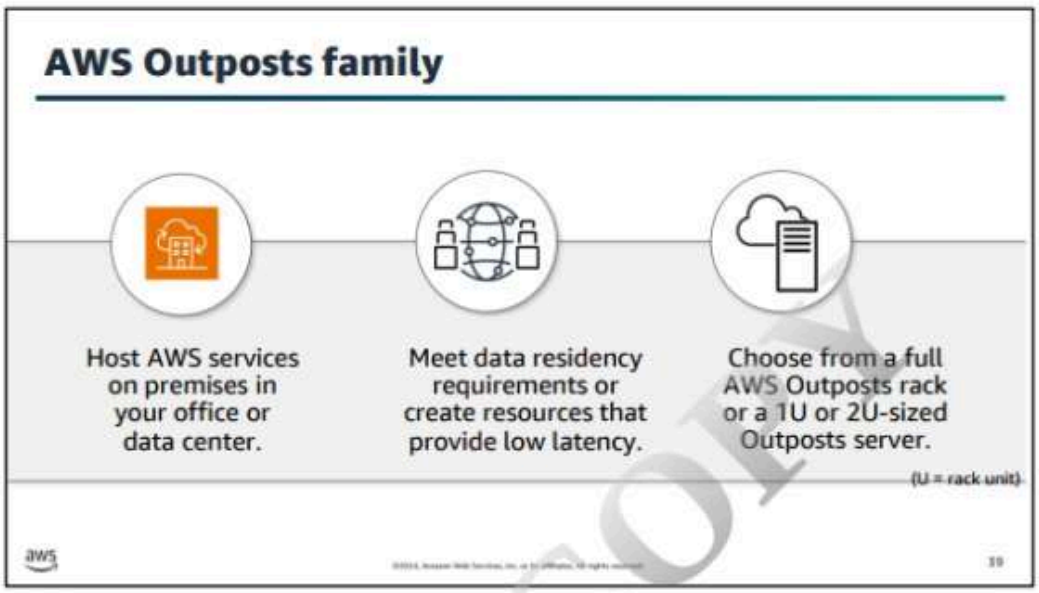
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The network engineer asks, "Does AWS support any services running on premises to meet our latency and residency requirements?" The engineering team must keep certain applications in their own data center for compliance and performance. The engineer wants you to explain how they can manage local resources using the same services that exist in the AWS Cloud.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



To best understand AWS Outposts, consider this challenge: As you migrate applications to AWS, are there any applications that must remain on premises?

These applications might need to generate near real-time responses to end-user applications. Or they might need to communicate with other on-premises systems or control on-site equipment. Examples include workloads running on factory floors for automated operations in manufacturing, real-time patient diagnosis or medical imaging, and content and media streaming.

You need a solution to securely store and process customer data that must remain on premises or in countries outside of an AWS Region. You need a solution to run data-intensive workloads and process data locally, or when you want closer controls on data analysis, backup, and restore.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Outposts rack and Outposts servers

Outposts rack*

- Scale up to 96 42U–standard racks
- Pool compute and storage capacity between multiple Outposts racks
- Get more service options

Outposts servers

- Place in your own rack
- Choose from:
  - 1U Graviton-based processor
  - 2U Intel Xeon Scalable processor

*Outposts rack requires an AWS Enterprise Support plan

40

With Outposts, you can extend the AWS Cloud to an on-premises data center. Outposts come in different form factors, each with separate requirements. Verify that your site meets the requirements for the form factor that you're ordering.

When you order an *Outposts rack*, you can choose from a variety of Outposts configurations. Each configuration provides a mix of EC2 instance types and Amazon Elastic Block Store (Amazon EBS) volumes.

To fulfill the Outposts rack order, AWS will schedule a date and time with you. You will also receive a checklist of items to verify or provide before the installation. The team will roll the rack to the identified position, and your electrician can power the rack. The team will establish network connectivity for the rack over the uplink that you provide, and they will configure the rack's capacity.

With *Outposts servers*, you can order hardware at a smaller scale while still providing your AWS services on premises. You can choose from ARM-based or Intel-based options. Not all services available in Outposts racks are supported in Outposts servers.

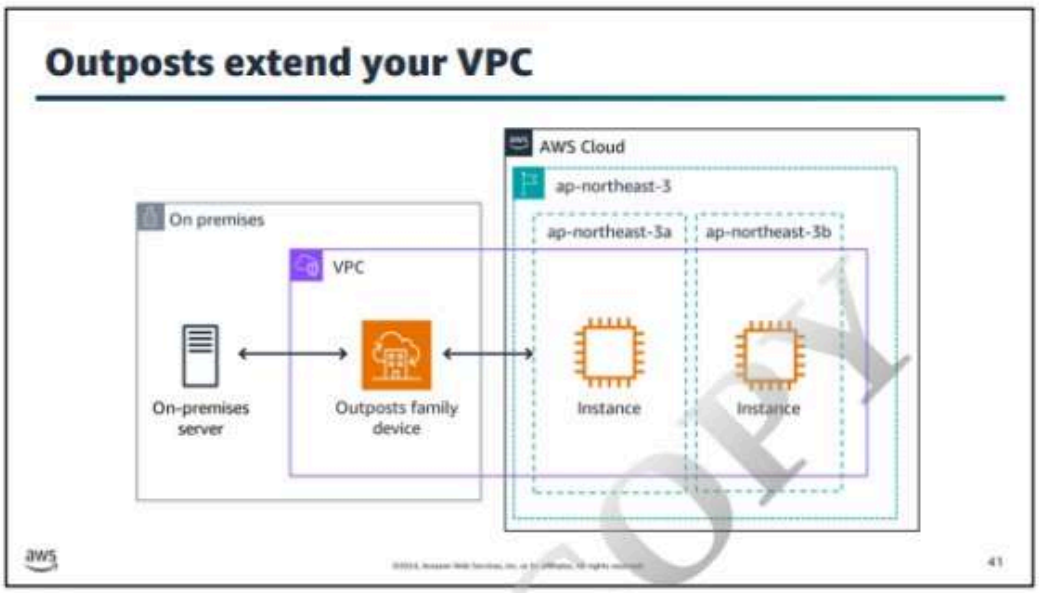
Outposts servers are delivered directly to you and installed by either your own onsite personnel or a third-party vendor. Once connected to your network, AWS will remotely provision compute and storage resources.

For more information about Outpost site requirements, see "Site requirements for Outposts racks" in the *AWS Outposts User Guide* at https://docs.aws.amazon.com/outposts/latest/userguide/outposts-requirements.html.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



A VPC spans all Availability Zones in its AWS Region. You can extend any VPC in the Region to your Outpost by adding an Outpost subnet.

The diagram shows the relation between your Outposts and the AWS Cloud. Your Outposts family device (Outposts rack or Outposts server) is configured on premises on your network. Other on-premises devices such as a local server can communicate with Outposts. Your Outposts device is linked to a specific Availability Zone—in this case, ap-northeast-3a. The subnets have been removed from the diagram for simplicity. Adding an Outposts subnet extends your VPC to your on-premises network.

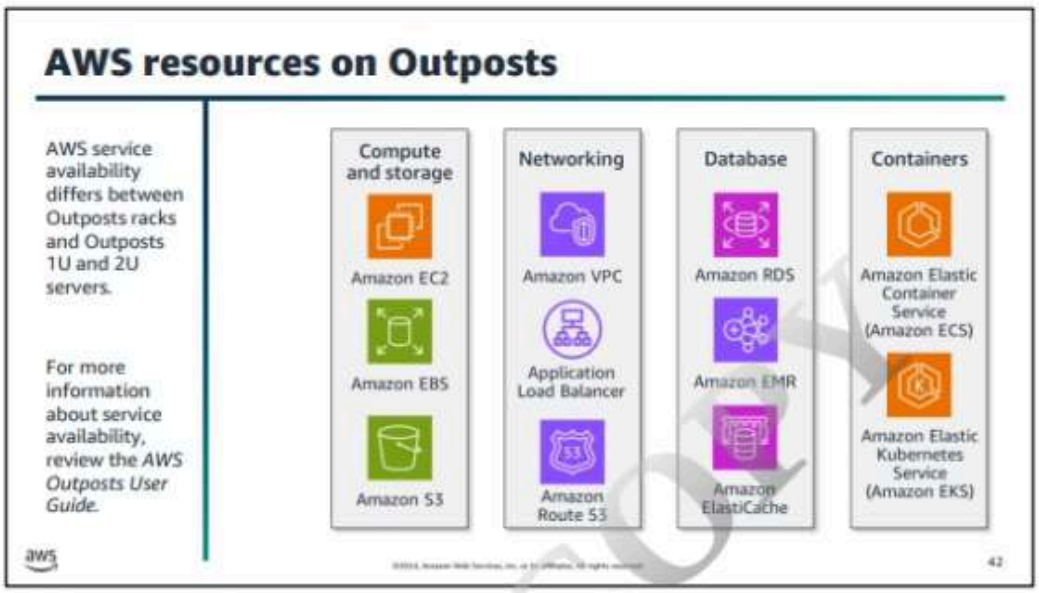
Outposts support multiple subnets. You choose the EC2 instance subnet when you launch the EC2 instance in your Outpost. You cannot choose the underlying hardware where the instance is deployed, because the Outpost is a pool of AWS compute and storage capacity.

Each Outpost can support multiple VPCs that can have one or more Outpost subnets. You create Outpost subnets from the VPC CIDR range where you created the Outpost. You can use the Outpost address ranges for resources, such as EC2 instances that reside in the Outpost subnet. AWS does not directly advertise the VPC CIDR, or the Outpost subnet range, to your on-premises location.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



You can create resources on your Outpost to support low-latency workloads that must run in close proximity to on-premises data and applications.

Some AWS resources that can be used on Outposts include:

- Amazon EC2
- Amazon EBS
- Amazon S3
- Amazon VPC
- Application load balancer
- Amazon Route 53
- Amazon Relational Database Service (Amazon RDS)
- Amazon EMR
- Amazon ElastiCache
- Amazon ECS
- Amazon EKS

For more information and the full list of services available on Outposts, see "AWS resources on Outposts" in the *AWS Outposts User Guide* at https://docs.aws.amazon.com/outposts/latest/userguide/what-is-outposts.html#services.

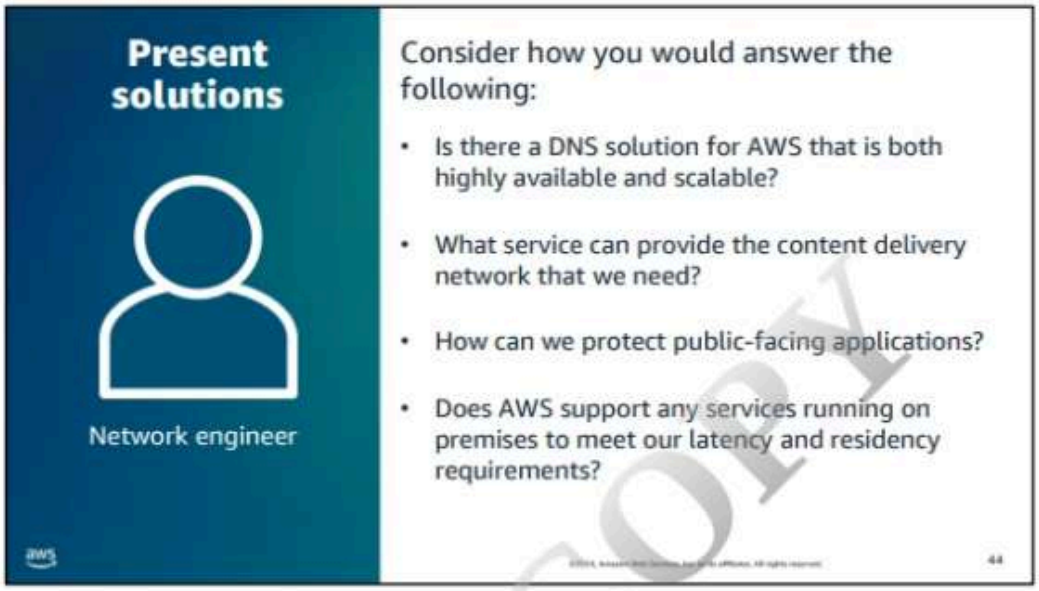
DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Imagine you are now ready to talk to the network engineer and present solutions that meet their architectural needs. Think about how you would answer the questions from the beginning of the lesson.

Your answers should include the following solutions:

- Use Amazon Route 53, a highly available and scalable cloud DNS web service. It is designed to give developers and businesses a reliable and cost-effective way to route end users to internet applications. Route 53 operates at AWS edge locations and provides you with multiple routing options based on your use case.
- The Amazon CloudFront service is a content delivery network (CDN) in the AWS Cloud. It caches your data in edge points of presence (PoP) worldwide.
- Services such as AWS WAF and AWS Shield protect you from DDoS and other attacks. Automated mitigation techniques are built into Shield Standard for Layer 4 and Layer 3 attacks, and AWS WAF protects the application layer (Layer 7) with rules and ACLs.
- You can request either an AWS Outposts rack or Outposts servers to meet your latency and residency needs. Choose the product that provides you with the services that you need. Consider cost optimization, and verify requirements before creating an Outposts site at your office or data center.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Module review

In this module, you learned about the following:

- ✓ Edge fundamentals
- ✓ Route 53
- ✓ Amazon CloudFront
- ✓ DDoS protection
- ✓ AWS Outposts

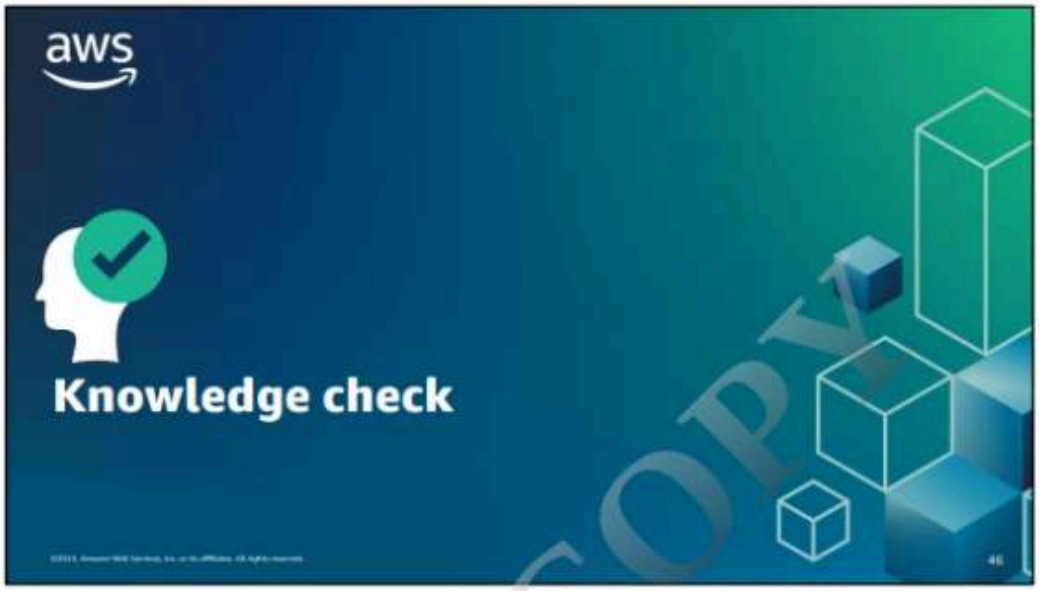
Next, you review the following:

- Knowledge check
- Lab introduction

45

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 1

What are the potential benefits of implementing a CloudFront distribution? (Select TWO.)

Choice	Response
A	Increased application security
B	Two global static IP addresses
C	Automatic redundancy for all application content
D	Reduced latency for access to application content
E	On-premises data caching

47

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 1 and answer

What are the potential benefits of implementing a CloudFront distribution? (Select TWO.)

Choice	Response
A correct	Increased application security
B	Two global static IP addresses
C	Automatic redundancy for all application content
D correct	Reduced latency for access to application content
E	On-premises data caching

The answers are A and D, increased application security and reduced latency for access to application content.

CloudFront has built-in integrations with AWS Shield and AWS WAF, providing DDoS protection. CloudFront is a global content delivery network (CDN) service that accelerates delivery of your websites, APIs, video content, or other web assets.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 2

Which of the following services offer DDoS protection? (Select TWO.)

Choice	Response
A	AWS Outposts
B	Amazon EC2
C	Network Load Balancer
D	AWS Shield
E	AWS WAF

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 2 and answer

Which of the following services offer DDoS protection? (Select TWO.)

Choice	Response
A	AWS Outposts
B	Amazon EC2
C	Network Load Balancer
D correct	AWS Shield
E correct	AWS WAF

The answers are D and E, AWS Shield and AWS WAF.

AWS Shield – AWS provides AWS Shield Standard and AWS Shield Advanced for protection against DDoS attacks. AWS Shield Standard is automatically included at no extra cost beyond what you already pay for AWS WAF and your other AWS services.

For more information, see "AWS Shield" in the *AWS WAF, AWS Firewall Manager, and AWS Shield Advanced Developer Guide* at https://docs.aws.amazon.com/waf/latest/developerguide/shield-chapter.html.

AWS WAF – Use AWS WAF web ACLs to help minimize the effects of a DDoS attack.

For more information, see "AWS WAF" in the *AWS WAF, AWS Firewall Manager, and AWS Shield Advanced Developer Guide* at https://docs.aws.amazon.com/waf/latest/developerguide/waf-chapter.html.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 3

A network engineer wants to route 80 percent of web traffic to the ap-southeast-2 Region. The remaining 20 percent of traffic will be directed to the eu-west-1 Region. Which Route 53 routing policy is the best choice for this use case?

Choice	Response
A	Simple routing
B	Weighted routing
C	Geoproximity routing
D	Geolocation routing
E	Multivalue answer routing

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 3 and answer

A network engineer wants to route 80 percent of web traffic to the ap-southeast-2 Region. The remaining 20 percent of traffic will be directed to the eu-west-1 Region. Which Route 53 routing policy is the best choice for this use case?

Choice	Response
A	Simple routing
B correct	Weighted routing
C	Geoproximity routing
D	Geolocation routing
E	Multivalue answer routing

The answer is B, weighted routing.

With weighted routing, you can assign weights to a resource record set to specify the frequency with which different responses are served.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 4

What is a benefit of choosing Outposts servers when compared to an Outposts rack?

Choice	Response
A	More AWS services are available on Outposts servers.
B	You can pool compute and storage capacity between multiple Outposts servers.
C	A rack-mountable server is available in a location with space constraints.
D	You don't need to host your Outposts servers in a data center.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 4 and answer

What is a benefit of choosing Outposts servers when compared to an Outposts rack?

Choice	Response
A	More AWS services are available on Outposts servers.
B	You can pool compute and storage capacity between multiple Outposts servers.
C correct	A rack-mountable server is available in a location with space constraints.
D	You don't need to host your Outposts servers in a data center.

The answer is C, A rack-mountable server is available in a location with space constraints.

AWS will deliver Outposts servers directly to you, and you can either have your onsite personnel or an AWS preferred third-party contractor install them. After the Outposts servers are connected to your network, AWS will remotely provision compute and storage resources so you can start launching applications.

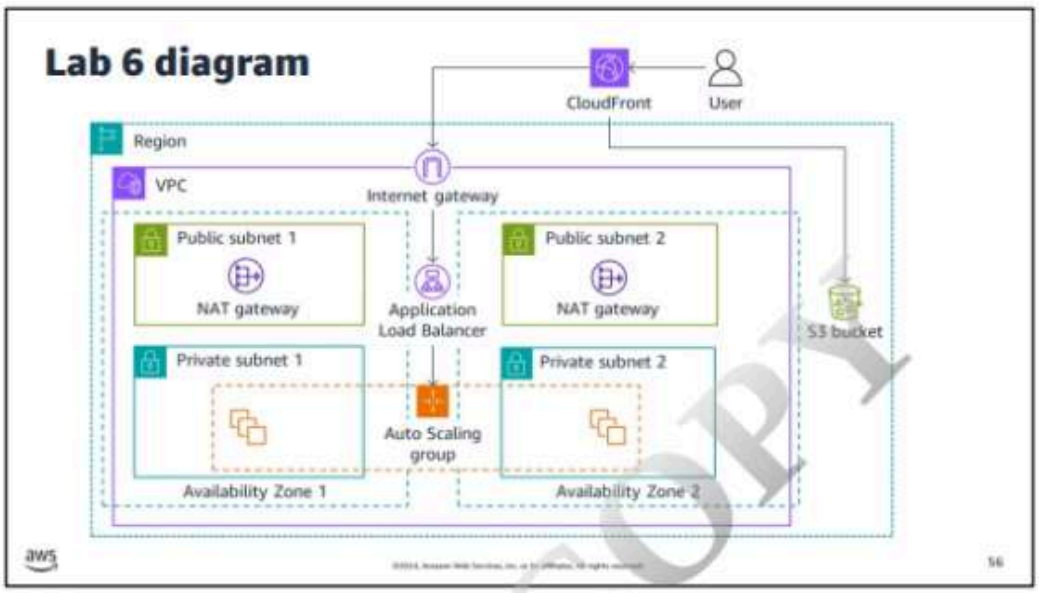
DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Lab 6 uses one Region and one VPC. There is a public and private subnet in each of two Availability Zones. Each public subnet has a NAT gateway. A user outside the Region interacts with CloudFront distributions. One distribution delivers content from an Auto Scaling group through an Application Load Balancer. The group's EC2 instances are launched in two private subnets. A second CloudFront distribution delivers content with an origin of an S3 bucket.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Lab tasks

Task 1: Explore an existing CloudFront distribution.

Task 2: Create an Amazon S3 bucket.

Task 3: Configure the Amazon S3 *LabBucket* for public access.

Task 4: Upload an object into the bucket and test the public access.

Task 5: Add the bucket as an additional origin to the distribution.

Task 6: Secure the bucket with a CloudFront origin access identity.

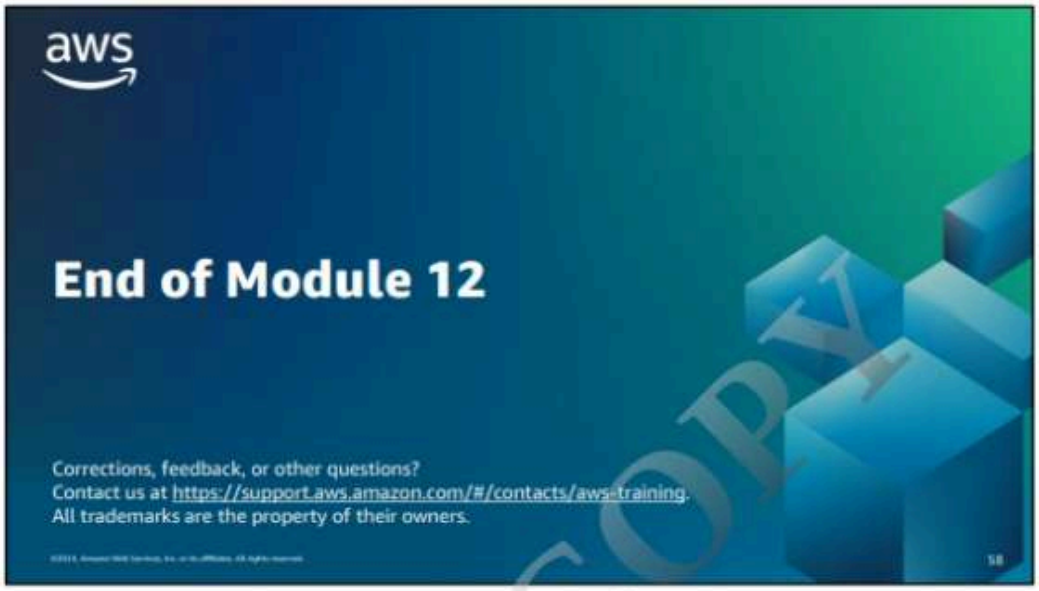
Task 7: Test direct access to the file in the bucket using the S3 URL.

Task 8: Test access to the object in the bucket using the CloudFront distribution.

57

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Poll question

Which backup and recovery components do you work with?

A. Offsite backup storage

B. Colocation facilities

C. Tape retrievals

D. On-premises backup

E. All of the above

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Module overview

- Business request
- Disaster planning
- AWS Backup
- Recovery strategies
- Present solutions
- Knowledge check
- Capstone lab: Build an AWS Multi-Tier Architecture

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Imagine you are meeting with the chief technology officer, who is creating a backup and disaster recovery (DR) plan. They are interested in learning more about the tools and strategies available in the cloud, which will protect data and minimize downtime. During this module, you learn about topics that answer these questions.

At the end of this module, you meet with the chief technology officer and present some solutions.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

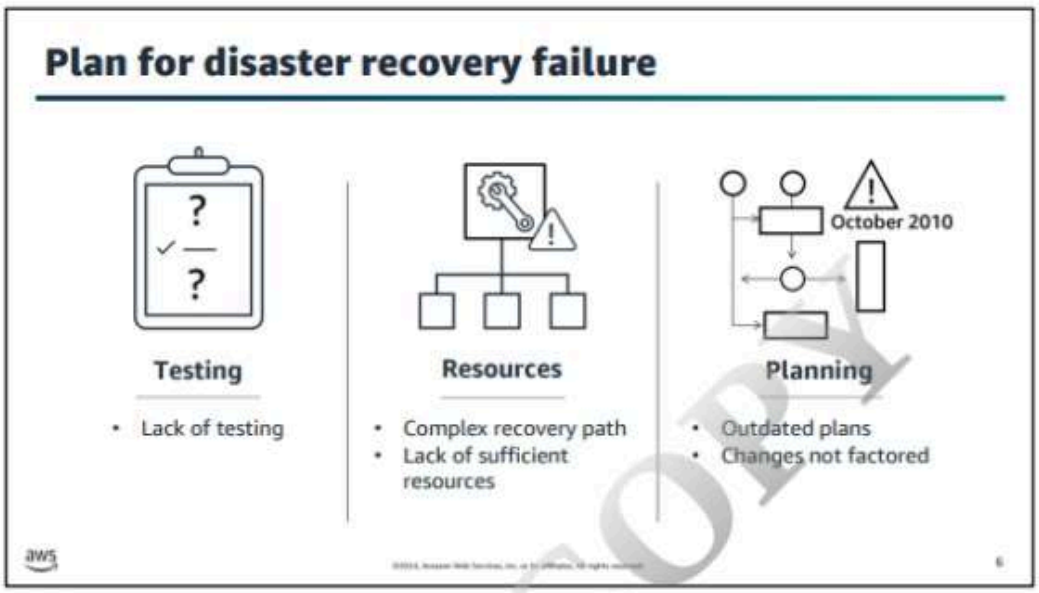
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The chief technology officer asks, "What strategies can we use to protect ourselves in the event of a disaster?" With many resources moving to the cloud, the company needs to better understand the DR features and strategies available for the AWS services they will be using. The company is asking you to summarize these features.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Not all disaster recovery (DR) plans are created equal, and many fail. Testing, resources, and planning are vital components of a successful DR plan.

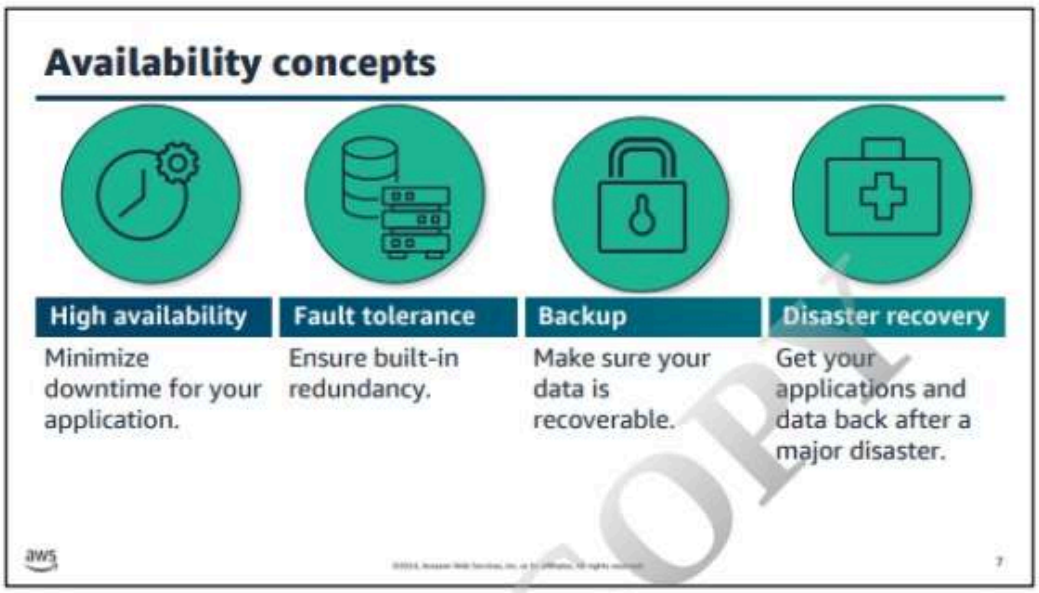
Testing
Test your DR plan to validate the implementation. Regularly test failover to your workload's DR Region to verify that you are meeting recovery objectives. Avoid developing recovery paths that you rarely run. For example, you might have a secondary data store that is used for read-only queries. Plan to fail over to your secondary data store when your primary fails.

Resources
Regularly run your recovery path in production. This will validate the recovery path and help you verify that resources are sufficient for operation throughout the event. Changes in your configuration in production must be mirrored in your DR network.

Planning
The only recovery that works is the path you test frequently. The capacity of the secondary resources, which might have been sufficient when you last tested, might no longer be able to tolerate your load. This is why it is best to have a small number of recovery paths. Establish recovery patterns, and regularly test them.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Production systems typically come with defined or implicit objectives in terms of uptime. A system is highly available when downtime is minimized in the event of a failure. Downtime is not eliminated, but it should be reduced to seconds or minutes.

High availability minimizes downtime and cost. It protects against failure. It keeps your business running with very little downtime and with fast recovery at low cost.

Fault tolerance is often confused with high availability, but fault tolerance refers to the built-in redundancy of an application's components to prevent service interruption. However, relying on fault tolerance does come at a higher cost than relying on only high availability.

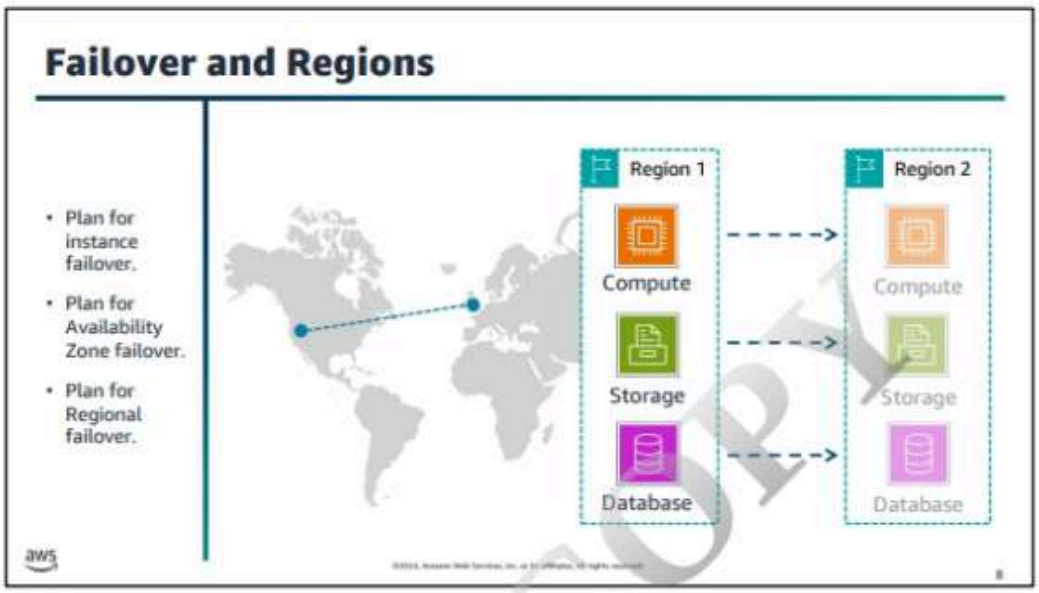
Backup is critical to protecting data and providing business continuity. At the same time, it can be a challenge to implement well. The pace at which data is generated is growing exponentially. The density and durability of local disks is not benefiting from the same growth rate. Enterprise backup has become its own industry.

Disaster recovery is often overlooked as a component of availability. If a natural disaster makes one or more of your components unavailable or destroys your primary data source, can you restore service quickly and without lost data?

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS is available in multiple Regions around the globe. You can choose the most appropriate location for your DR site, in addition to the site where your system is fully deployed. It is highly unlikely for a Region to be unavailable. However, it is possible for a Region to become unavailable if a very large-scale event—for instance, a natural disaster—impacts the Region.

AWS maintains a page that inventories current products and services offered by Region. AWS maintains a strict Region isolation policy so that any large-scale event in one Region will not impact any other Region. We encourage our customers to take a similar multi-Region approach to their strategy. Each Region should be able to be taken offline with no impact to any other Region.

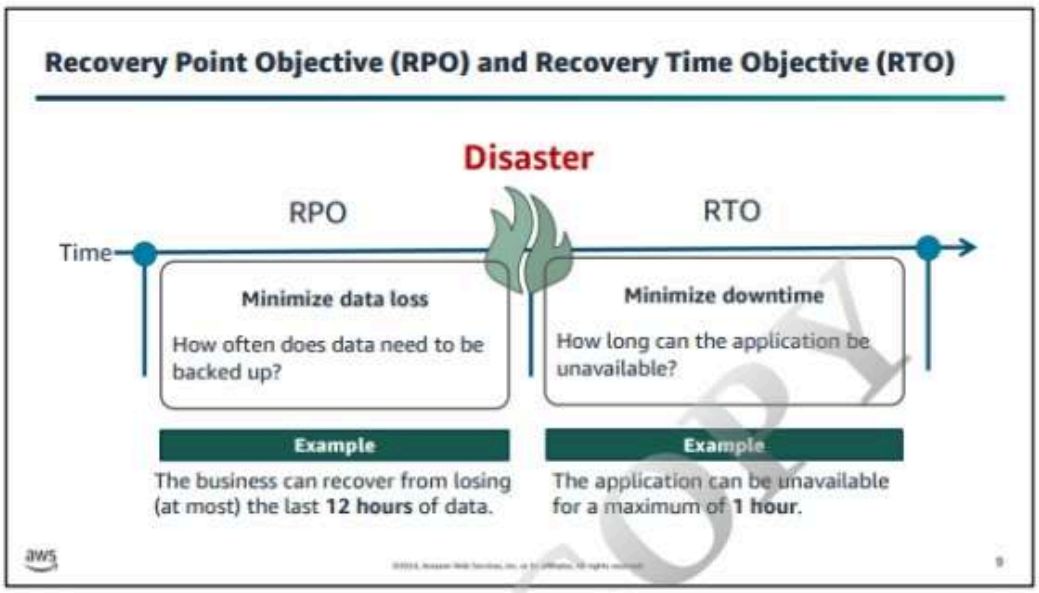
If you have an AWS Direct Connect circuit to any AWS Region in the United States (US), it will provide you with access to all Regions in the US, including AWS GovCloud (US), without that traffic going through the public internet. Also, consider how you deploy applications. If you deploy to each Region separately, you can isolate that Region in case of a disaster and transfer all of your traffic to another Region.

If you are deploying new applications and infrastructure rapidly, you might want to have an active-active Region. Imagine you deploy something that causes a Region's applications to misbehave or to be unavailable. You can remove the Region from the active record set in Amazon Route 53, identify the root cause, and roll back the change before reactivating the Region.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Recovery Point Objective (RPO) is the acceptable amount of data loss measured in time. For example, if a disaster occurs at 1 PM (13:00) and the RPO is 12 hours, the system should recover all data that was in the system before 1 AM (01:00) that day. At most, data loss will span 12 hours—between 1 PM and 1 AM.

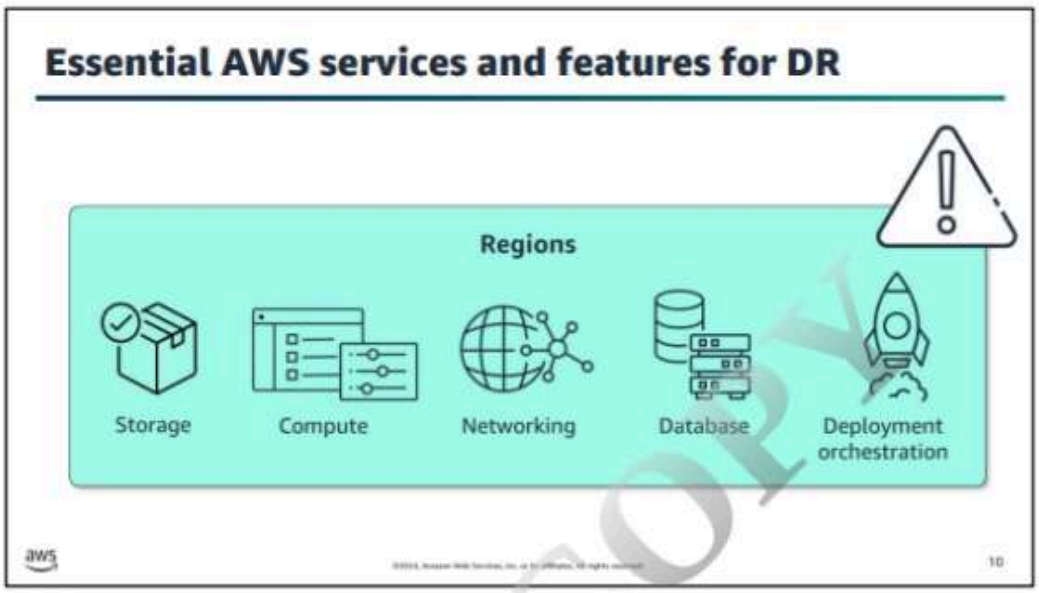
Recovery Time Objective (RTO) is the time it takes after a disruption to restore a business process to its service level, as defined by the operational level agreement (OLA). For example, if a disaster occurs at 1 PM (13:00) and the RTO is 1 hour, the DR process should restore the business process to the acceptable service level by 2 PM (14:00).

A company typically decides on an acceptable RPO and RTO based on the financial impact to the business when systems are unavailable. The company determines financial impact by considering many factors, such as the loss of business and damage to its reputation because of downtime and the lack of systems availability. IT organizations plan solutions to provide cost-effective system recovery based on the RPO within the timeline and the service level established by the RTO.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Before discussing the various approaches to DR, it is important to review the AWS services and features that are the most relevant to it. This section provides a summary.

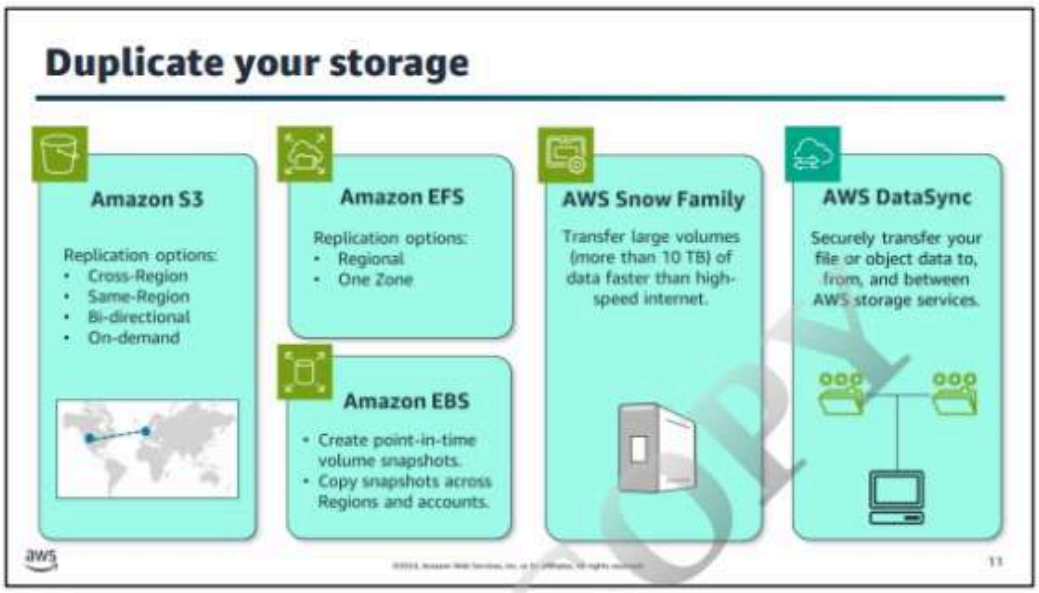
When planning for DR, consider the services and features that support data migration and durable storage. For some of the scenarios that involve either a scaled-down or a fully scaled deployment of your system in AWS, compute resources will also be required.

During a disaster, you need to either provision new resources or fail over to existing preconfigured resources. These resources include code and content. But they can also include other pieces, such as DNS entries, network firewall rules, and virtual machines or instances.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



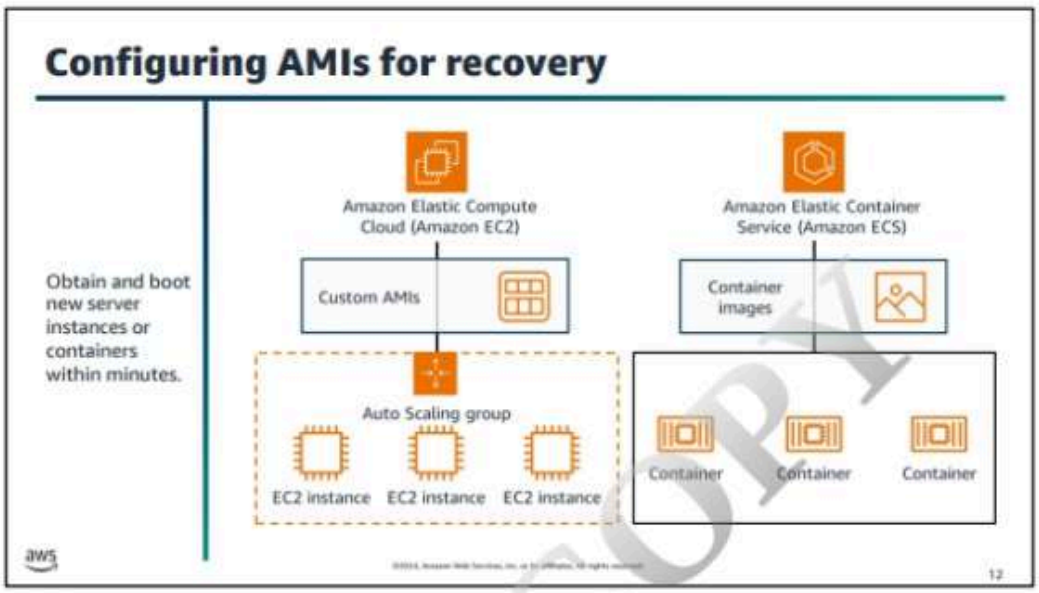
When building your backup strategy, make a plan to duplicate your storage by using a solution that meets your recovery needs.

AWS offers the following options for duplicating data in storage:

- Amazon Simple Storage Service (Amazon S3) provides highly durable storage designed for mission-critical and primary data storage. Amazon S3 offers several options for replicating objects:
  - Live replication to automatically replicate new and updated objects
    - Cross-Region Replication (CRR)
    - Same-Region Replication (SRR)
    - Bi-directional replication
  - On-demand replication to replicate existing objects
- Amazon Elastic File System (Amazon EFS) provides serverless, fully elastic file storage so that you can share file data without provisioning or managing storage capacity and performance. Amazon EFS offers the following options for replicating file systems:
  - Regional: Replicates data and metadata redundantly across all Availability Zones within the AWS Region
  - One Zone: Replicates data and metadata redundantly within a single Availability Zone
- Amazon Elastic Block Store (Amazon EBS) gives you the ability to create point-in-time snapshots of data volumes. When the snapshot status is *Completed*, you can copy it from one AWS Region to another, or within the same Region.
- The AWS Snow Family is a family of data transport solutions that move terabytes (TB) to petabytes of data into and out of AWS by using storage devices. It is designed to be secure for physical transport. Snow Family devices help you retrieve a large quantity of data stored in Amazon S3 much quicker than using high-speed internet.
- Use AWS DataSync to securely transfer your file or object data to, from, and between AWS storage services. DataSync copies files over the internet or a Direct Connect connection.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



In the context of DR, it's critical to be able to rapidly create virtual machines that you control. By launching instances in separate Availability Zones, you can protect your applications from the failure of a single location.

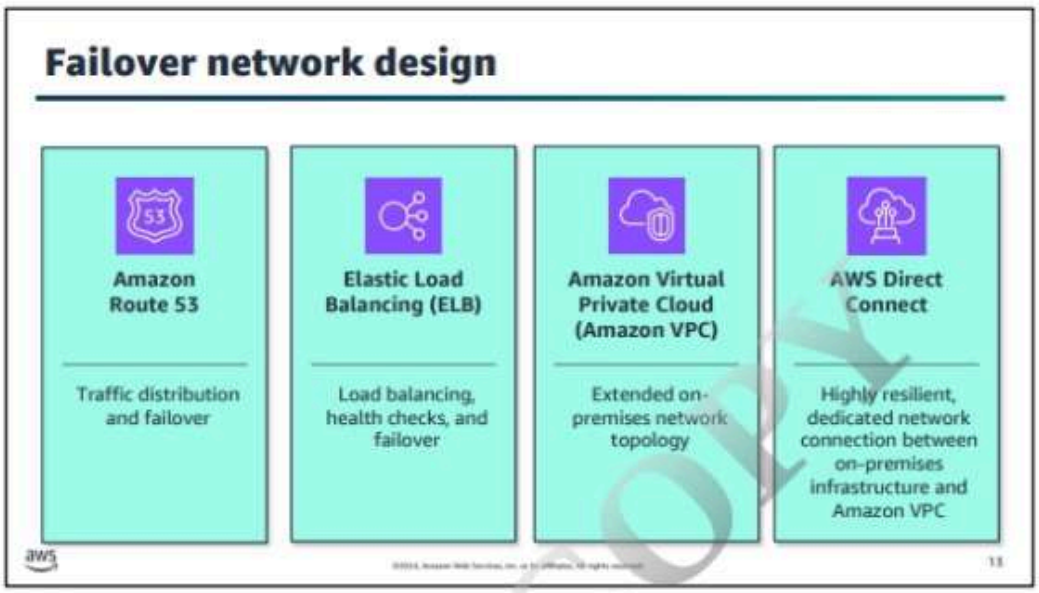
You can arrange for automatic recovery of an Amazon Elastic Compute Cloud (Amazon EC2) instance when a system status check of the underlying hardware fails. The instance will be rebooted (on new hardware if necessary). But it will retain its instance ID, IP address, Elastic IP addresses, Amazon EBS volume attachments, and other configuration details. For the recovery to be complete, you need to make sure that the instance automatically starts up any services or applications as part of its initialization process.

Amazon Machine Images (AMIs) are preconfigured with operating systems. Some preconfigured AMIs also include application stacks. You can also configure your own AMIs. In the context of DR, AWS strongly recommends that you configure and identify your own AMIs so that they can launch as part of your recovery procedure. Your AMIs should be preconfigured with your operating system of choice and the appropriate pieces of the application stack.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



When you are dealing with a disaster, you might need to fail over to another site. This requires modifying network settings to initiate the failover. It might also require adjusting network settings within the failover site. AWS offers several services and features that you can use to manage and modify network settings:

- Route 53 includes a number of global load balancing capabilities. This can be effective when you are dealing with DNS endpoint health checks, the ability to failover between multiple endpoints, and static websites hosted in Amazon S3.
- Elastic Load Balancing (ELB) automatically distributes incoming application traffic across multiple EC2 instances. You achieve greater fault tolerance in your applications by ELB providing the load balancing capacity that is needed in response to incoming application traffic. You can pre-allocate your load balancer to identify its DNS name and simplify running your DR plan. Route 53 and ELB are highly available resources by default.
- In the context of DR, you can use Amazon Virtual Private Cloud (Amazon VPC) to extend your existing network topology to the cloud. This can be especially appropriate when recovering enterprise applications that are typically on the internal network.
- Set up a dedicated network connection from your premises to AWS with Direct Connect. This can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience than internet-based connections.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Database backup and replicas

Amazon Relational Database Service (Amazon RDS)

- Take a snapshot of data and save it in a separate Region.
- Use Multi-AZ cluster deployment to build a resilient DR strategy by using standby instances and read replicas.
- Retain automated backups.

Amazon DynamoDB

- Back up full tables in seconds.
- Use point-in-time recovery to continuously back up tables for up to 35 days.
- Initiate backups with a single click in the console or a single API call.
- Use global tables for fast local performance for globally distributed applications.

14

With Amazon Relational Database Service (Amazon RDS), you can plan for high availability by using the following methods:

- Create a Multi-AZ database (DB) instance deployment, which creates a primary instance and a standby instance to provide failover support. However, the standby instance does not serve traffic.
- Create a Multi-AZ DB cluster deployment, which creates two standby instances that can also serve read traffic.
- Use a snapshot to create a read replica in a different Region. This replica can be promoted to primary in the event of disaster.
  - Promoting a read replica requires a reboot and has a higher RTO than failing over to a standby instance.
- Save a manual DB snapshot or a DB cluster snapshot in a separate Region.
- Share a manual snapshot with up to 20 other AWS accounts.

Amazon RDS read replicas for MySQL and MariaDB now support Multi-AZ deployments. By running a secondary copy of a database, you can build a resilient DR strategy and simplify your database engine upgrade process.

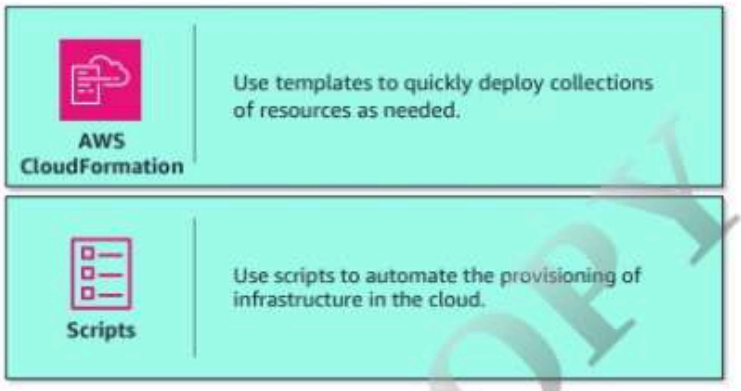
Global tables build on the Amazon DynamoDB global footprint to provide you with the following:

- A fully managed, multi-Region, and multi-active database that replicates your DynamoDB tables.
- A replica of your DynamoDB tables across your choice of AWS Regions. Unlike read replicas and secondary instances, DynamoDB treats all replicas in a global table as a single unit, with every table having the same table name and primary key. If one replica fails, you can access data by using replicas in other Regions.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Templates and scripts



Time is valuable in DR. Use AWS CloudFormation templates and scripting to deploy your resources efficiently. Use CloudFormation to define your infrastructure and deploy it consistently across AWS accounts and Regions. CloudFormation uses predefined pseudo-parameters to identify the AWS account and AWS Region in which it is deployed. You can implement condition logic in your CloudFormation templates to deploy the scaled-down version of your infrastructure in the DR Region.

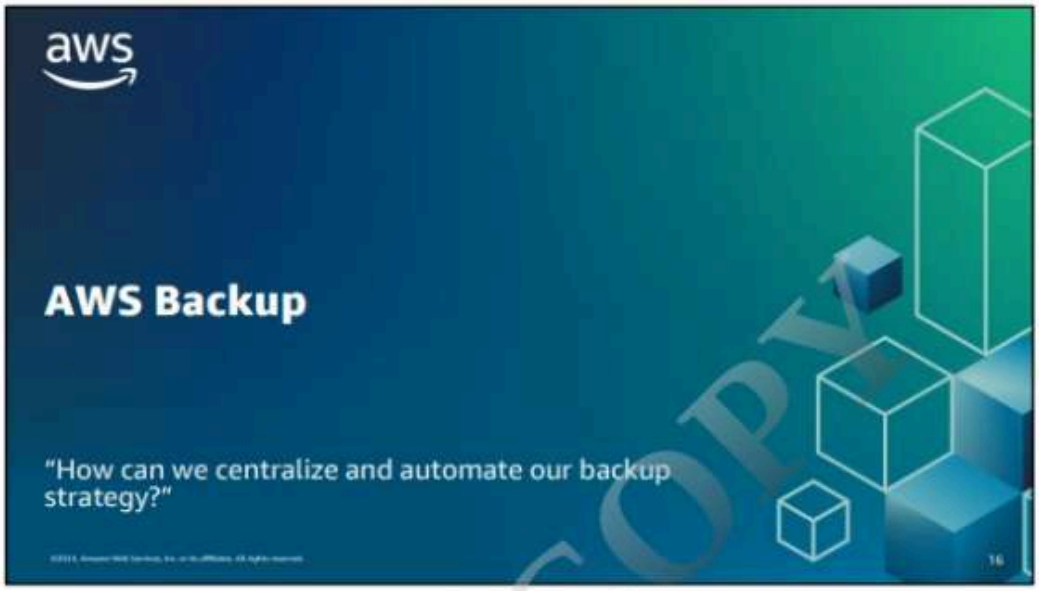
For EC2 instance deployments, an AMI supplies information about hardware configuration and installed software. You can implement an Amazon EC2 Image Builder pipeline that creates the AMIs you need. Copy these to both your primary and backup Regions to provide everything you need to redeploy or scale out your workload in a new Region.

To start EC2 instances or provision other AWS resources, you can create scripts by using AWS Command Line Interface (AWS CLI) or AWS SDK.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

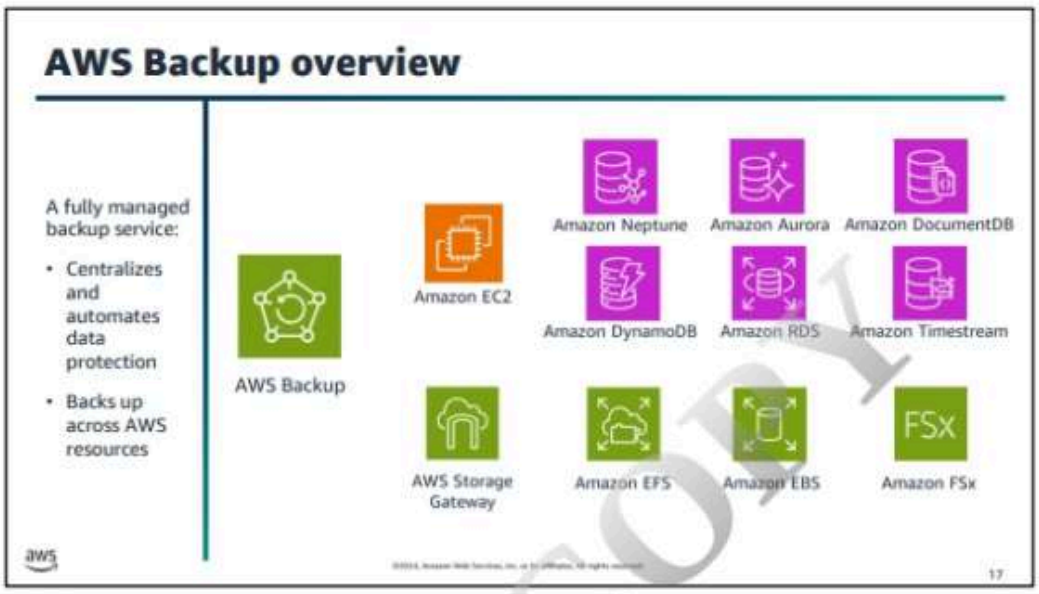


The chief technology officer asks, "How can we centralize and automate our backup strategy?" To increase the efficiency and reliability of their backup strategy, the company would like to manage backups from a single location. The company is asking you to propose a solution that meets this requirement.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS Backup is a fully managed backup service that helps you centralize and automate the backup of data across AWS services. AWS Backup also helps customers support their regulatory compliance obligations and meet business continuity goals.

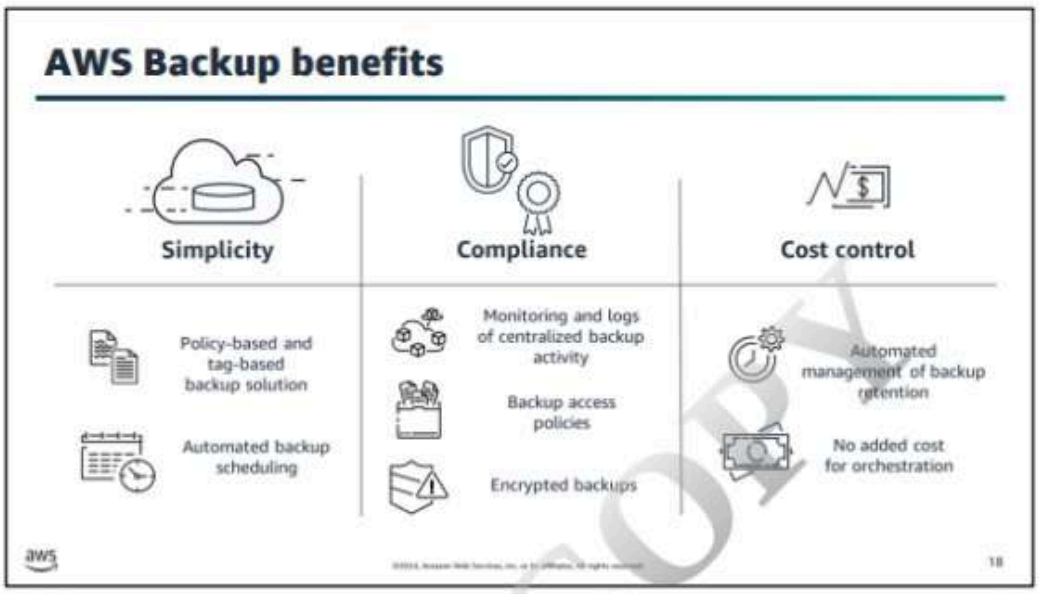
AWS Backup works with AWS Organizations. AWS Backup centrally deploys data protection policies to configure, manage, and govern your backup activity. It works across your AWS accounts and resources.

For a full list of AWS services supported by AWS Backup, see "Supported AWS Resources and Applications" at https://docs.aws.amazon.com/aws-backup/latest/devguide/whatisbackup.html#supported-resources.

For more information about managing backups, see "AWS Backup – Automate and Centrally Manage Your Backups" in the *AWS News Blog* at https://aws.amazon.com/blogs/aws/aws-backup-automate-and-centrally-manage-your-backups/.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Simplicity: AWS simplifies your backup plan. Unlike other tools, AWS Backup does not need custom backup scripts. It provides a central place to manage and monitor your backups. The AWS Backup plan is a set of rules that define your backup. The rules include when to start the backups, the duration of the backup window, and the retention period. One of the core capabilities of AWS Backup is the ability to use tags to affect the backup of resources you'd like to protect.

Compliance: You can enforce backup policies, encrypt your backups, protect your backups from manual deletion, and prevent changes to your backup lifecycle settings. You can use the consolidated logs of backup activity across AWS services to perform compliance audits. AWS Backup complies with Payment Card Industry (PCI) and International Organization for Standardization (ISO) standards and is eligible under Health Insurance Portability and Accountability Act (HIPAA).

Cost control: AWS Backup reduces the risk of downtime, which can negatively impact your business. You can also reduce operating costs by reducing time spent on manual configuration and by automating backups.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Understanding how AWS Backup works

You implement your backup strategy with AWS Backup by creating backup plans.

1. Create AWS Backup plan.
- Schedule
- Lifecycle
- Vault
- Tags for the backups

2. Assign resources.
- Select the resources that will use this plan.
  - Assign tags.
  - Use resource IDs.
- Specify the AWS Identity and Access Management (IAM) role.

3. Manage and monitor backups.
- AWS Backup automatically runs the backup plan.
- Centrally manage backups.
- Centrally monitor backup activity.

AWS Backup works with AWS Organizations to manage backup policies across AWS accounts.

19

When you create a backup plan, you specify the following:
- Schedule: Set the frequency of the backups and the window of time during which to conduct backups.
- Lifecycle: Determine when a backup is moved to cold storage and when a backup expires.
- Vault: AWS Backup keeps backups in an AWS Backup vault. You specify which backup vault your backup plan uses.
  - When you create a backup vault, you assign an AWS Key Management Service (AWS KMS) encryption key to encrypt backups that do not have their own encryption methods.
- Tags for backup: You specify tags that will be assigned to backups created by this plan.

You assign the resources to back up with your backup plan. You also define which AWS Identity and Access Management (IAM) role AWS Backup will use to gain the needed access to these resources. You can assign resources by using the following methods:
- Tags: You can provide a tag value. All AWS resources with that tag will be backed up by using this plan.
- Resource IDs: You can use resource IDs for specific resources, such as a specific DynamoDB table or EBS volume.

AWS Backup works with other AWS services, such as Amazon CloudWatch, Amazon EventBridge, AWS CloudTrail, and Amazon Simple Notification Service (Amazon SNS), to monitor its workloads.

For more information about using these services to monitor activity in AWS Backups, see "Monitoring AWS Backup" in the *AWS Backup Developer Guide* at https://docs.aws.amazon.com/aws-backup/latest/devguide/monitoring.html.

For more information about managing backup policies across accounts with AWS Organizations, see "Managing AWS Backup Resources Across Multiple AWS Accounts" in the *AWS Backup Developer Guide* at https://docs.aws.amazon.com/aws-backup/latest/devguide/manage-cross-account.html.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The chief technology officer asks, "Which disaster recovery strategy minimizes downtime but is also cost-effective?" The company is exploring configurations that will deliver business continuity and minimize downtime in the event of an error. Some of these solutions are more expensive than others. The company is asking you to summarize these strategies and their relative cost.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

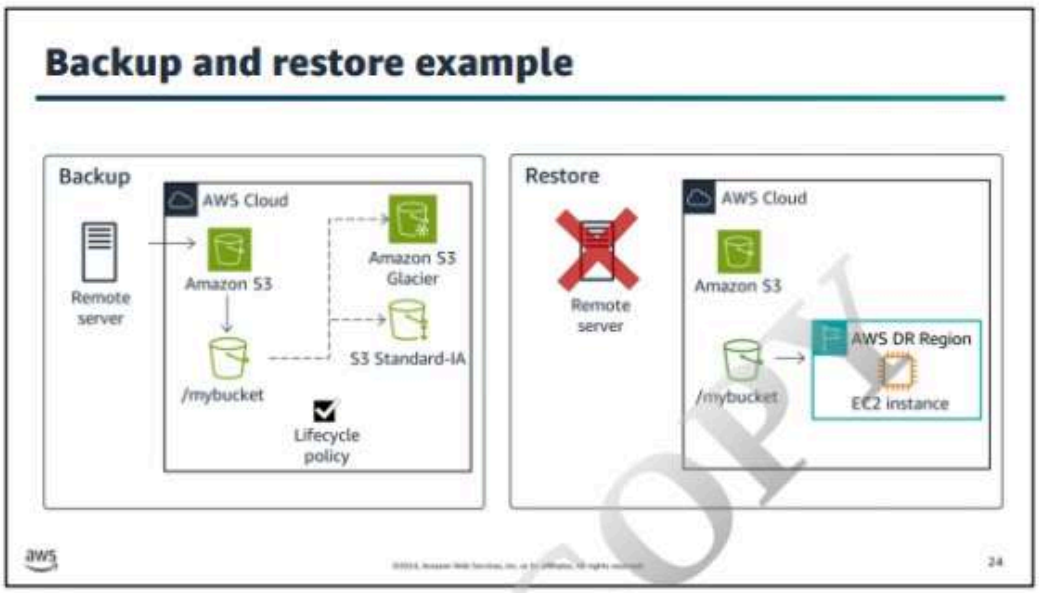
This section outlines four DR scenarios that highlight the use of AWS and then compares AWS with traditional DR methods (sorted from highest to lowest RTO and RPO) as follows:

1. Backup and restore
2. Pilot light
3. Warm standby
4. Multi-site active/active

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



In most traditional environments, data is backed up to tape and sent offsite regularly. If you use this method, it can take a long time to restore your system in the event of a disruption.

Amazon S3 is an ideal destination for quick access to your backup. You typically transfer data to and from Amazon S3 through the network, and it is therefore accessible from any location. You can also use a lifecycle policy to move older backups to progressively more cost-efficient storage classes over time.

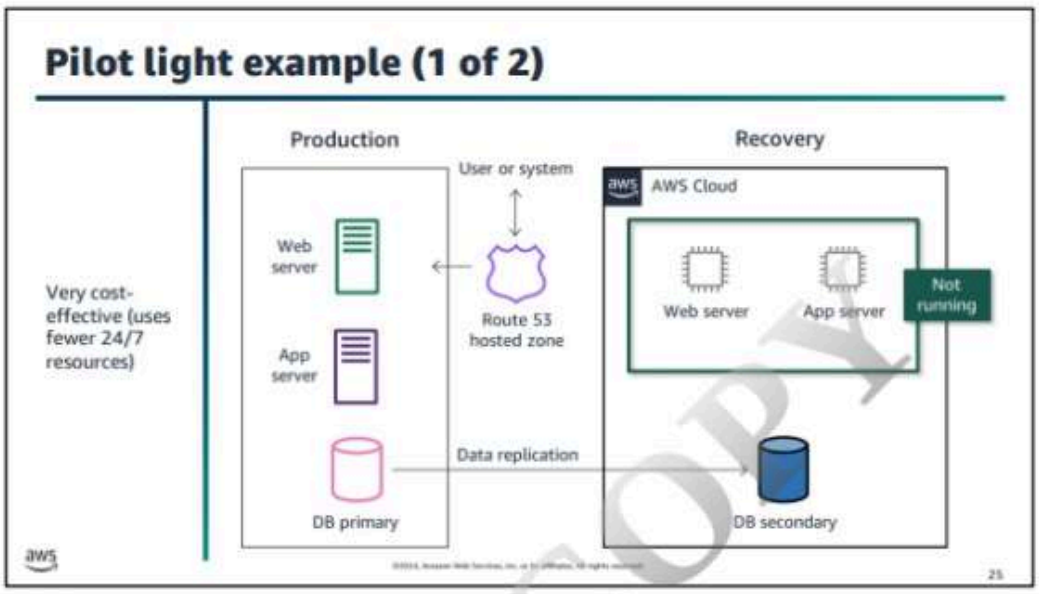
In this example, data from a remote server is backed up to Amazon S3 into /mybucket. A lifecycle policy moves each backup to S3 Standard-Infrequent Access (S3 Standard-IA) once it has been in /mybucket for some time. Later, the backup is moved to S3 Glacier.

If the remote server fails, you can restore services by deploying a disaster recovery VPC. Use CloudFormation to automate the deployment of core networking. Create an EC2 instance by using an AMI that matches your remote server. Then restore your systems by retrieving your backups from Amazon S3. You then adjust DNS records to point to AWS.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



With the *pilot light* approach, you replicate your data from one environment to another and provision a copy of your core workload infrastructure. In this example, you replicate the production web server, application server, and database in a recovery environment. Route 53 routes all traffic to the production environment.

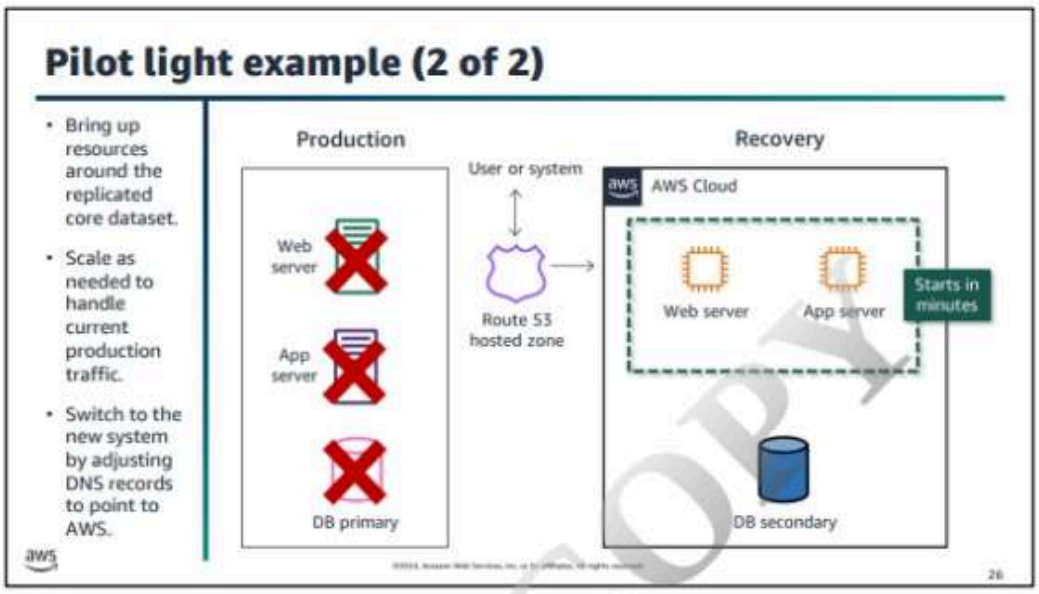
Resources required to support data replication and backup, such as databases and object storage, are always on. Other elements, such as application servers, are loaded with application code and configurations but are switched off. These elements are used only during testing or when disaster recovery failover is invoked. Unlike the backup and restore approach, your core infrastructure is always available. You always have the option to quickly provision a full-scale production environment by switching on and scaling out your application servers.

The pilot light architecture is relatively inexpensive to implement. In the preparation phase of DR, it is important to consider the use of services and features that support data migration and durable storage.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



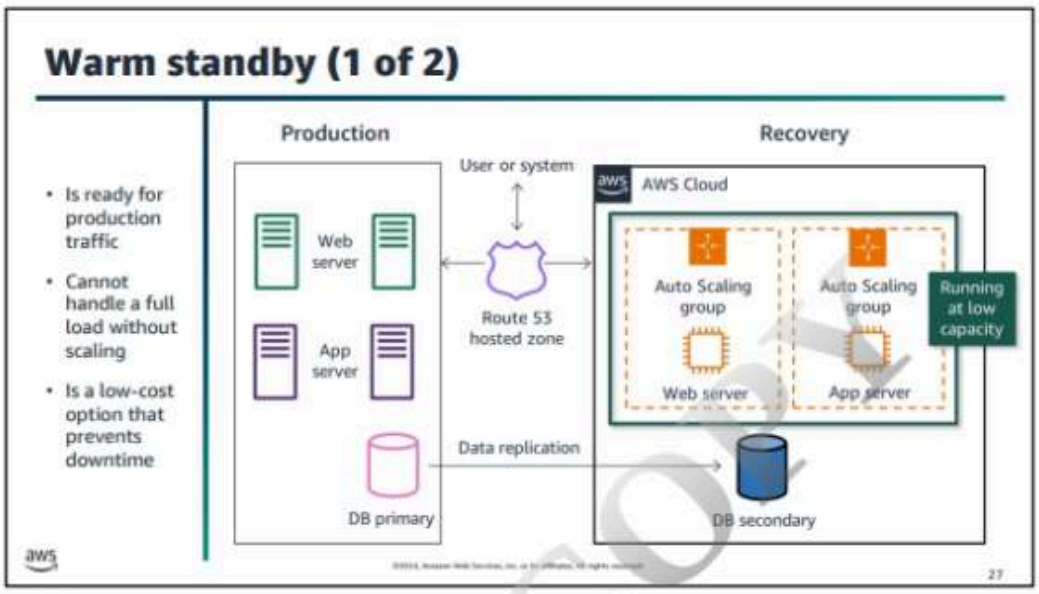
When disaster strikes, the servers in the recovery environment start up. Route 53 then begins sending production traffic to the recovery environment. The essential infrastructure pieces include DNS, networking features, and various Amazon EC2 features.

For more information about pilot light solutions, see "Disaster Recovery (DR) Architecture on AWS, Part III: Pilot Light and Warm Standby" in the *AWS Architecture Blog* at https://aws.amazon.com/blogs/architecture/disaster-recovery-dr-architecture-on-aws-part-iii-pilot-light-and-warm-standby/.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



The warm standby approach involves creating a scaled down but fully functional copy of your production environment in a recovery environment. By identifying your business-critical systems, you can fully duplicate these systems on AWS and have them always on. This decreases the time to recovery because you do not have to wait for resources in the recovery environment to start up.

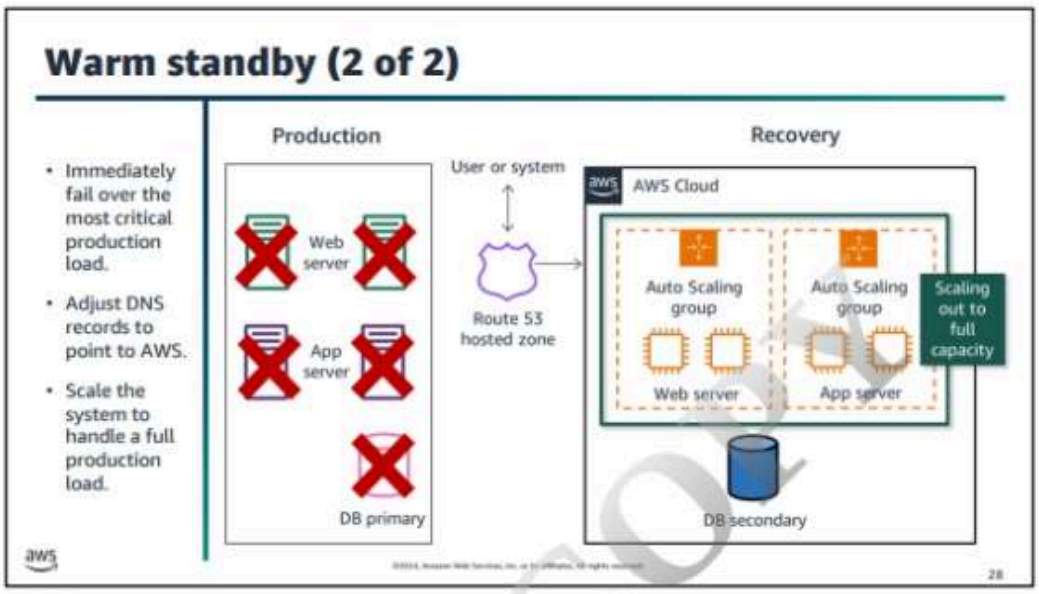
In this example, you replicate the web servers and application servers in the recovery environment as Auto Scaling groups. The Auto Scaling groups can run the minimum number of instances and smallest EC2 instance sizes possible. This solution is not scaled to take a full production load, but it is fully functional. You can use it for nonproduction work, such as testing, quality assurance, and internal use. Use Route 53 to distribute requests between the production and recovery environments. You copy the database into the recovery environment and use data replication to keep it current.

In the diagram, there are two systems running: the production environment and a low-capacity system running in the recovery environment. Warm standby in the recovery environment requires that all necessary components run continuously but are not scaled for production traffic. It is a best practice to do continuous testing by using a subset of production traffic to the DR site.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

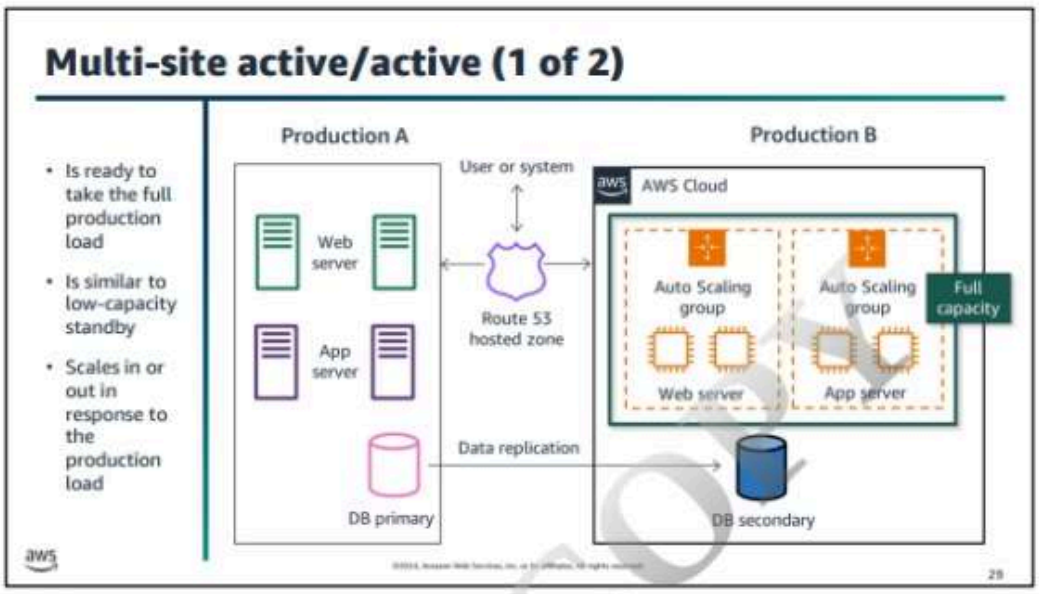


If the production environment is unavailable, Route 53 switches over to the recovery environment. The recovery environment automatically scales its capacity out in the event of a failover from the primary system. For your critical loads, warm standby RTO is as long as it takes to fail over. For all other loads, it takes as long as it takes you to scale up. The RPO depends on the replication type.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



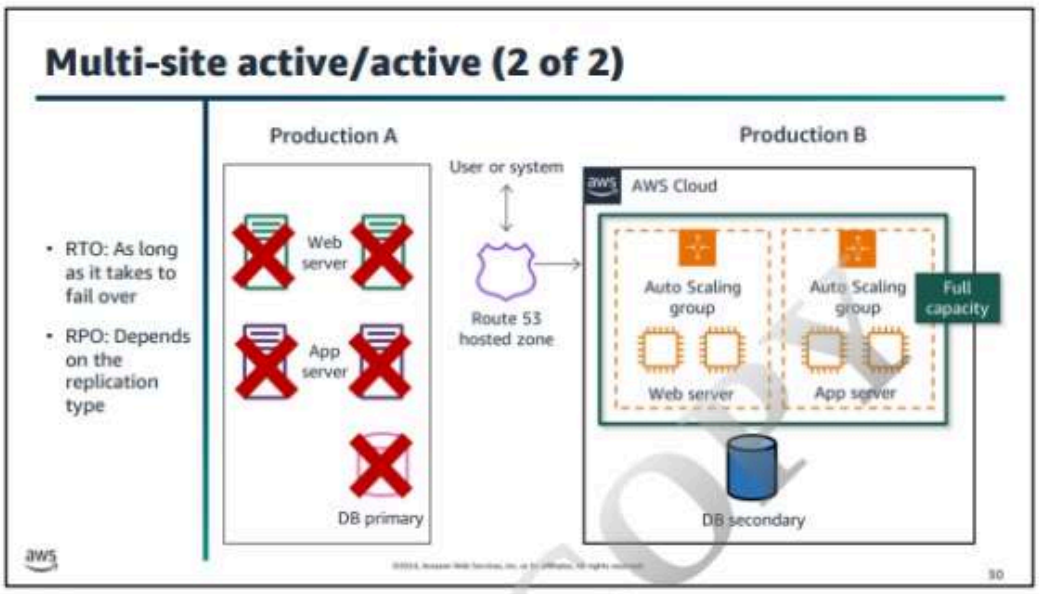
In an active/active configuration, a multi-site solution runs in two environments. In this example, both the Production A and Production B environments run web servers, application servers, and databases to handle production traffic. Route 53 routes traffic between the two environments.

This architecture potentially has the least amount of downtime. It has more costs associated with it because more environments are running. You can use a DNS service that supports weighted routing, such as Route 53, to route production traffic to different sites that deliver the same application or service.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



In a disaster situation in the Production A environment, you can adjust the DNS weighting and send all traffic to the Production B environment. The capacity of the AWS service can be rapidly increased to handle the full production load. You can use Amazon EC2 Auto Scaling to automate this process. You might need some application logic to detect the failure of the primary database services and cut over to the parallel database services running in AWS.

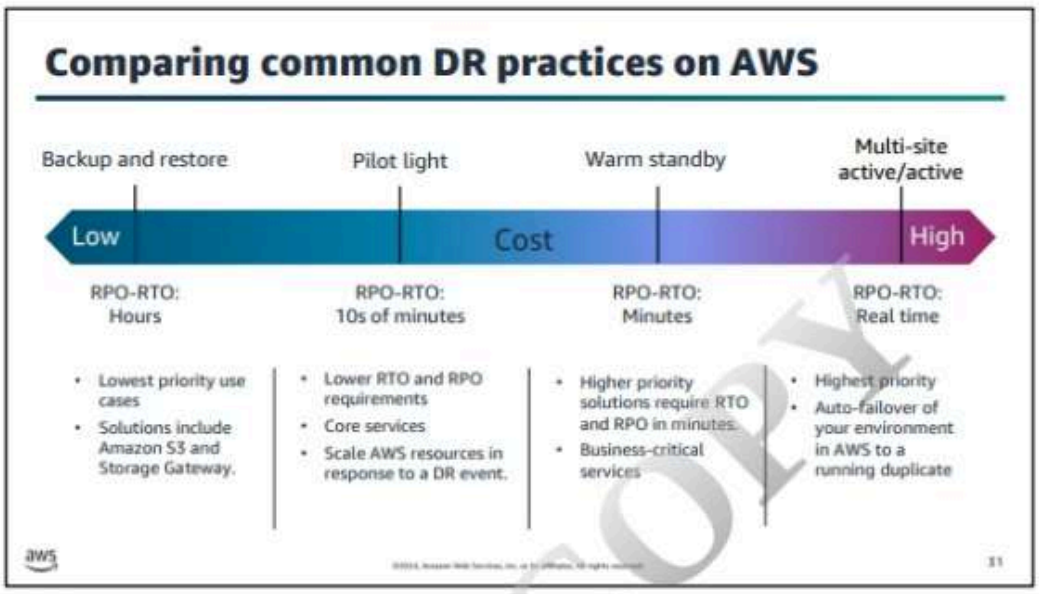
This pattern potentially has the least amount of downtime. It has more costs associated with it because more systems are running. The cost of this scenario is determined by how much production traffic is handled by AWS during normal operation. In the recovery phase, you pay for only what you use for the duration that the DR environment is required at full scale. To further reduce cost, purchase Amazon EC2 Reserved Instances for AWS servers that must be always on.

For more information about multi-site active/active solutions, see "Disaster Recovery (DR) Architecture on AWS, Part IV: Multi-Site Active/Active" in the *AWS Architecture Blog* at https://aws.amazon.com/blogs/architecture/disaster-recovery-dr-architecture-on-aws-part-iv-multi-site-active-active/.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



You deliver business continuity by making sure that critical business functions continue to operate or recover quickly despite serious disasters. Applications can be placed on a spectrum of complexity.

The figure shows a spectrum for the four scenarios, arranged by how quickly a system can be available to users after a DR event. With AWS, you can operate each of these DR strategies in a cost-effective way. These are just examples of possible approaches. Other variations and combinations of these are also possible.

In case of disaster, both pilot light and warm standby offer the capability to limit data loss (RPO). Both offer sufficient RTO performance that limits downtime. Between these two strategies, you have a choice of optimizing for RTO or for cost.

DO NOT COPY javierartiga.espublico@gmail.com

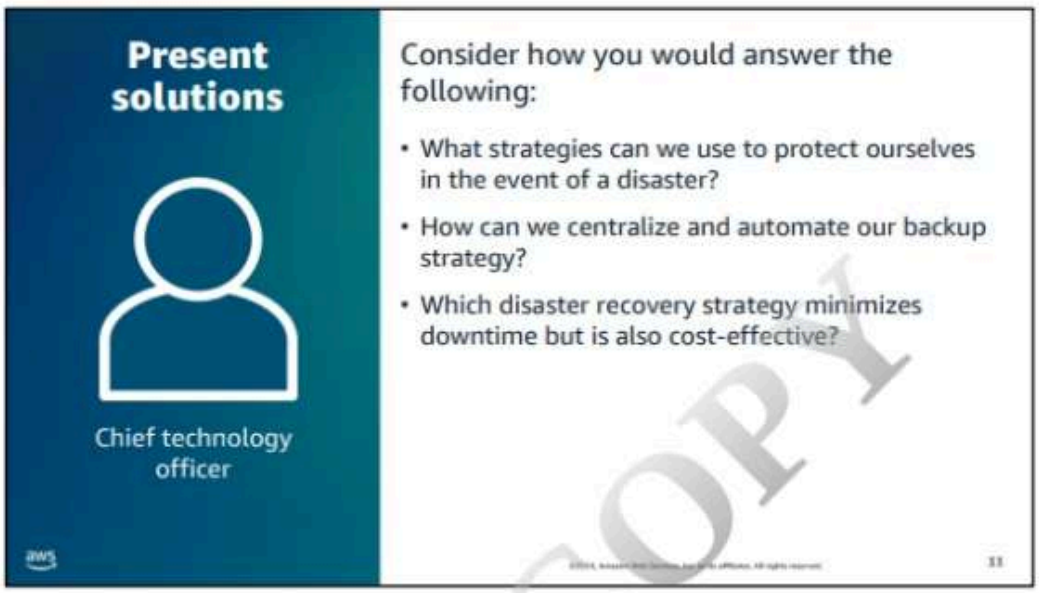
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Imagine you are now ready to talk to the chief technology officer and present solutions that meet their architectural needs. Think about how you would answer the questions from the beginning of the lesson.

Your answers should include the following solutions:

- Some of the ways the company can protect itself in the event of a disaster include the following:
  - Duplicate data by using different AWS storage services.
  - Create quickly deployable AMIs to launch compute resources.
  - Design the network with multiple failover mechanisms that will route traffic away from failed components.
  - Use database snapshots and backups.
- AWS Backup is a fully managed backup service that you can use to centralize and automate the backup of data across AWS services.
- The company should choose between backup and restore, pilot light, low-capacity standby on AWS, and multi-site active/active DR solutions.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Module review

In this module, you learned about the following:

- ✓ Disaster planning
- ✓ AWS Backup
- ✓ Recovery strategies

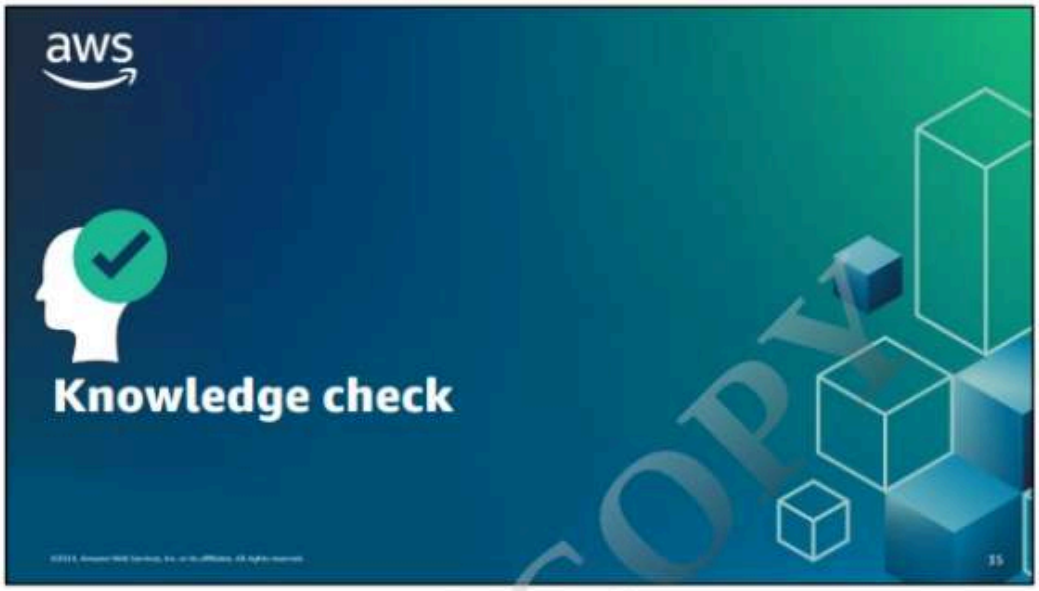
Next, you will review the following:

- Knowledge check
- Capstone lab introduction

14

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 1

Which disaster recovery model offers a Recovery Time Objective (RTO) in minutes at the lowest cost?

Choice	Response
A	Warm standby
B	Pilot light
C	Backup and restore
D	Multi-site active/active

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 1 and answer

Which disaster recovery model offers a Recovery Time Objective (RTO) in minutes at the lowest cost?

Choice	Response
A	Warm standby
B correct	Pilot light
C	Backup and restore
D	Multi-site active/active

The correct answer is B: pilot light.

The pilot light architecture is relatively inexpensive to implement. Other elements, such as application servers, are loaded with application code and configurations but are switched off. These elements are used only during testing or when disaster recovery failover is invoked. Unlike the backup and restore approach, your core infrastructure is always available. You always have the option to quickly provision a full-scale production environment by switching on and scaling out your application servers.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 2

Which metric defines how often data must be backed up?

Choice	Response
A	Recovery Time Objective (RTO)
B	Recovery Point Objective (RPO)
C	Available storage
D	Amount of data

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 2 and answer

Which metric defines how often data must be backed up?

Choice	Response
A	Recovery Time Objective (RTO)
B correct	Recovery Point Objective (RPO)
C	Available storage
D	Amount of data

The correct answer is B: Recovery Point Objective (RPO).

RPO is the acceptable amount of data loss measured in time. The RPO determines the maximum age of a backup before there is a risk of losing an unacceptable amount of data.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 3

Which options are features of AWS Backup? (Select THREE.)

Choice	Response
A	Encrypted backups
B	Ability to work across every AWS service
C	Ability to work across multiple services
D	Automated failover to read replicas
E	Incremental backups
F	Automated machine conversion

40

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 3 and answers

Which options are features of AWS Backup? (Select THREE.)

Choice	Response
A correct	Encrypted backups
B	Ability to work across every AWS service
C correct	Ability to work across multiple services
D	Automated failover to read replicas
E correct	Incremental backups
F	Automated machine conversion

The correct answers are A, C, and E: encrypted backups, ability to work across multiple services, and incremental backups.

Encrypted backups: Set the AWS KMS encryption key used to encrypt backups in the vault.

Ability to work across multiple services: AWS Backup works across your AWS accounts. It includes Amazon EC2 instances, Amazon EBS volumes, Amazon RDS databases, DynamoDB tables, Amazon EFS, FSx for Lustre, FSx for Windows File Server, and Storage Gateway volumes. The protection for Amazon RDS databases includes Aurora clusters.

Incremental backups: AWS Backup stores your periodic backups incrementally. The first backup of an AWS resource backs up a full copy of your data. For each successive incremental backup, only the changes to your AWS resources are backed up.

For more information about AWS Backup features, see "AWS Backup features" at https://aws.amazon.com/backup/features/.

DO NOT COPY javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 4

What is the best way to make an existing Amazon RDS DB instance highly available and minimize Recovery Time Objective (RTO)?

Choice	Response
A	Run a secondary copy of the DB instance in another Region.
B	Run a Multi-AZ DB instance in the same Region.
C	Create a read replica in another Region.
D	Create a read replica in the same Region.

42

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Knowledge check question 4 and answer

What is the best way to make an existing Amazon RDS DB instance highly available and minimize Recovery Time Objective (RTO)?

Choice	Response
A	Run a secondary copy of the DB instance in another Region.
B correct	Run a Multi-AZ DB instance in the same Region.
C	Create a read replica in another Region.
D	Create a read replica in the same Region.

The correct answer is B: Run a Multi-AZ DB instance in the same Region.

Multi-AZ is a feature of Amazon RDS that provisions a standby replica in another Availability Zone. The replica is kept in sync with your primary Amazon RDS instance.

In case of an infrastructure failure, Amazon RDS performs an automatic failover to the standby. In this way, you can resume database operations as soon as the failover is complete. The DB instance that failed is replaced with a new standby replica. Failover to a standby instance is faster than promoting a read replica because a read replica has to reboot.

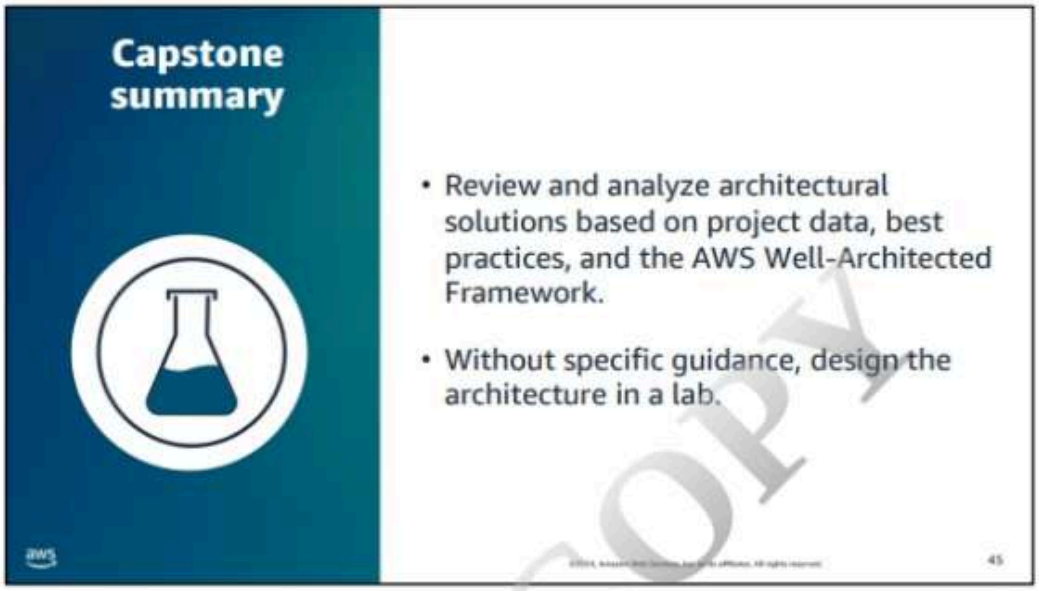
DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

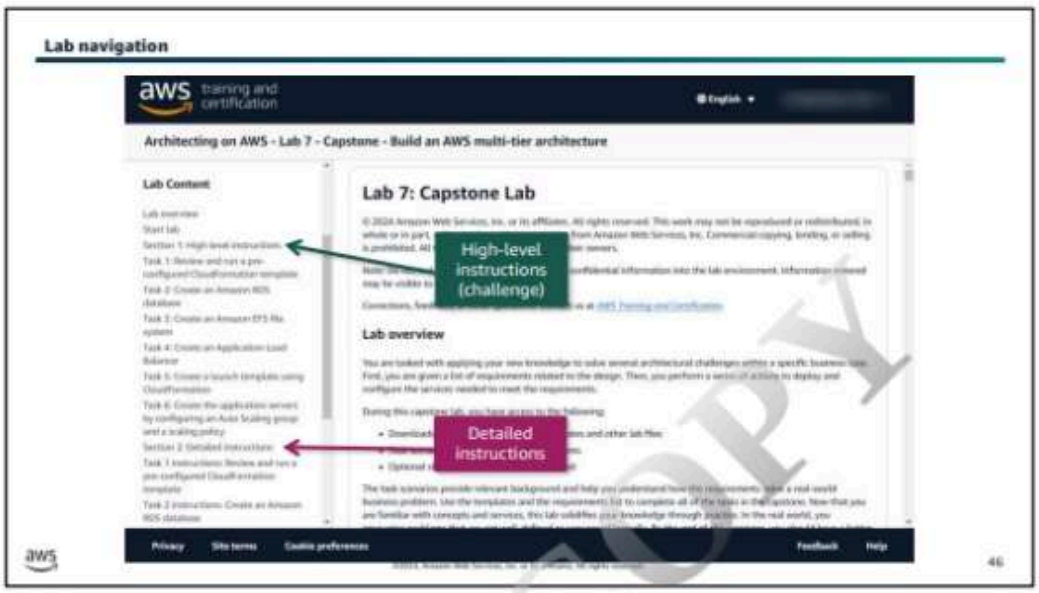


The capstone lab is the final project for this course.

During the lab, you receive a scenario that discusses a business need. Review the requirements and use what you have learned in this course to complete the list of tasks.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Find the capstone lab in the same place in which you completed labs 1–6.

The capstone lab is organized into the following sections:

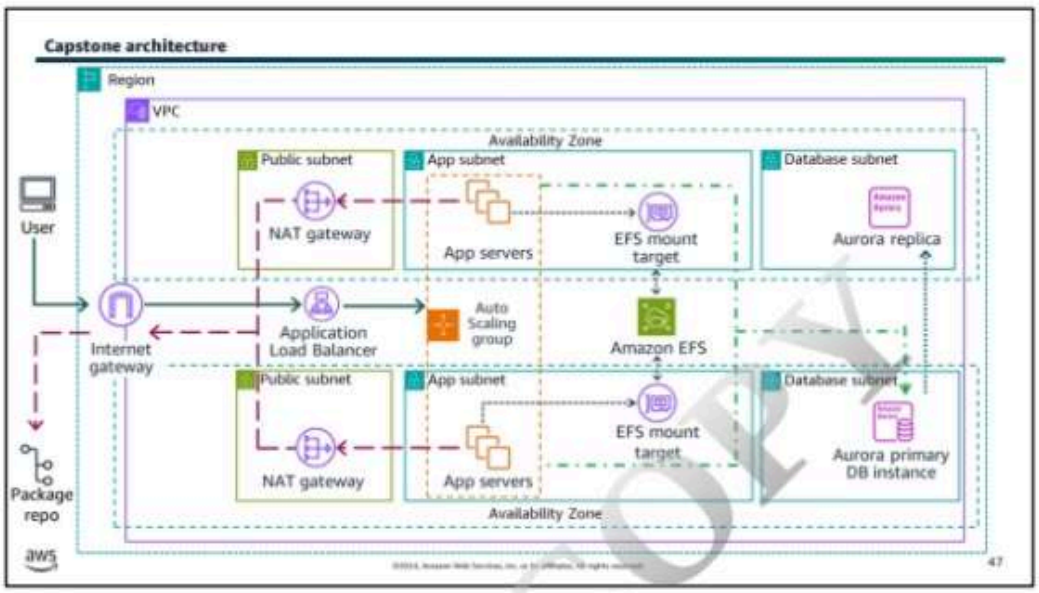
- **Lab overview:** Introduction to the lab.
- **Objectives:** Learning objectives of the capstone lab.
- **Start lab:** Instructions to begin the lab.
- **Lab scenario:** Description of what needs to be solved in the lab.
- **Section 1 – High-level instructions:** The primary instructions for the capstone lab. The instructions have an open-ended structure to attempt to challenge you and assess what you learned in this course.
- **Section 2 – Detailed instructions:** In-depth instructions for each task in the lab. Use the step-by-step instructions to complete that task only if necessary, and then move back to the next task in the high-level instructions.

Use the navigation menu on the left side of the screen to move between sections.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



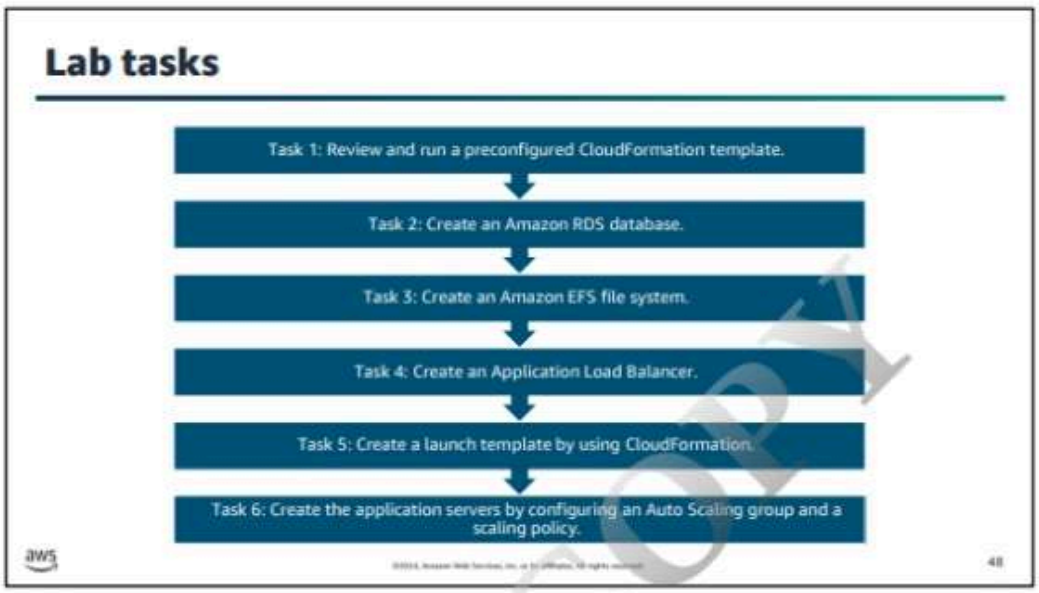
You build this multi-tier architecture in the capstone lab.

[Image description: The diagram shows a single AWS Region with one virtual private cloud (VPC) and two Availability Zones. Each Availability Zone contains a public subnet, an application subnet, and a database subnet. A unidirectional arrow outside the VPC routes from the remote user through an internet gateway to an Application Load Balancer. The Application Load Balancer is depicted as not residing in a subnet or an Availability Zone, but it does reside in both public subnets. The arrow continues to the Auto Scaling group that has application servers in the application subnets of both Availability Zones. Each application server communicates with an Amazon EFS mount target in its own subnet to reach the EFS file system, which does not reside inside the VPC and is shown this way for simplicity. All application servers communicate with an Amazon Aurora primary DB instance in one of the database subnets. The other database subnet holds the Aurora replica. A unidirectional arrow points between the Aurora primary and Aurora replica. A unidirectional arrow points from the application servers and travels through each Availability Zone's NAT gateway. The NAT gateways are in the public subnet of each Availability Zone. The arrows travel through each NAT gateway, exit the VPC through the internet gateway, and reach the remote package repository. The arrows in the diagram are shown as unidirectional to indicate where the request originated. Assume that the response from the target travels the path in reverse unless otherwise indicated. **End description].**

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

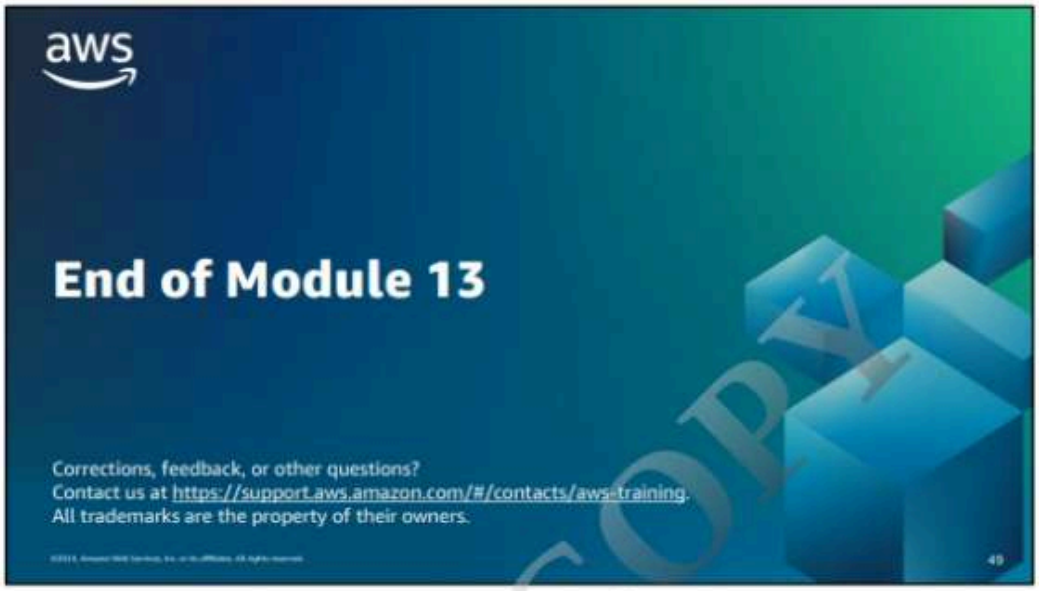


The lab is organized into six tasks. Remember that you can view the more detailed instructions if you need more assistance with a task step.

Begin reading the instructions for the capstone lab, and have fun building.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Course objective

- Identify Amazon Web Services (AWS) services, compare features, and explore best practices to architect resilient, secure, and highly available IT solutions on AWS.

Review the course objective and the summary of topics covered in this class.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Day 1 summary

- Architecting Fundamentals
- Lab 1: Explore and Interact with the AWS Management Console and AWS Command Line Interface
- Account Security
- Networking 1
- Compute
- Lab 2: Build Your Amazon VPC Infrastructure

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Day 2 summary

- Storage
- Database Services
- Lab 3: Create a Database Layer in Your Amazon VPC Infrastructure
- Monitoring and Scaling
- Lab 4: Configure High Availability in Your Amazon VPC
- Automation
- Containers

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Day 3 summary

- Networking 2
- Serverless
- Lab 5: Build a Serverless Architecture
- Edge Services
- Lab 6: Configure an Amazon CloudFront Distribution with an Amazon S3 Origin
- Backup and Recovery
- Capstone lab: Build an AWS Multi-Tier Architecture

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS Certification helps learners to build credibility and confidence by validating their cloud expertise with an industry-recognized credential. Certification helps organizations to identify skilled professionals who can lead cloud initiatives by using Amazon Web Services (AWS).

The slide shows the AWS certifications that are currently available. To earn an AWS Certification, you must earn a passing score on a proctored exam. For role-based certifications, the following are the recommended experience levels with AWS Cloud services for each certification level:

- **Professional:** Two years of comprehensive experience designing, operating, and troubleshooting solutions by using the AWS Cloud
- **Associate:** One year of experience solving problems and implementing solutions by using the AWS Cloud
- **Foundational:** Six months of fundamental AWS Cloud and industry knowledge

Specialty certifications focus on a particular technical domain. The recommended experience for taking a specialty exam is technical experience in the domain as specified in the exam guide.

AWS does not publish a list of all services or features that are covered in a certification exam. However, the exam guide for each exam lists the current topic areas and objectives that the exam covers. For more information and to review the exam guides and other preparation materials, see AWS Certification exam preparation at https://aws.amazon.com/certification/certification-prep/.

This information is current as of October 2024. However, exams are frequently updated, and the details regarding which exams are available—and what is tested by each exam—are subject to change. For more information about the latest AWS certification exam information, see AWS Certification at https://aws.amazon.com/certification/.

You are required to update your certification (or recertify) every 3 years. For more information, see AWS Recertification at https://aws.amazon.com/certification/recertification/.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

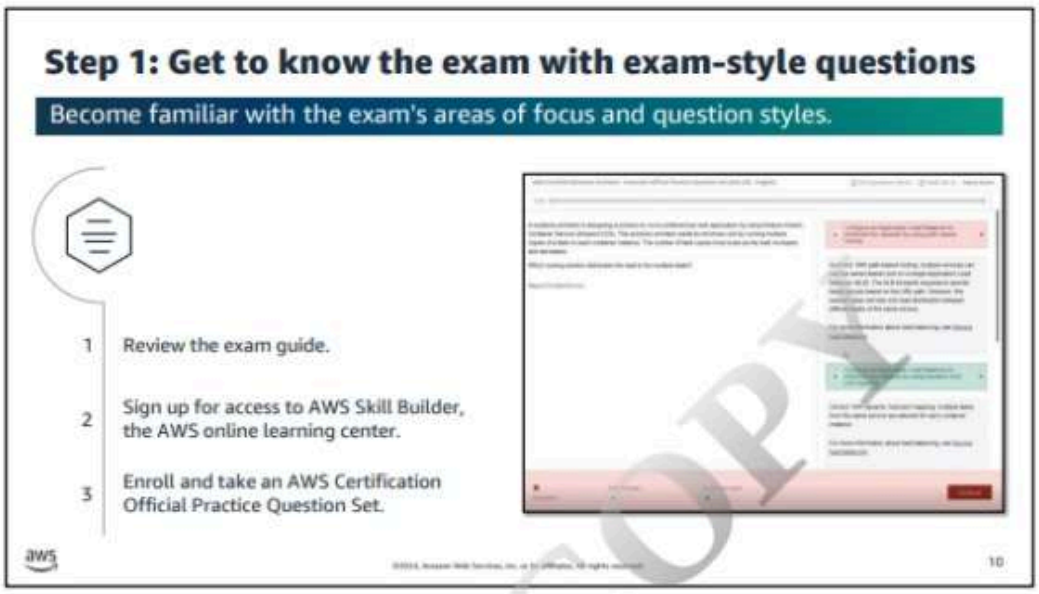
Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

This course includes content that might be related to an AWS Certification exam. To continue preparing for the exam, follow these four steps.

To see specific details in the four-step plan for each exam, scan the QR code to explore the preparation section of AWS Skill Builder at https://skillbuilder.aws/#prepare-for-exam.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Step 1 is getting to know the exam with exam-style questions.

We recommend reviewing the exam guide, which you can find from the four-step plan in the previous slide.

Official practice question sets feature 20 questions developed by AWS to demonstrate the style of certification exam questions. Take the question set to understand exam-style questions. The practice questions include detailed feedback and recommended resources to help you prepare for your exam.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

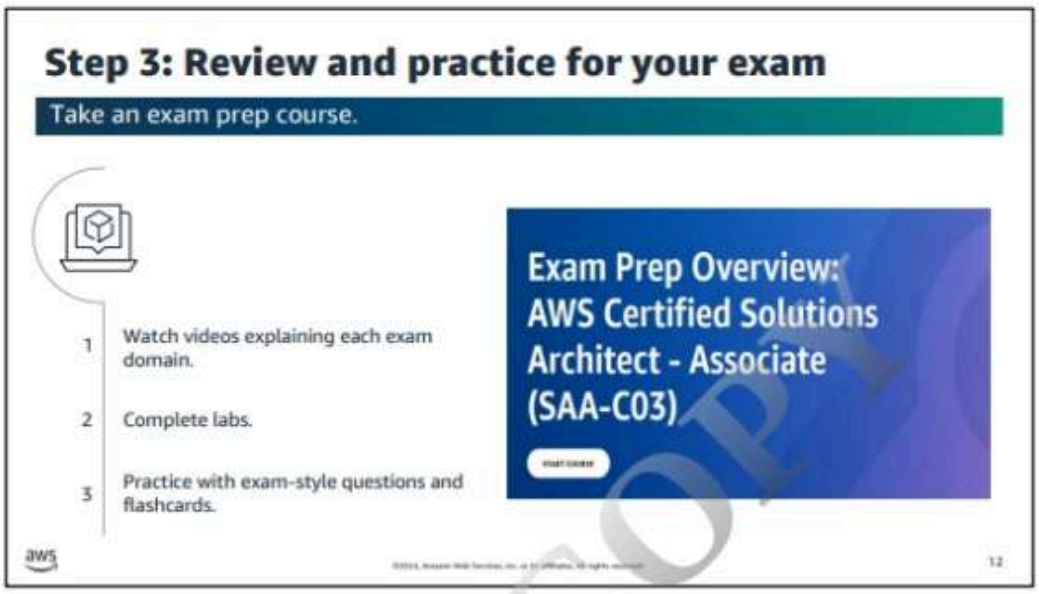
Step 2 is brushing up on exam topics. The four-step plan for each exam includes specific recommendations.

While there is no exact training required to take an AWS Certification, we offer suggestions for digital courses, builder labs, AWS Cloud Quest, and AWS Jams. Subscribe to AWS Skill Builder to unlock all of these training opportunities.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Step 3 is to review and practice.

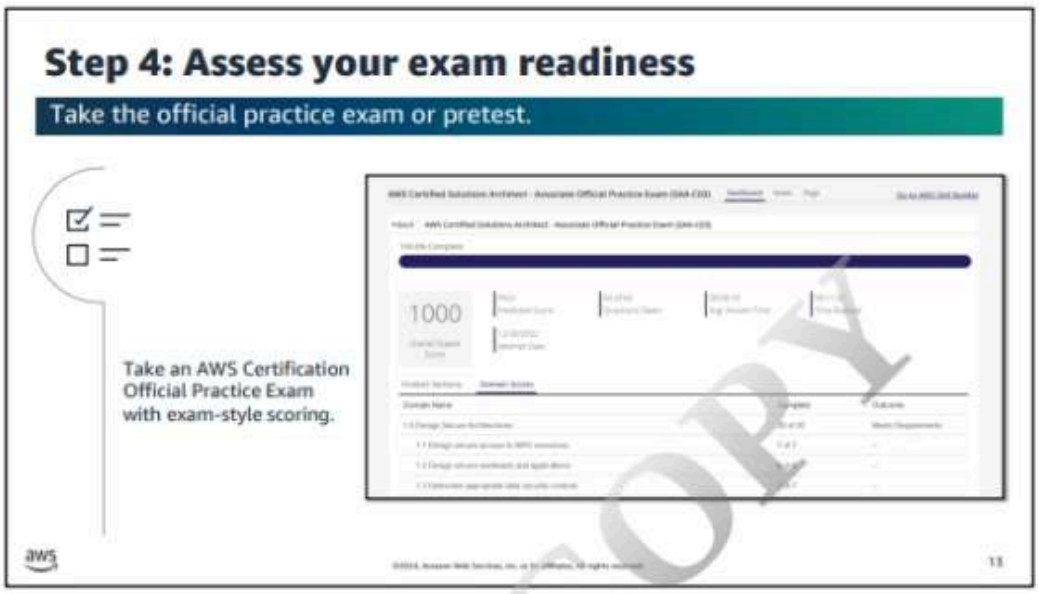
Prepare for the exam by exploring the exam's topic areas and how they align to developing on AWS and to specific areas of study. Gauge your understanding of topics and concepts from each task statement grouped by domain. Reinforce your knowledge and identify learning gaps with hands-on lab exercises and explanations of exam-style questions. Follow the instructor as they review exam-style questions. Learn test-taking strategies to identify incorrect responses.

Note: Some materials require an AWS Skill Builder subscription.

DO NOT COPY javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Step 4 is assessing your exam readiness.

Depending on the exam, you will take a pretest or practice exam to gauge your exam readiness. Questions in both assessments have the same style, depth, and rigor as the actual exam questions. Both assessments include detailed feedback and score reports. The only difference is that the practice exam has exam-style scoring, and the pretest provides standard percentage scoring.

DO NOT COPY
javierartiga.espublico@gmail.com

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS offers flexible, convenient options for taking exams.

For more information and to choose the exam option that works best for you, see Schedule an exam at https://aws.amazon.com/certification/certification-prep/testing/.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Continue your learning with AWS Skill Builder, our online learning center.

Are you ready to achieve your goals at your pace? Free digital training on AWS Skill Builder offers more than 600 on-demand courses and learning plans so that you can build the skills you need, your way.

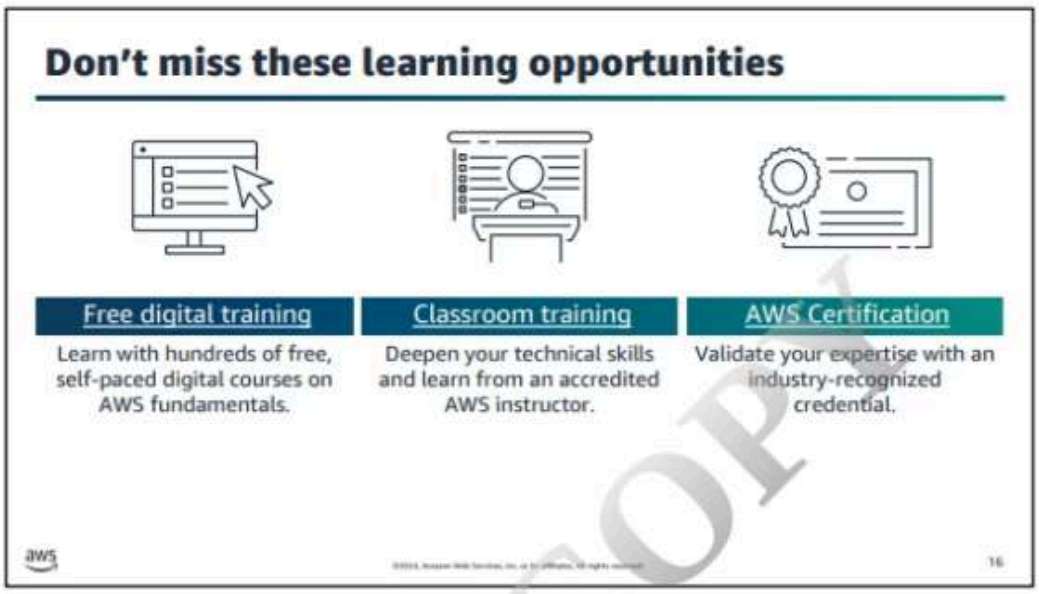
Want to build problem-solving cloud skills by using an interactive, engaging experience? A Skill Builder subscription offers access to self-paced labs, practice exams, role-based games, and real-world challenges to accelerate your learning.

To learn more and get started, see AWS Skill Builder at https://aws.amazon.com/training/digital.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



AWS Training and Certification is an organization dedicated to expanding and deepening knowledge of AWS and driving proliferation in the use of AWS services. Our programs are designed for customers, AWS Partners, and AWS employees. Over the past several months, we have rolled out several new courses, training labs, and certifications to our customers and partners.

Expand your AWS Cloud skills. For more information, see the following resources:

- **Digital training** – https://explore.skillbuilder.aws/
- **Classroom training** – https://aws.amazon.com/training
- **AWS Certification** – https://aws.amazon.com/certification
- **AWS Workshops** – https://workshops.aws/
- **Tech Talks** – https://aws.amazon.com/events/online-tech-talks/on-demand/
- **AWS Ramp-Up Guides** – https://aws.amazon.com/training/ramp-up-guides/

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: javierartiga.espublico@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DO NOT COPY
javierartiga.espublico@gmail.com

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.