Cloud Computing

There are 3 types of cloud computing model

- Infrastructure as a service (IaaS) – managed up to the OS
- Platform as a service (PaaS) – managed up to the code
- Software as a service (SaaS) – pure consumption model

Cloud Computing

- There are 4 types of cloud deployment:
- Public Cloud or simple “Cloud” – e.g. AWS, Azure, GCP
- Hybrid Cloud – mixture of public and private clouds
- Private Cloud (on-premise) – managed in your own data centre, e.g. Hyper-V, OpenStack, VMware
- Multicloud – use private/public clouds from multiple providers

Cloud Computing

Fundamentals of pricing:

- Compute – CPU/RAM and duration
- Storage – quantity of data stored or allocated
- Outbound data transfer – data leaving an AWS Region

Cloud Computing

The AWS Global Infrastructure is made up of:

- AWS Regions
    - A region is a geographical area
    - Each region consists of 2 or more availability zones
    - Isolated from other AWS Regions

Availability Zones

- Availability Zones are physically separate and isolated from each other
- AZs span one or more data centers
- Each AZ is designed as an independent failure zone

Cloud Computing

Local Zones

- AWS Local Zones place compute, storage, database, and other select AWS services closer to end-users
- Extension of an AWS Region where you can run your latency sensitive applications

Edge Locations and Regional Edge Caches

- Edge locations are Content Delivery Network (CDN) endpoints for CloudFront
- There are many more edge locations than regions
- Regional Edge Caches sit between your CloudFront Origin servers and the Edge Locations
- A Regional Edge Cache has a larger cache-width than each of the individual Edge Locations

Cloud Computing

The AWS shared responsibility model defines customer/AWS responsibilities

- AWS are responsible for “Security of the Cloud”
    - AWS is responsible for protecting the infrastructure that runs all of the services offered in the AWS Cloud
    - This infrastructure is composed of the hardware, software, networking, and facilities that run AWS Cloud services

Customers are responsible for “Security in the Cloud”

- For EC2 this includes network level security, operating system patches and updates, IAM user access management, and client and server-side data encryption

Cloud Computing

The 6 advantages of cloud:

- Trade capital expense for variable expense • Benefit from massive economies of scale
- Stop guessing about capacity
- Increase speed and agility
- Stop spending money running and maintaining data centres
- Go global in minutes

Cloud Computing

Identity and Access Management (AWS IAM)

AWS IAM

- AWS Identity and Access Management (IAM) is a web service that helps you securely control access to AWS resources
- You use IAM to control who is authenticated (signed in) and authorized (has permissions) to use resources
- Users are individual accounts you log in with
- Users have NO permissions by default
- Groups are used for organizing users and applying policies
- Policies are used for defining permissions
- Roles are used for delegating permissions and are assumed by services

AWS IAM

- Users log in to the AWS Management Console with a user name and password
- Access keys are used for CLI/API access (programmatic)
- Access keys consist of an access key ID and secret access key
- The root user is the user that created the account
- Root users have full permissions and cannot be restricted
- Multi-factor authentication (MFA)uses a second factor in addition to a password – typically a code generated on a device

AWS IAM

- Service Control Policies (SCPs) are a feature of AWS Organizations
- SCPs control the maximum available permissions in an AWS account
- SCPs do not grant permissions

AWS IAM

IAM Best Practices:

- Lock away your AWS account root user access keys
- Create individual IAM users
- Use groups to assign permissions to IAM users
- Grant least privilege
- Get started using permissions with AWS managed policies
- Use customer managed policies instead of inline policies
- Use access levels to review IAM permissions
- Configure a strong password policy for your users
- Enable MFA

AWS IAM

IAM Best Practices ctd:

- Use roles for applications that run on Amazon EC2 instances
- Use roles to delegate permissions
- Do not share access keys
- Rotate credentials regularly
- Remove unnecessary credentials
- Use policy conditions for extra security
- Monitor activity in your AWS account

Cloud Computing

AWS Compute Services

Saravanan Chandrasekar

AWS Compute Services

Amazon EC2

- Amazon Elastic Compute Cloud (Amazon EC2) is a web service with which you can run virtual server “instances” in the cloud
- Amazon EC2 instances can run the Windows, Linux, or MacOS operating systems
- Amazon Machine Image (AMI) is used to launch an EC2 instance – consists of EBS snapshot, permissions and configuration

AWS Compute Services

Amazon EC2 Metadata and User Data

- User data is data that is supplied by the user at instance launch in the form of a script
- Instance metadata is data about your instance that you can use to configure or manage the running instance
- User data and metadata are not encrypted
- Instance metadata is available at http://169.254.169.254/latest/meta-data

AWS Compute Services

- Access keys can be used on EC2 instances to gain permissions to other AWS services
- Access keys are stored in plaintext so this is not secure
- Better to use IAM roles whenever possible and avoid access keys

AWS Compute Services

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

AWS Compute Services

Amazon LightSail

- Amazon Lightsail is great for users who do not have deep AWS technical expertise as it make it very easy to provision compute services
- Amazon Lightsail provides compute, storage, and networking capacity and capabilities to deploy and manage websites, web applications, and databases in the cloud
- Best suited to projects that require a few dozen instances or fewer
- Provides a simple management interface
- Good for blogs, websites, web applications, e-commerce etc.
- Can deploy load balancers and attach block storage

AWS Compute Services

Amazon Elastic Container Service (ECS)

- ECS is used for running Docker containers in the cloud
- ECS containers are known as tasks

EC2 launch type:

- You managed EC2 instances which are the hosts for running the tasks

Fargate launch type:

- AWS manage the underlying compute, cluster, and scaling
- Amazon Elastic Container Registry (ECR) is a private container image registry

AWS Storage Services

Saravanan Chandrasekar

AWS Storage Services

Amazon Elastic Block Store (EBS)

- EBS volume data persists independently of the life of the instance
- EBS volumes do not need to be attached to an instance
- You can attach multiple EBS volumes to an instance
- You can use multi-attach to attach a volume to multiple instances but with some constraints
- EBS volumes must be in the same AZ as the instances they are attached to
- Root EBS volumes are deleted on termination by default
- Extra non-boot volumes are not deleted on termination by default

AWS Storage Services

- Snapshots capture a point-in-time state of an instance
- Snapshots are stored on S3
- If you make periodic snapshots of a volume, the snapshots are incremental
- EBS volumes are AZ specific but snapshots are region specific

AWS Storage Services

Data Lifecycle Manager (DLM)

• DLM automates the creation, retention, and deletion of EBS snapshots and EBS-backed AMIs

DLM helps with the following:

• Protects valuable data by enforcing a regular backup schedule

• Create standardized AMIs that can be refreshed at regular intervals

• Retain backups as required by auditors or internal compliance

• Reduce storage costs by deleting outdated backups

• Create disaster recovery backup policies that back up data to isolated accounts

AWS Storage Services

Instance Store Volumes

- Instance store volumes are high performance local disks that are physically attached to the host computer on which an EC2 instance runs
- Instance stores are ephemeral which means the data is lost when powered off (non-persistent)
- Instances stores are ideal for temporary storage of information that changes frequently, such as buffers, caches, or scratch data

AWS Storage Services

Amazon Elastic File System (EFS)

- File-based storage system
- Uses the NFS protocol
- Can connect many EC2 instance concurrently
- EC2 instances can be connected from multiple AZs
- Only available for Linux instances
- Can connect instances from other VPCs

AWS Storage Services

Amazon Simple Storage Service (S3)

- You can store any type of file in S3
- Files can be anywhere from 0 bytes to 5 TB
- There is unlimited storage available
- S3 is a universal namespace so bucket names must be unique globally
- However, you create your buckets within a REGION
- It is a best practice to create buckets in regions that are physically closest to your users to reduce latency

AWS Storage Services

Six S3 Storage Classes

- S3 Standard (durable, immediately available, frequently accessed)
- S3 Intelligent-Tiering (automatically moves data to the most cost- effective tier)
- S3 Standard-IA (durable, immediately available, infrequently accessed)
- S3 One Zone-IA (lower cost for infrequently accessed data with less resilience)
- S3 Glacier (archived data, retrieval times in minutes or hours)
- S3 Glacier Deep Archive (lowest cost storage class for long term retention)

AWS Storage Services

Additional S3 Features

- Transfer acceleration – speeds up uploads using CloudFront
- Requester pays – the account requesting the objects pays
- Events – can trigger notifications to SNS, SQS and Lambda
- Static website hosting – setup a static website • Encryption – encrypt objects in the bucket
- Replication – replicate within (SRR) or across (CRR) Regions

AWS Storage Services

S3 Versioning

- Versioning is a means of keeping multiple variants of an object in the same bucket
- Use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket
- Versioning-enabled buckets enable you to recover objects from accidental deletion or overwrite

AWS Storage Services

S3 Glacier

- Extremely low cost and you pay only for what you need with no commitments of upfront fees
- Two classes Glacier and Glacier Deep Archive
- Three options for access to archives, listed in the table below: Expedited Standard Bulk Data access time (Glacier) 1-5 minutes 3-5 hours 5-12 hours Data access time (Deep Archive) N/A 12 hours 48

AWS Storage Services

S3 Object Lock

• Store objects using a write-once-read-many (WORM) model • Prevent objects from being deleted or overwritten for a fixed time or indefinitely

S3 Glacier Vault Lock

• Also used to enforce a WORM model

• Can apply a policy and lock the policy from future edits • Use for compliance objectives and data retention

AWS Storage Services

AWS Storage Gateway

- Hybrid cloud storage service
- Access cloud storage from on-premises applications
- Enables access to proprietary object storage (S3) using standard protocols

Use cases:

- Moving backups to the cloud
- Using on-premises file shares backed by cloud storage • Low latency access to data in AWS for on-premises applications • Disaster recovery

AWS Storage Services

Storage Gateway offers three different types of gateways:

- File Gateway - provides file system interfaces to on-premises servers
- Volume Gateway - provides block-based access for onpremises servers
- Tape Gateway - provides a virtual tape library that is compatible with common backup software (block and file interfaces)

DNS, Elastic Load Balancing, and Auto Scaling

Saravanan Chandrasekar

DNS, Elastic Load Balancing, and Auto Scaling

Amazon Route 53

- Route 53 is the AWS Domain Name Service
- Route 53 performs three main functions:
  - Domain registration – Route 53 allows you to register domain names
  - Domain Name Service (DNS) – Route 53 translates name to IP addresses using a global network of authoritative DNS servers
  - Health checking – Route 53 sends automated requests to your application to verify that it's reachable, available and functional

DNS, Elastic Load Balancing, and Auto Scaling

Amazon Route 53 Routing Policies

- Simple – IP address associated with name
- Failover – if primary is down, route to secondary
- Geolocation – route based on geographic location of request
- Geoproximity – route to closes Region withing geo area
- Latency – use lowest latency route to resources
- Multivalue answer – returns several IP addresses
- Weighted – relative weights (e.g. 80%/20%)

DNS, Elastic Load Balancing, and Auto Scaling

Amazon EC2 Auto Scaling

- Automates scaling of EC2 instances
- Launches and terminates EC2 instances based on demand
- Helps to ensure that you have the correct number of EC2 instances available to handle the application load
- Amazon EC2 Auto Scaling provides elasticity and scalability
- You create collections of EC2 instances, called an Auto Scaling group (ASG)

DNS, Elastic Load Balancing, and Auto Scaling

Amazon EC2 Auto Scaling

- Responds to EC2 status checks and CloudWatch metrics
- Can scale based on demand (performance) or on a schedule
- Scaling policies define how to respond to changes in demand
- Scaling policies include:
    - Target Tracking – Attempts to keep the group at or close to the metric
    - Simple Scaling – Adjust group size based on a metric
    - Step Scaling – Adjust group size based on a metric – adjustments vary based on the size of the alarm breach
    - Scheduled Scaling – Adjust the group size at a specific time

DNS, Elastic Load Balancing, and Auto Scaling

Amazon Elastic Load Balancing

- ELB automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses
- ELB can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones
- ELB features high availability, automatic scaling, and robust security necessary to make your applications fault tolerant

DNS, Elastic Load Balancing, and Auto Scaling

There are two types of Elastic Load Balancer (ELB) that may feature on the exam:

- Application Load Balancer (ALB) – layer 7 load balancer that routes connections based on the content of the request
- Network Load Balancer (NLB) – layer 4 load balancer that routes connections based on IP protocol data

Application Services

Application Services

- With serverless there are **no instances** to manage
- You don’t need to provision hardware
- There is no management of operating systems or software
- Capacity provisioning and patching is handled automatically
- Provides automatic scaling and high availability

Application Services

- Serverless services include:
- AWS Lambda
- AWS Fargate
- Amazon EventBridge
- AWS Step Functions
- Amazon SQS
- Amazon SNS
- Amazon API Gateway
- Amazon S3
- Amazon DynamoDB

Application Services

AWS Lambda Functions

- AWS Lambda executes code only when needed and scales automatically
- You pay only for the compute time you consume (you pay nothing when your code is not running)

Benefits of AWS Lambda:

- No servers to manage
- Continuous scaling
- Millisecond billing
- Integrates with almost all other AWS services

Application Services

Amazon Simple Queue Service (SQS)

- SQS offers a reliable, highly-scalable, hosted queue for storing messages in transit between computers
- SQS is used for distributed/decoupled applications
- SQS uses a message-oriented API
- SQS uses pull based (polling) not push based

Application Services

Amazon MQ

- Message broker service
- Similar to Amazon SQS
- Based on Apache Active MQ and RabbitMQ
- Used when customers require industry standard APIs and protocols
- Useful when migrating existing queue-based applications into the cloud

Application Services

Amazon Simple Notification Service (SNS)

- Publisher / subscriber model
- Amazon SNS is used for building and integrating loosely coupled, distributed applications
- Provides instantaneous, push-based delivery (no polling)
- Uses simple APIs and easy integration with applications
- Offered under an inexpensive, pay-as-you-go model with no up-front costs

Application Services

AWS Step Functions

- AWS Step Functions makes it easy to coordinate the components of distributed applications as a series of steps in a visual workflow
- You can quickly build and run state machines to execute the steps of your application in a reliable and scalable fashion

Application Services

Amazon Simple Workflow Service (SWF)

- Coordinate work across distributed application components
- Create distributed asynchronous systems as workflows
- Best suited for human-enabled workflows like an order fulfilment system or for procedural requests
- AWS recommends that for new applications customers consider Step Functions instead of SWF

Application Services

Amazon EventBridge

- Serverless event bus
- Used for building event-driven architectures
- Ingests data and routes it to target AWS services

Application Services

Amazon API Gateway

- Publish APIs on AWS
- Create RESTful and WebSocket APIs
- Fully managed service
- Forward connections to AWS services and on-premises applications

Amazon VPC, Networking, and Hybrid

Saravanan Chandrasekar

Amazon VPC, Networking, and Hybrid

Amazon Virtual Private Cloud (VPC)

- A VPC is a virtual network dedicated to your AWS account
- Analogous to having your own DC inside AWS
- It is logically isolated from other virtual networks in the AWS Cloud
- Provides complete control over the virtual networking environment
- You can launch your AWS resources, such as Amazon EC2 instances, into your VPC

Amazon VPC, Networking, and Hybrid

Amazon Virtual Private Cloud (VPC)

- When you create a VPC, you must specify a range of IPv4 addresses for the VPC in the form of a Classless Inter-Domain Routing (CIDR) block; for example, 10.0.0.0/16
- A VPC spans all the Availability Zones in the region
- You have full control over who has access to the AWS resources inside your VPC
- By default you can create up to 5 VPCs per region
- A default VPC is created in each region with a subnet in each AZ

Amazon VPC, Networking, and Hybrid

Security Groups

- Firewall for EC2 instances
- Operate at the instance level
- Support allow rules only
- Stateful

Amazon VPC, Networking, and Hybrid

Network Access Control Lists (ACLs)

- Firewall at the subnet level
- Support allow and deny rules
- Stateless
- Process rules in order

Amazon VPC, Networking, and Hybrid

IP addresses

- Public IP is dynamic and lost when instance is stopped
- Public IPs cannot be moved between instances
- Private IPs are attached to all EC2 instances
- Private IPs are retained when the instance is stopped
- Elastic IPs are static public addresses
- Elastic IPs are retained when the instance is stopped
- Elastic IPs can be moved between instances
- Elastic IPs are chargeable if not used

Amazon VPC, Networking, and Hybrid

NAT Instances and Gateways

- Used for accessing the internet from private subnets
- Deployed in public subnets
- Must update the route table in private subnets
- NAT instances are managed by you
- NAT gateways are managed by AWS

Amazon VPC, Networking, and Hybrid

VPC Peering

- Used to route between VPCs using private IP addresses

AWS Managed VPN

- Virtual private network (VPN) connection between onpremises sites and AWS
- Uses the public Internet

AWS Direct Connect

- Private connection from on-premises to AWS
- Avoids the public Internet

Amazon VPC, Networking, and Hybrid

AWS Transit Gateway

- Connects VPCs and on-premises networks through a central hub
- Simplifies network configuration

AWS Outposts

- Deploy AWS infrastructure on-premises and connect AWS services
- Can extend a VPC into the on-premises environment
- Supports several AWS services

Deployment and Automation

Saravanan Chandrasekar

Deployment and Automation

Amazon CloudFront

- CloudFront is a content delivery network (CDN) that allows you to store (cache) your content at “edge locations” located around the world
- This allows customers to access content more quickly and provides security against DDoS attacks
- CloudFront can be used for data, videos, applications, and APIs
- CloudFront reduces latency for global users

Deployment and Automation

AWS Global Accelerator

- Routes connections to application endpoints (EC2/ELB) in multiple Regions
- Improves the availability and performance of applications with local or global users
- Uses the AWS global network to optimize the path from users to applications, improving the performance of TCP and UDP traffic

Deployment and Automation

AWS Global Accelerator vs CloudFront

- Both use the AWS global network and edge locations
- CloudFront improves performance for cacheable content and dynamic content
- GA improves performance for a wide range of applications over TCP and UDP
- GA proxies connections to applications in one or more AWS Regions
- GA provides failover between AWS Regions

Deployment and Automation

AWS CloudFormation

- Infrastructure is provisioned consistently, with fewer mistakes (human error)
- Less time and effort than configuring resources manually
- Free to use (you're only charged for the resources provisioned)
- A template is a YAML or JSON template used to describe the end-state of the infrastructure you are either provisioning or changing
- CloudFormation creates a Stack based on the template
- Can easily rollback and delete the entire stack as well

Deployment and Automation

AWS Cloud Development Kit (CDK)

- Open-source software development framework to define your cloud application resources using familiar programming languages
- Preconfigures cloud resources with proven defaults using constructs
- Provisions your resources using AWS CloudFormation
- Enables you to model application infrastructure using TypeScript, Python, Java, and .NET
- Use existing IDE, testing tools, and workflow patterns

Deployment and Automation

AWS Elastic Beanstalk

- Managed service for web applications on Amazon EC2 instances and Docker containers
- Deploys an environment that can include Auto Scaling, Elastic Load Balancing and databases
- Considered a Platform as a Service (PaaS) solution
- Allows full control of the underlying resources
- Code is deployed using a ZIP file, WAR file or Git repository

Deployment and Automation

AWS X-Ray

- AWS X-Ray helps developers analyze and debug production, distributed applications, such as those built using a microservices architecture

AWS OpsWorks

- AWS OpsWorks is a configuration management service that provides managed instances of Chef and Puppet
- Updates include patching, updating, backup, configuration and compliance management

Databases and Analytics

Saravanan Chandrasekar

Databases and Analytics

Amazon Relational Database Service (RDS)

- RDS uses EC2 instances, so you must choose an instance family/type
- Relational databases are known as Structured Query Language (SQL) databases
- RDS is an Online Transaction Processing (OLTP) type of database
- Easy to setup, highly available, fault tolerant, and scalable
- Common use cases include online stores and banking systems
- Can encrypt your Amazon RDS instances and snapshots at rest
- Encryption uses AWS Key Management Service (KMS)

Databases and Analytics

Amazon Relational Database Service (RDS)

- Amazon RDS supports the following database engines:
- SQL Server, Oracle, MySQL Server, PostgreSQL, Aurora, MariaDB
- Scales up by increasing instance size (compute and storage)
- Read replicas option for read heavy workloads (scales out for reads/queries only)
- Disaster recovery with Multi-AZ option

Databases and Analytics

Amazon Aurora

- Amazon Aurora is an AWS database offering in the RDS family
- Amazon Aurora is a MySQL and PostgreSQLcompatible relational database built for the cloud
- Amazon Aurora features a distributed, fault-tolerant, selfhealing storage system that auto-scales up to 128TB per database instance

Databases and Analytics

Amazon DynamoDB

- Fully managed NoSQL database service
- Key/value store and document store
- It is a non-relational, key-value type of database
- Fully serverless service
- Push button scaling

Databases and Analytics

Amazon DynamoDB features**DynamoDB Feature Benefit**

- Serverless - Fully managed, fault tolerant, service
- Highly available - 99.99% availability SLA – 99.999% for Global Tables!
- NoSQL type of database with Name / Value structure - Flexible schema, good for when data is not well structured or unpredictable
- Horizontal scaling - Seamless scalability to any scale with push button scaling or Auto Scaling
- DynamoDB Accelerator (DAX) - Fully managed in-memory cache for DynamoDB that increases performance (microsecond latency)
- Backup - Point-in-time recovery down to the second in last 35 days; On-demand backup and restore
- Global Tables - Fully managed multi-region, multi-master solution

Databases and Analytics

Amazon RedShift

- RedShift is a SQL based data warehouse used for analytics applications
- RedShift is a relational database that is used for Online

Analytics Processing (OLAP) use cases

- RedShift uses Amazon EC2 instances, so you must choose an instance family/type
- RedShift always keeps three copies of your data
- RedShift provides continuous/incremental backups

Databases and Analytics

Amazon EMR

- Managed cluster platform that simplifies running big data frameworks including Apache Hadoop and Apache Spark
- Used for processing data for analytics and business intelligence
- Can also be used for transforming and moving large amounts of data
- Performs extract, transform, and load (ETL) functions

Databases and Analytics

Amazon ElastiCache

- Fully managed implementations Redis and Memcached
- ElastiCache is a key/value store
- In-memory database offering high performance and low latency
- Can be put in front of databases such as RDS and DynamoDB

Databases and Analytics

Amazon Athena

- Athena queries data in S3 using SQL
- Can be connected to other data sources with Lambda
- Data can be in CSV, TSV, JSON, Parquet and ORC formats
- Uses a managed Data Catalog (AWS Glue) to store information and schemas about the databases and tables

Databases and Analytics

AWS Glue

- Fully managed extract, transform and load (ETL) service
- Used for preparing data for analytics
- AWS Glue runs the ETL jobs on a fully managed, scale-out Apache Spark environment
- Works with data lakes (e.g. data on S3), data warehouses (including RedShift), and data stores (including RDS or EC2 databases)

Databases and Analytics

Amazon Kinesis Data Streams

- • Producers send data which is stored in shards for up to 7 days
- • Consumers process the data and save to another service

Amazon Kinesis Data Firehose

- • No shards, completely automated and elastically scalable
- • Saves data directly to another service such as S3, Splunk, RedShift, or Elasticsearch

Amazon Kinesis Data Analytics

- • Provides real-time SQL processing for streaming data

Databases and Analytics

AWS Data Pipeline	• Processes and moves data between different AWS compute and storage services	• Save results to services including S3, RDS, DynamoDB, and EMR	Amazon QuickSight
• Business intelligence (BI) service	• Create and publish interactive BI dashboards for Machine Learning-powered insights	Amazon Neptune	• Fully managed graph database service

Databases and Analytics

Amazon DocumentDB

- • Fully managed document database service (non-relational)
- • Supports MongoDB workloads
- • Queries and indexes JSON data

Amazon QLDB

- • Fully managed ledger database for immutable change history
- • Provides cryptographically verifiable transaction logging

Amazon Managed Blockchain

- • Fully managed service for joining public and private networks
- using Hyperledger Fabric and Ethereum

Management and Governance

Management and Governance

AWS Organizations

- Allows you to consolidate multiple AWS accounts into an organization that you create and centrally manage
- Available in two feature sets:
    - Consolidated Billing
    - All features
- Includes root accounts and organizational units
- Policies are applied to root accounts or OUs
- Consolidated billing includes:
    - Paying Account – independent and cannot access resources of other accounts
    - Linked Accounts – all linked accounts are independent

Management and Governance

AWS Control Tower

- Simplifies the process of creating multi-account environments
- Sets up governance, compliance, and security guardrails for you
- Integrates with other services and features to setup the environment for you including:
    - AWS Organizations, SCPs, OUs, AWS Config, AWS CloudTrail, Amazon S3, Amazon SNS, AWS CloudFormation, AWS Service Catalog, AWS Single Sign-On (SSO)

Management and Governance

AWS Systems Manager

- Manages many AWS resources including Amazon EC2, Amazon S3, Amazon RDS etc.
- Systems Manager Components:
- Automation – uses documents to run automations
- Run Command – run commands on EC2 instances
- Inventory – gather inventory information
- Patch Manager – manage patching schedules and installation
- Session Manager – connect securely without SSH or RDP
- Parameter Store – store secrets and configuration data securely

Management and Governance

AWS Service Catalog

- Allows organizations to create and manage catalogs of IT services that are approved for use on AWS
- Allows you to centrally manage commonly deployed IT services
- IT services can include virtual machine images, servers, software, and databases and multi-tier application architectures
- Enables users to quickly deploy only the approved IT services they need

Management and Governance

AWS Config

- Fully-managed service for compliance management
- Helps with compliance auditing, security analysis, resource change tracking and troubleshooting

Management and Governance

Trusted Advisor

- Online resource that helps to reduce cost, increase performance and improve security by optimizing your AWS environment
- Provides real time guidance to help you provision your resources following best practices
- Advises you on Cost Optimization, Performance, Security, and Fault Tolerance

Management and Governance

AWS Personal Health Dashboard

- Provides alerts and remediation guidance when AWS is experiencing events that may impact you
- Gives you a personalized view into the performance and availability of the AWS services underlying your AWS resources
- Also provides proactive notification to help you plan for scheduled activities

Management and Governance

Service Health Dashboard

- Shows you current status of AWS services
- Not personalized

AWS Cloud Security and Identity

Saravanan Chandrasekar

AWS Cloud Security and Identity

AWS Directory Services

Directory Service	Service Description	Use Case
AWS Directory Service for Microsoft Active Directory	AWS-managed full Microsoft AD running on Windows Server 2012 R2	Enterprises that want hosted Microsoft Active Directory
AD Connector	Allows on-premises users to log into AWS services with their existing AD credentials	Single sign-on for on-premises employees
Simple AD	Low scale, low cost, AD implementation based on Samba	Simple user directory, or you need LDAP compatibility

AWS Cloud Security and Identity

AWS Systems Manager Parameter Store

- Provides secure, hierarchical storage for configuration data management and secrets management
- You can store data such as passwords, database strings, and license codes as parameter values
- You can store values as plaintext (unencrypted data) or ciphertext (encrypted data)
- You can then reference values by using the unique name that you specified when you created the parameter

AWS Cloud Security and Identity

AWS Secrets Manager

- Similar to Parameter Store
- Allows native and automatic rotation of keys
- Fine-grained permissions
- Central auditing for secret rotation

AWS Cloud Security and Identity

AWS Certificate Manager (ACM)

- Create, store and renew SSL/TLS X.509 certificates
- Single domains, multiple domain names and wildcards
- Integrates with several AWS services including:
  - Elastic Load Balancing
  - Amazon CloudFront
  - AWS Elastic Beanstalk
  - AWS Nitro Enclaves
  - AWS CloudFormation

AWS Cloud Security and Identity

AWS Key Management Service (KMS)

- Used for creating and managing encryption keys
- Gives you centralized control over the encryption keys used to protect your data
- KMS is integrated with most other AWS services
- Easy to encrypt the data you store in these services with encryption keys you control

AWS Cloud Security and Identity

AWS CloudHSM

- Cloud-based hardware security module (HSM)
- Generate and use your own encryption keys on the AWS Cloud
- Manage your own encryption keys using FIPS 140-2 Level 3 validated HSMs
- CloudHSM runs in your VPC

AWS Cloud Security and Identity

AWS CloudTrail

- CloudTrail logs API activity for auditing
- By default, management events are logged and retained for 90 days
- A CloudTrail Trail logs any events to S3 for indefinite retention
- Trail can be within Region or all Regions
- CloudWatch Events can be triggered based on API calls in CloudTrail
- Events can be streamed to CloudWatch Logs

AWS Cloud Security and Identity

VPC Flow Logs

- Flow Logs capture information about the IP traffic going to and from network interfaces in a VPC
- Flow log data is stored using Amazon CloudWatch Logs
- Flow logs can be created at the following levels:
  - VPC
  - Subnet
  - Network interface

AWS Cloud Security and Identity

Elastic Load Balancing Access Logs

- Capture detailed information about requests sent to the load balancer
- Use to analyze traffic patterns and troubleshoot issues
- Can identify requester, IP, request type etc.
- Can be optionally stored and retained in S3

S3 Access Logs

- Provides detailed records for the requests that are made to a bucket
- Details include the requester, bucket name, request time, request action, response status, and error code (if applicable)
- Disabled by default

AWS Cloud Security and Identity

Amazon Detective

- Analyze, investigate, and quickly identify the root cause of potential security issues or suspicious activities
- Automatically collects data from AWS resources
- Uses machine learning, statistical analysis, and graph theory
- Data sources include VPC Flow Logs, CloudTrail, and GuardDuty

AWS Cloud Security and Identity

AWS GuardDuty

- Intelligent threat detection service
- Detects account compromise, instance compromise, malicious reconnaissance, and bucket compromise
- Continuous monitoring for events across:
  - AWS CloudTrail Management Events
  - AWS CloudTrail S3 Data Events
  - Amazon VPC Flow Logs
  - DNS Logs

AWS Cloud Security and Identity

Amazon Macie

- • Macie is a fully managed data security and data privacy service
- • Uses machine learning and pattern matching to discover, monitor, and help you protect your sensitive data on Amazon

S3

- • Macie enables security compliance and preventive security

AWS Cloud Security and Identity

AWS WAF

- AWS WAF is a web application firewall
- Create rules that block common web exploits like SQL injection and cross site scripting
- The rules are known as Web ACLs

AWS Shield

- AWS Shield is a managed Distributed Denial of Service (DDoS) protection service
- Safeguards web application running on AWS with always-on detection and automatic inline mitigations

AWS Cloud Security and Identity

AWS Artifact

- AWS Artifact provides on-demand access to AWS' security and compliance reports and select online agreements
- Reports available in AWS Artifact include:
    - Service Organization Control (SOC) reports
    - Payment Card Industry (PCI) reports

AWS Cloud Security and Identity

AWS Security Hub

- Provides a comprehensive view of security alerts and security posture across AWS accounts
- Aggregates, organizes, and prioritizes security alerts, or findings, from multiple AWS services

AWS Security Bulletins

- Security and privacy events affecting AWS services are published (also has an RSS feed)

AWS Cloud Security and Identity

AWS Trust & Safety Team

- Contact the AWS Trust & Safety team if AWS resources are being used for:
  - Spam
  - Port scanning
  - Denial-of-service attacks
  - Intrusion attempts
  - Hosting of objectionable or copyrighted content
  - Distributing malware

AWS Cloud Security and Identity

Penetration Testing

- Penetration testing is the practice of testing one's own application's security for vulnerabilities by simulating an attack
- AWS allows penetration testing without prior approval for 8 AWS services

Architecting for the Cloud

Saravanan Chandrasekar

AWS Well-Architected Framework

- Helps you understand the pros and cons of decisions you make while building systems on AWS
- Based on 5 pillars:

Operational Excellence Pillar

- Support development and run workloads effectively
- Gain insight into workload operations
- Continuously improve processes and procedures to deliver business value

AWS Well-Architected Framework

Best practices for operational excellence:

- Perform operations as code
- Make frequent, small, reversible changes
- Refine operations procedures frequently
- Anticipate failure
- Learn from all operational failures

AWS Well-Architected Framework

Security Pillar

- Protect data, systems, and assets to take advantage of cloud technologies to improve your security
- Best practices for security:
    - Implement a strong identity foundation
    - Enable traceability
    - Apply security at all layers
    - Automate security best practices
    - Protect data in transit and at rest
    - Keep people away from data
    - Prepare for security events

AWS Well-Architected Framework

Reliability Pillar

- Ensuring a workload can perform its intended function correctly and consistently when it's expected to
- This includes the ability to operate and test the workload through its total lifecycle

Best practices for reliability:

- Automatically recover from failure
- Test recovery procedures
- Scale horizontally to increase aggregate workload availability
- Stop guessing capacity
- Manage change in automation

AWS Well-Architected Framework

Performance Efficiency Pillar

- The ability to use computing resources efficiently to meet system requirements, and to maintain that efficiency as demand changes and technologies evolve
- Best practices for performance efficiency:
    - Democratize advanced technologies
    - Go global in minutes
    - Use serverless architectures
    - Experiment more often
    - Consider mechanical sympathy

AWS Well-Architected Framework

Cost Optimization Pillar

- The ability to run systems to deliver business value at the lowest price point
- Best practices for cost optimization:
  - Implement Cloud Financial Management
  - Adopt a consumption model
  - Measure overall efficiency
  - Stop spending money on undifferentiated heavy lifting
  - Analyze and attribute expenditure

Accounts, Billing and Support

Accounts, Billing and Support

Pay-as-you-go

- Easily adapt to changing business needs
- Improved responsiveness to change
- Adapt based on needs, not forecasts
- Reduce risk over over positioning of missing capacity

Save when you reserve

- Invest in reserved capacity (e.g. RDS and EC2)
- Save up to 75% compared to on-demand (pay-as-you-go)
- The more you pay upfront the greater the discount

Accounts, Billing and Support

Pay less by using more

- Pay less using volume-based discounts
- Tiered pricing means the more you use the lower the unit pricing

Accounts, Billing and Support

On-Demand Standard rate - no discount; no commitments; dev/test, short-term, or unpredictable workloads	**Reserved** 1 or 3-year commitment; up to 75% discount; steady-state, predictable workloads and reserved capacity
Spot Instances Bid for unused capacity; up to 90% discount; can be terminated at any time; workloads with flexible start and end times	**Dedicated Instances** Physical isolation at the host hardware level from instances belonging to other customers; pay per instance
Dedicated Hosts Physical server dedicated for your use; Socket/core visibility, host affinity; pay per host; workloads with server-bound software	**Savings Plans** Commitment to a consistent amount of usage (EC2 + Fargate + Lambda); Pay by $/hour; 1 or 3-year commitment

Accounts, Billing and Support

Amazon S3 Pricing

- Storage class – e.g. Standard or IA
- Storage quantity – data volume stored in your buckets on a per GB basis
- Number of requests – the number and type of requests

  Lifecycle transitions requests – moving data between storage classes
- Data transfer – data transferred out of an S3 region is charged

Accounts, Billing and Support

- Three options for access to archives, listed in the table below:

	Expedited	Standard	Bulk
Data access time	1-5 minutes	3-5 hours	5-12 hours
Data retrievals	$0.03 per GB	$0.01 per GB	$0.0025 per GB
Retrieval requests	On-Demand: $0.01 per request Provisioned: $100 per Provisioned Capacity Unit	$0.050 per 1,000 requests	$0.025 per 1,000 requests

Accounts, Billing and Support

Amazon EBS Pricing

- Volumes – volume storage for all EBS volumes type is charged by the amount of GB provisioned per month
- Snapshots – based on the amount of space consumed by snapshots in S3

Amazon RDS Pricing

- Clock hours of server uptime – amount of time the DB instance is running
- Database characteristics – e.g. database engine, size and memory class
- Database purchase type – e.g. On-Demand, Reserved

Accounts, Billing and Support

Amazon DynamoDB

Charged for reading, writing, and storing data

On-demand capacity mode

- Charged for reads and writes
- No need to specify how much capacity is required
- Good for unpredictable workloads

Provisioned capacity mode

- Specify number of reads and writes per second
- Can use Auto Scaling
- Good for predictable workloads
- Consistent traffic or gradual changes

Accounts, Billing and Support

AWS Lambda

- Number of requests
- Duration of request – rounded up to the nearest millisecond
- Price is dependent on the amount of memory allocated to the function

Accounts, Billing and Support

AWS Organizations

- Consolidated billing has the following benefits:
- One bill – You get one bill for multiple accounts
- Easy tracking – You can track the charges across multiple accounts and download the combined cost and usage data
- Combined usage – You can combine the usage across all accounts in the organization to share the volume pricing discounts and Reserved Instance discounts
- No extra fee – Consolidated billing is offered at no additional cost

Accounts, Billing and Support

AWS Budgets

- Set Custom Budgets - set custom usage and reservation budgets
- Configure Alerts – receive alerts when you exceed or are forecast to exceed your alert thresholds
- Integrated with other AWS services – Includes Cost Explorer Chatbot, and Service Catalog

Accounts, Billing and Support

AWS Cost Explorer

- Free tool that allows you to view charts of your costs
- Cost Explorer can be used to discover patterns in how much you spend on AWS resources over time and to identify cost problem areas

AWS Cost & Usage Report

- Publish AWS billing reports to an Amazon S3 bucket
- Reports break down costs by:
    - Hour, day, month, product, product resource, tags

Accounts, Billing and Support

AWS Price List API

- Query the prices of AWS services
- Price List Service API
- AWS Price List API

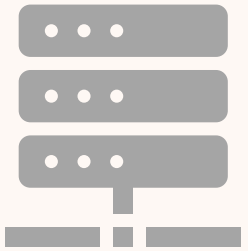
Migration, Machine Learning and More

Saravanan Chandrasekar

Migration, Machine Learning and More

AWS Migration Hub

- Provides a single location to track the progress of application migrations across multiple AWS and partner solutions



AWS Database Migration Service (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

Migration, Machine Learning and More

AWS Server Migration Service (SMS)

- • Migrates servers and virtual machines to Amazon EC2
- • Agentless service which makes it easier and faster for you to migrate thousands of on-premises workloads to AWS
- • Automate, schedule, and track incremental replications of live server volumes

AWS DataSync

- • Online data transfer service
- • Transfer data between on-premises and AWS storage services

Migration, Machine Learning and More

Snowball Family	Snowball Edge Compute Optimized	Snowball Edge Storage Optimized	Snowcone
• • AWS Snowball and Snowmobile are used for migrating large volumes of data to AWS	• • Provides block and object storage and optional GPU • • Edge computing use cases	• • Provides block storage and Amazon S3-compatible object storage • • Use for local storage and large-scale data transfer	• • Small device used for edge computing, storage and data transfer • • Can transfer data offline or online with AWS DataSync agent

Migration, Machine Learning and More

Snowball Family

- Uses a secure storage device for physical transportation
- Snowball (80TB) (50TB) “petabyte scale”
- Snowball Edge (100TB) “petabyte scale”
- Snowmobile – “exabyte scale” with up to 100PB per Snowmobile

Migration, Machine Learning and More

AWS Rekognition

- • Add image and video analysis to your applications
- • Identify objects, people, text, scenes, and activities in images and videos

Amazon Transcribe

- • Add speech to text capabilities to applications
- • Recorded speech can be converted to text before it can be used in applications

Migration, Machine Learning and More

Amazon Translate

- • Neural machine translation service that delivers fast, high-quality, and affordable language translation
- • Localize content such as websites and applications for your diverse users

Amazon Sagemaker

- • Helps data scientists and developers to prepare, build, train, and deploy high-quality machine learning (ML) models

Migration, Machine Learning and More

Amazon Comprehend

- • Natural-language processing (NLP) service
- • Uses machine learning to uncover information in unstructured data

Amazon Lex

- • Conversational AI for Chatbots
- • Build conversational interfaces into any application using voice and text

Amazon Polly

- • Turns text into lifelike speech
- • Create applications that talk, and build entirely new categories of speech-enabled products

Migration, Machine Learning and More

Amazon Workspaces

- Managed Desktop-as-a-Service (DaaS) solution
- Provision either Windows or Linux desktops

AWS AppStream 2.0

- Fully managed non-persistent application streaming service
- Alternative to popular products such as Citrix XenApp

AWS Worklink

- Provides secure, one-click access to your internal websites and web apps using mobile phone browsers
- Does not require VPN client or App

Migration, Machine Learning and More

AWS WorkDocs

- Fully managed, secure content creation, storage, and collaboration service
- Create, edit, and share content that’s centrally stored on AWS

AWS IoT Core

- Lets you connect IoT devices to the AWS cloud without the need to provision or manage servers