# Amazon Cloud Computing

The on demand delivery of IT resources over the internet with pay as you go pricing.

# Amazon EC2 Service

Amazon EC2 (Elastic Compute Cloud) is a web service that provides secure, resizable virtual servers (instances) in the cloud. It allows users to run applications, store data, and scale resources up or down as needed, paying only for the compute power they use.

## Types

1. General Purpose: Balanced CPU, memory, and storage for a variety of workloads, such as web servers and app development.

2. Compute Optimized: High CPU performance for compute intensive tasks like gaming servers and batch processing.

3. Memory Optimized: Extra memory for data intensive applications, such as large databases and in memory caching.

4. Storage Optimized: High disk performance for tasks that require fast local storage, like big data and NoSQL databases.

5. Accelerated Computing: GPU and specialized hardware for machine learning, graphics rendering, and complex scientific computations.

## Pricing

Amazon EC2 offers flexible pricing options, which include:

1. On Demand : Pay only for the time the instance is running, with no upfront cost.

Example : Launch a virtual server for a few hours to test an app, paying hourly or per second pricing.

2. Savings Plans : Commit to a 1 or 3 year term to get discounted rates (up to 72% savings) on On Demand prices.

Example : A business commits to a 1 year term for a 30% discount on its web server costs.

3. Reserved Instances : Reserve instances for a 1 or 3 year term, reducing costs (up to 75% savings).

Example : A company reserves an instance for a 3 year term to save on a consistently needed database server.

4. Spot Instances : Use spare capacity at a reduced price (up to 90% off) but can be interrupted when demand rises.

Example : A data analyst uses spot instances to perform large computations overnight at a lower cost.

5. Dedicated Hosts : Rent physical servers for security or compliance requirements.

Example : A financial company needs dedicated servers for compliance, paying for exclusive hardware.

These pricing models give flexibility based on usage patterns and budget needs.

Here’s a simple overview of each:

1. \*\*Elastic Load Balancer (ELB)\*\*:

- Distributes incoming traffic across multiple EC2 instances to improve application availability and fault tolerance.

- Types include Application Load Balancer (ALB), Network Load Balancer (NLB), and Gateway Load Balancer (GLB), each designed for different use cases.

2. \*\*Simple Queue Service (SQS)\*\*:

- A fully managed message queuing service that lets you decouple and scale microservices or distributed systems.

- It ensures messages are stored reliably until they’re processed, allowing for asynchronous communication between services.

3. \*\*Simple Notification Service (SNS)\*\*:

- A fully managed publish-subscribe messaging service that sends notifications to multiple subscribers at once.

- Useful for sending alerts, pushing updates, or triggering workflows across distributed systems.

Together, these services help manage, scale, and streamline distributed application architecture on AWS.

# Selecting Region

Compliance: Location bounded data 🡪 (data must not get out of UK)🡪 Choose London.

Proximity: The nearest the faster.

### **Available services within a Region**

### **Pricing**

# **Amazon CloudFront**

a content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to users worldwide with low latency and high transfer speeds. It caches content at edge locations globally, reducing the load on origin servers and speeding up content delivery to users closer to these locations.

# Edge Location

An **edge location** is a site that Amazon CloudFront uses to store cached copies of your content closer to your customers for faster delivery.

# **Amazon Outposts**

a fully managed service that brings AWS infrastructure, services, and tools to your on-premises data center or facility. It allows you to run AWS services locally, giving you a consistent hybrid experience with low-latency access to on-premises applications.

# AWS Elastic Beanstalk

a platform-as-a-service (PaaS) that helps you deploy and manage applications without needing to handle the underlying infrastructure. You simply upload your code, and Elastic Beanstalk automatically manages the deployment, load balancing, scaling, and monitoring for you.

# AWS CloudFormation

a service that allows you to model, provision, and manage AWS resources using code. By defining infrastructure as code (IaC) in templates (JSON or YAML), CloudFormation automates the setup of complex AWS environments, making it easy to create, update, and replicate resources in a consistent manner.

VPC: Your private network in AWS to control and isolate resources.

Subnet: IP range within a VPC; public subnets connect to the internet, private stay internal.

Direct Connect: Private, fast, and secure link between your data center and AWS.

* **Security Groups**: Virtual firewalls for controlling inbound and outbound traffic to AWS resources. They allow or deny access at the instance level.
* **Network ACLs (Access Control Lists)**: Optional layer of security for controlling traffic at the subnet level. They allow or deny IP-based traffic rules for entire subnets.

Use **Security Groups** for instance-specific access and **ACLs** for broader subnet security.

* **DNS (Domain Name System)**: System for translating domain names (like example.com) into IP addresses so browsers can load resources.
* **Route 53**: AWS’s scalable DNS and domain registration service. It routes traffic to AWS resources and external websites and can also be used for health checks and load balancing.
* **CDN (Content Delivery Network)**: Network of servers that caches and delivers content from locations closer to users, reducing latency.

Together, **Route 53** handles routing, while a **CDN** like CloudFront delivers content quickly to global users.

4o

Top of Form

Bottom of Form