NAI Deployment

Product Code: CNS-INF-A-SVC-DEP-GPT

At-a-Glance

Phase: Deploy

- The Nutanix Enterprise Artificial Intelligence (NAI) Deployment service accelerates the deployment of a comprehensive inference endpoint management product designed to streamline and optimize your AI model orchestration experience. NAI allows you to select, deploy, and manage large language models (LLMs) on a Kubernetes cluster. This offer is ideal for the Deploy stage of the Nutanix Enterprise AI Solution journey.

Related Services

- AI/ML Planning Workshop
- AI/ML Design Workshop
- Infrastructure Deployment
- NKP Deployment
- NUS Deployment

Service Scope

- Highly skilled consultants with strong domain expertise and rich experience deploy NAI on a supported Kubernetes platform. After the deployment of the solution, the consultant demonstrates the LLM with sample data.

This service includes the following activities:

- Review the NCI cluster configuration that runs a supported Kubernetes Platform, including:
  - GPU Support
  - GPU operator installed
  - NKP version
  - NUS files installed with CSI Setup in NKP
  - NKP load balancer set up with FQDN
  - SSL certification (secure)
- Install NAI on the supported Kubernetes cluster, including:
  - Add and update Nutanix helm repository
  - Set up Hugging face
  - Importing LLM
  - Configuring Endpoint
  - Demonstrating the LLM with sample application

Limitations

- For each quantity purchased, deployment is limited to 1 on-premises NCI cluster
- Excludes training a new LLM
- Excludes creation or updates to existing design documentation
- Excludes NCI cluster, NKP, and NUS deployment

Supported Hypervisors

- Nutanix AHV

Supported Kubernetes Platforms

- Nutanix Kubernetes Platform (NKP)

Prerequisites

- Fully supported and functional on-premises NCI cluster that meets all product requirements for NCI, NKP, NUS, and a supported GPU

  Note: For information on the requirements for NCI Clusters, see Field Installation Overview in the *Field Installation Guide* on the Nutanix Support Portal.

  For information on the requirements for deploying NKP see Basic Installations by Infrastructure in the *Nutanix Kubernetes Platform Guide* on the Nutanix Support Portal

  For information on the requirements for NAI, see Nutanix Enterprise AI Requirements in the *Nutanix Enterprise AI Guide*.

  For information on NUS Files Prerequisites, see Prerequisites in *Nutanix Files User's Guide* on the Nutanix Support Portal.

- Completed Pre-Install Questionnaire

Required Product Licenses

- Nutanix Cloud Infrastructure (NCI) Ultimate Edition
- Nutanix Enterprise AI (NAI)
- Nutanix Kubernetes Platform (NKP) Pro or Ultimate Edition
- Nutanix Unified Storage (NUS) Pro Edition

Deliverables

- Project Kickoff
- Project Schedule
- Project Status Report(s)
- Deployment
- Usable endpoint
- Project Closeout

Duration

Typically up to 2 days

Related Products

- Nutanix Cloud Infrastructure (NCI)
- Nutanix AI (NAI)
- Nutanix Kubernetes Platform (NKP)
- Nutanix Unified Storage (NUS)

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 that can be viewed at https://www.nutanix.com/support-services/consulting-services/terms-and-conditions