

# How DatabricksIQ boosts your productivity



---

**Weston Hutchins, Sr. Manager, Product**  
**Richard Tomlinson, Director of Product Marketing**

# Introduction

# Data Lakehouse

An open, unified foundation  
for all your data



# Generative AI

Easily scale and use data and AI

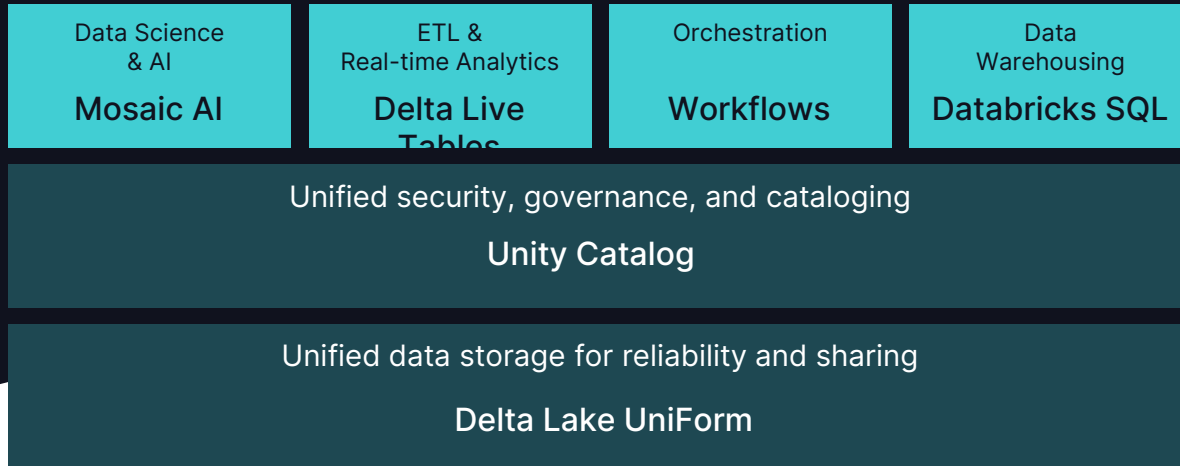


# Data Intelligence Platform

Democratize data + AI across  
your entire organization



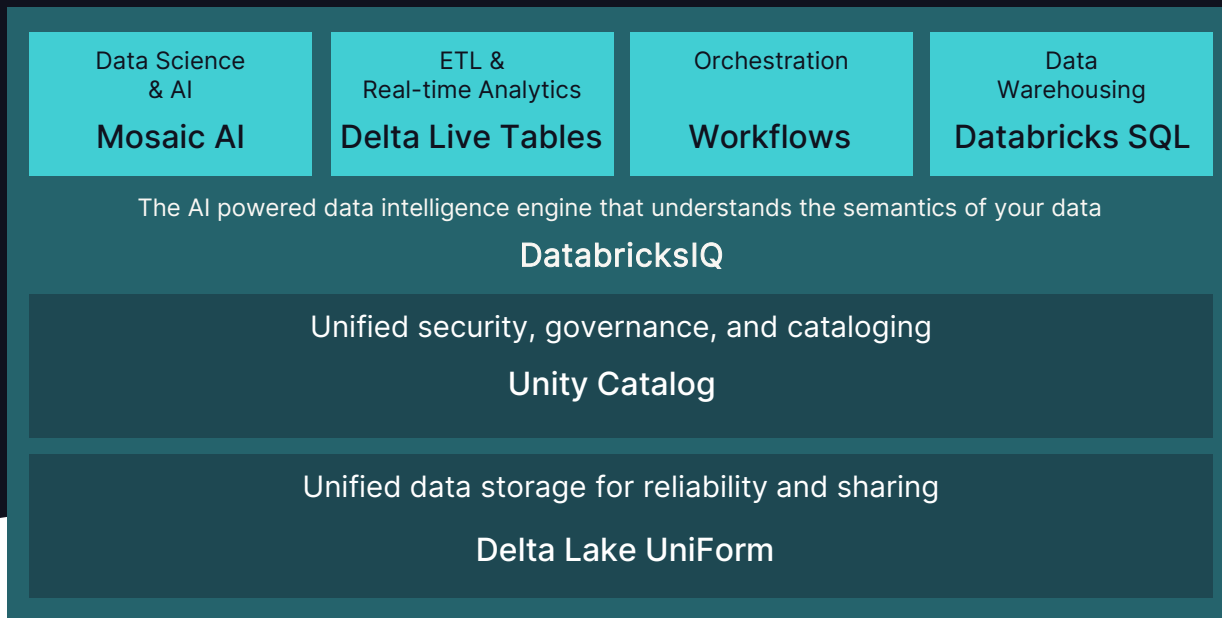
# Databricks Data Intelligence Platform



## Open Data Lake

All Raw Data  
(Logs, Texts, Audio, Video, Images)

# Databricks Data Intelligence Platform



## Open Data Lake

All Raw Data  
(Logs, Texts, Audio, Video, Images)

# Databricks Data Intelligence Platform

## Mosaic AI

Create, tune, and serve custom LLMs

## Delta Live Tables

Automated data quality

## Workflows

Job cost optimized based on past runs

## Databricks SQL

Text-to-SQL  
Text-to-Viz

The AI powered data intelligence engine that understands the semantics of your data

## DatabricksIQ

### Unity Catalog

Securely get insights in natural language

### Delta Lake UniForm

Data layout is automatically optimized based on usage patterns

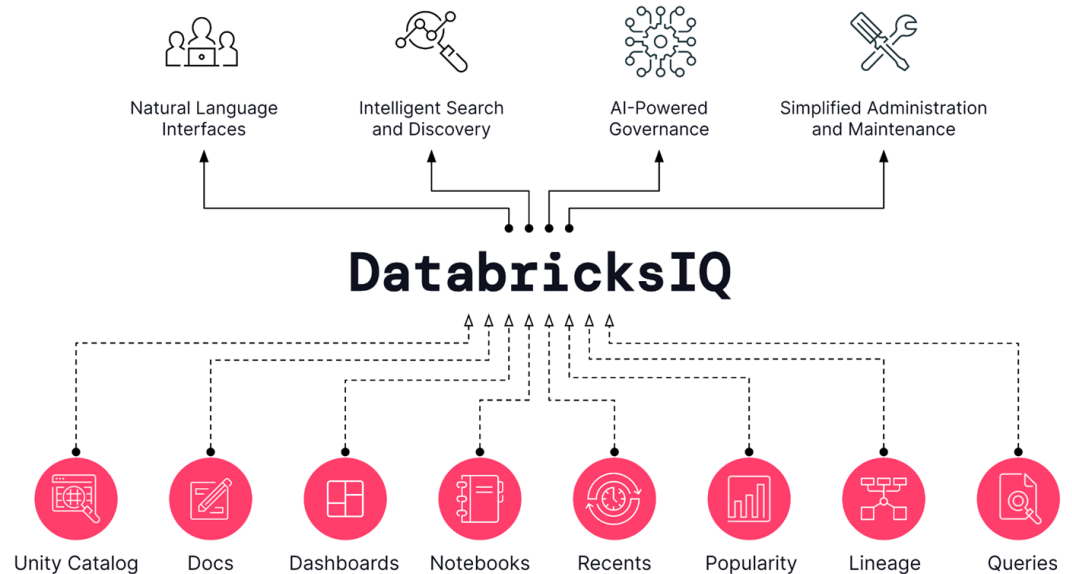
## Open Data Lake

All Raw Data  
(Logs, Texts, Audio, Video, Images)

# DatabricksIQ

## The AI-Powered Data Intelligence Engine

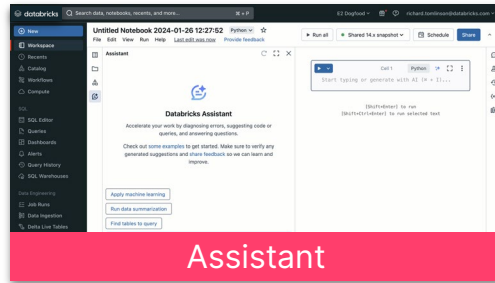
- ▶ AI that uniquely understands your data and usage patterns
- ▶ Boosts productivity for everyone through natural language interfaces
- ▶ Enables the automatic optimization of your data and AI workloads



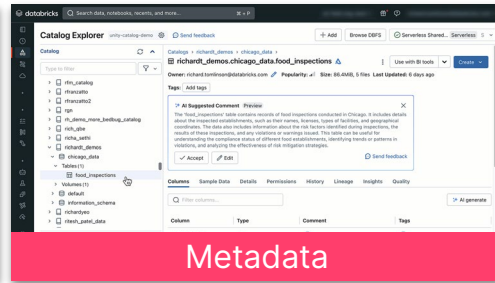
# Databricks IQ-Powered Tools

Boosts productivity for everyone through natural language interfaces

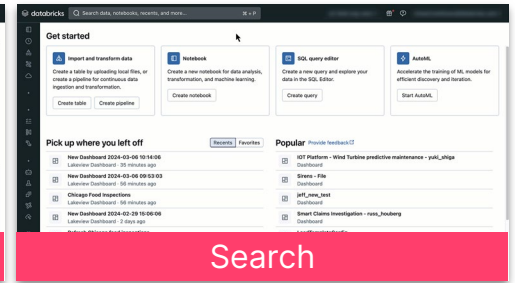
For Data Teams



Assistant

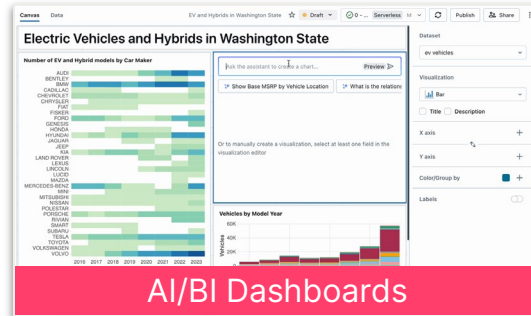


Metadata

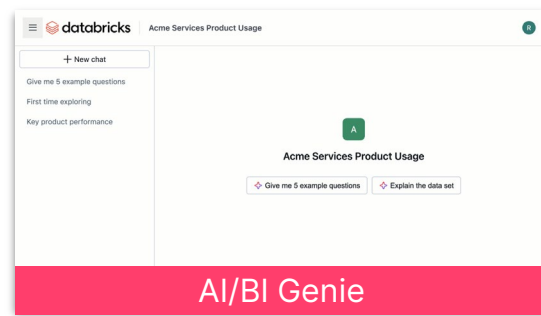


Search

For Business Teams



AI/BI Dashboards



AI/BI Genie



# DatabricksIQ Ingredients



# LLMs need great retrieval to work

Search is the foundation of DatabricksIQ and is the key to having relevant, accurate responses

 weston.hutchins@databricks.com

who are my top customers in 2024?

 Assistant

To answer this question, I would need information about your customers and their transactions or interactions with your business in 2024. This could include data on purchases, order values, frequency of transactions, or any other metric you use to define "top customers."

If you have a table or dataset that includes customer transactions for 2024, please provide the table name and relevant column details. Without this information, I cannot provide a specific solution.



# Retrieval in the Assistant

## Context is king

- In-product context (e.g. notebook content)
- Databricks documentation (hybrid vector search)
- Table search (using Databricks APIs)
- Query table (parsing SQL query and using UC metadata API)
- DataFrame schemas
- Previous conversations



# Prompt Engineering

*You are the Databricks assistant. The user is diagnosing an error in the Databricks notebook and is looking for a solution. This can either be an explanation of the problem with the fixed code or steps to execute outside the code to resolve the problem.*

*Here is the cell that was executed: ```python **[[ CELL CONTENT ]]**```*

*Here are the cells that come before this cell: **[[ CELLS BEFORE ]]***

*Here are the cells that come after: **[[ CELLS AFTER ]]***

*Error message: **[[ ERROR MESSAGE ]]***

*Here are some documentation pages that could help solve the problem:*

***[[ DOC 1 ]]***

***[[ DOC 2 ]]***



# What about the models?

# Delete a column in Delta Lake?

## Fine-tuning with Databricks examples

### Vanilla OSS Model

Note: In Databricks, you cannot directly delete a column from a table. Instead, you need to create a new table with the desired columns and then drop the original table.

Here's an example of how you can achieve this:

Step 1: Create a new table with the desired columns

```
CREATE TABLE main.eng.logs_new AS
SELECT <all the other columns>
FROM main.eng.logs;
```

Step 2: Drop the original table

```
DROP TABLE main.eng.logs;
```

Step 3: Rename the new table to the original table name

```
ALTER TABLE main.eng.logs_new RENAME TO main.eng.logs;
```

...

# Delete a column in Delta Lake?

## Fine-tuning with Databricks examples

### Vanilla OSS Model X

Note: In Databricks, you cannot directly delete a column from a table. Instead, you need to create a new table with the desired columns and then drop the original table.

Here's an example of how you can achieve this:

Step 1: Create a new table with the desired columns

```
CREATE TABLE main.eng.logs_new AS
SELECT <all the other columns>
FROM main.eng.logs;
```

Step 2: Drop the original table

```
DROP TABLE main.eng.logs;
```


Step 3: Rename the new table to the original table name

```
ALTER TABLE main.eng.logs_new RENAME TO main.eng.logs;
```

...

# Delete a column in Delta Lake?

## Fine-tuning with Databricks examples

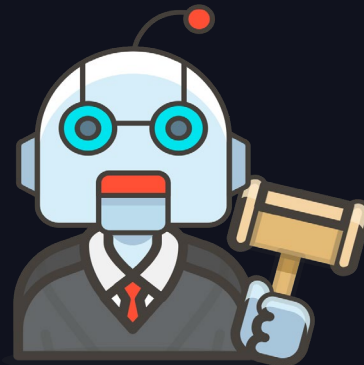
| Vanilla OSS Model <span style="color: red;">✗</span>  | Fine-tuned on Databricks documentation <span style="color: green;">✓</span>   |
|---|---|
| <p>Note: In Databricks, you cannot directly delete a column from a table. Instead, you need to create a new table with the desired columns and then drop the original table.</p> <p>Here's an example of how you can achieve this:</p> <p>Step 1: Create a new table with the desired columns</p> <pre>CREATE TABLE main.eng.logs_new AS SELECT &lt;all the other columns&gt; FROM main.eng.logs;</pre> <p>Step 2: Drop the original table</p> <pre>DROP TABLE main.eng.logs;</pre> <p>Step 3: Rename the new table to the original table name</p> <pre>ALTER TABLE main.eng.logs_new RENAME TO main.eng.logs;</pre> <p>...</p> | <p>SQL:</p> <pre>ALTER TABLE main.eng.logs DROP COLUMN _partition_backfill_version;</pre>  |



# Evaluation

## Measuring quality

- Hard to find **realistic** eval sets
- **Supervised** LLM-as-judge
- In-product **side-by-side** results
- **Human-in-the-loop** labeling and curation



# Latency

## Quality vs. responsiveness

- Is it better to have a faster answer with occasionally lower accuracy or a slower, more accurate answer? **It depends!**
- Diagnose error = Accuracy is key
- AI Autocomplete = Latency is key



# DatabricksIQ Ingredients

## High-quality AI experiences

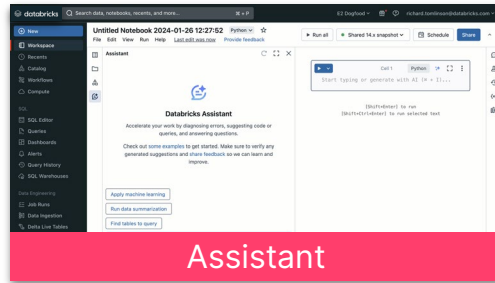
- Retrieval
- Prompt Engineering
- Fine-tuned Models
- Evaluation
- Latency
- And many more: UX, personalization, knowledge graphs, NL parsing, ...

# Boost productivity for data teams

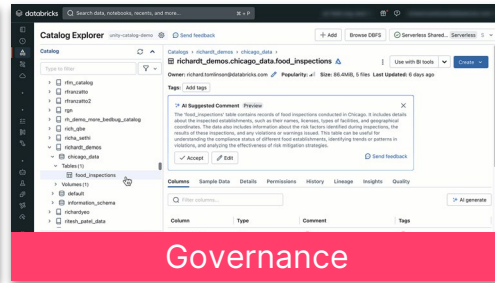
# Databricks IQ-Powered Tools

Boosts productivity for everyone through natural language interfaces

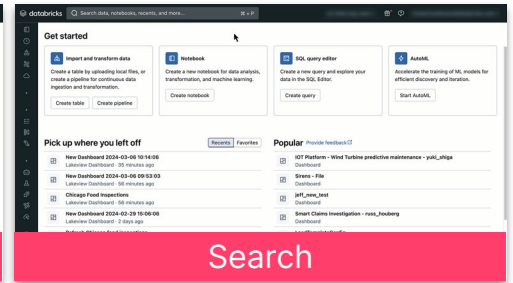
For Data Teams



Assistant

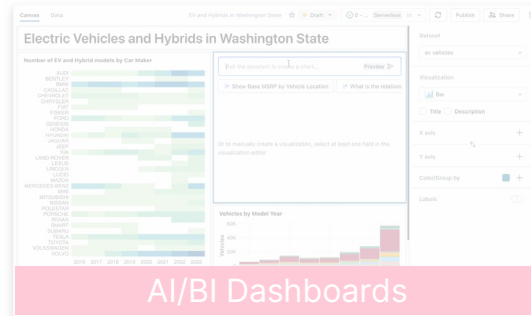


Governance

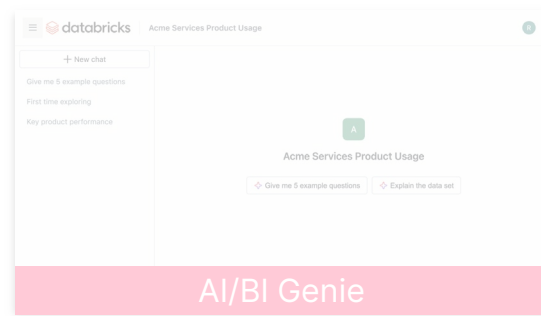


Search

For Business Teams



AI/BI Dashboards

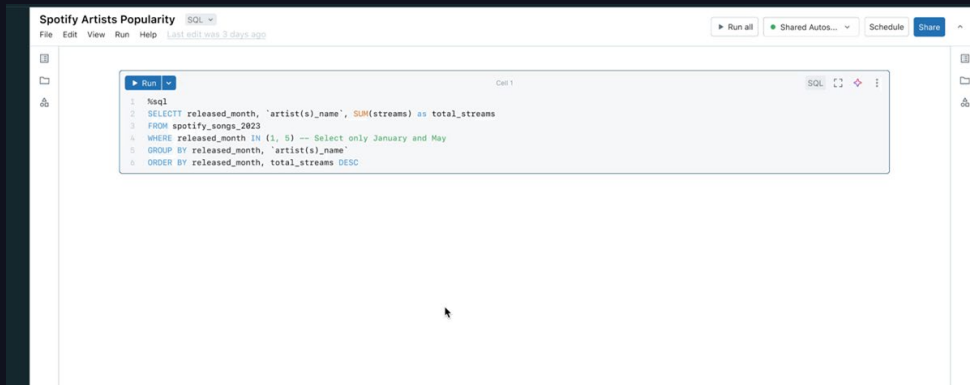


AI/BI Genie

# Demo

# Coming enhancements

- Auto-fix errors
- More inline experiences (e.g., rename, refactors, quick refinement)
- Performance optimization
- Diagnose job errors
- Catalog Explorer help



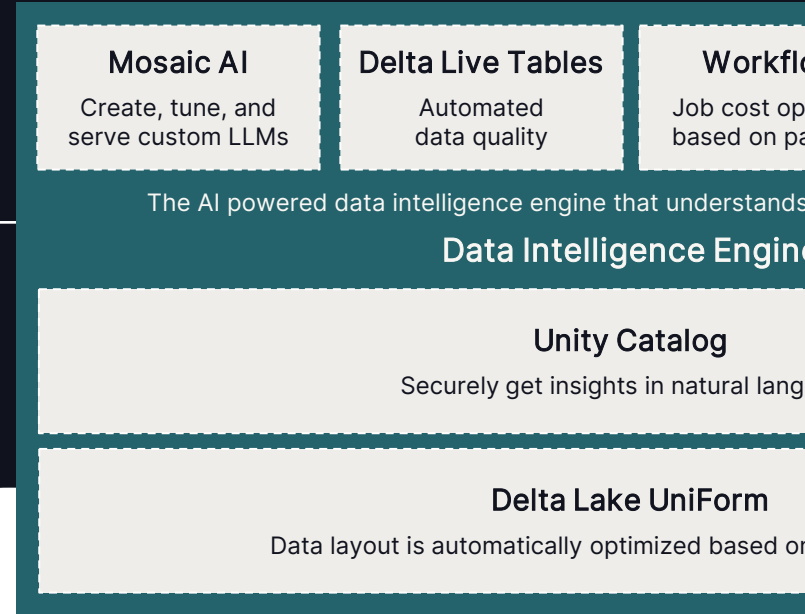
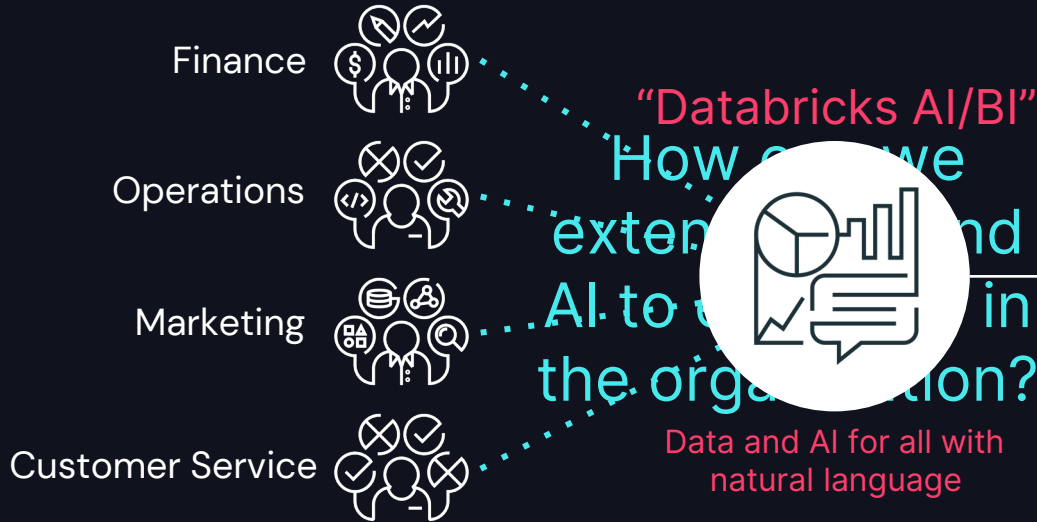
The screenshot shows the Databricks SQL interface. The title bar reads "Spotify Artists Popularity" and "SQL". The menu bar includes "File", "Edit", "View", "Run", and "Help". The main editor area contains a SQL query:

```
1 --sql
2 SELECT released_month, 'artist(s)_name', SUM(streams) as total_streams
3 FROM spotify_songs_2023
4 WHERE released_month IN (1, 5) -- Select only January and May
5 GROUP BY released_month, 'artist(s)_name'
6 ORDER BY released_month, total_streams DESC
```

# Boost productivity for business teams



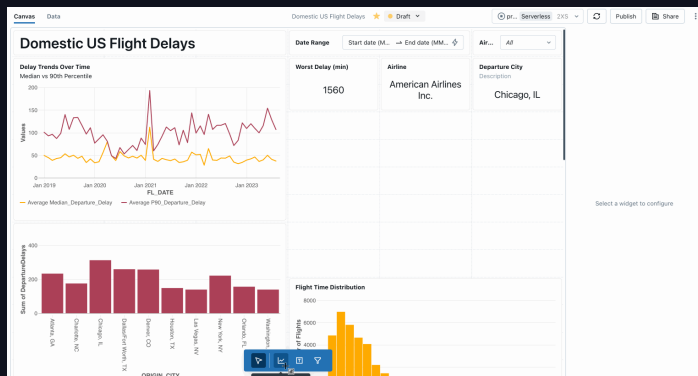
# Databricks Data Intelligence Platform



**Open Data Lake**  
All Raw Data  
(Logs, Texts, Audio, Video, Images)

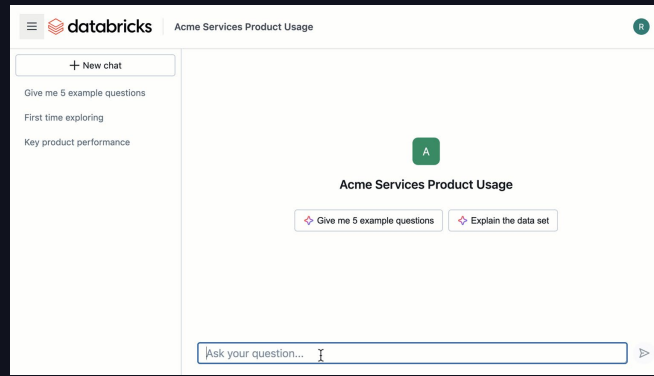
# Databricks AI/BI

Intelligent analytics for real-world data



Dashboards

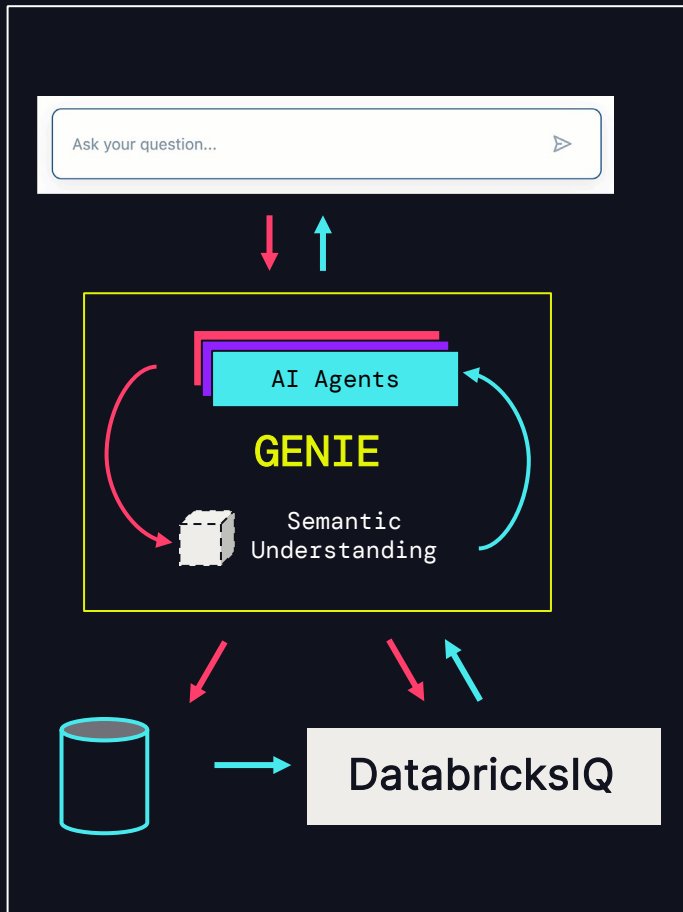
Generally Available



Genie

Public Preview





# Compound AI System that continuously learns your unique data and semantics

Ensemble of specialized AI agents working in concert

Leverage tools and context

- Unity Catalog metadata
- Query history (all workloads)
- Related assets (Notebooks, Dashboards, Queries)

Seeks clarifications when needed

Remembers and learns

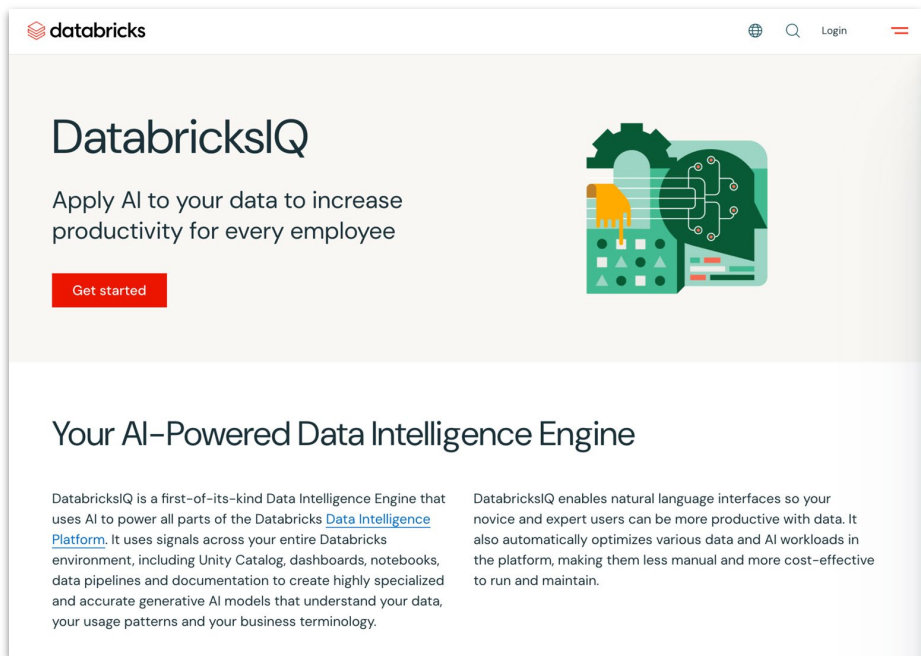
# Demo

Want to learn  
more about  
DatabricksIQ?

# Visit

[www.databricks.com/product/databricksIQ](https://www.databricks.com/product/databricksIQ)

And read the blogs (links on web page)



**DatabricksIQ**

Apply AI to your data to increase productivity for every employee

[Get started](#)

## Your AI-Powered Data Intelligence Engine

DatabricksIQ is a first-of-its-kind Data Intelligence Engine that uses AI to power all parts of the Databricks [Data Intelligence Platform](#). It uses signals across your entire Databricks environment, including Unity Catalog, dashboards, notebooks, data pipelines and documentation to create highly specialized and accurate generative AI models that understand your data, your usage patterns and your business terminology.

DatabricksIQ enables natural language interfaces so your novice and expert users can be more productive with data. It also automatically optimizes various data and AI workloads in the platform, making them less manual and more cost-effective to run and maintain.

## Data Intelligence Platforms

Our viewpoint on how AI will fundamentally change data platforms and how data will change enterprise AI

by [Michael Armbrust](#), [Adam Conway](#), [Ali Ghodsi](#), [Naveen Rao](#), [Arsalan Tavakoli-Shiraji](#), [Patrick Wendell](#), [Reynold Xin](#) and [Matei Zaharia](#)

November 15, 2023 in [Platform Blog](#)

### Adding Intelligence to Databricks Search

New intelligence features enhance Databricks Search, simplifying the discovery of all the assets you need for your data and AI projects.

by [Xiaolei Shi](#), [Joy Xie](#), [Manzar Kazj](#), [Chi Zhang](#), [Ritendra Datta](#), [Richard Tomlinson](#) and [Justin Kim](#)

March 4, 2024 in [Platform Blog](#)

### How we improved DatabricksIQ LLM quality for AI-generated table comments

Swapping in SOTA open-source models and automating quality evaluation with minimal effort

# Enabling DatabricksIQ-Powered Tools

## Partner-powered AI assistive features

Enables access to product features, such as the Databricks Assistant, that use model partner providers for this workspace. Model partners do not access or retain your data.

Default (enabled) ▾ Save

Ensure your Admins  
have enabled DBIQ  
features

## Previews

Manage previews. Learn more about use of [previews](#). More previews will continue to be added to this page. Preview enablement or disablement changes take effect in a few minutes.

### Databricks Assistant Autocomplete Public Preview

Allow workspace users to see real-time AI-based autocomplete suggestions

On

[Documentation](#) | [Provide feedback](#)

### Genie Spaces Public Preview

No code experience for business users powered by LLMs. Analysts set up spaces which business users can use to ask questions. Requires Unity Catalog and the Databricks Assistant.

On

[Provide feedback](#)

# Thanks !

