AWS Instance Cases

Date: March 1st, 2025

Case Study 1: General Purpose Instances

Business Case: A startup is developing a web application that serves dynamic content to users. They need a balanced compute, memory, and networking option for hosting their website.

Recommended Instance Type: AWS EC2 T3 or M5 instances.

Scenario A: A SaaS-based project management tool that allows users to collaborate in real time. The application requires a balance of CPU and RAM without over-provisioning resources.

Scenario B: A small e-commerce website with moderate traffic that needs scalable performance without excessive costs.

Instance Series: T3, T4g, M5, M6i

Case Study 2: Compute Optimized Instances

Business Case: A financial services company runs risk analysis simulations that require high compute power.

Recommended Instance Type: AWS EC2 C5 or C6i instances.

Scenario A: A hedge fund runs Monte Carlo simulations for portfolio risk assessment, requiring fast CPU performance without excessive memory allocation.

Scenario B: An AI-powered chatbot application that performs extensive natural language processing (NLP) in real-time.

Instance Series: C5, C6i, C7g

Case Study 3: Memory Optimized Instances

Business Case: A healthcare company processes large genomic datasets to analyze patient DNA and detect disease markers.

Recommended Instance Type: AWS EC2 R5 or X2idn instances.

Scenario A: A bioinformatics company loads entire DNA sequences into memory for real-time analysis, needing high memory capacity and low latency.

Scenario B: An in-memory caching solution for a high-traffic social media platform to improve data retrieval speeds.

Instance Series: R5, R6g, X2idn, X2iedn

Case Study 4: Accelerated Computing Instances

Business Case: A gaming company is developing an AI-powered NPC engine that requires GPU acceleration for machine learning inference.

Recommended Instance Type: AWS EC2 P4 or G5 instances.

Scenario A: A deep-learning-based character animation system generates realistic movements in a 3D environment, utilizing GPUs for fast inference.

Scenario B: A video rendering and encoding service that processes large media files at high speed.

Instance Series: P4, P5, G5, G6

Case Study 5: Storage Optimized Instances

Business Case: A video streaming service needs to store and process large amounts of media files efficiently.

Recommended Instance Type: AWS EC2 I3 or D3 instances.

Scenario A: A media company processes and serves high-resolution video content with fast SSD-backed storage to ensure low-latency access for millions of users.

Scenario B: A log analytics platform that processes and indexes real-time log data for security monitoring.

Instance Series: I3, I4i, D3, D2

Case Study 6: HPC Optimized Instances

Business Case: A scientific research lab runs climate simulations that require massive computational power and high-speed interconnects.

Recommended Instance Type: AWS EC2 Hpc6id or C6gn instances.

Scenario A: A meteorology department performs complex weather modeling that requires parallel computing on large datasets, needing optimized HPC resources.

Scenario B: An automotive company running computational fluid dynamics (CFD) simulations to optimize car aerodynamics.

Instance Series: Hpc6id, Hpc7g, C6gn

Case Study 7: Instance Features

Business Case: A fintech company is scaling its infrastructure and needs to select the right AWS instance type based on networking, burstable performance, and cost-effectiveness.

Recommended Approach: Leverage AWS EC2 Instance Features like enhanced networking (ENA), burstable performance (T-series), and dedicated hosts.

Scenario A: A digital bank uses T4g instances with burstable CPU for intermittent heavy loads, ensuring efficient cost management.

Scenario B: A logistics company requiring enhanced networking for real-time vehicle tracking and fleet management.

Instance Series: T4g, M6i, C6i, Dedicated Hosts

Case Study 8: Measuring Instance Performance

Business Case: An e-commerce company needs to benchmark the best AWS instance type for its recommendation engine.

Recommended Approach: Utilize AWS CloudWatch metrics and AWS Compute Optimizer to assess CPU, memory, and network performance.

Scenario A: An online retail platform runs A/B tests comparing M6i and C6i instances to determine which delivers the best product recommendation latency under peak load conditions.

Scenario B: A travel booking website analyzing the impact of different instance types on search response times under heavy traffic.

Instance Series: M6i, C6i, R5, X2idn

Expectation On Each Case Study

1. Over view of recommended instance (Introduction detailing about the instance type based on series)
2. Characteristics
3. Why they are suitable
4. Consideration detailing the incident series
5. Comparison and selection
6. Key consideration supporting the business case
7. Conclusion