Hitachi Vantara

Solution Profile

Hitachi Data Lakehouse: Powering Trusted AI and Analytics at Scale

Analyze data in place with consistent governance and control

Analyze Data in Place

Simplify analytics by running high-performance SQL directly on governed data using open tables and federated access, without duplicating pipelines

Scale With Control

Scale analytics and AI workloads independently while enforcing life-cycle policy, workload placement and predictable performance as data grows

Build Trusted Outcomes

Apply consistent governance, lineage, data quality and retention so analytics and AI remain accurate, auditable and compliant across datasets

Too Many Systems, Too Many Pipelines, and Still No Trusted View.

As your data grows — across unstructured files, objects, operational systems, SaaS platforms and multiple clouds — complexity becomes part of your daily reality. What once lived in a few well-understood systems now spans data warehouses, data lakes, file shares and cloud services — each with different access models, metadata and governance rules. Individually, those decisions often made sense. Together, they leave you coordinating across too many systems just to answer routine analytical questions.

To keep AI and analytics moving, you often end up moving data. Each new use case brings another pipeline, ingestion job or reshaping effort to fit the next engine. Over time, those pipelines pile up: They become harder to manage and easier to break. Failures slow delivery, cloud egress fees climb and teams still struggle to provide a consistent view of the data that analytics and models rely on.

At the same time, confidence in the data weakens. Lineage is difficult to trace end to end. Data quality issues surface late, after dashboards or models are already in use. Sensitive or regulated data appears in places it shouldn't, making retention, auditability and compliance harder to enforce. When audits or security questions arise, answers rely more on manual coordination than built-in control.

As regulatory pressure increases and cyber threats go after your data, more data locations only create more exposure.

Without reducing movement, enforcing governance where data lives, and controlling cost as you scale, duplication grows and control weakens.

60%

By 2027, 60% of AI use cases will fall short of their intended impact due to weak data governance.

— Gartner

Hitachi Data Lakehouse: Powering Trusted AI and Analytics at Scale

AI and Analytics Need a New Operating Model

Modern AI and analytics require an operating model that reduces rework when the next question hits. To move faster, you need to work directly on data, with fewer hand-built pipelines sitting between you and an answer. That means fewer copies, less dependency on fragile ingestion chains, and fewer changes that ripple downstream.

That model starts with running analytics where data already lives. Open table formats provide consistent, versioned datasets for exploration and production. In-place SQL enables fast analysis across analytical and operational data without loading or restructuring. Federated access extends this approach across distributed systems, so you can query broadly without waiting on new pipelines.

Governance can't come later. Lineage, data quality signals, retention and immutability must be enforced at the foundation so analytics and AI scale without added risk. When access, governance and life-cycle controls work together, you move faster without sacrificing trust, cost control or operational stability.

A Unified, Governed Foundation for AI and Analytics

Every AI and analytics initiative depends on a data foundation you can trust to be available and governed as new use cases emerge. With Hitachi Data Lakehouse, that foundation is built on Hitachi Virtual Storage Platform One (VSP One), which brings block, file and object storage together under a single architecture, so your data operates as one coordinated system.

Within that unified foundation, each capability is designed around how your business runs. Object storage forms the analytical backbone, holding Apache Iceberg® Tables, Amazon S3® Tables, curated datasets and the historical depth your analytics and models depend on. File services support day-to-day engineering work, including pipelines, notebooks, logs, features and collaborative development. Block storage delivers predictable, low-latency performance for transactional systems, metadata services and compute-intensive engines that feed analytics and AI.

On top of the storage foundation, governance and life-cycle intelligence run continuously. Pentaho Data Catalog gives you visibility into what data exists, how it is classified, where it came from and whether it is fit for use — surfacing quality and lineage issues before they propagate into reports, features or training datasets. Pentaho Data Optimizer applies policy-driven lifecycle decisions, helping control growth, reduce waste and align storage placement with value.

Analytics completes the system. Hitachi Advanced Database runs high-performance SQL directly on object and operational data, while Zetaris provides federated "query-in-place" access, empowering you to analyze, integrate and manage data across distributed sources without moving or duplicating it. You ingest once, govern with clarity and move faster with confidence as AI and analytics scale.

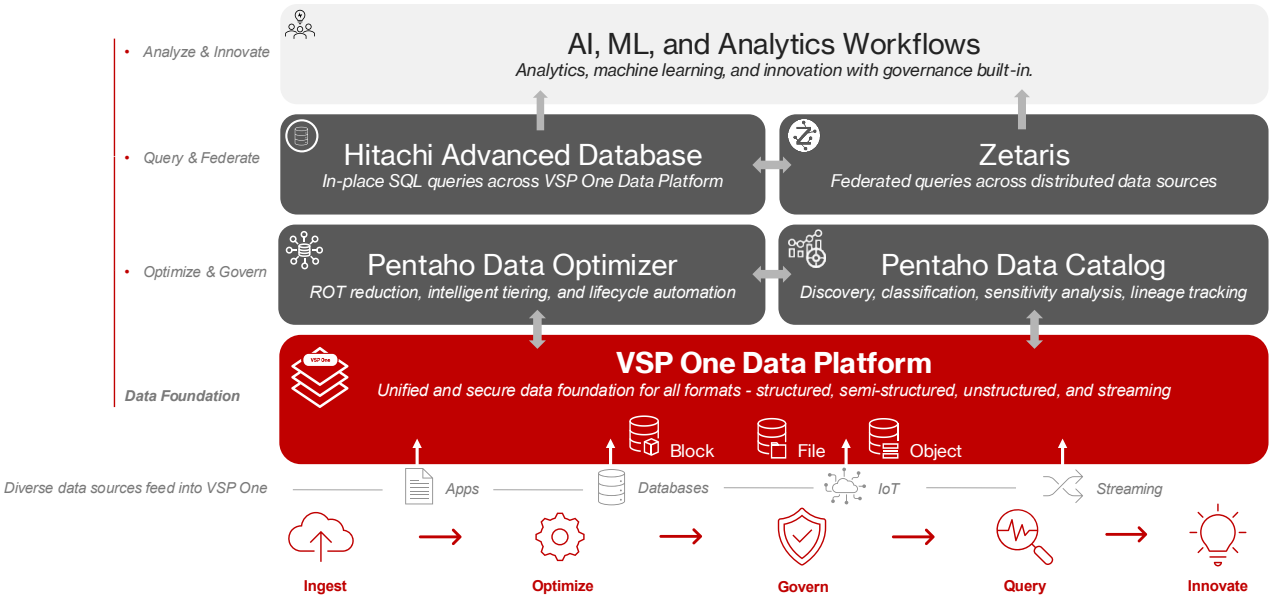

Fig. 1 Hitachi Data Lakehouse integrates storage, lifecycle management and analytics services into a unified, governed data foundation.

Features

Unified Storage Across All Data Types

Work with block, file and object data on one platform, supporting structured, semi-structured, unstructured and streaming datasets with retention, immutability and replication controls.

Reduce Platform Sprawl and Simplify Daily Operations

Open Tables for Analytics and AI

Create and analyze Iceberg and S3 Tables directly on object storage, maintaining versioned datasets that support reproducibility and audit needs.

Trust Versioned Data for Models and Reporting

Life-Cycle Automation and Cost Control

Identify redundant, obsolete and low-value data and apply policy-driven tiering and archival, keeping growth predictable and storage aligned to value.

Control Growth and Protect Analytics Budgets

Built-In Governance and Data Quality

Automatically discover, classify and track lineage while surfacing freshness, consistency and completeness issues early, before analytics or models rely on flawed data.

Improve Trust in Analytics and AI Outputs

In-Place SQL and Federated Analytics

Run SQL directly on object and operational data and query distributed sources in place, avoiding brittle ETL pipelines and unnecessary data copies.

Answer Questions Faster Without Rebuilding Pipelines

Unified Workspaces for Data Teams

Support notebooks, logs, features and intermediate outputs on shared storage, so engineering and analytics workflows stay aligned and productive.

Keep Teams Moving Without Workflow Friction

> *"Enterprise data platforms are rapidly becoming AI native and lakehouse centric, serving both human decision-makers and autonomous agents with equal rigor and transparency. Vendors that tightly integrate performance, governance, and AI-driven automation across personas will shape the next stage of data warehouse and lakehouse innovation."*
>
> **Marlanna Bozicevich**, Research Director – Data Analytics & AI, IDC

Solution Profile

Hitachi Data Lakehouse: Powering Trusted AI and Analytics at Scale

Where Hitachi Data Lakehouse Delivers Value

Hitachi Data Lakehouse drives the AI, analytics, and data management initiatives your teams run every day:

- **Trusted AI Training**
  Build fraud, forecasting and vision models on curated, versioned datasets with lineage and reproducibility.
- **Real-Time Analytics**
  Run dashboards, anomaly detection and event-driven analytics on live and historical data in place.
- **Governance and Retention**
  Enforce lineage, sensitivity controls, immutability and audit-ready retention across datasets.
- **Cost-Efficient Data Management**
  Automate life-cycle placement for archives, telemetry, logs and sensor data.
- **Simplified Data Engineering**
  Reduce pipeline sprawl by federating sources and blending multiformat data.
- **Operational Performance**
  Support databases and analytics with predictable, low-latency services.

Why Hitachi Leads in On-Prem Lakehouse Architecture

Choosing a lakehouse means choosing who you trust to run the data behind analytics, AI and the business. Hitachi brings decades of experience engineering platforms that support availability, performance and governance. Hitachi Data Lakehouse is designed as one integrated system, so storage, governance and query access work together from day one, without gaps or operational guesswork.

With Hitachi Data Lakehouse, you get:

- ***The first on-prem data lakehouse*** with native S3 Tables and Apache Iceberg.
- ***One unified data foundation*** across block, file, object and open tables.
- ***Built-in governance and life-cycle intelligence*** that protect analytics and AI data.
- ***In-place SQL and federated access*** across operational and analytical data.
- ***AI- and analytics-ready integration*** validated to work as one system.
- ***Guaranteed enterprise resilience and efficiency,*** with 100% Availability and 4:1 Effective Capacity backed by enterprise block architecture for always-on, cost-efficient lakehouse data.

Explore how to build a unified, governed data foundation for AI and analytics with Hitachi Data Lakehouse.

About Hitachi Vantara

Hitachi Vantara is transforming the way data fuels innovation. A wholly owned subsidiary of Hitachi, Ltd., we're the data foundation the world's leading innovators rely on. Through data storage, infrastructure systems, cloud management and digital expertise, we build the foundation for sustainable business growth.

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact

HV-BTD-SP-Hitachi-Data-Lakehouse-9Feb26-A