

Denodo and Databricks: A Unified Strategy for Maximizing the Value of Your Data Ecosystem

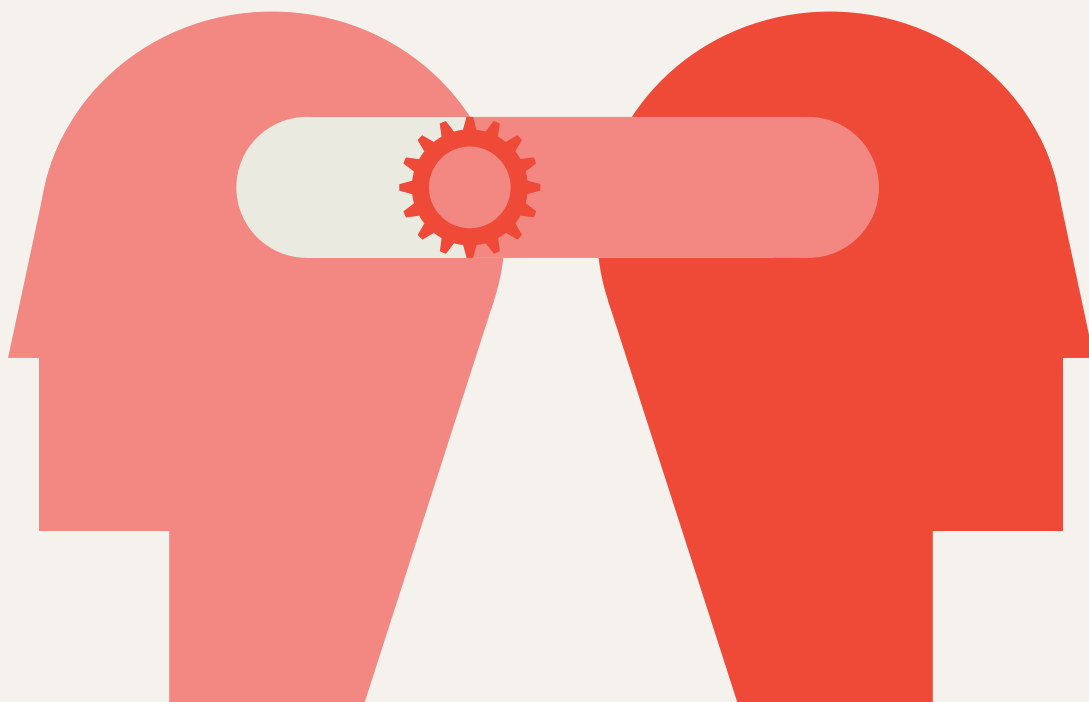
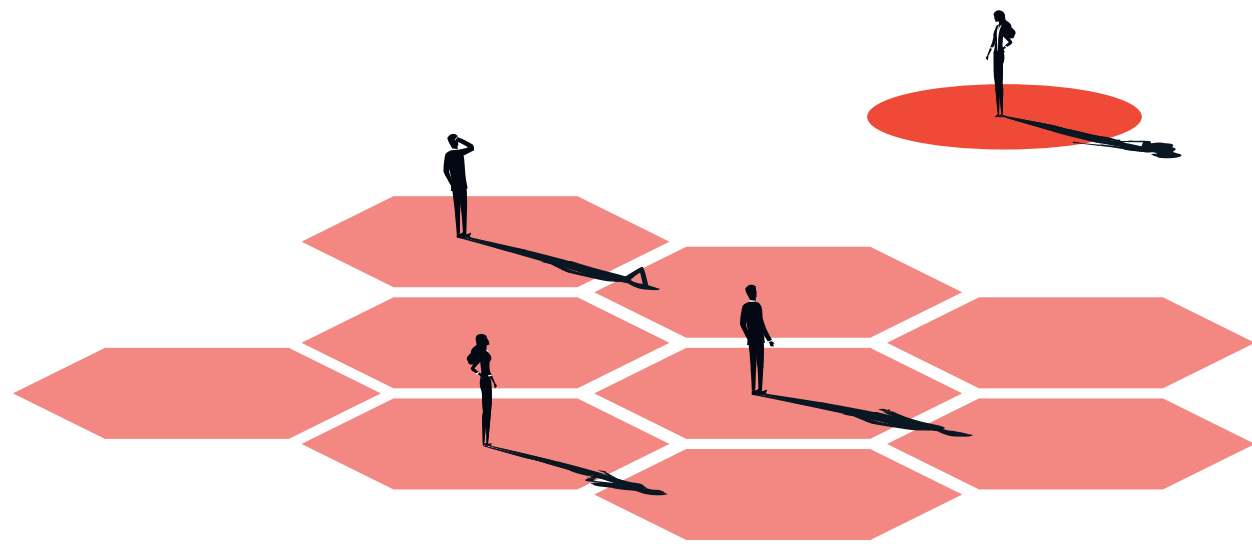


Table of Contents

Introduction	3
Enhancing the Lakehouse with Logical Data Delivery	4
A Stronger Foundation for AI and Analytics	5
Future-Ready Architecture, Today	6
Conclusion	6
Denodo and Databricks Reference Architecture	7





Introduction

Organizations today are navigating a rapidly evolving data landscape, where success depends on delivering timely, trusted insights to every corner of the business. Whether powering AI-driven innovation, enabling self-service analytics, or improving operational efficiency, modern enterprises need more than just storage and processing, they need agility, security, and intelligent data access.

Databricks has emerged as a leading platform for unified analytics, delivering a powerful lakehouse architecture that supports large-scale data engineering, machine learning, and advanced analytics. Many organizations have embraced Databricks to consolidate and simplify their analytical environments, and rightly so. It provides the scalability, openness, and performance required for modern workloads.

As these environments grow, the question becomes: how can we get even more value from our Databricks investment, faster, and across more of the business?

The Denodo Platform offers a compelling answer. By adding a logical data management layer on top of Databricks and other enterprise systems, Denodo helps organizations deliver curated, AI-ready data to every user and application, without compromising security, governance, or performance.

Together, Databricks and Denodo create a modern, enterprise-ready data ecosystem that brings the best of centralized analytics and decentralized data access.

Enhancing the Lakehouse with Logical Data Delivery

Databricks offers exceptional capabilities for centralized data storage, high-performance processing, and machine learning model development. However, as data architectures mature, many organizations find themselves needing broader capabilities to operationalize that data across distributed teams, diverse tools, and real-time business use cases.

This is where Denodo complements and extends the Databricks lakehouse. By introducing a logical data layer, Denodo delivers governed, business-friendly data from across the entire enterprise.

DELIVERING BUSINESS-READY DATA WITH SEMANTIC CONSISTENCY

While Databricks excels at managing data at scale, exposing that data in a form that business users and AI models can easily interpret often requires additional effort. Denodo addresses this challenge by creating a semantic layer that provides consistent business definitions, documentation, and lineage across all data. This makes it easier for users to find, understand, and trust the data they consume, without needing to navigate raw datasets or custom SQL logic.

With Denodo, the valuable data already managed in Databricks becomes even more accessible and impactful across the organization.

EMPOWERING SELF-SERVICE AND EXPANDING DATA ACCESS

Databricks provides powerful tools for technical users, data scientists, engineers, and developers who are building models, pipelines, and analytics on the lakehouse. However, to fully democratize data and extend access to business analysts and other non-technical users, additional abstraction and usability are often needed.

This is where Denodo's data marketplace experience delivers significant value. Designed with the data consumer in mind, it delivers a user-friendly, marketplace-like experience that simplifies data exploration and access. Users can easily search for and evaluate available data products, request access, review endorsements, report data quality issues, leverage AI-powered data preparation, and collaborate directly with data owners, all within a governed and intuitive interface.

To further streamline the experience, the built-in Denodo Assistant leverages generative AI and active metadata to simplify and accelerate common tasks. It supports natural language querying, recommends relevant datasets, and assists with query construction and data preparation. By reducing friction and automating routine steps, it boosts productivity for both business users and data engineers, while also promoting faster onboarding, clearer data understanding, and more consistent use of trusted data across the organization.

While Databricks Unity Catalog plays a critical role in managing and securing data within the lakehouse, Denodo complements it by extending those governance and discovery capabilities across all data, regardless of location, whether in the lakehouse, legacy systems, SaaS platforms, or on-premises databases. Importantly, Denodo integrates seamlessly with Unity Catalog and other enterprise catalogs, providing a unified experience for users and a consistent layer of access control and metadata across the full data landscape.

Together, Unity Catalog and the data consumer-focused marketplace experience provided by Denodo enable organizations to safely scale self-service access, ensuring that users of all skill levels can confidently find, understand, and use trusted data without adding complexity or compromising governance.



EXTENDING GOVERNANCE ACROSS THE ENTERPRISE

Databricks includes strong capabilities for managing security and access controls within its environment. However, many enterprises operate in hybrid or multi-cloud environments, where critical data resides outside the lakehouse, whether in on-prem systems, SaaS applications, or partner platforms.

Denodo enhances enterprise-wide governance by enabling centralized policy enforcement across all systems, regardless of where the data is stored. Fine-grained access controls can be defined once and applied consistently across the entire ecosystem, helping to maintain compliance and trust as data flows across domains and users.

SUPPORTING REAL-TIME OPERATIONAL AND AI USE CASES

While Databricks supports high-performance batch and streaming analytics, some operational use cases, such as real-time personalization, AI agent responses, or dashboard refreshes, require instantaneous access to live data in source systems. Rather than replicating all data into the lakehouse, Denodo allows selective, on-demand access to real-time data from across the enterprise.

This complements Databricks' architecture by enabling real-time responsiveness for applications and users, while controlling costs and reducing data duplication.

ACCELERATING VALUE AND OPTIMIZING PERFORMANCE

Denodo helps organizations get more value from Databricks, faster, by intelligently optimizing how data is accessed and delivered. Through advanced query acceleration techniques—including caching and aggregate-aware optimizations—Denodo improves performance while reducing the load on underlying systems. Its built-in FinOps dashboards provide detailed visibility into data usage and associated costs, helping teams make informed decisions about resource allocation and system efficiency.

Building on this foundation, the Denodo Assistant uses generative AI and active metadata to automate performance tuning tasks that traditionally require manual effort. By analyzing query patterns, it can proactively recommend summary caches and optimize execution strategies, ensuring that high-impact queries run efficiently—even as workloads and data volumes grow. This AI-driven automation simplifies operations for data engineers and enhances overall system responsiveness.

Additionally, to manage advanced acceleration scenarios with dependencies across tables, Denodo manages a Directed Acyclic Graph (DAG) to automatically detect and execute in the correct order the corresponding replication and ELT workflows used behind the scenes for those tasks.

A Stronger Foundation for AI and Analytics

Databricks offers a robust platform for building and scaling AI and analytics. Denodo helps operationalize those efforts by making the right data available, securely and in the right context, to every model, user, and application.

Whether supporting data scientists building predictive models, analysts creating executive dashboards, or developers powering intelligent applications, the combination of Databricks and Denodo ensures the data is:



Trusted and governed



Rich in business meaning



Delivered in real time



Available where it's needed, without replication

Future-Ready Architecture, Today

Both Databricks and Denodo embrace open formats, scalable architectures, and integration with modern data ecosystems. Importantly, Denodo introduces a level of flexibility that protects organizations as they evolve.

By decoupling data access from underlying storage or processing engines, Denodo allows organizations to change or expand back-end systems, whether adding cloud platforms, modernizing legacy systems, or reorganizing around data domains, without disrupting data consumers or workflows.

This architectural flexibility safeguards prior investments in Databricks while enabling continued innovation and adaptability.

Reference Architecture: How Denodo and Databricks Work Together

In a typical deployment, Databricks serves as the centralized analytical platform, managing data storage, preparation, and model execution. Denodo overlays this environment, creating a virtualized semantic layer that spans not only Databricks but also other enterprise systems, cloud databases, file systems, and applications.

This logical layer provides consistent data views, enforces access policies, and exposes data through standard query languages, APIs, self-service tools, and an AI SDK that simplifies the building of AI Agents. It integrates with governance platforms, supports notebook environments like Jupyter, and enables both real-time and batch access depending on use case.

The Denodo Platform's deep integration with Databricks leverages SparkSQL for seamless collaboration and uses its embedded Presto MPP engine to support open table formats such as Delta Lake, Iceberg, and Parquet. It is also compatible with Databricks Unity Catalog, extending fine-grained governance across the open data ecosystem. Together, these capabilities enable a unified and governed architecture that supports everything from high-performance analytics to real-time operational use cases. The following diagram illustrates how the two platforms work together to deliver a modern, scalable data environment.

Conclusion

Databricks is an exceptional platform for modern data processing, and many organizations are rightfully making it a cornerstone of their data strategy. As these strategies mature, the opportunity is not to replace or rework the lakehouse, but to amplify its value, to extend its reach, improve its accessibility, and accelerate its impact.

That's where Denodo comes in.

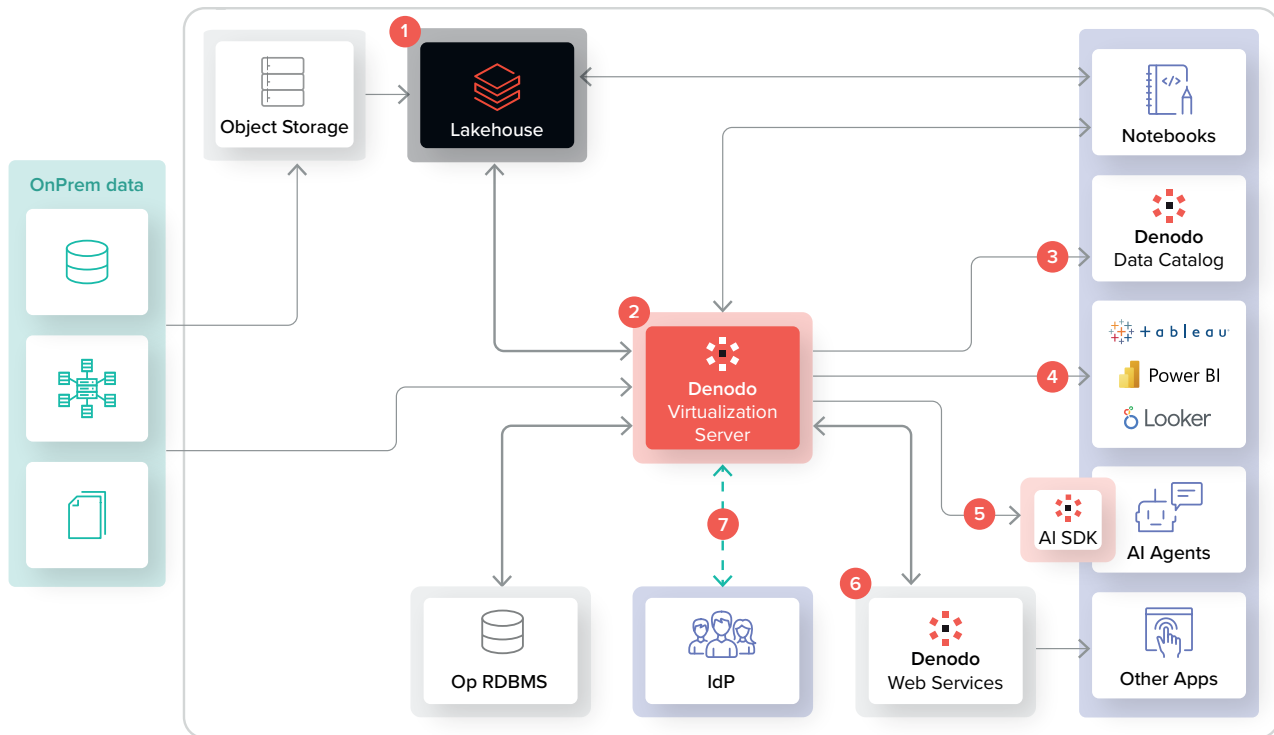
By complementing Databricks with Denodo's logical data management capabilities, organizations can create a data ecosystem that is not only scalable and performant, but also agile, governed, and business-ready. The combined solution empowers every user, technical or not, to engage with data confidently, securely, and in real time.

Together, Denodo and Databricks deliver the best of both worlds: a strong analytical foundation and a responsive, intelligent data delivery layer that unlocks the full potential of your enterprise data.



Denodo and Databricks Reference Architecture

IN AWS, MICROSOFT AZURE, AND GOOGLE CLOUD



CLOUD SERVICES

Object Storage	Op RDBMS	IdP
Amazon S3	Amazon Aurora	AWS SSO
ADLS	Azure SQL	Microsoft Entra ID
Google Cloud File System	Google Cloud Spanner	Google Identity Platform

- 1. Modern Analytics Foundation** – Databricks' Delta Lake innovation and investment in open data management frameworks have made the lakehouse model a dominant approach in modern analytics architectures.
- 2. Unified Data Access** – The Denodo Platform complements Databricks by adding a semantic layer that unifies access to all data, starting with the lakehouse and extending to operational systems, cloud platforms, on-premises databases, and hybrid or multi-cloud environments.
- 3. Data Discovery & Cataloging** – Denodo's Data Catalog offers a marketplace-style experience where data consumers can explore, discover, and access data with ease. It includes rich business semantics, clear data relationships, contextual metadata, view and field documentation, and data profiling, helping both end users and LLMs quickly find, understand, and trust relevant data.
- 4. Seamless Analytics Integration** – With Denodo's universal semantic layer, data consumers can access a single, centralized location for all their data, regardless of format or location. This eliminates the need to maintain semantics separately across multiple analytics tools.
- 5. GenAI & AI Agent Enablement** – The Denodo AI SDK simplifies GenAI workflows through REST API calls, abstracting LLM orchestration with tools like LangChain and vector database embedding. By removing complexity, it enables anyone familiar with REST to build AI Agents or Chatbots without requiring specialized AI expertise.
- 6. API-Based Data Access** – Applications that require API-based access to enterprise data can leverage Denodo's native support for REST, OData, and GraphQL, ensuring flexible and scalable data integration.
- 7. Governance & Security Enforcement** – Denodo supports a wide range of Identity Providers and offers fine-grained access control with ABAC (Attribute-Based Access Control) and RBAC (Role-Based Access Control) policies, integrating metadata from tools like Collibra to enforce sensitive data protections using PII tagging and other classifications.



Visit: databricks.com
Follow: [@databricks](https://twitter.com/databricks)



Visit: denodo.com
Follow: [@Denodo](https://twitter.com/Denodo)