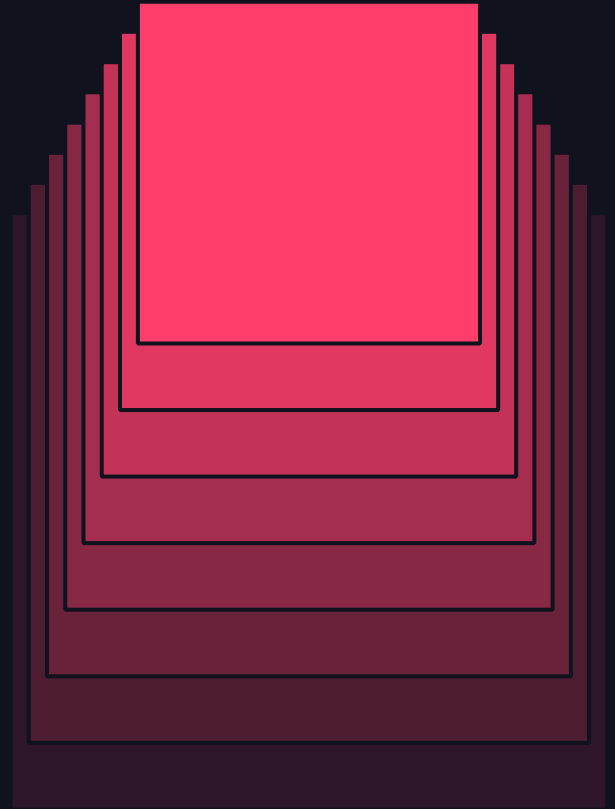


AI/BI Genie - Intro and Best Practices

Chao Cai, Sr. Director of Engineering
Hanlin Sun, Associate Product Manager



Product safe harbor statement

This information is provided to outline Databricks' general product direction and is for **informational purposes only**. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward-looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all

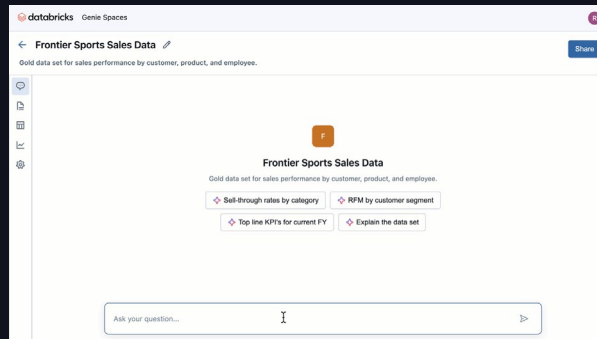
Databricks

AI/BI
Intelligent analytics
for real-world data

Generally Available
Dashboards



Public Preview
Genie



Governed and Secured with Unity Catalog



Introducing AI/BI Genie

Now in Public Preview

The screenshot displays the Databricks AI/BI Genie interface. At the top, it shows the Databricks logo and 'Genie Spaces'. The main heading is 'Frontier Sports Sales Data' with a 'Share' button. Below this, a description reads: 'Gold data set for sales performance by customer, product, and employee.' A central orange square with the letter 'F' is positioned above the data set title. Below the title, the same description is repeated. Four suggested queries are listed in rounded rectangular buttons: 'Sell-through rates by category', 'RFM by customer segment', 'Top line KPI's for current FY', and 'Explain the data set'. At the bottom, there is a text input field with the placeholder 'Ask your question...' and a search icon.

Natural Language Analytics

Direct Guidance/Control

Learns Over Time



Agenda

- Motivation for Genie
- Genie Fundamentals
 - How to Set Up Genie
 - End User Experience
 - How Genie Learns Over Time
- Live Demo

Motivation for Genie



Our mission is to democratize data + AI

for everyone



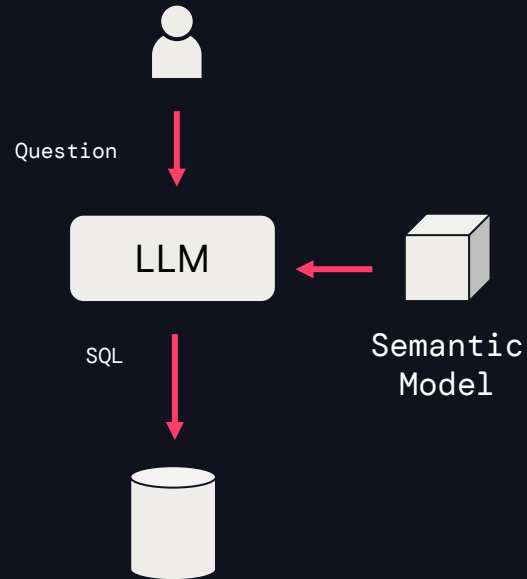
The Challenge with BI

The screenshot shows the Databricks web interface. At the top, there is a search bar with the text "Search data, notebooks, recents, and more...". Below the search bar, the main content area displays a dashboard titled "1. DSQL Metrics (SE, #Customers, #Users)". The dashboard contains two large charts: a stacked area chart on the left and a line chart on the right. The left chart shows three stacked areas in blue, yellow, and green, representing different metrics over time. The right chart shows three lines in blue, yellow, and green, also representing metrics over time. The interface includes a sidebar on the left with various navigation icons and a top navigation bar with the Microsoft Azure and Databricks logos.

A vertical strip of various data visualizations, including line charts, bar charts, and tables, arranged in a grid-like fashion. A red arrow points to the top right corner of this strip.

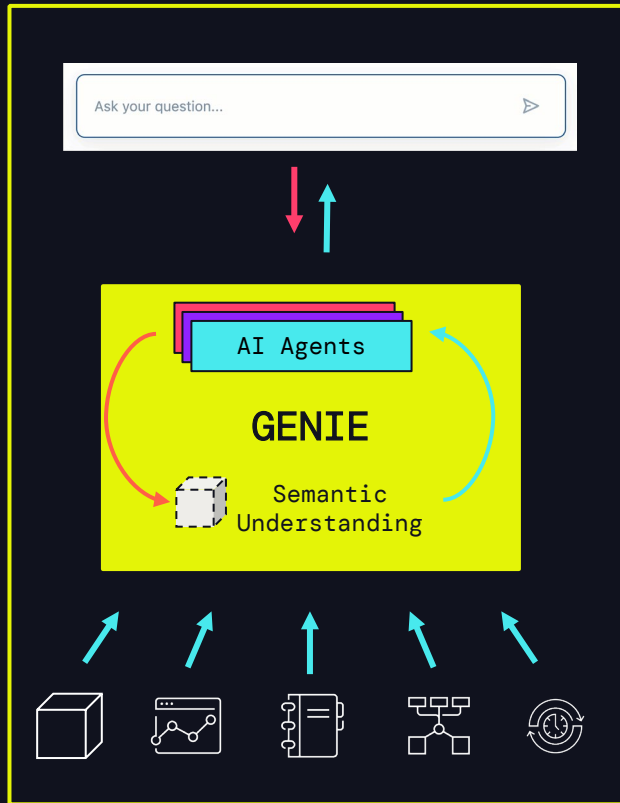


Conventional Text2SQL



Pre-modeled
semantics

We designed Genie to...



1. Continuously learn your data + semantics

1. Include controls for guidance + oversight

1. Be easy and secure for business users

"AI/BI Genie is remarkable in its ability to *understand the concepts and relationships* that tie our operational and engagement data in ways that are *unique to Accolade*.

Our product managers, account managers and power users can now *self-serve questions unanswered by dashboards...*"

-Kapil Ashar,
VP of Enterprise Data and Clinical Platforms



Genie Fundamentals



Genie Availability & Access

- Available in **ungated Public Preview** on AWS and Azure (GCP coming soon!)
 - Workspace admins can enable under Previews
 - Available on all non-GFM regions
- Data used by Genie must be in Unity Catalog
- Serverless or Pro SQL Warehouses required
- Accessible for users with SQL entitlement
- **Free outside of warehouse costs!**

Genie Fundamentals

Data teams set up topic specific Genie spaces

Business users can get answers to net-new data questions

Genie learns over time leveraging Data Intelligence Platform

The screenshot shows the Databricks Genie Spaces interface. At the top, it says 'databricks Genie Spaces' and 'Frontier Sports Sales Data'. Below the title, it states 'Gold data set for sales performance by customer, product, and employee.' A 'Share' button is in the top right. On the left, there is a sidebar with a '+ New chat' button and a list of chat topics: 'How's my business doing?', 'Show revenue by FQ', 'Revenue analysis', and 'First exploration'. The main area features a large orange square with the letter 'F' and the title 'Frontier Sports Sales Data'. Below the title, it repeats 'Gold data set for sales performance by customer, product, and employee.' There are four buttons with diamond icons: 'RFM by customer segment', 'Top line KPI's for current FY', 'Sell-through rates by category', and 'Explain the data set'. At the bottom, there is a text input field labeled 'Ask your question...' with a right-pointing arrow.

How to Set Up Genie



Genie Set Up Process



Focused Topic + Data

- Each Genie space should be topic-specific
- Bring in well-documented, clean Unity Catalog tables+metadata



Add Instructions

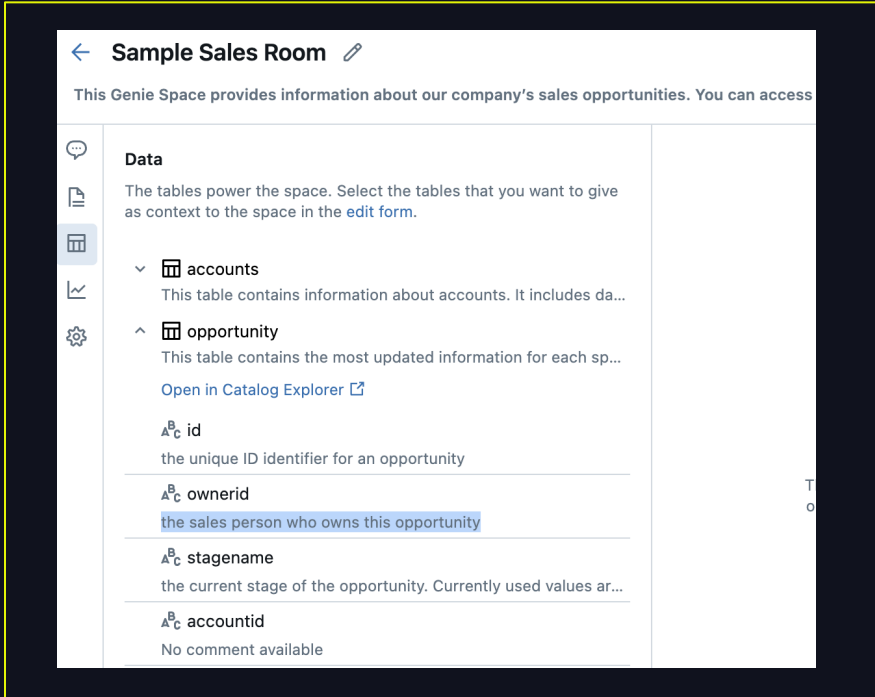
- Authors can guide + teach Genie
- General Instructions and sample SQL statements



Accuracy Validation

- Authors + trusted business users can test expected questions
- Make needed improvements to Instructions

Focused Topic + Data



The screenshot shows the Unity Catalog interface for a 'Sample Sales Room'. The title is 'Sample Sales Room' with an edit icon. Below the title is a description: 'This Genie Space provides information about our company's sales opportunities. You can access...'. On the left is a sidebar with icons for chat, documents, tables, charts, and settings. The main content area is titled 'Data' and contains the following text: 'The tables power the space. Select the tables that you want to give as context to the space in the edit form.' Below this are two table entries:

- accounts**: This table contains information about accounts. It includes da...
- opportunity**: This table contains the most updated information for each sp...
[Open in Catalog Explorer](#)

Under the 'opportunity' table, several columns are listed:

- id**: the unique ID identifier for an opportunity
- ownerid**: the sales person who owns this opportunity
- stagename**: the current stage of the opportunity. Currently used values ar...
- accountid**: No comment available

- Genie works on all types of data topics
- Bring in a focused set of tables
- Unity Catalog tables/views need to be well documented
- SUGGESTED: Create views to remove excess columns and clean data

General Instructions

- Gives data teams the power to give direct guidance
- Define your unique jargon, logic, concepts, KPIs, etc.
- Instructions are iterable over time

Assistant instructions

Add general instructions and example queries on how you want the space assistant to behave.

General Instructions

- win/won opportunities refer to the stagename value "5. Closed Won"
- lost opportunities refer to the stagename value "X. Closed Lost"
- An open opportunity is defined as not being "5. Closed Won" or "X. Closed Lost"
- Sales are calculated only on the opportunities where the stagename is a win/won opportunity. Sales people performance is only evaluated on their won opportunities.
- When users ask for sales reps, they want the name of the sales rep not just the ID
- Expected value of an opportunity is calculated by multiplying the probability by the amount for an opportunity.
- When an user asks about past purchases, check the opportunityhistory for that customer and the number of won opportunity updates there are

Join relationships:

- opportunity.id is the primary key, opportunityhistory.opportunityid is the foreign key
- accounts.id is the primary key, opportunity.accountid is the foreign key. A customer's region info is in the accounts table.

Example SQL Queries

Add example queries that the assistant can learn from.

what were the biggest closed deals in 2023 and who were the sales reps for them?

```
SELECT o.name AS deal_name, u.name AS sales_rep_name, o.amount, o.cl...
```

which sales reps converted open opportunities to closed opportunities at the highest percentage?

```
SELECT u.name AS sales_rep_name, COUNT(o.id) AS total_opportunities,...
```

Sample SQL Statements

Hanlin Sample Sales Room

This query retrieves the accounts and opportunities that were closed in January 2023 and had updates in the same month. It specifically looks for opportunities that were closed successfully (stagename = '5. Closed Won') and had updates that were not closed won (oh.stagename != '5. Closed Won'). The results are sorted by the closedate of the opportunities.

	account_name	opportunity_name	closedate
1	Jetpulse	Jetpulse - New Business	2023-01-04
2	Topdrive	Topdrive - New Business	2023-01-05
3	Pixonyx	Pixonyx - New Business	2023-01-05
4	Aimbo	Aimbo - New Business	2023-01-06
5	Twinte	Twinte - New Business	2023-01-06
6	Meevee	Meevee - New Business	2023-01-06
7	Feedbug	Feedbug - New Business	2023-01-07
8	Eabox	Eabox - New Business	2023-01-07
9	Brainlounge	Brainlounge - New Business	2023-01-07
10	Oyundu	Oyundu - New Business	2023-01-07
11	Gabvina	Gabvina - New Business	2023-01-08

95 rows

Save as instruction

```
1 SELECT DISTINCT
2   a.name AS account_name,
3   o.name AS opportunity_name,
4   o.closedate
5 FROM
6   dataroom_sample_views_foobar.shared_views.opportunity o
7 JOIN dataroom_sample_views_foobar.shared_views.opportunityhistory oh ON o.id = oh.o_id
8 JOIN dataroom_sample_views_foobar.shared_views.accounts a ON o.accountid = a.id
9 WHERE
10  o.stagename = '5. Closed Won'
```

- Sample SQL statements directly teach the model how to answer specific questions
- “Save as Instruction” on verified SQL statements teaches Genie for future questions

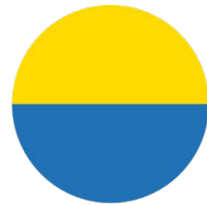
End User Experience



"We recently made AI/BI Genie available to our key stakeholders so that they are able to ask and answer questions about electricity markets and our asset performance in live discussions without being reliant on a data analyst."

- Philipp Cüppers, Team Lead Energy Markets and Asset Optimisation, Vattenfall Hydro Germany

VATTENFALL



Genie <> End Users



Answer net-new questions



Get tabular or visual results



Provide feedback to teach Genie

Monitoring Page

The screenshot shows the Databricks Genie Spaces interface. At the top, it says 'databricks Genie Spaces' and 'Hanlin Sample Sales Room'. Below that, there's a 'Monitoring' section with a sub-header 'Monitor the activity within your space. Understand the errors, feedback, and results the space is returning for your users'. Underneath is a 'Messages feed' with filters for 'Last 30 days', 'Rating', 'User', and 'Status'. The main content is a table with columns: Status, Question, Rating, User, and Created. The table lists 15 entries, all with a status of 'Completed'. The questions range from 'what are all my opportunities that started in december fy 2023?' to 'when was the latest opportunity that was closed won or lost?'. The users listed are 'Hanlin Sun' and 'chao cai@databricks.com'. The page has a pagination control at the bottom showing '50 / page'.

Status	Question	Rating	User	Created
Completed	what are all my opportunities that started in december fy 2023?		Hanlin Sun	2024-05-28 15:4...
Completed	what about just in December FY 2023?		Hanlin Sun	2024-05-28 15:3...
Completed	which sales reps converted open opportunities to closed opportunities at the highest percentage?		Hanlin Sun	2024-05-28 15:3...
Completed	which quarter in 2023 saw the most opportunities go from open to won?		Hanlin Sun	2024-05-28 15:3...
Completed	what were the biggest closed deals in 2023 and who were the sales reps for them?		Hanlin Sun	2024-05-28 15:2...
Completed	what were the biggest closed deals in 2023 and who were the sales reps for them?		Hanlin Sun	2024-05-28 15:1...
Completed	What is the expected value of all open opportunities in Dec 2023?		Hanlin Sun	2024-05-28 15:1...
Completed	wghat is the most recent sales opportunity?		Hanlin Sun	2024-05-28 15:0...
Completed	what is the average?		Hanlin Sun	2024-05-28 14:5...
Completed	what is the smallest?		Hanlin Sun	2024-05-28 14:5...
Completed	what is thee biggest sales opportunity I've ever had?		Hanlin Sun	2024-05-28 14:5...
Completed	What is the expected value of all open opportunities in Dec 2023?		chao cai@databricks.com	2024-05-28 14:1...
Completed	What is the expected value of all open opportunities in Jan 2024?		chao cai@databricks.com	2024-05-28 14:1...
Completed	when was the first opportunity that was not closed won or lost?		chao cai@databricks.com	2024-05-28 13:5...
Completed	when was the latest opportunity that was closed won or lost?		chao cai@databricks.com	2024-05-28 13:5...

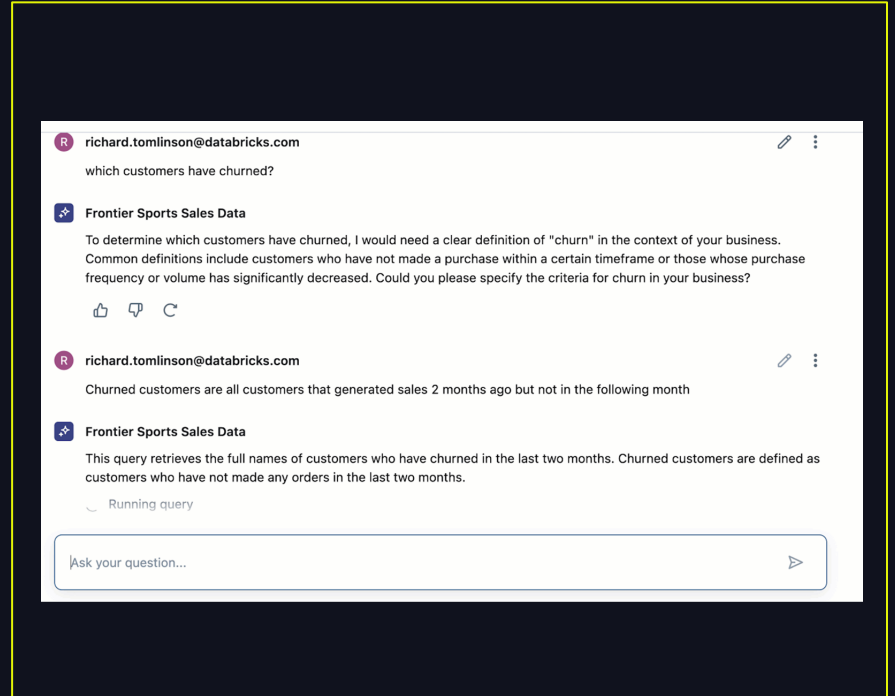
- Proactively understand what your business users are asking
- Step through how Genie behaved in a conversation thread
- Make improvements to your Instructions for future questions

Genie Learns Over Time




Genie Asks Follow Ups

- When a question isn't clear, Genie is designed to ask follow ups/clarifications
- Users can add and save new semantic knowledge for Genie to handle future questions




Data Intelligence Platform

- Unity Catalog table metadata is automatically added to Genie's knowledge base
- Create Genie from Dashboard
- **COMING:** automatically learning from your workspace's Query History, Dashboards, Queries, Notebooks, etc.



Data

The tables power the space. Select the tables that you want to give as context to the space in the edit form.

^  **accounts**
This table contains information about accounts. It includes data on the industry, name, type, r...
[Open in Catalog Explorer](#)

industry
The industry of the account

id
The unique identifier of the account

name
The name of the account

TYPE
The type of account

region_hq__c
The headquarters region of the account

region__c
The region of the account

company_size_segment__c
The size segment of the company (either SMB, MM, or ENT)

1.2 annualrevenue
The annual revenue of the account

DEMO



Advanced Features (Coming Soon)

More ways to iterating and refine Genie for your team's needs

Certified Answers

When you have a **mission critical** question where you want 100% confidence that the answer will always be produced in a specific way, you can add a Certified Answer to prescribe this.

Canonical Metrics

For metrics where rollups are tricky to get right (e.g. **unique users**), we are introducing a way to precisely inform the space to query the right tables in the right ways.

Accuracy Evaluation

Dedicated experience for authors to test questions with Genie before sharing, including curating an ongoing validation question set that can be evaluated against the current set of instructions.

Learn more at the summit!



Databricks
Events App



Tells us what you think

- We kindly request your valuable feedback on this session.
- Please take a moment to rate and share your thoughts about it.
- You can conveniently provide your feedback and rating through the **Mobile App**.



What to do next?

- Discover more related sessions in the mobile app!
- Visit the Demo Booth: Experience innovation firsthand!
- More Activities: Engage and connect further at the Databricks Zone!



Get trained and certified

- Visit the Learning Hub Experience at **Moscone West, 2nd Floor!**
- Take complimentary certification at the event; come by the Certified Lounge
- Visit our Databricks Learning website for more training, courses and workshops! databricks.com/learn



Q&A

