

Microsoft Partner Project Ready

Implement with Impact

Modern Data Platform with Azure Databricks

<Speaker name or subtitle>

<Date>

Day 3 of 3



Course Plan and Learning Objectives



Day 1

Module 1 - Introduction to Azure Databricks

- Azure Databricks: A Data Intelligent Platform
- Why Azure Databricks
- Decision guide: Azure Databricks vs. Microsoft Fabric

Module 2 - Migration to Azure Databricks

- Microsoft Cloud Adoption Framework for Azure
- Migration strategies
- Data landing zones
- Migration scenarios

Interactive Simulated Lab Experience

- End-to-End Streaming Pipeline with Lakeflow Declarative Pipelines in Azure Databricks

Day 2

Module 3 - Integration with Azure

- Seamless integration with Microsoft Azure services
- Connect to Azure Data Lake Storage (ADLS) Gen2 and Blob Storage
- Leverage Azure Databricks for Azure Cosmos DB Operations
- Secret management with Azure Key Vault
- Connect Azure Databricks to Azure Event Hubs

Module 4 - Integration with Microsoft Fabric and Power BI

- Data Intelligence with Azure Databricks and Microsoft Fabric
- Connect Power BI to Azure Databricks
- Integration with Azure Data Factory
- Mirroring Azure Databricks Unity Catalog

Interactive Simulated Lab Experience

- Setup and use Unity Catalog for Data Management in Azure Databricks
- Real-Time Streaming with Azure Databricks and Azure Event Hubs

Day 3

Module 5 - Integration with Azure AI Foundry

- Azure Databricks connector in Azure AI Foundry
- Mosaic AI and machine learning on Azure Databricks
- Query Generative AI model serving endpoints
- Databricks Assistant, AI/BI Genie and AI Functions on Azure Databricks
- Chat with LLMs and prototype GenAI apps using AI Playground
- Build and optimize agents on your data with Agent Bricks

Module 6 - Security and Governance

- Integrate Azure Databricks with Microsoft Purview
- Integration of Azure Databricks Unity Catalog with Microsoft Purview

Module 7 - Well-architected for Azure Databricks

- Lakehouse implementation: Principles and best practices
- Azure Databricks well-architected framework

Interactive Simulated Lab Experience

- Responsible AI with Large Language Models using Azure Databricks and Azure OpenAI
- Connect to and manage Azure Databricks in Microsoft Purview

05

Integration with Azure AI Foundry

Azure Databricks connector in Azure AI Foundry

Azure OpenAI



gpt-4

gpt-4.1

gpt-4.1-nano

gpt-4.1-mini

gpt-4o

Bring
models
to data

Lakeflow
Ingest, ETL,
Streaming



Databricks SQL
Data
Warehousing



AI/BI
Business
Intelligence



Mosaic AI
Artificial
Intelligence



Knowledge of your data

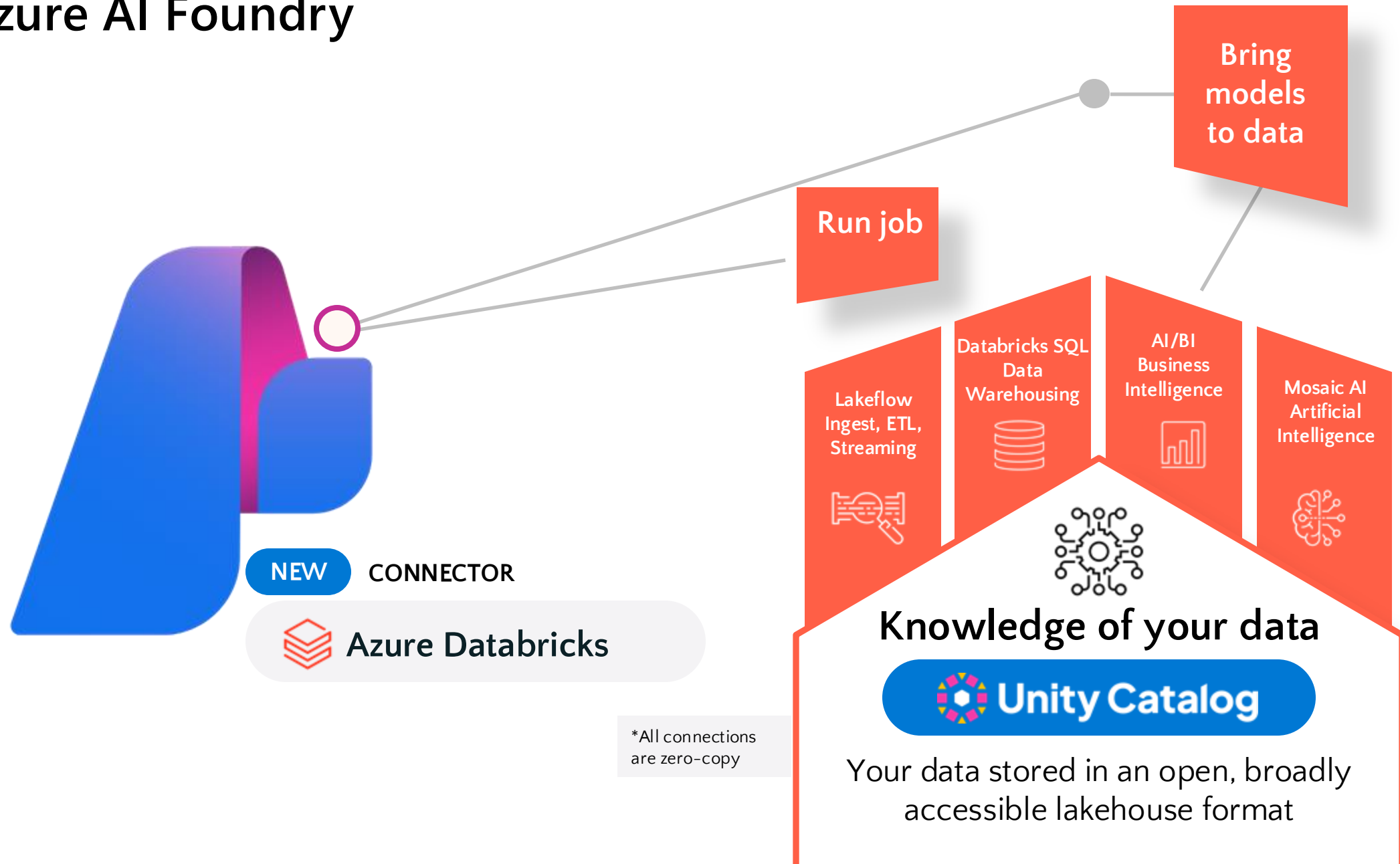
 **Unity Catalog**

Your data stored in an open, broadly
accessible lakehouse format

*All connections
are zero-copy



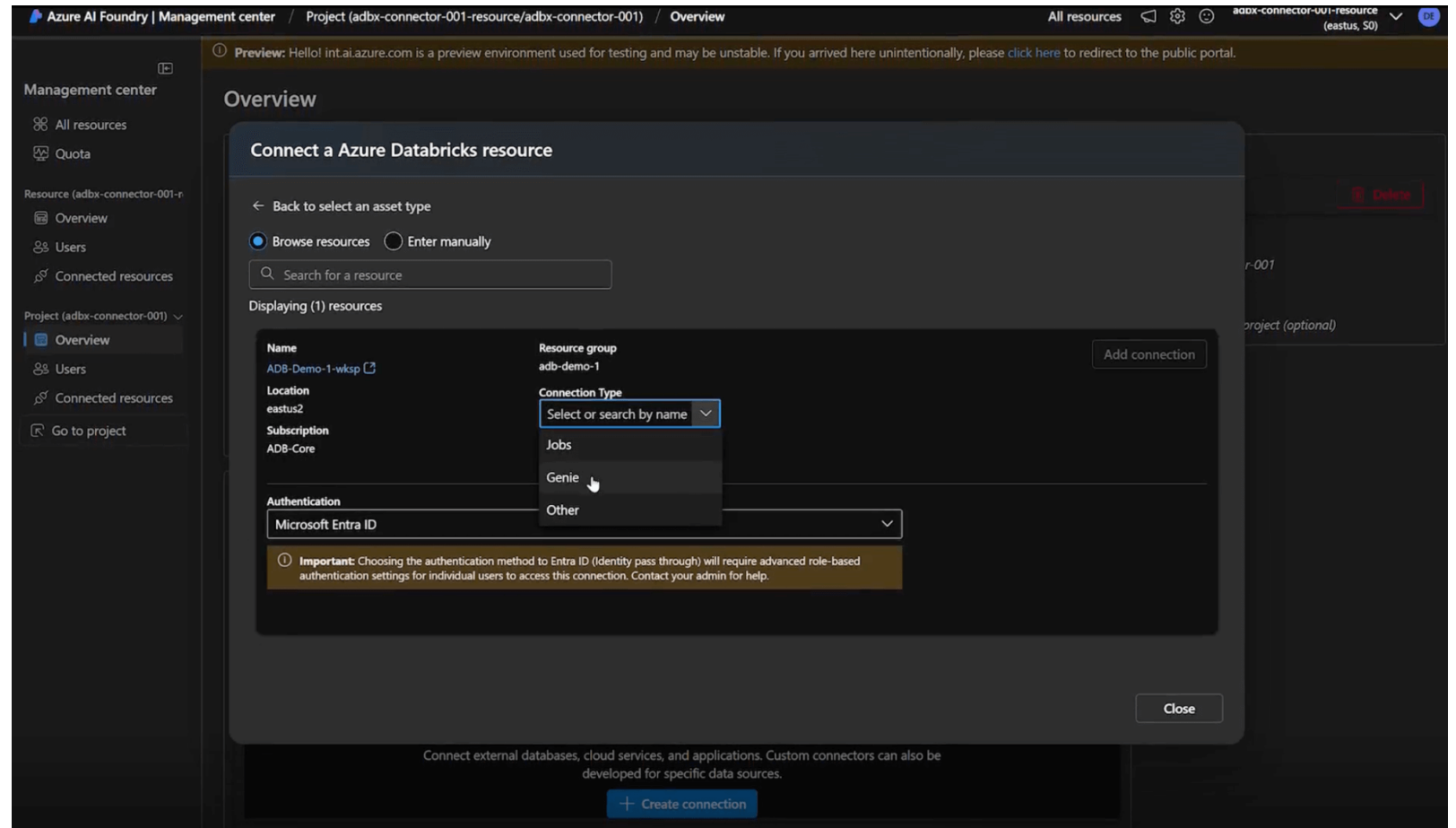
Azure AI Foundry



Azure Databricks connector in Azure AI Foundry

Build enterprise-grade AI agents that reason over real-time Azure Databricks data while being governed by Unity Catalog

These agents will also be enriched by the responsible AI capabilities of Azure AI Foundry



Mosaic AI and machine learning on Azure Databricks

Mosaic AI

Create domain-specific
agentic applications

Lakeflow
Ingest, ETL,
Streaming



Databricks SQL
Data
Warehousing



AI/BI
Business
Intelligence



Mosaic AI
Artificial
Intelligence



Knowledge of your data

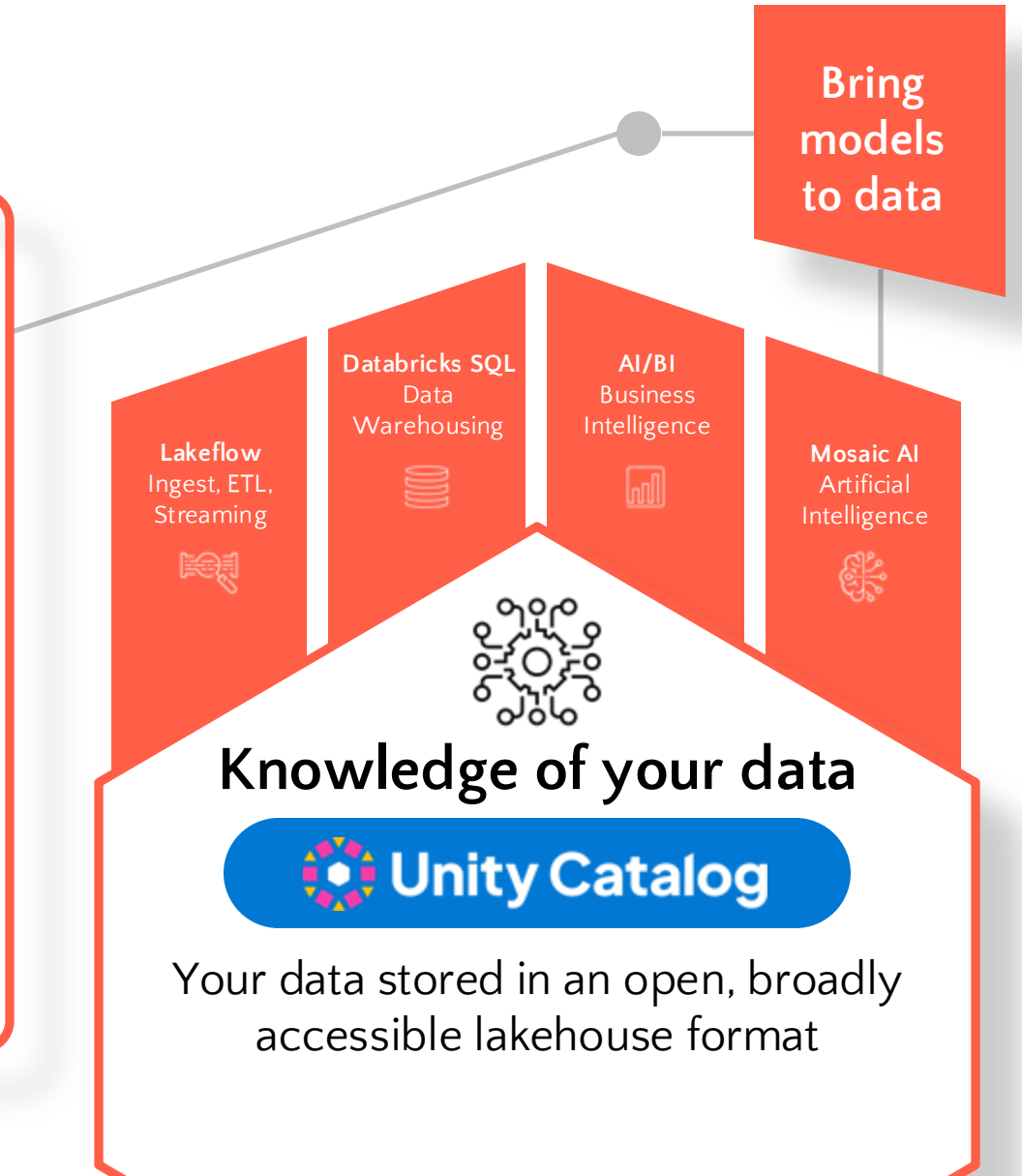


Unity Catalog

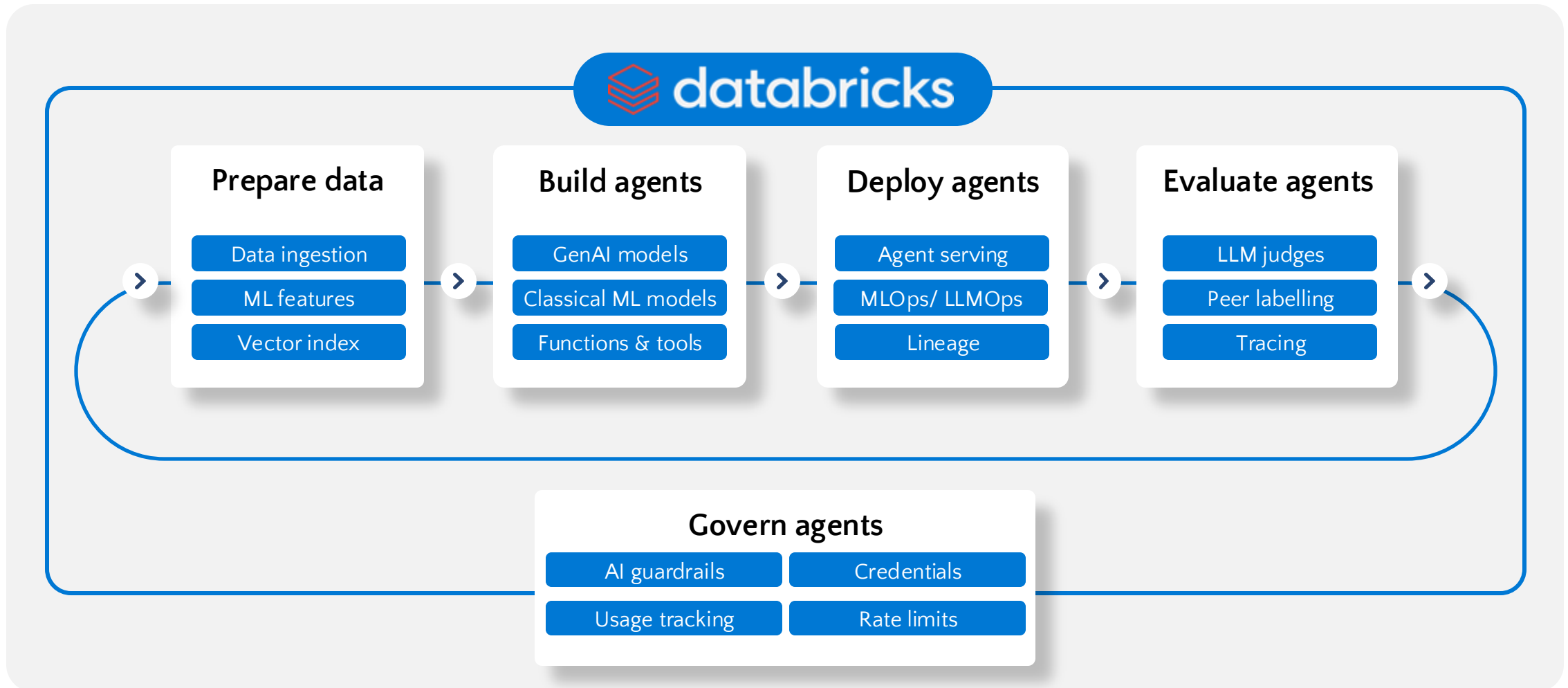
Your data stored in an open, broadly
accessible lakehouse format

RAG, finetune, agentize your data

Connect it with any model on the cloud

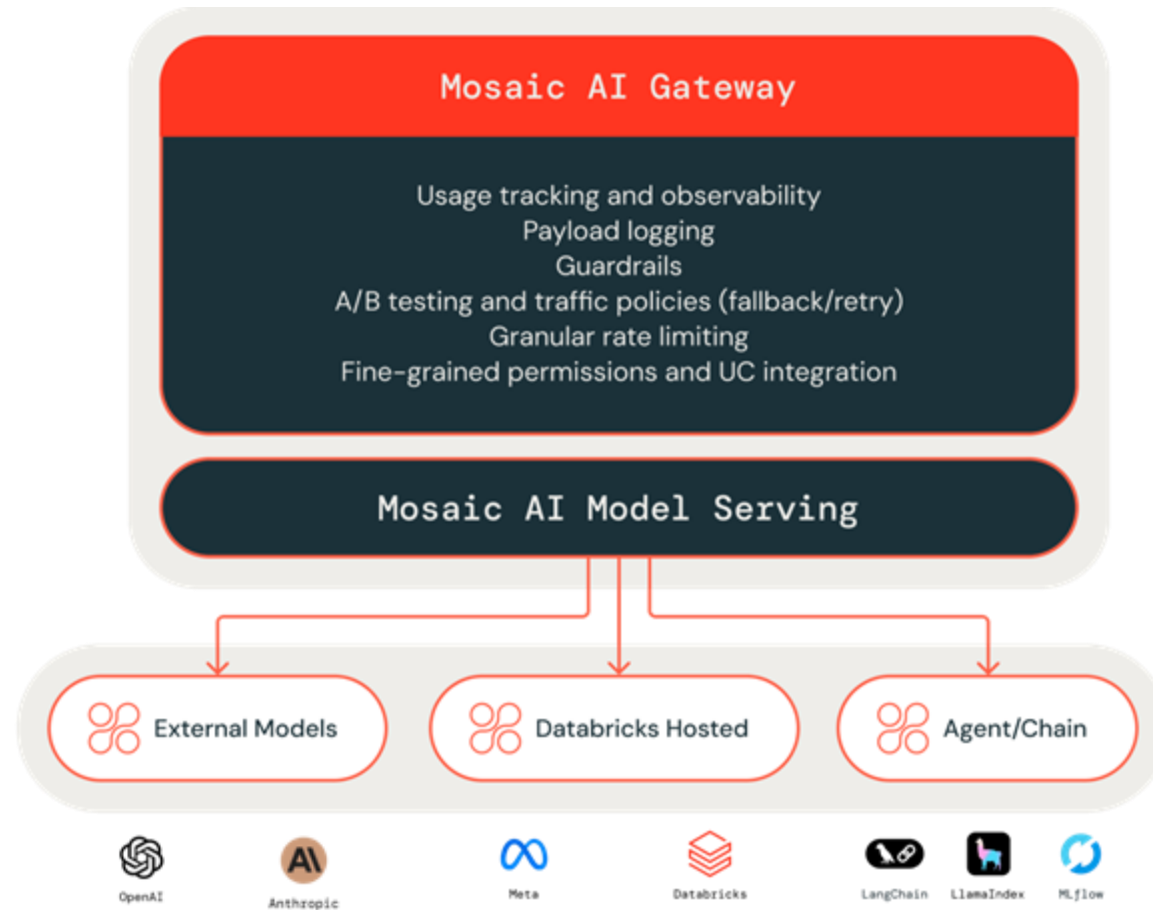


Unified platform to build AI apps & agents



AI Gateway & Foundation Models

- Manage and govern all GenAI models, agents, and chains across the enterprise
- Support for Azure OpenAI, Anthropic
- Permission and rate limiting to control who has access and how much access



Batch and real-time AI serving

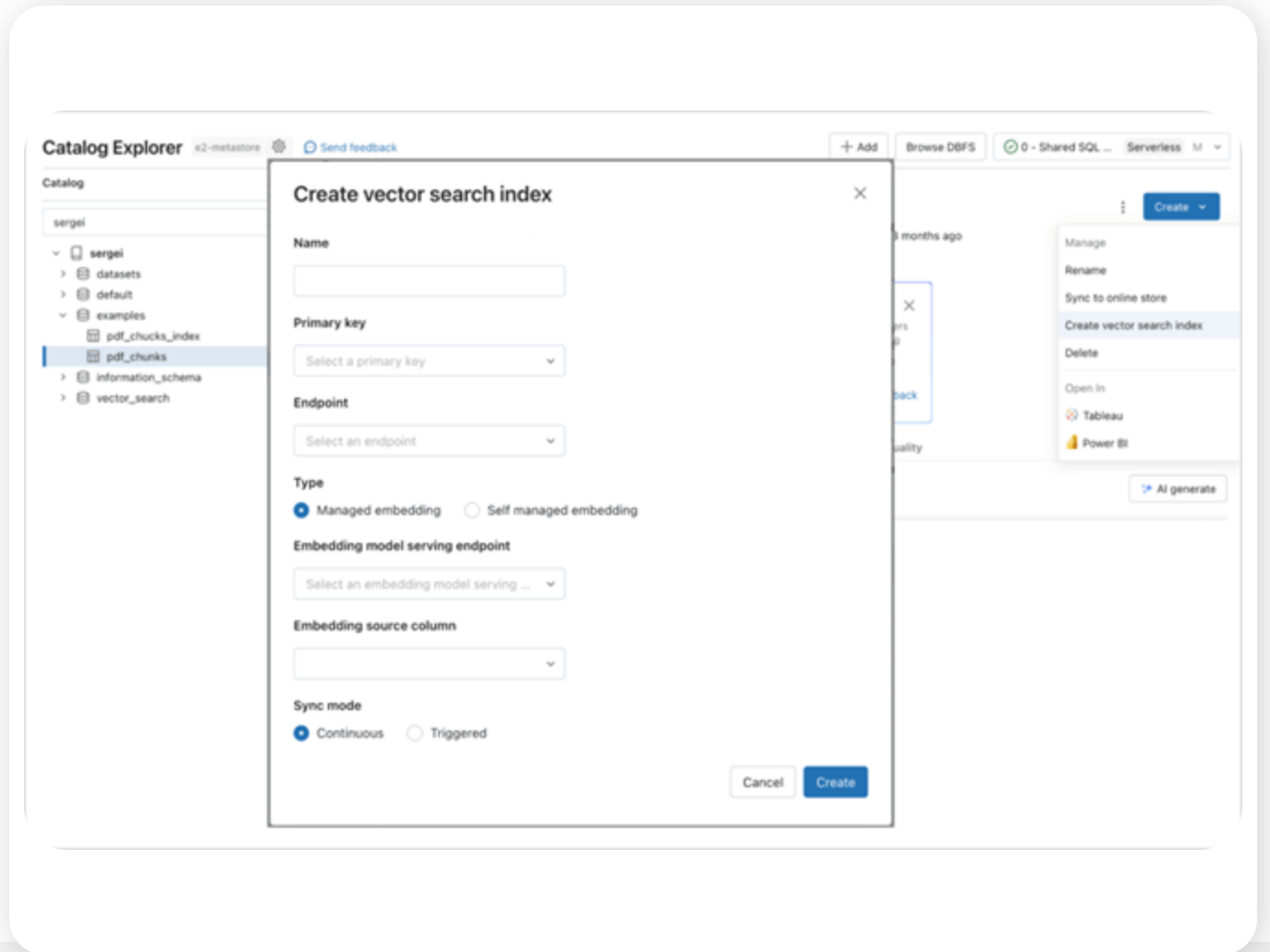
- Single solution to deploy, govern, and manage AI models and agents for batch and real-time workloads
- Deploy both custom and GenAI models as REST APIs
- Manage models hosted by other providers like Azure OpenAI (ChatGPT) and AWS (Anthropic)

The screenshot displays the 'Serving endpoints' interface for Foundation Model APIs. It features a sidebar with navigation icons, a top header with a 'Provide feedback' link, and a main content area. Two model cards are shown at the top: 'Llama3 70B Chat' (Chat • Pay-per-token) and 'GTE Large (En)' (Embedding • Pay-per-token), both with 'Preview' buttons and 'Query' links. Below these is a search bar labeled 'Filter endpoints' and a filter dropdown set to 'Created by me'. A table lists the endpoints with columns for Name, Status, and Served entities.

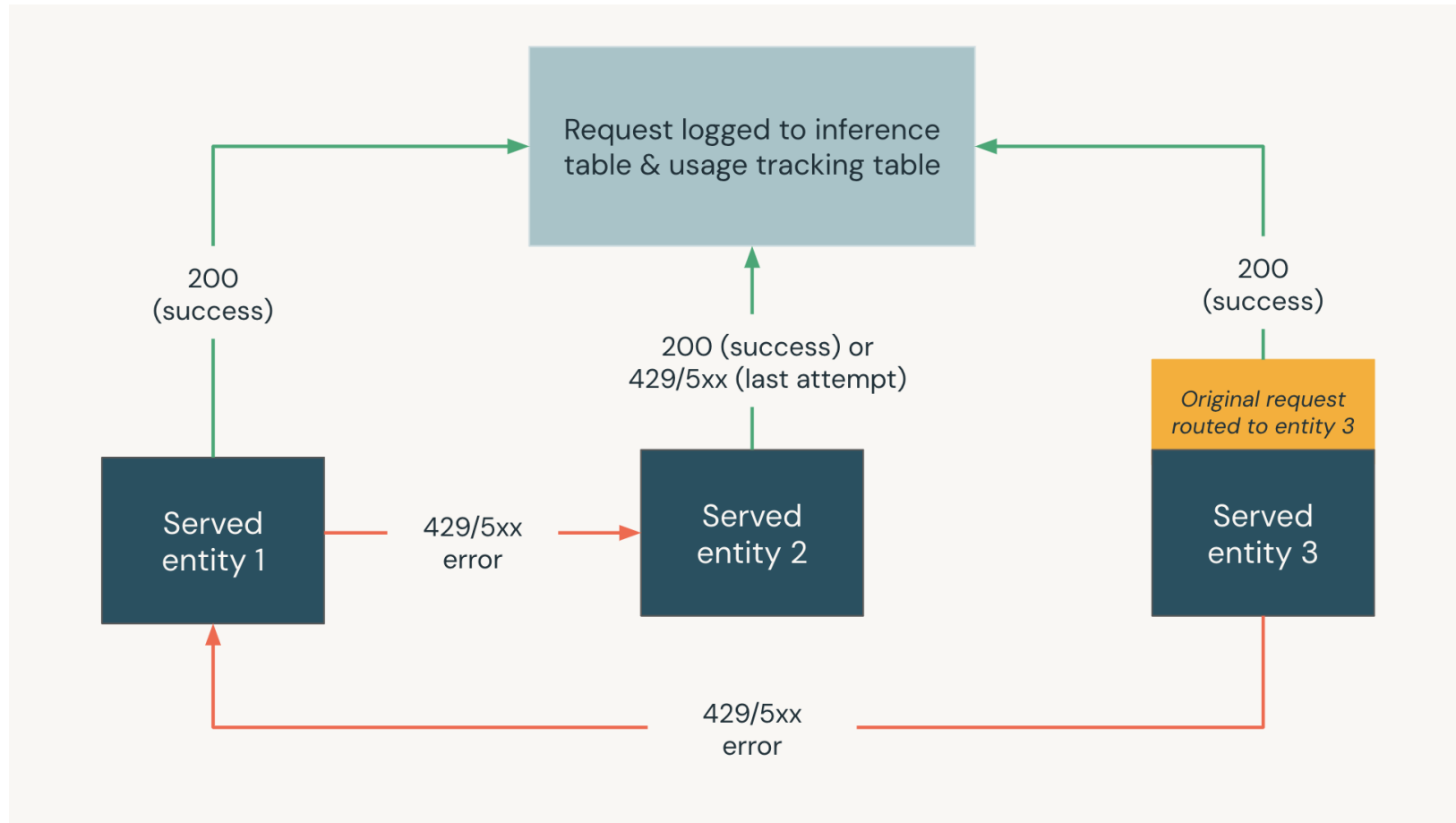
Name	Status	Served entities
databricks-dbrx-instruct	✓ Ready	DBRX Instruct
databricks-meta-llama-3...	✓ Ready	Meta Llama 3 70B Instruct
databricks-mixtral-8x7b-i...	✓ Ready	Mixtral-8x7B Instruct
External GPT 4 Chat	✓ Ready	GPT-4.0 (Open AI)
External GPT 3.5 Chat	✓ Ready	GPT-3.5-Turbo
Visionary-Proft-Boost-55	✗ Not ready	

Vector Search and Feature Store

- Feature and Function Serving that serves structured data to RAG applications
- Vector Search to build RAG applications with unstructured data
- Automated real-time pipelines that synchronize with source data in the data lake without any intervention



Example | AI Gateway on model serving endpoints



Demo

Create and query a vector search index


External models in Mosaic AI Model Serving

External models allow you to streamline the usage and management of various large language model (LLM) providers, such as

OpenAI

Anthropic

Python

 Copy

```
import mlflow.deployments

client = mlflow.deployments.get_deploy_client("databricks")
client.create_endpoint(
    name="openai-completions-endpoint",
    config={
        "served_entities": [{
            "name": "openai-completions",
            "external_model": {
                "name": "gpt-3.5-turbo-instruct",
                "provider": "openai",
                "task": "llm/v1/completions",
                "openai_config": {
                    "openai_api_key": "{{secrets/my_openai_secret_scope/openai_api_key}}"
                }
            }
        }]
    }
)
```

Anthropic | Configure a serving endpoint

To serve and query external models you need to configure a serving endpoint

The *external_model* field defines the model to which this endpoint forwards requests

Python

Copy

```
import mlflow.deployments

client = mlflow.deployments.get_deploy_client("databricks")

client.create_endpoint(
    name="anthropic-completions-endpoint",
    config={
        "served_entities": [
            {
                "name": "test",
                "external_model": {
                    "name": "claude-2",
                    "provider": "anthropic",
                    "task": "llm/v1/completions",
                    "anthropic_config": {
                        "anthropic_api_key": "{{secrets/my_anthropic_secret_scope/anthropic_api_key}}"
                    }
                }
            }
        ]
    }
)
```

Anthropic | Configure the provider for an endpoint

Configuration Parameter	Description	Required
<i>anthropic_api_key</i>	The Azure Databricks secret key reference for an Anthropic API key. If you prefer to paste your API key directly, see <i>anthropic_api_key_plaintext</i>	You must provide an API key using one of the following fields: <i>anthropic_api_key</i> or <i>anthropic_api_key_plaintext</i>
<i>anthropic_api_key_plaintext</i>	The Anthropic API key provided as a plaintext string. If you prefer to reference your key using Azure Databricks Secrets, see <i>anthropic_api_key</i>	You must provide an API key using one of the following fields: <i>anthropic_api_key</i> or <i>anthropic_api_key_plaintext</i>

Anthropic | Query a model endpoint

- You can send scoring requests to the endpoint using the OpenAI client, the REST API or the MLflow Deployments SDK
- The given example queries the claude-2 completions model hosted by Anthropic using the OpenAI client

Python

```
import os
import openai
from openai import OpenAI

client = OpenAI(
    api_key="dapi-your-databricks-token",
    base_url="https://example.staging.cloud.databricks.com/serving-endpoints"
)

completion = client.completions.create(
    model="anthropic-completions-endpoint",
    prompt="what is databricks",
    temperature=1.0
)
print(completion)
```


Mosaic AI Model Training

Streamlines and unifies
the process of training
and deploying traditional
ML models through

AutoML

Foundation Model Fine-
tuning workloads

Mosaic AI Model Training Preview

General

Task

☒ Chat Completion
Finetune your model on chat logs between a user and an AI assistant

☐ Continued Pre-training
Train your model with additional text data to add new knowledge to a model

☐ Instruction Finetuning
Finetune your model on structured prompt-response data to adapt the model to a new task

Select Foundation Model

Models trained in Model Training may be subject to license requirements and/or use policies. [Learn more](#)

Llama 3 8B Instruct

Training data

Training data*

Select a dataset or enter a link to a Hugging Face dataset, for example databricks/databricks-dolly-15k

Browse training data

Browse

Model registration

Register to location*

Select the location in Unity Catalog where you want to register the trained model

Select a catalog

Choose schema

Model name

ift-meta-llama-3-8b-instruct-zdnknm

Advanced options >

Optional. You can change hyperparameters, evaluation prompts and resume training from previous checkpoint.

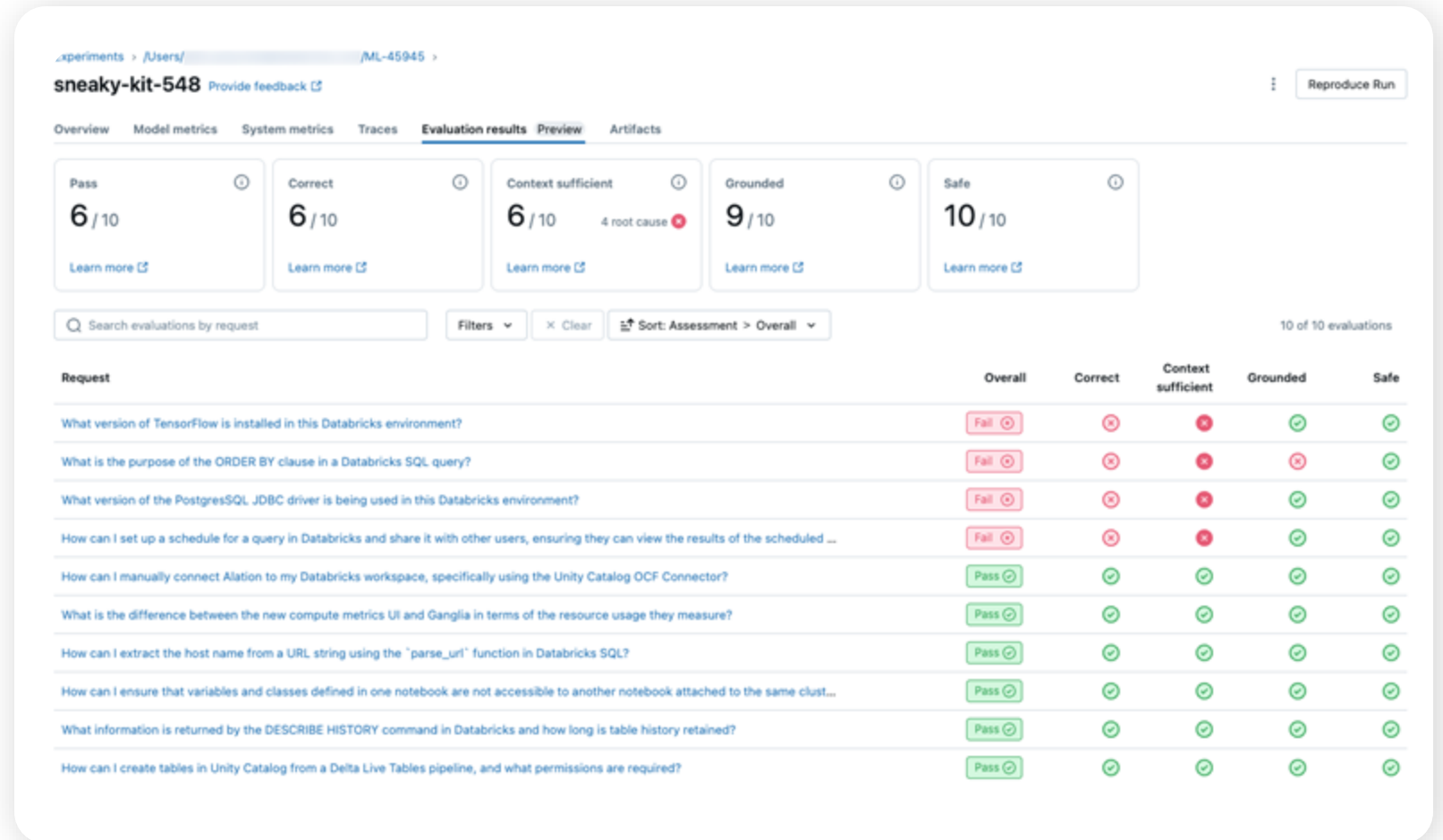
By clicking "Start Training", you acknowledge that your use of Databricks to finetune Meta Llama 3 is governed by the [Meta Llama 3 Community License](#) in addition to the Databricks [terms of service](#)

Agent evaluation

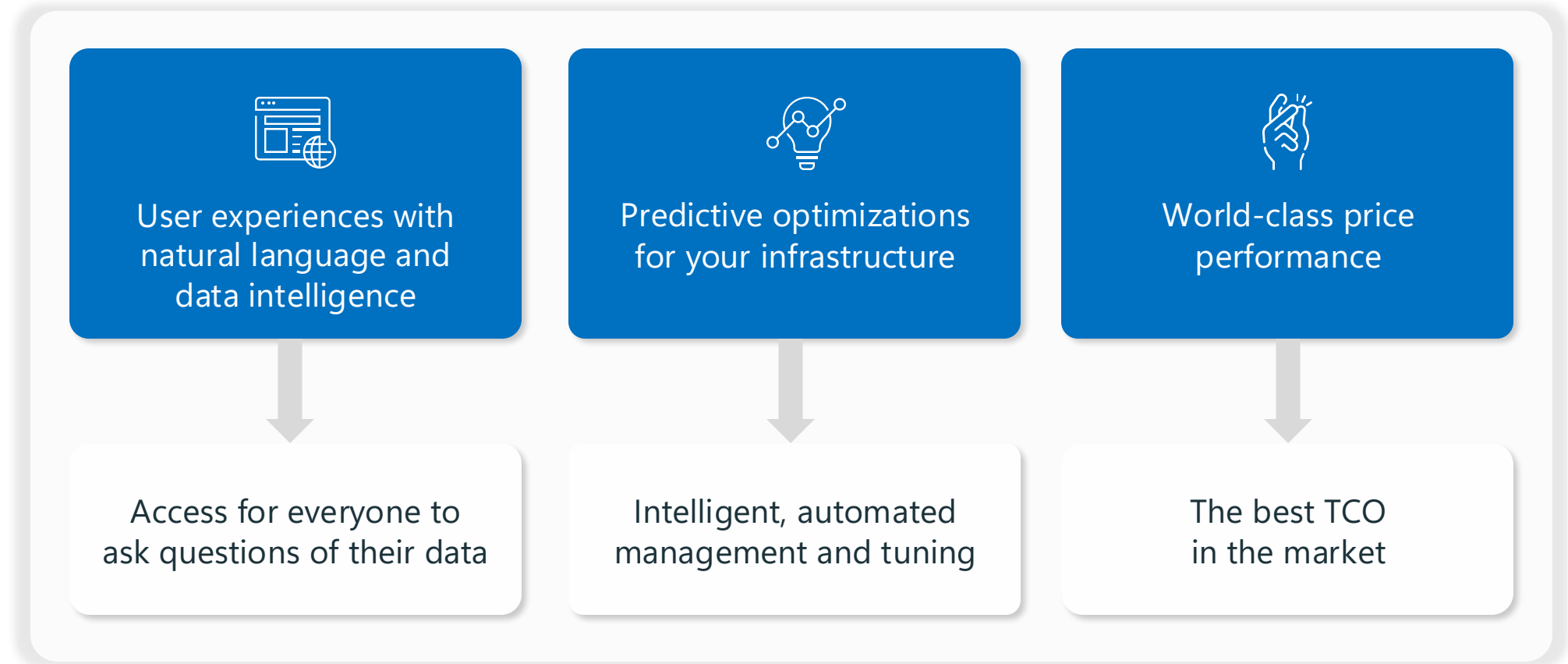
Evaluate the quality, cost, and latency of Agents, including RAG applications and chains

Human and machine evaluators which use advanced, research-backed techniques

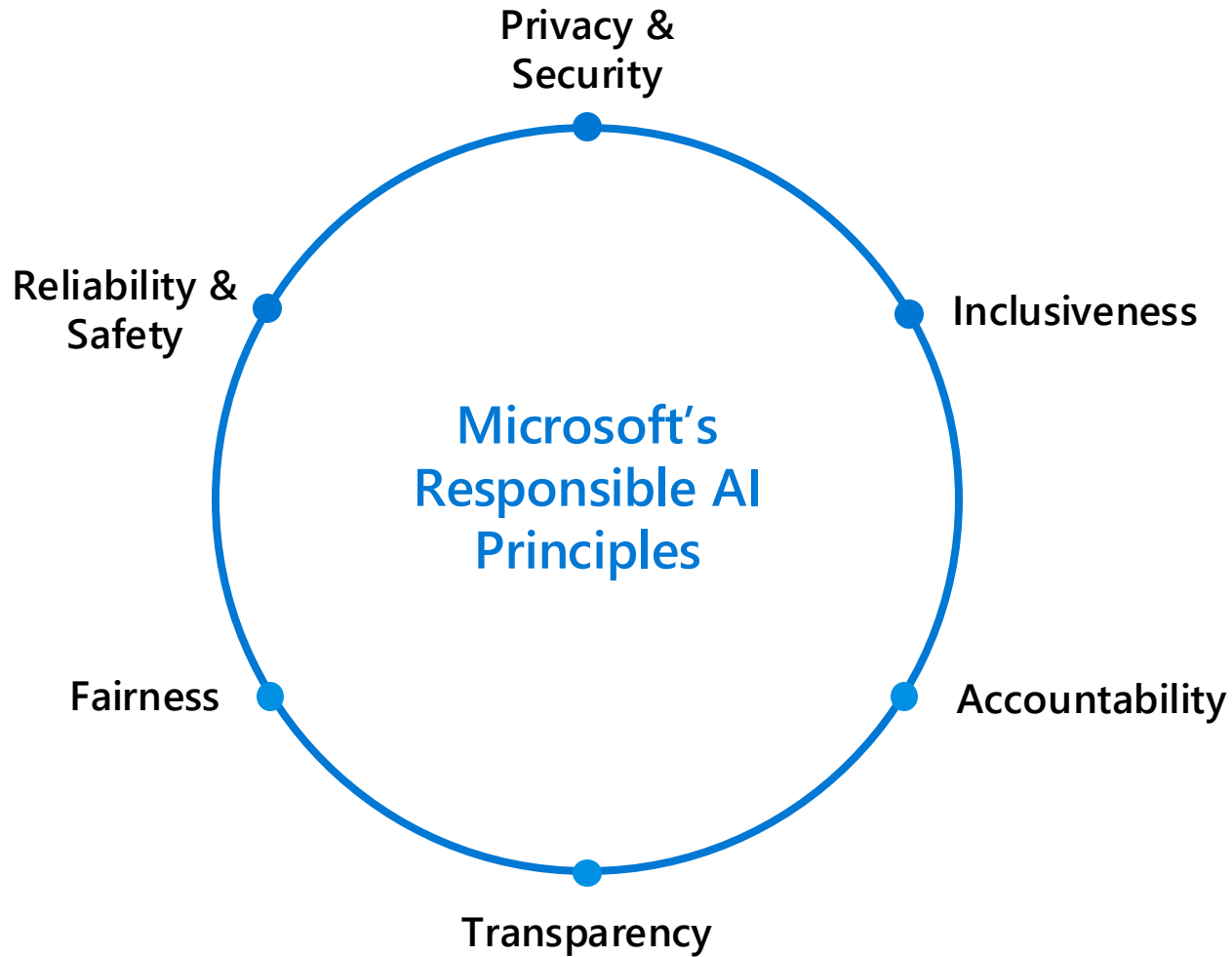
Synthetic data generation API creates high-quality evaluation data based on enterprise data



Intelligent Data Warehousing



Microsoft's Responsible AI Principles



Building blocks to enact principles



Tools and processes



Training and practices



Rules



Governance

Query Generative AI model serving endpoints

Query Generative AI models

- | Mosaic AI Model Serving supports [Foundation Models APIs](#) and [external models](#) for accessing Generative AI models
- | Model Serving uses a unified OpenAI-compatible API and SDK for querying them

Python

```
from databricks.sdk import WorkspaceClient

w = WorkspaceClient()
openai_client = w.serving_endpoints.get_open_ai_client()

response = openai_client.chat.completions.create(
    model="databricks-dbrx-instruct",
    messages=[
        {
            "role": "system",
            "content": "You are a helpful assistant."
        },
        {
            "role": "user",
            "content": "What is a mixture of experts model?",
        }
    ],
    max_tokens=256
)
```

Model Serving | Options for sending scoring requests to endpoints

| OpenAI client

| Serving UI

| REST API

| MLflow Deployments
SDK

| Databricks Python SDK

| SQL function

Demo

Query a chat completion model

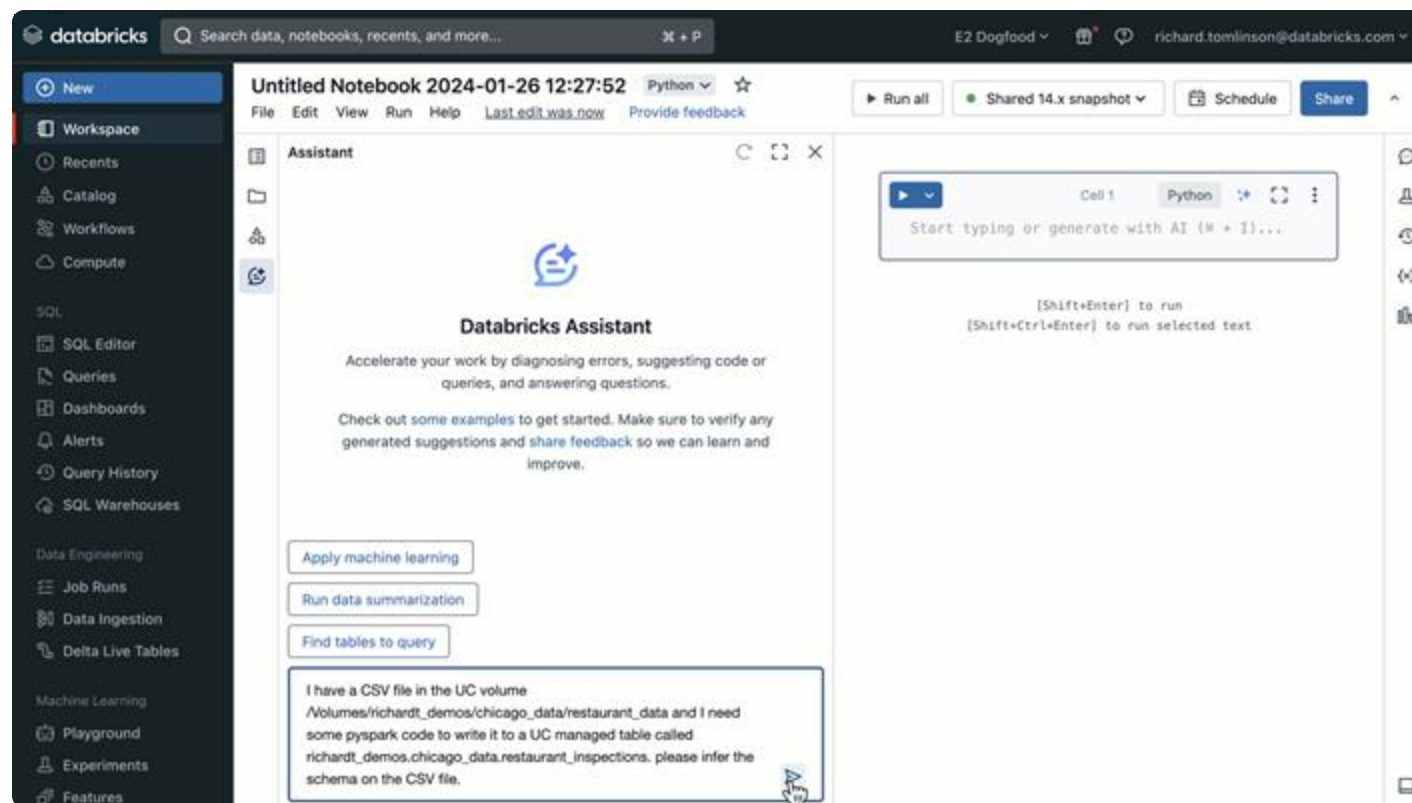
Databricks Assistant, AI/BI Genie and AI Functions on Azure Databricks

Databricks Assistant

Boost productivity with your context-aware AI assistant



- Generate, fix and optimize code using natural language
- Leverages UC metadata, code cells, libraries, runtime, and more as context to improve accuracy
- Explains, diagnoses, and fixes issues from within a cell within seconds



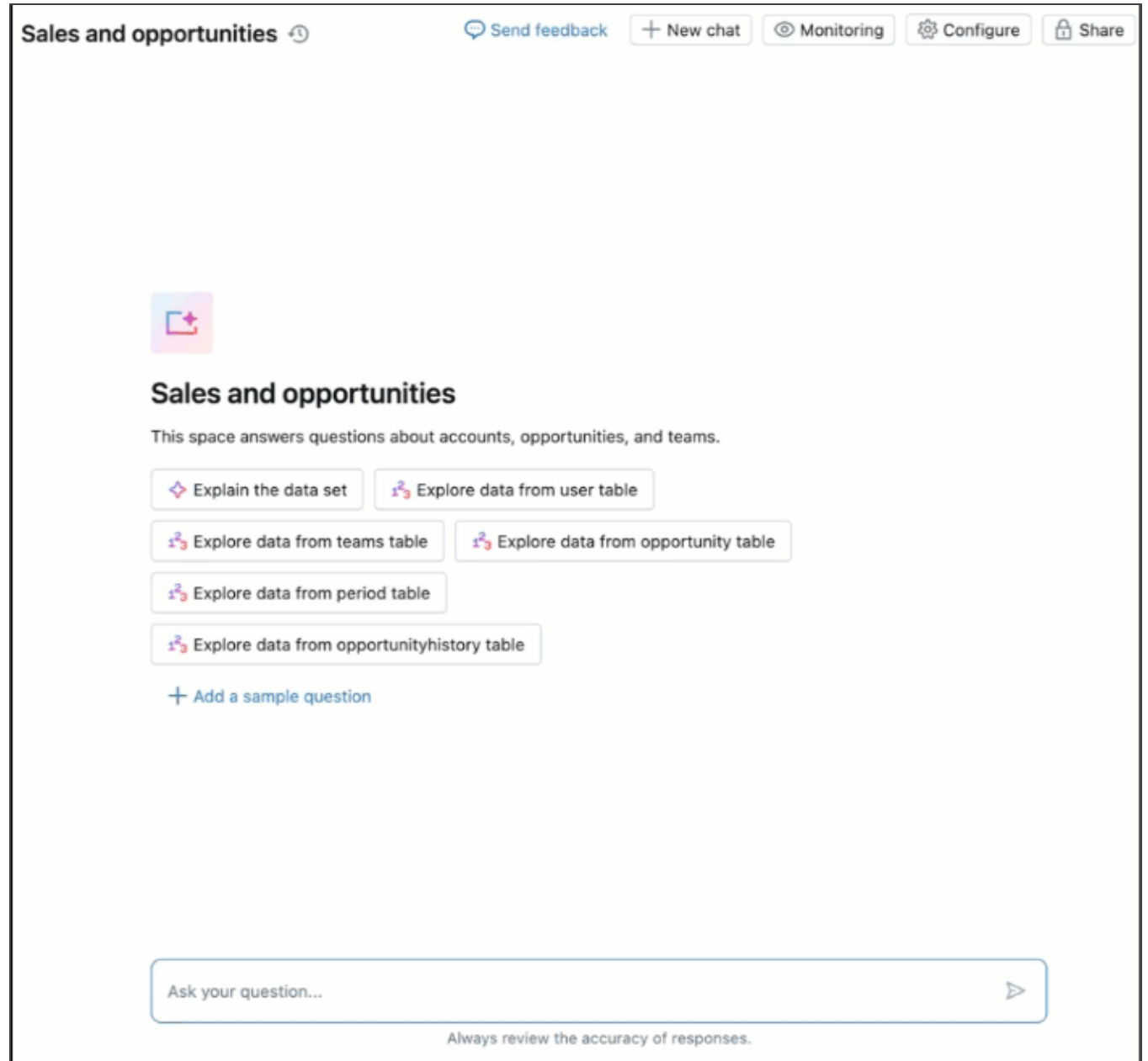
**Databricks Assistant Quick Fix
has a 75% acceptance rate**

 Over 150K users assisted every month! 

AI/BI Genie space

A chat interface for business users to ask natural-language questions about their data

Configure Genie spaces with datasets, sample queries, and text guidelines to help Genie translate business questions into analytical queries



Generate responses with Genie

Genie uses the following components to generate responses:

Table metadata

Column names and descriptions

Example SQL queries

SQL functions

Instructions

Prompt and responses history

order history ☆

+ New chat

History

Configure

Monitoring

This report shows the number of customers from each nation, calculated by combining customer data with their corresponding nation information. customer nation

25 rows [Add as instruction](#) [Show code](#)

	n_name	customer_count
1	EGYPT	29,933
2	FRANCE	30,062
3	ALGERIA	29,776
4	IRAN	30,257
5	INDIA	30,234
6	JAPAN	29,966
7	KENYA	29,712
8	RUSSIA	30,088
9	MOZAMBIQUE	29,690
10	CHINA	29,949
11	JORDAN	29,951

Distribution of Nations in Customer Table

n_name

INDONESIA

IRAN

INDIA

IRAQ

ETHIOPIA

GERMANY

UNITED STATES

UNITED KINGDOM

RUSSIA

ARGENTINA

Is this correct? Your response makes Genie better over time.

Yes

Fix it

Request review

What is the total number of customers from each nation?

Which nation has the highest average account b

Trusted assets in Genie

Trusted assets are predefined functions and example queries meant to provide verified answers to questions that you anticipate from users

When a user submits a question that invokes a trusted asset, it's indicated in the response, adding an extra layer of assurance to the accuracy of the results

What are my open opportunities in EMEA



Genie

✓ Trusted

Found a function that might match your question: `open_opps_in_region` ⓘ

✓ Show less

Addresses questions about the pipeline in the specified regions by returning a list of all the open opportunities. If no region is specified, returns all open opportunities. Example questions: "What is the pipeline for APAC and EMEA?", "Open opportunities in APAC"

• **regions:** EMEA

				📄	⬇	↻
	^A _C Oppld	^A _C Region	^A _C Oppor			
1	320dbe8e20c449ec9eb821423a7c828f	EMEA	Zoonoodl			
2	bf932f6d2a02410c875bf95540e61c9c	EMEA	Zoombear			
3	ca75fc84dd7246d7b5fc2fc3ac659a19	EMEA	Zoomdog			
4	6a0c761b5ba64da583c4f11d1a8c648e	EMEA	Eimbee -			
5	d9d0355e1da24f708ee081845c231179	EMEA	Rhynyx - I			
6	c23a8393a4804fe59aa86a9d3797746d	EMEA	Aibox - N			
7	ef0a0e8bdd80478dab70503ce93474e4	EMEA	Mymm - N			
8	e044df9c272a4a2b9e2893e3e697e932	EMEA	Kwideo - I			
9	46da9c0000784c04bfdbd1ba1806927f	EMEA	Realmix -			
10	3f0bf73bf05941bc9467759efb96fc27	EMEA	Skilith - N			
11						

932 rows

AI Functions on Azure Databricks

AI Functions invoke a state-of-the-art Generative AI model from Databricks Foundation Model APIs to perform tasks like,

- Sentiment analysis
- Classification
- Translation

SQL

```
> SELECT ai_analyze_sentiment('I am happy');  
positive  
  
> SELECT ai_analyze_sentiment('I am sad');  
negative
```

SQL

```
> SELECT ai_translate('Hello, how are you?', 'es');  
"Hola, ¿cómo estás?"  
  
> SELECT ai_translate('La vida es un hermoso viaje.', 'en');  
"Life is a beautiful journey."
```

Chat with LLMs and prototype GenAI apps using AI Playground

LangChain on Azure Databricks for LLM development

LangChain is a software framework designed to help create applications that utilize large language models (LLMs)

With LangChain integrations you can

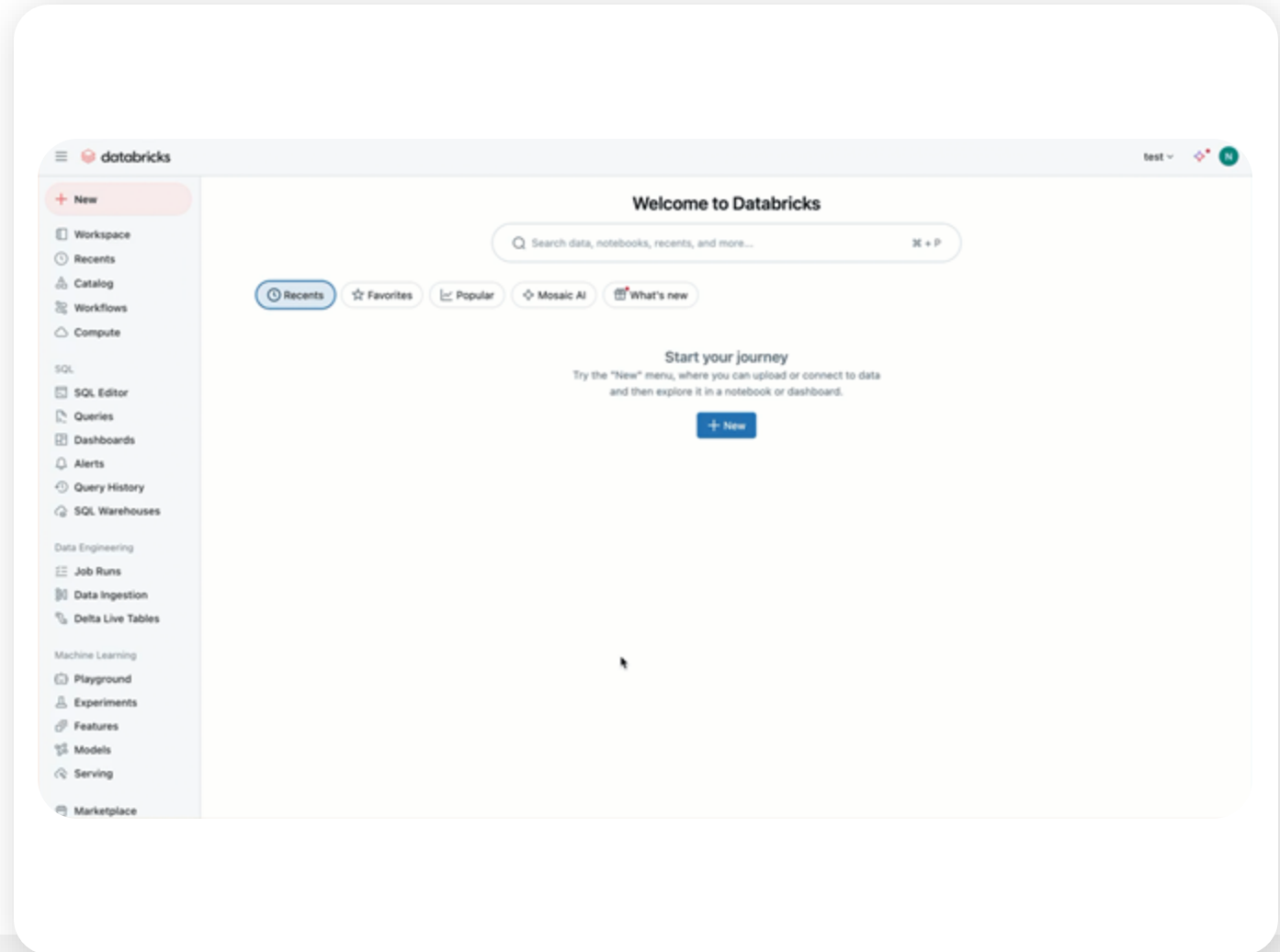
Seamlessly load data from a PySpark DataFrame with the PySpark DataFrame loader

Interactively query your data using natural language with the Spark DataFrame Agent or Databricks SQL Agent

Wrap your Azure Databricks served model as a large language model (LLM) in LangChain

AI Playground

- Interactive environment where users can experiment with various models simultaneously and compare results
- Test any model endpoint hosted in Databricks or outside
- Automatically tracks prompt engineering experiments to build evaluation datasets



Prototype Generative AI agents in AI Playground

Use the AI Playground to give the tool to an agent and interact with it to validate and test behavior

The AI Playground provides a sandbox to prototype tool-calling agents





Select an endpoint

Pay-per-token

Provisioned Throughput


External


Custom Agent

Name	Type
Llama 4 Maverick databricks-llama-4-maverick	Pay-per-token
✓ Meta Llama 3.3 70B Instruct databricks-meta-llama-3-3-70b-instruct	 Pay-per-token
Claude 3.7 Sonnet databricks-claude-3-7-sonnet	 • Pay-per-token
Meta Llama 3.1 8B Instruct databricks-meta-llama-3-1-8b-instruct	Pay-per-token
Meta Llama 3.1 405B Instruct databricks-meta-llama-3-1-405b-instruct	 • Pay-per-token
Shutterstock ImageAI, Powered by Databricks databricks-shutterstock-imageai	 • Pay-per-token
DBRX Instruct	

Meta Llama 3.3 70B Instruct


Ready

 Tools enabled

 Pay-per-token

Llama 3.3 is a state-of-the-art 70B parameter dense language model trained and released by Meta. The model supports a context length of 128K tokens. The model is optimized for multilingual dialogue use cases and aligned with human preferences for helpfulness and safety. It is not intended for use in languages other than English. Meta Llama 3.3 is licensed under the Meta Llama 3.3 Community License, Copyright © Meta Platforms, Inc. All Rights Reserved. Customers are responsible for ensuring compliance with applicable model licenses.

Model

 **meta-llama-3.3-70b-instruct**
Databricks

Price per 1M tokens
21.429 DBUs

Cancel

Use endpoint

**Build and optimize agents on your data with
Agent Bricks**

Building Agents today is challenging



Evaluation is difficult



Too many knobs

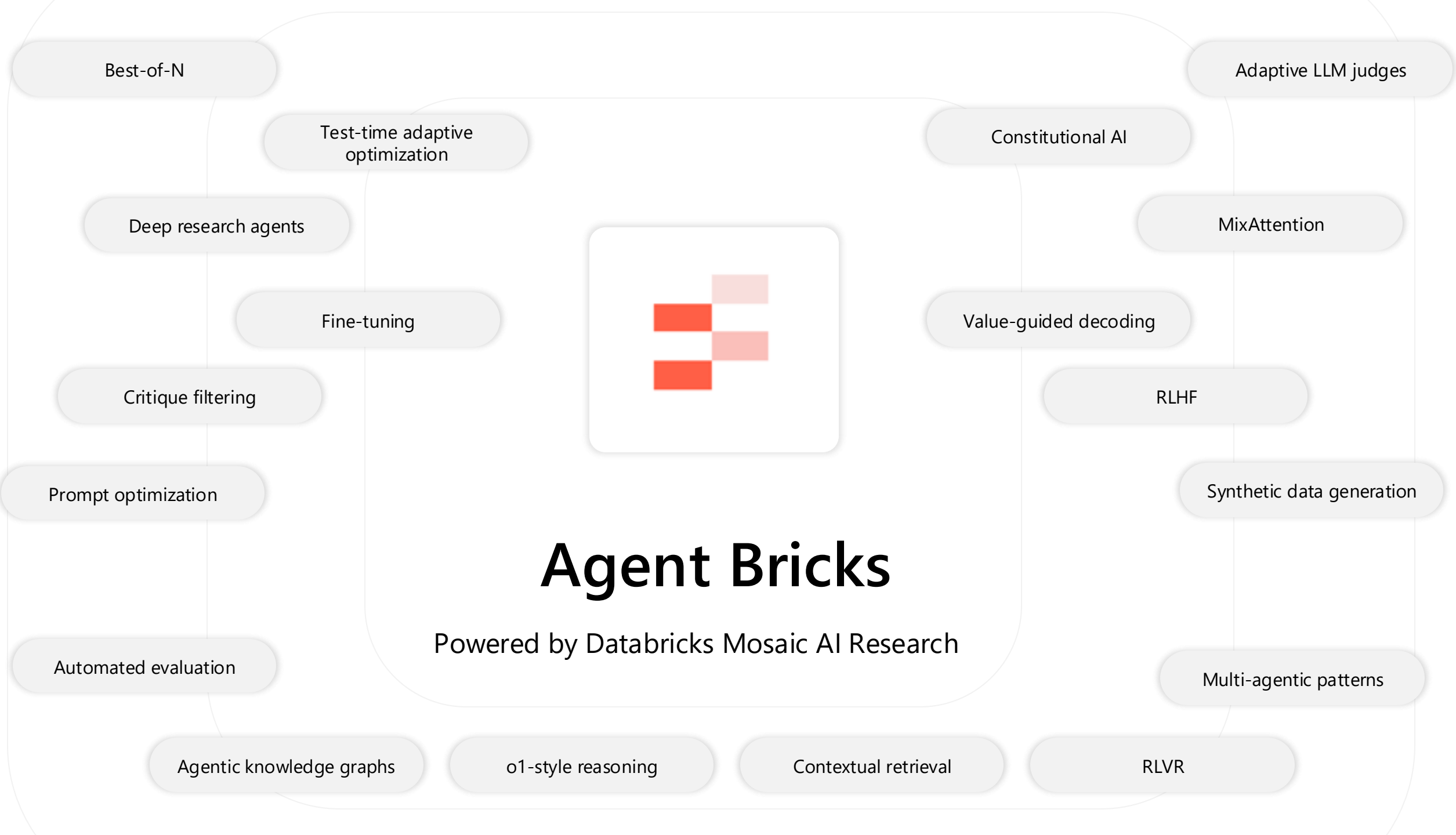


Trade-off cost versus
quality



Agent Bricks

Powered by Databricks Mosaic AI Research



Best-of-N

Test-time adaptive
optimization

Constitutional AI

Adaptive LLM judges

MixAttention

Value-guided decoding

RLHF

Synthetic data generation

Multi-agentic patterns

RLVR

Contextual retrieval

o1-style reasoning

Agentic knowledge graphs

Automated evaluation

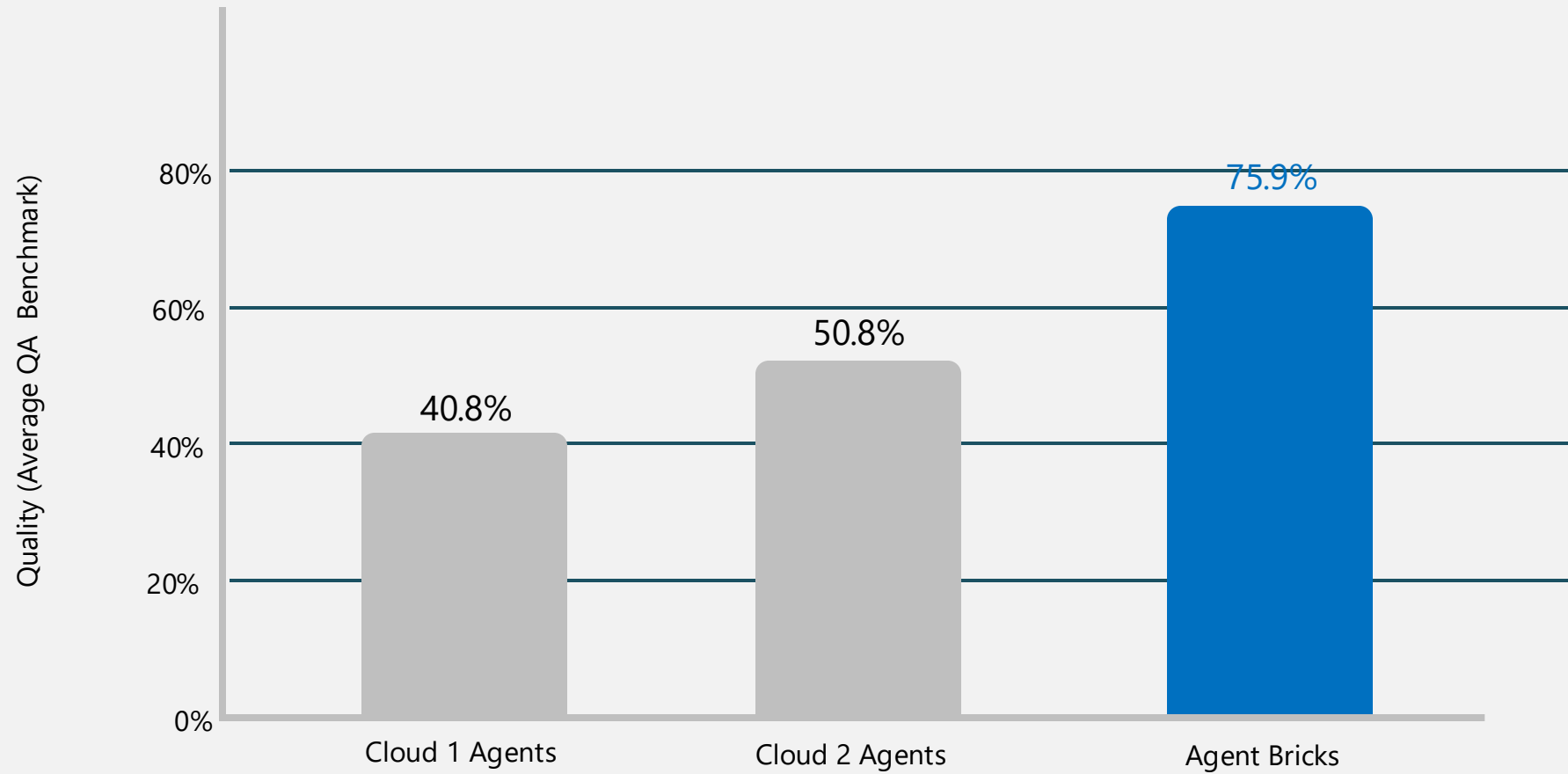
Prompt optimization

Critique filtering

Deep research agents

Fine-tuning

High Quality Out of the Box



Use Cases

Agent Bricks

Information Extraction Agent

Problem: Data is buried

Value: Turn unstructured text into structured fields like names, dates, and entities—no labeled data needed.

Custom LLM Agent

Problem: Teams rely on generic models or manual effort to tailor text for their domain.

Value: Transform text for tasks like summarization, classification, or rewriting—optimized for your industry.

Knowledge Assistant Agent

Problem: Chatbots give vague or incorrect answers

Value: Turn unstructured text into structured fields like names, dates, and entities—no labeled data needed.

Multi-Agent Supervisor

Problem: Complex tasks break down when a single model tries to do everything at once.

Value: Deliver fast, accurate answers grounded in your enterprise data.

Information Extraction

Extract key information from documents to structured outputs

Example use cases:

Legal documents or contracts

Extract key details (parties, terms, dates) weekly to verify important discrepancies.

Customer reviews

Identify and categorize product names daily to analyze trending products and customer sentiment.

Build your agent



This product is in Beta. Expect ongoing refinements and enhancements.



Recommendation

1 of 5 < >

title

Specify the exact format (e.g., title case, uppercase) and any character limits.

Dismiss

Review results

Review sample agent inputs and outputs to validate performance. Adjust model instructions or schema definitions based on your analysis to improve accuracy.

Custom LLM

Convert unstructured data into actionable insights

Example use cases:

Summarize customer service calls

Representatives can understand past customer touchpoints to improve interactions.

Generate reports

Distill key insights and save employee time.

Tailor a chat model to enterprise policies

An LLM will adhere to enterprise guidelines when replying to chats.

Create marketing material

Generate engaging headlines aligned with company styles

Build your agent

 This product is in Beta. Expect ongoing refinements and enhancements.

 **Recommendation** 10 remaining

Provide feedback to improve responses

Feedback will be used to improve quality automatically upon the next update.

Input	Markdown	Agent response	Text
-------	----------	----------------	------

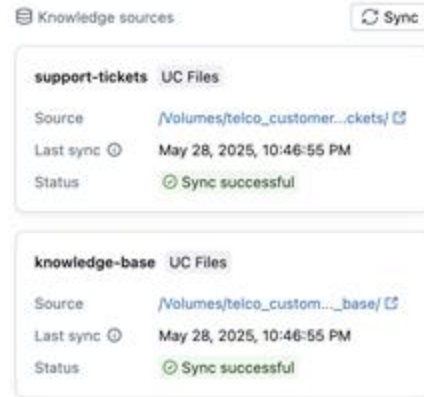
Markdown Content: Use the `DESCRIBE TABLE EXTENDED` statement to describe the table	Being an expert in using `DESCRIBE TABLE EXTENDED` entails understanding SQL syntax, practical...
See full text	See full text

Is this a good response?

☒ Yes

☐ No

Use Knowledge Assistant for the answers you need



**End-to-end ingestion
handled for you**

File parsing, ingestion, chunking, and
embedding all handled for you

Improve quality with labeled data ⓘ This product is in Beta. Expect ongoing refinements and enhancements.

ID	Question	Guidelines	Feedback records
938a9377-2a12-4a	My phone won't connect to wifi. How do I fix it?	Include how to contact support at the end of r	2 records
9b3105a-a062-4a	Is iPhone or Android better?	Be neutral on the topic of device manufacture	2 records
c444b30-a8d4-4a	I can't make calls but data works	Don't put support ticket numbers in the device Don't forget to include warnings	2 records
d7cfeae6-cab0-4d	What does Error Code AC-4032 mean?	Show them how to check the account settings	2 records

**Agent learning &
evaluation built-in**

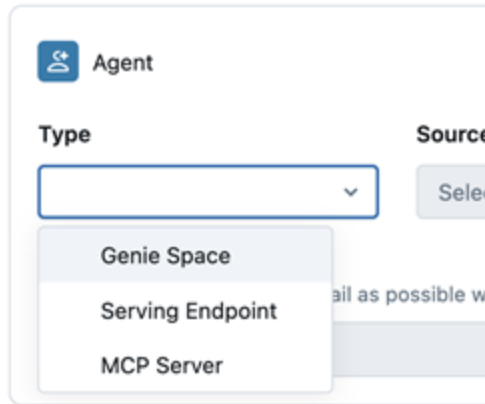
Gather feedback from SMEs &
improve the model over time



**Ready to scale endpoint
out-of-the-box**

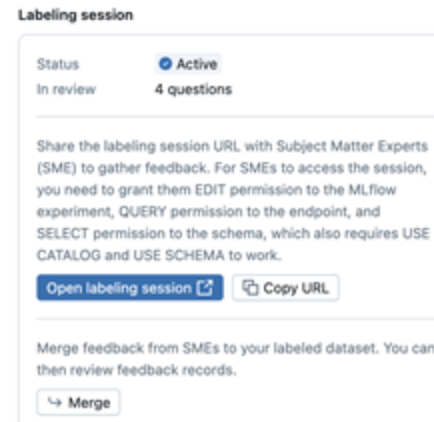
Easily embed your endpoint wherever
you need it - including a Databricks
App

Multi-Agent Supervisor makes the possibilities endless



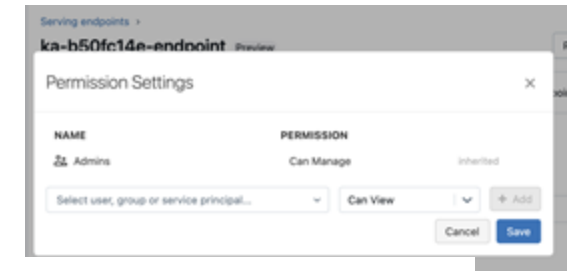
Flexible for your use case

Build agents that can reason over all your data – structured or unstructured



Agent Learning & evaluation built-in

Gather feedback from SMEs & improve the model over time



Run As Setting

When sharing the space with other users, select whether to embed the publishing user's credentials, or to use the viewer's credentials

- ☒ Viewer
- ☐ Embedded Credentials

Governance & access built-in

Leverage existing permissions with each agent so you only expose data they have access to

Coming up next...



Day 1

Module 1 - Introduction to Azure Databricks

- Azure Databricks: A Data Intelligent Platform
- Why Azure Databricks
- Decision guide: Azure Databricks vs. Microsoft Fabric

Module 2 - Migration to Azure Databricks

- Microsoft Cloud Adoption Framework for Azure
- Migration strategies
- Data landing zones
- Migration scenarios

Interactive Simulated Lab Experience

- End-to-End Streaming Pipeline with Lakeflow Declarative Pipelines in Azure Databricks

Day 2

Module 3 - Integration with Azure

- Seamless integration with Microsoft Azure services
- Connect to Azure Data Lake Storage (ADLS) Gen2 and Blob Storage
- Leverage Azure Databricks for Azure Cosmos DB Operations
- Secret management with Azure Key Vault
- Connect Azure Databricks to Azure Event Hubs

Module 4 - Integration with Microsoft Fabric and Power BI

- Data Intelligence with Azure Databricks and Microsoft Fabric
- Connect Power BI to Azure Databricks
- Integration with Azure Data Factory
- Mirroring Azure Databricks Unity Catalog

Interactive Simulated Lab Experience

- Setup and use Unity Catalog for Data Management in Azure Databricks
- Real-Time Streaming with Azure Databricks and Azure Event Hubs

Day 3

Module 5 - Integration with Azure AI Foundry

- Azure Databricks connector in Azure AI Foundry
- Mosaic AI and machine learning on Azure Databricks
- Query Generative AI model serving endpoints
- Databricks Assistant, AI/BI Genie and AI Functions on Azure Databricks
- Chat with LLMs and prototype GenAI apps using AI Playground
- Build and optimize agents on your data with Agent Bricks

Module 6 - Security and Governance

- Integrate Azure Databricks with Microsoft Purview
- Integration of Azure Databricks Unity Catalog with Microsoft Purview

Module 7 - Well-architected for Azure Databricks

- Lakehouse implementation: Principles and best practices
- Azure Databricks well-architected framework

Interactive Simulated Lab Experience

- Responsible AI with Large Language Models using Azure Databricks and Azure OpenAI
- Connect to and manage Azure Databricks in Microsoft Purview

Thank You!