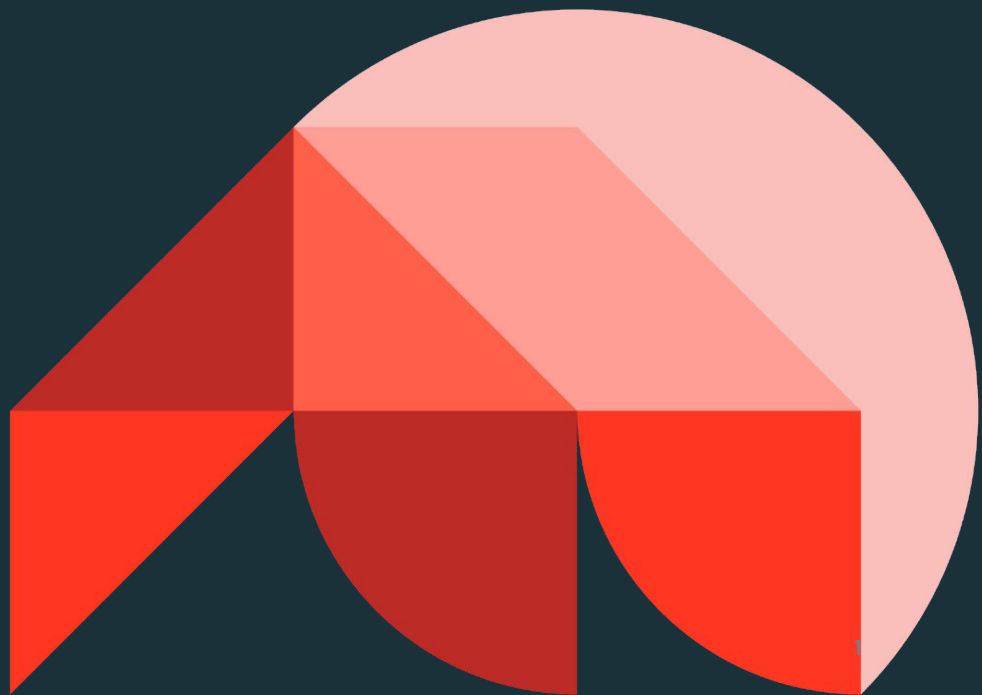


Genie

Best Practices



Databricks Data Intelligence Platform

100% serverless

Disaster recovery

Cost controls

Enterprise security

Mosaic AI
Artificial intelligence



Databricks SQL
Data warehousing



Workflows/DLT
Ingest, ETL, streaming



AI/BI
Business intelligence



Lakehouse

 **Unity Catalog**

 **DELTA LAKE**

ICEBERG

 **Parquet**

The challenge with BI

Traditional reporting



Querying with natural language



For everyone

AI/BI



AI/BI Dashboards

Create and iterate visualizations with AI-assisted authoring

Collaborate and securely share with ease

Frontier Sporting Goods - Sales & Marketing Dashboard

Country:

Territory:

Category:

Model:

Revenue:

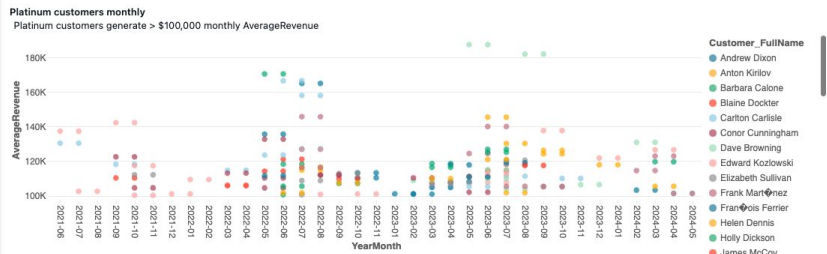
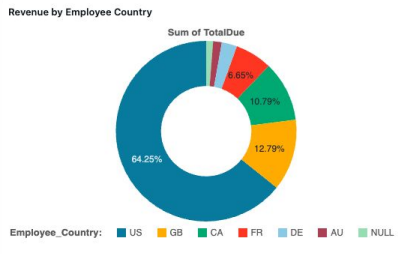
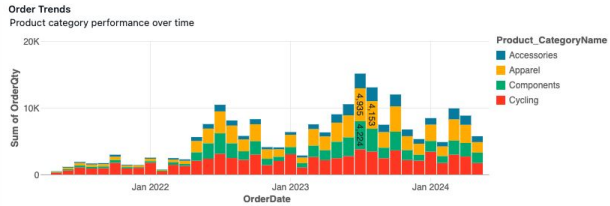
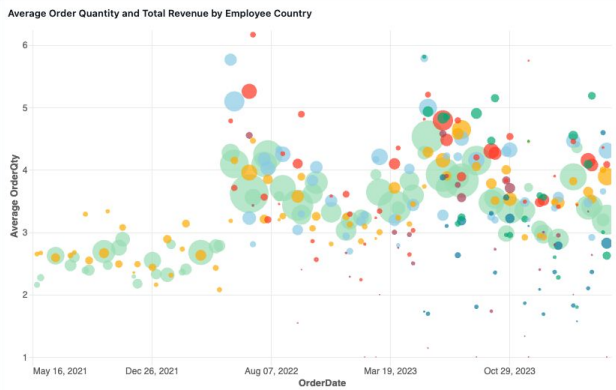
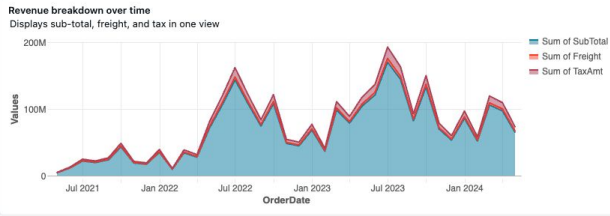
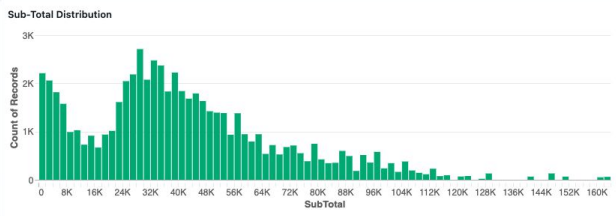
Order Date:

Total Orders: 214.52K

Total Revenue: 2.86B

Std Discount: 0.03

Launch Genie [Ask follow up questions](#)



Genie: Assistant with context knowledge

Allow practitioners to get insights without needing any pre-existing knowledge of your data layer

- **Self-service Q&A**—go beyond dashboards and converse directly with your data
- **AI with Data Intelligence**—accurate answers from AI with unique knowledge of your data that improves through user feedback
- **Governed and secure**—fully governed interactions that adhere to your policies established in Unity Catalog

The screenshot shows the Databricks Genie Spaces interface. At the top, it says 'databricks Genie Spaces' and 'Frontier Sports Sales Data'. Below that, it describes the data set: 'Gold data set for sales performance by customer, product, and employee.' There are four suggested queries: '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 send button.



Genie

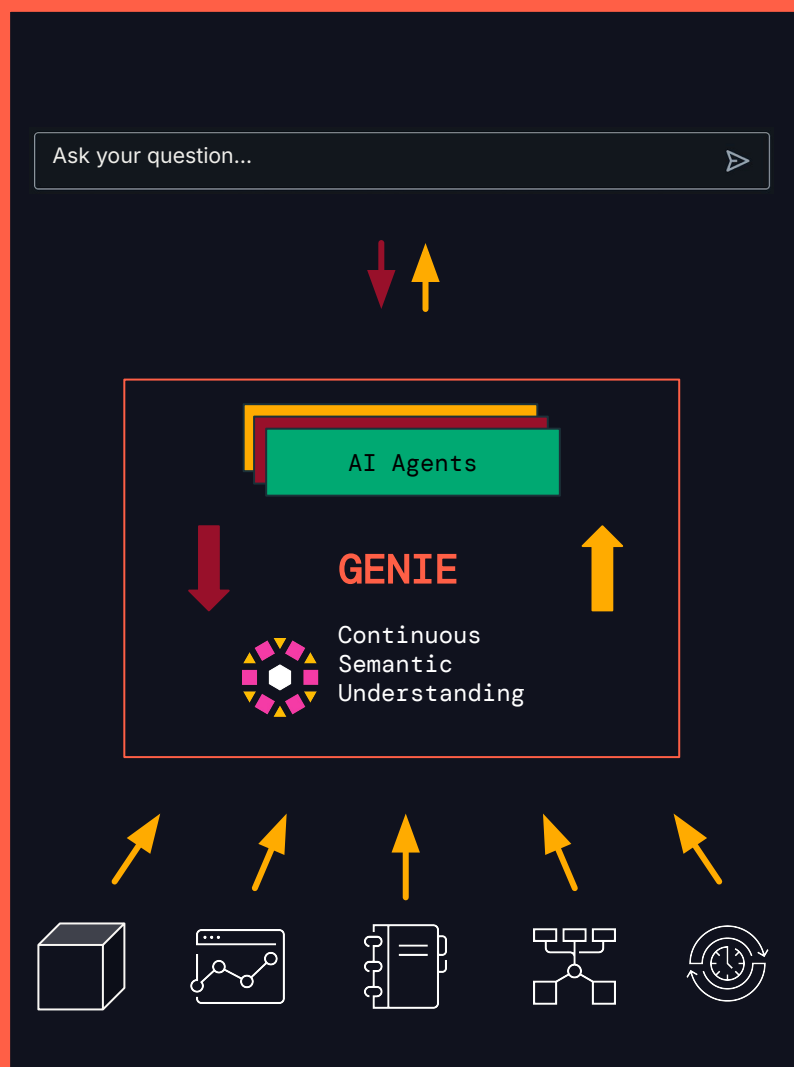
Ensemble of specialized AI agents working in concert

Leverage tools and context

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

Seeks clarifications when needed

Remembers and learns



How to set up Genie

Genie setup process



Focused topic and data

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



Add instructions

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



Accuracy validation

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



Start with Clean Data

Focused set of UC tables and remove unnecessary columns (e.g. Salesforce opportunity schema, Workday HR data, etc)

Add comments that describe your column, including example values and synonyms business users might use

Define PK/FKs in Unity Catalog or pre-join tables into materialized views

The screenshot displays the 'Data' section of the Unity Catalog interface. It lists two tables: 'accounts' and 'opportunity'. The 'opportunity' table is expanded, showing its columns and their descriptions. A tooltip is visible over the 'stagename' column description, providing example values.

Data
The tables power the space. Select the tables that you want to give as context to the space in the [edit form](#).

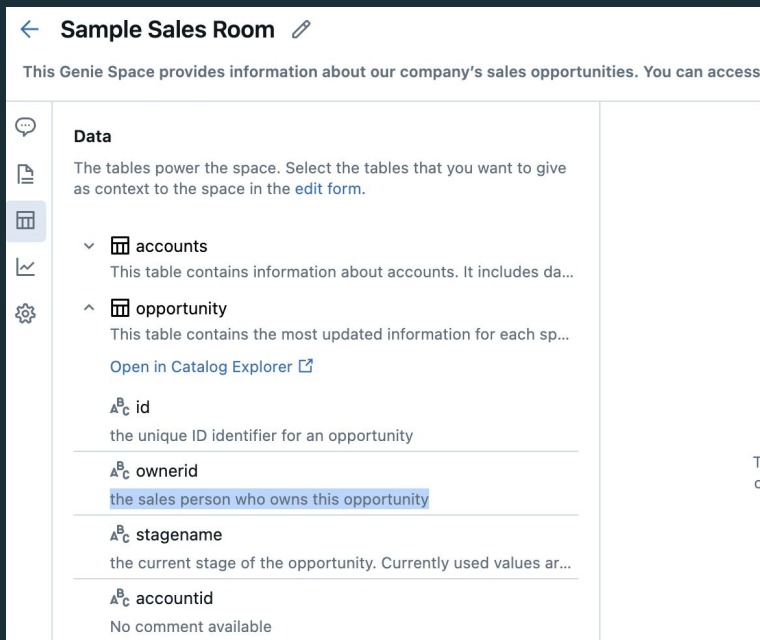
- > **accounts**
This table contains information about accoun...
- ▼ **opportunity**
This table contains the most updated informa...
[Open in Catalog Explorer](#)


id	the unique ID identifier for an opportunity
ownerid	the sales person who owns this opportunity
stagename	the current stage of the opportunity. Currentl...
accountid	No comment available
type	No comment available

the current stage of the opportunity. Currently used values are: "1. Discovery", "2. Demo", "3. Validation", "4. Procure", "5. Closed Won", "X. Closed Lost".

Focused topic + data

Curate UC + Start small




← **Sample Sales Room** 

This Genie Space provides information about our company's sales opportunities. You can access

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

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



Add instructions

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.

Discard

Save

Example SQL Queries

Add example queries that the assistant can learn from.

+ Add example query

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,...
```



Add instructions

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	Gahvine	Gahvine - New Business	2023-01-08

95 rows

Save as instruction

Hide generated code

```
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



Accuracy validation

Monitor page

The screenshot shows the 'Monitoring' section of a Databricks Genie Space. The page title is 'Hanlin Sample Sales Room' with a subtitle 'This room is to analyze Sample CRM data'. Below the title is a 'Messages feed' section with filters for 'Last 30 days', 'Rating', 'User', and 'Status'. The main content is a table of monitoring messages.

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 the 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



Accuracy validation

Benchmarks

The screenshot shows the 'Sample Sales Room' interface in Genie Spaces. It features a 'Benchmarks' section with a '+ Add benchmark' button and a table of benchmark questions and their SQL answers. The table has two columns: 'Question' and 'SQL Answer'. Each row includes a question, a copy icon, a trash icon, and an edit icon. The SQL answers are truncated and include a 'Show more' link.

Question	SQL Answer
Who are my best 5 salespeople by closed sales?	<pre>SELECT u.name AS sales_rep_name, ROUND(SUM(o.amount), 2) AS total_sales FROM dataroom_sample_views_foobar.shared_views.opportunity o JOIN dataroom_sample_views_foobar.shared_views.user u ON _</pre> Show more
Who is my best rep in asia by sales?	<pre>SELECT u.name AS sales_rep_name, ROUND(SUM(o.amount), 2) AS total_sales FROM dataroom_sample_views_foobar.shared_views.opportunity o JOIN dataroom_sample_views_foobar.shared_views.user u ON _</pre> Show more
who are my most active sales reps by 2024 by # of won opportunities?	<pre>SELECT u.name AS sales_rep_name, COUNT(o.id) AS won_opportunities FROM dataroom_sample_views_foobar.shared_views.opportunity o JOIN dataroom_sample_views_foobar.shared_views.user u ON _</pre> Show more
Who is my largest customer in europe?	<pre>SELECT a.name AS customer_name, ROUND(SUM(o.amount), 2) AS total_sales FROM dataroom_sample_views_foobar.shared_views.accounts a JOIN dataroom_sample_views_foobar.shared_views.opportunit_</pre> Show more

Build a list of benchmark questions and gold-standard SQL that your Genie space should be able to answer

Test for the different phrasings and versions of questions business users may ask

Track your accuracy over time as you adjust instructions, comments, etc.

Benchmarks

Benchmark results 78% accurate (7/9)

who are my most active sales reps by...
Which business users lost the most...
which mid market accounts have...
Who is my largest customer in europe?
Who are my best 5 salespeople by...
test
Which sales reps had the most lost...
Who is my best rep in asia by sales?
Which accounts have we generated th...

which mid market accounts have generated the most sales globally?

Assessment: Good

Question
which mid market accounts have generated the most sales globally?

Response

Model output

```
SELECT
  a.name AS account_name,
  ROUND(SUM(o.amount), 2) AS total_sales
FROM
  dataroom_sample_views.foobar.shared_views.accounts a
  JOIN dataroom_sample_views.foobar.shared_views.oppportunity
WHERE
  o.stagename = 'S_Closed Won'
  AND a.company_size_segment_c = 'MM'
GROUP BY
  a.name
ORDER BY
  total_sales DESC
LIMIT
  5
```

	a.name account_name	1.2 total_sales
1	Mynte	145991.19
2	Meeveo	128001.64
3	Quiltb	121003.31

Ground truth

```
SELECT
  a.name AS account_name,
  ROUND(SUM(o.amount), 2) AS total_sales
FROM
  dataroom_sample_views.foobar.shared_views.accounts a
  JOIN dataroom_sample_views.foobar.shared_views.oppportunity
WHERE
  o.stagename = 'S_Closed Won'
  AND a.company_size_segment_c = 'MM'
GROUP BY
  a.name
ORDER BY
  total_sales DESC
LIMIT
  5
```

	a.name account_name	1.2 total_sales
1	Mynte	145991.19
2	Meeveo	128001.64
3	Quiltb	121003.31

Benchmark Run

Genie spaces > **Sample Sales Room**

This Genie Space provides information about our company's sales opportunities. You can access aggregated data insights on sales opportunities by sales reps, regions, and sales...

Benchmarks

Evaluations Questions

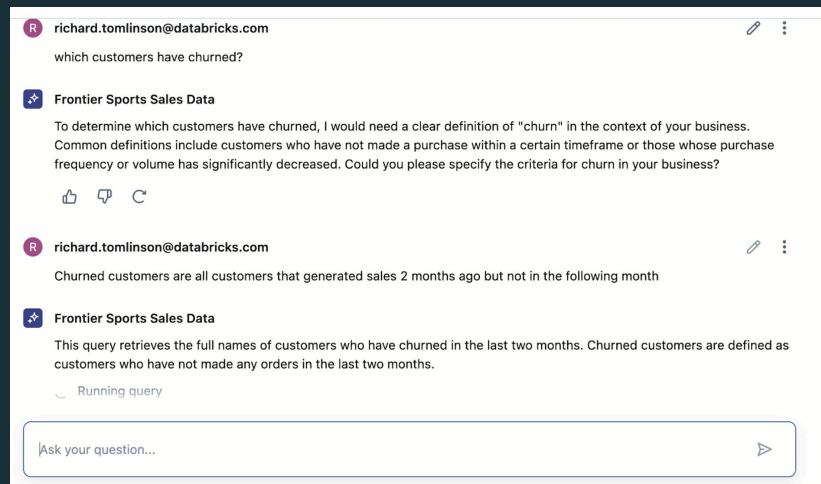
Evaluation name	Execution status	Accuracy
2024-08-29 14:16:21	Completed	100% (9/9)
2024-08-22 10:37:34	Completed	78% (7/9)
2024-08-14 16:26:50	Completed	89% (8/9)
2024-08-13 15:36:47	Paused	--
2024-08-08 14:54:01	Manual review needed (3)	--
2024-08-08 14:44:28	Completed	83% (5/6)
2024-08-06 10:36:59	Manual review needed (1)	--
2024-08-05 15:21:18	Completed	83% (5/6)
2024-07-31 22:32:48	Completed	83% (5/6)
2024-07-31 22:30:20	Completed	67% (4/6)

Benchmark History

Accuracy validation

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



The screenshot displays a chat conversation within a white interface. At the top, a user with the email 'richard.tomlinson@databricks.com' asks, 'which customers have churned?'. Below this, a system message from 'Frontier Sports Sales Data' explains that to determine churned customers, a clear definition of 'churn' is needed, providing common definitions and asking for clarification. The user then responds with, 'Churned customers are all customers that generated sales 2 months ago but not in the following month'. A second system message from 'Frontier Sports Sales Data' states that the query will retrieve the full names of customers who have churned in the last two months based on the provided definition. At the bottom, there is a text input field with the placeholder 'Ask your question...' and a send button.

Troubleshooting Cheat Sheet

- Genie isn't joining my tables correctly!
 - Pre-join your tables into a materialized view or define PK / FK relationships in UC.
Alternatively add instructions on how to relate your tables.
- Genie isn't using the right column values in my WHERE clause!
 - Enumerate them in column comments if your column is categorical
- Genie is "ignoring" my general Instructions!
 - Try adding example SQL statements and removing unnecessary Instructions that could be confusing Genie
- I'm hitting the token limit!
 - Remove unnecessary columns, streamline column descriptions, prune example SQL
- Detailed troubleshooting guide [linked here](#)
- NOTE: we are working to make it easier to deal with each of these challenges as well!

Ask For Review

ties. You can access aggregated data insights on sales opportunities by sales reps, regions, and sales targets, and...

H Hanlin Sun

Biggest open opportunities?

Genie

This query retrieves the names of the biggest open opportunities, along with their amounts, the associated account names, and the names of the sales representatives handling them. The results are sorted in descending order based on the opportunity amount and limited to the top 5 opportunities.

	🔗 opportunity_name	1.2 amount	🔗 account_name	🔗 sales_rep_name
1	Yata - New Business	12002.27	Yata	Orin Gramer
2	Tekfly - New Business	12001.76	Tekfly	Jeremias Ropars
3	Skalith - New Business	12001.55	Skalith	Consalve Fosdick
4	Realpoint - New Business	12001.53	Realpoint	Gideon Malletratt
5	Fiveclub - New Business	12001.44	Fiveclub	Donnajean Meak

5 rows



Add as instruction



> Show generated code

Request a review from the room admin

Does the answer look right to you?

Cancel

Send for review

For questions that business users aren't sure about, they can mark it for review by data teams

Data teams can then edit SQL or mark it as correct, and the business user will then be notified



What questions can I ask?

Ask precise questions based on the available data

✓ Questions Genie can answer

- What were the sales this quarter?
- Which customers generated the most revenue?
- Which sales representatives generated the most revenue?
- What were the top-performing marketing campaigns in 2024?

✗ Questions Genie cannot answer

- Why were sales low this quarter?
- How can I get more revenue from Customer M?
- What incentives should I offer my sales team to generate more revenue?
- How can I improve my marketing strategy?



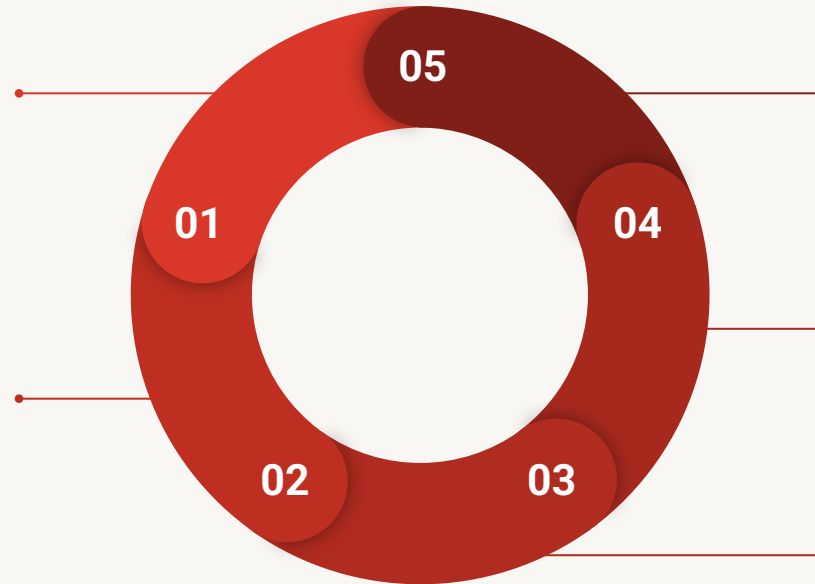
Best Practices for Genie Spaces

Data Preparation

Curate **relevant**, topic-specific **datasets**.
Add detailed **metadata** (descriptions, business context).

Semantic Modeling

Define entity relationships (**primary/foreign keys**)
Add contextual **descriptions** to tables/columns
Use proper data types (e.g., dates, strings)



Monitoring and Feedback

Encourage users to **rate responses** (thumbs-up/down).
Monitor "Messages Feed" for common queries/issues.
Continuously refine based on **feedback**.

Testing and Iteration

Test with **common business questions**.
Review generated SQL queries for accuracy.
Use benchmark questions with expected outputs.

Instructions and Examples

Teach business logic with clear instructions
Add **example SQL queries** for complex logic



Genie setup process



Focused topic and data

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



Add instructions

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



Accuracy validation

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



Appendix





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](#) 

A^BC industry
The industry of the account

A^BC id
The unique identifier of the account

A^BC name
The name of the account

A^BC TYPE
The type of account

A^BC region_hq__c
The headquarters region of the account

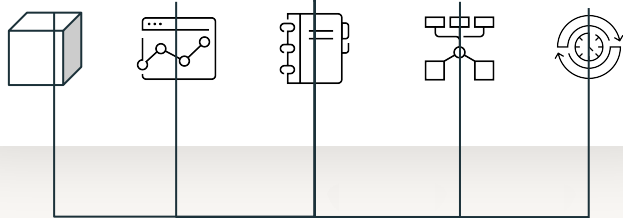
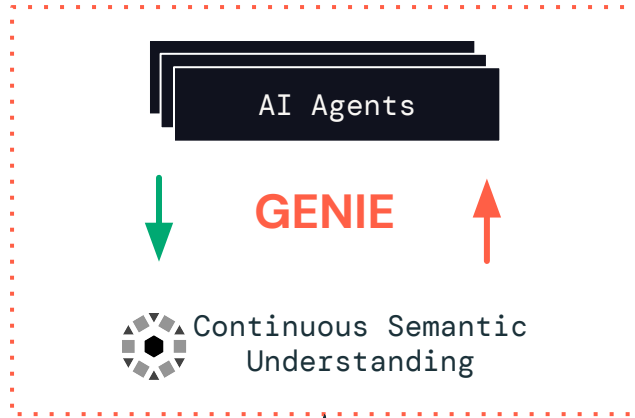
A^BC region__c
The region of the account

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

1.2 annualrevenue
The annual revenue of the account



Ask your question... >



An agent system 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)
- Semantic models when available

Seeks clarifications when needed

Remembers and learns



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

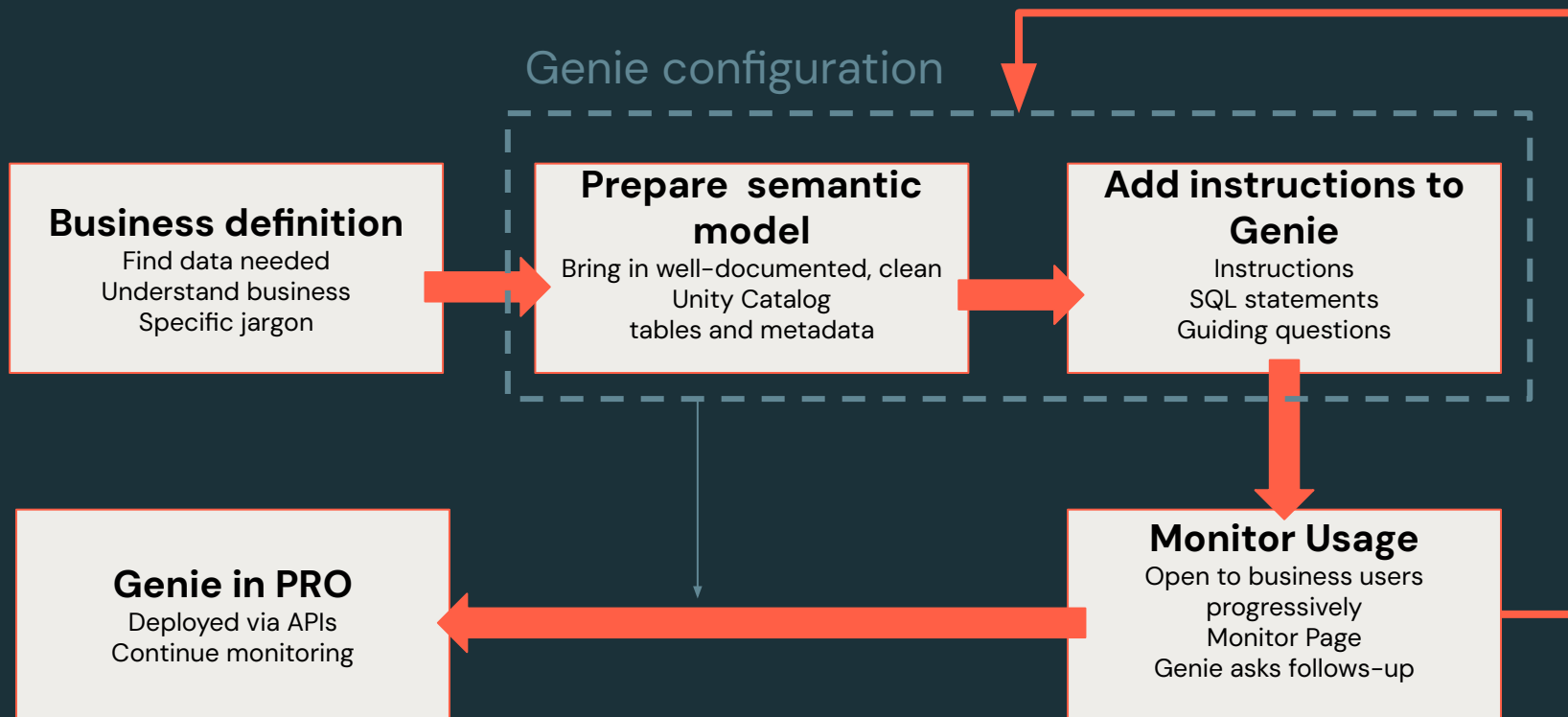
For metrics where rollups are tricky to get right (e.g., unique users), we are introducing a way to precisely inform the space to query the right tables in the right ways

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



Genie from scratch to Production



Useful links

- [Apps with Lakebase](#)
- [Databricks Apps example](#)
- [Agent Bricks Demo](#)
- [Lakeflow Connect Demo](#)
- [Genie Demo](#)

Notebooks:

- [Databricks AI/BI: Supply Chain Optimization](#)
- [Orchestrated and Run DBT jobs](#)

