

E-BOOK

Maximize the ROI of Your Databricks Lakehouse with Stardog

STARDOG





Introduction

With the ability to store a variety of data types and handle large data volumes, the Databricks lakehouse has become the modern enterprise data architecture.

Organizations choose the lakehouse to streamline data management, derive insights, and accelerate machine learning. A semantic layer that harmonizes the lakehouse with external data improves ROI, connects relevant elements, and provides greater context.

By connecting enterprise data and overlaying business semantics, Stardog's Enterprise Knowledge Graph platform facilitates agile data operations, reduces integration costs, increases lakehouse ROI, and helps generate powerful insights.

In this book, we'll discuss what you can accomplish by combining Databricks with Stardog, the leading Enterprise Knowledge Graph platform.

Chapter 1.

Semantically enrich your consolidated data for more robust analysis.

Innovative companies are helping organizations co-locate data from across the organization, and use that data to leverage the benefits of AI. The result is a reduced need to maintain expensive ETL pipelines against costly on-prem warehouses.

While access to all the data is a reality, it's not available to the users that need rapid insights to keep pace with change

Hooking up a BI tool to Databricks might seem to accomplish that. But it doesn't reduce latency, promote collaboration, enable self-service, or infer new insights.

The idea of a semantic layer is not new.

It has been around for over 30 years, often helping companies build purpose-built dashboards.

But, adoption has been slow due to the embedded nature of that layer in a rigid, complex BI system. Those systems have the same limitations as a physical system—they can't model data the way it's related in the real world.

A Semantic Layer operates between data storage and analytics, representing a logically-enriched view of information as interrelated business concepts, and results from using a knowledge graph.

Stardog Transforms Boehringer Ingelheim's Data Lake

The world's largest privately held pharmaceutical company uses the Stardog Enterprise Knowledge Graph platform as a semantic layer over its data lake, making information easier to navigate, query and analyze.

Enhanced discovery

Bioinformaticians use Stardog to navigate up to **90% of the company's R&D information**, with all relevant data logically connected. It allows users to search for specifics and explore how individual elements relate to one another.

Faster time to insights

The new linked data dictionary lets users fetch data **directly from the data lake** for analysis.

Simplifying complex data analysis

Increasing data connectedness enables Boehringer to **answer complex questions quickly**.

[Learn More](#)



A knowledge graph powers a semantic data layer that enables exploration of connections across your data universe with business context. Specifically, users can:

1. **Ask questions based on business concepts and the relationships between them.** Those concepts map to the underlying metadata that can facilitate rapid pipeline development for sharing data across applications.
2. **Run fast and flexible federated queries between Databricks and data in other sources** in support of ad hoc data analysis. Linking and querying data in and outside of Databricks enables just-in-time analytics for richer, faster insights without data sprawl.
3. **Save time and money spent on data wrangling and movement**, easily sharing your findings through visualization and enabling self-service analytics inside existing investments.

Chapter 2.

Enable federated querying and analytics on your lakehouse.

DevOps promises to increase the speed with which data can be provisioned to support production applications and ad hoc analysis.

Many organizations recognize the inefficiencies of data engineers readying data for analysis. Automating some of that sounds ideal, but the ability to bring data together is often not systematized.

An enterprise knowledge graph **logically connects enterprise data**, and maps key elements in a way that makes sense to the business.

Analytics applications that point to a semantic data layer powered by an enterprise knowledge graph provide better, faster services by reducing latency.

Stardog returns queries in milliseconds

that currently take tens of seconds when run directly against a data lake without a semantic layer.

Chapter 3.

Accelerate your machine-learning projects.

Machine learning (ML) strategy generally focuses on ensuring data is accessible, reusable, interpretable, *and* high quality, which is often a challenge.

A rule-based AI approach includes everything from making intelligent inferences about schemas to assembling techniques for text analytics or Natural Language Processing (NLP).

The best data science projects combine various sources of data, facilitated by solutions like Databricks. When combining data sources and datasets, ontologies and context help complete machine-learning projects. Platforms like Stardog assist tremendously.

Knowledge graphs make it easier to connect data sources and feed better data into ML algorithms. You can maximize internal content by laying the foundation for AI and semantic applications, ultimately showing relationships between data.

23% of organizations deploy graph techniques in their artificial intelligence (AI) projects.*

23%

77%

Large technology companies, public-sector organizations, financial solutions providers, and healthcare lead the way, using graph technologies to enhance data search, information retrieval, and recommendations.

How Knowledge Graphs and Machine Learning Work Together for Better Outcomes

Stardog leverages AI to solve common challenges. Stardog Voicebox makes knowledge graph adoption simple. See how.

[Read Our Blog](#)

Enable Highly Productive Data Workers

Instead of spending time cleansing data, a knowledge graph lets data scientists train models directly on unified data.

Stardog's Inference Engine helps resolve conflicting data definitions without impacting underlying data. The engine intelligently applies your business and domain rules at query time.

Work Seamlessly with Your Existing Tools

Knowledge graphs that contain virtualization can maintain data accuracy and security of existing tools. When a knowledge graph is a true data layer, you don't need to change anything else.

Get Started in AI and Machine Learning Quickly

An enterprise knowledge graph platform can help you predict, classify, or compare something,

Stardog's built-in predictive analytics and similarity search support rapid model development and iteration for data analysis. You can make intelligent predictions based on patterns you extract from data.

Infer New Facts

Machine learning complements logical reasoning. Inference expresses all implied and predicated connections between your data sources, creating a richer, more accurate view.

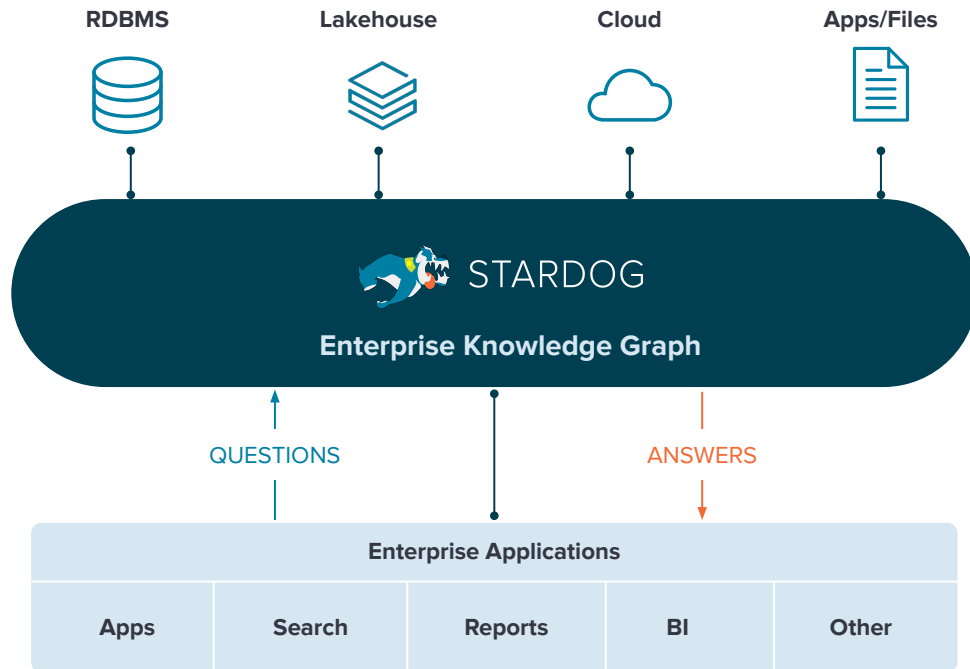
Adding Stardog's inference engine allows you to capture and infer new facts.

Chapter 4.

Connect data outside your lakehouse for a complete information landscape.

While a lakehouse provides the ideal view of the data lake, it doesn't usually capture an organization's full data landscape.

Adding a semantic layer to your data lake provides scalability and lower cost-per-GB storage, while speeding up unification. This combination generates a better ROI than data warehouses lacking the semantic layer.



Implementation: Big-Picture Thinking, One Step at a Time

By nature, many enterprise technology implementations are large and complex. Without a clear focus on business benefits, projects easily stagnate.

An enterprise knowledge graph offers a flexible, extensible data model. And, everything doesn't need to be modelled at once. Organizations can prioritize by determining which business questions are most urgent.

When prioritizing, focus on delivering business goals rather than technical capabilities. Tackling specific data domains while focusing on business objectives enables organizations to direct resources productively.

“

Our primary objective is to provide data at a higher quality and relieve the heavy lifting up front so our data scientists **can actually work with the data.**”

Head of IT Research,
Global Pharmaceutical Company

It's difficult to follow our instincts and explore data to find the root cause of issues.

Data sources define compounds in different ways, even though they're the same entity.

How do we detect fraud across the entire network of actors, assets, controls, and systems?

Chapter 5.

Make sense of real-world complexity.

Many specialized systems help individuals organize their day-to-day responsibilities. But, several factors compound complexities.

Data lakes consolidate enterprise data, but don't necessarily help **make sense of that data**. To do that, organizations need to model complexity.

Enterprise knowledge graphs capture complexity in an intuitive manner because they aren't encumbered by location or structure. They model data so it makes sense to the business by capturing the relationship between elements, making it easier to tackle complex use cases.

For supply chain analysis, organizations can make complex associations of data across the operation, and instantly see connections to enable scenario planning, identify needed changes, and boost resiliency.

“Data and analytics leaders rely on graphs to quickly answer complex business questions which require contextual awareness and an understanding of the nature of connections and strengths across multiple entities. By 2025, **graph technologies will be used in 80% of data and analytics innovations**, up from 10% in 2021, facilitating rapid decision making across the organization.”

Gartner.

Chapter 6.

Solve the Data and AI Challenge.

Getting fast and accurate data answers to make critical business decisions is **time consuming, inefficient, and slows growth.**

Data takes time

The data estate is fragmented, time intensive, and costly to obtain, prepare, and manage.

Deep insights are slow, hampering innovations and slowing growth.

No AI strategy without data strategy

AI with security, privacy, and accuracy hinges on data quality and establishing guardrails.

Current GenAI Solutions are hallucinating and don't provide traceability for regulatory compliance.

**Data is
fragmented and
sourced without
business context
or meaning**

“Accuracy is critical. We need to trust the data and ensure we remain compliant.”

RISK & COMPLIANCE EXECUTIVE, BANKING

“It takes IT weeks, if not months, to answer critical business questions.”

MANAGING DIRECTOR, CAPITAL MARKETS

“My team can only see answers they have access to — so it's imperative to have the right level of permissions and entitlements for each group of knowledge workers.”

DATA & ANALYTICS EXECUTIVE,
PHARMACEUTICAL

Stardog Voicebox on the Databricks lakehouse is intelligent data that is business ready.



Stardog Voicebox

An LLM-powered UX that allows any knowledge worker to chat with any enterprise data to discover hidden connections that are fast, accurate, secure and hallucination free

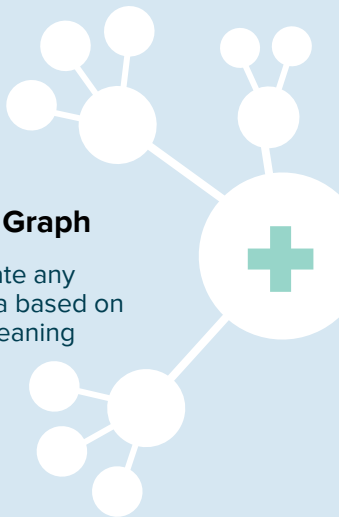


Stardog Voicebox is the World's First AI Data Assistant.

Anyone can ask **any question** about **any data** and get a fast, accurate and explainable answer.

Enterprise Knowledge Graph

Able to integrate any enterprise data based on its business meaning



Generative AI

Uses Generative AI to enable universal data access for knowledge workers to answer questions and create both insight and foresight



Conclusion

A Databricks lakehouse consolidates data that has amassed in silos. However, the biggest ROI comes when that data—along with data that must live outside the lakehouse—can be leveraged for demonstrated business value.

The combination of Databricks and Stardog helps data leaders:

- Unify data through business meaning
- Connect to all relevant data
- Enable data exploration and discovery
- Democratize Data + AI with business ready insights.

Accelerate Your Lakehouse Investment With Stardog

The Stardog Enterprise Knowledge Graph platform connects data based on business meaning to get better insight faster. Organizations like Boehringer Ingelheim, Schneider Electric, and NASA rely on Stardog's flexible, reusable semantic data layer to accelerate insights from data lakes, data warehouses, or any enterprise data source. Learn more at stardog.com.

STARDOG

Stardog created Voicebox, the world's first AI data assistant that allows any knowledge worker to get answers that are accurate, timely, secure, and hallucination-free. No code. No dashboards. Just answers and insights for everyone.

www.stardog.com

