# AWS Solutions Associate Architect Associate

Page136

## Notes

### Intro

#### Intro to AWS

### Computing

#### Elastic Compute Cloud (EC2)

**Security Groups:** Only contain Allow rules

External firewall for EC2 instances (if a request is blocked by SG, instance will never know)

Security groups rules can reference a resource by IP or Security Group

Default SG: inbound traffic from the same SG is allowed; all outbound traffic is allowed

New SG: all inbound traffic is blocked; all outbound traffic is allowed

Since security groups are stateful, you can apply any changes to an incoming rule and it will be automatically applied to the outgoing rule.

A security group can be attached to multiple instances and vice versa

Bound to a VPC (and hence to a region)

**Purchasing Options**

On-demand Instances: Pay per use (no upfront payment); Highest cost; No long-term commitment; Recommended for short-term, uninterrupted and unpredictable workloads

Standard Reserved Instances: Reservation Period: 1 year or 3 years; Recommended for steady-state applications (like database); Sell unused instances on the Reserved Instance Marketplace

Convertible Reserved Instances: Can change the instance type; Lower discount; Cannot sell unused instances on the Reserved Instance Marketplace

Scheduled Reserved Instances: reserved for a time window (ex. everyday from 9AM to 5PM)

Spot Instances: Work on a bidding basis where you are willing to pay a specific max hourly rate for the instance. Your instance can terminate if the spot price increases.

Spot blocks are designed not to be interrupted

Good for workloads that are resilient to failure; Distributed jobs (resilient if some nodes go down); Batch jobs

Dedicated Hosts: Server hardware is allocated to a specific company (not shared with other companies); 3 year reservation period; Billed per host; Useful for software that have BYOL (Bring Your Own License) or for companies that have strong regulatory or compliance needs

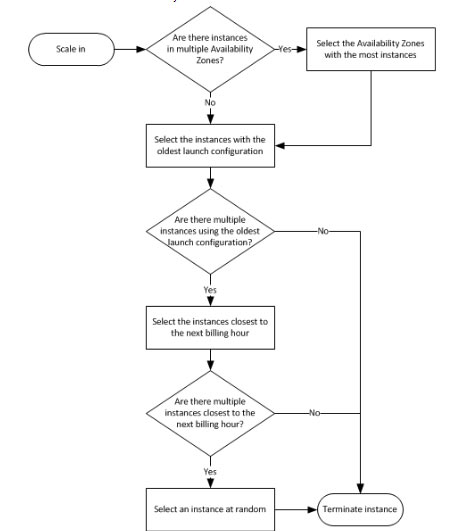
Dedicated Instances: Dedicated hardware; Billed per instance; No control over instance placement

On-Demand Capacity Reservations: Ensure you have the available capacity in an AZ to launch EC2 instances when needed; Can reserve for a recurring schedule (ex. everyday from 9AM to 5PM); No need for 1 or 3-year commitment (independent of billing discounts); Need to specify the following to create capacity reservation: - AZ - Number of instances - Instance attributes

#### Elastic Load Balancer (ELB)

Elastic Load Balancing distributes incoming application traffic across multiple targets

**Application Load Balancer (ALB):** Load balancing to multiple applications (target groups) based on the request parameters. Operates at Layer 7 (HTTP, HTTPS and WebSocket). Security Groups can be attached to ALBs. Termination policy: basically choose the oldest launch configuration:



**Network Load Balancer (NLB):** NLB is designed to handle TCP/UDP traffic and provides high availability and low latency. No security groups can be attached to NLBs.

**Cross-zone Load Balancing**

Allows ELBs in different AZ containing unbalanced number of instances to distribute the traffic evenly across all instances in all the AZ registered under a load balancer. With Application Load Balancers, cross-zone load balancing is always enabled.

#### Auto Scaling Group (ASG)

Auto Scaling can be used to automatically adjust the number of Amazon EC2 instances based on demand.

**Scaling Policies**

Scheduled Scaling: Scale based on a schedule; Used when the load pattern is predictable

Simple Scaling: Scale to certain size on a CloudWatch alarm. need to wait for the cooldown period to complete before initiating additional scaling activities Ex. when CPU > 90%, then scale to 10 instances

Step Scaling: Scale incrementally in steps using CloudWatch alarms; Specify the instance warmup time to scale faster Ex. when CPU > 70%, then add 2 units and when CPU < 30%, then remove 1 unit

Target Tracking Scaling: trigger a scaling activity immediately without waiting for the cooldown period to expire Ex. maintain CPU usage at 40%

Predictive Scaling: Historical data is used to predict the load pattern using ML and scale automatically

#### Lambda

AWS Lambda is a serverless compute service that's more suitable for event-driven, short-lived functions. AWS Lambda allows you to define environment variables that can be securely stored with encryption.

Lambda functions can absorb reasonable bursts of traffic for approximately 15-30 minutes. Lambda can scale faster than the regular Auto Scaling feature of Amazon EC2, Amazon Elastic Beanstalk, or Amazon ECS.

#### High Performance Computing (HPC)

Definition: High Performance Computing (HPC) refers to the utilization of advanced hardware and software to achieve exceptional computational capabilities, allowing for accelerated processing of intricate tasks and simulations.

Use Case: HPC finds application in fields like climate modeling, molecular simulations, and financial risk analysis, where massive data sets and complex calculations require rapid and efficient processing to gain insights and make informed decisions.

#### Workflows

### Storage

|  |  |  |
| --- | --- | --- |
| Database Type | Use Cases | AWS Service |
| Relational | Traditional applications, ERP, CRM, e-commerce, rigid schema, complex queries | Amazon RDS, Amazon Aurora, Amazon Redshift |
| Key-value | High-traffic web apps, e-commerce systems, gaming applications | Amazon DynamoDB |
| In-memory | Caching, session management, gaming leaderboards, geospatial applications | Amazon ElastiCache for Memcached, Amazon ElastiCache for Redis |
| Document | Content management, catalogs, user profiles | Amazon DocumentDB (with MongoDB compatibility) |
| Wide column | High-scale industrial apps for equipment maintenance, fleet management, and route optimization | Amazon Keyspaces (for Apache Cassandra) |
| Graph | Fraud detection, social networking, recommendation engines | Amazon Neptune |
| Time series | IoT applications, DevOps, industrial telemetry | Amazon Timestream |
| Ledger | Systems of record, supply chain, registrations, banking transactions | Amazon QLDB |

#### Instance Store

#### Elastic Block Storage (EBS)

Amazon EBS provides block-level storage(traditional filesystems and databases) volumes that are attached to Amazon EC2 instances(1 instance at a time). It offers different types of volumes optimized for various workloads, including General Purpose SSD, Provisioned IOPS SSD, Throughput Optimized HDD, and Cold HDD. EBS volumes are designed for use cases where you need block storage that can be attached to EC2 instances and provide specific performance characteristics.

Amazon EBS can deliver performance for workloads that require the lowest-latency access to data from a single EC2 instance. You can also increase EBS storage for up to 16TB or add new volumes for additional storage.

By default, upon instance termination, the root EBS volume is deleted and any other attached EBS volume is not deleted (can be over-ridden using DeleteOnTermination attribute)

**General Purpose SSD (gp3 and gp2):** Balanced performance for various workloads with gp2 providing burstable IOPS for light workloads.

**Provisioned IOPS SSD (io1 and io2):** Optimized for high and consistent I/O performance, ideal for databases and transaction-intensive applications.

Supports EBS Multi-attach (not supported by other types) Storage: 4 GB - 16 TB

Max IOPS: 64,000 for Nitro EC2 instances & 32,000 for non-Nitro

50 lOPS per GB (64,000 IOPS at 1,280 GB)

**io2 Block Express:** Provides exceptionally high IOPS and low latency, suitable for the most demanding performance needs. Storage: 4 GiB - 64 TB Sub-millisecond latency

Max IOPS: 256,000 1000 lOPS per GB

**Throughput Optimized HDD (st1):** Best for large sequential data transfers, suited for big data and data warehousing.

**Cold HDD (sc1):** Lowest cost storage option for infrequently accessed data, suitable for archival and backup purposes.

**Snapshots** occur asynchronously; the point-in-time snapshot is created immediately, but the status of the snapshot is pending until the snapshot is complete. While it is completing, an in-progress snapshot is not affected by ongoing reads and writes to the volume hence, you can still use the EBS volume normally.

Data Lifecycle Manager (DLM) can be used to automate the creation, retention, and deletion of snapshots of EBS volumes. Combined with the monitoring features of Amazon CloudWatch Events and AWS CloudTrail, Amazon DLM provides a complete backup solution for EBS volumes at no additional cost.

#### Elastic File System (EFS)

Amazon EFS is a scalable file storage service that can be mounted directly to multiple EC2 instances. It's designed for applications that require shared access to file-based storage and need to scale easily. EFS is particularly useful for scenarios where multiple instances need to access the same data, such as web servers in a cluster.

Amazon EFS is a fully-managed service that makes it easy to set up and scale file storage in the Amazon Cloud. Tens, hundreds, or even thousands of Amazon EC2 instances can access an Amazon EFS file system at the same time, and Amazon EFS provides consistent performance to each Amazon EC2 instance.

#### Simple Storage Service (S3)

Amazon S3 is an object storage service that offers scalable and durable storage for a wide range of data types. It's designed to store and retrieve large amounts of unstructured data, such as images, videos, backups, and more. S3 data is accessible over the internet via APIs. S3 offers features like object versioning, lifecycle policies, and access control to manage data retention, access permissions, and storage costs efficiently.

S3 and S3 Glacier are the preferred AWS services for the storage layer for your data lake. S3 is a service where the underlying infrastructure hosting the service is managed for you. if you upload an object to S3, the likelihood that that object would be corrupted or lost over time is incredibly low.

**Server Side Encryption(SSE):** data encryption at rest

**SSE-S3**: Keys managed by S3; AES-256 encryption. HTTP or HTTPS can be used. **Use Case**: Suitable for scenarios where you want a simple and hassle-free way to ensure data at rest is encrypted without the need to manage encryption keys yourself.

**SSE-C:** Keys managed by the client.Client sends the key in HTTPS headers for encryption/decryption (S3 discards the key after the operation); HTTPS must be used as key (secret) is being transferred. **Use Case**: Useful when you want more control over the encryption keys and want to manage them yourself while still benefiting from server-side encryption.

**SSE-KMS**: KMS manage the encryption keys.KMS manages the master keys. HTTP or HTTPS can be used. KMS provides control over who has access to what keys as well as audit trails. **Use Case**: Ideal when you want granular control over encryption keys, automated key rotation, and detailed audit logs of key usage and management.

**Client-Side Encryption**: Keys managed by the client; Client encrypts the object before sending it to S3 and decrypts it after retrieving it from S3.

**Access Management**

Amazon S3 objects are private by default.

User based security: IAM policies define which API calls should be allowed for a specific user.

Resource based security (Bucket Policy): Grant public access to the bucket; Force objects to be encrypted at upload; Cross-account access; Object Access Control List (ACL) - applies to the objects while uploading; Bucket Access Control List (ACL) - access policy that applies to the bucket.

**S3 Static Websites**

To host an S3 static website on a custom domain using Route 53(domain registrar, automatically

configure Route 53 as the DNS service for the domain), the bucket name should be the same as your domain or subdomain.

**Storage Classes**

Standard

Infrequent access: For data that is infrequently accessed, but requires rapid access when needed

Lower storage cost than Standard but cost on retrieval

Can move data into IA from Standard only after 30 days

Two types:

Standard IA 99.9% Availability

One-Zone IA 99.5% Availability Data is lost if AZ fails Storage for infrequently accessed data that can be easily recreated

Glacier: long-term archival storage. The Vault Lock feature in Amazon Glacier allows you to set a policy to lock the vault, preventing data deletion for a specified period of time.

Intelligent tiering

**S3 Select**

Select a subset of data from S3 using SQL queries (server-side filtering, filter the contents); Less network cost; Less CPU cost on the client-side

**Performance**

Performance Optimizations

1. Multi-part Upload: parallelizes upload; recommended for files > 100MB; must use for files > 5GB

2. Byte-range fetches: Parallelize download requests by fetching specific byte ranges in each request; Better resilience in case of failures since we only need to refetch the failed byte range and not the whole file

3. S3 Transfer Acceleration: Speed up upload and download for large objects (>1GB) for global users; Data is ingested at the nearest edge location and is transferred over AWS private network (uses CloudFront internally)

**S3 Notification Events**

The Amazon S3 notification feature enables you to receive notifications when certain events happen in your bucket. Amazon S3 event notifications are designed to be delivered at least once and to one destination only.

Amazon S3 supports the following destinations where it can publish events:

- Amazon Simple Notification Service (Amazon SNS) topic

- Amazon Simple Queue Service (Amazon SQS) queue

- AWS Lambda

**MFA Delete**

MFA required to: permanently delete an object version; suspend versioning on the bucket

Bucket Versioning must be enabled

Can only be enabled or disabled by the root user

#### Relational Database Service (RDS)

Amazon RDS is a managed database service that simplifies the setup, operation, and scaling of relational databases. It supports popular database engines like MySQL, PostgreSQL, MariaDB, Oracle, and Microsoft SQL Server. Automatic failover by promoting the standby replica only occurs if the primary database is the one that is affected.

When you create or modify your DB instance to run as a Multi-AZ deployment, Amazon RDS automatically provisions and maintains a synchronous standby replica in a different Availability Zone. Updates to your DB Instance are synchronously replicated across Availability Zones to the standby in order to keep both in sync and protect your latest database updates against DB instance failure.

**Read Replicas:** Allows us to scale the read operation (SELECT) on RDS. You can create a read replica as a Multi-AZ DB instance. A standby of the replica will be created in another AZ for failover support for the replica. Asynchronous replication.

**RDS Events**

RDS events only provide operational events on the DB instance (not the data)

To capture data modification events, use native functions or stored procedures to invoke a Lambda function.

**Monitoring**

CloudWatch Metrics for RDS: Gathers metrics from the hypervisor of the DB instance: CPU Utilization; Database Connections; Freeable Memory

Enhanced Monitoring: Gathers metrics from an agent running on the RDS instance: OS processes; RDS child processes; Used to monitor different processes or threads on a DB instance.

Enhanced Monitoring metrics are useful when you want to see how different processes or threads on a DB instance use the CPU. CloudWatch gathers metrics about CPU utilization from the hypervisor (hypervisor layer performs a small amount of work) for a DB instance, and Enhanced Monitoring gathers its metrics from an agent on the instance.

**RDS Proxy**

RDS Proxy helps you manage a large number of connections from Lambda to an RDS database by establishing a warm connection pool to the database.

“Too many connections” error-> all available connections are in use-> option1. Upgrade with higher memory (more memory, more connection) option2. RDS proxy connection pool to reuse connection.

**DynamoDB Accelerator (DAX)**

DAX is primarily used for read performance improvement of your DynamoDB table from milliseconds response time to microseconds.

#### Aurora

Amazon Aurora is a MySQL and PostgreSQL-compatible relational database engine that offers high performance, availability, and durability. Aurora is designed for applications that require high performance and availability with the benefits of managed services.

**Endpoints**

Amazon Aurora typically involves a cluster of DB instances instead of a single instance. When you connect to an Aurora cluster, the host name and port that you specify point to an intermediate handler called an endpoint.

Writer Endpoint (Cluster Endpoint): Always points to the master (can be used for read/write); Each Aurora DB cluster has one cluster endpoint

Reader Endpoint: Provides load-balancing for read replicas only (used to read only); If the cluster has no read replica, it points to master (can be used to read/write); Each Aurora DB cluster has one reader endpoint

Custom Endpoint: Used to point to a subset of replicas; Provides load-balanced based on criteria other than the read-only or read-write capability of the DB instances like instance class (ex, direct internal users to low-capacity instances and direct production traffic to high-capacity instances)

**Aurora Serverless**

Amazon Aurora Serverless is an on-demand, auto-scaling configuration for Amazon Aurora. An Aurora Serverless DB cluster is a DB cluster that automatically starts up, shuts down, and scales up or down its compute capacity based on your application's needs. Aurora Serverless provides a relatively simple, cost- effective option for infrequent, intermittent, sporadic or unpredictable workloads.

Take note that a non-Serverless DB cluster for Aurora is called a provisioned DB cluster. Aurora Serverless clusters and provisioned clusters both have the same kind of high-capacity, distributed, and highly available storage volume. When you work with Amazon Aurora without Aurora Serverless (provisioned DB clusters), you can choose your DB instance class size and create Aurora Replicas to increase read throughput. If your workload changes, you can modify the DB instance class size and change the number of Aurora Replicas. This model works well when the database workload is predictable, because you can adjust capacity manually based on the expected workload.

Amazon Aurora Global Database is designed for globally distributed applications, allowing a single Amazon Aurora database to span multiple AWS regions. Aurora Global Database supports storage-based replication that has a latency of less than 1 second. If there is an unplanned outage, one of the secondary regions you assigned can be promoted to read and write capabilities in less than 1 minute. This feature is called Cross-Region Disaster Recovery. An RPO of 1 second and an RTO of less than 1 minute provides you a strong foundation for a global business continuity plan.

**Aurora Events**

Invoke a Lambda function from an Aurora MySQL-compatible DB cluster with a native function or a stored procedure (same as RDS)

Used to capture data changes whenever a row is modified

#### DynamoDB

Amazon DynamoDB Accelerator (DAX) is a fully managed, highly available, in-memory cache that can reduce Amazon DynamoDB response times from milliseconds to microseconds, even at millions of requests per second.

DynamoDB Streams: Ordered stream of notifications of item-level modifications (create/update/delete) in a table

The partition key portion of a table's primary key determines the logical partitions in which a table's data is stored. In general, you will use your provisioned throughput more efficiently as the ratio of partition key values accessed to the total number of partition key values increases. One example for this is the use of partition keys with high-cardinality attributes, which have a large number of distinct

values for each item.

#### ElastiCache

Using Redis AUTH command can improve data security by requiring the user to enter a password before they are granted permission to execute Redis commands on a password-protected Redis server.

#### FSx

Amazon FSx provides fully managed file systems that are compatible with Windows and Lustre.

Amazon FSx for Windows File Server is a fully managed Microsoft Windows file system with full support for the SMB protocol, Windows NTFS, Microsoft Active Directory (AD) Integration.

Amazon FSx For Lustre is a high-performance file system for fast processing of workloads. Lustre is a popular opensource parallel file system which stores data across multiple network file servers to maximize performance and reduce bottlenecks.

#### Redshift

Amazon Redshift is a fast, scalable data warehouse that makes it simple and cost-effective to analyze all your data across your data warehouse and data lake. Redshift delivers **10x** faster performance than other data warehouses by using machine learning, massively parallel query execution, and **columnar** storage on high-performance disk. Integrated with Business Intelligence (**BI**) tools such as QuickSight or Tableau. Used for Online Analytical Processing (**OLAP**) and high-performance querying.

Amazon Redshift Spectrum is a feature of Amazon Redshift that enables you to query and analyze all of your data in Amazon S3 using the open data formats you already use, with no data loading or transformations needed.

#### Neptune

#### ElasticSearch

### Data Migration & Sync

#### Snow Family

#### Database Migration Service (DMS)

#### Storage Gateway

Bridge between on-premises data and S3 for Hybrid Cloud

**File Gateway**: Provides a file interface to Amazon S3, allowing you to access S3 objects as files, suitable for backup, archiving, and file sharing scenarios.

A file gateway supports a file interface into Amazon Simple Storage Service (Amazon S3) and combines a service and a virtual software appliance. By using this combination, you can store and retrieve objects in Amazon S3 using industry standard file protocols such as Network File System (NFS) and Server Message Block (SMB).

**Volume Gateway (Cached Volumes):** Allows you to store frequently accessed data locally while keeping a copy in Amazon S3, ideal for extending on-premises storage and providing low-latency access to data.

**Volume Gateway (Stored Volumes):** Provides block storage volumes that can be mounted as iSCSI devices, with the data stored directly in Amazon S3, well-suited for disaster recovery and backup.

**Tape Gateway**: Presents a virtual tape library interface, enabling you to store backup data in Amazon S3 and retrieve it using virtual tapes, suitable for archival and long-term data retention.

**Amazon FSx File Gateway:** Extends on-premises file systems into Amazon FSx for Windows File Server, providing low-latency access to shared file storage in the cloud and enabling hybrid cloud use cases.

#### DataSync

Move large amounts of data from your on-premises NAS or file system via NFS or SMB protocol to AWS over the public internet using TLS. AWS DataSync makes it simple and fast to move large amounts of data online between on-premises storage and Amazon S3, Amazon Elastic File System (Amazon EFS), or Amazon FSx for Windows File Server.

#### AppSync

#### Transfer Family

### Networking

#### Route 53

**Amazon Route 53**: Amazon Route 53 is a scalable and highly available Domain Name System (DNS) web service that helps route user requests to endpoints in a reliable manner. **Use case**: register domain names; route internet traffic to the resources for your domain; check the health of your resources

**Simple routing policy** – Use to route internet traffic to a single resource that performs a given function for your domain, for example, a web server that serves content for the example.com website.

**Failover routing policy** – Use when you want to configure active-passive failover.

**Geolocation routing policy** – Use when you want to route internet traffic to your resources based on the location of your users.

**Geoproximity routing policy** – Use when you want to route traffic based on the location of your resources and, optionally, shift traffic from resources in one location to resources in another.

**Latency routing policy** – Use when you have resources in multiple locations and you want to route traffic to the resource that provides the best latency.

**IP-based routing policy** – Use when you want to route traffic based on the location of your users, and have the IP addresses that the traffic originates from.

**Multivalue answer routing policy** – Use when you want Route 53 to respond to DNS queries with up to eight healthy records selected at random.

**Weighted routing policy** – Use to route traffic to multiple resources in proportions that you specify.

#### API Gateway

Amazon API Gateway is a fully managed service that makes it easy to create, publish, and maintain secure APIs at scale. APIs are the front door to backend applications and services. It allows the users to access the data using an API.

Throttling limits can be set for standard rates and bursts. Amazon API Gateway tracks the number of requests per second. Any requests over the limit will receive a 429 HTTP response. The client SDKs generated by Amazon API Gateway retry calls automatically when met with this response.

Serverless REST APIs; Invoke Lambda functions using REST APIs (API gateway will proxy the request to lambda); Can be integrated with any HTTP endpoint in the backend or any AWS API

#### Virtual Private Cloud (VPC)

**CIDR** (Classless Inter-Domain Routing): CIDR notation is used to specify IP address ranges. For example, 10.0.0.0/16 represents a range of IP addresses. **Use Case**: Defining address spaces for VPC subnets and IP ranges.

**Subnets:** Subnets are network segments within a VPC, allowing you to partition resources. **Use Case**: Organizing resources, such as placing web servers in one subnet and databases in another.

**Internet Gateway (IGW):** IGW provides a path for internet communication to/from VPC resources. **Use Case**: Enabling public internet access for resources like web servers.

**Bastion Hosts**: A bastion host is a special purpose computer on a network specifically designed and configured to withstand attacks. If you have a bastion host in AWS, it is basically just an EC2 instance. It should be in a public subnet with either a public or Elastic IP address with sufficient RDP(windows) or SSH(linux) access defined in the security group. Users log on to the bastion host via SSH or RDP and then use that session to manage other hosts in the private subnets.

**NAT Instance/Gateway**: Network Address Translation (NAT) instances or gateways allow private subnets to initiate outbound traffic to the internet. **Use Case**: Allowing instances in private subnets to access the internet for software updates.

**DNS Resolution in VPC**: DNS resolution enables instances to resolve domain names to IP addresses. **Use Case**: Ensuring that instances can access resources using domain names.

**Network Access Control List (NACL**): NACLs are stateless firewalls for controlling traffic in/out of subnets. To enable the connection to a service running on an instance, the associated network ACL must allow both inbound traffic on the port that the service is listening on as well as allow outbound traffic from ephemeral ports. **Use Case**: Controlling inbound/outbound traffic at the subnet level.

**VPC Peering**: VPC peering connects two VPCs, enabling direct communication between them. **Use Case**: Sharing resources between VPCs, such as cross-account communication.

**VPC Endpoints**: VPC endpoints enable private connections to AWS services without traversing the internet. **Use Case**: Securely accessing AWS services (e.g., S3, DynamoDB) from within a VPC.

A gateway endpoint is a gateway that you specify in your route table to access Amazon S3 from your VPC over the AWS network. Interface endpoints extend the functionality of gateway endpoints by using private IP addresses to route requests to Amazon S3 from within your VPC, on-premises, or from a different AWS Region.

**VPC Flow Logs:** VPC Flow Logs capture information about IP traffic going to and from network interfaces in a VPC. **Use Case**: Monitoring and troubleshooting network traffic and security.

**IPv6 Support:** VPCs can now support IPv6 addresses in addition to IPv4. **Use Case:** Providing connectivity for IPv6-enabled resources and clients.

A security group acts as a virtual firewall for your instance to control inbound and outbound traffic. When you launch an instance in a VPC, you can assign up to five security groups to the instance. Security groups act at the instance level, not the subnet level. Therefore, each instance in a subnet in your VPC can be assigned to a different set of security groups. The proper CIDR notation should be used. The /32 denotes one IP address and the /0 refers to the entire network. Take note that the SSH protocol uses TCP and port 22.

#### PrivateLink

#### Site-to-Site VPN

By default, instances that you launch into a virtual private cloud (VPC) can't communicate with your own network. You can enable access to your network from your VPC by attaching a virtual private gateway to the VPC, creating a custom route table, updating your security group rules, and creating an AWS managed VPN connection.

Although a VPN connection refers to the connection between your VPC and your own network. AWS supports Internet Protocol security (IPsec) VP N connections.

A customer gateway is a physical device or software application on your side of the VPN connection. To create a VPN connection, you must create a customer gateway resource in AWS, which provides information to AWS about your customer gateway device. Next, you have to set up an Internet-routable IP address (static) of the customer gateway's external interface.

**VPN CloudHub**

Low-cost hub-and-spoke model for network connectivity between a VPC and multiple on-premise data centers

#### Direct Connect (DX)

AWS Direct Connect is primarily used to establish a dedicated network connection from your premises to AWS.

#### Transit Gateway

Transit Gateway connects your VPC and on-premises networks through a central hub. It acts as a cloud router that allows you to integrate multiple networks.

#### Networking Costs

#### Reachability Analyzer

#### Traffic Mirroring

### Messaging

#### Simple Queue Service (SQS)

In Amazon SQS, you can configure the message retention period to a value from 1 minute to 14 days. The default is 4 days. Once the message retention limit is reached, your messages are automatically deleted. A single Amazon SQS message queue can contain an unlimited number of messages. However, there is a 120,000 limit for the number of inflight messages for a standard queue and 20,000 for a FIFO queue.

Always remember that the messages in the SQS queue will continue to exist even after the EC2 instance has processed it, until you delete that message.

**Configurations**

Short polling works for scenarios that require higher throughput.

Long polling poll the queue for longer; Decreases the number of API calls made to SQS (cheaper); Polling time: 1 sec to 20 sec; Can be enabled at the queue level or at the consumer level by using WaitTimeSeconds parameter in ReceiveMessage API.

ReceiveMessageWaitTimeSeconds: By default, its value is zero which means it is using Short polling. If it is set to a value greater than zero, then it is Long polling.

#### Simple Notification Service (SNS)

Used to broadcast messages; Pub-Sub model (publisher publishes messages to a topic, subscribers listen to the topic); Instant message delivery (does not queue messages)

Standard Topics

Subscribers can be:

SQS queues; HTTP / HTTPS endpoints; Lambda functions; Emails (using SNS)

SMS & Mobile Notifications; Kinesis Data Firehose (KDF) to send the data into S3 or Redshift

#### Kinesis

A new shard iterator is returned by every GetRecords request (as NextShardIterator), which you then use in the next GetRecords request (as ShardIterator). Typically, this shard iterator does not expire before you use it. However, you may find that shard iterators expire because you have not called GetRecords for more than 5 minutes, or because you've performed a restart of your consumer application. If the shard iterator expires immediately before you can use it, this might indicate that the DynamoDB table used by Kinesis does not have enough capacity to store the lease data. This situation is more likely to happen if you have a large number of shards. To solve this problem, increase the write capacity assigned to the shard table.

**AWS Kinesis Firehose** is for real-time data streaming and continuous data delivery

#### Amazon MQ

Amazon MQ is a managed message broker service for Apache ActiveMQ that makes it easy to set up and operate message brokers in the cloud. Connecting your current applications to Amazon MQ is easy because it uses industry standard APIs and protocols for messaging, including JMS, NMS, AMQP, STOMP, MQTT, and WebSocket. Using standards means that in most cases, there's no need to rewrite any messaging code when you migrate to AWS. Doesn’t “scale” as much as SQS or SNS because it is provisioned.

#### EventBridge

### Access Management

#### Identity & Access Management (IAM)

Groups are collections of users and have policies attached to them.

IAM User has limited permission to the account. Root User has full access to the account. IAM Roles: Collection of policies for AWS services

Take note that you can assign an IAM Role to the users or groups from your Active Directory once it is integrated with your VPC via the AWS Directory Service AD Connector.

IAM database authentication works with MySQL and PostgreSQL. you use an authentication token.

#### Cognito

#### Security Token Service (STS)

Used to grant limited and temporary access to AWS resources; Token is valid for up to 1h

Allows IAM Users to assume an IAM Role

single sign-on (SSO) approach to temporary access: provide users access to AWS without creating new AWS identities for them and requiring them to sign in with a separate user name and password.

#### Identity Federation in AWS

#### AWS Organizations

AWS Organizations helps you centrally manage and govern multiple AWS accounts. It allows you to create a hierarchy of accounts, making it easier to manage and apply policies across accounts.

Manage multiple AWS accounts under an organization: 1 master account, member accounts

**Service Control Policies (SCPs**): SCPs are a feature of AWS Organizations that allow you to set fine-grained permissions for the AWS services and actions that can be performed within each account.

Whitelist or blacklist IAM actions at the OU or Account level

Does not apply to the Master Account

#### AWS Directory Services

Used to extend the AD(Active Directory) network by involving services like EC2 to be a part of the AD to share login credentials.

#### Single Sign-On (SSO)

### Distribution

#### CloudFront

Amazon CloudFront, a content delivery network (CDN). CloudFront helps reduce the load on the web tier by serving static content from edge locations around the world.

Lambda@Edge is a feature of Amazon CloudFront that lets you run code closer to users of your application, which improves performance and reduces latency. By using Lambda@Edge and Kinesis together, you can process real-time streaming data so that you can track and analyze globally-distributed user activity on your website and mobile applications, including clickstream analysis.

Match Viewer is an Origin Protocol Policy which configures CloudFront to communicate with your origin using HTTP or HTTPS, depending on the protocol of the viewer request.

Amazon CloudFront delivers your content from each edge location and offers the same security as the Dedicated IP Custom SSL feature. SNI Custom SSL works with most modern browsers.

**Signed URL / Cookies**

CloudFront signed URLs and signed cookies provide the same basic functionality: they allow you to control who can access your content. If you want to serve private content through CloudFront and you're trying to decide whether to use signed URLs or signed cookies, consider the following:

Use signed URLs for the following cases:

- You want to use an RTMP distribution. Signed cookies aren't supported for RTMP distributions.

- You want to restrict access to individual files

- Your users are using a client (for example, a custom HTTP client) that doesn't support cookies.

Use signed cookies for the following cases:

- You want to provide access to multiple restricted files

- You don't want to change your current URLs

#### Global Accelerator

### Monitoring & Audit

#### CloudWatch

Amazon CloudWatch is a monitoring and management service that helps you track and optimize the performance of your AWS resources by collecting metrics, setting up alarms, and managing logs. It's used for real-time monitoring, resource optimization, and centralized logging.

The unified CloudWatch Logs agent is a single agent that can collect both Windows and Linux logs from EC2 instances and push them to CloudWatch Logs. The CloudWatch Logs agent will automatically collect and send logs to CloudWatch Logs without the need for custom scripts or manual configuration.

CloudWatch Logs Insights enables you to interactively search and analyze your log data in Amazon CloudWatch Logs.

CloudWatch has available Amazon EC2 Metrics for you to use for monitoring. CPU Utilization identifies the processing power required to run an application upon a selected instance. Network Utilization identifies the volume of incoming and outgoing network traffic to a single instance. Disk Reads metric can be used to determine the speed of the application. However, there are certain metrics that are not readily available in CloudWatch such as Memory utilization, Disk swap utilization, Disk space utilization, Page file utilization, Log coll.

#### CloudTrail

AWS CloudTrail is a service that enables governance, compliance, operational auditing, and risk auditing of your AWS account. With CloudTrail, you can log, continuously monitor, and retain account activity related to actions across your AWS infrastructure.

AWS CloudTrail is an auditing and security service that records detailed activity logs of actions taken on your AWS resources. It's essential for tracking changes, investigating security incidents, ensuring compliance, and providing an audit trail for user and API activity. CloudTrail provides detailed event history of AWS account activity, including actions taken through the AWS Management Console, AWS SDKs, command line tools, and other AWS services. It's a key service for monitoring and auditing actions in your AWS environment.

#### Config

#### X-Ray

AWS X-Ray helps you debug and analyze your microservices applications with request tracing so you can find the root cause of issues and performance.

#### Trusted Advisor

AWS Trusted Advisor is a service that provides recommendations for optimizing your AWS environment. AWS Trusted Advisor is an online tool that provides you with real-time guidance to help you provision your resources following AWS best practices.

Cost Optimization - recommendations that can potentially save you money by highlighting unused resources and opportunities to reduce your bill.

Security - identification of security settings that could make your AWS solution less secure.

Fault Tolerance - recommendations that help increase the resiliency of your AWS solution by highlighting redundancy shortfalls, current service limits, and over-utilized resources.

Performance -recommendations that can help to improve the speed and responsiveness of your applications.

Service Limits - recommendations that will tell you when service usage is more than 80% of the service limit.

#### CostExplorer

### Containerization

#### Elastic Container Service (ECS)

Amazon Elastic Container Service (Amazon ECS) is a fully managed container orchestration service that allows you to easily run, manage, and scale Docker containers. It's suitable for managing containerized batch processing workloads.

Amazon ECS is an AWS proprietary container service. This means that it is not an open-source platform.

Amazon ECS enables you to inject sensitive data into your containers by storing your sensitive data in either AWS Secrets Manager secrets or AWS Systems Manager Parameter Store parameters and then referencing them in your container definition. This feature is supported by tasks using both the EC2 and Fargate launch types.

Amazon ECS doesn't support resource-based policies such as S3 bucket policy.

**AWS Fargate** is a serverless compute engine for containers

#### Elastic Kubernetes Service (EKS)

Amazon EKS is a portable, extensible, and open-source platform for managing containerized workloads and services. Kubernetes is considered cloud-agnostic because it allows you to move your containers to other cloud service providers.

Amazon EKS provisions and scales the Kubernetes control plane, including the API servers and backend persistence layer, across multiple AWS availability zones for high availability and fault tolerance.

Amazon EKS automatically detects and replaces unhealthy control plane nodes and provides patching for the control plane. Amazon EKS is integrated with many AWS services to provide scalability and security for your applications. These services include Elastic Load Balancing for load distribution, IAM for authentication, Amazon VPC for isolation, and AWS CloudTrail for logging.

#### Elastic Container Registry (ECR)

Amazon ECR is a fully-managed Docker container registry

### Deployment

#### CloudFormation

#### Elastic Beanstalk

AWS Elastic Beanstalk provides a fully managed platform-as-a-service (PaaS) that abstracts away the infrastructure management tasks. Elastic Beanstalk is designed to host web applications. Elastic Beanstalk offers automatic scaling, allowing your application to scale up or down based on demand. Elastic Beanstalk supports a variety of platforms and languages, making it versatile for different application types.

#### Serverless Application Model (SAM)

#### Continuous Integration Continuous Delivery (CICD)

#### Blue-Green Deployment

### Parameters & Encryption

#### Key Management Service (KMS)

#### SSM Parameter Store

Serverless; Used to store parameters & secrets ;;Parameter versioning

Seamless Encryption with KMS for encryption and decryption of stored secrets

Parameters are stored in hierarchical fashion

#### Secrets Manager

#### CloudHSM

AWS provisions dedicated encryption hardware (Hardware Security Module); Use when you want to manage encryption keys completely; HSM device is stored in AWS (tamper resistant, FIPS 140-2 Level 3 compliance)

4 reasons for using HSM:

1. keys that are explicitly required to be protected in a single-tenant HSM or in an HSM over which you have direct control.
2. keys that are required to be stored in an HSM that has been validated to FIPS 140-2 level 3 overall
3. need the ability to immediately remove key material from AWS KMS and to prove you have done so by independent means
4. be able to audit all use of your keys independently of AWS KMS or AWS CloudTrail

#### OpsWorks

### Analytics

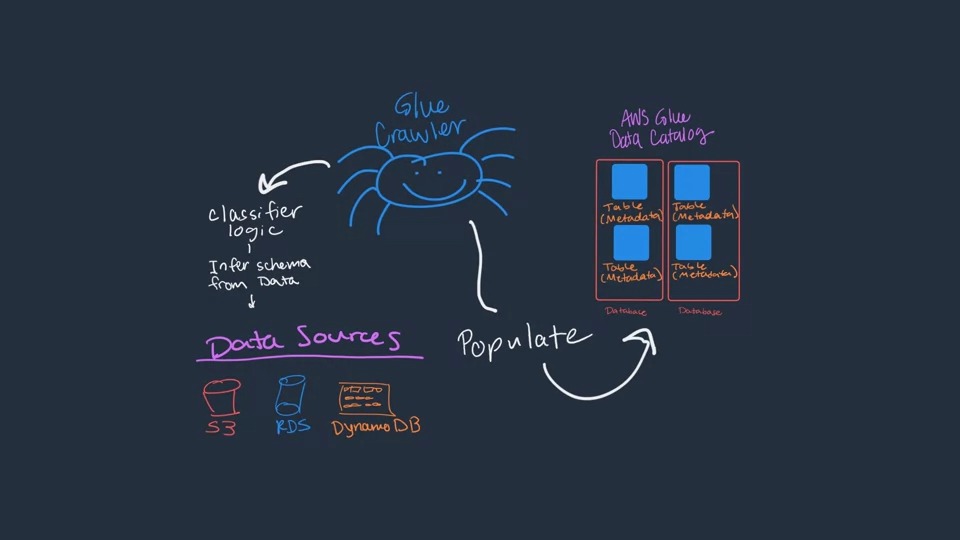
#### Athena

Amazon Athena is an interactive query service that allows you to analyze data in Amazon S3 using standard SQL queries without the need for infrastructure setup(Athena is serverless) or data movement. It's particularly suitable for running ad-hoc queries on large datasets stored in data lakes.

#### Elastic Map Reduce (EMR)

#### Glue

AWS Glue is a fully managed serverless data processing and cataloging service. Serverless Extract, Transform & Load (ETL) service. Used to prepare & transform data for analytics. Used to get data from a store, process and put it in another store (could be the same store). Glue Data Crawlers crawl databases and collect metadata which is populated in Glue Data Catalog



#### Lake Formation

AWS Lake Formation is a service that you can use to set up a secure data lake in days. A data lake is a centralized, curated, and secured repository that stores all your data, both in its original form and prepared for analysis. You can use a data lake to break down data silos and combine different types of analytics to gain insights and guide better business decisions.

AWS Lake Formation allows you to consolidate data from multiple AWS accounts into a single data lake. AWS Lake Formation provides fine-grained access control using policies, which allows you to grant access to users from different accounts based on their roles.

### Cloud Security

#### AWS Shield

In addition to the network and transport layer protections that come with Standard, AWS Shield Advanced provides additional detection and mitigation against large and sophisticated DDoS attacks, near real-time visibility into attacks, and integration with AWS WAF, a web application firewall.

#### Web Application Firewall (WAF)

AWS WAF is a web application firewall that lets you monitor the HTTP and HTTPS requests.

Protects your application from common layer 7 web exploits such as SQL Injection and Cross-Site Scripting (XSS)

Can only be deployed on: Application Load Balancer; API Gateway; CloudFront

WAF contains Web ACL (Access Control List) containing rules to filter requests based on: IP addresses; HTTP headers; HTTP body; URI strings ;Size constraints (ex. max 5kb) ;Geo-match (block countries) ;Rate-based rules (to count occurrences of events per IP) for DDoS protection(set up rate limiting for requests coming to your web application)

Associating the web ACL (Access Control List) with an Application Load Balancer ensures that the filtering occurs at the load balancer level before the requests reach the EC2 instances.

#### Firewall Manager

AWS Firewall Manager is mainly used to simplify your AWS WAF administration and maintenance tasks across multiple accounts and resources.

AWS Firewall Manager just simplifies your AWS WAF and AWS Shield Advanced administration and maintenance tasks across multiple accounts and resources.

#### GuardDuty

GuardDuty is a threat detection service that continuously monitors for malicious activity and unauthorized behavior to protect your AWS accounts and workloads.

#### Inspector

Inspector is basically an automated security assessment service that helps improve the security and compliance of applications deployed on AWS.

#### Macie

Amazon Macie is an ML-powered security service that helps you prevent data loss by automatically discovering, classifying, and protecting sensitive data stored in Amazon S3. Amazon Macie uses machine learning to recognize sensitive data such as personally identifiable information (PII) or intellectual property, assigns a business value, and provides visibility into where this data is stored and how it is being used in your organization.

Amazon Macie continuously monitors data access activity for anomalies, and delivers alerts when it detects risk of unauthorized access or inadvertent data leaks. Amazon Macie has ability to detect global access permissions inadvertently being set on sensitive data, detect uploading of API keys inside source code, and verify sensitive customer data is being stored and accessed in a manner that meets their compliance standards.

### Disaster Recovery

#### Disaster Recovery

#### AWS Backup

### Sharing

#### Resource Access Manager (RAM)

AWS Resource Access Manager (RAM) is a service that enables you to easily and securely share AWS resources with any AWS account or within your AWS Organization.

### Misc

#### Amazon WorkSpaces

#### Well Architected Framework

#### Extras

Using Reserved Instances for critical batch processing workloads provides cost savings over on-demand instances. Spot Instances offer significant cost savings compared to on-demand instances but are suitable for non-critical workloads due to their potential for interruption.

Cross-origin resource sharing (CORS) defines a way for client web applications that are loaded in one domain to interact with resources in a different domain. With CORS support, you can build rich client-side web applications with Amazon S3 and selectively allow cross-origin access to your Amazon S3 resources.

Amazon WorkDocs is simply a fully managed, secure content creation, storage, and collaboration service. With Amazon WorkDocs, you can easily create, edit, and share content. And because it's stored centrally on AWS, you can access it from anywhere on any device.

Field-Level Encryption allows you to securely upload user-submitted sensitive information to your web servers

## Architecture

Serverless ToDo List App

Serverless Blogging Website

Serverless Premium Stock Video Website

DDoS Protection

Cost Effective Highly Available Monolithic Architecture

## Theory

Concepts

Important Ports

Shared Responsibility Model

## **AWS analytics services**

|  |  |  |
| --- | --- | --- |
| **Solution areas** | **Use cases** | **AWS service** |
| [Advanced analytics](https://aws.amazon.com/big-data/datalakes-and-analytics/#Advanced_analytics) | Interactive analytics | [Amazon Athena](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) |
| Big data processing | [Amazon EMR](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) |
| Data warehousing | [Amazon Redshift](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) |
| Real-time analytics | [Amazon Kinesis Data Analytics](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) |
| Operational analytics | [Amazon OpenSearch Service](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) |
| Dashboards and visualizations | [Amazon QuickSight](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) |
| Visual data preparation | [Amazon Glue DataBrew](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) |
| [Data management](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) | Real-time data movement | [Amazon Managed Streaming for Apache Kafka (Amazon MSK)](https://aws.amazon.com/msk/?c=a&sec=srvm) |  [Amazon Kinesis Data Streams](https://aws.amazon.com/kinesis/data-streams/?c=a&sec=srvm) | [Amazon Kinesis Data Firehose](https://aws.amazon.com/kinesis/data-firehose/?c=a&sec=srvm) |  [Amazon Kinesis Video Streams](https://aws.amazon.com/kinesis/video-streams/?c=a&sec=srvm) | [AWS Glue](https://aws.amazon.com/glue/?c=a&sec=srvm) |
| Data governance | [Amazon DataZone](https://aws.amazon.com/datazone/) | [AWS Glue](https://aws.amazon.com/glue/?c=a&sec=srvm) | [AWS Entity Resolution](https://aws.amazon.com/entity-resolution/) |  [AWS Lake Formation](https://aws.amazon.com/lake-formation/?c=a&sec=srvm&loc=2) | [Amazon S3](https://aws.amazon.com/s3/?c=a&sec=srvm) |  [AWS Identity and Access Management](https://aws.amazon.com/iam/) | [AWS Data Exchange](https://aws.amazon.com/data-exchange/?c=a&sec=srvm) |  [AWS Clean Rooms](https://aws.amazon.com/clean-rooms/) |
| Object storage for data lakes | [Amazon S3](https://aws.amazon.com/s3/?c=a&sec=srvm) | [AWS Lake Formation](https://aws.amazon.com/lake-formation/?c=a&sec=srvm&loc=0) |
| Backup and archive for data lakes | [Amazon S3 Glacier](https://aws.amazon.com/glacier/?c=a&sec=srvm) | [AWS Backup](https://aws.amazon.com/backup/?c=a&sec=srvm) |
| Data catalog | [AWS Glue](https://aws.amazon.com/glue/?c=a&sec=srvm) | [AWS Lake Formation](https://aws.amazon.com/lake-formation/?c=a&sec=srvm&loc=2) |
| Third-party data | [AWS Data Exchange](https://aws.amazon.com/data-exchange/?c=a&sec=srvm) | [AWS Clean Rooms](https://aws.amazon.com/clean-rooms/) |
| [Predictive analytics & machine learning](https://aws.amazon.com/big-data/datalakes-and-analytics/#Solution_areas) | Frameworks and interfaces | [AWS Deep Learning AMIs](https://aws.amazon.com/machine-learning/amis/?c=a&sec=srvm) |
| Platform services | [Amazon SageMaker](https://aws.amazon.com/sagemaker/?c=a&sec=srvm) |