Building a Scalable and Resilient Data Platform *Nutanix Edition*

Oct 2025

Capgemini | NUTANIX

Contents

Executive Summary

Enterprise data platforms have emerged as the cornerstone of digital transformation, with the market expanding from $110.53 billion in 2024 to a projected $221.58 billion by 2030 (https://www.grandviewresearch.com/industry-analysis/enterprise-data-management-market, https://www.imarcgroup.com/enterprise-data-management-market)

Enterprises are overwhelmed with but lack actionable insight. Teams juggle transactional systems, data lakes, real-time streams, machine learning/Enterprise AI pipelines, and governance obligations—often across multiple clouds and datacenters. A modern data platform provides a consistent foundation to ingest, store, process, govern, and serve data—reliably, securely, and costeffectively.

This paper explains the state of enterprise data today, defines a data platform "in a nutshell," outlines why it's needed, details core non-functional requirements (hybrid/multi-cloud, security, performance, integrity/backup), and gives a how-to blueprint using Nutanix products. It closes with processes and best practices for DevOps, DataOps, and MLOps so the platform stays healthy as it scales.

State of Enterprise Data

- **Fragmentation everywhere:** SaaS apps, on-prem databases, data lakes, event streams, file shares, and shadow spreadsheets. Data silos plague 68% of organizations as the primary concern, with large enterprises managing an average of 360+ software tools where only onethird integrate effectively (https://www.dataversity.net/data-strategy-trends-in-2025-from-silosto-unified-enterprise-value/). The data fragmentation creates an economic waste with underutilized data and un-analyzed information that businesses can leverage to make transformative decisions. Data silos create duplicated tooling, brittle pipelines, security gaps, slow time-to-insight and increased cost.
- **Data volume explosion:** Accelerating, with enterprise data growing from 1 petabyte in 2020 to 2.02 petabytes in 2022—a 42.2% annual growth rate that shows no signs of slowing (https://www.statista.com/statistics/1186304/total-enterprise-data-volume-location/)
- **Polyglot compute:** VMs, containers, serverless endpoints, and GPUs—each with different storage, networking, and security needs.
- **Mixed workloads:** Batch ETL/ELT, CDC (Change Data Capture), streaming and batch analytics, BI dashboards, AI/ML fine-tuning & inference, reverse ETL/operational analytics.
- **Governance pressure:** Focus on data sovereignty, privacy (PII – Personally Identifiable Information/PHI – Protected Health Information), auditability, lineage, retention, and eDiscovery. regulatory compliance costs continue escalating, with GDPR violations alone costing companies over $3 billion in 2023, while 75% of enterprises expect storage capacity increases due to privacy regulations. (https://cybernews.com/tech/tiktok-instagram-gdpr-fines-childrendata/, https://www.storagenewsletter.com/2024/09/20/total-enterprise-storage-systemsinstalled-base-expected-to-grow-at-30-9-cagr-between-2020-2025/)
- **FinOps reality:** Unpredictable cloud bills, stranded capacity on-prem, and complex chargeback/showback.

Enterprises require comprehensive data platforms supporting unified data strategies, real-time decision-making, and organization-wide data democratization. This transformation demands balancing security and accessibility while maintaining governance and compliance across increasingly complex hybrid environments.

Modern Data Platform in Nutshell

A modern data platform must bring together compute, storage, networking, security, and data services into a unified fabric that supports transactional and analytical workloads at scale. It should be simple to operate, flexible to extend, and secure by design—ready for both today's enterprise data challenges and tomorrow's AI-driven demands.

They represent a fundamental architectural evolution from traditional silos of file servers, relational databases, and monolithic data warehouses to distributed cloud-native frameworks. The decoupling of data and applications define modern platform architecture, enabling elastic scaling and cost optimization impossible with traditional systems. Data fabric architecture is an advanced data management framework that connects, integrates, and governs data from many sources, enabling seamless access across both hybrid and multi-cloud environments. It is designed to remove data silos and offers uniform data access, analytics, and governance without forcing data to be centrally located. Data mesh patterns implement decentralized, domain-oriented approaches where business units own their data products while leveraging self-serve infrastructure platforms. These patterns support scalability across volume (petabyte scale), velocity (high-frequency streaming), variety (diverse data types), and users (concurrent access) while maintaining performance through query optimization, data partitioning, and advanced compression.

A **modern data platform** is a productized foundation (people + process + technology) that delivers these capabilities as shared, self-service building blocks for transactional and analytical data.

- **Ingest:** Supports high-throughput and low-latency ingestion of transactional data via batch, micro-batch, streaming, and change data capture (CDC) mechanisms.
- **Store:** Optimized for operational workloads with tiered storage (hot/warm/cold) across object, file, and block formats, ensuring durability and performance for frequently accessed transactional datasets.
- **Process:** Enables real-time and near-real-time data processing using SQL engines, Spark, Trino, Flink, and GPU acceleration where needed—tailored for operational analytics and transactional transformations.
- **Transact:** Provides robust support for transactional workloads with ACID-compliant relational databases, low-latency SQL engines, and real-time data access for mission-critical applications.
- **Serve:** Delivers operational insights through BI dashboards, APIs, and visualization tools, enabling fast access to transactional metrics and KPIs.
- **Govern:** Ensures trust and compliance with comprehensive data governance—cataloging, lineage tracking, data quality enforcement, access control, and auditability for transactional data flows.
- **Protect:** Safeguards operational data with enterprise-grade backup, snapshots, replication, disaster recovery (DRaaS), immutability, and ransomware protection.
- **Operate:** Facilitates efficient platform operations with observability, cost governance, automation, and infrastructure-as-code (IaC) for managing transactional data pipelines and services.

Challenges: The Need for a Modern Data Platform for Transactional & Analytical Workloads

Traditional data architectures, whether standalone relational databases for specific applications or centralized enterprise data warehouses are ill-equipped to handle the demands of both modern

transactional and analytical workloads. These legacy systems struggle with the velocity, concurrency, and real-time responsiveness required by today's digital applications. Organizations now need cloud-native, elastic platforms that support high-throughput transactional processing, low-latency access, and unified operations across hybrid environments.

Technical Challenges

- **Siloed operational data & tool fragmentation**: Disconnected systems and tightly coupled pipelines hinder agility and reuse across transactional applications.
- **Rigid scaling models**: Inability to independently scale computing and storage for transactional workloads leads to performance bottlenecks and resource inefficiencies.
- **Hybrid data patterns**: Transactional and analytical systems must support structured and semistructured data with real-time, near real-time, and batch ingestion and processing.
- **Performance variability**: Transactional systems face unpredictable spikes in workload, leading to latency in reads/writes, inconsistent query response times, and throughput degradation.
- **Security exposure**: Mixed-tenant environments and distributed networks increase the blast radius for breaches, especially in systems handling sensitive operational data.
- **Environment drift**: Inconsistent configurations across dev/test/prod and across clouds complicate deployment and reliability of transactional applications.
- **Resilience gaps**: Inadequate backup and disaster recovery strategies result in inconsistent RPO/RTO and weak defenses against ransomware and data corruption.

Business Challenges

- **Delayed time-to-value**: Operational data is often locked in silos, slowing down analytics, ML, and decision-making tied to real-time business processes.
- **Opaque and rising costs**: Hidden costs from data egress, duplication, and idle infrastructure (e.g., orphaned clusters) impact ROI for transactional platforms.
- **Compliance risk**: Operational systems must meet stringent requirements for data retention, access control, and residency—often across jurisdictions.
- **Real-time expectations**: Businesses increasingly demand instant insights and responsiveness from digital and AI-driven applications, which legacy data systems cannot deliver.

What is Needed in Data Platform

Hybrid/multi-cloud or private cloud integration

- **Portability:** Support transactions, run analytics & ML and server end-users where data live; avoid forced large-scale data movement.
- **Consistency:** Common storage semantics, identity, policy, and automation across sites.
- **Brokerage:** Place workloads based on cost, data gravity, and compliance.

Security & compliance

- **Zero-trust by design:** Least-privilege RBAC/ABAC, micro-segmentation, private service endpoints.
- **Data protection controls:** At-rest and in-flight encryption, key management, Object lock/WORM, immutable snapshots, data classification and masking.
- **Operational controls:** Audit logs, posture management, policy-as-code, vulnerability management.
- **Regulatory Requirements:** GDPR for processing EU resident data, HIPAA for healthcare, PCIDSS for finance and SCO2, NIST for hybrid and private clouds.
- **Ethical AI:** Personal info protection, data privacy, explainability, data traceability, data bias removal

Performance & scalability

- **Right storage for right workload:** Block for databases, file for engines needing POSIX/NFS/SMB, object for lakes & AI artifacts.
- **Tiering & locality:** Move cold data to cheap tiers automatically; keep hot data close to computing.
- **Elastic computing:** Mix VMs and Kubernetes, scale independently from storage.

Data integrity & backup

- **End-to-end integrity:** Data durability, checksums, ACID/transactional tables with quality gates.
- **Protection:** Snapshots, replication, near-sync/metro, air-gapped/offsite copies, validated backup and restores.
- **3-2-1(-1-0):** Three copies, two media, one offsite (+1 offline/immutable, 0 verified errors).

Why Build Data Platform with Nutanix

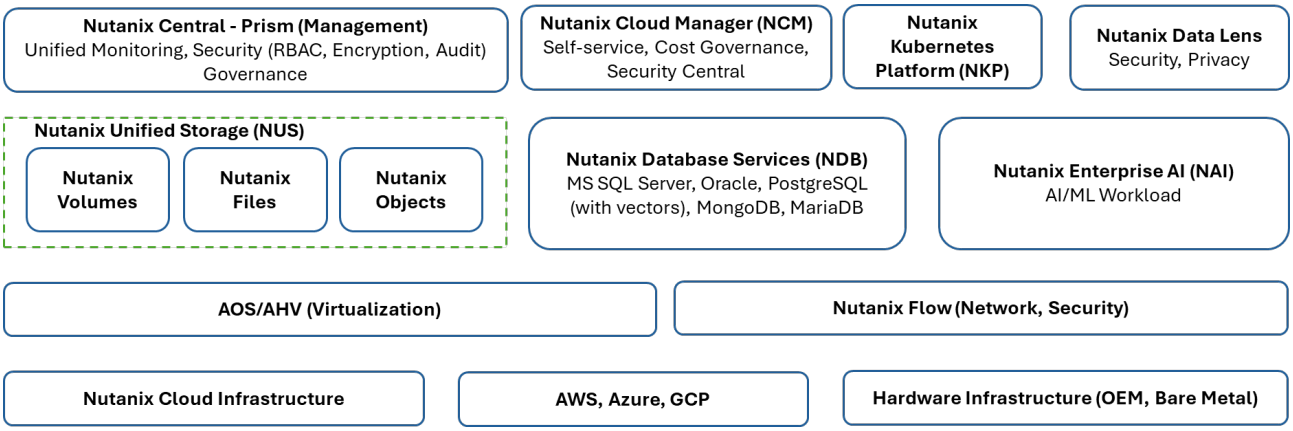
Nutanix platform offers features to simplify datacenter management, streamline operations, and offer a unified approach to running applications and data across hybrid and multi-cloud environments. Its key strengths include hyperconverged infrastructure, unified storage, automated operations, and robust security features. This results in reduced complexity, increased efficiency, and cost savings for organizations.

The following are the key drivers for building data platforms with Nutanix:

- **Simplified data platform management:** Nutanix pioneered HCI, converging compute, storage, and networking into a single, software-defined platform, reducing hardware silos and simplifying management. Nutanix provides a single platform for managing all aspects of the infrastructure, including storage, compute, networking, and virtualization. Nutanix automates many routine tasks, such as provisioning, patching, and scaling, freeing up IT staff for more strategic initiatives.

- **Superior performance:** Distributed storage architecture and features like de-duplication, compression, and tiering optimize data access and application performance.
- **Data resiliency:** Features a self-healing design that protects data and ensures business continuity, even in the face of disruptions. It provides built-in backup and replication capabilities, along with disaster recovery options like synchronous and asynchronous replication ensuring business continuity and data protection.
- **Workload Characteristics:** A modern data platform must support a spectrum of workload needs — from high-performance, latency-sensitive databases and analytics, to long-term archival and batch processing. Nutanix makes this alignment seamless: transactional and real-time analytics workloads can run on high-performance clusters with Nutanix Database Service (NDB) and Files, while archival and batch workloads are efficiently handled with Nutanix Objects and tiering policies. This ensures the right performance and cost profile for each workload without adding complexity.
- **Cost efficiency:** Nutanix can significantly reduce infrastructure costs by consolidating resources and optimizing resource utilization. Nutanix's robust design and automation capabilities help reduce unplanned downtime, minimizing business disruptions and saving cost.
- **Developer productivity:** Accelerate developer productivity through a platform engineering approach that integrates DevSecOps and automation into software development and testing workflows. Enable seamless provisioning of dev/test/prod environments with DBaaS support for both relational and document databases. Leverage API-driven access, automated patching and updates, elastic scaling, and built-in cost optimization to streamline operations and reduce overhead.
- **Enterprise AI:** Nutanix delivers an enterprise-grade AI data platform optimized for ultra-low latency and high-throughput data access. By leveraging Nutanix Unified Storage, organizations can seamlessly feed structured and unstructured data into AI/ML pipelines. The platform provides end-to-end AI tooling, from data ingestion to model deployment, tightly integrated with Prism for full-stack observability, governance, and lifecycle management. This ensures secure, performant, and scalable AI operations across hybrid and multi-cloud environments–ideal for real-time inference, continuous training, and cost-efficient scaling.
- **Hybrid and multi-cloud flexibility:** Nutanix provides a consistent operating experience across on-premises, public cloud, and edge locations, simplifying management of hybrid and multicloud environments. It facilitates smooth application and data migration to the cloud without the need for costly refactoring. Nutanix offers flexible deployment options and portable licenses, enabling organizations to scale resources and adapt to changing business needs.
- **Security:** Nutanix includes robust security features like access control, data protection, encryption, and self-healing systems, ensuring business continuity and cyber resilience.
- **Compliance and regulatory requirements:** Nutanix data platforms enforces policies for data residency, encryption, and auditability to meet industry and regional compliance standards.

Nutanix Data Platform Software Stack



How to Build Data Platform with Nutanix

Nutanix offers a unified, software-defined approach that addresses critical enterprise challenges through integrated data services, AI-ready infrastructure, and simplified operations.

Core Nutanix Building Blocks

- **Nutanix Cloud Infrastructure (NCI)** with Distributed Storage, AHV Hypervisor, networking, and security: Hyperconverged compute and storage foundation (linear scale, RDMA-ready performance). Also supported is the option of disaggregated (3rd party) storage with the Nutanix NCI-Compute Stack.
- Nutanix Unified Storage (NUS):
  - **Objects** (S3-compatible) for data lakes, ML artifacts, backups; supports objectlock/immutability.
  - **Files** (NFS/SMB) for POSIX-style analytics engines, data warehouse, and user shares.
  - **Volumes** (iSCSI block) for applications requiring block level access.
  - **Data Lens** for data governance, security, protection against anomalies and compliance risks
- **Nutanix Kubernetes Platform (NKP):** Managed Kubernetes for modern data/ML stacks; CSI & CNI integrations.
- **Nutanix Database Service (NDB):** Implements Database-as-a-Service for hybrid multi-cloud environments supporting SQL databases (Oracle, SQL Server, MySQL, PostgreSQL) and NoSQL (MongoDB) with Vector database support for AI workloads. Lifecycle automation for databases (provision, clone, patch, time-machine).
- **Nutanix Flow** Network Security & Flow Virtual Networking: Micro-segmentation, VPCs, and traffic policy.
- **Nutanix Cloud Manager (NCM)** including Self-Service & Cost Governance: Infrastructure-as-aCode (IaaC)/app automation and FinOps.
- **Prism Central & Nutanix Central:** Fleet management, operations, and SaaS-based oversight.
- **Data Protection & DR** (snapshots, replication, NearSync, Metro) + Nutanix Mine ecosystem for integrated backup and protection.

- **Nutanix Enterprise AI (NAI):** LLM deployment with NVIDIA NIM microservices integration and OpenAI-compatible APIs.

These products create a solid foundation to implement data platform patterns. Below are production-ready patterns and how to realize them on Nutanix.

Design Pattern 1: Data Platform for Transactional Workload with OLTP

Here is an architecture for deploying Online Transaction Processing (OLTP) workloads on the Nutanix Cloud Platform. OLTP systems are foundational to business operations, demanding exceptional performance, sub-millisecond latency, stringent data integrity (ACID compliance), and continuous availability.

Key Features

Purpose and Workload Profile: OLTP systems primarily designed for processing database transactions, which include inserting, updating, or deleting data characterized by short atomic but high-volume transactions, milli-second latency.

Performance: The processing times for OLTP transactions measured in milliseconds or less, with updates occurring in real-time, triggered by users or applications. Nutanix supports highperformance with localized I/O, the utilization of all-flash storage, and an optimized data path.

Concurrency: OLTP systems must inherently support many concurrent users and transactions, maintain data integrity and minimize response times for any database changes.

Data Integrity: An important aspect of OLTP is ACID (Atomicity, Consistency, Isolation, Durability) compliance. It is typically achieved through advanced consensus protocols and distributed transaction management, ensuring that transactions are processed reliably and accurately.

Database Management: Nutanix Database Service (NDB) supports a broad range of enterprise databases including Oracle, Microsoft SQL Server, PostgreSQL, MongoDB, MySQL, and MariaDB. Delivered as a fully managed Database-as-a-Service (DBaaS), NDB simplifies operations across hybrid and multi-cloud environments with built-in automation, high availability, and multi-cluster support.

Key capabilities include:

- **Accelerated Provisioning:** Rapid deployment of databases along with associated compute, storage, networking, and policy configurations—reducing provisioning cycles from days to minutes. High availability (HA) and disaster recovery (DR) configurations are built-in.
- **Standardized Templates:** Predefined provisioning templates minimize environmental drift and enforce consistent deployment aligned with enterprise best practices.
- **Thin Cloning for Test/Dev:** Instantly clone production databases for testing or QA without impacting live workloads. Includes support for data scrubbing and post-processing.
- **Time Machine Recovery:** Restore or clone databases from historical snapshots to support rollback, recovery, or environment recreation.
- **Integrated Disaster Recovery:** Native support for DR drills, failover, switchover, and switchback operations ensures business continuity.
- **Security & Compliance:** Role-based access control (RBAC), encryption, audit logging, and policydriven controls meet enterprise security and regulatory requirements.
- **Automated Lifecycle Management:** Streamline patching, scaling, and backup operations to reduce manual overhead and improve reliability.

- **API-Driven Automation:** Full support for GUI, CLI, and API workflows enables self-service provisioning and DevOps-centric automation for database operations.

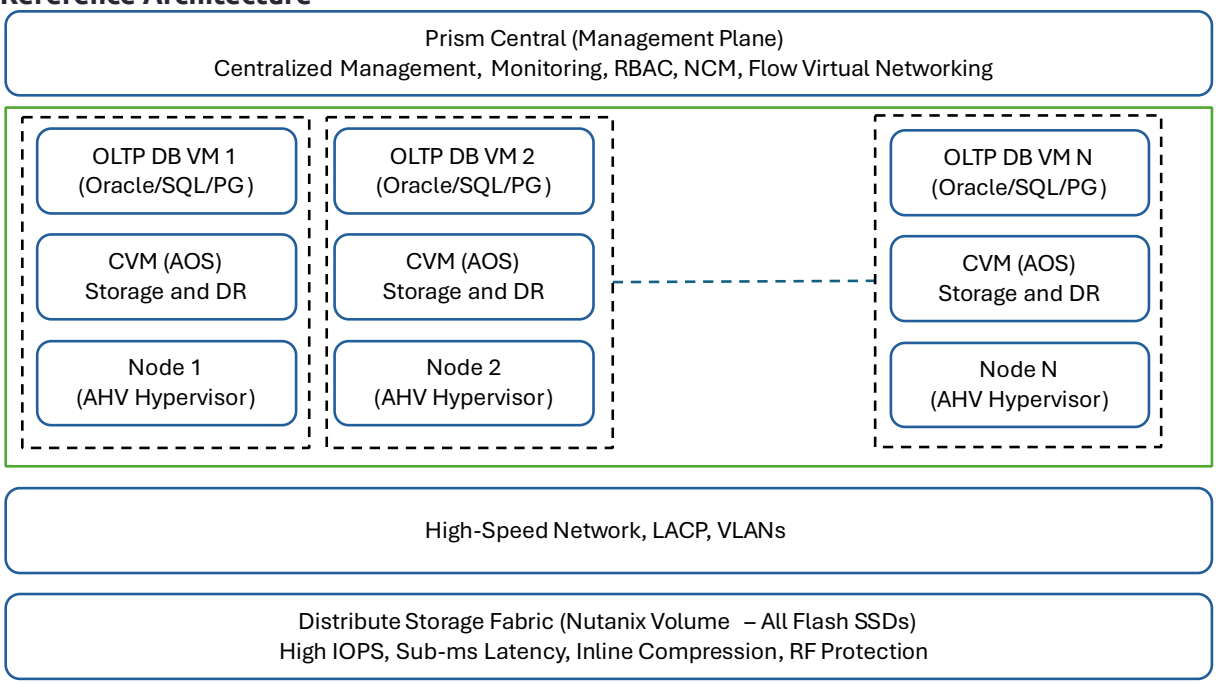
Availability and Data Protection: High availability commonly achieved through multiple data backups and resilient infrastructure designs. Nutanix integrated features such as snapshots, remote replication, and metro-level availability contribute to this resilience.

Scalability: The Nutanix solution offers flexibility to start with modest deployment and scale both performance and capacity non-disruptively as organizational needs evolve.

OLTP Requirements vs Nutanix Capability/Feature

Requirement	Nutanix Capability/Feature	Description
High Performance	All-Flash Nodes, Data Locality, NVMe-oF, Nutanix Volumes	Sub-millisecond latency, localized I/O, optimized for high IOPS, intelligent caching
Sub-millisecond Latency	Nutanix Volumes, All-Flash Storage, CVM Optimization	Optimized data path and caching for rapid transaction processing.
High Concurrency	AOS Distributed Storage, Scalable Architecture	Supports large volumes of concurrent transactions without sacrificing consistency.
Data Integrity (ACID)	AOS Replication Factor, NDB	Ensures transactional consistency and accuracy through robust data protection and database service features.
High Availability	Replication Factor, Metro Availability, Snapshots, Self-healing	Native data protection, disaster recovery, and automated failover capabilities.
Scalability	Nutanix HCI Architecture, NDB	Start small and scale compute, storage, and database instances nondisruptively.
Real-time Processing	Optimized I/O Path, In-memory Caching	Data locality, write optimization and Dynamic cache management
Security	Nutanix Flow, Role-Based Access Control (RBAC)	Software-defined firewalls (microsegmentation) and granular access control for critical assets.
Simplified Management	Nutanix Prism Central	Unified interface for configuring, monitoring, and managing the virtual infrastructure, including OLTP databases.

Reference Architecture



Design Pattern 2: Data Platform with DevSecOps

Nutanix offers a comprehensive, enterprise-ready modern data platform that unifies RDBMS management, advanced DevSecOps capabilities, and hybrid cloud operations through a single integrated stack. This integrated approach eliminates traditional data platform complexity by providing unified storage services, comprehensive Infrastructure-as-a-Code automation, and built-in DevSecOps tooling within a single, scalable architecture maintaining with comprehensive compliance coverage for HIPAA, SOC 2, ISO 27001, and NIST.

Key Features

Database Management: Nutanix Database Service (NDB) supports Oracle, Microsoft SQL Server, PostgreSQL with pgvector, MySQL, MariaDB, MongoDB. The service delivers automated lifecycle management, high availability configurations, and multi-cluster support.

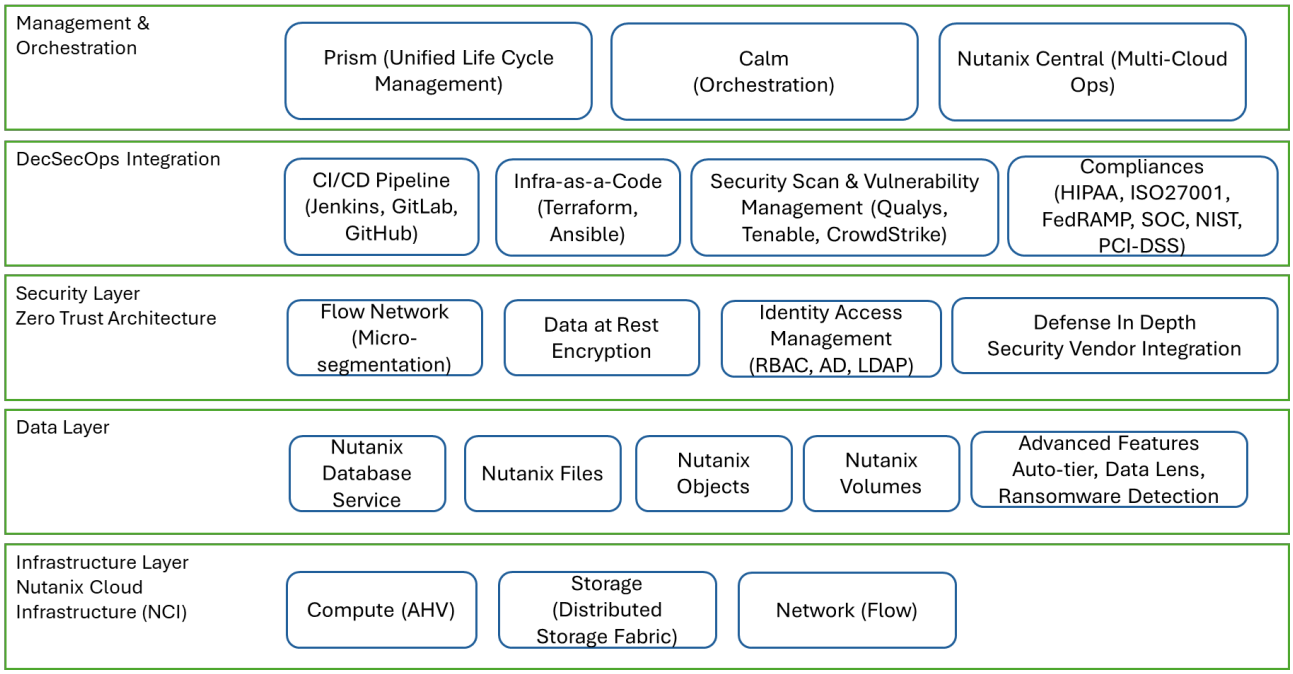
Infra-as-a-Code: Supports multiple toolchains including Terraform providers, Ansible collections, and native Self-Service orchestration. The Terraform provider offers comprehensive resource coverage supporting VM provisioning, network management, storage configuration, and database lifecycle operations. Ansible integration provides configuration management capabilities complementing infrastructure provisioning workflows.

CI/CD pipeline integration: Supports Jenkins plugins, GitLab webhook support, and comprehensive REST API coverage enabling API-first automation strategies. The platform supports popular DevOps workflows including automated testing, deployment, and infrastructure management.

Security scanning and vulnerability detection: Integrates with Qualys, Tenable, and CrowdStrike for comprehensive vulnerability assessment and remediation. The platform provides one-click patching capabilities through Nutanix LCM Full-stack Update Manager, enabling automated vulnerability remediation with metadata-guided resolution processes.

Network micro-segmentation: Flow Network Security provides application-centric security policies with category-based tagging systems that follow workloads across environments. The system supports zero-trust networking with default-deny policies and comprehensive east-west traffic control integrated directly into the AHV hypervisor.

Reference Architecture



Design Pattern 3: Data Platform for Batch Processing.

Nutanix provides a robust data platform designed for handling batch processing, leveraging core Nutanix Cloud Infrastructure providing the hyperconverged infrastructure foundation for Nutanix Unified Storage (NUS) and Nutanix Database Services (NDB)

Key Features

Purpose and Workload Profile: Batch and Online Analytical typically include Extract-Transform-Load (ETL) processes, data aggregation, and periodic reporting, which require high throughput rather than low latency.

Scalable Data Lake: Nutanix Unified Storage (NUS) serves as the underlying storage layer, offering unified file, block, and object storage with S3 compatibility for data lakes.

Performance: Nutanix Unified Storage (NUS) provides Sequential I/O for large datasets ensuring high throughput.

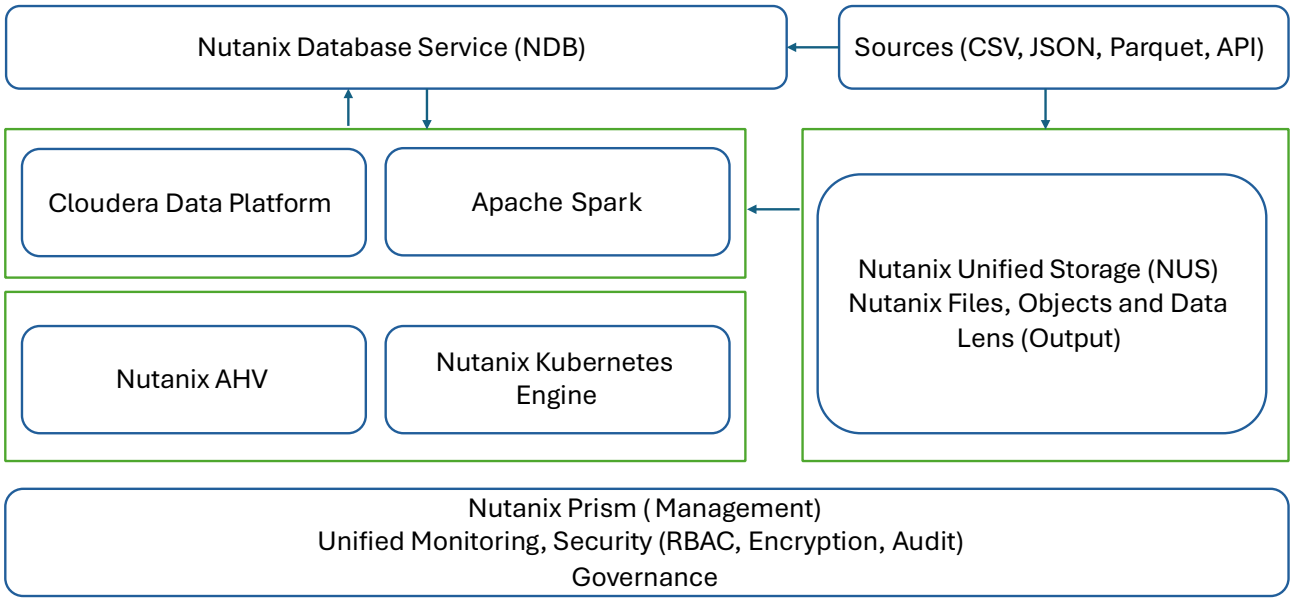
FinOps & Cost Control: NUS provides Tier data based on access frequency (hot/cold storage). Nutanix management tools provide intelligent sizing recommendations, identifying underutilized VMs or storage in Nutanix clusters.

Batch and Online Analytical Processing features with Nutanix

Component	Description	Role in Batch processing

Nutanix HCI (AHV+AOS)	Hyperconverged platform converging compute, storage, and virtualization. Also, Compression, deduplication, erasure coding, snapshots.	Provides scalable nodes for running VMs/containers with data locality for efficient batch processing. Reduces storage footprint for large batch datasets and enables fast recovery.
Nutanix Unified Storage (NUS)	Software-defined storage consolidating file, block, and object protocols with data services like analytics and lifecycle management.	Acts as a data lake for storing large OLAP datasets, supporting S3 for big data tools.
Cloudera Data Platform (CDP)	Integrated big data platform with Hadoop, Spark, Hive, and Impala.	Handles batch workloads (Spark/MapReduce) and OLAP queries (Impala for interactive, Hive for batch).
Apache Spark	In-memory processing engine.	Executes batch ETL jobs efficiently on Nutanix's high-throughput storage.
Nutanix Prism	Centralized management console.	Enables infrastructure management, one-click upgrades, monitoring, and scaling for the entire platform.
Nutanix Database Service (NDB)	MS SQL Server, PostgreSQL, Oracle, MongoDB storing data for batch processing	Multi-cluster and HA support for scalability, performance, high availability

Reference Architecture



Design Pattern 4: Data Platform for Real Time Workload with OLAP

Key Features

Purpose and Workload Profile:

OLAP systems fundamentally designed for multi-dimensional analytical queries that require complex aggregations, drill-down capabilities, and slice-and-dice operations across massive datasets.

Performance:

Unlike OLTP systems that prioritize low-latency transactions, OLAP workloads favor throughput over latency processing sequential I/O patterns optimized for reading large volumes of historical data. Nutanix delivers high performance with parallel execution over all-flash storage, intelligent caching, and data locality. Store hot analytical tables on fast tiers; keep cold data in cost-efficient object storage with automatic lifecycle policies.

Concurrency:

Hundreds to thousands of simultaneous dashboard and API queries. Scale out computing independently from storage using VMs or Nutanix Kubernetes Platform (NKP).

Availability and Data Protection:

Platform-level snapshots, replication, NearSync/Metro availability for critical datasets, plus ObjectLock/WORM for ransomware resilience. Validated runbooks for restore and failover.

Scalability:

Start small and scale linearly. Expand computing (VMs/K8s nodes), storage (Objects/Files/Volumes), databases (SQL Server, Oracle, MySQL, PostgreSQL, MariaDB, MongoDB) and network domains nondisruptively. Support multi-cluster and hybrid deployments.

Governance & Security:

Unified identity, RBAC/ABAC, encryption in flight/at rest, Nutanix Flow micro-segmentation, audit logs, catalog/lineage integration, and policy-as-code.

FinOps & Cost Control:

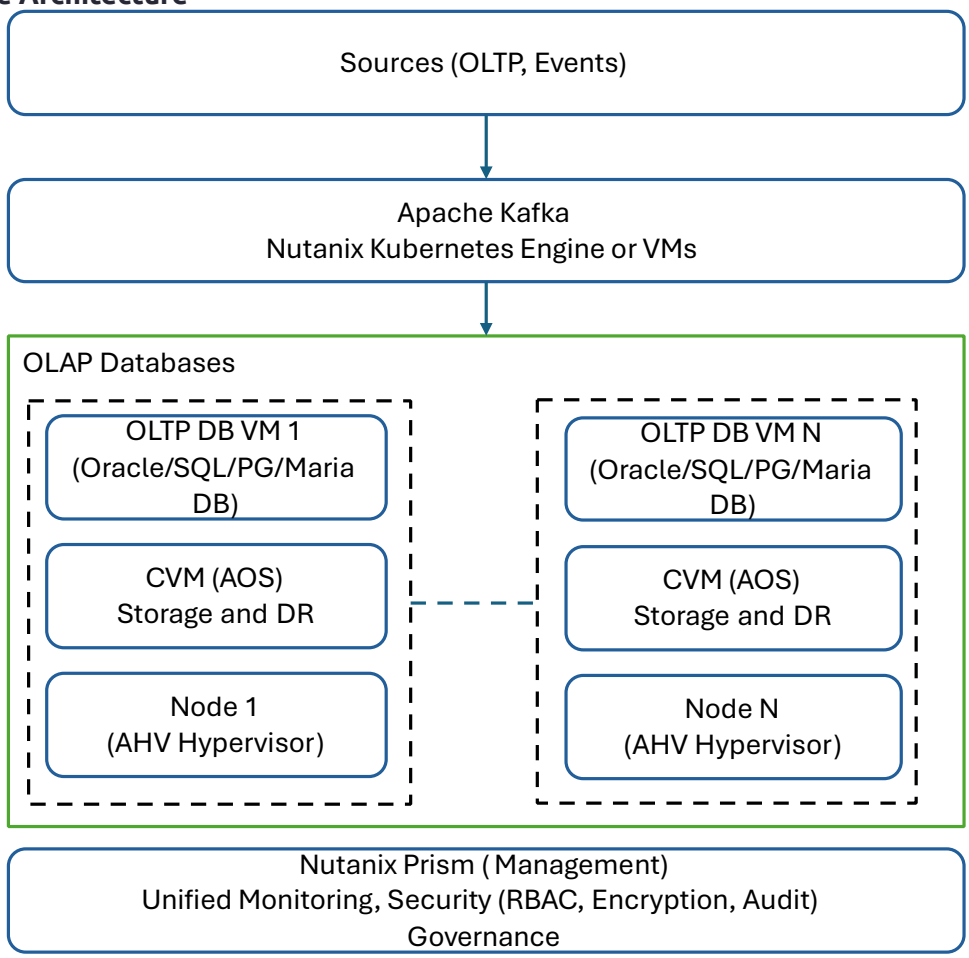
Tiering (hot/warm/cold), auto-suspend dev/test clusters, showback / chargeback via Nutanix Cloud Manager (NCM) Cost Governance.

Real Time OLAP Features with Nutanix Database Service

Feature / Database	SQL Server	Oracle	PostgreSQL	MariaDB
Primary OLAP support	Yes, via SQL Server Analysis Services (SSAS) providing native OLAP cubes, multidimensional and tabular models	Yes, with Oracle OLAP Option, and powerful analytical functions and materialized views	Supports OLAP via extensions like CUBE/ROLLUP, materialized views, and foreign data wrappers	OLAP supported using ColumnStore engine and Xpand distributed SQL for analytics
Real-time data ingestion & querying	Supports real-time analytics with in-memory OLAP cubes and DirectQuery for live connections	Real-time analytics using materialized views and inmemory features	Good support for streaming / CDC data with tools like Debezium, enabling near-realtime analytics	Supports realtime via via MaxScale, CDC, and streaming integrations

Feature / Database	SQL Server	Oracle	PostgreSQL	MariaDB
Multidimensional analysis	Full multidimensional OLAP cubes, drilldown, slice-and-dice, KPIs, hierarchies	Strong multidimensional OLAP with cube design and drilldown features	Supports multidimensional grouping with extensions and queries	Supported via distributed SQL and columnar storage
Query language	T-SQL with extensions for OLAP; MDX for cubes	PL/SQL with OLAP analytical functions	SQL with analytical extensions (CUBE, ROLLUP)	SQL with OLAP extensions and still MySQL compatible
Materialized views / preaggregation	Yes, with SSAS and indexed/materialized views	Yes, extensive materialized views and caching	Materialized views supported	Materialized views and caching in ColumnStore
Scalability for OLAP	Scales vertically, supports distributed query with PolyBase	Scales well vertically and in RAC for analytics workloads	Scales well on Nutanix; horizontal scaling via extensions	Designed for scale-out with Xpand and ColumnStore
Integration with BI tools	Strong integration with Power BI, Excel, Tableau	Strong ecosystem including Oracle BI, Tableau, Power BI	Broad support: Tableau, Power BI, Grafana	Integrates with Tableau, Power BI, Grafana
Real-time streaming / CDC support	Supports CDC, realtime ETL, and Streaming Analytics	Supports CDC with GoldenGate for real-time replication	CDC via Debezium and logical replication	Supports realtime streaming via MaxScale
Columnar storage for analytics	Available with inmemory column store indexes	Advanced columnar compression in Exadata and other offerings	Via extensions like cstore_fdw	MariaDB ColumnStore engine provides columnar analytics
Suitability for real-time OLAP	Highly suitable with dedicated OLAP tools and in-memory acceleration	Highly suitable for enterprise analytics, real-time materialized views	Good opensource solution with extensions and streaming support	Growing OLAP focus with distributed SQL and ColumnStore

Reference Architecture



Design Pattern 5: Enterprise AI Platform

The solution provides detailed guidance for building a production-ready Enterprise AI platform leveraging the **Nutanix Kubernetes Platform (NKP)** with **Nutanix Enterprise AI (NAI)**, **NVIDIA GPU acceleration**, and modern MLOps practices to deliver a scalable, secure, and cost-effective enterprise ML platform to support concurrent users performing ML fine-tuning and inference workloads.

Key Features

Scalable Infrastructure: Nutanix AHV and NKP for running AI applications.

AI Storage: Nutanix Files for high-performance data storage, Nutanix Objects for S3 compatible artifact storage.

GPU Hardware: NVIDIA GPU with GPU passthrough for high-performance and vGPU for multi-tenant effective resource sharing.

AI Providers Tool Integration: Cloud-native AI platform with NVIDIA NIM integration, Hugging Face model catalog access, and multi-cloud deployment support.

MLOps: MLflow Enterprise for distributed model registry, experiment tracking, and lifecycle management. Kubeflow for comprehensive ML workflow orchestration with pipelines, training operators, and model serving capabilities.

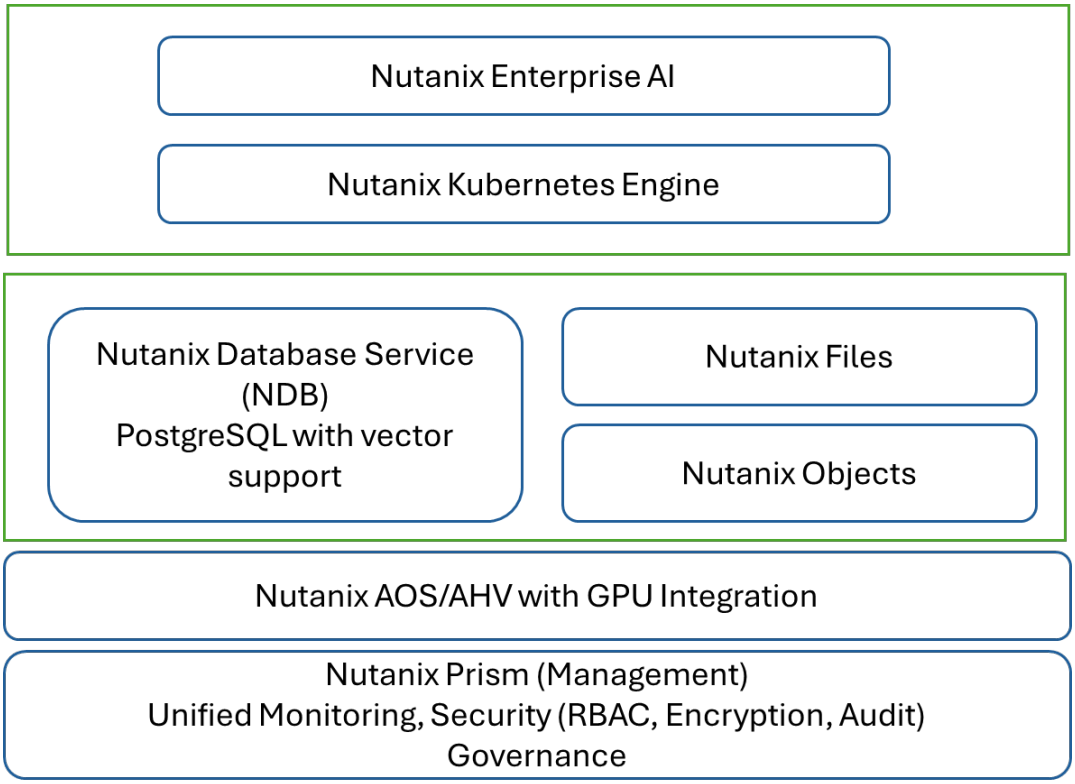
ML Databases: Nutanix Database Service (NDB) PostgreSQL with vector support for storing vector embedding to enable RAG based pipelines.

Security: Flow network policies per team; secrets management; signed containers.

ML Platform Features with Nutanix

Platform Requirement	Nutanix Feature/Product	Key Capabilities
Scalable GPU Compute for Thousands of Users	Nutanix AHV with GPU Support	Hypervisor with native GPU passthrough and vGPU capabilities
Container Orchestration	Nutanix Kubernetes Platform (NKP)	GPU-aware scheduling, Built-in AI Navigator assistant, Fleet management capabilities, Integrated with NVIDIA GPU Operator
High-Performance Training Data Storage	Nutanix Files 5.0	Distributed file system with NFS/SMB support
Model & Artifact Storage	Nutanix Objects	S3-compatible object storage platform, Versioning and lifecycle policies, Integration with MLflow/Kubeflow
Enterprise AI Platform	Nutanix Enterprise AI (NAI)	Cloud-native AI platform with model catalog, NVIDIA NIM integration, Hugging Face model access, LLM deployment and serving, multi-cloud support.
MLOps & Model Management	Integrated MLflow + Kubeflow	Complete ML lifecycle management, Experiment tracking, Model registry and versioning, Pipeline orchestration, A/B testing and canary deployments
ML Databases	PostgreSQL with vector support	Nutanix Database Services with store, query, and index high-dimensional vectors. Data protection with application consistent snapshots and cloning.
Multi-tenant Isolation	Nutanix Flow Security Networking & Nutanix Flow Virtual Networking	Virtual Private Cloud, micro-segmentation and security policies, application-centric policies, network isolation, traffic visualization, threat detection
Cost Management	Nutanix Cost Governance	Usage tracking and chargeback, Resource usage tracking, Chargeback/showback, Budget alerts, Optimization recommendations
Infrastructure Monitoring	Prism with AI/ML	Predictive analytics, Anomaly detection, Capacity planning, Performance optimization

Reference Architecture



Processes for Data Platform

Building a data platform involves structured phases to ensure it meets business needs while prioritizing security, governance, protection, performance, and cost efficiency. Below is detailed definition of the processes to follow for building and operating a data platform using Nutanix products.

Planning

- Identify stakeholders (e.g., data engineers, analysts, compliance officers) and gather requirements for data sources, volume, velocity, variety, and use cases.
- Define key objectives and business goals for data platform.
- Discover existing data and applications. Define new data applications to uncover new insights.

Design and Architecture

- Design a layered architecture: Ingestion layer (e.g., ETL tools), storage layer (e.g., data lakes/warehouses), processing layer (e.g., Spark for analytics), and consumption layer (e.g., APIs/dashboards).
- Incorporate security by design: Use role-based access control (RBAC), encryption at rest/transit, and multi-factor authentication.

- Ensure governance: Implement data cataloging, metadata management, and lineage tracking to maintain data quality and traceability.
- Plan for data protection: Include anonymization/masking for sensitive data and backup strategies.
- Optimize for performance: Choose scalable storage and databases based on OLTP, OLAP and ML platform considerations.

Implementation and Development

- Build in iterations: Start with a minimum viable platform (MVP) focusing on core data pipelines.
- Integrate security tools: Deploy firewalls, intrusion detection, and vulnerability scanning.
- Set up governance frameworks: Define data policies, auditing logs, and approval workflows for schema changes.
- Implement protection measures: Use differential privacy or tokenization for PII; ensure regular data backups and disaster recovery tests.
- Control costs: Monitor resource usage during development and implement tagging for cost allocation.

Testing and Validation

- Perform unit, integration, and load testing to verify performance under peak loads.
- Conduct security penetration testing and compliance audits.
- Test data protection with breach simulations and recovery drills.

Deployment

- Use CI/CD pipelines for automated, secure deployments.
- Roll out in stages (e.g., dev/staging/prod) with monitoring for security and performance issues.
- Document all processes for governance and handover.

Monitoring and Maintenance

- Continuously monitor key metrics like performance (latency, throughput), security (anomaly detection, data breaches) and cost (resource consumption)
- Schedule regular maintenance: Patch updates, backups, application upgrades
- Automate alert and notification for key events related to performance and cost.
- Identify performance bottlenecks and create an action plan for optimization.
- Conduct daily security scans and weekly audits to detect vulnerabilities.

- Enforce governance: Review access logs, permission and policy checks. • Compliance check ISO 27001, SOC2, GDPR, NIST if applicable

Call to Action - Implementation roadmap (90-day starter)

In 90 days, the goal is to build a secure, scalable and high performant data platform for advanced analytics, AI, and multi-cloud operations. Days 0–30 focus on deploying Nutanix clusters, securing VMs, and organizing data into Bronze/Silver/Gold zones. Days 31–60 enhance capabilities with Spark, MLflow, or streaming platforms like Kafka, alongside robust backup/DR solutions. By Days 61–90, enterprises monitor with Prism, onboard additional applications, and conduct DR drills, ensuring a scalable, AI-ready data ecosystem.

Days 0–30

- Identify data owners and profile current data application landscape.
- Stand up Nutanix cluster(s) with **NCI**; enable **Objects, Files, Volumes**.
- Deploy **NKP** or **AHV** based VMs and baseline security (Flow policies, private networking, IAM integration).
- Land initial data via batch and CDC; establish **Bronze/Silver/Gold** zones on Objects.
- Deploy databases based on data types (structured, unstructured or vector) using Nutanix Database Services **(NDB)** for OLTP or OLAP workload.
- Define Role Based Access Control (**RBAC**)

Days 31–60

- Add Spark or Cloudera Data Platform on NKP in case of OLAP workload.
- Enable MLflow or KubeFlow in case of machine learning/AI workload.
- Add streaming (Kafka/Flink) if needed for streaming data analytics.
- Implement backup/DR tiers (immutable objects, snapshots/replication, NearSync for critical sets).
- Roll out NCM Self-Service blueprints and GitOps for data apps.

Days 61–90

- Use **Prism** for monitoring and audits.
- Onboard data application from other business domain; define chargeback/showback; codify SLOs and quality gates.
- Conduct DR drill and ransomware recovery rehearsal.

Conclusion

A modern data platform turns scattered infrastructure and tools into a coherent product that teams can trust. Nutanix provides a powerful substrate—unified storage, consistent Kubernetes/VM compute, integrated security, lifecycle automation, and robust protection—on which to implement

proven data patterns (Lakehouse, Streaming, Operational Data Store, Machine Learning, and Log Analytics). Pair these capabilities with disciplined DevOps/DataOps/MLOps and you'll deliver reliable, governed, and cost-effective data products on enterprise scale.

Publication year

Oct 2025

About Nutanix

Nutanix is a global leader in cloud software, offering a unified platform to run applications and manage data across on-premises datacenters, public clouds, and edge locations. Our softwaredefined architecture combines compute, storage, virtualization, and networking, and adapts to various hardware and cloud options. With built-in flexibility, scalability, and resilience, Nutanix helps you simplify operations, boost performance, and maintain control.

About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 340,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in AI, generative AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2024 global revenues of €22.1 billion. Make it real. | www.capgemini.com