May 2025 | eBook

Transforming Healthcare IT

The Essential Guide to Navigating AI and Cloud-Native Deployments

Run AI and cloud-native apps anywhere across hybrid multicloud environments to improve patient care

NUTANIX

Table of Contents

How Nutanix Solves AI and Cloud Native Challenges
GPT-in-a-Box 4
Nutanix Enterprise AI 4

Faster Code and Content Creation
Considerations 5
Challenges 5
How Nutanix helps 6
Use case: Government agency tackles document analysis 6

Enrich the Customer Experience
Considerations 7
Challenges 7
How Nutanix helpss 8
Use case: Satellite firm uses GPT-in-a-Box for chatbots and AI assistantss 8

Strengthen Security with AI
Considerations 9
Challenges 9
How Nutanix helps 10

Accelerate the Deployment of Cloud Native Apps
Considerations 11
Challenges 11
How Nutanix helps 11
Use-case: One-stop cloud native with Nutanixs 11

Key Capabilities Checklist
Faster code and content creation 12
Enrich the customer experience 12
Strengthen security with AI 12
Accelerate the deployment of cloud native apps 12

Get Started

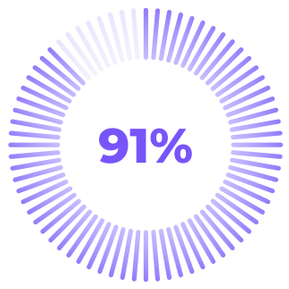
Introduction

Healthcare organizations everywhere are launching AI initiatives to improve clinical decision-making, streamline operations, enhance patient care experiences, strengthen cybersecurity, and accelerate the deployment of cloud-native applications.

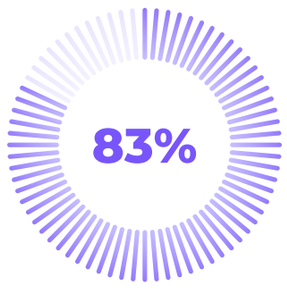
However, many healthcare providers and IT leaders are still asking the same fundamental question: Where do we start?

AI offers a wide range of healthcare apps – from clinical documentation and diagnostic support to operational optimization and personalized patient outreach. But with so many possibilities, it can be difficult to identify the ones that will have the greatest positive impact on your organization.

Consider AI a priority



Agree their IT infrastructure needs to be improved to support AI



Plan to increase investment in egde strategy to support AI

Choosing the right infrastructure to support AI in production environments is no small feat. Should you deploy AI workloads in your on-premises datacenter, in the cloud or at the edge – closer to patient care delivery? What acceleration hardware is required to support real-time inference? How will you manage sensitive patient data, compliance requirements and the AI lifecycle?

AI-powered healthcare apps running in cloud-native environments are typically containerized. This represents a shift in deployment and data models for IT teams that are accustomed to managing VMs. The ecosystem evolves rapidly, skill gaps are real, and configuring these environments can be a significant challenge.

This eBook will help you choose use cases and successfully deploy AI and other cloud-native apps that move the needle for your healthcare organization. It explores use cases in four domains:

- **Code and content creation.** Adopt GenAI to increase productivity.
- **Patient care experience.** Deliver a better, more personalized experience for patients.
- **Security.** Protect the security and privacy of your business, employees and patients.
- **Cloud native apps.** Develop and run AI and other apps using the latest cloud-native tools.

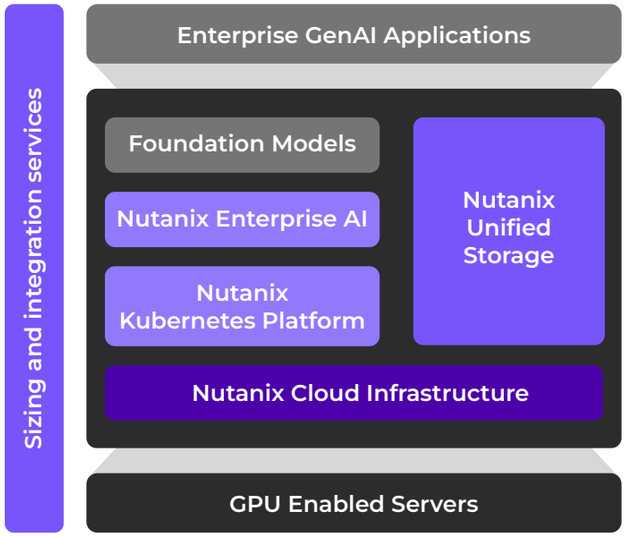
How Nutanix Solves AI and Cloud Native Challenges

GPT-in-a-Box

GPT-in-a-Box from Nutanix is a turnkey software-defined solution designed to seamlessly integrate AI models into your operations while keeping data and apps under your control.

Advantages include:

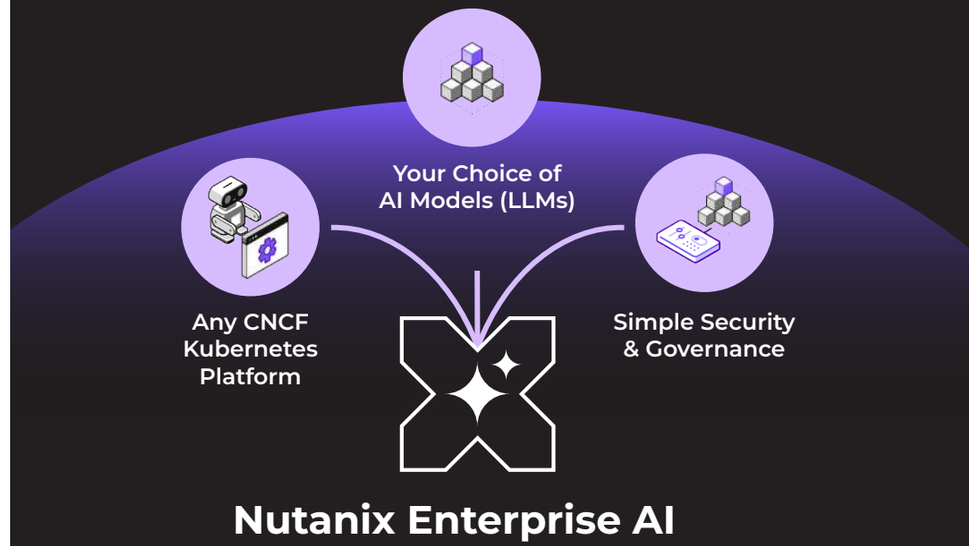
- **Full AI stack.** Deploy a curated set of large language models (LLMs) using leading open-source AI frameworks.
- **Delivered anywhere.** From datacenter to edge to public cloud.
- **Built-in data services.** Fine-tune and run LLMs on optimized storage while maintaining control of data and apps.
- **Inference at the edge.** Deploy everything you need to run AI models for inference at edge locations. A compact footprint overcomes site limitations and helps control costs.
- **Lower TCO.** Reduce TCO with automation, dynamic resource allocation and consolidation to optimize infrastructure costs.
- **Deploy and adapt with ready-to-use AI models and LLMs.** Validated AI models that can be deployed quickly from the edge to cloud and speed-up time-to-value.
- **Deploy and operate secure APIs for AI apps.** Easily create secure role-based APIs to connect apps to AI models.



Nutanix Enerprise AI

A key part of GPT-in-a-Box, Nutanix Enterprise AI (NAI) offers end point APIs, including NVIDIA NIM and Hugging Face, so it's easy to deploy a wide range of GenAI models on-premises and in the cloud. NAI includes a simple UI, role-based access controls and untethered operations.

While NAI runs great on Nutanix infrastructure solutions, Nutanix infrastructure is not required. You can deploy NAI on the infrastructure of your choice.



Faster Code and Content Creation

To get started with AI, identify use cases that will streamline business operations and improve the efficiency of internal teams. Common examples include:

- **Co-pilot for content.** Deploy AI tools to help generate better content in less time.
- **Co-pilot for code.** Increase software team productivity with the latest AI tools.
- **Intelligent document processing.** Leverage natural language processing to extract, classify and interpret information from documents and speed-up business processes.

Many organizations – especially those in regulated industries and the public sector – can't or prefer not to use shared, cloud-based services from vendors such as Microsoft, OpenAI or Google for reasons of cost, security and compliance. But by supporting code and content creation use cases from private infrastructure that you control, you can deliver the capabilities employees need to be successful, while avoiding the risks created when well-meaning users upload sensitive or proprietary company data to public AI services.

Considerations

Use cases for code and content creation generally rely on pretrained GenAI models. To get started, you'll need to deploy the necessary Full AI stack. Deploy a curated set of large language models (LLMs) using leading open-source AI frameworks.

- Inference for the latest foundation models from organizations like Meta, Microsoft, Mistral, and Google. Nutanix has partnered with Hugging Face, a popular resource for open-source models.

You may also need to support:

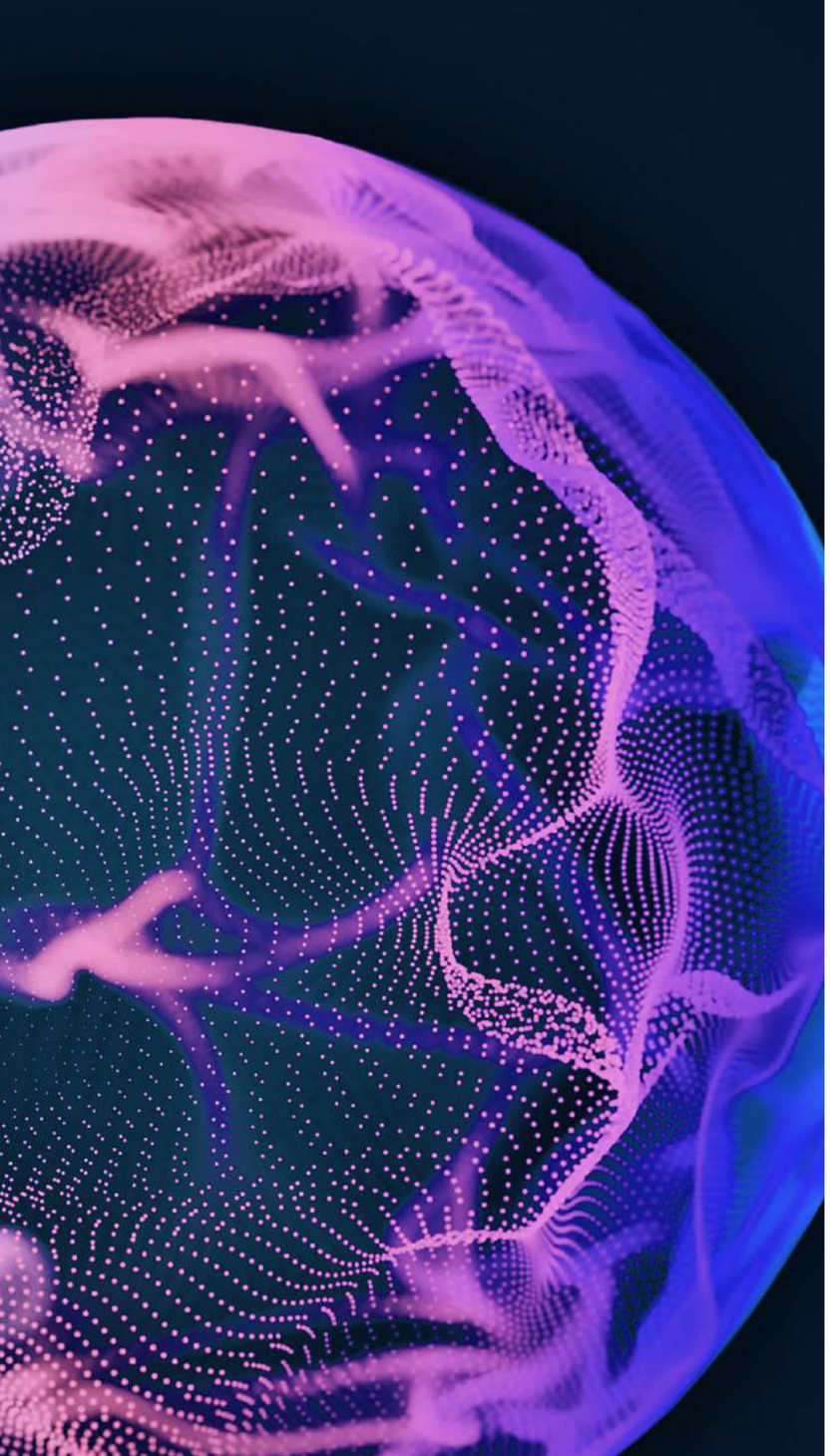
- Fine-tuning, a form of deep learning, of your chosen models using company data.
- Retrieval augmented generation (RAG) so your model has access to selected internal data organized in a database.

This Nutanix Validated Design can help you get started with fine-tuning or RAG.

Challenges

The challenges associated with code and content creation use cases include:

- **Complexity** You may need to deploy AI everywhere you operate, including datacenters, public clouds and at the edge. Technology differences between thes e disparate environments increase operational complexity.
- **Specialized hardware.** You'll need it to accelerate inference – GPUs from NVIDIA or CPUs optimized for inference – and support AI training (GPUs). Because this hardware is expensive and difficult to come by, it's important to keep utilization high.
- **AI lifecycle management.** Because AI apps and the models they are built on constantly evolve, containers and Kubernetes are preferred to run and scale AI models. This provides faster deployments and updates, high availability and efficient resource utilization. However, there may be a skills gap if containers and Kubernetes are new to your organization and it might be difficult to find people with the cloud native skills.



How Nutanix helps

Nutanix provides the AI tools your employees need in less time with less effort using a secure, scalable and cost-effective platform. With Nutanix, you can accelerate your AI initiatives across hybrid multicloud environments while maintaining control over your data.

Nutanix advantages

Nutanix addresses several key challenges of code and content creation use cases.

- **Take advantage of hybrid multicloud.** Run the latest AI models anywhere without significant operational changes in each location.
- **Efficiently utilize specialized hardware.** Get more from costly GPUs and other specialized hardware through efficient resource sharing.
- **Leverage containers and Kubernetes**. Using GPT-in-a-Box to simplify Kubernetes, Nutanix makes it easy to get started with cloud native by enabling you to use your preferred tools. At the same time, NAI lets you effortlessly deploy AI models on the infrastructure and Kubernetes of your choice.

Use case: Government agency tackles document analysis

When a government agency needed to decrease the time spent finding legislative data in a repository with millions of documents in 24 languages, it turned to Nutanix. GPT in-a-box made the agency's AI solution easy to deploy and eliminated the need to build a solution from scratch

In addition to GPT-in-a-Box with NVIDIA GPUs, key solution elements included:

- DocsGPT, an AI-powered chatbot that improves the user experience for documentation use cases.
- Red Hat OpenShift, the full-stack Red Hat Kubernetes solution fully validated for the Nutanix platform.
- A custom LLM from Mistral.

Enrich the Customer Experience

GenAI can improve and personalize the customer experience, elevate customer satisfaction and stoke innovation – all while boosting operating efficiency and reducing customer service costs.

Popular customer experience use-cases include:

- **Customer support chatbots.** Fine-tune LLM with company data to create chatbots that can handle customer inquiries, prompt customer responses and assist with complex issues.
- **Sentiment analysis.** Analyze customer feedback, identify trends and proactively address issues before they escalate.
- **Personalization.** Analyze vast amounts of data about customer preferences, behaviors and interactions to create personalized recommendations, emails, social media posts, and other tailored content.

Considerations

Customer experience use cases like the examples above typically utilize GenAI foundation models combined with either fine-tuning or RAG. You should plan to provide infrastructure support for:

- **Inference** for your chosen foundation models.
- **Training and fine-tuning** of your models using company data.
- **RAG** so that your chosen model has access to selected internal data organized in a database.

Because of the amount of data involved, customer experience use cases will increase data storage needs and data management complexity.

Challenges

As you plan for your customer experience use-case, you may encounter:

- **Hybrid multicloud**. You may need to deploy AI everywhere you operate. Technology differences between disparate environments can increase operational complexity.
- **Specialized hardware.** You'll need specialized hardware to accelerate inference – GPUs from NVIDIA or CPUs optimized for inference – and support AI training (GPUs).
- **Containers.** Containers and Kubernetes are generally preferred to run and scale AI models and ensure high availability and efficient resource utilization.

You should also expect significantly greater data management challenges. In the 2023 Nutanix State of Enterprise AI Report, 100% of organizations required additional skills, including those necessary to manage AI models and datasets during the AI lifecycle.

How Nutanix helps

Nutanix delivers AI tools to improve the customer experience, such as faster time-to-problem-resolution and responses tailored to customer context. With Nutanix, you can accelerate AI initiatives across hybrid multicloud environments while simplifying data management and maintaining 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

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.

Use case: Satellite firm uses GPT-in-a-Box for chatbots and AI assistants

A growing satellite operator offers multi-mission satellite services in over 160 countries. With an AI strategy focused on productivity, efficiency, security, and better decision-making, the company leveraged the Nutanix AI solution powered by GPT-in-a-Box to create chatbots and AI assistants. One of the first solutions enabled chat-to-documents, resulting in an AI assistant with deep knowledge of the company.

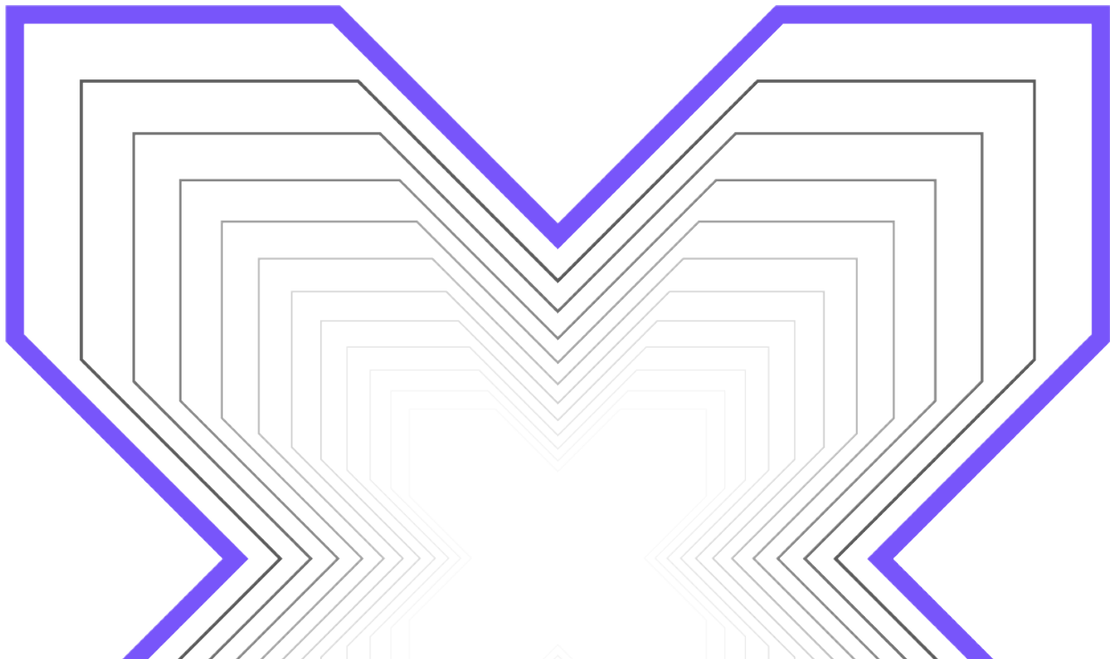
> Nutanix has been instrumental in facilitating infrastructure provisioning, software deployment, and ongoing support, ensuring a seamless execution of our ongoing AI initiatives.”
>
> **Chief operations officer**

Strengthen Security with AI

With cybercrime continuing to rise, security use cases are a priority and many organizations are turning to AI to detect and respond to bad actors.

Examples of AI-enabled security use cases include:

- **Fraud detection.** Analyze historical transaction data to create AI models that can identify anomalies in transaction patterns and accurately detect attempts at fraud.
- **Threat detection.** Leverage AI to analyze vast amounts of data and identify patterns in areas like anomaly detection, threat intelligence and vulnerability assessment.
- **Alert enrichment.** Use AI to automatically gather and analyze context for security alerts so you can set priorities, reduce false positives and speed-up incident response.
- **Automatic policy creation.** Dynamically generate and update security policies based on ongoing analysis of the environment, threat landscape and industry best practices.



Considerations

AI capabilities that satisfy these use cases are now being included in the latest security and data protection tools.

Use cases like fraud detection may require you to acquire or create your own AI model. It will require your team to apply greater effort through significant experimentation, regular training and fine-tuning.

Currently, GenAI only plays an adjunct role in most security use cases. For example, GenAI is often used to create synthetic examples of fraudulent transactions to increase the signal in training datasets used for fraud detection models.

For fraud detection or other internally developed security software, you'll need an infrastructure that supports:

- **Inference** for AI models.
- **Training and fine-tuning** of models using company data.

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

Because of the amount of data involved, security use cases can increase data storage needs and data management complexity.

Challenges

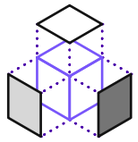
As you plan for security use cases, expect to encounter these challenges:

- **Hybrid multicloud.** You will need to deploy AI-enabled security software everywhere you operate.
- **Specialized hardware.** This is required to accelerate inference – GPUs from NVIDIA or CPUs optimized for inference – and support AI training (GPUs).
- **Containers.** Containers and Kubernetes are generally preferred to run and scale AI models and ensure high availability and efficient resource utilization.

You should also expect significantly greater data management challenges. Many organizations lack the skills necessary to manage AI models and datasets throughout the AI lifecycle.

How Nutanix helps

Nutanix helps you deliver AI security tools that proactively safeguard your business. With Nutanix, you can accelerate AI security initiatives across hybrid multicloud environments while simplify in data management.



Complete data services

Security use cases are data intensive. Your ability to move and manage data efficiently between the datacenter, cloud and the edge can be critical when it comes to real-time threat detection.

Nutanix eliminates data management challenges with software-defined data services that simplify storage operations across the datacenter, cloud and edge.

Nutanix combines the data services you need with the enterprise-class capabilities you expect. We eliminate the need for separate file, block and object storage systems while reducing cost and complexity.

A hardened platform

With Nutanix, security begins with a robust software foundation built for hybrid multicloud architectures. Nutanix starts with a hardened software platform for hyperconverged infrastructure and builds on that foundation.

Our capabilities strengthen your security posture by enabling you to detect and respond quickly to threats that can lead to data loss and business disruption. You can count on Nutanix to provide a secure foundation for AI and other critical workloads.

Accelerate the Deployment of Cloud Native Apps

While enterprise IT teams have deep expertise in virtualization, they may have little or no experience with the nuances of Kubernetes, containers and microservices. You'll need to modernize your methods and skills if you're going to succeed in deploying containerized applications at scale while avoiding Day 2 operational barriers.

Considerations

Cloud native techniques enable loosely coupled systems that are resilient, manageable and observable. Combined with robust automation, they allow engineers to make high-impact changes frequently and predictably with minimal effort. barriers.

Challenges

Configuring and managing a cloud native environment – including storage, networking, monitoring, and Kubernetes lifecycle management – are formidable challenges. Anyone deploying cloud native for the first time is likely to face challenges in four key areas:

- **Deployment and management.** Designing, deploying and operating a full-stack solution requires a lot of time and expertise, especially across datacenters, clouds and edge.
- **Expertise.** Expertise is in short supply. Many solutions in the cloud native arena are open source, making operational support difficult to find.
- **Ecosystem.** Cloud native ecosystems are growing fast and evolving, making it critical to identify technology partners that can help you separate signal from noise.
- **Data.** Container instances are transient and require flexibility to address persistent storage needs for file, object and block storage and different types of databases.

How Nutanix helps

Nutanix makes it incredibly easy to deploy a consistent end-to-end production-ready cloud native environment anywhere – in the datacenter, across multiple clouds and at edge locations.

Runs everywhere

Whether you run cloud native workloads on-premises, in a public cloud, at a service provider datacenter, or at the edge, Nutanix has you covered with consistent management and security across all environments.

Nutanix Kubernetes Platform

Nutanix Kubernetes Platform (NKP) simplifies platform engineering by reducing operational complexity without locking you into a single Kubernetes option. Strategic partnerships with Red Hat, Microsoft, Google Cloud, and AWS enable you to use your preferred Kubernetes platform.

The benefits of NKP include:

Comprehensive data services

- Nutanix Data Services for Kubernetes (NDK)
- Resilience
- Tightly integrated
- Easy upgrades

Use-case: One-stop cloud native with Nutanix

A pharmaceutical company wanted to integrate and connect everything it needed for its cloud native environment. The IT team was concerned about its ability to manage the disparate environment because it was difficult to get support and guidance when problems arose.

With Nutanix, the IT team created a feature-rich, scalable and manageable cloud native environment. Today, the team saves substantial time and effort with a resilient and cost-effective virtualization platform for the Nutanix-validated Red Hat OpenShift Plus solution.

With Nutanix, the company gained a trusted partner that shares best practices and roadmaps and provides help when and where it is needed.

Key Capabilities Checklist

Companies that are just getting started with AI and cloud native apps should consider choosing from straightforward code and content, customer experience, security, and cloud native use cases.

Faster code and content creation

Code and content use cases such as co-pilots rely on private deployment of 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 GPT-in-a-Box, a turnkey software-defined solution that seamlessly integrates GenAI models into your operations.

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 RAG
- ☑ The infrastructure and skills to manage AI models and datasets throughout the AI lifecycle
- ☑ Support for containerized apps at scale

Nutanix provides infrastructure to accelerate your AI initiatives across hybrid multicloud environments while simplifying data management and allowing you to maintain control over data protection and privacy.

Strengthen security with AI

Enterprises increasingly rely on AI to ensure security, and security use cases are on the priority list for many. 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.

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.

Get Started

Nutanix provides everything you need to tackle the use cases that are most critical to your organization. By eliminating the need to integrate disparate hardware and software, Nutanix delivers a scalable, flexible and easy-to-manage AI platform that can save you months of effort.

Don’t let anything stand between you and greater AI and cloud native success. Take an AI and cloud native test drive to see the Nutanix difference for yourself.

Take an AI Test Drive

Take a Cloud Native Test Drive

Visit nutanix.com/ai to learn more. You can also contact us at info@nutanix.com or send us a request at www.nutanix.com/demo to set up your own customized briefing.

NUTANIX

info@nutanix.com | www.nutanix.com | @nutanix

©2025 Nutanix, Inc. All rights reserved. Nutanix, the Nutanix logo and all product and service names mentioned herein are registered trademarks or trademarks of Nutanix, Inc. in the United States and other countries. All other brand names mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s).
EPC-DefinitiveGuidetoBCDR-eBook-FY25Q2-010825