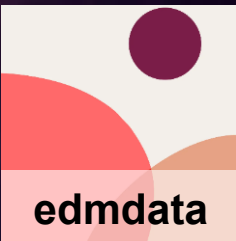


Azure Databricks – Democratizing Data with AI/BI Genie

Wilson Mok

May 12, 2025





Wilson Mok

Sr. Data architect & Consultant

 /wilson-mok

 /wilson-mok

 @the-analytics-lab



My experiences includes:

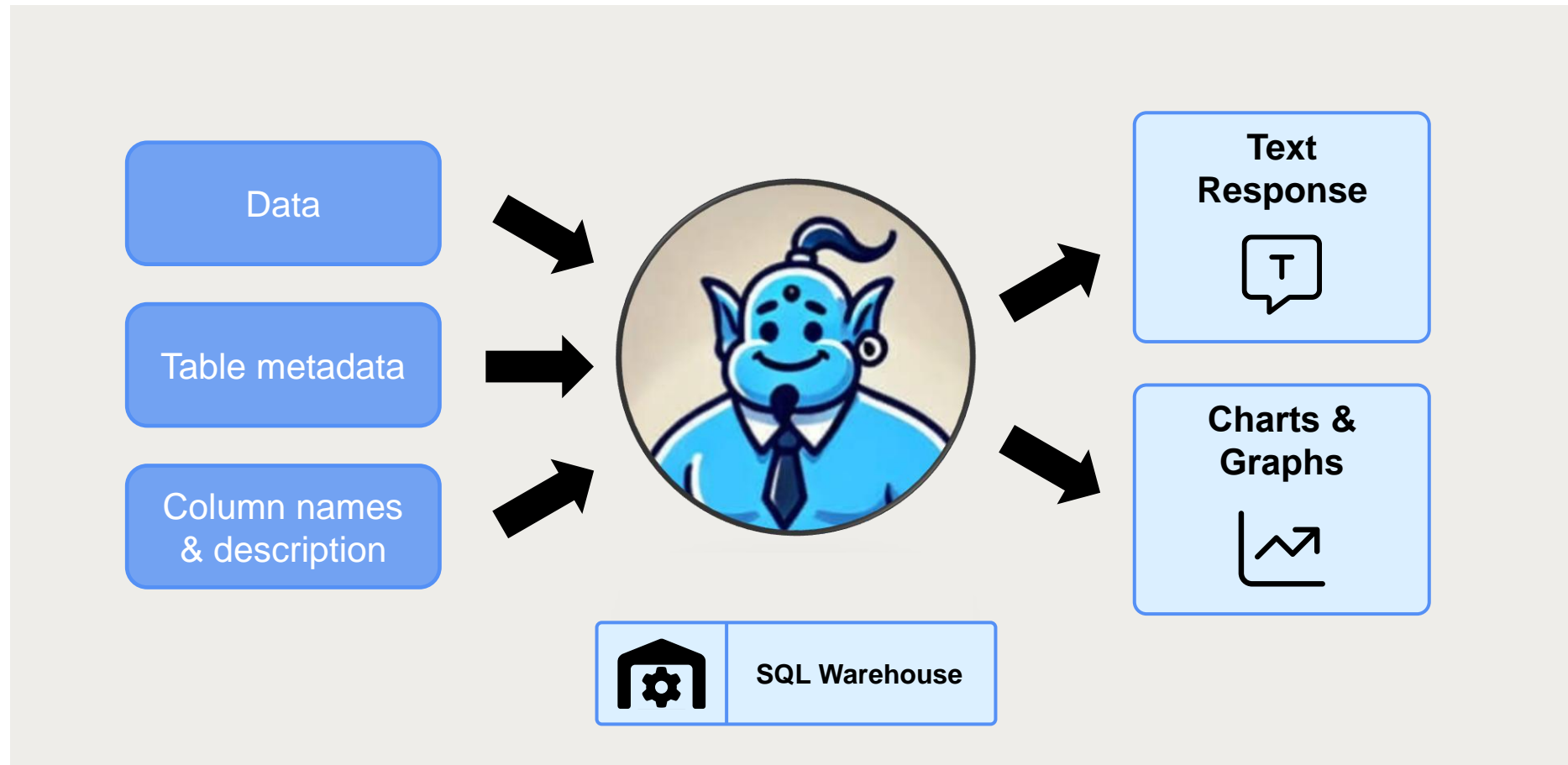
- Avanade
- CAE
- Air Canada



Expected Audience and Presentation Structure

- This session, we will cover:
 - Recap of AI/BI Genie's core capabilities and how it works.
 - Customization techniques to tailor Genie's response to deliver consistent and trustworthy insights.
 - Discuss the user and admin workflows.
 - Apply the best practices and techniques to create our own Genie.
- Fundamental knowledge: Basic SQL and Databricks. Experienced with AI/BI Genie would be beneficial but not required.

Quick Recap: How Genie works and What we built



Enabling Self Servicing: Six steps to deliver Accurate, Consistent, and Business-Aware Insights



Metadata

Value Dictionaries

Custom Instructions

Trusted Assets

Benchmarking

Feedback & Monitoring

Metadata: Help Genie Understand How Your Data Connects

Purpose: Improve Genie's generate accurate queries by clarifying how data relates through business-friendly metadata.

How: Leverage table description, column naming standards, column comments and table relationships (PK/FK) hints.

Tips

- ✓ Use descriptive, consistent column names (e.g. user_id, product_id, review_date).
- ✓ Add clear, business-friendly column comments. (e.g. Date when the review was submitted)
- ✓ Define PK/FK relationships. They serve as hints/guidance and not enforcement.

The screenshot shows a data catalog interface with tabs for Overview, Sample Data, Details, Permissions, History, Lineage, Insights, and Quality. The 'Overview' tab is selected, displaying a description of the 'Silver' table and a table of its columns.

Column	Type	Comment	Tags	Column masking rule
review_id PK	string	Unique identifier for the review.		
user_id FK	string	ID of the user who wrote the review.		
business_id FK	string	ID of the business being reviewed.		
review_date	date	Date when the review was submitted.		

Enabling Self Servicing: Six steps to deliver Accurate, Consistent, and Business-Aware Insights



Metadata

Value
Dictionaries

Custom
Instructions

Trusted
Assets

Benchmarking

Feedback &
Monitoring

Value Dictionaries: Help Genie Interpret Business Terms Accurately

Purpose: Map domain-specific terms to specific column values in string fields so Genie can understand (e.g. customer_segment).

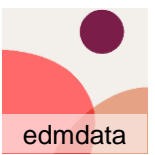
How: Genie automatically creates and manages list of values for eligible string columns.

Tips

- ✓ Enable dictionaries only for meaningful, well-maintained values. Refresh them regularly to keep value sets up to date.
 - ⚠ Dictionaries are managed by Genie. It cannot be edit it.
- ✓ For non-string columns, use Custom Instructions.
- ✓ Stay within limit of value dictionaries: [Public Preview](#)
 - 30 columns/Genie, 1024 distinct values, 127 max char.

Columns (14)			
Name		Description	
business_id	PK	Unique business id...	Add value dictionary
name		Name of the busin...	Value dictionary added
city		City where the busi...	Value dictionary added
state		State where the bu...	Value dictionary added
latitude		Latitude coordinate...	Add value dictionary

Enabling Self Servicing: Six steps to deliver Accurate, Consistent, and Business-Aware Insights



Metadata

Value
Dictionaries

Custom
Instructions

Trusted
Assets

Benchmarking

Feedback &
Monitoring

Custom Instructions: Guide Genie's Behaviour with Context, Tone and Examples.

Purpose: Shape Genie's response by defining how it should behave, what to prioritize, how to response and what to avoid.

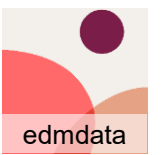
How: Use Markdown to define Genie's behavior in a structured way.

Tips

- ✓ Organize using Markdown formatting: Heading, bullet lists.
- ✓ Be specific and concise – less is more when communicating intent (e.g. the fiscal year starts on May).
- ✓ Tailor language and expectations to your audience.
- ✓ Define any common terms and include reporting standards.

A screenshot of a web application interface for configuring custom instructions. The interface has a top navigation bar with "Context" and "Settings" tabs, and a close button (X) in the top right. Below the navigation bar are three tabs: "Data", "Instructions" (which is selected and highlighted in blue), and "SQL Queries". The main content area is titled "General Instructions" and includes a subtitle "Add general instructions on how you want Genie to behave." Below this is a text input area containing several examples of instructions in a light blue monospace font, such as "* MCA stands for My Company Abbreviation" and "* Countries in the country_code column are stored with two characters (e.g. US, IT)". At the bottom of the text area, it says "(You can use markdown text)".

Enabling Self Servicing: Six steps to deliver Accurate, Consistent, and Business-Aware Insights



Metadata

Value
Dictionaries

Custom
Instructions

Trusted
Assets

Benchmarking

Feedback &
Monitoring

Trusted Assets: Prioritize Curated Logic to Help Genie Deliver Accurate Responses.

Purpose: Ensure Genie uses vetted queries and functions to produce responses users can rely on.

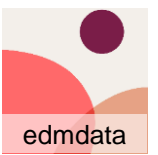
How: Add “Trusted” SQL Queries and Functions with parameters for Genie to prioritize when answering questions.

Tips

- ✓ Trust only production-ready, business-approved assets.
- ✓ Use parameters for repeatable logic.
- ✓ Regularly review Trusted assets to ensure they are valid and relevant.

A screenshot of a web application interface for configuring 'Trusted Assets'. The interface has a top navigation bar with 'Context' and 'Settings' tabs. Below this is a section with three tabs: 'Data', 'Instructions', and 'SQL Queries', with 'SQL Queries' being the active tab. The main content area is titled 'SQL queries & functions' and includes a sub-header 'Add example queries that Genie can learn from.' and a '+ Add' button. Below this is a table with two columns: 'Name' and 'Type'. The table contains one entry: a question mark icon followed by the text 'What is the average price range for busin' and the word 'Query'.

Enabling Self Servicing: Six steps to deliver Accurate, Consistent, and Business-Aware Insights



Metadata

Value
Dictionaries

Custom
Instructions

Trusted
Assets

Benchmarking

Feedback &
Monitoring

Benchmarking: Measure Genie's response accuracy using Business-Driven Prompts

Purpose: Test the overall response accuracy using a pre-defined set of test questions.

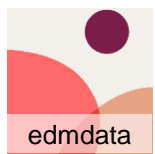
How: Create realistic questions the corresponding expected SQL query. Genie will compare its response with the provided query output

Tips

- ✓ Phase each question 3 to 5 different ways to simulate realistic language variability from users.
- ✓ Include questions known to be ambiguous or frequently misunderstood.
- ✓ Repeat benchmarks regularly to monitor performance over time.

Evaluations Questions (2)	
Question	SQL Answer
What is the max, min and avg stars for	<pre>SELECT max(stars) as max_stars, min(stars) as min_stars, avg(stars) as avg_stars FROM `demo_catalog`.`demo_yelp_academic`.`silver_business` ... 3 more lines</pre>
What is the top 5 reviews for the best	<pre>WITH top_5_reviews AS (SELECT b.name, r.text FROM `demo_catalog`.`demo_yelp_academic`.`silver_business` ... 16 more lines</pre>

Enabling Self Servicing: Six steps to deliver Accurate, Consistent, and Business-Aware Insights



Metadata

Value
Dictionaries

Custom
Instructions

Trusted
Assets

Benchmarking

Feedback &
Monitoring

Feedback & Monitoring: Improve Genie Over Time by Capturing and Acting on User Input.

Purpose: Creates a continuous improvement loop by tracking user's feedback to improve performance.

How: Users provide feedback on responses. These signals are tracked and monitored to help identify areas where Genie refinements are required.

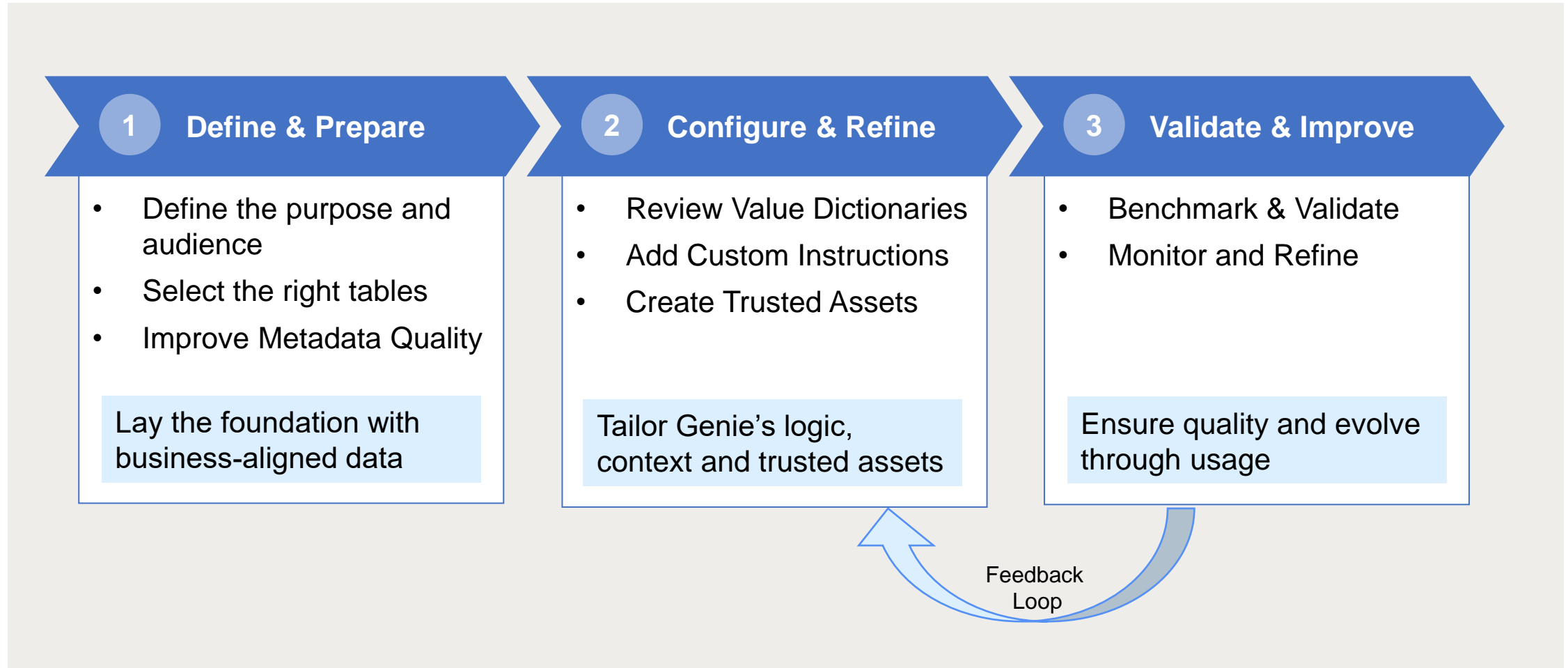
Tips

- ✓ Encourage users to give feedback (Yes, Fix it, or Request Review). They are crucial in helping improve the Genie.
- ✓ Establish a review process to "Close the loop".

A screenshot of the Genie interface. At the top, a light blue box contains the query: "What is the average stars for businesses in Sherwood Park, Alberta?". Below this, a text block explains: "This query calculates the average star rating of businesses located in Sherwood Park, Alberta. It focuses on the ratings provided by users to assess the overall quality of businesses in that specific area." A small icon and the text "silver_business" are visible. Below the text, it says "1 row" with a download icon, "Add as instruction", and "Show code". A table with one row is displayed, showing "1.2 average_stars" in the header and "3.5" in the data row. At the bottom, a feedback prompt asks "Is this correct? Your response makes Genie better over time." with three buttons: "Yes" (checked), "Fix it", and "Request review".

	1.2 average_stars
1	3.5

Demo – Lifecycle of Setup and Maintaining a Genie Space



Key Takeaways: Building a Trusted, Self-Serve Genie Space

- To democratize data, users need to “talk to the data” using natural language.
 - Genie enables this – but only works well with thoughtful setup.
- Metadata, Value Dictionaries and Trusted Assets help Genie interpret questions and generate reliable answers.
- Custom instructions align Genie’s tone, logic, and priorities with business context.
- Benchmarking and Feedback loops are essential to continuously improve the data companion’s accuracy and usability.
- The code is in my Github repo:
 - Link: <https://github.com/wilson-mok/demo/tree/main/2025/EdmData/2-Azure-Databricks-Democratizing-Data-with-AI-BI-Genie>

Thank you



LinkedIn