AI/ML Design

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

At-a-Glance

Stage: Design

The Artificial Intelligence/Machine Learning (AI/ML) Design provides a focused design service to ensure readiness for deploying AI workloads based on a Generative Pre-trained Transformer (GPT). 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

- AI/ML Planning
- NAI Deployment
- AI/ML Strategy & Optimization Series

Service Scope

The AI/ML Design includes a series of collaborative sessions led by certified Nutanix consultants with deep expertise in AI workload design. 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.

This service includes the following activities:

- Gather and document solution requirements, constraints, assumptions, dependencies, and decisions in a series of design sessions
- Develop AI workload design on NCI platform.
  - Gather and document solution requirements, constraints, assumptions, dependencies, and decisions in a series of design sessions
  - Gather AI-specific requirements, constraints, assumptions, dependencies, and decisions
  - 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
  - Look at the client's current Data Governance
  - Develop NCI cluster design
  - Gather GPU requirements based on Use Cases
  - Design virtual networking, including integration with the physical network
  - Design virtual storage, including container layout, compression, and de-duplication
  - Design NAI – Number of instances and which LLM should be used based on the use cases
- Validate NCI/NAI sizing, based 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/technical 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 AI/ML inference use case.
- Management and other cluster designs require a separate Infrastructure Design for each additional cluster
- Excludes Nutanix Kubernetes Platform (NKP), Nutanix Database Service (NDB), and Nutanix Unified Storage (NUS) design
- The design service is limited to a single physical site
- Excludes AI/ML workloads running on NAI Bare Metal or NAI Standalone

  Note: For AI/ML workloads running on NAI Bare Metal or NAI Standalone, a custom statement of work (SOW) is required

Supported Hypervisors

- Nutanix AHV

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

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 5 days

Delivery Type

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