When Virtualization Trumps Public Cloud - The Case for Predictable, Secure, High-Performance AI

The public cloud has revolutionized computing, offering unmatched flexibility, scalability and a pay-as-you-go model. However, there are situations where a virtualized private cloud environment may actually be the better choice, especially when it comes to running AI and machine learning workloads. Here's why:

Predictable, Steady-State Loads

Many AI and analytics applications, once deployed to production, have relatively stable and predictable resource requirements. For these steady-state workloads, the scalability of public cloud is often not needed.
Virtualization allows you to precisely provision the compute, memory and storage resources to match the workload. This avoids overprovisioning and delivers better cost efficiency compared to public cloud's pay-per-use model for flat or predictable usage.

Use cases:

- Payroll processing systems that run on a fixed schedule (e.g. bi-weekly or monthly) and have consistent resource requirements based on the number of employees. Virtualizing these workloads allows consolidation while maintaining steady performance.
- Batch jobs and ETL (extract, transform, load) pipelines that process a known quantity of data on a recurring basis, such as end-of-day financial transactions or daily data warehouse updates. The predictable nature makes them good candidates for virtualization.

Higher Performance and Lower Latency

AI workloads like deep learning and large-scale data processing are extremely sensitive to system performance. Even small amounts of latency can significantly slow down model training and inference.
With virtualization, the hardware is purpose-built and dedicated to your workloads. There is no "noisy neighbor" effect from other tenants. You have full control over the infrastructure stack and can tune it for optimal performance.
Virtualized environments can deliver the highest levels of performance and lowest latency compared to public cloud, which is critical for keeping AI applications running at peak speed.

Use cases:

- High-frequency trading platforms that require the lowest possible latency to execute trades faster than competitors. Virtualizing the trading application on dedicated, high-performance infrastructure delivers consistent ultra-low latency.
- Real-time bidding engines for digital advertising that must process and respond to ad requests within strict time constraints (100-200ms). Virtualization allows fine-tuned resource allocation to meet latency SLAs.
- Online gaming servers that must quickly process moves from a large number of simultaneous players to provide a responsive experience. Virtualized gaming servers can be optimized for high I/O performance and low network latency.

Tighter Security and Control

With the rapid adoption of AI, new risks are emerging - from attacks on machine learning models to potential for malicious use of AI. For enterprises in regulated industries, the security and governance of AI systems is paramount.
Virtualization provides the ability to deploy AI workloads in your own data centers or in a private hosted environment. This gives you full control over security, privacy and compliance. You can implement security best practices like data encryption, access controls, network isolation and real-time monitoring across the entire AI lifecycle.

Keeping sensitive data and AI models in a private virtualized environment helps meet strict data residency requirements and industry-specific regulations. It also reduces the risk of intellectual property theft.

Use cases:

- Payment processing applications that handle sensitive cardholder data and must comply with strict PCI DSS regulations. Virtualizing payment workloads in a tightly-controlled environment helps meet security and compliance requirements.
- Healthcare systems storing electronic medical records (EMR) that are subject to data privacy laws like HIPAA. Virtualization provides an additional layer of isolation and security to protect patient information.
- Government and military applications dealing with classified information that cannot be put in shared public cloud environments. Secure on-premises virtualization gives them full control over the infrastructure.

The Best of Both Worlds

While virtualization has unique advantages for AI, this doesn't mean completely avoiding public cloud. The ideal approach for many enterprises is a hybrid model - deploying steady-state, mission-critical AI workloads on virtualized infrastructure while leveraging public cloud for development, testing and burst capacity.
Virtualization platforms now offer tools to unify management across on-premises and cloud environments. This allows enterprises to pick the optimal platform for each AI use case while maintaining consistent operations.

As AI becomes critical for more and more business processes, enterprises need an AI infrastructure that delivers predictable performance, strong security and cost efficiency at scale. For a significant portion of production AI deployments, virtualization will likely be the most suitable choice.