1. **How many number of ec2 instances can be attached to a multi-attach EBS volume?**

   Ans-> 16 EC2 instances at a time.

2. **Why use a Load Balancer?**

   Ans->
   - Spread load across multiple downstream instances
   - Expose a single point of access (DNS) to your application
   - Seamlessly handle failures of downstream instances
   - Do regular health checks to your instances
   - Provide SSL termination (HTTPS) for your websites
   - Enforce stickiness with cookies
   - High availability across zones
   - Separate public traffic from private traffic

3. **Latency difference between NLB and ALB**

   Ans-> 100ms for NLB and 400ms for ALB

4. **when application needs to be accessed only with few specific static IPs**

   Ans-> think of NLB

5. **What is the use of X-Forwarded-For header, X-Forwarded-Port header and X-Forwarded-Proto header?**

   Ans->
   - ***X-Forwarded-For header*** - this is used to get the IP address of clients connected to your website.
   - ***X-Forwarded-Port header*** - this is used to get the port number of clients connected to your website.
   - ***X-Forwarded-Proto header*** - this is used to get the protocol of clients connected to your website.

6. **Health check supports which protocols?**

   Ans-> Health checks support the TCP, HTTP and HTTPS Protocols.

7. **Which protocols does Application Load Balancer supports?**

   Ans-> HTTP, HTTPS and WebSocket

8. **NLB supports which health checks?**

   Ans-> NLB supports HTTP health checks as well as TCP and HTTPS

9. **Which load balancer uses GENEVE protocol on port 6081?**

Ans-> Gateway Load Balancer

10. **What are the different types of cookies available in Sticky sessions?**

Ans-> Application-based Cookies & Duration-based Cookies

11. **While creating custom application-based cookie in your Application Load Balancer. Which cookie name cannot we use?**

Ans-> The following cookie names are reserved by the ELB (**AWSALB**, **AWSALBAPP**, **AWSALBTG**),
so, we cannot use these cookie names.

12. **Explain about Cross-Zone Load Balancing pricing and enabling.**

Ans->

- **ALB**-> enable by default and no charges for inter AZ data
- **NLB & GWLB** -> disable by default and you pay extra charges ($) for inter AZ data if enabled.
- **CLB** -> disable by default and no charges for inter AZ data if enabled.

13. **The Load Balancer uses which SSL/TLS certificate?**

Ans-> It uses an X.509 SSL/TLS certificate.

14. **Which AWS service is used to manage certificates?**

Ans-> ACM (AWS Certificate Manager)

15. **What is SNI?**

Ans-> SNI (Server Name Indication) solves the problem of loading multiple SSL certificates onto one web server (to serve multiple websites)

It's a "newer" protocol and requires the client to indicate the hostname of the target server in the initial SSL handshake.

The server will then find the correct certificate, of return the default one.

Note--> It's only work for ALB and NLB (newer generation), CloudFront.

It doesn't work with CLB (older generation).

16. For Disaster recovery, can we set up Read Replicas as Multi AZ?

Ans-> Yes, we can.

Note: Read Replicas are ASYNC replication and Multi AZs are SYNC replication.

17. How do we make a RDS database from Single-AZ to Multi-AZ?

Ans->

- It's zero-downtime operation (no need to stop the DB).
- Just click on "modify" for the database and enable Multi-AZ.

18. What are the things that can we do while using RDS Custom for Oracle and Microsoft SQL Server?

Ans-> In RDS Custom we have access to the underlying database and OS so we can do the followings:

- Configure settings
- Install patches
- Enable native features
- Access the underlying EC2 Instance using SSH or SSM Session Manager

Note:-> De-activate Automation Mode to perform your customization, better to take a DB snapshot before.

19. In which amount Aurora storage grows automatically?

Ans-> Aurora storage automatically grows in increments of 10GB, up to 128TB.

a.

20. How many replicas can Aurora have?

Ans-> Aurora can have up to 15 replicas and the replication process is faster than MySQL (sub 10ms replica lag).

21. How much Aurora is costlier than compared to RDS?

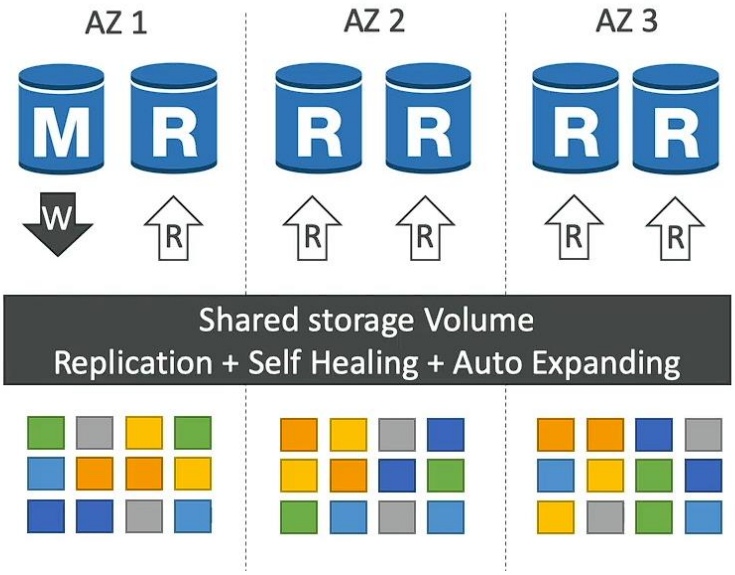
Ans-> Aurora costs 20% more than RDS but is more efficient.

22. Explain Aurora High Availability and Read Scaling.

Ans->

- 6 copies of your data across 3 AZ:
  - 4 copies out of 6 needed for writes
  - 3 copies out of 6 needs for reads
  - Self-healing with peer-to-peer replication

    - Storage is striped across 100s of volumes
- One Aurora Instance takes writes (master)
- Automated failover for master in less than 30 seconds
- Master + up to 15 Aurora Read Replicas serve reads
- Support for Cross Region Replication

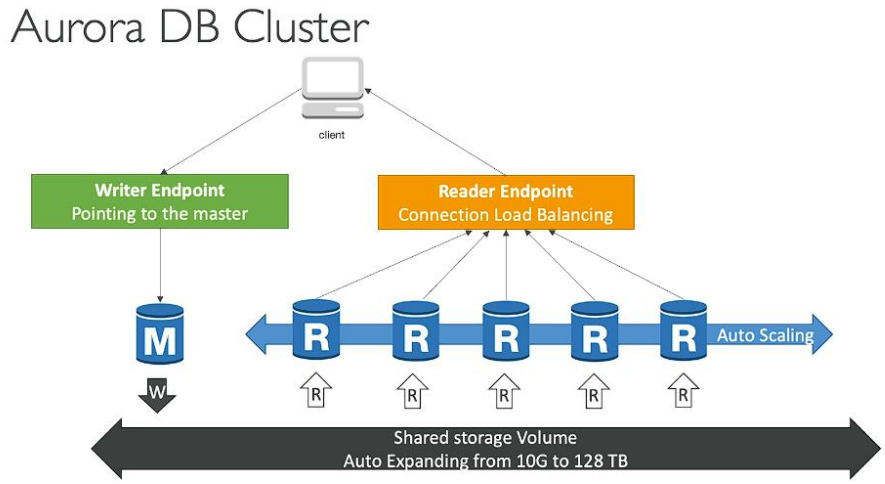


23. What is Writer Endpoint in Aurora DB Cluster?

Ans-> The Writer Endpoint is exclusively dedicated to writing operations like INSERT, UPDATE, or DELETE queries. It uniquely points to the current read-write primary instance within the cluster. The singular writer endpoint ensures that all write operations are directed to the primary instance, maintaining data consistency in the database.

24. What is Reader Endpoint in Aurora DB Cluster?

Ans-> The **Reader Endpoint** designed for read operations, such as SELECT queries, it dynamically distributes read traffic across up to 15 read replicas in an Aurora DB Cluster. This feature enhances the scalability and responsiveness of the database, particularly useful for scenarios with read-heavy workloads. By utilizing the reader endpoint, users can leverage multiple replicas to efficiently handle read requests and improve overall performance.



25. What are the main features of Aurora?

Ans-> The main features of Aurora are as follows:

- Automatic fail-over
- Backup and recovery
- Isolation and security
- Industry compliance
- Push-button scaling
- Automated Patching with Zero Downtime
- Advanced Monitoring
- Routine Maintenance
- Backtrack: restore data at any point of time without using backups.

26. What is Aurora Custom Endpoints?

Ans-> **Custom Endpoints** allow you to create custom read and write endpoints for your Aurora database clusters. These endpoints can be configured to route traffic to specific instances within the cluster, providing flexibility in directing queries to meet specific performance or business requirements. For example, when we want to run analytical queries on specific replicas. The Reader Endpoint is generally not used after defining Custom Endpoints

27. What is Aurora Serverless?

Ans->

- Aurora Serverless is designed to automatically adjust database capacity based on actual usage.

- It is a cost-effective (pay per second) option for infrequent or unpredictable workloads, as it scales up or down in response to demand.
- No capacity planning is needed here.

28. What is Global Aurora?

Ans-> **Global Aurora** is used to create read replicas in multiple AWS regions for an Aurora database cluster. By doing so, it provides low-latency global read access, improves disaster recovery capabilities, and allows efficient distribution of read traffic across different geographical locations. This feature is particularly beneficial for applications with a global user base, offering both performance optimization and enhanced resilience.

Aurora Cross Region Read Replicas:

- Useful for disaster recovery.
- Simple to put in place.

Aurora Global Database (Recommended):

- 1 primary region (read/write).
- Up to 5 secondary (read-only) regions, replication lag is less than 1 second.
- Up to 16 Read Replicas per secondary region.
- Promoting another region (for disaster recovery) has an RTO (Recovery Time Objective) < 1 minute.
- Typical cross-region replication takes less than 1 second.

29. Which type of Aurora takes less than 1 second for cross-region replication?

Ans-> Global Aurora.

30. What is Aurora Machine Learning?

Ans-> It Enables you to add ML-based predictions to your application via SQL. It is simple, optimized, and secure integration between Aurora and AWS ML services.

Supported services:

- Amazon SageMaker (use with any ML model)
- Amazon Comprehend (for sentiment analysis)

You don't need to have ML experience

Use Cases: fraud detection, ads targeting, sentiment analysis, product recommendations.

31. Is there any trick to reduce cost while using RDS database?

Ans-> To reduce cost, while using RDS database, first you can create a RDS database and then after using it for 2hrs, take a snapshot of it and after that delete the actual RDS

database and start using the snapshot instead, cause its way cheaper than RDS, and whenever you feel like to create database then just create the database through the snapshot --> this is a trick to reduce cost while using RDS.

32. How does RDS backup works?

Ans->

Automated backups:

- Daily full backup of the database (during the backup window)
- Transaction logs are backed-up by RDS every 5 minutes
- => ability to restore to any point in time (from oldest backup to 5 minutes ago)
- 1 to 35 days of retention, set 0 to disable automated backups.

Manual DB Snapshots:

- Manually triggered by the use
- Retention of backup for as long as you want

33. How does Aurora backup works?

Ans->

Automated backups:

- 1 to 35 days (cannot be disabled)
- point-in-time recovery in that timeframe

Manual DB Snapshots:

- Manually triggered by the user
- Retention of backup for as long as you want

34. How does RDS & Aurora Restore works?

Ans-> Restoring a RDS/Aurora backup from a snapshot, it will create a new database

Restoring MySQL RDS database from S3:

- Create a backup of your on-premises database
- Store it on Amazon s3 (object storage)
- Restore the backup file onto a new RDS instance running MySQL

Restoring MySQL Aurora cluster from S3:

- Create a backup of your on-premises database using Percona XtraBackup
- Store the backup file on Amazon S3
- Restore the backup file onto a new Aurora cluster running MySQL

35. Which protocol does Aurora Database Cloning uses?

Ans-> Aurora Database Cloning uses ***copy-on-write protocol***.

copy-on-write protocol: it is a mechanism used in database cloning where the cloned database initially shares the same underlying storage as the source database. When either the source or the clone modifies a data block, a new copy of that block is created and modified, leaving the original block unchanged. This approach allows for rapid and efficient creation of database clones, with minimal additional storage costs and performance impact, as only modified data blocks are duplicated.

36. Explain about RDS & Aurora Security.

Ans->

At-rest encryption:

- Database master & replicas encryption using AWS KMS - must be defined at launch time
- if the master is not encrypted, the read replicas cannot be encrypted
- To encrypt an un-encrypted database, go through a DB snapshot & restore as encrypted

In-flight encryption: TLS-ready by default, use the AWS TLS root certificates client-side

IAM Authentication: IAM roles to connect to your database (instead of username/pw)

Security Groups: Control Network access to your RDS/Aurora DB

No SSH available except on RDS custom

Audit Logs can be enabled and sent to CloudWatch log for longer retention

37. What is RDS Proxy?

Ans-> RDS Proxy is an AWS service that acts as a intermediary layer between your application and a relational database (like Amazon RDS or Aurora). It helps applications efficiently manage database connections, improving scalability and availability. This can enhance performance by reducing the load on the database and providing connection pooling capabilities.

Some key points about RDS Proxy:

- Its fully managed database proxy for RDS
- It improves database efficiency by reducing the stress on database resources (e.g., CPU, RAM) and minimize open connection (and timeouts)
- Its serverless, autoscaling, highly available (multi-AZ)

- It reduces RDS & Aurora failover time by up to 66%
- Supports RDS (MySQL, PostgreSQL, MariaDB, MS SQL Server) and Aurora (MySQL, PostgreSQL)
- No code changes required for most apps
- Enforce IAM Authentication for DB, and securely store credentials in AWS Secrets Manager
- RDS Proxy is never publicly accessible (must be accessed from VPC)

38. What are the key points of ElastiCache?

Ans->

- Its a managed in-memory caching service
- It is to get managed Redis or Memcached
- Helps reduce load off of databases for read intensive workloads
- it helps make your application stateless
- Automatic failover for Redis clusters
- Scalability with easy addition/removal of cache nodes
- AWS takes care of OS maintenance/patching, optimizations, setup, configuration, monitoring, failure recovery and backups.

39. What are the steps for User Session Store in ElastiCache Solution Architecture?

Ans->

- User logs into any of the applications among multiple applications
- The application writes the session data into ElastiCache
- The user hits another instance of application
- The instance retrieves the data, and the user is already logged in

40. What are the key differences between Redis and Memcached?

Ans->

Redis: supports complex data structures
Memcached: primarily handles simple key-value pairs

Redis: Have backup and restore features
Memcached: no backup and restore

Redis: offers optional data persistence
Memcached: does not have built-in persistence

Redis: Well-suited for advanced data structures, sorting, and range queries
Memcached: best for simple key-value storage and retrieval

41. Explain 3 main types of patterns for ElastiCache.

Ans->**Lazy-Load Pattern:**

Fetches data from the cache only when needed, minimizing unnecessary database calls and improving response time. This is particularly useful for scenarios where not all data needs to be preloaded into the cache.

Write-Through Pattern:

Involves writing data to both the cache and the database simultaneously. This ensures that the cache remains up to date with the latest database changes. This pattern is beneficial when the balance between read and write operations is significant.

Session Store Pattern:

The cache is utilized to store and manage user session information. This includes data such as user preferences, authentication tokens, and other session-related details.

42. Mention some important ports to remember.

Ans->

Important ports:

- FTP: 21
- SSH: 22
- SFTP: 22 (same as SSH)
- HTTP: 80
- HTTPS: 443

vs RDS Databases ports:

- PostgreSQL: 5432
- MySQL: 3306
- Oracle RDS: 1521
- MSSQL Server: 1433
- MariaDB: 3306 (same as MySQL)
- Aurora: 5432 (if PostgreSQL compatible) or 3306 (if MySQL compatible)

43. Explain about DNS Terminologies.

Ans ->

- Domain Registrar: Amazon Route 53, GoDaddy, etc.
- DNS Records: A, AAAA, CNAME, NS, etc.
- Zone File: Contains DNS records

- Name Server: Resolves DNS queries (Authoritative or Non_Authoritative)
- Top Level Domain (TLD): .com, .us, .in, .gov, .org, etc.
- Second Level Domain (SLD): amazon.com, google.com, etc.

44. What are the contents of a record of Route 53?

Ans->

- Domain/subdomain Name - e.g., example.com
- Record Type - e.g., A or AAAA
- Value - e.g., 12.34.56.78
- Routing Policy - How Route 53 responds to queries
- TTL (Time To Live) - amount of time the record cached at DNS Resolvers

45. What are the record types do Route 53 supports?

Ans->

- A -> Maps a hostname to IPV4
- AAAA -> Maps a hostname to IPV6
- CNAME -> Maps a hostname to another hostname

  The target is a domain name which must have an A or AAAA record

  Can't create a CNAME for record for the top node of a DNS namespace (Zone Apex)

  Example: you can't create CNAME for example.com, but you can create CNAME for www.example.com

- NS -> Name server for Hosted Zone

  Control how traffic is routed for a domain

- (advanced) CAA/DS/MX/NAPTR/PTR/SOA/TXT/SPF/SRV

46. What is Route 53 Hosted Zones?

Ans-> Hosted Zone is a container for records that define how to route traffic to a domain and its subdomains. There are 2 types of Hosted Zones:

Public Hosted Zones -> Contains records that specify how to route traffic on the internet (public domain names). Example: application1.mypublicdomain.com

Private Hosted Zones -> Contains records that specify how you route traffic within one or more VPCs (private domain names). Example: application1.company. Internal

Note: In AWS you pay $0.50 per month per hosted zone.

47. What is TTL in the context of Route 53?

Ans-> **TTL (Time To Live)** in AWS Route 53 specifies the duration for which DNS resolvers and caches can store a DNS record before it expires and needs to be refreshed.

- When a DNS resolver queries Route 53 for a particular DNS record, it receives both associated with that record. The resolver then caches this information locally for the duration specified by the TTL.
- During this time, subsequent queries for the same DNS record from clients or other resolvers can be answered directly from the cache without needing to query Route 53 again. This helps reduce the load on Route 53 and speeds up DNS resolution for clients.
- Once the TTL expires, the resolver discards the cached record and must query Route 53 again to get the latest record data. Route 53 then returns the updated record along with a new TTL value, and the caching process begins anew.
- The TTL value is set in seconds and can be configured for each individual DNS record in Route 53.

48. What are the main differences between CNAME and Alias?

Ans->

CNAME:-

- Points a hostname to any other hostname. (app.mydomain.com => example.anything.com)
- Only for NON-ROOT DOMAIN (aka.something.mydomain.com)

Alias:-

- Points a hostname to an AWS Resource (app.mydomain.com => examle.amazonaws.com)
- Works for ROOT DOMAIN and NON-ROOT DOMAIN (aka.mydomain.com)
- Free of charge
- Native health check

49. What are Alias Records?

Ans->

- Maps a hostname to an AWS resource
- An extension to DNS functionality
- Automatically recognizes changes in the resource's IP addresses
- Unlike CNAME, it can be used for the top node of a DNS namespace (Zone Apex),
- e.g.: example.com
- Alias Record is always of type A/AAAA for AWS resources (IPv4/IPv6)

- You can't set the TTL, its set automatically by Route 53

50. What are the Alias Records Targets?

Ans->

- Elastic Load Balancers
- CloudFront Distributions
- API Gateway
- Elastic Beanstalk environments
- S3 Websites
- VPC Interface Endpoints
- Global Accelerator
- Route 53 record in the same hosted zone
- You cannot set an ALIAS record for an EC2 DNS name

51. What is the Route 53 Routing Policies?

Ans->

1. Simple Routing Policy
2. Weighted Routing Policy
3. Latency based Routing Policy
4. Failover Routing Policy
5. Geolocation Routing Policy
6. IP-based Routing Policy
7. Multi-Value Answer Routing Policy
8. Geoproximity (using Route 53 Traffic Flow feature) Routing Policy

52. Key points about Route 53 - Simple Routing Policy.

Ans->

- Typically, route traffic to a single resource
- Can specify multiple values in the same record
- If multiple values are returned a random one is chosen by the client
- When Alias enabled, specify only one AWS resource
- Can't be associated with Health Checks

53. Key points about Route 53 - Weighted Routing Policy.

Ans->

- Control the % of the request that go to each specific resource
    - Assign each record a relative weight:
    - traffic (%) = Weight for a specific record/Sum of all the weights for all records
- Weights don't need to sum up to 100
- DNS records must have the same name and type
- Can be associated with Health Checks
- Use cases: load balancing between regions, testing new application versions etc.
- Assign a weight of 0 to a record to stop sending traffic to a resource
- If all records have weight of 0, then all records will be returned equally

54. Key points about Route 53 - Latency-based Routing Policy.

Ans->

- Redirect to the resource that has the least latency close to us
- Latency is based on traffic between users and AWS Regions
- Can be associated with Health Checks (has a failover capability)
- Configurable TTL for DNS caching
- Ideal for globally distributed applications
- Provides monitoring and metrics for performance tracking
- Automatically adjusts routing for unhealthy regions

55. Key points about Route 53 - Failover Routing Policy.

Ans->

- Failover routing is like having a backup plan for your website or app.
- You set up a main server/instance (primary) and a backup server/instance (secondary).
- It utilizes DNS records for primary and secondary resources.
- Responds to DNS queries with secondary resource IP during failover.
- Commonly used for disaster recovery and high availability

56. Key points about Route 53 - Geolocation Routing Policy.

Ans->

- Geolocation routing is like a GPS for directing internet traffic.
- Routes traffic based on user's geographic location.
- Different from Latency-based!
- Customizes responses with multiple sets of DNS records.
- Utilizes Route 53's global network for precise targeting worldwide.
- Ideal for services with regional content or compliance requirements.
- Includes fallback mechanism for unspecified locations.

- Use Cases: website localization, restrict content distribution, load balancing, etc.
- Can be associated with Health Checks.

57. Key points about Route 53 - IP-based Routing Policy.

Ans->

- Ip-based routing is like sorting mail based on postal codes.
- It directs users to different places on the internet based on their IP addresses.
- You can customize where users go depending on their IP address.
- You provide a list of CIDRs for your users and the corresponding endpoints/location (user-IP-to-endpoint mapping)
- Use cases: Optimize performance, reduce network costs
- Example: route end users from a particular ISP to a specific endpoint

58. Key points about Route 53 - Multi-Value Routing Policy.

Ans->

- Multi-Value Routing Policy is like having multiple doors open for visitors to enter a building.
- It's good for making sure there's always a way in, even if one door is closed.
- Associates Multiple values with a single DNS record.
- Returns multiple values in DNS responses,
- DNS resolver randomly selects one of the values for use.
- Provides basic load balancing across multiple resources.
- Can be associated with Health Checks (return only value for healthy resources).
- Up to 8 healthy records are returned for each Multi-Value query.

  Note: Multi-Value is not a substitute for having an ELB.

59. Key points about Route 53 - GeoProximity Routing Policy.

Ans->

- Route traffic to your resources based on the geographic location of users and resources.
- Ability to shift more traffic to resources based on the defined bias.
- To change the size of the geographic region, specify bias values:
  - To expand (1 to 99) -> more traffic to the resource
  - To shrink (-1 to -99) -> less traffic to the resource

Resources can be:

- AWS resources (specify AWS region)
- Non-AWS resources (specify Latitude and Longitude)

Note: You must use Route 53 Traffic Flow (advanced) to use this feature

60. Key points about Route 53 - Health Checks.

Ans->

- HTTP Health Checks are only for public resources
- Health Check => Automated DNS Failover, and Health checks are total 3 types:
  - ***Health check that monitors an endpoint*** (application, server, other AWS resource).
  - ***Health Checks that monitor other health checks*** (Calculated Health Checks).
  - ***Health checks that monitor CloudWatch Alarms*** (full control !!) - e.g., throttles of DynamoDB, alarms on RDS, custom metrics, etc. (helpful for private resources).
- Health Checks are integrated with CloudWatch metrics.
- Supports various check types (HTTP, HTTPS, TCP, ICMP).
- Geo-distributed checks for global perspective.

61. Key points about the Health Checks that Monitor an Endpoint

Ans->

About 15 global health checkers will check the endpoint health

- Healthy/Unhealthy Threshold - 3 (default)
- Interval - 30 sec (can set to 10 sec - higher cost)
- Supported protocols: HTTP, HTTPS and TCP
- If > 18% of health checkers report the endpoint is healthy, Route 53 considers it Healthy. Otherwise, It's Unhealthy
- Ability to choose which locations you want Route 53 to use

Health Checks pass only when the endpoint responds with the 2xx and 3xx status codes.

Health Checks can be setup to pass/fail based on the text in the first 5120 bytes of the response.

Configure your router/firewall to allow incoming requests from Route 53 Health Checkers.

62. Key points about the Route 53 - Calculated Health Checks

Ans->

- Combine the results of multiple Health Checks (Childs) into a single Health Check (Parent)
- You can use OR, AND, or NOT to combine them
- Can monitor up to 256 Child Health Checks
- Specify how many of the Health Checks need to pass to make the parent pass
- Usage: perform maintenance to your website without causing all health checks to fail

63. How to do Health Checks for Private Hosted Zones?

Ans -> Route 53 health checkers are outside the VPC. So, they can't access private endpoints (private VPC or on-premises resource).

So, in this case, to do Health Check for a Private Hosted Zone, we can create a CloudWatch Metric and associate a CloudWatch Alarm, then create a Health Check that checks the alarm itself.

64. What is Golden AMI?

Ans -> A Golden AMI (Amazon Machine Image) is a pre-set, secure, and optimized image with the operating system, necessary software, and configurations for deploying applications. It ensures that all instances are consistent, secure, and ready to use, making deployments faster and easier. This helps organizations maintain uniformity, enhance security, and reduce the time and effort needed to manage AWS infrastructure.

65. What is Bootstrap using User Data?

Ans -> Bootstrapping using user data in AWS allows you to automatically run scripts or commands when an EC2 instance launches. This script, specified in the "user data" field, can install software, configure settings, and perform initial setup tasks. For example, a user data script for a Linux instance to install and start Apache web server might look like this:

```bash
#!/bin/bash
yum update -y
yum install -y httpd
systemctl start httpd
systemctl enable httpd
```

66. Elastic Beanstalk Overview.

Ans -> AWS Elastic Beanstalk is a Platform as a Service (PaaS) that simplifies the process of deploying, managing, and scaling web applications and services. It supports several programming languages, frameworks and platforms, such as:

- Go
- Java SE
- Java with Tomcat
- .NET Core on Linux
- .NET on Windows Server
- Node.js
- PHP
- Python
- Ruby
- Packer Builder
- Single Container Docker
- Multi-Container Docker
- Preconfigured Docker

Elastic Beanstalk automatically handles the underlying infrastructure provisioning, load balancing, auto-scaling, and monitoring, allowing developers to focus solely on writing code.

Note: Beanstalk is free, but you pay for the underlying instances.

67. What are the Components of Elastic Beanstalk?

Ans ->

Application

An Elastic Beanstalk application is a logical collection of Elastic Beanstalk components, including environments, versions, and configurations. It serves as a container for environments and application versions.

Application Version

An application version refers to a specific, labeled iteration of deployable code for an application. Each version is stored in Amazon S3 and is referenced when deploying to an environment.

Environment

An environment is a collection of AWS resources running an application version. Each environment runs only one application version at a time, but you can run multiple environments, such as **dev**, **test**, and **prod**, for the same application. The environment tier determines the type of application Elastic Beanstalk will run. The two main tiers are:

- **Web Server Environment**: For applications (web applications) that process/handle and respond to HTTP(S) requests.
- **Worker Environment**: For applications that processes/handle background jobs and tasks.

Environment Configuration: Parameters defining environment behavior.

Configuration Files (.ebextensions): Customize environment via YAML or JSON files.

Instances: Hosts application using Amazon EC2.

Load Balancer: Distributes traffic across EC2 instances.

Auto Scaling Group: Adjusts EC2 instances based on demand.

RDS: Optional relational database service.

Elastic Load Balancing (ELB): Distributes traffic to ensure load balance.

Amazon S3: Stores application versions and logs.

Amazon CloudWatch: Provides monitoring and logging.

Elastic Beanstalk CLI (EB CLI): Command-line interface for management.

Summary Diagram:

68. What is the difference between 2 environments: Web Server Tier & Worker Tier?

Ans ->

Web Server Tier: Designed for applications that process and respond to HTTP(S) requests.
Worker Tier: Intended for applications that handle background tasks or asynchronous processing.

Web Server Tier: Ideal for web applications, APIs, or any service that handles user-facing traffic.
Worker Tier: Useful for processing jobs such as data processing, image resizing, or batch processing.

Web Server Tier: Automatically provisions resources like load balancers and auto-scaling groups to manage incoming web traffic.
Worker Tier: Typically decoupled from user interaction and often perform tasks independently.

Web Server Tier: Ensures high availability and scalability by distributing traffic across multiple EC2 instances.
Worker Tier: Scales independently from the web tier to optimize resource allocation for background tasks.

69. Key points about Amazon S3 Buckets.

Ans ->

- Amazon S3 allows people to store object (files) in "buckets" (directories)
- Buckets must have a globally unique name (across all regions all accounts)
- Naming Convention:
    - No uppercase, No underscore
    - 3-63 characters long
    - Not an IP
    - Must start with lowercase letter or number
    - Must NOT start with the prefix xn--
    - Must NOT end with the suffix -s3alias
- Buckets are defined at the region level
- S3 looks like a global service but buckets are created in a region.

70. What are the contents of an Amazon S3 Object?

Ans ->

- Objects (files) have a Key.
- The key is composed of prefix + object name
- e.g.- s3://my-bucket/my_folder/my_file.txt [prefix -> "s3://my-bucket/my_folder/", object name -> my_file.txt]
- Max object size is 5TB.
- If uploading object size is more than 5GB, then AWS recommends "multi-part uploads".
- Metadata (list of text key/value pairs - system or user metadata).
- Tags (Unicode key/value pair - up to 10) - useful for security/lifecycle
- Version ID (if versioning is enabled).

71. Explain about Amazon S3 - Security.

Ans ->

User-Based Security:

- ***IAM Policies:***
    - User Policies: Permissions for individual IAM users.
    - Group Policies: Permissions for groups of IAM users.
    - Role Policies: Permissions for IAM roles assumed by users or services.

Resource-Based Security:

- ***Bucket Policies:***

    - JSON-based policies attached to S3 buckets.
- ***Access Control Lists (ACLs):***
    - Bucket ACLs: Define basic read/write permissions for specific AWS accounts or predefined groups at the bucket level.
    - Object ACLs: These are the same as Bucket ACLs but less common.

- ***Bucket and Object-Level Permissions:***
    - Bucket-Level: Control access to operations on the bucket itself (e.g., listing objects, configuring bucket settings).
    - Object-Level: Control access to operations on individual objects (e.g., uploading, downloading, deleting).

Note: An IAM principal can access an S3 object if:

- The user IAM permissions ALLOW it OR the resource policy ALLOWS it
- And there's no explicit DENY

Encryption: encryption objects in Amazon S3 using encryption keys.

72. Explain about S3 Bucket Policy with an example.

Ans -> Bucket policies in Amazon S3 are JSON-based documents attached directly to S3 buckets. They define who can access the resources within the bucket and what actions they can perform on those resources. Here's an explanation with an example:

Example:

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": "*",
      "Action": "s3:GetObject",
      "Resource": "arn:aws:s3:::example-bucket/*"
    }
  ]
}
```

- **Version:** Specifies the version of the policy language being used (always set to "2012-10-17").
- **Statement:** Contains one or more statements defining permissions.
- **Effect:** Indicates whether the statement allows or denies access ("Allow" in this case).

- **Principal:** Specifies the entity to which the policy applies. In this example, it applies to all users ("*").
- **Action:** Defines the actions allowed (e.g., "s3" To allow reading objects).
- **Resource:** Specifies the resources to which the policy applies. Here, it applies to all objects within the "example-bucket" bucket.

Use Cases:

- Granting public read access to objects for website hosting.
- Restricting access to specific IP addresses or AWS accounts.
- Enforcing encryption requirements for objects stored in the bucket.

73. Key points about Amazon S3 - Static Website Hosting.

Ans ->

- S3 can host static websites and have them accessible on the internet.
- The website URL will be (depending on the region):
  - http://bucket-name.s3-website-aws-region.amazonaws.com
    OR
  - http://bucket-name.s3-website.aws-region.amazonaws.com
- If you get a 403 Forbidden error, make sure the bucket policy allows public reads!

74. Key points about Amazon S3 - Versioning.

Ans ->

Versioning Control and Management: Versioning can be enabled or suspended on a bucket. Once versioning is enabled, it cannot be completely disabled, only suspended, ensuring a continuous record of object versions.

Object Versioning: Each object gets a unique version ID upon modification, allowing multiple versions of an object to coexist and be referenced for retrieval and operations.

Data Protection: Protects against accidental overwrites and deletions, with the ability to restore deleted objects using their version IDs.

Note:

- Any file that is not versioned prior to enabling versioning will have version "null".
- Suspending versioning does not delete the previous versions.

75. Key points about Amazon S3 - Replication (CRR & SRR).

Ans ->

1. Amazon S3 supports Cross-Region Replication (CRR) and Same-Region Replication (SRR) to automatically replicate objects for enhanced data availability and disaster recovery.
2. Enabling versioning on both source and destination buckets is required for replication.
3. Replication is asynchronous, so there may be a delay, but it ensures high data durability and availability.
4. Costs include data transfer fees for CRR and storage costs for replicated objects, with SRR incurring only storage costs.
5. Permissions for Amazon S3 replication are managed using IAM policies and bucket policies to control access and actions.

Use Cases:

- CRR: Ideal for geographically distributed applications, lower latency access, compliance, data sovereignty, and disaster recovery.
- SRR: Useful for maintaining compliance with data residency requirements, log aggregation, and for live replication between production and test accounts.

76. Amazon S3 - Replication (Notes).

Ans ->

- After you enable replication, only new objects are replicated
- Optionally, you can replicate existing objects using S3 Batch Replication, which replicates existing objects and objects that failed to replication.
- For DELETE operations
    - Can replicate delete markers from source to target (optional setting)
    - Deletion with a version ID is not replicated (to avoid malicious deletes)
- There is no "chaining" of replication
    - if bucket 1 has replication into bucket 2, which has replication into bucket 3
    - Then objects created in bucket 1 are not replicated to bucket 3

77. Explain about Amazon S3 - Durability and Availability.

Ans ->

Durability:

- High durability (99.999999999%, 11 9's) of objects across multiple AZ
- If you store 10,000,000 objects with amazon S3, you can on average expect to incur a loss of a single object once every 10,000 years
- This is applicable for all storage classes

Availability:

- Measures how readily available a service is
- Varies depending on storage class
- Example: S3 standard has 99.99% availability = not available 53 minutes a year

78. What are the different types Amazon S3 Storage Classes available?

Ans ->

1. Amazon S3 Standard - General Purpose
2. Amazon S3 Standard-Infrequent Access (IA)
3. Amazon S3 Intelligent Tiering
4. Amazon S3 One Zone-Infrequent Access
5. Amazon S3 Glacier Instant Retrieval
6. Amazon S3 Glacier Flexible Retrieval
7. Amazon S3 Glacier Deep Archive

79. Key points about Amazon S3 Standard - General Purpose

Ans ->

- 99.99% Availability
- Used for Frequently accessed data
- Low latency and high throughput
- Sustain 2 concurrent facility failures

Use Cases: Big Data analytics, mobile & gaming application, content distribution etc.

80. Key points about Amazon S3 Storage Classes - Infrequent Access

Ans ->

- For data that is less frequently accessed, but required rapid access when needed
- Lower cost than S3 Standard

Amazon S3 Standard-Infrequent Access (S3 Standard-IA)

- 99.9% Availability
- Use Cases: Disaster Recovery, Backups

Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA)

- High durability (99.999999999%) in a single AZ; data lost when AZ is destroyed
- 99.5% Availability
- Use Cases: Storing secondary backup copies of on-premises data, or data that you can recreate.

81. Key points about S3 Glacier Storage Classes

Ans ->

- Low-cost object storage meant for archiving/backup

- Pricing: Price for storage + object retrieval cost

Amazon S3 Glacier Instant Retrieval:

- Millisecond retrieval, great for data accessed once a quarter
- Minimum storage duration of 90 days (about 3 months)

Amazon S3 Glacier Flexible Retrieval (formerly Amazon S3 Glacier):

- Expedited (1 to 5 minutes), Standard (3 to 5 hours), Bulk (5 to 12 hours) - free
- Minimum storage duration of 90 days (about 3 months)

Amazon S3 Glacier Deep Archive - for long term storage:

- Standard (12 hours), Bulk (48 hours)
- Minimum storage duration of 180 days (about 6 months)

82. Key points about Amazon S3 Intelligent-Tiering.

Ans ->

- Small monthly monitoring and auto-tiering fee
- Moves objects automatically between Access Tiers based on usage
- There are no retrieval charges in S3 Intelligent-Tiering

- ***Frequent Access tier (automatic):*** default tier
- ***Infrequent Access tier (automatic):*** Objects not accessed for 30 days
- **Archive Instant Access tier (automatic):** Objects not accessed for 90 days
- ***Archive Access tier (optional):*** configurable from 90 days to 700+ days
- ***Deep Archive tier (optional):*** config. from 180 days to 700+ days

83. Explain about Amazon S3 - Lifecycle Rules.

Ans ->

Amazon S3 - Lifecycle Rules are JSON-based configurations to manage object transitions and expirations.

Rule Components:

Transition Action:

- Move objects to different storage classes after a set time

  Examples:

  - Move objects to Standard IA class 60 days after creation.
  - Moving to Glacier for archiving after 6 months.

Expiration Action:

- Permanently deletes objects after a set time.

  Examples:

  - Access logs can be set to delete after 365 days
  - Can be used to delete old version of files (if versioning is enabled).
  - Can be used to delete incomplete Multi-Part uploads.

Prefix/Tags:

- Apply rules to specific objects based on prefixes or tags.

  Examples:

  - Rules can be created for a certain prefix (e.g., S3://mybucket/mp3/*)
  - Rules can be created for certain object Tags (e.g., Department: Finance).

84. Amazon S3 - Lifecycle Rules (Scenario 1)

Your application on EC2 creates images thumbnails after profile photos are uploaded to Amazon S3. These thumbnails can be easily recreated, and only need to be kept for 60 days. The source images should be able to be immediately retrieved for these 60 days, and afterwards, the user can wait up to 6 hours. How would you design this?

Ans ->

- S3 source images can be on Standard, with a lifecycle configuration to transition them to Glacier after 60 days (about 2 months).
- S3 thumbnails can be on One-Zone IA, with a lifecycle configuration to expire them (delete them) after 60 days (about 2 months).

85. Amazon S3 - Lifecycle Rules (Scenario 2)

A rule in your company states that you should be able to recover your deleted S3 objects immediately for 30 days, although this may happen rarely. After this time, and for up to 365 days, deleted objects should be recoverable within 48 hours. How would you design this?

Ans ->

- Enable S3 versioning to have object versions, so that "deleted objects" are in fact hidden by a "delete marker" and can be recovered.
- Transition the "noncurrent versions" of the object to Standard IA

- Transition afterwards the "noncurrent versions" to Glacier Deep Archive

86. Key points about Amazon S3 Analytics - Storage Class Analysis.

Ans ->

- Analyzes storage access patterns.
- Recommends transitioning infrequently accessed data to lower-cost storage classes.
- Identifies data for transition to S3 Standard-IA, S3 intelligent-Tiering, and other supported storage classes.
- Does not directly support for One-Zone IA or Glacier
- Generates daily reports on data access.
- Can be applied to entire buckets or specific prefixes/tags.

87. Explain about Amazon S3 - Requester Pays.

Ans ->

Amazon S3's Requester Pays feature allows you to configure your S3 bucket so that the requester, not the bucket owner, covers data transfer and request costs. This is useful for sharing data without incurring access costs.

Feature Overview:

- Cost Transfer: Shifts data transfer and request costs to the requester.
- Data Sharing: Ideal for public data where the owner wants to avoid access costs.

Billing and Costs:

- Data Transfer: Requesters pay for data transfer.
- Request Costs: Requesters pay for GET, PUT, LIST, and other requests.
- Bucket Owner Costs: Owners remain responsible for storage costs.

Use Cases:

- Public Data Sets: Share large data sets without bearing access costs.
- Collaborative Projects: Multiple parties access shared data, covering their own costs.
- Third-Party Access: Third parties access data for analysis or processing.

88. Explain about S3 Event Notifications.

Ans ->

Amazon S3 Event Notifications allows you to receive notifications when certain events happen in your S3 bucket. This feature can be used to trigger workflows, alert systems, or other automated processes in response to changes in your bucket.

Feature Overview:

- Trigger Actions: Notify or trigger actions based on specific events in an S3 bucket.
- Events: Can be set up for object creation, deletion, restoration, replication, and more.

Supported Destinations:

- Amazon SNS: Send notifications to SNS topics.
- Amazon SQS: Send messages to SQS queues.
- AWS Lambda: Trigger Lambda functions to process events.

Event Types:

- s3:ObjectCreated: Triggers on object creation (e.g., Put, Post, Copy).
- s3:ObjectRemoved: Triggers on object deletion (e.g., Delete).
- s3:ObjectRestore: Triggers on object restoration from Glacier.
- s3:Replication: Triggers on replication events.

Use Cases:

- Data Processing: Automatically trigger data processing workflows with Lambda.
- Monitoring and Alerting: Set up alerts for specific changes in your S3 bucket.
- Automation: Automate tasks like metadata updates, log processing, replication or backup tasks.

89. Amazon S3 - Event Notifications with Amazon EventBridge.

Ans ->

Amazon S3 Event Notifications can be integrated with Amazon EventBridge to provide advanced event-driven architecture capabilities. This integration allows you to create more complex and customizable event handling workflows.

- **Advanced filtering** options with JSON rules (metadata, object size, name, etc.)
- **Multiple Destinations** - ex Step Functions, Kinesis Streams / Firehose...
- **EventBridge Capabilities** - Archive, Replay Events, Reliable delivery.

90. Explain about Amazon S3 - Baseline Performance.

Ans ->

- Amazon S3 automatically scales to high request rates, latency 100-200 ms
- Your application can achieve at least 3,500 PUT/COPY/POST/DELETE or 5,500 GET/HEAD request per second per prefix in a bucket.
- There are no limits to the number of prefixes in a bucket.

Example (object path => prefix):

- bucket/folder1/sub_folder1/file => /folder1/sub_folder1
- bucket/folder1/sub_folder2/file => /folder1/sub_folder2
- bucket/1/file => /1/
- bucket/2/file => /2/

Note: If you spread reads across all four prefixes evenly, you can achieve 22,000 requests per second for GET and HEAD.

91. Amazon S3 - Performance for uploading, transferring and downloading files.

Ans ->

Multi-Part upload:

- Recommended for files > 100MB,
- Must use for files > 5GB
- Can help parallel uploads (speed up transfers)

S3 Transfer Acceleration:

- Increase transfer speed by transferring file to an AWS edge location which will forward the data to the S3 bucket in the target region.
- Compatible with multi-part upload.

Byte-Range Fetches:

- Breaks down file downloads into smaller chunks called byte ranges.
- Each chunk is specified by its starting and ending byte positions.
- Allows for efficient downloading of large files by fetching only required segments.
- Utilizes the HTTP Range header in GET requests to specify the byte range.
- Ideal for scenarios like video streaming or downloading large files where retrieving the entire file at once may be impractical or unnecessary.

92. Explain about Amazon S3 Select and Glacier Select.

Ans ->

S3 Select:

- Let's you retrieve specific data from objects stored in Amazon S3.
- Allows querying data directly within S3 using SQL expressions.
- Reduces the amount of data transferred, improving query performance and cost-effectiveness.
- Ideal for applications needing to analyze or process specific data subsets without downloading entire objects.

Glacier Select:

- Similar to S3 Select but works on data stored in Amazon Glacier.
- Enables querying archived data directly without needing to restore the entire archive.
- Saves time and costs by selectively retrieving only the data needed for analysis or processing.
- Useful for extracting insights or performing analytics on large datasets stored in Glacier archives.

93. What are the things that we can do with Amazon S3 - Batch Operations?

Ans ->

- Perform bulk operations on existing S3 objects with a single request, examples:
  - Modify object metadata & properties
  - Copy object between S3 buckets
  - Encrypt un-encrypted objects
  - Modify ACLs, tags
  - Restore objects from S3 Glacier
  - Invoke Lambda function to perform custom action on each object.
- A job consists of a list of objects, the action to perform, and optional parameters
- S3 Batch Operations manages retries, tracks progress, sends completion notifications, generates reports, etc.
- You can use S3 Inventory to get object list and use S3 Select to filter your objects.

94. Key points about Amazon S3 - Storage Lens.

Ans ->

With the help of Amazon S3 - Storage Lens we can:

- Understand, analyze, and optimize storage across entire AWS Organization
- Discover anomalies, identify cost efficiencies, and apply data protection best practices across entire AWS Organization (30 days usage & activity metrics)
- Aggregate data for Organization, specific accounts, regions, buckets, or prefixes
- Default dashboard or create your own dashboards
- Can be configured to export metrics daily to an S3 bucket (in CSV or Parquet format).

95. Key points about Storage Lens - Default Dashboard.

Ans ->

- Visualize summarized insights and trends for both free and advanced metrics
- Default dashboard shows Multi-Region and Multi-Account data
- Preconfigured by Amazon S3
- Can't be deleted, but can be disabled

96. What are the different types of Storage Lens - Metrics available? Explain them.

Ans ->

Summary Metrics:

- general insights about your S3 storage
- StorageBytes, ObjectCount...
- Use Cases: identify the fastest-growing (or not used) buckets and prefixes.

Cost-Optimization Metrics:

- Provide Insights to manage and optimize your storage costs
- NonCurrentVersionStorageBytes, IncompleteMultipartUploadStorageBytes...
- Use Cases: identify buckets with incomplete multipart uploaded older than 7 days, identify which objects could be transitioned to lower-cost storage class.

Data-Protection Metrics:

- Provide insights for data protection features
- VersioningEnabledBucketCount, MFADeleteEnabledBucketCount, SSEKMSenabledBucketCount, CrossRegionReplicationRuleCount...
- Use Cases: identify buckets that aren't following data-protection best practices

Access-management Metrics:

- Provide insights for S3 Object Ownership
- ObjectOwnershipBucketOwnerEnforcedbucketCount...
- Use Cases: identify which Object Ownership settings your buckets use.

Event Metrics:

- Provide insights for S3 Event Notifications
- EventNotificationEnabledBucketCount (identify which buckets have S3 Event Notifications configured).

Performance Metrics:

- Provide insights for S3 Transfer Acceleration
- TransferAccelerationEnabledBucketCount (identify which buckets have S3 Transfer Acceleration enabled)

Activity Metrics:

- Provide insights about how your storage is requested
- AllRequests, GetRequests, PutRequests, ListRequests, BytesDownloaded...

Detailed Status Code Metrics:

- Provide insights for HTTP status codes
- 200OKStatusCount, 403ForbiddenErrorCount, 404NotFoundErrorCount...

97. Storage Lens - Free vs. Paid.

Ans ->

Free Metrics:

- Automatically available for all customers
- Contains around 28 usage metrics
- Data is available for queries for 14 days

Advanced Metrics and Recommendations:

- Additional paid metrics and features
- Advanced Metrics = Activity, Advanced Cost Optimization, Advanced Data Protection, Status Code
- CloudWatch Publishing - Access metrics in CloudWatch without additional charges
- Prefix Aggregation - Collect metrics at the prefix level
- Data is available for queries for 15 months.

98. Key points about Amazon S3 - Object Encryption.

Ans ->

You can encrypt objects in S3 buckets using one of 4 methods:

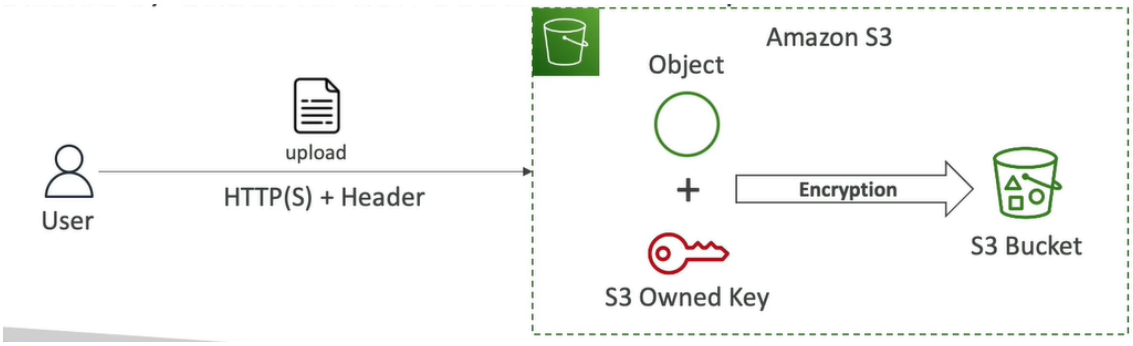
- **Server-Side Encryption (SSE)**

- **Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3)** - Enabled by Default
    - Encrypts S3 objects using keys handled, managed, and owned by AWS
- **Server-Side Encryption with KMS Keys stored in AWS KMS (SSE-KMS)**
    - leverage AWs Key Management Service (AWS KMS) to manage encryption keys
- **Server-Side Encryption with Customer-Provided Keys (SSE-C)**
    - When you want to manage your own encryption keys
- **Client-Side Encryption**

99. Key Points about Amazon S3 Encryption - SSE-S3.

Ans ->

- Encryption using keys handled, managed, and owned by AWS
- Object is encrypted server-side
- Encryption type is AES-256
- Must set header "x-amz-server-side-encryption":"AES256"
- Enabled by default for new buckets & new objects

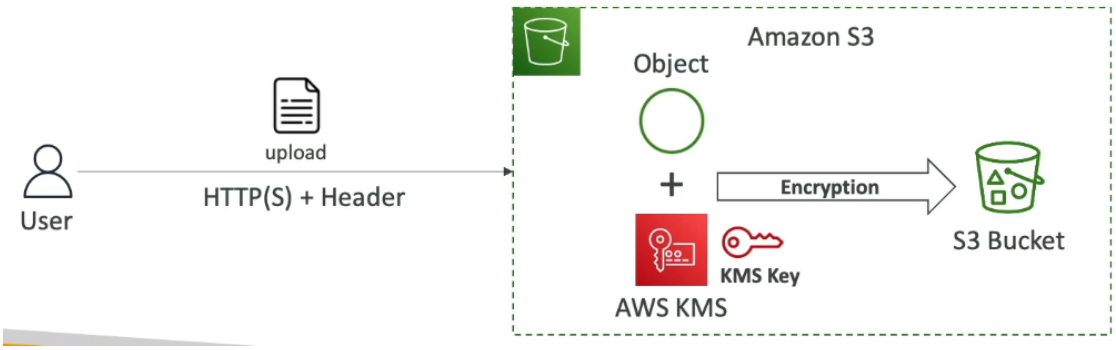


100. Key points about Amazon S3 Encryption - SSE-KMS.

Ans ->

- Encryption using keys handled and managed by AWS KMS (Key Management Service)

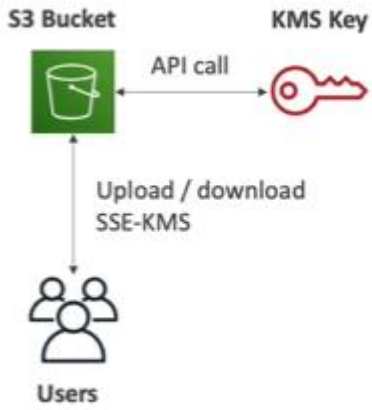
- KMS advantages: user control + audit key usage using CloudTrail
- Object is encrypted server side
- Must set header "x-amz-server-side-encryption":"aws:kms"



101. What are the Limitations of using SSE-KMS?

Ans ->

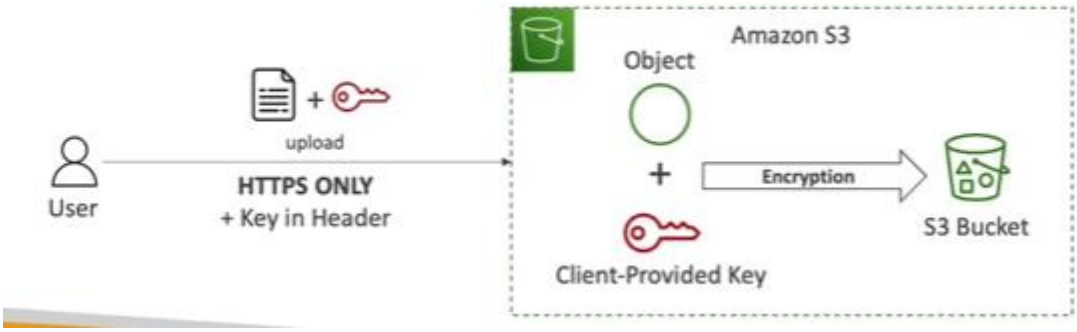
- If you use SSE-KMS, you may be impacted by the KMS limits.
- When you upload, it calls the GenerateDataKey KMS API
- When you download, it calls the Decrypt KMS API
- count towards the KMS quota per second (5500, 10000, 30000 req/s based on region)
- You can request a quota increase using the Service Quotas Console



102. Key points about Amazon S3 Encryption - SSE-C

Ans ->

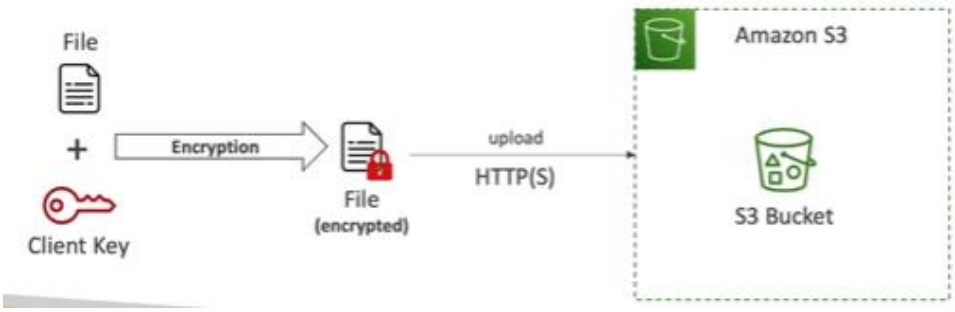
- Server-Side Encryption using keys fully managed by the customer outside of AWS
- Amazon S3 does NOT store the encryption key you provide
- HTTPS must be used
- Encryption key must be provided in HTTP headers, for every HTTP request made



103. Key points about S3 encryption - Client-Side Encryption

Ans ->

- Use client libraries such as Amazon S3 Client-Side Encryption Library
- Clients must encrypt data themselves before sending to Amazon S3
- Clients must decrypt data themselves when retrieving from amazon S3
- Customer fully manages the keys and encryption cycle



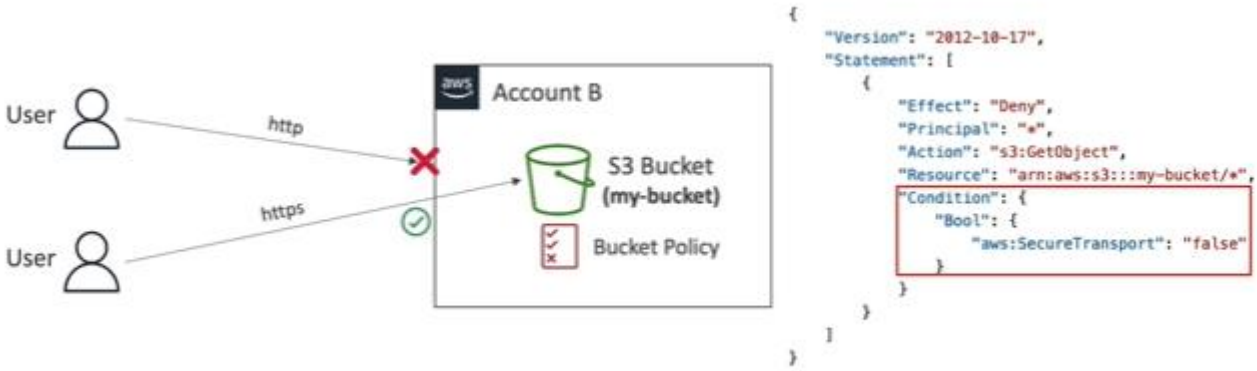
104. Amazon S3 - Encryption in transit (SSL/TLS)

Ans ->

- Encryption in flight is also called SSL/TLS
- Amazon S3 exposes to endpoints:
  - HTTP Endpoint - non encrypted
  - HTTPS Endpoint - encryption in flight
- HTTPS is recommended
- HTTPS is mandatory for SSE-C
- Most clients would use the HTTPS endpoint by default

105. Amazon S3 - Force Encryption in Transit aws:Secure Transport.

Ans ->



106. Amazon S3 - Default Encryption vs. Bucket Policies

Ans ->

- SSE-S3 encryption is automatically applied to new objects stored in S3 bucket
- Optionally, you can "force encryption" using a bucket policy and refuse any API call to PUT an S3 object without encryption headers (SSE-KMS or SSE-C).

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Deny",
            "Action": "s3:PutObject",
            "Principal": "*",
            "Resource": "arn:aws:s3:::my-bucket/*",
            "Condition": {
                "StringNotEquals": {
                    "s3:x-amz-server-side-encryption": "aws:kms"
                }
            }
        }
    ]
}
```

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Deny",
            "Action": "s3:PutObject",
            "Principal": "*",
            "Resource": "arn:aws:s3:::my-bucket/*",
            "Condition": {
                "Null": {
                    "s3:x-amz-server-side-encryption-customer-algorithm": "true"
                }
            }
        }
    ]
}
```

- Note: Bucket Policies are evaluated before "Default Encryption"

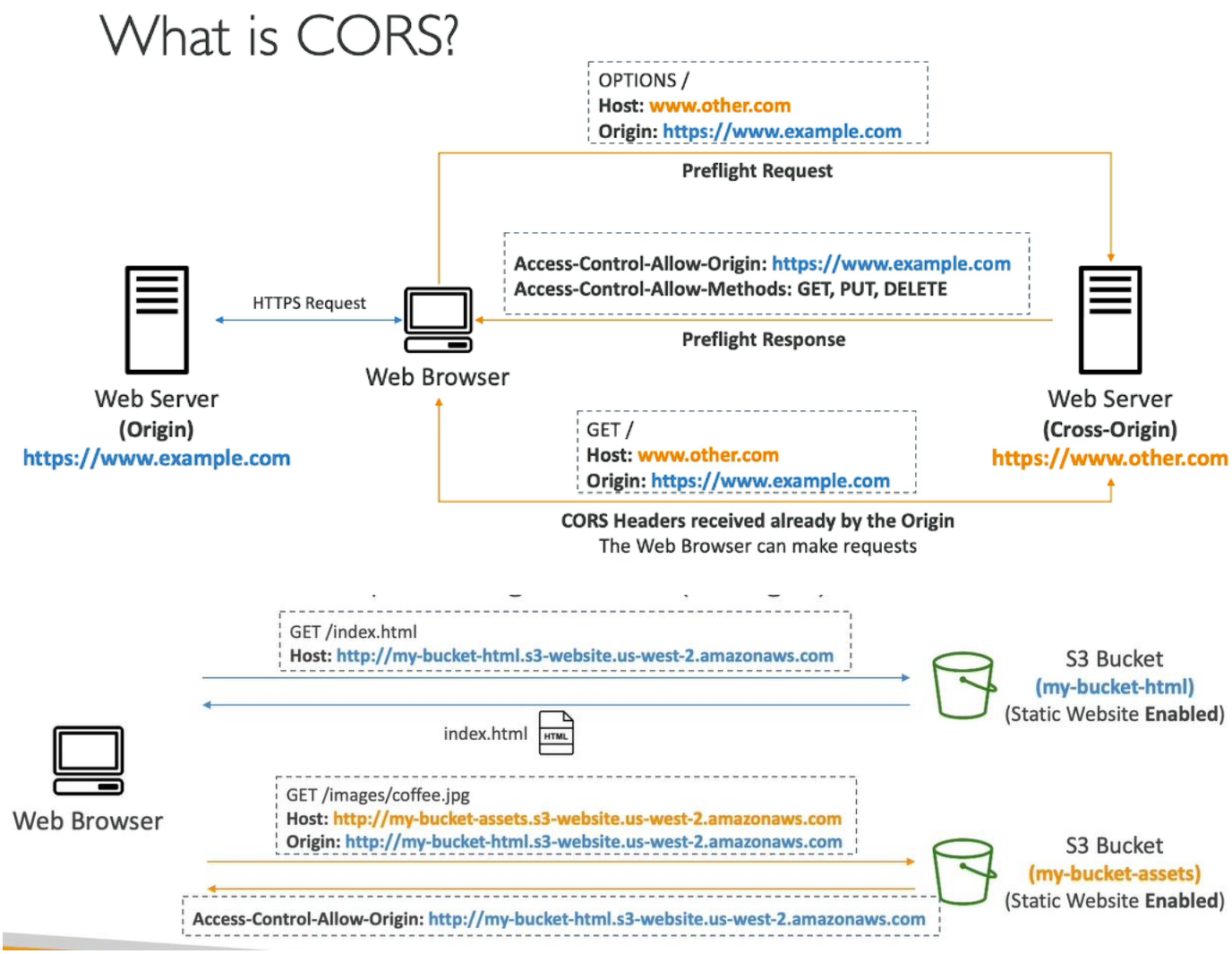
107. What is CORS?

Ans ->

Cross-Origin Resource Sharing (CORS) in AWS S3 lets your web applications access resources stored in S3 buckets from different domains. By setting up CORS rules, you control which websites can access your S3 data and how they can do it. This ensures that only approved websites can interact with your S3 resources, keeping your data safe while enabling useful cross-domain features for your web applications.

- **Origin = Scheme (protocol) + host (domain) + port**
  - example: https://www.example.com (implied port is 443 for HTTPS, 80 for HTTP)
- Web Browser based mechanism to allow requests to other origins while visiting the main origin

- Same origin: http://example.com/app1 & http://example.com/app2
- Different origin: http://www.example.com & http://other.example.com
- The request won't be fulfilled unless the other origin allows for the request, using CORS Headers (example: Access-Control-Allow-Origin).



108. Key points about Amazon S3 - MFA Delete.

Ans ->

- **MFA (Multi-Factor Authentication)** - force users to generate a code on a device (usually a mobile phone or hardware) before doing important operations on S3
- **MFA will be required to:**
  - permanently delete an object version
  - Suspend Versioning on the bucket
- **MFA won't be required to:**
  - Enable Versioning
  - List deleted versions
- To use MFA Delete, Versioning must be enabled on the bucket

- **Only the bucket owner (root account) can enable/disable MFA Delete.**

109. Key points about S3 Access Logs.

Ans ->

- For audit purpose, you may want to log all access to S3 buckets
- Any request made to S3, from any account, authorized or denied, will be logged into another S3 bucket.
- That data can be analyzed using data analysis tools...
- The target logging bucket must be in the same AWS region
- The log Format is at: https://docs.aws.amazon.com/AmazonS3/latest/dev/LogFormat.html

Note:

- Do not set your logging bucket to be the monitored bucket
- It will create a logging loop, and your bucket will grow exponentially.

110. What is Amazon S3 - Pre-Signed URLs?

Ans ->

Amazon S3 Pre-Signed URLs allow temporary access to S3 objects without needing AWS credentials. They can be used to share private files securely or allow uploads to an S3 bucket by generating a URL with specified permissions and an expiration time.

- **Generate pre-signed URLs using the S3 Console, AWS CLI or SDK**
- **URL Expiration:**
  - S3 Console -> 1 min up to 720 mins (12 hours)
  - AWS CLI -> configure expiration with -expires-in parameter in seconds (default 3600 secs, max 604800 secs ~ 168 hours)
- Users given a pre-signed URL inherit the permissions of the user that generated the URL for GET/PUT.
- **Examples:**
  - Allow only logged-in users to download a premium video from your S3 bucket.
  - Allow an ever-changing list of users to download files by generating URLs dynamically
  - Allow temporarily a user to upload a file to a precise location in you S3 bucket.

111. What is Amazon S3 Glacier Vault Lock?

Ans ->

Amazon S3 Glacier Vault Lock lets you set and lock policies to keep your data safe and compliant with regulations that require write-once-read-many (WORM) storages. Once you lock a policy, it can't be changed or deleted, ensuring your data stays untouched for the required period. This feature helps you easily meet legal requirements and protect your data without managing complex systems.

- Adopt a **WORM (Write Once Read Many)** model
- Create a Vault Loc Policy
- Lock the policy for future edits (can no longer be changed or deleted by any user)
- Helpful for compliance and data retention.

112. What is S3 Object Lock?

Ans ->

S3 Object Lock is a feature that allows you to store objects using a **write-once-read-many (WORM)** model. This ensures that the data cannot be deleted or overwritten for a specified period or indefinitely.

Block an object version deletion for a specified amount of time

Retention mode:

- **Compliance:**
  - Object versions can't be overwritten or deleted by any user, including the root user
  - Objects retention modes can't be changed, and retention periods can't be shortened
- **Governance:**
  - Most users can't overwrite or delete an object version or alter its lock settings
  - Some users have special permissions to change the retention or delete the object

Retention Period: Protect the object for a fixed period, it can be extended

Legal Hold:

- Protect the object indefinitely, independent from retention period
- can freely placed and removed using the s3:PutObjectLegalHold IAM permission

113. What is the difference between S3 Glacier Vault Lock and S3 Object Lock?

Ans ->

- **S3 Glacier Vault Lock:** Enforce compliance for archived data in Glacier.
- **S3 Object Lock:** Ensure immutability for S3 objects.

- **S3 Glacier Vault Lock:** It uses amazon S3 Glacier storage class.
- **S3 Object Lock:** It uses amazon S3 (Standard, Intelligent-Tiering, Standard-IA, One Zone-IA).

- **S3 Glacier Vault Lock:** Meets regulations like SEC Rule 17a-4(f) and CFTC Rule 1.31.
- **S3 Object Lock:** Supports various WORM compliance needs.

- **S3 Glacier Vault Lock:** Set retention periods in policies, which become immutable once locked.
- **S3 Object Lock:** Governance Mode (some modifications allowed) and Compliance Mode (no modifications).

- **S3 Glacier Vault Lock:** Doesn't support Legal Holds.
- **S3 Object Lock:** Supports Legal Holds.

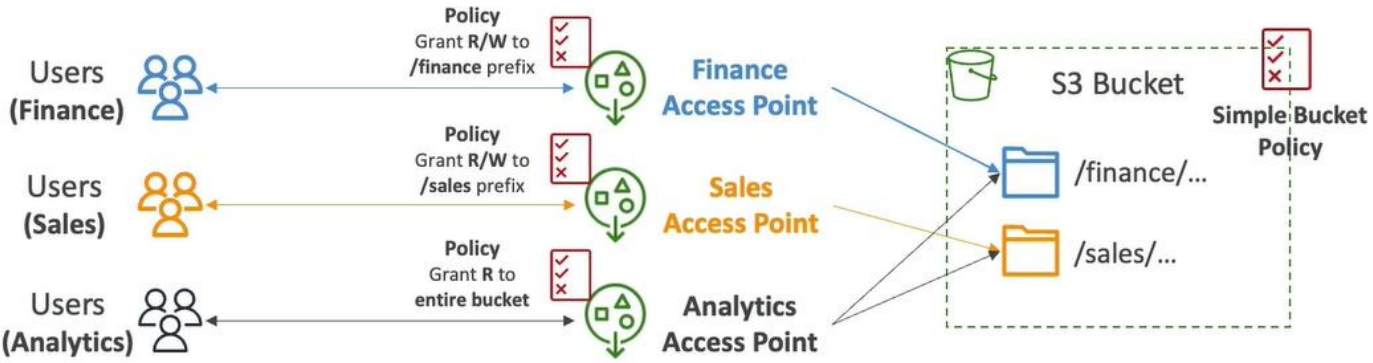
- **S3 Glacier Vault Lock:** Data retrieval takes minutes to hours.
- **S3 Object Lock:** Standard S3 retrieval times.

114. What is Amazon S3 - Access Points?

Ans ->

Amazon S3 Access Points are a feature that simplifies managing data access at scale for applications using shared datasets in S3. Instead of managing a single bucket policy, you can create multiple access points, each with its own policy, to provide specific permissions for different application or teams. This makes it easier to control access to your data, especially in environments with multiple users or applications.

- Access Points simplify security management for S3 buckets
- Each Access Point has:
  - its own DNS name (Internet Origin or VPC Origin)
  - an access point policy (similar to bucket policy) - manage security at scale.

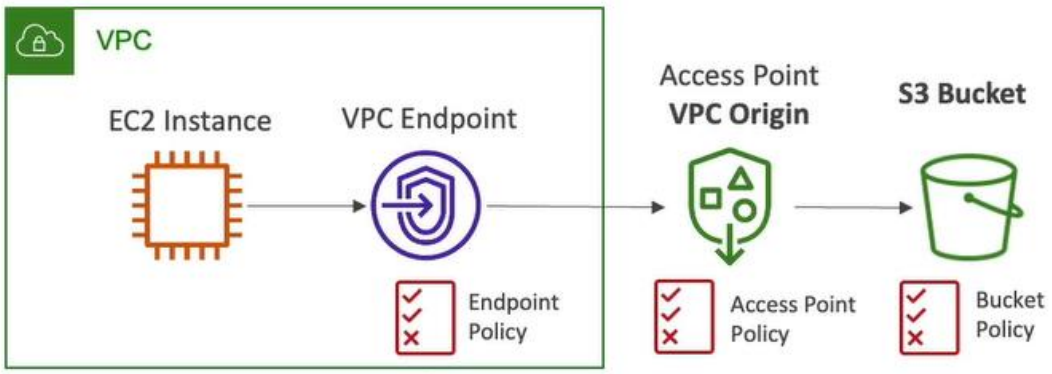


115. What is Amazon S3 - Access Points - VPC Origin?

Ans ->

Amazon S3 Access Points with VPC Origin provide a way to securely access S3 data from within a specific VPC. This ensures that S3 data can only be accessed through the specified VPC, adding an extra layer of security by isolating the data access within your private network.

- We can define the access point to be accessible only from within the VPC
- You must create a VPC Endpoint to access the Access Point (Gateway or Interface Endpoint)
- The VPC Endpoint Policy must allow access to the target bucket and Access Point.



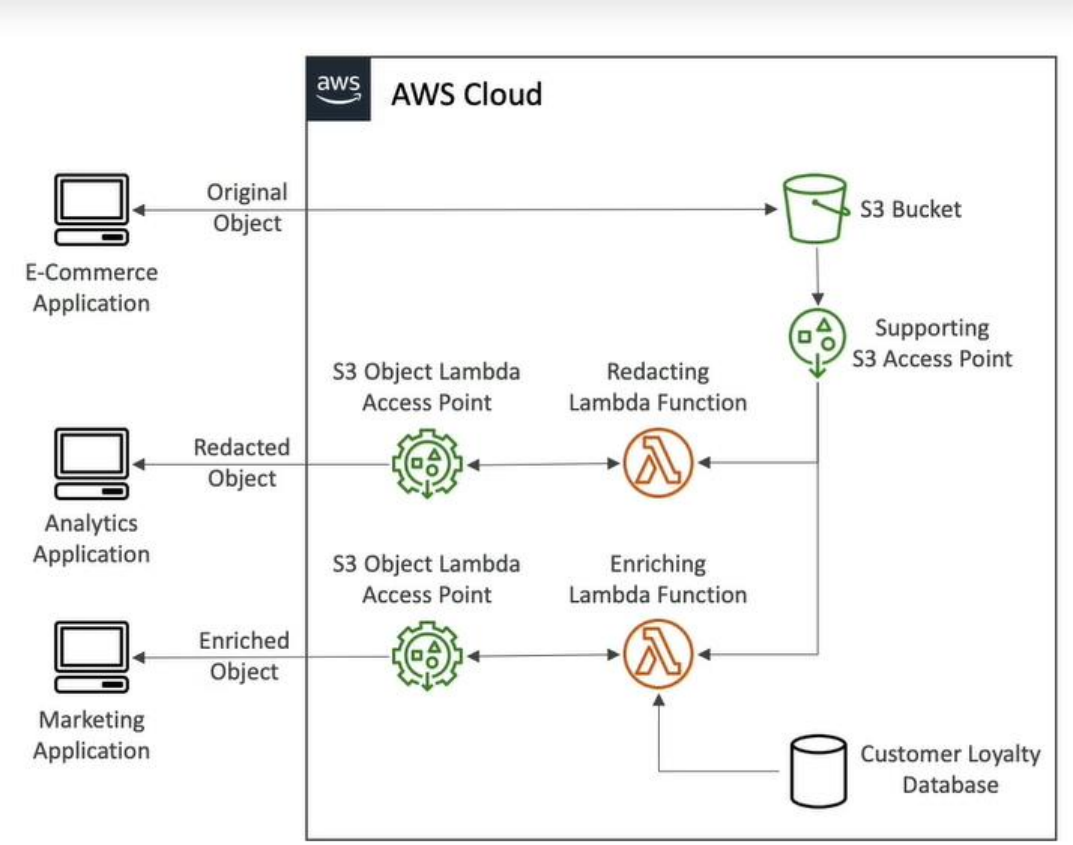

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Principal": "*",
            "Action": [
                "s3:GetObject"
            ],
            "Effect": "Allow",
            "Resource": [
                "arn:aws:s3:::awsexamplebucket1/*",
                "arn:aws:s3:us-west-2:123456789012:accesspoint/example-vpc-ap/object/*"
            ]
        }
    ]
}
```

116. What is Amazon S3 Object Lambda?

Ans ->

Amazon S3 Object Lambda allows you to process and transform data as it is retrieved from S3 using AWS Lambda functions. This means you can automatically modify the data on-the-fly without needing to store multiple versions of the data or perform pre-processing.

- Use AWS Lambda Functions to change the object before it is retrieved by the caller application.
- Only one S3 bucket is needed, on top of which we create S3 Access Point and S3 Object Lambda Access Points.
- **Use Cases:**
  - Redacting personally identifiable information (PII) for analytics or non-production environments.
  - Converting across data formats, such as converting XML to JSON.
  - Resizing and watermarking images on the fly using caller-specific details, such as the user who requested the object.



117. Key points about AWS CloudFront.

Ans ->

Amazon CloudFront is a Content Delivery Network (CDN) that caches and delivers web content, such as HTML, images, and videos, to users globally with low latency by using a network of edge locations.

- Content Delivery Network (CDN)
- Improves read performance, content is cached at the edge location
- Improves user experience
- **216 Point** of Presence globally (edge locations)
- DDoS protection (because worldwide), integration with Shield, AWS Web Application Firewall

118. Give some examples of CloudFront - Origins.

Ans ->

S3 bucket:

- For distributing files and caching them at the edge
- Enhanced security with CloudFront Origin Access Control (OAC)
- OAC is replacing Origin Access Identity (OAI)
- CloudFront can be used as an ingress (to upload files to S3)

HTTP/S Servers:

- This can be any web server that supports HTTP or HTTPS, including EC2 instances, Elastic Load Balancers, or servers outside of AWS

AWS Lambda Functions:

- Allows you to run code to generate content dynamically without provisioning servers.

Amazon API Gateway:

- For creating, publishing, maintaining, monitoring, and securing REST, HTTP, and WebSocket APIs.

Custome Origin (HTTP):

- Application Load balancer
- EC2 instance

- S3 website (must first enable the bucket as a statis S3 website)
- Any HTTP backend you want

119. What are the differences between CloudFront and S3 CRR?

Ans ->

CloudFront:

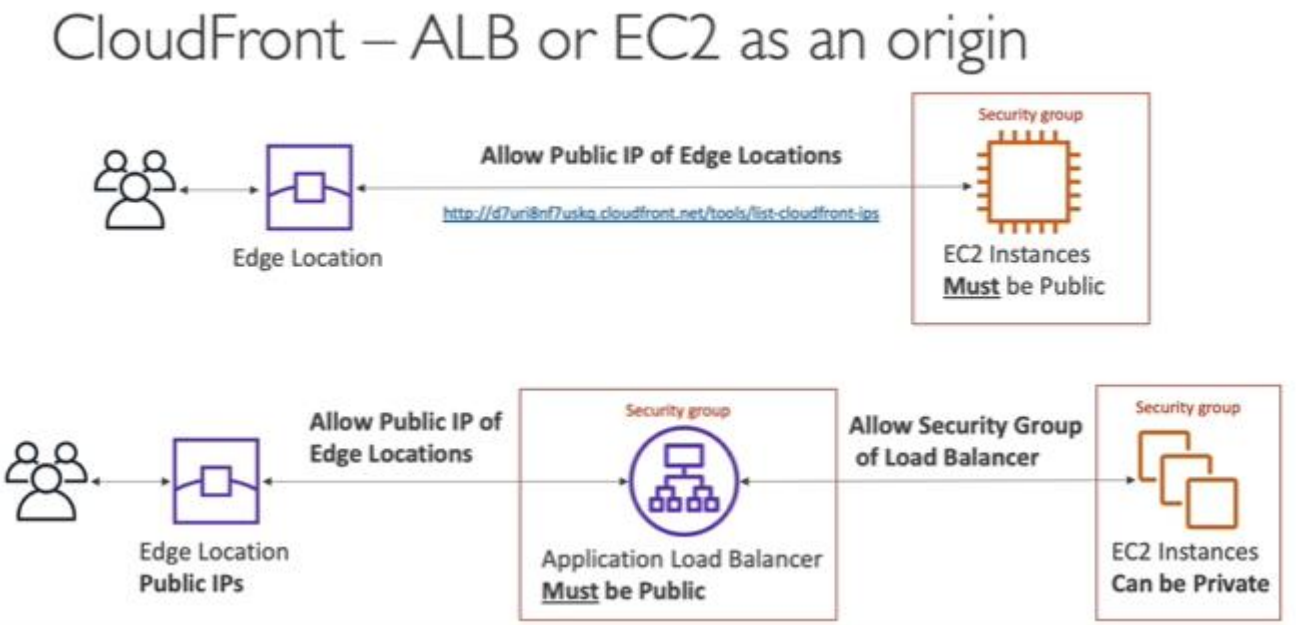
- It uses Global Edge network which has globally 216 points of presence.
- Files are cached for a TTL (maybe a day)
- Great for static content that must be available everywhere

S3 Cross Region Replication (CRR):

- Must be setup for each region you want replication to happen
- Files are updated in near real-time Read only (caching not supported)
- Great for dynamic content that needs to be available at low latency in few regions.

120. CloudFront - ALB or EC2 as an Origin.

Ans ->



121. Key points about CloudFront Geo Restriction.

Ans ->

- You can restrict who can access your distribution
  - **Allowlist:** Allow your users to access your content only if they're in one of the countries on a list of approved countries.
  - **Blocklist:** Prevent your users from accessing your content if they're in one of the countries on a list of banned countries.
- The "country" is determined using a **3rd party Geo-IP database**
- **Use case:** Copyright Laws to control access to content

122. CloudFront - Pricing classes.

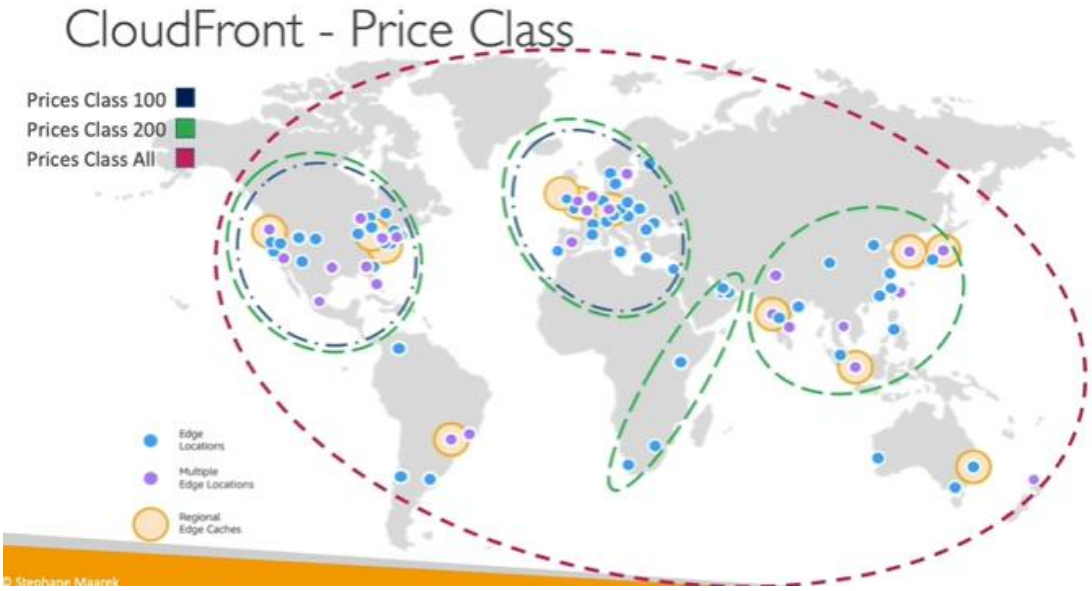
Ans ->

- CloudFront Edge locations are all around the world
- The cost of data out per edge location varies

Per Month	United States, Mexico, & Canada	Europe & Israel	South Africa, Kenya, & Middle East	South America	Japan	Australia & New Zealand	Hong Kong, Philippines, Singapore, South Korea, Taiwan, & Thailand	India
First 10TB	$0.085	$0.085	$0.110	$0.110	$0.114	$0.114	$0.140	$0.170
Next 40TB	$0.080	$0.080	$0.105	$0.105	$0.089	$0.098	$0.135	$0.130
Next 100TB	$0.060	$0.060	$0.090	$0.090	$0.086	$0.094	$0.120	$0.110
Next 350TB	$0.040	$0.040	$0.080	$0.080	$0.084	$0.092	$0.100	$0.100
Next 524TB	$0.030	$0.030	$0.060	$0.060	$0.080	$0.090	$0.080	$0.100
Next 4PB	$0.025	$0.025	$0.050	$0.050	$0.070	$0.085	$0.070	$0.100
Over 5PB	$0.020	$0.020	$0.040	$0.040	$0.060	$0.080	$0.060	$0.100

- You can reduce the number of edge locations for cost reduction
- Three price classes:
  1. Price Class All: all regions – best performance
  2. Price Class 200: most regions, but excludes the most expensive regions
  3. Price Class 100: only the least expensive regions

Edge Locations Included Within	United States, Mexico, & Canada	Europe & Israel	South Africa, Kenya, & Middle East	South America	Japan	Australia & New Zealand	Hong Kong, Philippines, Singapore, South Korea, Taiwan, & Thailand	India
Price Class All	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Price Class 200	Yes	Yes	Yes	x	Yes	x	Yes	Yes
Price Class 100	Yes	Yes	x	x	x	x	x	x



123. What is CloudFront - Cache Invalidations?

Ans ->

CloudFront cache invalidation is the process of removing files from the CloudFront cache before they expire. This is useful when you update content on your origin server and need to ensure that the latest version is served to users immediately.

Invalidation Requests:

you create invalidation requests specifying the file paths that need to be removed from the cache. You can invalidate specific files (e.g., /images/logo.png) or use wildcard characters to invalidate multiple files at once (e.g., /images/*, or just "*" for all files).

124. What is Unicast IP and Anycast IP?

Ans ->

Unicast IP:

A unicast IP address is an address that uniquely identifies a single host or device in a network. Data sent to a unicast address is delivered only to the specific device identified by that address.

Anycast IP:

An anycast IP address is an address assigned to multiple devices, with the goal of routing data to the nearest or best destination based on the routing protocol's decision.

125. Explain about Amazon Global Accelerator.

Ans ->

Amazon Global Accelerator is a service that improves the performance and availability of your applications by using the AWS global network infrastructure to route user traffic through the fastest and most reliable paths. It provides static **anycast IP addresses** that act as a fixed entry point for your application, ensuring lower latency, faster speeds, and automatic traffic redirection if an endpoint becomes unavailable, thus maintaining high availability and resilience.

Key points:

- Leverage the AWS internal network to route to your application
- 2 Anycast IP are created for your application
- The Anycast IP send traffic directly to Edge Locations
- The Edge locations send the traffic to your application

126. Some important points about AWS Global Accelerator.

Ans ->

- Works with Elastic IP, EC2 instances, ALB, NLB, public or private

Consistent Performance:

- Intelligent routing to lowest latency and fast regional failover
- No issue with client cache (because the IP doesn't change)
- Uses internal AWS network

Health Checks:

- Global Accelerator performs a health check of your applications
- Helps make your application global (failover less than 1 minute for unhealthy)
- Great for disaster recovery (thanks to the health checks)

Security:

- Only 2 external IP need to be whitelisted
- DDOS protection thanks to AWs Shield

127. Global Accelerator Data Transfer pricing.

Ans ->

Fixed fee: For every full or partial hour when an accelerator runs in your account, you are charged $0.025 until it is deleted.

		Destination (AWS edge location)								
		United States, Mexico & Canada	Europe, Israel, & Türkiye	Asia Pacific*	South Korea	India, Philippines, Thailand & Vietnam	Australia & New Zealand	Middle East	South America	South Africa, Kenya & Nigeria
Source (AWS Regions)	United States & Canada	$0.015 /GB	$0.015 /GB	$0.035 /GB	$0.035 /GB	$0.035 /GB	$0.105 /GB	$0.035 /GB	$0.040 /GB	$0.079 /GB
	Europe & Israel	$0.015 /GB	$0.015 /GB	$0.033 /GB	$0.035 /GB	$0.033 /GB	$0.105 /GB	$0.035 /GB	$0.043 /GB	$0.067 /GB
	Asia Pacific*	$0.012 /GB	$0.043 /GB	$0.010 /GB	$0.058 /GB	$0.032 /GB	$0.080 /GB	$0.055 /GB	$0.049 /GB	$0.065 /GB
	South Korea	$0.017 /GB	$0.020 /GB	$0.043 /GB	$0.043 /GB	$0.067 /GB	$0.080 /GB	$0.050 /GB	$0.049 /GB	$0.060 /GB
	India	$0.025 /GB	$0.023 /GB	$0.023 /GB	$0.072 /GB	$0.023 /GB	$0.094 /GB	$0.074 /GB	$0.057 /GB	$0.058 /GB
	Australia	$0.070 /GB	$0.070 /GB	$0.074 /GB	$0.091 /GB	$0.101 /GB	$0.007 /GB	$0.084 /GB	$0.088 /GB	$0.093 /GB
	Middle East	$0.032 /GB	$0.029 /GB	$0.029 /GB	$0.047 /GB	$0.029 /GB	$0.082 /GB	$0.057 /GB	$0.061 /GB	$0.067 /GB
	South America	$0.032 /GB	$0.035 /GB	$0.041 /GB	$0.056 /GB	$0.059 /GB	$0.083 /GB	$0.057 /GB	$0.024 /GB	$0.067 /GB
	South Africa	$0.043 /GB	$0.038 /GB	$0.045 /GB	$0.057 /GB	$0.058 /GB	$0.091 /GB	$0.067 /GB	$0.071 /GB	$0.010 /GB

128. What are the similarities and differences between AWS CloudFront and Global Accelerator?

Ans ->

Similarities:

- They both use the AWS global network and its edge locations around the world
- Both services integrate with AWS Shield for DDoS protection

Differences:

- **CloudFront:**
  - improves performance for both cacheable content (such as images and videos)
  - Dynamic content (such as API acceleration and dynamic site delivery)
  - Content is served at the edge
- **Global Accelerator:**
  - Improves performance for a wide range of applications over TCP or UDP

- Proxying packets at the edge to applications running in one or more AWS Regions.
- Good fit for non-HTTP use cases, such as gaming (UDP), IoT (MQTT), or Voice over IP
- Good for HTTP use cases that requires static IP addresses
- Good for HTTP use cases that required deterministic, fast regional failover

129. Explain about AWS Snow Family.

Ans ->

The AWS Snow Family consists of highly-secure, portable physical devices used to collect and process data at the edge (aka Edge Computing) and migrate large amounts data into and out of AWS. It includes:

- **AWS Snowcone:** A small, portable device for edge computing, storage, and data transfer, handling up to 8 TB of data
- **AWS Snowball:** A rugged device available in two versions, Snowball Edge Storage Optimized and Snowball Edge Compute Optimized, for transferring up to petabytes of data and providing edge computing capabilities.
- **AWS Snowmobile:** A 45-foot-long ruggedized shipping container that transport up to 100 PB of data, designed for large-scale data migrations.

These devices help customers migrate data, perform edge computing, and ensure data security during transit.

130. Key points about AWS Snowcone & Snowcone SSD.

Ans ->

- Small, portable computing, anywhere, rugged & secure, withstands harsh environments
- Light weight (4.5 pounds, 2.1 kg)
- Device used for edge computing, storage, and data transfer
- Snowcone - 8 TB of HDD Storage
- Use Snowcone where Snowball does not fit (space-constrained environment)
- Must provide your own battery/cables

- Can be sent back to AWS offline or connect it to internet and use AWS DataSync to send data.

131. Key points about Snowball Edge (for data transfers).

Ans ->

- Physical data transport solution: move TBs or PBs of data in or out of AWS.
- Alternative to moving data over the network (and paying network fees)
- Pay per data transfer job
- Provide block storage and Amazon S3-compatible object storage
- **Snowball Edge Storage Optimized:**
  - 80 TB of HDD or 210 TB of NVMe capacity for block volume and S3 compatible object storage.
- **Snowball Edge Compute Optimized:**
  - 42 TB of HDD or 28 TB of NVMe capacity for block volume and S3 compatible object storage.
- Use cases: large data cloud migrations, DC decommission, disaster recovery

132. Key points about AWS Snowmobile.

Ans ->

- Transfer exabytes of data (1 EB = 1,000 PB = 1,000,000 TBs)
- Each Snowmobile has 100 PB of capacity (use multiple in parallel)
- High security: temperature controlled, GPS, 24/7 video surveillance
- Better than Snowball if you transfer more than 10PB

133. AWS Snow Family for Data Migrations.

Ans ->

Snowcone

Snowball Edge

Snowmobile

	Snowcone & Snowcone SSD	Snowball Edge Storage Optimized	Snowmobile
Storage Capacity	8 TB HDD 14 TB SSD	80 TB - 210 TB	< 100 PB
Migration Size	Up to 24 TB, online and offline	Up to petabytes, offline	Up to exabytes, offline
DataSync agent	Pre-installed		

134. What is the process of using AWS Snow Family?

Ans ->

- Request Snowball devices from the AWS console for delivery
- Install the snowball client/AWS OpsHub on your servers
- Connect the snowball to your servers and copy files using the client
- ship back the device when you're done (goes to the right AWS facility)
- Data will be loaded into an S3 bucket
- Snowball is completely wiped

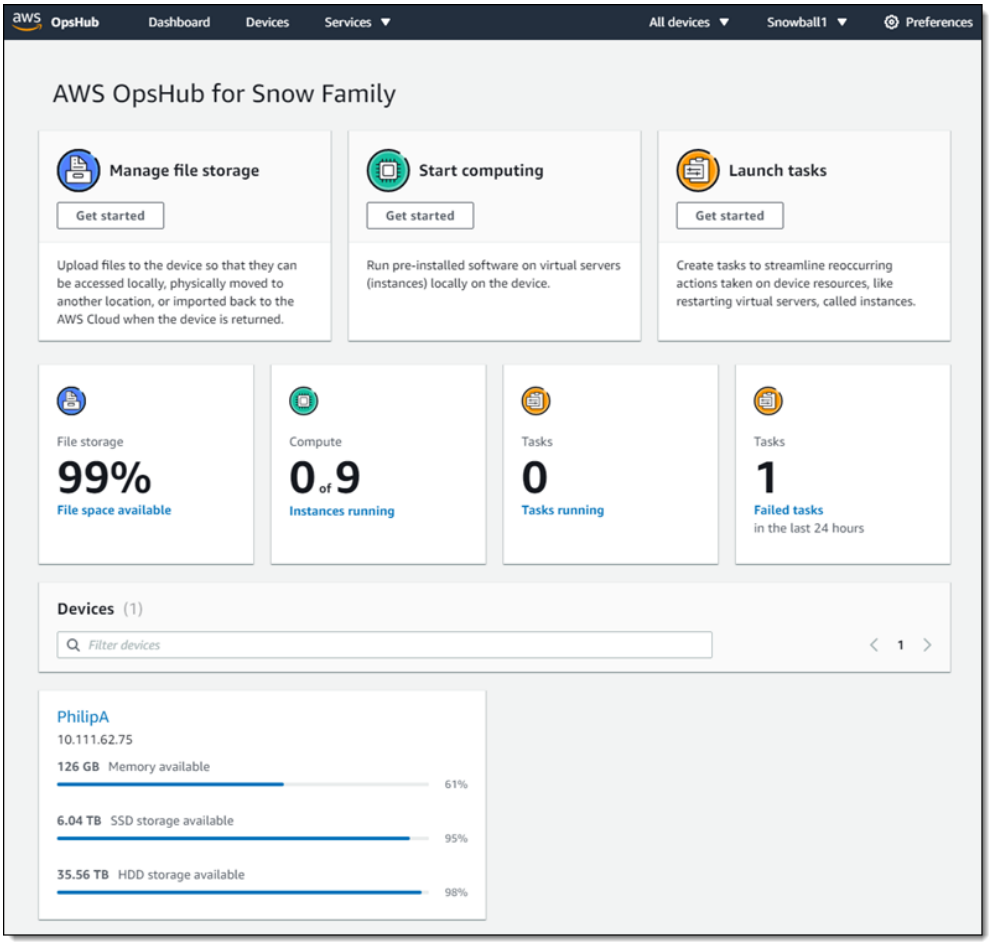
135. What is AWS OpsHub?

Ans ->

AWS OpsHub is a centralized operational hub designed to simplify management and monitoring of AWS resources across multiple accounts and regions. It provides a unified view of operational data, enabling administrators to automate tasks, monitor performance, and manage configurations more efficiently. OpsHub integrates with various AWS services, offering insights into resource utilization, cost management, and compliance, thereby enhancing operational efficiency and ensuring better governance across complex AWS environments.

OpsHub for Snow Family:

- Historically, to use Snow Family devices, you needed a CLI (Command Line Interface tool)
- Today, you can use AWS OpsHub (a software you install on your computer / laptop) to manage your Snow Family Device
    - Unlocking and configuring single or clustered devices Transferring files
    - Launching and managing instances running on Snow Family Devices
    - Monitor device metrics (storage capacity, active instances on your device)
    - Launch compatible AWS services on your devices (ex: Amazon EC2 instances, AWS DataSync, Network File System (NFS).



136. What is Edge Computing?

Ans ->

Edge computing is a distributed computing model that brings computation and data storage closer to the location where it is needed, typically at or near the data source. This approach reduces latency, improves response times, and decreases bandwidth usage by processing data locally rather than in a centralized cloud. It's commonly used in applications like IoT, autonomous vehicles, and real-time analytics.

- Process data while it's being created on an edge location
  - A truck on the road, a ship on the sea, a mining station underground
- These locations may have:
  - Limited / no internet access
  - Limited / no easy access to computing power
- We setup a Snowball Edge / Snowcone device to do edge computing
- Use cases of Edge Computing:
  - Preprocess data Machine
  - learning at the edge
  - Transcoding media streams

- Eventually (if need be) we can ship back the device to AWS (for transferring data for example)

137. AWS Snow Family for Edge Computing.

Ans ->

- **Snowcone & Snowcone SSD (smaller):**
  - 2 CPUs, 4 GB of memory, wired or wireless access
  - USB-C power using a cord or the optional battery
- **Snowball Edge - Compute Optimized:**
  - 104 vCPUs, 416 GiB of RAM
  - Optional GPU (useful for video processing or machine learning)
  - 28 TB NVMe or 42 TB HDD usable storage
  - Storage Clustering available (up to 16 nodes)
- **Snowball Edge - Storage Optimized:**
  - Up to 40 vCPUs, 80 GiB of RAM, 80 TB storage
  - Up to 104 vCPUs, 416 GiB of RAM, 210 TB NVMe storage
- All: Can run EC2 Instances & AWS Lambda functions (using AWS IoT Greengrass)
- Long-term deployment options: 1 and 3 years discounted pricing

138. Can Snowball directly import to Glacier?

Ans ->

- Snowball cannot import to Glacier directly
- You must use Amazon S3 first, in combination with an S3 lifecycle policy

139. What is Amazon FSx?

Ans ->

Amazon FSx is a fully managed AWS service offering scalable, high-performance file systems for various workloads. It's used for simplifying storage management with features like automatic backups, data replication, and integration with AWS services.

Amazon FSx launch 3rd-party high-performance file systems on AWS.

It supports:

- FSx for Windows File Server
- FSx for Lustre
- FSx for NetApp ONTAP
- FSx for OpenZFS

140. Key points about Amazon FSx for Windows File Server.

Ans ->

- FSx for Windows is fully managed Windows file system share drive
- Supports SMB protocol & Windows NTFS
- Microsoft Active Directory integration, ACLs, user quotas
- Can be mounted on Linux EC2 instances
- supports Microsoft's Distributed File Systems (DFS) Namespaces (group files across multiple FS)
- **Performance:**
  - Scale up to 10s of GB/s, millions of IOPS, 100s PB of data
  - Storage Options:
    - SSD - latency sensitive workloads (databases, media processing, data analytics, ...)
    - HDD - broad spectrum of workloads (home directory, CMS, ...)
  - Can be configured to be Multi-AZ (high availability)
  - Can be accessed from your on-premises infrastructure (VPN or Direct Connect)
  - Data is backed-up daily to S3

141. Amazon FSx for Lustre.

Ans ->

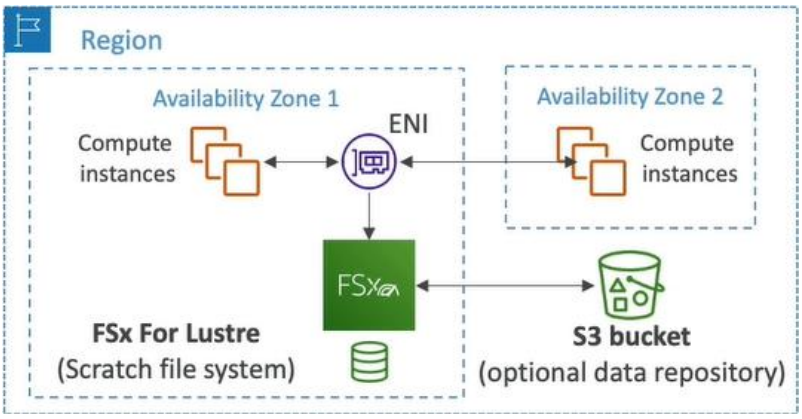
- Lustre is a type of parallel distributed file system, for large-scale computing The name Lustre is derived from "Linux" and "Cluster".
- Performance:
  - Machine Learning, High Performance Computing (HPC)
  - Video Processing, Financial Modeling, Electronic Design Automation
  - Scales upto 100s GB/s, millions of IOPS, sub-ms latencies
  - Storage Options:
    - SSD - low-latency, IOPS intensive workloads, small & random file operations
    - HDD - throughput-intensive workloads, large & sequential file operations

- Seamless integration with S3
  - Can "read S3" as a file system (through FSx)
  - Can write the output of the computations back to S3 (through FSx)
- Can be used from on-premises servers (VPN or Direct Connect)

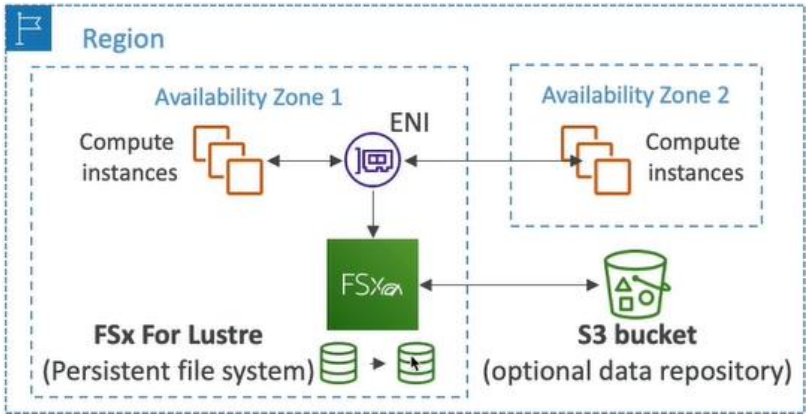
142. FSx File System Deployment Options:

Ans ->

- **Scratch File System:**
  - Temporary storage
  - Data is not replicated (doesn't persist if file server fails)
  - High burst (6x faster, 200MBps per TiB)
  - Usage: short-term processing, optimize costs



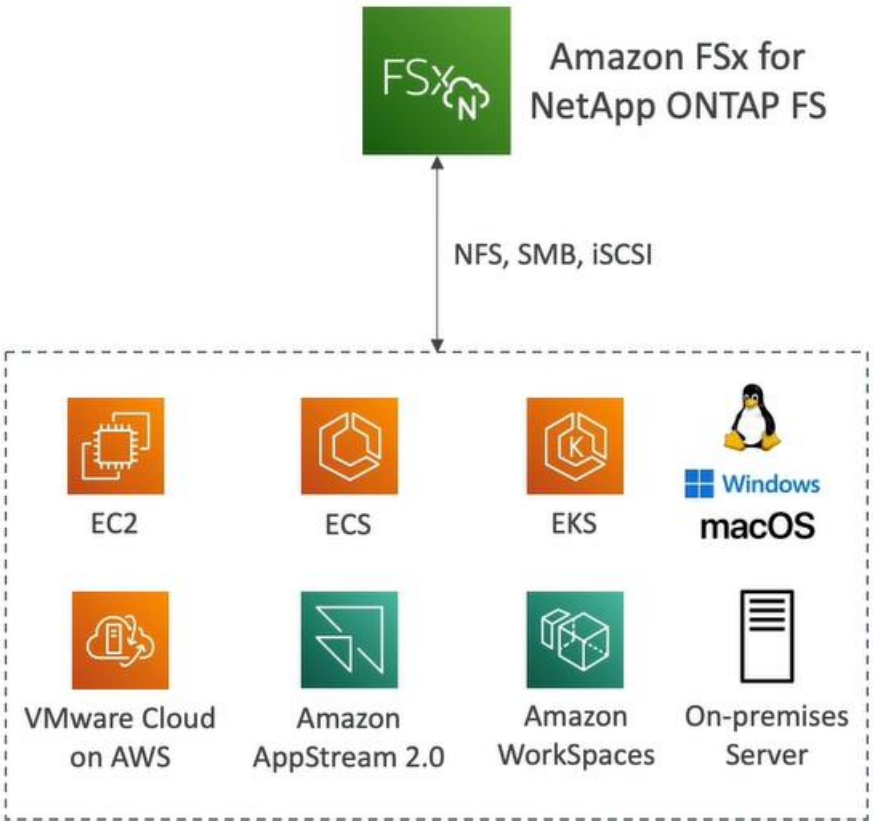
- **Persistent File System:**
  - Long-term storage
  - Data is replicated within same AZ
  - Replace failed files within minutes
  - Usage: long-term processing, sensitive data



143. Key points about Amazon FSx for NetApp ONTAP.

Ans ->

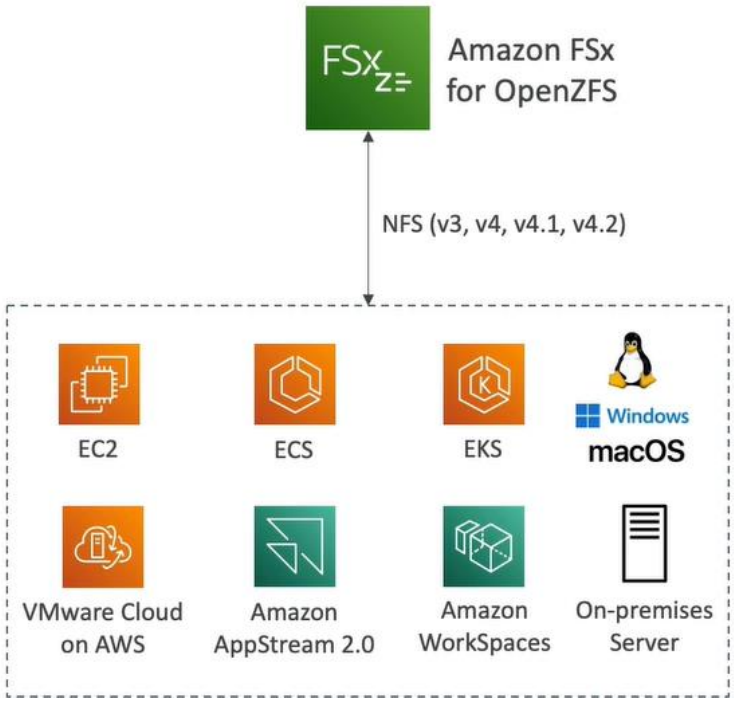
- Managed NetApp ONTAP on AWS
- File System compatible with NFS, SMB, ISCSI protocol
- Move workloads running on ONTAP or NAS to AWS
- Works with:
  - Linux
  - Windows
  - MacOS
  - VMware Cloud on AWS
  - Amazon Workspaces & AppStream 2.0
  - Amazon EC2, ECS and EKS
- Storage shrinks or grows automatically
- Snapshots, replications, low-cost, compression and data de-duplication
- Point-in-time instantaneous cloning (helpful for testing new workloads)



144. Amazon FSx for OpenZFS.

Ans ->

- Managed OpenZFS file system on AWS
- File System compatible with NFS (v3, v4, v4.1, v4.2)
- Move workloads running on ZFS to AWS
- Works with:
  - Linux
  - Windows
  - MacOS
  - VMware Cloud on AWs
  - Amazon Workspaces 7 AppStream 2.0
  - Amazon EC2, ECS and EKS
- Up to 1,000,000 IOPS with < 0.5ms latency
- Snapshots, compression and low-cost
- Point-in-time instantaneous cloning (helpful for testing new workloads)



145. What are the differences between all these FSx file systems?

Ans ->

Feature	FSx for Windows File Server	FSx for Lustre	FSx for NetApp ONTAP	FSx for OpenZFS
Primary Use Case	Windows-based applications	High-performance computing, ML	Enterprise applications, advanced data management	Linux/Unix workloads, advanced data protection
Protocols Supported	SMB	Lustre	NFS, SMB, iSCSI	NFS, SMB
Key Features	Native Windows compatibility, Active Directory integration, DFS, Windows ACLs	High throughput, low latency, Amazon S3 integration	NetApp ONTAP features (deduplication, compression, snapshots)	Snapshots, clones, data integrity verification
Data Transfer Integration	-	Amazon S3	-	-

Ideal For	Applications requiring Windows file system features	HPC, ML workloads needing fast data access	Enterprise data management needs	Workloads needing robust data protection and management

146. What is AWS Storage Gateway?

Ans ->

AWS Storage Gateway is a hybrid cloud storage service that enables on-premises applications to seamlessly use AWS cloud storage. It provides a bridge between on-premises data and cloud storage, allowing you to securely store data in the AWS cloud for scalable and cost-effective storage.

Types of Storage Gateway:

- S3 File Gateway
- FSx File Gateway
- Volume Gateway
- Tape Gateway

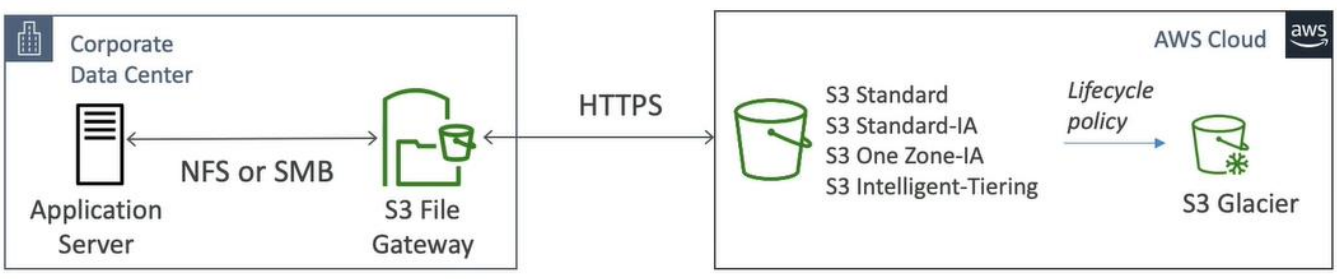
147. Key points about Amazon S3 File Gateway.

Ans ->

Amazon S3 File Gateway allows on-premises applications to access Amazon S3 as if it were a local file system. It provides a seamless connection between on-premises environments and Amazon S3 by presenting a file interface and storing files as objects in S3.

Key features:

- Configured S3 buckets are accessible using the NFS and SMB protocol
- Most recently used data is cached in the file gateway for quick response
- Supports S3 Standard, S3 Standard IA, S3 One Zone IA, S3 Intelligent Tiering
- Transition to S3 Glacier using a Lifecycle Policy
- Bucket access using IAM roles for each File Gateway
- SMB Protocol has integration with Active Directory (AD) for user authentication



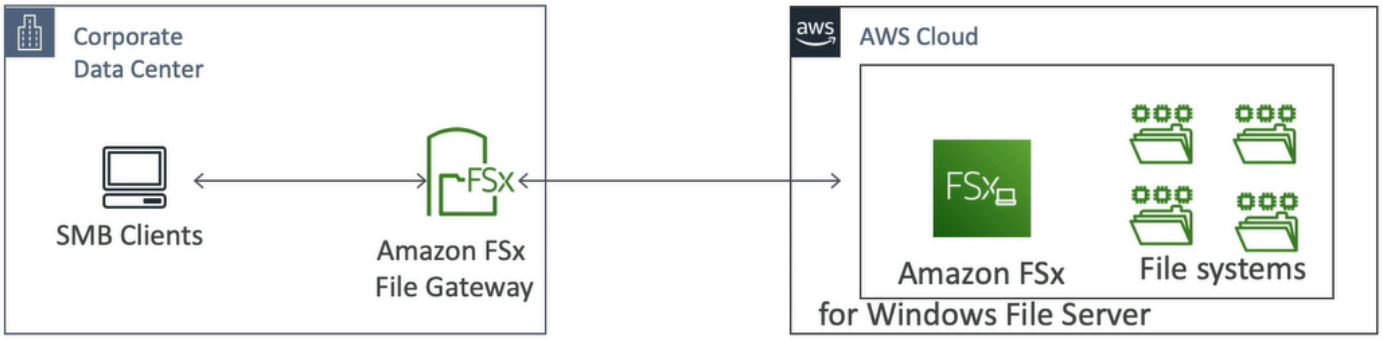
148. Key points about Amazon FSx File Gateway.

Ans ->

Amazon FSx File Gateway provides a way to access Amazon FSx for Windows File Server file shared from on-premises environments. It allows you to seamlessly integrate on-premises applications with FSx file systems, making it easy to use AWS cloud storage as part of your existing workflows.

Key Features:

- Native access to Amazon FSx for Windows File Server
- Local cache fore frequently accessed data
- Windows native compatibility (SMB, NTFS, Active Directory...)
- Useful for group file shares and home directories



149. Key points about Volume Gateway.

Ans ->

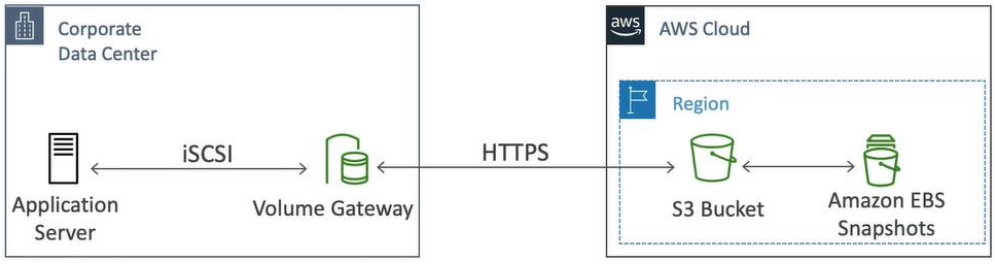
Volume Gateway provides block storage to on-premises applications using ISCSI protocol, while securely storing data in the AWS cloud. Backed by EBS snapshots which can help restore on-premises volumes. It offers two modes: Cached Volumes and Stored Volumes.

Cached Volumes:

- Frequently accessed data is cached locally for low latency access
- Primary data is stored in Amazon S3, reducing on-premises storage requirements.
- Provides scalable, cost-effective cloud storage with local performance.

Stored Volumes:

- Entire dataset is stored on-premises, with asynchronous backups to Amazon S3.
- Ensures low-latency access to all data while providing durable off-site backups.
- Ideal for environments needing fast local access to complete datasets.



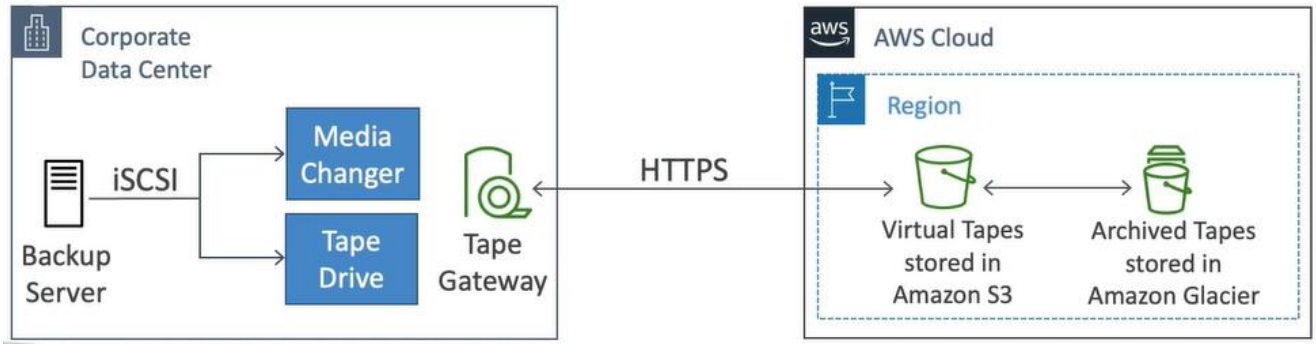
150. Key points about Tape Gateway.

Ans ->

The Tape Gateway allows you to use AWS cloud storage for backup and archival processes, emulating traditional tape-based storage. It integrates with existing backup applications, presenting cloud storage as Virtual Tape Libraries (VLTs).

Key features:

- Some companies have backup processes using physical tapes (!)
- With Tape Gateway, companies use the same processes but, in the cloud
- Virtual Tape Library backed by Amazon S3 and Glacier
- Back up data using existing tape-based processes (and ISCSI interface)
- Works with leading backup software vendors



151. Explain about Storage Gateway - Hardware Appliance

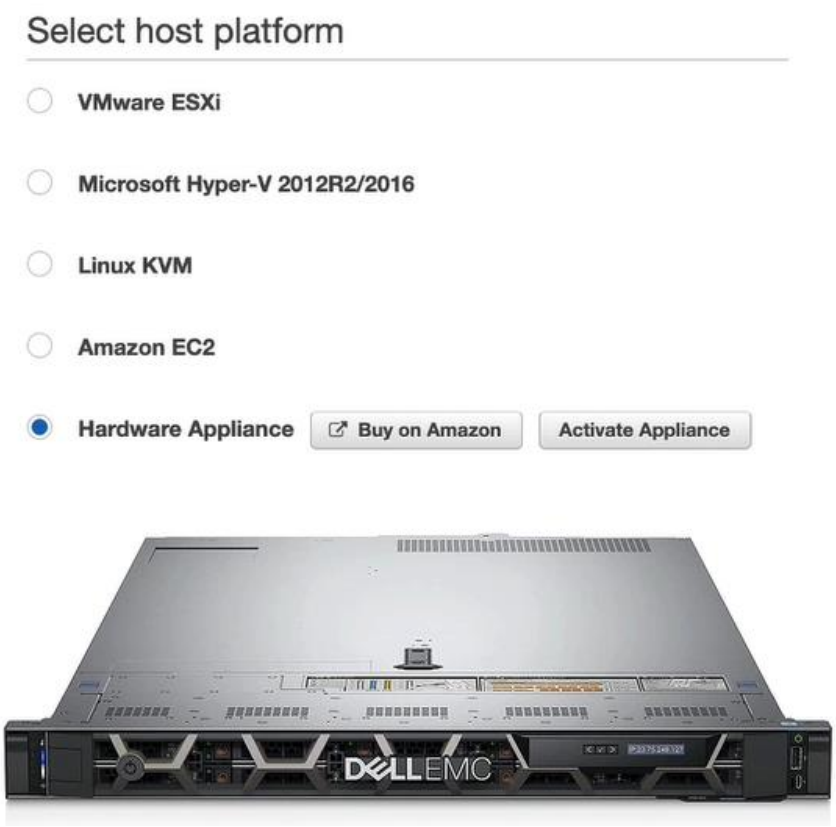
Ans ->

The Storage Gateway - Hardware Appliance is a physical device provided by AWS that integrates seamlessly with AWS Storage Gateway services. It offers a plug-and-play solution for organizations that need to connect their on-premises environments to AWS cloud storage without needing to set up their own hardware or virtual machines.

Key features:

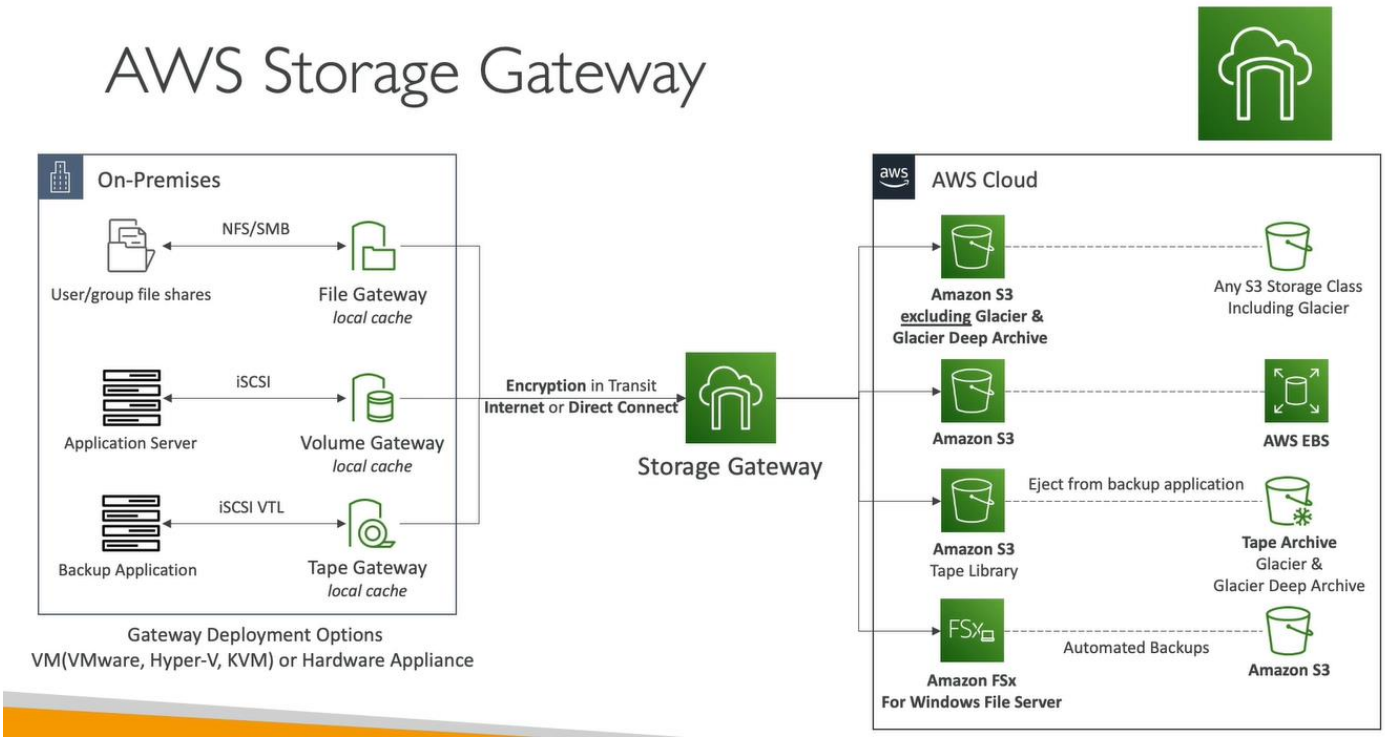
- Works with File Gateway, Volume Gateway, Tape Gateway

- Has the required CPU, memory, network, SSD cache resources
- Managed through the AWS Management Console, providing a unified interface for monitoring and configuring storage resources.
- Helpful for daily NFS backups in small data centers
- You can buy it from amazon.com



152. AWS Storage Gateway overall diagram.

Ans ->

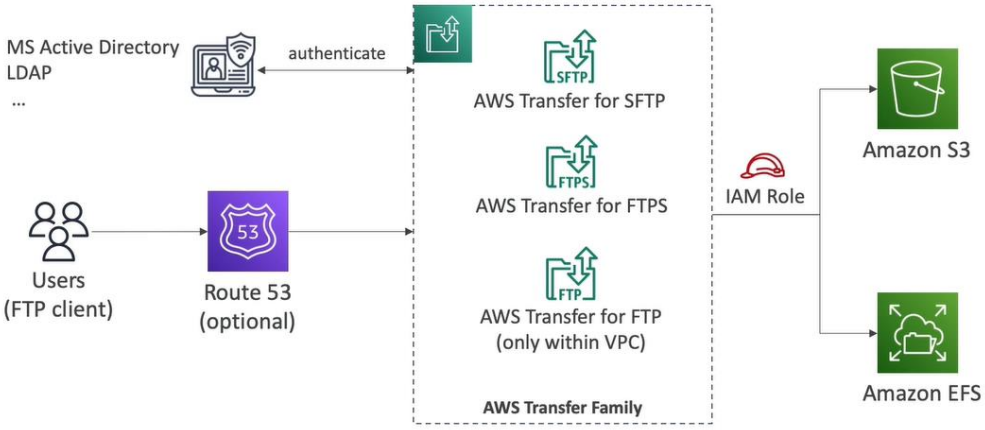


153. What is Amazon Transfer Family?

Ans ->

The AWS Transfer Family is a suite of fully managed services that enable the transfer of files into and out of AWS storage services using standard file transfer protocols.

- **Supported Protocols:**
  - AWS Transfer for **FTP** (File Transfer Protocol)
  - AWS Transfer for **FTPS** (File Transfer Protocol over SSL)
  - AWS Transfer for **SFTP** (Secure File Transfer Protocol)
- Managed infrastructure, Scalable, Reliable, Highly Available (multi-AZ)
- Pay per provisioned endpoint per hour + data transfers in GB
- Store and manage user's credentials within the service
- Integrated with existing authentication systems (Microsoft Active Directory, LDAP, Okta, Amazon Cognito, custom)
- Usage: Sharing files, public datasets, CRM, ERP, ...



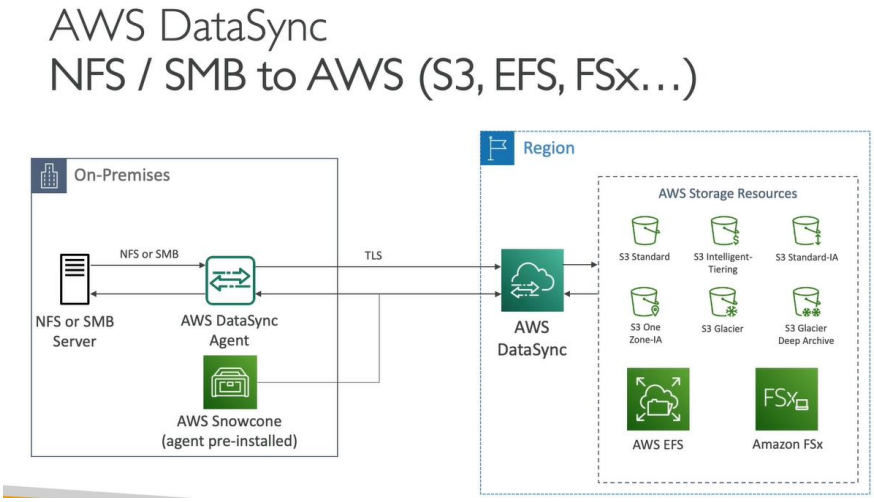
154. What is AWS DataSync?

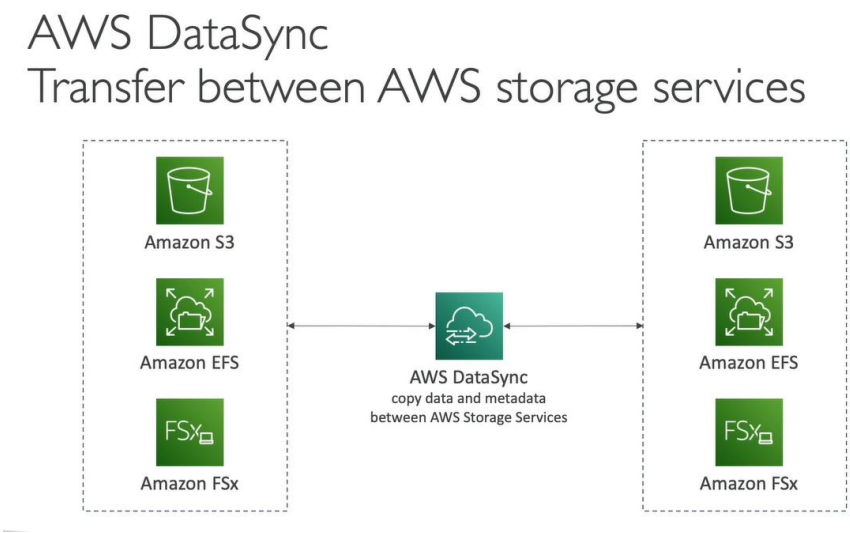
Ans ->

AWS DataSync simplifies and accelerates the process of transferring data to and from AWS, making it an ideal solution for data migrations, backups, disaster recovery, and hybrid cloud storage scenarios.

Key features:

- Move large amount of data to and from:
  - On-premises / other cloud to AWS (NFS, SMB, HDFS, S3, API...) - needs agent
  - AWS to AWS (different storage services) - no agent needed
- Can synchronize to:
  - Amazon S3 (any storage class - including Glacier)
  - Amazon EFS
  - Amazon FSx (Windows, Lustre, NetApp, OpenZFS...)
- Replication tasks can be scheduled hourly, daily, weekly
- File permissions and metadata are preserved (NFS, POSIX, SMB...)
- One agent task can use 10 Gbps, you can setup a bandwidth limit





155. Quick comparison between different AWS Storage services.

Ans ->

- **S3:** Object Storage
- **S3 Glacier:** Object Archival
- **EBS volumes:** Network storage for one EC2 instance at a time
- **Instance Storage:** Physical storage for your EC2 instance (high IOPS)
- **EFS:** Network File System for Linux instances, POSIX file system
- **FSx for Windows:** Network File System for Windows servers
- **FSx for Lustre:** High Performance Computing Linux file system
- **FSx for NetApp ONTAP:** High OS Compatibility
- **FSx for OpenZFS:** Managed ZFS file system
- **Storage Gateway:** S3 & FSx File Gateway, Volume Gateway (cache & stored), Tape Gateway
- **Transfer Family:** FTP, FTPS, SFTP interface on top of Amazon S3 or Amazon EFS
- **DataSync:** Scheduled data sync from on-premises to AWS, or AWS to AWS
- **Snowcone/Snowball/Snowmobile:** to move large amount of data to the cloud, physically
- **Database:** for specific workloads, usually with indexing and querying

156. What is Amazon SQS - Standard Queue?

Ans ->

Amazon SQS Standard Queue is a fully managed message queuing service that enables decoupling and scaling of microservices, distributed systems, and serverless applications.

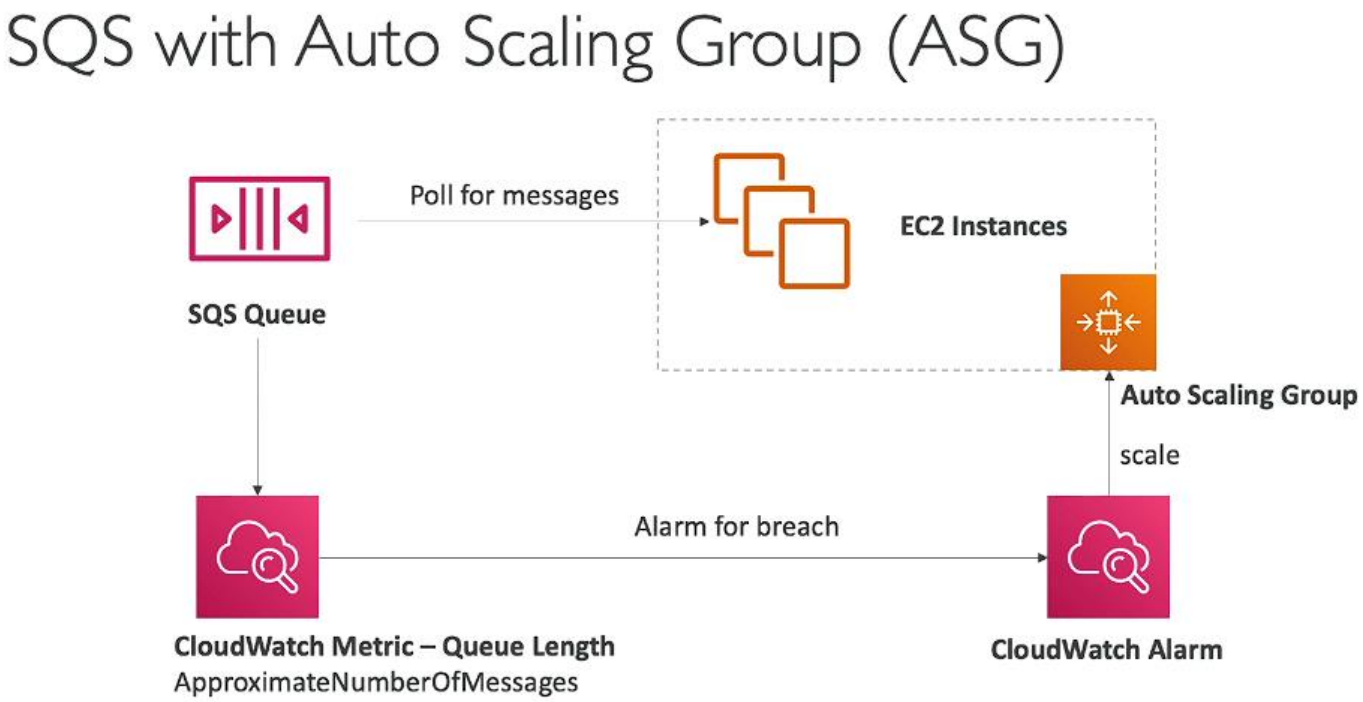
Key features:

- **Unlimited Throughput:** Supports a nearly unlimited number of transactions per second.
- **Default retention of messages:** minimum 4 days, maximum of 14 days
- Low latency (<10ms on publish and receive)
- Limitation of 256KB per message sent
- **At-Least-Once Delivery:** Ensures each message is delivered at least once but may deliver more than once.
- **Best-Effort Ordering:** Messages are generally delivered in the order sent but can have out of order messages.
- **Decoupling:** Enhances scalability and reliability by decoupling producing and consuming components.
- **Easy Integration:** Works seamlessly with Lambda, ECS, SNS, and S3.
- **Flexible Processing:** Supports both single and multiple consumer patterns.

157. SQS with Auto Scaling Group (ASG).

Ans ->

Using Amazon SQS with an Auto Scaling Group allows your application to dynamically adjust the number of EC2 instances based on the queue's message volume, ensuring efficient processing and cost management.



158. SQS to decouple between application tiers.

Ans ->

Using Amazon SQS to decouple applications tiers improves scalability, reliability, and fault tolerance.

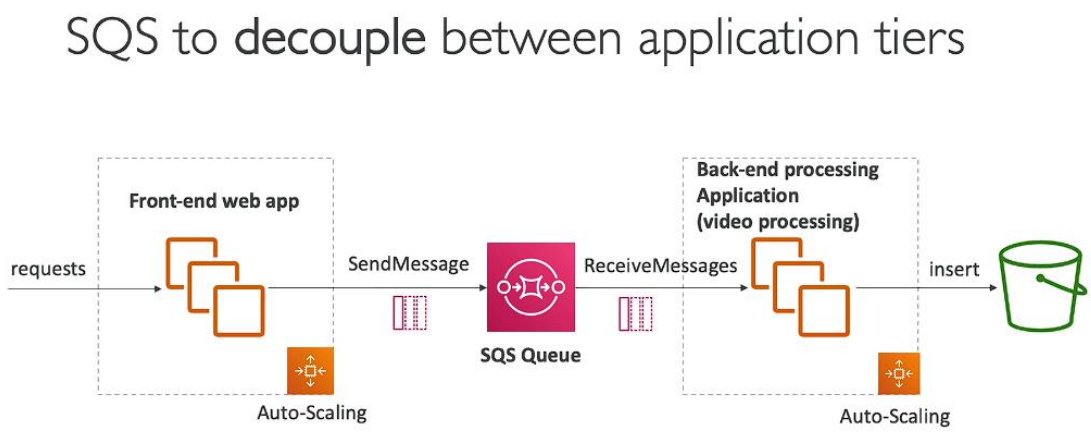
Key Steps to Decouple Application Tiers with SQS:

Identify Tiers: Determine the tiers in your application that need decoupling, such as front -end, back-end, and database layers.

Create SQS Queues: Set up SQS queues to act as buffers between these tiers.

Send Messages: The producer tier (e.g., front-end) sends messages to the SQS queue.

Receive Messages: The consumer tier (e.g., back-end) retrieves and processes messages from the queue and insert it in a S3 bucket.



159. Key points about Amazon SQS - Security.

Ans ->

- **Encryption:**
  - In-flight encryption using HTTPS API
  - At-rest encryption using KMS keys
  - Client-side encryption if the client wants to perform encryption/decryption itself
- **Access Controls:** IAM policies to regulate access to the SQS API
- **SQS Access Policies (similar to S3 bucket policies):**
  - Useful for cross-account access to SQS queues
  - Useful for allowing other services (SNS, S3…) to write to an SQS queue

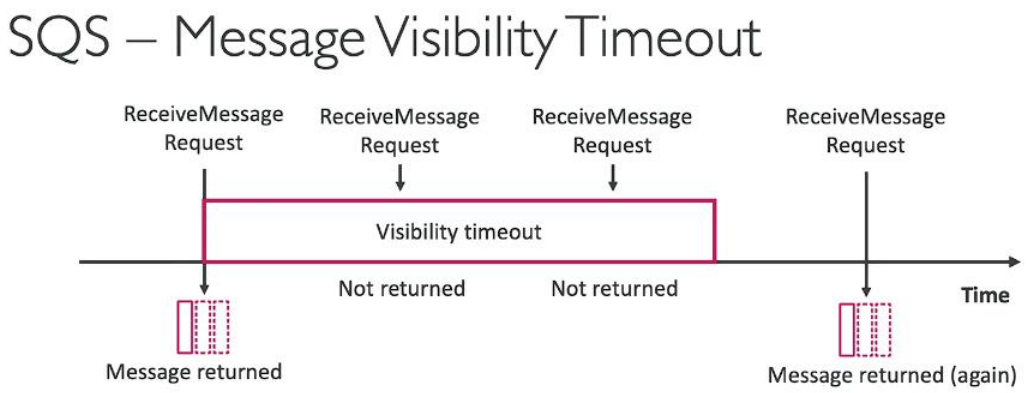
160. Key points about SQS - Message Visibility Timeout.

Ans ->

The **SQS Message Visibility Timeout** is a setting that determines how long a message remains invisible to other consumers once it has been retrieved from the queue by a

consumer. This timeout period ensures that only the consumer processes a message at a time and helps prevent message duplication.

- After a message is polled by a consumer, it becomes invisible to other consumers
- By default, the "message visibility timeout" is 30 seconds (can be adjusted up to 12 hours)
- That means the message has 30 seconds to be processed
- After the message visibility timeout is over, the message is "visible" in SQS

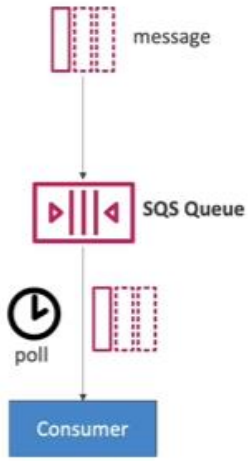


- If a message is not processed within the visibility timeout, it will be processed twice
- A consumer could call the ChangeMessageVisibility API to get more time
- If visibility timeout is high (hours), and consumer crashes, re-processing will take time
- If visibility timeout is too low (seconds), we may get duplicates

161. Key points about Amazon SQS - Long Polling.

Ans ->

- When a consumer requests messages from the queue, it can optionally "wait" for messages to arrive if there are none in the queue
- This is called Long Polling
- Long Polling decreases the number of API calls made to SQS while increasing the efficiency and latency of your application.
- The wait time can be between 1 sec to 20 sec (20 sec is preferable).
- Long Polling is preferable to Short Polling
- Long Polling can be enabled at the queue level or at the API level using WaitTimeSeconds.



162. A way to optimize the number of API calls and decrease the latency.

Ans -> Amazon SQS - Long Polling

163. Key points about Amazon SQS - FIFO Queue.

Ans ->

Amazon SQS - FIFO (First In First Out) Queues are suitable for applications that require strict messages ordering and exactly-once processing, such as financial transactions, order processing, and task management systems.

- **Ordered Processing:** Ensures messages are processed in the order they are sent.
- **Exactly-Once Delivery:** Guarantees each message is delivered exactly once.
- **Attributes:** Supports MessagesGroupId and MessageDeduplicationId for message grouping and deduplications control.
- **Suffix:** Identified by the ".fifo" suffix in the queue name.
- **Limited throughput:** With batching 3000 msg/s and without batching 300 msg/s.

164. SQS as a buffer to database writes.

Ans ->

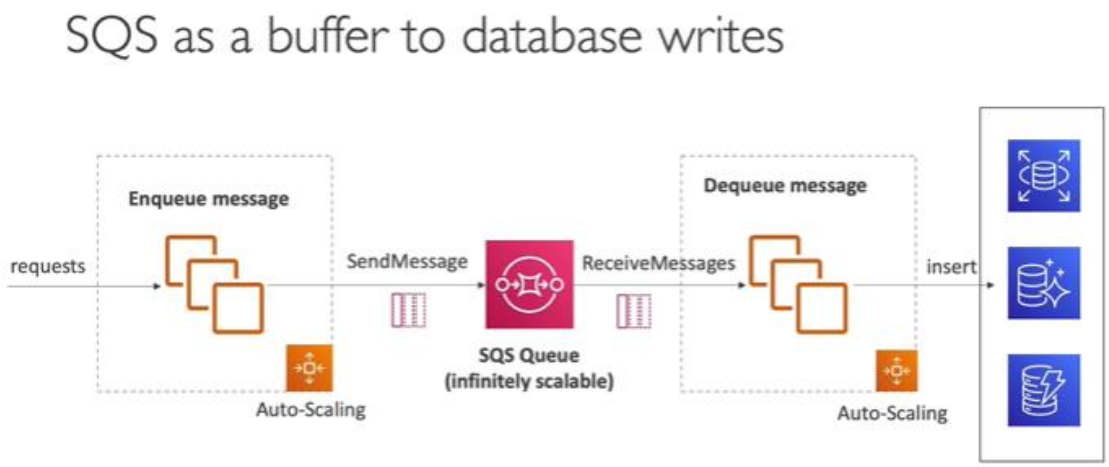
Using Amazon SQS as a buffer for database writes means:

Buffer: It holds write requests temporarily before they go to the database.

Smooths Out: It helps handle sudden increases in requests without overwhelming the database.

Safety Net: Protects data integrity by queuing writes during database downtime.

Efficiency: Allows processing multiple writes at once, improving how the database handles tasks.



165. AWS SNS - How to publish?

Ans ->

Topic Publish (using the SDK)

- Create a topic
- Create a subscription (or many)
- Publish to the topic

Direct Publish (for mobile apps SDK)

- Create a platform application
- Create a platform endpoint
- Publish to the platform endpoint
- Works with Google GCM, Apple APNS, Amazon ADM...

166. Key points about Amazon SNS - Security.

Ans ->

Encryption:

- In-flight encryption using HTTPS API

- At-rest encryption using KMS keys
- Client-side encryption if the client wants to perform encryption/decryption itself.

Access Controls:

- IAM Policies to regulate access to the SNS API

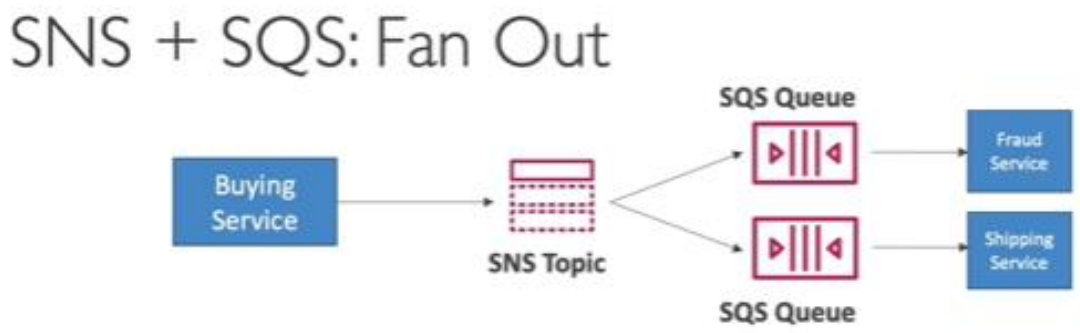
SNS Access Policies (similar to S3 bucket policies):

- Useful for cross-account access to SNS topics
- Useful for allowing other services (S3...) to write to an SNS topic

167. Explain about SNS + SQS: Fan Out pattern.

Ans ->

The **SNS + SQS: Fan Out pattern** is a setup, where a SNS topic is created and multiple SQS queues are subscribed to this topic. When a message is published to the SNS topic, it is automatically sent to all the subscribed SQS queues, allowing for parallel processing by different consumers. This pattern enhances scalability, decouples producers and consumers, and ensures reliable message distribution.



- Push once in SNS, receive in all SQS queues that are subscribers
- Fully decoupled, no data loss
- SQS allows for: data persistence, delayed processing and retires of work Ability to add more SQS subscribers over time
- Make sure your SQS queue access policy allows for SNS to write

168. Explain about SNS + SQS: Fan Out pattern with FIFO Topic.

Ans ->

The Fan Out pattern with FIFO Topic ensures ordered and exactly-once message delivery.

key points:

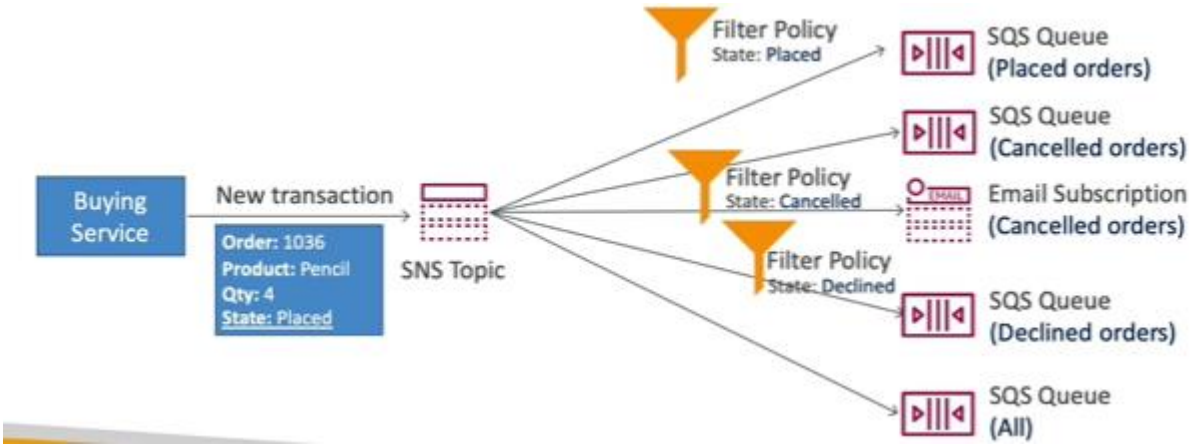
- **Similar features as SQS FIFO:**
  - Ordering by Message Group ID (all messages in the same group are ordered)
  - Deduplication using a Deduplication ID or Content Based Deduplication
- Can have SQS Standard and FIFO queues as subscribers
- Limited throughput (same throughput as SQS FIFO)

169. Key Points about SNS - Message Filtering.

Ans ->

Amazon SNS message filtering allows you to selectively route messages to specific subscribers based on messages attributes. This feature helps reduce the number or messages each subscriber receives enabling more efficient processing.

- JSON policy used to filter messages sent to SNS topic's subscriptions
- If subscription doesn't have a filter policy, it receives every message



170. Explain about Amazon Kinesis.

Ans ->

Amazon Kinesis is a suite of services designed to collect, process, and analyze real-time, streaming data. It enables you to handle large amounts of data from various sources, such as website clickstreams, database event streams, social media feeds, IT logs, and more.

Key Components:

- **Kinesis Data Stream:**
  - Purpose: Capture, process, and store data streams in real-time.
  - Use Cases: Real-time analytics, machine learning, log and event data collection.
- **Kinesis Data Firehose:**
  - Purpose: Load streaming data into AWS data stores.
  - Destinations: Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, Splunk.
- **Kinesis Data Analytics:**
  - Purpose: Analyze streaming data using SQL or Apache Flink.
  - Use Cases: Real-time dashboards, anomaly detection, streaming ETL.
- **Kinesis Video Streams:**
  - Purpose: Stream video data to AWS for analytics machine learning, and other processing.
  - Use Cases: Security monitoring, video analytics, IoT application.

171. Explain about Kinesis Data Streams.

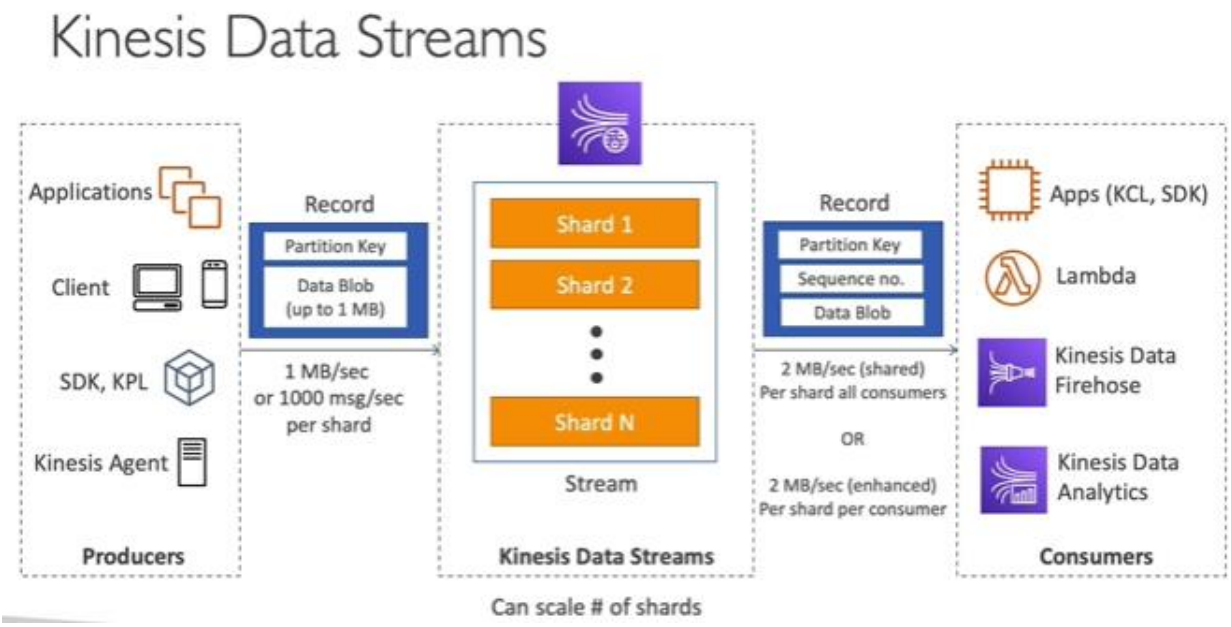
Ans ->

Amazon Kinesis Data Streams is a scalable, real-time service for collecting, processing, and storing large streams of data. It is designed to handle continuous data ingestion and is ideal for applications that require low-latency access to streaming data.

Key features:

- Retention between 1 day to 365 days
- Ability to reprocess (reply) data
- Once data is inserted in Kinesis, it can't be deleted (immutability)
- Data that shares the same partition goes to the same shard (ordering)
- **Producers:** AWS SDK, Kinesis Producers Library (KPL), Kinesis Agent

- **Consumers:**
  - Write your own: Kinesis Client Library (KCL), AWS SDK
  - Managed: AWS Lambda, Kinesis Data Firehose, Kinesis Data Analytics
- **Scalability:** Scales by adding shards, each supporting 1 MB/sec write and 2 MB/sec read throughput



172. Key points about Kinesis Data Streams - Capacity Modes.

Ans ->

Provisioned mode:

- Requires manually specifying the number of shards provisioned to handle data throughput, scale manually or using API
- Each shard gets 1 MB/sec in (or 1000 records per second)
- Each shard gets 2 MB/sec out (classic or enhanced fan-out consumer)
- You pay per shard provisioned per hour

On-demand mode:

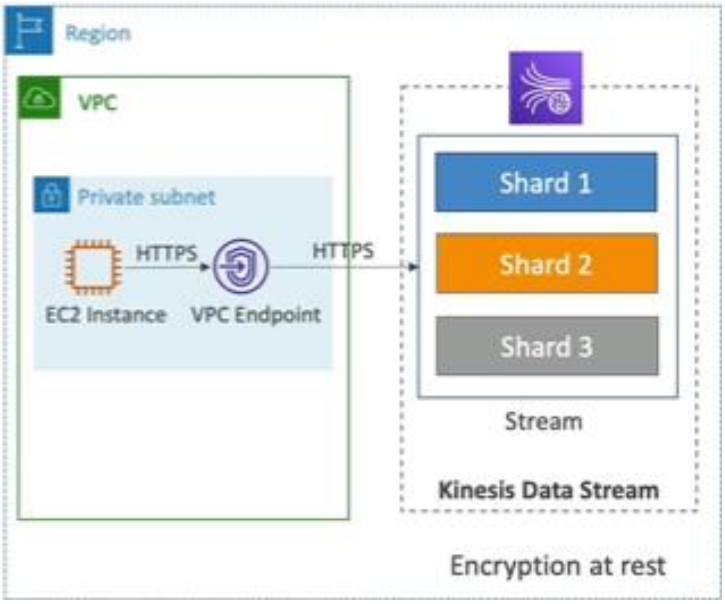
- No need to provision or manage the capacity
- Default capacity provisioned (4 MB/s in or 4000 records per second)
- Scales automatically based on observed throughput peak during the last 30 days
- Pay per stream per hour & data in/out per GB

173. Key points about Kinesis Data Streams Security.

Ans ->

- Control access/authorization using IAM policies

- Encryption in flight using HTTPS endpoints
- Encryption at rest using KMS
- You can implement encryption/decryption of data on client side (harder)
- VPC Endpoints available for Kinesis to access within VPC
- Monitor API calls using CloudTrail



174. Explain about Kinesis Data Firehose.

Ans ->

Amazon Kinesis Data Firehose is a fully managed service that allows reliable, **near real-time** delivery of streaming data into data lakes, data stores, and analytics services. It automates the process of ingesting, **transforming (if needed (with Lambda functions))**, and delivering streaming data with ease and efficiency. Kinesis Data Firehose handles scaling, monitoring, and error handling, making it simpler for developers to load streaming data into AWS without managing infrastructure.

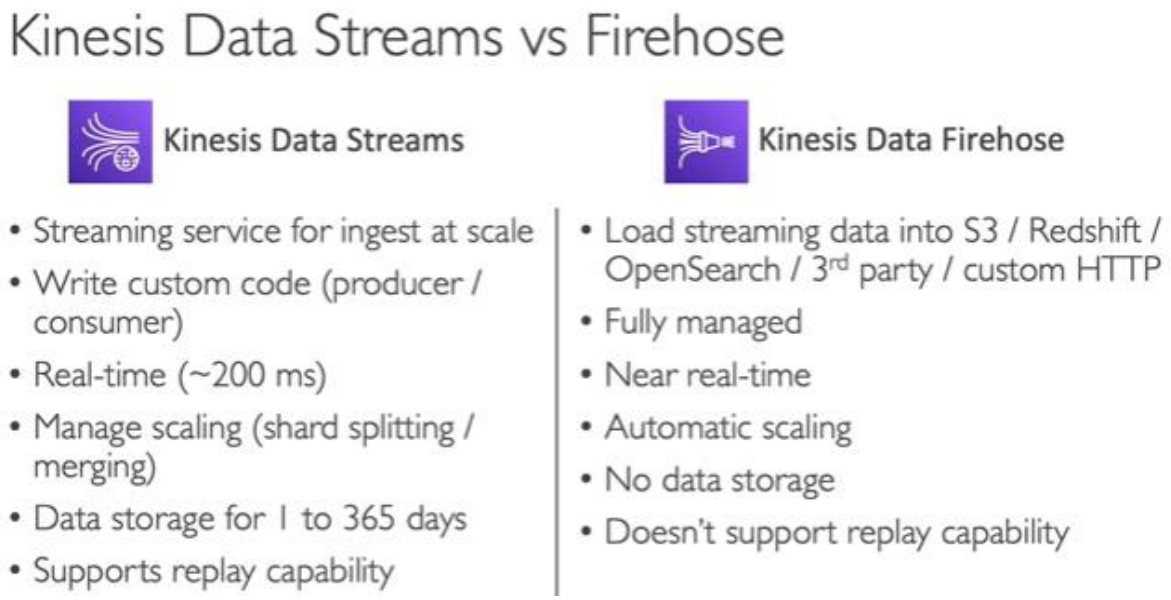
Key features:

- Fully Managed Services, no administration, automatic scaling, serverless
- You can send data to:
  - **AWS:** Redshift / Amazon S3 / OpenSearch
  - **3rd party partner:** Splunk / MongoDB / DataDog / NewRelic / ...
  - **Custom:** send to any HTTP endpoint
- Pay only for the data which is going through Firehose
- Near Real Time:
  - Buffer interval: 0 seconds (no buffering) to 900 seconds
  - Buffer size: minimum 1 MB
- Supports many data formats, conversions, transformations, compression
- Supports custom data transformations using AWS Lambda

- Can send failed or all data to a backup S3 bucket

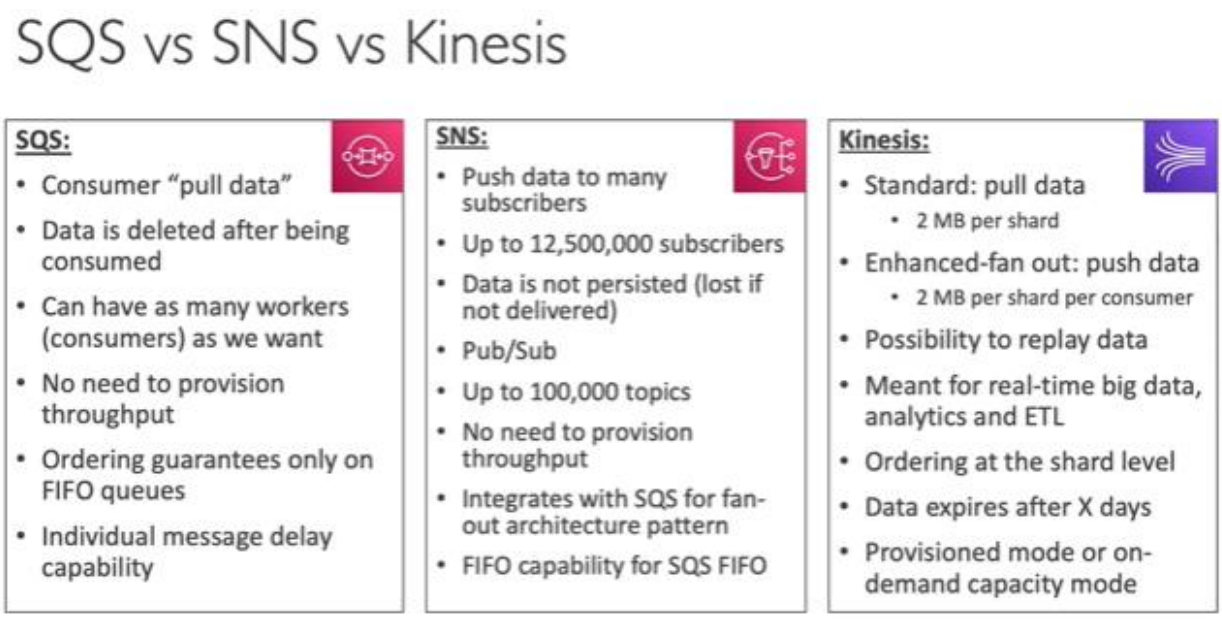
175. What is the difference between Kinesis Data Streams and Kinesis Data Firehose?

Ans ->



176. What are the differences between SQS vs SNS vs Kinesis?

Ans ->



177. Explain about ordering data into Kinesis.

Ans ->

When working with Amazon Kinesis Data Streams, ordering data is essential for applications that require processing records in a specific sequence.

Shards:

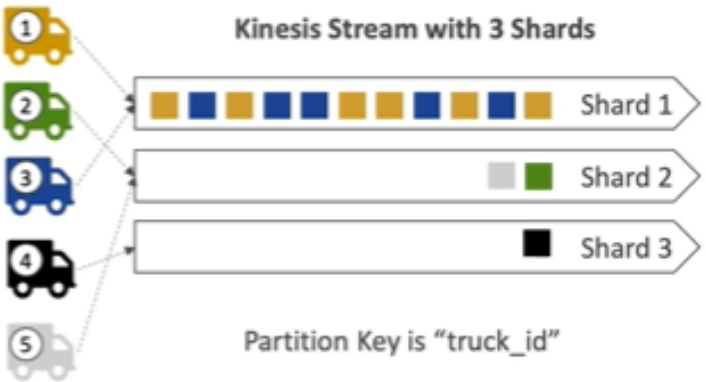
- A stream is divided into shards, each acting as a partition for data records.
- Data within a shard is ordered by the sequence in which records are added.

Partition Key:

- A partition key is used to determine which shard a data record is sent to.
- Records with the same partition key will always be directed to the same shard, preserving order within that shard.

Sequence Number:

- Each record within a shard is assigned a unique sequence number upon ingestion.
- Sequence numbers help maintain the order of records within a shard.

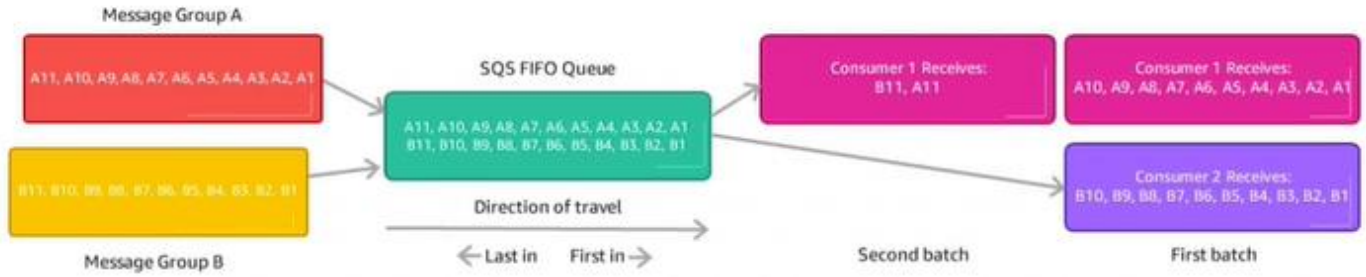


178. Explain about ordering data into SQS.

Ans ->

- For SQS standard, there is no ordering.
- For SQS FIFO, if you don't use a Group ID, message are consumed in the order they are sent, with only one consumer

- You want to scale the number of consumers, but you want messages to be "grouped" when they are related to each other
- Then you use a Group ID (similar to Partition Key in Kinesis)



179. What are the differences between Kinesis vs SQS data ordering?

Ans ->

Let's assume 100 trucks, 5 kinesis shards, 1 SQS FIFO

- **In the case of Kinesis Data Streams:**
  - On average you'll have 20 trucks per shard
  - Trucks will have their data ordered within each shard
  - The maximum number of consumers in parallel we can have is 5
  - Can receive up to 5 MB/s of data
- **In the case of SQS FIFO:**
  - You only have one SQS FIFO queue
  - You will have 100 Group ID
  - You can have up to 100 Consumers (due to the 100 Group ID)
  - You can have up to 300 messages per second (or 3000 if using batching)

180. Explain about Amazon MQ.

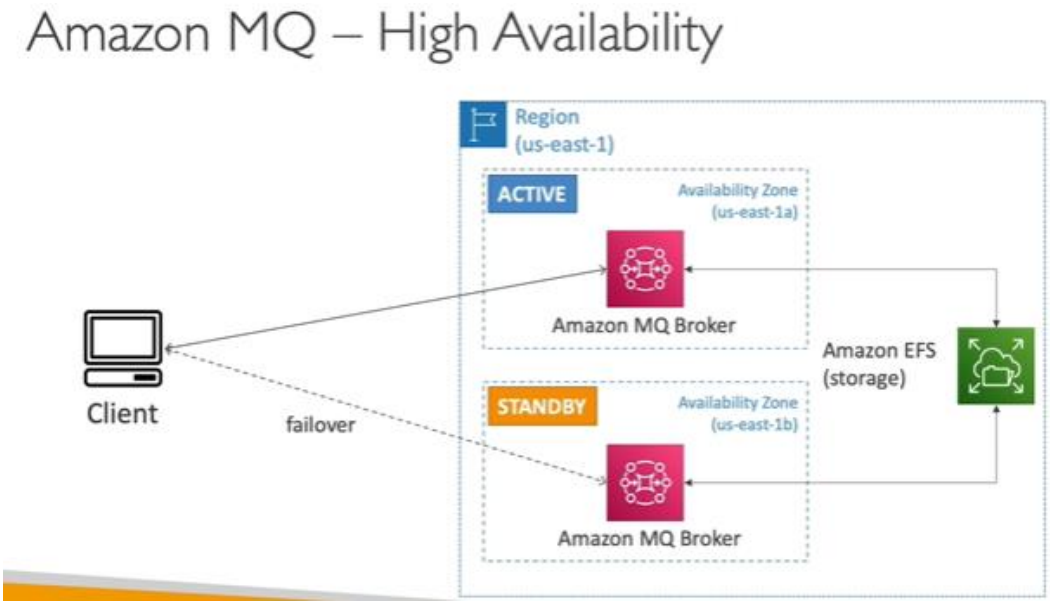
Ans ->

- SQS and SNS are "cloud-native" services: proprietary protocols from AWS
- Traditional applications running from on-premises may use open protocols such as: MQTT, AMQP, STOMP, Openwire, WSS

- When migrating to the cloud, instead of re-engineering the application to use SQS and SNS, we can use Amazon MQ.

Amazon MQ is a managed message broker service for **Apache ActiveMQ** and **RabbitMQ** that makes it easy to set up and operate message brokers in the cloud. It allows applications to communicate and exchange information using various messaging protocols, helping decouple and coordinate distributed systems.

- Amazon MQ doesn't "scale" as much as SQS/SNS
- Amazon MQ runs on servers, can run in Multi-AZ with failover
- Amazon MQ has both queue feature (~SQS) and topic features (~SNS).



181. What is Docker?

Ans ->

Docker is a tool that helps developers create, deploy, and run applications inside small, portable containers. These containers package everything the application needs to run including the code, libraries, and settings, so it works consistently across different environments.

Key points:

- Docker is a software development platform to deploy apps
- Apps are packaged in containers that can be run on any OS
- App run the same, regardless of where they're run
    - Any machine
    - No compatibility issue
    - Predictable behavior
    - Less work
    - Easier to maintain and deploy

        - Works with any language, any OS, any technology
    - Use cases: microservices architecture, lift-and-shift apps from on-premises to the AWS cloud

182. Where are Docker images stored?

Ans ->

Docker images are stored in Docker Repositories. The default repository is Docker Hub https://hub.docker.com, which is a public repository managed by Docker.

In AWS these images are stored, manage and deploy by a service called ECR (Elastic Container Registry). It provides both Private as well as Public Repository, for Public Repository it is Amazon ECR Public Gallery https://gallery.ecr.aws

183. Docker VS Virtual Machines.

Ans ->

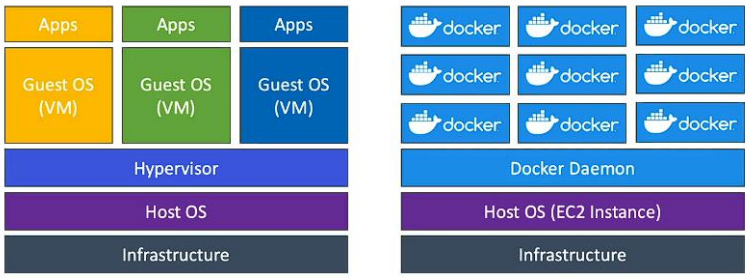
Docker and **virtual machines (VMs)** are both technologies for running applications in isolated environments, but they do so in different ways:

- Docker: is lightweight, shares the host system's kernel.
- VMs: is heavyweight, includes a full operating system.

- Docker: Containers start quickly.
- VMs: take longer to start.

- Docker: it has Process-level isolation means, runs on the same OS as the host.
- VMs: it has hardware-level isolation means, each VM acts as a separate machine.

- Docker: Uses fewer resources.
- VMs: Requires more resources due to full OS overhead.



184. Docker Containers Management on AWS.

Ans ->

Amazon Elastic Container Service (Amazon ECS):

- **Fully Managed:** Simplifies the process of running, stopping, and managing containers.
- **Integration:** Seamlessly integrates with other AWS services like IAM, ELB, and CloudWatch

Amazon Elastic Kubernetes Service (Amazon EKS):

- **Managed Kubernetes:** Provides a managed Kubernetes service to run and orchestrate Docker containers.
- **Flexibility:** Supports Kubernetes-native tools and APIs.

AWS Fargate

- **Serverless Containers:** Allows you to run containers without managing servers. Works with ECS and EKS
- **Cost-efficient:** Pay only for the resources used by your containers.

Amazon ECR:

- **Secure Storage:** Stores, manages, and deploys Docker container images securely.
- **Integration:** Integrates with ECS, EKS, and Fargate for seamless container deployment.

185. Explain about Amazon ECS - EC2 Launch Type.

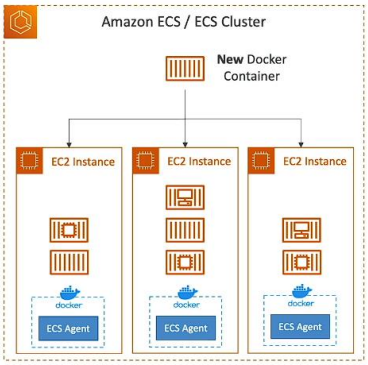
Ans ->

The **Amazon ECS (Elastic Container Service) EC2 launch type** allows you to run your Docker containers on a cluster of Amazon EC2 instances that you manage.

Key points:

- **Managed Instances:** You provision and manage the EC2 instances in your ECS cluster.
- **Control:** Provides more control over the underlying infrastructure, including instance types, scaling policies, and network settings.
- **Scaling:** You are responsible for scaling the EC2 instances based on the workload.
- **Cost:** Charges are based on the EC2 instances, storage, and other AWs resources you use.

Note: Each EC2 Instance must run the ECS Agent to register in the ECS Cluster



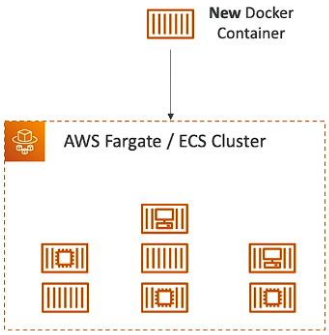
186. Explain about Amazon ECS - Fargate Launch Type.

Ans ->

The **Amazon ECS (Elastic Container Service) Fargate launch type** allows you to run Docker containers without managing the underlying infrastructure.

Key points:

- No EC2 instance management needed
- AWS just runs ECS Tasks for you based on the CPU/RAM you need
- To scale, just increase the number of tasks. Simple - no more EC2 instances
- Pay only for resources used
- Built-in isolation between tasks
- Works seamlessly with AWS services like VPC, IAM, and CloudWatch.



187. Explain about Amazon ECS - IAM Roles for ECS.

Ans ->

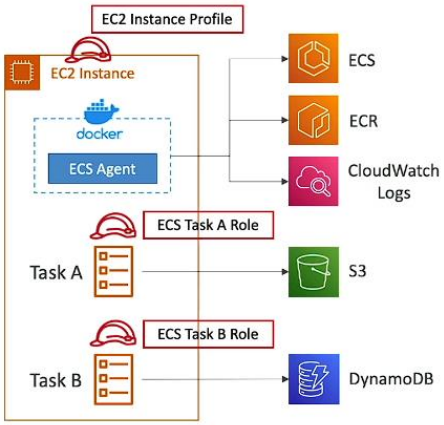
IAM roles for Amazon ECS allow you to control and manage permissions for ECS tasks and services.

EC2 Instance Profile (EC2 Launch Type only):

- Used by the ECS agent
- Makes API calls to ECS service
- Send container logs to CloudWatch Logs
- Pull Docker image from ECR
- Reference sensitive data in Secrets Manager or SSM Parameter Store

ECS Task Role:

- Allows each task to have a specific role Use different roles for the different ECS Services you run
- Task Role is defined in the task definition



188. Explain about Amazon ECS - Load Balancer Integration.

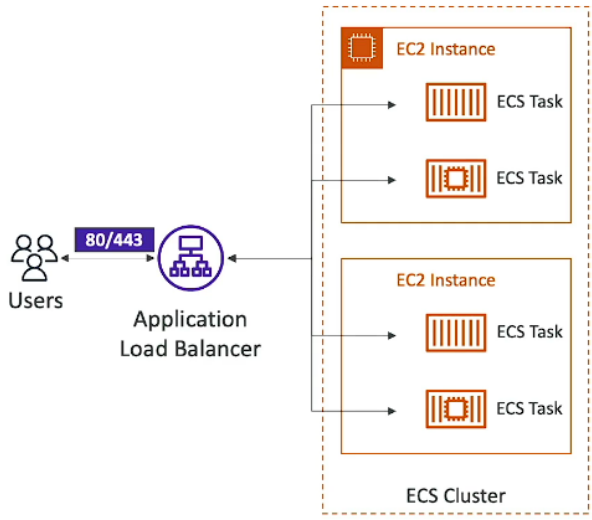
Ans ->

Integrating Amazon ECS with load balancer allows you to distributer incoming traffic across containers or services running in ECS.

- **Application Load balancer (ALB):** Routes traffic based on content of the request. Supported and works for most use cases.

- **Network Load Balancer (NLB):** Routes traffic based on IP protocol data. Recommended only for high throughput / high performance use cases, or to pair it with AWS Private Link.

- **Classic Load Balancer:** Older load balancer routing system supported but not recommended (no advanced features - no Fargate).



189. Explain about Amazon ECS - Data Volumes (EFS).

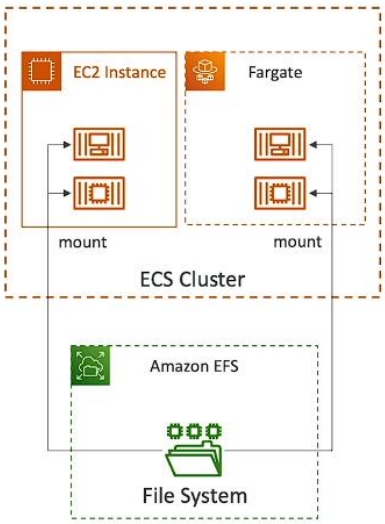
Ans ->

Amazon ECS supports data volumes through Amazon EFS, allowing containers to share persistent data across tasks and instances.

- **EFS Integration:** ECS can mount EFS file systems as volumes in container enabling multiple containers to access the same data simultaneously.
- **Flexibility:** Works for both EC2 and Fargate launch types
- **Persistent Storage:** Ensures data persists beyond the lifecycle of any single task or container instance.

Use cases: Persistent multi-AZ shared storage for your containers

Note: Amazon S3 cannot be mounted as a file system.



190. Key points about ECS Service Auto Scaling.

Ans ->

- Automatically increase/decrease the desired number of ECS tasks
- Amazon ECS Service Auto Scaling uses AWS Application Auto Scaling
    - ECS Service Average CPU Utilization
    - ECS Service Average Memory Utilization - Scale on RAM
    - ALB Request Count Per Target - metric coming from ALB

- **Target Tracking Scaling:** Adjusts the number of tasks to maintain a target metric (e.g., CloudWatch metric), such as CPU utilization or memory usage.
- **Step Scaling:** Increases or decreases the number of tasks in response to specified metric thresholds (e.g., specified CloudWatch Alarm).
- **Scheduled Scaling:** Scale based on a schedule date/time, accommodating predictable changes in demand.

- ECS Service Auto Scaling (task level) ≠ EC2 Auto Scaling (EC2 instance level)
- FarGate Auto Scaling is much easier to setup (because Serverless)

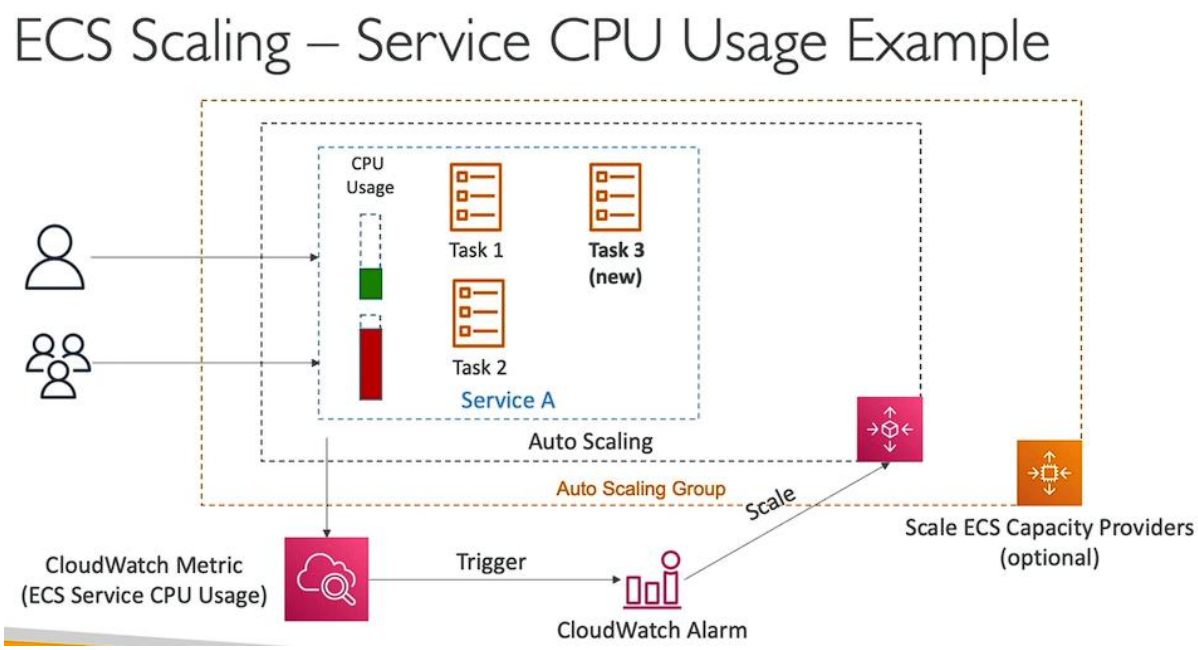
191. Key points about EC2 Launch Type - Auto Scaling EC2 Instances.

Ans ->

- Accommodate ECS Service Scaling by adding underlying EC2 Instances.
- **Auto Scaling Group Scaling:**
    - Scale your ASG based on CPU Utilization
    - Add EC2 instances over time
- **ECS Cluster Capacity Provider:**
    - Used to automatically provision and scale the infrastructure for your ECS Tasks
    - Capacity Provider paired with an Auto Scaling Group
    - Add EC2 Instances when you're missing capacity (CPU, RAM...)

192. ECS Scaling - Service CPU Usage Example.

Ans ->



193. Key points about Amazon ECR.

Ans ->

- ECR = Elastic Container Registry
- Store and manage Docker images on AWS
- Private and Public repository (Amazon ECR Public Gallery: https://gallery.ecr.aws)
- Fully integrated with ECS, backed by Amazon S3
- Access is controlled through IAM (permission errors => policy)
- Supports image vulnerability scanning, versioning, image tags, image lifecycle, ...

194. What is Amazon EventBridge?

Ans ->

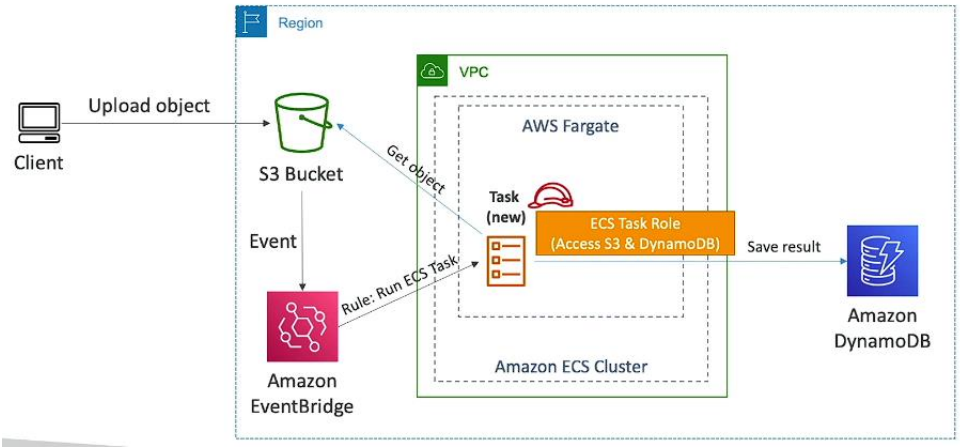
Amazon EventBridge is a serverless event bus that simplifies connecting applications and AWS services by routing data from various sources like AWS services, custom apps, or third-party SaaS apps.

You can set up rules to filter and route events, such as data changes or system state updates, to targets like AWS Lambda, SNS, SQS, and others. This enables you to create scalable, event-driven architectures that automatically respond to events in real-time, without the need to manage infrastructure.

EventBridge supports various event sources and allows you to build flexible and decoupled systems, improving application integration and responsiveness.

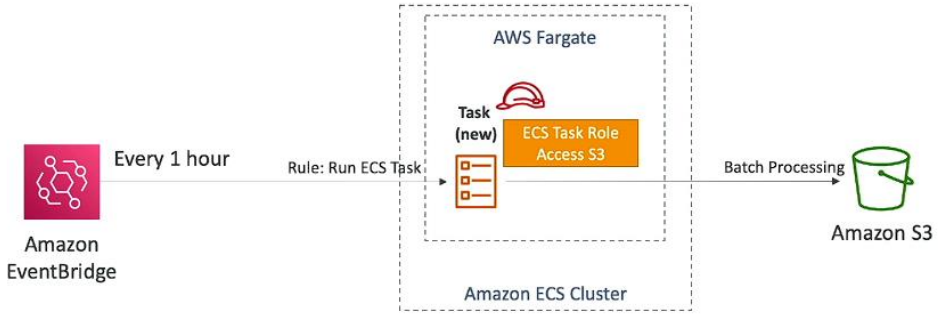
195. Solutions Architecture 1: ECS tasks invoked by EventBridge.

Ans->



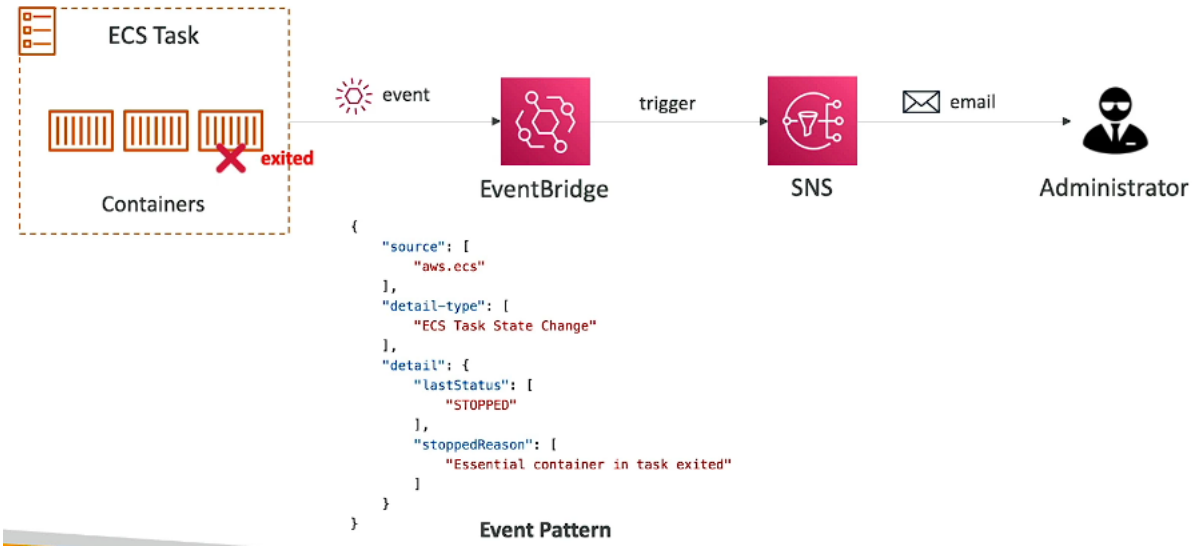
196. Solutions Architecture 2: ECS tasks invoked by EventBridge Schedule

Ans ->



197. Solutions Architecture 3: ECS - Intercept Stopped Tasks using EventBridge.

Ans ->

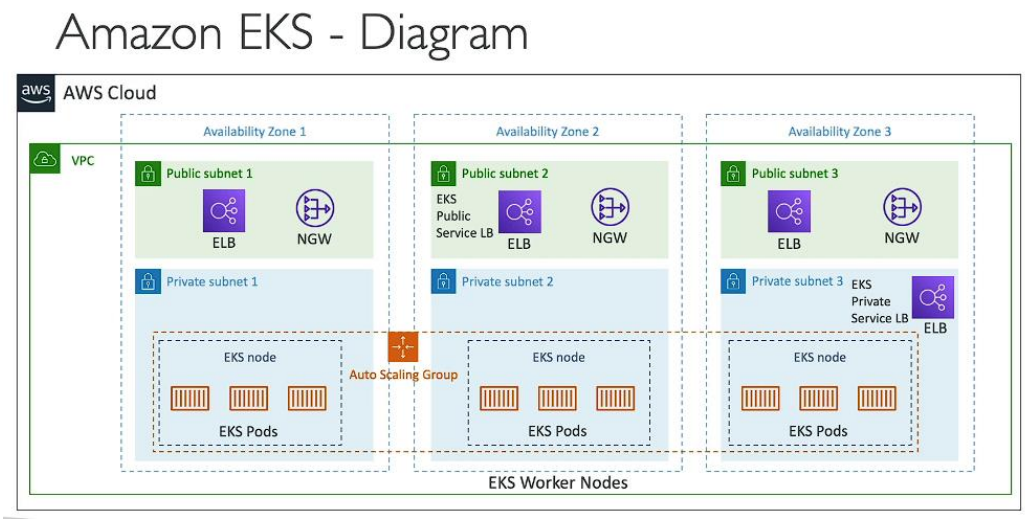


198. Key points about Amazon EKS.

Ans ->

Amazon EKS (Elastic Kubernetes Service) is a managed service that simplifies running Kubernetes on AWS, providing a secure and scalable control plane and seamless integration with AWS services. It enables you to deploy, manage, and scale containerized applications without the overhead of managing Kubernetes infrastructure, making it easier to build and run reliable, event-driven, and scalable applications.

- It is a way to launch managed Kubernetes clusters on AWS
- Kubernetes is an open-source system for automatic deployment, scaling and management of containerized (usually Docker) application
- It's an alternative to ECS, similar goal but different API
- EKS supports EC2 if you want to deploy worker nodes or Fargate to deploy serverless containers
- Use case: if your company is already using Kubernetes on-premises or in another cloud, and wants to migrate to AWS using Kubernetes.
- Kubernetes is cloud-agnostic (can be used in any cloud - Azure, GCP, ...)



199. Explain Amazon EKS - Node Types.

Ans ->

Amazon EKS (Elastic Kubernetes Service) supports different types of nodes to run your Kubernetes workloads. The two primary node types are **EC2 instances** and **AWS Fargate**.

EC2 Instances:

- **Managed Node Groups:**
  - EKS-managed node groups simplify the provisioning and lifecycle management of EC2 instances.
  - Automatically handle updates and scaling of nodes within the group.
  - Nodes are configured with Amazon Machine Images (AMIs) optimized for EKS.
  - Supports On-Demand or Spot Instances

- **Self-Managed Nodes:**
  - Requires you to manually handle provisioning, scaling, and updates.
  - Allows for customization of instance configurations and the use of custom AMIs.
  - Supports On-Demand or Spot Instances

AWS Fargate:

- Runs Kubernetes pods without managing the underlying EC2 instances.
- Automatically scales and manages the compute resources required by the pods.
- Pay only for the vCPU and memory resources consumed by the pods.

200. Explain about Amazon EKS - Data Volumes.

Ans ->

Amazon EKS (Elastic Kubernetes Service) supports various types of data volumes to manage persistent storage for your Kubernetes applications. These volumes can be used to store application data, configuration files, logs, and more, ensuring data persistence and durability across pod restarts and failures.

- Need to specify StorageClass manifest on your EKS cluster
- Leverages a Container Storage Interface (CSI) compliant driver

Support for:

- Amazon EBS
- Amazon EFS (works with Fragate)
- Amazon FSx for Lustre
- Amazon FSx for NetApp ONTAP

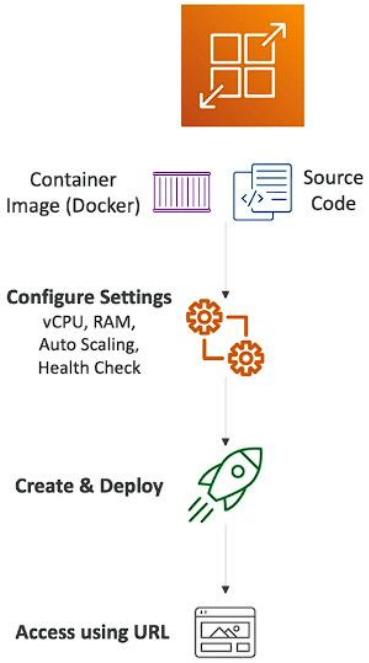
201. What is AWS App Runner?

Ans ->

AWS App Runner is a fully managed service designed to help developers quickly deploy and run containerized applications and APIs at scale without needing to manage infrastructure. App Runner abstracts away the complexities of container orchestration and infrastructure management, allowing developers to focus on writing code and delivering applications.

Key points:

- Fully managed service that makes it easy to deploy web applications and APIs at scale
- No infrastructure experience required
- Start with your source code or container image
- Automatically builds and deploy the web app
- It provides automatic scaling, highly availability, load balancer, encryption
- VPC access support
- Connect to database, cache, and message queue services
- Use cases: web apps, APIs, microservices, rapid production deployments.



202. What is Serverless?

Ans ->

- Serverless is a new paradigm in which the developers don't have to manage servers anymore...
- They just deploy code
- They just deploy functions
- Initially, Serverless == FaaS (Function as a Service)
- Serverless was pioneered by AWS Lambda but now also includes anything that's managed: "databases, messaging, storage, etc."
- Serverless does not mean there are no servers, it means you just don't manage / provision / see them.

203. What are the serverless services in AWS?

Ans ->

- AWS Lambda
- DynamoDB
- AWS Cognito
- AWS API Gateway
- Amazon S3
- AWS SNS & SQS
- AWS Kinesis Data Firehose
- Aurora Serverless

- Step Functions
- Fargate

204. What is AWS Lambda and what are the benefits of using it?

Ans ->

AWS Lambda is a serverless compute service that lets you run code without provisioning or managing servers. Lambda automatically scales applications by running code in response to event, such as changes to data in an Amazon S3 bucket or updates to a DynamoDB table.

Benefits of using Lambda:

- Easy Pricing:
  - Pay per request and compute time
  - Free tier of 1,000,000 (1M) AWS Lambda request and 400,000 GBs of compute time
- Integrated with the whole AWS suite of services
- Integrated with many programming languages
- Easy monitoring through AWS CloudWatch Easy to get more resources per functions (Up to 10GB of RAM)
- Increasing RAM will also improve CPU and network.

205. AWS Lambda language support.

Ans ->

AWS Lambda supports multiple programming languages, allowing developers to choose the best language for their specific use case. Additionally, Lambda supports Custom Runtime APIs, enabling the use of any additional programming languages.

Supported languages:

- Node.js (JavaScript)
- Python
- Java (Java 8 compatible)
- C# / PowerShell
- Golang
- Ruby
- Custom Runtime API (Community supported, example: Rust)
- **Lambda Container Image:**
  - The container image must implement the Lambda Runtime API
  - ECS / Fargate is preferred for running arbitrary Docker images

206. AWS Lambda Pricing: example.

Ans ->

You can find overall pricing information here:

https://aws.amazon.com/lambda/pricing/

Pay per calls:

- First 1,000,000 (1M) requests are free
- $0.20 per 1 million requests thereafter ($0.0000002 per request)

Pay per duration: (in increment of 1 ms)

- 400,000 GB-seconds of compute time per month FREE
- == 400,000 seconds if function is 1 GB RAM
- == 3,200,000 seconds if function is 128 MB RAM
- After that $1.00 for 600,000 GB-seconds

It is usually very cheap to run AWS Lambda so it's very popular

207. Key points about AWS Lambda Limits to Know - per region.

Ans ->

Execution:

- Memory allocation: 128 MB - 10 GB (1 MB increments)
- Maximum execution time: 900 seconds (15 minutes)
- Environment variables (4 KB)
- Disk capacity in the "function container" (in /tmp): 512 MB to 10 GB
- Concurrency execution: 1000 (can be increased)

Deployment:

- Lambda function deployment size (compressed .zip): 50 MB
- Size of uncompressed deployment (code + dependencies): 250 MB
- Can use the /tmp directory to load other files at startup

- Size of environment variables: 4 KB

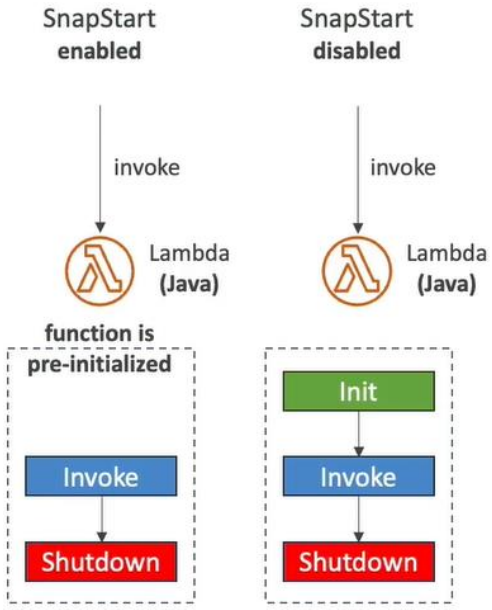
208. What is AWS Lambda SnapStart?

Ans ->

AWS Lambda SnapStart is a feature designed to reduce the cold start latency of java functions. It does this by initializing the function's execution environment ahead of time, taking a snapshot of the initialized environment, and then using this snapshot to quickly start new instances of the function.

Key points:

- Improves your Lambda functions performance up to **10x faster at no extra cost for Java 11 and above**
- When enabled, function is invoked from a pre-initialized state (no function initialization from scratch)
- **When you publish a new version:**
  - lambda initializes your function
  - Takes a snapshot of memory and disk state of the initialized function
  - Snapshot is cached for low-latency access.



Lambda Invocation Lifecycle Phases

209. What is CloudFront Functions?

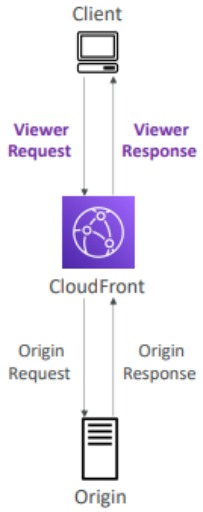
Ans ->

AWS CloudFront Functions allows you to run **lightweight JavaScript code** at CloudFront edge locations to customize and optimize the delivery of your content. It's

ideal for tasks like HTTP header manipulation, URL rewriting, access control, and A/B testing, providing improved performance, scalability, and cost-effectiveness for enhancing user experiences.

Key points:

- Lightweight functions written in JavaScript
- For high-scale, latency-sensitive CDN customizations
- Sub-ms startup times, millions of requests/seconds
- **Used to change Viewer requests and responses:**
  - **Viewer Request:** after CloudFront receives a request from a viewer
  - **Viewer Response:** before CloudFront forwards the response to the viewer
- Native feature of CloudFront (manage code entirely within CloudFront)



210. What is Lambda@Edge?

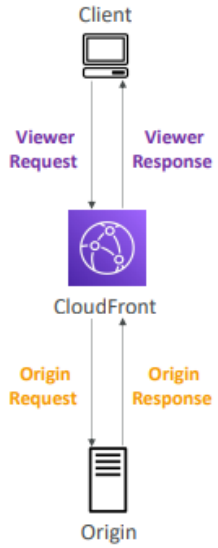
Ans ->

AWS Lambda@Edge is a feature of Amazon CloudFront that allows you to run serverless functions at CloudFront edge locations worldwide. It lets you customize and optimize content delivery by executing code in response to various CloudFront events, improving performance, reducing latency, and providing a flexible, scalable solution for enhancing user experiences.

Key points:

- Lambda functions written in NodeJS or Python
- Scales to 1000s of requests/second
- **Used to change CloudFront requests and responses:**
  - **Viewer Request -** after CloudFront receives a request from a viewer
  - **Origin Request -** before CloudFront forward the request to the origin
  - **Origin Response -** after CloudFront receives the response from the origin
  - **Viewer Response -** before CloudFront forwards the response to the viewer

- Author your functions in one AWS Region (us-east-1), then CloudFront replicates to its locations



211. What are the differences between CloudFront Functions vs. Lambda@Edge?

And ->

CloudFront Functions vs. Lambda@Edge

	CloudFront Functions	Lambda@Edge
Runtime Support	JavaScript	Node.js, Python
# of Requests	**Millions** of requests per second	**Thousands** of requests per second
CloudFront Triggers	- Viewer Request/Response	- Viewer Request/Response - Origin Request/Response
Max. Execution Time	< 1 ms	5 – 10 seconds
Max. Memory	2 MB	128 MB up to 10 GB
Total Package Size	10 KB	1 MB – 50 MB
Network Access, File System Access	No	Yes
Access to the Request Body	No	Yes
Pricing	Free tier available, 1/6th price of @Edge	No free tier, charged per request & duration

212. Explain about CloudFront Functions vs. Lambda@Edge - Use Cases

Ans ->

CloudFront Functions use cases:

- Cache key normalization
  - Transform request attributes (header, cookies, query strings, URL) to create an optimal Cache Key

- Header manipulation
    - Insert/modify/delete HTTP headers in the request or response
- URL rewrites or redirects
- Request authentication & authorization
    - Create and validate user-generated tokens (e.g., JWT) to allow/deny requests.

Lambda@Edge use cases:

- Longer execution time (several ms)
- Adjustable CPU or memory
- Your code depends on a 3rd libraries (e.g., AWS SDK to access other AWS services)
- Network access to use external services for processing
- File system access or access to the body of HTTP requests

213. Combined use cases of CloudFront Functions & Lambda@Edge.

Ans ->

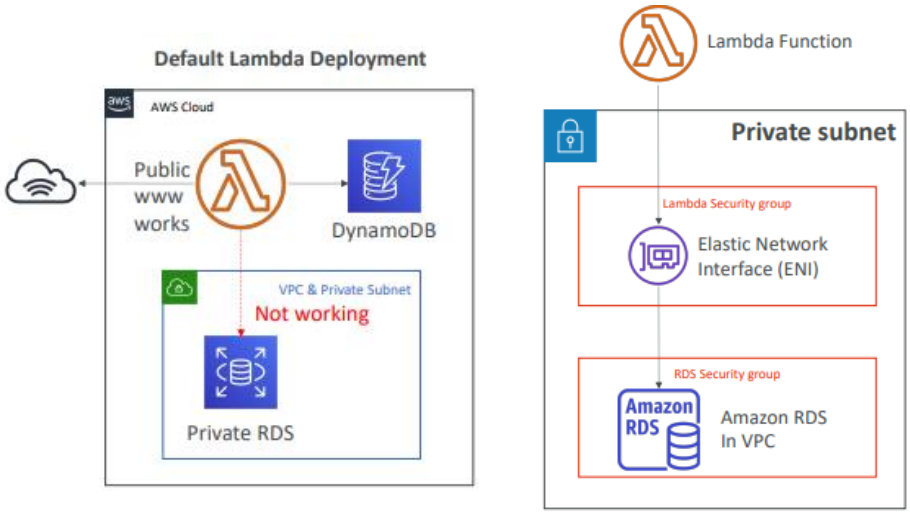
- Website Security and Privacy
- Dynamic Web Application at the Edge
- Search Engine Optimization (SEO)
- Intelligently Route Across Origins and Data Centers
- Bot Mitigation at the Edge
- Real-time Image Transformation
- A/B Testing
- User Authentication and Authorization
- User Prioritization
- User Tracking and Analytics

214. Lambda Default behavior with VPC.

Ans ->

- By default, your Lambda function is launched outside your own VPC (in an AWS-owned VPC).
- Therefore, it cannot access resources in your VPC (RDS, ElastiCache, internal ELB...).

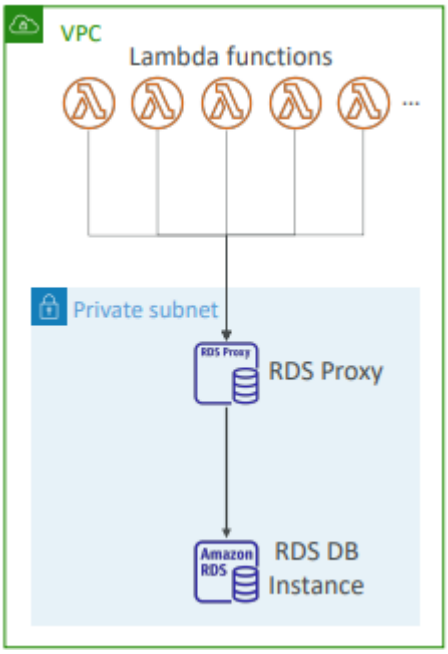
- However, if you need your Lambda function to access resources within your VPC, you can configure it to connect to your specific VPC by defining the VPC ID, the Subnets and the Security Groups.
- Lambda will create an ENI (Elastic Network Interface) in your subnets.



215. Lambda with RDS Proxy.

Ans ->

- If lambda functions directly access your database, they may open too many connections under high load.
- To fix this issue, Lambda can connect to Amazon RDS databases using RDS Proxy, to efficiently manage database connections, reducing connection overhead, improving performance, and enhancing scalability and availability.
- **RDS Proxy:**
  - Improve scalability by pooling and sharing DB connections.
  - Improve availability by reducing the failover time 66% and preserving connections
  - Improve security by enforcing IAM authentication and storing credentials in Secrets Manager
- The Lambda function must be deployed in your VPC, because RDS Proxy is never publicly accessible.



216. Invoking Lambda from RDS & Aurora.

Ans ->

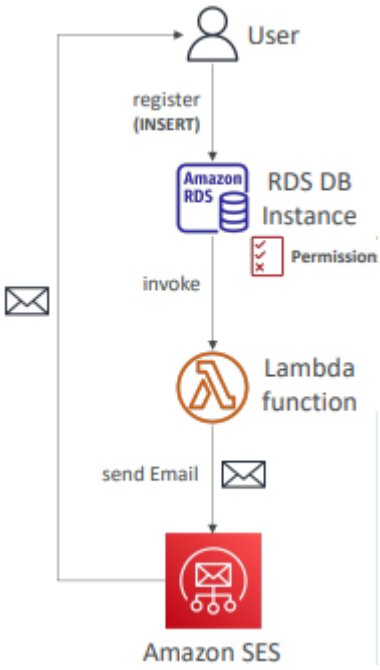
Amazon RDS and Aurora can invoke **AWS Lambda functions** in response to database events, enabling real-time processing, notifications, and automated workflows. This is achieved through database triggers and requires an IAM role with the necessary permissions to invoke the Lambda function.

- Invoke Lambda functions from within your DB instance
- Allows you to process **data events** from within a database
- Supported for **RDS for PostgreSQL** and **Aurora MySQL**
- **Must allow outbound traffic to your Lambda function** from within you DB instance (Public, NAT GW, VPC Endpoints)
- **DB instance must have the required permissions to invoke the Lambda function** (Lambda Resource-based Policy and IAM Policy)

Real-Time Data Processing: Automatically process or transform data in real-time as changes occur in the database.

Notifications: Send notifications or alerts based on specific database changes.

ETL Processes: Trigger extract, transform, and load (ETL) operations automatically.



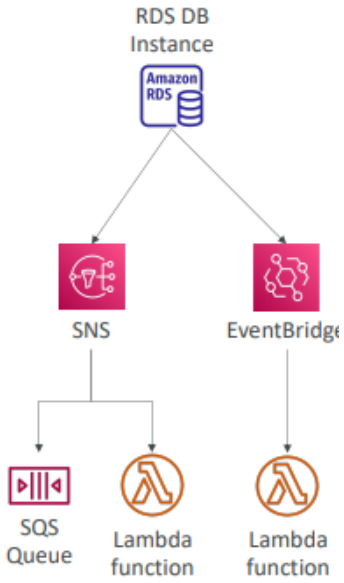
217. Explain about RDS Event Notifications.

Ans ->

Amazon RDS Event Notification allows you to receive notifications about important events related to your RDS instances, snapshots, parameter groups, and security groups. Amazon RDS Event Notifications can be sent via **Amazon SNS**, allowing you to receive notifications through email, SMS, or other endpoints. For more advanced event handling, you can integrate RDS with **Amazon EventBridge**.

Key points:

- Notifications that tells information about the DB instance itself (created, stopped, start, ...)
- You don't have any information about the data itself
- Subscribe to the following event categories: **DB instance, DB snapshot, DB Parameter Group, DB Security Group, RDS Proxy, Custom Engine Version**
- **Near real-time** events (up to 5 minutes)
- Send notifications to SNS or subscribe to events using EventBridge



218. What is Amazon DynamoDB?

Ans ->

Amazon DynamoDB is a fully managed, fast, and scalable NoSQL database service that handles large volumes of data with low-latency performance. It offers automatic scaling, built-in security, and additional features like global tables and DynamoDB Streams, making it suitable for a wide range of applications.

Key points:

- Fully managed, highly available with replication across multiple AZs
- NoSQL database - not a relational database - with transaction support
- Scales to massive workloads, distributed database
- Millions of requests per seconds, trillions of rows, 100s of TB of storage
- Fast and consistent in performance **(single-digit millisecond)**
- Integrated with IAM for security, authorization and administration
- Low cost and auto-scaling capabilities
- No maintenance or patching, always available
- Standard & Infrequent Access (IA) Table Class

219. Some more key points about DynamoDB.

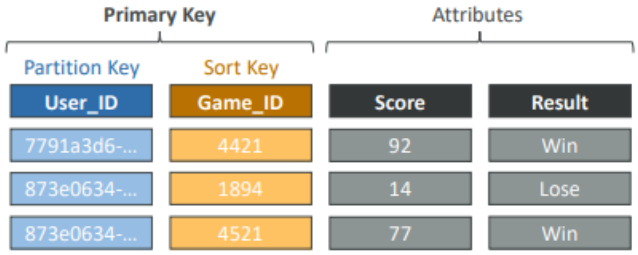
Ans ->

- DynamoDB is made of **Tables**
- Each table has a **Primary key** (must be decided at creation time)
- Each table can have an infinite number of items (=rows)
- Each item has **attributes** (can be added over time or can be null)
- Maximum size of an **item is 400KB**
- Data types supported are:

- **Scalar Types -** String, Number, Binary, Boolean, Null
- **Document Types -** List Map
- **Set Types -** String Set, Number Set, Binary Set

- **Therefore, in DynamoDB you can rapidly evolve schemas**

DynamoDB – Table example



220. Explain about DynamoDB - Read/Write Capacity Modes.

Ans ->

DynamoDB offers two capacity modes: **Provisioned Mode** for predictable workloads with pre-allocated RCUs and WCUs, and **On-Demand Mode** for unpredictable workloads with pay-per-request pricing. Provisioned Mode supports auto-scaling, while On-Demand Mode provides flexibility and scalability without capacity planning.

Key points:

- **Provisioned Mode (default):**
  - You specify the number of reads/writes per second
  - You need to plan capacity beforehand
  - Pay for **provisioned** Read Capacity Units (RCU) & Write Capacity Units (WCU)
  - Possibility to add auto-scaling mode for RCU & WCU
- **On-Demand Mode:**
  - Read/writes automatically scale up/down with your workloads
  - No capacity planning needed
  - Pay for what you use (more expensive)
  - Great for **unpredictable** workloads, **steep sudden spikes**

221. What is DynamoDB Accelerator (DAX)?

Ans ->

DynamoDB Accelerator (DAX) is an in-memory caching service designed to improve the performance of DynamoDB queries by reducing read latency from milliseconds to microseconds.

Key points:

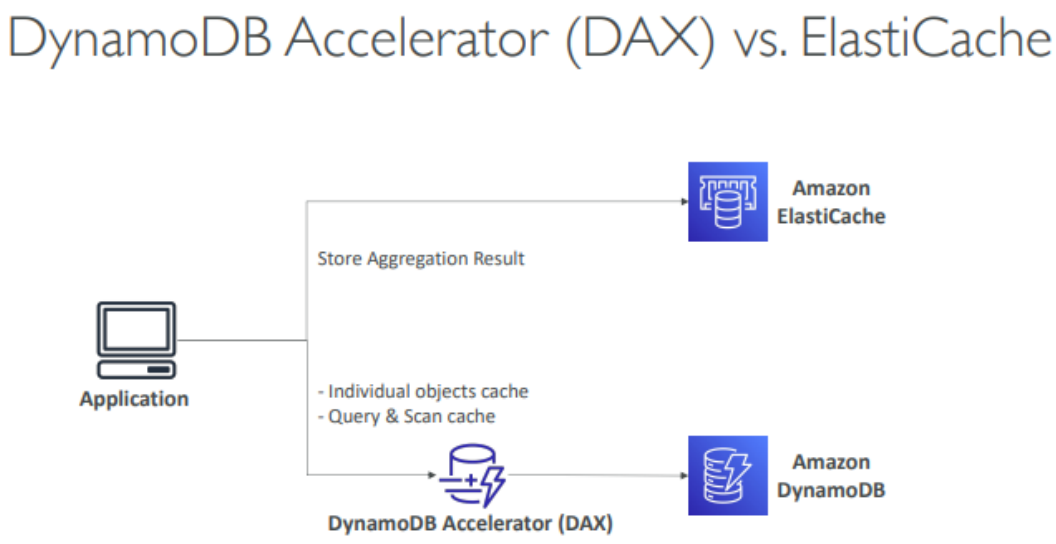
- Fully-managed, highly available, seamless in-memory cache for DynamoDB
- **Help solve read congestion by caching**
- **Microseconds latency for cached data**
- Doesn't require application logic modification (compatible with existing DynamoDB APIs)
- 5 minutes TTL for cache (default)

222. Why should you use DAX over ElastiCache?

Ans ->

Use **DynamoDB Accelerator (DAX)** if you need fast, seamless caching solution specifically for DynamoDB that requires minimal changes to your code and automatically handles data updates, it can store individual objects cache and also it can query and scan cache.

Choose **ElastiCache** for a flexible cache that works with multiple databases, supports advanced features like pub/sub messaging, complex data structures, and is ideal for broader use cases, it can store aggregation result.



223. Explain about DynamoDB - Stream Processing.

Ans ->

DynamoDB Stream Processing captures real-time changes to items (create/update/delete) in a table and can be used for analytics, data replication, and event-driven architectures. Streams integrate seamlessly with AWS Lambda for

automatic, serverless processing, and support various view types to capture the data needed.

Use cases:

- React to changes in real-time (welcome email to users)
- Real-time usage analytics
- Insert into derivative tables
- Implement cross-region replication
- Invoke AWS Lambda on changes to your DynamoDB table

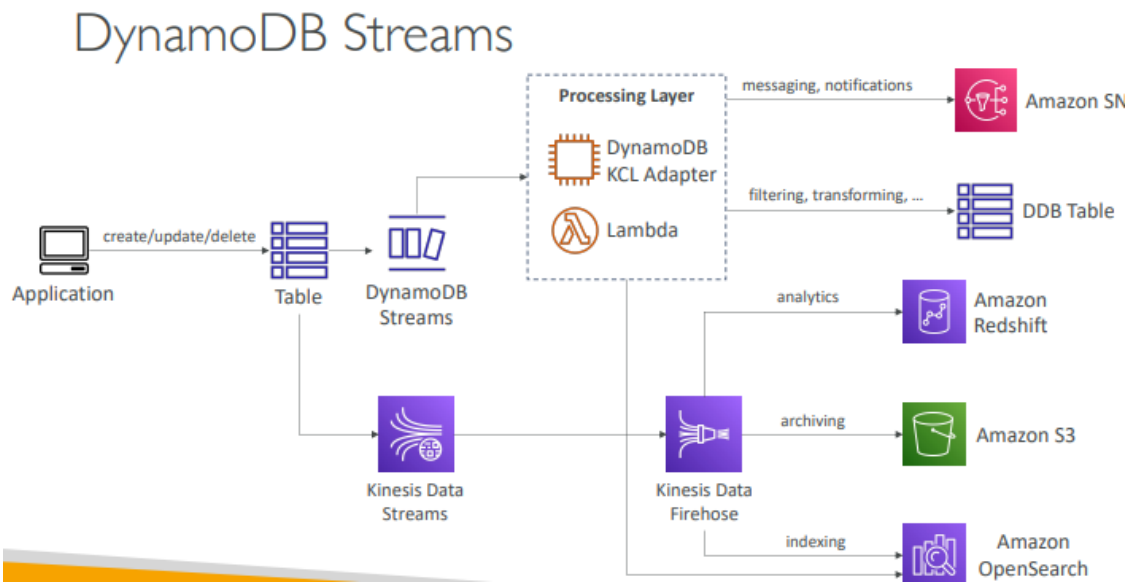
It supports 2 types of streaming:

1. DynamoDB Streams:

- 24 hours retention
- Limited # of consumers
- Process using AWS Lambda Triggers, or DynamoDB Stream Kinesis adapter

2. Kinesis Data Streams (newer):

- 1 year retention
- High # of consumers
- Process using AWS Lambda, Kinesis Data Analytics, Kinesis Data Firehose, AWS Glue Streaming ETL



224. Explain about DynamoDB Global Tables.

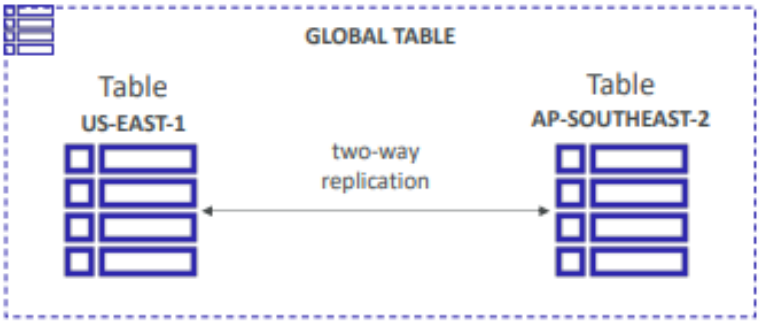
Ans ->

DynamoDB Global Tables provide multi-region replication for your tables, ensuring low-latency access, high availability, and fault tolerance. They support various

consistency models and are easy to set up and manage, making them ideal for globally distributed applications.

Key points:

- Make a DynamoDB table accessible with **low latency** in multiple-regions
- Active-Active replication
- Application can **READ** and **WRITE** to the table in any region
- **Must enable DynamoDB Streams as a pre-requisite**



225. Explain about DynamoDB - Backups for Disaster Recovery.

Ans ->

Continuous backups using point-in-time recovery (PITR):

- Optionally enabled for the last **35 days**
- Point-in-time recovery to any time within the backup window
- The recovery process creates a new table

On-demand backups:

- Full backups for long-term retention, until explicitly deleted
- Doesn't affected performance or latency
- Can be configured and managed in AWS Backup (enables cross-region copy)
- The recovery process creates a new table

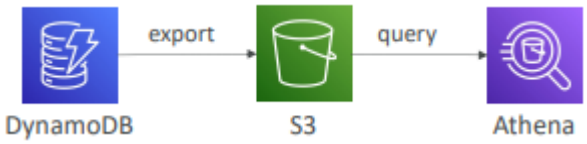
226. DynamoDB - Integration with Amazon S3

Ans ->

Export to S3 (must enable PITR):

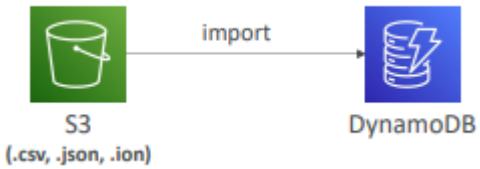
- Works for any point of time in the last 35 days
- Doesn't affect the read capacity of your table
- Perform data analysis on top of DynamoDB

- Retain snapshots for auditing
- ETL on top of S3 data before importing back into DynamoDB
- Export in DynamoDB **JSON** or **ION** format



Import from S3:

- Import **CSV**, DynamoDB **JSON** or **ION** format
- Doesn't consume any write capacity
- Creates a new table
- Import errors are logged in CloudWatch Logs



227. What is AWS API Gateway?

Ans ->

AWS API Gateway allows you to build and manage APIs easily. For example, if you run an online store and want to connect your mobile app to your product database, API Gateway can create an API to securely access and manage product data, handle large numbers of requests, and monitor performance. This makes it simpler to manage and maintain your app's connectivity to backend services.

Key points:

- AWS Lambda + API Gateway: No infrastructure to manage
- Support for the **WebSocket** Protocol
- Handle API **versioning** (v1, v2, ...)
- Handle different **environments** (dev, test, prod...)
- Handle **security** (Authentication and Authorization)
- Create API keys, handle request throttling
- Swagger / Open API import to quickly define APIs
- Transform and validate requests and responses

- Generate SDK and API specifications
- Cache API responses

228. Explain about AWS API Gateway - High Level Integration.

Ans ->

Lambda Function:

- Invoke Lambda function
- Easy way to expose REST API backed by AWS Lambda

HTTP:

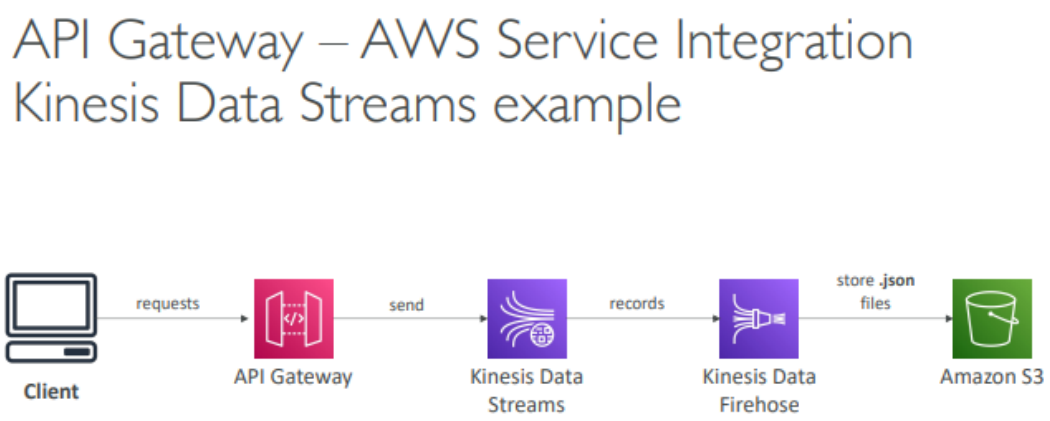
- Expose HTTP endpoints in the backend
- Example: internal HTTP API on premise, Application Load Balancer...
- Why to use? Ans -> Add rate limiting, caching, user authentications, API keys, etc...

AWS Service:

- Expose any AWS API through the API Gateway
- Example: start an AWS Step Function workflow, post a message to SQS
- Why to use? Ans -> Add authentication, deploy publicly, rate control...

229. API Gateway - AWS Service Integration Kinesis Data Streams example diagram.

Ans ->



230. Explain about AWS API Gateway - Endpoint Types.

Ans ->

Edge-Optimized (default):

- Best suited for global clients.
- Routes requests through CloudFront Edge locations to improve latency.
- The API Gateway is hosted in a single region.

Regional:

- Ideal for clients within the same region.
- Allows manual integration with CloudFront, offering more control over caching strategies and distribution.

Private:

- Accessible only within your VPC through an interface VPC endpoint (ENI).
- Access is controlled by a resource policy.

231. Explain about AWS API Gateway - Security.

Ans ->

User Authentication via:

- IAM Roles, which are beneficial for internal applications.
- Cognito, providing identity management for external users, such as mobile app users.
- Custom Authorizer, allowing you to implement your own authorization logic.

HTTPS security for Custom Domain Names is ensured through AWS Certificate Manager (ACM) integration:

- For **Edge-Optimized** endpoints, the certificate should be in the us-east-1 region.
- For **Regional endpoints**, the certificate should be in the same region as the API Gateway.
- Additionally, you must configure a CNAME or A-alias record in Route 53.

232. What is AWS Step Functions?

Ans ->

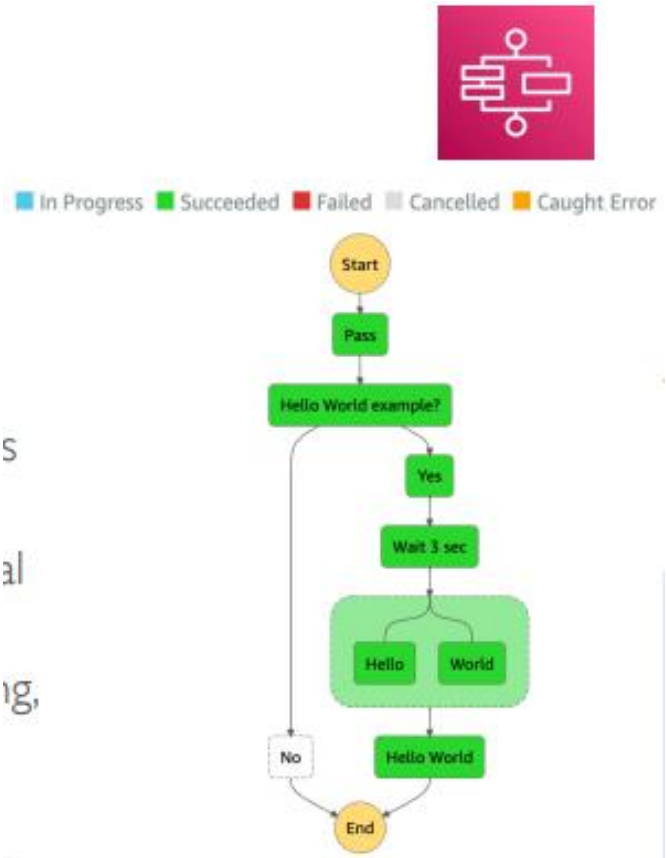
AWS Step Functions is a service that helps you create and manage workflows by connecting different AWS services into a sequence of steps. It automates and coordinates tasks, ensuring each step run in the right order and handles errors. It

features a visual workflow editor that lets you design, visualize, and monitor the flow of tasks, making it easier to manage and coordinate complex processes.

Key points:

- Build serverless visual workflow to orchestrate your Lambda functions
- **Features:** sequence, parallel, conditions, timeouts, error handling, ...
- Can integrate with EC2, ECS, On-premises servers, API Gateway, SQS queues, etc...
- Possibility of implementing human approval feature

Use cases: order fulfillment, data processing, web applications, any workflow



233. What is Amazon Cognito?

Ans ->

Amazon Cognito is a service that helps you add user **sign-up, sign-in, and access control** to your web and mobile apps quickly and easily. It provides used authentication, authorization, and user management, supporting features like multi-factor authentication and social identity providers (e.g., Facebook, Google). This allows you to secure your application without building you own authentication system.

Amazon Cognito provides two main types of pools for managing user authentications and access:

1. Cognito User Pools:

- Sign in functionality for app users
- Integrate with API Gateway & Application Load Balancer

2. Cognito Identity Pools (Federated Identity):

- Provide AWS credentials to users so they can access AWS resources directly
- Integarte with Cognito User Pools as an identity provider

Note: while going for the exam, for Cognito focus on the **key words**: "**hundreds of users", "mobile users", authentication with SAML**"

234. Explain in detail about Cognito User Pools (CUP) - User Features.

Ans ->

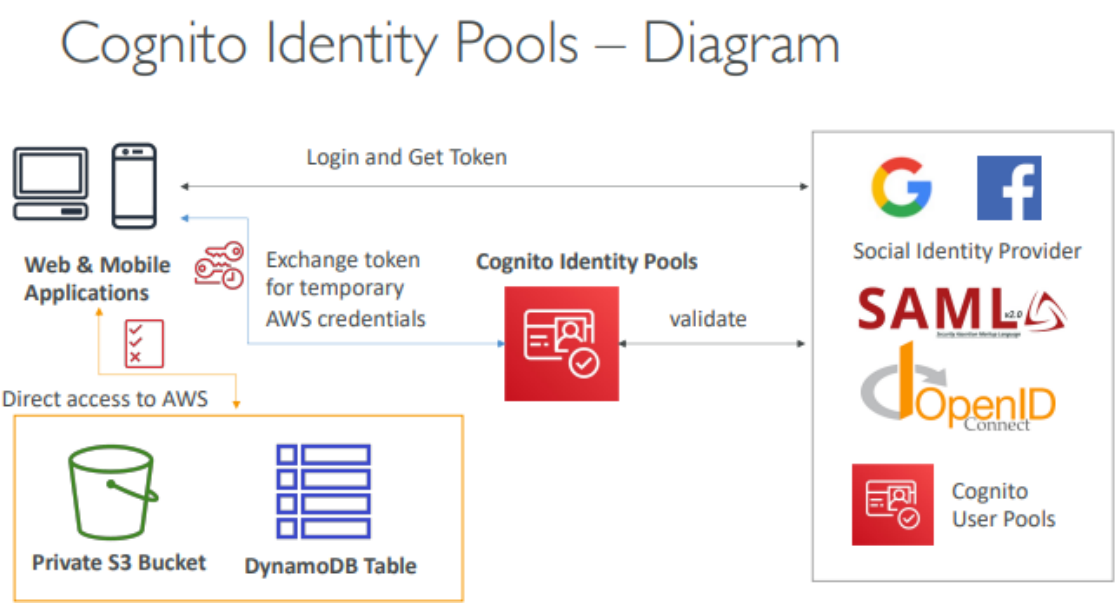
- Create a serverless database of user for your web & mobile apps
- **Simple login:** Username (or email) / password combination
- Password reset
- Email & Phone Number Verification
- Multi-factor authentication (MFA)
- **Federated Identities:** users from Facebook, Google, SAML...
- CUP integrates with API Gateway and Application Load Balancer

235. Explain in detail about Cognito Identity Pools (Federated Identities).

Ans ->

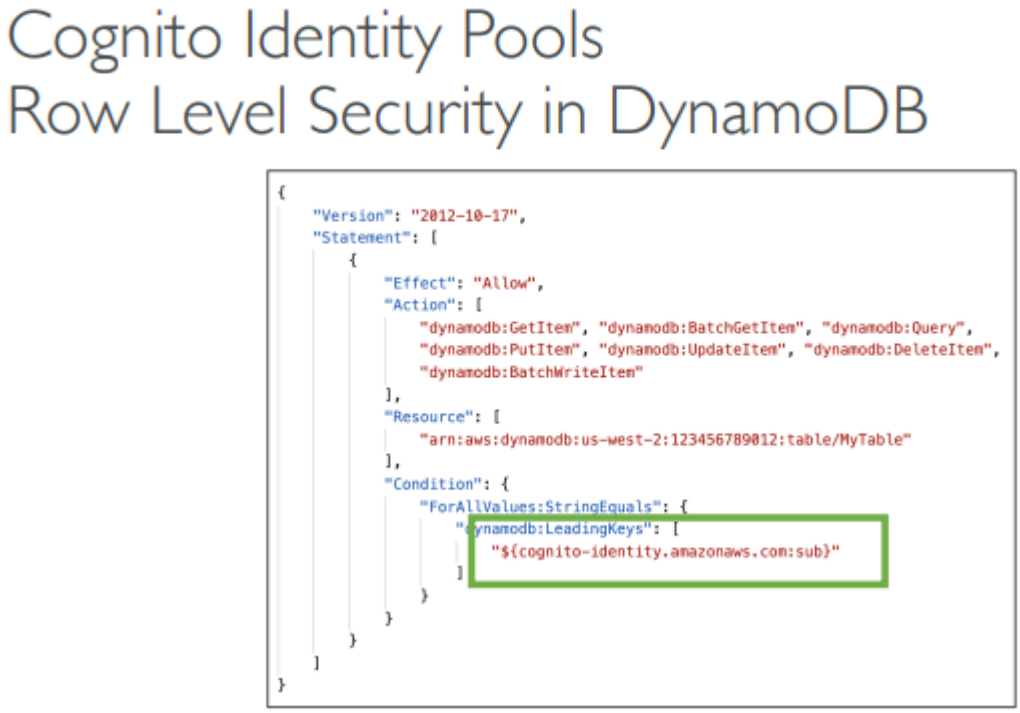
- **Get identities for "users" so they obtain temporary AWS credentials**

- Users source can be Cognito User Pools, 3rd party logins, etc...
- **Users can then access AWS services directly or through API Gateway**
- The IAM policies applied to the credentials are defined in Cognito
- They can be customized based on the user_id for fine grained control
- **Default IAM roles** for authenticated and guest users



236. Cognito Identity Pools Row Level Security in DynamoDB.

Ans ->

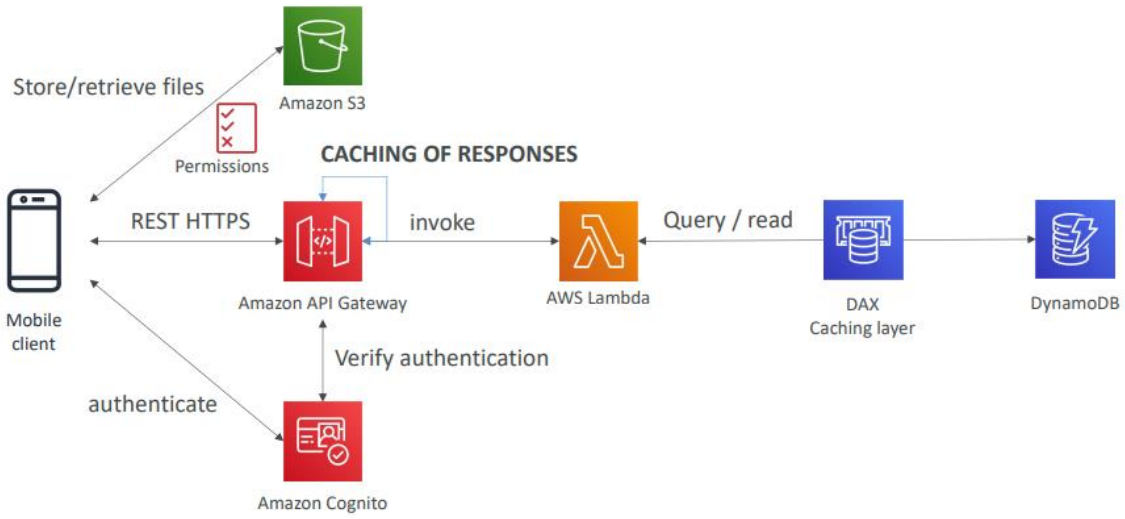


237. Mobile Application: Todo List

- **Requirements:**

- **Expose as REST API with HTTPS**
- **Serverless architecture**
- **Users should be able to directly interact with their own folder in S3**
- **Users should authenticate through a managed serverless service**
- **Then users can write and read to-dos, but they mostly read them**
- **The database should scale, and have some high read throughput**

Ans ->



Explanation of the above diagram:

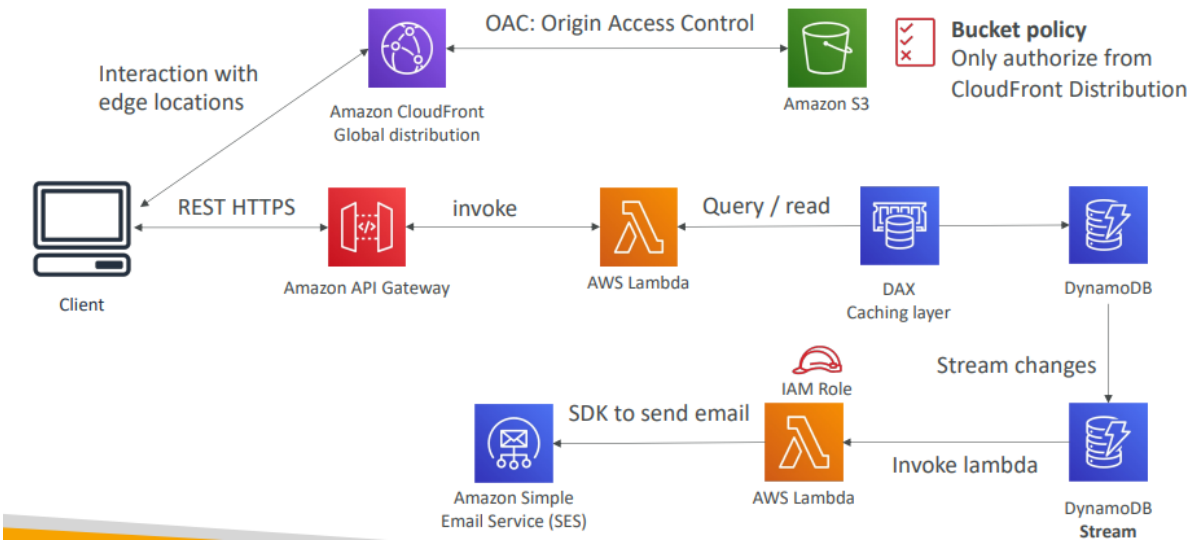
- Serverless REST API: HTTPS, API Gateway, Lambda, DynamoDB
- Using Cognito to generate temporary credentials to access S3 bucket with restricted policy. App users can directly access AWS resources this way. Pattern can be applied to DynamoDB, Lambda...
- Caching the reads on DynamoDB using DAX
- Caching the REST requests at the API Gateway level
- Security for authentication and authorization with Cognito

238. Serverless hosted website: MyBlog.com

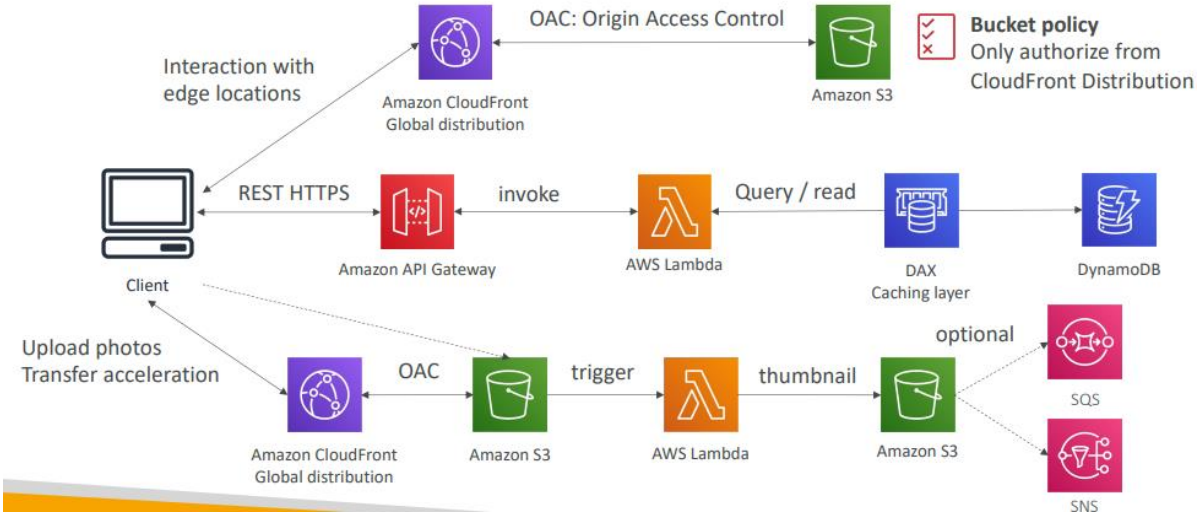
Requirements:

- **This website should scale globally**
- **Blogs are rarely written, but often read**
- **Some of the website is purely static files, the rest is a dynamic REST API**
- **Caching must be implemented where possible**
- **Any new users that subscribe should receive a welcome email**
- **Any photo uploaded to the blog should have a thumbnail generated**

Ans ->



Thumbnail Generation flow



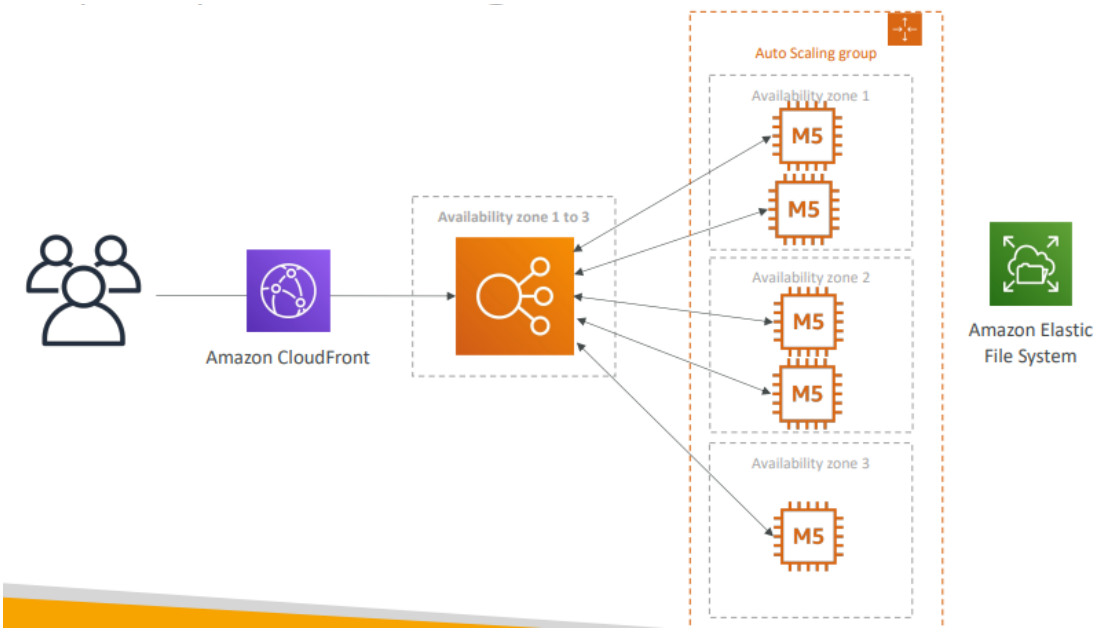
AWS Hosted Website Summary:

- We've seen static content begin distributed using CloudFront with S3
- The REST API was serverless, didn't need Cognito because public
- We leveraged a Global DynamoDB table to serve the data globally
- (We could have use Aurora Global Database)
- We enabled DynamoDB streams to trigger a Lambda function
- The lambda function had an IAM role which could use SES
- SES (Simple Email Service) was used to send emails in a serverless way
- S3 can trigger SQS/SNS/Lambda to notify of events

239. Sofware updates offloading.

- **We have an application running on EC2, that distributes software updates once in a while**
- **When a new software update is out, we get a lot of requests, and the content is distributed in mass over the network. It's very costly**
- **We don't want to change out application, but want to optimize out cost and CPU, how can we do it?**

Ans ->



By using CloudFront:

- No changes to architecture
- Will cache software update files at the edge
- Software update files are not dynamic, they're static (never changing)
- Our EC2 instances aren't serverless
- But CloudFront is, and will scale for us
- Our ASG will not scale as much, and we'll save cost tremendously in EC2
- We'll also save in availability, network bandwidth cost, etc.
- Easy way to make an existing application more scalable and cheaper!

240. Give an overview of AWS Database types.

Ans ->

Relational Databases (RDBMS/SQL/OLTP):

- **Amazon RDS and Amazon Aurora:** Ideal for complex queries and joins.

NoSQL Databases:

- **Amazon DynamoDB:** Stores data in a JSON-like format with key-value pairs.
- **Amazon ElastiCache:** Key-value store for caching.
- **Amazon Neptune:** Graph database for relationship data.
- **Amazon DocumentDB:** Compatible with MongoDB.
- **Amazon Keyspaces:** Compatible with Apache Cassandra.

Object Storage:

- **Amazon S3:** For storing large objects.
- **Amazon Glacier:** For backups and archival storage.

Data Warehousing (SQL Analytics/BI):

- **Amazon Redshift:** OLAP database for analytics.
- **Amazon Athena:** Query service for data in S3.
- **Amazon EMR:** Big data processing.

Graph Databases:

- **Amazon Neptune:** Visualizes relationships between data.

Ledger Databases:

- **Amazon Quantum Ledger Database (QLDB):** Immutable and transparent ledger.

Time Series Databases:

- **Amazon Timestream:** Optimized for time series data.

241. Amazon RDS - Summary.

Ans ->

- Managed PostgreSQL / MySQL / Oracle / SQL Server / DB2 / MariaDB/ Custom
- Provisioned RDS instance Size and EBS Volume Type & Size
- Auto-scaling capability for Storage
- Support for Read Replicas and Multi AZ
- Security through IAM, Security groups, KMS, SSL in transit
- Automated Backup with Point in time restore feature (up to 35 days)
- Manual DB Snapshot for longer-term recovery
- Managed and Scheduled maintenance (with downtime)
- Support for IAM Authentication Integration with Secrets Manager
- RDS Custom for access to and customize the underlying instance (Oracle & SQL Server)

Use Case: Store relational datasets (RDBMS / OLTP), perform SQL queries, transactions

242. Amazon Aurora - Summary.

Ans ->

- Compatible API for PostgreSQL / MySQL, separation of storage and compute
- **Storage:** data is stored in 6 replicas, across 3 AZ - highly available, self-healing, auto-scaling
- **Compute:** Cluster of DB instance across multiple AZ, auto-scaling of Read Replicas
- **Cluster:** Custom endpoints for writer and reader DB instances
- Same security / monitoring / maintenance features as RDS
- Know the backup & restore options for Aurora
- **Aurora Serverless -** for unpredictable / intermittent workloads, no capacity planning
- **Aurora Global:** up to 15 DB read Instances in each region, < 1 second storage replication
- **Aurora Machine Learning:** perform ML using SageMaker & Comprehend on Aurora
- **Aurora Database Cloning:** new cluster from existing one, faster than restoring a snapshot

Use case: same as RDS, but with less maintenance / more flexibility / more performance / more features

243. Amazon ElastiCache - Summary.

Ans ->

- Managed Redis / Memcached (similar offering as RDS, but for caches)
- In-memory data store, sub-millisecond latency
- Select an ElastiCache instance type (e.g., cache.m6g.large)
- Support for clustering (Redis) and Multi AZ, Read Replicas (sharding)

- Security through IAM, Security Groups, KMS, Redis Auth
- Backup / Snapshot / Point in time restore feature
- Managed and Scheduled Maintenance
- Requires some application code changes to be leveraged

Use Case: Key/Value store, Frequent reads, less writes, cache results for DB queries, store session data for websites, cannot use SQL.

244. Amazon DynamoDB - Summary.

Ans ->

- AWS proprietary technology, managed serverless NoSQL database, millisecond latency
- **Capacity modes:** provisioned capacity with optional auto-scaling or on-demand capacity
- Can replace ElastiCache as a key/value store (storing session data for example, using TTL feature)
- Highly Available, Multi AZ by default, Read and writes are decoupled, transaction capability
- DAX cluster for read cache, microsecond read latency
- Security, authentication and authorization is done through IAM
- **Event Processing:** DynamoDB Streams to integrate with AWS Lambda, or Kinesis Data Streams
- **Global Table feature:** active-active setup
- Automated backups up to 35 days with PITR (restore to new table), or on-demand backups
- Export to S3 without using RCU within the PITR window, import from S3 without using WCU
- Great to rapidly evolve schemas

Use Case: Serverless applications development (small documents 100s KB), distributed serverless cache

245. Amazon S3 - Summary.

Ans ->

- S3 is a... key/value store for objects
- Great for bigger objects, not so great for many small objects
- Serverless, scales infinitely, max object size is 5 TB, versioning capability
- **Tiers:** S3 Standard, S3 Infrequent Access, S3 Intelligent, S3 Glacier + lifecycle policy
- **Features:** Versioning, Encryption, Replication, MFA-Delete, Access Logs...
- **Security:** IAM, Bucket Policies, ACL, Access Points, Object lambda, CORS, Object/Vault Lock
- **Encryption:** SSE-S3, SSE-KMS, SSE-C, client-side, TLS in transit, default encryption
- Batch operations on object using S3 Batch, listing files using S3 Inventory
- **Performance:** Multi-part upload, S3 Transfer Acceleration, S3 Select
- **Automation:** S3 Event Notifications (SNS, SQS, Lambda, EventBridge)

Use Cases: Static files, key value store for big files, static website hosting

246. What is AWS DocumentDB?

Ans ->

AWS DocumentDB is a fully managed, scalable, and highly available **NoSQL** document database service designed to be **compatible** with **MongoDB**. It simplifies database management by handling maintenance, backups, and scaling, while ensuring high performance and security with features like automated backups, fault tolerance, and encryption. Designed for applications that require flexible, schema-less data storage, AWS DocumentDB provides a reliable solution for managing and querying **JSON-like documents**.

Key points:

- **Just like Aurora is an "AWS-implementation" of PostgreSQL / MySQL, DocumentDB is the same for MongoDB (which is a NoSQL database)**
- MongoDB is used to store, query, and index JSON data
- Similar "deployment concepts" as Aurora
- Fully Managed, highly available with replication across 3 AZ
- DocumentDB storage automatically grows in increments of 10GB
- Automatically scales to workloads with millions of requests per seconds.

247. What is AWS Neptune?

Ans ->

AWS Neptune is a fully managed **graph database** service designed to work with highly connected datasets. It supports popular graph models like property graphs and RDF, enabling you to build and run applications that work with complex relationships. Neptune is highly available, scalable, and secure, with features like automated backups, replication, and encryption.

Key points:

- Fully managed graph database
- A popular graph dataset would be a social network:
  - Users have friends
  - Posts have comments
  - Comments have likes from users
  - Users share and like posts...
- Highly available across 3 AZ, with up to 15 read replicas
- Build and run applications working with highly connected datasets - optimized for these complex and hard queries
- Can store up to **billions** of relations and query the graph with milliseconds latency
- Highly available with replications across multiple AZs

Use Cases: Great for knowledge graphs (Wikipedia), fraud detection, recommendations engines, social networking.

248. What is Amazon Neptune - Streams?

Ans ->

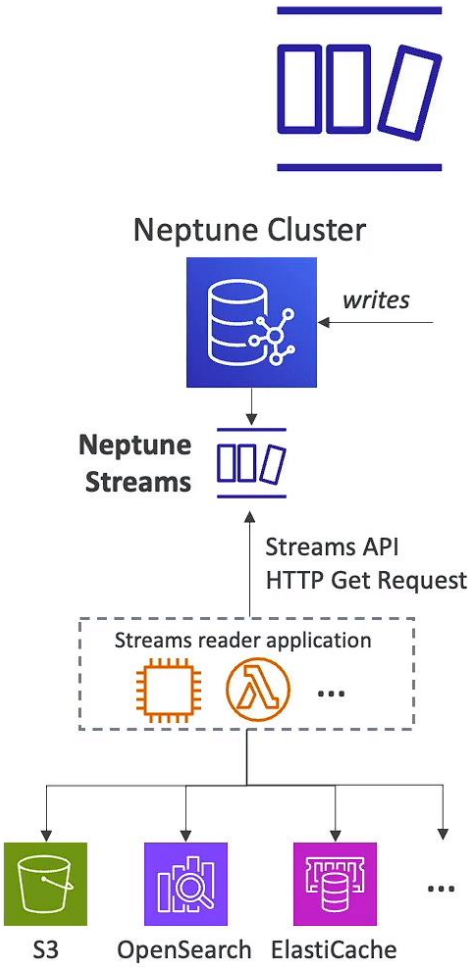
Amazon Neptune Streams is a feature that captures changes to your Neptune graph database in a time-ordered sequence, allowing you to track and process updates in near real-time. This is useful for applications that need to react to data changes, such as triggering notifications, synchronizing with other data stores, or performing analytics on the change data. Neptune Streams ensures that all changes are recorded and can be processed reliably and efficiently.

Key points:

- Real-time ordered sequence of every change to your graph data
- Changes are available immediately after writing
- No duplicates, strict order
- Streams data is accessible in an HTTP REST API

Use Cases:

- Send notifications when certain changes are made
- Maintain your graph data synchronized in another data store (e.g., S3, OpenSearch, ElastiCache)
- Replicate data across regions in Neptune



249. What is Amazon Keyspaces (for Apache Cassandra)?

Ans ->

Amazon Keyspaces (for Apache Cassandra) is a scalable, highly available, and fully managed database service designed to be compatible with Apache Cassandra. It enables you to run Cassandra workloads on AWS without managing the underlying infrastructure, providing automatic scaling, maintenance, and security features.

Key points:

- A managed Apache Cassandra-compatible database service
- Apache Cassandra is an **open-source NoSQL distributed database**
- Serverless, Scalable, highly available, fully managed by AWS
- Automatically scale tables up/down based on the application's traffic

- Tables are replicated 3 times across multiple AZ
- **Using the Cassandra Query Language (CQL)**
- **Single-digit millisecond** latency at any scale, 1000s of requests per second
- **Capacity:** On-demand mode or provisioned mode with auto-scaling
- Encryption, backup, Point-In-Time Recovery (PITR) up to 35 days

Use cases: store IoT devices info, time-series data, ...

250. What is Amazon QLDB?

Ans ->

Amazon Quantum Ledger Database (QLDB) is a fully managed ledger database that provides a transparent, immutable, and cryptographically verifiable transaction log. It is serverless and designed to track each application data change and maintain a complete and verifiable history of changes over time. QLDB is ideal for use cases requiring a trusted and auditable log of transactions, such as financial ledgers, supply chain tracking, and identity management.

Key points:

- **A ledger is a book recording financial transactions**
- Fully managed, Serverless, High available, Replications across 3 AZ
- Used to review history of all the changes made to your application data over time
- **Immutable system:** no entry can be removed or modified, cryptographically verifiable
- 2-3x better performance than common ledger blockchain frameworks, manipulate data using SQL
- Difference with Amazon Managed Blockchain: no decentralization component, in accordance with financial regulation rules

251. What is Amazon Timestream?

Ans ->

Amazon Timestream is a fully managed time series database service designed to collect, store, and analyze time series data, such as log and sensor data, efficiently and at scale. It is optimized for time series workloads, offering features like data lifecycle management. Timestream supports SQL queries for data analysis.

- Fully managed, fast, scalable, serverless time series database

- Automatically scales up/down to adjust capacity
- Store and analyze **trillions** of events per day
- 1000s times faster & 1/10th the cost of relational databases
- Scheduled queries, multi-measure records, SQL compatibility
- **Data storage tiering:** recent data kept in memory for fast access and historical data kept in a cost-optimized storage
- Built-in time series analytics functions (helps you identify patterns in your data in near real-time)
- Encryption in transit and at rest

Use cases: IoT apps, operational applications, real-time analytics, ...

252. What is AWS Athena?

Ans ->

Amazon Athena is an interactive query service provided by Amazon Web Services (AWS) that allows you to analyze data stored in Amazon S3 using standard SQL queries. It is a serverless service, which means you don't need to manage any infrastructure.

Key points:

- Serverless query service to analyze data stored in Amazon S3
- Uses standard SQL Language to query the files (Built on Presto)
- Supports CSV, JSON, ORC, Avro, and Parquet
- Pricing: $5.00 per TB of data scanned
- Commonly used with Amazon Quicksight for reporting/dashboards

Use cases: Business intelligence / analytics / reporting, analyze & query VPC Flow Logs, ELB Logs, CloudTrail trails, etc...

Exam Tip: analyze data in S3 using serverless SQL, think of Athena

253. Key points about Amazon Athena - Performance Improvement.

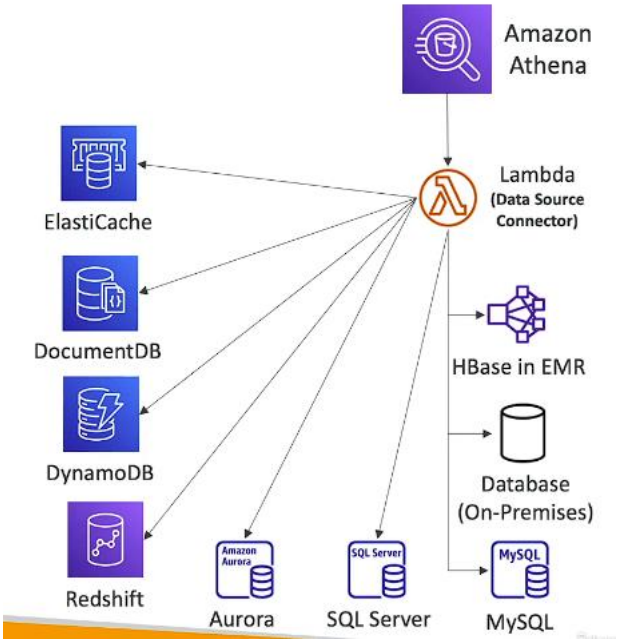
Ans ->

- **Use columnar data** for cost-savings (less scan)
- Apache Parquet or ORC is recommended
- Huge performance improvement
- **Use Glue to convert** your data to Parquet or ORC format
- Compress data for smaller retrievals (bzip2, gzip, lz4, snappy, zlip, zstd...)
- Partition datasets in S3 for easy querying on virtual columns
- S3://yourBucket/pathToTable
    - /<PARTITION_COLUMN_NAME>=<VALUE>
        - /PARTITION_COLUMN_NAME>=<VALUE>
            - /<PARTITION_COLUMN_NAME>=<VALUE>
                - /etc...
- **Example:** s3://athena-example/flight/parquet/year=1991/month=1/day=1/
- **Use larger files (>128MB) to minimize overhead**

254. What is Amazon Athena - Federated Query?

Ans ->

- **Amazon Athena Federated Query** allows you to run SQL queries across data stored in relational, non-relational, object and custom data sources (AWS or on-premises).
- **Uses Data Source Connectors that run on AWS Lambda to run Federated Queries** (e.g., CloudWatch Logs, DynamoDB, RDS, ...)
- Store the results back in Amazon S3 for later analysis



255. What is Amazon RedShift?

Ans ->

Amazon Redshift is a fully managed data warehouse service provided by Amazon Web Services (AWS). It enables you to run complex queries and perform analytics on large datasets, making it ideal for big data applications. Redshift is designed to handle petabyte-scale data and can provide high-performance query execution.

Key points:

- Redshift is based on **PostgreSQL**, but **it's not** used for **OLTP**
- **It's OLAP** - Online Analytical Processing (analytics and data warehousing)
- **10x better** performance than other data warehouses, scale to PBs of data
- **Columnar storage of data** (instead of row based) & parallel query engine
- Pay as you go based on the instance provisioned
- Has a SQL interface for performing the queries
- BI tools such as **Amazon Quicksight** or **Tableau** integrate with it
- **Compared to Athena:** faster queries / joins / aggregations thanks to indexes

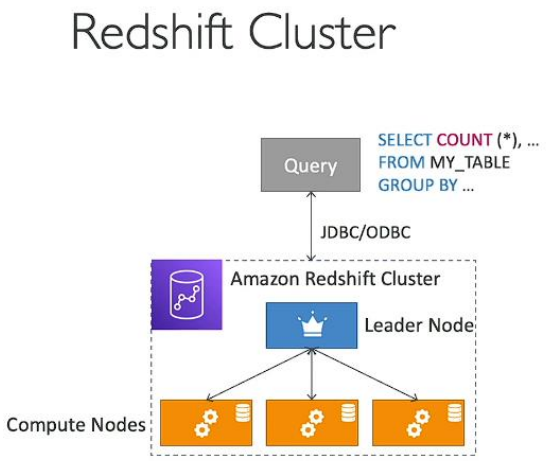
256. Key points about Redshift Cluster.

Ans ->

A Redshift cluster is a set of Amazon Redshift compute resources, organized into one or more nodes, that work together to provide a highly scalable and fast data warehousing solution. Each cluster includes a **leader node**, which coordinates query execution and communication with client applications, and one or more **compute nodes**, which handle data storage and processing. The cluster leverages **massively parallel processing (MPP)** and **columnar storage** to deliver high performance for complex analytical queries on large datasets, making it an ideal choice for data warehousing, business intelligence, and big data analytics.

Key points:

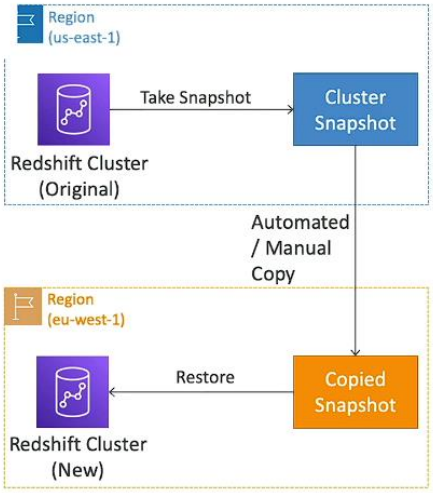
- **Leader node:** for query planning, results aggregation
- **Compute node:** for performing the queries, send results to leader
- You provision the node size in advance
- You can use Reserved Instances for cost savings



257. Key points about Redshift - Snapshots & Disaster Recovery.

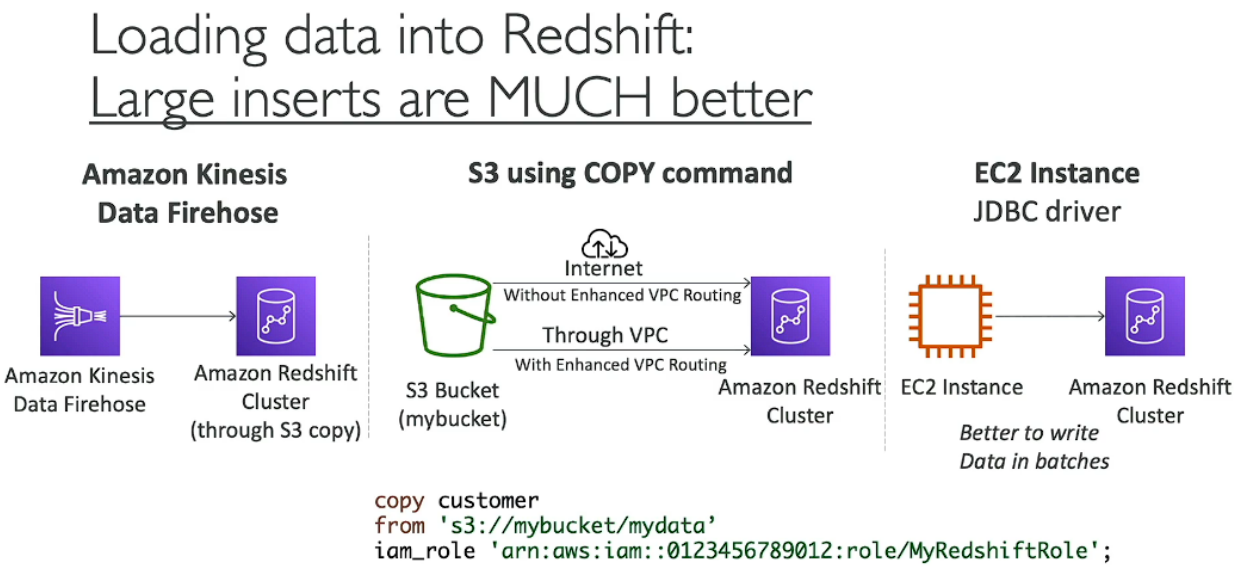
Ans->

- Redshift has "Multi-AZ" mode for some clusters
- Snapshots are **point-in-time backups** of a cluster, **stored internally in S3**
- Snapshots are incremental (only what has changed is saved)
- You can restore a snapshot into a **new cluster**
- **Automated:** every 8 hours, every 5 GB, or on a schedule. Set retention
- **Manual:** snapshot is retained until you delete it
- You can configure Amazon Redshift to automatically copy snapshots (automated or manual) or a cluster to another AWS Region



258. Loading data into Redshift: Large inserts are MUCH better.

Ans ->

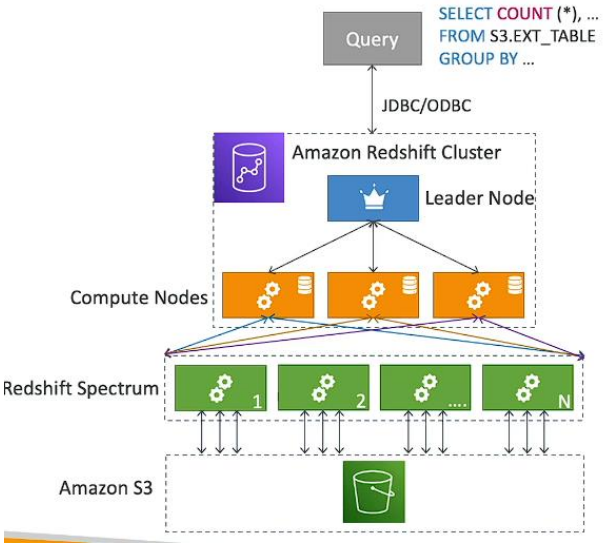


259. What is Amazon Redshift Spectrum?

Ans ->

Amazon Redshift Spectrum is a feature of Amazon Redshift that allows you to run SQL queries directly against exabytes of data in Amazon S3 without having to load the data into Redshift tables. It enables you to extend the analytics capabilities of Redshift beyond the data stored in your Redshift cluster to include your S3 data lake.

- Query data that is already in S3 without loading it
- Must have a **Redshift cluster available** to start the query
- The query is then submitted to thousands of Redshift Spectrum nodes



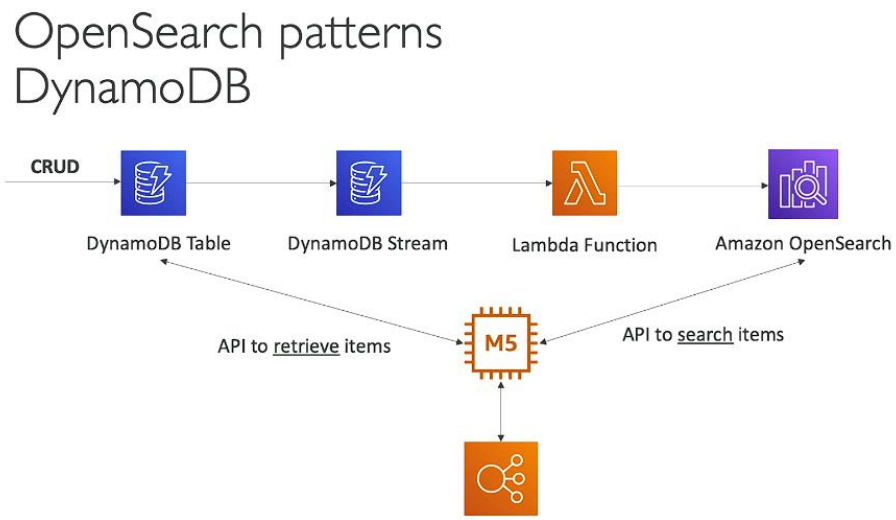
260. What is Amazon OpenSearch?

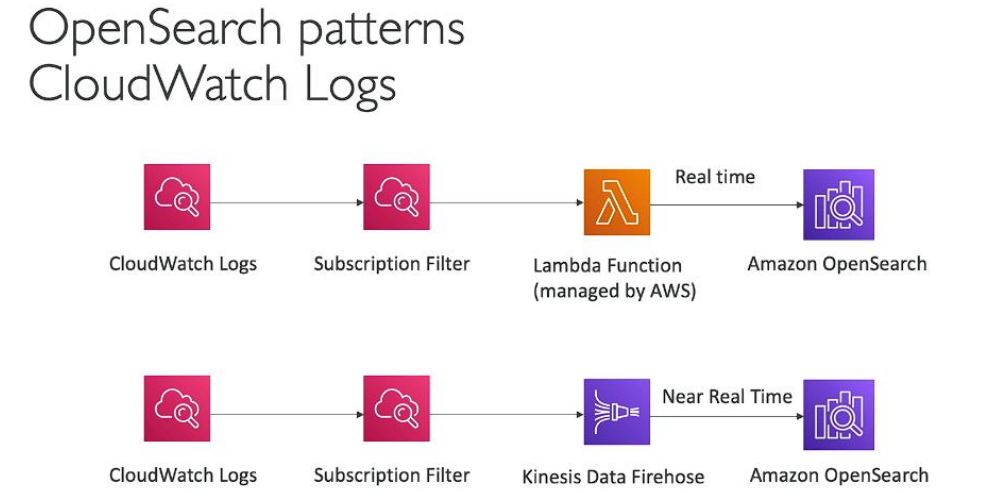
Ans ->

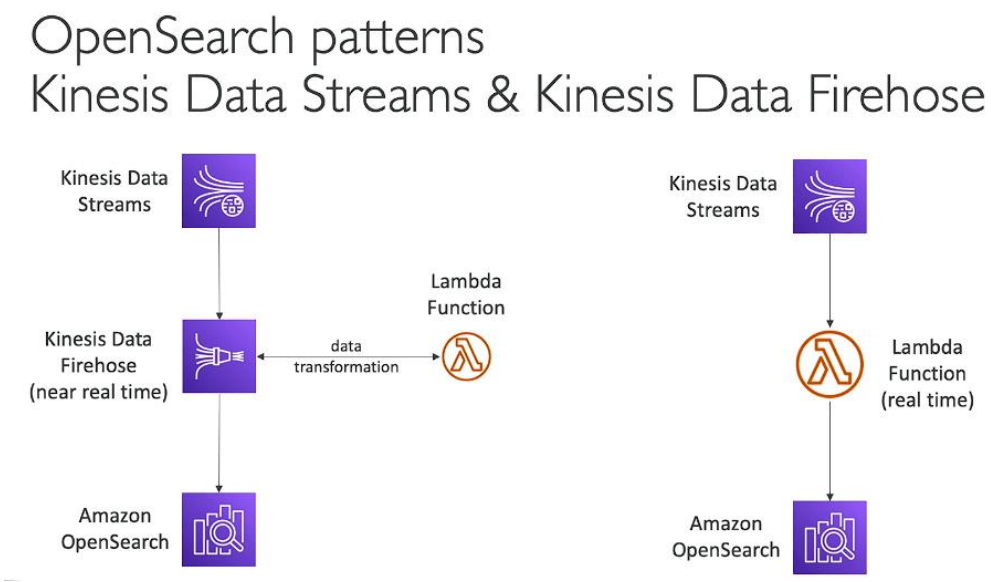
Amazon OpenSearch Service (formerly known as **ElasticSearch**) is a managed service that makes it easy to deploy, operate, and scale OpenSearch, which is a powerful search and analytics engine. It lets you search, analyze, and visualize large amounts of data in real-time without worrying about the underlying infrastructure. It's great for things like monitoring logs, analyzing application data, and building search features into your applications.

Key points:

- **Amazon OpenSearch is successor to Amazon ElasticSearch**
- In DynamoDB, queries only exist by primary key or indexes...
- With OpenSearch, you can search **any field**, even **partially matches**
- It's common to use OpenSearch as a complement to another database
- **Two modes:** managed cluster or serverless cluster
- **Does not natively support SQL (can be enabled via a plugin)**
- Ingestion from Kinesis Data Firehose, AWS IoT, and CloudWatch Logs
- Security through **Cognito and IAM, KMS encryption, TLS**
- Comes with OpenSearch Dashboards (visualization)







261. What is Amazon EMR?

Ans ->

Amazon EMR (Elastic MapReduce) is a cloud service that makes it easy to process and analyze large amounts of data using popular big data frameworks such as **Apache Hadoop, Spark, and Hive**. It allows you to quickly set up and scale clusters of compute resources, enabling you to perform large-scale data processing tasks like log analysis, web indexing, data transformations, machine learning, and more, without managing the underlying infrastructure.

Key points:

- EMR helps creating Hadoop clusters (Big Data) to analyze and process vast amount of data
- The clusters can be made of **hundreds of EC2 instances**
- EMR comes bundled with **Apache Spark, HBase, Presto, Flink...**
- EMR takes care of all the provisioning and configuration
- Auto-scaling and integrated with Spot instances

Use cases: data processing, machine learning, web indexing, big data...

262. Key points about Amazon EMR - Node types & purchasing.

Ans ->

- **Master Node:** Manage the cluster, coordinate, manage health - long running
- **Core Node:** Run tasks and store data - long running
- **Task Node (optional):** Just to run tasks - usually Spot
- **Purchasing options:**
  - **On-demand:** reliable, predictable, won't be terminated
  - **Reserved (min 1 year):** cost savings (EMR will automatically use if available)
  - **Spot instances:** cheaper, can be terminated, less reliable
- Can have long-running cluster, or transient (temporary) cluster

263. What is Amazon QuickSight?

Ans ->

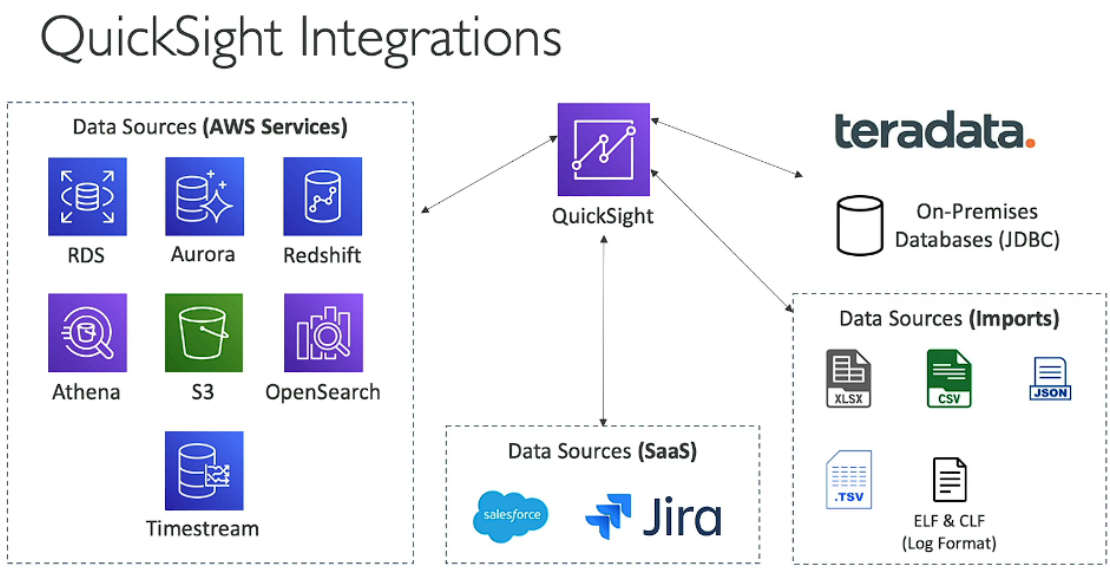
Amazon QuickSight is a cloud-based **business intelligence (BI)** service provided by AWS that enables you to create and share interactive dashboards, reports, and visualizations. It allows you to easily connect to various data sources, including AWS data services, SaaS applications, and on-premises databases, to analyze your data and derive insights. QuickSight uses machine learning to provide advanced analytics features like anomaly detection, forecasting, and natural language querying.

Key points:

- Serverless machine learning-powered business intelligence service to create interactive dashboards
- Fast, automatically scalable, embeddable, with **per-session pricing**

- **Use cases:**
  - Business analytics
  - Building visualization
  - Perform ad-hoc analysis

    - Get business insights using data
- Integrated with RDS, Aurora, Athena, Redshift, S3...
- **In-memory computation using SPICE engine** if data is imported into QuickSight
- **Enterprise edition:** Possibility to setup Column-Level security (CLS)



264. Key points about Amazon QuickSight - Dashboard & Analysis.

Ans ->

- Define users (standard versions) and Groups (enterprise version)
    - **These users & groups only exist within QuickSight, not IAM!!**
- **A dashboard:**
    - is a read-only snapshot of an analysis that you can share
    - preserves the configuration of the analysis (filtering, parameters, controls, sort)
- You can **share** the analysis or the dashboard with **Users or Groups**
- **To share a dashboard, you must first publish it**
- Users who see the dashboard can also see the underlying data

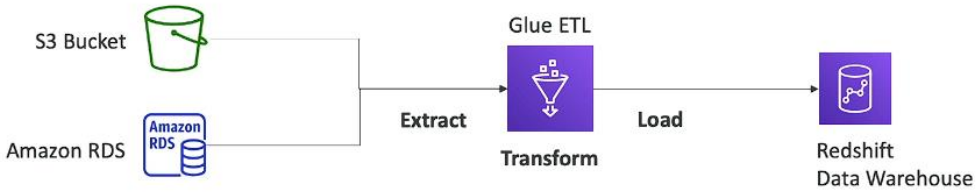
265. What is AWS Glue?

Ans ->

Amazon Glue is a fully managed **extract, transform, and load (ETL)** service provided by AWS. It helps you prepare and transform your data for analytics, machine learning, and application development. Glue automates much of the work involved in data preparation, including data discovery, schema inference, and job scheduling. It integrates seamlessly with other AWS services, enabling you to easily move and transform data across various sources like Amazon S3, RDS, and Redshift. With Glue, you can create and run ETL jobs without needing to provision or manage the underlying infrastructure.

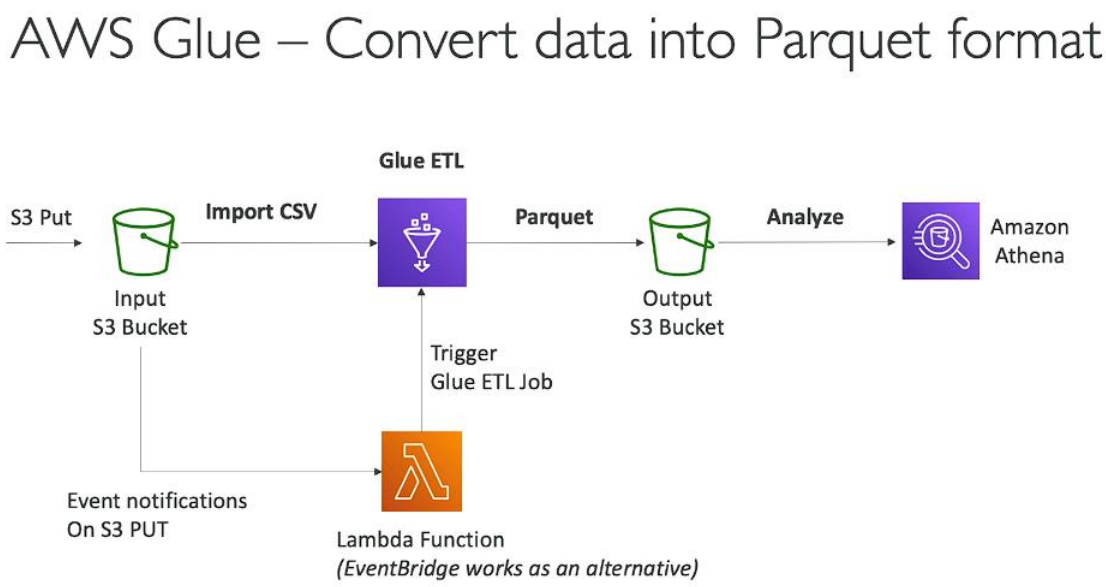
Key points:

- managed extract, transform, and load (ETL) service
- Useful to prepare and transform data for analytics
- Fully serverless service



266. AWS Glue - Convert data into Parquet format.

Ans ->



267. Glue Data Catalog: catalog of datasets.

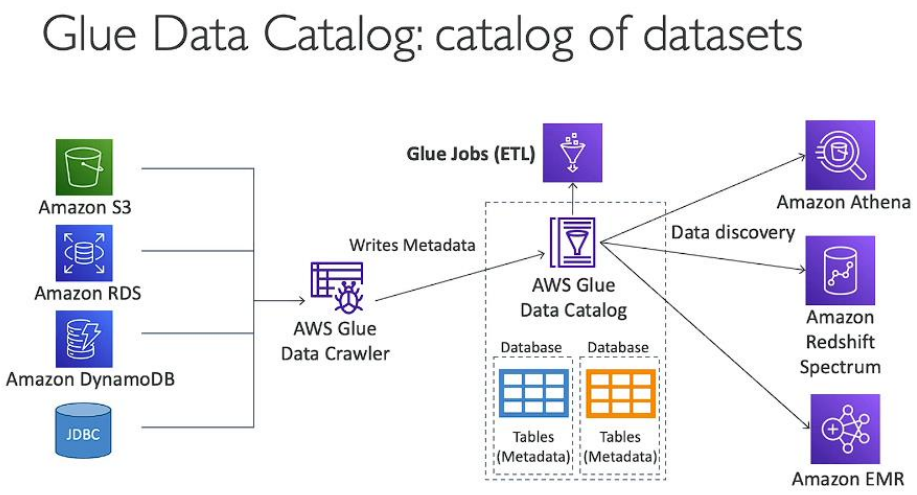
Ans ->

Glue Data Catalog:

The AWS Glue Data Catalog is a centralized metadata repository that stores information about data sources, their schemas, and transformations, enabling easy data discovery and management for analytics and ETL operations.

Glue Data Crawler:

A Glue Data Crawler is a component of AWS Glue that automatically scans your data sources, such as S3, and infers schemas to populate the AWS Glue Data Catalog, making your data ready for analysis and ETL operations.



In the above diagram, the **Glue Data Crawler** automatically discovers and catalogs metadata about your data sources, which is then stored in the **AWS Glue Data Catalog**. This cataloged metadata can be utilized by ETL jobs, Amazon Athena, Amazon Redshift Spectrum, and Amazon EMR for data analysis and processing.

268. AWS Glue - things to know at a high-level.

Ans ->

- **Glue Job Bookmarks:** prevent re-processing old data
- **Glue Elastic Views:**
  - Combine and replicate data across multiple data stores using SQL
  - No custom code, Glue monitors for changes in the source data, serverless
  - Leverages a "virtual table" (materialized view)
- **Glue DataBrew:** clean and normalize data using pre-built transformation
- **Glue Studio:** new GUI to create, run and monitor ETL jobs in Glue

- **Glue Streaming ETL (built on Apache Spark Structured Streaming):** compatible with Kinesis Data Streaming, Kafka, MSK (managed Kafka).

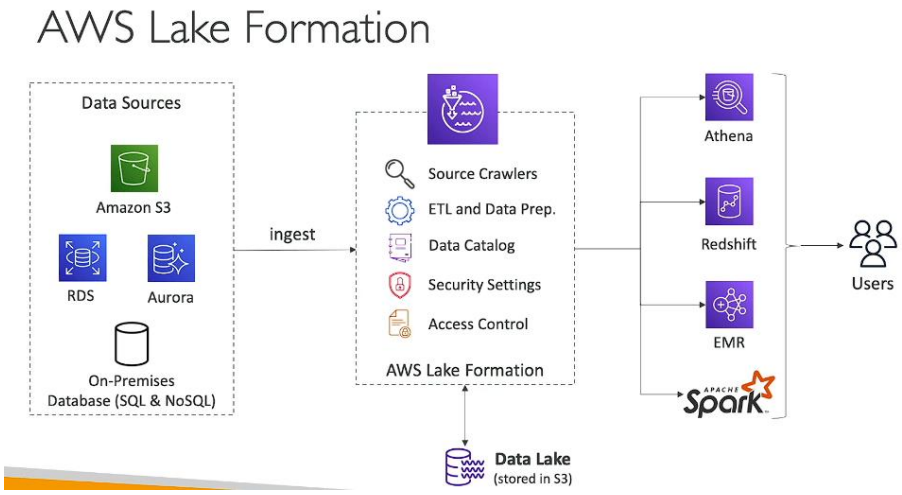
269. What is AWS Lake Formation?

Ans ->

AWS Lake Formation is a service that helps you set up a secure data lake in Amazon S3 easily. It simplifies collecting, organizing, cleaning, and securing your data so you can quickly start analyzing it with tools like Amazon Athena and Redshift. It takes care of many of the complex tasks involved in managing data, making it faster and easier to get insights from your data.

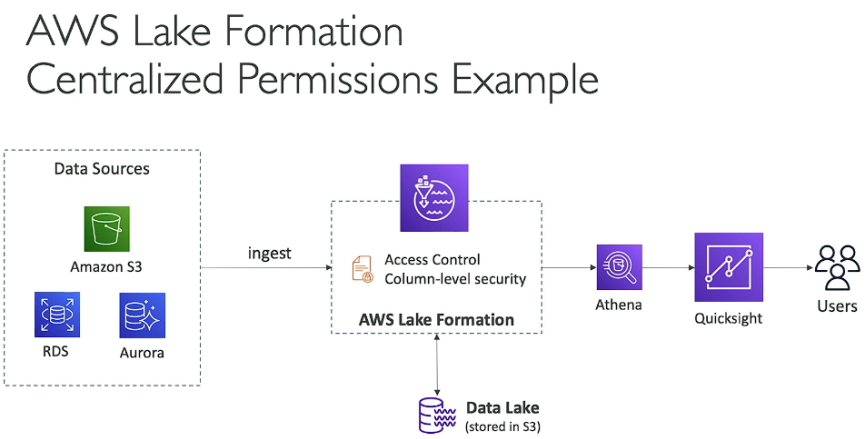
Key points:

- **Data lake = central place to have all your data for analytics purposes**
- Fully managed service that makes it easy to setup a data lake in days
- Discover, cleanse, transform, and ingest data into your Data Lake
- It automates many complex manual steps (collecting, cleansing, moving, cataloging data, ...) and de-duplicate (using ML Transforms)
- Combine structured and unstructured data in the data lake
- **Out-of-the-box source blueprints:** S3, RDS, Relational & NoSQL DB...
- Fine-grained Access Control for your applications (**row and column-level**)
- **Built on top of AWs Glue**



270. AWS lake Formation - Centralized Permissions Example.

Ans ->



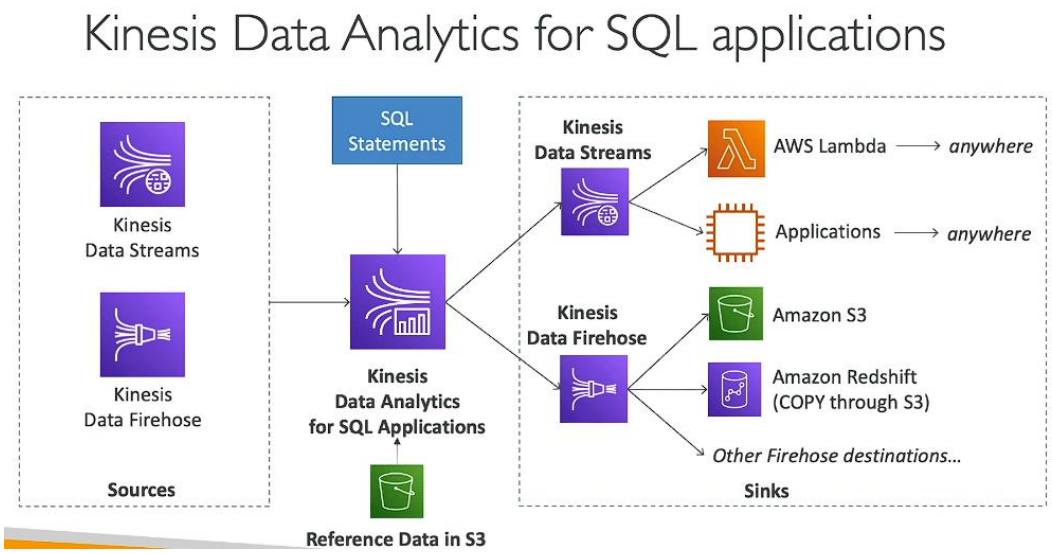
271. What is Kinesis data Analytics for SQL applications?

Ans ->

Kinesis Data Analytics (currently renamed as **AWS Managed Apache Flink) for SQL Applications** is a service that lets you use SQL to process and analyze real-time streaming data. It enables you to run continuous SQL queries on data as it flows through, helping you quickly gain insights and respond to changes in your data streams. This is useful for tasks like monitoring logs, analyzing sensor data, and creating real-time dashboards.

Key points:

- Real-time analytics on Kinesis Data Streams & Firehose using SQL
- Add reference data from Amazon S3 to enrich streaming data
- Fully managed, no servers to provision
- Automatic scaling
- Pay for actual consumption rate
- **Output:**
    - **Kinesis data streams:** create streams out of the real-time analytics queries
    - **Kinesis Data Firehose:** send analytics query results to destinations
- **Use cases:**
    - Time-series analytics
    - Real-time dashboards
    - Real-time metrics



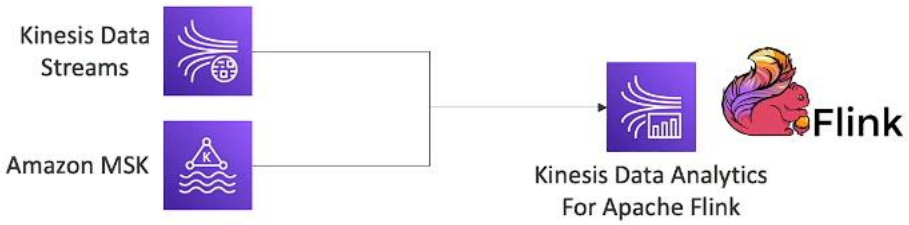
272. What is Kinesis Data Analytics for Apache Flink?

Ans ->

Kinesis Data Analytics for Apache Flink (right now it is renamed as "**Amazon Managed Apache Flink**") is a managed service that allows you to use Apache Flink, a powerful open-source framework for stream processing, to analyze and process real-time streaming data. With this service, you can build sophisticated applications that perform complex operations like filtering, aggregating, and transforming data as it flows through the system.

Key points:

- Use Flink (**Java, Scala or SQL**) to process and analyze streaming data
- Run any Apache Flink application on a managed cluster on AWS
  - provisioning compute resources, parallel computation, automatic scaling
  - application backups (implemented as checkpoints and snapshots)
  - Use any Apache Flink programming features
  - **Flink does not read from Firehose (use Kinesis Analytics for SQL instead)**



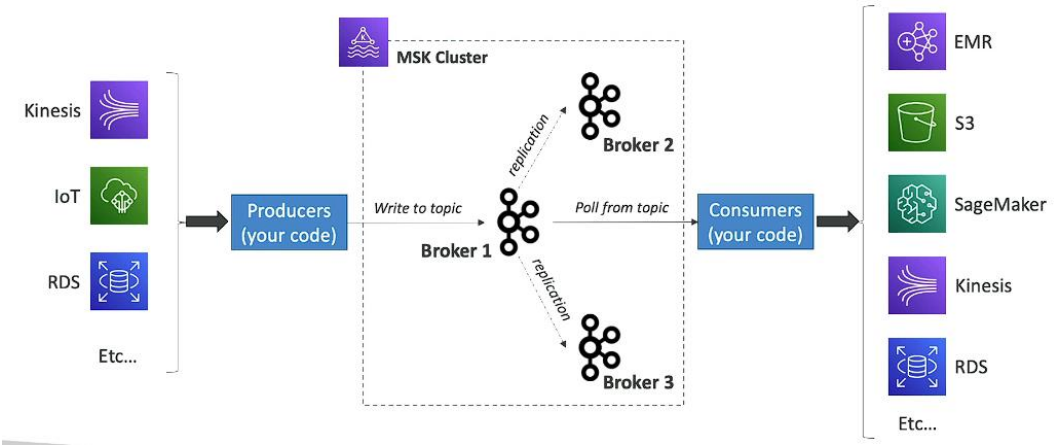
273. What is Amazon Managed Streaming for Apache Kafka (Amazon MSK).

Ans ->

Amazon Managed Streaming for Apache Kafka (Amazon MSK) is a fully managed service that makes it easy to build and run applications that use Apache Kafka to process and analyze streaming data. Apache Kafka is an open-source platform for building real-time streaming data pipelines and applications.

Key points:

- **Alternative to Amazon Kinesis**
- Fully managed Apache Kafka on AWS
  - Allow you to create, update, delete clusters
  - MSK creates & manages Kafka brokers nodes & zookeeper nodes for you
  - Deploy the MSK cluster in your VPC, multi-AZ (up to 3 for HA)
  - Automatic recovery from common Apache Kafka failures
  - Data is stored on EBS volumes for as long as you want.
- **MSK serverless:**
  - Run Apache Kafka on MSK without managing the capacity
  - MSK automatically provisions resources and scales compute & storage



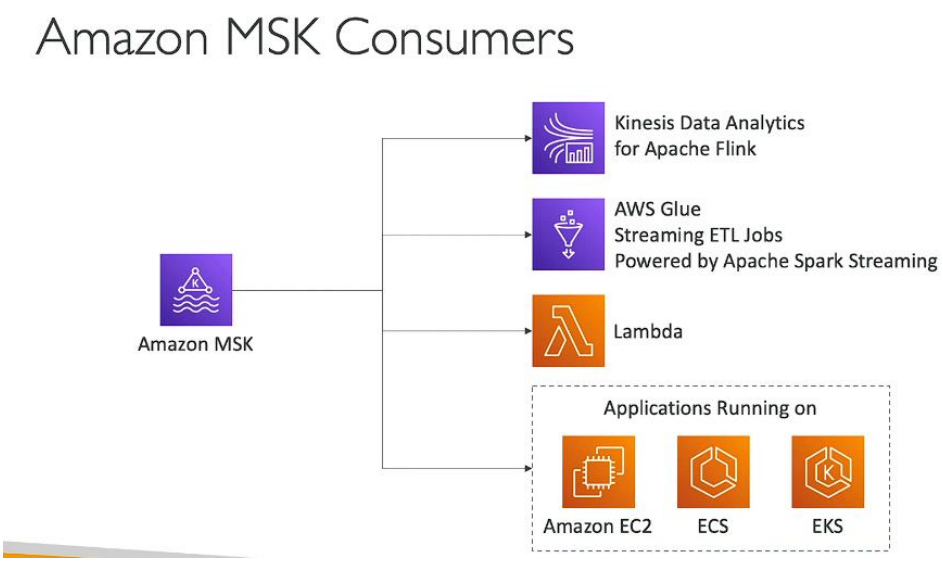
274. Kinesis Data Streams vs. Amazon MSK.

Ans ->

- **Message Size:** Both have a default 1 MB limit, but MSK can be configured for larger messages.
- **Data Structuring:** Kinesis uses shards, while MSK uses partitions.
- **Dynamic Scaling:** Kinesis supports shard splitting and merging; MSK can only add partitions.
- **Encryption:** Both provide TLS for in-transit encryption and KMS for at-rest encryption, but MSK also supports plaintext transmission.

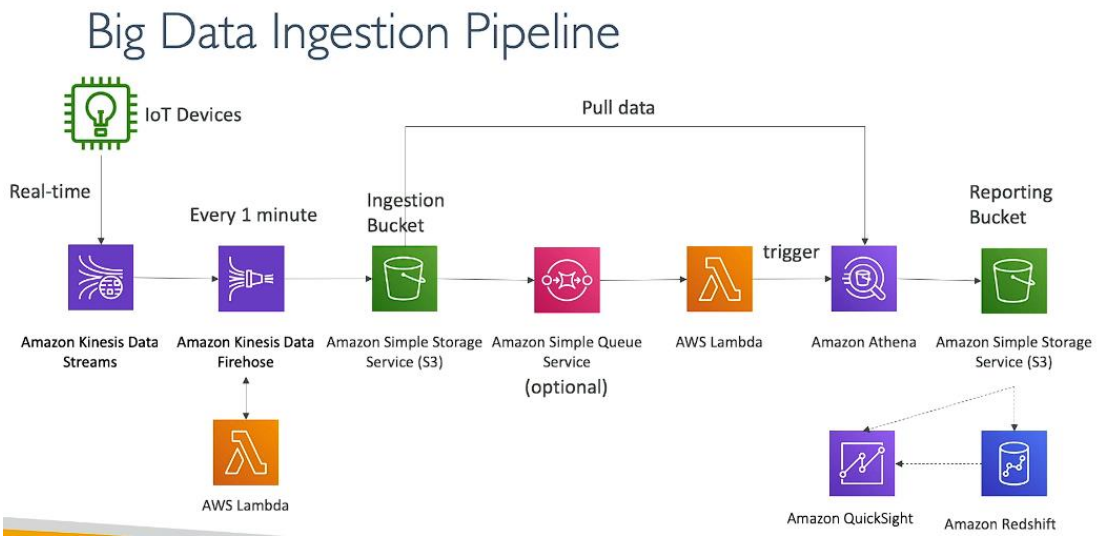
275. Amazon MSK Consumers.

Ans ->



276. Big Data Ingestion Pipeline discussion.

Ans ->



A big data ingestion pipeline refers to the process and architecture for efficiently capturing, storing, and processing large volumes of data from various sources in real-time or batch mode.

- IoT Core allows you to harvest data from IoT devices
- Kinesis is great for real-time data collection
- Firehose helps with data delivery to S3 in near real-time (1 minute)
- lambda can help Firehose with data transformations
- Amazon S3 can trigger notifications to SQS
- Lambda can subscribe to SQS (we could have connecter S3 to Lambda)
- Athena is a serverless SQL service and results are stored in S3

- The reporting bucket contains analyzed data and can be used by reporting tool such as AWS QuickSight, redshift, etc...

277. What is Amazon Rekognition?

Ans ->

Amazon Rekognition is a cloud-based service that uses deep learning to analyze images and videos. It can identify objects, people, text, scenes, and activities, as well as detect inappropriate content. It's commonly used for applications like facial recognition, image and video search, and content moderation.

Key points:

- Fin objects, people, text, scenes in images and videos using ML
- Facial analysis and facial search to do user verification, people counting
- Create a database of "familiar faces" or compare against celebrities
- **Use cases:**
  - Labeling
  - Content Moderation
  - Text Detection
  - Face Detection and Analysis (gender, age range, emotions...)
  - Face Search and Verification
  - Celebrity Recognition
  - Pathing (ex: for sports game analysis)

278. Key points about Amazon Rekognition - Content Moderation.

Ans ->

- Detect **content that is inappropriate, unwanted, or offensive** (image or videos)
- Used in social media, broadcast media, advertising, and e-commerce situations to create a safer user experience
- **Set a Minimum Confidence Threshold for items that will be flagged**
- Flat sensitive content for manual review in **Amazon Augmented AI (A2I)**
- Help complies with regulations.

279. What is Amazon Transcribe?

Ans ->

Amazon Transcribe is a cloud-based service that automatically converts speech into written text. It uses advanced machine learning to provide accurate and easy-to-read transcriptions from audio and video files, supporting various languages and dialects.

Key points:

- Automatically convert speech to text
- Uses a deep learning process called **automatic speech recognition (ASR)** to convert speech to text quickly and accurately
- Automatically remove **Personally Identifiable Information (PII)** using **Redaction**
- Supports Automatic language Identification for **multi-lingual audio**
- **Use cases:**
  - transcribe customer service calls
  - generating subtitles for videos
  - converting meeting recordings into text for documentation
  - generate metadata for media assets to create a fully searchable archive

280. What is Amazon Polly?

Ans ->

Amazon Polly is a cloud-based service that converts text into lifelike speech. Using advanced deep learning technologies, it generates natural-sounding human speech in multiple languages and voices.

Key points:

- Turn text into lifelike speech using deep learning
- Allowing you to create applications that talk
- **Use cases:**

- creating interactive voice responses
- providing text-to-speech functionality for applications
- generating audio versions of text content
- enhancing accessibility for visually impaired users

281. What is Amazon Polly - Lexicon?

Ans ->

Amazon Polly - Lexicons allow you to define how specific words or phrases should be **pronounced**. By creating custom **pronunciation dictionaries**, you can ensure that Polly speaks specialized terms, proper names, or acronyms correctly. Lexicons are defined using the **Pronunciation Lexicon Specification (PLS) format**.

Key points:

- Customize the pronunciation of words with Pronunciation lexicons
- **Stylized words:** S4m1m => "Samim"
- Acronyms: AWS => "Amazon Web Services"
- Upload the lexicons and use them in the **SynthesizeSpeech** operation

282. What is Amazon Polly - SSML?

Ans ->

Amazon Polly - SSML (Speech Synthesis Markup Language) is a markup language that lets you control various aspects of speech output. With SSML tags, you can adjust speech rate, volume, pitch, and pronunciation, as well as insert pauses, emphasize certain words, and add other effects. This provides finer control over how the synthesized speech sounds, making it more natural and expressive.

Key points:

- Generate speech from plain text or from documents marked up with Speech **Synthesis Markup Language (SSML)** - enables more customizations:
  - emphasizing specific words or phrases
  - using phonetic pronunciation
  - including breathing sounds, whispering
  - using the Newscaster speaking style

283. What is Amazon Translate?

Ans ->

Amazon Translate is a cloud-based machine translation service that automatically translates text between different languages. It uses advanced deep learning models to

provide fast, high-quality translations for a variety of applications, such as translating websites, documents, and real-time communication.

- natural and accurate language translation
- Amazon Translate allows you to localize content - such as websites and applications - for international users, and to easily translate large volumes of text efficiently.

284. What is Amazon Lex?

Ans ->

Amazon Lex is a service for building **conversational interfaces** into any application using voice and text. It uses the same deep learning technologies that power **Amazon Alexa** to enable you to create chatbots and voice assistants.

With **Amazon Lex**, you can define **intents (what users want to achieve)**, **slots (pieces of information needed to fulfill the intent)**, and **utterances (phrases users might say to express the intent)**. This allows you to build applications with natural language understanding and automatic speech recognition, making it easier to interact with users in a conversational manner.

- **Automatic Speech Recognition (ASR)** to convert speech to text
- Natural language Understanding to recognize the intent of text, callers
- **helps build chatbots, call center bots**

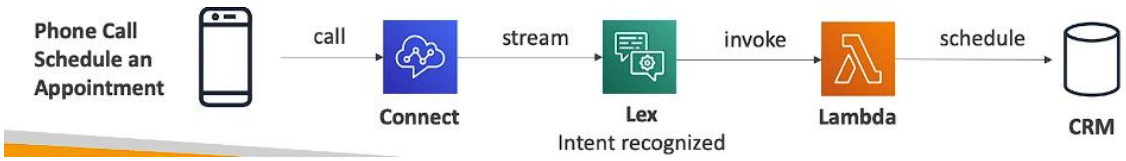
285. What is Amazon Connect?

Ans ->

Amazon Connect is a cloud-based **contact center service** that enables businesses to provide customer service and support through various communication channels, including **voice, chat,** and **email**. Amazon Connect offers features like automatic call distribution, interactive voice response (IVR), real-time and historical analytics.

- Receive calls, create flows, cloud-based virtual contact center
- Can integrate with other CRM systems or AWS
- **No upfront payments, 80% cheaper than traditional contact center solutions**

Amazon Connect integrates seamlessly with Amazon Lex to enable conversational AI in contact centers.



286. What is Amazon Comprehend?

Ans ->

Amazon Comprehend is a **Natural Language Processing (NLP)** service provided by AWS that uses machine learning to analyze and understand text. It can perform tasks like **sentiment analysis, entity recognition, key phrase extraction, language detection, and topic modeling**. Essentially, it helps you extract **insights** and understand the content of **text data**, making it useful for applications like customer feedback analysis, content categorization, etc.

Key points:

- For Natural Language Processing - NLP
- Fully managed and serverless service
- Uses machine learning to find insights and relationships in text
- Language of the text
- Extracts key phrases, places, people, brands, or events
- Understands how positive or negative the text is
- Analyzes text using tokenization and parts of speech
- Automatically organizes a collection of text files by topic
- **Sample use cases:**
    - analyze customer interactions (emails) to find what ideas to a positive or negative experience
    - Create and group articles by topics that Comprehend will uncover

287. What is Amazon Comprehend Medical?

Ans ->

Amazon Comprehend Medical is an AWS service that leverages natural language processing to analyze **unstructured medical text, such as clinical notes and patient records**. It extracts and identifies key medical entities, relationships, and protected health information (PHI), enabling healthcare professionals and researchers to automate data processing and **gain valuable insights from medical documentation**.

Key points:

- Amazon Comprehend Medical detects and returns useful information in unstructured clinical text:
  - **Physician's notes**
  - **Discharge summaries**
  - **Test results**
  - **Case notes**
- Use NLP to detect **Protected Health Information (PHI)** - **DetectPHI API**
- Store your documents in amazon S3, analyze real-time data with Kinesis Data Firehose, or use Amazon Transcribe to transcribe patient narratives into text that can be analyzed by Amazon Comprehend medical.

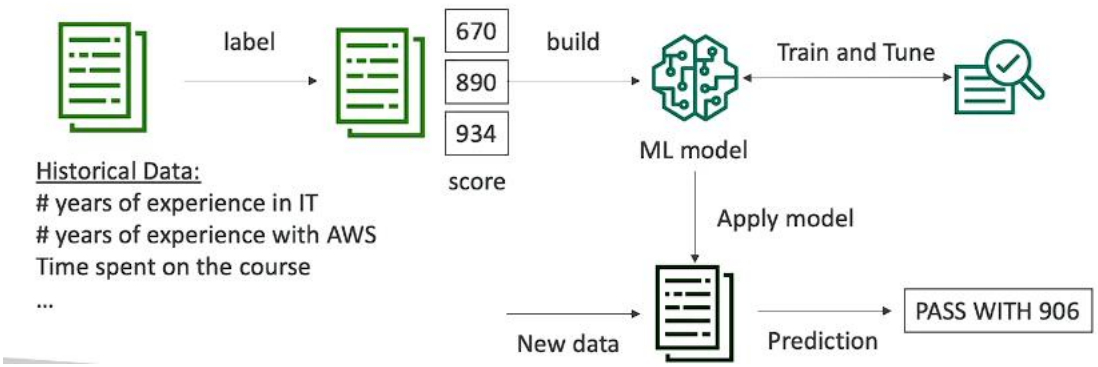
288. What is Amazon SageMaker?

Ans ->

Amazon SageMaker is a fully managed AWS service that simplifies the machine learning lifecycle by providing tools for building, training, and deploying models. It includes features like pre-built algorithms, Jupyter notebooks for development, automatic model tuning, and managed training and deployment, enabling developers and data scientists to efficiently create and scale machine learning models.

Key points:

- Fully managed service for developers / data scientists to **build ML models**
- Typically, difficult to do all the processes in one place + provision servers
- Machine learning process (simplified): predicting your exam score:



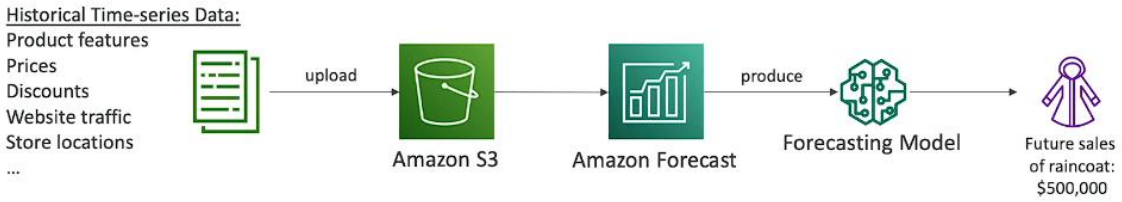
289. What is Amazon Forecast?

Ans ->

Amazon Forecast is a fully managed service from AWS that uses machine learning to generate accurate **time series forecasts**. It can predict future values based on historical data, helping businesses with demand planning, inventory management, and financial forecasting. The service automates the forecasting process, providing users with highly accurate predictions by leveraging machine learning algorithms and the extensive data capabilities of AWS.

Key points:

- Fully managed service that uses ML to deliver highly accurate forecasts
- Example: predict the future sales of a raincoat
- 50% more accurate than looking at the data itself
- Reduce forecasting time from months to hours
- Use cases: Product Demand Planning, Financial Planning, Resource Planning....



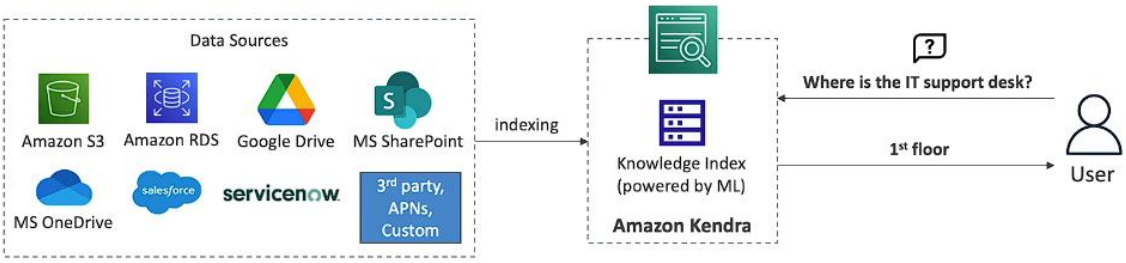
290. What is Amazon Kendra?

Ans ->

Amazon Kendra is a powerful search service that helps organizations **find information** quickly and accurately from their **internal documents and databases**. By using machine learning, it understands the context and meaning of your questions, so it can provide relevant results. Whether your data is in databases, file systems, intranet sites, or applications, Kendra can search across all these sources.

Key points:

- Fully managed **document search service** powered by machine Learning
- Extract answers from within a document (text, pdf, HTML, PowerPoint, MS Word, FAQs...)
- Natural language search capabilities
- **Learn from user interactions/feedback** to promote preferred results (Incremental Learning)
- Ability to manually fine-tune search results (importance of data, freshness, custom, ...)



291. What is Amazon Personalize?

Ans ->

Amazon Personalize is an AWS service that enables developers to create **personalized recommendations** for users, similar to Amazon's own recommendation system. It integrates with websites and applications to provide real-time, tailored content and product suggestions based on user behavior and preferences, enhancing user experience and engagement.

Key points:

- Fully managed ML-service to build apps with real-time personalized recommendations
- **Example:** personalized product recommendations/re-ranking, customized direct marketing
    - **Example:** User bought gardening tools, provide recommendations on the next one to buy
- Same technology used by Amazon.com
- Integrates into existing websites, applications, SMS, email marketing systems, ...
- Implement in days, not months (you don't need to build, train, and deploy ML solutions)
- **Use cases:** retail stores, media and entertainment...

292. What is Amazon Textract?

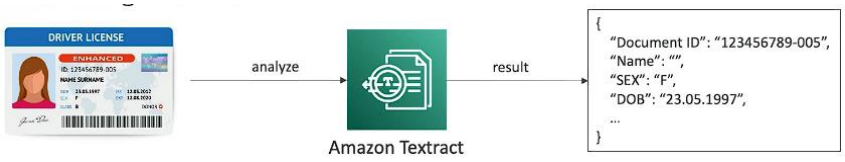
Ans ->

Amazon Textract is an AWS service that uses machine learning to automatically **extract text, handwriting, and data from scanned documents**. It goes beyond simple optical character recognition (OCR) to identify and understand the structure of documents, such as forms and tables, enabling users to efficiently process and analyze

their document-based data. This service helps automate data extraction, reducing manual effort and improving accuracy in document processing tasks.

Key points:

- Automatically extracts text, handwriting, and data from any scanned documents using AI and ML



- Extract data from forms and tables
- Read and process any type of document (PDFs, images, ...)
- **Use cases:**
  - Financial Services (e.g., invoices, financial reports)
  - Healthcare (e.g., medical records, insurance claims)
  - Public Sector (e.g., tax forms, ID documents, passports)

293. AWS Machine Learning - Summary.

Ans ->

- **Rekognition:** face detection, labeling, celebrity recognition
- **Transcribe:** audio to text (ex: subtitles)
- **Polly:** text to audio
- **Translate:** translations
- **Lex:** build conversational bots - chatbots
- **Connect:** cloud contact center
- **Comprehend:** natural language processing
- **SageMaker:** machine learning for every developer and data scientists
- **Forecast:** build highly accurate forecasts
- **Kendra:** ML-powered document search engine
- **Personalization:** real-time personalized recommendations
- **Textract:** detect text and data in documents

294. What is AWS CloudWatch?

Ans ->

AWS CloudWatch is a monitoring and observability service that **provides insights into AWS resources and applications**, allowing users to collect and track metrics, monitor log files, set alarms, and respond automatically to changes in the AWS environment. It helps ensure operational health by providing real-time visibility into resource utilization, performance, and operational data, enabling proactive management and troubleshooting of AWS infrastructure and applications.

295. What is AWS CloudWatch Metrics?

Ans ->

CloudWatch Metrics is a feature of AWS CloudWatch that tracks and records data points, like CPU usage or memory usage, from your AWS resources. It helps you monitor the performance and health of your applications by showing how different parts of your system are working over time, and lets you set alerts if something goes wrong.

Key points:

- CloudWatch provides metrics **for every service in AWS**
- Metric is a variable to monitor (CPUUtilization, NetworkIn...)
- Metrics belong to namespaces
- Dimension is an attribute of a metric (instance id, environment, etc...)
- Up to 30 dimensions per metric
- Metrics have timestamps
- Can create CloudWatch dashboards of metrics
- Can create CloudWatch Custom Metrics (for the RAM for example)

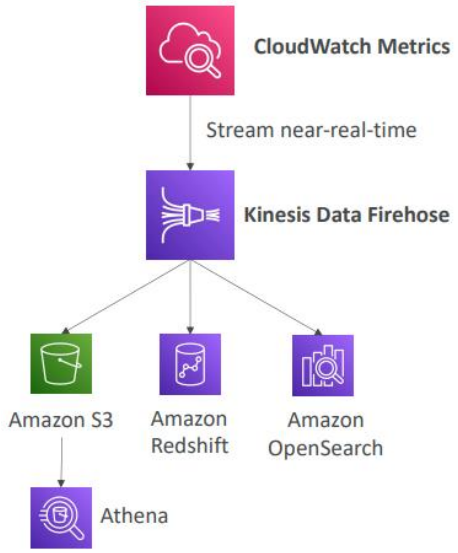
296. What is AWS CloudWatch Metric Streams?

Ans ->

CloudWatch Metrics Streams is a feature that allows you to continuously stream your CloudWatch metrics data and send them to other AWS services or third-party tools in near real-time. This enables you to analyze, visualize, and store your metrics outside of CloudWatch, providing more flexibility in how you use and process this data. It's useful for advanced analytics, long-term storage, or integrating with other monitoring systems.

Key points:

- Continually stream CloudWatch metrics to destination of your choice, with near-real-time delivery and low latency.
- Amazon Kinesis Data Firehose (and then its destinations)
- **3rd party services provider:** Datadog, Dynatrace, New Relic, Splunk, Zabbix, Sumo Logic...
- Option to filter metrics to only stream a subset of them.



297. What is AWS CloudWatch Logs?

Ans ->

AWS CloudWatch Logs is a service that helps you monitor, store, and access log files from various AWS resources, such as EC2 instances, Lambda functions, and other AWS services. It allows you to collect and centralize logs from multiple sources, search and filter the logs, and even set up alarms based on specific log patterns. This makes it easier to troubleshoot issues, monitor the performance of your applications, and maintain operational visibility.

Key points:

- **Log Groups:** arbitrary name, usually representing an application
- **Log stream:** instances within application / log files / containers
- Can define log expiration policies (never expire, 1 day to 10 years...)
- CloudWatch Logs can send logs to:
  - Amazon S3 (exports)
  - Kinesis Data Streams
  - Kinesis Data Firehose
  - AWS Lambda
  - OpenSearch
- Log are encrypted by default
- Can setup KMS-based encryption with your own keys

298. What are the different Log sources for AWS CloudWatch Logs?

Ans ->

- SDK, CloudWatch Logs Agent, CloudWatch Unified Agent
- **Elastic Beanstalk:** collection of logs from application
- **ECS:** collection from containers
- **AWS Lambda:** collection from function logs
- **VPC Flow Logs:** VPC specific logs
- API Gateway
- CloudTrail based on filter
- **Route53:** Log DNS queries

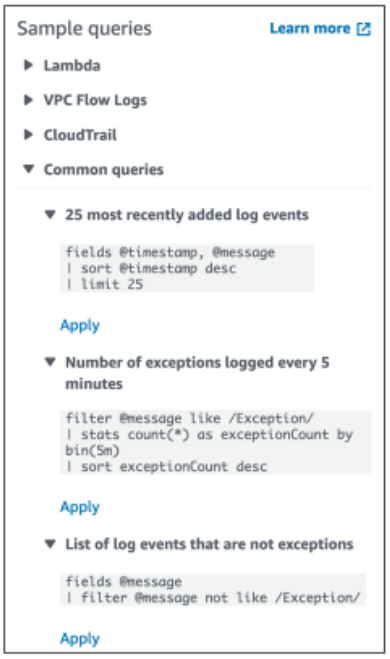
299. What is CloudWatch Logs Insights?

Ans ->

CloudWatch Logs Insights is a tool within AWS CloudWatch that lets you quickly query and analyze log data using a simple query language. It helps you filter, aggregate, and visualize logs in real-time, making it easier to troubleshoot issues and understand your system's performance.

Key points:

- Search and analyze log data stored in CloudWatch Logs
- **Example:** find a specific IP inside a log, count occurrences of "ERROR" in your logs...
- Provides a purpose-built query language
  - Automatically discovers fields from AWS services and JSON log events
  - Fetch desired event fields, filter based on conditions, calculate aggregate statistics, sort events, limit number of events...
  - Can save queries and add them to CloudWatch Dashboards
- Can query multiple Log Groups in different AWS accounts
- It's a query engine, not a real-time engine



300. What is AWS CloudWatch Logs - S3 Export.

Ans ->

CloudWatch Logs - S3 Export is a feature that allows you to export your log data from AWS CloudWatch Logs to an Amazon S3 bucket. This enables you to store logs for long-term archival, further analysis, or integration with other data processing tools. It's useful for retaining logs beyond the retention period offered by CloudWatch Logs and for meeting compliance or auditing requirements.

Key points:

- Log data can take **up to 12 hours** to become available for export
- The API call is **CreateExportTask**
- Not near-real or real-time... use Logs Subscriptions instead

301. What is CloudWatch Logs Subscriptions?

Ans ->

CloudWatch Logs Subscriptions let you **stream log data in real-time** from CloudWatch Logs to other AWS services or external systems. This is achieved through a subscription filter that captures log events and sends them to destinations like Amazon Kinesis, Amazon Elasticsearch Service, or AWS Lambda.

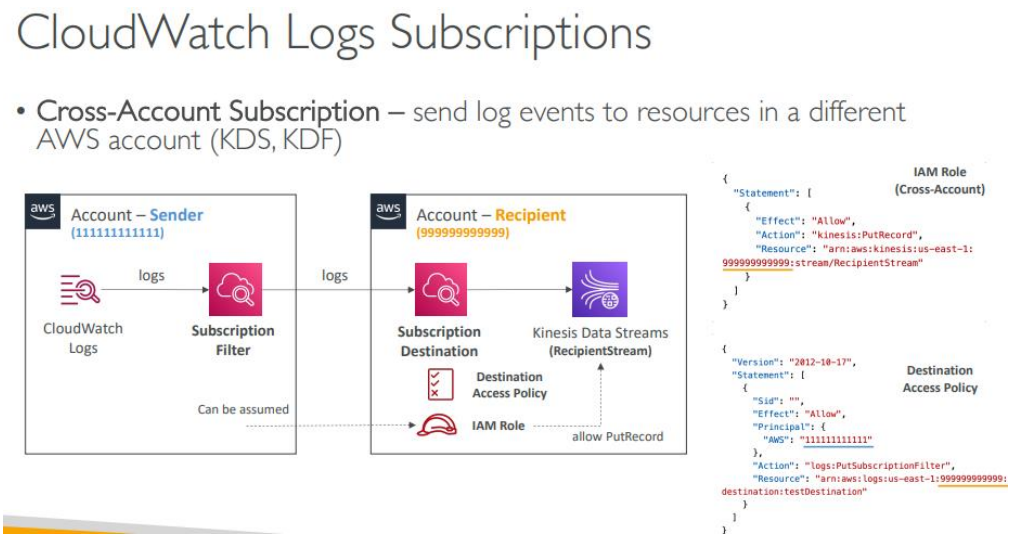
Key points:

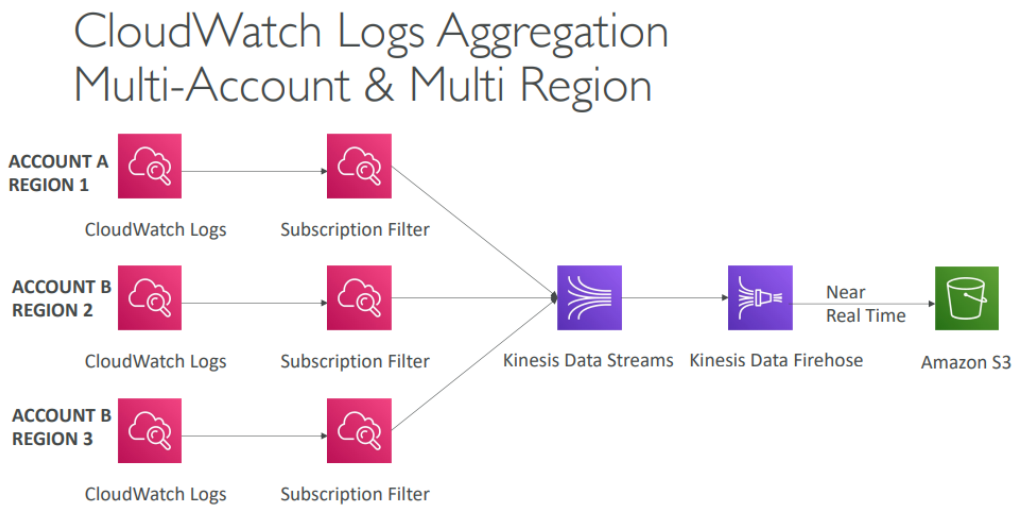
- Get a real-time log events from CloudWatch Logs for processing and analysis
- Send to Kinesis Data Streams, Kinesis Data Firehose, or Lambda
- **Subscription Filter** - filter which logs are events delivered to your destination

302. CloudWatch Logs - Cross-Account Subscription.

Ans ->

Cross-Account Subscription - send log events to resources in a different AWS account (KDS, KDF)





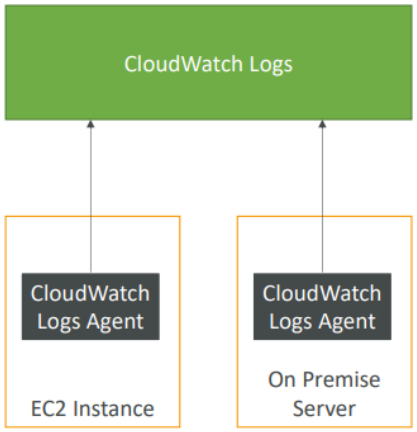
303. CloudWatch Logs for EC2.

Ans ->

CloudWatch Logs for EC2 lets you send log files from your EC2 instances to AWS CloudWatch. By setting up a **CloudWatch Agent** on your EC2, you can automatically send system and application logs to a central place. This makes it easier to monitor your logs, spot problems, and get alerts if something goes wrong, helping you keep your applications running smoothly.

Key points:

- By default, no logs from your EC2 machine will go to CloudWatch
- You need to run a CloudWatch Logs Agent on EC2 to push the log files you want
- Make sure IAM permissions are correct
- The CloudWatch Logs Agent can be setup on-premises too.



304. What is CloudWatch Agent and CloudWatch Logs Agent?

Ans ->

CloudWatch Agent is a tool that collects both system metrics (like CPU and memory usage) and logs from your servers and virtual machines. It's versatile, works on different operating systems, and can be customized for detailed monitoring.

CloudWatch Logs Agent is a simpler tool focused solely on collecting and sending log data to CloudWatch Logs. It's useful for basic log management but **doesn't collect system metrics**.

305. CloudWatch Logs Agent and unified Agent.

Ans ->

Logs Agent is a tool focused solely on collecting log data from servers and sending it to CloudWatch Logs. It's simple to set up and is mainly used for basic log management.

- Old version of the agent
- Can only send to CloudWatch Logs

Unified Agent (CloudWatch Agent) is more advanced, capable of collecting both logs and system metrics like CPU and memory usage. It offers greater flexibility and is ideal for comprehensive monitoring across various environments.

- Collect additional system-level metrics such as RAM, processes, etc...
- Collect logs to send to CloudWatch Logs
- Centralized configuration using SSM Parameter Store

306. What are the CloudWatch Unified Agent - Metrics?

Ans ->

- Collected directly on your Linux server / EC2 instance
- **CPU** (active, guest, idle, system, user, steal)
- **Disk metrics** (free, used, total), **Disk IO** (write, reads, bytes, iops)
- **RAM** (free, inactive, used, total, cached)
- **Netstat** (number of TCP and UDP connections, net packets, bytes)
- **Processes** (total, dead, bloqued, idle, running, sleep)
- **Swap Space** (free, used, used %)
- **Reminder:** out-of-the box metrics for EC2 - disk, CPU, network (high level)

307. Explain about CloudWatch Alarms.

Ans ->

AWS CloudWatch Alarms are used to monitor metrics from various AWS services and applications, allowing users to set thresholds on metrics like CPU utilization or request

counts. When these thresholds are breached, CloudWatch triggers alarms that can notify via Amazon SNS, execute Auto Scaling actions, or invoke AWS Lambda functions to automate responses. This proactive monitoring helps maintain the health, performance, and availability of AWS resources by alerting users to potential issues before they impact operations.

Key points:

- Alarms are used to trigger notifications for any metric
- Various options (sampling, %, max, min, etc...)
- **Alarm States:**
  - OK
  - INSUFFICIENT_DATA
  - ALARM
- **Period:**
  - Length of time in seconds to evaluate the metric
  - **High resolution custom metrics:** 10 sec, 30 sec or multiples of 60 sec

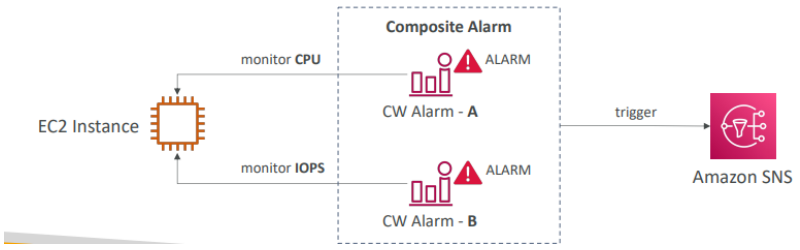
308. What is CloudWatch Alarms - Composite Alarms?

Ans ->

Amazon CloudWatch Composite Alarms are a feature within AWS CloudWatch that allow you to create alarms based on the state of multiple other alarms. Instead of monitoring a single metric or threshold, composite alarms let you aggregate and evaluate the states of multiple alarms to trigger a new alarm.

Key points:

- CloudWatch Alarms are on a single metric
- Composite Alarms are monitoring the states of multiple other alarms
- AND and OR conditions
- Helpful to reduce "alarm noise" by creating complex composite alarms



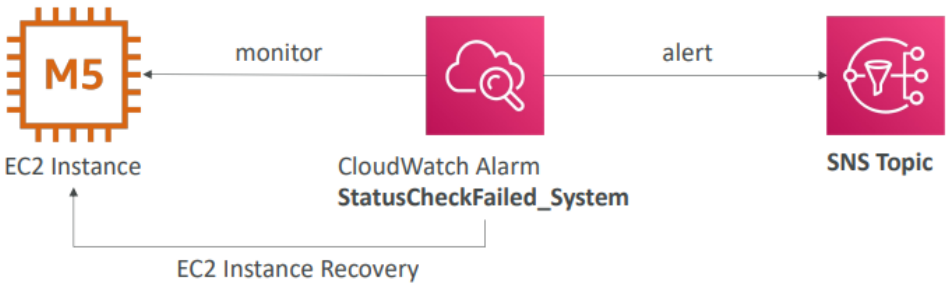
309. EC2 Instance Recovery

Ans ->

Status Check:

- Instance status = check the EC2 VM

- System status = check the underlying hardware



Recovery: Same Private, Public, Elastic IP, metadata, placement group

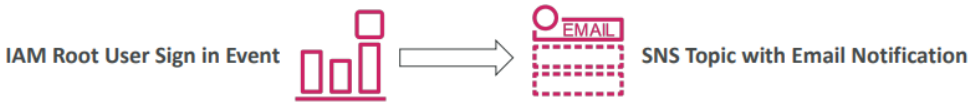
310. What is Amazon EventBridge?

Ans ->

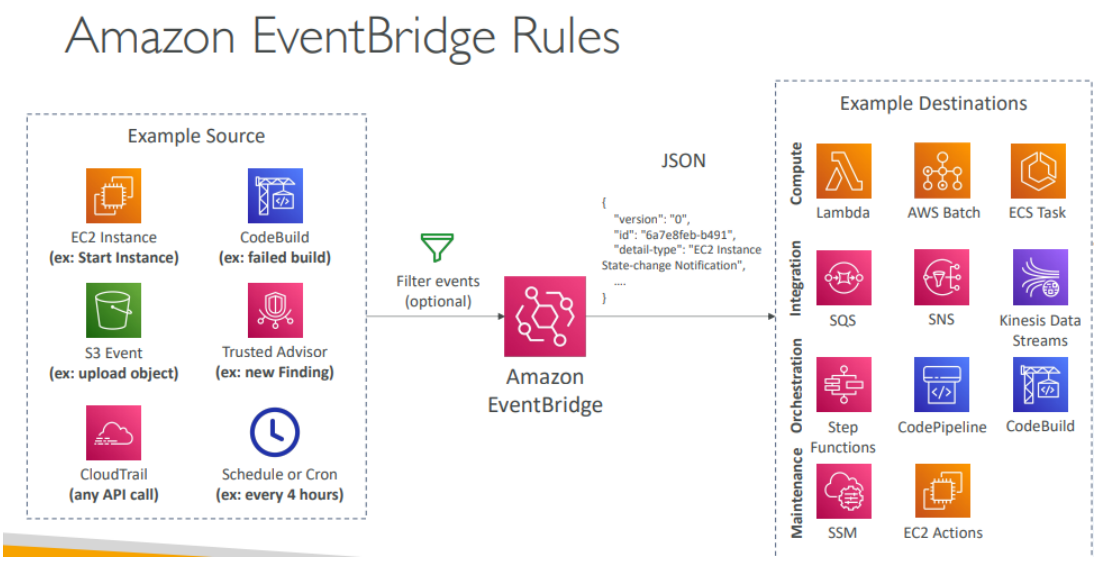
Amazon EventBridge (formerly known as CloudWatch Events) is a serverless event bus service from AWS that lets you set up automatic responses to events, like when data changes or certain actions happen. It helps you connect different parts of your application by routing these events to where they need to go, like triggering a function or sending a notification, all without having to manage the servers yourself.

Schedule: Cron jobs (scheduled scripts)

Event Pattern: Event rules to react to a service doing something



Trigger Lambda functions, send SQS/SNS messages...



311. What is Event Bus?

Ans ->

An **event bus** is a system that routes and manages events within an application. It acts like a central hub where events are published and then distributed to various components or services that need to respond to those events. This allows different parts of an application to communicate and react to changes or actions in a decoupled and scalable way.

Key points:

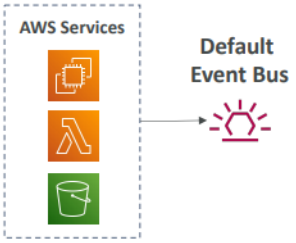
- Event buses can be accessed by other AWS accounts using Resource-based Policies
- You can archive events (all/filter) sent to an event bus (indefinitely or set period)
- Ability to replay archived events

312. How many types of Event Buses are there in AWS EventBridge?

Ans ->

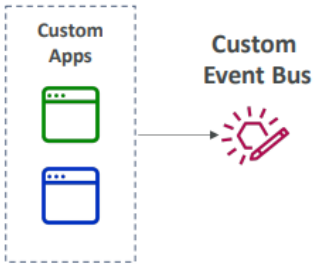
In AWS EventBridge, there are three main types of event buses:

- **Default Event Bus:** This is automatically created for you and captures events from AWS services. It handles events from services like S3, EC2, Lambda, and more.



- **Partner Event Buses:** These are provided by AWS SaaS partners and receive events from integrated SaaS applications. They enable you to integrate and react to events from third-party services within EventBridge.

- **Custom Event Buses:** You can create these to handle events from your own applications or specific sources. Custom event buses allow you to define and manage events that are specific to your use case.



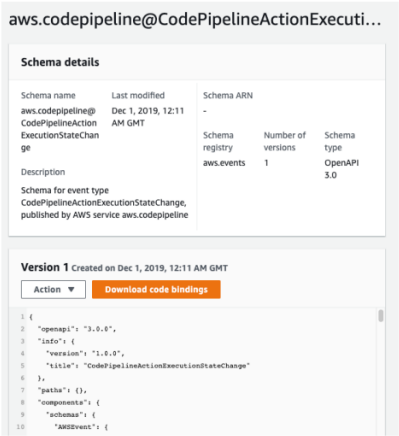
312. What is Amazon EventBridge - Schema Registry?

Ans ->

The **AWS EventBridge Schema Registry** is a feature that helps you manage and **understand the structure of events** flowing through your EventBridge event buses. It automatically detects and stores the structure of your events, lets you view and update these structures, and even helps generate code based on them. This makes it easier to work with and ensure the quality of the data your application handles.

Key points:

- EventBridge can analyze the events in your bus and infer the schema
- The Schema Registry allows you to generate code for your application, that will know in advance how data is structured in the event bus
- Schema can be versioned



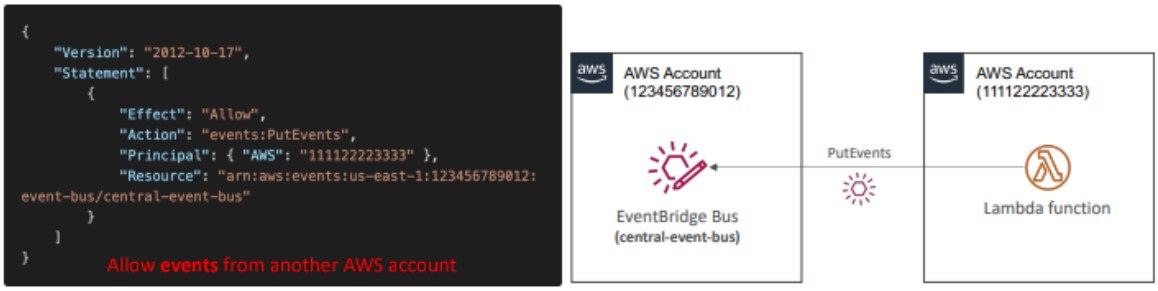
213. What is Amazon EventBridge - Resource-based Policy?

Ans ->

An **Amazon EventBridge resource-based policy** is a set of permissions that controls which AWS services or accounts can interact with your EventBridge resources, like event buses. It specifies who can put events on your bus, read from it, or perform other actions, helping you manage access and ensure only authorized entities can use your event bus.

Key points:

- Manage permissions for a specific Event Bus
- **Example:** allow/deny events from another AWS account or AWS region
- **Use case:** aggregate all events from your AWS Organization in a single AWS account or AWS region



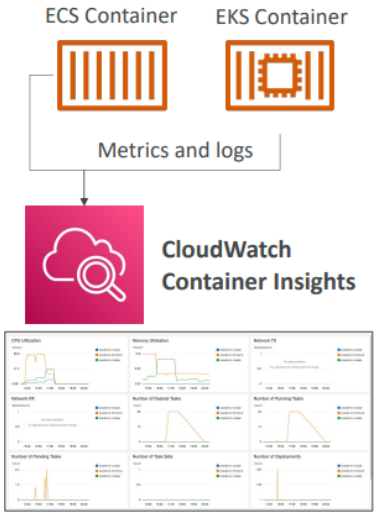
314. What is CloudWatch Container Insights?

Ans ->

CloudWatch Container Insights provides monitoring and troubleshooting capabilities for containerized applications. It collects, aggregates, and visualizes metrics, logs, and performance data from containers and container orchestration platforms like Amazon ECS, EKS, and Docker. This helps you gain insights into container health, resource usage, and application performance.

Key points:

- Collect, aggregate, summarize metrics and logs from containers
- Available for containers on…
    - Amazon ECS
    - Amazon EKS
    - Kubernetes platforms on EC2
    - Fargate (both for ECS and EKS)
- **In Amazon EKS and Kubernetes, CloudWatch Insights is using a containerized version of the CloudWatch Agent to discover containers**



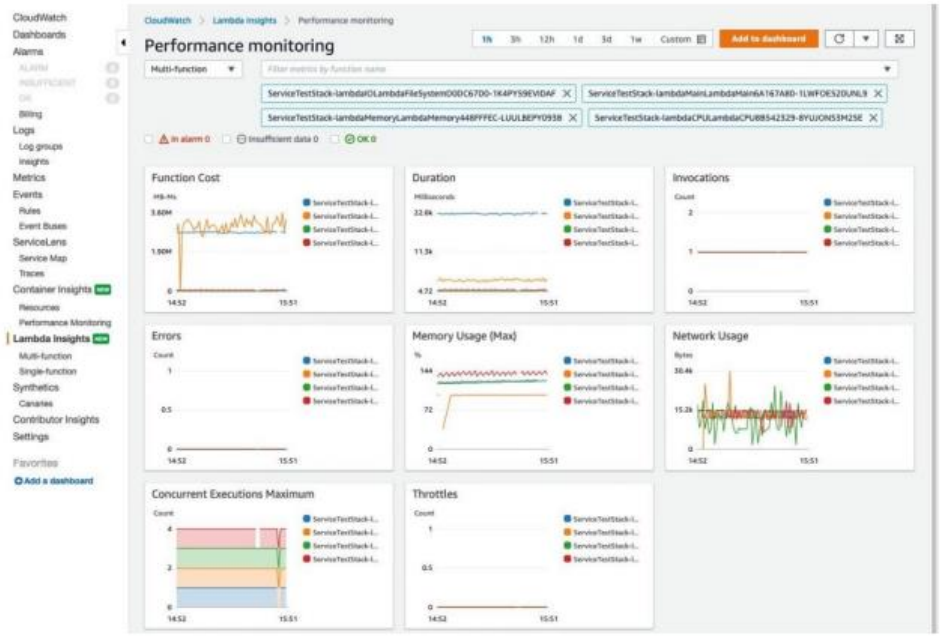
315. What is CloudWatch Lambda Insights?

Ans ->

CloudWatch Lambda Insights provides detailed monitoring and performance data for AWS Lambda functions. It offers in-depth visibility into Lambda functions by collecting and visualizing metrics, logs, and execution data. With Lambda Insights, you can track performance metrics like invocation counts, duration, errors, and request/response details, as well as diagnose issues with enhanced logging and tracing. This helps you understand how your Lambda functions are performing and quickly identify and troubleshoot problems.

Key points:

- Monitoring and troubleshooting solution for serverless applications running on AWS Lambda
- Collects, aggregates, and summarizes system-level metrics including CPU time, memory, disk, and network
- Collects, aggregates, and summarizes diagnostic information such as cold starts and Lambda worker shutdowns
- Lambda Insights is provided as a Lambda Layer



316. What is CloudWatch Contributor Insights?

Ans ->

CloudWatch Contributor Insights analyzes and visualizes high-cardinality data, like logs from applications or services, to identify **top contributors to performance issues or usage patterns**. It helps you understand which entities, such as users or IP addresses, are driving the most traffic or causing the most errors, making it easier to diagnose and address performance bottlenecks.

Key points:

- Analyze log data and create time series that display contributor data.
  - See metrics about the **top-N contributors**
  - The total number of unique contributors, and their usage.
- This helps you find top talkers and understand who or what is impacting system performance.
- Works for any AWS-generated logs (VPC, DNS, etc...)
- For example, you can **find bad hosts**, identify the **heaviest network users**, or find the URLs that generate the most errors.
- You can build your rules from scratch, or you can also use sample rules that AWS has created - leverages your CloudWatch Logs

317. What is CloudWatch Application Insights?

Ans ->

CloudWatch Application Insights helps you monitor and manage the health of your applications by automatically discovering and visualizing application resources, metrics, and logs. It provides insights into application performance and dependencies,

detects anomalies, and helps you diagnose issues quickly with integrated dashboards and alerts.

Key points:

- Provides automated dashboards that show potential problems with monitored applications, to help isolate ongoing issues.
- Your application run on Amazon EC2 instances with select technologies only (Java, .NET, Microsoft IIS Web Server, databases...)
- And you can use other AWS resources such as Amazon EBS, RDS ELB, ASG, Lambda, SQS, DynamoDB, S3 bucket, ECS, EKS, SNS, API Gateway...
- Powered by SageMaker
- Enhanced visibility into your application health to reduce the time it will take you to troubleshoot and repair your applications
- Findings and alerts are sent to Amazon EventBridge and SSM OpsCenter.

318. CloudWatch Insights and Operational Visibility.

Ans ->

CloudWatch Container Insights:

- ECS, EKS, Kubernetes on EC2, Fargate, needs agent for Kubernetes
- Metrics and logs

CloudWatch Lambda Insights:

- Detailed Metrics to troubleshoot serverless applications

CloudWatch Contributors Insights:

- Find "Top-N" Contributors through CloudWatch Logs

CloudWatch Application Insights:

- Automatic dashboard to troubleshoot your application and relate AWS services

319. What is AWS CloudTrail?

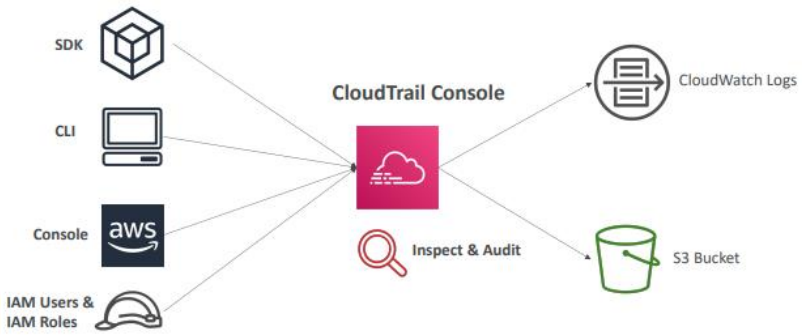
Ans ->

AWS CloudTrail is a service that logs and monitors **all API calls and activities** across your AWS account, providing a detailed record of actions taken by users, roles, and

AWS services. It helps with **security**, **compliance**, and operational **auditing** by enabling you to track changes to resources, detect unusual activity, and analyze trends over time. CloudTrail's logs can be stored, searched, and analyzed to ensure transparency and accountability in your AWS environment.

Key points:

- **Provides governance, compliance and audit for your AWS Account**
- CloudTrail is enabled by default!
- Get an history of events / API calls made within your AWS Account by:
  - Console
  - SDK
  - CLI
  - AWS Services
- Can put logs from CloudTrail into CloudWatch Logs or S3
- A trail can be applied to All Regions (default) or a single Region.
- If a resource is deleted in AWS, investigate CloudTrail first!



320. What are CloudTrail Events?

Ans ->

Management Events:

- Operations that are performed on resources in your AWS account
- **Examples:**
  - Configuring security (IAM AttachRolePolicy)
  - Configuring rules for routing data (Amazon EC2 CreateSubnet)
  - Setting up logging (AWS CloudTrail CreateTrail)
- **By default, trails are configured to log management events.**
- Can separate Read Events (that don't modify resources) from Write Events (that may modify resources)

Data Events:

- **By default, data events are not logged (because high volume operations)**

- Amazon S3 object-level activity (ex: GetObject, DeleteObject, PutObject): can separate Read and Write Events
- AWS Lambda function execution activity (the Invoke API)

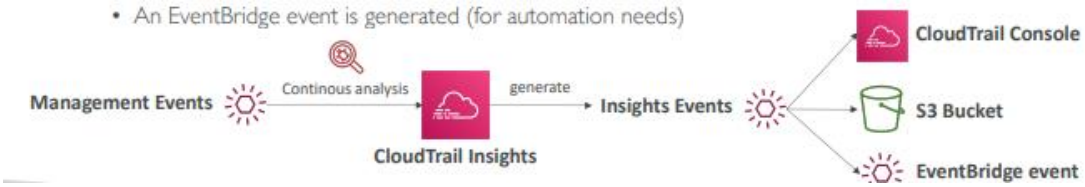
321. What is CloudTrail Insights?

Ans ->

CloudTrail Insights automatically detects and highlights unusual or anomalous activities within your AWS account. It analyzes your normal account activity patterns and identifies deviations, such as spikes in resource provisioning or unusual API call patterns. This helps you quickly spot potential security issues or operational problems, allowing for faster investigation and resolution.

Key points:

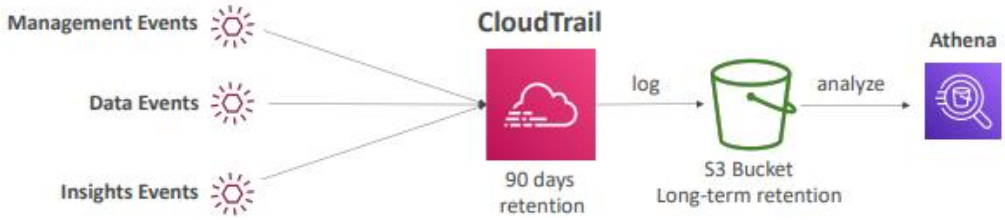
- Enable CloudTrail Insights to detect unusual activity in your account:
  - inaccurate resource provisioning
  - hitting services limits
  - Bursts of AWS IAM actions
  - Gaps in periodic maintenance activity
- CloudTrail Insights analyzes normal management event to create a baseline
- And then continuously analyzes write events to detect unusual patterns:
  - Anomalies appear in the CloudTrail console
  - Event is sent to Amazon S3
  - An EventBridge event is generated (for automation needs)



322. Key points about CloudTrail Events Retention.

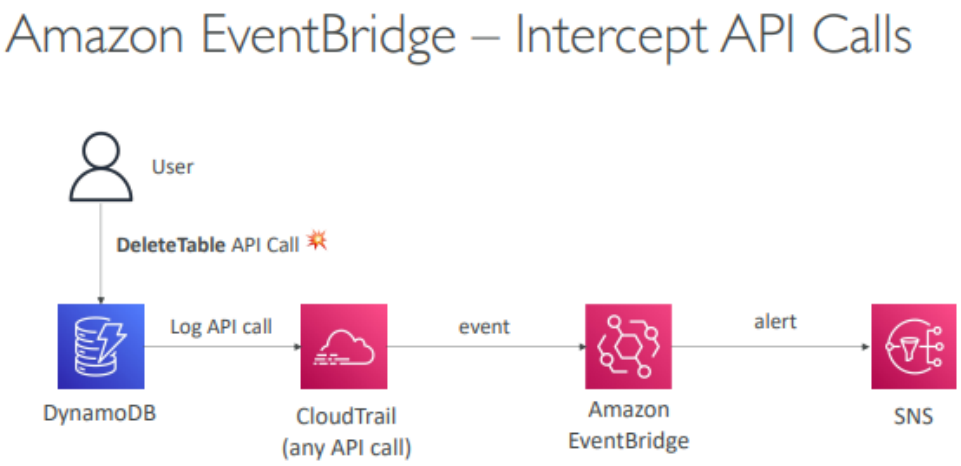
Ans ->

- Events are stored for **90 days** in CloudTrail
- To keep events beyond this period, log them to S3 and use Athena to query those data.



323. Amazon CloudTrail integration with EventBridge.

Ans ->



324. What is AWS Config?

Ans ->

AWS Config is a service that continuously monitors and records the configuration of your AWS resources, providing a detailed history of changes. It helps you assess compliance, track resource relationships, and set rules to automatically check if your environment meets security and policy standards.

Key points:

- Helps with **auditing** and recording **compliance** of your AWS resources
- Helps record configurations and changes over time
- **Questions that can be solved by AWS Config:**
  - Is there unrestricted SSH access to my security groups?
  - Do my buckets have any public access?
  - How has my ALB configuration changed over time?
- You can retrieve alerts (SNS notifications) for any changes
- AWS Config is a per-region service
- Can be aggregated across regions and accounts
- Possibility of storing the configuration data into S3 (analyzed by Athena)

325. Key points about AWS Config Rules.

Ans ->

AWS Config Rules are customizable rules that you can create to automatically check the compliance of your AWS resources with specific policies or best practices. These rules continuously evaluate the configuration of your resources and trigger notifications if they detect non-compliance or changes.

Key points:

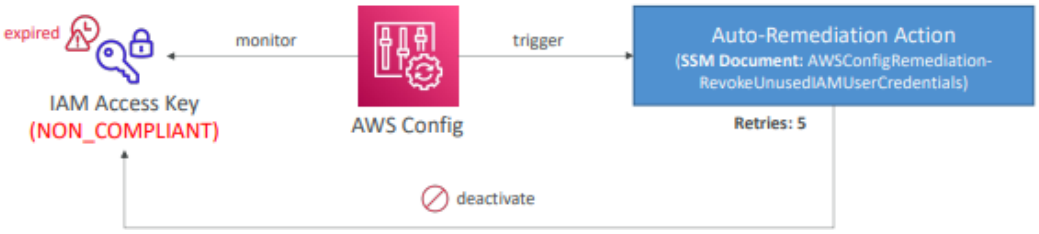
- Can use AWS managed config rules (over 75 rules are there to use)
- Can make custom config rules (must be defined in AWS Lambda)
  - Ex: evaluate if each EBS disk is of type gp2
  - Ex: evaluate if each EC2 instance is t2.micro
- Rules can be evaluated / triggered:
  - For each config change
  - And / or: at regular time intervals
- **AWS config Rules does not prevent actions from happening (no deny)**

Pricing: no free tier, $0.003 per configuration item recorded per region, $0.001 per config rule evaluation per region.

326. Key points about AWS Config Rules - Remediations.

Ans ->

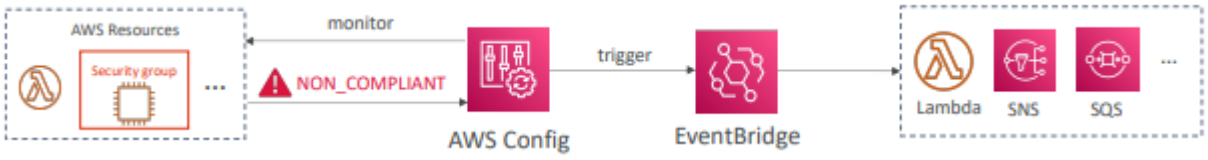
- Automate remediation of non-compliant resources using **SSM Automation Documents**
- Use AWS-Managed Automation Documents or create custom Automation Documents
  - Tip: you can create custom Automation Documents that invokes Lambda function
- You can set **Remediation retries** if the resource is still non-compliant after auto-remediation



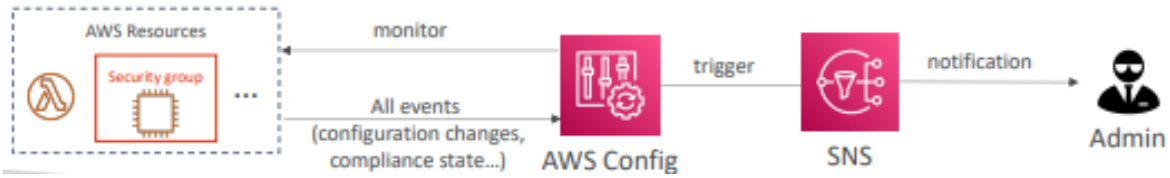
327. Key points about Config Rules - Notifications.

Ans ->

Use EventBridge to trigger notifications when AWS resources are non-compliant



Ability to send configuration changes and compliance state notifications to SNS (all events - use SNS Filtering or filter at client-side)



328. CloudWatch VS CloudTrail VS Config

Ans ->

CloudWatch:

- Performance monitoring (metrics, CPU, network, etc...) & dashboards
- Events & Alerting
- Log Aggregation & Analysis

CloudTrail:

- Record API calls made within your account by everyone
- Can define trails for specific resources
- Global Service

Config:

- Record configuration changes
- Evaluate resources against compliance rules
- Get timeline of changes and compliance

329. For an Elastic Load Balancer

Ans ->

CloudWatch:

- Monitoring Incoming connections metric
- Visualize error codes as a % over time
- Make a dashboard to get an idea of your load balancer performance

Config:

- Track security group rules for the Load balancer
- Track configuration changes for the Load Balancer
- Ensure an SSL certificate is always assigned to the Load Balancer (compliance)

CloudTrail:

- Track who made any changes to the Load Balancer with API calls

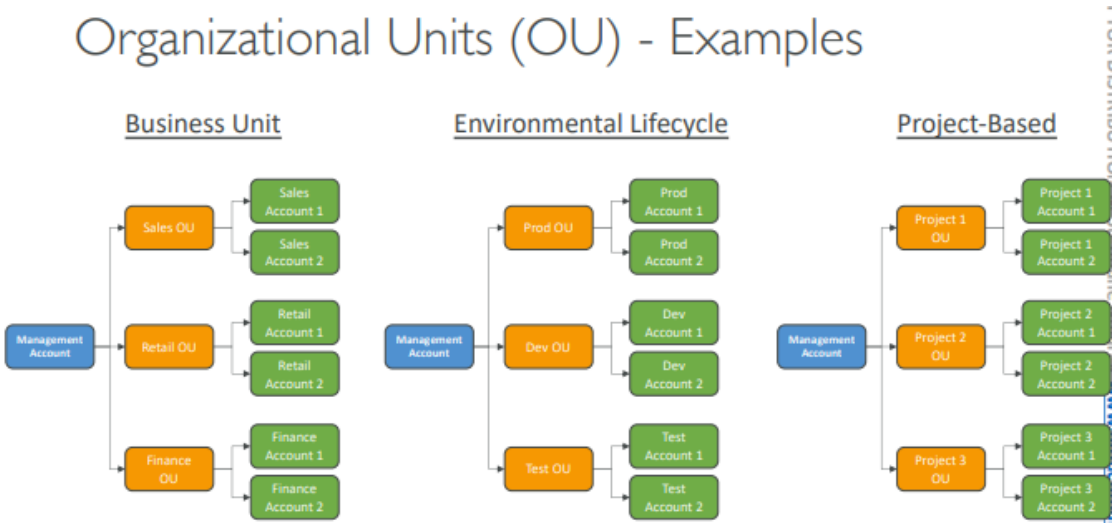
330. What is AWS Organizations?

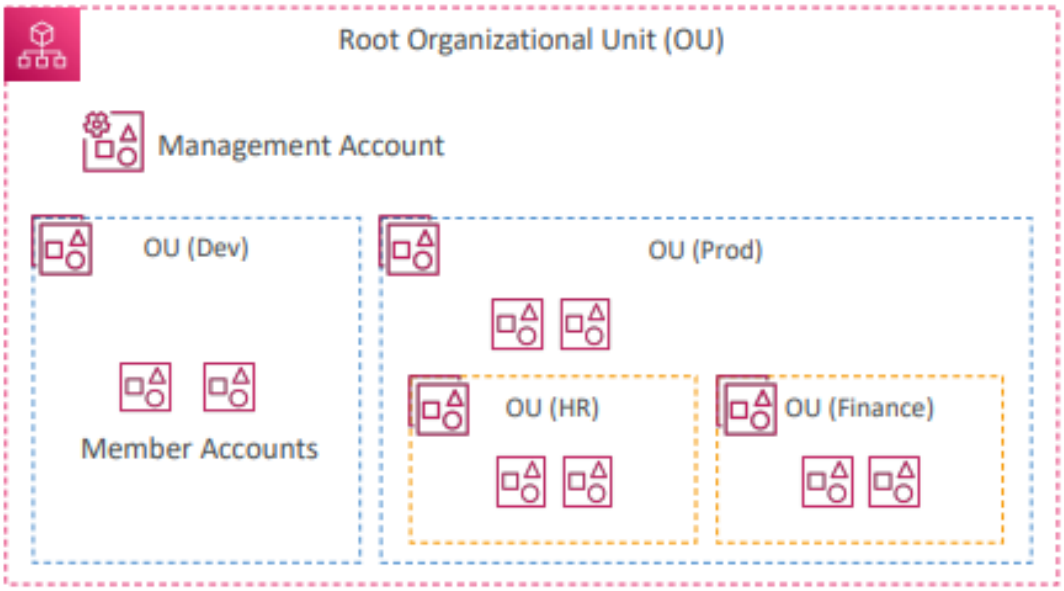
Ans ->

AWS Organizations is a service that lets you centrally manage multiple AWS accounts under a single organization. It allows you to group accounts, apply policies, and consolidate billing, making it easier to control security and costs across your AWS environment. This service is ideal for managing permissions, compliance, and costs efficiently in organizations with multiple AWS accounts.

Key points:

- Global service
- Allows to manage multiple AWS accounts
- The main account is the management account
- Other accounts are member accounts
- Member accounts can only be part of one organization
- Consolidated Billing across all accounts - single payment method
- Pricing benefits from aggregated usage (volume discount for EC2, S3...)
- Shared reserved instances and Savings Plans discounts across accounts
- API is available to automate AWS account creation





331. What are the advantages of AWS Organizations?

Ans ->

- Easily manage multiple AWS accounts from a single console
- Multi Account vs One Account Multi VPC
- Use tagging standards for billing purposes
- Consolidate billing across accounts to optimize costs
- Enable CloudTrail on all accounts, send logs to central S3 account
- Send CloudWatch Logs to central logging account
- Apply Service Control Policies (SCPs) to enforce security across accounts
- Establish Cross Account Roles for Admin purposes
- Easily add and manage new accounts as your organization grows

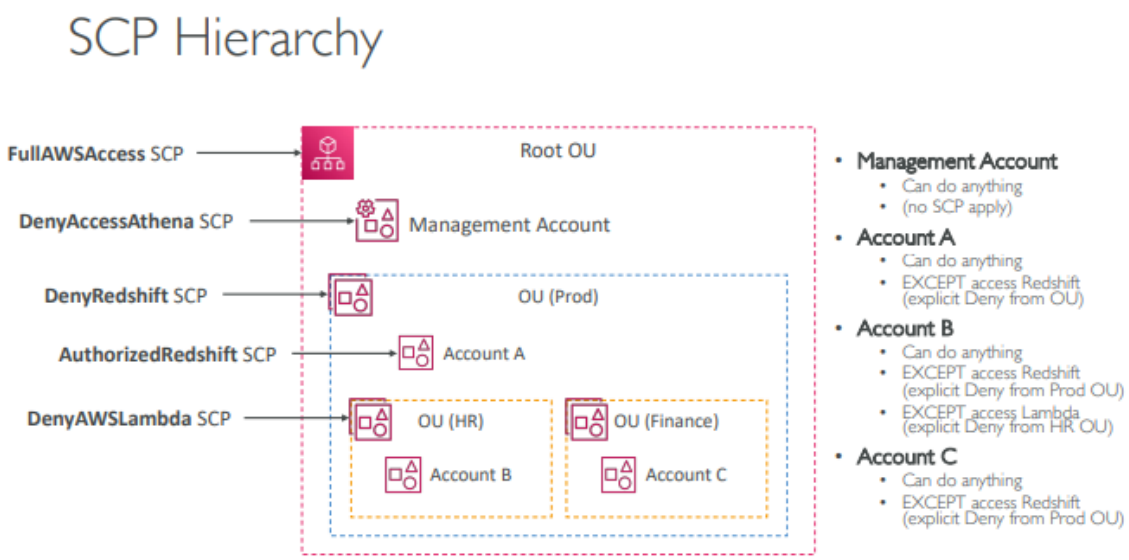
332. Explain about Service Control Policies (SCP).

Ans ->

Service Control Policies (SCPs) in AWS Organizations are policies that define the maximum permissions allowed for accounts in an organizational unit (OU) or for individual accounts. SCPs allow you to enforce specific security and compliance guidelines by restricting what actions users and roles can perform within those accounts, providing an additional layer of access control across your organization.

Key points:

- IAM policies applied to OU or Accounts to restrict Users and Roles
- **They do not apply to the management account (full admin power)**
- Must have an explicit allow from the root through each OU in the direct path to the target account (does not allow anything by default - like IAM)

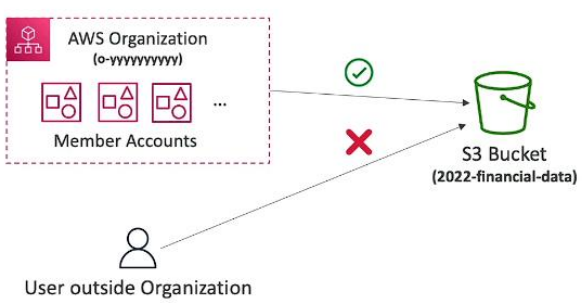


333. Explain about Resource Policies & AWS:PrincipalOrgID.

Ans ->

Resource Policies in AWS are JSON-based policies attached directly to resources like S3 buckets or KMS keys, defining who can access the resource and what actions they can perform.

AWS:PrincipalOrgID is a condition key in these policies that restricts access to resources by allowing only requests from accounts within a specific AWS Organization, ensuring access is limited to trusted accounts within your organization.




```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": ["s3:PutObject", "s3:GetObject"],
            "Resource": "arn:aws:s3:::2022-financial-data/*",
            "Condition": {
                "StringEquals": {
                    "aws:PrincipalOrgID": ["o-yyyyyyyyyy"]
                }
            }
        }
    ]
}
```

334. IAM for S3 (object level permission).

Ans ->

IAM for S3

- s3:ListBucket permission applies to arn:aws:s3:::test
- => bucket level permission

- s3:GetObject, s3:PutObject, s3:DeleteObject applies to arn:awn:s3:::test/*
- => object level permission

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": ["s3:ListBucket"],
            "Resource": "arn:aws:s3:::test"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": "arn:aws:s3:::test/*"
        }
    ]
}
```

335. IAM Roles vs Resource-Based Policies.

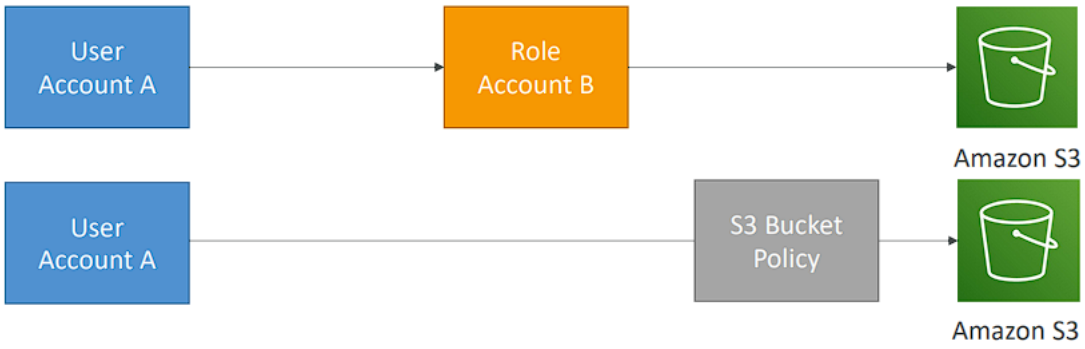
Ans ->

- **IAM Roles:** Roles are assigned to AWS entities like users, services, or applications, allowing them to assume a specific set of permissions temporarily. Roles are used

to delegate access across AWS accounts or services without sharing long-term credentials.

- **Resource-Based Policies:** These policies are directly attached to AWS resources like S3 buckets, SNS topics, or Lambda functions, specifying who can access the resource and what actions they can perform. They can grant cross-account access without requiring the use of roles.

In essence, IAM Roles are attached to entities to define what they can do, while Resource-Based Policies are attached to resources to define who can access them.



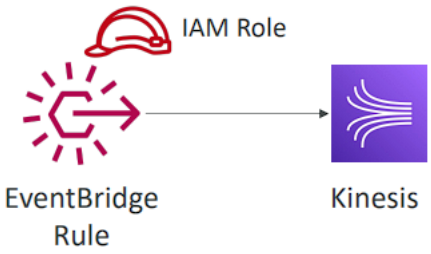
336. Amazon EventBridge - Security.

Ans ->

When a rule runs, it needs permissions on the target

Resource-based policy applies to: Lambda, SNS, SQS, S3 buckets, API Gateway...

IAM Role applies to: Kinesis streams, Systems Manager Run Command, ECs task...



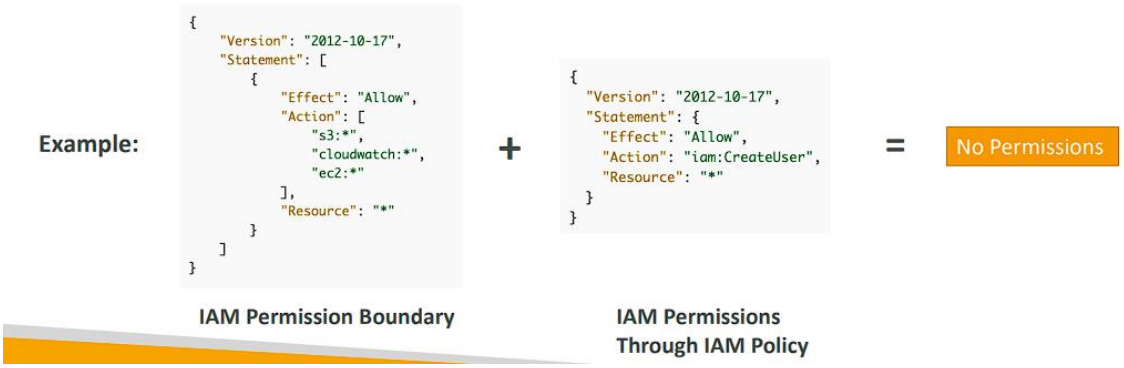
337. Explain about IAM Permission Boundaries.

Ans ->

IAM Permission Boundaries are a way to set the maximum permissions that an IAM role or user can have in AWS. When a permission boundary is applied, it acts as a ceiling, limiting what actions the user or role can perform, even if other attached policies grant broader permissions.

This is useful for controlling the scope of access, especially when delegating permission management, as it allows you to enforce strict limits on what users or roles can do, regardless of the permissions granted by other policies they may have.

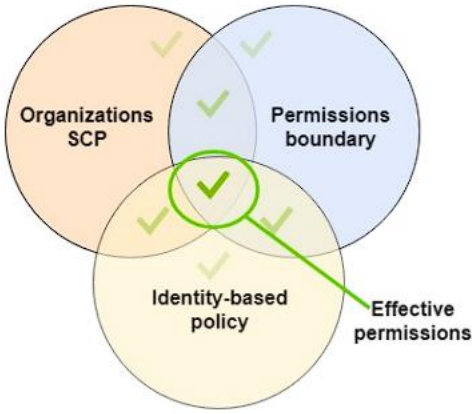
- IAM Permission Boundaries are **supported for users and roles (not groups)**
- Advanced feature to use a managed policy to set the maximum permissions an IAM entity can get.



338. What are the use cases of IAM Permission Boundaries?

Ans ->

- Can be used in combination of AWS Organizations SCP

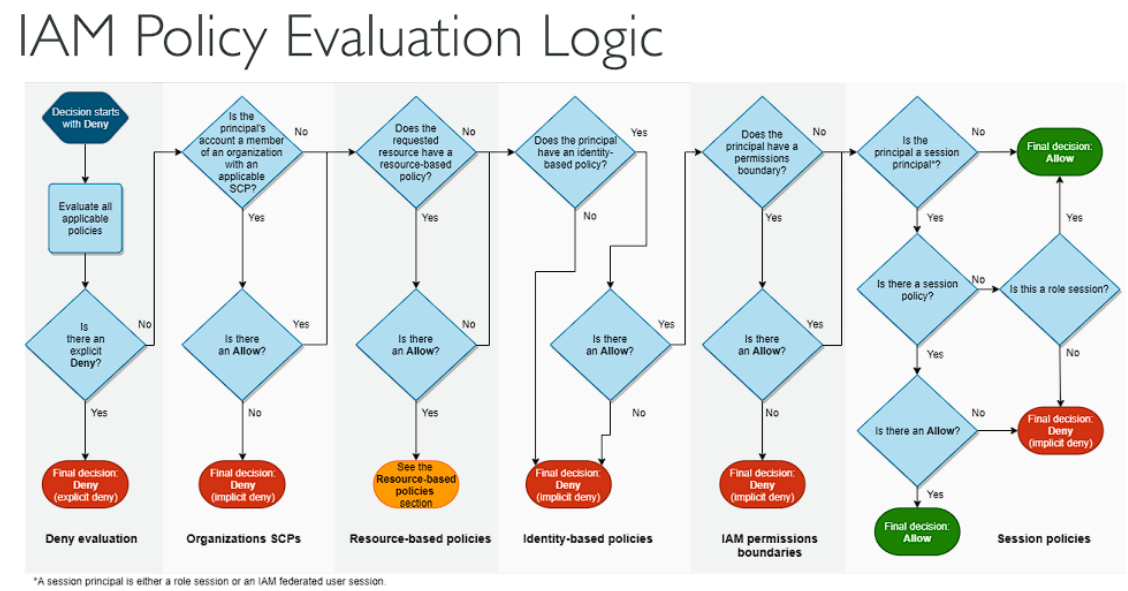


https://docs.aws.amazon.com/IAM/latest/UserGuide/access_policies_boundaries.html

- Delegate responsibilities to non-administrators within their permission boundaries, for example create new IAM users.
- Allow developers to self-assign policies and manage their own permissions, while making sure they can't "escalate" their privileges (=make themselves admin)
- Useful to restrict one specific user (instead of a whole account using Organizations & SCP)

339. IAM Policy Evaluation logic Flow Chart.

Ans ->



https://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_evaluation-logic.html

340. Explain the below given IAM Policy.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": "sqs:*",
      "Effect": "Deny",
      "Resource": "*"
    },
    {
      "Action": [
        "sqs:DeleteQueue"
      ],
      "Effect": "Allow",
      "Resource": "*"
    }
  ]
}
```

Ans ->

Can you perform sqs:CreateQueue?: No, because the policy has an **explicit deny**.

Can you perform sqs:DescribeInstances?: No, because the policy has an **explicit deny**, and it doesn't matter whether you've added broader permission in the next block or not.

341. What is AWS IAM Identity Center?

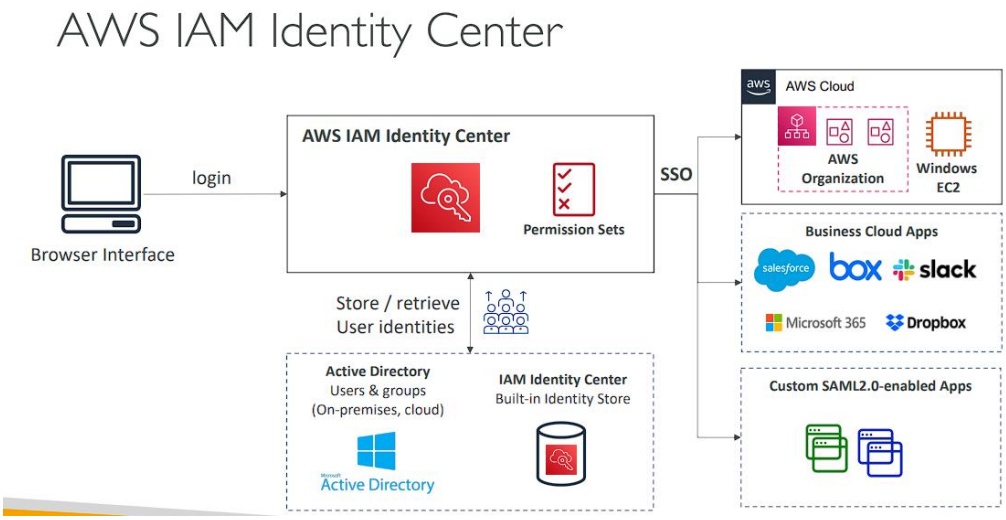
Ans ->

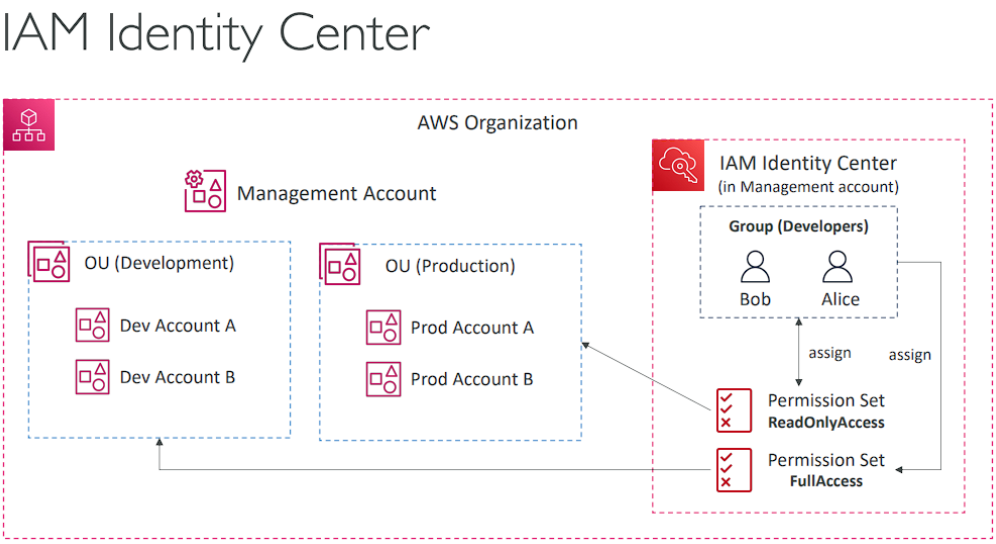
AWS IAM Identity Center (formerly known as AWS Single Sign-On or AWS SSO) is a service that helps you manage and simplify access to AWS accounts and applications. It lets users log in once to access multiple accounts and services without needing to enter their credentials again. By managing user permissions centrally, it makes it easier to control who can access what, integrates with existing identity systems, and provides tools for tracking user activity and ensuring security.

Key points:

- **One login (single sign-on) for all your:**
  - AWS accounts in AWS Organizations
  - Business cloud applications (e.g., Salesforce, Box, Microsoft 365...)
  - SAML2.0-enabled applications
  - EC2 Windows Instances

- **Identity providers:**
  - Built-in identity store in IAM Identity Center
  - **3rd party:** Active Directory (AD), OneLogin, Okta...





342. AWS IAM Identity Center Fine-grained Permissions and Assignments.

Ans ->

Multi-Account Permissions:

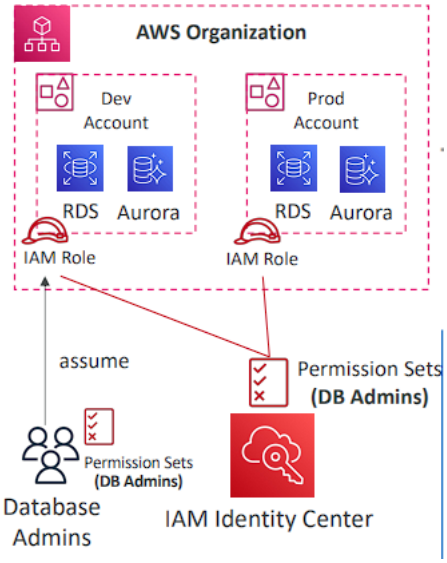
- Manage access across AWS accounts in your AWS Organization
- Permission Sets - a collection of one or more IAM Policies assigned to users and groups to define AWS access

Application Assignments:

- SSO access to many SAML 2.0 business applications (Salesforce, Box, Microsoft 365...)
- Provide required URLs, certificates, and metadata

Attribute-Based Access Control (ABAC):

- Fine-grained permissions based on user's attributes stores in IAM Identity Center Identity Store
- **Example:** cost center, title, locale, ...
- **Use case:** Define permissions once, then modify AWS access by changing the attributes



343. What is Microsoft Active Directory (AD)?

Ans ->

Microsoft Active Directory is a system used by organizations to manage and organize their computer networks. It helps administrators control who has access to what resources, such as files and printers, and ensures that users can log in and use their accounts securely. Think of it as a big digital directory that keeps track of all the users, computers, and permissions in a network, making it easier to manage everything from one central place.

Key points:

- Found on any Windows Server with AD Domain Services
- **Database of objects:** User Accounts, Computers, Printers, File Shares, Security Groups
- Centralized security management, create account, assign permissions
- Objects are organized in trees
- A group of trees is a forest

344. Explain about AWS Directory Services.

Ans ->

AWS Managed Microsoft AD:

- A fully managed service that provides a Microsoft Active Directory (AD) in the AWS cloud.
- Create your own AD in AWS, manage users locally, supports MFA
- Establish "trust" connections with your on-premise AD

AD Connector:

- A directory gateway that connects your AWS environment with an existing on-premises Microsoft Active Directory.
- Directory Gateway (proxy) to redirect to on-premise AD, supports MFA
- Users are managed on the on-premise AD

Simple AD:

- A simpler, cost-effective directory service that provides basic Active Directory-compatible features.
- AD-compatible managed directory on AWS
- Cannot be joined with on-premise AD

345. Explain about AWS Control Tower.

Ans ->

AWS Control Tower is a service that helps you set up and manage multiple AWS accounts easily. It provides a set of automated tools and best practices to create a secure, well-organized AWS environment. With Control Tower, you can quickly establish accounts, enforce company policies, and keep track of your AWS setup from a **central dashboard**, making it simpler to maintain good practices and compliance across your entire AWS infrastructure.

- Easy way to set up and govern a secure and compliant multi-account AWS environment based on best practices
- AWS Control Tower uses AWS Organizations to create accounts

Benefits:

- Automate the setup of your environment in a few clicks
- Automate ongoing policy management using guardrails
- Detect policy violation and remediate them
- Monitor compliance through an **interactive dashboard**

346. What is AWS Control Tower - Guardrails?

Ans ->

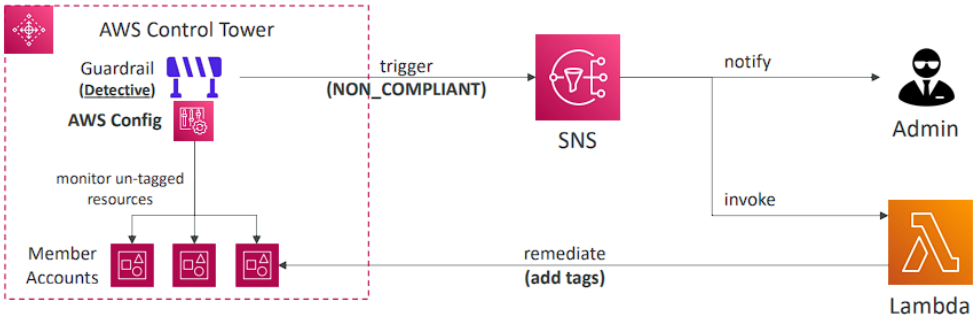
In **AWS Control Tower**, **Guardrails** are a set of policies designed to ensure that your AWS environment adheres to best practices and compliance standards. They come in two main types: **Preventive Guardrails** and **Detective Guardrails**.

Preventive Guardrails:

- **Purpose:** To prevent non-compliant actions from occurring in your AWS environment.
- using SCPs (e.g., Restrict Regions across all your accounts)

Detective Guardrails:

- **Purpose:** To identify and alert you about non-compliant actions or configurations that have already occurred.
- using AWS Config (e.g., identify untagged resources)



347. Explain about AWS KMS.

Ans ->

AWS Key Management Service (KMS) is a managed service that lets you create and control **cryptographic keys used to encrypt and decrypt data** across your AWS resources. It integrates with services like S3, EBS, and RDS, providing encryption for data at rest and in transit. KMS also supports automatic key rotation, granular access control using IAM and resource policies, and detailed logging of key usage through AWS CloudTrail, making it essential for securing sensitive data and meeting compliance requirements.

Kye points:

- AWS manages encryption keys for us
- Fully integrated with IAM for authorization
- Easy way to control access to your data
- Able to audit KMS Key usage using CloudTrail
- Seamlessly integrated into most AWS services (EBS, S3, RDS, SSM...)
- **Never ever store your secrets in plaintext, especially in your code!**
    - KMS Key Encryption also available through API calls (SDK, CLI)
    - Encrypted secrets can be stored in the code / environment variables

348. Explain about KMS Keys Types.

Ans ->

- **KMS Keys, formerly known as KMS Customer Master Key**
- **Symmetric (AES-256 keys):**
  - Single encryption key that is used to Encrypt and Decrypt
  - AWS services that are integrated with KMS use Symmetric CMKs
  - You never get access to the KMS Key unencrypted (must call KMS API to use)
- **Asymmetric (RSA & ECC key pairs):**
  - Public key (Encrypt) and Private key (Decrypt) pair
  - Used for Encrypt/Decrypt, or Sign/Verify operations
  - The public key is downloadable, but you can't access the Private Key unencrypted
  - **Use case:** encryption outside of AWS by users who can't call the KMS API

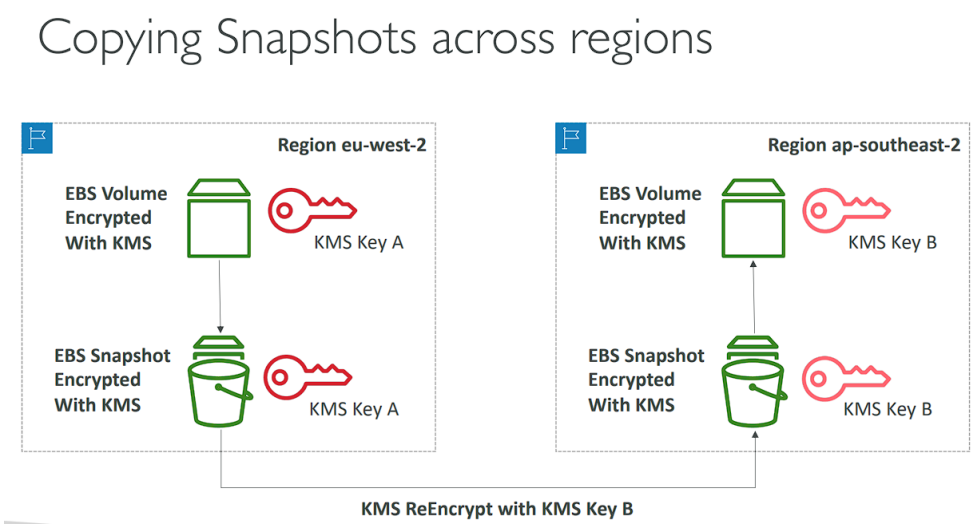
349. Types of KMS keys.

Ans ->

- **AWS Owned Keys (free):** SSE-S3, SSE-SQS, SSE-DDB (default key)
- **AWS Managed Key (free):** aws/service-name, example: aws/rds or aws/ebs
- **Customer Managed Keys create in KMS:** $1 / month
- **Customer Managed Keys imported (must be symmetric key):** $1 / month
  - +pay for API call to KMS ($0.03 / 10000 calls)
- **Automatic Key rotation:**
  - **AWS-managed KMS Key:** (must be enabled) automatic & on-demand
  - **Imported KMS Key:** only manual rotation possible using alias

350. Process of copying KMS encrypted Snapshots across regions.

Ans ->



351. Explain about KMS Key Policies.

Ans ->

- Control access to KMS keys, "similar" to S3 bucket policies
- **Difference:** you cannot control access without them

- **Default KMS key Policy:**
  - Created if you don't provide a specific KMS Key Policy
  - Complete access to the key to the root user = entire AWS account
- **Custom KMS Key Policy:**
  - Define users, roles that can access the KMS key
  - Define who can administer the key
  - Useful for cross-account access of your KMS key

351. Process of Copying Snapshots across accounts.

Ans ->

- Create a Snapshot, encrypted with your own KMS Key (Customer Managed Key)
- Attach a KMS Key Policy to authorize cross-account access
- Share the encrypted snapshot
- (in target) Create a copy of the Snapshot, encrypt it with a CMK in your account
- Create a volume from the snapshot

```
{
    "Sid": "Allow use of the key with destination account",
    "Effect": "Allow",
    "Principal": {
      "AWS": "arn:aws:iam::TARGET-ACCOUNT-ID:role/ROLENAME"
    },
    "Action": [
      "kms:Decrypt",
      "kms:CreateGrant"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "kms:ViaService": "ec2.REGION.amazonaws.com",
        "kms:CallerAccount": "TARGET-ACCOUNT-ID"
      }
    }
}
```

KMS Key Policy

352. What are KMS Multi-Region Keys?

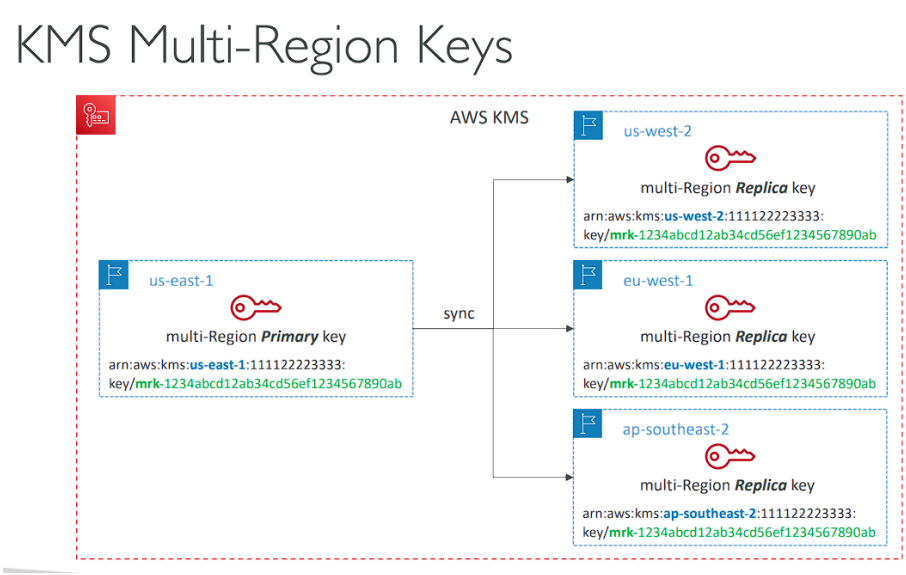
Ans ->

KMS Multi-Region Keys are encryption keys that you can use across different AWS regions. They help ensure your data is secure and accessible, no matter where it's stored or processed, by allowing you to manage one key for use in multiple regions. This makes it easier to keep your encryption consistent and reliable globally.

Key points:

- Identical KMS keys in different AWS Regions that can be used interchangeably
- Multi-Region keys have the same key ID, key material, automatic rotation...
- Encrypt in one Region and decrypt in other Regions
- No need to re-encrypt or making cross-Region API calls
- KMS Multi-Region are NOT global (Primary + Replicas)
- Each Multi-Region key is managed independently

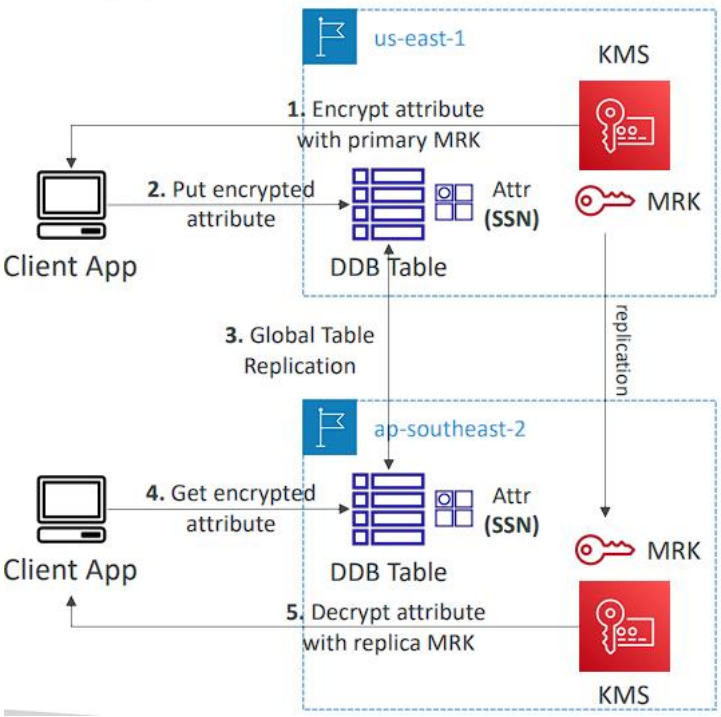
Use cases: global client-side encryption, encryption on Global DynamoDB, Global Aurora



353. DynamoDB Global Tables and KMS Multi-Region Keys Client-Side encryption.

Ans ->

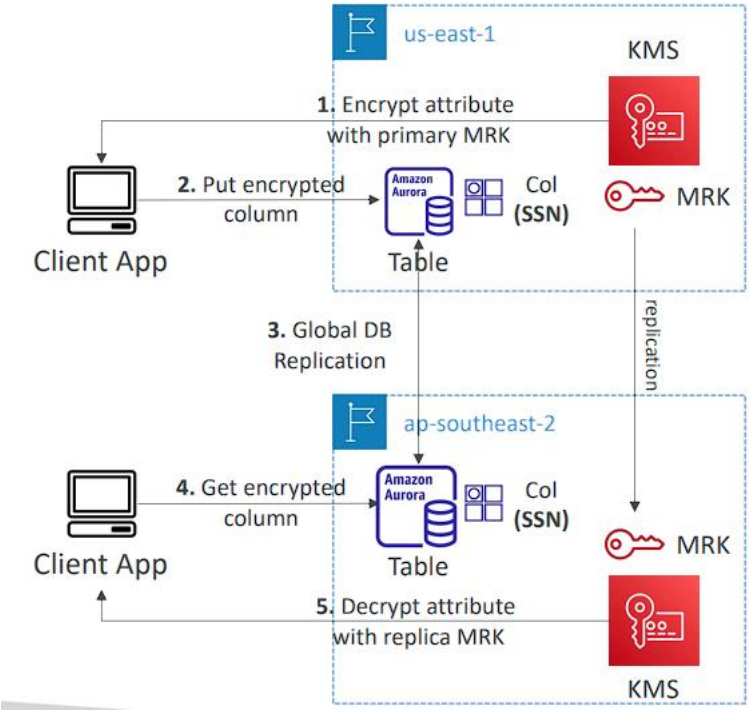
- We can ***encrypt specific attributes*** client-side in our DynamoDB table using the **Amazon DynamoDB Encryption Client**
- Combined with Global Tables, the client-side encrypted data is replicated to other regions
- If we use a multi-region key, replicated in the same region as the DynamoDB Global table, then clients in these regions can use low-latency API call to KMS in their region to decrypt the data client-side
- **Using client-side encryption we can protect specific fields and guarantee only decryption if the client has access to an API key**



354. Global Aurora and KMS Multi-Region Keys Client-Side encryption.

Ans ->

- **We can encrypt specific attributes client-side in our Aurora table using the AWS Encryption SDK**
- Combined with Aurora Global Tables, the client-side encrypted data is replicated to other regions
- If we use a multi-region key, replicated in the same region as the Global Aurora DB, then clients in these regions can use low-latency API calls to KMS in their region to decrypt the data client-side
- Using client-side encryption we can protect specific fields and guarantee only decryption if the client has access to an API key, **we can protect specific fields even from database admins**.



355. S3 Replication Encryption Consideration.

Ans ->

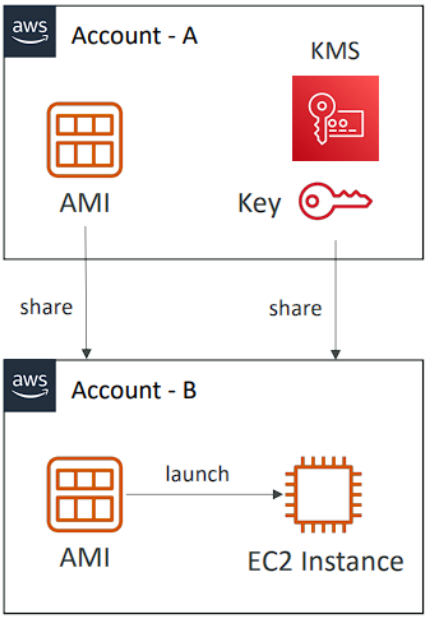
- Unencrypted objects and objects encrypted with SSE-S3 are replicated by default
- Objects encrypted with SSE-C (Customer Provided Key) can be replicated
- **For objects encrypted with SSE-KMS**, you need to enable the options:
  - Specify which KMS Key to encrypt the objects within the target bucket
  - Adapt the KMS Key Policy for the target key

        - An IAM Role with **kms:Decrypt** for the source KMS Key and **kms:encrypt** for the target KMS Key
        - You might get KMS throttling errors, in which case you can ask for a Service Quotas increase
    - **You can use multi-region AWS KMS Keys, but they are currently treated as independent keys by amazon S3 (the object will still be decrypted and then encrypted)**

356. Encrypted AMI Sharing Process via KMS.

Ans ->

- AMI in Source Account is **encrypted with KMS key from Source Account**
- Must modify the image attribute to **add a Launch Permission** which corresponds to the **specified target AWS account**
- Must **share** the KMS Keys used to encrypted the snapshot the AMI references with the target account / IAM Role
- The IAM Role/User in the target account must have the permission to **DescribeKey**, **ReEncrypted**, **CreateGrant**, **Decrypt**
- When launching an EC2 instance from the AMI, optionally the target account can specify a new KMS key in its own account to re-encrypt the volumes



357. What is AWS SSM Parameter Store?

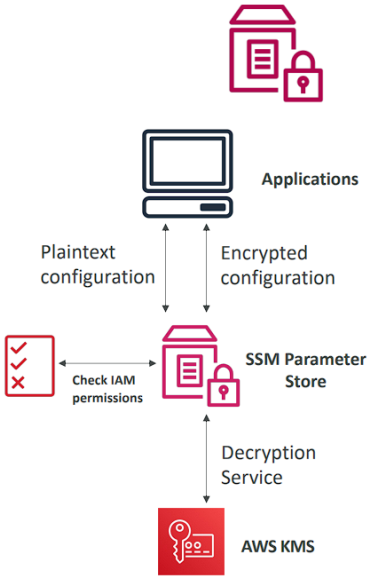
Ans ->

AWS Simple Systems Manager (SSM) Parameter Store is a secure service for storing and managing **configuration data** and **secrets** like **API keys** and **database strings**. It

allows you to retrieve these securely within your applications or AWS services, supports encryption with AWS KMS, and offers features like versioning and access control.

Key points:

- **Secure storage for configuration and secrets**
- Optional Seamless Encryption using KMS
- Serverless, scalable, durable, easy SDK
- **Version tracking of configurations / secrets**
- Security through IAM
- Notifications with Amazon EventBridge
- Integration with CloudFormation

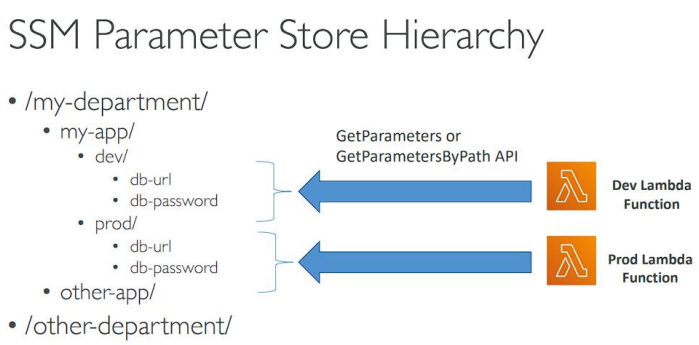


358. Explain about SSM Parameter Store Hierarchy.

Ans ->

In **AWS SSM Parameter Store**, parameters can be organized using a hierarchical structure, which allows you to logically group and manage related parameters. The hierarchy is created using a path-like naming convention with forward slashes (/).

- /application/
  - /application/dev/
    - /application/dev/db/username
    - /application/dev/db/password
  - /application/prod/
    - /application/prod/db/username
    - /application/prod/db/password



- **To get secrets of Secret Manager:**
  - /aws/reference/secretsmanager/secret_ID_in_Secrets_Manager
- **To access public parameters:**
  - /aws/service/ami-amazon-linux-latest/amzn2-ami-hvm-x86_64-gp2 (public)

359. Standard VS Advanced parameter tiers.

Ans ->

	Standard	Advanced
Total number of parameters allowed (per AWS account and Region)	10,000	100,000
Maximum size of a parameter value	4 KB	8 KB
Parameter policies available	No	Yes
Cost	No additional charge	Charges apply
Storage Pricing	Free	$0.05 per advanced parameter per month

360. Key points about Parameters Policies (for advanced Parameters).

Ans ->

In **AWS SSM Parameter Store**, Parameter Policies are **used with advanced parameters** to add more control and governance over how parameters are managed and accessed. These policies allow you to set rules for parameters, such as expiration and notification actions, ensuring better management of sensitive data.

- Allow to **assign a TTL to a parameter** (expiration data) to force updating or deleting sensitive data such as passwords
- Can assign multiple policies at a time

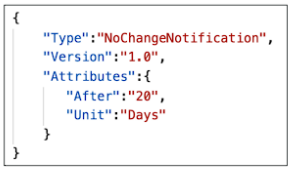
Expiration (to delete a parameter)

```
{
    "Type":"Expiration",
    "Version":"1.0",
    "Attributes":{
        "Timestamp":"2020-12-02T21:34:33.000Z"
    }
}
```

ExpirationNotification (EventBridge)

```
{
    "Type":"ExpirationNotification",
    "Version":"1.0",
    "Attributes":{
        "Before":"15",
        "Unit":"Days"
    }
}
```

NoChangeNotification (EventBridge)



```
{
    "Type":"NoChangeNotification",
    "Version":"1.0",
    "Attributes":{
        "After":"20",
        "Unit":"Days"
    }
}
```

361. What is AWS Secrets Manager?

Ans ->

AWS Secrets Manager is a service for securely storing, managing, and retrieving sensitive information like API keys and database credentials. It automates secret rotation and access, ensuring enhanced security and simplifying the management of secrets across your AWS environment.

Key points:

- Newer service, **meant for storing secrets**
- Capability to force rotation of secrets every X days
- Automate generation of secrets on rotation (uses Lambda)
- Integration with Amazon RDS (MySQL, PostgreSQL, Aurora)
- **Secrets are encrypted using KMS**
- **Mostly meant for RDS integration**

362. Key points about AWS Secrets Manager - Multi-Region Secrets.

Ans ->

AWS Secrets Manager's Multi-Region Secrets feature lets you replicate and manage secrets like passwords and API keys across multiple AWS regions. This ensures your secrets are available and consistent globally, providing resilience in case a region goes down. The replication is automated, so you don't need to manually synchronize secrets between regions, and the same security settings apply across all locations.

Key points:

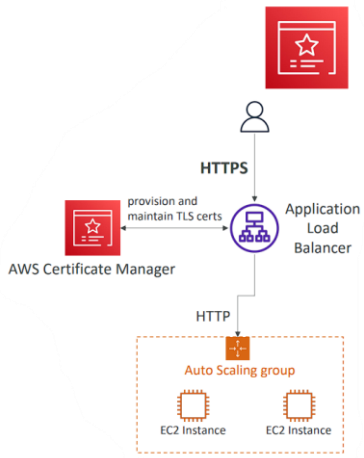
- Replicate Secrets across multiple AWS Regions
- Secrets Manager keeps read replicas in sync with the primary Secret
- Ability to promote a read replica Secret to a standalone Secret
- **Use cases:** multi-region apps, disaster recovery strategies, multi-region DB...

363. Explain about AWS Certificate manager (ACM).

Ans ->

AWS Certificate Manager (ACM) is a service that makes it easy to manage **SSL/TLS certificates** for your AWS-based websites and applications. ACM handles the process of **issuing**, **renewing**, and **deploying** certificates, allowing you to secure your data in transit with minimal effort. It supports both public and private certificates, integrates seamlessly with AWS services like Elastic Load Balancing, CloudFront, and API Gateway, and helps you manage certificates without the need to manually track expiration dates or renewals.

- Easily provision, manage, and deploy TLS Certificates
- Provide in-flight encryption for websites (HTTPS)
- Supports both public and private TLS certificates
- Free of charge for public TLS certificates
- Automatic TLS certificate renewal
- Integrations with (load TLS certificates on):
  - Elastic Load Balancers (CLB, ALB, NLB)
  - CloudFront Distributions
  - APIs on API Gateway
- Cannot use ACM with EC2 (can't be extracted)



364. What is the process of ACM - Requesting Public Certificates?

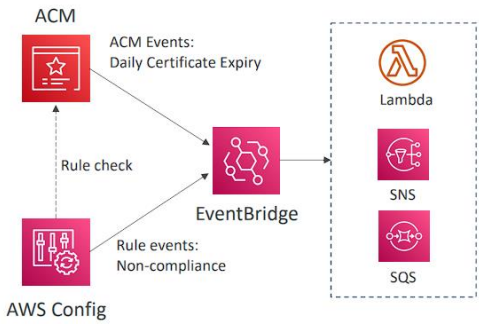
Ans ->

- **List domain names to be included in the certificate:**
  - **Fully Qualified Domain Name (FQDN):** crop.example.com
  - **Wildcard Domain:** *.example.com
- **Select Validation Method: DNS Validation or Email validation**
  - DNS validation is preferred for automation purposes
  - Email validation will send emails to contact addresses in the WHOIS database
  - DNS Validation will leverage a CNAME record to DNS config (ex: Route 53)
- It will take a few hours to get verified
- The Public Certificate will be enrolled for automatic renewal
  - **ACM automatically renews ACM-generated certificates 60 days before expiry**

365. What is the process of ACM - Importing Public Certificates?

Ans ->

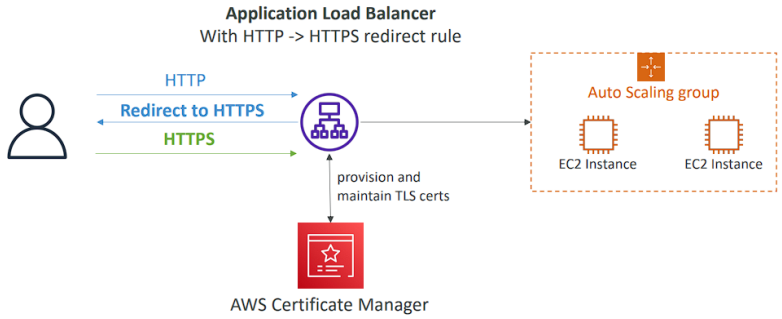
- Option to generate the certificate outside of ACM and then import it
- **No automatic renewal**, must import a new certificate before expiry
- **ACM sends daily expiration events starting 45 days prior to expiration**
  - The # of days can be configured
  - Events are appearing in EventBridge
- **AWS Config** has a managed rule named **acm-certificate-expiration-check** to check for expiring certificates (configurable number of days)



366. ACM - Integration with ALB.

Ans ->

ACM – Integration with ALB



367. Explain about API Gateway - Endpoint Types.

Ans ->

Edge-Optimized (default): For global clients

- Requests are routed through the CloudFront Edge locations (improves latency)
- The API Gateway still lives in only one region

Regional:

- For clients within the same region
- Could manually combine with CloudFront (more control over the caching Strategies and the distribution)

Private:

- Can only be accessed from you VPC using an interface VPC endpoint (ENI)
- Use a resource policy to define access

368. ACM - Integration with API Gateway.

Ans ->

Create a Custom Domain Name in API Gateway

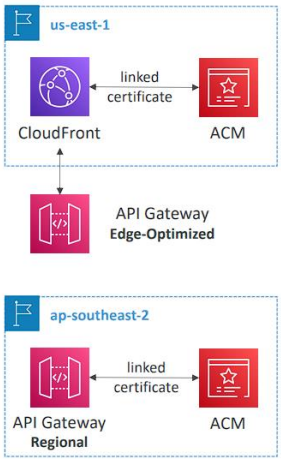
Edge-Optimized (default): For global clients

- Requests are routed through the CloudFront Edge locations (improves latency)
- The API Gateway still lives in only one region
- **The TLS Certificate must be in the same region as CloudFront, in us-east-1**
- Then setup CNAME or (better) A-Alias record in Route 53

Regional:

- For clients within the same region
- **The TLS Certificate must be imported on API Gateway, in the same region as the API Stage**

- Then setup CNAME or (better) A-Alias record in Route 53



369. What is AWS WAF?

Ans ->

AWS Web Application Firewall (WAF) is a security service that protects your web applications from common web exploits and attacks, such as **SQL injection and cross-site scripting (XSS)**. It allows you to create custom rules to filter and monitor HTTP and HTTPS requests based on conditions like IP addresses, query strings, and HTTP headers.

Key points:

- Protects your web application from common web exploits (**operates at Layer 7**)
- **Layer 7 is HTTP (vs Layer 4 is TCP/UDP)**
- **Can only be deployed on:**
  - Application Load Balancer
  - API Gateway
  - CloudFront
  - AppSync GraphQL API
  - Cognito User Pool

370. Key points about AWS WAF ACL.

Ans ->

A WAF ACL (Access Control List) in AWS WAF is a set of rules that controls which web requests are allowed or blocked, helping to protect your applications from attacks.

Key points:

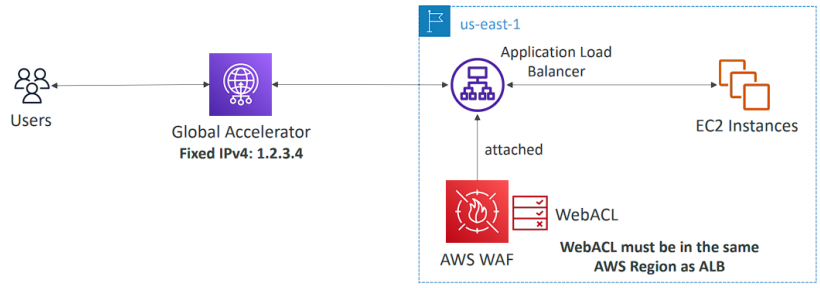
- **Define Web ACL (Web Access Control List) Rules:**
  - **IP Set:** up to 10,000 IP addresses - use multiple Rules for more IPs

        - HTTP headers, HTTP body, or URI strings Protects from common attack - **SQL injection and Cross-Site Scripting (XSS)**
        - Size constraints, geo-match (block countries)
        - **Rate-based rules** (to count occurrences of events) - for DDoS protection
    - **Web ACL are Regional except for CloudFront**
    - A rule group is a reusable set of rules that you can add to a web ACL

371. How to get a fixed IP while using WAF with a Load Balancer.

Ans ->

- **WAF does not support the Network Load balancer (Layer 4)**
- We can **use Global Accelerator** for fixed IP and WAF on the ALB



372. Explain about AWS Shield.

Ans ->

AWS Shield is a managed Distributed Denial of Service (DDoS) protection service that safeguards your applications running on AWS. It provides always-on detection and automatic inline mitigations that minimize application downtime and latency.

AWS Shield comes in two tiers:

- **AWS Shield Standard:**
    - Free service that is activated for every AWS customer
    - Provides protection from attacks such as SYN/UDP Floods, Reflection attacks and other layer 3/4 attacks
- **AWS Shield Advanced:**
    - Optional DDoS mitigation service (**$3000 per month per organization**)
    - Protect against more sophisticated attack on **Amazon EC2, Elastic Load Balancer (ELB), Amazon CloudFront, AWS Global Accelerator,** and **Route 53**
    - 24/7 access to AWS DDoS Response team (DRP)
    - Shield Advanced automatic application layer DDoS mitigation automatically creates, evaluates and deploys AWS WAF rules to mitigate layer 7 attacks

373. Explain about AWS Firewall Manager.

Ans ->

AWS Firewall Manager is a security management service that helps you **centrally configure and manage firewall rules** across your AWS accounts and resources. It works with **AWS WAF**, **AWS Shield**, and **VPC security groups**, allowing you to apply consistent security policies and protections across your entire organization.

Key points:

- Manage rules in all accounts of an AWS Organization
- **Security policy:** common set of security rules
  - WAF rules (Application Load Balancer, API Gateways, CloudFront)
  - AWS Shield Advanced (ALB, CLB, NLB, Elastic IP, CloudFront)
  - Security Groups for EC2, Application Load Balancer and ENI resources in VPC
  - AWS Network Firewall (VPC Level)
  - Amazon Route 53 Resolved DNS Firewall
  - Policies are created at the region level
- **Rules are applied to new resources as they are created (good for compliance) across all and future accounts in your organization.**

374. WAF vs Firewall Manager vs Shield

Ans ->

- **WAF, Shield and Firewall Manager are used together for comprehensive protection**
- Define your Web ACL rules in WAF
- For granular protection of your resources, WAF alone is the correct choice
- If you want to use AWS WAF across accounts, accelerate WAF configuration, automate the protections of new resources, use Firewall Manager with AWS WAF
- Shield Advanced adds additional features on top of AWS WAF, such as dedicated support from the Shield Response Team (SRT) and advanced reporting.

375. AWS Best Practices for DDoS Resiliency Edge Location Mitigation (BP1, BP3).

Ans ->

BP1 - CloudFront:

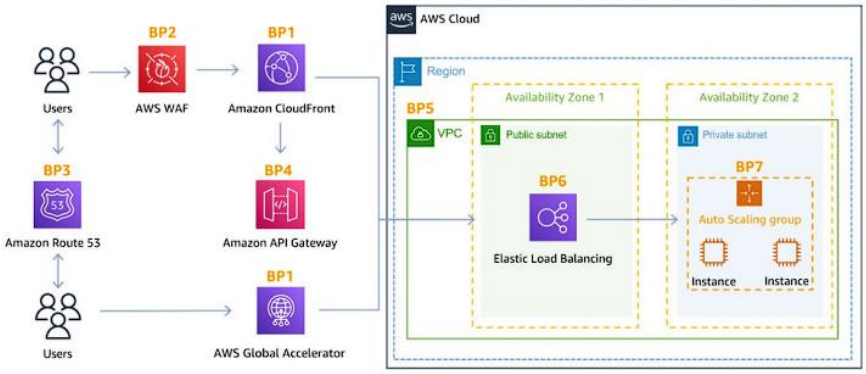
- Web Application delivery at the edge
- Protect from DDoS Common Attacks (SYN floods, UDP reflection...)

BP1 - Global Accelerator:

- Access your application from the edge Integration with Shield for DDoS protection
- Helpful if your backend is not compatible with CloudFront

BP3 - Route 53:

- Domain name Resolution at the edge
- DDoS Protection mechanism



376. AWS Best Practices for DDoS Resiliency Best Practices for DDoS mitigation.

Ans ->

Infrastructure layer defense (BP1, BP3, BP6):

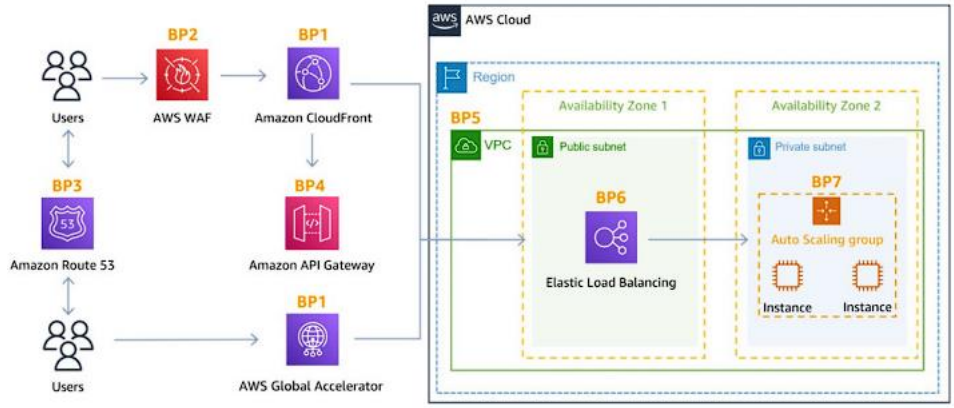
- Protect Amazon EC2 against high traffic
- That includes using Global Accelerator, Route 53, CloudFront, Elastic Load Balancing

Amazon EC2 with Auto Scaling (BP7):

- Helps scale in case of sudden traffic surges including a flash crowd or a DDoS attack

Elastic Load balancing (BP6):

- Elastic Load Balancing scales with the traffic increase and will distribute the traffic to many EC2 instances



377. AWS Best Practices for DDoS Resiliency Application layer Defense

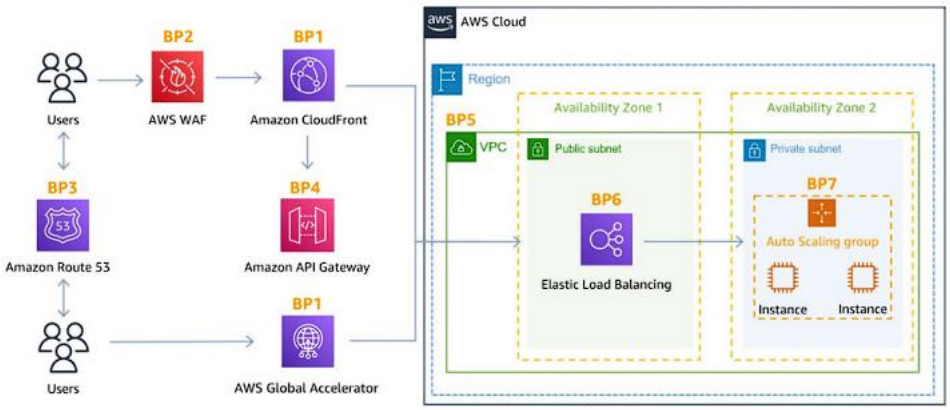
Ans ->

Detect and filter malicious web requests (BP1, BP2):

- CloudFront cache static content and serve from edge locations, protecting your backend
- AWS WAF is used on top of CloudFront and Application Load Balancer to filter and block requests based on request signatures
- WAF rate-based rules can automatically block the IPs of bad actors
- use managed rules on WAF to block attacks based on IP reputation, or block anonymous IPs
- CloudFront can block specific geographies

Shield Advanced (BP1, BP2, BP6):

- Shield Advanced automatic application layer DDoS mitigation automatically creates, evaluates and deploys AWS WAF rules to mitigate layer 7 attacks



378. AWS Best Practices for DDoS Resiliency Attack surface reduction

Ans ->

Obfuscating AWS resources (BP1, BP4, BP6):

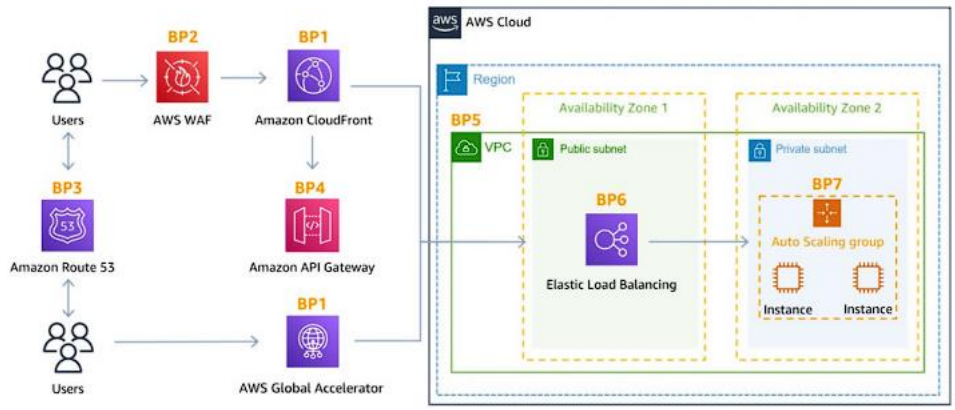
- Using CloudFront API gateway, Elastic Load balancing to hide your backend resources (Lambda functions, EC2 instances)

Security groups and Network ACLs (BP5):

- Use security groups and NACLs to filter traffic based on specific IP at the subnet or ENI-level
- Elastic IP are protected by AWS Shield Advanced

Protecting API endpoints (BP4):

- Hide EC2, Lambda, elsewhere
- Edge-optimized mode, or CloudFront + regional mode (more control for DDoS)
- WAF + API Gateway: burst limits, headers filtering, use API keys.



379. What is Amazon GuardDuty?

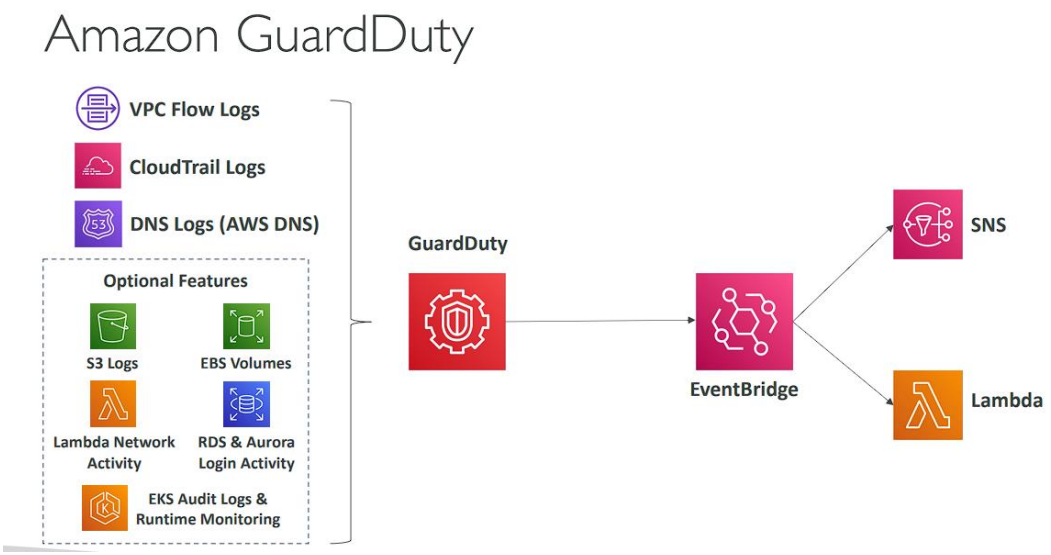
Ans ->

AWS GuardDuty is a managed threat detection service that automatically monitors your AWS environment for suspicious activity, such as unauthorized access or unusual behavior, and alerts you if something looks wrong. It helps keep your AWS resources safe by detecting potential threats without needing complex setup.

Key points:

- Intelligent Threat discovery to protect your AWS Account
- Use Machine Learning algorithms, anomaly detection, 3rd party data
- One click to enable (**3 days trial**), no need to install software
- Input data includes:
  - **Cloud Trail Events Logs** - unusual API calls, unauthorized deployments
    - **Cloud Trail Management Events** - create VPC subnet, create trail,...
    - **CloudTrail S3 Data Events** - get object, list objects, delete object,...
  - **VPC Flow Logs** - unusual internal traffic, unusual IP address

    - **DNS Logs** - compromised EC2 instances sending encoded data within DNS queries
    - **Optional Features** - EKS Audit Logs, RDS & Aurora, EBS, Lambda, S3 Data Events...
    - Can setup EventBridge rules to be notified in case of findings
- Event Bridge rules can target AWS Lambda or SNS
- **Can protect against CryptoCurrency attacks (has a dedicated "finding" for it)**



380. What is Amazon Inspector?

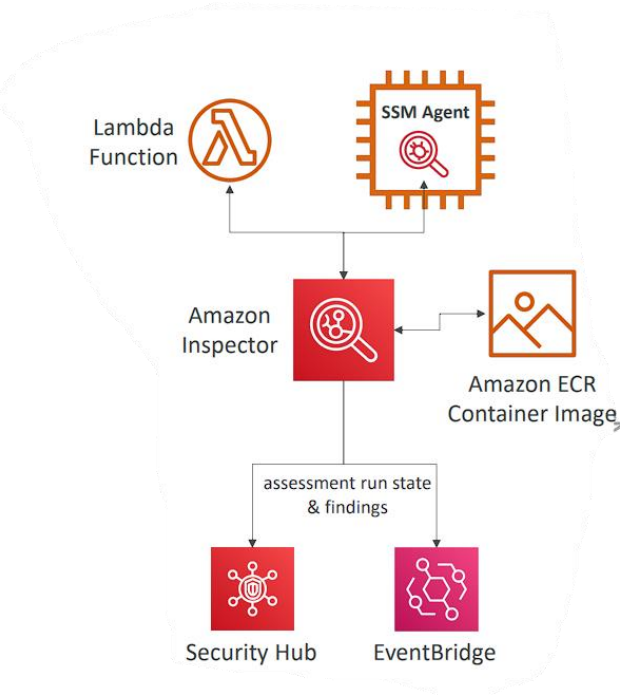
Ans ->

Amazon Inspector is a security assessment service that automatically scans your AWS resources (only **EC2 instances, Container images** and **Lambda Functions**), for vulnerabilities and deviations from best practices. It identifies potential security issues, such as unpatched software or weak configurations, and provides detailed findings to help you improve the security of your environment.

Key points:

- **Automated Security Assessments**
- **For EC2 instances:**
    - Leveraging the AWS System Manager (SSM) agent
    - Analyze against unintended network accessibility
    - Analyze the running OS against known vulnerabilities

- **For Container Images push to Amazon ECR:**
  - Assessment of Container Images as they are pushed
- **For Lambda Functions:**
  - Identifies software vulnerabilities in function code and package dependencies
  - Assessment of functions as they are deployed
  - Reporting & integration with AWS Security Hub
  - Send findings to Amazon EventBridge



381. What does Amazon Inspector evaluate?

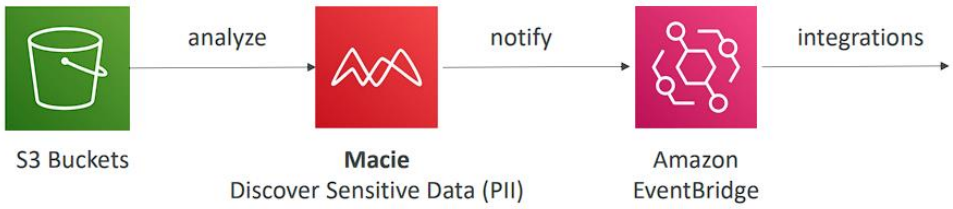
Ans ->

- **Remember:** only for **EC2 instances**, **Container Images** & **Lambda functions**
- Continuous scanning of the infrastructure, only when needed
- Package vulnerabilities (EC2, ECR & Lambda) - database of CVE
- Network reachability (EC2)
- A risk score is associated with all vulnerabilities for prioritization

382. What is Amazon Macie?

Ans ->

- **Amazon Macie is a fully managed data security and data privacy service** that uses machine learning and pattern matching to discover and **protect your sensitive data in AWS**.
- Macie helps identify and alert you to sensitive data, such as **personally identifiable information (PII)**.



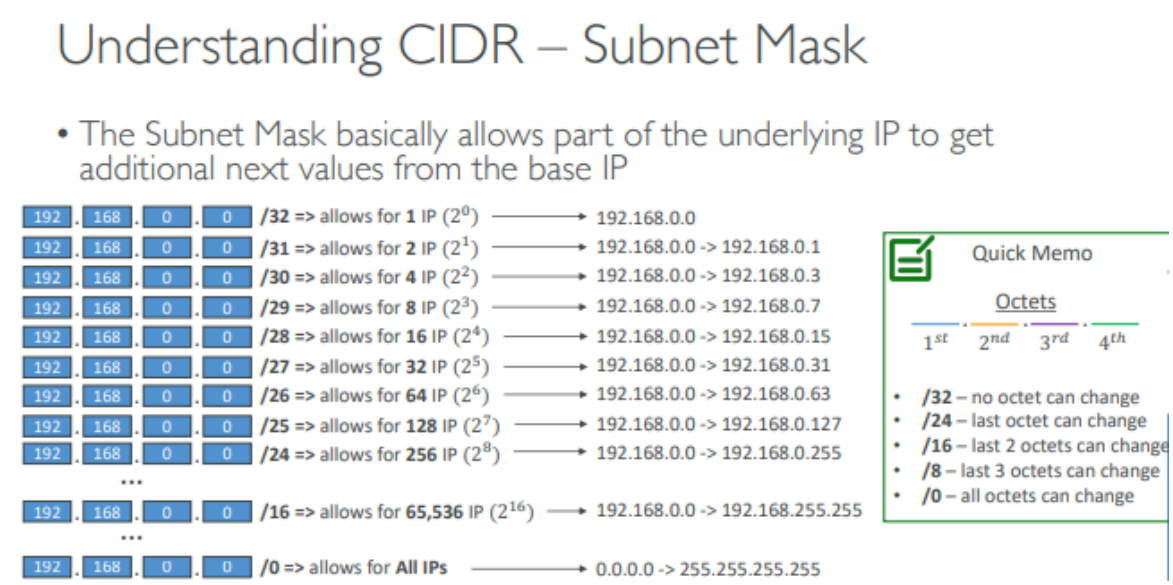
382. Explain about CIDR - IPv4 & Subnet Mask

Ans ->

- **Classless Inter-Domain Routing (CIDR)** - a method for allocating IP addresses
- Used in Security Groups rules and AWS networking in general

IP version	Type	Protocol	Port range	Source	Description
IPv4	SSH	TCP	22	122.149.196.85/32	–
IPv4	HTTP	TCP	80	0.0.0.0/0	–

- **They help to define an IP address range:**
  - We've seen **WW.XX.YY.ZZ/32** => **one IP**
  - We've Seen 0.0.0.0/0 => all IPs
  - But we can define: 192.168.0.0/26 => 192.168.0.0 - 192.168.0.63 (64 IP addresses)
- **A CIDR consists of two components:**
  - **Base IP:**
    - Represents an IP contained in the range (WW.XX.YY.ZZ)
    - **Example:** 10.0.0.0, 192.168.0.0, ...
  - **Subnet Mask:**
    - Defines how many bits can change in the IP
    - **Example:** /0, /8, /24, /32
    - /8 => 255.0.0.0
    - /16 => 255.255.0.0
    - /24 => 255.255.255.0
    - /32 => 255.255.255.255



383. Public vs. Private IP (IPv4)

Ans ->

The **Internet Assigned Numbers Authority (IANA)** established certain blocks of IPv4 addresses for the use of private (LAN) and public (Internet) addresses

Private IP can only allow certain values:

- 10.0.0.0 - 10.255.255.255 (10.0.0.0/8) => in big networks
- 172.16.0.0 - 172.31.255.255 (172.16.0.0/12) => AWS default VPC in that range
- 192.168.0.0 - 192.168.255.255 (192.168.0.0/16) => e.g., home networks

All the rest of the IP addresses on the Internet are **Public IP**.

384. Key points about VPC in AWS - IPv4.

Ans ->

- **VPC = Virtual Private Cloud**
- You can have multiple VPCs in an AWS region (max. 5 per region - soft limit)
- **Max. CIDR per VPC is 5, for each CIDR:**
  - Min. size is /28 (16 IP addresses)
  - Max. size is /16 (65536 IP addresses)
- **Because VPC is private, only the Private IPv4 ranges are allowed:**
  - 10.0.0.0 - 10.255.255.255 (10.0.0.0/8)
  - 172.16.0.0 - 172.31.255.255 (172.16.0.0/12)
  - 192.168.0.0 - 192.168.255.255 (192.168.0.0/16)
- **Your VPC CIDR should NOT overlap with your other networks (e.g., corporate)**

385. Key points about VPC - Subnet (IPv4).

Ans ->

- **AWS reserves 5 IP addresses (first 4 & last 1) in each subnet.**
- **These 5 IP addresses are not available for use and can't be assigned to an EC2 instance**
- **Example:** if CIDR block 10.0.0.0/24, then reserved IP addresses are:
  - 10.0.0.0 -> Network Address
  - 10.0.0.1 -> reserved by AWS for the VPC router
  - 10.0.0.2 -> reserved by AWS for mapping to Amazon-provided DNS
  - 10.0.0.3 -> reserved by AWS for future use
  - 10.0.0.255 -> Network Broadcast Address. AWS does not support broadcast in a VPC; therefore the address is reserved.
- **Exam Tip:** if you need 29 IP addresses for EC2 instances:
  - You can't choose a subnet of size /27 **(2^5 = 32 IP addresses, 32-5 = 27 < 29)**
  - You need to choose a subnet of size /26 **(2^6 = 64 IP addresses, 64-5 = 59 > 29)**

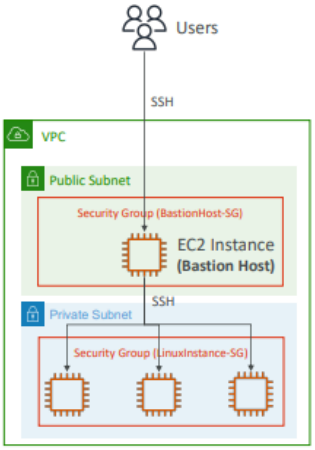
386. What are Bastion Hosts?

Ans->

A **Bastion Host** in AWS is a secure instance placed in a public subnet that acts as a gateway for accessing instances in a private subnet. It allows administrators to connect to private resources **through SSH or RDP**, while the private instances remain inaccessible from the internet. This setup enhances security by minimizing direct exposure to the internet and is often complemented by strict access controls and logging practices.

Key points:

- We can use a **Bastion Host to SSH or RDP into our private EC2 instances**
- The bastion is in the public subnet which is then connected to all other private subnets
- Bastion Host security group must allow inbound from the internet on port 22 from restricted CIDR, for example the public CIDR of your corporation
- Security Group of the EC2 Instances must allow the Security Group of the Bastion Host, or the private IP of the Bastion host



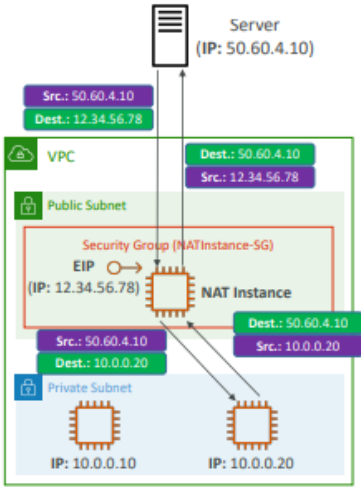
387. What is NAT Instance?

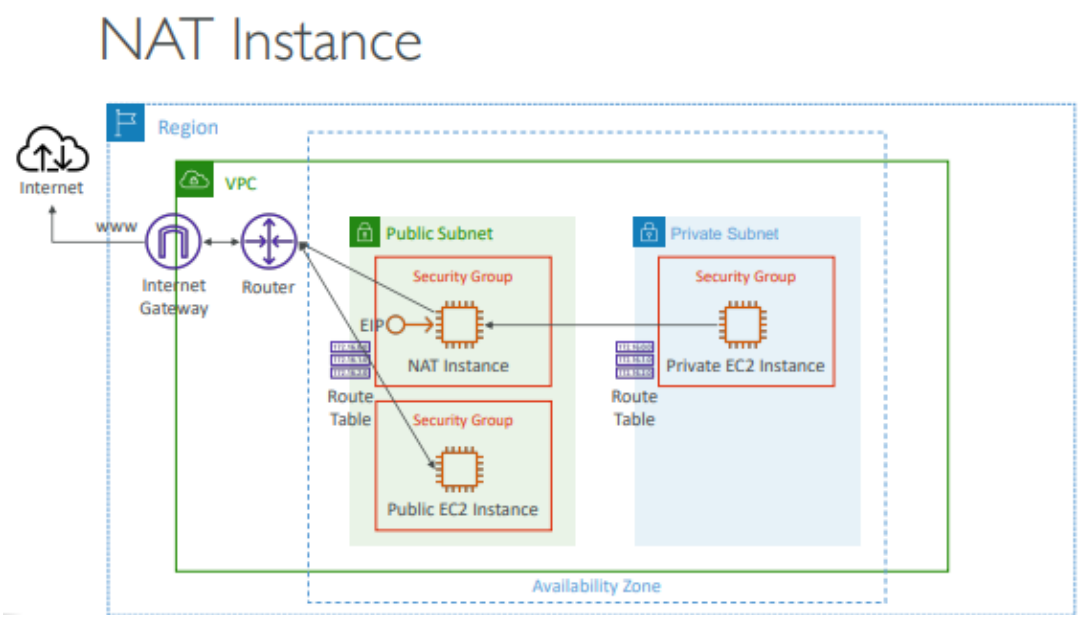
Ans ->

A **NAT Instance** in AWS lets private instances access the internet by acting as a bridge that translates their private IP addresses to a public one. This allows private instances to reach the web for updates and services without being directly exposed to incoming internet traffic.

Key points:

- **NAT = Network Address Translation**
- Allows EC2 instances in private subnets to connect to the Internet
- Must be launched in a public subnet
- **Must disable EC2 setting:** Source / Destination Check
- Must have Elastic IP attached to it
- Route Tables must be configured to route traffic from private subnets to the NAT Instance.





388. Some comments about NAT Instances.

Ans ->

- Pre-configured Amazon Linux AMI is available
  - Reached the end of standard support on December 31, 2020
- Not highly available / resilient setup out of the box
  - You need to create an ASG in multi-AZ + resilient user-data script
- Internet traffic bandwidth depends on EC2 instance type
- **You must manage Security Groups & rules:**
  - **Inbound:**
    - Allow HTTP / HTTPS traffic coming from Private Subnets
    - Allow SSH from your home network (access is provided through Internet Gateway)
  - **Outbound:**
    - Allow HTTP / HTTPS traffic to the Internet

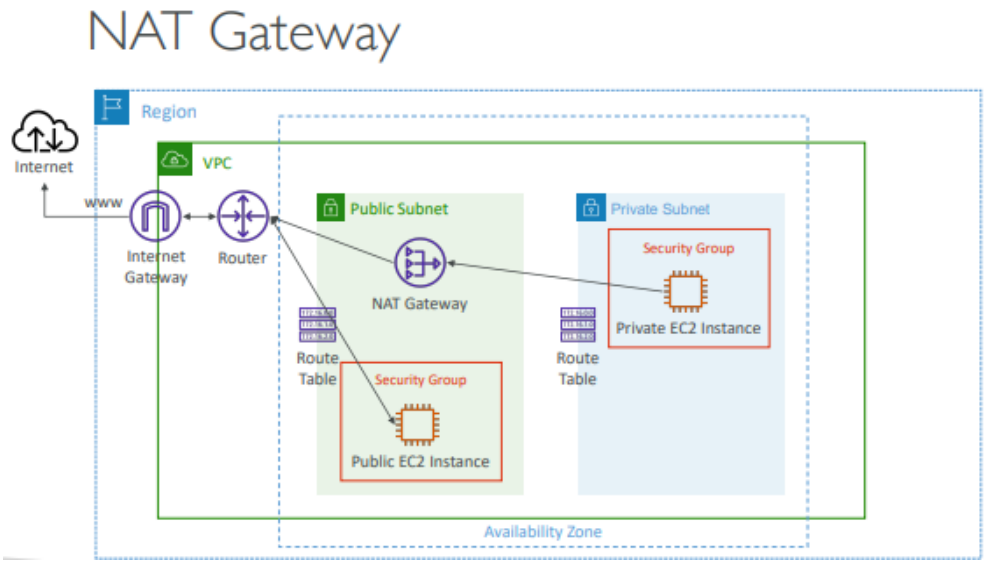
389. What is NAT Gateway?

Ans ->

A **NAT Gateway** in AWS is a managed service that enables instances in a private subnet to access the internet while keeping them shielded from incoming traffic. It automatically handles network address translation (NAT), scales with traffic, and is easier to set up and maintain compared to a NAT Instance.

Key points:

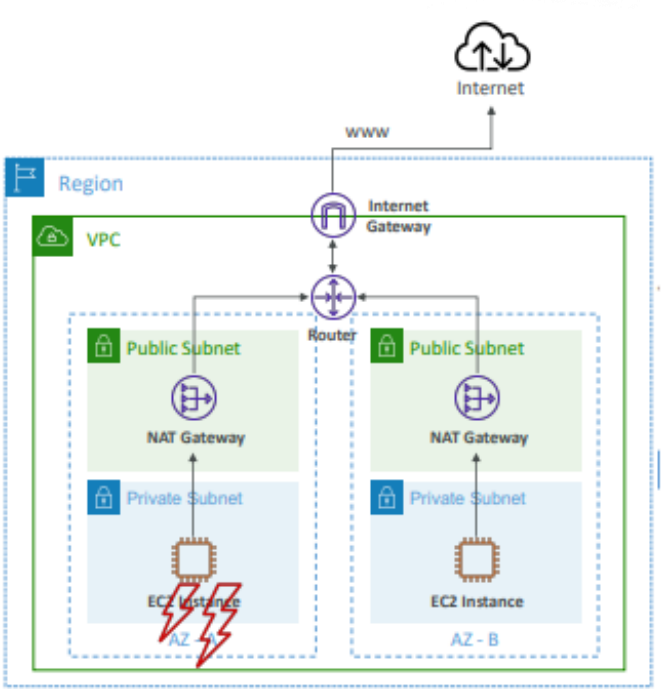
- AWS-managed NAT, higher bandwidth, high availability, no administration
- Pay per hour for usage and bandwidth
- NATGW is created in a specific Availability Zone, uses an Elastic IP
- Can't be used by EC2 instance in the same subnet (only from other subnets)
- Required an IGW (Private Subnet => NATGW => IGW)
- 5 Gbps of bandwidth with automatic scaling up to 100 Gbps
- No Security Groups to manage / required



390. NAT Gateway with High Availability.

Ans ->

- NAT Gateway is resilient within a single Availability Zone
- Must create multiple NAT Gateway in multiple AZs for fault-tolerance
- There is no cross-AZ failover needed because if an AZ goes down it doesn't need NAT



391. What are the differences between NAT Gateway and NAT Instance?

Ans ->

NAT Gateway vs. NAT Instance

	NAT Gateway	NAT Instance
Availability	Highly available within AZ (create in another AZ)	Use a script to manage failover between instances
Bandwidth	Up to 100 Gbps	Depends on EC2 instance type
Maintenance	Managed by AWS	Managed by you (e.g., software, OS patches, ...)
Cost	Per hour & amount of data transferred	Per hour, EC2 instance type and size, + network $
Public IPv4	✓	✓
Private IPv4	✓	✓
Security Groups	✗	✓
Use as Bastion Host?	✗	✓

More at: https://docs.aws.amazon.com/vpc/latest/userguide/vpc-nat-comparison.html

392. Explain about Security Groups and NACLs - Incoming and Outgoing Requests.

Ans ->

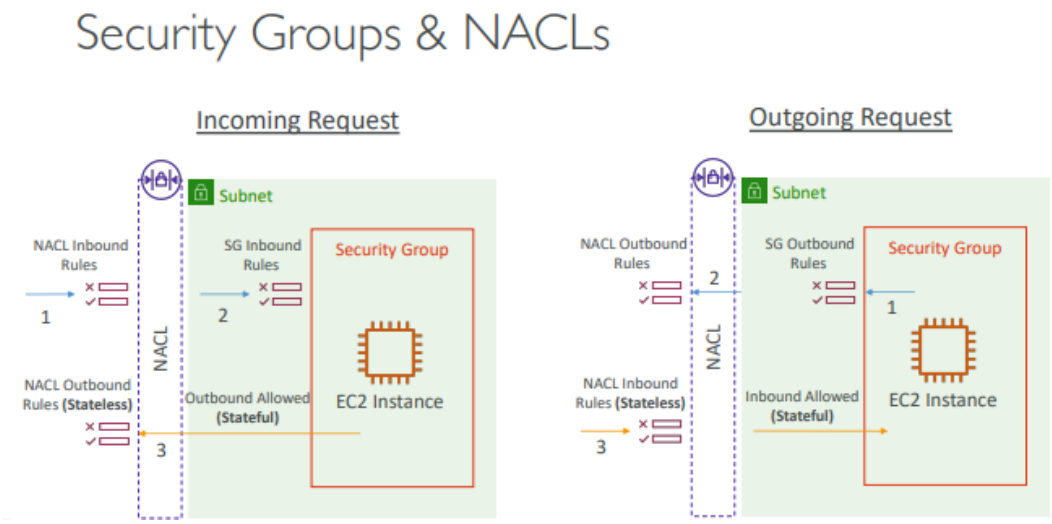
In AWS, Security Groups and Network ACLs (NACLs) control traffic to and from your resources like EC2 instances, with key differences in how they manage state.

Security Groups:

- **Inbound and Outbound Rules:** Security Groups allow or deny incoming and outgoing traffic at the instance level. By default, all outbound traffic is allowed, but you can restrict it.
- **Stateful:** Security Groups are stateful, meaning if you allow incoming traffic, the corresponding outgoing response is automatically allowed, simplifying management.

Network ACLs (NACLs):

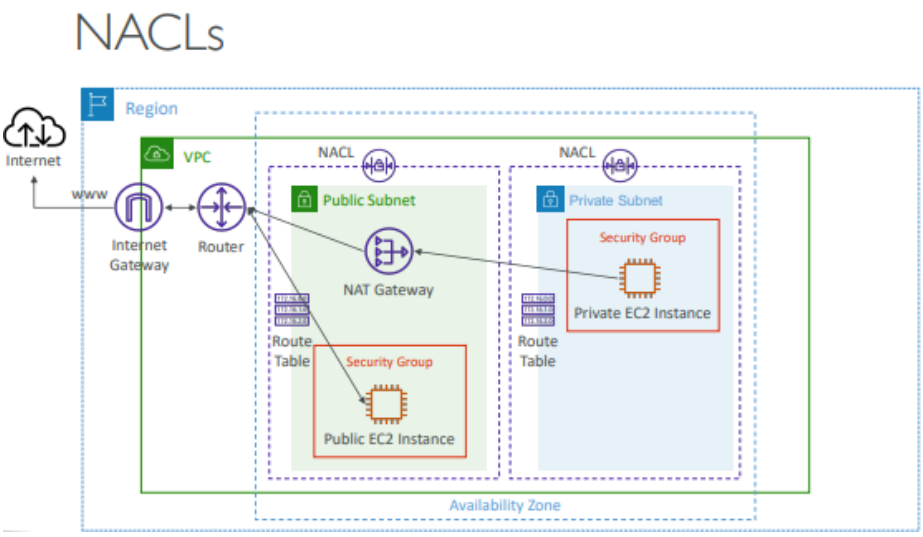
- **Inbound and Outbound Rules:** NACLs control traffic at the subnet level, allowing or denying traffic based on ordered rules.
- **Stateless:** NACLs are stateless, so each request and response is handled independently. You must explicitly set rules for both directions.



393. Key points about Network Access Control List (NACL).

Ans ->

- NACL are like a firewall which control traffic from and to subnets
- One NACL per subnet, new subnets are assigned the Default NACL
- **You define NACL Rules:**
  - Rules have a number (**1-32766**), higher precedence with a lower number
  - First rule match will drive the decision
  - **Example:** if you define **#100 ALLOW 10.0.0.10/32** and **#200 DENY 10.0.0.10/32**, the IP address will be allowed because 100 has a higher precedence over 200
  - The last rule is an asterisk (*) and denies a request in case of no rule match
  - AWS recommends adding rules by increment of 100
- **Newly created NACLs will deny everything**
- NACL are a great way of **blocking a specific IP address at the subnet level**



394. Explain about Default NACL.

Ans ->

The **default NACL** in AWS allows all inbound and outbound traffic by default and is automatically associated with all new subnets in a VPC. It's stateless, meaning inbound and outbound traffic are managed separately. You can modify its rules, but you cannot delete it.

- Accepts everything inbound/outbound with the subnets it's associated with.
- **Do NOT modify the Default NACL, instead create custom NACLs**

Default NACL for a VPC that supports IPv4

Inbound Rules

Rule #	Type	Protocol	Port Range	Source	Allow/Deny
100	All IPv4 Traffic	All	All	0.0.0.0/0	ALLOW
*	All IPv4 Traffic	All	All	0.0.0.0/0	DENY

Outbound Rules

Rule #	Type	Protocol	Port Range	Destination	Allow/Deny
100	All IPv4 Traffic	All	All	0.0.0.0/0	ALLOW
*	All IPv4 Traffic	All	All	0.0.0.0/0	DENY

395. Explain about Ephemeral Ports.

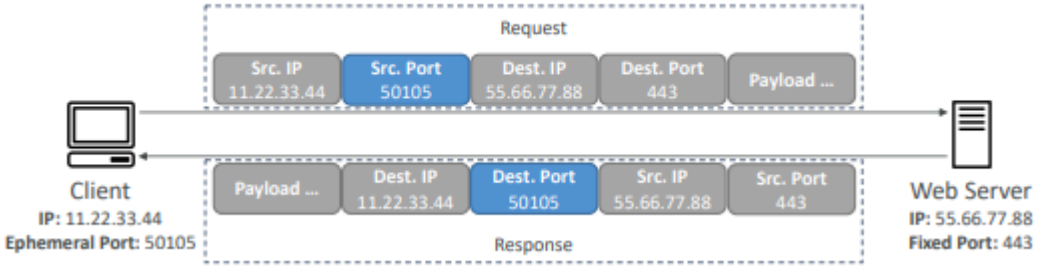
Ans ->

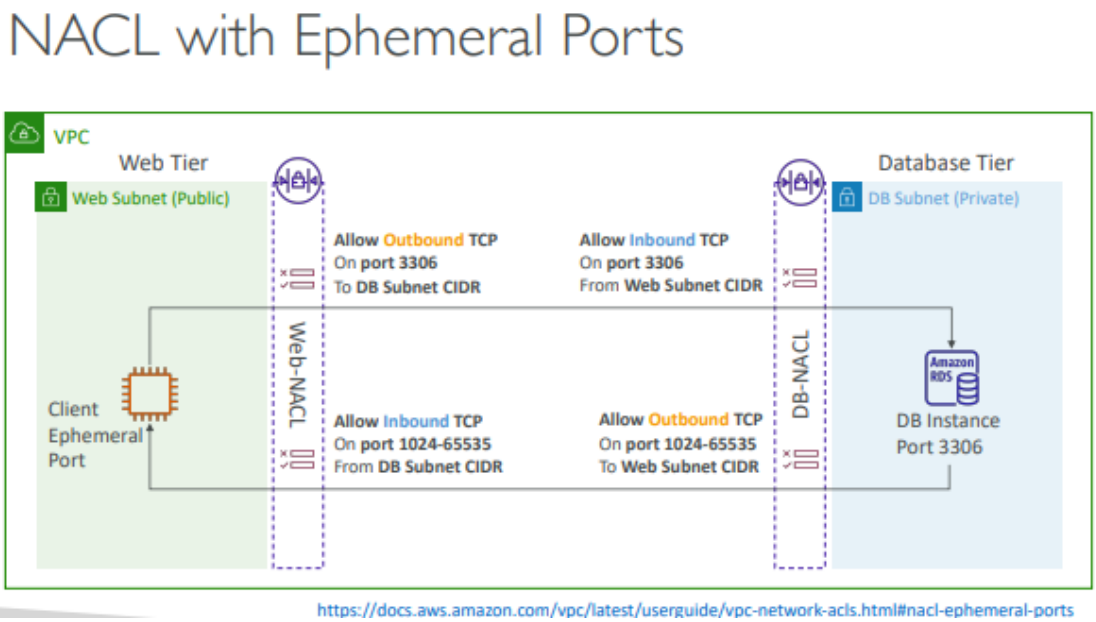
Ephemeral ports are temporary, short-lived ports automatically assigned by an operating system for the client side of a network connection. When a client (like your computer or an app) connects to a server, it uses an ephemeral port as the source port for that connection.

These ports are typically in a high range (e.g., **49152 to 65535**) and are released once the connection is closed, making them available for reuse in future connections. They help manage multiple connections simultaneously without conflicts.

Key points:

- **For any two endpoints to establish a connection, they must use ports**
- Clients connect to a defined port, and expect a response on an ephemeral port
- Different Operating Systems use different port ranges, examples:
- **IANA & MS Windows 10 -> 49152 - 65535**
- **Many Linux kernels => 32768 - 60999**





396. Security Group VS NACLs.

Ans ->

Security Group	NACL
Operates at the instance level	Operates at the subnet level
Supports allow rules only	Supports allow rules and deny rules
Stateful: return traffic is automatically allowed, regardless of any rules	**Stateless:** return traffic must be explicitly allowed by rules (think of ephemeral ports)
All rules are evaluated before deciding whether to allow traffic	Rules are evaluated in order (lowest to highest) when deciding whether to allow traffic, first match wins
Applies to an EC2 instance when specified by someone	Automatically applies to all EC2 instances in the subnet that it's associated with

NACL Examples: https://docs.aws.amazon.com/vpc/latest/userguide/vpc-network-acls.html

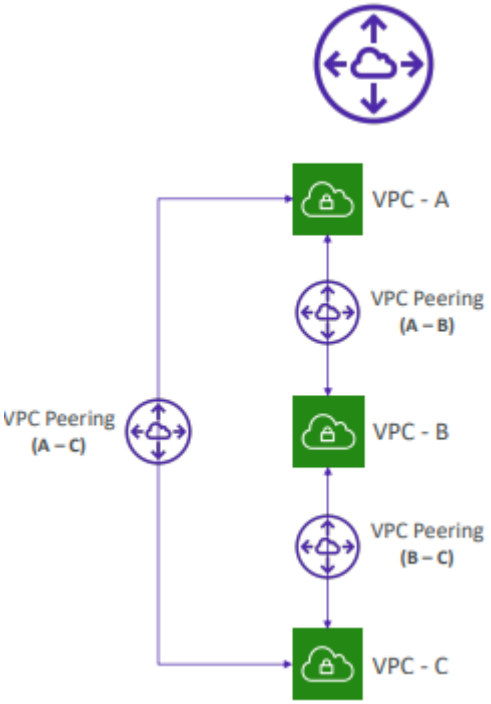
397. Explain about VPC Peering.

Ans ->

VPC Peering in AWS is a networking feature that allows two VPCs to communicate directly with each other using private IP addresses. It's useful for connecting resources across VPCs securely and efficiently. Peering connections are one-to-one, require non-overlapping IP ranges, and **do not support transitive routing or automatic communication with other VPCs not directly peered.**

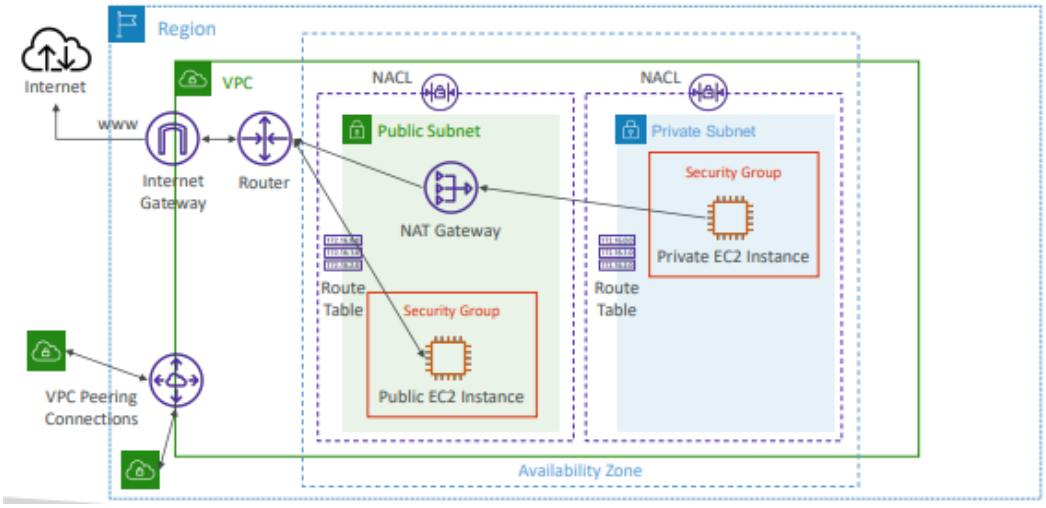
Key points:

- Privately connect two VPCs using AWS's network
- Make them behave as if they were in the same network
- Must not have overlapping CIDRs
- VPC Peering connection is NOT transitive (must be established for each VPC that need to communicate with one another)
- You must update route tables in each VPC's subnets to ensure EC2 instances can communicate with each other
- You can create VPC Peering connection between VPCs in different AWS accounts/regions
- You can reference a security group in a peered VPC (Works cross accounts - same region)



Type	Protocol	Port range	Source
HTTP	TCP	80	sg-04991f9af3473b939 / default
HTTP	TCP	80	[redacted] / sg-027ad1f7865d4be76

Account ID



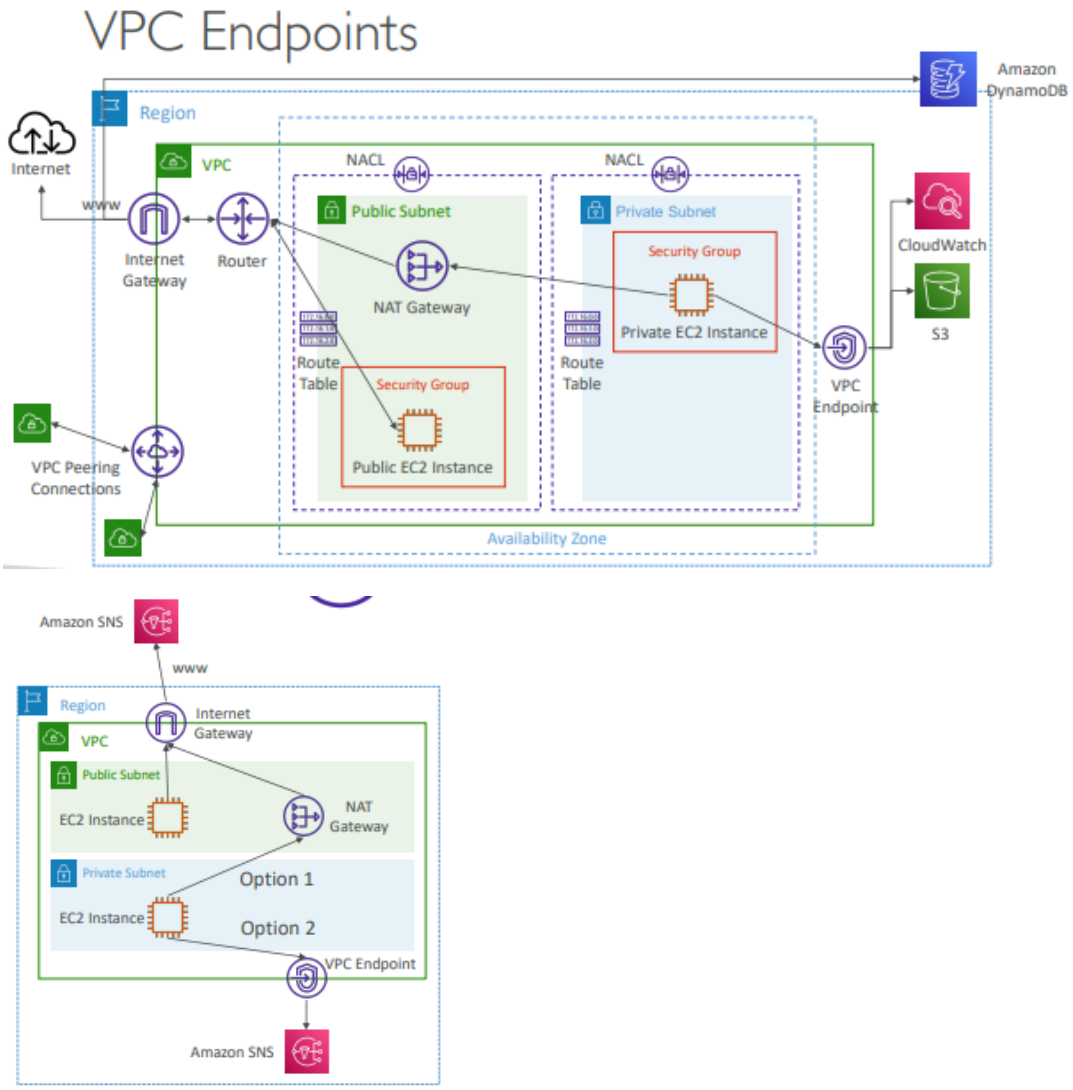
399. Explain about VPC Endpoints (AWS PrivateLink).

Ans ->

VPC Endpoints enable secure, private connections between your VPC and AWS services without using the internet. They keep traffic within the AWS network, improving security and potentially reducing costs and latency.

Key points:

- Every AWS service is publicly exposed (public URL)
- VPC Endpoints (powered by AWS PrivateLink) allows you to connect to AWS services using a private network instead of using the public Internet
- They're redundant and scale horizontally
- **They remove the need of IGW, NATGW, ..., to access AWS Services.**
- **In case of issues:**
  - Check DNS Setting Resolution in your VPC
  - Check Route Tables



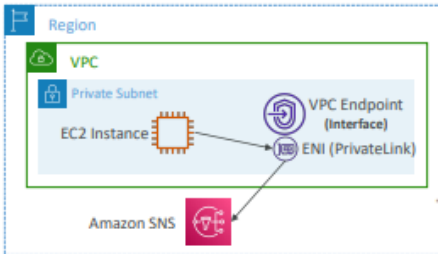
400. What are the different types of VPC Endpoints?

Ans ->

Interface Endpoints (powered by PrivateLink):

These are used to connect to AWS services over private IP addresses. They create an Elastic Network Interface (ENI) in your VPC with a private IP address that you can use to access the service.

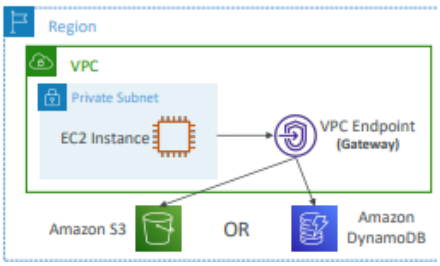
- Provisions an ENI (private IP address) as an entry point (must attach a Security Group)
- Supports most AWS services
- $ per hour + $ per GB of data processed



Gateway Endpoints:

These are used **specifically for S3 and DynamoDB.** They add a target to your route tables that directs traffic destined for the specified service to the VPC endpoint, ensuring that traffic remains within the AWS network.

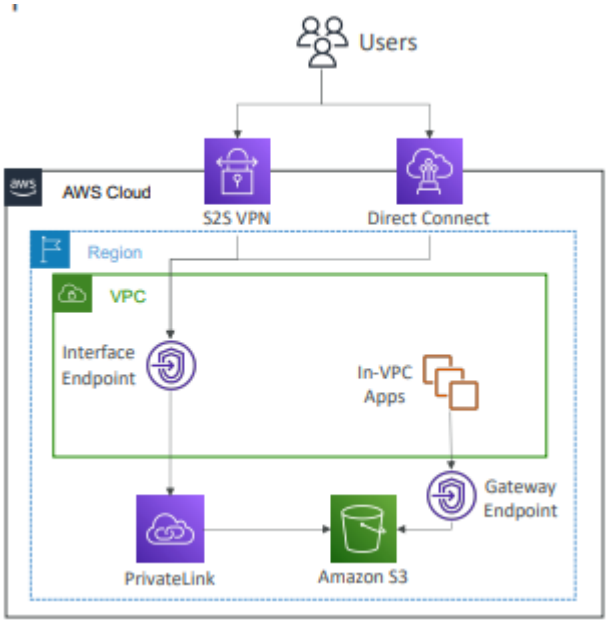
- Provisions a gateway and must be used as a target in a route table (does not use security groups)
- Supports both S3 and DynamoDB
- Free



401. Gateway or Interface Endpoint for S3?

Ans ->

- Gateway is most likely going to be preferred all the time at the exam
- **Cost: free for Gateway, $ for Interface Endpoint**
- Interface Endpoint is preferred where access is required from on-premises (Site to Site VPN or Direct Connect), a different VPC or a different region



402. What are VPC Flow Logs?

Ans ->

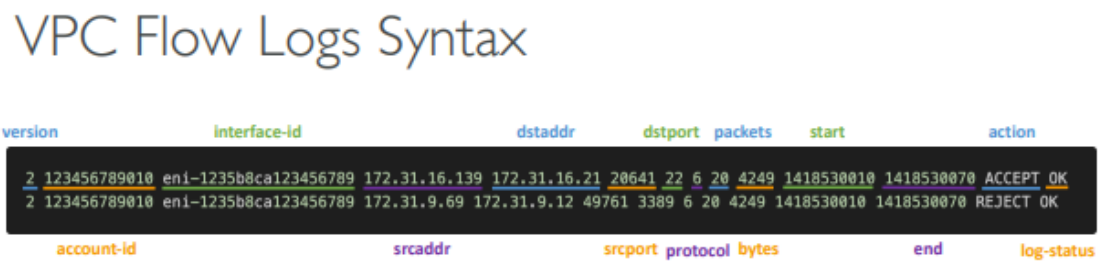
VPC Flow Logs capture detailed information about network traffic within your Virtual Private Cloud (VPC), including data like IP addresses, ports, and protocols. They help monitor network performance, troubleshoot issues, and enhance security by providing insights into traffic patterns.

Key points:

- **Capture information about IP traffic going into your interfaces:**
  - VPC Flow Logs
  - Subnet Flow Logs
  - Elastic Network Interface (ENI) Flow Logs
- Helps to monitor & troubleshoot connectivity issues
- Flow logs data can go to S3, CloudWatch Logs, and Kinesis Data Firehose
- Captures network information from AWS managed interfaces too: ELB, RDS, ElasticCache, Redshift, WorkSpaces, NATGW, Transit Gateway...

403. Explain about VPC Flow Logs Syntax.

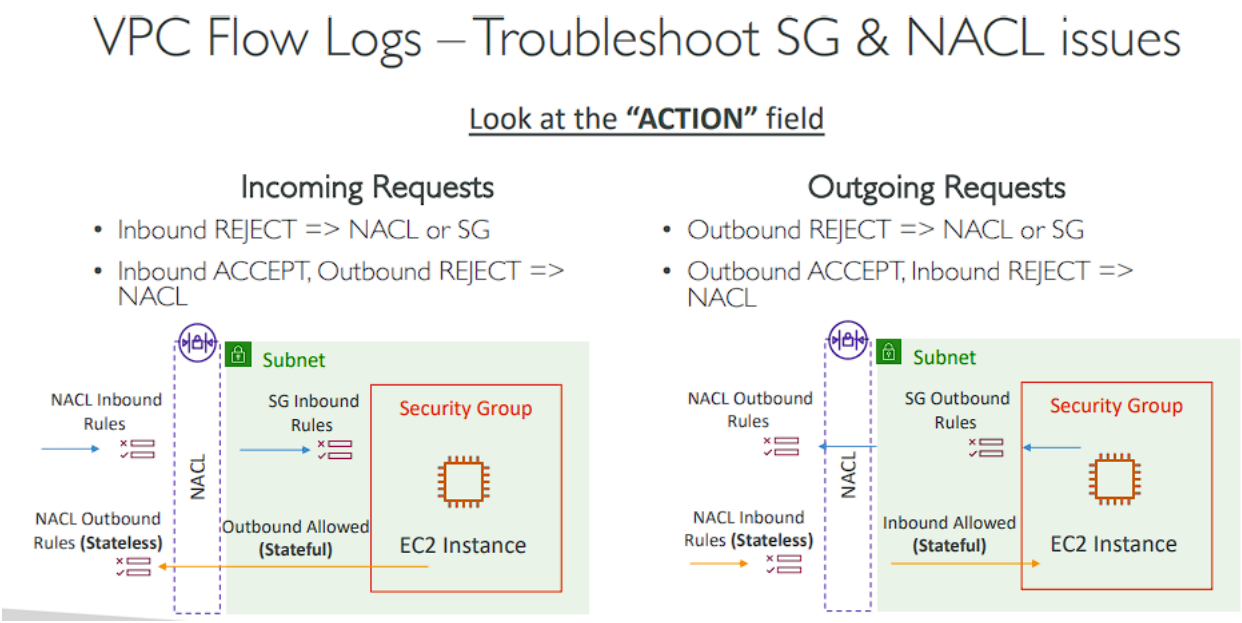
Ans ->



- **srcaddr & dstaddr** -> help identify problematic IP
- **srcport & dstport** -> help identity problematic ports
- **Action** - success or failure of the request due to Security Group / NACL
- Can be used for analytics on usage patterns, or malicious behavior
- **Query VPC Flow Logs using Athena on S3 or CloudWatch Logs Insights**
- **Flow Logs examples:** https://docs.aws.amazon.com/vpc/latest/userguide/flow-logs-records-examples.html

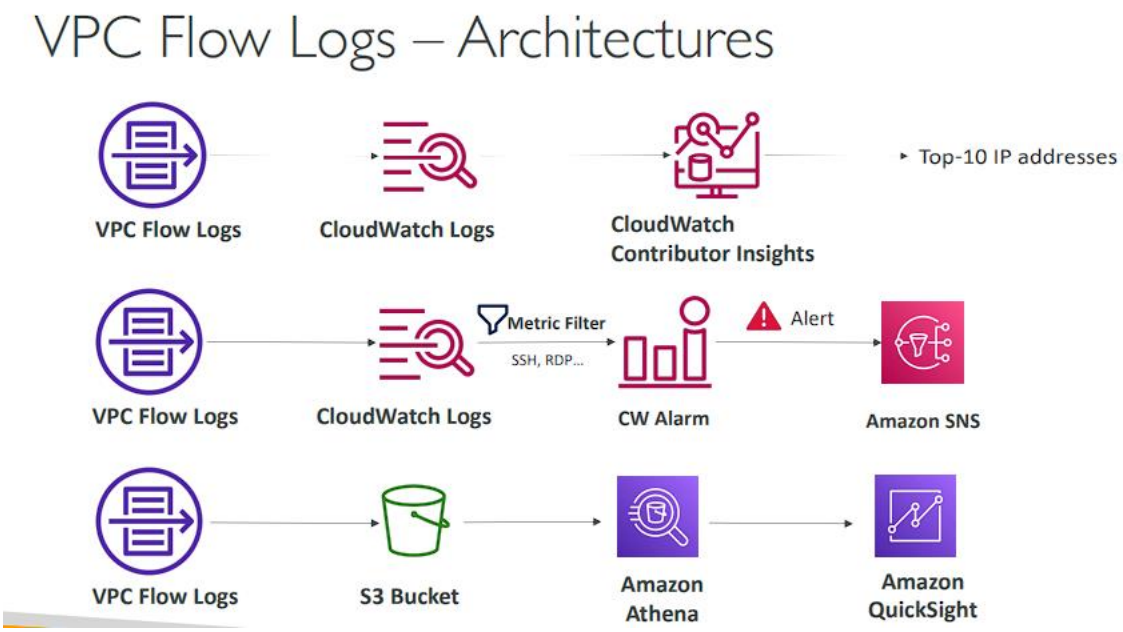
404. VPC Flow Logs - Troubleshoot SG & NACL issues.

Ans ->



405. VPC Flow Logs - Architectures.

Ans ->



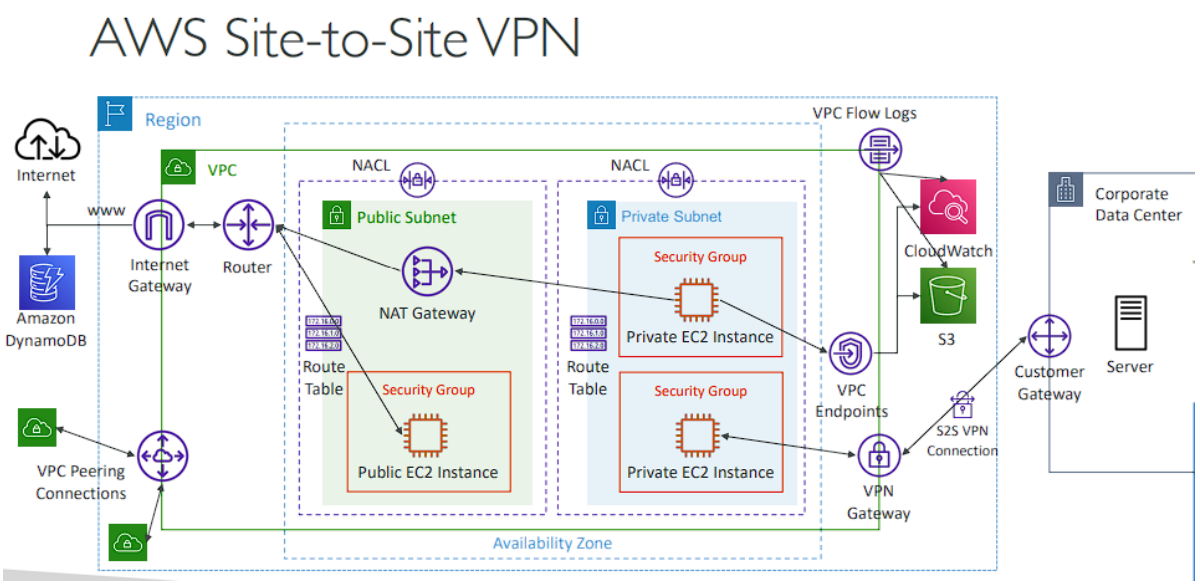
406. Explain about AWS Site-to-Site VPN.

Ans ->

A **Site-to-Site VPN** in AWS is a secure connection between your on-premises network and your AWS Virtual Private Cloud (VPC) over the internet. It allows your on-premises systems to communicate securely with resources in your AWS VPC as if they were on the same local network.

Key Components:

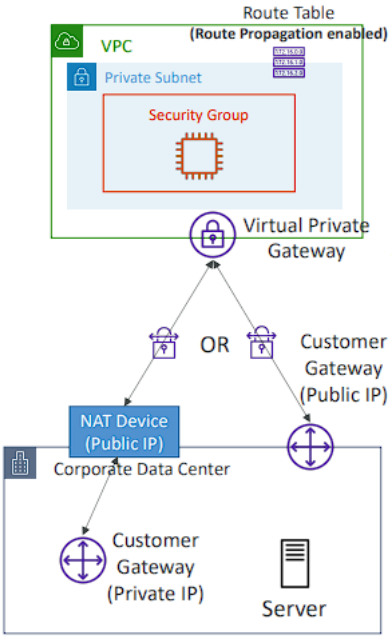
- **Virtual Private Gateway (VGW):**
  - This is the AWS side of the VPN connection. It attaches to your VPC and handles the VPN traffic coming from the on-premises network.
  - VPN concentrator on the AWS side of the VPN connection
  - VGW is created and attached to the VPC from which you want to create the Site-to-Site VPN connection
  - Possibility to customize the ASN (Autonomous System Number)
- **Customer Gateway (CGW):**
  - This represents your on-premises router or firewall. It manages the VPN connection on your local network side.
  - Software application or physical device on customer side of the VPN connection



407. Key points about Site-to-Site VPN Connections:

Ans ->

- **Customer Gateway Device (On-premises)**
  - **What IP address to use?**
    - Public Internet-routable IP address for your Customer gateway device
    - If it's behind a NAT device that's enabled for NAT traversal (NAT-T), use the public IP address of the NAT device
- **Important step:** enable **Route Propagation** for the **Virtual Private Gateway** in the route table that is associated with your subnets
- If you need to ping your EC2 instances from on-premises, make sure you add the **ICMP protocol** on the inbound of your security groups



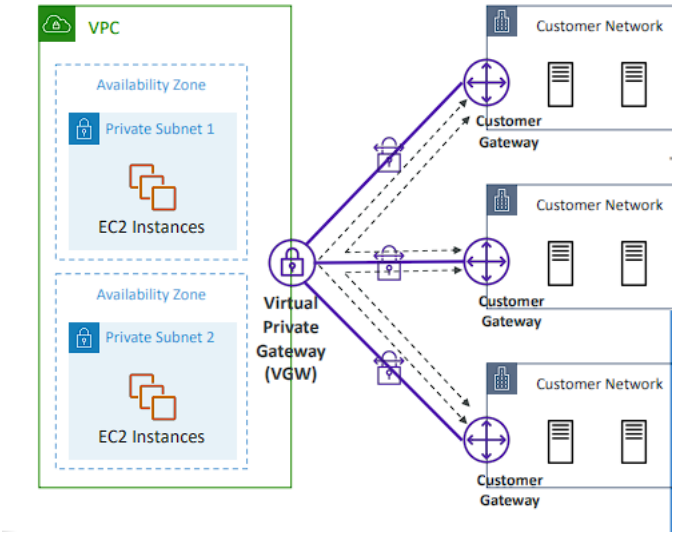
408. Explain about AWS VPN CloudHub.

Ans ->

AWS VPN CloudHub allows you to connect multiple on-premises sites (like branch offices) to each other and to your AWS VPC using a **hub-and-spoke model**. Each site connects via a **Site-to-Site VPN** to the VPC, enabling secure communication between them over the internet. It's cost-effective and supports dynamic routing, making it ideal for multi-site connectivity and disaster recovery scenarios.

Key points:

- Provide secure communication between multiple sites, if you have multiple VPN connections
- Low-cost **hub-and-spoke model** for primary or secondary network connectivity between different locations (VPN only)
- It's a VPN connection so it goes over the public Internet
- To set it up, connect multiple VPN connections on the same VGW, setup dynamic routing and configure route tables



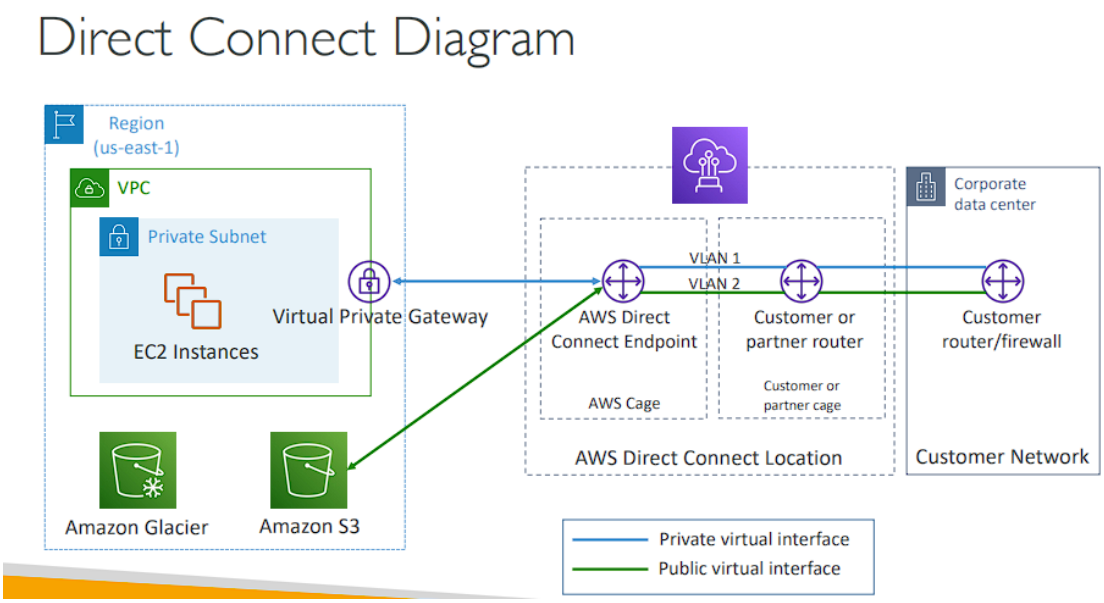
409. What is Amazon Direct Connect (DX)?

Ans ->

Amazon Direct Connect (DX) is a service that provides a dedicated, private network connection from your data center or office to AWS, offering lower latency, higher bandwidth, and more secure data transfer compared to standard internet connections.

Key points:

- Provides a dedicated private connection from a remote network to your VPC
- Dedicated connection must be setup between your Data Center and AWS Direct Connect locations
- You need to setup a Virtual Private Gateway on your VPC
- Access public resources (S3) and private (EC2) on same connection
- **Use Cases:**
  - Increase bandwidth throughput - working with large data sets - lower cost
  - More consistent network experience - applications using real-time data feeds
  - Hybrid Environments (on prem + cloud)
- Supports both IPv4 and IPv6

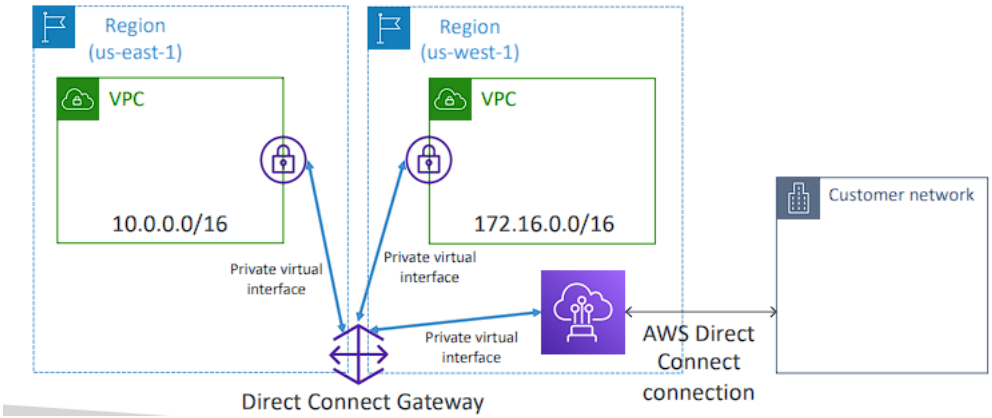


410. What is Direct Connect Gateway?

Ans ->

A **Direct Connect Gateway** allows you to **connect a single Amazon Direct Connect link** to multiple VPCs across different AWS regions, simplifying network management and providing inter-region connectivity.

- If you want to setup a Direct Connect to one or more VPC in many different regions (same account), you must use a Direct Connect Gateway.



411. Explain about Direct Connect - Connection Types.

Ans ->

Amazon Direct Connect offers two primary types of connections:

Dedicated Connections:

- **Description:** A dedicated, physical connection between your data center or office and an AWS Direct Connect location.
- **Capacity:** Typically, available in 1 Gbps, 10 Gbps, and sometimes 100 Gbps options.
- Request made to AWS first, then completed by AWS Direct Connect Partners

Hosted Connections:

- **Description:** A connection provided by an AWS Direct Connect Partner, which aggregates multiple customers' connections over a shared physical link to AWS
- **Capacity:** Available in 50 Mbps, 100 Mbps, 200 Mbps, 500 Mbps, 1 Gbps, and 10 Gbps options, capacity can be added or removed on demand.
- Connection requests are made via AWS Direct Connect Partners

Lead times are often longer than 1 month to establish a new connection

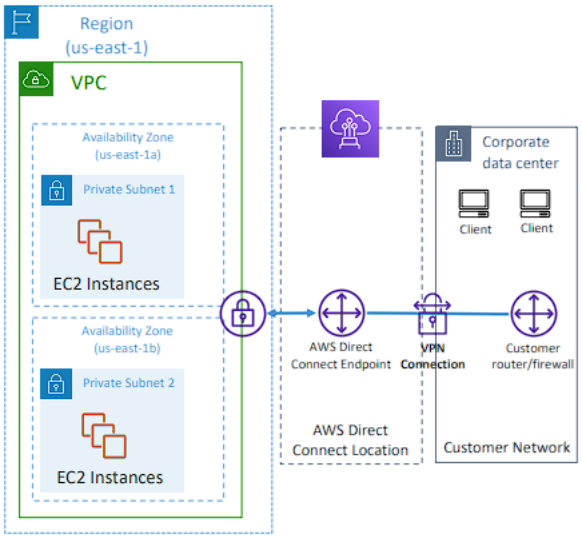
412. Explain about Direct Connect - Encryption.

Ans ->

Amazon Direct Connect does not provide built-in encryption. To secure data, you can **use IPsec VPN for encryption** over Direct Connect or encrypt data at the application level.

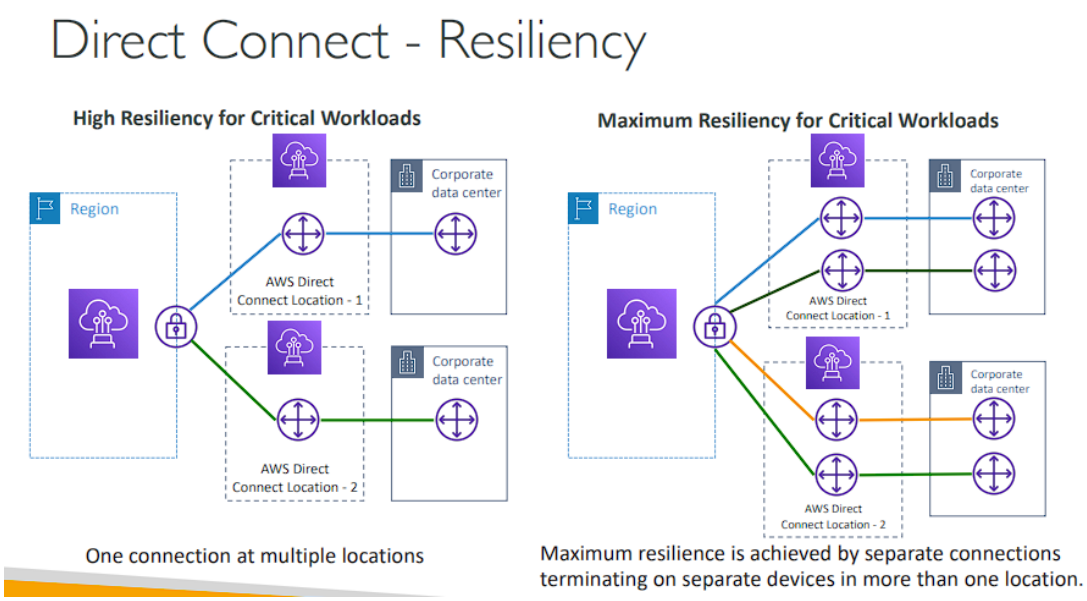
Key points:

- Data in transit is not encrypted but is private
- AWS Direct Connect + VPN provides an IPsec-encrypted private connection
- Good for an extra level of security, but slightly more complex to put in place



413. Explain about Direct Connect - Resiliency.

Ans ->



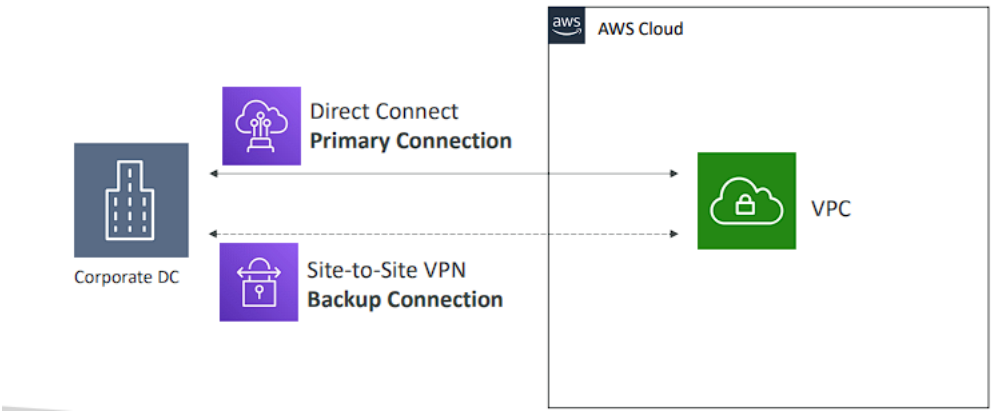
414. Site-to-Site VPN connection as a backup connection.

Ans ->

Using a site-to-site VPN connection as a backup for Amazon Direct Connect is a common practice to ensure high availability and continuity of connectivity. Here's how it works:

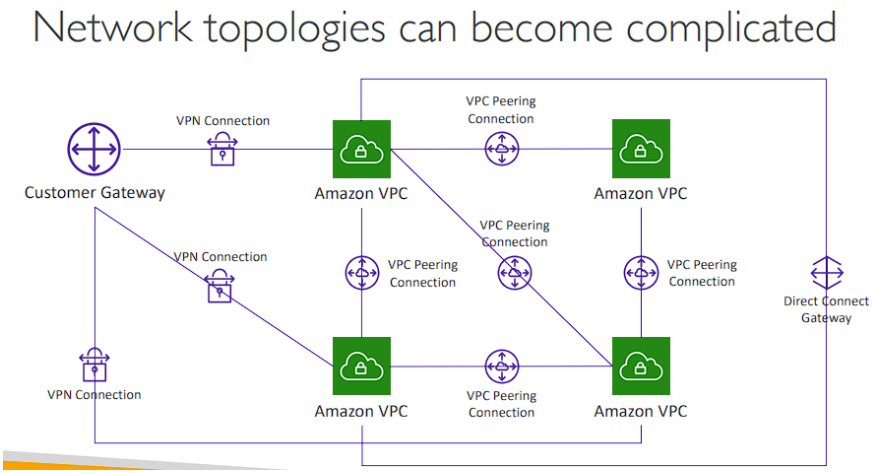
1. **Primary Connection:** Your main connection is Amazon Direct Connect, which provides a high-bandwidth, low-latency, and secure link between your on-premises network and AWS.
2. **Backup Connection:** You set up a site-to-site VPN connection over the public internet as a secondary link. This VPN tunnel provides a fallback option if your Direct Connect link becomes unavailable.

In case Direct Connect fails you can set up a backup Direct Connect connection (expensive), or a Site-to-Site connection (less expensive).



415. What is AWS Transit Gateway?

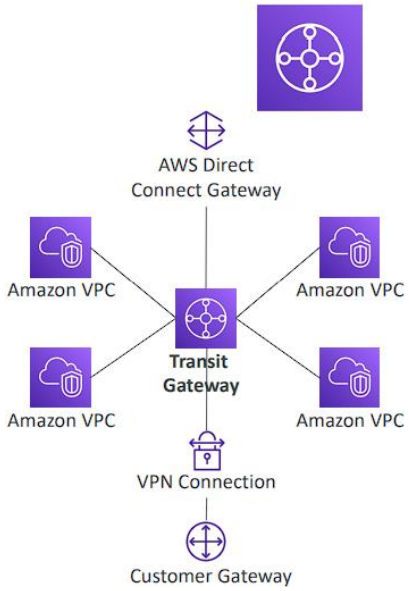
Ans ->



AWS Transit Gateway is a service that connects and manages multiple VPCs, on-premises networks, and VPNs through a central hub, simplifying and scaling your network architecture. It supports inter-region connectivity, allowing VPCs in different regions to communicate. With Transit Gateway, you can efficiently manage routing, implement network segmentation, and integrate with AWS Direct Connect and Site-to-Site VPN for a comprehensive network solution.

Key points:

- For having transitive peering between thousands of VPCs and on-premises, hub-and-spoke (star) connection
- Regional resource, can work cross-region
- Share cross-account using Resource Access Manager (RAM)
- You can peer Transit Gateways across regions
- **Route Tables:** limit which VPC can talk with other VPC
- Works with Direct Connect Gateway, VPN connections
- **Supports IP Multicast (not supported by any other AWS services)**



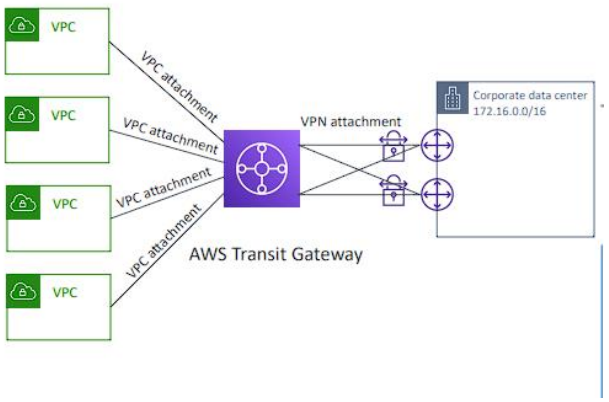
416. How to increase the bandwidth of Site-to-Site VPN using Transit Gateway: ECMP?

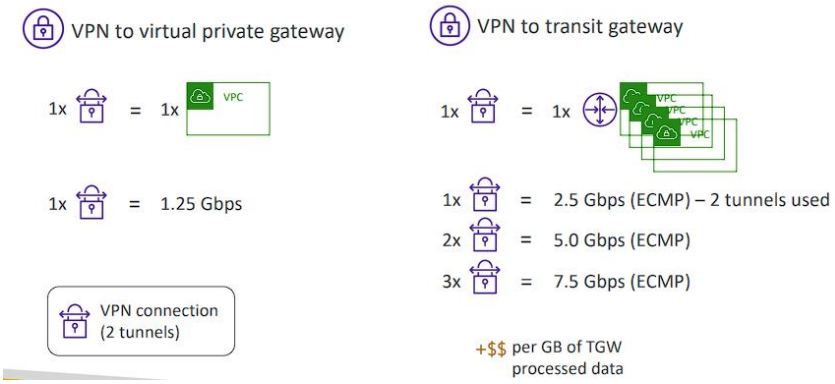
Ans ->

To increase the bandwidth of a Site-to-Site VPN using AWS Transit Gateway, enable Equal-Cost Multi-Path (ECMP) routing. ECMP allows AWS Transit Gateway to distribute traffic across multiple VPN connections, effectively increasing the total available bandwidth by balancing the load. Ensure your on-premises device supports ECMP for optimal traffic distribution.

Routing strategy to allow to forward a packet over multiple best path

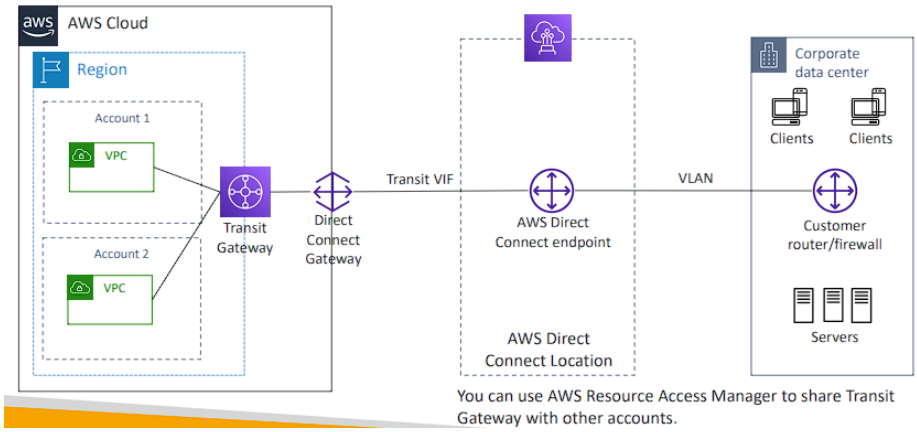
Use case: create multiple Site-to-Site VPN connections to increase the bandwidth of your connections to AWS.





417. Transit Gateway - Share Direct Connect between multiple accounts.

Ans ->



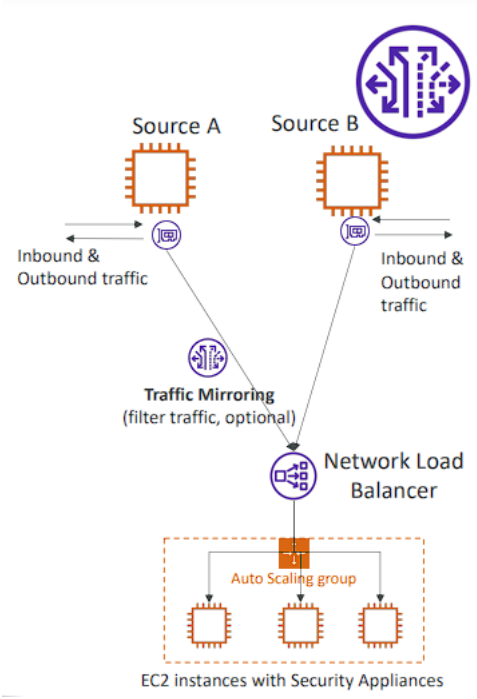
418. Explain about VPC - Traffic Mirroring

Ans ->

VPC Traffic Mirroring in AWS allows you to capture and view network traffic from your VPC by copying it to a specified destination. This helps you monitor and analyze network activity for security issues, performance problems, and troubleshooting, by sending a copy of the traffic to a tool or instance where you can examine it.

Key points:

- Allows you to capture and inspect network traffic in your VPC
- Route the traffic to security appliances that you manage
- **Capture the traffic:**
  - From (Source) - ENIs
  - To (Targets/Destination) - an ENI or Network Load balancer
- Capture all packets or capture the packets of your interest (optionally, truncate packets)
- Source and Target can be in the same VPC or different VPCs (VPC Peering)
- **Use cases:** content inspection, threat monitoring, troubleshooting, ...



419. If you cannot launch an EC2 instance in your subnet then what could be the possible reason behind this and what will be the solution?

Ans ->

- IPv4 cannot be disabled for your VPC and subnets
- So, if you cannot launch an EC2 instance in your subnet then:
  - It's not because it cannot acquire an IPv6 (the space is very large)
  - It's because there are no available IPv4 in your subnet
- **Solution:** create a new IPv4 CIDR in your subnet

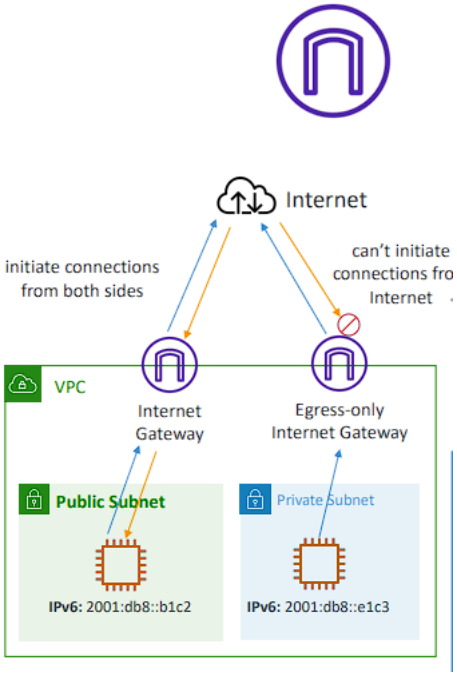
420. Explain about Egress-only Internet Gateway.

Ans ->

An **Egress-only Internet Gateway** is an AWS resource that allows instances in a VPC to access the internet for outbound traffic while blocking inbound traffic. It's specifically designed for IPv6 traffic, enabling instances to make requests to the internet without receiving any direct responses or inbound connections.

Key points:

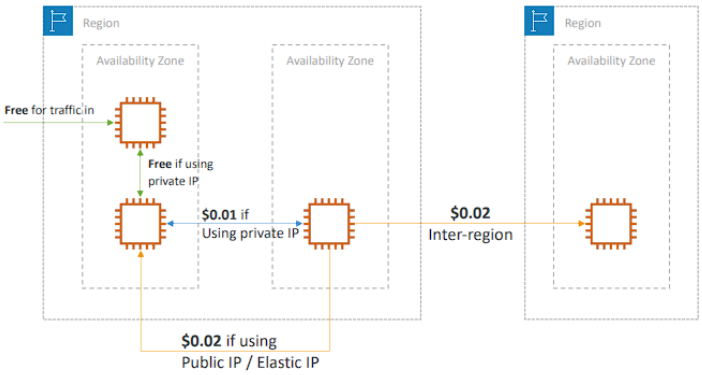
- **Used for IPv6 only**
- similar to a NAT Gateway but for IPv6
- Allows instances in your VPC outbound connections over IPv6 while preventing the internet to initiate an IPv6 connection to your instances
- You must update the Route Tables



421. Networking Costs in AWS per GB - Simplified.

Ans ->

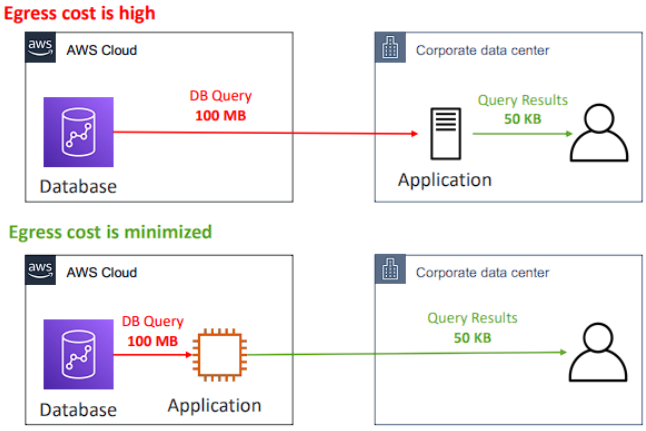
- Use Private IP instead of Public IP for good savings and better network performance
- Use same AZ for maximum savings (at the cost of high availability)



422. Minimizing egress traffic network cost.

Ans ->

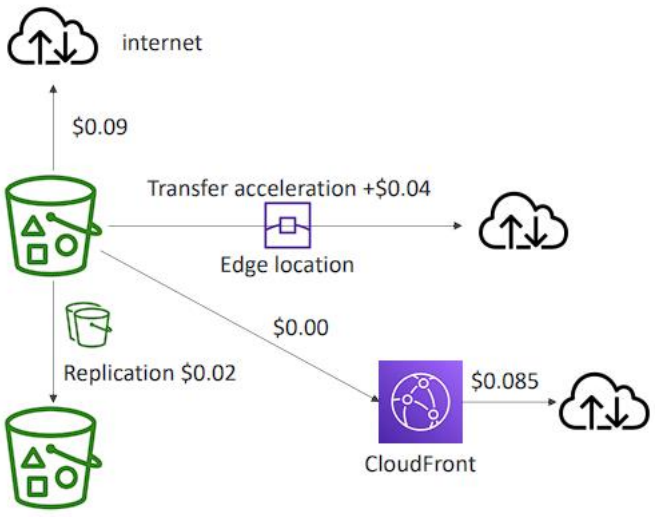
- **Egress traffic:** outbound traffic (from AWS to outside)
- **Ingress traffic:** inbound traffic - from outside to AWS (typically free)
- Try to keep as much internet traffic within AWS to minimize costs
- Direct Connect location that are co-located in the same AWS Regions result in lower cost for egress network



423. S3 Data Transfer Pricing - Analysis for USA.

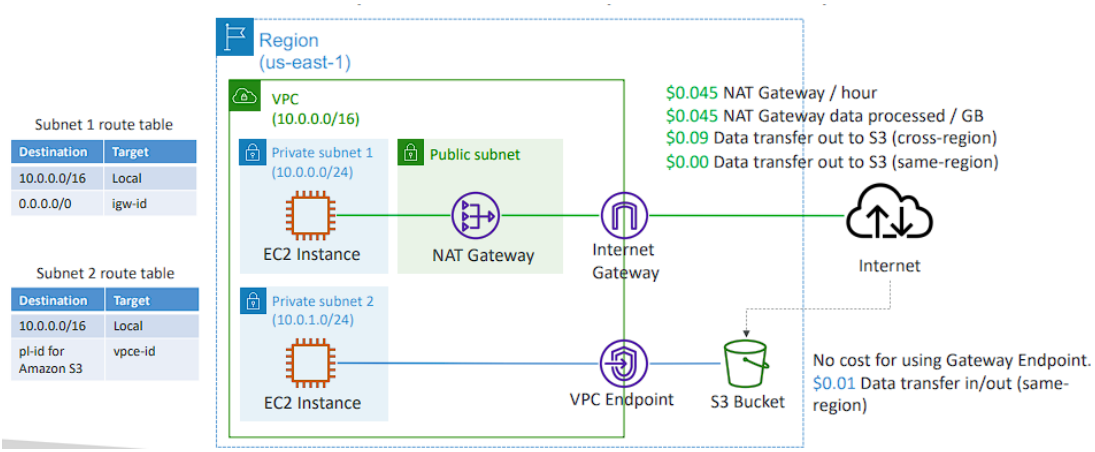
Ans ->

- **S3 ingress:** free
- **S3 to Internet:** $0.09 per GB
- **S3 Transfer Acceleration:**
  - Faster transfer times (50 to 500% better)
  - Additional cost on top of Data Transfer Pricing: +$0.04 to $0.08 per GB
- **S3 to CloudFront:** $0.00 per GB (Free)
- **CloudFront to Internet:** $0.085 per GB (slightly cheaper than S3)
  - Caching capability (lower latency)
  - Reduce costs associated with S3 Requests Pricing (7x cheaper with CloudFront)
- **S3 Cross Region Replication:** $0.02 per GB



424. Pricing: NAT Gateway vs Gateway VPC Endpoint.

Ans ->



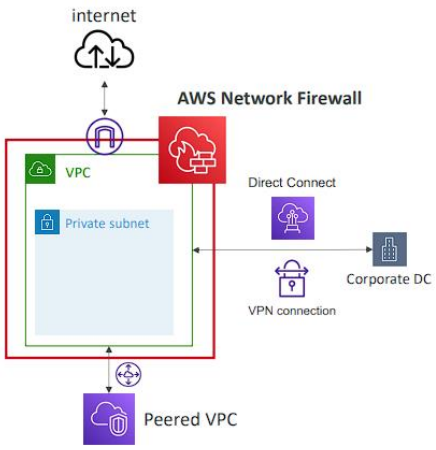
425. What is AWS Network Firewall?

Ans ->

AWS Network Firewall is a fully managed security service that allows you to easily deploy and manage network protection for your Amazon VPC. It provides both stateful and stateless packet filtering, traffic inspection, and intrusion prevention/detection to secure your cloud workloads. With AWS Network Firewall, you can define custom rules to control inbound and outbound traffic, protect your network from unwanted access or threats, and meet compliance needs.

Key points:

- Protect your entire Amazon VPC
- From Layer 3 to Layer 7 protection
- Any direction, you can inspect
  - VPC to VPC traffic
  - Outbound to internet
  - Inbound from internet
  - To/from Direct Connect & Site-to-Site VPN
- Internally, the AWS Network Firewall uses the AWs Gateway Load Balancer
- Rules can be centrally managed cross-account by AWs Firewall Manager to apply to many VPCs.



426. Explain about Network Firewall - Fine Grained Controls.

Ans ->

- Supports 1000s of rules
  - IP & port - example: 10,000s of IPs filtering
  - Protocol - example: block the SMB protocol for outbound communications
  - Stateful domain list rule groups: only allow outbound traffic to *.mycorp.com or third-party software repo
  - General pattern matching using regex
- **Traffic filtering:** Allow, drop, or alert for the traffic that matches the rules
- **Active flow inspection** to protect against network threats with intrusion-prevention capabilities (like Gateway Load Balancer, but all managed by AWS)
- Send logs of rule matches to Amazon S3, CloudWatch Logs, Kinesis Data Firehose.

427. What is RPO and RTO in AWS Disaster recovery?

Ans ->

RPO (Recovery Point Objective) and **RTO (Recovery Time Objective)** are two key metrics in disaster recovery.

RPO (Recovery Point Objective):

- It refers to the maximum acceptable amount of data loss measured in time. In simpler terms, it's the amount of time between the last backup and a potential failure. If your RPO is 4 hours, for example, this means you can afford to lose up to 4 hours of data in the event of a disaster.
- In AWS, this can be managed through regular backups, snapshots, or data replication to minimize data loss.

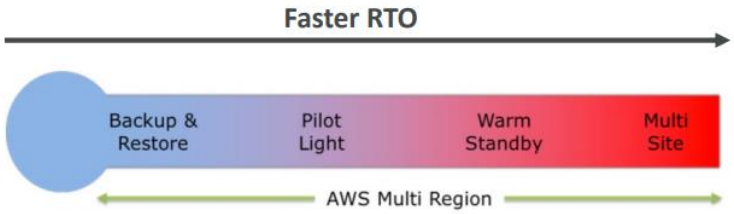
RTO (Recovery Time Objective):

- This is the maximum acceptable time it takes to restore services after a failure or disaster. It defines how quickly you need to recover to resume normal operations.
- AWS offers services like Elastic Disaster Recovery, RDS with automatic failover, and auto-scaling to help you achieve shorter RTOs.

428. What are the different types of Disaster Recovery Strategies?

Ans ->

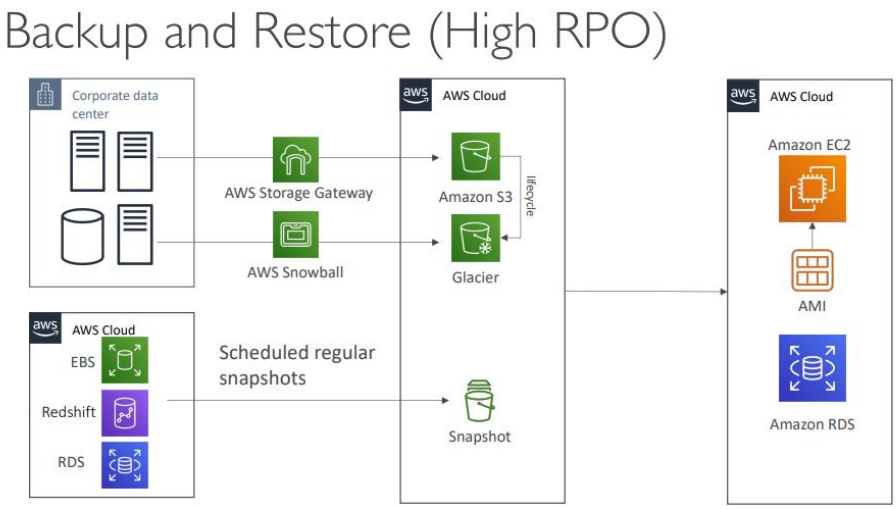
- Backup and Restore
- Pilot Light
- Warm Standby
- Hot Site / Multi Site Approach



429. Explain about Backup and Restore - Disaster Recovery Strategy.

Ans ->

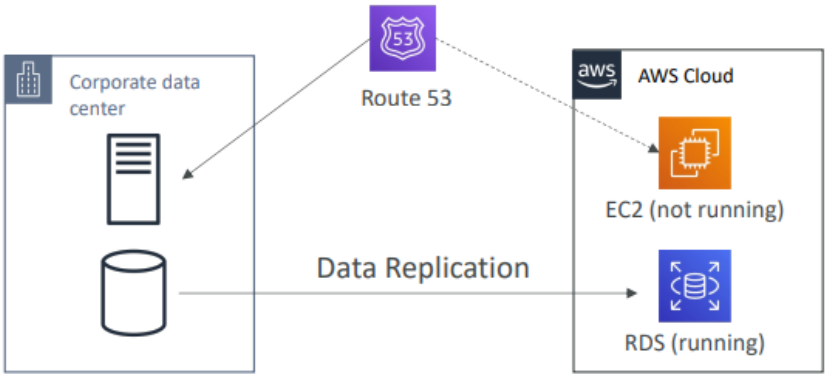
- **Overview:** Data is periodically backed up and stored (e.g., to S3), and in the event of a disaster, the system is restored from these backups.
- **RTO/RPO:** High RTO, High RPO (recovery can take longer, and data loss since the last backup might be significant).
- **Cost:** Low (only paying for storage and infrequent recovery operations).
- **AWS Services:** S3, Glacier, AWS Backup.



430. Explain about Pilot Light - Disaster Recovery Strategy.

Ans ->

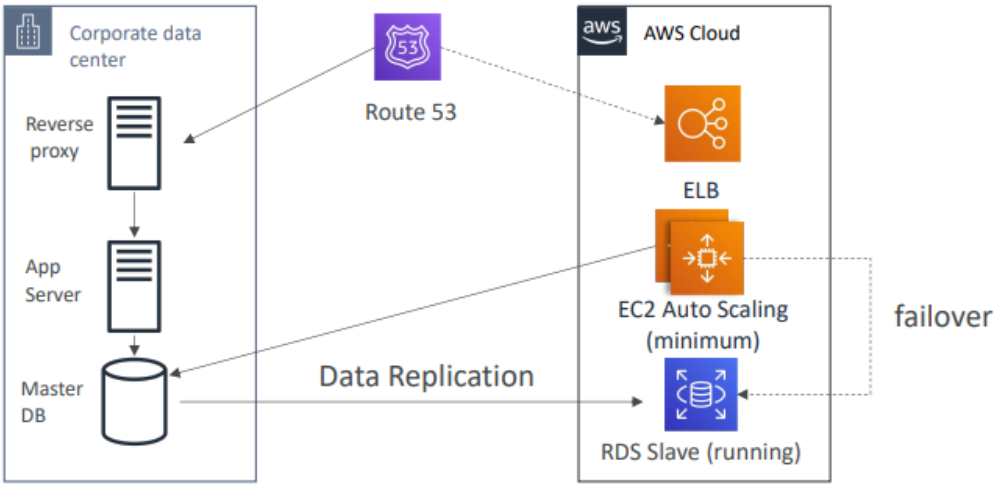
- **Overview:** A minimal version of the environment is always running (critical core services only), and in the event of a disaster, you can "scale up" by turning on additional resources to restore full functionality.
- **RTO/RPO:** Medium RTO, Low RPO (recovery time is faster because critical systems are pre-configured).
- **Cost:** Moderate (you maintain a small portion of the environment).
- **AWS Services:** EC2, RDS, AMIs, CloudFormation.
- Faster than Backup and Restore as critical systems are already up



431. Explain about Warm Standby - Disaster Recovery Strategy.

Ans ->

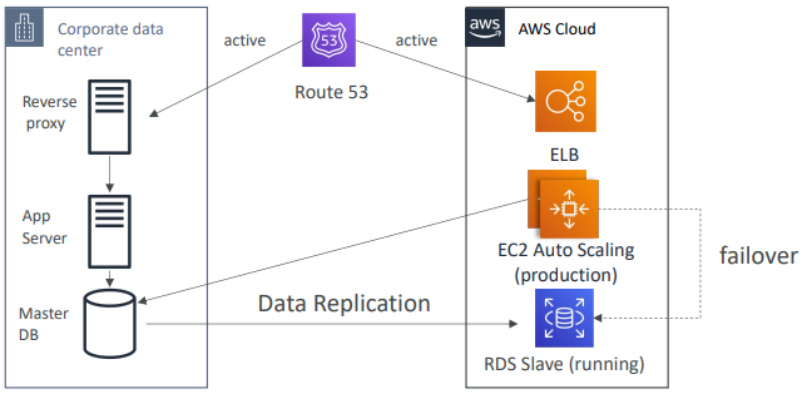
- **Overview:** A scaled-down but fully functional version of the system is running. During a disaster, you scale it up to handle the full production load.
- **RTO/RPO:** Low RTO, Low RPO (since a live environment exists, recovery is faster).
- **Cost:** Higher than Pilot Light, as a part of the system is already operational.
- **AWS Services:** Elastic Load Balancing (ELB), Auto Scaling, EC2, RDS.



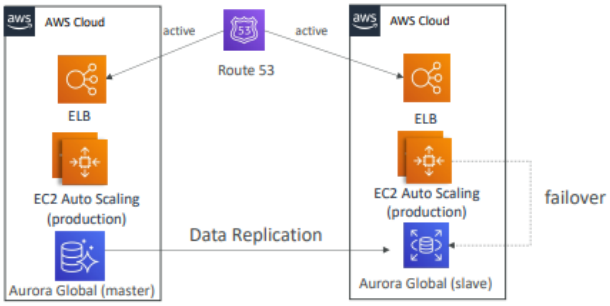
432. Explain about Multi Site / Hot Site Approach - Disaster Recovery Strategy.

Ans ->

- **Overview:** Full production workloads run in two or more locations (data centers or regions), so in case of a disaster, the load can be balanced or shifted to the unaffected location.
- **RTO/RPO:** Near-Zero RTO, Near-Zero RPO (because the system is fully operational in multiple locations).
- **Cost:** High (maintaining fully redundant infrastructure).
- **AWS Services:** Route 53, Global Accelerator, Auto Scaling, S3 Cross-Region Replication, DynamoDB Global Table



All AWS Multi Region



433. Disaster Recovery Tips.

Ans ->

Backup:

- EBS Snapshots, RDS automated backups / Snapshots, etc...
- Regular pushes to S3 / S3 IA / Glacier, Lifecycle Policy, Cross Region Replication
- From On-Premise: Snowball or Storage Gateway

High Availability:

- Use Route 53 to migrate DNS over from Region to Region
- RDS Multi-AZ, ElastiCache Multi-AZ, EFS, S3
- Site to Site VPN as a recovery from Direct Connect

Replication:

- RDS Replication (Cross Region), AWS Aurora + Global Databases
- Database replication from on-premise to RDS
- Storage Gateway

Automation:

- CloudFormation /Elastic Beanstalk to re-create a whole new environment
- Recovery / Reboot EC2 instances with CloudWatch if alarms fail
- AWS Lambda functions for customized automations

Chaos:

- Netflix has a "simian-army" for randomly terminating EC2 instances

434. Explain about AWS DMS.

Ans ->

AWS DMS (Database Migration Service) is a managed service that simplifies and streamlines the migration of databases to and from AWS. It supports a variety of database engines, allowing for minimal downtime during migrations and enabling continuous data replication. DMS also provides schema conversion to ensure compatibility between source and target databases.

Key points:

- Quickly and securely migrate databases to AWS, resilient, self-healing
- The source database remains available during the migration
- **Supports:**
  - **Homogeneous migrations:** ex -> Oracle to Oracle
  - **Heterogeneous migrations:** ex -> Microsoft SQL Server to Aurora
- Continuous Data Replication using CDC
- You must create an EC2 instance to perform the replication tasks

435. DMS Sources and Targets

Ans ->

SOURCES:

- On-Premises and EC2 instances databases: Oracle, MS SQL Server, MySQL, MariaDB, PostgreSQL, MongoDB, SAP, DB2
- Azure: Azure SQL Database
- Amazon RDS: all including Aurora
- Amazon S3
- DocumentDB

TARGETS:

- On-Premises and EC2 instances databases: Oracle, MS SQL Server, MySQL, MariaDB, PostgreSQL, SAP
- Amazon RDS
- Redshift, DynamoDB, S3
- OpenSearch Service
- Kinesis Data Streams
- Apache Kafka

- DocumentDB & Amazon Neptune
- Redis & Babelfish

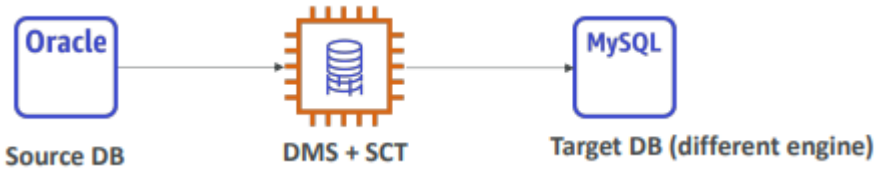
436. What is AWS SCT?

Ans ->

AWS SCT (Schema Conversion Tool) is a service that helps users convert database schemas from one database engine to another. It is particularly useful when migrating databases to AWS, as it assists in translating the database structure, including tables, views, indexes, and stored procedures, to be compatible with the target database system. This tool is often used in conjunction with AWS DMS for comprehensive database migration projects.

Key points:

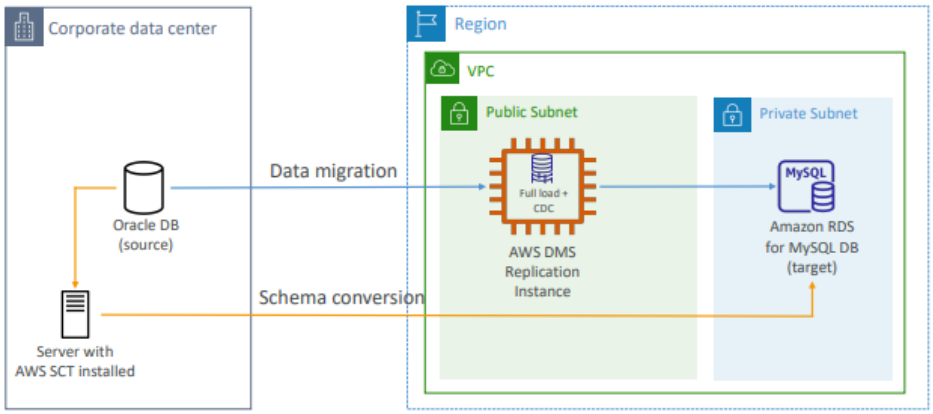
- Convert you Database's Schema from one engine to another
- **Example OLTP:** (SQL Server or Oracle) to MySQL, PostgreSQL, Aurora
- **Example OLAP:** (Teradata or Oracle) to Amazon Redshift



- **You do not need to use SCT if you are migrating the same DB engine**
  - **Ex:** On-Premises PostgreSQL => RDS PostgreSQL
  - The DB engine is still PostgreSQL (RDS is the platform)

437. DMS - Continuous Replication.

Ans ->



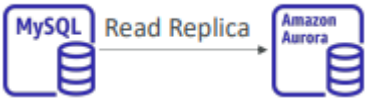
438. RDS MySQL or External MySQL to Aurora MySQL Migrations.

Ans ->

RDS MySQL to Aurora MySQL

- **Option 1:**
  - DB Snapshots from RDS MySQL restored as MySQL Aurora DB

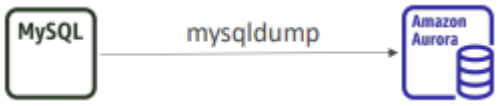
- **Option 2:**
  - Create an Aurora Read Replica from your RDS MySQL, and when the replication lag is 0, promote it as its own DB cluster (can take time and cost $)



External MySQL to Aurora MySQL

- **Option 1:**
  - Use Percona XtraBackup to create a file backup on Amazon S3
  - Create an Aurora MySQL DB from Amazon S3

- **Option 2:**
  - Create an Aurora MySQL DB
  - Use the mysqldump utility to migrate MySQL into Aurora (slower than S3 method)



Use DMS if both databases are up and running.

439. RDS PostgreSQL or External PostgreSQL to Aurora PostgreSQL Migrations.

Ans ->

RDS PostgreSQL to Aurora PostgreSQL

- **Option 1:** DB Snapshots from RDS PostgreSQL resorted as PostgreSQL Aurora DB

- **Option 2:** Create an Aurora Read Replica from your RDS PostgreSQL, and when the replication lag is 0, promote it as its own DB cluster (can take time and cost $)

External PostgreSQL to Aurora PostgreSQL

- Create a backup and put it in Amazon S3
- Import it using the aws_s3 Aurora extension

Use DMS if both databases are up and running

440. On-Premises strategy with AWS.

Ans ->

Ability to download Amazon Linux 2 AMI as a VM (.iso format)

- VMWare, KVM, VirtualBox (Oracle VM), Microsoft Hyper-V

VM Import / Export

- Migrate existing applications into EC2
- Create a DR repository strategy for your on-premises VMs
- Can export back the VMs from EC2 to on-premises

AWS Application Discovery Service (ADS)

- Gather information about your on-premises servers to plan a migration
- Server utilization and dependency mappings
- Track with AWS Migration Hub

AWS Database Migration Service (DMS)

- replication On-premises => AWS, AWS => AWS, AWS => On-premises
- Works with various database technologies (Oracle, MySQL, DynamoDB, etc...)

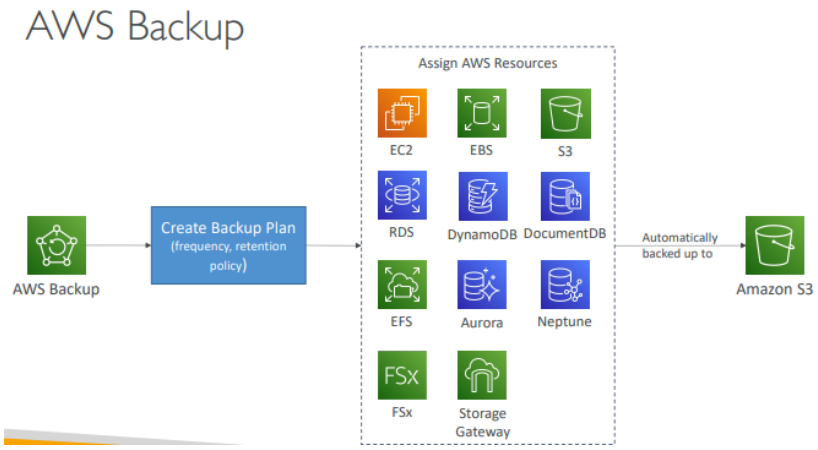
AWS Server Migration Service (SMS)

- Incremental replication of on-premises live servers to AWS

441. What is AWS Backup?

Ans ->

- Fully managed service
- Centrally managed and automate backups across AWS services
- No need to create custom scripts and manual processes
- **Supported services:**
  - Amazon EC2 / Amazon EBS
  - Amazon S3
  - Amazon RDS (all DBs engines) / Amazon Aurora / Amazon DynamoDB
  - Amazon DocumentDB / Amazon Neptune
  - Amazon EFS / Amazon FSx (Lustre & Windows Files Server)
  - AWS Storage Gateway (Volume Gateway)
  - etc...
- Supports cross-region backups
- Supports cross-account backups
- Supports PITR for supported services
- On-Demand and Scheduled backups
- Tag-based backup policies
- **You create backup policies known as Backup Plans:**
  - Backup frequency (every 12 hours, daily, weekly, monthly, cron expression)
  - Backup window
  - Transition to Cold Storage (Never, Days, Weeks, Months, Years)
  - Retention Period (Always, Days, Weeks, Months, Years)



442. Explain about AWS Backup Vault Lock.

Ans ->

Backup Vault Lock in AWS Backup enforces immutable retention policies, preventing the deletion or modification of backups for a specified period. This feature helps ensure compliance with regulations and protects against data loss while providing audit capabilities to track access and changes.

Key points:

- Enforce a **WORM (Write Once Read Many)** state for all the backups that you store in your AWS Backup Vault
- **Additional layer of defense to protect your backups against:**
  - Inadvertent or malicious delete operations
  - Updates that shorten or alter retention periods
  - Once a vault lock is applied, backup recovery points cannot be deleted or modified until the retention period expires.
  - Even the root user cannot delete backups when enabled

443. Explain about AWS Application Discovery Service (ADS).

Ans ->

AWS ADS helps organizations plan their cloud migration by automatically identifying and collecting information about on-premises applications, including their configurations and dependencies. It provides data visualization and integrates with migration tools to facilitate informed decision-making during the migration process.

Key points:

- Plan migration projects by gathering information about on-premises data centers Server utilization data and dependency mapping are important for migrations

- **Agentless Discovery (AWS Agentless Discovery Connector)**
  - VM inventory, configuration, and performance history such as CPU, memory, and disk usage
- **Agent-based Discovery (AWS Application Discovery Agent)**
  - System configuration, system performance, running processes, and details of the network connections between systems
- Resulting data can be viewed within AWS Migration Hub

444. What is AWS Migration Hub?

Ans ->

AWS Migration Hub is a service that provides a centralized platform to track and manage application migrations to AWS. It offers visibility into the migration process, helping organizations monitor progress and make informed decisions about their migration strategies.

Key Features:

- **Centralized Tracking:** Monitor the status of applications being migrated across multiple AWS and partner migration tools.
- **Integration with Migration Tools:** Works seamlessly with services like AWS Application Discovery Service, AWS Database Migration Service, and third-party migration tools.
- **Migration Planning:** Helps organizations assess application readiness and determine the best migration strategy.

445. Explain about AWS Application Migration Service (MGN)

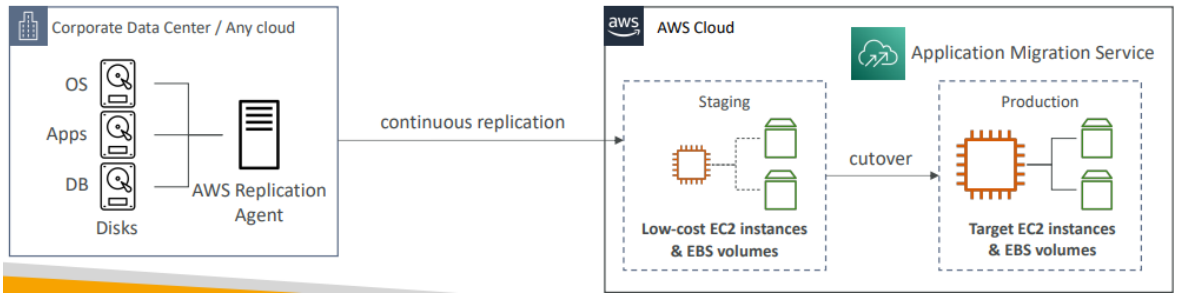
Ans ->

AWS Application Migration Service (MGN) is a service designed to simplify and accelerate the migration of on-premises applications to AWS. It automates the conversion of applications into AWS-native formats, allowing organizations to quickly and efficiently migrate workloads without significant re-architecting.

Key points:

- The "AWS evolution" of CloudEndure Migration, replacing AWS Server Migration Service (SMS)
- Lift-and-shift (rehost) solution which simplify migrating applications to AWS

- Converts your physical, virtual, and cloud-based servers to run natively on AWS
- Supports wide range of platforms, Operating Systems, and databases
- Minimal downtime reduced costs



446. Transferring large amount of data into AWS.

Ans ->

Example: transfer 200 TB of data in the cloud. We have a 100 Mbps internet connection.

Over the internet / Site-to-Site VPN:

- Immediate to setup
- Will take 200(TB)*1000(GB)*1000(MB)*8(Mb)/100 Mbps = 16,000,000s = 185 days

Over direct connect 1 Gbps:

- Long for the one-time setup (over a month)
- Will take 200(TB)*1000(GB)*8(Gb)/1 Gbps = 1,600,000s = 18.5 days

Over Snowball:

- Will take 2 to 3 snowballs in parallel
- Takes about 1 week for the end-to-end transfer
- Can be combined with DMS

For on-going replication / transfers: Site-to-Site VPN or DX with DMS or DataSync

447. What is VMware Cloud on AWS?

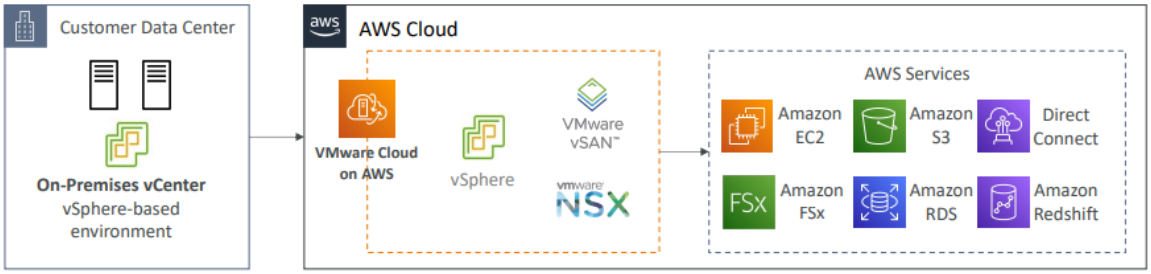
Ans ->

VMware Cloud on AWS is a hybrid cloud service that lets organizations run their VMware applications on AWS. It makes it easy to move workloads between on-premises VMware setups and the cloud, using familiar VMware tools. The service allows for flexible scaling of resources to meet demand and provides access to various

AWS services, all while maintaining a consistent VMware environment. This setup helps businesses enhance flexibility and innovation while managing their IT resources efficiently.

Key points:

- Some customers use VMware cloud to manage their on-premises Data Center
- They want to extend the Data Center capacity to AWS, but keep using the VMware Cloud Software
- **Use Cases:**
  - Migrate your VMware vSphere-based workloads to AWS
  - Run your production workloads across VMware vSphere-based private, public, and hybrid cloud environments
  - Have a disaster recovery strategy



448. To achieve Enhanced Networking in EC2 what are the things we can do?

Ans ->

Option 1:

- Enhanced Networking on Amazon EC2 using SR-IOV (Single Root I/O Virtualization) provides higher bandwidth, lower latency, and lower CPU utilization for your network traffic.

Option 2:

- Use Elastic Network Adapter (ENA), It supports higher bandwidth (up to 100 Gbps), lower latency, and improved packet per second (PPS) performance.

Option 3:

- Intel 82599 VF (up to 1 Gbps -LEGACY): For Intel-based instances, ensure the driver is installed if using older AMIs.

449. What is ENA?

Ans ->

Elastic Network Adapter (ENA) is a high-performance network interface for Amazon EC2 instances that provides enhanced networking capabilities, including higher

bandwidth (up to 100 Gbps), lower latency, and improved packet processing. Designed for a wide range of applications, ENA enables efficient data transfer and scalability, making it ideal for workloads such as big data processing, web applications, and distributed systems.

450. What is EFA?

Ans ->

Elastic Fabric Adapter (EFA) is a network interface designed by AWS specifically for **High-Performance Computing (HPC)** applications and workloads that require low-latency, high-throughput communication. EFA provides enhanced networking capabilities beyond what is offered by standard Elastic Network Adapters (ENA), making it suitable for tightly-coupled workloads, such as machine learning, computational fluid dynamics, and other distributed applications.

Key points:

- Improved ENA for HPC, only works for Linux
- Great for inter-node communications, tightly coupled workloads
- Leverages Message Passing Interface (MPI) standard
- Bypasses the underlying Linux OS to provide low-latency, reliable transport

451. What are the differences between ENA and EFA?

Ans ->

- **Purpose:**
  - **ENA:** General-purpose high-performance networking.
  - **EFA:** High-performance computing (HPC).
- **Performance:**
  - **ENA:** High bandwidth (up to 100 Gbps), low latency.
  - **EFA:** Ultra-low latency, high throughput.
- **RDMA Support:**
  - **ENA:** No.
  - **EFA:** Yes.
- **Use Cases:**
  - **ENA:** Suitable for web hosting, databases, and big data processing.
  - **EFA:** Ideal for HPC applications, machine learning, and simulations.
- **Instance Types:**
  - **ENA:** Available on most current generation EC2 instances.

    - **EFA:** Limited to specific HPC-optimized instances.
- **Packet Handling:**
    - **ENA:** Handles moderate packets per second (PPS) performance.
    - **EFA:** Designed for high PPS, ideal for inter-instance communication.
- **Networking Model:**
    - **ENA:** Standard TCP/IP networking.
    - **EFA:** Optimized for tightly-coupled workloads.

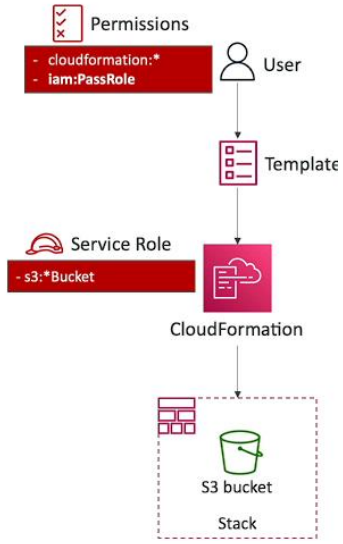
452. Explain about AWS CloudFormation - Service Role.

Ans ->

In **AWS CloudFormation**, a **service role** is a special IAM role used to manage AWS resources during stack operations. When you create or update a CloudFormation stack, AWS needs permissions to perform actions on your behalf, such as creating EC2 instances, creating S3 buckets, etc. The service role defines what actions CloudFormation can perform and which AWS resources it can access.

Key points:

- IAM role that allows CloudFormation to create/update/delete stack resources on your behalf
- Give ability to users to create/update/delete the stack resources even if they don't have permission to work with the resources in the stack
- **Use cases:**
    - You want to achieve the least privilege principle
    - But you don't want to give the user all the required permissions to create the stack resources
    - User must have iam:PassRole permissions



453. What is AWS SES?

Ans ->

AWS SES (Simple Email Service) is a cloud-based service that allows businesses and developers to send emails, whether for notifications, marketing, or bulk messages. It's easy to scale, so you can send a few or millions of emails. You only pay for what you use, and SES ensures emails are more likely to reach inboxes, not spam. It supports features like domain authentication, and you can also receive emails through it.

Key points:

- Fully managed service to send emails securely, globally and at scale
- Allows inbound/outbound emails
- Reputation dashboard, performance insights, anti-spam feedback
- Provides statistics such as email deliveries, bounces, feedback loop results, email open
- Supports DomainKeys Identified Mail (DKIM) and Sender Policy Framework (SPF)
- **Flexible IP deployment:** shared, dedicated, and customer-owned IPs
- Send email using your application using AWS Console, APIs, or SMTP
- **Use cases:** transactional, marketing and bulk email communications

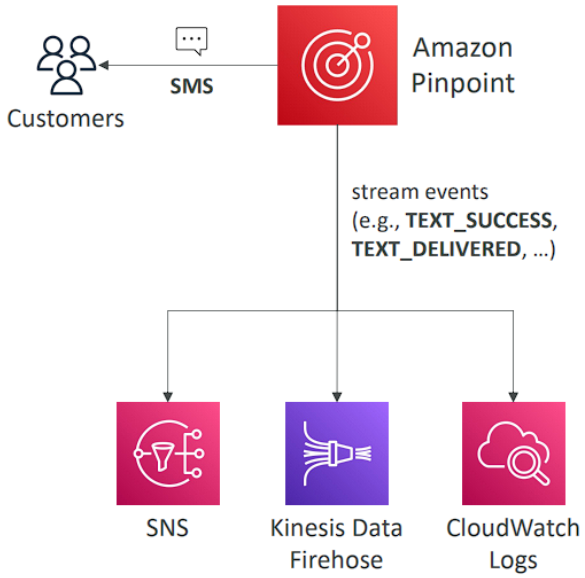
454. What is Amazon Pinpoint?

Ans ->

AWS Pinpoint is a service for engaging customers via email, SMS, push notifications, and voice. It helps businesses run marketing campaigns and send personalized messages at scale. With built-in analytics, you can track and improve customer targeting based on user behavior, making it ideal for sending promotions, app notifications, or transactional messages.

Key points:

- Scalable 2-way (outbound/inbound) marketing communications service
- Supports email, SMS, push, voice, and in-app messaging
- Ability to segment and personalize messages with the right content to customers
- Possibility to receive replies
- Scales to billions of messages per day
- **Use cases:** run campaigns by sending marketing, bulk, transactional SMS messages
- **Versus Amazon SNS or Amazon SES:**
  - In SNS & SES you managed each message's audience, content, and delivery schedule
  - In Amazon Pinpoint, you create message templates, delivery schedules, highly-targeted segments, and full campaigns



455. Explain about AWS SSM (Systems Manager) Session Manager.

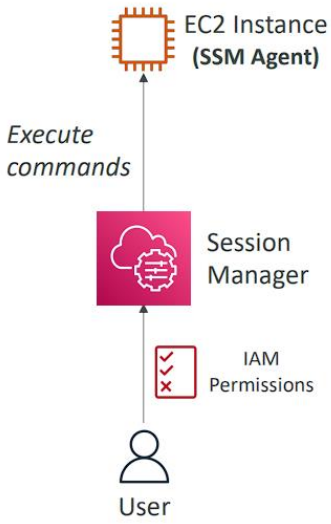
Ans ->

AWS SSM (Systems Manager) Session Manager is a tool that lets you securely access and manage EC2 instances without needing SSH or RDP. It provides a browser-based or AWS CLI interface for accessing instances, enhancing security by eliminating the need for open inbound ports or managing SSH keys. Session Manager allows you to run commands, troubleshoot, and manage your instances directly from AWS, with all activity logged for auditing purposes.

Key points:

- Allows you to start a secure shell on your EC2 and on-premises servers
- No SSH access, bastion hosts, or SSH keys needed
- No port 22 needed (better security)
- Supports Linux, macOS, and Windows

- Send session log data to S3 or CloudWatch Logs



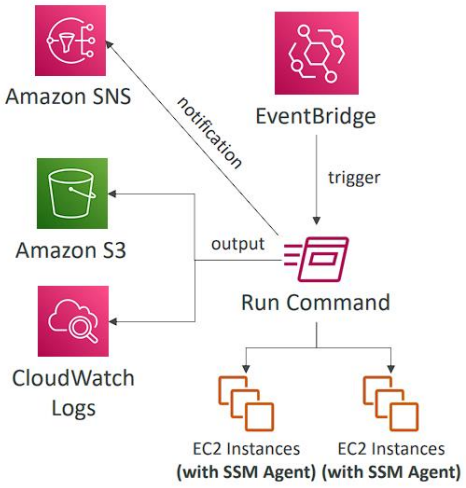
456. Explain about Systems Manager - Run Command.

Ans ->

AWS Systems Manager Run Command is a feature that allows you to remotely execute commands on your EC2 instances or on-premises servers without needing to log in to them. You can use it to automate administrative tasks, install software, run scripts, or apply patches across multiple instances at once. Run Command enhances security by allowing command execution without opening inbound ports or using SSH, and it logs all actions for auditing and compliance.

Key points:

- Execute a document (=script) or just run a command
- Run command across multiple instances (using resource groups)
- No need for SSH
- Command Output can be shown in the AWS Console, sent to S3 bucket or CloudWatch Logs
- Send notifications to SNS about command status (In progress, Success, Failed, ...)
- integrated with IAM & CloudTrail
- Can be invoked using EventBridge



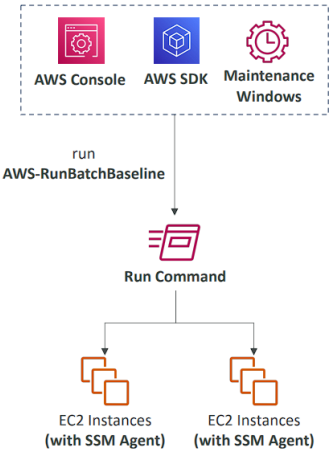
457. Explain about Systems Manager - Patch Manager.

Ans ->

AWS Systems Manager Patch Manager is a tool that automates the process of patching your EC2 instances or on-premises servers with security updates and bug fixes. It helps ensure your systems are up-to-date by scheduling and applying patches for operating systems and applications. Patch Manager supports patch baselines to define which patches to install and when, reducing the risk of vulnerabilities while allowing you to maintain control over the patching process. It also provides reports on patch compliance for auditing purposes.

Kye points:

- Automates the process of patching managed instances
- OS updates, applications updates, security updates
- Supports EC2 instances and on-premises servers
- Supports Linux, macOS, and Windows
- Patch on-demand or on a schedule using Maintenance Windows
- Scan instances and generate patch compliance report (missing patches)



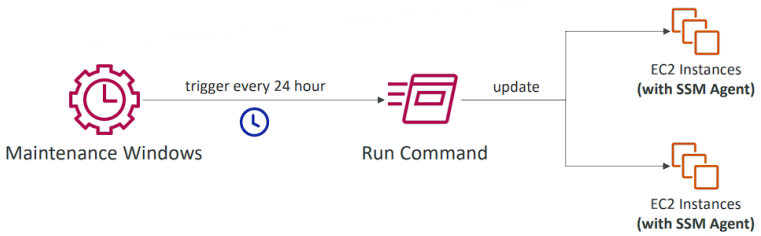
458. Explain about System Manager - Maintenance Windows.

Ans ->

AWS Systems Manager Maintenance Windows allow you to schedule recurring maintenance tasks, such as patching, updates, and scripts, for your EC2 instances or other AWS resources. You can define time windows when these tasks should run, helping to minimize disruption to operations. It ensures that maintenance activities happen in a controlled manner, coordinating tasks across multiple systems and reducing the risk of unplanned downtime. Maintenance Windows also offer flexibility in setting task priorities and deadlines, ensuring efficient management of routine operations.

Key points:

- Defines a schedule for when to perform actions on your instances
- **Example:** OS patching, updating drivers, installing software, ...
- **Maintenance Window contains:**
  - Schedule
  - Duration
  - Set of registered instances
  - Set of registered tasks



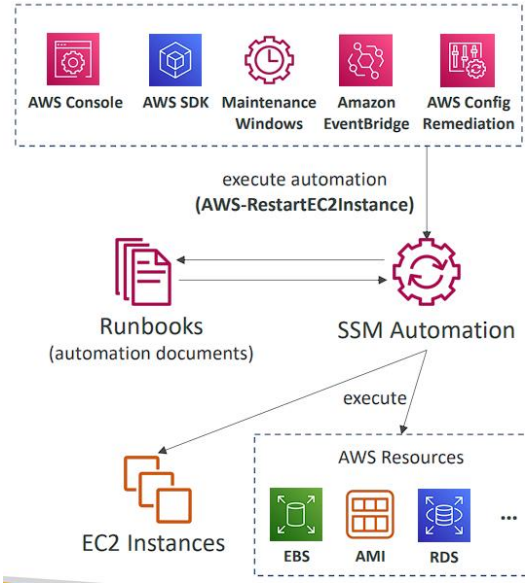
459. Explain about Systems Manager - Automation.

Ans ->

AWS Systems Manager Automation automates common tasks like starting/stopping EC2 instances, patching, and backups. It uses runbooks to streamline repetitive tasks, ensuring consistency and reducing manual effort. It also integrates with other AWS services and logs actions for auditing.

Key points:

- Simplifies common maintenance and deployment tasks of EC2 instances and other deployment tasks of EC2 instances and other AWS resources
- **Examples:** restart instances, create an AMI, EBS snapshot
- Automation Runbook - SSM Documents to define actions performed on your EC2 instances or AWS resources (pre-defined or custom)
- **Can be triggered using:**
  - Manually using AWS Console, AWS CLI or SDK
  - Amazon EventBridge
  - On a schedule using Maintenance Windows
  - By AWS Config or rules remediations



460. What is AWS Cost Explorer?

Ans ->

AWS Cost Explorer is a tool that helps you visualize, understand, and manage your AWS spending. It provides detailed insights into your usage patterns, costs, and forecasts, allowing you to track and analyze expenses across AWS services. With Cost Explorer, you can set custom time ranges, view cost trends, and identify areas where you can optimize and save on your AWS bills.

Key points:

- Visualize, understand, and manage your AWS costs and usage over time Create custom reports that analyze cost and usage data.
- Analyze your data at a high level: total costs and usage across all accounts
- Or monthly, hourly, resource level granularity
- Choose an optimal Savings Plan (to lower prices on your bill)
- Forecast usage up to 12 months based on previous usage

461. Explain about AWS Cost Anomaly Detection.

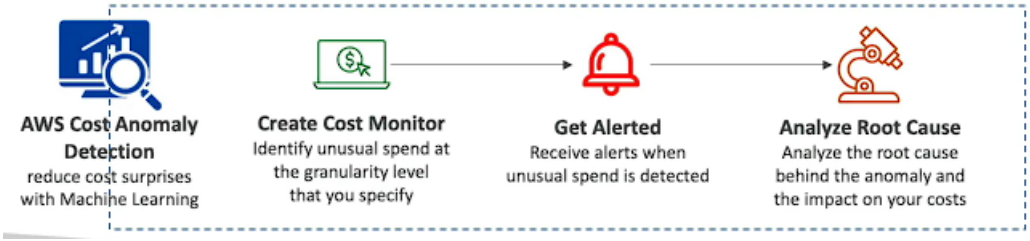
Ans ->

AWS Cost Anomaly Detection is a service that helps identify unexpected or unusual spending patterns in your AWS usage. It uses machine learning to automatically detect anomalies in your cost and usage data, alerting you when there's a spike or drop outside your normal spending trends. This helps you quickly identify potential issues,

like misconfigured resources or unexpected usage, and take action to prevent cost overruns.

Key points:

- Continuously monitor your cost and usage using ML to detect unusual spends
- It learns your unique, historic spend patterns to detect one-time cost spike and/or continuous cost increases (you don't need to define thresholds)
- Monitor AWS services, member accounts, cost allocation tags, or cost categories
- Sends you the anomaly detection report with root-cause analysis
- Get notified with individual alerts or daily/weekly summary (using SNS)



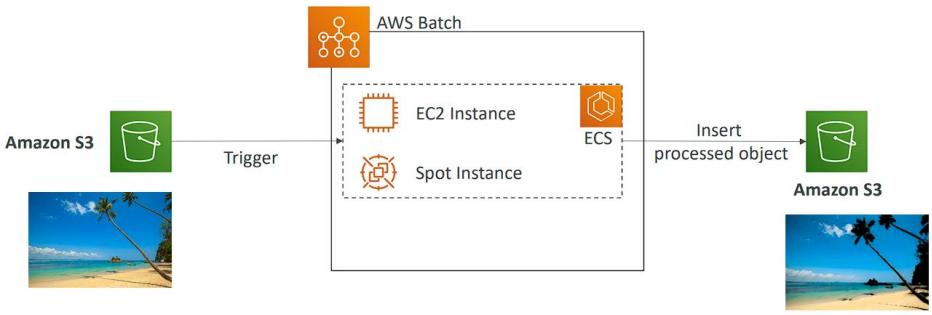
462. What is AWS Batch?

Ans ->

AWS Batch is a service that enables you to run batch computing jobs at any scale. It automatically provisions the necessary compute resources, such as EC2 instances or Spot Instances, based on the volume and requirements of your jobs. AWS Batch manages job scheduling, execution, and scaling, allowing you to focus on processing workloads without worrying about infrastructure setup. It's ideal for use cases like data processing, simulations, and large-scale computations.

Key points:

- Fully managed batch processing at any scale
- Efficiently run 100,000s of computing batch jobs on AWS
- A "batch" job is a job with a start and an end (opposed to continuous)
- Batch will dynamically launch EC2 instances or Spot Instances
- AWS Batch provisions the right amount of compute / memory
- You submit or schedule batch jobs and AWS Batch does the rest!
- Batch jobs are defined as Docker images and run on ECS
- Helpful for cost optimizations and focusing less on the infrastructure



463. AWS Batch VS Lambda.

Ans ->

Lambda:

- Time limit (max 15 mins)
- Limited runtimes
- Limited temporary disk space
- Serverless

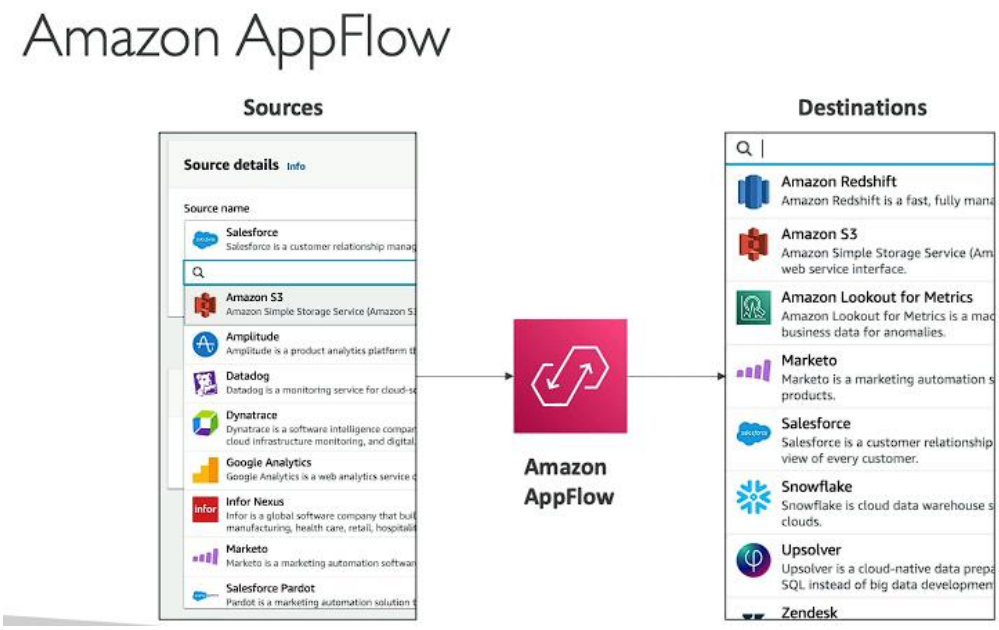
Batch:

- No time limit
- Any runtime as long as it's packaged as a Docker image
- Rely on EBS / instance store for disk space
- Relies on EC2 (can be managed by AWS)

464. What is Amazon AppFlow?

Ans ->

- Fully managed integration service that enables you to securely transfer data between Software-as-a-Service (SaaS) applications and AWS
- **Sources:** Salesforce, SAP, Zendesk, Slack, and ServiceNow
- **Destinations:** AWS services like Amazon S3, Amazon Redshift or non-AWS such as SnowFlake and Salesforce
- **Frequency:** on a schedule, in response to events, or on demand
- Data transformation capabilities like filtering and validation
- **Encrypted** over the public internet or privately over AWS PrivateLink
- Don't spend time writing the integrations and leverage APIs immediately



465. Explain about AWS Amplify.

Ans ->

AWS Amplify is a development platform that helps you build, deploy, and host full-stack web and mobile applications quickly. It provides tools and services for frontend and backend development, including features like authentication, APIs, storage, and hosting. Amplify streamlines the process of building apps by automating deployment and scaling, making it easy for developers to focus on writing code while managing infrastructure and updates. It's commonly used for developing modern, serverless applications.

Key points:

- A set of tools and services that helps you develop and deploy scalable full stack we and mobile applications
- Authentication, Storage, API (REST, GraphQL), CI/CD, PubSub, Analytics, AI/ML Predictions, Monitoring...
- Connect your source code from GitHub, AWS CodeCommit, Bitbucket, GitLab, or upload directly