AI/ML Design for Cisco

Product Code: CNS-NAI-A-AIML-DSGN-SD-INP-CSCO

At-a-Glance

Stage: Design

The Artificial Intelligence/Machine Learning (AI/ML) Design for Cisco engagement provides a focused design service to ensure readiness for deploying AI workloads on Nutanix Cloud Infrastructure (NCI) within Cisco HCI UCS environments.

Designed as a strategic step in your AI/ML journey, this engagement helps organizations plan and validate architecture for inference use cases, ensuring interoperability, security, and scalability for future growth.

This engagement focuses on:

- Capturing solution requirements, constraints, assumptions, and key decisions
- Designing AI workload architecture on NCI, including GPU sizing and configuration
- Validating integration with identity services (AD/LDAP), DNS, and data governance policies
- Planning for security measures such as encryption, SSL certificates, and syslog
- Aligning design with Nutanix recommended practices for performance and operational readiness

This service is especially valuable for organizations deploying AI/ML workloads on Cisco UCS platforms, ensuring optimized design for inference use cases and future scalability.

Related Services

- NCI Cluster Design for Cisco
- AI/ML Planning
- NAI Deployment
- AI/ML Strategy & Optimization Series

Service Scope

The AI/ML Design for Cisco includes a series of collaborative sessions led by certified Nutanix consultants with deep expertise in AI workload design and Cisco UCS environments. These sessions deliver immediate value by:

- Gathering and documenting solution requirements, constraints, dependencies, and decisions
- Designing AI workload architecture on NCI, including GPU selection and configuration for inference
- Validating NCI/NAI sizing based on customer-provided AI use case details
- Developing interoperability across NCI, NKP, and NAI platforms for security and scalability
- Assessing network requirements and designing virtual networking integrated with physical infrastructure
- Defining integration with AD/LDAP and DNS environments
- Planning for security hardening, including encryption, SSL certificates, and password policies

By the end of the engagement, customer teams will have a validated AI/ML design tailored to their operational goals, ready for deployment on Cisco UCS infrastructure.

This service includes the following activities:

- Gather and document AI-specific solution requirements, constraints, assumptions, dependencies, and decisions in a series of design sessions
- Develop AI workload design on NCI platform
  - Develop NCI/NKP/NAI interoperability, security, and scalability for future growth
  - Define integration with Active Directory (AD)/lightweight directory access protocol (LDAP) and domain name service (DNS) environments
  - Review the customer's current data governance
  - Validate existing NCI cluster design
  - Gather GPU requirements based on use cases
  - Design Nutanix AI (NAI) including the number of instances and which LLM should be used based on the use cases
- Validate NCI/NAI sizing based on AI use case details provided by the customer
- Discuss the GPU selection and configuration options for inference
- Assess the network requirements and design virtual networking, including integration with the physical network
- Validate cluster size and platform selection based on workload details provided by the customer
- Design security, including data-at-rest encryption, secure sockets layer (SSL) certificate, password complexity, and syslog

Project Management

Nutanix Project Management (PM) oversees Nutanix resources and aligns execution with your goals, scope, and timelines.

Core project management activities may include the following:

- Serve as a single point of contact for all project communication
- End-to-end Nutanix resource management
- Coordinate change window(s) and implementation schedules with customer
- Track and facilitate readiness and prerequisite completion
- Conduct Project Kickoff/Tech Readiness meeting(s)
- Integrate customer resources into the high-level project timeline
- Send status update(s)
- Manage timeline(s)
- Deliver created artifacts to the customer
- Facilitate project closeout activities

Limitations

- For each quantity purchased, infrastructure design is limited to a single physical site and a single AI/ML inference use case.
- Limited to Cisco HCI UCS environments
- Excludes design for NCI, Nutanix Kubernetes Platform (NKP), Nutanix Database Service (NDB), and Nutanix Unified Storage (NUS)

  Note: On-premises NCI design is included in the *NCI Cluster Design for Cisco* offer

  For AI/ML workloads running on Bare Metal or Public Cloud, a custom statement of work (SOW) is required

Project Management

- Excludes scheduling customer resources and activities
- Excludes detailed project plan (schedule) development and management
- Excludes responsibility for creating, managing, or delivering change management communications

Supported Hypervisors

- Nutanix AHV

Prerequisites

- None

Required Product Licenses

- None

Delivered Artifacts

Delivered Artifact	Description
Configuration Workbook	Captures all required configuration settings and decisions gathered during the design session to support accurate and consistent solution deployment.
Design Document	Captures the customer's solution architecture based on design session outcomes, detailing both high- and low-level designs. It documents requirements, constraints, assumptions, and risks, and provides clear rationale for design decisions to ensure the solution meets performance, availability, scalability, and other critical objectives.

Level of Effort

Typically up to 2 days

Delivery Type

Delivery Type	Delivery Activities
In-person	• In-person design session • Virtual documentation • Virtual project management **Note:** Any in-person project management activities provided solely at Nutanix's discretion.

Related Products

- Nutanix Cloud Infrastructure (NCI)

- Nutanix Enterprise AI (NAI)

- Nutanix Kubernetes Platform (NKP)

- Nutanix Unified Storage (NUS)

- Nutanix Database Service (NDB)

Terms and Conditions

This document contains the entire scope of the service offer. Anything not explicitly included above is out of scope. This service offer is subject to the Nutanix Services General Terms and Conditions, which can be viewed at https://www.nutanix.com/support-services/consulting-services/terms-and-conditions