EC2 Instance Type	What You Need To Know
General Purpose	• Offers a balanced ratio of CPU to memory • Good fit for general-purpose applications that use CPU and memory in equal proportions - for example, web servers with low to medium traffic or small to medium databases
Compute Optimized	• Optimized for CPU-intensive workloads • High ratio of CPU to memory • Good fit for web servers with medium traffic, batch preprocessing, network appliances, and application servers
Memory Optimized	• Comes with a high memory-to-CPU ratio • Great fit for production workloads - for example, database servers, relational database services, analytics, and larger in-memory caches
Storage Optimized	• Good choice for workloads that require heavy read/write operations and low latency • Thanks to high disk throughput and IO, storage-optimized instances come in handy for Big Data, SQL and NoSQL databases, data warehousing, and large transactional databases
Accelerated Computing	• Uses hardware accelerators (co-processors) to carry out functions like data pattern matching, graphics processing, and floating-point number calculations better than software running on CPUs • Great pick for machine learning (ML) and high-performance computing (HPC)
Inference Type	• Built to support machine learning applications • AWS EC2 Inf1 promises up to 30% higher throughput and 45% lower cost per inference than AWS EC2 G4 instances • Includes 16 AWS Inferentia chips, 2nd generation Intel® Xeon® Scalable processors, and up to 100 Gbps networking