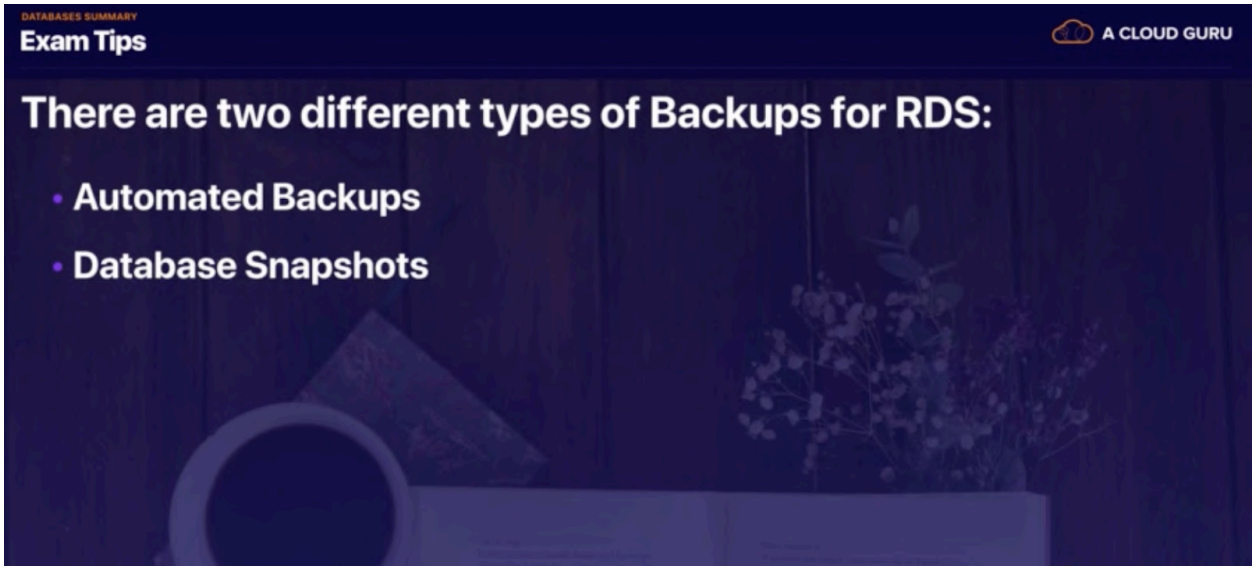
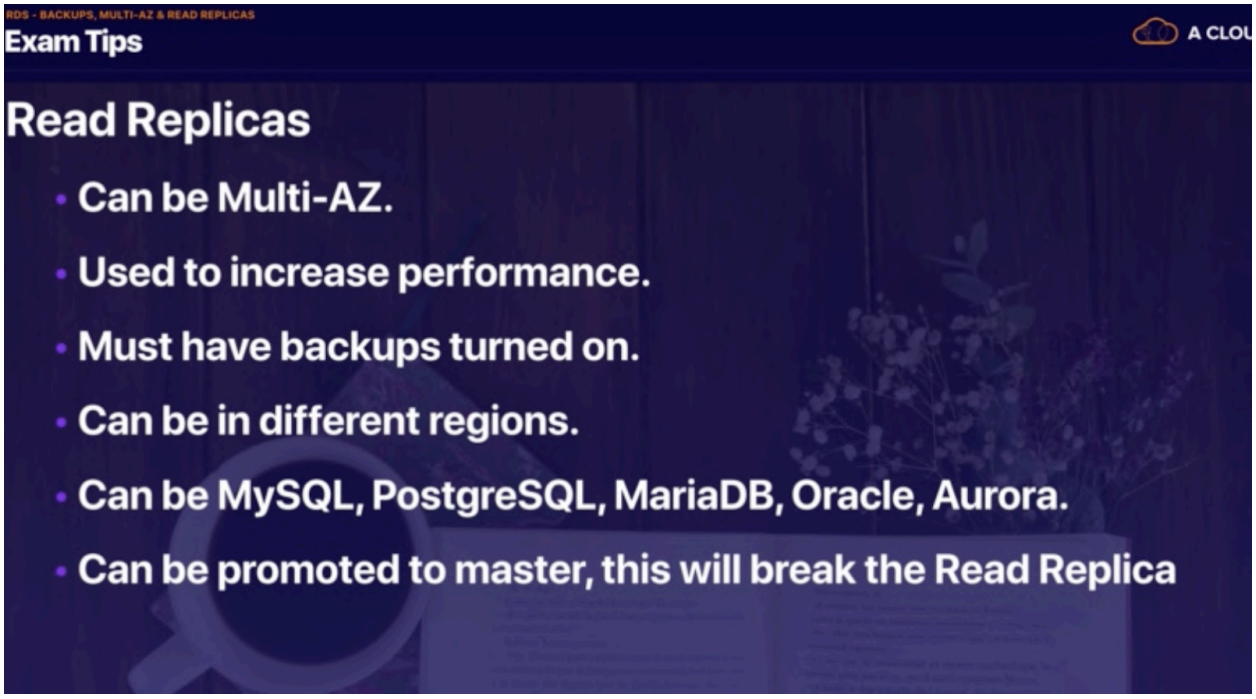
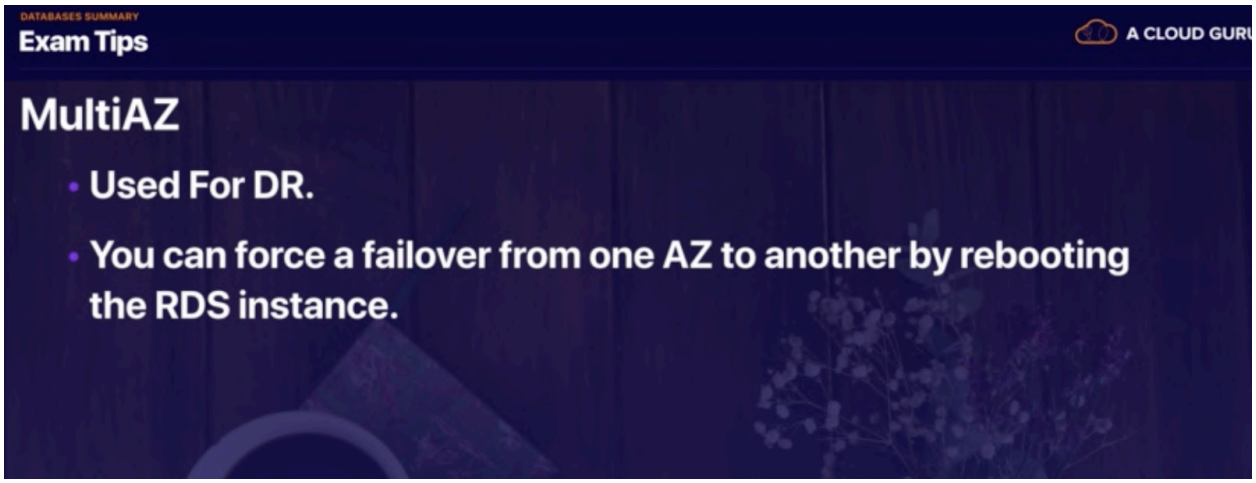
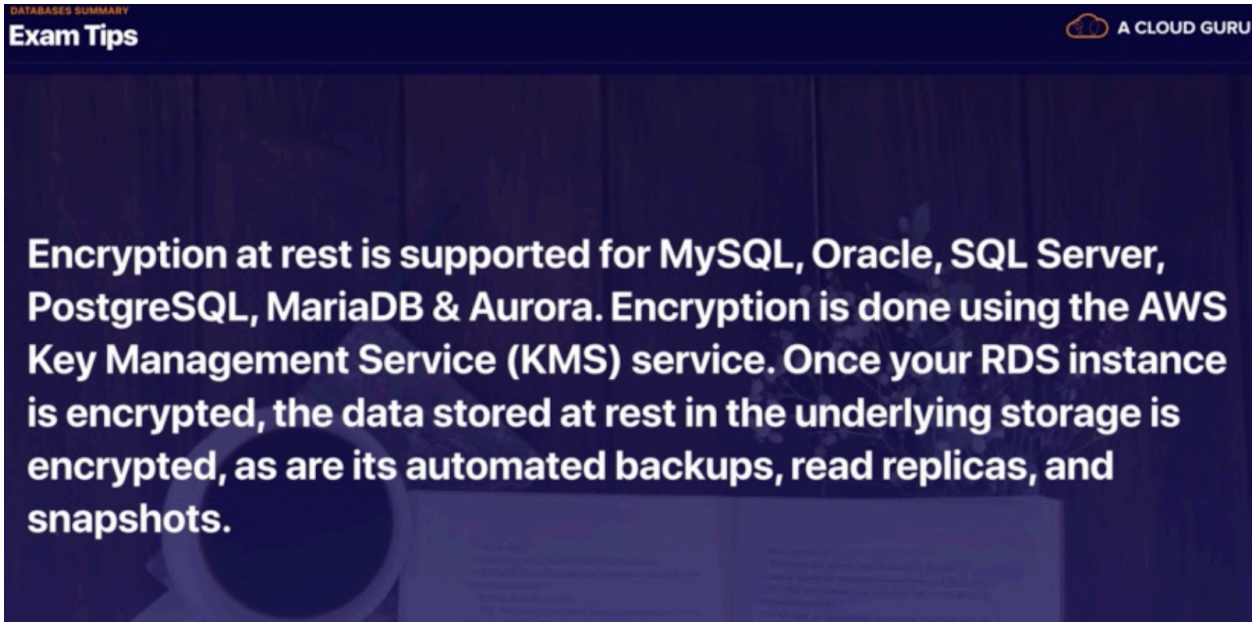
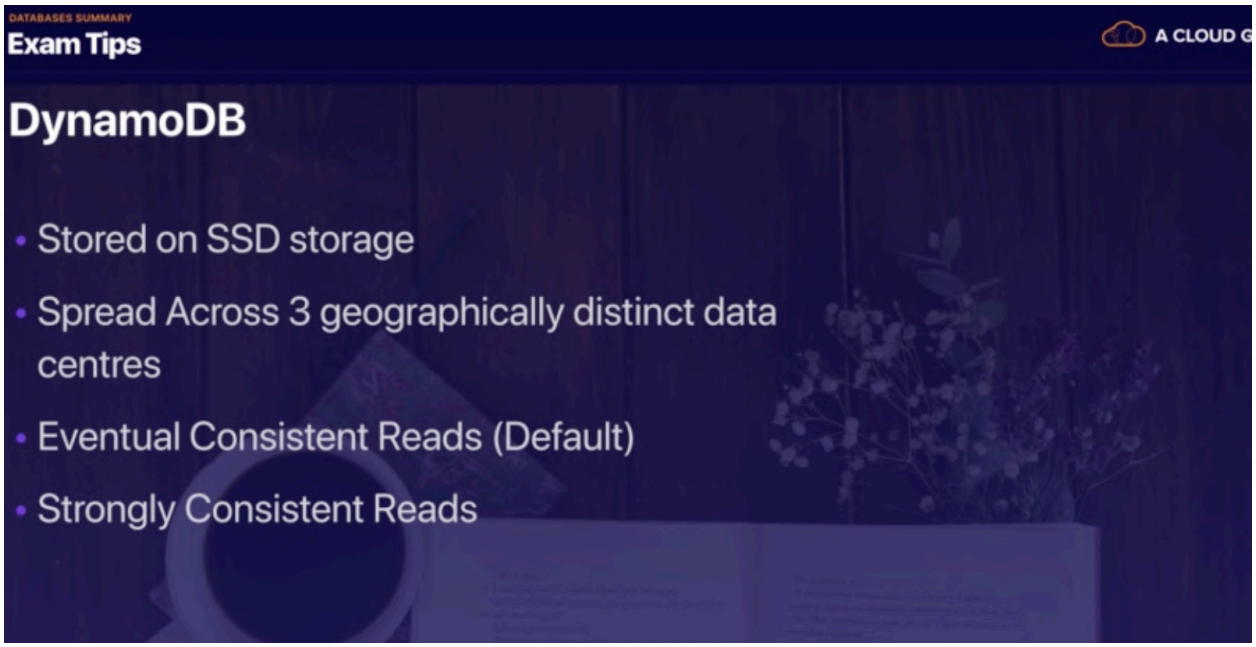

DATABASES SUMMARY
Exam Tips
A CLOUD GURU
There are two different types of Backups for RDS:
• Automated Backups
• Database Snapshots


RDS - BACKUPS, MULTI-AZ & READ REPLICAS
Exam Tips
A CLOU
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 SUMMARY
Exam Tips
A CLOUD GURU
MultiAZ
• Used For DR.
• You can force a failover from one AZ to another by rebooting the RDS instance.


DATABASES SUMMARY
Exam Tips
A CLOUD GURU
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 SUMMARY
Exam Tips
A CLOUD G
DynamoDB
• Stored on SSD storage
• Spread Across 3 geographically distinct data centres
• Eventual Consistent Reads (Default)
• Strongly Consistent Reads

DATABASES SUMMARY

Exam Tips

- Redshift is used for business intelligence.
- Available in only 1 AZ

DATABASES SUMMARY

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 SUMMARY

Exam Tips

A CL

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 SUMMARY

Exam Tips

A CLO

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 SUMMARY

Exam Tips

A CLO

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 SUMMARY

Exam Tips

A CLO

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 SUMMARY

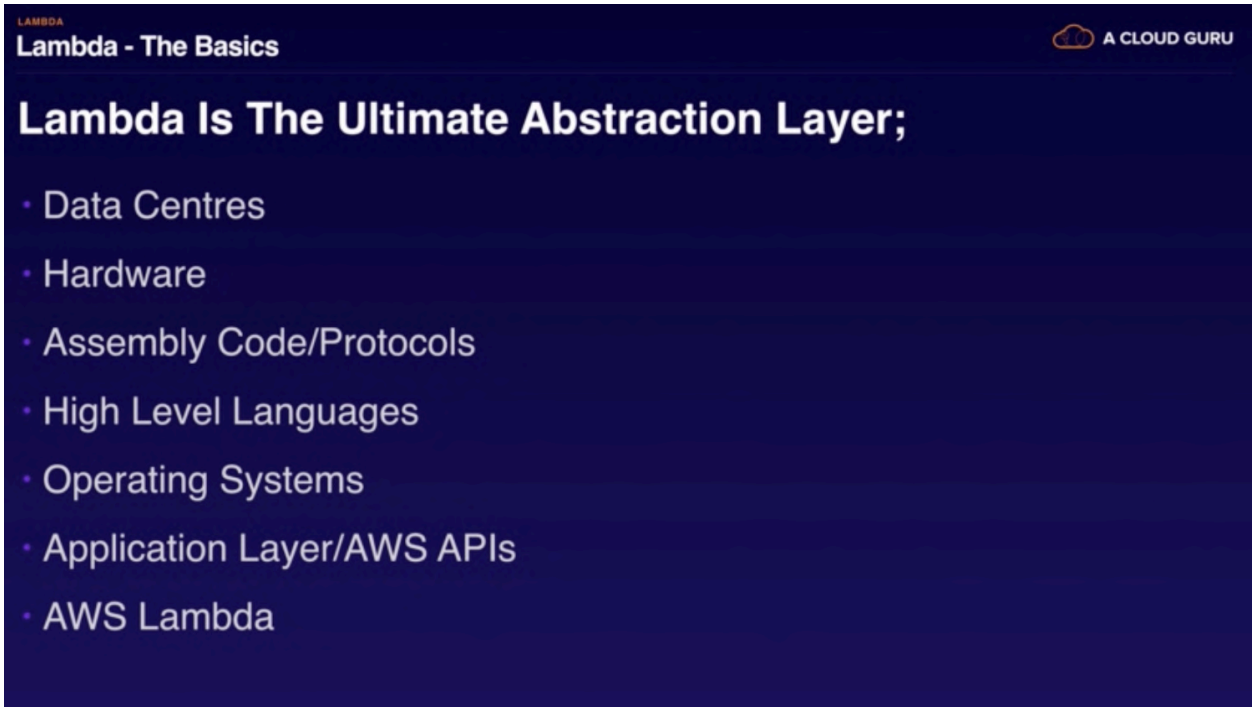
Exam Tips

A CLO

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

Serverless
Lambda

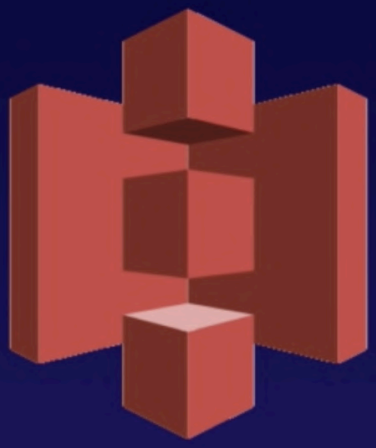


LAMBDA

What Is Lambda?

A CLOUD GURU

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

Lambda - The Basics

A CLOUD GURU

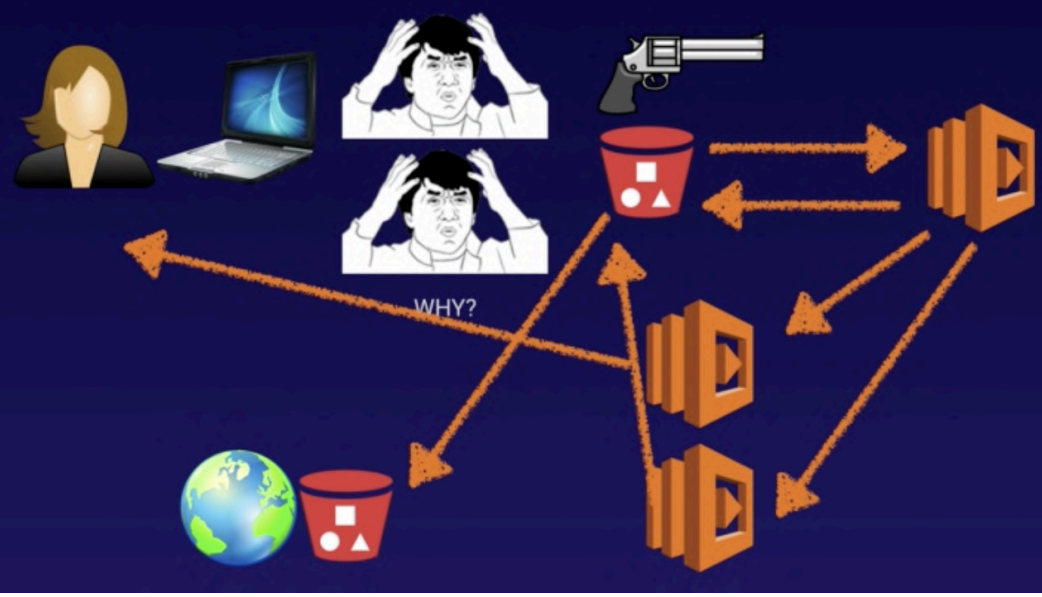
You can use Lambda in the following ways;

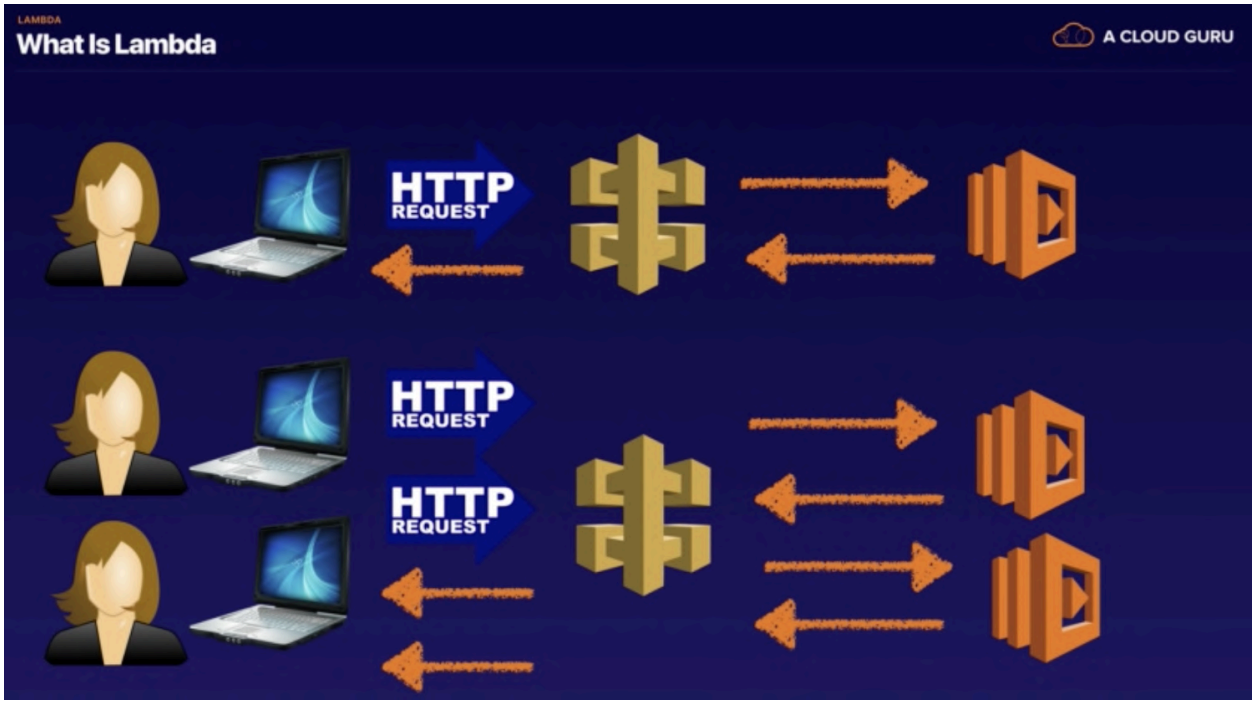
- As an event-driven compute service where AWS Lambda runs your code in response to events. These events could be changes to data in an Amazon S3 bucket or an Amazon DynamoDB table.
- As a compute service to run your code in response to HTTP requests using Amazon API Gateway or API calls made using AWS SDKs. This is what we use at A Cloud Guru.

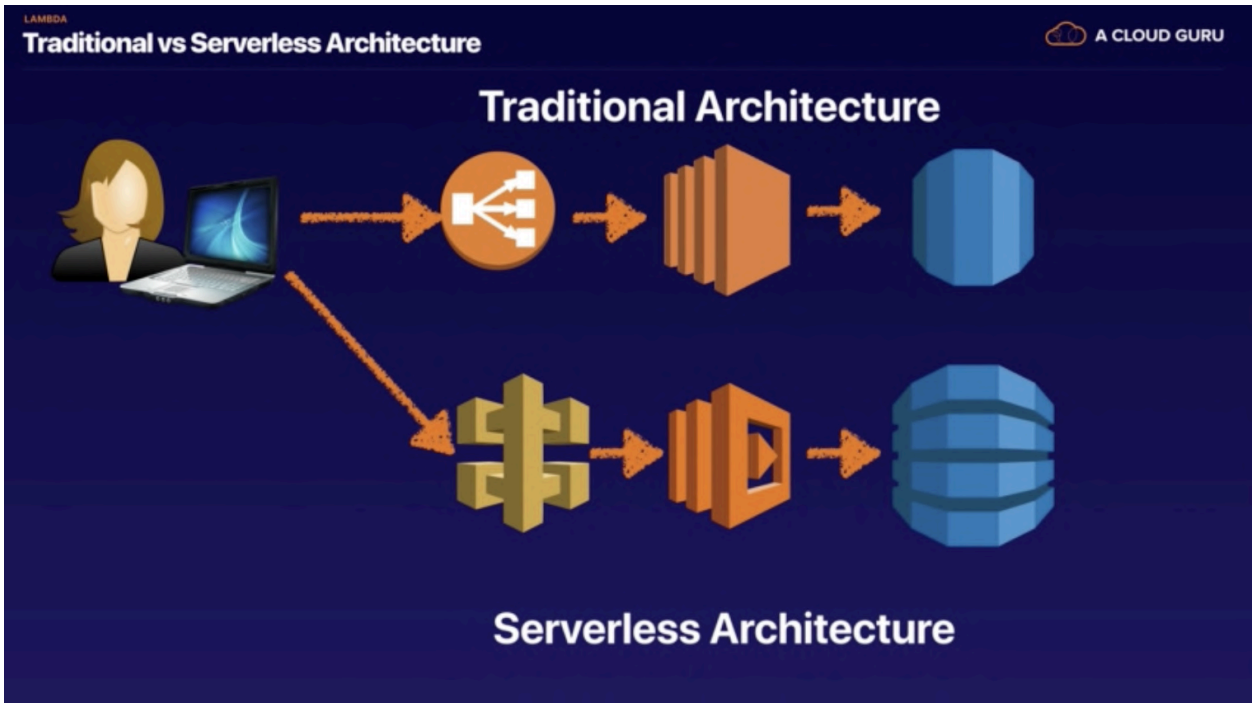
LAMBDA

What Is Lambda

A CLOUD GURU




LAMBDA
What Is Lambda
A CLOUD GURU
HTTP
REQUEST
HTTP
REQUEST
HTTP
REQUEST


LAMBDA
Traditional vs Serverless Architecture
A CLOUD GURU
Traditional Architecture
Serverless Architecture

What Languages Does Lambda Support

- Node.js
- Java
- Python
- C#
- Go
- PowerShell

LAMBDA

Lambda Pricing

A CLOUD GURU

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.

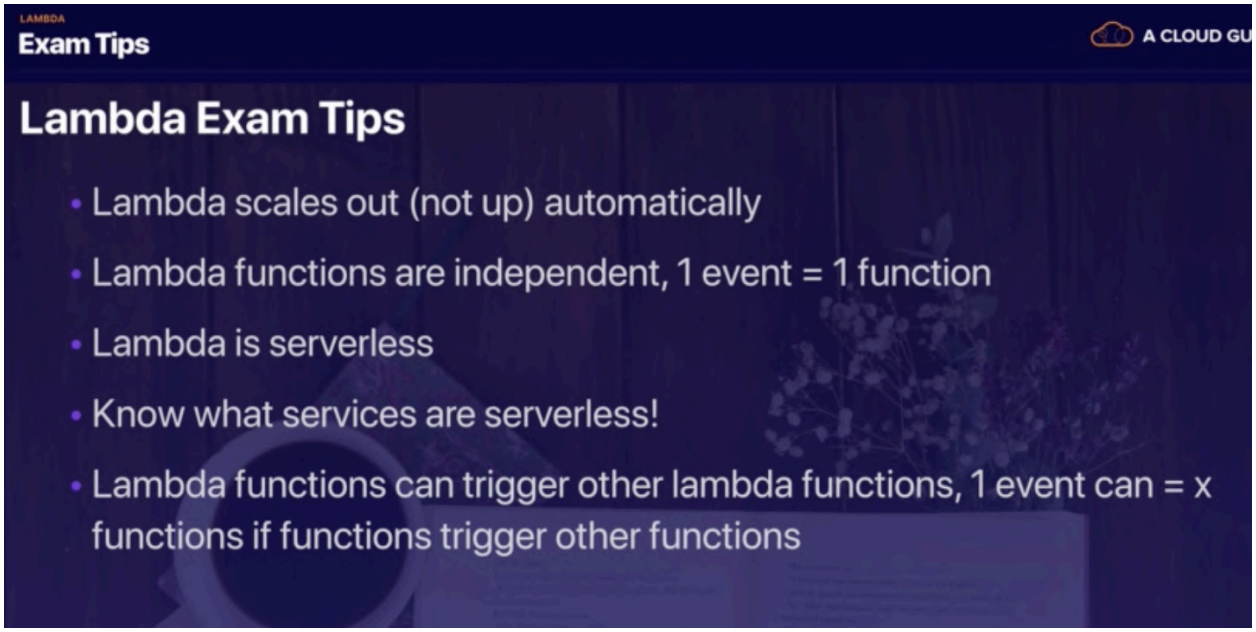
LAMBDA

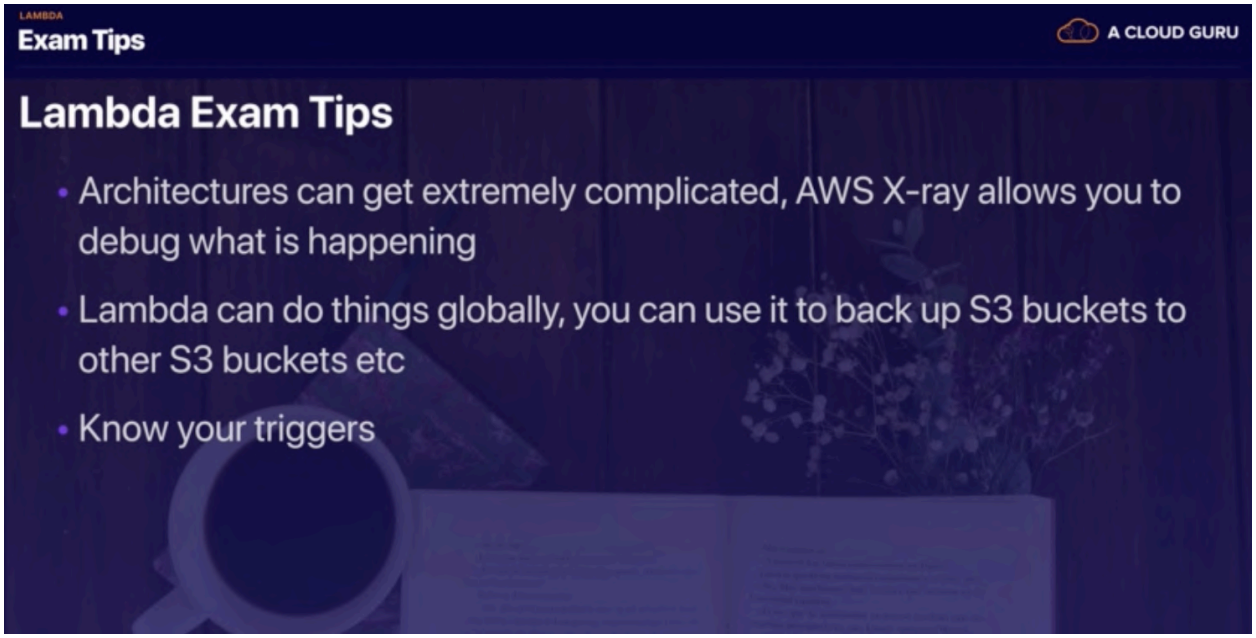
Lambda The Basics

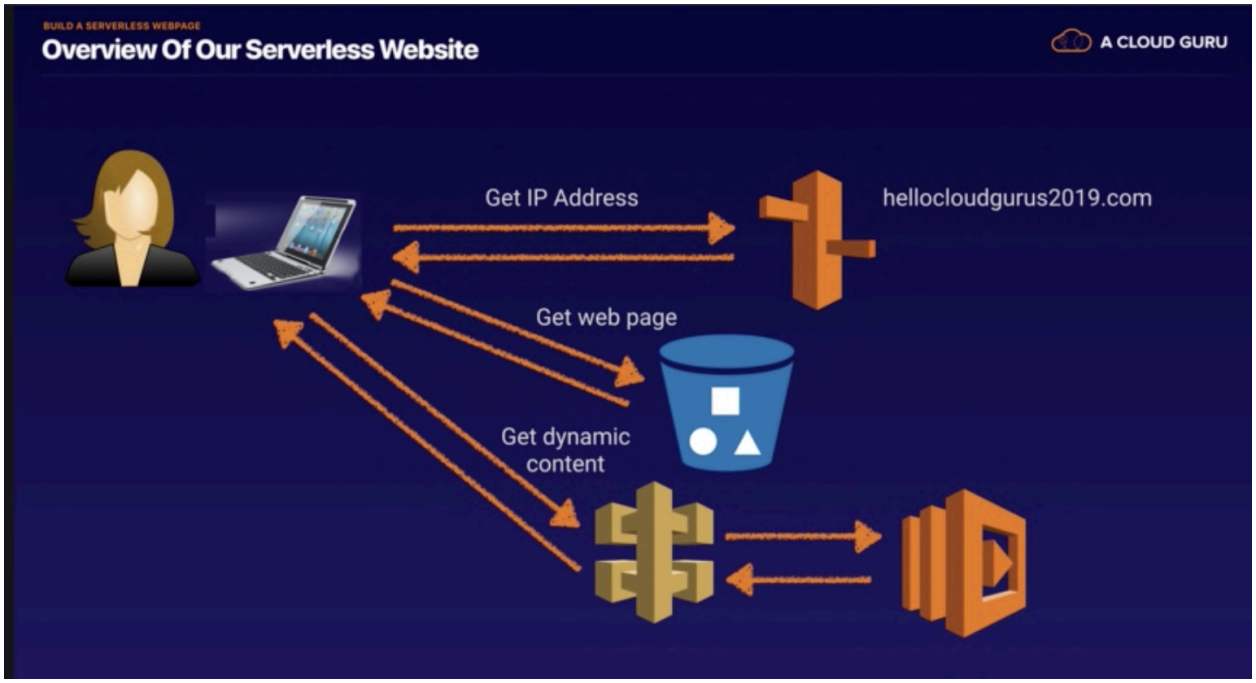
A CLOU

Why Is Lambda Cool?

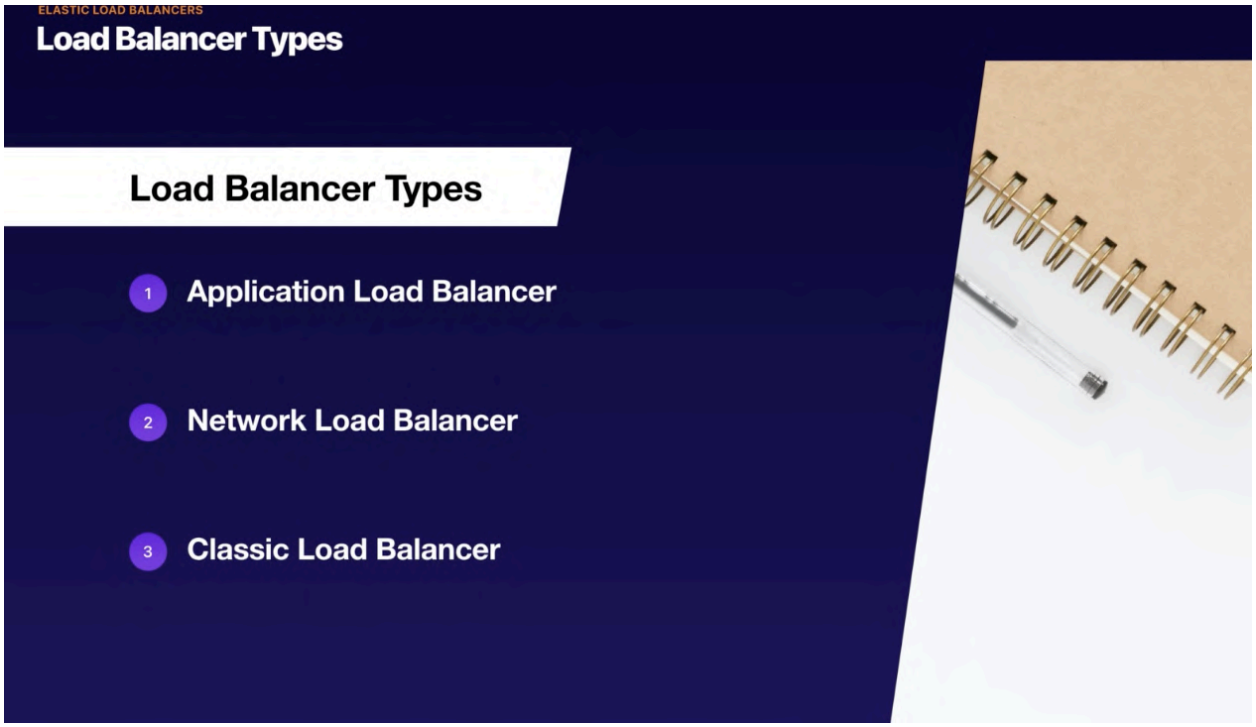
- NO SERVERS!
- Continuous Scaling
- Super super super cheap!


LAMBDA
Exam Tips
A CLOUD GU
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


LAMBDA
Exam Tips
A CLOUD GURU
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



HA Architecture



ELASTIC LOAD BALANCERS

Application Load Balancers

A CL

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

ELASTIC LOAD BALANCERS

Network Load Balancers

A CLOUD GURU

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

Use for extreme performance!

ELASTIC LOAD BALANCERS

Classic Load Balancers

A CLOUD GURU

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

ELASTIC LOAD BALANCERS

Errors in Load Balancing

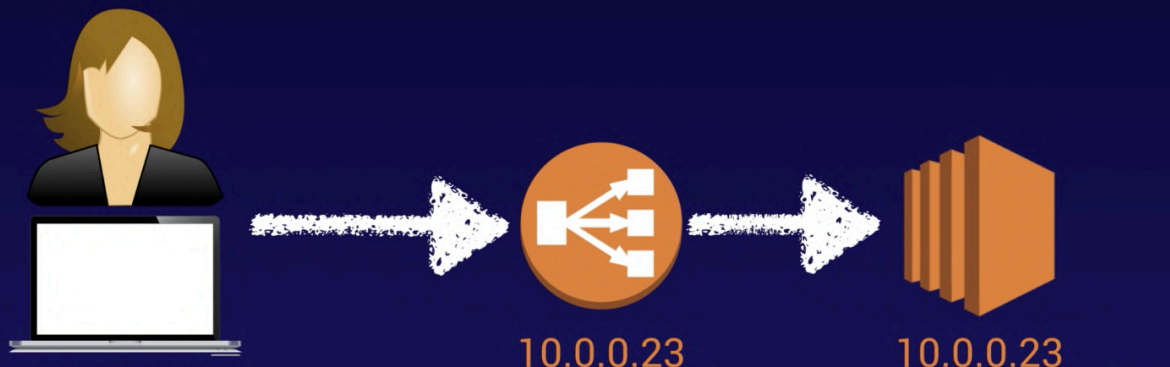
A CLOUD GU

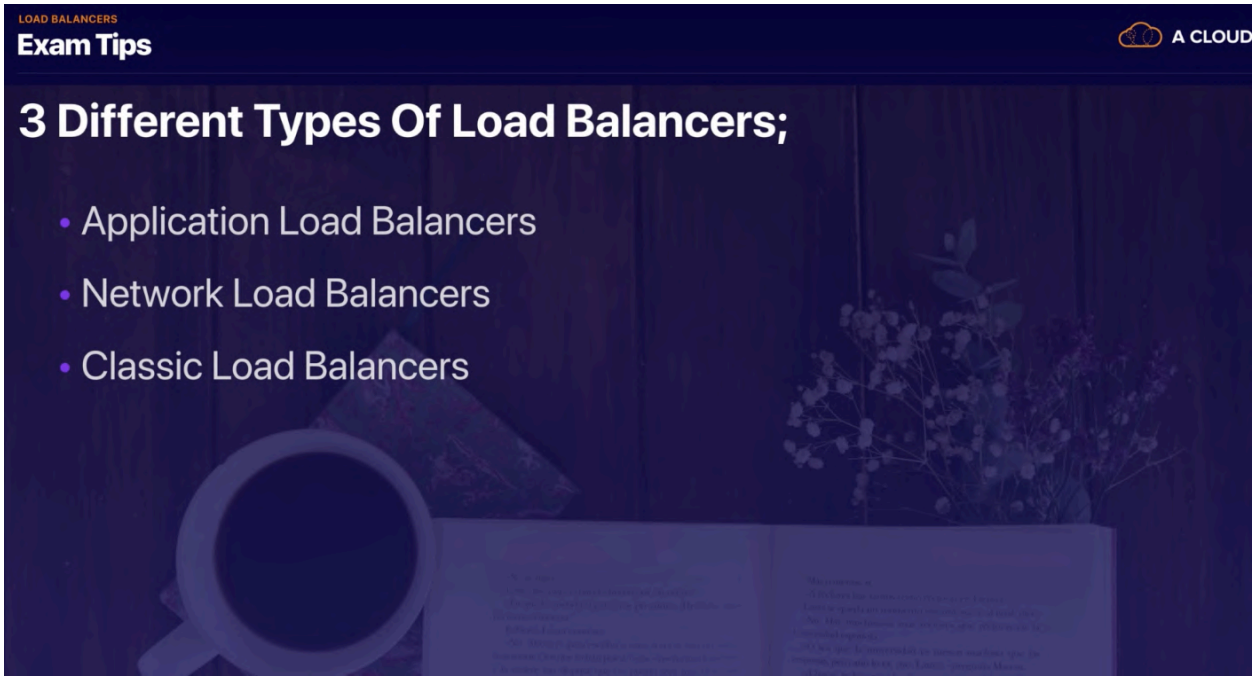
Classic Load Balancers -if your application stops responding, the ELB (Classic Load Balancer) responds with a 504 error.
This means that the application is having issues. This could be either at the Web Server layer or at the Database Layer.
Identify where the application is failing, and scale it up or out where possible.

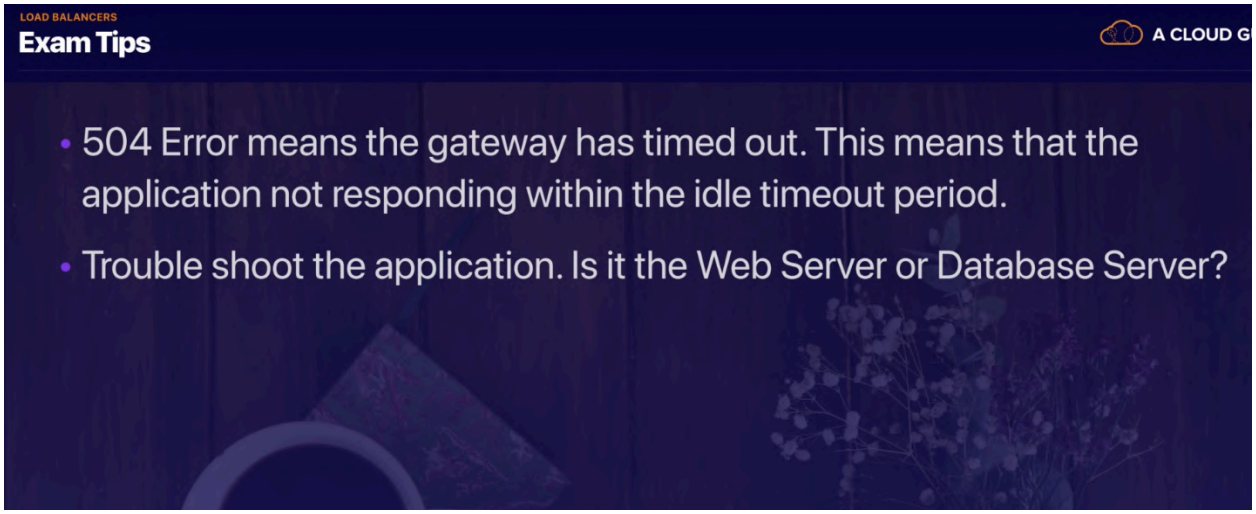
ELASTIC LOAD BALANCERS

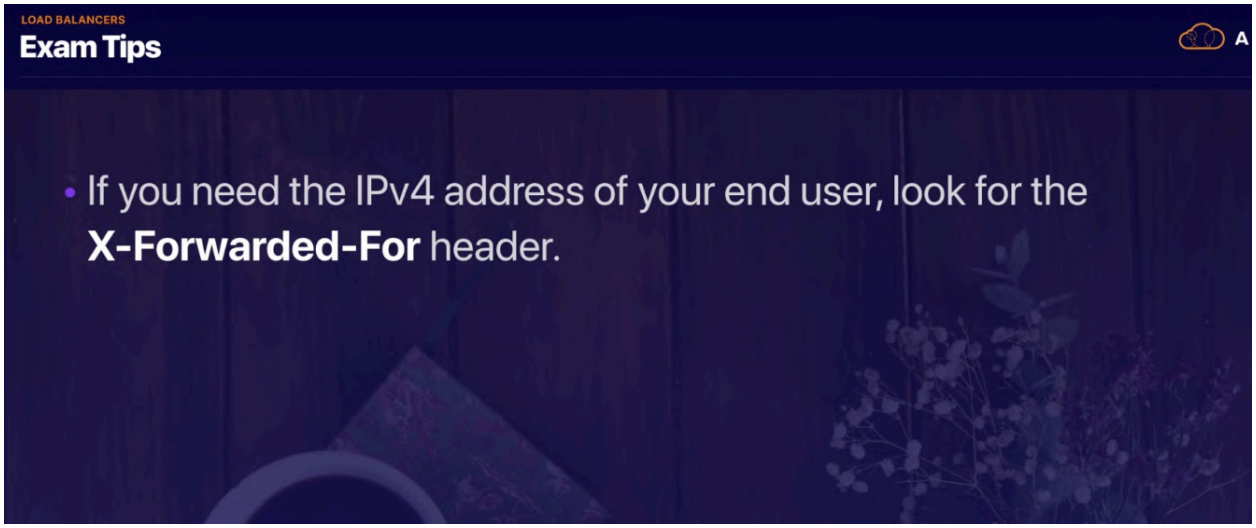
X-Forwarded-For Header

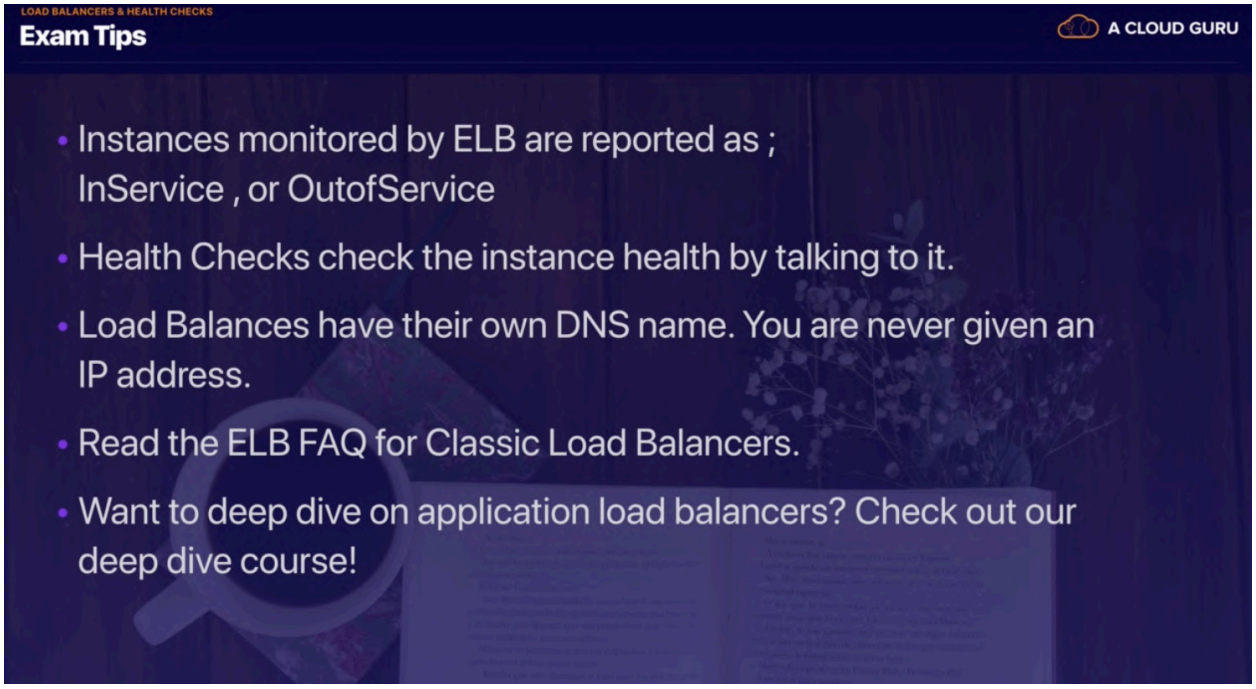
A CLOUD GURU




LOAD BALANCERS
Exam Tips
A CLOUD
3 Different Types Of Load Balancers;
• Application Load Balancers
• Network Load Balancers
• Classic Load Balancers


LOAD BALANCERS
Exam Tips
A CLOUD G
• 504 Error means the gateway has timed out. This means that the application not responding within the idle timeout period.
• Trouble shoot the application. Is it the Web Server or Database Server?



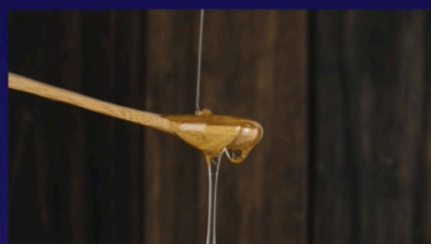


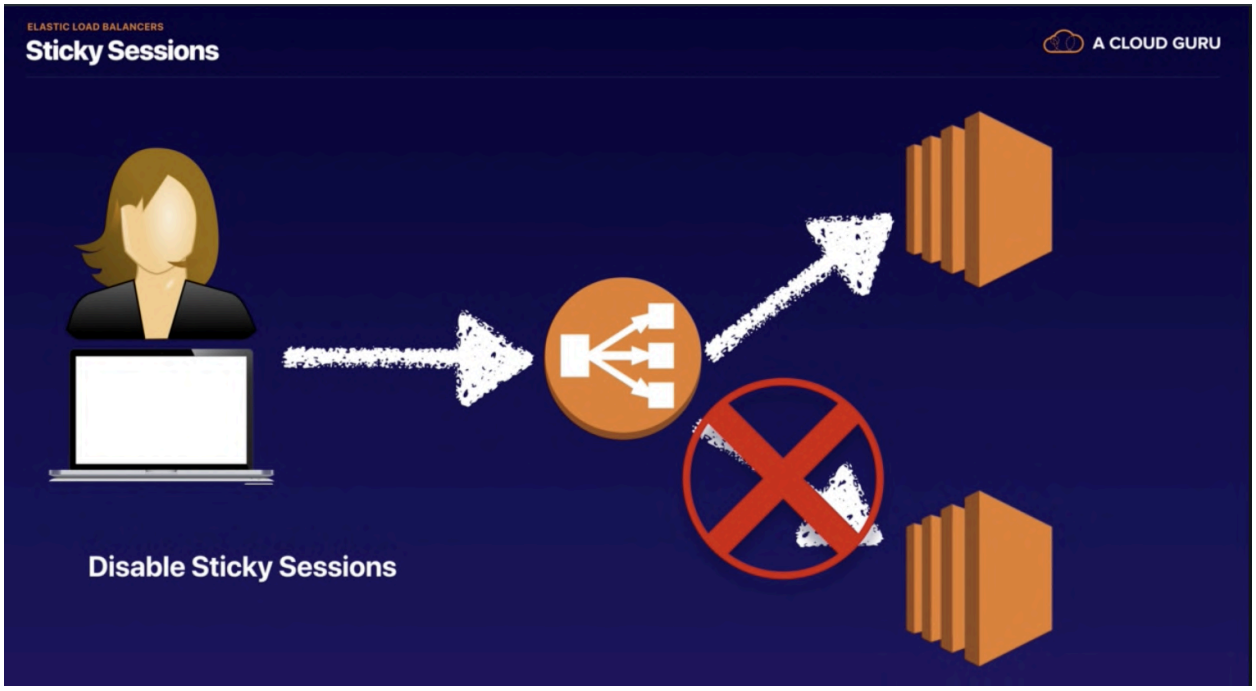
Advance LoadBalancer theory

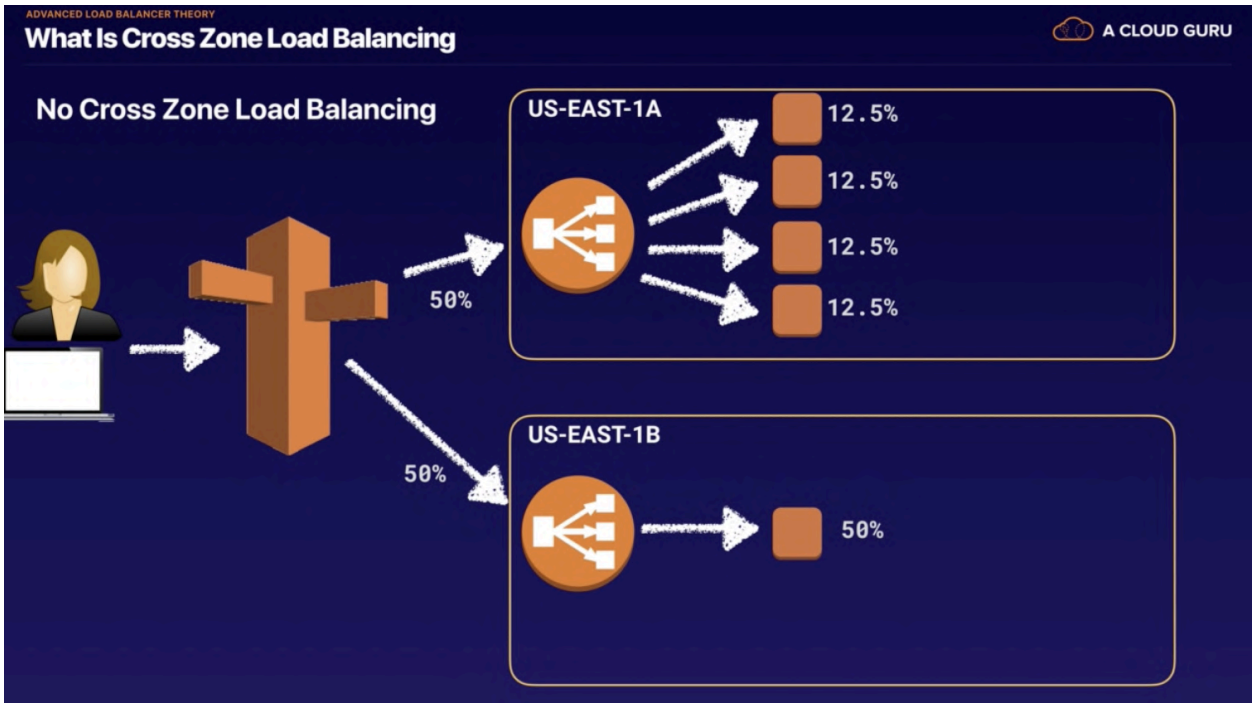
What Are Sticky Sessions?

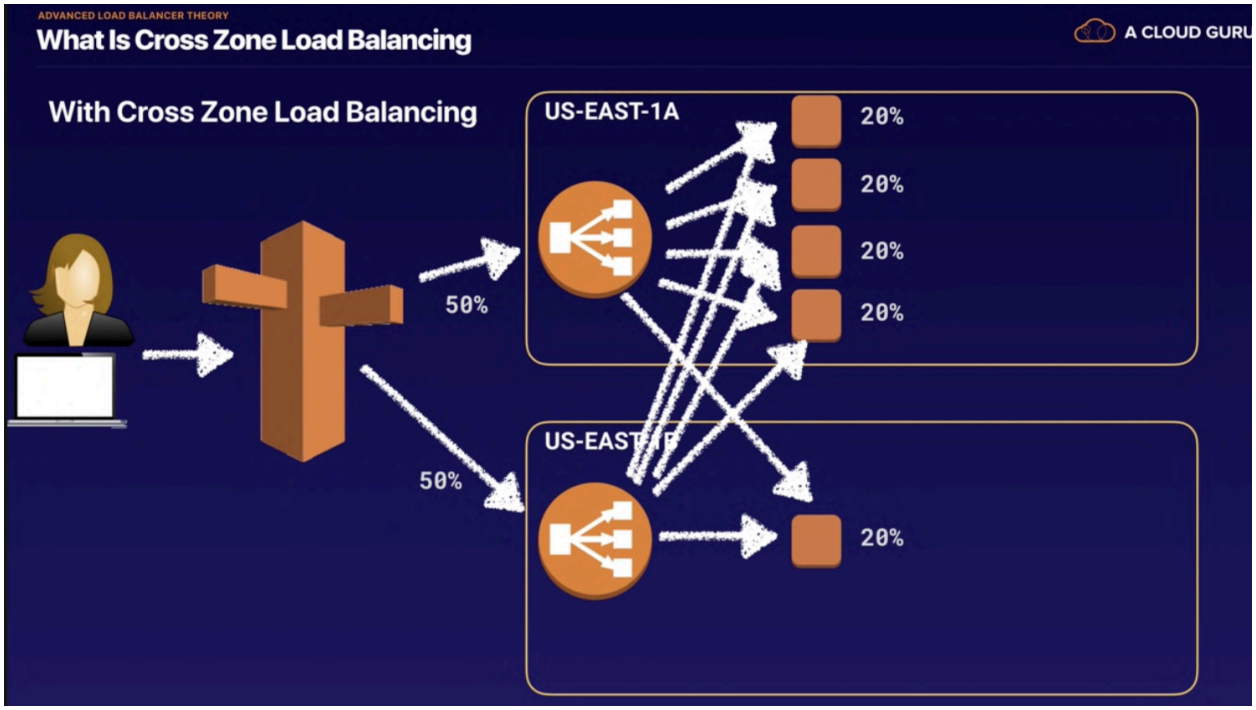
Classic Load Balancer routes each request independently to the registered EC2 instance with the smallest load.
Sticky sessions allow you to bind a user's session to a specific EC2 instance. This ensures that all requests from the user during the session are sent to the same instance.

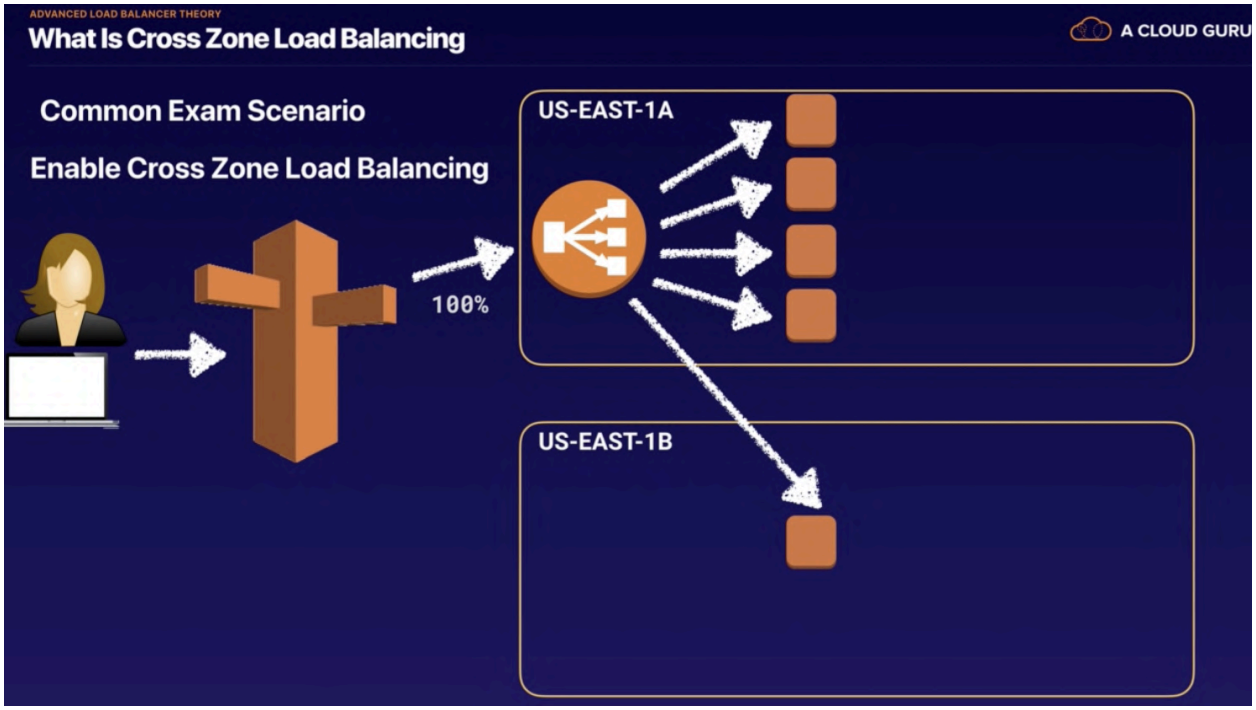
You can enable Sticky Sessions for Application Load Balancers as well, but the traffic will be sent at the Target Group Level.




ELASTIC LOAD BALANCERS
Sticky Sessions
A CLOUD GURU
Disable Sticky Sessions


ADVANCED LOAD BALANCER THEORY
What Is Cross Zone Load Balancing
A CLOUD GURU
No Cross Zone Load Balancing
US-EAST-1A
12.5%
12.5%
12.5%
12.5%
50%
50%
US-EAST-1B
50%

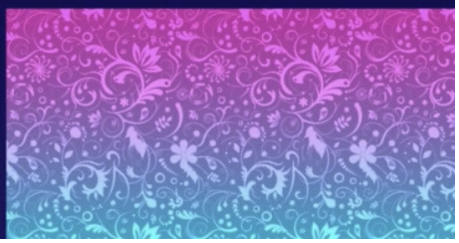

ADVANCED LOAD BALANCER THEORY
What Is Cross Zone Load Balancing
A CLOUD GURU
With Cross Zone Load Balancing
US-EAST-1A
20%
20%
20%
20%
50%
50%
20%

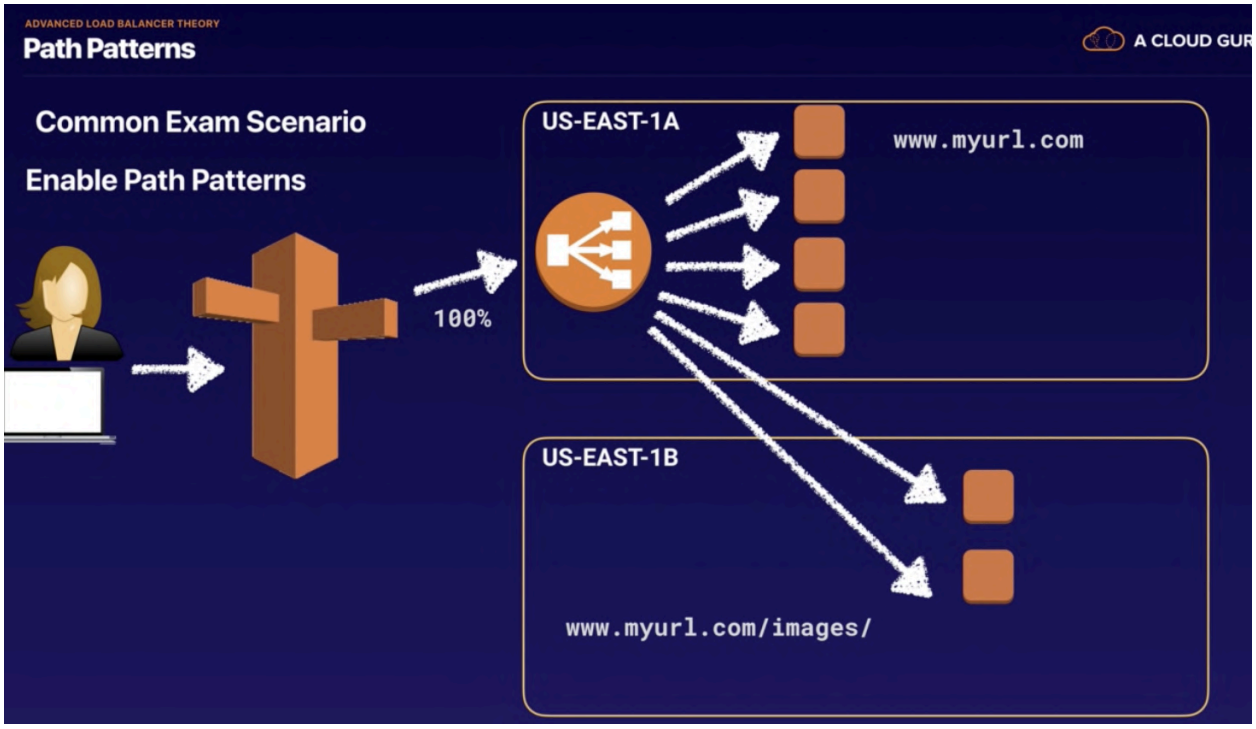

ADVANCED LOAD BALANCER THEORY
What Is Cross Zone Load Balancing
A CLOUD GURU
Common Exam Scenario
Enable Cross Zone Load Balancing
US-EAST-1A
100%
US-EAST-1B

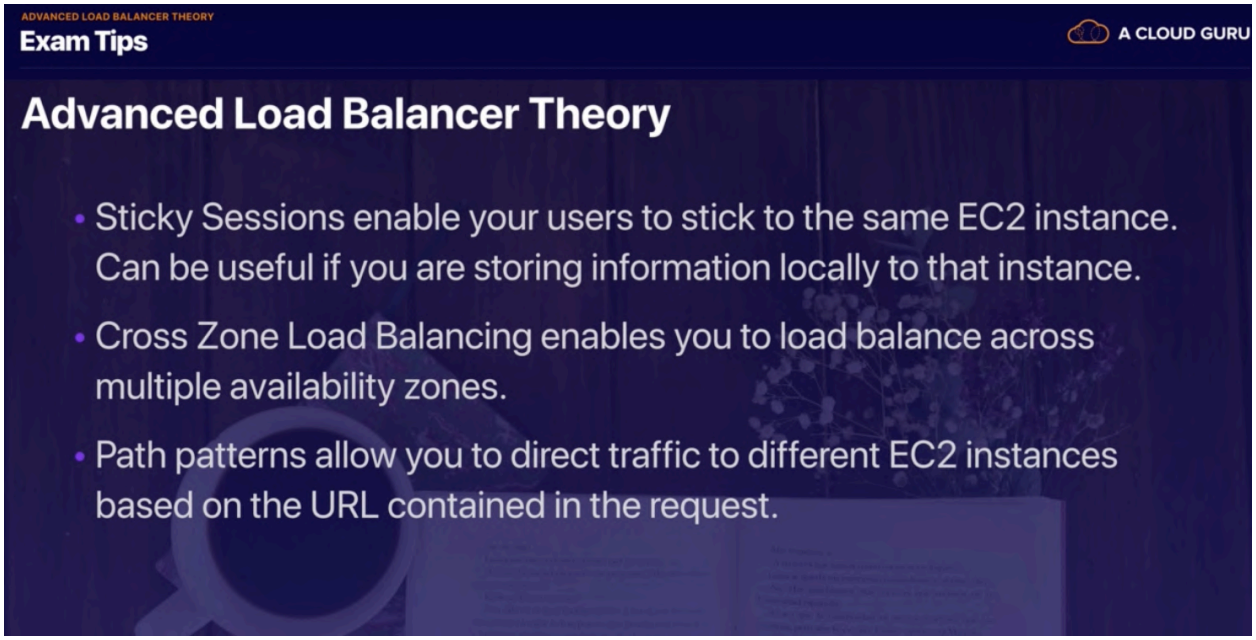
A CLOUD GURU

What Are Path Patterns?

You can create a listener with rules to forward requests based on the URL path. This is known as path-based routing. If you are running microservices, you can route traffic to multiple back-end services using path-based routing. For example, you can route general requests to one target group and requests to render images to another target group.



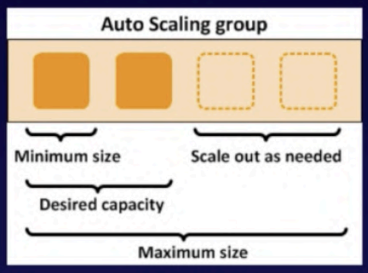




Autoscaling

A CLOUD GURU

Auto Scaling Has 3 Components



1. **Groups**
   Logical component. Webserver group or Application group or Database group etc.

2. **Configuration Templates**
   Groups uses a launch template or a launch configuration as a configuration template for its EC2 instances. You can specify information such as the AMI ID, instance type, key pair, security groups, and block device mapping for your instances.

3. **Scaling Options**
   Scaling Options provides several ways for you to scale your Auto Scaling groups. For example, you can configure a group to scale based on the occurrence of specified conditions (dynamic scaling) or on a schedule.

A CLOUD GURU

What are my scaling options?

- Maintain current instance levels at all times
- Scale manually
- Scale based on a schedule
- Scale based on demand
- Use predictive scaling

AUTOSCALING

Different Scaling Options

A CLOUD GURU

Maintain current instance levels at all times

You can configure your Auto Scaling group to maintain a specified number of running instances at all times.

To maintain the current instance levels, Amazon EC2 Auto Scaling performs a periodic health check on running instances within an Auto Scaling group.

When Amazon EC2 Auto Scaling finds an unhealthy instance, it terminates that instance and launches a new one.

AUTOSCALING

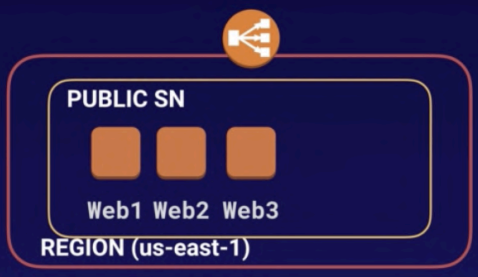
Different Scaling Options

A CLOUD GURU

Scale manually

Manual scaling is the most basic way to scale your resources, where you specify only the change in the maximum, minimum, or desired capacity of your Auto Scaling group.

Amazon EC2 Auto Scaling manages the process of creating or terminating instances to maintain the updated capacity.



Scale based on a schedule

Scaling by schedule means that scaling actions are performed automatically as a function of time and date.

This is useful when you know exactly when to increase or decrease the number of instances in your group, simply because the need arises on a predictable schedule.

AUTOSCALING

Different Scaling Options

A CLOUD GURU

Scale based on demand

A more advanced way to scale your resources - using scaling policies - lets you define parameters that control the scaling process.

For example, let's say that you have a web application that currently runs on two instances and you want the CPU utilization of the Auto Scaling group to stay at around 50 percent when the load on the application changes. This method is useful for scaling in response to changing conditions, when you don't know when those conditions will change. You can set up Amazon EC2 Auto Scaling to respond for you. We will do this in the next lab.

AUTOSCALING

Different Scaling Options

A CLOUD GUR

Use predictive scaling

You can also use Amazon EC2 Auto Scaling in combination with AWS Auto Scaling to scale resources across multiple services.

AWS Auto Scaling can help you maintain optimal availability and performance by combining predictive scaling and dynamic scaling (proactive and reactive approaches, respectively) to scale your Amazon EC2 capacity faster.

HA Architecture

HA ARCHITECTURE

Plan For Failure

A CLOUD GURU

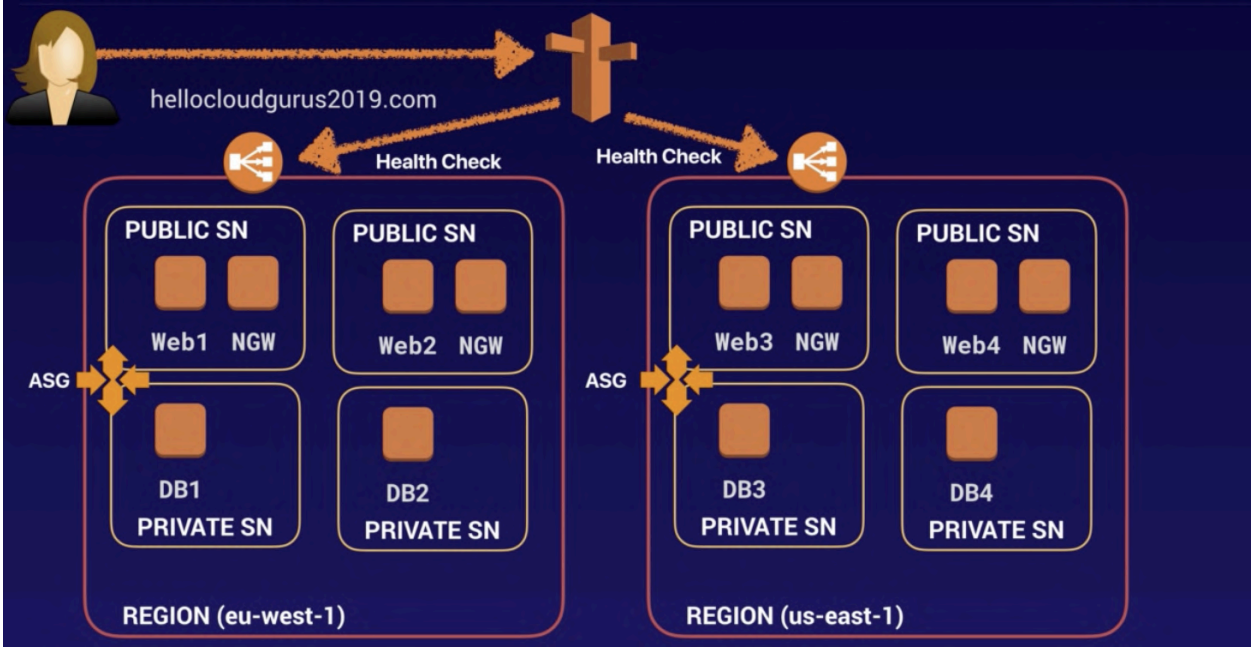
Everything fails. Everything.

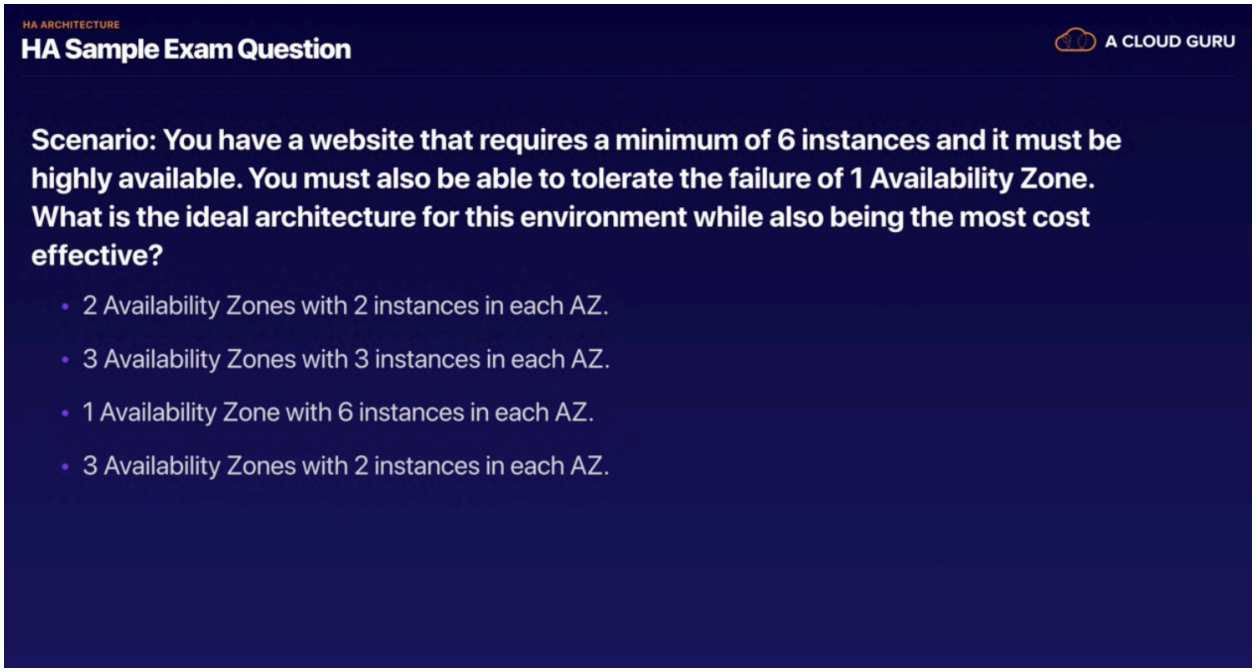
You should always plan for failure.

HA ARCHITECTURE

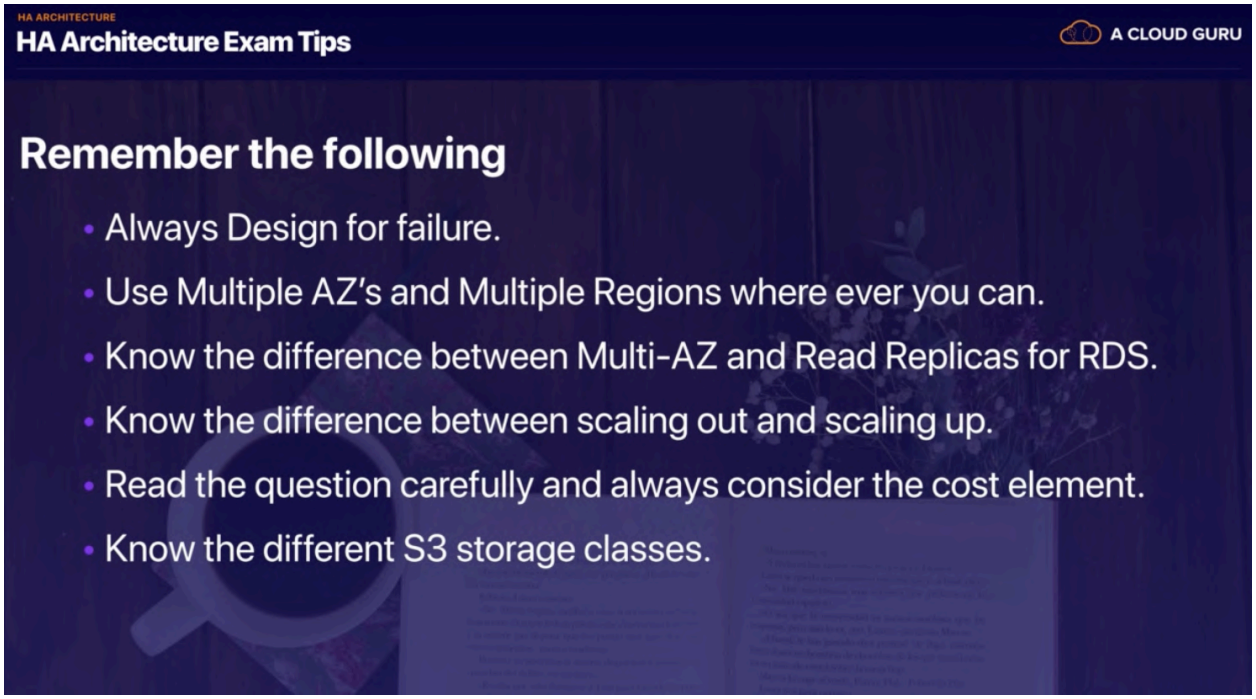
HA Architecture Example

A CLOUD GUR





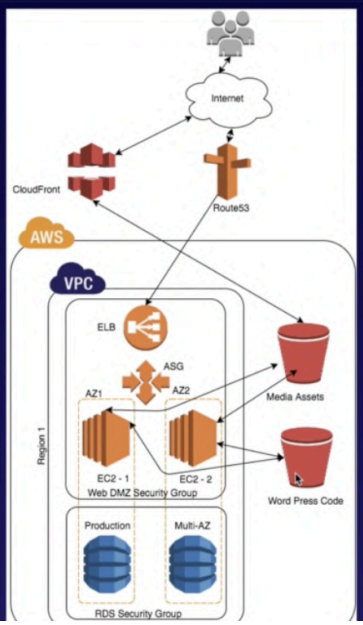
3 AZ with 2 instances in each AZ is wrong becoz if any one of AZ fails we will be left with only 4 instances.
Thus the correct ans is 3 AZ with 3 instances in each AZ.

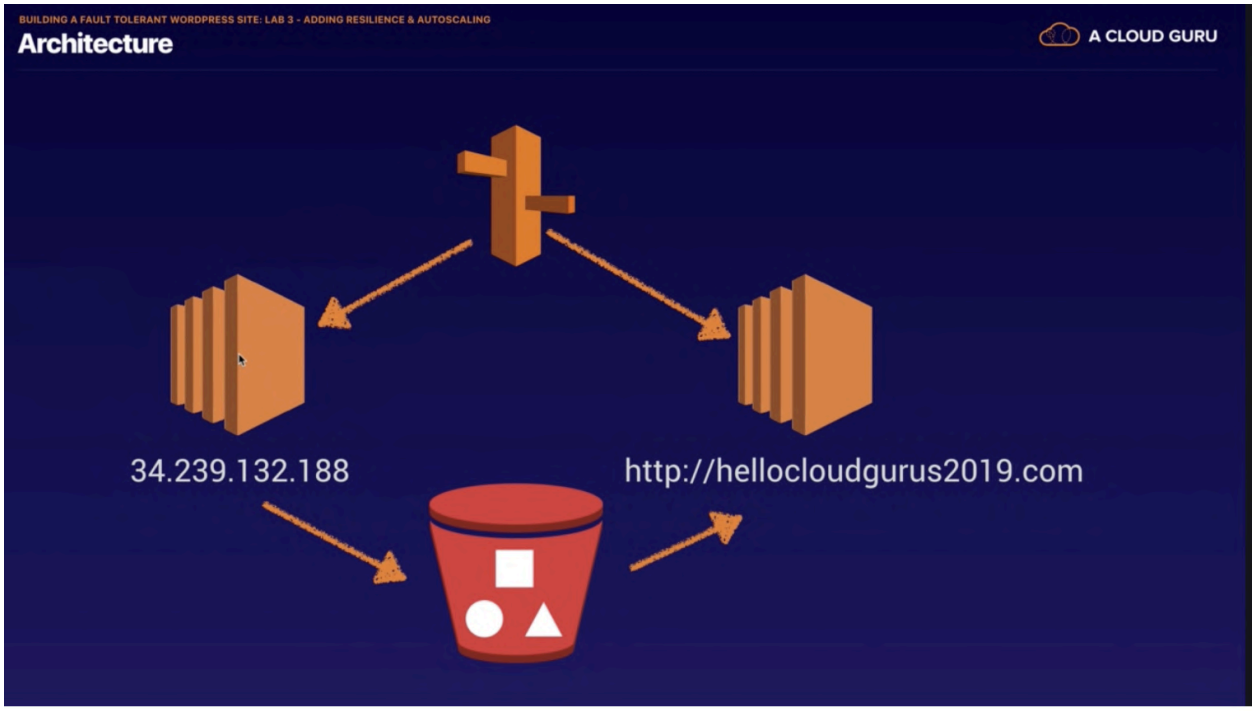


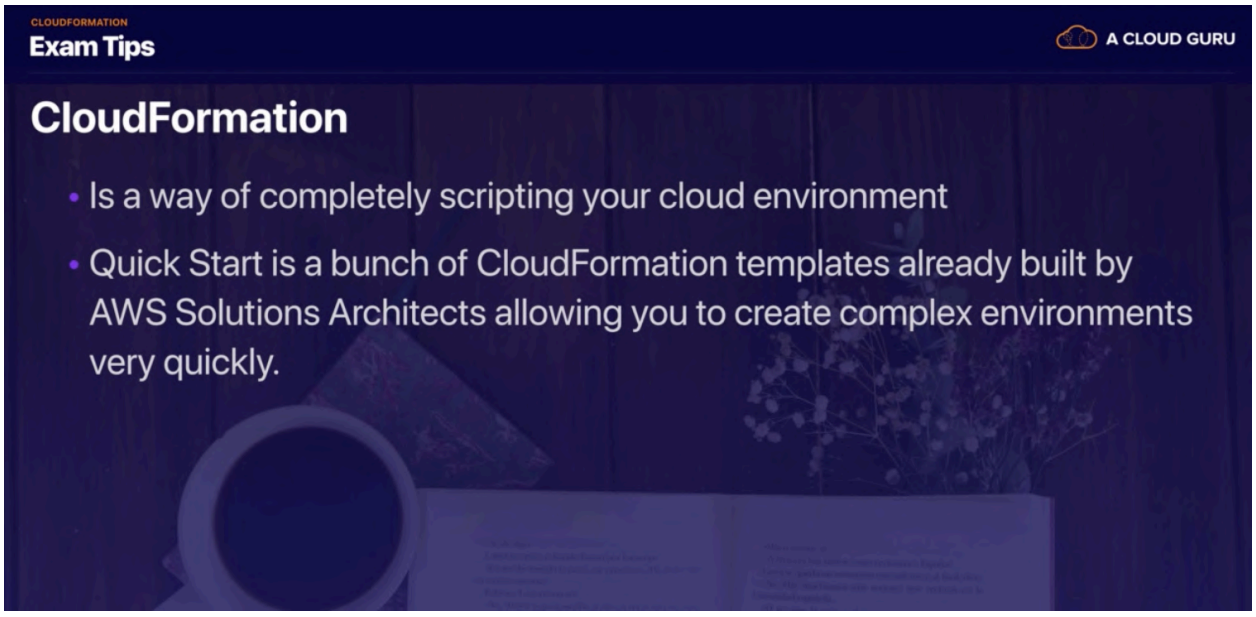
BUILDING A FAULT TOLERANT WORDPRESS SITE: LAB 1 - GETTING SETUP

Network Diagram

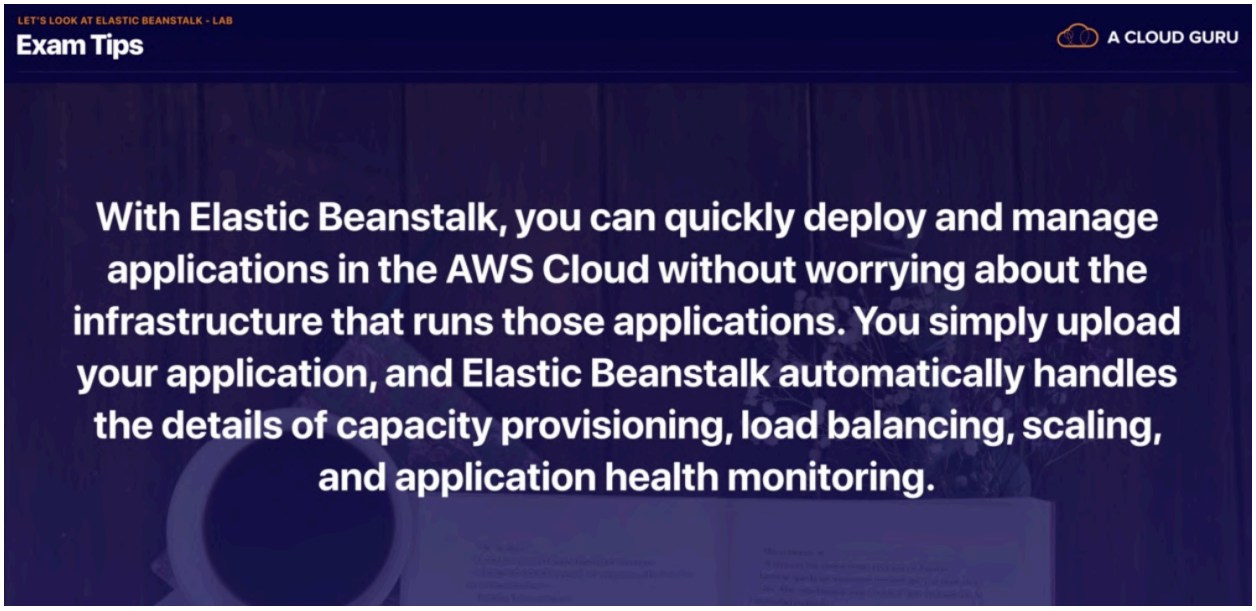
A CLOUD GURU







Elastic beanstalk



Architecture important resources

https://d1.awsstatic.com/whitepapers/architecture/AWS-Operational-Excellence-Pillar.pdf

https://d1.awsstatic.com/whitepapers/AWS_Cloud_Best_Practices.pdf

Quiz

Good job!

The ALB has functionality to distinguish traffic for different targets (mysite.co/accounts vs. mysite.co/sales vs. mysite.co/support) and distribute traffic based on rules for target group, condition, and priority.

Question 1:

You have a website with three distinct services, each hosted by different web server autoscaling groups. Which AWS service should you use?

- S3 static websites
- Elastic Load Balancers (ELB)
- Application Load Balancers (ALB)
- Classic Load Balancers (CLB)
- Network Load Balancers (NLB)

1.

Good job!

The Network Load Balancer is specifically designed for high performance traffic that is not conventional web traffic. The Classic LB might also do the job, but would not offer the same performance.

Question 2:

You manage a high-performance site that collects scientific data using a bespoke protocol over TCP port 1414. The data comes in at high speed and is distributed to an autoscaling group of EC2 compute services spread over three AZs. Which type of AWS load balancer would best meet this requirement?

- CloudFront combined with Lambda@Edge
- Elastic Load Balancers (ELB)
- Application Load Balancers (ALB)
- Network Load Balancers (NLB)

Good job!

In a question like this you need to evaluate if all the necessary services are in place. The glaring omission is that you have not built an autoscaling group to invoke the launch configuration you specified. The instance count and health check depend on instances being created by the autoscaling group. Finally, key pairs have no relevance to services running on the instance.

Question 3:

You have been tasked with creating a resilient website for your company. You create the Classic Load Balancer with a standard health check, a Route 53 alias pointing at the ELB, and a launch configuration based on a reliable Linux AMI. You have also checked all the security groups, NACLs, routes, gateways and NATs. You run the first test and cannot reach your web servers via the ELB or directly. What might be wrong?

- (●) The launch configuration is not being triggered correctly.
- ( ) The health check is not set up correctly.
- ( ) Your autoscaling group is set to zero instances.
- ( ) You have specified the wrong key pair and the servers cannot start the http service properly.

Good job!

Scaling out is where you have more of the same resource separately working in parallel (visualize services sitting side by side). Scaling up is where you make it bigger and bigger like and ugly tower with more floors being added after the initial design was finished.

Question 4:

If you are told that an EC2 instance is being changed to have more RAM, Is this considered scaling up or scaling out

- [ ] Scaling out
- [x] Scaling up

Good job!

Each word has a specific meaning and your ability to select the correct answer may depend on understanding the difference. Availability can be described as the % of a time period when the service will be able to respond to your request in some fashion.

Question 5:

In discussions about cloud services the words 'availability', 'durability', 'reliability' and 'resiliency' are often used. Which term is used to refer to the likelihood that you can access a resource or service when you need it?

- [x] Availability
- [ ] Durability
- [ ] Resiliency
- [ ] Reliability

Good job!

Each word has a specific meaning and your ability to select a correct answer may depend on understanding the difference. Durability refers to the on-going existence of the object or resource. Note that it does not mean you can access it, only that it continues to exist.

Question 6:

In discussions about cloud services the words 'availability', 'durability', 'reliability' and 'resiliency' are often used. Which term is used to refer to the likelihood that a resource will continue to exist until you decide to remove it?

- Availability
- Durability
- Resiliency
- Reliability

Good job!

Each word has a specific meaning and your ability to select the correct answer may depend on understanding the difference. Resiliency can be described as the ability to a system to self heal after damage or an event. Note that this does not mean that it will be available continuously during the event, only that it will self recover.

Question 7:

In discussions about cloud services the words 'availability', 'durability', 'reliability' and 'resiliency' are often used. Which term is used to refer to the likelihood that a resource ability to recover from damage or disruption?

- Availability
- Durability
- Resiliency
- Reliability

Good job!

Each word has a specific meaning and your ability to select a correct answer may depend on understanding the difference. Reliability is closely related to availability, however a system can be 'available' but not be working properly. Reliability is the probability that a system will work as designed. This term is not used much in AWS, but is still worth understanding.

Question 8:

In discussions about cloud services the words 'availability', 'durability', 'reliability' and 'resiliency' are often used. Which term is used to refer to the likelihood that a resource will work as designed?

- [ ] Availability
- [ ] Durability
- [ ] Resiliency
- [x] Reliability

Good job!

The key drivers here are availability and cost, so an awareness of cost is necessary to answer this. Full S3 is quite expensive at around $0.023 per GB for the lowest band. S3 standard IA is $0.0125 per GB, S3 OneZone-IA is $0.01 per GB, and Legacy S3-RRS is around $0.024 per GB for the lowest band. Of the offered solutions S3 One Zone-IA is the cheapest suitable option. Glacier cannot be considered as it is not intended for direct access, however it comes in at around $0.004 per GB. S3 has an availability of 99.99%, S3-IA has an availability of 99.9% while S3-1Zone-IA only has 99.5%.

Question 9:

You work for a major news network in Europe. They have just released a new mobile app that allows users to post their photos of newsworthy events in real-time. Your organization expects this app to grow very quickly, essentially doubling its user base each month. The app uses S3 to store the images, and you are expecting sudden and sizable increases in traffic to S3 when a major news event takes place (as users will be uploading large amounts of content.) You need to keep your storage costs to a minimum, and you are happy to temporally lose access to up to 0.1% of uploads per year. With these factors in mind, which storage media should you use to keep costs as low as possible?

- [x] S3 Standard-IA
- [ ] S3 Standard
- [ ] S3 – OneZone-Infrequent Access
- [ ] S3 – Reduced Redundancy Storage (RRS)
- [ ] Glacier
- [ ] S3 – Provisioned IOPS

Good job!

S3 OneZone-IA provides on-line access to files, while offering the same 11 9's of durability as all other storage classes. The trade-off is in the availability - 99.5% as opposed to 99.9%-99.99%. However in this brief as cost is more important than availability, S3 OneZone-IA is the logical choice . RRS is deprecated and new uses are strongly discouraged by AWS.

Question 10:

You work for a manufacturing company that operate a hybrid infrastructure with systems located both in a local data center and in AWS, connected via AWS Direct Connect. Currently, all on-premise servers are backed up to a local NAS, but your CTO wants you to decide on the best way to store copies of these backups in AWS. He has asked you to propose a solution which will provide access to the files within milliseconds should they be needed, but at the same time minimizes cost. As these files will be copies of backups stored on-premise, availability is not as critical as durability. Choose the best option from the following which meets the brief.

- Copy the files from the NAS to an S3 bucket configured as Standard class.
- Copy the files to an EC2 instance with a large EBS volume attached.
- Copy the files from the NAS to an S3 bucket with the One Zone-IA class
- Copy the files from the NAS to an S3 bucket with the Reduced Redundancy Storage class.

Good job!

The key point in the questions is that the data is non-replaceable and is frequently updated. The 1st excludes anything the has reduced durability, the second excluded anything with long recall, reduced availability, or billing based on infrequent access.

Question 11:

You need to use an object-based storage solution to store your critical, non-replaceable data in a cost-effective way. This data will be frequently updated and will need some form of version control enabled on it. Which S3 storage solution should you use?

- [x] S3
- [ ] S3 – IA
- [ ] S3 – OneZone-IA
- [ ] S3 – RRS
- [ ] Glacier

Good job!

Question 12:

In S3 the durability of my files is _____.

- 99.99%
- **99.999999999%**
- 99%
- 100%

Good job!

Question 13:

When you have deployed an RDS database into multiple availability zones, can you use the secondary database as an independent read node?

- **No.**
- Only in US-West-1.
- It depends on how you set it up.
- Yes.

Good job!

There is only one answer that is specific to Spread Placement Groups, and that is the final option. Whilst some of these answers are correct for either Cluster Placement Groups only, or for both Cluster and Spread Placement Groups, the question stated that only options specific to Spread Placement Groups should be chosen. This would rule out two options as they are true for both Spread & Cluster type placement groups. The Logical grouping of instances within a single Availability Zone is only true of Cluster Placement Groups and is also incorrect.

Question 14:

Placement groups can either be of the type 'cluster', 'spread', or 'partition'. Choose options from below which are only specific to Spread Placement Groups.

- Spread placement groups require a name that is unique within your AWS account for the region.
- An instance can be launched in one placement group at a time and cannot span multiple placement groups.
- A spread placement group is a logical grouping of instances within a single Availability Zone.
- A spread placement group is a group of instances that are each placed on distinct underlying hardware.

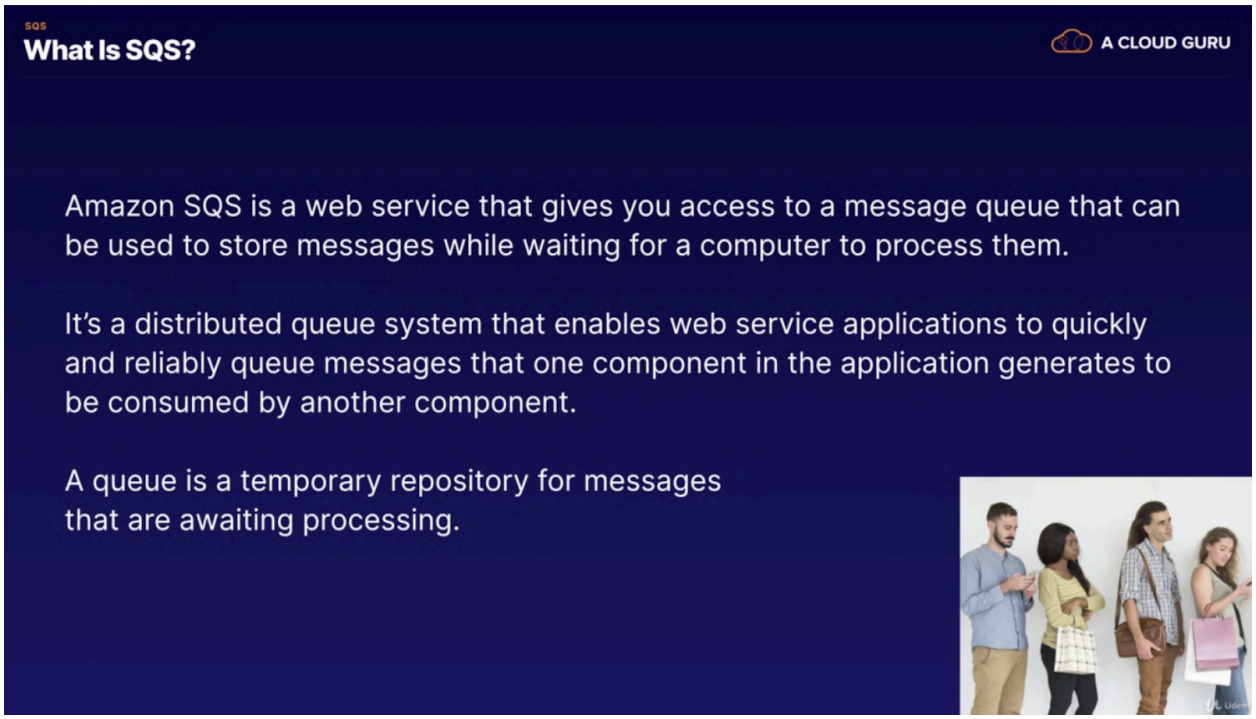
Good job!

Question 15:

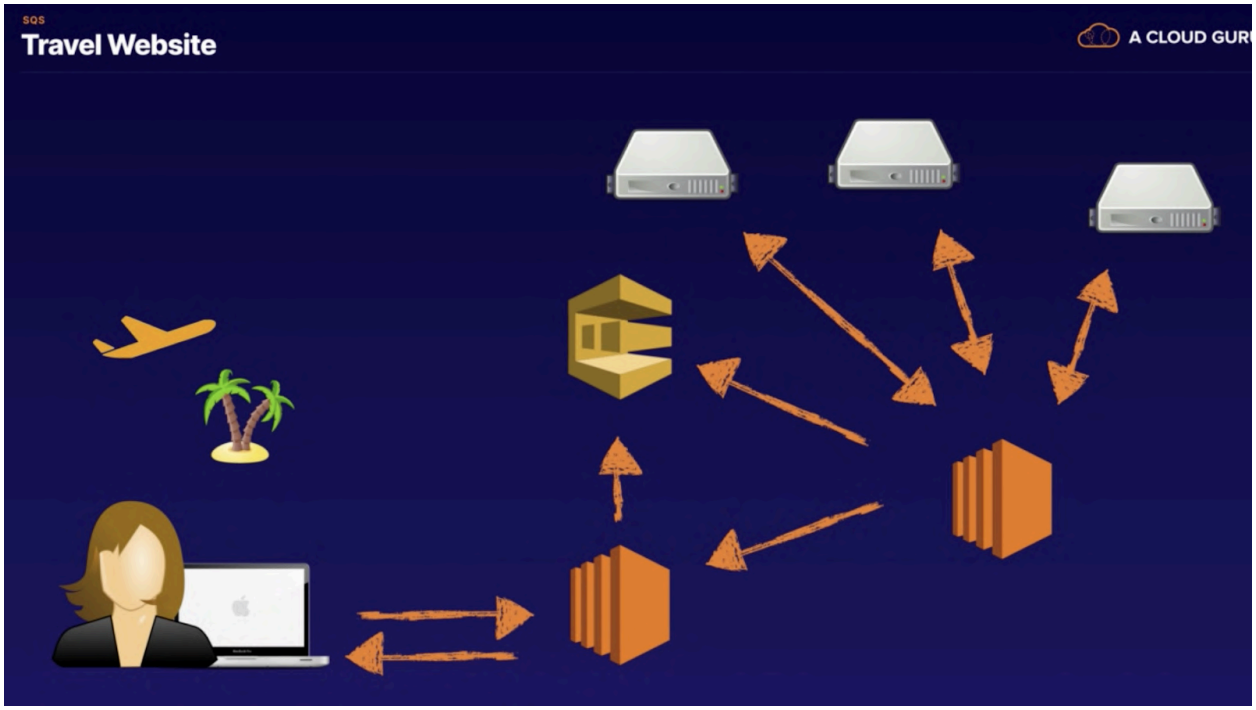
Can I "force" a failover for any RDS instance that has multi-AZ configured?

- (•) Yes.
- ( ) No.
- ( ) Only for Oracle RDS instances.

Application
SQS



Example


SQS
Travel Website
A CLOUD GUR

SQS

What Is SQS?

A CLOUD GURU

Using Amazon SQS, you can decouple the components of an application so they run independently, easing message management between components.

Any component of a distributed application can store messages in a fail-safe queue.

Messages can contain up to 256 KB of text in any format. Any component can later retrieve the messages programmatically using the Amazon SQS API.

SQS

What Is SQS?

A CLOUD GURU

The queue acts as a buffer between the component producing and saving data, and the component receiving the data for processing.

This means the queue resolves issues that arise if the producer is producing work faster than the consumer can process it, or if the producer or consumer are only intermittently connected to the network.

SQS

Queue Types

A CLOUD GURU

There are two types of queue:

- Standard queues (default)
- FIFO queues

SQS

Standard Queues

A CLOUD GURU

Amazon SQS offers standard as the default queue type. A standard queue lets you have a **nearly-unlimited number of transactions per second**. Standard queues guarantee that a message is delivered at least once.

Occasionally (because of the highly-distributed architecture that allows high throughput), more than one copy of a message might be delivered out of order.

However, standard queues provide best-effort ordering which ensures that messages are generally delivered in the same order as they are sent.

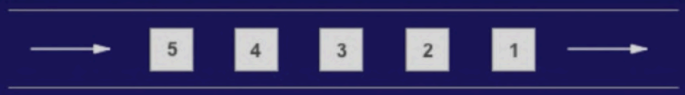
SQS

FIFO Queues

A CLOUD GURU

The FIFO queue complements the standard queue.

The most important features of this queue type are **FIFO (first-in-first-out) delivery** and **exactly-once processing**: the order in which messages are sent and received is strictly preserved and a message is delivered once and remains available until a consumer processes and deletes it; duplicates are not introduced into the queue.



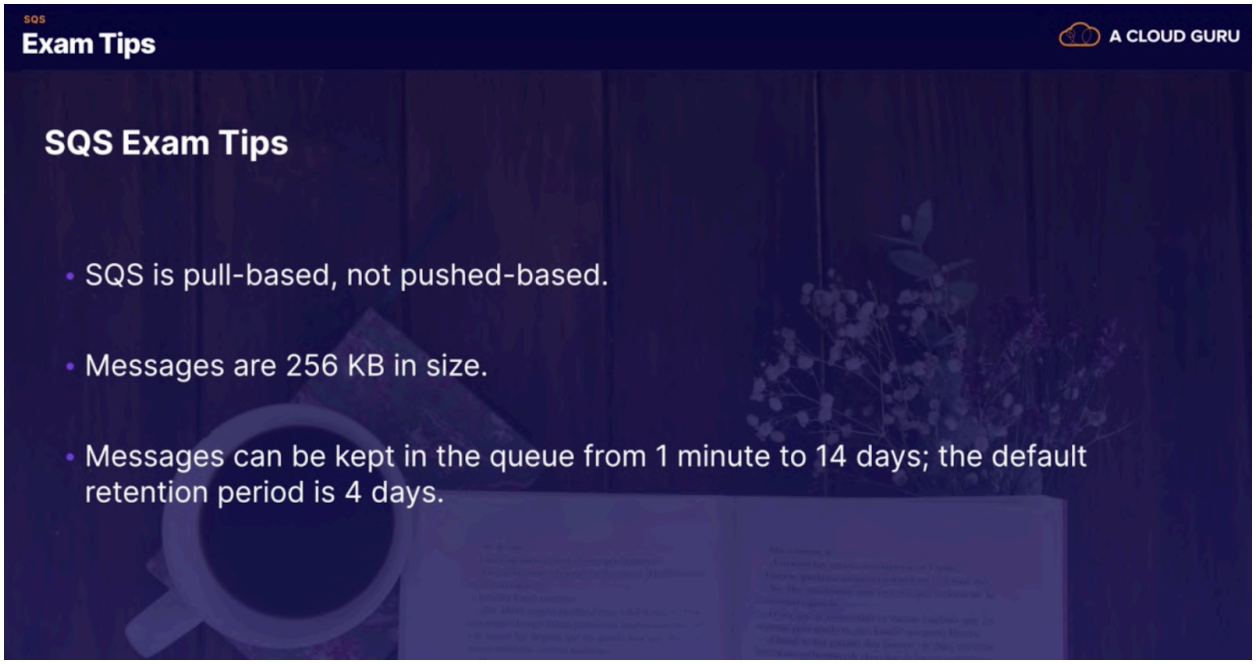
SQS

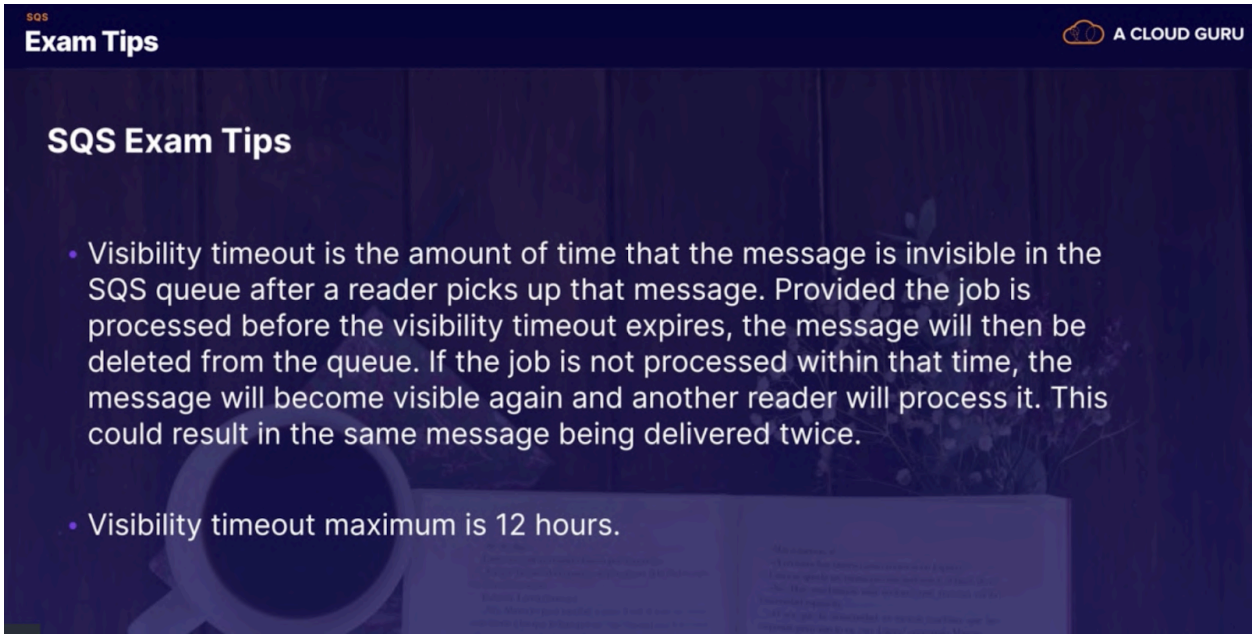
FIFO Queues

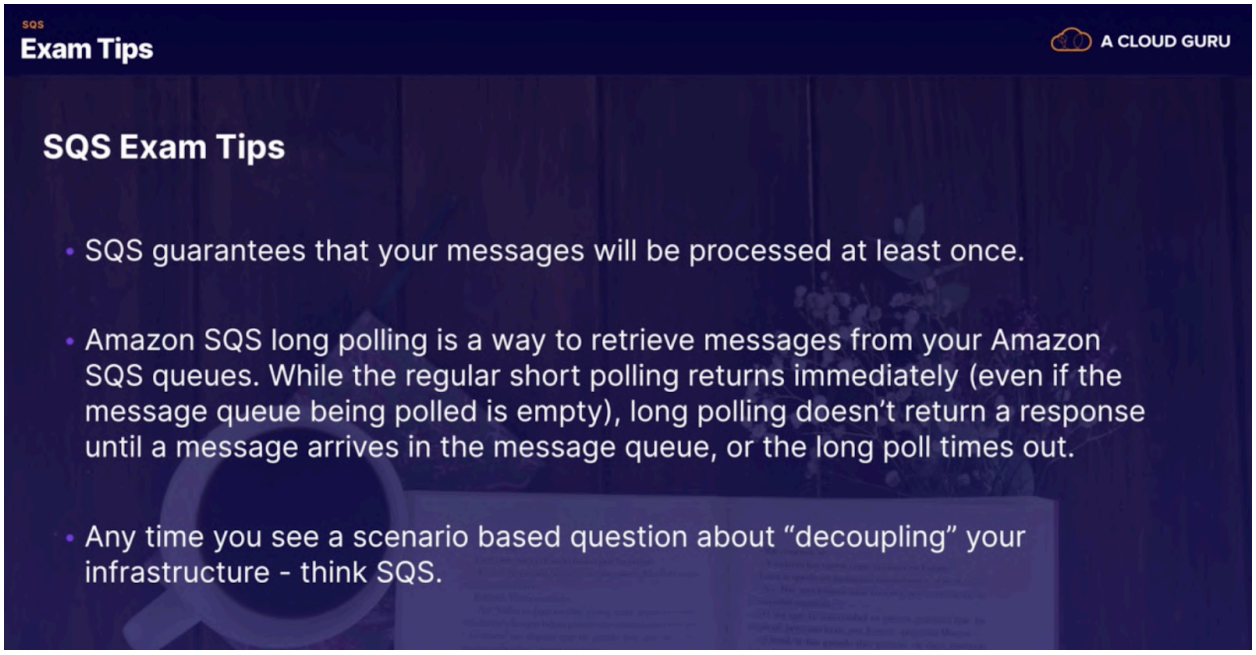
A CLOUD GURU

FIFO queues also support message groups that allow multiple ordered message groups within a single queue.

FIFO queues are limited to 300 transactions per second (TPS), but have all the capabilities of standard queues.


SQS Exam Tips
• SQS is pull-based, not pushed-based.
• Messages are 256 KB in size.
• Messages can be kept in the queue from 1 minute to 14 days; the default retention period is 4 days.


SQS
Exam Tips
A CLOUD GURU
SQS Exam Tips
• Visibility timeout is the amount of time that the message is invisible in the SQS queue after a reader picks up that message. Provided the job is processed before the visibility timeout expires, the message will then be deleted from the queue. If the job is not processed within that time, the message will become visible again and another reader will process it. This could result in the same message being delivered twice.
• Visibility timeout maximum is 12 hours.


SQS
Exam Tips
A CLOUD GURU
SQS Exam Tips
• SQS guarantees that your messages will be processed at least once.
• Amazon SQS long polling is a way to retrieve messages from your Amazon SQS queues. While the regular short polling returns immediately (even if the message queue being polled is empty), long polling doesn't return a response until a message arrives in the message queue, or the long poll times out.
• Any time you see a scenario based question about "decoupling" your infrastructure - think SQS.

SWF

SWF

What Is SWF?

A CLOUD GURU

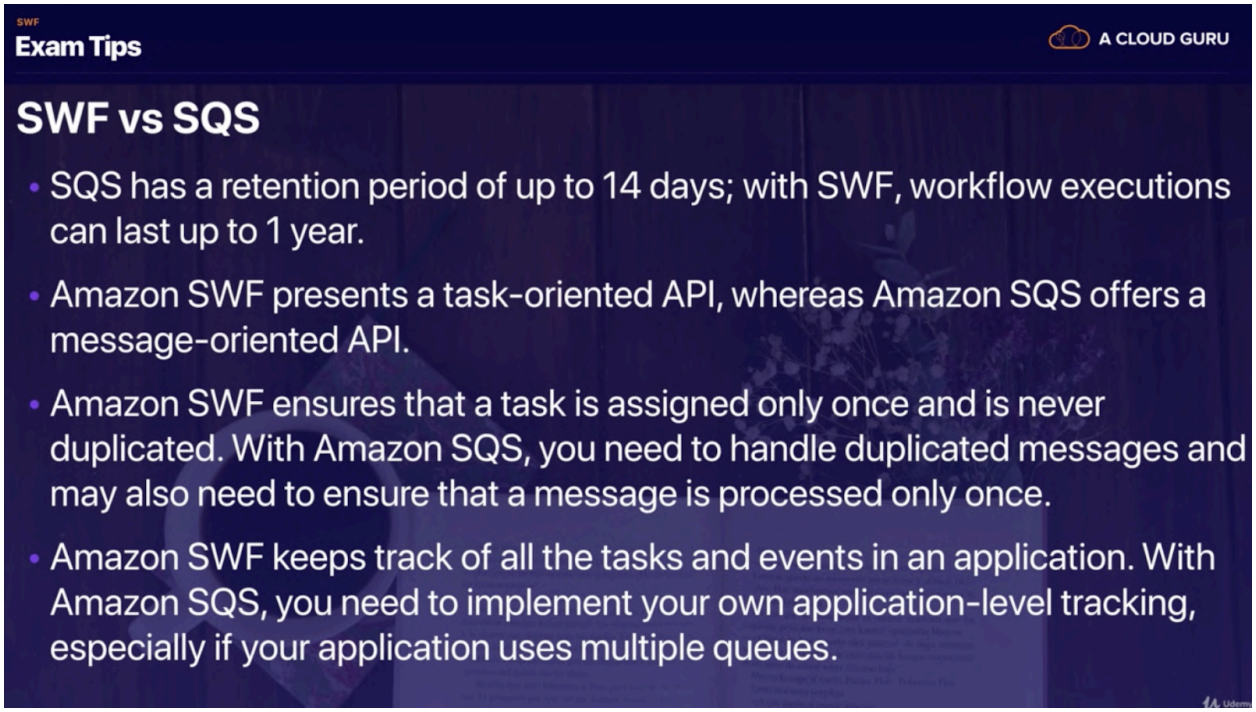
Amazon Simple Workflow Service (Amazon SWF) is a web service that makes it easy to coordinate work across distributed application components. SWF enables applications for a range of use cases, including media processing, web application back-ends, business process workflows, and analytics pipelines, to be designed as a coordination of tasks.

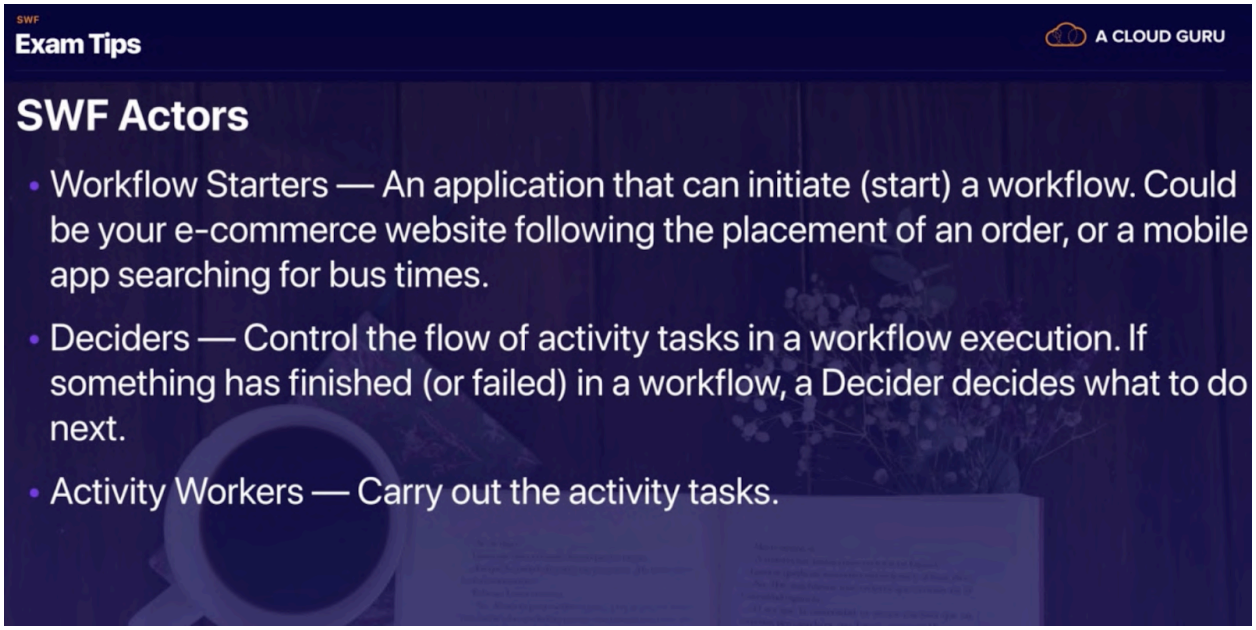
SWF

What Is SWF?

A CLOUD

Tasks represent invocations of various processing steps in an application which can be performed by executable code, web service calls, human actions, and scripts.


SWF
Exam Tips
A CLOUD GURU
SWF vs SQS
• SQS has a retention period of up to 14 days; with SWF, workflow executions can last up to 1 year.
• Amazon SWF presents a task-oriented API, whereas Amazon SQS offers a message-oriented API.
• Amazon SWF ensures that a task is assigned only once and is never duplicated. With Amazon SQS, you need to handle duplicated messages and may also need to ensure that a message is processed only once.
• Amazon SWF keeps track of all the tasks and events in an application. With Amazon SQS, you need to implement your own application-level tracking, especially if your application uses multiple queues.


SWF
Exam Tips
A CLOUD GURU
SWF Actors
• Workflow Starters — An application that can initiate (start) a workflow. Could be your e-commerce website following the placement of an order, or a mobile app searching for bus times.
• Deciders — Control the flow of activity tasks in a workflow execution. If something has finished (or failed) in a workflow, a Decider decides what to do next.
• Activity Workers — Carry out the activity tasks.

SNS

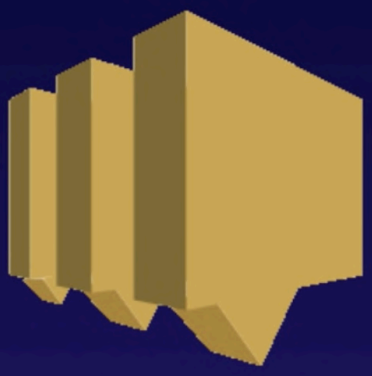
SIMPLE NOTIFICATION SERVICE

What Is SNS?

A CLOUD GURU

Amazon Simple Notification Service (Amazon SNS) is a web service that makes it easy to set up, operate, and send notifications from the cloud.

It provides developers with a highly scalable, flexible, and cost-effective capability to publish messages from an application and immediately deliver them to subscribers or other applications.



SIMPLE NOTIFICATION SERVICE

Push Notifications

A CLOUD GURU

Push notifications to Apple, Google, Fire OS, and Windows devices, as well as Android devices in China with Baidu Cloud Push.

Google

SIMPLE NOTIFICATION SERVICESNS

SQS Integration

A CLOUD GURU

Besides pushing cloud notifications directly to mobile devices, Amazon SNS can also deliver notifications by SMS text message or email to Amazon Simple Queue Service (SQS) queues, or to any HTTP endpoint.

SIMPLE NOTIFICATION SERVICE

What Is A Topic?

A CLOUD GURU

SNS allows you to group multiple recipients using topics. A topic is an "access point" for allowing recipients to dynamically subscribe for identical copies of the same notification.

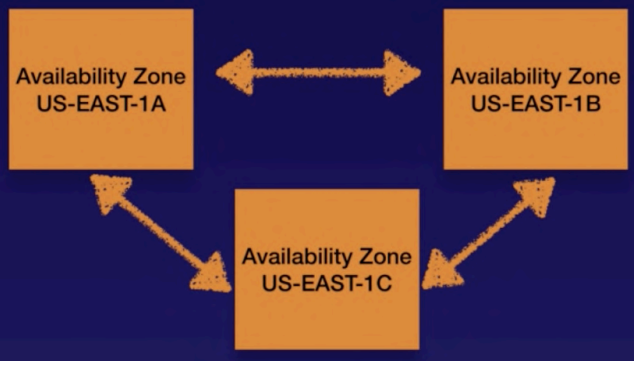
One topic can support deliveries to multiple endpoint types — for example, you can group together iOS, Android and SMS recipients. When you publish once to a topic, SNS delivers appropriately formatted copies of your message to each subscriber.

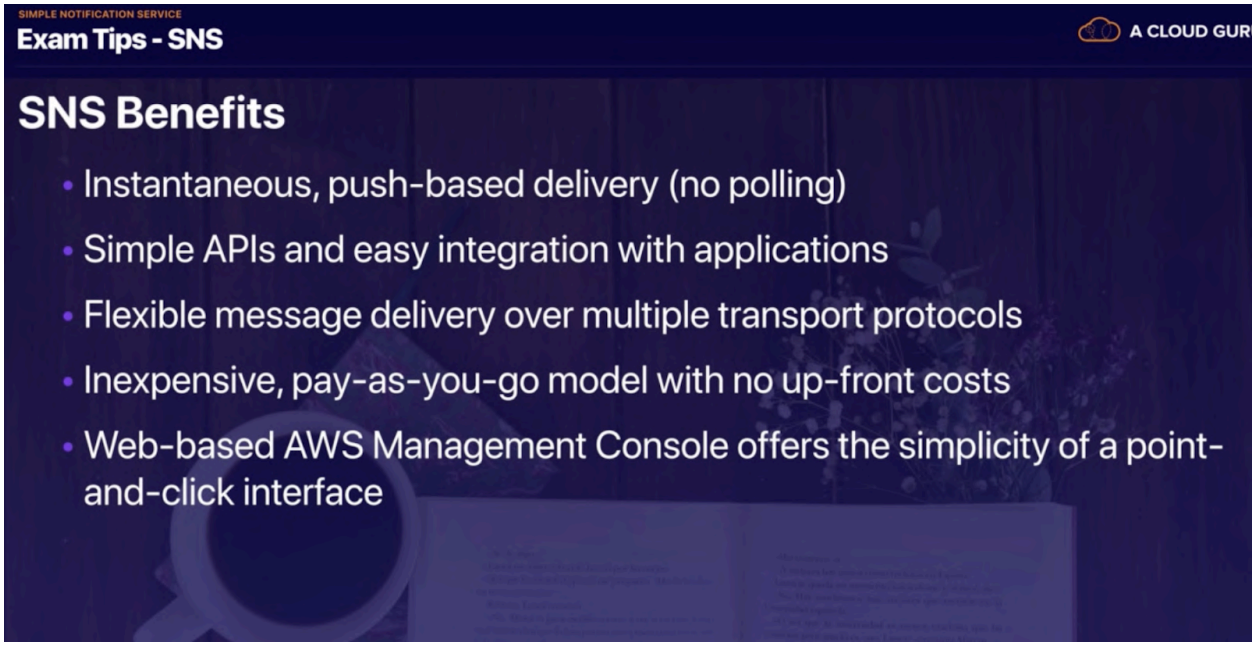
SIMPLE NOTIFICATION SERVICE

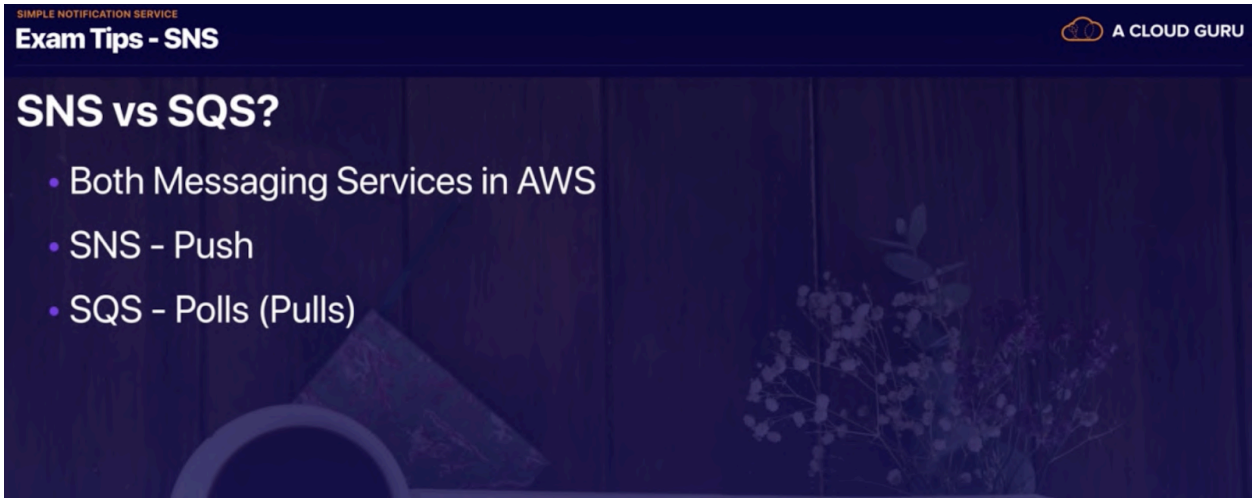
SNS Availability

A CLOUD GU

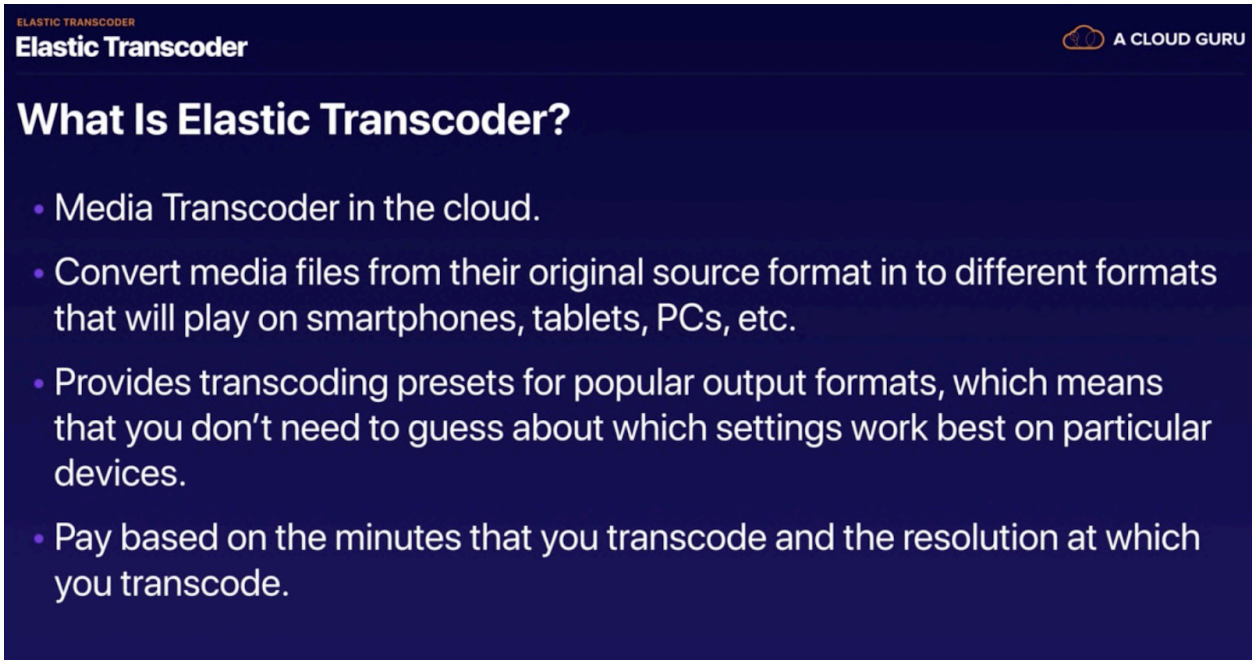
To prevent messages from being lost, all messages published to Amazon SNS are stored redundantly across multiple availability zones.

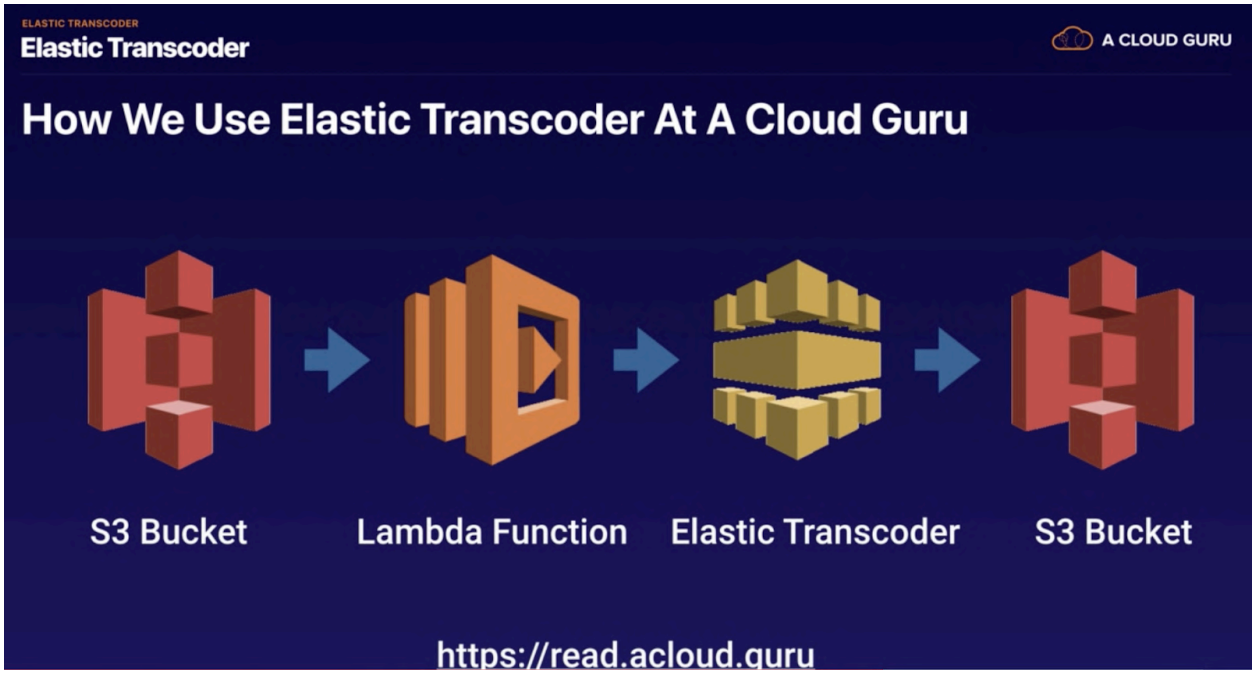


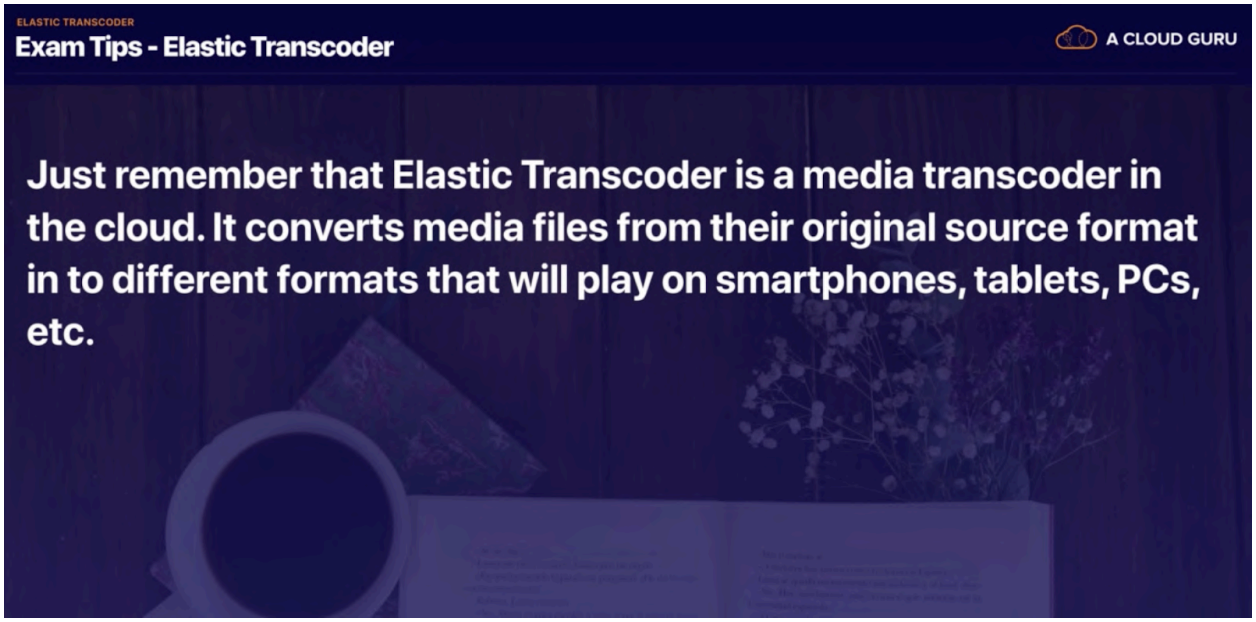

SIMPLE NOTIFICATION SERVICE
Exam Tips - SNS
A CLOUD GUR
SNS Benefits
• Instantaneous, push-based delivery (no polling)
• Simple APIs and easy integration with applications
• Flexible message delivery over multiple transport protocols
• Inexpensive, pay-as-you-go model with no up-front costs
• Web-based AWS Management Console offers the simplicity of a point-and-click interface


SIMPLE NOTIFICATION SERVICE
Exam Tips - SNS
A CLOUD GURU
SNS vs SQS?
• Both Messaging Services in AWS
• SNS - Push
• SQS - Polls (Pulls)

Elastic Transcoder




ELASTIC TRANSCODER
Elastic Transcoder
A CLOUD GURU
How We Use Elastic Transcoder At A Cloud Guru
S3 Bucket
Lambda Function
Elastic Transcoder
S3 Bucket
https://read.acloud.guru


ELASTIC TRANSCODER
Exam Tips - Elastic Transcoder
A CLOUD GURU
Just remember that Elastic Transcoder is a media transcoder in the cloud. It converts media files from their original source format in to different formats that will play on smartphones, tablets, PCs, etc.

API Gateway

API GATEWAY

What Is API Gateway?

A CLOUD GURU

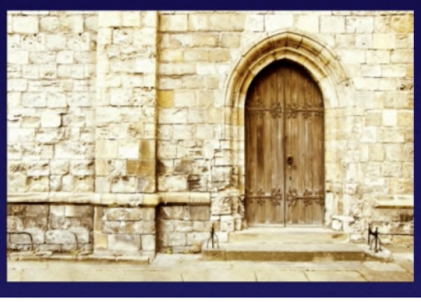
Amazon API Gateway is a fully managed service that makes it easy for developers to publish, maintain, monitor, and secure APIs at any scale.

API GATEWAY

What Is API Gateway?

A CLOUD GURU

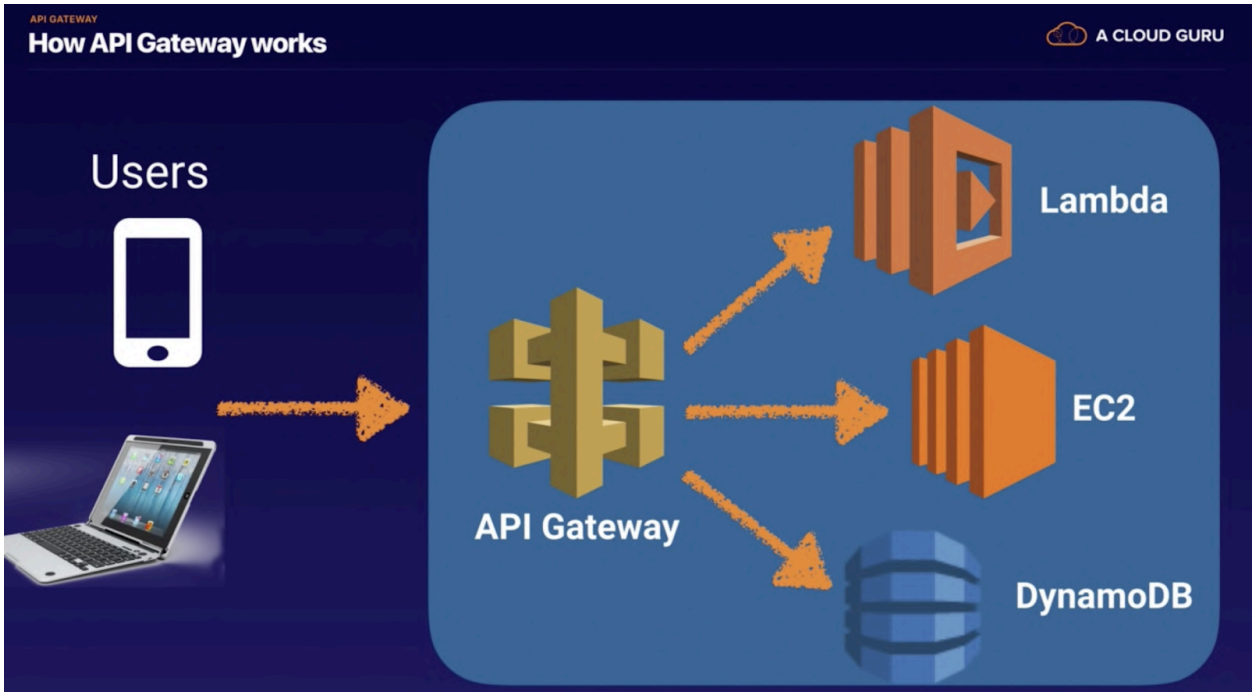
With a few clicks in the AWS Management Console, you can create an API that acts as a "front door" for applications to access data, business logic, or functionality from your back-end services, such as applications running on Amazon Elastic Compute Cloud (Amazon EC2), code running on AWS Lambda, or any web application.



API GATEWAY

How API Gateway works

A CLOUD GURU



API GATEWAY

API Gateway Options

A CLOUD G

What Can API Gateway Do?

- Expose HTTPS endpoints to define a RESTful API
- Serverless-ly connect to services like Lambda & DynamoDB
- Send each API endpoint to a different target
- Run efficiently with low cost
- Scale effortlessly
- Track and control usage by API key
- Throttle requests to prevent attacks
- Connect to CloudWatch to log all requests for monitoring
- Maintain multiple versions of your API

API GATEWAY

API Gateway Configuration

A CLOUD GURU

How Do I Configure API Gateway?

- Define an API (container)
- Define Resources and nested Resources (URL paths)
- For each Resource:
  - Select supported HTTP methods (verbs)
  - Set security
  - Choose target (such as EC2, Lambda, DynamoDB, etc.)
  - Set request and response transformations

API GATEWAY

API Gateway Deployments

A CLOUD GURU

How Do I Deploy API Gateway?

- Deploy API to a stage:
  - Uses API Gateway domain, by default
  - Can use custom domain
  - Now supports AWS Certificate Manager: free SSL/TLS certs

API GATEWAY

API Gateway Caching

A CLOUD GURU

You can enable API caching in Amazon API Gateway to cache your endpoint's response. With caching, you can reduce the number of calls made to your endpoint and also improve the latency of the requests to your API. When you enable caching for a stage, API Gateway caches responses from your endpoint for a specified time-to-live (TTL) period, in seconds. API Gateway then responds to the request by looking up the endpoint response from the cache instead of making a request to your endpoint.

API GATEWAY

Same Origin Policy

A CLOUD GURU

In computing, the same-origin policy is an important concept in the web application security model. Under the policy, a web browser permits scripts contained in a first web page to access data in a second web page, but only if both web pages have the same origin.

This is done to prevent Cross-Site Scripting (XSS) attacks.

- Enforced by web browsers.
- Ignored by tools like PostMan and curl.

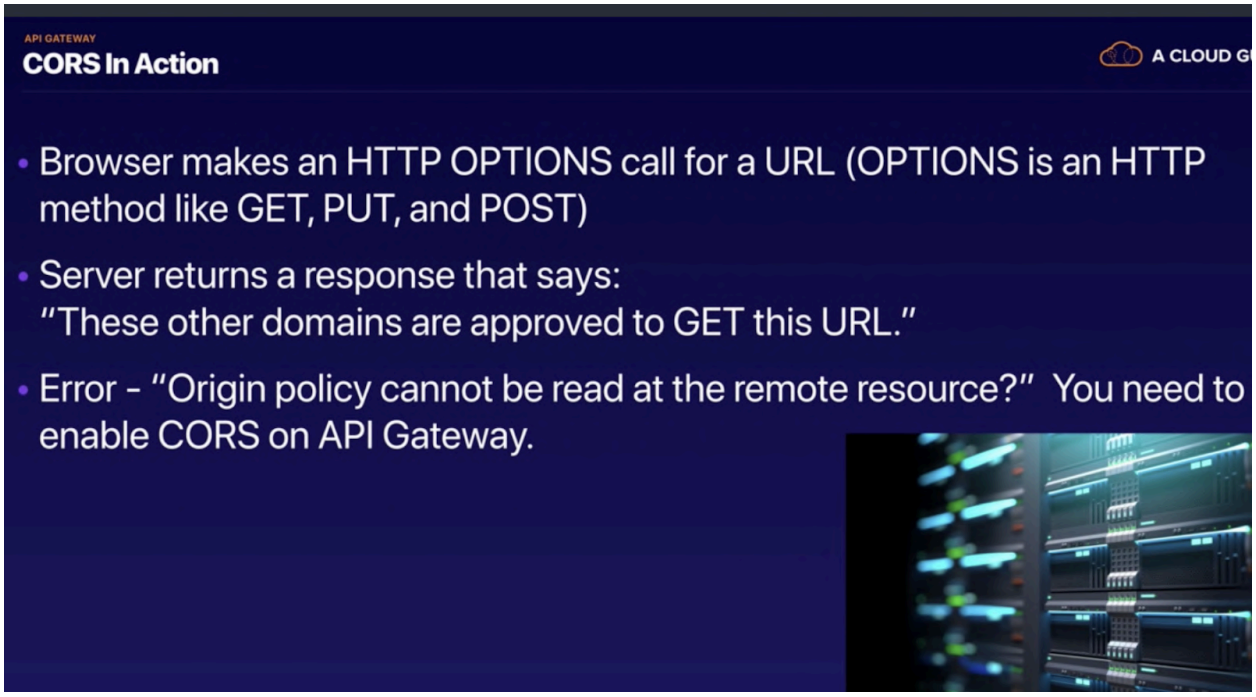
API GATEWAY

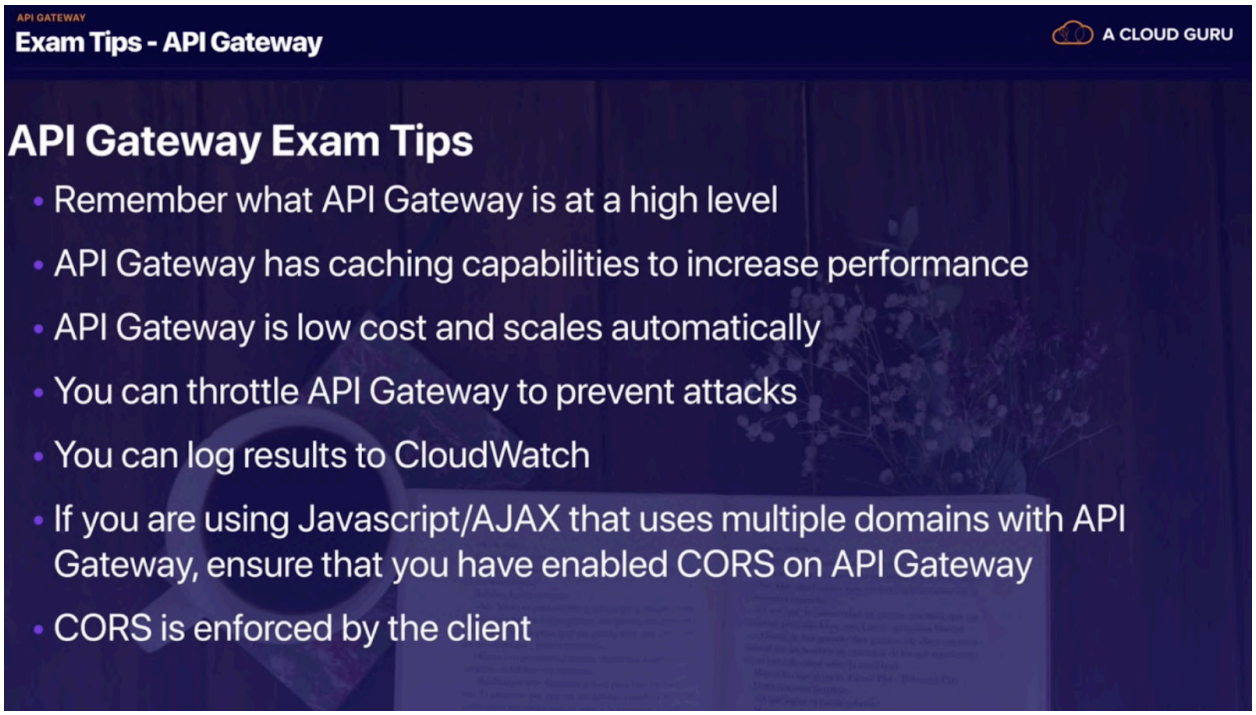
CORS Explained

A CLOUD GURU

CORS is one way the server at the other end (not the client code in the browser) can relax the same-origin policy.

Cross-origin resource sharing (CORS) is a mechanism that allows restricted resources (e.g. fonts) on a web page to be requested from another domain outside the domain from which the first resource was served.


API GATEWAY
CORS In Action
A CLOUD G
• Browser makes an HTTP OPTIONS call for a URL (OPTIONS is an HTTP method like GET, PUT, and POST)
• Server returns a response that says: "These other domains are approved to GET this URL."
• Error - "Origin policy cannot be read at the remote resource?" You need to enable CORS on API Gateway.


API GATEWAY
Exam Tips - API Gateway
A CLOUD GURU
API Gateway Exam Tips
• Remember what API Gateway is at a high level
• API Gateway has caching capabilities to increase performance
• API Gateway is low cost and scales automatically
• You can throttle API Gateway to prevent attacks
• You can log results to CloudWatch
• If you are using Javascript/AJAX that uses multiple domains with API Gateway, ensure that you have enabled CORS on API Gateway
• CORS is enforced by the client

Kinesis

KINESIS 101

What Is Streaming Data?

A CLOUD GURU

Streaming Data is data that is generated continuously by thousands of data sources, which typically send in the data records simultaneously, and in small sizes (order of Kilobytes.)

- Purchases from online stores (think amazon.com)
- Stock Prices
- Game data (as the gamer plays)
- Social network data
- Geospatial data (think uber.com)
- iOT sensor data

KINESIS 101

What Is Kinesis?

A CLOUD

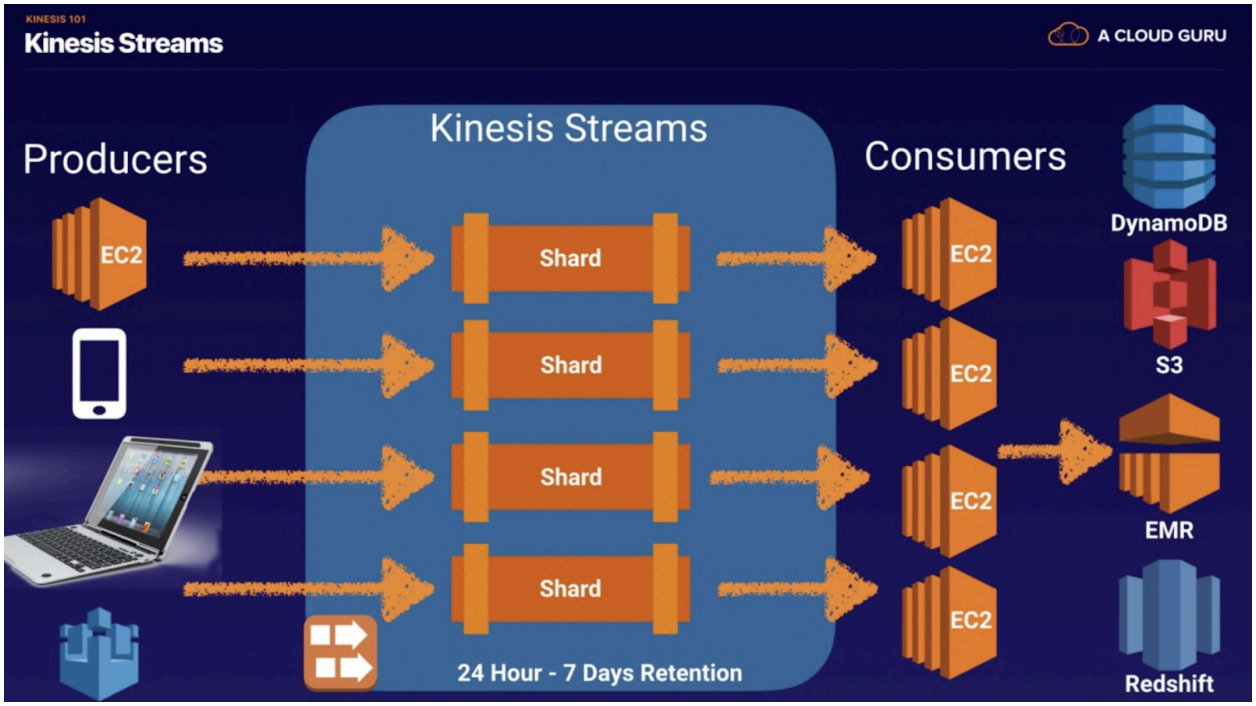
Amazon Kinesis is a platform on AWS to send your streaming data to. Kinesis makes it easy to load and analyze streaming data, and also providing the ability for you to build your own custom applications for you business needs.

KINESIS 101

Types of Kinesis

3 Different Types of Kinesis

- Kinesis Streams
- Kinesis Firehose
- Kinesis Analytics



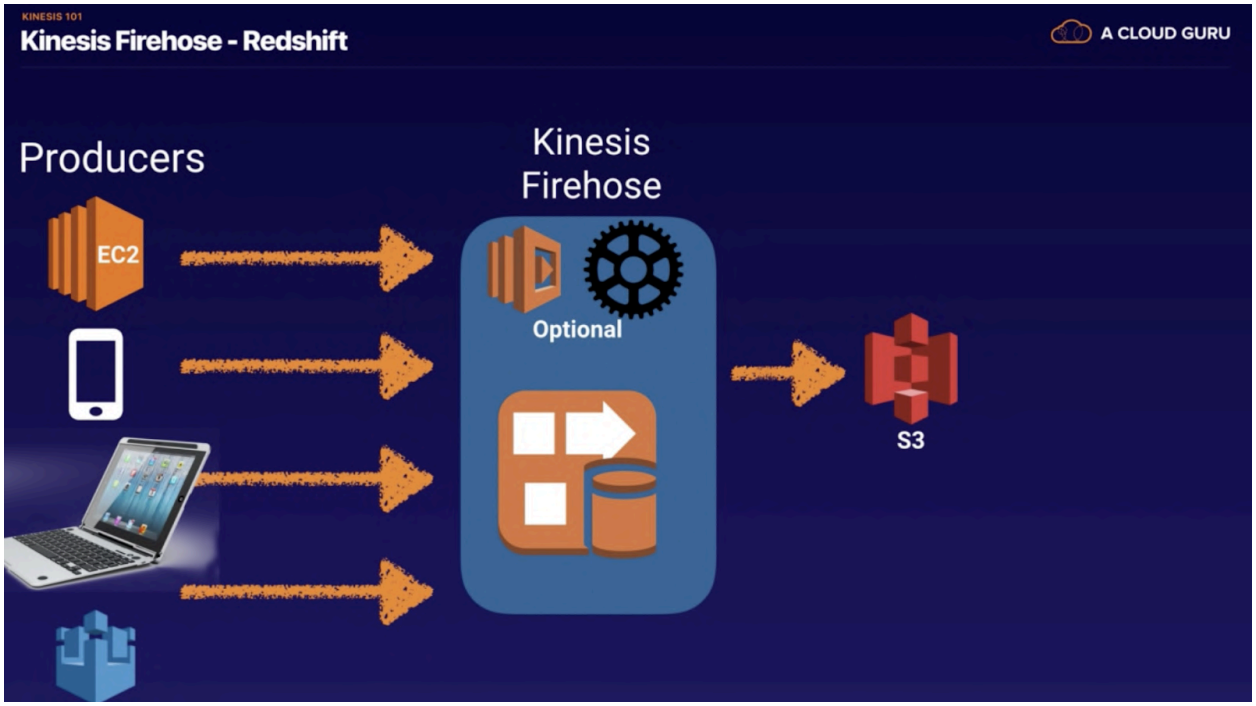
KINESIS 101

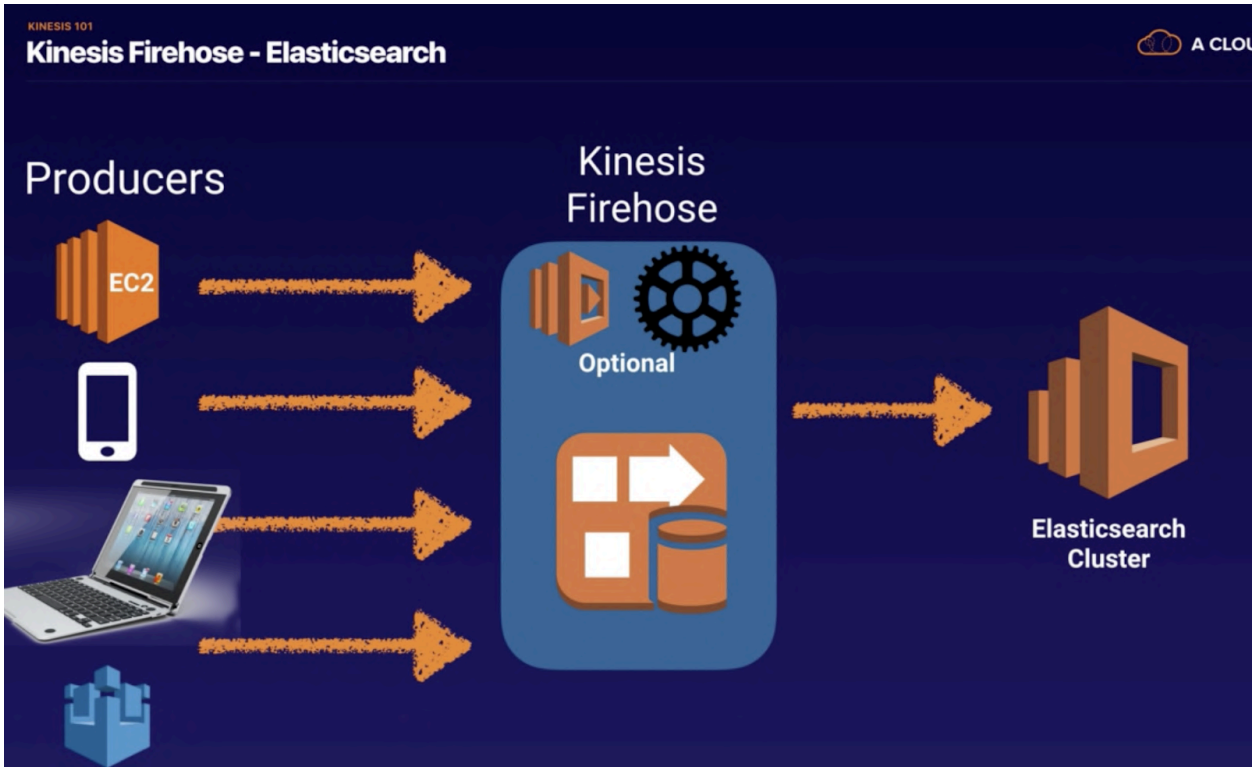
Types of Kinesis

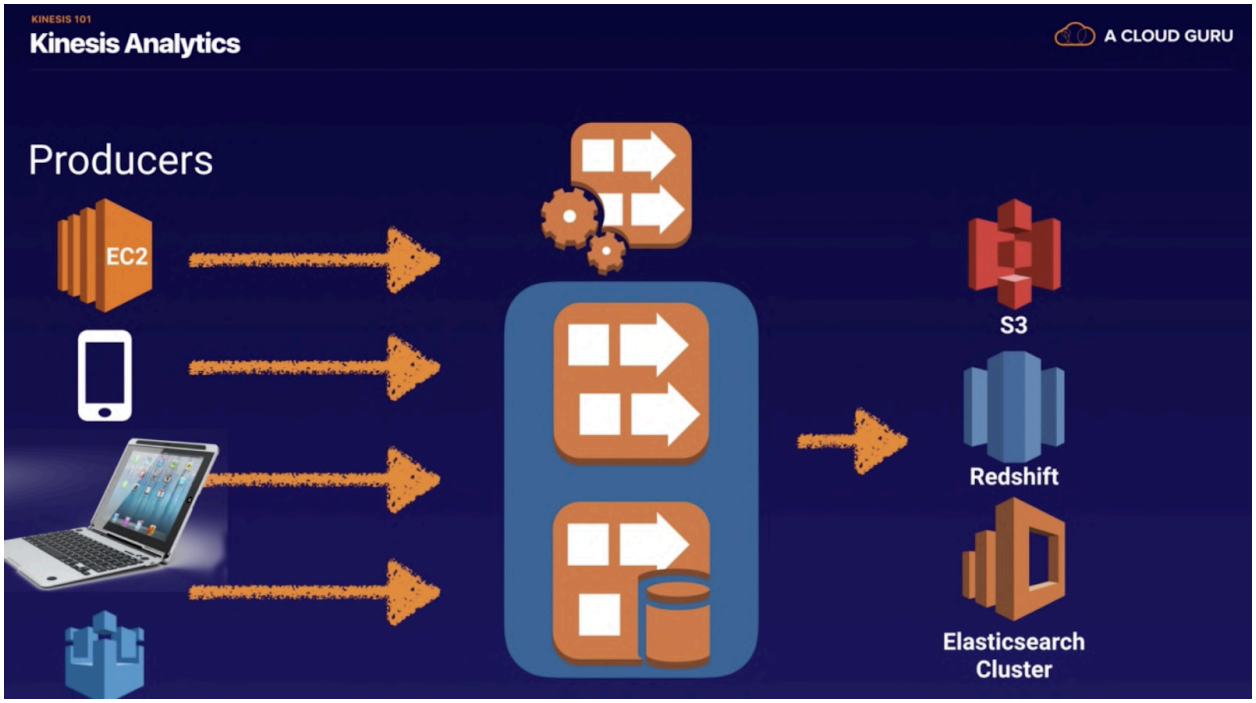
A CLOUD GURU

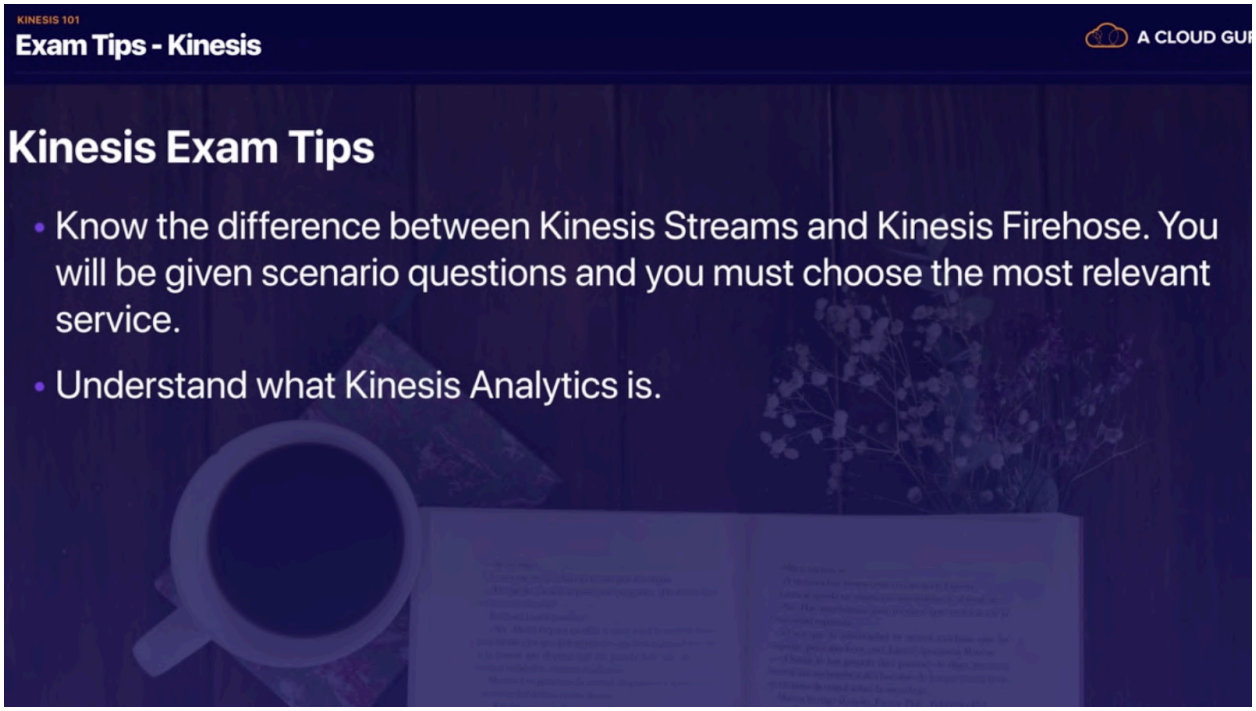
Kinesis Streams Consist Of Shards;

- 5 transactions per second for reads, up to a maximum total data read rate of 2 MB per second and up to 1,000 records per second for writes, up to a maximum total data write rate of 1 MB per second (including partition keys.)
- The data capacity of your stream is a function of the number of shards that you specify for the stream. The total capacity of the stream is the sum of the capacities of its shards.

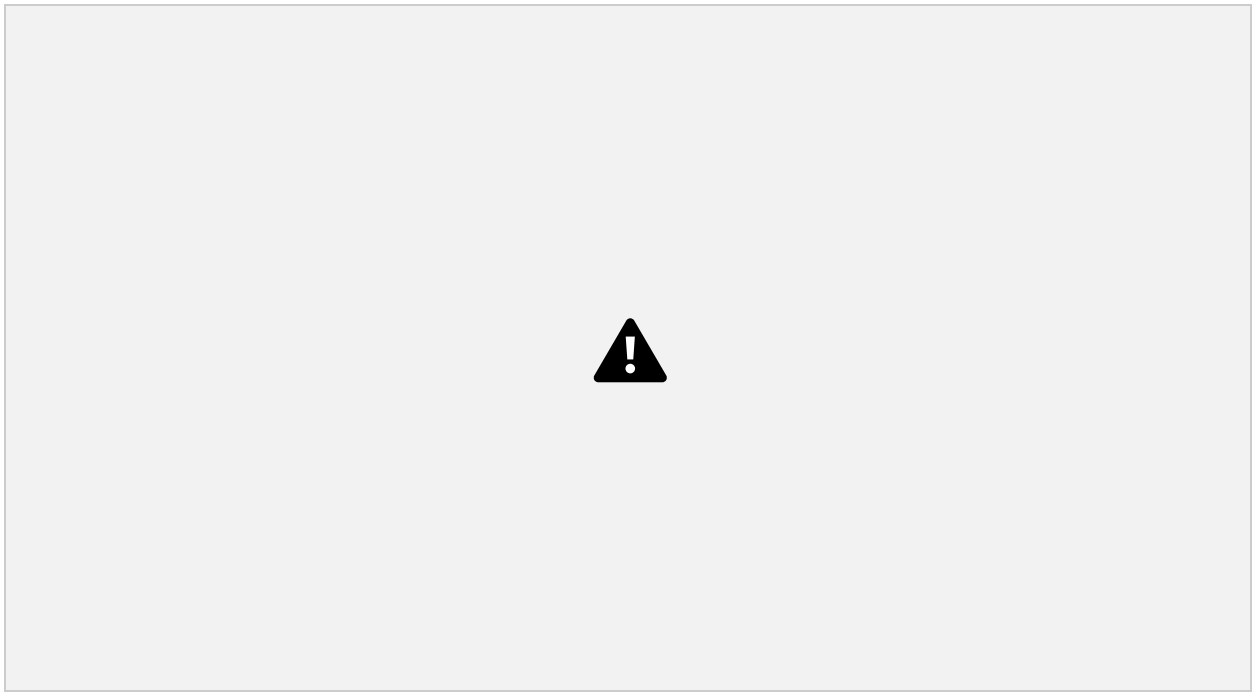

KINESIS 101
Kinesis Firehose - Redshift
A CLOUD GURU
Producers
Kinesis
Firehose
EC2
Optional
S3


KINESIS 101
Kinesis Firehose - Elasticsearch
Producers
Kinesis
Firehose
EC2
Optional
Elasticsearch
Cluster





WEB Identification Federation - Cognito



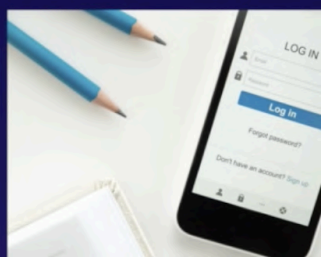
WEB IDENTITY FEDERATION & COGNITO

Amazon Cognito

A CLOU

Amazon Cognito provides Web Identity Federation with the following features:

- Sign-up and sign-in to your apps
- Access for guest users
- Acts as an Identity Broker between your application and Web ID providers, so you don't need to write any additional code.
- Synchronizes user data for multiple devices
- Recommended for all mobile applications AWS services.

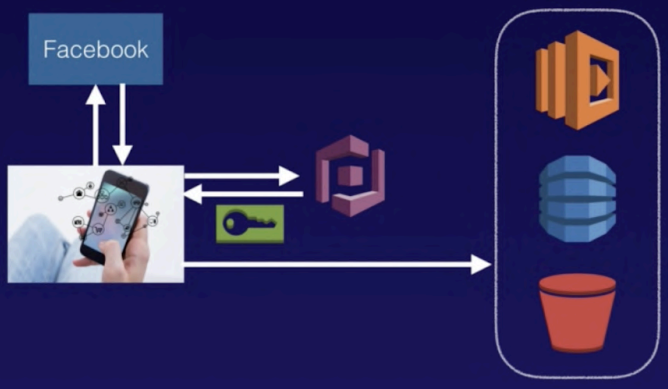


WEB IDENTITY FEDERATION & COGNITO

Amazon Cognito Use Cases

A CLOUD GURU

The recommended approach for Web Identity Federation using social media accounts like Facebook.



WEB IDENTITY FEDERATION & COGNITO

Amazon Cognito Use Cases

A CLOUD GURU

Cognito brokers between the app and Facebook or Google to provide temporary credentials which map to an IAM role allowing access to the required resources.

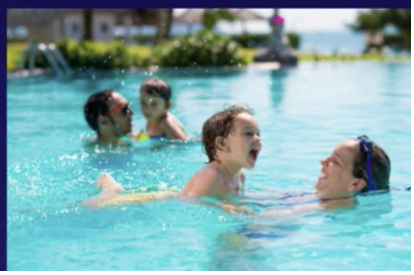
No need for the application to embed or store AWS credentials locally on the device and it gives users a seamless experience across all mobile devices.

WEB IDENTITY FEDERATION & COGNITO

Cognito User Pools

A CLOUD GURU

User Pools are user directories used to manage sign-up and sign-in functionality for mobile and web applications. Users can sign-in directly to the User Pool, or using Facebook, Amazon, or Google. Cognito acts as an Identity Broker between the identity provider and AWS. Successful authentication generates a JSON Web token (JWTs).



WEB IDENTITY FEDERATION & COGNITO

Cognito Identity Pools

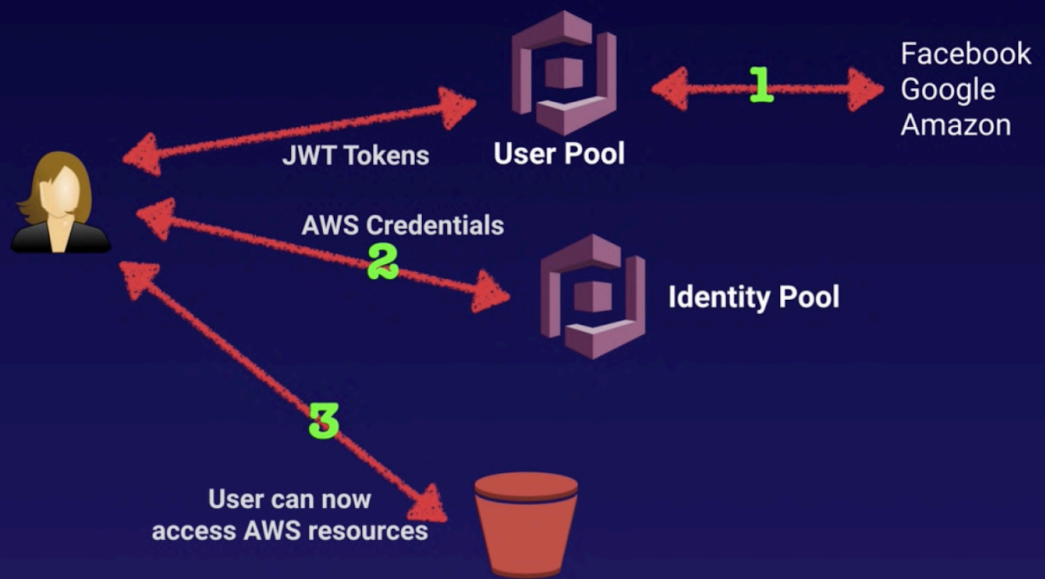
A CLOUD GURU

Identity Pools enable provide temporary AWS credentials to access AWS services like S3 or DynamoDB.

WEB IDENTITY FEDERATION & COGNITO

Amazon Cognito In Action

A CLOUD GURU

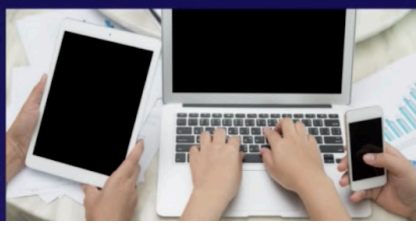


WEB IDENTITY FEDERATION & COGNITO

Cognito Synchronisation

A CLOUD GURU

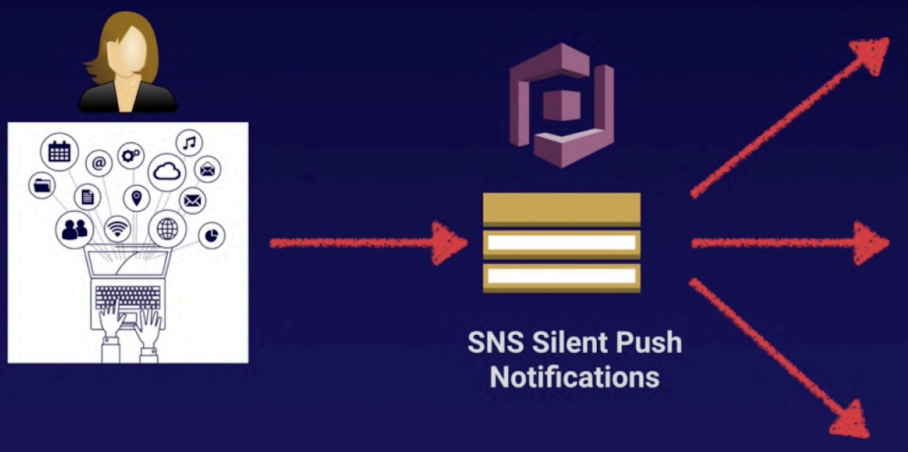
Cognito tracks the association between user identity and the various different devices they sign-in from. In order to provide a seamless user experience for your application, Cognito uses Push Synchronization to push updates and synchronize user data across multiple devices. Cognito uses SNS to send a notification to all the devices associated with a given user identity whenever data stored in the cloud changes.

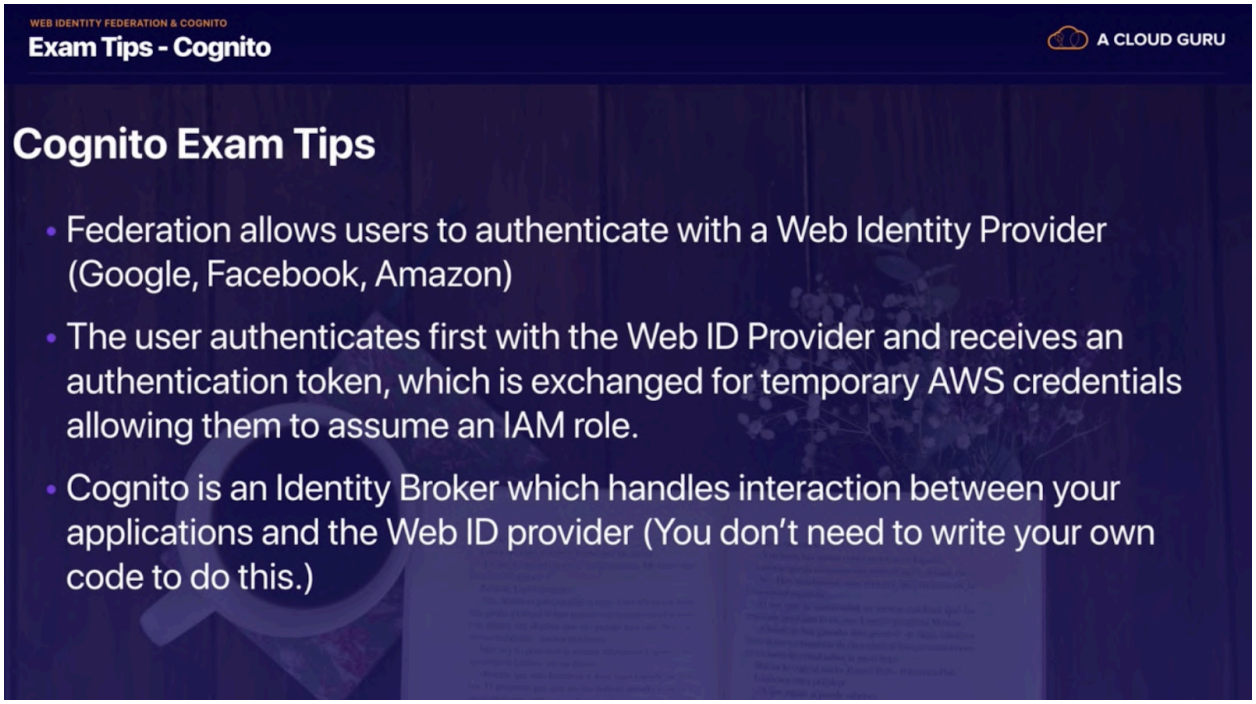


WEB IDENTITY FEDERATION & COGNITO

Push Synchronization Using SNS

A CLOUD GURU





Application Summary

SQS Exam Tips

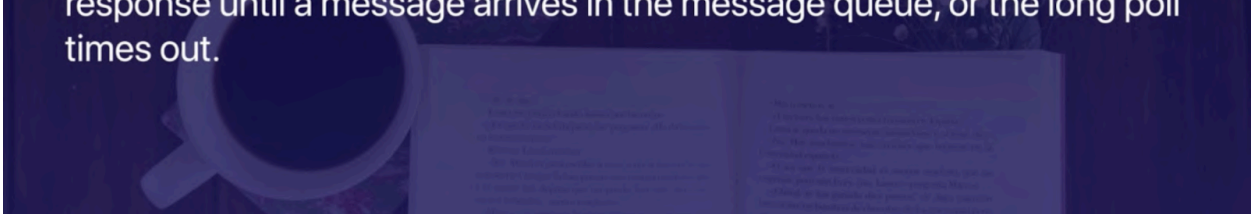
- SQS is a way to de-couple your infrastructure
- SQS is pull based, not pushed based.
- Messages are 256 KB in size.
- Messages can be kept in the queue from 1 minute to 14 days; the default retention period is 4 days.
- Standard SQS and FIFO SQS
- Standard order is not guaranteed and messages can be delivered more than once.
- FIFO order is strictly maintained and messages are delivered only once.

SQS Exam Tips

- Visibility Time Out is the amount of time that the message is invisible in the SQS queue after a reader picks up that message. Provided the job is processed before the visibility time out expires, the message will then be deleted from the queue. If the job is not processed within that time, the message will become visible again and another reader will process it. This could result in the same message being delivered twice.
- Visibility timeout maximum is 12 hours.

SQS Exam Tips

- SQS guarantees that your messages will be processed at least once.
- Amazon SQS long polling is a way to retrieve messages from your Amazon SQS queues. While the regular short polling returns immediately (even if the message queue being polled is empty), long polling doesn't return a response until a message arrives in the message queue, or the long poll times out.



SWF vs SQS

- SQS has a retention period of up to 14 days; with SWF, workflow executions can last up to 1 year.
- Amazon SWF presents a task-oriented API, whereas Amazon SQS offers a message-oriented API.
- Amazon SWF ensures that a task is assigned only once and is never duplicated. With Amazon SQS, you need to handle duplicated messages and may also need to ensure that a message is processed only once.
- Amazon SWF keeps track of all the tasks and events in an application. With Amazon SQS, you need to implement your own application-level tracking, especially if your application uses multiple queues.

SWF Actors - Exam Tips

- Workflow Starters — An application that can initiate (start) a workflow. Could be your e-commerce website following the placement of an order, or a mobile app searching for bus times.
- Deciders — Control the flow of activity tasks in a workflow execution. If something has finished (or failed) in a workflow, a Decider decides what to do next.
- Activity Workers — Carry out the activity tasks.

APPLICATIONS SUMMARY

Exam Tips - SNS

A CLOUD GURU

SNS Benefits

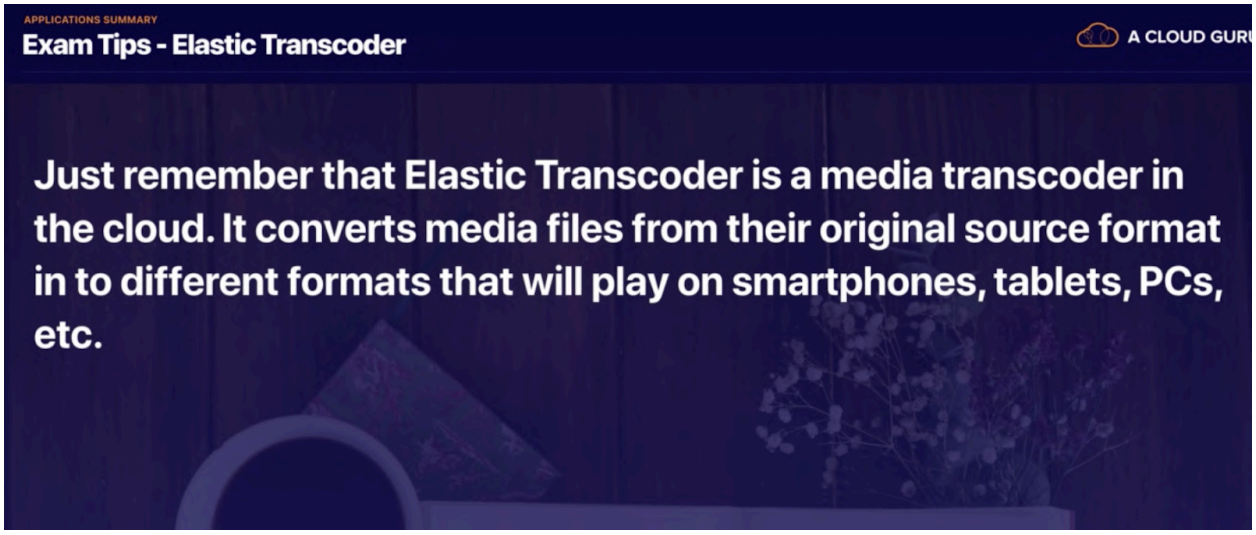
- Instantaneous, push-based delivery (no polling)
- Simple APIs and easy integration with applications
- Flexible message delivery over multiple transport protocols
- Inexpensive, pay-as-you-go model with no up-front costs
- Web-based AWS Management Console offers the simplicity of a point-and-click interface

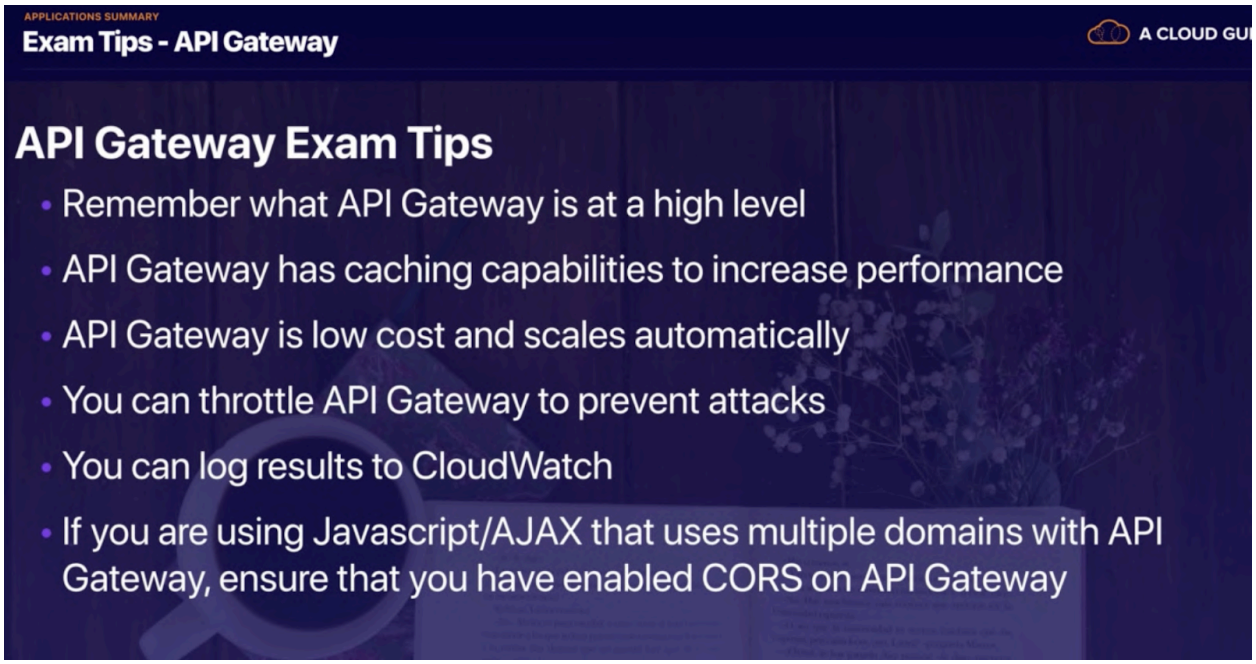
APPLICATIONS SUMMARY

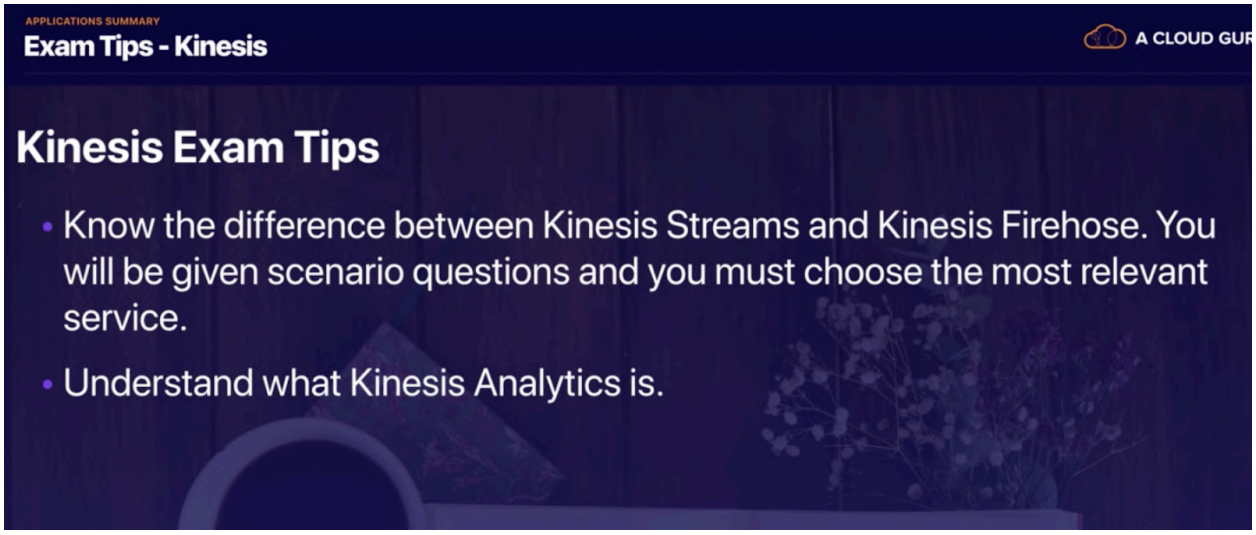
Exam Tips - SNS

SNS vs SQS?

- Both Messaging Services in AWS
- SNS - Push
- SQS - Polls (Pulls)


APPLICATIONS SUMMARY
Exam Tips - Elastic Transcoder
A CLOUD GUR
Just remember that Elastic Transcoder is a media transcoder in the cloud. It converts media files from their original source format in to different formats that will play on smartphones, tablets, PCs, etc.


APPLICATIONS SUMMARY
Exam Tips - API Gateway
A CLOUD GU
API Gateway Exam Tips
• Remember what API Gateway is at a high level
• API Gateway has caching capabilities to increase performance
• API Gateway is low cost and scales automatically
• You can throttle API Gateway to prevent attacks
• You can log results to CloudWatch
• If you are using Javascript/AJAX that uses multiple domains with API Gateway, ensure that you have enabled CORS on API Gateway


APPLICATIONS SUMMARY
Exam Tips - Kinesis
A CLOUD GUR
Kinesis Exam Tips
• Know the difference between Kinesis Streams and Kinesis Firehose. You will be given scenario questions and you must choose the most relevant service.
• Understand what Kinesis Analytics is.

Cognito Exam Tips

- Federation allows users to authenticate with a Web Identity Provider (Google, Facebook, Amazon)
- The user authenticates first with the Web ID Provider and receives an authentication token, which is exchanged for temporary AWS credentials allowing them to assume an IAM role.
- Cognito is an Identity Broker which handles interaction between your applications and the Web ID provider (You don't need to write your own code to do this.)

Cognito Exam Tips

- User pool is user based. It handles things like user registration, authentication, and account recovery.
- Identity pools authorise access to your AWS resources.