

July 2024 | eBook

How to Overcome the Challenges of Extending AI Applications Anywhere

There's a Simple, Cost-Effective Way to Achieve Success Across Hybrid Multicloud Environments.

NUTANIX



Introduction

Enterprises across all industries are accelerating their AI and cloud native application efforts to deliver the digital services needed to improve customer experience and keep pace with the competition. If a competitor leverages the latest generative AI (GenAI) capabilities in ways that your organization can't or won't, it puts your company at a disadvantage.

However, a recent survey of IT professionals makes clear that many teams aren't sure exactly how to approach AI—or the necessary cloud native capabilities that go with it.

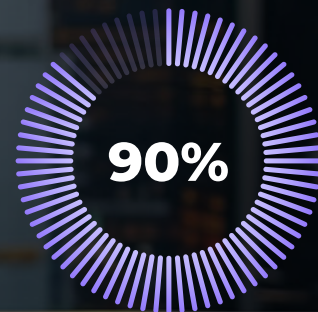


While **90%** of respondents say AI is a priority to their organization, there is still a significant amount of work to be done.

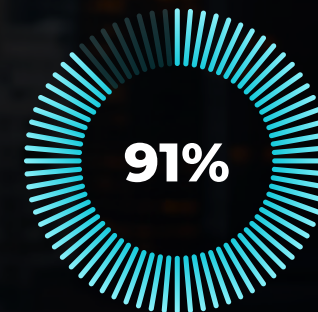
The nascent nature of this technology segment means there is a dearth of strategic best practices, established guardrails, or even reference architectures.

Many organizations are still determining which IT environments are best to run different parts of their AI processes and workloads.

[State of the Enterprise AI Report](#)



90% of respondents emphasize AI security and reliability, and most plan to enhance data protection.



Today's IT infrastructure wasn't built for AI. **91%** of respondents agree it's time to modernize.

Many organizations, especially those in regulated industries and the public sector, hesitate to build applications that leverage shared, cloud-based solutions from vendors such as Microsoft, OpenAI or Google due to:

Security

Protecting proprietary and confidential corporate data is paramount.

Predictability

Many enterprise applications require predictable latency and performance that cloud AI services may not deliver – not to mention unpredictable costs.

Compliance

Regulatory hurdles, data sovereignty requirements and data protection SLAs are a significant and growing concern.

However, building your own platform to satisfy cloud native needs such as AI has significant drawbacks due to challenges in three critical areas: *Infrastructure complexity, security and compliance, and lack of expertise.*

Infrastructure Complexity

Some infrastructure teams know very little about AI and data science, while data scientists know very little about infrastructure. This creates a huge challenge when choosing the right infrastructure to accelerate your AI efforts, support AI applications in production, and prepare your company for the future.

Lack of Expertise

In the [State of the Enterprise AI Report](#), **100%** of organizations said they require additional skills to support AI-related initiatives over the next 12 months. This includes the expertise to choose the right infrastructure to support training and inference, as well as the skills necessary to manage AI models and datasets throughout the AI lifecycle. Data management is cited as a frequent pain point.

For many, cloud native building blocks that underpin much of AI are also a barrier. Your team may have deep expertise in virtualization, but little or no experience with the nuances of Kubernetes, containers and microservices.

Security and Compliance

It can be a struggle to deploy and operate containerized GenAI or other AI-based applications without compromising compliance with IT policies for security, data protection, resilience, and other Day 2 operational needs.

Consistent, policy-based data protection and security across datacenters, edge locations and the cloud can be daunting. A simple, automated and secure approach is essential to efficiently move and manage data.

A Partner You Can Trust

Nutanix is the solution. The Nutanix Cloud Platform (NCP) enables you to deploy AI applications easily, cost-effectively and at scale. Nutanix delivers the application-aware data services, resource efficiency, and enterprise resiliency to support your AI and cloud native initiatives everywhere, while simplifying operations and delivering dramatically improved results.

This eBook explores the challenges enterprises face as they seek to deploy cloud native applications like AI and details how NCP simplifies your path to production and yields exceptional results in less time.

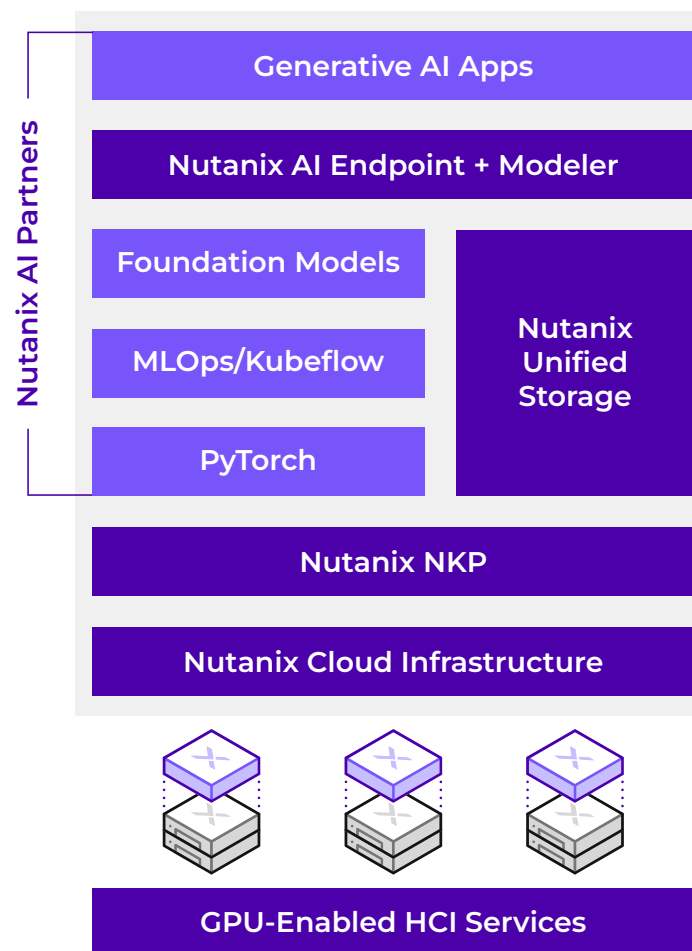
Nutanix Solves Your AI and Cloud Native Challenges

Nutanix offers a secure, scalable and cost-effective platform with built-in data services for your AI and cloud native applications. With Nutanix, you can accelerate your AI and cloud native initiatives across hybrid multicloud environments while maintaining control over data protection, recovery and privacy.

Nutanix has a laser-like focus on reducing infrastructure complexity and enabling hybrid multicloud operations. Built on our proven hyperconverged infrastructure (HCI), NCP provides agile, resilient and secure infrastructure that often exceeds your AI needs everywhere – from the edge to the core datacenter to the public cloud.

A common operating model and unified data management enable you to deliver consistent operations and accelerate AI and cloud native initiatives, by addressing the unique requirements in each environment.

In addition to proven infrastructure advantages, Nutanix offers unparalleled support and services to ensure you get the most from your investment. We've also forged meaningful partnerships with AI industry leaders, giving you certified support for the AI tools you need to succeed.



Simplified Operations

Operational complexity is another big challenge on the path to production deployment of AI. Many organizations perform a lot of AI experimentation in the cloud. But how do you take work that was done in the cloud and move it into the datacenter or operationalize it at the edge?

Turnkey AI

Nutanix offers a turnkey software-defined solution that seamlessly integrates GenAI and AI/ML applications into your organization while keeping data and applications under your control.

Advantages of Nutanix GPT-in-a-Box include:

Full AI Stack

Deploy a curated set of LLMs using leading open source AI frameworks.

Delivered Anywhere

From small-scale edge to large-scale private cloud.

Built-In Data Services

Fine-tune and run LLMs while maintaining control over data and applications.

Inference at the Edge

Deploy everything you need to run AI models for inference in edge locations. A compact footprint accommodates site limitations and helps control costs.

Lower TCO

Reduce TCO with automation, dynamic resource allocation and consolidation to optimize infrastructure costs.

With Nutanix, you can buy, build and modify AI models and data from any source and run them anywhere.

Simplified Operations (cont'd)

The Same Platform Everywhere

Nutanix removes infrastructure complications and limitations, giving you a headstart on the path to AI and cloud native. You can achieve seamless application, workload and data mobility across edge, datacenter and cloud environments.

A single control plane simplifies and streamlines common AI and cloud native workflows, while reducing the risks and the friction that result from operating across multiple, disparate environments.

Because NCP runs everywhere, your teams work efficiently across all environments.

Self-Service and Automation

Advanced self-service and automation capabilities ensure that the needs of developers, data scientists and application owners are met, while accelerating time to value without giving up IT control.

VMs and Containers

While many AI tools and applications run in containers, VMs may also be required. Enterprises require the ability to run VMs and containers well into the future to address the needs of AI, modern apps, and legacy applications.

Nutanix supports your VM- and container-based AI implementations on the same unified infrastructure. Choose your preferred Kubernetes® distribution and the latest tooling to support development and production needs for cutting-edge apps.

Advanced Security

Your AI strategy should enhance – not compromise – security. Nutanix takes a security-first approach to development that ensures all Nutanix software is security-hardened out of the box. Periodic scans self-heal Nutanix systems back to a secure baseline. The tight integration between compute, networking and storage in the Nutanix environment reduces the chance of exposed security gaps.

Nutanix offers a single dashboard to plan, manage and govern the security of AI and other workloads on any cloud – without the need for specialized cybersecurity training or additional personnel. Intelligent analysis automates incident response and helps ensure regulatory compliance.



Comprehensive Data Services

Your AI strategy should consider the differences in services between datacenters, clouds and edge locations—as well as the unique requirements of cloud native—while delivering the performance you expect and addressing regulatory and compliance requirements.

For example, while you may have block or file services available in your datacenter, in the cloud and at the edge, the details of operating those services, the data management capabilities available and the performance you can expect can be dramatically different from one location to the next.

Everything On A Single Platform

Nutanix eliminates this problem with software-defined data services that simplify storage operations while delivering the speed and flexibility needed to build modern AI applications and services and deploy them everywhere.

Any Nutanix cluster is able to support the data services you need with the enterprise-class capabilities you expect. You can also dedicate clusters to address large-scale data storage requirements. Either way, you eliminate the need for separate file, block and object storage systems while reducing cost and complexity.

With Nutanix, You Can:

- Deliver the same level of storage performance, reliability and control everywhere.
- Scale AI data quickly without disruption.
- Streamline lifecycle management of AI models and data.
- Protect AI applications from data leakage and data poisoning.
- Backup and restore centrally and locally.

Unified File, Block and Object

With unified file, block, and object storage options, Nutanix optimizes storage performance, simplifies MLOps and sharpens data security. Because of the importance of unstructured data in AI use cases, the scale-out, high-performance characteristics of Nutanix Files Storage and Nutanix Objects Storage are especially advantageous.

The Nutanix Data Lens™ data security solution enables you to assess and mitigate unstructured data security risks by identifying anomalous activity, auditing user behavior and adhering to monitoring requirements at the primary storage layer. Global data visibility across environments assists with intelligent decision-making and lifecycle management.

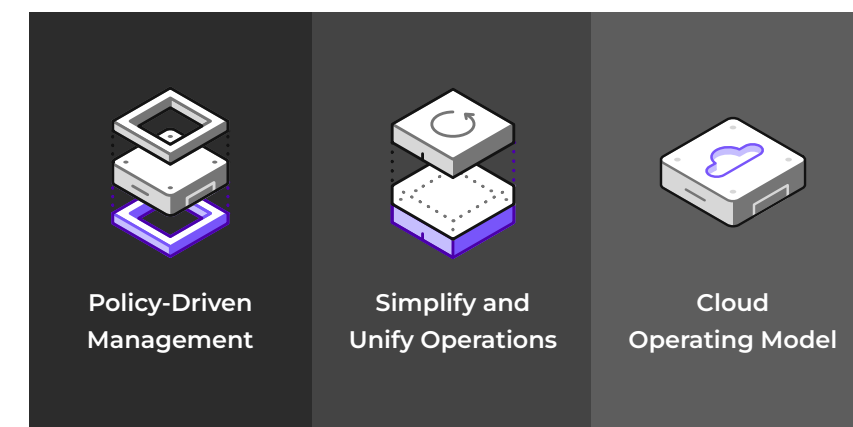
Database-As-A-Service

For structured data storage needs, Nutanix offers database-as-a-service that simplifies the deployment and management of popular databases like Microsoft SQL Server, Oracle, PostgreSQL, MongoDB, and MySQL.

Application-Aware Data Services

Nutanix Data Services for Kubernetes (NDK) extends Nutanix infrastructure capabilities to provide application-centric data services for containerized workloads. This enables developers to build and run containerized AI and other workloads without constantly scrambling for storage resources. With NDK, your platform engineering and development teams can: requirements at the primary storage layer. Global data visibility across environments assists with intelligent decision-making and lifecycle management.

- Manage, control and operate VMs and containerized applications as a single entity from a single platform.
- Tag AI and cloud native applications with specific policies and apply them globally.
- Leverage built-in data protection using Nutanix snapshots and disaster recovery processes.
- View all components of every application everywhere it is running.





Seamless Cloud Native Support

Teams that are accustomed to working in virtualized environments often find configuring and managing a cloud native environment – including storage, networking, monitoring, and Kubernetes lifecycle management—a formidable challenge. The deployment model is different, the data model is different, the entire ecosystem is evolving rapidly, and expertise is in short supply.

A Better Platform for Cloud Native

Nutanix offers the best platform for deploying and managing all your cloud native applications—including AI—efficiently and economically.

Unlike other solutions, Nutanix enables you to deliver and manage the same end-to-end production-ready cloud native environment everywhere—in the datacenter, cloud and edge. Our secure, resilient and self-healing platform simplifies planning, deploying and operating cloud native applications, like AI.

Simplified operations and comprehensive data services from Nutanix deliver profound benefits for cloud native environments. These benefits include:

Runs Everywhere

Whether you need to 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.

Greater Flexibility and Choice

Some solutions lock you into a single Kubernetes option, but your team might have experience with something else. Through strategic partnerships with Red Hat, Microsoft, Google Cloud, and AWS, Nutanix enables you to run the Kubernetes platform of your choice and seamlessly extend your investment and user experience to your on-premises Kubernetes environment.

Scalable

Nutanix HCI provides a highly available platform that can scale out or up to support training and inference for the most demanding AI models, modern microservices applications, and high-performance databases.

Resilient

Nutanix provides a highly available and resilient platform to ensure business continuity and minimize downtime. Nutanix self-healing capabilities ensure that failures are automatically remediated, which reduces the burden on IT staff and minimizes business disruptions caused by unplanned and planned downtime.

Integrated

By giving you the capabilities you need out of the box, Nutanix simplifies the cloud native stack and reduces tedious, time-consuming administrative tasks. Build and manage multicloud deployments quickly and easily by automating routine operational tasks. Built-in security and cost governance enable you to increase the efficiency and security of fast-growing cloud native application environments, while driving financial accountability and aiding regulatory compliance.

Easily Upgraded

With frequent Kubernetes releases and patches, lifecycle management can be a major headache. Nutanix takes the pain out of planning and executing Kubernetes infrastructure upgrades in the busiest and most complex environments.

Predictable and Controlled Costs

As your organization plans its AI and cloud native strategy, the ability to understand, predict and control costs should be built in from the start. A recent [Harvard Business Review article](#) recommends monitoring costs and integrating cost controls.

AI Cost Challenges

- GPUs are expensive resources that should be highly utilized not sitting idle.
- Cloud costs can rapidly get out of control.
- AI environments can span across datacenter to edge to cloud.

Cloud Native Cost Challenges

- Overprovisioned Kubernetes clusters cause sprawl and rising cloud costs.
- Clusters get started but are never retired.
- Complex, multicloud deployments make costs almost impossible to monitor and manage.

More Cost-Effective Infrastructure

The efficiencies built into NCP translate into a more cost-effective infrastructure for your AI and cloud native applications.

Unified Architecture

Nutanix HCI brings one-click simplicity to infrastructure management. Your team uses the same tools everywhere, which dramatically reduces time spent on management and the chance of errors.

Native Hypervisor

Nutanix AHV hypervisor eliminates the need to pay for virtualization.

Full License Mobility

Move licenses between locations to address changing operational requirements while avoiding sunk costs.

Efficient Cloud Operations

Applications and data running in the public cloud on Nutanix Cloud Clusters (NC2) can be up to **53%** lower cost than the native public cloud equivalent.

In a [recent study](#), Nutanix customers achieved average reduction TCO savings of **43%** and five-year ROI of **356%**.

Improved Cost Governance

Nutanix also offers industry-leading multicloud cost governance capabilities with full visibility from a single console. Advanced machine learning algorithms automatically detect cost anomalies and take fast, proactive measures to rightsize resources and help keep costs under control. Nutanix gives you cost metering and a single pane of glass for on-premises and public cloud infrastructure so you can see the big picture.

Nutanix capabilities enable you to:

- Maintain and control costs as your AI models grow.
- Scale containerized AI applications faster and with fewer resources.
- Match resources and spending on AI directly with business demands.
- Better address unique project needs across datacenter, edge, and cloud while lowering TCO.





Tools to Get Started

Private GPT

Maintain complete control over the privacy and security of proprietary company data with a private GenAI chatbot.

GenAI for Code

Boost developer productivity with GenAI-assisted code creation.

GenAI for Content

Enhance the productivity of your marketing and sales teams with the latest AI-powered tools for content creation.

AI-Assisted Document Understanding

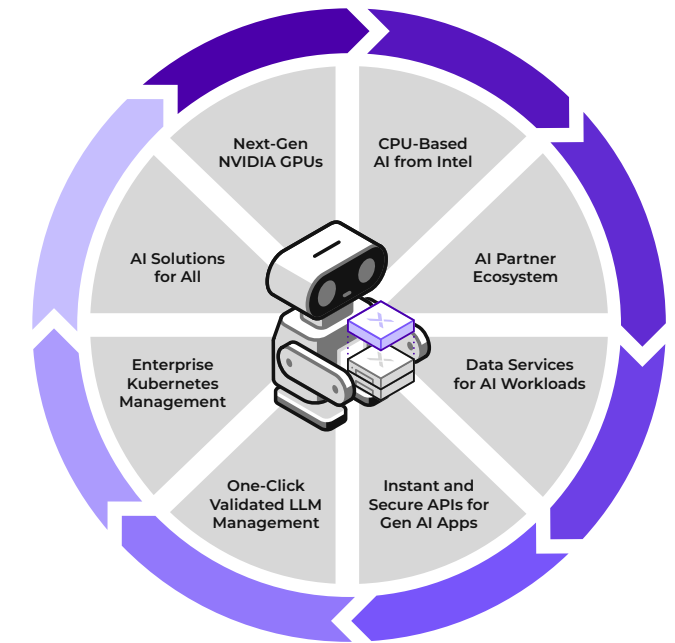
Extract, interpret, and process complex documents without exposing intellectual property or sensitive data.

Nutanix gives you everything you need to tackle the use cases that are most critical for your business and your industry, eliminating the need to integrate disparate hardware and software solutions to create an AI platform and saving months of effort.

Don't let anything stop you from achieving AI and cloud native success.

[Take a test drive](#) to see the Nutanix difference for yourself. Or visit [nutanix.com](https://www.nutanix.com) to learn more.

You can also contact Nutanix at info@nutanix.com or send us a request at www.nutanix.com/demo to set up a customized briefing.



NUTANIX

info@nutanix.com | www.nutanix.com | [@nutanix](https://twitter.com/nutanix)

©2024 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). AI+ModApps-Keyplay-eBook-FY24Q4 07092024