

CASE STUDY

Nutanix Transforms Enterprise Data Analytics with Dremio's Data Lakehouse Platform

At a Glance

The Customer

NUTANIX

Challenge

Nutanix faced complex hierarchical queries on 20+ million record datasets that severely impacted production database performance, while data silos across different teams made unified analytics nearly impossible.

Solution

Nutanix implemented Dremio's data lakehouse platform as their unified "Data as a Service" solution, leveraging Apache Iceberg integration and virtualization capabilities to eliminate performance bottlenecks while maintaining on-premises data sovereignty.

Results

- Query performance improved from minutes to milliseconds for complex hierarchical queries on 20+ million record datasets
- Production database load reduced by eliminating resource-intensive analytical workloads
- Real-time dashboard refresh capabilities enabled for customer-facing applications
- Cross-team data access democratized through unified platform serving multiple business units

The Customer

Nutanix began as a hardware startup and evolved into a comprehensive software company over three years of rapid growth. This transformation brought significant data challenges as the organization scaled across multiple teams, each maintaining their own databases and data sources in silos. With data scattered across MongoDB, MySQL, PostgreSQL, and various cloud storage systems, Nutanix found themselves struggling to deliver the unified analytics and real-time insights their growing business demanded.

The Challenge

Nutanix faced a complex web of data management challenges that threatened their ability to make data-driven decisions at scale. Their enterprise Jira data, spanning multiple projects and containing over 20 million records, required complex hierarchical queries that severely impacted production database performance. When analysts attempted to run recursive queries joining multiple large tables—some containing 6-8 million records each—the production systems would slow to a crawl or fail entirely with memory issues.

The company's growth had created data silos across different teams, with finance data in one system, application logs in another, and operational metrics scattered across various platforms. This fragmentation made it nearly impossible to correlate insights across business functions or create comprehensive dashboards for management decision-making. Additionally, their customer-facing applications required low-latency performance, but the existing infrastructure couldn't support real-time analytics without compromising core system performance.

Schema changes were constant as new software patches were released, requiring manual restructuring of tables and downstream applications. The lack of auto-schema evolution capabilities meant significant engineering overhead with each system update. Most critically, the team needed to shift high-compute analytical workloads away from production databases while maintaining the ability to refresh data on-demand for ad-hoc business requests.

The Solution

Nutanix implemented Dremio's data lakehouse platform as their unified "Data as a Service" solution, building it on a Kubernetes-based architecture integrated with Nutanix Object Storage for on-premises data sovereignty. The platform serves as a one-stop shop for data democratization, bringing together data from disparate sources through Dremio's virtualization capabilities.

The team leveraged Dremio's Apache Iceberg integration to migrate their massive enterprise Jira dataset—a complex view built from 17 different tables—into Iceberg format stored on Nutanix Object Storage. This transformation enabled auto-schema evolution, eliminating the manual overhead of managing structural changes. Dremio's reflections feature created optimized, distributed data representations that dramatically improved query performance while maintaining data freshness.

Using Dremio's virtual dataset (VDS) capabilities, Nutanix created business logic layers on top of their physical data sources (PDS), implementing role-based access control integrated with their RBAC system and OKTA authentication. This approach enabled project-based data spaces where team members automatically receive appropriate access permissions without manual intervention.

The platform connects seamlessly to Tableau, Power BI, and custom Python applications through Dremio's APIs, providing flexible consumption options for different user personas. Event-driven APIs were implemented with Airflow integration, allowing scheduled updates that intelligently check backend conditions before refreshing dependent reflections and downstream dashboards.

Results

The transformation delivered remarkable performance improvements that fundamentally changed how Nutanix approaches data analytics. Complex hierarchical queries that previously took minutes or failed entirely now execute in milliseconds, even when joining tables with 20+ million records. This performance breakthrough enabled real-time dashboard capabilities for customer-facing applications without impacting production system performance.

By shifting analytical workloads from production databases to the Dremio platform, Nutanix freed up significant computational resources, improving overall system performance for critical business applications. The unified platform eliminated data silos, enabling cross-functional analytics that combine finance, operational, and support data for comprehensive business insights.

The solution provided managers with unprecedented visibility into developer productivity and project progress through consolidated Jira analytics, enabling better resource allocation and project planning decisions. On-demand refresh capabilities empowered business users to access current data for ad-hoc analysis without requiring IT intervention, dramatically improving decision-making speed and accuracy across the organization.

ABOUT DREMIO

Dremio is the intelligent lakehouse platform for the business, serving hundreds of global enterprises, including Maersk, Amazon, Regeneron, NetApp, and S&P Global. Based on open-source technologies like Apache Iceberg and Apache Arrow, Dremio provides an open lakehouse architecture enabling the fastest time to insight and platform flexibility at a fraction of the cost.

Learn more at www.dremio.com.

Dremio and the Narwhal logo are registered trademarks or trademarks of Dremio, Inc. in the United States and other countries. Other brand names mentioned herein are for identification purposes only and may be trademarks of their respective holder(s). © 2025 Dremio, Inc. All rights reserved.