Key Capabilities Checklist

Successfully Run AI Models and Cloud Native Apps Anywhere

Introduction

Enterprises have launched AI initiatives with an eye on use cases to help generate code and content faster, improve the customer experience, strengthen security, and accelerate the deployment of cloud native apps.

However, many organizations are still wondering where to begin.

To get started, we'll focus on each of these highly sought-after use cases and explain the advanced capabilities you'll need to successfully navigate your AI and cloud native journey.

Accelerate the deployment of cloud native apps

Leading enterprises are adopting cloud-native development methods to accelerate delivery of digital services, remove friction from customer interactions and provide employees with the latest data-driven tools. Key capabilities include:

- ☑ Ability to run the same cloud native stack across datacenter, cloud and edge
- ☑ Flexibility to choose the right components for your specific app needs
- ☑ App-aware data services that work efficiently across your cloud native environment
- ☑ Integration to simplify and accelerate the deployment and management of a cloud native app stack that is tailored to your needs
- ☑ Partners that understand the cloud native ecosystem and can support the cloud native stack

Nutanix enables you to deliver and manage the same end-to-end production-ready cloud native environment across the datacenter, cloud and edge. Our secure, resilient and self-healing platform simplifies planning, deploying and operating cloud native apps anywhere.

Strengthen security

Enterprises are starting to rely on AI to fortify security, which has sent security use cases to the top of many priority lists. Key capabilities include:

- ☑ Internally developed security software
- ☑ Infrastructure to support inference with:
    - ☑ Appropriate accelerator hardware
    - ☑ Flexibility to deploy in all the locations across datacenter, cloud and edge
- ☑ Infrastructure to support AI model training and fine-tuning
- ☑ The infrastructure and skills to manage AI models and datasets throughout the AI lifecycle
- ☑ Support for containerized apps at scale

Commercial security software that runs on-premises may have specific infrastructure requirements you will need to satisfy while SaaS may have specific bandwidth requirements.

Nutanix provides infrastructure to accelerate your AI security initiatives across hybrid multicloud environments. Nutanix Security Central uses machine learning algorithms to analyze network traffic and recommend the right policies for specific environments.

Enrich the customer experience

Customer experience use cases personalize the customer experience, increase customer satisfaction, boost operating efficiency, and reduce customer service costs. Key capabilities include:

- ☑ Infrastructure to support inference with:
    - ☑ Appropriate accelerator hardware
    - ☑ Flexibility to deploy in all the locations across datacenter, cloud and edge
- ☑ Infrastructure to support AI model training and fine-tuning
- ☑ Ability to support retrieval augmented generation (RAG)
- ☑ The infrastructure and skills to manage AI models and datasets throughout the AI lifecycle
- ☑ Support for containerized apps at scale

Nutanix enables you to deliver AI tools to improve the customer experience, such as faster time-to-problem-resolution and responses tailored to customer context. And you'll maintain full control over data protection and privacy.

Nutanix GPT-in-a-Box

Customer experience use-cases typically rely on the latest GenAI models. Nutanix GPT-in-a-Box seamlessly integrates GenAI and other AI models into your operations while keeping data and applications under your control.

Nutanix Enterprise AI

Nutanix Enterprise AI (NAI), a key part of GPT-in-a-Box, offers endpoint APIs, including NVIDIA NIM and Hugging Face, making it easy to deploy a wide range of AI models on-premises and in the cloud. You can deploy NAI on Nutanix infrastructure or the infrastructure of your choice.

Complete data services

Performance and data management across on-premises, cloud and the edge can dramatically different. Nutanix overcomes this challenge by simplifying storage operations and providing the speed and flexibility to build AI apps and services and deploy them anywhere.

Faster code and content creation

Code and content use cases such as co-pilots rely on private deployment of generative AI (GenAI) models. Key capabilities include:

- ☑ Infrastructure to support inference with:
    - ☑ Appropriate accelerator hardware
    - ☑ Flexibility to deploy in all the locations across datacenter, cloud and edge
    - ☑ Ability to support containerized apps at scale

Nutanix accelerates code and content use cases with Nutanix GPT-in-a-Box, a turnkey software-defined solution that seamlessly integrates GenAI models into your operations.

Additional guidance

For more guidance, download the complete eBook, The Essential Guide to Navigating AI and Cloud Native Deployments.

This guide explains what you need to build and deploy AI models that speed-up code and content creation, enrich the customer experience, and strengthen security as well as accelerate the deployment of cloud native apps.

NUTANIX