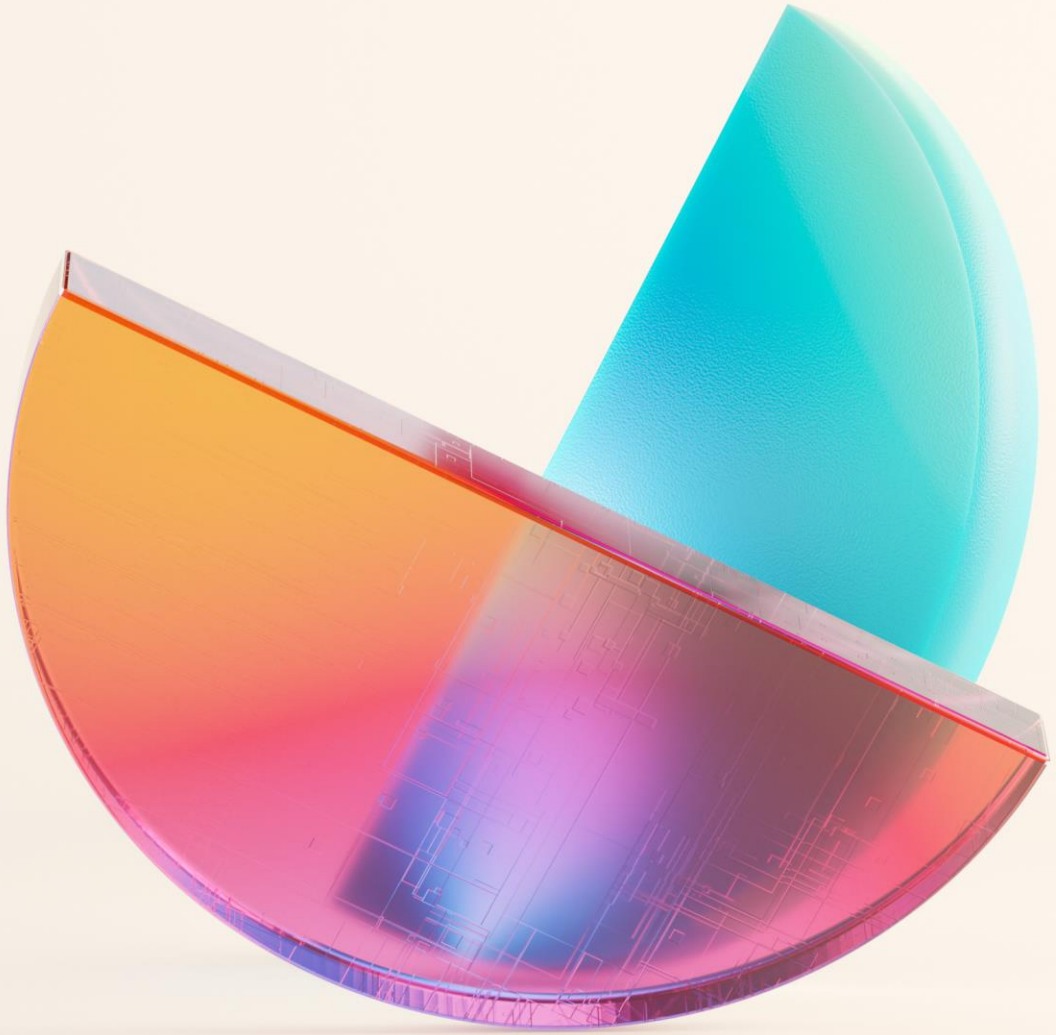
Microsoft | Microsoft Fabric



Microsoft Fabric

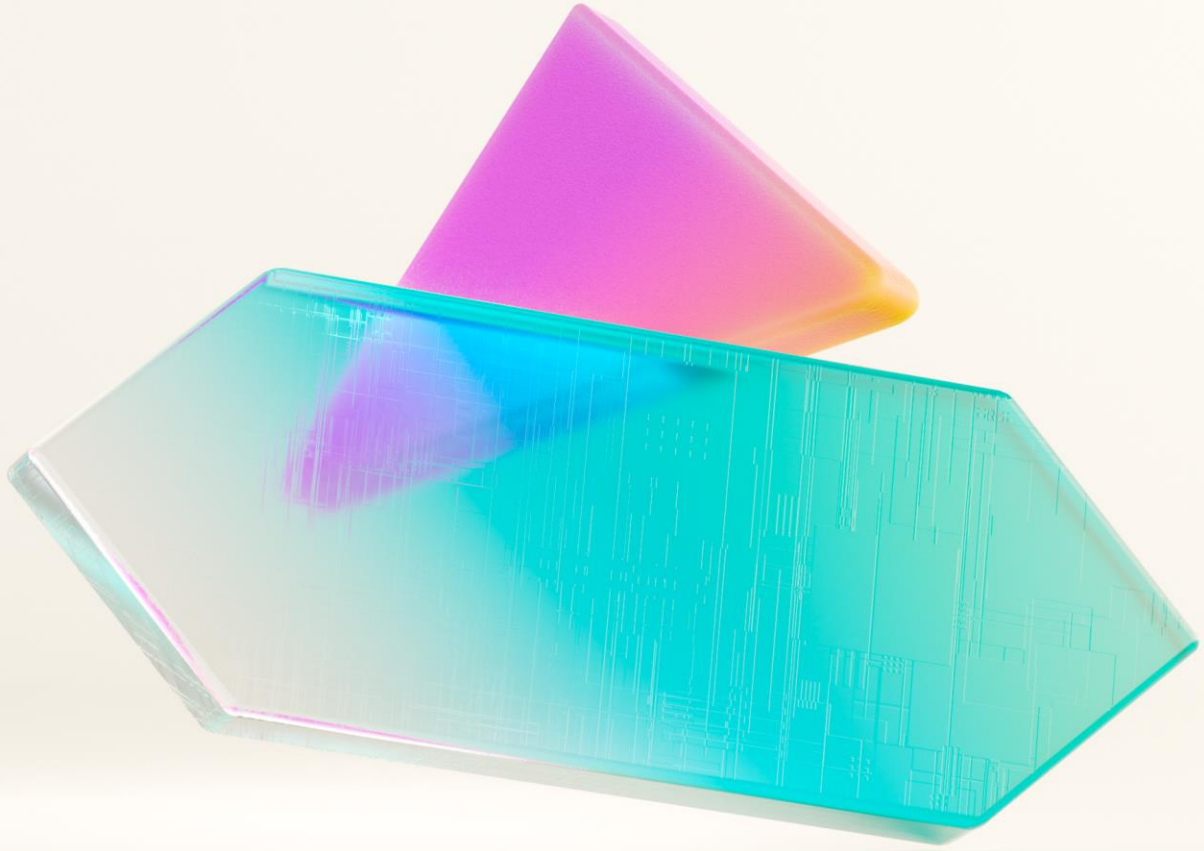
Fabric Analyst in a Day

Version: August 2024

<instructor>

Agenda (times are approximate and will be fluid with the class)

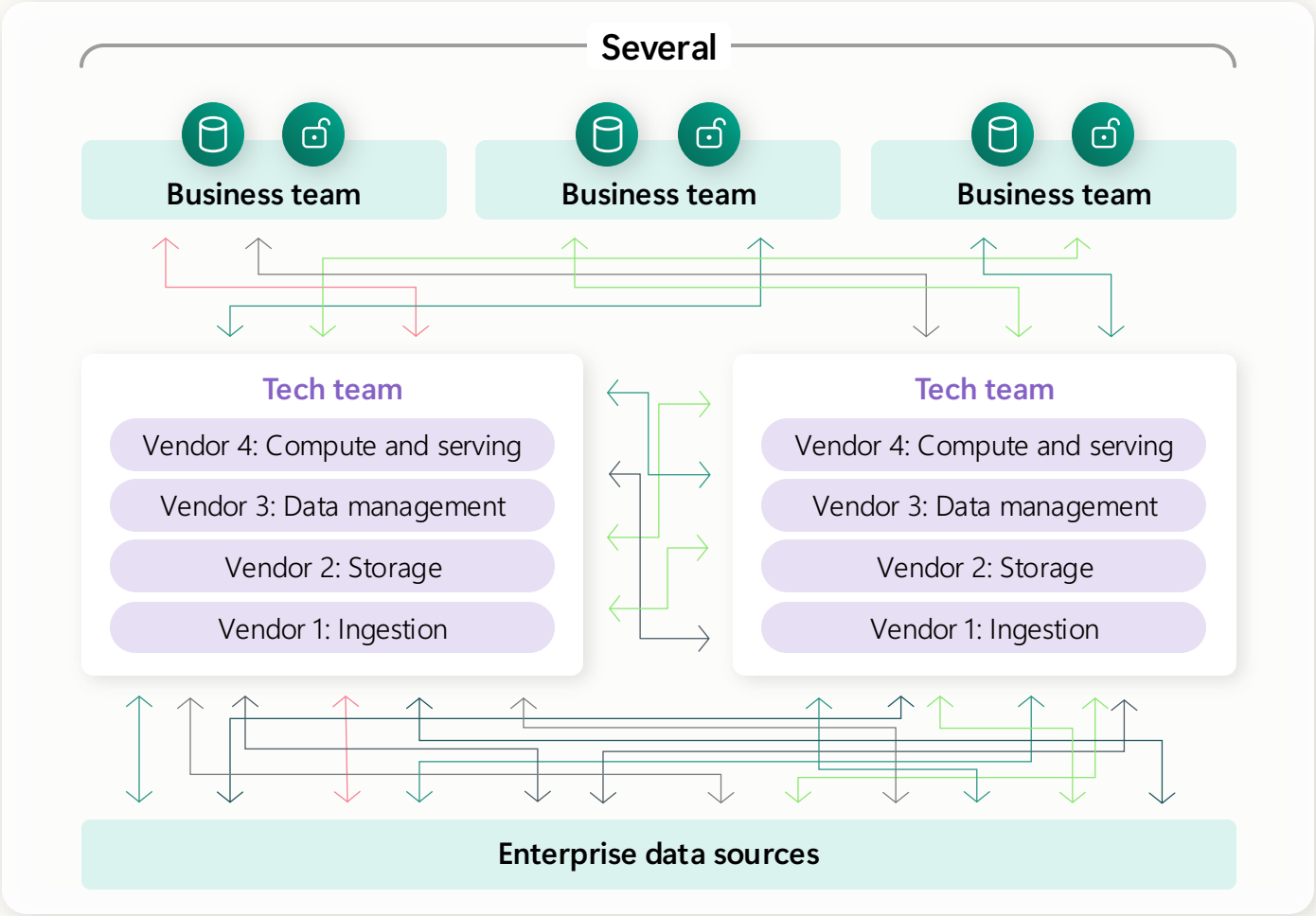
Morning	Presentation	Labs
09:00 AM – 10:00 AM	Introducing Microsoft Fabric	
10:00 AM – 10:30 AM	OneLake overview	Lab 1 – Problem Statement
10:30 AM – 10:45 AM	**Break**	
10:45 AM – 11:00 AM	Data Engineering	Lab 2 – Fabric Workspace
11:00 AM – 12:00 PM	Data Factory	Lab 3 – Introduction to Lakehouse Shortcut – Part 1
12:00 PM – 12:45 PM	**Break for Lunch**	
Afternoon	**Presentation**	**Labs**
12:45 PM – 01:45 PM		Lab 4 – Data Factory Experience – Part 2
01:45 PM – 02:45 PM	Data Warehouse	Lab 5 – Data Factory Experience – Part 3
02:45 PM – 03:00 PM	**Break**	
03:00 PM – 04:00 PM	Power BI	Lab 6 – Data Engineering Experience
04:00 PM – 04:30 PM	Forecast Model Demo	Lab 7 – Power BI Experience
04:30 PM – 04:45 PM	Data Activator Demo	Lab 7 – Clean Up
04:45 PM – 05:00 PM	Next steps and resources	



Introducing Microsoft Fabric

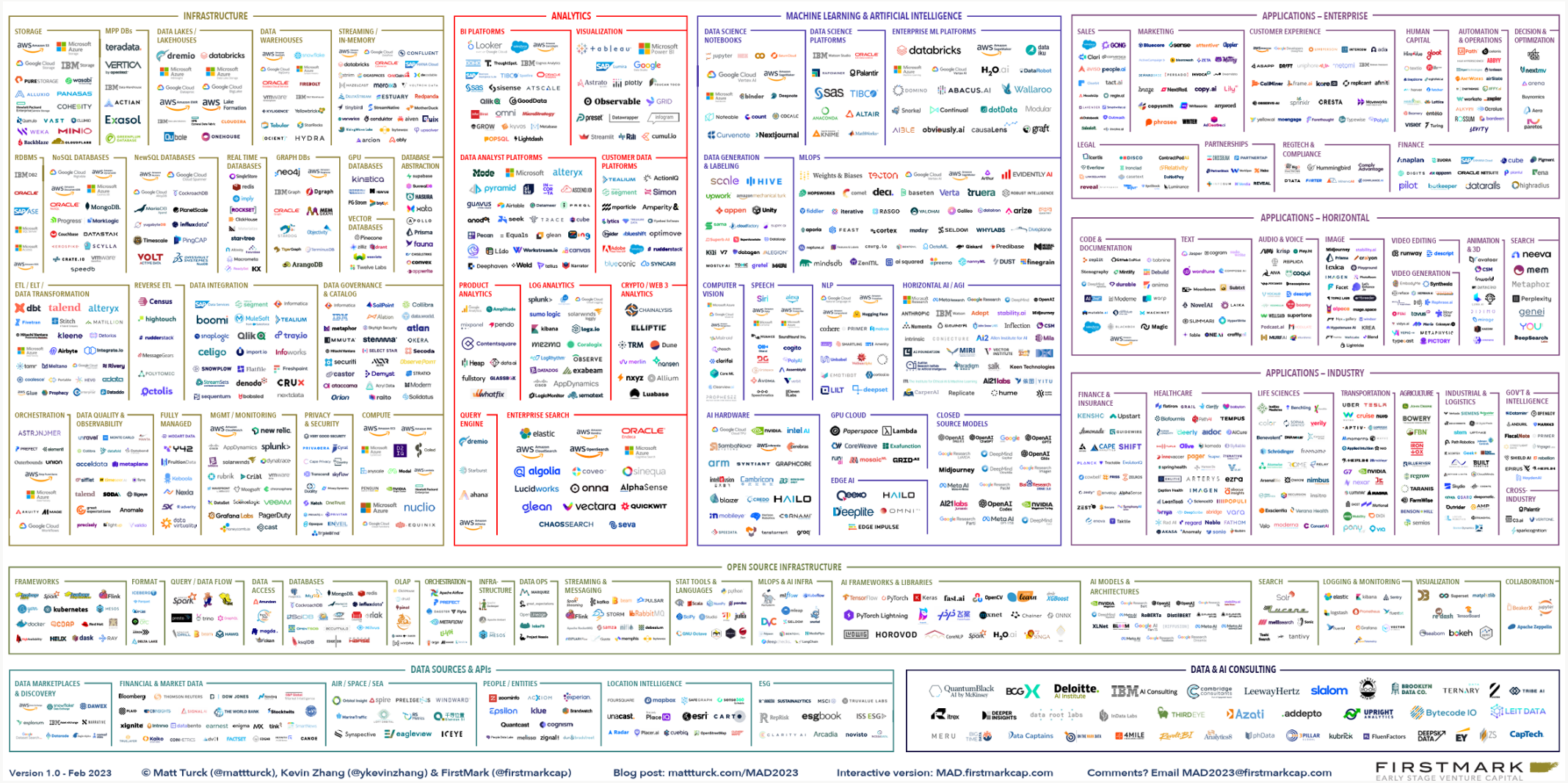
The starting line

A complex, organically evolved data estate

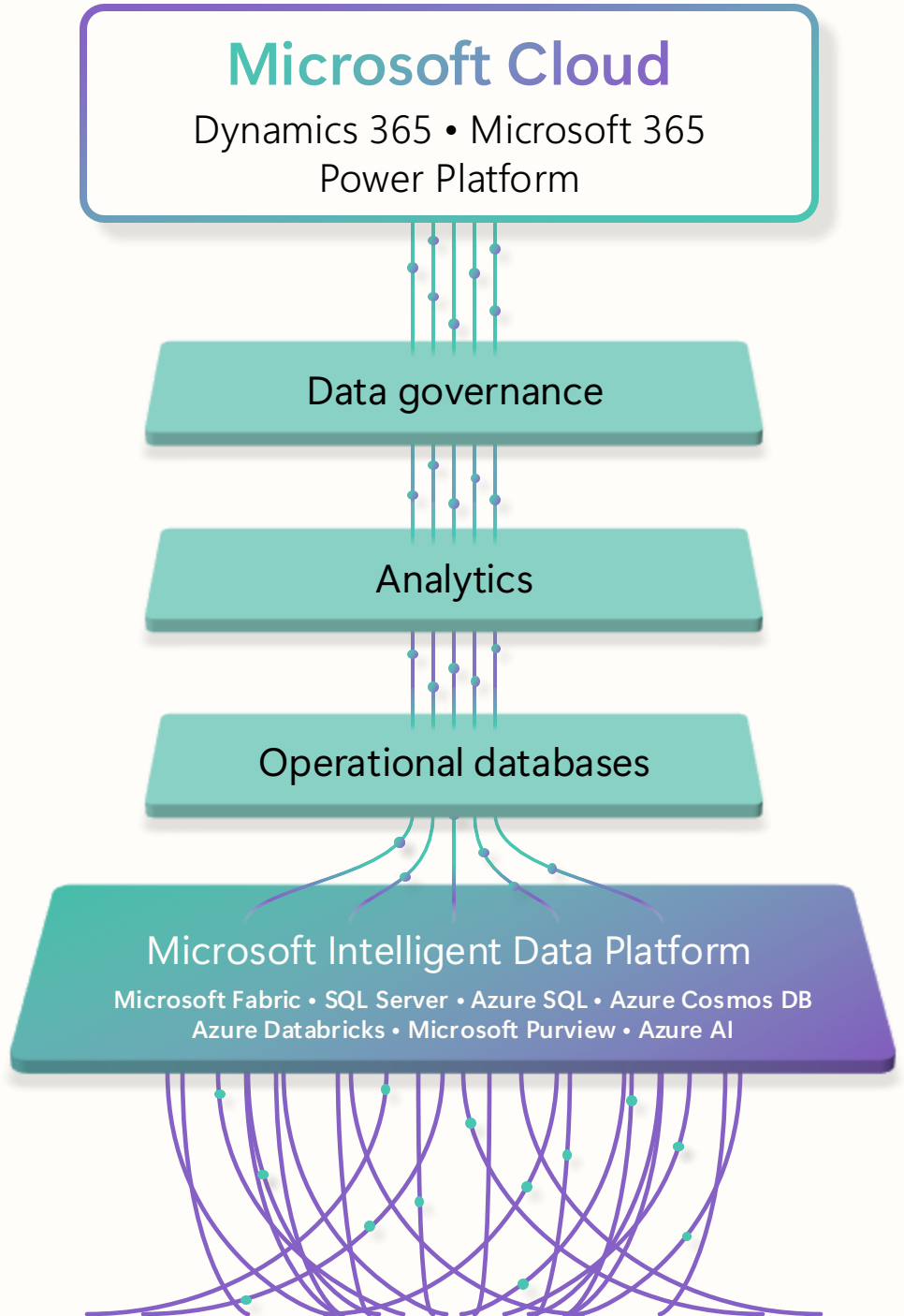


1. Data copies and infrastructure inefficiencies
2. Limited interoperability between vendor services
3. Data exposure risks

Customers enhancing their data estate face immense complexity



Leverage everything Microsoft has to offer



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

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

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

Empower every business user

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

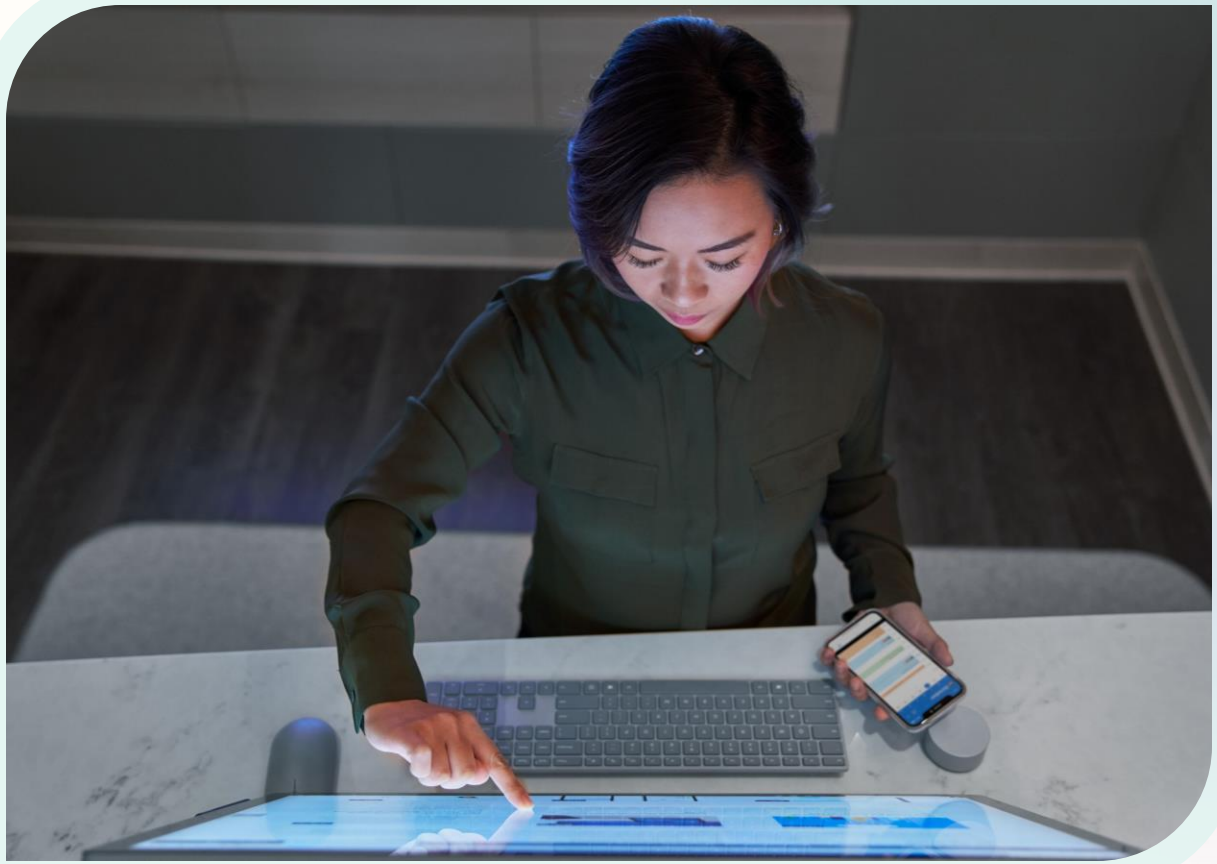
Fuel your AI innovation

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

Fuel your AI innovation

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

- Use conversational language with Copilot in Fabric to create dataflows and pipelines, write SQL statements, or even build machine learning models
- Simply describe what you need—including reports, summaries, and calculations—or ask a question, and Copilot in Power BI does the rest
- Use LangChain and Semantic Kernel to develop and scale custom generative AI models—right from your Fabric Notebook
- Deliver custom generative AI experiences like tailored Q&A on your data with AI skills



Unify your analytics on a complete platform

Give your data teams all the tools they need in a unified, governed, and secure experience that reduces the cost and effort of integration

- Empower data engineers, data scientists, analysts, and business users with role-specific tools in a SaaS platform built for collaboration
- Gain industry-leading, end-to-end security, governance, compliance, and visibility across the unified platform
- Simplify billing and reduce costs with a single pool of capacity and storage that can be used for every workload

Establish a trusted data foundation

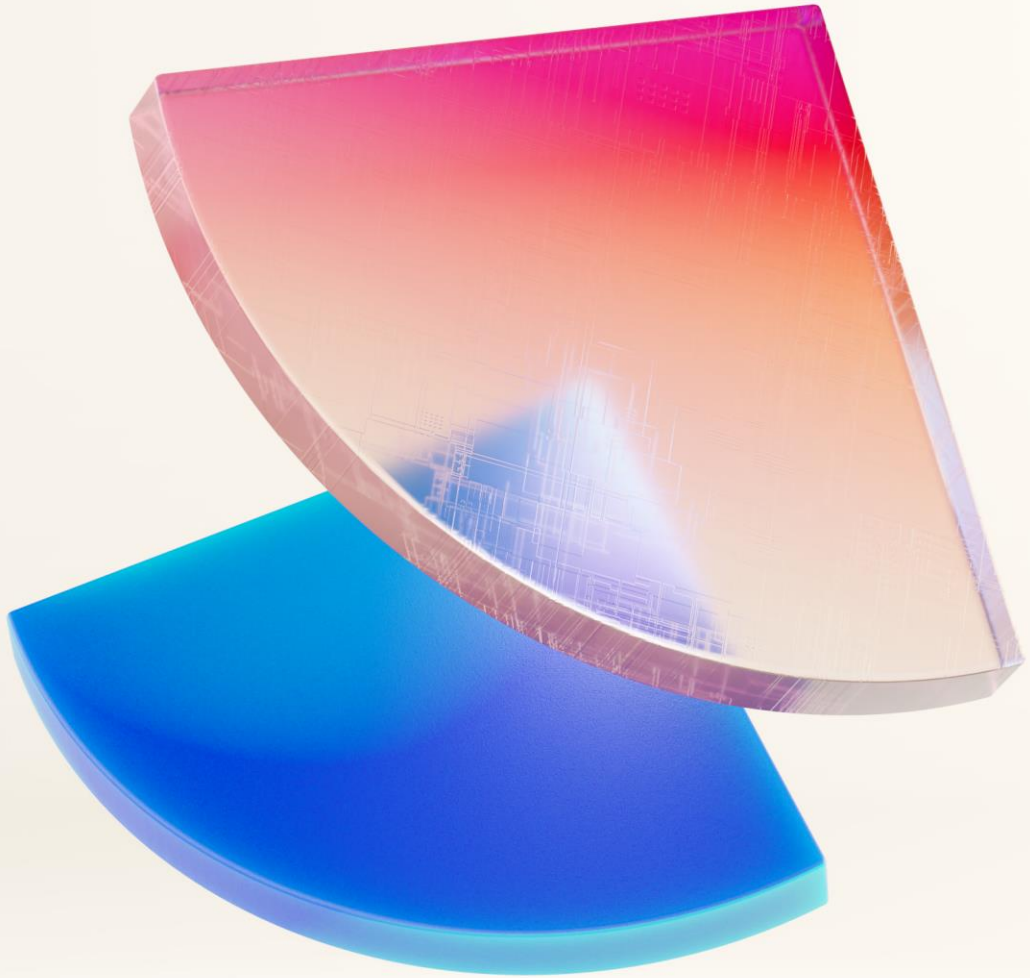
Integrate data from anywhere into a single, multi-cloud data lake for the entire organization, and work from the same copy of data across analytics engines

- Easily connect to data across clouds using “Shortcuts” to virtualize data in OneLake without having to move or duplicate the data
- Create, integrate, manage, and operate data lakes standardized on Delta Parquet format; the same open data format as Azure Databricks
- Intuitively organize your data in Microsoft Fabric’s data lake—OneLake—for central data discovery, sharing, governance, and compliance

Empower every business user

Empower everyone to uncover insights with the data they need, easy-to-use tools, and visuals embedded in the Microsoft 365 apps they use everyday

- Quickly go from data in a Lakehouse to insights in the hands of your business users
- Save time for analysts and provide up-to-date insights with Direct Lake mode, a blazing fast, real-time connection to your data in OneLake
- Foster a data-driven culture by seamlessly and securely embedded insights into Teams, Excel, PowerPoint, Outlook, and more with native integration



OneLake overview

OneLake for all data

"The OneDrive for data"

OneDrive
for documents

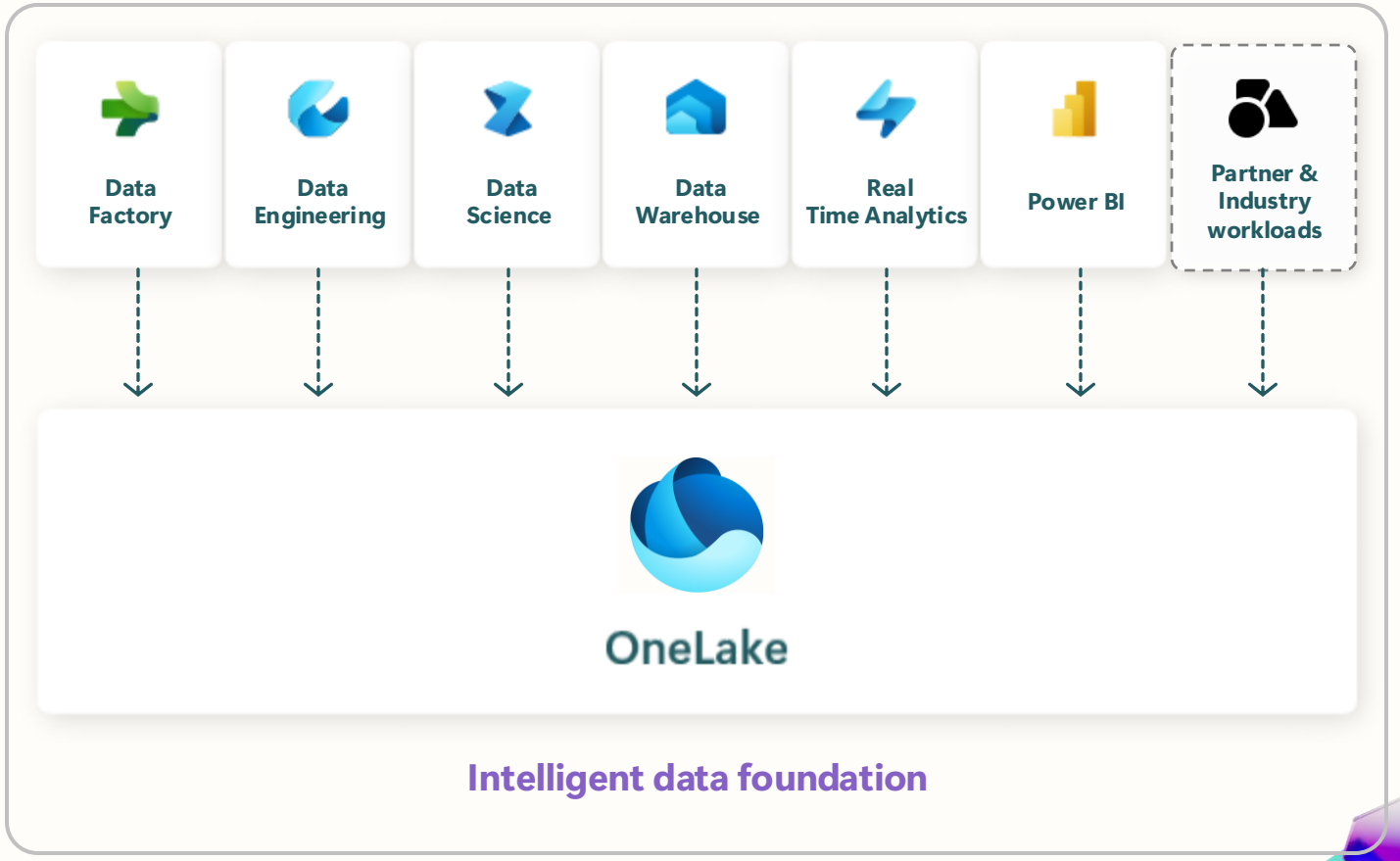
OneLake
for data

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

OneLake for all Data

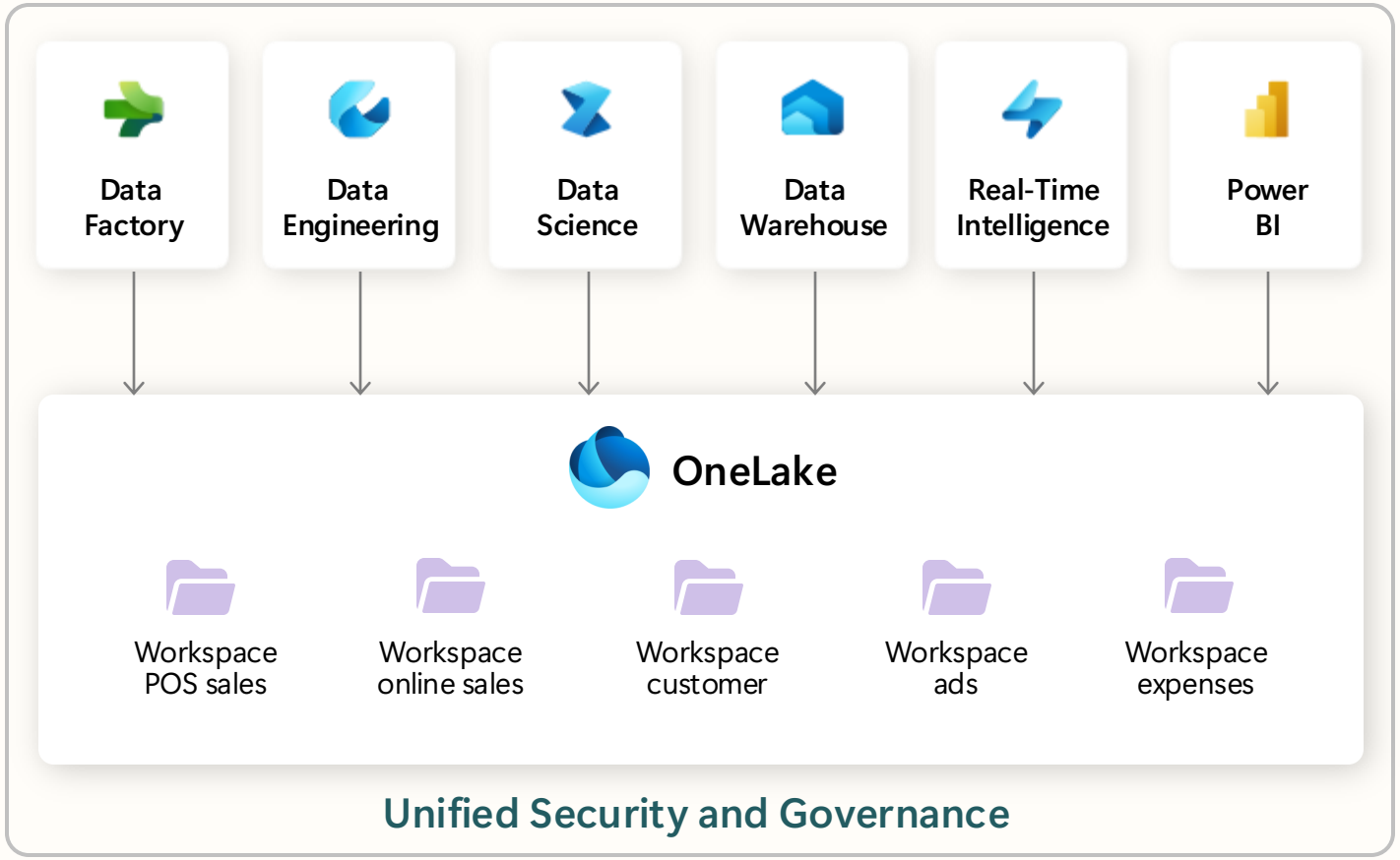
"The OneDrive for Data"

- 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"



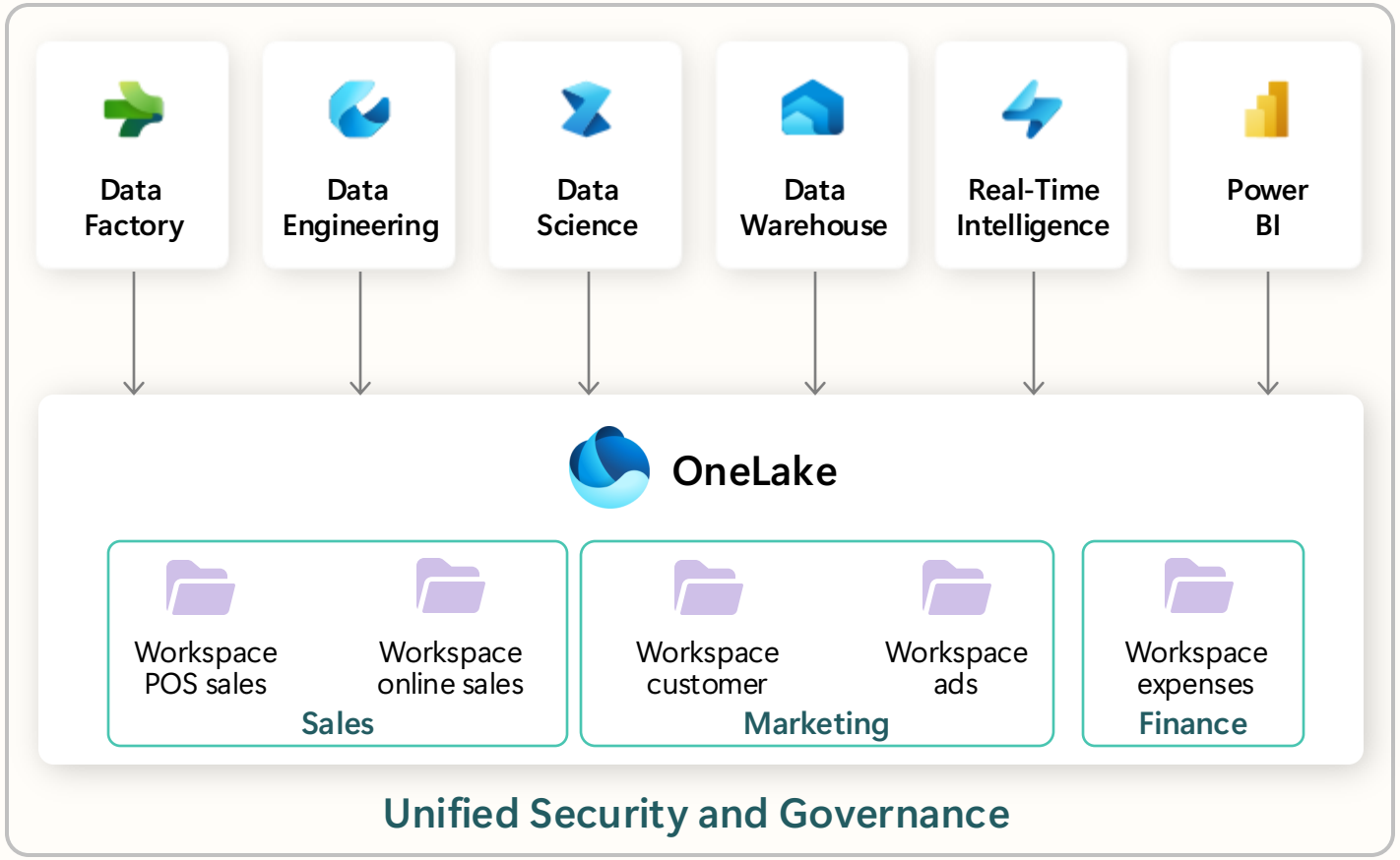
Provisioned automatically with the tenant.

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

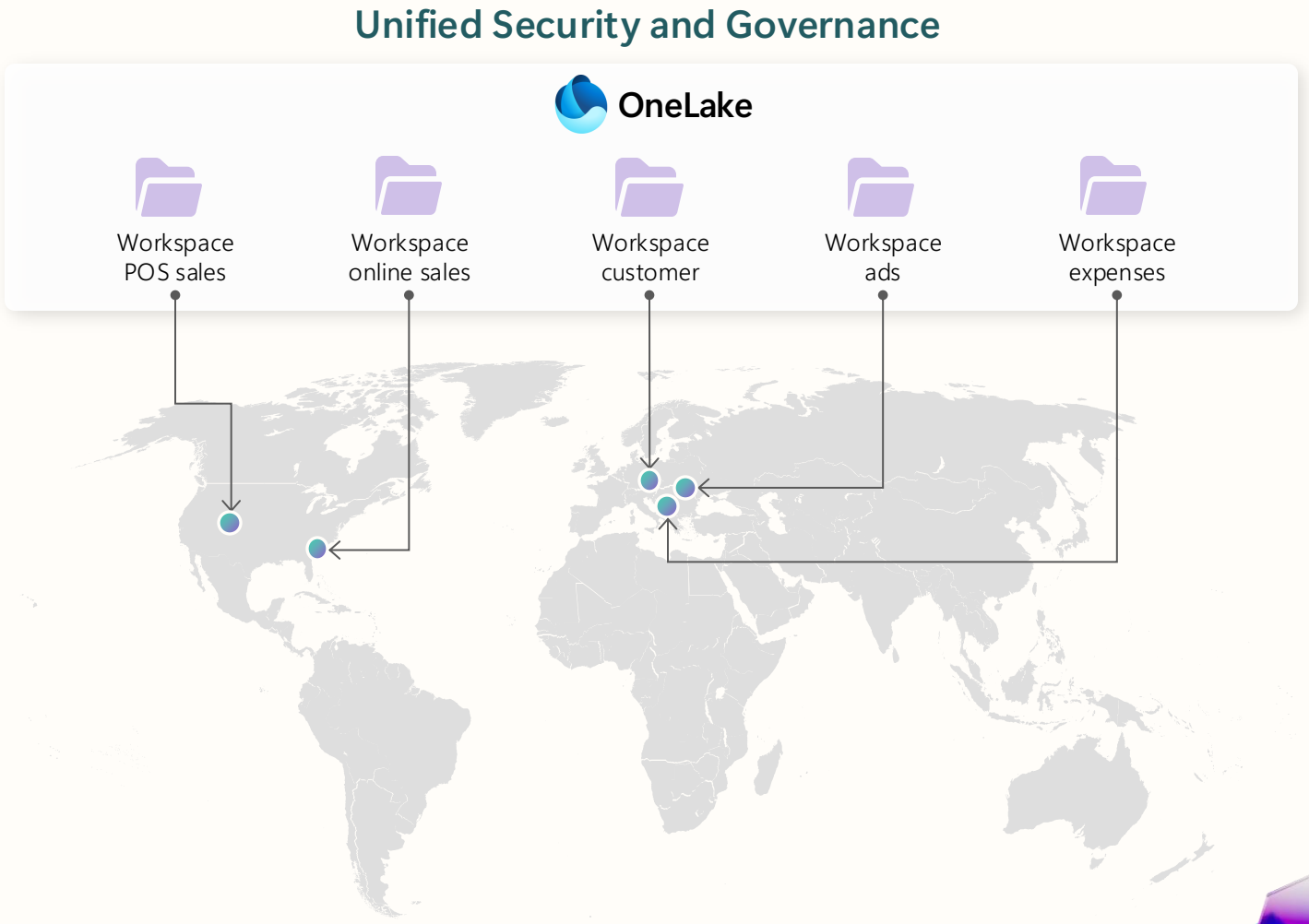
To achieve data residency requirements, workspaces can reside in different regions around the world while still being part of the same data lake.

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).

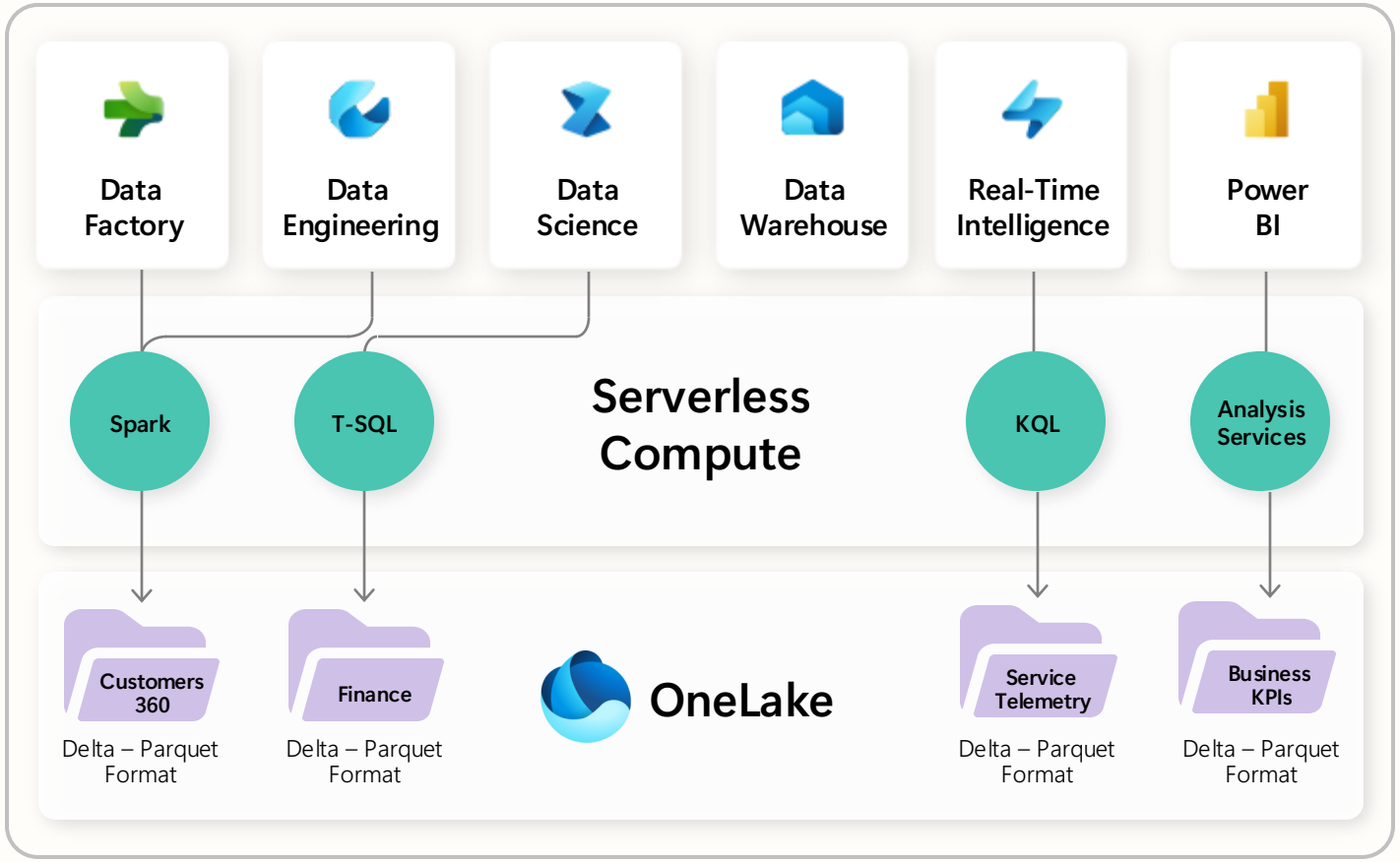
Underlying physical storage is virtualized away.

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



One Copy for all computes

Real separation of compute and storage



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

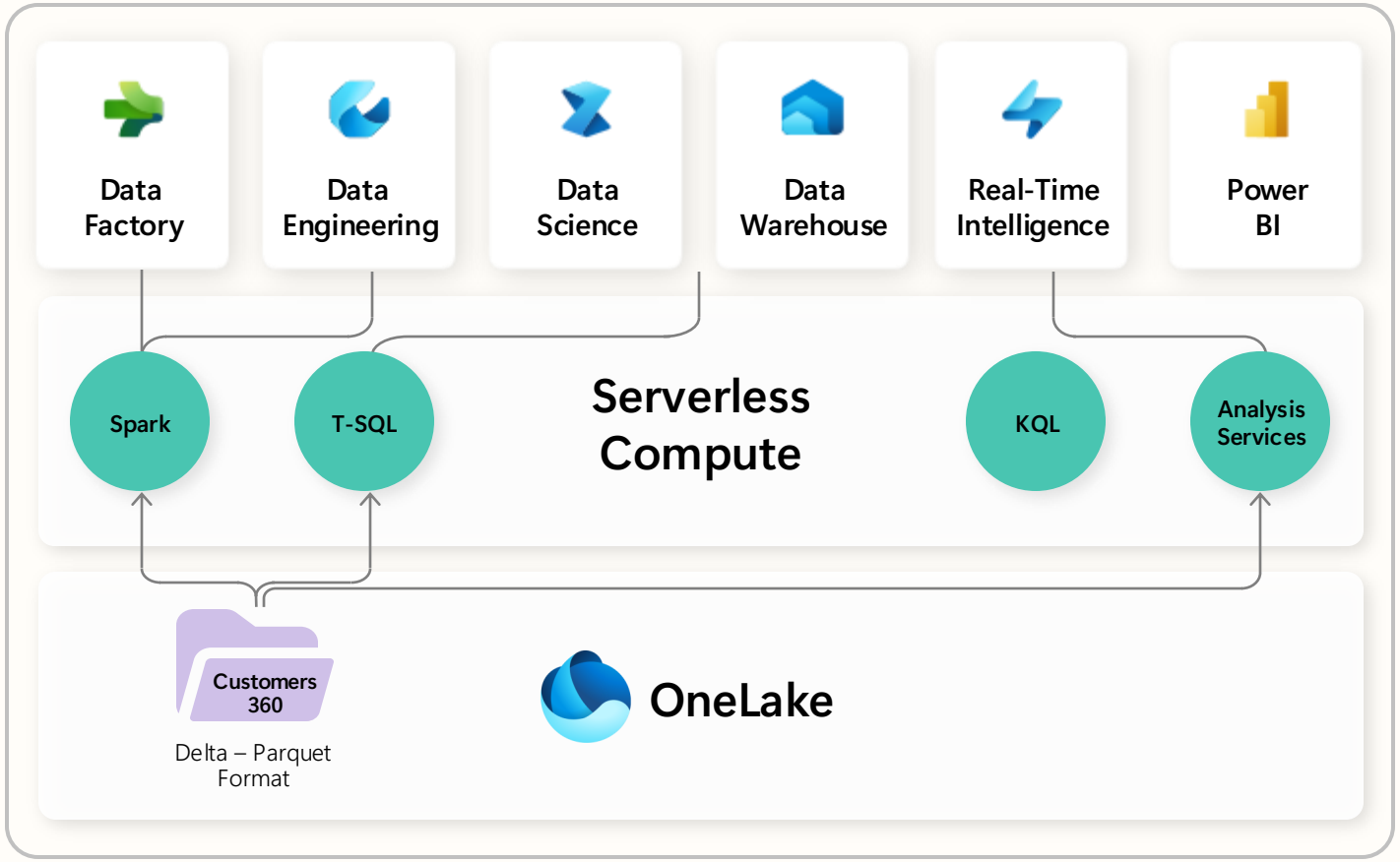
The data is stored in a single common format.

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 computes

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.

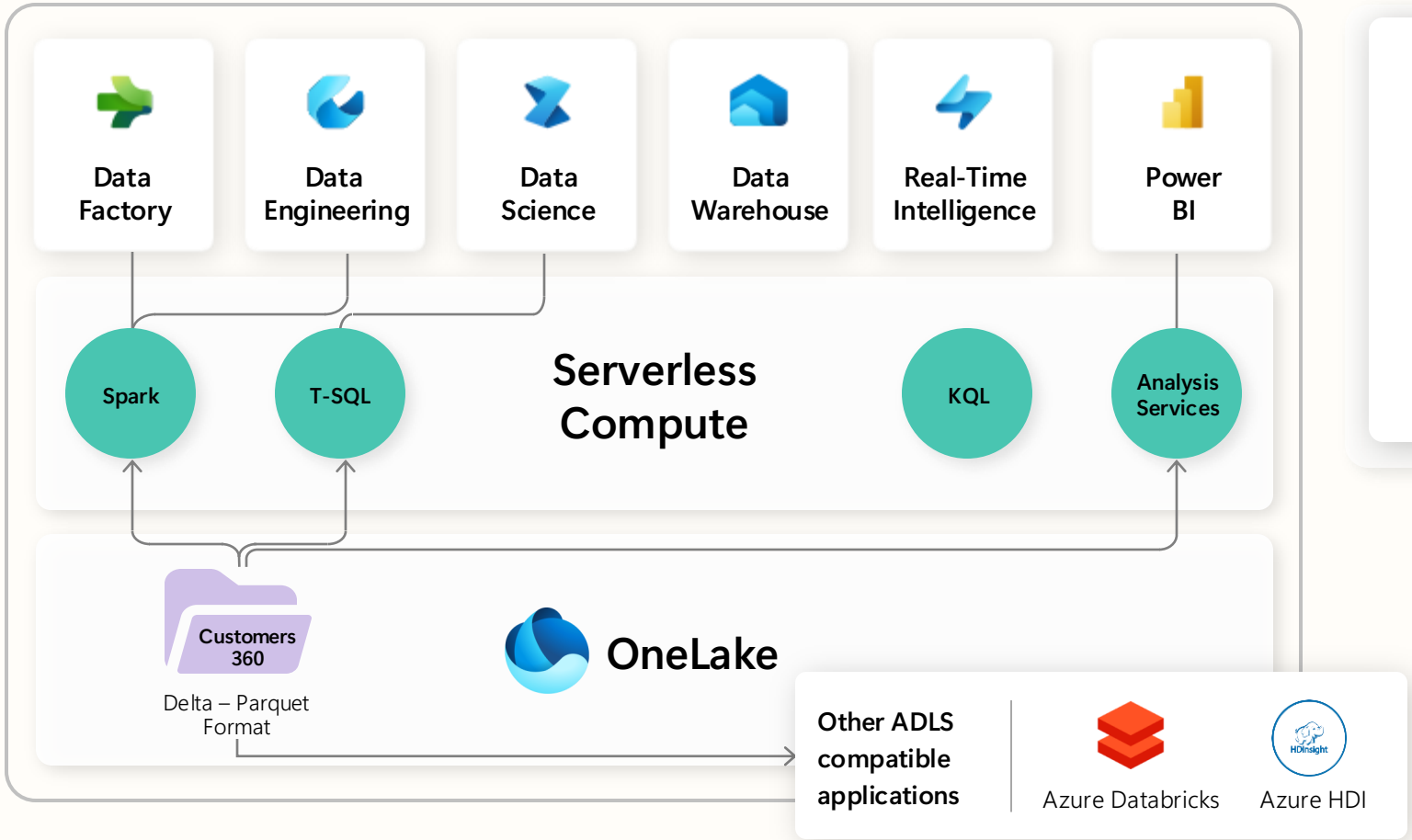
You are able to choose the right engine for the right job.

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



By supporting the ADLS Gen2 DFS APIs and SDKs, OneLake is compatible with existing ADLS applications.

Tenants will appear as one big storage account with different workspaces appearing as different containers with data organized into folders.

Underlying physical storage is virtualized away. OneLake ensures proper scale and performance.

Customer scenario

Fabrikam, Inc. is a wholesale novelty goods distributor. As a wholesaler, Fabrikam's customers are mostly companies who resell to individuals. Fabrikam sells to retail customers across the United States including specialty stores, supermarkets, computing stores, and tourist attraction shops. Fabrikam also sells to other wholesalers via a network of agents who promote the products on Fabrikam's behalf. While all Fabrikam's customers are currently based in the United States, the company is intending to push for expansion into other countries/regions.

You are a Data Analyst in the Sales team. You collect, clean, and interpret data sets to solve business problems. You also put together visualizations like charts and graphs, write reports, and present them to the decision-makers in the organization.

To draw valuable insights from the data, you pull data from multiple systems, clean it, and mash it up together. You pull data from the following sources:

- **Sales Data:** This data comes from the ERP system and is stored in an ADLS Gen2 database or Databricks. It gets updated at noon / 12 PM every day.
- **Supplier Data:** This data comes from different suppliers and is stored in a Snowflake database. It gets updated at midnight / 12 AM every day.
- **Customer Data:** This data comes from Customer Insights and is stored in Dataverse. The data is always up to date.
- **Employee Data:** This data comes from the HR system; it is stored as an export file in a SharePoint folder. It gets updated every morning at 9 AM.

Customer scenario continued

You are currently building a semantic model on Power BI Premium that pulls the data from the above source systems order to satisfy your reporting needs as well as provide end users with the ability to self-serve. You use Power Query to update your semantic model.

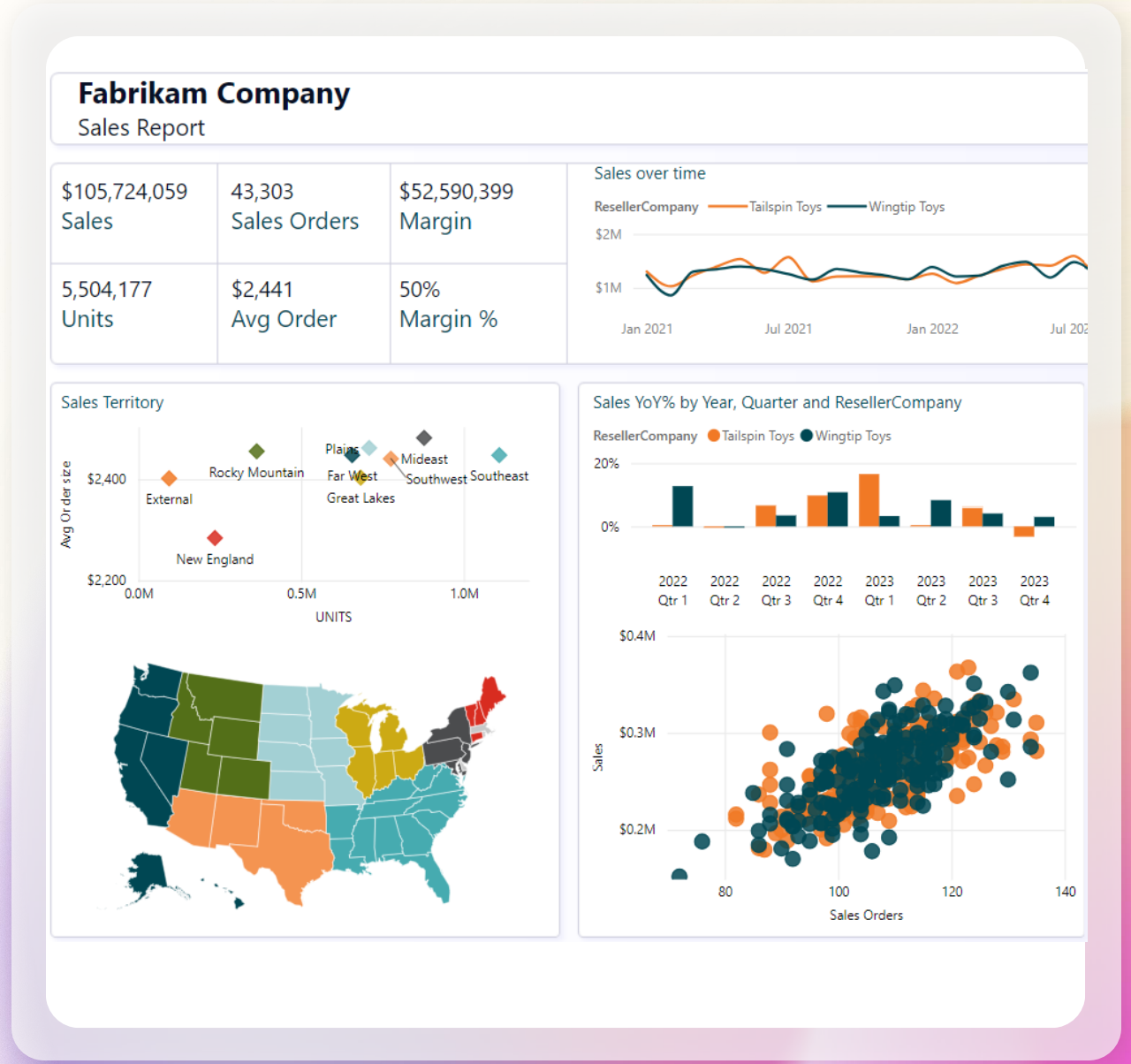
You are facing the following challenges:

- You need to refresh your semantic model at least three times a day to accommodate the different update times for the different data sources.
- Your refreshes take a long time as you need to do a full refresh every time to capture any updates that happened to the source systems.
- Any errors in any of the data sources you are pulling from will result in your semantic model refresh breaking. A lot of times the employee file doesn't upload on time resulting in your semantic model refresh breaking.
- It takes a very long time to make any changes to your data model as Power Query takes a long time to refresh your previews, given the large data sizes and complex transformations.
- You need a Windows PC to use Power BI Desktop even though the corporate standard is Mac.
- You heard about Microsoft Fabric, and decided to try to see if it will address your challenges.

Lab 1

By the end of this lab, you will have learned:

- How to set up Power BI Desktop in lab environment
- How to analyze Power BI Desktop Report
- How to review Power Queries to understand the data sources



15 minute break

Data Engineering

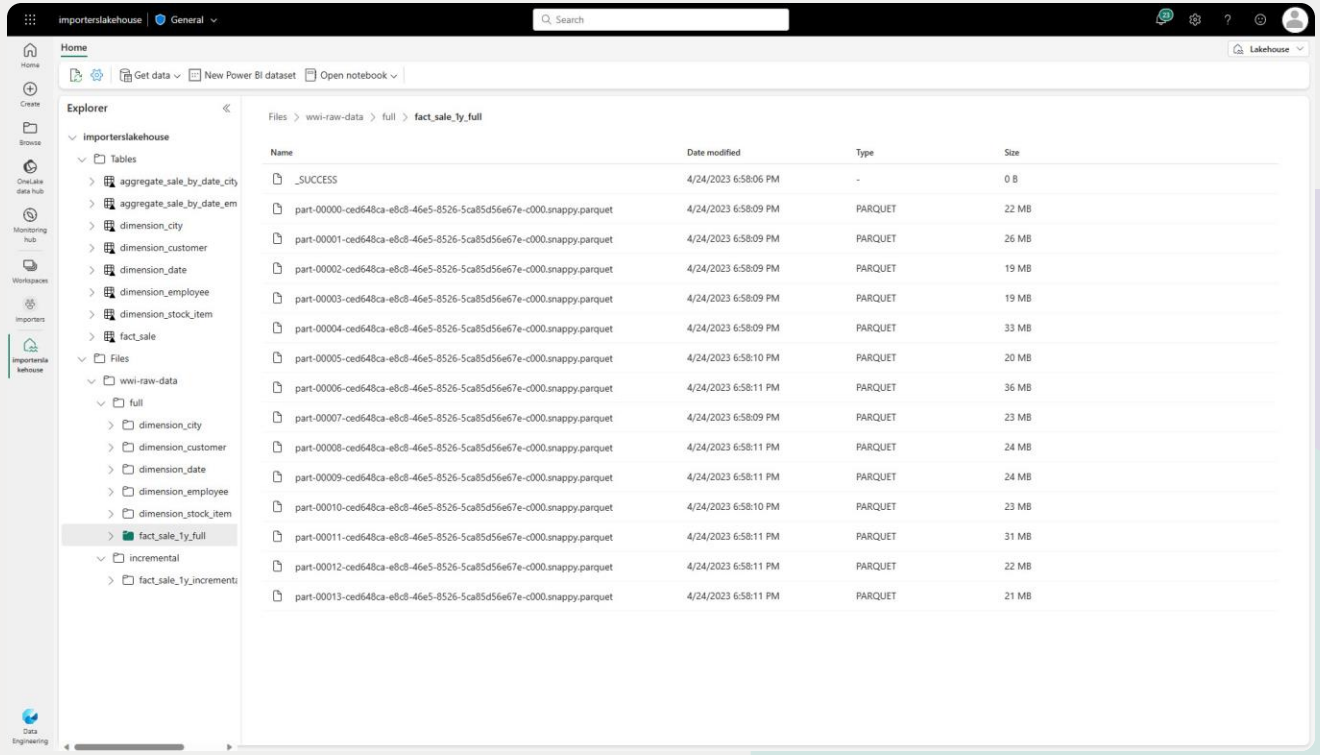
Lakehouse | Overview

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



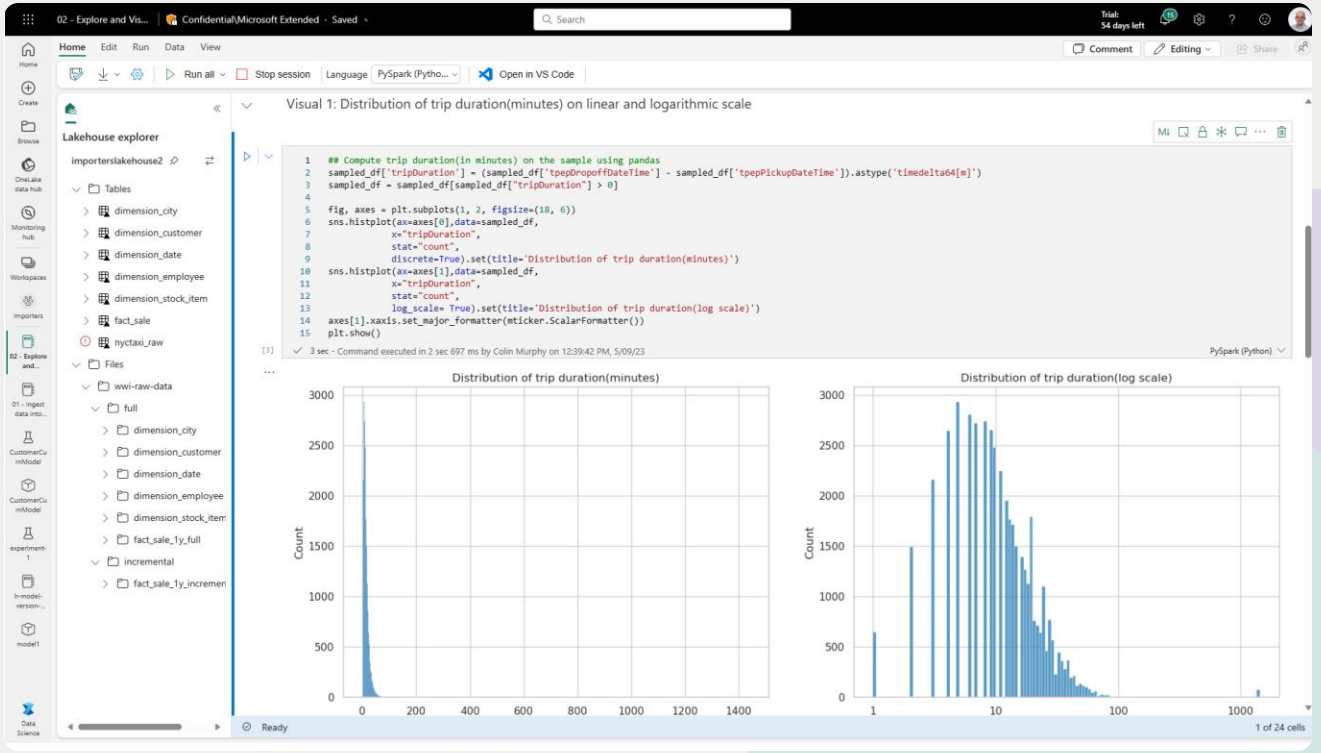
Notebook | Overview

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:

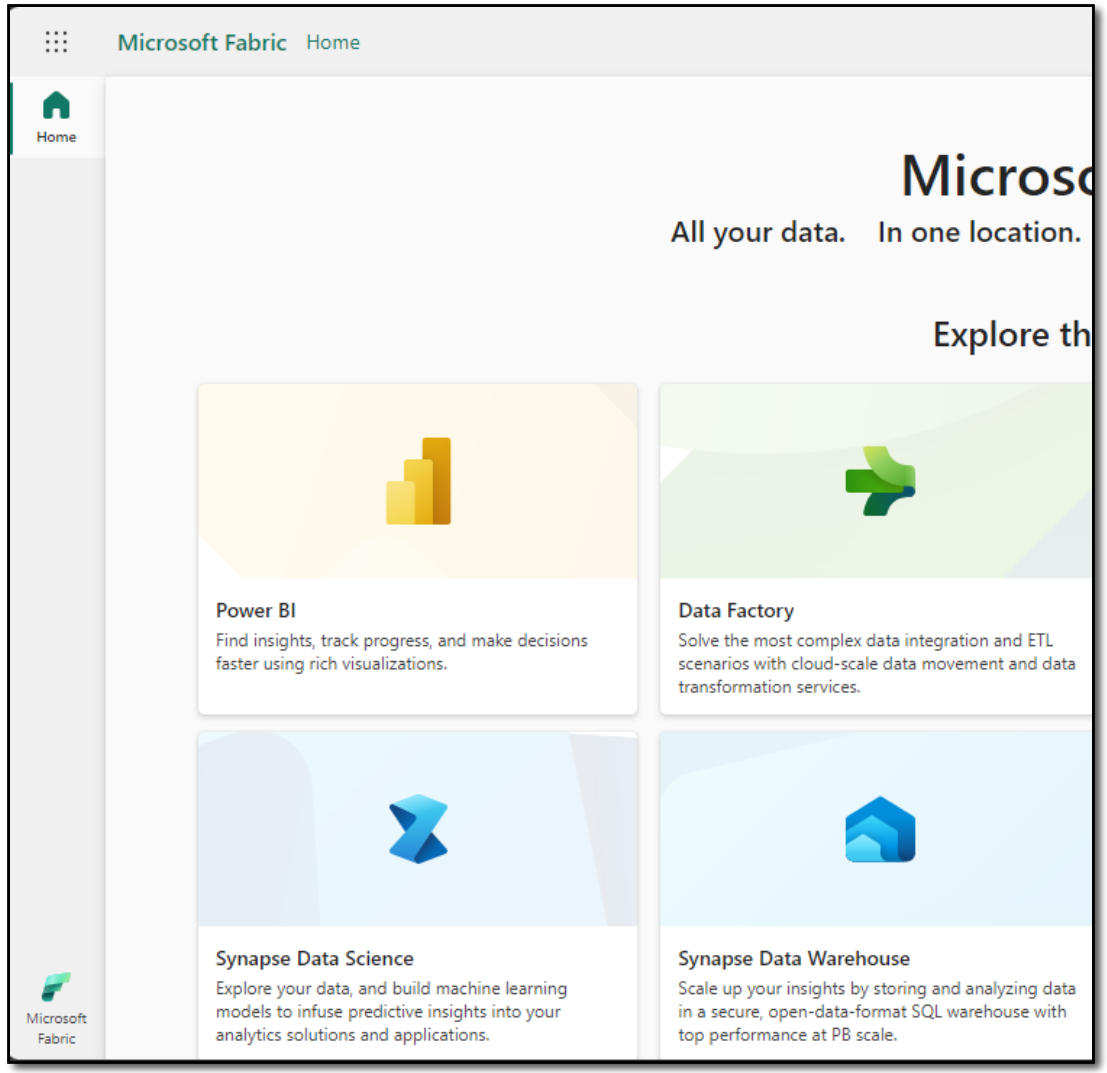
- Manage your Python and R libraries through in-line installs using commands like %pip install
- Advanced notebook development support with ability to reference notebooks in notebooks, and snapshots for easy troubleshooting
- In context monitoring complete with real time advice and error analysis
- Streamline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



Lab 2

By the end of this lab, you will have learned:

- How to create a Fabric workspace
- How to create a Lakehouse



Data Factory

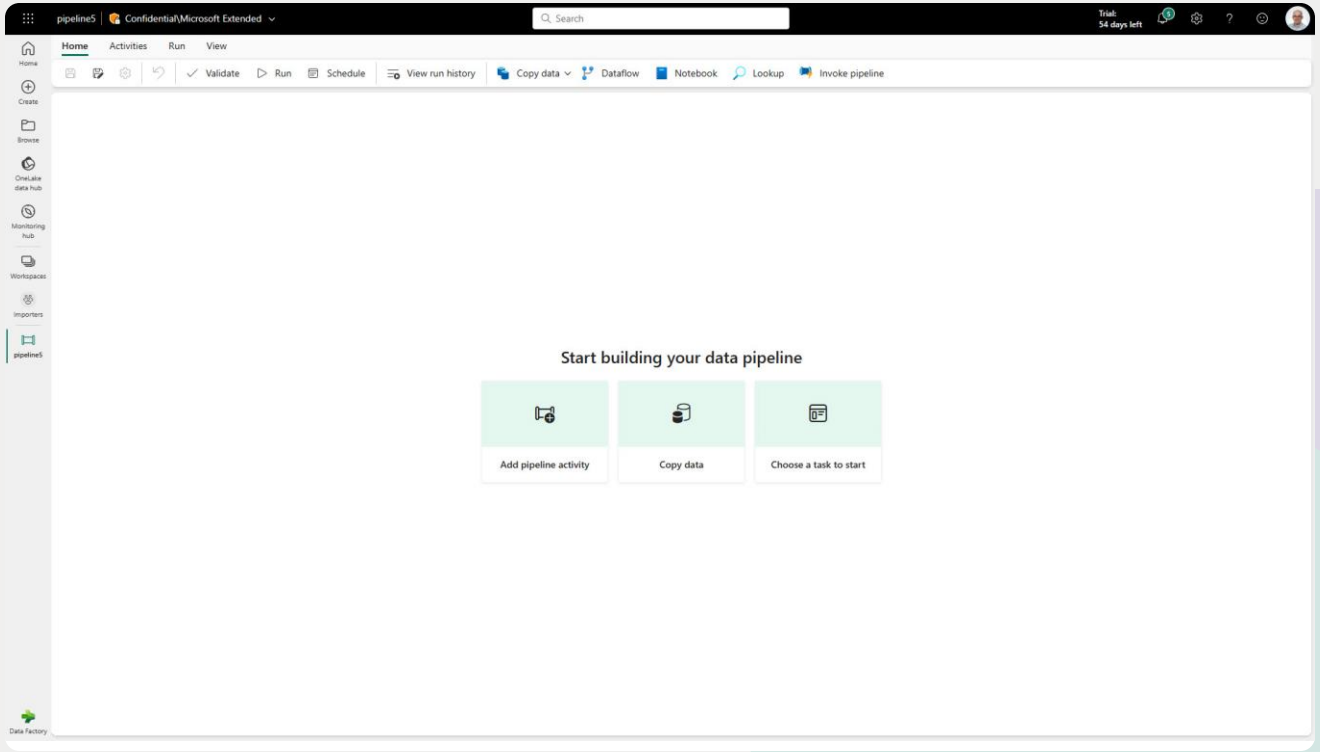
Data Factory

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

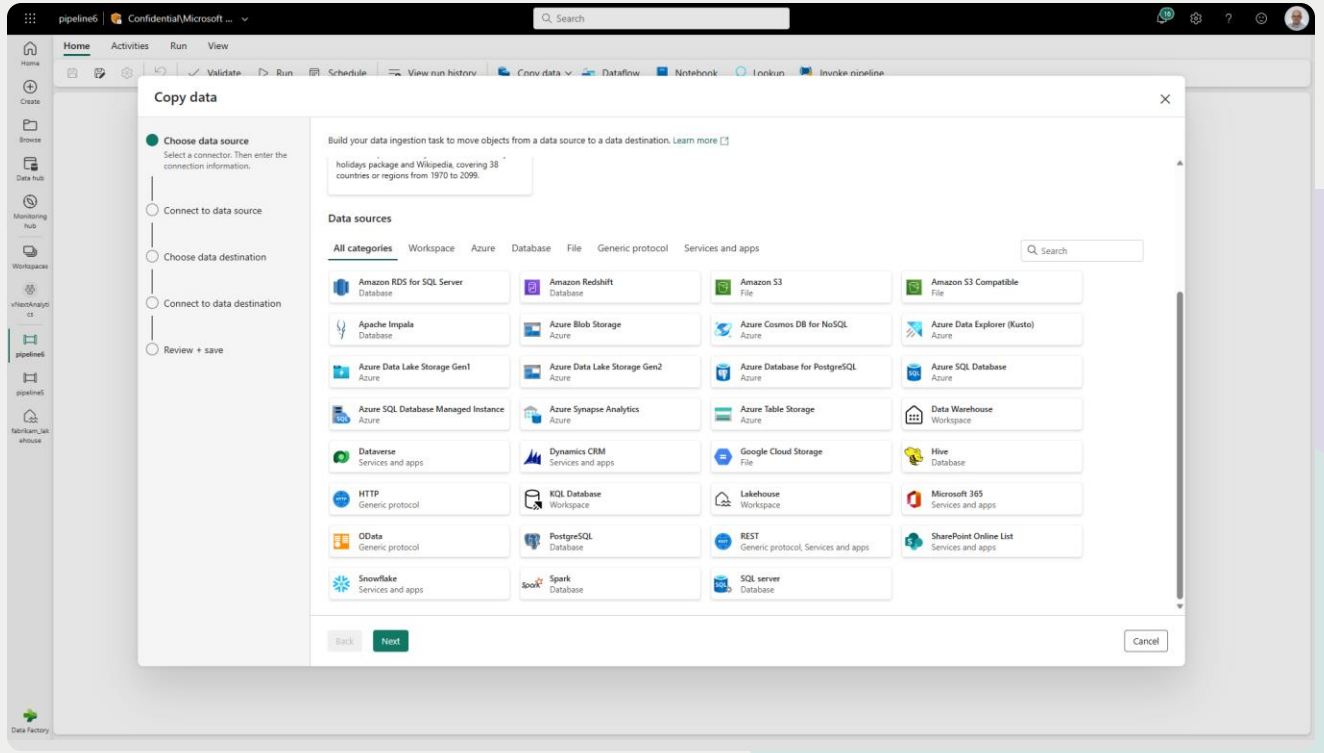


Autonomous ETL can unlock operational efficiencies and help orchestrate, monitor and manage pipeline performance.

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

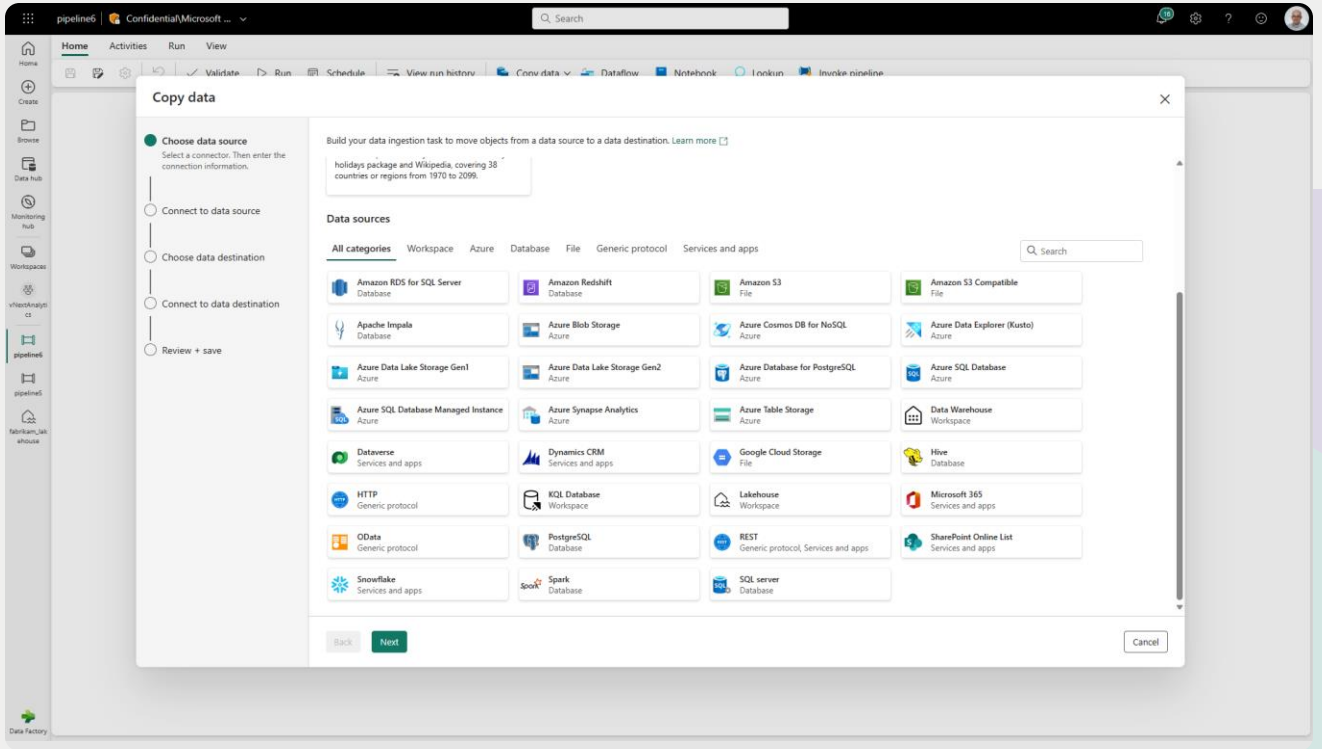


Data Pipelines | Connectors

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

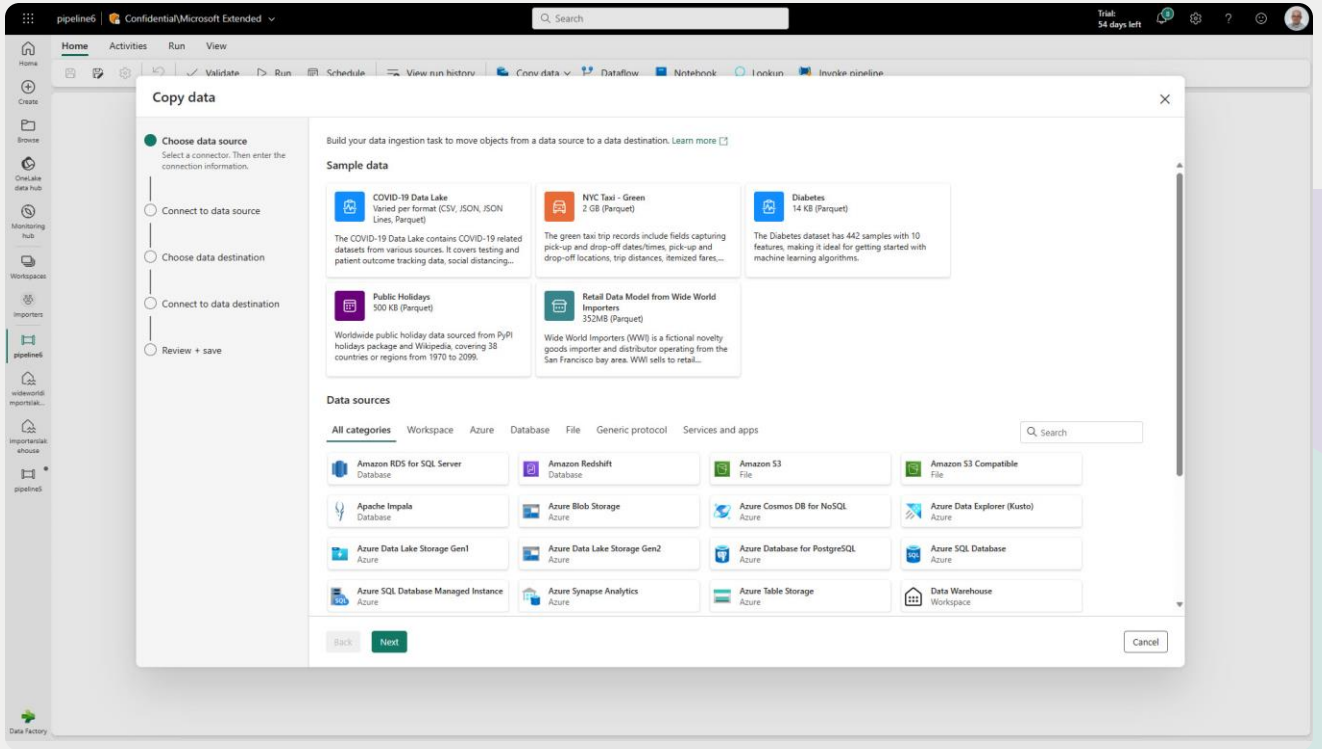


Data Pipelines | Sample data

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

Sample datasets:

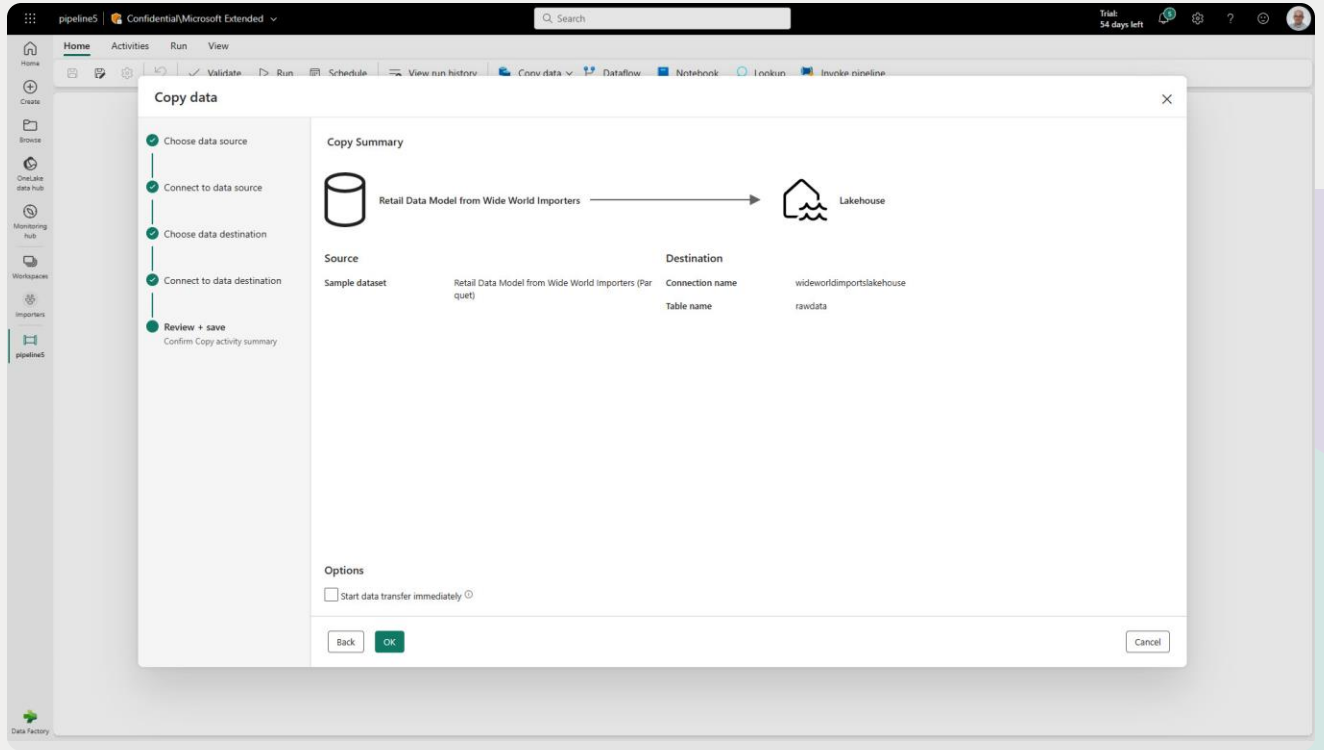
- COVID-19 Data Lake (CSV, JSON, JSON Lines, Parquet)
- NYC Tax – Green (2GB Parquet)
- Diabetes (14K Parquet)
- Public Holidays (500KB Parquet)
- Retail Data Model from Wide World Importers (352MB Parquet)



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

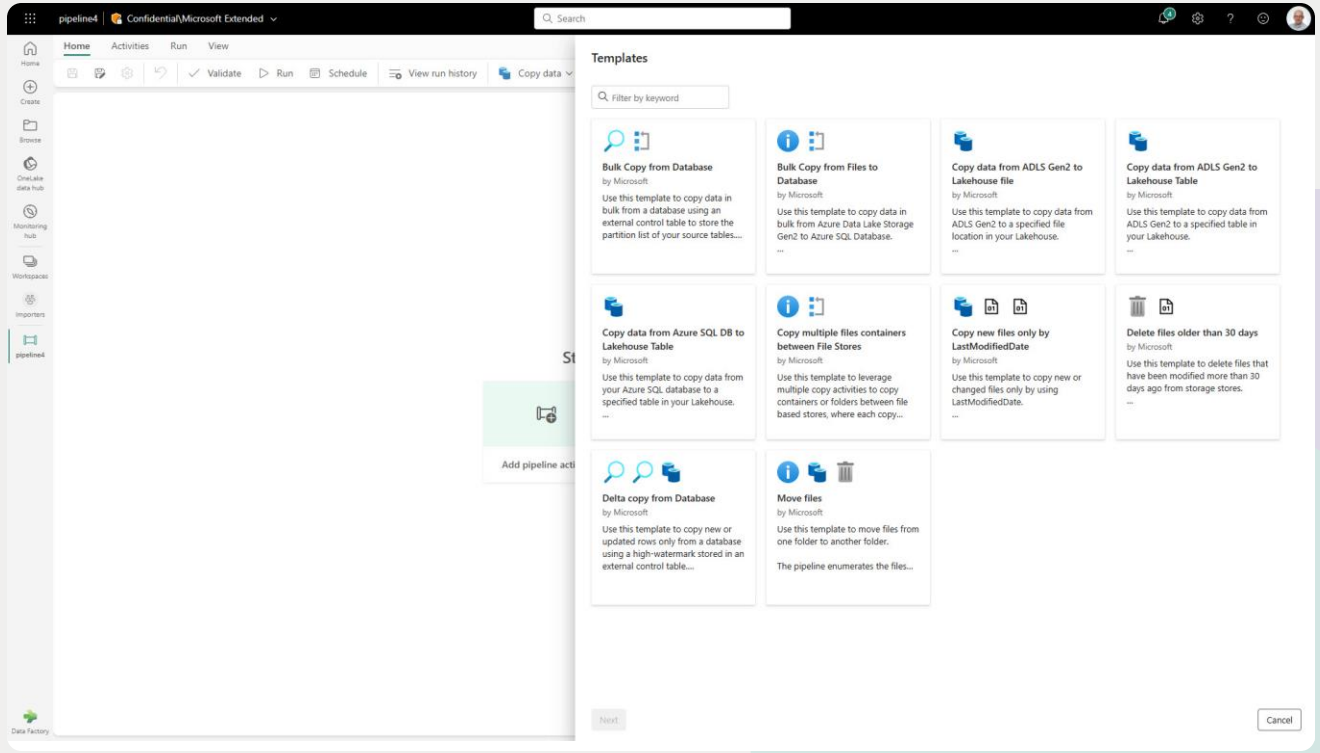
Quickly get started with data integration

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



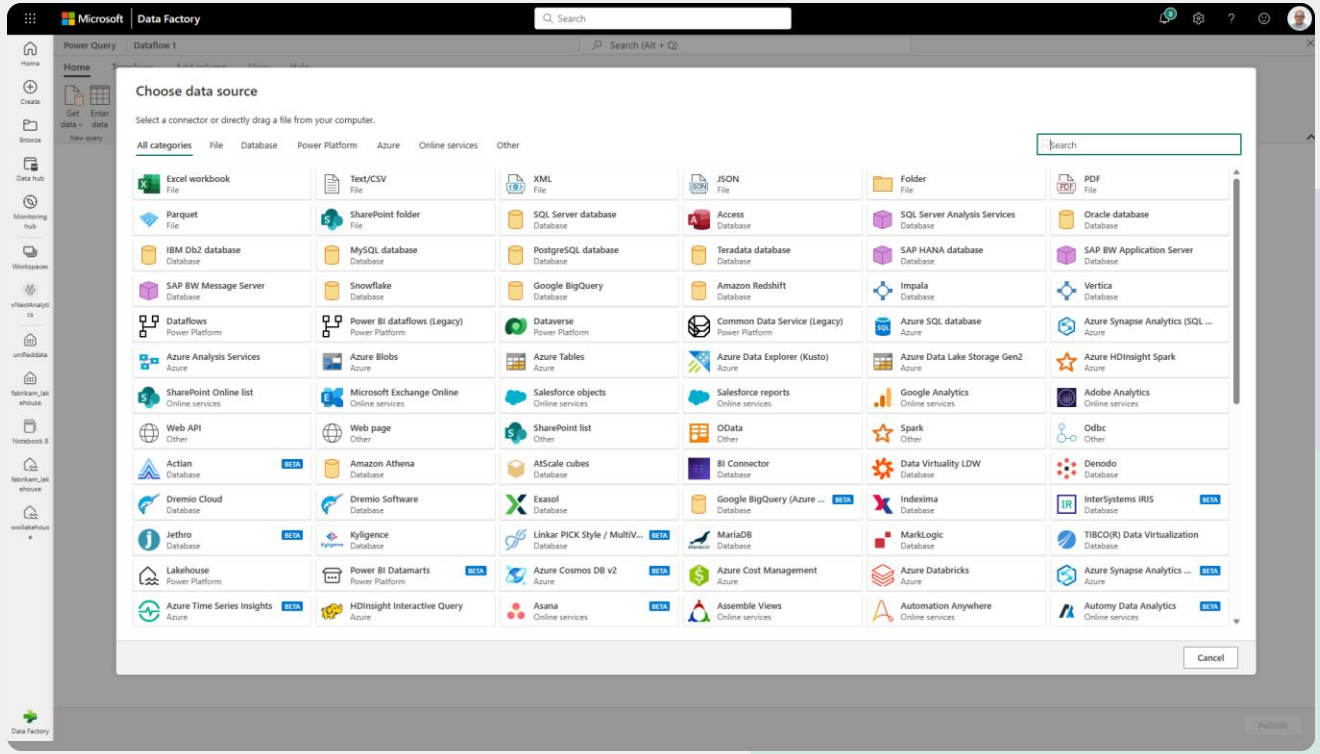
Dataflow

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.)

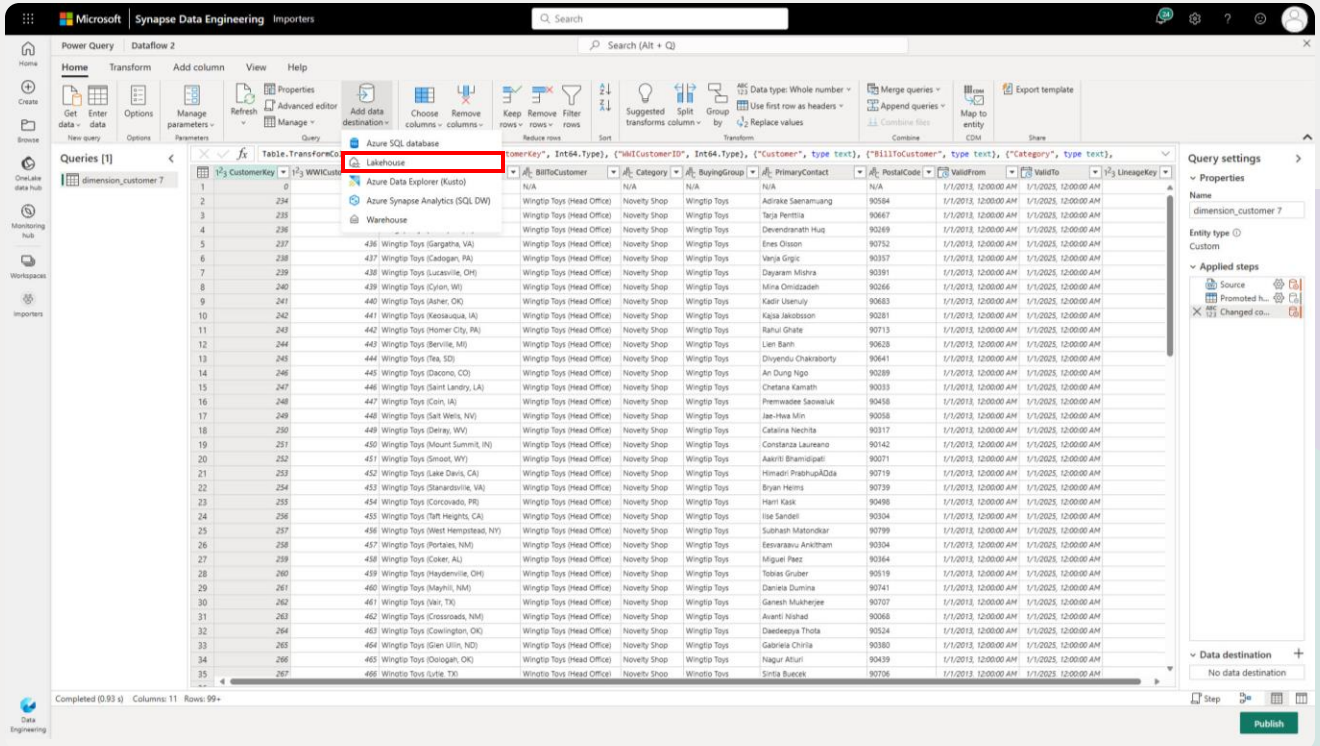


Dataflow | Output to Lakehouse

Simply write into a Lakehouse from a Dataflow

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

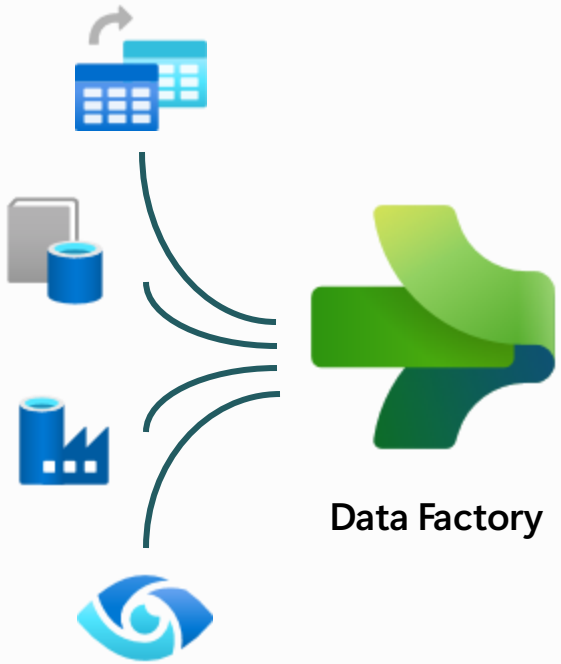
This requires the Lakehouse connector to be installed as a custom connector into your data gateway when loading data from on-premise



Unifying data in OneLake

Data Factory

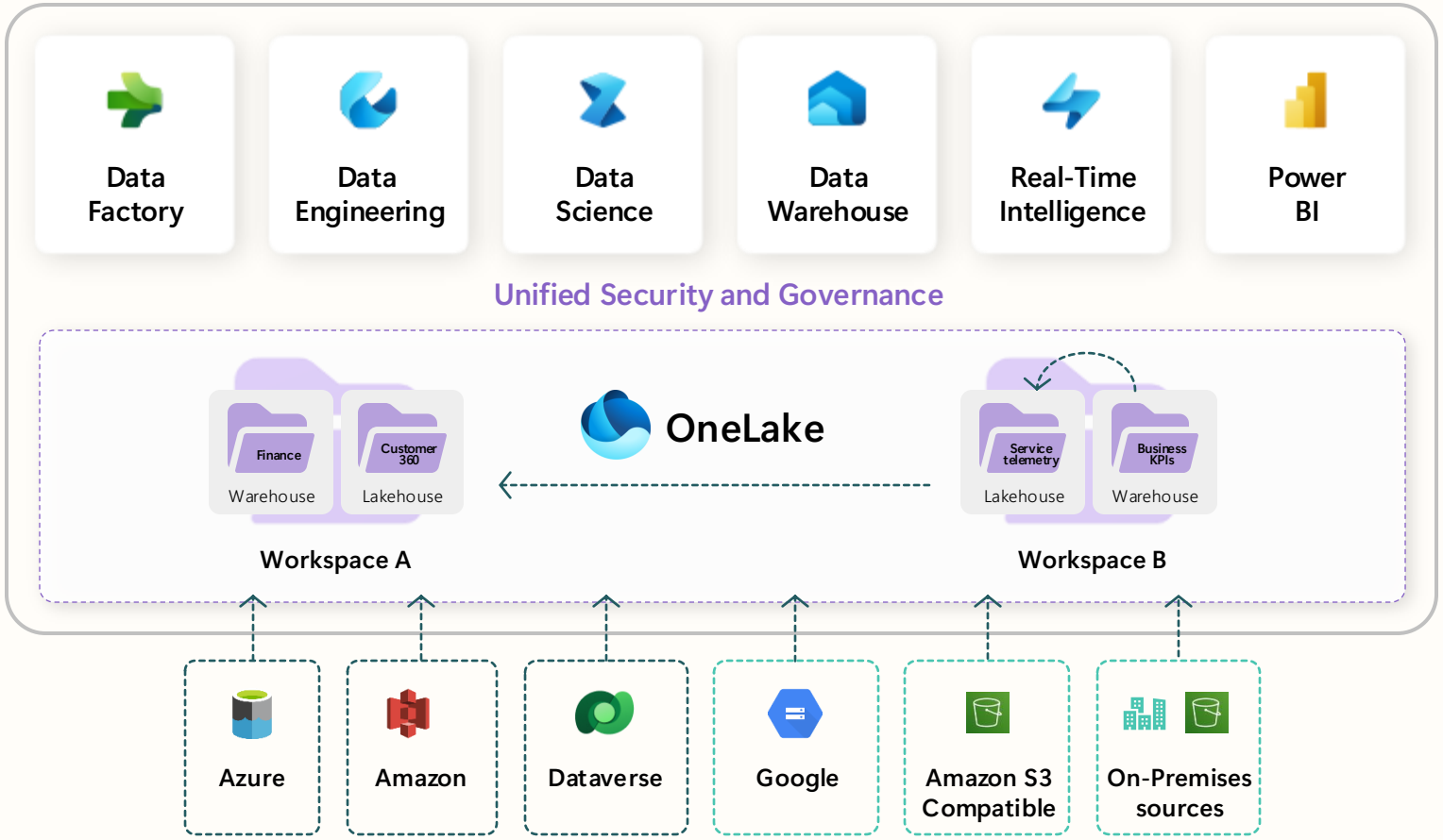
Seamlessly connect to more than 170+ data stores



Azure Database for PostgreSQL	Azure Databricks Delta Lake	Amazon RDS for Orade	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 Open Hub	SAP BW via MDX	SAP HANA	Orade 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 Cosmos Structured Stream	Azure Database for MariaDB	MySQL	Netezza	Orade	HTTP	Orade 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

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.

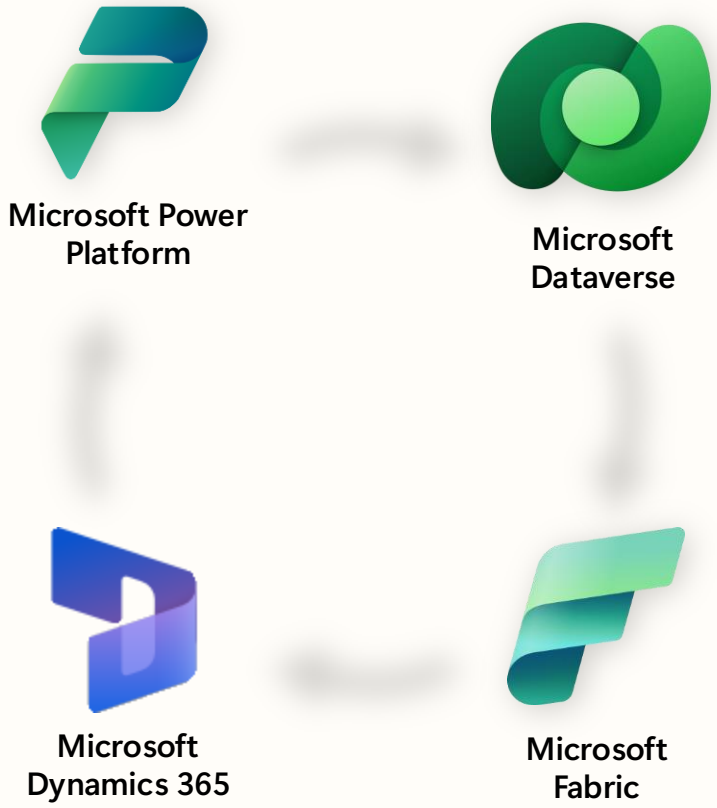
Create a shortcut within Fabric to consolidate data across items or workspaces without changing the ownership of the data. Data can be reused multiple times without data duplication.

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.

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

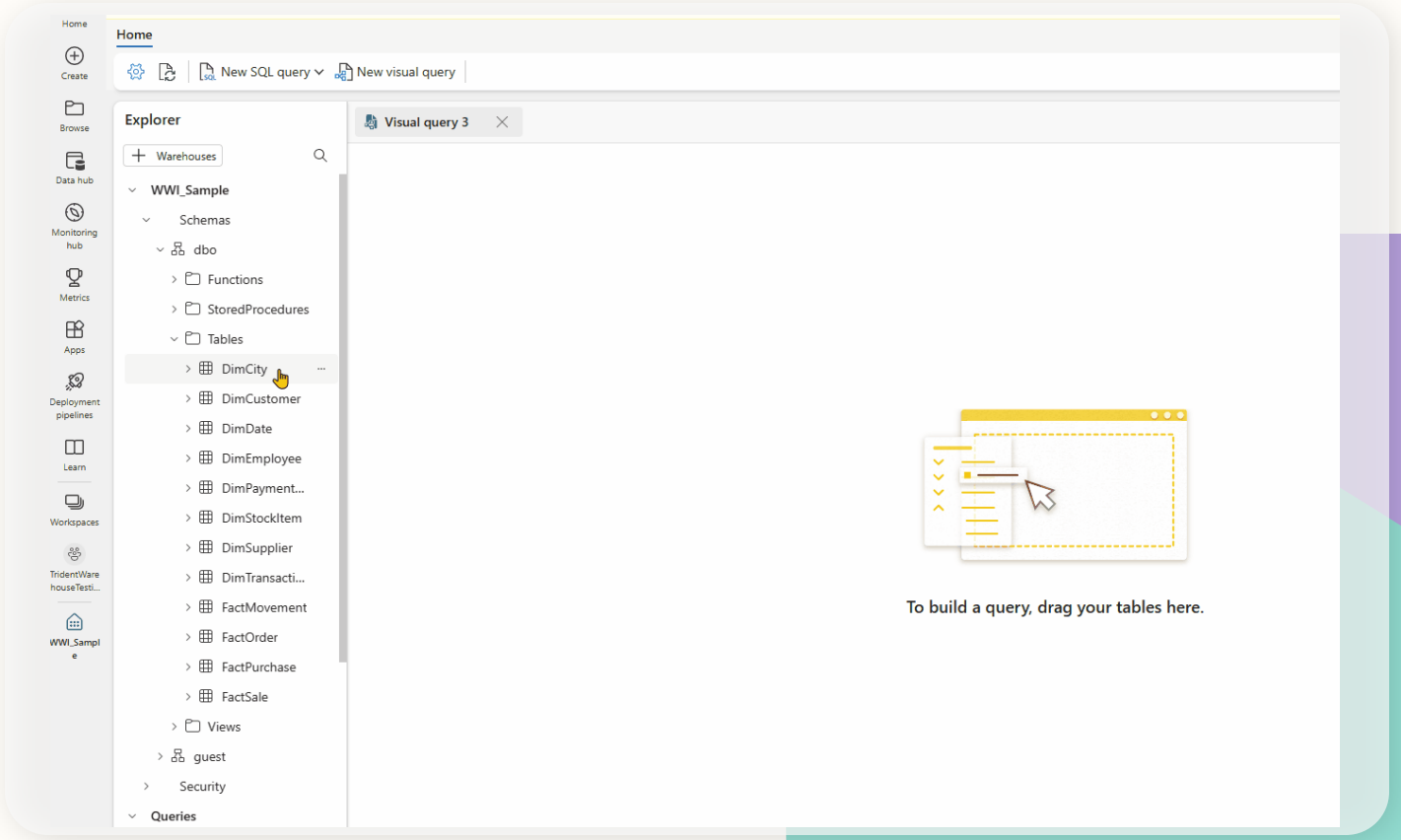
- No Copy. No ETL.
- Direct Connection via Dataverse
- Insights democratized to all low code apps and business using Fabrics 7 core workloads
- Makers informed by insights improves quality of applications
- Data is governed



Visual Query Editor

Key Capabilities:

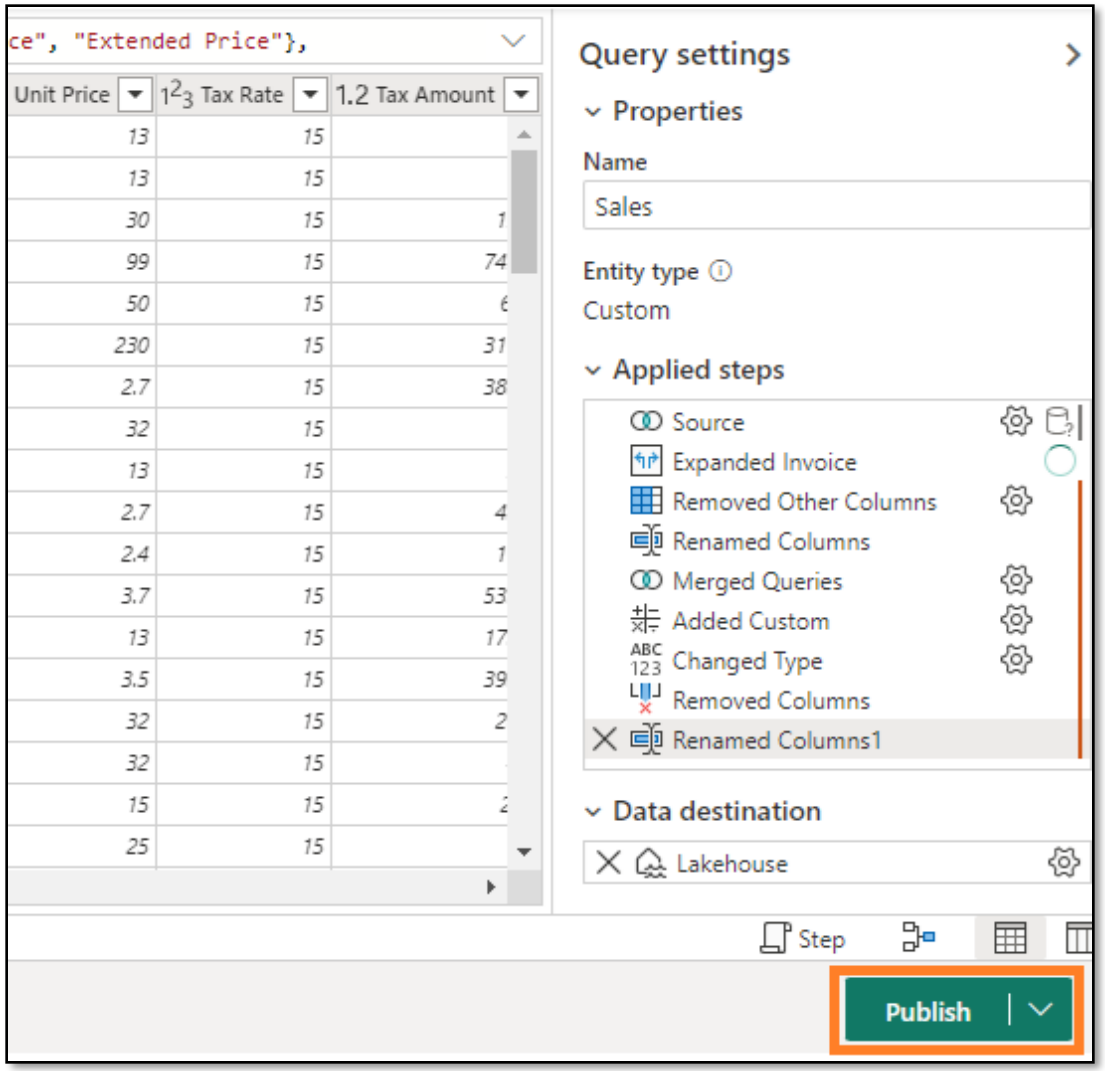
- Easy and efficient method to transform data
- No-code experience
- Interface like Power Query
- No data duplication
- Query executed on every call
- Direct Lake not available



Lab 3

By the end of this lab, you will have learned:

- How to create Shortcut to ADLS Gen2
- How to connect to create Visual Queries
- How to ingest data into Lakehouse

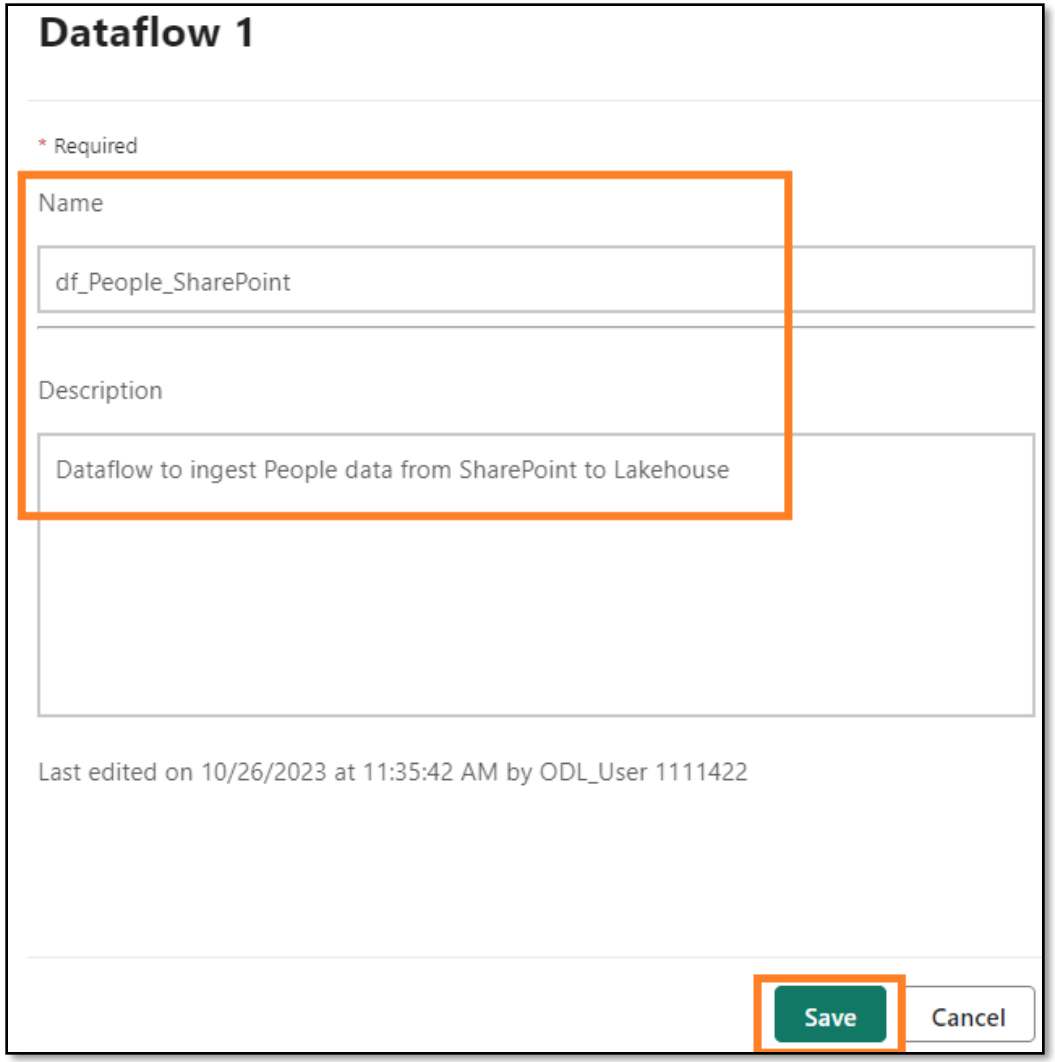


Lunch break

Lab 4

By the end of this lab, you will have learned:

- How to connect to SharePoint using Dataflow Gen2 and ingest data into Lakehouse
- How to connect to Snowflake using Dataflow Gen2 and ingest data into Lakehouse
- How to connect to Dataverse data by creating a Shortcut to existing Lakehouse



Data Warehouse

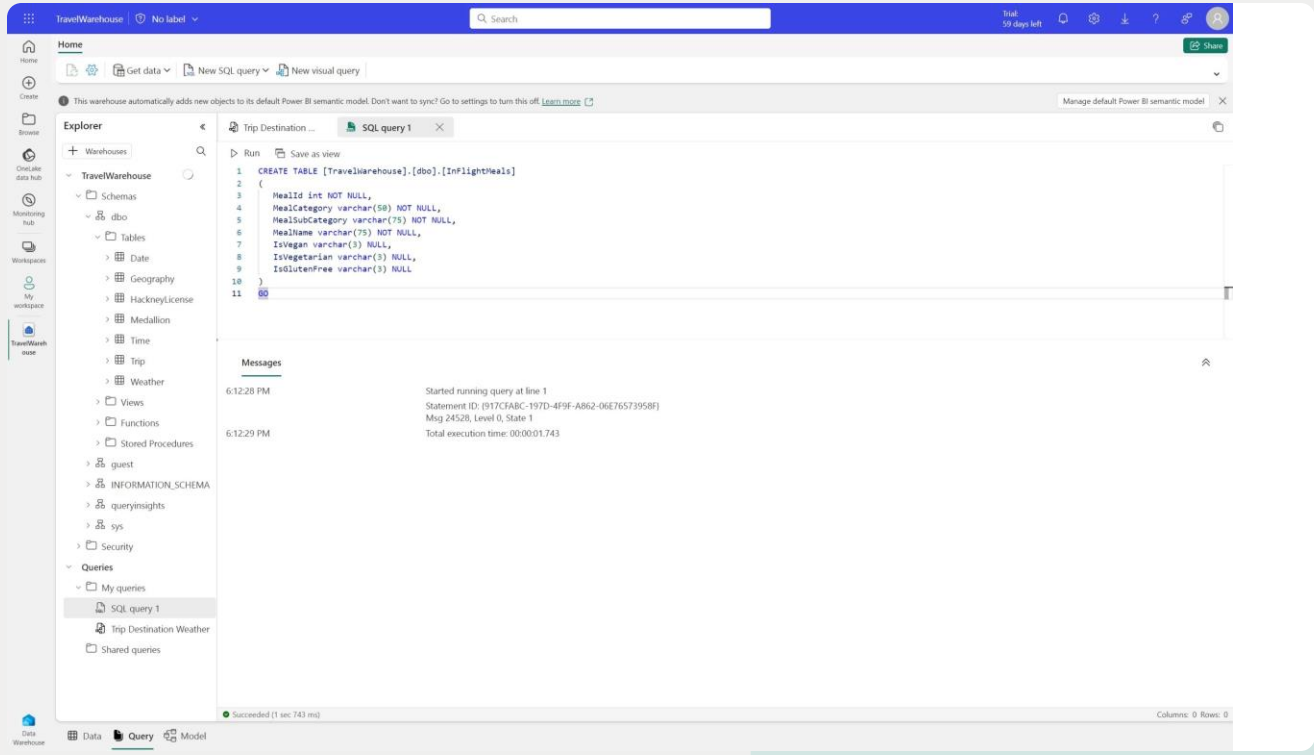
Data Warehouse | Overview

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



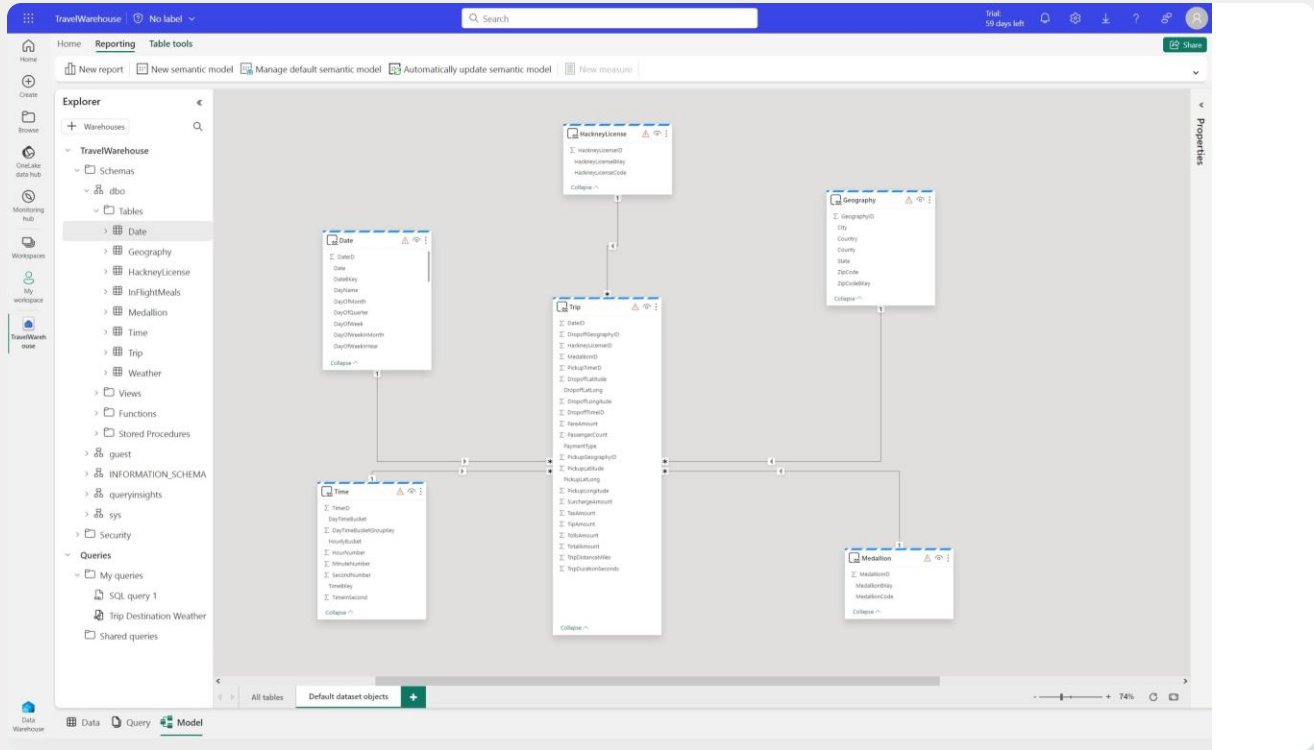
Fully integrated Power BI semantic layer

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



Secure by default

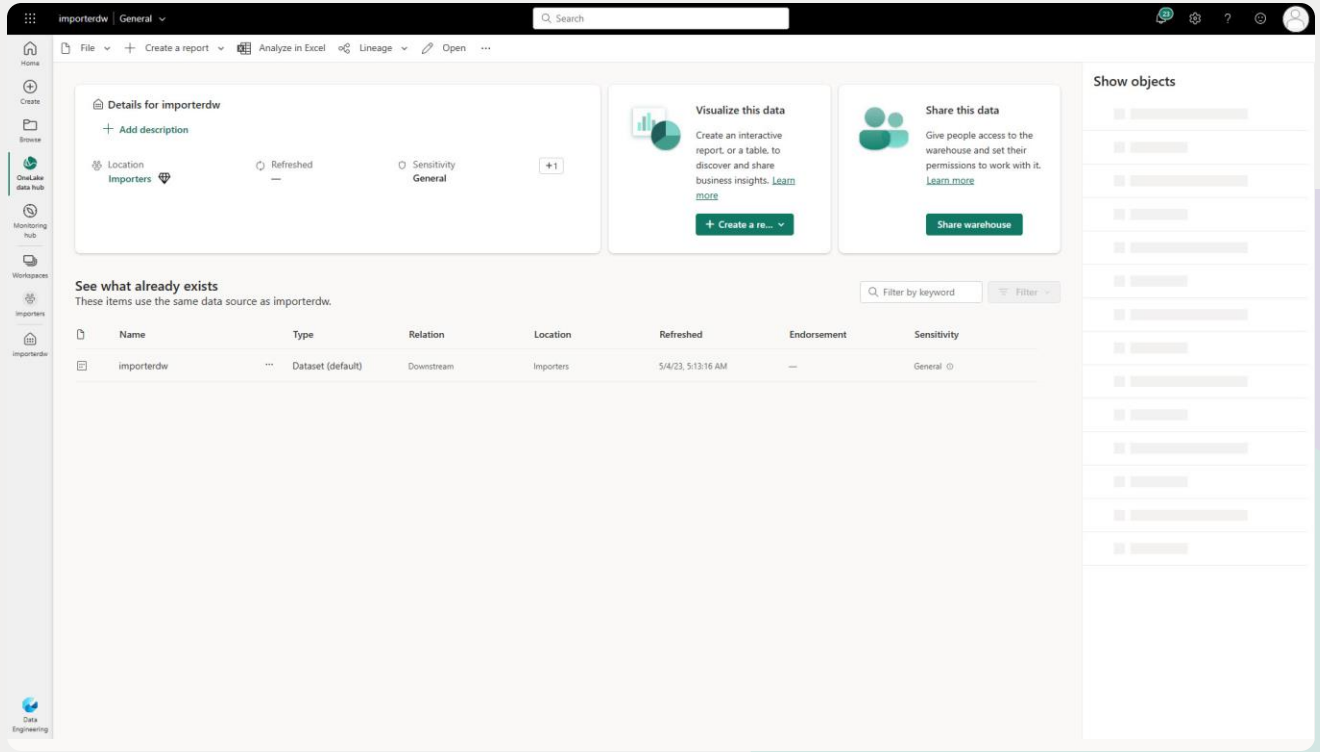
Keep data secure for any role accessing it and ensure peace of mind

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



Mirroring in Microsoft Fabric

Simplify near real-time intelligence

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

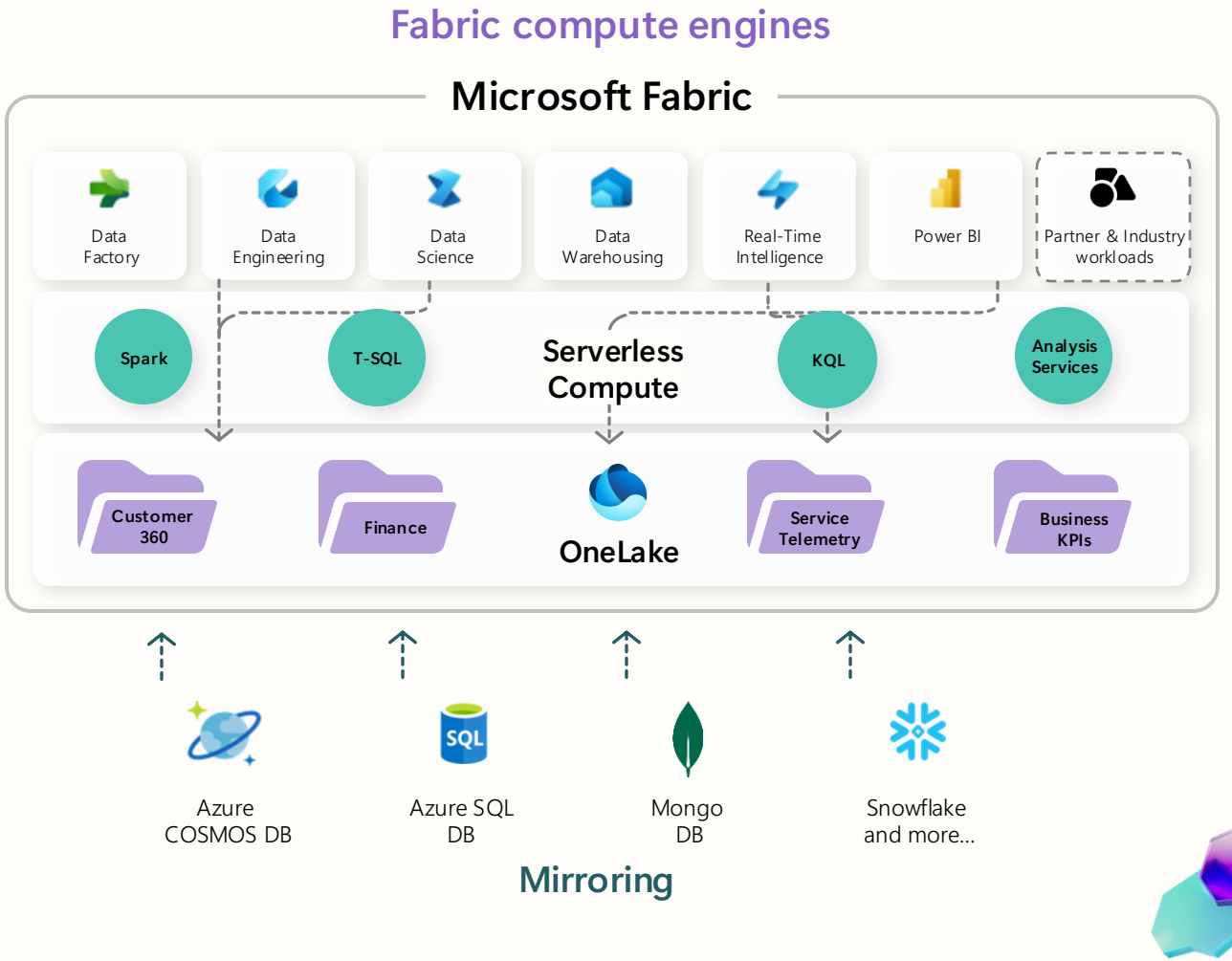
A full editing experience of the source database is available for the Mirrored database.

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.

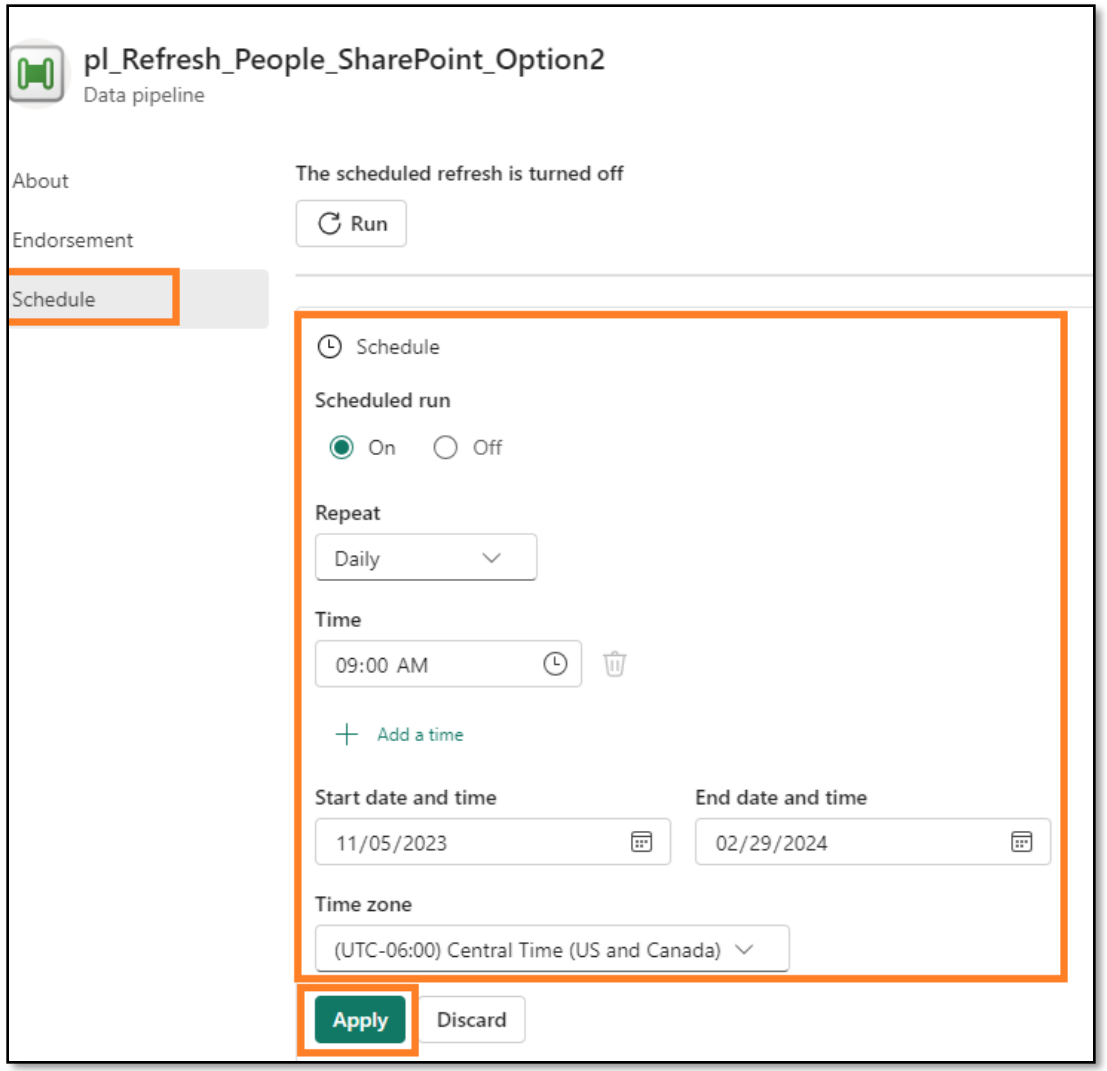
The replica protects operational databases from analytical queries.



Lab 5

By the end of this lab, you will have learned:

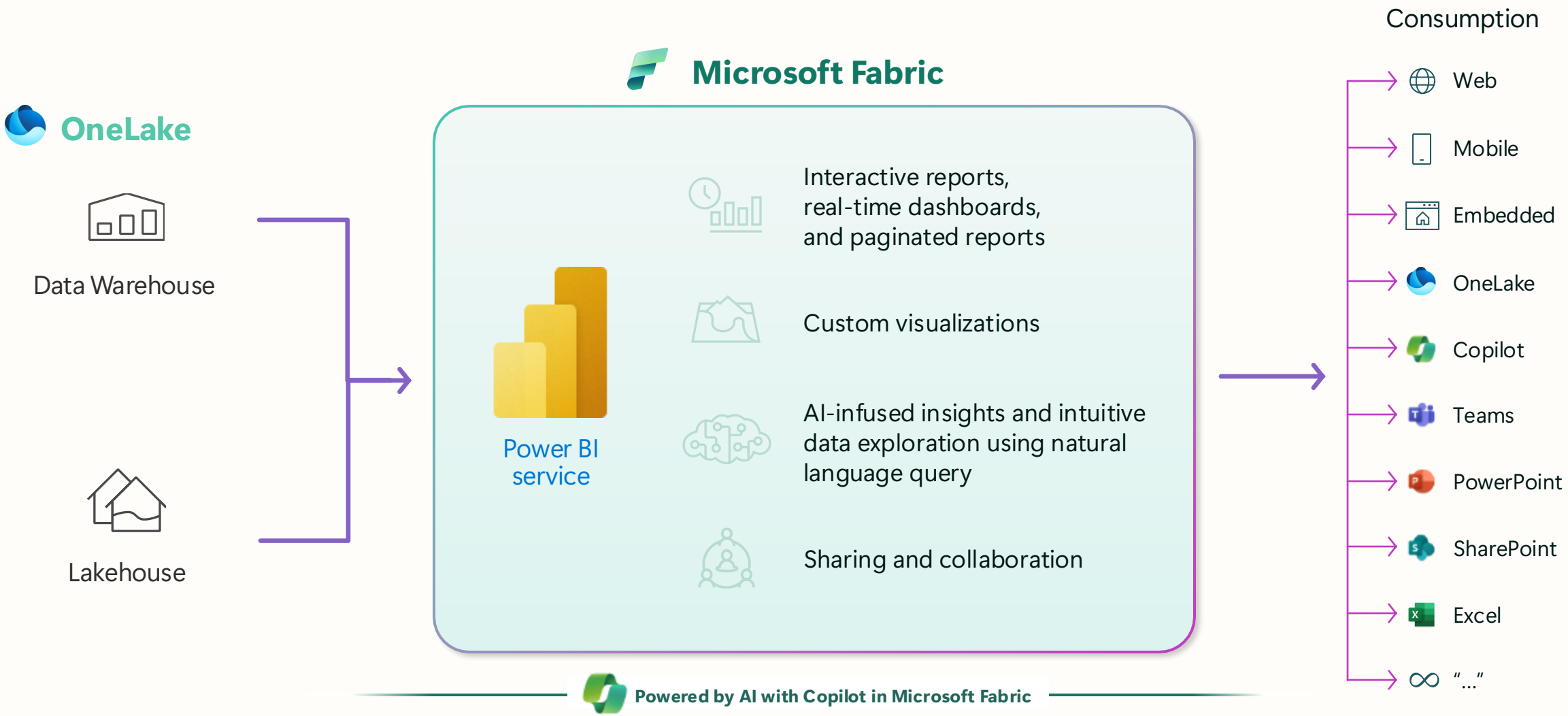
- How to configure a scheduled refresh of Dataflow Gen2
- How to create a Data Pipeline
- How to configure a scheduled refresh of a Data Pipeline



15-minute break

Power BI

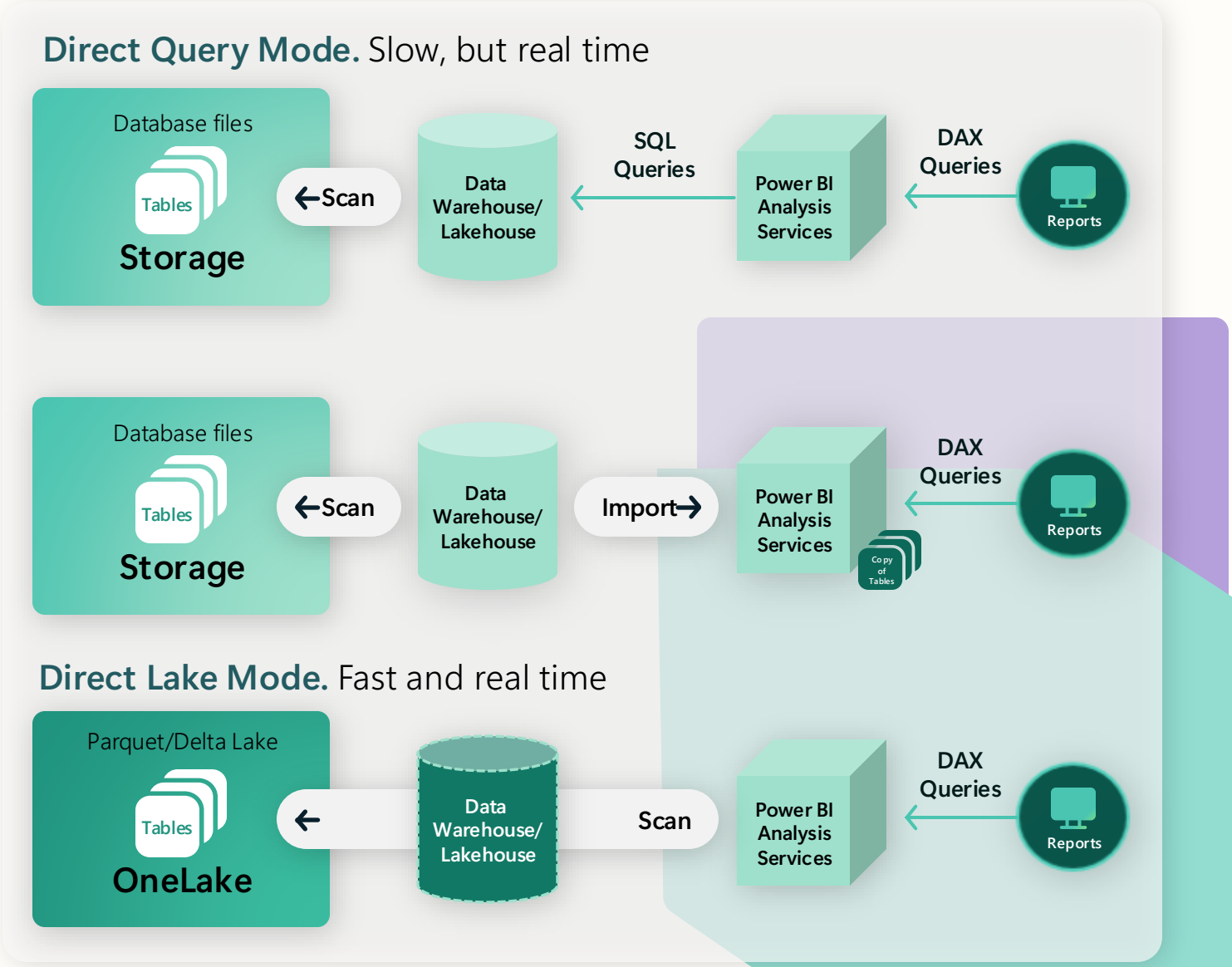
Power BI: The bridge between data and decisions



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



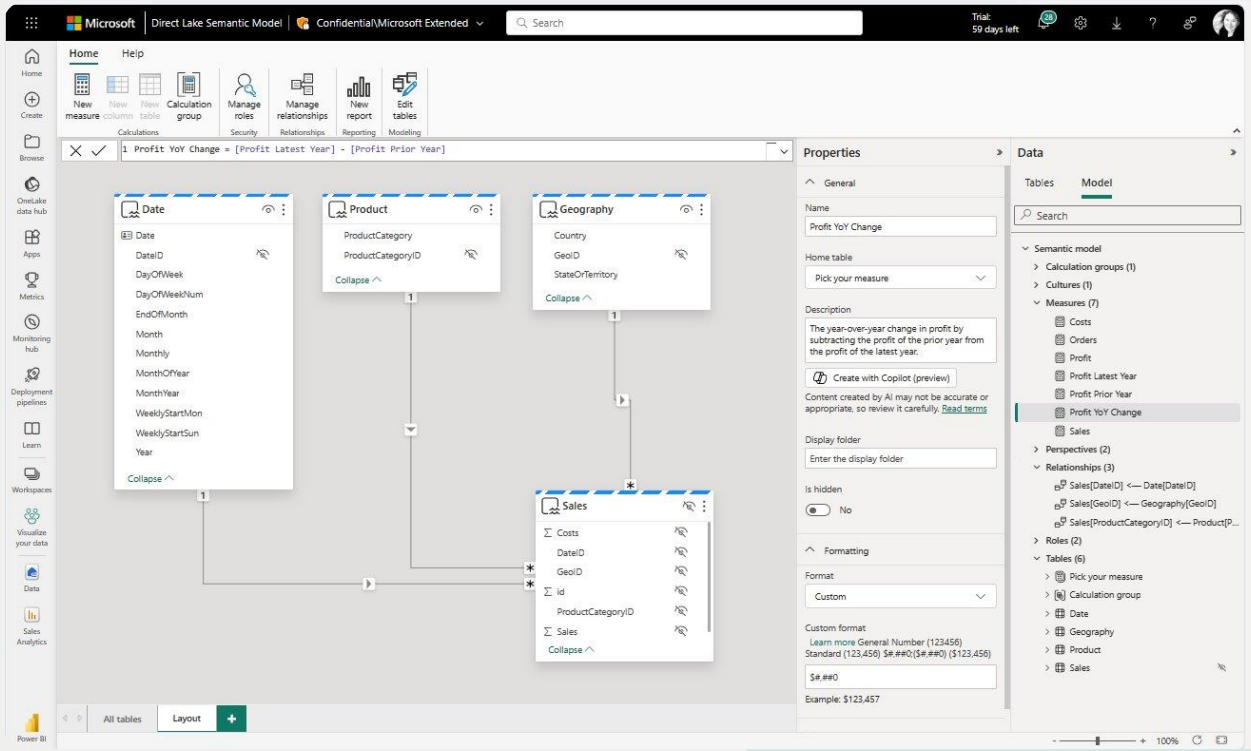
Power BI | Semantic Model

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



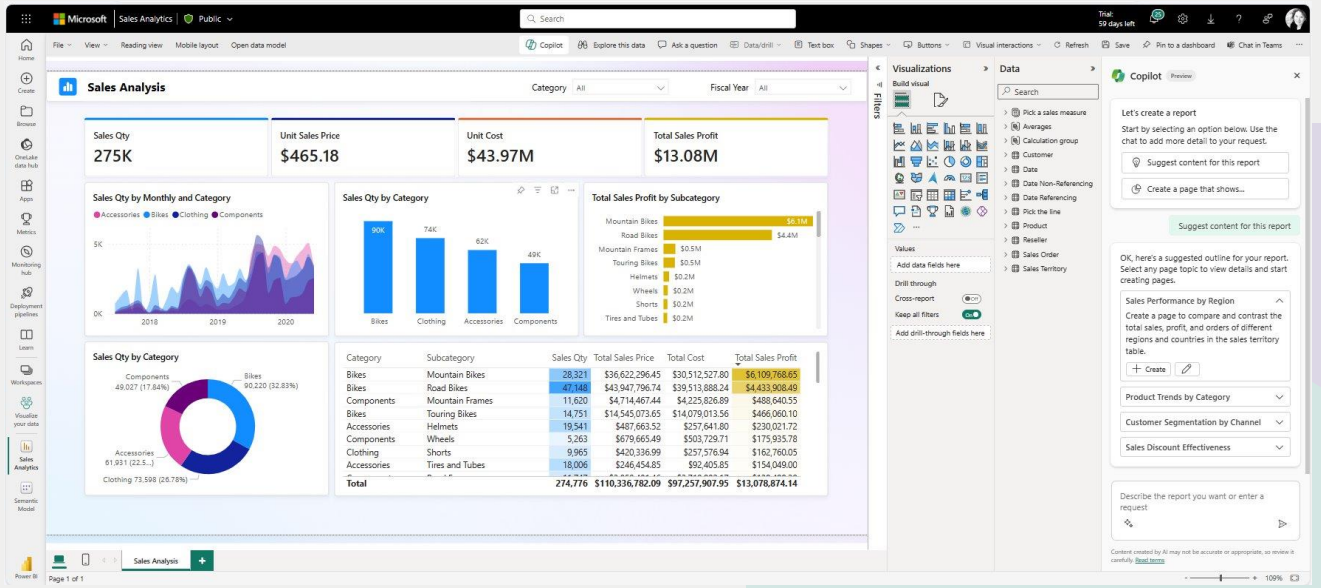
Power BI | Reports

Auto create Power BI reports from your semantic model

Blazing fast performance with Direct Lake

Key capabilities:

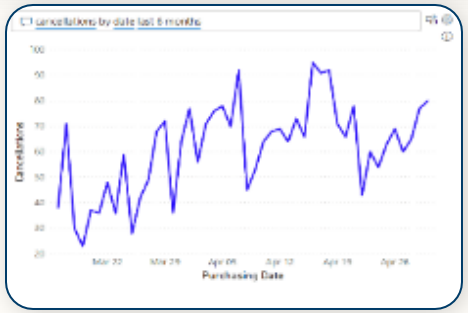
- Create an interactive report to discover and share business insights
- Use Copilot to help create, understand, and summarize reports
- Share interactively with Teams and PowerPoint
- View on phone or tablet with mobile-ready layouts for every report
- Explore data and find quick insights



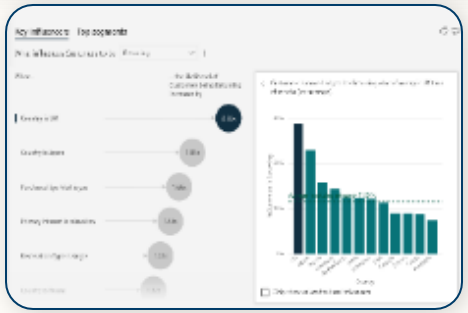
AI-powered analytics

The most complete AI capabilities in a BI product

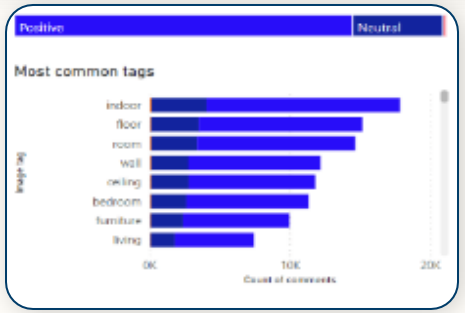
Information workers | Business analysts | Data scientists



Q&A

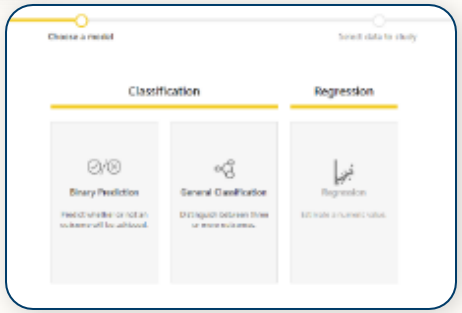


Key driver analysis

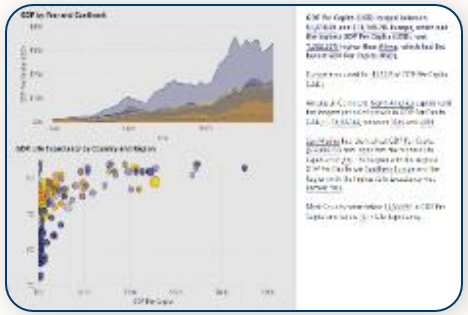


Key phrase extraction

Sentiment analysis



Create ML models



Smart narratives

Root cause analysis

Explore predictions

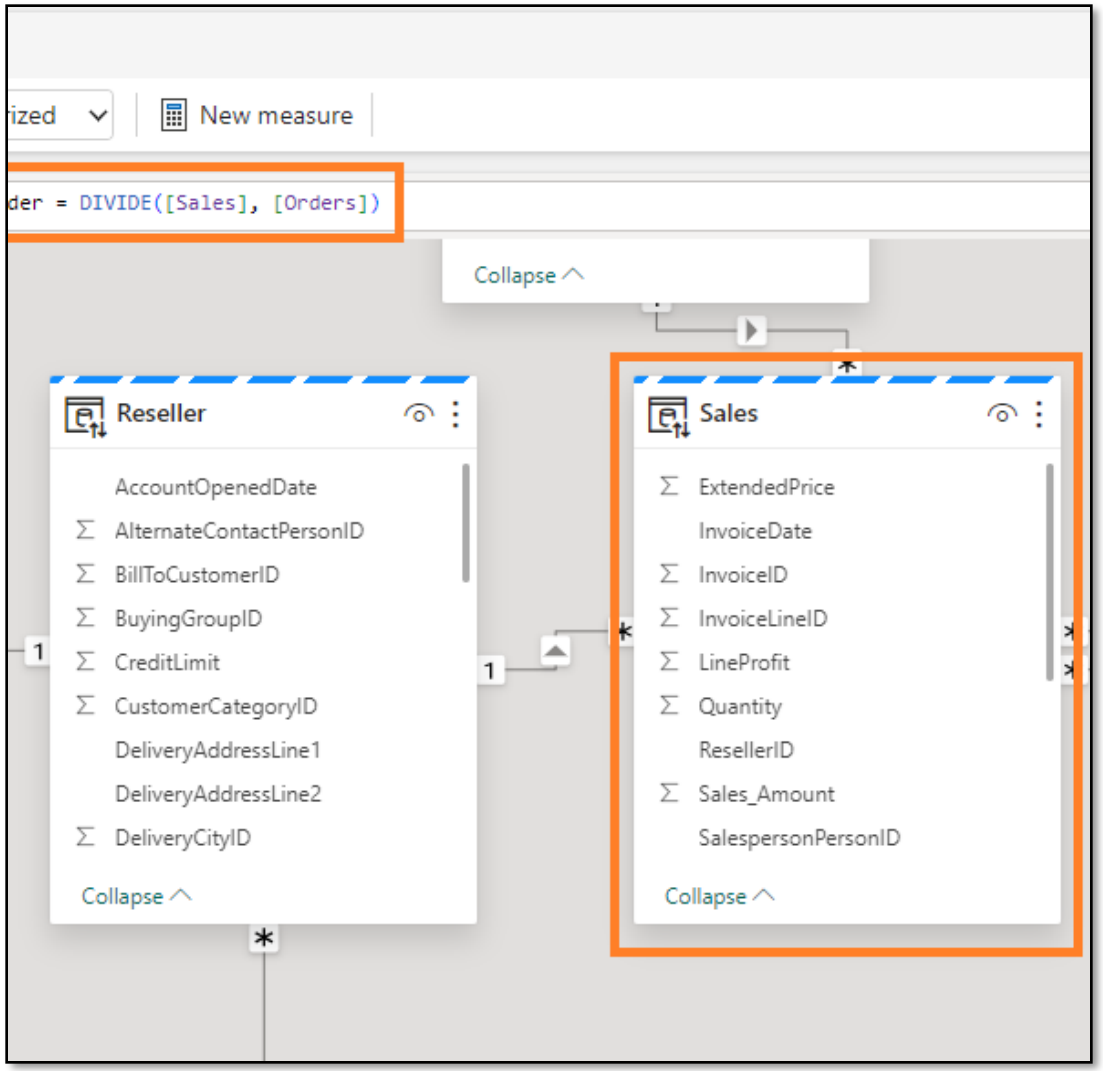
Python and R integration

Extend with Azure ML

Lab 6

By the end of this lab, you will have learned:

- How to create SQL view
- How to create Semantic model

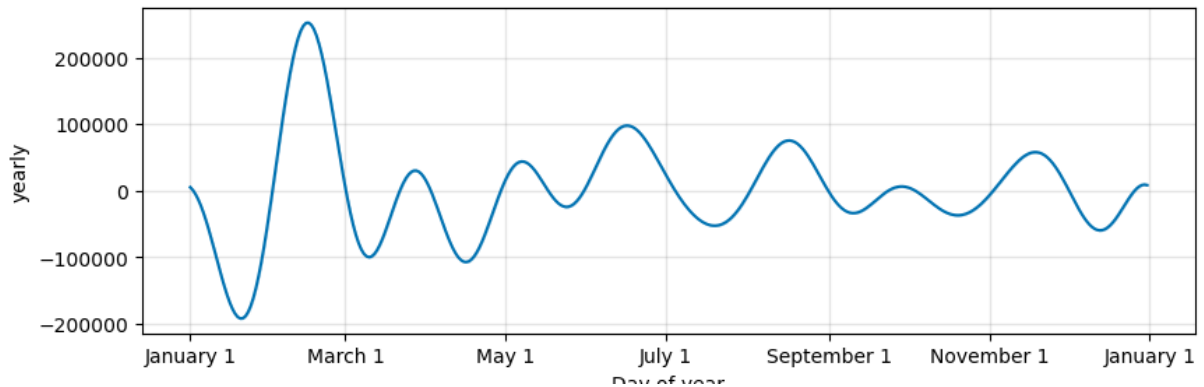


Forecasting Model Demo

Following functionality is highlighted in the demo

- How to create Notebooks
- Import and execute ML models
- Analyze results and highlight the ability to save the results in Lakehouse

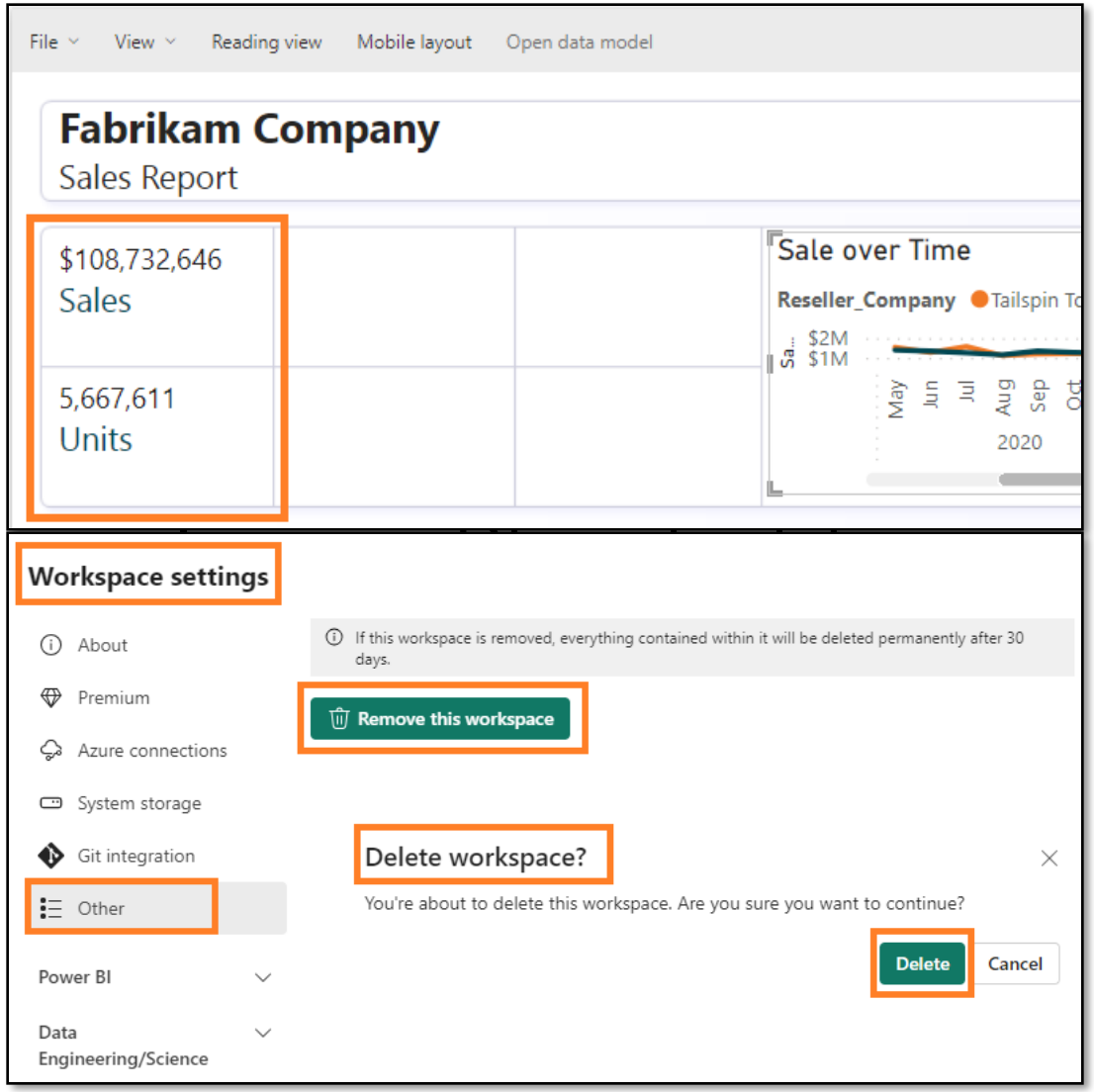
```python
from pyspark.sql import SparkSession
from pyspark.sql.functions import month, year, col
from prophet import Prophet
import pandas as pd

# Initialize Spark session
spark = SparkSession.builder.appName("Prophet Forecasting").getOrCreate()

# Load data from your specific Spark table
df = spark.sql("SELECT * FROM lh_FAIAD.Sales")

# Aggregate data to monthly level
monthly_df = df.withColumn("Month", month("InvoiceDate"))\
               .withColumn("Year", year("InvoiceDate"))\
               .groupBy("Year", "Month")\
               .sum("Quantity")\
               .orderBy("Year", "Month")

# Convert to Pandas DataFrame and prepare for Prophet
pandas_df = monthly_df.toPandas()
pandas_df['ds'] = pd.to_datetime(pandas_df[['Year', 'Month']].assign(DAY=1))
pandas_df['y'] = pandas_df['sum(Quantity)']

# Fit the Prophet model
model = Prophet(yearly_seasonality=True, weekly_seasonality=False,daily_seasonality=False)
model.fit(pandas_df[['ds', 'y']])
```



Lab 7

By the end of this lab, you will have learned:

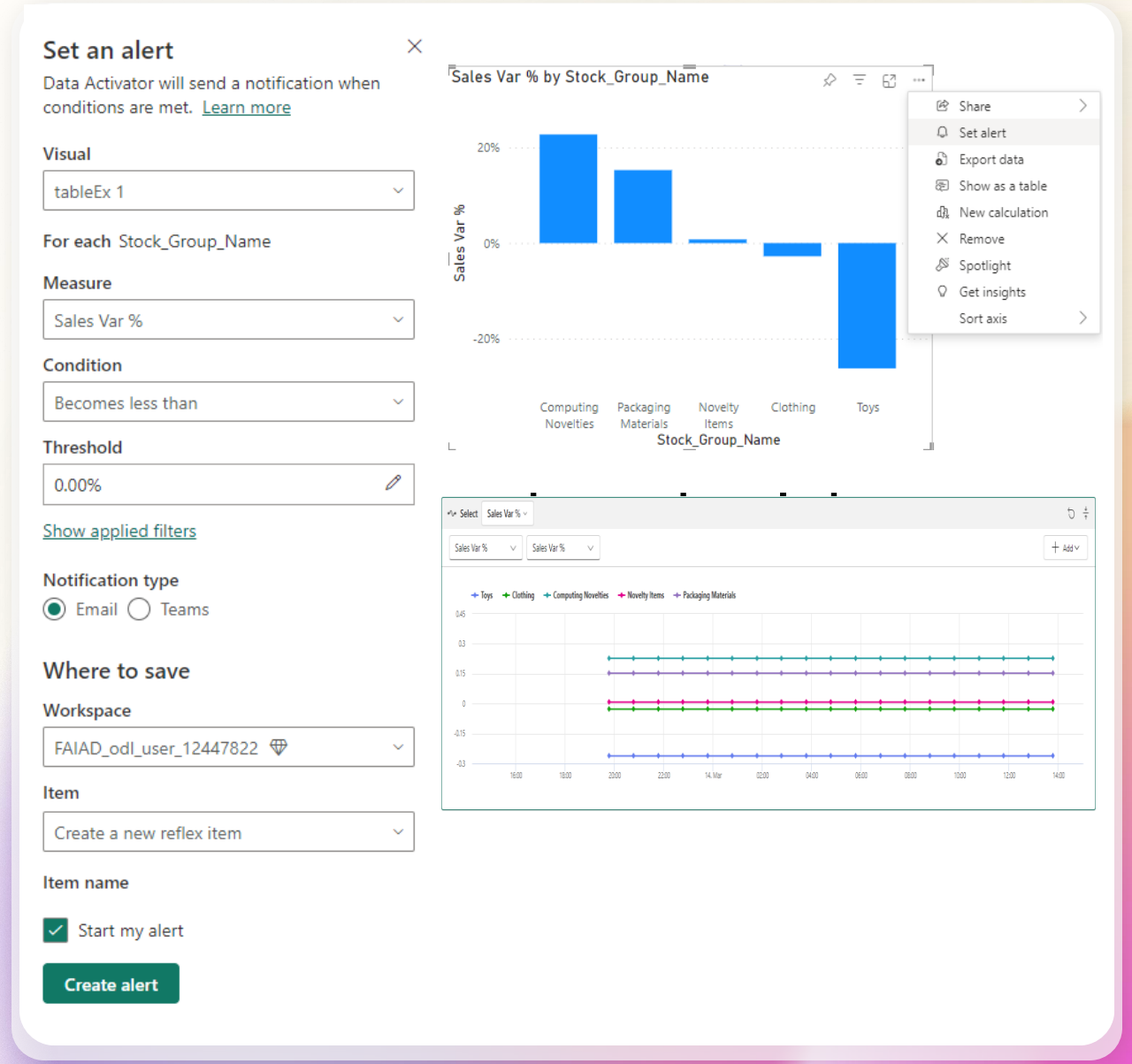
- How to auto-create a report
- How to build a report starting from a blank canvas
- How to connect Power BI Desktop to semantic model
- How to experience Direct Lake mode resulting in data automatically refreshing



Data Activator Demo

Following functionality is highlighted in the demo

- How to configure an alert
- How alerts are triggered
- Overview of the Reflex objects



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

End-to-end tutorials

Lakehouse tutorial
https://learn.microsoft.com/en-us/fabric/data-engineering/tutorial-lakehouse-introduction

Data Science tutorial
https://learn.microsoft.com/en-us/fabric/data-science/tutorial-data-science-introduction

Real-Time Analytics tutorial
https://learn.microsoft.com/en-us/fabric/real-time-analytics/tutorial-introduction

Data warehouse tutorial
https://learn.microsoft.com/en-us/fabric/data-warehouse/tutorial-introduction

Power BI tutorial
https://learn.microsoft.com/en-us/power-bi/fundamentals/fabric-get-started

Data Factory tutorial
https://learn.microsoft.com/en-us/fabric/data-factory/tutorial-end-to-end-introduction

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

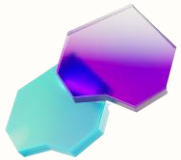
What is Data Warehouse?
Create a Warehouse
Query using SQL query editor
See more

Real-Time Intelligence

What is Real-Time Intelligence?
What is Event stream?
Create a database
See more

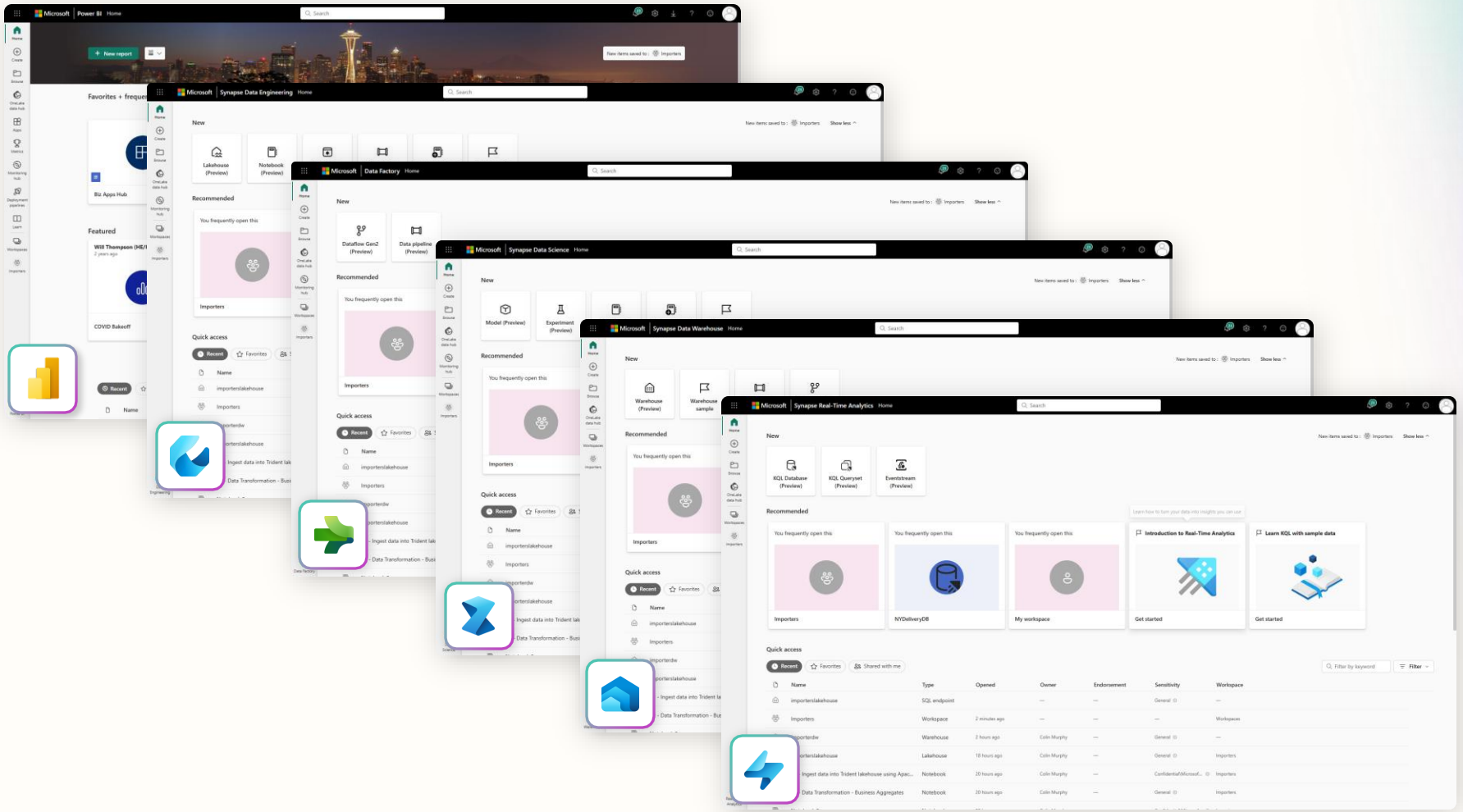
Power BI

Enable Microsoft Fabric for your organization
What is Power BI?
What is a datamart?
Azure and Power BI integration
See more



Appendix

Persona optimized workloads



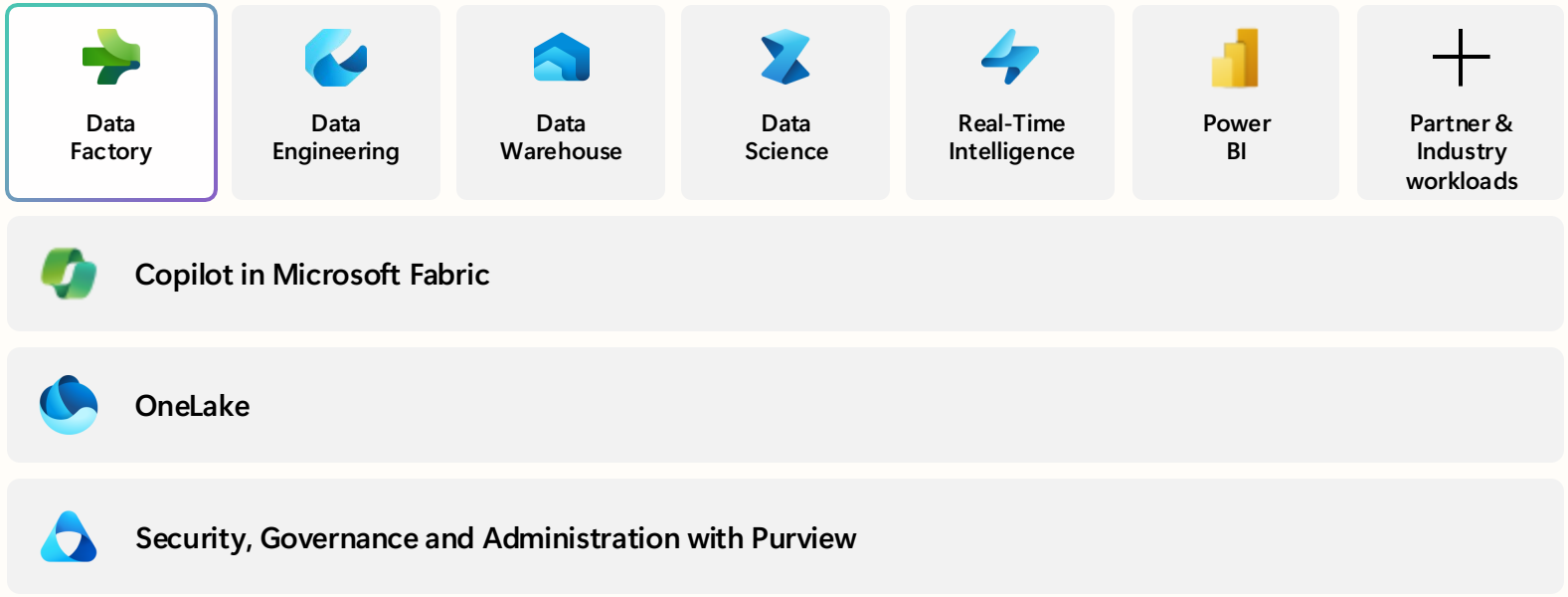
Seven key workloads for end-to-end analytics

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

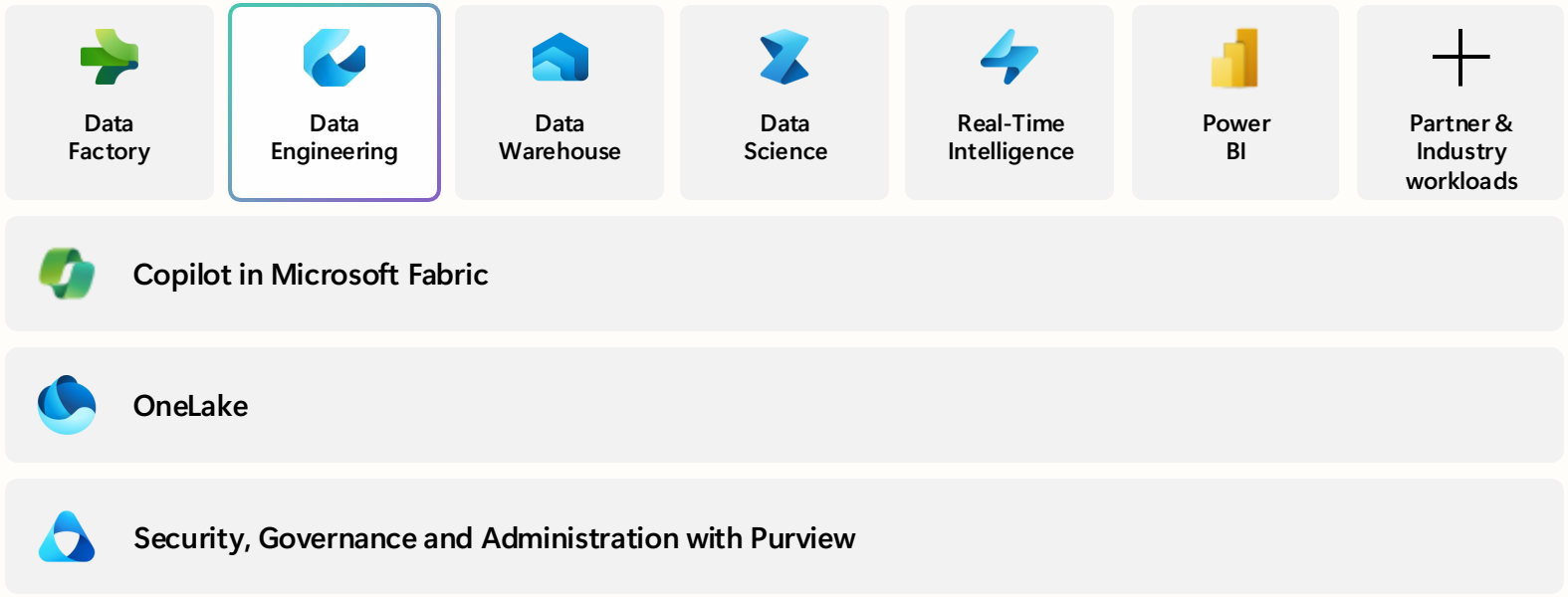
Dataflows and data pipelines bring together low-code, AI-based experiences, multi-cloud connectivity, and persistent data security and governance to help solve complex ETL scenarios for all developers



- 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

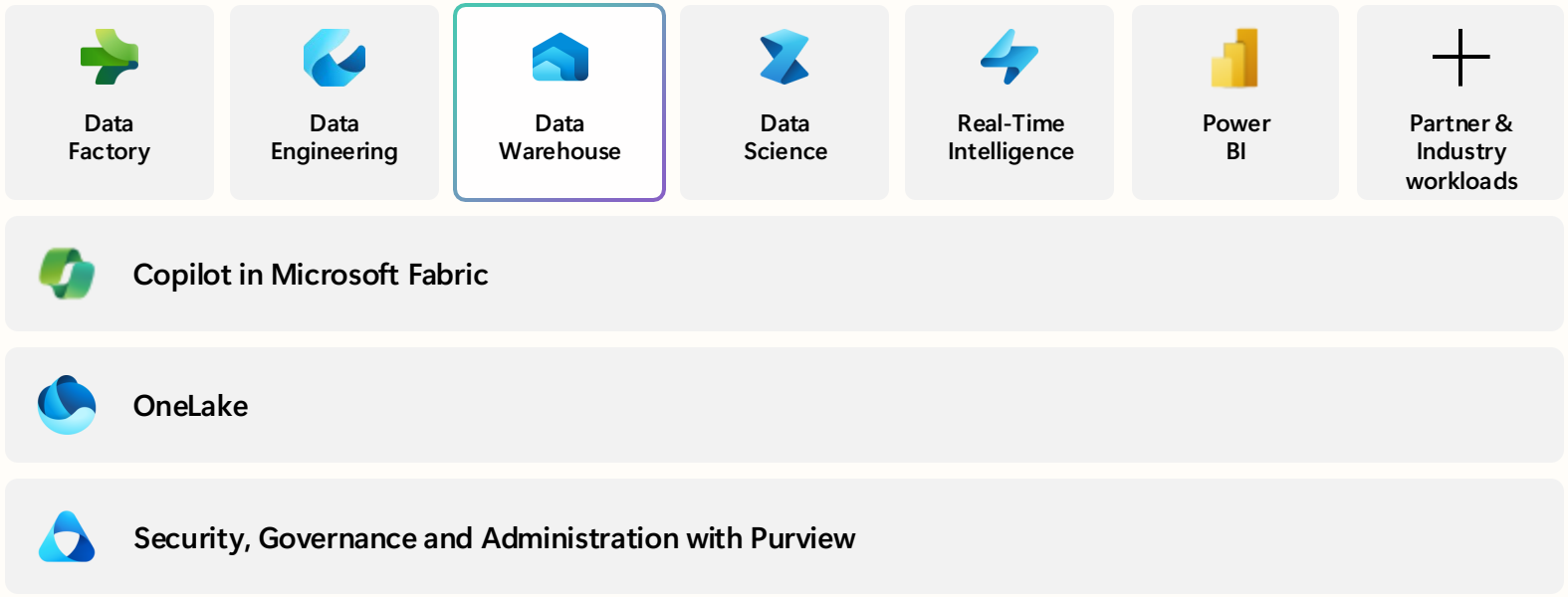
Build your data estate and empower data engineers with a world-class Spark platform, fully integrated with Data Factory, to transform and maintain infrastructures at scale



- 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

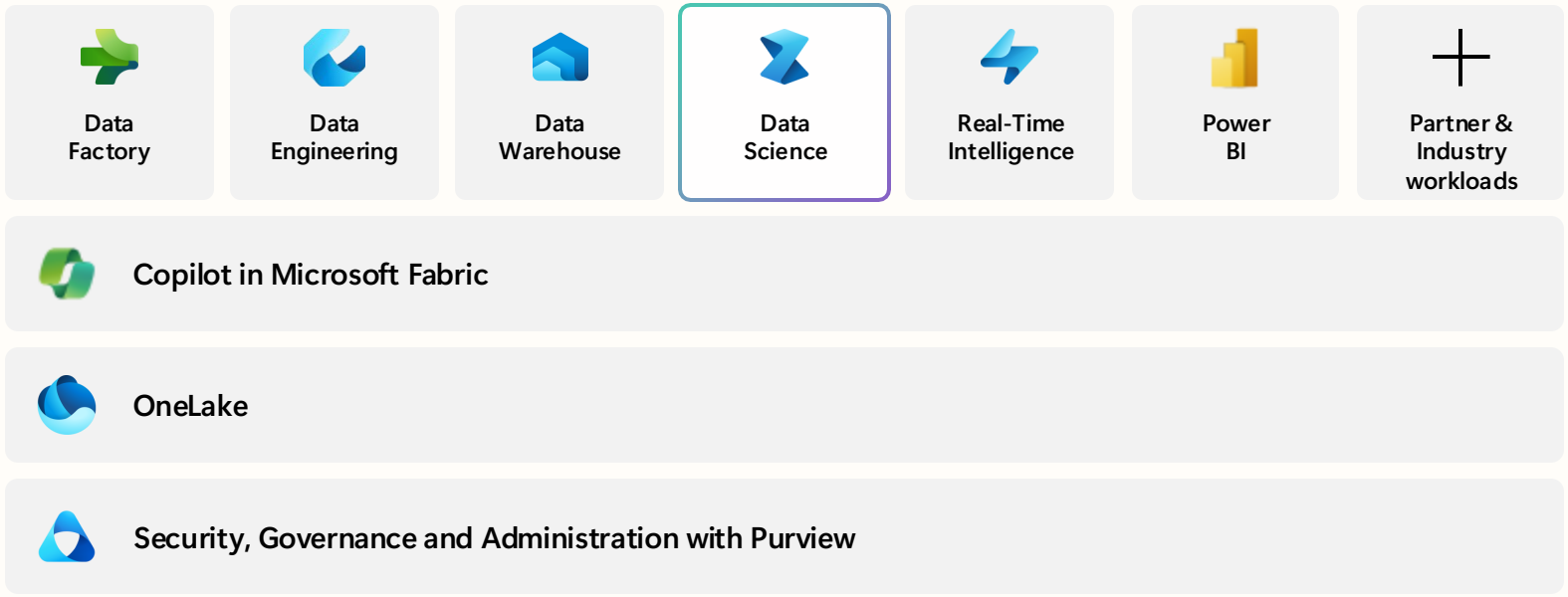
Achieve data platform goals with ease and cost efficiency, while empowering your developers and engineers of any skill level with accelerated reporting and insights



- 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

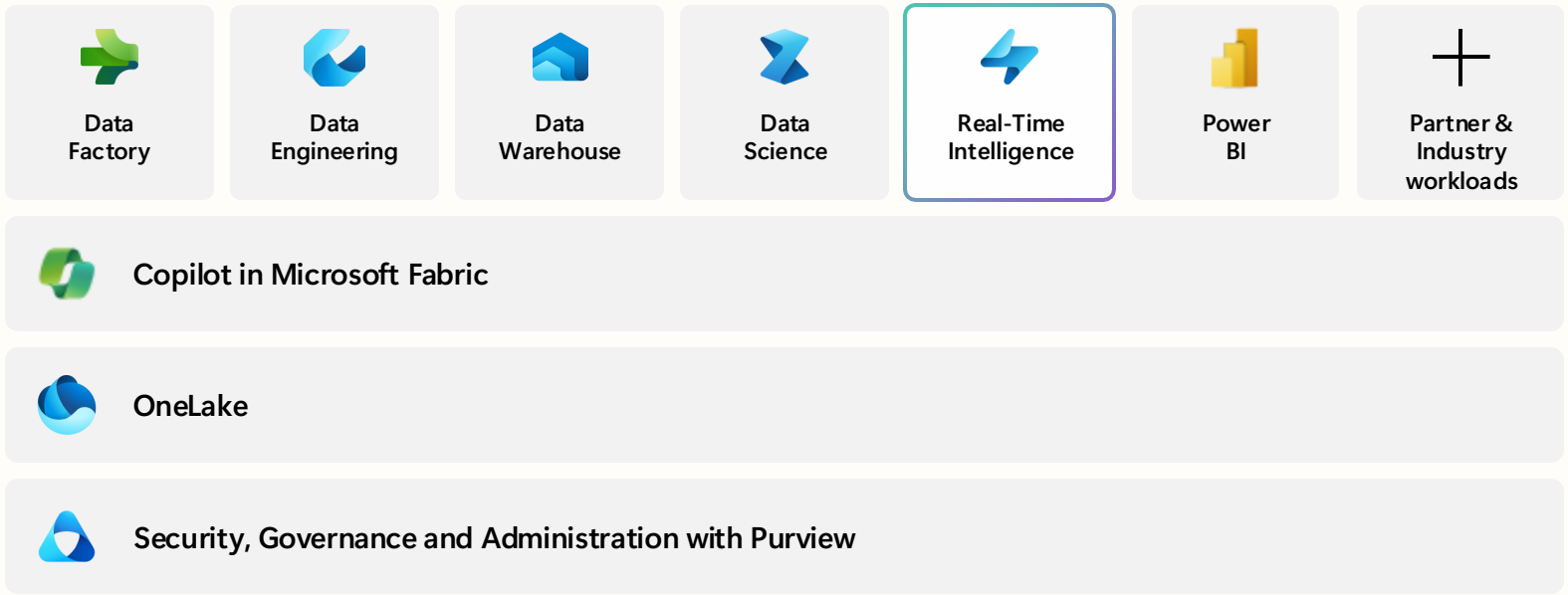
Build, deploy, and operationalize sophisticated AI and ML models with speed and at scale from your Lakehouse



- 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

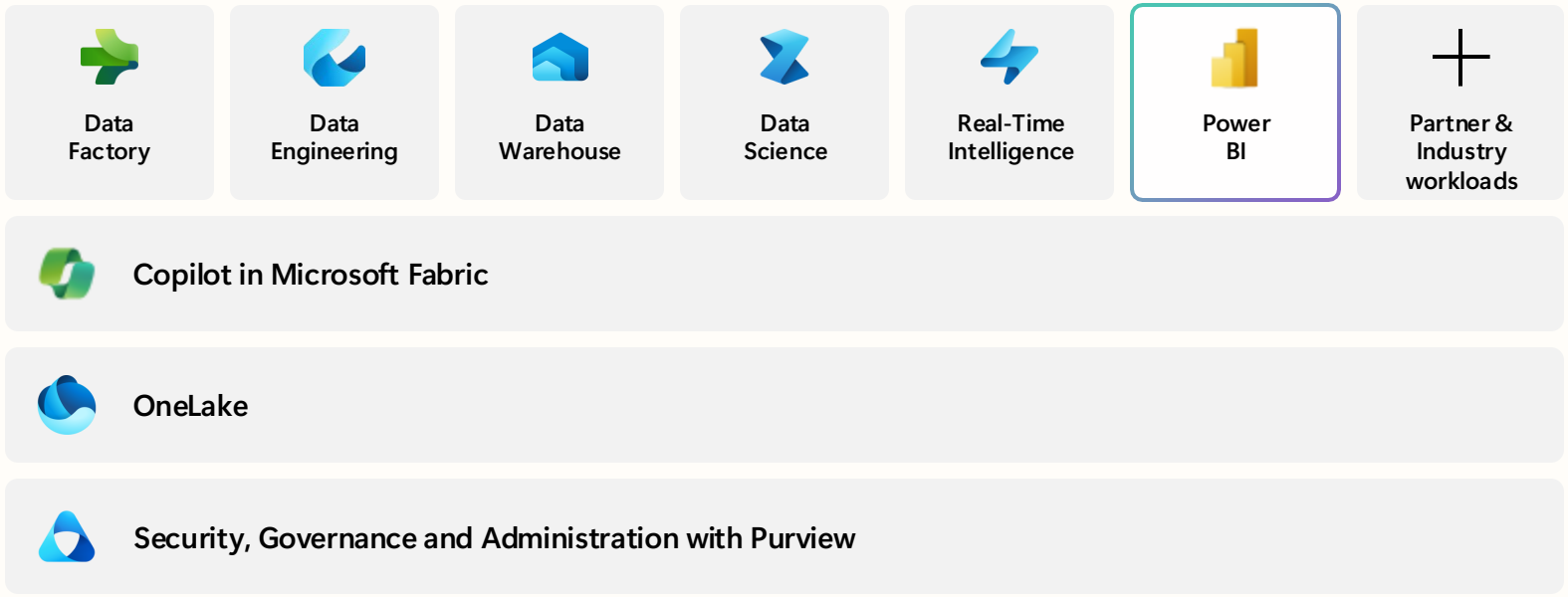
Explore data and turn insights into actions by performing real-time analysis across telemetry data to better predict, optimize, and improve data applications



- 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

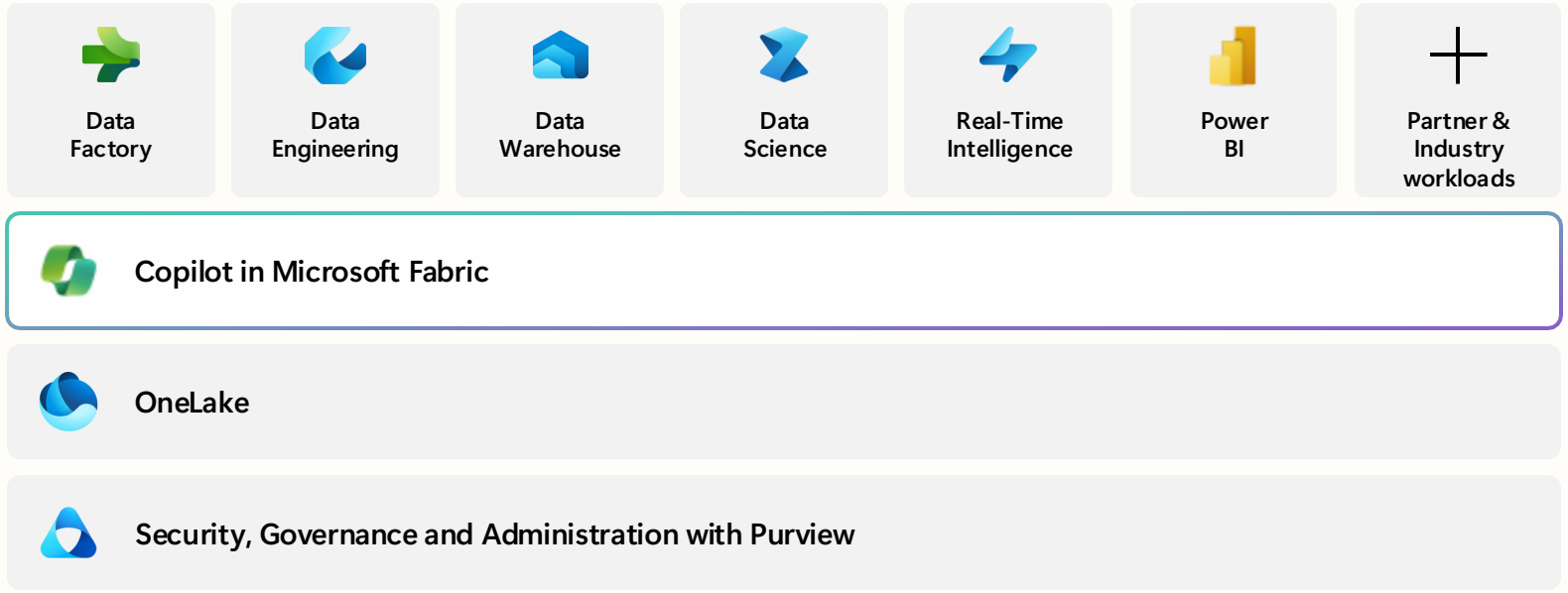
Uncover powerful insights with intelligence visuals, leverage data quickly and intuitively, and help achieve faster and better, data-based decisions with the industry-leading Power BI platform



- 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

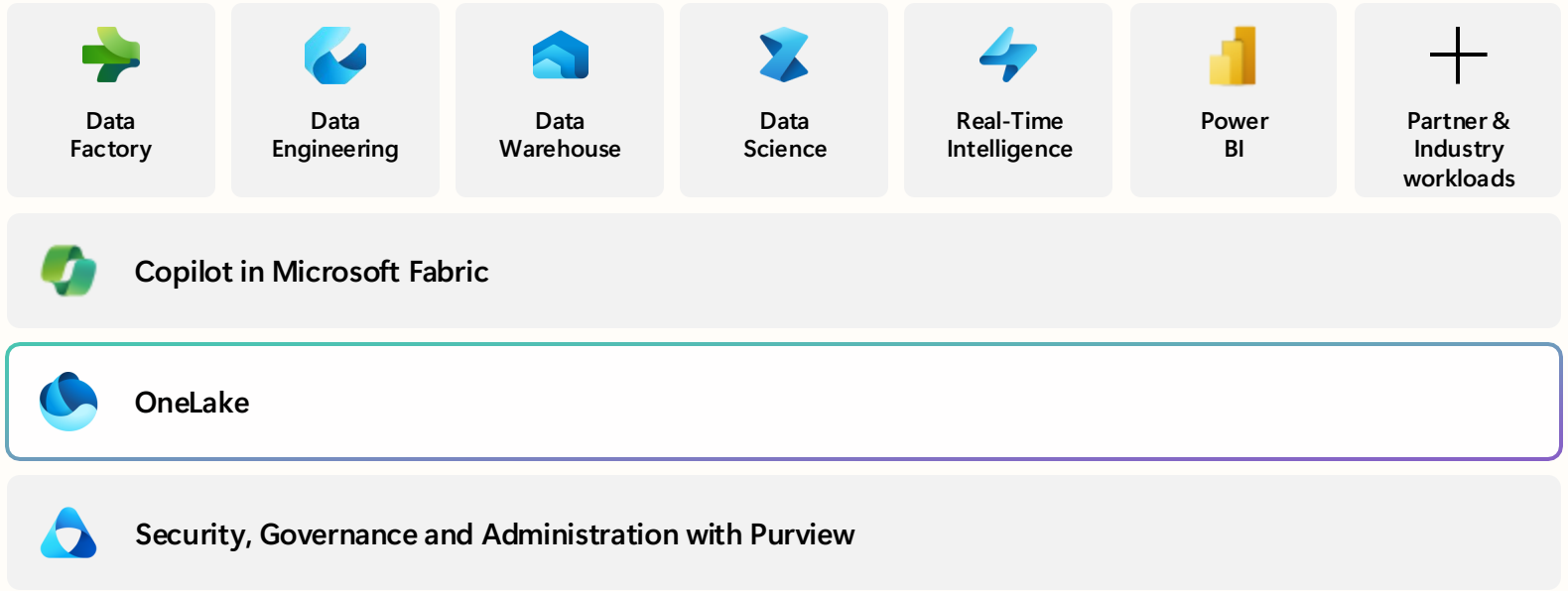
Use conversational language with Copilot in Fabric to create dataflows and pipelines, write SQL statements, build reports, and even build machine learning models



- 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

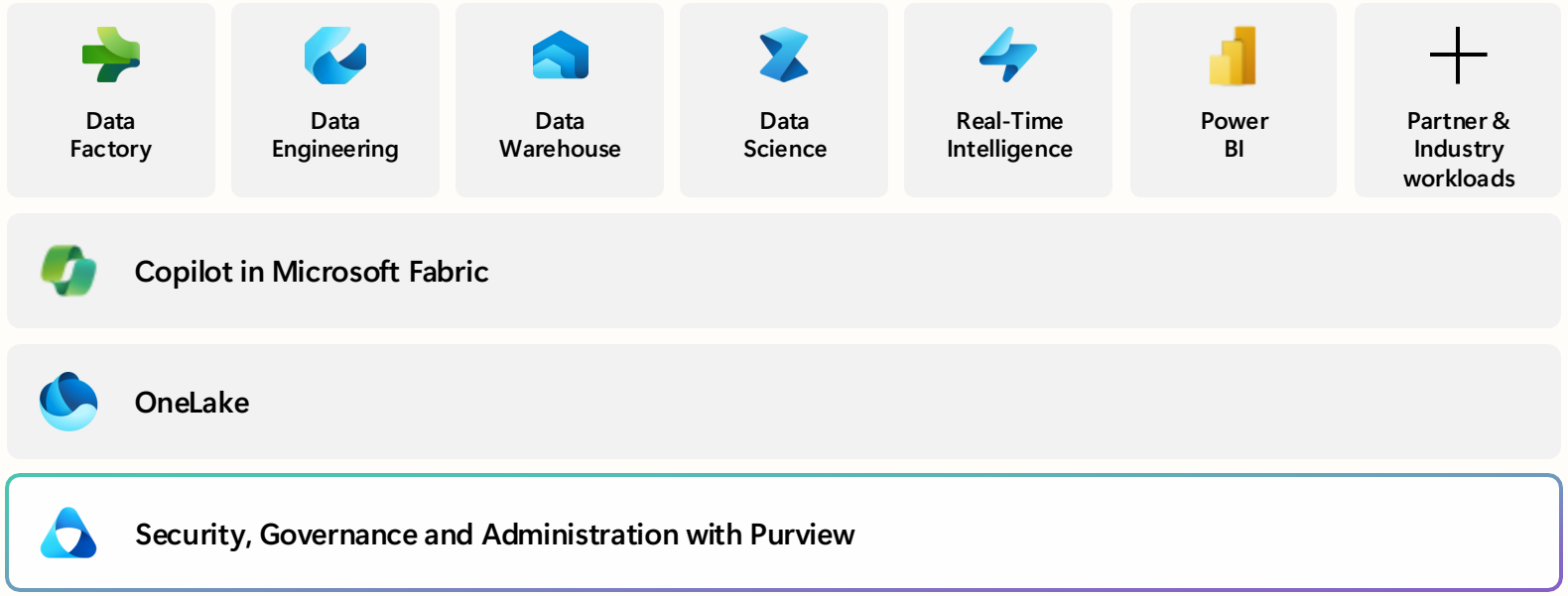
Manage and analyze all your data across your organization in a unified, secure, and centralized SaaS data lake for everyone with OneLake—the "OneDrive" for data



- 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