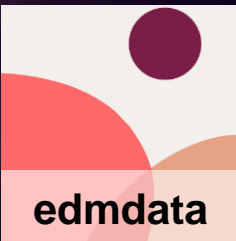


Getting Started with Azure Databricks

Wilson Mok

April 14, 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:
 - Introduction to Azure Databricks
 - Core Technologies Overview
 - Creating well-annotated metadata for your data assets
 - Explore Your Data using AI/BI Genie
- Fundamental knowledge: Basic SQL, Databricks would be beneficial but not required.

Databricks provides a Unified Data Intelligence Platform for Data and AI

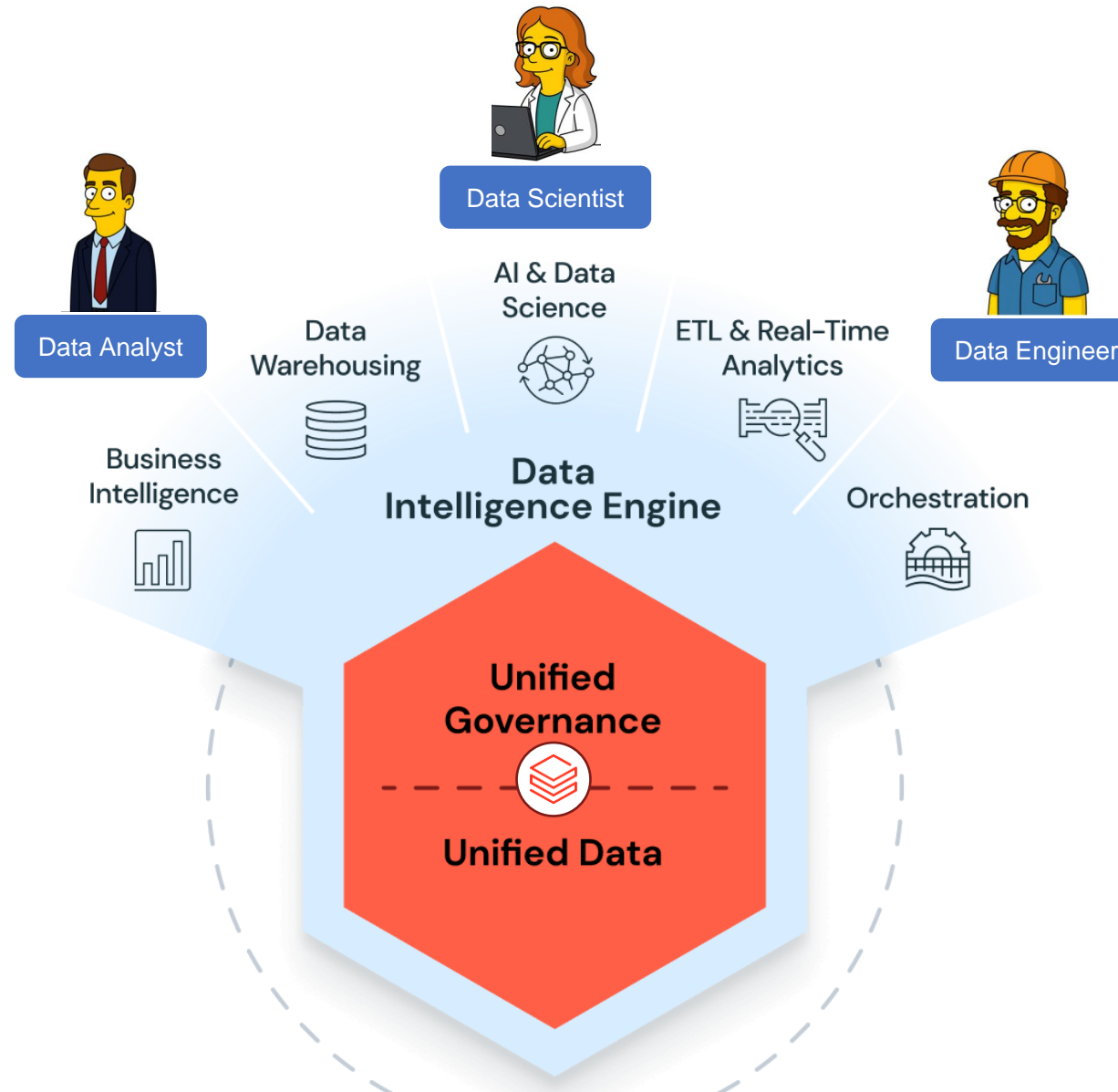
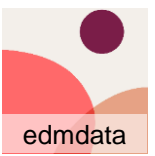
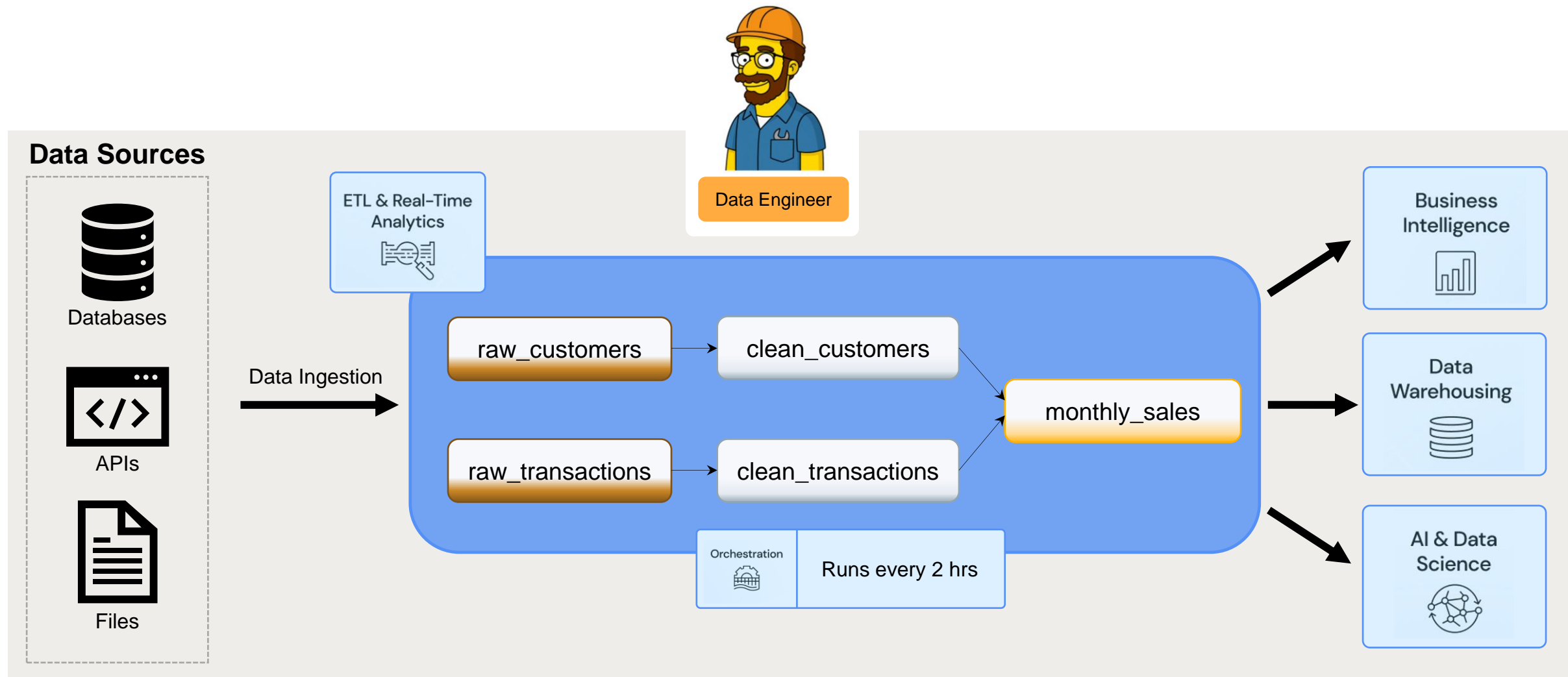
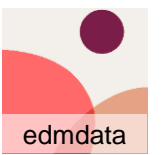
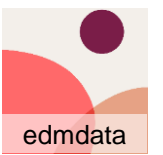


Image Credit: <https://www.databricks.com/product/data-intelligence-platform>

Data Engineers uses Databricks to build reliable data pipelines deliver fresh and actionable data to the business



Data Scientists experiment, track and deploy models faster using MLFlow, and Mosaic AI.



Data Scientist

AI & Data
Science

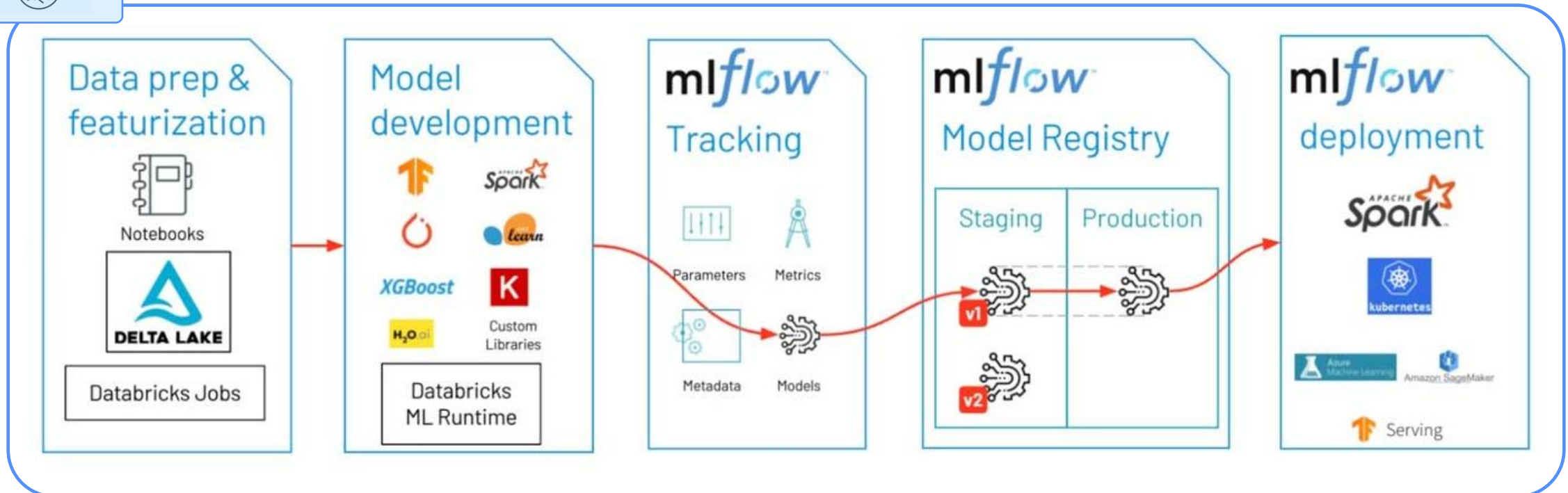


Image Credit: <https://www.databricks.com/blog/2021/06/24/three-principles-for-selecting-machine-learning-platforms.html>

Empowering Data Analysts to Query and Visualize Data using SQL

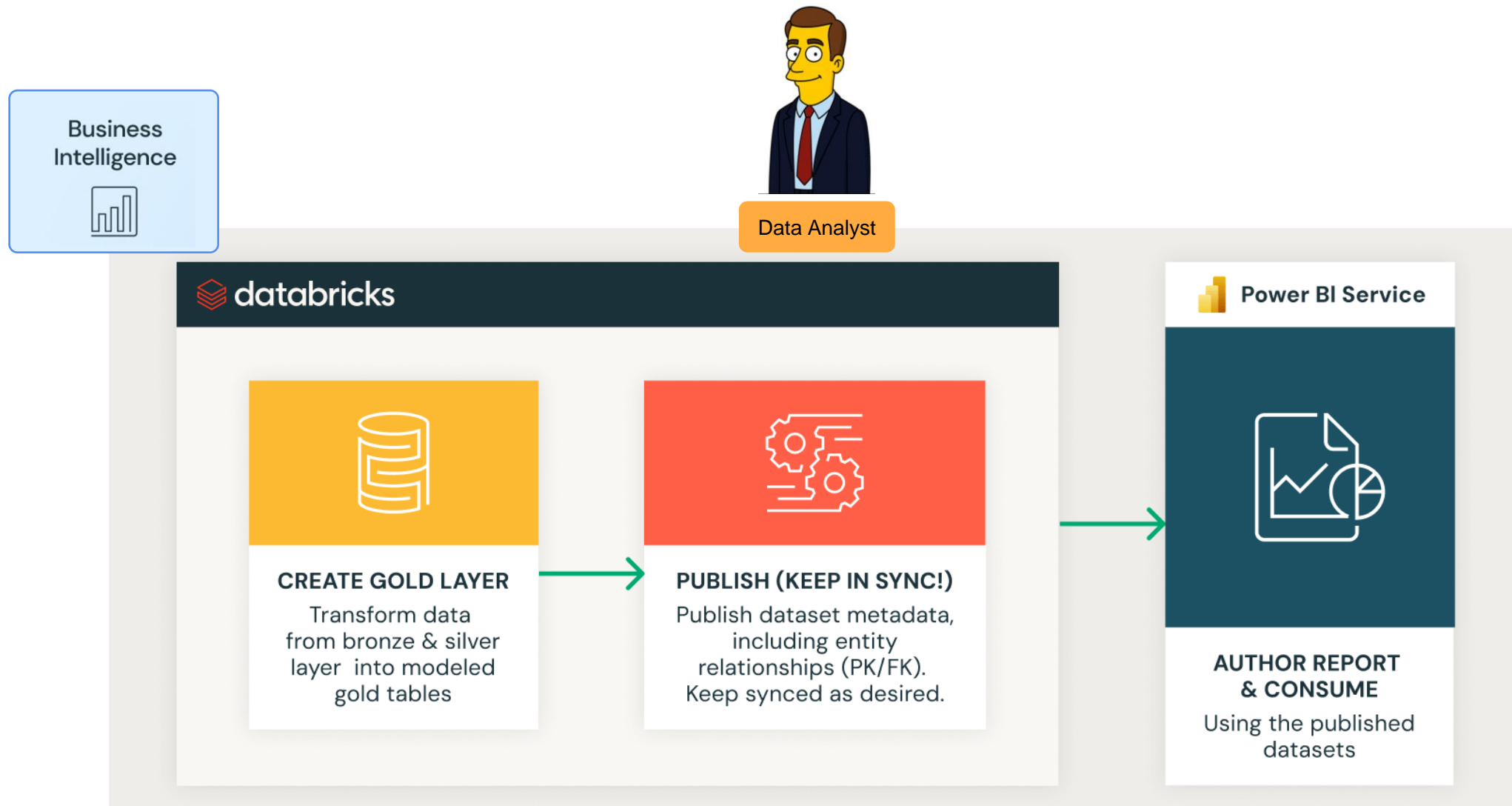
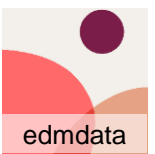


Image Credit: <https://www.databricks.com/blog/announcing-general-availability-publish-microsoft-power-bi-service-unity-catalog>

Empowering Data Analysts to Query and Visualize Data using SQL



Data Analyst

Data
Warehousing

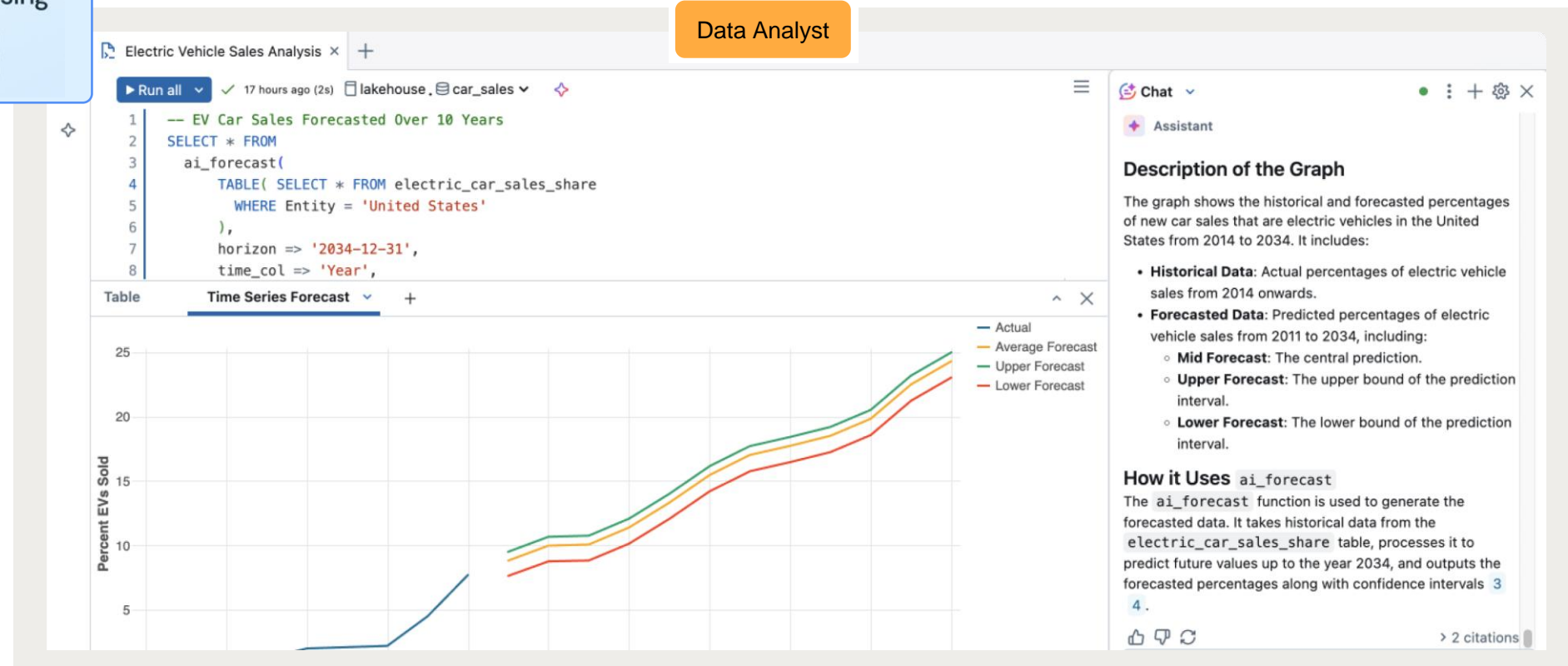
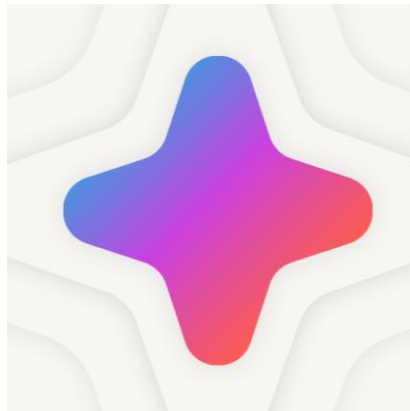
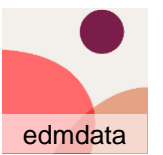


Image Credit: <https://www.databricks.com/product/databricks-sql>

Quick Tip: Databricks Assistant reduces the friction of Learning and Accelerates your productivity



Chat

Assistant

Description of the Graph

The graph shows the historical and forecasted percentages of new car sales that are electric vehicles in the United States from 2014 to 2034. It includes:

- **Historical Data:** Actual percentages of electric vehicle sales from 2014 onwards.
- **Forecasted Data:** Predicted percentages of electric vehicle sales from 2011 to 2034, including:
 - **Mid Forecast:** The central prediction.
 - **Upper Forecast:** The upper bound of the prediction interval.
 - **Lower Forecast:** The lower bound of the prediction interval.

How it Uses `ai_forecast`

The `ai_forecast` function is used to generate the forecasted data. It takes historical data from the `electric_car_sales_share` table, processes it to predict future values up to the year 2034, and outputs the forecasted percentages along with confidence intervals 3 4.

> 2 citations

Image Credit: <https://www.databricks.com/product/databricks-sql>

Key technologies that powers Databricks: Delta Lakehouse, Unity Catalog, Serverless Compute

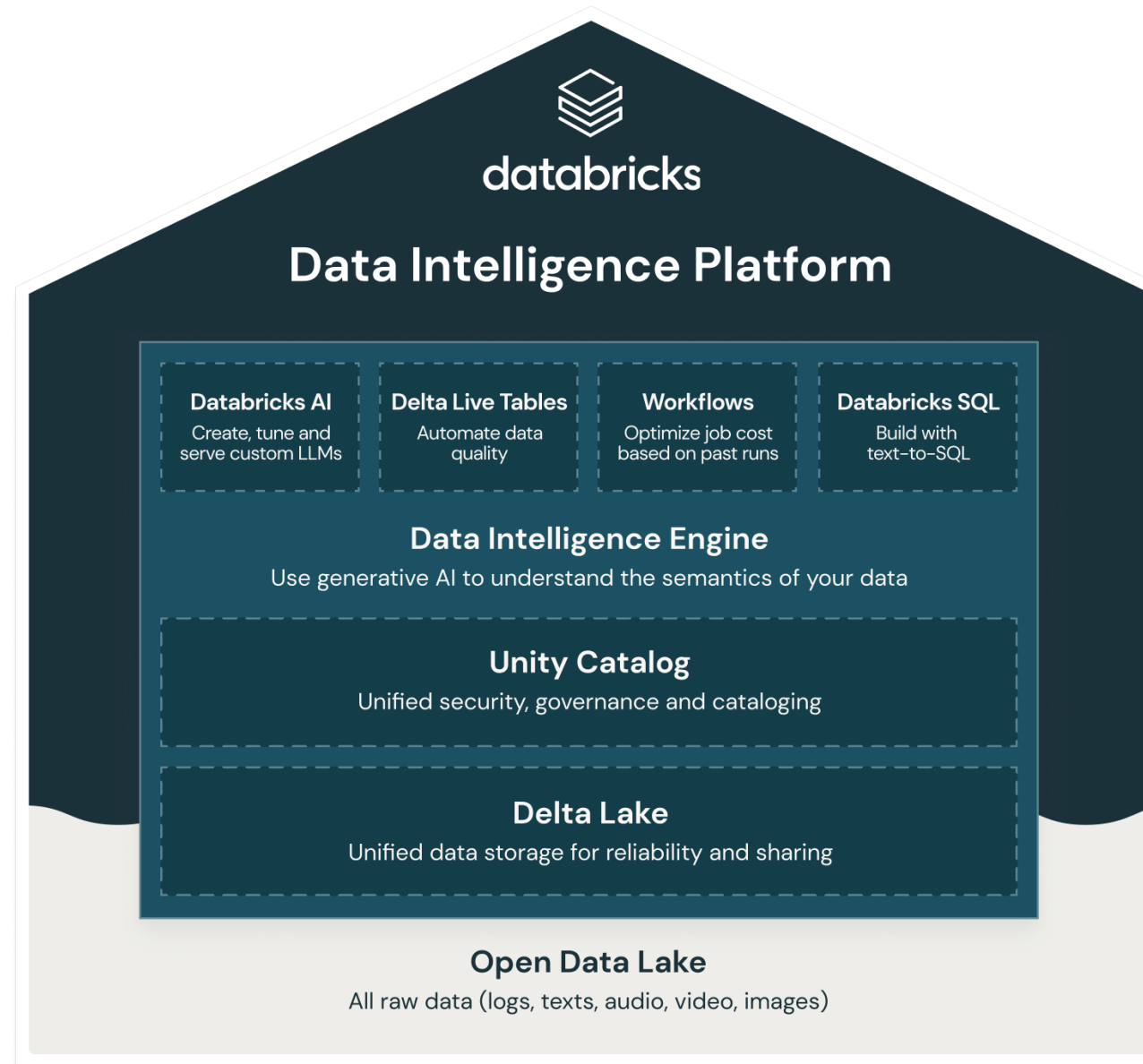
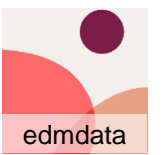
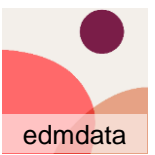


Image Credit: <https://www.databricks.com/product/data-intelligence-platform>

Demo 1 – Uploading data and Creating annotation for our dataset



+ New

Workspace

Recents

Catalog

Workflows

Compute

Marketplace

SQL

SQL Editor

Queries

Dashboards

Genie

Alerts

Query History

SQL Warehouses

Add data >

Create or modify table from file upload

sample_customers_data.csv uploaded 36.54KB

Create new table

Preview mode

Catalog demo_catalog

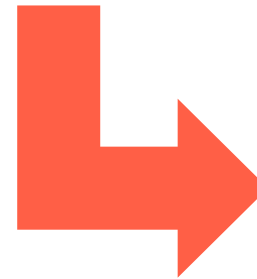
Schema sample_genie

Table name sample_customers_data

Advanced attributes

Previewing 50 rows, 4

| CustomerID | CustomerName | PersonID | ContactPersonName |
|------------|--------------------------------------|----------|----------------------|
| 181 | Tailspin Toys (Heilwood, PA) | 1361 | Aahlada Thota |
| 89 | Tailspin Toys (Victory Gardens, ...) | 1177 | Aakarsha Nookala |
| 451 | Wingtip Toys (Smoot, WV) | 2101 | Aakriti Bhamidipati |
| 832 | Aakriti Byrraju | 3032 | Aakriti Byrraju |
| 423 | Wingtip Toys (Baldwin City, KS) | 2045 | Aamdaal Kamasamudram |



Catalog Explorer > demo_catalog > sample_genie >

sample_customers_data

Open in a dashboard Create

Overview Sample Data Details Permissions History Lineage Insights Quality

About this table

Owner Demo - Business User

Data source Delta

Popularity

Last updated 36 seconds ago

Size 28.6KiB, 1 file

Tags

Add tags

Row filter

Add filter

AI Suggested Description

The sample_customers_data table contains information about customers, including their unique CustomerID, CustomerName, PersonID, and ContactPersonName. This data is crucial for tracking customer interactions, managing relationships, and providing personalized services. The table serves as a central repository for customer details, facilitating targeted marketing campaigns and efficient communication strategies. Understanding customer preferences and behavior is essential for driving business growth and enhancing customer satisfaction.

Accept Edit

Send feedback

Filter columns...

AI generate

| Column | Type | Comment | Tags | Column masking rule |
|--------------|--------|---------|------|---------------------|
| CustomerID | bigint | | | |
| CustomerName | string | | | |

Q & A 1 (5 minutes)

Having a Conversation with your data using AI/BI Genie



R Richard Tomlinson

Show me average price by model

Genie

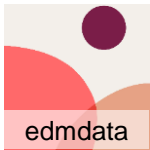
This query calculates the average unit price for each product model in the sales data, and then presents the results in descending order based on the average price.

| | ^A _C ModelName | 1.2 AveragePrice |
|---|-------------------------------------|------------------|
| 1 | Road-150 | 2146.96 |
| 2 | Mountain-100 | 1919.94 |
| 3 | Road-250 | 1408.24 |
| 4 | Touring-1000 | 1365.19 |
| 5 | Mountain-200 | 1301.73 |

Ask your question...



Demo 2 – Genie in Action



Sample - Sales Data ☆

Demo - Business User

What is the average revenue for each Month in Q1 2024?

Genie

This analysis provides the average revenue generated from invoices for each month within the first quarter of 2024. It combines invoice details from the invoice records with the corresponding total prices from individual invoice lines, focusing specifically on invoices created between January and March 2024. The results are then summarized to show the average revenue for each month. [sample_invoices_data](#) [sample_invoice_lines_data](#)

| | 1.1 Month | 1.2 AverageRevenue |
|---|-----------|--------------------|
| 1 | 2 | 837.68 |
| 2 | 1 | 848.24 |
| 3 | 3 | 822.33 |

filtered 3 rows/3 rows

Average Revenue for Each Month in Q1 2024

Add as instruction

Show code

Context Settings

Data

Instructions

Data

Link existing or upload new data to start discovering insights.

| Name | Type |
|---|-------|
| <div>sample_invoices_data</div> <div>demo_catalog.sample_genie</div> | Table |
| <div>sample_invoice_lines_data</div> <div>demo_catalog.sample_genie</div> | Table |
| <div>sample_customers_data</div> <div>demo_catalog.sample_genie</div> | Table |

After all this, we have run through...

- Databricks Unified Data Intelligence Platform bring Data Engineers, Data Scientist and Data Analysts together into one collaborative environment.
- Everyone can collaborate in the same notebook with their preferred programming language or use traditional SQL for queries and quick visualizations.
- Databricks Assistant and AI/BI Genie lower the barriers to learning, building and exploring data power by AI.
- A strong foundation is created through Unified Data (Delta Lake) and Unified Governance (Unity Catalog) enable trusted, governed access to your data assets.
- Data Intelligence Engine extend Databricks into generative AI and conversational analytics.
- The code is in my Github repo:
 - <https://github.com/wilson-mok/demo/...>

Training for Fabric, AI and Databricks



Now – May 28: [AI Skills Fest](#)

Microsoft | AI Skills Fest | Sessions | Register now

Microsoft AI Skills Fest

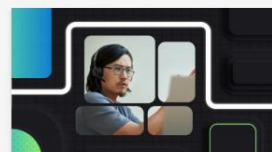
Now through May 28, 2025

Unlock the future with 50 days of AI discovery and learning

[Register now](#)

Deepen your skills with our lineup of curated experiences

Top picks by audience

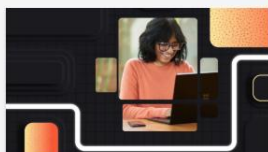


Tech professionals

Master the skills to create AI Agents

Explore how to build AI-driven agents using Azure AI Foundry and GitHub Copilot, integrating them into your applications and using advanced model fine-tuning techniques.

[Get started >](#)

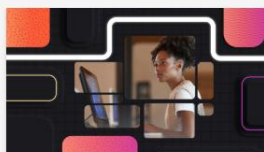


Business professionals

AI for managers by Microsoft and LinkedIn

Managers at all levels of the organization will learn practical ways to enhance their effectiveness with generative AI. This content leads to a professional certificate.

[Get started >](#)



Students

Minecraft Education - Fantastic Fairgrounds

Welcome to Fantastic Fairgrounds, an immersive Minecraft Education experience designed for learners ages 8 and older to build a strong foundation in AI Literacy. This single-player map is part of a broader AI Literacy progression.

[Get started >](#)



Public Sector

AI for public sector upskilling challenge

Equip yourself with AI skills to enhance efficiency, decision-making, and service delivery as a public servant. Learn how generative AI can be utilized in the public sector, responsible AI principles, and modern security operations to harness AI's transformative power.

[Get started >](#)

Now – Apr 30: [Virtual Learning Festival](#)



VIRTUAL EVENT

Databricks Learning Festival

Upskill yourself and be eligible for certification discount voucher

9-30 April 2025



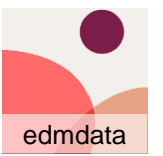
Q & A 2

Thank you



LinkedIn

What are the key difference between Azure Databricks and Microsoft Fabric?



| Feature/Aspect | Azure Databricks | Microsoft Fabric |
|-----------------------------|--------------------------------|--------------------------------------|
| Founded | 2013 (Azure: 2018) | 2023 |
| Deployment Model | PaaS – Infra setup is required | SaaS – No Infra setup is required |
| Cloud Agnostic | Azure, AWS, GCP | Azure Only |
| Governance framework | Unity Catalog | Microsoft Purview |
| Technology | SQL, Spark, Lakehouse | T-SQL, Spark, Lakehouse |
| Storage | Delta Lake (Open Source) | OneLake (integrated with Delta Lake) |
| Development style | Code first | Supports Low-code and Code |