

Cisco Compute Hyperconverged with Nutanix GPT-in-a-Box

Deploy GPT models quickly and with confidence

For many organizations, AI is the opportunity to achieve enhanced efficiency, cost reduction, improved decision-making, and superior customer experiences. Businesses aim to leverage AI to innovate and maintain a competitive edge by optimizing processes, enabling predictive analytics, ensuring quality control, and personalizing services.

The latest GPT models designed to understand and generate human-like text can enhance user experience, save time and money by automating tasks, and be utilized in a wide range of use cases making GPT models a highly versatile tool.

However, deploying AI-ready infrastructure can be a complex endeavor, and IT teams are being asked to quickly deploy new infrastructure for AI. Hence, organizations need validated, turnkey AI solutions that can easily and securely scale while still maintaining control of their data and applications.

The Cisco Compute Hyperconverged with Nutanix GPT-in-a-Box solution takes the complexity out of adopting generative AI by providing prescriptive steps for deploying the underlying infrastructure for Nutanix GPT-in-a-Box. This solution combines Cisco® servers and SaaS operations with Nutanix software, utilizing the most popular Large Language Models (LLMs) to produce a fully validated AI-ready platform that can simplify and jumpstart your AI initiatives from the data center to the edge.

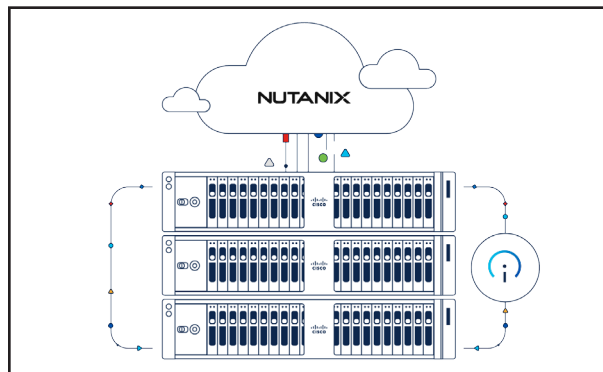


Figure 1. Cisco Compute Hyperconverged with Nutanix

Benefits

- **Streamline deployment** with a Cisco Validated Design that takes the guesswork out of adopting a GPT solution by using a tested and documented approach to help you design and deploy successfully
- **Simplify operations** with a cloud-operating model that delivers visibility, control, and consistency for hyperconverged systems across highly distributed environments
- **Protect your data** with a validated solution that has built-in redundancy and data-protection features to ensure high availability and data integrity

Solution components

Cisco Compute Hyperconverged with Nutanix GPT-in-a-Box solution is a reference architecture that combines:

- Nutanix GPT-in-a-Box software-defined solution
- Cisco Compute Hyperconverged C-Series servers
 - Cisco HCI C240 M7 nodes
- NVIDIA L40S GPU
- Cisco Intersight® standalone mode (no external fabric interconnect)
- Systems management
 - Nutanix Prism
- A range of the most popular large language models

Learn more

For information on Cisco Compute Hyperconverged with Nutanix, visit cisco.com/go/computehyperconverged

