WHIZLABS

AWS Certified: Solution Architect Associate

Cheat Sheet

Quick Bytes for you before the exam!

*The information provided in the Cheat Sheet is for educational purposes only; created in our efforts to help aspirants prepare for the **Exam AWS Solution Architect Associate certification**. Though references have been taken from **AWS documentation**, it's not intended as a substitute for the official docs. The document can be reused, reproduced, and printed in any form; ensure that appropriate sources are credited and required permissions are received.*

Index

—-Back to Index—

—-Back to Index—

—-Back to Index—

Compute

AWS EC2

What is AWS EC2?

EC2 (Elastic Compute Cloud) is a scalable virtual machine in the cloud.

- Automatically scales instances based on traffic.
- Eliminates hardware investment.
- Allows launching multiple servers with full control over security, networking, and storage.

Overview of Key Features in Amazon EC2

Feature	Description
Instance Type	Provides a range of instance types for various use cases. Determines the processor and memory configuration of your EC2 instance.
EBS Volume	- Stands for Elastic Block Storage. - Block-level storage assigned to a single EC2 instance. - Persists independently from running EC2 instances. **Types:** - General Purpose (SSD) - Cold Hard Disk Drive - Magnetic - Provisioned IOPS (SSD) - Throughput Optimized Hard Disk Drive
Instance Store	Ephemeral block-level storage for EC2 instances. Used for faster processing and temporary storage of applications.
AMI	- Stands for Amazon Machine Image. - Defines the OS, dependencies, libraries, and data for EC2 instances. - Enables launching multiple instances with the same configuration.
Security Group	- Virtual firewall for EC2 instances. - Controls ports and traffic. - Active at the instance level; Network ACLs operate at the subnet level. - Allows rules only, cannot deny. - Stateful design. - Outbound traffic allowed by default; inbound rules require definition.
Key Pair	- A set of security credentials (public and private keys) for identity verification when connecting to an instance. - Public key attached to the instance; private key remains with the user. - Access is granted when keys match. - Keep the private key secure.
Pricing	Different pricing options: - On-Demand - Reserved Instances - Savings Plan - Spot Instances
Tags	- Key-value pairs assigned to AWS resources. - Help identify and organize resources effectively.

AWS Batch

What is AWS Batch?

AWS Batch enables running thousands of computing jobs on AWS.

- Dynamically manages optimal compute resources (CPU, memory) based on job volume.
- Focuses on applications like shell scripts, Linux code, or Java programs.
- Supports execution on EC2 (including Spot instances) and AWS Fargate.

Key Concepts of AWS Batch

Component	Description
Jobs	Fundamental applications running on Amazon EC2 machines in containerized form.
Job Definitions	Define how the job should run, including IAM roles, vCPU requirements, and container properties.
Job Queues	Hold jobs until they are scheduled for execution.
Compute Environments	Linked to job queues and contain EC2 instances to run containerized applications. - **Managed**: AWS handles EC2 setup based on min/max vCPU and instance type. - **Unmanaged**: User manages their own ECS agent.
Scheduler	Manages the execution of jobs in the queue based on time and dependencies.

Best Practices

- **Use Fargate**: Ideal for running applications without managing EC2 infrastructure; AWS Batch handles it.
- **Use EC2**: Suitable for large-scale workloads requiring control over machine specifications (e.g., memory, CPU, GPU).
- Fargate jobs start faster as there's no lag in scaling out, unlike EC2, which may take time to launch instances.

Use Cases

- **Stock Markets & Trading**: Processes large-scale data daily for quick analytics and decision-making to drive growth.
- **Media & Entertainment**: Handles massive audio, video, and photo processing workloads, moving them to containers on AWS Batch.

Pricing

- AWS Batch is free; you only pay for the underlying resources like EC2 and Fargate.

AWS Elastic Beanstalk

What is AWS Elastic Beanstalk?

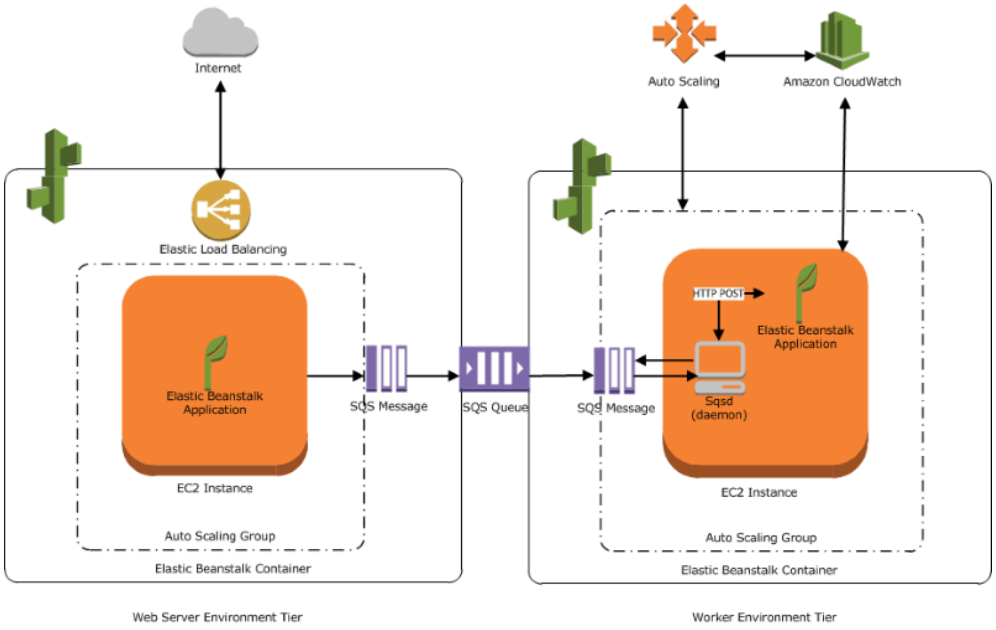
AWS Elastic Beanstalk is a compute service for deploying and scaling applications in popular programming languages. It allows developers to focus on code without managing infrastructure. Beanstalk offers:

- A quick and simple way to deploy applications.
- A user-friendly dashboard for application monitoring.
- Flexibility to select AWS resources like EC2 instances and pricing options based on your needs.

AWS Elastic Beanstalk supports two types of Environment:

Environment Types:

1. **Web Tier:** Handles HTTP/HTTPS requests using ELB and Auto Scaling.
2. **Worker Tier:** Processes background tasks like database cleanup and report generation via a Daemon that pulls tasks from SQS.



Key Components	Description
Elastic Load Balancer (ELB)	Distributes incoming traffic among EC2 instances in the Auto Scaling Group.
Auto Scaling Group	Dynamically adds/removes EC2 instances based on application load.
Host Manager	Manages logs, monitoring, and events on each EC2 instance.

AWS Lambda

What is AWS Lambda?

AWS Lambda is a **serverless compute service** that runs your code without the need for provisioning servers. It automatically scales with request count and follows a **pay-per-use model**, meaning no charges when the code isn't running. Lambda executes code for any application or backend service, triggered by events like updates in DynamoDB or S3 changes, or HTTP requests via API Gateway.

What is Serverless Computing?

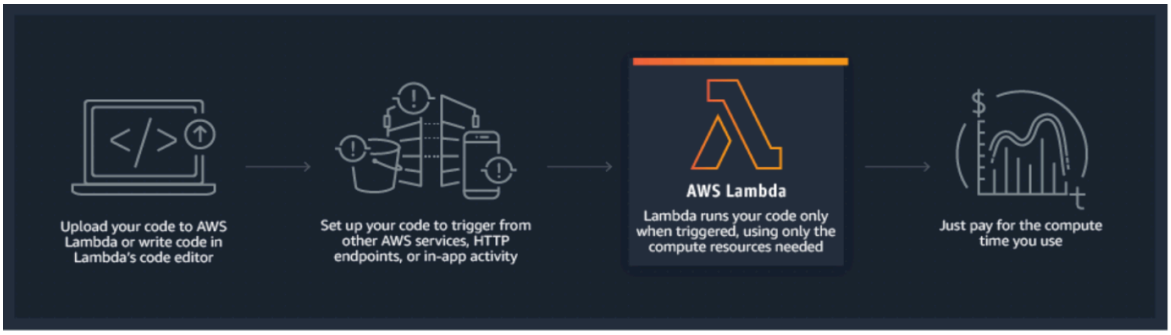
Serverless computing provides backend services on a **pay-per-use basis** without worrying about underlying infrastructure. Servers exist but are managed by the cloud vendor.

When to Use Lambda

- Ideal when you focus solely on your code while AWS manages compute resources like memory, CPU, and network.
- For custom compute management, consider EC2 or Elastic Beanstalk. Lambda abstracts server access and runtime customization.

How AWS Lambda Works

Aspect	Details
Lambda Functions	- Code is uploaded as a zip file or from an S3 bucket. - Functions are monitored via Amazon CloudWatch.
Lambda Layers	- Archive for additional code (e.g., libraries, dependencies, or runtimes). - Allows up to 5 layers per function. - Layers are immutable and can be shared publicly.
Lambda Events	- Entities that trigger functions, such as: - DynamoDB, SQS, SNS, CloudWatch, API Gateway, IoT, Kinesis.
Lambda@Edge	- Runs code closer to users through CloudFront, improving performance and reducing latency.



Supported Languages

- Node.js, Go, Java, Python, Ruby.

Pricing

- Charged based on **number of requests** and **execution duration** (per 100 ms).
- **Free Tier**: 1 million requests/month. 400,000 GB-seconds of compute time/month.

AWS Serverless Application Repository

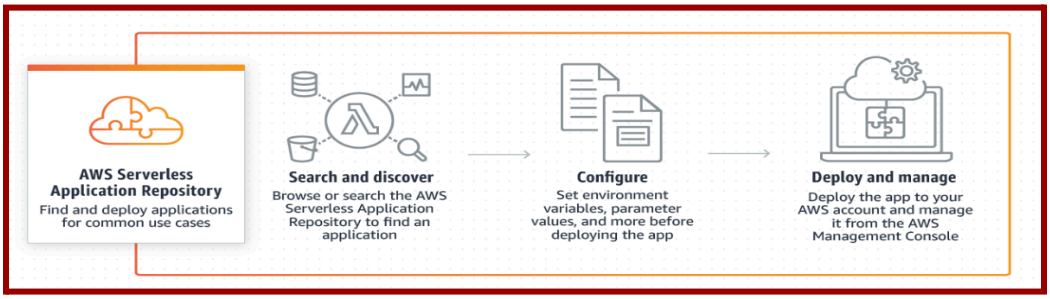
What is AWS Serverless Application Repository?

It is a managed repository for serverless applications. It is used by organizations and independent developers to store and share reusable applications.

Features:

Feature	Details
Applications Available	Alexa Skills, chatbots, IoT, and media processing from various publishers.
Licensing	AWS apps use the MIT license; public apps follow OSI standards.
Security	AWS reviews app permissions for customer clarity.
Application Sharing	Share within AWS Organization accounts; no cross-organization sharing allowed.
Publishing Process	Use AWS SAM for description, package via CLI, publish via CLI, SDK, or console.

See below diagram -



Use Case:

- Used for various AWS Alexa skills and integration with IoT devices.
- Used for chatbots that remove inappropriate messages, images from channels.
- Used in Twitter leadership boards.

Pricing:

- There is no charge for this service itself but you pay for the resources used in the application

AWS Fargate

What is AWS Fargate?

AWS Fargate is a serverless compute service for containers used with Amazon ECS and EKS. It simplifies running containers by eliminating the need to manage virtual machines like EC2.

Key Features	Description
Serverless Containers	Runs containers by specifying CPU, memory, and IAM policies.
Isolation	Fargate tasks have dedicated kernels, memory, CPU, and ENI, ensuring task isolation.
Task Limitations	Supports only specific ECS task definition parameters with some restrictions.
Kubernetes Integration	Schedules Kubernetes pods on Fargate using controllers. Security groups for EKS pods are unsupported.
Storage Support	- Amazon EFS for persistent storage. - Ephemeral storage for nonpersistent needs.

Benefits of AWS Fargate:

- **Focus on Application Development**: Fargate enables users to focus on building and operating applications rather than managing servers, security, scaling, and patching.
- **Automatic Scaling**: It automatically adjusts the compute environment to meet the container's resource requirements.
- **Built-in Integrations**: Fargate integrates seamlessly with other AWS services, including Amazon CloudWatch Container Insights for monitoring.

Pricing Details:

- **Cost Based on vCPU and Memory Usage**: Charges are incurred based on the amount of vCPU and memory consumed by the containerized applications.
- **Savings Plans**: Fargate's Savings Plans offer up to 50% savings in exchange for a one- or three-year long-term commitment.
- **Additional Charges**: Extra charges may apply if containers are used in conjunction with other AWS services.

Amazon Elastic Kubernetes Service(EKS)

What is Amazon Elastic Kubernetes Service (EKS)?

Amazon EKS is a fully managed service that allows users to deploy, manage, and scale Kubernetes applications on AWS or on-premises. It supports standard Kubernetes applications without the need for code modification.

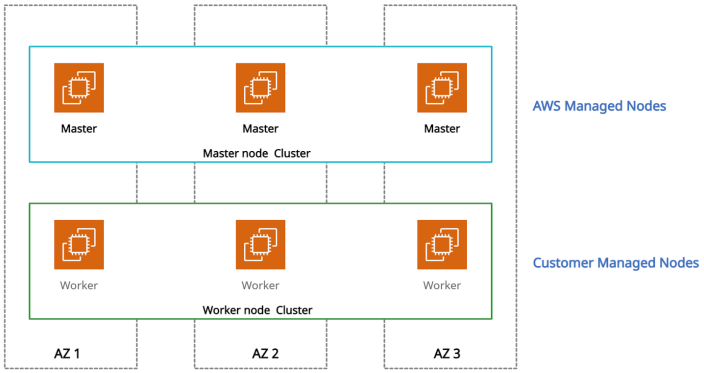
Key Components	
Amazon EKS Control Plane	Runs Kubernetes software (API server, etcd) with high availability across multiple AZs.
Amazon EKS Nodes	Worker nodes where Kubernetes pods run.
Cluster Creation Methods	
eksctl	Command-line tool for creating and managing clusters.
AWS Management Console & AWS CLI	Alternative methods for cluster creation.
Node Scheduling Methods	
Self-managed nodes	EC2 instances in Auto Scaling groups.
Amazon EKS Managed Node Groups	Automates node provisioning and lifecycle management.
AWS Fargate	Runs Kubernetes pods on Fargate without managing servers.
AWS Service Integrations	
Images	Amazon ECR for container images.
Load Balancing	AWS ELB for load balancing.
Authentication	AWS IAM for authentication.
Networking	Amazon VPC for networking and isolation.

Use Cases:

- **Hybrid Environments**: Manage Kubernetes clusters across on-premises and AWS.
- **Machine Learning**: Use Kubeflow with EC2 GPU instances for ML workflows.
- **Batch Workloads**: Execute Kubernetes jobs across EC2, Fargate, and Spot Instances.

Pricing:

- **EKS Cluster**: $0.10 per hour per cluster.
- **With EC2**: Charges for EC2 resources (e.g., instances, EBS volumes).
- **With Fargate**: Charges for CPU and memory from image download to pod termination.



—-Back to Index—

Amazon Elastic Container Service

What is Amazon Elastic Container Service (Amazon ECS)?

- A regional container orchestration service to execute, stop, and manage containers on a cluster.
- Allows containers to run smoothly across environments by combining code, dependencies, and system libraries.
- Containers are created from Docker images defined by a Dockerfile.
- Task definitions (in JSON format) specify which container images should run across clusters.

Key Concepts

- **ECS Cluster**: A combination of tasks or services running on EC2 instances or AWS Fargate.
- **Task Definition**: Defines the container images and configurations for tasks.
- **Service**: Maintains multiple tasks running simultaneously within a cluster.
- **Task**: Represents a single unit of work based on a task definition.
- **Container Agent**: Runs on ECS instances, manages tasks, and reports resource utilization.

ECS Integrations:

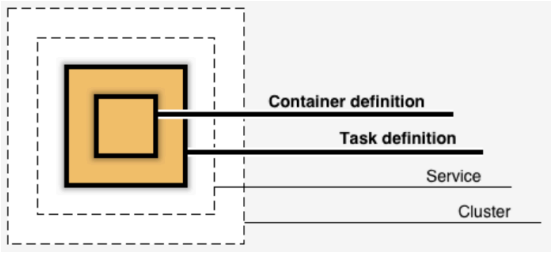
- AWS Identity and Access Management (IAM)
- Amazon EC2 Auto Scaling
- Elastic Load Balancing
- Amazon Elastic Container Registry (ECR)
- AWS CloudFormation
- AWS App Mesh (for traffic control, security, and observability)

Use Cases

- **Microservices**: Decomposes complex applications into smaller, independent services.
- **Batch Jobs**: Docker containers are ideal for processing short-lived, packaged jobs.

Pricing Details

- **Fargate Launch Type Model**: Pay for vCPU and memory resources used.
- **EC2 Launch Type Model**: Pay for AWS resources (EC2 instances, storage) to store and run applications.



ECS Cluster

Amazon Elastic Container Registry

What is Amazon Elastic Container Registry (ECR)?

- A managed service for storing, managing, sharing, and deploying container images and artifacts.
- Integrated with Amazon Elastic Container Service (ECS), Amazon Elastic Kubernetes Service (EKS), AWS Lambda, and AWS Fargate for simplified deployments.

Features

- **Container Storage**: Stores both user-created containers and container software from AWS Marketplace.
- **Integration**: Works seamlessly with ECS, EKS, Lambda, and Fargate for easy deployment.
- **IAM Integration**: AWS Identity and Access Management (IAM) allows resource-level access control for each repository.
- **Public and Private Repositories**: Supports both private (organization-specific) and public container image repositories.
- **Amazon ECR Public Gallery**: A separate portal for accessing all public repositories hosted on Amazon ECR Public.
- **Durability**: Stores images in Amazon S3, offering 99.999999999% (11 9's) data durability.
- **Cross-region and Cross-account Replication**: Supports replication for high availability applications.
- **Encryption**: Images are encrypted at rest using Amazon S3 server-side encryption or customer-managed AWS KMS keys. Data transfers are encrypted via HTTPS.
- **CI/CD Integration**: Works with continuous integration and delivery tools and third-party developer tools.
- **Lifecycle Policies**: Manages container image lifecycles efficiently.

Pricing Details

- **Free Tier**:
  - 500 MB-month of storage for private repositories for the first year.
  - 50 GB-month of storage for public repositories for the first year.
- **Public Repository Data Transfer**:
  - 500 GB per month can be transferred to the internet for free without sign-up.
  - 5 TB per month for free with AWS account sign-up or authentication to ECR.

Storage

Amazon S3

What is Amazon S3?

- **Amazon S3 (Simple Storage Service)** is an object storage service that allows users to store any type of data in a scalable, secure, and low-cost environment.

Basics of S3

- **Object-Based Storage**: Stores files as objects in **buckets**.
- **Buckets**: Folders for objects, with sizes ranging from 0 to 5 TB.
- **Bucket Naming**: Must be globally unique.
- **Upload Success**: Returns HTTP 200 code for successful uploads.
- **Consistency**: Strong consistency for new objects, overwrites, deletes, and list operations.
- **Privacy**: Objects are private by default.

Properties of Amazon S3

- **Versioning**: Keeps multiple versions of objects within the same bucket.
- **Static Website Hosting**: Hosts static websites without requiring server-side technology.
- **Encryption**: Supports encryption at rest using S3 Managed Keys (SSE-S3) or KMS Managed Keys (SSE-KMS).
- **Object Lock**: Prevents version deletion for a defined period, enabled during bucket creation.
- **Transfer Acceleration**: Speeds up file transfer using Amazon CloudFront's edge locations.

Permissions & Management

- **Access Control List (ACL)**: Grants read/write permissions to other AWS accounts.
- **Bucket Policy**: JSON-based access policies for advanced permissions.
- **CORS (Cross-Origin Resource Sharing)**: Allows cross-origin access to S3 resources.

Charges: **Factors Affecting Charges**:

- Storage
- Requests
- Storage Management (Lifecycle Policies)
- Transfer Acceleration
- Data Transfer

Miscellaneous Topics

- **Access Points**: Makes S3 accessible over the internet.
- **Lifecycle Policies**: Transition objects between storage classes based on lifecycle configuration.
- **Replication**: Replicates data across buckets, either within the same region or across different regions.

Storage class	Suitable for	Durability	Availability	Availability Zones	Min. storage days
S3 Standard	accessed data frequently	100%	99.99%	>= 3	None
S3 Standard-IA	accessed data infrequently	100%	99.90%	>= 3	30 days
S3 Intelligent-Tiering	Storage for unknown access patterns	100%	99.90%	>= 3	30 days
S3 One Zone-IA	Non-critical data	100%	99.50%	1	30 days
S3 Glacier	For long term Data Archival. e.g., 3 years – 5 years	100%	99.99%	>= 3	90 days
S3 Glacier Deep Archive	For long term Data Archival. e.g., 3 years – 5 years	100%	99.99%	>= 3	180 days
RRS (Reduced Redundancy Storage)	Frequently accessed for non-critical data but not recommended	99%	99.99%	>= 3	NA

AWS Backup

What is AWS Backup?

AWS Backup is a secure service that automates and manages data backup for AWS cloud resources and on-premises environments.

Features:

Feature	Description
Backup Management	Offers a backup console, APIs, and AWS CLI to manage backups for AWS resources (e.g., instances, databases).
Policy-Based Backup	Automates backup based on policies, tags, and resources.
Scheduled Backup Plans	Automates backups across AWS accounts and regions with customizable policies.
Incremental Backups	Reduces storage costs by performing full backups initially, followed by incremental backups.
Backup Retention Plans	Automatically retains and expires backups to optimize storage costs.
Backup Monitoring	Provides a dashboard in the AWS Backup console to track backup and restore activities.
Encryption	Supports separate encryption keys for multiple AWS resources.
Lifecycle Policies	Automates the transition of backups from Amazon EFS to cold storage.
Cross-Account Backup	Supports backup and restore across AWS accounts and organizations.
Cross-Region Backup	Enables backup and restore to different regions for disaster recovery and business continuity.
Monitoring & Auditing	Integrates with CloudWatch, CloudTrail, and SNS for monitoring, auditing, and notifications.

Use Cases

- **Hybrid Storage Backup**:
  - Uses AWS Storage Gateway volumes for secure, hybrid storage backup, compatible with Amazon EBS for restoring volumes.

Pricing

- **Charges**: Based on backup storage used and the amount of backup data restored.

Amazon EBS - Elastic Block Store

What is Amazon EBS?

Amazon Elastic Block Store (EBS) is a persistent block-level storage service for Amazon EC2 instances. It is AZ-specific, automatically replicated within its AZ for high availability and durability.

Types of EBS:

SSD-backed volumes (Solid State Drive)	**Optimized for transactional workloads** (small and frequent I/O) - IOPS	
Types SSD	**General Purpose SSD- gp2** (1 GiB — 16 TiB) IOPS : 3000 to 20000 Max / Volume	Boot volumes Development /Test Low-latency Apps Virtual Desktops
	Provisioned IOPS SSD (io1) low-latency or high-throughput Consistent IOPS (16,000+ IOPS) Transactional workloads	MongoDB / NoSQL MySQL / RDS Latency Critical Apps
HDD-backed volumes: (Magnetic Drive)	**Low-Cost throughput-intensive workloads** (Not Suitable for Low Latency(IOPS) -- i.e. booting)	
Types HDD	**Throughput Optimized HDD (st1)** Low Cost - Frequently accessed, throughput-intensive & Large-Sequential O/I -- 500 MB/s	Stream Processing Big Data Processing Data Warehouse
	Cold HDD (sc1) Lowest Cost - less frequently accessed data Throughput : 250 MiB/s	Colder Data requires fewer scans per day.

Features:

- **High Performance:** Single-digit millisecond latency.
- **Highly Scalable:** Scales to petabytes.
- **High Availability & Durability:** 99.999% uptime guarantee.
- **Encryption:** Seamless data encryption with AWS KMS.
- **Automated Backups:** Backups via EBS snapshots to S3 using lifecycle policies.
- **Quick Detach/Attach:** Easily detach from one EC2 instance and attach to another.

Pricing: Charges apply for provisioned capacity, snapshots, and data transfer between AZs/Regions.

EBS vs Instance Store:

- **Instance Store:** Ephemeral, temporary storage with high IOPS, data lost on instance stop/crash. Cannot create snapshots.
- **EBS:** Persistent, reliable storage that can be detached/reattached, boots faster, and supports snapshots.

—-Back to Index—

Amazon EFS - Elastic File Storage

What is Amazon EFS?

Amazon Elastic File System (EFS) is a scalable, fully managed file system based on NFS. It offers persistent storage, scales up to petabytes, and supports parallel access from thousands of EC2 instances. EFS is a regional service, automatically replicated across multiple Availability Zones for high availability and durability.

Types of EFS Storage Classes:

Standard Storage	For frequently accessed files.
Infrequent Access Storage (EFS-IA)	For files not accessed every day Cost-Optimized (costs only $0.025/GB-month) Use EFS Lifecycle Management to move the file to EFS IA

EFS Access and Performance Modes:

- **Performance Modes:**
  - *General Purpose:* Low latency, lower throughput.
  - *Max I/O:* High throughput, higher latency.
- **Throughput Modes:**
  - *Bursting (default):* Throughput grows with file system size.
  - *Provisioned:* Fixed throughput capacity.

Features:

- Fully managed, scalable, and durable NFSv4-based system.
- High availability, low latency (SSD-based).
- POSIX compliant.
- Access across AZs, regions, VPCs, and on-premises via Direct Connect/VPN.
- Lifecycle management for better cost-performance.
- Integrated with AWS DataSync, CloudWatch, CloudTrail.
- Supports encryption in transit (TLS) and at rest (KMS).
- Not suitable for boot volumes or highly transactional databases.

Use Cases:

- Mission-critical apps, microservices, containers, media storage, database backups, analytics, and machine learning.

Best Practices:

- Monitor with CloudWatch, track with CloudTrail.
- Leverage IAM for security, separate latency-sensitive workloads.

Pricing: Pay for storage, access mode, and backup storage used.

Amazon FSx for Windows File Server

Key Features:

Feature	Description
Storage Type	Supports HDD and SSD with high throughput and low latency.
Protocol	Uses **Server Message Block (SMB)** for file access.
Access	Connects to **EC2, ECS, WorkSpaces, AppStream 2.0,** and on-premises via **Direct Connect/VPN.**
High Availability	Multi-AZ deployment with active-standby replication.
Failover Management	Automatic synchronous data replication for seamless failover.
Migration	Uses **AWS DataSync** for migrating self-managed file systems.
Authentication	Identity-based authentication via **Microsoft Active Directory (AD).**
Encryption	**Data at Rest:** AWS KMS; **Data in Transit:** SMB Kerberos session keys.

Use cases:

- **Enterprise File Sharing**
  Enables shared access to multiple datasets across multiple users.
- **Application Migration**
  Supports seamless migration of self-managed applications using **AWS DataSync**.
- **Microsoft SQL Server Workloads**
  Handles **SQL Server Failover** and **data replication** for business-critical applications.
- **Media Processing**
  Ensures **low latency** and **high throughput** for media workloads.
- **Analytics & BI**
  Supports **high-performance analytics**, business intelligence, and data processing applications.

Price details:

- Charges are applied monthly based on the storage and throughput capacity used for the file system's file system and backups.
- The cost of storage and throughput depends on the deployment type, either single-AZ or multi-AZ.

Amazon FSx for Lustre

Key Features:

Feature	Description
Scalable Storage	Supports Lustre, a parallel and high-performance file system.
High Performance	Delivers **sub-millisecond latencies**, **millions of IOPS**, and **hundreds of GBps throughput**.
Storage Options	Offers both **SSD** and **HDD** choices.
Amazon S3 Integration	Uses **parallel data-transfer techniques** to process S3 data.
Automatic Updates	Syncs datasets in S3 as files, not objects, ensuring up-to-date data.
Unreplicated Systems	Allows selection of unreplicated file systems for short-term processing.
Machine Learning	Supports **Amazon SageMaker** for ML workloads.

Use cases:

- The workloads which require shared file storage and multiple compute instances use Amazon FSx for Lustre for high throughput and low latency.
- It is also applicable in media and big data workloads to process a large amount of data.

Price details:

- Charges are applied monthly in GB based on the storage capacity used for the file system.
- Backups are stored incrementally, which helps in storage cost savings.

Amazon S3 Glacier

Category	Description
Purpose	Long-term data archiving and backup.
Cost & Durability	Cheapest S3 storage class with 99.999999999% durability.
Data Types	Stores unlimited data (photos, videos, documents, TAR/ZIP files, data lakes, analytics, IoT, ML, compliance data).
Storage Distribution	Automatically distributes data across Availability Zones in a region.
Retrieval Options	Expedited: 1–5 minutes Standard: 3–5 hours Bulk: 5–12 hours

Features:

- **IAM Integration**: Grants user permissions for vault access.
- **CloudTrail Logging**: Tracks API call activities for auditing.
- **Vaults**: Store archives with options to create, delete, lock, list, retrieve, tag, and configure.
- **Access Policies**: Users can set policies for enhanced security.
- **Retrieval Jobs**: Uses **Amazon SNS** for job completion notifications.
- **S3 Glacier Select**: Queries specific archive objects instead of full retrievals.
- **Supported Format**: Works with **uncompressed CSV**, outputting results to S3.
- **SQL Support**: Uses **SELECT, FROM, WHERE** for queries.
- **Encryption**: Supports **SSE-KMS** and **SSE-S3**.
- **No Real-Time Retrieval**: Archives are not instantly accessible.

Use Cases:

- It helps to store and archive media data that can increase up to the petabyte level.
- Organizations that generate, analyze, and archive large data can make use of Amazon S3 Glacier and S3 Glacier Deep Archive storage classes.
- Amazon S3 Glacier replaces tape libraries for storage because it does not require high upfront cost and maintenance.

Price details:

- Free Usage Tier - Users can retrieve with standard retrieval up to 10 GB of archive data per month for free.
- Data transfer out from S3 Glacier in the same region is free.

AWS Snow Family

AWS Snow Family provides **secure and scalable** data migration and edge computing solutions for environments with limited or no internet connectivity. It includes **AWS Snowball, AWS Snowcone, and AWS Snowmobile**.

1. AWS Snowball

A rugged storage device for transferring large datasets between **Amazon S3** and on-premises storage.

Features

- Supports **50TB - 80TB** of data transfer per device.
- Uses **AWS Key Management Service (KMS)** for encryption.
- Managed via **AWS Snow Family Console** and job management API.
- Supports **parallelization** for faster transfers.
- Integrates with **AWS CloudTrail** (API logging) and **Amazon SNS** (notifications).

Variants

- **Snowball Edge Compute Optimized** – 40 vCPUs, block & object storage.
- **Snowball Edge Storage Optimized** – 52 vCPUs, block & object storage, optional GPU.

2. AWS Snowcone

A small, lightweight edge computing and data transfer device.

Features

- **8TB of usable storage**.
- Secure with **AWS KMS encryption**.
- Supports **Wi-Fi and wired** connectivity.
- Data transfers via **AWS DataSync or shipping to AWS**.
- **Battery-powered** for field deployments.

3. AWS Snowmobile

A **high-capacity** data transfer solution for **exabyte-scale** migrations.

Features

- **Up to 100PB** per Snowmobile.
- **Secure transport** with **GPS tracking, 24/7 monitoring, and military-grade encryption**.
- Transfers data **directly to AWS data centers**.
- Ideal for **large-scale cloud migrations**.

Service	Storage Capacity	Key Features	Use Cases	Pricing
AWS Snowball	50TB - 80TB	Secure, high-speed data transfer	Large-scale migrations, on-prem analytics	Per job, per day, region-based
AWS Snowcone	8TB	Small, portable, battery-powered	Edge computing, IoT data collection	Per job, per day, region-based
AWS Snowmobile	Up to 100PB	Exabyte-scale data migration	Data center migration, massive backups	Custom pricing

AWS Storage Gateway

A **hybrid cloud storage** service that integrates **on-premise storage** with AWS. Available as **AWS hardware or a virtual machine**.

Why Use AWS Storage Gateway?

- **Compliance & Licensing** – Meets regulatory requirements.
- **Cost Reduction** – Lowers storage and management costs.
- **Backup & Automation** – Simplifies application lifecycle and backups.
- **Hybrid Cloud & Migration** – Ensures seamless data transfer to AWS.

Types of AWS Storage Gateway

Type	Function	Key Features
Volume Gateway (iSCSI)	Virtual block storage for on-prem apps	Backups stored in Amazon S3 as snapshots
Stored Volume	Local storage with AWS S3 backup	Ensures high reliability and durability
Cached Volume	Hot data stored locally, rest in S3	Reduces storage costs
File Gateway (NFSv4 / SMB)	Object storage via S3	Mounts S3 as a virtual local file system
Tape Gateway (VTL)	Virtual tape storage for backup	Uses iSCSI-based Virtual Tape Library (VTL) for cost-effective archiving

Features

- **Cost-Effective** – Pay-as-you-go pricing.
- **Low Latency** – Local storage access for faster performance.
- **Hybrid Cloud** – On-prem control with cloud scalability.
- **Compliance & Security** – Meets regulatory and licensing needs.
- **Supports Standard Protocols** – NFS, SMB, iSCSI.

Use Cases

- **Backup & Disaster Recovery** – Reliable, scalable cloud backups.
- **Hybrid Cloud Storage** – On-premises storage integrated with AWS.
- **Cloud Migration** – Easy data movement between on-prem and AWS.

Pricing

- Based on **storage type and usage**. Pay only for what you use.

Database

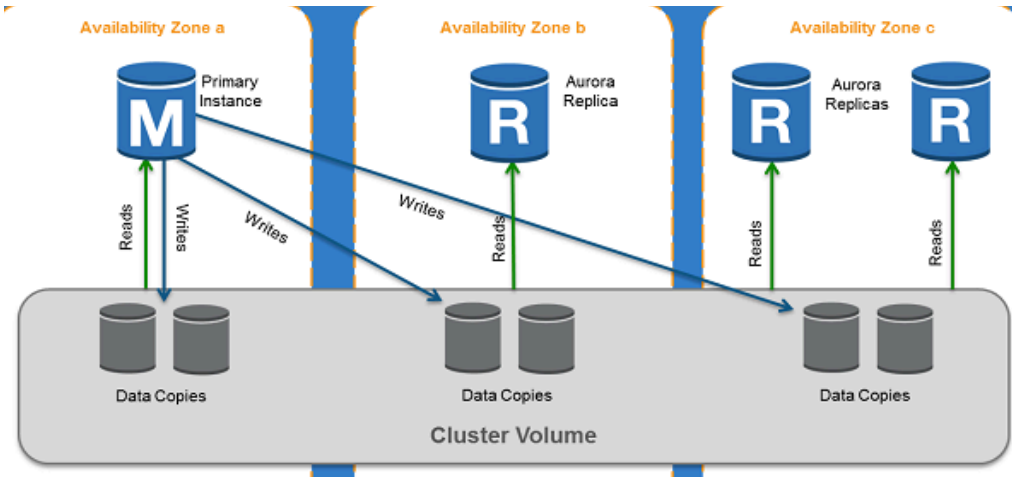
Amazon Aurora

What is Amazon Aurora?

Aurora is the fully managed RDS services offered by AWS**. It's only compatible with PostgreSQL/MySQL.** As per AWS, Aurora provides 5 times throughput to traditional MySQL and 3 times throughput to PostgreSQL.

Features:

- **Availability & Durability:**
  - Supported in regions with at least 3 AZs.
  - 99.99% availability with 6 copies of data (2 per AZ).
  - Up to 15 Read Replicas (RDS allows only 5).
  - Scales up to 128 TB per instance.
- **Aurora DB Cluster:**
  - **Primary DB Instance** – Handles read/write operations.
  - **Aurora Replica** – Read-only, auto-failover with <100 ms lag.
- **Security & Fault Tolerance:**
  - Data resides in VPC with AWS KMS encryption (at rest) and SSL (in transit).
  - **Fault tolerance:** Handles loss of 2 copies (write unaffected) and 3 copies (read unaffected).
  - **Self-healing storage:** Auto-detects and repairs disk errors.
- **Aurora Features:**
  - **Aurora Global Database** – Spans multiple regions for low-latency access and disaster recovery.
  - **Aurora Multi-Master** (MySQL only) – Enables write scaling across AZs, eliminating single points of failure.
  - **Aurora Serverless** – Auto-scales based on load, ideal for intermittent workloads.



Pricing: No upfront fees. On-demand costs more than reserved. Free backups (<1 day retention) and intra-AZ/inbound transfers. Outbound internet transfer is chargeable beyond 1 GB/month.

—-Back to Index—

Amazon DocumentDB

What is Amazon DocumentDB?

DocumentDB is a fully managed document database service by AWS which supports MongoDB workloads. It is highly recommended for storing, querying, and indexing JSON Data.

Features:

- It is compatible with MongoDB versions 3.6 and 4.0.
- All on-premise MongoDB or EC2 hosted MongoDB databases can be migrated to DocumentDB by using DMS (Database Migration Service).
- All database patching is automated in a stipulated time interval.
- DocumentDB storage scales automatically in increments of 10GB and maximum up to 64TB.
- Provides up to **15 Read replicas** with single-digit millisecond latency.
- All database instances are highly secure as they reside in VPCs which only allow a given set of users to access through security group permissions.
- It supports **role-based access control (RBAC).**
- Minimum **6 read copies of data is created in 3 availability zones making it fault-tolerant**.
- **Self-healing** – Data blocks and disks are continuously scanned and repaired automatically.
- All cluster snapshots are user-initiated and stored in S3 till explicitly deleted.

Best Practices:

- It reserves 1/3rd RAM for its services, so choose your instance type with enough RAM so that performance and throughput are not impacted.
- Setup Cloudwatch alerts to notify users when the database is reaching its maximum capacity.

Use Case:

- Highly beneficial for workloads that have flexible schemas.
- It removes the overhead of keeping two databases for operation and reporting. Store the operational data and send them parallel to BI systems for reporting without having two environments.

Pricing:

- Pricing is based on the instance hours, I/O requests, and backup storage.

Amazon DynamoDB

DynamoDB is a **serverless** NoSQL **key-value** and **document** database with **single-digit millisecond latency**. It handles **20M requests/sec** and **10T requests/day** while automatically managing data traffic across servers.

Key Features:

- **Scalability:** Supports automatic scaling and multi-region replication.
- **Flexible Schema:** Stores multi-valued attributes dynamically.
- **Primary Key Types:**
  - **Partition Key:** Unique identifier (e.g., Student_ID).
  - **Partition + Sort Key:** Composite key for better organization.
- **Indexes:**
  - **Global Secondary Index (GSI):** Different partition/sort key from the table.
  - **Local Secondary Index (LSI):** Same partition key, different sort key.

Performance & Acceleration:

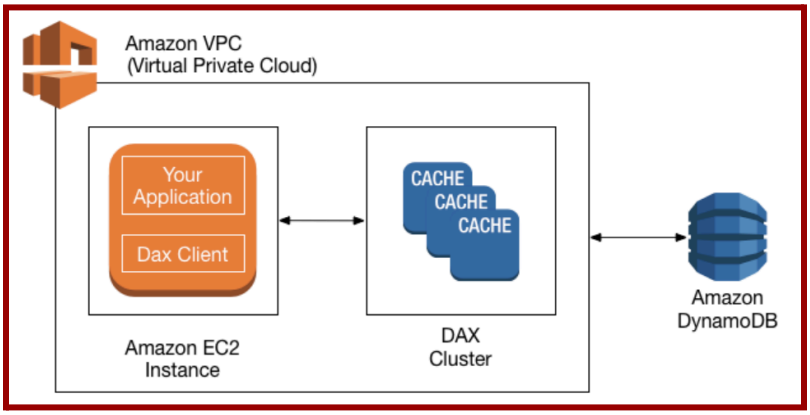
- **DynamoDB Accelerator (DAX):** In-memory caching for 10x performance boost (microseconds latency).
- Supports **horizontal scaling** (read replicas) and **vertical scaling** (node type changes).

Data Access & Operations:

- **Scan:** Retrieves multiple items but is slower than queries (up to **1MB** per operation).
- **Query:** Searches based on **primary key**, with optional **sort key** for filtering.
- **Streams:** Captures real-time item changes, retained for **24 hours**, accessed via **Lambda** or **KCL**.
- **Transactions:** ACID-compliant, supporting **up to 4MB** and **25 unique items** per transaction.

Consistency & Throughput Models:

- **Consistency:**
  - **Eventual Reads:** May return stale data but scales better.
  - **Strong Reads:** Always returns the latest data but has higher latency.
- **Throughput:**
  - **Read Capacity Unit (RCU):** 1 strong or 2 eventual reads (per 4KB).
  - **Write Capacity Unit (WCU):** 1 write per second (1KB).
  - **Provisioned Mode:** Pre-defined capacity for predictable workloads.
  - **On-Demand Mode:** Auto-scales for unpredictable workloads.



Pricing: **Pay-as-you-go** for disk space, data transfer, and provisioned throughput. Charges apply for **reserved capacity** and **on-demand usage**.

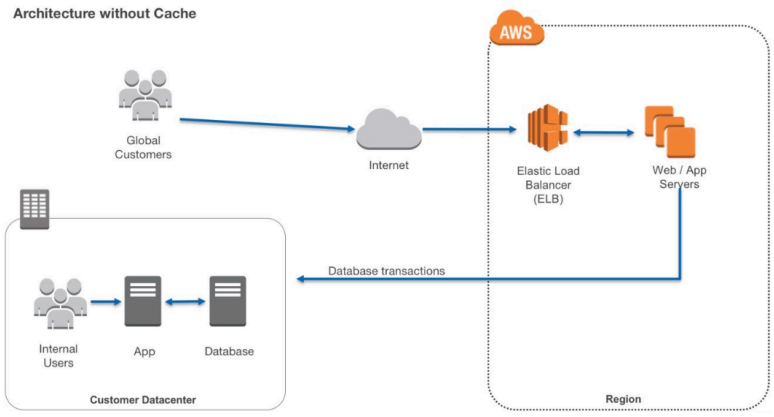
—-Back to Index—

Amazon ElastiCache

ElastiCache is a **fully managed in-memory data store** that boosts **read-heavy workloads** by reducing latency. It supports **Redis** and **Memcached** engines, offering faster performance than disk-based databases.

Key Features:

- **High Availability:** Ensures data access even during maintenance or outages.
- **Key-Value Storage:** Unlike databases, data is retrieved via key-value pairs.
- **Automatic Node Replacement:** Failed nodes are replaced automatically.



Memcached vs. Redis:

Feature	Memcached	Redis
Data Persistence	Volatile	Non-volatile
Data Types	Simple	Complex (strings, hashes, geospatial)
Multi-Threading	Yes	No
Scaling	Add/remove nodes	Add shards (primary + replicas)
Multi-AZ	Not supported	Supported via read replicas
Failover	Not supported	Auto-switch to replica

Best Practices:

- Session Storage: Use Redis for web sessions to ensure data persistence.
- Database Caching: Use Memcached with RDS for faster query performance.
- Live Polling & Gaming: Cache frequently accessed data in Memcached.
- Hybrid Approach: Combine RDS with ElastiCache for backend optimization.

Pricing:

- Charged per node hour (partial hours billed as full).
- Free data exchange within the same AZ.
- Available as on-demand or reserved nodes.

Amazon Keyspaces

What is Amazon Keyspaces (for Apache Cassandra)?

Keyspaces is an Apache Cassandra compatible database in AWS. It is fully managed by AWS, highly available, and scalable. Management of servers, patching is done by Amazon. It scales based on incoming traffic with virtually unlimited storage and throughput.

Features:

- Keyspaces is compatible with Cassandra Query Language (CQL). So your application can be easily migrated from on-premise to cloud.
- Two operation modes are available as below
    1. The On-Demand capacity mode is used when the user is not certain about the incoming load. So throughput and scaling are managed by Keyspaces itself. It's costly and you pay only for the resources you use.
    2. The Provisioned capacity mode is used when you have predictable application traffic. A user just needs to provide many max read/write per second in advance while configuring the database. It's less costly.
- There is no upper limit for throughput and storage.
- Keyspaces is integrated with Cloudwatch to measure the performance of the database with incoming traffic.
- Data is replicated across 3 Availability Zones for high durability.
- Point-in-Time-recovery (PITR) is there to recover data lost due to accidental deletes. The data can be recovered up to any second till 35 days.

Use Cases:

- Build Applications using open source Cassandra APIs and drivers. Users can use Java, Python, .NET, Ruby, Perl.
- Highly recommended for applications that demand a low latency platform like trading.
- Use cloud trail to check the DDL operations. It gives brief information on who accessed, when, what services were used and a response returned from AWS. Some hackers creeping into the database firewall can be detected here.

Pricing:

- Users only pay for the read and write throughput, storage, and networking resources.

Amazon Neptune

What is Amazon Neptune?

Amazon Neptune is a graph database service used as a web service to build and run applications that require connected datasets.

Key Features

Graph Models & Query Languages

- **Property Graph (PG)** – Uses **Apache TinkerPop Gremlin** for graph traversal.
- **Resource Description Framework (RDF)** – Uses **SPARQL** for querying data.

Performance & Scalability

- Supports **billions of connections** with **milliseconds query latency**.
- **Storage auto-scaling** to meet growing data needs.

High Availability & Fault Tolerance

- **Multi-AZ Deployment** – Ensures availability across **three AZs**.
- **Automatic Failover** – Supports **up to 15 low-latency read replicas**.
- **Fault-Tolerant Storage** – **Two copies** of data replicated across **three AZs**.

Backup & Security

- **Continuous Backup** to **Amazon S3** with **point-in-time recovery**.
- **Encryption** at **rest** and **in transit** for data security.

Amazon RDS

Amazon RDS Overview

Amazon RDS is a managed relational database service that simplifies operation, management, and scaling in the cloud. It automates tasks like patching, backups, and provisioning, offering cost-efficient scalability.

Supported Engines:

- MySQL, MariaDB, PostgreSQL – Open-source databases with easy AWS provisioning.
- MS SQL, Oracle – Commercial databases with provisioning and licensing options.
- Amazon Aurora – AWS-native MySQL/PostgreSQL-compatible engine, 5x faster than MySQL, 3x faster than PostgreSQL, supports 15 read replicas.

Instance Classes:

Type	Examples	Use Case
Standard	db.m6g, db.m5, db.m4	General-purpose workloads
Burstable	db.t3, db.t2	Baseline CPU with burst capability
Memory-Optimized	db.z1d, db.x1e, db.r5	Large datasets with high memory needs

High Availability & Performance:

Feature	Multi-AZ Deployment	Read Replicas
Replication	Synchronous	Asynchronous
Purpose	Disaster Recovery	Performance Enhancement
Scope	Two AZs in a region	Cross-AZ or cross-region
Failover	Automatic	Manual promotion

Storage Types:

- General Purpose (SSD): Baseline 3 IOPS/GiB, bursts up to 3,000 IOPS.
- Provisioned IOPS (SSD): High-performance storage, supports 1,000–30,000 IOPS.

Monitoring & Backups:

- Enhanced Monitoring: Disabled by default, incurs extra charges.
- Backups: Default retention 7 days (Console), 1 day (CLI/API), max 35 days.
- Manual Snapshots: 100 per region.

Pricing Factors:

- Active instances, storage, requests, backups, enhanced monitoring, cross-region replication.

Amazon Redshift

Amazon Redshift is a **fully managed, petabyte-scale data warehouse** offering **high scalability** at **low cost** ($1000 per TB per year).

Configuration:

- **Single Node** (160 GB)
- **Multi-Node:**
  - **Leader Node** – Manages queries & client connections.
  - **Compute Nodes** – Store data & execute queries (up to **128 nodes**).

Features:

- **Efficient Storage:** Uses compression techniques, no need for indexes or materialized views.
- **Massively Parallel Processing (MPP):** Distributes queries across nodes for fast performance.
- **High Durability:** Maintains **3 copies** of data (original, replica, and S3 backup).
- **Backup Retention: 1-day default, max 35 days**.
- **Disaster Recovery:** Can asynchronously replicate snapshots to another region.
- **Availability:** Runs in **one AZ**, but snapshots can be restored to new AZs.

Security:

- **In-transit Encryption:** SSL
- **At-rest Encryption:** AES-256
- **Key Management:** AWS-managed, **HSM**, or **AWS KMS**.

Use Cases:

- **Data ingestion from EMR, S3, DynamoDB** for BI tools.
- **Integrates with third-party libraries for querying.**

Pricing:

- **Compute Node Hours** – Charged per node-hour.
- **No charge for leader node hours.**

Security, Identity, & Compliance

AWS IAM

Identity and Access Management (IAM)

IAM is an AWS service that **securely controls access** to AWS resources by managing **users, groups, roles, and policies**.

Key Components

Principals

- **Root User** – Created with an AWS account, has full access.
- **IAM User** – Represents a person/service with assigned permissions.
- **IAM Group** – Collection of users with shared permissions.
- **IAM Role** – Temporary identity with policies, used by federated users or AWS services.

Policies

- **Identity-Based Policies** – Define access for users, groups, and roles.
    - **AWS Managed** – Predefined by AWS.
    - **Customer Managed** – Created by users for fine-grained control.
    - **Inline** – Directly attached to a user, group, or role.
- **Resource-Based Policies** – Control access at the resource level (e.g., S3 bucket).

Best Practices

✔ Grant **least privilege access**.
✔ Enable **MFA** for users.
✔ Monitor with **AWS CloudTrail**.
✔ Enforce **strong password policies**.
✔ **Use policy conditions for security.**

Pricing

IAM is **free**; charges apply only for AWS services used by IAM users.

Amazon Cognito

Amazon Cognito provides **authentication, authorization, and user management** for web and mobile apps. It supports **social sign-in** (Google, Facebook, Amazon) and **enterprise identity providers** (Microsoft AD via SAML).

Components

✔ **User Pools** – User directories offering **sign-up/sign-in**, **MFA**, **account takeover protection**, and **custom workflows** via AWS Lambda.

✔ **Identity Pools** – Provide **temporary AWS credentials** for accessing AWS resources via **Cognito user pools, third-party providers, OpenID, SAML, and developer-authenticated identities**.

✔ **Federated Identities** – Generate **temporary security credentials** for untrusted environments.

Features

✔ **Risk-based authentication** and protection from compromised credentials.

✔ Supports **OAuth 2.0, OpenID Connect, and SAML 2.0**.

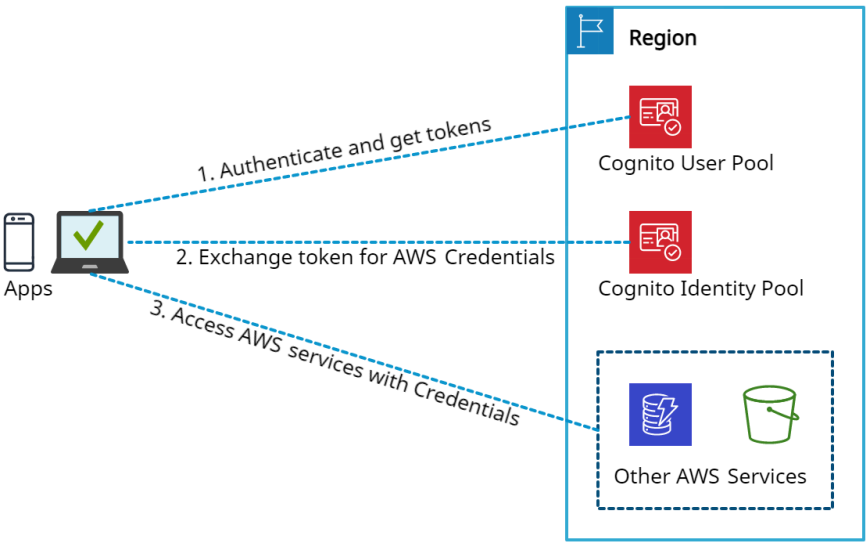
✔ Scales to **millions of users** with **SDKs for Android, iOS, and JavaScript**.

✔ **SMS/TOTP verification** (e.g., Google Authenticator).

Pricing

✔ Pay for **identity management and data synchronization** beyond the **free tier**.

✔ Volume-based pricing for **direct sign-ins and social identity providers**.



Amazon Cognito

AWS Directory Service

AWS Directory Service (AWS Managed Microsoft AD) enables **Microsoft Active Directory (AD) integration** with AWS services, supporting **authentication, schema extensions, and AD-dependent applications** like SharePoint and SQL Server.

Key Features

✔ **Trust relationships** – Extend on-premises AD authentication to AWS.
✔ **Patching without downtime** – Ensures high availability.
✔ **Supports Windows & Linux** domain-joining for EC2 instances.
✔ **Single Sign-On (SSO)** – Integrate AWS Managed AD with on-premises AD.

Limitations

❌ No **MFA**, **trust relationships**, **LDAPS communication**, or **PowerShell AD cmdlets**.

Additional Components

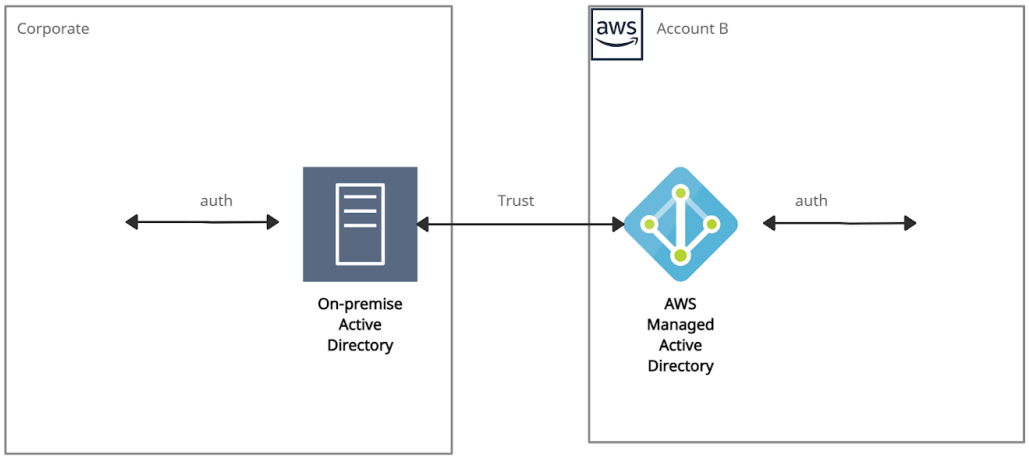
✔ **Amazon Cognito** – Provides **user authentication** via Cognito User Pools, supports **SAML-based federation** for external identities.
✔ **AD Connector** – Acts as a **gateway** redirecting directory requests to **on-premises AD**. Requires **VPN or Direct Connect**. Supports **MFA via RADIUS**.

Use Cases

✔ Sign in to AWS Cloud services with **AD credentials**.
✔ Provide **directory services** to AD-aware workloads.
✔ Enable **SSO** for Office 365 and cloud applications.
✔ Extend **on-premises AD to AWS** via **AD trusts**.

Pricing

✔ Varies by **region**.
✔ **Hourly charges** for shared directories.
✔ **Data transfer charges** for cross-region directory sharing.

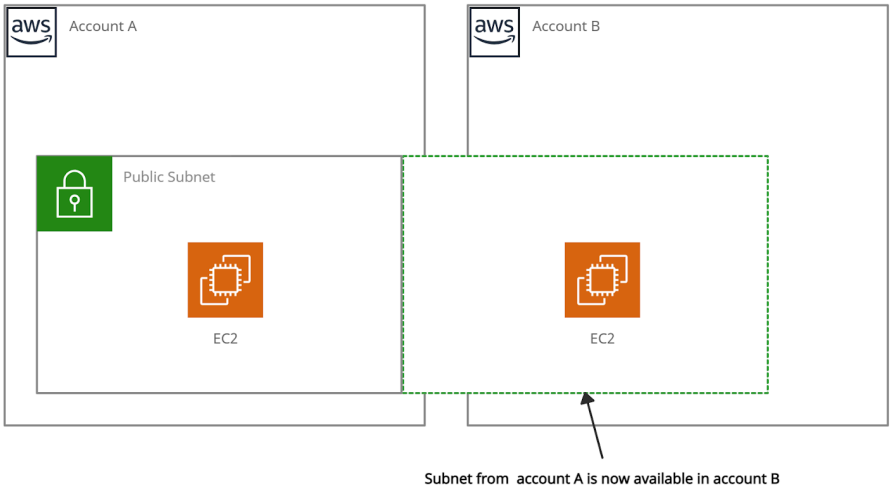


AWS Managed AD

—-Back to Index—

AWS Resource Access Manager

AWS Resource Access Manager (RAM) is a service that permits users to share their resources across AWS accounts or within their AWS Organization.



AWS Resource Access Manager

Resources that can be integrated with AWS RAM are:

- AWS App Mesh
- Amazon Aurora
- AWS Certificate Manager Private Certificate Authority
- AWS CodeBuild
- EC2 Image Builder
- AWS Glue
- AWS License Manager
- AWS Network Firewall
- AWS Outposts
- AWS Resource Groups

Benefits:

- The resource sharing feature of AWS RAM reduces customers' need to create duplicate resources in each of their accounts.
- It controls the consumption of shared resources using existing policies and permissions.
- It can be integrated with Amazon CloudWatch and AWS CloudTrail to provide detailed visibility into shared resources and accounts.
- Access control policies in AWS Identity & Access Management (IAM) and Service Control Policies in AWS Organizations provide security and governance controls to AWS Resource Access Manager (RAM).

Price details:

- The charges only differ based on the resource type. No charges are applied for creating resource shares and sharing your resources across accounts.

AWS Secrets Manager

AWS Secrets Manager securely stores, rotates, and retrieves sensitive credentials like database passwords, API keys, and OAuth tokens, replacing hardcoded secrets with API calls. It ensures encryption in transit and integrates with AWS KMS for encryption at rest.

Key Features

✔ Automated secret rotation for AWS databases (RDS, Aurora, Redshift, DocumentDB).
✔ Custom secret rotation via AWS Lambda for non-AWS services.
✔ Secure access control using IAM and resource-based policies.
✔ Monitoring & auditing with AWS CloudTrail and CloudWatch.

Access Methods

✔ AWS Console, CLI, SDKs, PowerShell, HTTPS API.

Use Cases

✔ Store encrypted secrets in SecretString/SecretBinary.
✔ Cache secrets using an open-source client for efficient access.
✔ Monitor changes with AWS Config.

Pricing

✔ Pay per stored secret and API calls.
✔ Additional charges for AWS KMS encryption keys.

AWS Security Hub

AWS Security Hub provides a **centralized view of security alerts and compliance status** across AWS accounts and services, helping organizations adhere to **security best practices**.

Key Features

✔ **Aggregates security alerts** from AWS services like:

- Amazon GuardDuty
- Amazon Inspector
- Amazon Macie
- AWS IAM Access Analyzer
- AWS Firewall Manager

✔ **Integrates with AWS Partner security solutions**.

✔ **Automated compliance checks** against:

- **PCI DSS** (Payment Card Industry Data Security Standard)
- **CIS AWS Foundations Benchmark** (43 best practices, e.g., IAM password policies).

✔ **Prioritizes findings** and suggests **remediation steps**.

Enabling Security Hub

✔ **AWS Management Console**
✔ **AWS CLI**
✔ **Infrastructure-as-Code tools (Terraform, etc.)**
✔ **Multi-region setup required** for full coverage.

Benefits:

- It collects data using a standard findings format and reduces the need for time-consuming data conversion efforts.
- Integrated dashboards are provided to show the current security and compliance status.

Price details:

- Charges applied for usage of other services that Security Hub interacts with, such as AWS Config items, but not for AWS Config rules that are enabled by Security Hub security standards.
- Using the Master account's Security Hub, the monthly cost includes the costs associated with all of the member accounts.
- Using a Member account's Security Hub, the monthly cost is only for the member account.
- Charges are applied only for the current Region, not for all Regions in which Security Hub is enabled.

AWS Key Management Service

AWS KMS is a **secure service** for creating and managing encryption keys, integrated with AWS services like Amazon S3 and EBS for **data encryption at rest**.

Key Concepts

✔ **Regional Keys** – Keys cannot be shared across regions.
✔ **Customer Master Keys (CMKs)** – Stores key metadata and is used for encryption.
✔ **Types of CMKs:**

- **Symmetric CMKs** – Single 256-bit key for encryption/decryption.
- **Asymmetric CMKs** – RSA/ECC key pairs for encryption/decryption or signing/verification.

CMK Management

✔ **Customer-Managed CMKs** – Fully controlled by users, visible in AWS KMS console.
✔ **AWS-Managed CMKs** – Created and managed by AWS, used by AWS services.

Envelope Encryption

✔ Encrypts plaintext with a **data key**, then encrypts the data key with a **master key**.
✔ Benefits:

- Protects data keys.
- Supports multiple master keys.
- Enhances security with multiple algorithms.

Features

✔ **Automatic key rotation** (yearly) without re-encryption.
✔ **AWS CloudTrail logging** for auditing.
✔ **Auto-scaling** to support encryption growth.
✔ **High availability** with multiple encrypted key copies.

Pricing

✔ **Free Tier** – 20,000 requests/month.
✔ **Customer-Managed CMKs** – $1/month per key.
✔ **AWS-Managed CMKs** – Free but limited to AWS service use.

AWS Certificate Manager (ACM)

AWS ACM provides, manages, renews, and deploys **SSL/TLS X.509 certificates** for secure web communications. Users can issue ACM certificates or import third-party certificates.

SSL Server Certificates

✔ **X.509 certificates** authenticate HTTPS transactions.
✔ Issued by a **Certificate Authority (CA)** and include server name, validity period, and public key.

Types of SSL Certificates

✔ **EV SSL** – Highest security, most expensive.
✔ **OV SSL** – Validates business credibility.
✔ **DV SSL** – Basic encryption.
✔ **Wildcard SSL** – Secures base domain + subdomains.
✔ **Multi-Domain SSL (MDC)** – Secures multiple domains/subdomains.
✔ **UCC** – Secures multiple domain names in one certificate.

Deployment Options

✔ **AWS Certificate Manager (ACM)** – Used for public certificates, deploys via **API Gateway, ELB, CloudFront**.
✔ **ACM Private CA** – Creates internal **PKI** to issue private certificates for internal authentication.

ACM-Integrated Services

✔ **Elastic Load Balancing, CloudFront, API Gateway, Elastic Beanstalk, AWS Nitro Enclaves, CloudFormation**.

Benefits

✔ **Automated creation & renewal** of SSL/TLS certificates.
✔ Simplifies **certificate issuance** and management.
✔ Ensures **data-in-transit security** and site authentication.

Pricing

✔ **Public ACM certificates** – Free when used with ACM-integrated services.
✔ **ACM Private CA** – Monthly charges for private CA operation and issued certificates.

Management & Governance

AWS Auto Scaling

AWS Auto Scaling monitors applications and automatically adjusts capacity for **consistent performance** and **cost efficiency**. It supports scaling for **EC2 Instances, ECS tasks, DynamoDB, and Aurora Read Replicas**.

Key Concepts

✔ **Launch Configuration vs. Launch Template**

- **Launch templates** offer the latest EC2 features and support on-demand & spot instances.
- **Launch configurations** lack some Auto Scaling features.

✔ **Lifecycle Hooks**

- Pause EC2 instances in **wait state** until an action completes or timeout ends.

✔ **Monitoring**

- **Health Checks** – Remove unhealthy instances from the target group.
- **CloudWatch Events & Metrics** – Track scaling actions and performance.
- **Notification Service** – Alerts for instance launch/termination.

Pricing

✔ **No additional cost** for Auto Scaling itself.
✔ **Pay only for the AWS resources used**.

AWS CloudFormation

AWS CloudFormation automates resource provisioning and management using **templates** to create, update, and delete stacks as a **single unit**. It integrates with **IAM for security** and **CloudTrail for API event tracking**.

Key Concepts

✔ **Templates** – JSON/YAML files for defining AWS resources.
✔ **Stack** – A collection of resources managed together.
✔ **Change Sets** – Preview changes before applying them.
✔ **Stack Updates** – Update only modified resources.
✔ **StackSets** – Manage stacks across multiple accounts/regions.
✔ **Nested Stacks** – Reuse common components within stacks.
✔ **CloudFormation Registry** – Supports third-party resource provisioning.

Pricing

✔ **Free for AWS resources**; charges apply for the services used.
✔ Supports *AWS::*, Alexa::*, and Custom:: namespaces**; others incur costs.
✔ **Free tier:** 1000 handler operations/month.
✔ **Paid operations:** $0.0009 per handler operation.

Example: EC2 Instance Template

```
EC2Instance:
  Type: AWS::EC2::Instance
  Properties:
    ImageId: 1234xyz
    KeyName: aws-keypair
    InstanceType: t2.micro
    SecurityGroups:
      - !Ref EC2SecurityGroup
    BlockDeviceMappings:
      - DeviceName: /dev/sda1
        Ebs:
          VolumeSize: 50
```

AWS CloudTrail

AWS CloudTrail enables operational and risk auditing by tracking account activity across AWS services. It records actions as events from users, roles, and AWS services via the Console, CLI, SDKs, and APIs.

Integrations

✔ Amazon S3 – Stores and retrieves log files.
✔ Amazon SNS & SQS – Notifies on log file delivery.
✔ Amazon CloudWatch & IAM – For monitoring and security.
✔ CloudTrail Insights – Detects unusual API activity.
✔ Log Retention – Past 90 days of events can be viewed/downloaded.

Event Types

✔ Management Events (e.g., CreateSubnet, CreateDefaultVpc)
✔ Data Events (e.g., GetObject, DeleteObject, PutObject)
✔ Insights Events (e.g., deleteBucket, AuthorizeSecurityGroupIngress)

CloudTrail vs. CloudWatch

✔ CloudTrail – Logs all AWS actions for auditing.
✔ CloudWatch – Monitors AWS services for performance & health.

Pricing

✔ First copy of management events per region is free.
✔ Additional copies: $2.00 per 100,000 events.
✔ Data events: $0.10 per 100,000 events.
✔ Insights events: $0.35 per 100,000 analyzed events.

Amazon CloudWatch

Amazon CloudWatch monitors and manages AWS applications and infrastructure by collecting logs, metrics, and events. It supports **EC2, RDS, DynamoDB, and custom log files**.

Access Methods

✔ **CloudWatch Console**
✔ **AWS CLI, API, SDKs**

Integrations

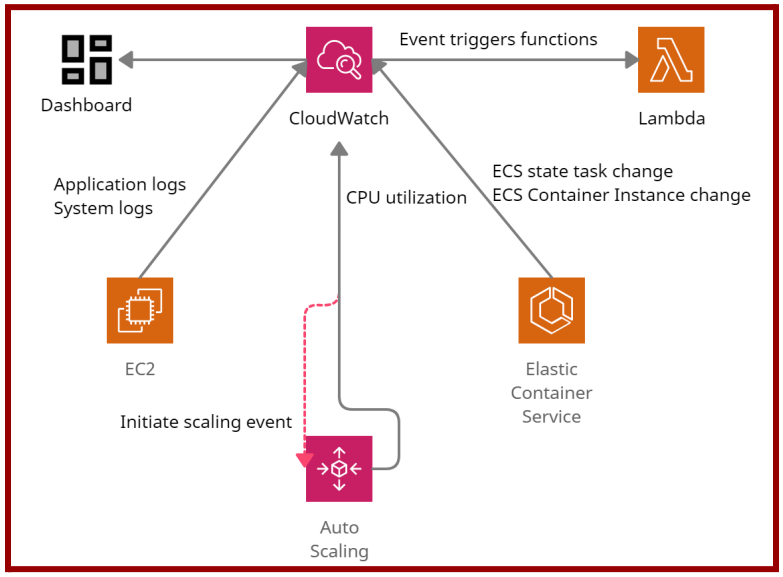
✔ **Amazon SNS** – Sends notifications
✔ **EC2 Auto Scaling** – Adjusts resources
✔ **AWS CloudTrail & IAM** – Security & auditing

Key Features

✔ **Custom Dashboards** – Visualize metrics
✔ **Alarms** – Trigger actions on threshold breaches
✔ **Cross-Account Visibility** – Unified monitoring across AWS accounts
✔ **Container Insights** – Monitors ECS, EKS, Kubernetes
✔ **Lambda Insights** – Tracks CPU, memory, disk, and network usage

CloudWatch Agent

✔ Collects **system-level metrics** from EC2/on-prem servers.
✔ Supports **StatsD (Linux/Windows)** and **collectd (Linux)** for custom metrics.
✔ Default namespace: **CWAgent** (configurable).



Amazon CloudWatch in action

AWS Config

AWS Config **monitors, evaluates, and records** AWS resource configurations, tracking changes over time.

Key Features

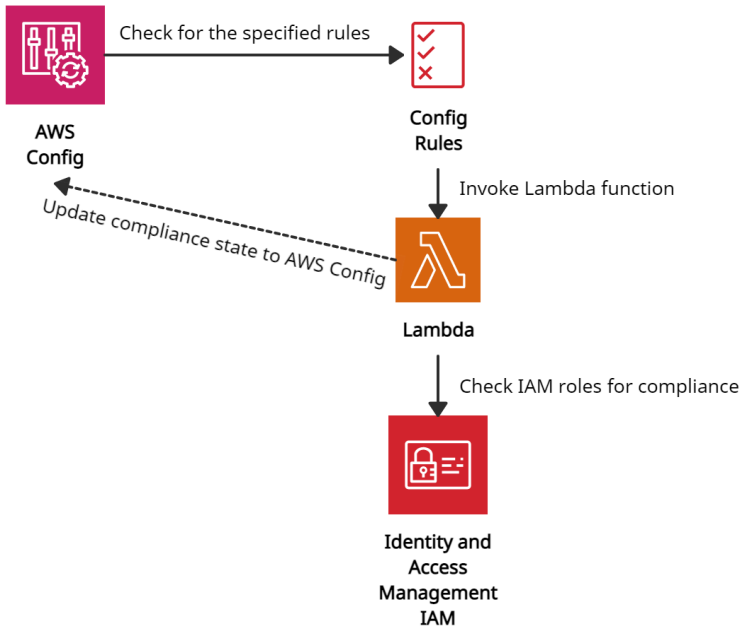
✔ **Configuration Snapshots** – Provides a complete resource inventory.
✔ **Change Tracking** – Records modifications and sends notifications.
✔ **Compliance Checks** – Uses **Managed** & **Custom Config Rules** (150 per region).
✔ **Integration** – Works with **IAM, S3, SNS, and CloudTrail** for auditing and alerts.
✔ **Aggregators** – Collects compliance data across multiple **accounts, regions, or AWS Organizations**.

Use Cases

✔ **Automates compliance checks** using Lambda-based custom rules.
✔ **Identifies security risks** and tracks historical configurations for analysis.

Pricing

✔ **$0.003 per configuration item recorded per region**.
✔ **Charges for Config rule evaluations** and **integrations with AWS services**.



AWS Config in action

AWS License Manager

AWS License Manager **manages software licenses** for AWS and on-premises environments, supporting **Bring-Your-Own-License (BYOL)** for vendors like Microsoft, SAP, Oracle, and IBM.

Key Features

✔ **Custom Licensing Rules** – Prevents violations with **hard (blocks) & soft (alerts) limits**.
✔ **Dashboard Control** – Provides visibility and enforcement of license usage.
✔ **Dedicated Host Management** – Optimizes allocation & capacity utilization.
✔ **Managed Entitlements** – Controls license assignments for users/workloads.
✔ **Cross-Account Management** – Uses **AWS Organizations** for license sharing.
✔ **Integration** – Works with **EC2, RDS, Systems Manager, IAM, Marketplace, CloudFormation, X-Ray**.

Pricing

✔ No additional charges; **AWS resources follow standard pricing**.

AWS Organizations

AWS Organizations **manages multiple AWS accounts**, enforcing security, governance, and cost tracking.

Access Methods

✔ **Console, CLI, SDKs, API, PowerShell**

Key Features

✔ **Security Boundaries** – Uses multiple member accounts.
✔ **Organizational Units (OUs)** – Groups accounts for better management.
✔ **Service Control Policies (SCPs)** – Enforces security & governance.
✔ **Cost Allocation Tags** – Tracks AWS costs per account.

Integration with AWS Services

✔ **CloudTrail** – Auditing & logging
✔ **Backup** – Backup monitoring
✔ **Control Tower** – Cross-account security & policy view
✔ **GuardDuty** – Threat detection
✔ **Resource Access Manager (RAM)** – Resource sharing

Member Account Migration

1️⃣ Remove from old Organization
2️⃣ Send invitation from new Organization
3️⃣ Accept invitation from member account

Pricing

✔ **Free service**; standard charges apply for AWS resources.
✔ **Consolidated billing** ensures volume discounts across accounts.

AWS Systems Manager

AWS Systems Manager **manages EC2 and on-premises systems at scale**, detects infrastructure issues, and automates patching for compliance. Works with **Windows & Linux**.

Key Features

✔ **Integration** – Works with CloudWatch & AWS Config
✔ **Software Discovery** – Audits installed software
✔ **Compliance Management** – Monitors patch levels & configurations
✔ **Resource Grouping** – Over 100 resource types into applications & units
✔ **Automated Workflows** – Reduces errors with centralized parameters
✔ **Security & Patching** – Runs commands & scheduled patching
✔ **Software Distribution** – Manages multiple versions safely

How it Works

- **SSM Agent** must be installed on controlled systems.
- **IAM Role** required for EC2 instances to allow SSM actions.

Pricing

✔ **App Config** – $0.2 per 1M API calls, $0.0008 per config received
✔ **Parameter Store** – Standard (Free), Advanced ($0.05/param/month)
✔ **Change Manager** – $0.296 per change request, $0.039 per 1K API calls

AWS Health dashboard

AWS Health Dashboard **provides real-time service events, scheduled changes, and account notifications**, enabling proactive issue resolution. Access updates via **AWS Health API (Premium Support), EventBridge, or the console**.

Key Features

✔ **Centralized Hub** – Integrates with 200+ AWS services for full visibility
✔ **Actionable Insights** – Helps troubleshoot & manage changes proactively
✔ **AWS Organizations Integration** – Consolidates service health across accounts
✔ **Automated Notifications** – Receive alerts via **EventBridge & ITSM tools**

Use Cases

✔ **Proactive Alerts** – Minimize disruptions with real-time updates
✔ **Lifecycle Event Monitoring** – Track planned events & resource-level actions
✔ **Efficient Event Tracking** – Automate monitoring via ITSM integrations
✔ **Incident Troubleshooting** – Identify AWS-related issues affecting applications

AWS Control Tower

AWS Control Tower **extends AWS Organizations** by providing governance and multi-account management with a **Landing Zone** based on AWS best practices.

Key Features

✔ **Preconfigured OUs** – Security (Audit & Log Archive), Sandbox, and Production
✔ **Identity Integration** – Supports AWS Identity Center, SAML IdPs, and Microsoft AD
✔ **Guardrails** – Preventive (SCP-based) & Detective (AWS Config & Lambda-based)
✔ **Centralized Dashboard** – Monitors accounts, OUs, and compliance policies
✔ **Account Factory** – Standardized account provisioning with pre-approved configurations

Use Cases

✔ **New or Existing AWS Organization Deployment** – Works with both setups
✔ **Automated Governance & Compliance** – Ensures policy enforcement across accounts

AWS Trusted Advisor

AWS Trusted Advisor provides best practice checks in Cost Optimization, Security, Fault Tolerance, Performance, and Service Limits to enhance AWS infrastructure.

Key Features

✔ Cost Optimization – Identifies unused/idle resources and recommends Reserved capacity
✔ Security – Checks S3 permissions, security groups, and NACLs for vulnerabilities
✔ Fault Tolerance – Suggests Auto Scaling, Multi-AZ, and backup configurations
✔ Performance – Recommends throughput optimization and monitors resource utilization
✔ Service Limits – Alerts when resource usage exceeds 80%

Use Cases

✔ Reduce Costs – Detect idle resources and optimize usage
✔ Enhance Security – Identify misconfigurations and access risks
✔ Improve Performance – Optimize resource allocation and scaling

Developer Tools

AWS Developer Tools

Developer tools overview

Feature	Details
Benefits	Faster releases, seamless AWS integration, simplified development, ML-based security & code quality.
Services	CI/CD, Infrastructure as Code, SDKs & CLI, IDEs, collaboration tools.
Use Cases	Automate deployments, streamline CI/CD, boost productivity, monitor performance.

AWS CodeBuild

Feature	Details
Description	Fully managed CI service for efficient builds & tests.
Benefits	No build queue waiting, auto-scaling, pay-as-you-go pricing.
Features	Easy setup, integrates with Jenkins/Git, automated builds.
Pricing	First 100 minutes free, charges based on usage.

AWS CodeDeploy

Feature	Details
Description	Automates deployments to EC2, Lambda, on-premises, ECS.
Deployment	Code, Lambda, web files, executables, scripts.
How It Works	Define revision → Configure YAML → Deploy.
Features	Rapid releases, zero-downtime deployment, status tracking.
Pricing	Free for EC2/Lambda, $0.02 per on-prem deployment.

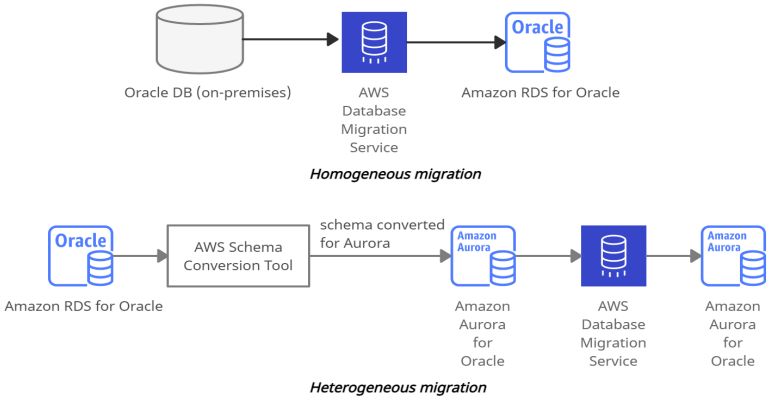
AWS X-Ray

Feature	Details
Description	Debugs & analyzes distributed applications.
Components	Daemon, Segments, Traces, Sampling, Service Graph.
Features	Supports multiple languages, AWS integrations, performance insights.
Pricing	$0.50 per 1M requests (beyond free tier).

Migration & Transfer

AWS Database Migration Service

Description: Migrates databases to AWS with minimal downtime, supporting both homogeneous and heterogeneous migrations.
Supported Sources: Oracle, SQL Server, PostgreSQL, Db2, SAP, MySQL, MariaDB, MongoDB, Amazon Aurora.
Supported Targets: Oracle, SQL Server, PostgreSQL, SAP ASE, MySQL, Amazon Redshift, Amazon S3, DynamoDB.
Features: Continuous replication, error handling, monitoring, scaling, and software patching.
Enhanced Migration: AWS DMS + AWS Schema Conversion Tool (AWS SCT) enables heterogeneous migrations.



AWS Database Migration Service

AWS Application Discovery Service

Description: Helps plan migrations by identifying servers, VMs, and network dependencies in on-prem data centers.
Discovery Methods:

- **Agentless:** Uses Collector for basic discovery, no agent installation needed.
- **Agent-based:** Provides detailed data collection with an installed agent.
  Integration: Works with AWS Partner Network (APN) and AWS Migration Hub.
  Use Cases: Server inventory discovery, network mapping, migration planning.
  Pricing: Free discovery; pay only for AWS resources used (e.g., S3, Athena, Kinesis).

AWS DataSync

Description: A managed service for secure, automated data migration between AWS, on-premises, edge locations, and other cloud providers.

Features

- Supports scheduling, bandwidth throttling, task filtering, and logging.
- Uses compression and parallel transfers for faster data movement.
- Provides in-flight (TLS) and at-rest encryption.
- Ensures data integrity verification.
- Integrates with CloudWatch, CloudTrail, and EventBridge.
- Pay-as-you-go pricing ($0.0125/GB).

Best Practices

- Evaluate tools, bandwidth, and source/destination before migration.
- Deploy & activate an **Agent** for on-premises to AWS transfers.
- Combine **DataSync** for archiving and **Storage Gateway** for local access.
- Use **Lambda** to trigger transfers when schedules are undefined.
- No Agent needed for AWS-to-AWS transfers (e.g., S3 → S3, S3 → EFS).

AWS Migration Hub

Description: A centralized platform for discovering, planning, and tracking application migrations, offering visibility across tools and processes.

Benefits

- **Streamlined Process:** Manage discovery, assessment, planning, execution, and tracking in one place.
- **Guided Expertise:** Use pre-built templates for faster migration.
- **Effective Resources:** Leverage specialized services for transformation.
- **Free to Use:** No cost for planning or tracking migrations.

Use Cases

- **Migration Planning:** Identify applications and develop migration strategies.
- **Migration Execution:** Use guided templates and services for seamless migration.
- **Application Modernization:** Refactor and manage applications efficiently.

Pricing

- **Free** for discovery, planning, and tracking migrations.
- Users pay for migration tools and AWS resources.
- **Refactor Spaces:**
  - 3 environments free for 3 months (2,160 hours/month).
  - After free tier: **$0.028/hour** per environment ($20/month if run continuously).
  - **API Requests:** $0.000002/request, with **500,000 free/month** in AWS Free Tier.

AWS Transfer Family

Category	Description
Service Overview	AWS Transfer Family is a fully managed, secure service that enables file transfers to/from AWS storage (S3, EFS) and on-premises systems using SFTP, FTPS, and FTP. It simplifies migration of file transfer workloads without impacting existing integrations.
SFTP	Secure File Transfer Protocol that uses SSH for encrypted file transfers.
FTPS	FTP over a TLS-encrypted channel, providing secure file transfers.
FTP	Standard File Transfer Protocol without encryption; typically used within a VPC via VPC endpoints for enhanced security.
Key Features	- Fully managed endpoints for S3 and EFS - Global high availability - Compliance with regional regulations - Pay-as-you-go pricing - Custom Identity Providers using API Gateway & Lambda
Use Cases	- Secure file transfers with IAM roles for S3 access - Migrating existing file transfer hostnames using Route 53 - Public access for SFTP/FTPS and restricted internal FTP access via VPC endpoints

Networking & Content Delivery

Amazon API Gateway

Category	Description
Overview	Amazon API Gateway creates, publishes, monitors, and secures APIs at any scale, powering both serverless and microservices architectures.
API Types	**REST APIs:** HTTP-based, stateless communication with standard methods (GET, POST, PUT, PATCH, DELETE). **WebSocket APIs:** Stateful, full-duplex communication.
Endpoint Types	**Edge-Optimized:** Global low latency with CloudFront. **Regional:** Optimized for same-region requests with CDN/WAF. **Private:** Restricted to VPC access.
Security	Secured using resource policies, IAM, Lambda authorizers, and Cognito user pools.
Integrations	Works with EC2, Lambda, CloudTrail, CloudWatch, AWS WAF, and X-Ray.
Pricing	Charges apply for API caching; auth failures, missing API keys, and throttled requests are free.

AWS Cloud Map

Category	Description
Overview	AWS Cloud Map is a service registry that tracks the names, locations, attributes, and health of cloud resources, enabling dynamic discovery.
How It Works	Resources are registered as service instances via the RegisterInstance API. Applications discover healthy instances using DiscoverInstances API or DNS.
Integrations	Supports various AWS resources (EC2, SQS, DynamoDB, S3, API Gateway) and is tightly integrated with Amazon ECS.
Features	Automates resource management, reduces manual tracking, supports health checks (Route 53 or third-party), and facilitates dynamic scaling.
Pricing	$0.10 per registered resource/month; $1.00 per million discovery API calls; plus additional Route 53 charges.

Amazon CloudFront

Category	Description
Overview	Amazon CloudFront is a CDN that securely delivers content worldwide with low latency and high transfer speeds by caching data at edge locations.
Integrations	Works with AWS services such as S3, EC2, ELB, Route 53, and Elemental Media Services.
Origins	Retrieves content from Amazon S3, EC2, ELB, or custom HTTP origins.
Edge Computing	Supports Lambda@Edge for custom code execution, dynamic load-balancing, and enhanced security at the edge.
Security	Provides HTTPS encryption (including field-level), AWS Shield Standard for DDoS protection, AWS WAF, and Origin Access Identity (OAI) to secure S3 content.
Access Controls	Uses signed URLs, signed cookies, and geo-restrictions to control access to content.
Pricing	Charged for data transfer out, HTTP/HTTPS requests, custom SSL certificates, field-level encryption, and Lambda@Edge execution. Free for inter-region transfers, ACM, and shared certificates.

AWS PrivateLink

Category	Description
Overview	AWS PrivateLink enables secure, private connectivity between VPCs and AWS services (or endpoint services) without using the public internet.
Endpoints	**Interface Endpoints:** Create an ENI in a subnet with a private IP for AWS service access. **Gateway Endpoints:** Route traffic to S3 and DynamoDB via the route table.
Use Cases	Connects service consumers to providers across VPCs securely, and supports secure on-premises migration via AWS Direct Connect or VPN.
Security & Integration	Enhances security by isolating traffic from the public internet and integrates with AWS Marketplace services, reducing exposure to DDoS and brute-force attacks.
Pricing	Charged based on the usage of endpoints.

AWS Transit Gateway

Overview:

- Central hub to interconnect multiple VPCs.
- Simplifies complex VPC peering and hybrid connectivity.
- Manages AWS routing configurations in one place.

Connectivity:

- Supports multiple Transit Gateways per region (cannot peer within a single region).
- Connects with AWS Direct Connect gateway (across different AWS accounts).
- Enables IPsec VPN connections via VPN attachments.
- Supports IPv6 CIDRs for VPC attachments.

Management:

- Create via AWS CLI, Management Console, or CloudFormation.
- Transit Gateway Network Manager monitors networking resources and remote branch connections.
- Allows multi-user gateway connections for redundancy.

Transit Gateway vs. VPC peering:

Transit Gateway	VPC peering
• It has an hourly charge per attachment in addition to the data transfer fees. • Multicast traffic can be routed between VPC attachments to a Transit Gateway. • It provides Maximum bandwidth (burst) of 50 Gbps per Availability Zone per VPC connection. • Security groups feature does not currently work with Transit Gateway.	• It does not charge for data transfer. • Multicast traffic cannot be routed to peering connections. • It provides no aggregate bandwidth. • Security groups feature works with intra-Region VPC peering.

AWS Direct Connect

Establishes a dedicated network connection from an on-premises environment to one or more AWS VPCs in the same region. Bypasses the public Internet for a more consistent network experience.

Virtual Interfaces:

- **Private VIF:** Connects to an Amazon VPC using private IP addresses.
- **Public VIF:** Connects to AWS services (except in China) using public IP addresses.

Connection Options:

- AWS Managed VPN, Direct Connect, Direct Connect + VPN, VPN CloudHub, Transit VPC, VPC Peering, AWS PrivateLink, VPC Endpoints.

Direct Connect Gateway:

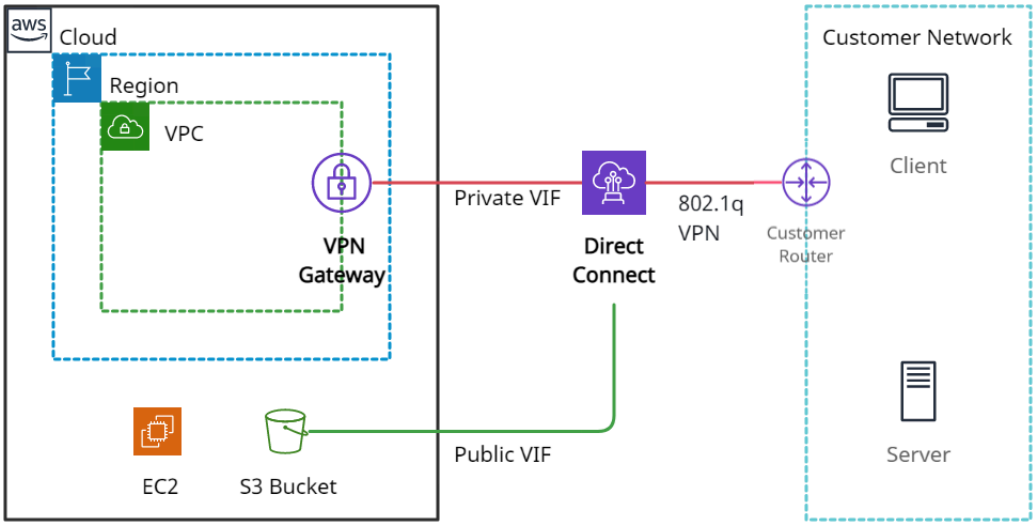
- Globally available service to connect multiple VPCs across regions or AWS accounts.
- Integrates with:
  - **Transit Gateway:** Connects multiple VPCs to an on-premises network within the same region.
  - **Virtual Private Gateway:** Provides edge routing for VPCs.

Key Features:

- Configurable via the AWS Management Console.
- Provides scalable dedicated connections (1 Gbps and 10 Gbps options).
- Ensures consistent connectivity with improved performance.

Pricing:

- Pay-as-you-go with no minimum fee.
- Charged per dedicated connection port hour (uniform globally, except Japan).
- Data Transfer OUT charges vary based on the AWS Region.



Amazon Direct Connect

AWS Elastic Load Balancer

Distributes incoming traffic across multiple targets (EC2, containers, Lambda, IPs) across one or more Availability Zones for high availability, scalability, and security.

Types:

- **Application Load Balancer:** Ideal for web apps; routes traffic based on request content.
- **Network Load Balancer:** Suited for high-performance apps; supports TCP, UDP, and TLS protocols.
- **Gateway Load Balancer:** Designed for third-party appliances (e.g., security, analytics).
- **Classic Load Balancer:** Legacy option for EC2; AWS recommends using ALB or NLB.

Key Components:

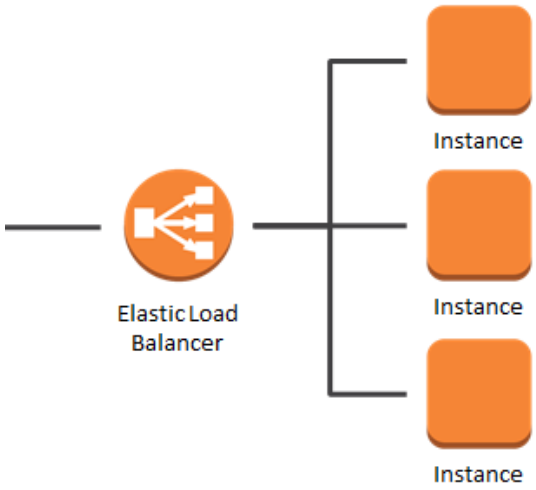
- **Listeners:** Monitor incoming requests on specified protocols/ports (HTTP, HTTPS).
- **Target Groups:** Define the destination for traffic (instances, IPs, Lambda functions).
- **Health Checks:** Regularly monitor target health and remove unhealthy targets.

Use Cases:

- Balancing traffic across multiple servers for web applications.
- Building hybrid cloud solutions by load balancing across AWS and on-premises resources.
- Supporting AWS migrations with auto-scaling and dynamic capacity management.

Pricing:

Billed hourly (or partial hour) plus based on Load Balancer Units (LCUs).



Amazon Route 53

- Managed DNS service that routes users to servers via domain names.
- Acts as a domain name registrar and DNS server.

Hosted Zones:

- **Public Hosted Zone:** Routes traffic on the Internet.
- **Private Hosted Zone:** Routes traffic within a VPC.

Record Types:

- Common: A (IPv4), AAAA (IPv6), CNAME, Alias.
- Others: CAA, MX, NAPTR, NS, PTR, SOA, SPF, SRV, TXT.

Routing Policies:

- **Simple:** Single resource routing (no health checks).
- **Weighted:** Routes based on assigned weights; supports health checks.
- **Failover:** Routes to secondary resource if primary fails.
- **Geo-location/Geo-proximity:** Routes based on geographic location.
- **Latency-based:** Routes to lowest-latency destination.
- **Multi-value Answer:** Distributes responses across multiple IPs.

Use Cases:

- Domain registration and DNS hosting.
- Managing public and private DNS zones.
- Routing based on performance, location, and health.

Pricing:

- No long-term contracts; annual fees for registered domains.
- Charges vary for different query types (standard, latency, geo, etc.).

Route53 CNAME vs. Alias

CNAME	Alias
• It points a hostname to any other hostname. • (app.mything.com -> abc.anything.com) • It works only for the non-root domains. • (abcxyz.maindomain.com) • Route 53 charges for CNAME queries. • It points to any DNS record that is hosted anywhere.	• It points a hostname to an AWS Resource. • (app.mything.com ->abc.amazonaws.com) • It works for the root domain and non-root domain. (maindomain.com) • Route 53 doesn't charge for Alias queries. • It points to an ELB, CloudFront distribution, Elastic Beanstalk environment, S3 bucket as a static website, or another record in the same hosted zone.

AWS VPC

A dedicated virtual network in AWS where you can launch and manage resources in an isolated environment.

Security:

- **Security Groups:**
  - *Default:* Allow all inbound/outbound traffic.
  - *Custom:* Block inbound by default, allow outbound.
- **Network ACLs:**
  - *Default:* Allow all traffic.
  - *Custom:* Deny all traffic by default.

Core Components:

- **Subnets:**
  - Logical IP address divisions.
  - *Public Subnet:* Has internet access via an Internet Gateway.
  - *Private Subnet:* No direct internet access; requires NAT for outbound connectivity.
- **Route Tables:**
  - Direct network traffic.
  - Public subnets use routes to an Internet Gateway; private subnets use NAT.
- **NAT Devices:**
  - **NAT Instance:** EC2 instance deployed in a public subnet for outbound IPv4 traffic.
  - **NAT Gateway:** Managed by AWS for scalable outbound connectivity.
- **DHCP Options Set:**
  - Automatically configures network parameters like domain name and DNS servers.
- **PrivateLink & Endpoints:**
  - Provide secure, private connectivity to AWS services without using the public internet.
- **Egress-Only Internet Gateway:**
  - Enables outbound-only IPv6 traffic.
- **VPC Peering:**
  - Connects two VPCs (within or across regions) for seamless resource communication.
- **VPN Connections:**
  - **AWS Site-to-Site VPN:** Securely connects on-premises networks to your VPC.
  - **AWS Client VPN:** Provides remote user connectivity to AWS resources.

Use Cases:

- Hosting public websites and multi-tier applications.
- Disaster recovery.
- Hybrid cloud setups and secure communication between different networks.

Pricing:

- VPC creation is free.
- Charges apply for NAT Gateway usage, data processing, and traffic mirroring.

Front End Web & Mobile

Service	Description	Key Features / Use Cases
AWS AppSync	Simplifies app development by creating secure, real-time GraphQL APIs connecting clients to backend data.	GraphQL schema, resolvers, real-time subscriptions, offline access, caching, conflict resolution.
AWS Amplify	Streamlines development, deployment, and management of full-stack web and mobile applications.	Supports React, Vue, etc.; built-in auth, storage (S3), serverless functions, CI/CD integration, scalable global deployments.
AWS Device Farm	Tests mobile and web apps on real devices and desktop browsers hosted on AWS.	Real device testing, parallel testing, advanced debugging (logs, videos), CI/CD integration, private device lab options.
Amazon EventBridge	A serverless event bus that routes real-time events from AWS, SaaS, and custom apps to various targets.	Event buses, rules, schema registry, supports 90+ AWS services as sources, 17+ targets, pay-as-you-go pricing for event ingestion.
AWS SNS	A flexible notification service for publishing messages to multiple endpoints (e.g., mobile, SMS, email).	Supports standard and FIFO topics; push notifications; integrates with Lambda, SQS, and more; ideal for app-to-person and app-to-app messaging.
Amazon SQS	A serverless message queuing service that decouples application components for asynchronous processing.	Offers Standard (unordered, at-least-once) and FIFO (ordered, exactly-once) queues, with delay and dead-letter queue support.
AWS Step Functions	Orchestrates serverless workflows by coordinating AWS services into state machines.	Standard and Express workflows, built-in retries, error handling, GUI monitoring, integration with Lambda, ECS, SQS, and more.
Amazon SWF	Manages and coordinates complex, distributed workflows with task scheduling and state management.	Workflow starters, deciders, and activity workers; supports long-running processes, concurrency, retries, and detailed logging.

Billing & Cost Management

AWS Cost Explorer

- **What It Is:** A UI tool for analyzing AWS cost and usage via graphs and reports.
- **Reports:** Cost & Usage and Reserved Instance (RI) utilization/coverage.
- **Data:** Up to 12 months of historical data, current month, and 12-month forecasts.
- **Access:** Billing & Cost Management console and API ($0.01 per API request).

AWS Budgets

- **What It Is:** A tool to set custom budgets and receive alerts for cost and usage.
- **Features:** Supports cost, usage, RI, and Savings Plans budgets with email/SNS alerts.
- **Access:** AWS Management Console, CLI, and API.
- **Pricing:** Monitoring is free; action-enabled budgets cost $0.10 per day after free quota.

AWS Cost & Usage Report (CUR)

- **What It Is:** Detailed report of AWS cost, usage, and resource metadata.
- **Delivery:** Report files sent to S3 up to three times per day.
- **Usage:** Analyze using Athena, Redshift, or QuickSight.
- **Access:** CUR API for creation, retrieval, and deletion.

Reserved Instance Reporting

- **What It Is:** Reports on RI utilization and coverage via Cost Explorer.
- **Features:** Visual charts and tables to monitor RI performance.
- **Pricing:** $0.01 per request for recommendation data retrieval.

AWS Management Console

- **What It Is:** A web-based interface for managing AWS services.
- **Features:** Access to all AWS service consoles, recent services, region selection, and a search function.
- **Availability:** Also available as a mobile app for Android and iOS.

Machine Learning

AI Models: Types

1. Computer Vision Models

- **Amazon Rekognition**: This service provides pre-trained models for image and video analysis, including capabilities like object detection, facial recognition, and scene detection.

2. Natural Language Processing (NLP) Models

- **Amazon Comprehend**: Used for analyzing text, Amazon Comprehend can perform sentiment analysis, entity recognition, and language detection.
- **Amazon Translate**: Provides real-time translation services between different languages.
- **Amazon Lex**: Powers conversational interfaces, enabling the creation of chatbots that can interact through voice and text.
- **Amazon Polly**: Converts written text into lifelike speech in multiple languages.

3. Speech Recognition Models

- **Amazon Transcribe**: This service converts speech into text, making it useful for transcriptions, subtitles, and more.

4. Document Processing Models

- **Amazon Textract**: Extracts text, tables, and other data from scanned documents, making it easier to process and analyze paper-based information.

5. Recommendation and Forecasting Models

- **Amazon Personalize**: Delivers personalized recommendations by analyzing user behavior and preferences.
- **Amazon Forecast**: Utilizes time-series data to predict future trends, such as sales forecasts or inventory needs.

6. Search and Information Retrieval Models

- **Amazon Kendra**: An enterprise search service that uses machine learning to provide relevant search results across documents and data sources.

7. Custom Machine Learning Models

- **Amazon SageMaker**: A comprehensive platform for building, training, and deploying custom machine learning models. It supports a wide range of algorithms and frameworks.

8. Generative AI Models

- **Amazon Bedrock**: A service that provides access to foundational models for generative AI, allowing users to create custom applications like text generation or image creation.
- **SageMaker JumpStart**: Offers pre-trained models and solutions for generative AI tasks, which can be fine-tuned for specific needs.

9. Edge AI Models

- **AWS IoT Greengrass ML Inference**: Enables machine learning inference on edge devices, allowing models to be deployed in environments where real-time processing is critical.

10. Hybrid AI Models

- **Amazon Neptune ML**: Integrates machine learning with graph databases, enabling advanced data analysis and knowledge graph applications.

Analytics

Amazon Athena

- **What:** Serverless interactive SQL query service for analyzing data stored in S3.

- **Key Features:**
  - Runs ANSI SQL queries on various formats (CSV, JSON, ORC, Avro, Parquet).
  - Integrates with Amazon QuickSight.
- **Pricing:** Billed per data scanned; DDL commands are free; costs reduced via compression, partitioning, and columnar formats.

Amazon EMR

- **What:** Managed cluster service for processing and analyzing big data with frameworks like Hadoop, Spark, Hive, and Flink.
- **Key Features:**
  - Scalable clusters on EC2; decouples compute (clusters) from storage (S3).
  - Supports machine learning, clickstream analysis, and real-time streaming.
- **Access:** EMR Console, CLI, SDK, and API.

AWS Glue

- **What:** Serverless ETL service for extracting, transforming, and loading data.
- **Key Features:**
  - Automates data cataloging and code generation (Python/Scala).
  - Processes semi-structured data using dynamic frames.
- **Use Cases:** Data migration, integration, and preparation for analytics.

Amazon Managed Service for Apache Flink

- **What:** Managed service for real-time stream processing using Apache Flink.
- **Key Features:**
  - No infrastructure management; high-speed, low-latency processing.
  - Multi-AZ deployments and API-driven lifecycle management.
- **Pricing:** Based on Kinesis Processing Units (KPUs) with additional costs for orchestration and storage.

Amazon Data Firehose

- **What:** Serverless service for capturing, transforming, and delivering streaming data.
- **Key Features:**
  - Synchronously replicates data across AZs.
  - Supports transformation, compression, and encryption.
  - Delivers data to S3, Redshift, Elasticsearch, and Splunk.
- **Latency:** Minimum 60 seconds or when 32 MB of data is collected.

Amazon Kinesis Data Streams

- **What:** Real-time data streaming service to collect, process, and analyze streaming data.
- **Key Features:**
  - Scalable shards (default retention 1 day, extendable to 7 days).

- Captures data from sources like websites, events, and social media.
- Each shard offers 1MB/sec input and 2MB/sec output capacity.

Comparison of Amazon Data Firehose & Amazon Kinesis Data Streams

Category	Amazon Data Firehose	Amazon Kinesis Data Streams
Purpose	Ingests, transforms, and delivers data to AWS destinations.	Collects and processes streaming data for custom applications.
Management	Fully managed; auto-buffers and delivers data.	Requires manual management of shards and consumers.
Latency	Near real-time (min 60 sec or on size threshold).	Low latency; processes data immediately.
Use Cases	Simple ingestion and loading into S3, Redshift, etc.	Real-time analytics and custom stream processing.
Data Retention	Temporary buffering only.	Configurable retention (24 hours to 7 days).
Integration	Directly integrates with AWS storage/analytics services.	Integrates with custom consumers via APIs and Kinesis Client Library.
Overhead	Minimal operational overhead.	Higher operational management required.

AWS Lake Formation:

- **What:** Service to create, manage, and secure data lakes.
- **Key Features:**
  - Automates data lake creation on S3.
  - Uses the AWS Glue Data Catalog for metadata management.
  - Integrates with IAM for fine-grained access control.
- **Integration:** Works with CloudWatch, CloudTrail, EMR, Redshift Spectrum, and Athena.

Amazon Managed Streaming for Apache Kafka (Amazon MSK)

- **What:** Managed service for running Apache Kafka clusters.
- **Key Features:**
  - Simplifies cluster management (server replacement, patching, ZooKeeper maintenance, multi-AZ replication).
  - Provides security via IAM, encryption at rest, and Kafka ACLs.

- **Integration:** Connects with AWS Glue, Kinesis Data Analytics, and Lambda.

Amazon OpenSearch Service

- **What:** Managed service to deploy, operate, and scale Elasticsearch/OpenSearch clusters.
- **Key Features:**
  - Direct access to Elasticsearch APIs.
  - Integrated with Kibana (visualization) and Logstash (log ingestion).
  - Auto-scales and auto-replaces failed nodes.
- **Pricing:** Billed per EC2 instance hour and attached storage; free data transfer within AZs.

Amazon QuickSight

- **What:** Scalable, cloud-based business intelligence (BI) service for interactive dashboards.
- **Key Features:**
  - Connects to diverse data sources.
  - Offers auto-forecasting, anomaly detection, and natural language query (Amazon Q).
  - Provides enterprise-grade security (SSO, row-level security, encryption).
- **Pricing:** Flexible per-user or capacity-based models; reader fees start at ~$3/month; no upfront licensing costs.