



Microsoft

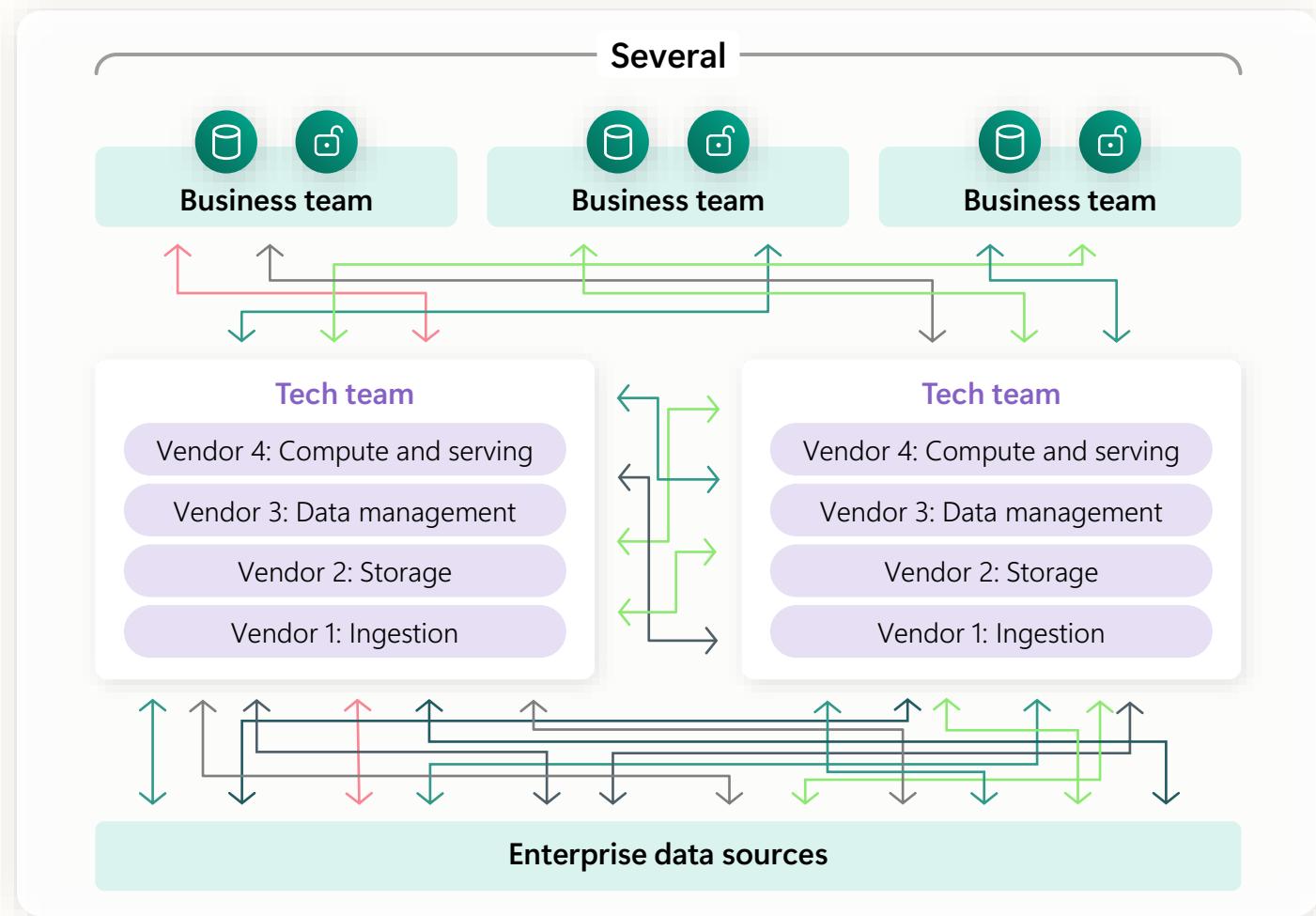
Microsoft Fabric

Microsoft Fabric

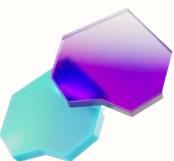
Manjunath Suryanarayana

The starting line

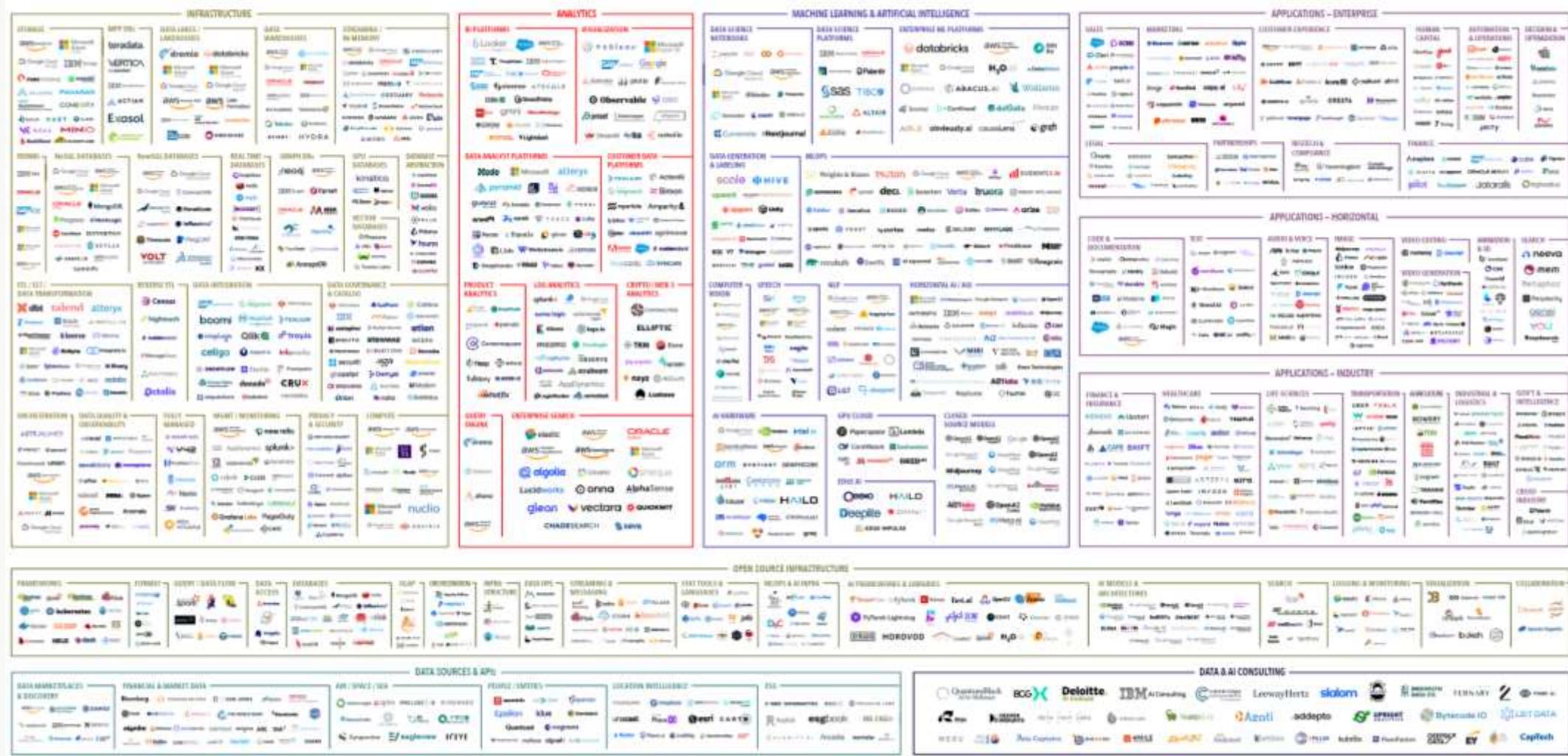
A complex, organically evolved data estate



- 1 Data copies and infrastructure inefficiencies
- 2 Limited interoperability between vendor services
- 3 Data exposure risks



Customers enhancing their data estate face immense complexity





Microsoft Fabric

The data platform for the era of AI

From

To

Multiple analytics services

» Unified stack

Disconnected data sources

» All the data in one place

Isolated application

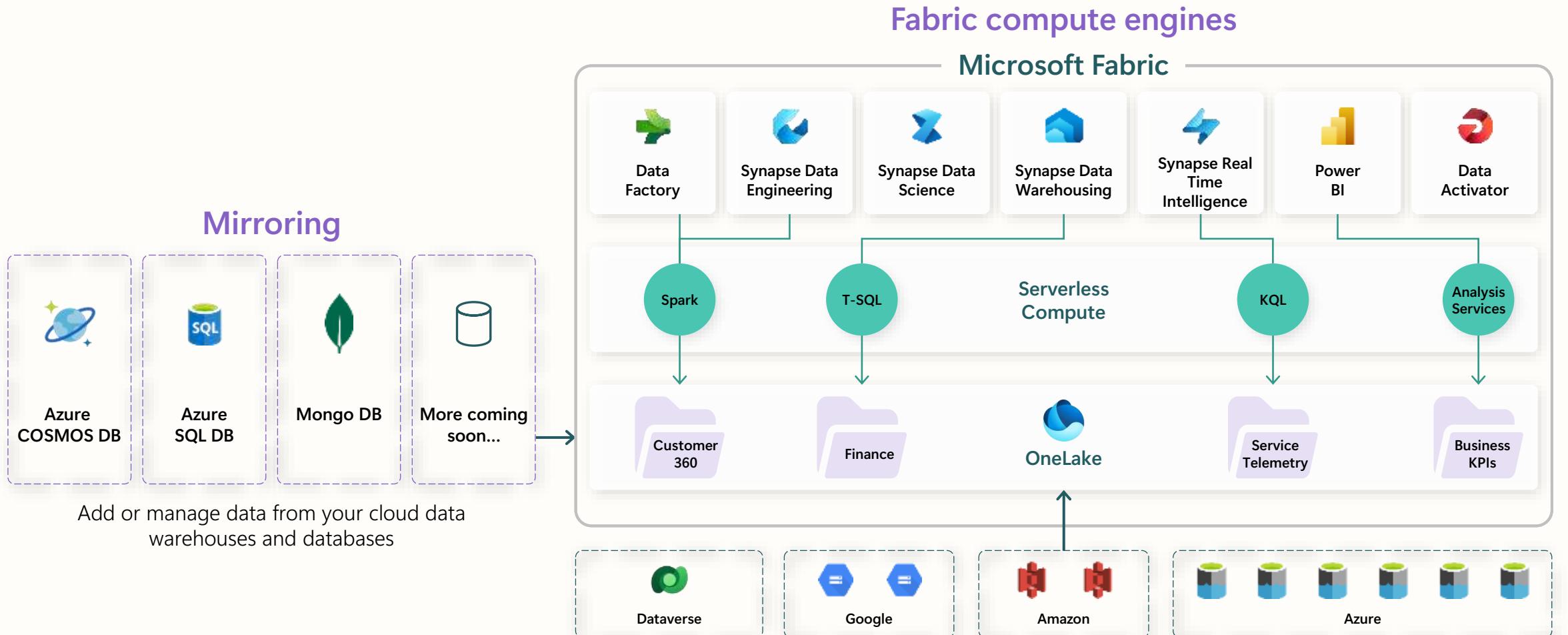
» Entire estate

Gen AI bolt on

» Gen AI built in



Bring in data from anywhere



Multi-cloud shortcuts

Virtualize data in OneLake from Azure, Dataverse, Google, and Amazon S3 without moving or duplicating it



Microsoft Fabric



Data
Factory



Data
Engineering



Data
Warehouse



Data
Science



Real-Time
Intelligence



Power BI



Partner &
Industry
workloads



Copilot in Fabric



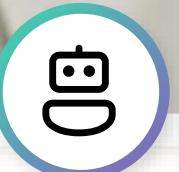
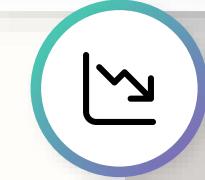
OneLake



Microsoft Purview



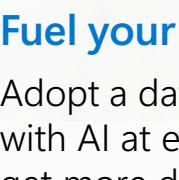
Microsoft Fabric Capabilities



Unify your analytics on a complete, governed platform

Reduce the cost and effort of integration with a unified, secure, and governed platform

Empower every business user
Empower everyone to uncover insights accessible data, easy-to-use tools, and visuals embedded apps they use everyday



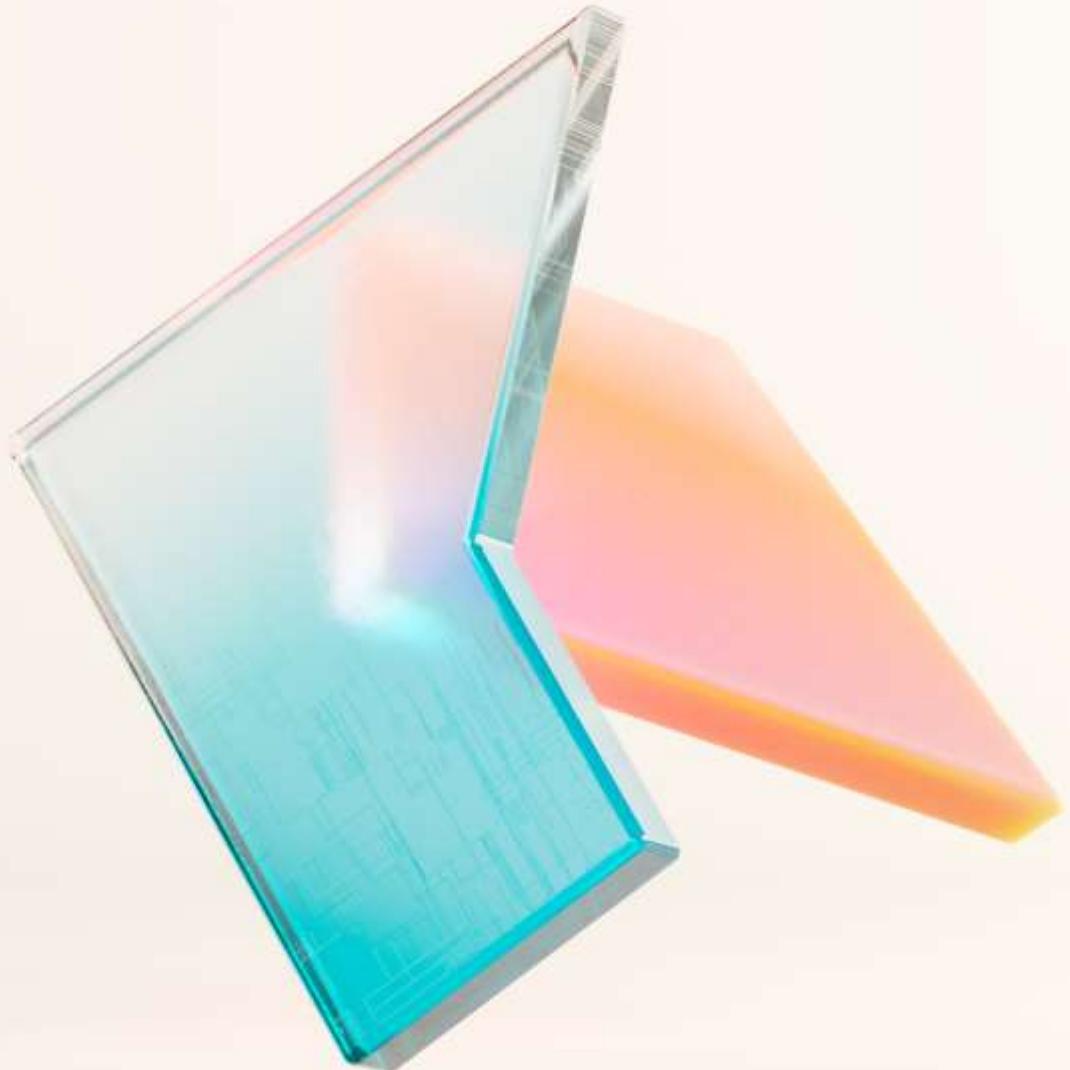
Establish a trusted data foundation

Connect to any data from a single, multi-cloud data lake and use a single copy of data across engines



Fuel your AI innovation

Adopt a data platform infused with AI at every layer to help you get more done, faster



Workload overview

Seven key workloads for end-to-end analytics

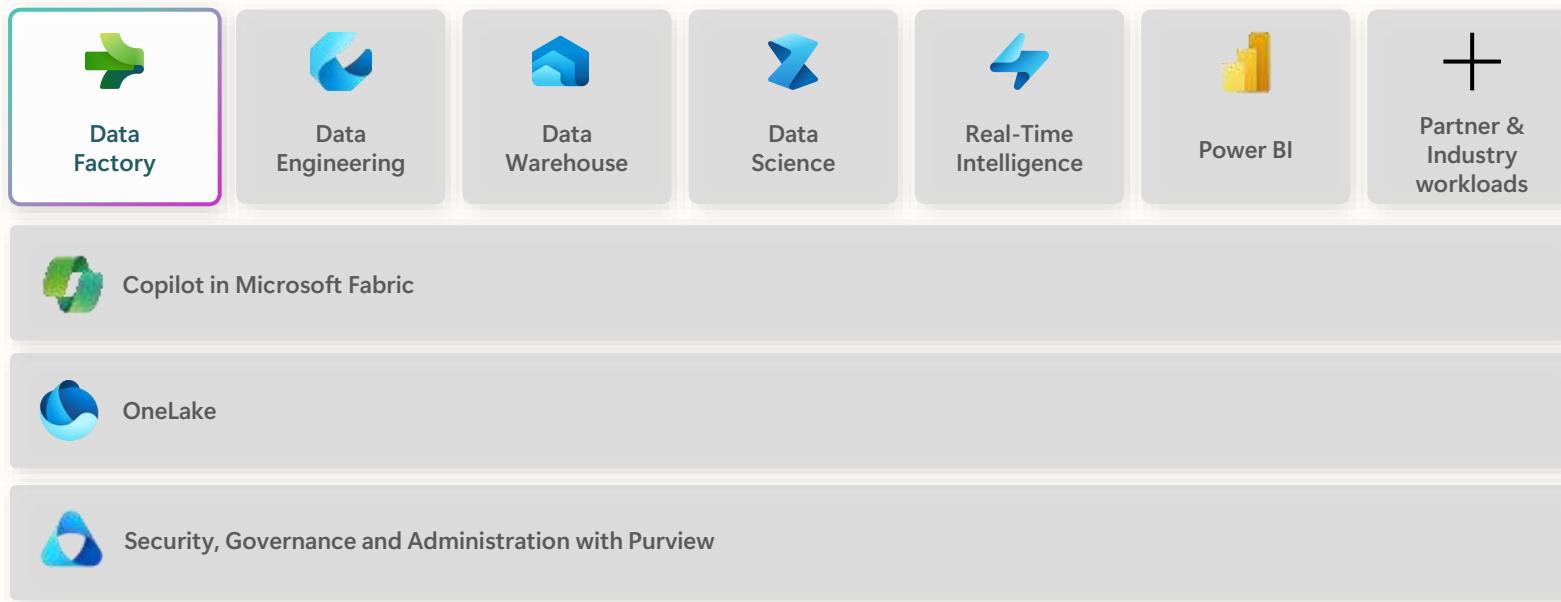
Workloads are designed to target specific personas and tasks, yet work together seamlessly in a unified platform via OneLake to enable creators to collaboratively do their best work

 Data Factory	Unify your data estate with a data integration experience and 300+ data transformations to easily solve the most complex ETL scenarios
 Data Engineering	Enable data engineers to design, build, and maintain infrastructures at scale using World-class Spark platform with great authoring experiences to
 Data Warehouse	Provide industry-leading SQL performance and scale, fully separating compute from storage for independently scaling and natively storing data in open Parquet/Delta Lake
 Data Science	Empower data scientists and analysts to quickly build, deploy, and operationalize sophisticated AI directly within Fabric
 Real Time Intelligence	Ingest streaming data with high granularity, dynamically transform streaming data, query data in real-time for instant insights, and trigger actions
 Power BI	Make better, data-driven decisions with the world's leading business intelligence platform that turns unrelated sources of data into coherent, interactive insights



Data Factory workload

Dataflows and data pipelines bring together low-code, AI-based experiences, multi-cloud connectivity, and persistent data security and governance to help solve complex ETL scenarios for all developers

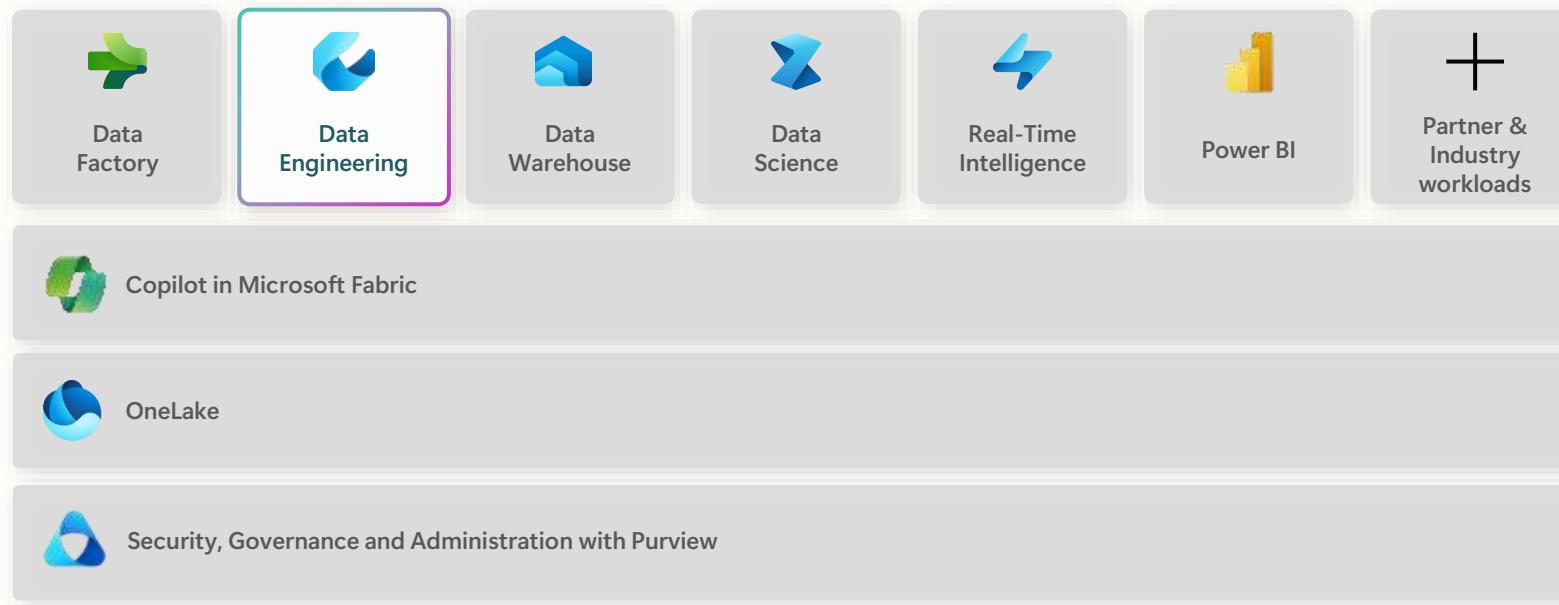


- 200+ native data source connectors
- 300+ data transformations in dataflows designer to transform data more easily
- Cloud-scale data movement with Data Factory
- Low-code interface for ingesting data from hundreds of data sources using Dataflows Gen2
- Out-of-the-box rich data orchestration capabilities to compose flexible workflows
- Powerful, enterprise-grade Data Factory workload with the best of ADF and Power Query together



Data Engineering workload

Build your data estate and empower data engineers with a world-class Spark platform, fully integrated with Data Factory, to transform and maintain infrastructures at scale

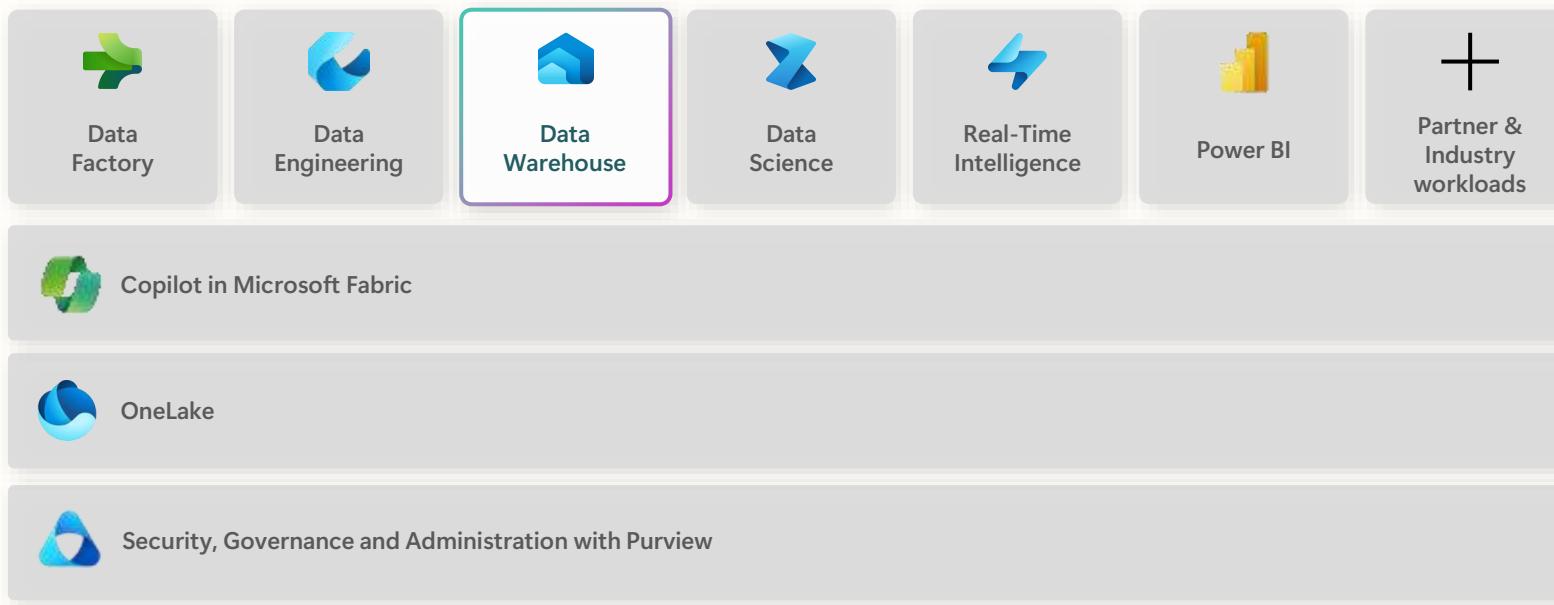


- ❯ Pro and low-code authoring experience
- ❯ Schedule and orchestrate data transformations with notebooks and Spark jobs
- ❯ Use notebooks to write code for data ingestion, preparation, and transformation
- ❯ Launch clusters on demand and dynamically scale in, scale out, pause, and resume
- ❯ Perform code-free interactive data exploration and add to your data pipeline



Data Warehouse workload

Achieve data platform goals with ease and cost efficiency, while empowering your developers and engineers of any skill level with accelerated reporting and insights

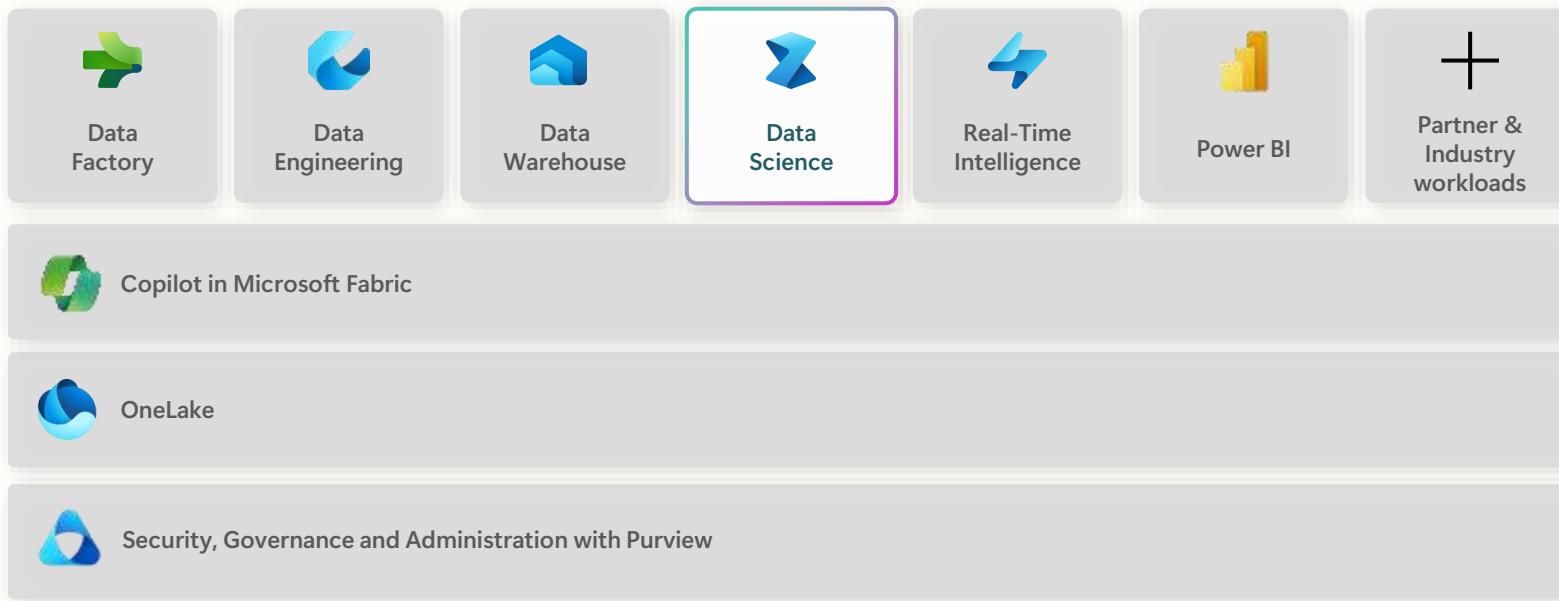


- Industry-leading SQL performance and scale
- Fully separate compute storage that can independently scale
- Natively store data in open Parquet/Delta Lake
- TLS 1.2 encryption protects connections for granular security across data platform
- Cross-database querying for fast insights and zero data duplication



Data Science workload

Build, deploy, and operationalize sophisticated AI and ML models with speed and at scale from your Lakehouse

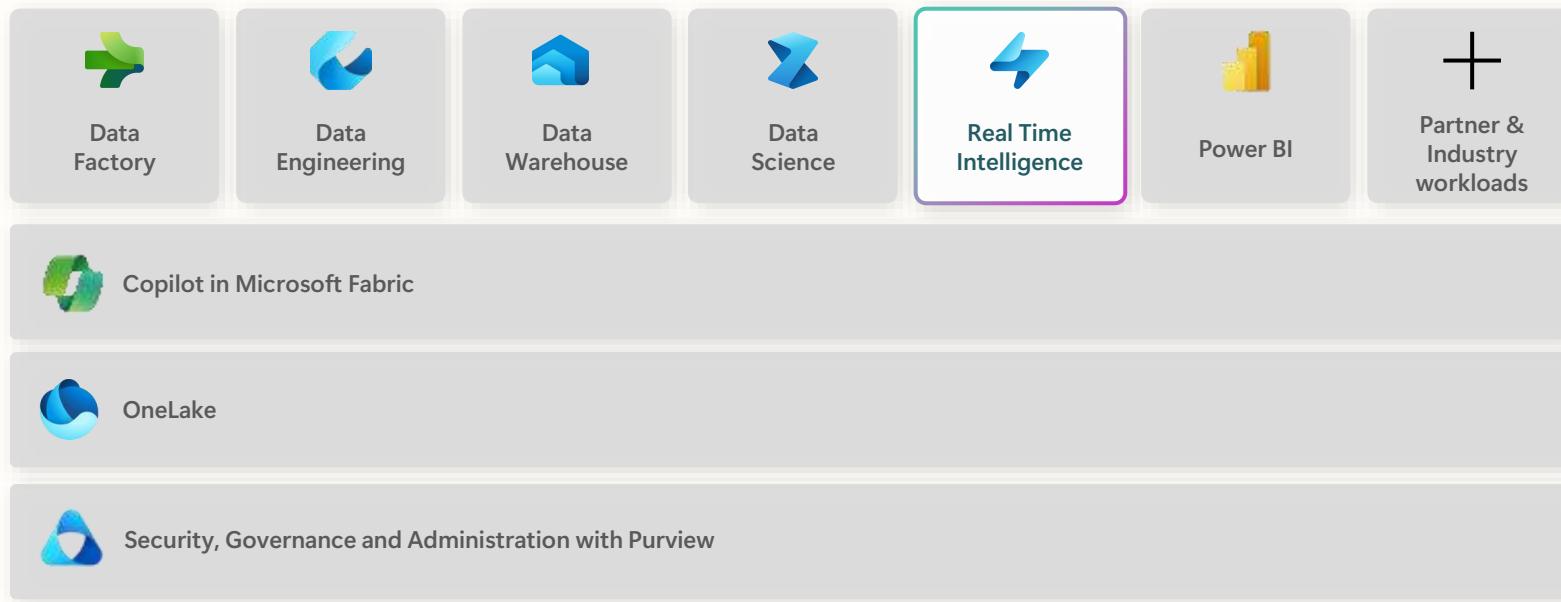


- Access data from multiple sources and store data and insights in Lakehouse(s)
- Leverage data science capabilities for model prediction at scale. Iterate, build, and track machine learning experiments using ML flow
- Perform exploration, experimentation, modeling, featurization and serving of predictive insights by leveraging built-in experiences
- Collaborate with others via Notebook, Power BI, and Lakehouses in real-time



Real-Time Intelligence workload

Explore data and turn insights into actions by performing real-time analysis across telemetry data to better predict, optimize, and improve data applications

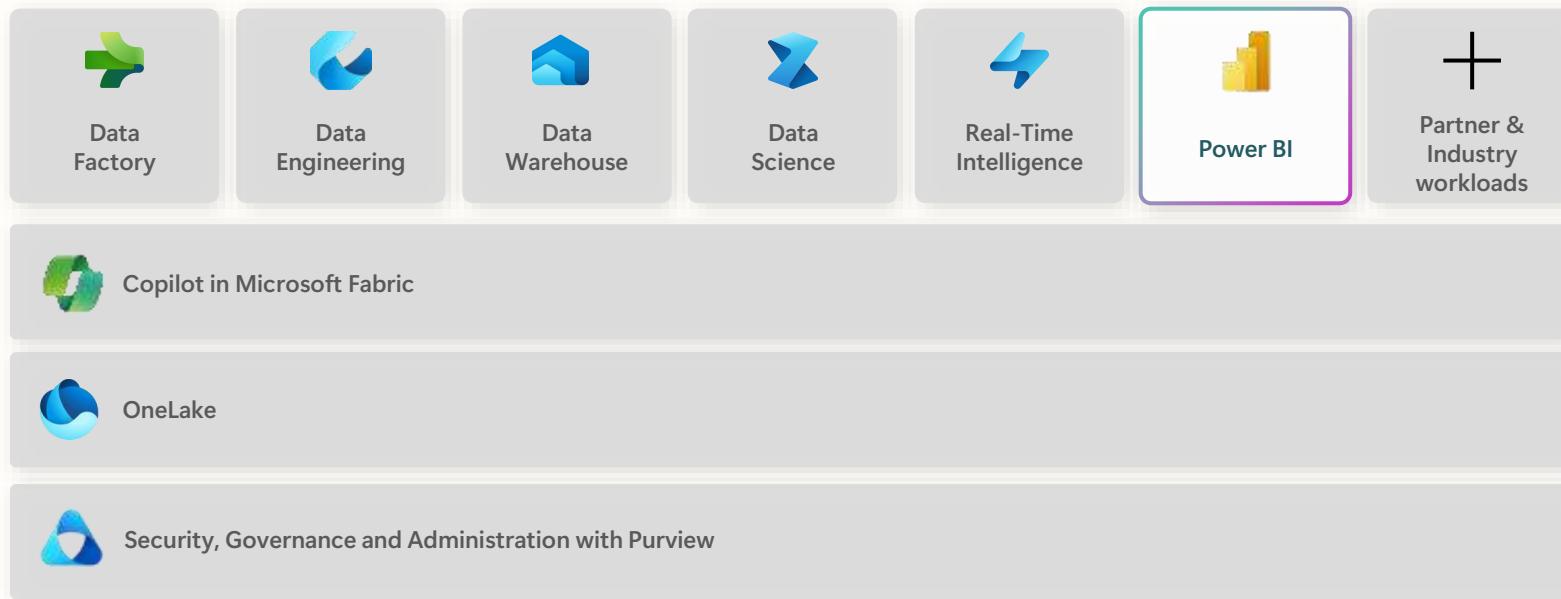


- Ingest, transform, query, visualize, and act on data in real time.
- Simple ingestion, curation and processing of streaming data in the Real-Time Hub, a single data estate for data in motion
- No-, low-, and pro-code experiences for everything from business insight discovery to complex stream processing
- Create triggers on changing data to act automatically when conditions are met
- Streamline analysis of event streaming data with Copilot in Fabric



Power BI workload

Uncover powerful insights with intelligence visuals, leverage data quickly and intuitively, and help achieve faster and better, data-based decisions with the industry-leading Power BI platform

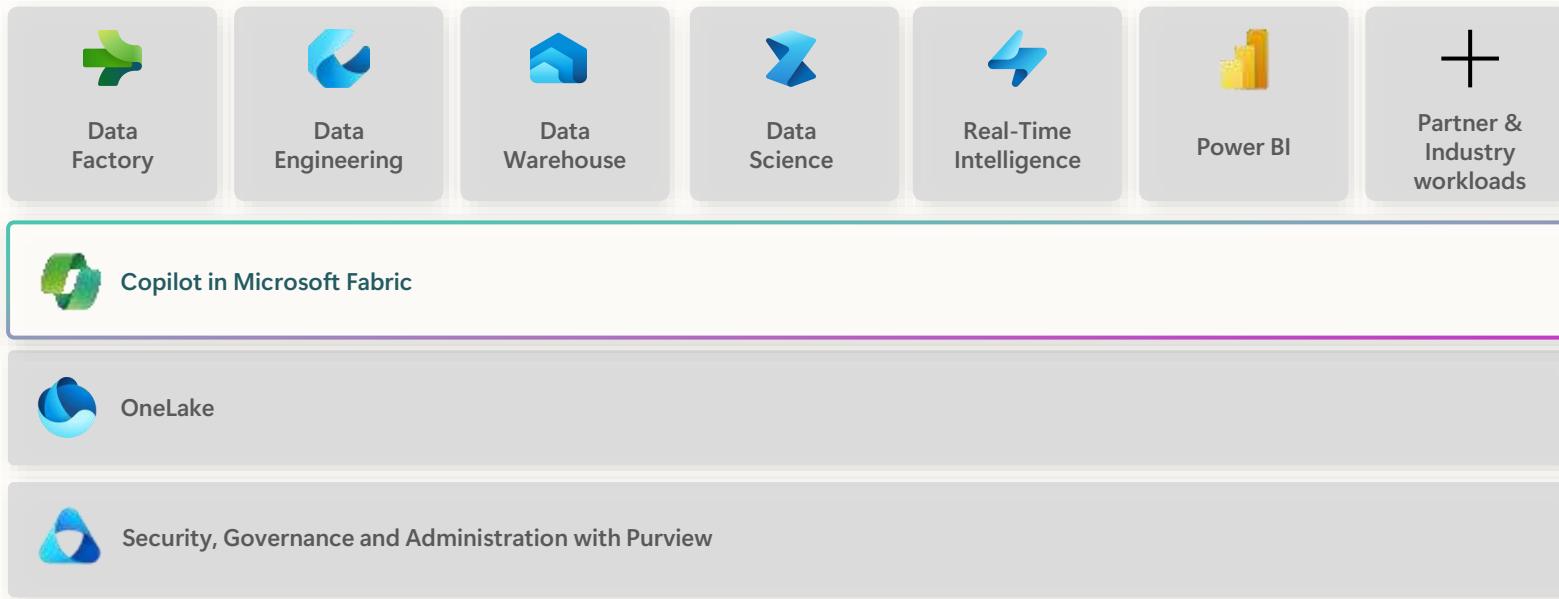


- Easy-to-use drag and drop canvas and visualizations for insightful and engaging report-building in seconds
- Native Integration with Microsoft 365
- Built-in AI capabilities and visuals illuminate hidden patterns, opportunities and anomalies with the click of a button
- Connect to, index, and certify datasets in the Power BI data hub
- Build governed databases, like data models or data marts, in a trusted and secure hub



Copilot in Microsoft Fabric

Use conversational language with Copilot in Fabric to create dataflows and pipelines, write SQL statements, build reports, and even build machine learning models

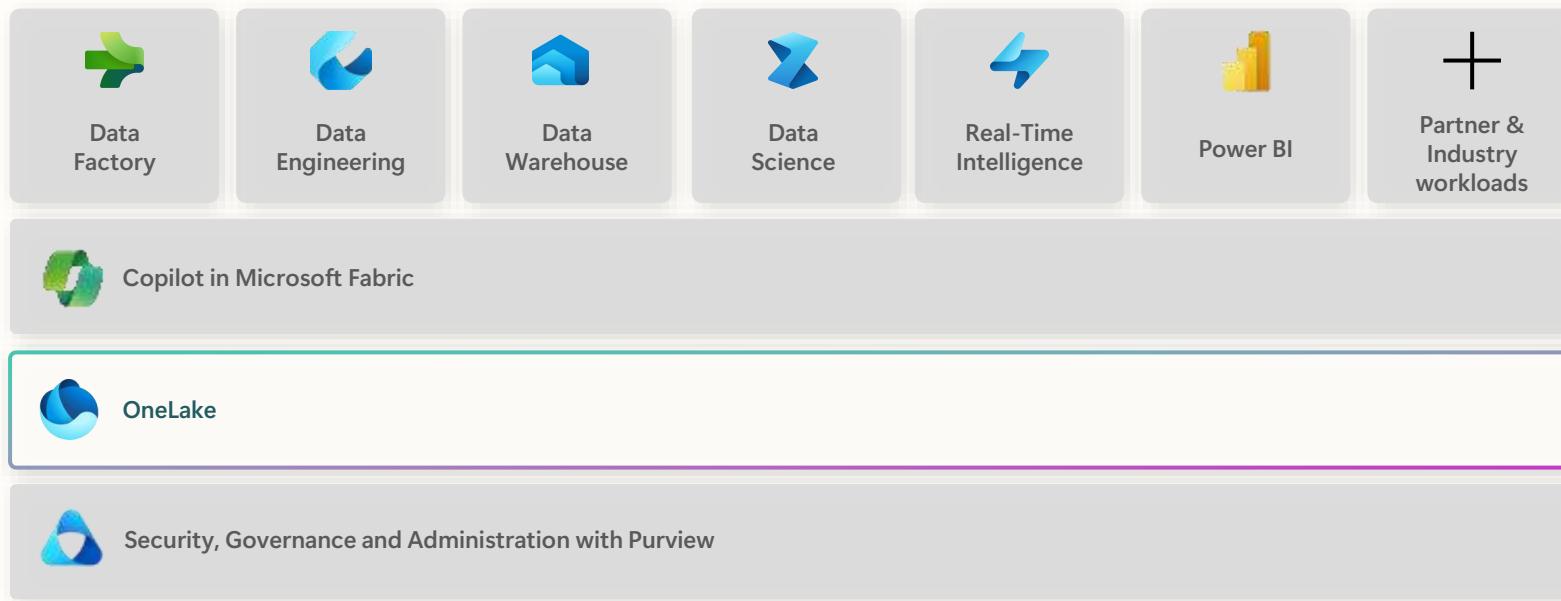


- ❯ Chat with AI assistant and request help handling data analysis
- ❯ Code more efficiently with intelligent code completion and generated code explanations
- ❯ More quickly enrich, model, analyze, and explore data all through natural languages
- ❯ Create Power BI reports automatically and summarize your insights for streamlined productivity
- ❯ Access industry-standard code templates to facilitate building robust data pipelines



Unified data foundation with OneLake

Manage and analyze all your data across your organization in a unified, secure, and centralized SaaS data lake for everyone with OneLake—the “OneDrive” for data

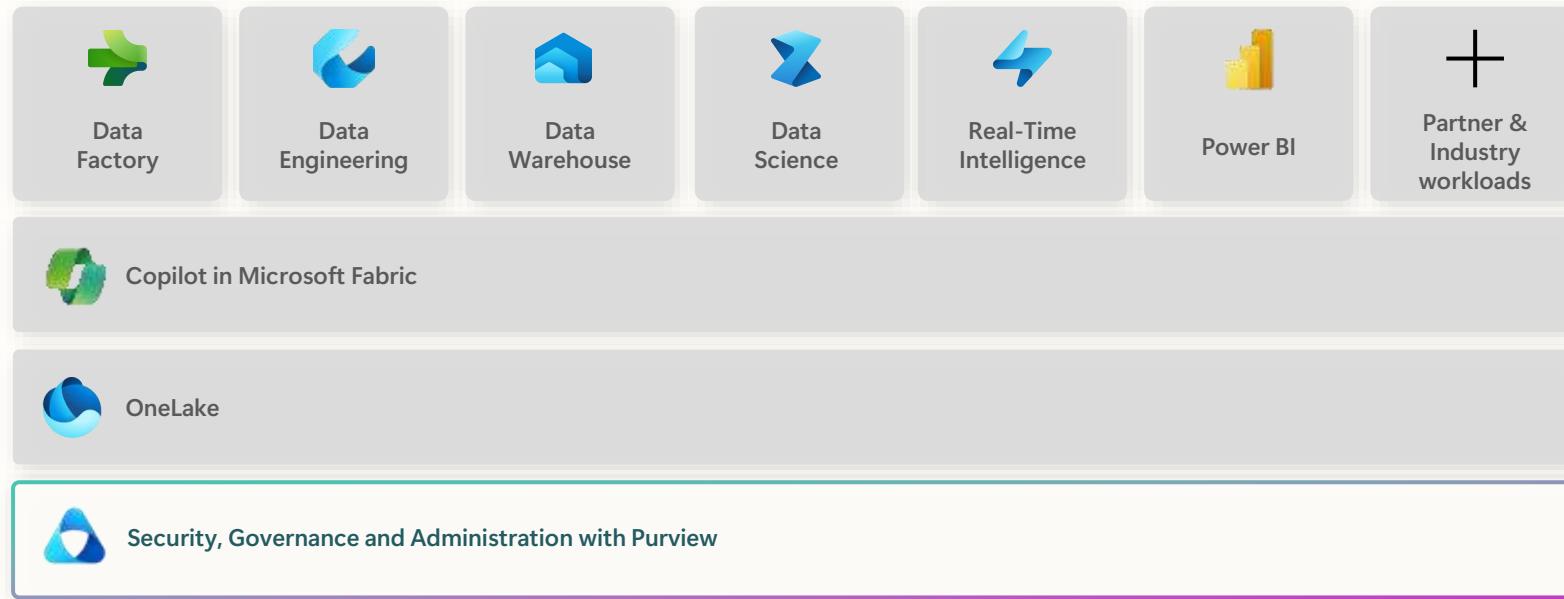


- A single and open, logical SaaS lake for the whole organization
- OneLake supports any type of file, structured or unstructured
- One copy of data for use with multiple analytical engines
- Enable virtualization of data without duplication using shortcuts
- All workloads automatically store their data in OneLake in Delta Parquet format
- Data in OneLake is automatically indexed for discovery, sharing, governance, and compliance



Security, Governance and Admin in Microsoft Fabric

Manage, secure, and govern all your data in Microsoft Fabric and beyond



- Reduce the effort needed to defend and control your entire analytics platform with out-of-the-box security and governance
- Secure your network from any intrusion, ensure only the right people have access to the right data, and maintain compliance with even the strictest requirements
- Enable different parts of the organization to take ownership of their data while still contributing to the same data lake
- Certify datasets to promote usage of the most accurate data across the organization
- Maintain the flexibility to use the partner and third-party solutions you want



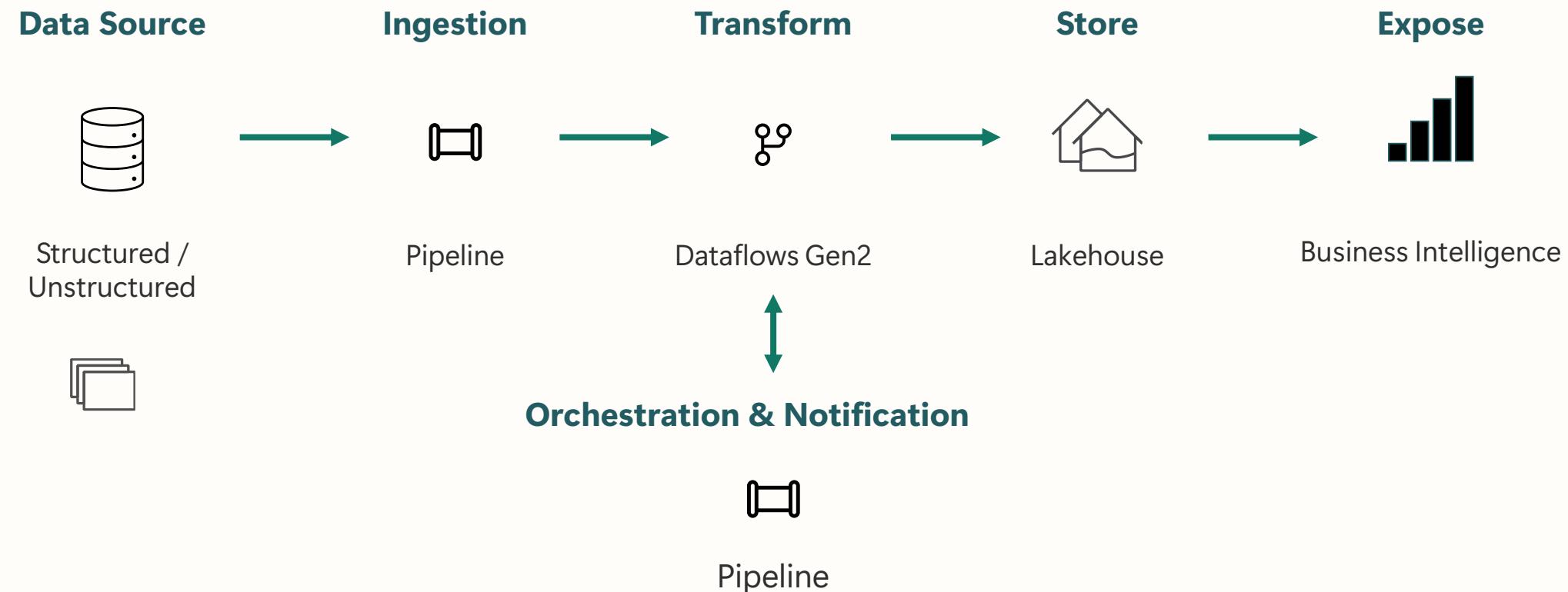
Data Factory





Data Factory scenario

End-to-end analytics scenario





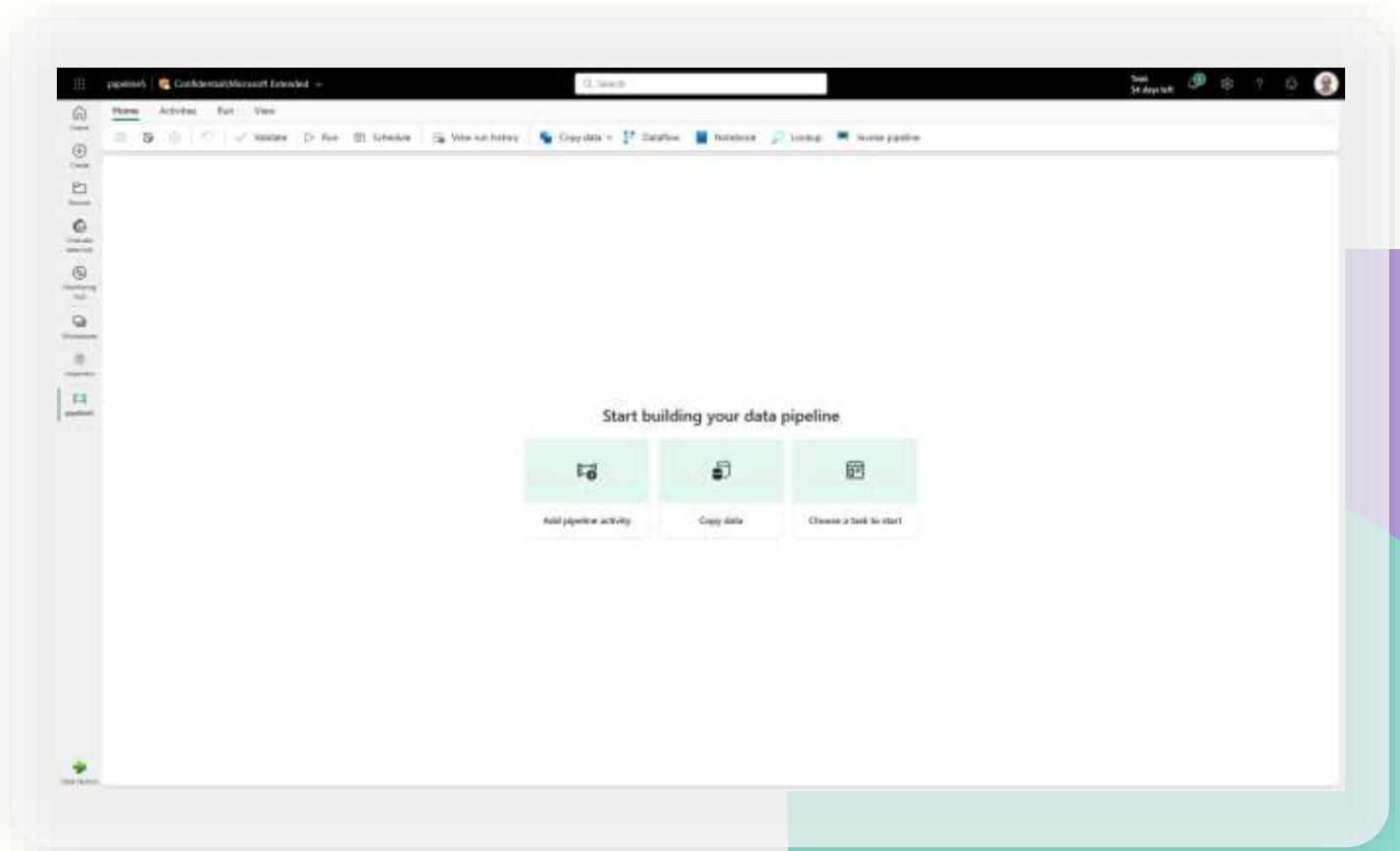
Data Factory

Data Factory in Microsoft Fabric provides cloud-scale data movement and data transformation services that allows you to solve the most complex ETL scenarios

Core to Data Factory are Data Pipelines and Dataflows to give users the option to a low-code, collaborative and enterprise scale approach for their ETL process.

Key Capabilities:

- Latest capabilities:
- Output destination to Lakehouse
- 170+ connectors available in Data Factory
- Pipeline Lakehouse copy assist
- Create data pipeline in Lakehouse portal
- Pipeline templates
- Pipeline support for Spark notebooks
- Service principle auth support



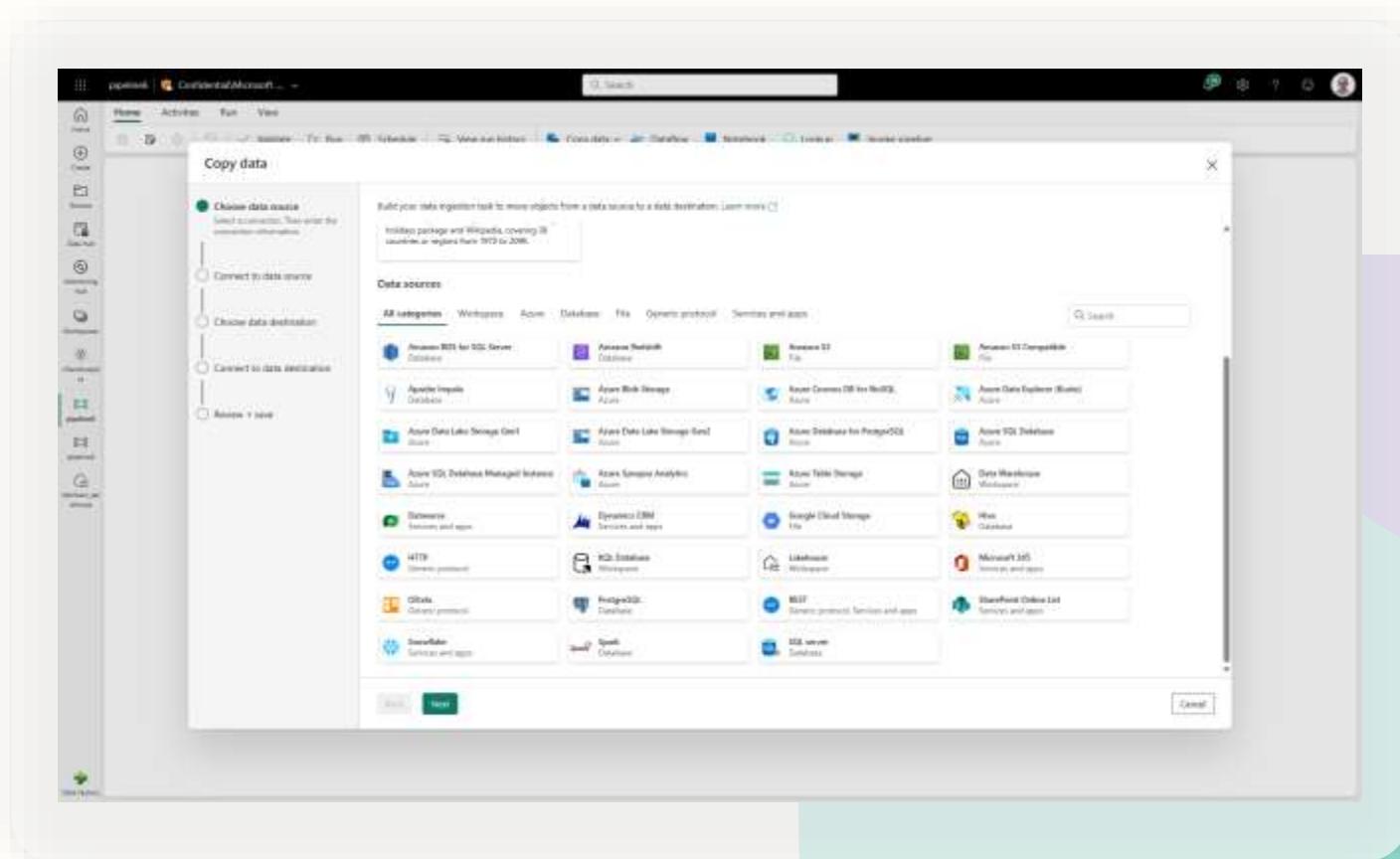
Autonomous ETL can unlock operational efficiencies and help orchestrate, monitor and manage pipeline performance.



Data Pipelines

Data Pipelines enable powerful workflow capabilities at cloud-scale like building complex workflows, moving PB-size data, and defining sophisticated control flow pipelines

Data pipelines can be used to build complex ETL and data factory workflows that can perform a number of different tasks at scale. Additionally, control flow capabilities are built into pipelines so you can build workflow logic which provide loops and conditional



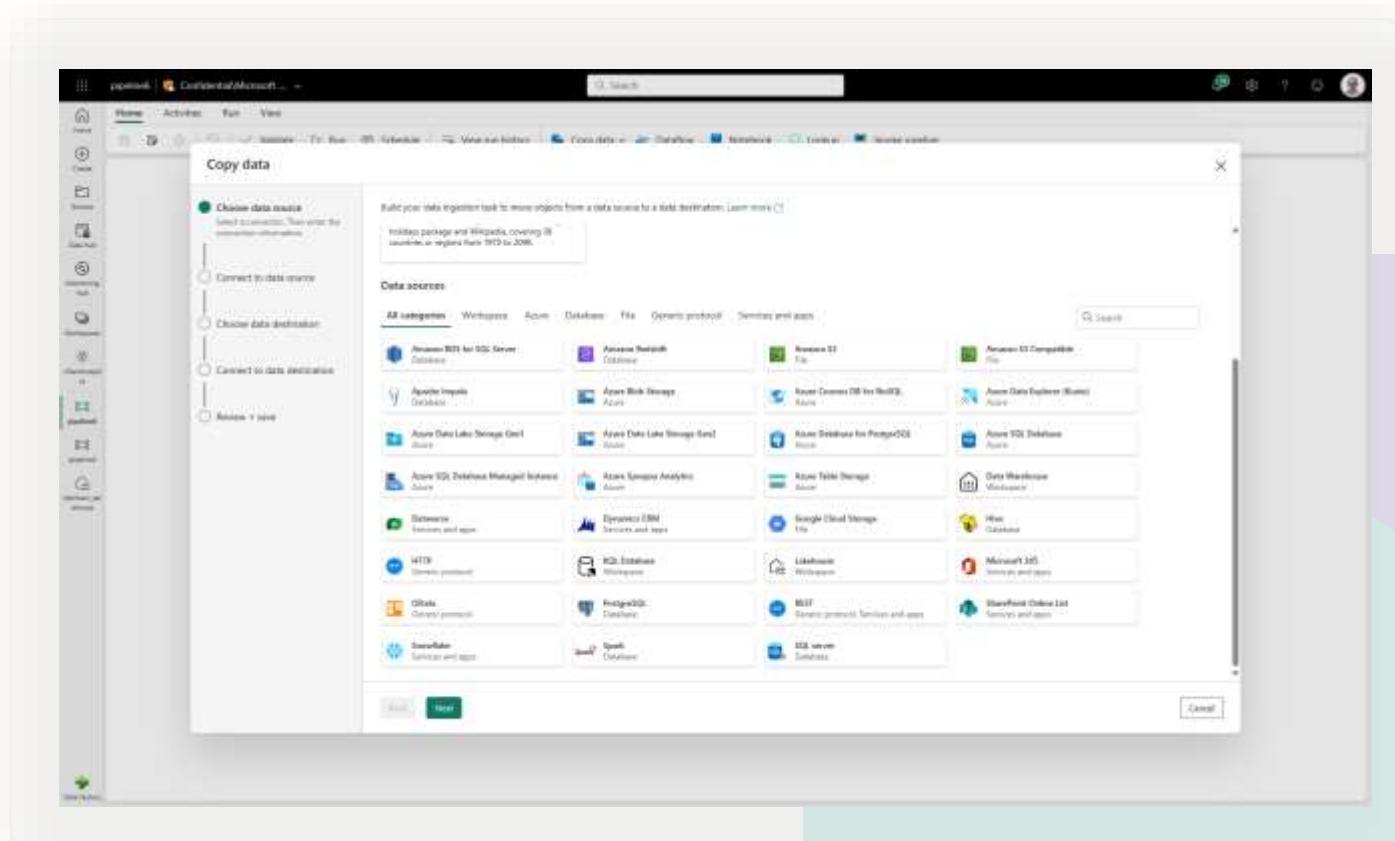


Data Pipelines | Connectors

New **Connectors** provide a low-code interface for ingesting data from a variety of data sources

Connectors:

- Warehouse Connector; connect to existing Azure
- Lakehouse connector
- 100+ connectors in the copy activity
- Access to on-premises data
- Access protected data inside of a VNET





Data Pipelines | Sample data

Sample Datasets helps new users get started quickly, building out their ELT processes using Data Pipelines

Sample datasets:

- COVID-19 Data Lake (CSV, JSON, JSON Lines, Parquet)
- NYC Tax – Green (2GB Parquet)
- Diabetes (14K Parquet)
- Public Holidays (500KB Parquet)
- Retail Data Model from Wide World Importers (352MB Parquet)

The screenshot shows the Microsoft Fabric Data Pipelines interface. A modal window titled "Copy data" is open, showing the "Choose data source" step. It lists four sample datasets:

- COVID-19 Data Lake (CSV, JSON, JSON Lines, Parquet)
- NYC Tax - Green (2GB Parquet)
- Diabetes (14K Parquet)
- Public Holidays (500KB Parquet)
- Retail Data Model from Wide World Importers (352MB Parquet)

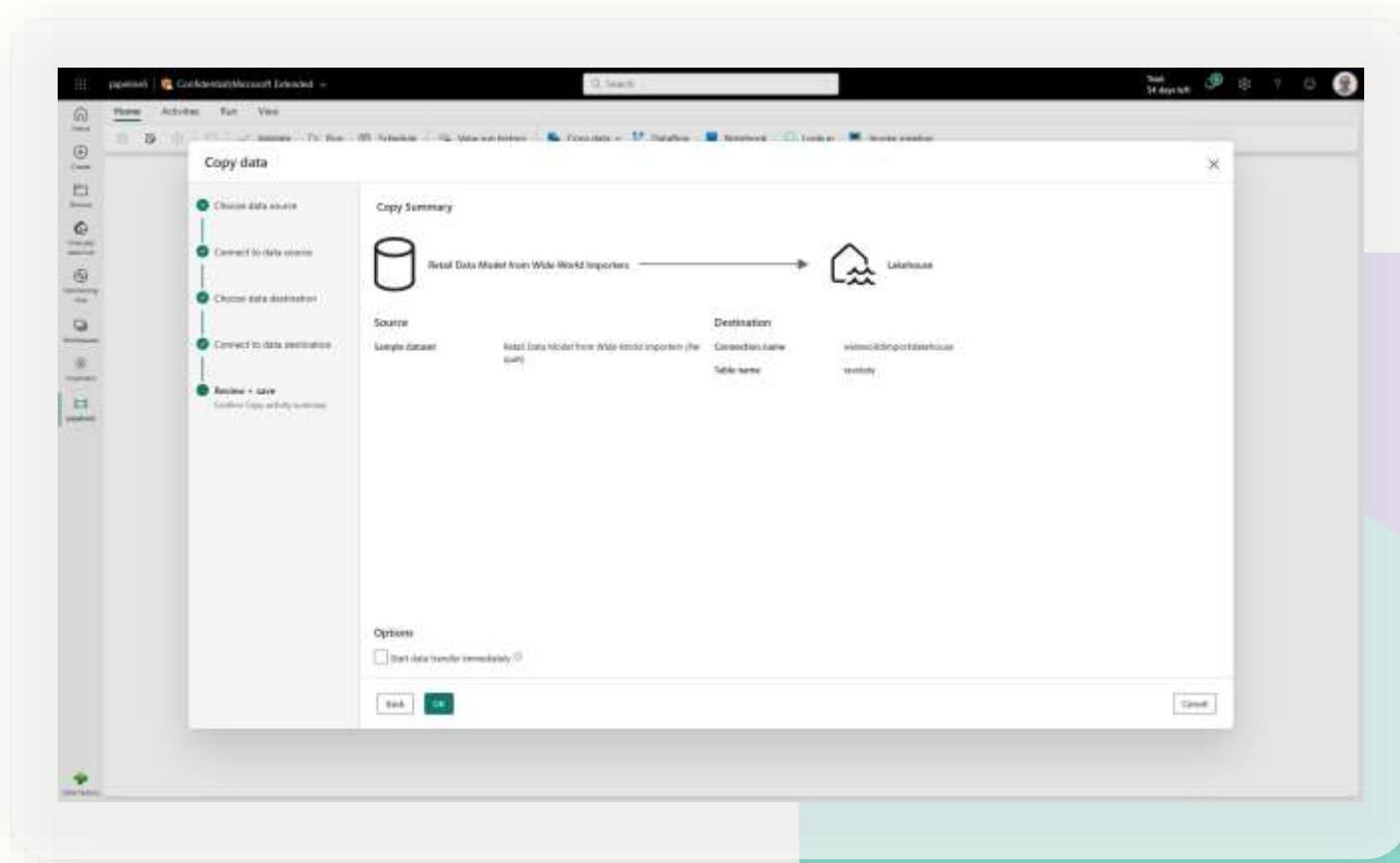
Below the datasets, a section titled "Data sources" lists various data storage options, including Amazon RDS for MySQL, Amazon Redshift, Amazon S3, Amazon Athena, Apache Impala, Azure Blob Storage, Azure Cosmos DB for NoSQL, Azure Data Explorer (Read-Only), Azure Data Lake Storage Gen1, Azure Data Lake Storage Gen2, Azure Database for PostgreSQL, Azure Table Storage, and Data Workbooks.



Data Pipelines | Lakehouse copy assist

Simply copying data to a Lakehouse with copy assist capabilities within the Data Pipeline

Additionally, users can create a Data Pipeline without having to leave the Lakehouse portal





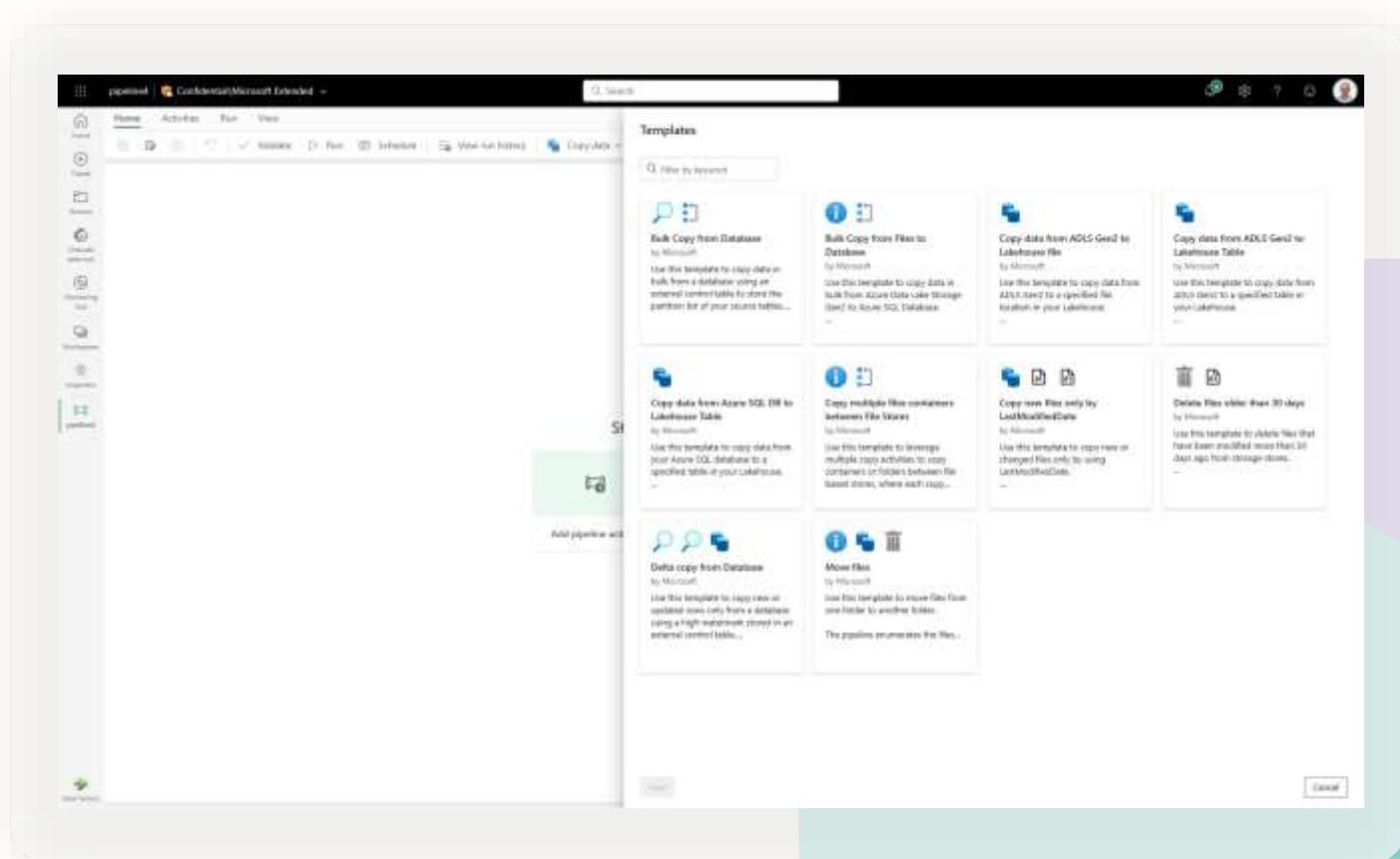
Data Pipelines | Templates

Quickly get started with data integration

Template help reduce development time by providing an easy way to create pipeline for common data integration scenarios.

Available Data Pipeline Templates:

- Bulk copy from Database
- Bulk copy from File to Database
- Copy data from ADLS Gen2 to Lakehouse file
- Copy from ADLS Gen2 to Lakehouse Table
- Copy data from Azure AQL DB to Lakehouse Table
- Copy multiple files containers between File Stores
- Copy new files only by Last Modified Date
- Delete files older than 30 days
- Delta copy from Database
- Move files





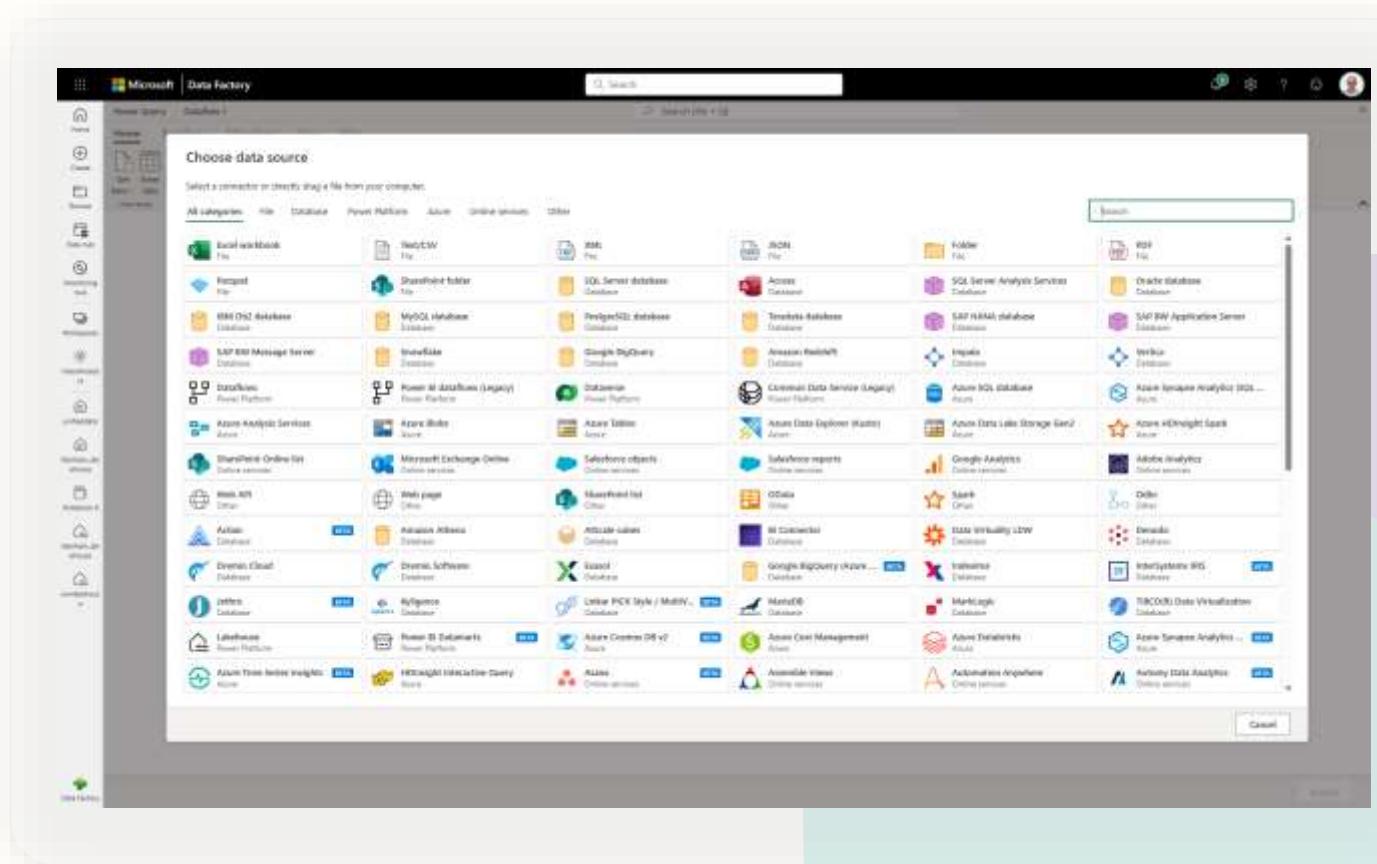
Dataflow

Dataflow provides a low-code interface for ingesting data from hundreds of data sources

Dataflow quickly and easily unify disparate data sources, establish a more collaborative analytics approach, and promote more informed, agile decision making.

Key Capabilities:

- Accelerate data transformation with code-free data flows
- Scale out using Fabric compute and Data Factory fast copy
- Load results of data transformations into multiple destinations (Azure SQL Databases, Lakehouse, etc.)





Dataflow | Output to Lakehouse

Simply write into a Lakehouse from a Dataflow

Users select the Lakehouse output destination from the list and configure the connection.

This requires the Lakehouse connector to be installed as a custom connector into your data gateway when loading data from on-premise

The screenshot shows the Microsoft Synapse Data Engineering interface. The top navigation bar includes 'Home', 'Transforms', 'Add column', 'Time', and 'Help'. Below the navigation bar, there are several tabs: 'Power BI', 'Synapse Analytics (SQL DWH)', 'Azure Data Explorer (Kusto)', 'Azure Synapse Analytics (SQL DWH)', and 'Lakehouse'. The 'Lakehouse' tab is highlighted with a red box. On the left, there's a sidebar with 'Queries [1]', 'Task ID: 10000000000000000000000000000000', and a 'Run' button. The main area displays a table with columns: 'Index', 'Name', 'Type', 'Last modified', and 'Description'. The table contains numerous rows, each representing a different table or view, such as 'Wingtip Trax Head Office', 'Wingtip Trax Head Office', 'Inventory', 'Wingtip Trax', '2023-01-12T08:00:00Z', and 'Inventory table'. A 'Query settings' pane on the right shows 'Name: dimension_customer ?' and 'Entity type: Dimension'. It also lists 'Applied steps': 'Source', 'Processors', and 'Target'. The bottom right corner of the interface has a purple decorative shape.

Unifying data in OneLake

Data Factory



Azure Database for PostgreSQL	Azure Databricks Delta Lake	Amazon RDS for Oracle	Amazon RDS for SQL Server	Amazon Redshift	Phoenix	PostgreSQL	Presto	Magento (Preview)
Azure SQL Database	Azure SQL Database Managed Instance	Apache Impala	Azure SQL Database Managed Instance	DB2	SAP BW	SAP BW via MDX	SAP HANA	Oracle Eloqua (Preview)
Azure Table Storage	MongoDB Atlas	Drill	Google AdWords	Google BigQuery	SAP Table	SQL server	Spark	PayPal (Preview)
Azure Cosmos DB (MongoDB API)	Azure Cosmos DB (SQLAPI)	Greenplum	HBase	Hive	Amazon S3	Amazon S3 Compatible	FTP	SAP Cloud For Customer
Azure Data Lake Storage Gen1	Azure Data Lake Storage Gen1 for Cosmos Structured Stream	Informix	MariaDB	Microsoft Access	File system	Google Cloud Storage (S3APD)	HDFS	Salesforce Marketing Cloud
Azure Data Lake Storage Gen1 for MariaDB Structured Stream	Azure Database for MariaDB	MySQL	Netezza	Oracle	HTTP	Oracle Cloud Storage (S3AP)	SFTP	Shopify (Preview)
teradata	VERTICA	ODBC	OData	REST	Amazon Marketplace Web Service	Concur (Preview)	Dataverse (Common Data Service for App)	Web Table
Jira	Kusto	SharePoint Online List	Dynamics 365	Dynamics AX	Dynamics CRM	Cassandra	Couchbase (Preview)	MongoDB

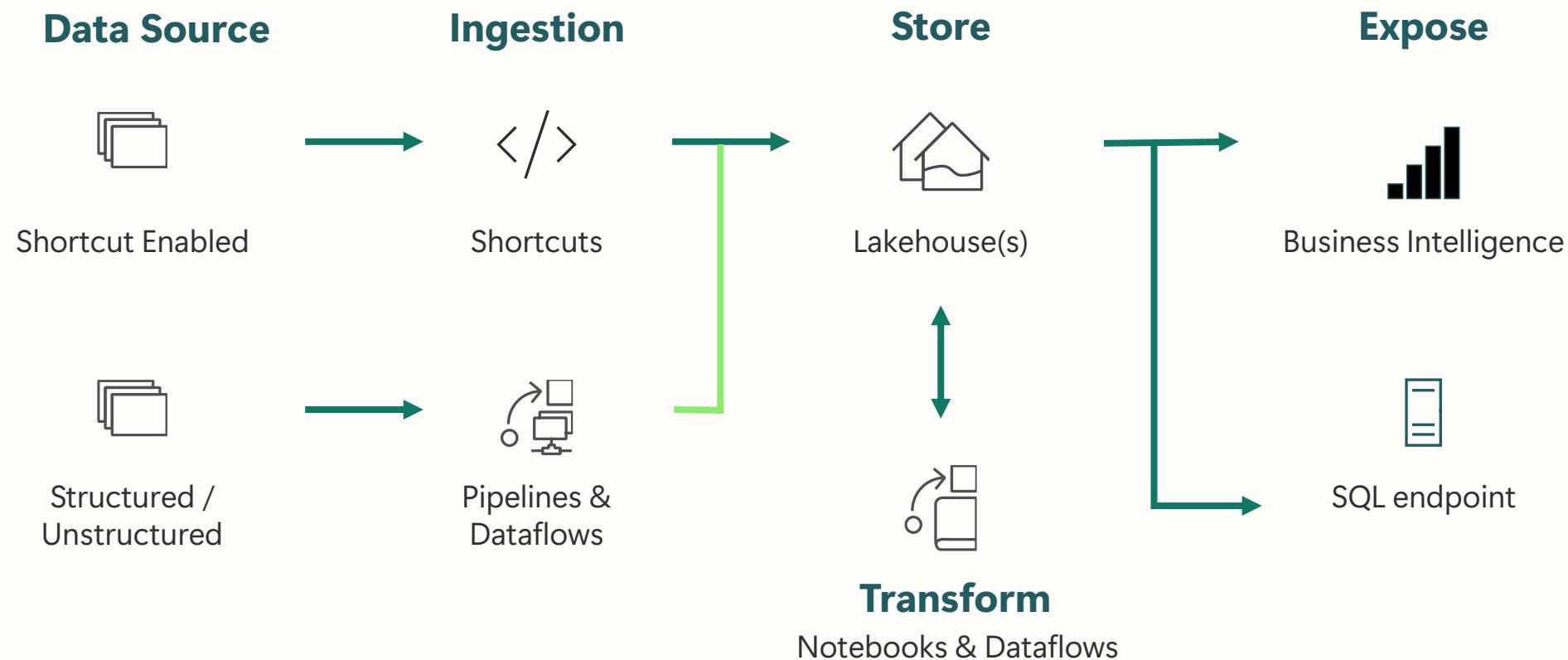
Data Engineering





Lakehouse scenario

End-to-end analytics scenario





Lakehouse | Overview

Store, manage and analyze all your data in a single location & easily share across the entire enterprise

Quickly and easily create a Lakehouse without having to provision and configure compute, storage and networking

Key Capabilities:

- Flexible and scalable solution that enables organizations to handle large data volumes of all types and sizes
- Built-in SQL endpoint unlocks data warehouse capabilities on top of your Lakehouse with no data movement
- Use 'direct lake' mode to build reports in seconds directly on top of the data lake with blazing fast performance
- Easily ingest data into the Lakehouse through a variety of methods
- Share your Lakehouse as a data product with consumers

The screenshot shows the Microsoft Fabric Explorer interface. On the left, there's a navigation sidebar with icons for Home, Data, Compute, Storage, Monitoring, and Workbooks. The main area has tabs for 'Home', 'Get data', 'New Power BI dataset', and 'Open notebook'. The 'Home' tab is selected. In the center, there's a tree view under 'ImportersLakehouse' with nodes for 'Tables' (including dimension_city, dimension_customer, dimension_date, dimension_employee, dimension_stock_item, fact_order, fact_order_line, fact_sales, fact_sales_line, fact_sales_tx, fact_sales_tx_line, fact_sales_tx_line_invoiced), 'File' (with a 'new-new-data' folder containing 'part_by_full'), and 'Incremental' (with 'fact_sales_tx_invoiced'). To the right is a table listing files in the 'part_by_full' folder. The table columns are 'Name', 'Date modified', 'Type', and 'Size'. The files listed are: SUCCESS, parn-00000-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00001-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00002-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00003-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00004-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00005-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00006-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00007-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00008-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00009-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00010-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00011-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00012-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet, parn-00013-ced40ca-4bd8-45e9-8526-5ca82f56f7e-1000.parquet.



Spark Runtime

Get started with running Spark in a matter of seconds

Instance performance with starter pools of Spark compute readily available

Key Capabilities:

- High concurrency mode enables users to share Spark sessions across multiple notebooks
- Attaching a notebook to an existing session results in lightning-fast session start up speed <5 seconds.
- Delta Lake enables 'OneCopy' allowing all Microsoft Fabric compute engines to seamlessly share data
- All Microsoft Fabric items automatically write Delta with V-Order optimization providing lightning-fast analytics for reporting with Power BI

The screenshot shows the Microsoft Synapse Data Engineering workspace interface. On the left, the 'Importers' blade is open, displaying a list of notebooks and experiments. The notebooks listed include: 01 - Input data into Delta Lake using Apache Flink, 01 - Create Delta Tables, 02 - Data Transformation - Business Aggregation, 03 - Explore and Visualize Data using Notebooks, 04 - Perform Data Cleaning and processing using Flink, 05 - Train and track machine learning models, 06 - Perform batch scoring and save predictions to a database, 07 - Create Delta Tables, 1 - Input data into Delta Lake using Apache Flink, 2 - Data Transformation - Business Aggregation, 2 - Explore and Visualize Data using Notebooks, Create Table, 03 - Calculate Stock Summary, Customer Summary Report, Customer Card Model, Customer Card Model, and Data Transformation. The 'Spark compute' settings dialog is open on the right, showing configuration options for a 'Default pool for workspace'. It includes sections for 'Pool details' (Node family: Memory optimized, Node size: Medium, Number of nodes: 1-10), 'Runtime Version' (Set to '1.1 (Spark 3.3, Delta 2.2)'), 'Automatically track machine-learning experiments and models' (On), and 'Spark properties' (448).



Spark Runtime | Performance

Continuous optimized & built-in autotuning

Delivering increased performance by default

Key Capabilities:

- **Increased performance** with Fabric runtime 1.1, which brings major updates such as upgrading Spark to 3.3.1, Delta to 2.2 and Python to 3.10

- **Partition Cache** - enhance your query performance by storing filtered partition information in a session-level cache (~10% performance boost).

- **Cross join multiple scalar subqueries** – merge them in a single plan to reduce computation time (~5% performance improvement)

The screenshot shows the Microsoft Fabric Data Explorer interface. On the left, the Lakehouse Explorer sidebar displays a tree view of tables and files under the 'lakehouse' database, including 'dimension_cpy', 'dimension_customer', 'dimension_date', 'dimension_employee', 'dimension_product', 'hyperlane', and 'raw_cdc_data'. The main workspace shows a Python script titled '1.mr - Command executed at 12:01 PM 10/18/2023'. The script contains code to save cleansed and prepared data to a lakehouse delta table. Below the script, a table titled 'Spark Jobs (5 of 6 succeeded)' lists five completed jobs with details like ID, Description, Status, Stages, Task, Duration, Processed, Data read, and Data written. A purple decorative shape is visible on the right side of the slide.

```
1. # Save Cleansed and prepared data to lakehouse delta table
2. TABLE_NAME = "hyperlane_prc"
3. import delta.tables as dt
4. print(dt.DeltaTable.list("evergreen").format("delta").save(f"Tables/{TABLE_NAME}"))
5. print(f"Spark database saved to delta table: {TABLE_NAME}")
6. final_note: str = f"Automation process to save data to lakehouse delta table has been completed successfully"
7. print(final_note)
8. print("-----")
9. print("-----")
```

ID	Description	Status	Stages	Task	Duration	Processed	Data read	Data written
job_14	Janitorial tasks performed at 10:49 AM 10/18/2023	0/0 succeeded	0/0	0/0 succeeded	< 1 sec	0.0 sec	16.44 KB	0.0
job_15	Save of湖州湖数据完成	0/0 succeeded	0/0	0/0 succeeded	1 min 9 sec	43,460,000 rows	121,932.0 KB	0.0 KB
job_16	Building a String	0/0 succeeded	0/0	0/0 succeeded	< 1 sec	0.0 sec	0.0	0.0
job_17	Building a String	0/0 succeeded	0/0	0/0 succeeded	< 1 sec	0.0 sec	18.49 KB	16.71 KB
job_18	Building a String	0/0 succeeded	0/0	0/0 succeeded	< 1 sec	0.0 sec	18.71 KB	4.25 KB



Spark Runtime | Monitoring

Monitor the progress & built-in analysis of Spark applications, browse historical jobs, and view related activities

The Monitoring Hub serves as a centralized portal for all jobs including your Spark applications

Key Capabilities:

- Dive into a more detailed analysis exploring Spark jobs, task information
- View all related items along with their respective snapshots and associated code input & output parameters
- Spark UI and History Server integration offers native open-source capabilities to augment Spark monitoring and debugging

The screenshot shows the Microsoft Synapse Data Engineering Monitoring hub interface. The top navigation bar includes 'Monitoring hub' and a search bar. On the left, there's a sidebar with icons for Home, Code, Databricks, Monitoring hub, Workspaces, and Monitoring dashboards. The main content area is titled 'Monitoring hub' with the sub-instruction 'Monitoring hub is a station to view and track active activities across different products.' Below this is a table with columns: Name, Status, Item Type, Start time, Duration, Submitter, Location, and Job type. The table lists several completed and in-progress jobs, such as 'NotebookBugDash_47628516-d7c-40fe-91dd...' and 'NotebookBugDash_a1921dd_9032_477f-ad01-4...', each with details like duration (e.g., 10m 5s, 8m 14s) and submitter (e.g., Jerry Jiang). A purple decorative shape is visible on the right side of the slide.

Name	Status	Item Type	Start time	Duration	Submitter	Location	Job type
NotebookBugDash_47628516-d7c-40fe-91dd...	In progress	Notebook	2:36 AM, 4/26/23	1m 4s	Jerry Jiang	Monitoring Bugla...	SparkSession
NotebookBugDash_a1921dd_9032_477f-ad01-4...	Completed	Notebook	2:17 AM, 4/26/23	10m 5s	Jerry Jiang	Monitoring Bugla...	SparkSession
NotebookBugDash_9500443d-e31b-4634-8234...	Completed	Notebook	2:01 AM, 4/26/23	8m 14s	Jerry Jiang	Monitoring Bugla...	SparkSession
NotebookBugDash_295427b7-ba40-4ebc-b2b5...	Completed	Notebook	1:58 AM, 4/26/23	49s	Jerry Jiang	Monitoring Bugla...	SparkSession
NotebookBugDash_41677391-dafe-435c-b406...	Completed	Notebook	1:49 AM, 4/26/23	6m 47s	Jerry Jiang	Monitoring Bugla...	SparkSession
NotebookBugDash_4734668e-2a99-4853-897b...	Completed	Notebook	1:23 AM, 4/26/23	2m 49s	Jerry Jiang	Monitoring Bugla...	SparkSession
Notebook_10_be34b4c1-6ff7-4528-ab66-c0f5b...	Completed	Notebook	12:39 AM, 4/26/23	4m 6s	Jerry Jiang	Monitoring Bugla...	SparkSession
Notebook_10_c3dd316e-6221-46d0-8fa7-57e93...	Completed	Notebook	12:23 AM, 4/26/23	1m 48s	Jerry Jiang	Monitoring Bugla...	SparkSession
NotebookBugDash_732067b2-4135-4ba9-b4b6...	Completed	Notebook	12:22 AM, 4/26/23	21m 20s	Jerry Jiang	Monitoring Bugla...	SparkSession
Bugdash11_TableLoad_ab31a3cc-e1b0-4a0d-974d...	Completed	Lakehouse	11:37 PM, 4/25/23	2hr 25s	Jerry Jiang	Monitoring Bugla...	SparkSession
Bugdash11_TableLoad_530ce5dc-a758-4ff6-9614...	Completed	Lakehouse	11:36 PM, 4/25/23	2hr 25s	Jerry Jiang	Monitoring Bugla...	SparkSession
Notebook_20_8ff7d36-10e7-4ba1-ac97-983988...	Completed	Notebook	11:35 PM, 4/25/23	5m 1s	Jerry Jiang	Monitoring Bugla...	SparkSession



Configure your workspace | Environment

Create multiple environments to configure and customize your workspace

Key Capabilities:

- Select your preferred Spark runtimes
- Configure your compute for a better resources allocation within a workspace
- Install your libraries from public repository (PiPl, Conda) or upload local custom-built libraries
- Use customized environments by attaching them to a Notebook or Spark job definition

The screenshot shows the Microsoft Fabric Environment configuration interface. The top navigation bar includes 'Microsoft' and 'Environment' tabs, a 'Confidential/Microsoft Extended' workspace indicator, a 'Saved' dropdown, a search bar, and a trial status message ('Trial 59 days left'). The left sidebar has a 'Home' icon, a 'Create' button, a 'Workspaces' icon, a 'Pins + Fabric' icon, and a 'Environments' section with a 'Recent Environments' dropdown and a 'Data Engineering' icon. The main content area is titled 'Public Libraries' and displays a table of installed libraries. The table columns are 'Library', 'Version', 'Source', and 'Status'. One row is visible: 'missingno' at version '0.5.2' from 'PyPI' with a 'Success' status. A 'Filter by name:' input field is also present.

Library	Version	Source	Status
missingno	0.5.2	PyPI	Success



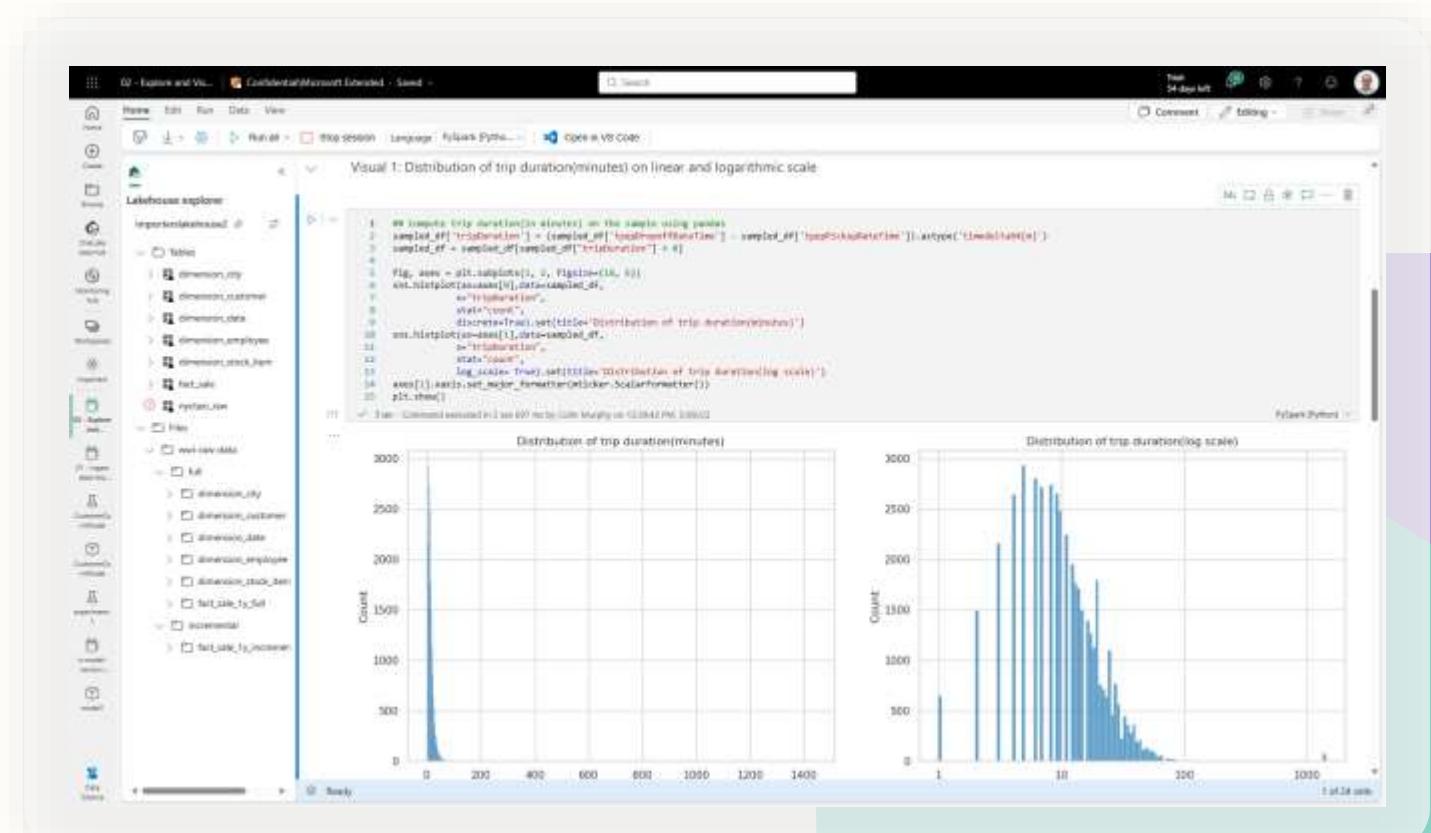
Notebook | Overview

Immersive authoring experience for data developers

Rich notebook capabilities including native Lakehouse integration, real-time collaboration with commenting support, auto-save support, lightweight scheduling and pipeline integration

Key Capabilities:

- Manage your Python and R libraries through in-line installs using commands like %pip install
- Advanced notebook development support with ability to reference notebooks in notebooks, and snapshots for easy troubleshooting
- In context monitoring complete with real time advice and error analysis
- Streamline data prep without giving up the power of reproducibility of Python





VS Code | Integration

Empower pro developers to use their tools of choice for their development

Launch the Fabric VS Code extension, navigate through your workspace and work with Notebooks, Spark Jobs and Lakehouses directly in the IDE

Key Capabilities:

- Run and debug notebook cells or standard batch jobs on the remote Spark cluster
- Have the flexibility to work offline with a local environment and push your changes online when needed

The screenshot shows the Microsoft Fabric VS Code extension interface. The left sidebar displays a 'Notebook' workspace with a single notebook titled 'Marketing Campaign.ipynb'. The main area shows two code cells. The first cell contains Python code to load a parquet file from a LakeHouse:

```
# Step 2: Load parquet file from LakeHouse
df = spark.read.parquet("files/back-additional-FULL.parquet")
# df.out is a spark DataFrame containing parquet data from "files/back-additional-FULL.parquet".
df.show(10)
```

The second cell contains Python code to use a summary tool to identify quality issues and clean the data:

```
# df_noMissing = df.filter(~df.previous.isna())
# df_result = df_noMissing.withColumn("Age", df_noMissing["Age"].cast("int"))
# df_noMissing = df_noMissing.withColumn("previous", df_noMissing["previous"].cast("int"))
# df_result.show(10)
```

Below the code cells, the terminal output shows:

```
Check if synapse spark kernel env has created... True
```

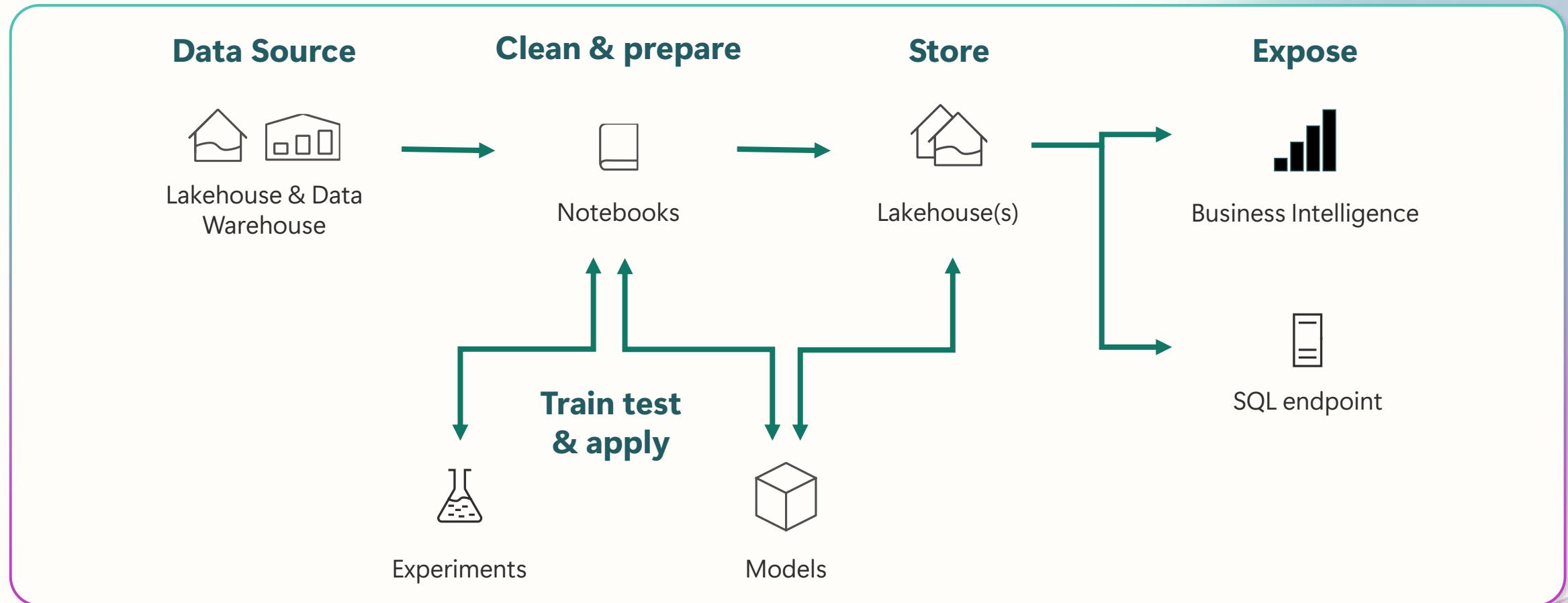
Data Science





Data science scenario

End-to-end analytics scenario





Machine learning with MLFlow

Built-in model & experiment tracking enables data scientists to track and compare their different experiment runs and model versions

Train, evaluate, and score machine learning models by using built-in Experiment and Model artifacts

Key Capabilities:

- Seamless integration with MLFlow for experiment tracking and model registration/deployment
- Automatically capture model metrics & parameters with built-in support for MLFlow auto-plogging
- ML integration with Azure OpenAI service makes it easier to use Apache Spark to process millions of prompts and scale workflow
- Model registry powered by AzureML

Name	Type	Owner	Reflected	Last run	Assessment	Scalability	Included in app
housing-mlflow-4	Supervised	Model Dev	Unassessed	...
model-housing-v1	None	Model Dev	Unassessed	...
Resolute-Delta	None	Model Dev	Unassessed	...
Search	None	None	None	None	None	None	None
Tensor	None	None	None	None	None	None	None
Tensor	None	None	None	None	None	None	None



Model scoring

Easily enrich your data stored in the Lakehouse with machine learning predictions

Leverage Predict to score data at scale and write it back to your Fabric Lakehouse

Key Capabilities:

- Predict can be invoked from the secure boundaries of a Fabric notebook using APIs for maximum flexibility
- Leverage Predict directly from the UI using a guided experiment to deploy models in a matter of clicks
- Seamless integration with the Lakehouse, OneLake and Direct Lake meaning it is super simple to get your predictions into Power BI reports

Input column name	Model input name	Model input data type
Median (double)	Median	double
HouseAge (double)	HouseAge	double
Available (double)	Available	double
Available (double)	Available	double
Population (double)	Population	double
AvgOccup (double)	AvgOccup	double
Latitude (double)	Latitude	double
Longitude (double)	Longitude	double



Get started with Data Science in Fabric

Get started with a rich set of data science samples for a variety of common use cases.

End-to-end examples of how to solve common problems with Machine Learning in MS Fabric

Newly added R samples and Quick tutorials

Use a sample

End-to-end workflows (Python) End-to-end workflows (R) Quick tutorials

Book recommendation Recommend books to readers using collaborative filtering.	Fraud detection Detect credit card fraud.	Time series Forecast New York City property sales by season.
Title genre classification Predict whether a book in the British Library is fiction or nonfiction.	Uplift modelling Estimate the causal impact of treatments on behavior.	Machine failure Predict when manufacturing equipment will fail.
Customer churn Predict when a bank customer will stop doing business with the bank.	Sales forecasting Predict sales numbers for product categories at a superstore.	



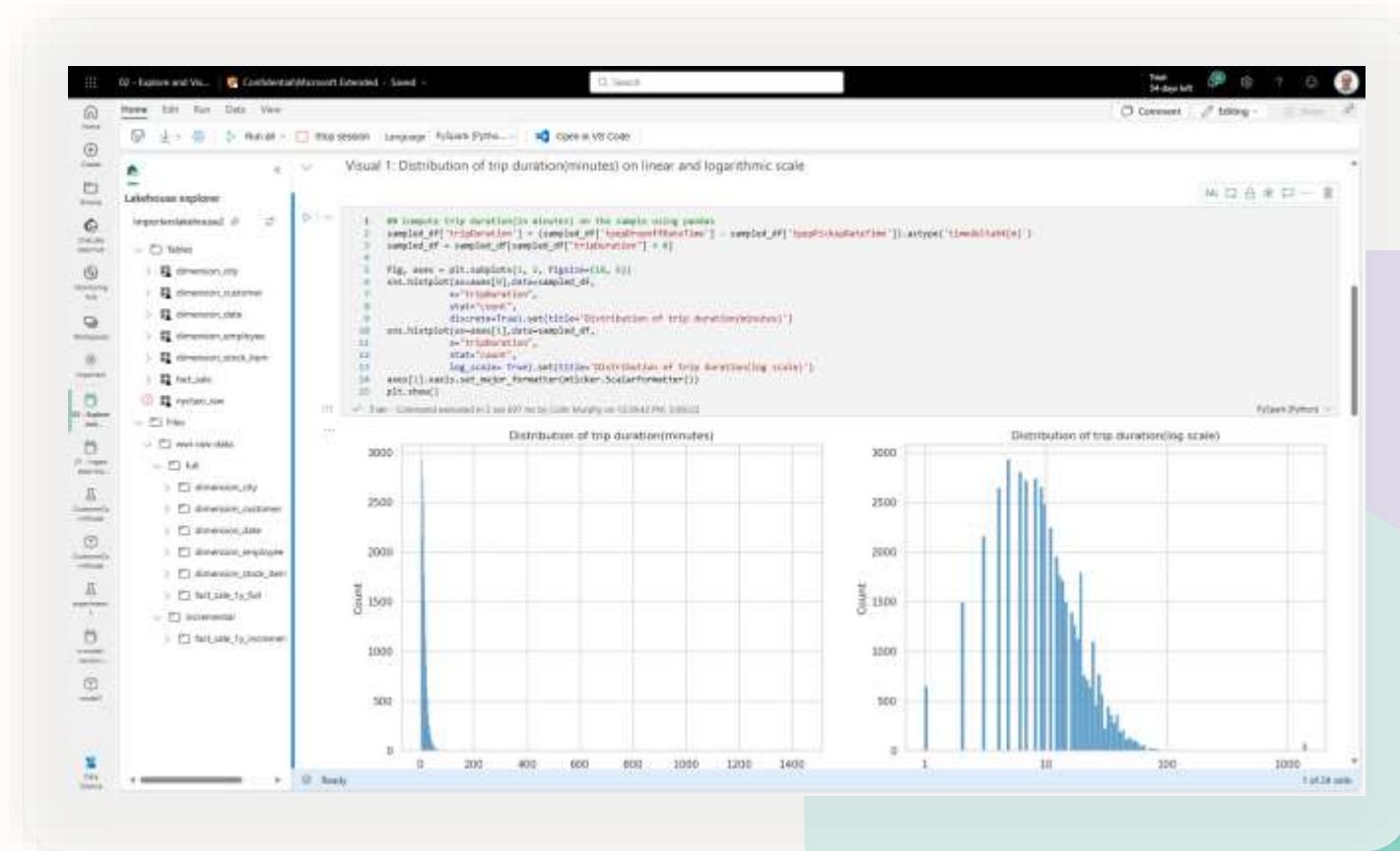
Explore, clean and prepare

Faster exportation, cleansing and transformation of data with built-in experiences as part of the data science workload

Data Science workloads supports data cleansing, transformation, exploration, and featurization by leveraging built-in experiences on Spark as well as Python based tools

Key Capabilities:

- Accelerate tedious and mundane tasks with Data Wrangler, a code tool to prepare data and generate Python code
- Simplify data analytics, capture and exploit data semantics as users perform transformations on datasets
- Leverage pre-installed open-source libraries for building visualizations like matplotlib, seaborn, Plotly, and more





Code generation with Data Wrangler

Streamline data prep without sacrificing the power and reproducibility of Python

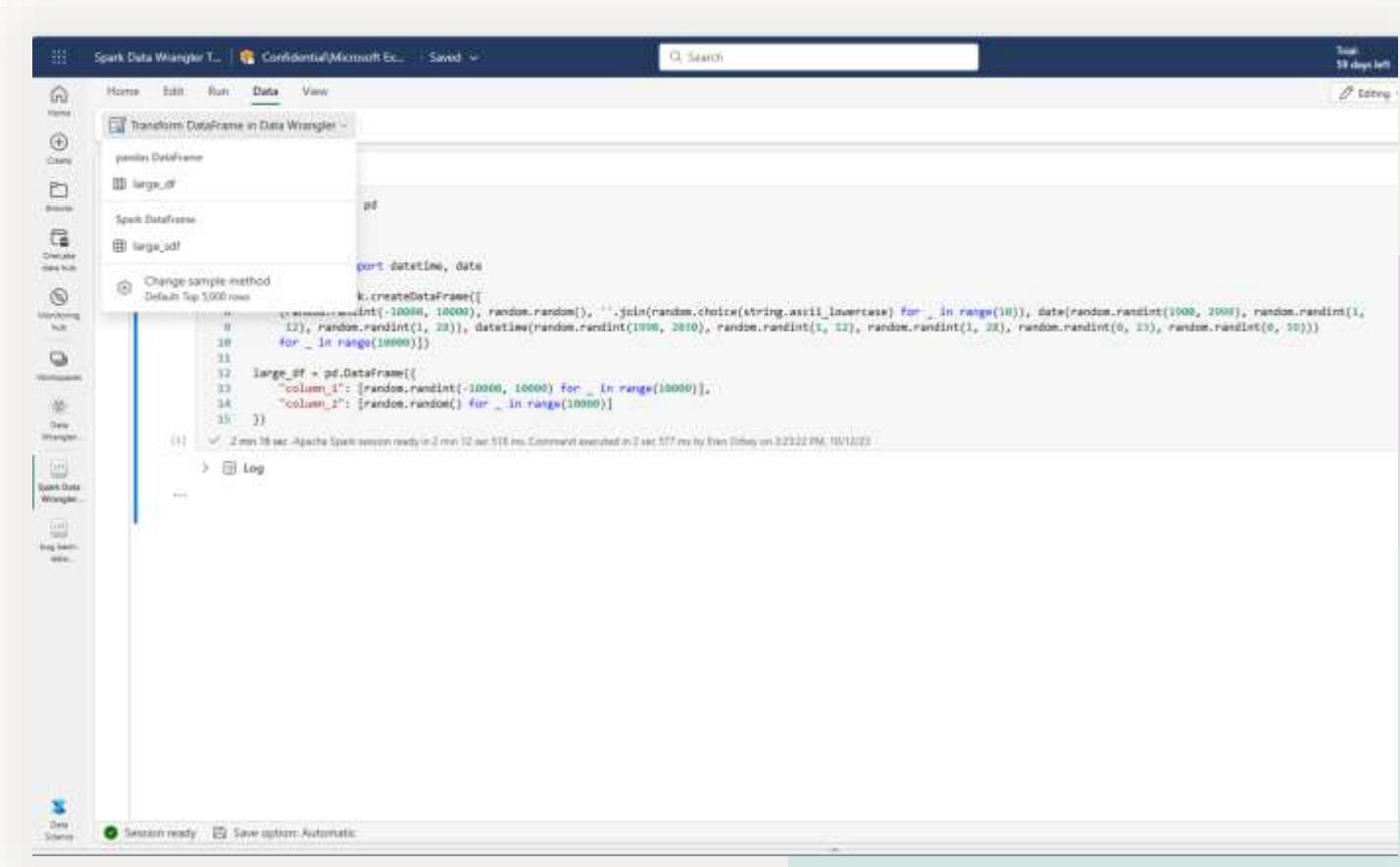
Dynamic data display

- Built-in stats and charts
- Pandas in a matter of clicks

Designed to onboard newer users and to accelerate pro developers

Spark dataframe support

- NL-to-code



The screenshot shows the Microsoft Fabric Data Wrangler interface. On the left, there's a sidebar with navigation links like Home, Edit, Run, Data, View, and a session status indicator "Session ready". The main area has tabs for "Transform DataFrame in Data Wrangler" and "Spark DataFrame". Under "Spark DataFrame", there are two sections: "large_df" and "large_sdf". A code editor window titled "Change sample method" displays Python code for generating a DataFrame:

```
1    pd = pd.read_csv("large.csv")
2    pd["date"] = pd["date"].apply(lambda x: datetime.strptime(x, "%Y-%m-%d"))
3    pd["date"] = pd["date"].dt.date
4
5    large_df = pd[["id", "date", "category", "value"]]
6
7    large_sdf = spark.createDataFrame(large_df)
```

Below the code editor, a log window shows the execution time: "2 min 18 sec - Apache Spark session ready in 2 min 12 sec 518 ms. Command executed in 2 sec 577 ms by Iain Timby on 3:23:22 PM 10/12/23".



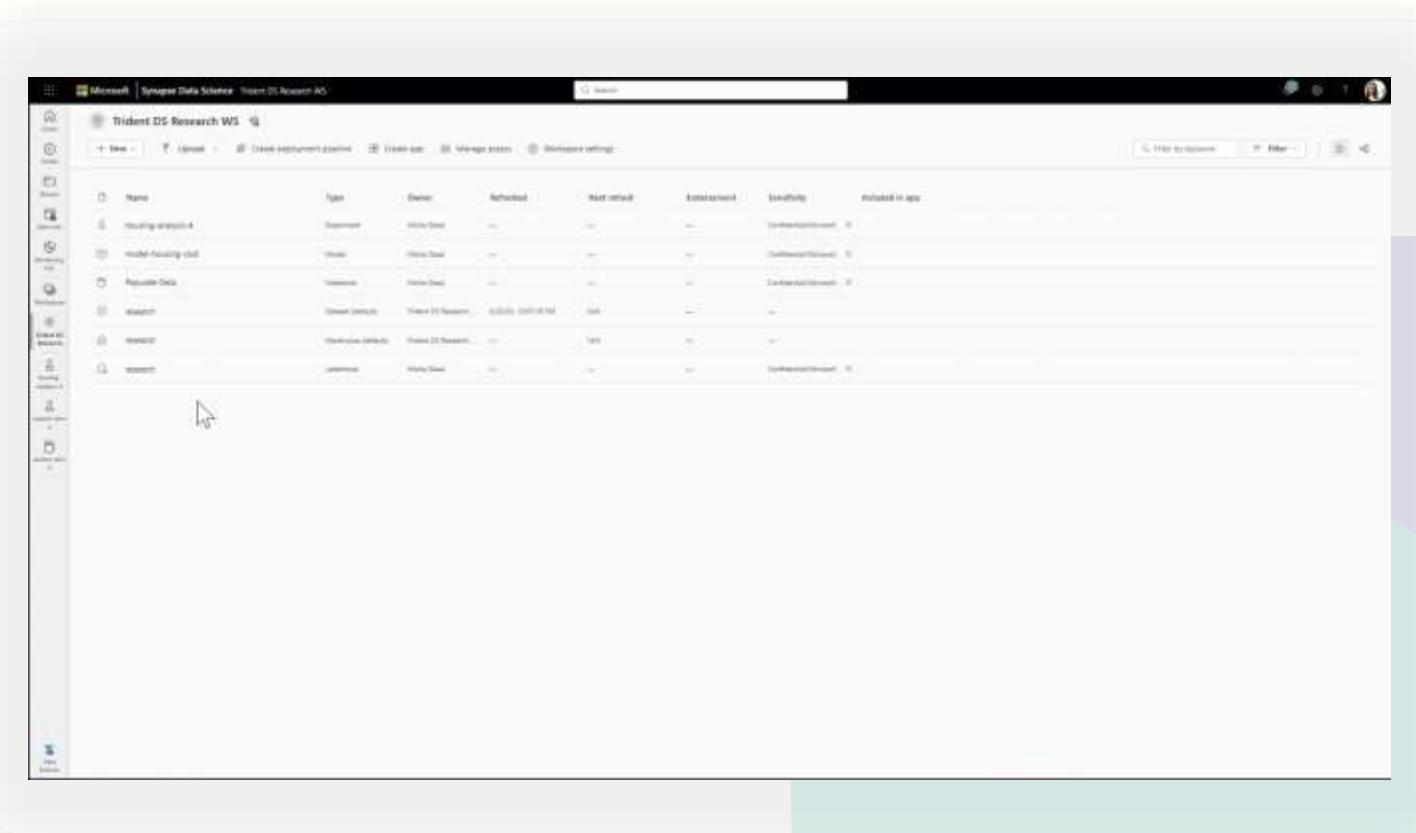
Models and experiments with MLFlow

Built-in model & experiment tracking enables data scientists to track and compare their different experiment runs and model versions

Automatically capture model metrics & parameters with built-in support for MLFlow auto-logging

Users can create and manage model and experiment items in Fabric

MLFlow model registry powered by Azure Machine Learning





Code-first batch scoring with PREDICT

Scalable batch scoring with PREDICT simplifies operationalization of ML models

Supports MIFlow models

Predictions served seamlessly from a Lakehouse to BI reports with zero lead time

The screenshot shows a Microsoft Fabric Data Explorer interface with a Python notebook open. The notebook title is "Step 4. Save Prediction Results" and the section title is "Prediction with Predict Transformer". The code uses the `spark.ml.PredictTransformer` to save predictions to a DataFrame named `test_spark`. Below the code is a table titled "Import results" showing 15 rows of prediction data.

Index	Time	W1	W2	W3	W4	W5	W6
0	0.000000	-0.000127001181817	0.0400000177100001	0.40300003820401	-1.03210476132	0.030001470664102	0.40000007234771
1	0.041516	0.000127000544415	-0.500117740544031	-0.0001770001700000	0.170002310000000	-1.00010000000000	-0.00010000000000
2	0.077420	-0.272504197541111	1.208017757070111	-0.277504197541111	-0.78000030101101	0.471204477700000	-1.07000000000000
3	0.120918	0.0001770001700000	0.111700000000000	1.170000000000000	0.210000000000000	0.122200000000000	0.00000000000000
4	0.164400	0.0001770001700000	-1.210000000000000	-0.0001770001700000	0.210000000000000	-1.02400000000000	0.00000000000000
5	0.210713	0.344001770001700	-1.040000000000000	0.111000000000000	-0.210000000000000	-0.210000000000000	-0.587000000000000
6	0.211463	2.110000000000000	-0.470000000000000	0.100000000000000	-0.220000000000000	-0.000000000000000	-1.00000000000000
7	0.258400	-0.520000000000000	0.000000000000000	0.000000000000000	0.500000000000000	0.500000000000000	0.500000000000000
8	0.309420	-0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000
9	0.360340	0.270000000000000	-0.700000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000
10	0.415268	0.520000000000000	-0.250000000000000	-0.250000000000000	-0.500000000000000	-0.500000000000000	-0.500000000000000
11	0.471320	-0.000000000000000	0.700000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000
12	0.530000	0.000000000000000	0.700000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000
13	0.590000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000
14	0.650000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000
15	0.710000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000	0.000000000000000



Low-code batch scoring with PREDICT

Get started with PREDICT directly from the UI, using a guided experience to deploy models in a matter of clicks

- Select source data
- Map it to your model's inputs
- Choose an output destination

We'll handle the rest and generate code for you!

The screenshot displays the Microsoft Fabric Model Management interface. On the left, a sidebar lists various model versions for 'model-housing-cost'. The main area is titled 'Apply model predictions' and shows a mapping process between 'Input table' columns and 'Model input names'. The 'Input table' is set to 'housing_analytics' and the 'Model version' is 'model-housing-cost, Version 10'. The 'Map input columns' section contains a table with 9 rows, each mapping a source column to a model input name and its data type (double). At the bottom, a preview of the 'Input schema' is shown with columns: Median, HouseAge, AveRooms, AveBedrms, Population, AveOccup, Latitude, and Longitude, all defined as double.

Input column name	Model input name	Model input data type
Median (double)	Median	double
HouseAge (double)	HouseAge	double
AveRooms (double)	AveRooms	double
AveBedrms (double)	AveBedrms	double
Population (double)	Population	double
AveOccup (double)	AveOccup	double
Latitude (double)	Latitude	double
Longitude (double)	Longitude	double



Low-code batch scoring with PREDICT

R language for Apache Spark:

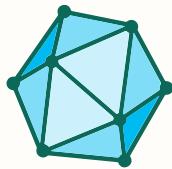
- Native support for SparkR, SparklyrR, Tidyverse, and popular visualization packages
- Key library management capabilities
- Available in both notebook and Spark Job Definitions

Enables data scientists to process data and develop ML models using familiar Spark and R interfaces

The screenshot shows the Microsoft Fabric Data Explorer interface. On the left, there's a sidebar with navigation links like Home, Edit, Run, Data, View, Layout, Standard, Contents, Keyboards, Live workspace, and Comment. Below that is a 'Lakeshouse explorer' section with a tree view of a project named 'Flight delay prediction'. The tree includes nodes for Preprocess, Insert pipeline, Data exploration, Data splitting, Create functions, Fit a model with a recipe, Train the model, Predict results, Evaluate the model, and Build a Power BI report. The main content area has a title 'Flight delay prediction' with a sub-section 'Prerequisites' containing instructions to attach notebooks to a Lakeshouse. Below that is a 'Install package' section with a command-line interface (CLI) window showing the installation of the 'microsoftflights3' package. The CLI output indicates the package was successfully installed in 10 ms. At the bottom, there's a 'Syncronous Editing' status bar.

ML (v.1.0)

Simple and distributed machine learning for Spark



Distributed model training

- LightGBM
- Responsible AI
- Causal Learning
- Explainable boosting machines
- Hundreds of algos!



MLFlow support

- Full integration with MLflow for ML models



Azure AI Services

- Easy to use Integration with Azure AI Services
- Scalable processing with pre-trained AI
- Quickly solve difficult ML tasks



Foundation models

- Integration with Azure Open AI foundation models and embeddings
- Ask questions about your data in English
- Use LLMs to add intelligence to your data

...and much more!



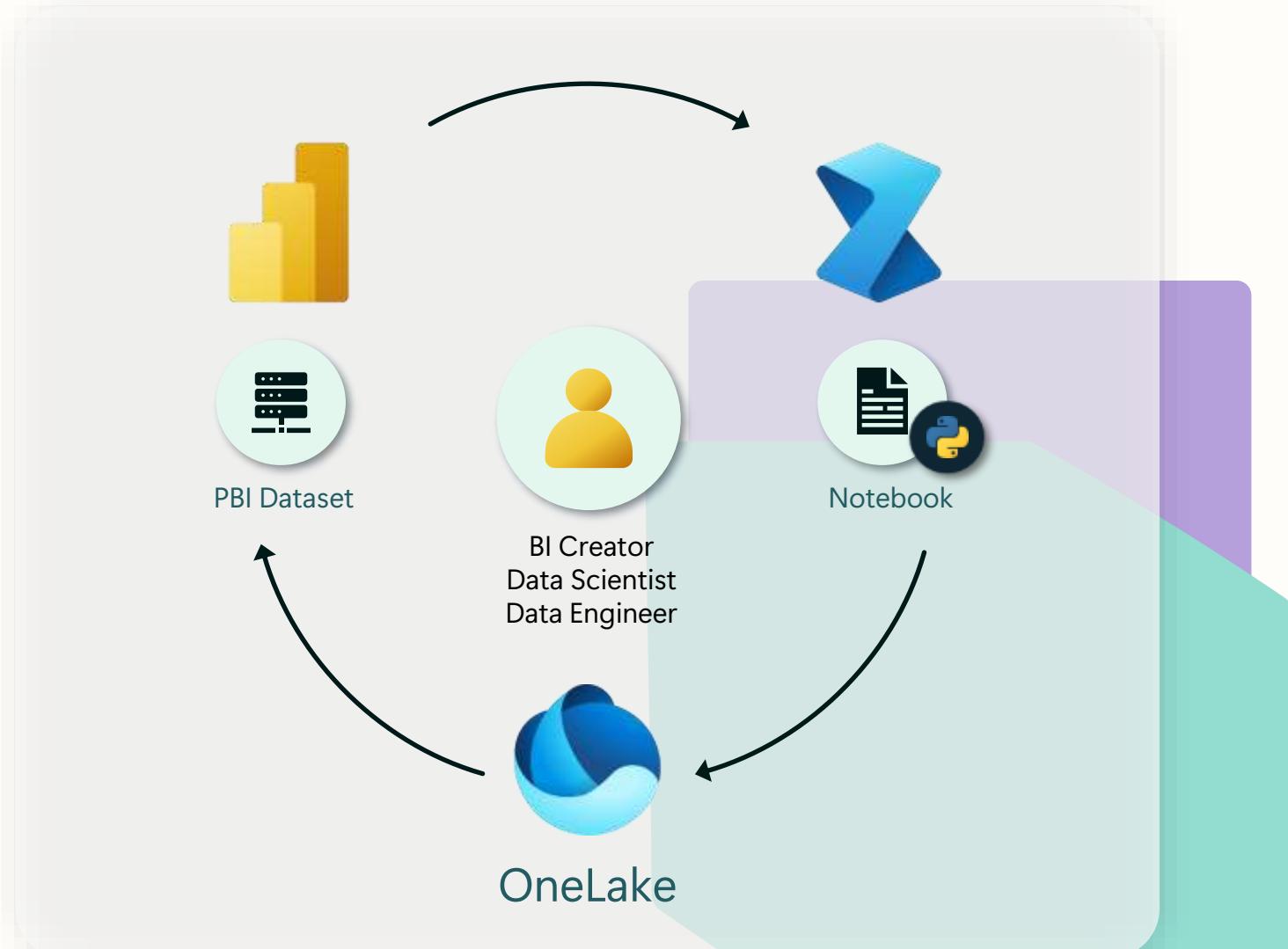
Semantic Link

Bridging Data Science and BI

Semantic Link, Python data access for Power BI

- Evaluate Power BI measures
- Read Power BI tables
- Support for Pandas and Spark
- Validate and test your data

...and propagate metadata to unlock more insights!



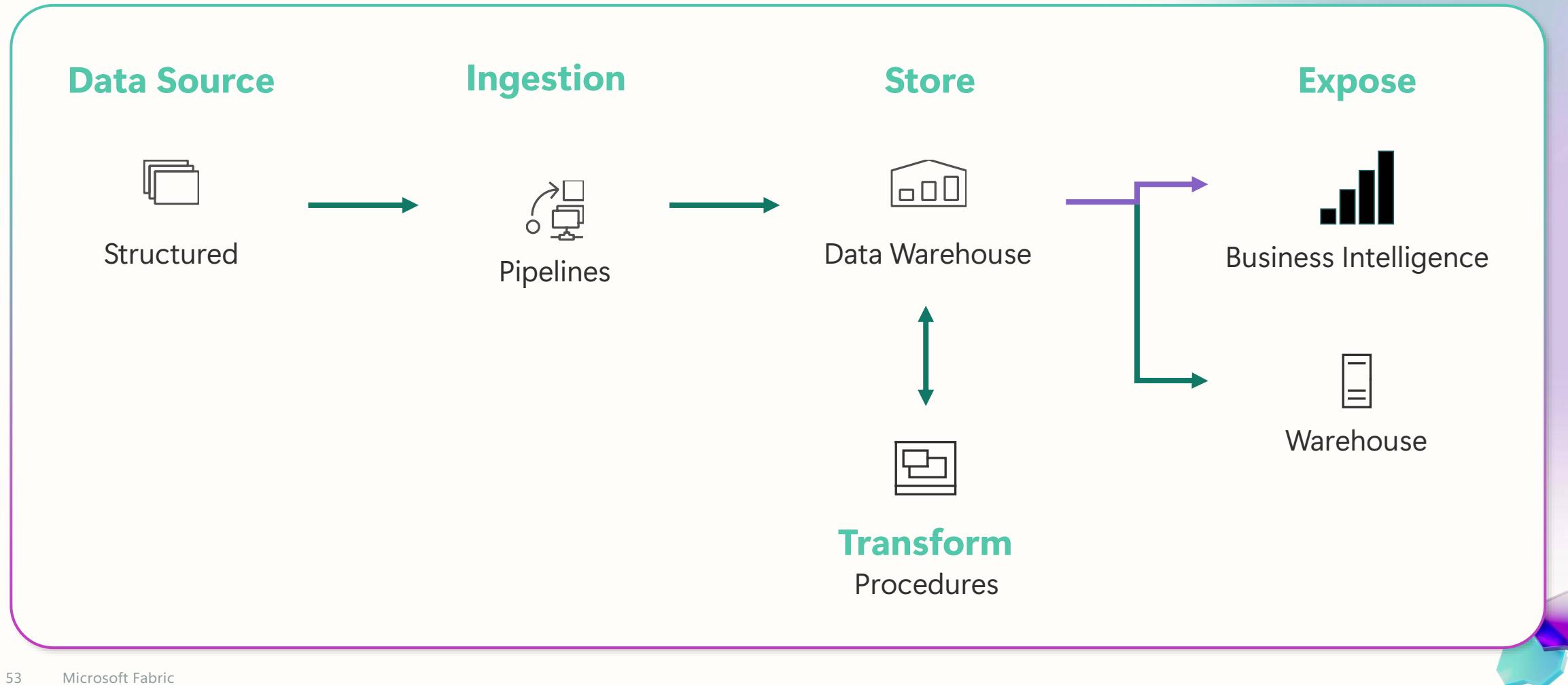
Data Warehouse





Data warehouse scenario

End-to-end analytics scenario





Data Warehouse | Overview

Enterprise scale data warehouse with open standard format

No knobs performance with minimal set-up and deployment, no configuration of compute or storage needed

Key Capabilities:

- Lake-centric warehouse stores data in OneLake in open Delta format with easy data recovery and management
- Use Fabric Mirroring for Zero-ETL integration of data from Azure SQL, Snowflake, or Azure Cosmos DB
- Data loading and transforms at scale, with full multi-table transactional guarantees provided by the SQL engine
- Virtual warehouses with cross-database querying and a fully integrated semantic layer
- Flexibility to build data warehouse or data mesh based on organizational needs and choice of no-code, low-code, or T-SQL for transformations

The screenshot shows the Microsoft Fabric Data Explorer interface. On the left, the Explorer sidebar lists databases (OneLake), schemas, tables, and other objects. A T-SQL query window titled "SQL query 1" is open, displaying the following code:

```
CREATE TABLE [OneLakeDatabase].[dbo].[HealthMetrics]
(
    HealthID NOT NULL,
    HealthCategory VARCHAR(50) NOT NULL,
    HealthSubCategory VARCHAR(50) NOT NULL,
    HealthName VARCHAR(100) NOT NULL,
    TargetValue DECIMAL,
    IntegrationVersion INT,
    DelimitedFromCSV WKS
)
```

The status bar at the bottom indicates "Completed 1 row(s) in 0 ms".



Data loading and transforms at scale

High scale data loading from any source.

Data managed by Warehouse compute engine provides full multi-table transaction guarantees

Key Capabilities:

- Integration with Data Factory Pipelines
- Data published to OneLake as Delta tables
- Support for multi-statement transactions
- Support snapshot isolation
- T-SQL support for:
 - COPY command
 - CTAS
 - INSERT ... SELECT
- Locking:
 - INSERT and SELECT require schema stability lock
 - UPDATE and DELETE do not lock tables

```
COPY INTO table_name.[(Column_list)]
FROM '<external_location>' [...n]
WITH
(
[FILE_TYPE = {'CSV' | 'PARQUET'} ]
[,CREDENTIAL = (AZURE CREDENTIAL) ]
[,ERRORFILE = '[http(s)://storageaccount/container]/errorfile_directory[/]]'
[,ERRORFILE_CREDENTIAL = (AZURE CREDENTIAL) ]
[,MAXERRORS = max_errors ]
[,COMPRESSION = { 'Gzip' | 'Snappy'}]
[,FIELDQUOTE = 'string_delimiter']
[,FIELDTERMINATOR = 'field_terminator']
[,ROWTERMINATOR = 'row_terminator']
[,FIRSTROW = first_row]
[,ENCODING = {'UTF8'|'UTF16'}]
);

CREATE TABLE [warehouseA].[dbo].[FactInternetSales_new]
AS
SELECT * FROM [warehouseB].[dbo].[FactInternetSales];

INSERT INTO dbo.EmployeeSales
SELECT sp.BusinessEntityID, c.LastName, sp.SalesYTD
FROM warehouseA.Sales.SalesPerson AS sp
INNER JOIN warehouseB.Person.Person AS c
    ON sp.BusinessEntityID = c.BusinessEntityID
WHERE sp.BusinessEntityID LIKE '2%'
ORDER BY sp.BusinessEntityID, c.LastName;
```



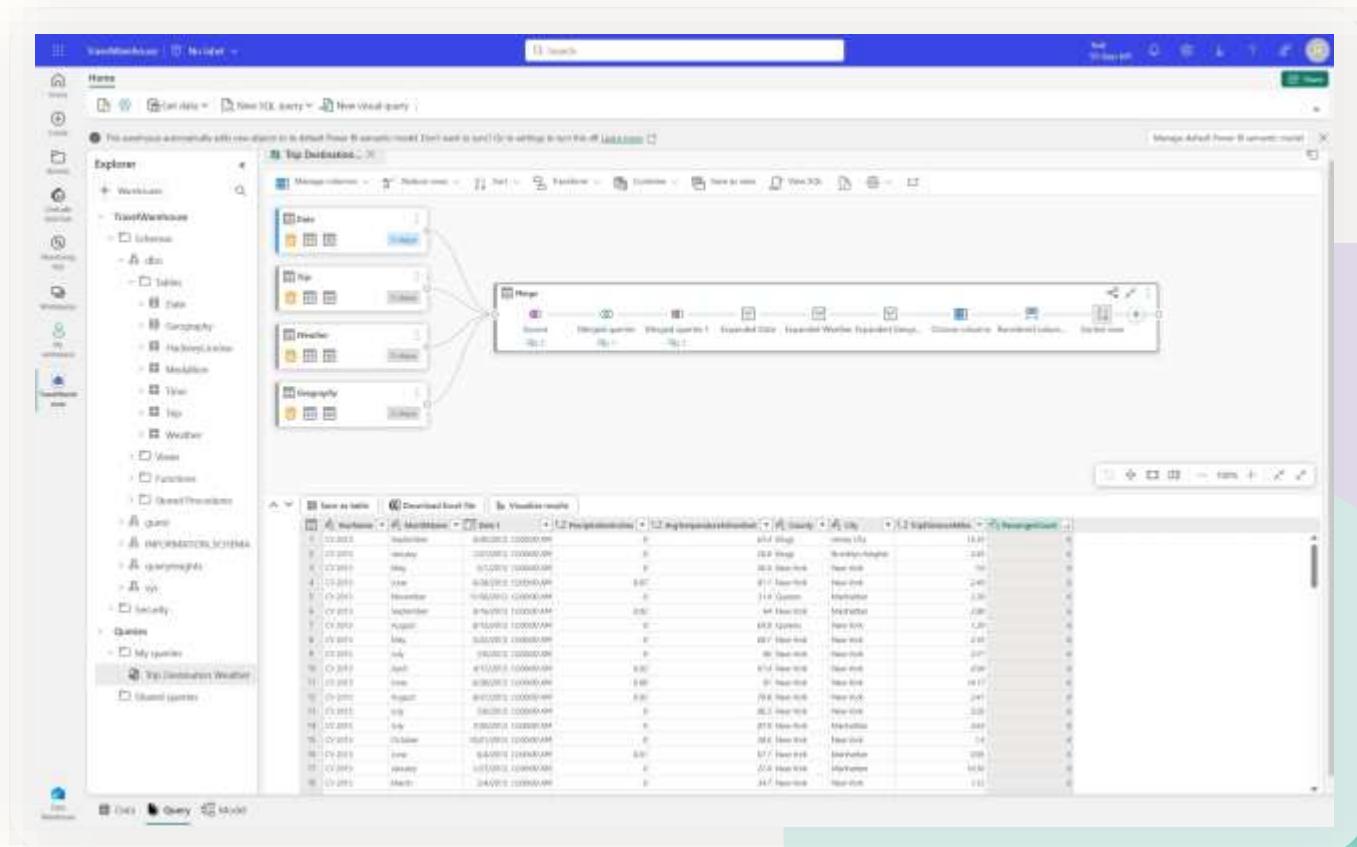
Virtual warehouses and cross database querying

Eliminate the need to duplicate data from any Lakehouse to any Warehouse

Customers can query data in the lake without the need for data movement, reducing costs

Key Capabilities:

- Build virtual warehouses by creating Lakehouses with shortcuts to data in the Lake
- Query across Warehouse and Lakehouse SQL Endpoint with zero data movement
- Build queries visually using the Visual Query Editor





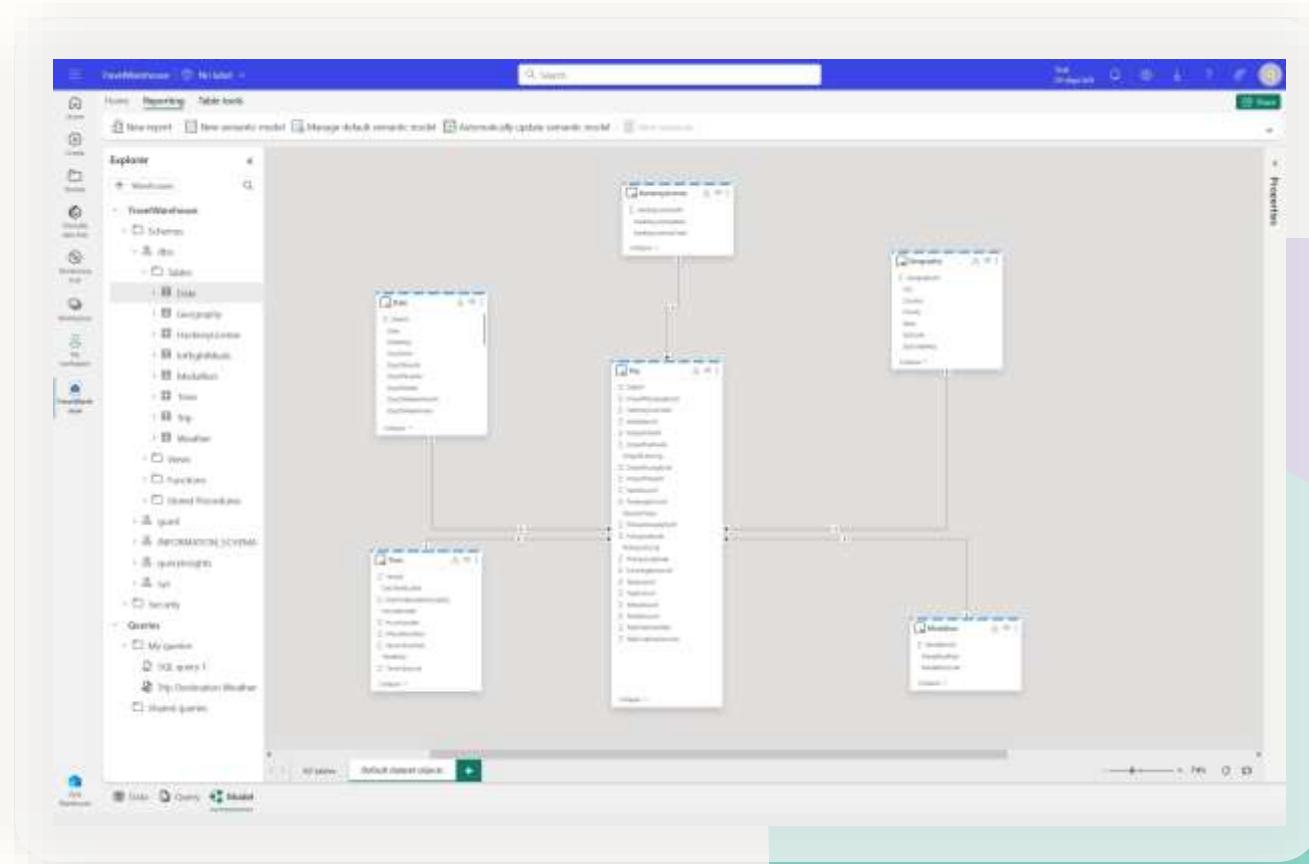
Fully integrated Power BI semantic layer

Reduce integration and gain insight from your data in seconds

Built-in Power BI enables everyone to visualize their data in seconds.

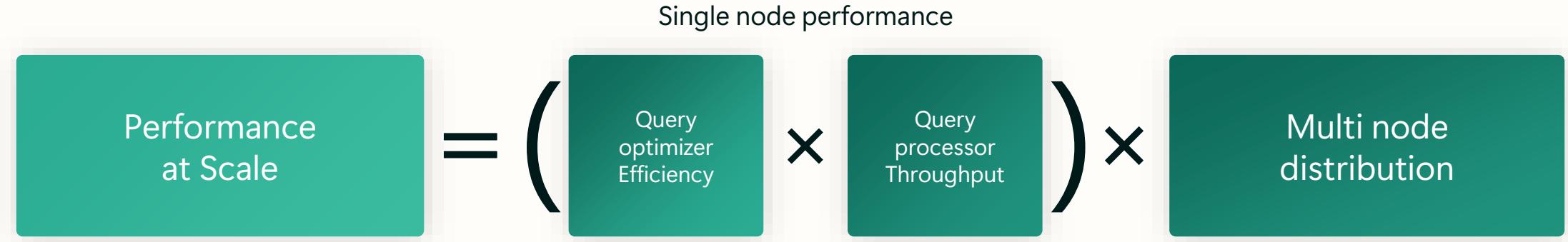
Key Capabilities:

- Auto-generated semantic models always in sync
- Default dataset in Direct Lake mode but automatically switches to Direct Query or Import mode as security or performance needs change
- Flexibility to add/remove tables to dataset
- Full web authoring experience for creating measures





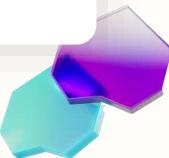
No Knobs Performance at scale with open data format



SQL Server QO The best query optimizer in the industry

VertiPaq The fastest columnar query processor in the world (originated from Power BI)

Polaris The most scalable distributed query processor in the world with the only published Petabyte scale benchmark (VLDB 13, August 2020)





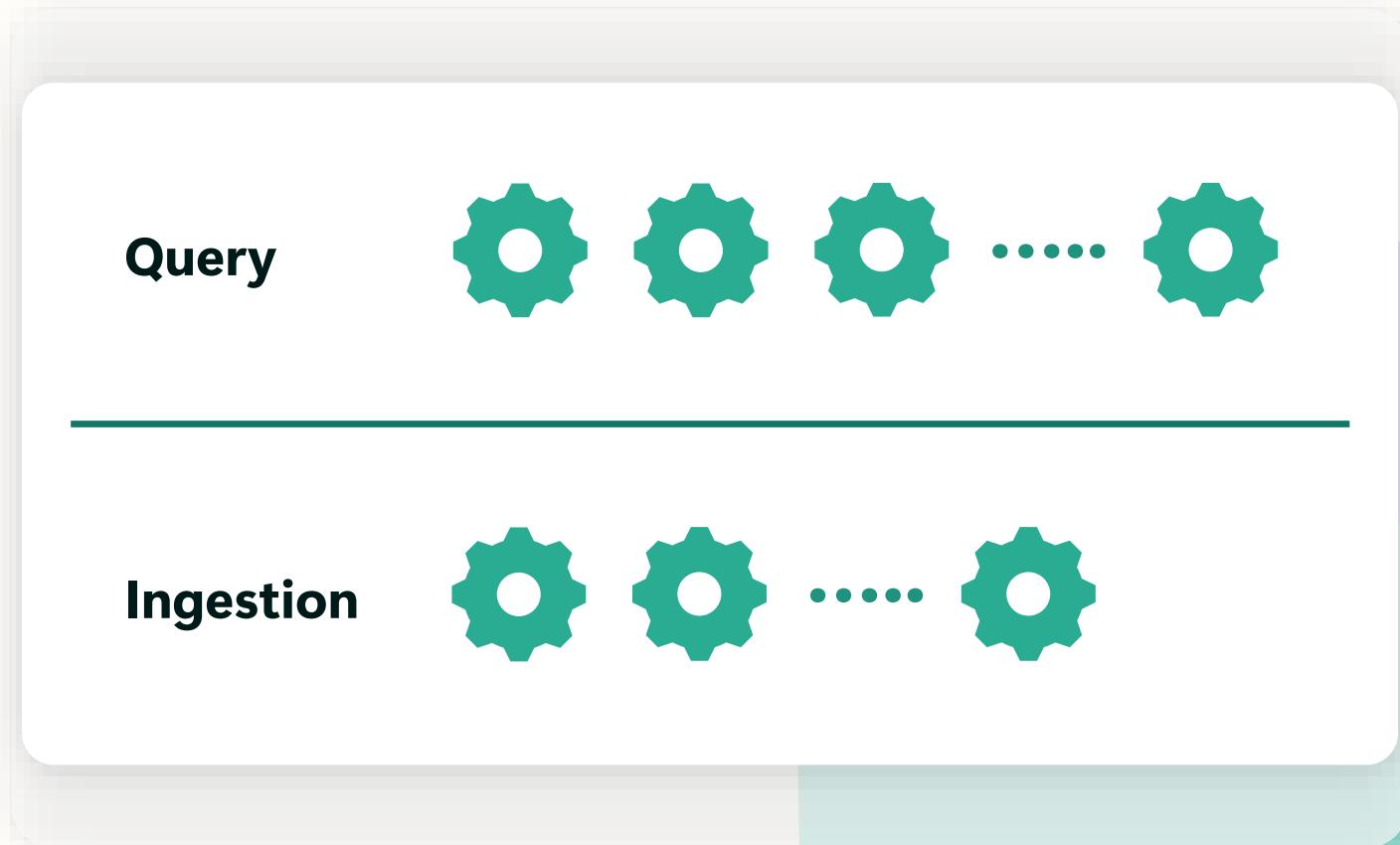
Autonomous workload management

Eliminate need to manually manage different types of workloads

Built-in autonomous workload management, provides workload performance isolation and scales based on demand without any setup

Key Capabilities:

- Workloads automatically detected and isolated based on SQL statement types
- Isolated compute used for loading data, providing predictable performance for ETL
- Distributed query processor auto-scales as demand increases





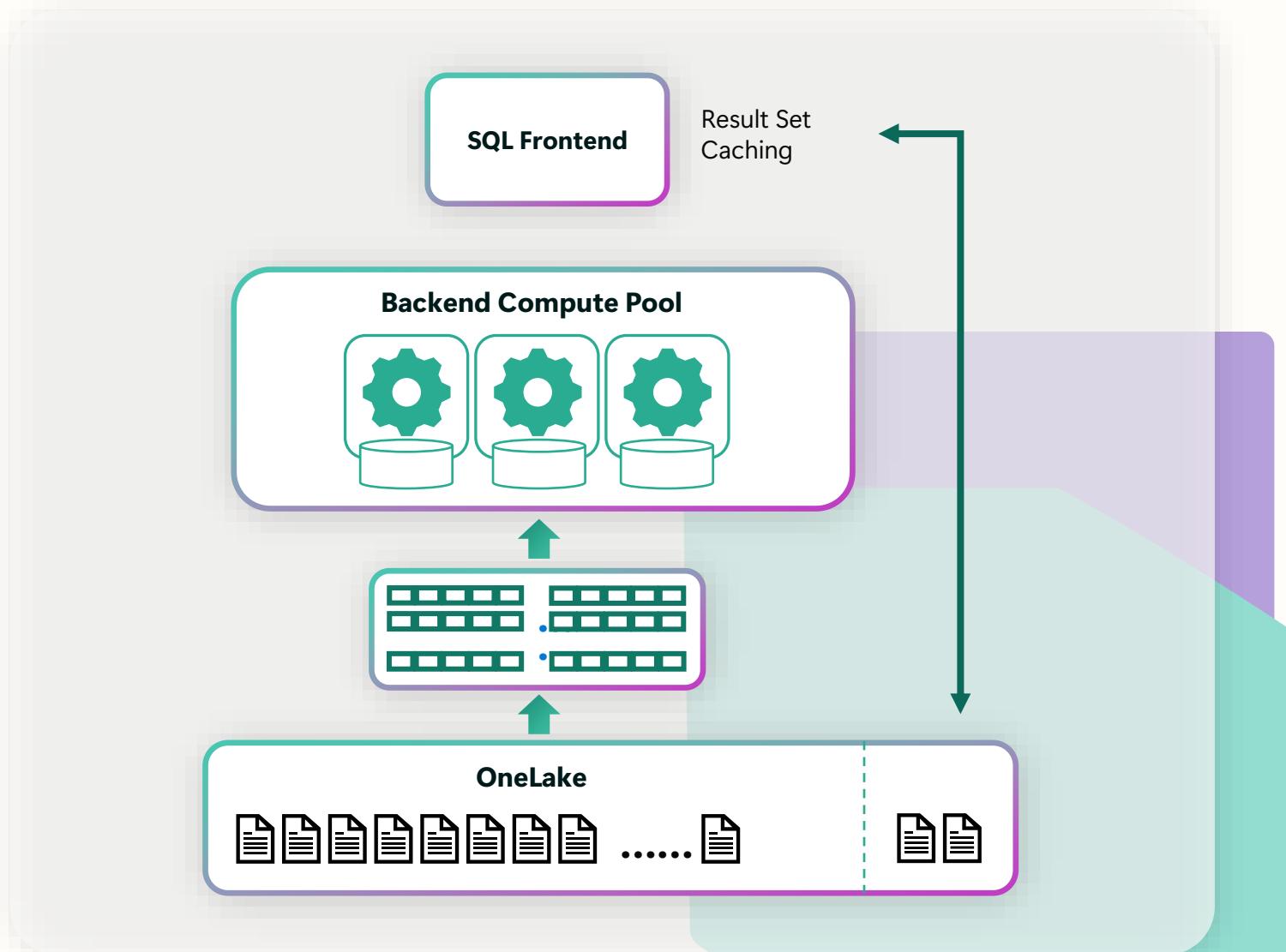
Automatic multi-tiered caching

Get best performance out of the box

Automatic caching at multiple tiers provides top performance for queries out-of-the-box and reduces management and tuning overhead

Key Capabilities:

- Memory caching offers the lowest latency and is designed to store the most frequently accessed data
- SSD caching reduces latency in fetching from OneLake
- Result set caching provides instant performance for repetitive queries





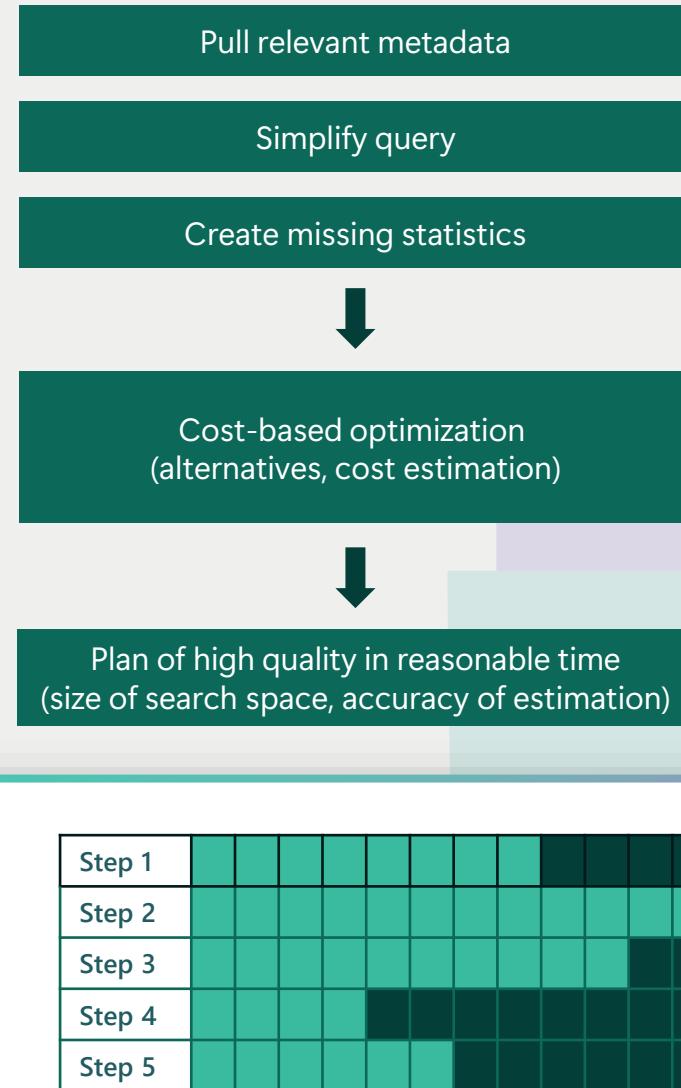
Unified query optimization

Get best performance out-of-the-box at lowest cost

Query plan provides top performance for queries out-of-the-box at cheapest cost

Key Capabilities:

- Query plan with optimal # of nodes for each step with efficient resource consumption automatically created
- Ability to create user-defined statistics manually
- Auto-create statistics ensures optimal query performance by creating statistics as needed by the query optimizer
- Auto-update statistics feature (pending) ensures optimal query performance maintenance free





Monitoring through dynamics management views

Get insights into SQL query lifecycle

Customers can view query performance and easily terminate any noisy or costly queries

Key Capabilities:

- Use Dynamic Management Views (DMVs) to:
 - Find open connections through `Sys.dm_exec_connections`
 - Find users through `Sys.dm_exec_sessions`
 - Fine queries executed by user through `Sys.dm_exec_requests`
- Use KILL command to terminate session

-- Identifies long-running queries in order of which query has taken the longest since it has arrived, and which user executed the request.

```
SELECT r.request_id, r.session_id,  
r.start_time, r.total_elapsed_time,  
s.login_name  
FROM sys.dm_exec_requests AS r  
JOIN sys.dm_exec_sessions AS s  
ON s.session_id = r.session_id  
WHERE r.status = 'running'  
ORDER BY r.total_elapsed_time DESC
```

-- Use the KILL command on the session with the long-running query.

```
KILL '[SESSION_ID WITH LONG-RUNNING  
QUERY]'
```



Secure by default

Keep data secure for any role accessing it and ensure peace of mind

Customers can secure their data using familiar constructs and ensure data is only visible to those authorized to do so

Key Capabilities:

- Workspace roles:** Workspace roles are used for collaboration with team. Add users to workspace with role assignment of Admin, Member, Contributor, Viewer
- Artifact permissions:** Artifact permissions are used for sharing for consumption of Warehouses. Provide access and share individual Warehouses with specific permissions
- Data security:** Use T-SQL, GRANT, REVOKE or DENY to secure any object within Warehouse. Users can be assigned to built-in custom roles.
- Sensitivity labels:** Apply sensitivity labels on your Warehouse to classify sensitive data.
- Granular security:** Implement row or column security or Dynamic Data Masking for granular data access control

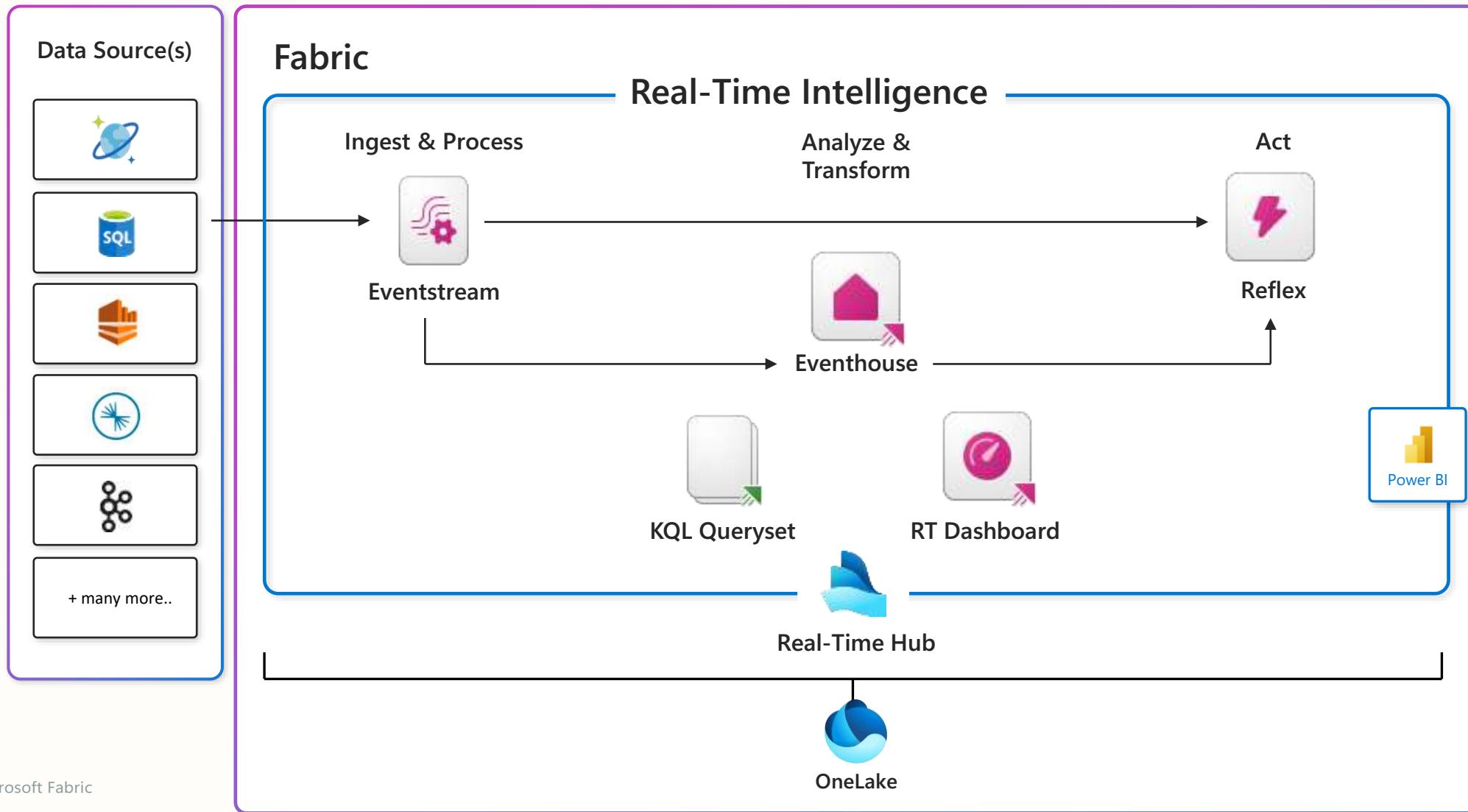
The screenshot shows the Microsoft Fabric interface with a focus on data security and sharing. On the left, there's a sidebar with navigation icons like Home, Search, and Import. The main area has a title bar 'Imported' and 'General'. Below the title bar, there's a section for 'Detail for Imported' with tabs for 'Unlocked' (selected), 'Released', and 'Security' (set to 'General'). To the right, there are two cards: 'Visualize this data' (Create an interactive report, or a table to discover and share business insights. Learn more) and 'Share this data' (Give people access to the warehouse and set their permissions to work with it. Learn more). Below these cards is a table titled 'See what already exists' with columns: Name, Type, Relation, Location, Refreshed, Enhanced, and Security. One row is shown: 'Imported' (Type: Dataset (built-in), Relation: Dimension, Location: Inserters, Refreshed: 5/12/2024 11:10:46 AM, Enhanced: --, Security: General). At the bottom, there's a 'Data Engineering' logo.

Real Time Intelligence



Real-Time Intelligence scenario

End-to-end analytics scenario



Ingest & process all data sources

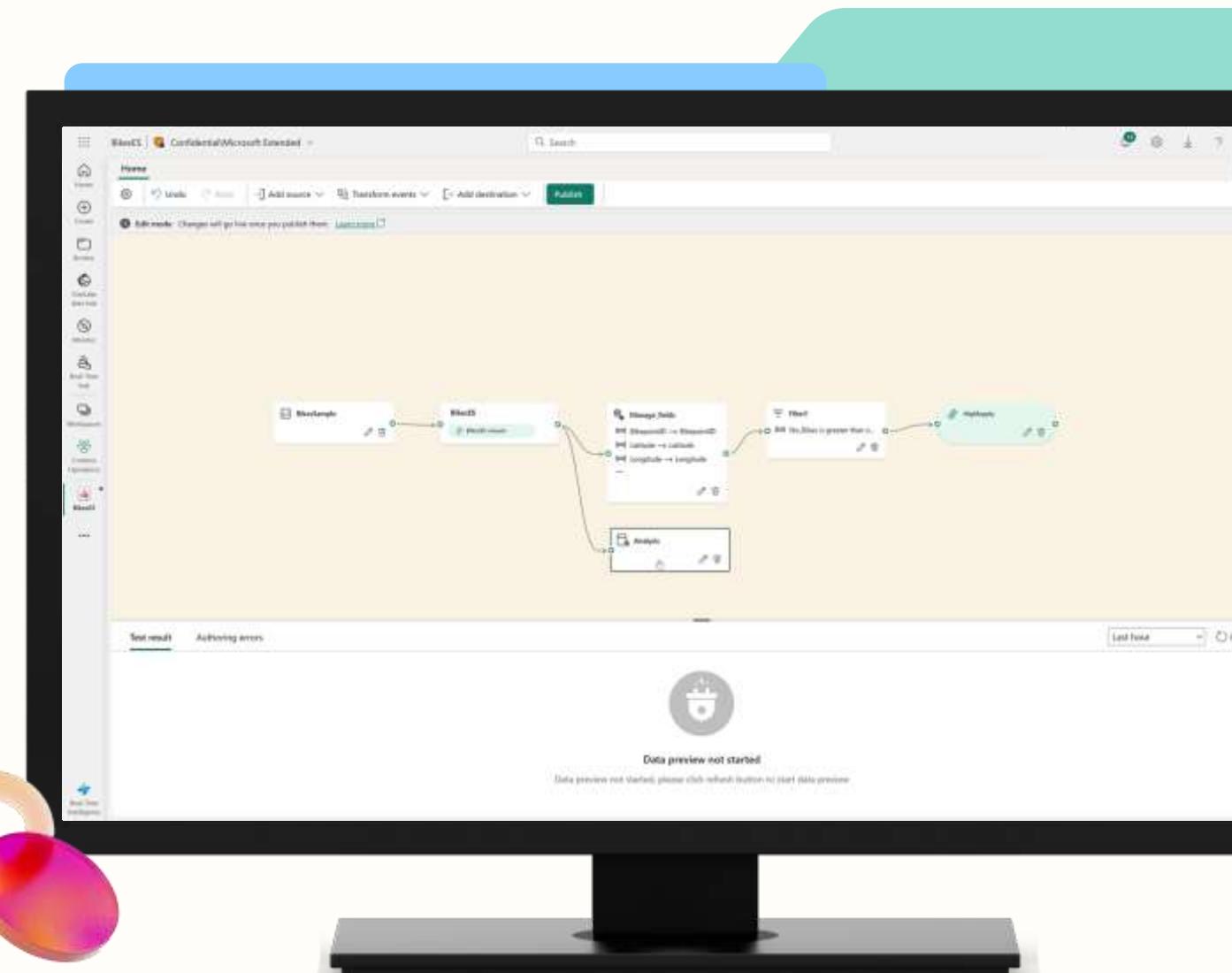
Ingest data from any source, in any data format

Capture, transform and route event data without writing any code

Access out-of-the-box connectors for streaming and event data sources

Process data and enrich real-time events

Route events to other Fabric and 3rd party entities



Analyze data event streams

Manage an unlimited amount of data, from gigabytes to petabytes, with unlimited scale

Use event houses to handle and analyze large volumes of real-time data streams

Monitor and manage multiple databases at once

Create KQL databases and querysets to run, view, and customize queries directly on raw data

Enhance organizational BI reports with enriched data



The screenshot displays the Microsoft Fabric Data Explorer interface. On the left, a sidebar shows a navigation tree with 'Manufacturing_kdb' selected, under 'KQL databases'. Below it are sections for 'Tables', 'DeliveryStreams', 'Metrics', and 'Data marts'. The main area shows 'Database details' for 'Manufacturing_kdb', including 'Created by', 'Region', 'Created on', and 'Last ingestion'. A 'Top tables' section lists tables like 'Deliveries', 'DeliveryStreams', 'Mile', 'DeliveryZones', and 'Bikes' with their sizes. Below that is a 'Recently updated batches' section with entries for 'Deliveries' and 'Deliveries'. To the right, a 'Explore your data' panel contains a KQL query editor with code snippets for 'YIELD_TABLE_4000', 'YIELD_TABLE_NINE', and 'take'. At the bottom, a table titled 'Bikes' is shown with columns: BikepointID, Street, Neighborhood, Latitude, Longitude, No_Bikes, and No_Visiting_Cycles. The table contains 100 records.

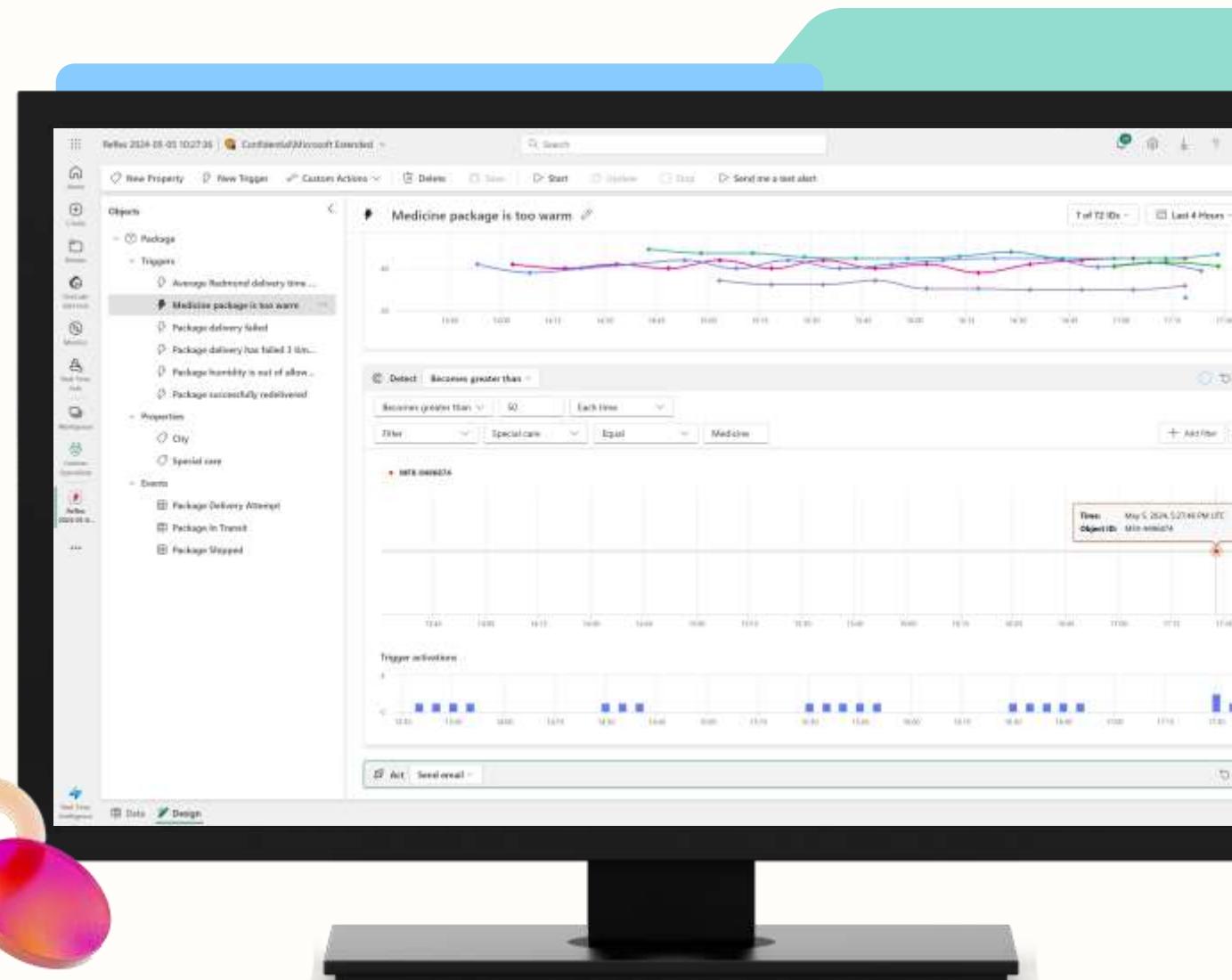
BikepointID	Street	Neighborhood	Latitude	Longitude	No_Bikes	No_Visiting_Cycles
BluePoint_101	Woodland Street	Mile End	51.520962	-0.120016	24	8
BluePoint_454	Concord Grove	Wentworth Road	51.517253	-0.131142	7	12
BluePoint_489	Bonhill Road	Mile End	51.518192	-0.093191	29	0
BluePoint_652	Upple Grove	Wentworth Road	51.472999	-0.133972	17	8
BluePoint_491	Queen Mary's	Mile End	51.522506	-0.091358	7	38
BluePoint_493	Maple Wood	Mile End	51.5236	-0.023287	13	29
BluePoint_714	Wenlock's Road	Wentworth Road	51.473178	-0.132729	15	12
BluePoint_723	Chelton Road	Mile End	51.5256	-0.090077	24	12
BluePoint_731	Tung Tolly Bridge	Mile End	51.530318	-0.042744	9	21
BluePoint_738	Arroll Road	Mile End	51.528208	-0.037491	7	23
BluePoint_750	Warford Wood	Mile End	51.521645	-0.092684	4	18
BluePoint_772	Mile End Station	Mile End	51.5184	-0.034881	10	9
BluePoint_781	Mile End Park Leisure Centre	Mile End	51.520956	-0.032569	24	23

Act quickly on top of data

Automatically take actions when patterns are detected in changing data

Drive actions on a per instance state that evolves over time

Act on data without needing a deep schema and semantic modeling



Access Real-Time Hub

Find and consume event sources in Real-Time Hub, a single catalog of all streaming sources

Use Get Events and all streaming connectors to easily bring data in

Subscribe to internal and external discrete events

Quickly connect experiences for Microsoft streaming sources like IoT Hub

Access system events emitted by Fabric and Azure storage

Ingest data streams from all clouds (e.g. AWS, Kinesis, Google Pub-Sub, etc.)



Name	Type	Owner	Location	Endorsement	Sensitivity
CustomCDC-goods-stream	Custom	Tessa Kloster (PALMER)	Contoso Operations	—	Microsoft Extended
new_event_stream-stream	Custom	Chandi Rupasinghe	SQL DB Native Bug (Fabric)	—	Microsoft Extended
CustomCDC-stream	Custom	Tessa Kloster (PALMER)	Contoso Operations	—	Microsoft Extended
CustomAppStream-stream	Custom	Shiv Kumar	DreamDemoCopilot	—	Microsoft Extended
ThermistorStream-stream	Custom	Shiv Kumar	DreamDemoCopilot	—	Microsoft Extended
Therminal	Custom	—	DreamDemoCopilot	—	Microsoft Extended
TherminalData	Custom	—	DreamDemoCopilot	—	Microsoft Extended
Sensors-stream	Custom	Tessa Kloster (PALMER)	Contoso Operations	—	Microsoft Extended
SubC_event_stream-stream	Custom	Rishabh Johnson	Power BI Admin and Governance (Fabric)	—	Microsoft Extended
CustomerSensors-stream	Custom	Tessa Kloster (PALMER)	Contoso Operations	—	Microsoft Extended
SQLCDC-customer-stream	Custom	Tessa Kloster (PALMER)	Contoso Operations	—	Microsoft Extended
Shipments-stream	Custom	Tessa Kloster (PALMER)	Contoso Operations	Permuted	Microsoft Extended
HighDemand	Custom	Tessa Kloster (PALMER)	Contoso Operations	Permuted	Microsoft Extended
HighSupply	Custom	Tessa Kloster (PALMER)	Contoso Operations	Permuted	Microsoft Extended
LowSupply	Custom	Tessa Kloster (PALMER)	Contoso Operations	Permuted	Microsoft Extended
OneRiverEvent-stream	Custom	Bert Cottier	Power BI Admin and Governance (Fabric)	—	Microsoft Extended
Salic_event_stream-stream	Custom	Bert Cottier	Power BI Admin and Governance	—	Microsoft Extended

Work across full event life cycle

Gain deeper insight into streams with data preview and usage history

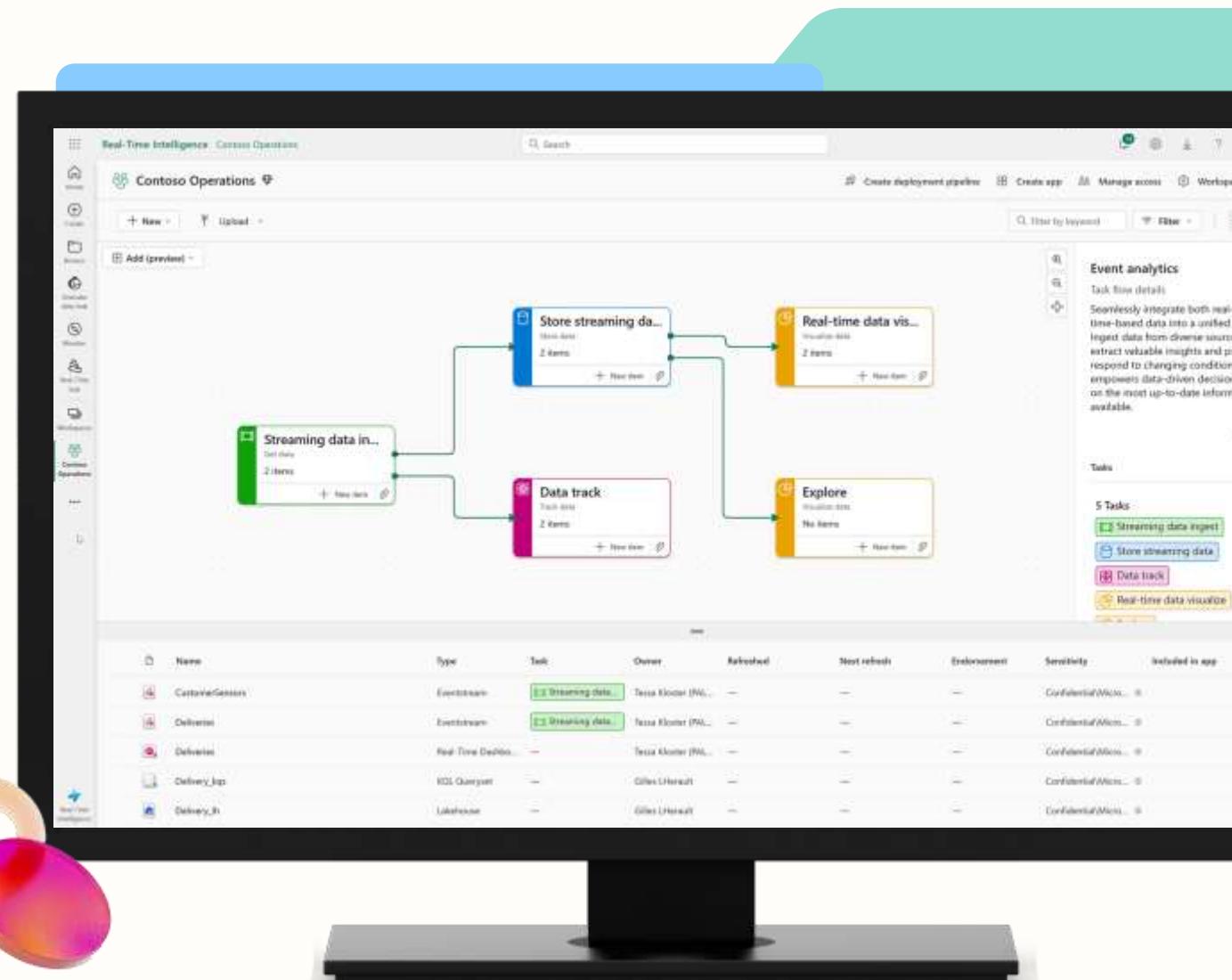
Create triggers through simple, embedded experiences

Open event streams to process and route events without writing any code

Land in Eventhouse for further analysis

Analyze and build lightweight models

Create derived and enriched streams



Consume data from anywhere

Find time-oriented data-in-motion readily available in minutes for use in OneLake

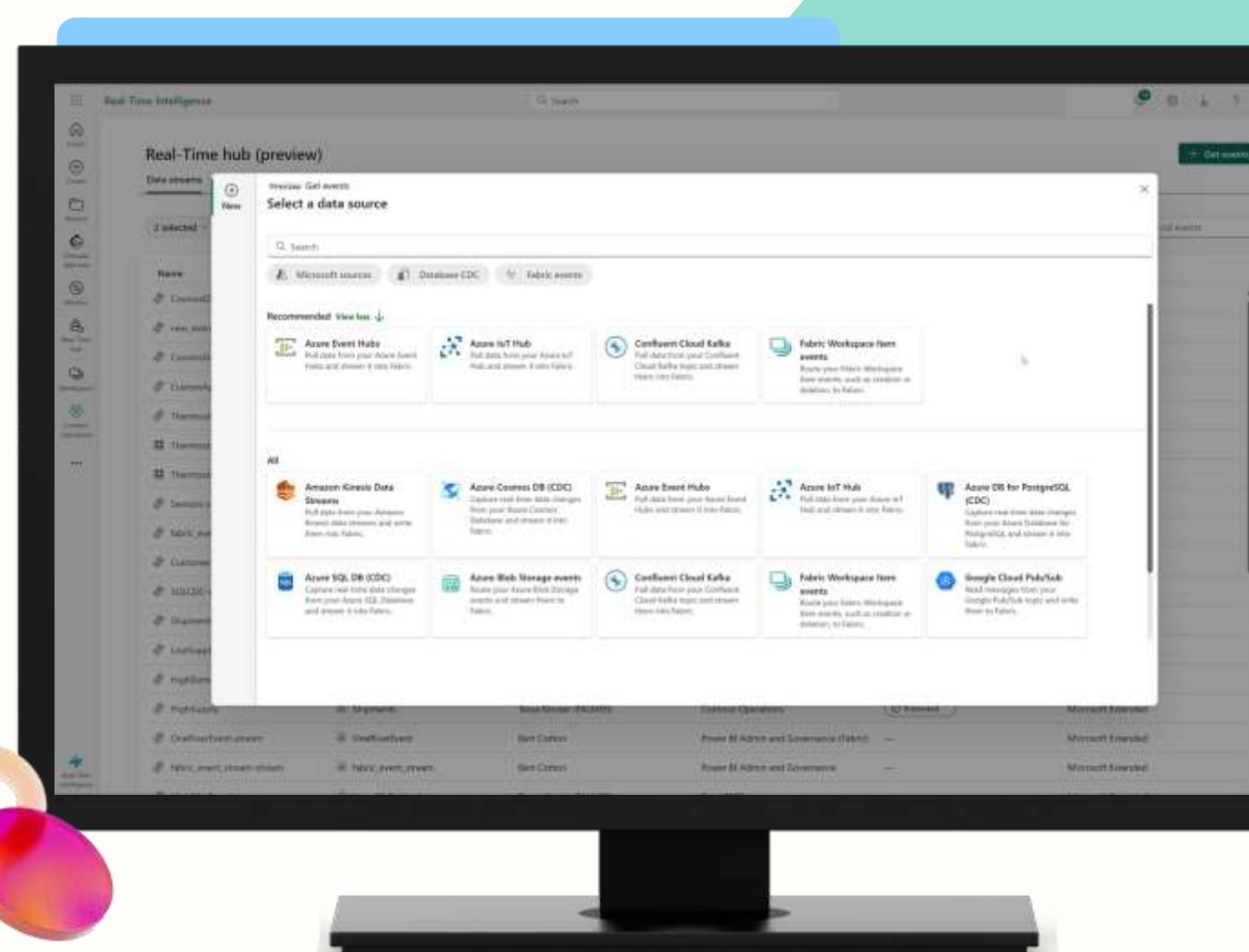
Discover events and seamlessly consume them from across organization

Simplify integration of stream processing frameworks

Expose and use well-established open-source APIs, standards, protocols and connectors

Maintain data ownership – data is not trapped in Microsoft's proprietary formats

Integrate seamlessly with other experiences in Microsoft Fabric



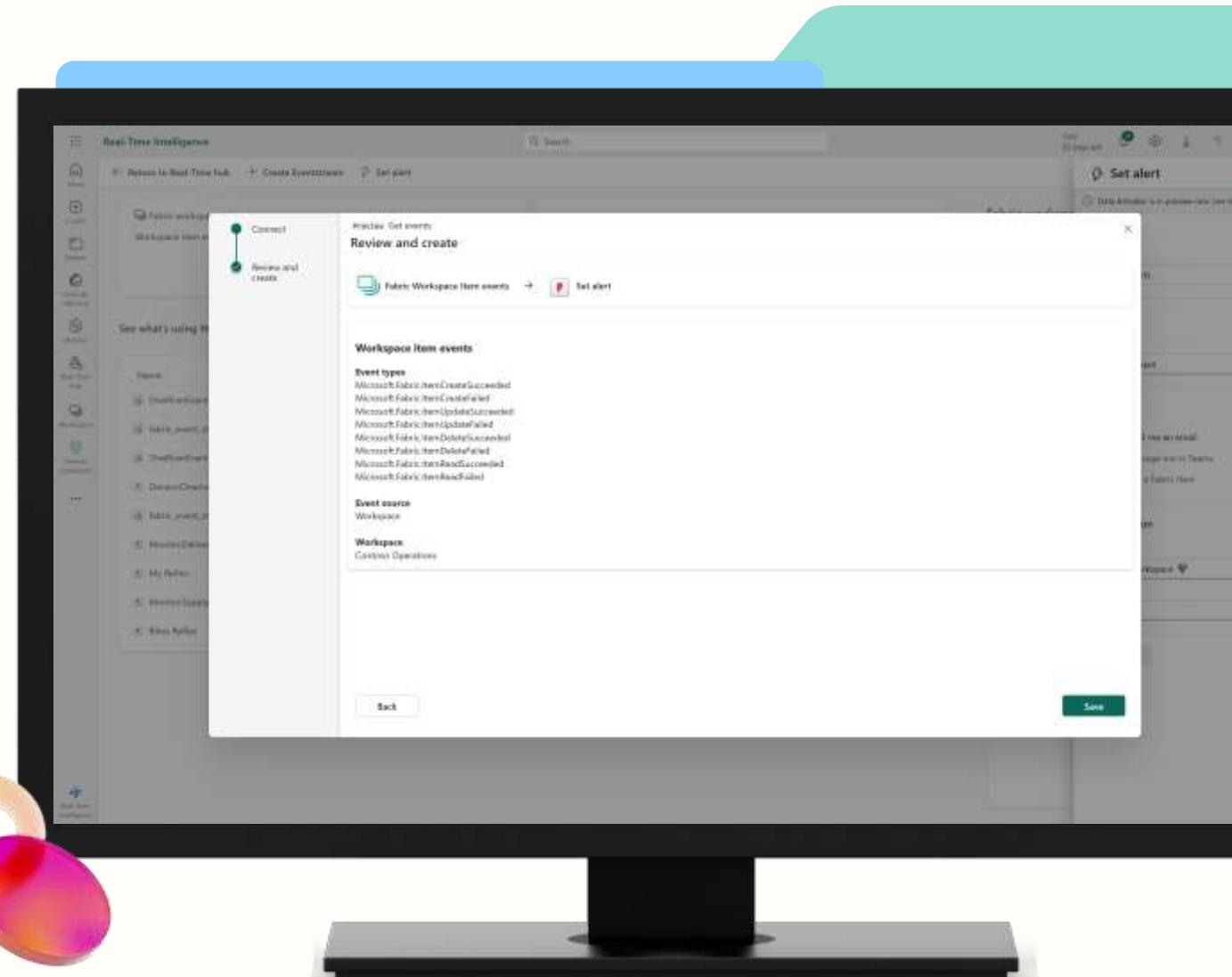
Rapid solution development – no-, low-, and pro-code experiences

Access wide range of experiences to quickly build production solutions

Leverage custom endpoints and APIs to integrate directly into existing systems

Empower developers with CI/CD tools within their workflows

Gain proven performance at scale across ingestion, querying and storage



Real-time AI insights – and more to come!

Leverage Copilot to automate routine tasks and act as an interactive aide

Easily generate KQL queries with copilot experience

Monitor data and automatically drive alerts when anomalies are detected

Use real-time insights to build and scale advanced ML models in Azure Machine Learning

Create generative AI experiences on top of your time-oriented data with Azure AI Studio

The screenshot displays the Microsoft Fabric Analytics interface. On the left, a sidebar lists various data sources: Functions, Bikes, Deliveries, DeliveryDrivers, DeliveryZones, Inventory, LoyaltyCustomer, NewTrading, SpecializedDelivery, and Trucks. The main area shows a KQL query being generated by Copilot:

```
26 | summarize totalBikes+sum(No_Bikes) by Neighborhood
27 // what is the total number of bikes by neighborhood and street
28 Bikes
29 | summarize totalBikes+sum(No_Bikes) by Neighborhood, Street
30
31 // what is the total number of bikes and total number of docks by neighborhood
32
33 Bikes
34 | summarize totalBikes+sum(No_Bikes), totalDocks+sum(No_Dock) by Neighborhood
35
36 Docks
37 | summarize totalDocks+sum(No_Docks) by Neighborhood
```

Below the query, a table titled "Table 1" shows data for neighborhoods, including Neighborhood, totalBikes, and totalDocks. The table includes 19 records from neighborhoods like Baltimore, Chelsea, Ward, Hill Top, Riverdale, Kingsbridge, Yonkers, West Chelsea, St. John's Wood, Bronx, Garfield, Mid Isl, and Bronx.

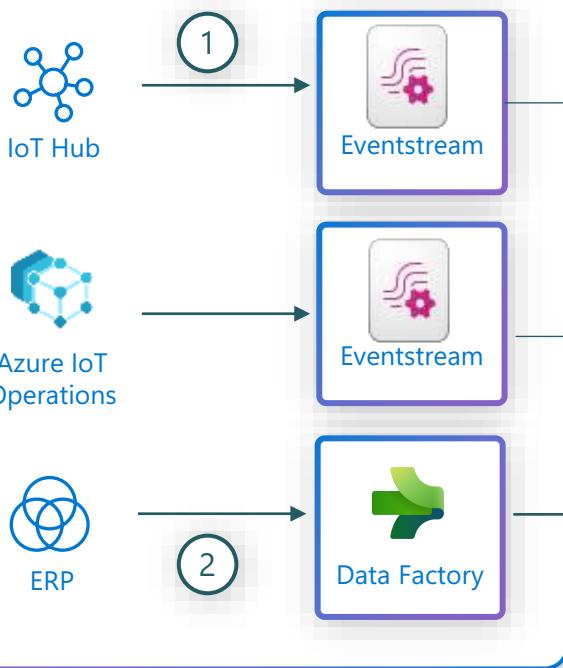
On the right, the Copilot pane asks, "What do you want to know about the data?" and provides two previous questions: "what is the total number of bikes and total number of docks by neighborhood" and "what is the total number of bikes by neighborhood".



An end-to-end Real-Time Intelligence experience – Connected Factory

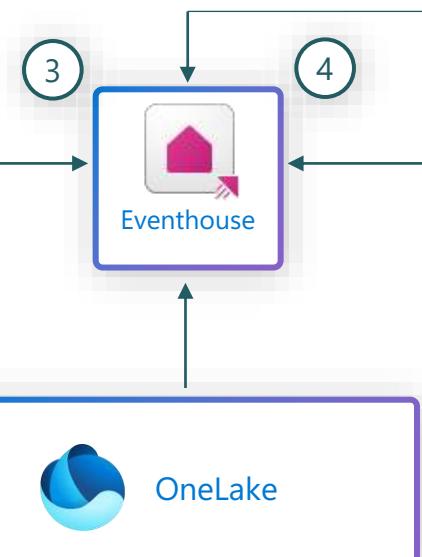
Ingest & process

1. Over 1M/1 HR IIoT events, 30k tags, from 40 factories are streamed in real time with sub-second latency
2. Contextualization data (OPC asset metadata, shift details, weather, component cost, etc.) mirrored into Fabric OneLake



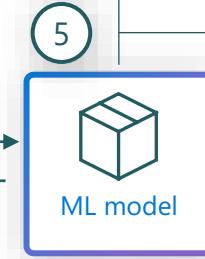
Analyze & transform

3. IIoT events are enriched in-motion with the Contextualization data and asset hierarchy from OneLake, providing clean and meaningful views
4. Enriched data is aggregated on the fly for long-term view



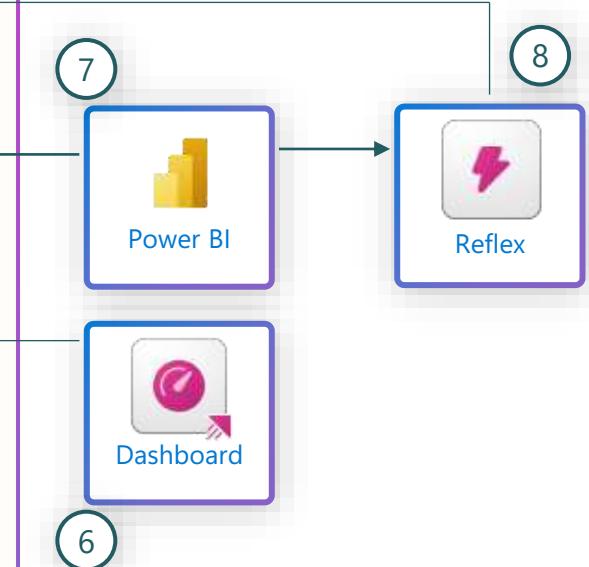
Train

5. Build, train, and score ML models in real time, persisting the scored data in OneLake



Visualize and activate

6. Real time dashboard of all factories with ability to easily shift from low granularity to high granularity view
7. Enhance organizational BI reports with high granularity enriched data
8. Notify users and activate automation in response to live indications from the factory floor

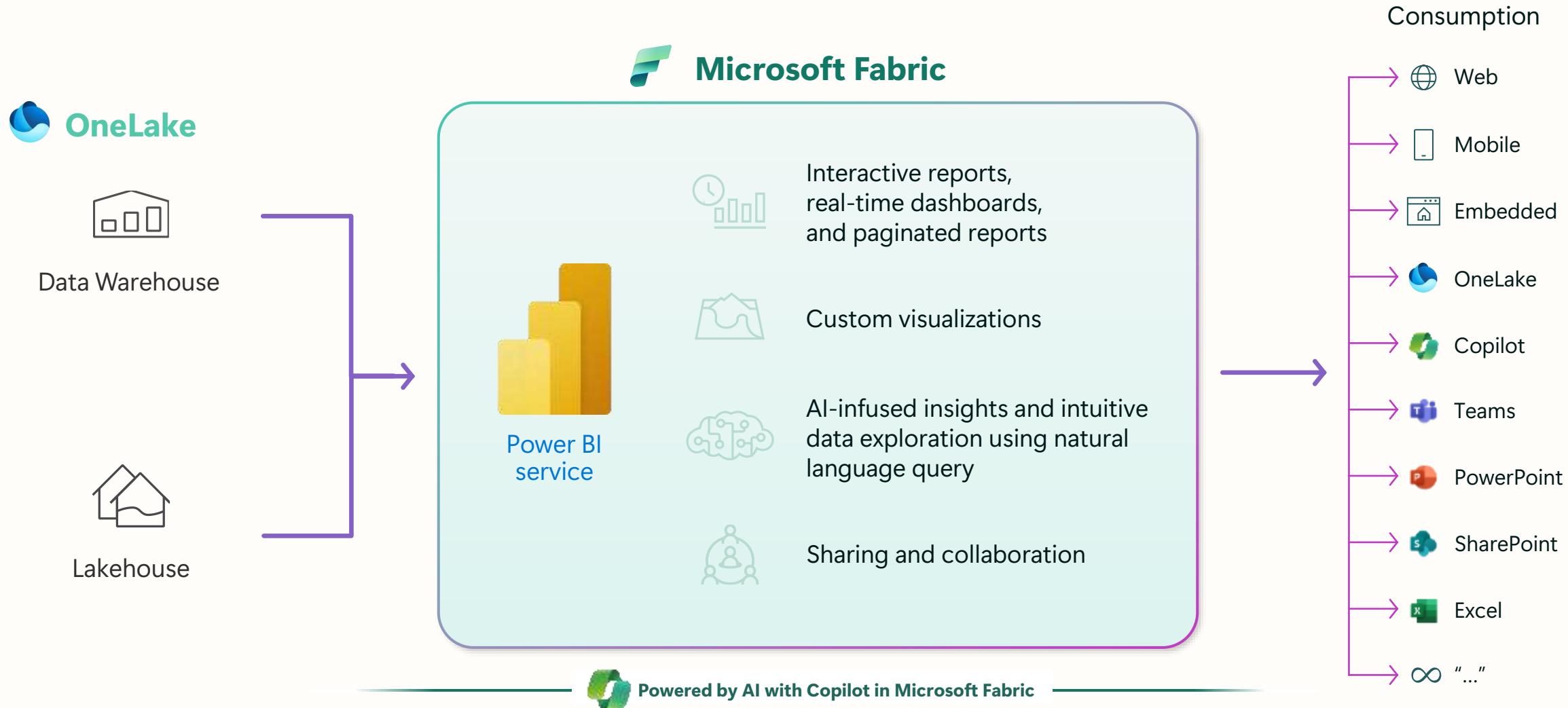


Power BI





Power BI: The bridge between data and decisions



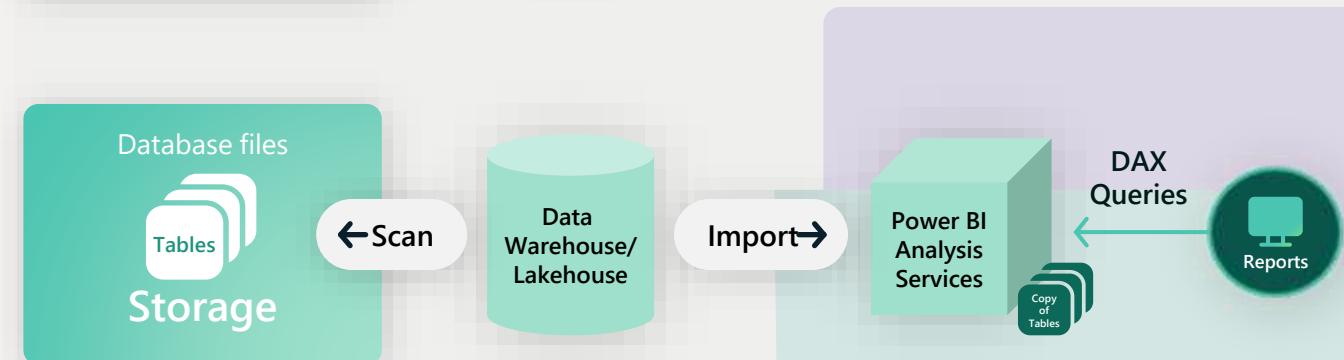
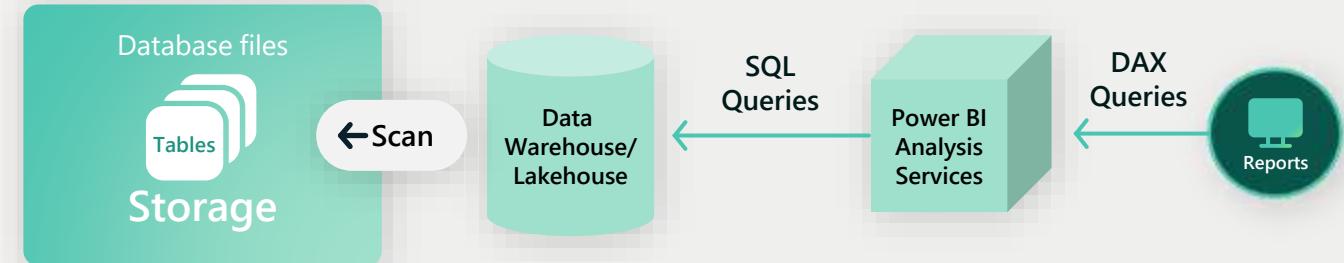


Power BI | Direct Lake Mode

Direct Lake is a fast-path to load the data from the lake straight into the Power BI engine, ready for analysis

Direct Lake is based on loading parquet-formatted files directly from a data lake without having to query a Lakehouse endpoint, and without having to import or duplicate data into a semantic model

Direct Query Mode. Slow, but real time



Direct Lake Mode. Fast and real time





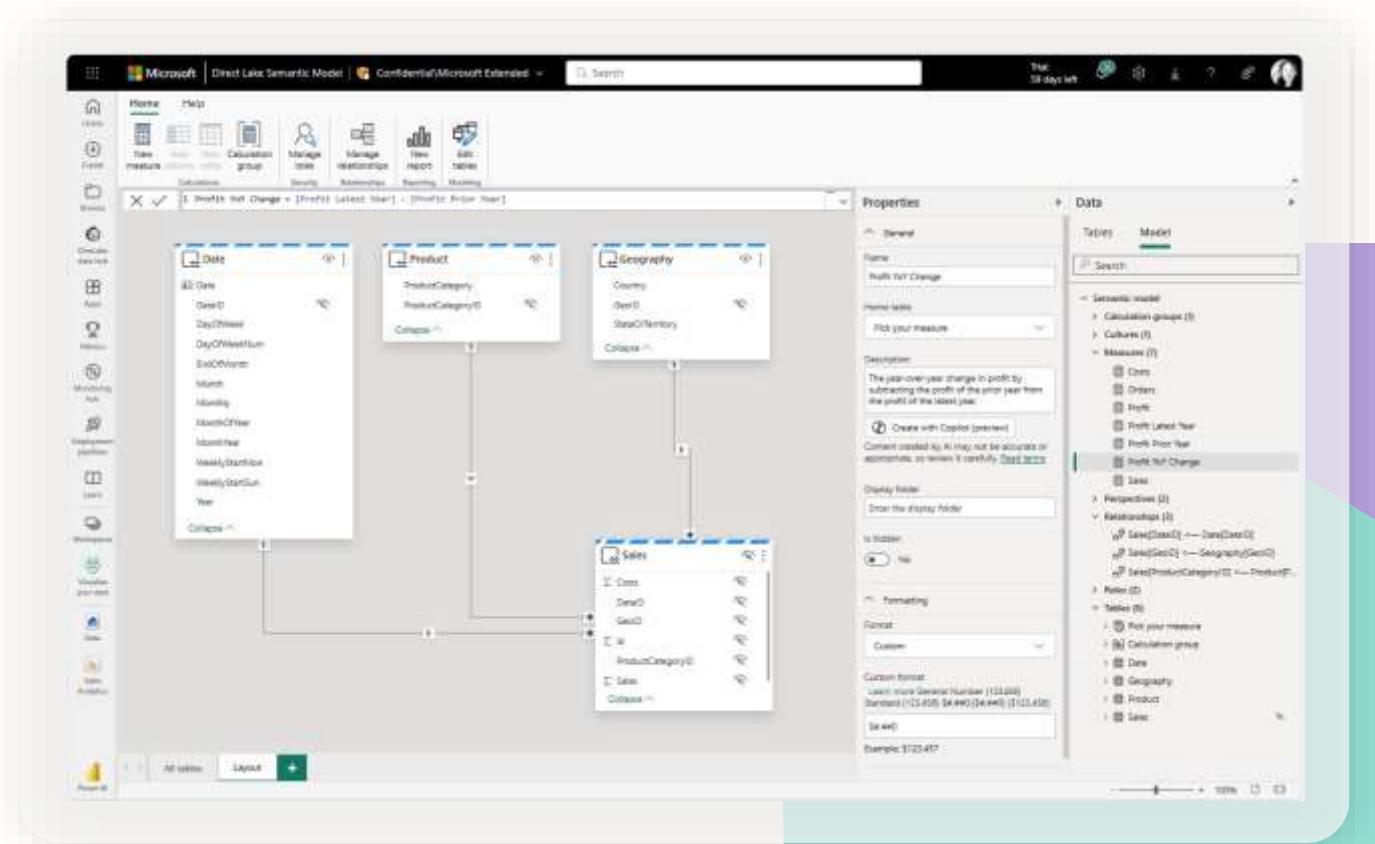
Power BI | Semantic Model

Model your data and quickly unlock insights

Power BI enables everyone to build semantic models they can use to explore data, visualize data in reports, and create scorecards

Key capabilities:

- Power BI semantic models in Fabric use Direct Lake mode to create lightning-fast reports on OneLake data
- Rich semantic modeling experience both online in browser and offline in Power BI Desktop, and in many community-built tools via XMLA endpoint
- Power BI semantic models give data tables meaning by creating relationships between tables and defining business logic in measures





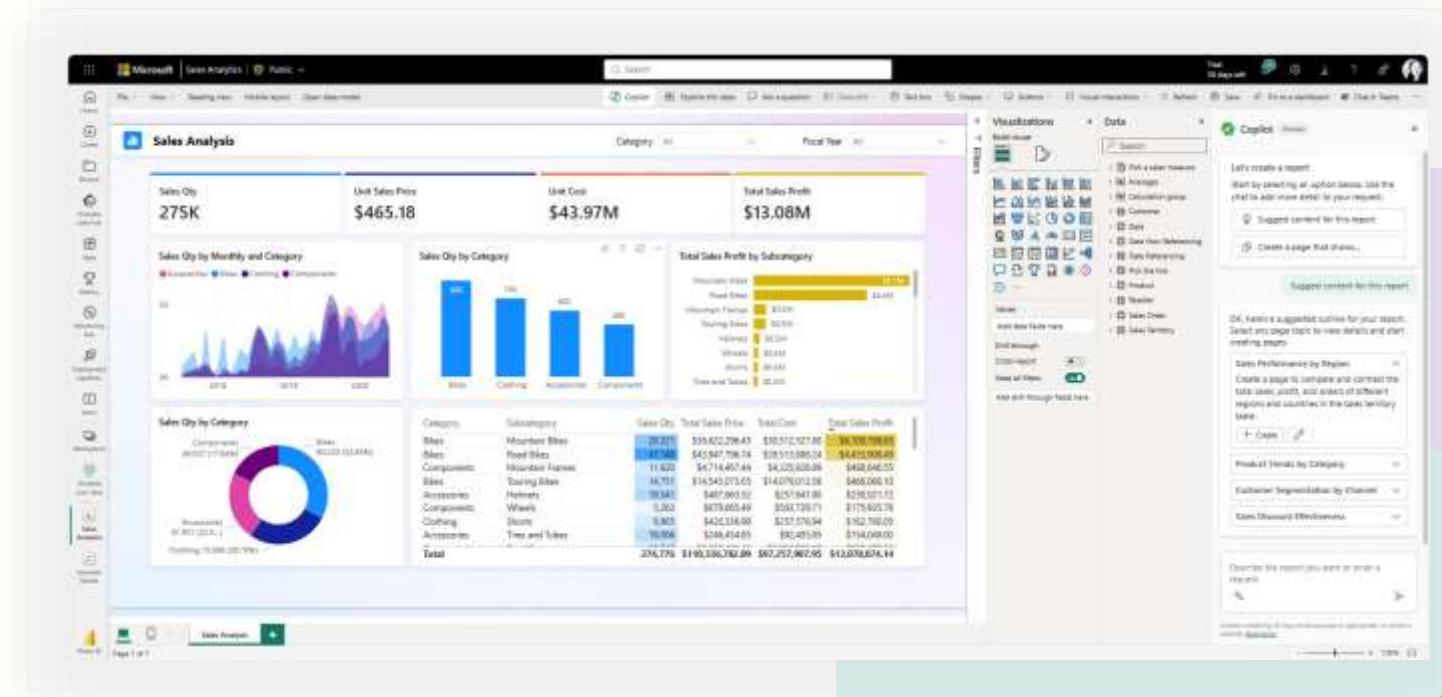
Power BI | Reports

Auto create Power BI reports from your semantic model

Blazing fast performance with Direct Lake

Key capabilities:

- Create an interactive report to discover and share business insights
- Use Copilot to help create, understand, and summarize reports
- Share interactively with Teams and PowerPoint
- View on phone or tablet with mobile-ready layouts for every report
- Explore data and find quick insights





AI-powered analytics

The most complete AI capabilities in a BI product



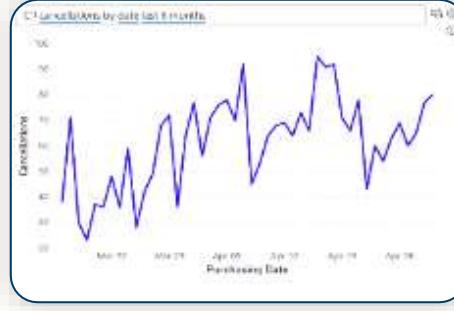
Information workers



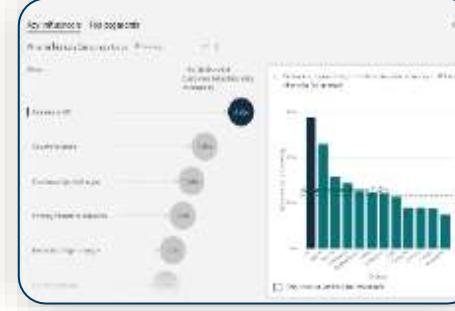
Business analysts



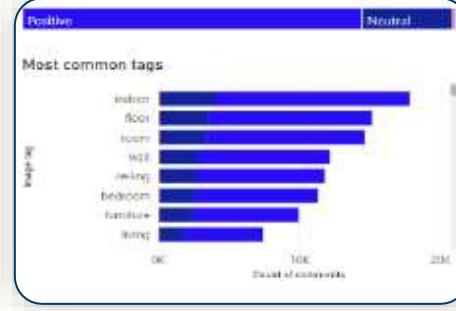
Data scientists



Q&A



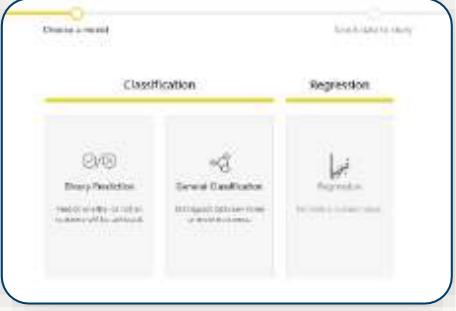
Key driver analysis



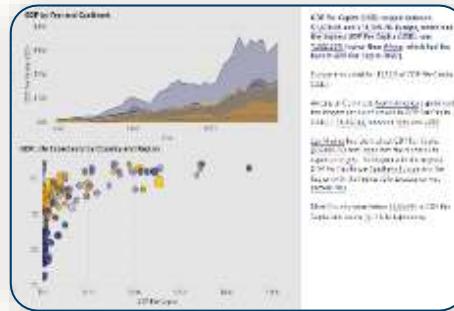
Key phrase extraction



Sentiment analysis



Create ML models



Smart narratives



Root cause analysis



Explore predictions



Python and R integration



Extend with Azure ML



The business analyst role

101010
010101
101010

Get data



Analyze



Visualize



Share



Collaborate



Power BI Desktop is a free companion authoring tool for the Power BI service, updated monthly with new features



Power BI | Get data

Easily connect, clean, and mashup data

101010
010101
101010

Connect to 275+ data sources, both on-premises and cloud



Shape, transform, and clean data for analysis



Live connectivity to on-premises and cloud data sources



Extend with custom data connectors for any data source



Prep your data using the familiar Power Query experience on the web





Power BI | Analyze

Build powerful models and flexibly measures

101010
010101
101010



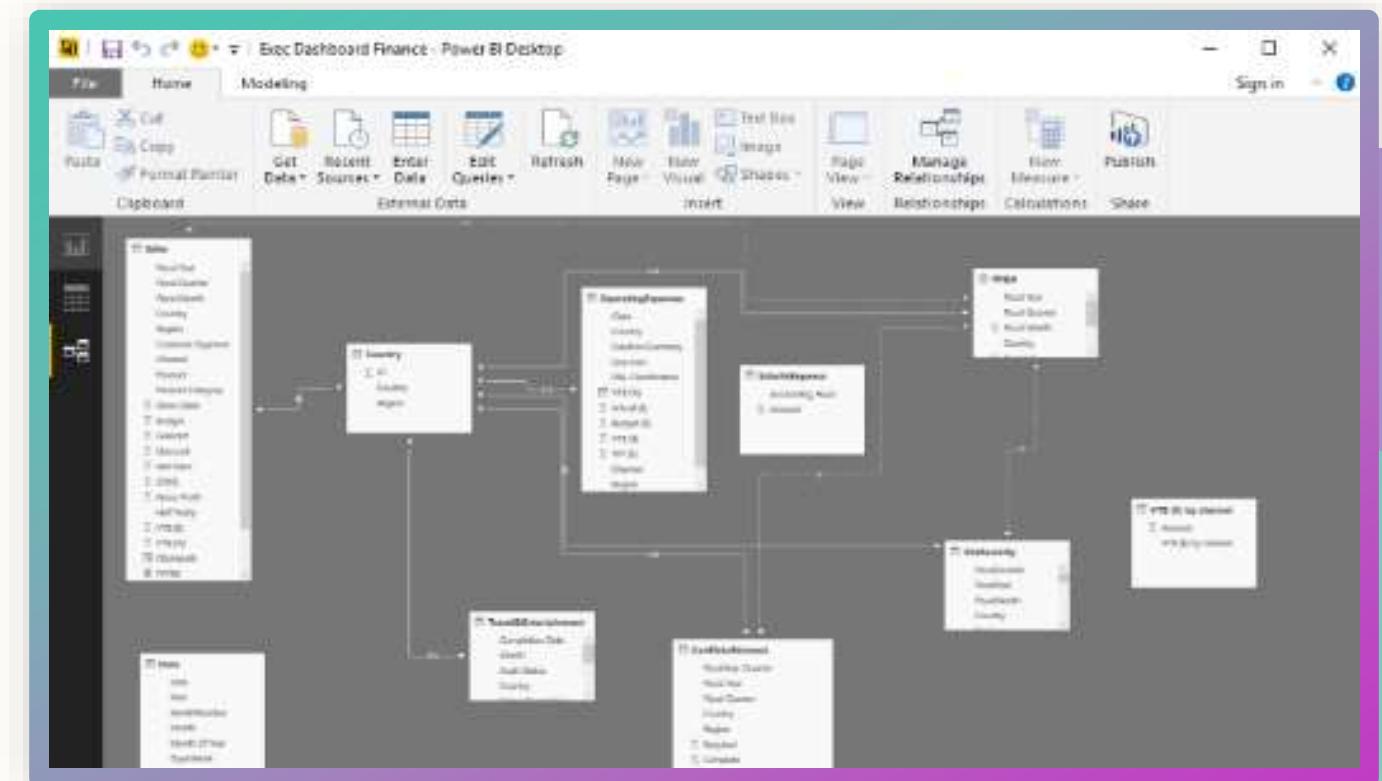
Prep your data using the familiar Power Query experience



Get started quickly with a common data model



Extend self-service prep to Azure Data Lake Storage





Power BI | Visualize

Create stunning interactive reports

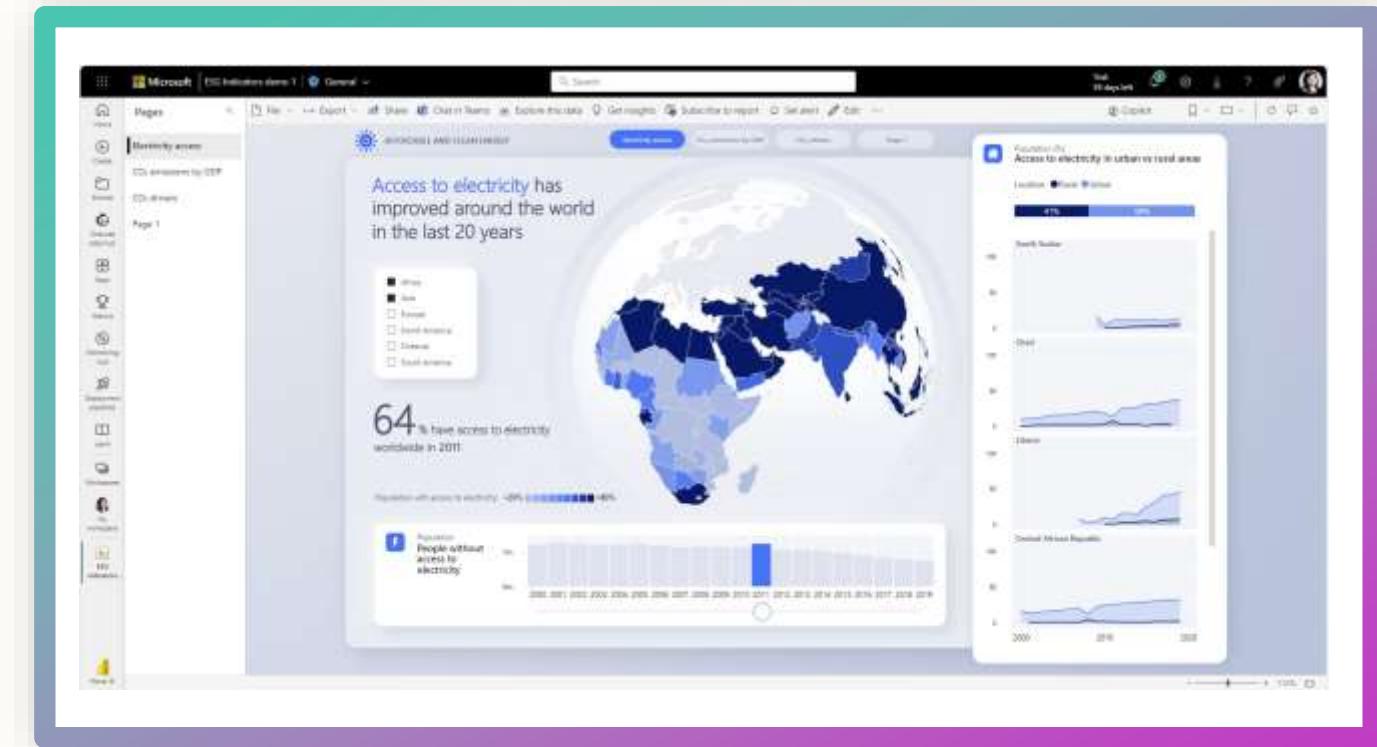
101010
010101
101010



Explore data across multiple interactive visualizations

Provide insights in the context of your business with Custom Visuals

Visualize your data story with bookmarks and custom navigation





Power BI | Share

Share insights with others

101010
010101
101010

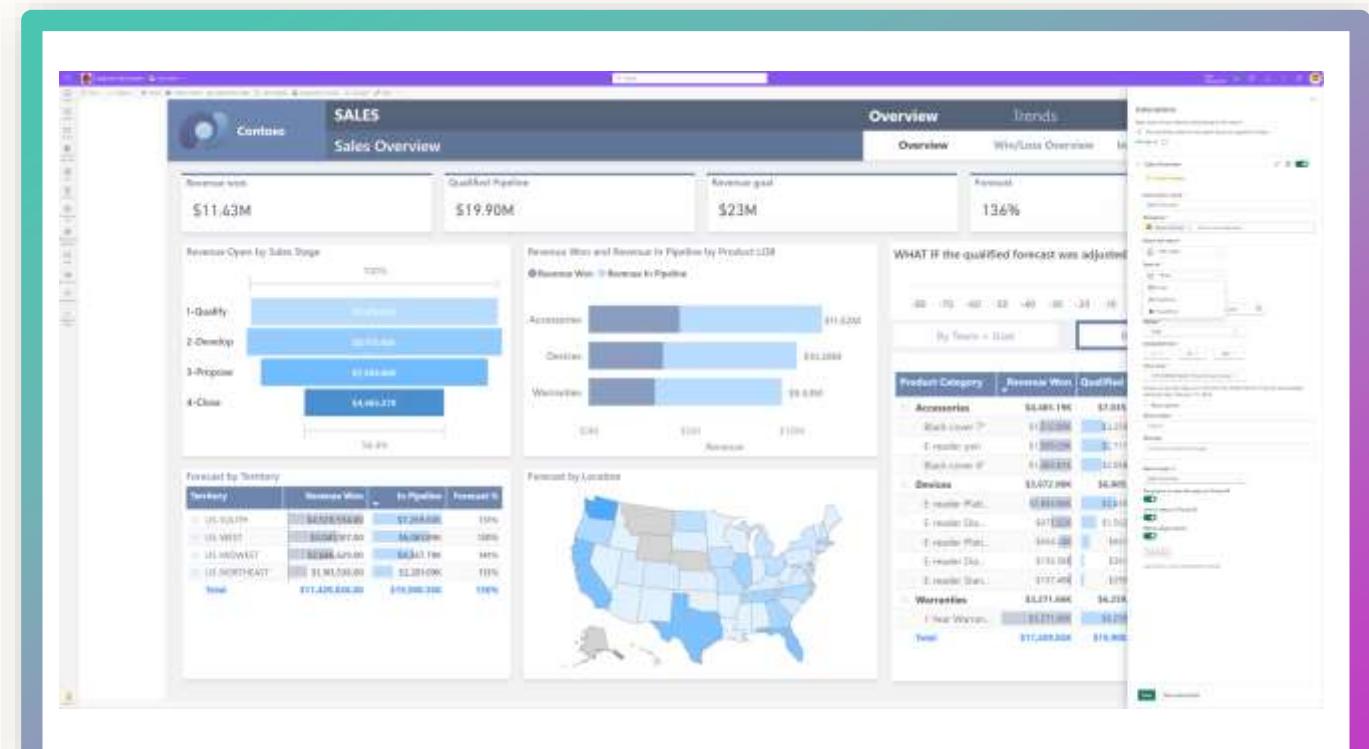


Publish directly to the cloud or on-premises

Regularly deliver reports, so everyone is up to date

Package your reports in apps for easy consumption and control

Manage your analytics content with admin and governance tools





Power BI | Collaborate

Empower your organization

101010
010101
101010



Provide access to reports and dashboards anywhere on any device

Enable users to ask questions and automatically discover insights from the data

Deliver insights through other services such as SharePoint and Teams

Connect live to your models in Excel and Power BI Desktop

The screenshot shows the Microsoft Teams Power BI app interface. At the top, there's a navigation bar with 'Power BI' selected, along with links for 'Home', 'Datasets', 'Create', 'Learn', and 'About'. A search bar is located above the main content area. The main area starts with a greeting 'Good afternoon, Allan' and a sub-header 'Find and share actionable insights to make data-driven decisions'. Below this is a section titled 'Favorites + frequent' featuring four tiles: 'Teams Analytics' (Microsoft App), 'Rockwell Financials' (Microsoft App), 'Finance' (Microsoft Workspace), and 'Finance_1' (Power BI Report). Further down is a 'Featured' section showing four reports: 'Rockwell Financials' (Rockwell Financials logo), 'Finance' (Finance logo), '#Finance Drilldown' (Finance Drilldown logo), and 'Alerts and Anomalies' (Alerts and Anomalies logo). Each report tile includes a small preview image and a timestamp indicating when it was featured.



Power BI | Developer Experiences

Unblock professional team development and deployment scenarios

101010
010101
101010

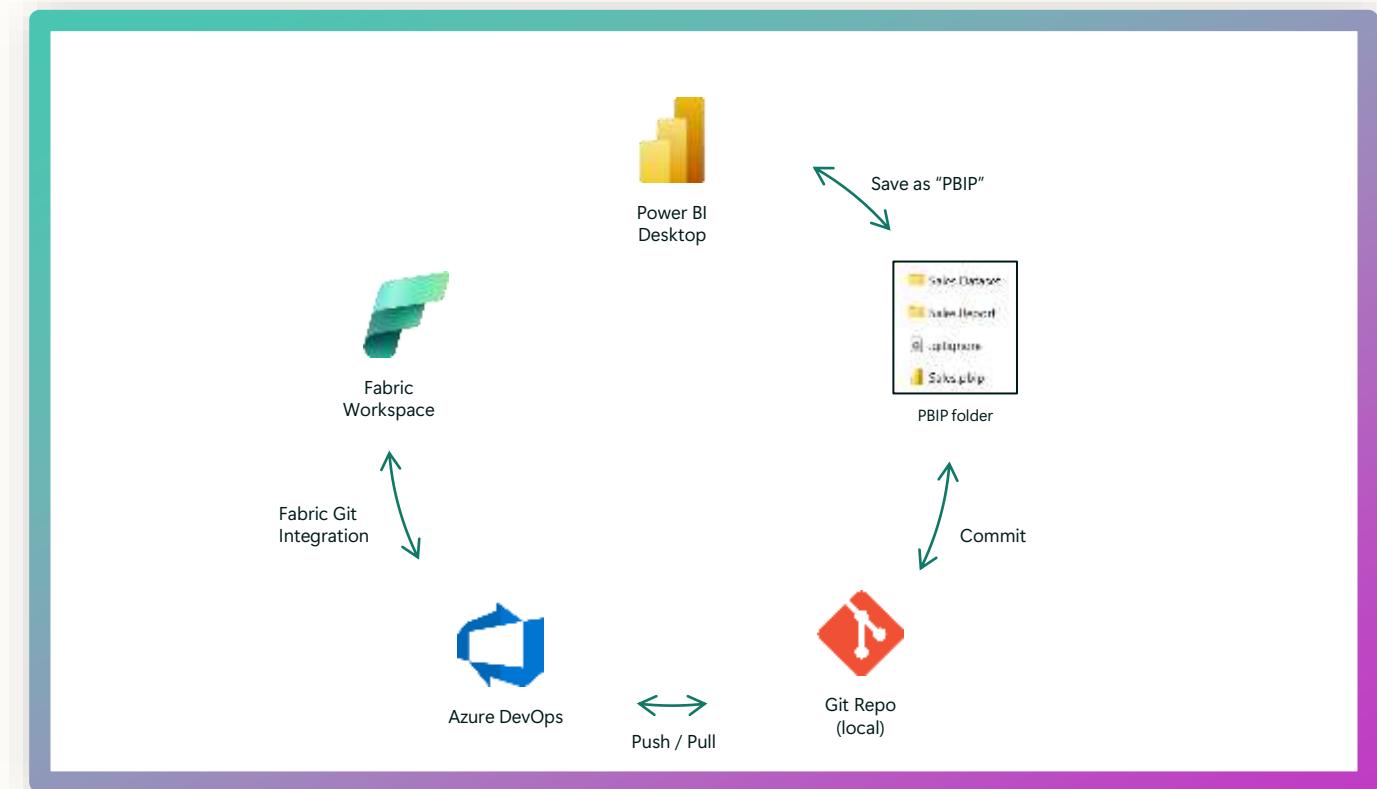


Save as Power BI Project files (PBIP),
text-based source control friendly
format

Track and backup development with Git

Co-develop using Azure DevOps

Automatically sync to Fabric using
Fabric Git Integration





Copilot in Microsoft Fabric overview



Copilot in Microsoft Fabric



Data Factory

Get intelligent code generation to transform data with ease and code explanations to help you better understand complex tasks



Data Engineering and Data Science

Quickly generate code in Notebooks to help work with Lakehouse data and get insights.



Data Warehouse

Write and explain T-SQL queries, or even make intelligent suggestions and fixes while you are coding



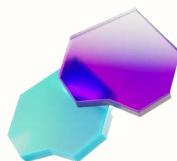
Real-Time Intelligence

Translate questions into KQL queries that you can execute.



Power BI

Quickly create report pages, natural language summaries, and generate synonyms.





Copilot in Fabric

The screenshot shows a Microsoft Fabric dashboard titled "Sales Overview" with the following key metrics:

- Revenue Won: \$7,720,093
- Close %: 28.3%
- Avg Days to Close: 84
- Opportunities Won: 334

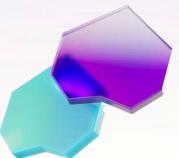
Below these metrics are three charts:

- Revenue Won by Month:** A line chart showing revenue over time from June 2022 to January 2023. The revenue starts around \$1M, dips slightly, then rises steadily to a peak of about \$2.5M in December 2022.
- Close % by Month:** A bar chart showing the close percentage for each month. The values fluctuate between 27% and 32%.
- Close % by Region:** A map of the United States where states are colored according to their close percentage. States like California and Texas appear darker blue (higher close %).
- Avg Days to Close by Month:** A bar chart showing the average days to close for each month. The values range from 96 to 137 days.

To the right of the dashboard is a "Copilot" sidebar:

- Create a report with Copilot:** A text input field with placeholder text: "Describe the report you want, in your own words, and Copilot will create it quickly."
- Help me build a sales report summarizing our key metrics and trends:** A button or suggestion card.
- Sales overview page added:** A message indicating a page has been added.
- Ask a question or type / for suggestions:** An input field with a microphone icon.
- AI-generated content can have mistakes. Make sure it's accurate and appropriate before using it. [Read review terms](#)

to finding insights and creating reports,





Copilot in Fabric | Power BI

Stay focused on your business outcomes and unlock insights in your data

 Create beautiful and insightful reports just by chatting with Copilot

 Define metrics and calculations for your data model just by describing them in natural language

 Use Copilot to find and summarize insights in your data

Copilot in Power BI





Copilot in Fabric | Data Factory

Easily integrate generative AI into your dataflows and pipelines using Copilot

Chat with Copilot to describe data transformations in natural language

Tap into generative AI capabilities from Azure Open AI as data transformation steps

Use Copilot to schedule and run and manage dataflows

Copilot in Data Factory



Copilot in Fabric | Data Engineering and Data Science

**Use Copilot to enrich,
model, analyze and explore
your data in notebooks**

 Work with Copilot to understand how best to analyze your data

 Chat with Copilot to create and configure ML models

 Write code faster with inline code suggestions from Copilot

 Use Copilot to summarize and explain code to understand how it works

Copilot in Data Science





Copilot in Fabric | Data Warehouse

Use Copilot to help write SQL queries, create tables, and even get data

Instantly explain queries with detailed comments next to the code

Quickly write new SQL queries and even get code suggestions as you write

Fix queries with a single click

The screenshot shows the Power BI Data Studio interface. On the left, there's an 'Explorer' sidebar with a tree view of database objects like 'Geography', 'HackneyLoc...', 'Mediation', 'Time', 'Trip', 'Weather', 'Views', 'Functions', 'StoredProcedure', 'guest', 'INFORMATION_SCHEMA', 'queryimights', 'sys', 'Security', and 'Queries'. Under 'Queries', there are 'My queries' (with a 'SQL query T' item) and 'Shared queries' (with items like 'F_AVG', 'Top 10-Cities', 'Trip Summary', and 'W_Impact'). The main area is titled 'Data preview' and shows a table with 1000 rows of data. The columns are labeled: ID, Date, abc_DateKey, abc_DayOfMonth, abc_DaySuffix, abc_DayName, abc_DayOfWeek, abc_DayOfWeekInMonth, abc_DayOfWeekInYear, abc_DayOfQuarter, abc_DayOfficer, and abc_Y. The first few rows of data are:

ID	Date	abc_DateKey	abc_DayOfMonth	abc_DaySuffix	abc_DayName	abc_DayOfWeek	abc_DayOfWeekInMonth	abc_DayOfWeekInYear	abc_DayOfQuarter	abc_DayOfficer	abc_Y	
1	20190317	2019-03-17 00:00:00.000000	03/17/2019	17	17th	Tuesday	3	3	11	11	76	3
2	20050317	2005-03-17 00:00:00.000000	03/17/2005	17	17th	Thursday	5	5	11	11	76	3
3	20010317	2001-03-17 00:00:00.000000	03/17/2001	17	17th	Saturday	7	3	11	11	76	3
4	20090317	2009-03-17 00:00:00.000000	03/17/2009	17	17th	Sunday	3	3	11	11	76	3
5	20040317	2004-03-17 00:00:00.000000	03/17/2004	17	17th	Wednesday	4	3	11	11	77	3
6	20030317	2000-03-17 00:00:00.000000	03/17/2000	17	17th	Friday	6	2	11	11	77	3
7	20120317	2012-03-17 00:00:00.000000	03/17/2012	17	17th	Saturday	7	1	11	11	77	3
8	20060317	2006-03-17 00:00:00.000000	03/17/2006	17	17th	Friday	6	2	11	11	76	3
9	20110317	2011-03-17 00:00:00.000000	03/17/2011	17	17th	Thursday	5	3	11	11	76	3
10	20100317	2019-03-17 00:00:00.000000	03/17/2019	17	17th	Wednesday	4	3	11	11	76	3
11	20070317	2007-03-17 00:00:00.000000	03/17/2007	17	17th	Saturday	7	3	11	11	76	3
12	20140314	2014-03-14 00:00:00.000000	03/14/2014	14	14th	Friday	6	2	7	7	45	3
13	20090314	2003-03-14 00:00:00.000000	03/14/2003	14	14th	Friday	6	2	7	7	45	3
14	20080314	2006-03-14 00:00:00.000000	03/14/2006	14	14th	Thursday	5	2	7	7	45	3
15	20110314	2011-03-14 00:00:00.000000	03/14/2011	14	14th	Monday	2	2	7	7	45	3
16	20050314	2007-03-14 00:00:00.000000	03/14/2007	14	14th	Monday	2	2	7	7	45	3
17	20010314	2001-03-14 00:00:00.000000	03/14/2001	14	14th	Wednesday	4	2	7	7	45	3
18	20080314	2006-03-14 00:00:00.000000	03/14/2006	14	14th	Tuesday	3	2	7	7	45	3
19	20020314	2002-03-14 00:00:00.000000	03/14/2002	14	14th	Thursday	5	2	7	7	45	3
20	20130314	2012-03-14 00:00:00.000000	03/14/2012	14	14th	Tuesday	3	2	7	7	45	3
21	20000314	2000-03-14 00:00:00.000000	03/14/2000	14	14th	Monday	2	2	7	7	45	3
22	20100314	2010-03-14 00:00:00.000000	03/14/2010	14	14th	Sunday	1	2	7	7	45	3
23	20070314	2007-03-14 00:00:00.000000	03/14/2007	14	14th	Wednesday	4	2	7	7	45	3
24	20130314	2013-03-14 00:00:00.000000	03/14/2013	14	14th	Thursday	5	2	7	7	45	3
25	20091031	2009-10-31 00:00:00.000000	10/31/2009	31	31st	Saturday	7	5	44	5	304	5
26	20121031	2012-10-31 00:00:00.000000	10/31/2012	31	31st	Wednesday	4	5	44	5	305	5
27	20071031	2007-10-31 00:00:00.000000	10/31/2007	31	31st	Wednesday	4	5	44	5	304	5
28	20021031	2002-10-31 00:00:00.000000	10/31/2002	31	31st	Thursday	5	5	44	5	304	5
29	20151031	2015-10-31 00:00:00.000000	10/31/2015	31	31st	Saturday	7	5	44	5	304	5



Copilot in Fabric | Real-Time Intelligence

Explore and analyze your real-time data with ease in Copilot

Ask questions about your real-time data in conversational language

Automatically translate it to a KQL query you can execute

Get the most from your time-series data stored in Eventhouse even if you're less familiar with KQL queries

The screenshot shows the Microsoft Fabric interface. On the left, there's a sidebar with various navigation options like Home, Help, Save, Copilot, Export to CSV, Database (Operations), and a search bar. The main area has tabs for Operations, Preview, Recall, Copy query, Pin to dashboard, KQL Tools, and Build PowerBI report. A code editor window displays the following KQL query:

```
15 // This query returns the number of ingestions per hour in the given table.  
16 YOUR_TABLE_HERE  
17 | summarize IngestionCount = count() by(ingestion_time(), 1h)  
18  
19  
20 // Use "take" to view a sample number of records in the table and check the data.  
21 Deliveries  
22 | take 100  
23  
24 // See the most recent data - records ingested in the last 24 hours.  
25 Deliveries  
26 | where ingestion_time() between (now(-1d) .. now())  
27
```

Below the code editor is a table titled "Table 1 Stats" with the following data:

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fwd_flag	PULocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax
1	2022-06-06 05:25:34	2022-06-06 05:26:14	2.0	0.3	5.0	N	132	132	1	0.8	1.25	0.0
1	2022-06-06 05:27:40	2022-06-06 05:28:02	2.0	0.3	5.0	N	132	132	1	68.0	1.25	0.0
2	2022-06-06 05:02:33	2022-06-06 05:23:20	1.0	14.76	3.0	N	211	1	1	57.0	0.5	0.0
2	2022-06-06 05:51:21	2022-06-06 05:54:18	1.0	6.7	1.0	N	90	186	1	43.0	0.5	0.5
2	2022-06-06 05:35:45	2022-06-06 06:31:00	3.0	18.38	2.0	N	132	230	2	52.0	0.0	0.5
1	2022-06-06 05:25:46	2022-06-06 05:57:22	1.0	9.9	99.0	N	61	142	1	39.2	0.0	0.5
1	2022-06-06 05:01:24	2022-06-06 05:08:03	2.0	0.9	1.0	N	249	114	2	6.0	3.0	0.5
1	2022-06-06 05:31:40	2022-06-06 05:44:19	1.0	2.8	1.0	N	106	140	1	11.5	3.0	0.5
2	2022-06-06 05:26:56	2022-06-06 05:48:55	6.0	9.3	1.0	N	132	76	2	28.0	0.5	0.5
2	2022-06-06 05:07:57	2022-06-06 05:37:16	2.0	16.84	2.0	N	233	132	2	52.0	0.0	0.5
1	2022-06-06 05:00:27	2022-06-06 05:19:40	2.0	54.8	1.0	N	132	179	1	39.0	1.75	0.5
2	2022-06-06 05:38:15	2022-06-06 05:44:38	1.0	13.8	1.0	N	7	140	1	7.0	0.5	0.5
2	2022-06-06 05:50:34	2022-06-06 05:52:58	1.0	1.02	1.0	N	229	170	1	5.0	0.5	0.5



LLM programming made easier

Powered by SynapseML

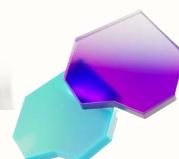
Empowering developers with native **Azure OpenAI model integration.**

Use **LangChain** and **Semantic Kernel** to develop and scale custom generative AI models – right from your Fabric Notebook.

Massively scalable, AI powered data processing made easier, through **SynapseML**.



The screenshot shows a Microsoft Fabric Notebook interface. The title bar reads "SynapseML-Azure 0... | Confidential/Microsoft Fab... - Saved -". The main area displays a Python code snippet under the heading "Prerequisites: package installation". The code imports synapse.ml.core and synapse.ml.openai, then imports langchain. A tooltip provides a brief description of the completion feature. Below this, under "Completions", there is a list of suggestions for "deployment_name": "text-davinci-003" and "code-davinci-002". Further down, a code completion tooltip appears over a line of code involving spark.createDataFrame and completion_type. The status bar at the bottom indicates "Not connected" and "Save options: Automatic".



Step 3: Create the LLM model

Leverage SynapseML and LangChain to initialize a conversational agent that utilizes the specified GPT-3.5 model hosted on Azure

```
1  llm = AzureChatOpenAI(  
2      deployment_name='gpt-35-turbo',  
3      model_name='gpt-35-turbo',  
4      temperature=0.1,  
5      verbose=False,  
6  )  
7  
8  template = """  
9      Your job is to determine the product category.  
10     Please use all information available in the dataset to determine the product category as if this is going to be sold on Amazon.  
11     Provide multiple categories separated by a comma if multiple categories are appropriate.  
12     If you are unsure or a category cannot be determined, say "Unknown".  
13     Write the category as a single word or short phrase.  
14     Examples:  
15     DC Cover Girls: Black Canary by Joëlle Jones Statue: Toys,  
16     Pacific Play Tent Agility Dog Training Chute: Pet Supplies."""  
17  
18  system_message = SystemMessage(content=template)  
19  human_template= "{text}"  
20  human_message_prompt = HumanMessagePromptTemplate.from_template(human_template)  
21  chat_prompt = ChatPromptTemplate.from_messages([system_message, human_message_prompt])  
22  chain = LLMChain(llm=llm, prompt=chat_prompt)
```

[6] ✓ <1 sec -Command executed in 417 ms by Nellie Gustafsson on 11:28:37 AM, 11/05/23

PySpark (Python)

And then I want to also instantiate a variable with my prompt.

ready Save option: Automatic

Selected Cell 10 of 15 cells





Generative AI on your data

Announcing Public Preview

Deliver custom generative AI experiences for **your data**



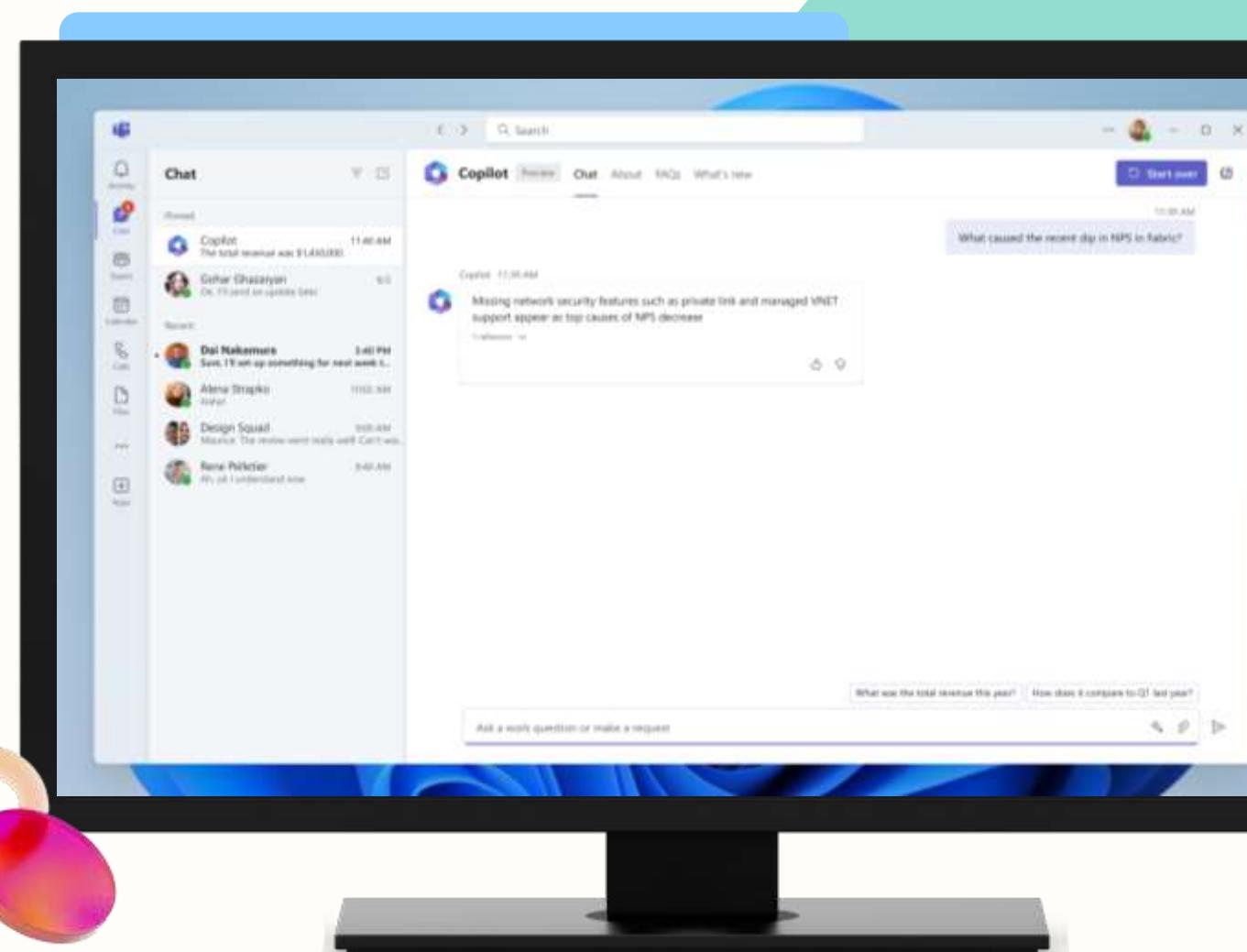
Enable custom Q&A on your data in Fabric



Define custom business semantics and grounding **unique to your org**



Scale the custom experiences to **M365 Chat, Copilot Studio, and Azure AI Studio**





AI skills in Microsoft Fabric

The screenshot shows the Microsoft Fabric Synapse Data Science Home interface. At the top, there's a navigation bar with the Microsoft DXT logo, the Synapse Data Science Home link, a search bar, and a trial status message ("Trial: 58 days left"). On the far right, there are user profile icons and a help icon.

The main area has a sidebar on the left with icons for Home, Create, Browse, OneLake data hub, Monitoring hub, Workspaces, and Data Science. The "Contoso Outdoors..." workspace is selected.

New section:

- Current workspace: Contoso Outdoors Sales
- Items will be saved to this workspace.
- Buttons: ML model, Experiment, Notebook, Environment (Preview), AI Skill (Preview), Import notebook, Use a sample.

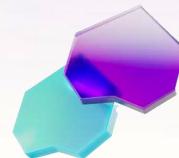
Recommended section:

- My workspace: You frequently open this. Shows a thumbnail of a person and an "Open" button.
- Notebook 6: You frequently open this. Shows a code editor icon and an "Open" button.
- Contoso Sales Excellence: You frequently open this. Shows a people icon and an "Open" button.
- Copilot Demo: You frequently open this. Shows a code editor icon and an "Open" button.

Quick access section:

- Recent (selected): Shows a list of recent workspaces.
- Favorites: Shows a list of favorite workspaces.
- Filter by keyword: A search bar.
- Filter: A dropdown menu.

Name	Type	Opened	Owner	Endorsement	Sensitivity	Location
Contoso Outdoors Sales	Workspace	now	—	—	—	Workspaces
Contoso Sales Excellence	Workspace	now	—	—	—	Workspaces



Copilot in Fabric pricing



Copilot in Fabric is limited to customers who have purchased Fabric capacity (F64 or higher) or Power BI Premium capacity (P1 and above) and is not included in the Fabric free account or trial or Power BI per user licenses



You can simply count Copilot usage against your existing Fabric or Power BI Premium capacity



Copilot usage is measured by the number of tokens processed. Tokens can be thought of as pieces of words. Approximately 1,000 tokens are about 750 words. Prices are calculated per 1,000 tokens, and input and output tokens are consumed at different rates

Operation in Metrics App	Description	Operation Unit of Measure	Consumption rate
Copilot in Fabric	The input prompt	Per 1,000 Tokens	400 CU seconds
Copilot in Fabric	The output completion	Per 1,000 Tokens	1,200 CU seconds

See full details here: <https://aka.ms/Copilot-Billing-Docs>



Copilot in Fabric status



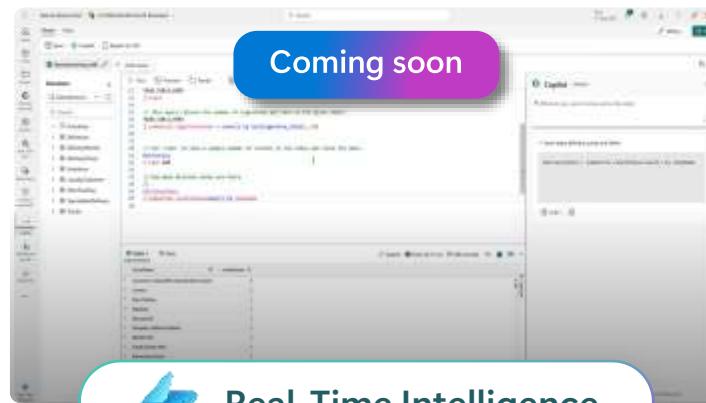
 Data Factory



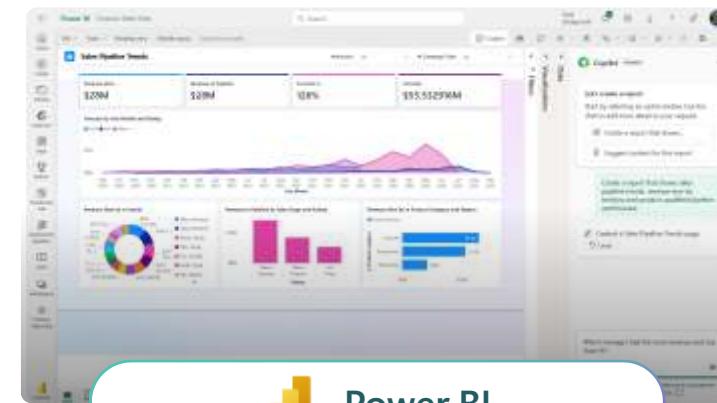
 Data Warehouse



 Data Science



 Real-Time Intelligence



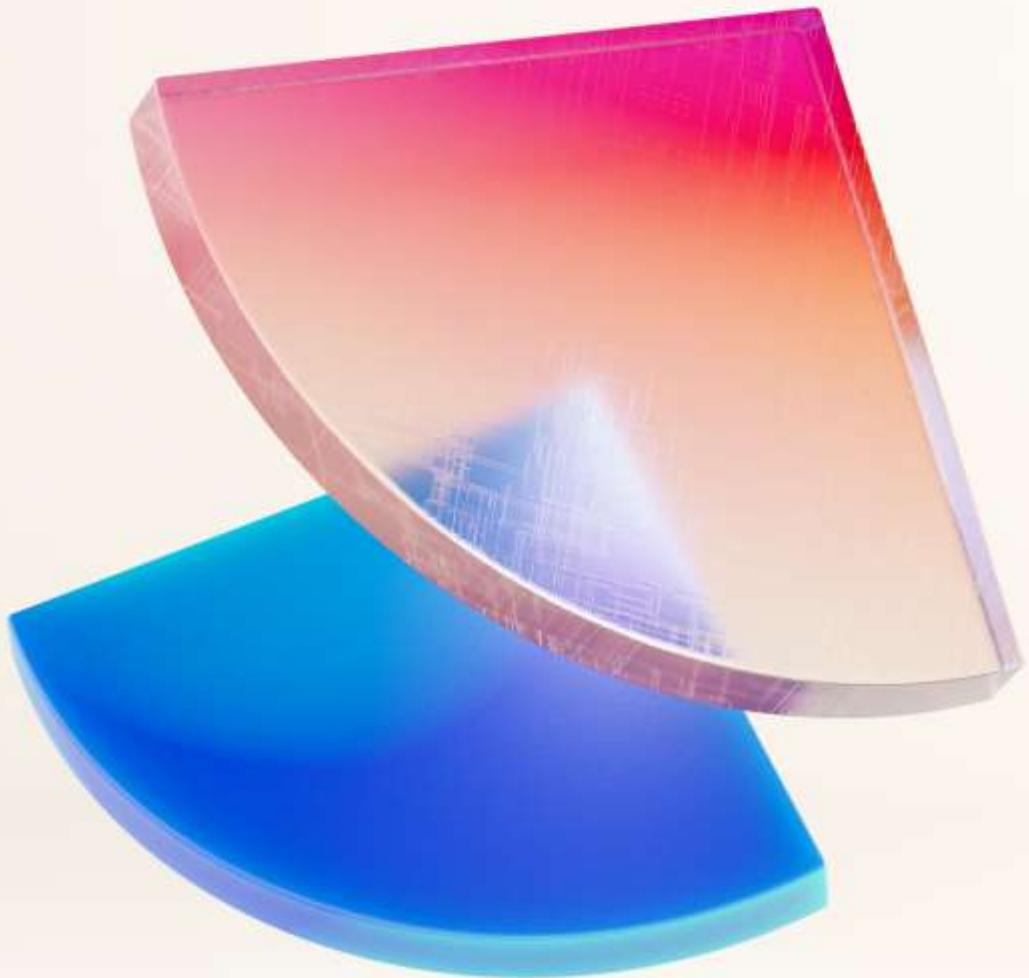
 Power BI

Last updated: May 2024

Enabling Copilot in Fabric

1. Your administrator needs to enable the tenant switch before you can start using Copilot. Administrators can read the article [Copilot tenant settings](#) for details.
2. Your F64 or P1 capacity needs to be in one of the regions listed in this article, [Fabric region availability](#).
3. If your tenant or capacity is outside the US or France, Copilot is disabled by default unless your Fabric tenant admin enables the [Data sent to Azure OpenAI can be processed outside your tenant's geographic region, compliance boundary, or national cloud instance](#) tenant setting in the Fabric Admin portal.
4. Copilot in Microsoft Fabric isn't supported on trial SKUs. Only paid SKUs (F64 or higher, or P1 or higher) are supported.





OneLake overview

OneLake for all data

“The OneDrive for data”



OneDrive
for documents



OneLake
for data

OneLake provides a data lake as a service without you needing to build it

OneLake for all Data

"The OneDrive for Data"

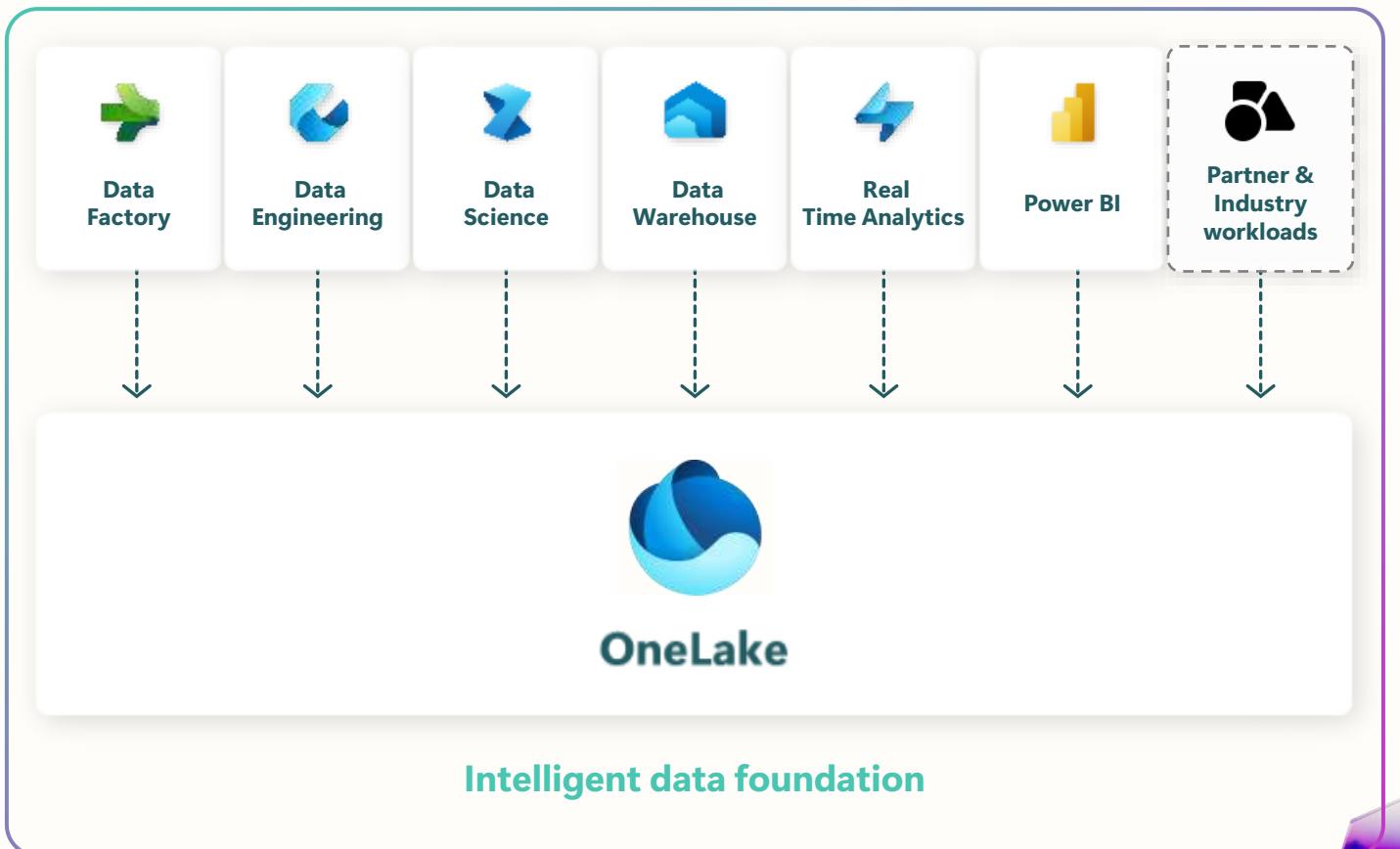
A single SaaS lake for the whole organization

Provisioned automatically with the tenant

All workloads automatically store their data in the OneLake workspace folders

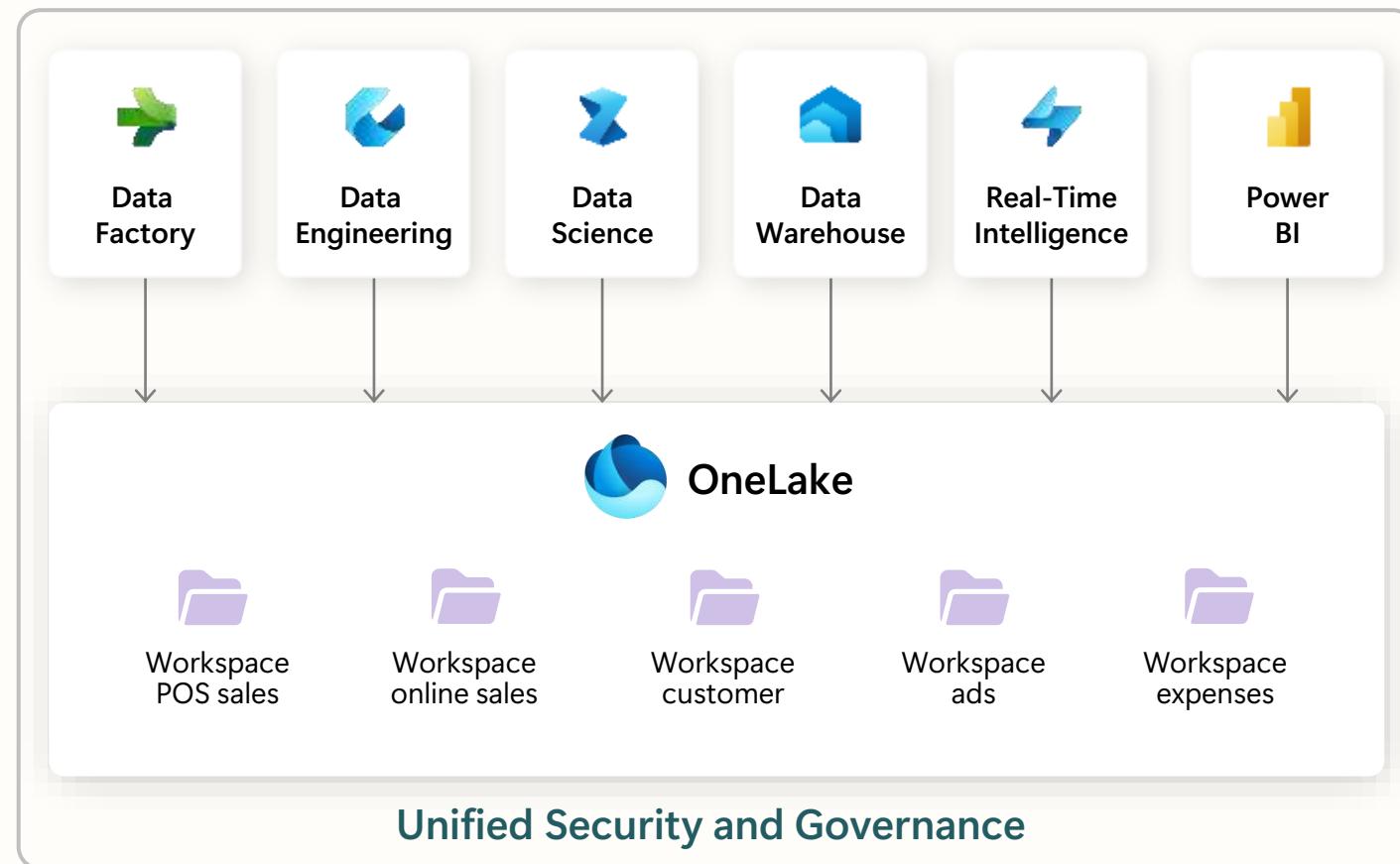
All the data is organized in an intuitive hierarchical namespace

The data in OneLake is automatically indexed for discovery, MIP labels, lineage, PII scans, sharing, governance and compliance



A single unified SaaS data lake

"No Silos"



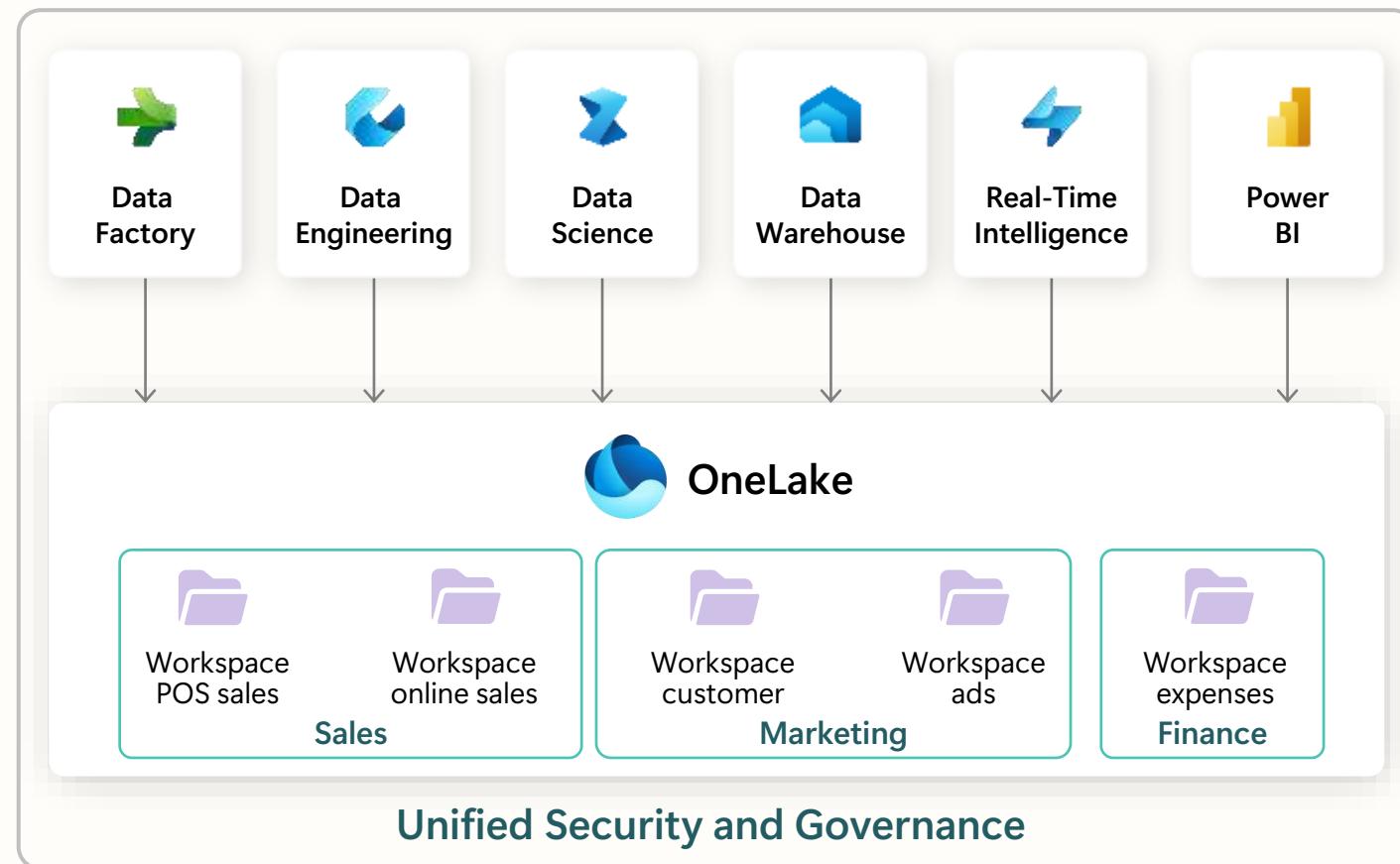
Provisioned automatically with the tenant.

Any data in OneLake works with out-of-the-box governance such as data lineage, data protection, certification, catalog integration, etc. All data is ultimately under the control of a tenant admin.

OneLake enables distributed ownership. Different workspaces allow different parts of the organization to work independently while still contributing to the same data lake. Each workspace can have its own administrator, access control, region and capacity for billing.

OneLake for all domains

OneLake gives a true data mesh as a service



Introducing domains as an integral part of Fabric:
A domain is a way to logically group together all the
data in an organization relevant to an area or field,
according to business needs.

Domains are defined with domain admins and
contributors who can associate workspaces and
group them together under a relevant domain.

Federated governance can be achieved by
delegating settings to domain admins, thus allowing
them to achieve more granular control over
their business area.

Domains simplify discovery and consumption of data
across the organization, thus allowing business
optimized consumption.

Avoid data swamps by endorsing certain data as
certified or promoted, thus encouraging reuse.

OneLake which logically spans the world

To achieve data residency requirements, workspaces can reside in different regions around the world while still being part of the same data lake.

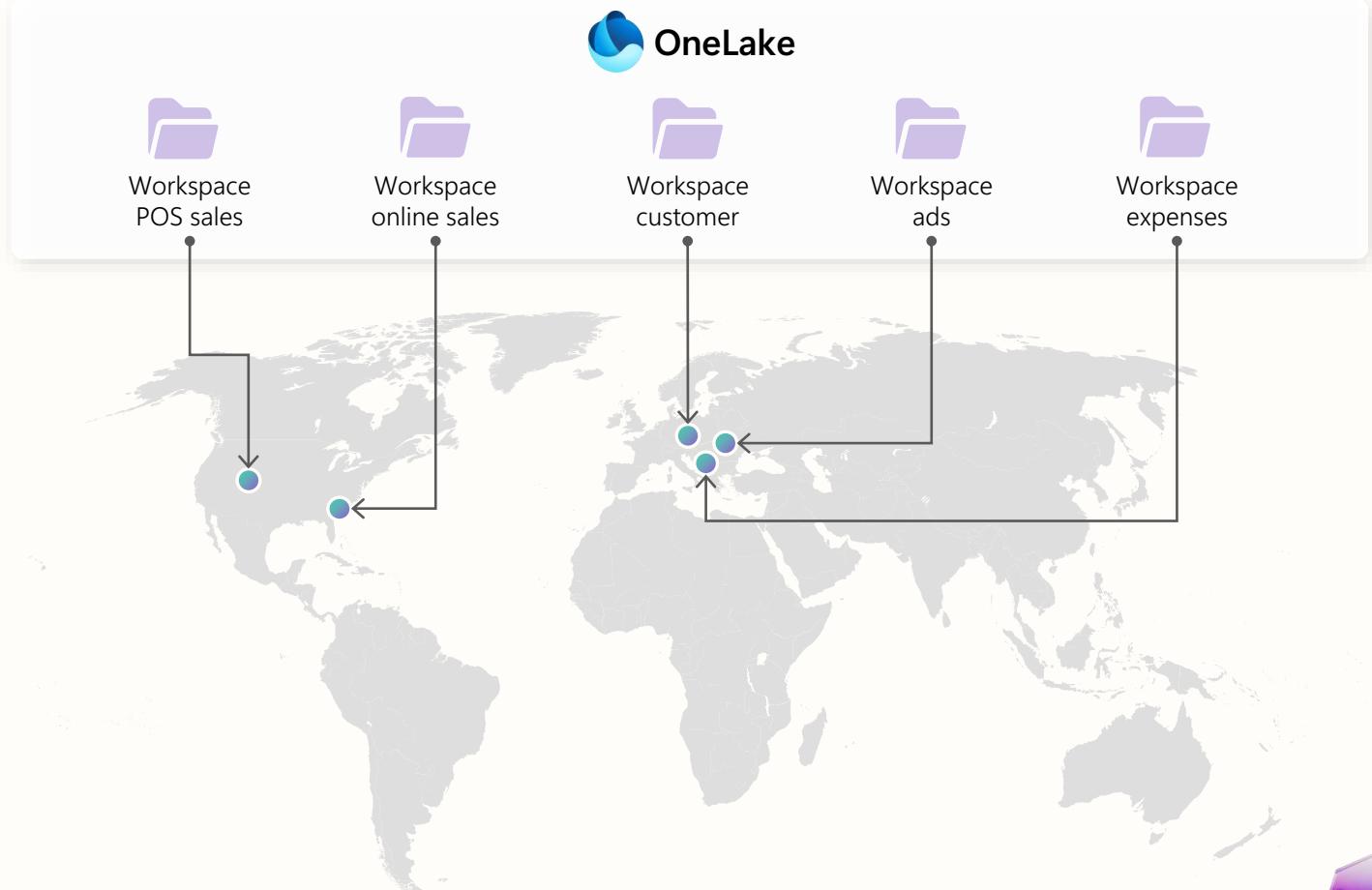
Data can reside in different regions without the overhead of managing different storage resources and without creating data silos.

OneLake provisions storage resources for each workspace to meet demand for scale (capacity, throughput and IOPS).

Underlying physical storage is virtualized away.

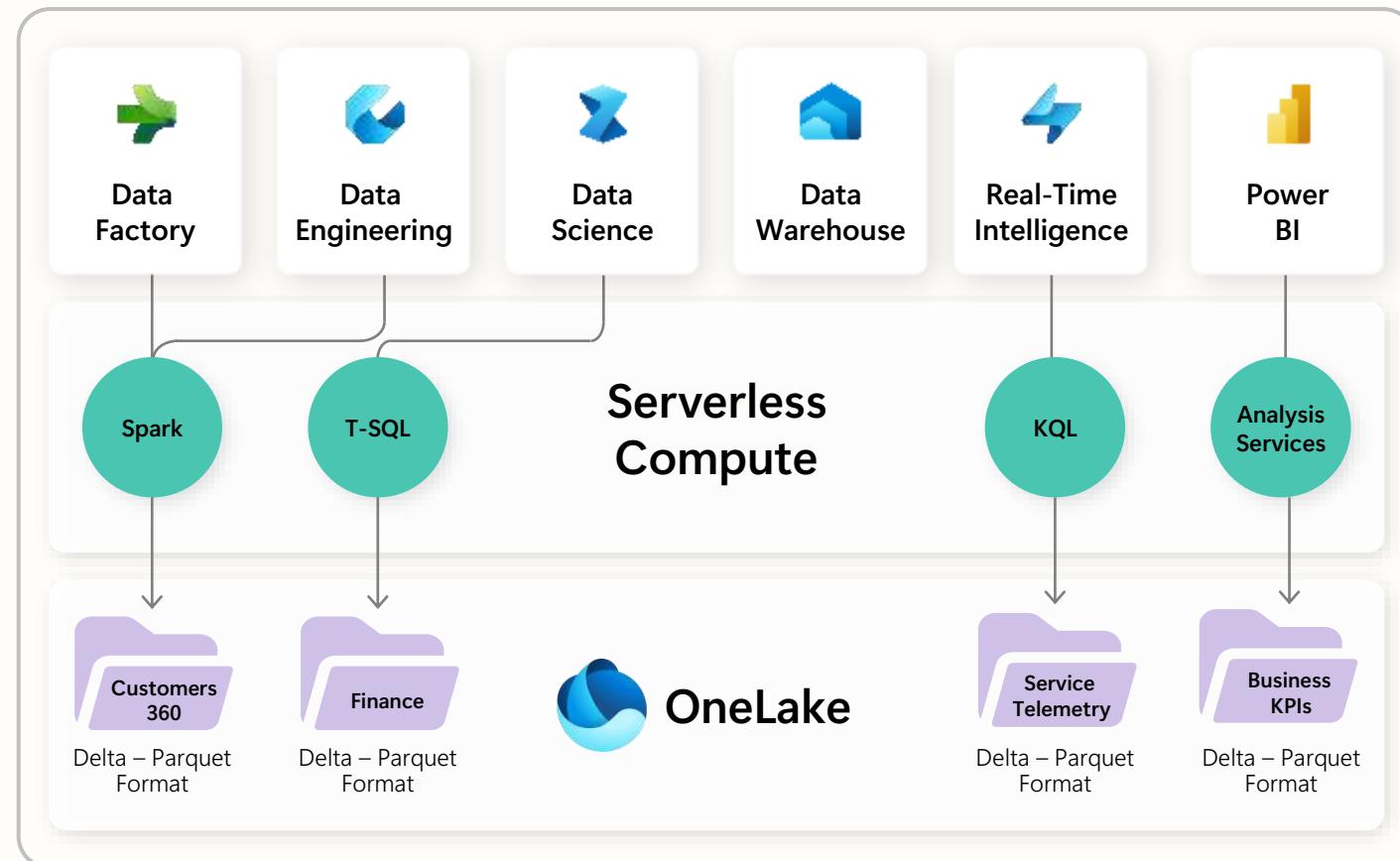
All storage is zone redundant by default with an option for Geo redundancy.

Unified Security and Governance



One Copy for all computers

Real separation of compute and storage



All the compute engines store their data automatically in OneLake as data items.

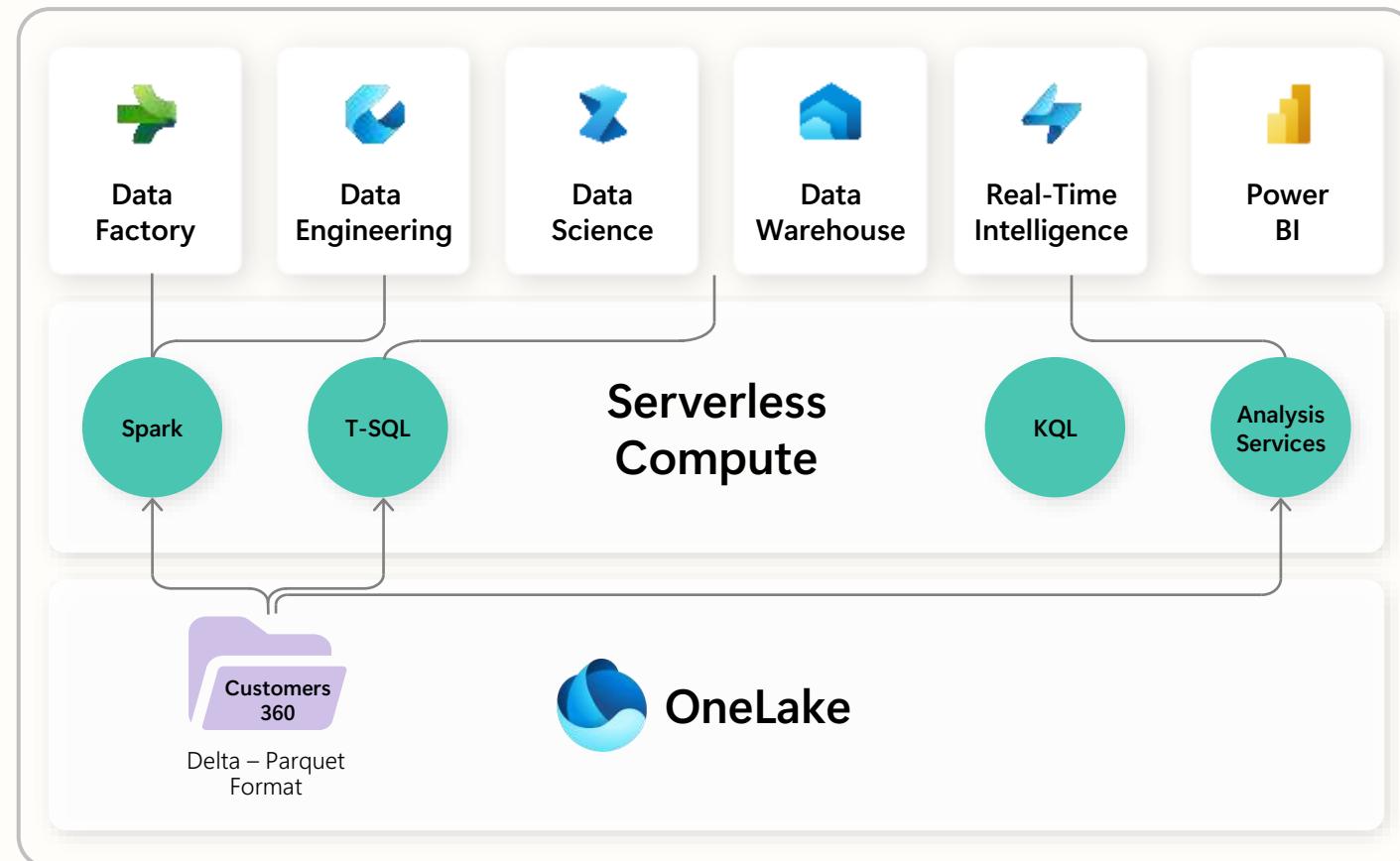
The data is stored in a single common format.

Delta – Parquet, an open standards format, and it is the storage format for all tabular data in Fabric.

All the compute engines have been fully optimized to work with Delta Parquet as their native format.

One Copy for all computers

One copy of data can be read by all engines



Once data is stored in the lake, it is directly accessible by all the engines without needing any import/export.

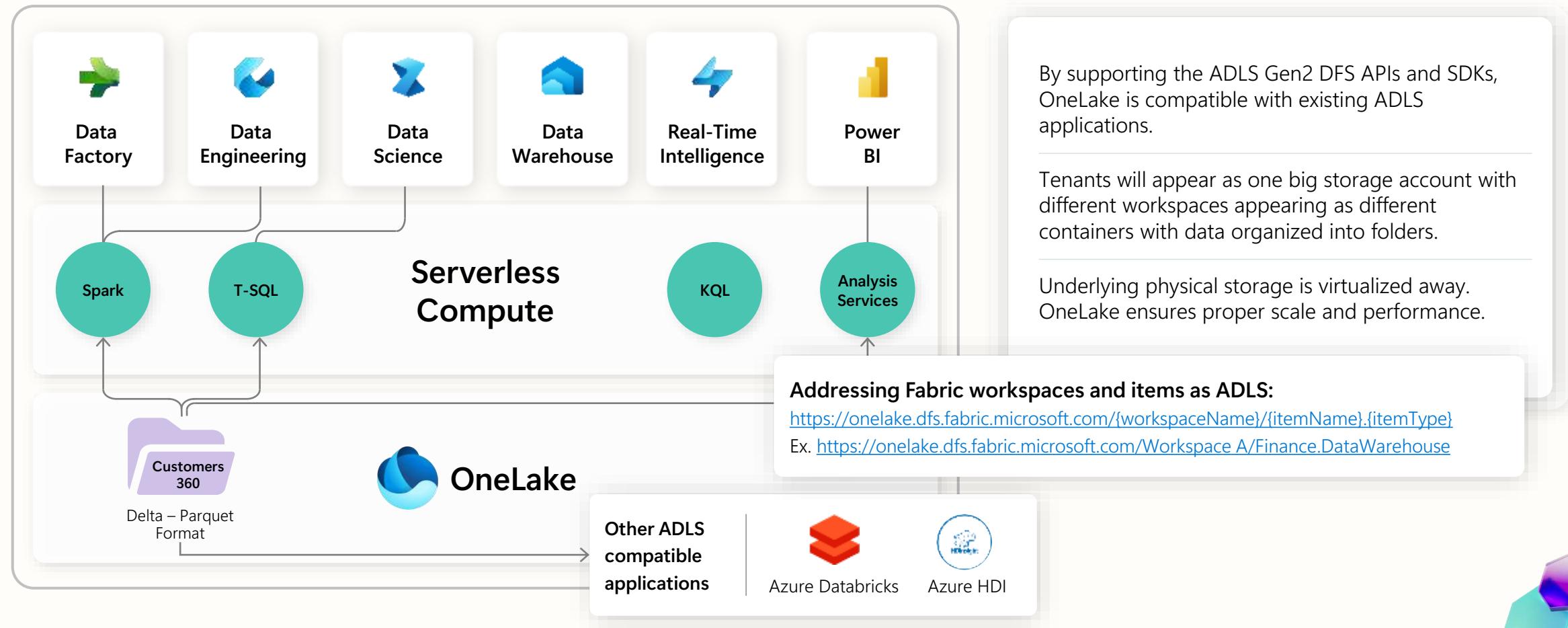
You are able to choose the right engine for the right job.

All the compute engines have been fully optimized to work with Delta Parquet as their native format.

Shared universal security model is enforced across all the engines (coming soon).

Open Access to data in OneLake

No lock-in with industry standard APIs and open file formats



Mirroring in Microsoft Fabric

Simplify near real-time intelligence

Fabric Mirroring enables adding existing databases and data warehouses to Fabric without any ETL.

A full editing experience of the source database is available for the Mirrored database.

Data is replicated into OneLake in Delta format and kept up-to-date in near-real-time.

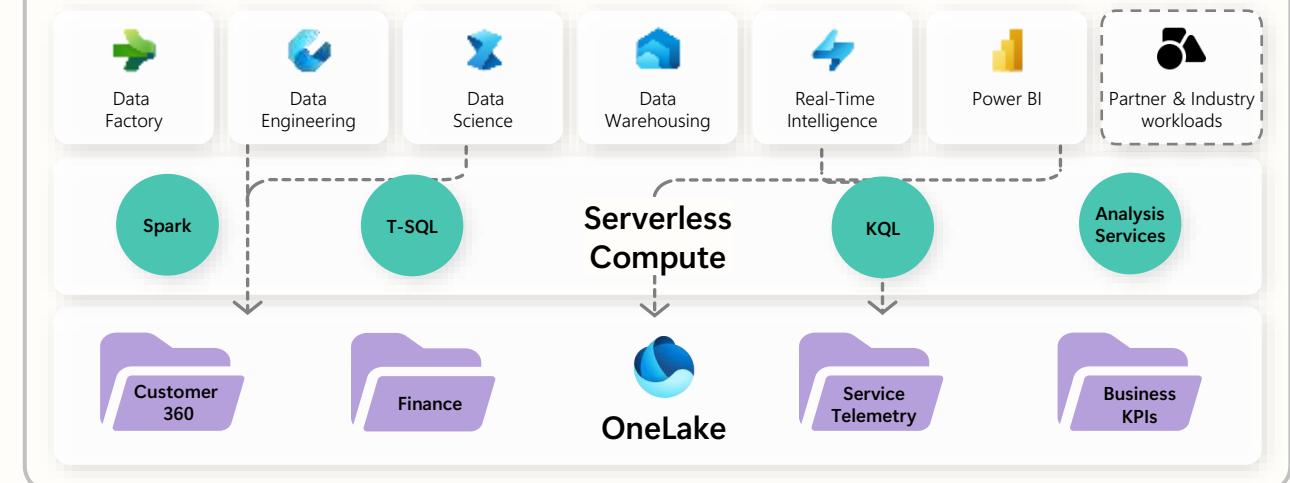
All the Fabric experiences instantly work with the OneLake replica.

Analysts and Data Scientists can work with real-time data.

The replica protects operational databases from analytical queries.

Fabric compute engines

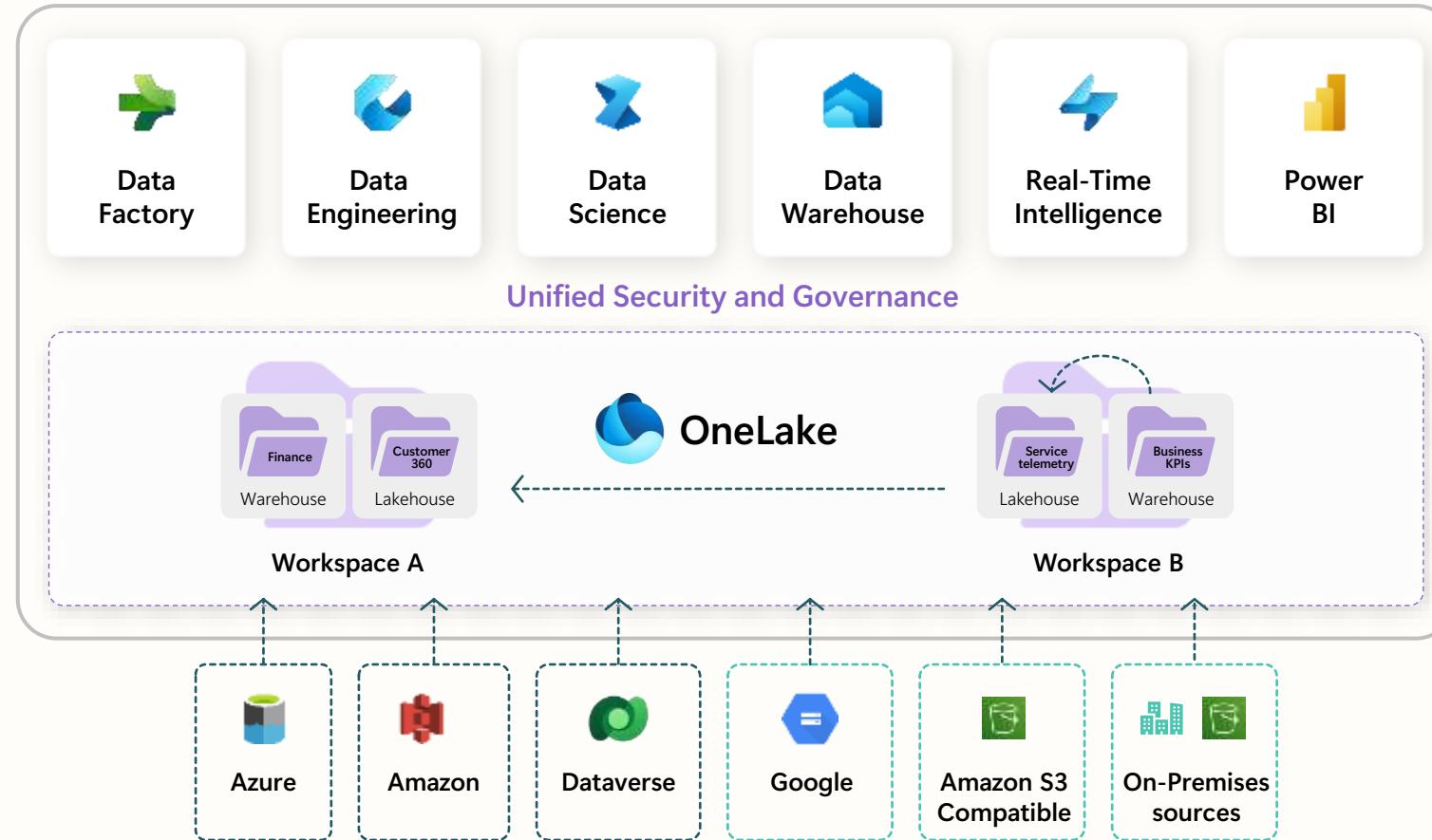
Microsoft Fabric



Mirroring

Shortcuts virtualize data across domains and clouds

No data movements or duplication



A shortcut is a symbolic link which points from one data location to another.

Create a shortcut to make data from a warehouse part of your lakehouse.

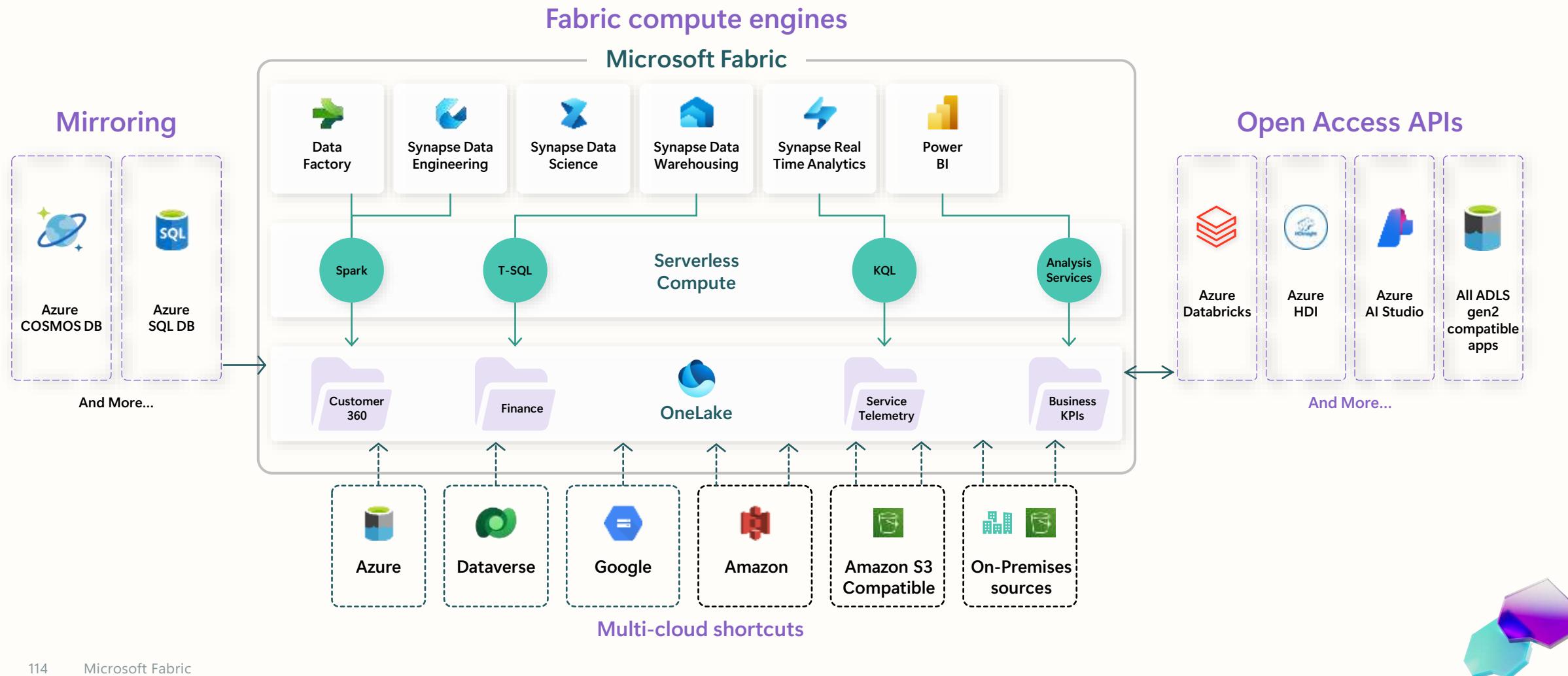
Create a shortcut within Fabric to consolidate data across items or workspaces without changing the ownership of the data. Data can be reused multiple times without data duplication.

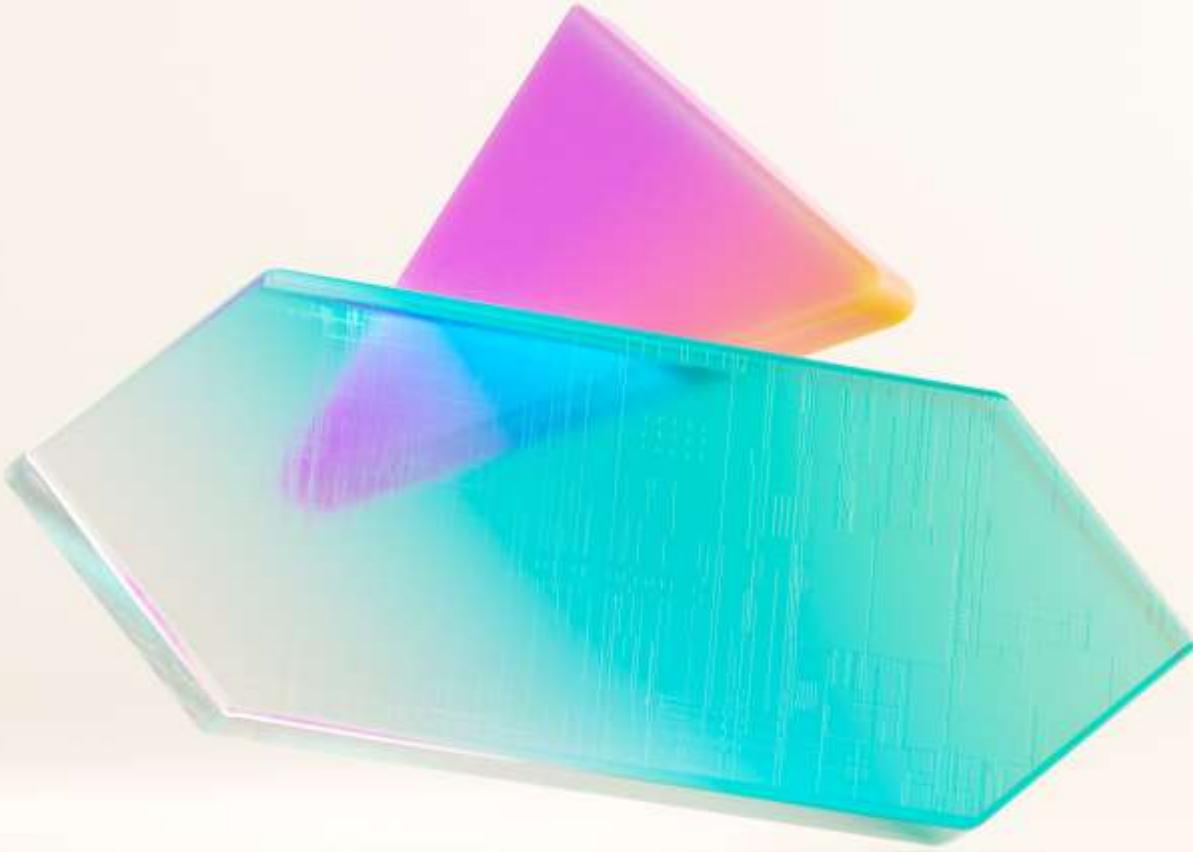
Existing ADLS Gen2 storage accounts and Amazon S3 buckets can be managed externally to Fabric and Microsoft while still being virtualized into OneLake with shortcuts.

All data is mapped to a unified namespace and can be accessed using the same APIs including the ADLS Gen2 DFS APIs.

All roads lead to OneLake

Creating Data Gravity in OneLake





Security,
governance, and
admin

Manage, govern, and secure data in Microsoft Fabric



Manage your data estate

Create a logical data mesh with domains to simplify management and increase data usage

Gain fine-grained control over your tenant, capacity, and workspace settings



Secure, protect & comply

Secure inbound and outbound network connections

Ensure only the right people have access internally

Maintain compliance with the strictest requirements



Encourage data discovery, trust, and use

Manage, catalog, and discover data in a logical, searchable data hub

Endorse and enrich high-quality data to create & promote curated, discoverable sources of truth

View the lineage for data items to identify origin and find other relevant sources



Monitor, uncover insights, and act

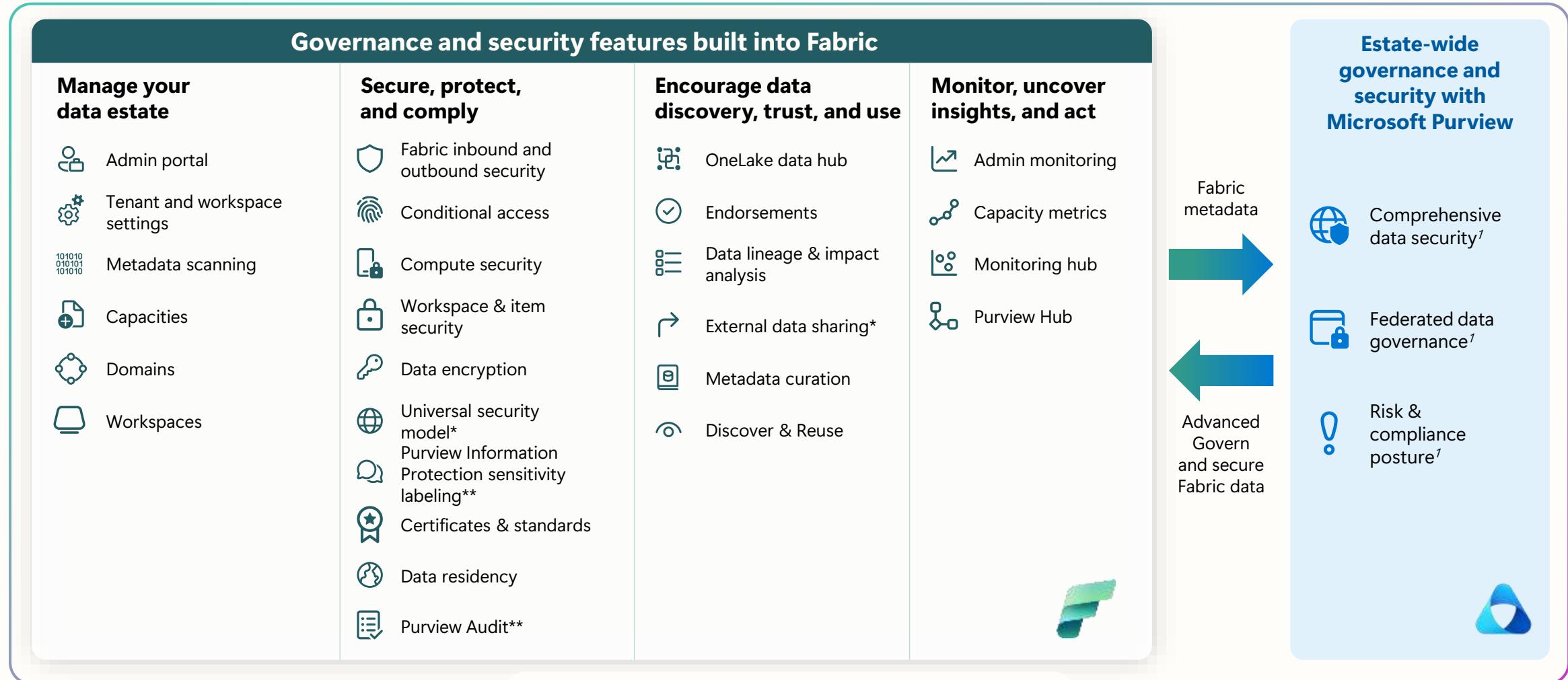
Track metrics important to admins such as usage, billing, and access

See relevant insights about sensitive data and item endorsement

Give users the ability to closely monitor their Fabric activities



Gain full visibility and control with industry-leading features



**Additional Microsoft Purview purchase required



Create a logical data mesh to simplify management and increase usage



Capacities

Easily manage multiple capacities in a single tenant to [better manage cost](#)



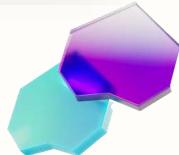
Domains and subdomains

Group data and workspaces together into a domain and sub domains to establish a [data mesh architecture](#)

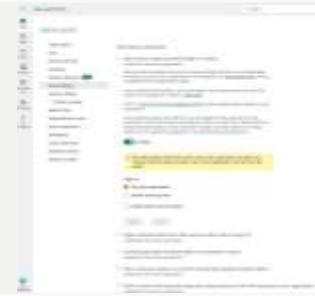


Workspaces

Group items like Lakehouses, warehouses, and reports into workspaces to organize and more easily manage items together



Gain fine-grained control over settings across your tenant



Admin portal

Centrally manage, review and apply **tenant, domain and capacity settings** for the entire Fabric tenant



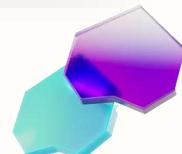
Tenant and workspace settings

Gain **fine-grained control** over the features that are made available to your organization and **govern workspaces, domains, and capacities** across your tenant



Connect to external tools

Easily extract information such as item name, owner, domain id, lineage, sensitivity label, and endorsement status for all items across Fabric, and connect to your 3rd party or homegrown tools



Secure inbound and outbound network connections



Fabric inbound security

Seamlessly secure **inbound** data access scenarios with Fabric's holistic identity and network security



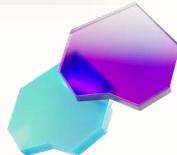
Service tags

Minimize the complexity of updating network security rules using Azure service tags to **group and manage IP addresses** for a service

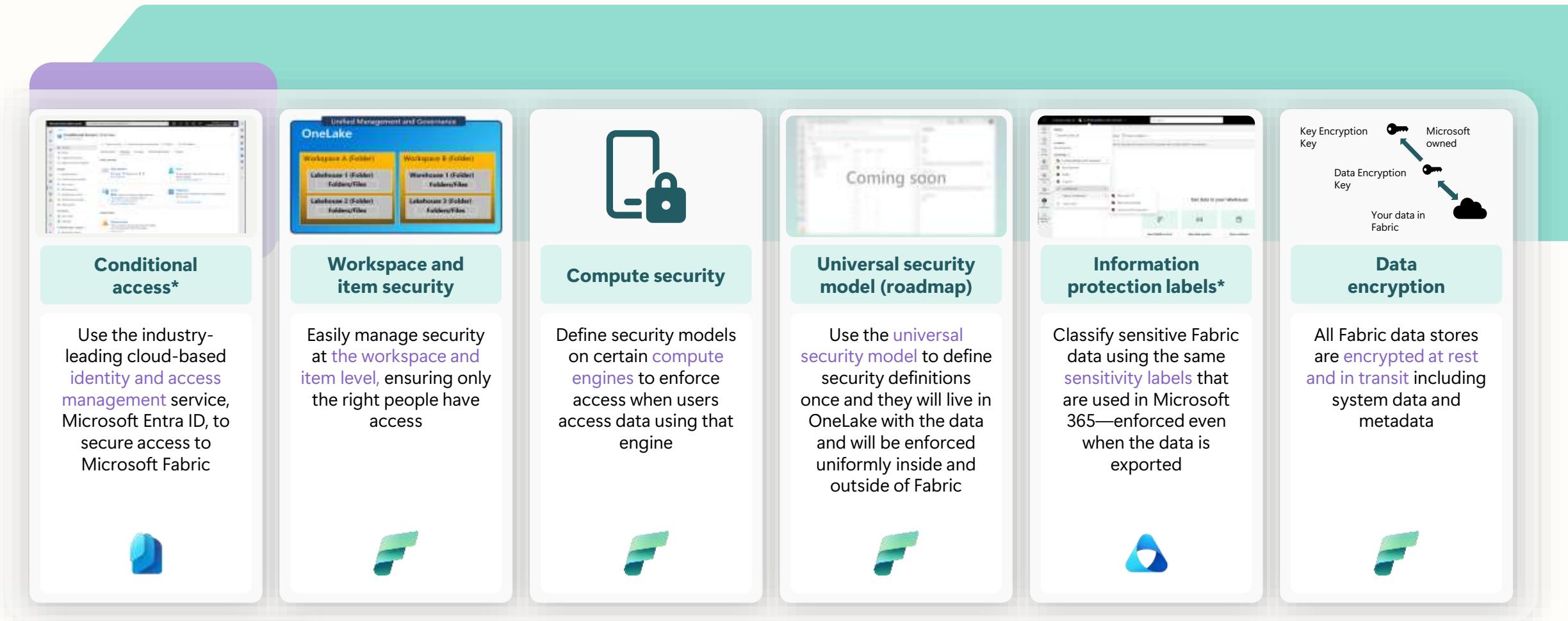


Fabric outbound security

Securely **connect to your protected data sources** and/or ingest data into OneLake



Ensure only the right people have access to the right data



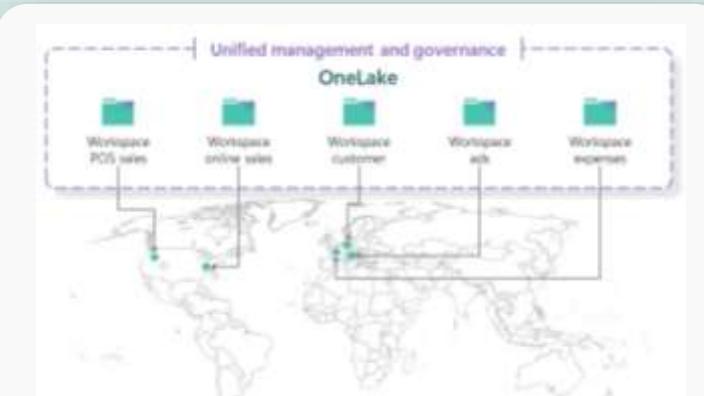
*Additional purchase required

Remain in compliance with even the strictest requirements



Certifications and standards

Microsoft Fabric, as a core online service, supports a wide range of compliance standards including GDPR, EUDB, HIPAA, ISO certifications and more



Data residency

Control where your data is stored and meet data residency requirements with over 54 data centers world-wide



A screenshot of a Microsoft Purview Audit interface. It shows a search bar at the top with filters for 'Audit' and 'Audit log'. Below is a table with columns for 'Audit log ID', 'Audit type', 'Audit status', and 'Audit date'. The table contains several rows of audit log entries.

Microsoft Purview Audit*

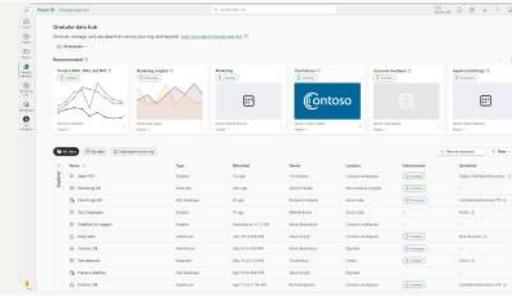
Log user activities from Microsoft Fabric to support security, forensic, and internal investigations



*Additional Microsoft Purview purchase required

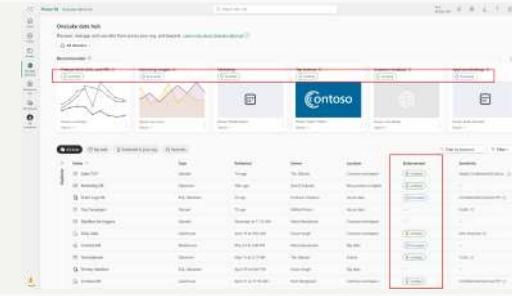


Effectively manage your data to create a trusted data hub



OneLake data hub

Find, explore, and [use](#) the Fabric data items in your organization that you have access to in one, central hub



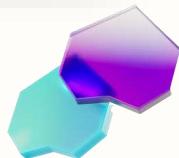
Endorsements

Endorse or mark items as discoverable high-quality content in Fabric to [create curated sources of truth](#) and increase discoverability and reuse



Data lineage and impact analysis

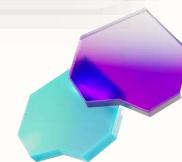
See lineage view of analytical projects to [see how data flows](#) through items and perform impact analysis to [assess impact of changes](#)



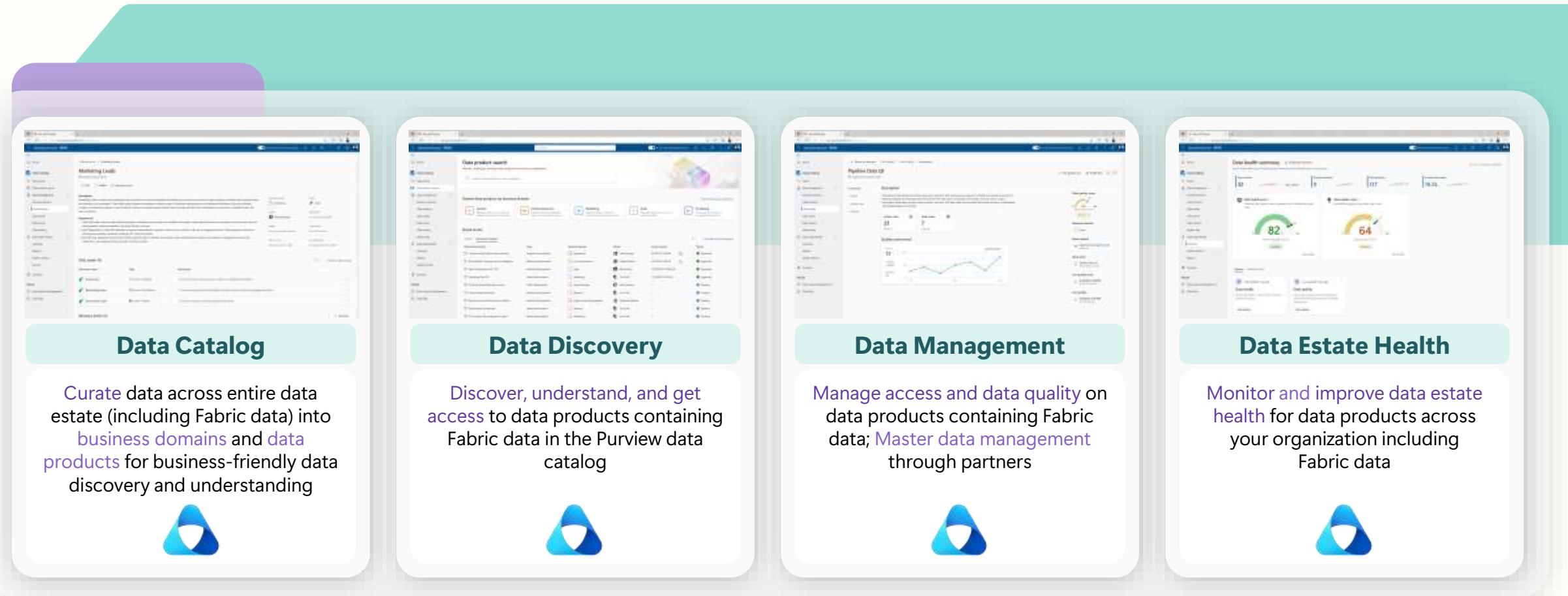
Easily monitor activity across your tenant and uncover relevant insights

The image displays four distinct Microsoft Fabric monitoring dashboards, each represented by a white card with a rounded bottom-right corner. Each card contains a screenshot of the dashboard, a title, and a descriptive text block.

- Admin monitoring**: Shows a dashboard with two main sections: "Fabric usage and Adoption / Activity Overview" and "Fabric Metrics". The first section includes a bar chart for "Total active users" and a table for "Active users by location". The second section includes a bar chart for "Total active datasets" and a table for "Active datasets by location".
- Monitoring hub**: Shows a dashboard titled "Fabric metrics" with a sidebar for "Fabric center". It includes sections for "Fabric usage", "Fabric metrics", and "Fabric logs". A red box highlights the "Fabric metrics" section.
- Capacity metrics**: Shows a dashboard with two main sections: "Fabric capacity and usage" and "Fabric metrics". The first section includes a line chart for "Fabric capacity utilization" and a table for "Fabric capacity utilization by location". The second section includes a line chart for "Fabric usage trends" and a table for "Fabric usage trends by location".
- Purview hub**: Shows a dashboard titled "Microsoft Purview hub summary" with sections for "Data governance", "Data protection", "Data lineage", and "Data quality". It includes a circular progress bar for "Data governance" and a table for "Data governance by location".



Govern entire data estate with Purview Data Governance

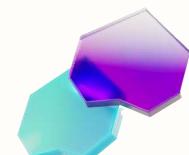


The slide displays four mobile device screens, each representing a different feature of Purview Data Governance:

- Data Catalog**: Curate data across entire data estate (including Fabric data) into business domains and data products for business-friendly data discovery and understanding.
- Data Discovery**: Discover, understand, and get access to data products containing Fabric data in the Purview data catalog.
- Data Management**: Manage access and data quality on data products containing Fabric data; Master data management through partners.
- Data Estate Health**: Monitor and improve data estate health for data products across your organization including Fabric data.

Each card features a blue triangle icon at the bottom.

Additional Microsoft Purview purchase required



API for GraphQL in Fabric

Bridging the gap between data and applications

SaaS Experience

Quick creation
Automated schema
Queries, mutations, relationships
Intelligent editor

Fabric data sources

Lakehouse
Warehouse
Mirrored databases

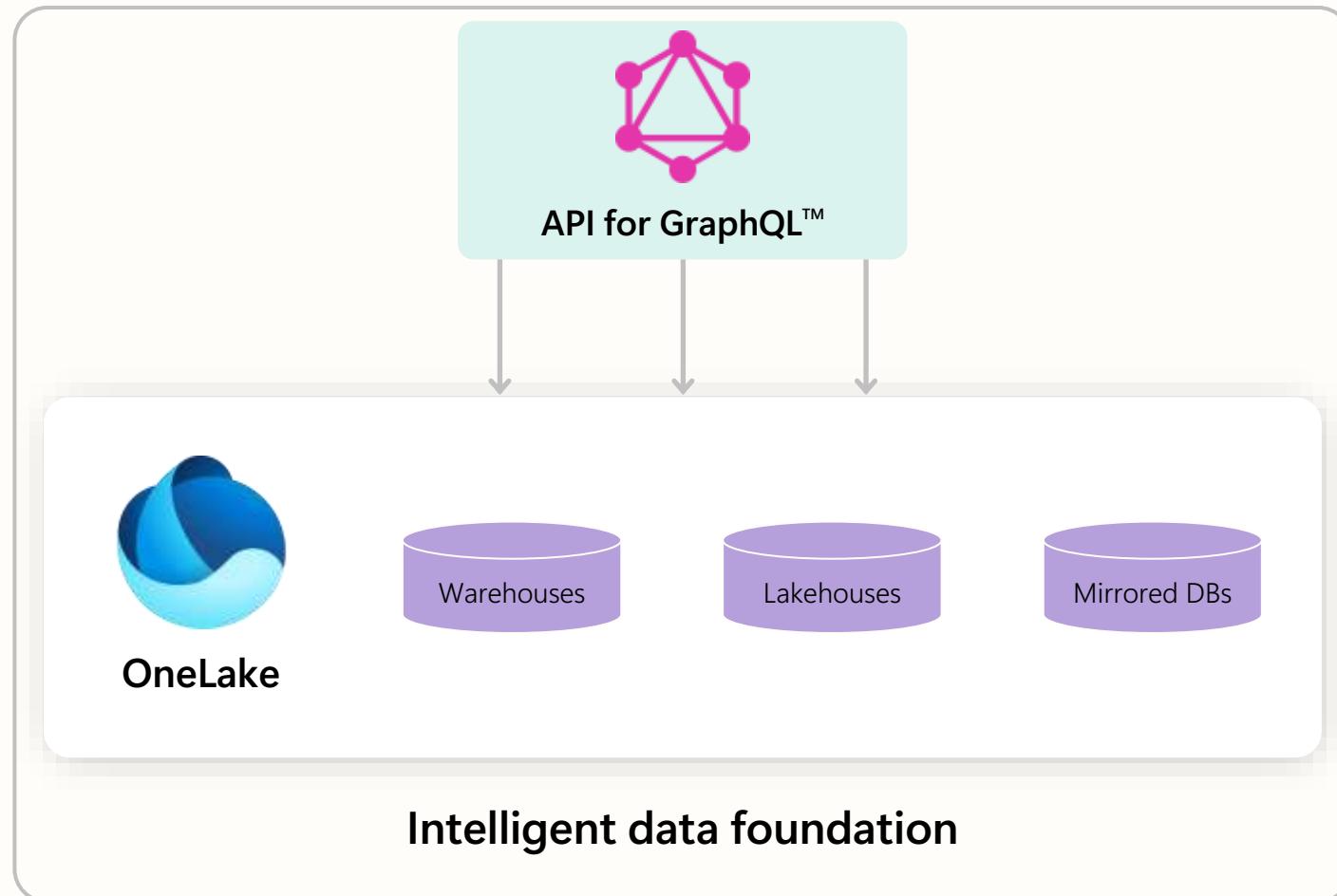
Fabric native

Passthrough auth
Workspace permissions
Lineage, endorsements
Capacity consumption

Documentation

API for GraphQL in Fabric

"The API for Accessing Fabric Data"



Consistent semantic layer API schema
abstracts database complexity

Single intelligent endpoint for all data

Flexible response structure for
fine-grained data retrieval

Simple, fast SaaS-ified experience

Fabric native: Security, governance, capacity
consumption

Takeaways

User Data Functions in Fabric

The “UDF” of Fabric

SaaS Experience

Seamless experience
Simple programming model
VSCode integration

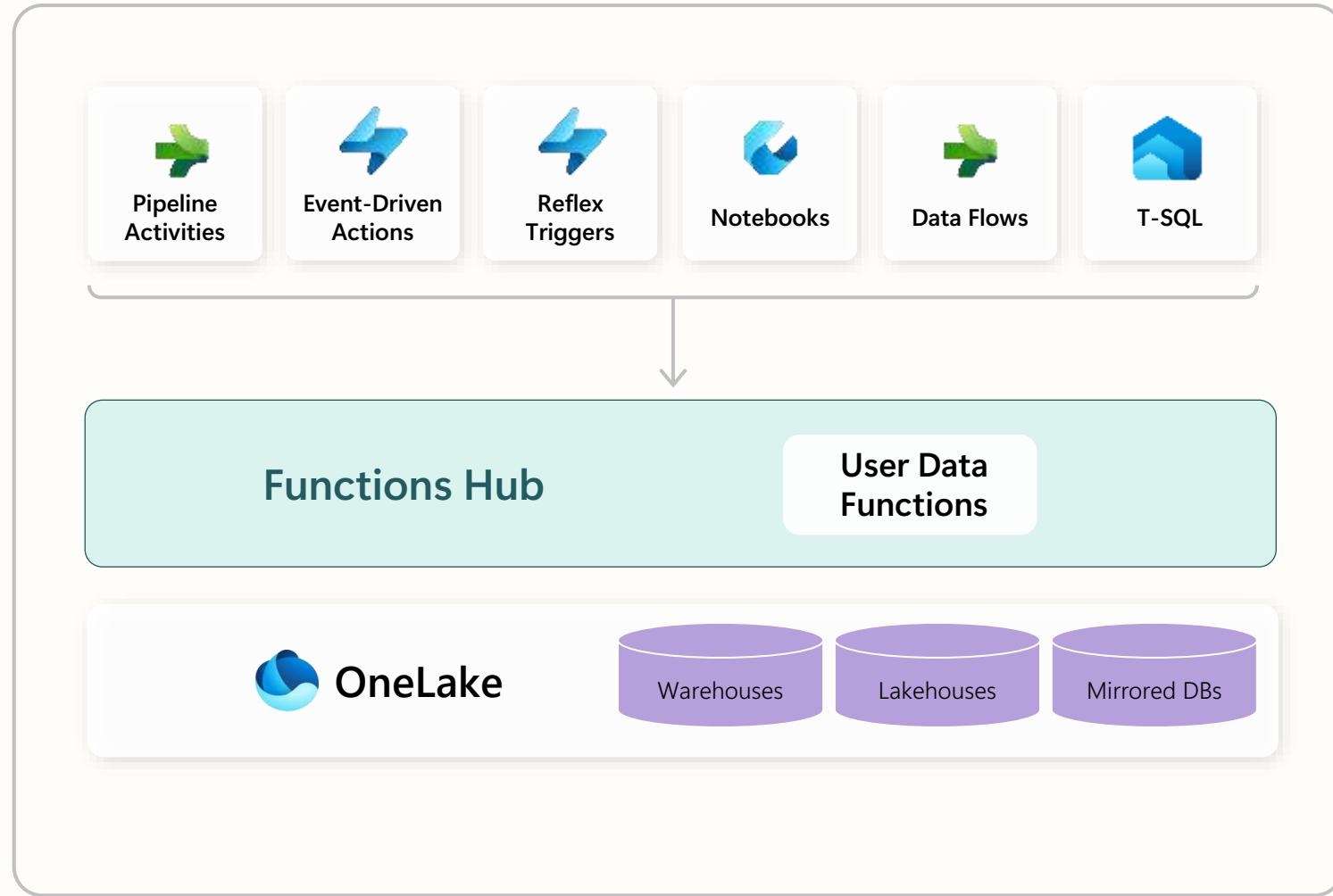
Fabric data sources

Lakehouse
Warehouse
Mirrored databases

Fabric native

Invokable from many Fabric items

User Data Functions in Microsoft Fabric



Invocable from many Fabric items

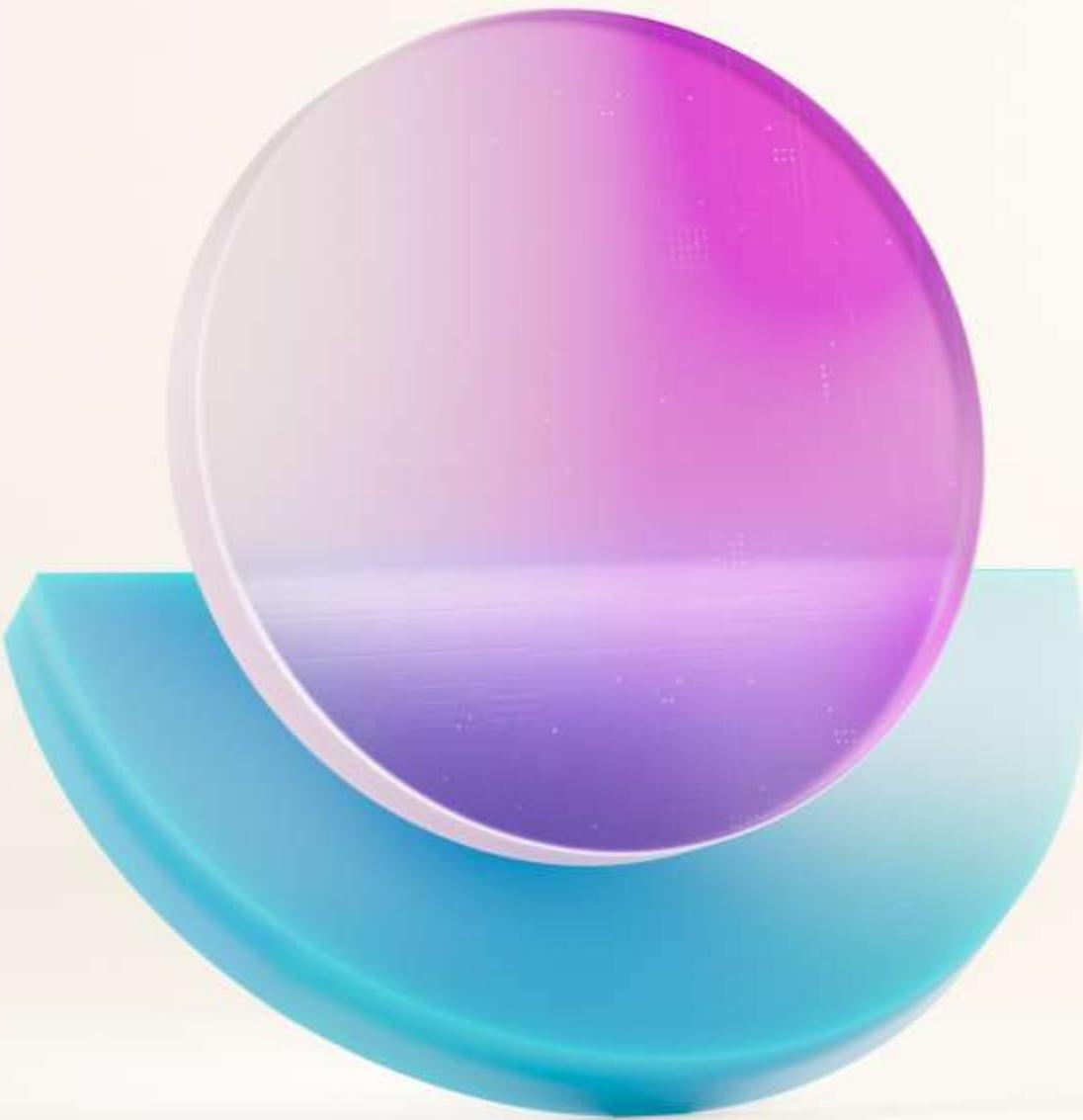
Simple programming model

Access to Fabric data sources

Developer-friendly experience

Reusable, discoverable via Hub

Security, governance, capacity consumption



Partner and
industry solutions
and integrations



Microsoft Fabric



Data
Factory



Data
Engineering



Data
Warehouse



Data
Science



Real-Time
Intelligence



Power BI



Partner &
Industry
workloads



Copilot in Fabric



OneLake



Microsoft Purview



Public Preview

Microsoft Fabric Workload Development kit

Expand, integrate, and scale your applications natively in Fabric and its ecosystem



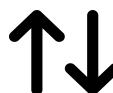
Accelerated
innovation



Consistent
experiences



Seamless
development



Effortless
integration

Introducing Workload Hub



Workload Hub: Where new capabilities enter Fabric



Allows users to start using your product within the Fabric environment

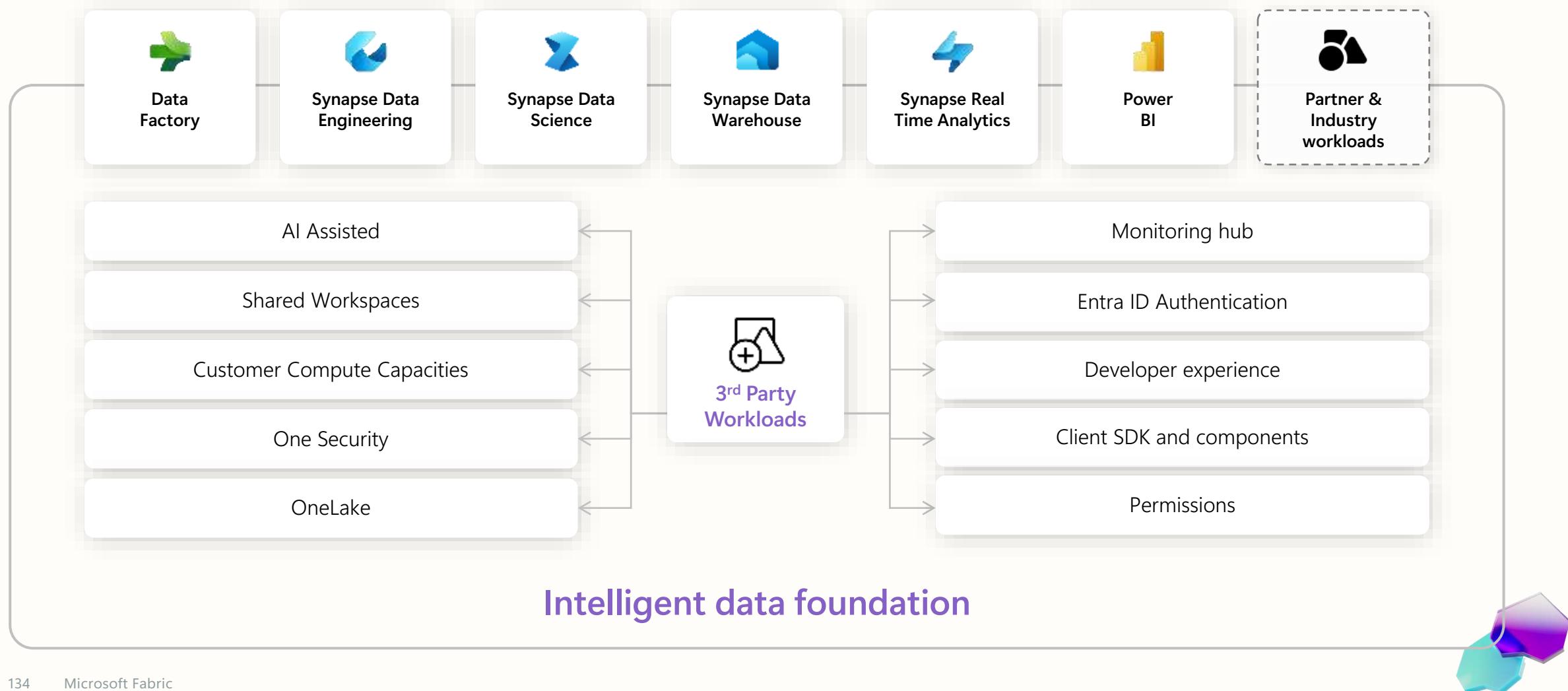


Seamless integration to improve workflow and productivity

The screenshot shows the Microsoft Fabric Workload Hub interface. On the left, there's a sidebar with navigation icons and a search bar. The main area displays the "OneLake Data Ingestion and Transformation Suite" workload. It includes a brief description, a large image of a server room, and a smaller image of a person working at a computer. Below the main image is a detailed description of the workload's features and benefits, such as "OneLake Data Ingestion: Simplify the import of diverse data into OneLake with no ETL required. OneLake is a fast, reliable, and cost-effective way to store and process data from various sources." To the right of the main workload card, there's a grid of other available workloads, each with a thumbnail, name, and a brief description.



Leverage Fabric workload services

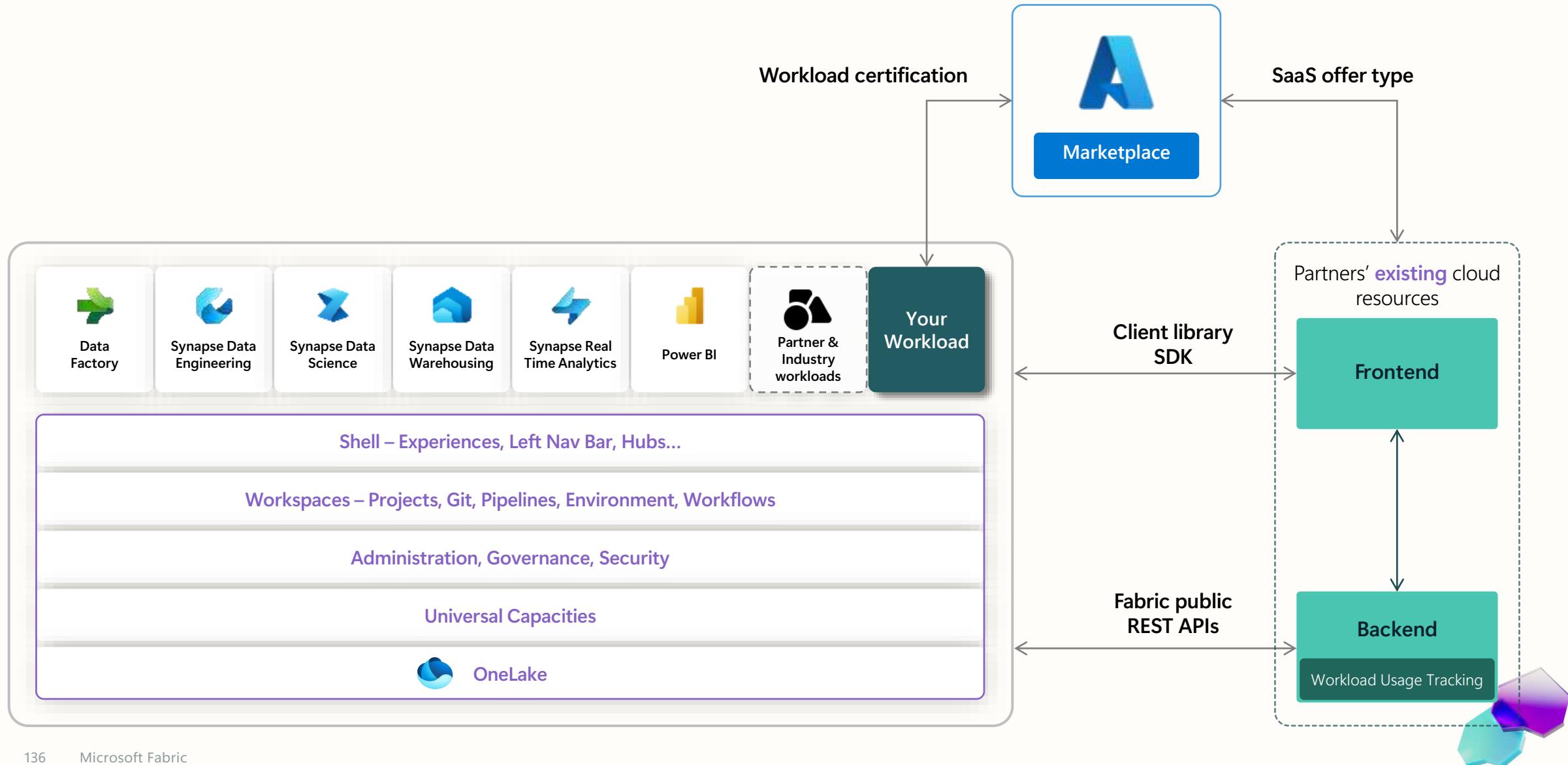


Native Fabric workload experience

Unmatched integration flexibility



Monetize via Azure Marketplace



3 Steps to release a Fabric workload



Develop locally

- Use the Workload Developer Kit.
- Interface with Fabric REST APIs.



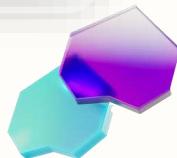
Test with Fabric

- Upload the manifest to the Fabric tenant.



Publish to Fabric

- Via Azure Marketplace.
- Through Partner Center.



Developer experience live demo

Develop a workload – Live code

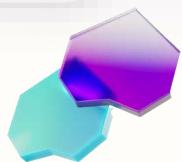
The image displays two side-by-side screenshots of the Microsoft Fabric developer interface, specifically the 'Workloads' section. Both screenshots show a grid of 12 workload cards, each with a title, icon, and a brief description.

Left Screenshot Workload Cards:

- Data Analytics**: Microsoft. Manages data to engage users and automated insights from AI and machine learning.
- Data Development**: Microsoft. Automates data pipelines using FlinkQL and serverless data flows for faster development.
- Data Engineering**: Microsoft. Creates a foundation, and can quickly build an AI-driven cloud-native data engineering stack to meet business needs.
- Data Factory**: Microsoft. Manages data integration and ETL processes with built-in governance and compliance.
- Data Science**: Microsoft. Provides a secure, fast, and hybrid environment for data science projects from your organization's confidential environments.
- Data Warehouse**: Microsoft. Helps you manage by extracting and analyzing data in a variety of sources from ETL, machine learning, big data, and more.
- Databases**: Microsoft. Manages databases with support for powerful security and access controls, along with native data.
- Frontend API playground**: Microsoft. Allows front-end teams to quickly experiment with Microsoft's latest technologies.
- Industry Solutions**: Microsoft. Provides industry-specific functionality, automation, and commerce-specific tools.
- Power BI**: Microsoft. Helps analysts, data managers, and power users quickly discover, analyze, and visualize data.
- Project Update Lakes**: Microsoft. Handles complex data integration requirements and supports real-time data processing.
- Real-Time Analytics**: Microsoft. Handles real-time data processing, and query via REST APIs and streams from 1 GB to 1 PB and more.

Right Screenshot Workload Cards:

- Cognitive Data Lake**: Microsoft. Manages data in storage, which can be used for AI and machine learning applications.
- Data Analytics**: Microsoft. Manages data to engage users and automated insights from AI and machine learning.
- Data Development**: Microsoft. Automates data pipelines using FlinkQL and serverless data flows for faster development.
- Data Engineering**: Microsoft. Creates a foundation, and can quickly build an AI-driven cloud-native data engineering stack to meet business needs.
- Data Factory**: Microsoft. Manages data integration and ETL processes with built-in governance and compliance.
- Data Science**: Microsoft. Provides a secure, fast, and hybrid environment for data science projects from your organization's confidential environments.
- Data Warehouse**: Microsoft. Helps you manage by extracting and analyzing data in a variety of sources from ETL, machine learning, big data, and more.
- Databases**: Microsoft. Manages databases with support for powerful security and access controls, along with native data.
- Frontend API playground**: Microsoft. Allows front-end teams to quickly experiment with Microsoft's latest technologies.
- Industry Solutions**: Microsoft. Provides industry-specific functionality, automation, and commerce-specific tools.
- Power BI**: Microsoft. Helps analysts, data managers, and power users quickly discover, analyze, and visualize data.
- Project Update Lakes**: Microsoft. Handles complex data integration requirements and supports real-time data processing.
- Real-Time Analytics**: Microsoft. Handles real-time data processing, and query via REST APIs and streams from 1 GB to 1 PB and more.



Just the beginning

Microsoft Fabric Workload Development Kit

				
				
				And more...

aka.ms/FabDevKit



Coming soon



Apache Iceberg

Shortcuts will automatically
translate Apache Iceberg
metadata to Delta Lake



Delta Lake

Unify your data estate with open formats support in OneLake

Expanded Partnership



Microsoft Fabric



Adobe

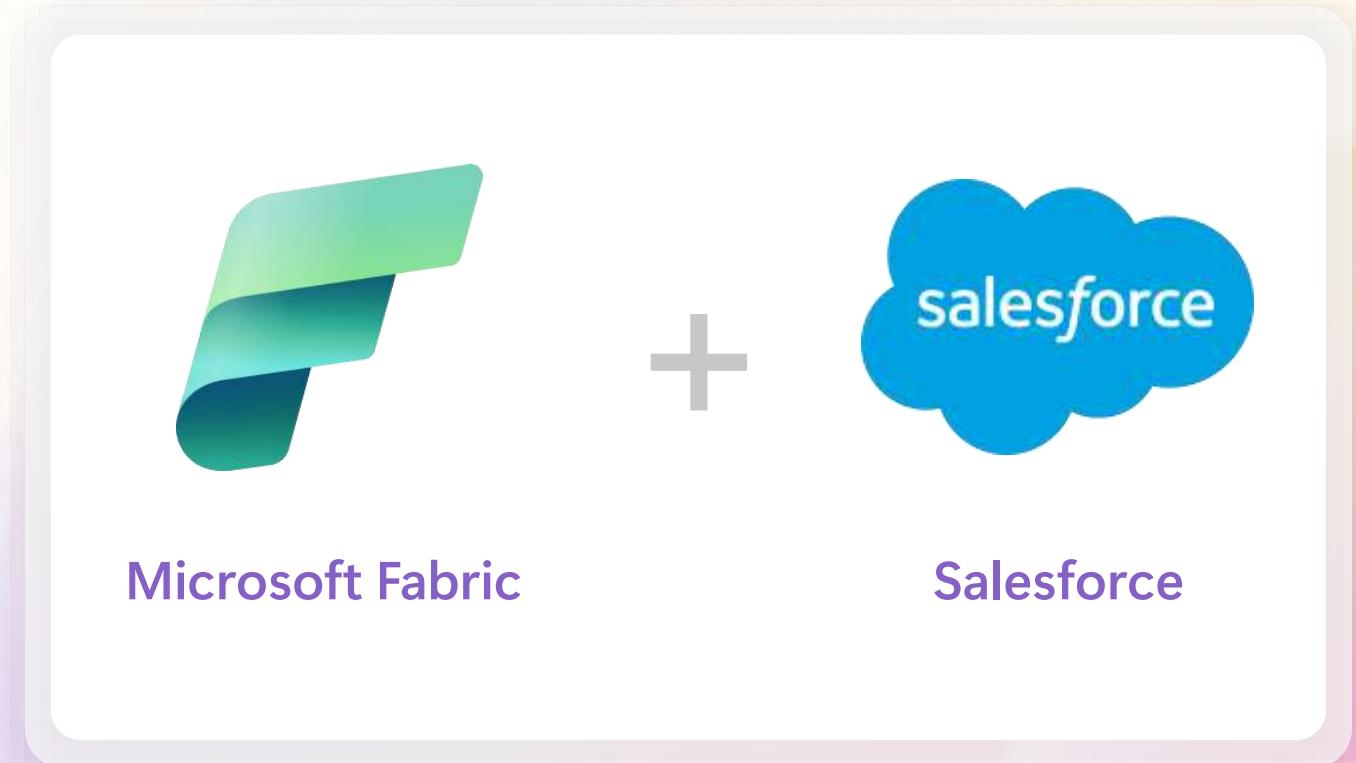
Adobe Experience Platform accesses data in OneLake via federated queries to Microsoft Fabric to create or enrich audiences for engagement.

Do more with your data with open standards

Unify your data in OneLake

Bi-directional data access

Seamless access from M365
and Copilot



Partner momentum

accenture

avanade

Capgemini

EY
Building a better working world

IBM

SONATA
SONATA SOFTWARE

tcs
TATA CONSULTANCY SERVICES

Cloud Services

SDK

MANDELBULB
TECHNOLOGIES

bakertilly

3Cloud

READYMIND
NEXT GENERATION SERVICES

iLink Digital

macaw

preludeSYS
IMPLICIT KNOWLEDGE

kanerika

Simpson Associates
The Data Analytics Company

INSPARK

QUADRANT

LANTERN

Fellowwind

OBUNGI

AdastrA

MAQ Software

ISV momentum

AVEVA

dbt Labs

Delphix

Epic

esri

Fivetran

Informatica

LSEG

mongoDB

Profisee

Qlik

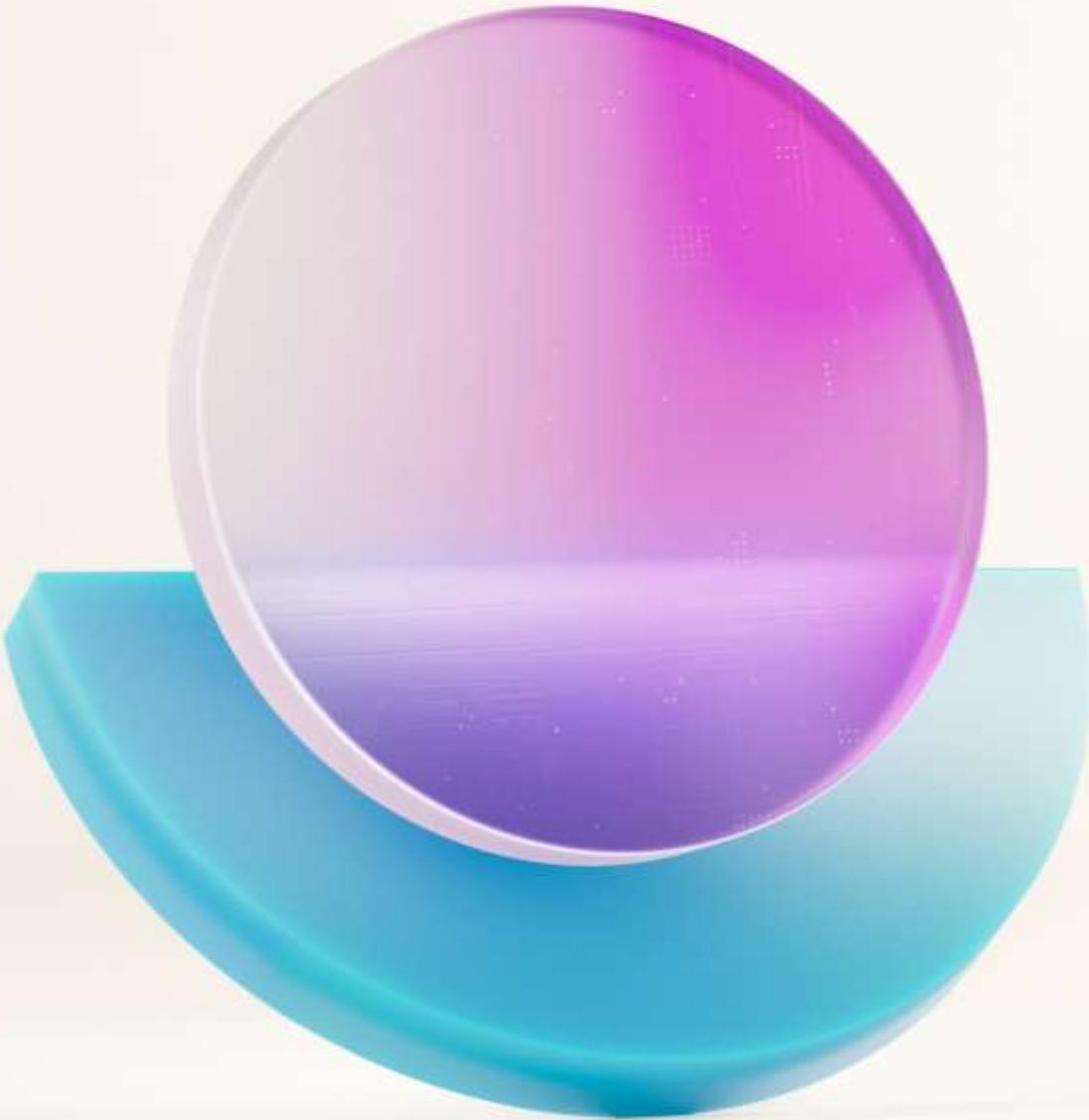
sas

striim

teradata.

yugabyteDB

+100s of partners

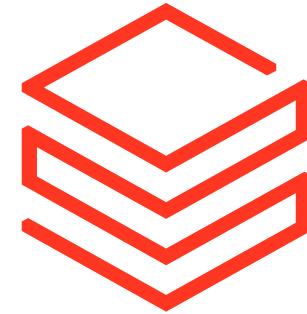


Microsoft Fabric +
Azure Databricks

Accelerate innovation on Microsoft platforms



Microsoft Fabric



Azure Databricks

Open and Governed Lakehouse



Microsoft Purview



Azure AI Studio



Azure Machine Learning



Azure AI Portfolio



Power Platforms



Copilot



Azure Cosmos DB



Azure SQL

Microsoft Fabric and Azure Databricks

DREAM demo architecture and integration

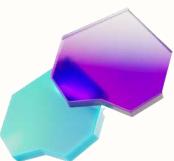


Microsoft Fabric with Azure Databricks DREAM Demo

- Click-by-click demo [here](#)
- Detailed demo video [here](#)
- All demo resources [here](#)

Modern Analytics with Azure Databricks DREAM demo

- Detailed demo video [here](#)
- Click-by-click demo [here](#)



Data transformation



Azure Databricks can be used as the compute engine across all layers of the medallion structure



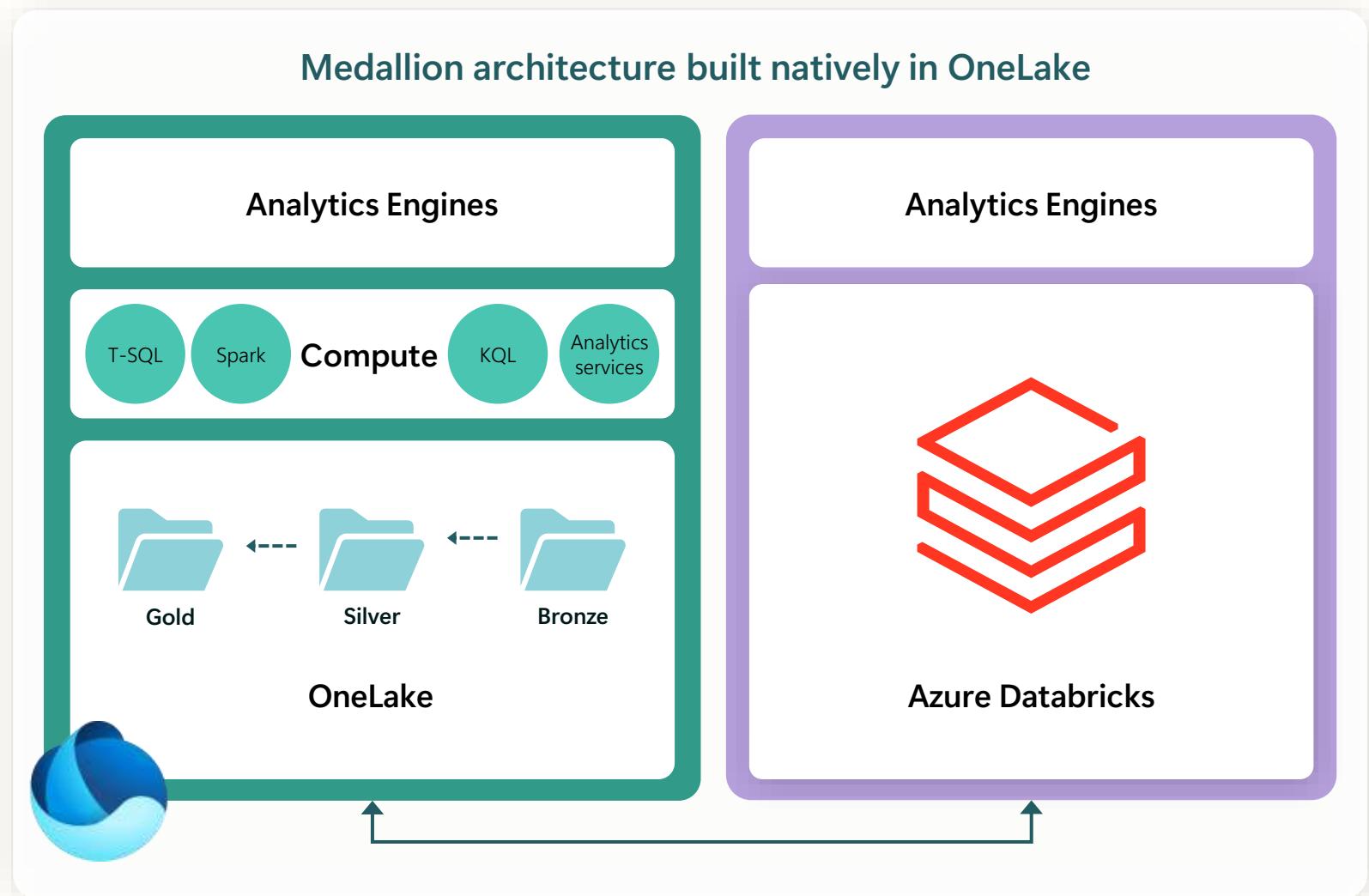
This single copy of data appears natively in other data products



Data products can be served to end users through Azure Databricks or any compute engine within Fabric



Shared APIs make it easy to start using OneLake data with any application, with minimal code changes



Enriched analytics



Create shortcuts to existing data and make it ready for consumption through OneLake



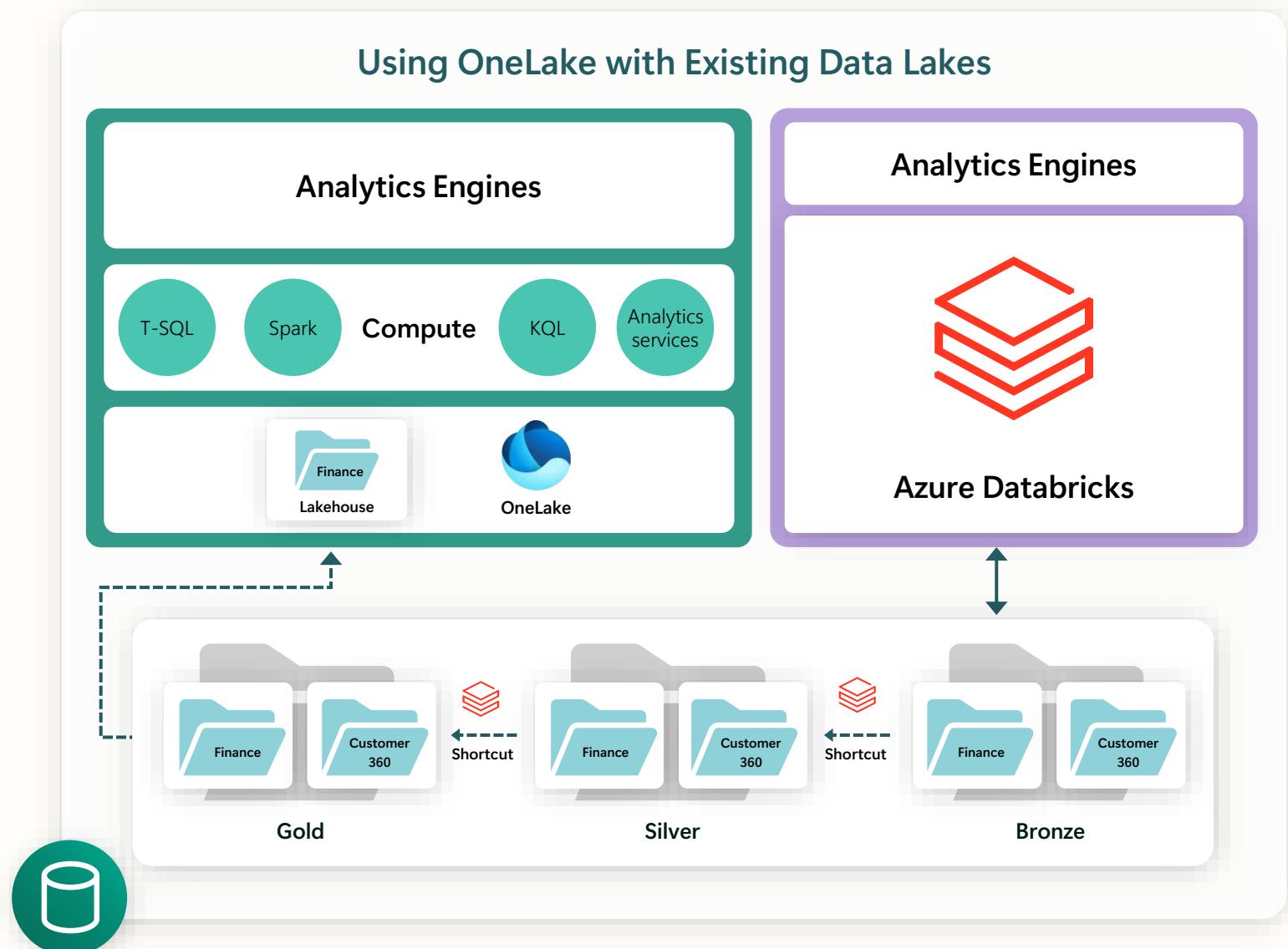
Use Direct Lake mode for blazing fast data queries in OneLake



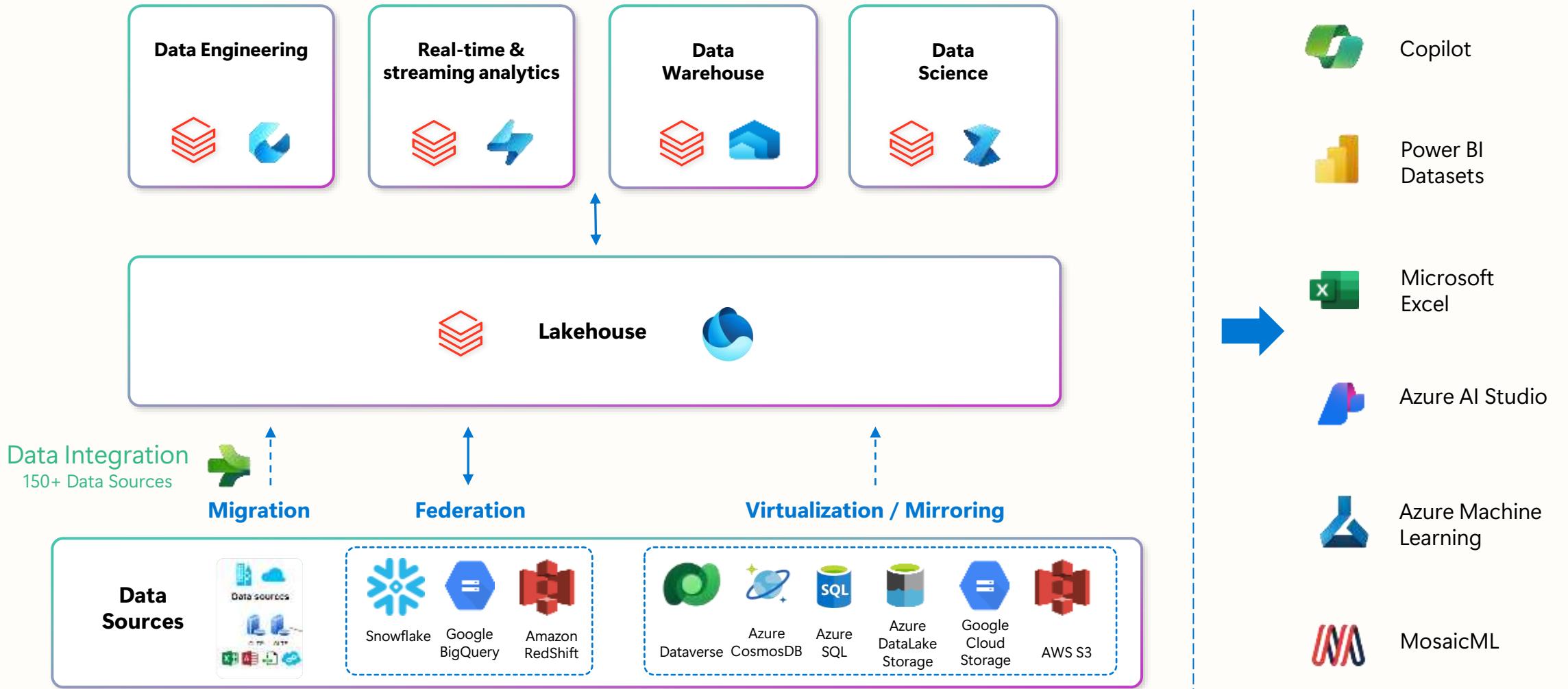
Update Azure Databricks notebooks to use OneLake endpoints for data



Keep paths consistent across experiences

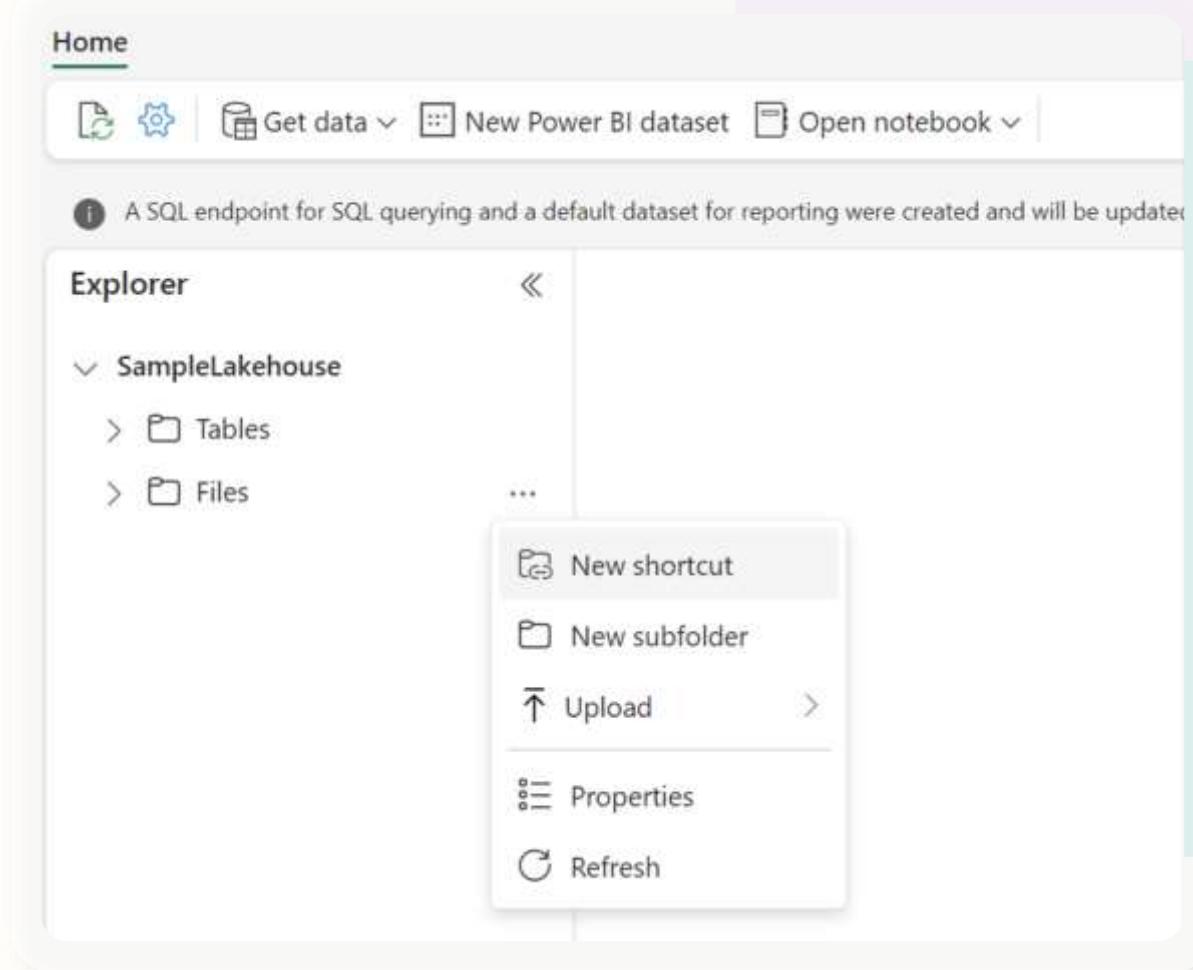


Winning the best of both worlds



Create a shortcut

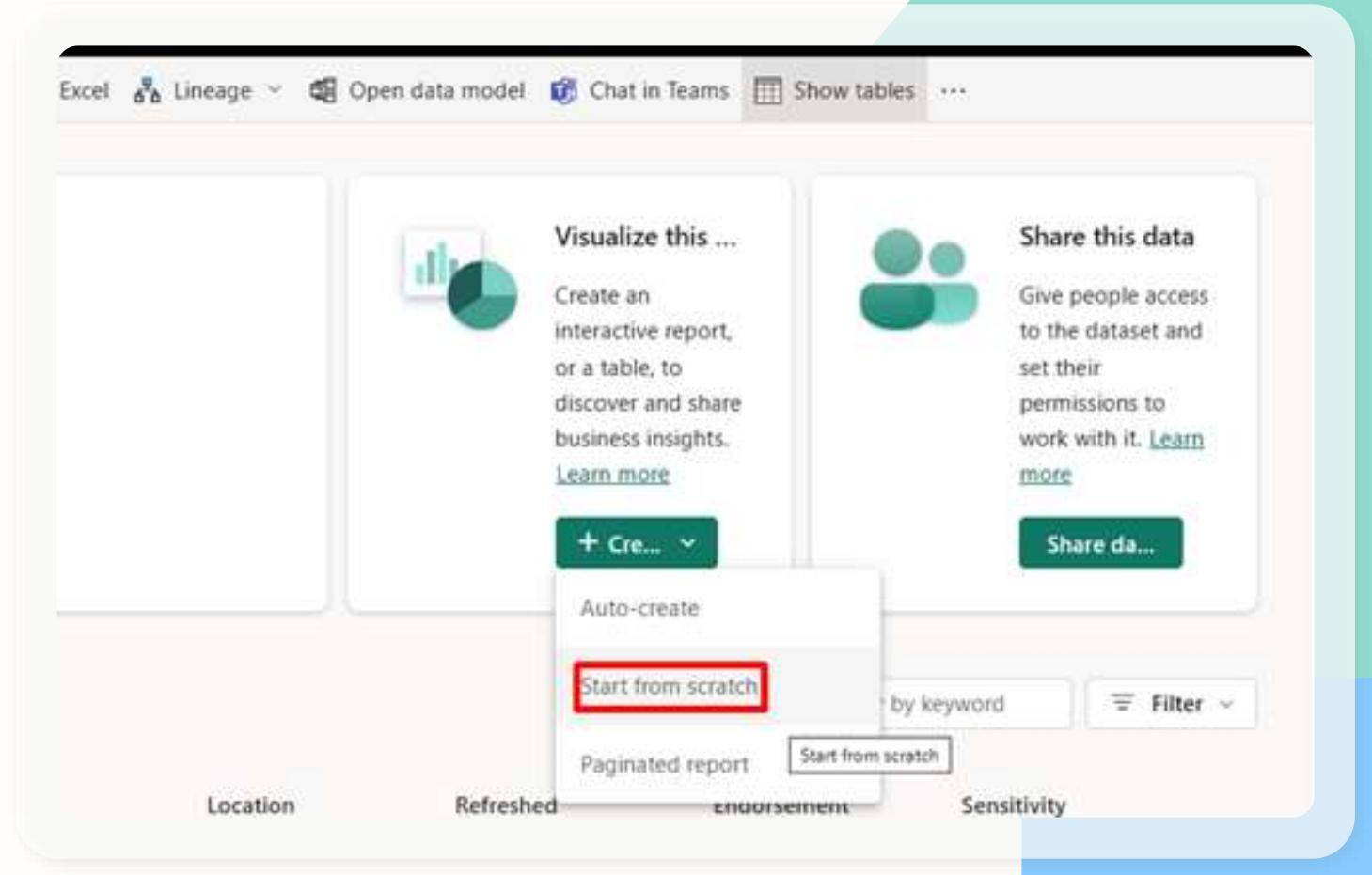
1. Create and open a lakehouse
2. Select **New Shortcut** from the menu
3. Choose Azure Data Lake Storage Gen2 (ADLS Gen2) as a source and connect
4. Name shortcut and enter sub path to desired storage folder
5. Connect



→ Learn more [here](#)

Explore your data

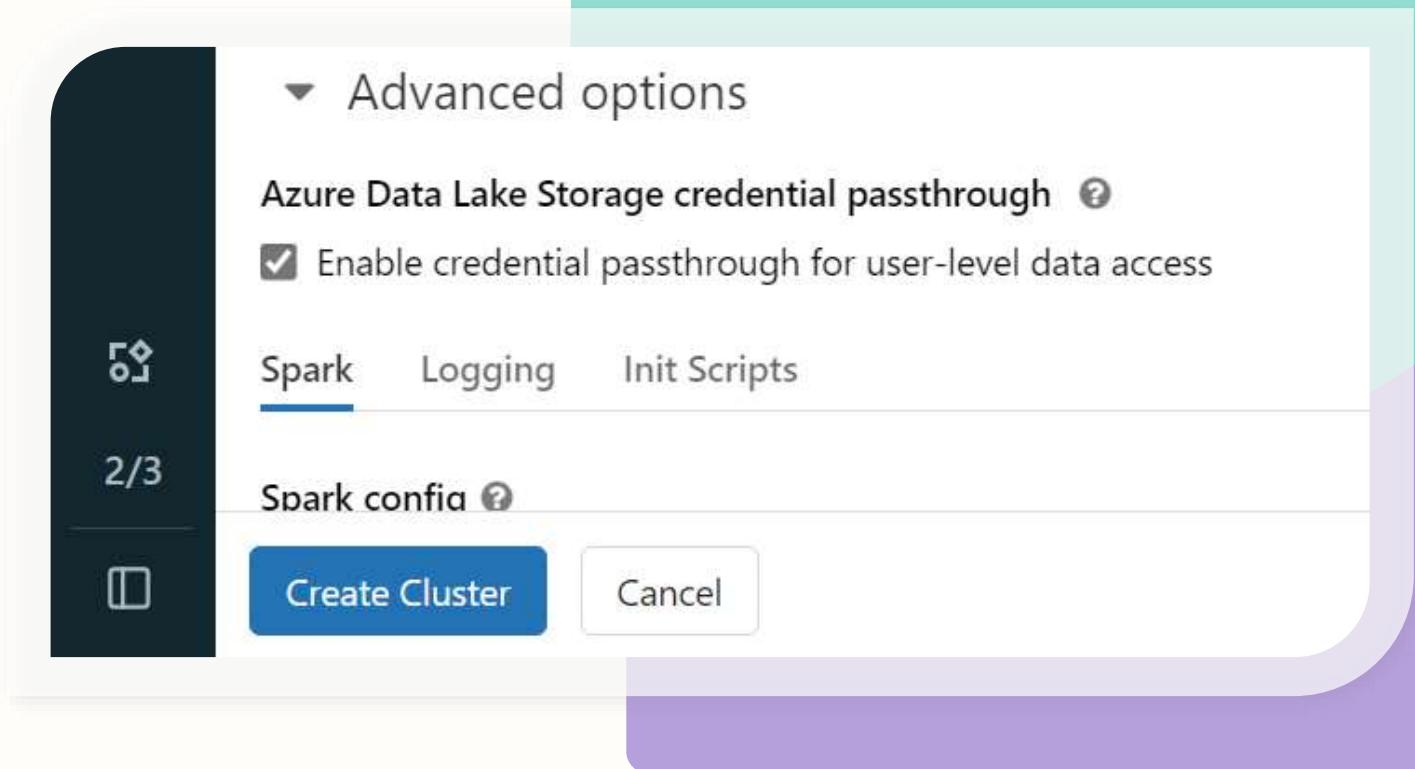
1. Open Azure Databricks notebook and create ADLS Gen2 delta table
2. Create shortcut to delta table from your Lakehouse
3. Select **New Semantic Model** and **Start from scratch**
4. Build Power BI report



→ Learn more [here](#)

Leverage Direct Lake

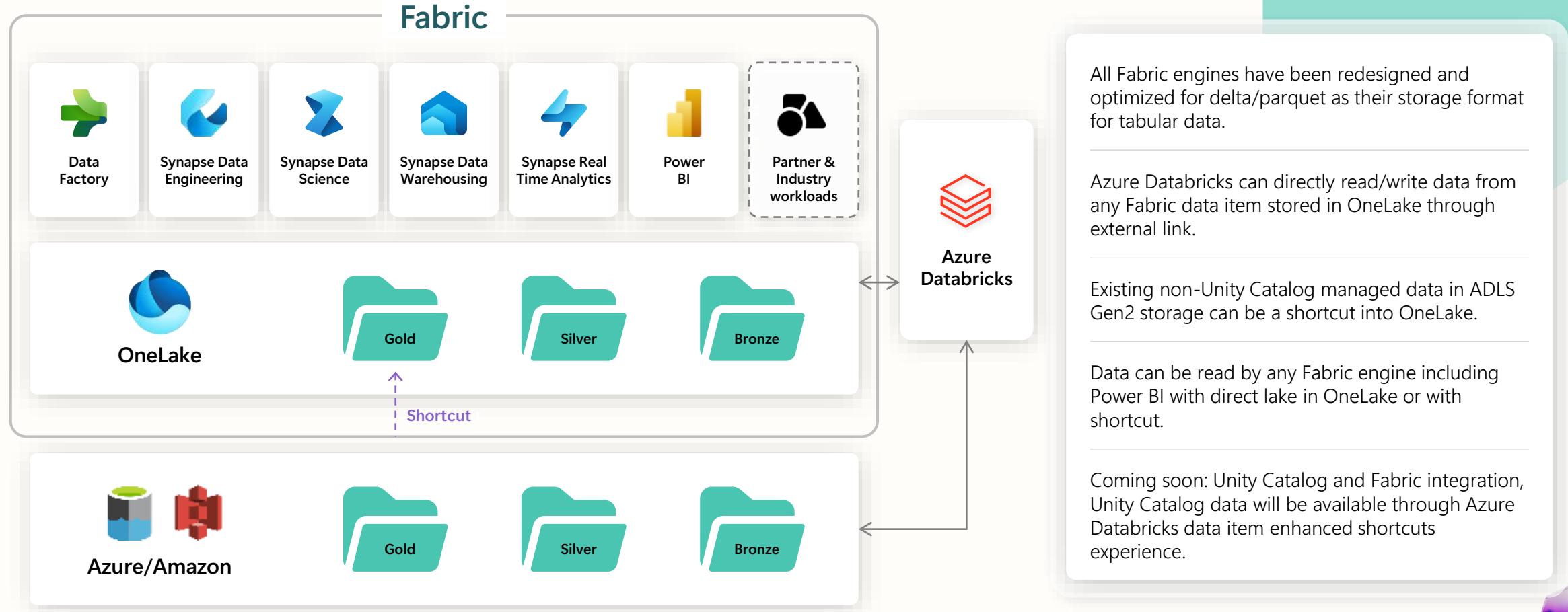
1. Create a cluster in Azure Databricks
2. Open notebook and connect to new cluster
3. Copy Azure Blob Filesystem (ABFS) path to Fabric Lakehouse Table
4. Load data into dataframe and filter as needed
5. Use OneLake path to write filtered dataframe to Fabric Lakehouse
6. Create a Power BI report targeting the data in the Lakehouse

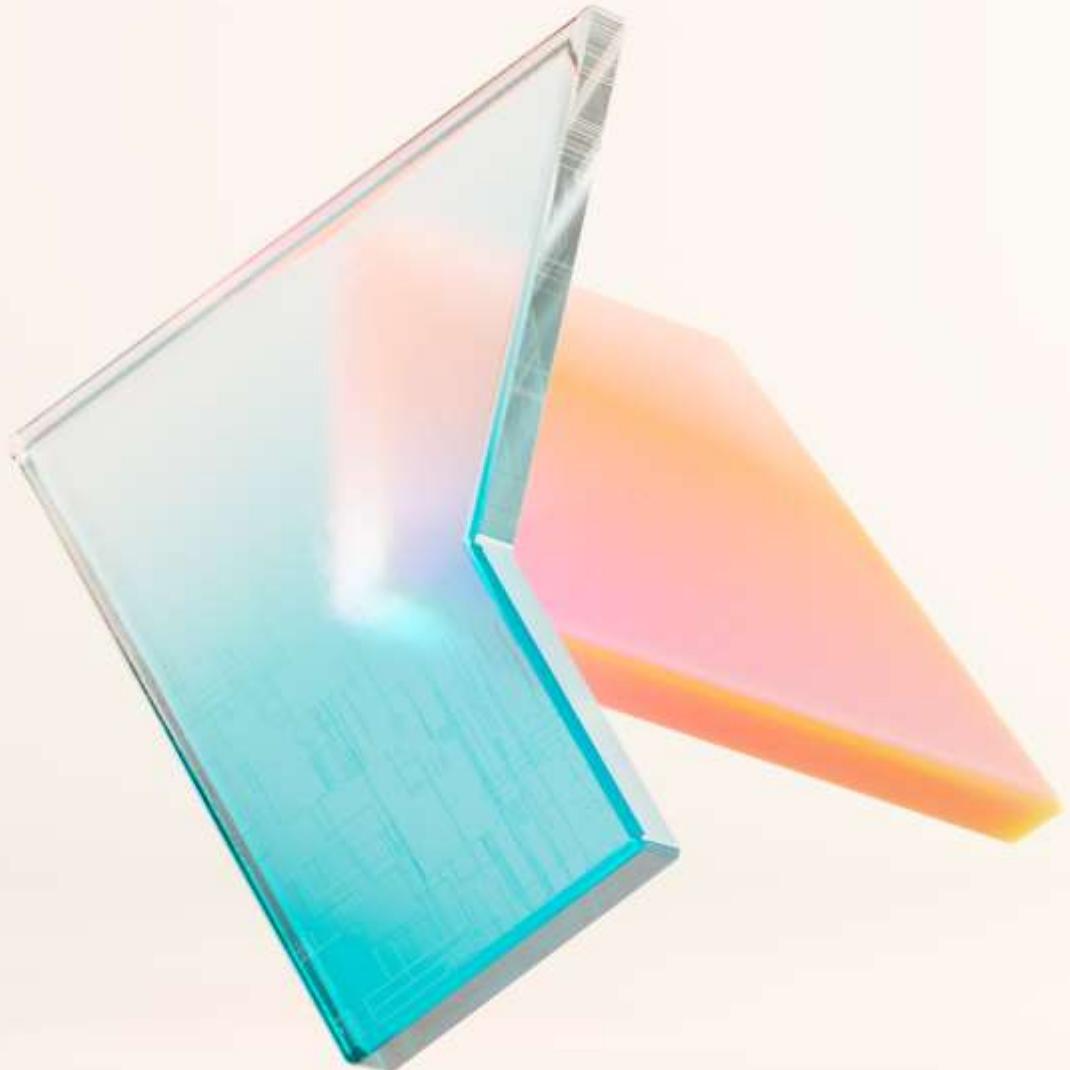


→ Learn more about OneLake integration [here](#) and Direct Lake [here](#)

Fabric and Azure Databricks

Better together through OneLake





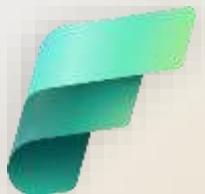
Product
integrations

Seamless integration of low code tools with unified analytics



Dataverse

The intelligent, scalable, low-code platform that manages data and business logic across Microsoft Dynamics 365 and Microsoft Power Platform.



Microsoft Fabric

The one-stop-shop for data integration, data engineering, real-time intelligence, data science, and business intelligence across all your organizational data.

Native Integration

Dataverse makes it easy to connect Power Platform and Dynamics 365 to Fabric

- No Copy. No ETL.
- Direct Connection via Dataverse
- Insights democratized to all low code apps and business using Fabrics 7 core workloads
- Makers informed by insights improves quality of applications
- Data is governed



Microsoft Power Platform



Microsoft Dataverse



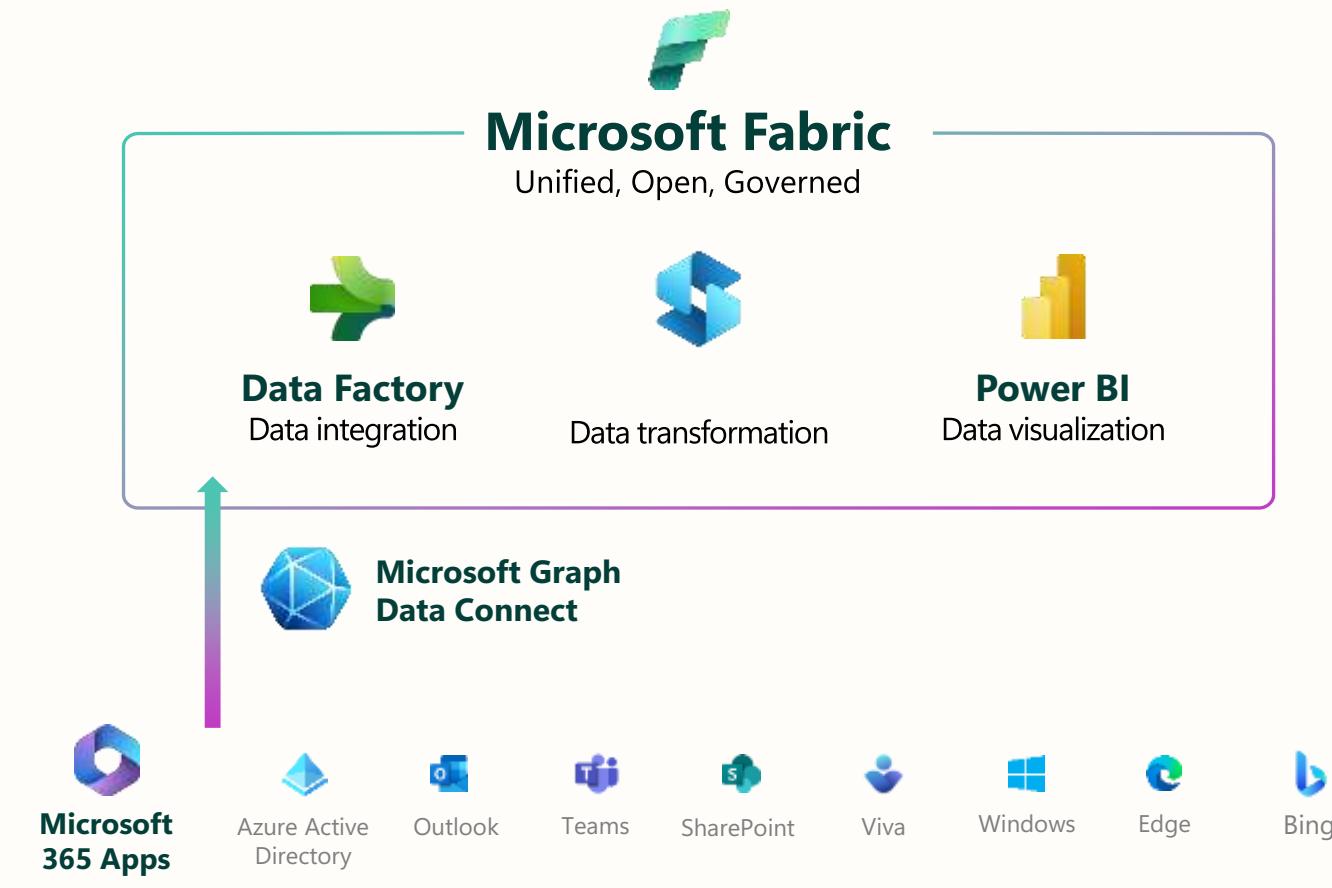
Microsoft Dynamics 365



Microsoft Fabric

Your intelligent hub for insights and democratized data

Unlock the power of your data quickly and securely with Microsoft Fabric and Microsoft 365



Seamless integration by Microsoft Fabric across Microsoft 365 and Microsoft Graph creates a single, end-to-end solution for all your data needs.

- **Empower everyone**, from data scientists to business professionals, to visualize, monitor and act on data
- **Draw intelligence** from Microsoft 365 data you already have in Outlook, Teams, SharePoint and more
- **Responsibly democratize** business intelligence with self-service tools and experiences
- **Provide an open and governed data foundation** with OneLake to eliminate siloes and optimize costs
- **Accelerate time to insights** with responsible ML and AI
- **Securely scale discovery** and access of analytics and data with built-in governance, powered by Microsoft Purview

Expanding list of Microsoft 365 datasets and metrics

Available 2023

Azure Active Directory	Outlook	Teams	SharePoint	Viva	Office 365
User profile	Messages, sent	Meeting chats	Doc sharing ¹	VI person report ^{1,2}	Meeting activity
Manager info	Contact	1:1 and 1:n chat	SharePoint groups ¹	VE messages ^{1,2}	Email activity
Direct reports	Calendar view	Std. channel msg.	Site usage ¹		Office app activity
Group members	Mailbox setting	Channel details	File usage		Teams activity
Group details	Mail folder	Call records	File activity		
Group owners	Event	Transcripts			
	Group message				
	Inbox message				
	To Do lists/tasks				
	Conference rooms				

Beyond



Govern your entire data estate with Fabric and Microsoft Purview



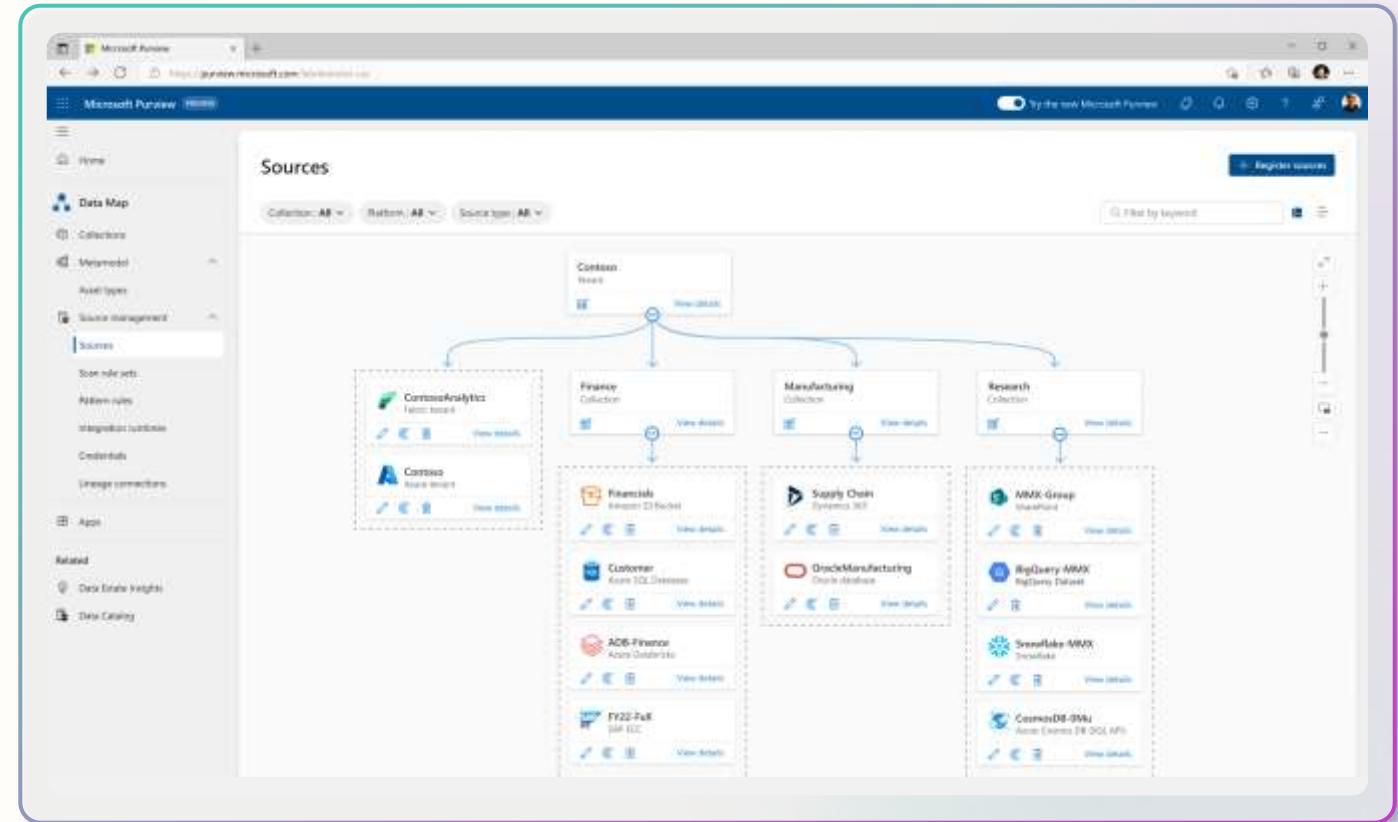
Try out Purview easily with [no set-up required](#) with Purview Data Map automatically provisioned and attached to every Fabric instance by default



Scan Fabric tenants at no extra cost and discover data assets across your Fabric landscape with data discovery, sensitive data classification, and end-to-end data lineage



Search and explore Fabric data assets in the Purview Data Catalog including the technical properties, sensitive data classification, and data lineage



Keep your data secure with Fabric and Microsoft Purview



Manually label Fabric items with sensitivity labels which follow data automatically as it flows from data lake to business users within Fabric and to Office



Preserve audit logs to meet regulatory requirements, support forensic investigations, and gain high-bandwidth access to data



Microsoft Fabric user activities are logged and available in the Microsoft Purview audit log – use the Fabric admin monitoring workspace to explore audit logs in Purview

The screenshot shows the Microsoft Fabric Data Lake interface. On the left, there's a sidebar with icons for Home, Create, Monitoring, Pipeline, Dataflows, and Admin. A file named 'Customer_Data_34' is selected. In the main area, there are fields for 'Name' (Customer_Data_34) and 'Location' (My workspace). Below these, a 'Sensitivity' dropdown is open, showing 'Confidential/Microsoft Extended' with several options: Non-Business (radio button), Public, General, Confidential (radio button checked), Highly Confidential (radio button), and Learn more. A tooltip for 'Confidential' shows three sub-options: Microsoft FTE (red), Microsoft Extended (blue), and Any User (No Protection) (grey). At the bottom right, there are four buttons: 'New Dataflow Gen2', 'New data pipeline', 'Open notebook', and 'New shortcut'. A purple rounded rectangle highlights the 'Confidential' section of the sensitivity dropdown.



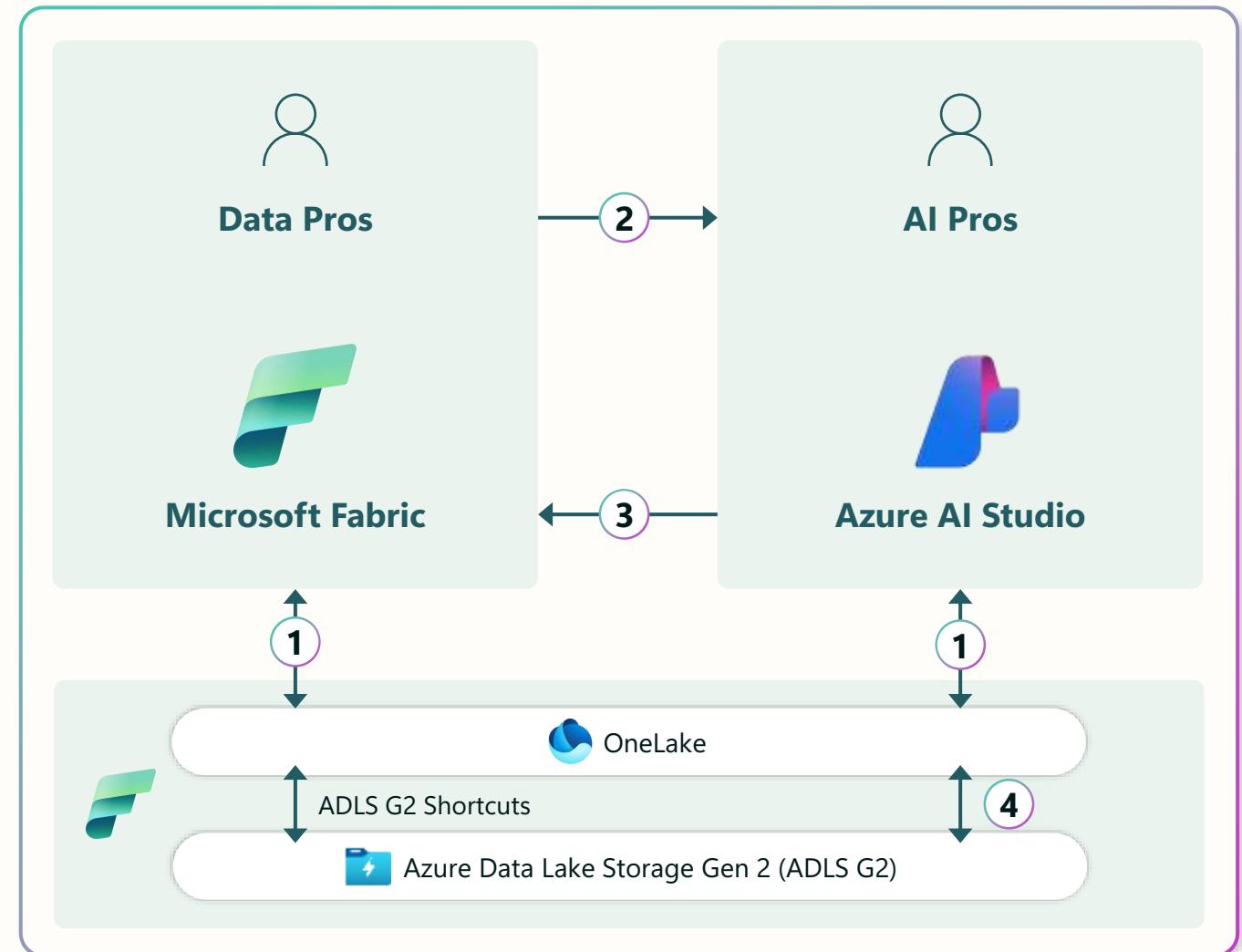
Fabric and Azure AI Studio integration

1 Data Pros and AI Pros access, use and process data across the AI lifecycle with OneLake as the connective tissue.

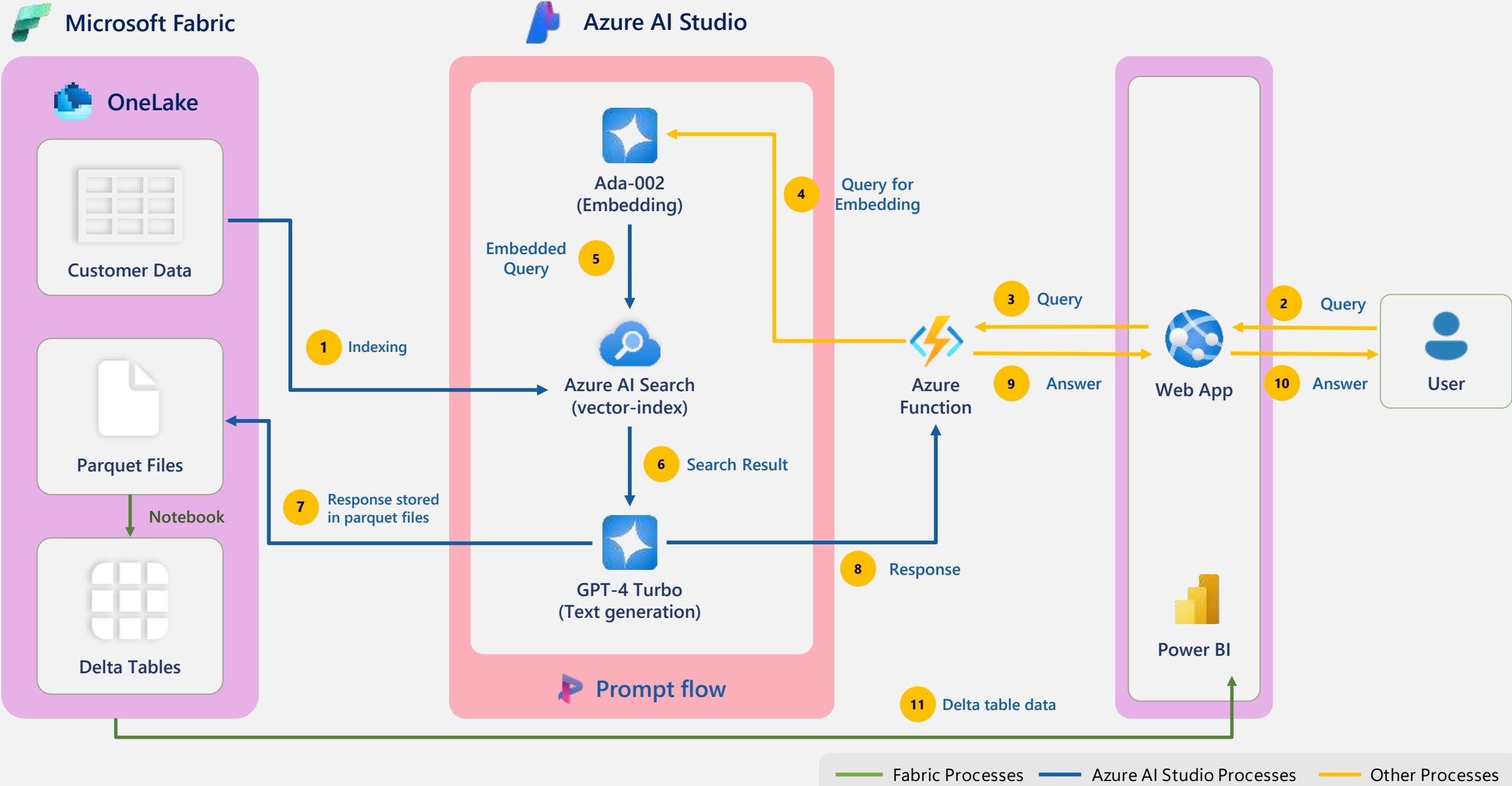
2 Data Pros can pre-process data and share AI-ready data with developers. Once in Azure AI Studio, AI developers can evaluate their models and create generative AI applications and custom copilot experiences.

3 AI Pros can make their models accessible in Fabric for Data Pros to employ

4 Data Pros can enrich their analytics workflows in their lakehouse and warehouse or serve through Power BI using Direct Lake mode.



Build your copilot in Azure AI Studio



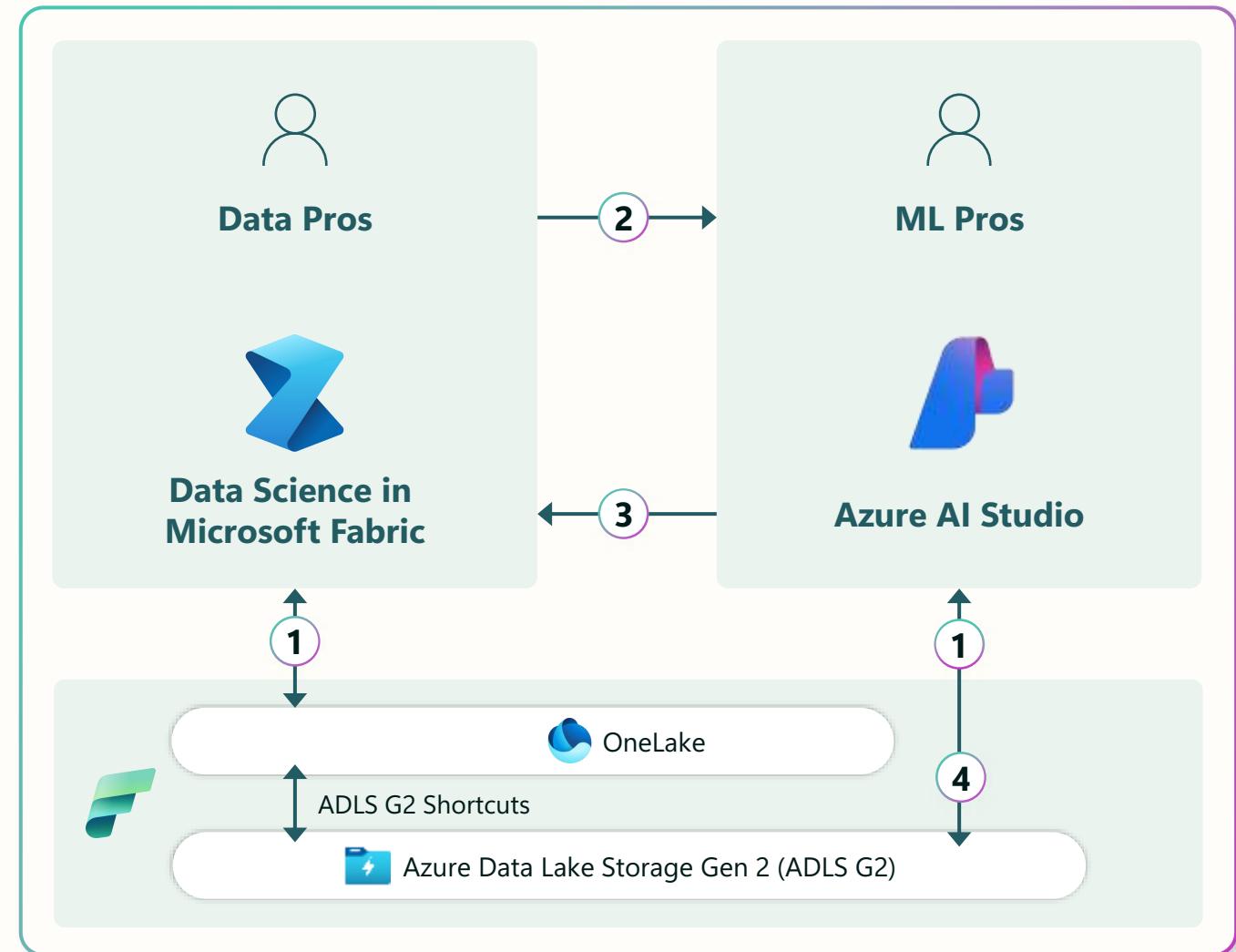
Fabric and Azure Machine Learning integration

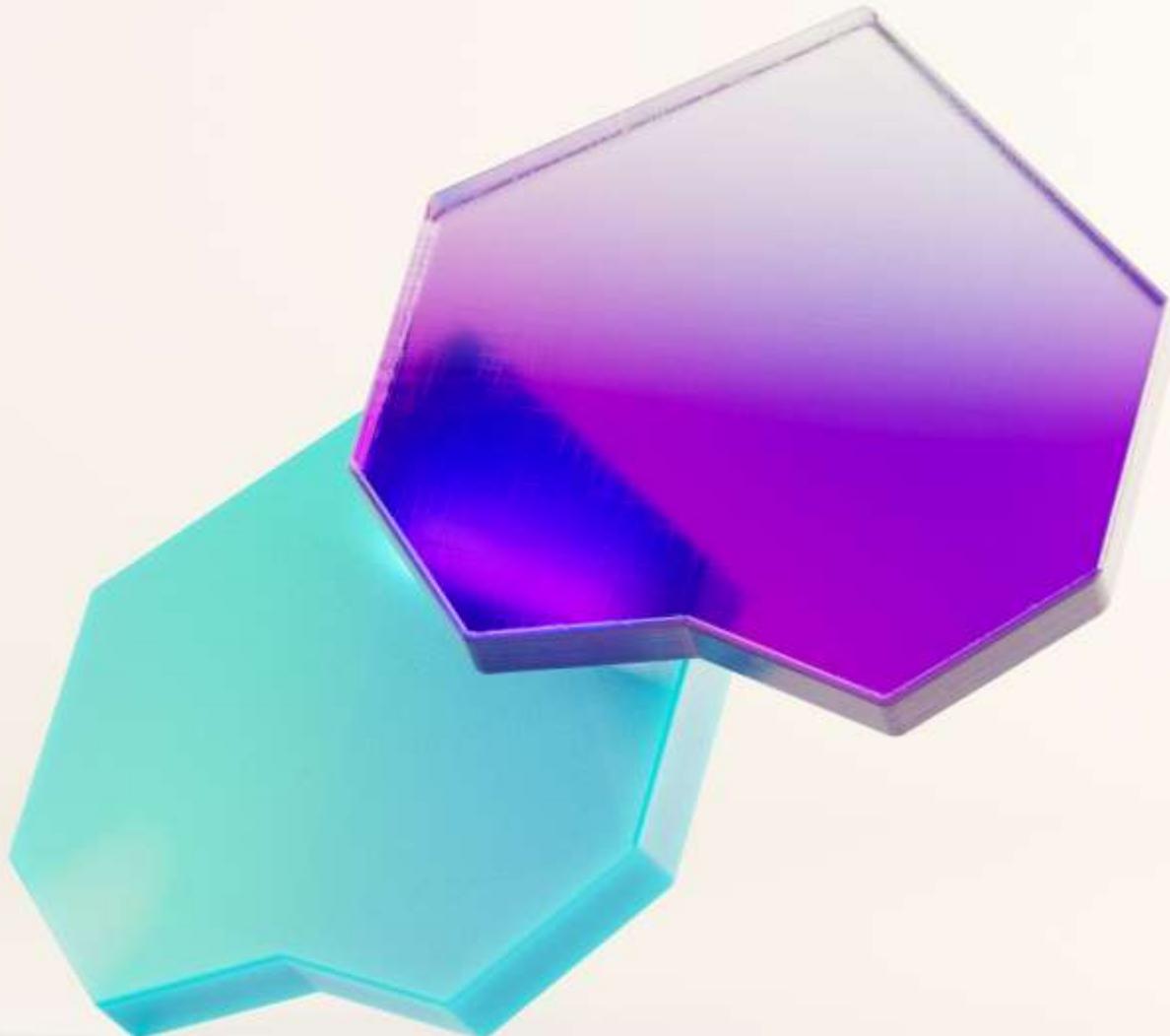
1 Data Pros and ML Pros access data items across all stages of the ML lifecycle by reading from and writing to OneLake. ADLS Gen 2 Shortcuts to OneLake can be used reference data from ADLS Gen 2 into OneLake.

2 Data Pros can train with Spark and access ML models in Fabric and deploy models using AzureML online and batch endpoints for production.

3 ML Pros can export models from AzureML to Fabric to enrich lakehouse or data warehouse serving data as part of analytics workflows.

4 Write predictions back to OneLake via Shortcuts to ADLS Gen2 and surface insights through PBI reports with Direct Lake mode.





Customer stories



Aurizon uses Microsoft Fabric to take advantage of its data streaming functionality and AI and ML capabilities

“

We're also benefiting greatly from the data streaming capability of Microsoft Fabric and consumption through Direct Lake in Power BI as an answer to HANA's virtual modeling capability, allowing us even greater scalability.

– Chris Nunn, Principal Data Engineer, Aurizon

Situation

Having already fitted majority of its fleet with telemetric sensors relaying data to Microsoft Fabric, Australian freight rail operator, Aurizon, wanted to reimagine its entire analytics estate and move from traditional data warehousing to a modern, scalable analytics platform, so it could get the most out of its data.

Solution

Instead of selecting several tools for this modernization effort, Aurizon chose Fabric as its single, unified data and analytics platform, in which the streaming architecture allows for direct data consumption without duplicating data (using Direct Lake mode in Power BI), replacing the need for Aurizon to create and maintain datasets in the future.

Impact

Having brought its enterprise data and telemetry data together in Microsoft Fabric, Aurizon has seen performance gains up to 240x. Aurizon plans to use Azure Machine Learning and Copilot in Microsoft Fabric to continue to drive better outcomes for customers using data and predictive analytics to transform business practices.

[Read full story here](#)





One NZ democratizes data access, enabling customer support with Microsoft Fabric Real-Time Intelligence



“

Previously, you needed to be a data engineer or scientist to access and understand customer information. Now we're making it user-friendly, so anyone can easily make data-driven decisions.

– Strathan Campbell, Channel Environment Technology Lead, One NZ

Situation

Seeking to improve customer experience, New Zealand telecom giant, One NZ, wanted to take its performance to the next level. Increasing data volumes were leading to delayed refresh rates and One NZ needed a solution with real-time data and analytics capabilities they could easily implement with their existing systems.

Solution

One NZ selected Microsoft Fabric's Real-Time Intelligence (RTI) and went from concept to delivered product in just two weeks. Using event streams to ingest data from external data sources and run analytics on top of its real-time streaming data, One NZ was able to generate immediate actions on their data for quicker decision-making.

Impact

With Microsoft Fabric's RTI, One NZ has improved overall customer experience. All teams have access to high quality data that can be used in ML models, plus, dashboards are updated 6x faster than before, so agents can identify customer behavior patterns and respond to customers more quickly than ever.

[Read full story here](#)





Milliman enables self-service actuarial modeling and mitigates business risk with Microsoft Fabric



“

Milliman is tremendously excited about Microsoft Fabric. We see that Microsoft Fabric will be critical for our professionals and insurance industry clients to realize the full potential of AI to protect the health and financial well-being of people everywhere.

— Ken Mungan, Chairman, Milliman

Situation

Milliman needed a modernized data analytics platform for actuaries to build custom data pipelines and quickly and accurately analyze terabytes of data.

Solution

Milliman leveraged Microsoft Fabric as their all-in-one analytics solution. They centralized all data and analytics workloads on a single, open, governed foundation, enabling actuaries with self-service analytics tools to build AI, analytics applications and ML models easily on a data lake-based architecture for near real-time insights.

Impact

Milliman’s new unified infrastructure expedites data access, while Fabric’s OneLake storage system means fewer data copies, making for more efficient reporting and reduced latency and cost.

[Read full story here](#)





ZEISS Group uses Microsoft Fabric to advance innovation with data analytics and AI

“

Data sharing is key, and eVA, supported by Fabric on Azure, could provide the foundation for a trusted analytics solution that we can easily share across our business units.

- Markus Morgner, Head of Enterprise Data Platform & Engineering, ZEISS Group

ZEISS Group is committed to scaling AI innovation for its advanced optical solutions, products and digital offerings.

Situation

The team needed a modern analytics solution to streamline its analytics workflows and provide internal business units with frictionless access to information.

Solution

ZEISS Group leveraged Microsoft Fabric to connect and combine all its business data in a unified and governed hub without duplication or movement.

Impact

Their new solution has allowed teams to increase use of AI and be more insight-driven, reducing the time to create new products and enabling business growth with self-service capabilities.



[Read full story here](#)

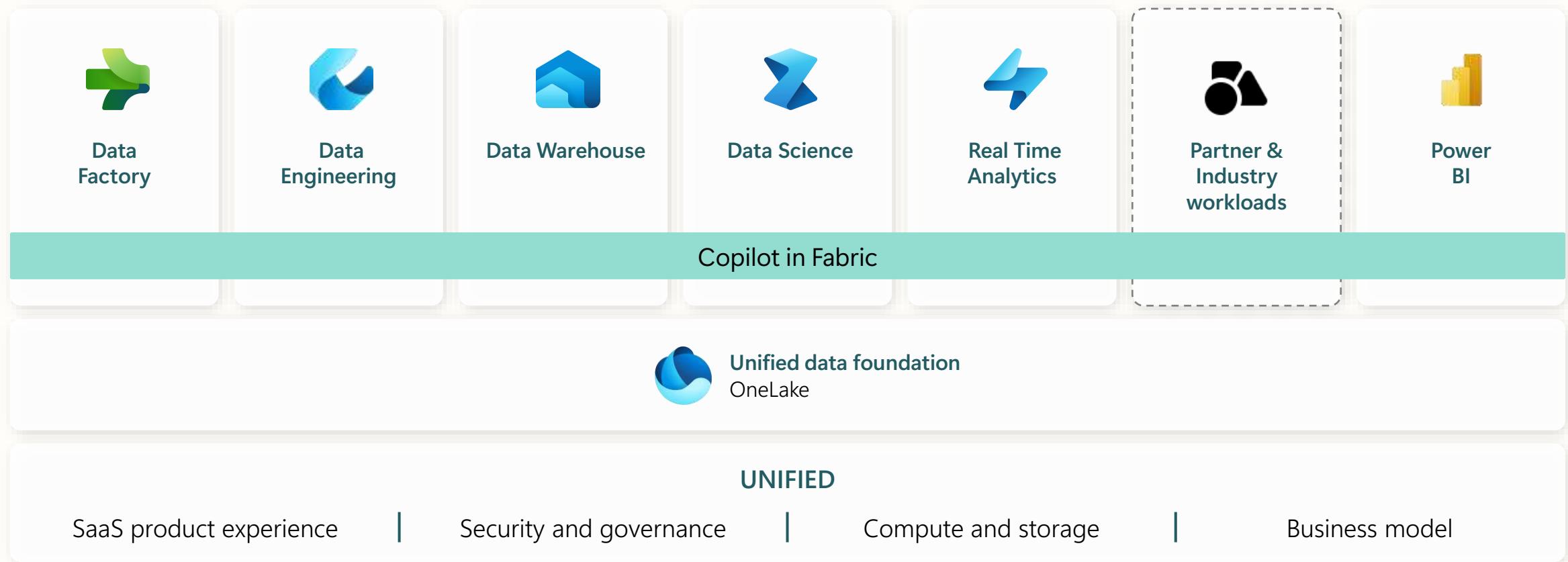




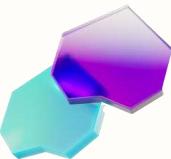
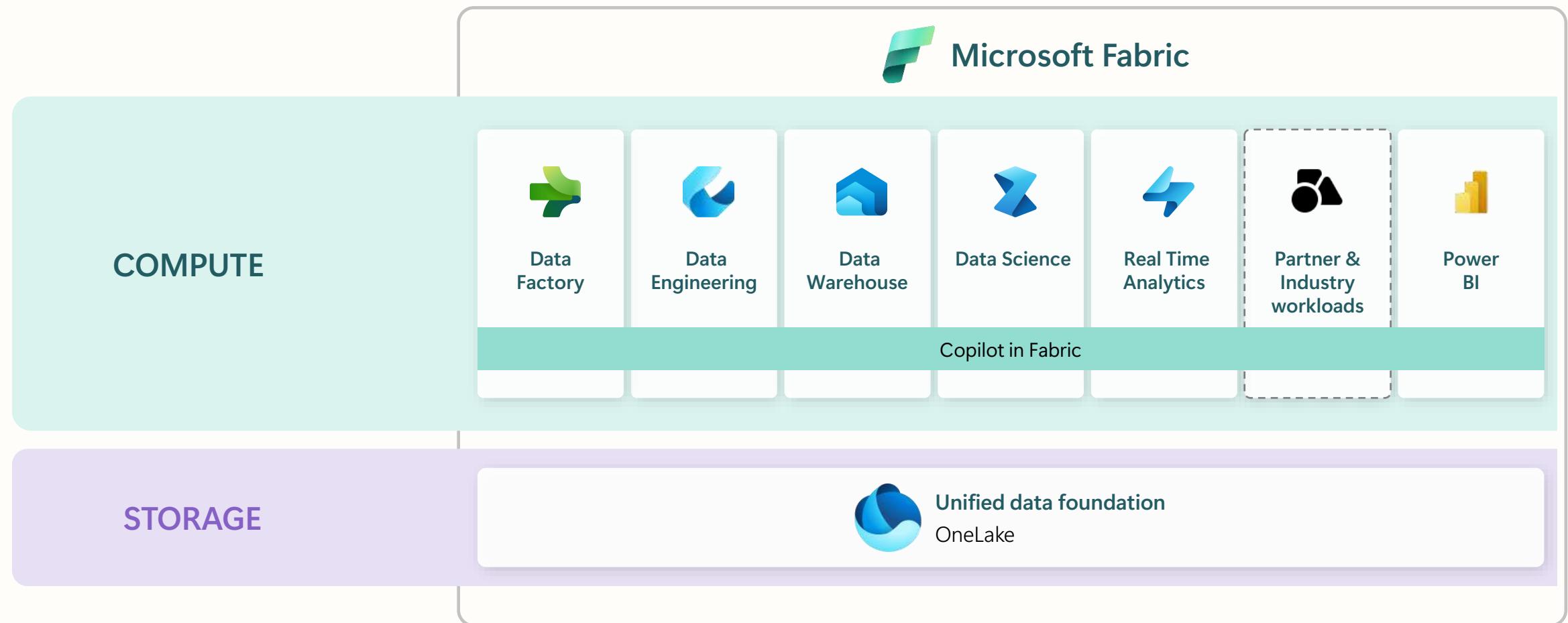
Licensing

Microsoft Fabric does it all—in a unified solution

An end-to-end analytics platform that brings together all the data and analytics tools that organizations need to go from the data lake to the business user



Microsoft Fabric business model overview



Microsoft Fabric simplicity

Microsoft Fabric is a unified product for all your data and analytics workloads. Rather than provisioning and managing separate compute for each workload, with Microsoft Fabric, your bill is determined by two variables: the amount of compute you provision and the amount of storage you use.



COMPUTE

A shared pool of capacity that powers all capabilities in Microsoft Fabric.

Pay-as-you-go and 1-year Reservation.



STORAGE

A single place to store all data.

Pay-as-you-go (\$ per GB/month).



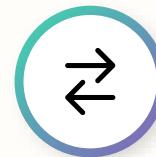
Microsoft Fabric capacity

A shared capacity that provides a set of capacity units (CUs) powering the compute for Data Warehouse, Data Integration, Data Science, Data Engineering, Real-Time Intelligence, Power BI, and Copilot.



Simplicity of purchasing

No need to purchase separate compute for each workload



Flexibility in use

With no pre-allocation required, the same set of CUs can be used for any workload



Cost savings

All CUs are pooled together and are not locked to an idle workload. [Smoothing](#) on compute usage also allows you to size capacity closer to the average usage than the peaks



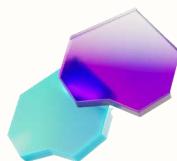
Value optimization

Adjust compute capacity by scaling up and down



Transparent monitoring

Centralized dashboard to monitor usage and costs in Microsoft Fabric utilization and metrics app



Fabric capacity pricing

Fabric capacity is priced uniquely across regions. The following table shows the pricing at US West 2 for reference. Fabric capacity can be purchased at Azure portal. Visit [Fabric pricing page](#) for more details.

1 CU pay-as-you-go price at US West 2 \$0.18/hour

SKU	Capacity unit (CU)	Pay-as-you-go (monthly)	Reservation (monthly) ~40.5% saving over Pay-as-you-go
F 2	2	\$262.8	\$156.334
F 4	4	\$525.6	\$312.667
F 8	8	\$1,051.2	\$625.334
F 16	16	\$2,102.4	\$1,250.667
F 32	32	\$4,204.8	\$2,501.334
F 64	64	\$8,409.6	\$5,002.667
F 128	128	\$16,819.2	\$10,005.334
F 256	256	\$33,638.4	\$20,010.667
F 512	512	\$67,276.8	\$40,021.334
F 1024	1024	\$134,553.6	\$80,042.667
F 2048	2048	\$269,107.2	\$160,085.334

Note:

- 1 CU PAYGO monthly rate calculation: $\$0.18 \times 730 = \131.4 . F2 = $\$131.4 \times 2 = \262.8
- 1 CU RI monthly rate calculation: Round $(\$0.18 \times (1 - 0.405) \times 730 \times 12,0) / 12 = \sim \78.166 ...F2 RI = $\sim \$78.166 \dots \times 2 = \sim \156.334

- Power BI Pro license is required for all Power BI Premium ("P") and Fabric Capacity ("F") SKUs to publish Power BI content to Microsoft Fabric. Enabling content consumers to review and interact with Power BI reports without additional paid per-user licenses is available at P1 and above (and F64 and above).

How it works

Capacity units (CUs) = Compute power

Capacity units (CUs) are units of measure that represent a pool of compute power needed. Compute power is required to run queries, jobs, or tasks.

CU Consumption

The CU consumption is highly correlated to the underlying compute effort needed for the tasks performed by the capability during the processing time.

Each capability, such as Power BI, Spark, Data Warehouse, with the associated queries, jobs, or tasks has a unique consumption rate.

Note: For additional documentation on purchasing experience and capacity models, please see here: aka.ms/fabriclicensing

Translation for existing Power BI Premium per capacity Customers

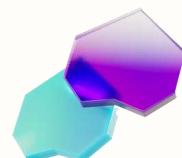
Today	
SKU	PBI vCore
P1	8
P2	16
P3	32
P4	64
P5	128



Microsoft Fabric Generally Available	
SKU	CU
F 64	64
F 128	128
F 256	256
F 512	512
F 1024	1024

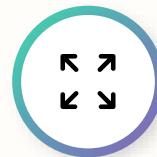
- Power BI Premium P SKUs get **access to the full Fabric capabilities**
- 8 CUs are providing the same compute power of 1 Power BI Premium v-core for Power BI workloads. That is, **F64 provides the equivalent compute power of Power BI Premium P1**

For additional documentation on Fabric licensing types, please see here: [Microsoft Fabric concepts – Microsoft Fabric | Microsoft Learn](#)



OneLake

A single unified SaaS data lake that comes automatically provisioned with every tenant without additional infrastructure to manage



Simplicity of purchasing

Automatically provisioned single storage service for all workloads

Single source of truth

Bring in data from anywhere and reduce costs from data duplication by virtualizing your cross-cloud data in OneLake

Maximize data investments

Allows linking your existing third-party storage systems natively and makes it available to Microsoft Fabric analytics workloads

Open access

Open format democratizes access to multiple analytical engines, reducing additional resources needed to set up

Simplified security

Simplified security, governance, and data discovery in one place to reduce maintenance overhead



OneLake pricing

OneLake is a data lake built into Microsoft Fabric and provides a single place to store all organizational data. Data storage is charged at a rate of \$ per GB per month and priced uniquely across regions.



Data Storage



Data Transfer & Internet Egress

Type	Pay-as-you-go price at West US 2
OneLake storage	\$0.023 per GB/month
OneLake BCDR storage	\$0.0414 per GB/month
OneLake cache	\$0.2 per GB/month

Cross-region data transfer network charges may apply based on source/destination of each storage access. Learn more at the [Bandwidth Pricing](#).

Note: OneLake cache is billed for KQL cache storage and Data Activator data retained.



Free storage for mirroring data

Enjoy free Mirroring storage for replicas up to a certain limit based on the purchased compute capacity SKU you provision. OneLake storage is billed only when the free Mirroring storage limit is exceeded, or the provisioned compute capacity is paused*

Capacity SKU	Free Mirroring storage for replicas (up to X TB)
F 2	2
F 4	4
F 8	8
F 16	16
F 32	32
F 64 / P1	64
F 128 / P2	128
F 256 / P3	256
F 512 / P4	512
F 1024 / P5	1024
F 2048	2048

*The free Mirroring storage can only be used for storing Mirroring replica data—no other data. Free Mirroring storage is also only for purchased capacities; it does not come with the Fabric free trial.



Copilot in Fabric pricing



Copilot in Fabric is limited to customers who have purchased Fabric capacity (F64 or higher) or Power BI Premium capacity (P1 and above) and is not included in the Fabric free account or trial or Power BI per user licenses



You can simply count Copilot usage against your existing Fabric or Power BI Premium capacity



Copilot usage is measured by the number of tokens processed. Tokens can be thought of as pieces of words. Approximately 1,000 tokens are about 750 words. Prices are calculated per 1,000 tokens, and input and output tokens are consumed at different rates

Operation in Metrics App	Description	Operation Unit of Measure	Consumption rate
Copilot in Fabric	The input prompt	Per 1,000 Tokens	400 CU seconds
Copilot in Fabric	The output completion	Per 1,000 Tokens	1,200 CU seconds

See full details here: <https://aka.ms/Copilot-Billing-Docs>



Appendix

Microsoft Fabric link library

Getting Started with Fabric

[What is Fabric?](#)
[Fabric website](#)
[Fabric trial](#)
[Fabric licenses](#)
[Buy a Fabric subscription](#)
[Navigate the Fabric portal](#)
[Workspaces in Fabric](#)
[Fabric Learning Pathway](#)
[Fabric Technical Documentation](#)
[Fabric Guided Tour](#)
[Fabric Industry Solutions](#)
[Fabric Community](#)
[See more](#)

Copilot in Microsoft Fabric

[Copilot for Data Science and Data Engineering](#)
[Copilot for Data Factory](#)
[Copilot for Power BI](#)

OneLake

[What is OneLake?](#)
[What are shortcuts?](#)
[Create a lakehouse with OneLake](#)
[See more](#)

Security, Governance, and Admin

[Fabric administration](#)
[Data governance and compliance](#)
[Security](#)
[See more](#)

End-to-End Tutorials

[Lakehouse tutorial](#)
[Data Science tutorial](#)
[Real-Time Intelligence tutorial](#)
[Data Warehouse tutorial](#)
[Power BI tutorial](#)
[Data Factory tutorial](#)

Other

[Azure Databricks trial](#)

Microsoft Fabric workload link library



Data Factory

[What is Data Factory?](#)
[Create your first pipeline](#)
[Create your first dataflow](#)
[Connectors](#)
[See more](#)



Data Engineering

[What is Data Engineering?](#)
[Create a Lakehouse](#)
[Create a Spark job definition](#)
[See more](#)



Data Science

[What is Data science?](#)
[Machine learning experiment](#)
[Use end-to-end AI samples](#)
[See more](#)



Data Warehouse

[What is Data Warehouse?](#)
[Create a Warehouse](#)
[Query using SQL query editor](#)
[See more](#)



Real-Time Intelligence

[What is Real-Time Intelligence?](#)
[What is Event stream?](#)
[Create a database](#)
[See more](#)



Power BI

[Enable Microsoft Fabric for your organization](#)
[What is Power BI?](#)
[What is a datamart?](#)
[Azure and Power BI integration](#)
[See more](#)