Microsoft Fabric

Data analytics for the era of AI

Andrea Benedetti

Sr Cloud Architect, Microsoft

in /in/abenedetti

@anBenedetti

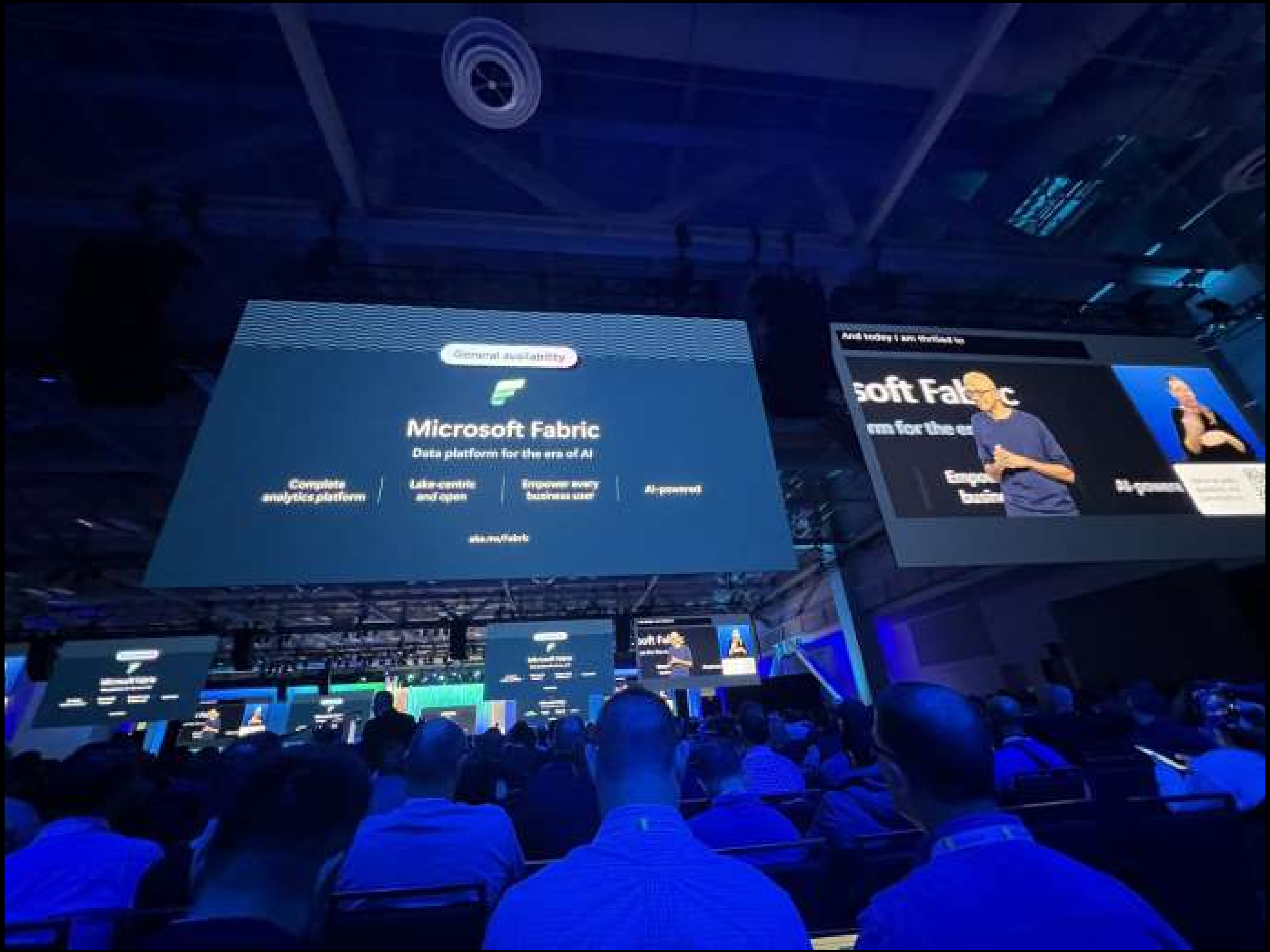
https://github.com/anbened

General Availability of Microsoft Fabric

Copilot in Microsoft Fabric is rolling out in public preview

Public Preview → over 25,000 orgs

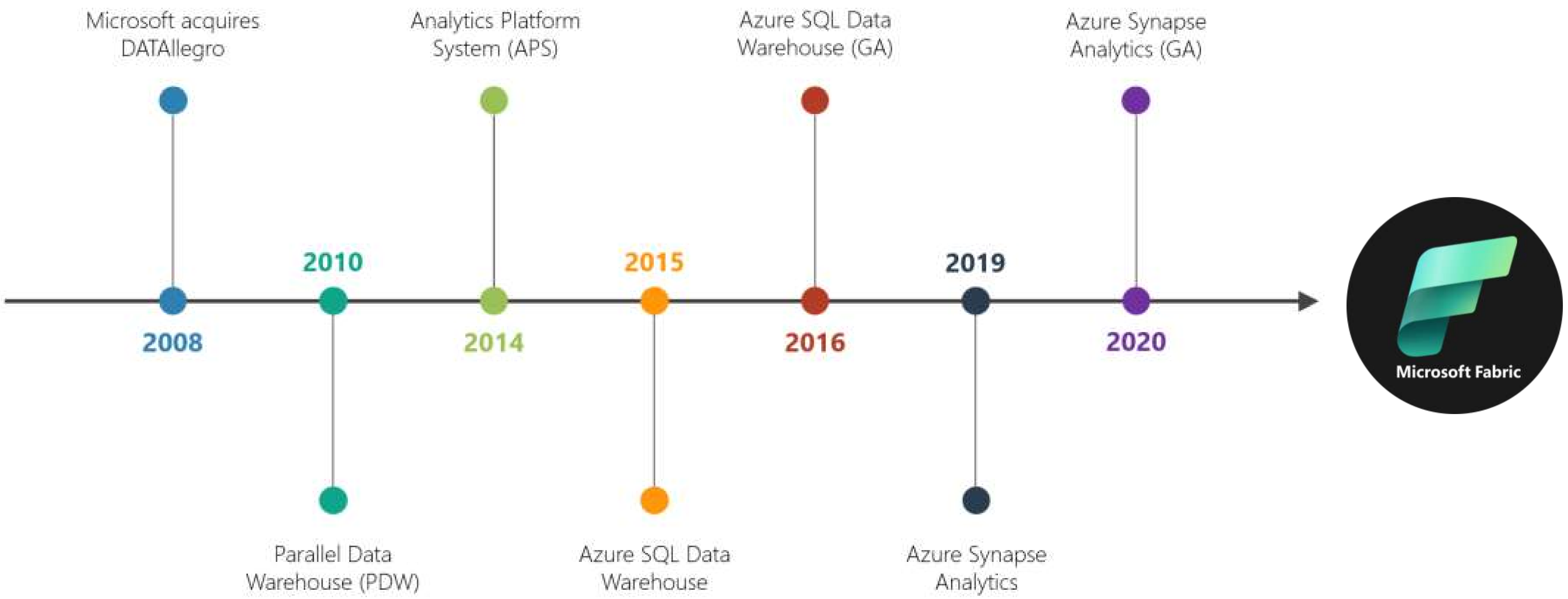
"Analytics for the era of AI" becoming GA


General availability
Microsoft Fabric
Data platform for the era of AI
Complete analytics platform
Lake-centric and open
Empower every business user
AI-powered
aka.ms/fabric

“Fabric has been our biggest data launch perhaps since SQL server”

Satya Nadella

Timeline


Microsoft acquires DATAllegro
2008
2010
Parallel Data Warehouse (PDW)
Analytics Platform System (APS)
2014
2015
Azure SQL Data Warehouse
Azure SQL Data Warehouse (GA)
2016
2019
Azure Synapse Analytics
Azure Synapse Analytics (GA)
2020
Microsoft Fabric

Ref: https://www.taygan.co/blog/2022/01/04/azure-synapse-analytics

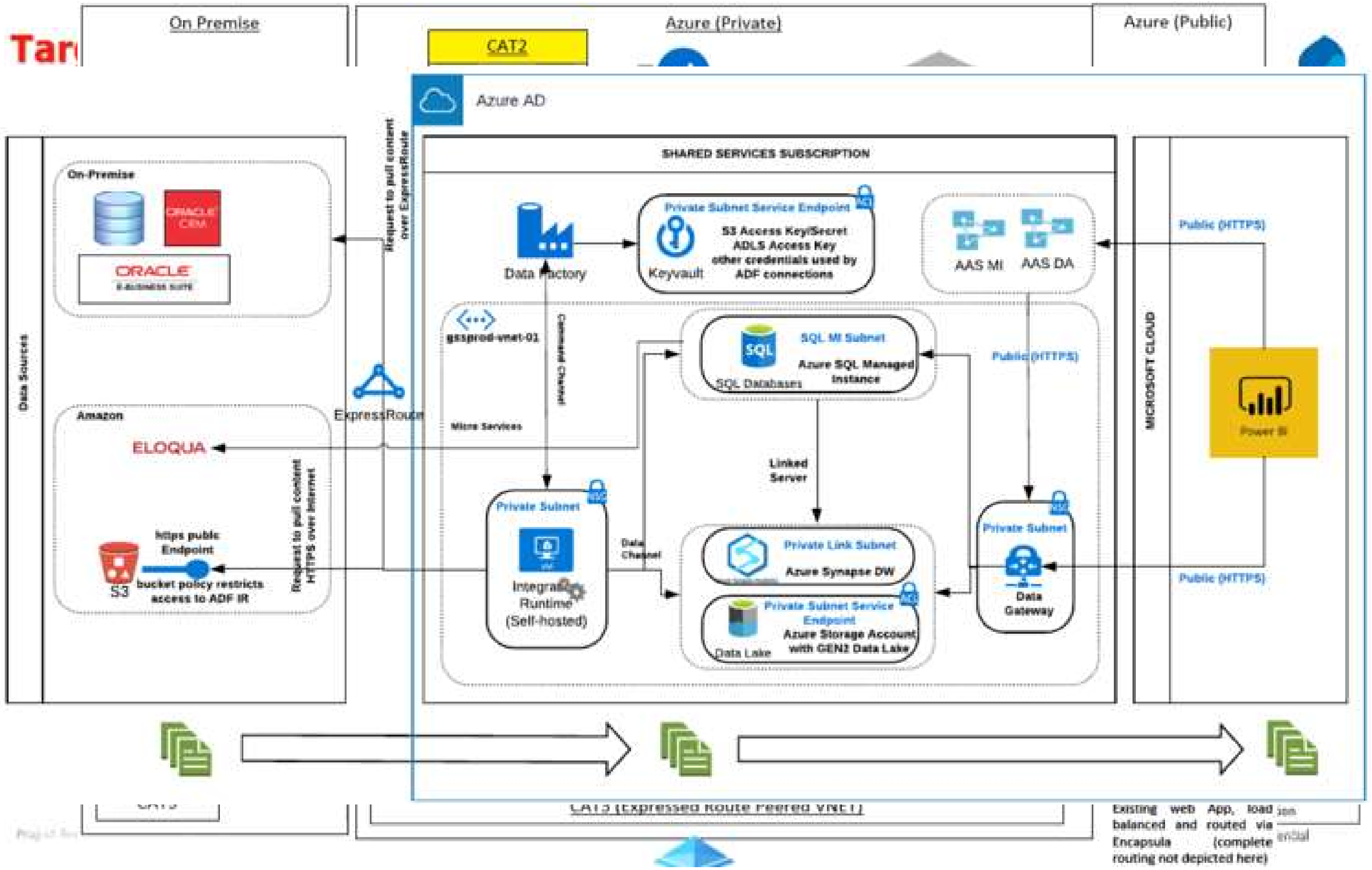
Scalable analytics are complex and fragmented

Every analytics project has many subsystems

Every subsystem need a different class of product

Products often come from multiple vendors

Integration at scale across products is complex, fragile, and expensive



Tar
On Premise
Azure (Private)
CAT2
Azure (Public)
Azure AD
SHARED SERVICES SUBSCRIPTION
Data Sources
On-Premise
ORACLE CRM
ORACLE E-BUSINESS SUITE
Amazon
ELOQUA
https public Endpoint
S3
bucket policy restricts access to ADF IR
Request to pull content HTTPS over Internet
Request to pull content over ExpressRoute
ExpressRoute
Data Factory
Private Subnet Service Endpoint
Keyvault
S3 Access Key/Secret
ADLS Access Key
other credentials used by ADF connections
AAS MI
AAS DA
gssprod-vnet-01
Command Channel
Micro Services
SQL MI Subnet
SQL Databases
Azure SQL Managed Instance
Linked Server
Private Subnet
Integration Runtime (Self-hosted)
Data Channel
Private Link Subnet
Azure Synapse DW
Private Subnet Service Endpoint
Data Lake
Azure Storage Account with GEN2 Data Lake
Private Subnet
Data Gateway
Public (HTTPS)
Public (HTTPS)
Public (HTTPS)
MICROSOFT CLOUD
Power BI
CAT3 (Expressed Route Peered VNET)
Existing web App, load balanced and routed via Encapsula (complete routing not depicted here)

Scalable analytics are complex and fragmented

Every analytics project has many subsystems

Every subsystem need a different class of product

Products often come from multiple vendors

Integration at scale across products is complex, fragile and expensive

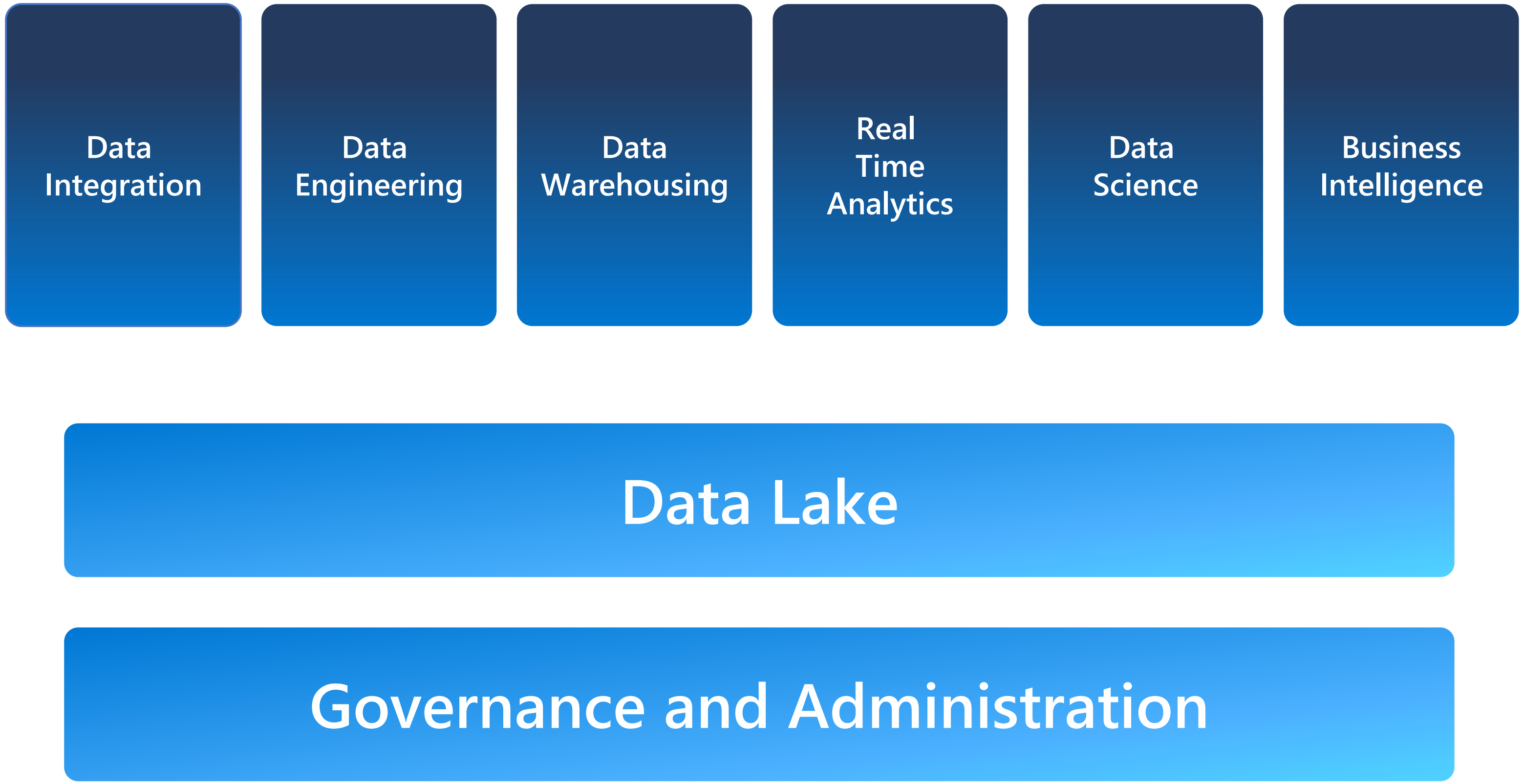
> "
>
> Simplify,
>
> I am the Chief Data Officer and don’t want to be the Chief Integration Officer.”

Every CDO, Every Enterprise

A silver lining?

Analytics systems have very predictable patterns

Microsoft has all the products with the right scale needed to build a complete analytics system



Data Integration

Data Engineering

Data Warehousing

Real Time Analytics

Data Science

Business Intelligence

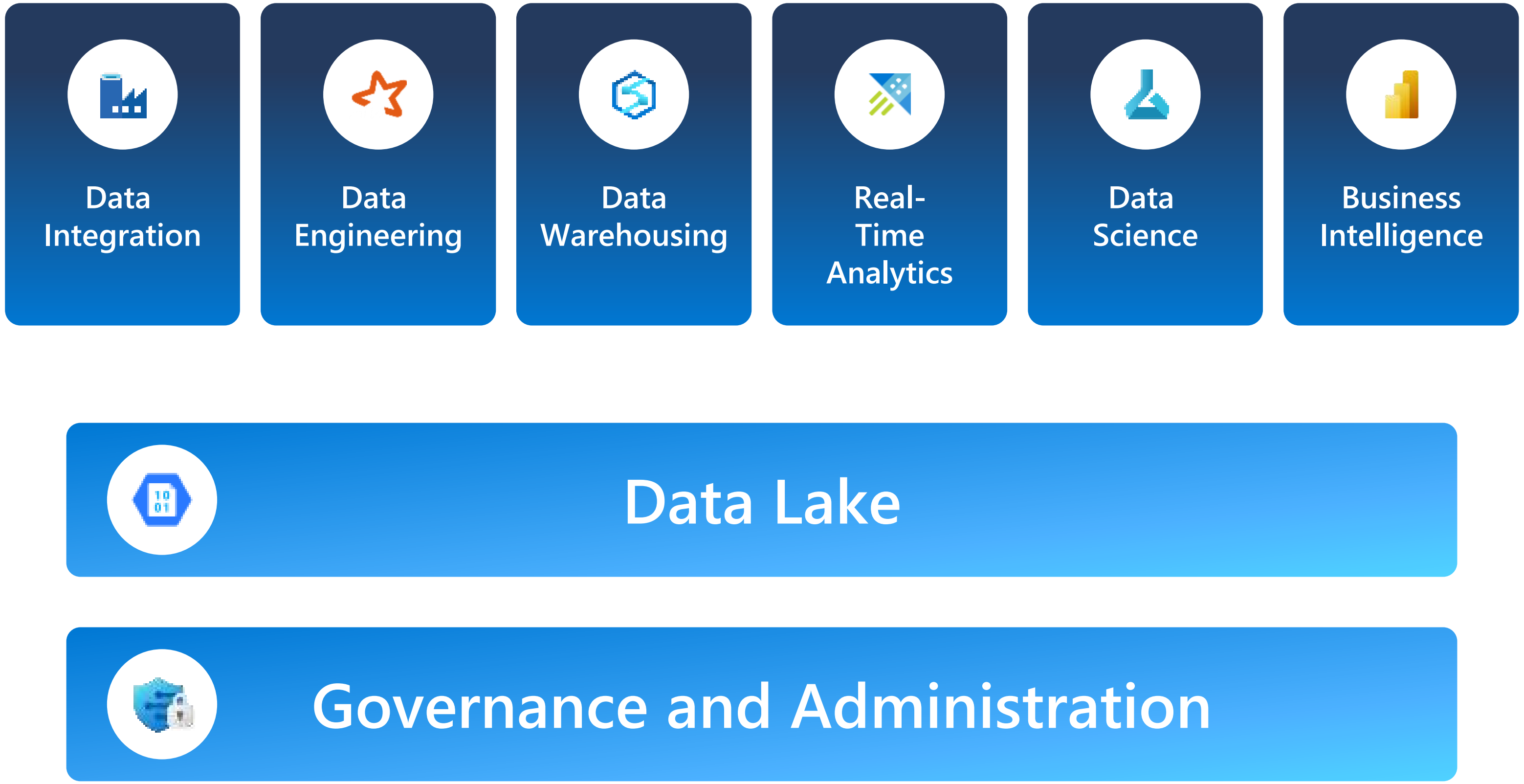
Data Lake

Governance and Administration

A silver lining?

Analytics systems have very predictable patterns

Microsoft has all the products with the right scale needed to build a complete analytics system



Data Integration

Data Engineering

Data Warehousing

Real-Time Analytics

Data Science

Business Intelligence

Data Lake

Governance and Administration

Still far too complex

Many Products

Different Experiences

Proprietary and Open

Dedicated and Serverless

PaaS and SaaS

Different Business Models

Steep Learning Curves

Deep Expertise Needed

High Integration Effort

Power BI

Synapse

Kusto

Azure AI

Data Factory

Spark

What we're hearing from our customers

How do I unify all our disparate sources of data into one single source of truth?

How do I improve democratization of data and insights with easy-to-use analytics tools?

How do I enable faster decision-making with scale, while optimizing governance and cost?

What is Fabric?

End-to-end analytics data fabric
From the data lake to the business user

Complete Analytics Platform

- Best of Breed
- Unified SaaS Solution
- Low Code Plus Pro Dev

Lake-centric and Open

- OneLake
- One Copy
- Always Synced

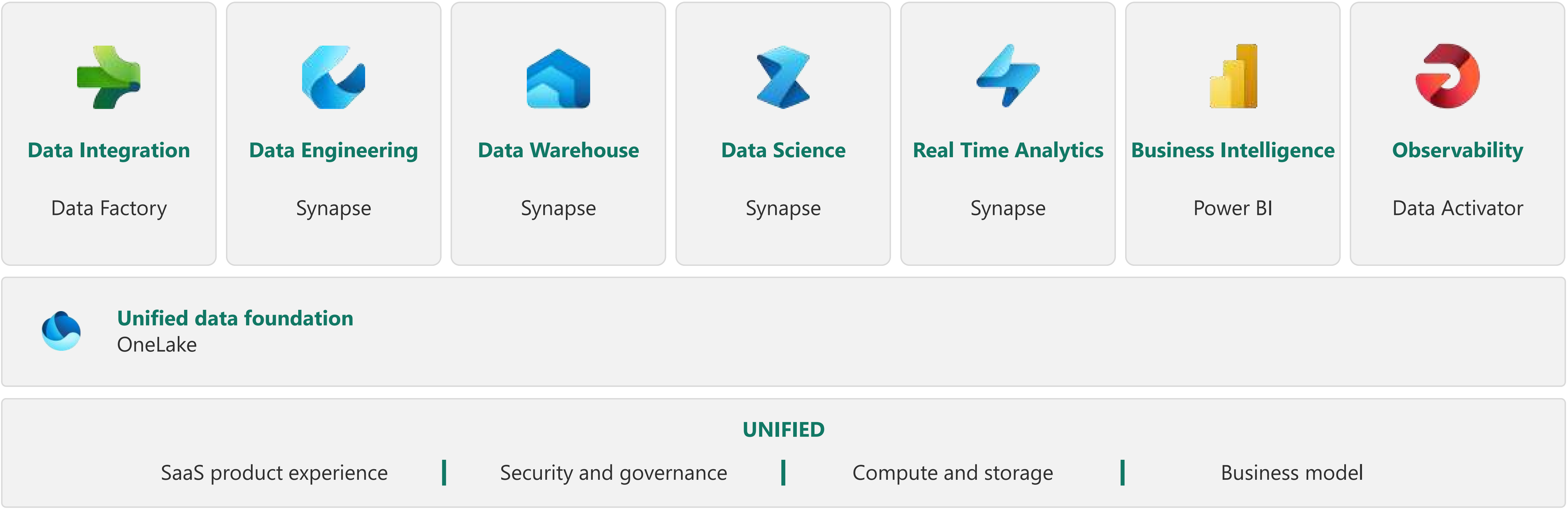
Empower Every Office User

- Familiar and Intuitive
- Built Into Office
- Insight to Action

Persistent Security and Governance

- End-to-End Visibility
- Always Governed
- Secure by Default

Microsoft Fabric



Data Integration
Data Factory

Data Engineering
Synapse

Data Warehouse
Synapse

Data Science
Synapse

Real Time Analytics
Synapse

Business Intelligence
Power BI

Observability
Data Activator

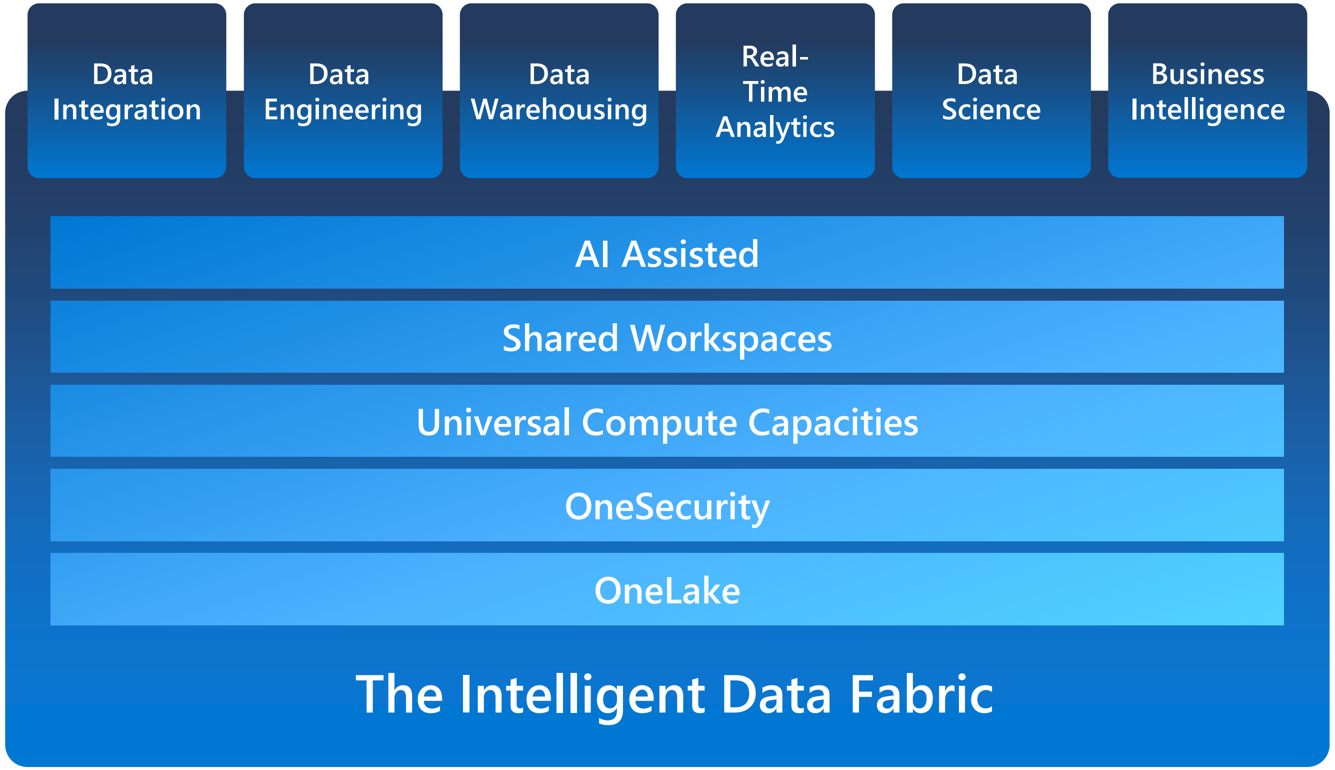
Unified data foundation
OneLake

UNIFIED

SaaS product experience | Security and governance | Compute and storage | Business model

End-to-end analytics data fabric
From the data lake to the business user

Microsoft Fabric



Data Integration

Data Engineering

Data Warehousing

Real-Time Analytics

Data Science

Business Intelligence

AI Assisted

Shared Workspaces

Universal Compute Capacities

OneSecurity

OneLake

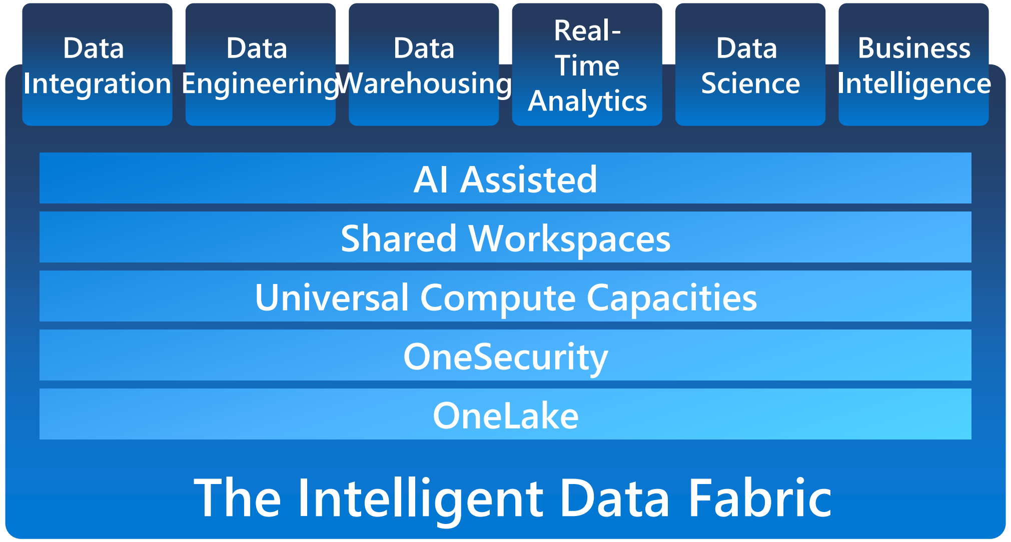
The Intelligent Data Fabric

Single...

Onboarding and trials
Sign-on
Navigation model
UX model
Workspace organization
Collaboration experience
Data Lake
Storage format
Data copy for all engines
Security model
CI/CD
Monitoring hub
Data Hub
Governance & compliance

What is Fabric? Recap

- Adds significant functionality to a Power BI tenant
- Compute in the form of Capacity Units (CU) → not a «server»
  - Power BI Premium P1 = F64
- Compute control abstracted away
- Storage = open-souce Delta Lake format



Data Integration

Data Engineering

Data Warehousing

Real-Time Analytics

Data Science

Business Intelligence

AI Assisted

Shared Workspaces

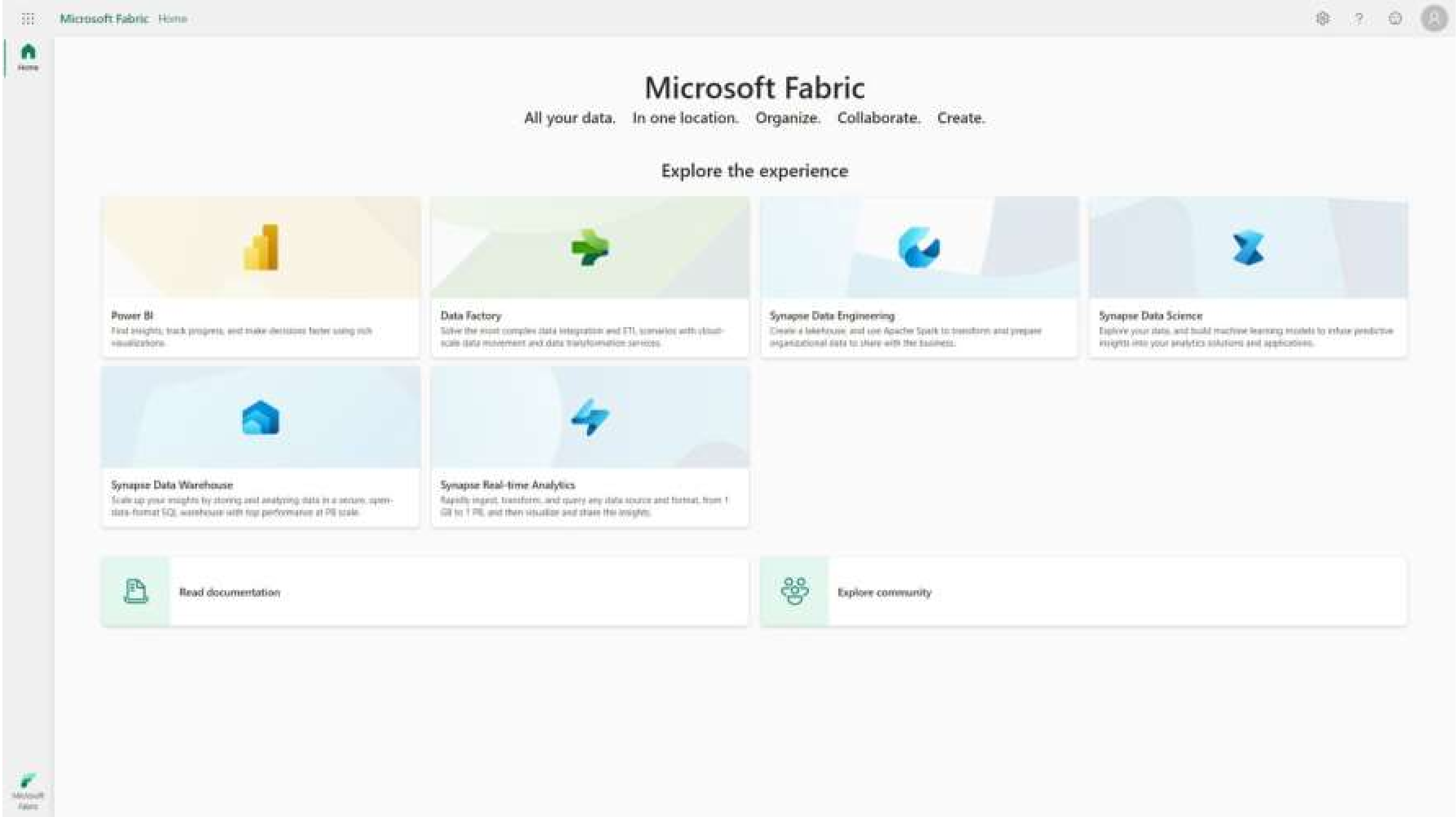
Universal Compute Capacities

OneSecurity

OneLake

The Intelligent Data Fabric

Persona Centric Experience



Microsoft Fabric

All your data. In one location. Organize. Collaborate. Create.

Explore the experience

Power BI
Find insights, track progress, and make decisions faster using rich visualizations.

Data Factory
Solve the most complex data integration and ETL scenarios with cloud-scale data movement and data transformation services.

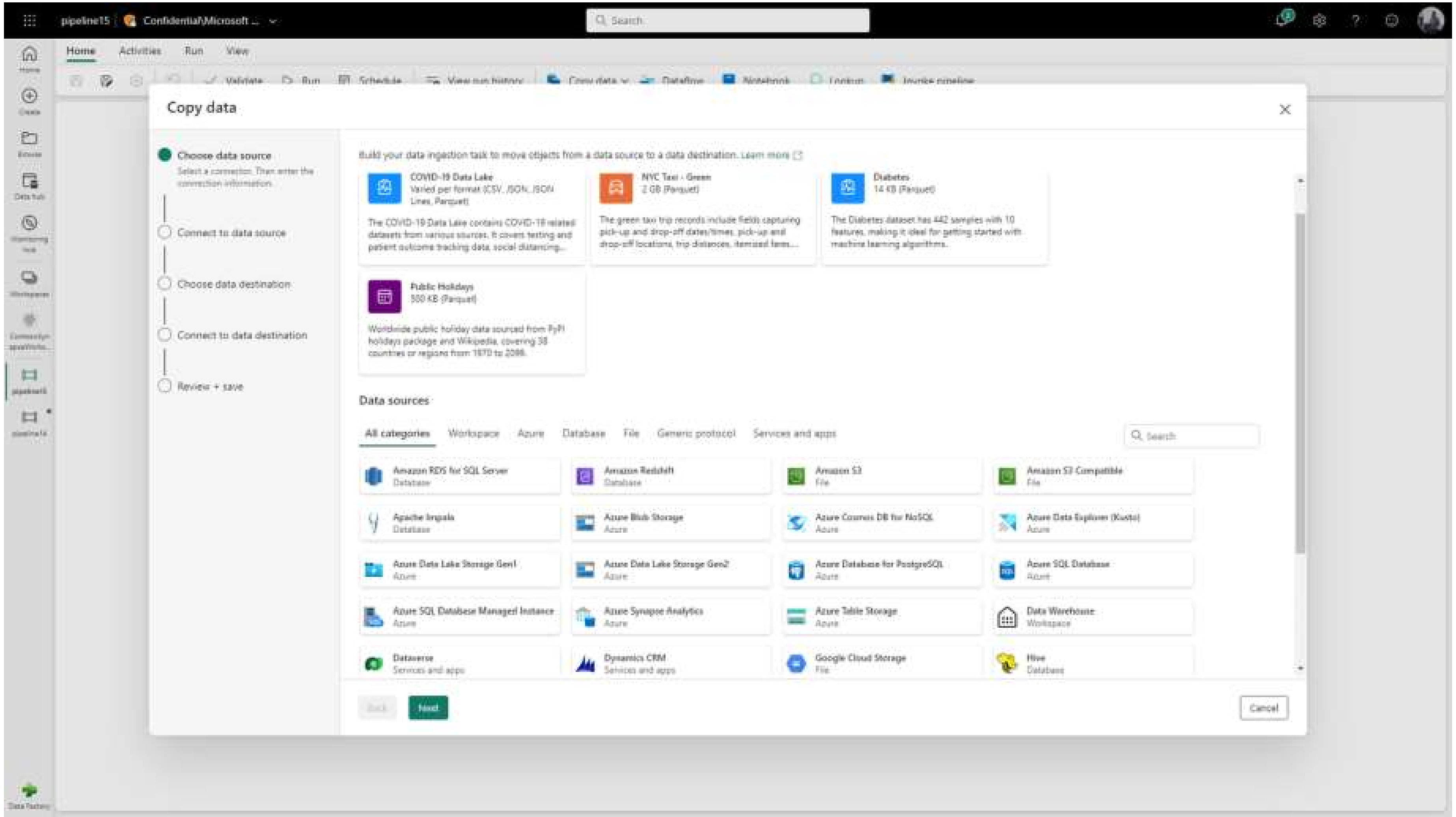
Synapse Data Engineering
Create a lakehouse and use Apache Spark to transform and prepare organizational data to share with the business.

Synapse Data Science
Explore your data, and build machine learning models to infuse predictive insights into your analytics solutions and applications.

Synapse Data Warehouse
Scale up your insights by storing and analyzing data in a secure, open-data-format SQL warehouse with top performance at PB scale.

Synapse Real-time Analytics
Rapidly ingest, transform, and query any data source and format, from 1 GB to 1 PB, and then visualize and share the insights.

Data Integration



Copy data

- Choose data source
  Select a connection. Then enter the connection information.
- Connect to data source
- Choose data destination
- Connect to data destination
- Review + save

Build your data ingestion task to move objects from a data source to a data destination. Learn more

COVID-19 Data Lake
Varied per format (CSV, JSON, JSON Lines, Parquet)
The COVID-19 Data Lake contains COVID-19 related datasets from various sources. It covers testing and patient outcome tracking data, social distancing...

NYC Taxi - Green
2 GB (Parquet)
The green taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares...

Diabetes
14 KB (Parquet)
The Diabetes dataset has 442 samples with 10 features, making it ideal for getting started with machine learning algorithms.

Public Holidays
500 KB (Parquet)
Worldwide public holiday data sourced from PyPI holidays package and Wikipedia, covering 38 countries or regions from 1970 to 2099.

Data sources

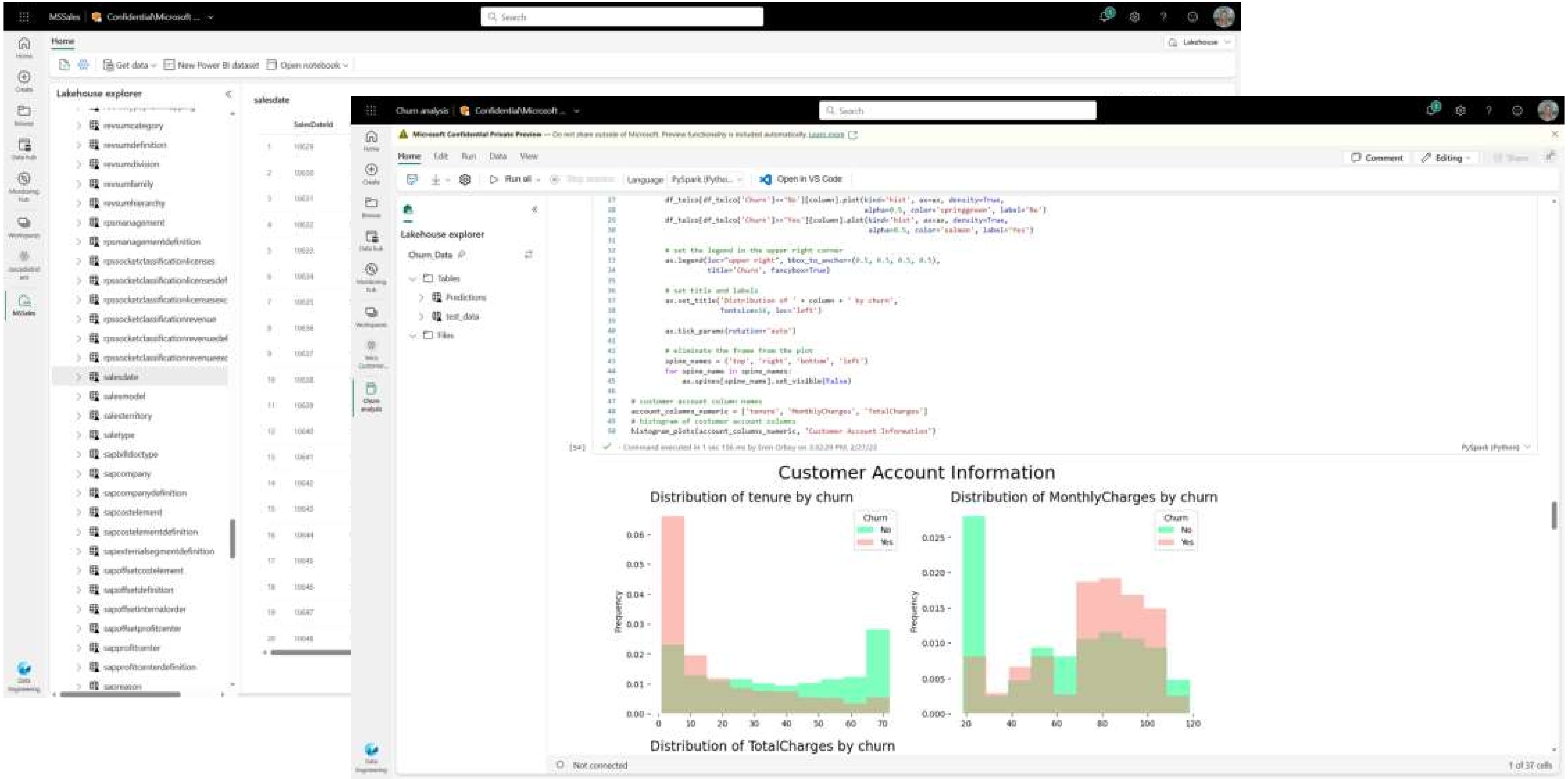
All categories | Workspace | Azure | Database | File | Generic protocol | Services and apps

Search

Amazon RDS for SQL Server Database	Amazon Redshift Database	Amazon S3 File	Amazon S3 Compatible File
Apache Impala Database	Azure Blob Storage Azure	Azure Cosmos DB for NoSQL Azure	Azure Data Explorer (Kusto) Azure
Azure Data Lake Storage Gen1 Azure	Azure Data Lake Storage Gen2 Azure	Azure Database for PostgreSQL Azure	Azure SQL Database Azure
Azure SQL Database Managed Instance Azure	Azure Synapse Analytics Azure	Azure Table Storage Azure	Data Warehouse Workspace
Dataverse Services and apps	Dynamics CRM Services and apps	Google Cloud Storage File	Hive Database

Back | Next | Cancel

Data Engineering



```
            df_telco[df_telco['Churn']=='No'][column].plot(kind='hist', ax=ax, density=True,
                                                            alpha=0.5, color='springgreen', label='No')
            df_telco[df_telco['Churn']=='Yes'][column].plot(kind='hist', ax=ax, density=True,
                                                             alpha=0.5, color='salmon', label='Yes')

            # set the legend in the upper right corner
            ax.legend(loc="upper right", bbox_to_anchor=(0.5, 0.5, 0.5, 0.5),
                      title='Churn', fancybox=True)

            # set title and labels
            ax.set_title('Distribution of ' + column + ' by churn',
                         fontsize=16, loc='left')

            ax.tick_params(rotation='auto')

            # eliminate the frame from the plot
            spine_names = ('top', 'right', 'bottom', 'left')
            for spine_name in spine_names:
                ax.spines[spine_name].set_visible(False)

    # customer account column names
    account_columns_numeric = ['tenure', 'MonthlyCharges', 'TotalCharges']
    # histogram of customer account columns
    histogram_plots(account_columns_numeric, 'Customer Account Information')
```

[54] ✓ - Command executed in 1 sec 156 ms by Eren Orbey on 3:32:29 PM, 2/27/23

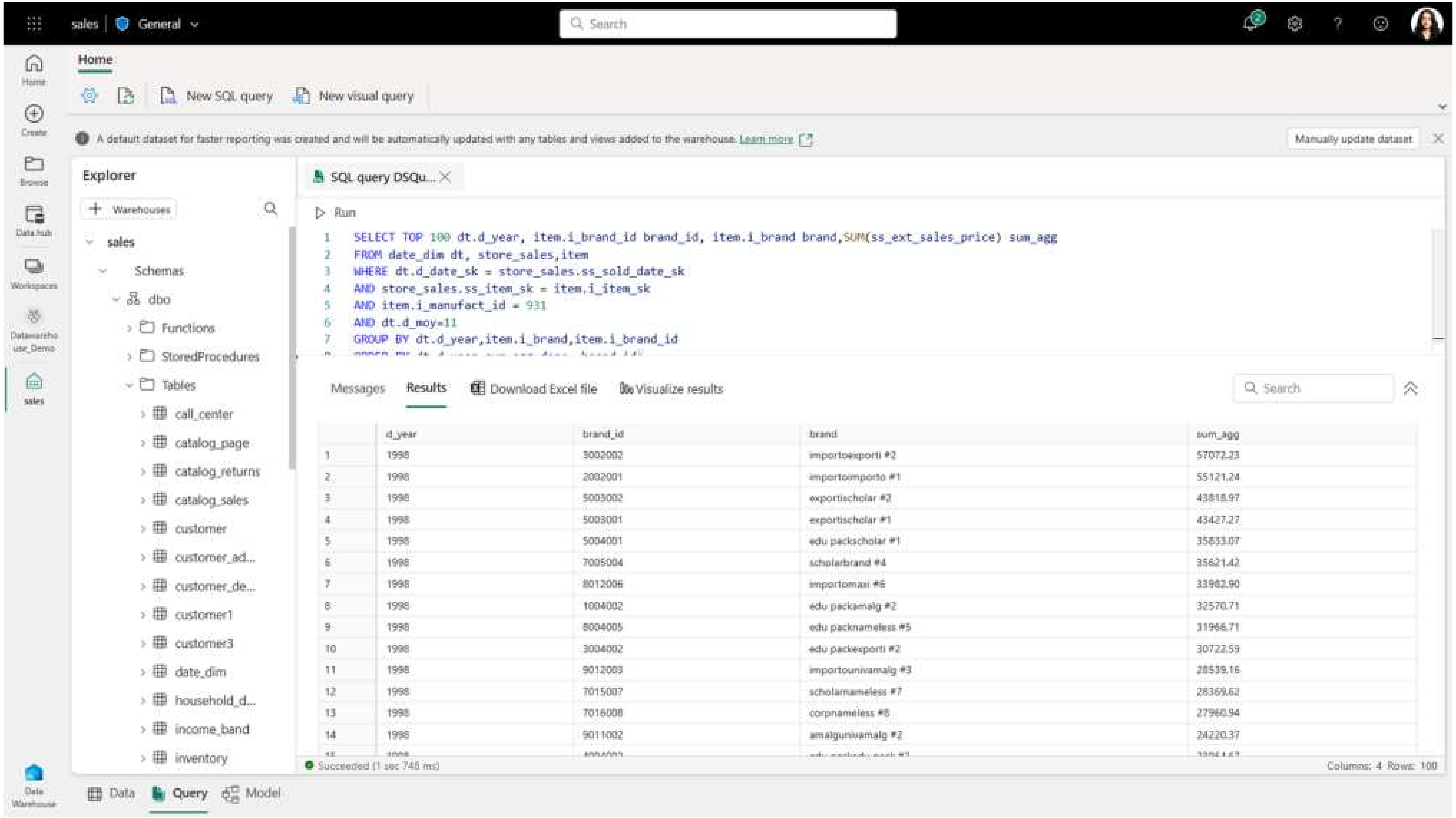
Customer Account Information

Distribution of tenure by churn

Distribution of MonthlyCharges by churn

Distribution of TotalCharges by churn

Data Warehouse



A default dataset for faster reporting was created and will be automatically updated with any tables and views added to the warehouse. Learn more

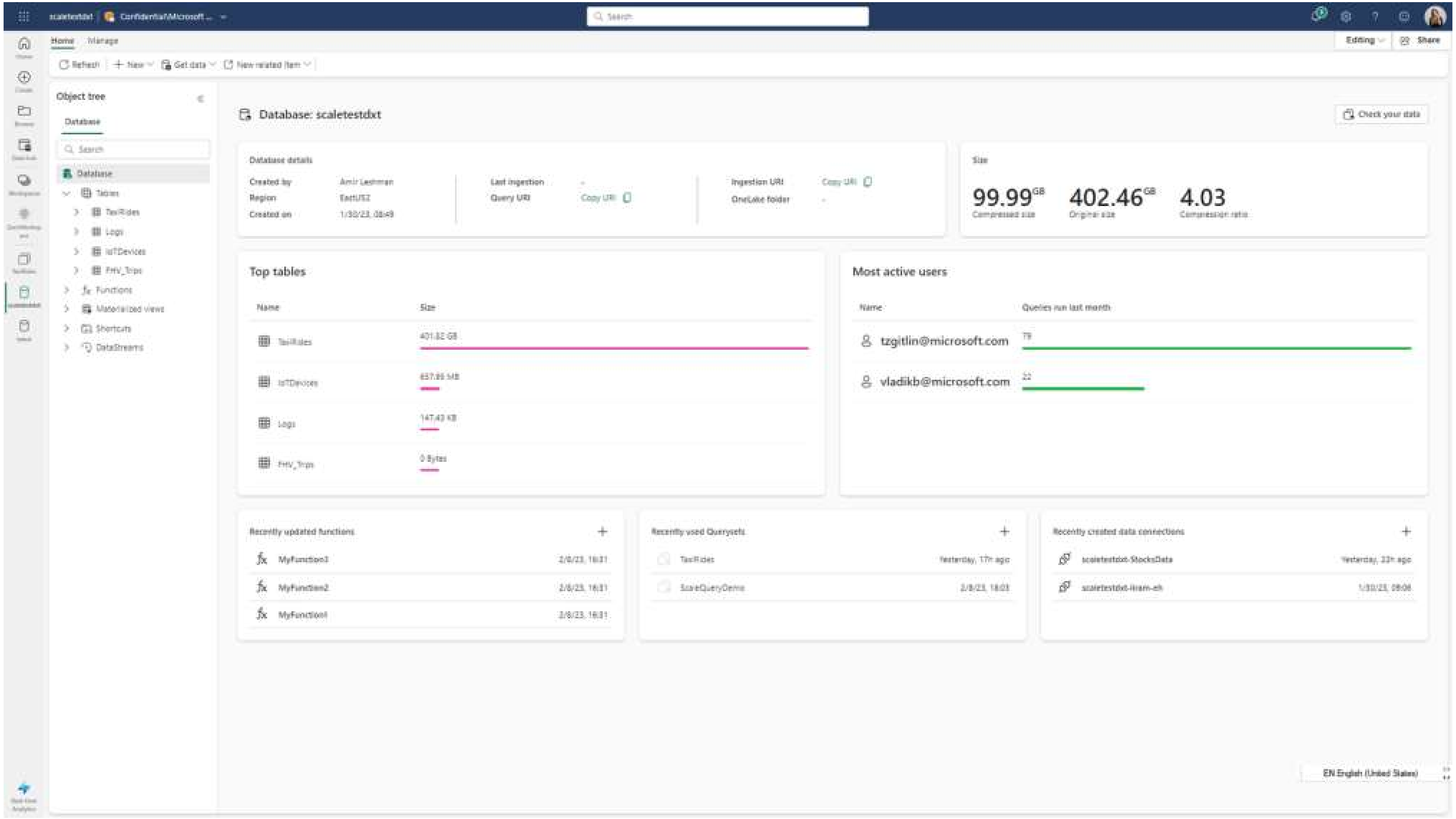
```sql
SELECT TOP 100 dt.d_year, item.i_brand_id brand_id, item.i_brand brand,SUM(ss_ext_sales_price) sum_agg
FROM date_dim dt, store_sales,item
WHERE dt.d_date_sk = store_sales.ss_sold_date_sk
AND store_sales.ss_item_sk = item.i_item_sk
AND item.i_manufact_id = 931
AND dt.d_moy=11
GROUP BY dt.d_year,item.i_brand,item.i_brand_id
```

	d_year	brand_id	brand	sum_agg
1	1998	3002002	importoexporti #2	57072.23
2	1998	2002001	importoimporto #1	55121.24
3	1998	5003002	exportischolar #2	43818.97
4	1998	5003001	exportischolar #1	43427.27
5	1998	5004001	edu packscholar #1	35833.07
6	1998	7005004	scholarbrand #4	35621.42
7	1998	8012006	importomaxi #6	33982.90
8	1998	1004002	edu packamalg #2	32570.71
9	1998	8004005	edu packnameless #5	31966.71
10	1998	3004002	edu packexporti #2	30722.59
11	1998	9012003	importounivamalg #3	28539.16
12	1998	7015007	scholarnameless #7	28369.62
13	1998	7016008	corpnameless #8	27960.94
14	1998	9011002	amalgunivamalg #2	24220.37

Succeeded (1 sec 748 ms)

Columns: 4 Rows: 100

Real-time Analytics



Database: scaletestdxt

Database details

Created by	Amir Leshman	Last ingestion	-	Ingestion URI	Copy URI
Region	EastUS2	Query URI	Copy URI	OneLake folder	-
Created on	1/30/23, 08:49				

Size

99.99 GB	402.46 GB	4.03
Compressed size	Original size	Compression ratio

Top tables

Name	Size
TaxiRides	401.82 GB
IoTDevices	657.95 MB
Logs	147.43 KB
FHV_Trips	0 Bytes

Most active users

Name	Queries run last month
tzgitlin@microsoft.com	79
vladikb@microsoft.com	22

Recently updated functions

MyFunction3	2/6/23, 16:31
MyFunction2	2/6/23, 16:31
MyFunction1	2/6/23, 16:31

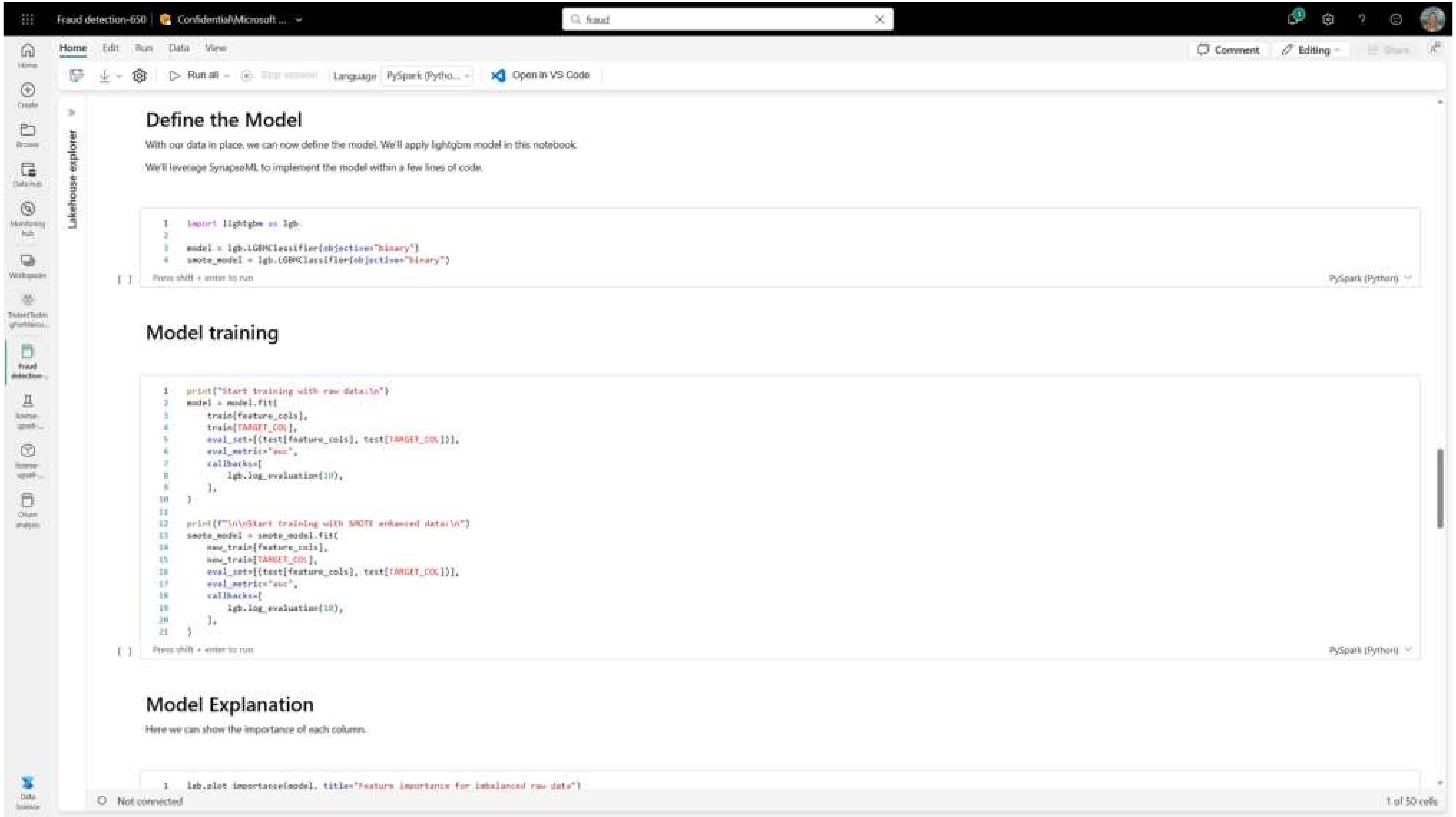
Recently used Querysets

TaxiRides	Yesterday, 17h ago
ScaleQueryDemo	2/8/23, 18:03

Recently created data connections

scaletestdxt-StocksData	Yesterday, 23h ago
scaletestdxt-itram-eh	1/30/23, 09:06

Data Science



Define the Model

With our data in place, we can now define the model. We'll apply lightgbm model in this notebook.

We'll leverage SynapseML to implement the model within a few lines of code.

```
import lightgbm as lgb

model = lgb.LGBMClassifier(objective="binary")
smote_model = lgb.LGBMClassifier(objective="binary")
```

Press shift + enter to run

Model training

```
print("Start training with raw data:\n")
model = model.fit(
    train[feature_cols],
    train[TARGET_COL],
    eval_set=[(test[feature_cols], test[TARGET_COL])],
    eval_metric="auc",
    callbacks=[
        lgb.log_evaluation(10),
    ],
)

print(f"\n\nStart training with SMOTE enhanced data:\n")
smote_model = smote_model.fit(
    new_train[feature_cols],
    new_train[TARGET_COL],
    eval_set=[(test[feature_cols], test[TARGET_COL])],
    eval_metric="auc",
    callbacks=[
        lgb.log_evaluation(10),
    ],
)
```

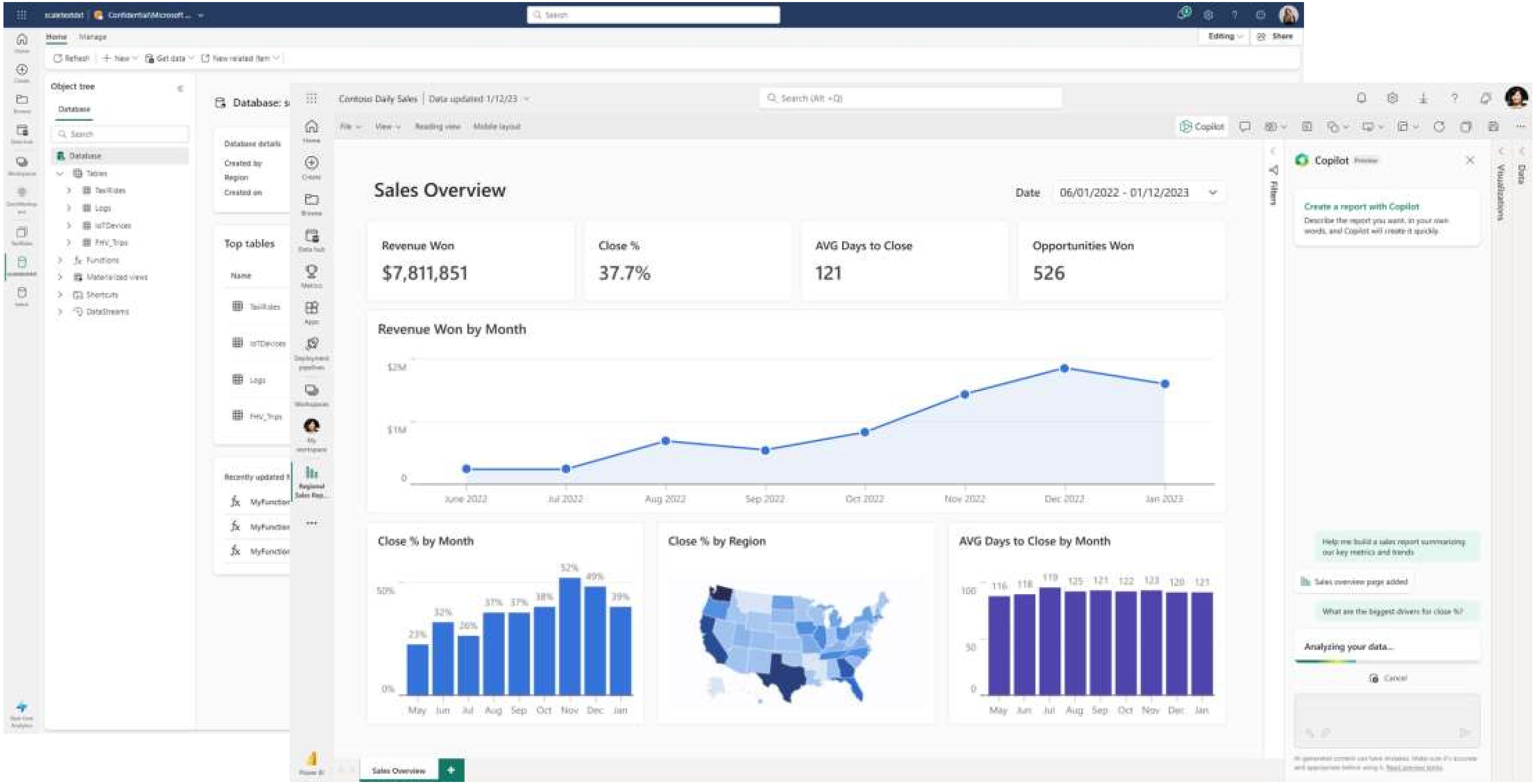
Press shift + enter to run

Model Explanation

Here we can show the importance of each column.

```
lgb.plot_importance(model, title="Feature importance for imbalanced raw data")
```

Power BI



Sales Overview

Date 06/01/2022 - 01/12/2023

Revenue Won $7,811,851

Close % 37.7%

AVG Days to Close 121

Opportunities Won 526

Copilot Preview

Create a report with Copilot

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

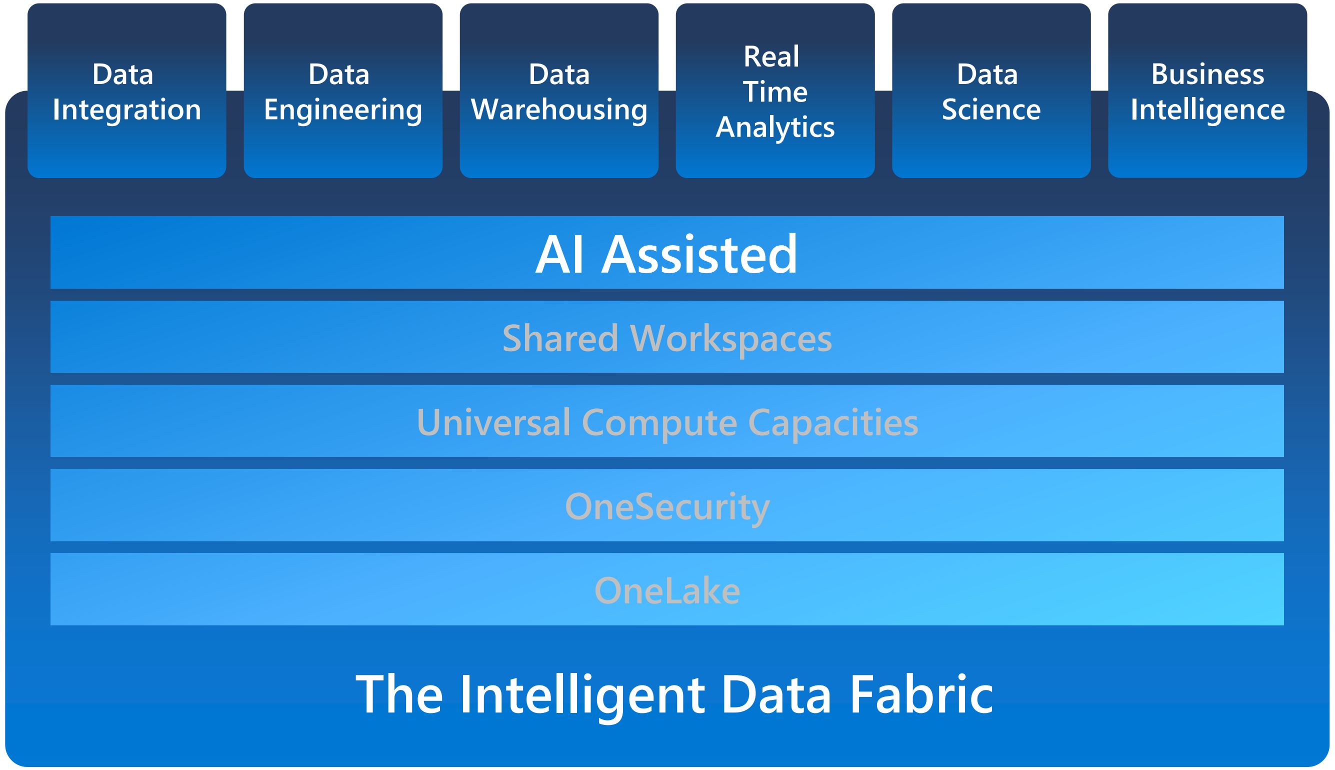
Sales overview page added

What are the biggest drivers for close %?

Analyzing your data...

Cancel

AI Assisted Creation in Microsoft Fabric



The Fabric platform will include built in Azure Open AI based assistant that will serve all the workloads

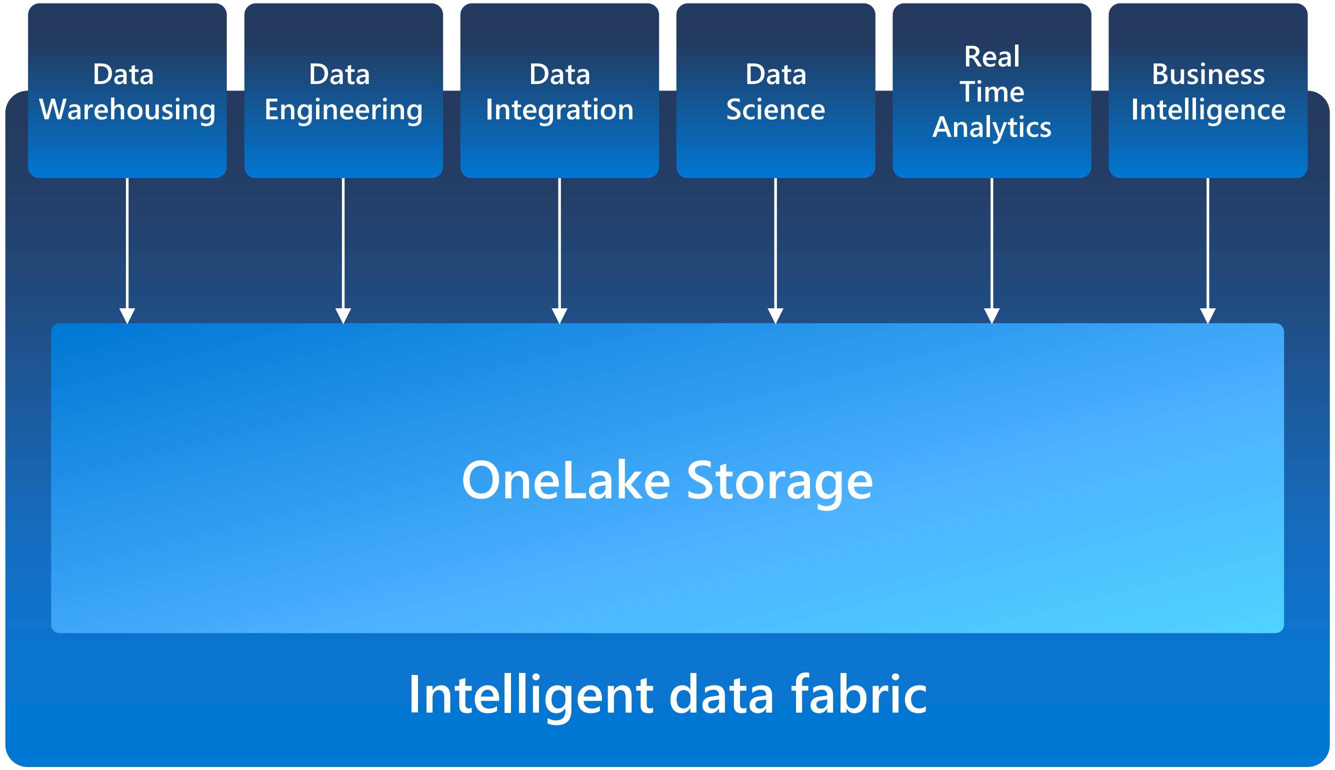
First GPT-based feature is already shipping in Power BI - NL2DAX – DAX calculation creation based on natural language prompts

Ongoing major ramp-up for pervasive AOAI based product-wide AI assistance

Lake-centric and open architecture

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

OneLake for all domains

A true data mesh across organization domains

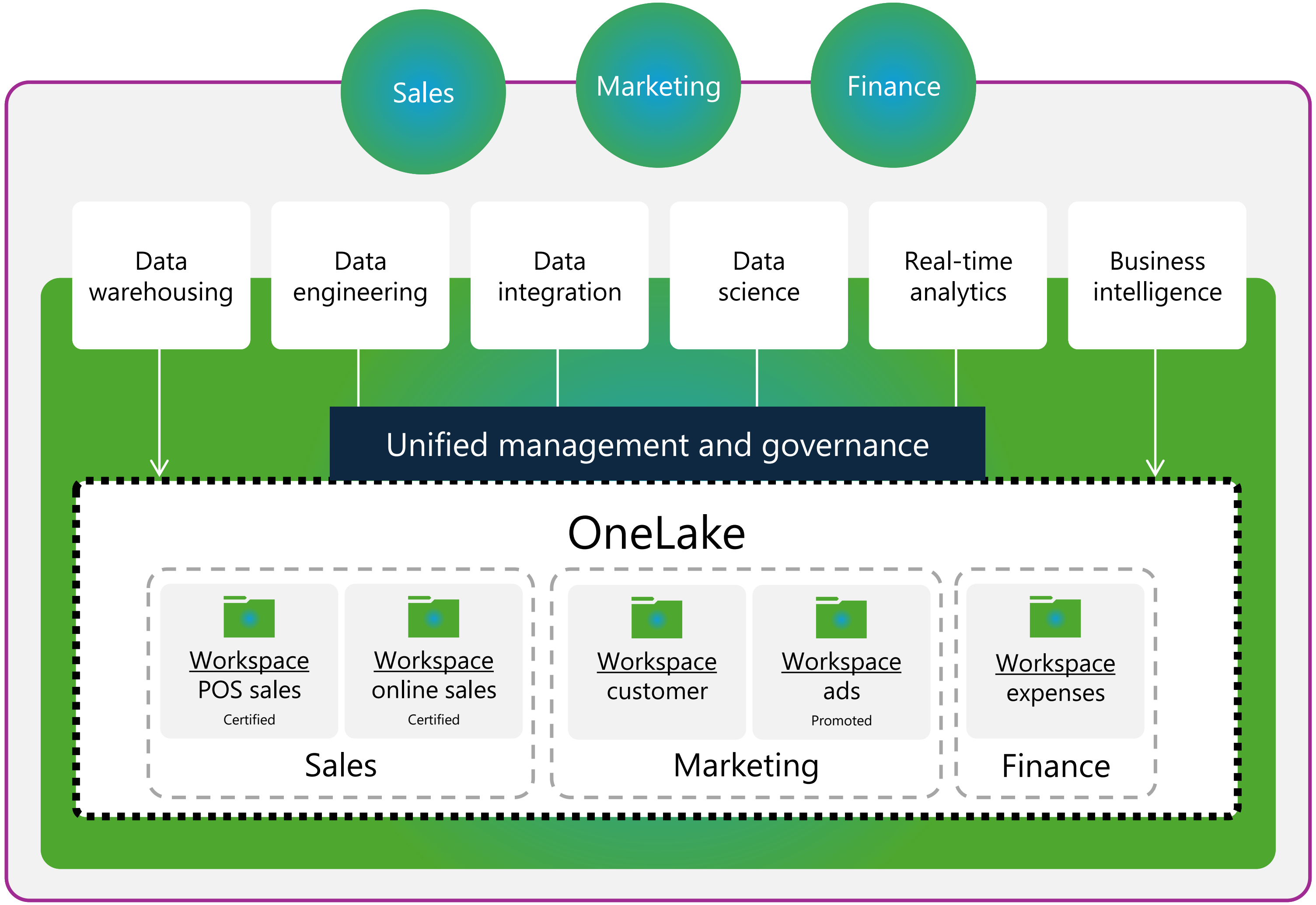
Introducing domains as an integral part of Fabric: A domain is a way to logically group together all the data in an organization relevant to an area or field, according to business needs

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

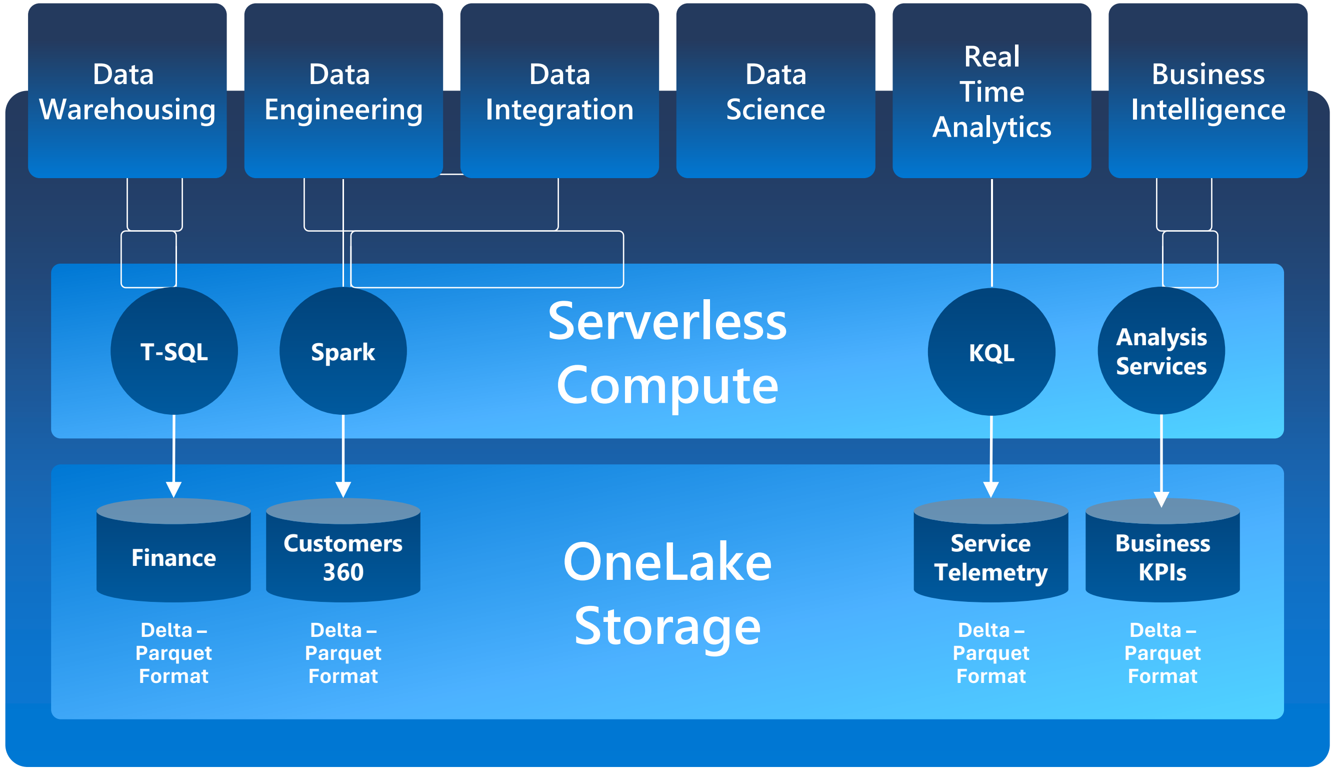
Domains simplify discovery and consumption of data across the organization, thus allowing business optimized consumption

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



One Copy for all computes

Real separation of compute and storage



All the compute engines store their data automatically in OneLake

The data is stored in a single common format

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

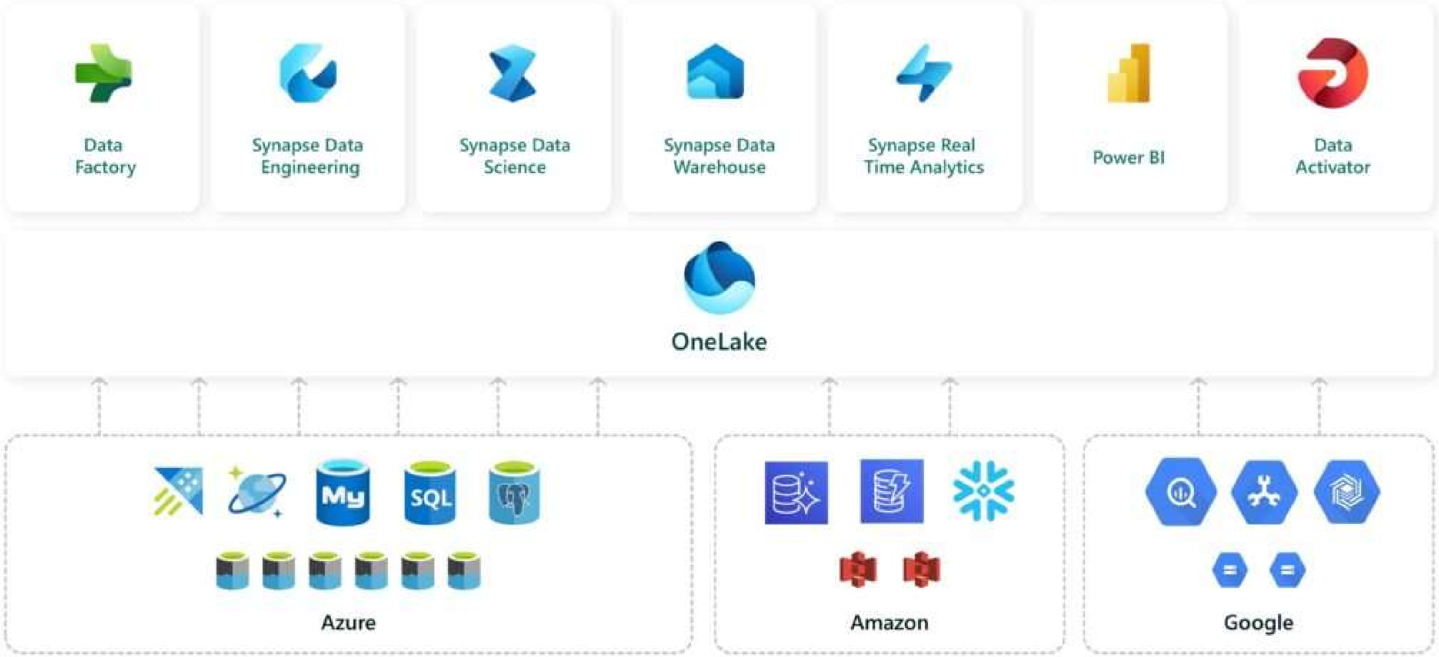
Once data is stored in the lake, it is directly accessible by all the engines without needing any import/export

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

Taking One Copy to the next level

Shortcuts

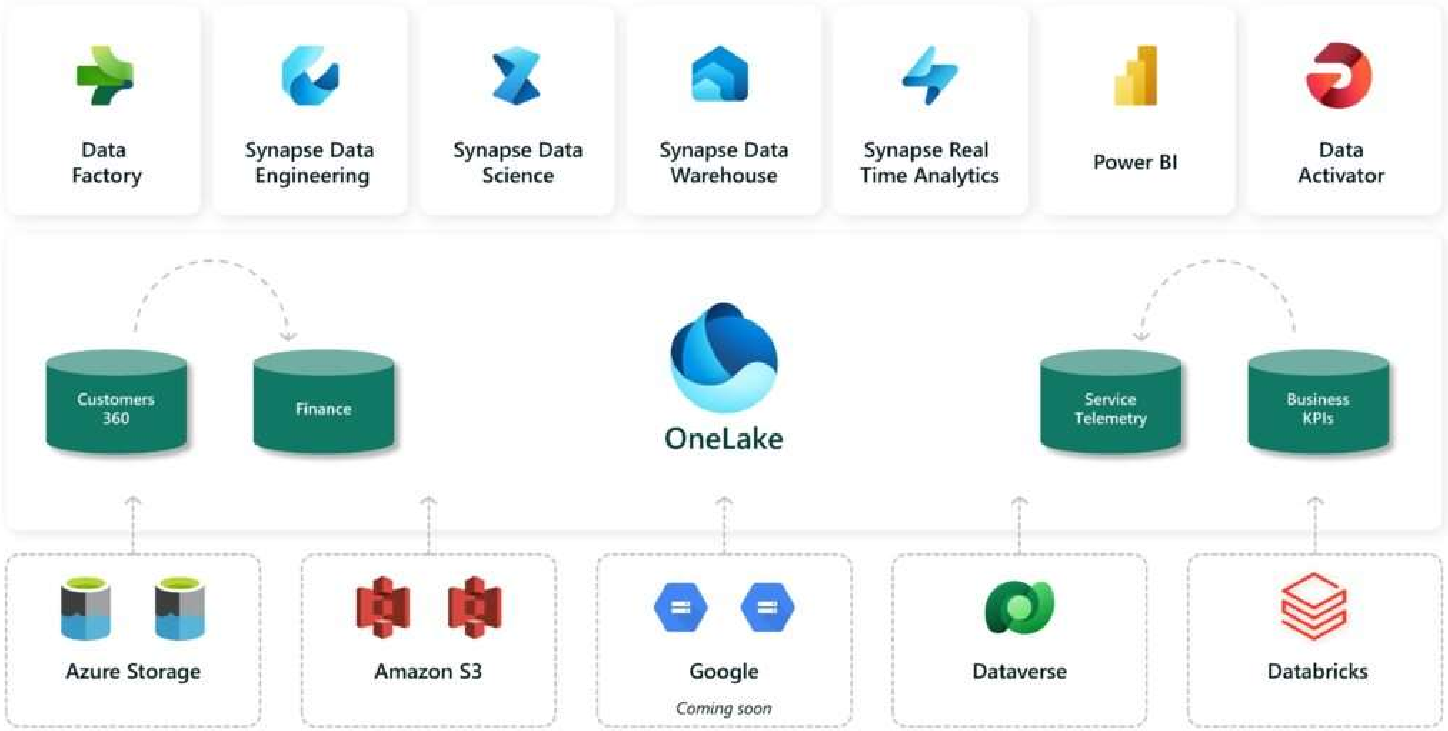


Coming soon

We will soon enable **Azure Cosmos DB**, **Azure SQL DB**, **Snowflake**, and **Mongo DB** customers to use mirroring to access their data in OneLake, with more data sources coming in 2024.

Taking One Copy to the next level

Shortcuts



Sharing data in OneLake is as easy as sharing files in OneDrive, removing the needs for data duplication

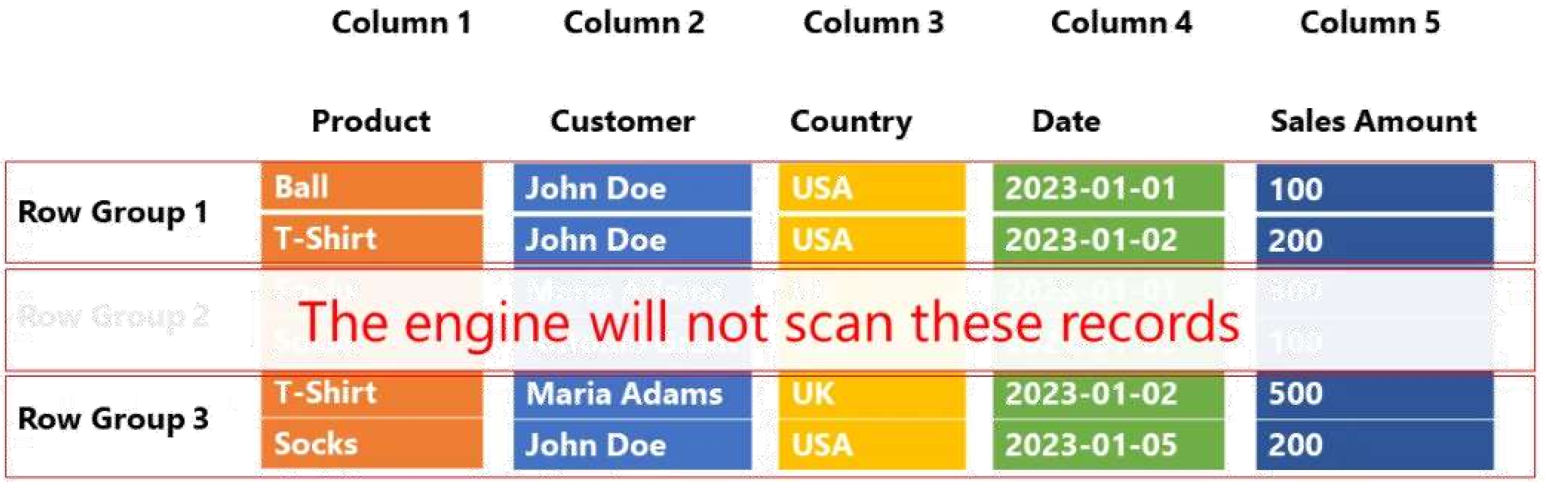
With **shortcuts**, data throughout OneLake can be composed together without any data movement

Shortcuts also allow instant linking of data already existing in Azure and in other clouds, without any data duplication and movement, making **OneLake the first multi-cloud data lake**

With support for industry standard APIs, OneLake data can be directly accessed by any application or service

Parquet file format

- Data compression
- Columnar Storage
- Language Agnostics
- Open-source format
- Support for complex data types
- Row-store vs Column-store



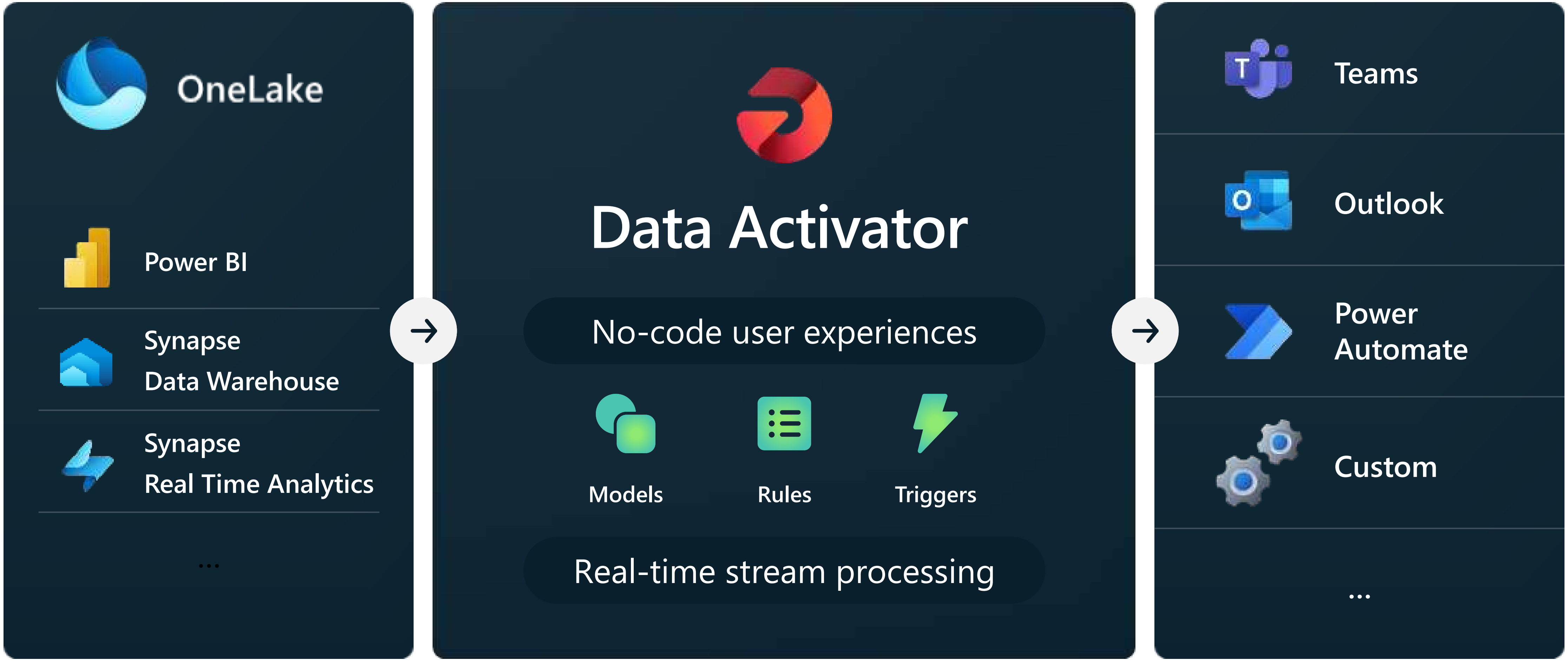
	Column 1	Column 2	Column 3	Column 4	Column 5
	Product	Customer	Country	Date	Sales Amount
Row Group 1	Ball	John Doe	USA	2023-01-01	100
	T-Shirt	John Doe	USA	2023-01-02	200
Row Group 2	Socks	Maria Adams	UK	2023-01-01	300
	Socks	[illegible]	[illegible]	[illegible]	100
Row Group 3	T-Shirt	Maria Adams	UK	2023-01-02	500
	Socks	John Doe	USA	2023-01-05	200

The engine will not scan these records

→ https://data-mozart.com/parquet-file-format-everything-you-need-to-know/

Data Activator

Trigger actions on all your data, from one place



OneLake

Power BI

Synapse Data Warehouse

Synapse Real Time Analytics

...

Data Activator

No-code user experiences

Models

Rules

Triggers

Real-time stream processing

Teams

Outlook

Power Automate

Custom

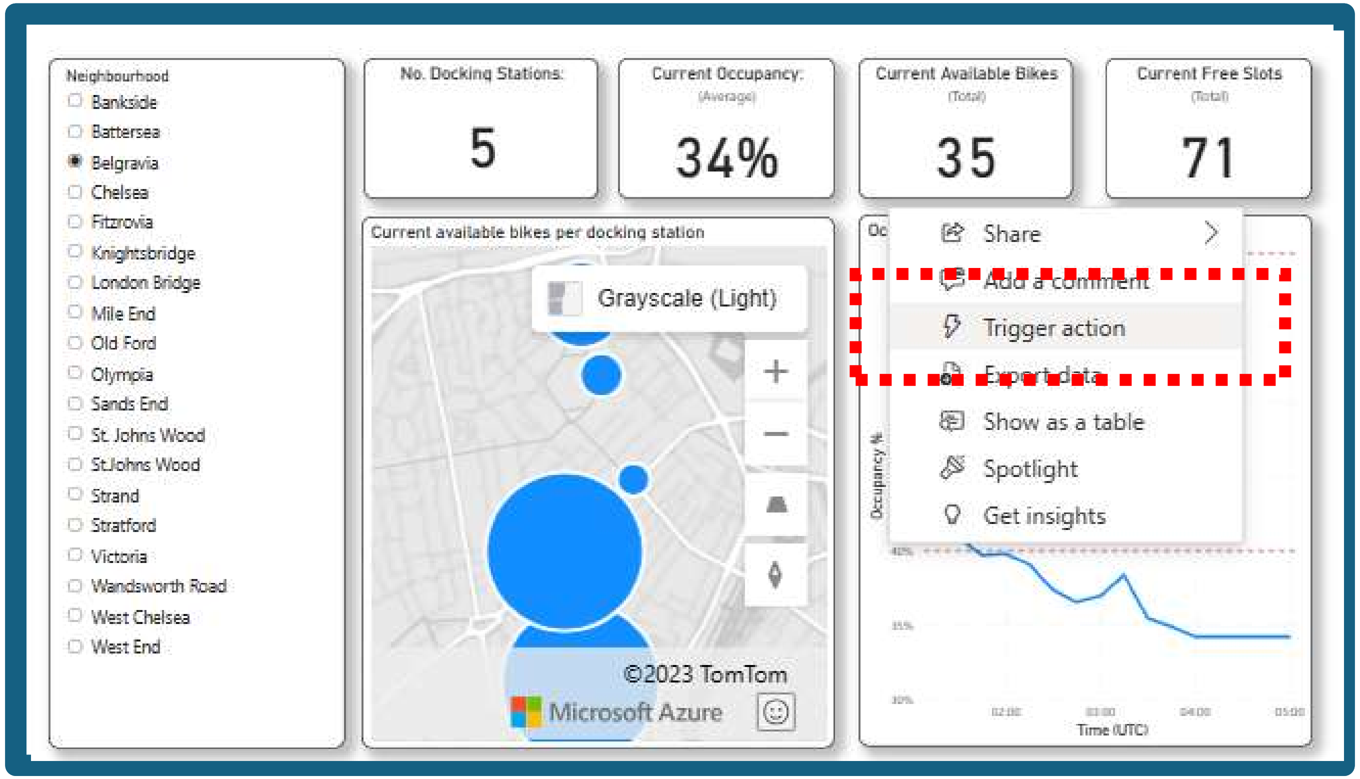
...

Trigger alerts & actions from Power BI visuals

Trigger alerts and actions from visuals in Power BI reports

Alert yourself, or define alerts that go to others in your organization.

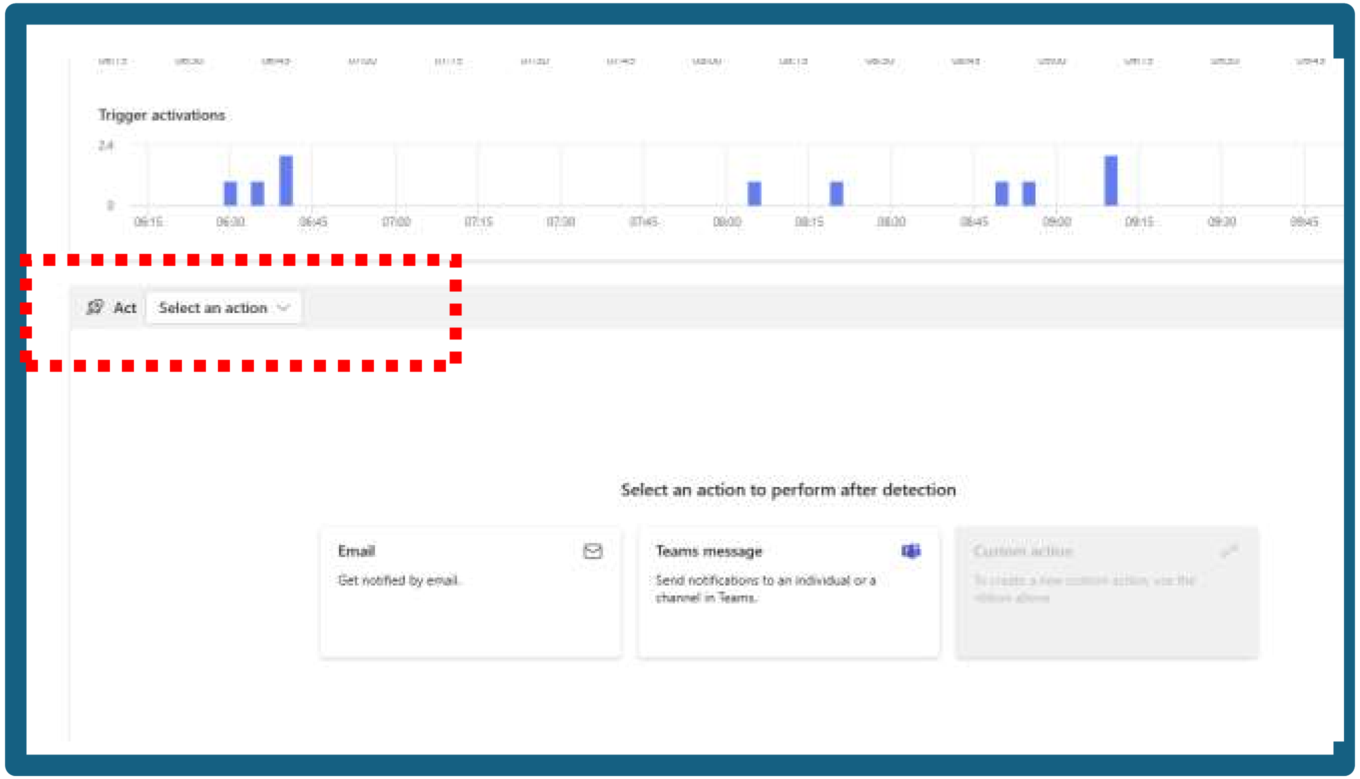
Supports both report consumers and report creators



Send alerts, or drive actions via Power Automate

Send alerts via Microsoft Teams, or via email when a trigger condition is met.

Trigger a Power Automate flow to connect to other alerting systems, to log a ticket in a ticketing system, and more



Trigger activations

Act Select an action

Select an action to perform after detection

Email
Get notified by email.

Teams message
Send notifications to an individual or a channel in Teams.

Custom action
To create a new custom action, use the ribbon above

Copilot & Explore

Copilot for...

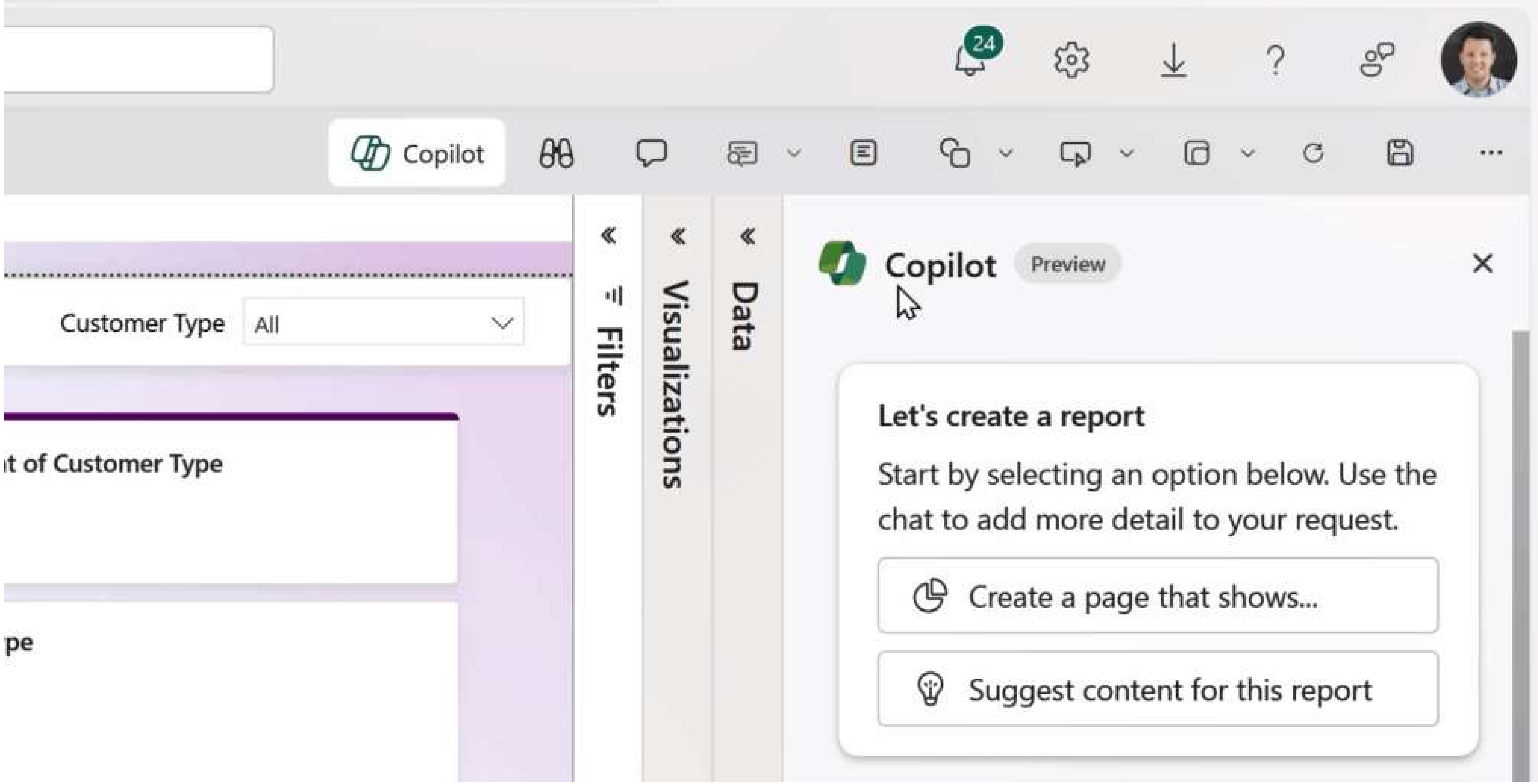
A new way to transform and analyze data, generate insights, and create visualizations and reports

Copilot and other **generative AI** features in preview

- Data Science and Data Engineering
- Data Factory
- Power BI

Copilot for...

A new way to transform and analyze data, generate insights, and create visualizations and reports

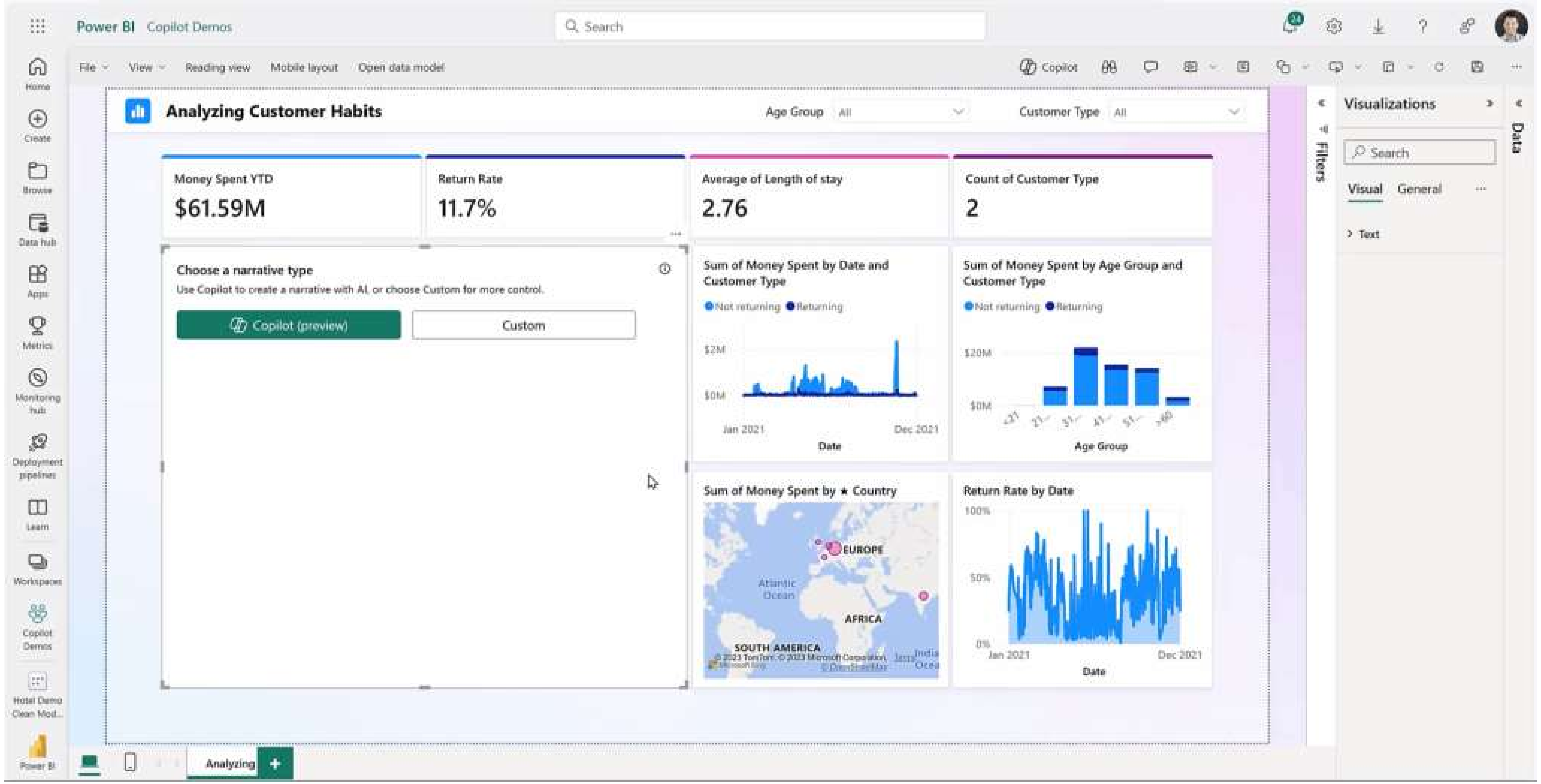


Let's create a report

Start by selecting an option below. Use the chat to add more detail to your request.

Copilot for...

A new way to transform and analyze data, generate insights, and create visualizations and reports



Choose a narrative type

Use Copilot to create a narrative with AI, or choose Custom for more control.

Money Spent YTD $61.59M

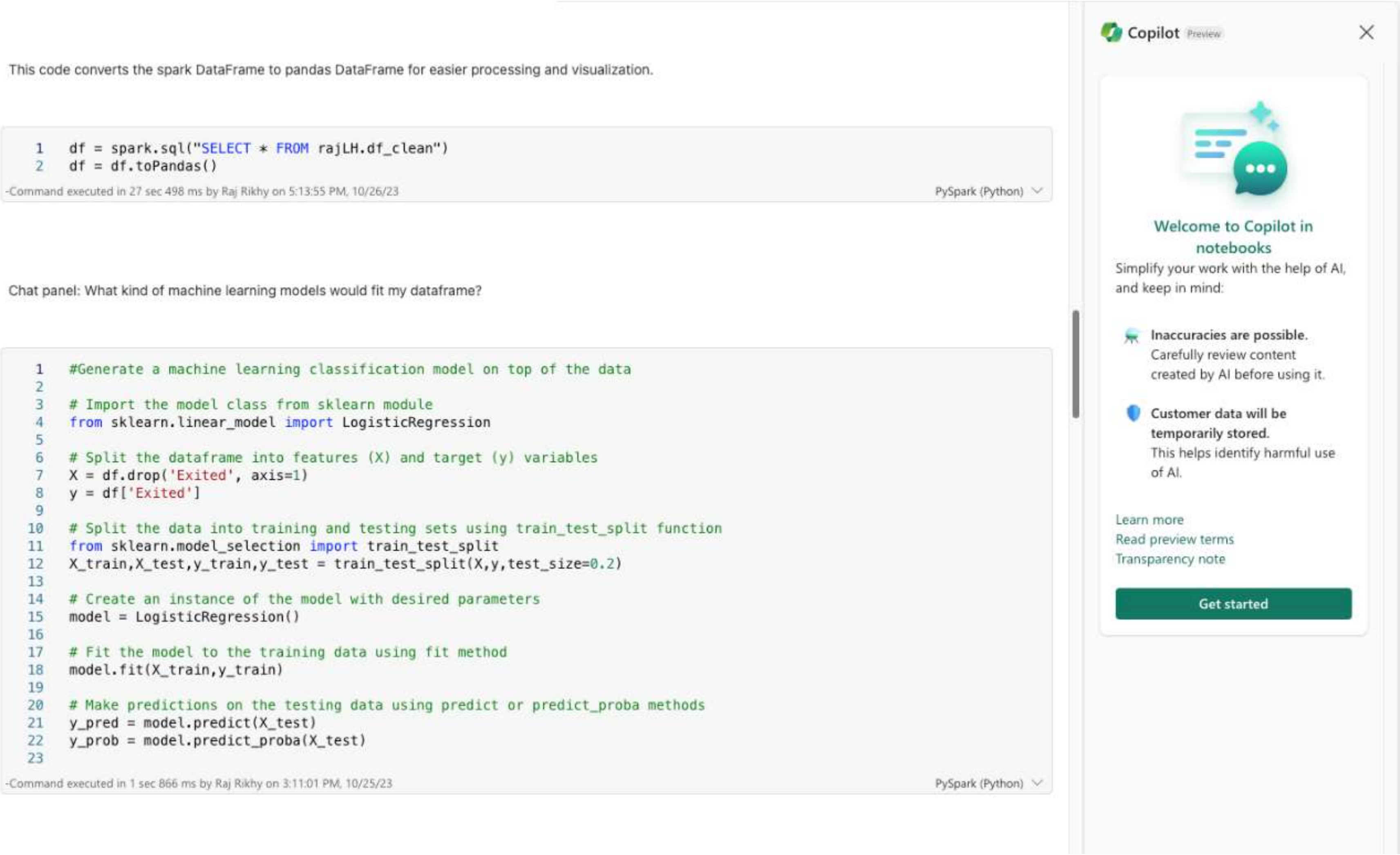
Return Rate 11.7%

Average of Length of stay 2.76

Count of Customer Type 2

Copilot for Data Science

Copilot Chat Panel



This code converts the spark DataFrame to pandas DataFrame for easier processing and visualization.

```
df = spark.sql("SELECT * FROM rajLH.df_clean")
df = df.toPandas()
```

-Command executed in 27 sec 498 ms by Raj Rikhy on 5:13:55 PM, 10/26/23

PySpark (Python)

Chat panel: What kind of machine learning models would fit my dataframe?

```
#Generate a machine learning classification model on top of the data

# Import the model class from sklearn module
from sklearn.linear_model import LogisticRegression

# Split the dataframe into features (X) and target (y) variables
X = df.drop('Exited', axis=1)
y = df['Exited']

# Split the data into training and testing sets using train_test_split function
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.2)

# Create an instance of the model with desired parameters
model = LogisticRegression()

# Fit the model to the training data using fit method
model.fit(X_train,y_train)

# Make predictions on the testing data using predict or predict_proba methods
y_pred = model.predict(X_test)
y_prob = model.predict_proba(X_test)

```

-Command executed in 1 sec 866 ms by Raj Rikhy on 3:11:01 PM, 10/25/23

PySpark (Python)

Welcome to Copilot in notebooks

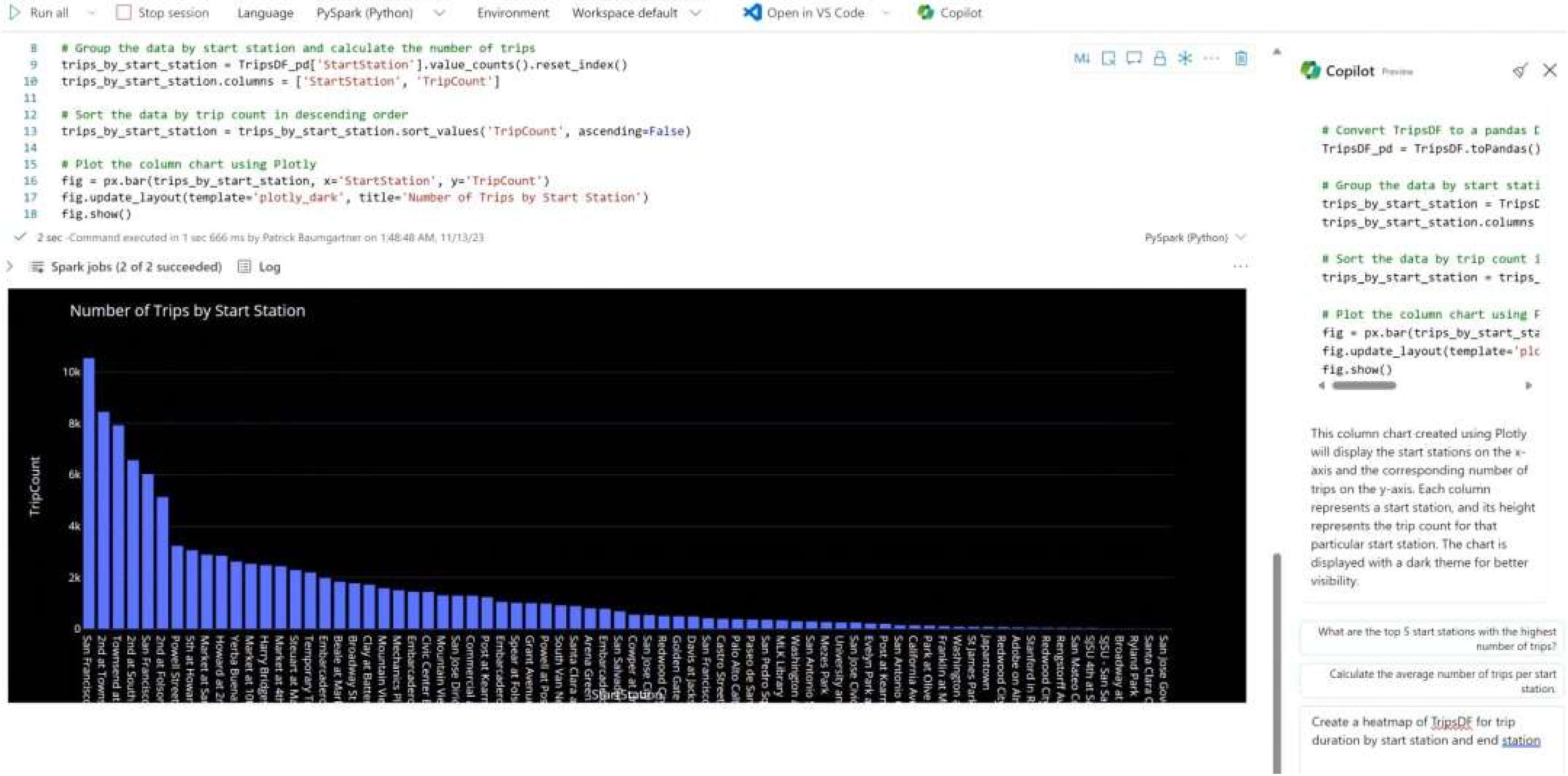
Simplify your work with the help of AI, and keep in mind:

Inaccuracies are possible. Carefully review content created by AI before using it.

Customer data will be temporarily stored. This helps identify harmful use of AI.

Copilot for Data Science

Copilot Chat Panel



```
# Group the data by start station and calculate the number of trips
trips_by_start_station = TripsDF_pd['StartStation'].value_counts().reset_index()
trips_by_start_station.columns = ['StartStation', 'TripCount']

# Sort the data by trip count in descending order
trips_by_start_station = trips_by_start_station.sort_values('TripCount', ascending=False)

# Plot the column chart using Plotly
fig = px.bar(trips_by_start_station, x='StartStation', y='TripCount')
fig.update_layout(template='plotly_dark', title='Number of Trips by Start Station')
fig.show()
```

✓ 2 sec -Command executed in 1 sec 666 ms by Patrick Baumgartner on 1:48:48 AM, 11/13/23

```
# Convert TripsDF to a pandas D
TripsDF_pd = TripsDF.toPandas()

# Group the data by start stati
trips_by_start_station = TripsD
trips_by_start_station.columns

# Sort the data by trip count i
trips_by_start_station = trips_

# Plot the column chart using P
fig = px.bar(trips_by_start_sta
fig.update_layout(template='plo
fig.show()
```

This column chart created using Plotly will display the start stations on the x-axis and the corresponding number of trips on the y-axis. Each column represents a start station, and its height represents the trip count for that particular start station. The chart is displayed with a dark theme for better visibility.

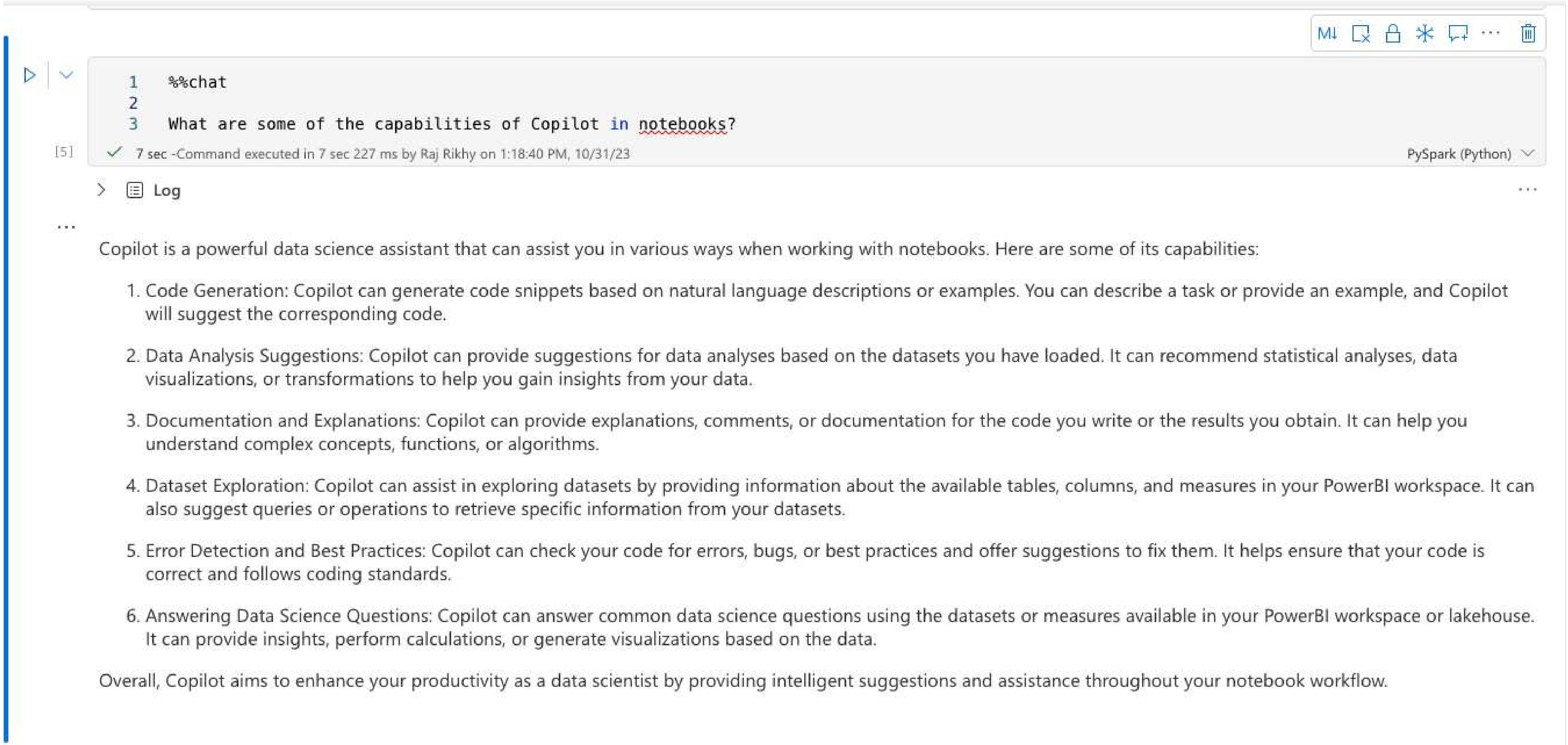
What are the top 5 start stations with the highest number of trips?

Calculate the average number of trips per start station.

Create a heatmap of TripsDF for trip duration by start station and end station

Copilot for Data Science

Chat Magics



```
%%chat

What are some of the capabilities of Copilot in notebooks?
```

✓ 7 sec -Command executed in 7 sec 227 ms by Raj Rikhy on 1:18:40 PM, 10/31/23

Copilot is a powerful data science assistant that can assist you in various ways when working with notebooks. Here are some of its capabilities:

1. Code Generation: Copilot can generate code snippets based on natural language descriptions or examples. You can describe a task or provide an example, and Copilot will suggest the corresponding code.

2. Data Analysis Suggestions: Copilot can provide suggestions for data analyses based on the datasets you have loaded. It can recommend statistical analyses, data visualizations, or transformations to help you gain insights from your data.

3. Documentation and Explanations: Copilot can provide explanations, comments, or documentation for the code you write or the results you obtain. It can help you understand complex concepts, functions, or algorithms.

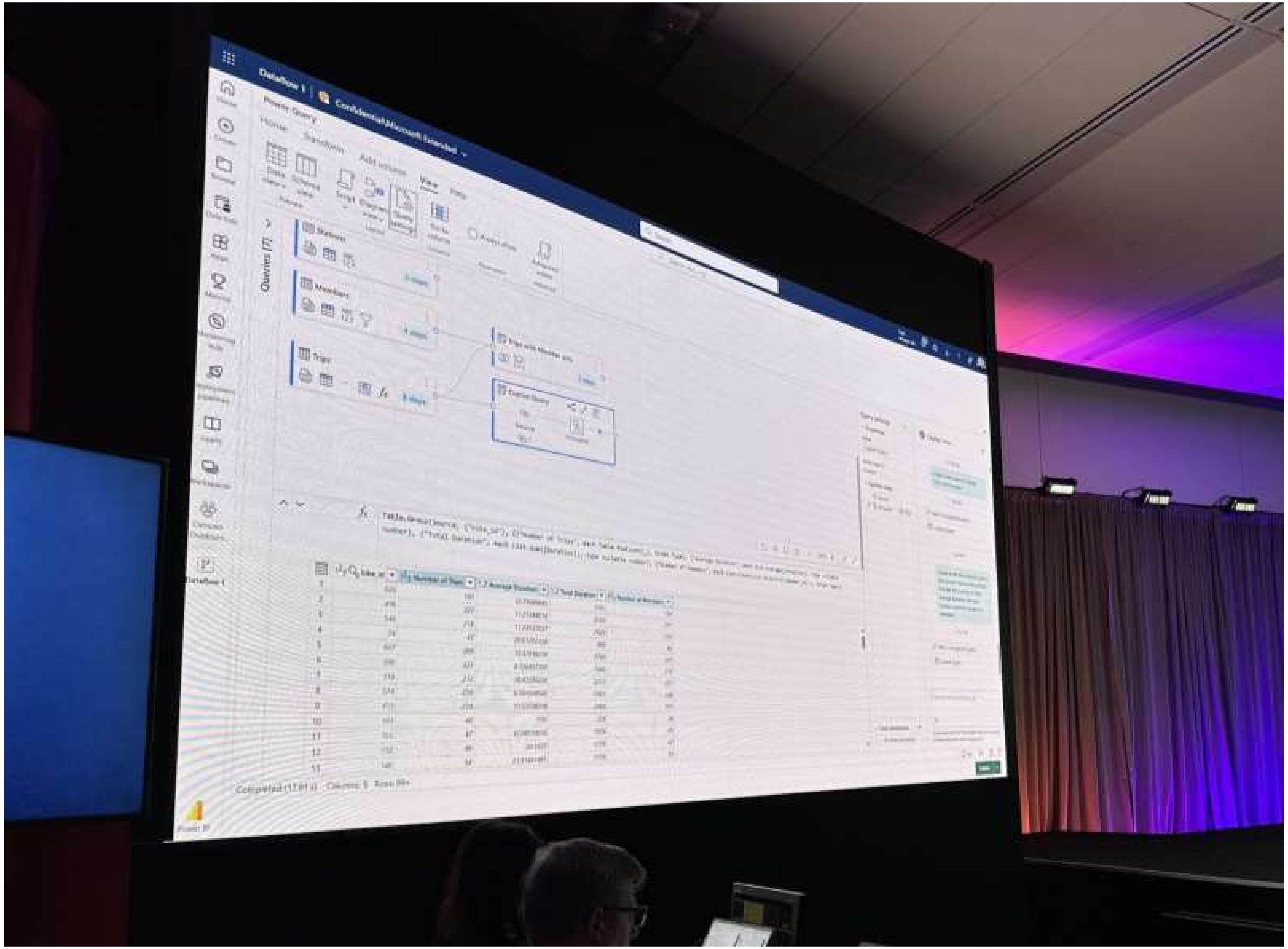
4. Dataset Exploration: Copilot can assist in exploring datasets by providing information about the available tables, columns, and measures in your PowerBI workspace. It can also suggest queries or operations to retrieve specific information from your datasets.

5. Error Detection and Best Practices: Copilot can check your code for errors, bugs, or best practices and offer suggestions to fix them. It helps ensure that your code is correct and follows coding standards.

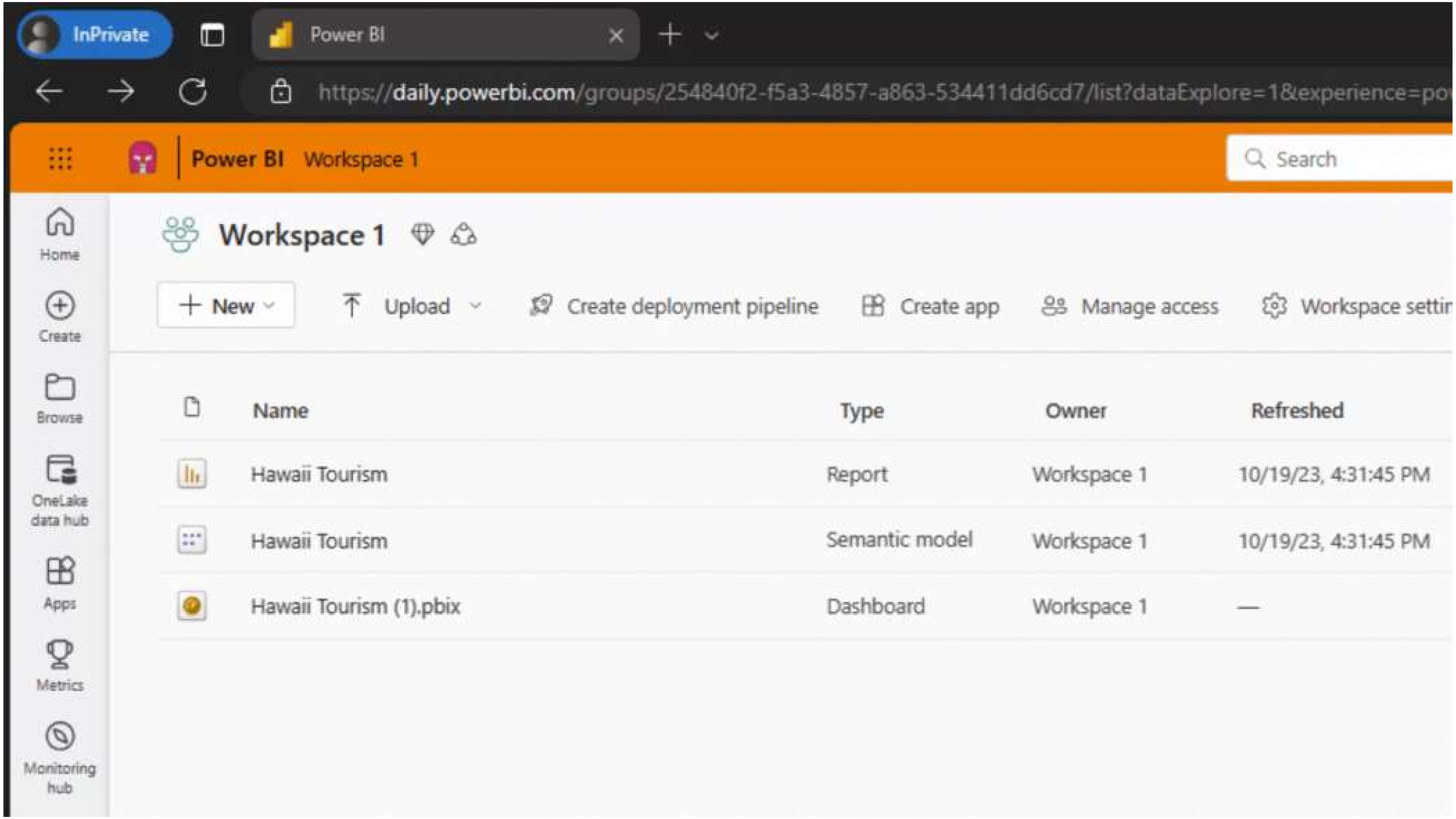
6. Answering Data Science Questions: Copilot can answer common data science questions using the datasets or measures available in your PowerBI workspace or lakehouse. It can provide insights, perform calculations, or generate visualizations based on the data.

Overall, Copilot aims to enhance your productivity as a data scientist by providing intelligent suggestions and assistance throughout your notebook workflow.

Copilot for DataFlow

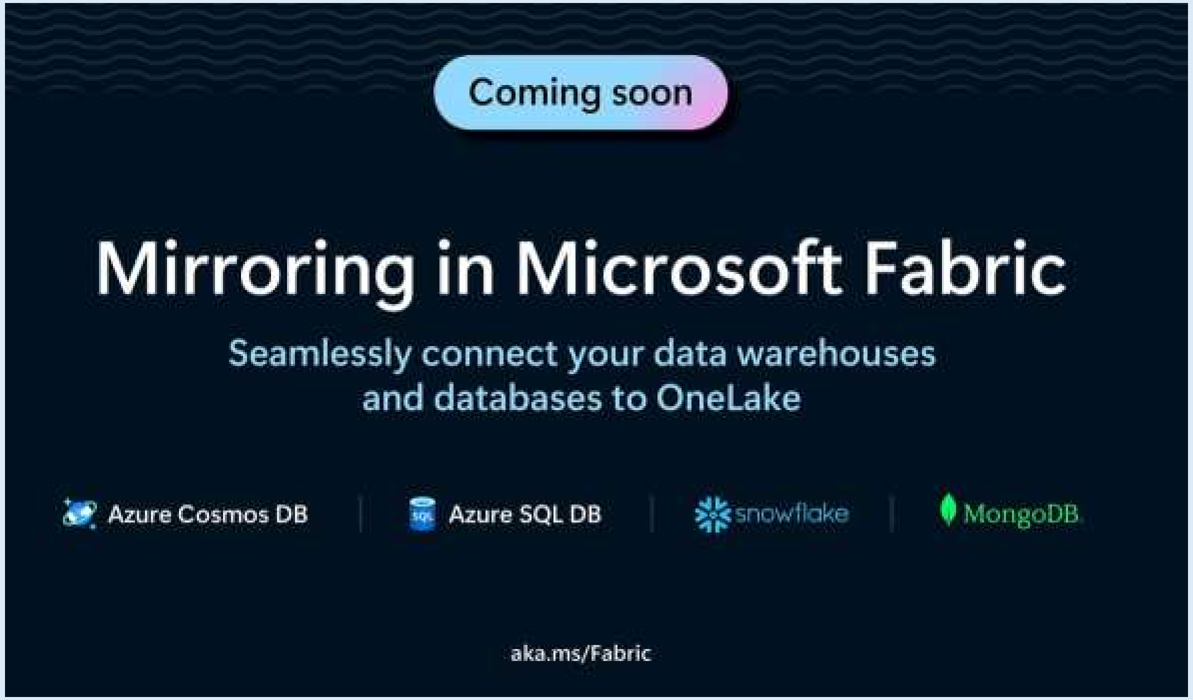


Quickly answer your data questions with Explore



Name	Type	Owner	Refreshed
Hawaii Tourism	Report	Workspace 1	10/19/23, 4:31:45 PM
Hawaii Tourism	Semantic model	Workspace 1	10/19/23, 4:31:45 PM
Hawaii Tourism (1).pbix	Dashboard	Workspace 1	—

Mirroring



Coming soon

Mirroring in Microsoft Fabric

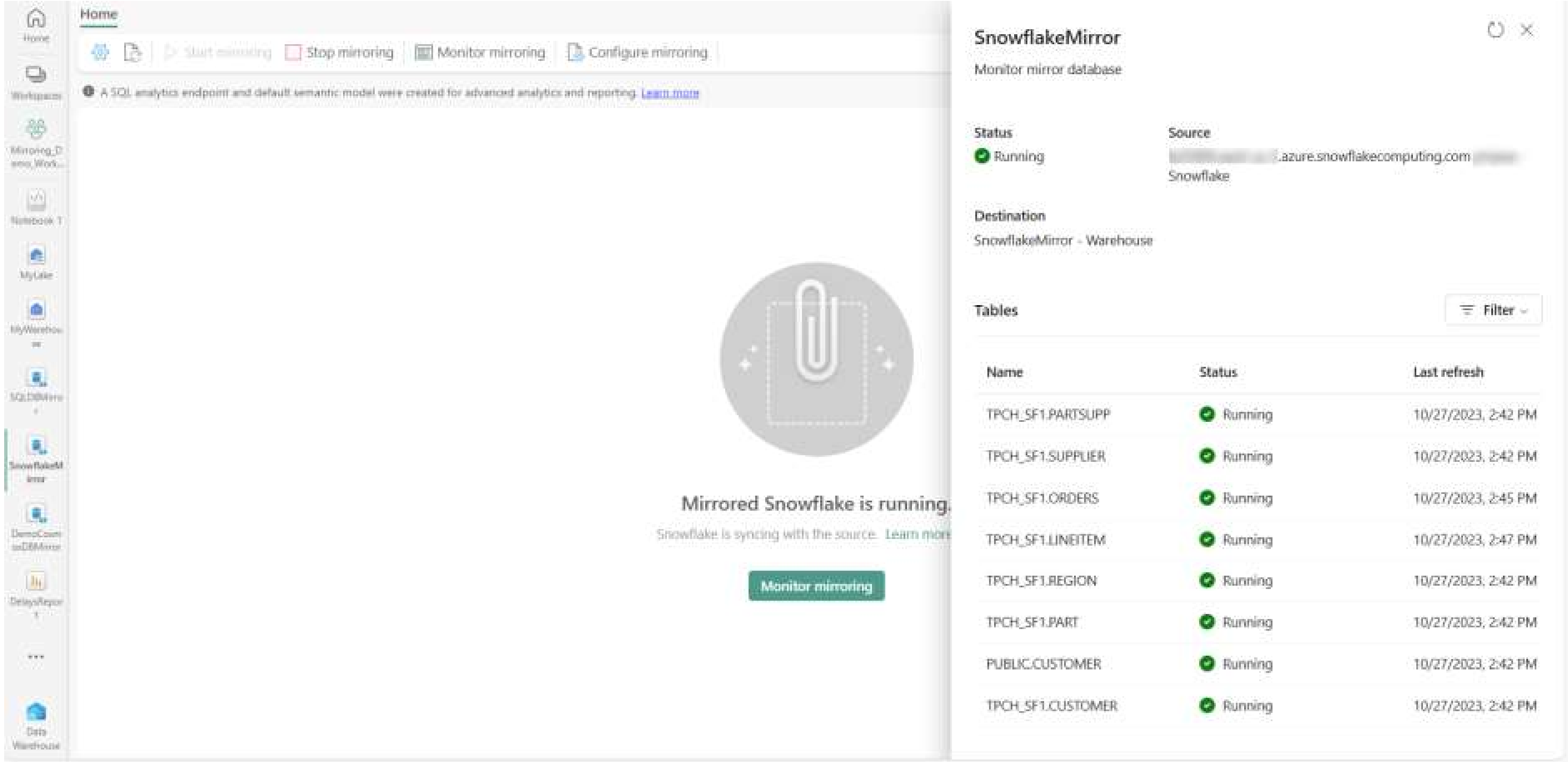
Seamlessly connect your data warehouses and databases to OneLake

Azure Cosmos DB | Azure SQL DB | snowflake | MongoDB.

aka.ms/Fabric

Mirroring in Microsoft Fabric

- By just providing connection details, your db is instantly available in Fabric as a Mirrored database
- **Real-time data replication**



A SQL analytics endpoint and default semantic model were created for advanced analytics and reporting. Learn more

Mirrored Snowflake is running.

Snowflake is syncing with the source. Learn more

Monitor mirroring

SnowflakeMirror

Monitor mirror database

Status: Running

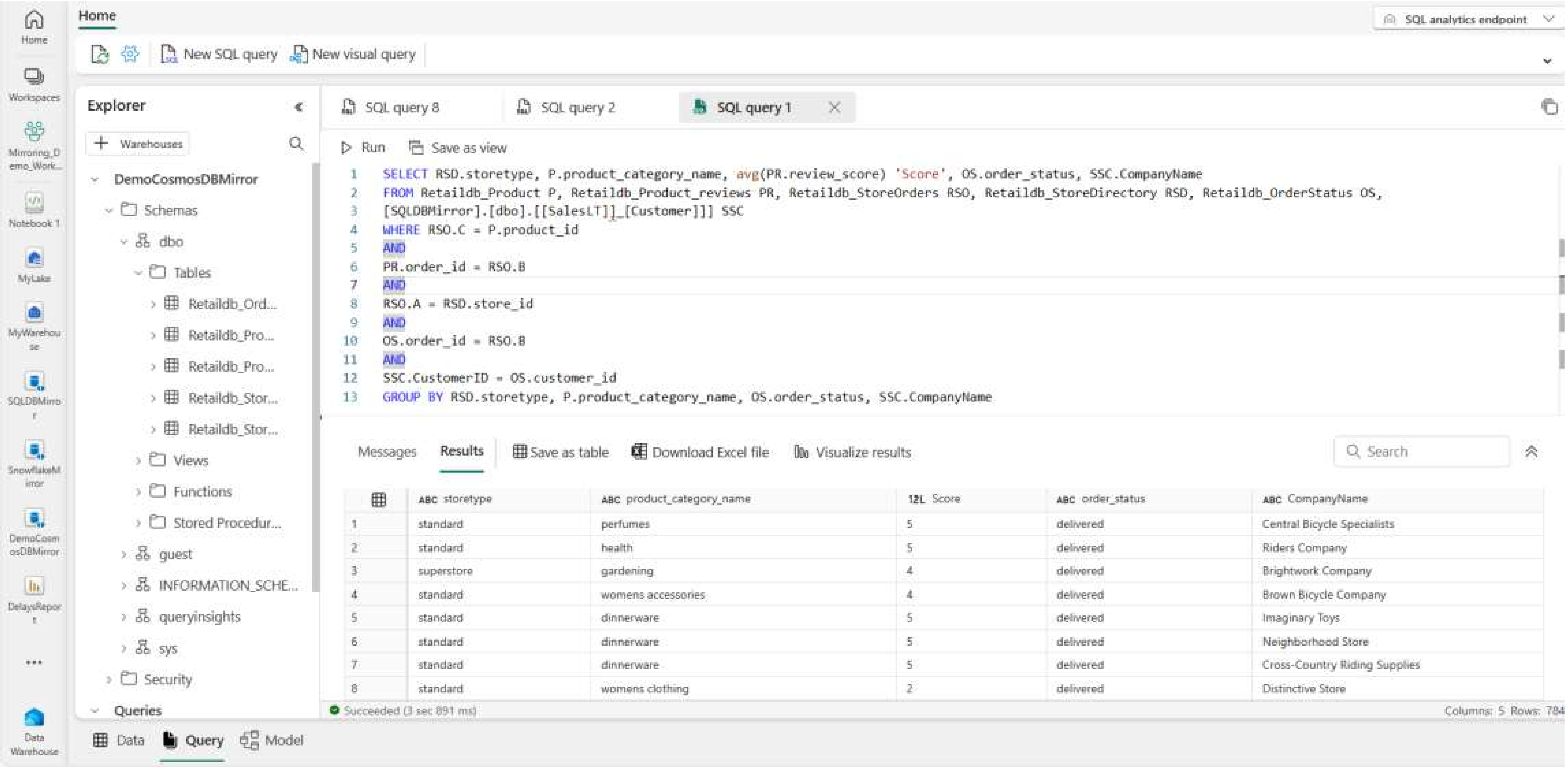
Source: [redacted].azure.snowflakecomputing.com [redacted]
Snowflake

Destination: SnowflakeMirror - Warehouse

Tables

Name	Status	Last refresh
TPCH_SF1.PARTSUPP	Running	10/27/2023, 2:42 PM
TPCH_SF1.SUPPLIER	Running	10/27/2023, 2:42 PM
TPCH_SF1.ORDERS	Running	10/27/2023, 2:45 PM
TPCH_SF1.LINEITEM	Running	10/27/2023, 2:47 PM
TPCH_SF1.REGION	Running	10/27/2023, 2:42 PM
TPCH_SF1.PART	Running	10/27/2023, 2:42 PM
PUBLIC.CUSTOMER	Running	10/27/2023, 2:42 PM
TPCH_SF1.CUSTOMER	Running	10/27/2023, 2:42 PM

Cross-joining Mirrored db, Warehouses, Lakehouses



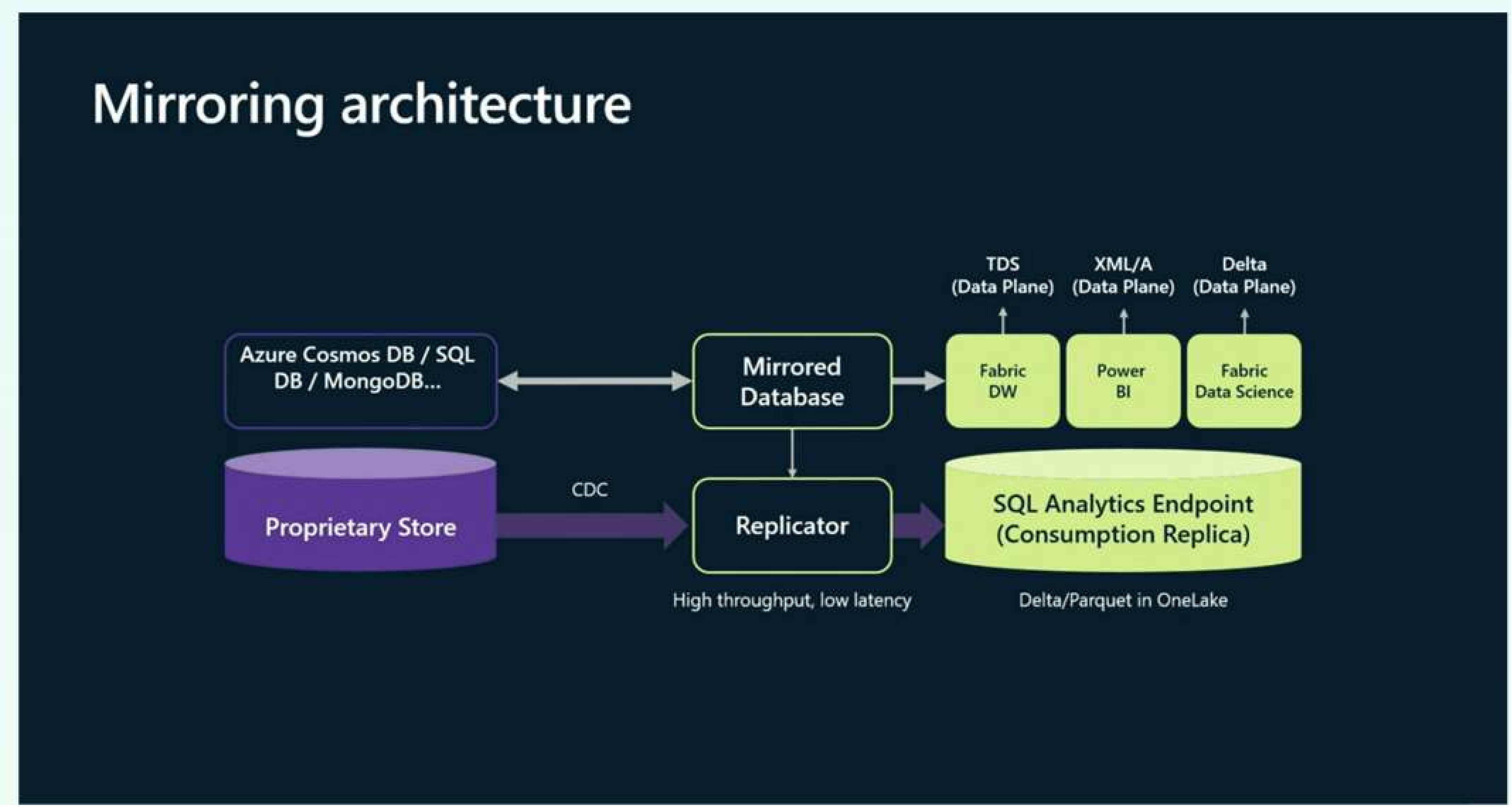
```sql
SELECT RSD.storetype, P.product_category_name, avg(PR.review_score) 'Score', OS.order_status, SSC.CompanyName
FROM Retaildb_Product P, Retaildb_Product_reviews PR, Retaildb_StoreOrders RSO, Retaildb_StoreDirectory RSD, Retaildb_OrderStatus OS,
[SQLDBMirror].[dbo].[[SalesLT]]_[Customer]]] SSC
WHERE RSO.C = P.product_id
AND
PR.order_id = RSO.B
AND
RSO.A = RSD.store_id
AND
OS.order_id = RSO.B
AND
SSC.CustomerID = OS.customer_id
GROUP BY RSD.storetype, P.product_category_name, OS.order_status, SSC.CompanyName
```

	storetype	product_category_name	Score	order_status	CompanyName
1	standard	perfumes	5	delivered	Central Bicycle Specialists
2	standard	health	5	delivered	Riders Company
3	superstore	gardening	4	delivered	Brightwork Company
4	standard	womens accessories	4	delivered	Brown Bicycle Company
5	standard	dinnerware	5	delivered	Imaginary Toys
6	standard	dinnerware	5	delivered	Neighborhood Store
7	standard	dinnerware	5	delivered	Cross-Country Riding Supplies
8	standard	womens clothing	2	delivered	Distinctive Store

Succeeded (3 sec 891 ms)

Columns: 5 Rows: 784

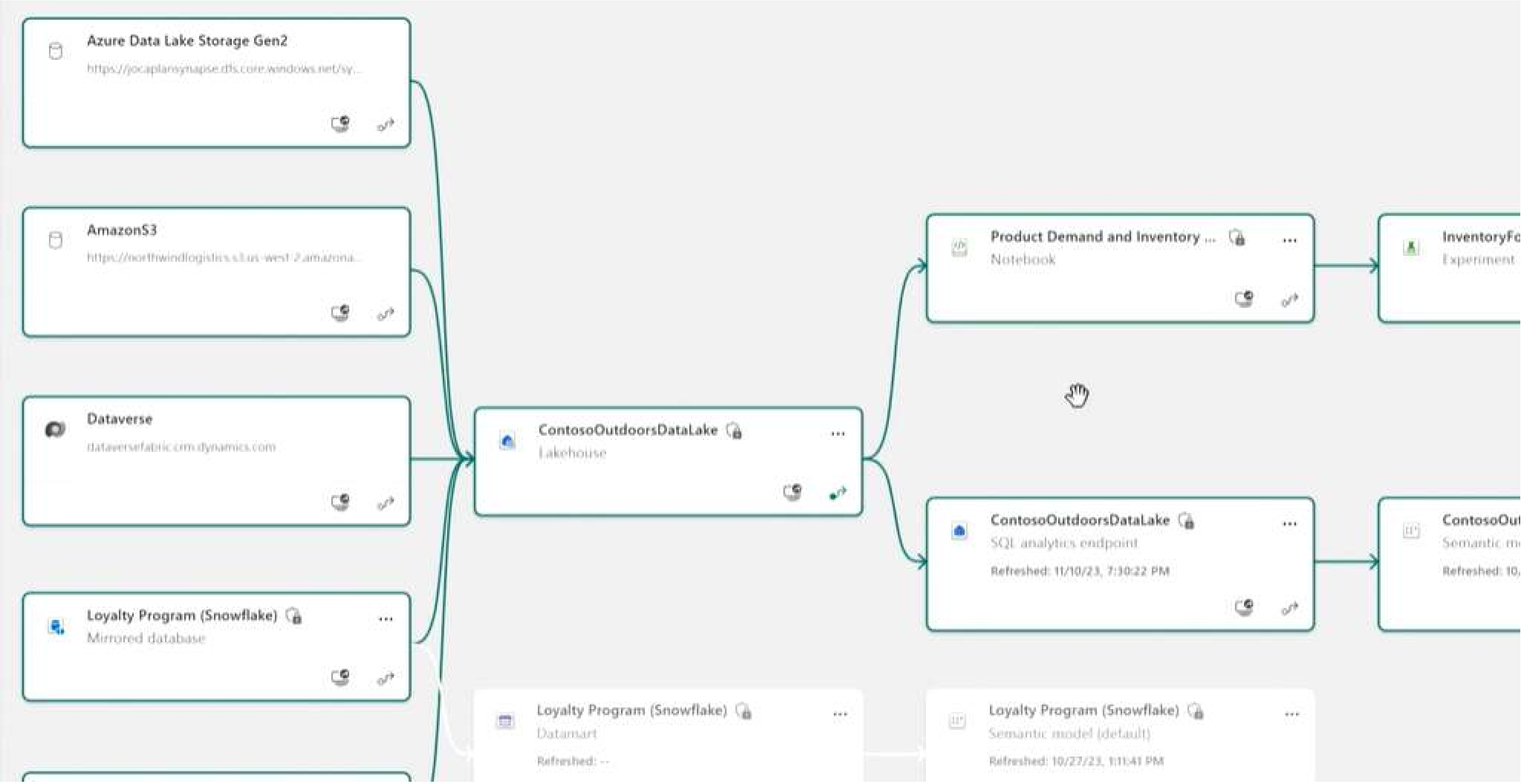
Mirroring Architecture



Mirroring architecture

Data Lineage

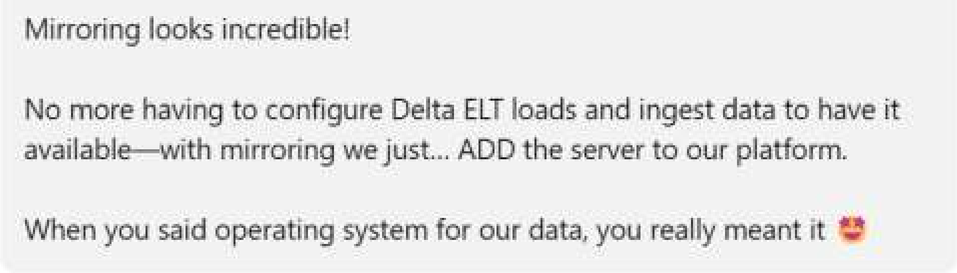
Includes and spans over to mirrored systems !


Azure Data Lake Storage Gen2
https://jocaplansynapise.dfs.core.windows.net/sy...
AmazonS3
https://northwindlogistics.s3.us-west-2.amazona...
Dataverse
dataversefabric.crm.dynamics.com
Loyalty Program (Snowflake)
Mirrored database
ContosoOutdoorsDataLake
Lakehouse
Product Demand and Inventory ...
Notebook
InventoryFo
Experiment
ContosoOutdoorsDataLake
SQL analytics endpoint
Refreshed: 11/10/23, 7:30:22 PM
ContosoOut
Semantic m
Refreshed: 10,
Loyalty Program (Snowflake)
Datamart
Refreshed: --
Loyalty Program (Snowflake)
Semantic model (default)
Refreshed: 10/27/23, 1:11:41 PM

Mirroring Availability

Azure Cosmos DB, **Azure SQL DB** and **Snowflake** customers will be able to use Mirroring to mirror their data in OneLake and unlock all the capabilities of Fabric Warehouse, Direct Lake Mode, Notebooks and much more.

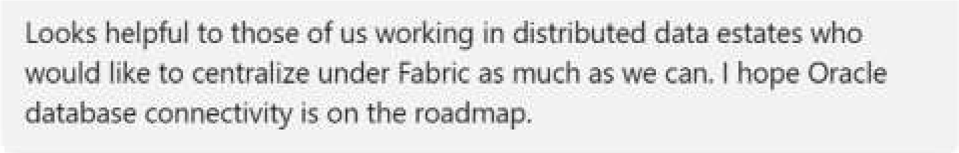
SQL Server, **Azure PostgreSQL**, **Azure MySQL**, **MongoDB** and other databases and data warehouses will be coming in CY24



Mirroring looks incredible!

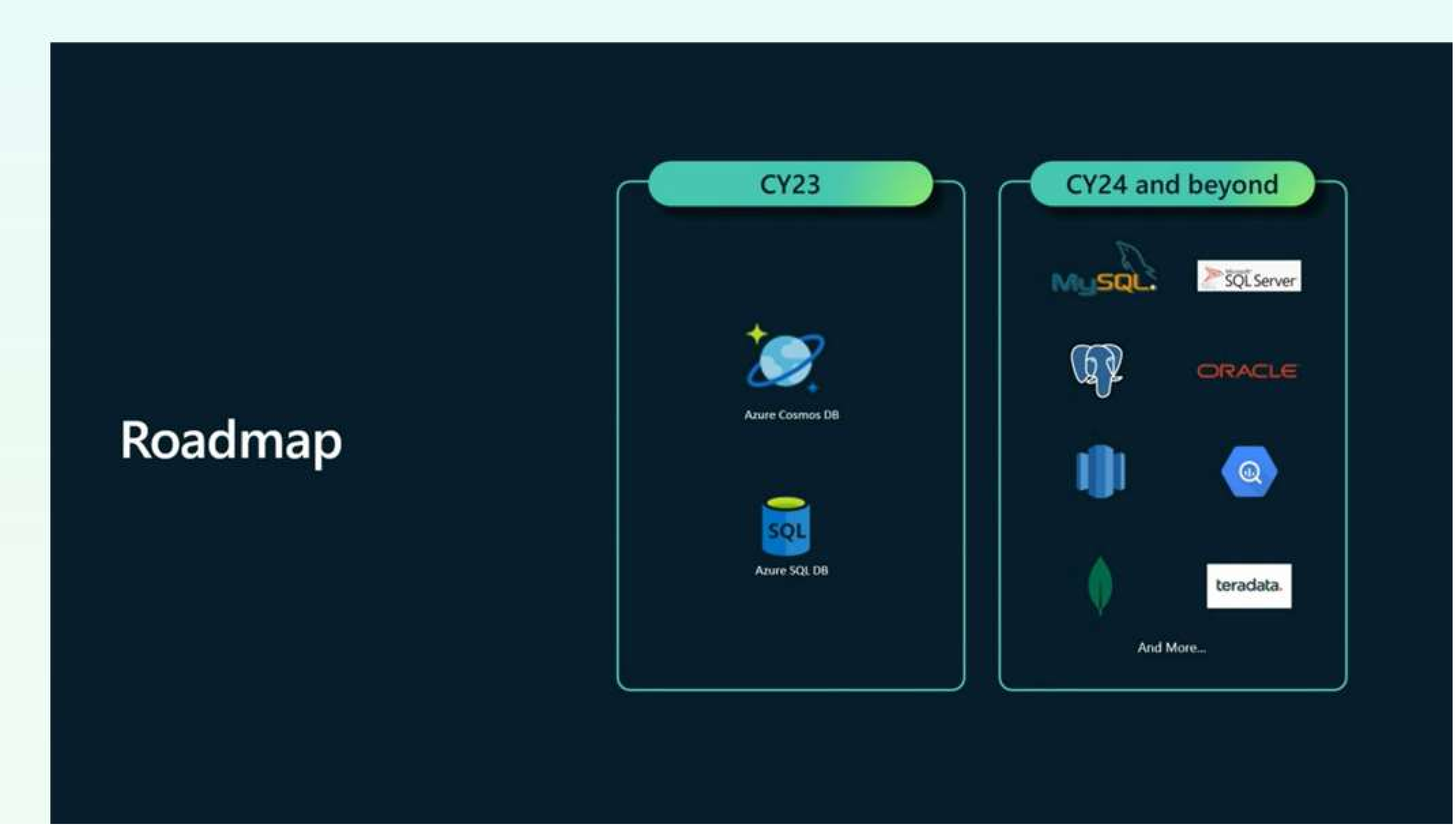
No more having to configure Delta ELT loads and ingest data to have it available—with mirroring we just... ADD the server to our platform.

When you said operating system for our data, you really meant it 😍



Looks helpful to those of us working in distributed data estates who would like to centralize under Fabric as much as we can. I hope Oracle database connectivity is on the roadmap.

Mirroring Roadmap



Roadmap

CY23

Azure Cosmos DB

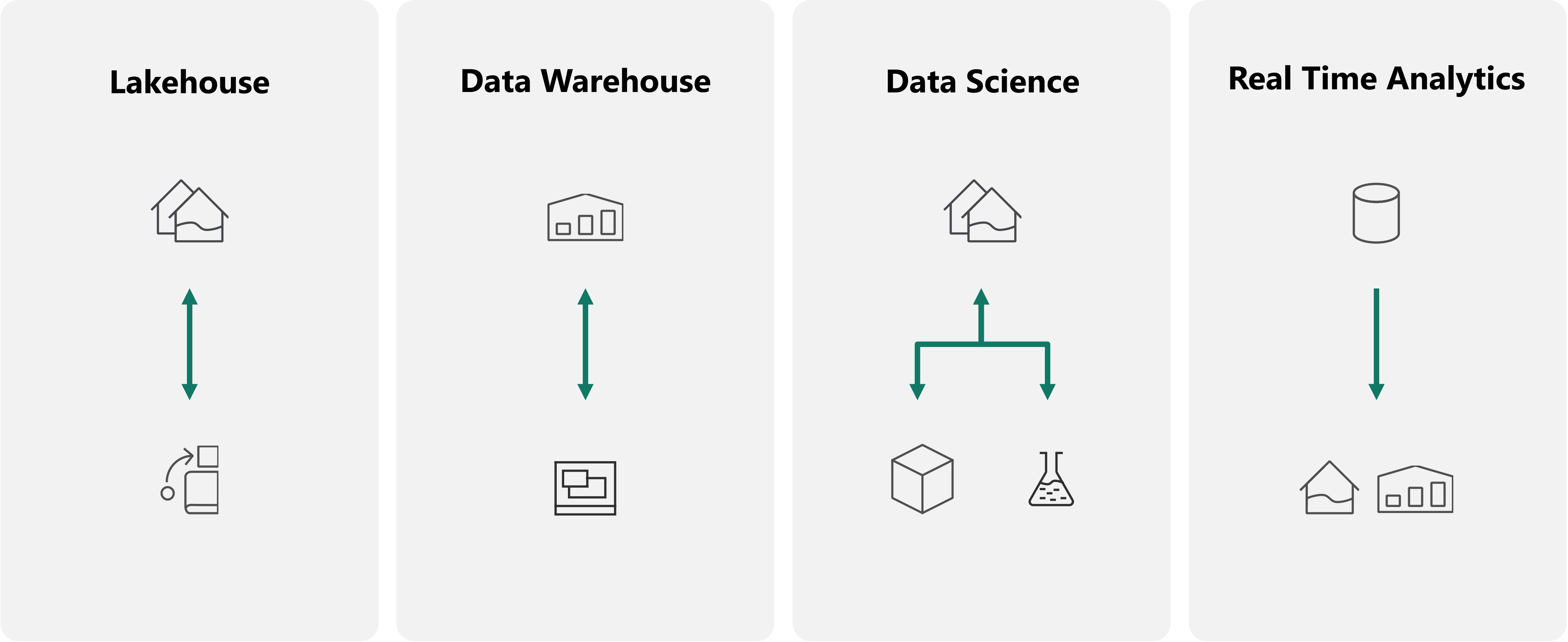
Azure SQL DB

CY24 and beyond

And More...

Scenarios

Common analytics scenarios



Lakehouse

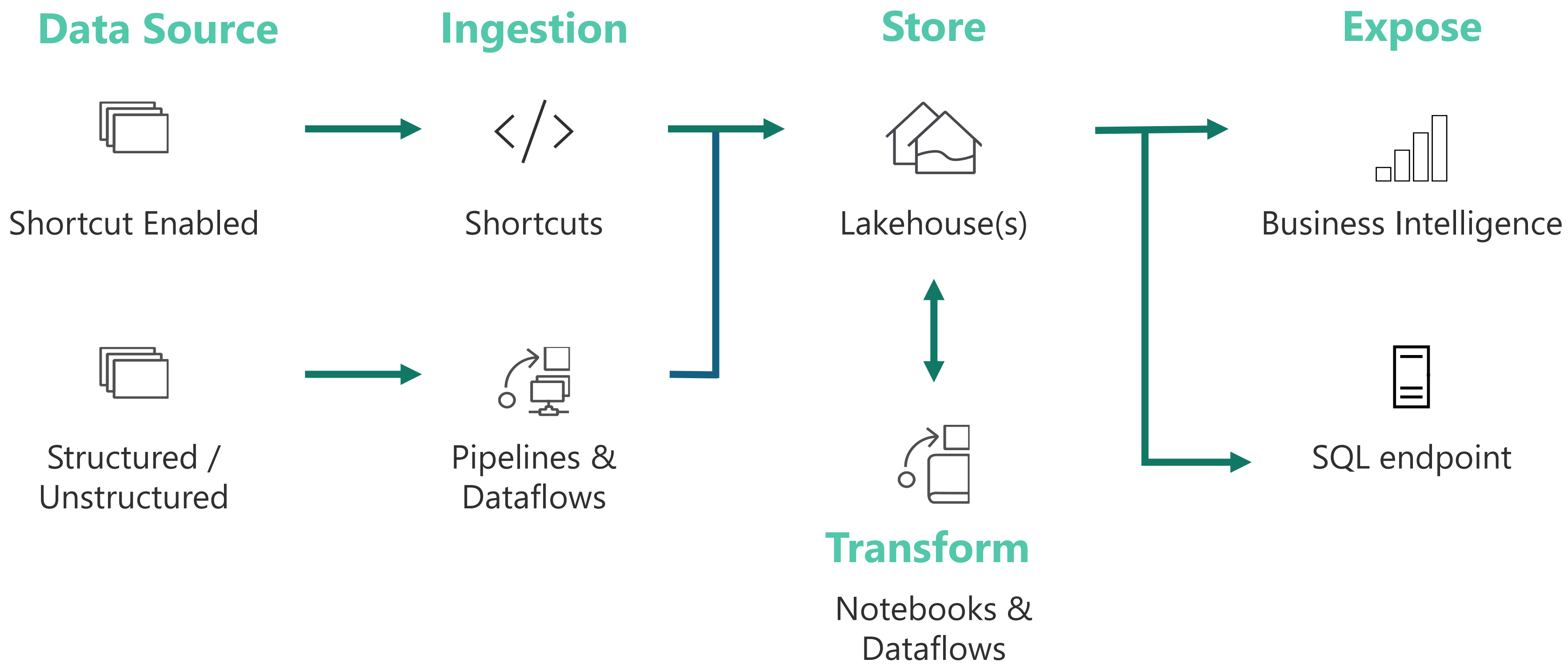
Data Warehouse

Data Science

Real Time Analytics

Lakehouse scenario

End-to-end analytics scenario



Data Source

Shortcut Enabled

Structured / Unstructured

Ingestion

Shortcuts

Pipelines & Dataflows

Store

Lakehouse(s)

Transform

Notebooks & Dataflows

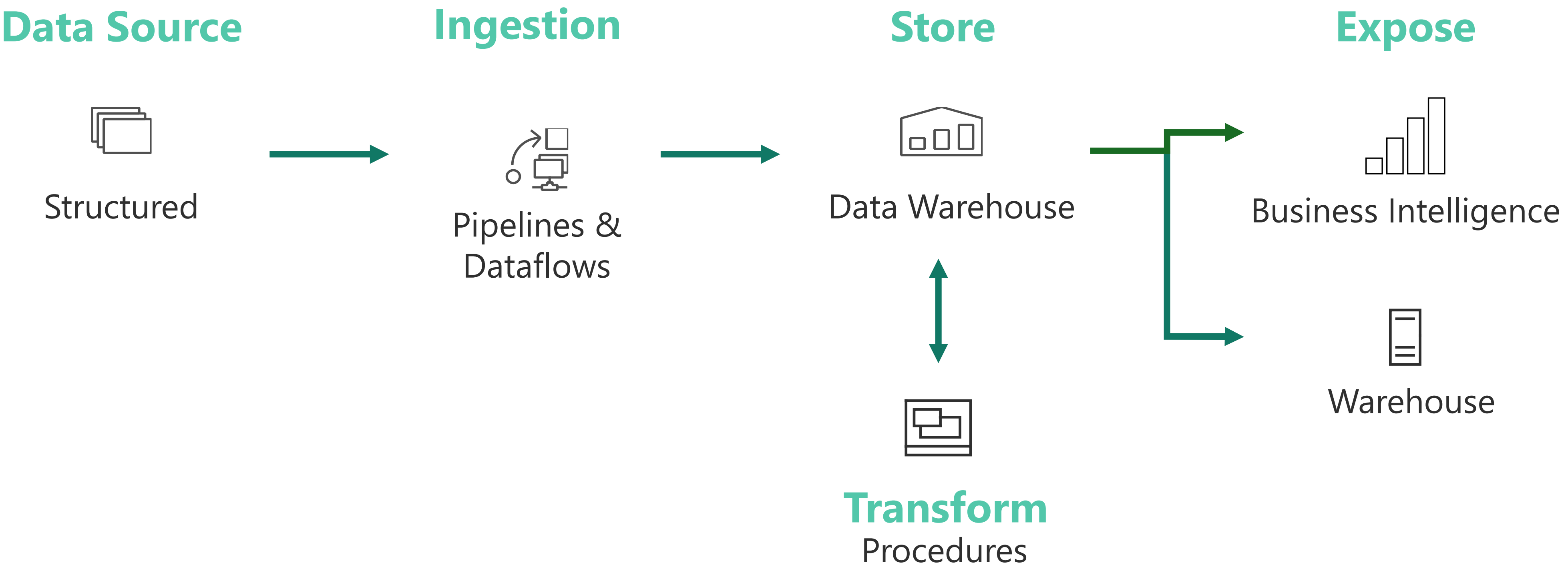
Expose

Business Intelligence

SQL endpoint

Data Warehouse scenario

End-to-end analytics scenario


Data Source

Structured

Ingestion

Pipelines & Dataflows

Store

Data Warehouse

Transform

Procedures

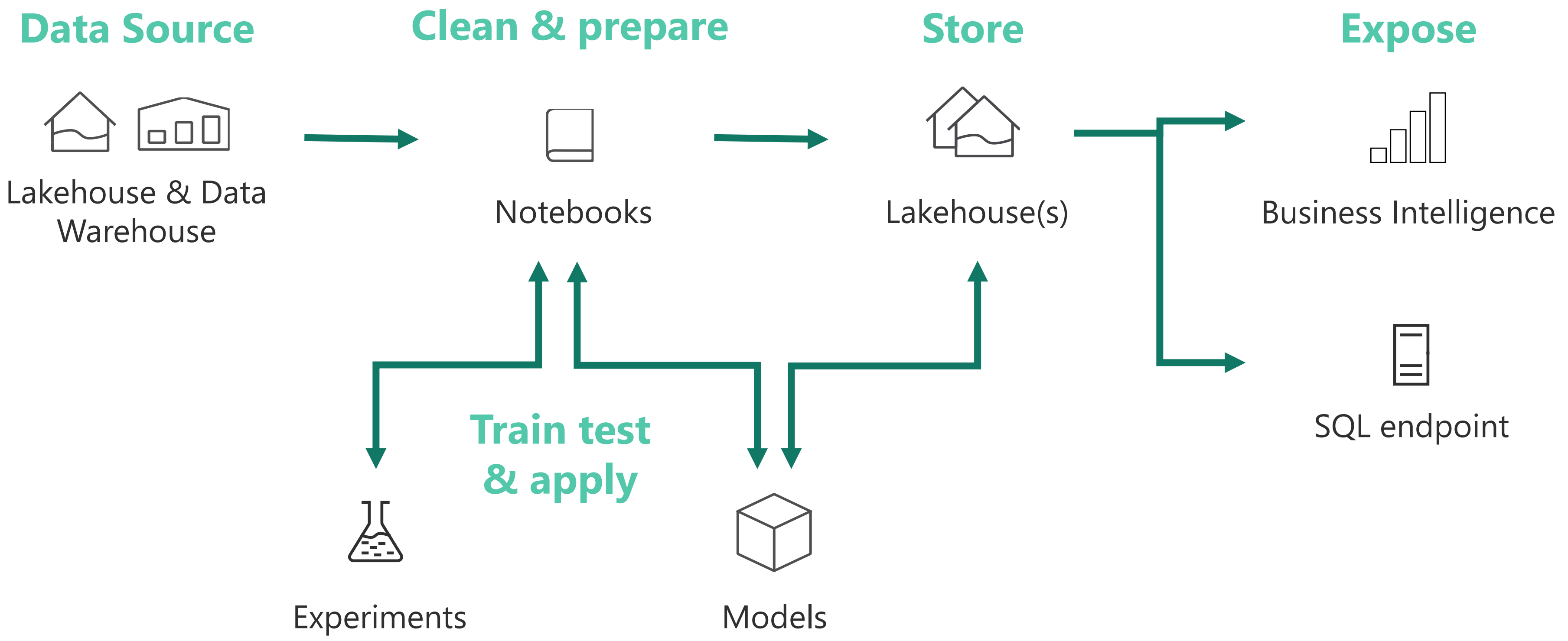
Expose

Business Intelligence

Warehouse

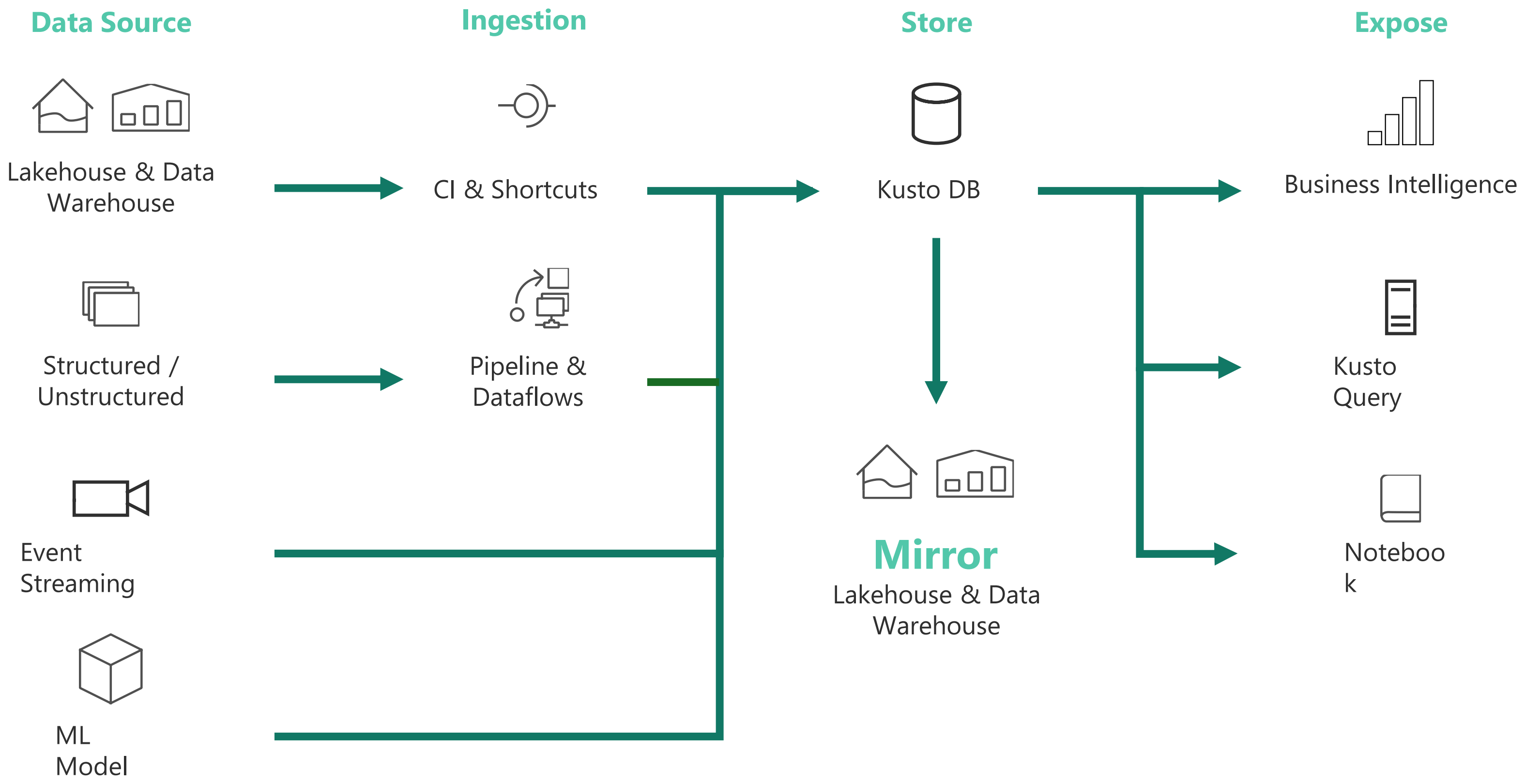
Data Science scenario

End-to-end analytics scenario



Real-time scenario

End-to-end analytics scenario

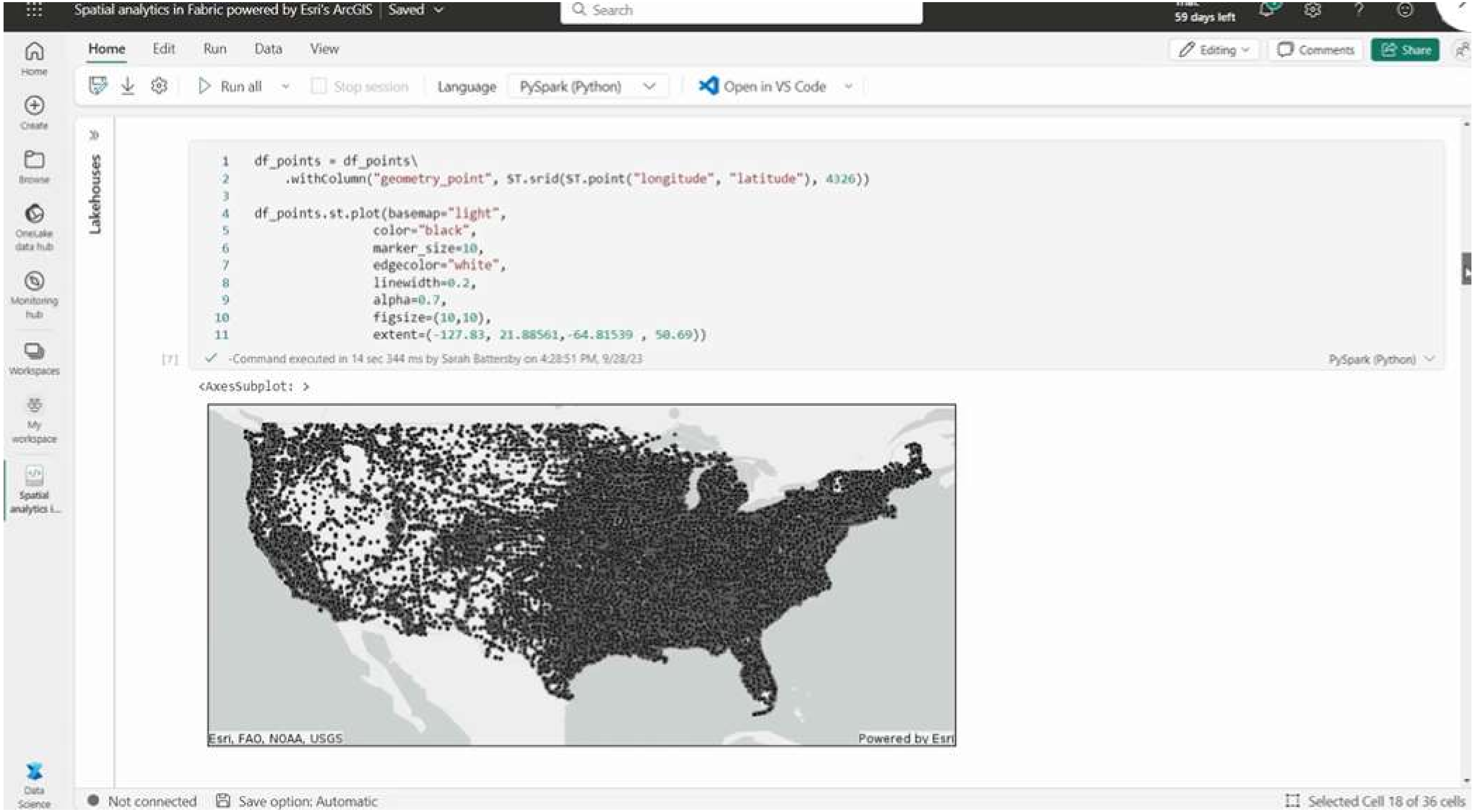


Spatial Analytics

Spatial Analytics

Esri announces integration with Microsoft to deliver leading-edge spatial analytics, allowing Microsoft Fabric customers valuable geospatial insights

Spatial Analytics



```
df_points = df_points\
    .withColumn("geometry_point", ST.srid(ST.point("longitude", "latitude"), 4326))

df_points.st.plot(basemap="light",
                  color="black",
                  marker_size=10,
                  edgecolor="white",
                  linewidth=0.2,
                  alpha=0.7,
                  figsize=(10,10),
                  extent=(-127.83, 21.88561,-64.81539 , 50.69))
```

[7] ✓ -Command executed in 14 sec 344 ms by Sarah Battersby on 4:28:51 PM, 9/28/23

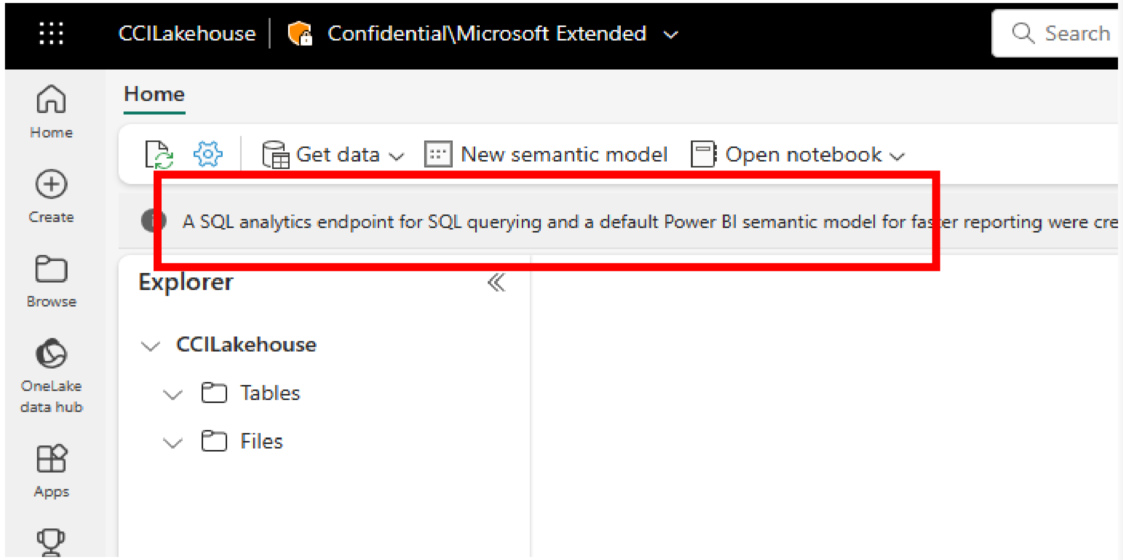
```
<AxesSubplot: >
```

● Not connected

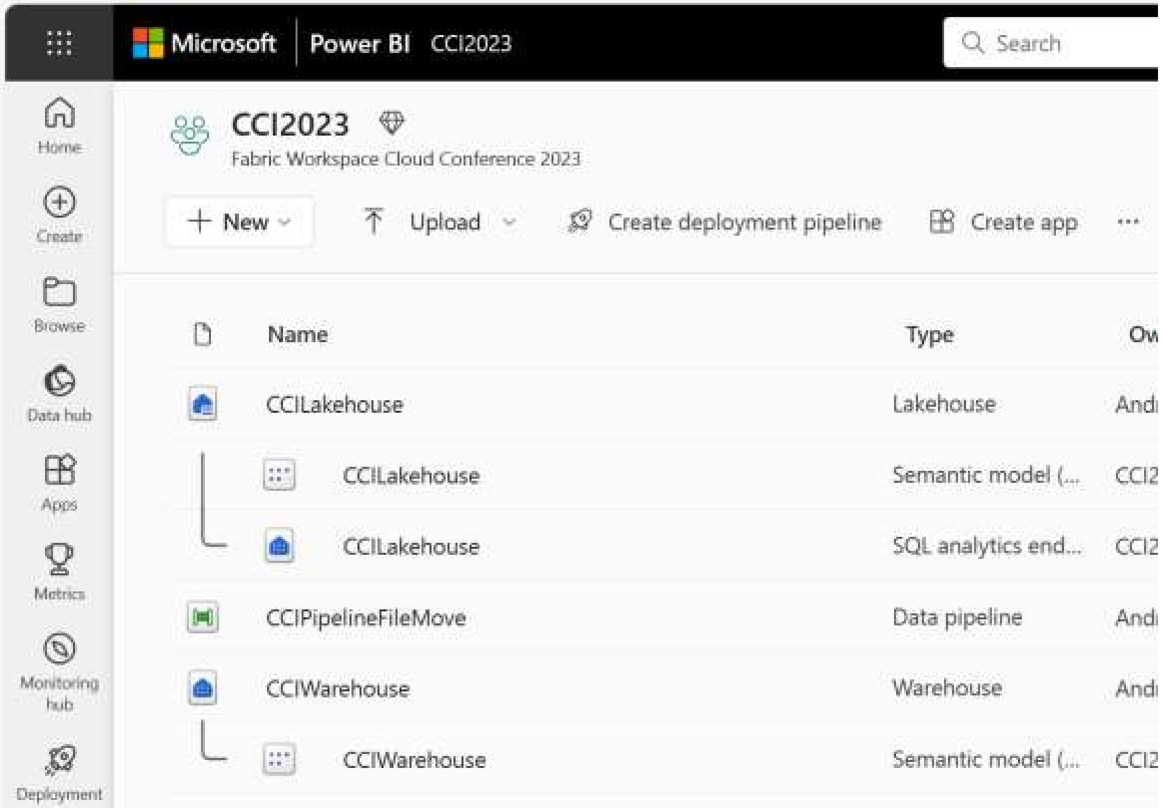
Save option: Automatic

Selected Cell 18 of 36 cells

Tools



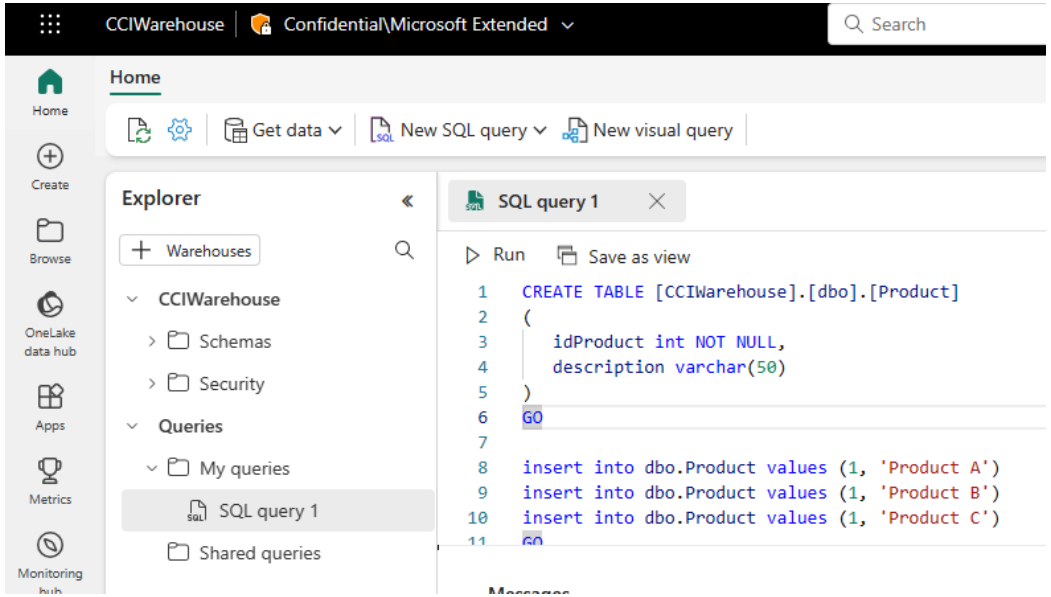
A SQL analytics endpoint for SQL querying and a default Power BI semantic model for faster reporting were cre



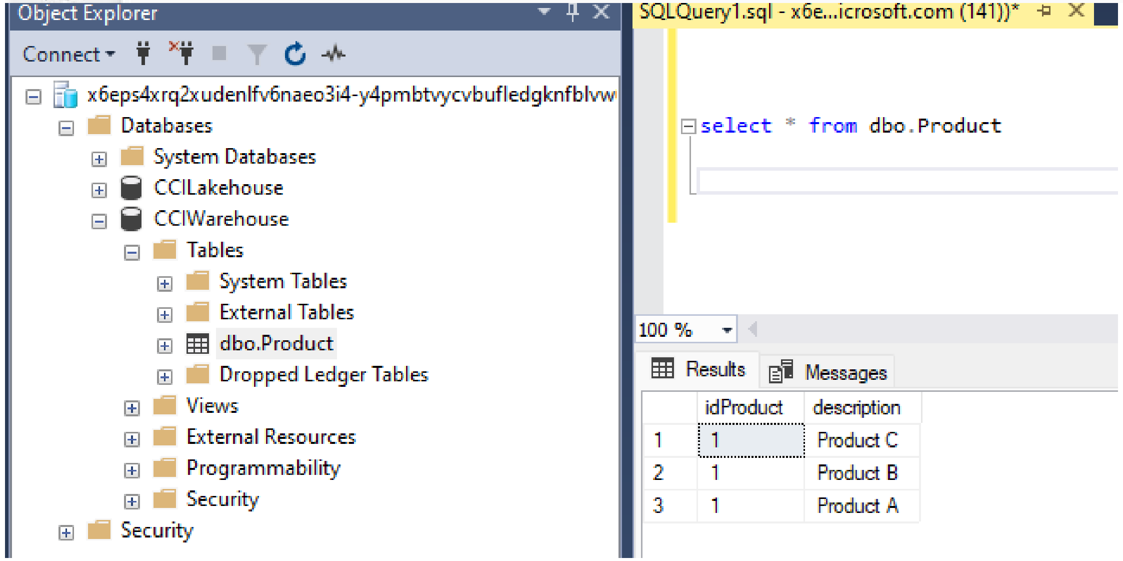
CCI2023

Fabric Workspace Cloud Conference 2023

Name	Type	Ow
CCILakehouse	Lakehouse	Andı
CCILakehouse	Semantic model (...	CCI2
CCILakehouse	SQL analytics end...	CCI2
CCIPipelineFileMove	Data pipeline	Andı
CCIWarehouse	Warehouse	Andı
CCIWarehouse	Semantic model (...	CCI2



```
CREATE TABLE [CCIWarehouse].[dbo].[Product]
(
    idProduct int NOT NULL,
    description varchar(50)
)
GO

insert into dbo.Product values (1, 'Product A')
insert into dbo.Product values (1, 'Product B')
insert into dbo.Product values (1, 'Product C')
GO
```



x6eps4xrq2xudenlfv6naeo3i4-y4pmbtvycvbufledgknfblvw
Databases
System Databases
CCILakehouse
CCIWarehouse
Tables
System Tables
External Tables
dbo.Product
Dropped Ledger Tables
Views
External Resources
Programmability
Security
Security

```
select * from dbo.Product
```

	idProduct	description
1	1	Product C
2	1	Product B
3	1	Product A

Lakehouse / Warehouse

- Structured, unstructured, semi-structured data across all ranges of file formats
- Data Engineer (Spark, Scala)
- Data Scientist (Python)
- Bronze, Silver, Gold layer

- DirectQuery, Import, new DirectLake (!) connection from Power BI Semantic Models

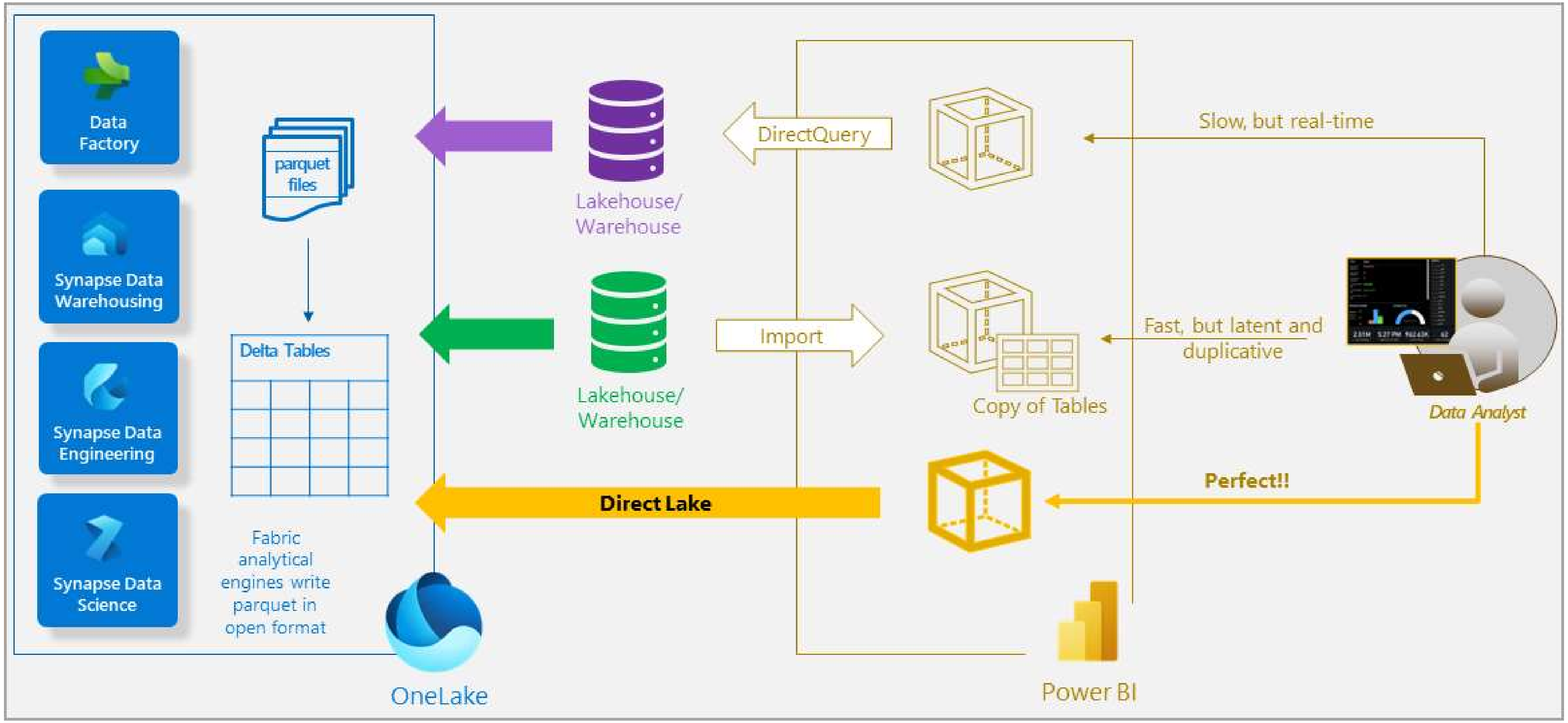
- Structured data in tables

- Data Warehouse Developer (SQL)

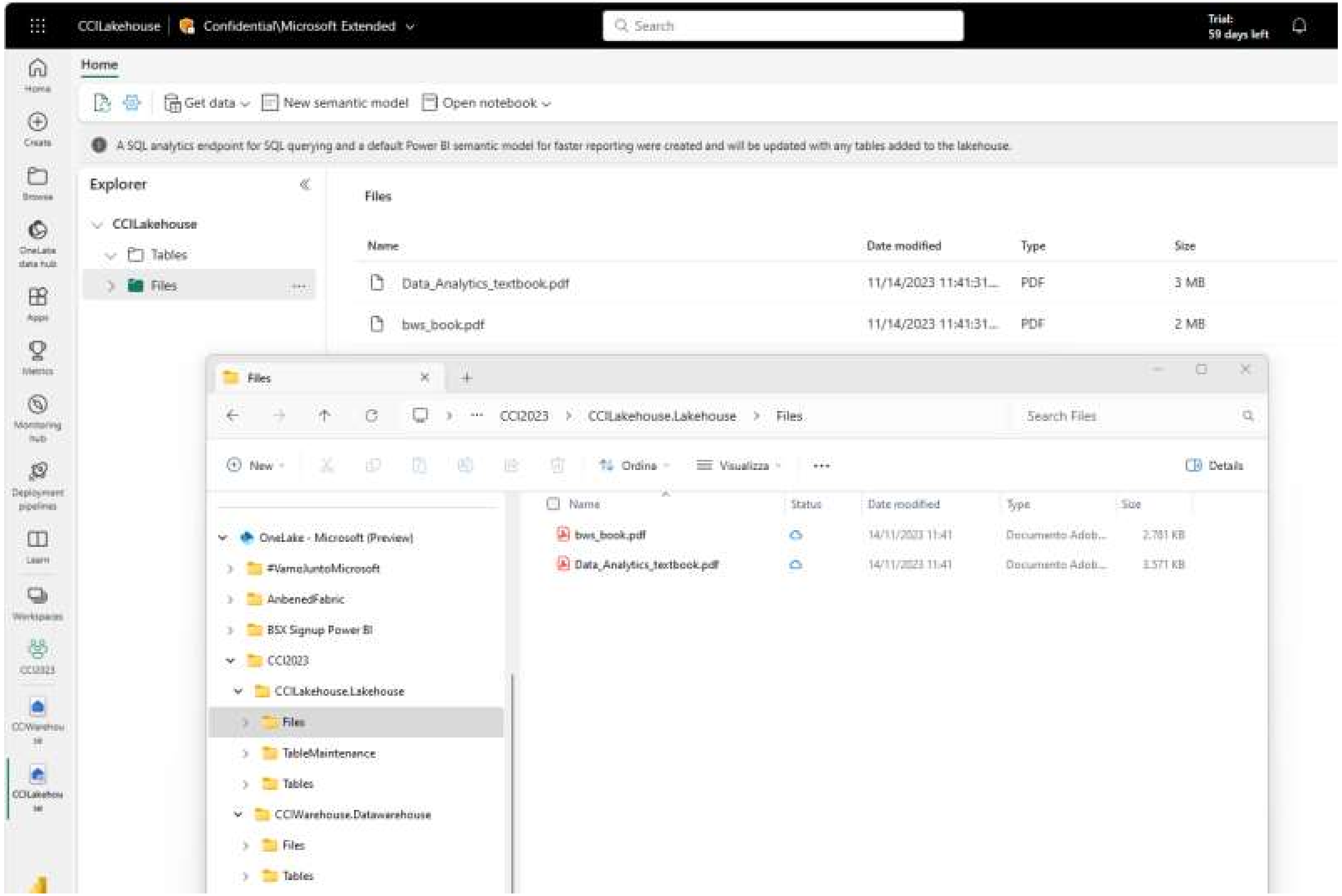
- Gold layer (can be used)

- DirectQuery, Import, DirectLake connection from Power BI Semantic Models

DirectLake


Data Factory
Synapse Data Warehousing
Synapse Data Engineering
Synapse Data Science
parquet files
Delta Tables
Fabric analytical engines write parquet in open format
OneLake
Lakehouse/ Warehouse
Lakehouse/ Warehouse
DirectQuery
Import
Direct Lake
Copy of Tables
Slow, but real-time
Fast, but latent and duplicative
Perfect!!
Data Analyst
Power BI

Lakehouse



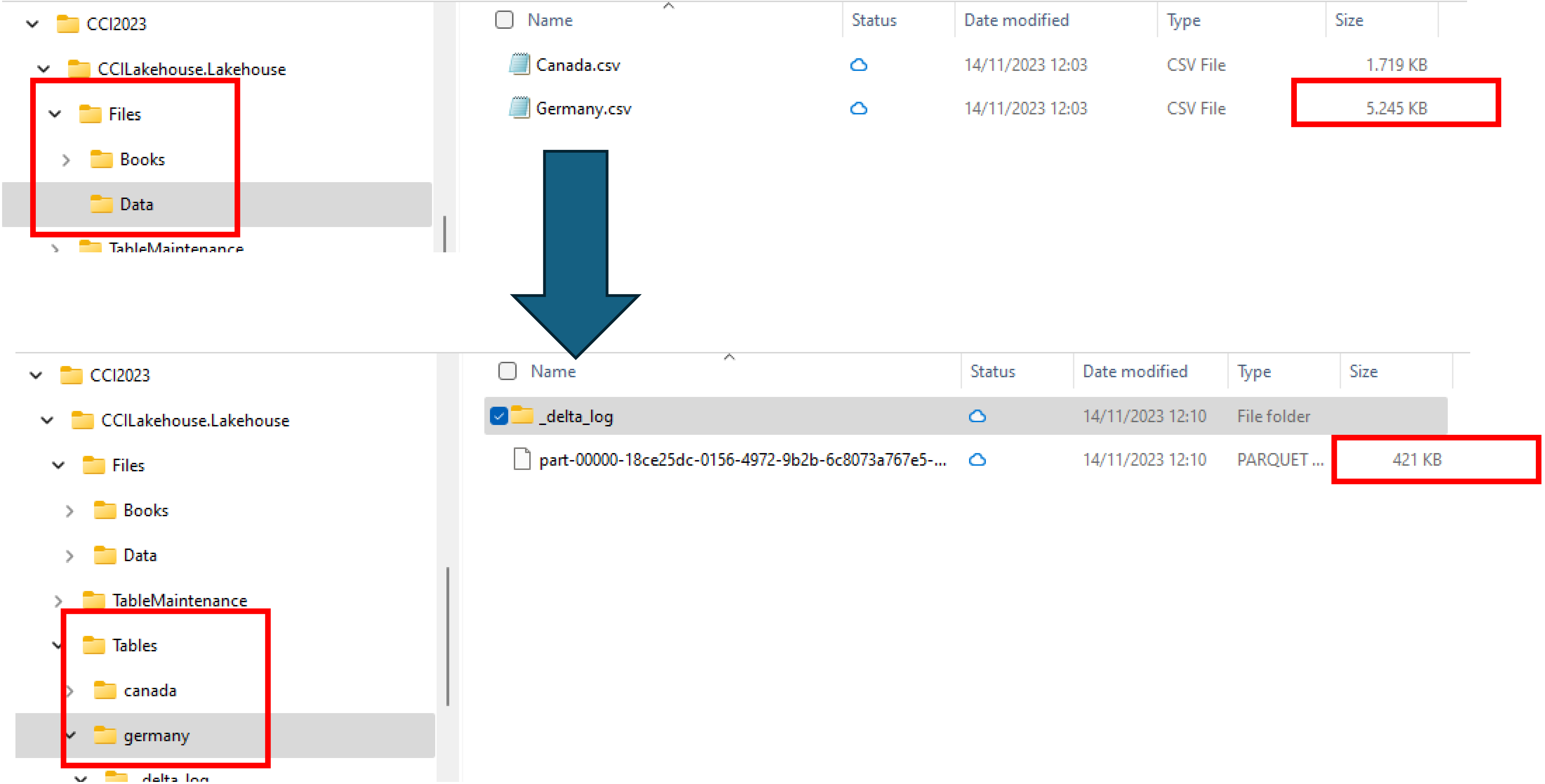
A SQL analytics endpoint for SQL querying and a default Power BI semantic model for faster reporting were created and will be updated with any tables added to the lakehouse.

Files

Name	Date modified	Type	Size
Data_Analytics_textbook.pdf	11/14/2023 11:41:31...	PDF	3 MB
bws_book.pdf	11/14/2023 11:41:31...	PDF	2 MB

Name	Status	Date modified	Type	Size
bws_book.pdf	☁	14/11/2023 11:41	Documento Adob...	2.781 KB
Data_Analytics_textbook.pdf	☁	14/11/2023 11:41	Documento Adob...	3.571 KB

Lakehouse



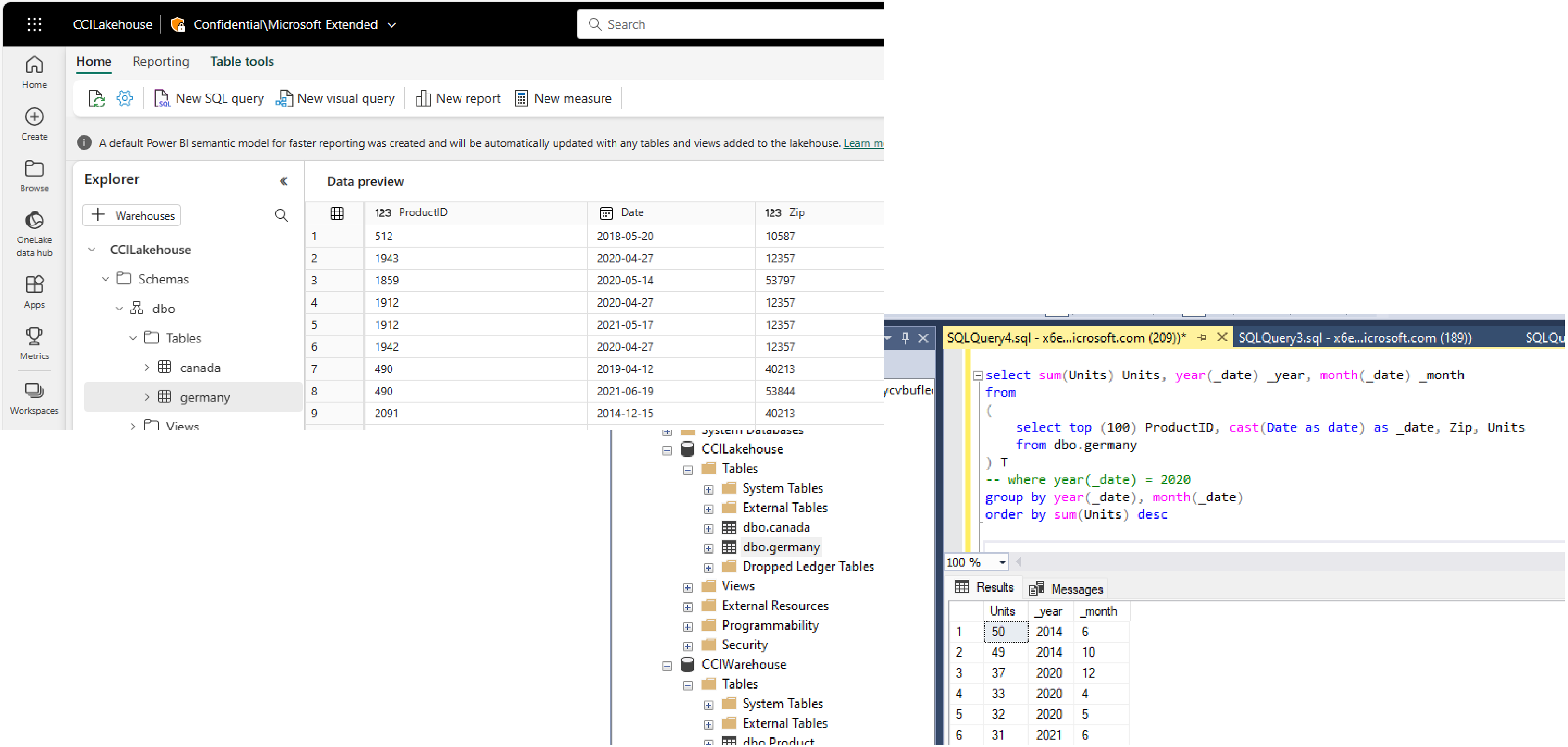
- CCI2023
  - CCILakehouse.Lakehouse
    - Files
      - Books
      - Data
    - TableMaintenance

□ Name	Status	Date modified	Type	Size
Canada.csv		14/11/2023 12:03	CSV File	1.719 KB
Germany.csv		14/11/2023 12:03	CSV File	5.245 KB

- CCI2023
  - CCILakehouse.Lakehouse
    - Files
      - Books
      - Data
    - TableMaintenance
    - Tables
      - canada
      - germany
        - _delta_log

□ Name	Status	Date modified	Type	Size
☑ _delta_log		14/11/2023 12:10	File folder	
part-00000-18ce25dc-0156-4972-9b2b-6c8073a767e5-...		14/11/2023 12:10	PARQUET ...	421 KB

Lakehouse



A default Power BI semantic model for faster reporting was created and will be automatically updated with any tables and views added to the lakehouse. Learn m

Data preview

	ProductID	Date	Zip
1	512	2018-05-20	10587
2	1943	2020-04-27	12357
3	1859	2020-05-14	53797
4	1912	2020-04-27	12357
5	1912	2021-05-17	12357
6	1942	2020-04-27	12357
7	490	2019-04-12	40213
8	490	2021-06-19	53844
9	2091	2014-12-15	40213

```
select sum(Units) Units, year(_date) _year, month(_date) _month
from
(
    select top (100) ProductID, cast(Date as date) as _date, Zip, Units
    from dbo.germany
) T
-- where year(_date) = 2020
group by year(_date), month(_date)
order by sum(Units) desc
```

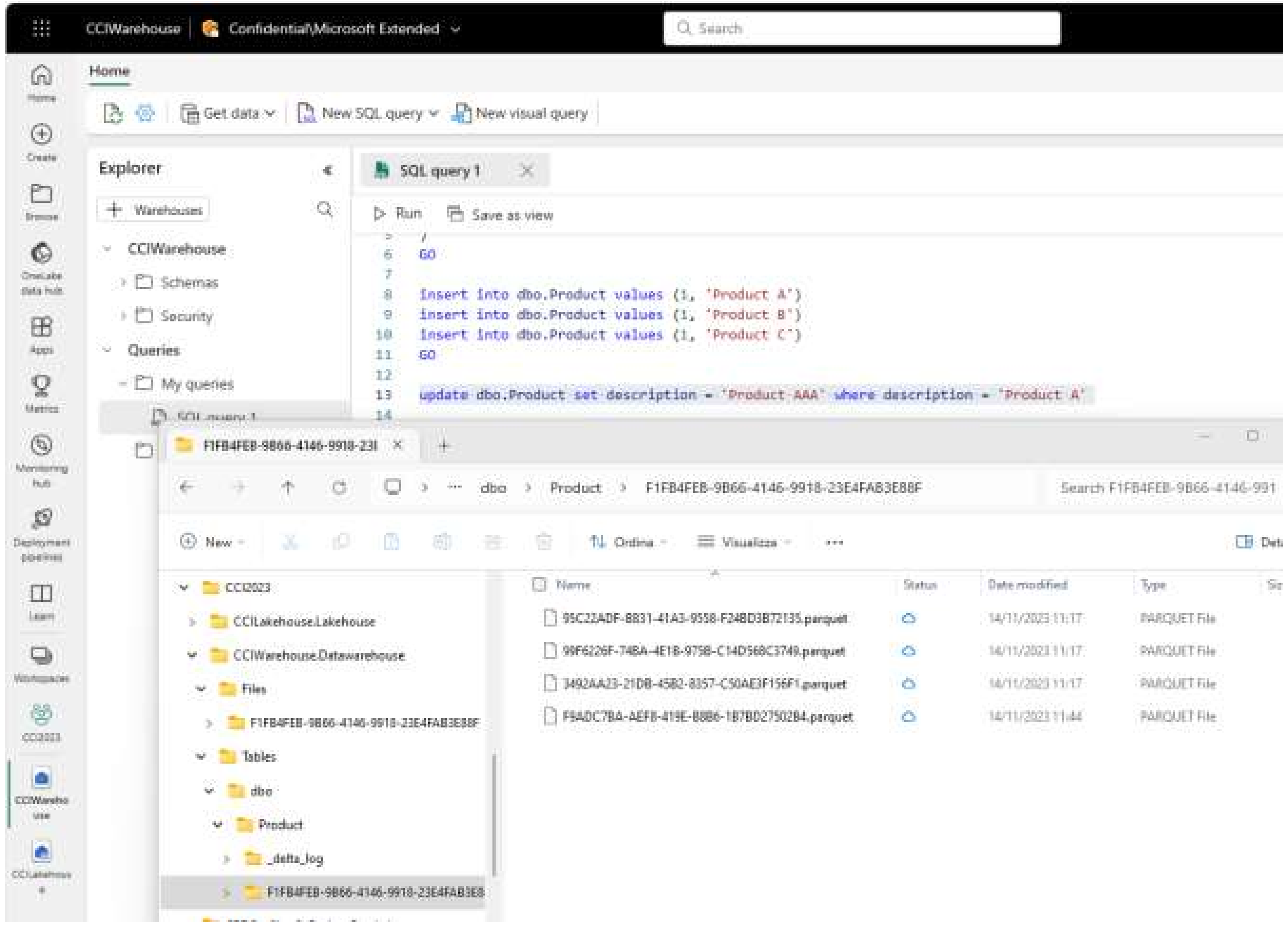
	Units	_year	_month
1	50	2014	6
2	49	2014	10
3	37	2020	12
4	33	2020	4
5	32	2020	5
6	31	2021	6

Warehouse

- Auto-scale out
- Fault tolerant
- Distributed compute engine

- Enhanced version of Serverless SQL Pool engine

- Native storage uses Delta Lake
  - Transaction Log
  - Metadata
  - Parquet file format



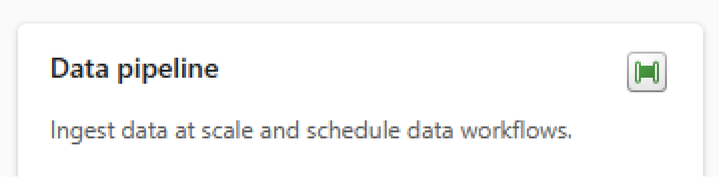
```
GO

insert into dbo.Product values (1, 'Product A')
insert into dbo.Product values (1, 'Product B')
insert into dbo.Product values (1, 'Product C')
GO

update dbo.Product set description = 'Product AAA' where description = 'Product A'
```

Name	Status	Date modified	Type
95C22ADF-8831-41A3-9558-F24BD3B72135.parquet	☁	14/11/2023 11:17	PARQUET File
99F6226F-74BA-4E18-9758-C14D568C3749.parquet	☁	14/11/2023 11:17	PARQUET File
3492AA23-21DB-45B2-8357-C50AE3F156F1.parquet	☁	14/11/2023 11:17	PARQUET File
F9ADC7BA-AEF8-419E-B8B6-1B7BD27502B4.parquet	☁	14/11/2023 11:44	PARQUET File

Loading

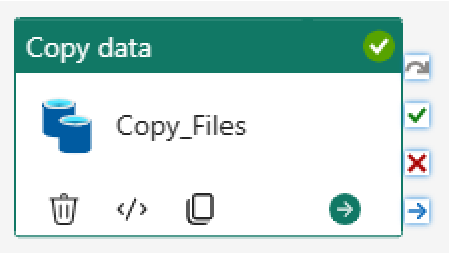
- SQL COPY INTO → from Azure Storage
- SQL CREATE TABLE AS … SELECT
- SQL INSERT INTO… VALUES / SELECT
- Pyodbc → from Notebooks

- Data Factory Data Pipeline

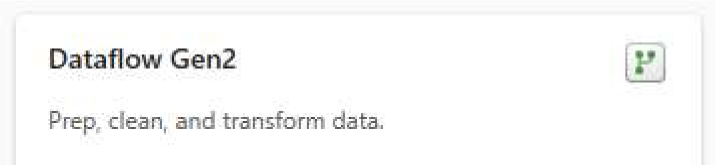


Data pipeline

Ingest data at scale and schedule data workflows.

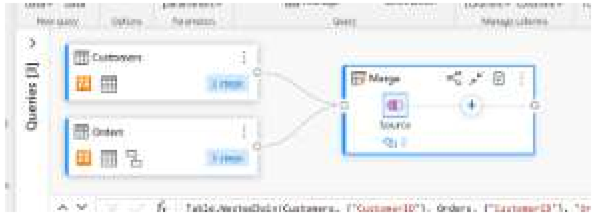


- Dataflows Gen2 (Power Query)



Dataflow Gen2

Prep, clean, and transform data.



```
Table.NestedJoin(Customers, {"CustomerID"}, Orders, {"CustomerID"}, "Or
```

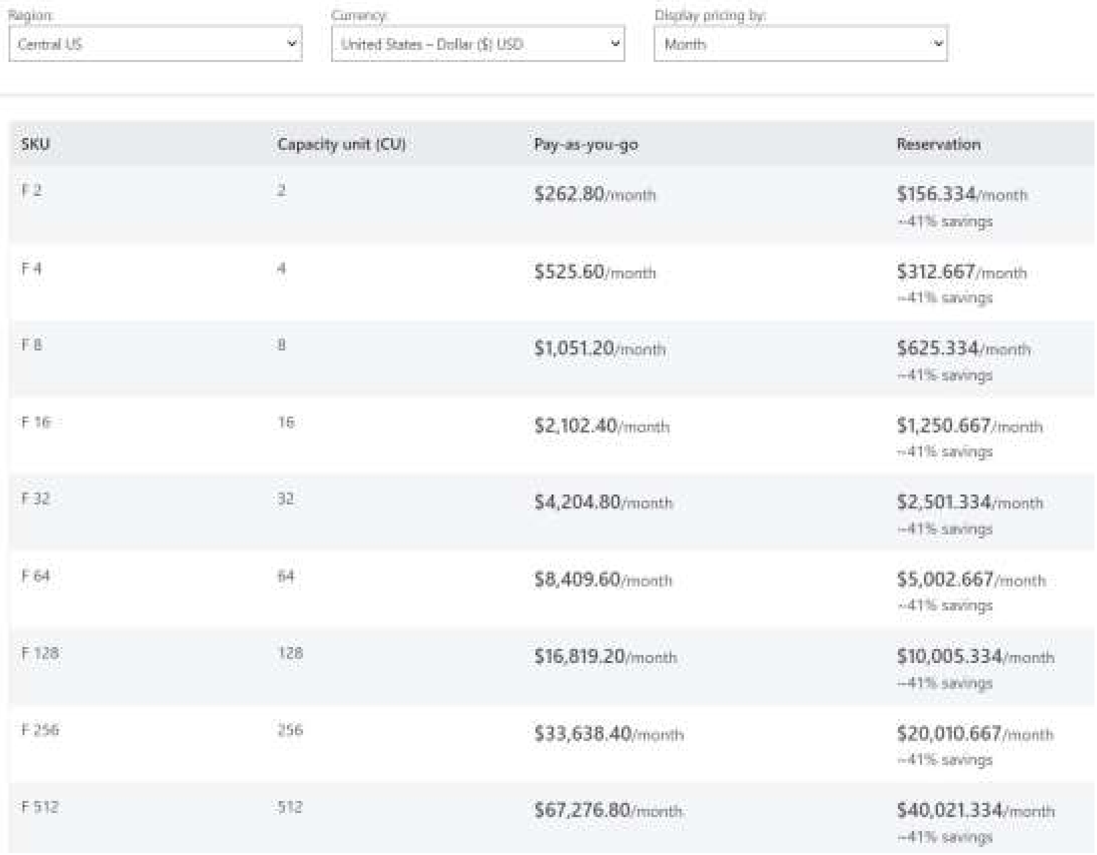
DEMO

If you are an existing Power BI Premium per capacity customer, you can already access Microsoft Fabric by simply turning on Fabric in your admin portal.

Economics

Microsoft Fabric Reserved Instances are here!

F2 reserved capacity that comes with all the workloads, all running at blazing speed, costs just $156/month.



Region: Central US

Currency: United States – Dollar ($) USD

Display pricing by: Month

SKU	Capacity unit (CU)	Pay-as-you-go	Reservation
F 2	2	$262.80/month	$156.334/month ~41% savings
F 4	4	$525.60/month	$312.667/month ~41% savings
F 8	8	$1,051.20/month	$625.334/month ~41% savings
F 16	16	$2,102.40/month	$1,250.667/month ~41% savings
F 32	32	$4,204.80/month	$2,501.334/month ~41% savings
F 64	64	$8,409.60/month	$5,002.667/month ~41% savings
F 128	128	$16,819.20/month	$10,005.334/month ~41% savings
F 256	256	$33,638.40/month	$20,010.667/month ~41% savings
F 512	512	$67,276.80/month	$40,021.334/month ~41% savings

Reservations

Reservation pricing → pre-commit to Fabric Capacity Units in one-year increments

- save up to 40.5 percent over the pay-as-you-go prices (excluding Power BI Capacity SKUs)
- You can pay for all your consumption, across every Fabric workload, with a single, unified bill and use the Capacity Metrics app to understand and track your usage

- You can choose to pay monthly or upfront
- You can increase the size of Fabric capacity reservations by purchasing additional reservations or by exchanging them for new ones
- You can exchange existing Azure Synapse Analytics Dedicated SQL pool (formerly SQL DW) reservations to Fabric capacity reservations
- Reservation for Fabric capacity doesn't cover storage or networking charges associated with the Microsoft Fabric usage; it only covers Fabric capacity usage

Notes

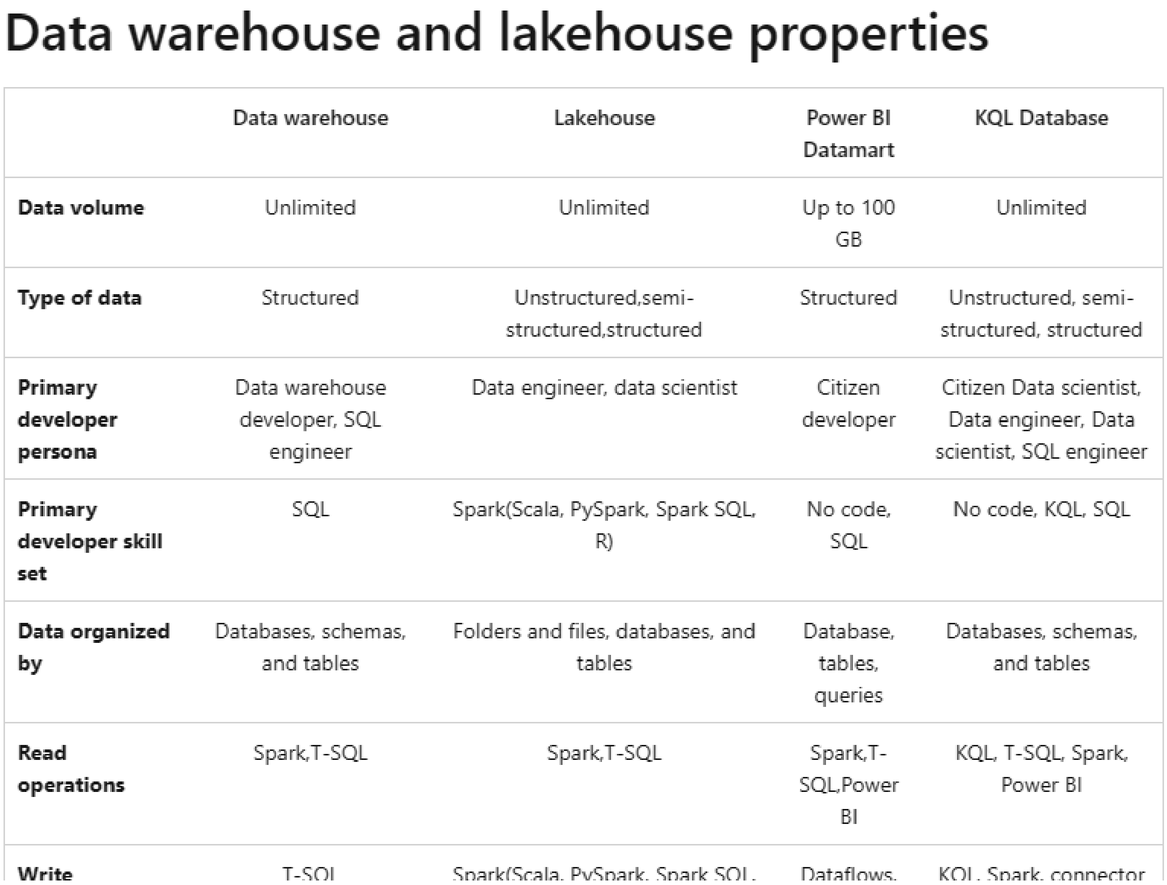
Regulatory compliance

Data residency

- Fabric will be available in every Azure region
- Data at rest: compliant with EUDB and other single-geo data residency regulations
- Multi-geo capacities allow control over content storage location in most Azure data centers world-wide

Microsoft pledges support for EU Data Boundary

Choose a data store

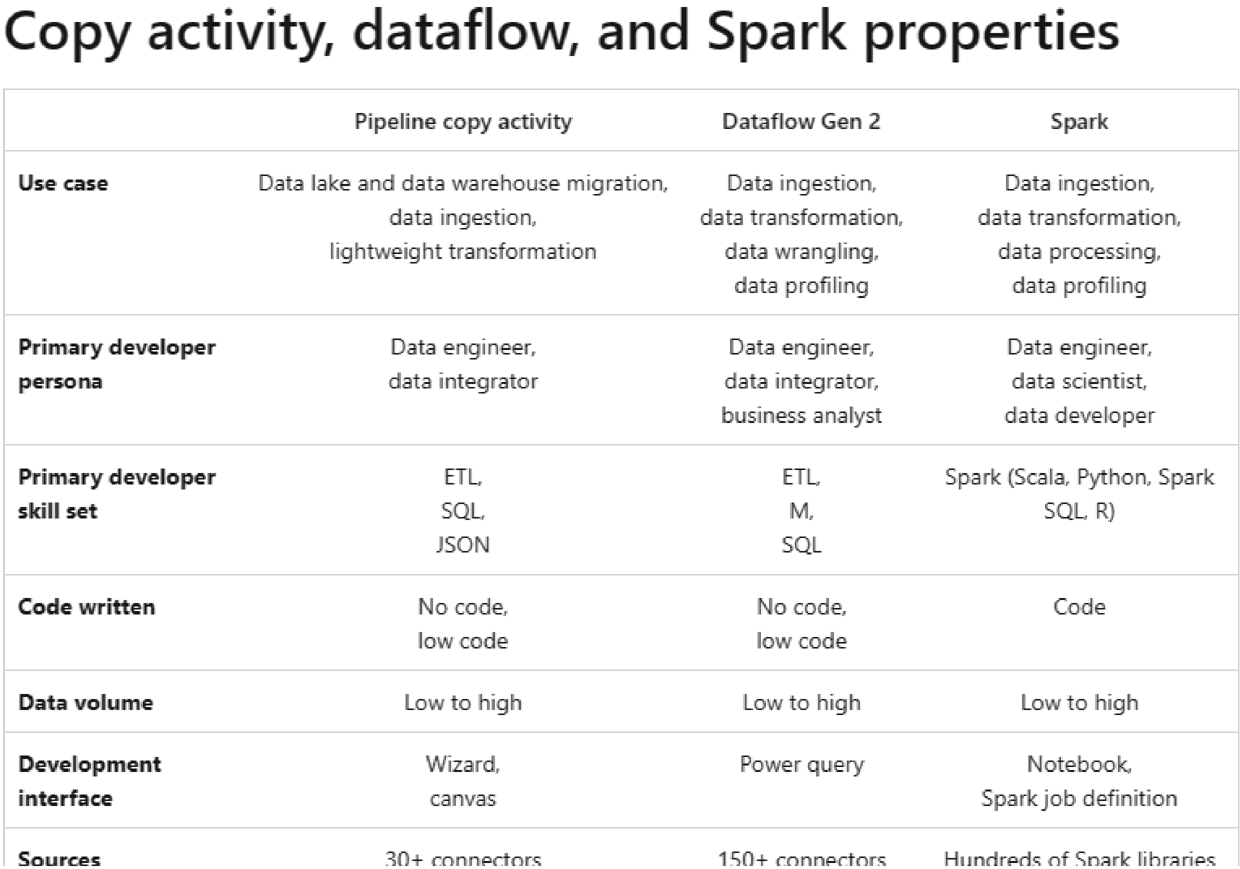


Data warehouse and lakehouse properties

	Data warehouse	Lakehouse	Power BI Datamart	KQL Database
Data volume	Unlimited	Unlimited	Up to 100 GB	Unlimited
Type of data	Structured	Unstructured,semi-structured,structured	Structured	Unstructured, semi-structured, structured
Primary developer persona	Data warehouse developer, SQL engineer	Data engineer, data scientist	Citizen developer	Citizen Data scientist, Data engineer, Data scientist, SQL engineer
Primary developer skill set	SQL	Spark(Scala, PySpark, Spark SQL, R)	No code, SQL	No code, KQL, SQL
Data organized by	Databases, schemas, and tables	Folders and files, databases, and tables	Database, tables, queries	Databases, schemas, and tables
Read operations	Spark,T-SQL	Spark,T-SQL	Spark,T-SQL,Power BI	KQL, T-SQL, Spark, Power BI
Write	T-SQL	Spark(Scala, PySpark, Spark SQL,	Dataflows,	KQL, Spark, connector

https://learn.microsoft.com/en-us/fabric/get-started/decision-guide-data-store

Copy activity, dataflow, or Spark



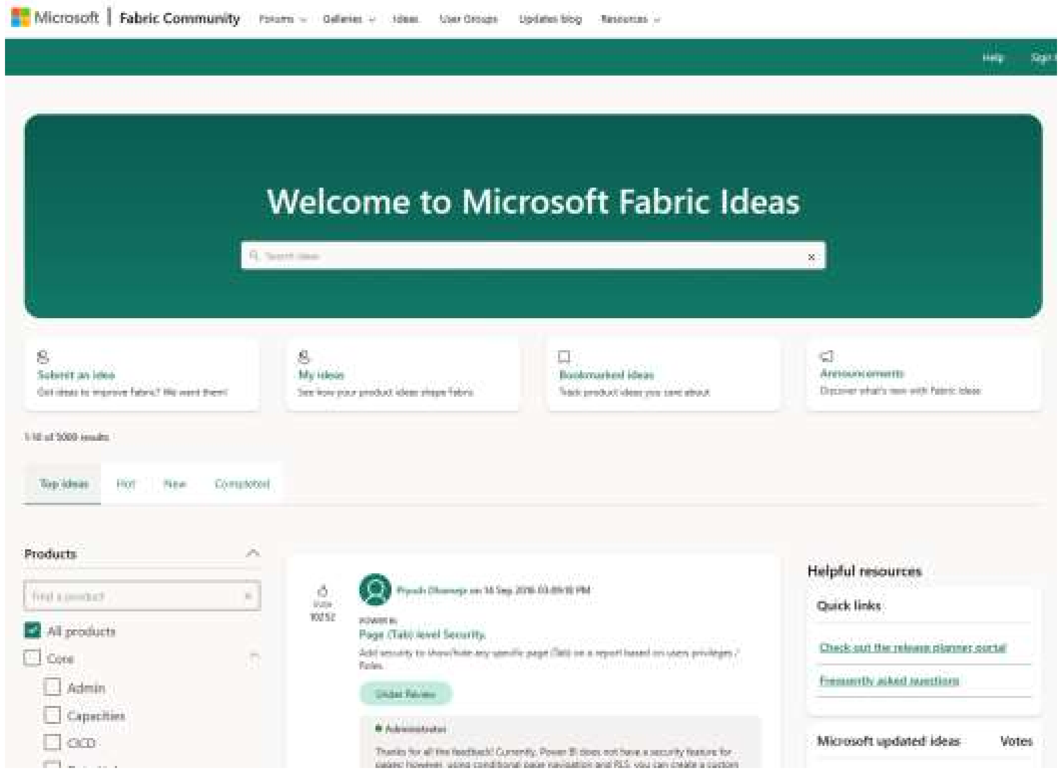
Copy activity, dataflow, and Spark properties

	Pipeline copy activity	Dataflow Gen 2	Spark
Use case	Data lake and data warehouse migration, data ingestion, lightweight transformation	Data ingestion, data transformation, data wrangling, data profiling	Data ingestion, data transformation, data processing, data profiling
Primary developer persona	Data engineer, data integrator	Data engineer, data integrator, business analyst	Data engineer, data scientist, data developer
Primary developer skill set	ETL, SQL, JSON	ETL, M, SQL	Spark (Scala, Python, Spark SQL, R)
Code written	No code, low code	No code, low code	Code
Data volume	Low to high	Low to high	Low to high
Development interface	Wizard, canvas	Power query	Notebook, Spark job definition
Sources	30+ connectors	150+ connectors	Hundreds of Spark libraries

https://learn.microsoft.com/en-us/fabric/get-started/decision-guide-pipeline-dataflow-spark

Fabric Community

https://ideas.fabric.microsoft.com/



Welcome to Microsoft Fabric Ideas

Submit an idea
Got ideas to improve Fabric? We want them!

My ideas
See how your product ideas shape Fabric

Bookmarked ideas
Track product ideas you care about

Announcements
Discover what's new with Fabric Ideas

1-10 of 5000 results

Products

☑ All products
□ Core
□ Admin
□ Capacities
□ CICD

Piyush Dhameja on 14 Sep 2016 03:09:16 PM

Vote 10252

POWER BI

Page (Tab) level Security.

Add security to show/hide any specific page (Tab) on a report based on users privileges / Roles.

Administrator

Thanks for all the feedback! Currently, Power BI does not have a security feature for pages; however, using conditional page navigation and RLS, you can create a custom

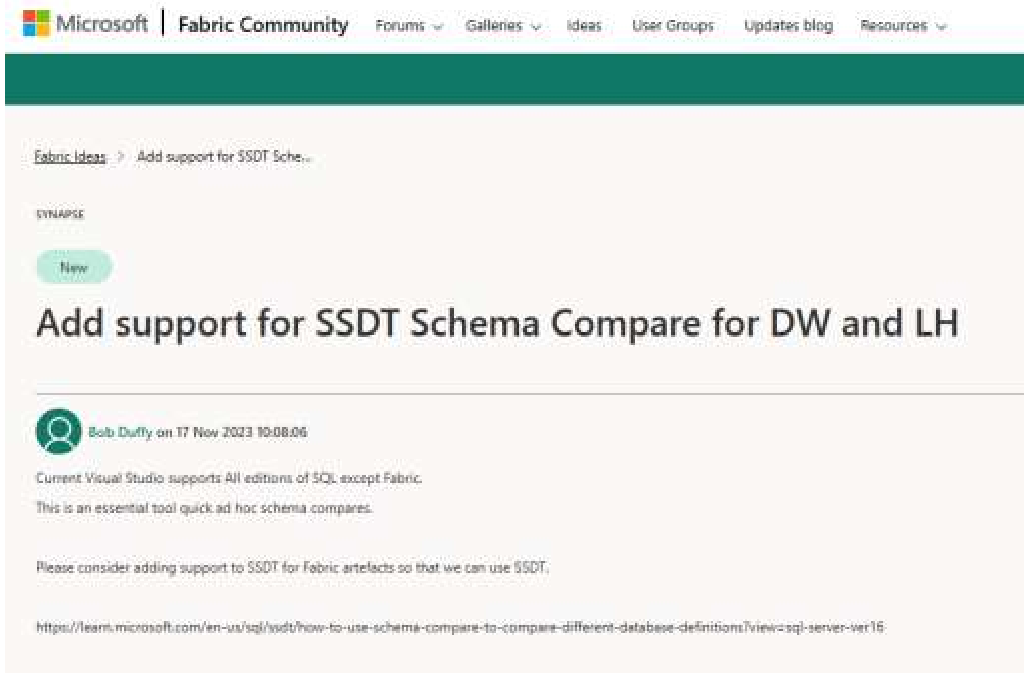
Helpful resources

Quick links

Check out the release planner portal

Frequently asked questions

Microsoft updated ideas | Votes



Fabric Ideas > Add support for SSDT Sche...

SYNAPSE

Add support for SSDT Schema Compare for DW and LH

Bob Duffy on 17 Nov 2023 10:08:06

Current Visual Studio supports All editions of SQL except Fabric.

This is an essential tool quick ad hoc schema compares.

Please consider adding support to SSDT for Fabric artefacts so that we can use SSDT.

https://learn.microsoft.com/en-us/sql/ssdt/how-to-use-schema-compare-to-compare-different-database-definitions?view=sql-server-ver16

Learn

An official **Microsoft certification** for implementing analytics solutions with Microsoft Fabric and Power BI

Microsoft
EXAM
EXAMS
Exam DP-600: Implementing Analytics Solutions Using Microsoft Fabric (beta)

Exam DP-600: Implementing Analytics Solutions Using Microsoft Fabric (beta) - Certifications | Microsoft Learn

Microsoft Fabric Community Resources

- Community Call to Action

✓ Try Microsoft Fabric for free: https://aka.ms/try-fabric

✓ Join the Fabric community: https://aka.ms/fabriccommunity

✓ Share and vote for ideas to improve Fabric: https://aka.ms/fabricideas

✓ Read and comment our blog: https://aka.ms/fabricblog

Learn More about Microsoft Fabric

- Product announcement: https://aka.ms/fabric
- Digital Event at Build (videos): https://aka.ms/build-with-analytics
- Product website: https://aka.ms/microsoft-fabric
- Documentation: https://aka.ms/fabric-docs
- Fabric e-book: https://aka.ms/fabric-get-started-ebook
- Microsoft Learn: https://aka.ms/learn-fabric
- End-to-end scenario tutorials: https://aka.ms/fabric-tutorials
- Fabric Notes: https://aka.ms/fabric-notes

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

GRAZIE!

SPEAKER

Andrea Benedetti

Sr Cloud Architect, Microsoft

in /in/abenedetti

@anBenedetti

https://github.com/anbened