

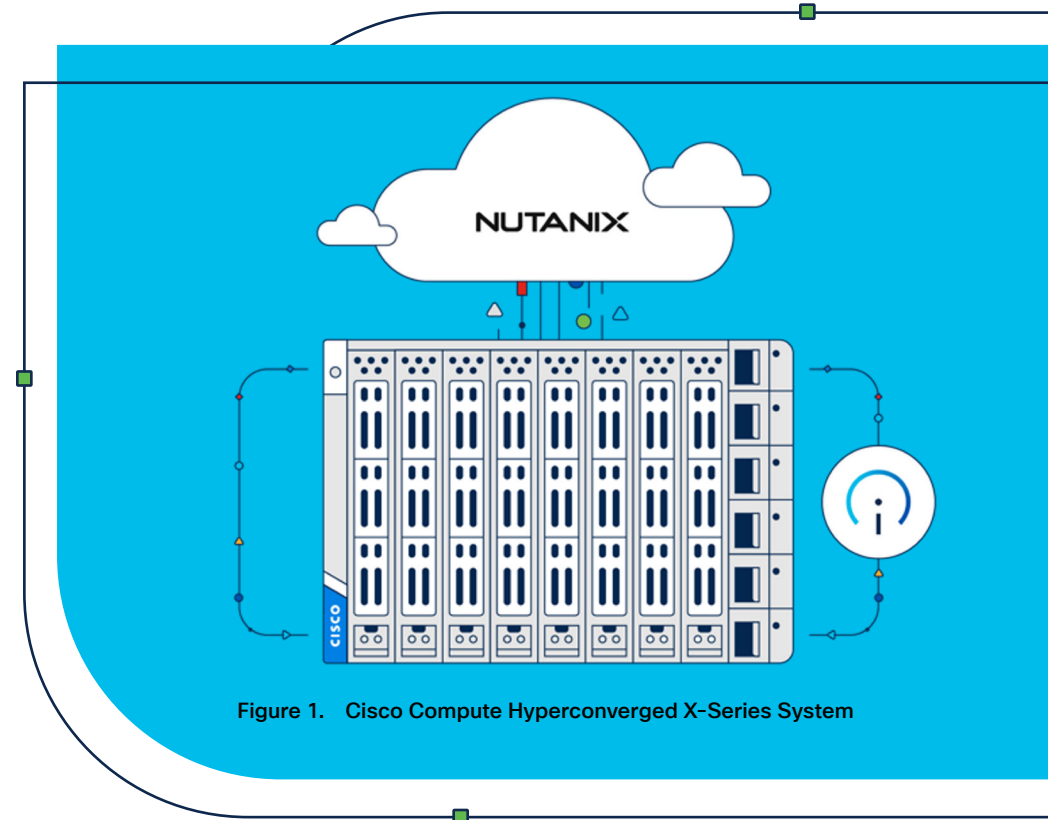
Compute Hyperconverged with Nutanix GPT-in-a-Box 2.0

Simplify the deployment and management of AI models

The Cisco® Compute Hyperconverged with Nutanix GPT-in-a-Box 2.0 solution takes the complexity out of adopting an enterprise AI platform. The solution provides prescriptive steps for deploying the underlying infrastructure for Nutanix GPT-in-a-Box.

Combining Cisco Compute Hyperconverged X-Series System and SaaS operations, this solution enables organizations to easily deploy, scale, and upgrade hyperconverged clusters with a more sustainable, future-ready solution, including Nutanix GPT-in-a-Box 2.0 that features new integrations with [NVIDIA NIM](#) inference microservices and the [Hugging Face](#) Large Language Models (LLMs) library.

The Cisco Compute Hyperconverged with Nutanix GPT-in-a-Box 2.0 is a full-stack reference architecture purpose-built to simplify enterprise AI adoption and run GenAI workloads in the data center and at the edge.



Benefits

- **Accelerate deployment** with a Cisco Validated Design that takes the guesswork out of adopting a GPT 2.0 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.
- **Ensure resiliency** with a validated solution that has built-in redundancy and data-protection features to ensure high availability and data integrity.
- **Increase agility** and respond to the dynamic AI needs of your business with a modular server that is easy to scale and includes support for future generations of processors, storage, accelerators, networking technologies, and SaaS innovations.

Solution components

Cisco Compute Hyperconverged with Nutanix GPT-in-a-Box 2.0 solution is a reference architecture that combines:

- Nutanix GPT-in-a-Box 2.0 with Nutanix Enterprise AI and Nutanix Kubernetes Platform
- Cisco Compute Hyperconverged X-Series System
 - NVIDIA L40S GPU
- Cisco Intersight® cloud-operations platform
- Integrations with [NVIDIA NIM](#) inference microservices and the [Hugging Face](#) Large Language Models (LLMs) library

Learn more

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