AI/ML Stack Design Workshop

Product Code: CNS-INF-A-WRK-DES-GPT

At-a-Glance

Phase: Design

The Nutanix Artificial Intelligence/Machine Learning (AI/ML) Stack Design Workshop offers IT teams in-depth and practical guidance to create a comprehensive design for AI/ML workloads based on a Generative Pre-trained Transformer (GPT) model on on-premises Nutanix Cloud Infrastructure (NCI) clusters. It covers various aspects such as performance, scalability, flexibility, integration, and operational needs. This workshop is beneficial during the Design stage of the Nutanix GPT-in-a-box Solution journey.

Service Scope

A series of design workshops are delivered by highly skilled consultants with strong domain expertise and rich experience to ensure that the solution requirements and required outcomes are identified. Design workshops require collaboration with key customer stakeholders from architecture, virtualization, and networking teams. After the design workshop, the consultant develops a Nutanix Design document for AI-ML workloads and a configuration workbook that addresses conceptual, logical, and physical NCI design elements. It also details requirements, constraints, assumptions, design decisions, identified risks, and mitigations

This service includes the following activities:

- Gather and document solution requirements, constraints, assumptions, dependencies, and decisions in a series of workshops
- Develop NCI cluster design for AI workloads
- Discuss the NVIDIA GPU selection and configuration options for training and inference stages
- 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 NUS Files or Objects data service
- Develop a plan for training a sample dataset and validation
- Design security including data-at-rest encryption, Secure Sockets Layer (SSL) certificate, password complexity, and syslog

Limitations

- The design is limited to a single AI/ML use case
- The design includes only one NUS Files or Objects data service

Supported Hypervisors

- Nutanix AHV

Prerequisites

- None

Required Product Licenses

- None

Deliverables

- Project Kickoff
- Project Schedule
- Project Status Report(s)
- Workshop
- Configuration Workbook
- Design Document
- Project Closeout

Duration

Typically up to 5 days

Related Products

- Nutanix Cloud Infrastructure (NCI)
- Nutanix Unified Storage (NUS)
- Nutanix Cloud Manager (NCM)

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

Tel 1 (855) 688-2649 · **Fax** 1 (408) 916-4039 · **Email** info@nutanix.com ·