

# Aws certified solutions architect associate

Ahmed Hosni

# 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
<b>Domain 1:</b> Design Resilient Architectures	30%
<b>Domain 2:</b> Design High-Performing Architectures	28%
<b>Domain 3:</b> Design Secure Applications and Architectures	24%
<b>Domain 4:</b> Design Cost-Optimized Architectures	18%
<b>TOTAL</b>	<b>100%</b>

# **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&region=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
- 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.

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



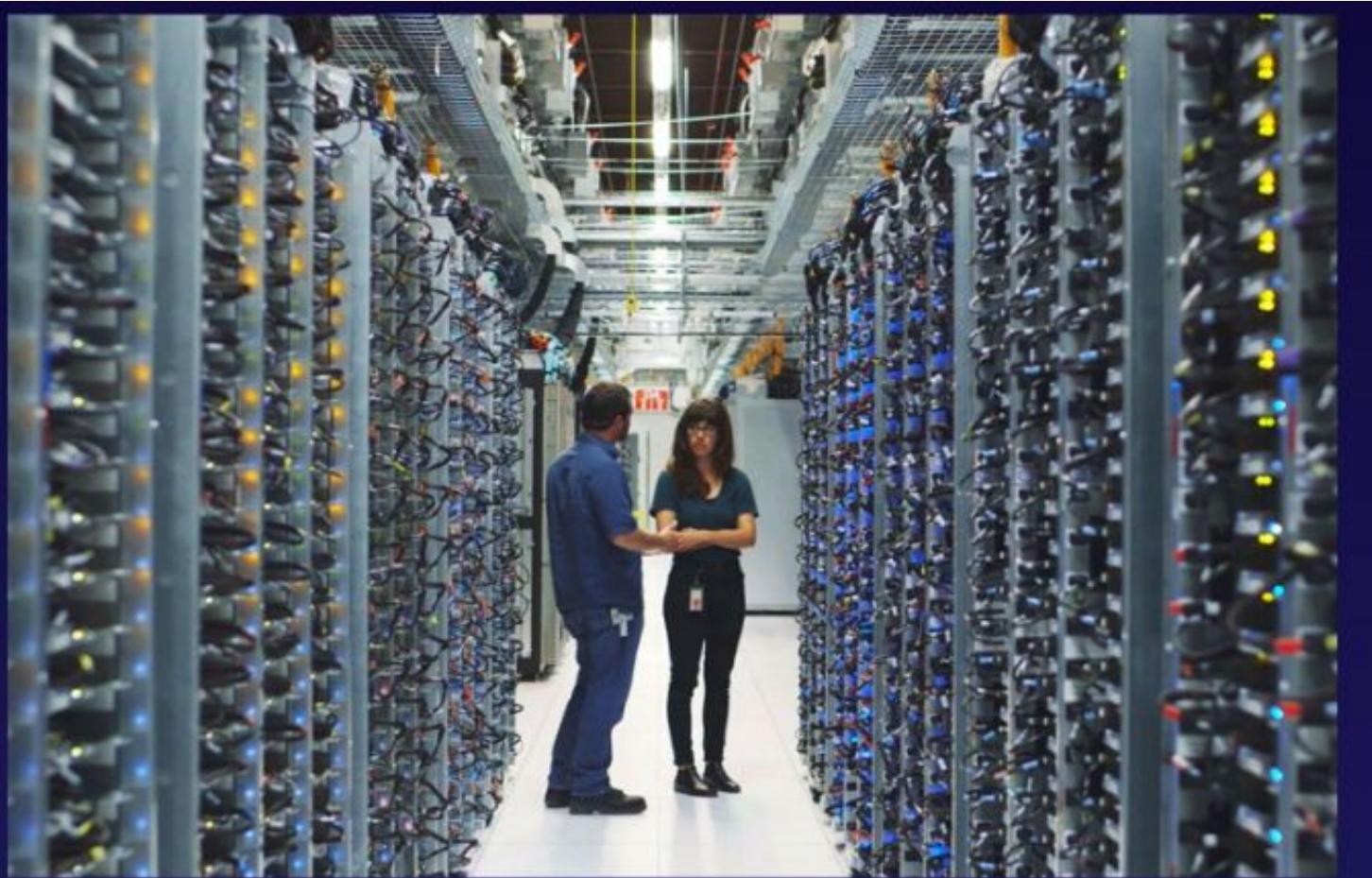
**19 Regions & 57 Availability Zones - December 2018**  
**5 More Regions & 15 More AZ's for 2019**

# AWS Global Infrastructure



**Think of an Availability Zone As A Data Center**

# AWS Global Infrastructure

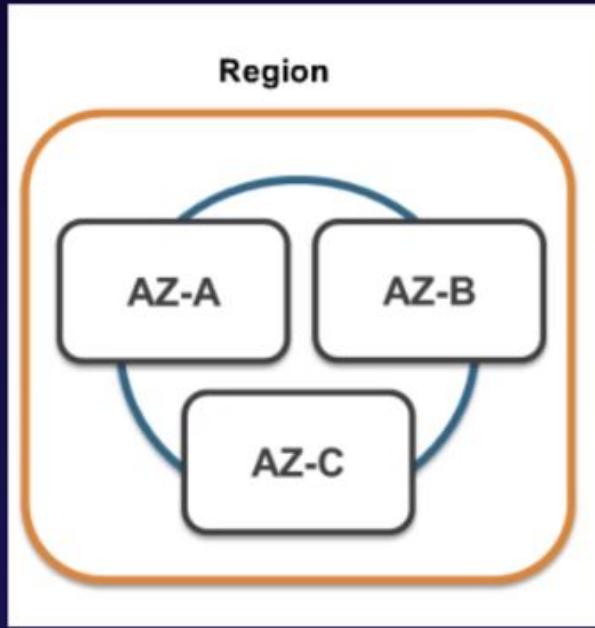


# AWS Global Infrastructure



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) Region<sup>2</sup>

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](#)

<sup>2</sup> 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 Region<sup>1</sup>

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)

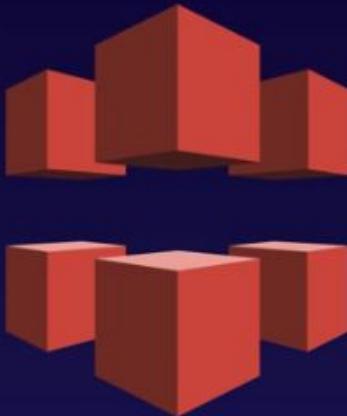
### AWS Edge Network Locations

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

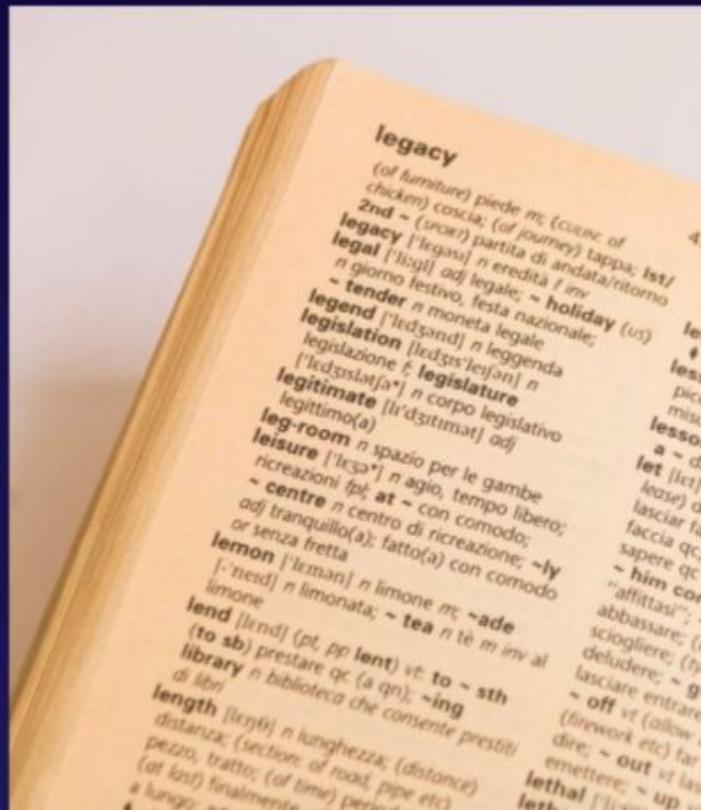
## Policies

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

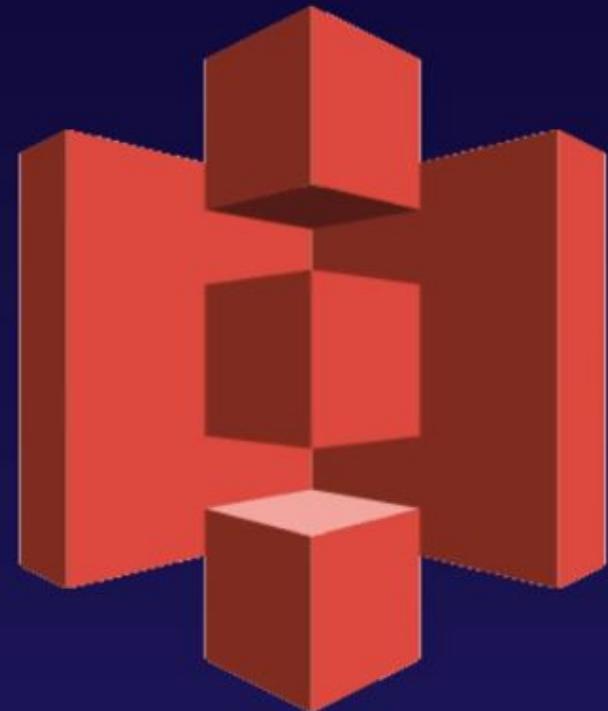
You create roles and then assign them to AWS Resources.



**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

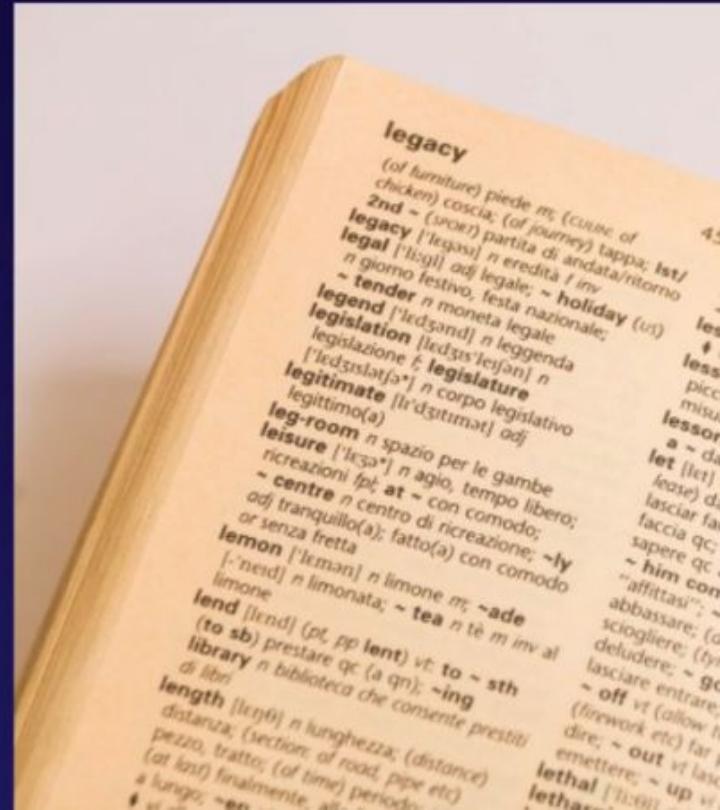
## S3 One Zone - IA

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



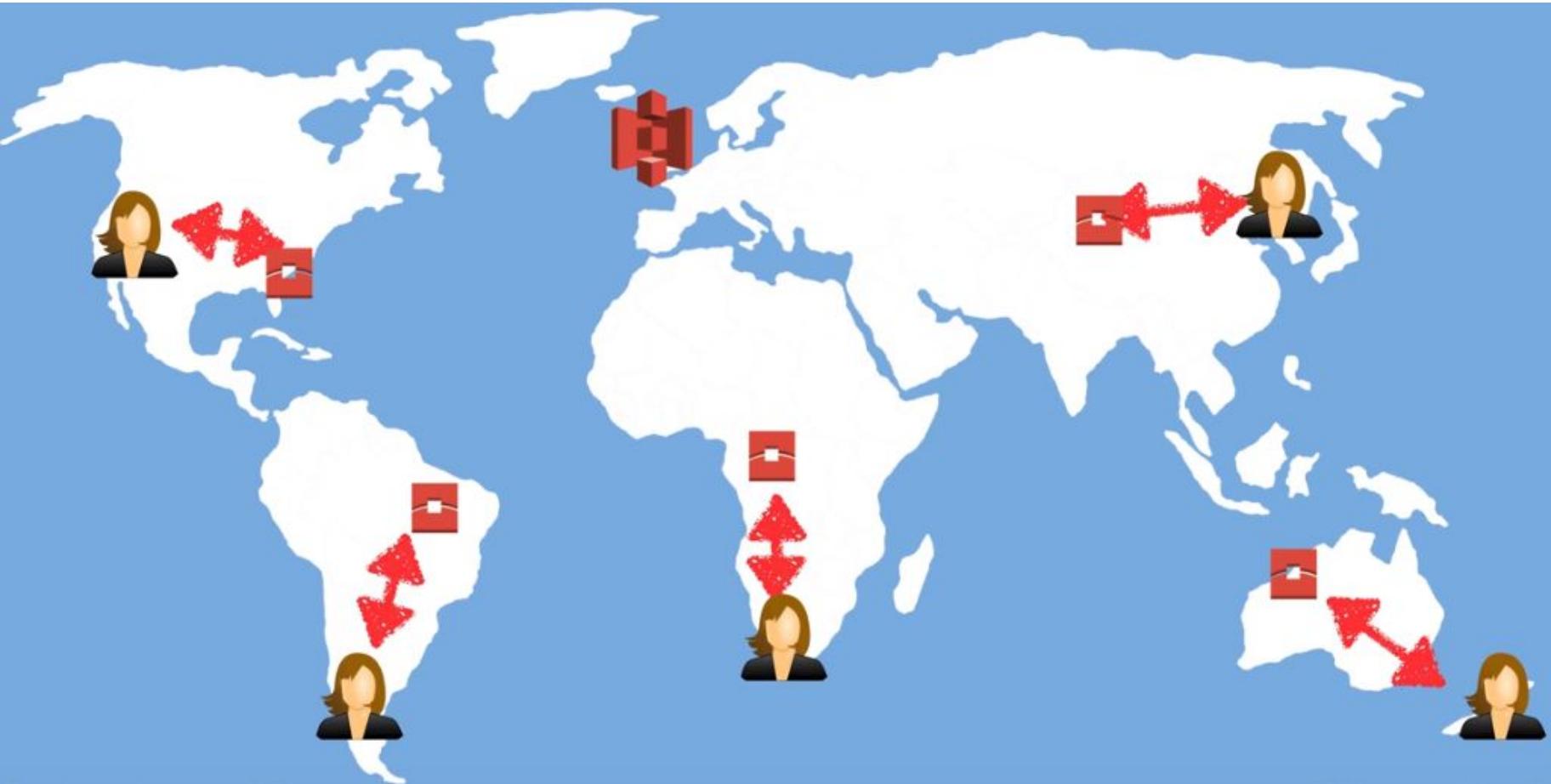
# S3 Transfer Acceleration

**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 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!

# **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

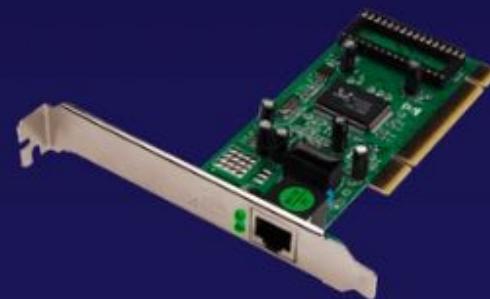
A network device that you can attach to your Amazon EC2 instance to accelerate High Performance Computing (HPC) and machine learning applications.



# ENI vs. ENA vs. EFA

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

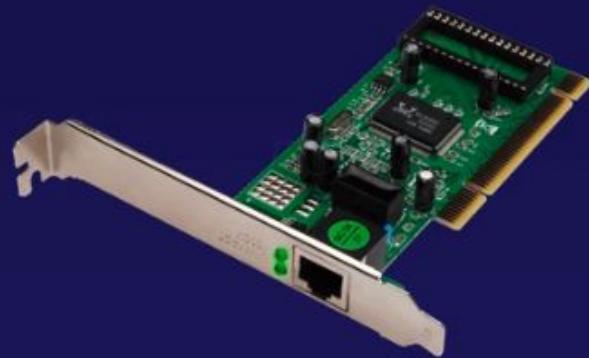
- A primary private IPv4 address from the IPv4 address range of your VPC
- One or more secondary private IPv4 addresses from the IPv4 address range of your VPC
- One Elastic IP address (IPv4) per private IPv4 address
- One public IPv4 address
- One or more IPv6 addresses
- One or more security groups
- 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.



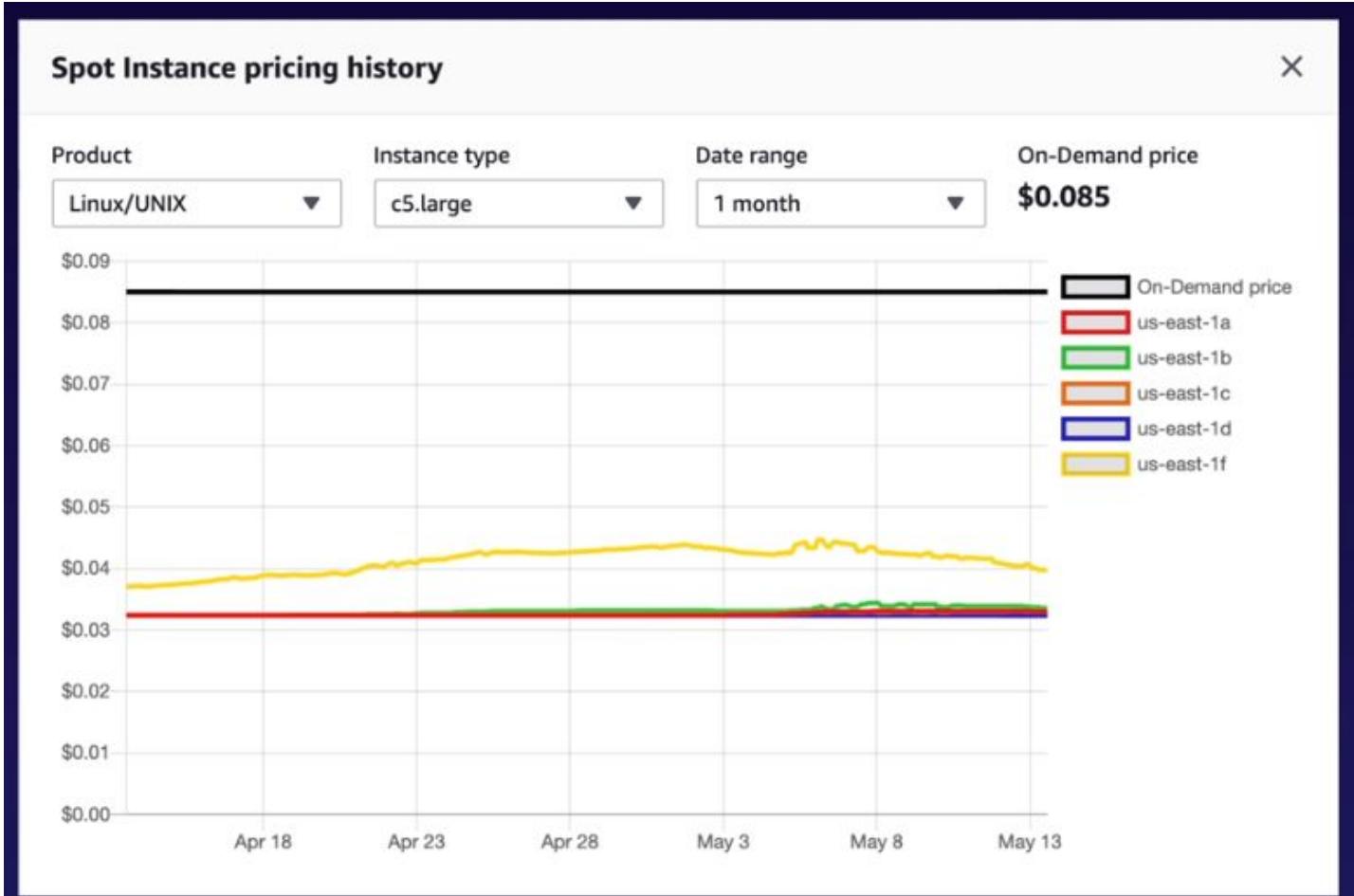
If the Spot price goes above your maximum, you have **two minutes** to choose whether to stop or terminate your instance.

## 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

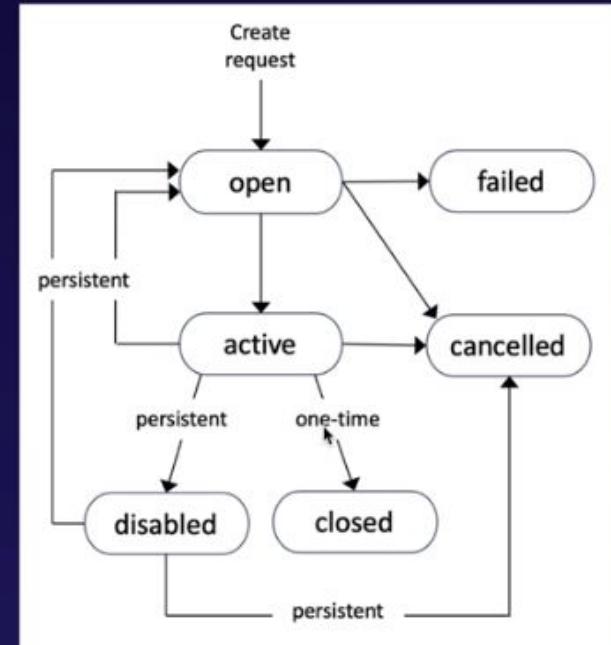
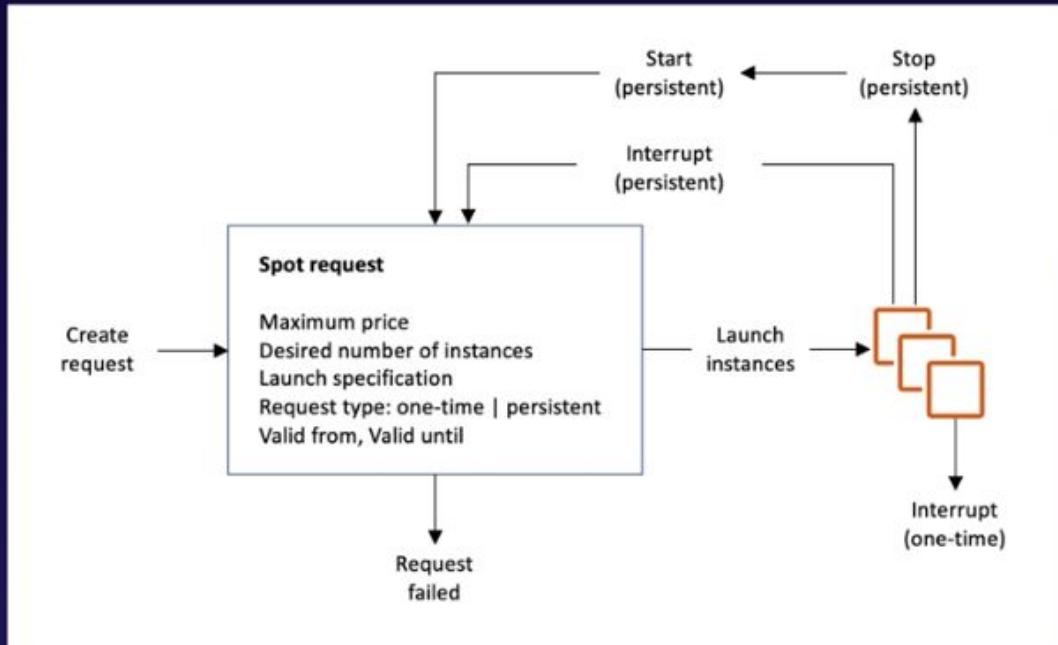
**Spot Instances are not good for:**

- ✓ 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.



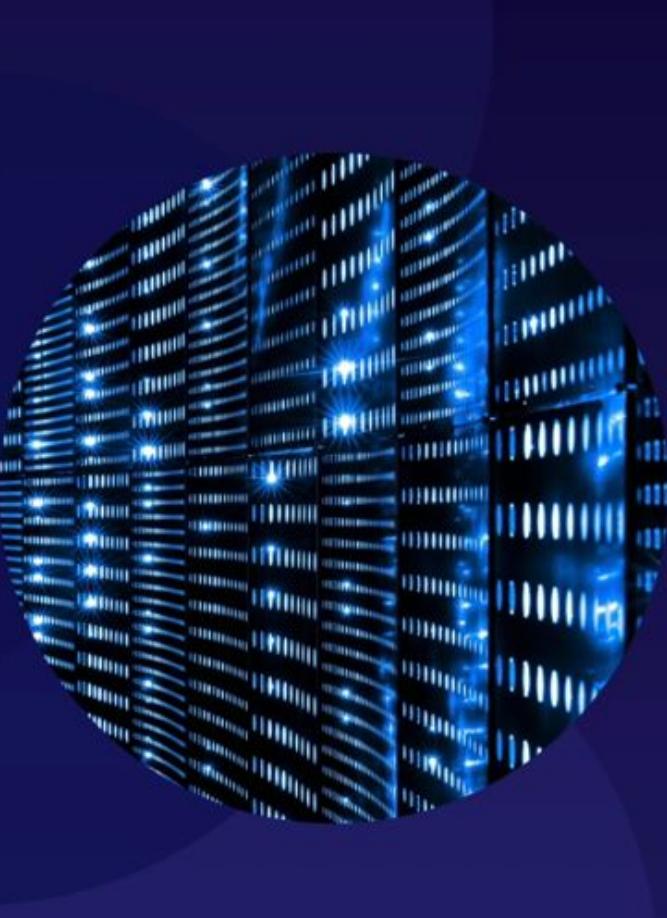
# **EC2 Hibernate**

# 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

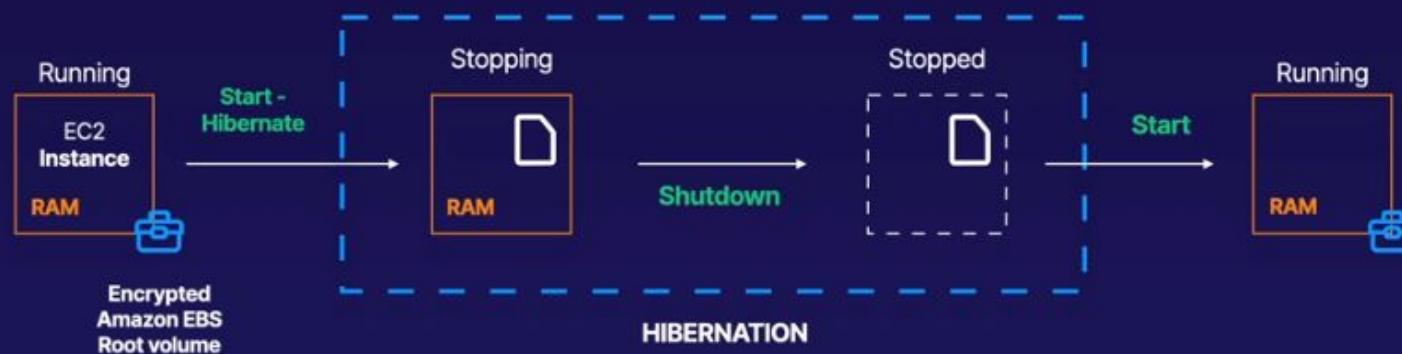
## EC2 Hibernate

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), 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). There are also tabs for 'Status Checks' and 'Monitoring'.

# 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.

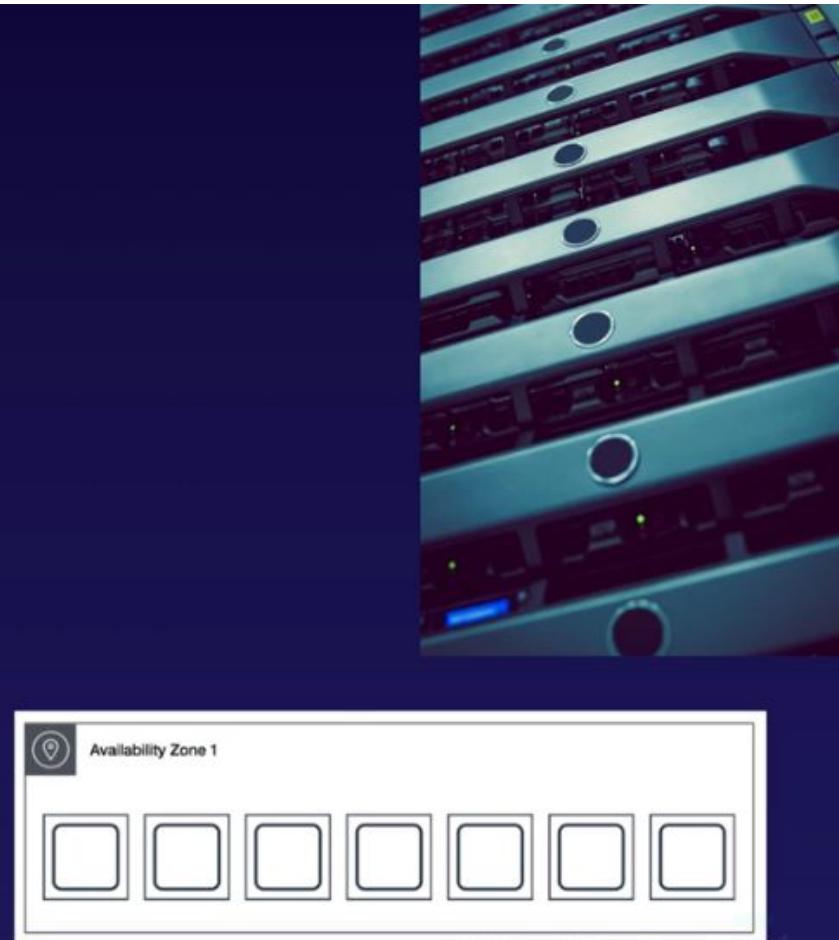


# Spread Placement Group

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

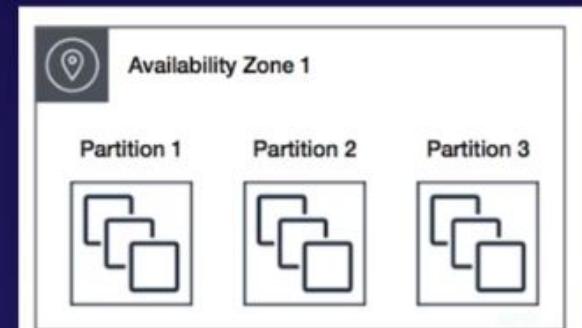
Spread placement groups are recommended for applications that have a small number of critical instances that should be kept separate from each other.

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.

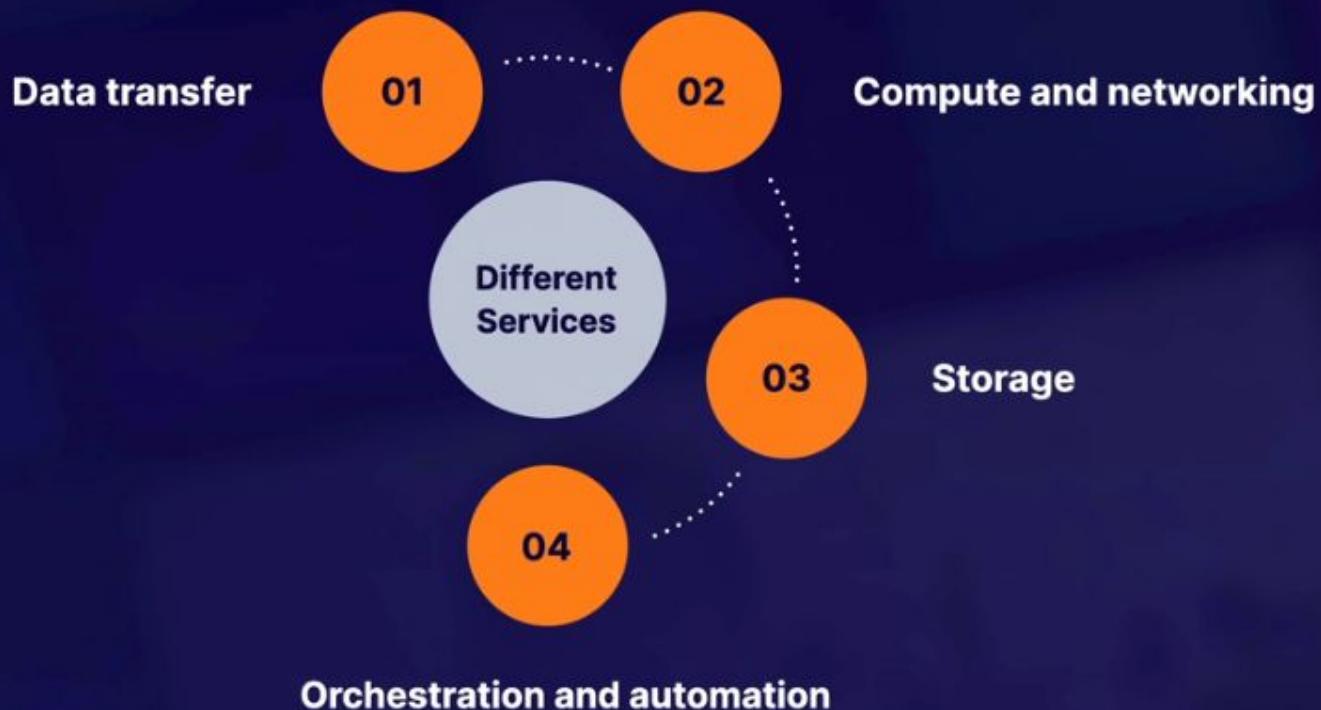
You can create a large number of resources in almost no time. You only pay for the resources you use — and, once finished, **you can destroy the resources**.

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

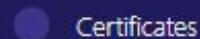
# EC2 Quizz

## QUESTION 18

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



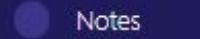
Tags



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



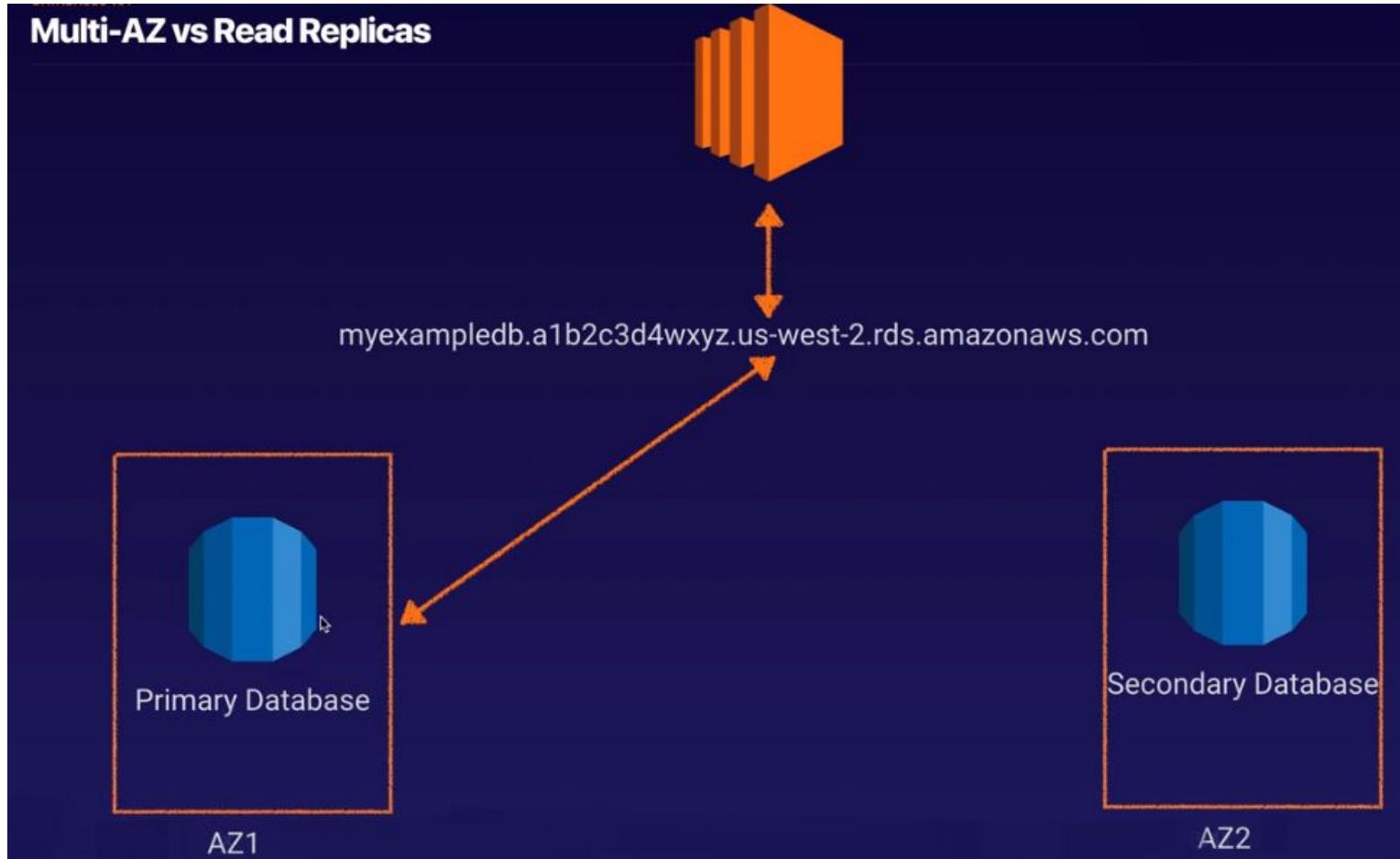
# Databases 101 :: Multi-AZ vs Read replicas

**RDS has two key features;**

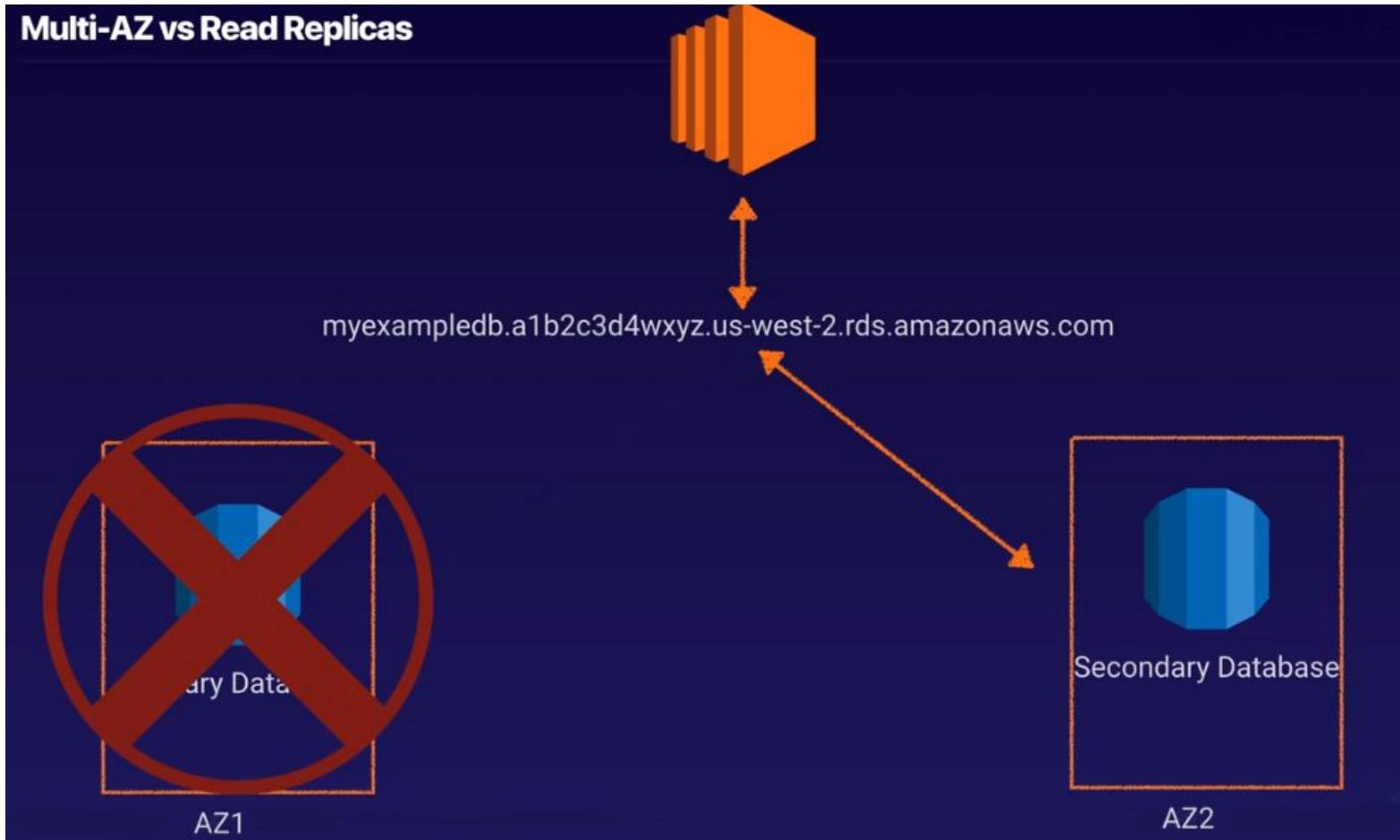
- Multi-AZ - For Disaster Recovery
- Read Replicas - For Performance



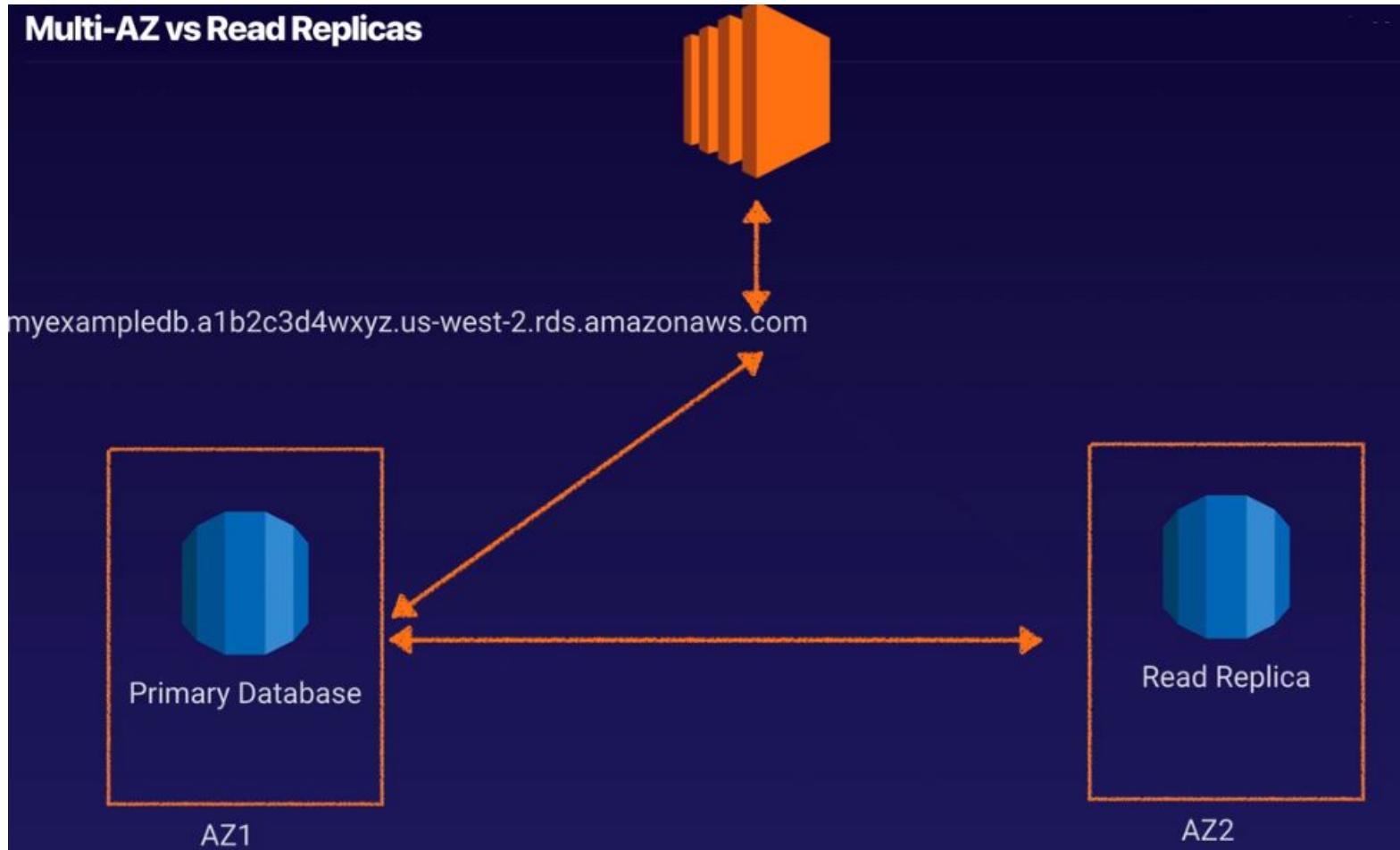
# Databases 101 :: Failover with Multi-AZ



# Databases 101 :: Failover with Multi-AZ

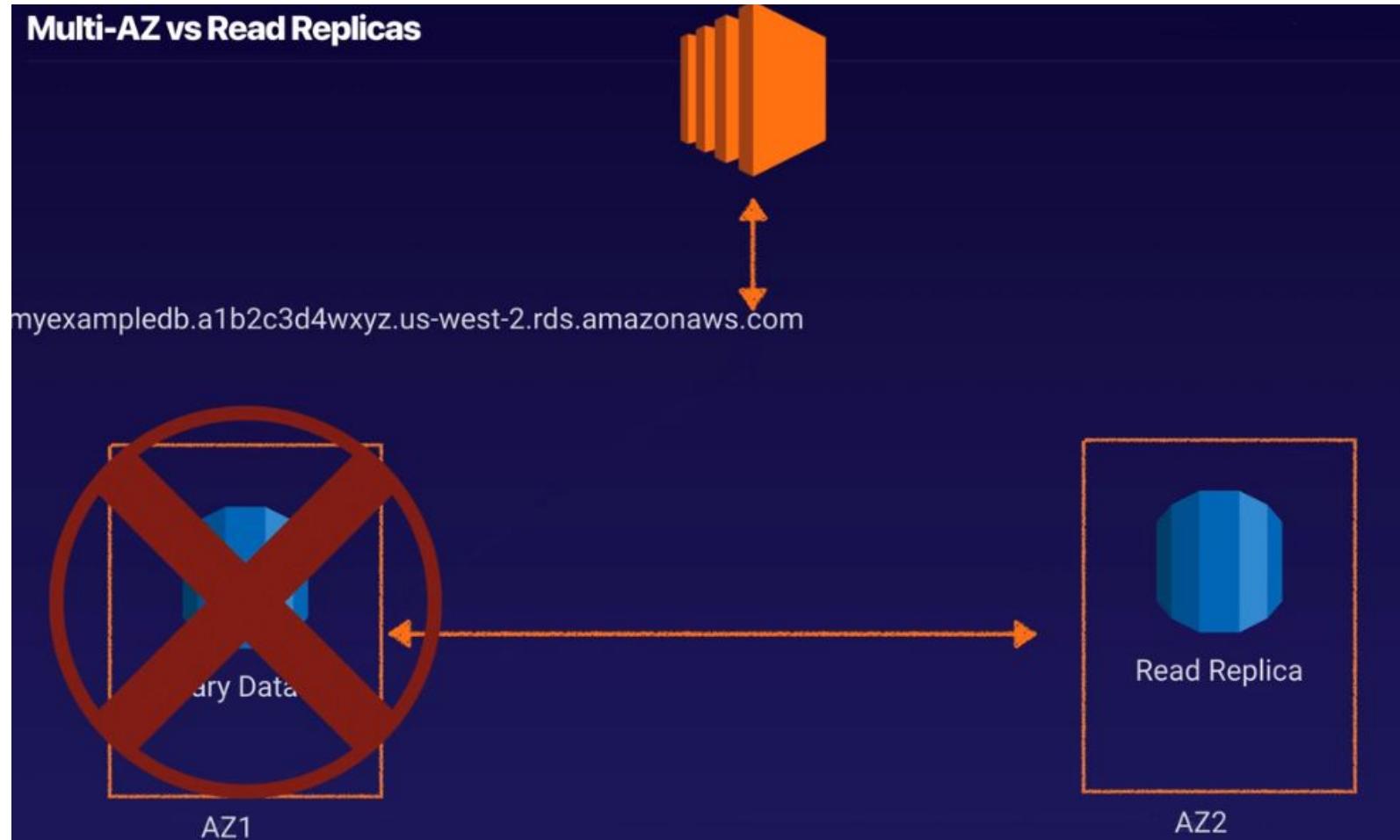


# Databases 101 :: Read replicas

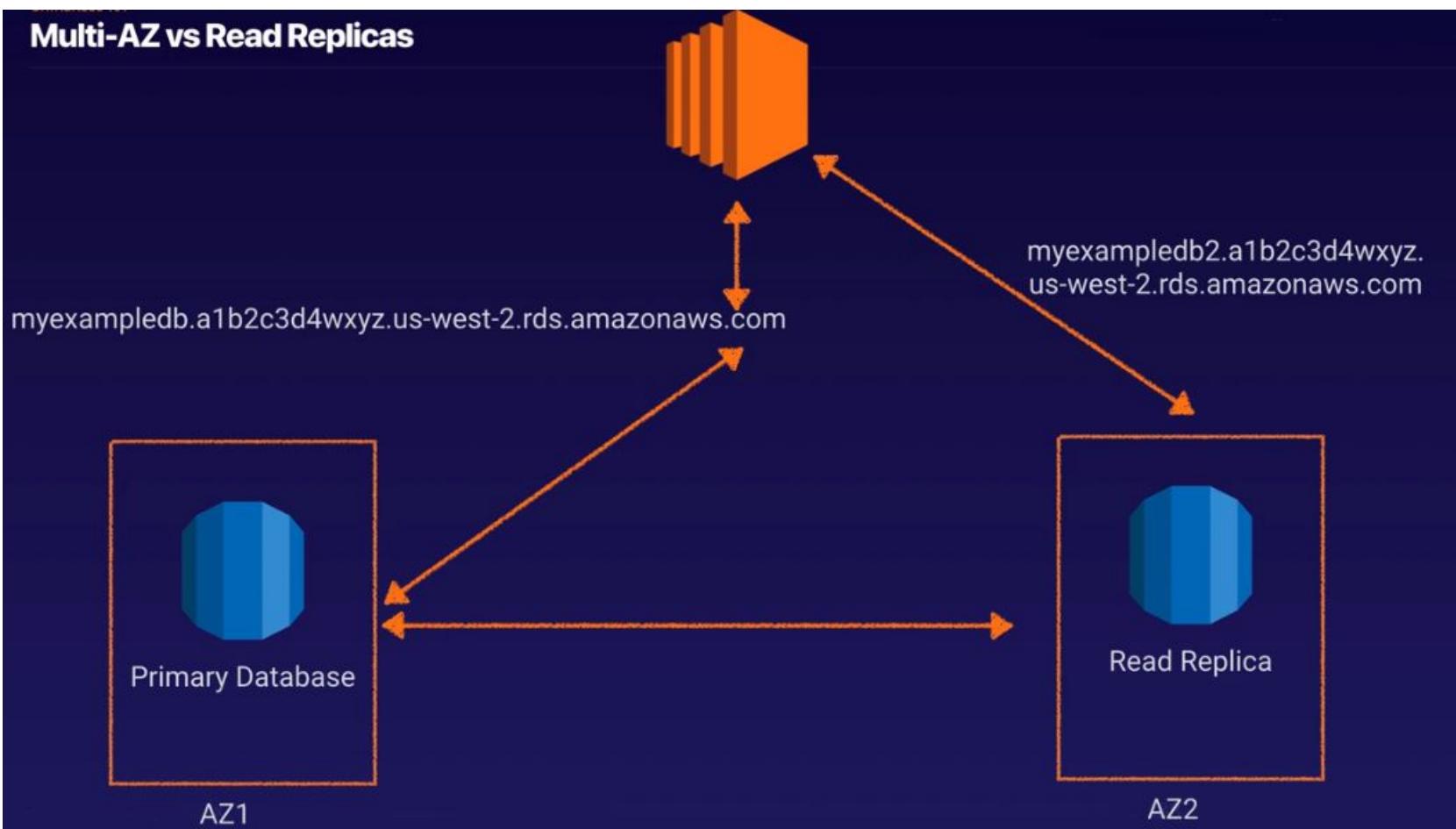


# Databases 101 :: Read replicas

## Multi-AZ vs Read Replicas



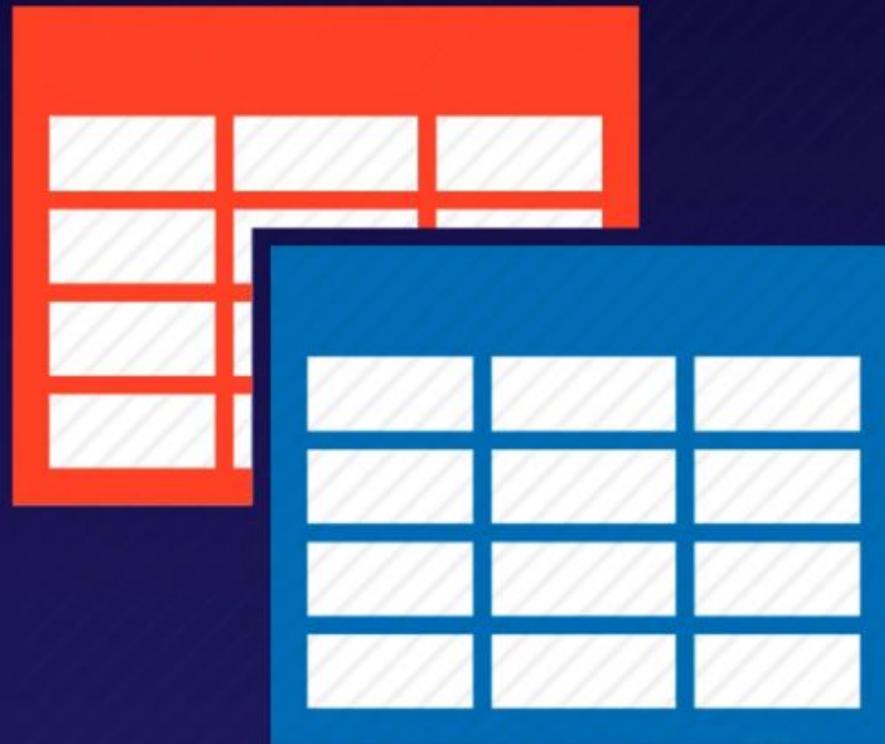
# Databases 101 :: Read replicas



# 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

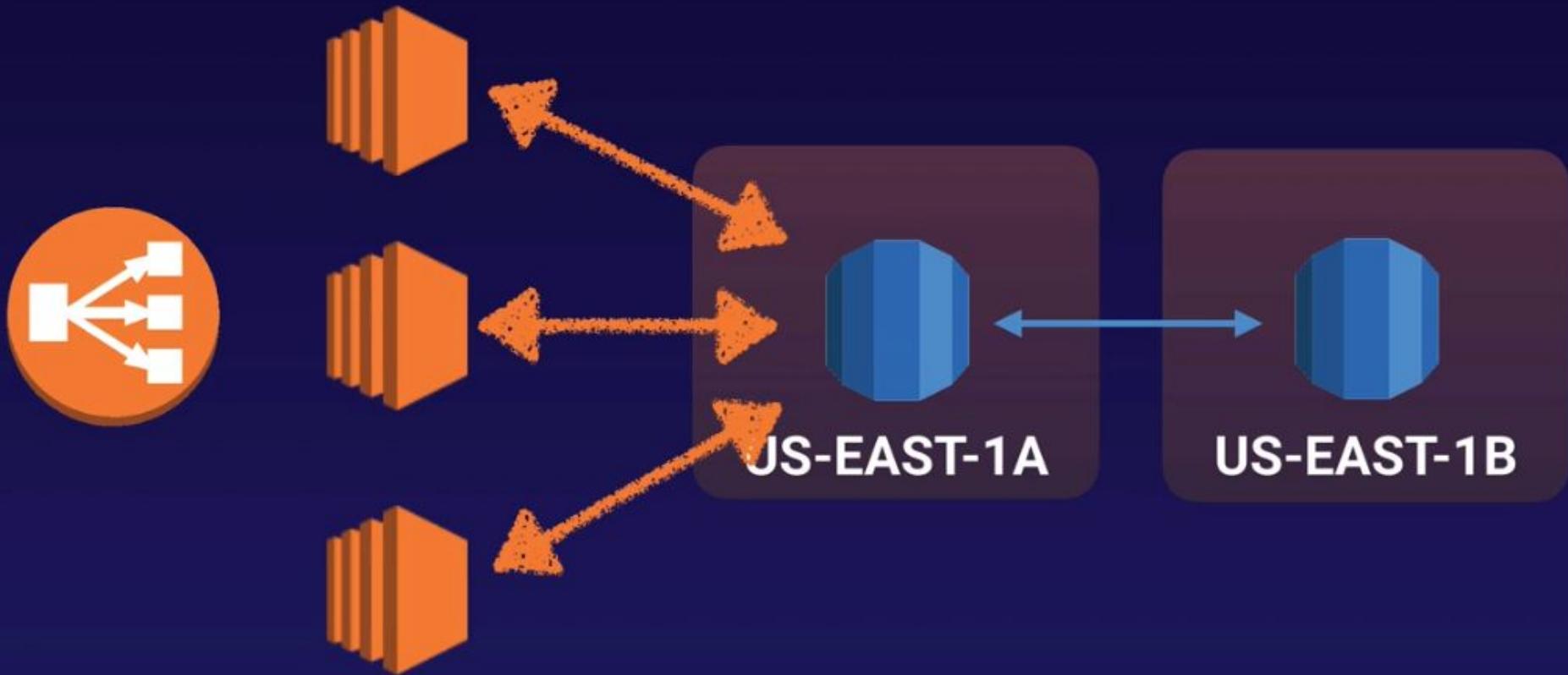
restored.eu-west-1.rds.amazonaws.com

# Encryption at Rest

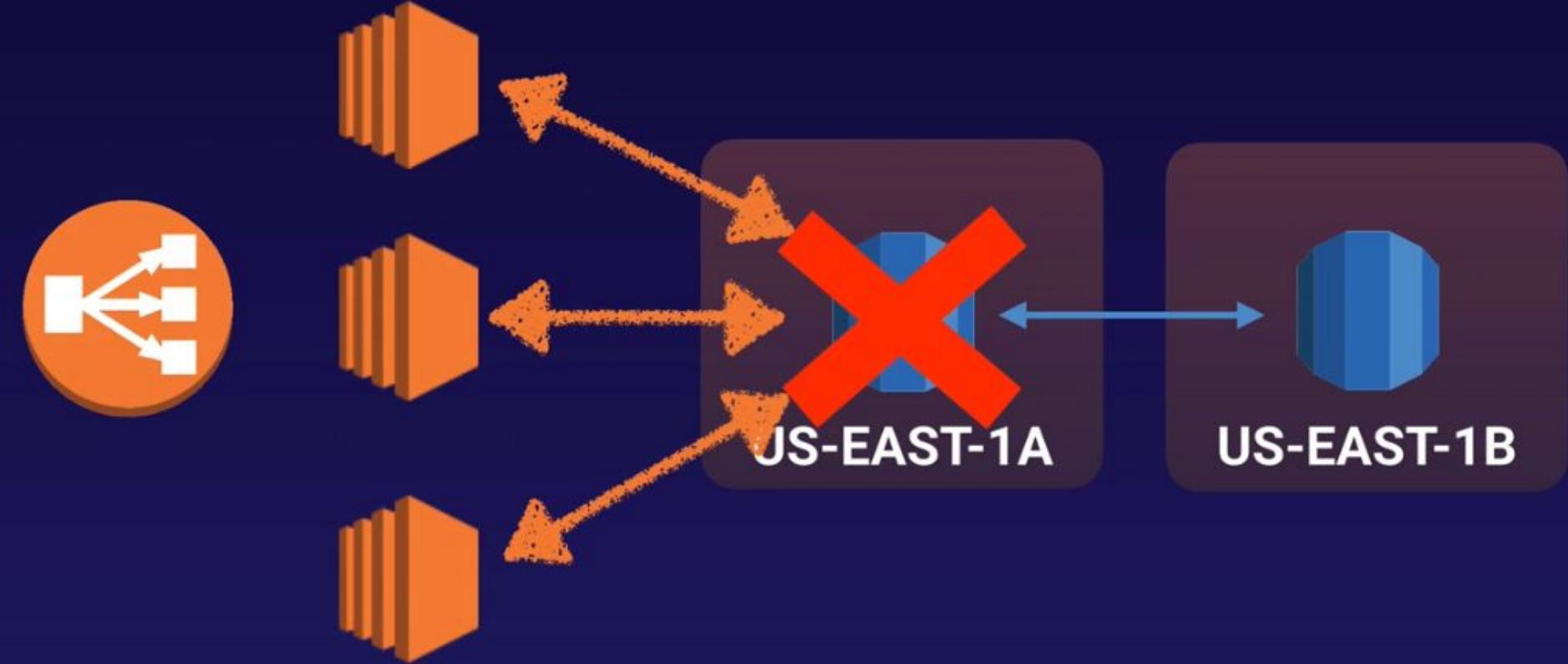
**Encryption at rest is supported for MySQL, Oracle, SQL Server, PostgreSQL, MariaDB & Aurora. Encryption is done using the AWS Key Management Service (KMS) service. Once your RDS instance is encrypted, the data stored at rest in the underlying storage is encrypted, as are its automated backups, read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

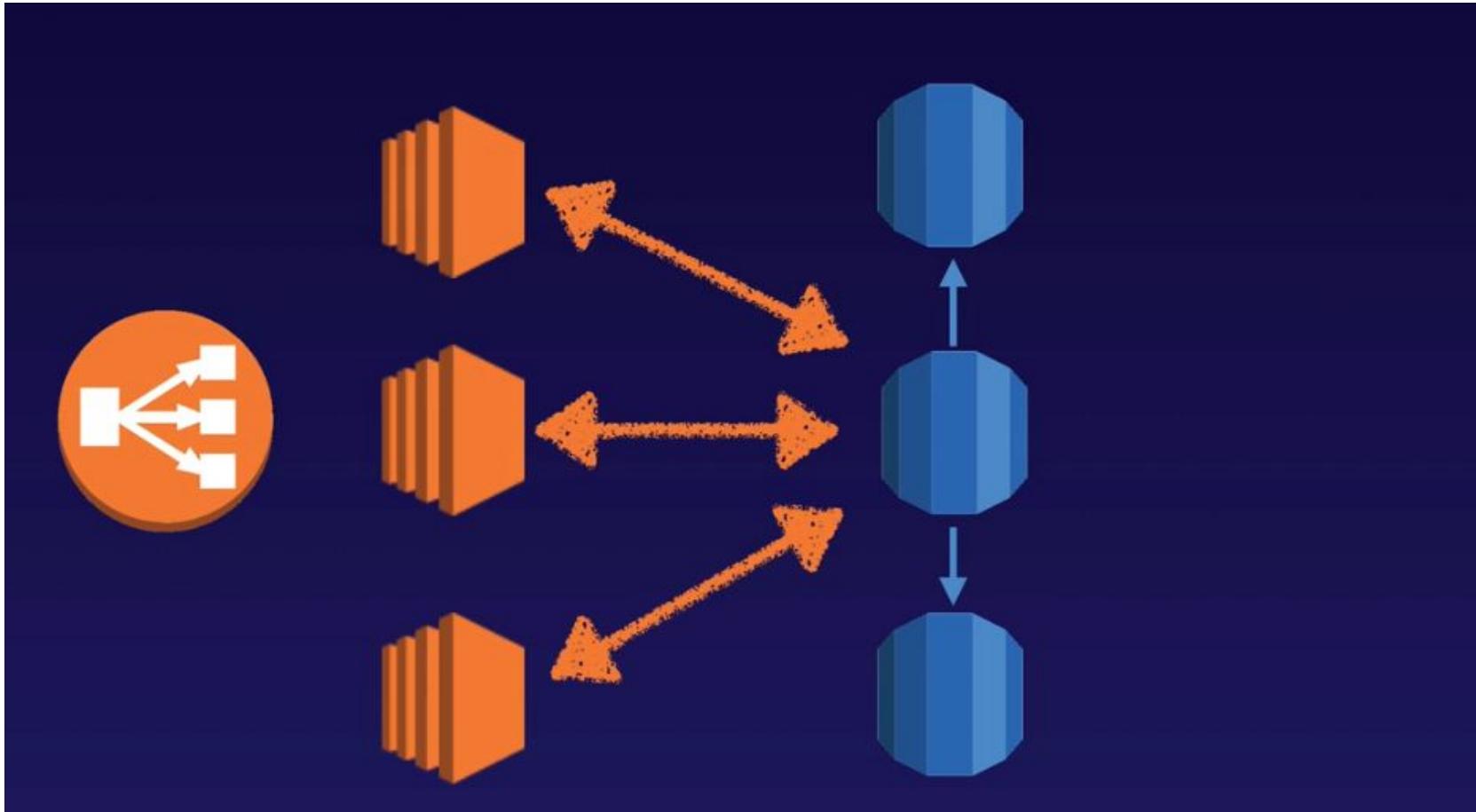
# What is Multi-AZ

Multi-AZ is available for the following databases

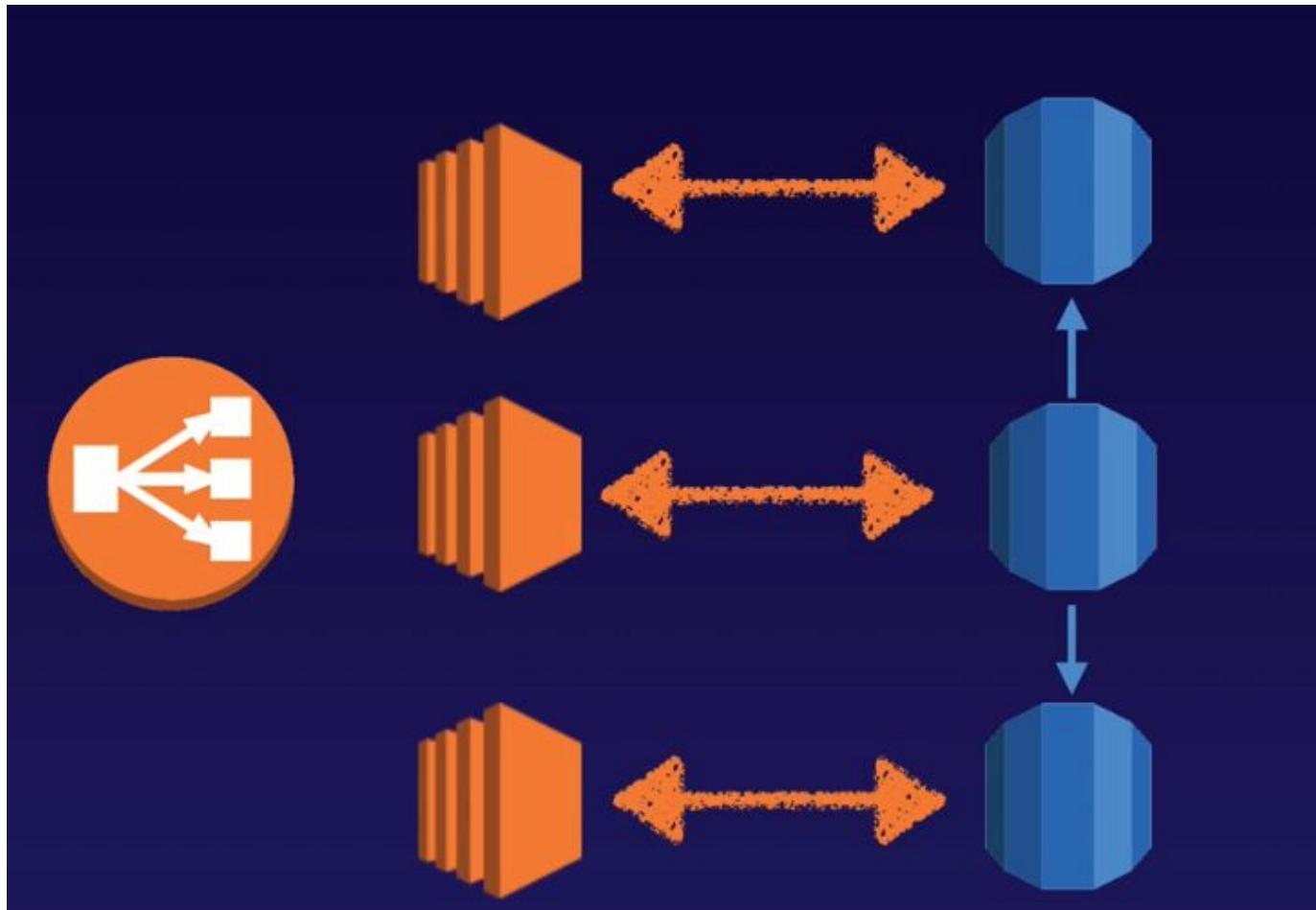
- SQL Server
- Oracle
- MySQL Server
- PostgreSQL
- MariaDB



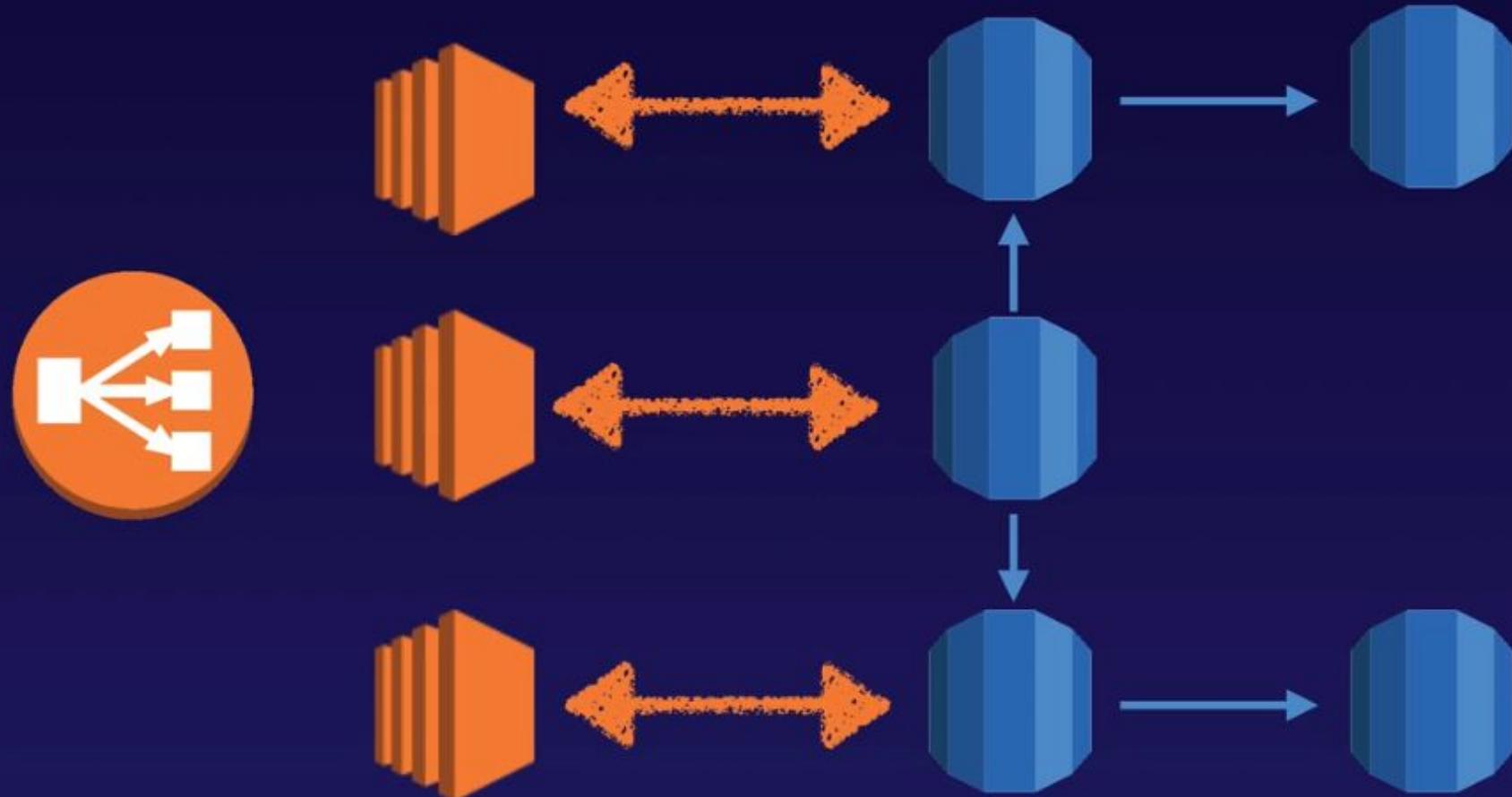
# What is A read replica



# What is A read replica

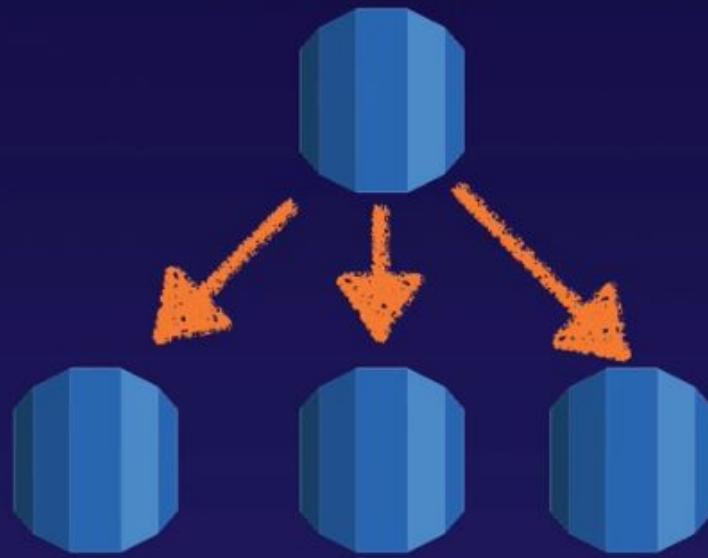


# What is A read replica



# 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**

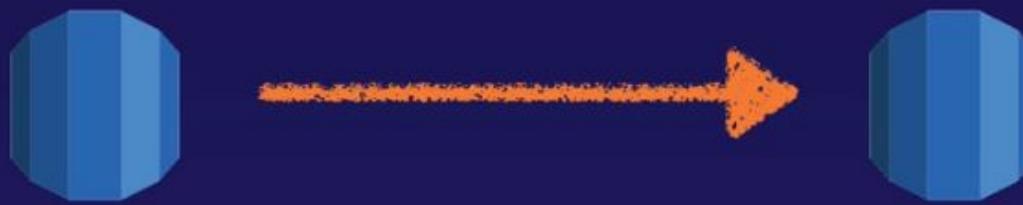
- MySQL Server
- PostgreSQL
- MariaDB
- Oracle
- Aurora



# Backup with RDS

## Things to know about Read Replicas;

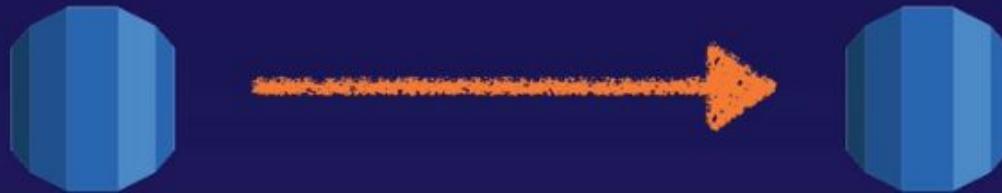
- Used for scaling, not for DR!
- Must have automatic backups turned on in order to deploy a read replica.
- You can have up to 5 read replica copies of any database.
- You can have read replicas of read replicas (but watch out for latency.)



# 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

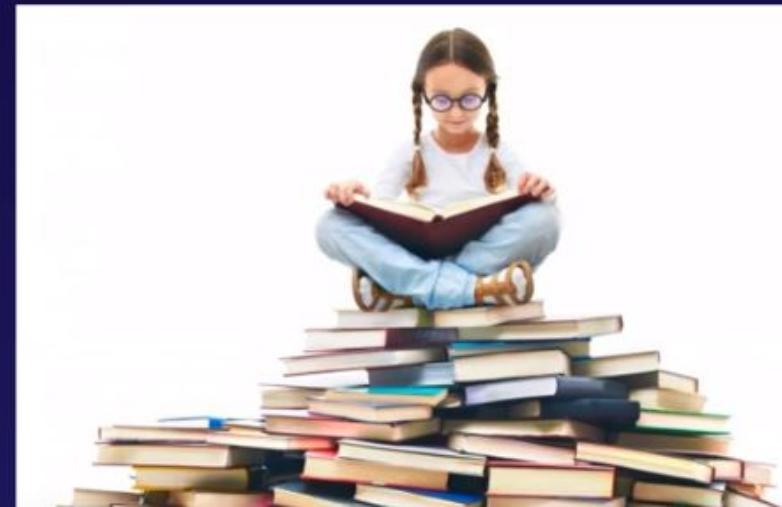


```
    document* item = el->FirstChildElement();
    GroupDesc* ElementDesc elDesc;
    #ifndef using sp_name = item->Attribute("name");
    #endif using spritename = item->Attribute("name");
    float x = boost::lexical_cast<float>(item->Attribute("x"));
    float y = boost::lexical_cast<float>(item->Attribute("y"));
    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

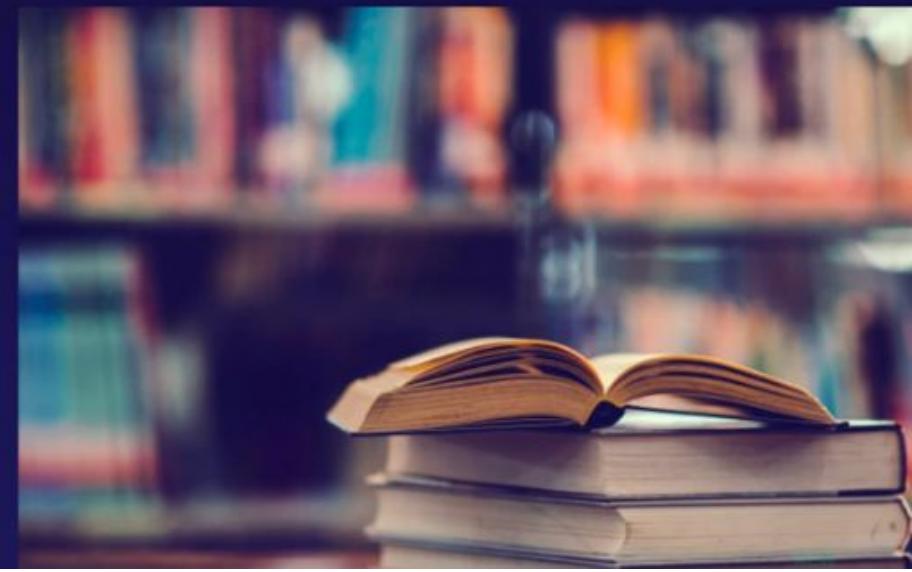
- Consistency across all copies of data is usually reached within a second. Repeating a read after a short time should return the updated data. (Best Read Performance)



# 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

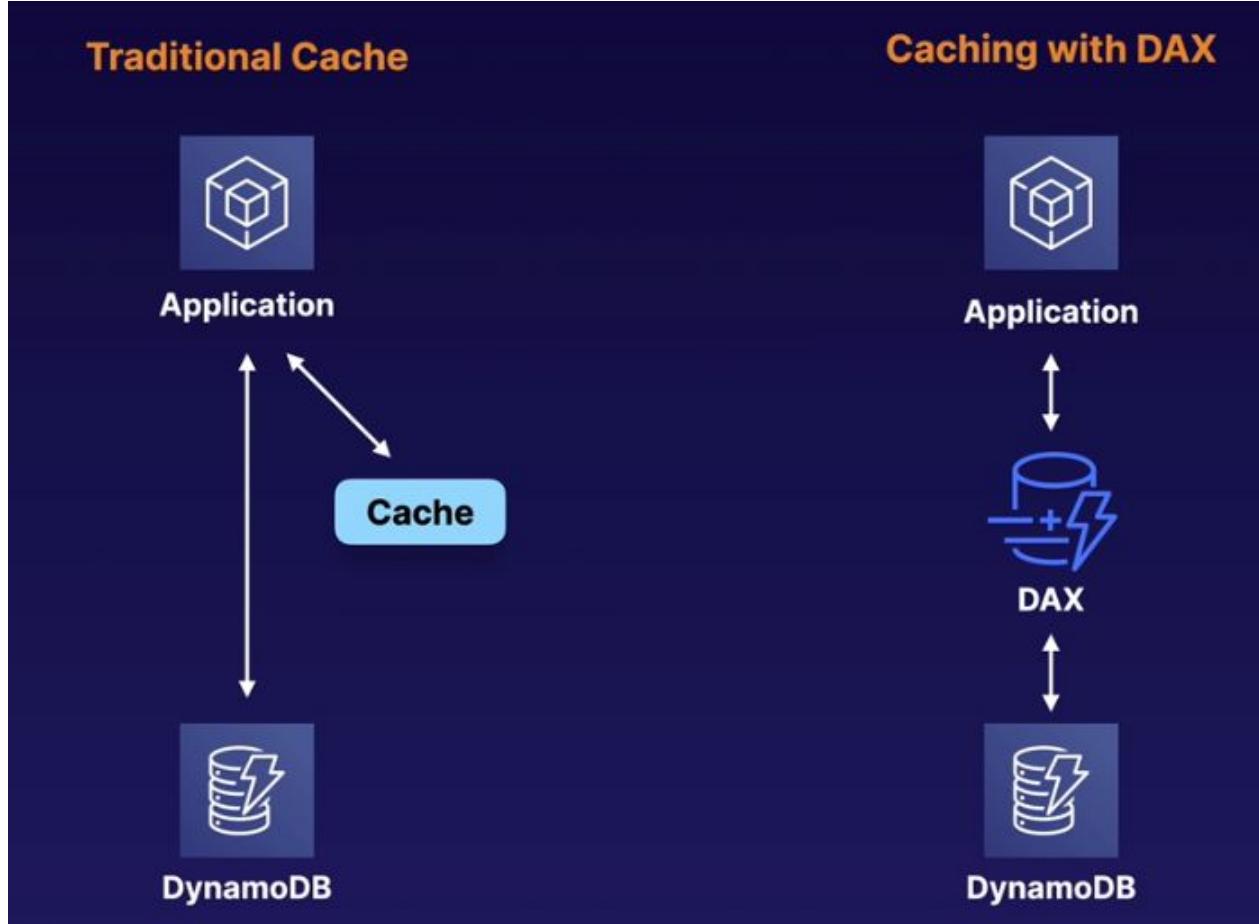
# **Advanced DynamoDB**

# 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

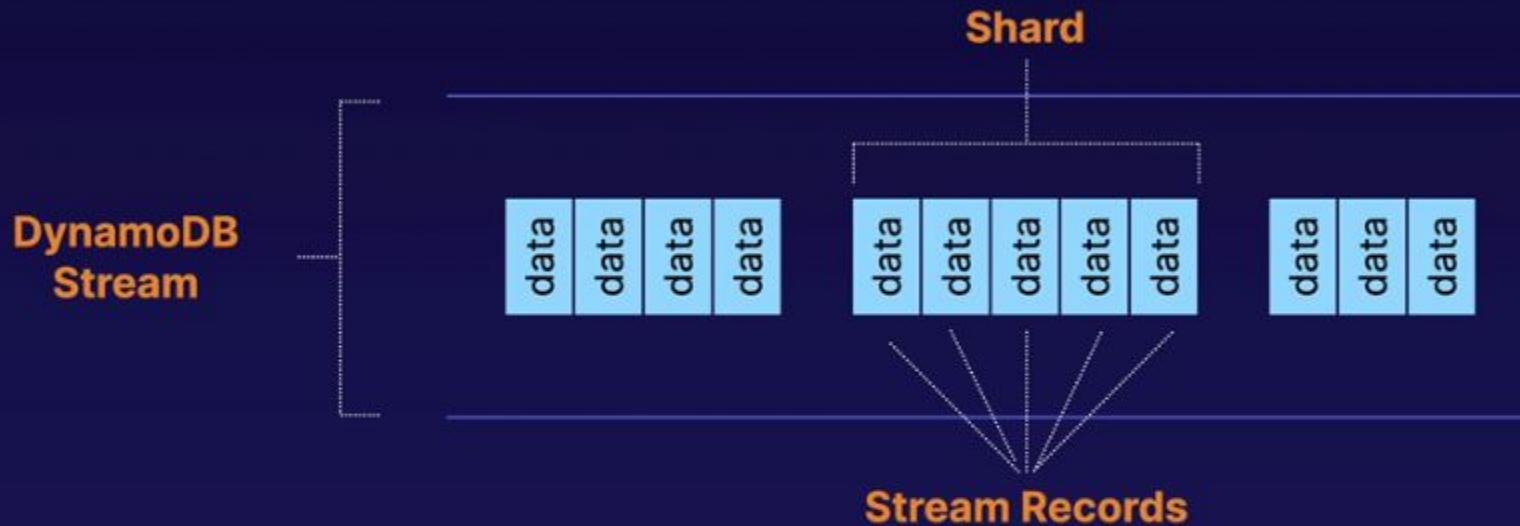
## Point-in-Time Recovery (PITR)

- 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

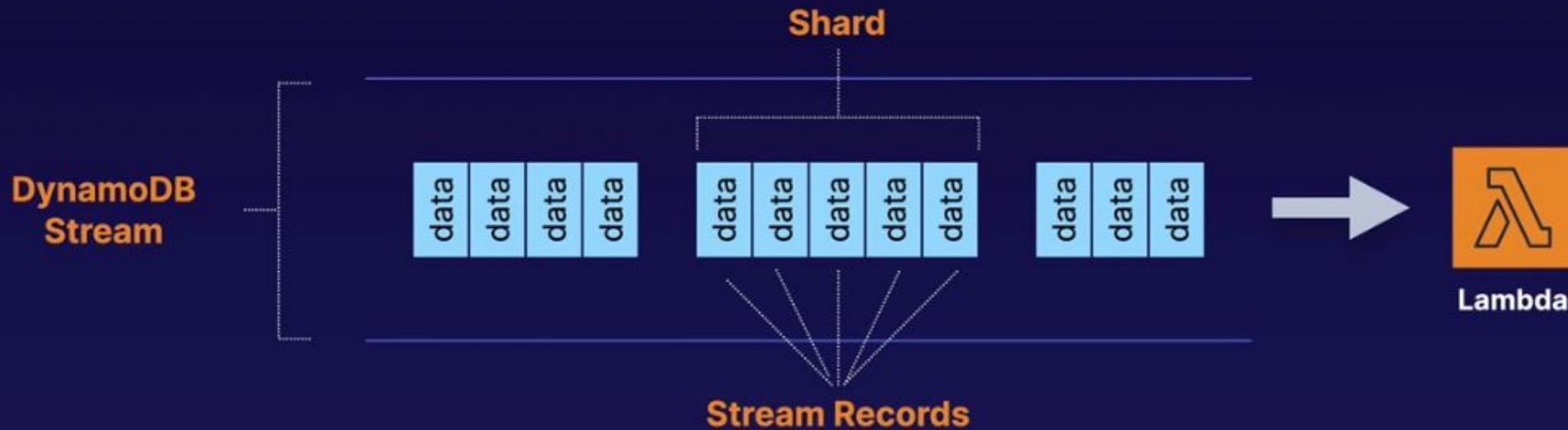
- Time-ordered sequence of item-level changes in a table



- 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", "Support", and user information "mrichman @ mark-richman".

The main menu includes "Delete table", "Overview", "Items", "Metrics", "Alarms", "Capacity", "Indexes", "Global Tables", "Backups", and "More".

A message states: "Global Tables enable you to use DynamoDB as a fully-managed, multi-region, multi-master database. Learn more".

A warning message: "⚠ 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.".

Table settings:

- KMS Customer: No (green checkmark)
- Streams: Disabled (red exclamation mark)
- Stream type: -
- Enable streams button

IAM role: AWS Lambda (selected)

Global Table settings:

- View type: New and old images - both the new and the old images of the item (radio button selected)
- Cancel and Enable buttons

Region settings:

- Add region and Delete region buttons
- Region Name dropdown
- Status, Read capacity units, Write capacity units, Auto Scaling, Endpoint tabs

Message at the bottom: "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", "Support", and user information "mrichman @ mark-richman".

The main menu includes "Create table", "Delete table", "Filter by table name" (with a search bar), "Choose a table group" (dropdown), "Actions" (dropdown), and "No Tables".

A message about DynamoDB: "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 Global Tables console. 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 console. The table list table has the following columns and data:

Name	Status	Partition key	Sort key

At the top, there are buttons for "Create table" and "Delete table". Below the table, there is a descriptive text box:

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 replication latency metrics for a global table.

**Add replica to global table**

Current region: US East (Ohio)

Global table version: 2019.11.21

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

Table details:

- 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

CloudWatch Metrics section:

- View all CloudWatch metrics
- Time Range: Last Hour
- Global table metrics: Replication latency (Milliseconds)

CloudWatch Metrics chart (Partial View):

Time	Latency (Milliseconds)
1	0.801
0.601	

Screenshot of the AWS DynamoDB Tables interface showing the creation of a new table.

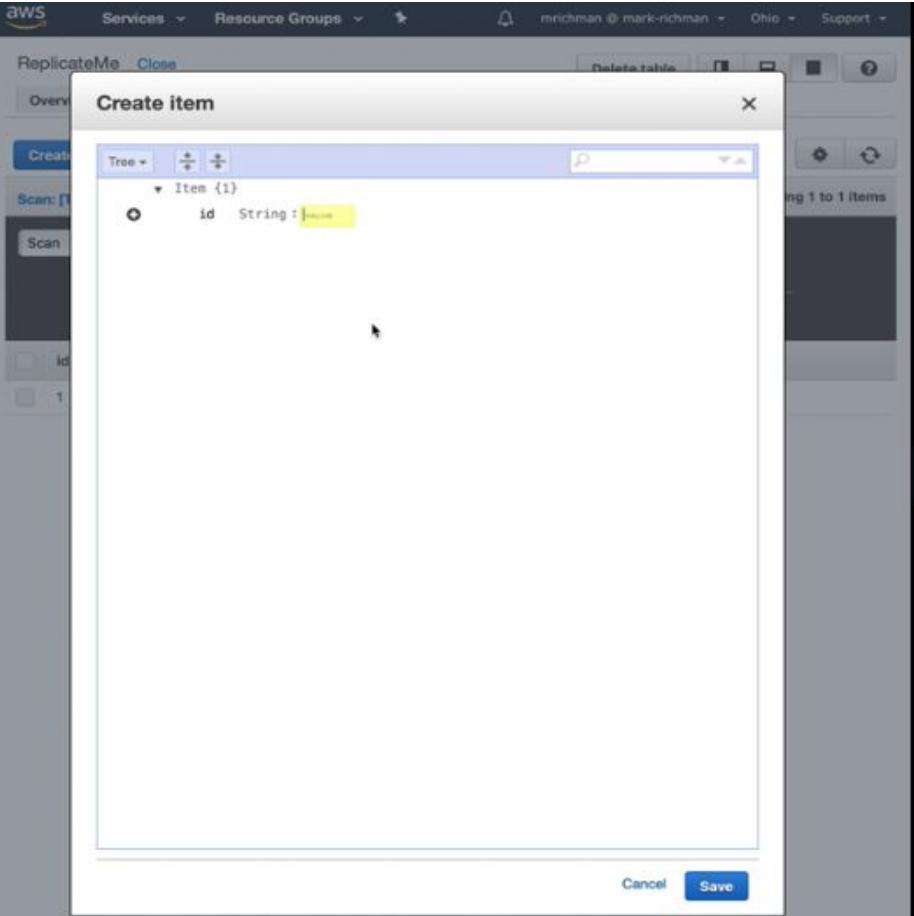
Create table

Filter by table name: ReplicateMe

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 one item with the following data:

	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

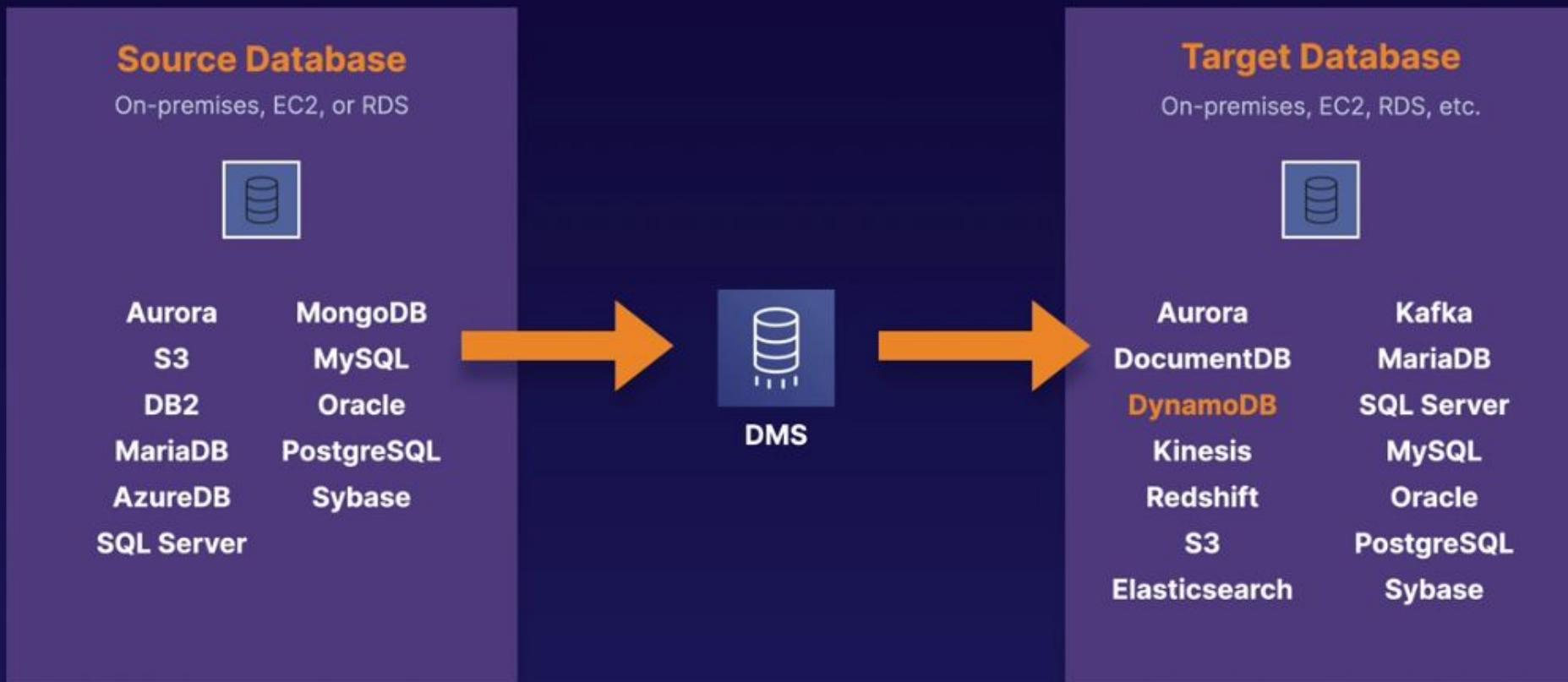
# Global tables

## 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:

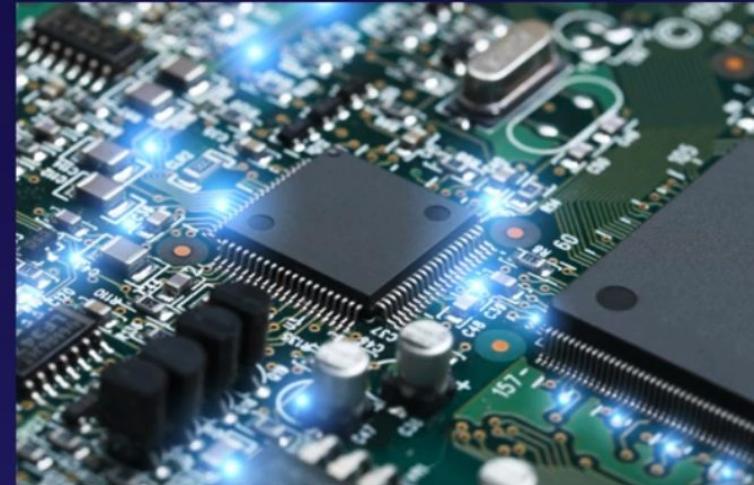
Columnar data stores can be compressed much more than row-based data stores because similar data is stored sequentially on disk. Amazon Redshift employs multiple compression techniques and can often achieve significant compression relative to traditional relational data stores. In addition, Amazon Redshift 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)



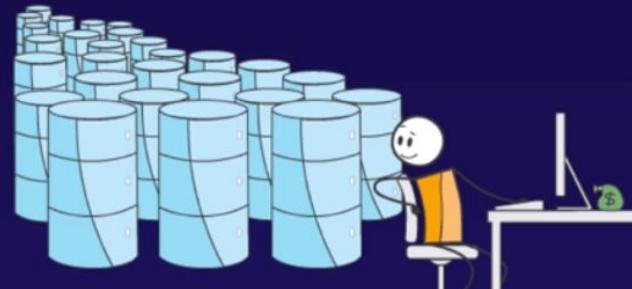
# Aurora

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

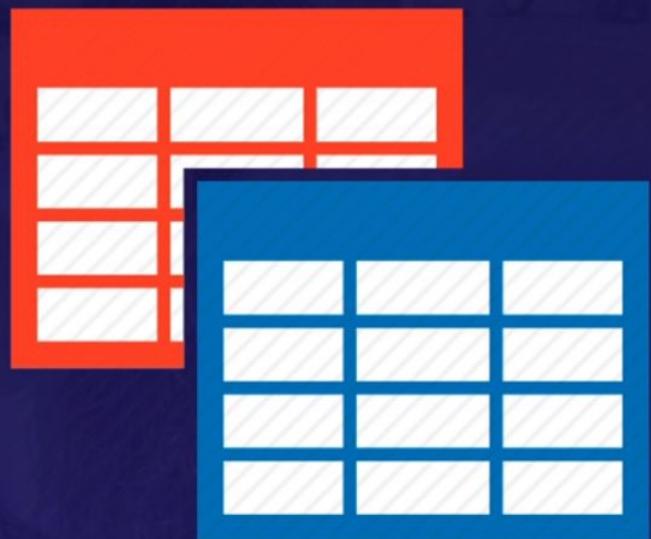
# Elasticache Exam tips

- 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

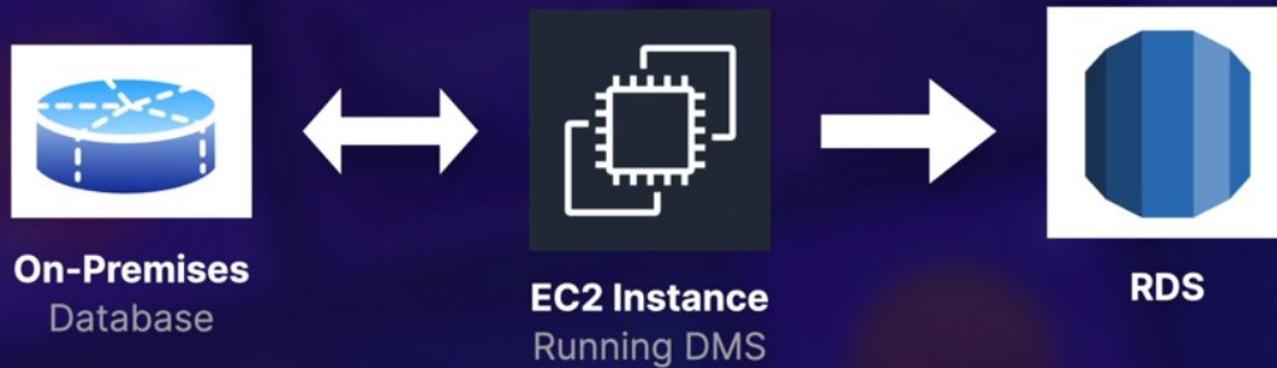
- ✓ On-premises and EC2 instances databases: Oracle, Microsoft SQL Server, MySQL, MariaDB, PostgreSQL, SAP, MongoDB, Db2
- ✓ Azure **SQL** Database
- ✓ Amazon **RDS** (including Aurora)
- ✓ Amazon S3

## 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

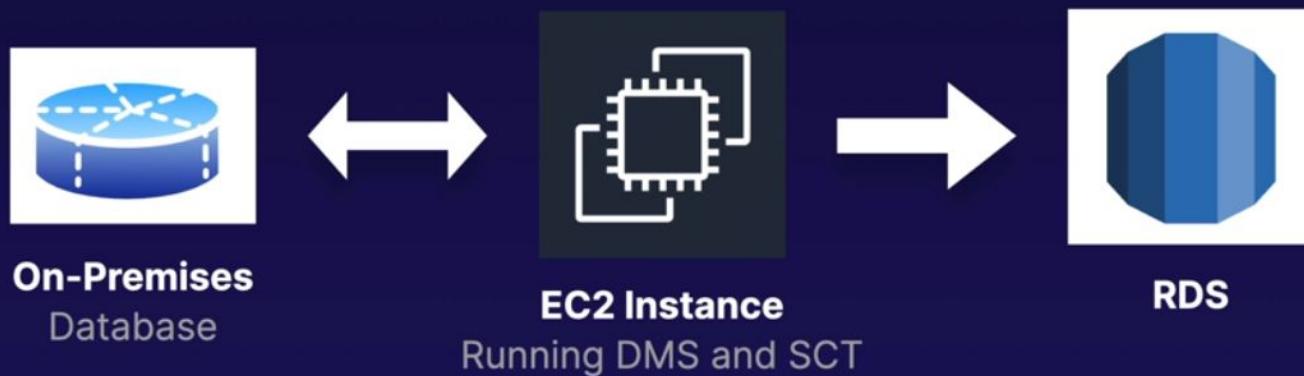
# DMS in action

## 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

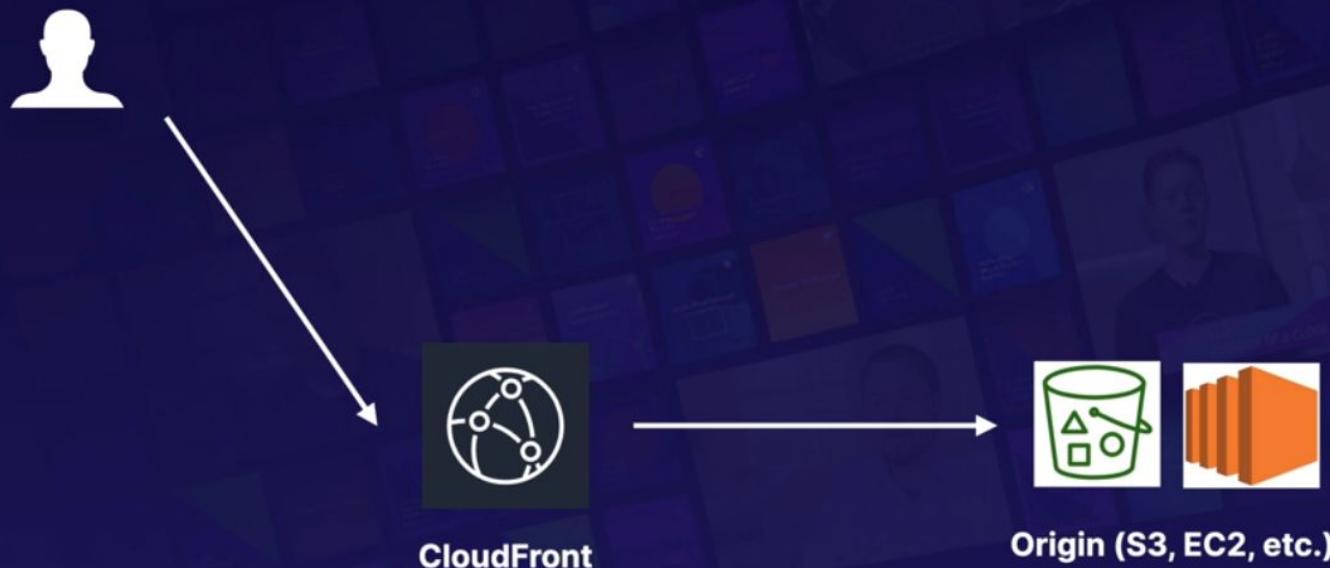
# Caching services

The following services have caching capabilities:

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

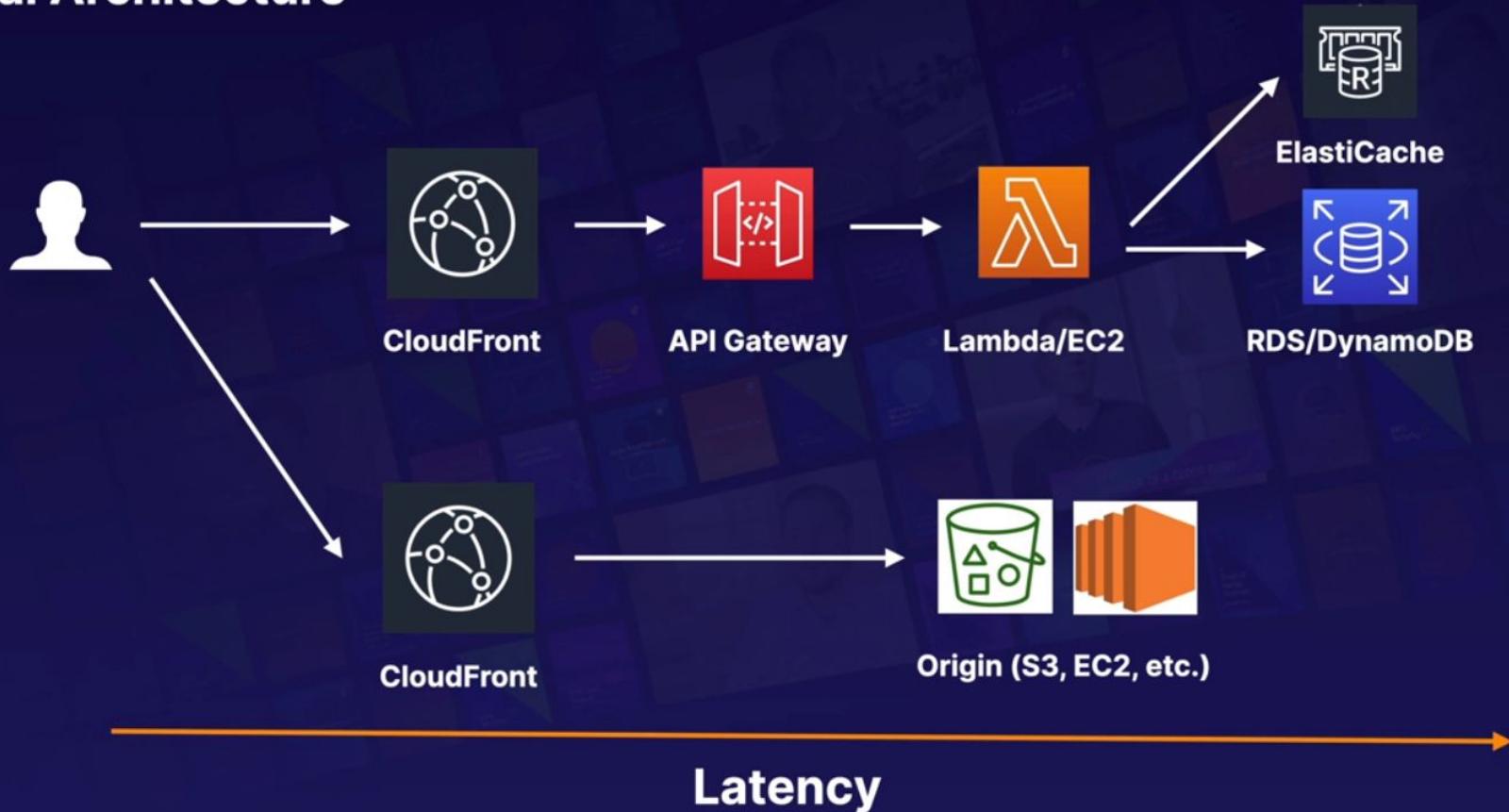
# Caching services

## 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.



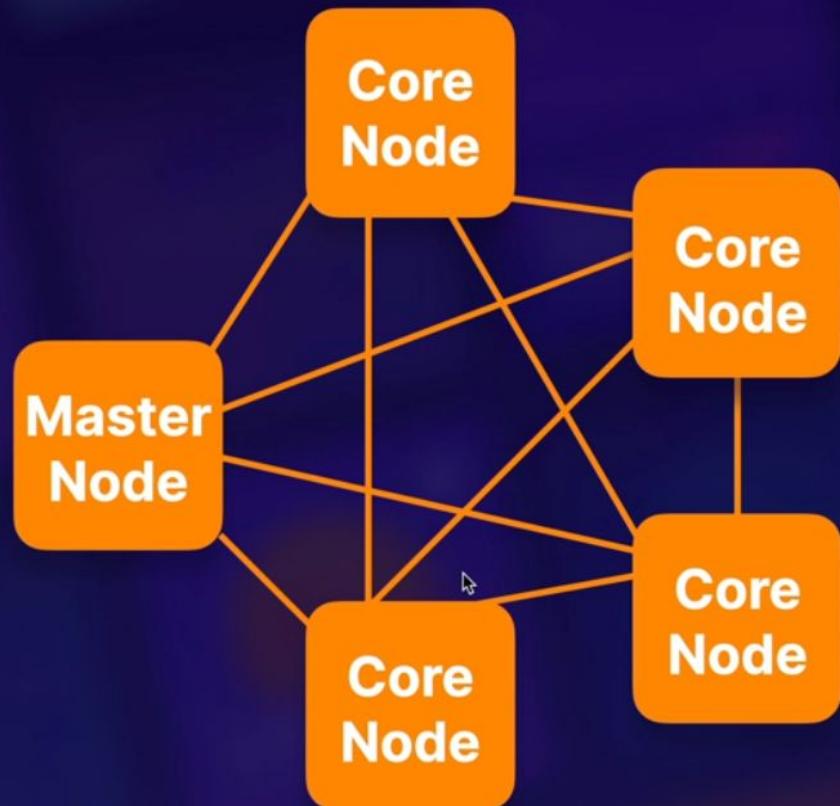
**Core node:** A node with software components that **runs tasks and stores data** in the Hadoop Distributed File System (HDFS) on your cluster. Multi-node clusters have at least one core node.



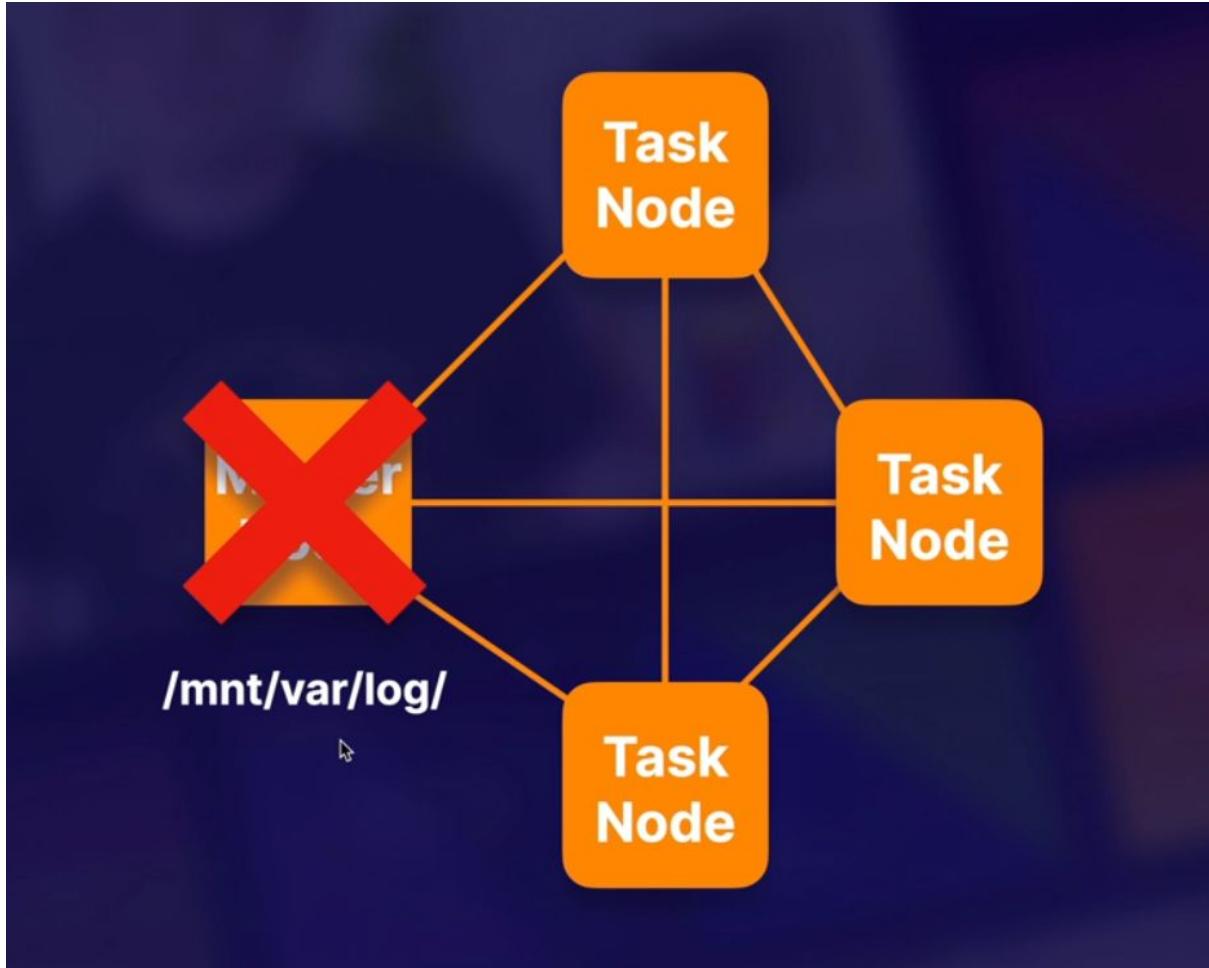
**Task node:** A node with software components that only runs tasks and **does not store data in HDFS**. Task nodes are **optional**.

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

• Redshift

Algunas de las más conocidas son por su uso en el comercio electrónico y la banca.

• MySQL es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• PostgreSQL es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• Oracle es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• Aurora es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• MariaDB es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• Redshift

Algunas de las más conocidas son por su uso en el comercio electrónico y la banca.

• MySQL es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• PostgreSQL es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• Oracle es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• Aurora es una base de datos relacional que se usa para aplicaciones web y bases de datos.

• MariaDB es una base de datos relacional que se usa para aplicaciones web y bases de datos.

# Databases Exam tips

## Elasticache

- Memcached
- Redis



Algo de algo

Leyendo una cosa es hacerla fría por las orejas.  
Lo que hay cuando la oyes, son páginas. De ahí las di-  
ctadas, las copias, los errores.

Rodríguez, Laura, 2008

Nosotros ésta pasa escribe a otros, y nos se viven en el mundo  
nuestros deseos. De lo que pasan, con sus miedos y deseos.  
Lo que pasa depende que nos pase lo que nos pase, que es cosa  
nuestra de dentro... somos en el campo.

Martínez, M., 2008

Algunas de las personas la anterior disposición a aprender tiene  
que ver con el deseo de ser mejores.

Algunas que están dispuestas a decir para luego de las pro-  
fesiones, las personas que tienen que ver con el deseo de ser

Más o menos, si

Y trae más que libro en contra de tierra en España.  
Leyendo quedas sin memoria, casi siempre, afinal, dice  
«No». Hay muchísimo más tristes que lecturas en la  
Universidad española.

Y es que la Universidad en muchos tristes que las  
imponen para que lo escribas, tanto a pregunta Marca.

Otro, te has pasado diez páginas de digo, mierda  
lamentar un bocón de la tristeza de lo que más me  
mejor sucede sobre la cosa fría.

Martínez, M., 2008

Algunas que tienen que ver con el deseo de ser mejores.

Leyendo quedas sin memoria.

# 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

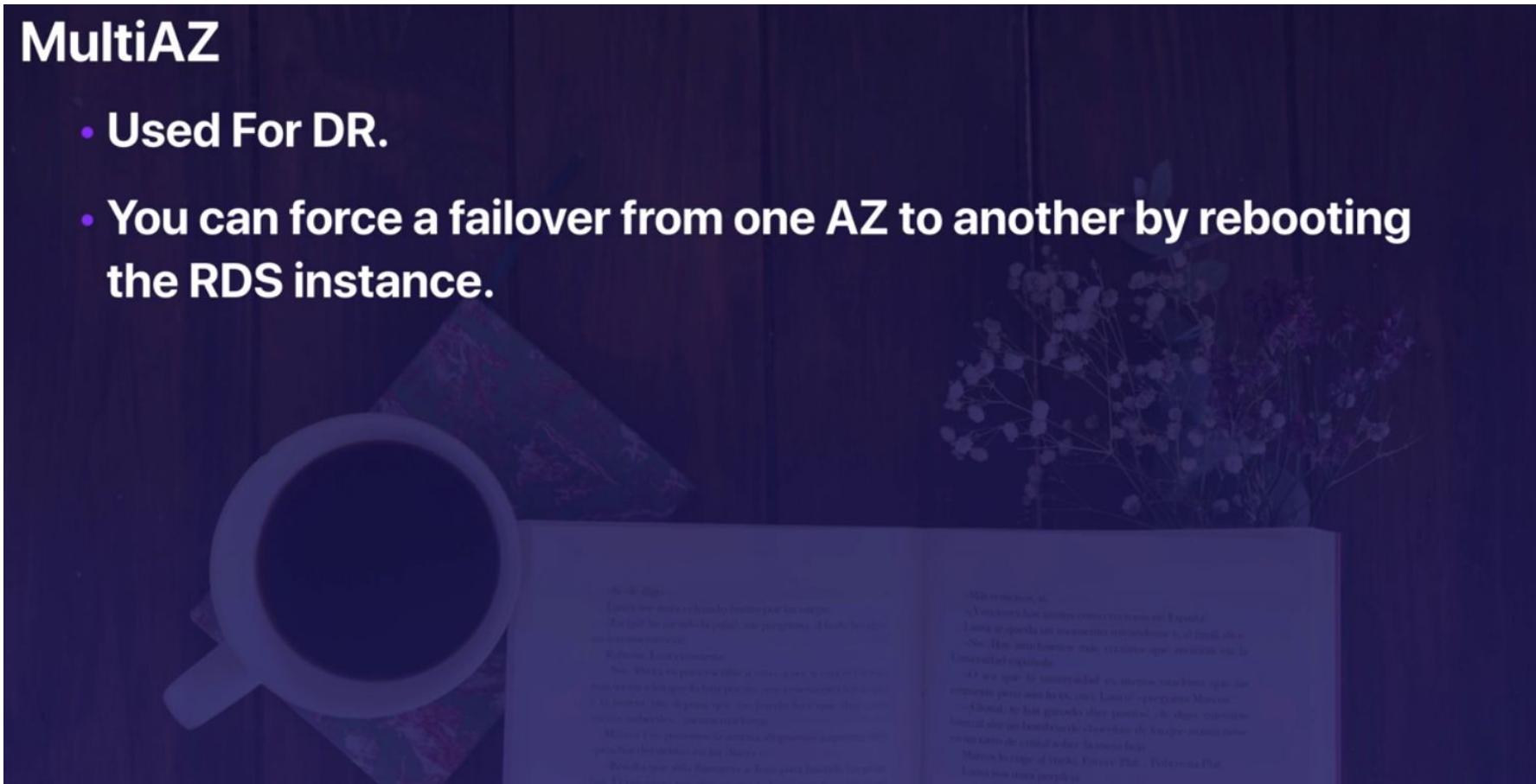
## Read Replicas

- 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

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

# 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**

# **CHAPTER 6**

# **Advanced IAM**

# **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

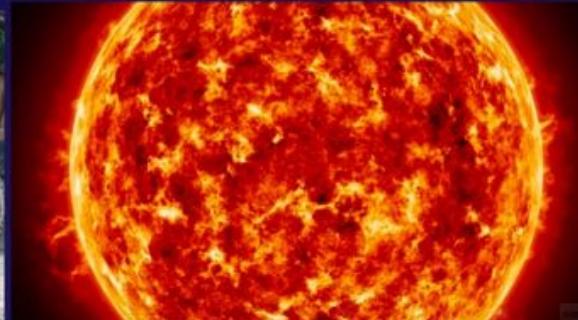
# DNS 101 :: IPV4 vs IPv6

## 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>

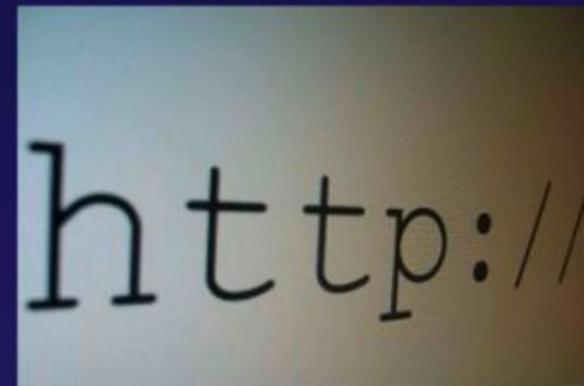


# DNS 101 :: Domain Registrars

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

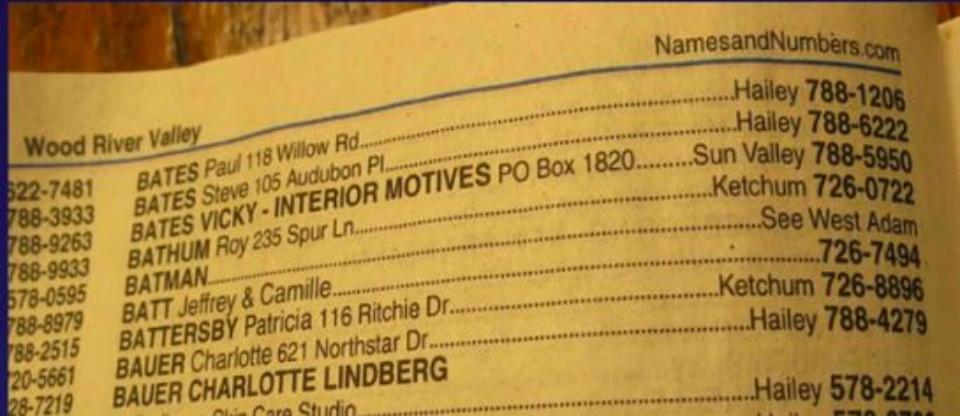
Type: CNAME – Canonical name

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



# **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.



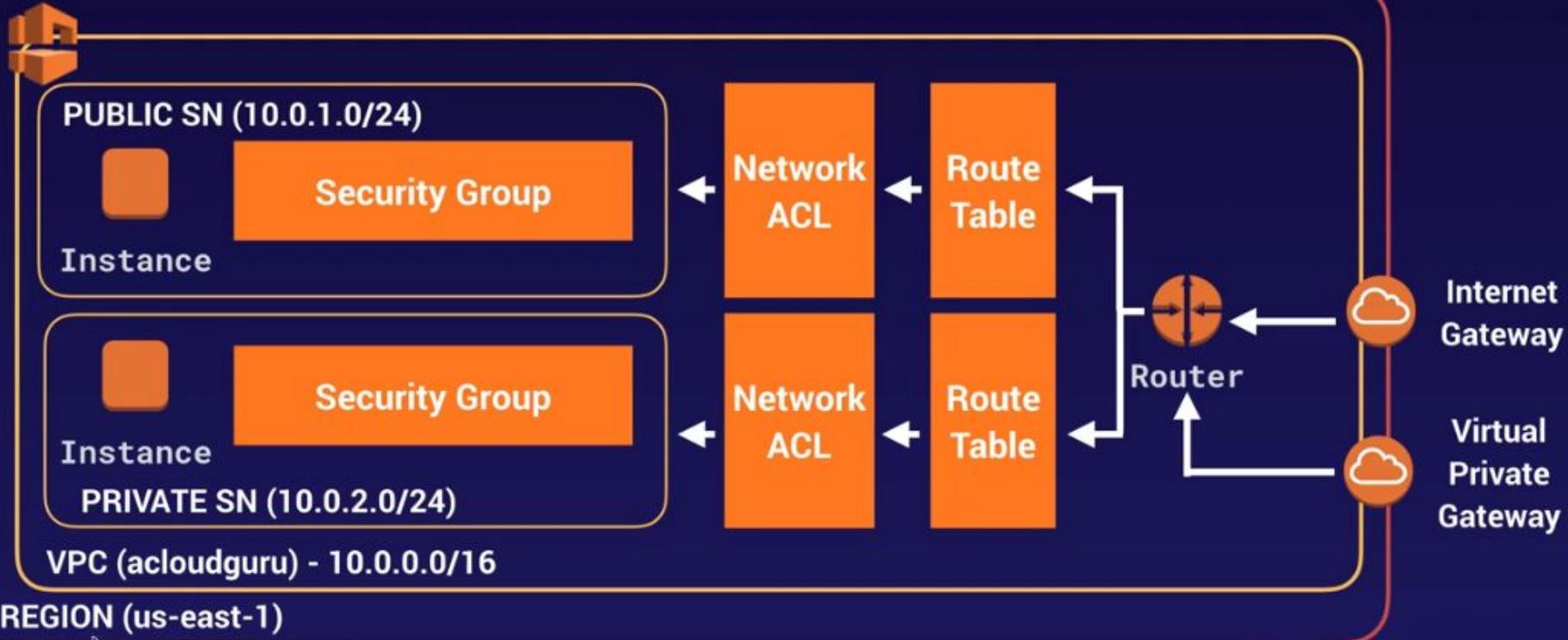
# VPC

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

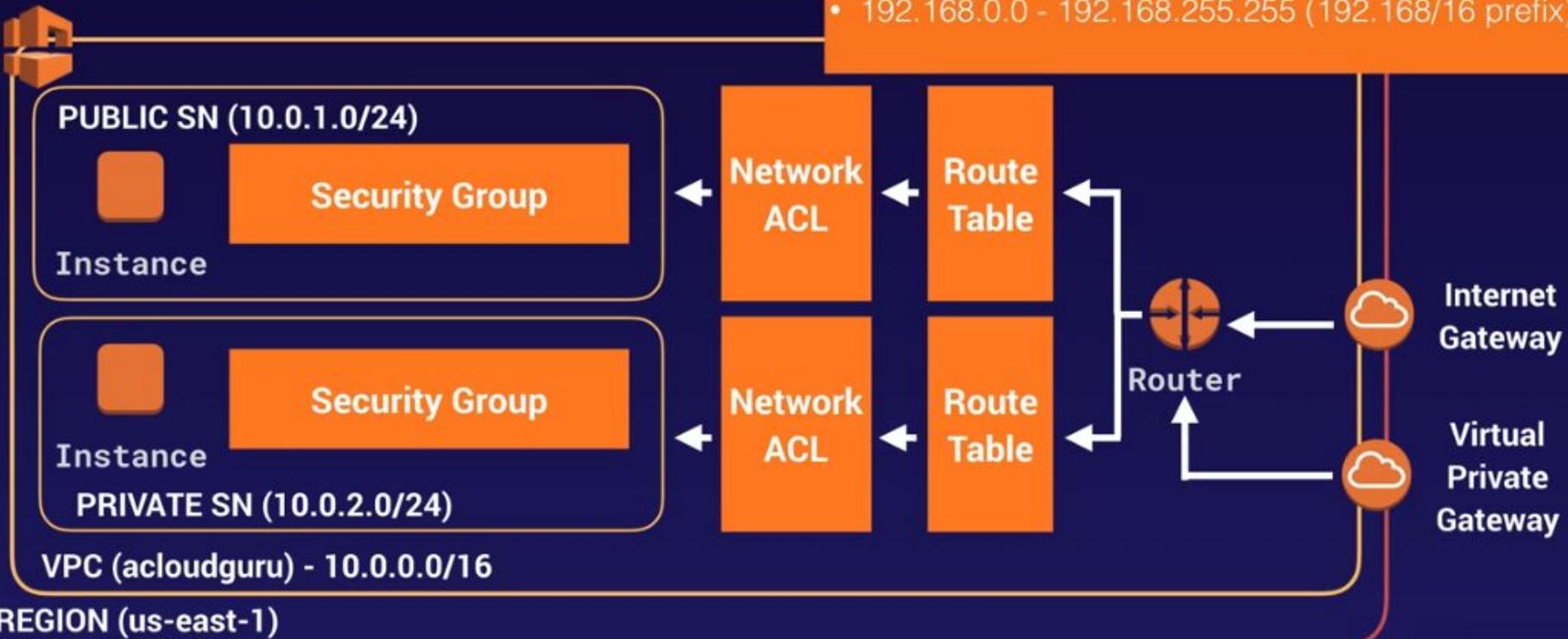
## VPC with Public & Private Subnet(s)



# 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)



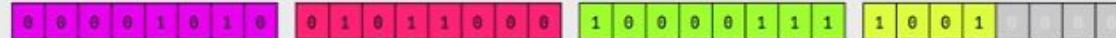
# VPC

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

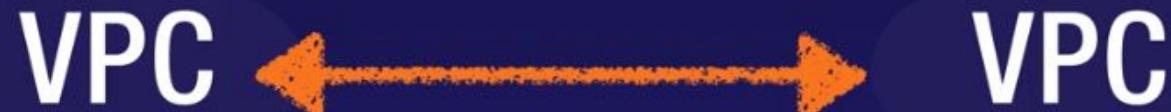
- Default VPC is user friendly, allowing you to immediately deploy instances.
- All Subnets in default VPC have a route out to the internet.
- Each EC2 instance has both a public and private IP address.

VPC

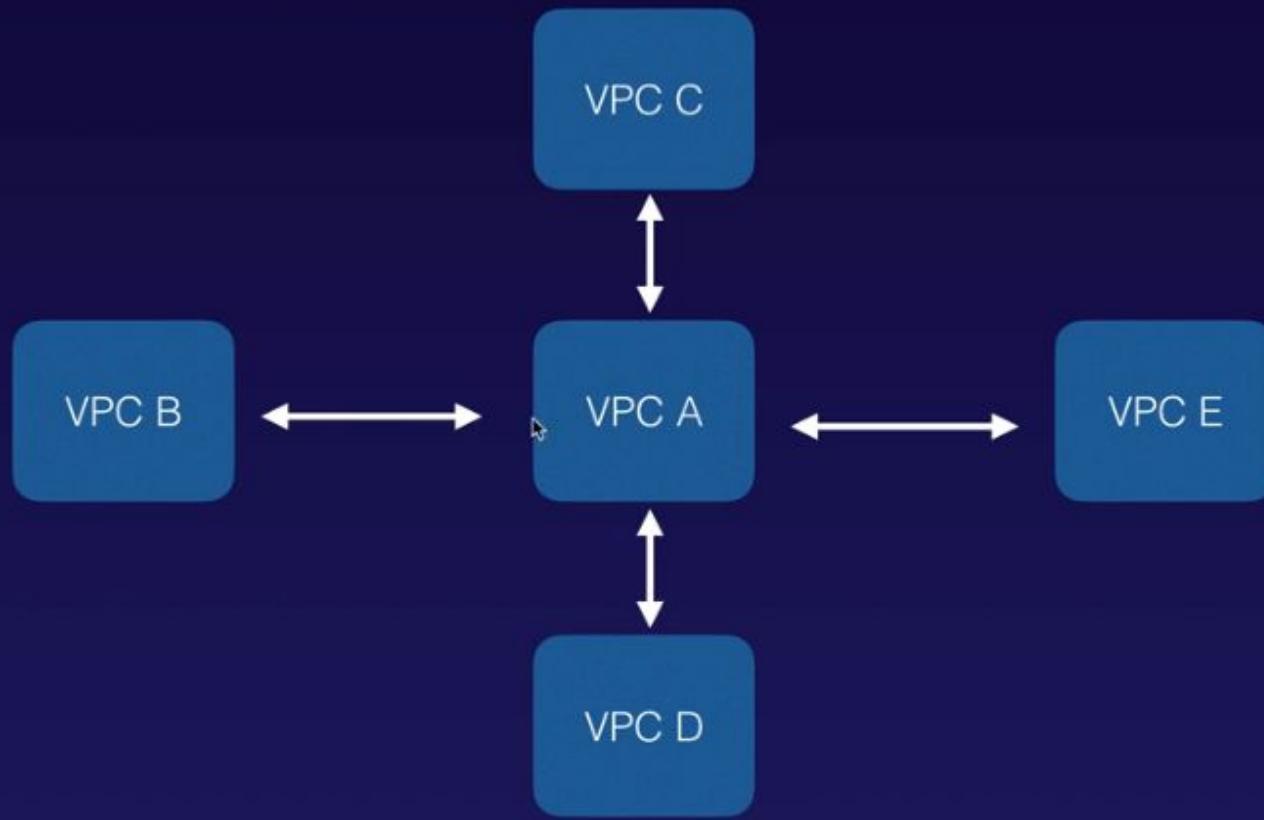
# VPC peering

## VPC Peering

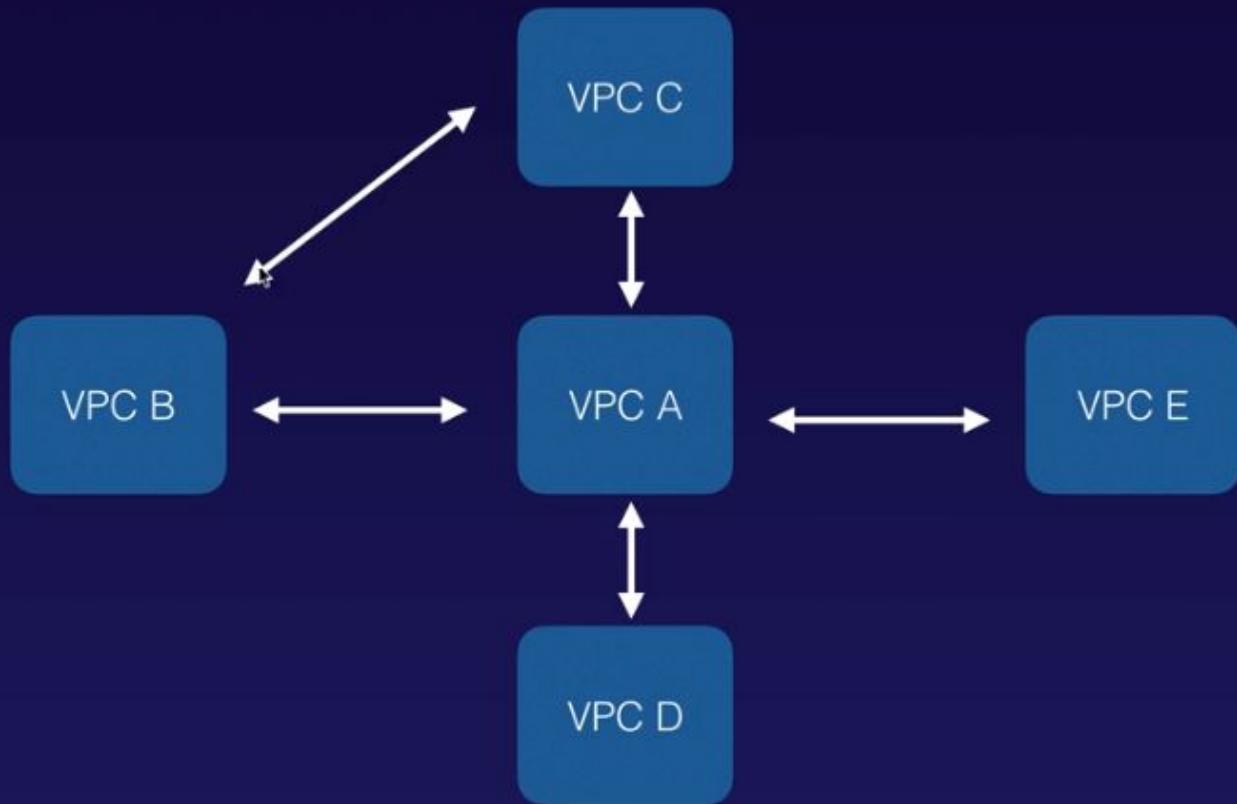
- Allows you to connect one VPC with another via a direct network route using private IP addresses.
- Instances behave as if they were on the same private network
- You can peer VPC's with other AWS accounts as well as with other VPCs in the same account.
- Peering is in a star configuration: ie 1 central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



# VPC peering



# VPC peering



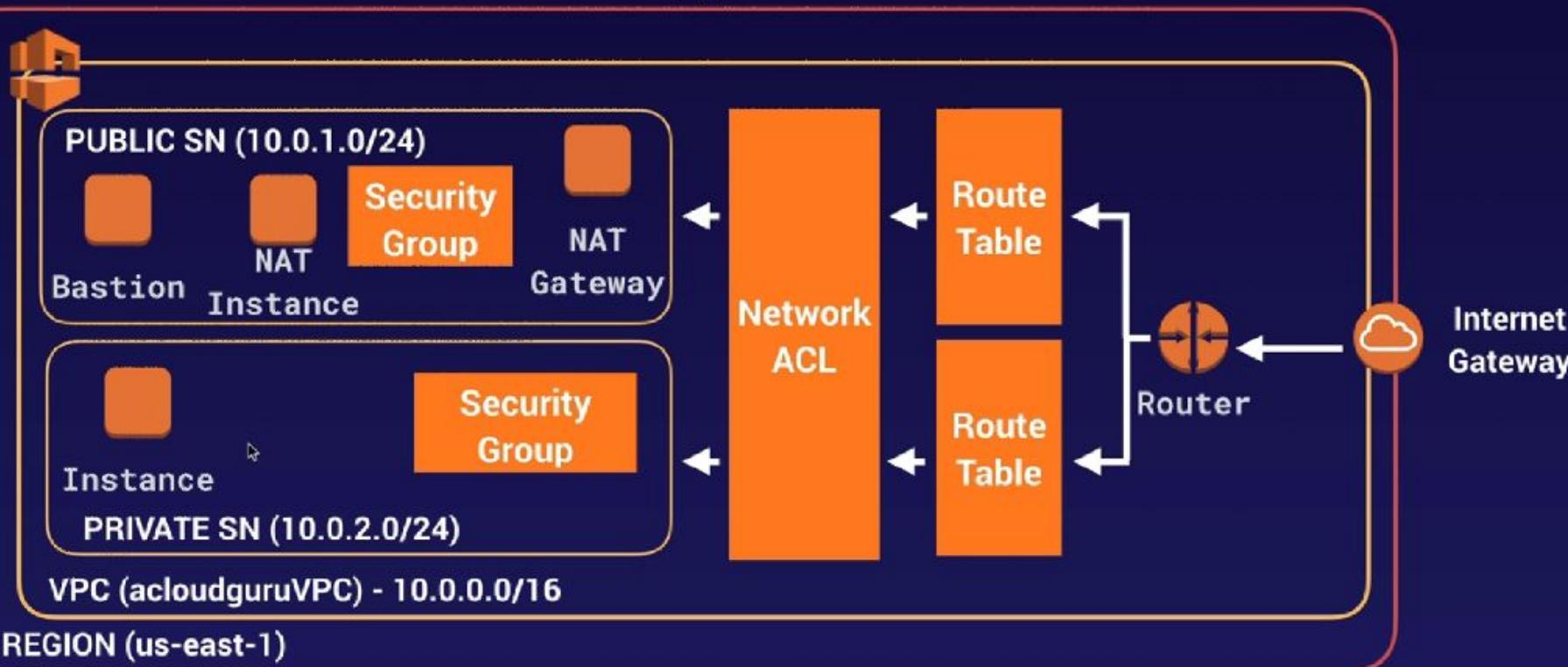
# Exam tips

## 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)



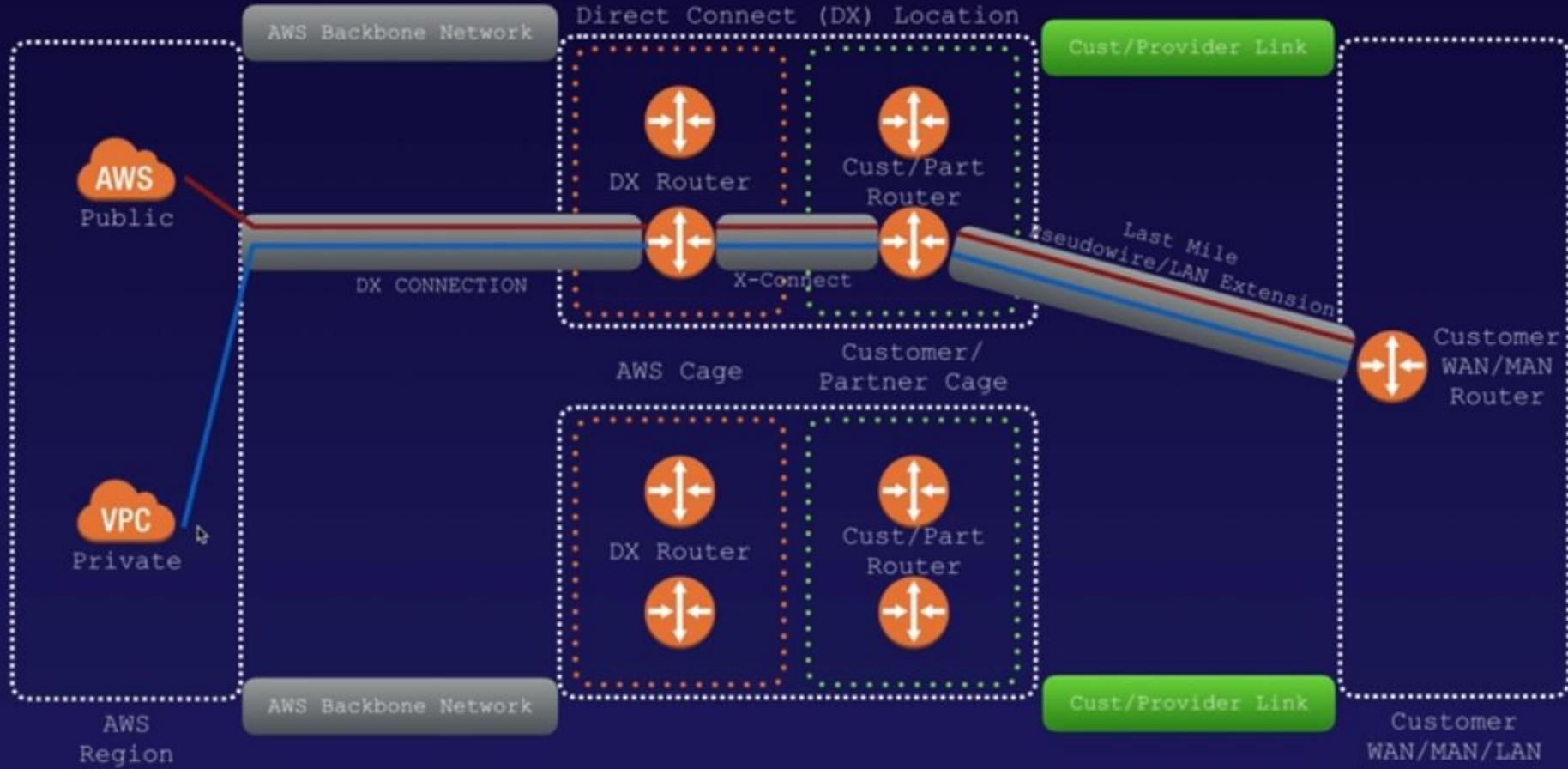
# What is Direct Connect?

## 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:

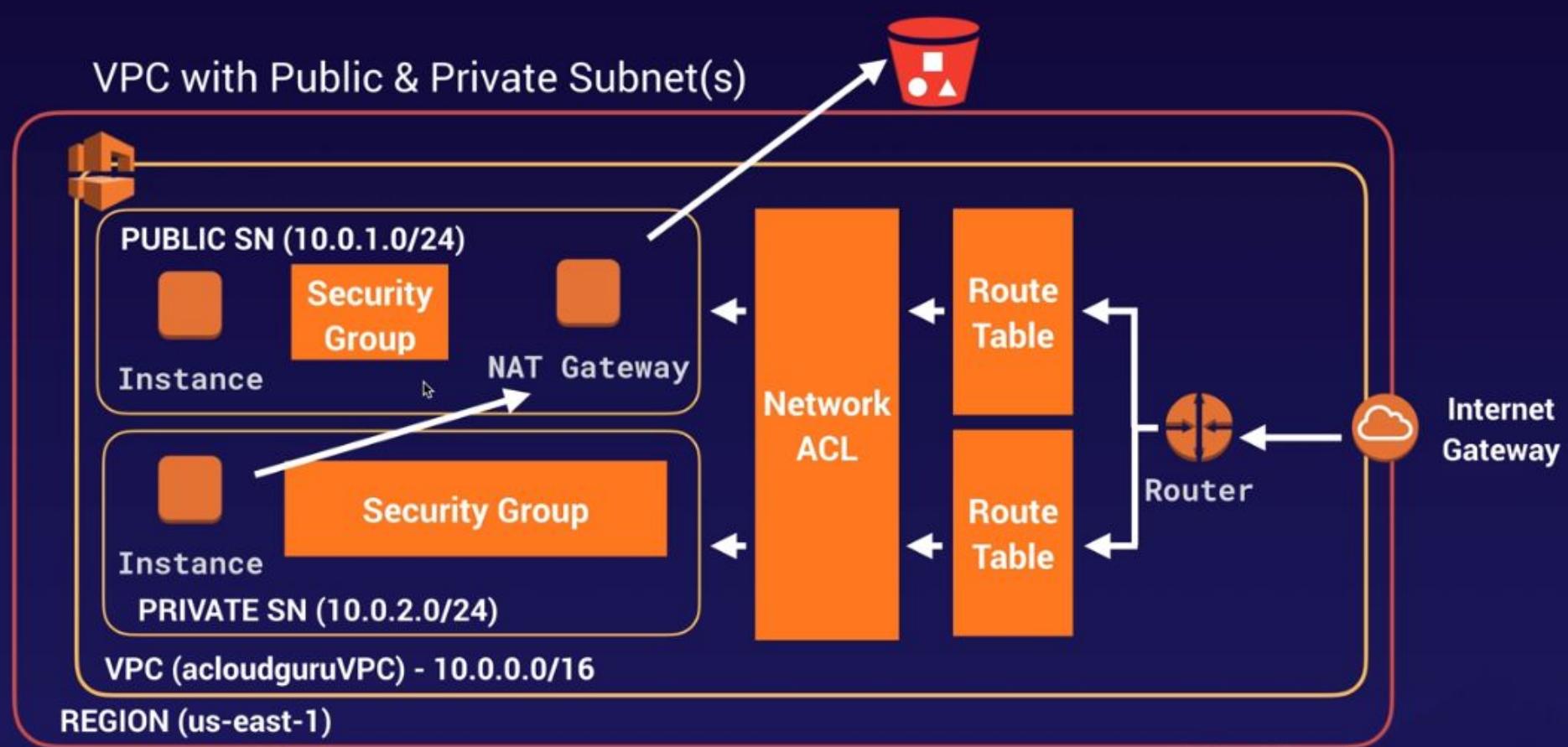
- Interface Endpoints
- Gateway Endpoints

# 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:

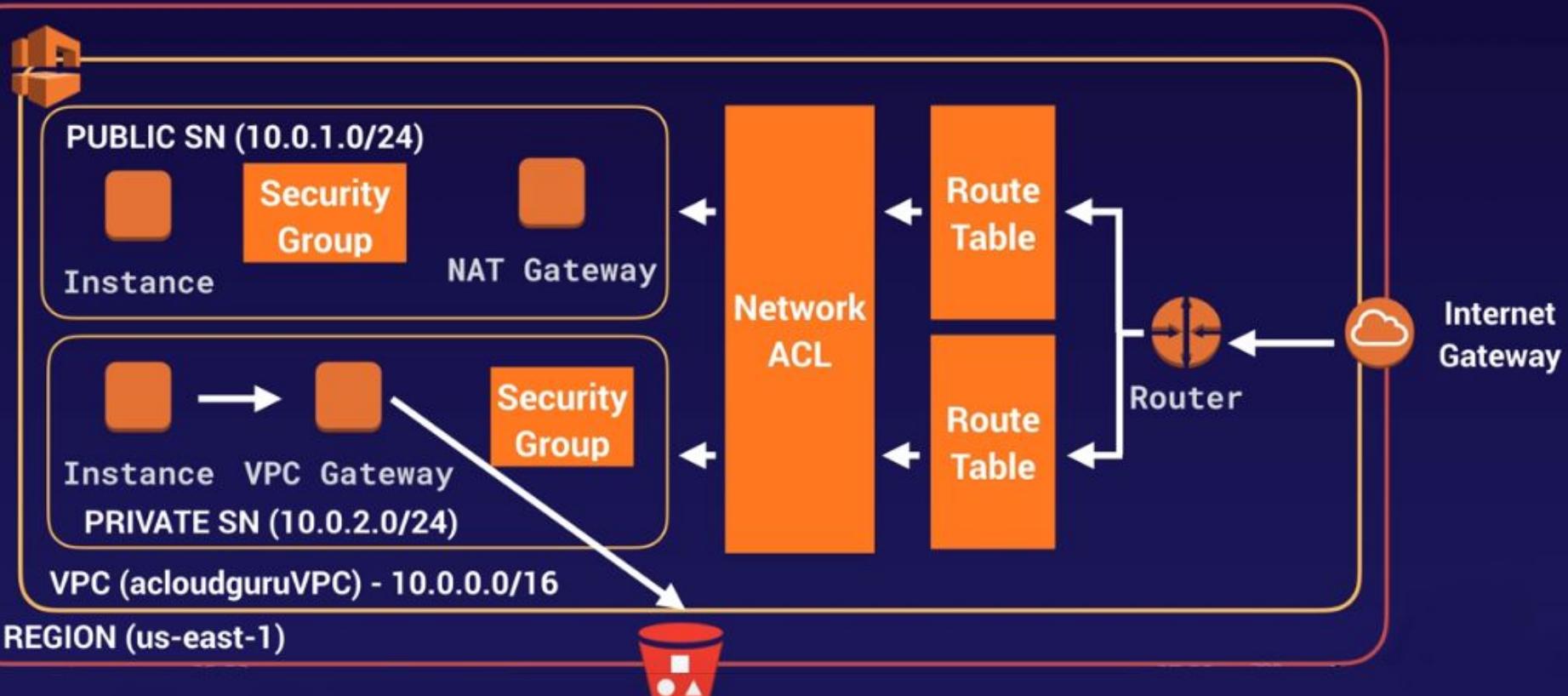
- Amazon API Gateway
- AWS CloudFormation
- Amazon CloudWatch
- Amazon CloudWatch Events
- Amazon CloudWatch Logs
- AWS CodeBuild
- AWS Config
- Amazon EC2 API
- Elastic Load Balancing API
- AWS Key Management Service
- Amazon Kinesis Data Streams
- Amazon SageMaker and Amazon SageMaker Runtime
- Amazon 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

# **CHAPTER 9**

# **HA Architecture**

# AWS PrivateLink

## **CHAPTER 10**

# **Applications**

# **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

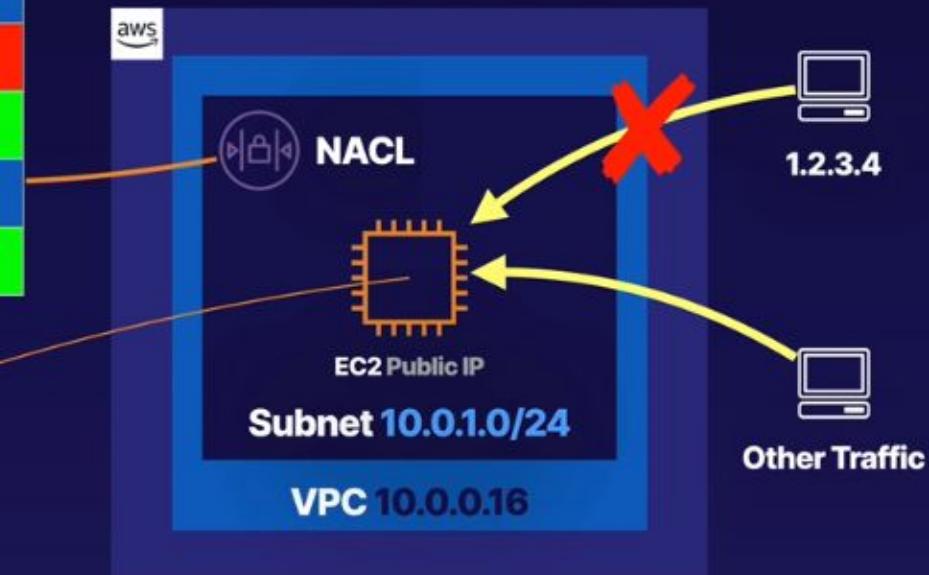


# Network Access Control (NACL)

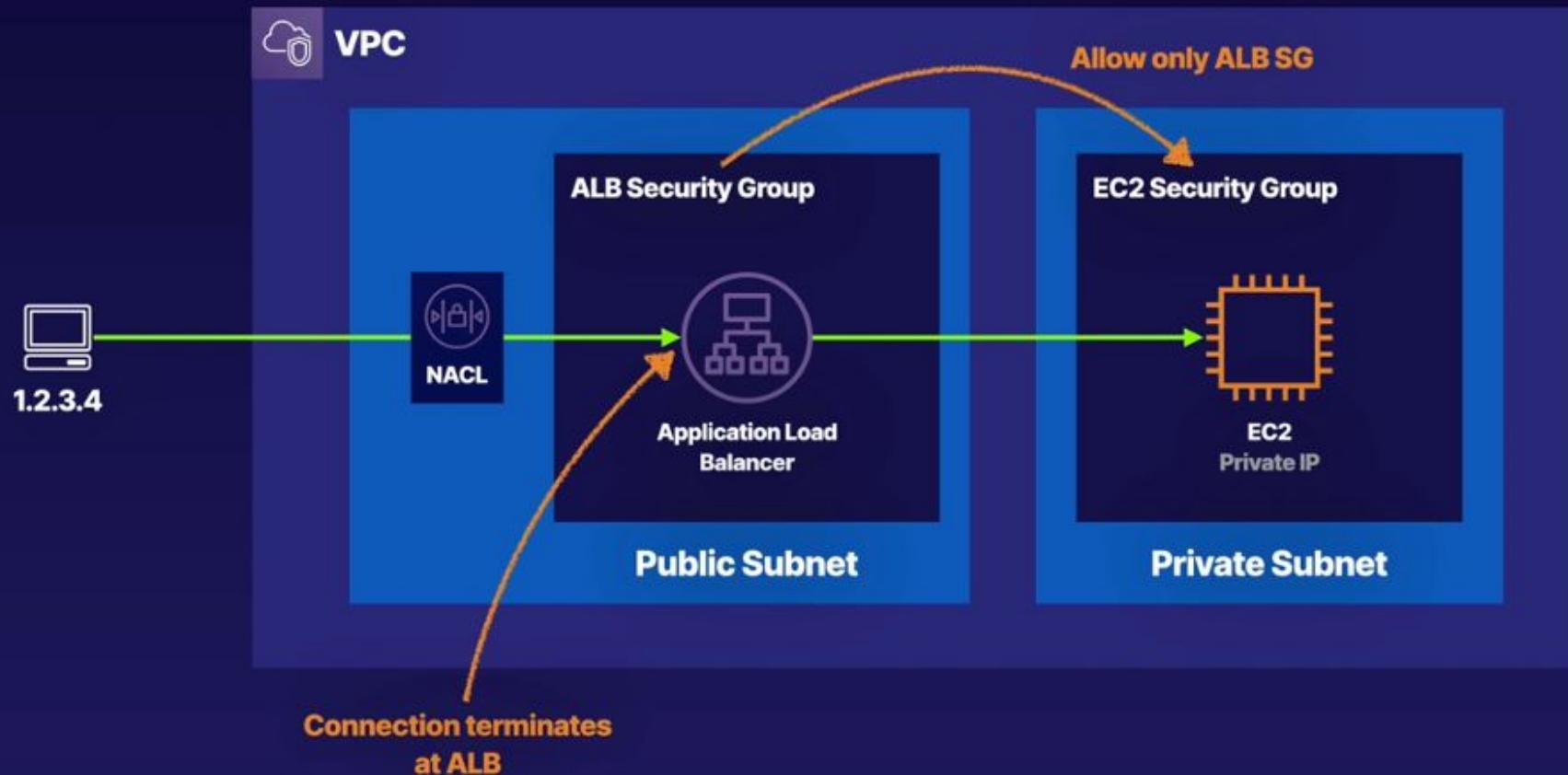
Inbound			
100	ALL Traffic	1.2.3.4	DENY
*	ALL Traffic	0.0.0.0/0	ALLOW
Outbound			
*	ALL Traffic	0.0.0.0/0	ALLOW

NACL Rules

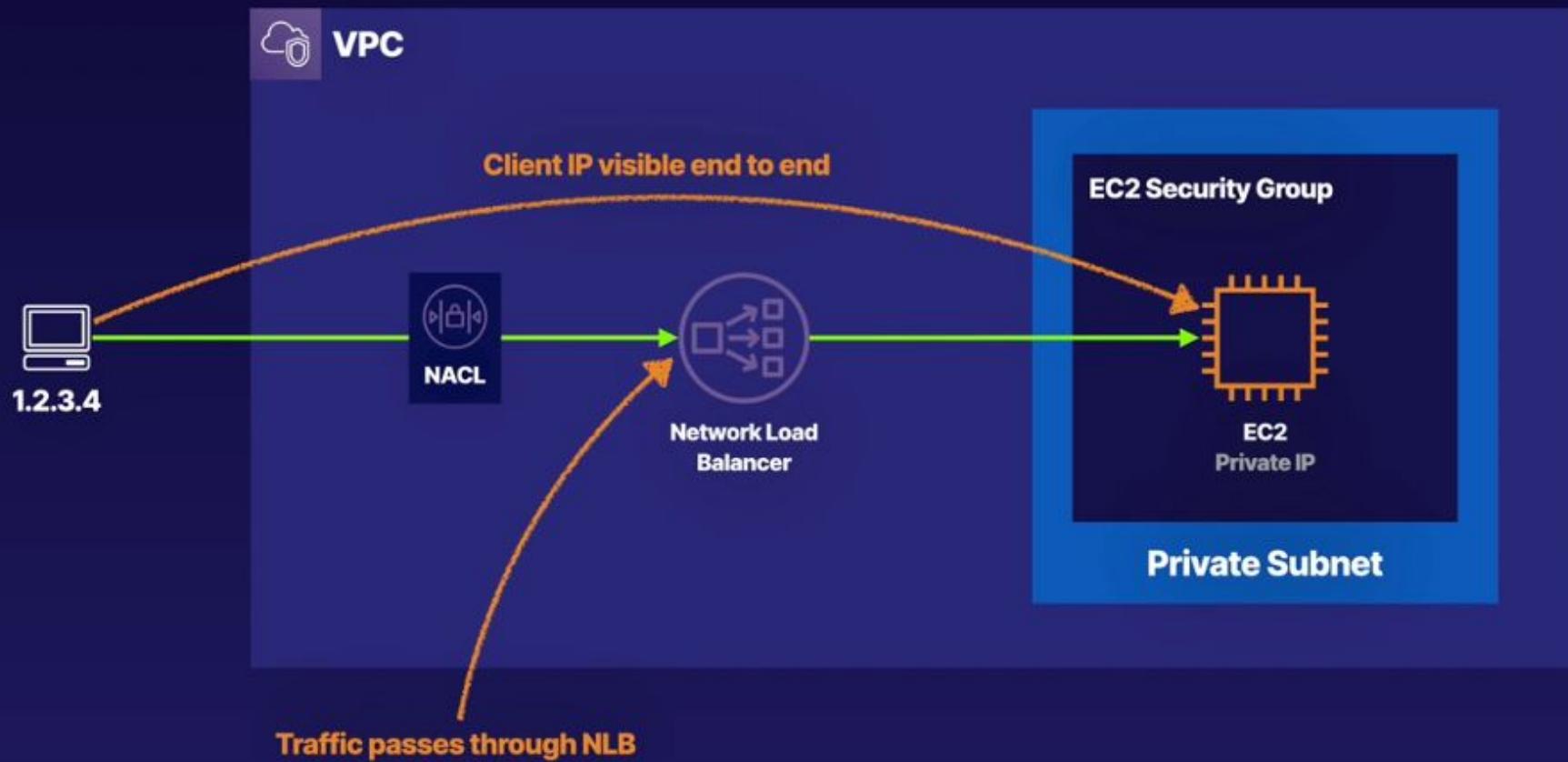
**Host-based** firewall  
e.g., firewalld, iptables, ufw, Windows Firewall



# Application Load Balancer (ALB)



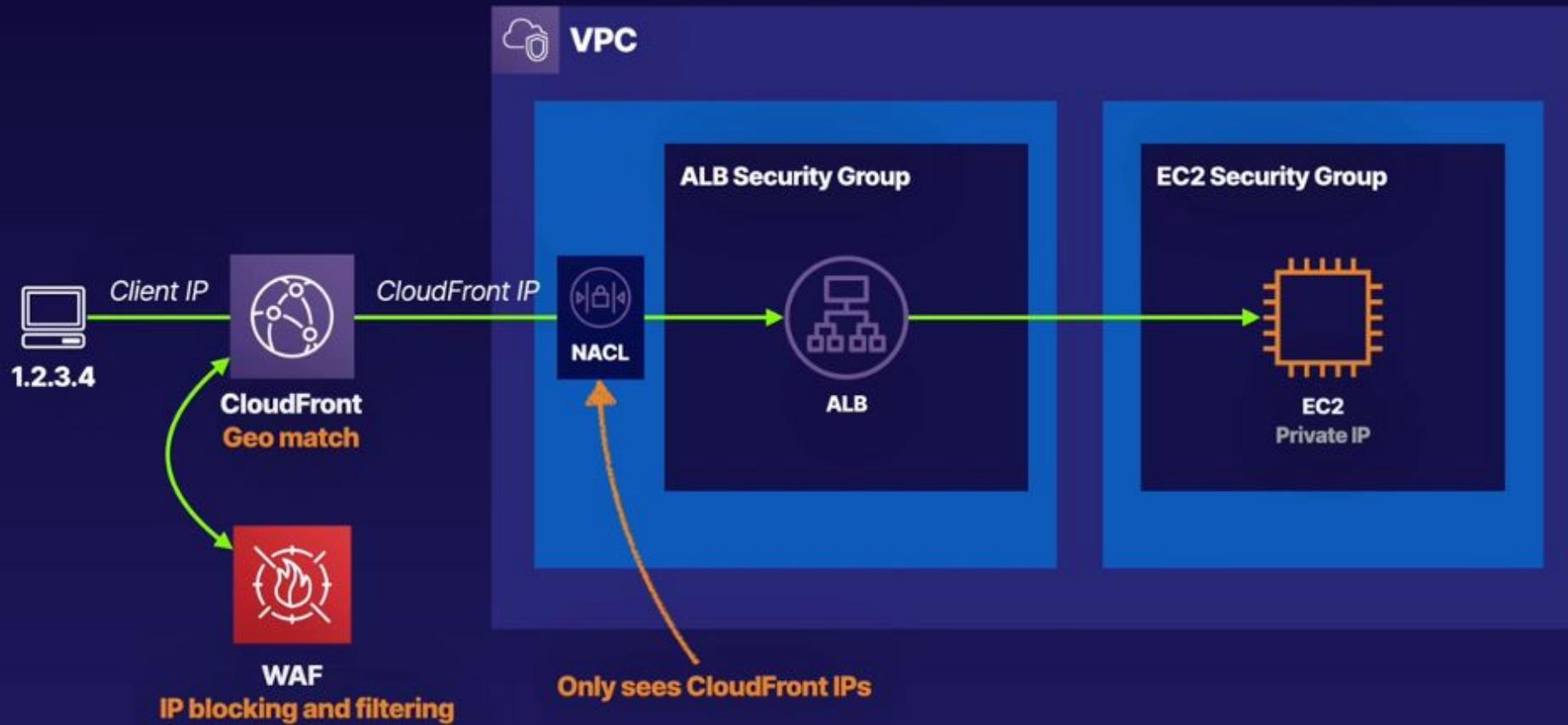
# Network Load Balancer (NLB)



# Web Application Firewall (WAF)



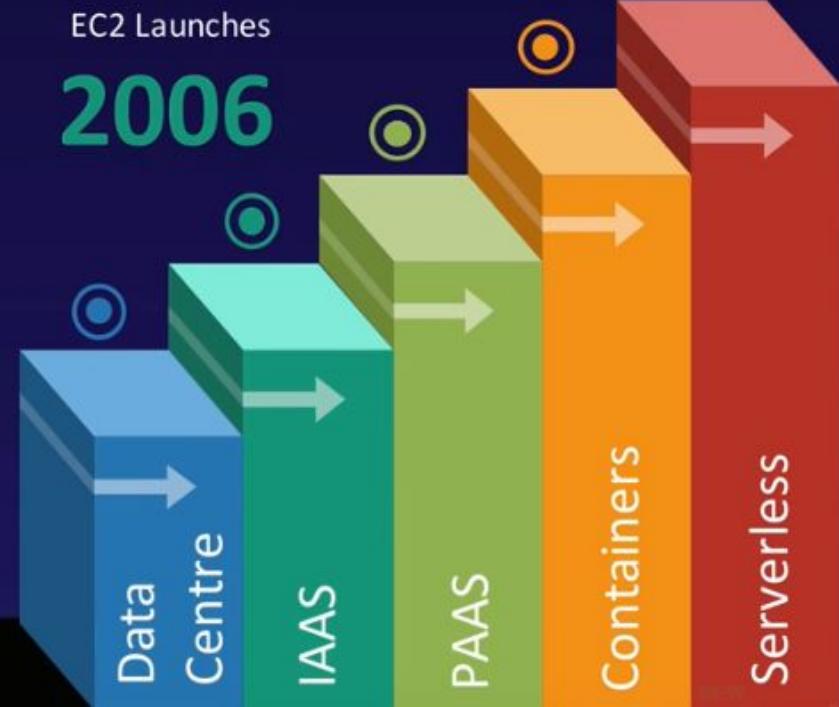
# WAF + CloudFront



# **CHAPTER 12**

# **Serverless**

# History



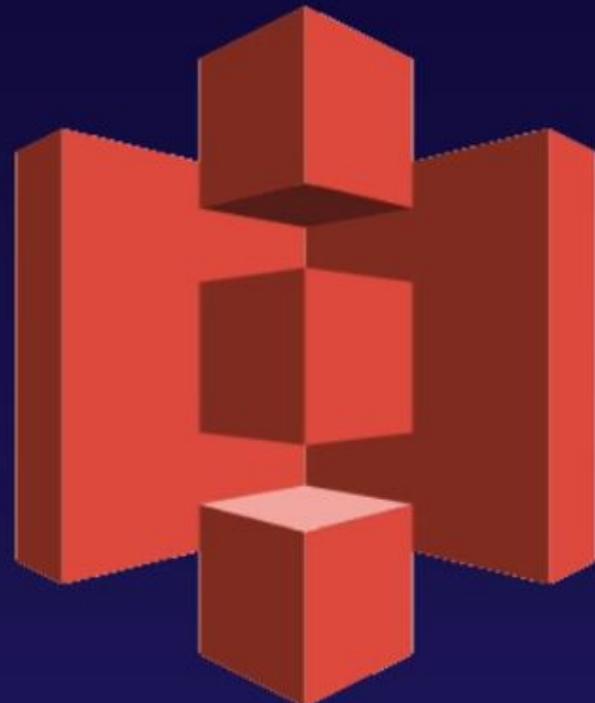
# History

**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.**

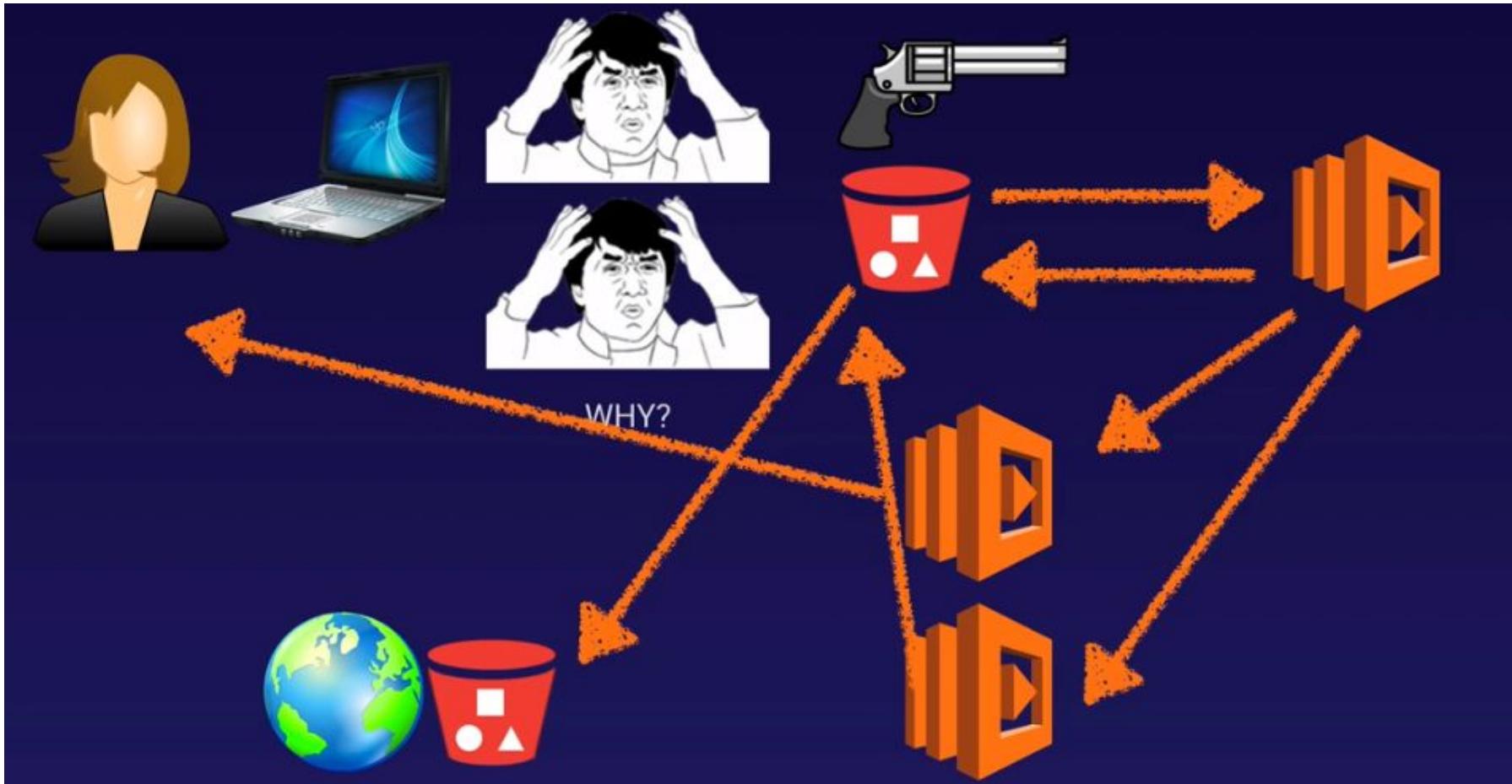


# Lambda - the basics

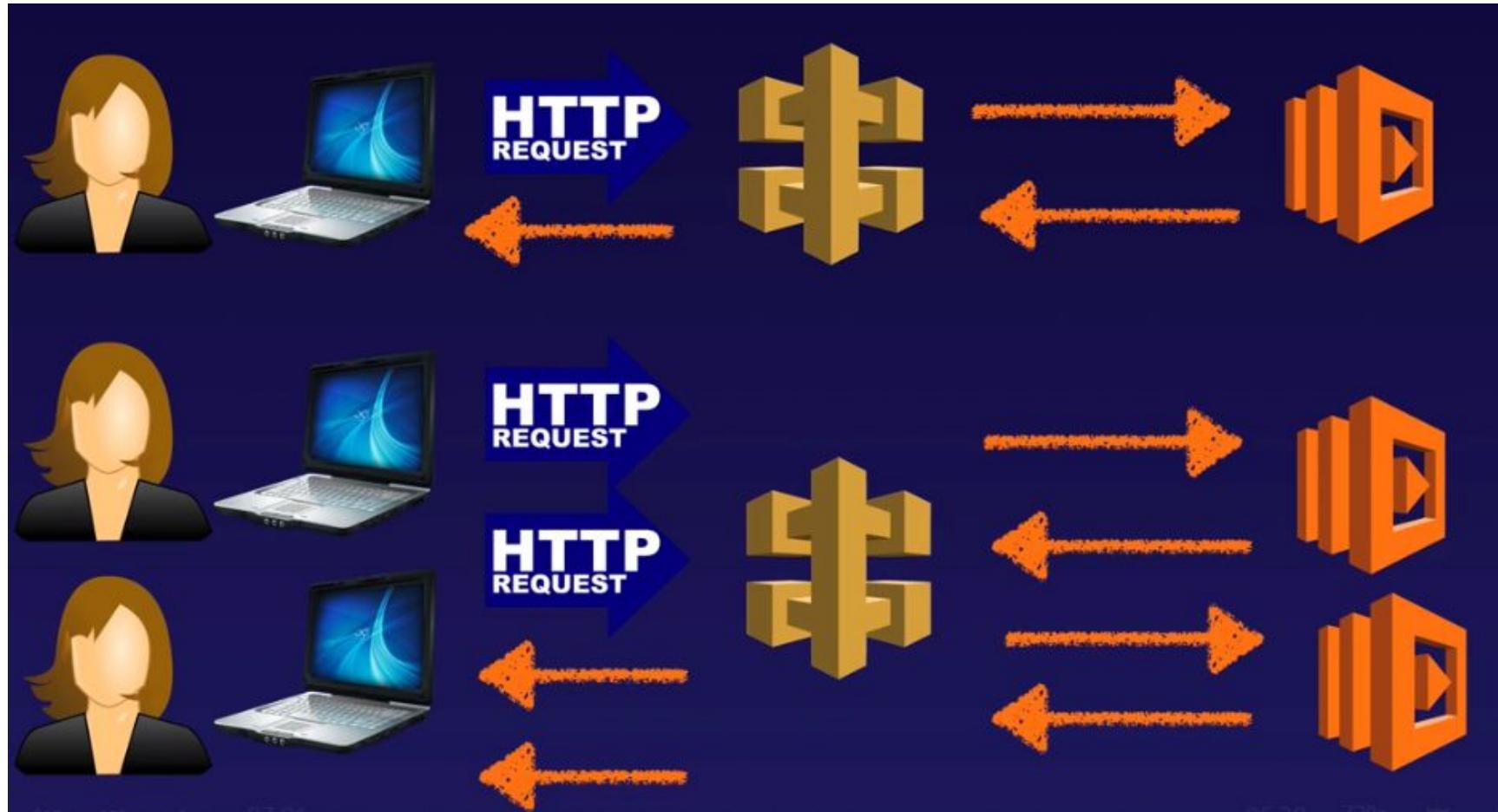
**You can use Lambda in the following ways;**

- 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.

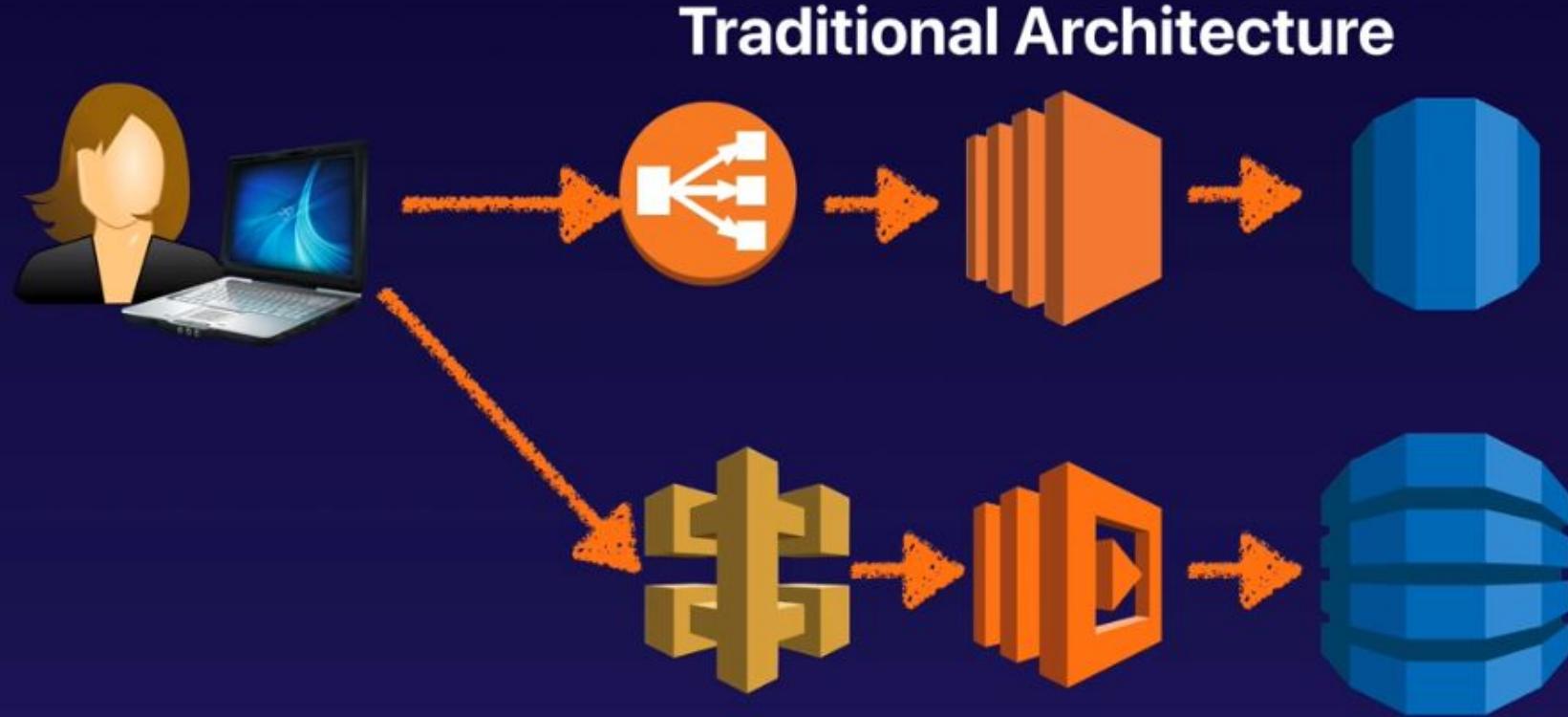
# What is Lambda?



# What is Lambda?



# What is Lambda?



**Serverless Architecture**

# 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

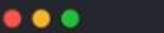
# **Serverless Application Model (SAM)**

# 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
```

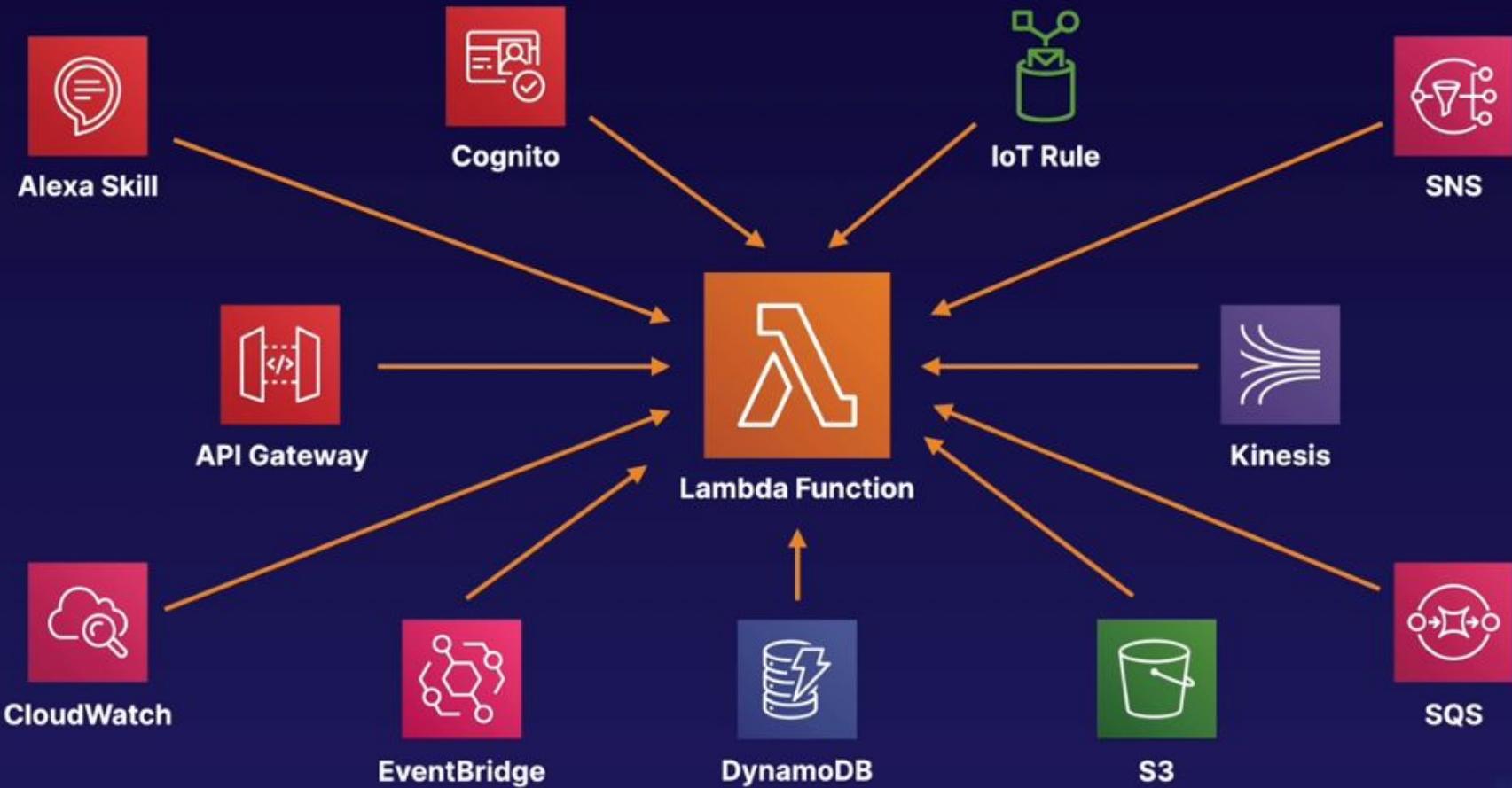
1 Tells CloudFormation this is a SAM template

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

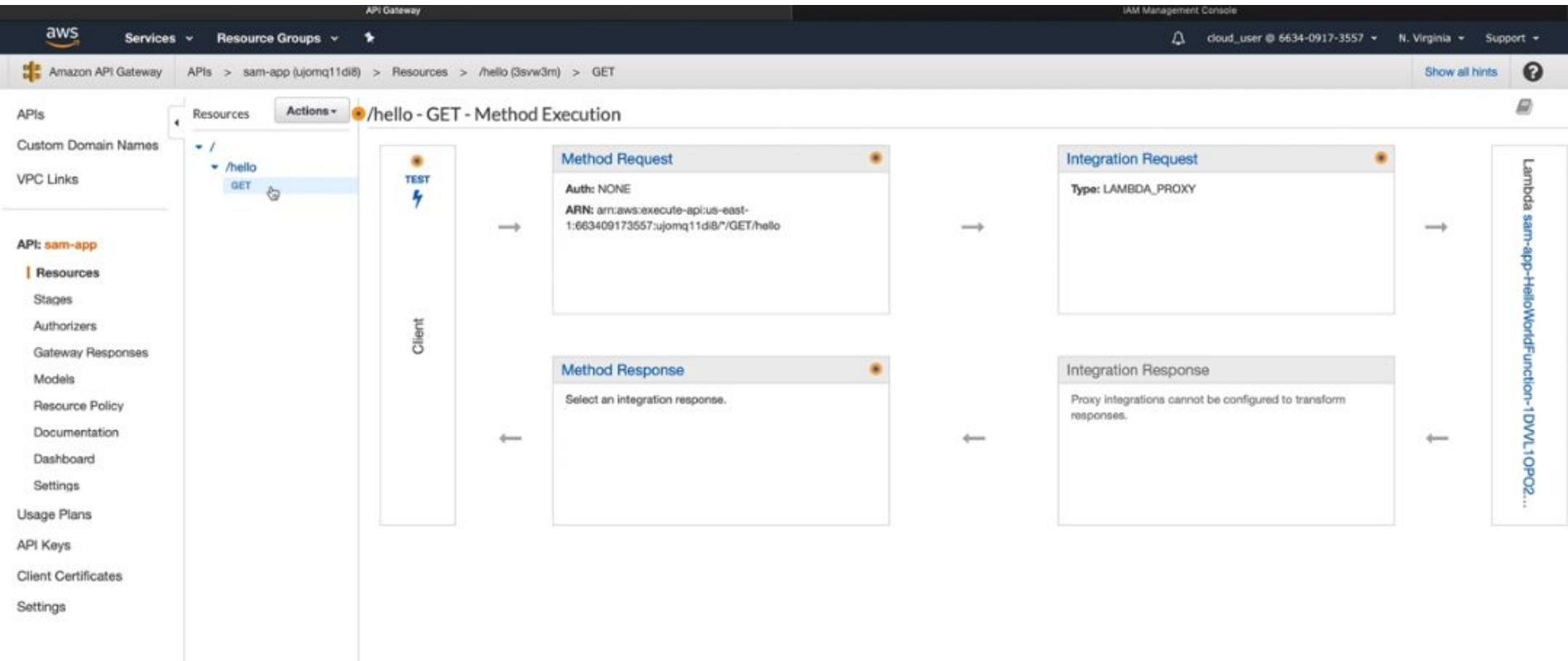
Key	HelloWorldApi
Description	API Gateway endpoint URL for Prod stage for Hello World function
Value	https://ujomqlid18.execute-api.us-east-1.amazonaws.com/Prod/hello/

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

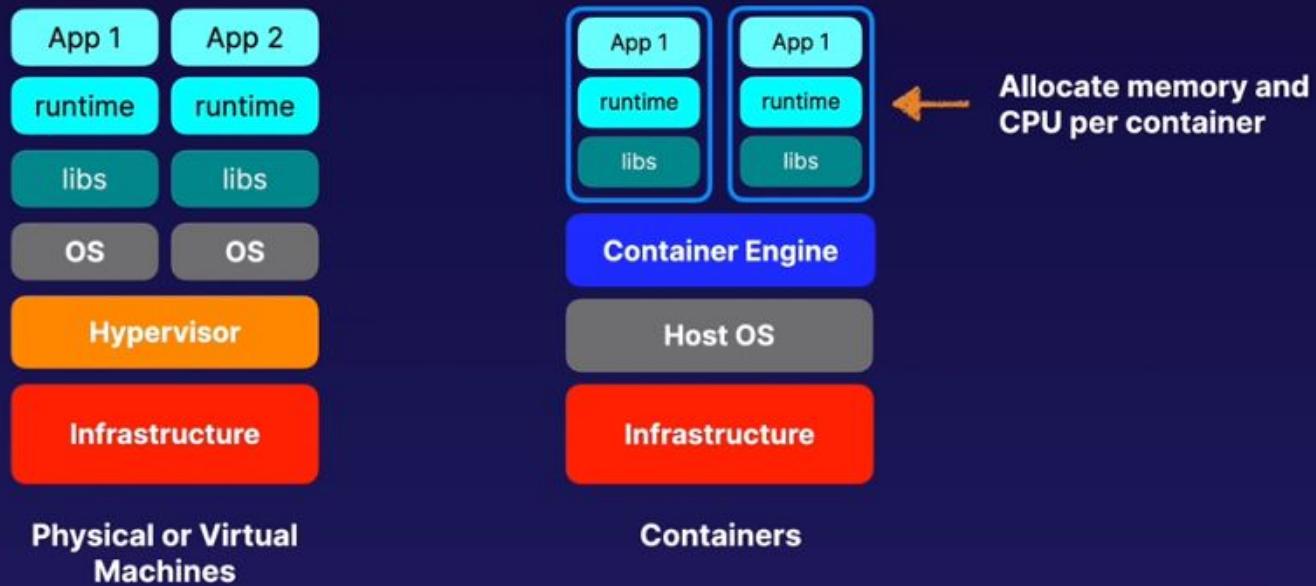
return {
    "statusCode": 200,
    "body": json.dumps({
        "message": "hello world",
        "location": ip.text.replace("\n", "")
    })
}
```

[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



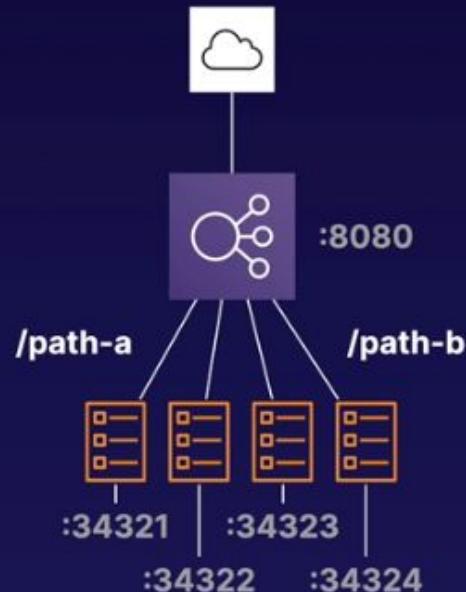
# ECR

- 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

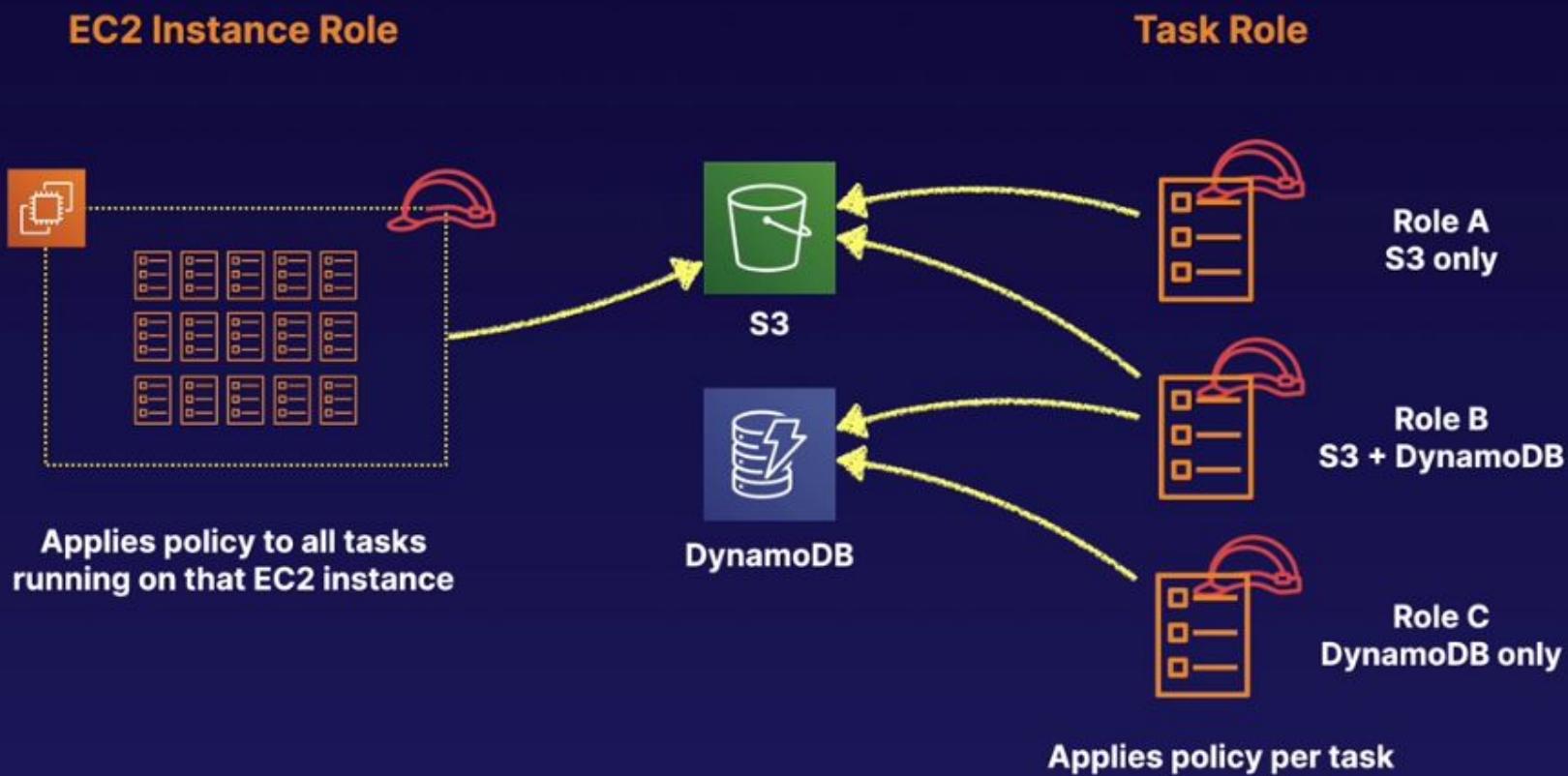


# ECS + ELB

- 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

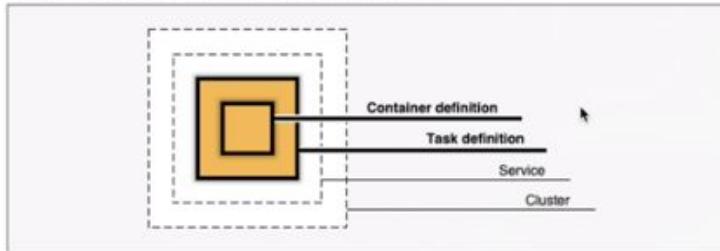
### Step 1: Container and Task

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

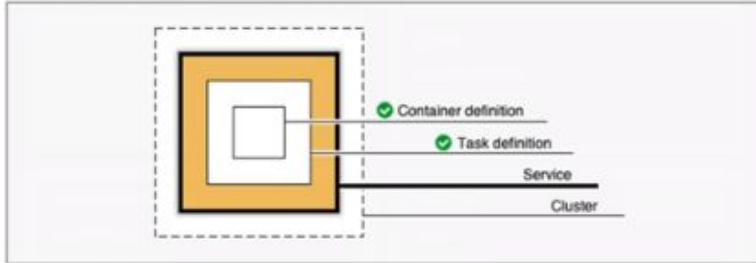
Step 1: Container and Task

**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-dfd... <a href="#">View details</a>	first-run-task-definition:1	RUNNING	RUNNING	service:sample-app-service	FARGATE	1.3.0
f972b175-b33c-48d2-9bcf-b43... <a href="#">View details</a>	first-run-task-definition:1	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

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

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

## 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).