

**Aws certified solutions architect
associate**

Outline

- **CHAPTER 1** - Introduction
- **CHAPTER 2** - AWS Overview
- **CHAPTER 3** - Identity and Access Management & S3
- **CHAPTER 4** - EC2
- **CHAPTER 5** - Databases on AWS
- **CHAPTER 6** - Advanced IAM
- **CHAPTER 7** - Route 53
- **CHAPTER 8** - VPCs
- **CHAPTER 9** - HA Architecture
- **CHAPTER 10** - Applications
- **CHAPTER 11** - Security
- **CHAPTER 12** - Serverless

CHAPTER 1

Introduction

Exam Blueprint

- **130 Minutes in Length**
- **65 Questions (this can change)**
- **Multiple Choice**
- **Results are between 100 - 1000 with a passing score of 720**
- **Aim for 70%**
- **Qualification is valid for 3 years**
- **Scenario based questions**

Exam - Content Outline

This exam guide includes weightings, test domains, and objectives only. It is not a comprehensive listing of the content on this examination. The table below lists the main content domains and their weightings.

Domain	% of Examination
Domain 1: Design Resilient Architectures	30%
Domain 2: Design High-Performing Architectures	28%
Domain 3: Design Secure Applications and Architectures	24%
Domain 4: Design Cost-Optimized Architectures	18%
TOTAL	100%

Exam - Recommended AWS Knowledge

- **1 year of hands-on experience designing available, cost-effective, fault-tolerant, and scalable distributed systems on AWS.**
- **Hands-on experience using compute, networking, storage, and database AWS services.**
- **Hands-on experience with AWS deployment and management services.**
- **Ability to identify and define technical requirements for an AWS-based application.**
- **Ability to identify which AWS services meet a given technical requirement.**
- **Knowledge of recommended best practices for building secure and reliable applications on the AWS platform.**
- **An understanding of the basic architectural principles of building in the AWS Cloud.**
- **An understanding of the AWS global infrastructure.**
- **An understanding of network technologies as they relate to AWS.**
- **An understanding of security features and tools that AWS provides and how they relate to traditional services.**

CHAPTER 2

AWS Overview

The History of AWS So Far

“Invention requires two things: 1. The ability to try a lot of experiments, and 2. not having to live with the collateral damage of failed experiments.”

Andy Jassy

CEO Amazon Web Services



The History of AWS So Far

- **2003** - Chris Pinkham & Benjamin Black present a paper on what Amazon's own internal infrastructure should look like. They suggested selling it as a service and prepared a business case.
- SQS officially launched in **2004**
- AWS Officially launched in **2006**
- **2007** over 180,000 developers on the platform
- **2010** all of amazon.com moved over
- **2012** First re:Invent Conference



The History of AWS So Far

- **2013** Certifications Launched
- **2014** Committed to achieve 100% renewable energy usage for its global footprint
- **2015** AWS breaks out its revenue: \$6 Billion USD per annum and growing close to 90% year on year
- **2016** Run rate of \$13 billion USD.
- **2017** AWS re:invent releases a host of Artificial Intelligent Services. Run rate hits \$27 Billion USD.
- **2018** AWS launch Machine Learning Specialty Certs. Heavy focus on automating AI & ML.
- **2019** Alexa Specialty Beta Certificate Launched. 10 Certs!



AWS Console 2015

AWS Services

Compute

- EC2** Virtual Servers in the Cloud
- Lambda** PREVIEW Run Code in Response to Events

Storage & Content Delivery

- S3** Scalable Storage in the Cloud
- Storage Gateway** Integrates On-Premises IT Environments with Cloud Storage
- Glacier** Archive Storage in the Cloud
- CloudFront** Global Content Delivery Network

Database

- RDS** MySQL, PostgreSQL, Oracle, SQL Server, and Amazon Aurora
- DynamoDB** Predictable and Scalable NoSQL Data Store
- ElastiCache** In-Memory Cache
- Redshift** Managed Petabyte-Scale Data Warehouse Service

Networking

- VPC** Isolated Cloud Resources
- Direct Connect** Dedicated Network Connection to AWS
- Route 53** Scalable DNS and Domain Name Registration

Administration & Security

- Directory Service** Managed Directories in the Cloud
- Identity & Access Management** Access Control and Key Management
- Trusted Advisor** AWS Cloud Optimization Expert
- CloudTrail** User Activity and Change Tracking
- Config** PREVIEW Resource Configurations and Inventory
- CloudWatch** Resource and Application Monitoring

Deployment & Management

- Elastic Beanstalk** AWS Application Container
- OpsWorks** DevOps Application Management Service
- CloudFormation** Templated AWS Resource Creation
- CodeDeploy** Automated Deployments

Analytics

- EMR** Managed Hadoop Framework
- Kinesis** Real-time Processing of Streaming Big Data
- Data Pipeline** Orchestration for Data-Driven Workflows

Application Services

- SQS** Message Queue Service
- SWF** Workflow Service for Coordinating Application Components
- AppStream** Low Latency Application Streaming
- Elastic Transcoder** Easy-to-use Scalable Media Transcoding
- SES** Email Sending Service
- CloudSearch** Managed Search Service

Mobile Services

- Cognito** User Identity and App Data Synchronization
- Mobile Analytics** Understand App Usage Data at Scale
- SNS** Push Notification Service

Enterprise Applications

- WorkSpaces** Desktops in the Cloud
- Zocalo** Secure Enterprise Storage and Sharing Service

Additional Resources

Getting Started
See our documentation to get started and learn more about how to use our services.

AWS Console Mobile App
View your resources on the go with our AWS Console mobile app, available from Amazon Appstore, Google Play, or iTunes.

AWS Marketplace
Find and buy software, launch with 1-Click and pay by the hour.

Service Health
 All services operating normally.
Updated: Dec 18 2014 10:31:00 GMT-0600
[Service Health Dashboard](#)

Set Start Page

© 2008 - 2014, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy Policy](#) [Terms of Use](#)

[Feedback](#)

AWS Console 2016

AWS Management Console X

https://console.aws.amazon.com/console/home?nc2=h_m_mc®ion=us-east-1

N. Virginia • Support •

Amazon Web Services

- Compute
 - EC2 Virtual Servers in the Cloud
 - Lambda Run Code in Response to Events
 - EC2 Container Service Run and Manage Docker Containers
- Storage & Content Delivery
 - S3 Scalable Storage in the Cloud
 - Elastic File System PREVIEW Fully Managed File System for EC2
 - Storage Gateway Integrates On-Premises IT Environments with Cloud Storage
 - Glacier Archive Storage in the Cloud
 - CloudFront Global Content Delivery Network
- Database
 - RDS MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
 - DynamoDB Predictable and Scalable NoSQL Data Store
 - ElastiCache In-Memory Cache
 - Redshift Managed Petabyte-Scale Data Warehouse Service
- Networking
 - VPC Isolated Cloud Resources
 - Direct Connect Dedicated Network Connection to AWS
 - Route 53 Scalable DNS and Domain Name Registration
- Administration & Security
 - Directory Service Managed Directories in the Cloud
 - Identity & Access Management Access Control and Key Management
 - Trusted Advisor AWS Cloud Optimization Expert
 - CloudTrail User Activity and Change Tracking
 - Config Resource Configurations and Inventory
 - CloudWatch Resource and Application Monitoring
- Deployment & Management
 - Elastic Beanstalk AWS Application Container
 - OpsWorks DevOps Application Management Service
 - CloudFormation Templated AWS Resource Creation
 - CodeDeploy Automated Deployments
- Analytics
 - EMR Managed Hadoop Framework
 - Kinesis Realtime Processing of Streaming Big Data
 - Data Pipeline Orchestration for Data-Driven Workflows
 - Machine Learning Build Smart Applications Quickly and Easily
- Application Services
 - SQS Message Queue Service
 - SWF Workflow Service for Coordinating Application Components
 - AppStream Low Latency Application Streaming
 - Elastic Transcoder Easy-to-use Scalable Media Transcoding
 - SES Email Sending Service
 - CloudSearch Managed Search Service
- Mobile Services
 - Cognito User Identity and App Data Synchronization
 - Mobile Analytics Understand App Usage Data at Scale
 - SNS Push Notification Service
- Enterprise Applications
 - WorkSpaces Desktops in the Cloud
 - WorkDocs Secure Enterprise Storage and Sharing Service
 - WorkMail PREVIEW Secure Email and Calendaring Service

Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

Create a Group Tag Editor

Additional Resources

Getting Started

See our documentation to get started and learn more about how to use our services.

AWS Console Mobile App

View your resources on the go with our AWS Console mobile app, available from Amazon Appstore, Google Play, or iTunes.

AWS Marketplace

Find and buy software, launch with 1-Click and pay by the hour.

AWS re:Invent - Register Now

Join us for keynote announcements, technical sessions, bootcamps and more.

Service Health

All services operating normally.

Updated: Jun 16 2015 00:13:00 GMT-0500

Service Health Dashboard

Feedback English

© 2006 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

AWS Console 2017



History

Console Home

Search services

Group

A-Z

Compute

- EC2
- EC2 Container Service
- Lightsail
- Elastic Beanstalk
- Lambda
- Batch

Developer Tools

- CodeCommit
- CodeBuild
- CodeDeploy
- CodePipeline

Analytics

- Athena
- EMR
- CloudSearch
- Elasticsearch Service
- Kinesis
- Data Pipeline
- QuickSight

Application Services

- Step Functions
- SWF
- API Gateway
- Elastic Transcoder

Storage

- S3
- EFS
- Glacier
- Storage Gateway

Management Tools

- CloudWatch
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog
- Trusted Advisor
- Managed Services

Artificial Intelligence

- Lex
- Polly
- Rekognition
- Machine Learning

Messaging

- SQS
- SNS
- SES

Database

- RDS
- DynamoDB
- ElastiCache
- Redshift

Security, Identity & Compli...

- IAM
- Inspector
- Certificate Manager
- Directory Service
- WAF & Shield
- Compliance Reports

Internet Of Things

- AWS IoT

Business Productivity

- WorkDocs
- WorkMail

Networking & Content Deli...

- VPC
- CloudFront
- Direct Connect
- Route 53

Game Development

- GameLift

Desktop & App Streaming

- WorkSpaces
- AppStream 2.0

Migration

- DMS
- Server Migration
- Snowball

Mobile Services

- Mobile Hub
- Cognito
- Device Farm
- Mobile Analytics
- Pinpoint

AWS Console

History

Console Home

Find a service by name or feature (for example, EC2, S3 or VM, storage).

Group A-Z

Compute

- EC2
- Lightsail ↗
- ECS
- EKS
- Lambda
- Batch
- Elastic Beanstalk
- ECR

Blockchain

- Amazon Managed Blockchain
- Satellite
- Ground Station

Analytics

- Athena
- EMR
- CloudSearch
- Elasticsearch Service
- Kinesis
- QuickSight ↗
- Data Pipeline
- AWS Glue
- MSK

Customer Engagement

- Amazon Connect
- Pinpoint
- Simple Email Service

Storage

- S3
- EFS
- FSx
- S3 Glacier
- Storage Gateway

Management & Governance

- CloudWatch
- AWS Auto Scaling
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog
- Systems Manager
- Trusted Advisor
- Managed Services
- Control Tower
- AWS License Manager
- AWS Well-Architected Tool

Security, Identity, & Compliance

- IAM
- Resource Access Manager
- Cognito
- Secrets Manager
- GuardDuty
- Inspector
- Amazon Macie ↗
- AWS Organizations
- AWS Single Sign-On
- Certificate Manager
- Key Management Service
- CloudHSM
- Directory Service
- WAF & Shield
- Artifact
- Security Hub

Business Applications

- Alexa for Business
- Amazon Chime ↗
- WorkDocs
- WorkMail

Desktop & App Streaming

- WorkSpaces
- AppStream 2.0

Internet Of Things

- IoT Core
- Amazon FreeRTOS
- IoT 1-Click
- IoT Analytics
- IoT Device Defender
- IoT Device Management
- IoT Events
- IoT Greengrass
- IoT SiteWise
- IoT Things Graph

Database

- RDS
- DynamoDB
- ElastiCache
- Neptune
- Amazon Redshift

Media Services

- Elastic Transcoder
- Kinesis Video Streams
- MediaConnect
- MediaConvert
- MediaLive

Migration & Transfer

- AWS Migration Hub
- Application Discovery Service

▲ close

AWS Console

AWS Management Console

AWS services

Find services
You can enter names, keyword or acronyms.

[▶ All services](#)

Build a solution

Get started with simple wizards and automated workflows.

Launch a virtual machine
With EC2
~2-3 minutes



Build a web app
With Elastic Beanstalk
~6 minutes



Build using virtual servers
With Lightsail
~1-2 minutes



Connect an IoT device
With AWS IoT
~5 minutes



Start a development project
With CodeStar
~5 minutes



Register a domain
With Route 53
~3 minutes



Access resources

 Access the AWS Management Console from your mobile device.

Explore AWS

Amazon Redshift
Fast, simple, cost-effective way to extend queries to your data.

Run Serverless Containers with AWS Fargate
AWS Fargate runs and scales your containers without having to manage servers or clusters. [Learn more.](#)

AWS Marketplace
Find, buy, and deploy popular software products that run on AWS. [Learn more.](#)

Scalable, Durable, Secure Backup & Restore with Amazon S3
Discover how customers are building backup & restore solutions on AWS that save money. [Learn more.](#)

US East (N. Virginia)

US East (Ohio)

US West (N. California)

US West (Oregon)

Asia Pacific (Mumbai)

Asia Pacific (Seoul)

Asia Pacific (Singapore)

Asia Pacific (Sydney)

Asia Pacific (Tokyo)

Canada (Central)

EU (Frankfurt)

EU (Ireland)

EU (London)

EU (Paris)

South America (São Paulo)

AWS - High Level Services

IOT		Game Development
Customer Engagement	Business Applications	Desktop & App Streaming
AR & VR	Application Integration	AWS Cost Management
Analytics	Security, Identity & Compliance	Mobile
Management & Governance	Media Services	Machine Learning
Robotics	Blockchain	Satellite
Migration & Transfer	Network & Content Delivery	Developer Tools
Compute	Storage	Databases

AWS Global Infrastructure

AWS Global Infrastructure

Global Infrastructure



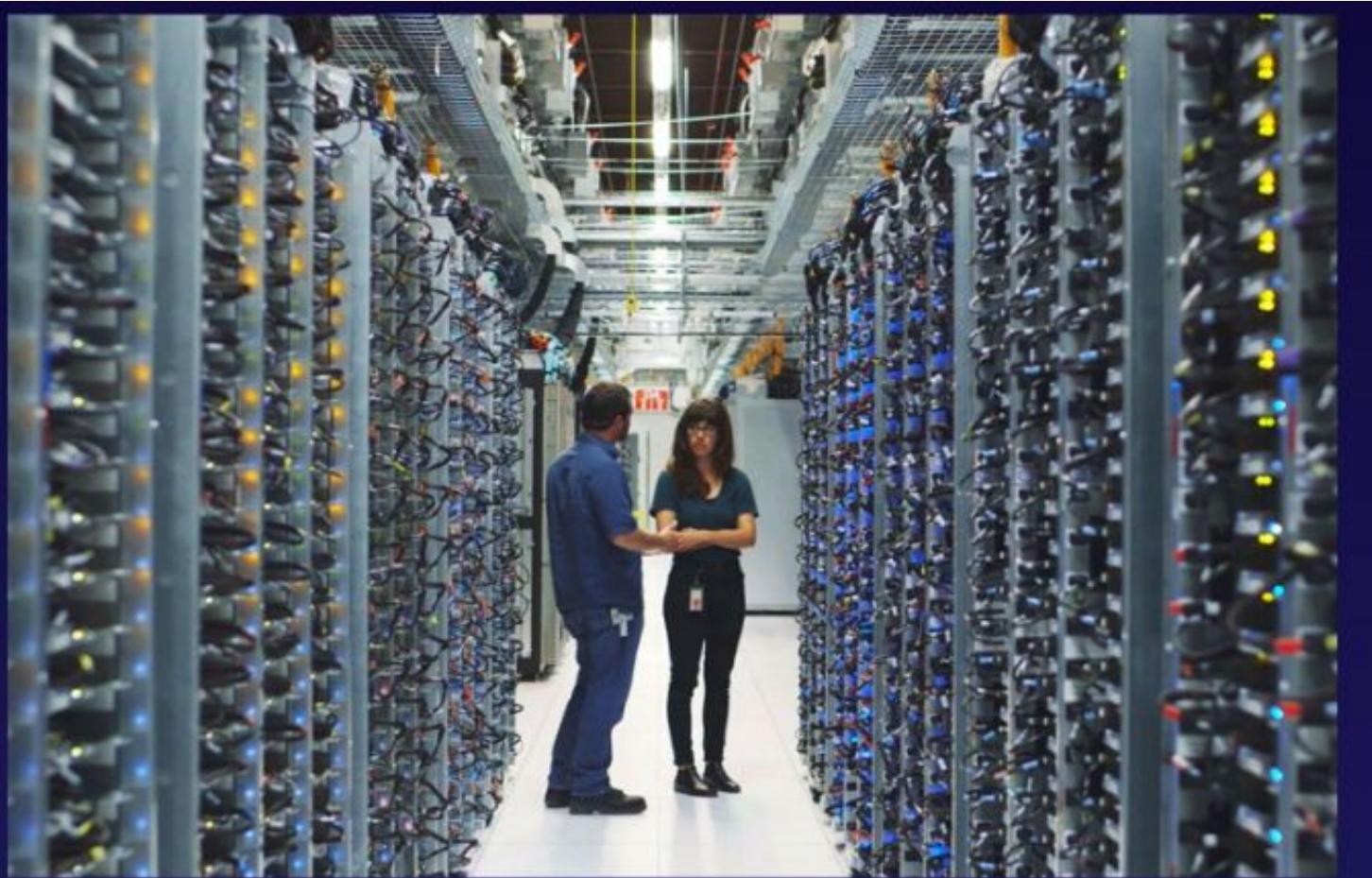
19 Regions & 57 Availability Zones - December 2018
5 More Regions & 15 More AZ's for 2019

AWS Global Infrastructure



Think of an Availability Zone As A Data Center

AWS Global Infrastructure

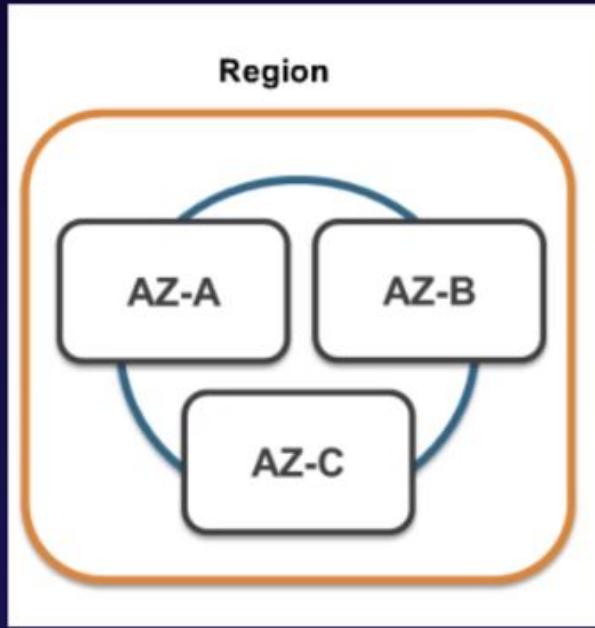


AWS Global Infrastructure



An Availability may be several data centres, but because they are close together, they are counted as 1 Availability Zone.

AWS Global Infrastructure



A Region is a geographical area. Each Region consists of 2 (or more) Availability Zones.

A Brief Look Today's Regions

North America



US East (Northern Virginia) Region

EC2 Availability Zones: 6

Launched 2006

US East (Ohio) Region

EC2 Availability Zones: 3

Launched 2016

US West (Oregon) Region

EC2 Availability Zones: 3

Launched 2011

US West (Northern California) Region

EC2 Availability Zones: 3*

Launched 2009

GovCloud (US-West) Region

EC2 Availability Zones: 3

Launched 2011

GovCloud (US-East) Region

EC2 Availability Zones: 3

Launched 2018

Canada (Central) Region

EC2 Availability Zones: 2

Launched 2016

Learn more at [AWS Canada](#)

AWS Edge Network Locations:

Edge locations - Ashburn, VA (3); Atlanta GA (3); Boston, MA; Chicago, IL (2); Dallas/Fort Worth, TX (5); Denver, CO (2); Hayward, CA; Hillsboro, OR; Jacksonville, FL; Los Angeles, CA (4); Miami, FL (3); Minneapolis, MN; Montreal, QC; New York, NY (3); Newark, NJ (3); Palo Alto, CA; Phoenix, AZ; Philadelphia, PA; San Jose, CA (2); Seattle, WA (3); South Bend, IN; St. Louis, MO; Toronto, ON

A Brief Look Today's Regions

South America



South America (São Paulo) Region

EC2 Availability Zones: 3*

Launched 2011

AWS Edge Network Locations

Edge locations - Rio de Janeiro (2), Brazil; São Paulo, Brazil (2)

Regional Edge Caches - São Paulo, Brazil

*New customers can access two EC2 Availability Zones in South America (São Paulo)

[See detailed list of offerings at all AWS locations](#)

A Brief Look Today's Regions

Europe / Middle East / Africa



Europe (Ireland) Region²

EC2 Availability Zones: 3
Launched 2007

Europe (Frankfurt) Region

EC2 Availability Zones: 3
Launched 2014

Europe (London) Region

EC2 Availability Zones: 3
Launched 2016

Europe (Paris) Region

EC2 Availability Zones: 3
Launched 2017

AWS Edge Network Locations

Edge locations - Amsterdam, The Netherlands (2); Berlin, Germany; Cape Town, South Africa; Copenhagen, Denmark; Dubai, United Arab Emirates; Dublin, Ireland; Frankfurt, Germany (8); Fujairah, United Arab Emirates; Helsinki, Finland; Johannesburg, South Africa; London, England (9); Madrid, Spain (2); Manchester, England; Marseille, France; Milan, Italy; Munich, Germany; Oslo, Norway; Palermo, Italy; Paris, France (4); Prague, Czech Republic; Stockholm, Sweden (3); Vienna, Austria; Warsaw, Poland; Zurich, Switzerland

Regional Edge Caches - Frankfurt, Germany; London, England

[See detailed list of offerings at all AWS locations](#)

² Europe (Ireland) Region is located in the Republic of Ireland.

A Brief Look Today's Regions

Asia Pacific



Asia Pacific (Singapore) Region

EC2 Availability Zones: 3

Launched 2010

Asia Pacific (Seoul) Region

EC2 Availability Zones: 2

Launched 2016

Asia Pacific (Tokyo) Region

EC2 Availability Zones: 4*

Launched 2011

Asia Pacific (Mumbai) Region

EC2 Availability Zones: 2

Launched 2016

Asia Pacific (Osaka) Local Region¹

EC2 Availability Zones: 1

Launched 2018

China (Beijing) Region

EC2 Availability Zones: 2

[Learn more at \[www.amazonaws.cn\]\(http://www.amazonaws.cn\)](http://www.amazonaws.cn)

Asia Pacific (Sydney) Region

EC2 Availability Zones: 3

Launched 2012

China (Ningxia) Region

EC2 Availability Zones: 3

[Learn more at \[www.amazonaws.cn\]\(http://www.amazonaws.cn\)](http://www.amazonaws.cn)

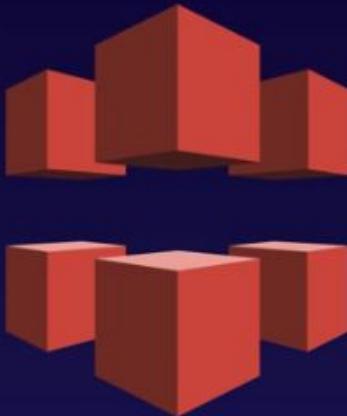
AWS Edge Network Locations

Edge locations - Bangalore, India; Chennai, India (2); Hong Kong SAR, China (3); Hyderabad, India (2); Kuala Lumpur, Malaysia; Manila, the Philippines; Melbourne, Australia; Mumbai, India (2); New Delhi, India (3); Osaka, Japan; Perth, Australia; Seoul, Korea (4); Singapore (3); Sydney, Australia; Taipei, Taiwan (3); Tokyo, Japan (9)

Regional Edge Caches - Mumbai, India; Seoul, Korea; Singapore; Sydney, Australia; Tokyo, Japan

*New customers can access three EC2 Availability Zones in Asia Pacific (Tokyo).

Edge Locations



Edge Locations are endpoints for AWS which are used for caching content. Typically this consists of CloudFront, Amazon's Content Delivery Network (CDN)

**There are many more Edge Locations than Regions.
Currently there are over 150 Edge Locations.**

What Do I Need to Know To Pass My Solutions Architect Exam?

Security, Identity & Compliance

Network & Content Delivery

Compute

Storage

Databases

AWS Global Infrastructure

Exam Tips

Understand the difference between a region, an Availability Zone (AZ) and an Edge Location.

- A Region is a physical location in the world which consists of two or more Availability Zones (AZ's).
- An AZ is one or more discrete data centers, each with redundant power, networking and connectivity, housed in separate facilities.
- Edge Locations are endpoints for AWS which are used for caching content. Typically this consists of CloudFront, Amazon's Content Delivery Network (CDN)

AWS Overview Quizz

AWS Overview Quizz

QUESTION 1

Which statement best describes Availability Zones?

- Two zones containing compute resources that are designed to automatically maintain synchronized copies of each other's data.
- Distinct locations from within an AWS region that are engineered to be isolated from failures.
- Restricted areas designed specifically for the creation of Virtual Private Clouds.
- A Content Distribution Network used to distribute content to users.

Good work!

An Availability Zone (AZ) is a distinct location within an AWS Region. Each Region comprises at least two AZs.

AWS Overview Quizz

QUESTION 2

What is an AWS region?

- A region is an independent data center, located in different countries around the globe.
- A region is a subset of AWS technologies. For example, the Compute region consists of EC2, ECS, Lambda, etc.
- A region is a collection of Edge Locations available in specific countries.
- A region is a geographical area divided into Availability Zones. Each region contains at least two Availability Zones.

Good work!

A region is a geographical area divided into Availability Zones. Each region contains at least two Availability Zones.

AWS Overview Quizz

QUESTION 3

Which of the following is correct?

of Regions > # of Availability Zones > # of Edge Locations

of Edge Locations > # of Availability Zones > # of Regions

of Availability Zones > # of Regions > # of Edge Locations

of Availability Zones > # of Edge Locations > # of Regions

Sorry!

Correct Answer

The number of Edge Locations is greater than the number of Availability Zones, which is greater than the number of Regions.

AWS Overview Quizz

QUESTION 4

Which of the below are compute service from AWS?

Choose 2



Good work!

Both Lambda and EC2 offer computing in the cloud. S3 is a storage offering while VPC is a network service.

AWS Overview Quizz

QUESTION 5

In which of the following is CloudFront content cached?

- Region
- Availability Zone
- Edge Location
- Data Center

Good work!

CloudFront content is cached in Edge Locations.

AWS Overview Quizz

QUESTION 6

Which of the below are factors that have helped make public cloud so powerful?

Choose 2

Traditional methods that are used for on-premise infrastructure always work just as well in cloud

No special skills required

Generally little upfront payment (Mostly pay as you go).

The ease of trying new solutions.

Good work!

Public cloud allows organizations to try out new ideas, new approaches and experiment with little upfront commitment. If it doesn't work out, organizations have the ability to terminate the resources and stop paying for them.

AWS Overview Quizz

QUESTION 7

What is an Amazon VPC?

Virtual Private Cloud

Virtual Private Compute

Virtual Public Cloud

Virtual Public Compute

Good work!

VPC stands for Virtual Private Cloud.

AWS Overview Quizz

QUESTION 8

Which of the following are a part of AWS' Network and Content Delivery services?

Choose 2

Cloudfront

EC2

VPC

RDS

Good work!

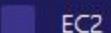
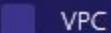
VPC allows you to provision a logically isolated section of the AWS where you can launch AWS resources in a virtual network. Cloudfront is a fast, highly secure and programmable content delivery network (CDN). EC2 provides compute resources while RDS is Amazon's Relational Database System.

AWS Overview Quizz

QUESTION 9

Which of the below are storage services in AWS?

Choose 2



Good work!

S3 and EFS both provide the ability to store files in the cloud. EC2 provides compute, and is often augmented with other storage services. VPC is a networking service.

AWS Overview Quizz

QUESTION 10

What does an AWS Region consist of?

- A collection of databases that can only be accessed from a specific geographic region.
- A console that gives you a quick, global picture of your cloud computing environment.
- A collection of data centers that is spread evenly around a specific continent.
- A distinct location within a geographic area designed to provide high availability to a specific geography.

Sorry!

Correct Answer

Each region is a separate geographic area. Each region has multiple, isolated locations known as Availability Zones.

AWS Overview Quizz

QUESTION 11

Which of the below are database services from AWS?

Choose 2

DynamoDB

RDS

EC2

S3

Good work!

RDS is a service for relational databases provided by AWS. DynamoDB is AWS' fast, flexible, no-sql database service. S3 provides the ability to store files in the cloud and is not suitable for databases, while EC2 is part of the compute family of services.

AWS Overview Quizz

QUESTION 12

The VPC service is a member of which group of AWS services in the 'All services' view of the AWS Portal?

Compute Services

Database Services

Networking & Content Delivery

Global Infrastructure

Sorry!

VPC is found in the "Networking & Content Delivery" section of the AWS Portal.

Correct Answer

A Virtual Private Cloud (VPC) is a virtual network dedicated to a single AWS account. It is logically isolated from other virtual networks in the AWS cloud. VPC is found in the "Networking & Content Delivery" section of the AWS Portal.

CHAPTER 3

Identity and Access Management & S3

What is IAM?

IAM allows you to manage users and their level of access to the AWS Console.

It is important to understand IAM and how it works, both for the exam and for administering a company's AWS account in real life.



Key features of IAM

Identity Access Management (IAM) offers the following features;

- Centralised control of your AWS account
- Shared Access to your AWS account
- Granular Permissions
- Identity Federation (including Active Directory, Facebook, Linkedin etc)
- Multifactor Authentication
- Provide temporary access for users/devices and services where necessary
- Allows you to set up your own password rotation policy
- Integrates with many different AWS services
- Supports PCI DSS Compliance



Terminology of IAM

1

Users

End Users such as people, employees of an organization, etc.

2

Groups

A collection of users. Each user in the group will inherit the permissions of the group.

3

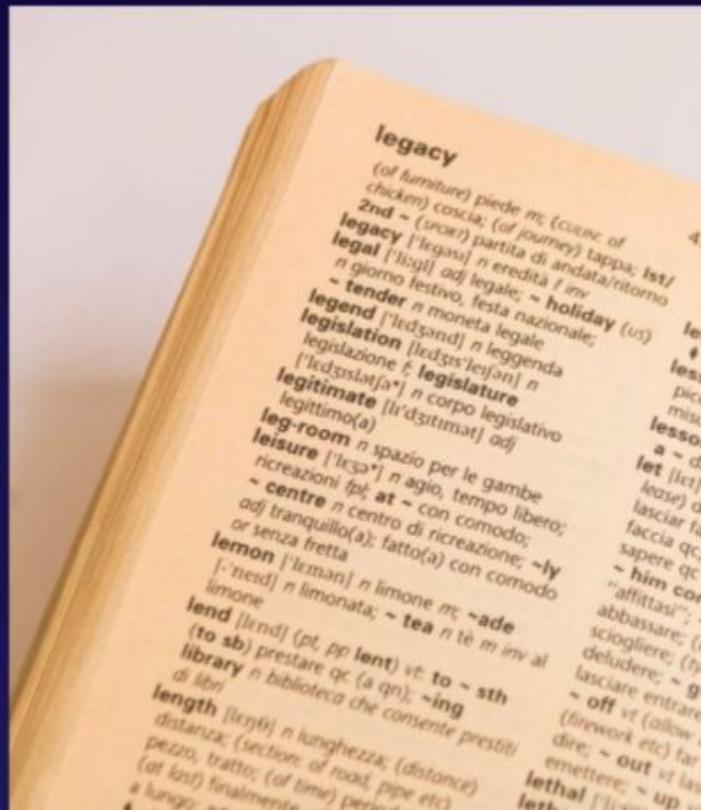
Policies

Policies are made up of documents, called Policy documents. These documents are in a format called JSON and they give permissions as to what a User/Group/Role is able to do.

4

Roles

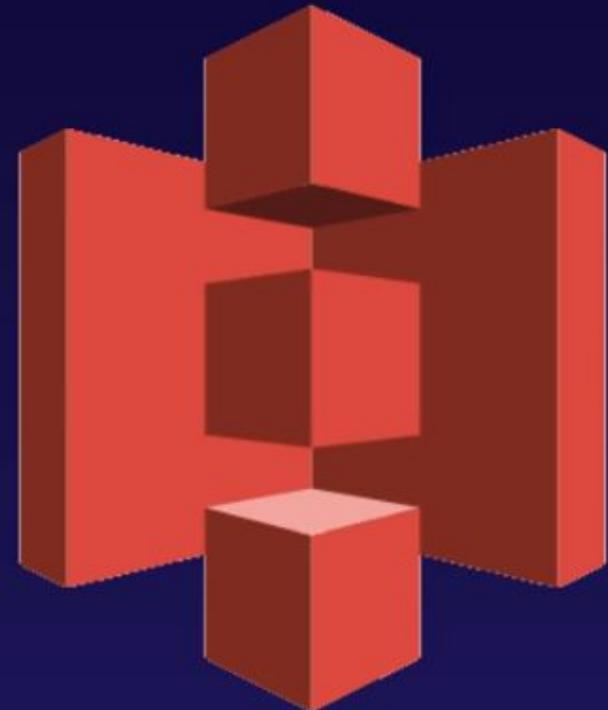
You create roles and then assign them to AWS Resources.



S3 101

What is S3?

S3 provides developers and IT teams with secure, durable, highly-scalable object storage. Amazon S3 is easy to use, with a simple web services interface to store and retrieve any amount of data from anywhere on the web.



What is S3 the basics

- S3 is a safe place to store your files.
- It is Object-based storage.
- The data is spread across multiple devices and facilities.



What is S3 the basics

The **basics of S3** are as follows;

- S3 is **Object-based** — i.e. allows you to upload files.
- Files can be from 0 Bytes to 5 TB.
- There is unlimited storage.
- Files are stored in Buckets.



What is S3 the basics

The **basics of S3** are as follows;

- **S3 is a universal namespace.** That is, names must be unique globally.

- <https://ghazelatechacademy.s3.amazonaws.com/>

- <https://ghazelatechacademy.eu-west-1.amazonaws.com/>

- When you upload a file to S3, you will receive a
- **HTTP 200 code** if the upload was successful.



S3 - Objects

S3 is Object based. **Think of Objects just as files.**

Objects consist of the following:

- Key (This is simply the name of the object)
- Value (This is simply the data and is made up of a sequence of bytes).
- Version ID (Important for versioning)
- Metadata (Data about data you are storing)
- Subresources;

Access Control Lists

Torrent



Data Consistency Model for S3

In Other Words;

- If you write a new file and read it immediately afterwards, you will be able to view that data.
- If you update **AN EXISTING file** or delete a file and read it immediately, you may get the older version, or you may not. Basically changes to objects can take a little bit of time to propagate.



S3 Guarantees

S3 has the following guarantees from Amazon;

- Built for 99.99% availability for the S3 platform.
- Amazon Guarantee 99.9% availability
- Amazon guarantees 99.999999999% durability for S3 information. (Remember 11 x 9s).



S3 Features

S3 has the following features;

- Tiered Storage Available
- Lifecycle Management
- Versioning
- Encryption
- MFA Delete
- Secure your data using **Access Control Lists** and **Bucket Policies**



S3 Storage Classes

1

S3 Standard

99.99% availability
99.99999999% durability,
stored redundantly across
multiple devices in multiple
facilities, and is designed to
sustain the loss of 2 facilities
concurrently.

2

S3 - IA

(Infrequently Accessed):
For data that is accessed
less frequently, but requires
rapid access when needed.
Lower fee than S3, but you
are charged a retrieval fee.

3

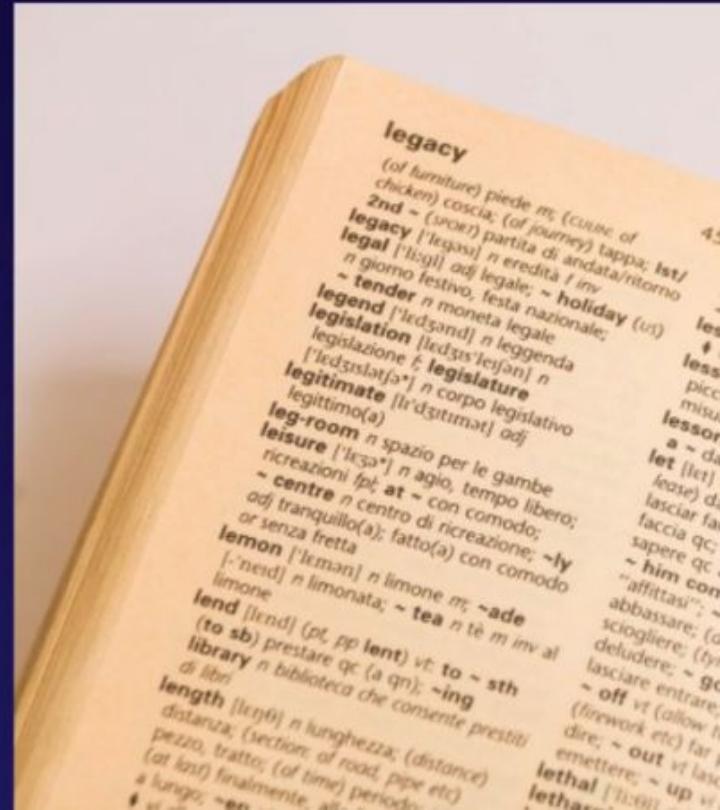
S3 One Zone - IA

For where you want a
lower-cost option for
infrequently accessed data,
but do not require the
multiple Availability Zone
data resilience.

4

S3 - Intelligent Tiering

Designed to optimize costs by
automatically moving data to the
most cost-effective access tier,
without performance impact or
operational overhead.



S3 Storage Classes

5

S3 Glacier

S3 Glacier is a secure, durable, and low-cost storage class for data archiving. You can reliably store any amount of data at costs that are competitive with or cheaper than on-premises solutions. Retrieval times configurable from minutes to hours.

6

S3 Glacier Deep Archive

S3 Glacier Deep Archive is Amazon S3's lowest-cost storage class where a retrieval time of 12 hours is acceptable.



S3 Comparison

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier	S3 Glacier Deep Archive**
Designed for durability	99.99999999% (11 9's)					
Designed for availability	99.99%	99.9%	99.9%	99.5%	N/A	N/A
Availability SLA	99.9%	99%	99%	99%	N/A	N/A
Availability Zones	≥3	≥3	≥3	1	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	40KB	40KB
Minimum storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	select minutes or hours	select hours

S3 Charges

You are charged for S3 in the following ways;

- Storage
- Requests
- Storage Management Pricing
- Data Transfer Pricing
- Transfer Acceleration
- Cross Region Replication Pricing



S3 Cross region replications



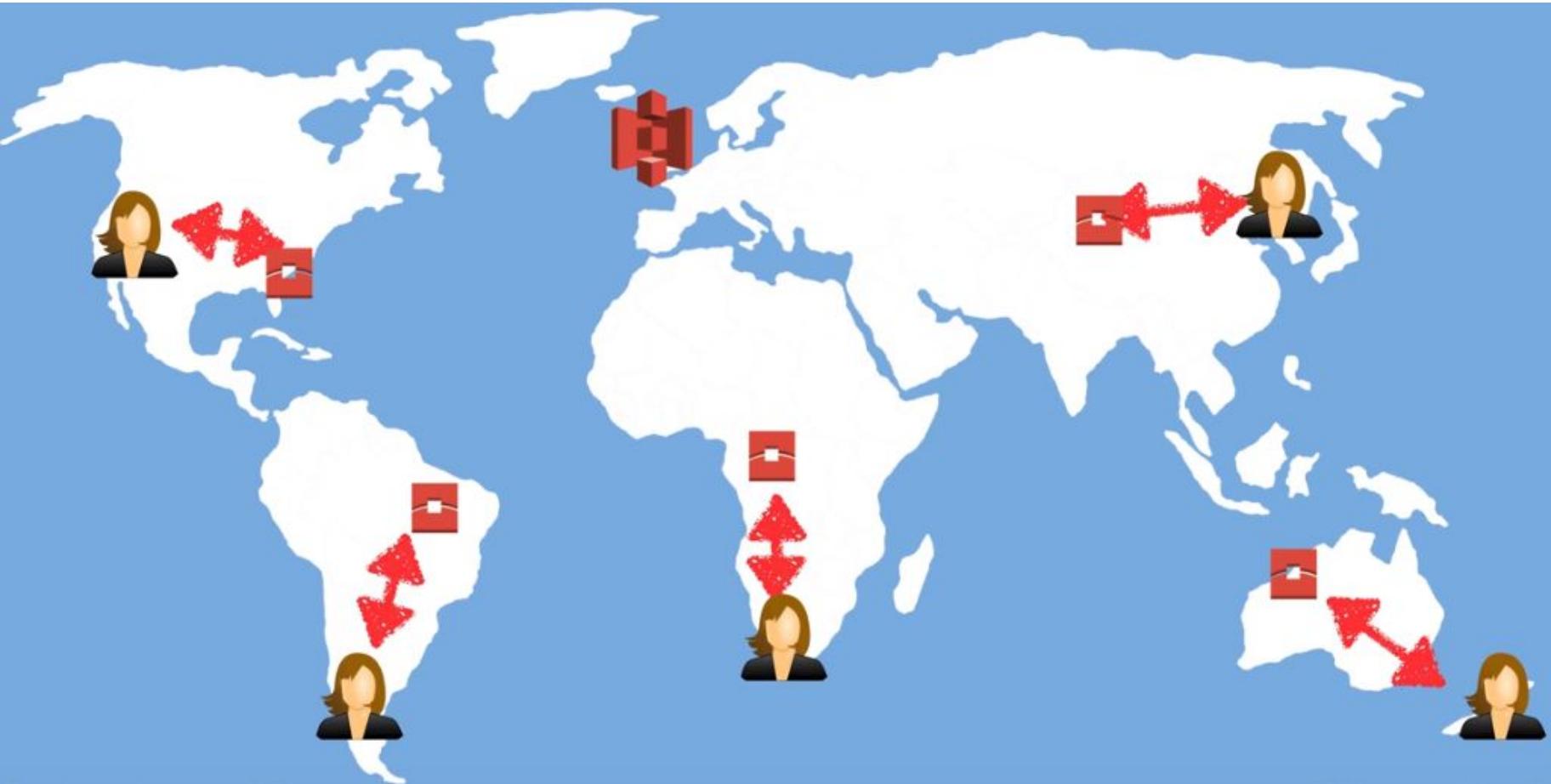
S3 Transfer Acceleration

Amazon S3 Transfer Acceleration enables fast, easy, and secure transfers of files over long distances between your end users and an S3 bucket.

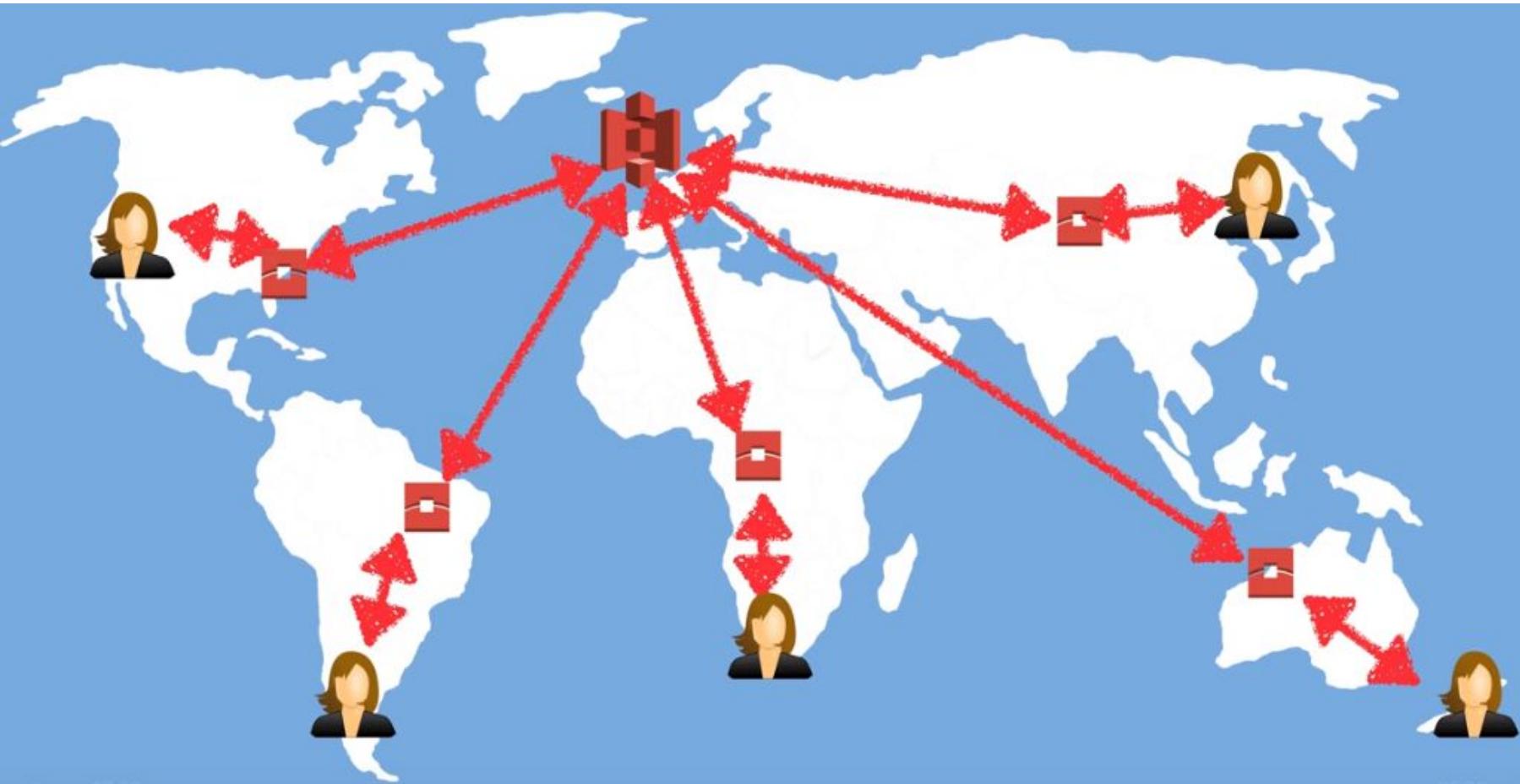
Transfer Acceleration takes advantage of Amazon CloudFront's globally distributed edge locations. As the data arrives at an edge location, data is routed to Amazon S3 over an optimized network path.



S3 Transfer Acceleration



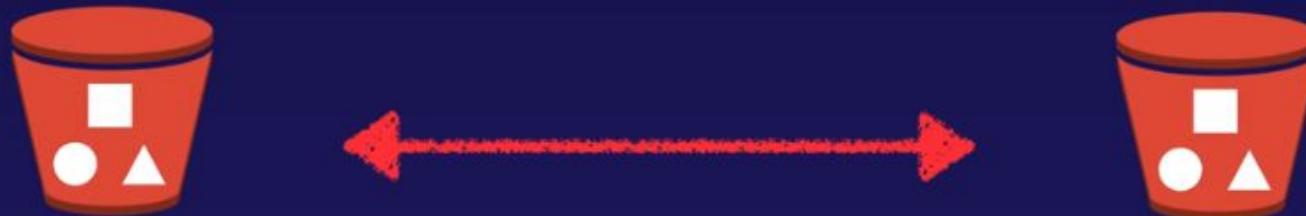
S3 Transfer Acceleration



S3 Cross-accounts Access

3 different ways to share S3 buckets across accounts

- Using Bucket Policies & IAM (applies across the entire bucket).
Programmatic Access Only.
- Using Bucket ACLs & IAM (individual objects). Programmatic Access Only.
- Cross-account IAM Roles. Programmatic AND Console access.



S3 Exam Tips

- Remember that S3 is **Object-based**: i.e. allows you to upload files.
- Files can be from 0 Bytes to 5 TB.
- There is unlimited storage.
- Files are stored in Buckets.
- **S3 is a universal namespace**. That is, names must be unique globally.

S3 Exam Tips

- Not suitable to install an operating system on.
- Successful uploads will generate a **HTTP 200** status code.
- You can turn on **MFA Delete**

S3 Exam Tips

The Key Fundamentals of S3 Are;

- Key (This is simply the name of the object)
- Value (This is simply the data and is made up of a sequence of bytes).
- Version ID (Important for versioning)
- Metadata (Data about data you are storing)
- Subresources;

Access Control Lists

Torrent

S3 Exam Tips

- Read after Write consistency for PUTS of new Objects
- Eventual Consistency for overwrite PUTS and DELETES (can take some time to propagate)

S3 Exam Tips

1

S3 Standard

99.99% availability
99.99999999% durability,
stored redundantly across
multiple devices in multiple
facilities, and is designed to
sustain the loss of 2 facilities
concurrently.

2

S3 - IA

(Infrequently Accessed):
For data that is accessed
less frequently, but requires
rapid access when needed.
Lower fee than S3, but you
are charged a retrieval fee.

3

S3 One Zone - IA

For where you want a
lower-cost option for
infrequently accessed data,
but do not require the
multiple Availability Zone
data resilience.

4

S3 - Intelligent Tiering

Designed to optimize costs
by automatically moving
data to the most cost-
effective access tier, without
performance impact or
operational overhead.

5

S3 Glacier

S3 Glacier is a secure, durable,
and low-cost storage class for
data archiving. Retrieval times
configurable from minutes to
hours.

6

S3 Glacier Deep Archive

S3 Glacier Deep Archive is
Amazon S3's lowest-cost
storage class where a
retrieval time of 12 hours is
acceptable.

S3 Exam Tips

- Read the S3 FAQs before taking the exam. It comes up A LOT!

S3 Exam Tips

Using **Versioning** With S3;

- Stores all versions of an object (including all writes and even if you delete an object)
- Great backup tool.
- Once enabled, **Versioning cannot be disabled**, only suspended.
- Integrates with **Lifecycle** rules
- Versioning's **MFA Delete** capability, which uses multi-factor authentication, can be used to provide an additional layer of security.



S3 Exam Tips - Lifecycle

- Automates moving your objects between the different storage tiers.
- Can be used in conjunction with versioning.
- Can be applied to current versions and previous versions.

S3 Exam Tips - Cross-Region Replication

- Versioning must be enabled on both the source and destination buckets.
- Files in an existing bucket are not replicated automatically.
- All subsequent updated files will be replicated automatically.
- Delete markers are not replicated.
- Deleting individual versions or delete markers will not be replicated.
- Understand what Cross Region Replication is at a high level.

S3 Security and Encryption

S3 basics

By default, all newly created buckets are **PRIVATE**. You can setup access control to your buckets using;

- **Bucket Policies**
- **Access Control Lists**

S3 buckets can be configured to create access logs which log all requests made to the S3 bucket. This can be sent to another bucket and even another bucket in another account.

S3 basics

Encryption In Transit is achieved by

- **SSL/TLS**

Encryption At Rest (Server Side) is achieved by

- **S3 Managed Keys - SSE-S3**
- **AWS Key Management Service, Managed Keys - SSE-KMS**
- **Server Side Encryption With Customer Provided Keys - SSE-C**

Client Side Encryption



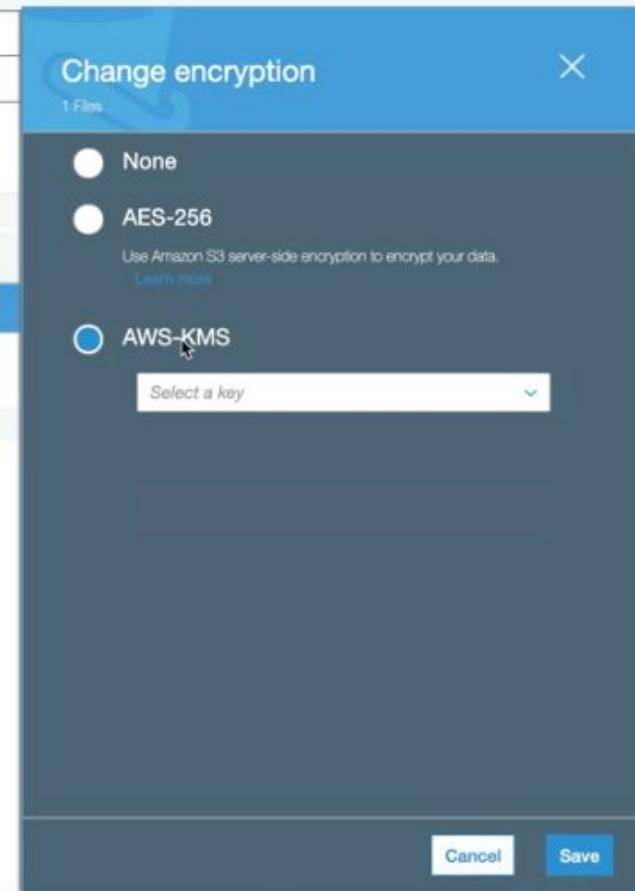
S3 basics

Overview Properties Permissions Management

Type a prefix and press Enter to search. Press ESC to clear.

Upload + Create folder Download Actions ▾

Name	Last modified
<input checked="" type="checkbox"/> fayeryan-replay.jpg	Jan 7, 2019 4:01:29 PM GMT+0000
<input type="checkbox"/> jeffbarr.jpg	Jan 7, 2019 2:40:56 PM GMT+0000



S3 Object Lock and Glacier Vault Lock

[SAA-C02]

S3 Object Lock?

You can use S3 Object Lock to store objects using a **write once, read many (WORM)** model. It can help you prevent objects from being deleted or modified for a fixed amount of time or indefinitely.

You can use **S3 Object Lock** to meet regulatory requirements that require WORM storage, or add an extra layer of protection against object changes and deletion.

S3 Object Lock Modes: Governance Mode

Governance Mode

In governance mode, **users can't overwrite or delete an object version or alter its lock settings** unless they have special permissions. With governance mode, you protect objects against being deleted by most users, but you can still grant some users **permission to alter the retention settings** or delete the object if necessary.



S3 Object Lock Modes: Compliance Mode

Compliance Mode

In compliance mode, **a protected object version can't be overwritten or deleted by any user**, including the root user in your AWS account. When an object is locked in compliance mode, its retention mode can't be changed and its retention period can't be shortened. Compliance mode ensures an object version **can't be overwritten or deleted** for the duration of the retention period.

S3 basics

Retention Periods

A retention period **protects an object version for a fixed amount of time**. When you place a retention period on an object version, Amazon S3 stores a timestamp in the object version's metadata to indicate when the retention period expires. After the retention period expires, the object version can be **overwritten or deleted** unless you also placed a legal hold on the object version.



S3 basics

Legal Holds

S3 Object Lock also enables you to place a legal hold on an object version. Like a retention period, **a legal hold prevents an object version from being overwritten or deleted**. However, a legal hold doesn't have an associated retention period and remains in effect until removed. Legal holds can be freely placed and removed by any user who has the s3 :PutObjectLegalHold permission.

Glacier Vault Lock

Glacier Vault Lock

S3 Glacier Vault Lock allows you to **easily deploy and enforce compliance controls for individual S3 Glacier vaults with a Vault Lock policy**. You can specify controls, such as WORM, in a Vault Lock policy and lock the policy from future edits. Once locked, the policy can no longer be changed.

Exam tips

- ✓ Use **S3 Object Lock** to store objects using a write once, read many (WORM) model.
- ✓ Object locks can be on individual objects or **applied across the bucket** as a whole.
- ✓ Object locks come in two modes: **governance mode** and **compliance mode**.



Exam tips

With **governance mode**, users can't overwrite or delete an object version or alter its lock settings unless they have special permissions.

With **compliance mode**, a protected object version can't be overwritten or deleted by any user, including the root user in your AWS account.

Exam tips



S3 Glacier Vault Lock allows you to easily deploy and enforce compliance controls for individual S3 Glacier vaults with a Vault Lock policy. You can **specify controls such as WORM in a Vault Lock policy and lock the policy from future edits**. Once locked, the policy can no longer be changed.

S3 Performance

[SAA-C02]

What are S3 prefix?

S3 Prefix

- ✓ *mybucketname/folder1/subfolder1/myfile.jpg > /**folder1/subfolder1***
- ✓ *mybucketname/folder2/subfolder1/myfile.jpg > /**folder2/subfolder1***
- ✓ *mybucketname/folder3/myfile.jpg > /**folder3***
- ✓ *mybucketname/folder4/subfolder4/myfile.jpg > /**folder4/subfolder4***

S3 performance



S3 Performance

S3 has extremely low latency. You can get the first byte out of S3 within **100-200 milliseconds**

You can also achieve a high number of requests: **3,500 PUT/COPY/POST/DELETE** and **5,500 GET/HEAD** requests per second per prefix.

S3 performance

S3 Performance

1

You can get better performance by spreading your reads across **different prefixes**. For example, if you are using **two prefixes**, you can achieve **11,000 requests per second**.

2

If we used all **four prefixes** in the last example, you would achieve **22,000 requests per second**.

KMS request Rates

S3 LIMITATIONS WHEN USING KMS

- If you are using **SSE-KMS** to encrypt your objects in S3, you must keep in mind the **KMS limits**.
- When you **upload** a file, you will call **GenerateDataKey** in the KMS API.
- When you **download** a file, you will call **Decrypt** in the KMS API.

KMS request Rates

S3 Limitations When Using KMS

- ✓ Uploading/downloading will count toward the **KMS quota**.
- ✓ Region-specific, however, it's either **5,500, 10,000, or 30,000 requests per second**.
- ✓ Currently, you **cannot** request a quota increase for KMS.



S3 performance: Uploads

Multipart Uploads

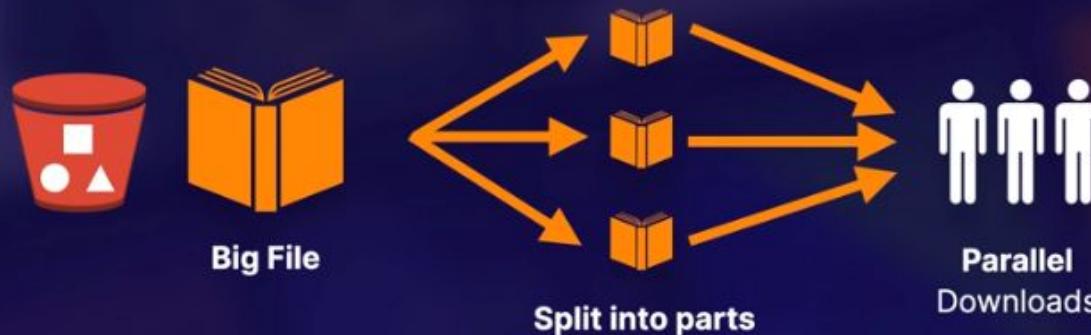
- Recommended for files **over 100 MB**
- Required for files **over 5 GB**
- Parallelize uploads (increases **efficiency**)



S3 performance: Downloads

S3 Byte-Range Fetches

- Parallelize **downloads** by specifying byte ranges.
- If there's a failure in the download, it's only for a specific byte range.



S3 performance: Downloads

S3 Byte-Range Fetches



Can be used to
speed up
downloads



Can be used to
just download
partial amounts of
the file (e.g.,
header
information)

Exam tips



mybucketname/folder1/subfolder1/myfile.jpg > /folder1/subfolder1



You can also achieve a high number of requests: **3,500 PUT/COPY/POST/DELETE** and **5,500 GET/HEAD** requests per second per prefix.



You can get better performance by spreading your reads across **different prefixes**. For example, if you are using **two prefixes**, you can achieve **11,000 requests per second**.

Exam tips

If you are using SSE-KMS to encrypt your objects in S3, you must keep in mind the KMS limits.



Uploading/downloading will count toward the **KMS quota**.



Region-specific, however, it's either **5,500, 10,000, or 30,000** requests per second.



Currently, you **cannot** request a quota increase for KMS.

Exam tips



Use **multipart uploads** to increase performance when **uploading files** to S3.



Should be used for any files **over 100 MB** and must be used for any file **over 5 GB**.



Use **S3 byte-range fetches** to increase performance when **downloading files** to S3.

S3 Select and Glacier Select

[SAA-C02]

What is S3 Select?

S3 Select enables applications to retrieve only a subset of data from an object by using simple SQL expressions. By using S3 Select to retrieve only the data needed by your application, you can **achieve drastic performance increases** — in many cases, you can get as much as a 400% improvement.



S3 Select example

Let's assume all your data is stored in S3 in zip files that contain CSV files. Without S3 Select, you would need to **download**, **decompress**, and **process the entire CSV** to get the data you needed.

S3 Select example

With S3 Select, you can use a **simple SQL expression** to return only the data from the store you're interested in instead of retrieving the entire object. This means you're dealing with an order of magnitude less data, which **improves the performance of your underlying applications**.



Glacier Select example

Some companies in **highly regulated industries** — e.g., financial services, healthcare, and others — write data directly to Amazon Glacier to satisfy compliance needs like SEC Rule 17a-4 or HIPAA. Many S3 users have lifecycle policies **designed to save on storage costs** by moving their data into Glacier when they no longer need to access it on a regular basis.

Glacier Select allows you to run SQL queries against Glacier directly.



Exam tips



Remember that S3 Select is used to **retrieve only a subset of data** from an object by using simple SQL expressions.



Get data by **rows or columns** using simple SQL expressions.



Save money on **data transfer** and increase speed.



AWS Organizations and Consolidated Billing

[SAA-C02]

Overview of AWS organization

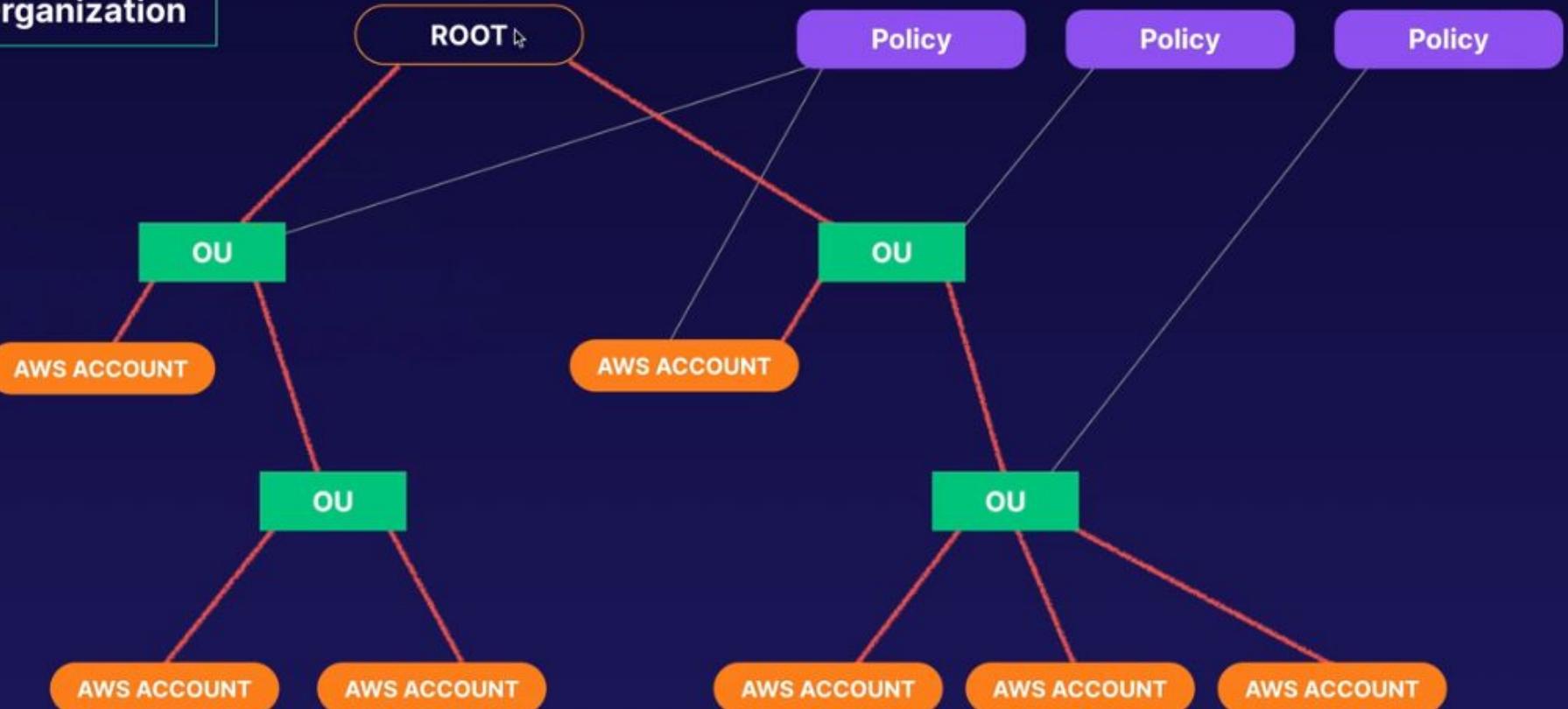
What is AWS Organizations?

“AWS Organizations is an account management service that enables you to consolidate multiple AWS accounts into an organization that you create and centrally manage.”



Overview of AWS organization

Organization



Consolidated Billing



Paying account is independent.
Cannot access resources of the other accounts.

All linked accounts are independent.

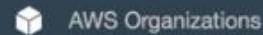
Consolidated Billing

Advantages of Consolidated Billing

- One bill per AWS account
- Very easy to track charges and allocate costs
- Volume pricing discount



Create new organization



Welcome

You can work with AWS Organizations in any of the following ways:

AWS Organizations Query API
AWS Management Console
AWS Command Line Tools
AWS SDKs

Invitations 0

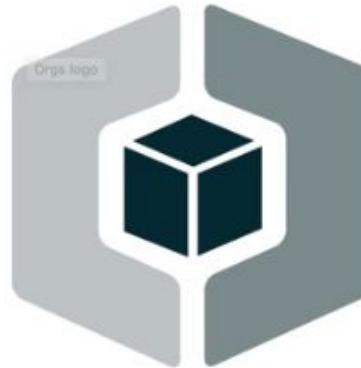
AWS Organizations

AWS Organizations enables you to centrally apply policy-based controls across multiple accounts in the AWS Cloud.

You can consolidate all your AWS accounts into an organization, and arrange all AWS accounts into distinct organizational units.

[Create organization](#)

[User guide](#)



1

Create accounts

Invite existing AWS accounts to join your organization or create new accounts based on your needs.
[Learn more](#)

2

Organize accounts

Model your organization by defining one or more organizational units that best represent your business requirements.
[Learn more](#)

3

Apply policies

You can assign a policy to your organizational units, or individual AWS accounts. [Learn more](#)

Invite account

AWS Organizations



How do you want to add an account to your organization?

Invite account

Invite an existing AWS account to join your organization.

Create account

Create an AWS account in this organization.

Email or account ID*

Enter multiple email addresses or account IDs separated by commas.

Notes

You may optionally include a note with your request.

* Required fields

Cancel

Invite

Exam tips

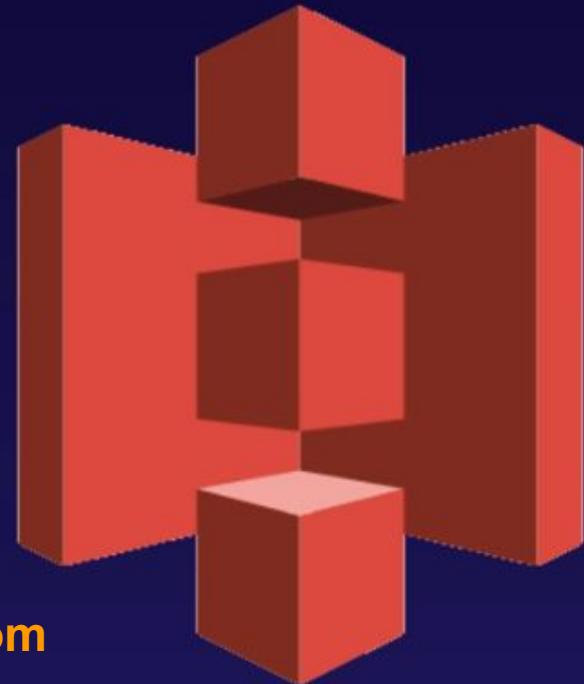
Some Best Practices With AWS Organizations

- Always enable multi-factor authentication on root account.
- Always use a strong and complex password on root account.
- Paying account should be used for billing purposes only. Do not deploy resources into the paying account.
- Enable/Disable AWS services using Service Control Policies (SCP) either on OU or on individual accounts.

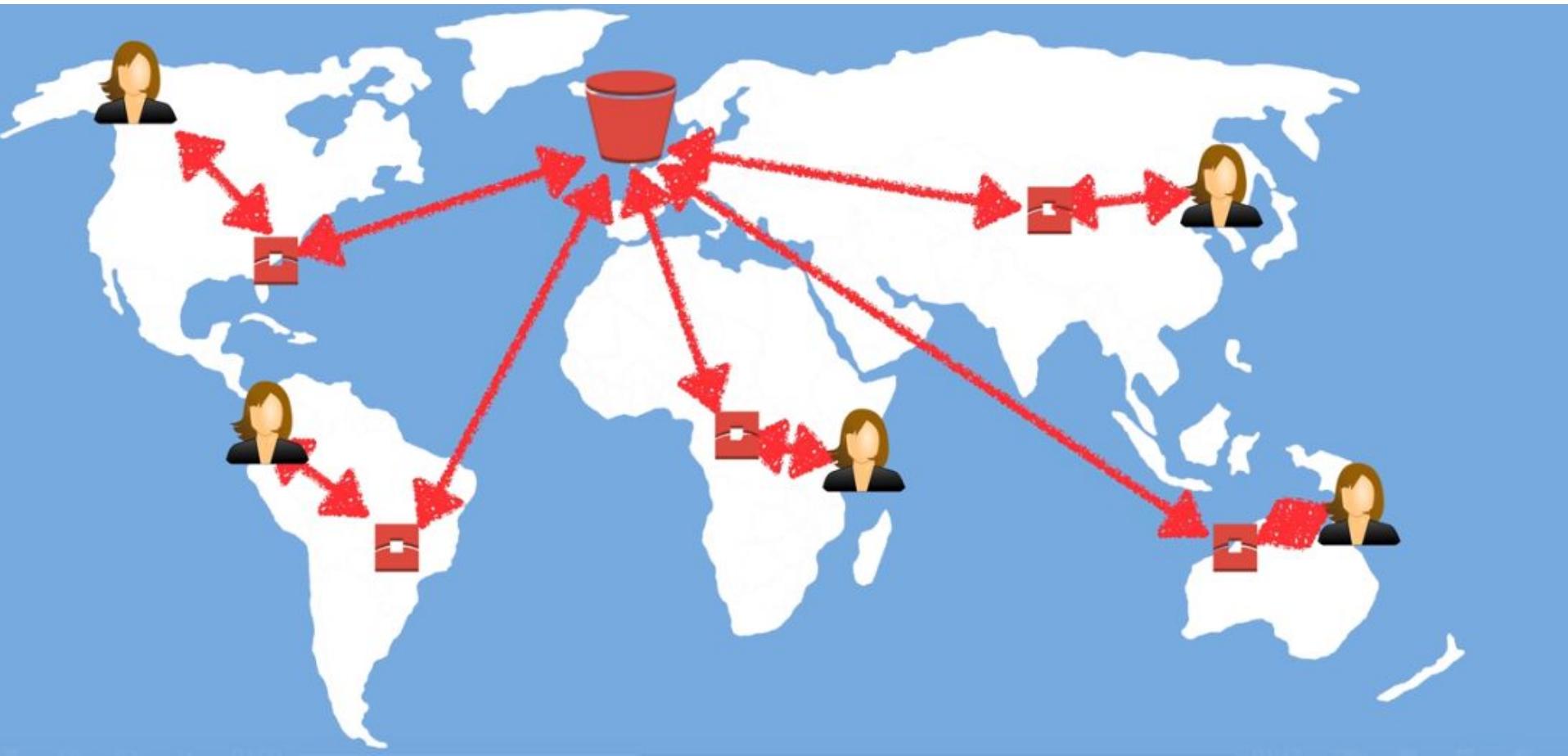
S3 Transfer Acceleration

S3 Transfer Acceleration

S3 Transfer Acceleration utilises the CloudFront Edge Network to accelerate your uploads to S3.
Instead of uploading directly to your S3 bucket, you can use a distinct URL to upload directly to an edge location which will then transfer that file to S3. You will get a distinct URL to upload to: **itpeac.s3-accelerate.amazonaws.com**



S3 Transfer Acceleration



S3 Transfer Acceleration tool



Amazon S3 Transfer Acceleration Speed Comparison

Upload speed comparison in the selected region

Virginia

(US-EAST-1)

S3 Direct Upload Speed

Pending

S3 Accelerated Transfer Upload Speed

Pending

This speed checker uses multipart uploads to transfer a file from your browser to various Amazon S3 regions with and without Amazon S3 Transfer Acceleration. It compares the speed results and shows the percentage difference for every region.

Note: In general, the farther away you are from an Amazon S3 region, the higher the speed improvement you can expect from using Amazon S3 Transfer Acceleration. If you see similar speed results with and without the acceleration, your upload bandwidth or a system constraint might be limiting your speed.

Upload speed comparison in other regions

San Francisco

(US-WEST-1)

Oregon

(US-WEST-2)

Dublin

(EU-WEST-1)

S3 Transfer Acceleration tool



Upload speed comparison in the selected region

Virginia

(US-EAST-1)



137% faster

S3 Direct Upload Speed



Uploaded complete

S3 Accelerated Transfer Upload Speed



Uploaded complete

This speed checker uses multipart uploads to transfer a file from your browser to various Amazon S3 regions with and without Amazon S3 Transfer Acceleration. It compares the speed results and shows the percentage difference for every region.

Note: In general, the farther away you are from an Amazon S3 region, the higher the speed improvement you can expect from using Amazon S3 Transfer Acceleration. If you see similar speed results with and without the acceleration, your upload bandwidth or a system constraint might be limiting your speed.

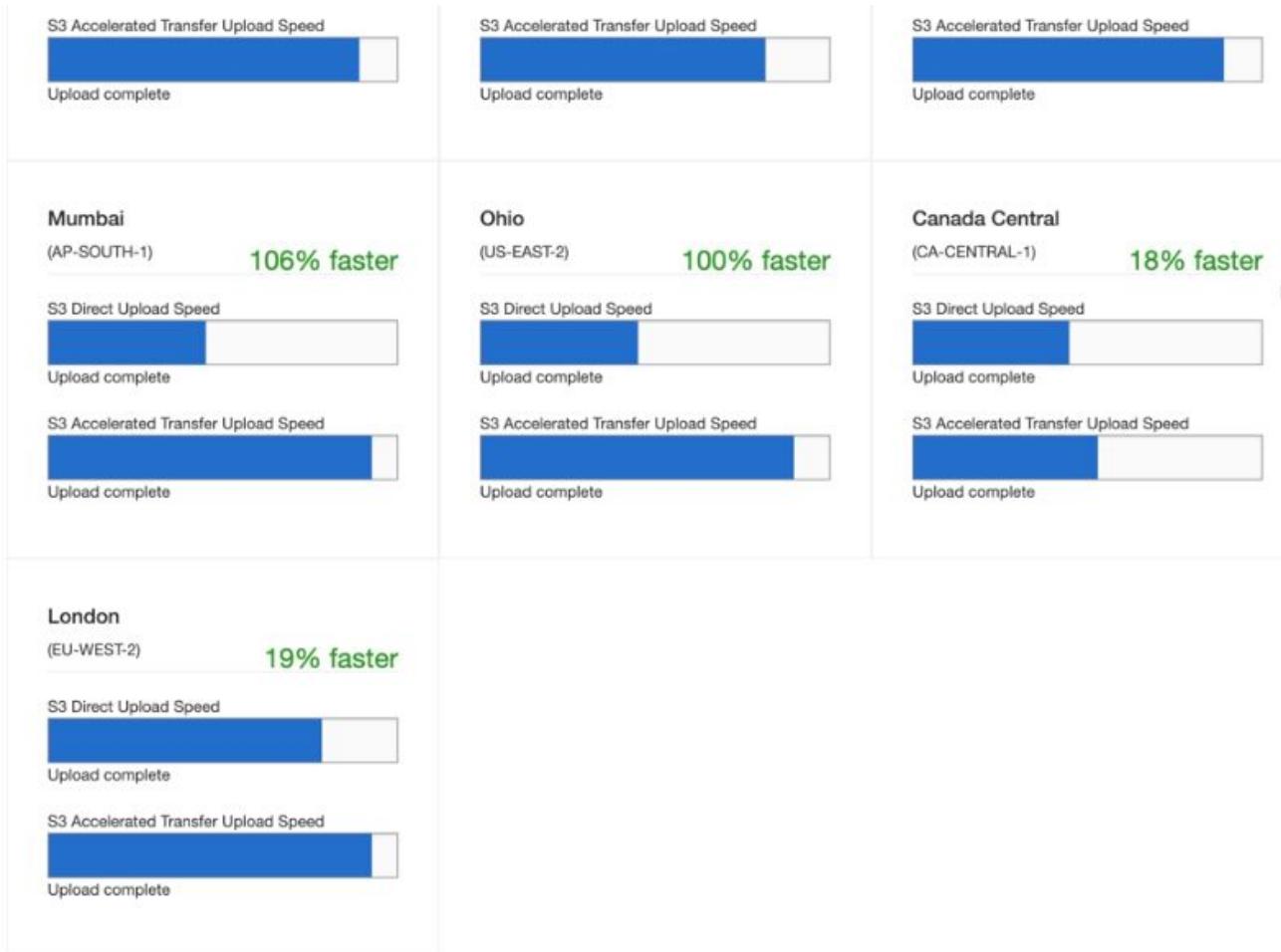
Upload speed comparison in other regions

San Francisco

Oregon

Dublin

S3 Transfer Acceleration tool



AWS DataSync

[SAA-C02]

What is Aws Data Sync?



Exam tips



Used to move **large amounts** of data from on-premises to AWS.



Used with **NFS-** and **SMB**-compatible file systems.



Replication can be done hourly, daily, or weekly.



Install the **DataSync agent** to start the replication.



Can be used to replicate **EFS to EFS**.

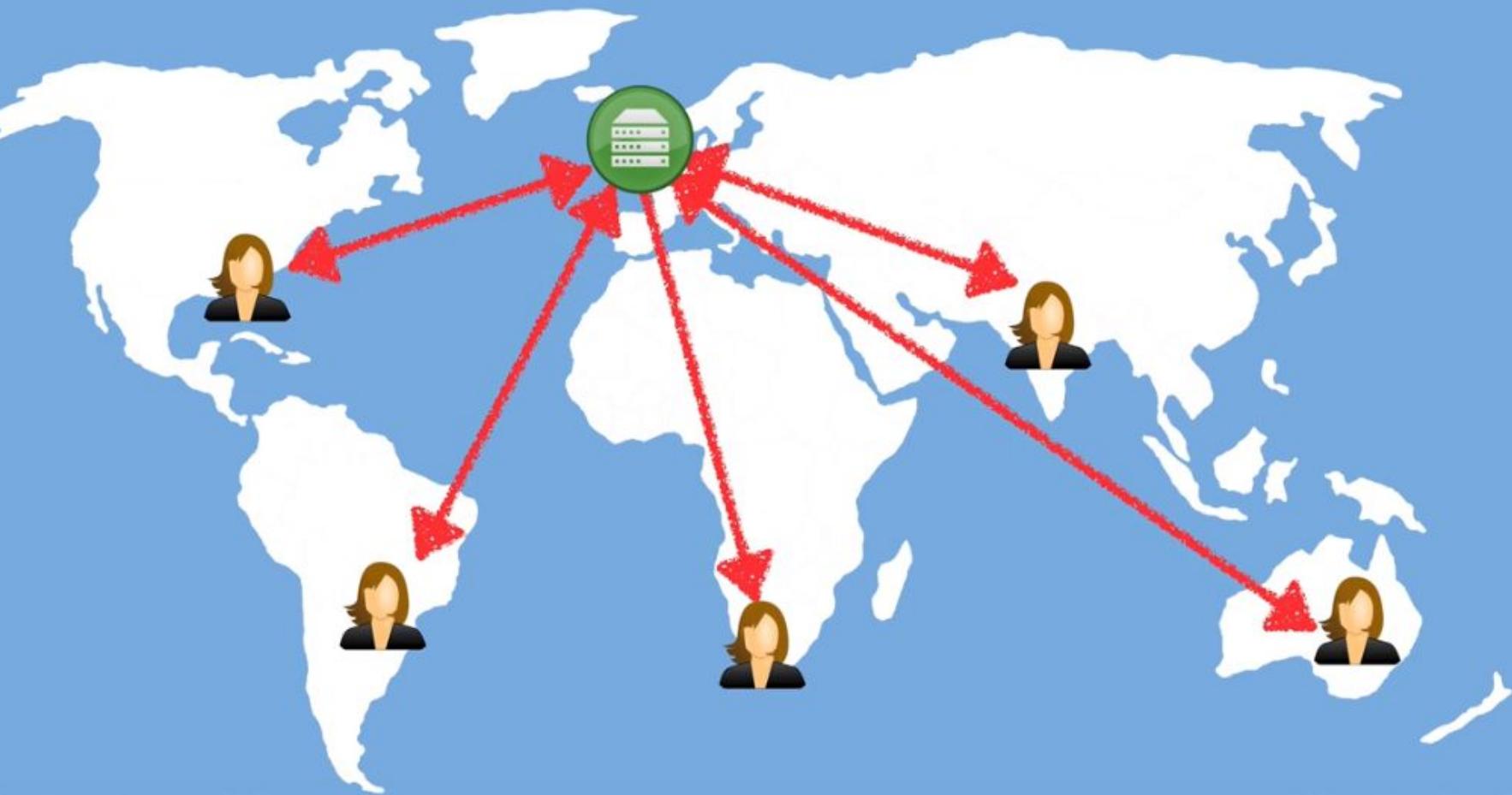
CloudFront

What is Cloudfront?

A content delivery network (CDN) is a system of distributed servers (network) that deliver webpages and other web content to a user based on the geographic locations of the user, the origin of the webpage, and a content delivery server.

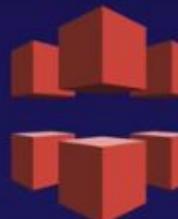


What is a CDN?

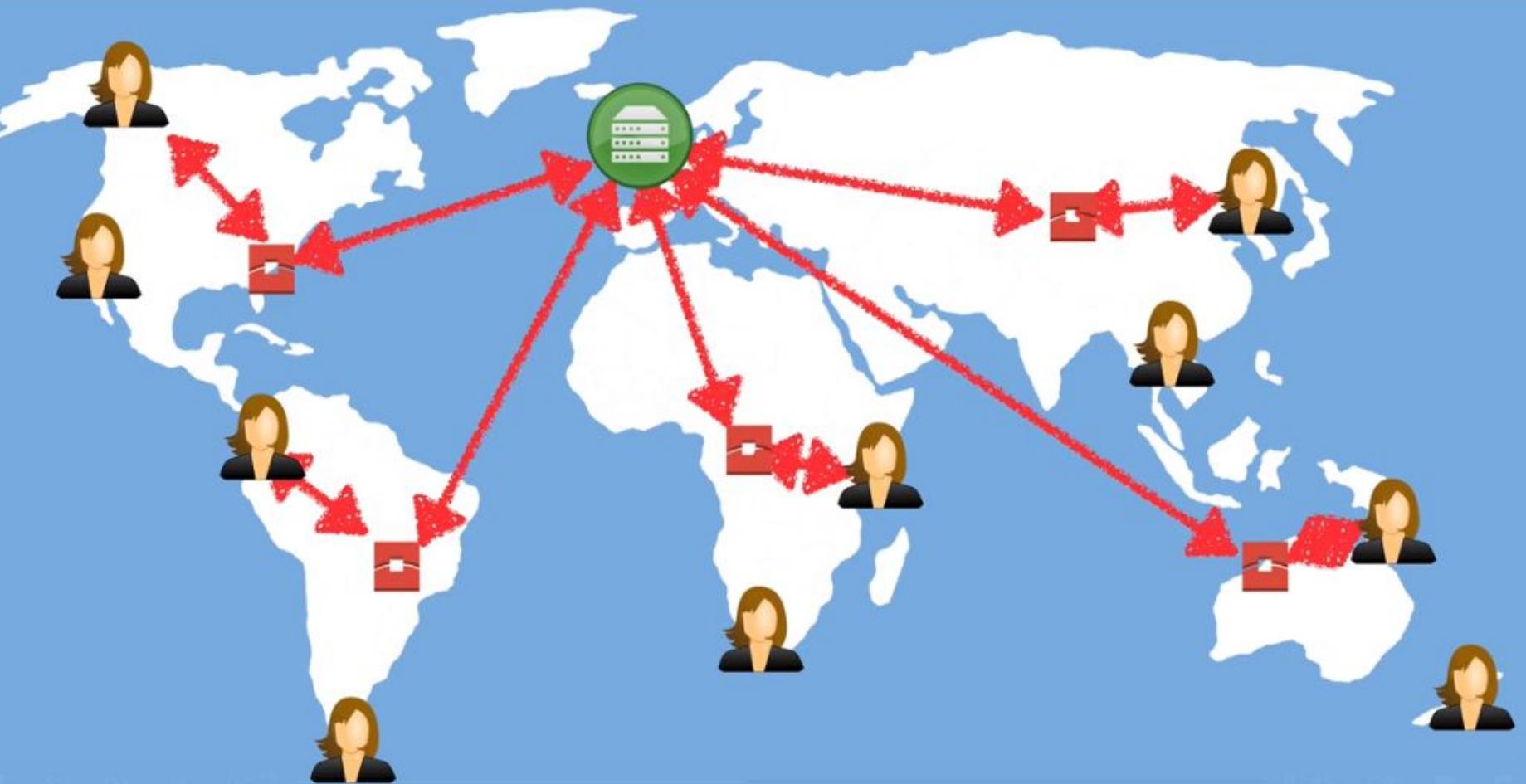


CloudFront - Key terminology

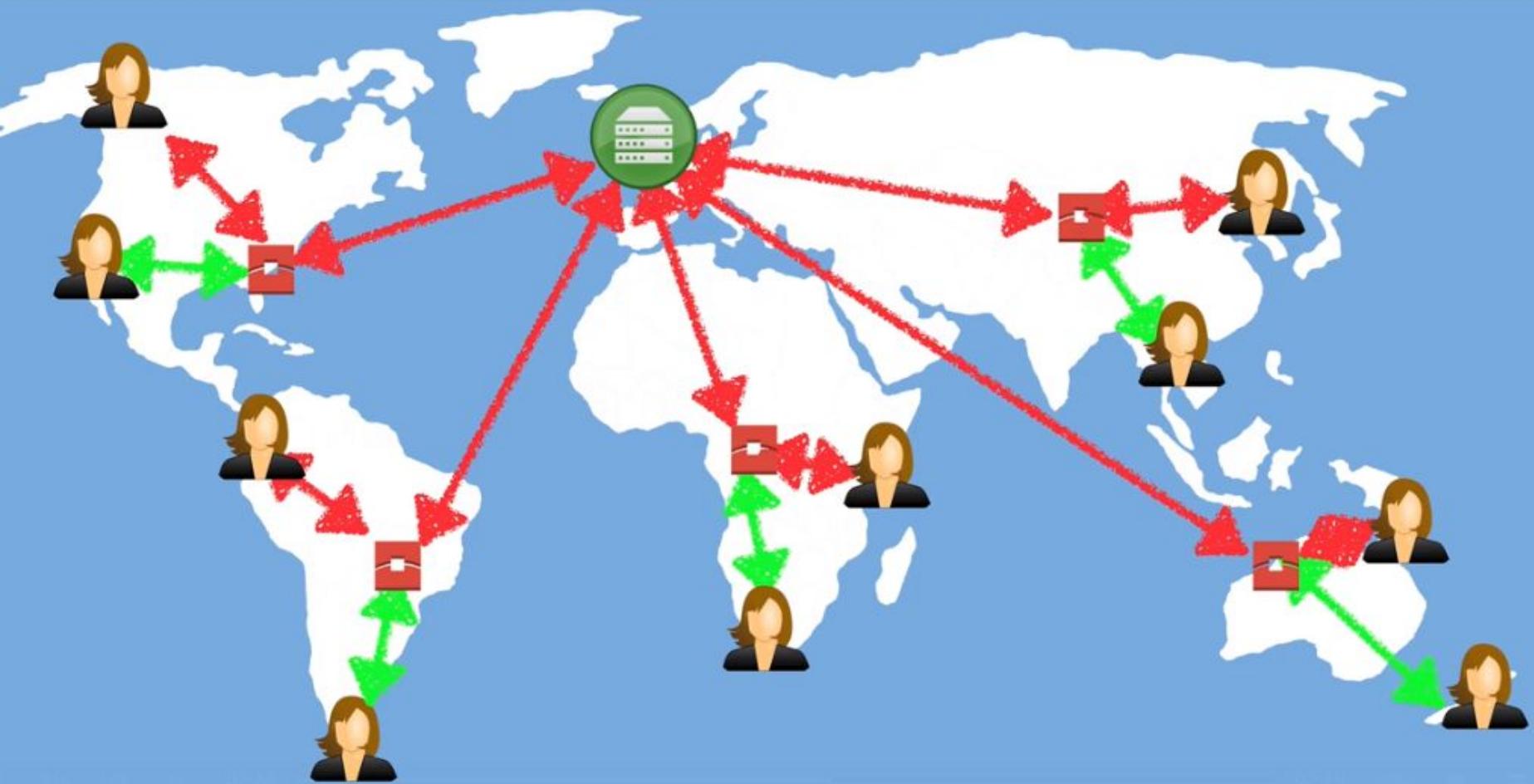
- **Edge Location** - This is the location where content will be cached. This is separate to an AWS Region/AZ.
- **Origin** - This is the origin of all the files that the CDN will distribute. This can be an S3 Bucket, an EC2 Instance, an Elastic Load Balancer, or Route53.
- **Distribution** - This is the name given the CDN which consists of a collection of Edge Locations.



CloudFront - How it works?



CloudFront - How it works?



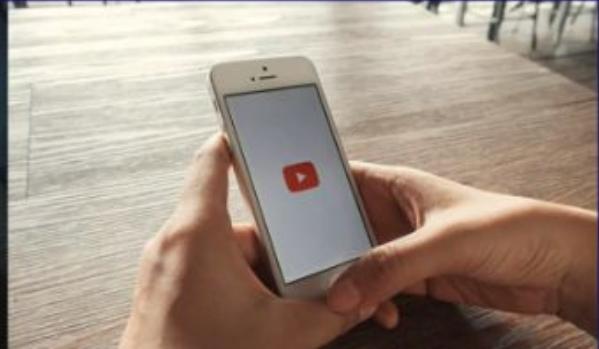
CloudFront - How it works?

Amazon CloudFront can be used to deliver your entire website, including dynamic, static, streaming, and interactive content using a global network of edge locations. Requests for your content are automatically routed to the nearest edge location, so content is delivered with the best possible performance.



CloudFront - How it works?

- Web Distribution - Typically used for Websites.
- RTMP - Used for Media Streaming.



Exam tips

- **Edge Location** - This is the location where content will be cached. This is separate to an AWS Region/AZ.
- **Origin** - This is the origin of all the files that the CDN will distribute. This can be either an S3 Bucket, an EC2 Instance, an Elastic Load Balancer, or Route53.
- **Distribution** - This is the name given the CDN which consists of a collection of Edge Locations.
- **Web Distribution** - Typically used for Websites.
- **RTMP** - Used for Media Streaming.

Exam tips

- Edge locations are not just READ only — you can write to them too. (ie put an object on to them.)
- Objects are cached for the life of the **TTL (Time To Live.)**
- You can clear cached objects, but you will be charged.

CloudFront Signed URLs and Cookies

[SAA-C02]

CloudFront review

What We Have Learned So Far



Edge Location

This is the location where content will be cached.
This is different than an **AWS Region/AZ**.



Web Distribution

Typically used for websites.



Distribution

This is the name given the **CDN**, which consists of a collection of edge locations.



RTMP

Used for media streaming.



Origin

This is the origin of all the files the CDN will distribute.
This can be either an S3 bucket, an EC2 instance, an Elastic Load Balancer, or Route 53.

URLs vs. Cookies

Use CloudFront Signed URLs or Signed Cookies

- 1 A signed URL is for individual files.
1 file = 1 URL.

- 2 A signed cookie is for multiple files.
1 cookie = multiple files.



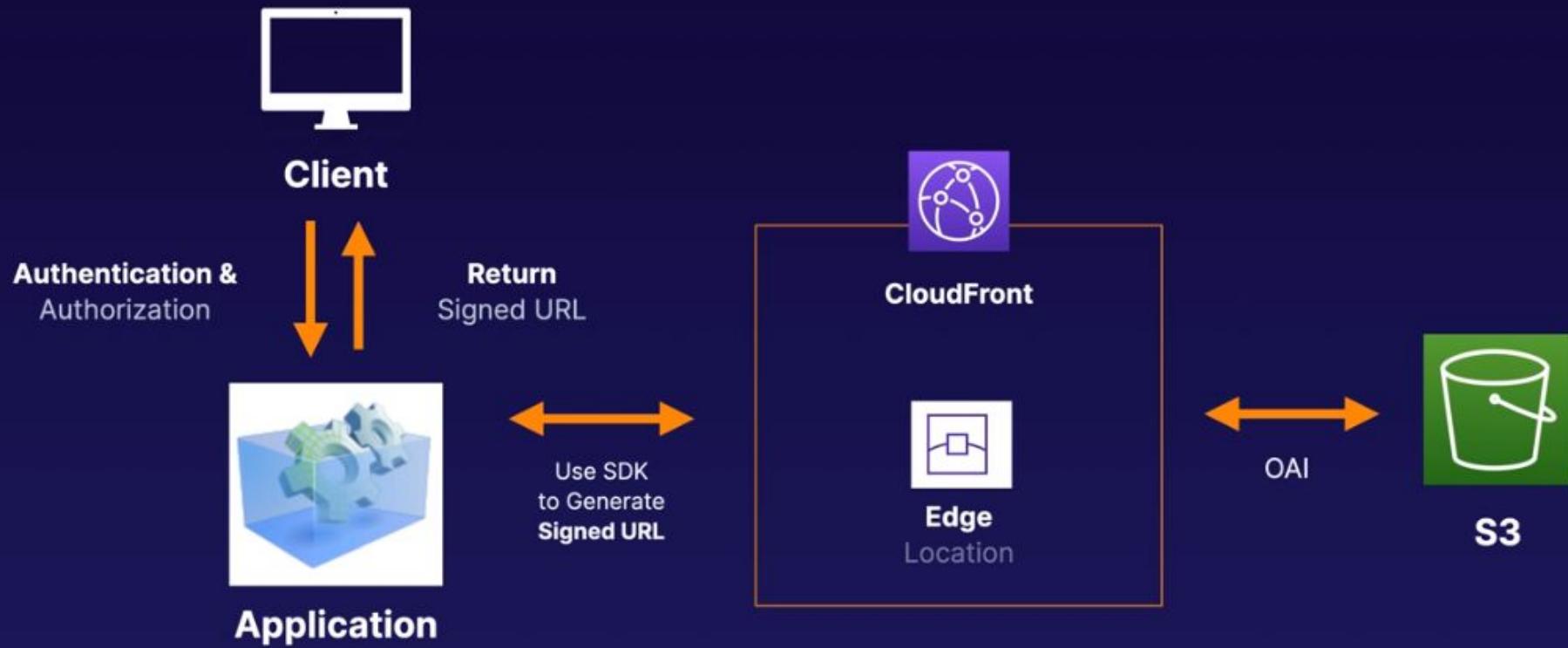
Signed URL Policies

When we create a signed URL or signed cookie, we attach a policy.

The policy can include:

- URL expiration
- IP ranges
- Trusted signers (which AWS accounts can create **signed URLs**)

How Signed URLs Work



CloudFront Signed URL Features

CloudFront Signed URL



Can have different origins.
Does not have to be **EC2**.



Key-pair is account wide and
managed by the root user



Can utilize **caching** features



Can filter by date, path, IP
address, expiration, etc.



S3 Signed URL Features

S3 Signed URL



Issues a request as the
IAM user who creates the
presigned URL



Limited
lifetime



Exam tips

- ✓ Use signed **URLs/cookies** when you want to secure content so that only the people you authorize are able to access it.
- ✓ A signed URL is for individual files. **1 file = 1 URL.**
- ✓ A signed cookie is for multiple files. **1 cookie = multiple files.**
- ✓ If your origin is EC2, then use CloudFront.

Snowball

[SAA-C02]

What is Snowball?



Snowball is a petabyte-scale data transport solution that uses secure appliances to transfer large amounts of data into and out of AWS. Using Snowball addresses common challenges with large-scale data transfers including high network costs, long transfer times, and security concerns. Transferring data with Snowball is simple, fast, secure, and can be as little as one-fifth the cost of high-speed Internet.

Snowball



Snowball comes in either a 50TB or 80TB size. Snowball uses multiple layers of security designed to protect your data including tamper-resistant enclosures, 256-bit encryption, and an industry-standard Trusted Platform Module (TPM) designed to ensure both security and full chain-of-custody of your data. Once the data transfer job has been processed and verified, AWS performs a software erasure of the Snowball appliance.

What is Snowball Edge?



AWS Snowball Edge is a 100TB data transfer device with on-board storage and compute capabilities. You can use Snowball Edge to move large amounts of data into and out of AWS, as a temporary storage tier for large local datasets, or to support local workloads in remote or offline locations.

Snowball Edge



Snowball Edge connects to your existing applications and infrastructure using standard storage interfaces, streamlining the data transfer process and minimizing setup and integration.

Snowball Edge can cluster together to form a local storage tier and process your data on-premises, helping ensure your applications continue to run even when they are not able to access the cloud.

What is Snowmobile?



AWS Snowmobile is an Exabyte-scale data transfer service used to move extremely large amounts of data to AWS. You can transfer up to 100PB per Snowmobile, a 45-foot long ruggedized shipping container, pulled by a semi-trailer truck. Snowmobile makes it easy to move massive volumes of data to the cloud, including video libraries, image repositories, or even a complete data center migration. Transferring data with Snowmobile is secure, fast and cost effective.

When should i use Snowball?

Available Internet Connection	Theoretical Min. Number of Days to Transfer 100TB at 80% Network Utilization	When to Consider AWS Import/Export Snowball?
T3 (44.736Mbps)	269 days	2TB or more
100Mbps	120 days	5TB or more
1000Mbps	12 days	60TB or more

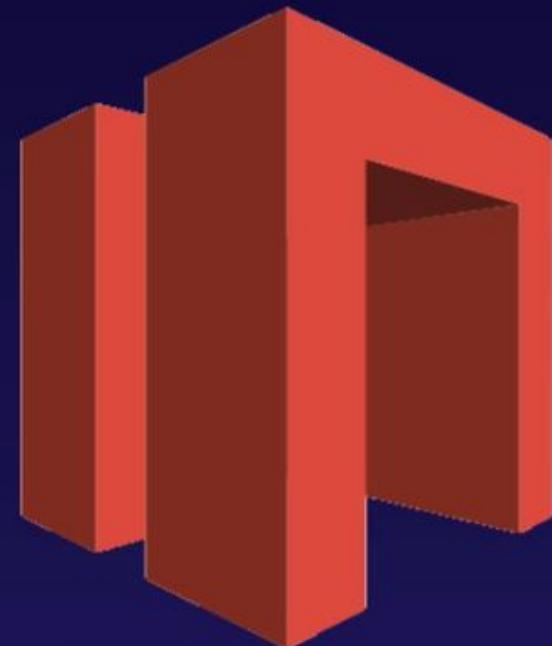
Exam tips

- Understand what Snowball is
- Snowball Can
- Import to S3
- Export from S3

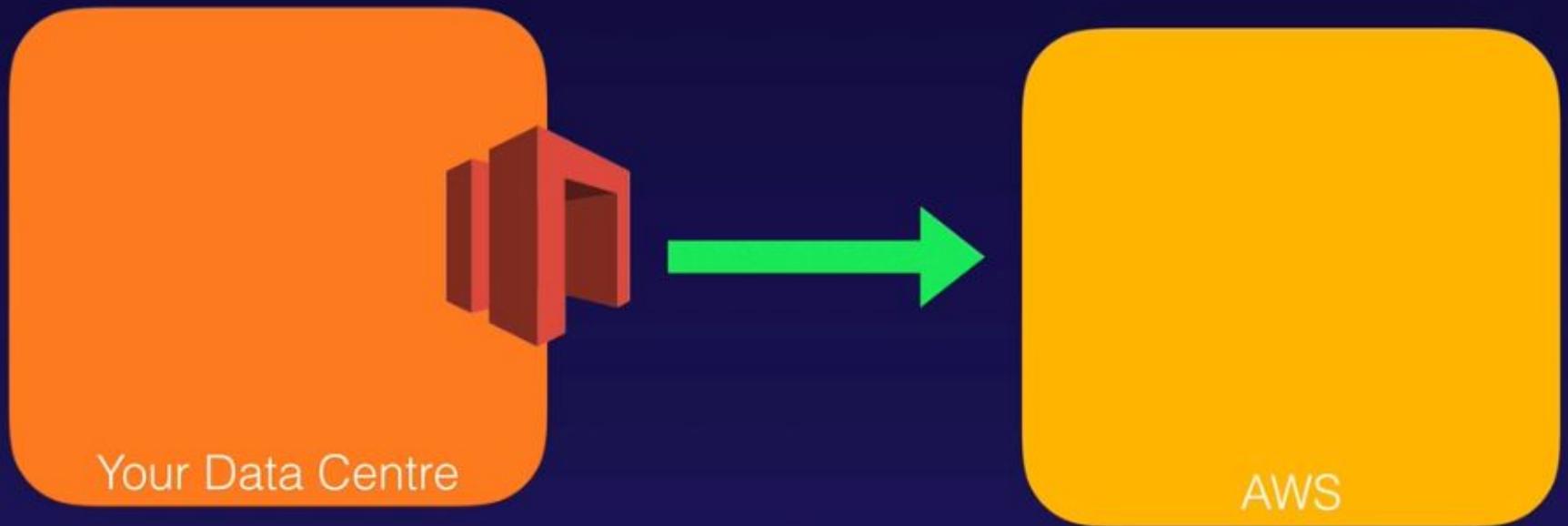
Storage Gateway

Storage Gateway

AWS Storage Gateway is a service that connects an on-premises software appliance with cloud-based storage to provide seamless and secure integration between an organization's on-premises IT environment and AWS's storage infrastructure. The service enables you to securely store data to the AWS cloud for scalable and cost-effective storage.



Storage Gateway



Storage Gateway

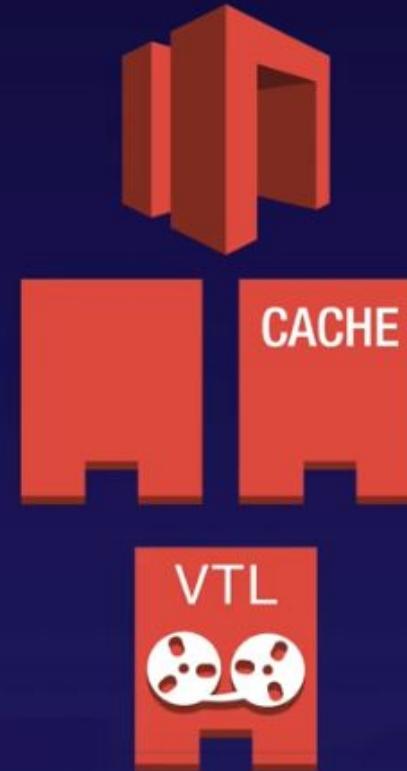


AWS Storage Gateway's software appliance is available for download as a virtual machine (VM) image that you install on a host in your datacenter. Storage Gateway supports either VMware ESXi or Microsoft Hyper-V. Once you've installed your gateway and associated it with your AWS account through the activation process, you can use the AWS Management Console to create the storage gateway option that is right for you.

Storage Gateway Types

The three different types of Storage Gateway are as follows;

- File Gateway (NFS & SMB)
- Volume Gateway (iSCSI)
 - Stored Volumes
 - Cached Volumes
- Tape Gateway (VTL)

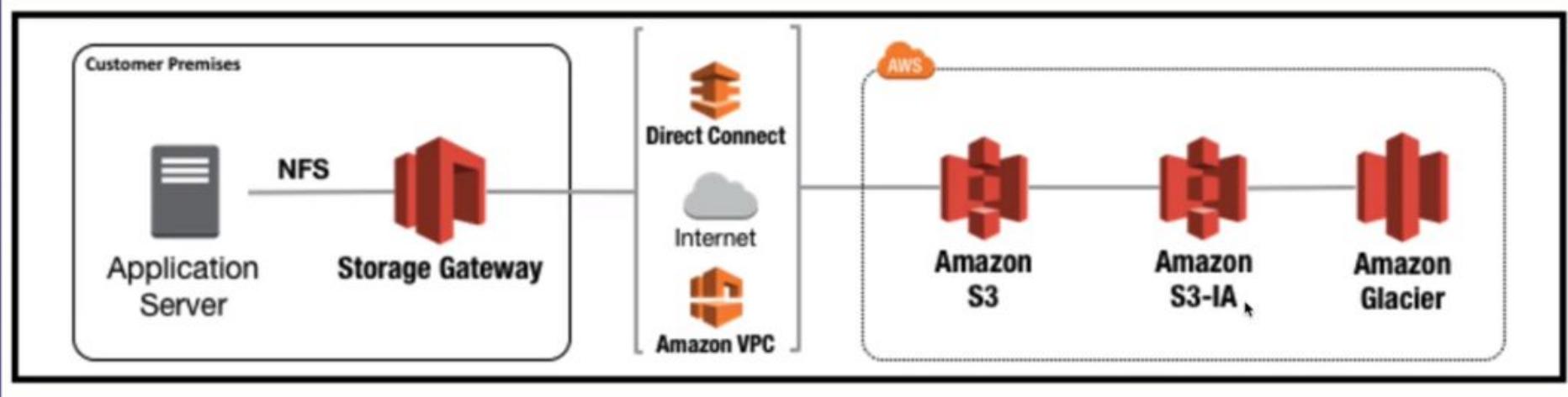


File Gateway



Files are stored as objects in your S3 buckets, accessed through a Network File System (NFS) mount point. Ownership, permissions, and timestamps are durably stored in S3 in the user-metadata of the object associated with the file. Once objects are transferred to S3, they can be managed as native S3 objects, and bucket policies such as versioning, lifecycle management, and cross-region replication apply directly to objects stored in your bucket.

File Gateway



Volume Gateway



The volume interface presents your applications with disk volumes using the iSCSI block protocol.

Data written to these volumes can be asynchronously backed up as point-in-time snapshots of your volumes, and stored in the cloud as Amazon EBS snapshots.

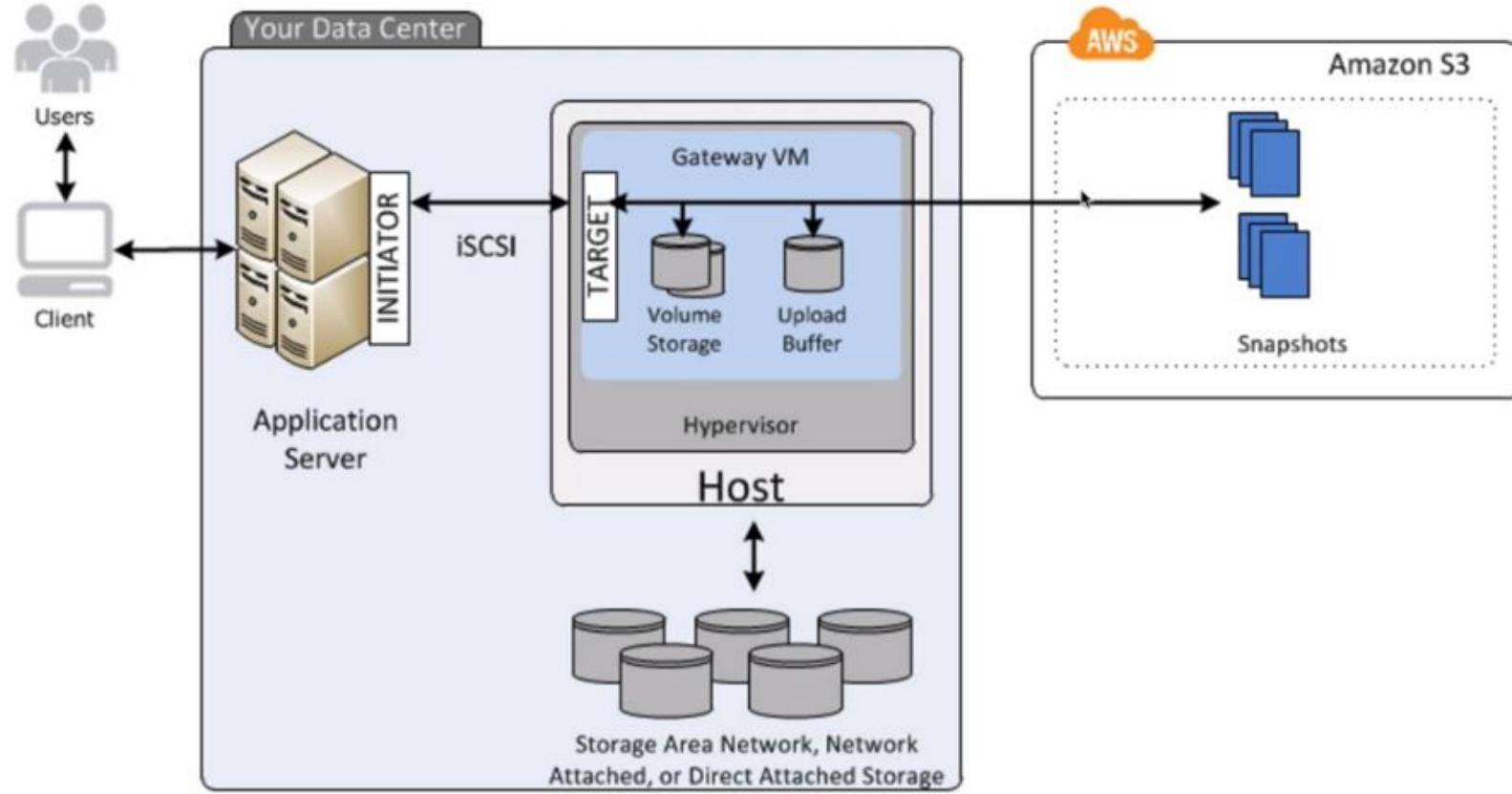
Snapshots are incremental backups that capture only changed blocks. All snapshot storage is also compressed to minimize your storage charges.

Volume Gateway - Stored Volumes



Stored volumes let you store your primary data locally, while asynchronously backing up that data to AWS. Stored volumes provide your on-premises applications with low-latency access to their entire datasets, while providing durable, off-site backups. You can create storage volumes and mount them as iSCSI devices from your on-premises application servers. Data written to your stored volumes is stored on your on-premises storage hardware. This data is asynchronously backed up to Amazon Simple Storage Service (Amazon S3) in the form of Amazon Elastic Block Store (Amazon EBS) snapshots. 1 GB - 16 TB in size for Stored Volumes.

Volume Gateway - Stored Volumes

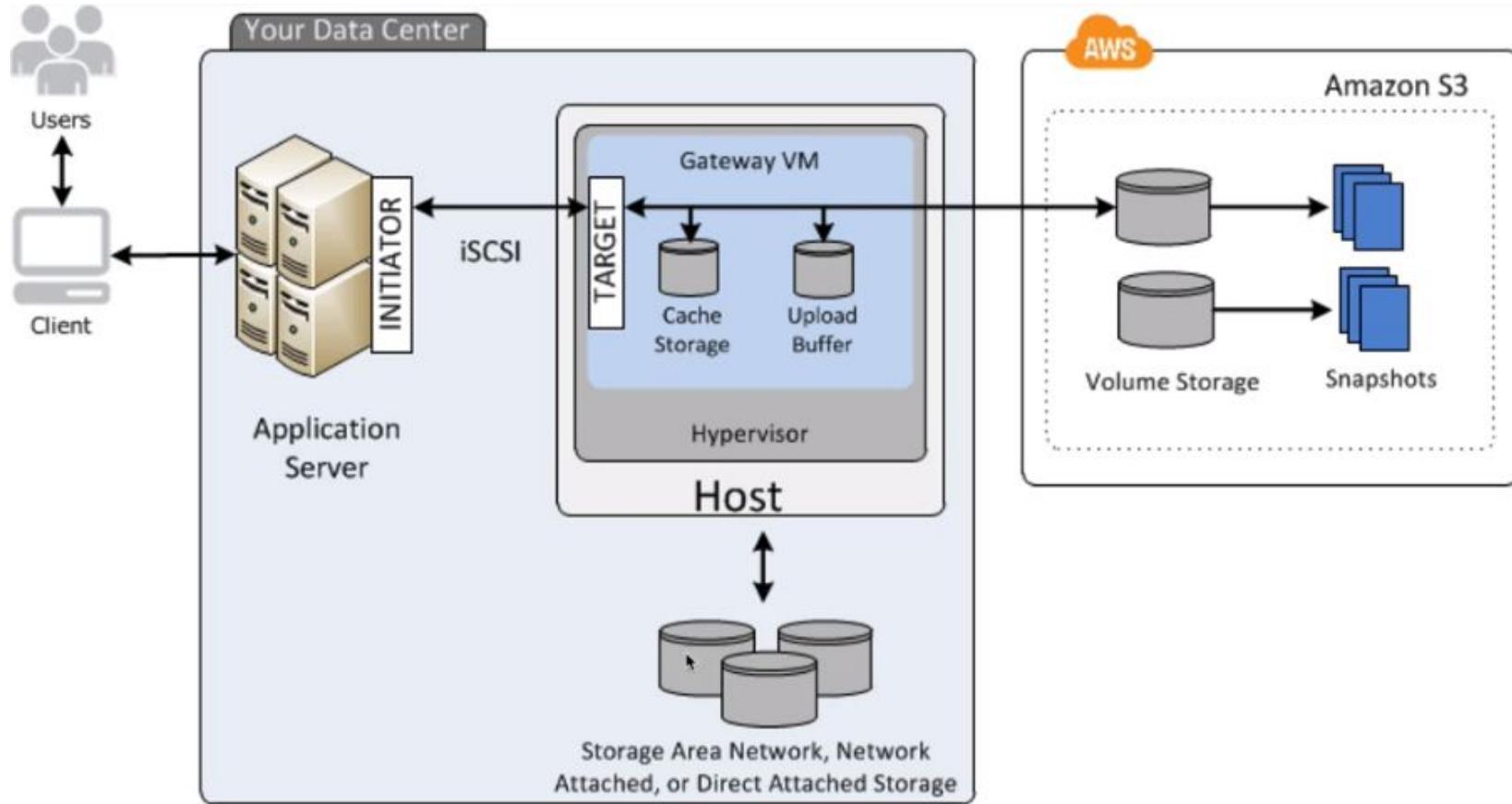


Volume Gateway - Cached Volumes



Cached volumes let you use Amazon Simple Storage Service (Amazon S3) as your primary data storage while retaining frequently accessed data locally in your storage gateway. Cached volumes minimize the need to scale your on-premises storage infrastructure, while still providing your applications with low-latency access to their frequently accessed data. You can create storage volumes up to 32 TiB in size and attach to them as iSCSI devices from your on-premises application servers. Your gateway stores data that you write to these volumes in Amazon S3 and retains recently read data in your on-premises storage gateway's cache and upload buffer storage. 1 GB - 32 TB in size for Cached Volumes.

Volume Gateway - Cached Volumes

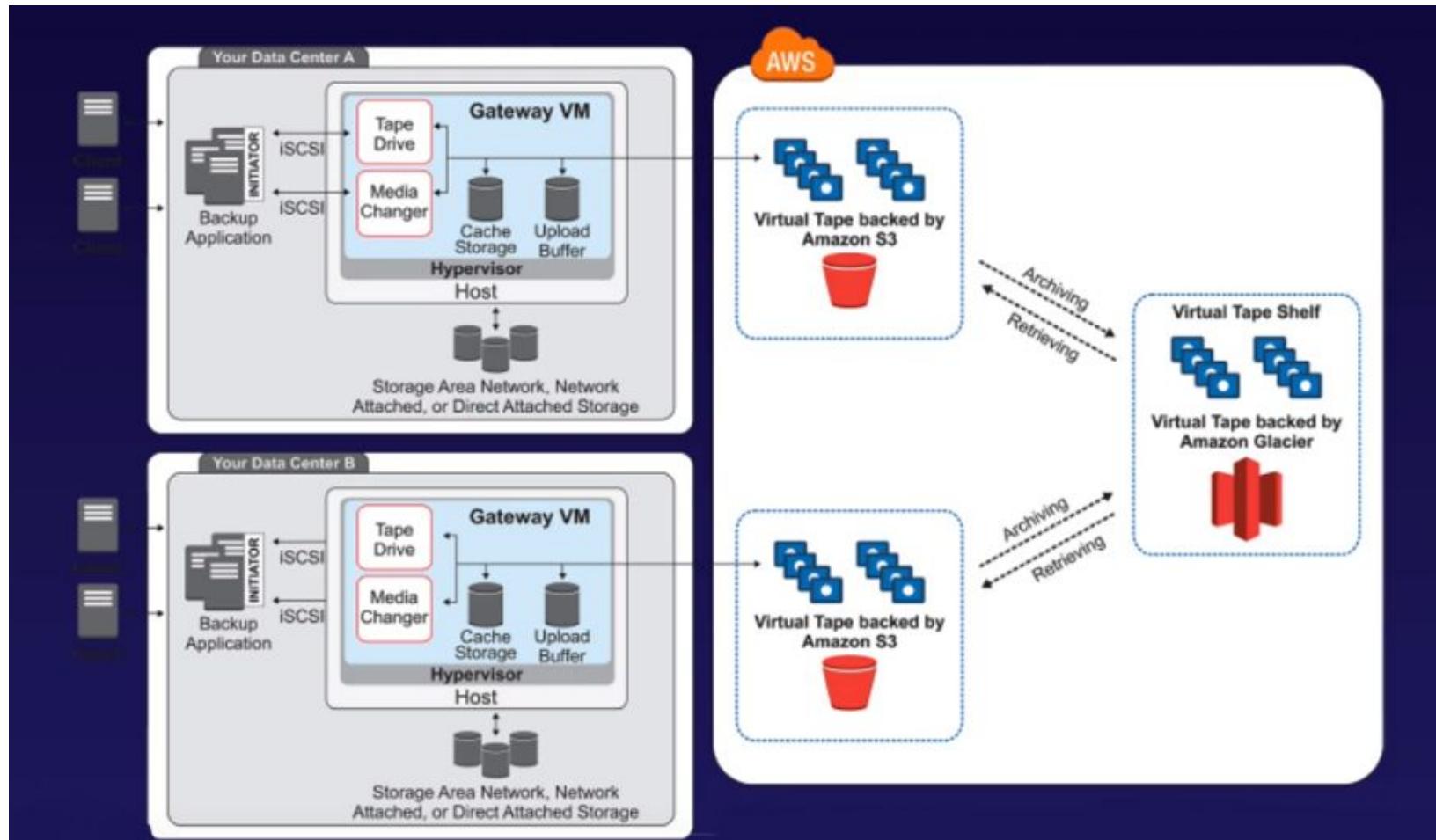


Tape Gateway



Tape Gateway offers a durable, cost-effective solution to archive your data in the AWS Cloud. The VTL interface it provides lets you leverage your existing tape-based backup application infrastructure to store data on virtual tape cartridges that you create on your tape gateway. Each tape gateway is preconfigured with a media changer and tape drives, which are available to your existing client backup applications as iSCSI devices. You add tape cartridges as you need to archive your data. Supported by NetBackup, Backup Exec, Veeam etc.

Tape Gateway



Exam tips

File Gateway

- File Gateway - For flat files, stored directly on S3.

Volume Gateway

- **Stored Volumes** - Entire Dataset is stored on site and is asynchronously backed up to S3.
- **Cached Volumes** - Entire Dataset is stored on S3 and the most frequently accessed data is cached on site.

Gateway Virtual Tape Library

Athena vs. Macie

[SAA-C02]

What is Athena?

Interactive query service which enables you to analyse and query data located in S3 using standard SQL

- Serverless, nothing to provision, pay per query / per TB scanned
- No need to set up complex Extract/Transform/Load (ETL) processes
- Works directly with data stored in S3



Athena Use cases

What Can Athena Be Used For?

- Can be used to query log files stored in S3, e.g. ELB logs, S3 access logs etc
- Generate business reports on data stored in S3
- Analyse AWS cost and usage reports
- Run queries on click-stream data



What is Macie?

What is PII (Personally Identifiable Information)?

- Personal data used to establish an individual's identity
- This data could be exploited by criminals, used in identity theft and financial fraud
- Home address, email address, SSN
- Passport number, driver's license number
- D.O.B, phone number, bank account, credit card number



What is Macie?

Security service which uses Machine Learning and NLP (Natural Language Processing) to discover, classify and protect sensitive data stored in S3

- Uses AI to recognise if your S3 objects contain sensitive data such as PII
- Dashboards, reporting and alerts
- Works directly with data stored in S3
- Can also analyze CloudTrail logs
- Great for PCI-DSS and preventing ID theft



Exam tips

Athena Exam Tips

Remember what Athena is and what it allows you to do:

- Athena is an interactive query service
- It allows you to query data located in S3 using standard SQL
- Serverless
- Commonly used to analyse log data stored in S3

Exam tips

Macie Exam Tips

Remember what Macie is and what it allows you to do:

- Macie uses AI to analyze data in S3 and helps identify PII
- Can also be used to analyse CloudTrail logs for suspicious API activity
- Includes Dashboards, Reports and Alerting
- Great for PCI-DSS compliance and preventing ID theft

Identity Access Management Quiz

Identity Access Management Quiz

QUESTION 1

A new employee has just started work, and it is your job to give her administrator access to the AWS console. You have given her a user name, an access key ID, a secret access key, and you have generated a password for her. She is now able to log in to the AWS console, but she is unable to interact with any AWS services. What should you do next?

- Tell her to log out and try logging back in again.
- Require multi-factor authentication for her user account.
- Ensure she is logging in to the AWS console from your corporate network and not the normal internet.
- Grant her Administrator access by adding her to an Administrators' group.

Good work!

By default new user accounts come with no permission to interact with services. These must be explicitly assigned by adding a Policy or adding them to a Group.

Identity Access Management Quiz

QUESTION 2

Every user you create in the IAM systems starts with ____.

Full Permissions

Partial Permissions

Inherited Permissions

No Permissions

Good work!

AWS systems are designed to be secure 1st. The system administrator needs to add permissions to allow accounts to take actions.

Identity Access Management Quiz

QUESTION 3

You are a solutions architect working for a large engineering company that are moving from a legacy infrastructure to AWS. You have configured the company's first AWS account and you have set up IAM. Your company is based in Andorra, but there will be a small subsidiary operating out of South Korea, so that office will need its own AWS environment. Which of the following statements is true?



You will need to configure Users and Policy Documents only once, as these are applied globally.



You will need to configure your users regionally, however your policy documents are global.



You will need to configure your policy documents regionally, however your users are global.



You will then need to configure Users and Policy Documents for each region, respectively.

Good work!

Identity Access Management Quiz

QUESTION 4

Power User Access allows ____.

- Read Only access to all AWS services and resources.
- Access to all AWS services except the management of groups and users within IAM.
- Full Access to all AWS services and resources.
- Users to inspect the source code of the AWS platform

Good work!

Identity Access Management Quiz

QUESTION 5

You are a security administrator working for a hotel chain. You have a new member of staff who has started as a systems administrator, and she will need full access to the AWS console. You have created the user account and generated the access key id and the secret access key. You have moved this user into the group where the other administrators are, and you have provided the new user with their secret access key and their access key id. However, when she tries to log in to the AWS console, she cannot. Why might that be?

You cannot log in to the AWS console using the Access Key ID / Secret Access Key pair.
 Instead, you must generate a password for the user, and supply the user with this password and your organization's unique AWS console login URL.

You have not yet activated multi-factor authentication for the user, so by default they will not be able to log in.

You have not applied the "log in from console" policy document to the user. You must apply this first so that they can log in.

Your user is trying to log in from the AWS console from outside the corporate network. This is not possible.

Sorry!

The question does not state that MFA is needed. While good practice it is not essential. The MFA verification step comes after the initial Account & Password login. However she does not have an Account & Password, only the programmatic access keys, so will never get to that step.

Identity Access Management Quiz

QUESTION 6

Which statement best describes IAM?



IAM allows you to manage users, groups, roles, and their corresponding level of access to the AWS Platform.



IAM allows you to manage users' passwords only. AWS staff must create new users for your organization. This is done by raising a ticket.



IAM allows you to manage permissions for AWS resources only.



IAM stands for Improvised Application Management, and it allows you to deploy and manage applications in the AWS Cloud.

Good work!

Correct. Key concepts; Users, Groups, Roles, and the level of access. And "...to the AWS Platform."

Identity Access Management Quiz

QUESTION 7

A __ is a document that provides a formal statement of one or more permissions.

Group

Role

Policy

User

Good work!

A Policy is a structured permissions statement

Identity Access Management Quiz

QUESTION 8

What is the default level of access a newly created IAM User is granted?

Administrator access to all AWS services.

Read-only access to all AWS services.

No access to any AWS services.

Power user access to all AWS services.

Good work!

By default new IAM Users have no permissions to AWS services. They must be explicitly granted.

Identity Access Management Quiz

QUESTION 9

You are a developer at a fast-growing startup. Until now, you have used the root account to log in to the AWS console. However, as you have taken on more staff, you will need to stop sharing the root account to prevent accidental damage to your AWS infrastructure. What should you do so that everyone can access the AWS resources they need to do their jobs?

Choose 2

Create a customized sign-in link such as "yourcompany.signin.aws.amazon.com/console" for your new users to use to sign in with.

Create individual user accounts with minimum necessary rights and tell the staff to log in to the console using the credentials provided.

Create an additional AWS root account for each new user.

Give your users the root account credentials so that they can also sign in.

Good work!

Read the AWS Security Best Practice white paper. Also note that the IAM account signin URL is different from the Root account signin URL

Identity Access Management Quiz

QUESTION 10

Which of the following is not a feature of IAM?

- IAM offers fine-grained access control to AWS resources.
- IAM offers centralized control of your AWS account.
- IAM integrates with existing active directory account allowing single sign-on.
- IAM allows you to set up biometric authentication, so that no passwords are required.

Good work!

AWS makes use of Accounts & Passwords, or Keys and Secret keys, and MFA, to prove identity. You may have a 3rd party device that uses BioMetrics to initiate and exchange of the password or secret key with AWS, but that is not an AWS / IAM service. <https://aws.amazon.com/iam/features/>

Identity Access Management Quiz

QUESTION 11

You have a client who is considering a move to AWS. In establishing a new account, what is the first thing the company should do?

- Set up an account via SQS (Simple Queue Service).
- Set up an account using their company email address.
- Set up an account via SNS (Simple Notification Service)
- Set up an account using Cloud Search.

Good work!

This email address is a key part of linking the AWS account to your company. Using a private email address may make it harder to establish ownership if you need help from AWS.

<https://aws.amazon.com/premiumsupport/knowledge-center/create-and-activate-aws-account/>
<https://aws.amazon.com/blogs/security/getting-started-follow-security-best-practices-as-you-configure-your-aws-resources/> Explanation

Identity Access Management Quiz

QUESTION 12

An application you are working on has a new app. The development team for this app requires access to a bucket that is located within your team's aws account. The other team requires programmatic and console level access to your team's bucket. How would you share this bucket with this other team's account?

Setting up a Resource Based Access Control List (ACL)

Setting up a cross account IAM Role

Setting up a resource-based policy

Setting up a shared IAM policy

Sorry!

Correct Answer

Setting up a cross account IAM role is currently the only method that will allow IAM users to access cross account S3 buckets both programmatically and via the AWS console.

Identity Access Management Quiz

QUESTION 13

What level of access does the "root" account have?

Administrator Access

Power User Access

No Access

Read-only Access

Good work!

Identity Access Management Quiz

QUESTION 14

In what language are policy documents written?



JSON



Python



Java



Node.js

Good work!

JavaScript Object Notation - is a human readable and easily parsed structured data format used to pass blocks of data into & between systems.

Identity Access Management Quiz

QUESTION 15

When you create a new user, that user ____.

- Will be able to log in to the console only after multi-factor authentication is enabled on their account.
- Will be able to interact with AWS using their access key ID and secret access key using the API, CLI, or the AWS SDKs assuming programmatic access was enabled.
- Will be able to log in to the console anywhere in the world, using their access key ID and secret access key.
- Will only be able to log in to the console in the region in which that user was created.

Good work!

To access the console you use an account and password combination. To access AWS programmatically you use a Key and Secret Key combination

Identity Access Management Quiz

QUESTION 16

What is an additional way to secure the AWS accounts of both the root account and new users alike?



Implement Multi-Factor Authentication for all accounts.



Store the access key id and secret access key of all users in a publicly accessible plain text document on S3 of which only you and members of your organization know the address.



Configure the AWS Console so that you can only log in to it from a specific IP Address range



Configure the AWS Console so that you can only log in to it from your internal network IP address range.

Good work!

MFA provides an additional requirement for the person signing on to prove that they are who they claim to be. Username & password are things you 'know' the MFA is something that you 'have'. e.g. you have the only device that can generate the token. <https://aws.amazon.com/iam/features/mfa/>

Identity Access Management Quiz

QUESTION 17

You have created a new AWS account for your company, and you have also configured multi-factor authentication on the root account. You are about to create your new users. What strategy should you consider in order to ensure that there is good security on this account.



Enact a strong password policy: user passwords must be changed every 45 days, with each password containing a combination of capital letters, lower case letters, numbers, and special symbols.



Give all users the same password so that if they forget their password they can just ask their co-workers.



Require users only to be able to log in using biometric authentication.



Restrict login to the corporate network only.

Good work!

Identity Access Management Quiz

QUESTION 18

Which of the following is not a component of IAM?

- Roles
- Users
- Organizational Units
- Groups

Good work!

S3 Quiz

S3 Quiz

QUESTION 1

Which of the following options allows users to have secure access to private files located in S3?

Choose 3



CloudFront Signed URLs



Public S3 buckets



CloudFront Origin Access Identity



CloudFront Signed Cookies

Good work!

There are three options in the question which can be used to secure access to files stored in S3 and therefore can be considered correct. Signed URLs and Signed Cookies are different ways to ensure that users attempting access to files in an S3 bucket can be authorised. One method generates URLs and the other generates special cookies but they both require the creation of an application and policy to generate and control these items. An Origin Access Identity on the other hand, is a virtual user identity that is used to give the CloudFront distribution permission to fetch a private object from an S3 bucket. Public S3 buckets should never be used unless you are using the bucket to host a public website and therefore this is an incorrect option.

S3 Quiz

QUESTION 2

What is the availability of objects stored in S3?

99.90%

99%

99.99%

100%

Good work!

S3 Quiz

QUESTION 3

What is the minimum file size that I can store on S3?

0 bytes

1MB

1KB

1 byte

Good work!

S3 Quiz

S3 - Infrequently Accessed Storage

S3 - One Zone-Infrequent Access

S3 - Reduced Redundancy Storage (RRS)

S3 - Provisioned IOPS

Glacier

Sorry!

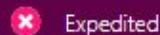
Correct Answer

The key driver here is cost, so an awareness of cost is necessary to answer this. Full S3 is quite expensive at around \$0.023 per GB for the lowest band. S3 standard IA is \$0.0125 per GB, S3 One-Zone-IA is \$0.01 per GB, and Legacy S3-RRS is around \$0.024 per GB for the lowest band. Of the offered solutions S3 One-Zone-IA is the cheapest suitable option. Glacier cannot be considered as it is not intended for direct access, however it comes in at around \$0.004 per GB. Of course you spotted that RRS is being deprecated, and there is no such thing as S3 - Provisioned IOPS. In this case OneZone IA should be fine as users will 'post' material but only the organization will access it and only to find relevant material. The question states that there is no concern if some material is lost.

S3 Quiz

QUESTION 5

You have been asked to set up a recovery process that generates the lowest possible cost for retrieving information from Glacier. There is petabytes of information that need to be retrieved if this process is to be enacted. Which of the following retrieval options would allow you to achieve this?



Sorry!

Correct Answer

Cost of retrieval of information from Glacier can go up dependent on how quickly you require the data and how much data is to be retrieved. Expedited retrievals allow you to quickly access your data stored in the S3 Glacier storage class when occasional urgent requests for a subset of archives are required, but at the highest cost. Standard retrievals allow you to access any of your archived objects within several hours, this is faster than bulk (averaging around 12 hours) but more expensive. Bulk retrievals are the lowest-cost retrieval option in Amazon S3 Glacier, enabling you to retrieve large amounts, even petabytes, of data inexpensively.

S3 Quiz

QUESTION 6

What is Amazon Glacier?



An AWS service designed for long term data archival.



It is a tool used to resurrect deleted EC2 snapshots.



A highly secure firewall designed to keep everything out.



A tool that allows you to "freeze" an EBS volume.



Good work!

S3 Quiz

QUESTION 7

You run a popular photo-sharing website that depends on S3 to store content. Paid advertising is your primary source of revenue. However, you have discovered that other websites are linking directly to the images in your buckets, not to the HTML pages that serve the content. This means that people are not seeing the paid advertising, and you are paying AWS unnecessarily to serve content directly from S3. How might you resolve this issue?



Use CloudFront to serve the static content.



Use security groups to blacklist the IP addresses of the sites that link directly to your S3 bucket.



Use EBS rather than S3 to store the content.



Remove the ability for images to be served publicly to the site and then use signed URLs with expiry dates.

S3 Quiz



S3 - IA



S3



S3 - 1Zone-IA



Glacier



S3 - RRS

Sorry!

While S3 - IA is low cost, the lowest reasonable solution is S3 - 1Zone-IA. The question specifies low cost and that the meme objects are not critical.

Correct Answer

S3 - OneZone-IA is the recommended storage for when you want cheaper storage for infrequently accessed objects. It has the same durability but less availability. There can be cost implications if you use it frequently or use it for short lived storage. Glacier is cheaper, but has a long retrieval time. RRS has effectively been deprecated. It still exists but is not a service that AWS want to sell anymore. RRS is also now more expensive than S3 Standard.

S3 Quiz

QUESTION 9

S3 has what consistency model for PUTS of new objects



Eventual Consistency



Write After Read Consistency



Usual Consistency



Read After Write Consistency

S3 Quiz

QUESTION 10

You are a solutions architect who works with a large digital media company. The company has decided that they want to operate within the Japanese region and they need a bucket called "testbucket" set up immediately to test their web application on. You log in to the AWS console and try to create this bucket in the Japanese region however you are told that the bucket name is already taken. What should you do to resolve this?

- Change your region to Korea and then create the bucket "testbucket".
- Bucket names are global, not regional. This is a popular bucket name and is already taken.
You should choose another bucket name.
- Raise a ticket with AWS and ask them to release the name "testbucket" to you.
- Run a WHOIS request on the bucket name and get the registered owners email address.
Contact the owner and ask if you can purchase the rights to the bucket.

S3 Quiz



S3 - IA



S3 - RRS



Glacier



S3



S3 - 1Zone-IA

Sorry!

Correct Answer

The need for immediate access is an important requirement along with cost. Glacier has a long recovery time at a low cost or a shorter recovery time at a high cost, and 1Zone-IA has a lower Availability level which means that it may not be available when needed.

S3 Quiz

QUESTION 12

S3 has eventual consistency for which HTTP Methods?

PUTS of new objects and UPDATES

UPDATES and DELETES

overwrite PUTS and DELETES

PUTS of new Objects and DELETES

S3 Quiz

QUESTION 13

One of your users is trying to upload a 7.5GB file to S3. However, they keep getting the following error message: "Your proposed upload exceeds the maximum allowed object size.". What solution to this problem does AWS recommend?



Design your application to use the Multipart Upload API for all objects.



Log in to the S3 console, click on the bucket and then click properties. You can then increase your maximum object size to 1TB.



Design your application to use large object upload API for this object.



Raise a ticket with AWS to increase your maximum object size.

S3 Quiz

QUESTION 14

What is the availability of S3-OneZone-IA?

 99.90%

 99.50%

 100%

 99.99%

 Sorry!

 Correct Answer

OneZone-IA is only stored in one Zone. While it has the same Durability, it may be less Available than normal S3 or S3-IA.

S3 Quiz

- <https://www.my-registered-domain-guru/index.html>
- <http://my-bucket.s3-website-ap-southeast-2.amazonaws.com/index.php>
- <https://my-bucket.s3.us-west-2.amazonaws.com/fastpuppy.csv>
- <https://my-bucket.amazonaws.com/lazycat.docx>
- <http://my-bucket.s3-website.us-east-2.amazonaws.com/index.htm>
- <https://s3.us-west-2.amazonaws.com/my-bucket/slowpuppy.tar>

Sorry!

Correct Answer

Virtual style puts your bucket name 1st, s3 2nd, and the region 3rd. Path style puts s3 1st and your bucket as a sub domain. Legacy Global endpoint has no region. S3 static hosting can be your own domain or your bucket name 1st, s3-website 2nd, followed by the region. AWS are in the process of phasing out Path style, and support for Legacy Global Endpoint format is limited and discouraged. However it is still useful to be able to recognize them should they show up in logs.

S3 Quiz

How many S3 buckets can I have per account by default?

100

10

50

20

Good work!

CHAPTER 4

EC2

EC2 101

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change.



EC2 Pricing model

1

On Demand

Allows you to pay a fixed rate by the hour (or by the second) with no commitment.

2

Reserved

Provides you with a capacity reservation, and offer a significant discount on the hourly charge for an instance. Contract Terms are 1 Year or 3 Year Terms.

3

Spot

Enables you to bid whatever price you want for instance capacity, providing for even greater savings if your applications have flexible start and end times.

4

Dedicated Hosts

Physical EC2 server dedicated for your use. Dedicated Hosts can help you reduce costs by allowing you to use your existing server-bound software licenses.



EC2 101

On Demand pricing is useful for;

- Users that want the low cost and flexibility of Amazon EC2 without any up-front payment or long-term commitment
- Applications with short term, spiky, or unpredictable workloads that cannot be interrupted
- Applications being developed or tested on Amazon EC2 for the first time



Reserved pricing is useful for;

- Applications with steady state or predictable usage
- Applications that require reserved capacity
- Users able to make upfront payments to reduce their total computing costs even further



EC2 101

Reserved Pricing Types

1 Standard Reserved instances

These offer up to 75% off on demand instances. The more you pay up front and the longer the contract, the greater the discount.

2 Convertible Reserved Instances

These offer up to 54% off on demand capability to change the attributes of the RI as long as the exchange results in the creation of Reserved Instances of equal or greater value.

3 Scheduled Reserved Instances

These are available to launch within the time windows you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, a week, or a month.



EC2 101

Reserved Pricing Types

1 Standard Reserved instances

These offer up to 75% off on demand instances. The more you pay up front and the longer the contract, the greater the discount.

2 Convertible Reserved Instances

These offer up to 54% off on demand capability to change the attributes of the RI as long as the exchange results in the creation of Reserved Instances of equal or greater value.

3 Scheduled Reserved Instances

These are available to launch within the time windows you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, a week, or a month.



EC2 101

Dedicated Hosts pricing is useful for;

- Useful for regulatory requirements that may not support multi-tenant virtualization.
- Great for licensing which does not support multi-tenancy or cloud deployments.
- Can be purchased On-Demand (hourly.)
- Can be purchased as a Reservation for up to 70% off the On-Demand price.



EC2 101

Spot pricing is useful for;

- Applications that have flexible start and end times
- Applications that are only feasible at very low compute prices
- Users with urgent computing needs for large amounts of additional capacity



EC2 Instances Types

Family	Specialty	Use case
F1	Field Programmable Gate Array	Genomics research, financial analytics, real-time video processing, big data etc
I3	High Speed Storage	NoSQL DBs, Data Warehousing etc
G3	Graphics Intensive	Video Encoding/ 3D Application Streaming
H1	High Disk Throughput	MapReduce-based workloads, distributed file systems such as HDFS and MapR-FS
T3	Lowest Cost, General Purpose	Web Servers/Small DBs
D2	Dense Storage	File servers/Data Warehousing/Hadoop
R5	Memory Optimized	Memory Intensive Apps/DBs
M5	General Purpose	Application Servers
C5	Compute Optimized	CPU Intensive Apps/DBs
P3	Graphics/General Purpose GPU	Machine Learning, Bit Coin Mining etc
X1	Memory Optimized	SAP HANA/Apache Spark etc
Z1D	High compute capacity and a high memory footprint.	Ideal for electronic design automation (EDA) and certain relational database workloads with high per-core licensing costs.
A1	Arm-based workloads	Scale-out workloads such as web servers
U-6tb1	Bare Metal	Bare metal capabilities that eliminate virtualization overhead

EC2 101- Mnemonic

- **F** - For FPGA
- **I** - For IOPS
- **G** - Graphics
- **H** - High Disk Throughput
- **T** - Cheap general purpose (think T2 Micro)
- **D** - For Density
- **R** - For RAM
- **M** - Main choice for general purpose apps
- **C** - For Compute
- **P** - Graphics (think Pics)
- **X** - Extreme Memory
- **Z** - Extreme Memory AND CPU
- **A** - Arm-based workloads
- **U** - Bare Metal





EC2 - Exam tips

- Termination Protection is **turned off** by default, you must turn it on.
- On an EBS-backed instance, the **default action is for the root EBS volume to be deleted** when the instance is terminated.
- EBS Root Volumes of your DEFAULT AMI's **CAN** be encrypted. You can also use a third party tool (such as bit locker etc) to encrypt the root volume, or this can be done when creating AMI's (lab to follow) in the AWS console or using the API.
- Additional volumes can be encrypted.



Security Groups - Exam tips

- All Inbound traffic is blocked by default.
- All Outbound traffic is allowed.
- Changes to Security Groups take effect immediately.
- You can have any number of EC2 instances within a security group.
- You can have multiple security groups attached to EC2 Instances.



Security Groups - Exam tips

- Security Groups are **STATEFUL**.
- If you create an inbound rule allowing traffic in, that traffic is automatically allowed back out again.
- You cannot block specific IP addresses using Security Groups, instead use Network Access Control Lists.
- You can specify allow rules, but not deny rules.

EBS 101

EBS 101

Amazon Elastic Block Store (EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability.



5 Different Types of EBS Storage;

- General Purpose (SSD)
- Provisioned IOPS (SSD)
- Throughput Optimised Hard Disk Drive
- Cold Hard Disk Drive
- Magnetic



EBS 101

Solid-State Drives (SSD)			Hard disk Drives (HDD)		
Volume Type	General Purpose SSD	Provisioned IOPS SSD	Throughput Optimized HDD	Cold HDD	EBS Magnetic
Description	General purpose SSD volume that balances price and performance for a wide variety of transactional workloads	Highest-performance SSD volume designed for mission-critical applications	Low cost HDD volume designed for frequently accessed, throughput-intensive workloads	Lowest cost HDD volume designed for less frequently accessed workloads	Previous generation HDD
Use Cases	Most Work Loads	Databases	Big Data & Data Warehouses	File Servers	Workloads where data is infrequently accessed
API Name	gp2	io1	st1	sc1	Standard
Volume Size	1 GiB - 16 TiB	4 GiB - 16 TiB	500 GiB - 16 TiB	500 GiB - 16 TiB	1 GiB-1 TiB
Max. IOPS**/ Volume	16,000	64,000	500	250	40-200



EBS - Exam tips

- Volumes exist on EBS. Think of EBS as a virtual hard disk
- Snapshots exist on S3. Think of snapshots as a photograph of the disk.
- Snapshots are point in time copies of Volumes.
- Snapshots are incremental — this means that only the blocks that have changed since your last snapshot are moved to S3.
- If this is your first snapshot, it may take some time to create.



EBS - Exam tips

- To create a snapshot for Amazon EBS volumes that serve as root devices, you should stop the instance before taking the snapshot.
- However you can take a snap while the instance is running.
- You can create AMI's from Snapshots
- You can change EBS volume sizes on the fly, including changing the size and storage type.
- Volumes will **ALWAYS** be in the same availability zone as the EC2 instance.



EBS - Exam tips

- To move an EC2 volume from one AZ to another, take a snapshot of it, create an AMI from the snapshot and then use the AMI to launch the EC2 instance in a new AZ.
- To move an EC2 volume from one region to another, take a snapshot of it, create an AMI from the snapshot and then copy the AMI from one region to the other. Then use the copied AMI to launch the new EC2 instance in the new region.



EBS - Exam tips

- Snapshots of encrypted volumes are encrypted automatically.
- Volumes restored from encrypted snapshots are encrypted automatically.
- You can share snapshots, but only if they are unencrypted.
- These snapshots can be shared with other AWS accounts or made public.
- You can now encrypt root device volumes upon creation of the EC2 instance.

EBS - Exam tips



- Create a Snapshot of the unencrypted root device volume
- Create a copy of the Snapshot and select the encrypt option
- Create an AMI from the encrypted Snapshot
- Use that AMI to launch new encrypted instances

AMI Types (EBS vs. Instance Store)

AMI Types

You can select your AMI based on:

- Region (see Regions and Availability Zones)
- Operating system
- Architecture (32-bit or 64-bit)
- Launch Permissions
- Storage for the Root Device (Root Device Volume

Instance Store (**EPHEMERAL STORAGE**)

EBS Backed Volumes



AMI Types

All AMIs are categorized as either backed by Amazon EBS or backed by instance store.

For EBS Volumes: The root device for an instance launched from the AMI is an Amazon EBS volume created from an Amazon EBS snapshot.

For Instance Store Volumes: The root device for an instance launched from the AMI is an instance store volume created from a template stored in Amazon S3.



AMI Types

- Instance Store Volumes are sometimes called Ephemeral Storage.
- Instance store volumes cannot be stopped. If the underlying host fails, you will lose your data.
- EBS backed instances can be stopped. You will not lose the data on this instance if it is stopped.
- You can reboot both, you will not lose your data.
- By default, both ROOT volumes will be deleted on termination. However, with EBS volumes, you can tell AWS to keep the root device volume.

ENI vs. ENA vs. EFA

ENI vs. ENA vs. EFA

ENI vs ENA vs EFA

1 ENI

Elastic Network Interface - essentially a virtual network card.

2 EN

Enhanced Networking. Uses single root I/O virtualization (SR-IOV) to provide high-performance networking capabilities on supported instance types.

3 Elastic Fabric Adapter

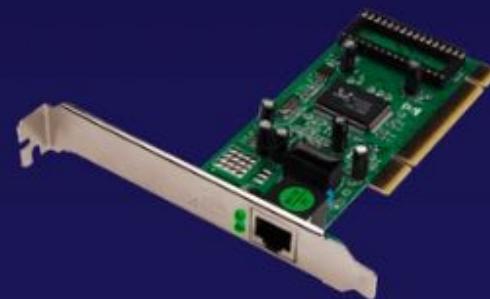
A network device that you can attach to your Amazon EC2 instance to accelerate High Performance Computing (HPC) and machine learning applications.



ENI vs. ENA vs. EFA

An ENI is simply a virtual network card for your EC2 instances. It allows:

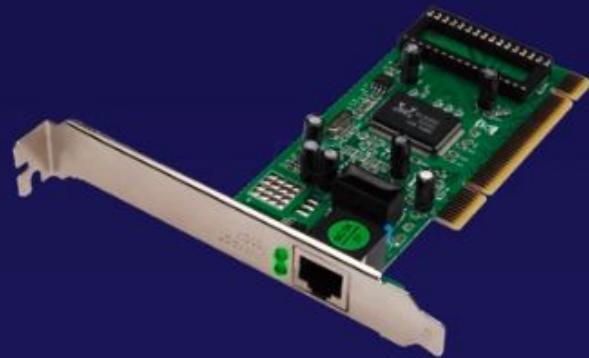
- A primary private IPv4 address from the IPv4 address range of your VPC
- One or more secondary private IPv4 addresses from the IPv4 address range of your VPC
- One Elastic IP address (IPv4) per private IPv4 address
- One public IPv4 address
- One or more IPv6 addresses
- One or more security groups
- A MAC address
- A source/destination check flag
- A description



Elastic Network Interface

Scenarios for Network Interfaces:

- Create a management network.
- Use network and security appliances in your VPC.
- Create dual-homed instances with workloads/roles on distinct subnets.
- Create a low-budget, high-availability solution.



Enhanced Networking

What is Enhanced Networking?

- It uses **single root I/O virtualization (SR-IOV)** to provide high-performance networking capabilities on supported instance types. SR-IOV is a method of device virtualization that provides higher I/O performance and lower CPU utilization when compared to traditional virtualized network interfaces.
- Enhanced networking provides higher bandwidth, higher packet per second (PPS) performance, and consistently lower inter-instance latencies. There is no additional charge for using enhanced networking.
- Use where you want good network performance.



Enhanced Networking

Depending on your instance type, enhanced networking can be enabled using:

- **Elastic Network Adapter (ENA)**, which supports network speeds of up to **100 Gbps** for supported instance types.

Or

- Intel 82599 **Virtual Function (VF)** interface, which supports network speeds of up to **10 Gbps** for supported instance types.
This is typically used on older instances.



In any scenario question, you probably want to choose ENA over VF if given the option.

Enhanced Networking



What is an Elastic Fabric Adapter?

- An **Elastic Fabric Adapter (EFA)** is a network device that you can attach to your Amazon EC2 instance to accelerate High Performance Computing (HPC) and machine learning applications.
- EFA provides lower and more consistent latency and higher throughput than the TCP transport traditionally used in cloud-based HPC systems.
- EFA can use OS-bypass. OS-bypass enables HPC and machine learning applications to bypass the operating system kernel and to communicate directly with the EFA device. It makes it a lot faster with a lot lower latency. Not supported with Windows currently, only Linux.

Enhanced Networking

In the exam you will be given different scenarios and you will be asked to choose whether you should use an ENI, EN or EFA.

- **ENI**

For basic networking. Perhaps you need a separate management network to your production network or a separate logging network and you need to do this at low cost. In this scenario use multiple ENIs for each network.

- **Enhanced Network**

For when you need speeds between 10Gbps and 100Gbps. Anywhere you need reliable, high throughput.

- **Elastic Fabric Adaptor**

For when you need to accelerate High Performance Computing (HPC) and machine learning applications *or* if you need to do an OS by-pass. If you see a scenario question mentioning HPC or ML and asking what network adaptor you want, choose EFA.

Spot Instances and Spot Fleets

Spot Instances and Spot Fleets



Amazon EC2 Spot Instances let you take advantage of unused EC2 capacity in the AWS Cloud. Spot Instances are available at up to a 90% discount compared to On-Demand prices. You can use **Spot Instances** for various stateless, fault-tolerant, or flexible applications, such as big data, containerized workloads, CI/CD, web servers, high-performance computing (HPC), and other test and development workloads.

Spot prices

To use **Spot Instances**, you must first decide on your maximum Spot price. The instance will be provisioned so long as the Spot price is **BELOW** your maximum Spot price.



The **hourly Spot price** varies depending on capacity and region.



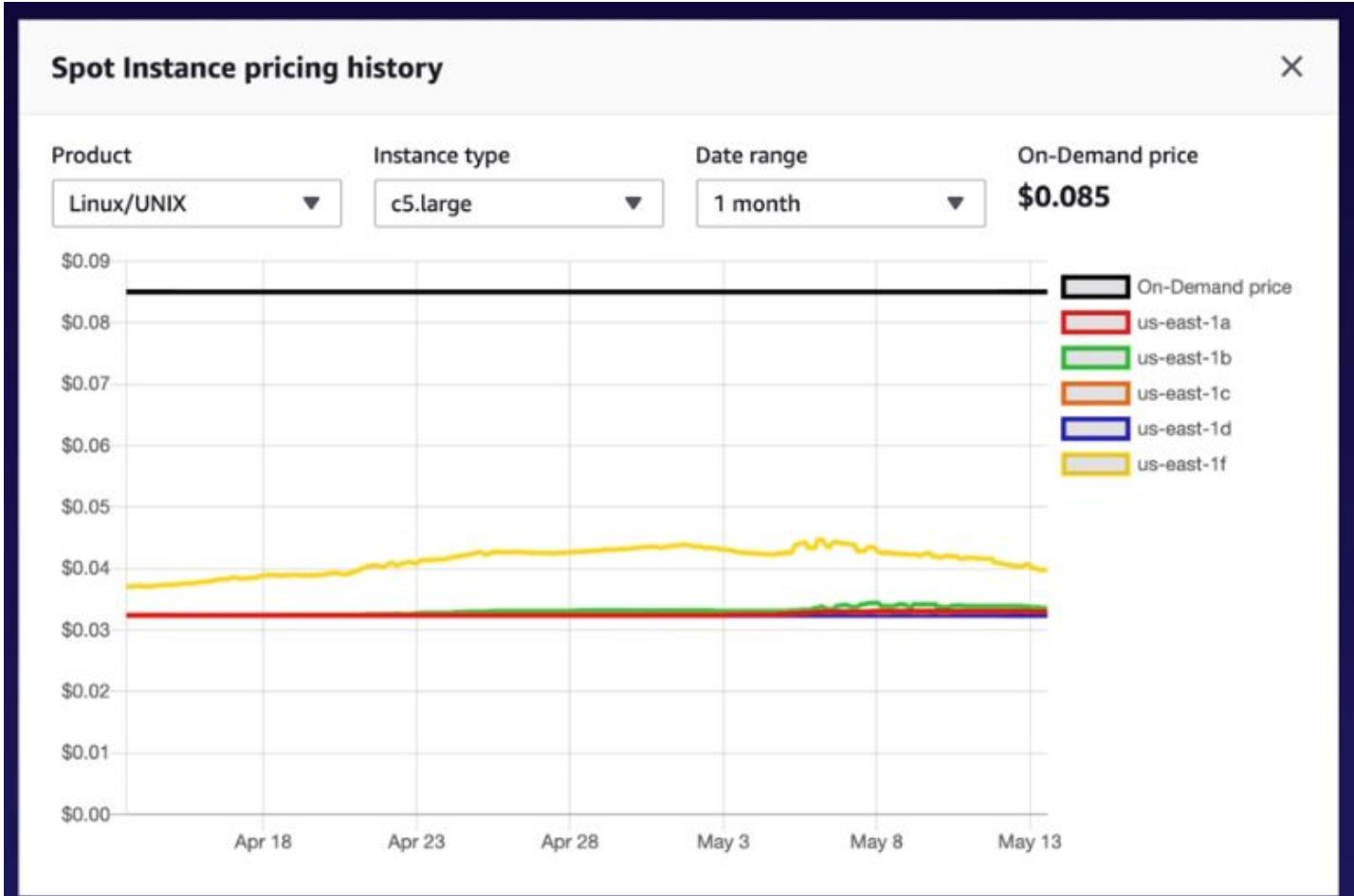
If the Spot price goes above your maximum, you have **two minutes** to choose whether to stop or terminate your instance.

Spot prices



You may also use a **Spot block** to stop your Spot Instances from being terminated even if the Spot price goes over your max Spot price. You can set Spot blocks for between **one to six hours** currently.

Spot prices



Spot instances

Spot Instances are useful for the following tasks:



Big data and analytics



Containerized
workloads



CI/CD and testing



Web services



Image and media
rendering



High-performance
computing

Spot instances

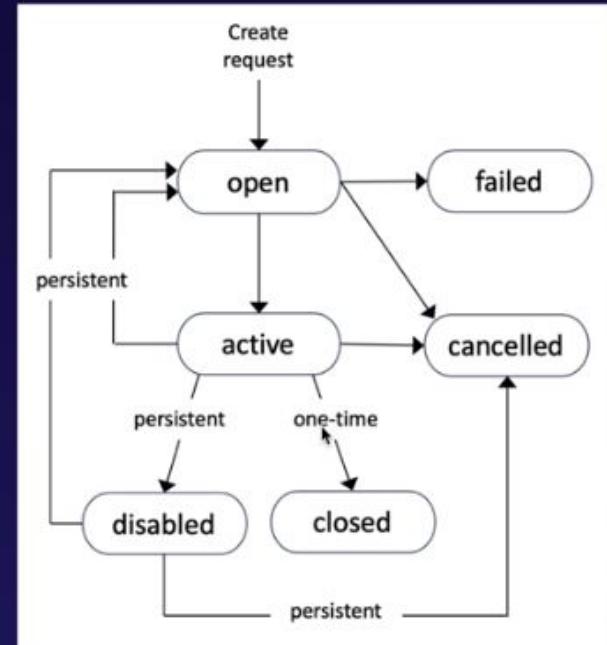
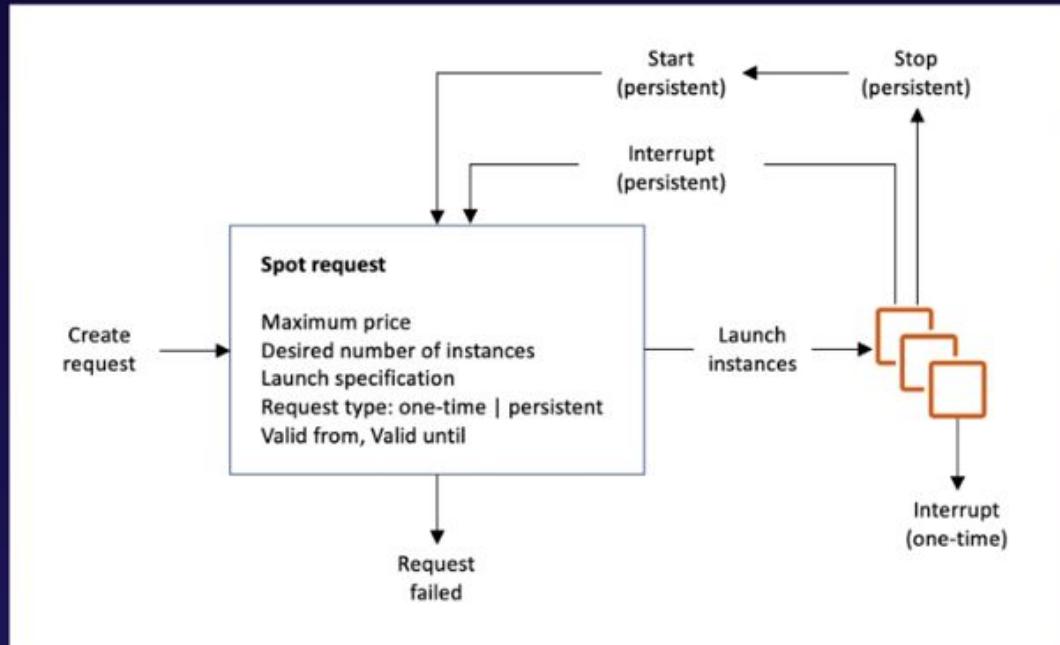
Spot Instances are not good for:

- ✓ Persistent workloads
- ✓ Critical jobs
- ✓ Databases



Terminating Spot instances

How to Terminate Spot Instances



Spot fleets

Spot Fleets

A Spot Fleet is a collection of Spot Instances and, optionally, On-Demand Instances.

The **Spot Fleet** attempts to launch the number of Spot Instances and On-Demand Instances to meet the target capacity you specified in the Spot Fleet request. The request for Spot Instances is fulfilled if there is available capacity and the **maximum price you specified in the request exceeds the current Spot price**. The Spot Fleet also attempts to maintain its target capacity fleet if your Spot Instances are interrupted.

Launch pools

Spot Fleets will try and match the target capacity with your price restraints.



- 1 Set up different launch pools. Define things like **EC2** instance type, operating system, and Availability Zone.
- 2 You can have **multiple** pools, and the fleet will choose the best way to implement depending on the strategy you define.
- 3 Spot fleets will **stop launching instances** once you reach your price threshold or capacity desire.

Strategies

You can have the following strategies with Spot Fleets.



capacityOptimized

The Spot Instances come from the pool with optimal capacity for the number of instances launching.



lowestPrice

The Spot Instances come from the pool with the lowest price. This is the default strategy.



diversified

The Spot Instances are distributed across all pools.



InstancePoolsToUseCount

The Spot Instances are distributed across the number of Spot Instance pools you specify. This parameter is valid only when used in combination with **lowestPrice**.

Exam Tips



Spot Instances save up to **90%** of the cost of On-Demand Instances.



Useful for any type of computing where you don't need **persistent storage**.



You can block Spot Instances from terminating by using **Spot block**.



A Spot Fleet is a collection of Spot Instances and, optionally, On-Demand Instances.



EC2 Hibernate

EBS Behaviors Reviewed

We have learned so far we can stop and terminate EC2 instances. If we stop the instance, the **data is kept on the disk (with EBS)** and will remain on the disk until the EC2 instance is started. If the instance is terminated, then by default **the root device volume will also be terminated**.



EBS Behaviors Reviewed



When we start our EC2 instance, the following happens:

- ✓ Operating system boots up
- ✓ User data script is run (**bootstrap scripts**)
- ✓ Applications start (can take some time)

EBS Behaviors Reviewed

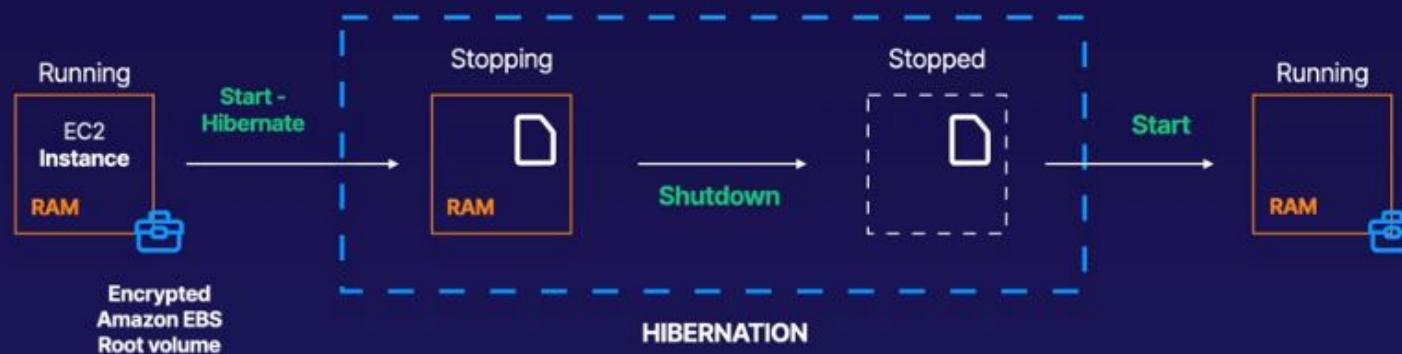
EC2 Hibernate

When you hibernate an EC2 instance, the operating system is told to perform hibernation (suspend-to-disk). Hibernation **saves the contents** from the instance memory (RAM) to your Amazon EBS root volume. We persist the instance's Amazon EBS root volume and any attached Amazon EBS data volumes.

Starting Your EC2 Instance with EC2 Hibernate

When you start your instance out of hibernation:

- The **Amazon EBS** root volume is restored to its previous state
- The **RAM** contents are reloaded
- The processes that were previously running on the instance are resumed
- Previously attached data volumes are **reattached and the instance retains its instance ID**



Starting Your EC2 Instance with EC2 Hibernate

With **EC2 Hibernate**, the instance boots much faster. The operating system does not need to reboot because the in-memory state (RAM) is preserved. This is useful for:

- 1 **Long-running processes**
- 2 **Services that take time to initialize**

Starting Your EC2 Instance with EC2 Hibernate

1. Choose AMI

2. Choose Instance Type

3. Configure Instance

4. Add Storage

5. Add Tags

6. Configure Security Group

7. Review

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

Number of Instances i

1

Launch into Auto Scaling Group i

Purchasing option i

Request Spot Instances

Network i

vpc-84fe10ef (default)

C Create new VPC

Subnet i

No preference (default subnet in any Availability Zone)

C Create new subnet

Auto-assign Public IP i

Use subnet setting (Enable)

Placement group i

Add instance to placement group

Capacity Reservation i

Open

C Create new Capacity Reservation

IAM role i

None

C Create new IAM role

Shutdown behavior i

Stop

Stop - Hibernate behavior i

Enable hibernation as an additional stop behavior

To enable hibernation, space is allocated on the root volume to store the instance memory (RAM). Make sure that the root volume is large enough to store the RAM contents and accommodate your expected usage, e.g. OS, applications. To use hibernation, the root volume must be an encrypted EBS volume. [Learn more](#)

Enable termination protection i

Protect against accidental termination

Monitoring i

Enable CloudWatch detailed monitoring

Additional charges apply.

Tenancy i

Shared - Run a shared hardware instance

[Cancel](#)

[Previous](#)

[Review and Launch](#)

[Next: Add Storage](#)

Starting Your EC2 Instance with EC2 Hibernate

2. Choose Instance Type

3. Configure Instance

4. Add Storage

5. Add Tags

6. Configure Security Group

7. Review

Add Storage

You will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or you can change the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. [Learn more about storage in Amazon EC2](#).

Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Encryption
/dev/xvda	snap-0cc421f413b5be1dd	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input checked="" type="checkbox"/>	Not Encrypted

When you hibernate, encrypt the root volume.

Eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage. Learn more about free usage tier eligibility and restrictions.

KMS Key	KMS Key ID
Aliases	
Not Encrypted	
(default) aws/ebs	alias/aws/ebs
MyKey	f085733c-5ebd-4232-b9f3-7821a85091e7

Starting Your EC2 Instance with EC2 Hibernate

The screenshot shows the AWS EC2 Instances page. On the left, there's a sidebar with navigation links like 'New EC2 Experience', 'EC2 Dashboard', 'Events', 'Tags', 'Reports', 'Limits', 'INSTANCES' (selected), 'Instances' (selected), 'Instance Types', 'Launch Templates', 'Spot Requests', 'Savings Plans', 'Reserved Instances', 'Dedicated Hosts', 'Capacity Reservations', 'IMAGES', 'AMIs', 'Bundle Tasks', 'ELASTIC BLOCK STORE', 'Volumes', 'Snapshots', 'Lifecycle Manager', and 'NETWORK & SECURITY', 'Security Groups'. The main content area shows a table of instances. One instance, named 'MyHibernate...', has its details expanded. The 'Actions' dropdown menu is open over this instance, showing options: 'Connect', 'Get Windows Password', 'Create Template From Instance', 'Launch More Like This', 'Instance State' (with 'Start', 'Stop', 'Stop - Hibernate' (selected), 'Reboot', and 'Terminate' options), 'Instance Settings', 'Image', 'Networking', and 'CloudWatch Monitoring'. Below the table, the instance details are shown: Instance: i-046ef7c4e39868653 (MyHibernateEC2), Public DNS: ec2-18-191-161-211.us-east-2.compute.amazonaws.com. The 'Description' tab is selected in the details view, showing fields: Instance ID (i-046ef7c4e39868653), Instance state (running), Instance type (t2.micro), Status Checks (2/2 checks ...), Monitoring (None), Tags, Description (Finding: Opt-in to AWS Compute Optimizer for recommendations. Learn more), Private DNS (ip-172-31-31-167.us-east-2.compute.internal), Public DNS (IPv4) (ec2-18-191-161-211.us-east-2.compute.amazonaws.com), IPv4 Public IP (18.191.161.211), IPv6 IPs (-), Elastic IPs, and Availability zone (us-east-2b).

EC2 Hibernate Exam tips

EC2 Hibernate preserves the in-memory RAM on persistent storage (EBS)

Much faster to boot up because you **do not need to reload the operating system**

Instance RAM must be less than **150 GB**

Instance families include C3, C4, C5, M3, M4, M5, R3, R4, and R5

Available for Windows, Amazon Linux 2 AMI, and Ubuntu

Instances can't be hibernated for more than **60 days**

Available for **On-Demand instances** and **Reserved Instances**



Cloudwatch 101

What is CloudWatch?

Amazon CloudWatch is a monitoring service to monitor your AWS resources, as well as the applications that you run on AWS.



CloudWatch in Nutshell

CloudWatch monitors performance.



What can CloudWatch monitor?

CloudWatch can monitor things like

- Compute
 - EC2 Instances
 - Autoscaling Groups
 - Elastic Load Balancers
 - Route53 Health Checks
- Storage & Content Delivery
 - EBS Volumes
 - Storage Gateways
 - CloudFront



CloudWatch & EC2

Host Level Metrics Consist of:

- CPU
- Network
- Disk
- Status Check



What is Cloud Trail?

AWS CloudTrail increases visibility into your user and resource activity by recording AWS Management Console actions and API calls. You can identify which users and accounts called AWS, the source IP address from which the calls were made, and when the calls occurred.



CloudWatch vs cloudtrail

- CloudWatch monitors performance.
- CloudTrail monitors API calls in the AWS platform.



Exam Tips

Remember;

- CloudWatch is used for monitoring performance.
- CloudWatch can monitor most of AWS as well as your applications that run on AWS.
- CloudWatch with EC2 will monitor events every 5 minutes by default.
- You can have 1 minute intervals by turning on detailed monitoring.
- You can create CloudWatch alarms which trigger notifications.
- CloudWatch is all about performance. CloudTrail is all about auditing.



What Can I do With CloudWatch?

- Dashboards - Creates awesome dashboards to see what is happening with your AWS environment.
- Alarms - Allows you to set Alarms that notify you when particular thresholds are hit.
- Events - CloudWatch Events helps you to respond to state changes in your AWS resources.
- Logs - CloudWatch Logs helps you to aggregate, monitor, and store logs.

Exam Tips



- Standard Monitoring = 5 Minutes
- Detailed Monitoring = 1 Minute

Amazon FSx for Windows and Amazon FSx for Lustre

Amazon FSx

"Amazon FSx for Windows File Server provides a fully managed native Microsoft Windows file system so you can easily move your Windows-based applications that require file storage to AWS. Amazon FSx is built on Windows Server."



How is Windows FSx different to EFS?

Windows FSx

- A managed Windows Server that runs Windows Server Message Block (SMB)-based file services.
- Designed for Windows and Windows applications.
- Supports AD users, access control lists, groups and security policies, along with Distributed File System (DFS) namespaces and replication.

EFS

- A managed NAS filer for EC2 instances based on Network File System (NFS) version 4.
- One of the first network file sharing protocols native to Unix and Linux.

Amazon FSx

"Amazon FSx for Lustre is a fully managed file system that is optimized for compute-intensive workloads, such as high-performance computing, machine learning, media data processing workflows, and electronic design automation (EDA)."

With Amazon FSx, you can launch and run a Lustre file system that can process massive data sets at up to hundreds of gigabytes per second of throughput, millions of IOPS, and sub-millisecond latencies."

How is Windows FSx different to EFS?

Lustre FSx

- Designed specifically for fast processing of workloads such as machine learning, high performance computing (HPC), video processing, financial modeling, and electronic design automation (EDA).
- Lets you launch and run a file system that provides sub-millisecond access to your data and allows you to read and write data at speeds of up to hundreds of gigabytes per second of throughput and millions of IOPS.

EFS

- A managed NAS filer for EC2 instances based on Network File System (NFS) version 4.
- One of the first network file sharing protocols native to Unix and Linux.

Windows FSx Exam Tips

In the exam you'll be given different scenarios and asked to choose whether you should use an EFS, FSx for Windows or FSx for Lustre.

- **EFS**

When you need distributed, highly resilient storage for Linux instances and Linux-based applications.

- **Amazon FSx for Windows**

When you need centralised storage for Windows-based applications such as Sharepoint, Microsoft SQL Server, Workspaces, IIS Web Server or any other native Microsoft Application.

- **Amazon FSx for Lustre**

When you need high-speed, high-capacity distributed storage. This will be for applications that do High Performance Compute (HPC), financial modelling etc. Remember that FSx for Lustre can store data directly on S3.

EC2 Placement Groups

EC2 Placement Groups

Three Types of Placement Groups;

- Clustered Placement Group
- Spread Placement Group
- Partitioned



Clustered Placement Group

A cluster placement group is a grouping of instances within a single Availability Zone. Placement groups are recommended for applications that need low network latency, high network throughput, or both.

Only certain instances can be launched in to a Clustered Placement Group.

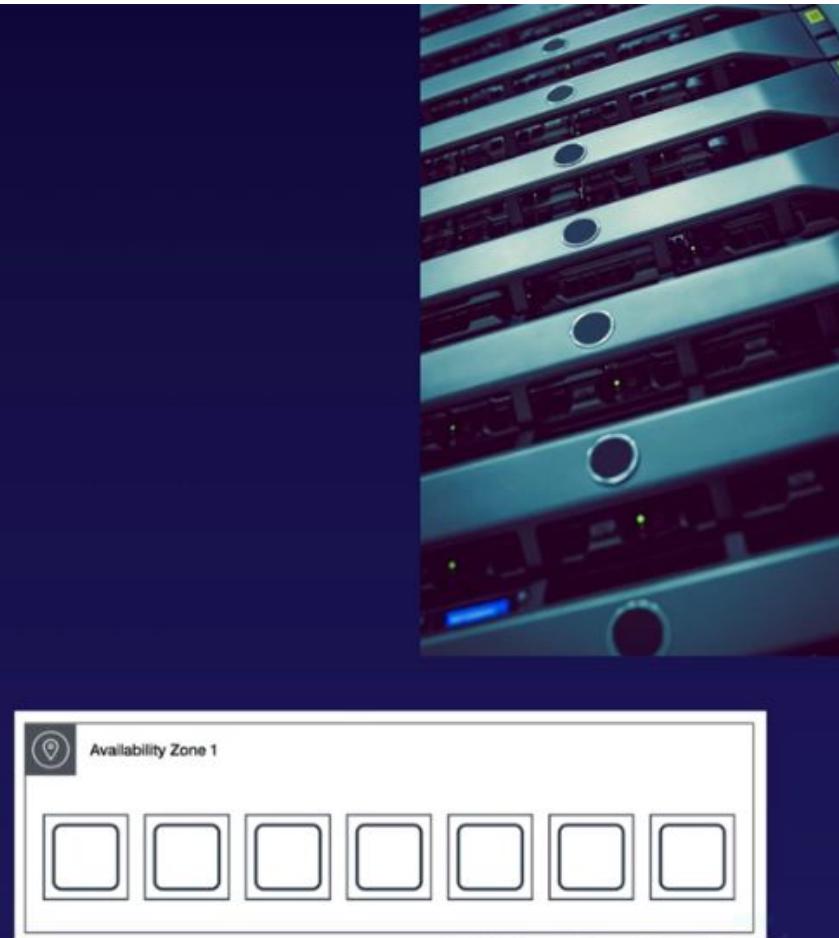


Spread Placement Group

A spread placement group is a group of instances that are each placed on distinct underlying hardware.

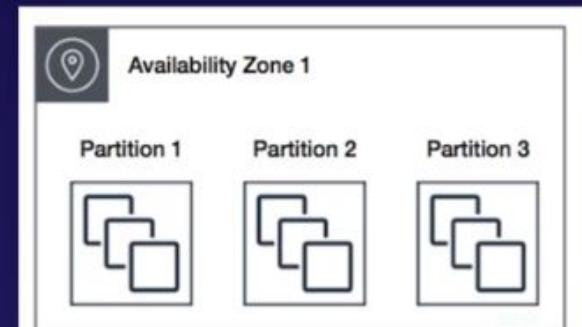
Spread placement groups are recommended for applications that have a small number of critical instances that should be kept separate from each other.

THINK INDIVIDUAL INSTANCES



Partitioned Placement Group

When using partition placement groups, Amazon EC2 divides each group into logical segments called partitions. Amazon EC2 ensures that each partition within a placement group has its own set of racks. Each rack has its own network and power source. No two partitions within a placement group share the same racks, allowing you to isolate the impact of hardware failure within your application.



THINK MULTIPLE INSTANCES

EC2 Placement Group

Three Types of Placement Groups;

- Clustered Placement Group
 - Low Network Latency / High Network Throughput
- Spread Placement Group
 - Individual Critical EC2 instances
- Partitioned
 - Multiple EC2 instances HDFS, HBase, and Cassandra

Exam Tips

- A clustered placement group can't span multiple Availability Zones.
- A spread placement and partitioned group can.
- The name you specify for a placement group must be unique within your AWS account.
- Only certain types of instances can be launched in a placement group (Compute Optimized, GPU, Memory Optimized, Storage Optimized)
- AWS recommend homogenous instances within clustered placement groups.
- You can't merge placement groups.
- You can move an existing instance into a placement group. Before you move the instance, the instance must be in the stopped state. You can move or remove an instance using the AWS CLI or an AWS SDK, you can't do it via the console yet.

HPC on AWS

HPC

It's never been easier to get started with **high-performance computing** (HPC) than in any other time in history — and AWS is the perfect place to perform it.

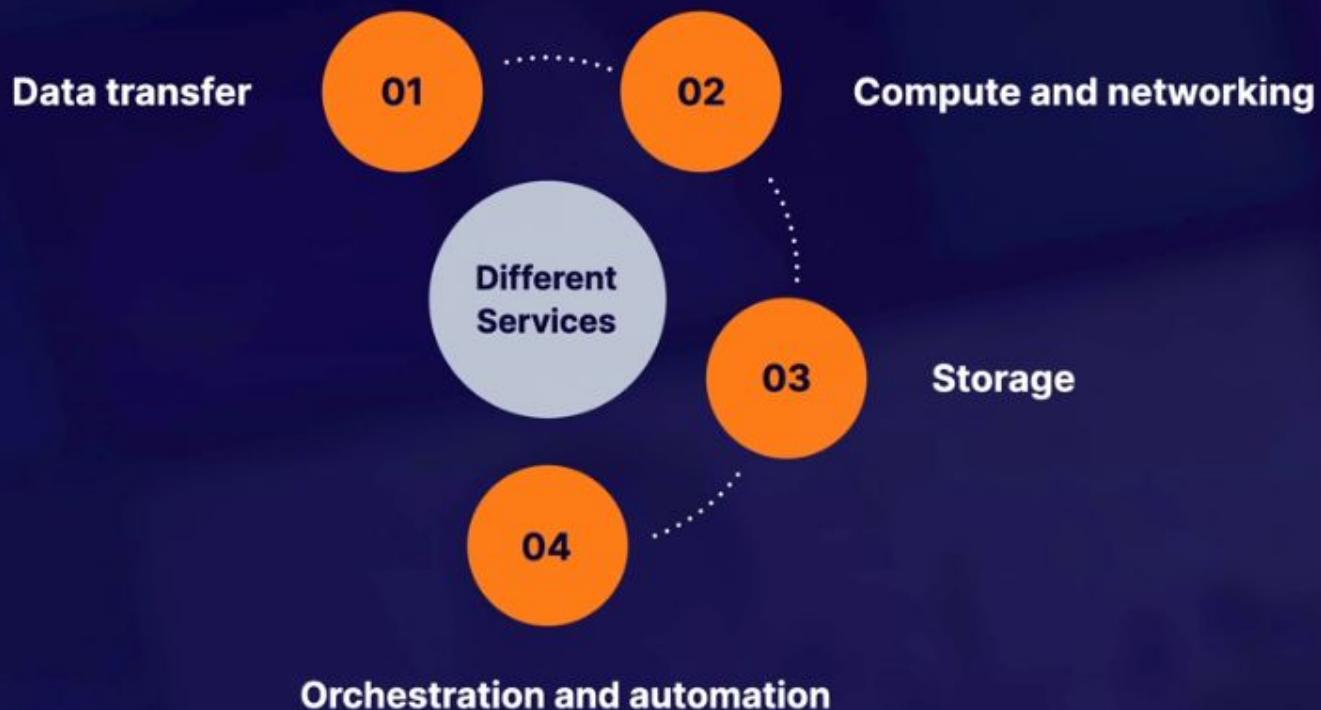
You can create a large number of resources in almost no time. You only pay for the resources you use — and, once finished, **you can destroy the resources**.

HPC is used for industries such as genomics, finance and financial risk modeling, machine learning, weather prediction, and even autonomous driving.



Achieving HPC on AWS

What are the different services we can use to achieve HPC on AWS?



HPC on AWS: Data Transfer

What are some ways we can get our data into AWS?

- ✓ Snowball, Snowmobile (**terabytes/petabytes** worth of data)
- ✓ **AWS DataSync** to store on S3, EFS, FSx for Windows, etc.
- ✓ Direct Connect

Data Transfer: AWS Direct Connect

AWS Direct Connect

AWS Direct Connect is a cloud service solution that makes it easy to establish a dedicated network connection from your premises to AWS. Using AWS Direct Connect, you can establish private connectivity between AWS and your data center, office, or colocation environment — which, in many cases, **can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience** than internet-based connections.



Data Transfer: Compute & Networking

What are the **compute and networking services** that allow us to achieve HPC on AWS?



EC2 instances that are GPU or CPU optimized



Enhanced networking



EC2 fleets (Spot Instances or Spot Fleets)



Elastic Network Adapters



Placement groups (cluster placement groups)



Elastic Fabric Adapters

Data Transfer: Enhanced Networking

What Is Enhanced Networking?

- It uses **single root I/O virtualization (SR-IOV)** to provide high-performance networking capabilities on supported instance types. SR-IOV is a method of device virtualization that provides **higher I/O performance** and **lower CPU utilization** when compared to traditional virtualized network interfaces.
- Enhanced networking provides higher bandwidth, higher packet per second (PPS) performance, and consistently lower inter-instance latencies. There is no additional charge for using enhanced networking.
- Use where you want good **network performance**.

Data Transfer: Enhanced Networking

Depending on your instance type, enhanced networking can be enabled using an:

- **Elastic Network Adapter (ENA)**, which supports network speeds of up to **100 Gbps** for supported instance types.

Or

- Intel 82599 **Virtual Function (VF)** interface, which supports network speeds of up to **10 Gbps** for supported instance types. This is typically used on older instances (**LEGACY**).

Note: In any **scenario question**, if given the option, you probably want to **choose ENA over VF**.

Data Transfer: Enhanced Networking

Depending on your instance type, enhanced networking can be enabled using an:

- **Elastic Network Adapter (ENA)**, which supports network speeds of up to **100 Gbps** for supported instance types.

Or

- Intel 82599 **Virtual Function (VF)** interface, which supports network speeds of up to **10 Gbps** for supported instance types. This is typically used on older instances (**LEGACY**).

Note: In any **scenario question**, if given the option, you probably want to **choose ENA over VF**.

Data Transfer: Enhanced Networking

Depending on your instance type, enhanced networking can be enabled using an:

- **Elastic Network Adapter (ENA)**, which supports network speeds of up to **100 Gbps** for supported instance types.

Or

- Intel 82599 **Virtual Function (VF)** interface, which supports network speeds of up to **10 Gbps** for supported instance types. This is typically used on older instances (**LEGACY**).

Note: In any **scenario question**, if given the option, you probably want to **choose ENA over VF**.

Data Transfer: Elastic Fabric Adapter



What is an Elastic Fabric Adapter?

- An **Elastic Fabric Adapter (EFA)** is a network device you can attach to your Amazon EC2 instance to accelerate HPC and machine learning applications.
- EFA provides **lower, more consistent latency** and **higher throughput** than the TCP transport traditionally used in cloud-based HPC systems.
- EFA can use **OS-bypass**, which enables HPC and machine learning applications to bypass the operating system kernel and communicate directly with the EFA device. It makes it a lot faster with much lower latency. It is not supported with Windows currently — only Linux.

Data Transfer: Storage

What are the **storage services** that allow us to achieve HPC on AWS?

Instance-attached storage:

- **EBS**: Scale up to 64,000 IOPS with Provisioned IOPS (PIOPS)
- **Instance Store**: Scale to millions of IOPS; low latency

Network storage:

- **Amazon S3**: Distributed object-based storage; not a file system
- **Amazon EFS**: Scale IOPS based on total size, or use Provisioned IOPS
- **Amazon FSx for Lustre**: HPC-optimized distributed file system; millions of IOPs, which is also backed by S3

Data Transfer: Orchestration & Automation

What are the orchestration and automation services that allow us to achieve HPC on AWS?

AWS Batch



AWS Batch enables developers, scientists, and engineers to easily and efficiently run hundreds of thousands of batch computing jobs on AWS.



AWS Batch supports multi-node parallel jobs, which allows you to run a single job that spans **multiple EC2 instances**.



You can easily schedule jobs and launch **EC2 instances** according to your needs.

Data Transfer: AWS ParallelCluster

AWS ParallelCluster

1

Open-source cluster management tool that **makes it easy for you to deploy and manage** HPC clusters on AWS.

2

ParallelCluster uses a **simple text file to model and provision all the resources needed** for your HPC applications in an automated and secure manner.

3

Automate creation of VPC, subnet, cluster type, and instance types.

Exam tips

We can achieve HPC on AWS through:



Data transfer



Compute and networking



Storage



Orchestration and automation

Exam tips

Data Transfer

- Snowball, Snowmobile (**terabytes/petabytes worth of data**)
- AWS DataSync to store on S3, EFS, FSx for Windows, etc.
- Direct Connect

Exam tips

Compute & Networking

- ✓ EC2 instances that are **GPU** or **CPU** optimized
- ✓ **EC2 fleets** (Spot Instances or Spot Fleets)
- ✓ Placement groups (cluster placement groups)
- ✓ Enhanced networking single root I/O virtualization (**SR-IOV**)
- ✓ Elastic Network Adapters or **Intel 82599 Virtual Function** (VF) interface
- ✓ Elastic Fabric Adapters

Exam tips

Storage

Instance-attached storage:

- **EBS:** Scale up to 64,000 IOPS with Provisioned IOPS (PIOPS)
- **Instance Store:** Scale to millions of IOPS; low latency

Network storage:

- **Amazon S3:** Distributed object-based storage; not a file system.
- **Amazon EFS:** Scale IOPS based on total size, or use Provisioned IOPS
- **Amazon FSx for Lustre:** HPC-optimized distributed file system; millions of IOPs, which is also backed by S3

Exam tips

Orchestration & Automation

- ✓ AWS Batch
- ✓ AWS ParallelCluster



AWS WAF

WAF?

AWS WAF is a web application firewall that lets you monitor the HTTP and HTTPS requests that are forwarded to Amazon CloudFront, an Application Load Balancer or API Gateway.

AWS WAF also lets you control access to your content.



WAF?

You can configure conditions such as what IP addresses are allowed to make this request or what query string parameters need to be passed for the request to be allowed.

Then the application load balancer or CloudFront or API Gateway will either allow this content to be received or to give a HTTP 403 Status Code.



WAF?

At its most basic level, AWS WAF allows 3 different behaviours:

- 1 Allow all requests except the ones you specify
- 2 Block all requests except the ones you specify
- 3 Count the requests that match the properties you specify



WAF?

Extra protection against web attacks using conditions you specify. You can define conditions by using characteristics of web requests such as:

- IP addresses that requests originate from.
- Country that requests originate from.
- Values in request headers.
- Strings that appear in requests, either specific strings or string that match regular expression (regex) patterns.
- Length of requests.
- Presence of SQL code that is likely to be malicious (known as SQL injection).
- Presence of a script that is likely to be malicious (known as cross-site scripting).

Exam tips

In the exam you will be given different scenarios and you will be asked how to block malicious IP addresses.

- Use AWS WAF
- Use Network ACLs - We will cover this in more detail in the VPC section of the course.

EC2 Summary

EC2 Exam Tips

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change.



What is EFS?

Amazon Elastic File System (Amazon EFS) is a file storage service for Amazon Elastic Compute Cloud (Amazon EC2) instances. Amazon EFS is easy to use and provides a simple interface that allows you to create and configure file systems quickly and easily. With Amazon EFS, storage capacity is elastic, growing and shrinking automatically as you add and remove files, so your applications have the storage they need, when they need it.



CLI Exam Tips

- You can interact with AWS from anywhere in the world just by using the command line (CLI).
- You will need to set up access in IAM
- Commands themselves are not in the exam, but some basic commands will be useful to know for real life.

CLI Exam Tips

- Roles are more secure than storing your access key and secret access key on individual EC2 instances.
- Roles are easier to manage.
- Roles can be assigned to an EC2 instance after it is created using both the console & command line.
- Roles are universal — you can use them in any region.

Instance Metadata Exam Tips

- Used to get information about an instance (such as public ip)
- curl <http://169.254.169.254/latest/meta-data/>
- curl <http://169.254.169.254/latest/user-data/>

EC2 Quizz

EC2 Quizz

QUESTION 1

When updating the policy used by an IAM Role attached to an EC2 instance, what needs to happen for the changes to take effect?

- Wait up to 15 minutes for the change to take effect
- Reboot the instance to force the change
- Reattach the IAM Role to the EC2 instance
- Nothing - It will take effect immediately

Good work!

Changes to IAM Policies take effect almost immediately (with maybe a few seconds delay). No substantial waiting time is required, nor changes to the system. This is because the IAM Policy exists in the AWS API, rather than on the instance itself. As a way to remember it in a scenario, if you think about a compromised system, you would need to revoke the access immediately, without waiting for changes to take effect.

EC2 Quizz

QUESTION 2

Can Spread Placement Groups be deployed across multiple Availability Zones?

Yes.

No.

Yes, but only using the AWS API.

Only in Us-East-1.

Sorry!

Correct Answer

Spread Placement Groups can be deployed across availability zones since they spread the instances further apart. Cluster Placement Groups can only exist in one Availability Zone since they are focused on keeping instances together, which you cannot do across Availability Zones

EC2 Quizz

QUESTION 3

Will an Amazon EBS root volume persist independently from the life of the terminated EC2 instance to which it was previously attached? In other words, if I terminated an EC2 instance, would that EBS root volume persist?

- Yes - It will always persist until deleted manually
- Yes - Unless 'Delete on Termination' is unchecked for the volume
- No - It will always be deleted immediately on termination
- Yes - But only for certain instance types

Correct Answer

You can control whether an EBS root volume is deleted when its associated instance is terminated. The default delete-on-termination behaviour depends on whether the volume is a root volume, or an additional volume. By default, the `DeleteOnTermination` attribute for root volumes is set to 'true.' However, this attribute may be changed at launch by using either the AWS Console or the command line. For an instance that is already running, the `DeleteOnTermination` attribute must be changed using the CLI.

EC2 Quizz

QUESTION 4

Can I delete a snapshot of an EBS Volume that is used as the root device of a registered AMI?

Only via the Command-Line.

Only using the AWS API.

No.

Yes.

Good work!

If the original snapshot was deleted, then the AMI would not be able to use it as the basis to create new instances. For this reason, AWS protects you from accidentally deleting the EBS Snapshot, since it could be critical to your systems. To delete an EBS Snapshot attached to a registered AMI, first remove the AMI, then the snapshot can be deleted

EC2 Quizz

QUESTION 5

When can you attach a IAM Role to an EC2 instance?

Anytime, without restriction

Anytime, only if there isn't already an attached IAM Role

Only during launch, and cannot be changed once the instance is launched

Anytime, but the instance must be stopped

Good work!

IAM Roles can be attached or detached from instances at any time, regardless of whether the instance is started or stopped. This is important to be able to do for security measures. Prior to early 2017, you would only be able to attach an IAM role at launch, and if you wanted to attach a role, you would have to terminate and re-launch the instance.

EC2 Quizz

QUESTION 6

What is the underlying Hypervisor for EC2?

Choose 2

Hyper-V

ESX

Nitro

Xen

OVM

Good work!

AWS originally used a modified version of the Xen Hypervisor to host EC2. In 2017, AWS began rolling out their own Hypervisor called Nitro

EC2 Quizz

QUESTION 7

To retrieve instance metadata or user data you will need to use the following IP Address:

- http://192.168.0.254
- http://127.0.0.1
- http://10.0.0.1
- http://169.254.169.254

Good work!

This IP Address is specific to AWS, where you can use it on any instance to acquire information about that instance. It is a specific type of address called a 'link-local address', and is only accessible from that particular instance. You can also disable the metadata service to prevent its misuse.

EC2 Quizz

QUESTION 8

Which of the following features only relate to Spread Placement Groups?

The name of your placement group must be unique within your AWS Account

The placement group can only have 7 running instances per Availability Zone

Instances must be deployed in a single Availability Zone

There is no charge for creating a placement group

Sorry!

Correct Answer

Spread placement groups have a specific limitation that you can only have a maximum of 7 running instances per Availability Zone and therefore this is the only correct option. Deploying instances in a single Availability Zone is unique to Cluster Placement Groups only and therefore is not correct. The last two remaining options are common to all placement group types and so are not specific to Spread Placement Groups.

EC2 Quizz

QUESTION 9

Where in the AWS Global Infrastructure are EC2 instance provisioned?

In Availability Zones

Globally

In Regions

Good work!

When you're setting up an EC2 instance, you select which subnet you'd like to place your EC2 instance in. Each subnet is tied to a specific availability zone. You cannot move an instance between Availability Zones, without setting up a copied version of the instance. Whilst they exist in Regions, they are not portable across the whole region, nor across the whole globe

EC2 Quizz

QUESTION 10

Which of the following provide the lowest cost EBS options?

Choose 2

Throughput Optimized (st1)

Provisioned IOPS (io1)

General Purpose (gp2)

Cold (sc1)

Sorry!

Correct Answer

Of all the EBS types, both current and of the previous generation, HDD based volumes will always be less expensive than SSD types. Therefore, of the options available in the question, the Cold (sc1) and Throughput Optimized (st1) types are HDD based and will be the lowest cost options.

EC2 Quizz

QUESTION 11

Standard Reserved Instances can be moved between regions



False



True

Good work!

Standard Reserved Instances cannot be moved between regions. You can choose if a Reserved Instance applies to either a specific Availability Zone, or an Entire Region, but you cannot change the region.

EC2 Quizz

QUESTION 12

In order to enable encryption at rest using EC2 and Elastic Block Store, you must ____.

Configure encryption when creating the EBS volume

Configure encryption using X.509 certificates

Configure encryption using the appropriate Operating Systems file system

Mount the EBS volume in to S3 and then encrypt the bucket using a bucket policy.

Good work!

The use of encryption at rest is default requirement for many industry compliance certifications. Using AWS managed keys to provide EBS encryption at rest is a relatively painless and reliable way to protect assets and demonstrate your professionalism in any commercial situation.

EC2 Quizz

QUESTION 13

Which AWS CLI command should I use to create a snapshot of an EBS volume?

aws ec2 deploy-snapshot

aws ec2 create-snapshot

aws ec2 fresh-snapshot

aws ec2 new-snapshot

Good work!

When looking at the AWS CLI, remember the verbs, like 'create', which are used as part of commands. This helps you build the necessary command in your head, without referring to the documentation. For example, we might a new image along with this snapshot. From this, we could understand that the command would likely be 'aws ec2 create-image'.

EC2 Quizz

QUESTION 14

Spread Placement Groups can be deployed across multiple Availability Zones



False



True

Good work!

Spread Placement Groups can be deployed across availability zones since they spread the instances further apart. Cluster Placement Groups can only exist in one Availability Zone since they are focused on keeping instances together, which you cannot do across Availability Zones

EC2 Quizz

QUESTION 15

When creating a new security group, all inbound traffic is allowed by default.

True

False

Good work!

There are slight differences between a normal 'new' Security Group and a 'default' security group in the default VPC. For a 'new' security group nothing is allowed in by default.

EC2 Quizz

QUESTION 16

EBS Snapshots are backed up to S3 in what manner?

EBS snapshots are NOT stored on S3.

Differentially

Incrementally

Exponentially

Sorry!

EBS snapshots use incremental backups and are stored in S3. Restores can be done from any of the snapshots. The original full snapshot can be safely deleted without impacting the ability to use the other related incremental backups.

Correct Answer

EBS snapshots use incremental backups and are stored in S3. Restores can be done from any of the snapshots. The original full snapshot can be safely deleted without impacting the ability to use the other related incremental backups.

EC2 Quizz

QUESTION 17

What type of storage are Amazon's EBS volumes based on?

Object-based

Block-based

File-based

Database-based

Good work!

EBS uses Block-based storage, where the data is stored on a virtual disk managed by the Operating System. EFS uses File-based storage, where the underlying filesystem is managed by AWS. S3 uses Object-based storage, where files are kept in a flat structure

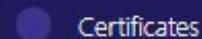
EC2 Quizz

QUESTION 18

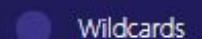
To help you manage your Amazon EC2 instances, you can assign your own metadata in the form of ____.



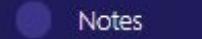
Tags



Certificates



Wildcards



Notes

Good work!

Tagging is a key part of managing an environment. Even in a lab, it is easy to lose track of the purpose of resources, and tricky determine why it was created and if it is still needed. This can rapidly translate into lost time and lost money.

EC2 Quizz

QUESTION 19

You need to know both the private IP address and public IP address of your EC2 instance. You should ____.

- Use the following command: AWS EC2 DisplayIP.
- Retrieve the instance User Data from <http://169.254.169.254/latest/user-data/>.
- Run IPCONFIG (Windows) or IFCONFIG (Linux).
- Retrieve the instance Metadata from <http://169.254.169.254/latest/meta-data/>.

Good work!

Instance Metadata and User Data can be retrieved from within the instance via a special URL. Similar information can be extracted by using the API via the CLI or an SDK. The ipconfig and ifconfig tools don't have the ability to see the Public IP Address directly, since it's attached dynamically inside the AWS Software Defined Network which has to be queried by the API

EC2 Quizz

QUESTION 28

The use of a cluster placement group is ideal __

- When you need to distribute content on a CDN network.
- When you need to deploy EC2 instances that require high disk IO.

 Your fleet of EC2 Instances requires low latency and high network throughput across multiple availability zones.

 Your fleet of EC2 instances requires high network throughput and low latency within a single availability zone.

Sorry!

Correct Answer

Cluster Placement Groups are primarily about keeping your compute resources within one network hop of each other on high speed rack switches. This is only helpful when you have compute loads with network loads that are either very high or very sensitive to latency.

EC2 Quizz

QUESTION 21

If an Amazon EBS volume is attached as an additional disk (not the root volume), can I detach it without stopping the instance?



Yes, although it may take some time.



No, you will need to stop the instance.

Good work!

Since the additional disk does not contain the operating system, you can detach it in the EC2 Console while the instance is running. However, any data on that drive would become inaccessible, and possibly cause problems for the EC2 instance.

EC2 Quizz

QUESTION 22

Is it possible to perform actions on an existing Amazon EBS Snapshot?

It depends on the region.

Yes, through the AWS APIs, CLI, and AWS Console.

No

EBS does not have snapshot functionality.

Good work!

EC2 Quizz

QUESTION 23

Which EC2 feature allows you to utilize SR-IOV?

- CloudWatch Agent
- Bootstrap Scripts (User Data)
- IAM Roles
- Enhanced Networking

Good work!

SR-IOV, or Single Root I/O Virtualization, is a feature of Enhanced Networking used to provide higher networking performance. On a normal EC2 instance, multiple EC2 instances may share a single physical network interface on the EC2 Host. SR-IOV effectively dedicates the interface to a single instance, and bypasses parts of the Hypervisor, allowing for better performance

EC2 Quizz

QUESTION 24

Which service would you use to run a general Windows File Server with minimal overhead?

EFS

S3

EBS Multi Attach

FSx for Windows

Good work!

Amazon FSx for Windows File Server provides a fully managed native Microsoft Windows file system so you can easily move your Windows-based applications that require shared file storage to AWS. EBS Multi Attach allows you to attach a volume to up to 16 instances, but would have issues across multiple availability zones, and could not use NTFS natively. EFS uses the NFS protocol, and is explicitly not supported on Windows. S3 is object-based storage, and would not be suitable as the backend for a file server.

CHAPTER 5

Databases on AWS

Databases 101 :: What is a relational Database

Relational databases are what most of us are all used to. They have been around since the 70's.
Think of a traditional spreadsheet:

- Database
- Tables
- Row
- Fields (Columns)



Databases 101 :: What is a relational Database

ID	First Name	Surname	Gender
1	Ryan	Kroonenburg	M
2	John	Adams	M
3	Julia	Clark	F
4	Danielle	Dustagheer	F

Databases 101 :: What is a relational Database

Relational databases on AWS;

- SQL Server
- Oracle
- MySQL Server
- PostgreSQL
- Aurora
- MariaDB



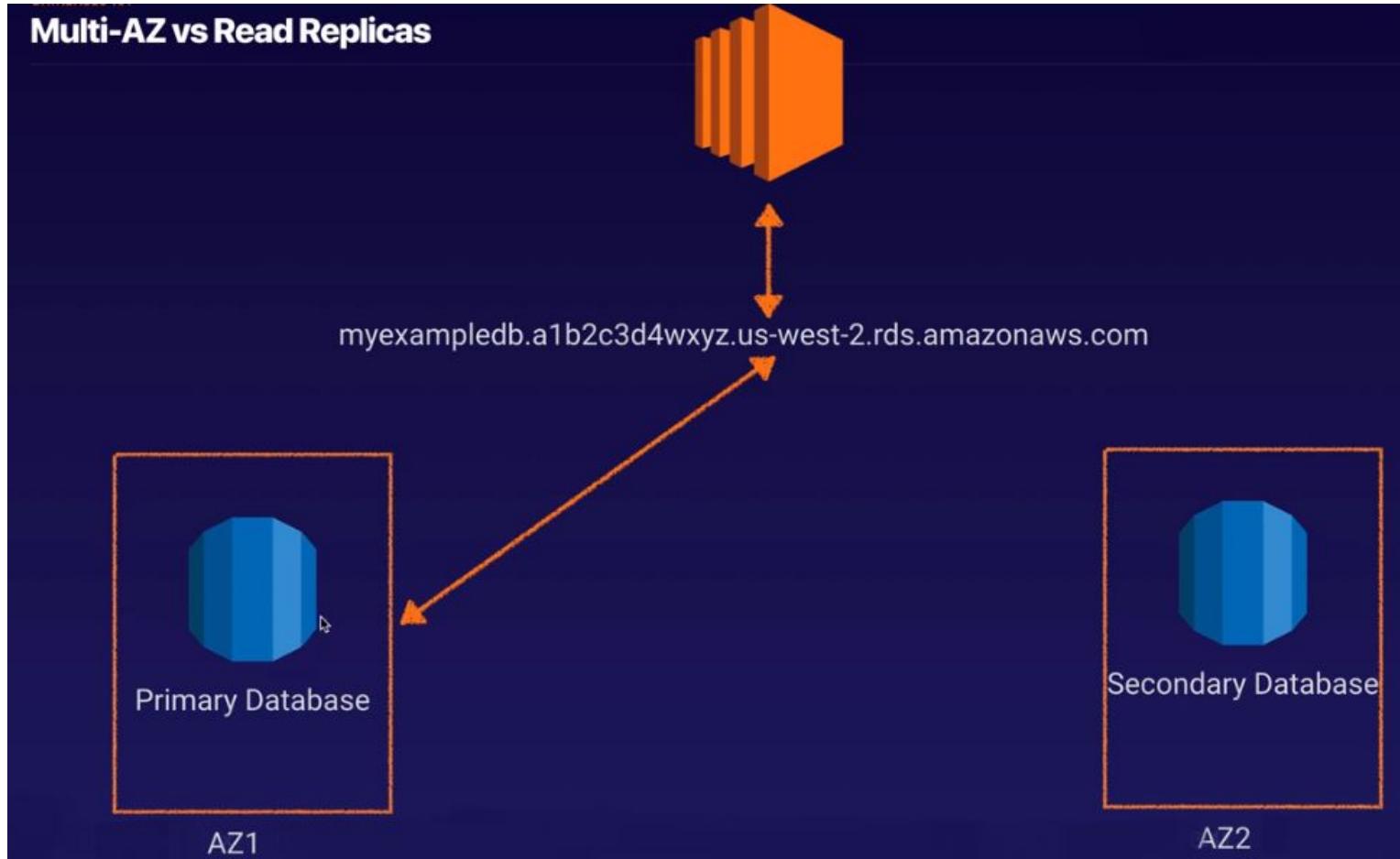
Databases 101 :: Multi-AZ vs Read replicas

RDS has two key features;

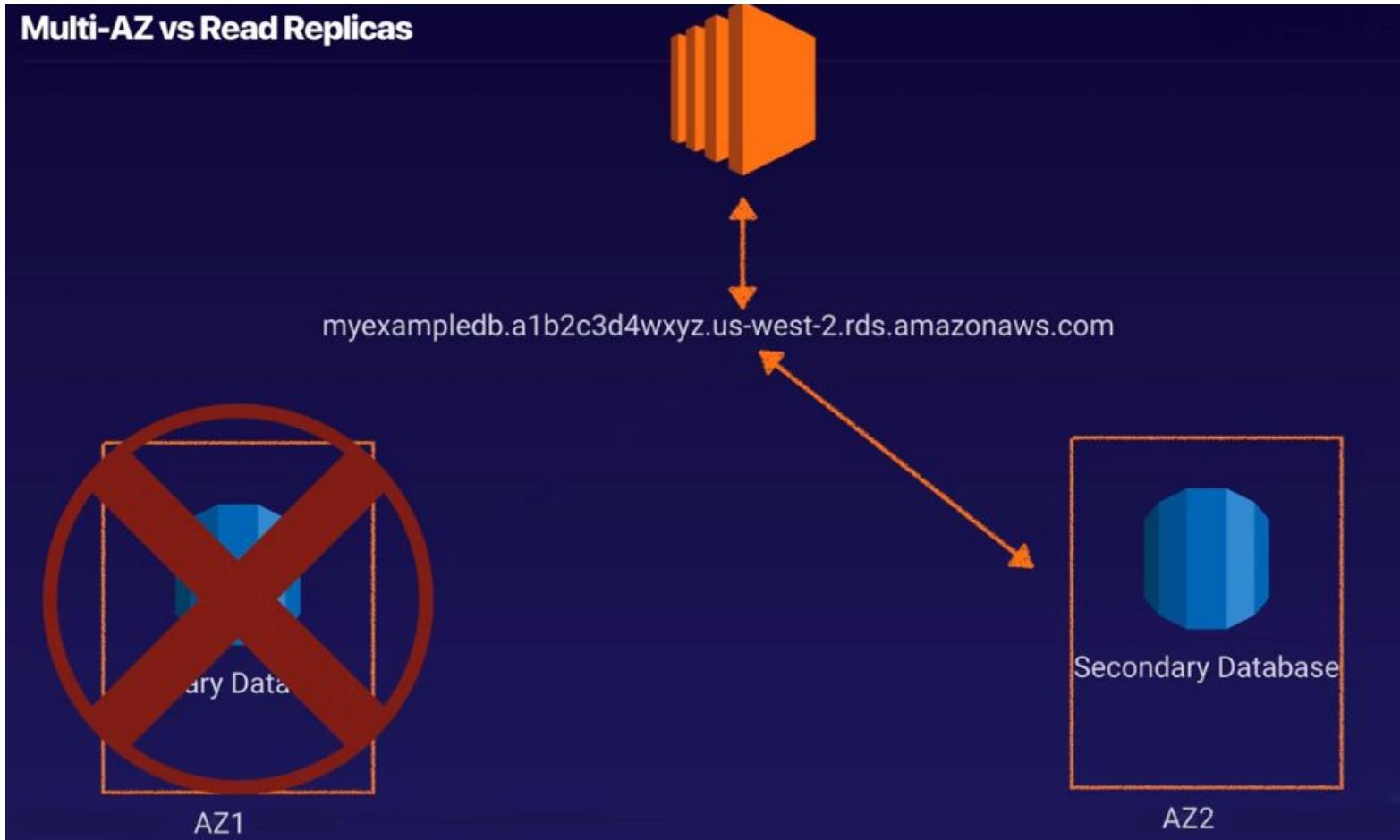
- Multi-AZ - For Disaster Recovery
- Read Replicas - For Performance



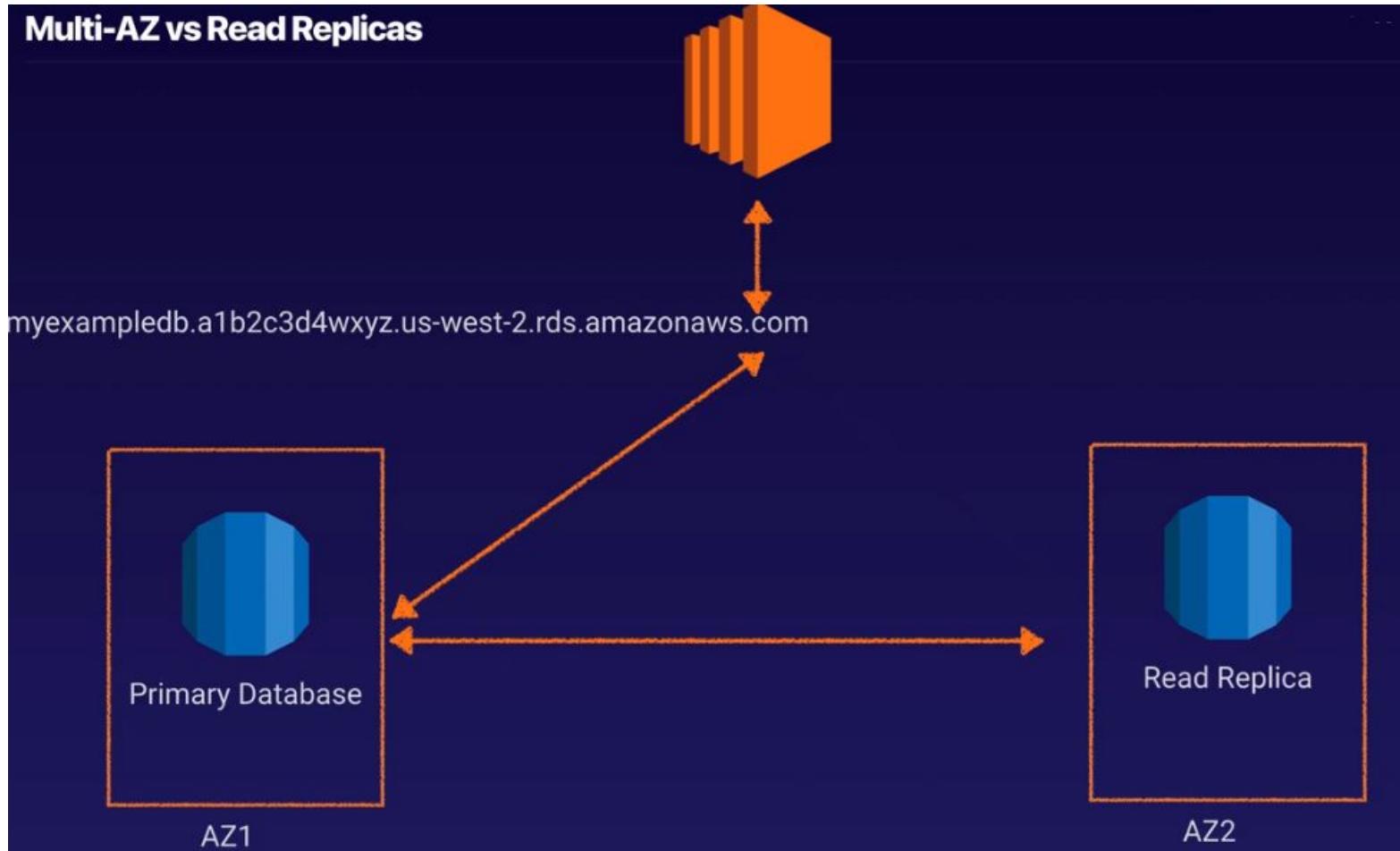
Databases 101 :: Failover with Multi-AZ



Databases 101 :: Failover with Multi-AZ

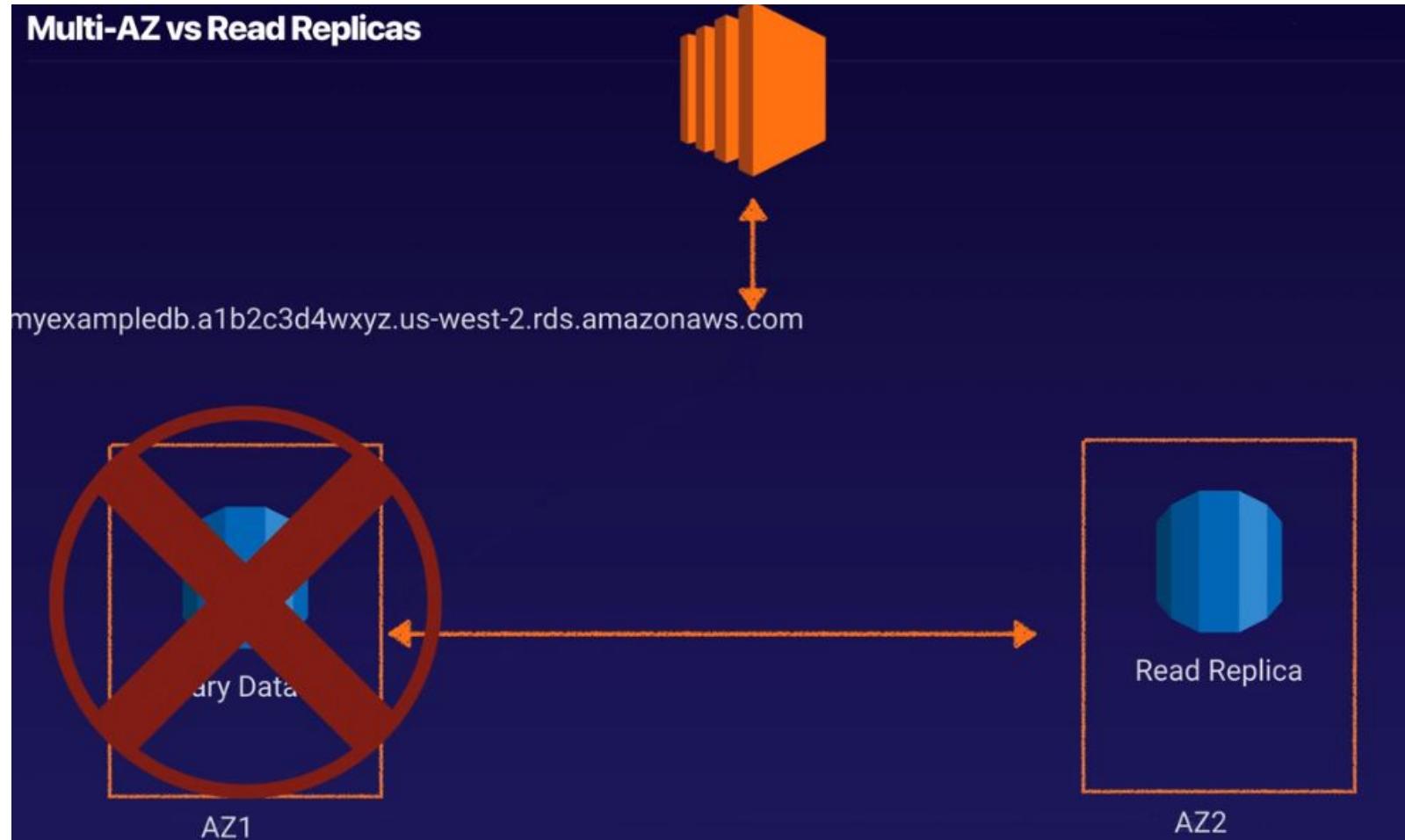


Databases 101 :: Read replicas

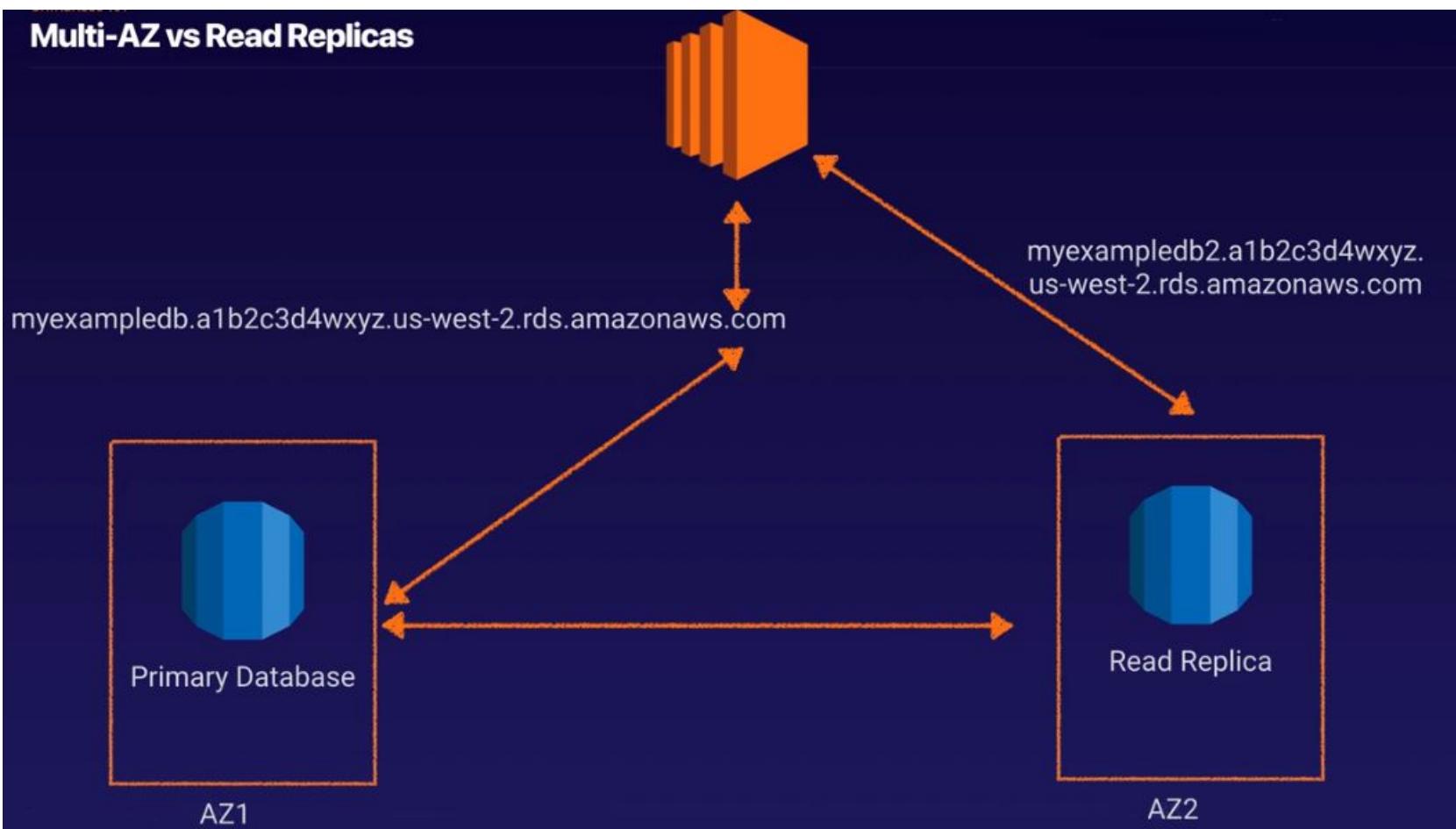


Databases 101 :: Read replicas

Multi-AZ vs Read Replicas



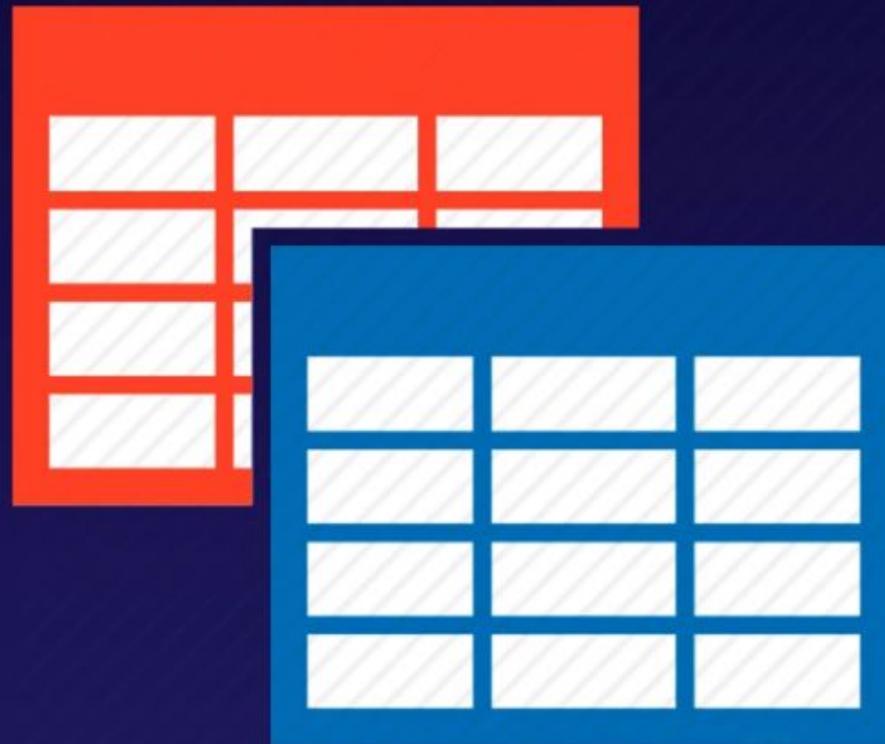
Databases 101 :: Read replicas



Databases 101

Non Relational Databases are as follows;

- Collection = Table
- Document = Row
- Key Value Pairs = Fields



Databases 101

```
{  
  "_id": "51262c865ca358946be09d77",  
  "firstname": "John",  
  "surname": "Smith",  
  "Age": "23",  
  "address": [  
    {"street": "21 Jump Street",  
     "suburb": "Richmond"}  
  ]  
}
```

Databases 101

Used for business intelligence. Tools like Cognos, Jaspersoft, SQL Server Reporting Services, Oracle Hyperion, SAP NetWeaver.

Used to pull in very large and complex data sets. Usually used by management to do queries on data (such as current performance vs targets etc)



Databases 101

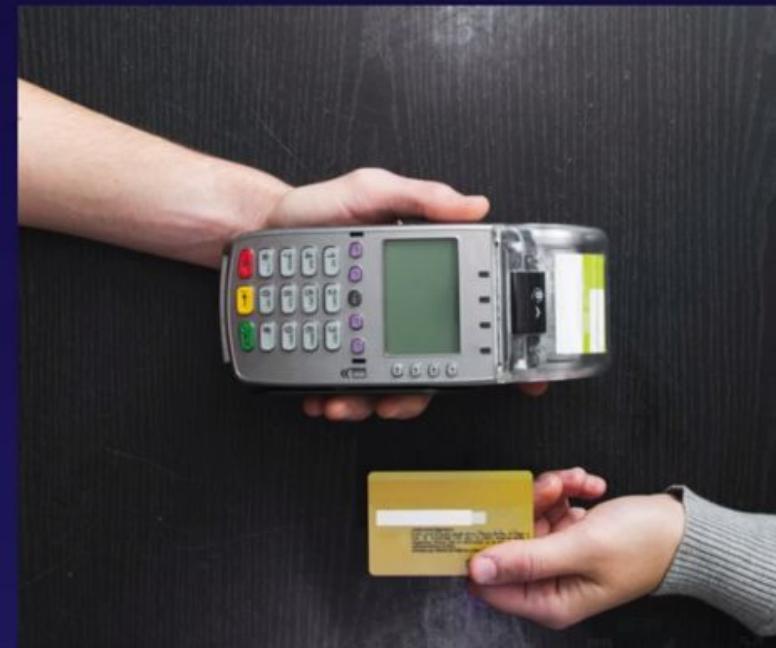
OLTP vs OLAP

Online Transaction Processing (OLTP) differs from OLAP Online Analytics Processing (OLAP) in terms of the types of queries you will run.

OLTP Example:

Order number 2120121

Pulls up a row of data such as Name, Date, Address to Deliver to, Delivery Status etc.



Databases 101

OLTP vs OLAP

OLAP transaction Example:

Net Profit for EMEA and Pacific for the Digital Radio Product.

Pulls in large numbers of records

Sum of Radios Sold in EMEA

Sum of Radios Sold in Pacific

Unit Cost of Radio in each region

Sales price of each radio

Sales price - unit cost.



Databases 101

OLTP vs OLAP

Data Warehousing databases use different type of architecture both from a database perspective and infrastructure layer.

**Amazon's Data
Warehouse
Solution Is Called
Redshift**



Databases 101 :: ElasticCache

ElastiCache is a web service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud. The service improves the performance of web applications by allowing you to retrieve information from fast, managed, in-memory caches, instead of relying entirely on slower disk-based databases.

ElastiCache supports two open-source in-memory caching engines:

Databases 101 :: ElasticCache

ElasticCache supports two open-source in-memory caching engines:

- Memcached
- Redis



redis

RDS: Backups, Multi-AZ, and Read Replicas

RDS: Backups, Multi-AZ, and Read Replicas

There are two different types of Backups for RDS:

- Automated Backups
- Database Snapshots



Automated Backups

Automated Backups allow you to recover your database to any point in time within a “retention period”. The retention period can be between one and 35 days. Automated Backups will take a full daily snapshot and will also store transaction logs throughout the day. When you do a recovery, AWS will first choose the most recent daily back up, and then apply transaction logs relevant to that day. This allows you to do a point in time recovery down to a second, within the retention period.



Automated Backups

Automated Backups are enabled by default. The backup data is stored in S3 and you get free storage space equal to the size of your database. So if you have an RDS instance of 10Gb, you will get 10Gb worth of storage.

Backups are taken within a defined window. During the backup window, storage I/O may be suspended while your data is being backed up and you may experience elevated latency.



Database snapshot

DB Snapshots are done manually (ie they are user initiated.) They are stored even after you delete the original RDS instance, unlike automated backups.



Restore Backups

Whenever you restore either an Automatic Backup or a manual Snapshot, the restored version of the database will be a new RDS instance with a new DNS endpoint.



original.eu-west-1.rds.amazonaws.com

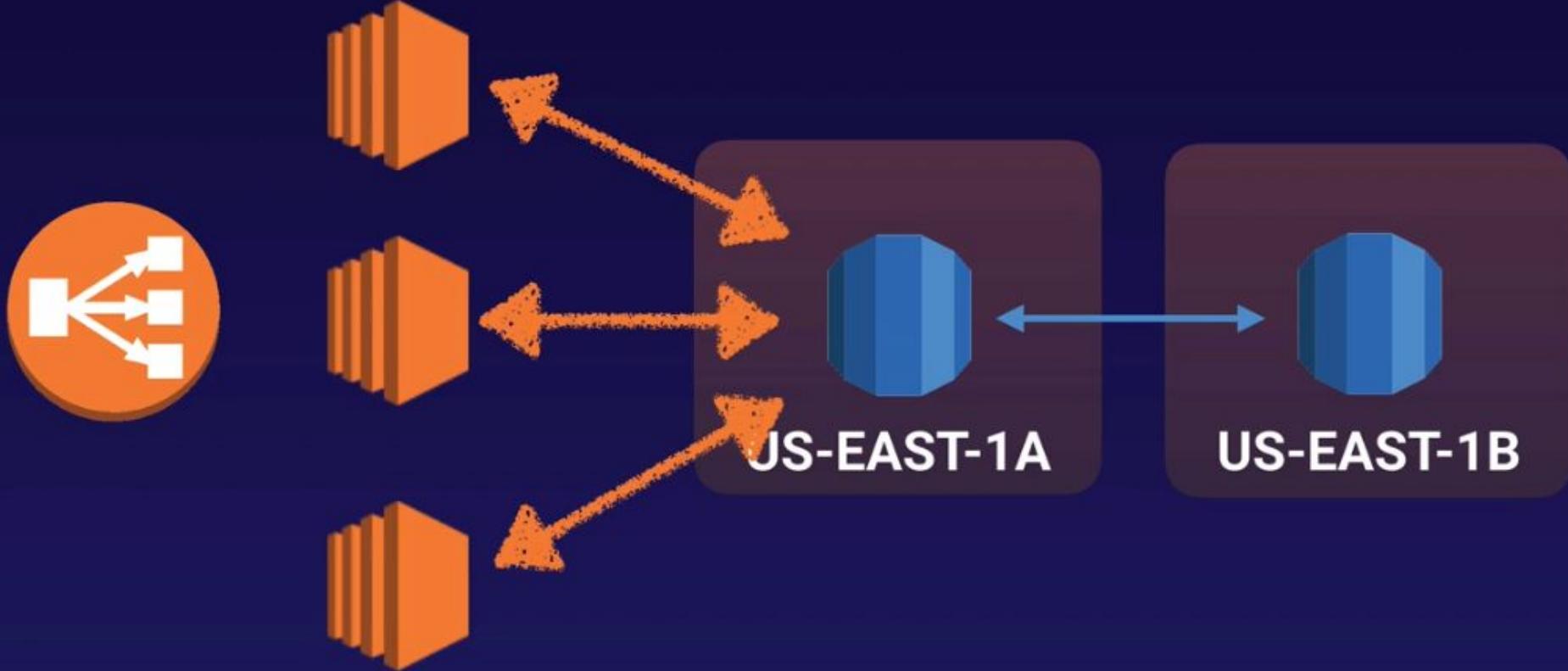
restored.eu-west-1.rds.amazonaws.com

Encryption at Rest

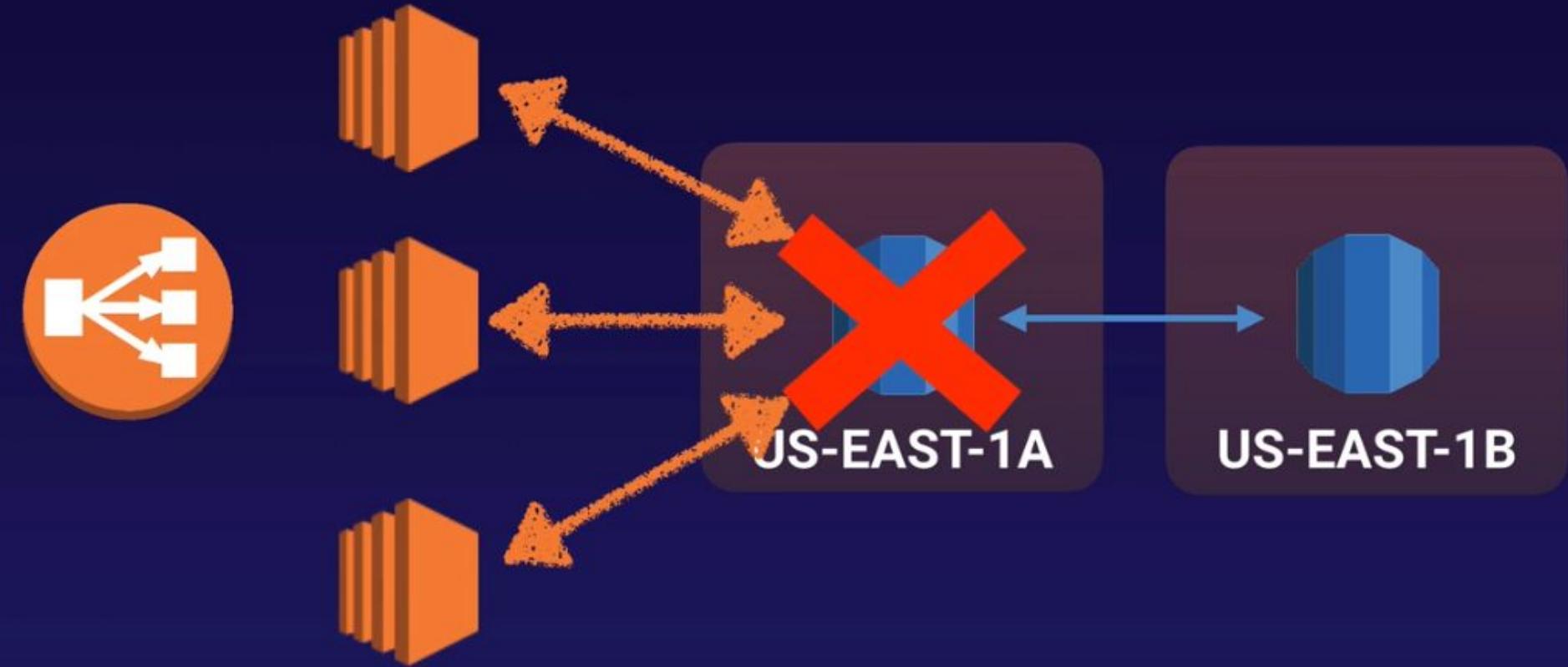
Encryption at rest is supported for MySQL, Oracle, SQL Server, PostgreSQL, MariaDB & Aurora. Encryption is done using the AWS Key Management Service (KMS) service. Once your RDS instance is encrypted, the data stored at rest in the underlying storage is encrypted, as are its automated backups, read replicas, and snapshots.



What is Multi-AZ



What is Multi-AZ



What is Multi-AZ

Multi-AZ allows you to have an exact copy of your production database in another Availability Zone. AWS handles the replication for you, so when your production database is written to, this write will automatically be synchronized to the stand by database.

In the event of planned database maintenance, DB Instance failure, or an Availability Zone failure, Amazon RDS will automatically failover to the standby so that database operations can resume quickly without administrative intervention.

What is Multi-AZ

Multi-AZ is for Disaster Recovery only



It is not primarily used for improving performance. For performance improvement, you need Read Replicas.

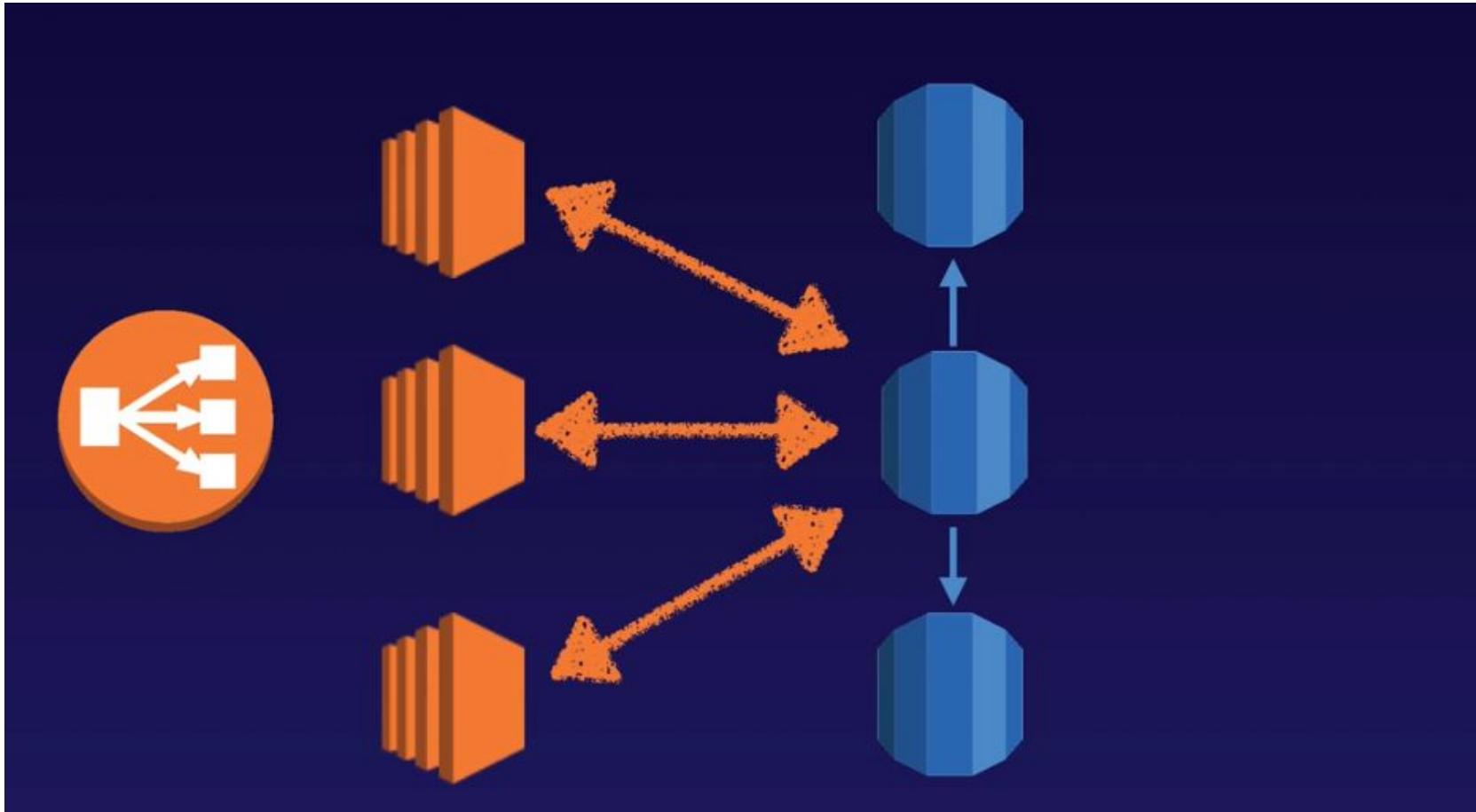
What is Multi-AZ

Multi-AZ is available for the following databases

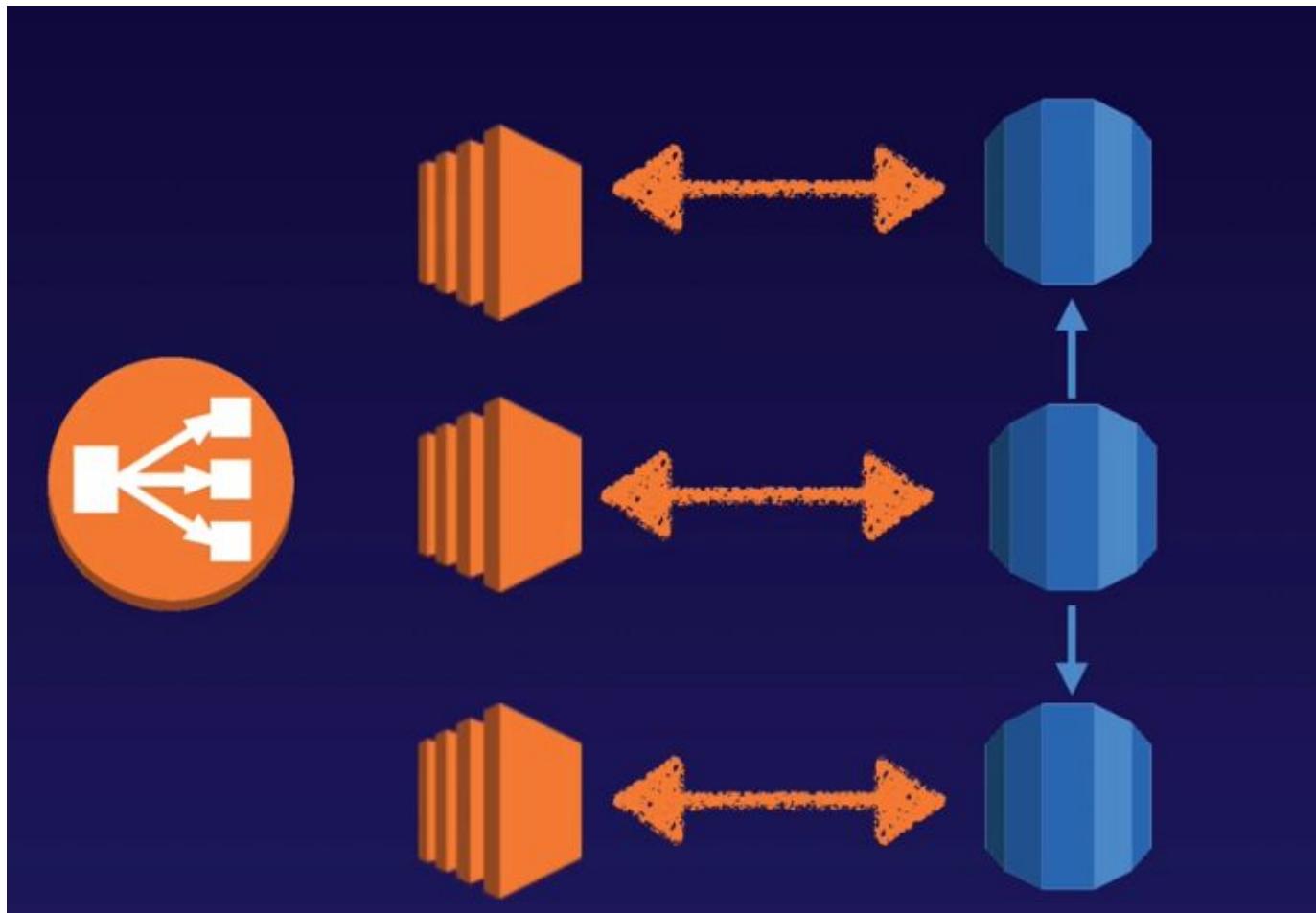
- SQL Server
- Oracle
- MySQL Server
- PostgreSQL
- MariaDB



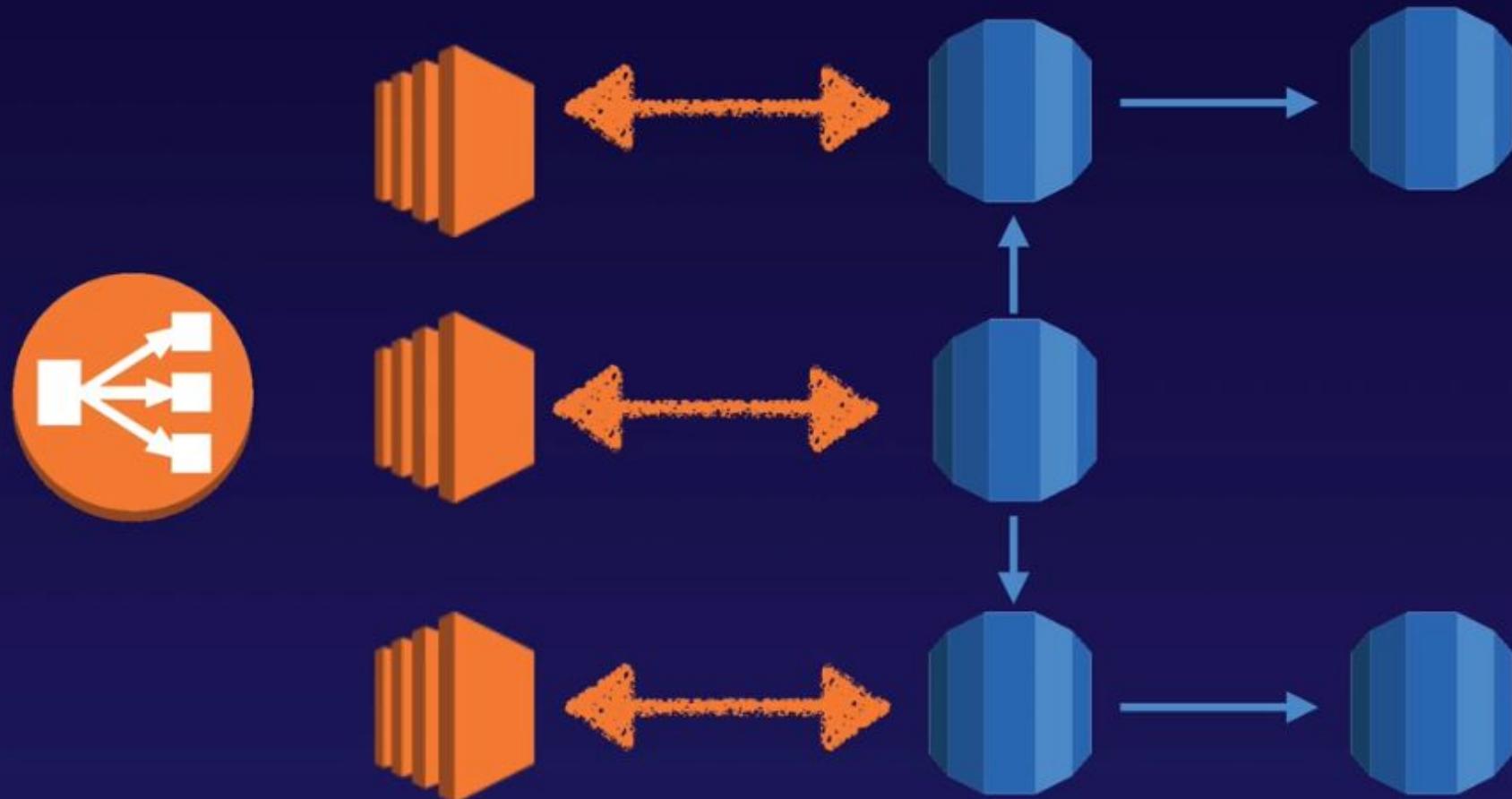
What is A read replica



What is A read replica

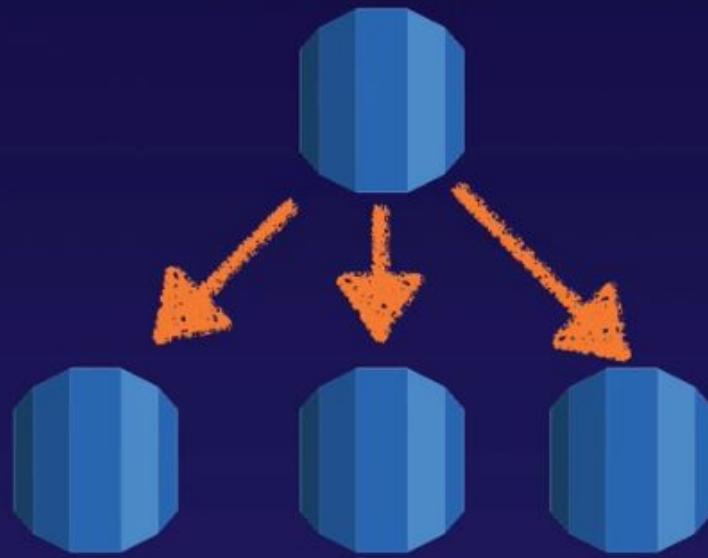


What is A read replica



What is A read replica

Read replicas allow you to have a read-only copy of your production database. This is achieved by using Asynchronous replication from the primary RDS instance to the read replica. You use read replicas primarily for very read-heavy database workloads.



Backup with RDS

Read Replicas are available for the following databases

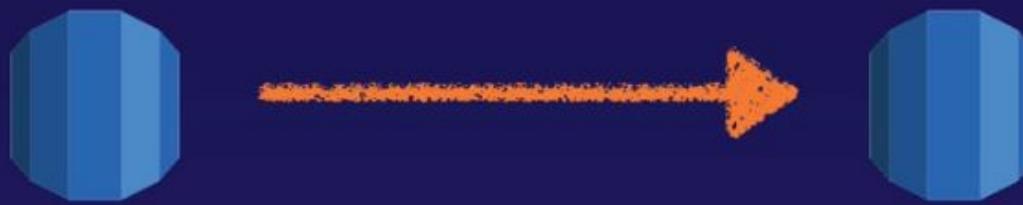
- MySQL Server
- PostgreSQL
- MariaDB
- Oracle
- Aurora



Backup with RDS

Things to know about Read Replicas;

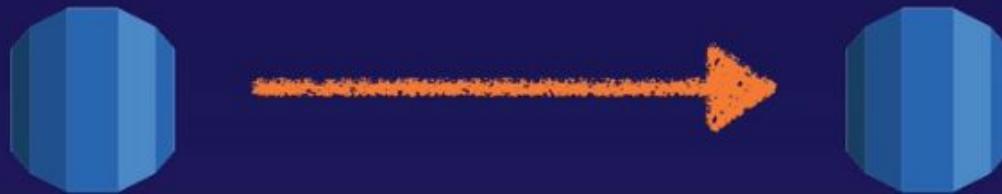
- Used for scaling, not for DR!
- Must have automatic backups turned on in order to deploy a read replica.
- You can have up to 5 read replica copies of any database.
- You can have read replicas of read replicas (but watch out for latency.)



Backup with RDS

Things to know about Read Replicas;

- Each read replica will have its own DNS end point.
- You can have read replicas that have Multi-AZ.
- You can create read replicas of Multi-AZ source databases.
- Read replicas can be promoted to be their own databases. This breaks the replication.
- You can have a read replica in a second region.



DynamoDB

What is DynamoDB?

Amazon DynamoDB is a fast and flexible NoSQL database service for all applications that need consistent, single-digit millisecond latency at any scale. It is a fully managed database and supports both document and key-value data models. Its flexible data model and reliable performance make it a great fit for mobile, web, gaming, ad-tech, IoT, and many other applications.



What is DynamoDB?

The basics of DynamoDB are as follows;

- Stored on SSD storage
- Spread across 3 geographically distinct data centres
- Eventual Consistent Reads (Default)
- Strongly Consistent Reads

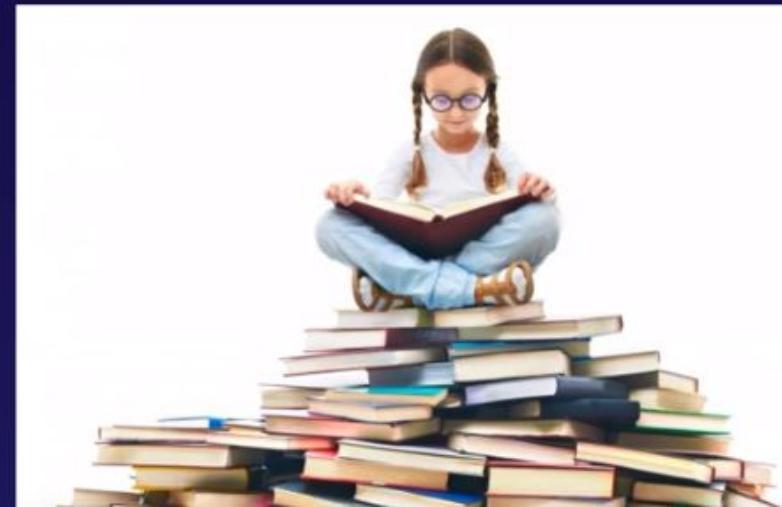


```
    document* item = el->FirstChild();
    GroupDesc* ElementDesc elDesc;
    #ifndef using sp_name = item->Attribute("name");
    #endif using spritename = item->Attribute("name");
    float x = boost::lexical_cast<float>(item->Attribute("x"));
    float y = boost::lexical_cast<float>(item->Attribute("y"));
    float offset = boost::lexical_cast<float>(item->Attribute("offset"));
    unsigned layer = 50; // default
    if (item->Attribute("layer")) {
        layer = boost::lexical_cast<unsigned>(item->Attribute("layer"));
    }
    name = sp_name;
```

DynamoDB reads

Eventual Consistent Reads

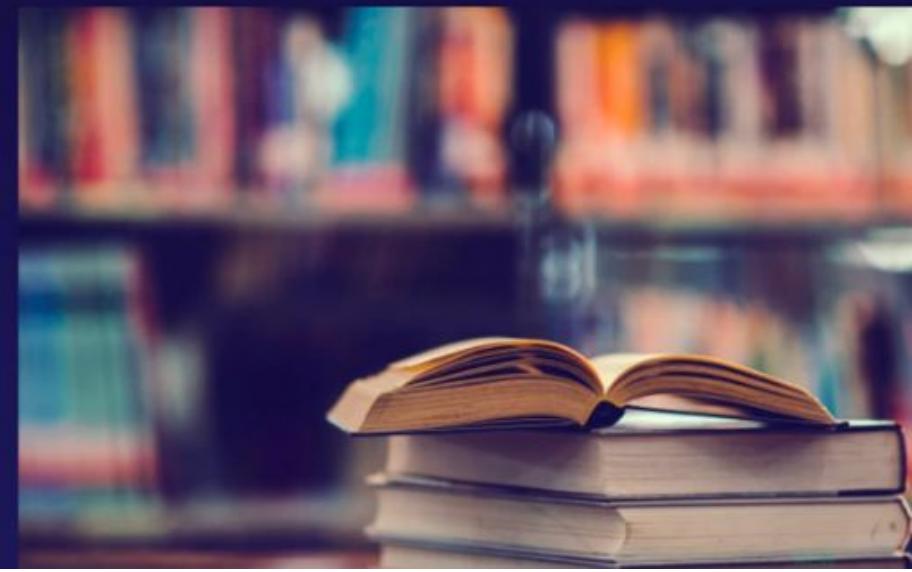
- Consistency across all copies of data is usually reached within a second. Repeating a read after a short time should return the updated data. (Best Read Performance)



DynamoDB reads

Strongly Consistent Reads

- A strongly consistent read returns a result that reflects all writes that received a successful response prior to the read.



DynamoDB Exam tips

The basics of DynamoDB are as follows;

- Stored on SSD storage
- Spread across 3 geographically distinct data centres
- Eventual Consistent Reads (Default)
- Strongly Consistent Reads

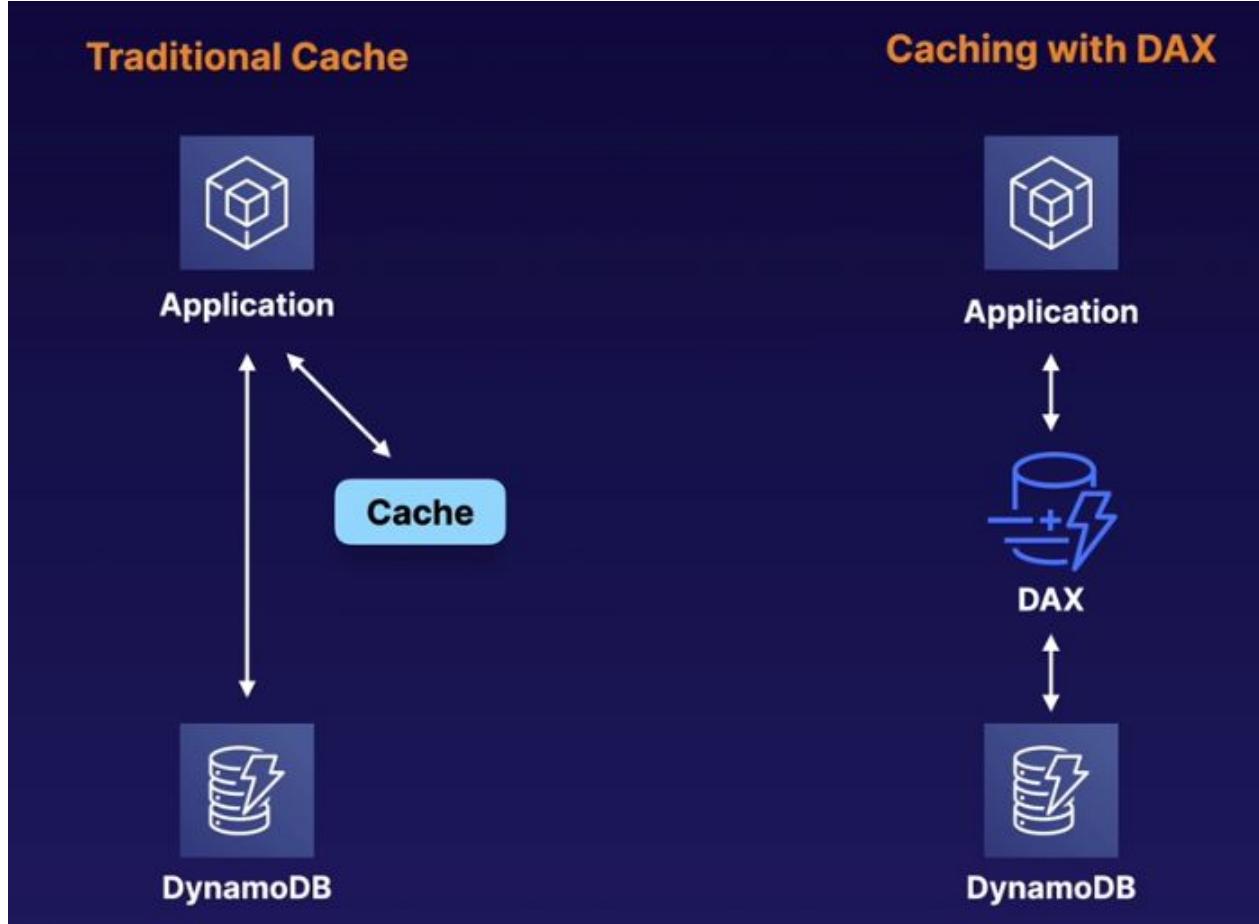
Advanced DynamoDB

DynamoDB Accelerator - DAX

- Fully managed, highly available, in-memory cache
- 10x performance improvement
- Reduces request time from milliseconds to **microseconds** — even under load
- No need for developers to manage caching logic
- Compatible with DynamoDB API calls



DynamoDB Accelerator - DAX



Transactions

- Multiple “all-or-nothing” operations
- Financial transactions
- Fulfilling orders
- Two underlying reads or writes — prepare/commit
- Up to 25 items or 4 MB of data



On-Demand capacity

- **Pay-per-request** pricing
- Balance cost and performance
- No minimum capacity
- No charge for read/write — only storage and backups
- **Pay more per request** than with provisioned capacity
- Use for new product launches



Backup and restore

On-Demand Backup and Restore

- Full backups at any time
- Zero impact on table performance or availability
- Consistent within seconds and **retained until deleted**
- Operates within same region as the source table



Backup and restore

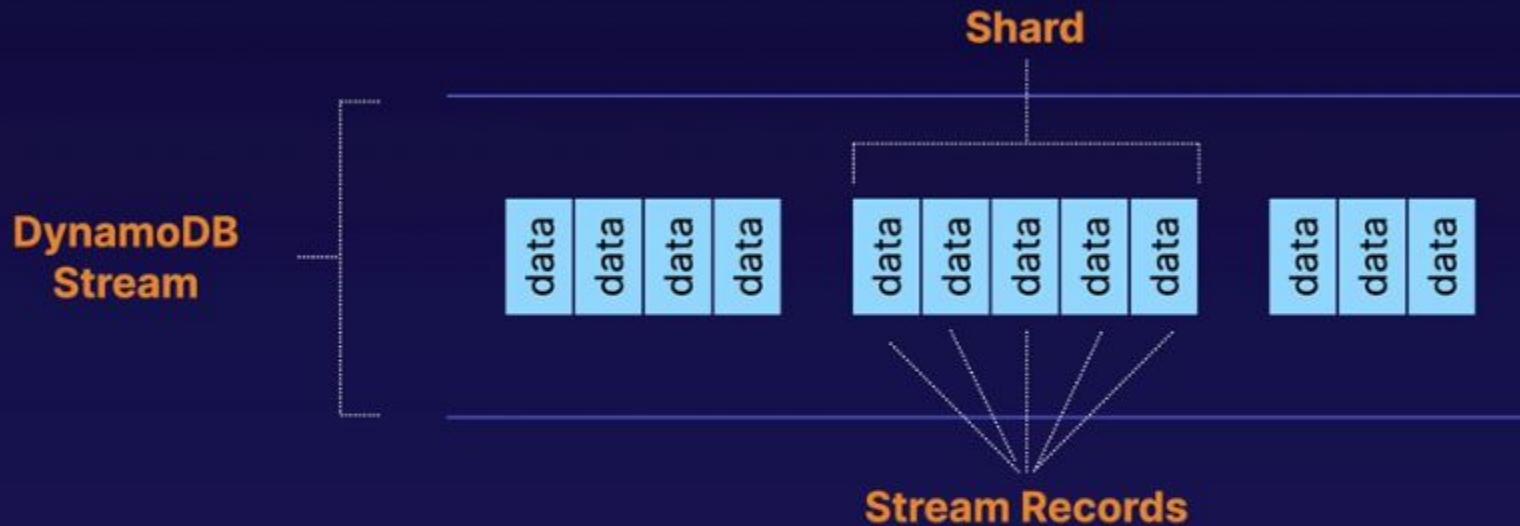
Point-in-Time Recovery (PITR)

- Protects against accidental writes or deletes
- Restore to any point in the last **35 days**
- Incremental backups
- Not enabled by default
- Latest restorable: **five minutes** in the past



Streams

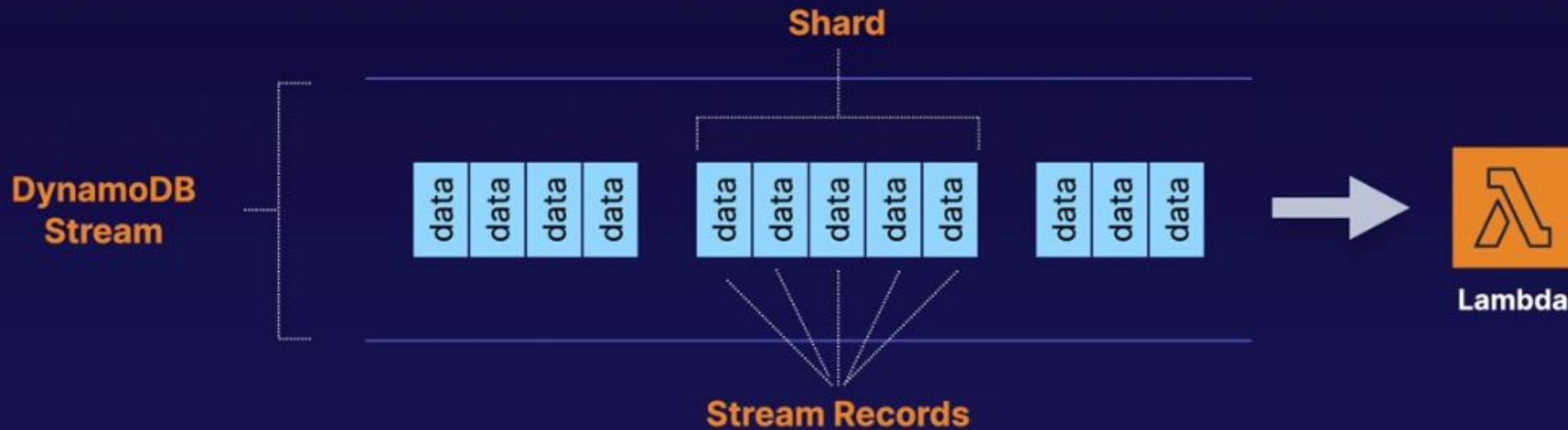
- Time-ordered sequence of item-level changes in a table



- Stored for **24 hours**
- Inserts, updates, and deletes

Streams

- Time-ordered sequence of item-level changes in a table



- Stored for **24 hours**
- Inserts, updates, and deletes
- Combine with Lambda functions for functionality like stored procedures

Streams

Managed Multi-Master, Multi-Region Replication

- Globally distributed applications
- Based on DynamoDB streams
- Multi-region redundancy for DR or HA
- No application rewrites
- Replication latency under **one second**



Streams

Screenshot of the AWS DynamoDB Streams interface for a global table named "ReplicateMe".

The top navigation bar shows "Services", "Resource Groups", "Ohio", and "Support". The main menu includes "Delete table", "Overview", "Items", "Metrics", "Alarms", "Capacity", "Indexes", "Global Tables", "Backups", and "More". A note states: "Global Tables enable you to use DynamoDB as a fully-managed, multi-region, multi-master database. Learn more".

A warning message in a yellow box says: "⚠ To create a global table, ensure that DynamoDB Streams are enabled. A table must meet the following requirements to become part of a global table." It lists requirements: "KMS Customer managed key: No", "Streams: Disabled", and "Stream type: -". A "Enable streams" button is present.

Below this, there's a modal window titled "Manage Stream" with the sub-section "View type" set to "New and old images - both the new and the old images of the item". It has "Cancel" and "Enable" buttons.

The main table area shows a single row for "ReplicateMe" with columns: "Name" (ReplicateMe), "Status" (Active), "Partition key" (not specified), and "Sort key" (not specified). A note at the bottom says: "You do not have a global table with this name. Click "Add region" to create one."

Screenshot of the AWS DynamoDB Tables interface.

The top navigation bar shows "Services", "Resource Groups", "N. California", and "Support". The main menu includes "Create table", "Delete table", "Filter by table name", "Choose a table group", "Actions", and "No Tables".

A note in a blue box says: "DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. DynamoDB allows you to create a database table that can store and retrieve any amount of data, and serve any level of request traffic. More info".

Streams

The screenshot shows the AWS DynamoDB Streams interface. A modal dialog titled "Add replica to global table" is open. It displays the following information:

- Current region: US East (Ohio)
- Global table version: 2019.11.21
- Region dropdown: US West (N. California)
- Status: Checking region
- Text: You are creating the 2019.11.21 version global table. If you need to create a version 2017.11.29 global table, please follow the steps outlined here: [Learn more](#)
- Buttons: Cancel and Create replica

Below the modal, the main table view shows a single row with the following details:

Name	Status	Partition key	Sort key

A message at the bottom states: "You do not have a global table with this name. Click 'Add region' to create one."

The screenshot shows the AWS DynamoDB Streams interface. The global table list table has the following columns and data:

Name	Status	Partition key	Sort key

Below the table, a descriptive text box states:

DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. DynamoDB allows you to create a database table that can store and retrieve any amount of data, and serve any level of request traffic. [More info](#)

Streams

aws

Services

Resource Groups



mrichman @ mark-richman

Ohio

Support

Rep

Ob

Glob

IAM

Glob

Cre

Ad

Add replica to global table

Current region: US East (Ohio)

Global table version: 2019.11.21

Replica in the intended region has been initiated successfully. It may take a few minutes for the replica to be created.

Go to table

Table name ReplicateMe
Region US West (N. California)
Primary partition key id (String)
Primary sort key -
Read/write capacity mode Provisioned
Provisioned read capacity units 5
Provisioned write capacity units 5
Auto Scaling READ_AND_WRITE
Stream enabled Yes
Encryption Type DEFAULT

[Close](#)

aws

Services

Resource Groups



mrichman @ mark-richman

N. California

Support

Create table

Delete table

Filter by table name

Choose a table group

Actions

No Tables

Name

Status

Partition key

Sort key

DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. DynamoDB allows you to create a database table that can store and retrieve any amount of data, and serve any level of request traffic. [More info](#)

Streams

Screenshot of the AWS CloudWatch Metrics interface showing the creation of a global table replica.

Add replica to global table

Current region: US East (Ohio)

Global table version: 2019.11.21

Replica status: Replica in the intended region has been initiated successfully. It may take a few minutes for the replica to be created.

Table name: ReplicateMe

Region: US West (N. California)

Primary partition key: id (String)

Primary sort key: -

Read/write capacity mode: Provisioned

Provisioned read capacity units: 5

Provisioned write capacity units: 5

Auto Scaling: READ_AND_WRITE

Stream enabled: Yes

Encryption Type: DEFAULT

[Go to table](#)

[Close](#)

View all CloudWatch metrics

Time Range: Last Hour

Global table metrics

Replication latency (Milliseconds)

1
0.801
0.601

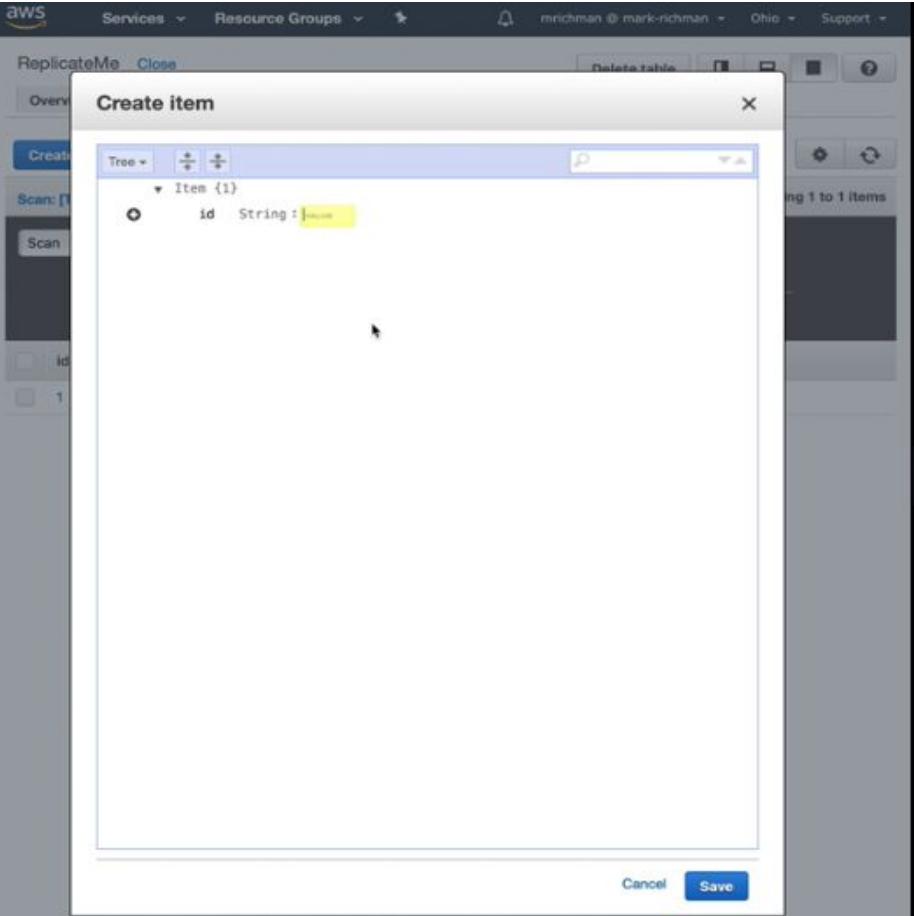
Screenshot of the AWS CloudWatch Metrics interface showing the creation of a global table replica.

Create table Delete table

Filter by table name Choose a table group Actions Viewing 1 of 1 Tables

Name	Status	Partition key	Sort key
ReplicateMe	Creating	id (String)	-

Streams



The screenshot shows the AWS DynamoDB 'ReplicateMe' table. The table has two items:

	id	message
	1	Hello

Streams

AWS Services Resource Groups Ohio Support

ReplicateMe Close Delete table

Overview Items Metrics Alarms Capacity Indexes Global Tables Backups More

Create item Actions

Scan [Table] ReplicateMe: id Add filter Start search

Viewing 1 to 2 items

	id	message
	1	Hello
	2	Cloud Gurus

AWS Services Resource Groups N. California Support

ReplicateMe Close Delete table

Overview Items Metrics Alarms Capacity Indexes Global Tables Backups More

Create item Actions

Scan [Table] ReplicateMe: id Add filter Start search

Viewing 1 to 2 items

	id	message
	1	Hello
	2	Cloud Gurus

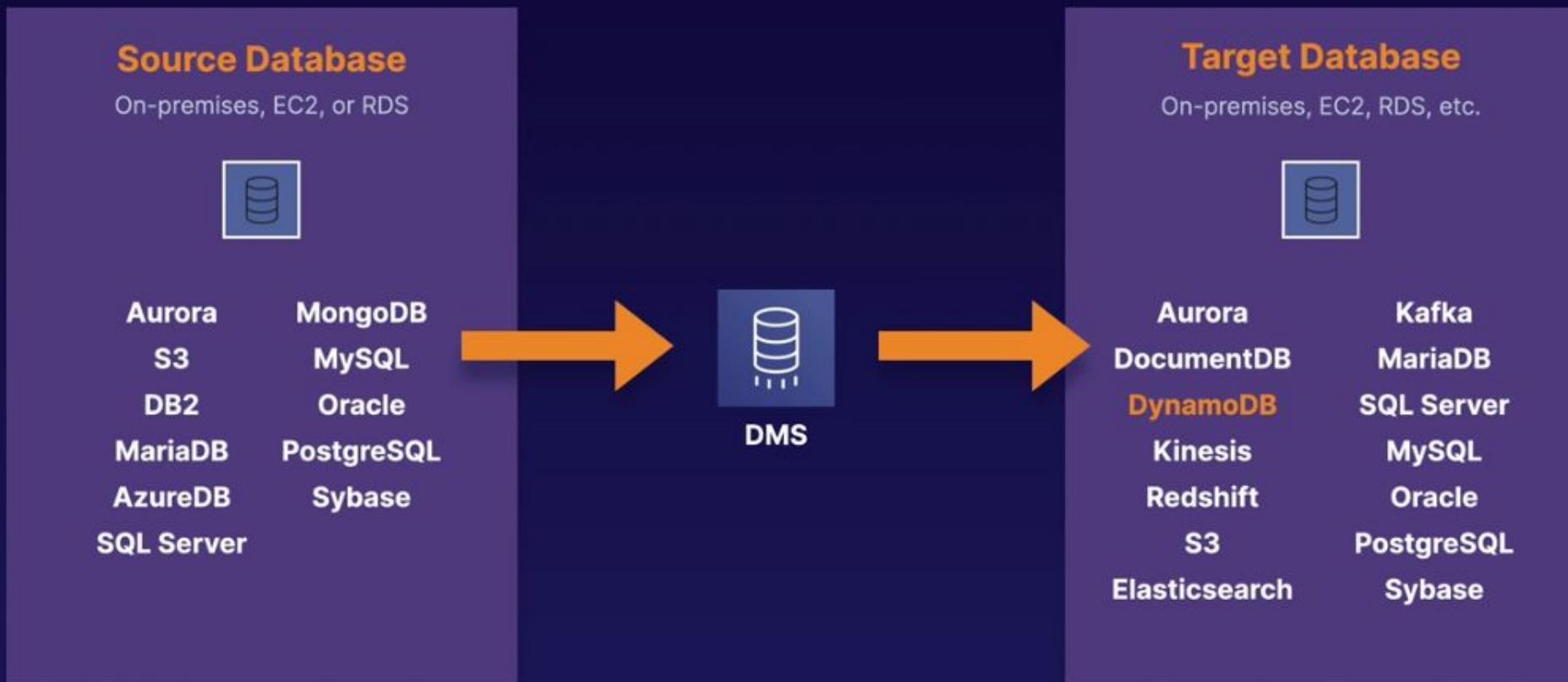
Global tables

Managed Multi-Master, Multi-Region Replication

- Globally distributed applications
- Based on DynamoDB streams
- Multi-region redundancy for DR or HA
- No application rewrites
- Replication latency under **one second**



Database Migration Service (DMS)



Source database remains operational

Security

- Encryption at rest using **KMS**
- Site-to-site VPN
- Direct Connect (DX)
- IAM policies and roles
- Fine-grained access
- CloudWatch and CloudTrail
- VPC endpoints



Redshift

Redshift

Amazon Redshift is a fast and powerful, fully managed, petabyte-scale data warehouse service in the cloud. Customers can start small for just \$0.25 per hour with no commitments or upfront costs and scale to a petabyte or more for \$1,000 per terabyte per year, less than a tenth of most other data warehousing solutions.



Redshift

OLAP transaction Example:

Net Profit for EMEA and Pacific for the Digital Radio Product.

Pulls in large numbers of records

Sum of Radios Sold in EMEA

Sum of Radios Sold in Pacific

Unit Cost of Radio in each region

Sales price of each radio

Sales price - unit cost.



Redshift

Data Warehousing databases use different type of architecture both from a database perspective and infrastructure layer.

**Amazon's Data
Warehouse
Solution Is Called
Redshift**



Redshift

Redshift can be configured as follows

- Single Node (160Gb)
- Multi-Node
 - Leader Node (manages client connections and receives queries.)
 - Compute Node (store data and perform queries and computations). Up to 128 Compute Nodes.



Redshift

Advanced Compression:

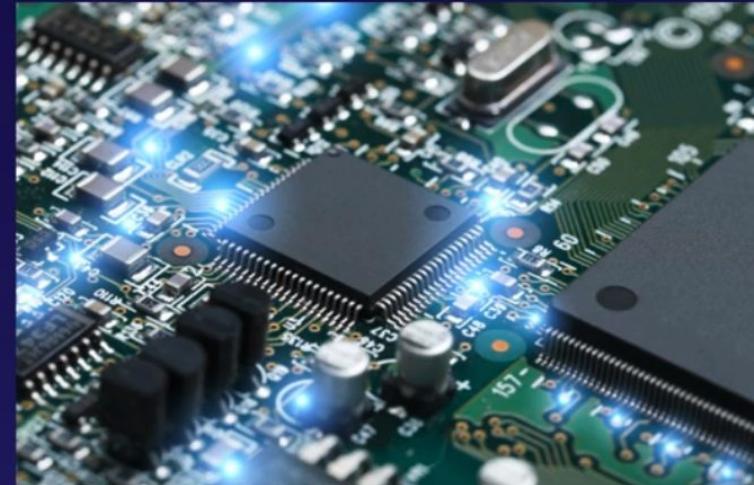
Columnar data stores can be compressed much more than row-based data stores because similar data is stored sequentially on disk. Amazon Redshift employs multiple compression techniques and can often achieve significant compression relative to traditional relational data stores. In addition, Amazon Redshift doesn't require indexes or materialized views, and so uses less space than traditional relational database systems. When loading data into an empty table, Amazon Redshift automatically samples your data and selects the most appropriate compression scheme.



Redshift

Massively Parallel Processing (MPP):

Amazon Redshift automatically distributes data and query load across all nodes. Amazon Redshift makes it easy to add nodes to your data warehouse and enables you to maintain fast query performance as your data warehouse grows.



Redshift

Backups

- Enabled by default with a 1 day retention period.
- Maximum retention period is 35 days.
- Redshift always attempts to maintain at least three copies of your data (the original and replica on the compute nodes and a backup in Amazon S3).
- Redshift can also asynchronously replicate your snapshots to S3 in another region for disaster recovery.



Redshift

Redshift is priced as follows;

- Compute Node Hours (total number of hours you run across all your compute nodes for the billing period. You are billed for 1 unit per node per hour, so a 3-node data warehouse cluster running persistently for an entire month would incur 2,160 instance hours. You will not be charged for leader node hours; only compute nodes will incur charges.)
- Backup
- Data transfer (only within a VPC, not outside it)



Redshift

Security Considerations:

- Encrypted in transit using SSL
- Encrypted at rest using AES-256 encryption
- By default RedShift takes care of key management.
 - Manage your own keys through HSM
 - AWS Key Management Service



Redshift

Redshift Availability:

- Currently only available in 1 AZ
- Can restore snapshots to new AZs in the event of an outage.



Redshift Exam tips

Backups

- Enabled by default with a 1 day retention period.
- Maximum retention period is 35 days.
- Redshift always attempts to maintain at least three copies of your data (the original and replica on the compute nodes and a backup in Amazon S3).
- Redshift can also asynchronously replicate your snapshots to S3 in another region for disaster recovery.

Aurora

Aurora

What is Aurora?

Amazon Aurora is a MySQL and PostgreSQL-compatible relational database engine that combines the speed and availability of high-end commercial databases with the simplicity and cost-effectiveness of open source databases.

Aurora

Amazon Aurora provides up to five times better performance than MySQL and three times better than PostgreSQL databases at a much lower price point, whilst delivering similar performance and availability.



Aurora

Things to know about Aurora

- 1 Start with 10GB, Scales in 10GB increments to 64TB (Storage Autoscaling)
- 2 Compute resources can scale up to 32vCPUs and 244GB of Memory.
- 3 2 copies of your data is contained in each availability zone, with minimum of 3 availability zones. 6 copies of your data.



Aurora

Scaling Aurora

- Aurora is designed to transparently handle the loss of up to two copies of data without affecting database write availability and up to three copies without affecting read availability.
- Aurora storage is also self-healing. Data blocks and disks are continuously scanned for errors and repaired automatically.



Aurora

Three Types of Aurora Replicas are available:

- Aurora Replicas (currently 15)
- MySQL Read Replicas (currently 5)
- PostgreSQL (currently 1)



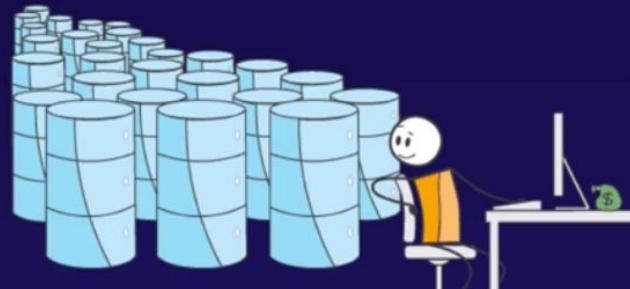
Aurora

Feature	Amazon Aurora Replicas	MySQL Replicas
Number of replicas	Up to 15	Up to 5
Replication type	Asynchronous (milliseconds)	Asynchronous (seconds)
Performance impact on primary	Low	High
Replica location	In-region	Cross-region
Act as failover target	Yes (no data loss)	Yes (potentially minutes of data loss)
Automated failover	Yes	No
Support for user-defined replication delay	No	Yes
Support for different data or schema vs. primary	No	Yes

Aurora

BACKUPS WITH AURORA

- Automated backups are always enabled on Amazon Aurora DB Instances. Backups do not impact database performance.
- You can also take snapshots with Aurora. This also does not impact on performance.
- You can share Aurora Snapshots with other AWS accounts.



Aurora

Amazon Aurora Serverless is an on-demand, autoscaling configuration for the MySQL-compatible and PostgreSQL-compatible editions of Amazon Aurora. An Aurora Serverless DB cluster automatically starts up, shuts down, and scales capacity up or down based on your application's needs.

Aurora

Aurora Serverless provides a relatively simple, cost-effective option for infrequent, intermittent, or unpredictable workloads.



Create an Aurora Read replica

RDS > Databases > Create aurora replica

Create Aurora read replica

Create an Amazon Aurora DB cluster that is a Read Replica of your source DB instance. The new DB cluster will have the same master username, master password, and database name as the source DB instance.

Instance specifications

DB engine
Aurora - compatible with MySQL 5.6.10a

DB instance class [Info](#)

Multi-AZ deployment
Specifies if the DB instance should have a standby deployed in another availability zone.

Create Replica in Different Zone
 No

Settings

DB instance identifier*
DB instance identifier. This is the unique key that identifies a DB instance. This parameter is stored as a lowercase string (e.g. mydbinstance).

Aurora Exam tips

- 2 copies of your data are contained in each availability zone, with minimum of 3 availability zones. 6 copies of your data.
- You can share Aurora Snapshots with other AWS accounts.
- 3 types of replicas available. Aurora Replicas, MySQL replicas & PostgreSQL replicas. Automated failover is only available with Aurora Replicas.
- Aurora has automated backups turned on by default. You can also take snapshots with Aurora. You can share these snapshots with other AWS accounts.
- Use Aurora Serverless if you want a simple, cost-effective option for infrequent, intermittent, or unpredictable workloads.

ElastiCache

Elasticache

ElastiCache is a web service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud. The service improves the performance of web applications by allowing you to retrieve information from fast, managed, in-memory caches, instead of relying entirely on slower disk-based databases.

Elasticache

ElastiCache supports two open-source in-memory caching engines:

- Memcached
- Redis



redis

Elasticache

Requirement	Memcached	Redis
Simple Cache to offload DB	Yes	Yes
Ability to scale horizontally	Yes	No
Multi-threaded performance	Yes	No
Advanced data types	No	Yes
Ranking/Sorting data sets	No	Yes
Pub/Sub capabilities	No	Yes
Persistence	No	Yes
Multi-AZ	No	Yes
Backup & Restore Capabilities	No	Yes

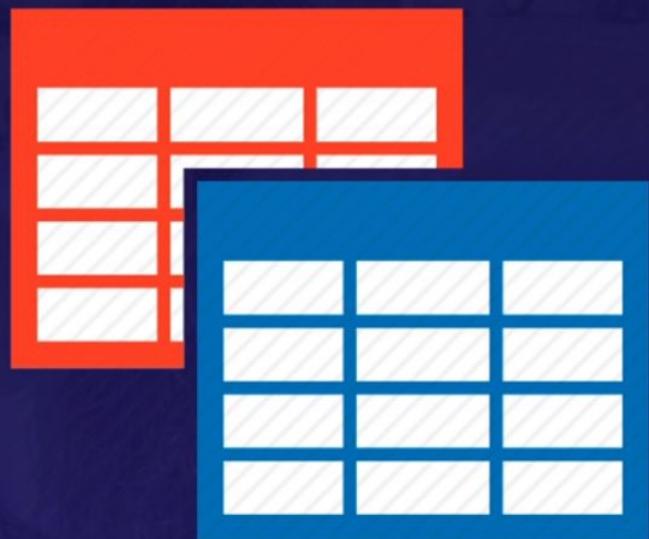
Elasticache Exam tips

- Use Elaticache to increase database and web application performance.
- Redis is Multi-AZ
- You can do back ups and restores of Redis

Database Migration Service

DMS?

AWS Database Migration Service (DMS) is a cloud service that makes it easy to migrate relational databases, data warehouses, NoSQL databases, and other types of data stores. You can use **AWS DMS** to migrate your data into the AWS Cloud, between on-premises instances (through an AWS Cloud setup), or between combinations of cloud and on-premises setups.



How does DMS work?

At its most basic level, **AWS DMS** is a server in the AWS Cloud that runs replication software. You create a source and target connection to tell AWS DMS where to extract from and load to. Then you schedule a task that runs on this server to move your data. AWS DMS creates the tables and associated primary keys if they don't exist on the target. You can **pre-create the target tables manually, or you can use AWS Schema Conversion Tool (SCT) to create some or all of the target tables, indexes, views, triggers, etc.**



Types of DMS migrations

Supports **homogenous** migrations:



And supports **heterogeneous** migrations:



Sources and targets

Sources

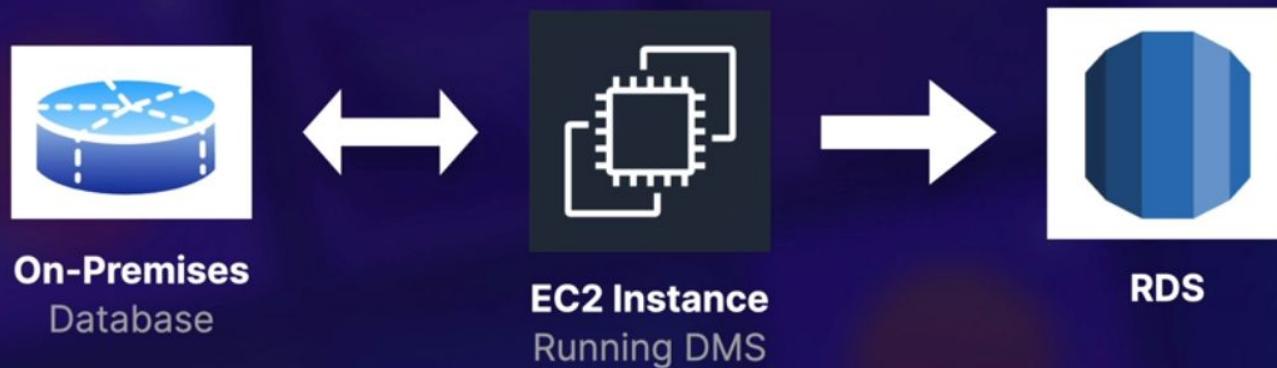
- ✓ On-premises and EC2 instances databases: Oracle, Microsoft SQL Server, MySQL, MariaDB, PostgreSQL, SAP, MongoDB, Db2
- ✓ Azure **SQL** Database
- ✓ Amazon **RDS** (including Aurora)
- ✓ Amazon S3

Targets

- ✓ On-premises and EC2 instances databases:
Oracle, Microsoft SQL Server, MySQL,
MariaDB, PostgreSQL, SAP
- ✓ RDS
- ✓ Redshift
- ✓ DynamoDB
- ✓ S3
- ✓ **Elasticsearch** service
- ✓ **Kinesis** Data Streams
- ✓ DocumentDB

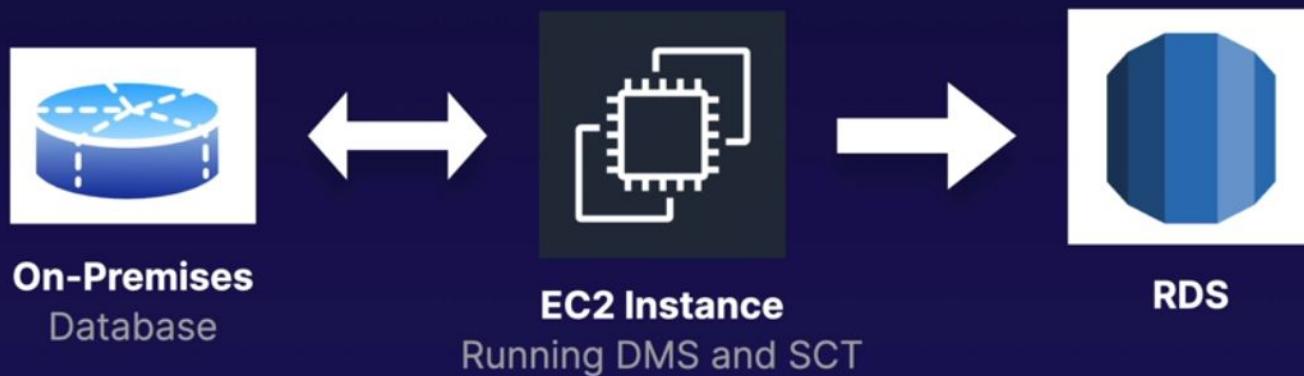
DMS in action

Solution Overview of DMS



SCT in Action

AWS Schema Conversion Tool



You do not need SCT if you are migrating to identical databases!

DMS Exam tips

Remember the following:



DMS allows you to **migrate databases** from one source to AWS.



The source can either be on-premises, or inside AWS itself or another cloud provider such as Azure.



You can do **homogenous** migrations (same DB engines) or **heterogeneous** migrations.



If you do a heterogeneous migration, you will need the **AWS Schema Conversion Tool** (SCT).

Caching Strategies on AWS

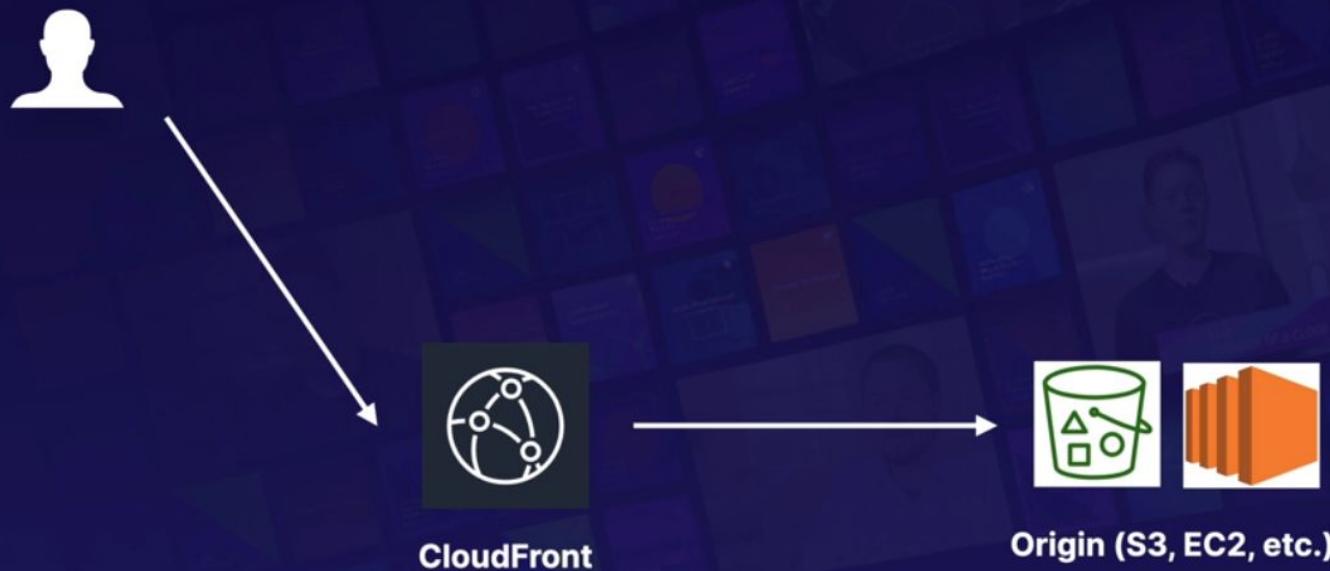
Caching services

The following services have caching capabilities:

- ✓ CloudFront
- ✓ API Gateway
- ✓ ElasticCache — **Memcached** and **Redis**
- ✓ DynamoDB Accelerator (DAX)

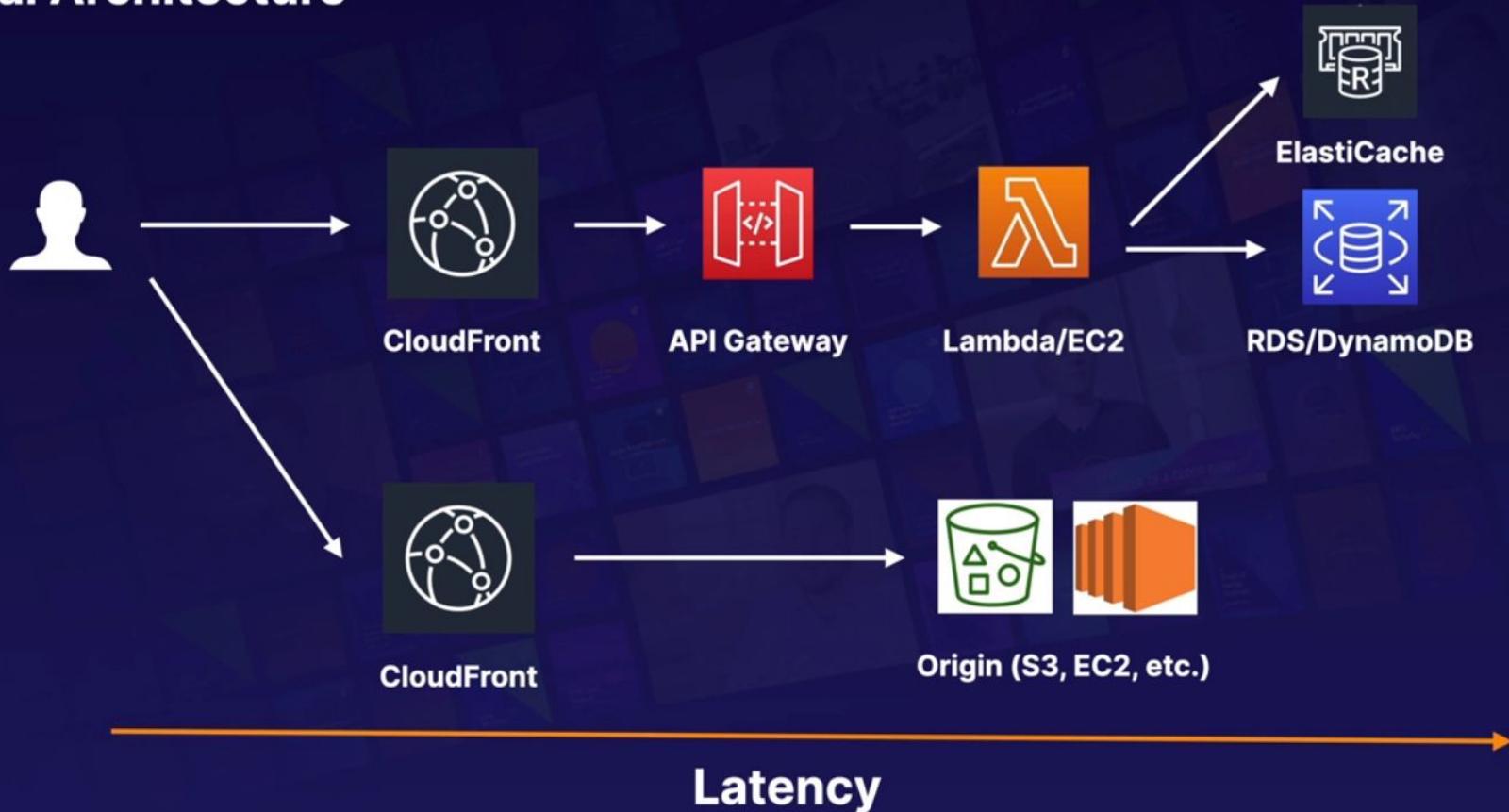
Caching services

Typical Architecture



Caching services

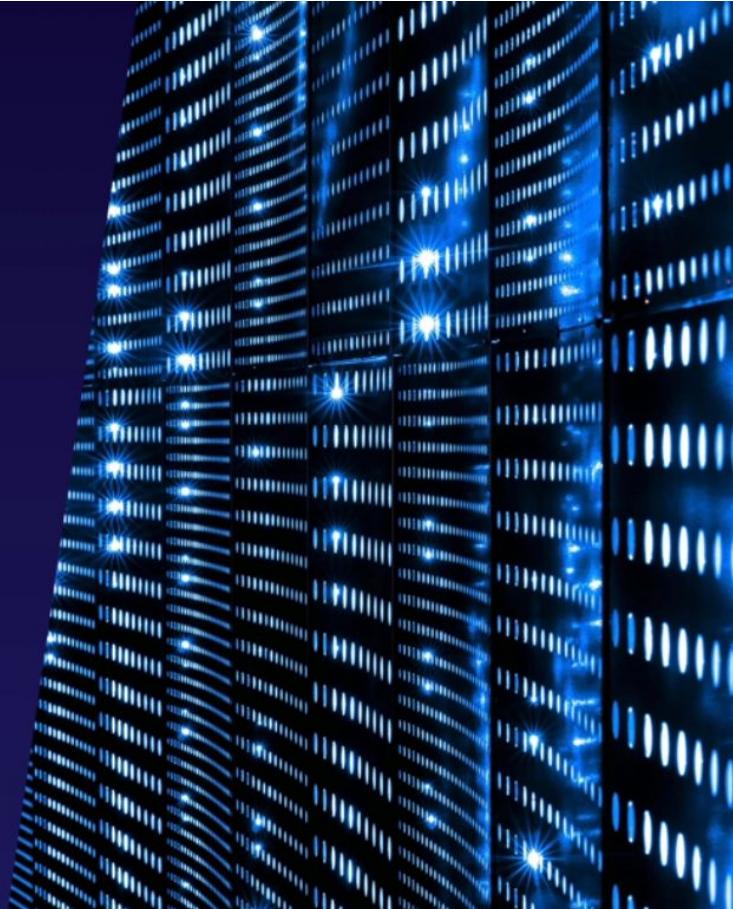
Typical Architecture



Caching Exam tips

Caching is a balancing act between **up-to-date, accurate information** and **latency**. We can use the following services to **cache on AWS**:

- 1 CloudFront
- 2 API Gateway
- 3 ElastiCache — **Memcached** and **Redis**
- 4 DynamoDB Accelerator (DAX)



EMR

EMR?

Amazon EMR is the industry-leading cloud big data platform for processing vast amounts of data using open-source tools such as Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi, and Presto. With EMR, you can run petabyte-scale analysis at **less than half the cost of traditional on-premises solutions** and over three times faster than standard Apache Spark.



EMR?

The central component of **Amazon EMR** is the cluster. A cluster is a collection of Amazon Elastic Compute Cloud (Amazon EC2) instances. Each instance in the cluster is called a node. Each node has a role within the cluster, referred to as the node type.

Amazon EMR also installs **different software components on each node type**, giving each node a role in a distributed application like Apache Hadoop.

EMR?

The node types in Amazon EMR are as follows:



Master node: A node that manages the cluster. The master node tracks the **status of tasks** and monitors the health of the cluster. Every cluster has a master node.



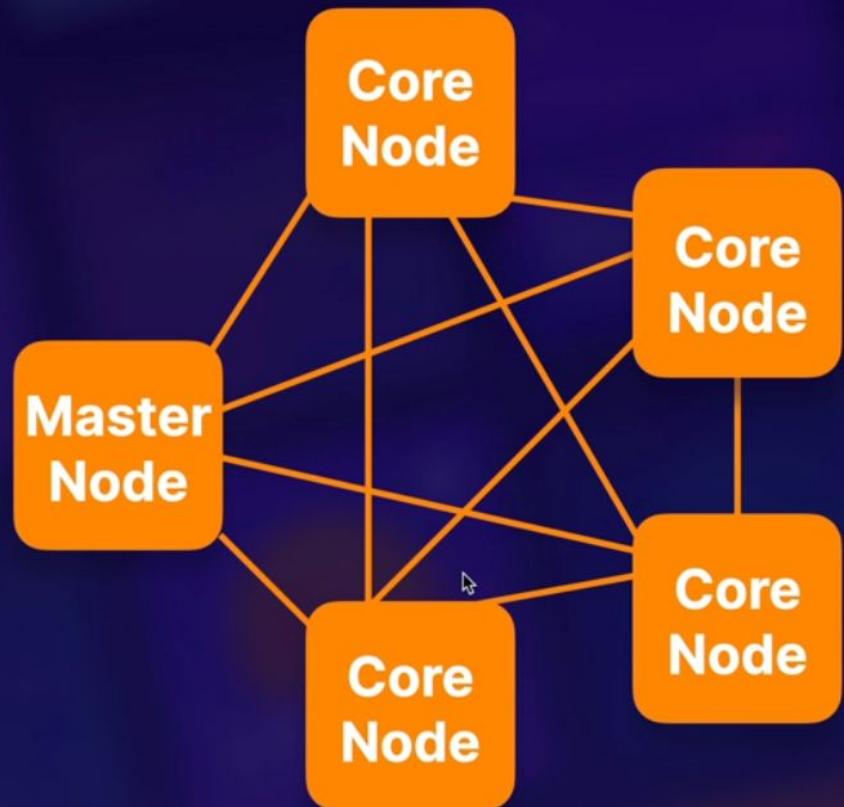
Core node: A node with software components that **runs tasks and stores data** in the Hadoop Distributed File System (HDFS) on your cluster. Multi-node clusters have at least one core node.



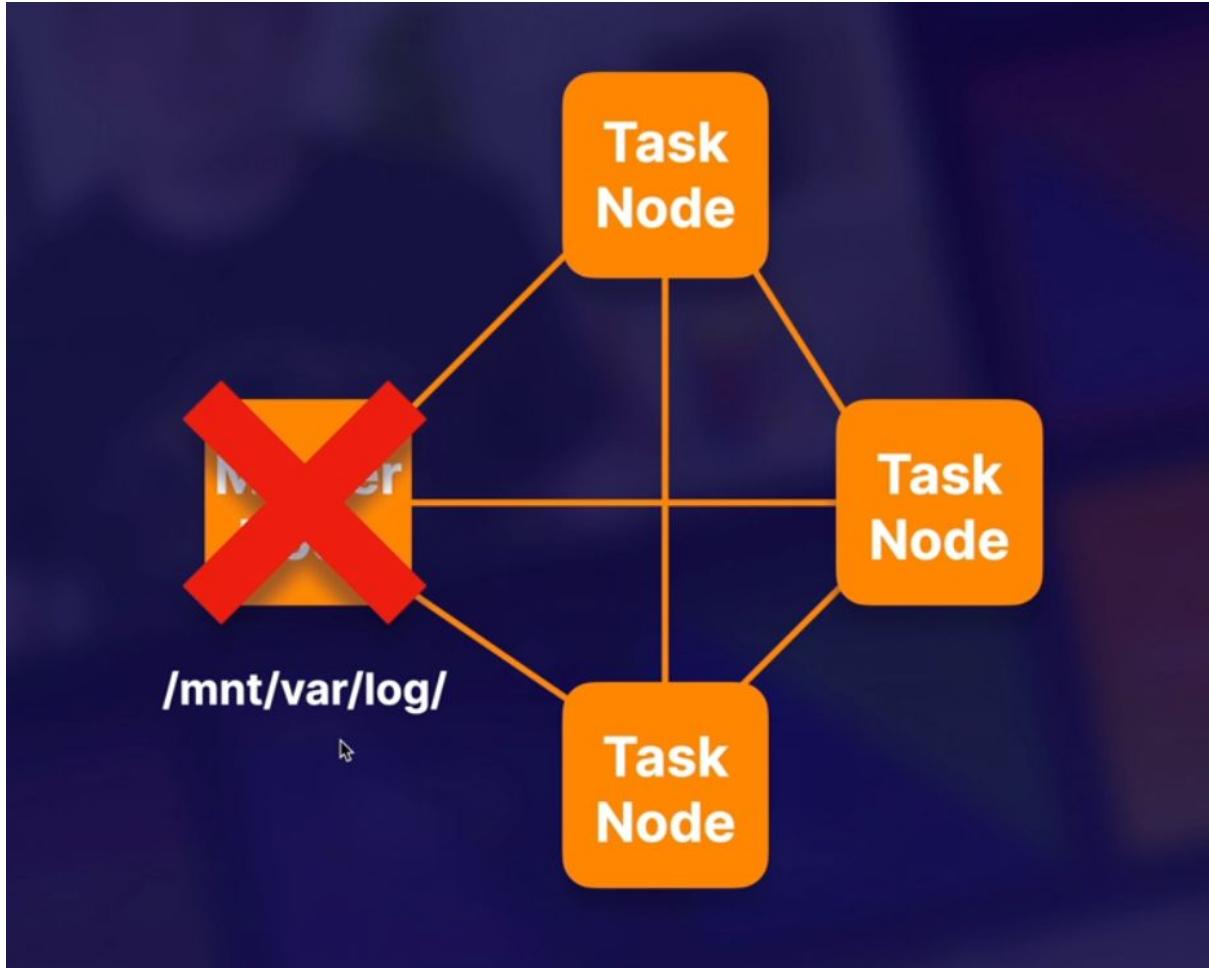
Task node: A node with software components that only runs tasks and **does not store data in HDFS**. Task nodes are **optional**.

EMR?

Amazon EMR Cluster



EMR?



EMR?



/mnt/var/log/

You can configure a cluster to **periodically archive the log files stored on the master node to Amazon S3**. This ensures the log files are available after the cluster terminates, whether this is through normal shutdown or due to an error. Amazon EMR archives the log files to Amazon S3 at **five-minute intervals**.

EMR Exam tips



EMR is used for
big data
processing.



Consists of a **master node**, a **core node**, and (optionally) a **task node**.



By default, log data is **stored on the master node**.



You can configure replication to S3 on **five-minute intervals for all log data from the master node**; however, this can only be configured when creating the cluster for the first time.

Databases Summary

Exam tips

Databases Exam tips

RDS (OLTP)

- SQL
 - MySQL
 - PostgreSQL
 - Oracle
 - Aurora
 - MariaDB

DynamoDB (No SQL)

Red Shift OLAP

Databases Exam tips

Elasticache

- Memcached
- Redis



Algo de algo

Leyendo una cosa es hacerlo también por las imágenes.
Es que las cosas de la vida, las personas... No las das ni te das
ni te das ni te das.

Algunas cosas cambian.

Algunas cosas no cambian. Algunas cosas se pierden, las personas
se pierden, las cosas se pierden.

Más o menos, sí.

Y más o menos, las cosas cambian en España.
Leyendo quedas un momento conmigo, al final, dice
«Sí». Has encontrado más tristes que felices en la
Universidad española.

«Sí» es que la Universidad es mucho más que las
imposiciones que nos lo dicen, como lo pregunta Mariano.

«Sí», se han pasado diez años, «el digo» me ha
lamentado un bonito de la universidad lo que más me
me ha hecho de vital sobre la cosa fría.

Muchas cosas han cambiado. Poco o mucho. Poco o mucho.
Leyendo una persona,



Databases Exam tips

Remember the following points;

- RDS runs on virtual machines
- You cannot log in to these operating systems however.
- Patching of the RDS Operating System and DB is Amazon's responsibility
- RDS is NOT Serverless
- Aurora Serverless IS Serverless

Databases Exam tips

There are two different types of Backups for RDS:

- **Automated Backups**
- **Database Snapshots**

Databases Exam tips

Read Replicas

- Can be Multi-AZ.
- Used to increase performance.
- Must have backups turned on.
- Can be in different regions.
- Can be MySQL, PostgreSQL, MariaDB, Oracle, Aurora.
- Can be promoted to master, this will break the Read Replica

Databases Exam tips

MultiAZ

- Used For DR.
- You can force a failover from one AZ to another by rebooting the RDS instance.

Databases Exam tips

Encryption at rest is supported for MySQL, Oracle, SQL Server, PostgreSQL, MariaDB & Aurora. Encryption is done using the AWS Key Management Service (KMS) service. Once your RDS instance is encrypted, the data stored at rest in the underlying storage is encrypted, as are its automated backups, read replicas, and snapshots.

Databases Exam tips

DynamoDB

- Stored on SSD storage
- Spread Across 3 geographically distinct data centres
- Eventual Consistent Reads (Default)
- Strongly Consistent Reads

Strongly Consistent Reads

Lleva una mayor latencia porque los cambios

deben ser sincronizados entre los tres

centros.

Eventual Consistency

No lleva una mayor latencia, pero la respuesta

es más rápida porque no se tienen que esperar

los cambios entre los tres centros.

Es útil para aplicaciones que necesitan

actualizaciones rápidas.

Strongly Consistent

Y eventualmente consistente (EBS) o eventualmente consistente (ECS).

Lleva una menor latencia porque no se tienen que esperar

los cambios entre los tres centros.

Es útil para aplicaciones que necesitan

actualizaciones rápidas.

Eventual Consistency (ECS)

Y eventualmente consistente (ECS).

Lleva una menor latencia porque no se tienen que esperar

los cambios entre los tres centros.

Es útil para aplicaciones que necesitan

Databases Exam tips

Redshift Backups

- Enabled by default with a 1 day retention period.
- Maximum retention period is 35 days.
- Redshift always attempts to maintain at least three copies of your data (the original and replica on the compute nodes and a backup in Amazon S3).
- Redshift can also asynchronously replicate your snapshots to S3 in another region for disaster recovery.

Databases Exam tips

Aurora

- 2 copies of your data are contained in each availability zone, with minimum of 3 availability zones. 6 copies of your data.
- You can share Aurora Snapshots with other AWS accounts.
- 3 types of replicas available. Aurora Replicas, MySQL replicas & PostgreSQL replicas. Automated failover is only available with Aurora Replicas.
- Aurora has automated backups turned on by default. You can also take snapshots with Aurora. You can share these snapshots with other AWS accounts.
- Use Aurora Serverless if you want a simple, cost-effective option for infrequent, intermittent, or unpredictable workloads.

Databases Exam tips

Elasticache

- Use Elasticache to increase database and web application performance.
- Redis is Multi-AZ
- You can do back ups and restores of Redis
- If you need to scale horizontally, use Memcached

Databases Quizz

Databases Quizz

QUESTION 1

Amazon's ElastiCache uses which two engines?

Redis & Memory

Redis & Memcached

MyISAM & InnoDB

Reddit & Memorush

Good work!

Databases Quizz

QUESTION 2

If you are using Amazon RDS Provisioned IOPS storage with a Microsoft SQL Server database engine, what is the maximum size RDS volume you can have by default?

1TB

6TB

500GB

32TB

16TB

Good work!

Databases Quizz

QUESTION 3

Which of the following AWS services is a non-relational database?

Redshift

DynamoDB

RDS

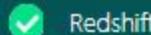
ElastiCache

Good work!

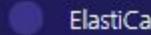
Databases Quizz

QUESTION 4

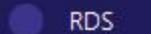
Which of the following is most suitable for OLAP?



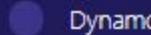
Redshift



ElastiCache



RDS



DynamoDB

Good work!

Redshift would be the most suitable for online analytics processing.

Databases Quizz

QUESTION 5

Which of the following data formats does Amazon Athena support?

Choose 3

Apache Parquet

XML

JSON

Apache ORC

Sorry!

Correct Answer

Amazon Athena is an interactive query service that makes it easy to analyse data in Amazon S3, using standard SQL commands. It will work with a number of data formats including "JSON", "Apache Parquet", "Apache ORC" amongst others, but "XML" is not a format that is supported.

Databases Quizz

QUESTION 6

If I wanted to run a database on an EC2 instance, which of the following storage options would Amazon recommend?



S3



EBS



RDS



Glacier

Good work!

Databases Quizz

QUESTION 7

When you add a rule to an RDS DB security group, you must specify a port number or protocol.



True



False

Sorry!

Correct Answer

Technically a destination port number is needed, however with a DB security group the RDS instance port number is automatically applied to the RDS DB Security Group.

Databases Quizz

QUESTION 8

If you want your application to check RDS for an error, have it look for an __ node in the response from the Amazon RDS API.

Incorrect

Abort

Exit

Error

Good work!

Databases Quizz

QUESTION 9

MySQL installations default to port number ____.

80

3389

1433

3306

Good work!

Databases Quizz

QUESTION 10

Which AWS service is ideal for Business Intelligence Tools/Data Warehousing?

- Elastic Beanstalk
- ElastiCache
- DynamoDB
- Redshift

Good work!

Databases Quizz

QUESTION 11

Which of the following DynamoDB features are chargeable, when using a single region?

Choose 2



Read and Write Capacity



The number of tables created



Storage of Data



Incoming Data Transfer

Sorry!

Correct Answer

There will always be a charge for provisioning read and write capacity and the storage of data within DynamoDB, therefore these two answers are correct. There is no charge for the transfer of data into DynamoDB, providing you stay within a single region (if you cross regions, you will be charged at both ends of the transfer.) There is no charge for the actual number of tables you can create in DynamoDB, providing the RCU and WCU are set to 0, however in practice you cannot set this to anything less than 1 so there always be a nominal fee associated with each table.

Databases Quizz

How many copies of my data does RDS - Aurora store by default?

3

2

1

6

Good work!

Databases Quizz

In RDS, what is the maximum value I can set for my backup retention period?

45 Days

15 Days

30 Days

35 Days

Good work!

Databases Quizz

Which set of RDS database engines is currently available?

PostgreSQL, MariaDB, MongoDB, Aurora

Oracle, SQL Server, MySQL, PostgreSQL

MariaDB, SQL Server, MySQL, Cassandra

Aurora, MySQL, SQL Server, Cassandra

Good work!

Databases Quizz

What data transfer charge is incurred when replicating data from your primary RDS instance to your secondary RDS instance?

The charge is the same as the standard data transfer charge.

There is no charge associated with this action.

The charge is half of the standard data transfer charge.

The charge is double the standard data transfer charge.

Sorry!

Databases Quizz

AWS's NoSQL product offering is known as ____.

MySQL

MongoDB

DynamoDB

RDS

Good work!

Databases Quizz

What happens to the I/O operations of a single-AZ RDS instance during a database snapshot or backup?

Nothing.

I/O operations to the database are sent to a Secondary instance of a Multi-AZ installation (for the duration of the snapshot.)

I/O operations will function normally.

I/O may be briefly suspended while the backup process initializes (typically under a few seconds), and you may experience a brief period of elevated latency.

Good work!

I/O may be briefly suspended while the backup process initializes (typically under a few seconds), and you may experience a brief period of elevated latency.

Databases Quizz

You are hosting a MySQL database on the root volume of an EC2 instance. The database is using a large number of IOPS, and you need to increase the number of IOPS available to it. What should you do?

- Use CloudFront to cache the database.
- Migrate the database to an S3 bucket.
- Add 2 additional EBS SSD volumes and create a RAID 0 volume to host the database.
- Migrate the database to Glacier.

Good work!

RAID 0 provides performance improvements compared with a single volume as data can be read and written to multiple disks simultaneously. 2 disks, each with a bandwidth of 4,000 IOPS will provide a combined bandwidth of 8,000 IOPS.

Databases Quizz

RDS Reserved instances are available for multi-AZ deployments.



True



False

Good work!

Databases Quizz

Which of the following is not a feature of DynamoDB?

Single availability zone by default

Data reads that are either eventually consistent or strongly consistent

The ability to perform operations by using a user-defined primary key

The primary key can either be a single-attribute or a composite

Sorry!

Correct Answer

DynamoDB is the AWS managed NoSQL database service. It has many features that are being added to constantly, making it a great service to use for many different requirements. The feature which was incorrect is DynamoDB only being single availability zone by default making this the correct answer. DynamoDB is distributed across three geographically distinct datacentres by default, all of the other options listed are valid features of DynamoDB.

Databases Quizz

When creating an RDS instance, you can select the Availability Zone into which you deploy it.

False

True

Good work!

Databases Quizz

Under what circumstances would I choose provisioned IOPS over standard storage when creating an RDS instance?



If you use online transaction processing in your production environment.

If you have workloads that are not sensitive to latency/lag.

If this was a test Database.

If your business was trying to save money.

Good work!

Provisioned IOPS becomes important when you are running production environments requiring rapid responses, such as those which run e-commerce websites. Without high performant responses from an RDS instance page loads of the website could suffer resulting in loss of business. If your workloads are not latency sensitive or you are running a test environment the additional cost of provisioned IOPS will not be cost beneficial to your project.

Databases Quizz

In RDS, changes to the backup window take effect ____.

Immediately

You cannot back up in RDS.

After 30 minutes

The next day

Good work!

Databases Quizz

With new RDS DB instances, automated backups are enabled by default?

False

True

Good work!

CHAPTER 6

Advanced IAM

AWS Directory Service

- Family of managed services
- Connect AWS resources with on-premises AD
- Standalone directory in the cloud
- Use existing corporate credentials
- SSO to any domain-joined EC2 instance



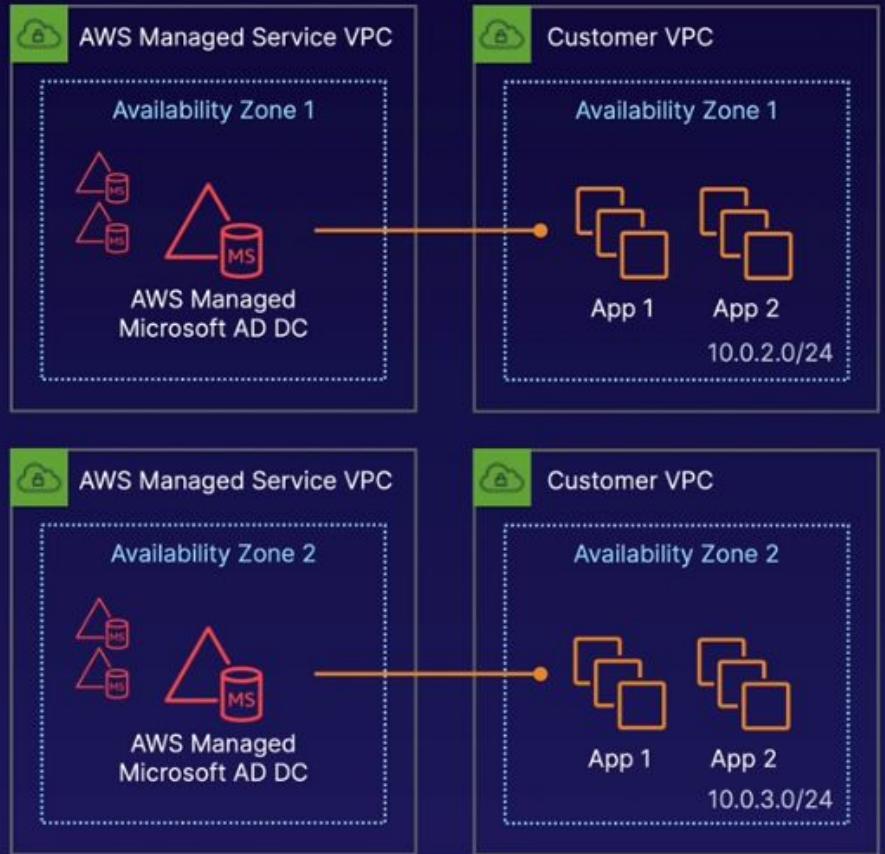
AWS Directory Service

- On-premises directory service
- Hierarchical database of users, groups, computers — **trees** and **forests**
- Group policies
- LDAP and DNS
- Kerberos, LDAP, and NTLM authentication
- Highly available



What is Active Directory?

- AD domain controllers (DCs) running Windows Server
- Reachable by applications in your VPC
- Add DCs for HA and performance
- Exclusive access to DCs
- Extend existing AD to on-premises using **AD Trust**



AWS Managed Microsoft AD

AWS

- Multi-AZ deployment
- Patch, monitor, recover
- Instance rotation
- Snapshot and restore

Customer

- Users, groups, GPOs
- Standard AD tools
- Scale out DCs
- Trusts (resource forest)
- Certificate authorities (LDAPS)
- Federation

AD Connector

- Standalone managed directory
- Basic AD features
- Small: <= 500; Large <= 5,000 users
- Easier to manage EC2
- Linux workloads that need LDAP
- Does not support **trusts** (can't join on-premises AD)



AD Connector

- Directory gateway (proxy) for on-premises AD
- Avoid caching information in the cloud
- Allow on-premises users to log in to AWS using AD
- Join EC2 instances to your existing AD domain
- Scale across multiple AD Connectors



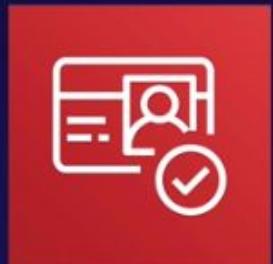
Cloud Directory

- Directory-based store for **developers**
- Multiple hierarchies with hundreds of millions of objects
- Use cases: org charts, course catalogs, device registries
- Fully managed service



Amazon Cognito User Pools

- Managed user directory for SaaS applications
- Sign-up and sign-in for web or mobile
- Works with **social media** identities



AD vs. Non-AD Compatible Services



AD Compatible

- Managed Microsoft AD
(a.k.a., Directory Service for Microsoft Active Directory)
- AD Connector
- Simple AD



Not AD Compatible

- Cloud Directory
- Cognito user pools

IAM Policies: Amazon Resource Name (ARN)

ARNs all begin with:

arn:partition:service:region:account_id:



And end with:

resource
resource_type/resource
resource_type/resource/qualifier
resource_type/resource:qualifier
resource_type:resource
resource_type:resource:qualifier

Examples:

arn:aws:iam::123456789012:user/mark
arn:aws:s3:::my-awesome_bucket/image.png
arn:aws:dynamodb:us-east-1:123456789012:table/orders
arn:aws:ec2:us-east-1:123456789012:instance/*

IAM Policies

- JSON document that defines permissions
- Identity policy
- Resource policy
- No effect until attached
- List of statements



IAM Policies

```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      1      ...  
    },  
    {  
      2      ...  
    },  
    {  
      3      ...  
    }  
  ]  
}
```

A policy document is
a list of statements.

Each statement matches
an AWS API request.

IAM Policies

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Sid": "SpecificTable",  
            "Effect": "Allow",  
            "Action": [  
                "dynamodb:BatchGet*",  
                "dynamodb:DescribeStream",  
                "dynamodb:DescribeTable",  
                "dynamodb:Get*",  
                "dynamodb:Query",  
                "dynamodb:Scan",  
                "dynamodb:BatchWrite*",  
                "dynamodb CreateTable",  
                "dynamodb>Delete*",  
                "dynamodb:Update*",  
                "dynamodb:PutItem"  
            ],  
            "Resource": "arn:aws:dynamodb:*:*:table/MyTable"  
        }  
    ]  
}
```

**Effect is either
Allow or Deny**

**Matched based
on their Action**

**And the Resource the
Action is against**

IAM Policies

IAM POLICIES

Evaluating IAM Policies



Exam Tips

- Not explicitly allowed == **implicitly denied**
- Explicit deny > everything else
- Only attached policies have effect
- AWS **joins** all applicable policies
- AWS-managed vs. customer-managed



IAM Policies

IAM POLICIES

Permission Boundaries

- Used to **delegate** administration to other users
- Prevent **privilege escalation** or **unnecessarily broad permissions**
- Control **maximum** permissions an IAM policy can grant
- Use cases:
 - Developers creating roles for Lambda functions
 - Application owners creating roles for EC2 instances
 - Admins creating ad hoc users



AWS Resource Access Manager (RAM)



AWS Resource Access Manager (RAM)
allows **resource sharing** between accounts.



Individual Accounts



AWS Organization

AWS Resource Access Manager (RAM)

Which AWS resources can I share using RAM?



App Mesh



Aurora



CodeBuild



EC2



EC2 Image
Builder



License
Manager



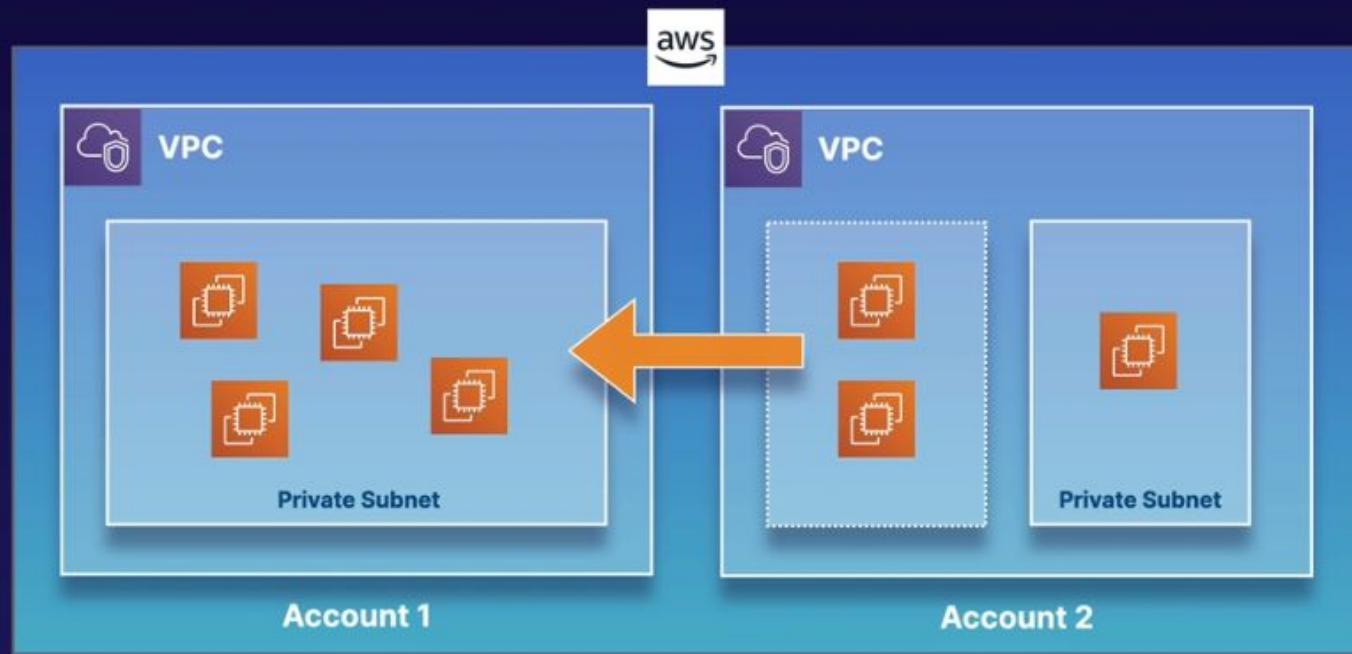
Resource
Groups



Route 53

AWS Resource Access Manager (RAM)

Example: Launch EC2 instances in a shared subnet.



AWS Single Sign-On



Single sign-on (SSO) service helps **centrally manage** access to AWS accounts and business applications.



Centrally manage
accounts

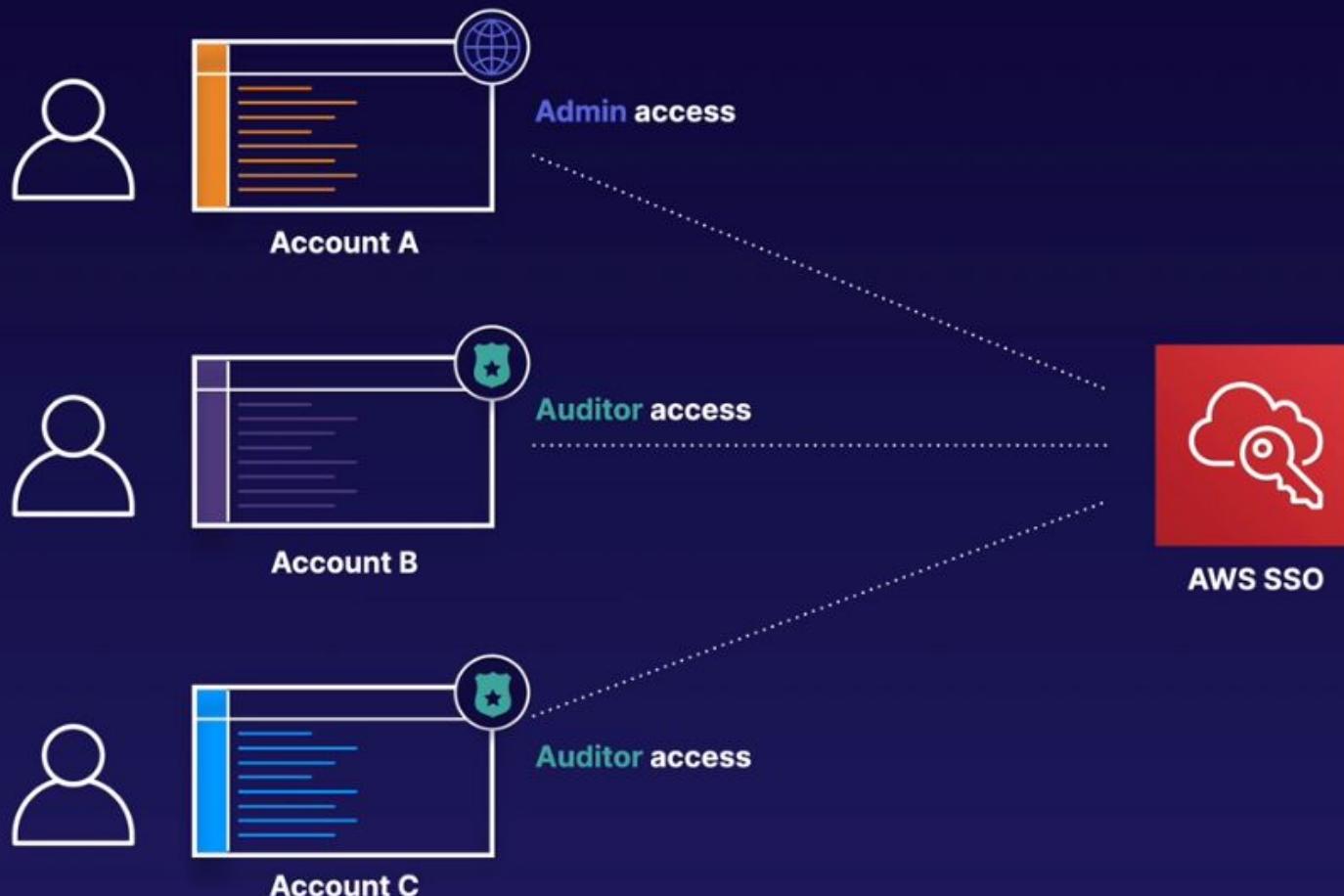


Use existing
corporate identities

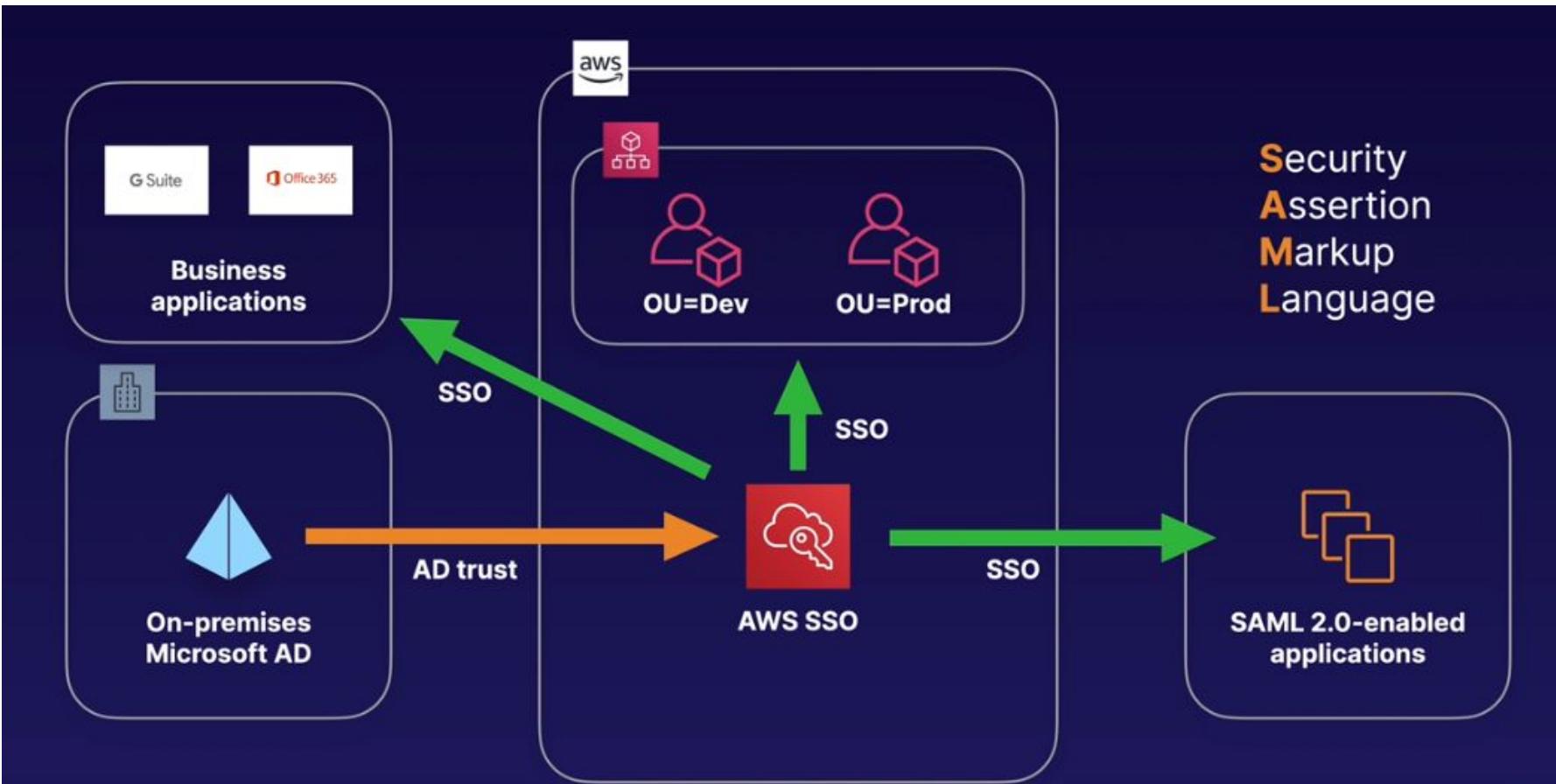


SSO access to
business applications

AWS SSO: Granular Account-Level Permissions



AWS Single Sign-On



CHAPTER 7

Route 53

DNS 101

If you've used the internet, you've used DNS. DNS is used to convert human friendly domain names (such as `http://acloud.guru`) into an Internet Protocol (IP) address (such as `http://82.124.53.1`).

IP addresses are used by computers to identify each other on the network. IP addresses commonly come in 2 different forms, IPv4 and IPv6.



NamesandNumbers.com	
<i>Wood River Valley</i>	
622-7481	BATES Paul 118 Willow Rd..... Hailey 788-1206
788-3933	BATES Steve 105 Audubon Pl..... Hailey 788-6222
788-9263	BATES VICKY - INTERIOR MOTIVES PO Box 1820..... Sun Valley 788-5950
788-9933	BATHUM Roy 235 Spur Ln..... Ketchum 726-0722
578-0595	BATMAN..... See West Adam
788-8979	BATT Jeffrey & Camille..... 726-7494
788-2515	BATTERSBY Patricia 116 Ritchie Dr..... Ketchum 726-8896
20-5661	BAUER Charlotte 621 Northstar Dr..... Hailey 788-4279
28-7219	BAUER CHARLOTTE LINDBERG.....
38-2317	Radiance Skin Care Studio.....
	BAUER Matt 3340 Woodside Blvd..... Hailey 578-2214
	BAUER Rich..... Hailey 578-0703
	BAUER Rich..... 720-0165

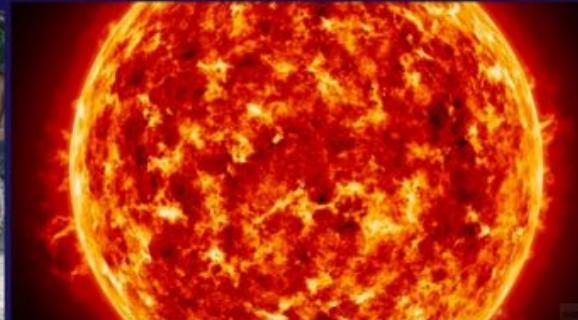
DNS 101 :: IPV4 vs IPv6

IPv4 Addresses are running out...

The IPv4 space is a 32 bit field and has over 4 billion different addresses (4,294,967,296 to be precise).

IPv6 was created to solve this depletion issue and has an address space of 128bits which in theory is

340,282,366,920,938,463,463,374,607,431,768,211,456 addresses or 340 undecillion addresses.



DNS 101 :: Top Level Domains

If we look at common domain names such as google.com, bbc.co.uk, acloud.guru etc., you will notice a string of characters separated by dots (periods). The last word in a domain name represents the “top level domain”. The second word in a domain name is known as a second level domain name (this is optional though and depends on the domain name).

- .com
- .edu
- .gov
- .co.uk
- .gov.uk
- .com.au



DNS 101 :: Top Level Domains

These top level domain names are controlled by the Internet Assigned Numbers Authority (IANA) in a root zone database which is essentially a database of all available top level domains. You can view this database by visiting:

<http://www.iana.org/domains/root/db>

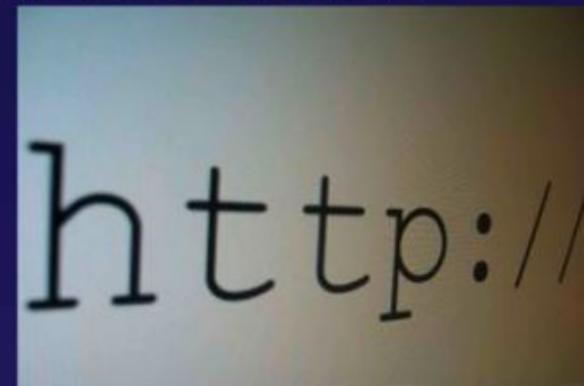


DNS 101 :: Domain Registrars

Because all of the names in a given domain name have to be unique there needs to be a way to organize this all so that domain names aren't duplicated. This is where domain registrars come in. A registrar is an authority that can assign domain names directly under one or more top-level domains. These domains are registered with InterNIC, a service of ICANN, which enforces uniqueness of domain names across the Internet. Each domain name becomes registered in a central database known as the WhoIS database.

Popular domain registrars include

Amazon, GoDaddy.com, 123-reg.co.uk etc.



DNS 101 :: Start of Authority Record (SOA)

The SOA record stores information about:

- The name of the server that supplied the data for the zone.
- The administrator of the zone.
- The current version of the data file.
- The default number of seconds for the time-to-live file on resource records.



DNS 101 :: NS Records

NS stands for Name Server Records

They are used by Top Level Domain servers to direct traffic to the Content DNS server which contains the authoritative DNS records.



DNS 101 :: TTL

What's an TTL?

The length that a DNS record is cached on either the Resolving Server or the users own local PC is equal to the value of the "Time To Live" (TTL) in seconds. The lower the time to live, the faster changes to DNS records take to propagate throughout the internet.



DNS 101 :: CName

What's a CName?

A Canonical Name (CName) can be used to resolve one domain name to another. For example, you may have a mobile website with the domain name `http://m.acloud.guru` that is used for when users browse to your domain name on their mobile devices. You may also want the name `http://mobile.acloud.guru` to resolve to this same address.

Create Record Set

Name: m.certifiedcloudpractitioner

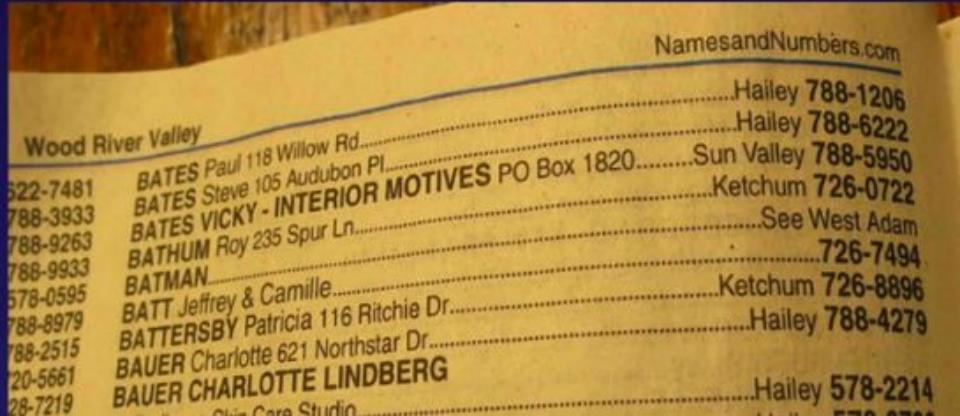
Type: CNAME – Canonical name

Alias: Yes No

TTL (Seconds): 300 1m 5m 1h 1d

Value: mobile

The domain name that you want to resolve to instead of the value in the Name field.
Example:

NamesandNumbers.com	
	
Wood River Valley	BATES Paul 118 Willow Rd..... Hailey 788-1206
522-7481	BATES Steve 105 Audubon Pl..... Hailey 788-6222
788-3933	BATES VICKY - INTERIOR MOTIVES PO Box 1820..... Sun Valley 788-5950
788-9263	BATHUM Roy 235 Spur Ln..... Ketchum 726-0722
788-9933	BATMAN..... See West Adam
578-0595	BATT Jeffrey & Camille..... 726-7494
788-8979	BATTERSBY Patricia 116 Ritchie Dr..... Ketchum 726-8896
788-2515	BAUER Charlotte 621 Northstar Dr..... Hailey 788-4279
720-5661	BAUER CHARLOTTE LINDBERG..... Hailey 578-2214
28-7219	Chin Care Studio.....

DNS 101 :: Alias Records

Alias Records

Alias records are used to map resource record sets in your hosted zone to Elastic Load Balancers, CloudFront distributions, or S3 buckets that are configured as websites.

Alias records work like a CNAME record in that you can map one DNS name (www.example.com) to another 'target' DNS name (elb1234.elb.amazonaws.com).



DNS 101 :: Alias Records

Alias Records

Key difference - A CNAME can't be used for naked domain names (zone apex record.) You can't have a CNAME for `http://acloud.guru`, it must be either an A record or an Alias.



DNS 101 :: Route53

Route53 Exam Tips

- ELBs do not have pre-defined IPv4 addresses; you resolve to them using a DNS name.
- Understand the difference between an Alias Record and a CNAME.
- Given the choice, always choose an Alias Record over a CNAME.

DNS 101 :: DNS Types

Common DNS Types

- SOA Records
- NS Records
- A Records
- CNAMEs
- MX Records
- PTR Records

Routing policies

The Following Routing Policies Are Available With Route53:

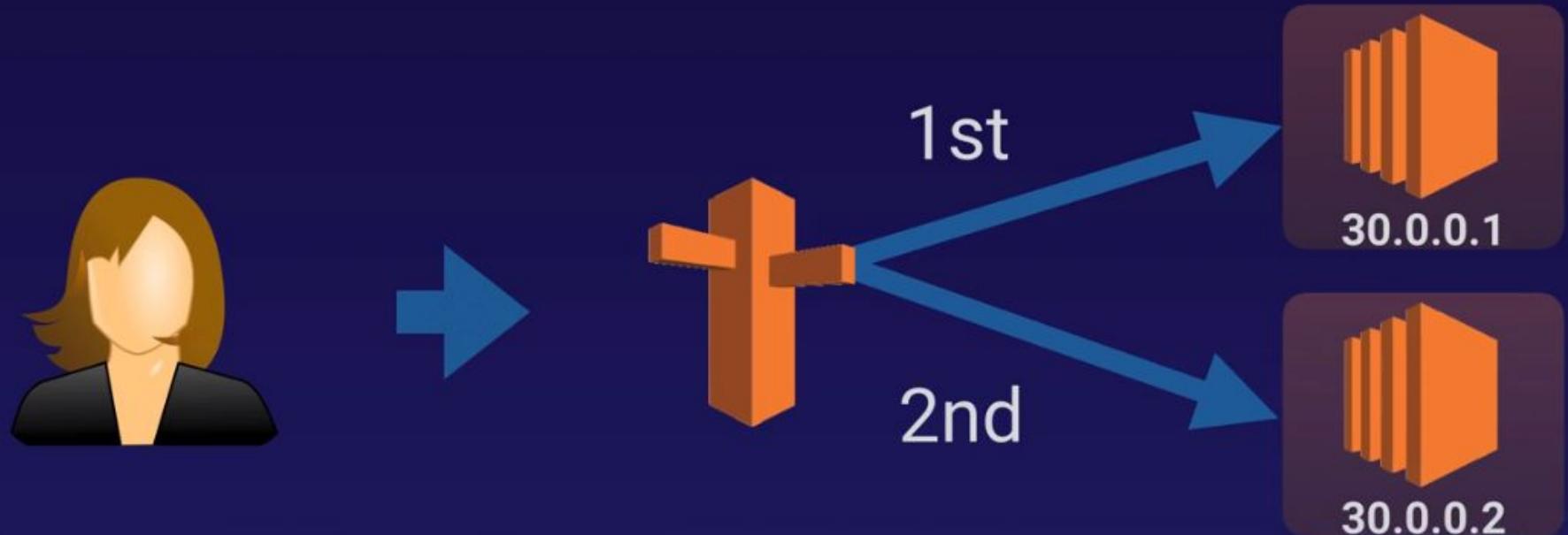
- Simple Routing
- Weighted Routing
- Latency-based Routing
- Failover Routing
- Geolocation Routing
- Geoproximity Routing (Traffic Flow Only)
- Multivalue Answer Routing



Routing policies

Simple Routing Policy

If you choose the simple routing policy you can only have one record with multiple IP addresses. If you specify multiple values in a record, Route 53 returns all values to the user in a random order.



Routing policies

Simple Routing Policy

If you choose the simple routing policy you can only have one record with multiple IP addresses. If you specify multiple values in a record, Route 53 returns all values to the user in a random order.

Routing policies

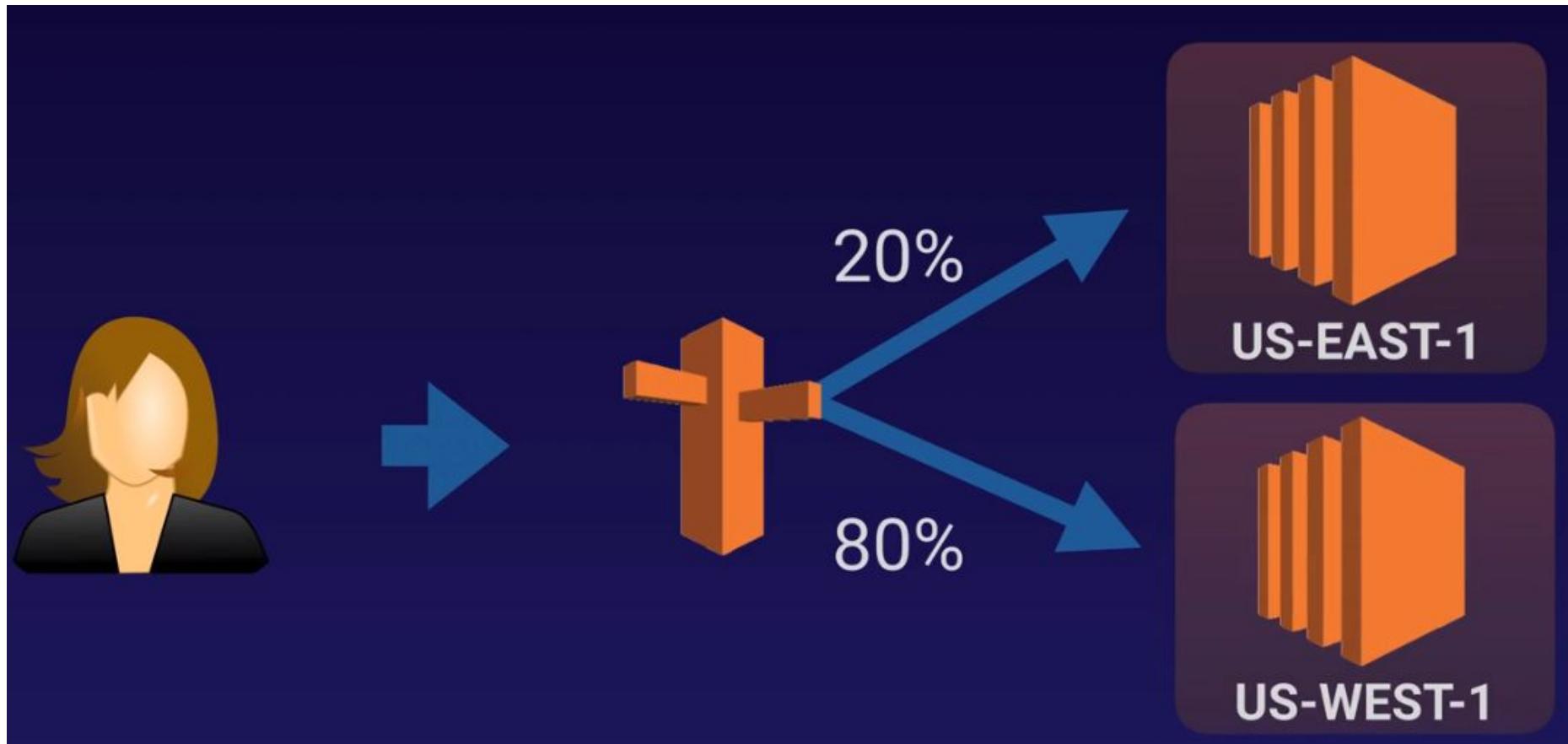
Weighted Routing Policy

Allows you split your traffic based on different weights assigned.

For example, you can set 10% of your traffic to go to US-EAST-1 and 90% to go to EU-WEST-1.



Routing policies



Routing policies

Health Checks

- You can set health checks on individual record sets.
- If a record set fails a health check it will be removed from Route53 until it passes the health check.
- You can set SNS notifications to alert you if a health check is failed.

Routing policies

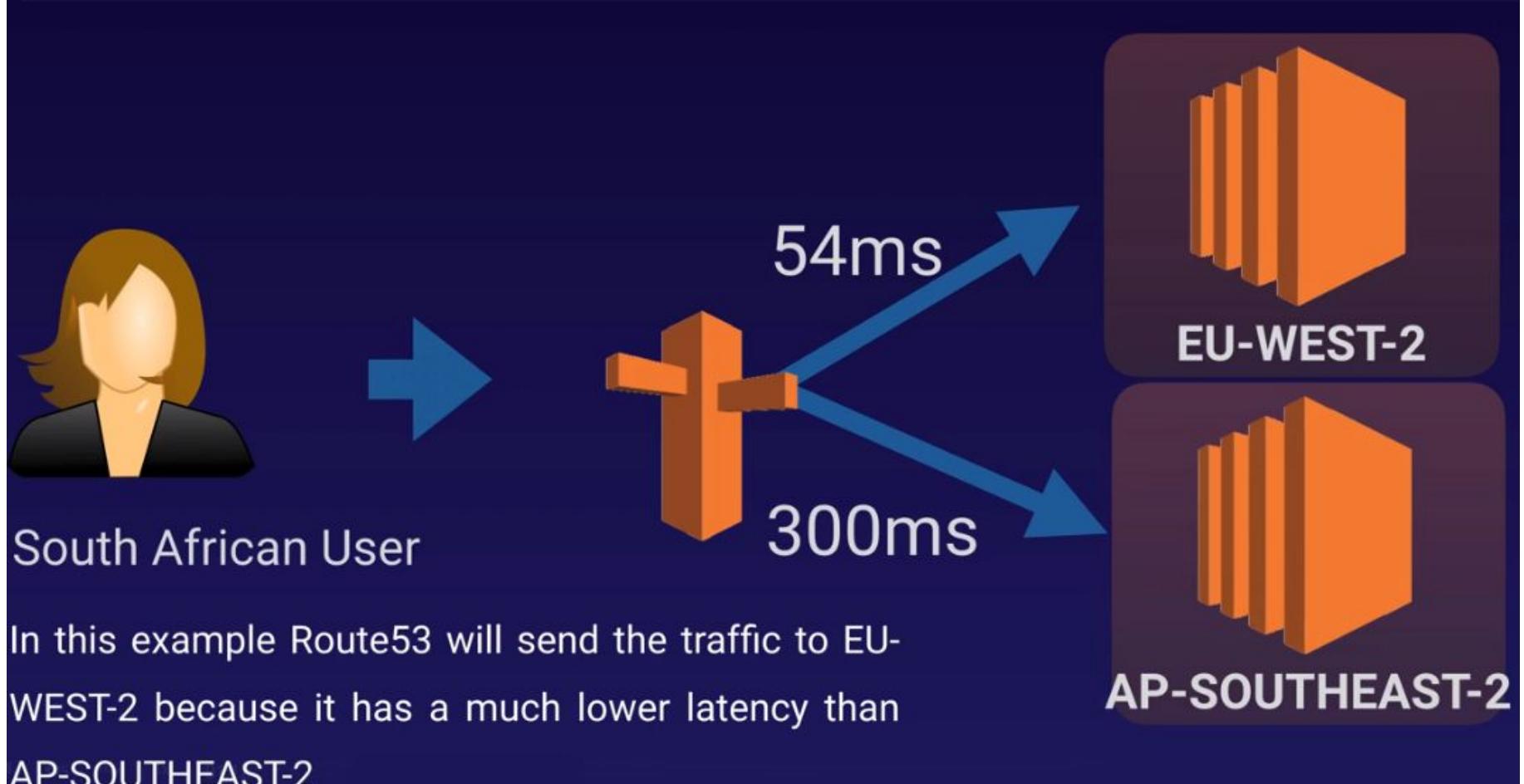
Latency-Based Routing

Allows you to route your traffic based on the lowest network latency for your end user (ie which region will give them the fastest response time).

To use latency-based routing, you create a latency resource record set for the Amazon EC2 (or ELB) resource in each region that hosts your website. When Amazon Route 53 receives a query for your site, it selects the latency resource record set for the region that gives the user the lowest latency. Route 53 then responds with the value associated with that resource record set.



Routing policies



Routing policies

Failover Routing Policy

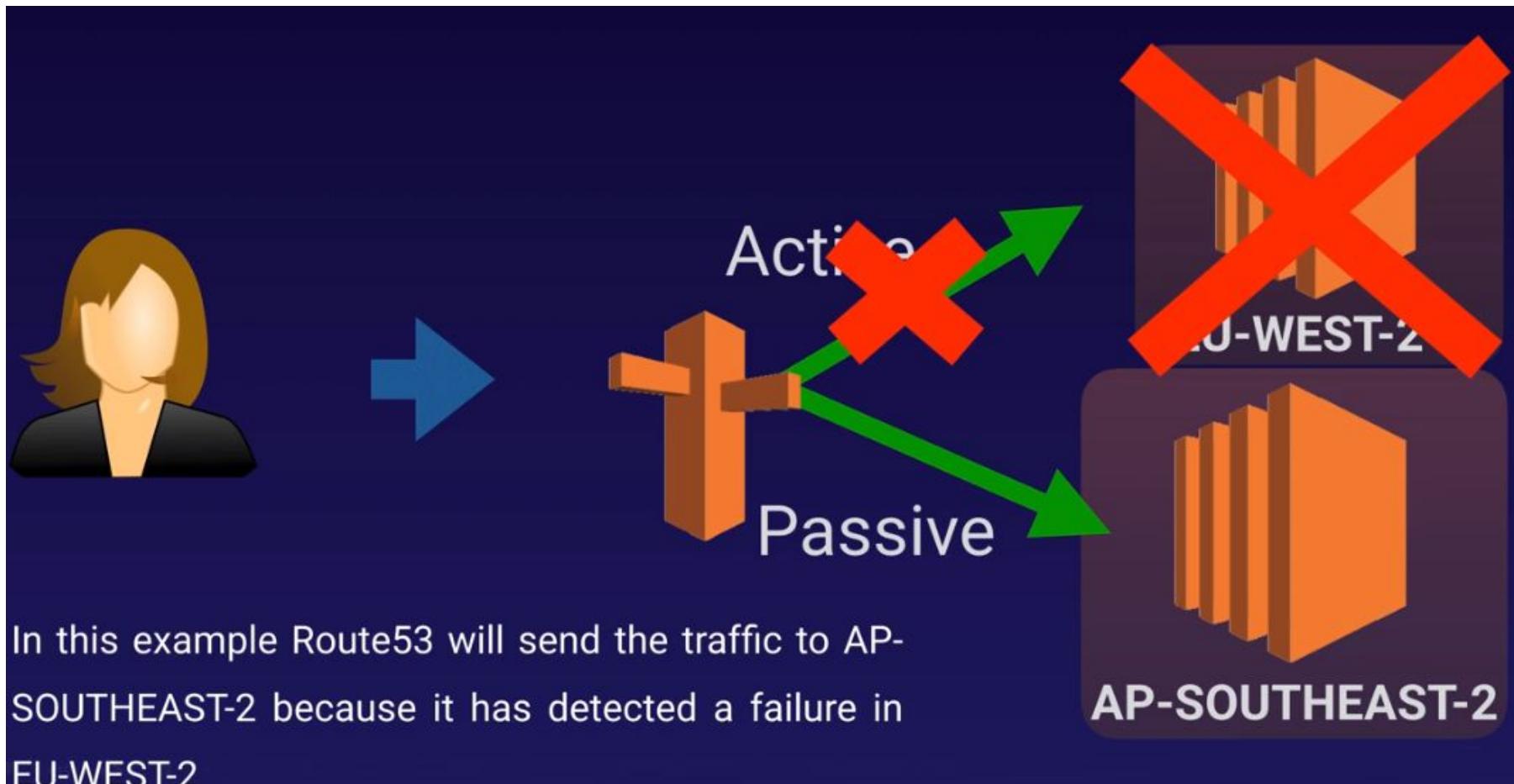
Failover routing policies are used when you want to create an active/passive set up. For example, you may want your primary site to be in EU-WEST-2 and your secondary DR Site in AP-SOUTHEAST-2.

Route53 will monitor the health of your primary site using a health check.

A health check monitors the health of your end points.



Routing policies



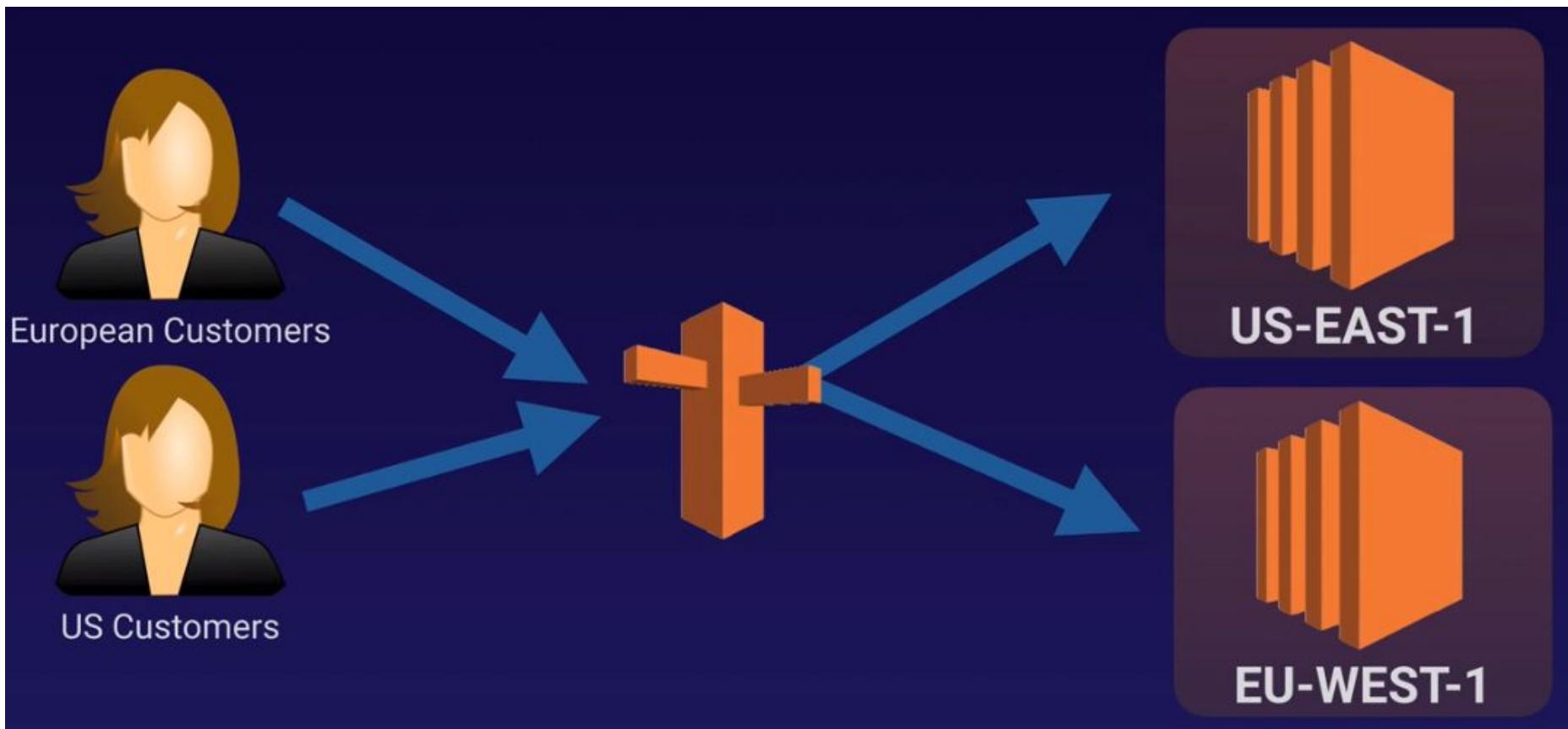
Routing policies

Geolocation Routing Policy

Geolocation routing lets you choose where your traffic will be sent based on the geographic location of your users (ie the location from which DNS queries originate). For example, you might want all queries from Europe to be routed to a fleet of EC2 instances that are specifically configured for your European customers. These servers may have the local language of your European customers and all prices are displayed in Euros.



Routing policies



Routing policies

Geoproximity Routing (Traffic Flow Only)

Geoproximity routing lets Amazon Route 53 route traffic to your resources based on the geographic location of your users and your resources. You can also optionally choose to route more traffic or less 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.

To use geoproximity routing, you must use Route 53 traffic flow.

Routing policies

Multivalue Answer Policy

Multivalue answer routing lets you configure Amazon 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 multivalue answer routing also lets you check the health of each resource, so Route 53 returns only values for healthy resources.

This is similar to simple routing however it allows you to put health checks on each record set.

Routing policies



CHAPTER 8

VPCs

VPC

Think of a VPC as a virtual data centre in the cloud.



VPC

Amazon Virtual Private Cloud (Amazon VPC) lets you provision a logically isolated section of the Amazon Web Services (AWS) Cloud where you can launch AWS resources in a virtual network that you define. You have complete control over your virtual networking environment, including selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways.



VPC

You can easily customize the network configuration for your Amazon Virtual Private Cloud. For example, you can create a public-facing subnet for your webservers that has access to the Internet, and place your backend systems such as databases or application servers in a private-facing subnet with no Internet access. You can leverage multiple layers of security, including security groups and network access control lists, to help control access to Amazon EC2 instances in each subnet.



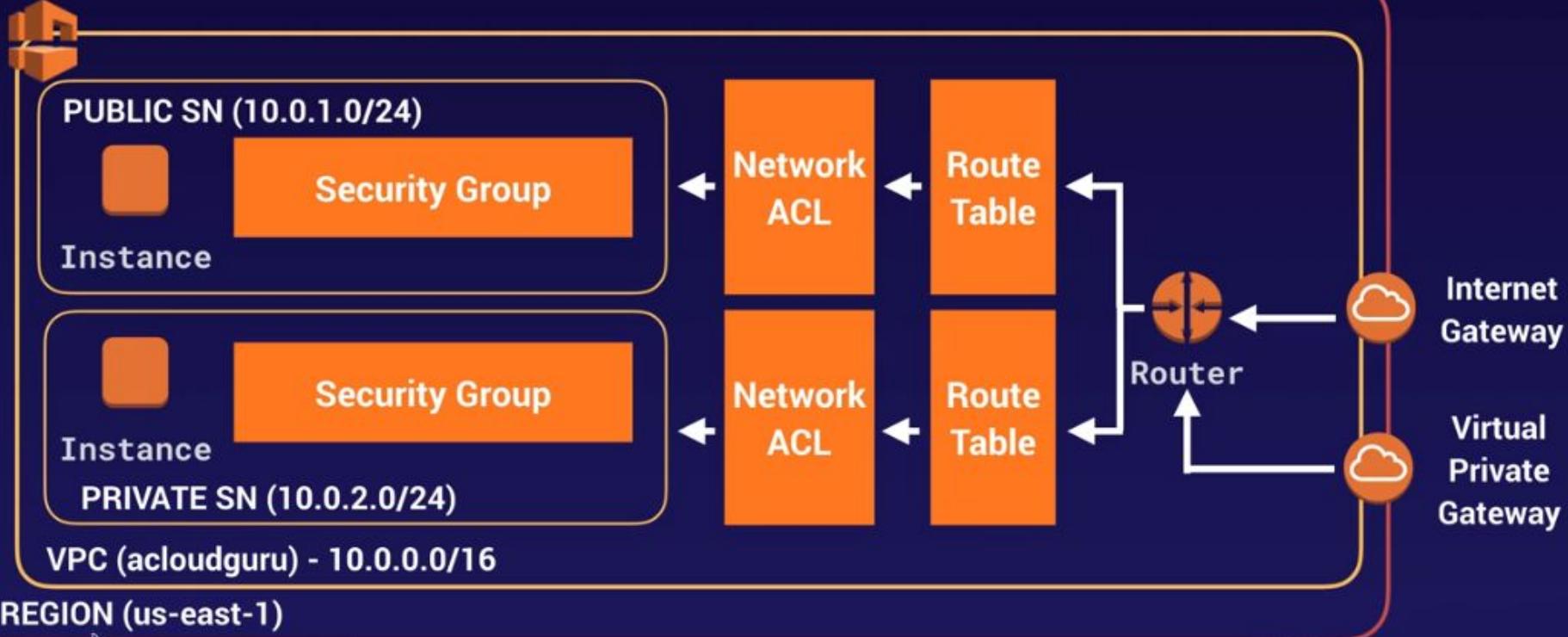
VPC

Additionally, you can create a Hardware Virtual Private Network (VPN) connection between your corporate datacenter and your VPC and leverage the AWS cloud as an extension of your corporate datacenter.



VPC

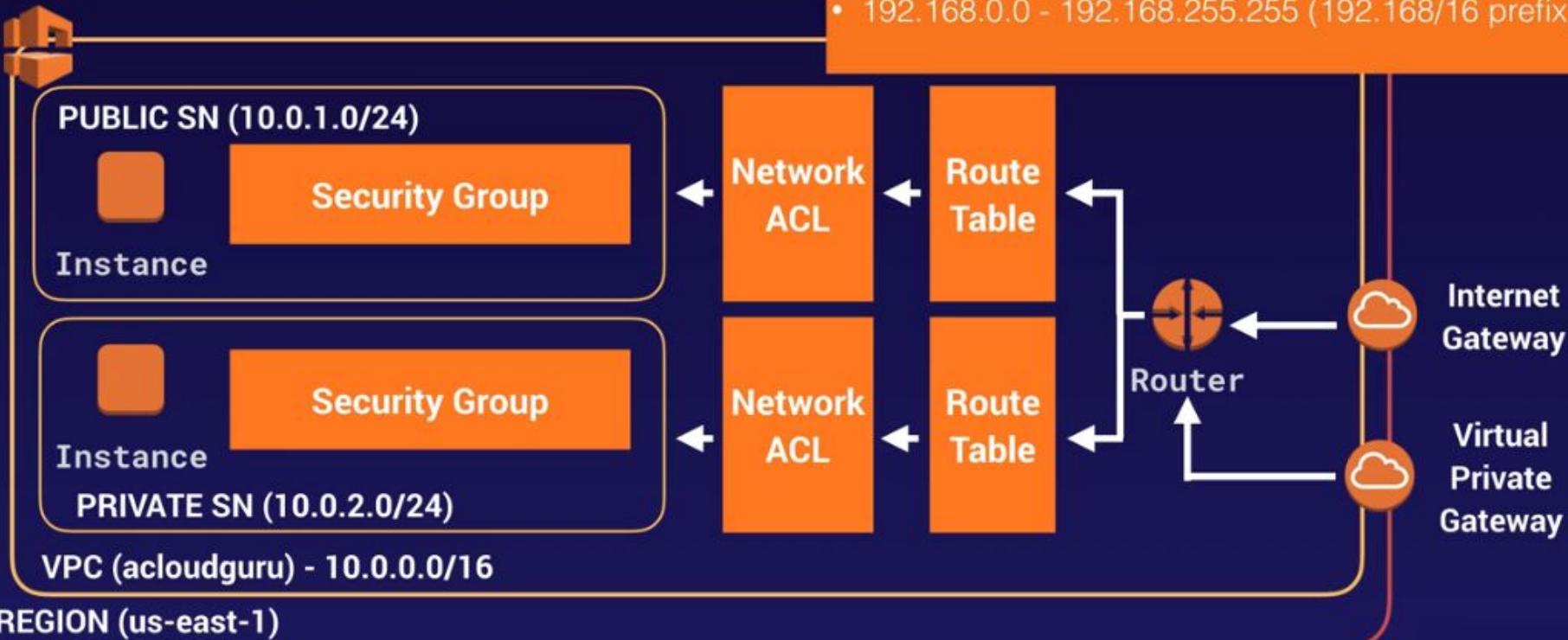
VPC with Public & Private Subnet(s)



VPC

VPC with Public & Private Subnet(s)

- 10.0.0.0 - 10.255.255.255 (10/8 prefix)
- 172.16.0.0 - 172.31.255.255 (172.16/12 prefix)
- 192.168.0.0 - 192.168.255.255 (192.168/16 prefix)



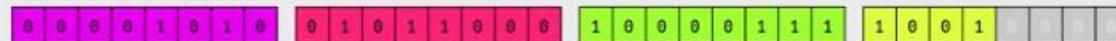
VPC

CIDR.xyz

AN INTERACTIVE IP ADDRESS AND CIDR RANGE VISUALIZER

CIDR is a notation for describing blocks of IP addresses and is used heavily in various networking configurations. IP addresses contain 4 octets, each consisting of 8 bits giving values between 0 and 255. The decimal value that comes after the slash is the number of bits consisting of the routing prefix. This in turn can be translated into a netmask, and also designates how many available addresses are in the block.

10 . 88 . 135 . 144 / 28



255.255.255.240
NETMASK

10.88.135.145
FIRST IP

10.88.135.158
LAST IP

16
COUNT

Created by [Yuval Adam](#). Source available on [Github](#).

VPC

What can we do with a VPC?

- Launch instances into a subnet of your choosing
- Assign custom IP address ranges in each subnet
- Configure route tables between subnets
- Create internet gateway and attach it to our VPC
- Much better security control over your AWS resources
- Instance security groups
- Subnet network access control lists (ACLs)



VPC peering

Default VPC vs Custom VPC

- Default VPC is user friendly, allowing you to immediately deploy instances.
- All Subnets in default VPC have a route out to the internet.
- Each EC2 instance has both a public and private IP address.

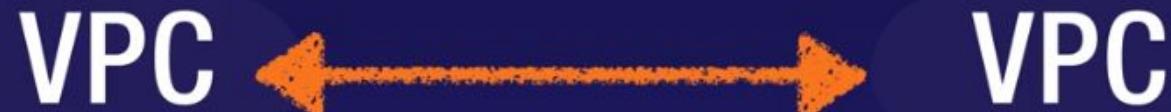
A large, stylized white 'VPC' logo is positioned in the bottom right corner of the slide. The letters are bold and blocky, with a slight shadow effect.

VPC

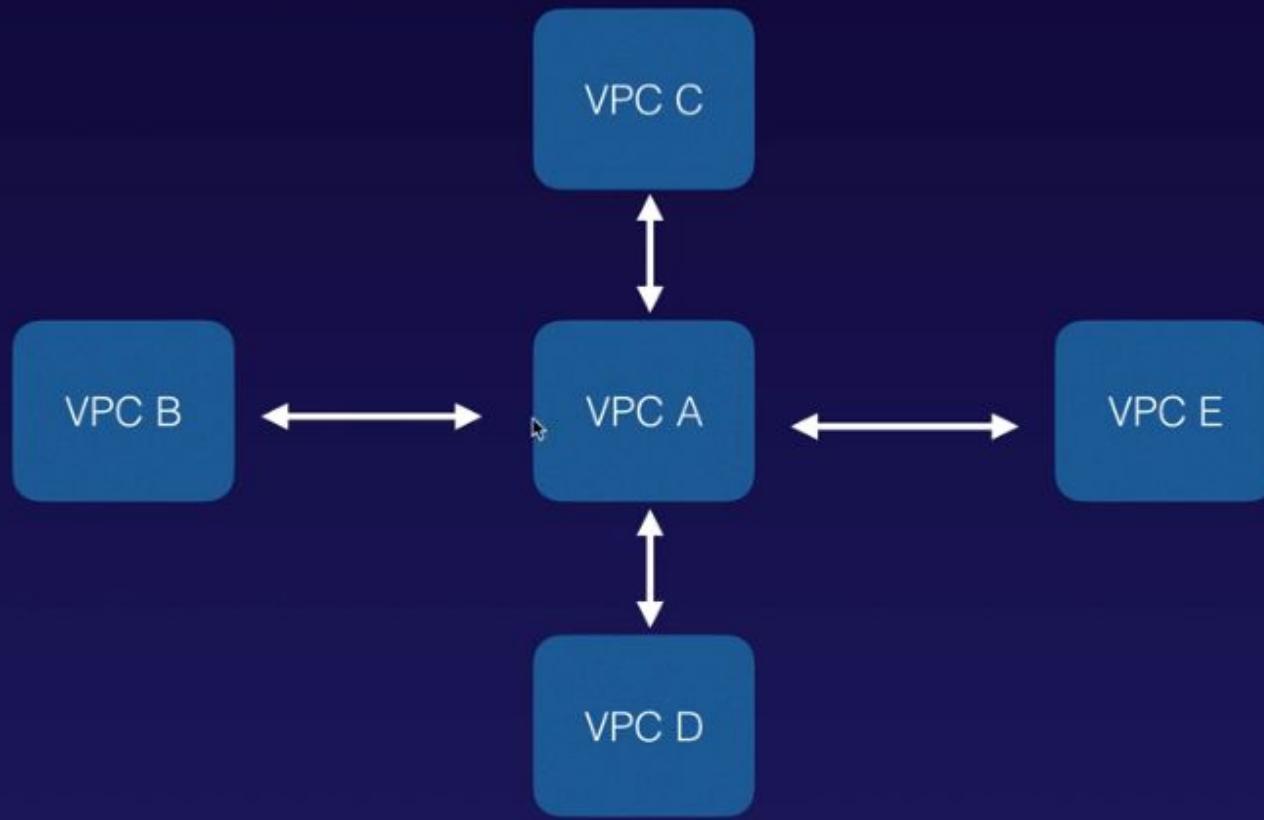
VPC peering

VPC Peering

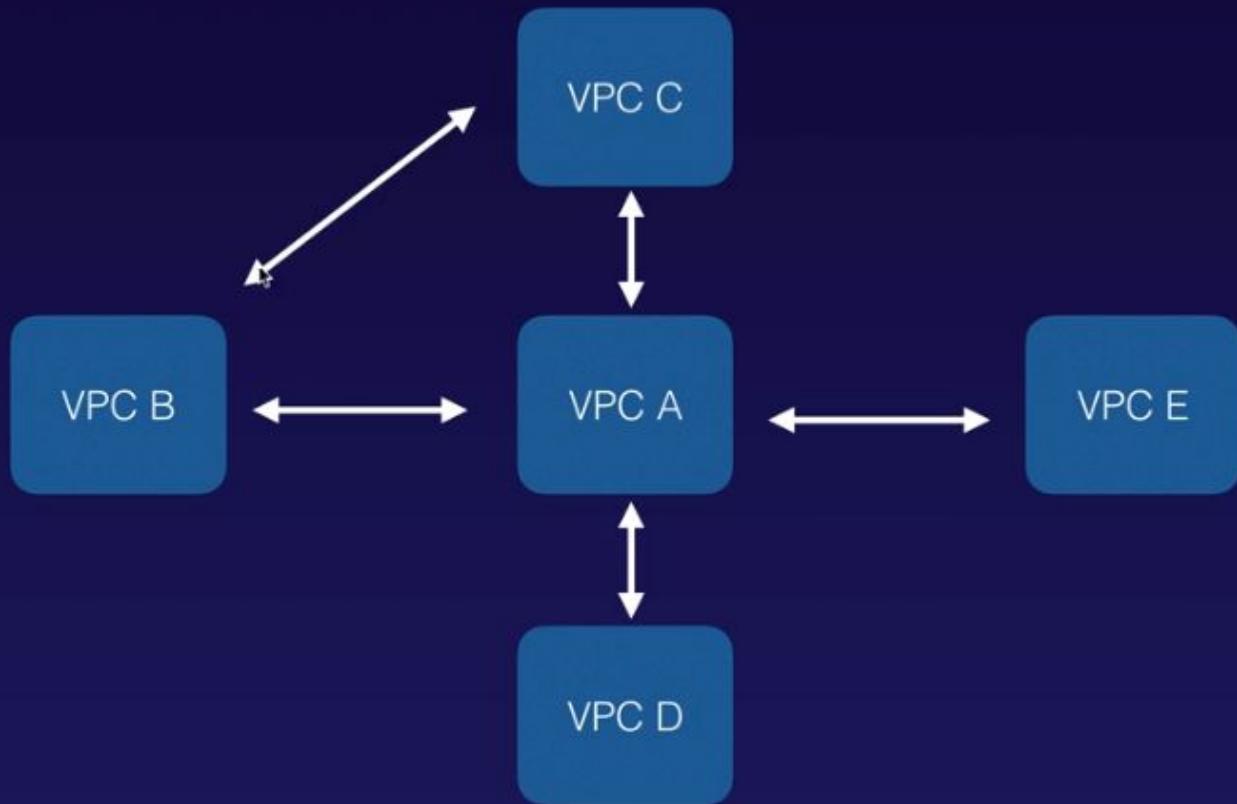
- Allows you to connect one VPC with another via a direct network route using private IP addresses.
- Instances behave as if they were on the same private network
- You can peer VPC's with other AWS accounts as well as with other VPCs in the same account.
- Peering is in a star configuration: ie 1 central VPC peers with 4 others.
NO TRANSITIVE PEERING!!!
- You can peer between regions.



VPC peering



VPC peering



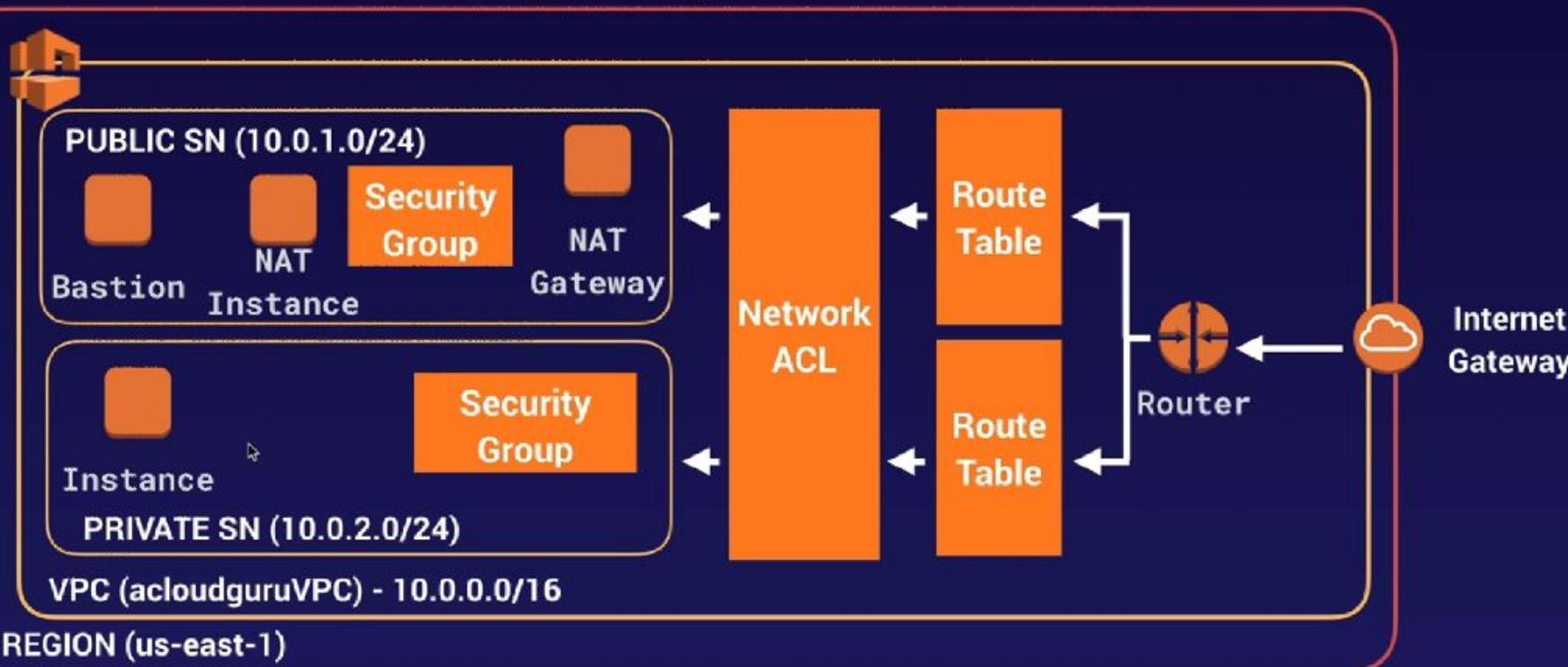
Exam tips

Remember the following:

- Think of a VPC as a logical datacenter in AWS.
- Consists of IGWs (Or Virtual Private Gateways), Route Tables, Network Access Control Lists, Subnets, and Security Groups
- 1 Subnet = 1 Availability Zone
- Security Groups are Stateful; Network Access Control Lists are Stateless
- NO TRANSITIVE PEERING

Bastions in actions

VPC with Public & Private Subnet(s)



Exam tips

Remember the following;

- A NAT Gateway or NAT Instance is used to provide internet traffic to EC2 instances in a private subnets.
- A Bastion is used to securely administer EC2 instances (Using SSH or RDP). Bastions are called Jump Boxes in Australia.
- You cannot use a NAT Gateway as a Bastion host.

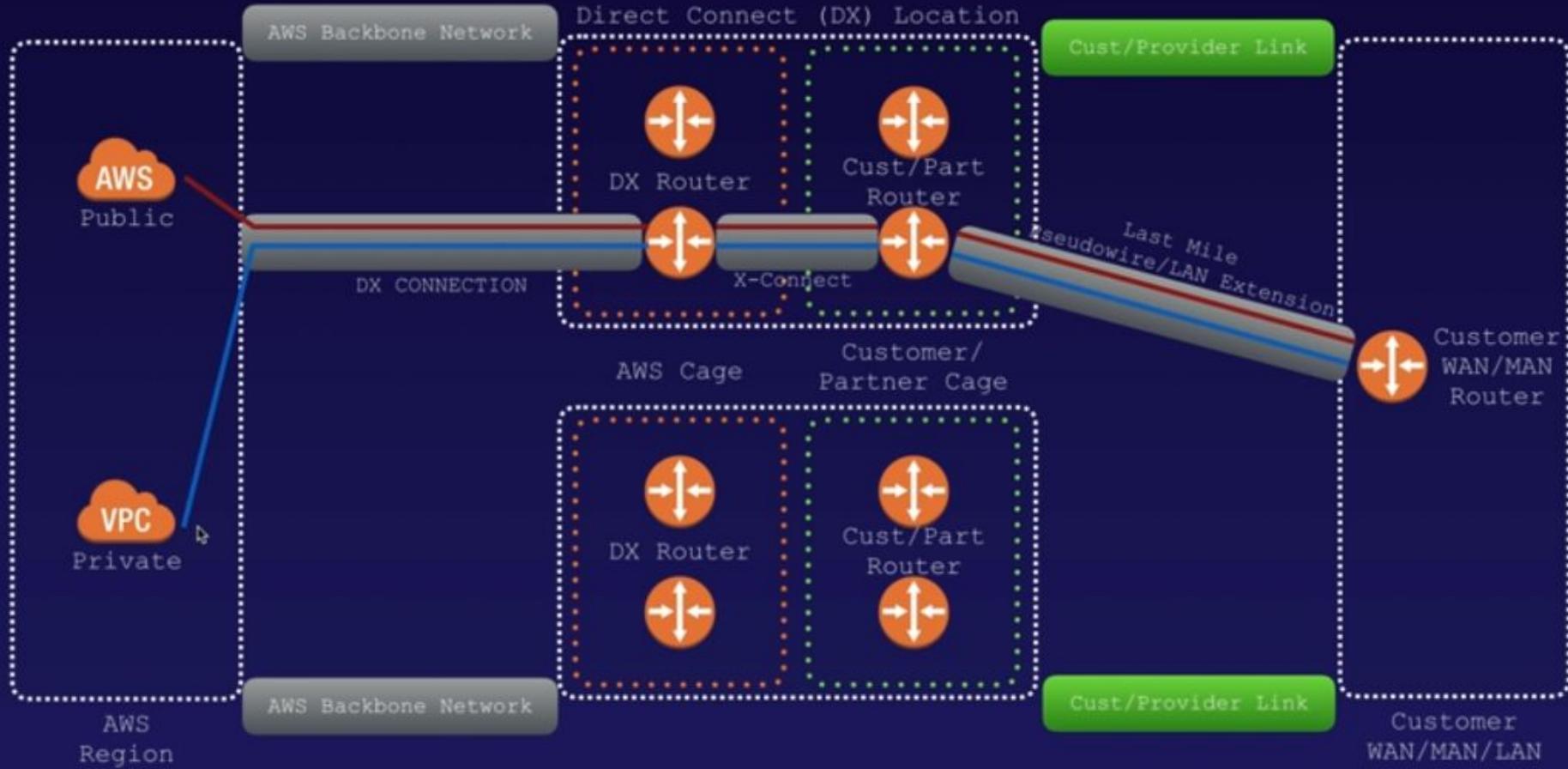
What is Direct Connect?

Direct Connect

AWS Direct Connect is a cloud service solution that makes it easy to establish a dedicated network connection from your premises to AWS. Using AWS Direct Connect, you can establish private connectivity between AWS and your datacenter, office, or colocation environment, which in many cases can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience than Internet-based connections.



Direct connect in Action



Setting Up Direct Connect - Steps

Steps to setting up Direct Connect

- Create a virtual interface in the Direct Connect console. This is a **PUBLIC Virtual Interface**.
- Go to the VPC console and then to VPN connections. Create a Customer Gateway.
- Create a Virtual Private Gateway
- Attach the Virtual Private Gateway to the desired VPC.
- Select VPN Connections and create new VPN Connection.
- Select the Virtual Private Gateway and the Customer Gateway
- Once the VPN is available, set up the VPN on the customer gateway or firewall.



Exam tips - Direct Connect

Remember the following:

- Direct Connect directly connects your data center to AWS
- Useful for high throughput workloads (ie lots of network traffic)
- Or if you need a stable and reliable secure connection.

Exam tips - Direct Connect

Remember the Steps to Creating a Direct Connect Connection.

- Create a virtual interface in the Direct Connect console. This is a PUBLIC Virtual Interface.
- Go to the VPC console and then to VPN connections. Create a Customer Gateway.
- Create a Virtual Private Gateway
- Attach the Virtual Private Gateway to the desired VPC.
- Select VPN Connections and create new VPN Connection.
- Select the Virtual Private Gateway and the Customer Gateway
- Once the VPN is available, setup the VPN on the customer gateway or firewall.

VPC endpoints

A VPC Endpoint:

A VPC endpoint enables you to privately connect your VPC to supported AWS services and VPC endpoint services powered by PrivateLink without requiring an internet gateway, NAT device, VPN connection, or AWS Direct Connect connection. Instances in your VPC do not require public IP addresses to communicate with resources in the service. Traffic between your VPC and the other service does not leave the Amazon network.

Endpoints are virtual devices. They are horizontally scaled, redundant, and highly available VPC components that allow communication between instances in your VPC and services without imposing availability risks or bandwidth constraints on your network traffic.



VPC endpoints

There are two types of VPC endpoints:

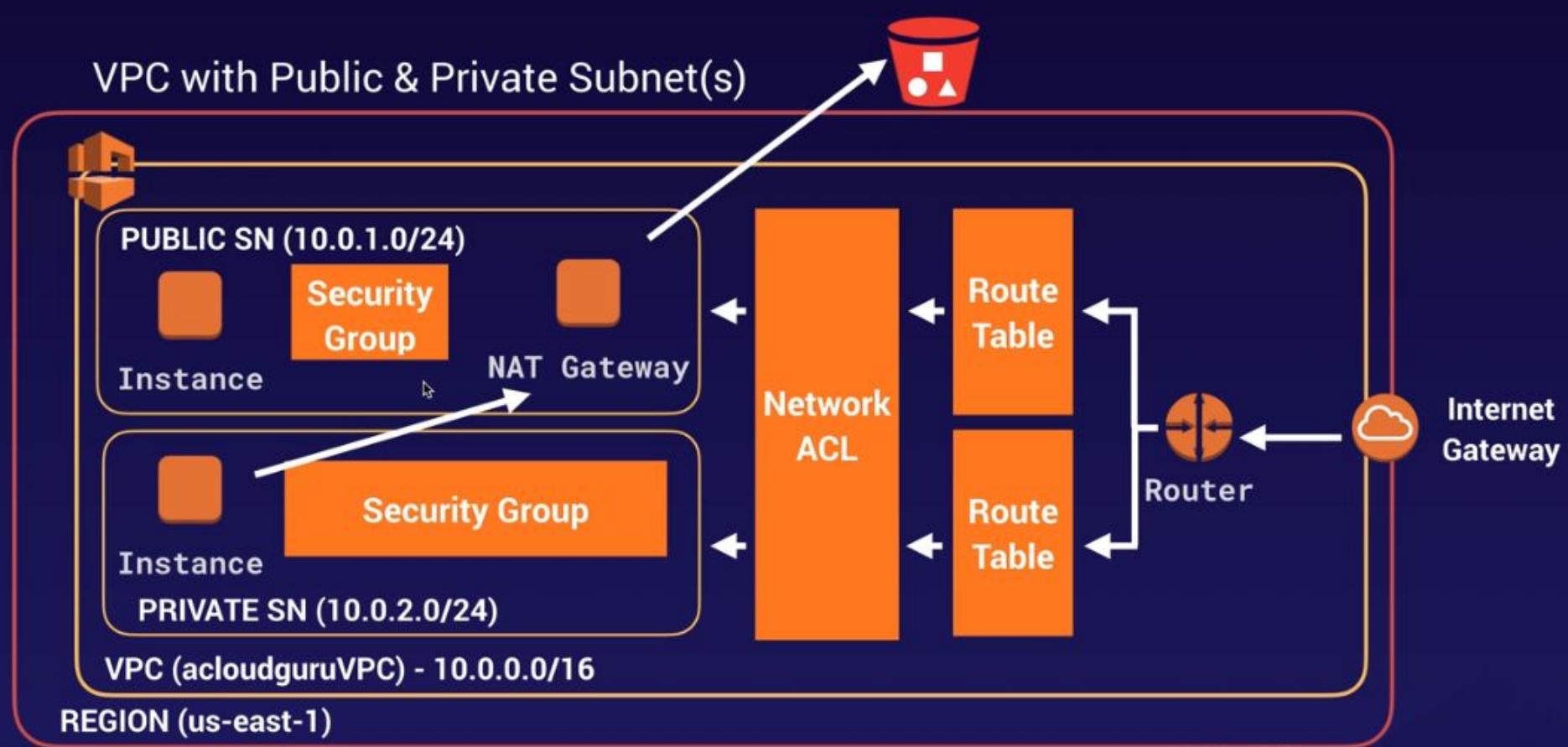
- Interface Endpoints
- Gateway Endpoints

VPC endpoints

An interface endpoint is an elastic network interface with a private IP address that serves as an entry point for traffic destined to a supported service. The following services are supported:

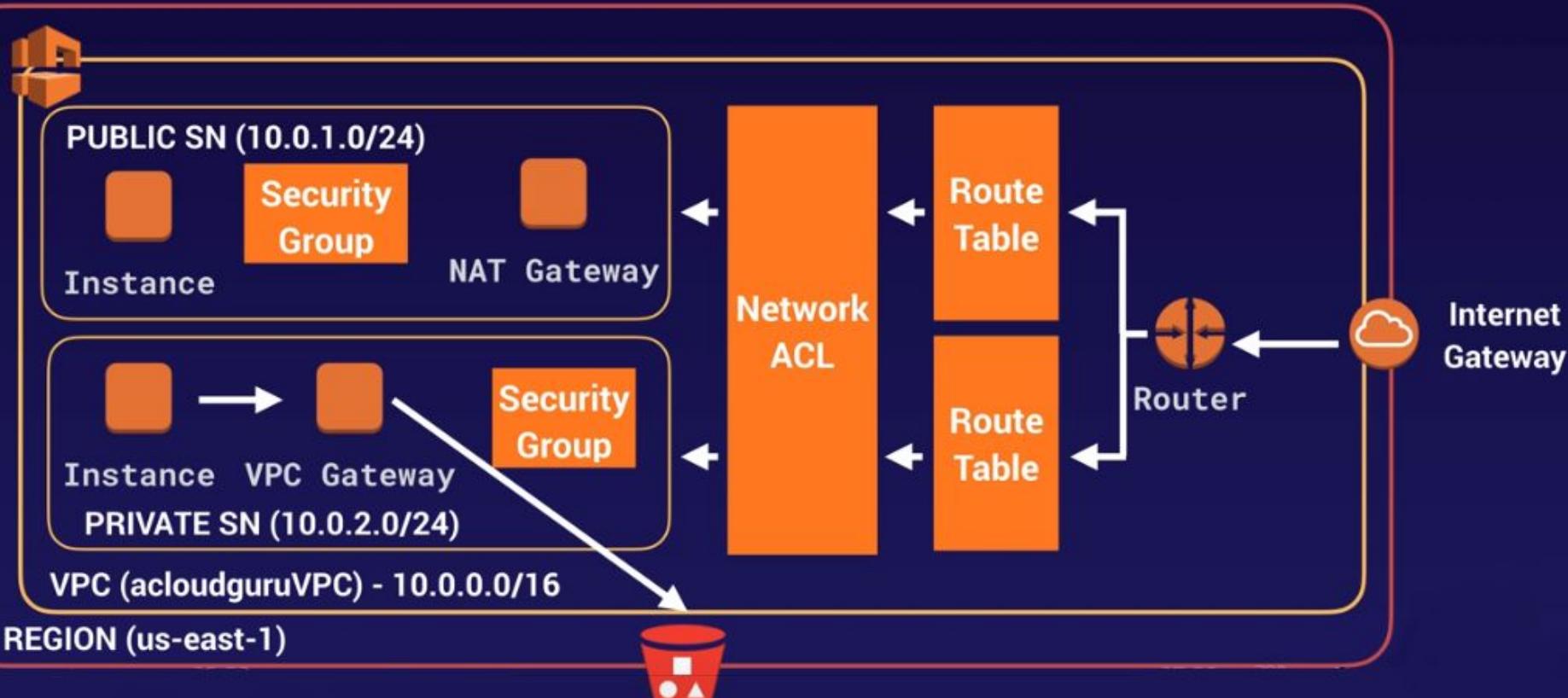
- Amazon API Gateway
- AWS CloudFormation
- Amazon CloudWatch
- Amazon CloudWatch Events
- Amazon CloudWatch Logs
- AWS CodeBuild
- AWS Config
- Amazon EC2 API
- Elastic Load Balancing API
- AWS Key Management Service
- Amazon Kinesis Data Streams
- Amazon SageMaker and Amazon SageMaker Runtime
- Amazon SageMaker Notebook Instance
- AWS Secrets Manager
- AWS Security Token Service
- AWS Service Catalog
- Amazon SNS
- Amazon SQS
- AWS Systems Manager
- Endpoint services hosted by other AWS accounts
- Supported AWS Marketplace partner services

VPC endpoints



Exam tips - VPC endpoints

VPC with Public & Private Subnet(s)



Exam tips - VPC endpoints

A VPC Endpoint:

A VPC endpoint enables you to privately connect your VPC to supported AWS services and VPC endpoint services powered by PrivateLink without requiring an internet gateway, NAT device, VPN connection, or AWS Direct Connect connection. Instances in your VPC do not require public IP addresses to communicate with resources in the service. Traffic between your VPC and the other service does not leave the Amazon network.

Endpoints are virtual devices. They are horizontally scaled, redundant, and highly available VPC components that allow communication between instances in your VPC and services without imposing availability risks or bandwidth constraints on your network traffic.

Exam tips - VPC endpoints

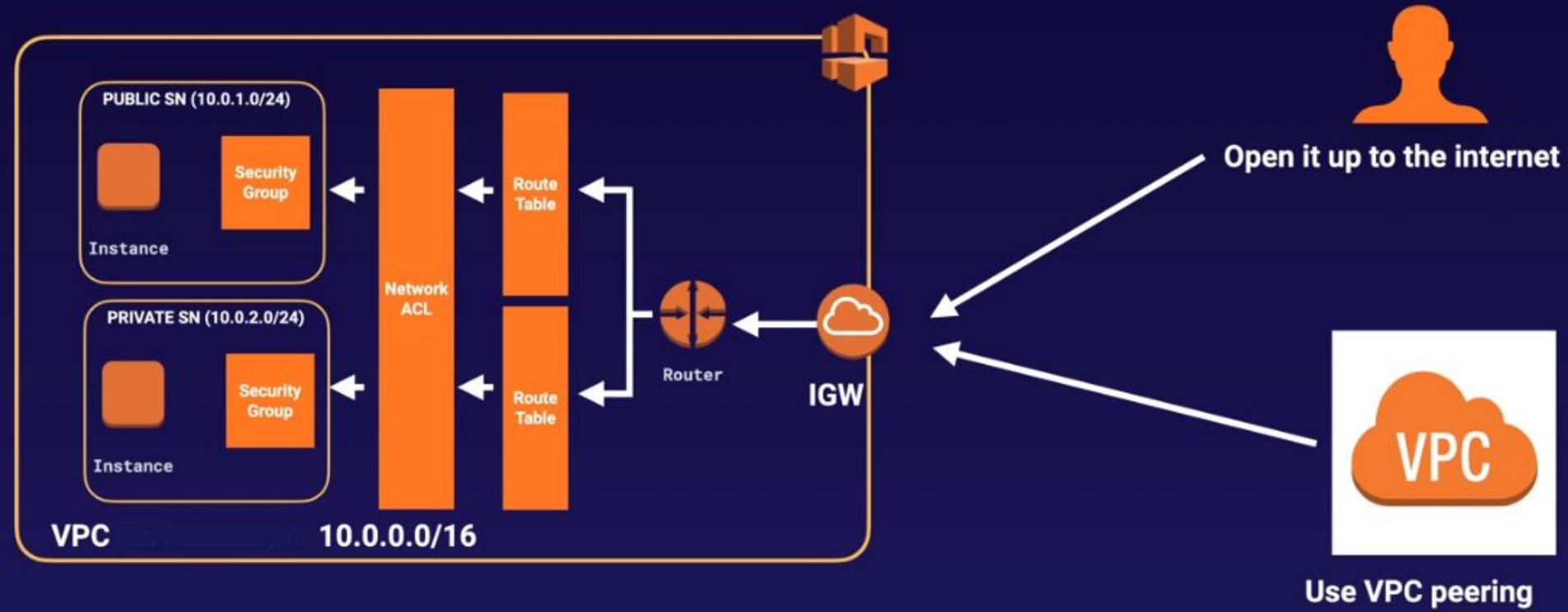
There are two types of VPC endpoints:

- Interface Endpoints
- Gateway Endpoints

Currently Gateway Endpoints Support:

- Amazon S3
- DynamoDB

AWS PrivateLink: Opening Your Services in a VPC to another VPC



AWS PrivateLink: Sharing Application Across VPCs

To open our applications up to other VPCs, we can either:

Open the VPC up to the internet:

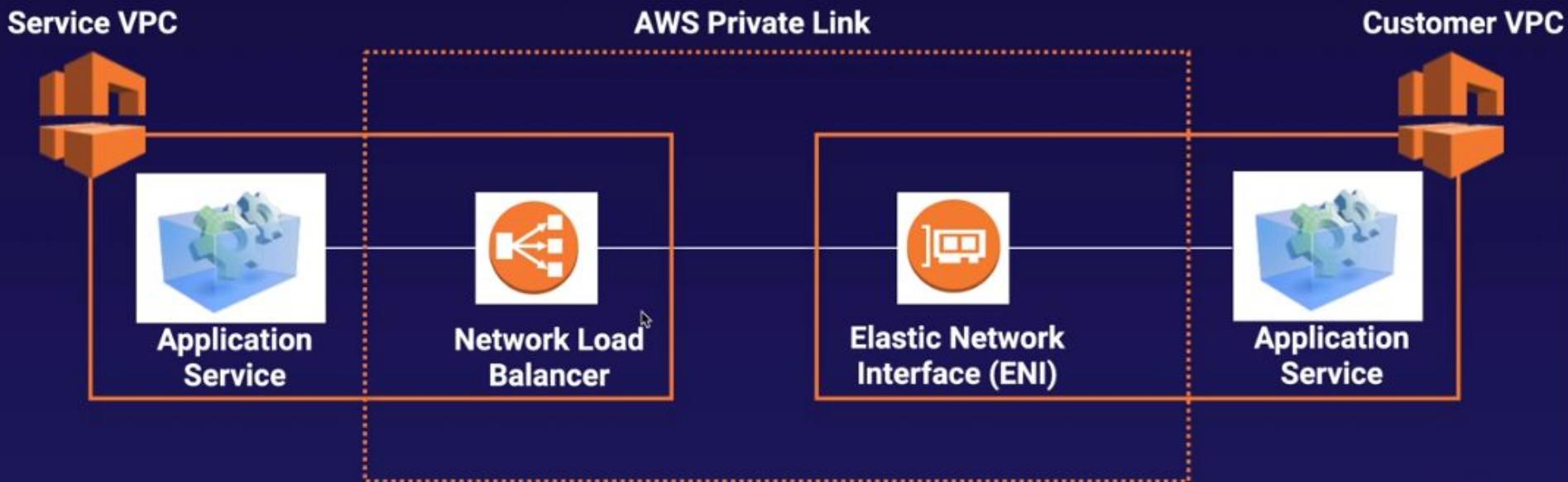
- Security considerations; everything in the public subnet is public.
- A lot more to manage.

Use VPC peering:

- You will have to create and manage many different peering relationships.
- The whole network will be accessible. This isn't good if you have multiple applications within your VPC.

AWS PrivateLink

- ✓ The best way to expose a service VPC to tens, hundreds, or thousands of customer VPCs
- ✓ Doesn't require VPC peering; no route tables, NAT, IGWs, etc.
- ✓ Requires a Network Load Balancer on the service VPC and an ENI on the customer VPC

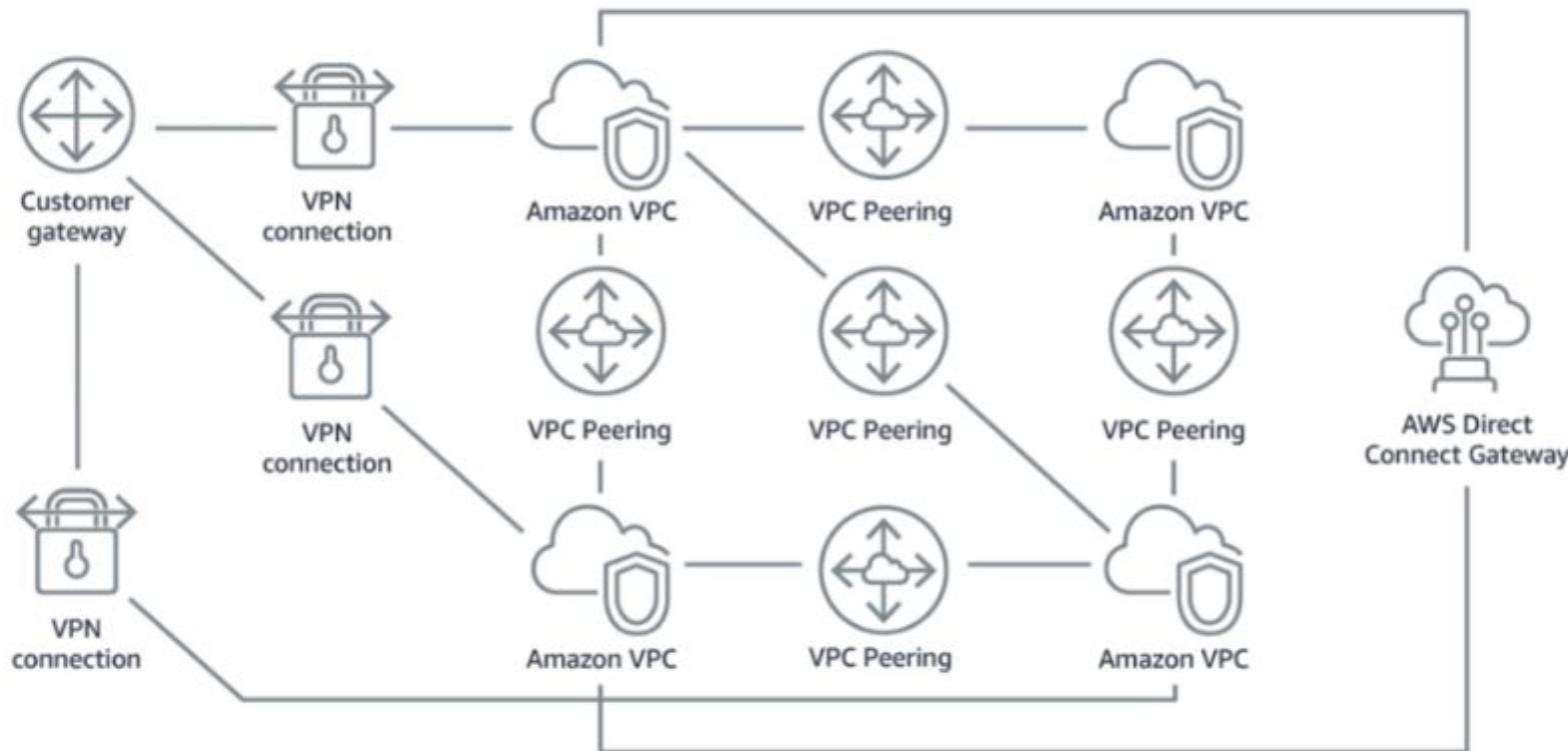


Exam tips: AWS PrivateLink

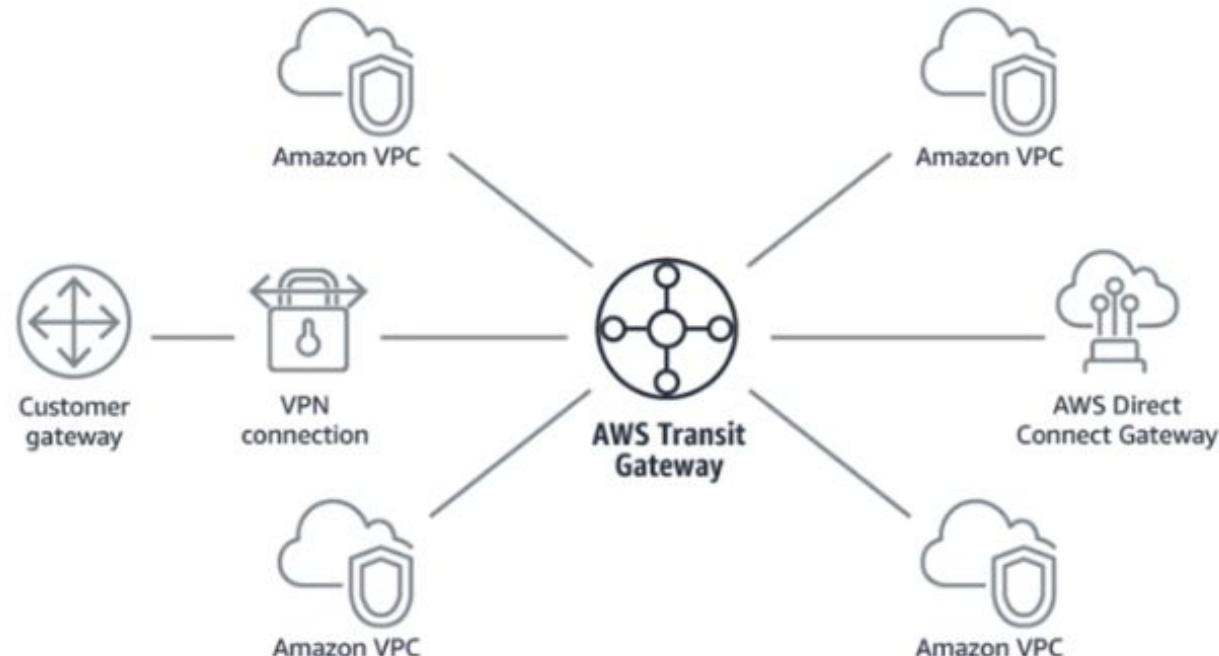
AWS PrivateLink

- If you see a question asking about peering VPCs to tens, hundreds, or thousands of customer VPCs, think of AWS PrivateLink.
- Doesn't require VPC peering; no route tables, NAT, IGWs, etc.
- Requires a Network Load Balancer on the service VPC and an ENI on the customer VPC

AWS Transit Gateway



Simplify your network topology

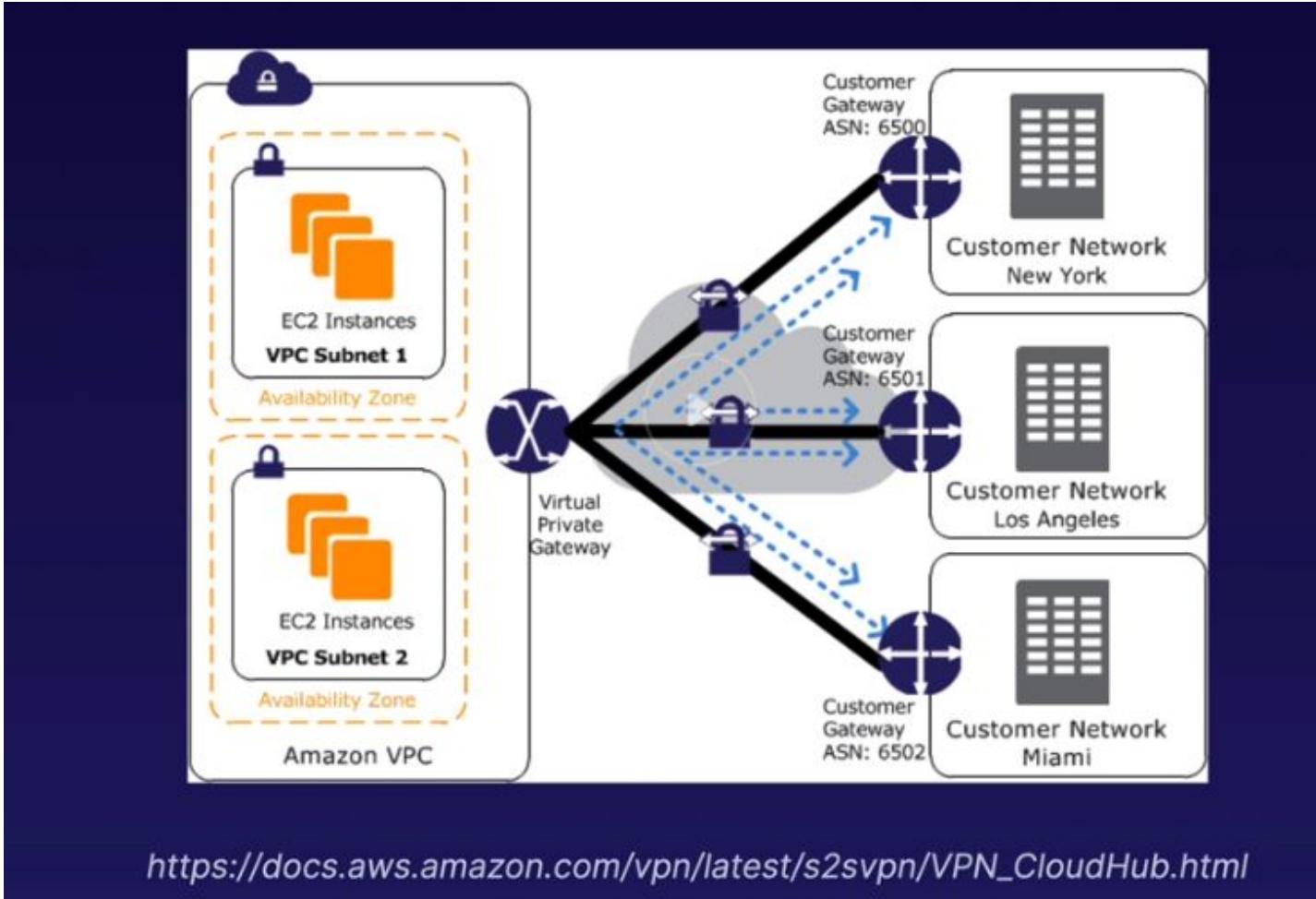


Exam tips: AWS Transit Gateway

Transit Gateway

- Allows you to have transitive peering between thousands of VPCs and on-premises data centers.
- Works on a hub-and-spoke model.
- Works on a regional basis, but you can have it across multiple regions.
- You can use it across multiple AWS accounts using RAM (Resource Access Manager).
- You can use route tables to limit how VPCs talk to one another.
- Works with Direct Connect as well as VPN connections.
- Supports **IP multicast** (not supported by any other AWS service).

AWS VPN CloudHub

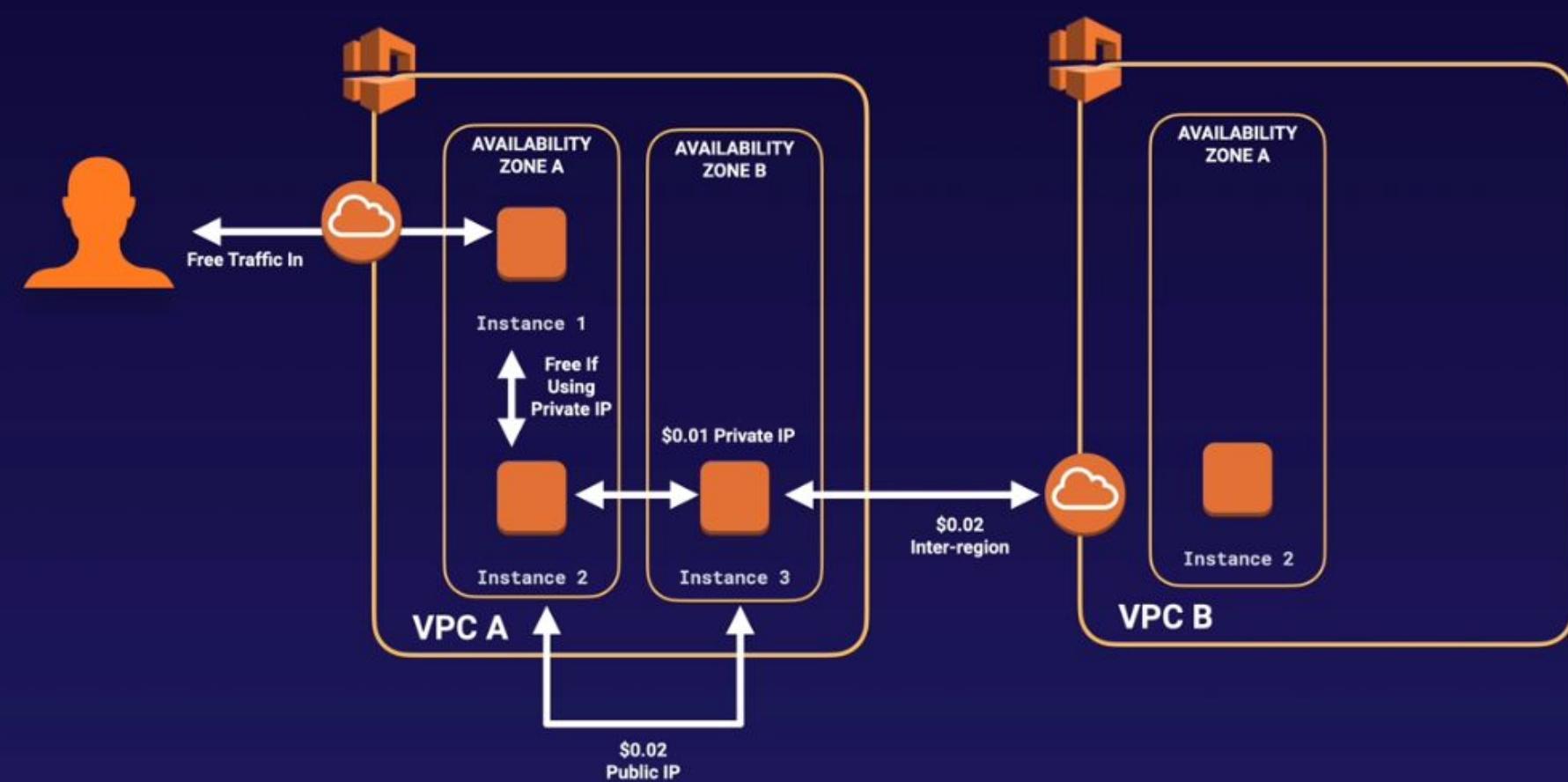


Exam tips: AWS VPN CloudHub

AWS VPN CloudHub

- If you have multiple sites, each with its own VPN connection, you can use AWS VPN CloudHub to connect those sites together.
- Hub-and-spoke model.
- Low cost; easy to manage.
- It operates over the public internet, but all traffic between the customer gateway and the AWS VPN CloudHub is encrypted.

AWS Network Costs



Exam tips: AWS Network Costs

AWS Network Costs

- Use private IP addresses over public IP addresses to save on costs. This then utilizes the AWS backbone network.
- If you want to cut all network costs, group your EC2 instances in the same Availability Zone and use private IP addresses. This will be cost-free, but make sure to keep in mind single point of failure issues.

CHAPTER 9

HA Architecture

Elastic Load Balancers

Load Balancer Types

- 1 Application Load Balancer
- 2 Network Load Balancer
- 3 Classic Load Balancer



Elastic Load Balancers

Application Load Balancers are best suited for load balancing of HTTP and HTTPS traffic. They operate at Layer 7 and are application-aware. They are intelligent, and you can create advanced request routing, sending specified requests to specific web servers.



Elastic Load Balancers

Network Load Balancers are best suited for load balancing of TCP traffic where extreme performance is required. Operating at the connection level (Layer 4), Network Load Balancer are capable of handling millions of requests per second, while maintaining ultra-low latencies.

Use for extreme performance!



Elastic Load Balancers

Classic Load Balancers are the legacy Elastic Load Balancers. You can load balance HTTP/HTTPS applications and use Layer 7-specific features, such as X-Forwarded and sticky sessions. You can also use strict Layer 4 load balancing for applications that rely purely on the TCP protocol.



Errors in Load balancing

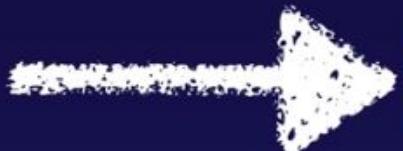
Classic Load Balancers -if your application stops responding, the ELB (Classic Load Balancer) responds with a 504 error. This means that the application is having issues. This could be either at the Web Server layer or at the Database Layer. Identify where the application is failing, and scale it up or out where possible.



X-Forward-For Header



124.12.3.231



10.0.0.23



10.0.0.23
X-Forwarded-For

124.12.3.231

What Are Sticky Sessions

What Are Sticky Sessions?

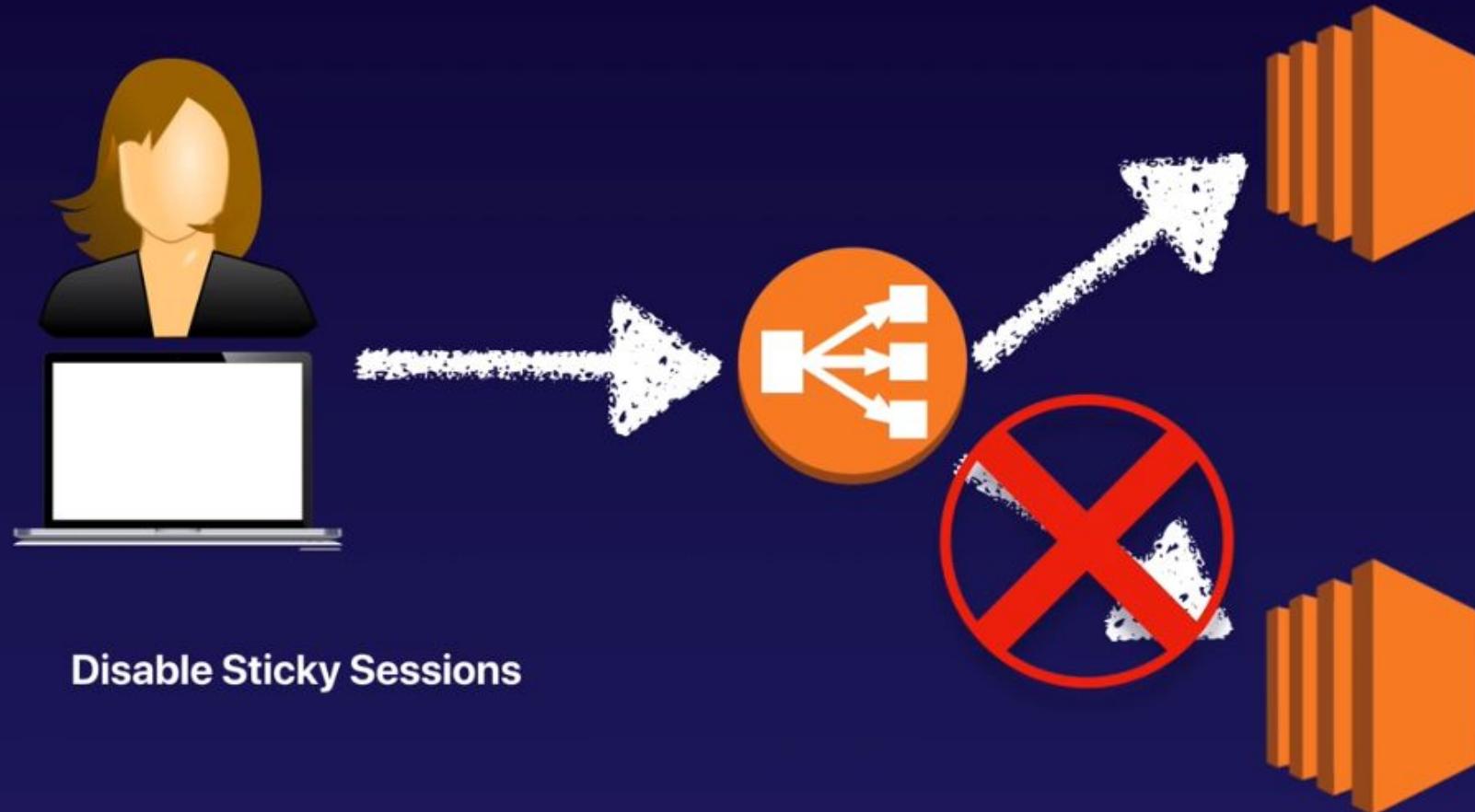
Classic Load Balancer routes each request independently to the registered EC2 instance with the smallest load.

Sticky sessions allow you to bind a user's session to a specific EC2 instance. This ensures that all requests from the user during the session are sent to the same instance.

You can enable Sticky Sessions for Application Load Balancers as well, but the traffic will be sent at the Target Group Level.

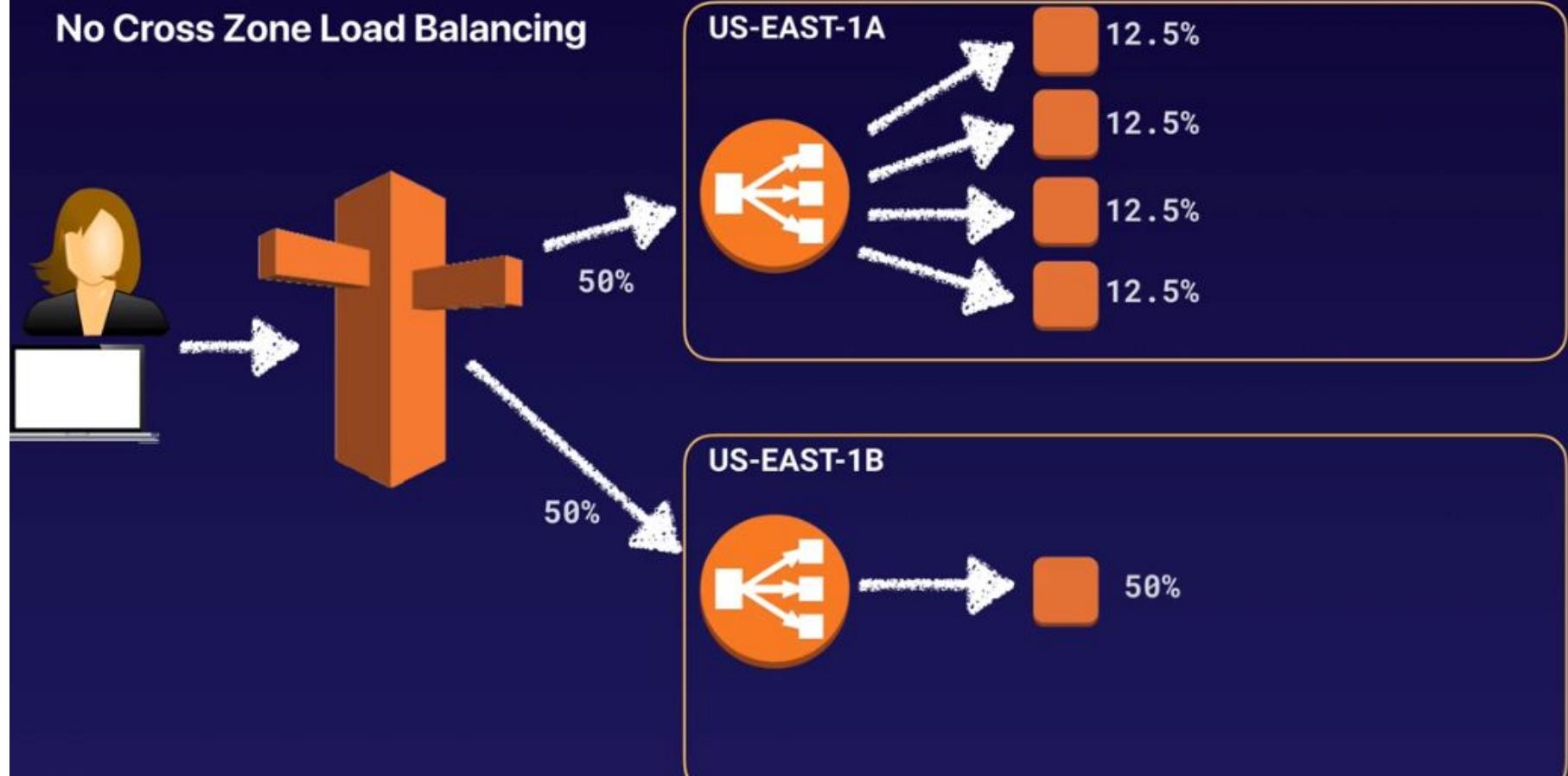


What Are Sticky Sessions



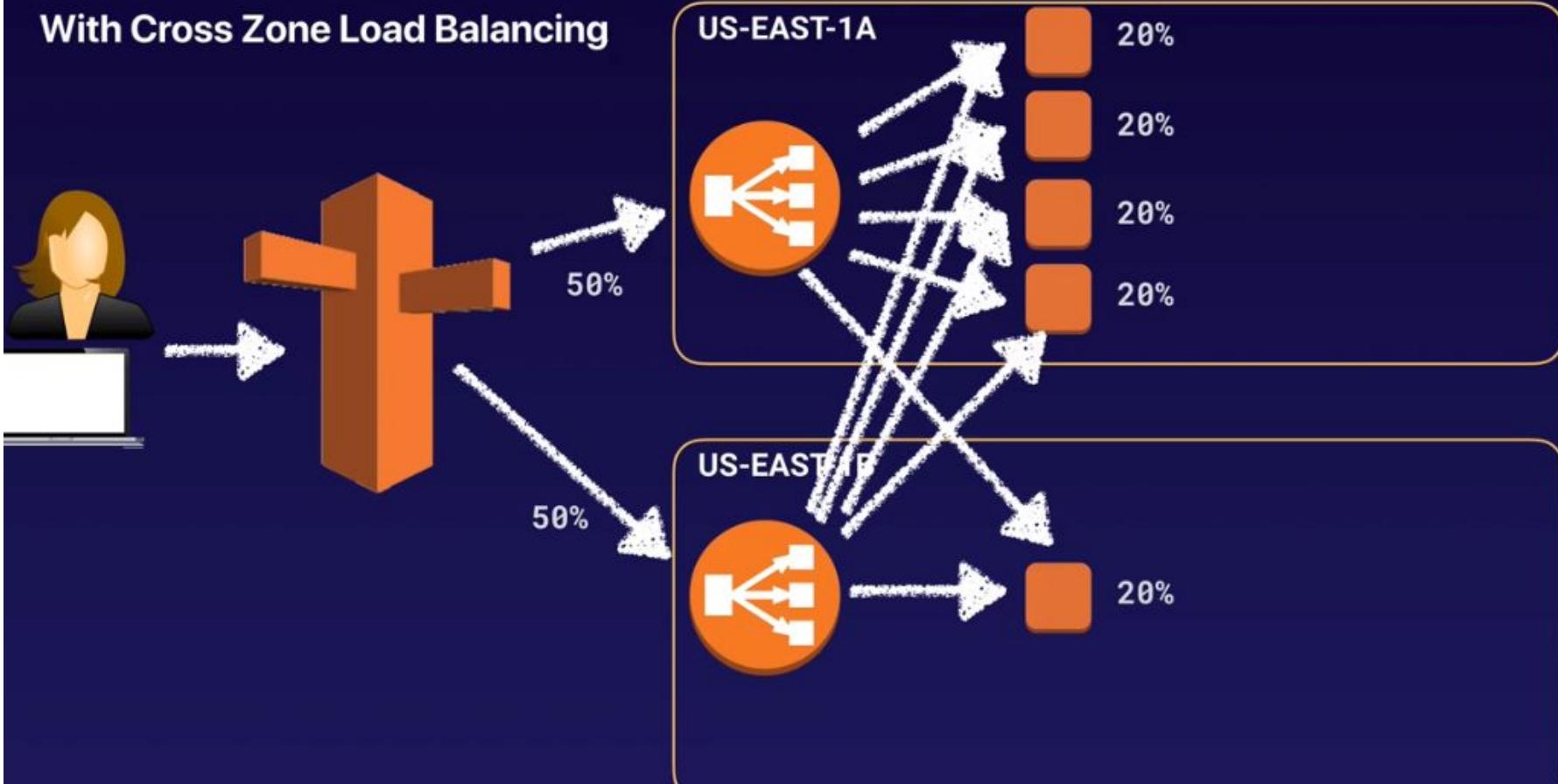
What Is Cross Zone Load Balancing

No Cross Zone Load Balancing



What Is Cross Zone Load Balancing

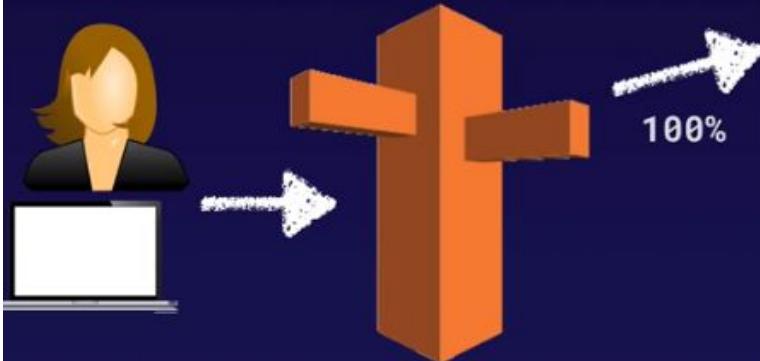
With Cross Zone Load Balancing



What Is Cross Zone Load Balancing

Common Exam Scenario

Enable Cross Zone Load Balancing



Path Patterns

What Are Path Patterns?

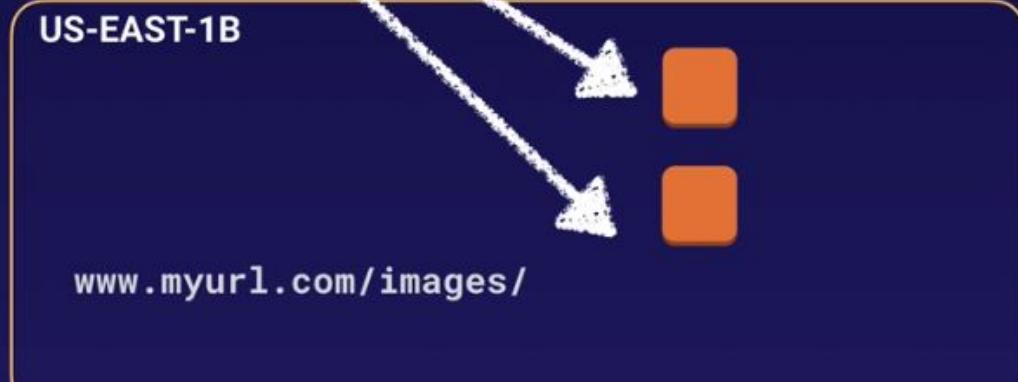
You can create a listener with rules to forward requests based on the URL path. This is known as path-based routing. If you are running microservices, you can route traffic to multiple back-end services using path-based routing. For example, you can route general requests to one target group and requests to render images to another target group.



Path Patterns

Common Exam Scenario

Enable Path Patterns



Exam Tips

Advanced Load Balancer Theory

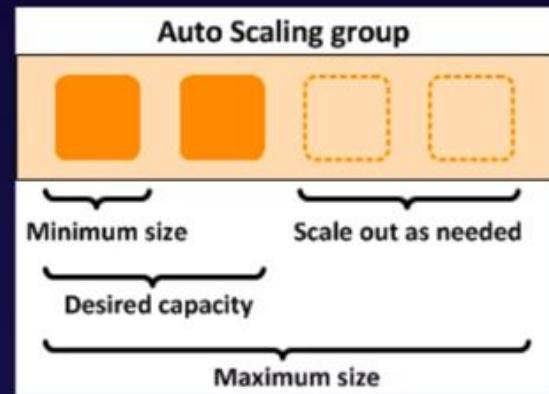
- Sticky Sessions enable your users to stick to the same EC2 instance.
Can be useful if you are storing information locally to that instance.
- Cross Zone Load Balancing enables you to load balance across multiple availability zones.
- Path patterns allow you to direct traffic to different EC2 instances based on the URL contained in the request.

Auto Scaling

Auto Scaling Has 3 Components

1 Groups

Logical component. Webserver group or Application group or Database group etc.



2 Configuration Templates

Groups uses a launch template or a launch configuration as a configuration template for its EC2 instances. You can specify information such as the AMI ID, instance type, key pair, security groups, and block device mapping for your instances.

3 Scaling Options

Scaling Options provides several ways for you to scale your Auto Scaling groups. For example, you can configure a group to scale based on the occurrence of specified conditions (dynamic scaling) or on a schedule.

Auto Scaling

What are my scaling options?

- Maintain current instance levels at all times
- Scale manually
- Scale based on a schedule
- Scale based on demand
- Use predictive scaling



Auto Scaling

Maintain current instance levels at all times

You can configure your Auto Scaling group to maintain a specified number of running instances at all times.

To maintain the current instance levels, Amazon EC2 Auto Scaling performs a periodic health check on running instances within an Auto Scaling group.

When Amazon EC2 Auto Scaling finds an unhealthy instance, it terminates that instance and launches a new one.



Auto Scaling

Scale manually

Manual scaling is the most basic way to scale your resources, where you specify only the change in the maximum, minimum, or desired capacity of your Auto Scaling group.

Amazon EC2 Auto Scaling manages the process of creating or terminating instances to maintain the updated capacity.



Auto Scaling

Scale based on a schedule

Scaling by schedule means that scaling actions are performed automatically as a function of time and date.

This is useful when you know exactly when to increase or decrease the number of instances in your group, simply because the need arises on a predictable schedule.



Auto Scaling

Scale based on demand

A more advanced way to scale your resources - using scaling policies - lets you define parameters that control the scaling process.

For example, let's say that you have a web application that currently runs on two instances and you want the CPU utilization of the Auto Scaling group to stay at around 50 percent when the load on the application changes. This method is useful for scaling in response to changing conditions, when you don't know when those conditions will change. You can set up Amazon EC2 Auto Scaling to respond for you. We will do this in the next lab.



Auto Scaling

Use predictive scaling

You can also use Amazon EC2 Auto Scaling in combination with AWS Auto Scaling to scale resources across multiple services.

AWS Auto Scaling can help you maintain optimal availability and performance by combining predictive scaling and dynamic scaling (proactive and reactive approaches, respectively) to scale your Amazon EC2 capacity faster.

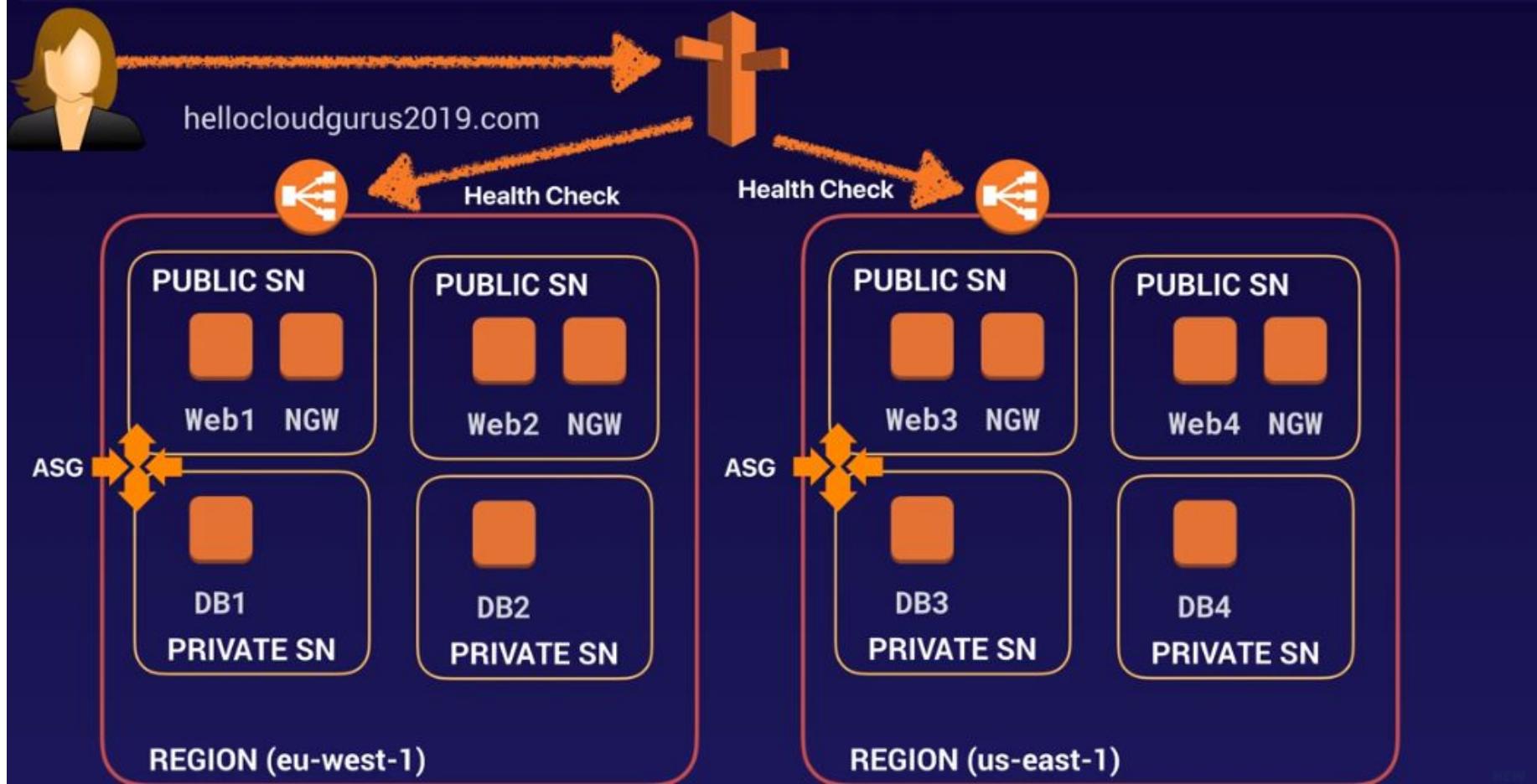


Exam Tips Auto Scaling

What are my scaling options?

- Maintain current instance levels at all times
- Scale manually
- Scale based on a schedule
- Scale based on demand
- Use predictive scaling

HA Architecture Sample



HA Architecture Sample

Scenario: You have a website that requires a minimum of 6 instances and it must be highly available. You must also be able to tolerate the failure of 1 Availability Zone. What is the ideal architecture for this environment while also being the most cost effective?

- 2 Availability Zones with 2 instances in each AZ.
- 3 Availability Zones with 3 instances in each AZ.
- 1 Availability Zone with 6 instances in each AZ.
- 3 Availability Zones with 2 instances in each AZ.

Exam Tips - HA Architecture

Remember the following

- Always Design for failure.
- Use Multiple AZ's and Multiple Regions where ever you can.
- Know the difference between Multi-AZ and Read Replicas for RDS.
- Know the difference between scaling out and scaling up.
- Read the question carefully and always consider the cost element.
- Know the different S3 storage classes.

Elastic Beanstalk



Create a web app

Create a new application and environment with a sample application or your own code. By creating an environment, you allow AWS Elastic Beanstalk to manage AWS resources and permissions on your behalf. [Learn more](#)

Application information

Application name

HelloCloudGurus

Up to 100 Unicode characters, not including forward slash (/).

Base configuration

Platform

PHP

Choose [Configure more options](#) for more platform configuration options.**Application code** Sample application

Get started right away with sample code.

 Upload your code

Upload a source bundle from your computer or copy one from Amazon S3.



Upload ZIP or WAR

[Cancel](#)[Configure more options](#)[Create application](#)

Elastic Beanstalk

All Applications > HelloCloudGurus > Hellocloudgurus-env (Environment ID: e-mkfij2mdwp)

Actions ▾



Creating Hellocloudgurus-env

This will take a few minutes..

10:58am Created EIP: 3.96.19.208

10:58am Created security group named:
awseb-e-mkfij2mdwp-stack-AWSEBSecurityGroup-57FB5L6QANKY

10:58am Environment health has transitioned to Pending. Initialization in progress (running for 5 seconds). There are no instances.

10:58am Using elasticbeanstalk-us-east-1-970845658441 as Amazon S3 storage bucket for environment data.

10:58am createEnvironment is starting.

Learn More

[Get started using Elastic Beanstalk](#)

[Modify the code](#)

[Create and connect to a database](#)

[Add a custom domain](#)

Featured

[Create your own custom platform](#)

Command Line Interface (v3)

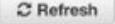
[Installing the AWS EB CLI](#)

[EB CLI Command Reference](#)

Elastic Beanstalk

All Applications > HelloCloudGurus > Hellocloudgurus-env (Environment ID: e-mkfj2mdwp, URL: Hellocloudgurus-env.q7yff7ji3t.us-east-1.elasticbeanstalk.com)

Actions ▾

Dashboard Overview 

Configuration Logs Health Monitoring Alarms Managed Updates Events Tags

Recent Events 

Time	Type	Details
2018-12-24 11:00:42 UTC+0000	INFO	Successfully launched environment: Hellocloudgurus-env
2018-12-24 11:00:42 UTC+0000	INFO	Application available at Hellocloudgurus-env.q7yff7ji3t.us-east-1.elasticbeanstalk.com .
2018-12-24 11:00:29 UTC+0000	INFO	Added instance [i-0465e2b9e2ce619f1] to your environment.
2018-12-24 11:00:29 UTC+0000	INFO	Environment health has transitioned from Pending to Ok. Initialization completed 15 seconds ago and took 2 minutes.
2018-12-24 11:00:03 UTC+0000	INFO	Waiting for EC2 instances to launch. This may take a few minutes.

Health Running Version Configuration

 Sample Application 

Ok Causes Upload and Deploy PHP 7.2 running on 64bit Amazon Linux/2.8.5 Change

Elastic Beanstalk

Congratulations!

Your AWS Elastic Beanstalk *PHP* application is now running on your own dedicated environment in the AWS Cloud

You are running  PHP version 7.2.11

What's Next?

- [AWS Elastic Beanstalk overview](#)
- [Deploying AWS Elastic Beanstalk Applications in PHP Using Eb and Git](#)
- [Using Amazon RDS with PHP](#)
- [Customizing the Software on EC2 Instances](#)
- [Customizing Environment Resources](#)

AWS SDK for PHP

- [AWS SDK for PHP home](#)
- [PHP developer center](#)
- [AWS SDK for PHP on GitHub](#)

Elastic Beanstalk

All Applications > HelloCloudGurus > Hellocloudgurus-env (Environment ID: e-mkfj2mdwp, URL: Hellocloudgurus-env.q7yffj3t.us-east-1.elasticbeanstalk.com)

Actions ▾

Dashboard

Configuration

Logs

Health

Monitoring

Alarms

Managed Updates

Events

Tags

Configuration overview

[Cancel](#)

[Review changes](#)

[Apply configuration](#)

Software

Rotate logs: disabled (default)
Log streaming: disabled (default)
Environment properties: 0

[Modify](#)

Instances

EC2 instance type: t2.micro
EC2 image ID: ami-02f14f1a619244d79
Monitoring interval: 5 minute
Root volume type: container default
Root volume size (GB): container default
Root volume IOPS: container default
Security groups: sg-0167cc761b7814b5c

[Modify](#)

Capacity

Environment type: single instance

[Modify](#)

Load balancer

This configuration does not contain a load balancer.

Rolling updates and deployments

Deployment policy: All at once
Rolling updates: disabled

[Modify](#)

Security

Service role: aws-elasticbeanstalk-service-role
Virtual machine key pair: --
Virtual machine instance profile: aws-elasticbeanstalk-ec2-role

[Modify](#)

Monitoring

Health reporting system: Enhanced
Ignore HTTP 4xx: disabled
Health event log streaming: disabled

Managed updates

Managed updates: disabled

Notifications

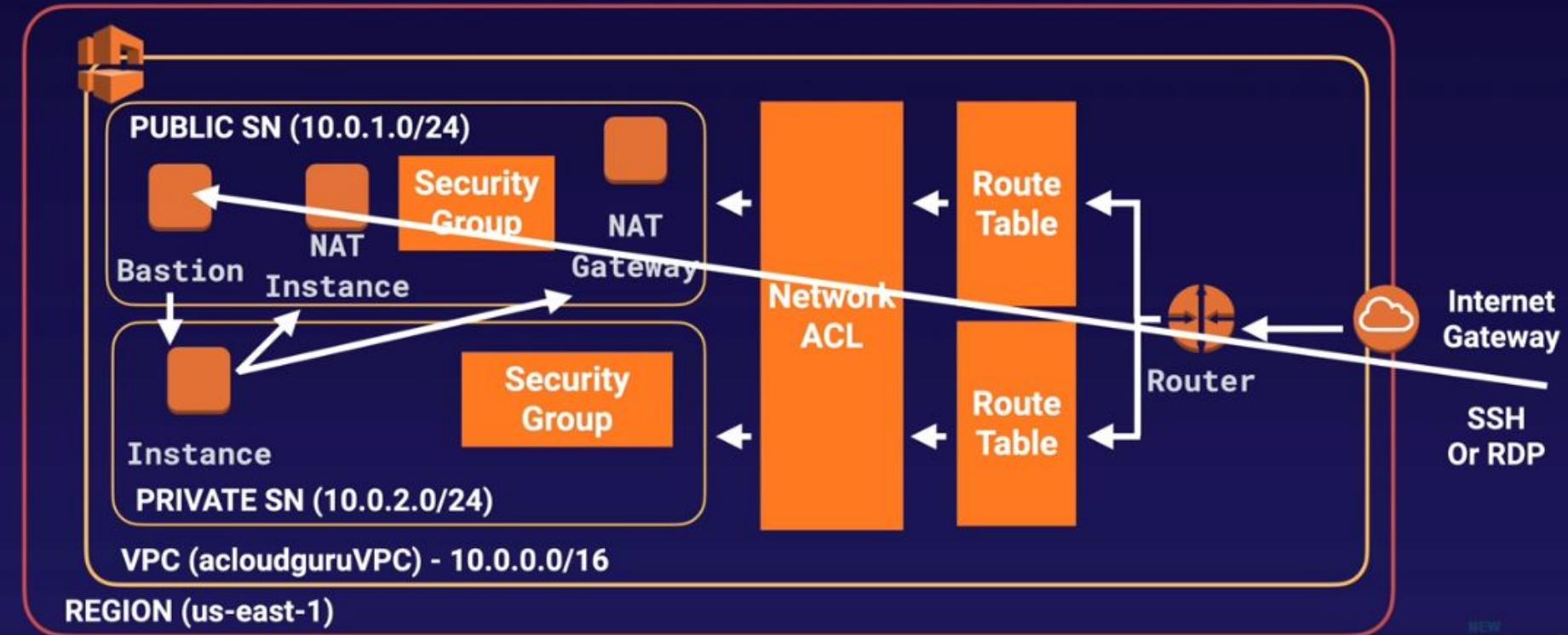
Email address: --

Elastic Beanstalk

With Elastic Beanstalk, you can quickly deploy and manage applications in the AWS Cloud without worrying about the infrastructure that runs those applications. You simply upload your application, and Elastic Beanstalk automatically handles the details of capacity provisioning, load balancing, scaling, and application health monitoring.

Bastions In Actions

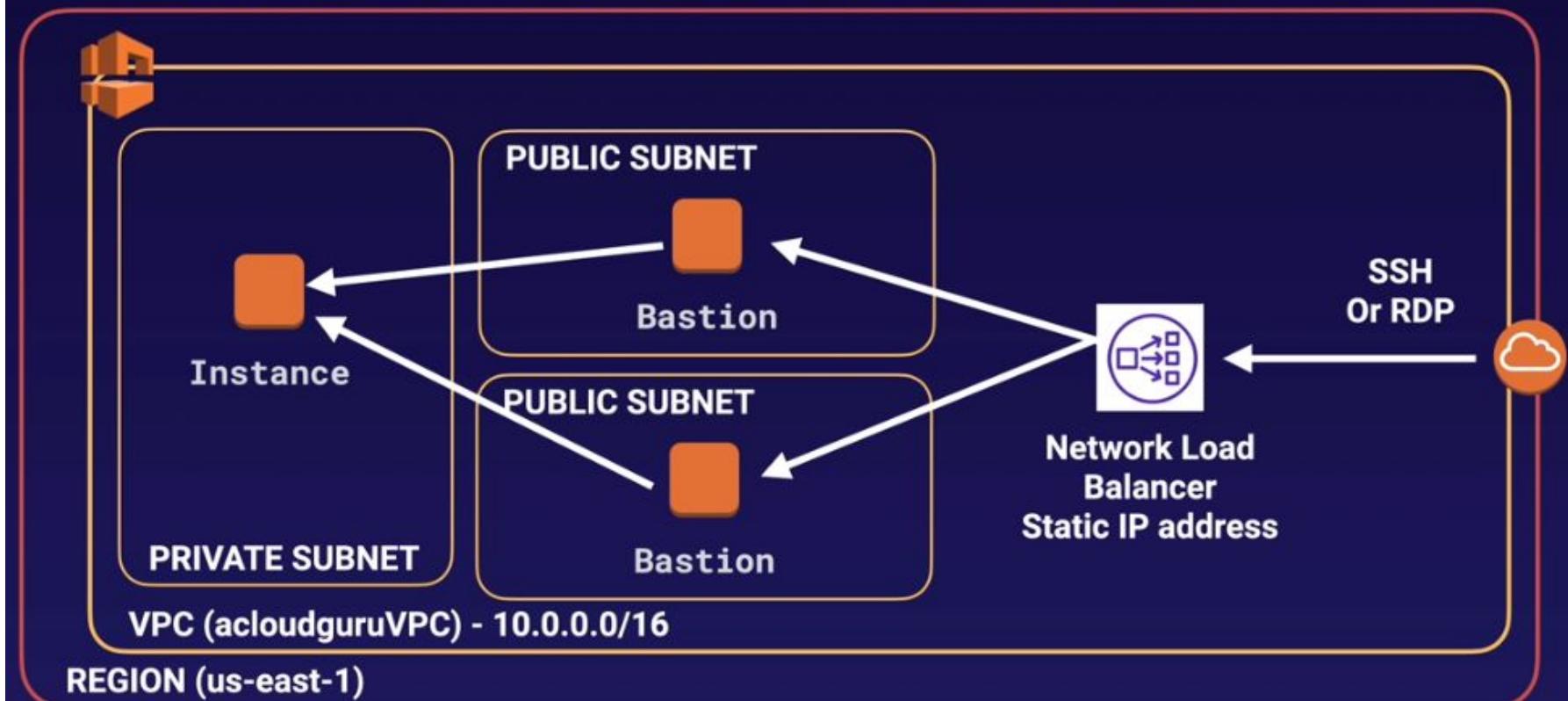
VPC with Public and Private Subnet(s)



Bastions In Actions

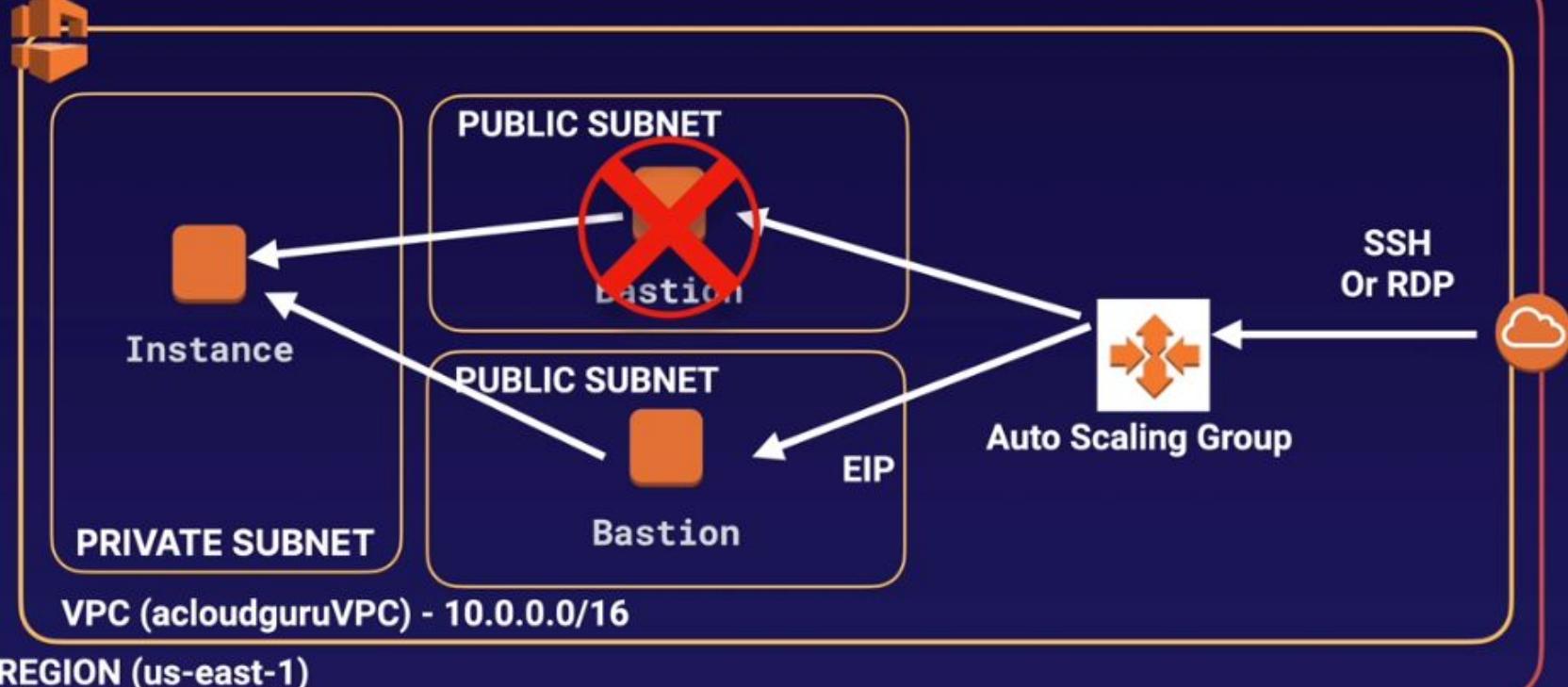
Scenario 1:

Two EC2 Instances, Two Availability Zones, Network Load Balancer



Bastions In Actions

Scenario 2:
One EC2 Instance, Two Availability Zones, Auto Scaling Group



Exam tips - Bastions In Actions

High Availability with Bastion Hosts

- Two hosts in two separate Availability Zones. Use a Network Load Balancer with static IP addresses and health checks to fail over from one host to the other.
- Can't use an Application Load Balancer, as it is layer 7 and you need to use layer 4.
- One host in one Availability Zone behind an Auto Scaling group with health checks and a fixed EIP. If the host fails, the health check will fail and the Auto Scaling group will provision a new EC2 instance in a separate Availability Zone. You can use a user data script to provision the same EIP to the new host. This is the cheapest option, but it is not 100% fault tolerant.

Bastions In Actions

You need to be aware of what high-level AWS services you can use on-premises for the exam:

- ✓ Database Migration Service (DMS)
- ✓ Server Migration Service (SMS)
- ✓ AWS Application Discovery Service
- ✓ VM Import/Export
- ✓ Download Amazon Linux 2 as an ISO

Bastions In Actions

Database Migration Service

- Allows you to move databases to and from AWS
- Might have your DR environment in AWS and your on-premises environment as your primary
- Works with most popular database technologies, such as Oracle, MySQL, DynamoDB, etc.

Supports **homogenous** migrations:



And supports **heterogeneous** migrations:



Bastions In Actions

Server Migration Service

- Server Migration Service supports incremental replication of your on-premises servers in to AWS.
- Can be used as a backup tool, multi-site strategy (on-premises and off-premises), and a DR tool.



Bastions In Actions

AWS Application Discovery Service does the following:

- AWS Application Discovery Service helps enterprise customers plan migration projects by gathering information about their on-premises data centers.
- You install the AWS Application Discovery Agentless Connector as a virtual appliance on VMware vCenter.
- It will then build a server utilization map and dependency map of your on-premises environment.
- The collected data is retained in encrypted format in an AWS Application Discovery Service data store. You can export this data as a CSV file and use it to estimate the Total Cost of Ownership (TCO) of running on AWS and to plan your migration to AWS.
- This data is also available in AWS Migration Hub, where you can migrate the discovered servers and track their progress as they get migrated to AWS.



Bastions In Actions

Using VM Import/Export

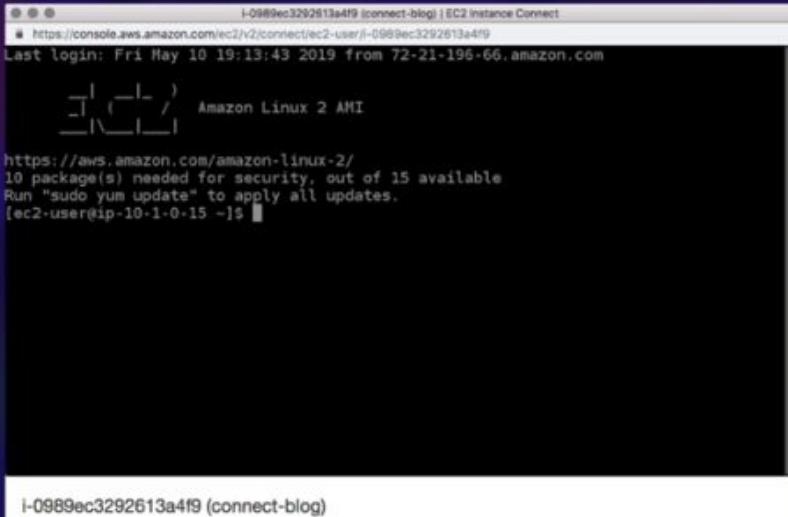
- Migrate existing applications in to EC2.
- Can be used to create a DR strategy on AWS or use AWS as a second site.
- You can also use it to export your AWS VMs to your on-premises data center.



Bastions In Actions

AWS allows you to download Amazon Linux 2 as an ISO:

- Works with all major virtualization providers, such as VMware, Hyper-V, KVM, VirtualBox (Oracle), etc.



The screenshot shows a terminal window titled "i-0989ec3292613a4f9 (connect-blog) | EC2 Instance Connect". The window displays a command-line interface for an Amazon Linux 2 AMI. The user has run the command "sudo yum update" and is awaiting further input at the prompt "[ec2-user@ip-10-1-0-15 ~]\$". The terminal also shows the user's last login information: "Last login: Fri May 10 19:13:43 2019 from 72-21-196-66.amazon.com". Below the terminal, the URL "https://aws.amazon.com/amazon-linux-2/" and a message about security packages are visible.

```
i-0989ec3292613a4f9 (connect-blog) | EC2 Instance Connect
https://console.aws.amazon.com/ec2/v2/connect/ec2-user/i-0989ec3292613a4f9
Last login: Fri May 10 19:13:43 2019 from 72-21-196-66.amazon.com
[ec2-user@ip-10-1-0-15 ~]$ https://aws.amazon.com/amazon-linux-2/
10 package(s) needed for security, out of 15 available
Run "sudo yum update" to apply all updates.
[ec2-user@ip-10-1-0-15 ~]$
```

Bastions In Actions

You need to be aware of what high-level AWS services you can use on-premises for the exam:

- ✓ Database Migration Service (DMS)
- ✓ Server Migration Service (SMS)
- ✓ AWS Application Discovery Service
- ✓ VM Import/Export
- ✓ Download Amazon Linux 2 as an ISO

CHAPTER 10

Applications

SQS

Amazon SQS is a web service that gives you access to a message queue that can be used to store messages while waiting for a computer to process them.

It's a distributed queue system that enables web service applications to quickly and reliably queue messages that one component in the application generates to be consumed by another component.

A queue is a temporary repository for messages that are awaiting processing.



SQS

Using Amazon SQS, you can decouple the components of an application so they run independently, easing message management between components.

Any component of a distributed application can store messages in a fail-safe queue.

Messages can contain up to 256 KB of text in any format.

Any component can later retrieve the messages programmatically using the Amazon SQS API.



SQS

The queue acts as a buffer between the component producing and saving data, and the component receiving the data for processing.

This means the queue resolves issues that arise if the producer is producing work faster than the consumer can process it, or if the producer or consumer are only intermittently connected to the network.



SQS

There are two types of queue:

- Standard queues (default)
- FIFO queues

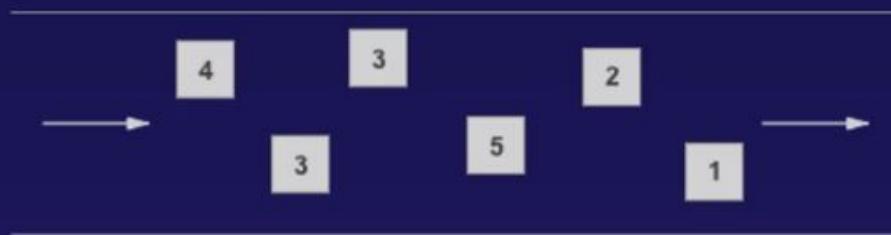


SQS

Amazon SQS offers standard as the default queue type. A standard queue lets you have a **nearly-unlimited number of transactions per second**. Standard queues guarantee that a message is delivered at least once.

Occasionally (because of the highly-distributed architecture that allows high throughput), more than one copy of a message might be delivered out of order.

However, standard queues provide best-effort ordering which ensures that messages are generally delivered in the same order as they are sent.



SQS

The FIFO queue complements the standard queue.

The most important features of this queue type are **FIFO (first-in-first-out) delivery** and **exactly-once processing**: the order in which messages are sent and received is strictly preserved and a message is delivered once and remains available until a consumer processes and deletes it; duplicates are not introduced into the queue.



SQS

FIFO queues also support message groups that allow multiple ordered message groups within a single queue.

FIFO queues are limited to 300 transactions per second (TPS), but have all the capabilities of standard queues.



SQS

SQS Exam Tips

- SQS is pull-based, not pushed-based.
- Messages are 256 KB in size.
- Messages can be kept in the queue from 1 minute to 14 days; the default retention period is 4 days.

SQS

SQS Exam Tips

- SQS guarantees that your messages will be processed at least once.
- Amazon SQS long polling is a way to retrieve messages from your Amazon SQS queues. While the regular short polling returns immediately (even if the message queue being polled is empty), long polling doesn't return a response until a message arrives in the message queue, or the long poll times out.
- Any time you see a scenario based question about “decoupling” your infrastructure - think SQS.

Simple Workflow Service

Amazon Simple Workflow Service (Amazon SWF) is a web service that makes it easy to coordinate work across distributed application components. SWF enables applications for a range of use cases, including media processing, web application back-ends, business process workflows, and analytics pipelines, to be designed as a coordination of tasks.



Simple Workflow Service

Tasks represent invocations of various processing steps in an application which can be performed by executable code, web service calls, human actions, and scripts.



Simple Workflow Service

SWF vs SQS

- SQS has a retention period of up to 14 days; with SWF, workflow executions can last up to 1 year.
- Amazon SWF presents a task-oriented API, whereas Amazon SQS offers a message-oriented API.
- Amazon SWF ensures that a task is assigned only once and is never duplicated. With Amazon SQS, you need to handle duplicated messages and may also need to ensure that a message is processed only once.
- Amazon SWF keeps track of all the tasks and events in an application. With Amazon SQS, you need to implement your own application-level tracking, especially if your application uses multiple queues.

Simple Workflow Service

SWF Actors

- Workflow Starters — An application that can initiate (start) a workflow. Could be your e-commerce website following the placement of an order, or a mobile app searching for bus times.
- Deciders — Control the flow of activity tasks in a workflow execution. If something has finished (or failed) in a workflow, a Decider decides what to do next.
- Activity Workers — Carry out the activity tasks.

Simple Notification Service

Amazon Simple Notification Service (Amazon SNS) is a web service that makes it easy to set up, operate, and send notifications from the cloud.

It provides developers with a highly scalable, flexible, and cost-effective capability to publish messages from an application and immediately deliver them to subscribers or other applications.



Simple Notification Service

Push notifications to Apple, Google, Fire OS, and Windows devices, as well as Android devices in China with Baidu Cloud Push.



Google



SQS Integration

Besides pushing cloud notifications directly to mobile devices, Amazon SNS can also deliver notifications by SMS text message or email to Amazon Simple Queue Service (SQS) queues, or to any HTTP endpoint.



SNS Topics

SNS allows you to group multiple recipients using topics. A topic is an “access point” for allowing recipients to dynamically subscribe for identical copies of the same notification.

One topic can support deliveries to multiple endpoint types — for example, you can group together iOS, Android and SMS recipients. When you publish once to a topic, SNS delivers appropriately formatted copies of your message to each subscriber.



Simple Notification Service Availability

To prevent messages from being lost, all messages published to Amazon SNS are stored redundantly across multiple availability zones.



Simple Notification Service

SNS Benefits

- Instantaneous, push-based delivery (no polling)
- Simple APIs and easy integration with applications
- Flexible message delivery over multiple transport protocols
- Inexpensive, pay-as-you-go model with no up-front costs
- Web-based AWS Management Console offers the simplicity of a point-and-click interface

Simple Notification Service

SNS vs SQS?

- Both Messaging Services in AWS
- SNS - Push
- SQS - Polls (Pulls)

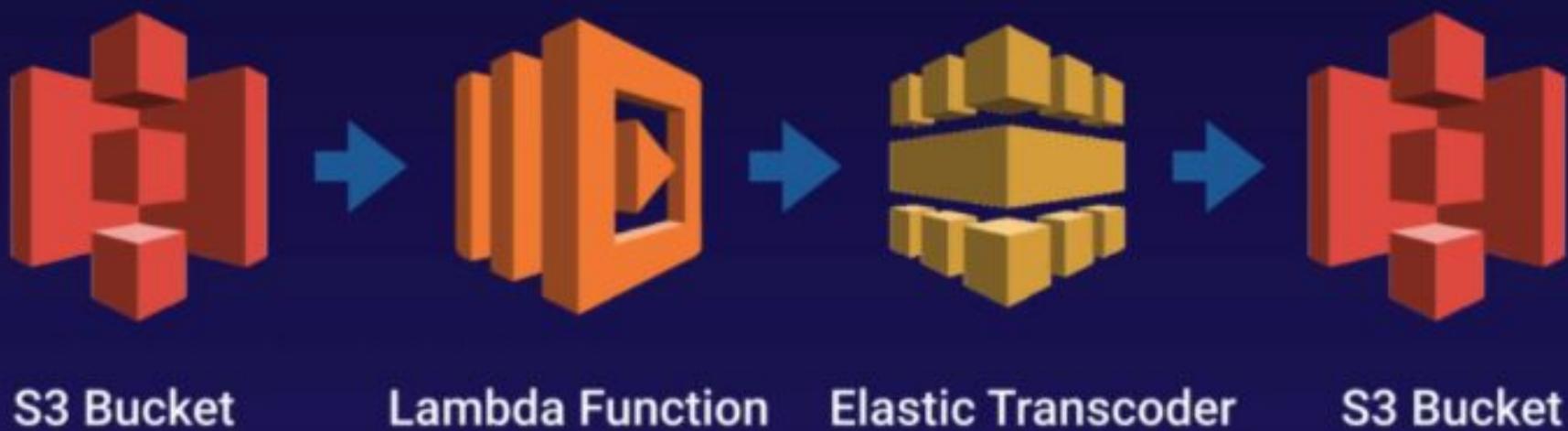
AWS Elastic Transcoder

What Is Elastic Transcoder?

- Media Transcoder in the cloud.
- Convert media files from their original source format in to different formats that will play on smartphones, tablets, PCs, etc.
- Provides transcoding presets for popular output formats, which means that you don't need to guess about which settings work best on particular devices.
- Pay based on the minutes that you transcode and the resolution at which you transcode.

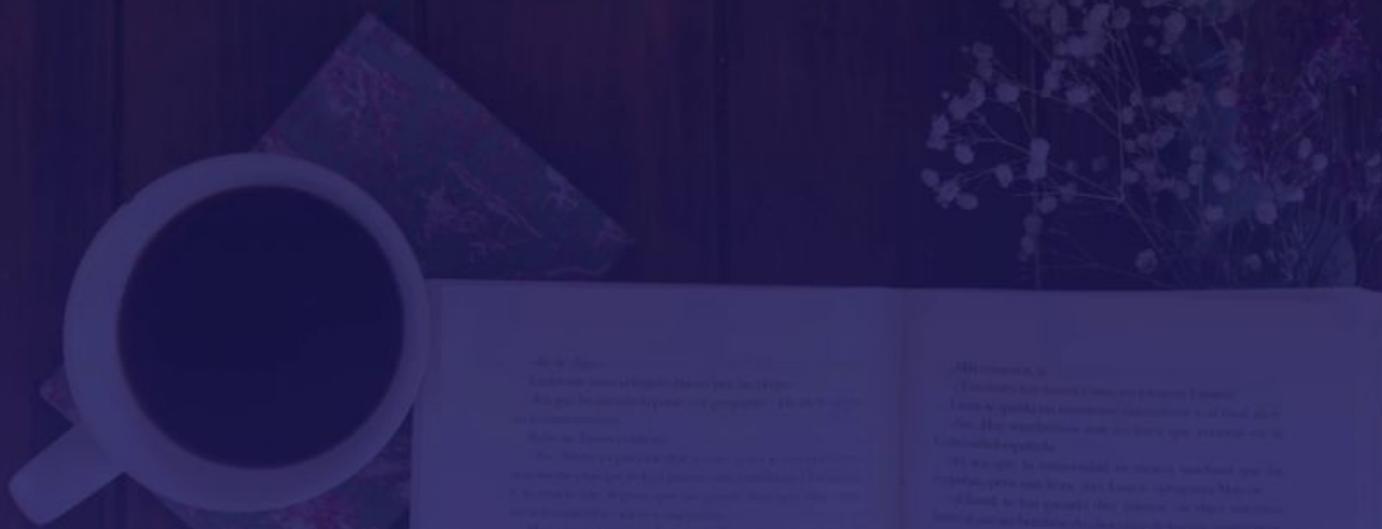
AWS Elastic Transcoder

How We Use Elastic Transcoder



AWS Single Sign-On

Just remember that Elastic Transcoder is a media transcoder in the cloud. It converts media files from their original source format in to different formats that will play on smartphones, tablets, PCs, etc.



AWS Streams

Streaming Data is data that is generated continuously by thousands of data sources, which typically send in the data records simultaneously, and in small sizes (order of Kilobytes.)

- Purchases from online stores (think amazon.com)
- Stock Prices
- Game data (as the gamer plays)
- Social network data
- Geospatial data (think uber.com)
- IoT sensor data



AWS Kinesis

Amazon Kinesis is a platform on AWS to send your streaming data to. Kinesis makes it easy to load and analyze streaming data, and also providing the ability for you to build your own custom applications for your business needs.



AWS Kinesis

3 Different Types of Kinesis

- Kinesis Streams



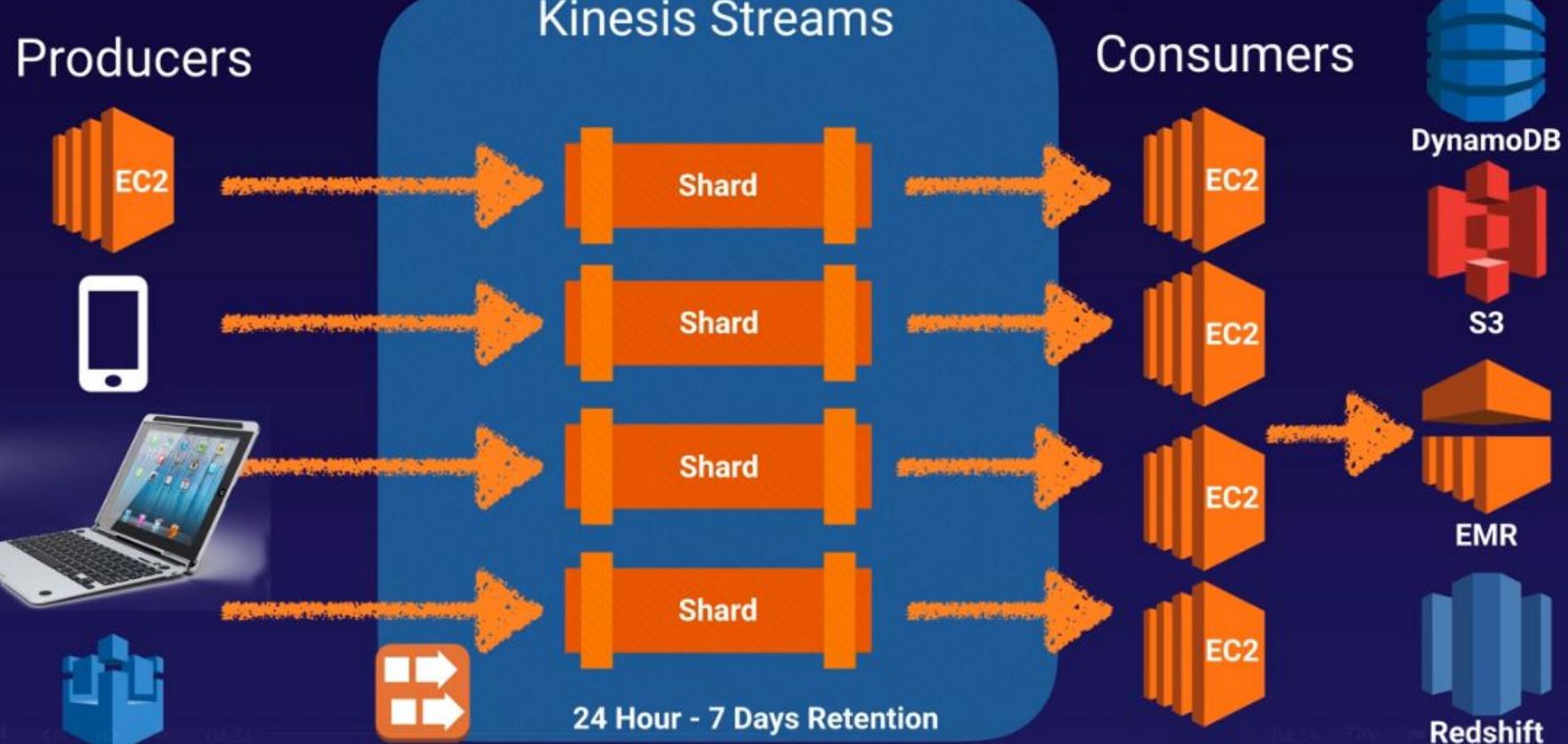
- Kinesis Firehose



- Kinesis Analytics



AWS Kinesis



AWS Kinesis

Kinesis Streams Consist Of Shards;

- 5 transactions per second for reads, up to a maximum total data read rate of 2 MB per second and up to 1,000 records per second for writes, up to a maximum total data write rate of 1 MB per second (including partition keys.)
- The data capacity of your stream is a function of the number of shards that you specify for the stream. The total capacity of the stream is the sum of the capacities of its shards.



AWS Kinesis

Kinesis Exam Tips

- Know the difference between Kinesis Streams and Kinesis Firehose. You will be given scenario questions and you must choose the most relevant service.
- Understand what Kinesis Analytics is.

CHAPTER 11

Security

Reducing Security Threats

Bad Actors

- Typically automated processes
- Content scrapers
- Bad bots
- Fake user agent
- Denial of service (DoS)



Reducing Security Threats

Benefits of Preventing Bad Actors

- Reduce security threats
- Lower overall costs

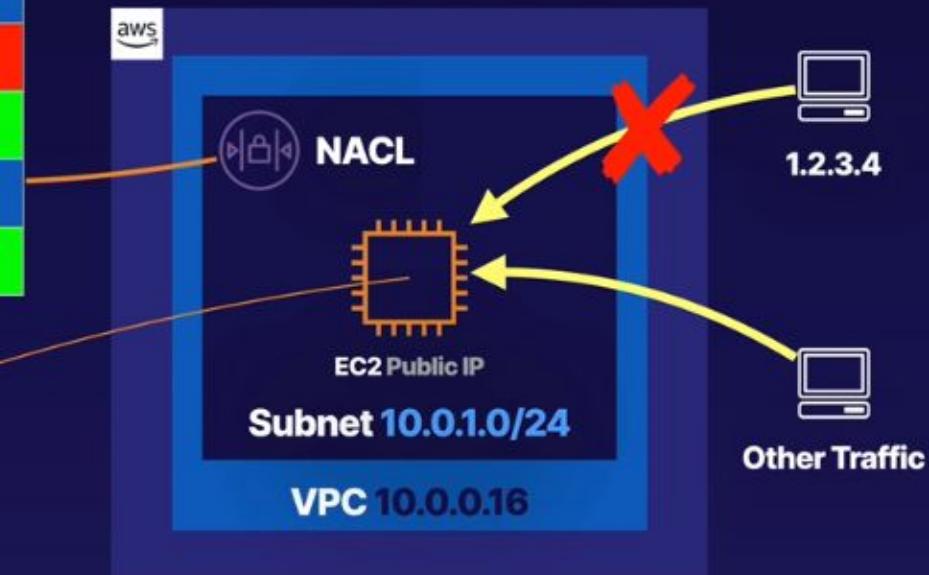


Network Access Control (NACL)

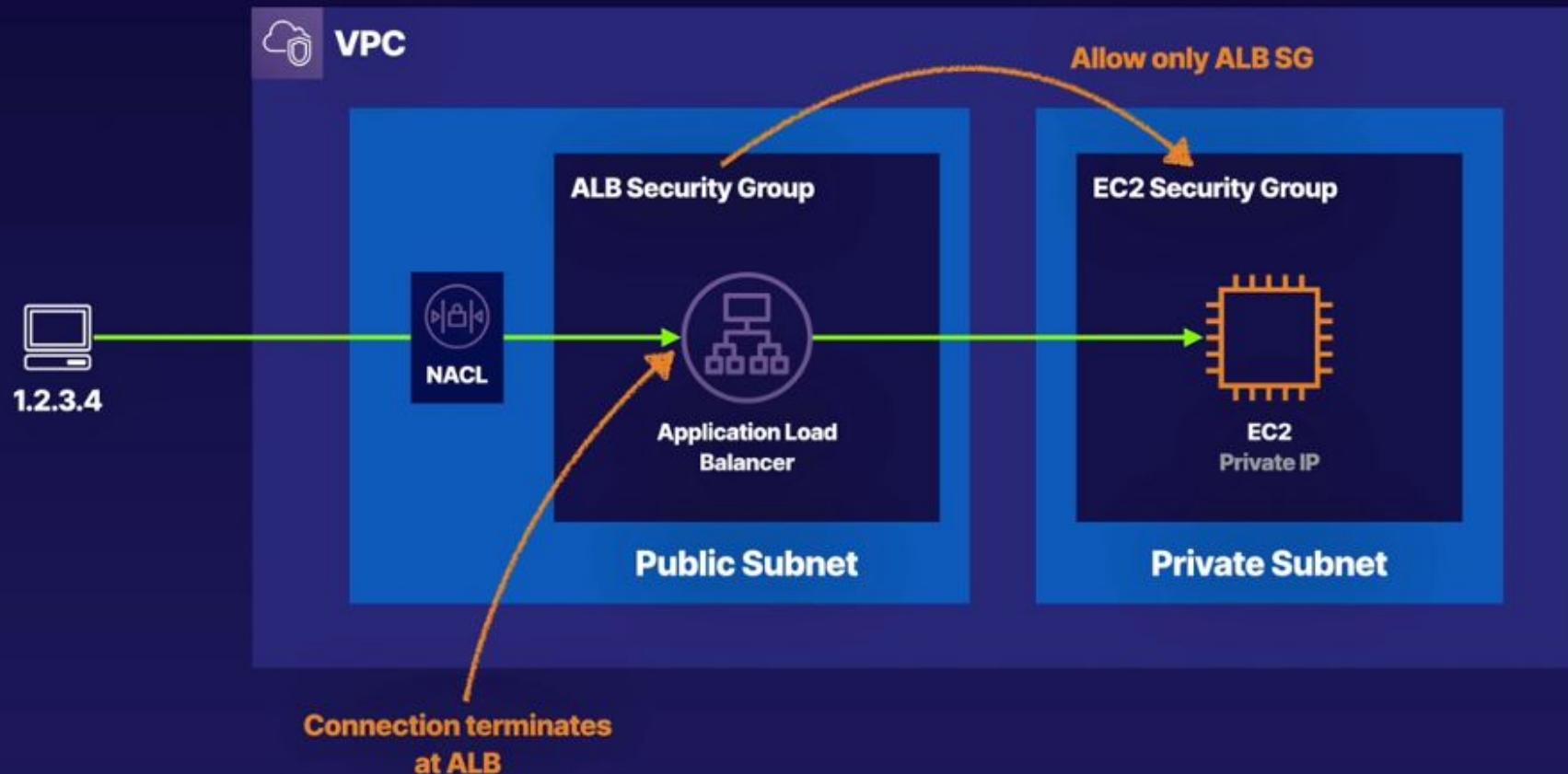
Inbound			
100	ALL Traffic	1.2.3.4	DENY
*	ALL Traffic	0.0.0.0/0	ALLOW
Outbound			
*	ALL Traffic	0.0.0.0/0	ALLOW

NACL Rules

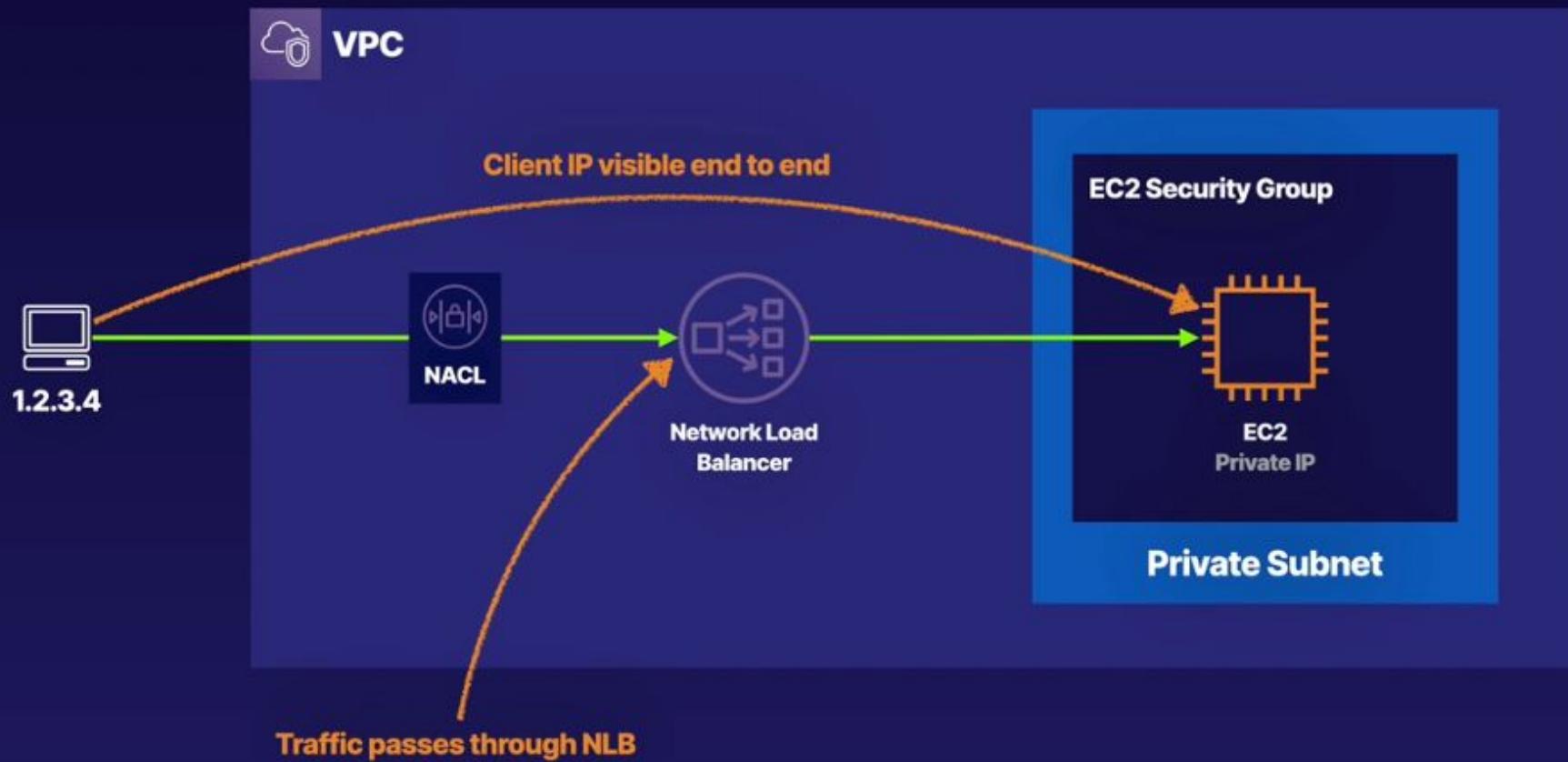
Host-based firewall
e.g., firewalld, iptables, ufw, Windows Firewall



Application Load Balancer (ALB)



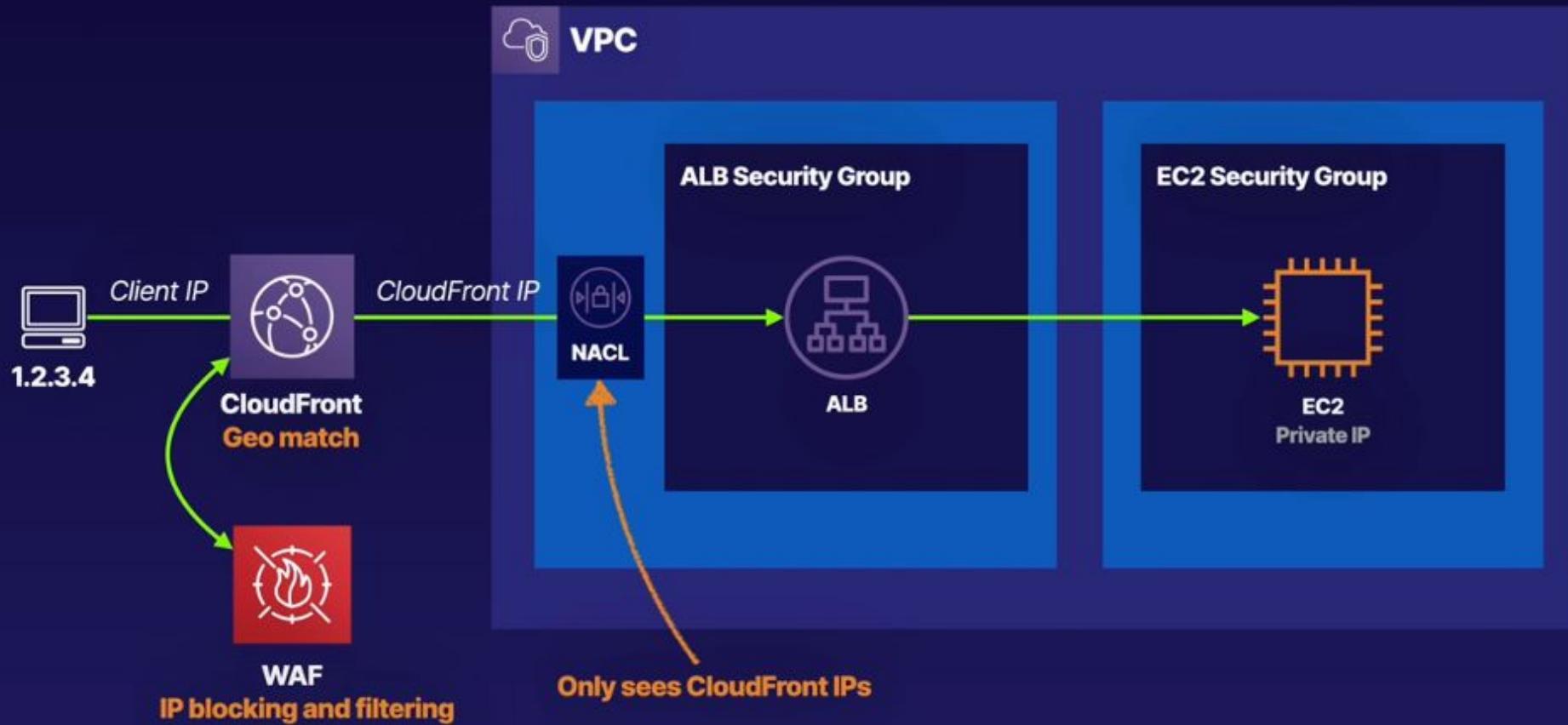
Network Load Balancer (NLB)



Web Application Firewall (WAF)



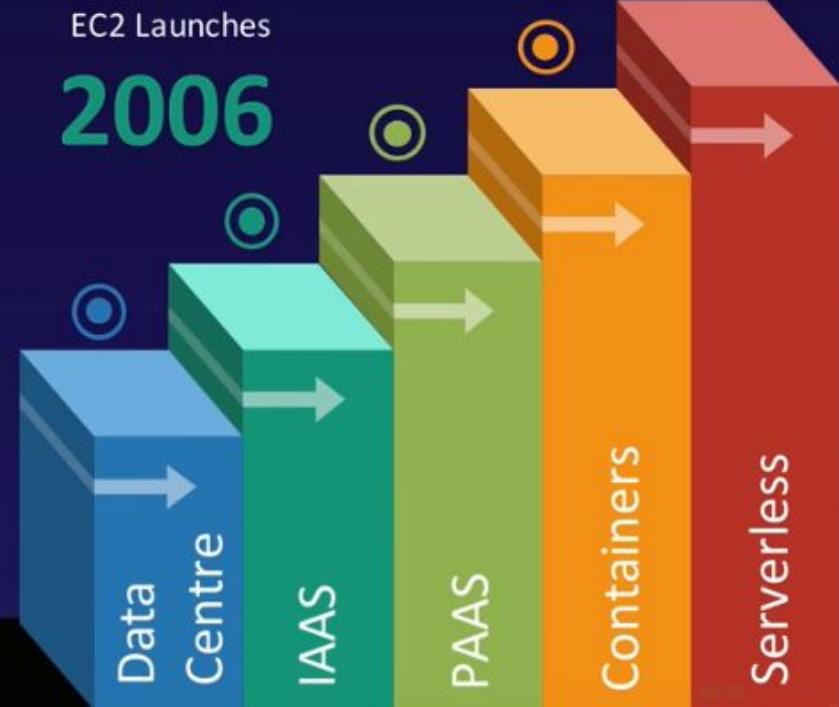
WAF + CloudFront



CHAPTER 12

Serverless

History



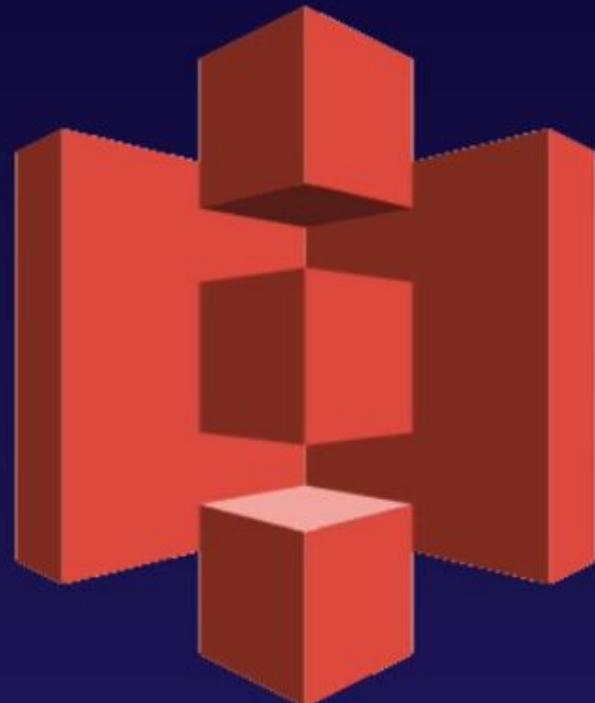
History

Lambda Is The Ultimate Abstraction Layer;

- Data Centres
- Hardware
- Assembly Code/Protocols
- High Level Languages
- Operating Systems
- Application Layer/AWS APIs
- AWS Lambda

What is Lambda?

AWS Lambda is a compute service where you can upload your code and create a Lambda function. AWS Lambda takes care of provisioning and managing the servers that you use to run the code. You don't have to worry about operating systems, patching, scaling, etc.

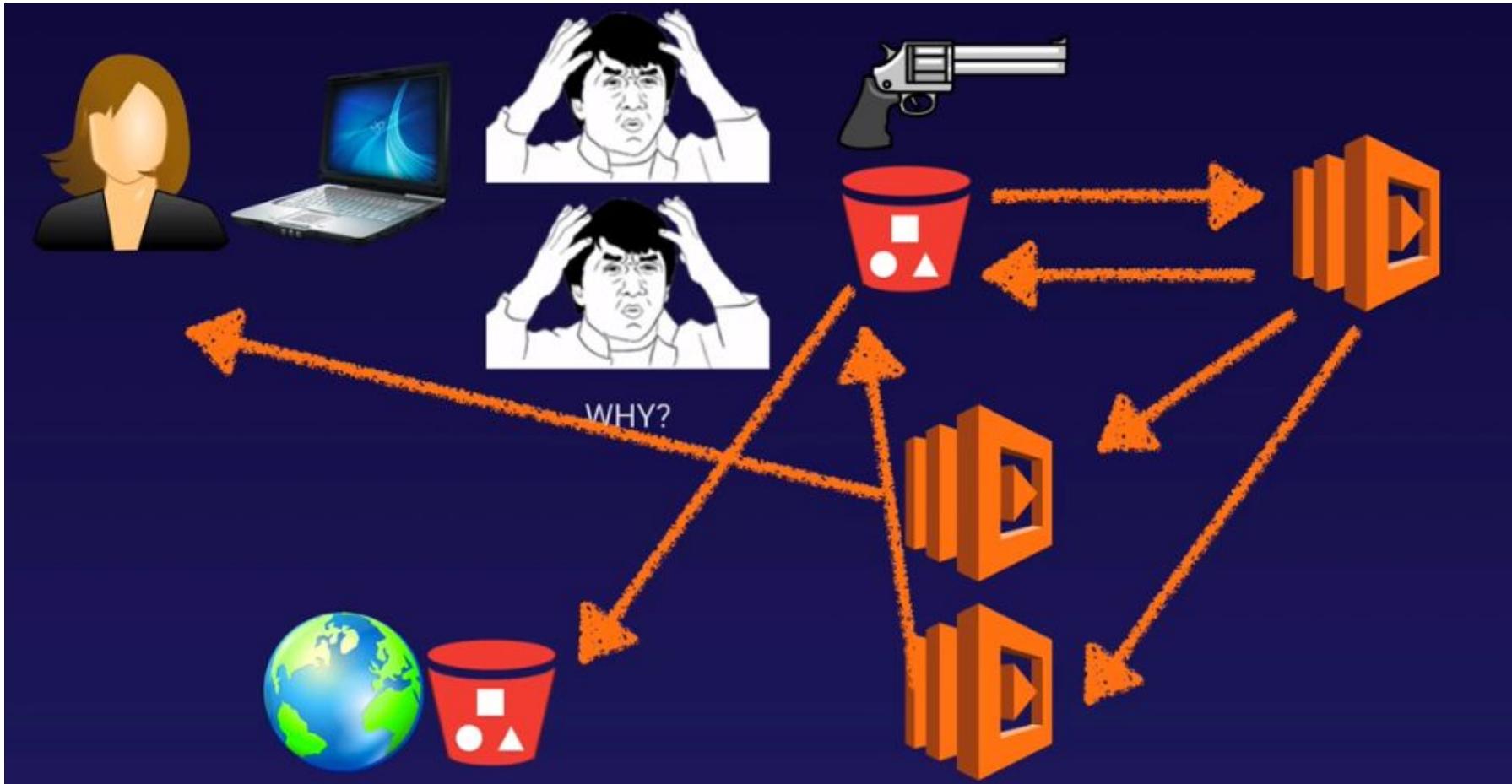


Lambda - the basics

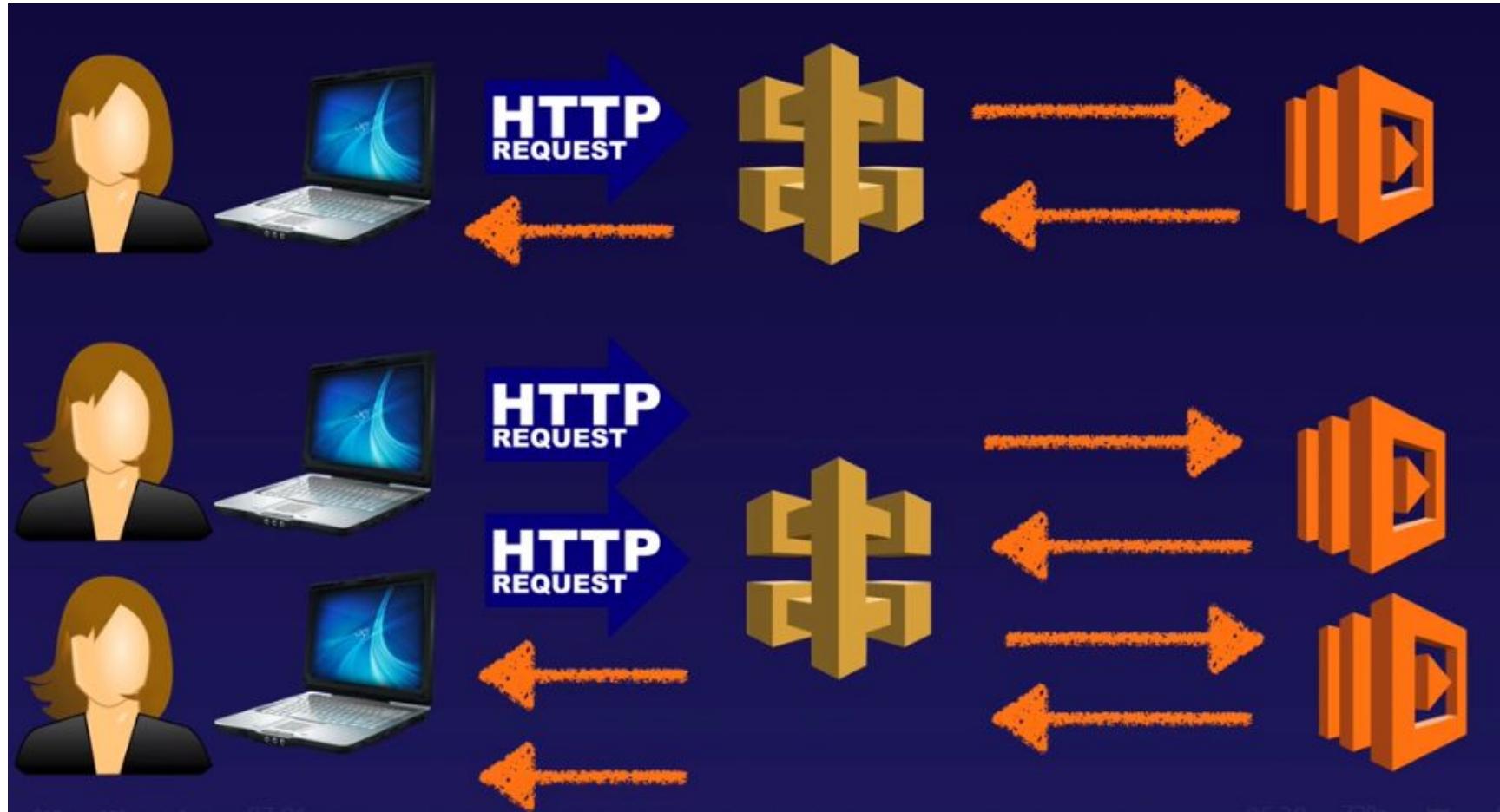
You can use Lambda in the following ways;

- As an event-driven compute service where AWS Lambda runs your code in response to events. These events could be changes to data in an Amazon S3 bucket or an Amazon DynamoDB table.
- As a compute service to run your code in response to HTTP requests using Amazon API Gateway or API calls made using AWS SDKs.

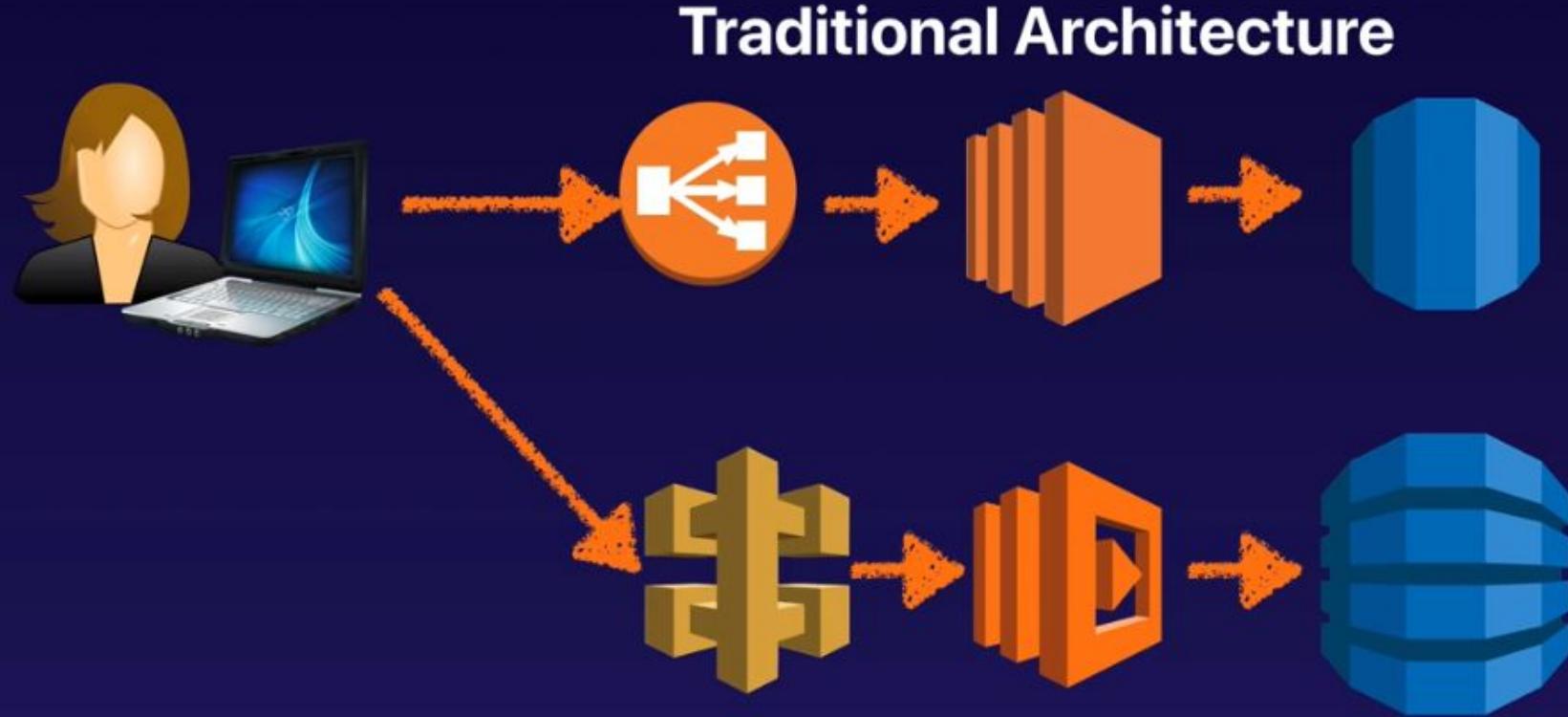
What is Lambda?



What is Lambda?



What is Lambda?



What is Lambda?

How Is Lambda Priced

1 Number Of Requests

First 1 million requests are free. \$0.20 per 1 million requests thereafter.

2 Duration

Duration is calculated from the time your code begins executing until it returns or otherwise terminates, rounded up to the nearest 100ms. The price depends on the amount of memory you allocate to your function. You are charged \$0.00001667 for every GB-second used.



What is Lambda?

Why Is Lambda Cool?

- NO SERVERS!
- Continuous Scaling
- Super super super cheap!



What is Lambda?

Lambda Exam Tips

- Lambda scales out (not up) automatically
- Lambda functions are independent, 1 event = 1 function
- Lambda is serverless
- Know what services are serverless!

What is Lambda?

Lambda Exam Tips

- Architectures can get extremely complicated, AWS X-ray allows you to debug what is happening
- Lambda can do things globally, you can use it to back up S3 buckets to other S3 buckets etc
- Know your triggers

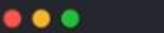
Serverless Application Model (SAM)

What is SAM?

- CloudFormation extension optimized for serverless applications
- New types: functions, APIs, tables
- Supports anything CloudFormation supports
- Run serverless applications locally
- Package and deploy using CodeDeploy



Anatomy of a SAM template



```
1 AWSTemplateFormatVersion: '2010-09-09'          1
2 Transform: AWS::Serverless-2016-10-31           1
3 Description: Hello World SAM Template           1
4
5 Globals:
6   Function:                                     2
7     Timeout: 3
8
9 Resources:
10 HelloWorldFunction:                           3
11   Type: AWS::Serverless::Function
12   Properties:
13     CodeUri: hello_world/
14     Handler: app.lambda_handler
15     Runtime: python3.8
16   Events:
17     HelloWorld:
18       Type: Api
19       Properties:
20         Path: /hello
21         Method: get
22
23 Outputs:
24 HelloWorldApi:                                4
25   Description: "API Gateway endpoint URL for Prod stage for Hello World function"
26   Value: !Sub "https://${ServerlessRestApi}.execute-api.${AWS::Region}.amazonaws.com/Prod/hello/"
27 HelloWorldFunction:
28   Description: "Hello World Lambda Function ARN"
29   Value: !GetAtt HelloWorldFunction.Arn
30 HelloWorldFunctionIamRole:
31   Description: "Implicit IAM Role created for Hello World function"
32   Value: !GetAtt HelloWorldFunctionRole.Arn
```

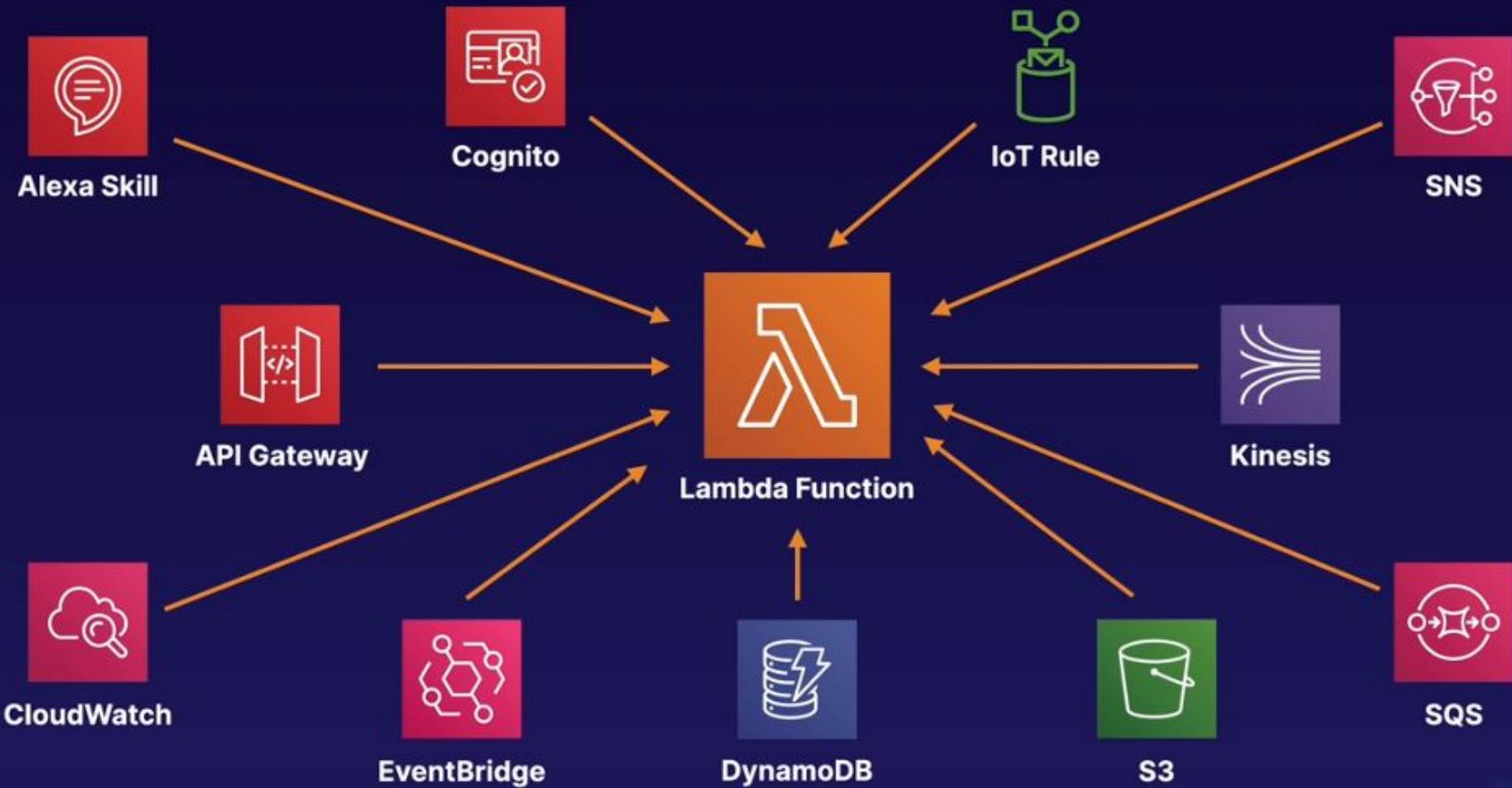
1 Tells CloudFormation this is a SAM template

2 Applies the same properties to all functions

3 Creates a Lambda function from local code. Also creates an API Gateway endpoint, mappings, and permissions.

4 Outputs relevant information

Lambda functions event sources



What is SAM?

```
Which runtime would you like to use?  
1 - nodejs12.x  
2 - python3.8  
3 - ruby2.7  
4 - go1.x  
5 - java11  
6 - dotnetcore3.1  
7 - nodejs10.x  
8 - python3.7  
9 - python3.6  
10 - python2.7  
11 - ruby2.5  
12 - java8  
13 - dotnetcore2.1  
  
Runtime: 8  
  
Project name [sam-app]:  
  
Cloning app templates from https://github.com/awslabs/aws-sam-cli-app-templates.git  
  
AWS quick start application templates:  
1 - Hello World Example  
2 - EventBridge Hello World  
3 - EventBridge App from scratch (100+ Event Schemas)  
4 - Step Functions Sample App (Stock Trader)  
Template selection: 1  
  
-----  
Generating application:  
-----  
Name: sam-app  
Runtime: python3.7  
Dependency Manager: pip  
Application Template: hello-world  
Output Directory: .  
  
Next steps can be found in the README file at ./sam-app/README.md
```

```
[ec2-user@ip-172-31-69-52 ~]$ ll  
total 0  
drwxrwxr-x 5 ec2-user ec2-user 108 Jun  1 14:00 sam-app  
[ec2-user@ip-172-31-69-52 ~]$ cd sam-app/  
[ec2-user@ip-172-31-69-52 sam-app]$ ll  
total 12  
drwxrwxr-x 2 ec2-user ec2-user   24 Jun  1 14:00 events  
drwxrwxr-x 2 ec2-user ec2-user   63 Jun  1 14:00 hello_world  
-rw-rw-r-- 1 ec2-user ec2-user 7490 Jun  1 14:00 README.md  
-rw-rw-r-- 1 ec2-user ec2-user 1623 Jun  1 14:00 template.yaml  
drwxrwxr-x 3 ec2-user ec2-user   18 Jun  1 14:00 tests  
[ec2-user@ip-172-31-69-52 sam-app]$ █
```

What is SAM?

```
AWSTemplateFormatVersion: '2010-09-09'
Transform: AWS::Serverless-2016-10-31
Description: >
    sam-app

Sample SAM Template for sam-app

# More info about Globals: https://github.com/awslabs/serverless-application-model/blob/master/docs/globals.rst
Globals:
    Function:
        Timeout: 3

Resources:
    HelloWorldFunction:
        Type: AWS::Serverless::Function # More info about Function Resource: https://github.com/awslabs/serverless-application-model/blob/master/versions/2016-10-31.md#awsserverlessfunction
        Properties:
            CodeUri: hello_world/
            Handler: app.lambda_handler
            Runtime: python3.7
        Events:
            HelloWorld:
                Type: Api # More info about API Event Source: https://github.com/awslabs/serverless-application-model/blob/master/versions/2016-10-31.md#api
                Properties:
                    Path: /hello
                    Method: get

Outputs:
    # ServerlessRestApi is an implicit API created out of Events key under Serverless::Function
    # Find out more about other implicit resources you can reference within SAM
    # https://github.com/awslabs/serverless-application-model/blob/master/docs/internals/generated_resources.rst#api
    HelloWorldApi:
        Description: "API Gateway endpoint URL for Prod stage for Hello World function"
        Value: !Sub "https://${ServerlessRestApi}.execute-api.${AWS::Region}.amazonaws.com/Prod/hello/"
    HelloWorldFunction:
        Description: "Hello World Lambda Function ARN"
        Value: !GetAtt HelloWorldFunction.Arn
```

What is SAM?

```
import json

import requests

def lambda_handler(event, context):
    """Sample pure Lambda function

    Parameters
    -----
    event: dict, required
        API Gateway Lambda Proxy Input Format

        Event doc: https://docs.aws.amazon.com/apigateway/latest/developerguide/set-up-lambda-proxy-integrations.html#api-gateway-simple-proxy-for-lambda-input-format

    context: object, required
        Lambda Context runtime methods and attributes

        Context doc: https://docs.aws.amazon.com/lambda/latest/dg/python-context-object.html

    Returns
    -----
    API Gateway Lambda Proxy Output Format: dict

        Return doc: https://docs.aws.amazon.com/apigateway/latest/developerguide/set-up-lambda-proxy-integrations.html
    """

    try:
        # ip = requests.get("http://checkip.amazonaws.com/")
        # except requests.RequestException as e:
        #     # Send some context about this error to Lambda Logs
        #     print(e)

        #     raise e

    return {
        "statusCode": 200,
        "app.py" 42L, 1151C 1x 05:09 ━━━━━━ 04:12 720p NEW 28,5 Top
```

What is SAM?

```
event: dict, required
    API Gateway Lambda Proxy Input Format

    Event doc: https://docs.aws.amazon.com/apigateway/latest/developerguide/set-up-lambda-proxy-integrations.html#api-gateway-simple-proxy-for-lambda-input-format

context: object, required
    Lambda Context runtime methods and attributes

    Context doc: https://docs.aws.amazon.com/lambda/latest/dg/python-context-object.html

Returns
-----
API Gateway Lambda Proxy Output Format: dict

    Return doc: https://docs.aws.amazon.com/apigateway/latest/developerguide/set-up-lambda-proxy-integrations.html#api-gateway-simple-proxy-for-lambda-output-format
"""

try:
    ip = requests.get("http://checkip.amazonaws.com/")
except requests.RequestException as e:
    # Send some context about this error to Lambda Logs
    print(e)

    raise e

return {
    "statusCode": 200,
    "body": json.dumps({
        "message": "hello world",
        "location": ip.text.replace("\n", "")
    }),
}
```

What is SAM?

```
[ec2-user@ip-172-31-69-52 sam-app]$ ll
total 12
drwxrwxr-x 2 ec2-user ec2-user 24 Jun 1 14:00 events
drwxrwxr-x 2 ec2-user ec2-user 63 Jun 1 14:03 hello_world
-rw-rw-r-- 1 ec2-user ec2-user 7490 Jun 1 14:00 README.md
-rw-rw-r-- 1 ec2-user ec2-user 1623 Jun 1 14:00 template.yaml
drwxrwxr-x 3 ec2-user ec2-user 18 Jun 1 14:00 tests
[ec2-user@ip-172-31-69-52 sam-app]$ sam build
Building function 'HelloWorldFunction'
Running PythonPipBuilder:ResolveDependencies
Running PythonPipBuilder:CopySource

Build Succeeded

Built Artifacts : .aws-sam/build
Built Template : .aws-sam/build/template.yaml

Commands you can use next
=====
[*] Invoke Function: sam local invoke
[*] Deploy: sam deploy --guided

[ec2-user@ip-172-31-69-52 sam-app]$ █
```

What is SAM?

```
[ec2-user@ip-172-31-69-52 sam-app]$ sam deploy --guided

Configuring SAM deploy
=====
Looking for samconfig.toml : Not found

Setting default arguments for 'sam deploy'
=====
Stack Name [sam-app]:
AWS Region [us-east-1]:
#Shows you resources changes to be deployed and require a 'Y' to initiate deploy
Confirm changes before deploy [y/N]:
#SAM needs permission to be able to create roles to connect to the resources in your template
Allow SAM CLI IAM role creation [Y/n]:
HelloWorldFunction may not have authorization defined, Is this okay? [y/N]: y
Save arguments to samconfig.toml [Y/n]: y

Looking for resources needed for deployment: Not found.
Creating the required resources...
```

What is SAM?

```
Looking for samconfig.toml : Not found

Setting default arguments for 'sam deploy'
=====
Stack Name [sam-app]:
AWS Region [us-east-1]:
#Shows you resources changes to be deployed and require a 'Y' to initiate deploy
Confirm changes before deploy [y/N]:
#SAM needs permission to be able to create roles to connect to the resources in your template
Allow SAM CLI IAM role creation [Y/n]:
HelloWorldFunction may not have authorization defined, Is this okay? [y/N]: y
Save arguments to samconfig.toml [Y/n]:

Looking for resources needed for deployment: Not found.
Creating the required resources...
Successfully created!

Managed S3 bucket: aws-sam-cli-managed-default-samclisourcebucket-1mqrhz409kxtt
A different default S3 bucket can be set in samconfig.toml

Saved arguments to config file
Running 'sam deploy' for future deployments will use the parameters saved above.
The above parameters can be changed by modifying samconfig.toml
Learn more about samconfig.toml syntax at
https://docs.aws.amazon.com/serverless-application-model/latest/developerguide/serverless-sam-cli-config.html

Deploying with following values
=====
Stack name      : sam-app
Region         : us-east-1
Confirm changeset   : False
Deployment s3 bucket : aws-sam-cli-managed-default-samclisourcebucket-1mqrhz409kxtt
Capabilities    : ["CAPABILITY_IAM"]
Parameter overrides : {}

Initiating deployment
=====
Uploading to sam-app/a56605ae6b10dbb4412f2b813a0a614f 534633 / 534633.0 (100.00%)
HelloWorldFunction may not have authorization defined.
Uploading to sam-app/34e7466018060df438d91a607f8d6bab.template 1090 / 1090.0 (100.00%)

Waiting for changeset to be created..
```

What is SAM?

CREATE_IN_PROGRESS	AWS::IAM::Role	HelloWorldFunctionRole	Resource creation Initiated
CREATE_IN_PROGRESS	AWS::IAM::Role	HelloWorldFunctionRole	-
CREATE_COMPLETE	AWS::IAM::Role	HelloWorldFunctionRole	-
CREATE_IN_PROGRESS	AWS::Lambda::Function	HelloWorldFunction	-
CREATE_IN_PROGRESS	AWS::Lambda::Function	HelloWorldFunction	Resource creation Initiated
CREATE_COMPLETE	AWS::Lambda::Function	HelloWorldFunction	-
CREATE_IN_PROGRESS	AWS::ApiGateway::RestApi	ServerlessRestApi	-
CREATE_IN_PROGRESS	AWS::ApiGateway::RestApi	ServerlessRestApi	Resource creation Initiated
CREATE_COMPLETE	AWS::ApiGateway::RestApi	ServerlessRestApi	-
CREATE_IN_PROGRESS	AWS::ApiGateway::Deployment	ServerlessRestApiDeployment47fc2d5f9d	-
CREATE_IN_PROGRESS	AWS::ApiGateway::Deployment	ServerlessRestApiDeployment47fc2d5f9d	Resource creation Initiated
CREATE_IN_PROGRESS	AWS::Lambda::Permission	HelloWorldFunctionHelloWorldPermissionProd	Resource creation Initiated
CREATE_IN_PROGRESS	AWS::Lambda::Permission	HelloWorldFunctionHelloWorldPermissionProd	-
CREATE_COMPLETE	AWS::ApiGateway::Deployment	ServerlessRestApiDeployment47fc2d5f9d	-
CREATE_IN_PROGRESS	AWS::ApiGateway::Stage	ServerlessRestApiProdStage	-
CREATE_IN_PROGRESS	AWS::ApiGateway::Stage	ServerlessRestApiProdStage	Resource creation Initiated
CREATE_COMPLETE	AWS::ApiGateway::Stage	ServerlessRestApiProdStage	-
CREATE_COMPLETE	AWS::Lambda::Permission	HelloWorldFunctionHelloWorldPermissionProd	-
CREATE_COMPLETE	AWS::CloudFormation::Stack	sam-app	-

CloudFormation outputs from deployed stack

Outputs

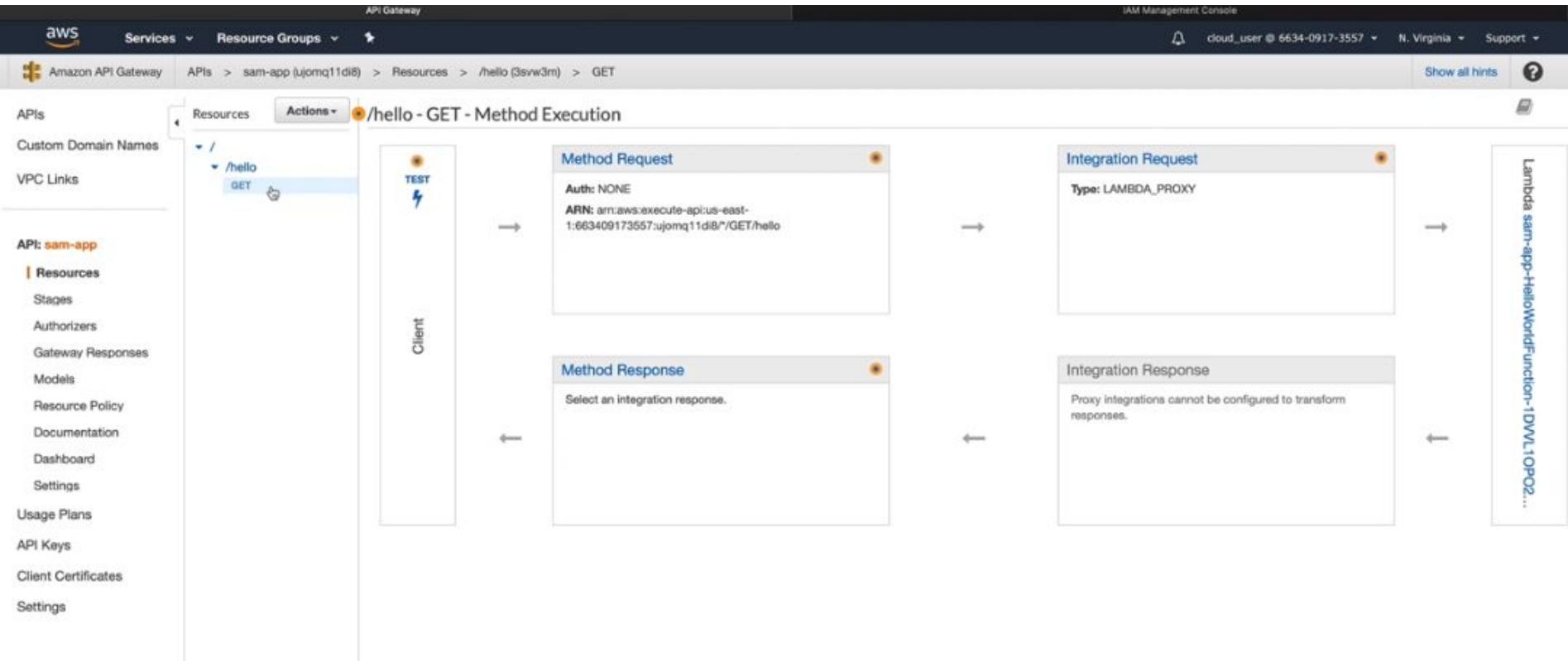
Key	HelloWorldFunctionIamRole
Description	Implicit IAM Role created for Hello World function
Value	arn:aws:iam::663409173557:role/sam-app>HelloWorldFunctionRole-JDV0F80CRJ10

Key	HelloWorldApi
Description	API Gateway endpoint URL for Prod stage for Hello World function
Value	https://ujomqlid18.execute-api.us-east-1.amazonaws.com/Prod/hello/

Key	HelloWorldFunction
Description	Hello World Lambda Function ARN
Value	arn:aws:lambda:us-east-1:663409173557:function:sam-app>HelloWorldFunction-1DVVL10P025I2

Successfully created/updated stack - sam-app in us-east-1

What is SAM?



What is SAM?

sam-app-HelloWorldFunction-1DVVL1OPO25I2

Throttle Qualifiers Actions Select a test event Test Save

Function code [Info](#)

Code entry type Edit code inline Runtime Python 3.7 Handler app.lambda_handler

File Edit Find View Go Tools Window Save Test

Environment sam-app-HelloWork certifi certifi-2020.4.5.1.dist-info chardet chardet-3.0.4.dist-info idna idna-2.9.dist-info requests requests-2.23.0.dist-info urllib3 urllib3-1.25.9.dist-info __init__.py app.py requirements.txt

app.py

```
context: object, required
    Lambda Context runtime methods and attributes
    Context doc: https://docs.aws.amazon.com/lambda/latest/dg/python-context-object.html

    Returns
    -----
    API Gateway Lambda Proxy Output Format: dict
    Return doc: https://docs.aws.amazon.com/apigateway/latest/developerguide/set-up-lambda-proxy-integrations.html

try:
    ip = requests.get("http://checkip.amazonaws.com/")
except requests.RequestException as e:
    # Send some context about this error to Lambda Logs
    print(e)

    raise e

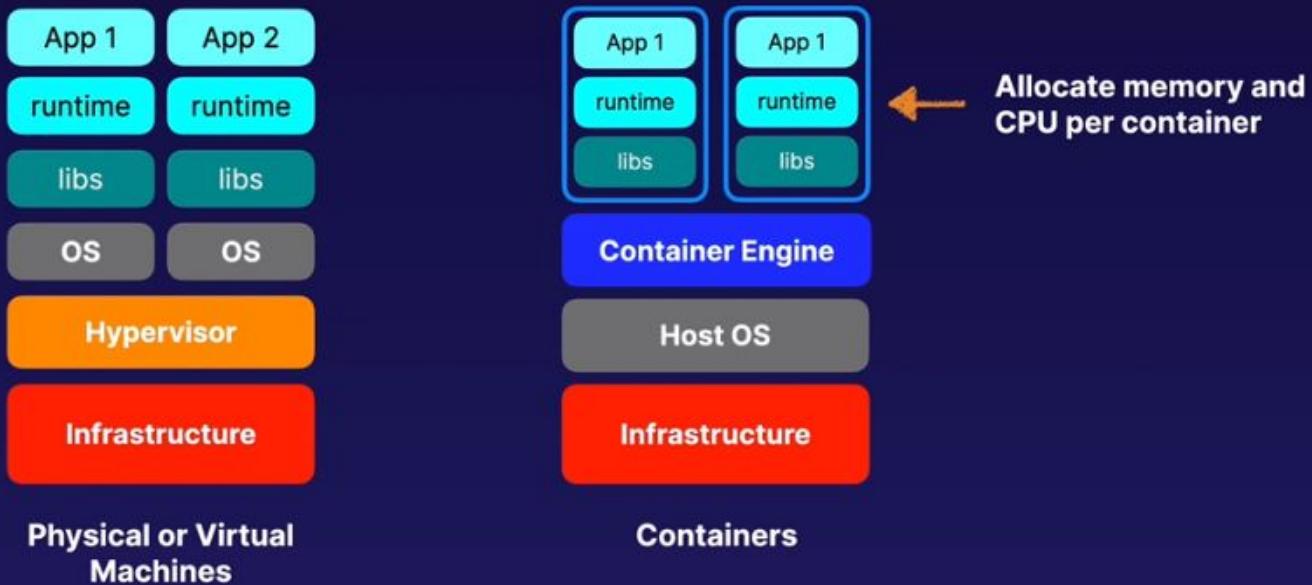
return {
    "statusCode": 200,
    "body": json.dumps({
        "message": "hello world",
        "location": ip.text.replace("\n", "")
    })
}
```

[ec2-user@ip-172-31-69-52 sam-app]\$ curl https://ujomq11di8.execute-api.us-east-1.amazonaws.com/Prod/hello/
{"message": "hello world"} "location": "18.232.59.90"}[ec2-user@ip-172-31-69-52 sam-app]\$

Elastic Container Service (ECS)

Elastic Container Service (ECS)

- A **container** is a package that contains an application, libraries, runtime, and tools required to run it
- Run on a container engine like **Docker**
- Provides the **isolation** benefits of virtualization with less overhead and faster starts than VMs
- Containerized applications are **portable** and offer a consistent environment



Elastic Container Service (ECS)

- Managed container **orchestration service**
- Create **clusters** to manage fleets of container deployments
- ECS manages EC2 or Fargate instances
- Schedules containers for optimal placement
- Defines rules for CPU and memory requirements
- Monitors resource utilization
- Deploy, update, roll back
- **FREE ... for real!**
- VPC, security groups, EBS volumes
- ELB
- CloudTrail and CloudWatch



ECS Terminology

Cluster

Logical collection of ECS resources — either ECS EC2 instances or Fargate instances

Task Definition

Defines your application. Similar to a Dockerfile but for running containers in ECS. Can contain multiple containers.

Container Definition

Inside a task definition, it defines the individual containers a task uses. Controls CPU and memory allocation and port mappings.

Task

Single running copy of any containers defined by a task definition. One working copy of an application (e.g., DB and web containers).

Service

Allows task definitions to be scaled by adding tasks. Defines minimum and maximum values.

Registry

Storage for container images (e.g., Elastic Container Registry (ECR) or Docker Hub). Used to download images to create containers.

Fargate

- **Serverless** container engine
- Eliminates need to provision and manage servers
- Specify and pay for resources per application
- Works with both **ECS** and **EKS**
- Each workload runs in its own kernel
- Isolation and security
- Choose EC2 instead if:
 - Compliance requirements
 - Require broader customization
 - Require GPUs



EKS

- Elastic Kubernetes Service
- K8s is **open-source** software that lets you deploy and manage containerized applications at scale
- Same toolset on-premises and in cloud
- Containers are grouped in **pods**
- Like ECS, supports both EC2 and Fargate
- Why use EKS?
 - Already using K8s
 - Want to migrate to AWS



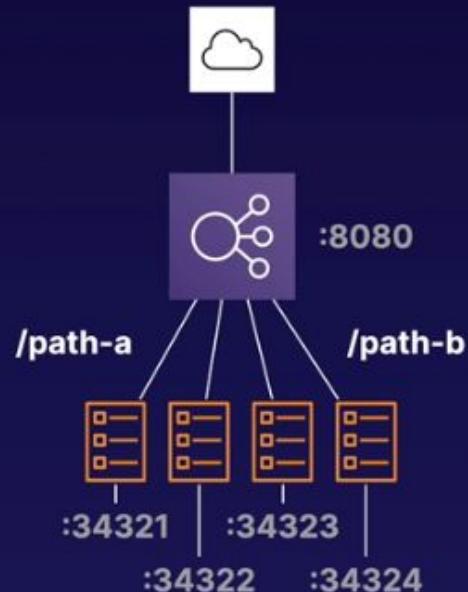
ECR

- Managed Docker container registry
- Store, manage, and deploy images
- Integrated with ECS and EKS
- Works with on-premises deployments
- Highly available
- Integrated with **IAM**
- Pay for storage and data transfer

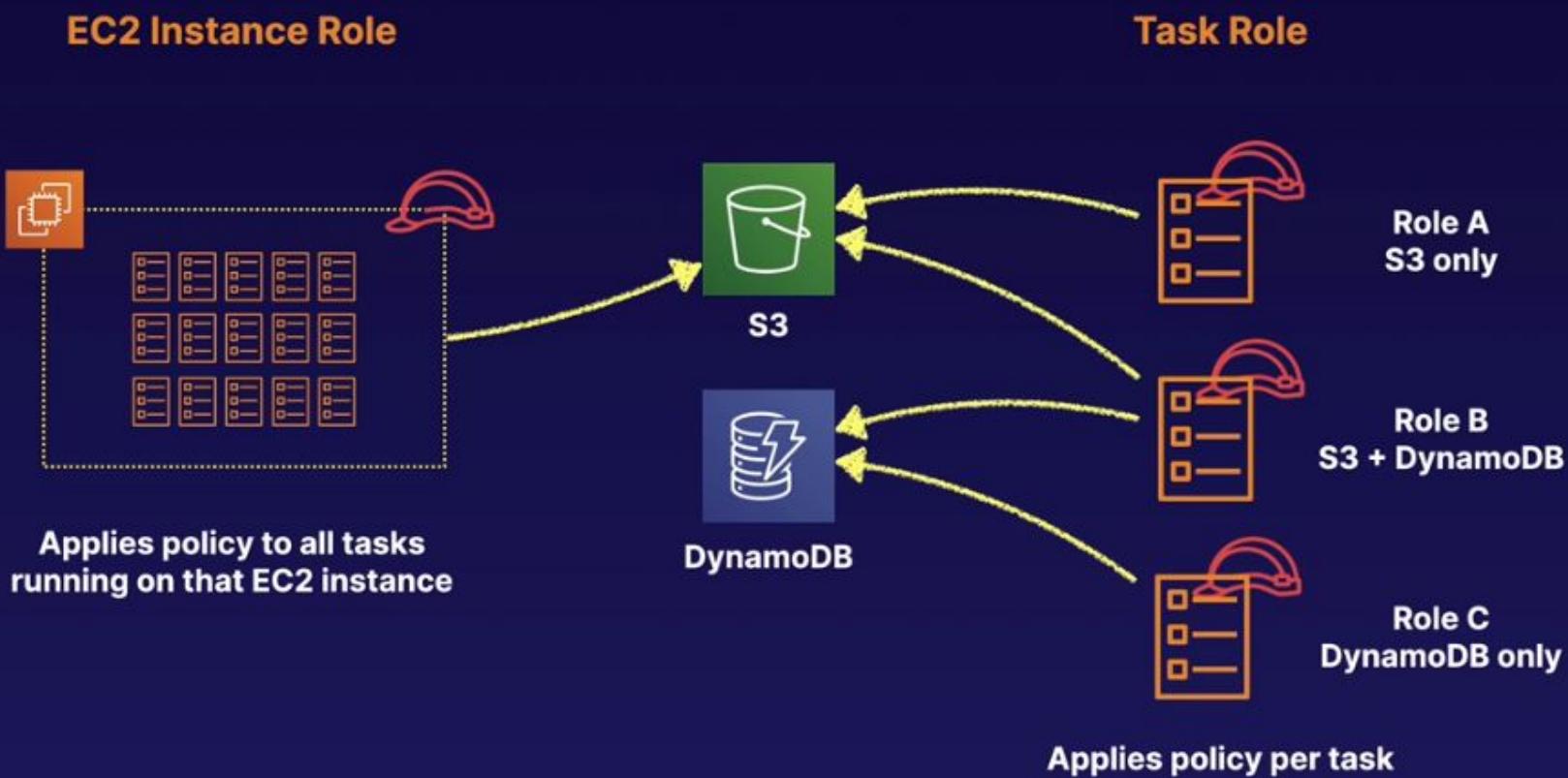


ECS + ELB

- Distribute traffic evenly across tasks in your service
- Supports ALB, NLB, CLB
- Use ALB to route HTTP/HTTPS (layer 7) traffic
- Use NLB or CLB to route TCP (layer 4) traffic
- Supported by both EC2 and Fargate launch types
- ALB allows:
 - Dynamic host port mapping
 - Path-based routing
 - Priority rules
- ALB is recommended over NLB or CLB



ECS security



ECS

Getting Started with Amazon Elastic Container Service (Amazon ECS) using Fargate

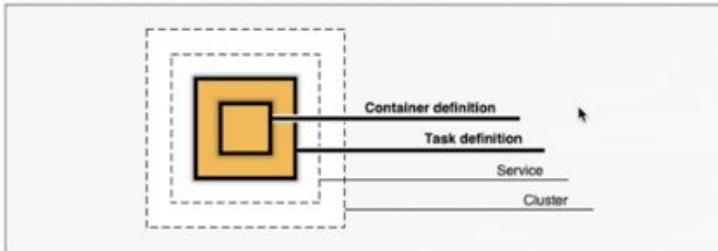
Step 1: Container and Task

Step 2: Service

Step 3: Cluster

Step 4: Review

Diagram of ECS objects and how they relate



Container definition

Edit

Choose an image for your container below to get started quickly or define the container image to use.

sample-app

image : httpd:2.4
memory : 0.5GB (512)
cpu : 0.25 vCPU (256)

nginx

image : nginx:latest
memory : 0.5GB (512)
cpu : 0.25 vCPU (256)

tomcat-webserver

image : tomcat
memory : 2GB (2048)
cpu : 1 vCPU (1024)

custom

Configure

image : --
memory : --
cpu : --

Task definition

Edit

A task definition is a blueprint for your application, and describes one or more containers through attributes. Some attributes are configured at the task level but the majority of attributes are configured per container.

ECS

Getting Started with Amazon Elastic Container Service (Amazon ECS) using Fargate

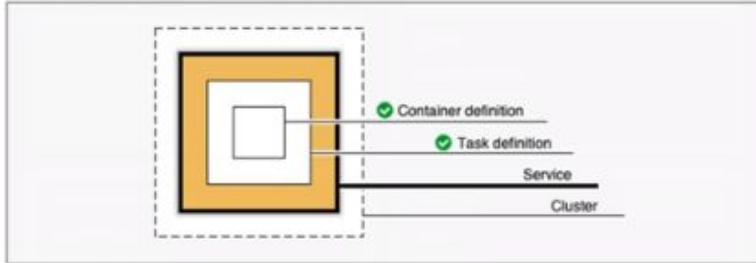
Step 1: Container and Task

Step 2: Service

Step 3: Cluster

Step 4: Review

Diagram of ECS objects and how they relate



Define your service

Edit

A service allows you to run and maintain a specified number (the "desired count") of simultaneous instances of a task definition in an ECS cluster.

Service name sample-app-service

Number of desired tasks 1

Security group Automatically create new

A security group is created to allow all public traffic to your service only on the container port specified.
You can further configure security groups and network access outside of this wizard.

Load balancer type None

Application Load Balancer

*Required

Cancel

Previous

Next

ECS

Getting Started with Amazon Elastic Container Service (Amazon ECS) using Fargate

Launch Status

We are creating resources for your service. This may take up to 10 minutes. When we're complete, you can view your service.

Back

[View service](#)

Enabled after service creation completes successfully

Additional features that you can add to your service after creation

Scale based on metrics

You can configure scaling rules based on CloudWatch metrics

Preparing service : 3 of 9 complete

ECS resource creation

Cluster default	complete
Task definition first-run-task-definition:1	complete
Service	pending

Additional AWS service integrations

Log group /ecs/first-run-task-definition	complete
CloudFormation stack	pending
VPC	pending
Subnet 1	pending
Subnet 2	pending
Security group	pending

ECS

Amazon ECS

Clusters

- Task Definitions
- Account Settings
- Amazon EKS
- Clusters
- Amazon ECR
- Repositories
- AWS Marketplace
- Discover software
- Subscriptions

Clusters > default > Service: sample-app-service

Service : sample-app-service

Cluster default Status ACTIVE Task definition first-run-task-definition:1 Desired count 1 Pending count 0 Running count 1

Service type REPLICA Launch type FARGATE Service role AWSServiceRoleForECS

Details Tasks Events Auto Scaling Deployments Metrics Tags Logs

Last updated on June 10, 2020 11:40:48 AM (0m ago)

Task status: Running Stopped

Task	Task Definition	Last status	Desired status	Group	Launch type	Platform version
da770152-8c0e-415f-82f4-d94...	first-run-task-definition:1	RUNNING	RUNNING	service:sample-app-service	FARGATE	1.3.0

ECS

Amazon ECS
Clusters
Task Definitions
Amazon EKS
Clusters
Amazon ECR
Repositories
AWS Marketplace
Discover software
Subscriptions

Clusters > default > Service: sample-app-service

Service : sample-app-service

Update Delete

Cluster	default	Desired count	2
Status	ACTIVE	Pending count	1
Task definition	first-run-task-definition:1	Running count	1
Service type	REPLICA		
Launch type	FARGATE		
Service role	AWSServiceRoleForECS		

Details Tasks Events Auto Scaling Deployments Metrics Tags Logs

Last updated on June 10, 2020 11:44:13 AM (0m ago)



Task status: Running Stopped

Filter in this page

1-2 > Page size 50 ▾

Task	Task Definition	Last status	Desired status	Group	Launch type	Platform version
f0ca747e-f643-4011-979d-dff... f972b175-b33c-48d2-9bcf-b43...	first-run-task-definition:1	RUNNING	RUNNING	service:sample-app-service	FARGATE	1.3.0
		PENDING	RUNNING	service:sample-app-service	FARGATE	1.3.0

Serverless Summary

Exam Tips

Lambda Exam Tips

- Lambda scales out (not up) automatically
- Lambda functions are independent, 1 event = 1 function
- Lambda is serverless
- Know what services are serverless!
- Lambda functions can trigger other lambda functions, 1 event can = x functions if functions trigger other functions

Exam Tips

Lambda Exam Tips

- Architectures can get extremely complicated, AWS X-ray allows you to debug what is happening
- Lambda can do things globally, you can use it to back up S3 buckets to other S3 buckets etc
- Know your triggers

Serverless Quizz

Exam Quizz

- The availability zone that DynamoDB is hosted in is down.
- You have written your function in Python which is not supported as a runtime environment for Lambda.
- Your lambda function does not have sufficient Identity Access Management (IAM) permissions to write to DynamoDB.
- The availability zone that Lambda is hosted in is down.

Sorry!

Correct Answer

Like any services in AWS, Lambda needs to have a Role associated with it that provide credentials with rights to other services. This is exactly the same as needing a Role on an EC2 instance to access S3 or DDB.

Exam Quizz

 Work with your web design team to create some web pages with embedded JavaScript to emulate your 5 most popular information web pages and sign up web pages.

 Create a stand by sign up server to use in case the primary fails due to load.

 Recreate your 5 most popular new customer web pages and sign up web pages on Lightsail and take advantage of AWS auto scaling to pick up the load.

 Create a duplicate sign up page that stores registration details in DynamoDB for asynchronous processing using SQS & Lambda.

 Upgrade your existing server from a 1xlarge to a 32xlarge for the duration of the campaign.

 Work with your web design team to create some web pages in PHP to run on a 32xlarge EC2 instance to emulate your 5 most popular information web pages and sign up web pages.

Sorry!

Correct Answer

A 500x increase is beyond the scope of a well designed single server system to absorb unless it is already hugely over-specialised to accommodate this sort of burst load. An AWS solution for this situation might include S3 static web pages with client side scripting to meet high demand of information pages. Plus use of a noSQL database to collect customer registration for asynchronous processing, and SQS backed by scalable compute to keep up with the requests. Lightsail does provide a scalable provisioned service solutions, but these still need to be designed and planned by you and so offer no significant advantage in this situation. A standby server is a good idea, but will not help with the anticipated 500x load increase.

Exam Quizz

QUESTION 3

You have created a serverless application to add metadata to images that are uploaded to a specific S3 bucket. To do this, your lambda function is configured to trigger whenever a new image is created in the bucket. What will happen when multiple users upload multiple different images at the same time?



Multiple instances of the Lambda function will be triggered, one for each image



A single Lambda functions will be triggered, that will process all images that have finished uploading one at a time



Multiple Lambda functions will trigger, one after the other, until all images are processed



A single Lambda functions will be triggered, which will process all images at the same time

Sorry!

Correct Answer

Each time a Lambda function is triggered, an isolated instance of that function is invoked. Multiple triggers result in multiple concurrent invocations, one for each time it is triggered

Exam Quizz

QUESTION 4

What AWS service can be used to help resolve an issue with a Lambda function?

AWS X-Ray

CloudTrail

DynamoDB

API Gateway

Sorry!

Correct Answer

AWS X-Ray helps developers analyze and debug production, distributed applications, such as those built using a microservices & serverless architectures

Exam Quizz

QUESTION 5

In which direction(s) does Lambda scale automatically?

None - Lambda does not scale automatically

Up and Out

Up

Out

Sorry!

Correct Answer

Lambda scales out automatically - each time your function is triggered, a new, separate instance of that function is started. There are limits, but these can be adjusted on request.

Exam Quizz

QUESTION 6

Lambda pricing is based on which of these measurements?

Choose 2

Duration of execution billed in fractions of seconds.

The amount of CPU assigned.

The amount of memory assigned.

Whether you choose an AMD or Intel processor.

Sorry!

Correct Answer

Lambda billing is based on both The MB of RAM reserved and the execution duration in 100ms units.

Exam Quizz

Which of the following services can invoke Lambda function directly?

Choose 3



S3



Amazon Lex



IAM



API Gateway



EC2



Kinesis Data Firehose

Sorry!

Correct Answer

ALB, Cognito, Lex, Alexa, API Gateway, CloudFront, and Kinesis Data Firehose are all valid direct (synchronous) triggers for Lambda functions. S3 is one of the valid asynchronous triggers.

Exam Quizz

A pricing model based on high level commodity measures such as on compute duration and storage capacity.

 The ability to run applications and services without thinking about servers or capacity provisioning.

 A native Cloud Architecture that allows customers to shift more operational responsibility to AWS.

 A marketing term for HaaS (Hosting as a Service).

 The use of Quantum computing to eliminate the need for physical servers.

Sorry!

Correct Answer

'Serverless' computing is not about eliminating servers, but shifting most of the responsibility for infrastructure and operation of the infrastructure to a vendor so that you can focus more on the business services, not how to manage the infrastructure that they run on. Billing does tend to be based on simple units, but the choice of services, intended usage pattern (RLs), and amount of capacity needed also influences the pricing.

Exam Quizz

QUESTION 9

As a DevOps engineer you are told to prepare complete solution to run a piece of code that required multi-threaded processing. The code has been running on an old custom built server based around a 4 core Intel Xeon processor. Which of these best describe the AWS compute services that could be used?



ECS, and EC2.



EC2, ECS, & Lambda.



Only a EC2 'Bare Steel' server.



None of the above.

Sorry!

Correct Answer

The exact ratio of cores to memory has varied over time for Lambda instances, however Lambda like EC2 and ECS supports hyper-threading on one or more virtual CPUs (if your code supports hyper-threading).