



KPLABS Course

AWS Certified Solutions Architect - Professional

Exam Preparation

ISSUED BY

Zeal

REPRESENTATIVE

instructors@kplabs.in

Module 1: Delegation

You should know how to architect the cross-account IAM role-based architecture.

- Create a cross-account IAM role in Account B (with Account A as trusted entities).
- Create a user in Account A, which wants to assume the role.
- Assign the AssumeRole policy with RoleArn of Account B role to a user in account A.
- Share the Sign-In link with the user.

Module 2: AWS Organization

Consolidated billing can be enabled via AWS organizations.

AWS Organizations allow us to set “Service Control Policies” to control access to the linked accounts.

For use-cases where minimal blast radius is required, a separate AWS account is a good practice with AWS Organization OU.

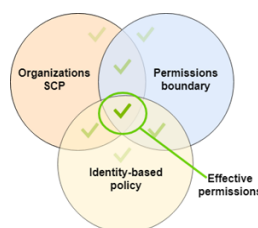
Two available feature sets: Consolidated Billing Features and All Features (SCP + Billing)

To “Allow Access”, you can make use of IAM Policies and not SCP Policies.

Module 3: Effective Permissions in AWS

The effective permissions for an entity are the permissions that are granted by all the policies associated with the user/role/account.

If anywhere there is an explicit deny, the final action will be Deny.



Module 4: Use-Cases IAM

Alice has created a Lambda function that monitors for CPU Utilization of EC2 instance. If CPU utilization is less than 10%, the Lambda function will stop the instance.

IAM Policy attached to Alice:

- CloudWatch Full Access
- Lambda Full Access
- EC2 ReadOnly Access.

Final Decision: Alice can setup Lambda, but it will not be able to stop the EC2.

Module 5: Types of Managed Policies

“PowerUser” managed policy by AWS provides full access to AWS services and resources, but does not allow management of Users and groups.

For Developers, instead of assigning Administrator access, PowerUser access should be given.

If you want to provide access to the developers to all the relevant services but only in one region, then you can approach it with the following scenario:

Attach PowerUser Access

Add an Explicit Deny policy that denies access to resources in all the regions except the approved one.

Module 6: Service Control Policy vs IAM Policy

IAM policies are attached to the IAM User/Group/Roles.

Root account does not have an IAM policy (has full access)

The specified actions from an attached SCP affect all IAM users, groups, and roles for an account, including the root account identity.

SCP's are not used to allow an IAM user to perform certain operations explicitly. Use IAM Policies instead.

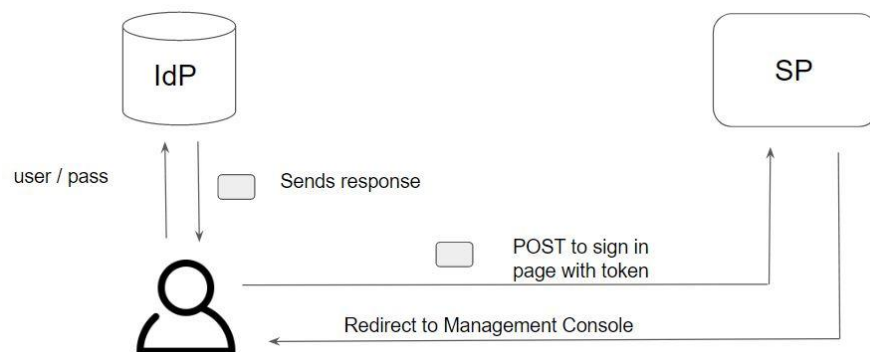
Module 7: Active Directory & Directory Service

Types	Description
AWS Managed Microsoft AD	Full-Fledged Microsoft Active-Directory. Standard Edition: For small and mid-size organizations (up to 5000 users) Enterprise Edition for Larger Deployments.
AD Connector	Used when on-premise users needs to access AWS service via AD.
Simple AD	Powered by Samba 4 Active Directory. Does not support features like trust relationships, MFA, LDAPS, FSMO role transfer. Small - Supports upto 500 users. Large - Supports upto 5000 users.

Module 8: SAML

Security Assertion Markup Language (SAML) is an open standard for exchanging authentication and authorization data between parties, in particular, between an identity provider and a service provider.

They are used primarily during Single Sign-On.



Module 9: Active Directory Federation Service (ADFS)

Active Directory Federation Services (ADFS) is a Single Sign-On (SSO) solution created by Microsoft.

Supports SAML for authentication.

Active Directory groups are associated with IAM roles.

All the users in the AD group can assume the appropriate IAM role.



Module 10: Amazon Cognito

Amazon Cognito provides authentication, authorization, and user management service for your web and mobile apps.

Supports authentication from social identity providers like Google, Amazon, Facebook.



Module 11: AWS Service Catalog

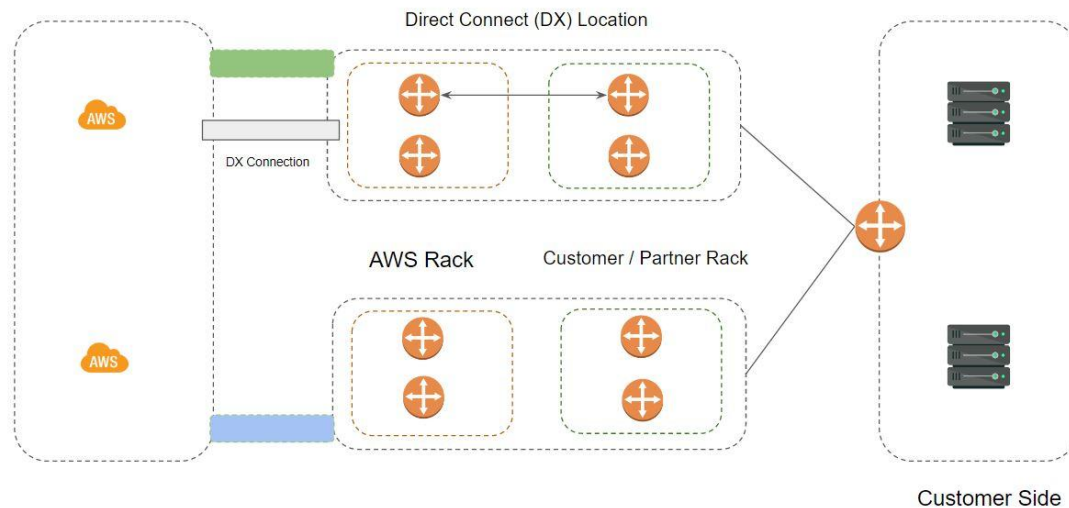
AWS Service Catalog enables organizations to create and manage catalogs of IT services that are approved for use on AWS.

Primarily used when you do not want users to create random resources and only use a standardized set of resources that are approved.

Service Catalog portfolio can also be shared across other AWS accounts.

Module 12: Direct Connect

You should know the process to set up the Direct Connect connection.



You should know about the VIF's (both public and private ones)

- i) Public VIF: Used to access public endpoints within the region [S3].
- ii) Private VIF: Used to access private endpoints like VPC.

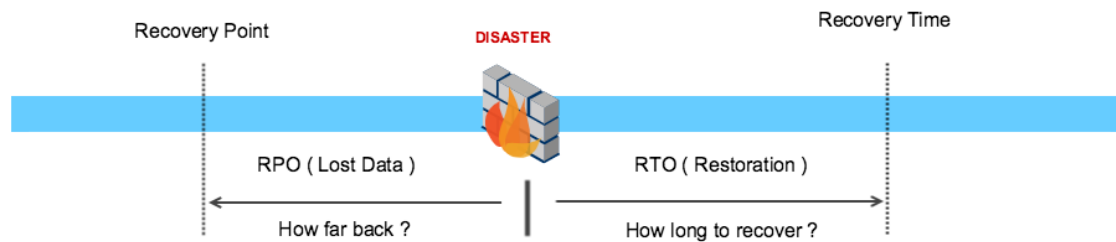
If use-case specifies high-bandwidth and highly available architectures, then two DX connections are required. No VPN in such cases.

If cost is the factor and a single point of failure is a concern for DX, then VPN can be introduced as a backup.

Module 13: RTO vs RPO

Recovery Time Objective (RTO) is the amount of time frame it takes for you to recover your infrastructure and business operations after the disaster has struck.

Recovery Point Objective (RPO) is concerned with data and the maximum tolerance period to which data can be lost.



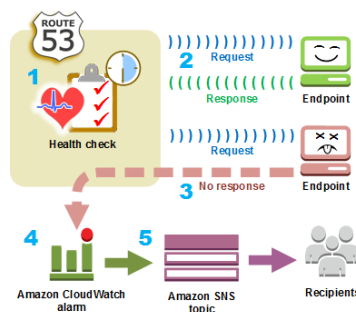
Module 14: Route53 Health Checks

Amazon Route 53 health checks monitor the health of your resources, such as web servers.

How Route53 determines if an endpoint is healthy?

- The endpoint must return an HTTP 2xx or 3xx status code.
- It must receive the response body from the endpoint
- Route 53 searches the response body for a string that you specify.

The string must appear entirely in the first 5,120 bytes of the response body, or the endpoint fails the health check.



Module 15: Route53 Routing Policies

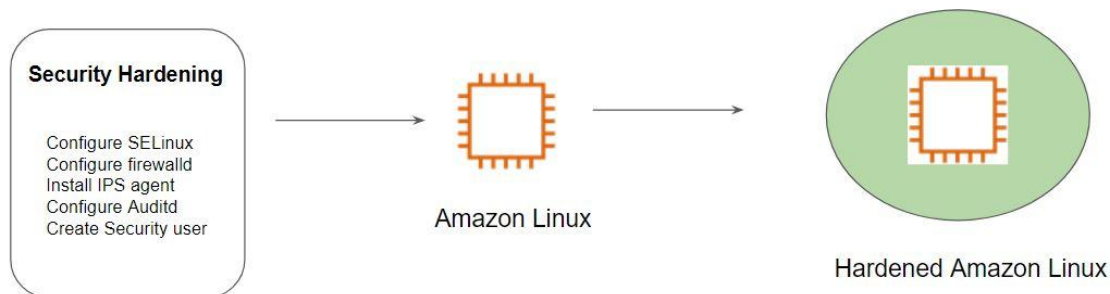
Route53 can also be used to host private hosted zones that are shared among multiple VPC's across AWS accounts.

Routing Policies	Description
Latency Based Routing	Use when you have resources in multiple AWS Regions, and you want to route traffic to the region that provides the best latency.
Geolocation routing policy	Use when you want to route traffic based on the location of your users.
Failover routing policy	Use when you want to configure active-passive failover.
Weighted routing policy	Use to route traffic to multiple resources in proportions that you specify.

Module 17: Amazon Machine Image (AMI)

Amazon Machine Image (AMI) is the master image from which new EC2 instances can be launched.

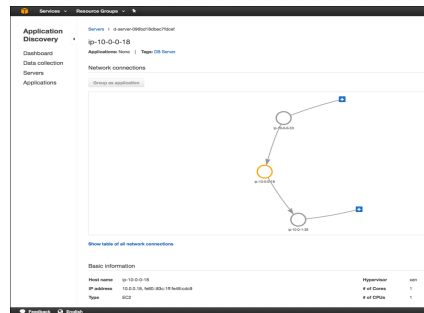
AMI is a better approach for use-cases where installing and configuring data takes a lot of time in the EC2 instances and needs to be repeated for each new EC2.



Module 18: Application Discovery Service

AWS Application Discovery Service helps enterprise customers plan migration projects by gathering information about their on-premises data centers.

Can support agent as well as the agentless architectures.

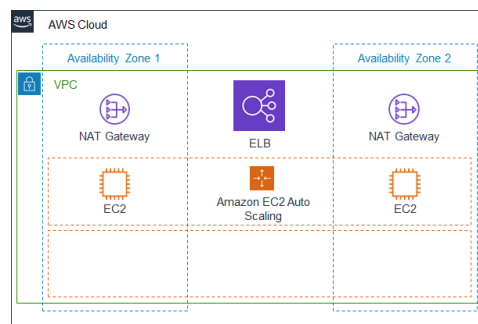


Module 19: Auto Scaling

You should know basic auto-scaling concepts like Launch Configuration, Auto-Scaling Groups.

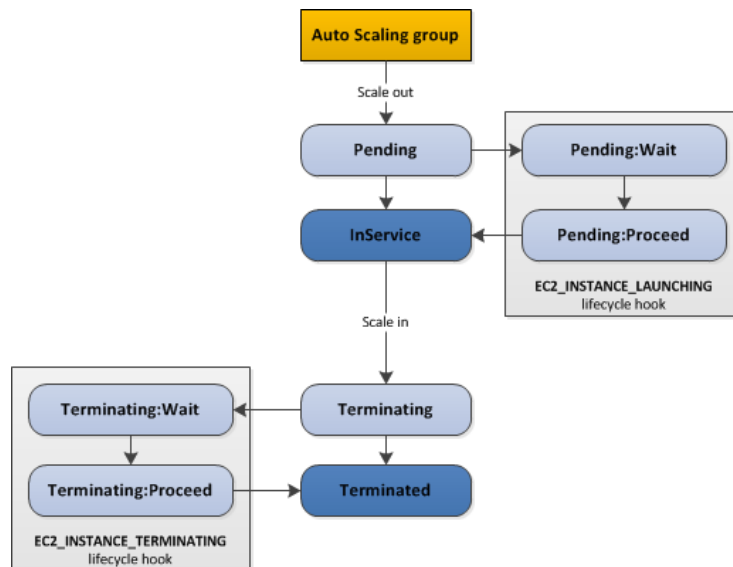
- Launch Configuration: Specify AMI ID, Key-Pair, Security Groups
- Auto-Scaling Groups: How many instances you want to run.

You cannot modify the launch config after it was created. For any updates, create a new launch config.



Module 20: Auto Scaling LifeCycle Hooks

Auto-Scaling LifeCycle hook allows us to perform a set of custom actions when an instance launches or terminates.



Module 21: Kinesis

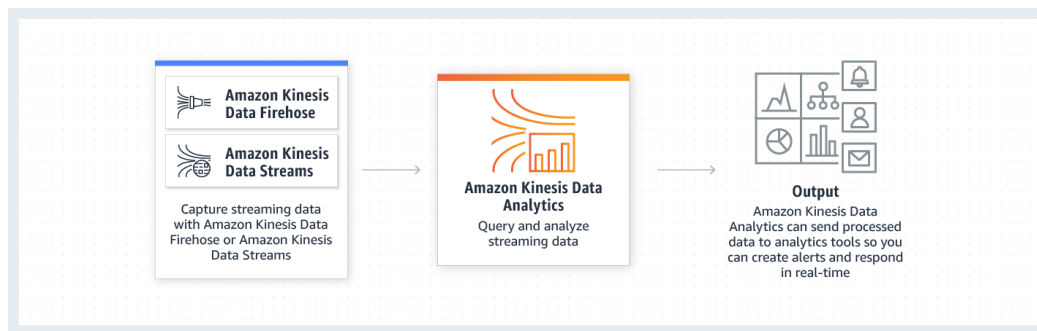
You should know different types of Kinesis streams and in which use-case a specific stream will be useful.

Sr No	Types of Kinesis	Description
1	Kinesis Data Stream	Captures, processes, and stores data streams in real-time
2	Kinesis Data Firehose	Allows to capture and deliver data to data stores in real-time. Primary to move data from point A to point B.
3	Kinesis Data Analytics	Analyze streaming data in real-time with SQL / Java code.
4	Kinesis Video Stream	Capture, processes and stores video streams.

Module 22: Kinesis Data Analytics

Amazon Kinesis Data Analytics is the easiest way to analyze streaming data, gain actionable insights.

For use-cases, where you want to capture anomalies in DynamoDB, you can make use of DynamoDB streams to send updates to Lambda and Lambda to store it in Kinesis Data Stream. Kinesis Data Analytics can then be used to analyze

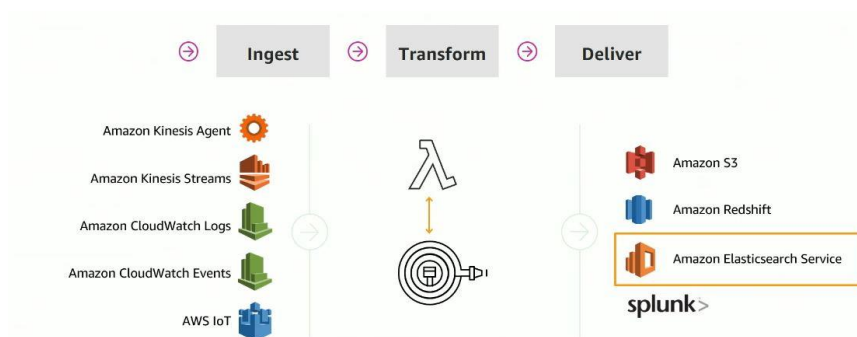


Module 23: Kinesis Data Firehose

Kinesis Firehose is primarily about delivery from point A to point B.

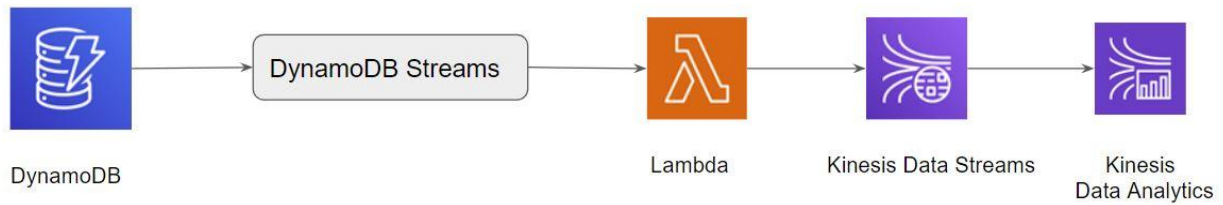
Use-cases where you want to ingest data to S3, instead of making use of EC2 instances, Kinesis Data Firehose can be a better option.

For visualization related use-cases, data can be sent to Firehose and then to ES. Kibana can be used for visualization.



Module 23: Use-Case - DynamoDB Streams & Kinesis

Use-Case: Detecting certain anomalies based on DynamoDB data.



Module 24: Load Balancing

Know the use-case where each type of load balancer type can be used.

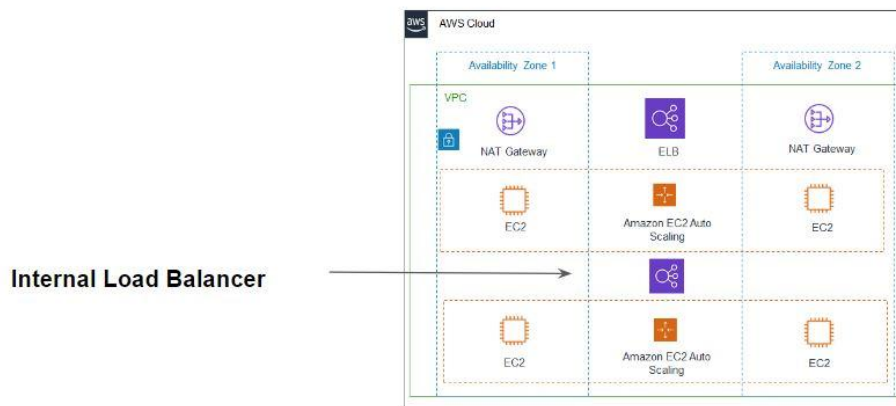
Load Balancer	Description
Classic Load Balancer (CLB)	For development / testing.
Application Load Balancer (ALB)	For Layer 7 Traffic.
Network Load Balancer (NLB)	Very fast performance. Can associate static IP address.

Module 25: Middle Tier Scaling Architectures - ELB

Allows separation of tiers.

Not publicly accessible.

Independent Scaling.

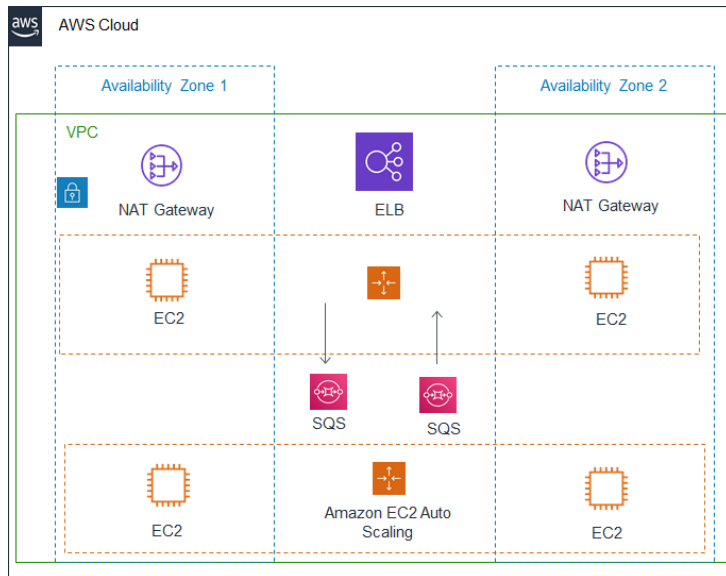


Module 25: Middle Tier Scaling Architectures - SQS

-
- Asynchronous Tasks.
- Unordered
- Single-direction only.
- “At-least once” delivery.

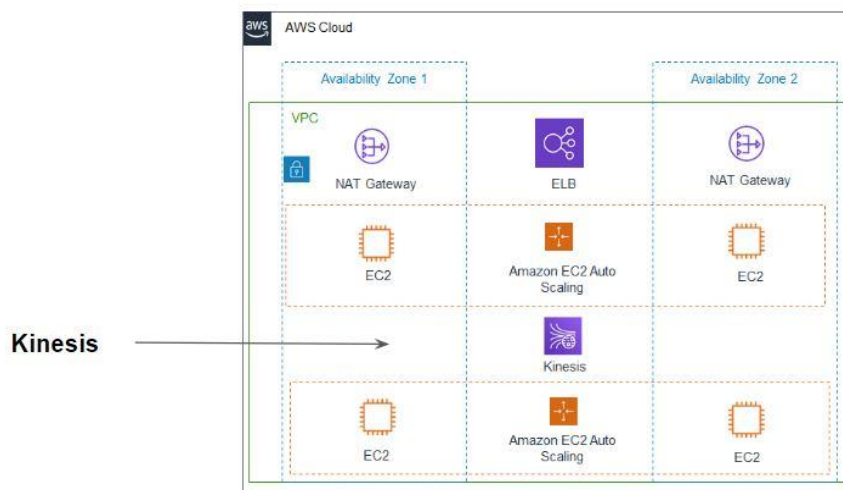


SQS Queue For Request-Response communication



Module 26: Middle Tier Scaling Architectures - Kinesis

- Ordered within a shard
- Single-direction only.



Module 27: EBS and Instance Store Volumes

EBS volumes:

- Persistent Storage.
- Size and Volume Types can be increased while being attached to EC2.

Instance Store:

- Size cannot be increased.
- It is not portable as is associated with the physical host.

Module 28: EBS - GP2 vs Provisioned IOPS



General Purpose SSD

Used for normal workloads which balances the price and performance.

Generally used for test, dev and prod environments under The normal workload.

Volume size from 1 GB to 16 TB

Referred by gp2

Provisioned IOPS SSD

Highest performance SSD volume designed for mission-critical application workloads.

Generally used for large database workloads like MongoDB, Cassandra, PostgreSQL, MySQL etc.

Volume size from 4GB to 16 TB

Referred by io1

Module 29: Storage Performance

EBS with PIOPS supports a maximum of 80,000 IOPS/instance while ephemeral storage can provide up to 1.6 million IOPS

Remember baseline of 3 IOPS per GB in GP2.

Pricing between PIOPS and GP2 plays a major role in the decision. In many situations, increasing the EBS size to allow more IOPS is more beneficial.

Example:

If there is a 1TB EBS volume with IOPS of 3000. You need 6000 IOPS.

- Extend EBS to 2TB to get 6000 IOPS = \$201
- EBS with PIOPS with 1TB and 6000 IOPS = \$518

Module 30: AWS S3

S3 Versioning allows users to keep multiple variants of an object in the same S3 bucket.

Cross-region replication allows an object to be replicated across S3 buckets between multiple regions.

Lifecycle policies allow you to automatically review objects within your S3 Buckets and have them moved to different S3 storage class or have the objects deleted from S3.

S3 Event Notifications can be used to notify when a specific type of event occurs. It can be integrated with Lambda for use-cases where Lambda needs to run every time a new file is uploaded to S3.

Storage Class:

At a high-level, be aware of S3 storage classes.

- Standard
- Intelligent-Tiering
- Standard-IA
- One Zone-IA
- Glacier
- Glacier Deep Archive
- Reduced Redundancy

Object Level Logging

CloudTrail supports logging Amazon S3 object-level API operations such as GetObject, DeleteObject, and PutObject. By default, this is not enabled. (Data events)

You can enable object-level logging for the S3 bucket for both the read and write events.

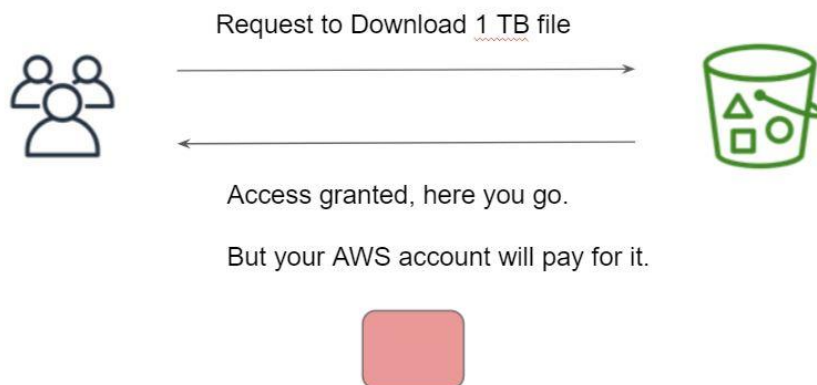
To detect and remediate objects which have been uploaded as a public object, the following steps can be used:

Enable Object Level logging.

CloudWatch Event to notify when a PutObject API call with public-read permission is detected.
Lambda function for remediation.

Requester Pays

With Requester Pays buckets, the requester instead of the bucket owner pays the cost of the request and the data download from the bucket. The bucket owner always pays the cost of storing data.



S3 Encryption

There are three ways in which we can Encrypt data in S3:-

- i) Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3)
- ii) Server-Side Encryption with AWS KMS-Managed Keys (SSE-KMS)
- iii) Server-Side Encryption with Customer-Provided Keys (SSE-C)

S3 Encryption Use-Case

For use-cases where you want the encryption keys to be highly-available and also requirement where control of keys on per-user-basis is required, the following options can be used:

AWS SSE with KMS

AWS SSE with Customer Managed Keys. These keys can be stored in CloudHSM.

For use-cases where data must be encrypted at transit (HTTPS), you can make use of bucket policy to enforce HTTPS connections only.

Module 31: AWS Certificate Manager

Important Pointer: Certificates are region-specific.

Example Scenario:

Medium Corp is deploying example.com in two regions (us-east-1 and ap-southeast-1) using latency based routing. They need SSL/TLS with ELB. How many certificates are required?

- Request 1 ACM Certificate for North Virginia Region.
- Request 1 ACM Certificate for Singapore Region.

Module 32: CloudFormation

CloudFormation StackSets allows us to deploy stacks across multiple AWS account / AWS regions from a single location.

The deletion policy attribute allows us to preserve or (in some case) backup a resource when it's stack is deleted.

There are two options when we use the deletion policy attribute.

- i) Retain: CloudFormation keeps the resource without deleting it.
- ii) Snapshot: CloudFormation creates a snapshot of the resource before deleting it.

Module 33: AppStream

AppStream 2.0 allows us to centrally manage our desktop application and securely deliver them to any computer.

Key Word: Desktop Application



Module 34: DynamoDB

Understand the distinction on when to use DynamoDB or when to use RDS.

If you need a relational database with ACID transactions → Use RDS

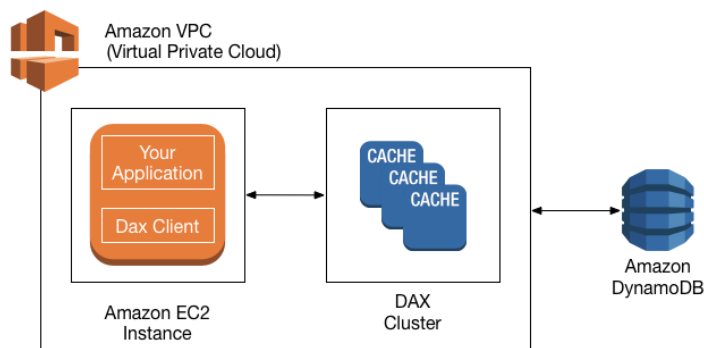
If you have unstructured data → Use DynamoDB

DynamoDB TTL allows items from tables to be automatically be deleted after a certain duration.

DynamoDB Accelerator

DynamoDB Accelerator (DAX) is a fully managed, clustered, in-memory cache for DynamoDB. It delivers up to 10x read performance improvement.

Suitable when you need a response time in microseconds and millions of requests per second.

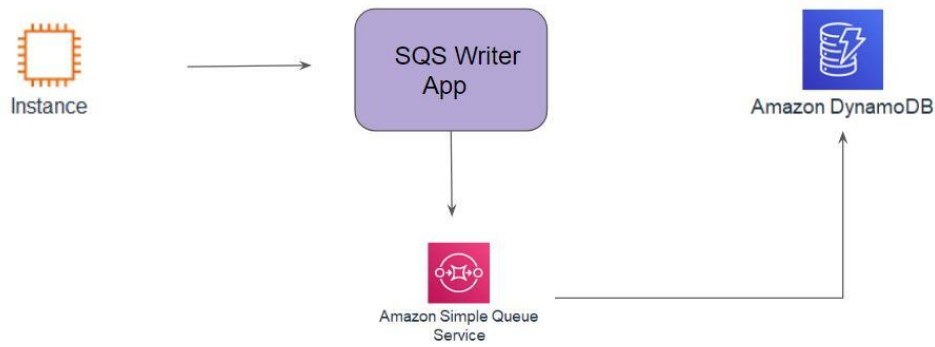


DynamoDB Cost Consideration

DynamoDB Throughput can lead to a huge amount of overall cost (RCU and WCU).

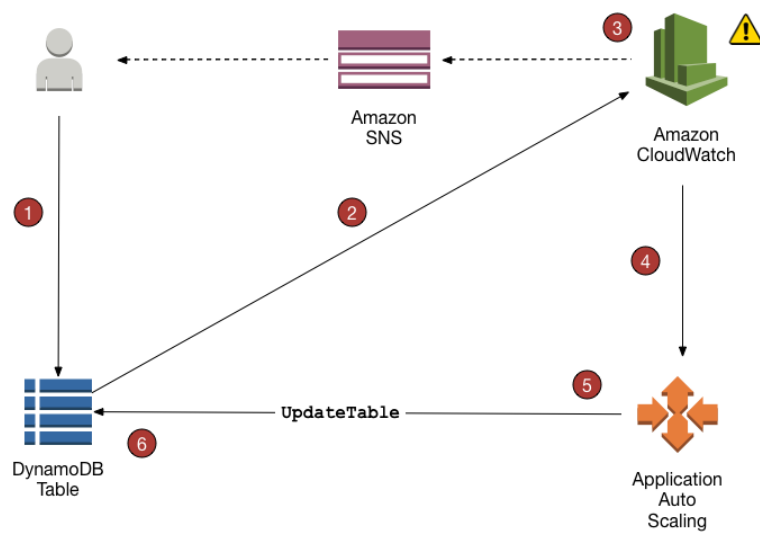
We can make use of SQS to deal with burst workloads with limited RCU and WCU

SQS approach is good if asynchronous operations is fine involves latency



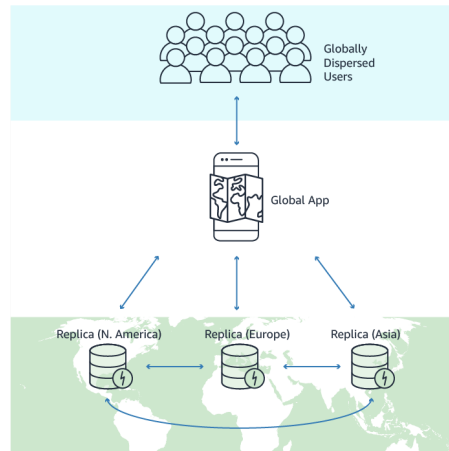
DynamoDB Auto Scaling:

DynamoDB Auto-Scaling allows us to scale up and scale down the throughput dynamically.



DynamoDB Global Tables

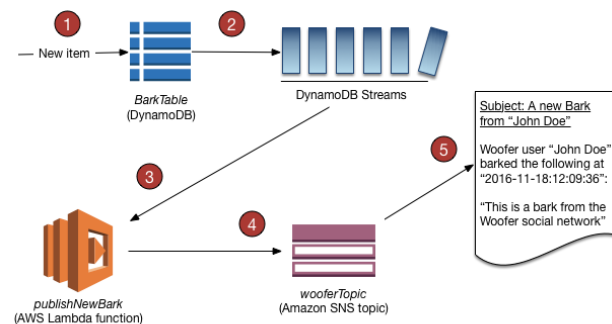
Global Tables fully automates the replication of your Amazon DynamoDB tables across your choice of AWS regions.



DynamoDB Streams:

DynamoDB Streams provides a time-ordered sequence of item-level changes in the DynamoDB table.

This allows us to achieve various use-cases related to continuous analytics, real-time notifications, and various others.



Miscellaneous Databases:

- Cassandra is a NoSQL database management system.
- You can consider to move it to DynamoDB as part of the migration.

Module 35: Relational Database Service

Supported Engines:

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

IBM DB2 family is not supported. For migration, it needs to be hosted in EC2.

For Amazon Aurora, the storage will automatically grow, up to 64 TB, in 10GB increments with no impact on database performance. There is no need to provision storage in advance.

Amazon RDS Read Replicas enable you to create one or more read-only copies of your database instance within the same AWS Region or in a different AWS Region. We can also promote the read-replica as the new master whenever required.

RDS Event Notification:

RDS Event Notification allows customers to get a notification by email, text message, or a call to an HTTP endpoint when an Amazon RDS event occurs using Amazon SNS.

Various categories of events are reported, some of these include:

Availability, Failure, Maintenance, Recovery, Restoration

The screenshot shows the 'Source type' dropdown set to 'Instances'. Below it, the 'Instances to include' section contains a search bar and a list of event categories. The categories listed are: configuration change, creation, deletion, failover, failure, low storage, maintenance, notification, read replica, recovery, and select event categories. The 'notification' category is currently selected and highlighted in blue.

Source type
Source type of resource this subscription will consume event from
Instances
Instances to include
Instances that this subscription will consume events from
configuration change
creation
deletion
failover
failure
low storage
maintenance
notification
read replica
recovery
select event categories

Use-Case Specific Workloads

Type	Description
RDS	<ul style="list-style-type: none">• Relational Database• ACID Transactions• OLTP
DynamoDB	<ul style="list-style-type: none">• NoSQL Database.• Unstructured data.• High I/O needs.
EC2	<ul style="list-style-type: none">• When customers need complete control.• Databases which are not supported by RDS.
S3	<ul style="list-style-type: none">• Data Blobs.
RedShift	<ul style="list-style-type: none">• Data Warehouse• OLAP

Multi-AZ

- Useful for High-Availability Aspect.
- Standby instance cannot be accessed.
- Automated failover taken care by AWS.
- DNS Names do not change on failover.
- Based on synchronous replication.

Read replicas:

- Useful for Scalability Aspect.
- Can be used for DR with Read Replica across the different region.
- Based on asynchronous replication.

Module 36: CodeX Solutions

Be aware of Code Commit, Code Build, Code Deploy, Code Pipeline.

Code Commit	Fully-managed source control service that hosts secure Git-based repositories.
Code Build	Build and test code with continuous scaling
Code Deploy	Deployment service that automates application deployments to Amazon EC2 instances, on-premises instances, serverless Lambda functions, or Amazon ECS services
Code Pipeline	CodePipeline, a continuous integration and release automation service for applications you want to release in the cloud.

Module 37: AWS OpsWorks

AWS OpsWorks manages infrastructure deployment for cloud administrators.

AWS OpsWorks is a global service, but we can only manage resources in the region where the OpsWorks stack is created.

Remember, when you get Chef, Puppet, AWS OpsWorks might be the right answer.

Module 38: Simple Notification Service

It integrates with various AWS services like CloudWatch for alarm functionality.

SNS cannot provide data every minute; it works at a 5-minute interval.

Module 39: AWS Config

AWS Config is a service that enables you to continually assess, audit, and evaluate the configurations of your AWS resources.

If the question states auditing, monitoring, then services like AWS Config, CloudTrail are good.

If there is a use-case where only pre-approved AMI should be used, AWS Config + Lambda can be used to audit and remediate.

Module 40: SQS

Category	Standard Queue	FIFO Queue
Message Order	Provide best-effort ordering which ensures that messages are generally delivered in the same order as they are sent.	FIFO queues offer first-in-first-out delivery
Message Delivery	Standard queues guarantee that a message is delivered at least once and duplicates can be introduced into the queue	Duplicates are not introduced into the queue
Transaction per Second	Allow a nearly-unlimited number of transactions per second	FIFO queues are limited to 300 transactions per second per API action

Module 41: VPC Endpoints

VPC Endpoint policies can be configured to allow access to required S3 buckets only.

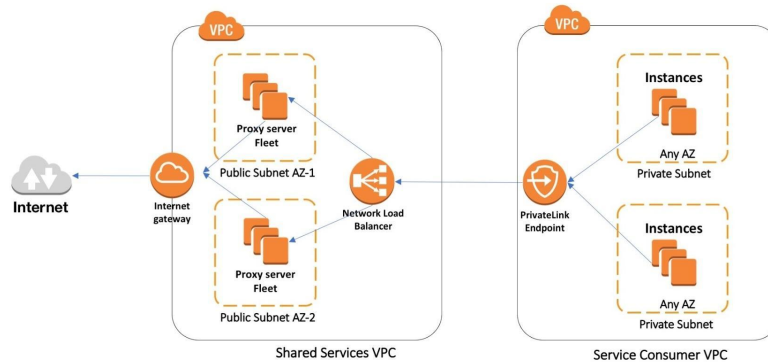
If you want only EC2 instance from specific VPC to connect to S3, then S3 VPC Endpoint can be used along with the S3 bucket policy to limit the access to the created endpoint.

```
"Statement": [{
  "Sid": "VPCE and SourceIP",
  "Effect": "Deny",
  "Principal": "*",
  "Action": "s3:*",
  "Resource": [
    "arn:aws:s3::awsexamplebucket",
    "arn:aws:s3::awsexamplebucket/*"
  ],
  "Condition": {
    "StringNotLike": {
      "aws:sourceVpce": [
        "vpce-1111111",
        "vpce-2222222"
      ]
    }
  }
}]
```

Module 42: Centralized Architecture for Internet Traffic Routing

To monitor and control the web traffic, AWS PrivateLink based architecture can be used.

This simplified architecture describes how we can use AWS PrivateLink to allow instances in one VPC (consumer VPC) to privately use the web proxy fleet in another VPC (shared services VPC).



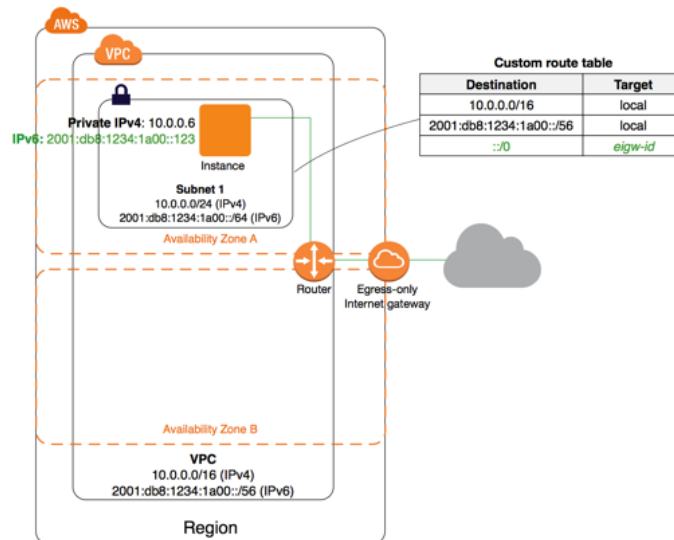
Module 43: NAT Gateways

You cannot route traffic to a NAT gateway through a VPC peering connection, a Site-to-Site VPN connection, or AWS Direct Connect. A NAT gateway cannot be used by resources on the other side of these connections.

It is recommended to replace NAT instances with NAT Gateways.

Module 44: Egress-Only Internet Gateway

Egress-Only Gateway allows EC2 instance with IPv6 address to access the internet directly but prevent the resource from the internet to directly initiate a new connection with the EC2 instance.

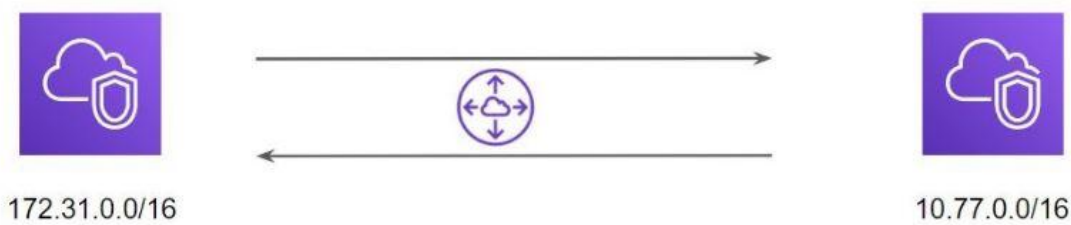


Module 45: VPC Peering

VPC peering is a network connection between two VPC that enables the communication between instances of both the VPC.

Maximum VPC Peering Limit = 125

Other approaches like PrivateLink endpoint with NLB can be used if there are a lot of accounts and services that need to communicate with each other across accounts.



Module 46: Serverless

If a question has a key-word of “serverless” for its use-case, make sure to select serverless services from the solution.

- AWS Lambda, SNS, SQS, Kinesis
- Lambda@Edge, Athena, Step Function
- Fargate, S3
- DynamoDB, API Gateway
- AWS EFS

Avoid services like EC2.

Module 47: Elastic Beanstalk

Elastic Beanstalk can be used if the team wants to reduce time in capacity-management and maintenance of infrastructure

Supported Platforms:

Java, .NET, PHP, Node.js, Python, Ruby, Go, Docker

Blue-Green deployments can be used with Elastic Beanstalk for the least disruptions.

Once an application is deployed to a new environment, make use of Swap Environment URLs

EB with Blue/Green is a good choice when developers want to deploy a new version to production quickly and with the least amount of disruption of service.

Module 48: EC2 Tenancy Attribute

Every EC2 instance that we launch in the VPC has a specific tenancy attribute associated with it. There are three primary tenancy attributes which are available:

Tenancy Attribute	Description
Shared	The EC2 instance runs on shared hardware.
Dedicated	EC2 instance runs on hardware which will only be shared between same account AWS instances.
Hosts	Instance runs on dedicated hosts with very granular level of hardware access.

Module 49: EC2 Auto-Recovery

You can create an Amazon CloudWatch alarm that monitors an Amazon EC2 instance and automatically recovers the instance if it becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair.

A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata.

When the StatusCheckFailed_System alarm is triggered, and the recover action is initiated.

Module 50: EC2 Pricing Models

Know when to use On-Demand, Spot, Reserved Instances

When there is mention of steady workloads, Reserved Instances are the right choice.

If you need a workload for a few hours, on-demand is the right choice.



If the workload is needed and if data is part of the SQS queue and you need a huge workload for small-time and abruptions are fine, spot instances can be the right choice.

Module 51: Scheduled Reserved Instance

Scheduled Reserved Instances (Scheduled Instances) enable you to purchase capacity reservations that recur on a daily, weekly, or monthly basis, with a specified start time and duration, for a one-year term.

The minimum required utilization is 1,200 hours per year.

Create a schedule

Starting on	<input type="text"/>		for duration	<input type="text" value="4"/>		hours
	<input type="checkbox"/> +/- 2 hours					
Recurring	<div>Daily ▼</div>					

Module 52: AWS WAF and AWS Shield

AWS WAF can protect against various Layer 7 attacks like SQL Injection, IP Address Whitelisting / Blacklisting, Blocking bad user-agent bots, and others.

AWS Shield and AWS Shield Advanced can be used for DDoS protection.

Migrating DNS to Route53 is also a good option if DDoS is in the consideration.

Module 53: ECS

Multiple applications can be deployed to an ECS cluster under an ALB. This can save costs.

Fargate can be used if you are unsure about the requirements related to the amount of EC2 instances that need to be configured for the jobs.

Security Group can be applied to the tasks, and we can make use of a task role in task definitions. The tasks can use this IAM role for AWS API calls.

Networking Mode	Description
None	The task's containers do not have external connectivity, and port mappings can't be specified in the container definition.
Bridge	The task utilizes Docker's built-in virtual network, which runs inside each container instance.
Host	The task bypasses Docker's built-in virtual network and maps container ports directly to the EC2 instance's network interface directly.
<u>awsvpc</u>	The task is allocated an elastic network interface.

Module 54: RedShift

RedShift only supports a single availability zone (AZ) deployments.

For Disaster Recovery scenarios, Best practice is to ensure that CloudFormation is used for the creation of RedShift cluster and automatic backups to S3 is enabled.

RedShift should be avoided as a data-storage platform, prefer S3 instead.

Module 55: Cost Allocation Tags

It is recommended to assign a tag to AWS resources.

Tagging helps in a variety of aspects including cost allocation, access control, automation.

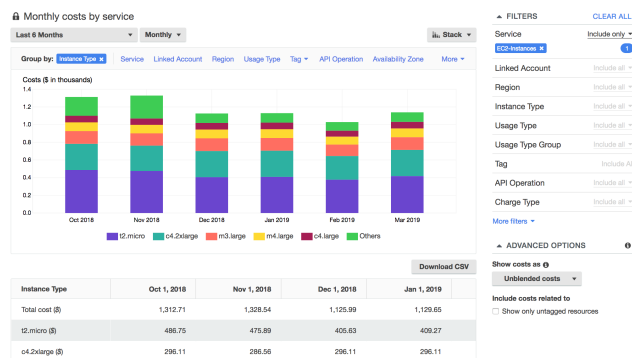
For use-cases where visibility is required into each team's spending in the organization, it is important to make use of tags for billing purposes.

Total Cost ▾	user:Owner ▾	user:Stack ▾	user:Cost Center ▾	user:Application ▾
0.95	DbAdmin	Test	80432	Widget2
0.01	DbAdmin	Test	80432	Widget2
3.84	DbAdmin	Prod	80432	Widget2
6.00	DbAdmin	Test	78925	Widget1
234.63	SysEng	Prod	78925	Widget1
0.73	DbAdmin	Test	78925	Widget1
0.00	DbAdmin	Prod	80432	Portal
2.47	DbAdmin	Prod	78925	Portal

Module 56: AWS Cost Explorer

It allows customers to visualize cost over a period of time.

Provides out of box reporting and tracking associated with Reserved Instances and it benefits across your AWS accounts.



Module 57: Storage Migration

AWS Import/Export	<ul style="list-style-type: none">• Customer ships their external disks to AWS.• AWS team plugs in and transfers the data to S3.
AWS Snowball	<ul style="list-style-type: none">• Rugged NAS which AWS ships to customers.• Customers can copy up to 80 TB of data and ships back to AWS.• AWS Team copies the data to S3.• Can take more than a week for the entire process. <p>Note: If customer wants to quickly transfer data within few days, Snowball might not be the right option.</p> <p>Option of Snowball + Internet can be used to copy base data and sync the deltas.</p>
AWS Snowmobile	<ul style="list-style-type: none">• 45-foot long ruggedized shipping container, pulled by a semi-trailer truck.• It supports Exabyte-scale storage.

Module 58: AWS DMS

AWS Database Migration Service helps you migrate databases to AWS quickly and securely.

The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database.

It can work along with a schema conversion tool for both homogeneous (same database) migration as well as heterogeneous (different database types) migrations.

DMS also supports the AWS ElasticSearch service as a target.

Be sure to request a Snowball Edge device (Snowball Edge Storage Optimized), because regular Snowball devices are not supported for AWS DMS.

Module 58: AWS SCT

The AWS Schema Conversion Tool makes heterogeneous database migrations predictable by automatically converting the source database schema and a majority of the database code objects, including views, stored procedures, and functions, to a format compatible with the target database.

SCT can also scan your application source code for embedded SQL statements and convert them as part of a database schema conversion project.

Module 59: AWS Elastic Filesystem (EFS)

Good for use-cases where we need storage solutions that can be attached to multiple hosts at a given instant of time.

To access EFS file systems from on-premises, you must have an AWS Direct Connect or AWS VPN connection between your on-premises datacenter and your Amazon VPC.

EFS can be mounted in both EC2 instances and on-premises servers for use-cases where centralized file share is required.

Catch-Word: Network Attached Storage (NAS)

Module 60: AWS Elastic Filesystem (EFS)

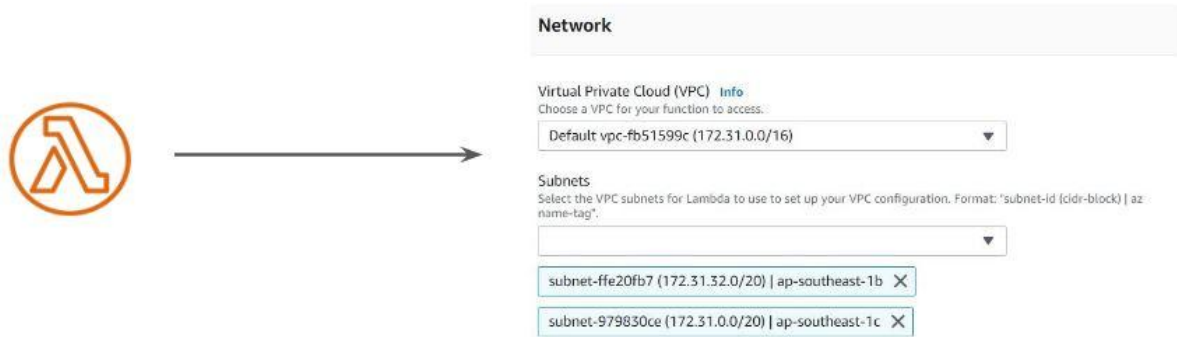
EFS - File Storage

EBS - Block Storage

S3 - Object Storage [Supports object operations via API]

Module 60: Lambda & VPC

When launching in a specific subnet in VPC, make sure that NAT gateway is attached in-case if you need internet access to the Lambda function.



Module 60: Lambda & IAM

Appropriate IAM role needs to be assigned to Lambda function if the function will be performing operations on EC2 resources.

Example:

- Lambda needs to stop EC2 instance at 12 AM every night, start at 9 AM every morning.
- IAM Role with Start and Stop EC2 instances should be associated with the Lambda function.

Module 60: Lambda & SQS Integration

SQS service can handle a huge spike in messages. Duplicates can be introduced in Standard queue.

The Lambda peak concurrency and function duration will determine how quickly you can process the messages sent to the queue (delay).

In programming, idempotency is the capacity of an application or component to identify repeated events and respond accordingly to prevent duplicated, inconsistent, or lost data.

You can architect your Lambda function idempotent to prevent inconsistencies

Module 61: Lambda@Edge

Lambda@Edge can also be used for authentication and authorization operations. Since it runs closest to the viewer, the overall latency is minimal.

For use-cases where you need to improve the overall login experience due to latency, you can consider using Lambda@Edge.

Module 62: EC2 Instance Types

Remember that T2 based instances are burstable and have moderate network performance. If the network is a bottleneck, you can move to instance types like M4 which has higher network performance.

Compute Intensive: C5 instances.

Memory Intensive: R5 instances.

Module 63: Storage Gateway

Understand the different types of storage gateways:

- File Gateway
- Tape Gateway
- Volume Gateway

Module 64: Systems Manager (SSM)

Run Command:

Allows to run set of command document on the target instance.

Patch Manager:

Deploy patches using Run Command in a specific maintenance window with the appropriate approval. Be aware of the entire process. Should also know about the maintenance window.

Parameter Store

A centralized place to store configuration data which includes credentials.

Module 65: Trusted Advisor

AWS Trusted Advisor analyzes your AWS environment and provides best practice recommendations in five major categories:

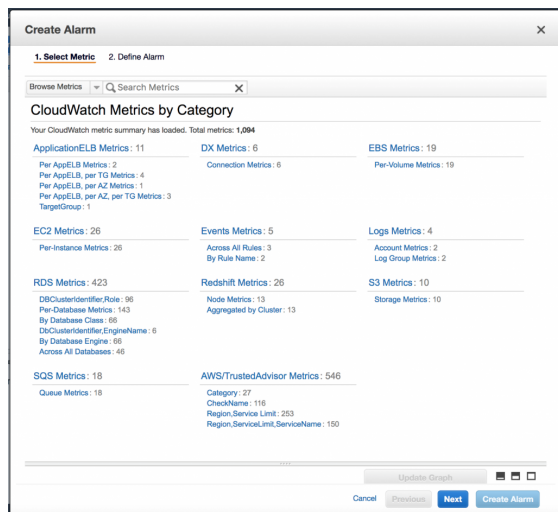
For checks like Cost Optimization, customers can choose to purchase Business support plan.



Module 66: Service Limits with Trusted Advisor & CloudWatch

Trusted Advisor exposes metrics in Amazon CloudWatch.

Business and Enterprise support customers can create customizable alarms for individual service limits.



Module 67: CloudWatch Logs

CloudWatch logs can be used to store server and application logs centrally.

Batch_size parameter specifies the max size of log events in a batch, in bytes. The default value is 1048576 bytes

If you want to immediately push the log message, you can reduce the batch_size to minimal.

Steps:

Assign appropriate IAM role to EC2 instance to push logs to CloudWatch.
Install the CloudWatch Agent.

Configure appropriate configuration and start CloudWatch Agent.

Module 68: Miscellaneous CloudWatch

The size of a PutMetricData request is limited to 8KB for HTTP GET requests and 40KB for HTTP POST requests.

You can assign up to 10 dimensions per metric.

Lambda automatically monitors the functions on your behalf and the following metrics are monitored:

Invocations, errors, duration, throttles, concurrent executions.

For Auto-Scaling, CloudWatch monitors 8 metrics and 1 Dimension.

CloudWatch supports basic as well as detailed monitoring. Basic Monitoring is 5 minutes and Detailed Monitoring is at a 1-minute interval.

The AWS SNS service sends data every 5 minutes. Thus, it supports only basic monitoring. The user cannot enable detailed monitoring with SNS.

The CloudWatch resources are always region-specific and they will have the endpoint as region-specific. If you want to fetch the metrics associated with the resources from the Mumbai region (ap-south-1), then the endpoint would:

`monitoring.ap-south-1.amazonaws.com/`

One-time charges and refunds related metrics are not sent to CloudWatch from Billing.

When you create a metric, it can take up to 2 minutes before you can retrieve statistics for the new metric using the `get-metric-statistics` command. However, it can take up to 15 minutes before the new metric appears in the list of metrics retrieved using the `list-metrics` command.

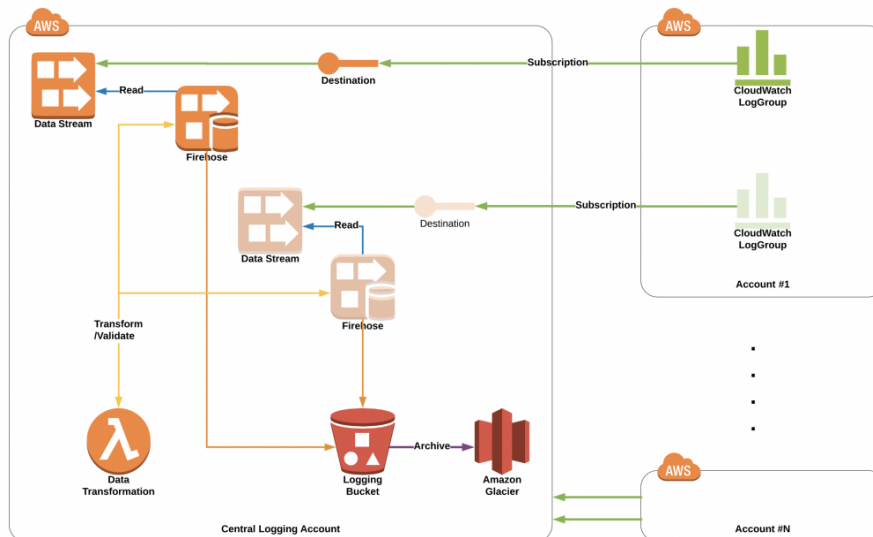
To fetch the metrics associated with Lambda function using CLI, the following commands can be used:

`list-metrics` and `get-metrics`

CloudWatch Metric filters can be used to identify the errors and CloudWatch Alarms can be used when it reaches the threshold.

Module 69: Cross-Account CloudWatch Logging

Create the Kinesis Data stream in Central Logging Account. Subscribe Kinesis Data stream to CloudWatch. Configure Kinesis Data stream as a source for Kinesis Firehose and push the data to the appropriate destination.

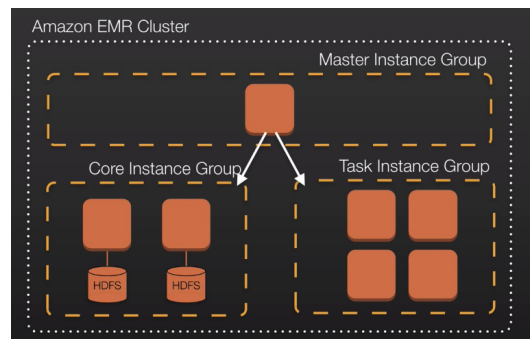


Module 70: EMR Components

Master nodes are responsible for the coordination and distribution of data and tasks among other nodes. It also keeps track of the status and checks the overall health of the cluster.

Core nodes contain both “Data Nodes” & “Task Tracker daemon”, thus it stores data based on HDFS and can also run tasks.

Task nodes only run the “TaskTracker” daemon and perform tasks.



Module 71: EMR Cost Considerations

We can make use of Reserved, On-Demand and Spot instances for components in EMR.

If EMR is going to run only once a day, then a scheduled reserved instance can also be used specifically with the master node.

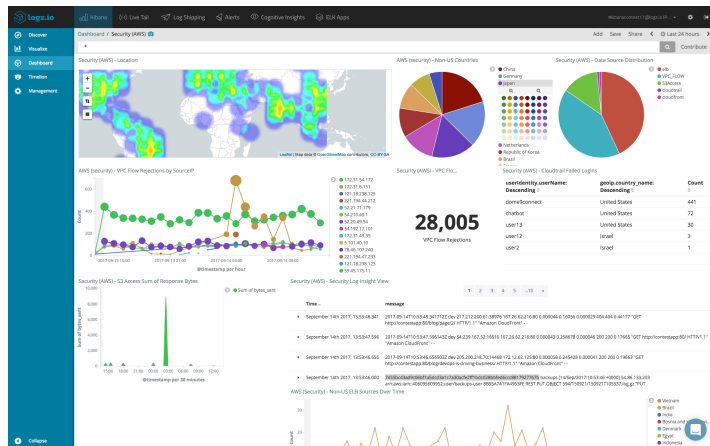
For additional tasks, a combination of on-demand with spot instances can be used for core and task node.

Module 72: AWS Elasticsearch Service

Amazon Elasticsearch Service is a fully managed service that makes it easy for you to deploy, secure, and operate Elasticsearch at scale with zero downtime.

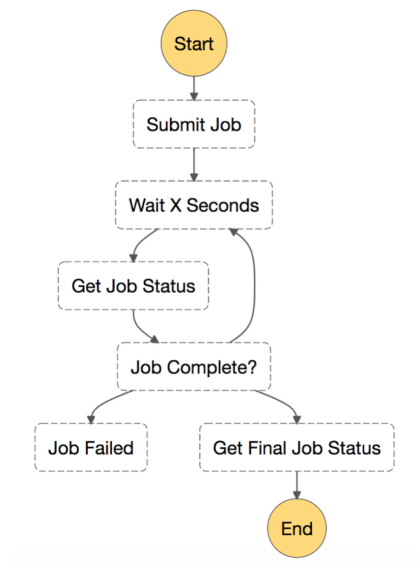
AWS DMS can be used to replicate data from on-premise ES to AWS ElasticSearch service.

Kibana can be used for data visualization.



Module 73: Step Function

Step Functions coordinates multiple AWS Lambda functions or other AWS resources based on user-defined workflow steps.



Module 74: AWS Batch

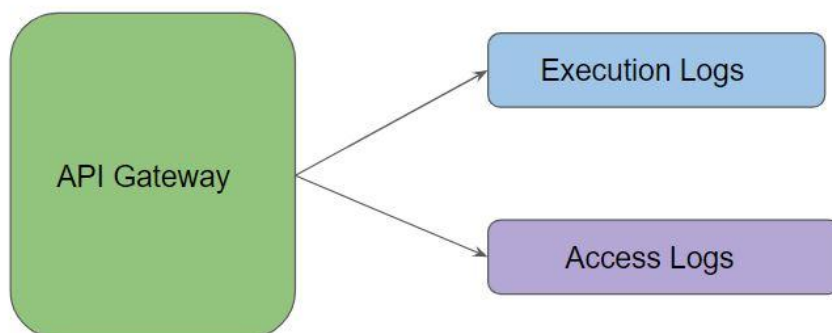
AWS Batch enables developers, scientists, and engineers to easily and efficiently run hundreds of thousands of batch computing jobs on AWS.

AWS Batch dynamically provisions the optimal quantity and type of compute resources (e.g., CPU or memory optimized instances) based on the volume and specific resource requirements of the batch jobs submitted.

Module 75: API Gateway Logging

Execution Logs records the API Gateway internal information as the request is processed.

Access Logs records details related to who has accessed the API.



Module 76: Troubleshooting Errors in API Gateway

An HTTP 504 status code (Gateway Timeout) indicates that when API Gateway forwarded a request to the origin, one of the following happened:

The origin returned an HTTP 504 status code to API Gateway.
The origin didn't respond before the request expired.

API Gateway will return an HTTP 504 status code if traffic is blocked to the origin by a firewall or security group, or if the origin isn't accessible.

AWS X-Ray, AWS CloudTrail, and Amazon CloudWatch are tools that Amazon API Gateway developers can use to trace, log, and monitor API execution and management operations.

Module 76: AWS Athena

AWS Athena is generally used for use-case where we want to analyze the logs from S3 like CloudTrail, VPC Flow Logs, and others with simple SQL statements in a serverless manner.

Any use-case that we might find in the exam was an analysis of logs stored in S3 without setting up in infrastructure, Athena might be the right option.

Module 76: AWS ElastiCache

ElastiCache is a fully managed AWS service that makes it easier to deploy, operate and scale an in-memory data-store or cache in the cloud.

Used very frequently for storing the sessions data centrally.

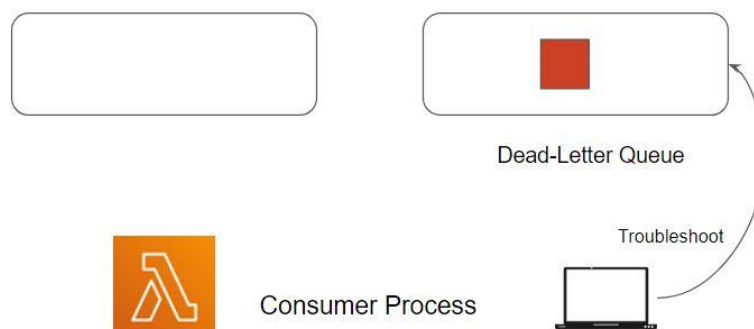
You can choose AWS ElastiCache - Redis if use-case requires replication, availability.

Redis also supports sub-millisecond latencies.

Module 77: SQS Dead-letter Queue

Dead-letter queues are useful for debugging your application or messaging system because they let you isolate problematic messages to determine why their processing doesn't succeed.

The messages are sent to the dead letter queue after exceeding the maximum receives.

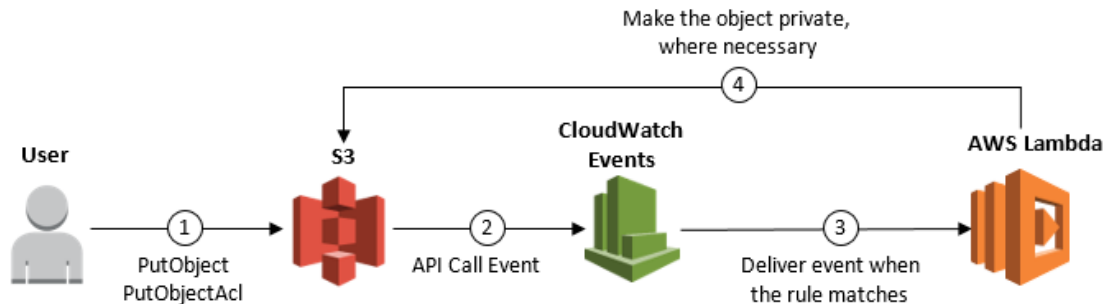


Module 78: S3 Object Level Logging

Allows customers to record object-level API operations on your S3 buckets.

Example Rule: Record all the PutObject API call with public-read permission.

Can automatically be remediated with CloudWatch Events and Lambda.



Module 78: Miscellaneous Pointers

API Gateway can be integrated with Cognito user pools to control access to the API.

In order to move an EC2 instance inside a placement group, stop the instance, modify the instance placement, and then restart the instance.