Unlocking Data Potential: Microsoft Fabric for Software Developers

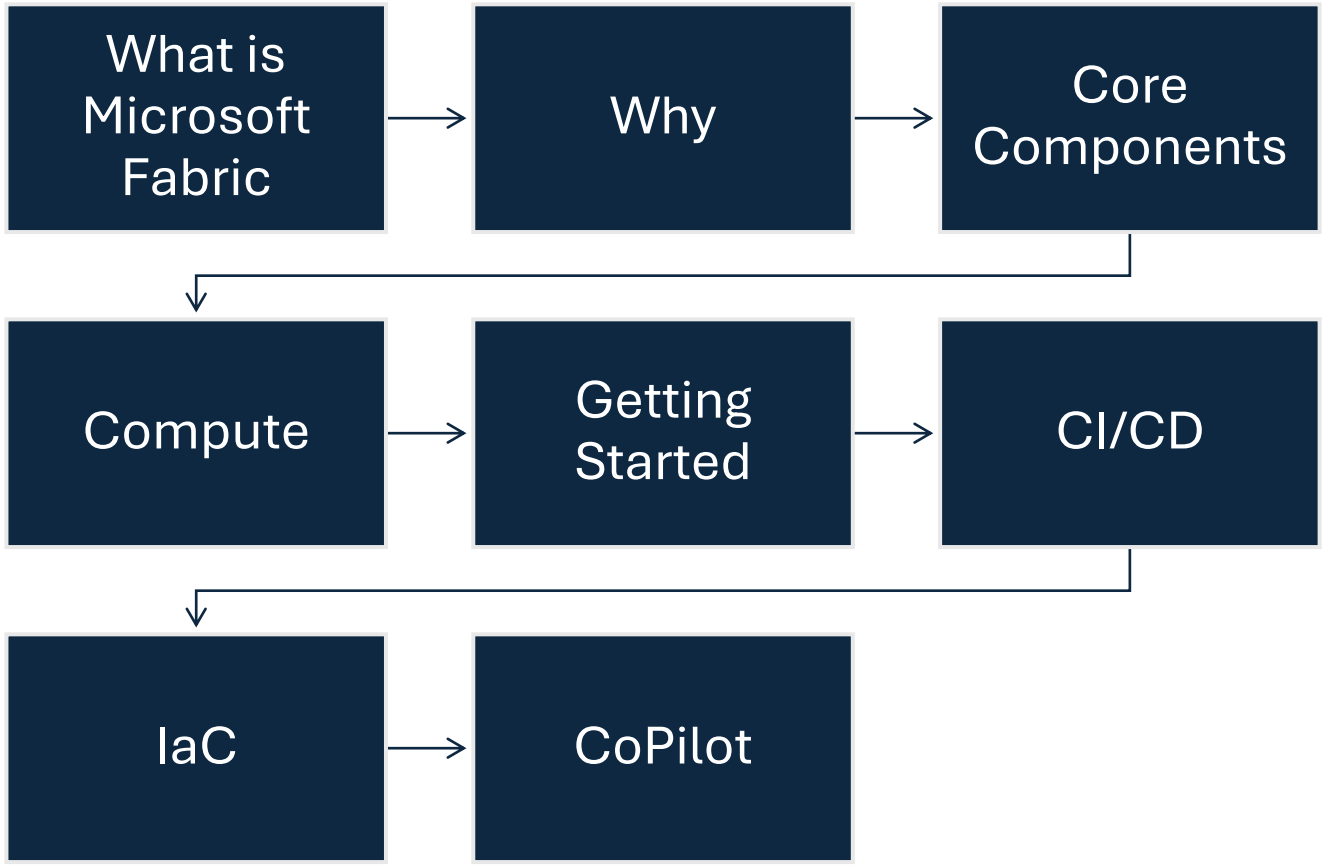
Who am I

- 18+ Years Experience
- Data Engineering & AI Consultant
- Intensive Software & Data Engineering Experience
- Microsoft Data Platform & AI MVP
- Public Speaker
- Community Organizer

@annawykes

Microsoft® Most Valuable Professional

Agenda


What is Microsoft Fabric
Why
Core Components
Compute
Getting Started
CI/CD
IaC
CoPilot

What is Microsoft Fabric

What is Microsoft Fabric

End-to-end, enterprise-ready analytics platform that unifies various data and analytics tools

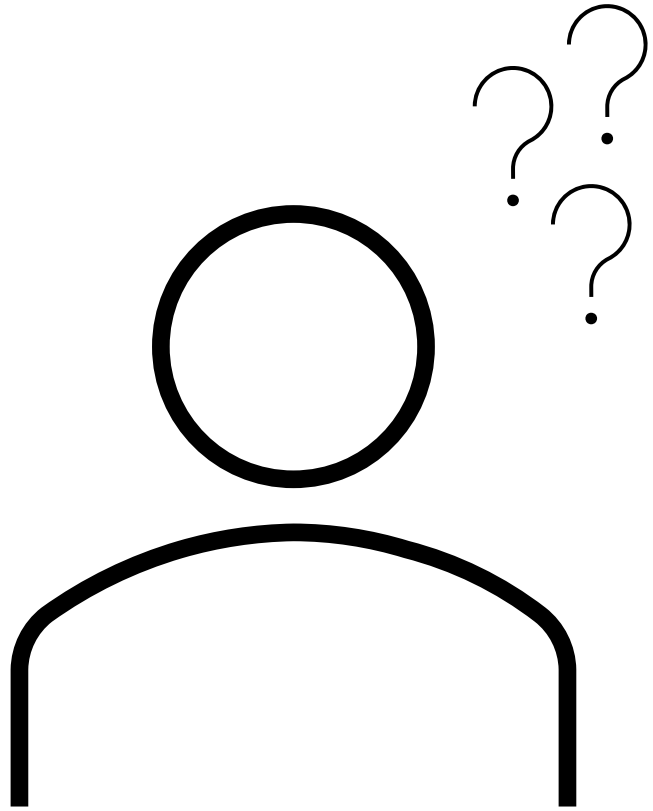
It includes capabilities such as:

- Data movement, ingestion, processing, and transformation
- Real-time event routing
- Report building and visualization

Why?

Online Analytical Processing (OLAP)

How do we best analyse and report on our data?



Initially, we relied on the Data Warehouse – It worked well but had limitations, especially with **loading and storing complex data types**.

As data grew larger and more varied, the **warehouse became too rigid and opinionated**.

We adopted Data Lakes to store data and used tools like **Databricks** for computation.

While Data Lakes offered flexibility, we missed some of the benefits of the **Data Warehouse**, especially **stability**.

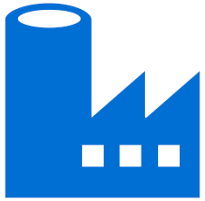
A key challenge: The lake lacked **A.C.I.D (Atomicity, Consistency, Isolation, and Durability) Transactions**

To solve these challenges, new technologies emerged, each claiming to be the ultimate **big data solution** while being **simple and user-friendly**.

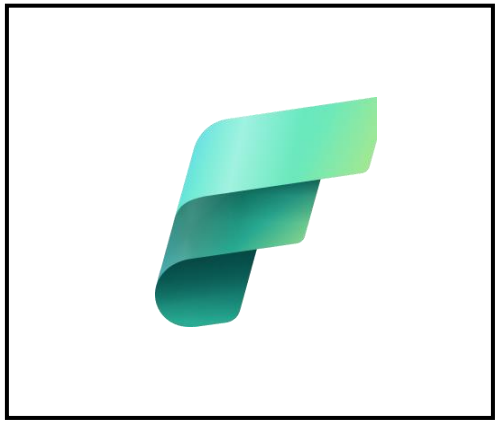
One of the latest solutions in this space is Microsoft Fabric

Combining them all together

Azure Data Lake Gen2



Data Factory



Microsoft Fabric

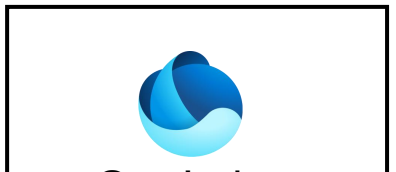
Azure ML

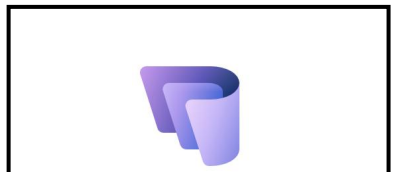
Power BI

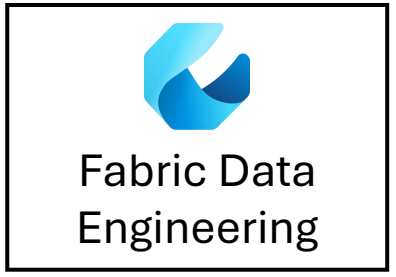
Synapse

Core Components

Core Fabric Components







Power BI

Let's you easily connect to your data sources to visualize and discover what's important and share that with anyone or everyone you want.

For more information, see What is Power BI?

https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview

Data Factory

Provides a modern data integration experience to ingest, prepare, and transform data from a rich set of data sources.

Incorporates the simplicity of Power Query, and you can use more than 200 native connectors to connect to data sources on-premises and in the cloud.

For more information, see What is Data Factory in Microsoft Fabric?

https://learn.microsoft.com/en-us/fabric/data-factory/data-factory-overview

OneLake

Unified, cloud-based data lake that serves as the storage foundation for **Microsoft Fabric**.

It is designed to provide a **centralized, single data store** across an organization, enabling seamless data access and management.

OneLake is like **OneDrive for your organization's data**, making it easier to store, access, and analyze data at scale within Microsoft Fabric

Databases

Developer-friendly transactional database such as Azure SQL Database, which allows you to easily create your operational database in Fabric.

Using the mirroring capability, you can bring data from various systems together into OneLake.

You can continuously replicate your existing data estate directly into Fabric's OneLake, including data from Azure SQL Database, Azure Cosmos DB, Azure Databricks, Snowflake, and Fabric SQL database.

Real-Time Intelligence

End-to-end solution for event-driven scenarios, streaming data, and data log

Enables the extraction of insights, visualization, and action on data in motion by handling data ingestion, transformation, storage, analytics, visualization, tracking, AI, and real-time actions.

The Real-Time hub in Real-Time Intelligence provides a wide variety of no-code connectors, converging into a catalog of organizational data that is protected, governed, and integrated across Fabric.

For more information, see What is Real-Time Intelligence in Fabric?.

https://learn.microsoft.com/en-us/fabric/fundamentals/microsoft-fabric-overview#real-time-hub-the-unification-of-data-streams

https://learn.microsoft.com/en-us/fabric/real-time-intelligence/overview

Fabric Data Engineering

Provides a Spark platform with great authoring experiences.

It enables you to create, manage, and optimize infrastructures for collecting, storing, processing, and analyzing vast data volumes.

Fabric Spark's integration with Data Factory allows you to schedule and orchestrate notebooks and Spark jobs. For more information, see What is Data engineering in Microsoft Fabric?

https://learn.microsoft.com/en-us/fabric/data-engineering/data-engineering-overview

Fabric Data Science

Enables you to build, deploy, and operationalize machine learning models from Fabric.

It integrates with Azure Machine Learning to provide built-in experiment tracking and model registry.

Data scientists can enrich organizational data with predictions and business analysts can integrate those predictions into their BI reports, allowing a shift from descriptive to predictive insights.

For more information, see What is Data science in Microsoft Fabric?

https://learn.microsoft.com/en-us/fabric/data-science/data-science-overview

Fabric Data Warehouse

Provides industry leading SQL performance and scale.

It separates compute from storage, enabling independent scaling of both components.

Additionally, it natively stores data in the open Delta Lake format.

For more information, see What is data warehousing in Microsoft Fabric?

https://learn.microsoft.com/en-us/fabric/data-warehouse/data-warehousing

Microsoft Fabric Compute

Compute Options in Microsoft Fabric

Microsoft Fabric offers multiple **compute options** to handle different data processing workloads efficiently. These compute engines are integrated within Fabric, allowing users to choose the best fit for their needs.

Spark Compute (Apache Spark)

Ideal for **big data processing, machine learning, and data engineering.**

Fully managed **Spark-as-a-Service** within Fabric.

Supports **notebooks, Python, Scala, SQL, and R**.

Enables distributed processing of large datasets.

SQL Compute (Data Warehouse & Lakehouse SQL)

Data Warehouse (DWH) SQL: Optimized for **structured, relational data processing**

Lakehouse SQL: Enables **SQL-based querying on Delta Lake storage**.

Supports **T-SQL** for analytics and transformations.

Best for **traditional SQL workloads and BI reporting**.

KQL Compute (Real-Time Analytics Engine)

Uses **Kusto Query Language (KQL)** for **real-time event processing and log analytics**.

Best for **streaming data analysis, IoT, and monitoring scenarios**.

High-performance, columnar storage for rapid querying.

Power BI Compute (Direct Query & Import Mode)

Embedded compute for **data visualization and analytics**.

Supports **Direct Query (live connection)** and **Import Mode (cached data)**.

Best for **interactive dashboards and reporting**.

Data Factory Compute (ETL & Data Pipelines)

Used for **data movement, transformation, and integration**.

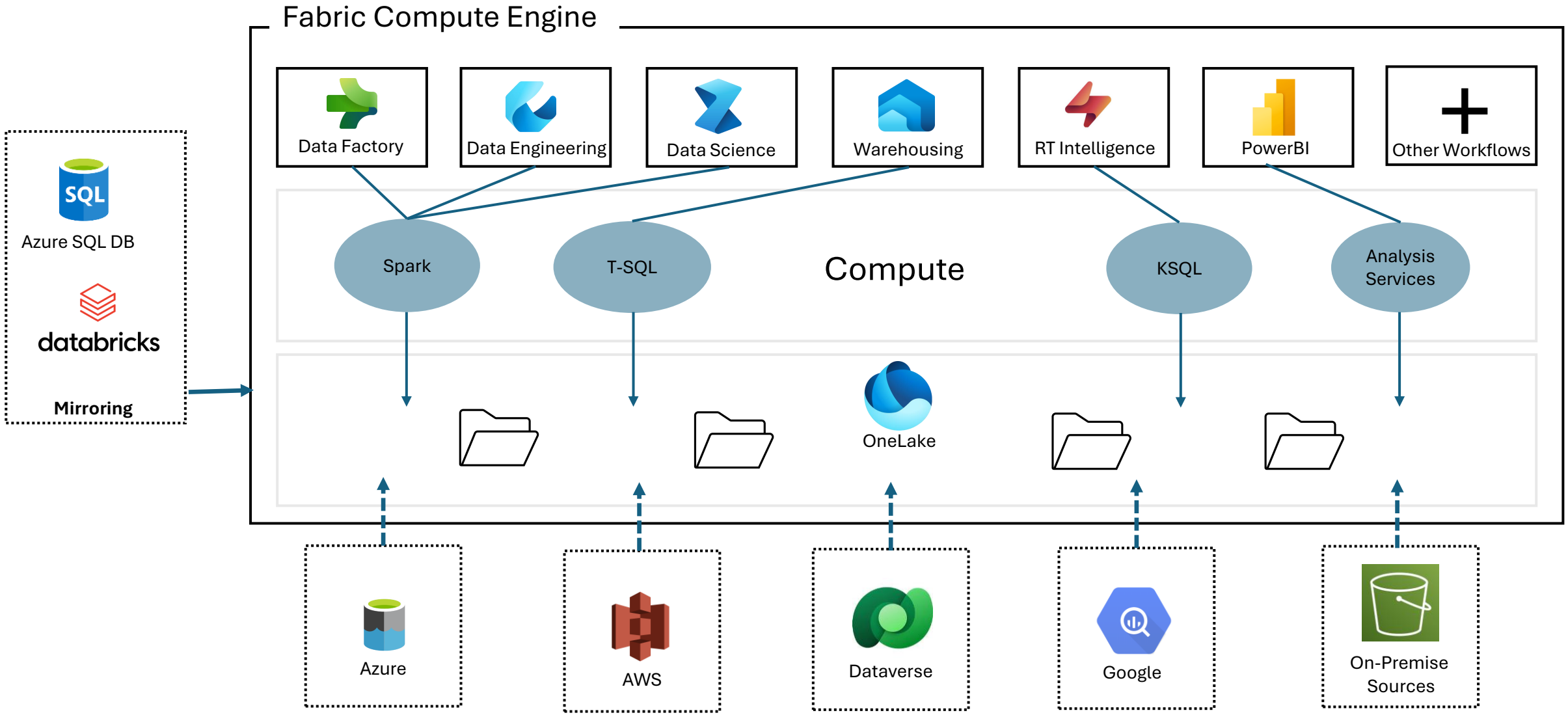
Supports **low-code/no-code pipelines** for ELT (Extract, Load, Transform).

Connects to **multiple data sources** within Fabric and external environments.

Choosing the Right Compute Engine

Each of these compute engines is **fully managed and scalable**, making Microsoft Fabric a **unified** analytics solution for different data needs.

Compute Option	Best For
Spark Compute	Big data processing, AI/ML workloads
SQL Compute	Structured data analysis, reporting, BI
KQL Compute	Real-time analytics, streaming data
Power BI Compute	Interactive dashboards, business intelligence
Data Factory Compute	ETL, data movement, transformation


Fabric Compute Engine
Data Factory
Data Engineering
Data Science
Warehousing
RT Intelligence
PowerBI
Other Workflows
Azure SQL DB
databricks
Mirroring
Spark
T-SQL
Compute
KSQL
Analysis Services
OneLake
Azure
AWS
Dataverse
Google
On-Premise Sources

Getting Started

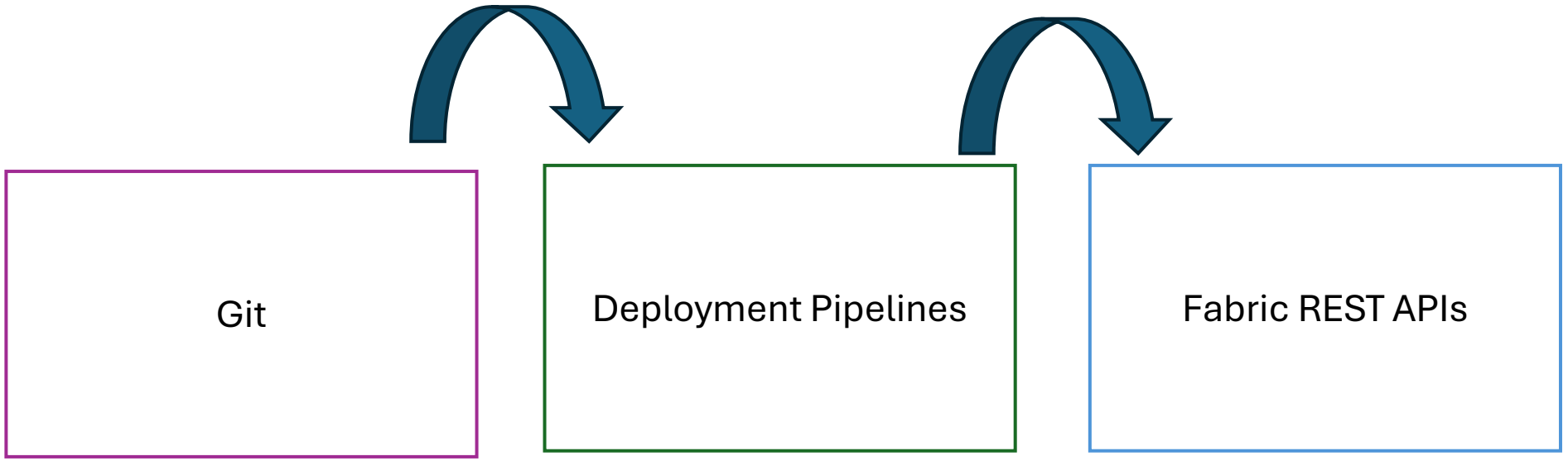
Two Options

Azure

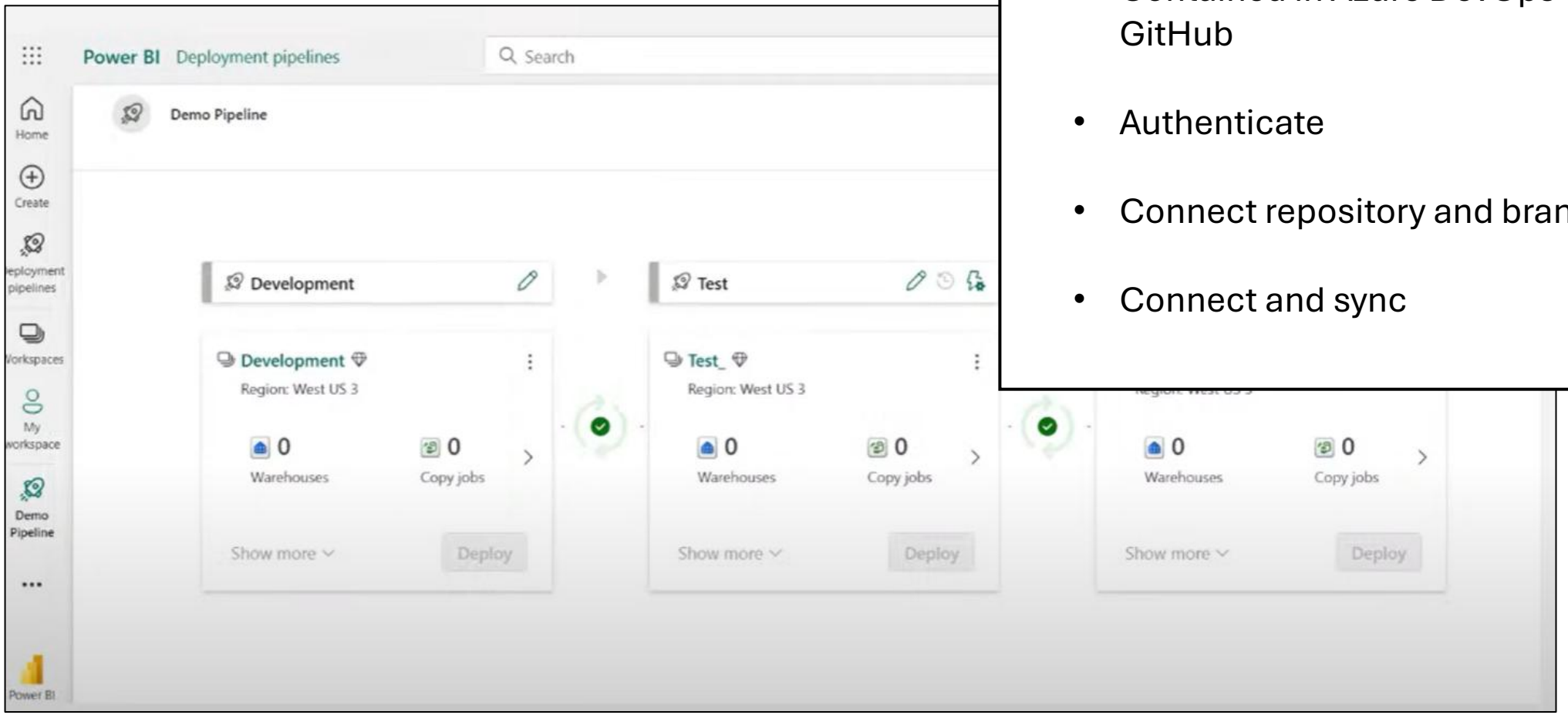
Power BI

CI/CD in Fabric

Rest API



Deployment Pipeline



Contained in Workspaces

- Contained in Azure DevOps or GitHub
- Authenticate
- Connect repository and branch
- Connect and sync

Core

Object Type	Git	API (CRUD)
Connection	n/a	Supported
Domain	n/a	Supported
Workspace	n/a	Supported
Capacity	n/a	List only
Deployment Pipeline	n/a	Get/List only
Workspace Spark Settings	?	Get/Update only

Data Factory and Engineering

Object Type	Git	API (CRUD)
Lakehouse	Partially	Covered
Notebook	Supported	Covered
Data pipeline	Supported	Limited
Dataflow Gen2	Preview	Not covered
(Spark) Environment	Supported	Covered
Spark Job Definition	Supported	Covered
API for GraphQL (preview)	Unsupported	Preview and limited
Azure Data Factory (preview)	Unsupported	Not covered
Apache Airflow job (preview)	Unsupported	Not covered
SQL Database	Supported	Covered

Database and Data Warehousing

Object Type	Git	API (CRUD)
SQL Database	Supported	Covered
Warehouse	Supported	Create/Update name only
Mirrored Azure SQL Database	Supported	Covered
Mirrored Snowflake	Supported	?
Mirrored Azure Comsos DB (preview)	??	?
Mirrored Azure SQL Managed Instance (preview)	??	?

Data Science

Object Type	Git	API (CRUD)
ML Model	Unsupported	Limited
ML Experiment	Unsupported	Limited

Real-Time Intelligence

Object Type	Git	API (CRUD)
Eventhouse	Supported	Covered
KQL Database	Supported	Covered
KQL Queryset	Supported	Covered
Eventstream	Supported	Covered
Activator (aka Reflex)	Supported	Covered
Real-Time (KQL) Dashboard	Supported	Covered

Power BI

Object Type	Git	API (CRUD)
Dataflow Gen 1	Unsupported	Unsupported
Paginated Reports (preview)	Supported	Covered
Scorecard	Unsupported	Unsupported
Dashboard	Unsupported	Limited
Sematic Model (default)	n/a	n/a
Sematic Model	Supported	Covered
Reports	Supported	Covered
Exploration (preview)	Unsupported	Not Covered
Datamart (preview)	Unsupported	Limited

Infrastructure as Code

Terraform Provider

Only released last year in preview

Supports the creation and management of the following resources:

Environment	Data Pipeline	Eventhouse	Eventstream	KQL Database
ML Experiment	ML Model	Notebook	Report	Semantic Model
Warehouse	Domain	Domain Role Assignment	Domain Workspace Assignment	Workspace
Workspace Role Assignment	Workspace Git	Spark Settings	Spark Pool	Environment Spark Settings

Pre-Requisites

- A Fabric capacity
- A service principal to be added to the fabric capacity as a capacity administrator
- Terraform >= 1.8x
- VS Code with Terraform extension
- Azure CLI for authentication with User context when the service principal is not supported

Considerations

The provider is quite basic in certain areas, such as the Warehouse resource, where only a display name can be passed.

Some resources still need improvement, but it's a strong first step toward a more robust approach to managing and deploying a Fabric-based data platform.

Getting Started

https://github.com/microsoft/fabric-terraform-quickstart

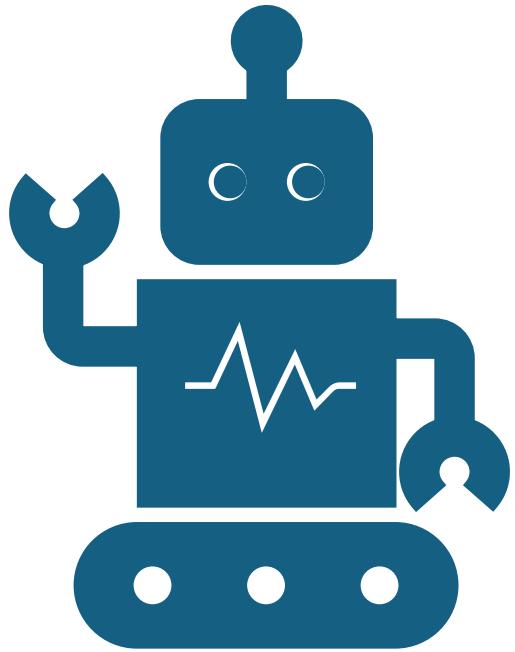
CoPilot, CoPilot, CoPilot

Before you can start using Copilot capabilities in Microsoft Fabric, you need to enable Copilot:

https://learn.microsoft.com/en-us/fabric/fundamentals/copilot-enable-fabric

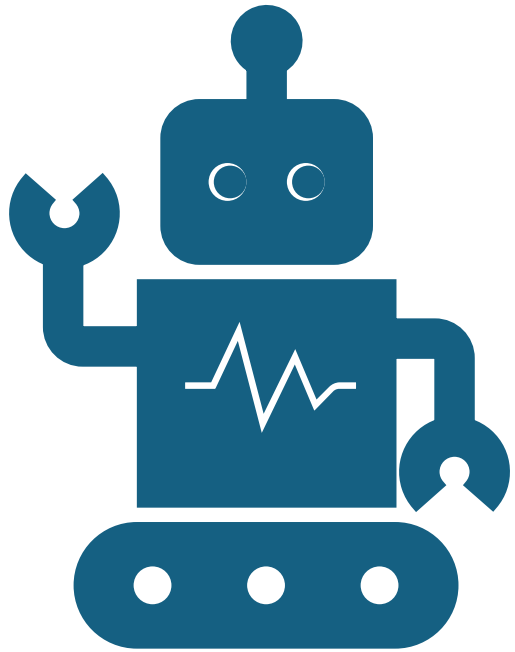
Copilot for Data Science and Data Engineering

- Copilot for Data Engineering and Data Science is an AI-powered toolset designed to enhance the workflow of data professionals.
- Intelligent code completion, automates repetitive tasks, and provides industry-standard code templates to streamline the development of data pipelines and analytical models.
- Delivers contextual code suggestions tailored to specific tasks, enabling more efficient and effortless coding.
- Serves as an interactive assistant, reducing workload and accelerating the transformation of raw data into valuable insights.



Copilot for Data Factory

- Copilot for Data Factory is an AI-powered toolset designed to assist data wranglers in optimizing their workflow.
- Enables seamless data transformation through code generation and offers code explanations to enhance understanding of complex tasks

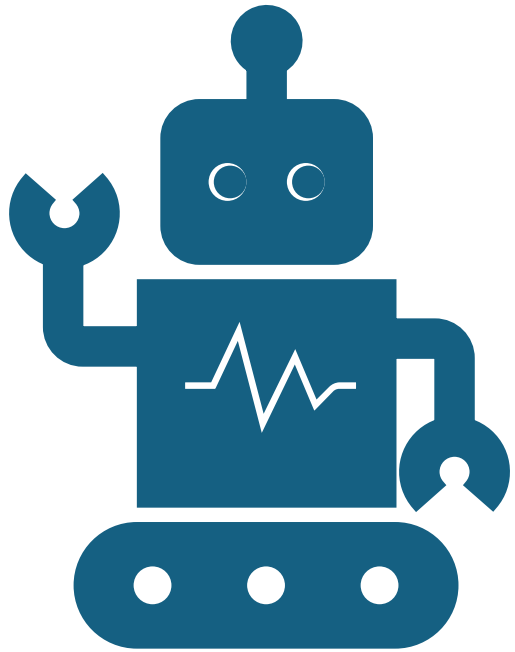


Copilot for Data Warehouse

Microsoft Copilot for Fabric Data Warehouse is an AI-powered assistant designed to simplify data warehousing tasks.

Its key features include

- Natural Language to SQL
- Code completion
- Quick actions
- Intelligent insights

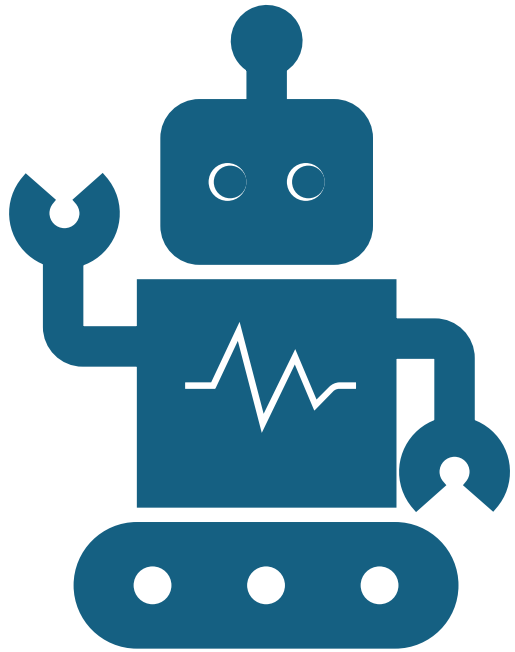


Copilot for Power BI

Power BI has introduced generative AI that allows you to create reports automatically by selecting the topic for a report or by prompting Copilot for Power BI on a particular topic.

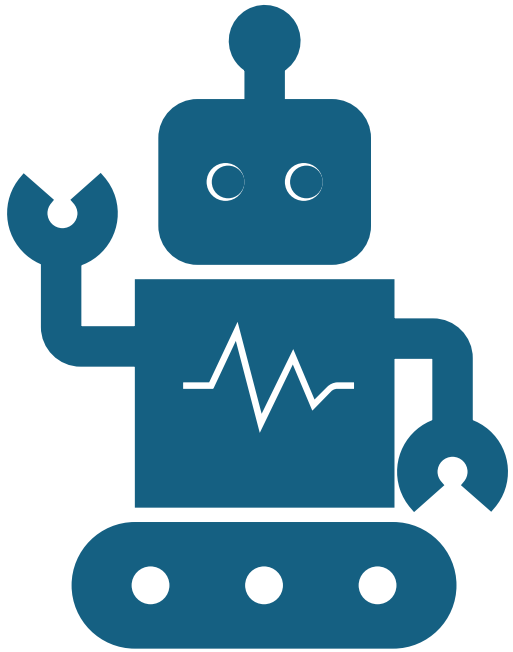
Key features include

- Generative AI in Power BI allows for automatic report creation.
- Report Generation – Create reports by selecting a topic or prompting Copilot for Power BI.
- Summarization – Generate a summary for the report page you just created.
- Enhanced Q&A – Copilot can generate synonyms to improve natural language interactions.



Copilot for Real-Time Intelligence

- Copilot for Real-Time Intelligence is an AI-powered tool for data exploration and insight extraction.
- Natural Language to KQL – Input questions about your data, and Copilot automatically converts them into Kusto Query Language (KQL) queries.
- Streamlined Analysis – Simplifies data analysis for both experienced KQL users and citizen data scientists.

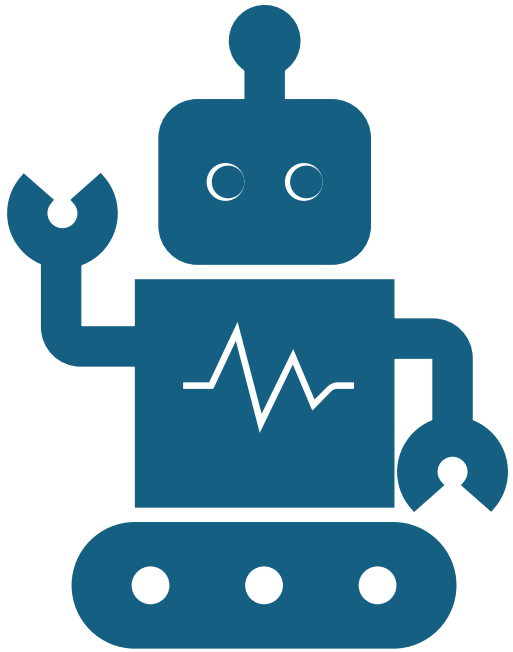


Copilot for SQL database

Copilot for SQL Database in Microsoft Fabric is an AI assistant for streamlining OLTP database tasks.

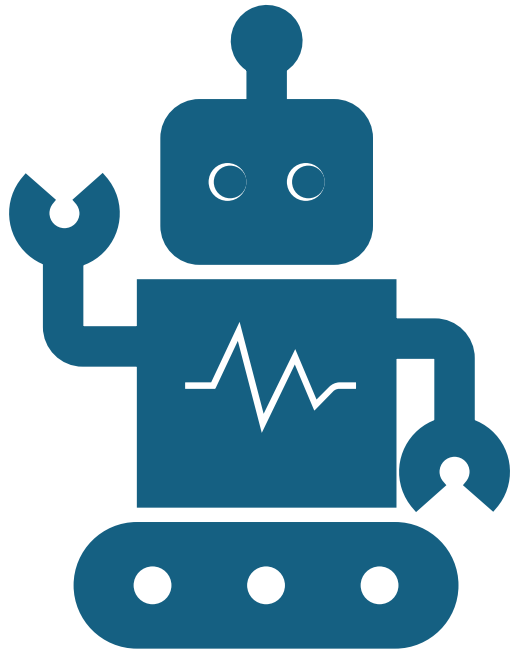
Key features

- Natural Language to SQL: Convert natural language queries into
- SQL.Code Completion: Assist with writing SQL code efficiently.
- Quick Actions: Enable fast actions for database management.
- Document-based Q&A: Answer questions based on database documentation.



Create Your Own CoPilot!!

- The Conversational Knowledge Mining Solution Accelerator is built on Microsoft Fabric, Azure OpenAI Service, and Azure AI Speech.
- It helps customers with large volumes of conversational data to leverage generative AI for identifying key phrases and operational metrics.
- This enables the discovery of valuable insights that can drive business impact.
- Learn more here: https://github.com/microsoft/Conversation-Knowledge-Mining-Solution-Accelerator



Summary

Microsoft Fabric is a **unified, end-to-end data analytics platform** that integrates multiple Microsoft data services into a single solution. It simplifies data management, processing, and analytics for organizations.

Key Features:

- **OneLake** – A centralized, organization-wide data lake for seamless storage and access.
- **Multiple Compute Engines** – Includes **SQL, Spark, and KQL** for diverse workloads.
- **Integrated Services** – Combines **Data Engineering, Data Factory, Data Science, Real-Time Analytics, and Business Intelligence**.
- **AI-Powered Analytics** – Enhances data insights with built-in AI capabilities.

Microsoft Fabric is essentially a cloud-based, all-in-one data platform designed to streamline data management, analytics, and AI-driven insights