



Databricks SQL: Intelligent Data Warehousing

Slide Deck



carahsoft.

For more information, contact Carahsoft or our reseller partners:

Databricks@carahsoft.com | 703-581-6693



Databricks SQL: Intelligent Data Warehousing

Cary Moore/ Senior SSA

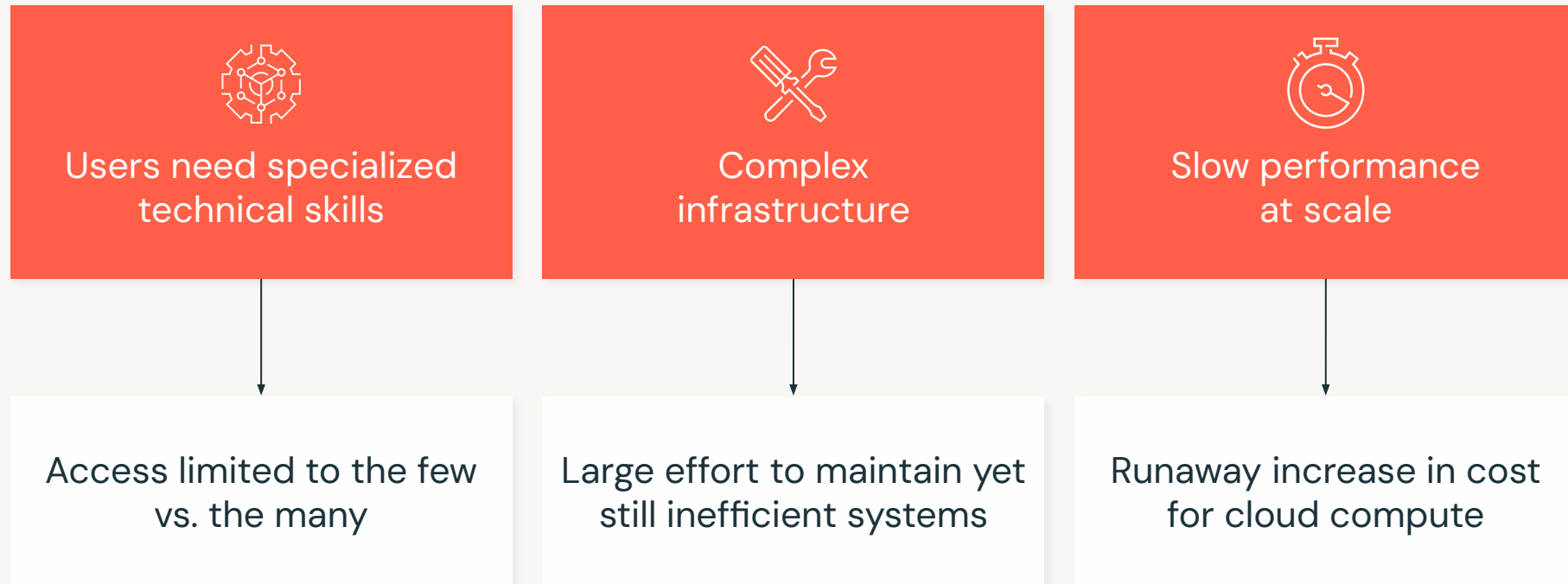
December, 2024

©2024 Databricks Inc. — All rights reserved



AI is changing everything —
including data warehousing

Today, data warehousing is complex and expensive



The new era of AI enables a new approach

Introducing intelligent data warehousing



Introducing **intelligent** data warehousing



User experiences with
natural language and
data intelligence



Access for everyone to
ask questions of their data



Predictive optimizations
for your infrastructure



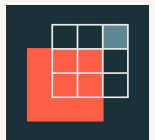
Intelligent, automated
management and tuning



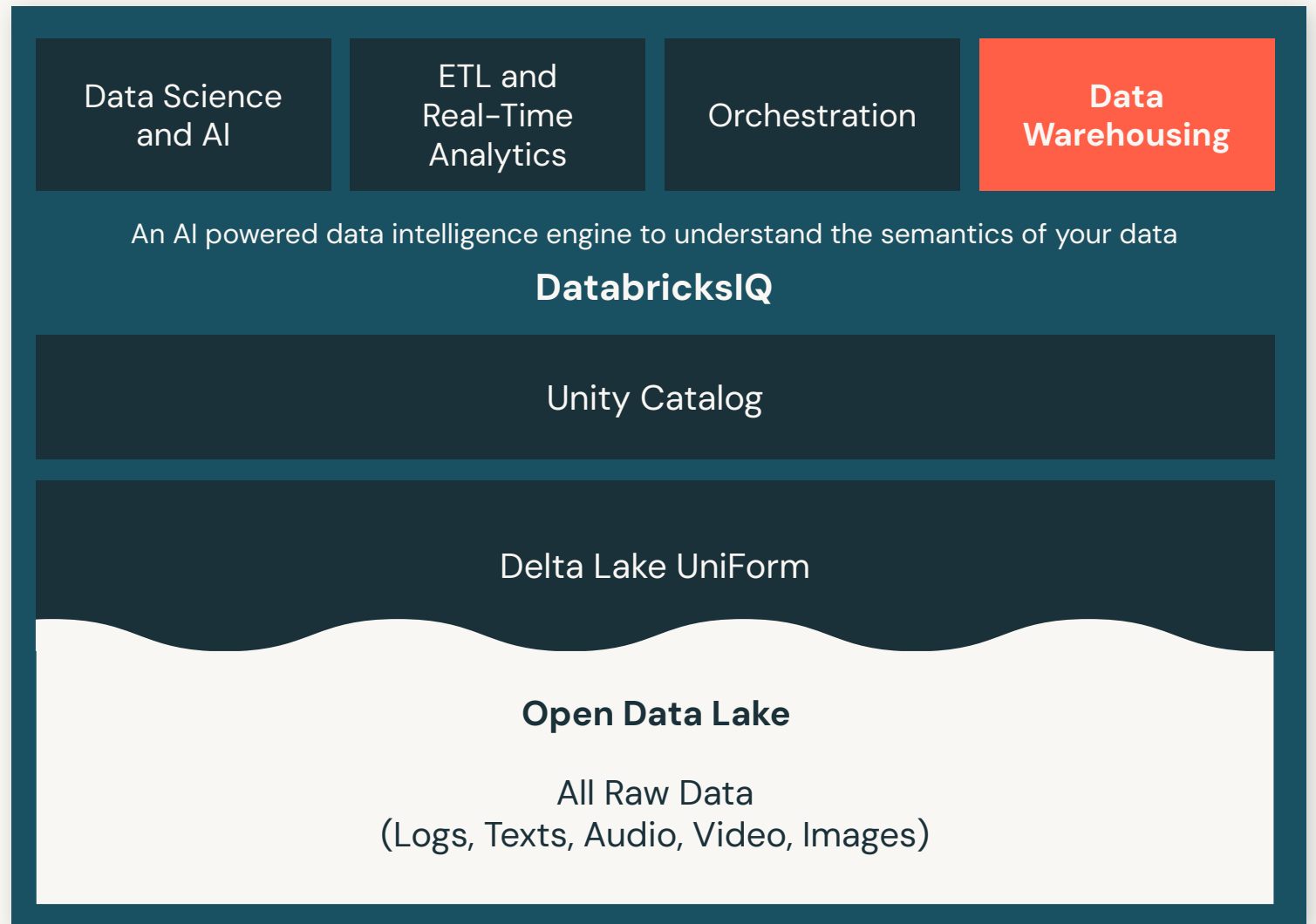
World-class price
performance



The best TCO
in the market



Databricks SQL intelligent data warehousing on the Data Intelligence Platform



Databricks SQL

Built on a **lakehouse architecture**

Unified architecture
for all your data, BI and AI

Reduce costs with the only warehouse that doesn't create a new data silo

Unified governance
for all your data + AI assets

Simplify governance with a single source of truth

Unified data teams
with SQL + Python for everyone

Improve productivity for people by working with SQL and Python together

Open data sharing
and marketplace

Improve collaboration by securely sharing data + AI assets with anyone



Databricks SQL

Data warehousing on a lakehouse architecture is now standard

What Is a Lakehouse?



by Ben Lorica, Michael Armbrust, Reynold Xin, Matei Zaharia and Ali Ghodsi

January 30, 2020 in [Engineering Blog](#)

Share this post



Over the past few years at Databricks, we've seen a new data management architecture that emerged independently across many customers and use cases: [the lakehouse](#). In this post we describe this new architecture and its advantages over previous approaches.

Data warehouses have a [long history](#) in decision support and business intelligence applications. Since its inception in the late 1980s, data warehouse technology continued to evolve and MPP architectures led to systems that were able to handle larger data sizes. But while warehouses were great for structured data, a lot of modern enterprises have to deal with unstructured data, semi-structured data, and data with high variety, velocity, and volume. Data warehouses are not suited for many of these use cases, and they are certainly not the most cost efficient.



2020

Databricks pioneered the lakehouse architecture



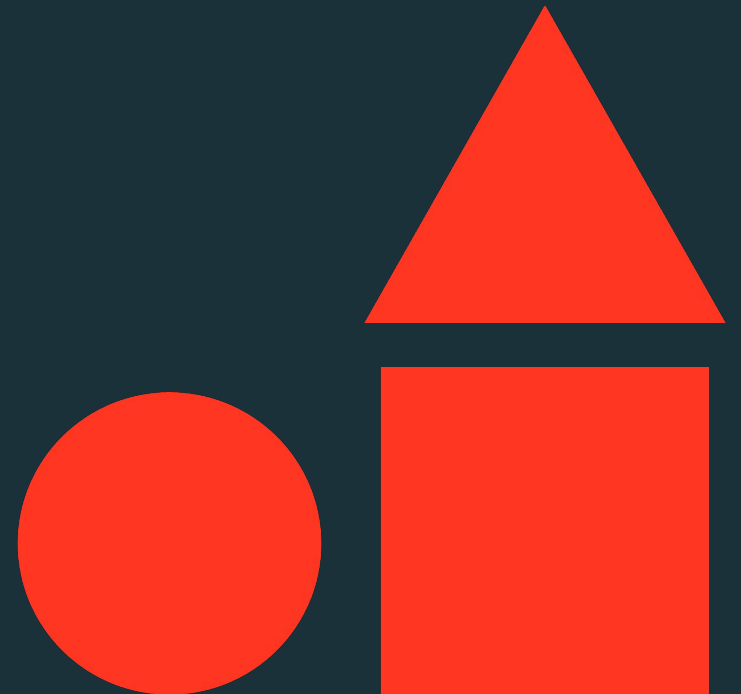
Today

74% of global enterprises have adopted lakehouse

MIT Technology Review Insights, 2023



Intelligent data warehousing with Databricks SQL



Three

reasons why
customers
love Databricks SQL

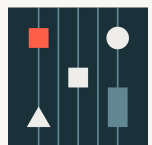
- 1 **Intelligent Experiences** with built-in understanding of your data
- 2 **Predictive Optimizations** for all your workloads
- 3 **Best Price/Performance** for the lowest TCO

1

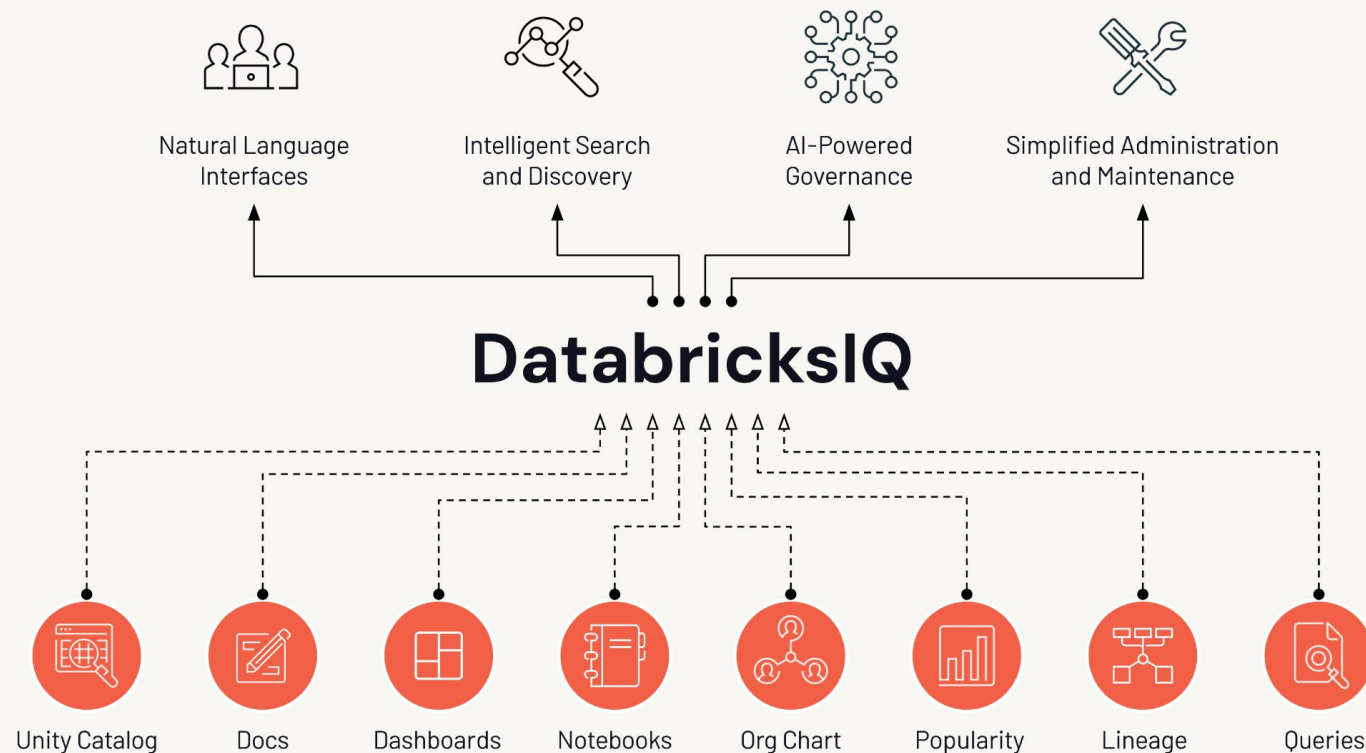
Intelligent experiences with built-in understanding of your data

DatabricksIQ

Apply AI to your data to increase productivity for every employee

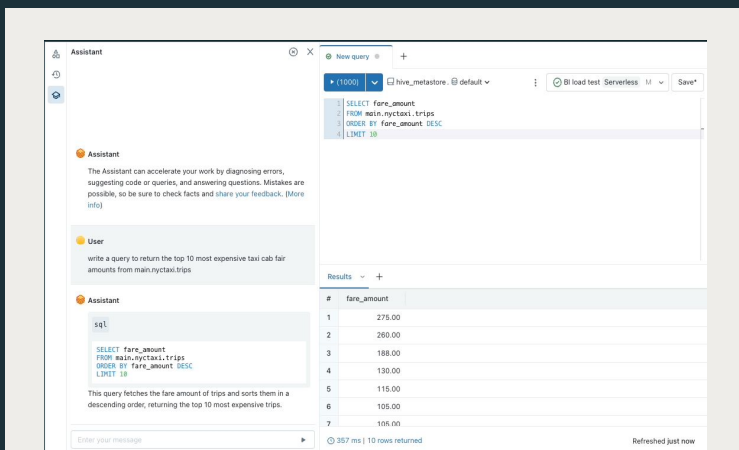


AI that understands **your data**, **your workloads** and **your business context** to power natural language interfaces and the predictive optimization of the platform.



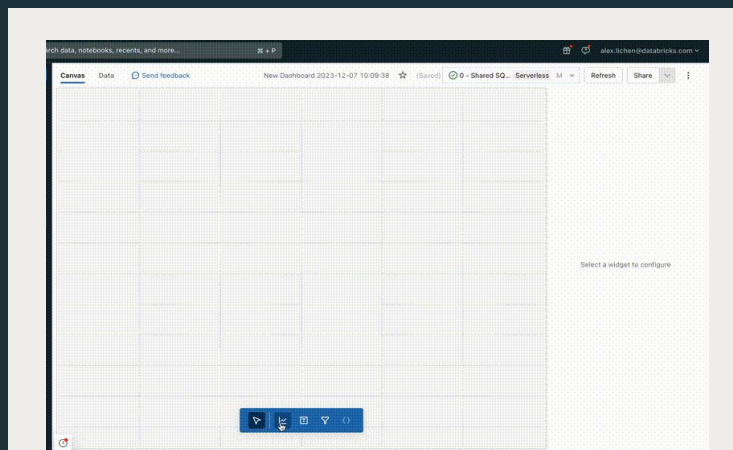
Intelligent experiences with natural language

Powered by DatabricksIQ



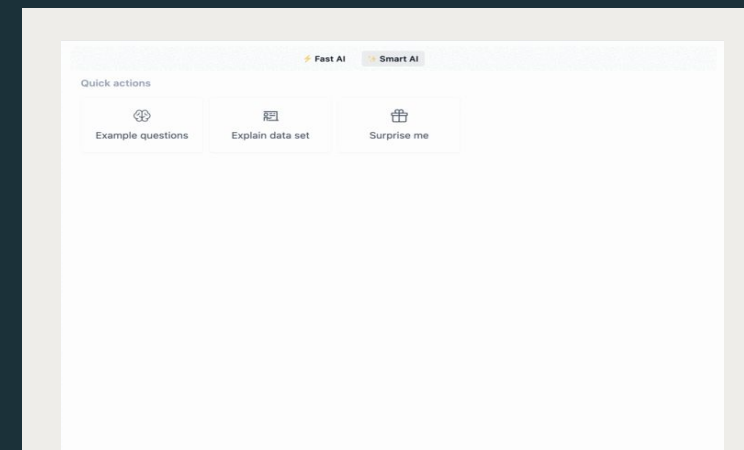
SQL Editor

Data Science / Engineers



Dashboards

Analysts



Project Genie

Coming Soon!

Business Users

Assistant in SQL Editor

Developers can create, explain and fix SQL code



Assistant in Query Editor

The screenshot shows the Databricks Assistant interface. On the left, the Assistant chat window displays a conversation where the user asks for a query to get the top 10 most expensive taxi cab fares. The Assistant responds with the following SQL query:

```
sql
SELECT fare_amount
FROM main.nyctaxi.trips
ORDER BY fare_amount DESC
LIMIT 10
```

The Assistant explains: "This query fetches the fare amount of trips and sorts them in a descending order, returning the top 10 most expensive trips." Below the chat is an input field labeled "Enter your message".

On the right, the SQL Editor shows the same query in a new query window. The results table is displayed below the query:

#	fare_amount
1	275.00
2	260.00
3	188.00
4	130.00
5	115.00
6	105.00
7	105.00

At the bottom of the results table, it shows "357 ms | 10 rows returned" and "Refreshed just now".

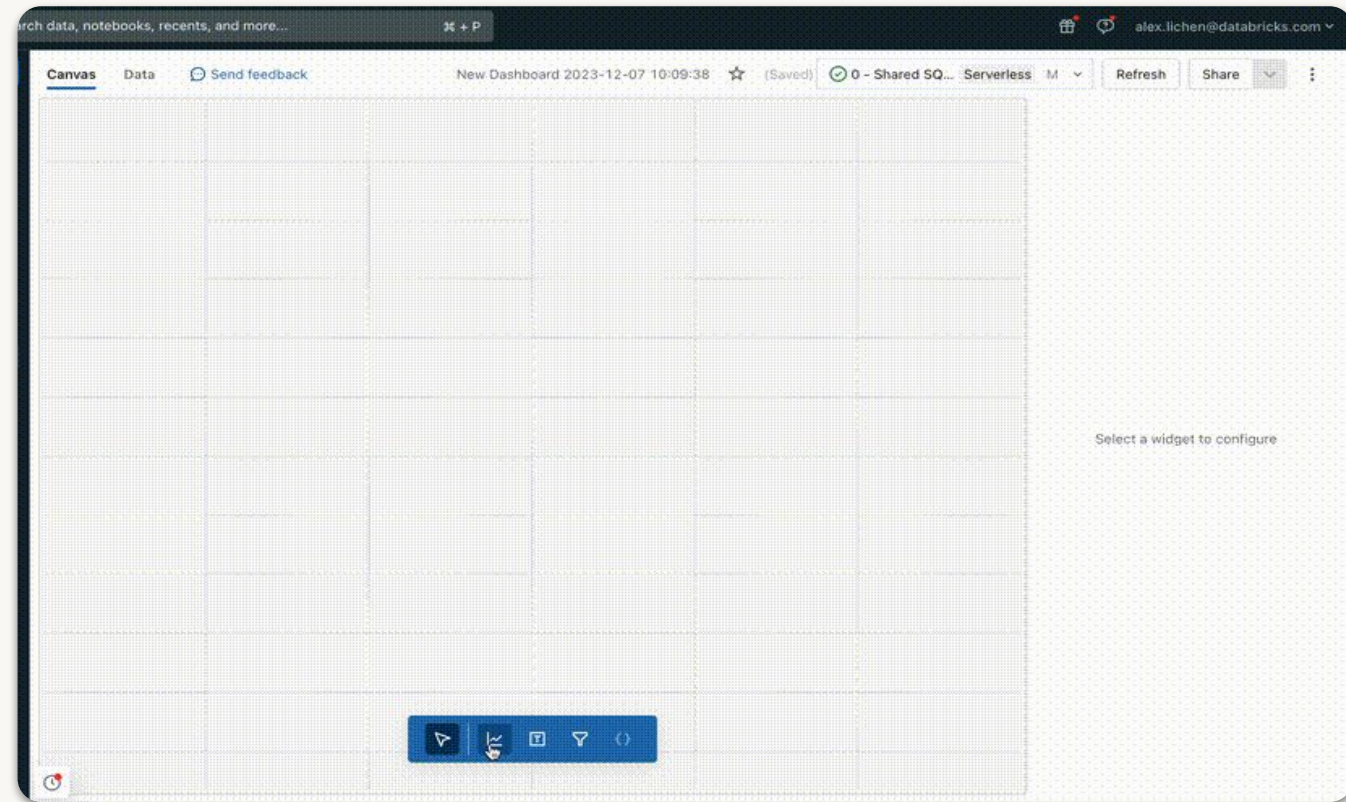


Assistant in Dashboards

Analysts can generate visualizations using natural language



Assistant in
Dashboards



Project Genie

Business users can query and visualize data using natural language



Project Genie

The screenshot shows the Databricks Project Genie interface. On the left is a sidebar with the Databricks logo, a search bar, and a list of items including 'EMEA Sales Analysis Data', 'Chats', 'Tables', and 'Instructions'. The main area has a '+ New chat' button and a message input field with the text 'Name all the unique stores with their footfall'. The user's name 'Vidhi Agrawal' is shown at the bottom left. The interface also features 'Quick actions' like 'Example questions', 'Explain data set', and 'Surprise me'.

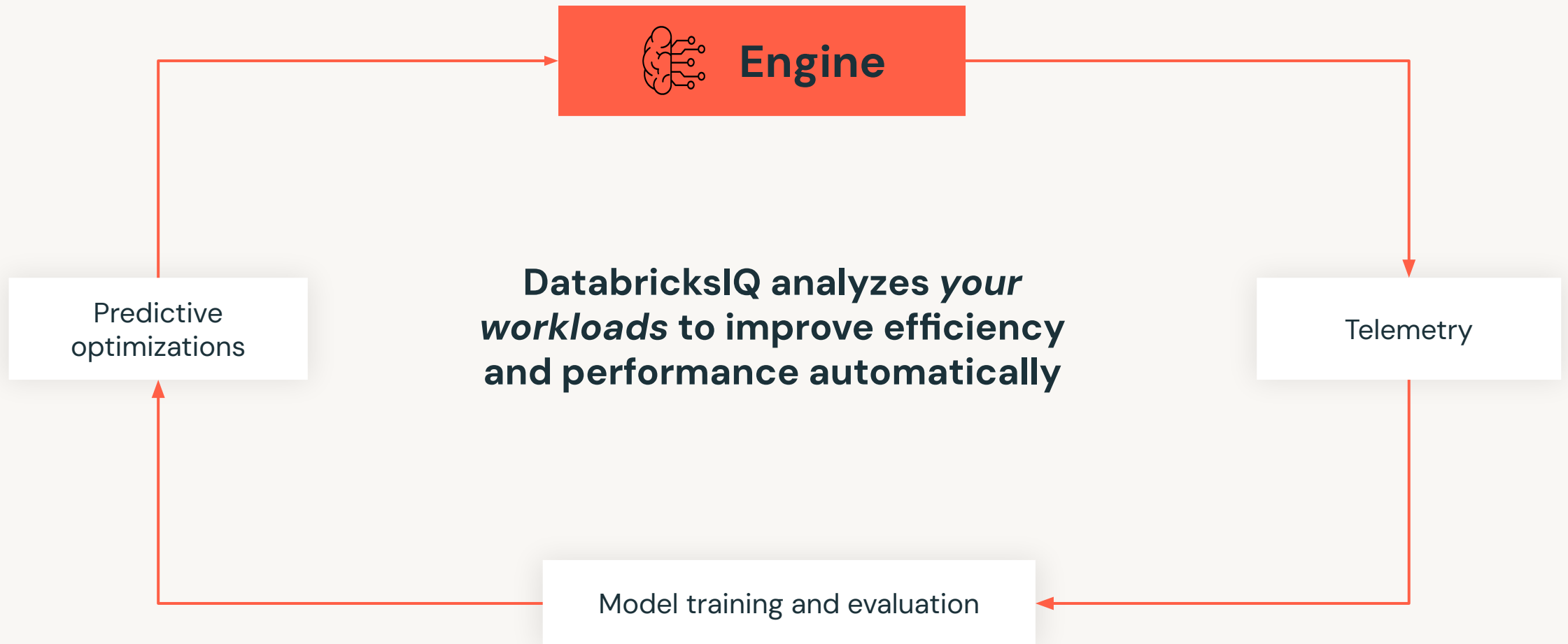


2

Predictive optimizations for all your workloads

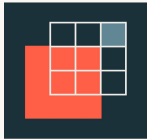
Predictive optimizations in our engine

Powered by DatabricksIQ

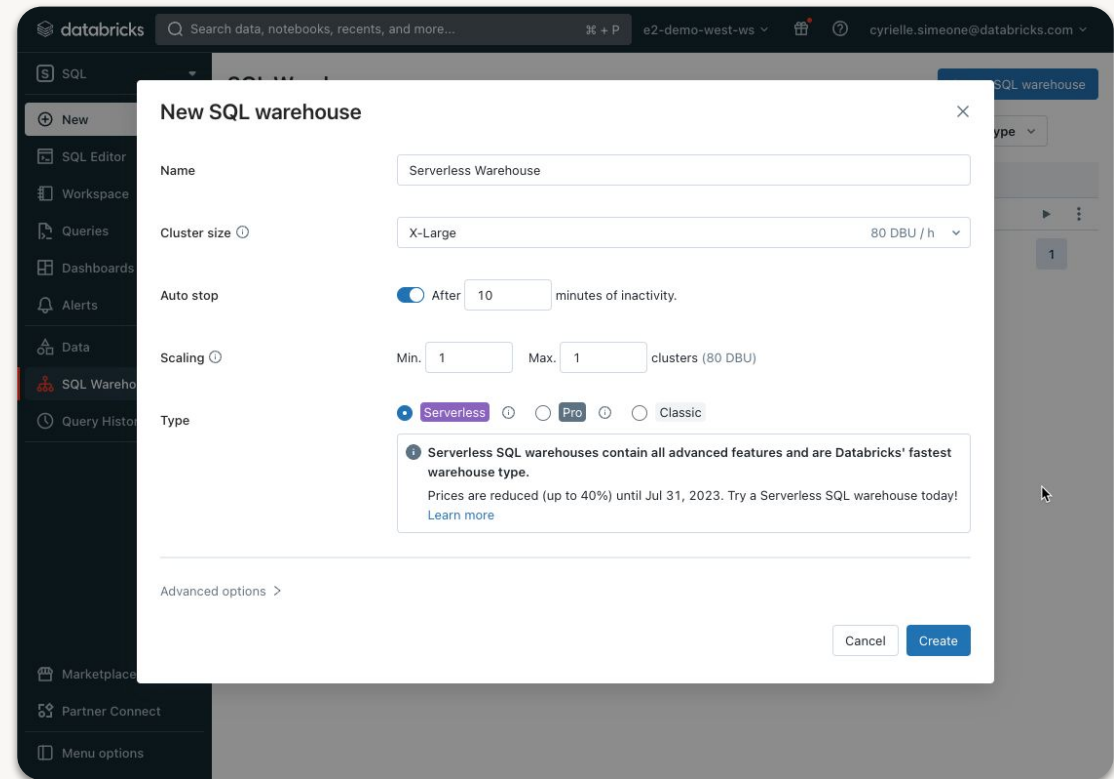


Databricks SQL Serverless

Instant, scalable compute for all your DW/BI workloads



Get the best performance, lower costs, and focus on delivering value rather than managing infrastructure.

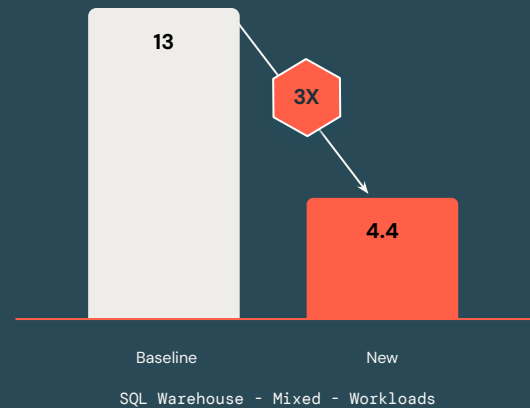


Predictive optimizations in our engine

Examples

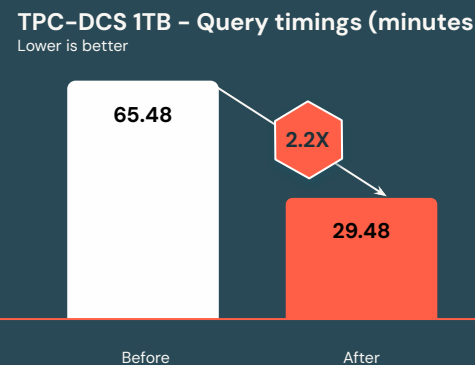
Intelligent Workload Management

Leverages machine learning to efficiently route queries and scale clusters to maximize cost/performance



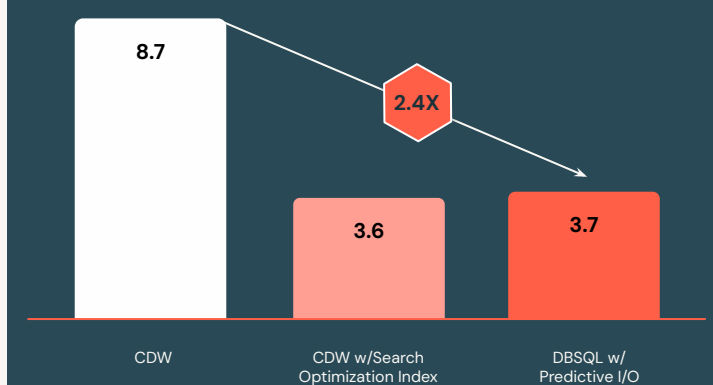
Automatic Data Layout

Eliminates knobs to optimize storage with ROI-based table maintenance algorithms



Indexless Indexing

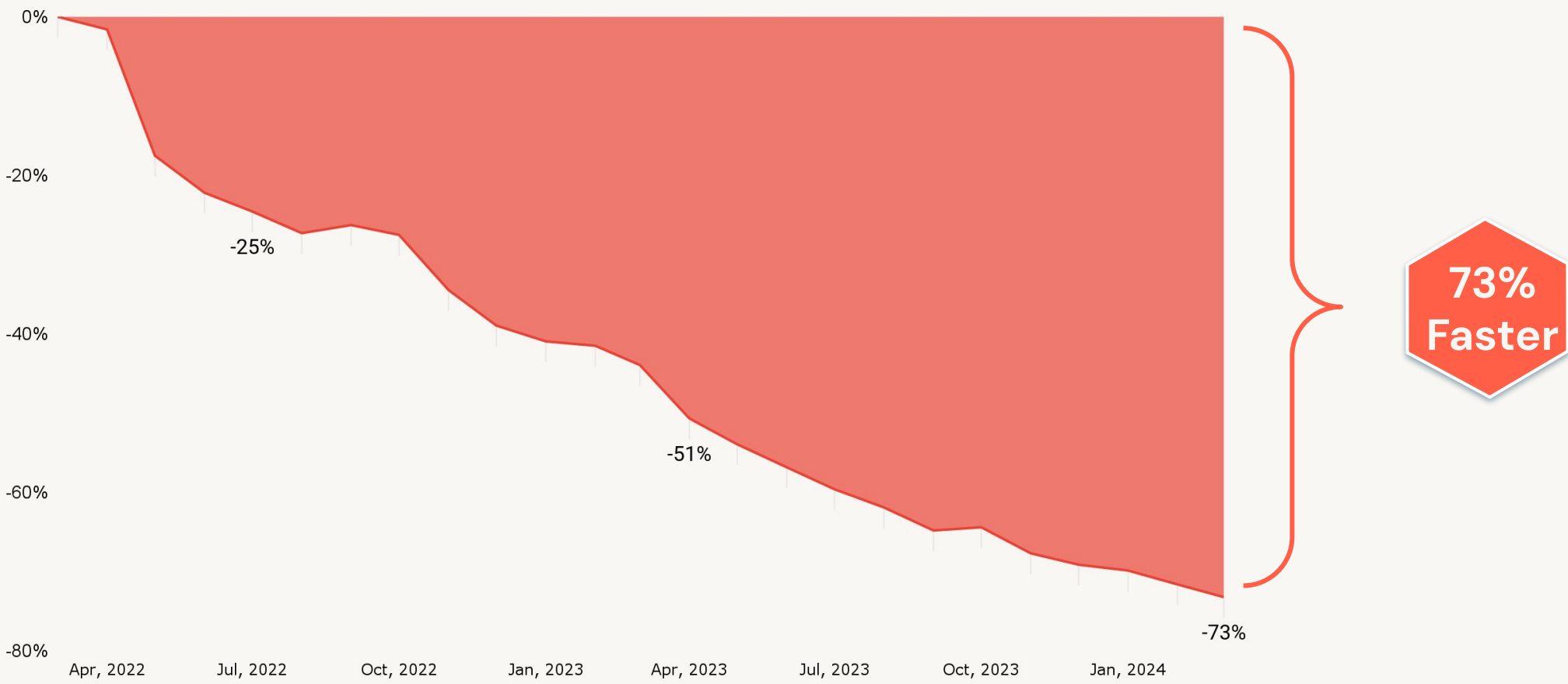
Predictive i/o delivers comparable performance **without** expensive search-optimized indexes



Performance improvements time

Customer BI queries – improved 73% over last 2 years

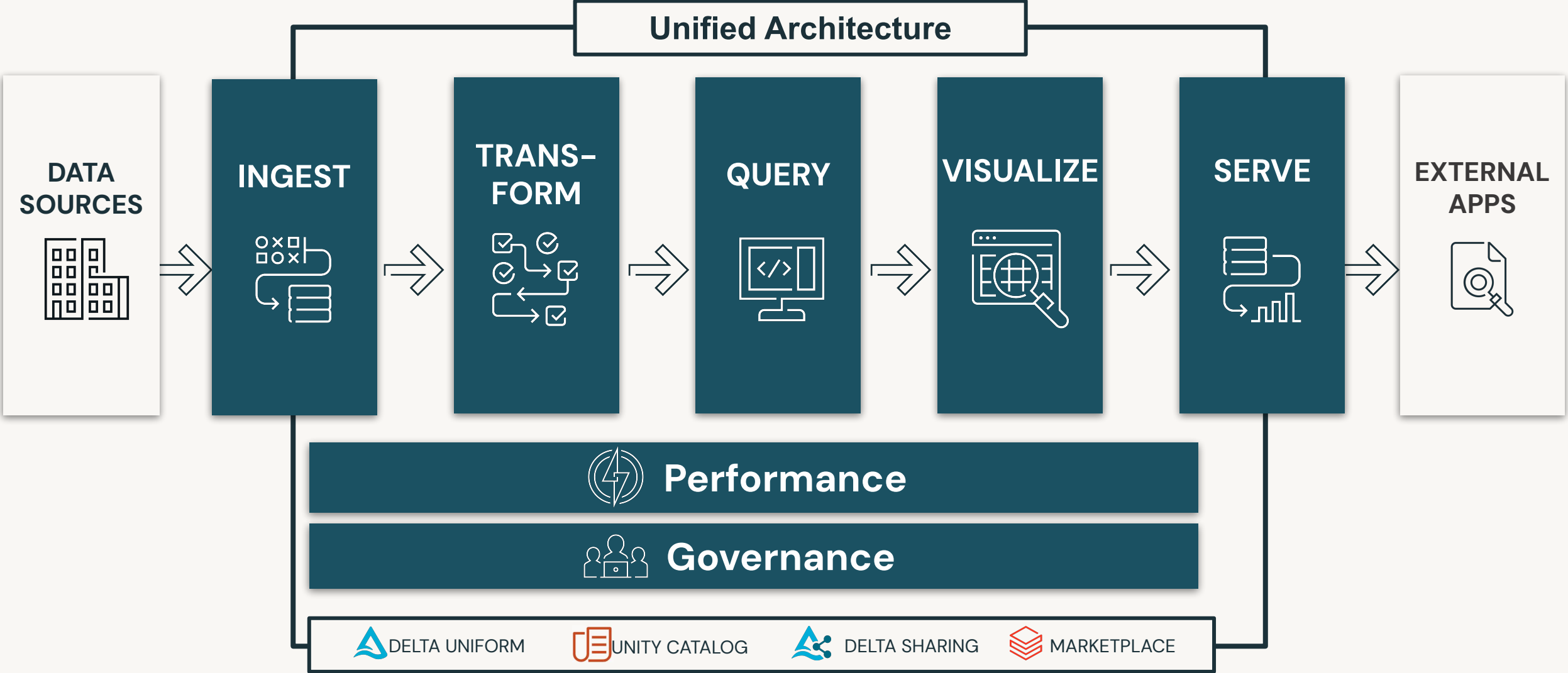
Performance Improvement over time



Your Data On Lakehouse

One Platform for ALL of your Analytics

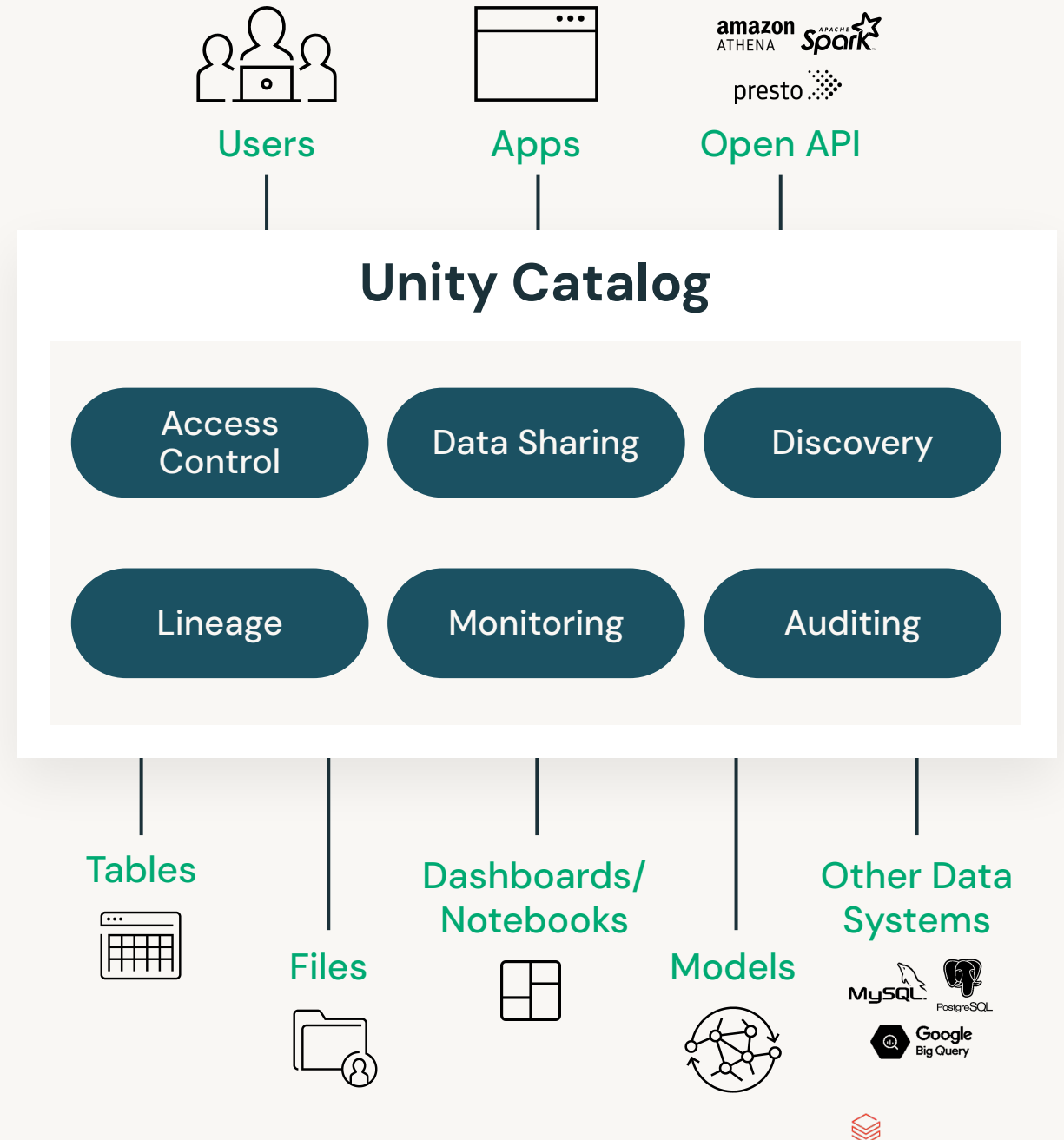
Complete Data Warehousing Solution



Databricks Unity Catalog

Unified governance, security and collaboration is the foundation for the

**Data Intelligence
Platform**



Lakehouse Federation

Discover, query, and govern all your data—no matter where it lives

Build a unified view of your data estate

Query and combine all data efficiently with a single engine

Safeguard data across data sources



Databricks recognized as a leader in the industry

Gartner MQ
Leader
Database Management Systems



Forrester Wave
Leader
Data Lakehouses

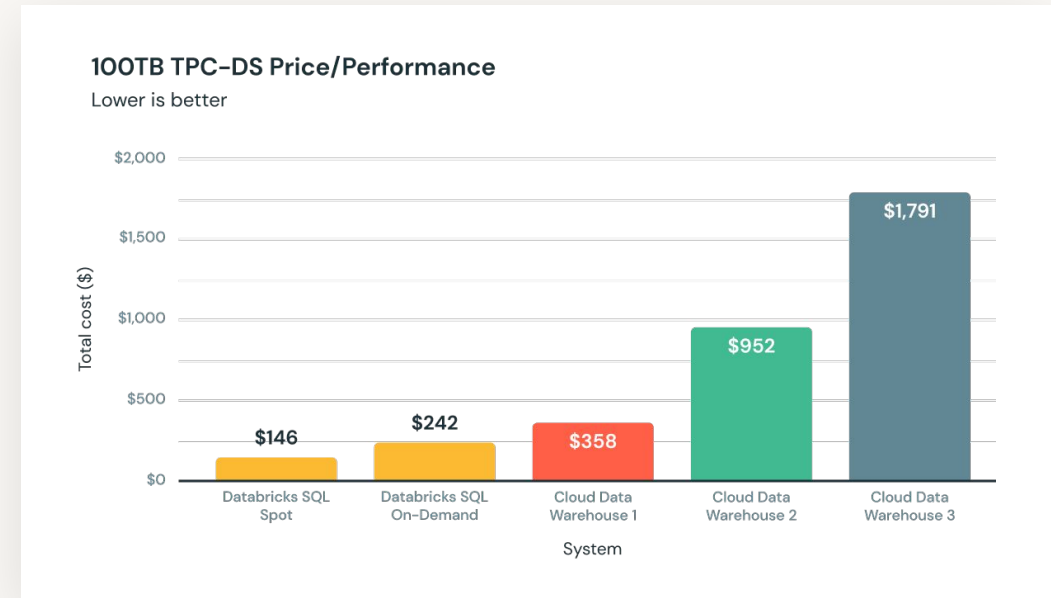
3

Best price/performance
for the lowest TCO

Photon Query Engine

Next-gen, vectorized query engine with world-record performance

Written from the ground up in C++, Photon takes advantage of modern hardware for faster queries, providing TPC world-record price/performance compared to other cloud data warehouses — all natively on your data lake.



SIGMOD 2022
Best Industry
Paper Award

Photon: A Fast Query Engine for Lakehouse Systems

Alexander Behm, Shoumik Palkar, Utkarsh Agarwal, Timothy Armstrong, David Cashman, Ankur Dave, Todd Greenstein, Shant Hovsepian, Ryan Johnson, Arvind Sai Krishnan, Paul Leventis, Ala Luszczak, Prashanth Menon, Mostafa Mokhtar, Gene Pang, Sameer Paranjpye, Greg Rahn, Bart Samwel, Tom van Bussel, Herman van Hovell, Maryann Xue, Reynold Xin, Matei Zaharia
photon-paper-authors@databricks.com
Databricks Inc.

ABSTRACT

Many organizations are shifting to a data management paradigm called the "Lakehouse" which implements the functionality of structured data warehouses on top of unstructured data lakes. This paradigm offers excellent performance on the raw un-

structured data. Traditionally, for the most demanding SQL workloads, enterprises have also moved a curated subset of their data into data warehouses to get high performance, going to a two-tier architecture. However, this two-tier architecture is expensive, as only a subset of data is available in the warehouse, and this data may be out of sync with the raw data due to issues in the extract, transform and load (ETL) process [19]. In response, many organizations are shifting to a data management approach called the Lakehouse [19], which implements data in a single tier, ACID transactions and rich analytics.



World-class performance & TCO

Meets or beats the price/performance of major CDWs across scales!

PERF

TPC-DS Elapsed Time (Lower is Better)



COST

TPC-DS Total Cost (Lower is Better)



SLED Customers Leading The Way

UNIVERSITY OF CALIFORNIA Office of the President



VANDERBILT UNIVERSITY MEDICAL CENTER



UC DAVIS HEALTH



saif Work. Life. Oregon.

UCLA Health

Education



State Government



UC San Diego Health

CENTER FOR AN INFORMED PUBLIC UNIVERSITY of WASHINGTON



WaTech Washington Technology Solutions



King County



Employment Security Department WASHINGTON STATE



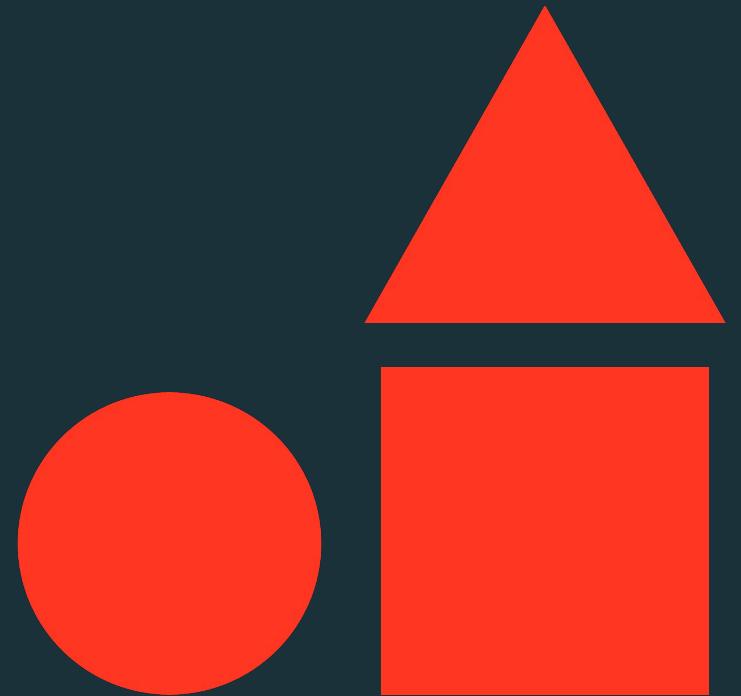
UCI Health



Local Government



Demo: Intelligent data warehousing with Databricks SQL



Q&A



Databricks SQL

Get started with intelligent data warehousing!



Try for free

www.databricks.com/try

\$400 free serverless credits

Learn more:

databricks.com/product/databricks-sql



Databricks SQL

Explore more resources

Virtual Event

[Watch →](#)



Reynold Xin
DATABRICKS



Erika Ehrli
DATABRICKS



Pearl Ubaru
DATABRICKS



Narinder Singh
SALESFORCE



Gaurav Saraf
DATABRICKS



Alex Lichen
DATABRICKS

databricks

VIRTUAL

Data Warehousing in the Era of AI

Using data intelligence to revolutionize data management

Demos

[Watch demos →](#)

- TUTORIAL: AI Functions: Query LLMs With DB SQL
- ON-DEMAND VIDEO: Data Ingestion using COPY INTO
- ON-DEMAND VIDEO: Data Warehousing on Databricks
- TUTORIAL: Data Warehousing With Identity, Primary Key and Foreign Key
- ON-DEMAND VIDEO: Databricks Serverless Security Overview Part 1: Journey of a DBSQL Query
- ON-DEMAND VIDEO: Databricks SQL
- ON-DEMAND VIDEO: Get Data Into Databricks from SQL / Oracle
- ON-DEMAND VIDEO: Getting Started on Databricks SQL for Administrators
- ON-DEMAND VIDEO: Tutorial - Databricks SQL Workspace | Databricks Academy







Thank you for downloading this Databricks presentation! Carahsoft serves as the Master Government Aggregator® and Distributor for Databricks, offering expertise in government procurement processes and practices with purchasing available via GSA, SEWP V, ITES-SW and other contract vehicles.

To learn how to take the next step toward acquiring Databricks' solutions, please check out the following resources and information:



For additional resources, please visit carah.io/DatabricksResources



For additional solutions, visit carah.io/DatabricksSolutions



To speak with our team directly, email Databricks@carahsoft.com or reach out at 703-581-6693.



To view our upcoming Databricks events, visit carah.io/DatabricksEvents



For additional Open Source solutions, visit carah.io/OpenSourceSolutions



To purchase, check out the contract vehicles available for procurement at carah.io/DatabricksContracts

carahsoft

For more information, contact Carahsoft or our reseller partners:
Databricks@carahsoft.com | 703-581-6693