
#FABCONSQLCON2026
FABCON
Microsoft Fabric
COMMUNITY CONFERENCE
SQLCON
Microsoft SQL
COMMUNITY CONFERENCE
ATLANTA MARCH 16 - 20, 2026

AI in Action: Microsoft Fabric for Data Science

Exploring Data Science using Microsoft Fabric to build AI technologies and building scalable intelligent systems

By Sadiq Ahmed
Quantum Technology Inc.

Linkedin.com/in/sadiqhahmed/

Session Outline

- Holistic Overview
- When Fabric is the better choice
- Why Fabric is effective for Data Science & AI
- Key Concepts
- AI & DS
- Kickstart Your Data Science Journey
- Data Ingestion
- Wrangling (Data Preparation)
- Exploration & Visualization
- Modeling
- Evaluation
- Deployment
- Things to know
- Take Away
- Thank You & Q&A

Holistic overview

This presentation is a Holistic Overview of the Fabric primary for Data Science & AI

When Fabric is the better choice

When

You want **SaaS-first design, tight Microsoft ecosystem integration,** and **unified architecture**.

You want **one platform** for data engineering, data science, ML, BI and Governance.

You want **serverless compute** with no cluster management.

Your team includes **multiple roles** like analysts and citizen data scientists.

You want integration with **Power BI**.

You want **fast time to value** with minimal setup & infrastructure management.

And More …

Why Fabric is effective for Data Science & AI

WHY

Its strength comes from combining **OneLake, Spark, ML tooling,** and **AI functions** into a single SaaS experience.

Deep integration with **Azure AI** and **OpenAI**

End-to-end workflow in one place

Low-code/no-code and **pro-code** together

Governance and **responsible** AI built in

Key Concepts of Microsoft Fabric

Key Concepts !

- **Fabric (Platform)**

A **unified SaaS platform** for analytics, data science, and AI, offering a single environment for all workloads.

- **OneLake (Centralized Storage Layer)**

A **unified data lake storing** all organizational data with one logical view, reducing duplication across regions and clouds.

- **Azure Data Lake Storage (ADLS) Foundation**

Built on ADLS, supports open formats (**Delta, Parquet, CSV, JSON**), and ensures scalability and compatibility.

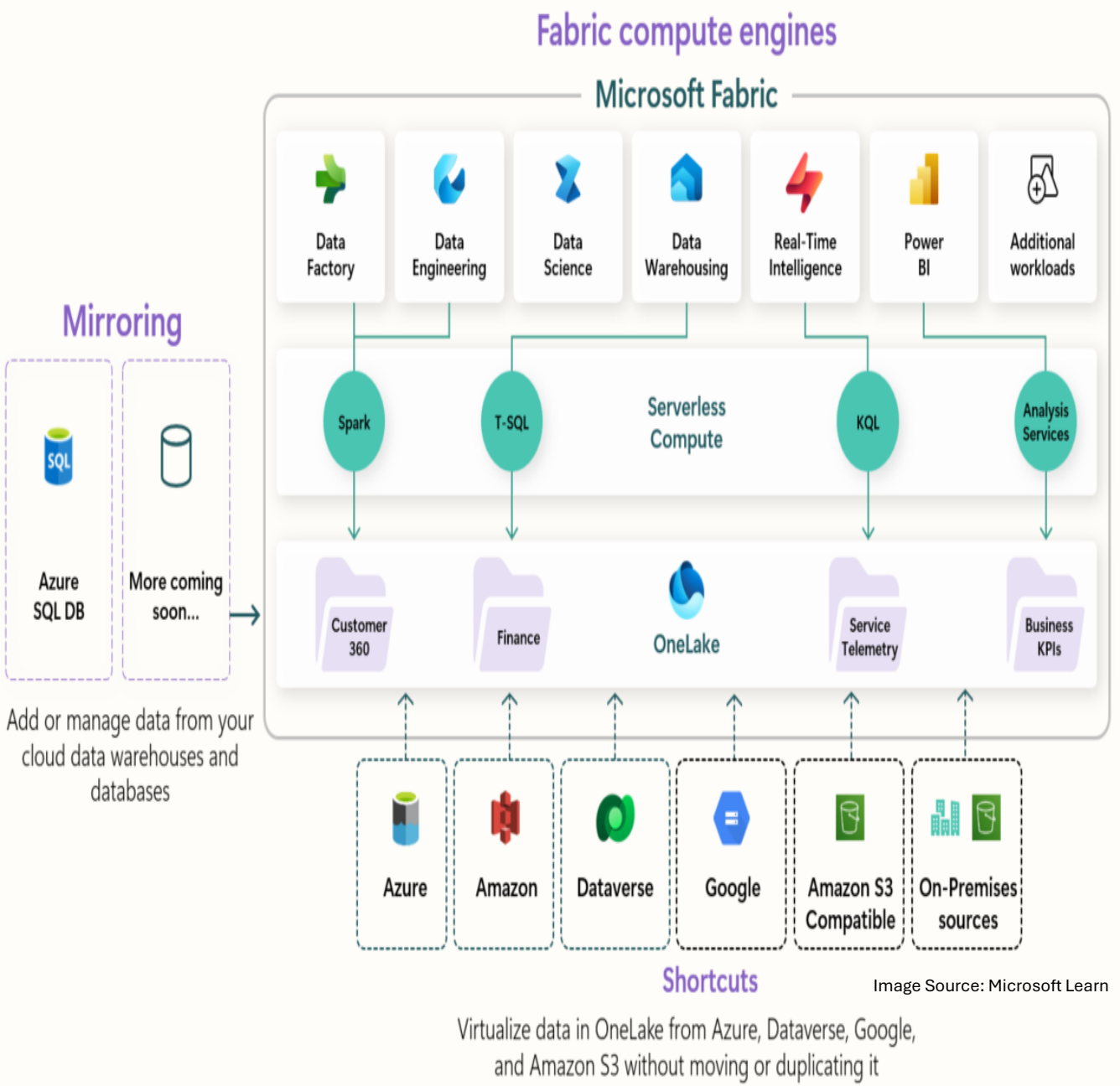


Image Source: Microsoft Learn

Key Concepts !

- **Lakehouse (Structured Container in OneLake)**

Organizes data as tables and files, blending data lake flexibility with warehouse schema, optimized for analytics and ML.

- **Compute Engines (Integration Layer)**

All compute engines (Data Engineering, Data Science, Real-Time Analytics, Power BI) store data in OneLake using **delta-parquet format** for seamless interoperability.

- **Shortcuts (External References)**

References to external files or storage outside OneLake enable access without copying and keep data synchronized.

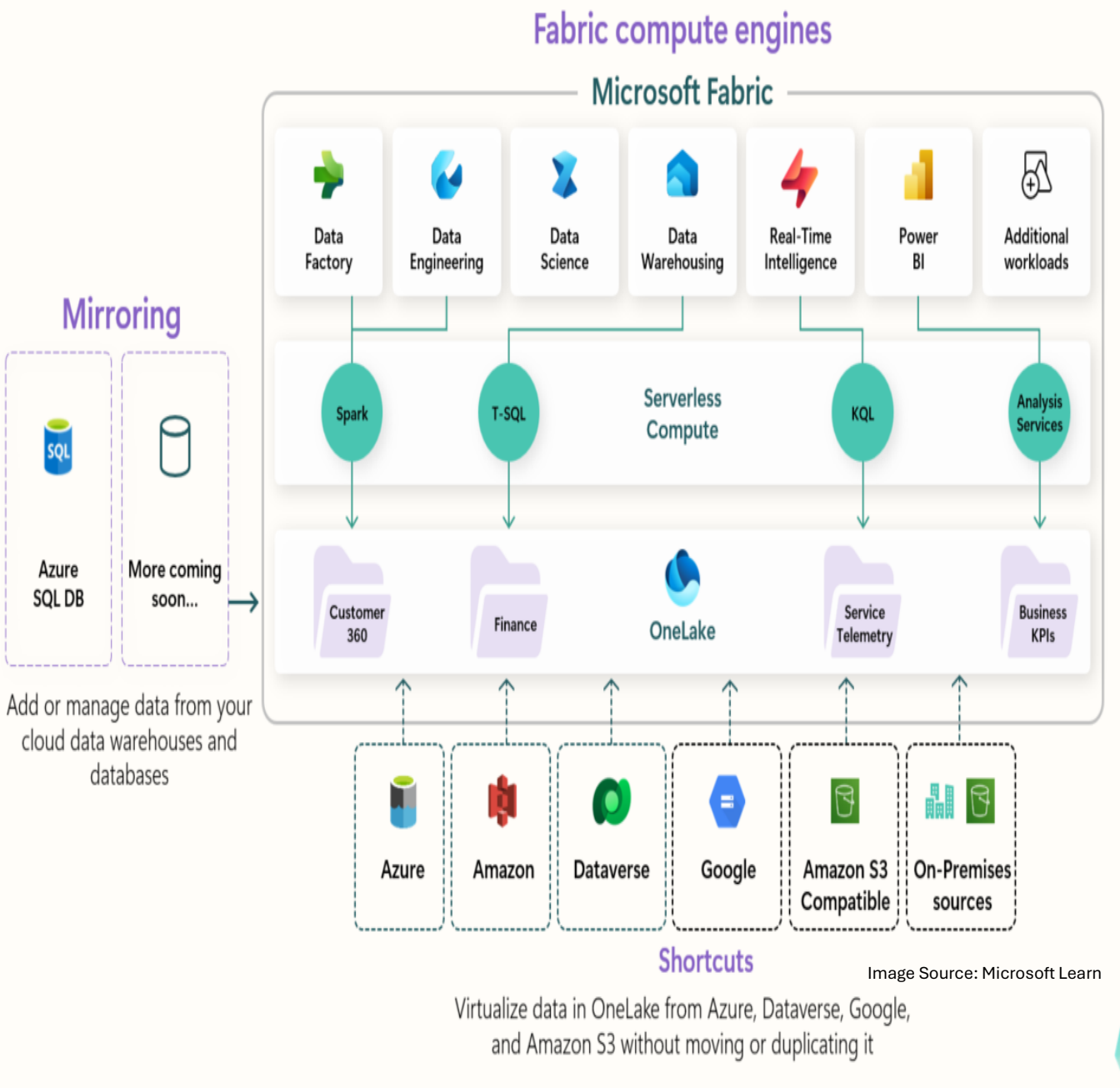


Image Source: Microsoft Learn

Key Concepts ! !

➢**Mirroring (External References)**

Mirroring in Fabric is near real-time **replication** of external databases into OneLake so you can query and analyze them directly.

➢**Workspace**

A workspace in Microsoft Fabric is a secure, **collaborative container** that organizes and manages all the data and analytics items for a specific project or team

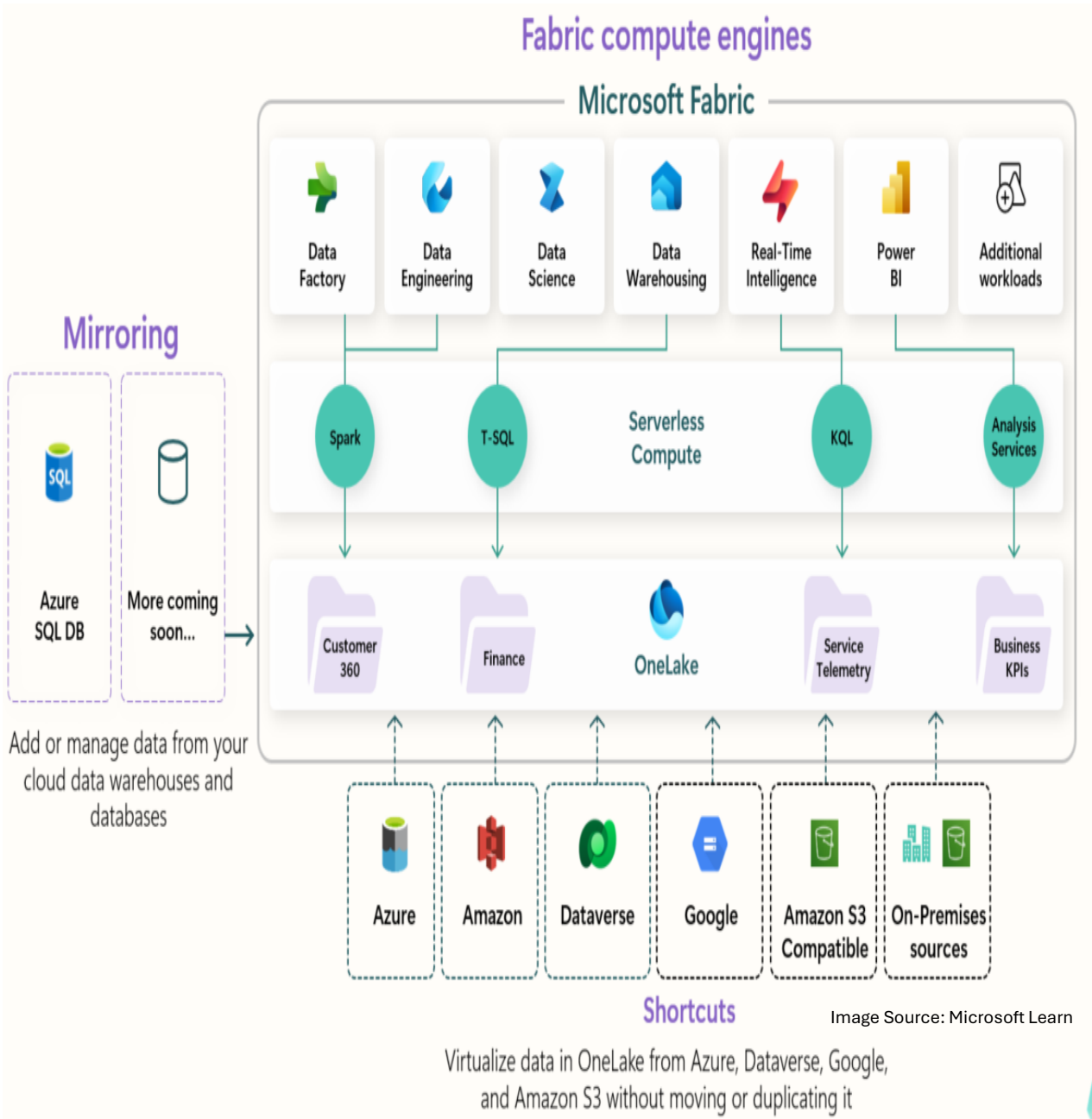


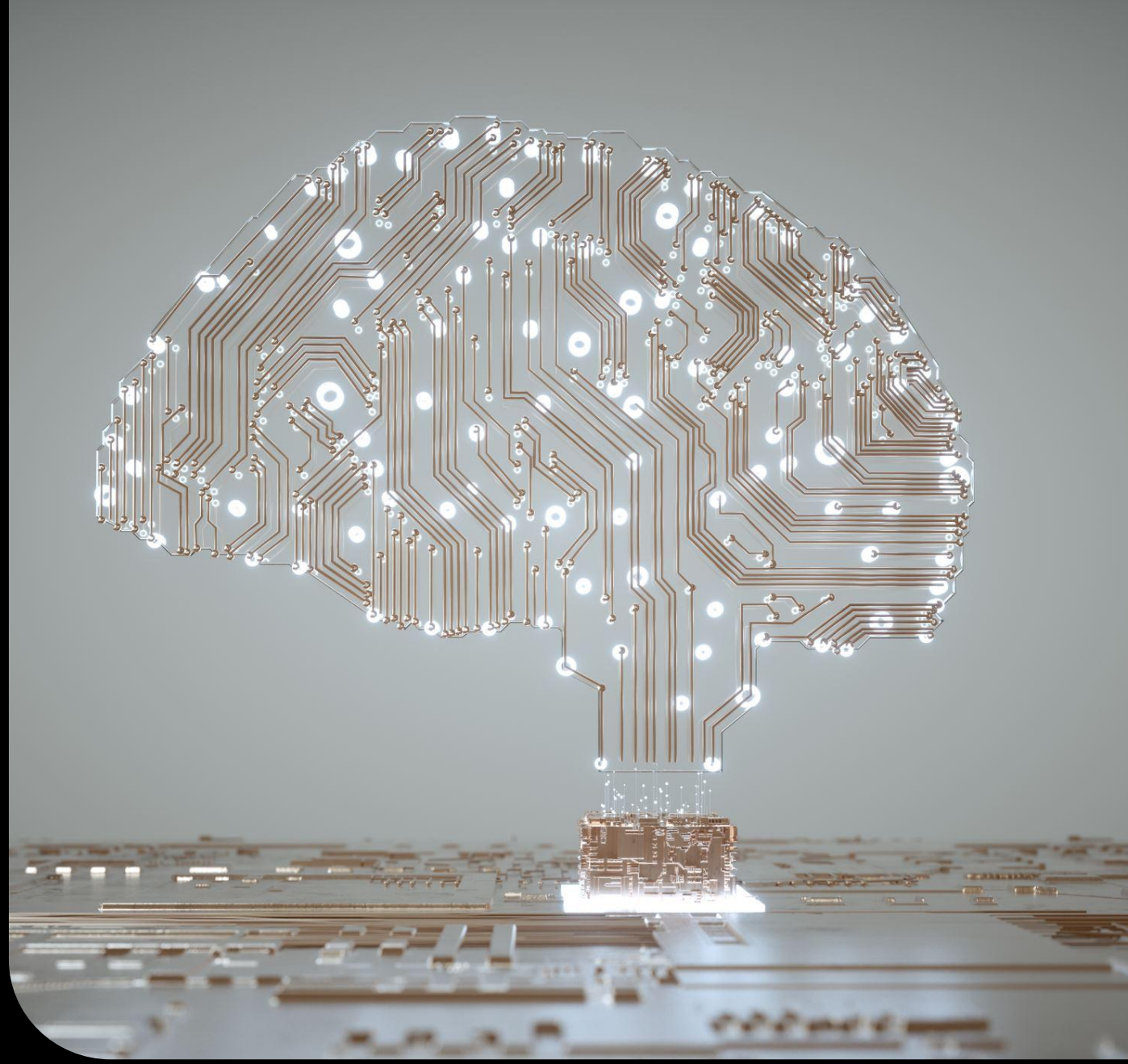
Image Source: Microsoft Learn

KEY TAKEAWAY:

CONCEPTS

- ✓ **Fabric** is the platform.
- ✓ **OneLake** is the unified storage layer.
- ✓ **Lakehouse** (Structured Container in OneLake)
- ✓ **ADLS + formats** are the foundation.
- ✓ **Compute engines** integrate seamlessly.
- ✓ **Shortcuts** extend access to external data.
- ✓ **Mirroring** : Replication of external system

Data Science & Artificial Intelligence



Relation : DS & AI

Data Science workflow = foundation (data prep, modeling, evaluation).

AI workflow = extension
(using ML + advanced AI capabilities).

In Fabric, you typically start with the Data Science steps and then expand into AI workflows if your project requires intelligent automation, generative AI, or advanced ML.

Data Science + AI/ML lifecycle Project

Generally, follows these steps:

Business stage

- Business problem definition (Business understanding)

Technical stages

- Data collection
- Data preparation
- Data exploration (**E**xploratory **D**ata **A**nalyst)
- Modeling
- Model evaluation
- Model deployment + Monitoring & Inference

Frameworks/Lifecycle/workflow /methodologies..

CRISP-DM
OSEMN
SEMMA
KDD Process
TDSP (Microsoft)
ASUM-DM (IBM)
AI-DSF
ASEMIC
ML Lifecycle (Google)
MLOps Lifecycle (Google/AWS/Azure)
NIST AI RMF
DataOps Lifecycle

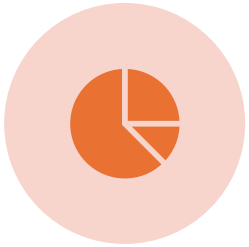
In General, Stages in Microsoft Fabric

Data Ingestion →

bringing raw data from OneLake, databases, or external sources

Data Wrangler →

cleaning, transforming, and preparing datasets for analysis



Data Exploration & Visualization →

charts, summaries, and quick insights

Modeling →

training ML models (classification, regression, clustering, etc.)

Evaluation →

testing model accuracy and performance

Deployment →

integrating models into Fabric pipelines for production use

Backbone of workflow

Six steps are for **Data Science in Fabric**, but they also serve as the **backbone of AI workflows**.

AI simply layers additional capabilities on top of them.

Kickstart Your Journey

ROLES INVOLVED:

- ✓ Data Engineer
- ✓ Data Analyst
- ✓ Data Scientist
- ✓ Machine Learning Engineer
- ✓ AI Engineer
- ✓ MLOps Engineer
- ✓ Database Administrator
- ✓ Data Architect
- ✓ Solution Architect
- ✓ Data Steward
- ✓ Governance/Compliance Officer
- ✓ Fabric Administrator / Workspace Admin
- ✓ Business Analyst
- ✓ BI Developer
- ✓ Real-Time Analytics Engineer
- ✓ DevOps Engineer
- ✓ Product Owner
- ✓ .....

USE CASES : Examples

AI-Driven Customer Insights

Organizations use AI to segment customers, enabling personalized marketing and enhanced customer experiences.

Predictive Analytics

Predictive analytics forecast like sales trends, helping businesses optimize inventory and increase revenue.

Operational Efficiency Gains

AI-driven insights enhance operational workflows, leading to improved efficiency and competitive advantage.

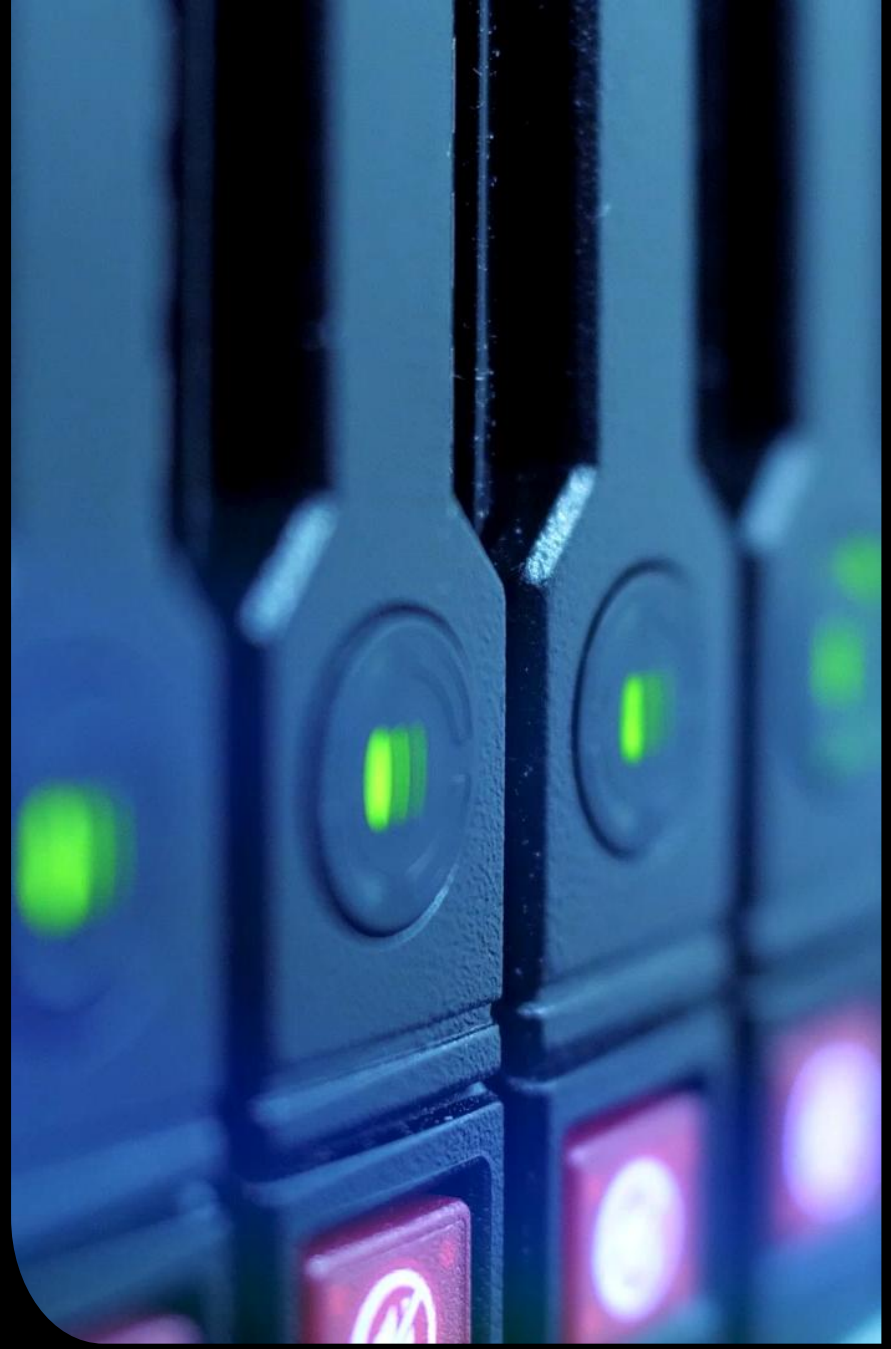
Stages in Microsoft Fabric

- **Data Ingestion** → bringing raw data from OneLake, databases, or external sources
- **Data Wrangler** → cleaning, transforming, and preparing datasets for analysis
- **Data Exploration & Visualization** → charts, summaries, and quick insights
- **Modeling** → training ML models (classification, regression, clustering, etc.)
- **Evaluation** → testing model accuracy and performance
- **Deployment** → integrating models into Fabric pipelines for production use

Demo

- Workspace
- Menu

Data Onboarding in Fabric



Data Ingestion

Data Ingestion

Ingestion data into OneLake (physically or virtually) so Fabric can analyze, transform, and use it.

Ways to Ingest Data ?

Pipelines
Dataflows
Notebooks
Real-time streams
Direct file uploads
Shortcuts (virtualization)
SQL loading
APIs
More..

Ways to Ingest Data ?
Code-Free Ingestion

Dataflows Gen2 (Power Query)

Data Pipelines (Data Factory in Fabric)

Copy Activity (inside Pipelines)

Drag-and-drop **file upload** into the Lakehouse

Shortcuts (virtualize external storage like ADLS, AWS S3, Google Cloud Storage)

Eventstreams (real-time ingestion)

Ways to Ingest Data ?
Code-Based Ingestion

Spark Notebooks
(PySpark, Scala, Spark SQL)

Spark Jobs

Pandas / Python inside notebooks

Delta Lake APIs

COPY INTO (SQL Warehouse)

Bulk Insert (SQL Warehouse)

Ways to Ingest Data ?
Streaming Ingestion

Eventstreams
(real-time ingestion into OneLake or KQL)

Kusto / Eventhouse connectors

Event Hubs / Kafka via Eventstream

What happens after Ingestion?

1. Fabric transforms raw information into **OneLake** (**Delta tables)**
2. Making them **searchable** via SQL and Spark,
3. Allows to build **Semantic Models**, and
4. Supporting **analytics, reporting, and machine learning** applications built on the underlying data.

Key Takeaway:
DATA INGESTION

Process:
Collect raw data from sources
(databases, APIs, files, external systems).

Goal:
Bring all relevant data into Fabric for analysis.

Fabric Tools/Features:

OneLake → unified data lake for all organizational data

Lakehouse → structured container for analytics-ready data

Data Factory pipelines → ETL/ELT workflows to move and transform data

More…

Connectors → More than **200** native connectors

Demo

Data Preparation with Data Wrangler

Wrangling (Data Preparation)

IMPORTANCE OF CLEAN, WELL-PREPARED DATA

Data Cleaning

Cleaning data removes noise and inconsistencies, essential for accurate machine learning results.

Data Transformation

Transforming data prepares it for modeling, boosting machine learning performance.

Model Performance Improvement

Well-prepared data enhances model accuracy and trustworthiness.

What Data Wrangler Does in Microsoft Fabric

Exploratory data analysis: Displays data in a grid-like interface with dynamic summary statistics.

Data cleaning operations: Built-in functions for handling missing values, duplicates, and formatting issues.

Transformations: Apply operations like filtering, grouping, or feature engineering with just a few clicks.

Code generation: Produces pandas or PySpark code that can be saved back into notebooks for reproducibility.

Visualization support: Offers charts and plots to quickly understand distributions and relationships.

Why It Matters

Accelerates preprocessing:
Saves time compared to manual coding.

Ensures reproducibility:
Generated code can be reused and shared.

Bridges raw data to ML models:
Clean data is essential for accurate predictions.

Scales with Fabric:
Works with both small datasets (pandas) and big data (PySpark).

Bridge

Data Wrangler in Microsoft Fabric is the **bridge between raw data** and **machine learning models**, making the data science process smoother, faster, and more reliable.

Key Takeaway:
WRANGLING (DATA PREPARATION)

Process:
Clean, transform, and structure the data.

Goal:
Make messy raw data usable for analysis and modeling.

Fabric Tools/Features:

Data Wrangler → interactive tool in notebooks for cleaning and transformation

Spark (PySpark) → distributed processing for large-scale data prep

Demo

Notebook Magic: Explore, Visualize, Discover

Exploration & Visualization

Understanding Data Through Visual Representation

Visualize

Visualize Data Relationships

- Explore connections within your data by using visualization tools.

Libraries

Use Key Libraries

- Take advantage of libraries like **matplotlib** and **seaborn** in Fabric notebooks to plot correlations, outliers, and trends.

Enhance

Enhance Analysis with Embedded Visuals

- Incorporate visuals directly into your workflow to make **analysis clearer** and highlight important **insights**.

Understanding Data Through Visual Representation

- Code Execution Integration
- Documentation and Explanation
- Handling Missing Values
- etc

KEY TAKEAWAY: EXPLORATION & VISUALIZATION

Process:
Analyze distributions, correlations, and anomalies.

Goal:
Understand the data before modeling.

Fabric Tools/Features:

Notebooks → Python/Spark for exploratory analysis

Visualization libraries → matplotlib, seaborn, Plotly

Power BI integration → interactive dashboards and reports

Demo

Train Smarter: MLflow in Action

Modeling

Train , Track & Manage

Microsoft Fabric enables data scientists to **train, track,** and **manage** machine learning models with notebooks and framework/libraries.

Training models inside Fabric

- Fabric notebooks use **Spark** compute, supporting **PySpark** and **Python**.

- Popular ML frameworks/libraries include:
  a. **Scikit-learn:** For classification, regression, and clustering models.
  b. **PyTorch and TensorFlow**: For deep learning in NLP and computer vision.
  c. **SynapseML**: For scalable machine learning pipelines.

- All works with Python and **Pandas DataFrames**.

Tracking experiments with MLflow

- Fabric supports **MLflow Experiments**, making it easy to log
  - parameters,
  - metrics,
  - artifacts, and
  - model versions.
- This streamlines run **comparison**, result reproduction, and team collaboration.
- MLflow acts as the main platform to **track** model performance over time.

Managing models in Fabric

- After training, models can be registered and versioned with MLflow's model registry.
- Fabric keeps models, code, data, and experiment history unified in one environment.

Key Takeaway: MODELING

Process:
Train machine learning models (classification, regression, clustering, etc.).

Goal:
Build predictive or descriptive models.

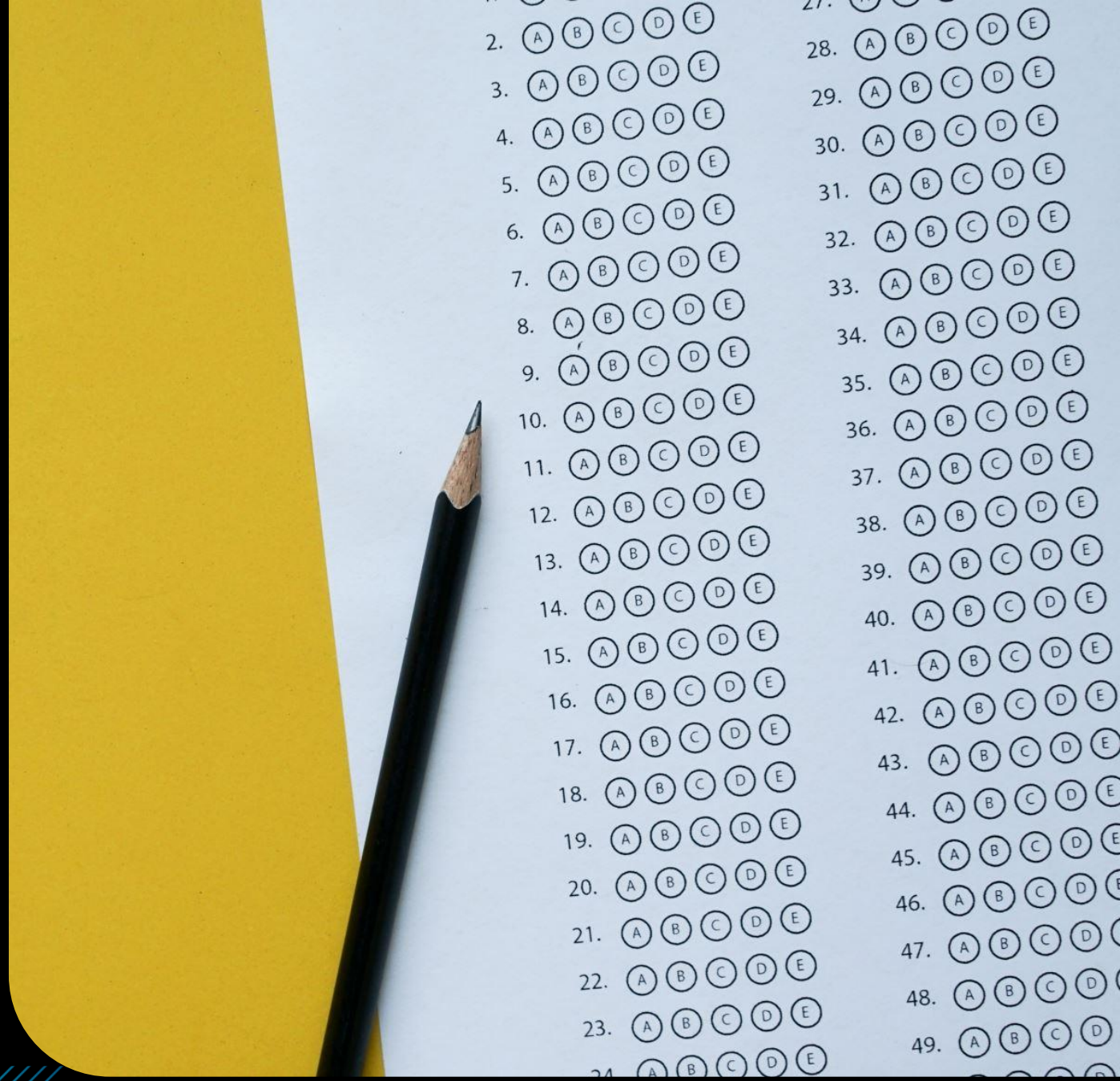
Fabric Tools/Features:

MLflow integration → experiment tracking and model management

Fabric notebooks → scikit-learn, TensorFlow, PyTorch, Spark MLlib

Demo

Model Evaluation with MLflow Metrics



Evaluation

BEST PRACTICES: REPRODUCIBILITY, EXPERIMENT COMPARISON

Ensuring Reproducibility

MLflow captures code, data, and parameters to ensure experiments can be reproduced accurately and reliably.

Experiment Comparison

Comparing different experiments allows selecting the best performing models confidently for deployment.

Key Takeaway: EVALUATION

Process:
Test models with metrics (accuracy, precision, recall, RMSE, F1 score).

Goal:
Validate performance and select the best model.

Fabric Tools/Features:

MLflow metrics tracking → compare runs and performance

Evaluation libraries → scikit-learn metrics, Spark MLlib evaluators

Demo

Predict at Scale: Deploy and Deliver

Deployment

DEPLOYING MODELS FOR PRODUCTION USE

Model Deployment Support

Fabric enables seamless deployment of machine learning models into production environments for enterprise use.

Real-Time Inference

Models deployed with Fabric can provide real-time inference for immediate data processing and decision making.

Batch Inference Capability

Batch inference is supported allowing models to process large datasets efficiently at scheduled intervals.

GENERATING BATCH PREDICTIONS INTO DELTA TABLES

Batch Prediction Generation

Batch predictions enable processing large datasets at once for efficient analytics and reporting.

Delta Tables Storage

Storing predictions in Delta tables ensures reliable, scalable data management and easy access.

Analytics Integration

Delta tables facilitate seamless integration with downstream analytics and reporting workflows.

Key Takeaway: DEPLOYMENT

Process:
Integrate the model into production pipelines or apps.

Goal:
Deliver predictions and insights into real-world systems.

Fabric Tools/Features:

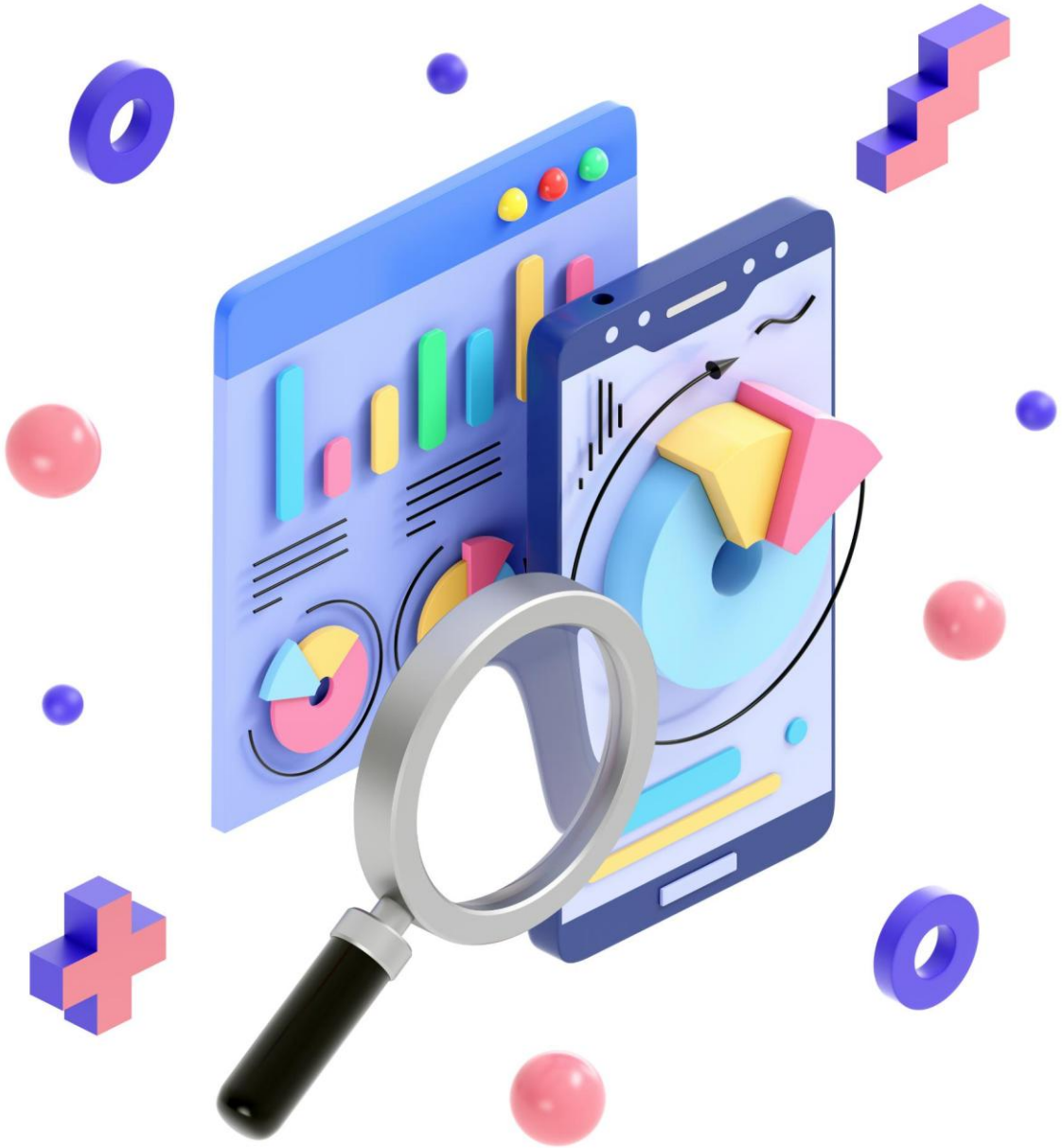
MLflow model registry → save, version, and manage models

Batch prediction pipelines → generate predictions at scale

Power BI dashboards → consume predictions in business reports

Demo

Additional Points



Things to know

How Fabric and AI Foundry Fit Together ?

PLATFORM	PRIMARY PURPOSE	STRENGTHS	WHEN YOU USE IT
Microsoft Fabric	Unified analytics + data engineering + ML	Data pipelines, Lakehouse, notebooks, ML training, governance	When preparing data, training models, and managing analytics
Microsoft AI Foundry	End-to-end AI **application** development	Prompt engineering, model catalog, evaluation, deployment, monitoring	When building, deploying, and operationalizing **AI** apps

Storage Account in Azure Portal ?

- Fabric uses OneLake, not Azure Blob Storage
- Exposes paths that look like ADLS Gen2 URLs, but they are virtual

OneLake & lakehouse

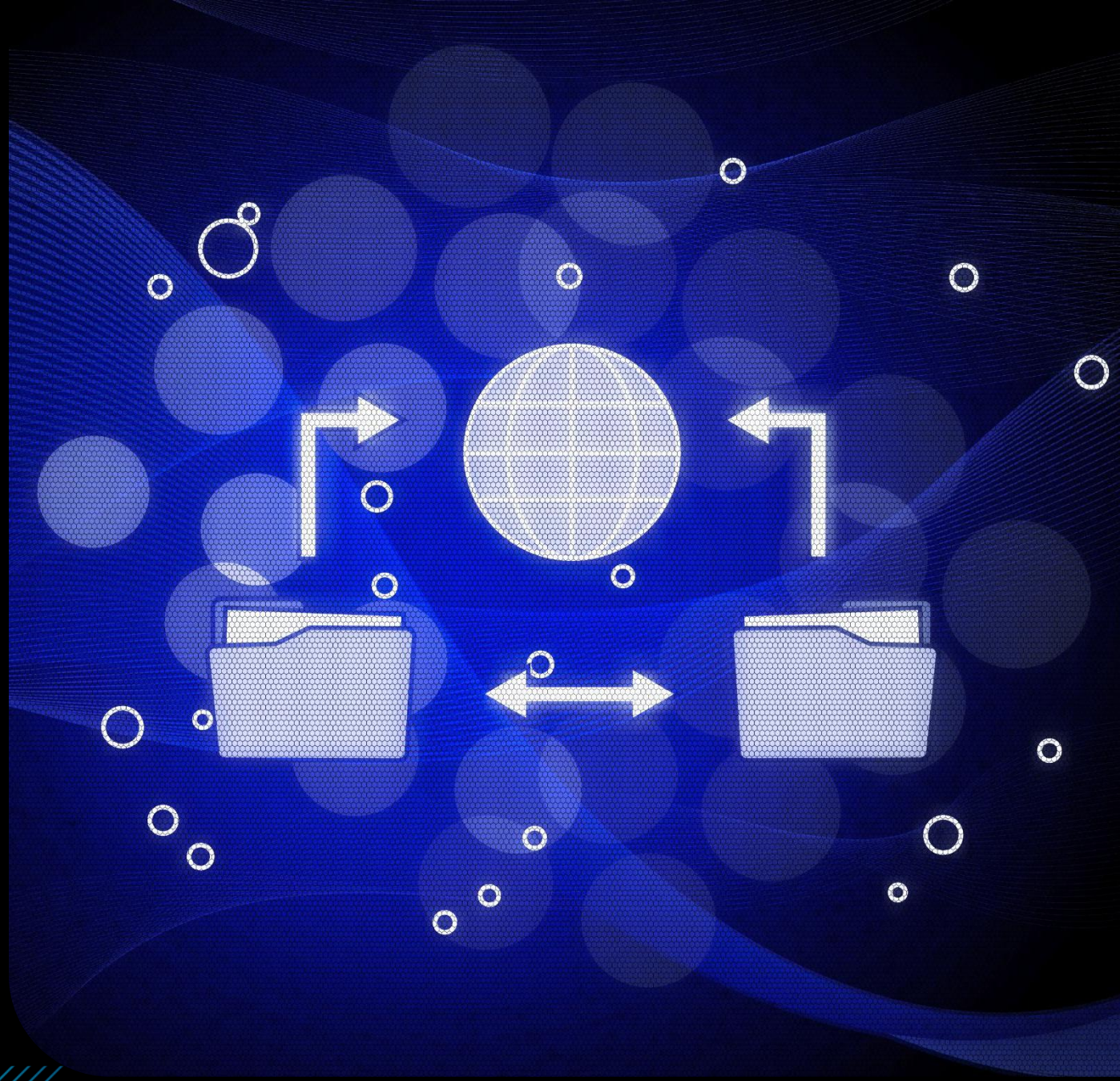
- **OneLake**
  - ✓Fabric's **unified data lake**
  - ✓Built on top of **Azure storage**
  - ✓Single **logic**

- **Lakehouse**
  - ✓A structured data container **organizes** information into tables and files.
  - ✓Inside **OneLake**
  - ✓Optimized for **analytics and machine learning**..

Usually!

- Ingest data → Lakehouse
- Transform data → Lakehouse (Delta tables)
- Expose tables → Semantic
- Model Build reports → Power BI

Medallion Architecture in Fabric



Medallion Architecture

Medallion Architecture

It is a multi-layered **Data Architectural Pattern** that organizes data into **Bronze, Silver, and Gold** layers to progressively improve quality, structure, and business value.

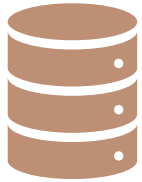
Medallion Architecture

- Medallion Architecture inside **Lakehouses** stored in **OneLake**.

Layer	Fabric Implementation	Typical Storage	Typical Tools/Engines	Purpose
Bronze	Lakehouse (raw zone)	Files, Delta tables	Data Factory Pipelines, Copy Activity, Dataflows Gen2	Ingest raw data
Silver	Lakehouse (clean zone)	Delta tables	Spark Notebooks, Dataflows Gen2, Pipelines	Transform & standardize
Gold	Lakehouse (curated zone)	Delta tables or warehouse or semantic model	Spark, SQL, Power BI modeling	Business-ready analytics

Take Away

Take Away : Stages



Data Ingestion →

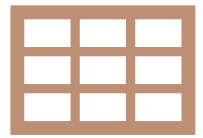
bringing raw data from OneLake, databases, or external sources

Data Wrangling →

cleaning, transforming, and preparing datasets for analysis

Exploration & Visualization →

charts, summaries, and quick insights

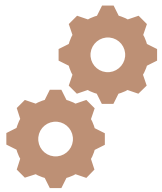


Modeling →

training ML models (classification, regression, clustering, etc.)

Evaluation →

testing model accuracy and performance



Deployment →

integrating models into Fabric pipelines for production use

Take Away : Stages- Tools/Features

Data Ingestion →

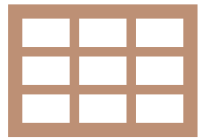
OneLake, Lakehouse, Data Factory (pipelines), Shortcuts, Mirroring

Data Preparation / Wrangling →

Data Wrangler, Spark (PySpark)

Exploration & Visualization →

Notebooks, Visualization libraries (matplotlib, seaborn, Plotly), Power BI

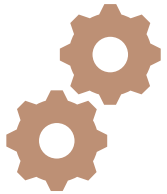


Modeling →

MLflow, Spark MLlib, scikit-learn, TensorFlow, PyTorch

Evaluation →

MLflow metrics tracking, scikit-learn metrics, Spark MLlib evaluators



Deployment →

MLflow model registry, Batch prediction pipelines, Power BI dashboards

Further study

To Learn more, please visit the following

Analytics end-to-end with Microsoft Fabric

https://learn.microsoft.com/en-us/azure/architecture/example-scenario/dataplate2e/data-platform-end-to-end

THANK YOU

Connect and continue the discussion through professional networking.

www.linkedin.com/in/sadiqhahmed/

Sound off.
The mic is all yours.
Influence the product roadmap.

Join the Fabric User Panel

Share your feedback directly with our Fabric product group and researchers.

https://aka.ms/JoinFabricUserPanel

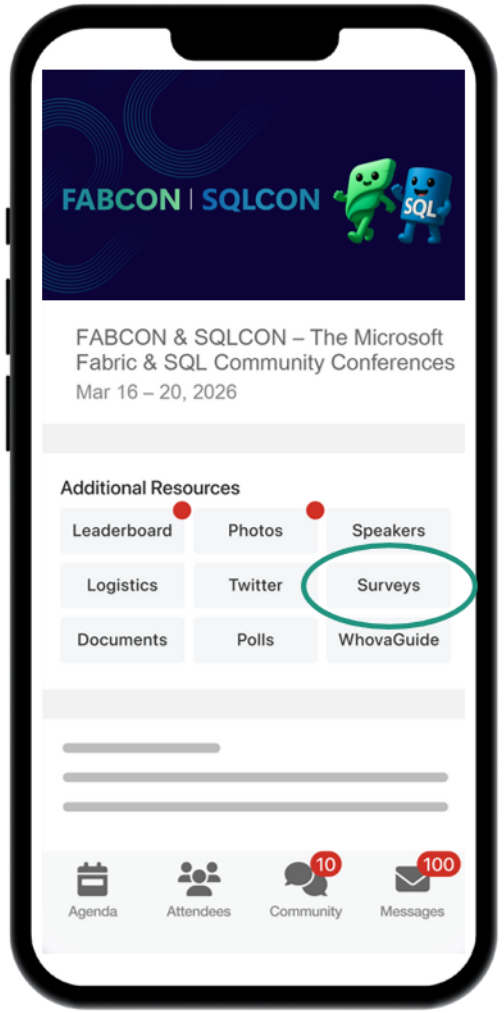
Join the SQL User Panel

Influence our SQL roadmap and ensure it meets your real-life needs

https://aka.ms/JoinSQLUserPanel

How was the session?

Complete Session Surveys in Whova for your chance to WIN PRIZES!



Two Fabric Certifications, One FREE Exam Included

Attendees can take the Fabric Analytics Engineer or Fabric Data Engineer exam for free. Be part of the 2 fastest growing role-based certifications in Microsoft history.

Request your voucher by March 31, 2026.

https://aka.ms/GetDataCertified