



FlexPod Datacenter with SUSE Rancher for AI Workloads Design Guide



FlexPod Datacenter

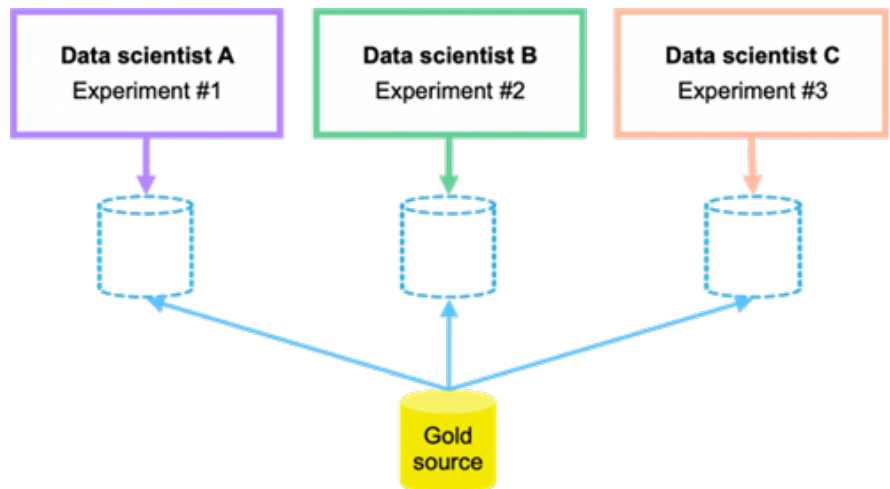
FlexPod® Datacenter is a validated approach for deployment of Cisco and NetApp® technologies to create shared private and public cloud infrastructure. Developed through the partnership of Cisco and NetApp, FlexPod integrates compute, storage, and network design best practices, helping you overcome IT challenges by offering documented design guidance and support throughout the deployment stages. The solution is adaptable, incorporating technology and product innovations in management, computing, storage, and networking.

SUSE Rancher

This Cisco Validated Design (CVD) focuses on FlexPod Datacenter for SUSE Rancher, aiming to help your organization with your digital transformation, application modernization, and artificial intelligence and machine learning (AI/ML) practices. As hybrid cloud operations become prevalent, the solution covers network connections to the public cloud, Kubernetes cluster management, and workload distribution across on-premises and public clouds.

Enterprise Container Management

SUSE Rancher Enterprise Container Management (ECM) is an enterprise-ready Kubernetes container platform, facilitating automated operations across various deployments. When combined with FlexPod, it simplifies your container infrastructure deployment and management. Ansible integration automates FlexPod infrastructure deployment and SUSE Rancher installation, enabling scalable and agile infrastructure programming.



Empower your data scientists while controlling storage costs

The NetApp DataOps Toolkit enables your data scientists to quickly create efficient copies of datasets, regardless of size. These copies, known as clones, can be easily reformatted and manipulated while preserving the original dataset.

This process, powered by the NetApp FlexClone® feature, streamlines dataset cloning operations, so your data scientists can carry out such tasks in seconds. Your data scientists can also save a space-efficient, read-only copy of a dataset by using NetApp Snapshot™ technology. This feature facilitates dataset versioning, and it maintains traceability from dataset to model, which is crucial in regulated industries. The NetApp DataOps Toolkit simplifies and accelerates such tasks that once took days or weeks with the help of a storage administrator.

Future-ready programmable infrastructure

Key advantages of integrating FlexPod Datacenter with SUSE Rancher ECM include programmable infrastructure, use of the latest hardware and software innovations, storage modernization with NetApp all-flash arrays, innovative cloud operations, and future-ready design for technologies like liquid cooling and high-wattage CPUs. The solution also incorporates Cisco Intersight for monitoring, orchestration, and workload optimization, with a modular nature for easy upgrades to additional services.

NVIDIA GPU support

Several NVIDIA GPUs are supported in FlexPod, powered by Cisco UCS. Each GPU is designed for specific applications.

The NVIDIA A100 Tensor Core GPU, based on the Ampere architecture, introduces Multi-Instance GPU (MIG) for optimal resource utilization. The H100 Tensor Core GPU, built on the Hopper architecture, offers acceleration for various math precisions.

The L4 Tensor Core GPU is a passively cooled card that's optimized for inference at scale, and the L40S GPU, based on the Ada Lovelace architecture, is described as the most powerful universal GPU for data centers.

You can complete the solution with NVIDIA AI Enterprise. This end-to-end AI software platform has over 50 frameworks, pretrained models, and development tools that help you accelerate your data science pipeline and simplify your deployment of production AI.

Summary

Choose this innovative FlexPod solution for your organization to deploy an enterprise-ready container platform with:

- Programmable infrastructure and automation
- AI/ML capabilities, including GPU support
- Industry-leading hybrid cloud operations

To learn more, read the [FlexPod Datacenter with SUSE Rancher for AI Workloads Design Guide](#).

To see what else FlexPod can do for you, [check out the rest of our design guides](#).