EC2 Pricing Models in AWS

There are 5 main pricing models for EC2:

1. On-Demand Instances

- Pay-as-you-go model: no long-term commitment.
- Best for: short-term workloads or unpredictable traffic.
- Highest cost per hour, but very flexible.

2. Reserved Instances (RIs)

- Commitment-based (1 or 3 years).
- Up to 75% cheaper than On-Demand.
- Three types:
    - **Standard RIs** – fixed usage.
    - **Convertible RIs** – change instance type.
    - **Scheduled RIs** – for specific time windows.

3. Spot Instances

- Unused EC2 capacity at up to 90% discount.
- Can be terminated by AWS anytime if demand increases.
- Ideal for fault-tolerant or batch jobs.

4. Savings Plans

- Commit to a consistent amount of usage ($/hr) for 1 or 3 years.
- More flexible than RIs (can change instance types, OS, regions).
- Two types:
    - Compute Savings Plan
    - EC2 Instance Savings Plan

5. Dedicated Hosts

- Physical servers fully dedicated to your use.
- Helps meet compliance or licensing needs (e.g., BYOL).
- Expensive but offers complete hardware control.

EC2 Instance Types / Families

EC2 instance types are grouped into **families** based on **use case and hardware optimization**:

Family	Use Case	Examples
General Purpose	Balanced CPU, memory, networking	t3, t4g, m5, m6i
Compute Optimized	High-performance CPU tasks	c5, c6g, c7i
Memory Optimized	In-memory databases, large caches	r5, r6g, x2idn
Storage Optimized	High storage I/O (e.g., big data, OLAP)	i3, i4i, d3en
Accelerated Computing	GPUs, FPGAs (ML, HPC)	p4, g5, inf2, f1
High Memory	SAP HANA, in-memory DBs	u-6tb1, u-12tb1
Mac Instances	Build, test Apple macOS apps	mac1, mac2

Example Format:

Instance type m5.large

- Family: m5 (General purpose)

- Size: large (determines vCPU, RAM)