

What are the types of AWS EC2 instances and their typical use cases?

What is Amazon EC2?

Amazon EC2 (Elastic Compute Cloud) is a web service that provides resizable compute capacity in the cloud. It allows users to run virtual servers (called instances) on-demand, helping scale applications up or down quickly and efficiently.

What are the types of EC2 instances?

EC2 instances are categorized by **use case**. Main types include:

- General Purpose Balanced compute, memory (e.g., t3, m6g)
- Compute Optimized For compute-intensive workloads (e.g., c6q)
- **Memory Optimized** For RAM-heavy tasks (e.g., r6q, x2idn)
- Storage Optimized High IOPS or throughput (e.g., i4i, d3en)
- Accelerated Computing GPU/FPGA workloads (e.g., p4, inf2)

1. General Purpose Instances

Best for: Balanced compute, memory, and networking.

Use Cases:

- Web Servers: Hosting websites or REST APIs (e.g., a company website using Apache or Nginx).
- **Dev/Test Environments**: Developers testing small applications or Docker containers.
- Microservices: Deploying light-weight containers in ECS/EKS.

✓ Scenario:

A startup launches its MVP web app using t3.medium because it's low-cost and good enough for 1000 daily users.

2. Compute Optimized Instances

Best for: Applications that need a lot of CPU power.

```
C5 | C5a |
C5ad | C5d |
C5n | C6a |
C6g | C6gd |
C6gn | C6i |
C6id | C6in |
C7a | C7g |
C7gd | C7gn |
C7i | C7i-flex |
C8g | C8gd
```

Use Cases:

- High-performance web servers
- Batch processing
- Scientific modeling
- Media encoding/transcoding

✓ Scenario:

An e-commerce platform runs a product recommendation engine that requires fast CPU cycles. They use c5.2xlarge to handle millions of real-time calculations during peak sale hours.

3. Memory Optimized Instances

Best for: Applications with high memory needs (RAM).

```
R5 | R5a | R5ad | R5b | R5d |
R5dn | R5n | R6a | R6g | R6gd |
R6i | R6idn | R6in | R6id | R7a |
R7g | R7gd | R7i | R7iz | R8g |
R8gd | U-3tb1 | U-6tb1 | U-9tb1 |
U-12tb1 | U-18tb1 | U-24tb1 |
U7i-6tb | U7i-8tb | U7i-12tb |
U7in-16tb | U7in-24tb | U7in-
32tb | U7inh-32tb | X2gd | X2idn |
X2iedn | X2iezn | X8g | z1d
```

Use Cases:

- In-memory databases like Redis or Memcached
- Real-time big data analytics
- High-performance databases like SAP HANA, MySQL, or PostgreSQL
- Machine learning model training

Scenario:

A bank uses r5.4xlarge to run its PostgreSQL databases for storing customer transactions. They chose memory-optimized instances to avoid disk I/O bottlenecks.

4. Storage Optimized Instances

Best for: High disk throughput or low-latency storage access.

```
D3 | D3en | I3en | I4g |

Examples I4i | I7ie | I8g | Im4gn

| Is4gen
```

Use Cases:

- Data warehouses
- Log processing
- Real-time analytics with large datasets
- NoSQL databases like Cassandra or MongoDB

✓ Scenario:

A video streaming company processes logs from millions of devices. They use i3en.large with fast NVMe SSDs to ingest and analyze logs in real time.

5. Accelerated Computing Instances

Best for: GPU/FPGA use, AI/ML, deep learning, and HPC.

Examples p4, p3, g5, inf2, f1

Use Cases:

- Training deep learning models with TensorFlow, PyTorch
- Video rendering and 3D simulations
- Autonomous driving algorithms
- Crypto mining and genomics research

✓ Scenario:

An AI startup is training a computer vision model to detect road signs from dashcam footage. They use a p4d instance with NVIDIA A100 GPUs to cut model training time from 3 days to 8 hours.

Quick Summary Table:

Instance Type	Best For	Example Use Case
General Purpose	Balanced workloads	Web servers, microservices
Compute Optimized	High CPU performance	Recommendation engines, API servers
Memory Optimized	RAM-heavy apps	Databases, analytics, caching
Storage Optimized	Fast local storage / IOPS	Data warehouses, logging pipelines
Accelerated Computing	AI/ML, GPU-intensive tasks	Deep learning, video rendering, HPC

Which instance type is best suited for balanced workloads like web servers and microservices?

- A. Compute Optimized
- B. Memory Optimized
- C. General Purpose
- D. Storage Optimized
- Correct Answer: C. General Purpose

You are building a recommendation engine that requires high CPU performance. Which instance type should you choose?

- A. Memory Optimized
- B. Compute Optimized
- C. General Purpose
- D. Accelerated Computing
- Correct Answer: B. Compute Optimized

Which instance type is designed for applications that require large amounts of RAM, such as databases and analytics?

- A. General Purpose
- B. Accelerated Computing

- C. Memory Optimized
- D. Storage Optimized
- **✓** Correct Answer: C. Memory Optimized

For which use case is a storage optimized instance most suitable?

- A. Running a web application
- B. Performing deep learning
- C. Managing a data warehouse
- D. Hosting a microservice
- Correct Answer: C. Managing a data warehouse

If your workload includes deep learning or video rendering, which instance type should you select?

- A. Compute Optimized
- B. Storage Optimized
- C. General Purpose
- D. Accelerated Computing
- ✓ Correct Answer: D. Accelerated Computing