EC2 instance

General
General purpose instances provide a balance of compute, memory and networking resources. Can be used for multiple purposes.

Compute Optimized
Optimal for compute-bound applications that benefit from high-performance processors. Are ideal for high-performance compute-intensive applications.

Memory Optimized
Designed to deliver fast performance for workloads that process large datasets in memory.

Accellerated Computing
Uses hardware accellerators, or coprocessors, to perform some functions more efficiently than is possible in software running on CPUs. Graphic, Floating points calculations.

Storage Optimized
Designed for workloads that require high, sequential read and write access to large datasets on local storage.

EC2 load balancing

Elastic Load Balancer
automatically distribute incoming application traffic across multiple resources.

Messaging

SNS (Simple Notifications Services)
Used for Publisher/Subscriber pattern. Subscribers can be: web servers, emails, Lambda and more.

SQS (Simple Queue Services)
Use it to send messages between systems. A queue receives a message that can be consumed or deleted.

EC2 pricing

On-demand
Short-term and irregular workloads. You pay only when the instance runs, no commitment.

EC2 Saving Plans
Commit upfront for 1 or 3 years, you can save up to 66% compared to on-demand.

Reserved Instances
Applies a billing discount to on-demand. You can purchase 1 or 3 years Reserved or Convertible instances.

Spot Instances
Ideal for flexible workloads, it provides unused EC2 instances and you can save up to 90%. The instance can be claimed by AWS at any time with 2 minutes warning.

Dedicated Hosts
Physical hardware dedicated to you. You can reuse your own Software Licenses. Most expensive solution.

EC2 scaling

Auto-scaling
Automatically add or remove instances in response to changes like higher demand.

Dynamic Scaling
Scale instances based on dynamic demand.

Predictive Scaling
Scales instances based on predicted demand (time of the day).

Serveless Computing

AWS Lambda
Is a service that lets you run code without provisioning a server. You pay only for what you consume.

Containers
Package code and dependencies into a single object.

Containers

Amazon Elastic Container Service (ECS)
Highly available and scalable high-performance container management system that enables you to scale containerazed applications on AWS.

Supports Docker.

Amazon Elastic Kubernets Service (EKS)
Fully managed services to run Kubernetes on AWS. Use in alternative of ECS if you want to leverage the additional options available in Kubernetes.

AWS Fargate
Serverless compute engine for containers. Can be used with ECS and EKS and provisions or manage servers for your containers.

AWS Cloud Practitioner

Infrastructure Reliability

AWS Regions

Region
A region is a geographical area composed by multiple data-center. They are interconnected via Fiber connection.

Choose a Region
- Compliance requirements (where data should live)
- Proximity (how close you are)
- Feature availability (not all regions have same services)
- Pricing (some locations are more expensive to operate i.e. Brazil)

Availability Zones
A single data center or group of data centers, within a Region. ***Regional Scope Service*** when it runs across all Zones of a Region (Load Balancer)

Edge Locations
AWS Cloudfront is a CDN used to store cached copies of customer content closer to the customer than an availability zone. An Edge location is separate than a region.

AWS Outposts
An AWS Region operated by AWS inside Customer hardware building.

Connectivity

VPC (Virtual Private Cloud)
Isolated Network within your AWS, divided by multiple Subnet.

IGW (Internet Gateway)
A Front door which expose your VPC to the internet.

Virtual Private Gateway
A VPN Connection between a well-know private network and an AWS VPC. Can be affected by internet bandwith and public traffic. Shared with other customers.

AWS Direct Connect
A direct doorway to the AWS VPC. A completely private directly connected to your VPC. This is customer dedicated.

Provisioning

API
Application Programming Interface. In AWS anything you do is an API. From creation to manipulation of resources.

AWS Management Console
Browser based that allows you to manage and create resources. View Bills and other non-technical resources.

AWS Command Line Interface
Makes API call using a terminal on your machine, for automation and repetable actions.

AWS SDK
Allows you to interact using programming languages, which allows programmer to interact with AWS.

AWS Elastic Beanstalk
Serverless EC2 instances, which execute customer code/ configuration and auto-build the environment. You do not have to provision the resources.

AWS CloudFormation
Infrastructure as code tool, using JSON or YAML. Cloud formation templates allows you to define what to build.

Subnet and Network Access

Subnet
It is a section of the VPC used to group resources based on certain security rules:
- **Public subnet**
  It is used to make the resources visible outside the VPC
- **Private subnet**
  It is used to make the resources visible only within the VPC

ACL (Access Control Lists)
It is a virtual firewall which control inbound and outbound traffic at the subnet level. The control is **stateless**, which means is conducted for every single packet/ip, everytime. ***By default all traffic is allowed***.

Secure Groups
A secure group is a virtual firewall used to control inbound and outbound traffic, solely for a specific EC2 instance. ***By default, all inbound traffic is denied***. It is **stateful**, it remembers decisions made on previous incoming packets.

Global networking

Route 53
It is a service which redirect requests from a DNS Alias into a resource on your VPC.

Cloud Front
Hosted on Hedge Location, it is a CDN service which host hedge content and deliver it to users based on nearest location.

EBS (Elastic Block storage)

Instance store
It is a ***temporary*** vritual drive mounted to the EC2 instance which gets destroyed as soon as the instance is stopped or destroyed.

EBS (Elastic Block Store)
It is a dedicate virtual drive which can be mounted to an EC2 instance and does not get destroyed. Used for data that needs to be persisted.

- Snapshot is the technology used to have incremental back-ups on an EBS

EBS vs S3

- Block storage, good for incremental change of a file
- Constant changes, on the same file
- Web enabled, good for web hosting static files
- Regional distributed (no need of backup)
- Objects, occasionally change

EBS vs EFS (Elastic File system)

- Does not scale, fixed size
- Specific to the same zone of an EC2 instance
- Scales automatically if data increase
- Regional, can be used by all Availability Zones
- Multiple EC2 can access it

Relational Databases Services (RDS)

Managed service which automate hardware, database setup, patching and backups.

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

Amazon Aurora
Enterprise Relational Database which support MySQL and PostgreSQL. 5 times faster than MySQL and 3 times than PostgreSQL. Replicate 6 copies within 3 Availability Zones and Backup over S3.

Database Migration Services (AWS DMS)

Allows you to migrate any sort of database. It can migrate (and keep operational) from an on-prem database to an AWS RDS one, which does not have to be the same

Amazon S3 (Simple Storage Service)

Used to store Files in binary format. You can group files into **Buckets** and max file size is of **5TB**.
Available ***classes***:

- **S3 standard**
  Used for frequent data access
  Stores the data in minimum 3 availability zones
- **S3 Standard-IA (Infrequent Access)**
  Infrequent data access
  Lower storage price, higher retrieval price
- **S3 One Zone-IA (One Zone Infrequent Access)**
  Only one availability zone
  Lower storage price
- **S3 Intelligent Tiering**
  Unknown data storage strategy
  Small monthly fee for monitoring
- **S3 Glacier**
  Good for archive (Backup)
  Retrieval is slow (minutes to hours)
- **S3 Glacier Deep Archive**
  Deep archiving and data retention
  Takes up to 12 hours for retrieval

Amazon DynamoDB

Serverless, NoSQL Database used to store non relational data. It contains **Tables** which store **Items** that have **Attributes**.

It replicates across Availability Zones and highly performant. Very good for high rate access and dynamic schema requirements. It is fully managed by AWS. It auto-scale to ensure best performances but also best costs at all the time.

Amazon Redshift

Data warehouse service used for big data analytics. It can collect data from many sources and understand relationships and data trends.

Amazon Redshift

Additional Database services:

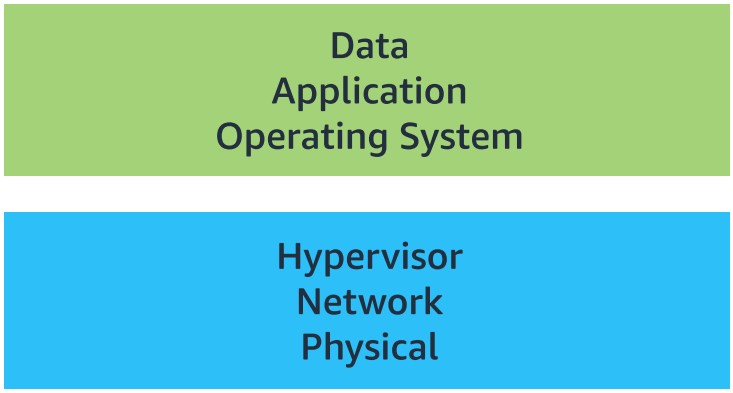
- **Amazon DocumentDB**
  Document service to support MongoDB
- **Amazon Neptune**
  Graph database service for highly connected apps, social networking
- **Amazon Quantum Ledger Database (QLDB)**
  Immutable Ledger where any entry cannot be removed from the audit
- **Amazon ElasticCache**
  Cache layer in MemCache or Redis
- **Amazon DynamoDB Accellerator**
  In-memory cache for DynamoDB

Shared Responsibility model

It is a model where AWS and the Customers share responsibility about securing AWS and their products.

Customers
Customer Data, Platforms, Applications, Operating systems, Network and Firewall, Encryption and Traffic.

AWS
Software, Compute, Storage, Database, Hardware provided, Regions, Availability Zones and Edge Locations



Identity Access Management (IAM)

- **Root user**
  The user created when you generate an AWS account. It has unlimited privileges
- **IAM User**
  It is an identity that represents a physical person or an application interacting with AWS services
- **IAM Group**
  A collection of IAM users with certain IAM policies associated to them
- **IAM Policy**
  It is a JSON document that allows or denies access to resources and services. It uses the ***least privileges concept***
- **Roles**
  It is a temporary IAM user which grant temporary access to certain permissions (***breaking glass***)

Additional Security

AWS Key Management Service (AWS KMS)
Allows encryption at REST and in TRANSIT. The customer can decide the type of encryption and the level of management with IAM.

AWS WAF
Web application firewall that lets you monitor web traffic and network requests. It uses ACL (Access Control List) to allow or deny incoming traffic. Controlled via IP addresses

Amazon Inspector
Executes automated security assessments and provide a list of findings marked with various priorities.

Amazon Guard Duty
Is a service that provides intelligent threat detection for AWS infrastructure. It monitors network activity and account behaviours. On Account level

AWS Organizations

A central location to manage multiple AWS accounts. You can manage:

- Centralize management of AWS accounts
- Consolidated billing for all AWS accounts
- Hirerachical grouping for security and budget constraints
  - Organization Units (OU)
- Control AWS APIs with SCP (Service Control Policies)
  - Restrict what a User and Role can do in an AWS account: OUs or Individual Accounts

Compliance

AWS Artifact
It is a service that provides on-demand access to AWS security and compliance reports for audit.

- **AWS Artifact Agreements**
  Special agreements signed between Client and AWS, for customers subject to specific regulations
- **AWS Artifacts Reports**
  Compliance reports from third-party auditors related to AWS compliance status

Denial of service attacks (DDoS)

A DDoS try to overwhelm your application in order to shut it down.
AWS WAF
Web Application firewall to filter traffic marked as “bad actor”. Proactive defend your system
AWS Shield Standard
Automatically protect customers at no cost
AWS Shield Advanced
Paid service which provide more details on DDoS attachs and integrates with many AWS services.

Amazon CloudWatch

Web service which enables you to monitor and manage various metrics. Metrics are sent from AWS services into AWS CloudWatch.

CloudWatch Alarms
Alert triggered when a CloudWatch threeshold is reached for a specific metric.

- Dashboard can be used to monitor those metrics and alerts

Amazon CloudTrail

It records all APIs calls to your AWS account. Update time is about **15 minutes** and it contains all audit information related to who and when your AWS API has been called.

CloudTrail Insight is used to detect unusual APIs calls, to ensure no unexpected activity is happening on your account automation, for example.

Amazon Trusted Advisor

Web service that inspects and AWS environment and provides in real-time reccomendations in accordance to AWS best practices. You can hook **Alerts actions**.

- Cost optimization
- Performance
- Security
- Fault tolerance
- Service limits

AWS Cloud Practitioner

Pricing
Support

Amazon Free Tier

There are three types of free:

- **Always Free**
  Free forever to all customers. For example 1 million of AWS lambda requests
- **12 Months Free**
  The are free for the first 12 months after sign-up
- **Trials**
  Specific to a service which can be free the first 90 days, for example

Consolidation

AWS Organizations allow you to consolidate multiple AWS Account billings. You can receive a single bill for all the AWS Account within the same AWS Organization.

Beware, **max 4 accounts** can be within an Organization.

AWS Budget

AWS budget ***updates 3 times*** a day. You can set alerts if your budget is reaching the limit or exceeding it. Pro-active if you are going to spend more than planned.

AWS Marketplace

It is a curated digital catalog which streamline services you can use on AWS with various payment options.

- You don't need to maintain core images, for example by using one-click deployment

Categories:

- Business Applications
- Data & Analytics
- DevOps
- Infrastructure Software
- Internet of Things (IoT)
- Machine Learning
- Migration
- Security

Pricing Models

There are three types of models:

- **Pay for what you use**
  You pay the services that you are using, without a binding contract
- **Reserve**
  You reserve a certain amount of services and reduce up to 72% the costs
- **Volume based discounts**
  Some pricing models are tiered to allow you pay less

AWS Billing
AWS Billing and Cost Management is a dashboard which allow you to control your current spending and forecasts but also previous invoices and budgets alerts.

AWS Cost Explorer

It allows you to visualize, understand and manage AWS costs and usage. You can apply filters and groups.

AWS Support

Basic Support
Provided 24/7 to any customer for free:

- Documentation and Whitepapers
- Support forums
- AWS Trusted Advisor
- AWS Personal Health Dashboard

Developer Support

- Basic
- Open unrestricted number of Support cases

Business Support

- AWS Trusted Advisor
- Phone access with 4 hours SLA and 1 hour SLA for system down
- Access to Infra event management (like large events)

Enterprise Support

- Mission Critical
- 15 minutes SLA
- TAM (Dedicated Technical Account Manager)
- TAM are specialized in monitoring customer and help in architecture reviews

AWS Cloud Adoption Framework (AWS CAF)

AWS CAS organizes on a high level, 6 perspectives, which are areas of focus and responsibilities. Each perspective has inputs which will generate an AWS CAF Adoption Framework.

- **Business Perspective**
  Ensures the IT aligns with the Business needs and deliver key results
- **People Perspective**
  Supports the development of an organization-wide change due to cloud adoption
- **Governance Perspective**
  Focus on the skills required to align IT strategy with business strategy
- **Platform Perspective**
  Includes principles and patterns for implementing new cloud solutions
- **Security Perspective**
  Ensure the organization meets the security objectives and audability requirements
- **Operations Perspective**
  Help to enable, run, use, operate and recover IT workloads to the SLA agreed with business

AWS Snow Family

A collection of physical devices that help the physically move exabytes of data into AWS datacenters:

- **AWS Snowcone**
  Small, rugged and secure data transfer device
  2CPUs, 4GB RAM and carries 8TB of data
- **AWS Snowball**
  - **Snowball Edge Storage Optimized**
    80TB HDD/S3 and 1TB SSD
    Compute 40 vCPUs and 80GB RAM
  - **Snowball Edge Compute Optimized**
    42TB HDD/S3 and 7.68TB SSD
    Compute 52 vCPUs and 208GB RAM
    NVIDIA V100 GPU
- **AWS Snowmobile**
  Exabyte data transfer service. 100 petabytes of data for each Snowmobile

Migration strategies (6R)

When moving to Cloud, there are 6 major strategies that can be implemented:

- **Rehosting**
  Lift-and-shift. Move an application without changes
- **Replatforming**
  Lift-thinker-shift. Make some cloud optimizations to realize a benefit
- **Refactoring**
  Reimaging the application by using cloud-native features
- **Repurchasing**
  Move from a traditional license to a SaaS model. For example a CRM
- **Retaining**
  Keep critical application into the source environment. No migration
- **Retiring**
  Decommissioning application that are no longer needed

AWS Innovation

Multiple paths are available in AWS to bring innovation.

Serverless Applications
Applications that do not require provision, maintain or administration of servers.

Artificial Intelligence
- *Amazon Transcribe* - convert speech to text
- *Amazon Comprehend* - discover patterns in text
- *Amazon Fraud Detector* - identify frauds online
- *Amazon Lex* - chabots

Machine Learning
SageMaker empower to build, train and deploy ML models

AWS Well-Architected Framework

Composed by 5 pillars, it helps on how to design an operate a reliable, secure, efficient and cost-effective AWS Cloud.

- **Operational excellence**
  The ability to run and monitor systems
- **Security**
  First priority, semplifies integrity of data and encrypting of content
- **Reliability**
  Ability to recover, acquire additional resources and mitigate disruptions
- **Performance efficiency**
  Ability to use compute resources efficiently to meet system requirements
- **Cost optimization**
  It is the ability to run systems to deliver business value