Azure Databricks: Building your own AI-Powered Data Assistance

Wilson Mok

Feb 26, 2026

EDS
Edmonton Data Society

Wilson Mok

Director, Data and AI

/wilson-mok

/wilson-mok

@the-analytics-lab

My experiences includes:
- Avanade
- CAE
- Air Canada

Expected Audience and Presentation Structure

✅ What We'll Cover:

– Sign-up for a free Databricks account.

– Prepare the data for our Data Assistant.

– What is a Genie and how it helps users get answers from data using natural language.

– Refine Genie's response to improve the response accuracy of the Data Assistant.

⚙️ Fundamental knowledge: Basic SQL and Python.

– Experienced with Databricks and AI/BI Genie would be beneficial but not required.

Databricks Free Edition: What it is & When to use it?

What it can do

- Sandbox for learning or personal projects

- Same tool as the enterprise version with some limited capabilities

- Easy to start: Sign-up and go.

Link: https://learn.microsoft.com/en-us/azure/databricks/getting-started/free-edition

Limitations

- Limited computing power

- One Databricks workspace per account

- Data is stored in the Databricks default storage

- Inactive accounts for a prolonged period might be deleted

- No security customization

- No SLA (its free!)

To really learn the platform, we need to work with real data

Yelp Open Dataset
A real-world dataset designed for educational use.

yelp * data licensing

6,990,280
Reviews

150,346
Businesses

11
Metropolitan areas

200,100
Pictures

Link: https://business.yelp.com/data/resources/open-dataset/

Demo 1: Preparing data for Data Assistant

Businesses

Reviews

Users

Tips

databricks
FREE EDITION
EXPLORE DATABRICKS
TOOLS FOR FREE

Delta tables

6

Q & A 1

What is Genie and How Does It Work?

Image Credit: https://www.databricks.com/blog/onboarding-your-new-aibi-genie
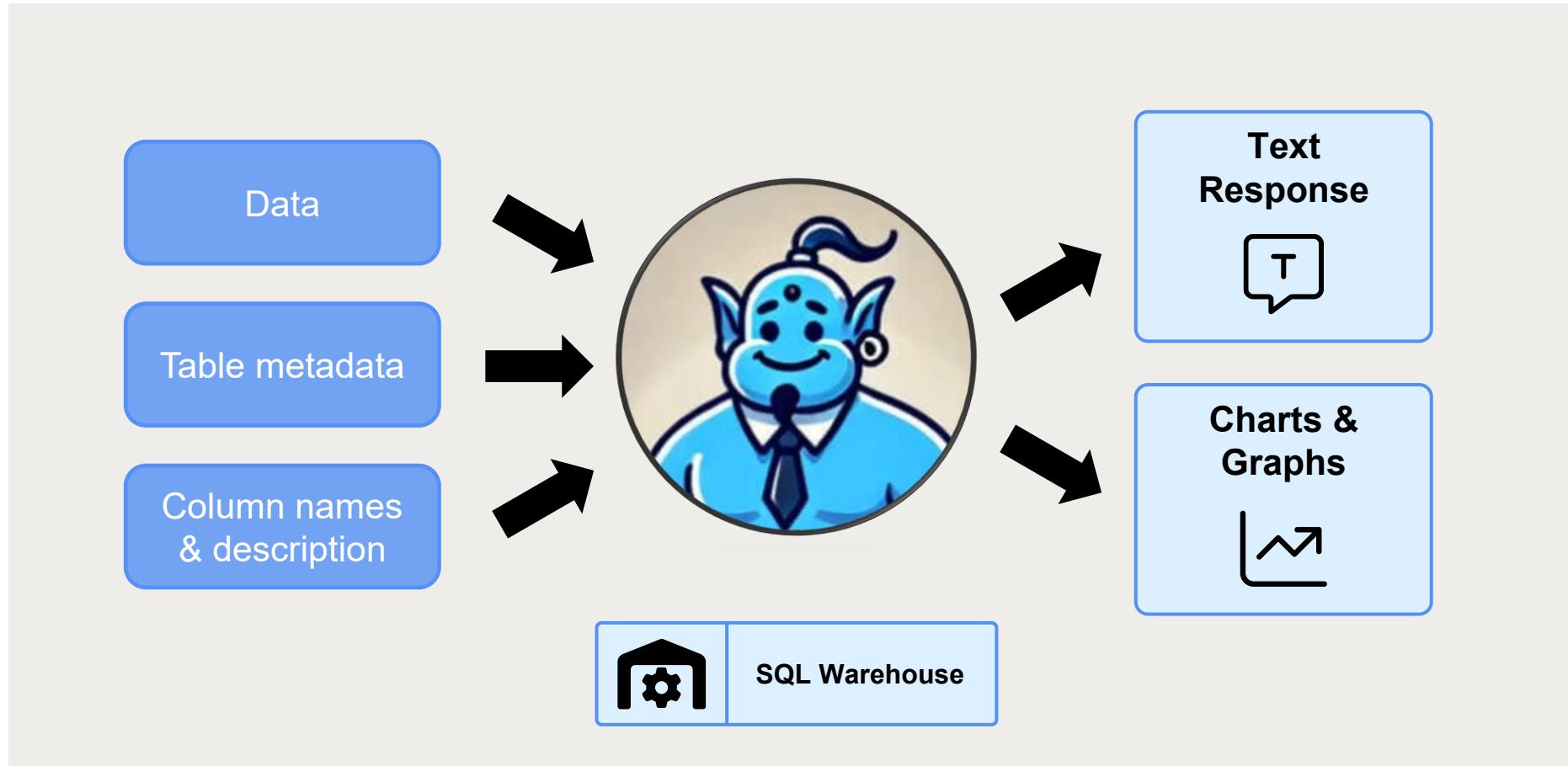
How Do We Make Genie Accurate and Trusted?

Genie Lifecycle

1. Define & Prepare
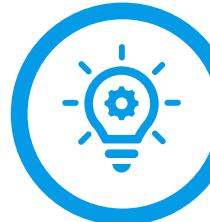
Lay the foundation with business-aligned data

- Define the purpose and audience
- Select the right tables
- Improve Metadata Quality

2. Configure

Tailor Genie's logic, context and trusted assets

- Review Value Dictionaries
- Add Custom Instructions
- Create Trusted Assets

3. Validate

Quality control and validate results

- Improve accuracy through Benchmark testing.
- Validate results to measure accuracy.

4. Monitor & Refine

Monitor user responses and apply feedback

- Monitor user feedback
- Continuous Improvement through refining configurations.

Demo 3: Let's Build One

Genie Lifecycle

1. Define & Prepare

Lay the foundation with business-aligned data

- Define the purpose and audience
- Select the right tables
- Improve Metadata Quality

2. Configure

Tailor Genie's logic, context and trusted assets

- Review Value Dictionaries
- Add Custom Instructions
- Create Trusted Assets

3. Validate

Quality control and validate results

- Improve accuracy through Benchmark testing.
- Validate results to measure accuracy.

4. Monitor & Refine

Monitor user responses and apply feedback

- Monitor user feedback
- Continuous Improvement through refining configurations.

Key Takeaways: Unlocking Unstructured Data for Analytics

- Databricks Free Edition lets you experiment and learn using real-work data and tool.

- Raw data must be structured and cleaned before it becomes usable for analytics or AI.

- Genie translates business questions into SQL (delta tables and metadata).

- Clear metadata, defined scope, and business alignments are key to improve Genie's responses.

- Continuous improvement is key to create and maintain a dependable data assistant.

- The code is in my Github repo:

  - https://github.com/wilson-mok/demo/tree/main/2026/EDS/1-Trusted-Data-Assistant/

Thank you

LinkedIn