Creating a Genie space

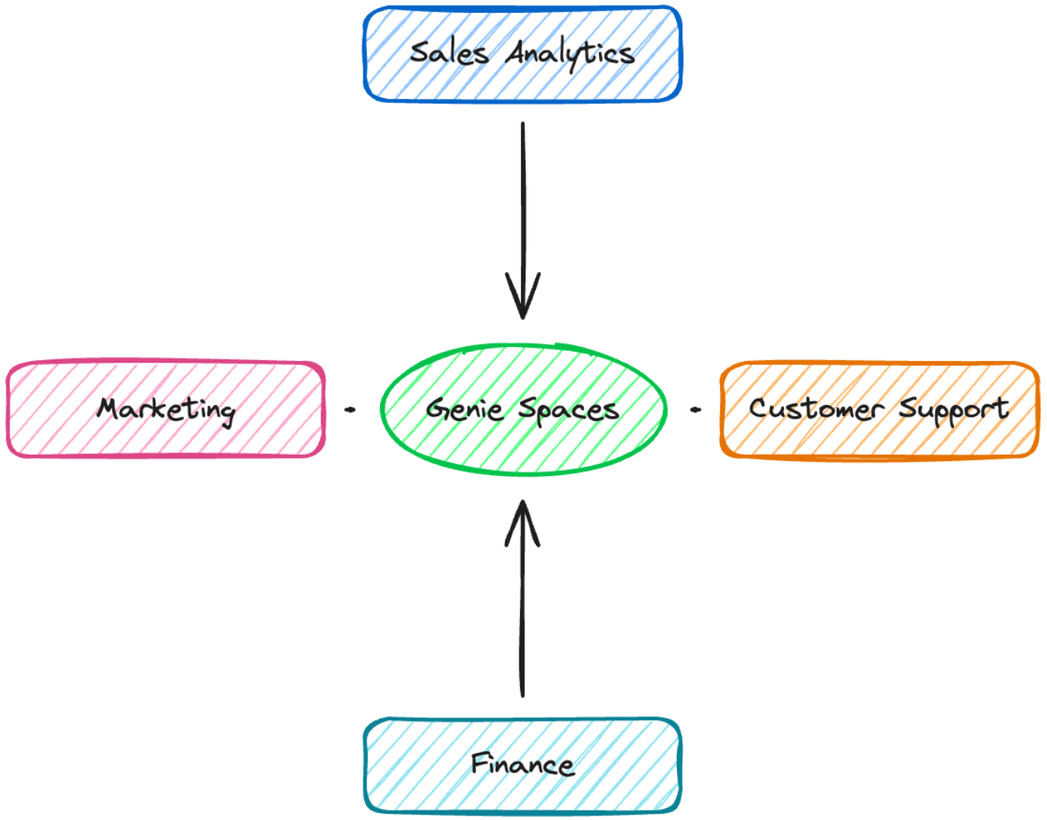
INTRODUCTION TO DATABRICKS GENIE



Gang Wang
Senior Data Scientist

datacamp

When to create a space

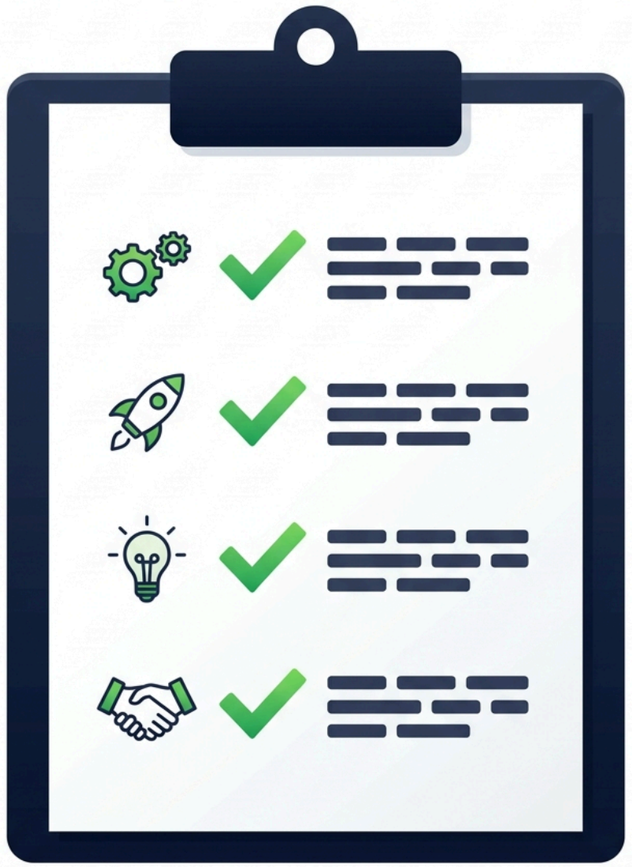


One space per domain

Each Genie space serves a specific business area:

- Franchise Sales Analytics
- Supplier Performance
- Inventory and Operations
- Regional Performance

Prerequisites

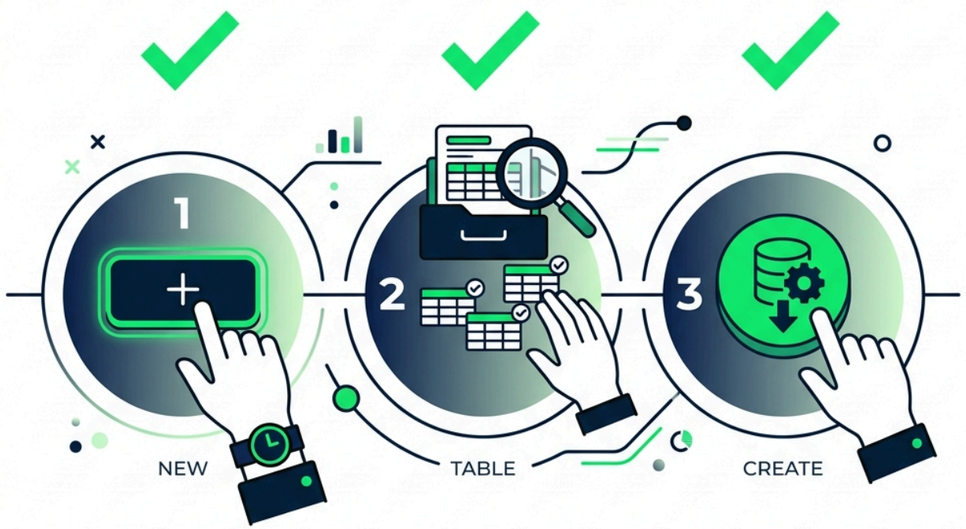


Minimum viable space

- Data registered in **Unity Catalog**
- **Table descriptions** added
- **Column comments** (no cryptic names!)
- An active **SQL Warehouse**

Genie only knows what you invite in - **clear names and descriptions matter**

Space creation in three steps

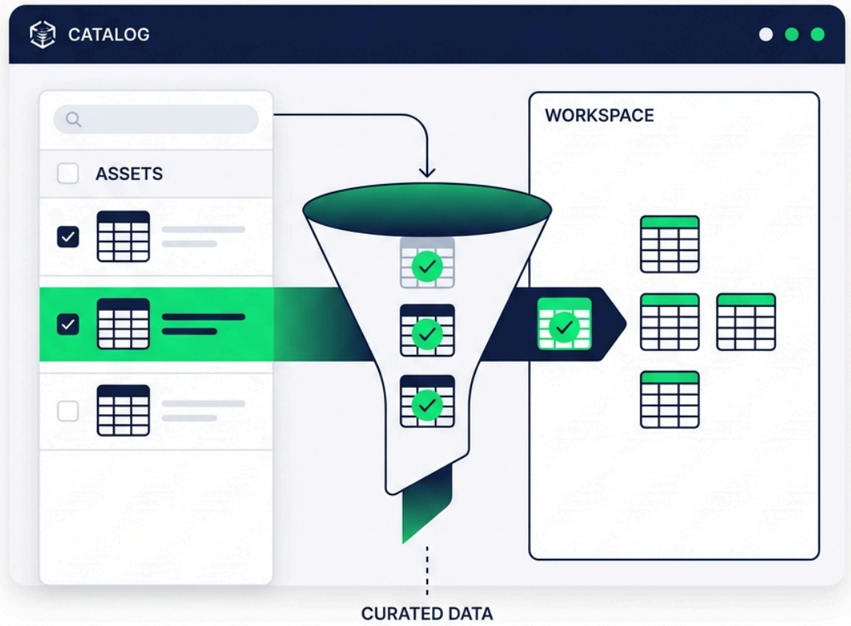


1. Click **Genie** in the sidebar, then **New**
2. Browse **Unity Catalog** and select your data tables
3. Click **Create**

Three components working together:

- **SQL Warehouse** - runs the queries
- **Unity Catalog** - the library of table names and metadata
- **Genie space** - the chat window where it all connects

Example

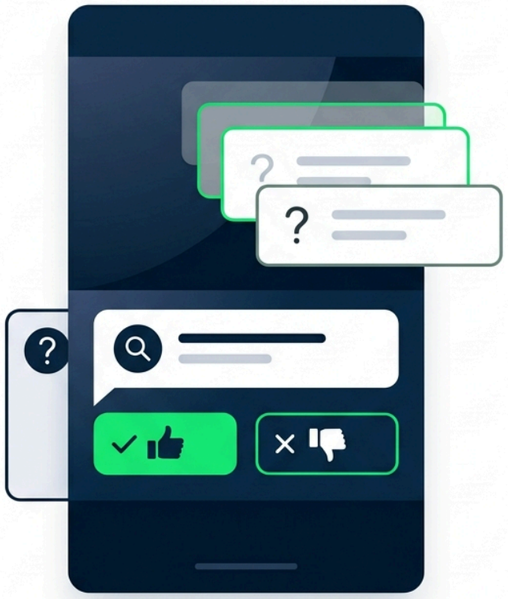


"Bakehouse Franchise Analytics"

- `sales_transactions` - Transaction-level order data
- `sales_suppliers` - Supplier information

Tip: 5 relevant tables > 50 random tables. Too much data leads to hallucinations.

Reviewing suggested queries



- Genie searches for popular workspace queries on your tables
- A notification appears if relevant queries are found
- **Accept** good matches - they become example SQL queries
- **Reject** irrelevant ones

Testing your new space



Verify with simple questions

> "Show me total sales by franchise"

What to check:

- Tables are connected correctly
- Genie understands relationships
- Results make sense

Common mistakes



Avoid these pitfalls:

- **The data dump** - 50+ tables "just in case" leads to ambiguous joins and hallucinations
- **Granularity mismatch** - Mixing daily summaries with transaction-level data causes double-counting
- **The ghost warehouse** - Under-provisioned or aggressive auto-stop causes frustrating delays

Let's practice!

INTRODUCTION TO DATABRICKS GENIE

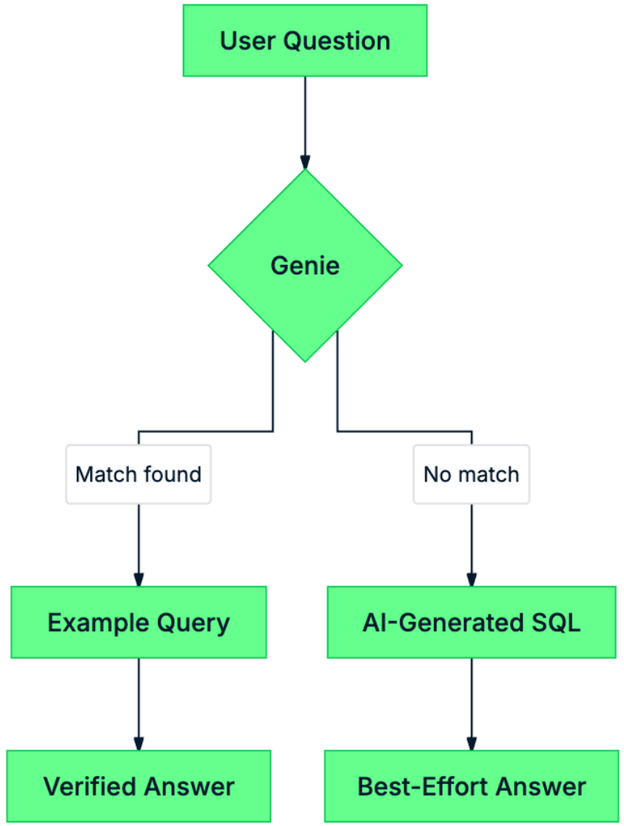
Adding example queries

INTRODUCTION TO DATABRICKS GENIE



Gang Wang
Senior Data Scientist

Why example queries matter



Example queries are pre-written SQL

Add examples when:

- A calculation is non-standard (e.g., Gross Margin)
- Logic involves multiple joins across tables
- Business-specific filters apply (e.g., "Active User" = `last_login > 30 days`)

Like SOPs

Static vs parameterized examples

Static example

> "Total sales in Q4 2024?"

```
SELECT SUM(amount)
FROM sales_transactions
WHERE quarter = 'Q4 2024'
```

Fixed business fact, one-time reference

Parameterized example

> "Sales in {{city}}"

```
SELECT SUM(amount)
```

Same logic, values change - London, Paris, Tokyo

About Data Instructions

Text | Joins | SQL Expressions | SQL Queries

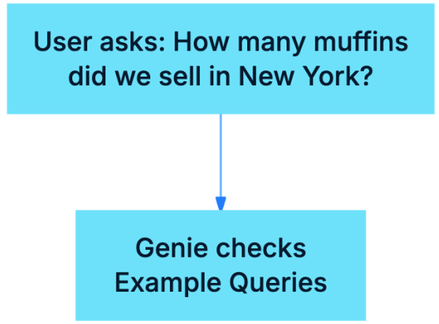
What were total sales last quarter?

```sql
SELECT
  SUM(`totalPrice`) as total_sales,
  COUNT(DISTINCT `transactionID`) as transaction_count,
  COUNT(DISTINCT `customerID`) as unique_customers,
  MIN(`dateTime`) as period_start,
  MAX(`dateTime`) as period_end
FROM
  `samples`.`bakehouse`.`sales_transactions`
```

How Genie routes to examples

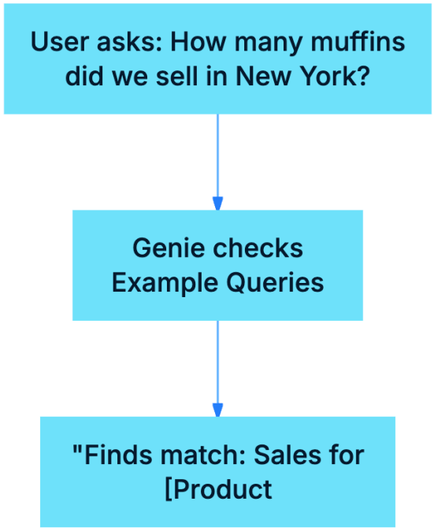
User asks: How many muffins did we sell in New York?

How Genie routes to examples



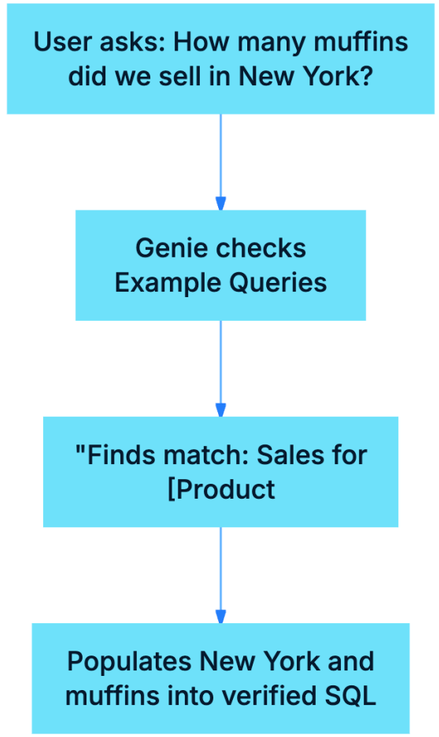
Step 1: Genie checks your Example Queries

How Genie routes to examples



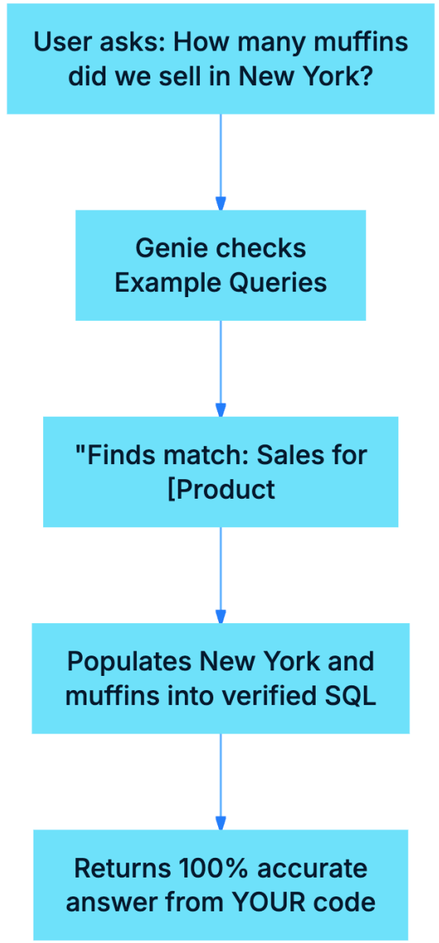
Step 2: Finds match: "Sales for **[Product]** in **[Location]**"

How Genie routes to examples



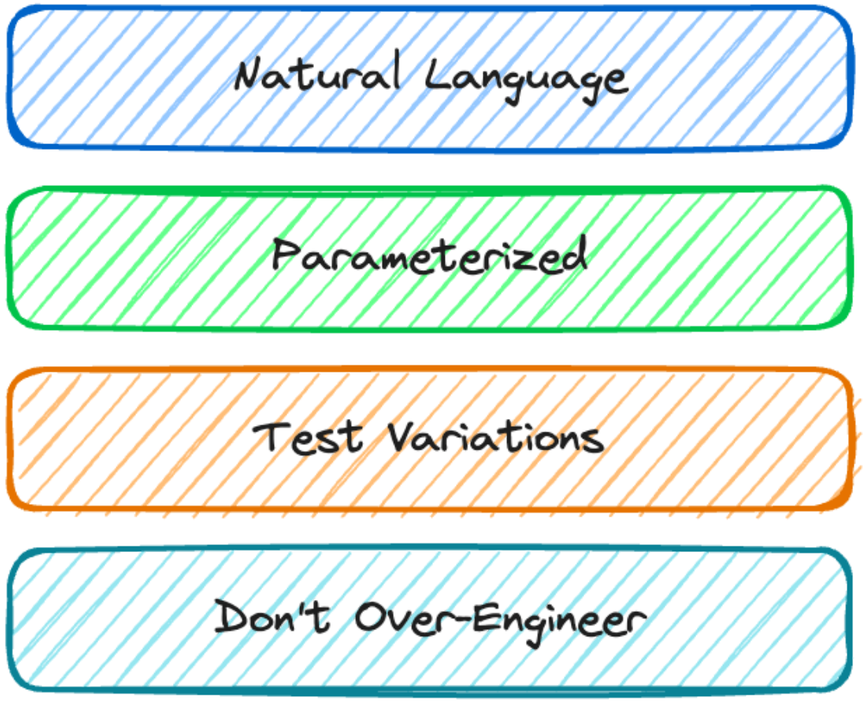
Step 3: Populates "New York" and "muffins" into your verified SQL template

How Genie routes to examples



Result: 100% accurate answer from **YOUR code**, not AI guesswork

Best practices



Good vs. bad examples

Good	Anti-Pattern
Focused, documented, common business terms	`SELECT *` (teaches nothing)
Complex join or calculated field (e.g., Gross Margin)	Easy queries like `SELECT count(*)` (clutters system)
Business rule enforcement (e.g., approved coffee suppliers)	One example per trivial question

Let's practice!

INTRODUCTION TO DATABRICKS GENIE

Guiding users with instructions

INTRODUCTION TO DATABRICKS GENIE



Gang Wang
Senior Data Scientist

datacamp

The power of instructions



Instructions as "Guardrails"

Instructions should:

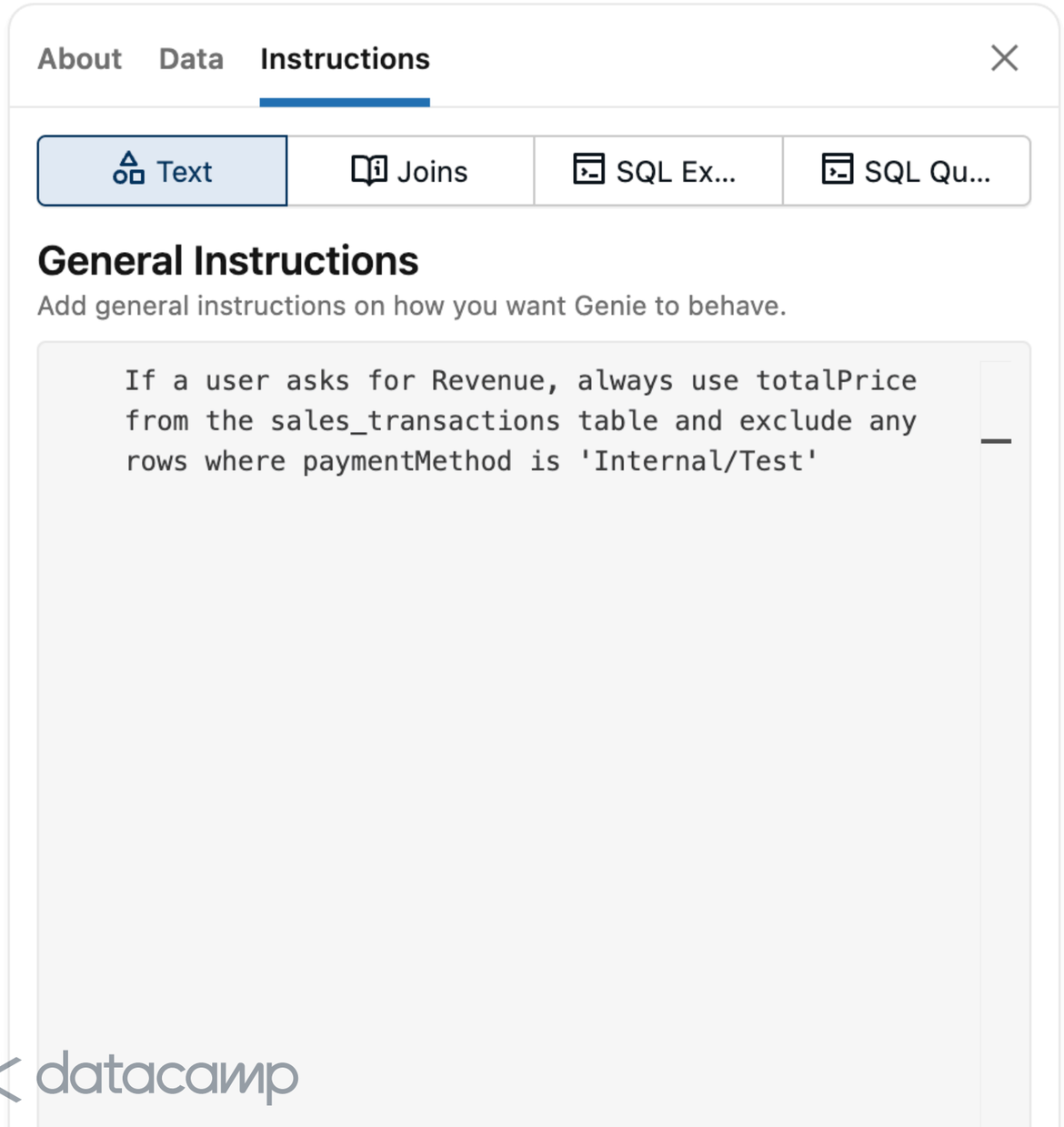
- Explain **how the business uses** the data, not just what it is
- Resolve ambiguity and enforce business logic
- Guide Genie toward consistent, accurate answers

Types of instructions

Four types of effective instructions:

Type	Good	Bad
Metric definitions	"Use `(totalPrice - unitPrice * quantity) / totalPrice` for Gross Margin"	"Calculate profit correctly"
Synonyms	"franchise, store, and shop all refer to `sales_franchises` "	"Refer to the franchise table for store info"
Data quality	"Exclude rows where `paymentMethod` is TEST or VOID"	"Don't show test data"
Formatting	"Round financial figures to two decimal places with $"	"Format numbers nicely"

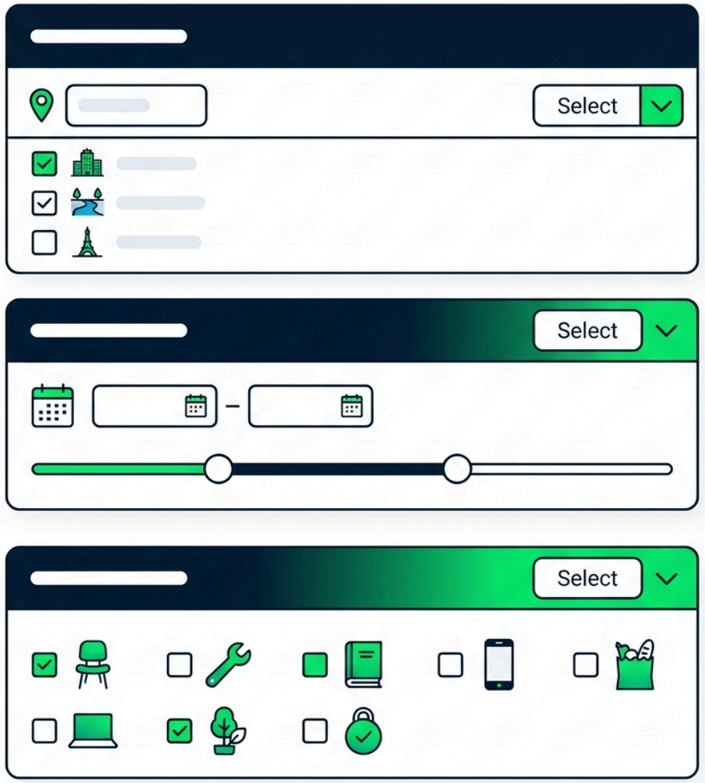
Adding instructions



Adding an instruction:

1. Open **Space settings**
2. Find the **Instructions** section
3. Write clear, specific guidance
4. Genie applies it to **all queries** in the space

Filters for common parameters



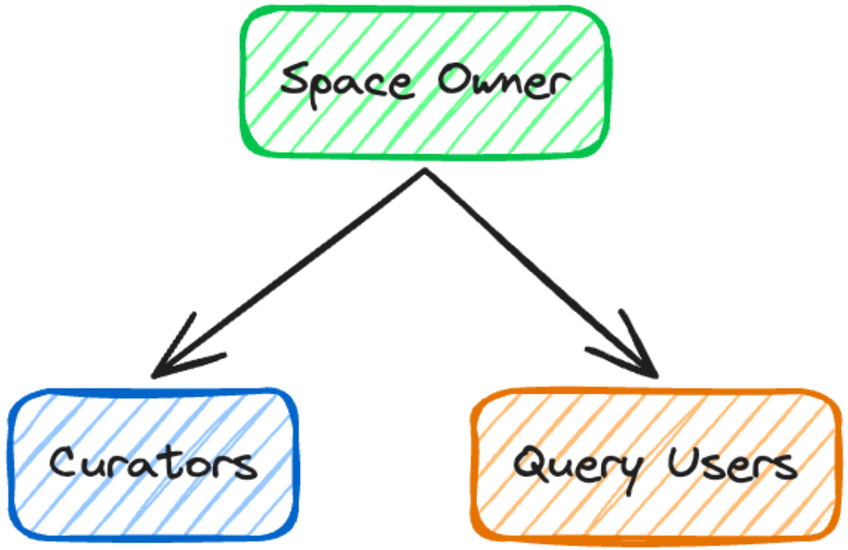
Create a dropdown filter for:

- High-frequency categories
- Fields with complex spellings (City, Date)

Let users type when:

- One-off logic or open-ended comparisons
- e.g., "Compare London to Paris"

Sharing your space



Permission levels

Role	Can Do
Owner	Full control, delete space
Curator	Edit examples, instructions
User	Query only

Tip: Start with query-only access, promote curators as needed

Let's practice!

INTRODUCTION TO DATABRICKS GENIE