How Denodo Enables an Enterprise Data Fabric for the Age of AI

A Technical Whitepaper
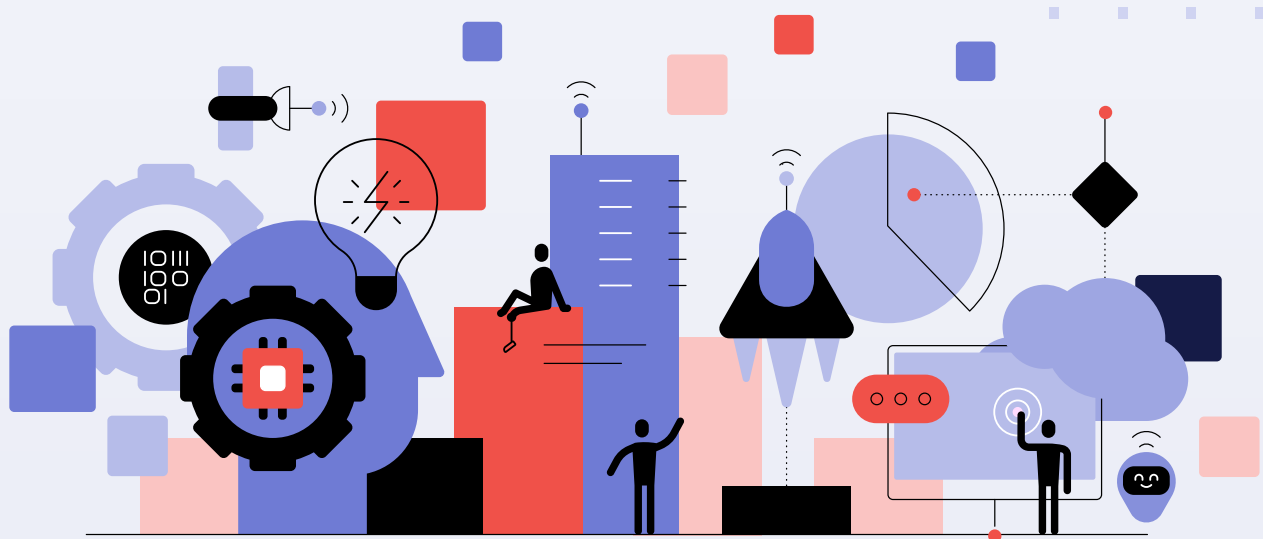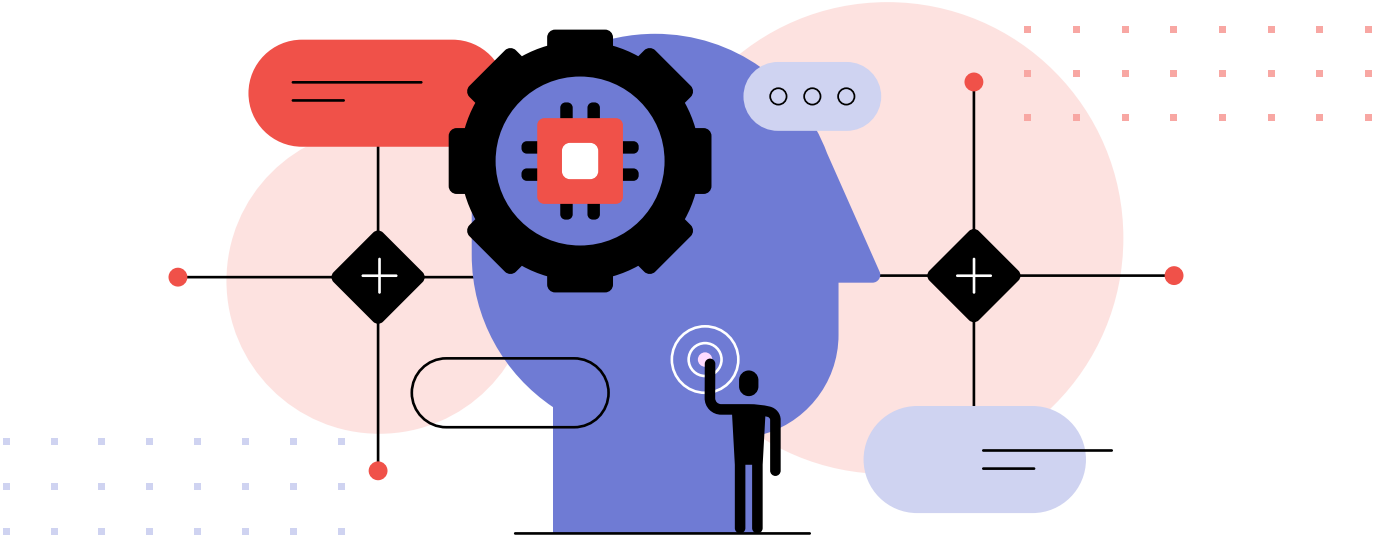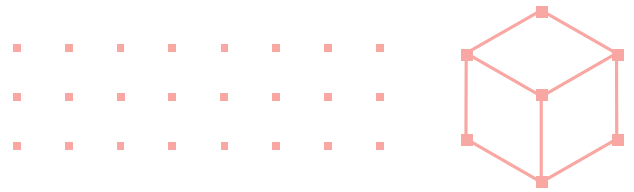
Table of Contents

Abstract

According to Gartner, "A data fabric is a data management design concept for attaining flexible, reusable and augmented data integration pipelines and services in support of various operational and analytics use cases delivered across multiple deployment and orchestration platforms. Data fabrics support a combination of different data integration styles and utilize active metadata, knowledge graphs, semantics and machine learning to augment data integration design and delivery."[1] Based on this definition and Gartner's extended research as published in "What CIOs Need to Know About Data Fabric," we believe that data fabric is not a standalone product but a composable framework, and that it is one that unifies disparate sources, enforces governance, and delivers real-time, trusted data for analytics, operations, and AI applications. This whitepaper explores how Denodo enables an enterprise data fabric: a metadata-driven platform that federates data without requiring replication, enabling seamless integration in hybrid and multi-cloud ecosystems. We also discuss how data fabric is an ideal architectural framework for enterprise AI.

Drawing on the latest advancements in the Denodo Platform, we delve into the technical architecture through its four foundational pillars: semantic unification, personalized self-service, federated governance, and real-time data delivery with cost control. These pillars incorporate core components such as a data marketplace, active metadata, AI-based recommendations, a semantic layer, orchestration via data operations (DataOps), and multi-cloud provisioning. Also, through support for frameworks like retrieval-augmented generation (RAG) and tools like the Denodo AI SDK, the Denodo Platform enables semantically rich, governed data for large language models (LLMs) and AI applications.[2] Real-world customer implementations illustrate these capabilities, while ROI analysis from recent studies highlights up to 345% ROI over three years, driven by reduced integration times and optimized cloud costs.[3] Ultimately, a Denodo-enabled enterprise data fabric transforms data fragmentation into an intelligent, AI-ready foundation, accelerating enterprise innovation.
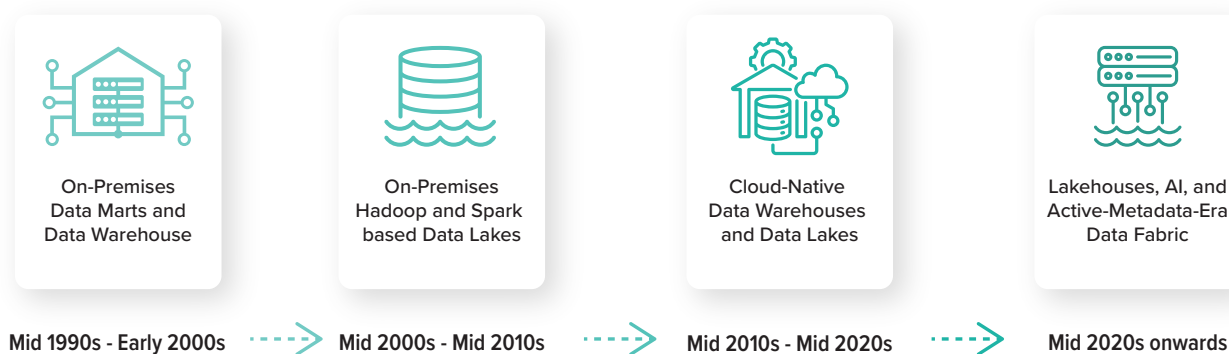
Introduction

In today's enterprise landscape, the proliferation of AI-driven initiatives, cloud-native applications, and IoT devices has elevated data to become a strategic asset, yet also a profound operational challenge. Data now spans legacy on-premises systems, expansive cloud data lakes, software-as-a-service (SaaS) ecosystems, and edge computing environments, each governed by distinct formats, protocols, and access paradigms. While modern data lakehouse architectures, such as those offered by Snowflake or Databricks, have advanced storage and processing efficiencies, they often fall short in providing the semantic unification, real-time federation, and enterprise-wide governance essential for delivering contextual, trustworthy data to analytics platforms and AI systems.

The traditional era of the enterprise data warehouse (EDW) has yielded to a more dynamic paradigm, fueled by big data paradigms, machine learning (ML), and the rise of self-service business intelligence. Citizen developers and data scientists now demand agile, on-demand access to diverse datasets, even as IT teams grapple with escalating system complexity under constrained budgets. Cloud and SaaS adoptions have mitigated some silos but have also amplified data distribution, creating hybrid ecosystems rife with fragmentation. This evolution imposes multifaceted demands: advanced analytics, ML, and AI require increasingly intricate data pipelines, while specialized data stores and formats from EDWs and data marts to relational databases, NoSQL stores, REST APIs, and real-time social feeds undermine the pursuit of a single version of the truth. Meanwhile, a broadening array of stakeholders, including BI analysts, data stewards, security professionals, and domain experts, each bring unique skillsets and access needs.

Compounding these pressures are regulatory imperatives such as the General Data Protection Regulation (GDPR) in the European Union; the California Consumer Privacy Act (CCPA) in the United States; the Personal Data Protection Act (PDPA) across jurisdictions such as Singapore, Thailand, and Indonesia; China's Personal Information Protection Law (PIPL) and Data Security Law (DSL); and India's Digital Personal Data Protection Act. Others include environmental, social and governance (ESG) regulations, anti-money laundering initiatives, and various industry regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the International Financial Reporting Standards. These, alongside evolving cyber and privacy threats, necessitate heightened compliance and governance standards.

Multi-cloud transitions further fragment physical data assets, demanding architectures that balance flexibility with uninterrupted business continuity. Securing and orchestrating such ecosystems manually proves error-prone and inefficient, underscoring the need for an intelligent intermediary.

Enter the enterprise data fabric, an architectural blueprint that addresses these imperatives through active metadata, semantic modeling, and automated orchestration. It establishes a common access stratum that conceals deployment intricacies, furnishing a cohesive logical layer for consumption. This architecture accommodates diverse integration modalities: including data virtualization for real-time federation and governed access, extract-transform-load (ETL) for batch processing, and streaming for event-driven scenarios, tailored to analytical and operational contexts alike. At Denodo, we believe that by infusing semantics into data elements and their interrelations, it renders them more intuitive for manipulation and analysis. Broad-spectrum governance, documentation, and security mechanisms instill confidence, while automation, powered by active metadata and AI, streamlines development, operations, and utilization. Crucially, this architecture provides **the trusted, connected, data foundation enterprises need to scale AI responsibly and effectively**.

On-Premises Data Marts and Data Warehouse	On-Premises Hadoop and Spark based Data Lakes	Cloud-Native Data Warehouses and Data Lakes	Lakehouses, AI, and Active-Metadata-Era Data Fabric
Mid 1990s - Early 2000s	**Mid 2000s - Mid 2010s**	**Mid 2010s - Mid 2020s**	**Mid 2020s onwards**

The Denodo Platform enables an enterprise data fabric that realizes this vision as a logical, composable layer that federates data in real time. Anchored in semantic unification, it complements existing lakehouses by bridging gaps in federation, metadata orchestration, and delivery flexibility. Without mandating data replication or vendor lock-in, it empowers organizations to cultivate AI-ready data products, enabling downstream AI models to operate on consistent, governed, and high-quality data, fostering accelerated insights and reducing total cost of ownership (TCO). This whitepaper elucidates the technical underpinnings through Denodo's four pillars (semantic unification, personalized self-service, federated governance, and real-time data delivery with cost control), as well as the AI synergies, multi-cloud enablers, and proven deployments that position Denodo as the pragmatic foundation for an AI-augmented data ecosystem.

## Complementing Lakehouses to Build an Enterprise Data Fabric

While modern lakehouse platforms, such as Delta Lake, Databricks, and Snowflake's offerings, excel at unifying storage and processing for analytical workloads, they represent only one component of a comprehensive enterprise data fabric. A true enterprise data fabric demands an architecture of composable layers: active metadata for orchestration, knowledge graphs for semantics, integration engines for federation, governance for trust, and flexible consumption interfaces for delivery. Treating a lakehouse as the entire fabric risks leaving critical gaps unaddressed, hindering scalability, real-time access, and AI readiness.

Lakehouses excel at large-scale batch ETL/ELT processing and compute optimization but often lack a global semantic layer that can virtualize and unify access across heterogeneous systems without physical data movement.[3] Metadata and governance are typically fragmented across lakehouses, data warehouses, SaaS applications, and external catalogs, resulting in inconsistent discovery, trust gaps, and brittle access controls. Delivery patterns remain focused on analytics-centric batch processing, with limited native capabilities for low-latency operational APIs, real-time streaming joins, or federated queries across legacy systems. Integrating operational or transactional systems usually requires additional data pipelines, increasing both cost and latency.

To enable trusted AI and context-aware analytics, lakehouses must be augmented with a semantic modeling layer that captures business entities, relationships, and glossaries through active metadata, enabling consistent meaning and reducing hallucinations in AI applications.[3]

Denodo addresses these gaps as the fabric's semantic and orchestration layer, interoperating natively with lakehouses to deliver enterprise-wide unification.[3] Its federated semantic layer abstracts sources into business-aligned views, preserving meaning without data movement or duplication. Active metadata bridges catalogs and policies, enabling consistent lineage and runtime governance across the estate. Flexible delivery, spanning data virtualization, pushdown execution, streaming, and on-demand caching, adapts to SLAs, from exploratory analytics to production APIs. A financial operations (FinOps) dashboard optimizes costs by tracking resource consumption, enabling administrators to monitor and control cloud costs.

This composable approach accelerates modernization at low risk: organizations leverage lakehouse strengths for heavy lifting while Denodo orchestrates the fabric, yielding reductions in compute/storage spend and faster AI project delivery.[3] By evolving incrementally and layering federation atop existing investments, enterprises avoid rip-and-replace pitfalls, building an intelligent architecture that scales with use cases, including high impact AI workloads.
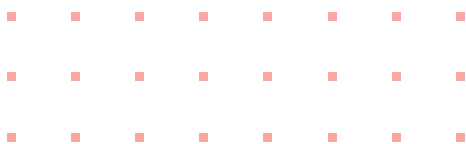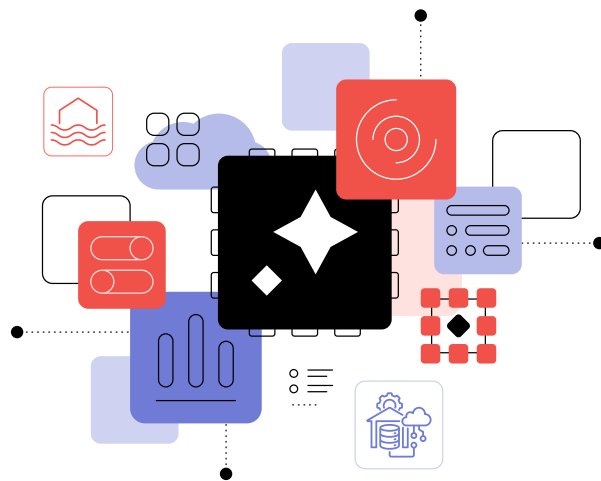
The Four Pillars of Denodo's Support for Enterprise Data Fabric

In supporting enterprise data fabric, Denodo is architected around four interdependent pillars, as discussed above: semantic unification, personalized self-service, federated governance, and real-time data delivery with cost control. These pillars integrate core technical components including data virtualization, execution engines, augmented catalogs, active metadata, AI recommendations, semantic layers, and DataOps, to form a resilient, self-optimizing system tailored for distributed, AI-driven environments. In this section, we'll take a close look at each pillar in turn.

PILLAR 1: SEMANTIC UNIFICATION OF ALL DATA

At the heart of Denodo lies its semantic layer: a metadata-driven abstraction layer that unifies data from heterogeneous sources into a consistent, governed business model. This pillar bridges technical schemas and business vocabularies, enabling uniform interpretation across departments and eliminating conflicting definitions or redundant pipelines. Supporting relational, NoSQL, cloud, and streaming data, it presents a single logical view of enterprise assets, defining reusable business entities such as "Customer," "Product," or "Transaction" once, for ubiquitous application.

Technically, the semantic layer uses extended metadata, including classifications, glossaries, and lineage, for entity resolution and context preservation. Active metadata management complements this by capturing runtime signals like query patterns and performance statistics, stored in a central repository for ML-driven optimization. The Denodo Assistant leverages this intelligence for AI-based recommendations, enabling natural language querying (NLQ) and dataset discovery that evolves with usage.

This unification transforms fragmented data, enabling a self-learning, enterprise data fabric, reducing wrangling efforts and accelerating AI readiness. As organizations scale, the layer's inheritable policies and automated schema-mapping provide consistency without vendor lock-in, fostering reusable data products for analytics and operations.

CASE STUDY: PIERRE FABRE LABORATORIES

Pierre Fabre Laboratories, a French multinational pharmaceutical and cosmetics company, relied on a legacy on-premises data warehouse since 1998 to manage large data volumes. The company sought to modernize for a strategic shift from pharmacy sales to **direct-to-customer engagement**, but its systems — Oracle database, Cognos, Business Objects, Tibco Spotfire, and Tableau (with a 2015 MapR data lake) — created silos and limited reuse, and it lacked central governance.[4]

The Denodo Platform powered the migration of on-premises systems to Microsoft Azure and served as the foundation for a combined data fabric/data mesh. It created a logical abstraction layer delivering user-friendly, governed, semantically unified views for seamless reuse across business analysts and developers.

> *We moved from a siloed, on-premises approach to an approach that is better governed, better architected, and based on a data fabric that supports our entire data lifecycle.*
>
> — **Wassim Bouaziz**, IT Director of Integration, Pierre Fabre Laboratories

A new **data office** and IT department collaborated to build the cloud platform, establishing departmental data governance organizations.

WITH DENODO, PIERRE FABRE LABORATORIES WAS ABLE TO:

- Leverage data-marketplace functionality to accelerate access to data, reducing the workload of business users and enabling decisions to be made more quickly

- Easily identify trusted data, to offer users a unified view of authoritative information

- Better understand customer behavior and the impact of advertising and brand awareness campaigns

- Support the seamless sharing of data between business domains, data reusability, and data governance

- Reduce Snowflake costs by 10%

- Rely less on extract, transform, and load (ETL) processes, reducing time-to-data by 75%

- Develop new data products quickly and at a lower cost

PILLAR 2: PERSONALIZED DATA SELF-SERVICE

Modern data consumers demand intuitive, on-demand access, and Denodo enables this for a data fabric through an augmented data marketplace that curates governed data products for diverse personas. This pillar empowers self-service discovery, previewing, and querying, enriched with business metadata, usage analytics, and AI-driven suggestions, to increase relevance and shorten the cycle time for AI experimentation and insight generation.

The data marketplace serves as the cornerstone, integrating with external catalogs like Collibra for hybrid visibility and supporting collaborative glossaries. Users search via semantic affinities, preview samples, and access views with role-based controls, while the Denodo Assistant provides NLQ and transformation recommendations based on historical patterns.

This personalization democratizes data without compromising governance, shifting the focus from toil to insights and enabling citizen integrators to innovate at scale.

CASE STUDY: SHIMIZU CORPORATION

Shimizu Corporation, founded in 1804, is a leading construction company in Japan renowned for building the tallest structures. The construction industry in Japan faces a declining workforce due to an aging population and a lack of interest from younger generations, highlighting the need for more efficient operational strategies. Rising costs of construction materials due to unstable international conditions increased project expenses. Regulations in April 2024 mandated stricter control over overtime, prompting Shimizu to prepare better compliance strategies.

On the technology side, the company faced fragmentation due to siloed organizational structures (business, design, quotation, construction, compliance across the architecture, public works, and engineering), which hindered efficient data sharing and collaboration. Also, managing data quality, freshness, standardization, and governance was challenging.
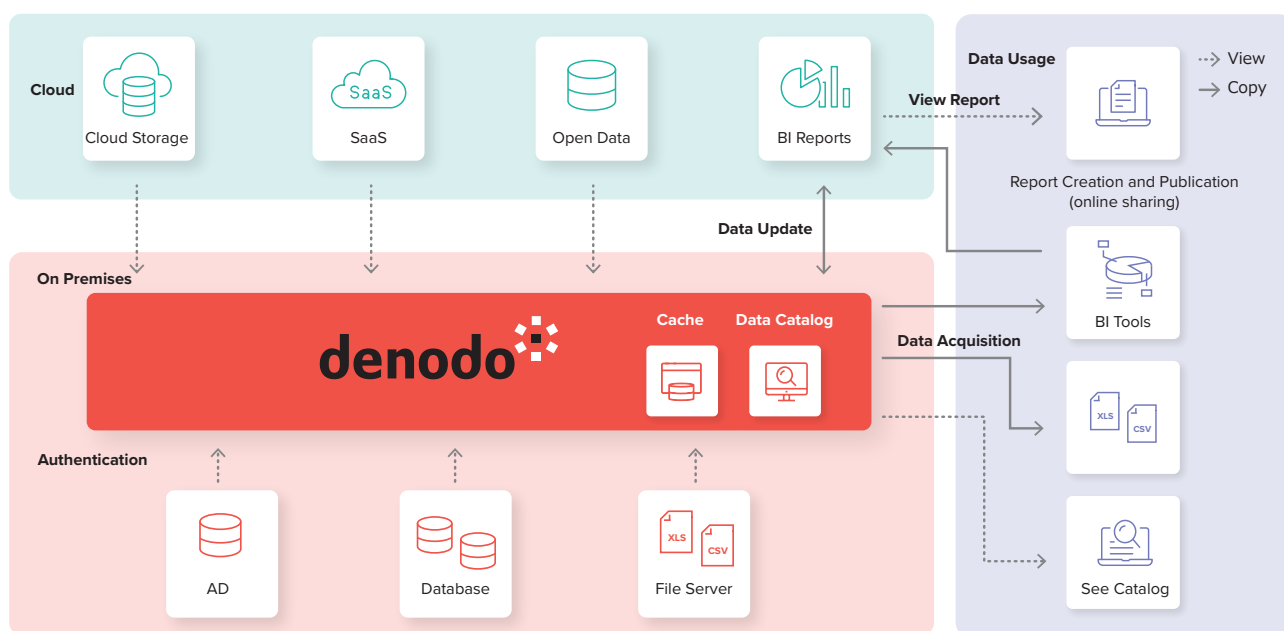
Shimizu corporation enables data sharing and self-service.

To address these challenges, Shimizu Corporation adopted the Denodo Platform, which enabled the company to centralize data accessibility and integrate data seamlessly without physical movement. The implementation of Denodo addressed several critical needs:

Enhanced Data Accessibility and Governance: Data became easily accessible both inside and outside the organization with governed access, allowing for quicker and more efficient utilization.

Improved Data Sharing and Utilization Mindset: The platform enabled user-level data sharing and quick discovery of data via the Denodo Data Catalog, reducing silos and promoting utilization.

Reduced Lead Time and Cost Savings: Reports that previously took 1-2 weeks to deliver now can be delivered in less than a day.

Additionally, the in-house development of reports resulted in substantial cost savings that would have otherwise gone to vendor services.

Low-Code Development: Facilitated low-code application development, enabling departments to create their own applications, reducing the burden on the Digital Strategy Promotion Office.

PILLAR 3: FEDERATED GOVERNANCE WITH CENTRALIZED OVERSIGHT

In complex data landscapes, governance must balance decentralization with control, and Denodo achieves this through a federated model that centrally manages policies, lineage, and access rules yet enforces them dynamically at runtime. The platform evaluates and applies governance rules while queries are executed, without replicating data, so every domain consumes consistent, compliant, and up-to-date policies regardless of where the underlying data resides. This pillar integrates security into the fabric's core, managing access and use of trusted data for various use cases in the organization, supporting the creation of data products and audit trails across environments.

Runtime enforcement applies fine-grained controls — row-level filtering, column masking, and role-based access — via inheritable metadata, while active metadata propagates changes automatically. This is increasingly essential, as organizations operationalize AI and must keep sensitive or regulated data — used in model training and inference — in compliance. DataOps pipelines, managed by Denodo Solution Manager (a module within the Denodo Platform), automate testing and promotions through dev-test-prod, integrating with DevOps tools for compliance. As data proliferates, these capabilities facilitate regulatory compliance, while maintaining oversight, while also satisfying emerging AI governance expectations such as transparency, lineage, and accountable data usage.

Federated governance thus sustains agility and trust, mitigating risks in hybrid deployments.

CASE STUDY: ONE OF THE WORLD'S LARGEST MINING COMPANIES

One of the world's largest mining companies, which operates across Australia, the Americas, and Asia, managed hundreds of data systems across on-premises and cloud platforms. Integrating this distributed data for analytics and reporting was slow and complex, and it raised data sovereignty and compliance challenges — especially as regional laws restricted the transfer of operational data across borders.

> *We were going to create a data mart. Instead, with Denodo, we had a much quicker, more agile approach.*
>
> — **Data manager**, at the mining company

The company deployed the **Denodo Platform** as a **logical data management layer**, with clusters in **Australia, the United States, and Chile**. This enabled users to access governed, integrated views of global data **without physically moving raw datasets**. Only aggregated or filtered query results travel securely between regions, preserving compliance with data localization requirements.

The organization gained a governed, agile data access layer that reduced replication, improved visibility across global operations, and facilitated compliance with regional data regulations.

PILLAR 4: REAL-TIME UNIFIED DATA DELIVERY WITH COST CONTROL

Denodo's runtime and query optimization framework orchestrate multiple delivery styles, including virtualization, intelligent query pushdown, selective materialization (caching and summaries), and streaming. Queries are executed at the most efficient layer available. The platform's cost-based optimizer uses statistics from base views to estimate execution costs and choose plans that apply pushdown, reorder joins, select join algorithms, or precomputed summaries for acceleration.

To accelerate analytics at scale, Denodo includes an embedded Presto-based MPP engine that runs on Kubernetes and supports distributed query execution against data lakes. This engine is designed for big-data scenarios in which in-platform execution reduces latency for analytic workloads.

The platform provides broad, out-of-the-box connectivity with over 200 data sources supporting JDBC, ODBC, cloud stores, message systems, and SaaS endpoints. Streaming sources such as Kafka and CDC pipelines can be integrated into the same semantic model used for batch queries and APIs.

Operational deployments can be containerized and orchestrated with Kubernetes. Denodo provides tooling for running its components and the embedded MPP engine in cloud environments, enabling administrators to scale clusters and automate lifecycle operations.

For cost visibility and FinOps control, Denodo provides detailed usage and access metrics that help teams understand data consumption patterns. By minimizing data movement and enabling localized virtualization across clouds, organizations can reduce avoidable egress and storage costs. FinOps dashboards provide transparency to monitor consumption and optimize materialization decisions.

Together, these capabilities enable organizations to deliver governed data services for both analytics and operational APIs without requiring replication. The result is an enterprise data fabric that supports hybrid and multi-cloud architectures, simplifies real-time data patterns, and provides operational control over performance and cloud spend, forming the high-quality, real-time data backbone essential for production-grade AI systems.

CASE STUDY: DNB BANK

DNB Bank, Norway's largest financial services group and one of the largest in the Nordic region, offers a full range of services including loans, savings, insurance, and pension products for retail and corporate customers. The bank managed a highly complex data landscape with more than 40 data sources across multiple on-premises warehouses (Oracle, Teradata) and an AWS-based data lake containing over nine billion transactional records.

To increase agility and reduce the time data scientists spent on data preparation, DNB sought a solution that could unify access to distributed data while maintaining strict compliance with regulations such as the GDPR.

DNB deployed the **Denodo Platform** to enable an enterprise data fabric, creating a unified semantic layer that connected disparate systems across the enterprise. The platform became the foundation of DNB's Insight Platform for Analytics — a centralized environment in which data and code coexist to drive agile, compliant innovation.

Through Denodo, data scientists gained governed, self-service access to curated datasets directly from tools such as **JupyterLab** and **RStudio**, enabling faster model development and experimentation. Denodo's runtime optimizer enabled efficient query execution across on-premises and cloud systems, while centralized lineage, auditing, and fine-grained access controls enforced consistent governance and GDPR compliance.

KEY OUTCOMES:

- Unified more than 40 heterogeneous data sources into a governed semantic layer

- Enabled over 200 active data scientists and analysts to build and deploy models on demand

- Powered APIs serving personalized pricing, churn prediction, and next-best-offer models

- Supported compliance use cases including anti-fraud, cybercrime, and anti–money laundering (AML)

- Streamlined "Right of Access" reporting for GDPR through Denodo's federated governance capabilities

- Accelerated analytics delivery while reducing operational overhead through real-time access to curated data
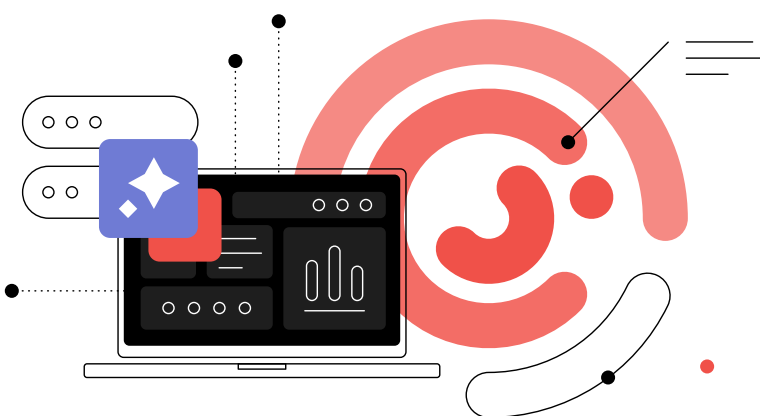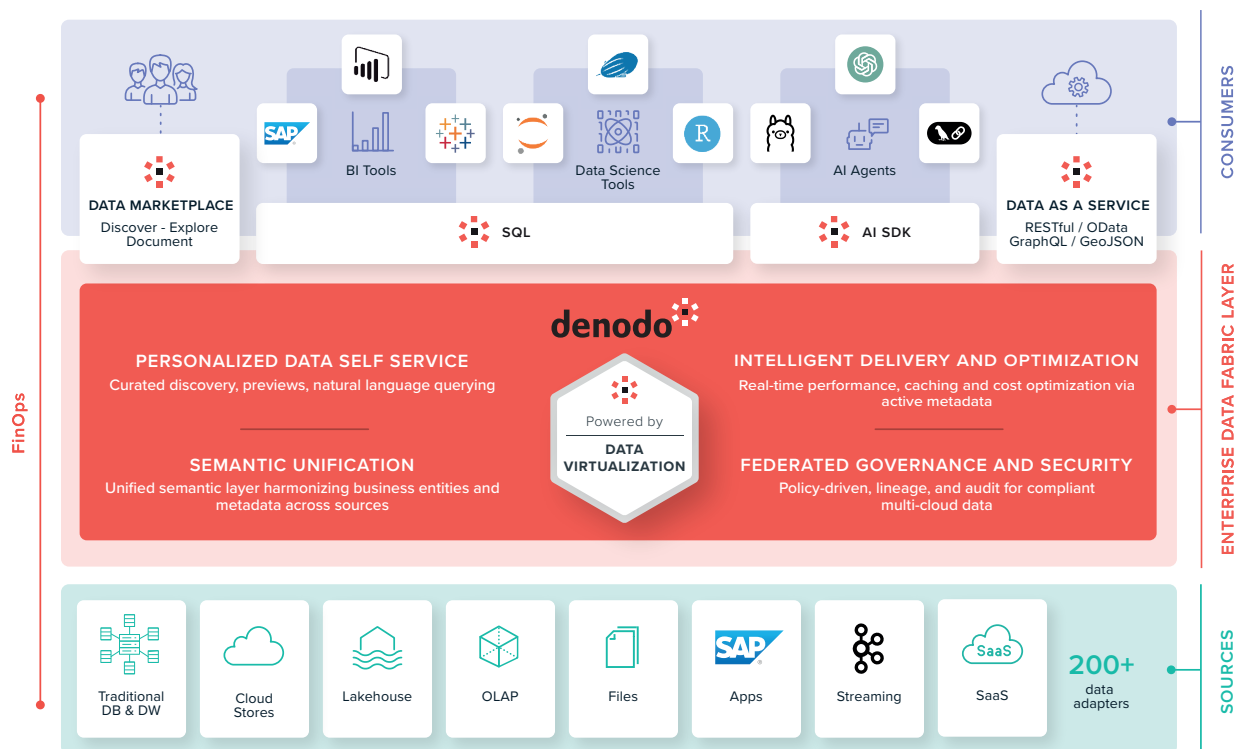
Denodo enables an enterprise data fabric.

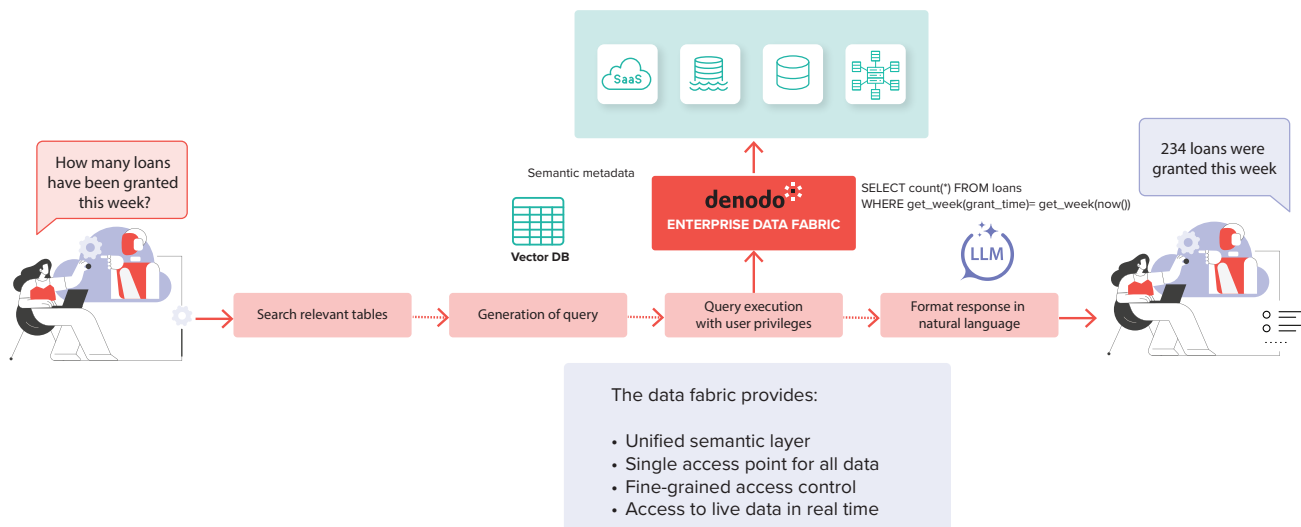# Denodo-Enabled Enterprise Data Fabric and AI

In the age of generative AI, the effectiveness of a data fabric depends on its ability to deliver semantically enriched, governed, and explainable data to LLMs and intelligent agents. Denodo-enabled enterprise data fabric can accomplish this purpose. Its semantic layer acts as an intelligent mediation layer that abstracts distributed, heterogeneous sources into business-aligned views enriched with contextual metadata. This lets AI systems operate on consistent, trusted, and traceable information.

A key enabler of this capability is RAG. In a Denodo-powered architecture, unified semantic views and active metadata supply LLMs with governed, verifiable context, through a RAG framework. Using the Denodo AI SDK, descriptive and structural metadata, such as business terms, glossaries, and lineage, can be vectorized and integrated with external vector databases. This supports embedding-based retrieval while maintaining live connectivity to federated data sources, eliminating ingestion overhead and keeping AI responses grounded in real-time enterprise data.[9]

NLQ extends this accessibility to business users. Through the Denodo Assistant, users can pose questions in plain language, which are translated into SQL using LLMs. The Assistant leverages the semantic model to interpret intent and resolve ambiguities; for example, mapping a phrase such as "customer churn risk" to the correct entities and metrics. Governance-aware suggestions guide users on data sensitivity and access scopes, while the data fabric enforces the actual policies, including role-based filtering, data masking, and auditing. This separation of query assistance and governance enforcement preserves both flexibility and compliance.

For advanced analytical reasoning, Denodo DeepQuery extends this paradigm by orchestrating multi-step analyses across distributed sources with full lineage and traceability. It enables complex reasoning chains for AI agents and data scientists, combining multiple queries or models into an explainable, auditable workflow. Governance is integral throughout the AI lifecycle. Runtime policies in the Denodo Platform enforce security, privacy, and differential access, enabling alignment with global regulations such as the EU AI Act, the U.S. Executive Order on Safe, Secure, and Trustworthy AI, and data protection laws including GDPR, CCPA, PDPA, and PIPL.

The [Denodo AI SDK](#), an open-source framework, simplifies the integration of the data fabric with generative AI ecosystems. It provides APIs for embedding, orchestration, and security, supporting MCP-compliant multi-agent systems for autonomous decisioning. The SDK integrates seamlessly with LangChain, LLM frameworks, and external vector databases, enabling developers to build Denodo Query RAG (a proprietary version of the RAG framework) endpoints that connect AI agents directly to governed, real-time data within the Denodo Platform.

Denodo-enabled data fabric facilitates powerful AI applications.

The Key Benefits of Denodo-Enabled Enterprise Data Fabric

Denodo-enabled enterprise data fabric yields profound advantages, accelerating data velocity while curtailing complexity and expenditure. By federating sources in real time, it slashes integration timelines by up to 75%, liberating engineering teams from pipeline maintenance to higher-value pursuits. Governance, once a siloed burden, becomes intrinsic. Dynamic policies propagate across views, facilitating compliance without impeding agility. For AI initiatives, semantic enrichment curtails model retraining cycles and hallucination risks, fostering trustworthy deployments.

Operational resilience amplifies through multi-cloud agnosticism, where location-optimized routing and automated provisioning minimize latency and vendor entanglements. Self-service discovery via the augmented marketplace empowers analysts and scientists alike, yielding 3-4x faster insights and collaborative innovation. Collectively, these attributes not only enhance decision-making but also fortify the data estate against evolving threats, positioning organizations for sustained competitive edge.[3]

Conclusion:

Denodo-enabled enterprise data fabric represents a foundational shift in enterprise data architecture: a composable, AI-native platform that federates distributed sources into a semantically unified, governed, and real-time continuum. Through its four technical pillars: **semantic unification, personalized self-service, federated governance, and real-time delivery with cost contro**l; it eliminates data silos, accelerates AI workloads, and optimizes multi-cloud spend.

An independent analysis by Veqtor8 confirms the business impact of the Denodo Platform: organizations achieve **345% ROI over three years, 3–4x faster time-to-insight**, and **75% reduction in integration effort**, all while maintaining strict compliance and lineage traceability.[3]

As generative AI becomes mission-critical, Denodo equips enterprise architects with a proven, low-risk path to AI readiness by delivering trusted, real-time data to LLMs and agentic systems without replication or disruption. **By embedding governance, semantics, and real-time access into a unified fabric, Denodo enables AI initiatives to scale reliably and responsibly**.

The result is not just operational efficiency, but a strategic data foundation that transforms complexity into competitive advantage.

References

1. Gartner: What CIOs Need to Know About Data Fabric
2. Denodo: Semantic Layers for the AI Era
3. Veqtor8: The ROI of Using the Denodo Platform alongside the Modern Data Lakehouse
4. Denodo: Pierre Fabre Laboratories
5. Denodo: Transforming Construction with Data Fabric (Shimizu)
6. Denodo: DNB Bank
7. Denodo/Accenture: Multi-Cloud Data Integration Made Simple
8. Denodo: Enabling GenAI Success
9. Denodo AI SDK User Manual
10. From Chatbots to Agents: Navigating the Future of AI

denodo