Amazon EC2 Instance Types

Introduction

Amazon EC2 (Elastic Compute Cloud) provides a wide selection of instance types optimized to fit different use cases.
Each instance type offers different compute, memory, storage, and networking capacities, and is grouped into instance families based on these capabilities.

General Purpose Instances

These provide a balance of compute, memory, and networking resources. They are ideal for applications that use these resources in equal proportions.
Examples:
- t4g, t3, t3a: Low-cost, burstable performance.
- m7g, m6g, m5, m5a: Balanced performance for most workloads.

Compute Optimized Instances

Designed for compute-intensive workloads such as batch processing, high-performance web servers, and scientific modeling.
Examples:
- c7g, c6g, c6i, c5: Ideal for compute-bound applications.

Memory Optimized Instances

Optimized to deliver fast performance for workloads that process large data sets in memory.
Examples:
- r7g, r6g, r5, x2idn, u-6tb1.metal: Ideal for in-memory databases, real-time big data analytics.

Amazon EC2 Instance Types

Storage Optimized Instances

Designed for workloads that require high, sequential read and write access to very large data sets on local storage.

Examples:

- i4i, i3, i3en, d3: Good for NoSQL databases, distributed file systems.

Accelerated Computing Instances

Use hardware accelerators, or co-processors, to perform functions such as floating point number calculations, graphics processing, or data pattern matching more efficiently.

Examples:

- p5, p4, inf2, trn1: Ideal for ML, GPU computing, graphics workloads.