



Major Service Provider in Asia Expands AI Service Offerings and Market Reach with Cisco Secure AI Factory with NVIDIA

For mobile network operators, success is defined not only by delivering a superior service experience, but by constantly innovating to help customers apply the latest technologies to meet new challenges. When a customer reached out with a new service request, a leading service provider in Asia saw a unique opportunity to enhance its AI service offerings.

The customer was seeking a service provider that could offer GPU-as-a-Service, to support its new initiative to increase the availability of AI infrastructure. To meet the customer's needs, the service provider would need to dramatically expand its own AI capacity. Serving a large customer base, the organization required a full-stack solution that would encompass compute, networking, security, and storage. The new solution would need to roll out fast and be flexible to offer many options for monetization.

A complete, one-stop solution from Cisco

After considering options from a variety of vendors, the service provider chose a Cisco Secure AI Factory with NVIDIA deployed via Cisco Cloud Reference architecture. At the heart of the networking component are Cisco Nexus 9000 400G switches with Nexus Dashboard. The new Cisco Cloud Reference Architecture is based on NCP design tenets for Cisco Silicon One-based switches, with an add-on license for deeper integration with NVIDIA Spectrum-X Ethernet technology. This add-on license enables the NVIDIA Spectrum-X adaptive routing capability to work seamlessly with Cisco Nexus 9300 Series Switches and NVIDIA Spectrum-X Ethernet SuperNICs. The combined solution delivers low latency and end-to-end congestion management to significantly enhance network performance for AI networks and accelerate job completion.

Cisco Nexus Dashboard provides a unified management dashboard that integrates smoothly with the provider's existing infrastructure and orchestration platform, to simplify and automate operations. The compute functionality of the solution was built around Cisco UCS C885A M8 servers.

The full-stack solution also includes Cisco AI Defense, Cisco Hypershield, Cisco Secure Firewall 4200 Series appliances, and Cisco Virtual Firewall, which deliver robust protection against cybersecurity risks.

Although the service provider considered competitive solutions across the compute, networking, and storage stacks, Cisco was unique in its ability to offer a full stack solution, including VAST Data as the storage partner. VAST offered an advanced multi-tenancy architecture, and inspired confidence with its roadmap for a data pipeline for AI.

Flexible capabilities to support new monetization opportunities

As the service provider was working under an aggressive implementation deadline, Cisco's ownership of the entire stack was a key differentiator. Splitting up the compute, network, and storage stacks across different vendors was likely to slow deployment, and the complete single-vendor solution gave the provider confidence in its ability to put its new AI service offerings into production fast.

In addition to the new GPU-as-a-Service offering, the service provider is developing two additional services. An AI cloud offering will support a variety of AI-based use cases for customers, while an enterprise cloud service will offer public cloud services capable of competing with hyperscalers.

The Cisco team proposed a unified infrastructure approach in which the enterprise and AI cloud services share the same infrastructure stack. The organization is initially targeting three industry segments, including financial services, government, and manufacturing. Cisco Secure AI Factory with NVIDIA provides a validated reference architecture for the stack.

Working closely with Cisco, the service provider has rapidly expanded its data center to build out an AI capacity of 1,000 GPUs in just a few months, with more expansion expected in the future.

“We focused on our customer’s pain points and replicated the enterprise reference architecture that includes compute, networking, security, and storage. That did a lot to build confidence. Our competitors couldn’t offer such an integrated full-stack solution.”

Cisco Account Executive