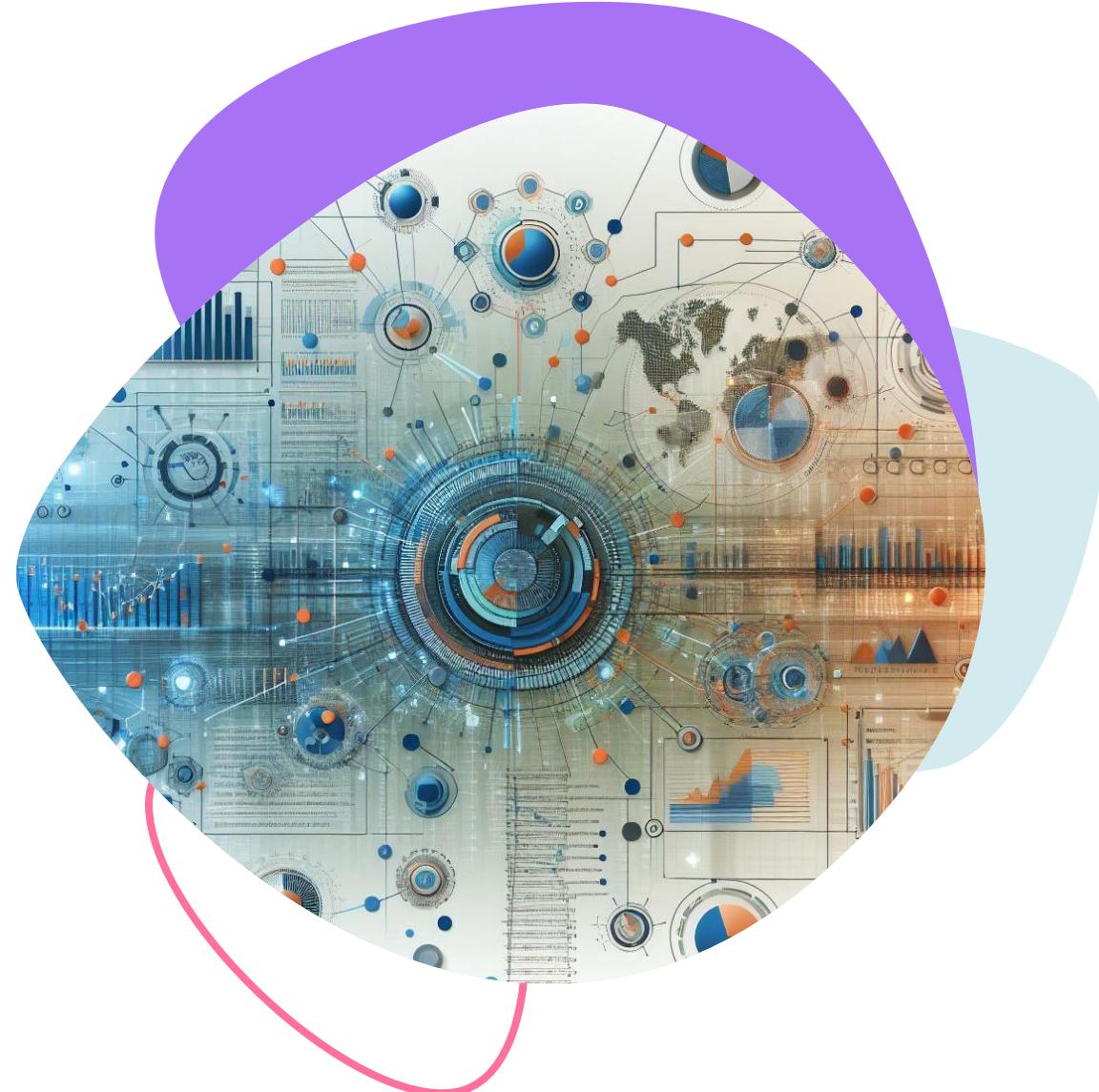




A Day in a Microsoft Fabric

Nikola Ilic

Data Mozart, Microsoft Data Platform MVP



Thank you!



Many thanks to our sponsors, without whom such an event would not be possible.





Agenda

- Overview
- Fabric UI and terminology
- Core components
- Parquet/Delta file formats
- Fabric for Power BI professionals
- Fabric experiences

Timeline

09:00 Introduction, expectations, overview

09:15 What is Microsoft Fabric and why do(n't) we need it?

09:30 Microsoft Fabric terminology and UI

09:50 Microsoft Fabric core components

10:50 Coffee break

11:00 Understanding Delta & Parquet file format

11:20 OneLake in Microsoft Fabric

11:50 Lakehouse or Warehouse: Which one is for me? (Part I)

12:30 Lunch break

Timeline (continued...)

- 13:30 Lakehouse or Warehouse: Which one is for me? (Part II)
- 13:50 Microsoft Fabric for Power BI professionals
- 14:50 Microsoft Fabric experiences (Data Engineering, Data Warehousing, RTA, Data Science) Part I
- 15:30 Coffee break
- 15:45 Microsoft Fabric experiences (Data Engineering, Data Warehousing, RTA, Data Science) Part II
- 16:05 Microsoft Fabric licensing options
- 16:25 Summary, Q&A and quiz



Nikola Ilic

Consultant & Trainer



www.data-mozart.com



@DataMozart

learn.data-mozart.com

- *I'm making music from the data!*
- Power BI and SQL addict, blogger, speaker...
- Father of 2, Barca & Leo Messi fan...

Setting the expectations

- Not a deep dive!
- No labs and hands-on
- Understand Fabric essentials
- Comfortable going deeper into specific areas/experiences



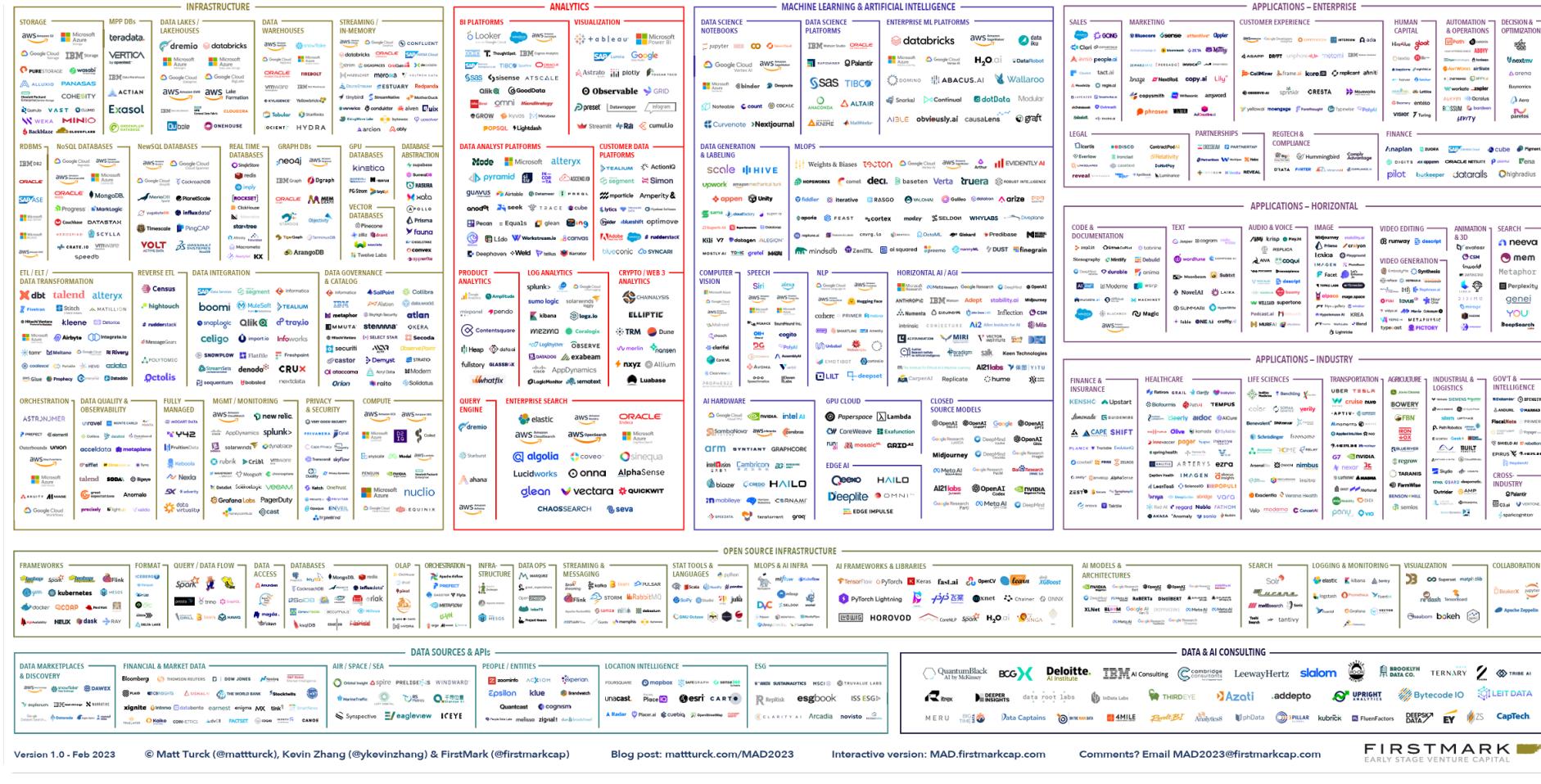


“

What is Microsoft
Fabric and why
do(n’t) we need it?

”

The 2023 ML, AI, and Data Landscape



Unification

“Unify,
I am the Chief Data Officer and don’t
want to be the Chief Integration Officer.”

How to achieve this **Unification** ?

Data Access
“Lake-centric
and open
architecture”

A complete
analytic
platform in
SaaS

Being more
productive with
AI & Copilot



Microsoft Fabric In 30 Seconds...

From

One component

Single database

Gen AI wired in

To

Unified stack

All the data

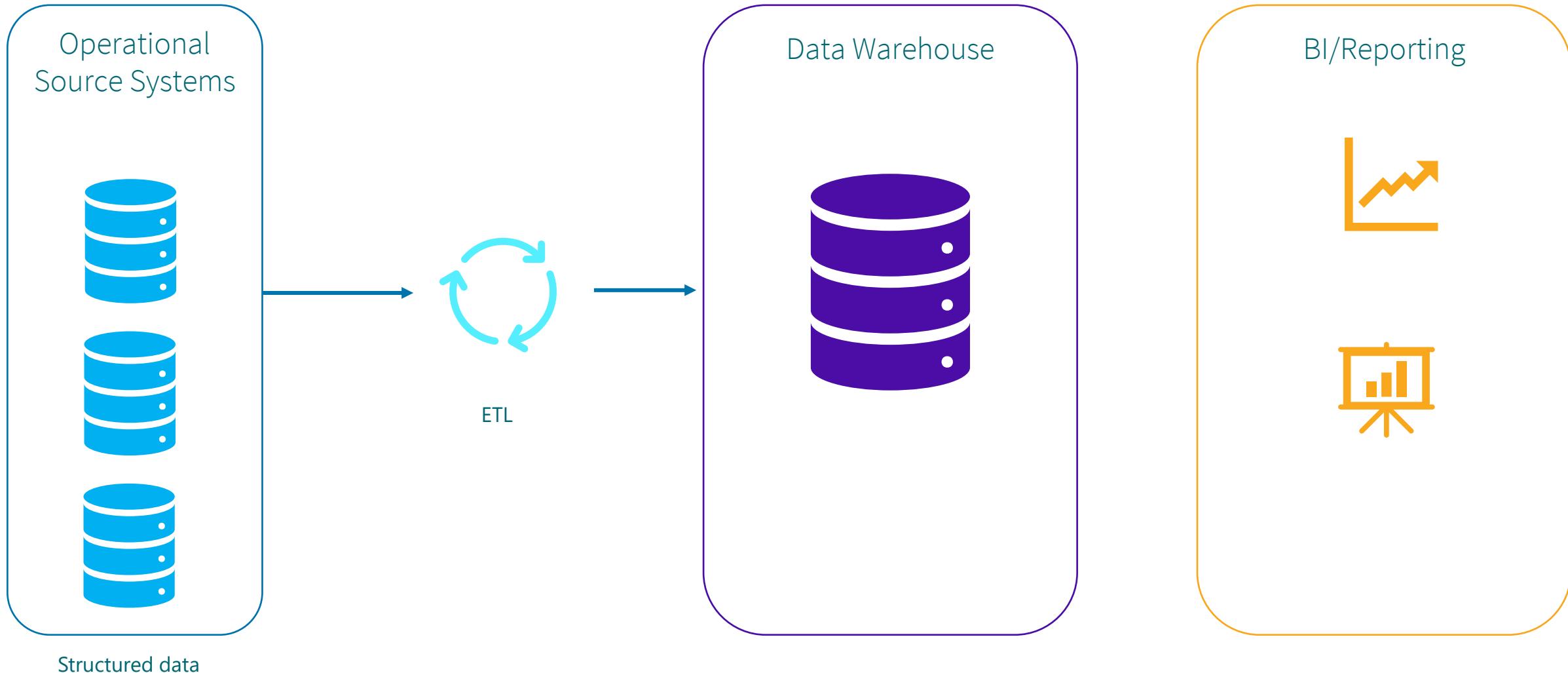
Gen AI built in



History Lessons



Traditional Data Warehousing

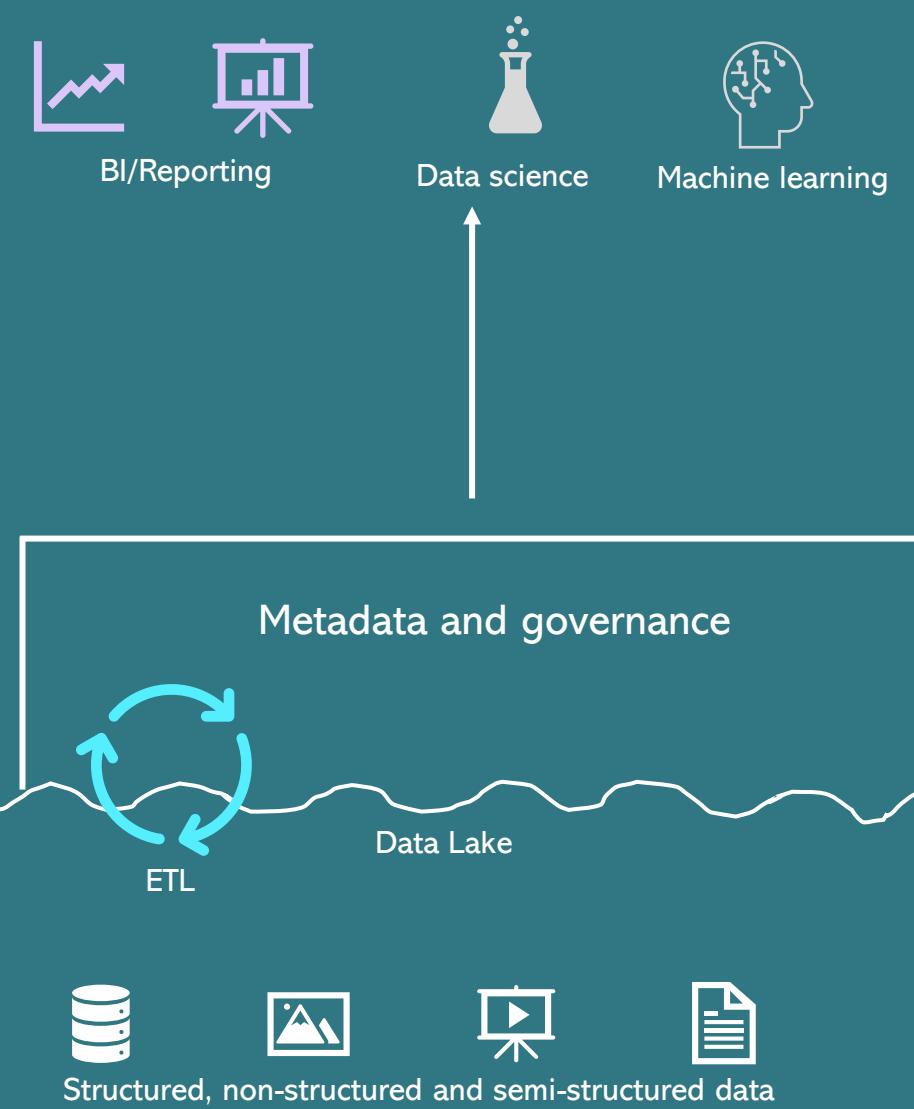
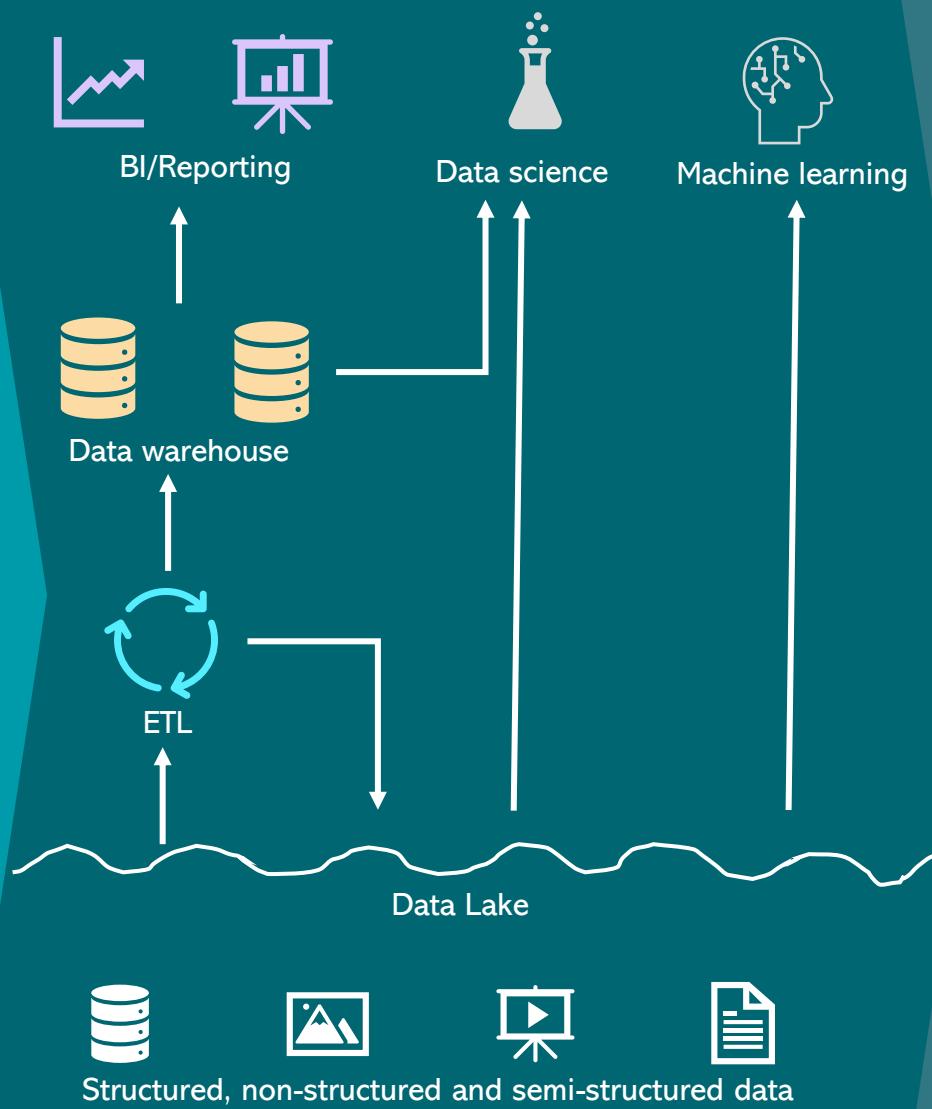


Traditional Data Warehousing



The background of the image is a soft-focus photograph of a mountain range. In the foreground, there is a large, calm body of water, possibly a lake or a wide river, reflecting the surrounding misty mountains. The overall atmosphere is serene and somewhat ethereal.

Modern Data Platform



What is a Data Lakehouse?



Flexibility

Cost-
efficiency

Scalability



Data
management

ACID
transactions



HUBADVISER

Data Lakehouse key features

Transaction support

Schema enforcement

BI support

Decoupled storage and compute

Open format support

Diverse data types

Diverse workloads

Streaming End-to-end

How to Choose the Right Product Architecture

Data platform users' skillset



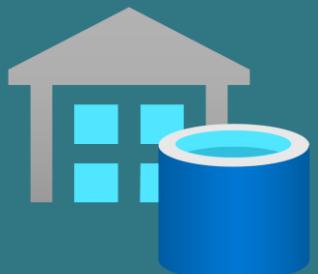
Analysts



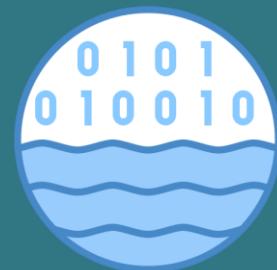
ML engineers/data scientists



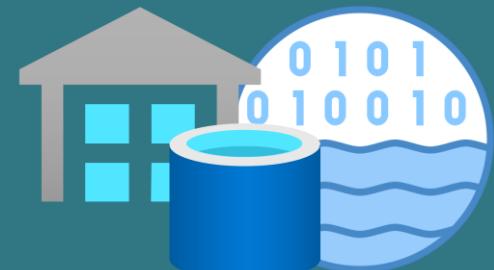
Analysts & ML engineers/data scientists



Data Warehouse



Data Lake



Data Lakehouse

How to Choose the Right Product

Small vs Large Organizations

How small is small?

Less than 300.000\$/Year

1M\$+/Year

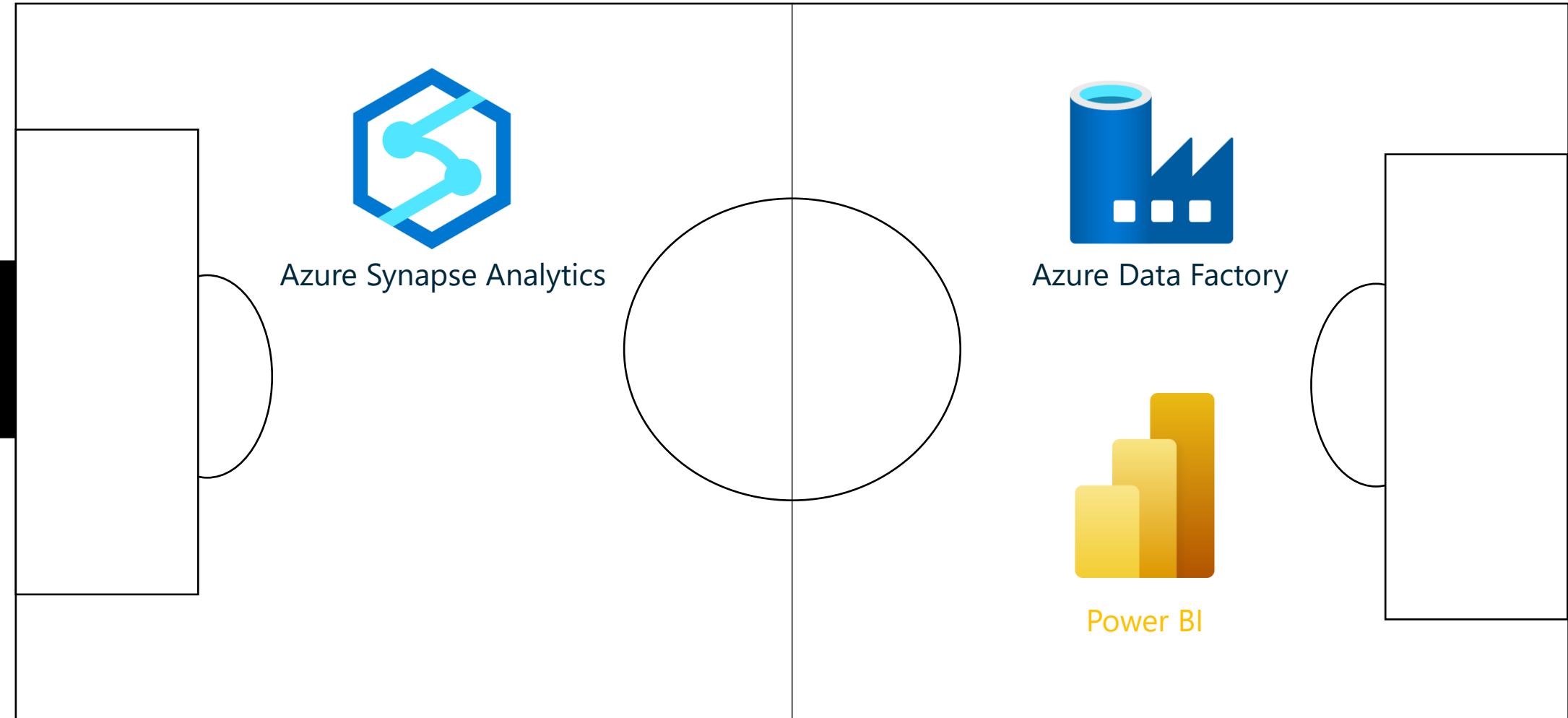


Single cloud provider

Decentralized vs standardized
data platform



What is a Microsoft Fabric?





“Players” in Microsoft Fabric



Why DON'T we need Microsoft Fabric?



Current solutions are adequate



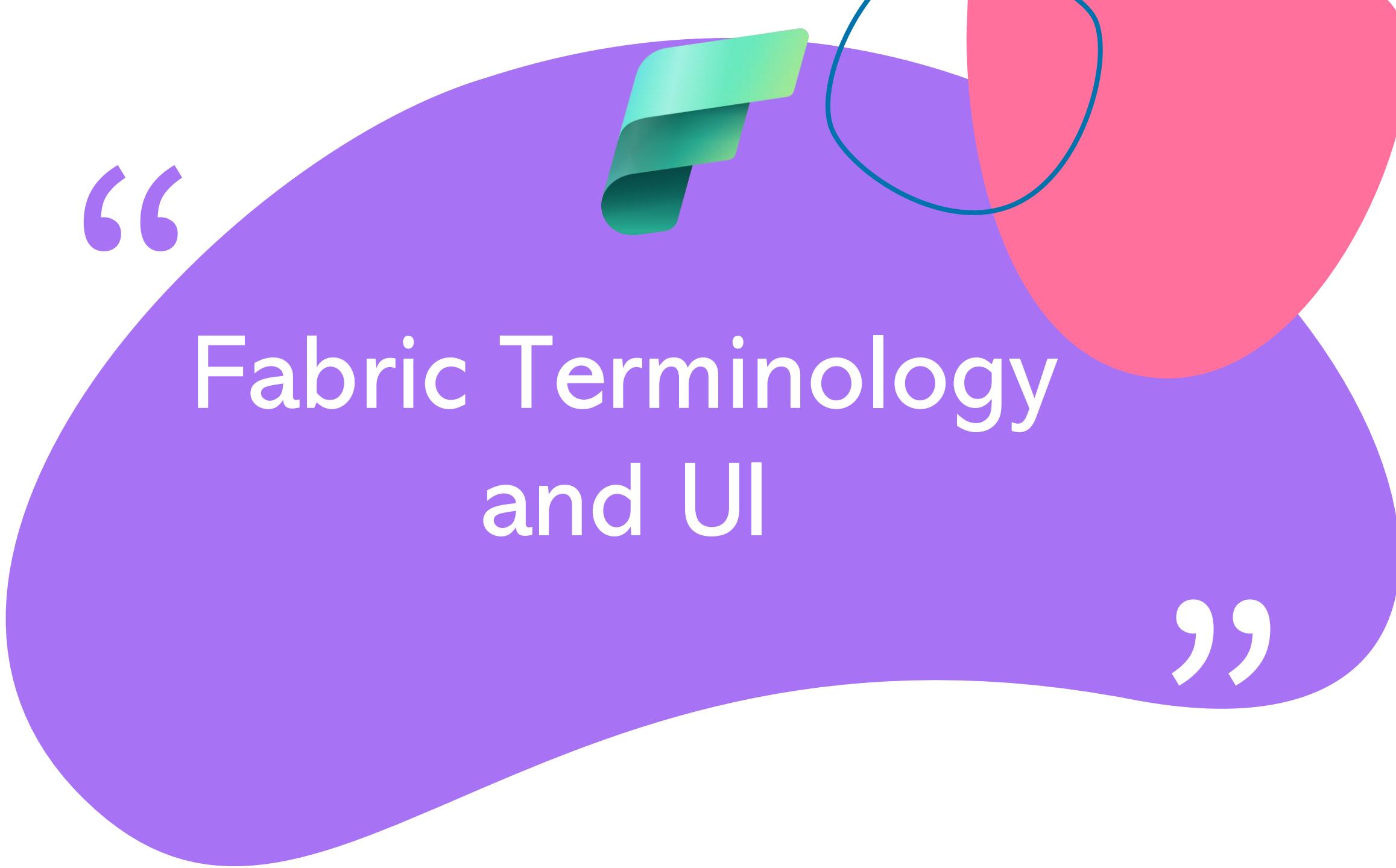
**Integration challenges with
non-Microsoft systems**



Existing vendor agreements



**Limited use for advanced
analytics**

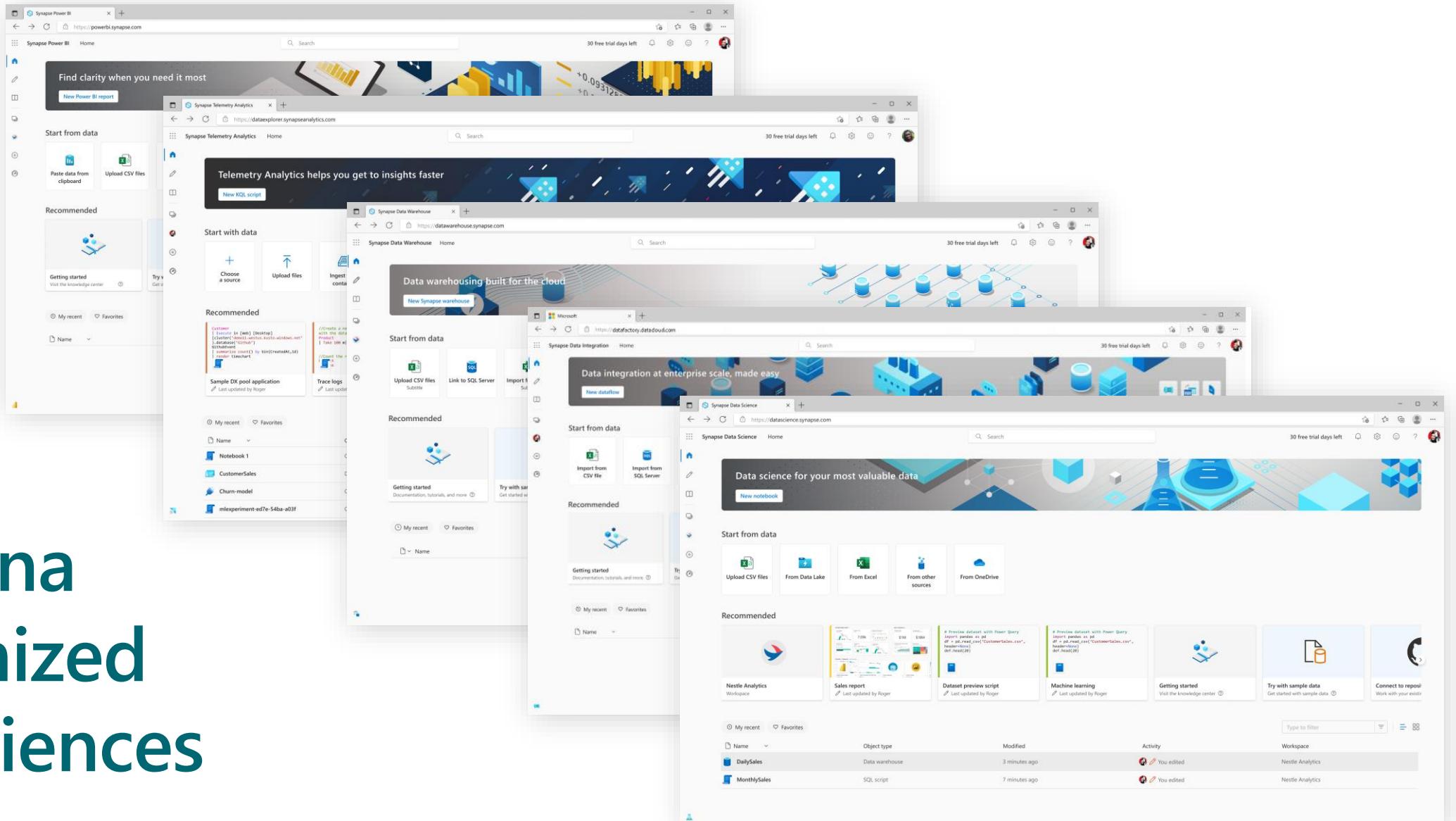


“

Fabric Terminology and UI

”

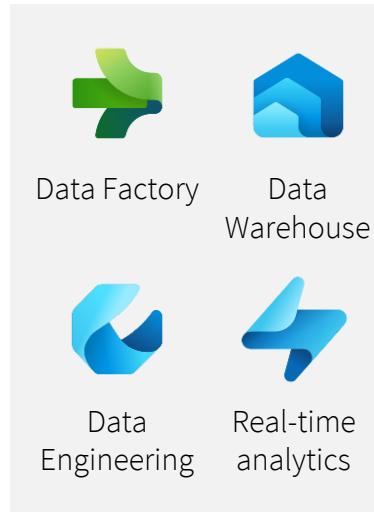
Persona optimized experiences



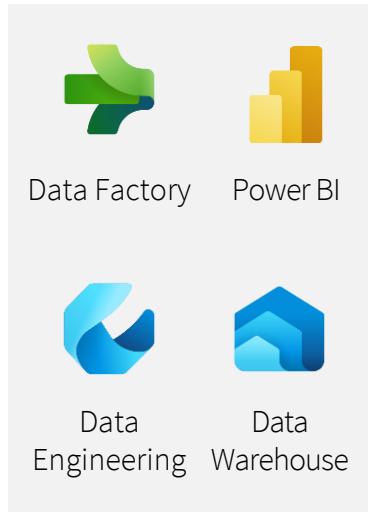
Data teams in the era of Fabric

Data professionals to work together!

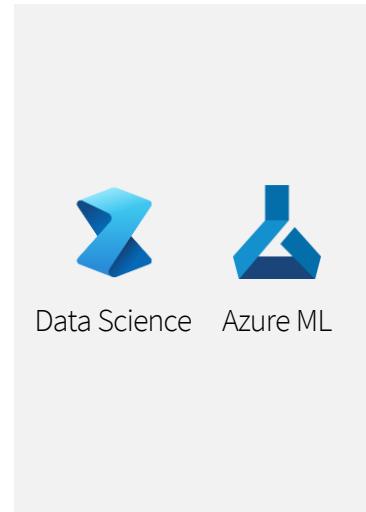
Data Engineers



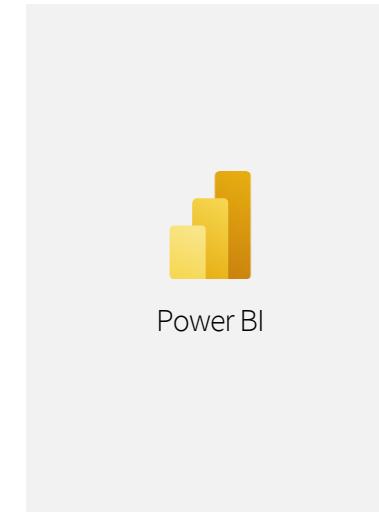
Analytics Engineers



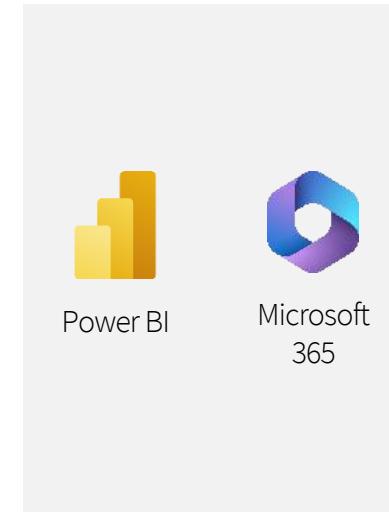
Data Scientists



Data Analysts



Decision Makers



Data Stewards



Microsoft Fabric Terminology

Tenant - organization's fabric instance

Capacity - compute power

Domain - logical grouping of workspaces

Workspace - collaborative container

Items - units of experience

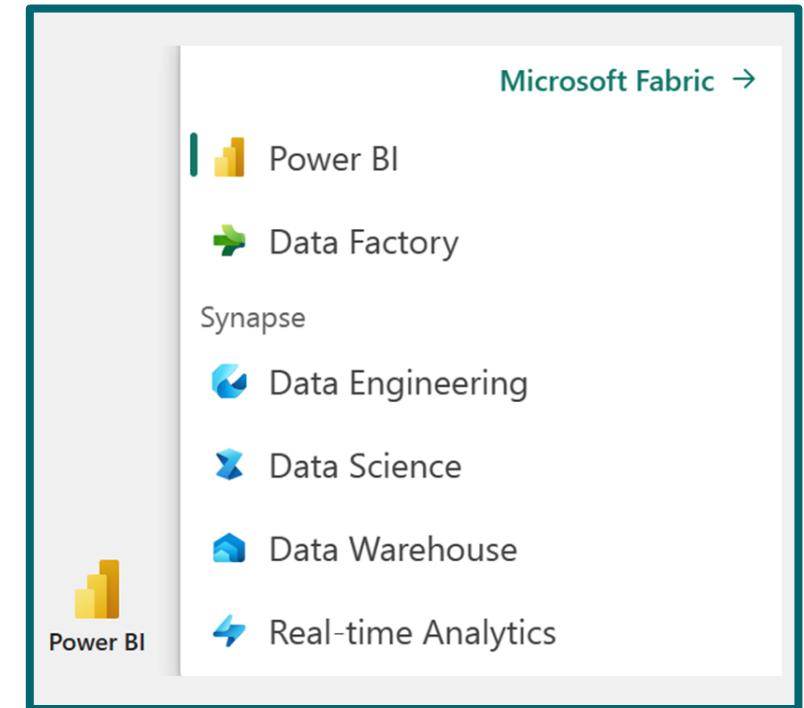


Enable Microsoft Fabric

Fabric **must be enabled** in the tenant by either:

- Fabric admin (FKA Power BI admin)
- Power Platform admin
- Microsoft 365 admin

Once enabled, workspace(s) must be assigned to Premium per capacity or Fabric license mode (P or F SKU)





Fabric Core Components

“

”

Microsoft Fabric Core Components

						
Data Integration Data Factory	Data Engineering Synapse	Data Warehouse Synapse	Data Science Synapse	Real Time Analytics Synapse	Business Intelligence Power BI	Observability Data Activator

- **Data Factory:** Data integration combining Power Query with the scale of Azure Data Factory to move and transform data
- **Synapse Data Engineering:** Data engineering with a Spark platform for data transformation at scale
- **Synapse Data Warehouse:** Data warehousing with SQL performance and scale to support data use
- **Synapse Data Science:** Data Science with Azure Machine Learning and Spark for model training and execution tracking
- **Synapse Real-Time Analytics:** Real-time analytics to query and analyze large volumes of data in real-time
- **Power BI:** Business intelligence for translating data into decisions
- **Data Activator:** Real-time detection and monitoring of data that can trigger notifications and actions when it finds specified patterns in data

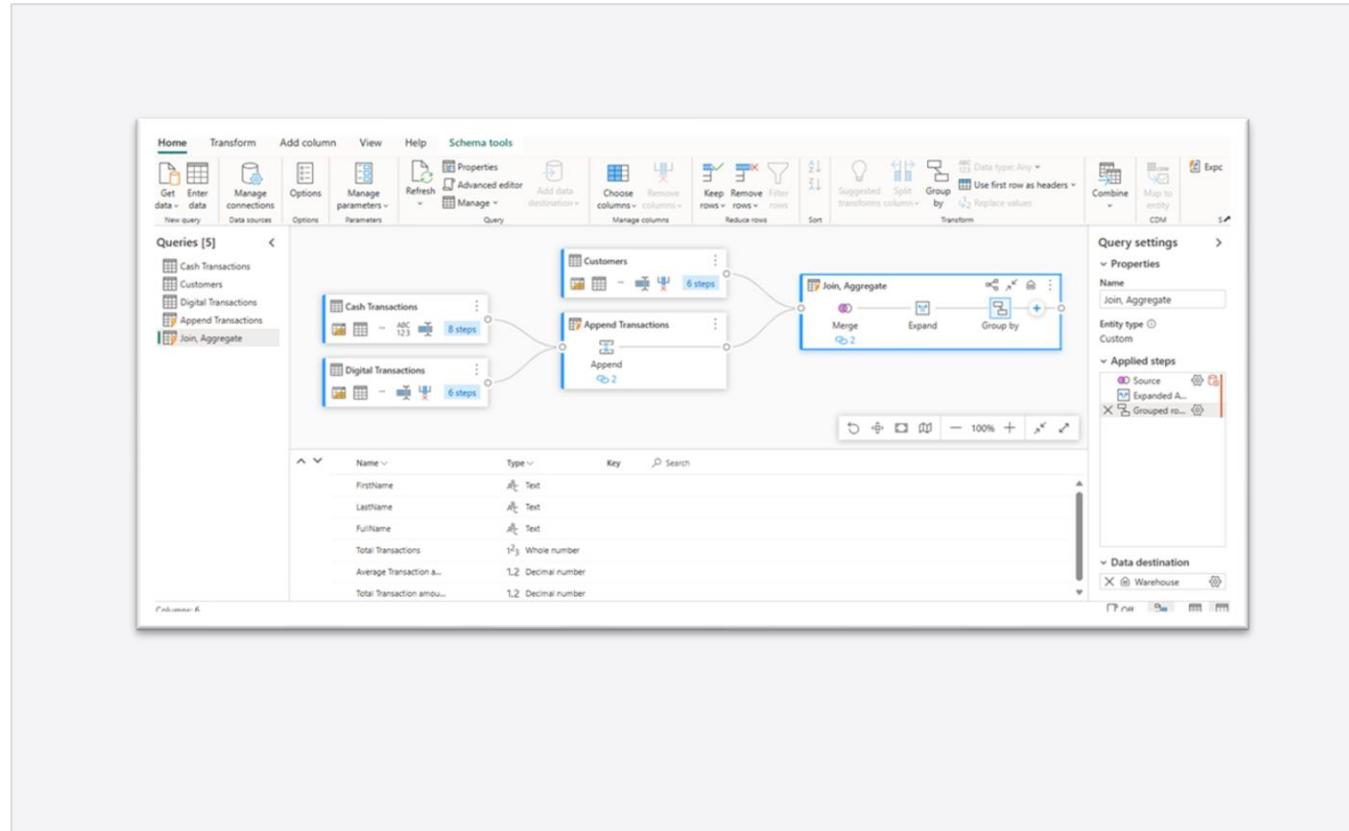
Data Integration in Microsoft Fabric



Data Integration

Dataflows (Gen2)

- Low-code graphical environment for defining ETL solutions
- Extract data from multiple sources, transform it, and load it into a destination
- Run dataflows independently or as an activity in a Pipeline



Dataflow Gen2 Pros and Cons



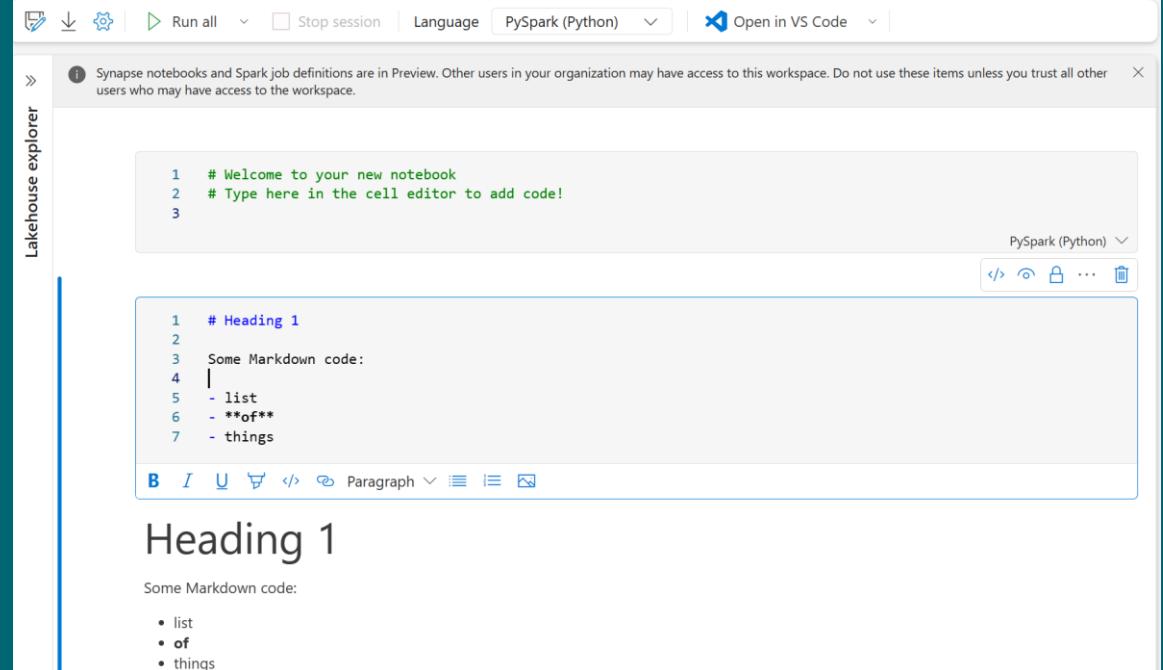
- Allow self-service users access to a subset of data warehouse separately
- Optimize performance – extract once, reuse multiple times
- Reduce complexity – expose only dataflows to larger analyst groups
- Ensure consistency and quality of data by enabling users to clean and transform data before loading it to a destination



- Row-level security isn't supported
- Fabric capacity workspace is required

Fabric Notebooks

- Code (PySpark, Scala, R, Spark SQL)
- Markdown (comments)
- Run or freeze individual or multiple cells
- Ingest and transform
- Support automation



The screenshot shows a Fabric Notebook interface with the following content:

```
1 # Welcome to your new notebook
2 # Type here in the cell editor to add code!
3
```

```
1 # Heading 1
2
3 Some Markdown code:
4
5 - list
6 - **of**
7 - things
```

Below the code cells, there is a heading and some list items:

Heading 1

Some Markdown code:

- list
- of
- things

Connect to external data sources

Use a SAS token

- Fabric shortcuts offer easy connections to external sources
- When not available, notebooks can programmatically connect to and ingest data from an **external** source

```
# Azure Blob Storage access info  
  
blob_account_name = "azureopendatastorage"  
blob_container_name = "nyctlc"  
blob_relative_path = "yellow"  
blob_sas_token = "<yourSASToken>"  
  
# Construct the path for connection  
  
wasbs_path =  
f'wasbs://{{blob_container_name}}@{{blob_account_name}}.blob.core.windows.net/{{blob_relative_path}}  
?{{blob_sas_token}}'  
  
# Read parquet data from Azure Blob Storage path  
  
blob_df = spark.read.parquet(wasbs_path)  
  
# Show the Azure Blob DataFrame  
blob_df.show()
```

Write data in a lakehouse

Write to a Parquet file

Parquet format allows the lakehouse to distribute and optimize performance in the Spark engine

```
# Write DataFrame to Parquet file format  
  
parquet_output_path =  
    "dbfs:/FileStore/your_folder/your_file_name"  
df.write.mode("overwrite").parquet(parquet_o  
utput_path)  
print(f"DataFrame has been written to  
Parquet file: {parquet_output_path}")
```

```
# Write DataFrame to Delta table
```

```
delta_table_name = "your_delta_table_name"  
df.write.format("delta").mode("overwrite").s  
aveAsTable(delta_table_name)  
print(f"DataFrame has been written to Delta  
table: {delta_table_name}")
```

Write to a Delta table

Optimize Delta table writes

- Use **Delta format** for durability and scale.
- Optimize read and write with **V-Order** and **optimized write** options.

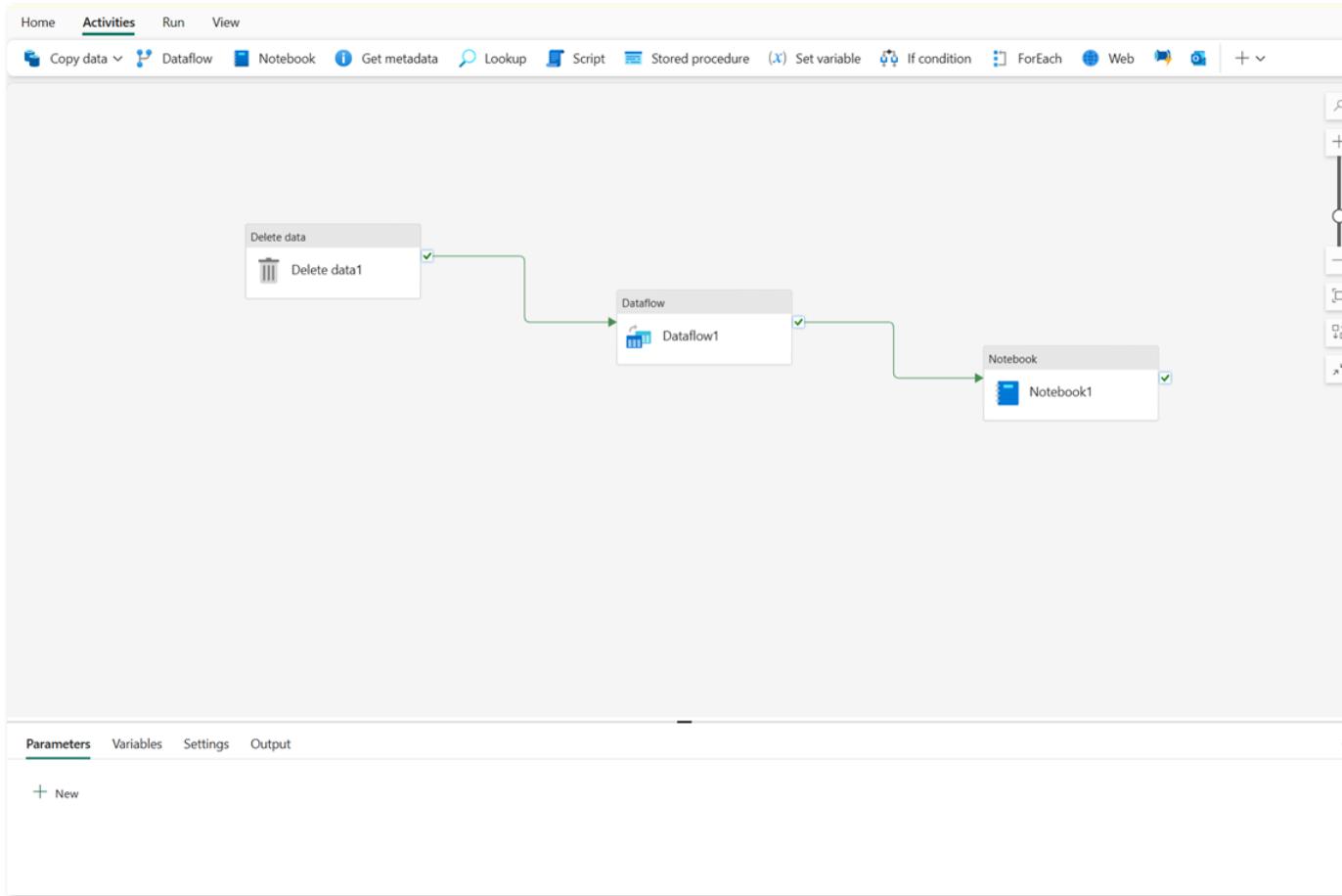
```
# Use format and save to load as a Delta table
table_name = "nyctaxi_raw"
filtered_df.write.mode("overwrite").format("delta").save(f"Tables/{table_name}")

# Confirm load as Delta table
print(f"Spark DataFrame saved to Delta table: {table_name}")

# Enable V-Order
spark.conf.set("spark.sql.parquet.vorder.enabled", "true")

# Enable automatic Delta optimized write
spark.conf.set("spark.microsoft.delta.optimizeWrite.enabled", "true")
```

Fabric Pipelines



Pipeline concepts:

- Activities
- Data transformation
- Control flow
- Parameters
- Pipeline runs

Common Activities – Copy Data

The screenshot shows the Azure Data Factory interface. On the left, a 'Copy data' wizard window is open, guiding the user through selecting a data source and destination. On the right, the main workspace shows a pipeline canvas with a 'Copy data' activity named 'Copy_Product_Data'. A large purple arrow points from the wizard window towards the pipeline canvas, indicating the transition from configuration to execution.

Copy data

Choose data source

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

Sample data

Data sources

All categories Workspace Azure Database File Generic protocol Services and apps

Home Activities Run View

Copy data

Copy_Product_Data

Destination

Data store type

Workspace

Workspace data store type

Lakehouse

Root folder

Table name *

WWI_Products

Advanced

1. Use the copy data tool

2. Edit the settings below the pipeline canvas

Common Activities – Pipeline Templates

Use templates to build common pipeline scenarios.

The screenshot shows the Azure Data Factory interface for creating a new pipeline. On the left is a navigation sidebar with options like Home, Create, Browse, OneLake Data Hub, Monitoring Hub, Workspaces, and a specific workspace named DP-601. The main area has a header "Start building your data pipeline" with three buttons: "Add pipeline activity", "Copy data", and "Choose a task to start". The third button is highlighted with a red box. To the right is a "Templates" section displaying nine data transfer and processing templates, each with a preview icon, title, and brief description. At the bottom are "Next" and "Cancel" buttons.

Templates

Start building your data pipeline

Add pipeline activity

Copy data

Choose a task to start

Bulk Copy from Database by Microsoft

Bulk Copy from Files to Database by Microsoft

Copy data from ADLS Gen2 to ADLS Gen2 by Microsoft

Copy data from ADLS Gen2 to Lakehouse Table by Microsoft

Copy data from Azure SQL DB to Lakehouse Table by Microsoft

Copy multiple files containers between File Stores by Microsoft

Copy new files only by LastModifiedDate by Microsoft

Delete files older than 30 days by Microsoft

Delta copy from Database by Microsoft

Move files by Microsoft

Next

Cancel

Microsoft Fabric Core Components



Data Engineering

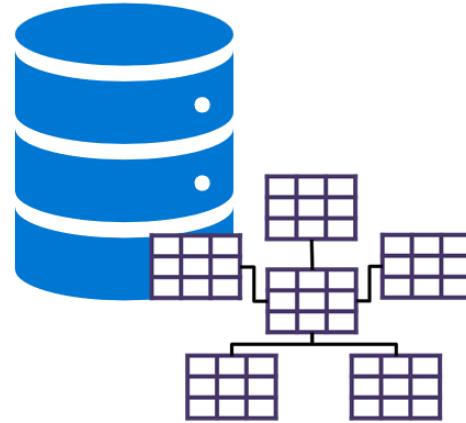
Synapse

Fabric “Houses” - Lakehouse



Data **Lakehouse**

- Scalable, distributed file storage
- Flexible schema-on-read semantics
- Big data technology compatibility



Data **Warehouse**

- Relational schema modeling
- SQL-based querying
- Proven basis for reporting and analytics



Fabric Lakehouse in a Nutshell

The screenshot shows the Microsoft Fabric Data Engineering interface. On the left, there's a navigation sidebar with icons for Home, Create, Browse, OneLake data hub, Monitoring hub, Workspaces, and DP-601. Under DP-601, there's a sub-item for DP601_Bronze. At the bottom of the sidebar is a Data Engineering icon.

The main area displays the "Details for DP601_Bronze" page. It includes:

- Owner:** Owner (DP-601)
- Dataset (default):** DP601_Bronze
- SQL endpoint:** DP-601
- Lakehouse:** DP601_Bronze
- Location:** DP-601
- Refreshed:** 7/5/23, 4:45:00 PM
- Sensitivity:** Confidential\Microsoft Extended

Below the details, there are two cards:

- Visualize this data:** Create an interactive report, or a table, to discover and share business insights. [Learn more](#)
- Share this data:** Give people access to the dataset and set their permissions to work with it. [Learn more](#)

A "Create a report" button is also present.

At the bottom, there's a section titled "See what already exists" which lists items sharing the same data source:

Name	Type	Relation	Location	Refreshed	Endorsement	Sensitivity
DP601_Bronze	SQL endpoint	Upstream	DP-601	—	—	Confidential\Microsoft Extend...
DP601_Bronze	Lakehouse	Upstream	DP-601	—	—	Confidential\Microsoft Extend...

At the bottom right, there are "Filter by keyword" and "Filter" buttons.



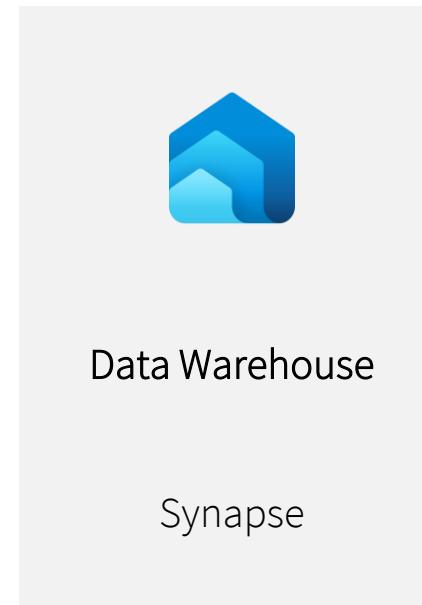
Working With Fabric Lakehouse

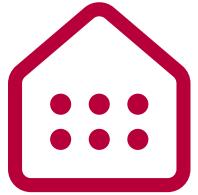
Tools and techniques to explore and transform data

- ★ Apache Spark
 - ⌚ Notebooks
 - ✳️ Spark Job Definitions
- 🏠 SQL analytics endpoint
- ⌚ Dataflows (Gen2)
- ⌚ Data Pipelines

Visualize with Power BI

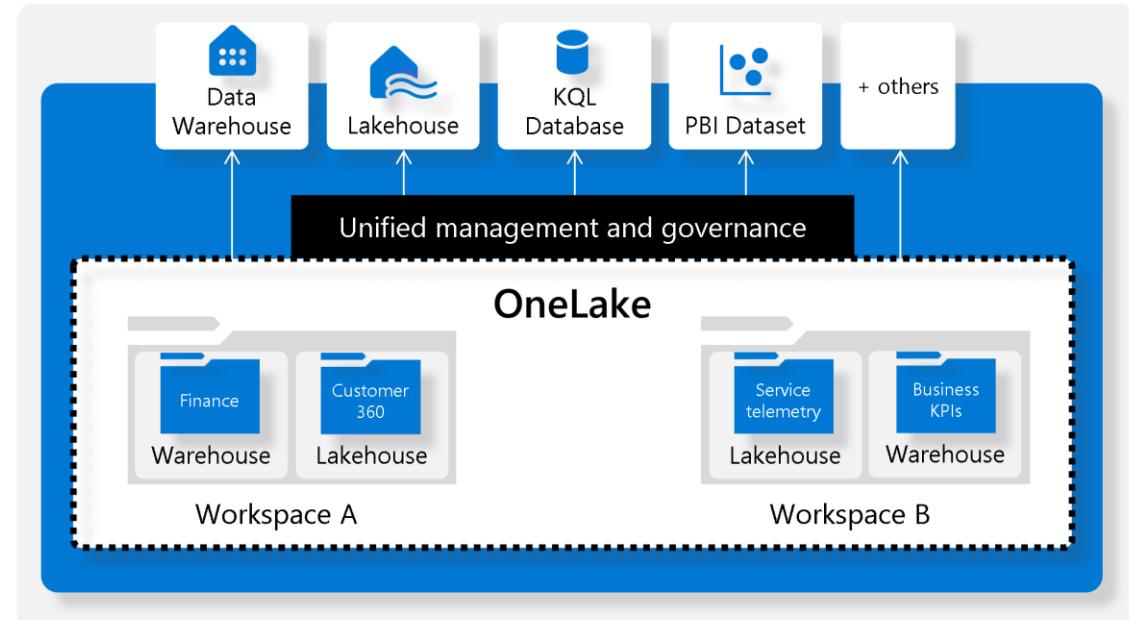
Microsoft Fabric Core Components





Fabric Warehouse in a Nutshell

- Centered on the single data lake (OneLake)
- Powered by Synapse Analytics
- (Almost) Fully supports T-SQL
- Parquet file format

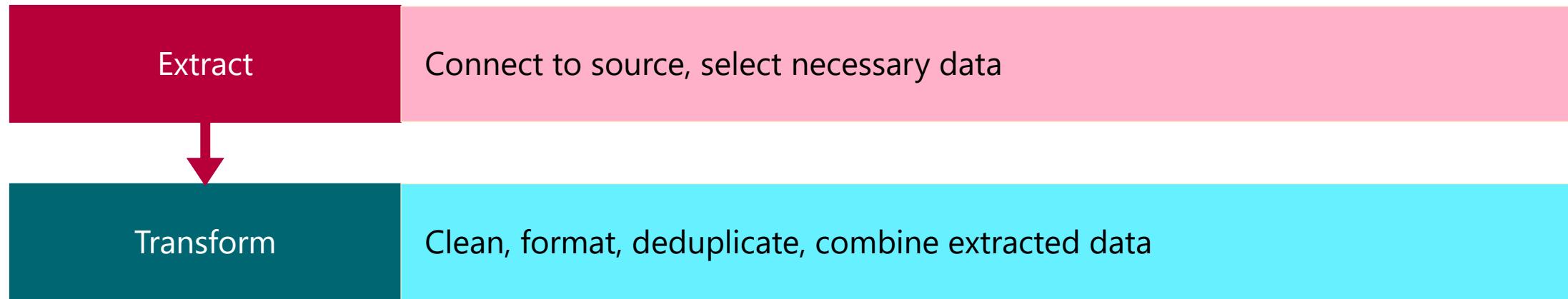


Understand ETL (Extract, Transform and Load)

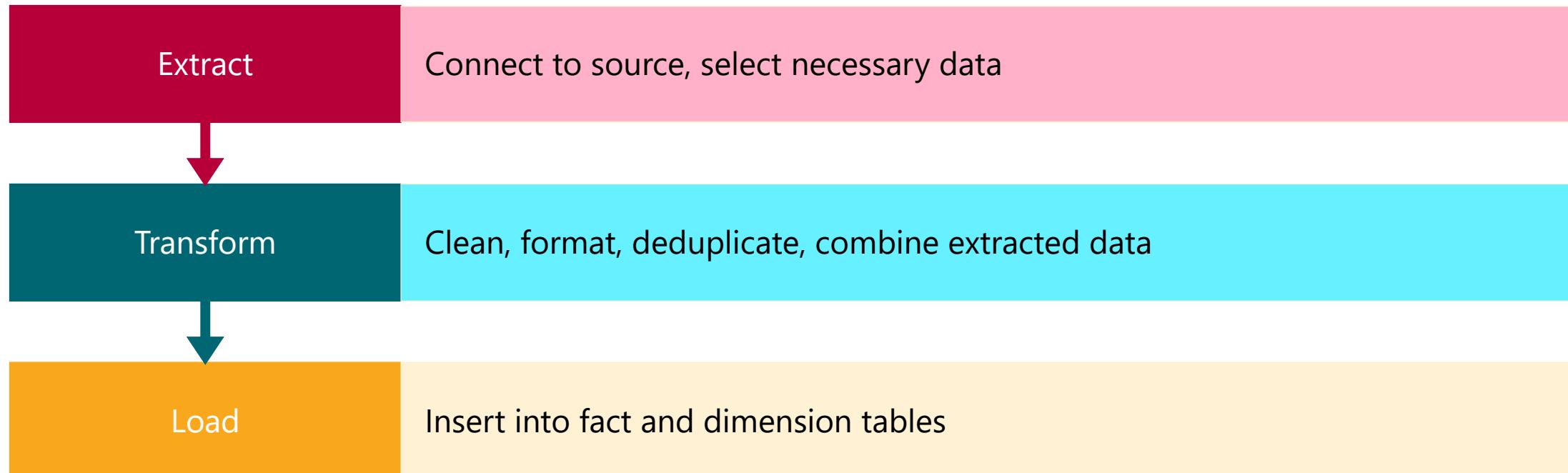
Extract

Connect to source, select necessary data

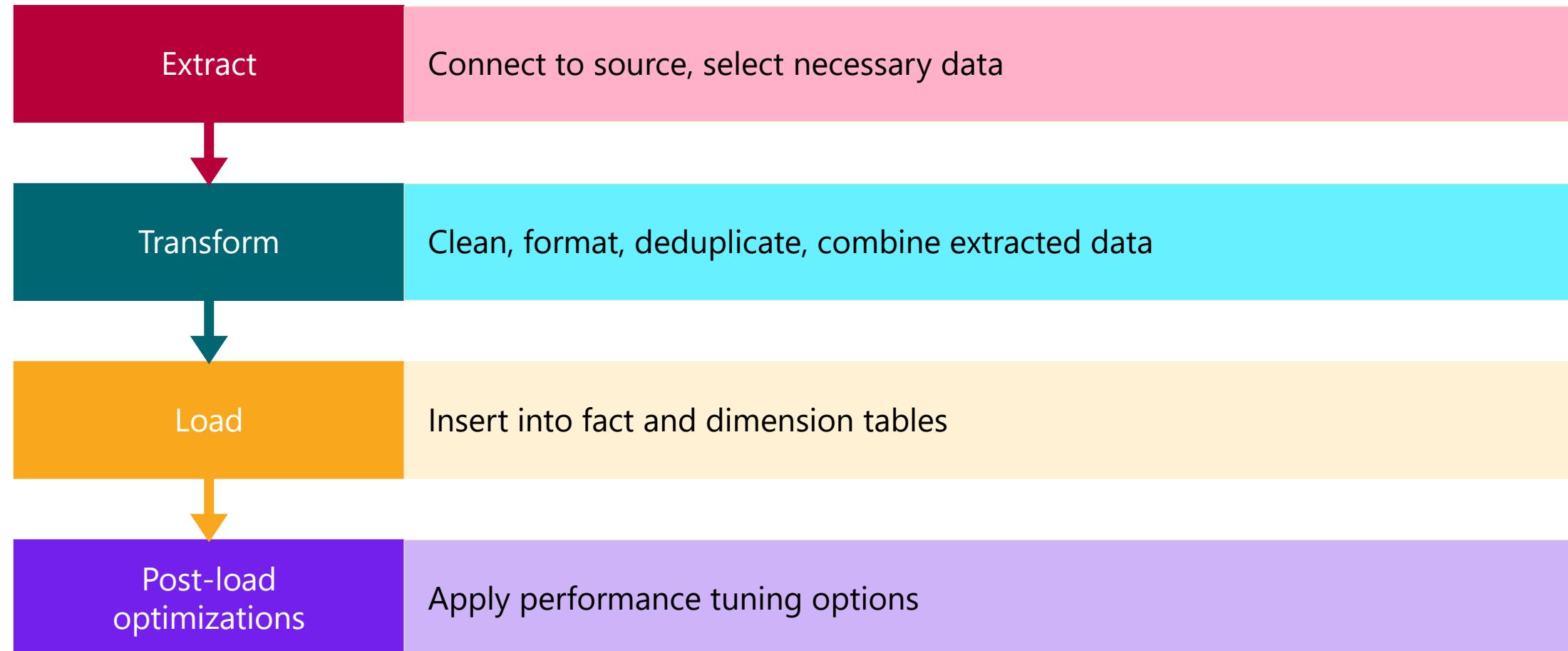
Understand ETL (Extract, Transform and Load)



Understand ETL (Extract, Transform and Load)



Understand ETL (Extract, Transform and Load)



Data Loading Process

Full loading

- ✓ Truncate table and load data
- ✓ Longer load time
- ✓ No history stored
- ✓ Best for initial load or refresh

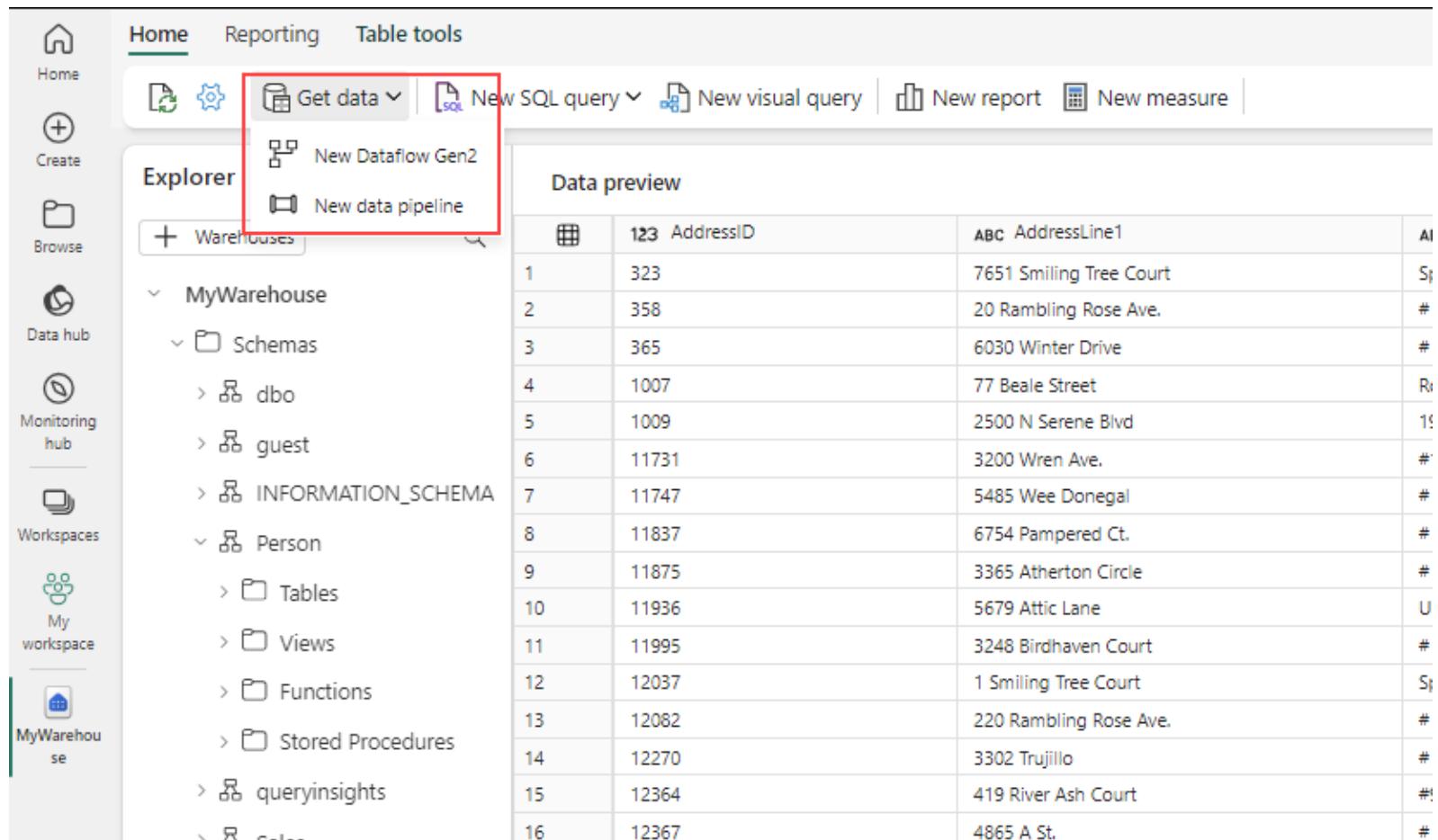
Incremental loading

- ✓ Append data to tables
- ✓ Faster updates
- ✓ Preserves history (timestamp)
- ✓ Best for frequent updates

Load data with Fabric pipelines

Orchestrate activities

- Copy Data
- Schedule
- Parameters
- Delete Data



The screenshot shows the Power BI desktop application interface. The top navigation bar includes 'Home', 'Reporting', and 'Table tools'. Below the navigation bar is a ribbon with several icons: 'Get data' (highlighted with a red box), 'New SQL query', 'New visual query', 'New report', and 'New measure'. The main workspace is titled 'MyWarehouse' under 'My workspace'. On the left, there's a sidebar with icons for Home, Create, Browse, Data hub, Monitoring hub, Workspaces, My workspace, and MyWarehouse. The 'MyWarehouse' section is expanded, showing 'Schemas' (dbo, guest, INFORMATION_SCHEMA), 'Tables' (Person), 'Views', 'Functions', 'Stored Procedures', 'queryinsights', and 'color'. To the right of the schema tree is a 'Data preview' table with two columns: 'AddressID' and 'AddressLine1'. The table contains 16 rows of data.

	AddressID	AddressLine1
1	323	7651 Smiling Tree Court
2	358	20 Rambling Rose Ave.
3	365	6030 Winter Drive
4	1007	77 Beale Street
5	1009	2500 N Serene Blvd
6	11731	3200 Wren Ave.
7	11747	5485 Wee Donegal
8	11837	6754 Pampered Ct.
9	11875	3365 Atherton Circle
10	11936	5679 Attic Lane
11	11995	3248 Birdhaven Court
12	12037	1 Smiling Tree Court
13	12082	220 Rambling Rose Ave.
14	12270	3302 Trujillo
15	12364	419 River Ash Court
16	12367	4865 A St.

Load data with T-SQL

Use COPY statement

- External Azure storage
- Specify file format
(PARQUET / CSV)
- Error handling
- Multiple files

```
-- Load as CSV
COPY my_table
FROM
'https://myaccount.blob.core.windows.net/myblobcontainer/
folder0/*.csv,
https://myaccount.blob.core.windows.net/myblobcontainer/f
older1/'
WITH (FILE_TYPE = 'CSV',
      CREDENTIAL=(IDENTITY= 'Shared Access Signature',
SECRET='<Your_SAS_Token>')
      FIELDTERMINATOR = '|')

-- Load as PARQUET
COPY INTO test_parquet
FROM
'https://myaccount.blob.core.windows.net/myblobcontainer/
folder1/*.parquet'
WITH (CREDENTIAL=(IDENTITY= 'Shared Access Signature',
SECRET='<Your_SAS_Token>'))
```

Load from other items

Combine into one warehouse

- Requires three-part-naming
- CREATE TABLE AS SELECT
- INSERT...SELECT

```
CREATE TABLE  
[analysis_warehouse].[dbo].[combined_data]  
AS  
SELECT  
FROM [sales_warehouse].[dbo].[sales_data] sales  
INNER JOIN [social_lakehouse].[dbo].[social_data]  
social  
ON sales.[product_id] = social.[product_id];  
  
INSERT INTO  
[analysis_warehouse].[dbo].[combined_data]  
SELECT  
    sales.product_id  
FROM [sales_warehouse].[dbo].[sales_data] sales  
INNER JOIN [social_lakehouse].[dbo].[social_data]  
social  
ON sales.product_id = social.product_id;
```

Load data with Dataflows Gen2

Ingest and transform with Power Query

- ✓ Connect to data source
- ✓ Transform data – leverage copilot
- ✓ Append or replace
- ✓ Add data destination
 - ✓ Lakehouse
 - ✓ Warehouse
 - ✓ Azure SQL Database
 - ✓ Azure Data Explorer (Kusto)
 - ✓ Azure Synapse Analytics (SQL DW)

Query settings >

Properties

Name

customer-churn 1

Entity type ⓘ

Custom

Applied steps

 Source	
 Promoted headers	
 Changed column type	
 Custom	

Data destination +

No data destination

“

Coffee Break



”



“

Understanding Delta
and Parquet File
Format

”



Parquet File Format

Data compression

Columnar storage

Language agnostic

Open-source format

Support for complex
data types





Row-Based Storage

	Product	Customer	Country	Date	Sales Amount
Row 1	Ball	John Doe	USA	2023-01-01	100
Row 2	T-Shirt	John Doe	USA	2023-01-02	200
Row 3	Socks	Maria Adams	UK	2023-01-01	300
Row 4	Socks	Antonio Grant	USA	2023-01-03	100
Row 5	T-Shirt	Maria Adams	UK	2023-01-02	500
Row 6	Socks	John Doe	USA	2023-01-05	200





Row-Based Storage

	Product	Customer	Country	Date	Sales Amount
Row 1	Ball	John Doe	USA	2023-01-01	100
Row 2	T-Shirt	John Doe	USA	2023-01-02	200
Row 3	Socks	Maria Adams	UK	2023-01-01	300
Row 4	Socks	Antonio Grant	USA	2023-01-03	100
Row 5	T Shirt	Maria Adams	UK	2023-01-02	500
Row 6	Socks	John Doe	USA	2023-01-05	200





Row-Based Storage

	Product	Customer	Country	Date	Sales Amount
Row 1	Ball	John Doe	USA	2023-01-01	100
Row 2	T-Shirt	John Doe	USA	2023-01-02	200
Row 3	Socks	Maria Adams	UK	2023-01-01	300
Row 4	Socks	Antonio Grant	USA	2023-01-03	100
Row 5	T-Shirt	Maria Adams	UK	2023-01-02	500
Row 6	Socks	John Doe	USA	2023-01-05	200





Column-Based Storage

Product	Customer	Country	Date	Sales Amount
Ball	John Doe	USA	2023-01-01	100
T-Shirt	John Doe	USA	2023-01-02	200
Socks	Maria Adams	UK	2023-01-01	300
Socks	Antonio Grant	USA	2023-01-03	100
T-Shirt	Maria Adams	UK	2023-01-02	500
Socks	John Doe	USA	2023-01-05	200





**Parquet is a columnar
format that stores the data
in row groups!**



Parquet Storage

	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	Antonio Grant	USA	2023-01-03	100
Row group 3	T-Shirt	Maria Adams	UK	2023-01-02	500
	Socks	John Doe	USA	2023-01-05	200





Projection and Predicate(s)

Projection = SELECT

Predicate(s) = WHERE

	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 Socks	John Doe	UK USA	2023-01-03	300
		Maria Adams	UK	2023-01-02	500
Row group 3	T-Shirt	Maria Adams	UK	2023-01-05	200
	Socks	John Doe	USA		

The engine will skip scanning these records!



@DataMozart

Row-Based Storage: 5 Columns + 6 Rows



	Product	Customer	Country	Date	Sales Amount
Row 1	Ball	John Doe	USA	2023-01-01	100
Row 2	T-Shirt	John Doe	USA	2023-01-02	200
Row 3	Socks	Maria Adams	UK	2023-01-01	300
Row 4	Socks	Antonio Grant	USA	2023-01-03	100
Row 5	T Shirt	Maria Adams	UK	2023-01-02	500
Row 6	Socks	John Doe	USA	2023-01-05	200



Column-Based Storage: 2 Columns + 6 Rows

Product	Customer	Country	Date	Sales Amount
Ball	John Doe	USA	2023-01-01	100
T-Shirt	John Doe	USA	2023-01-02	200
Socks	Maria Adams	UK	2023-01-01	300
Socks	Antonio Grant	USA	2023-01-03	100
T-Shirt	Maria Adams	UK	2023-01-02	500
Socks	John Doe	USA	2023-01-05	200

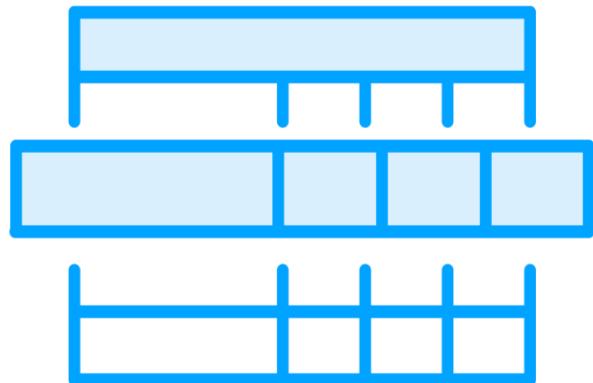




Column Storage With Row Groups: 2 Columns + 4 Rows

	Column 1	Column 2	Column 3	Column 4	Column 5
	Product	Customer	Country	Date	Sales Amount
	Ball	John Doe	USA	2023-01-01	100
	T-Shirt	John Doe	USA	2023-01-02	200
	Socks Socks	The engine will skip scanning these records!			
	T-Shirt	Maria Adams	UK	2023-01-02	500
	Socks	John Doe	USA	2023-01-05	200

How Parquet “Knows” Which Row Group to Scan?



Parquet contains metadata

- Data about data

Min and max values

Footer

- Format version
- Schema information
- Column metadata

Performance tip!

- Merge multiple smaller files into one bigger
- A few hundred MBs



@DataMozart

Can It Be Better Than This?



Data compression

1. Dictionary encoding

Ball	0
T-Shirt	1
Socks	2
Socks	2
T-Shirt	1
Socks	2

0	Ball
1	T-Shirt
2	Socks

2. Run-Length encoding

Long arm T-shirt with application on the neck



@DataMozart



Can It Be Better Than THIS?!



Delta format



Parquet format on steroids!

- ✓ Versioning of Parquet files
- ✓ Stores transaction log
- ✓ Tracks all changes



@DataMozart



“

OneLake in Microsoft Fabric

”



OneLake in a Nutshell



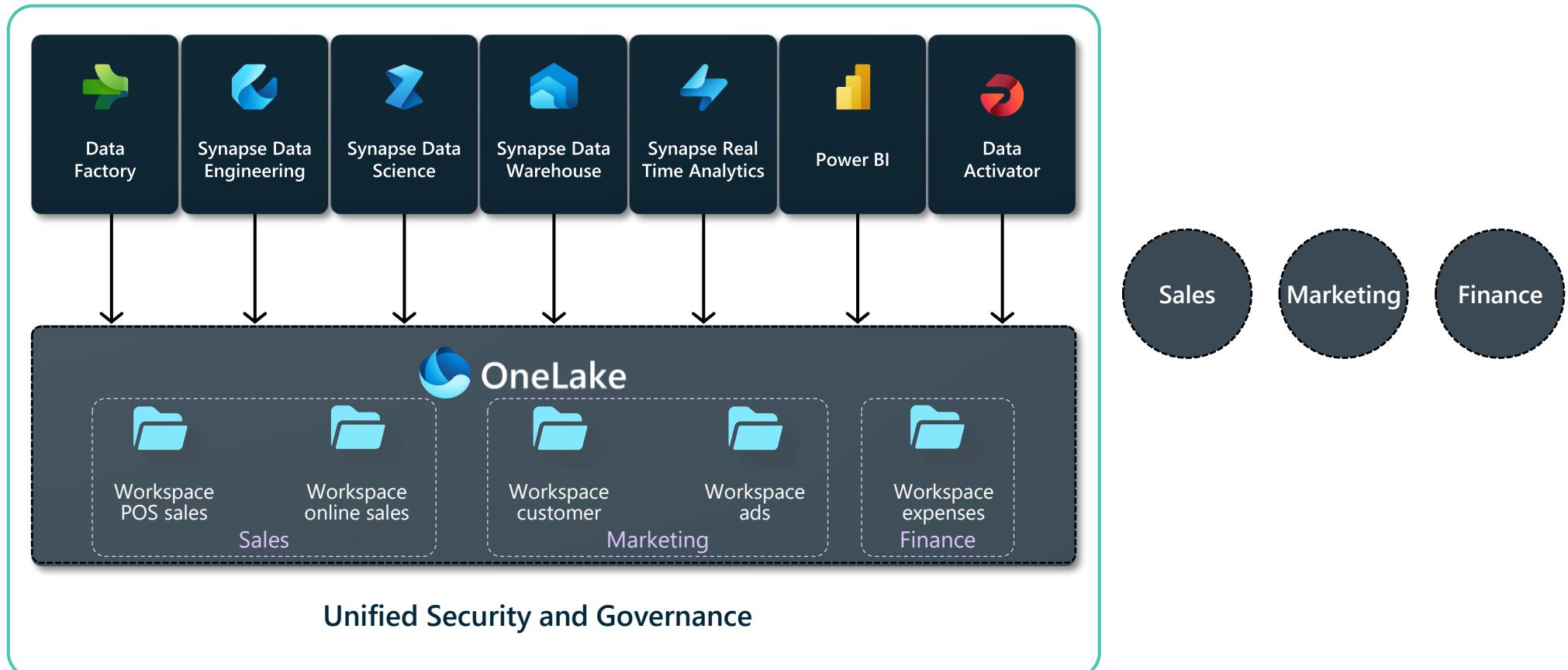
- ✓ SaaS for the entire organization (no silos)
- ✓ Foundation for all Fabric data artifacts
- ✓ No need to move/duplicate data
- ✓ Full access through APIs
- ✓ Logical abstraction of disparate sources



For All Domains



OneLake provides a true data mesh as a service

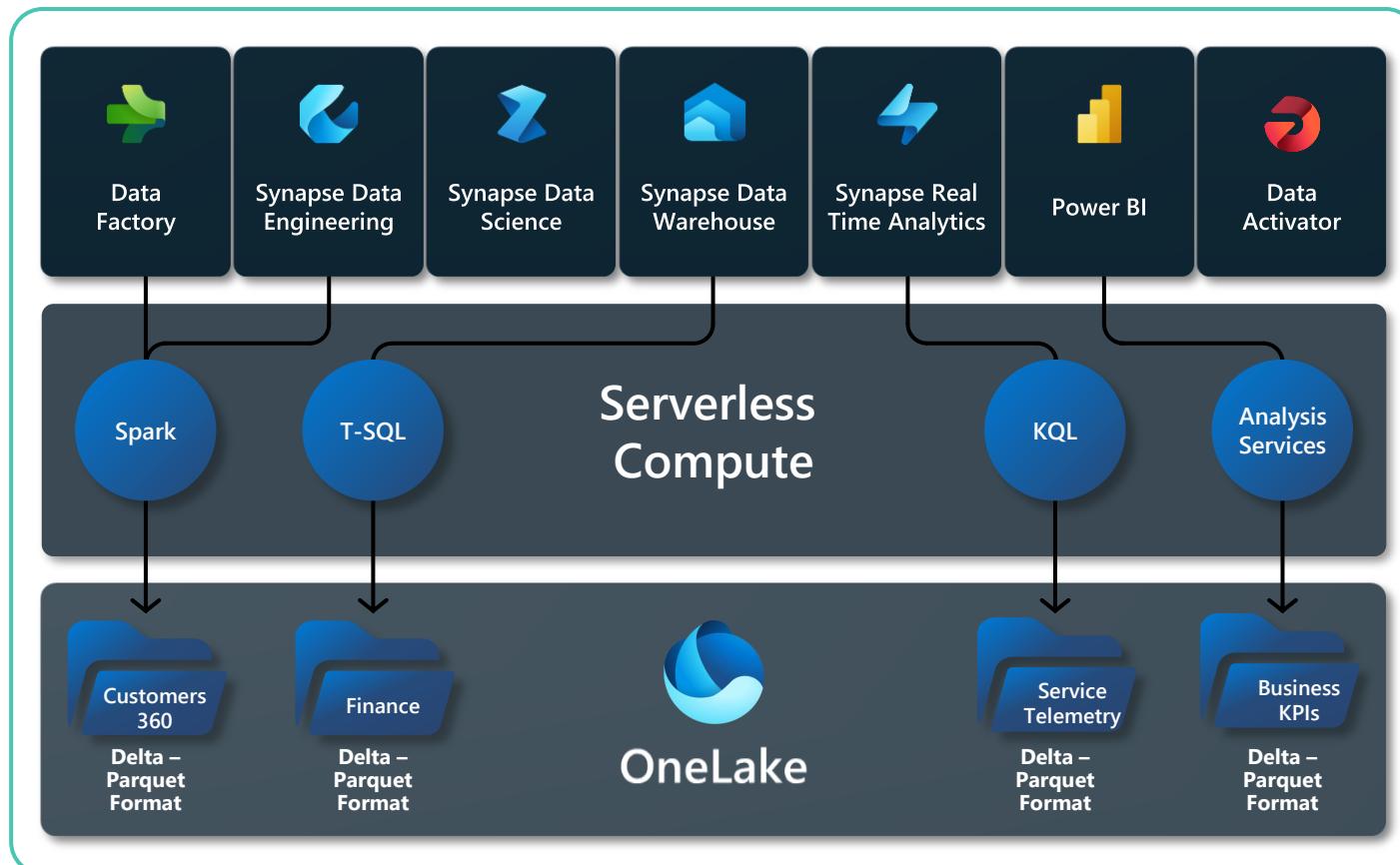




One Copy for all computers



Real separation of compute and storage

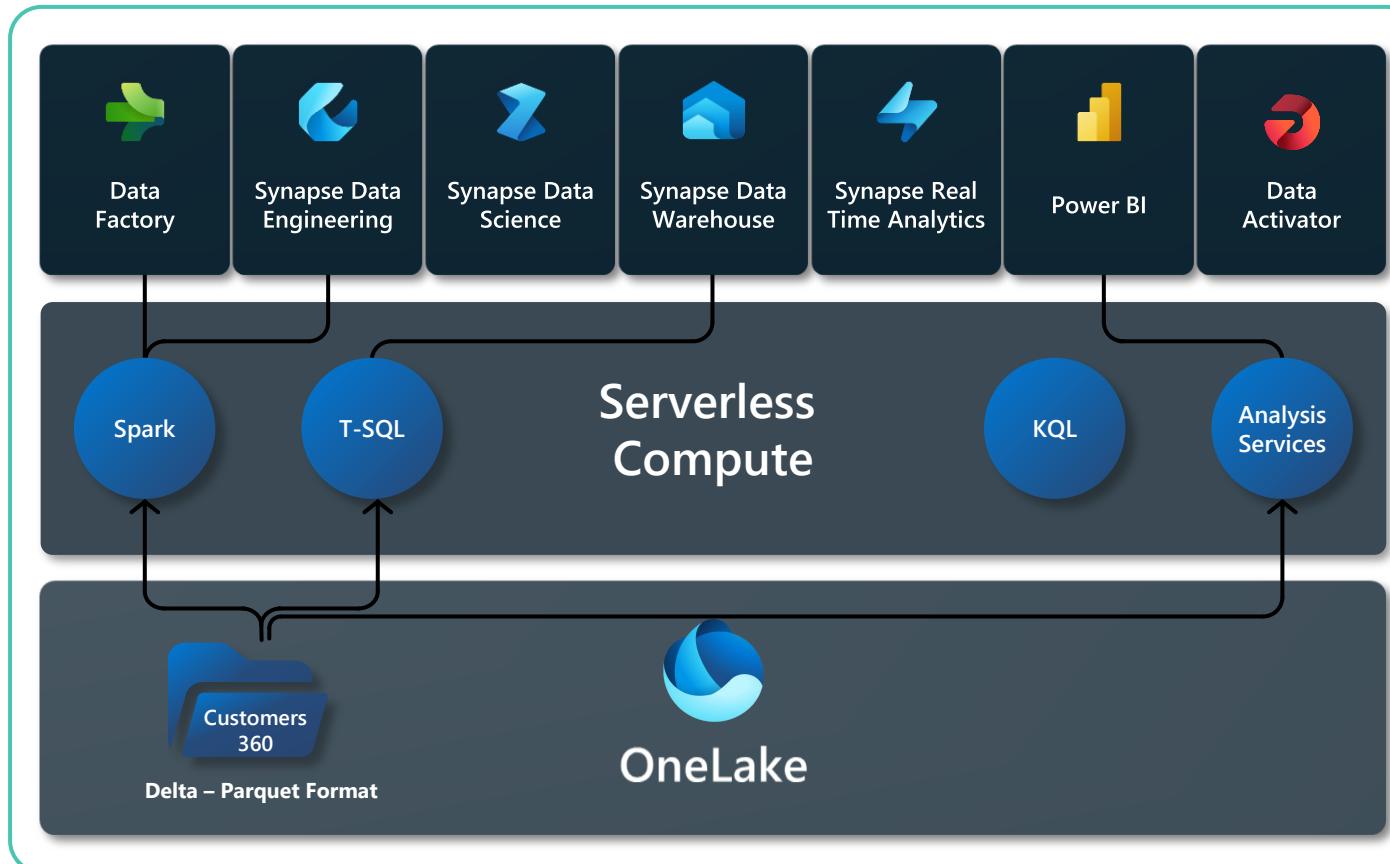




One Copy for all computers



One copy of data can be read by all engines





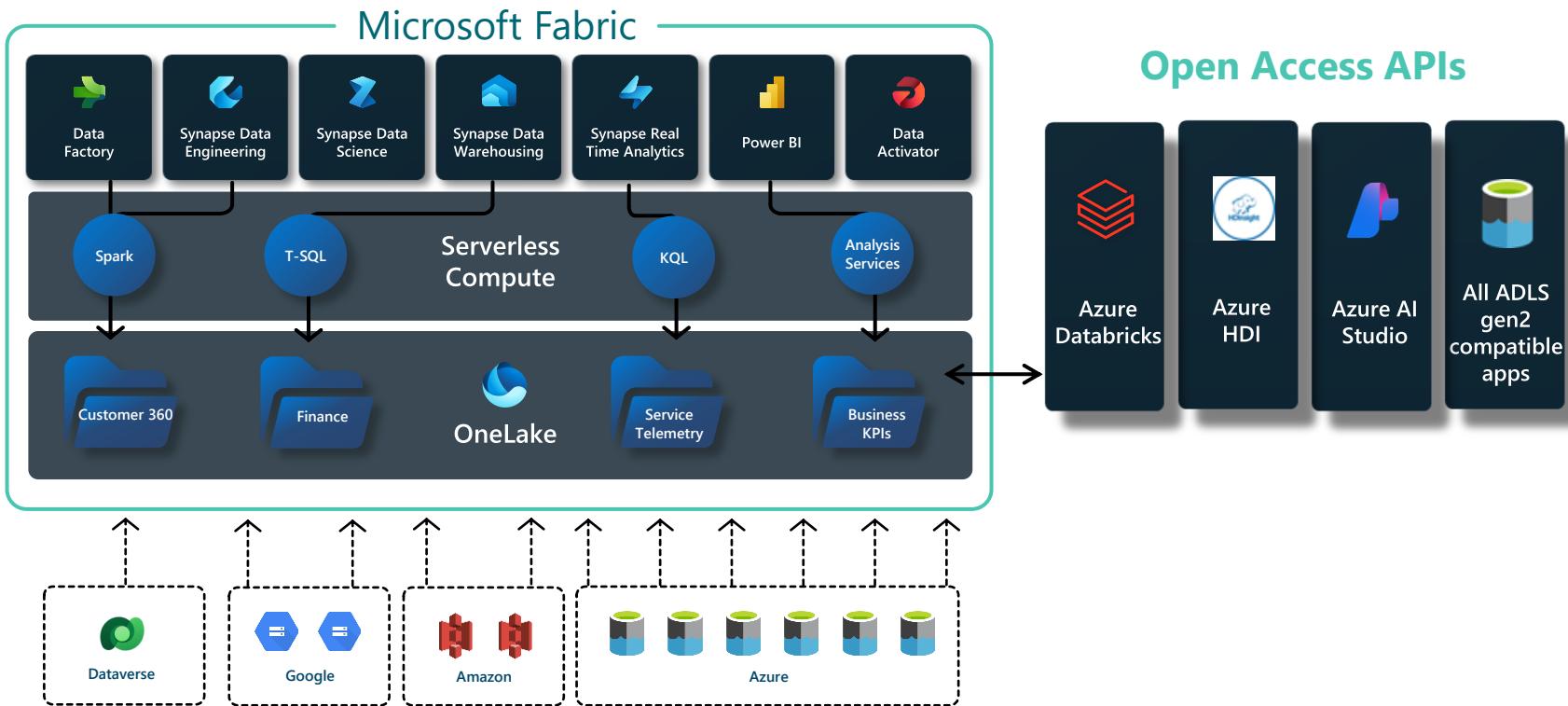
All roads lead to...

Rome

OneLake



Fabric compute engines



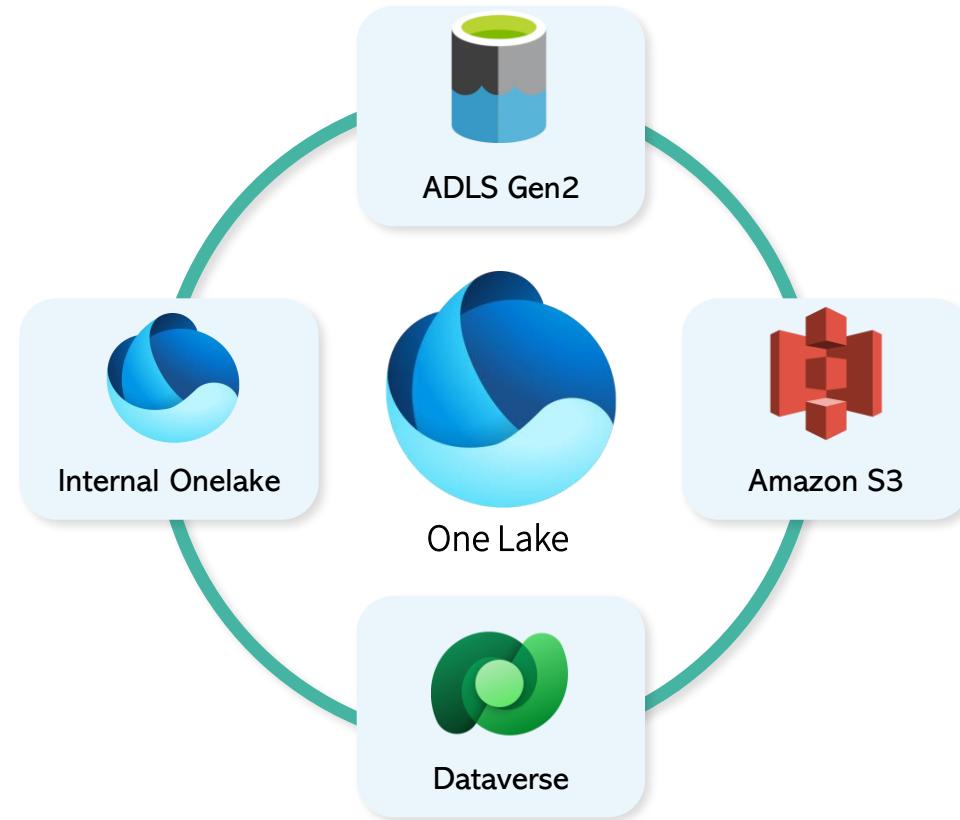
Multi-cloud shortcuts



OneLake Shortcuts

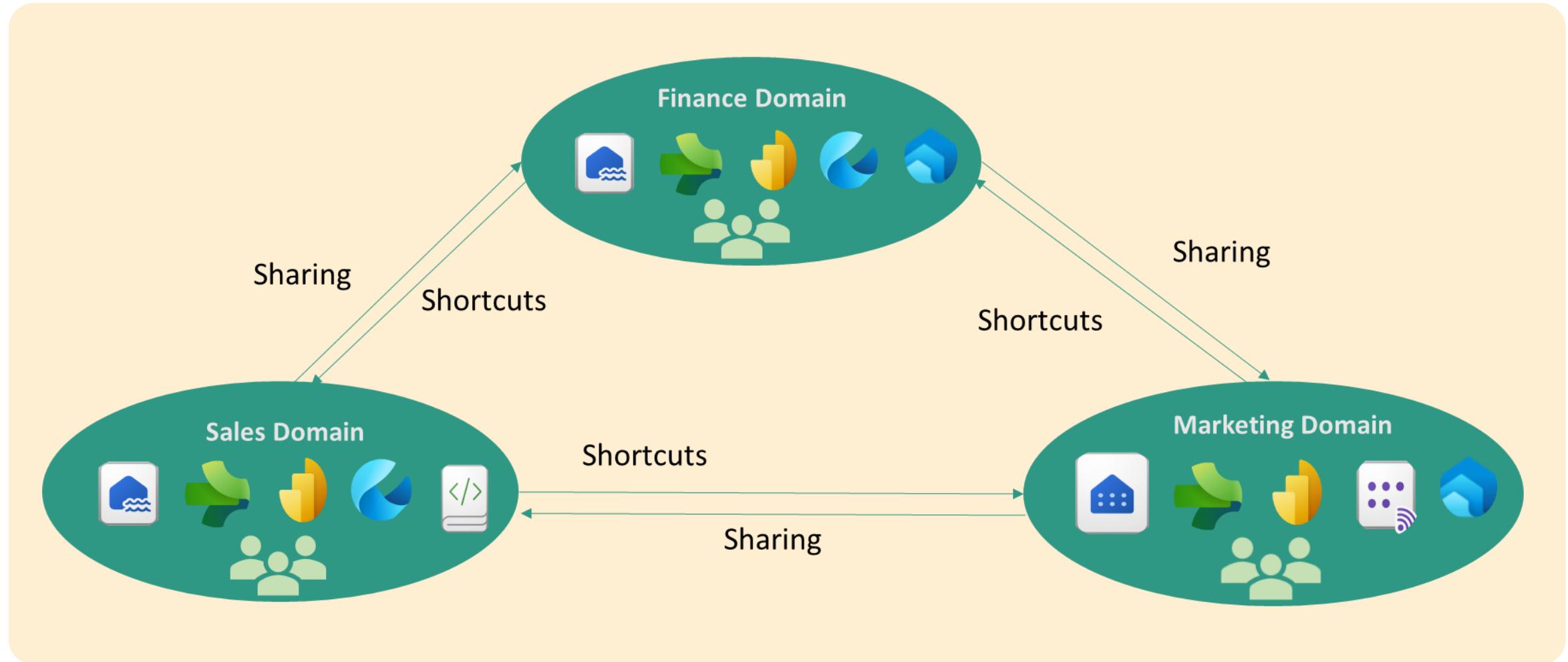


Shortcuts –
Like Windows Explorer?





Life is easier with Shortcuts



Benefits of using Shortcuts

- OneLake for all Domains
- Seamlessly query multiple data sources
- Data sharing capabilities made easy
- Unlocking the potential of Data Mesh
- Effective governance and data security



“

Lakehouse or
Warehouse: Which
One Is For Me?

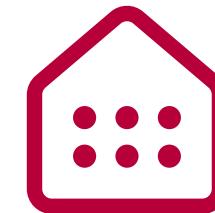
”



Two Types of SQL-“Houses” in Fabric



SQL Endpoint of the Lakehouse



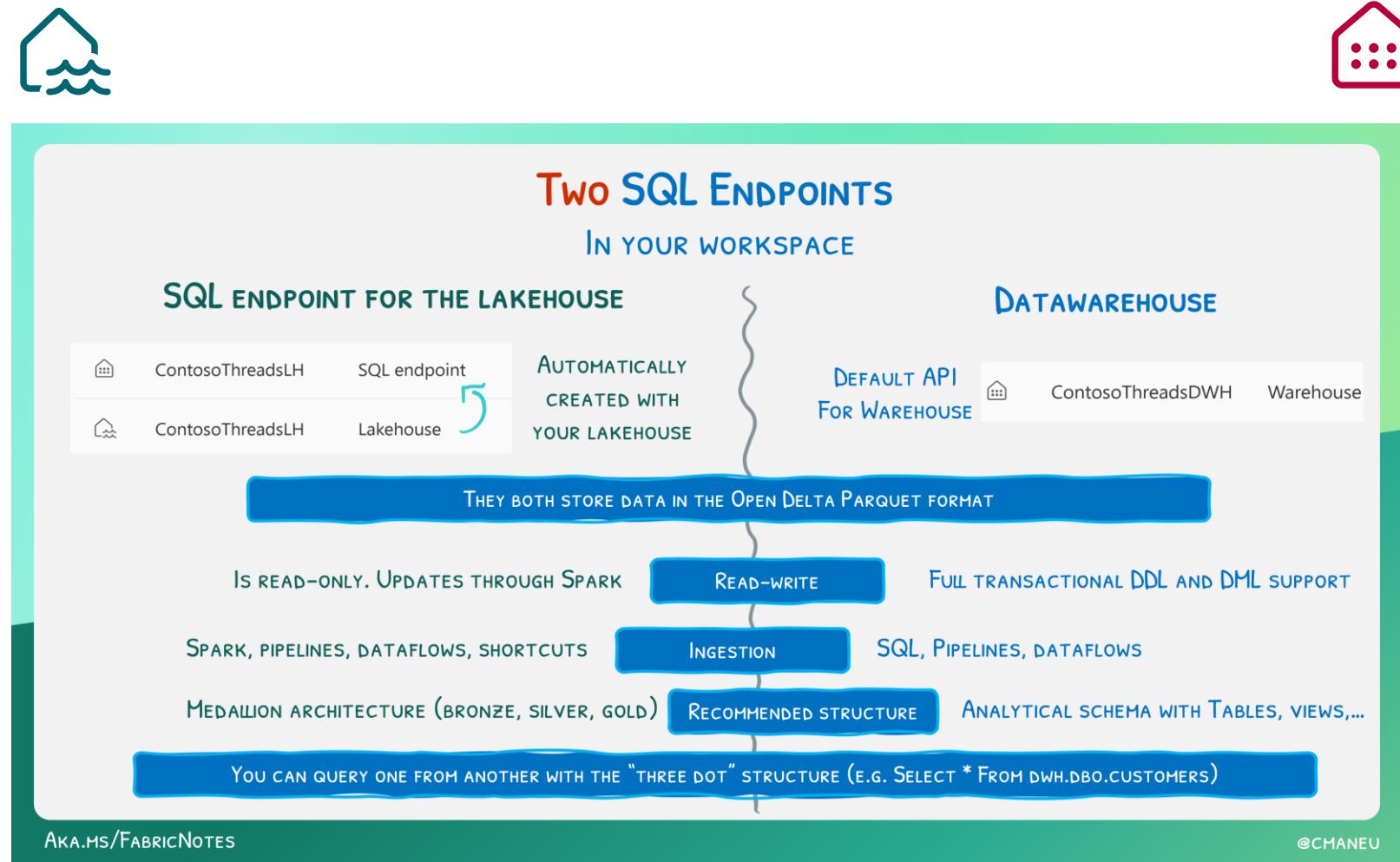
Synapse Data Warehouse

- ✓ Automatically generated
- ✓ Supports ONLY read operations
- ✓ Views, inline TVFs, procs...
- ✓ Manage permissions

- ✓ Full transactional support
- ✓ DDL/DML operations
- ✓ Traditional data warehousing workloads

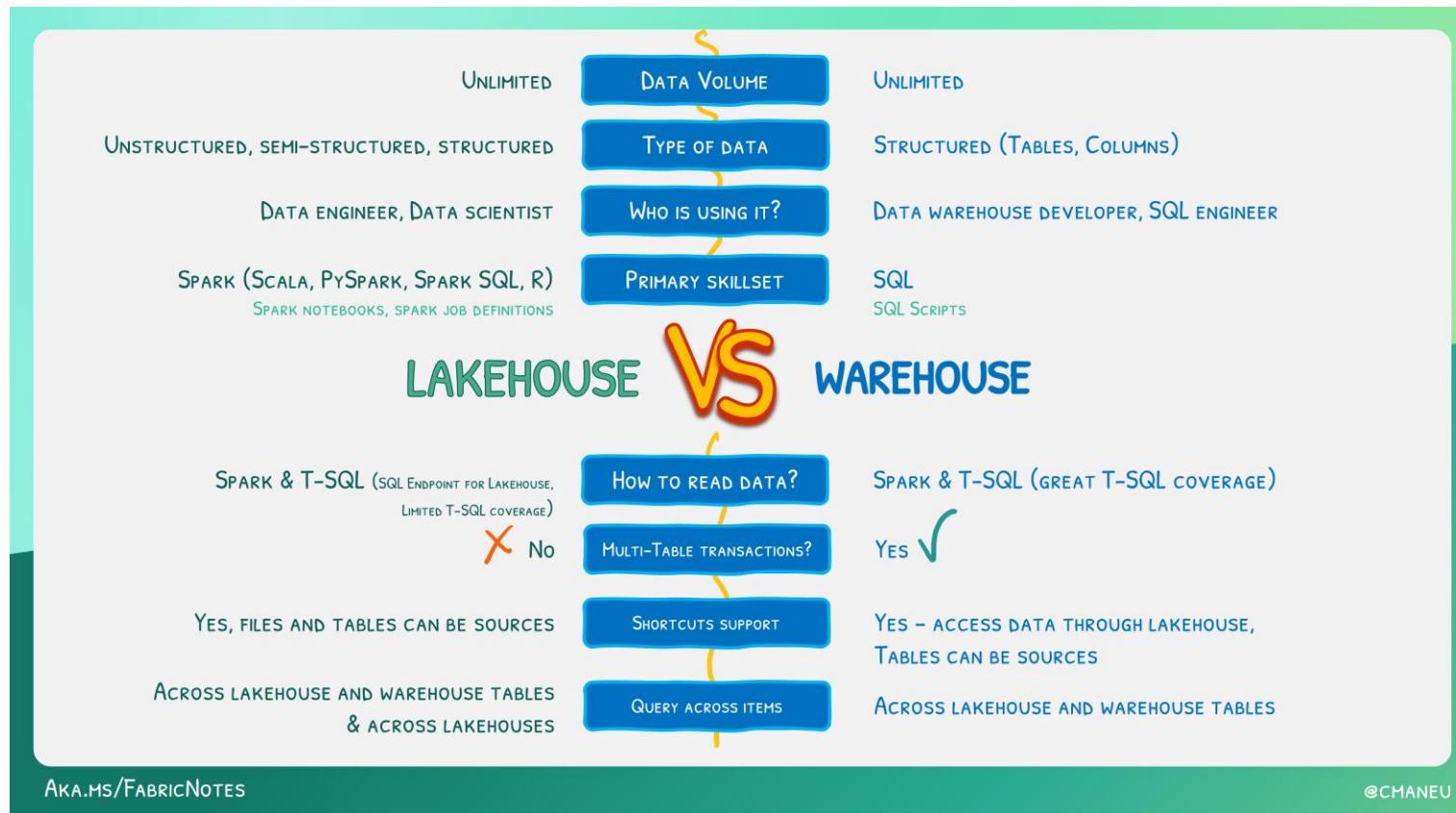


Tale of Two SQL Endpoints in Fabric





Do I Need a Warehouse or Lakehouse?



Credit: Christopher Maneu

Understanding the Workload Perspective



Not only personas/skillset

Spark



Polaris

- ✓ Enhanced version of Synapse Spark
- ✓ No “A” (Atomicity) guaranteed – transactions managed on table level
- ✓ Various languages (Python, R...)
- ✓ Data streaming and Machine learning
- ✓ Extend with OneLake Shortcuts

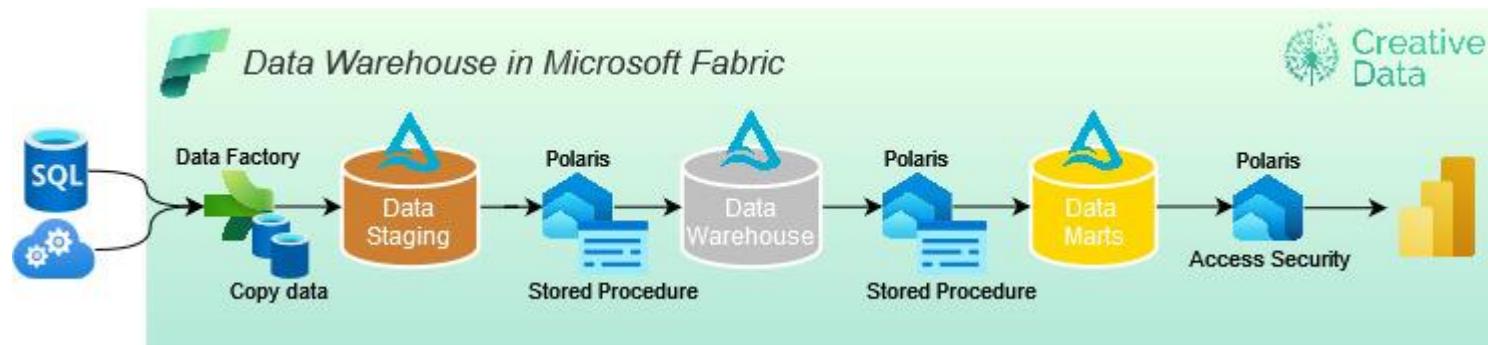
- ✓ Powering Synapse Serverless
- ✓ Now enables data WRITE
- ✓ Provides a “feeling” of RDBMS
- ✓ Perfect fit for batch workloads (SPs)
- ✓ Limited T-SQL functionality

Understanding the Workload Perspective

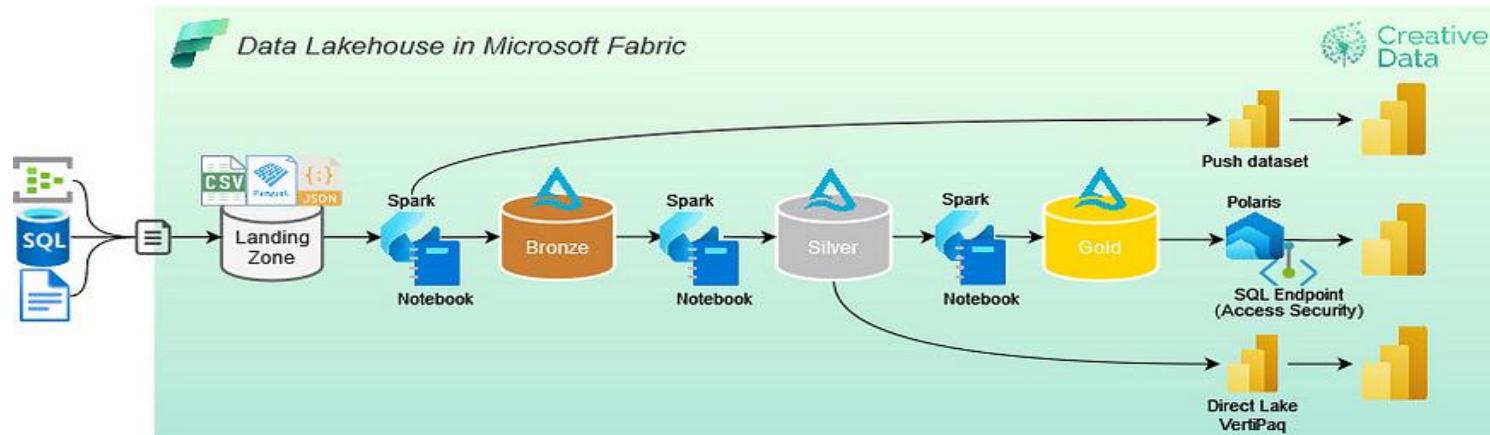
Common scenario



Polaris



Spark



Understanding the Workload Perspective

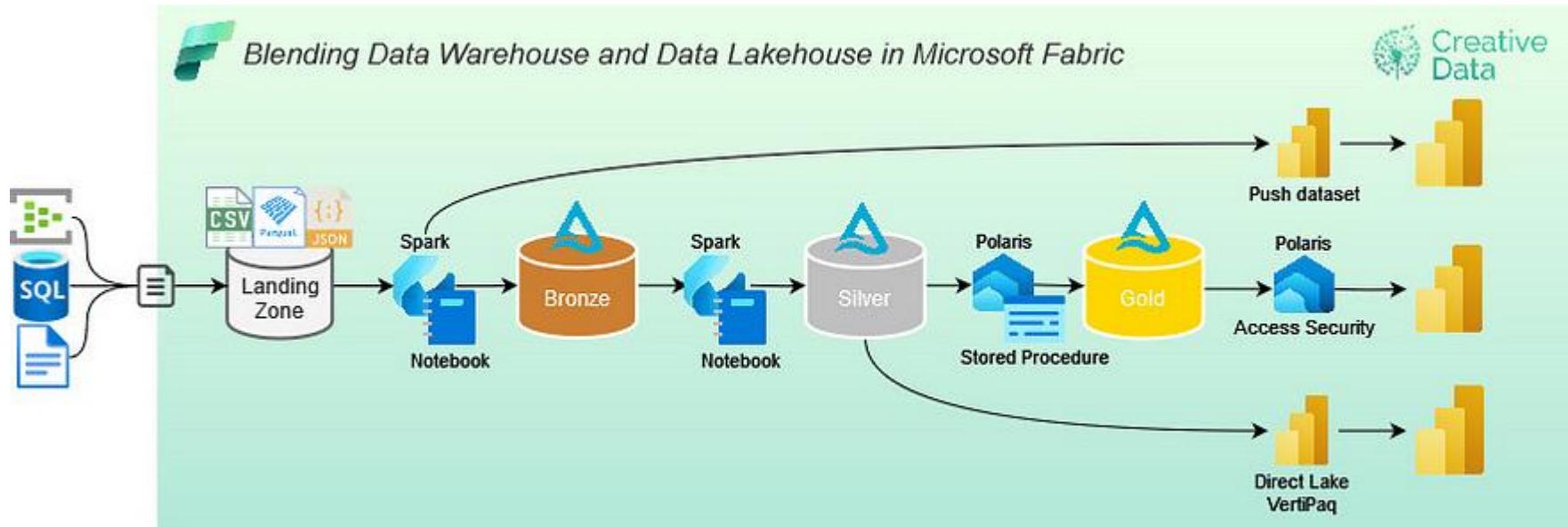
Better together!



Polaris



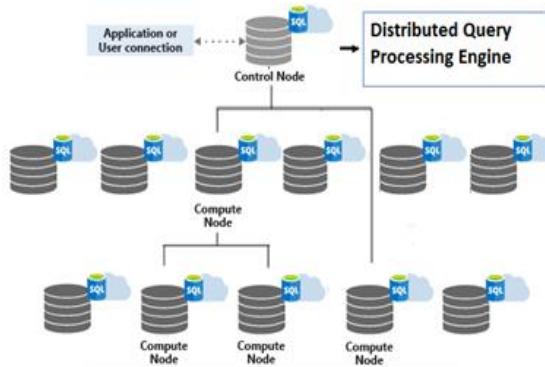
Spark





What Happens Behind the Scenes?

Polaris Engine (Synapse Serverless SQL)



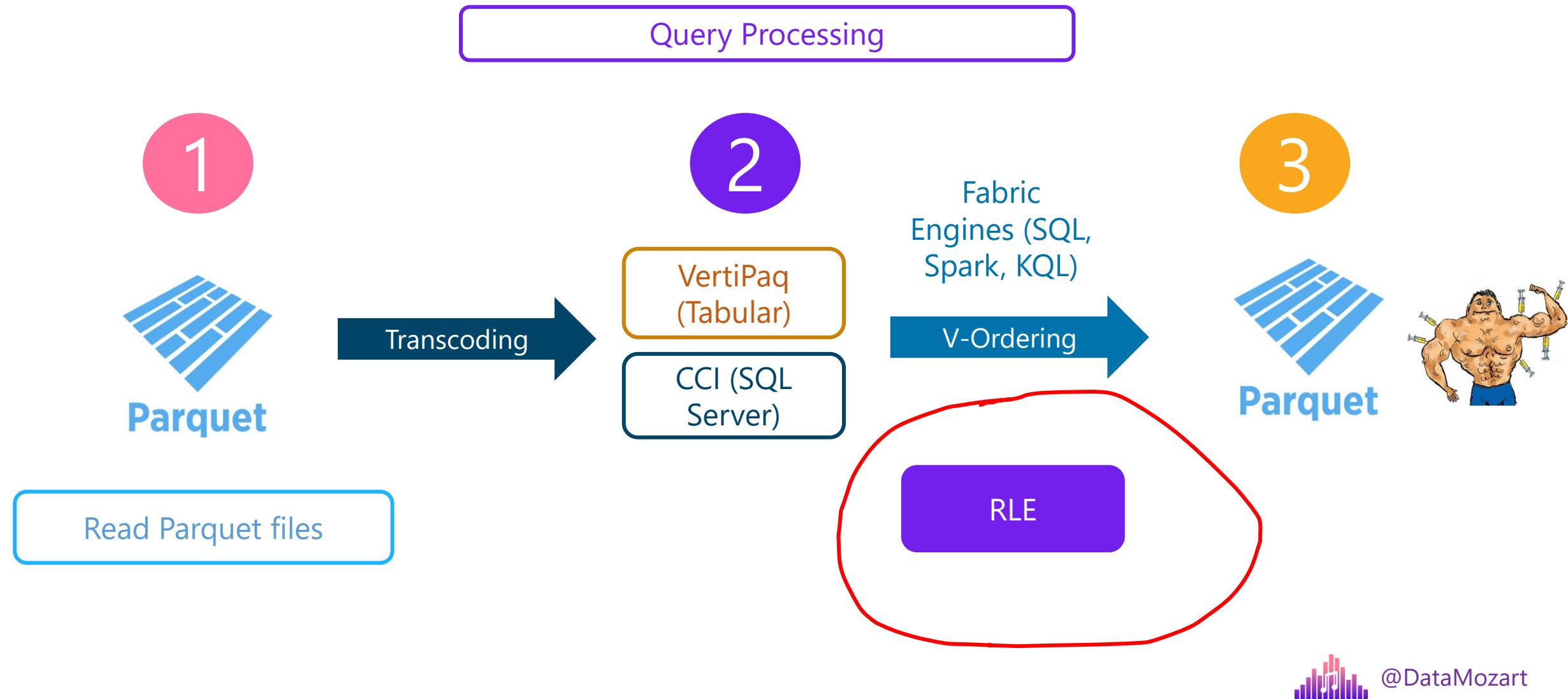
SQL Server Query Optimizer

VertiPaq

[Polaris Engine Whitepaper](#)



What Happens Behind the Scenes?





Organizing the Fabric Lakehouse



Bronze

Raw



Silver

Validated



Gold

Enriched/Curated

Medallion Architecture Fundamentals



Bronze

- ✓ Land data from external sources in its original state
- ✓ Serve as a repository of the historical archive of source data
- ✓ Contains unvalidated data
- ✓ Stores the data in Parquet/Delta

Medallion Architecture Fundamentals



Silver

- ✓ Conformed and cleaned data from the bronze layer
- ✓ Ad-hoc analysis, machine learning workloads
- ✓ Contains enriched and validated data
- ✓ Data model normalized to a 3rd normal form
- ✓ Stores the data in Delta/Parquet

Medallion Architecture Fundamentals



Gold

- ✓ Structured and organized data for specific project requirements
- ✓ Data additionally cleaned and refined
- ✓ Complex business logic and specific calculations
- ✓ Data model is a Kimball-style star schema
- ✓ Stores the data preferably in Delta, alternatively in Parquet

Organizing the Fabric Lakehouse

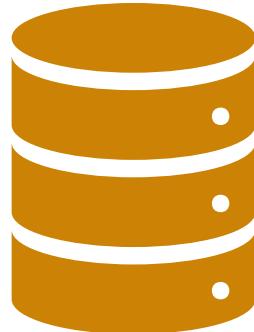
- How much data are you working with?
- How complex are the transformations you need to make?
- How often will you need to move data between layers?
- What tools are you most comfortable with?

 Note Dataflows transform data, pipelines orchestrate data, notebooks can do both

Implement a Medallion Architecture in Fabric

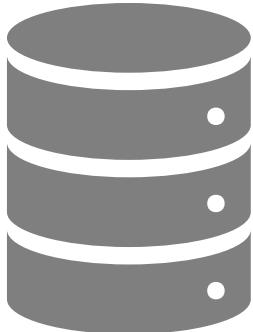
Bronze

Ingest raw data



Silver

Cleanse and validate data



Gold

Additional transformations
and modeling



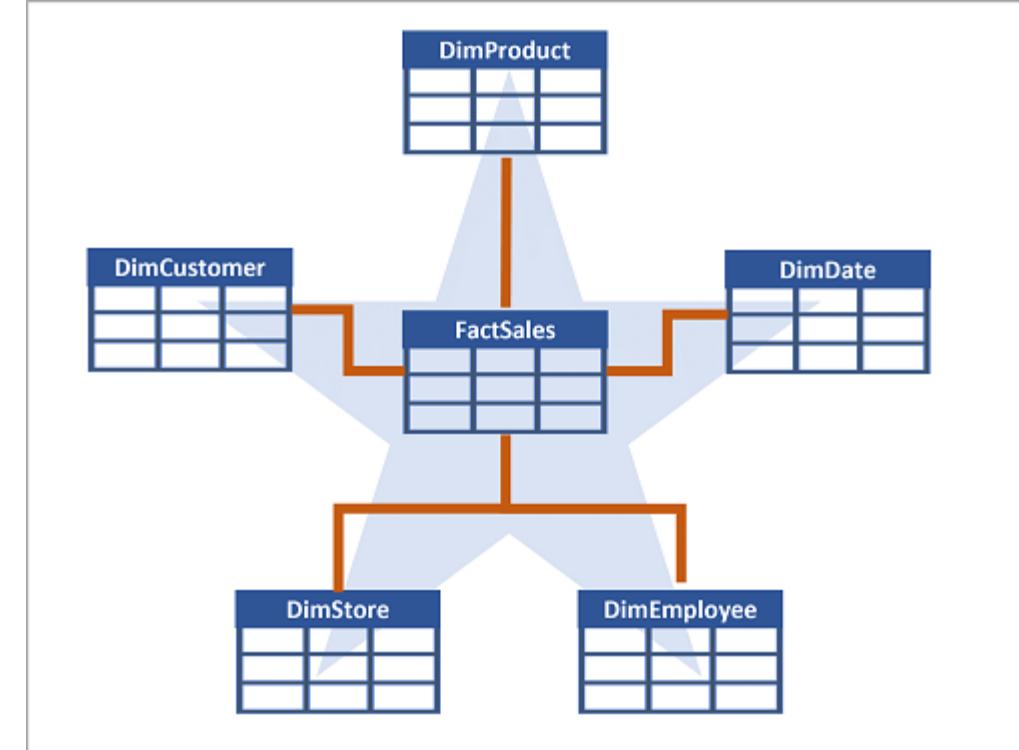
Pipelines, dataflows, notebooks

Dataflows or notebooks

SQL analytics endpoint or
semantic model

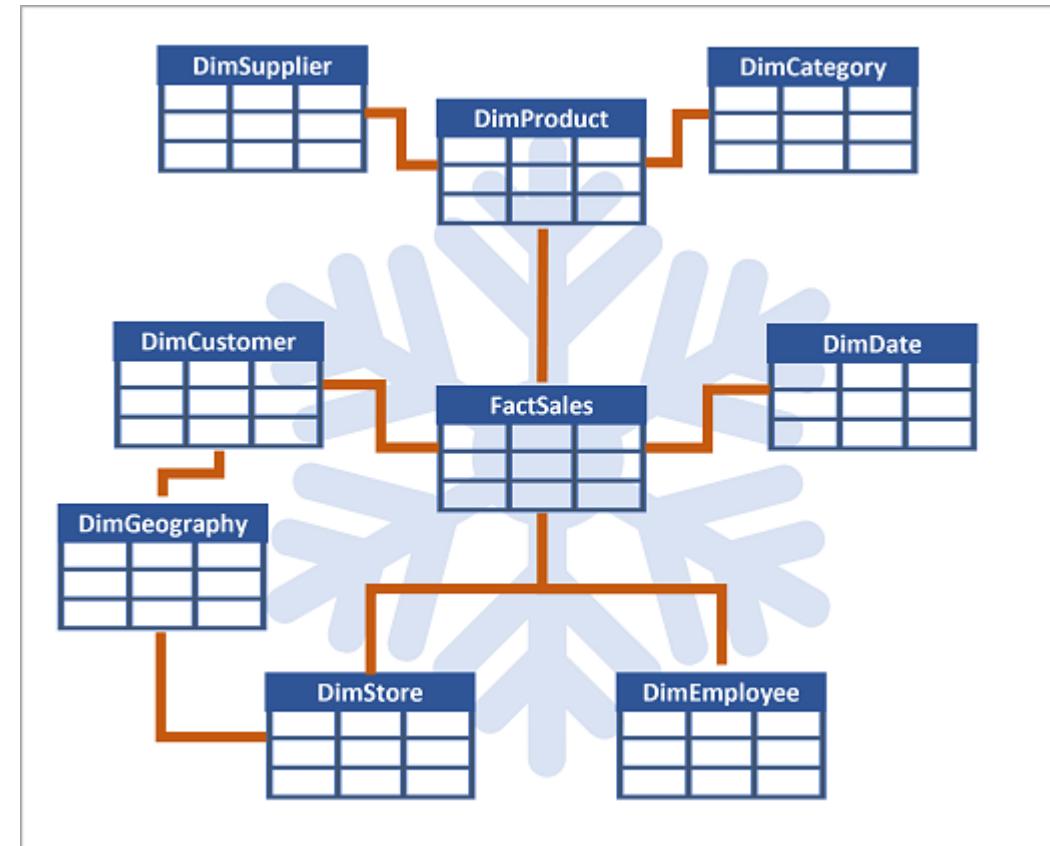
Dimensional Modeling Fundamentals

- Fact tables
- Dimension tables
- Unique keys
 - *Surrogate key*
 - *Alternate key*



Dimensional Modeling Beyond Star Schema

- Star schema, further denormalized
- More granular dimensions



Dimension Tables Extended

Calendar dimension

- Extensive date table
- Ideal for aggregation
- Columns may include:
 - Year
 - Quarter
 - Month
 - Day
 - ...

Slowly changing dimension (SCD)

- Changes to attributes
- Analyze changes over time
- Changes may include:
 - Customer address
 - Product price

“



Lunch Break



”



“

Fabric for Power BI Professionals

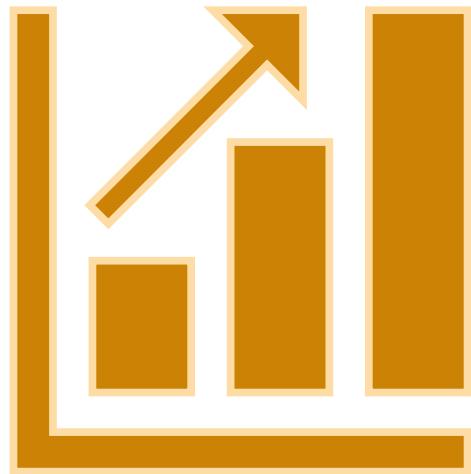
”

I'm a Power BI Professional...

What should I do now?!



Understanding Enterprise-Scale Data



Massive

Complex

Diverse

Rapid

Valuable

Understanding the Importance of Scalability

Flexibility

Accommodate change

Data growth

Handle data volume increase with acceptable report performance

Reduced complexity

Less complex and easier to manage

Designing for Scalability

- ✓ Build with data modeling best practices in mind
- ✓ Determine your needs for Direct Lake/ Import/ Direct Query
- ✓ Enable large semantic model storage feature

Data Modeling Best Practices

Choose the appropriate storage mode

- Import
- DirectQuery
- Direct Lake

Implement data modeling best practices

- “Roche’s Maxim”



Direct Lake for Power BI

Revolutionary feature!

✓ Prerequisites

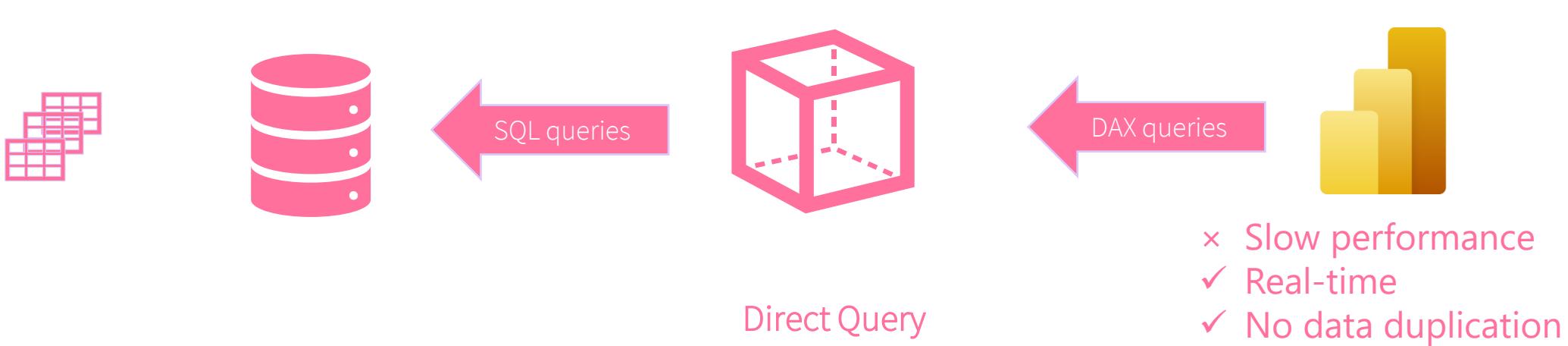
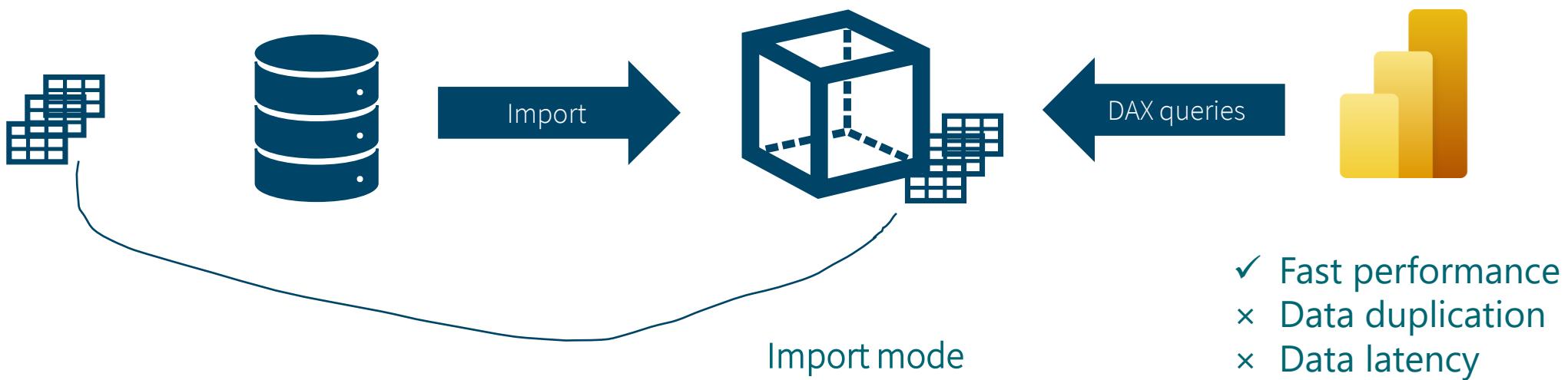
- ✓ Fabric F Capacity/Power BI Premium
- ✓ Lakehouse + SQL Endpoint (for DQ fallback)/Warehouse
- ✓ Delta tables
- ✓ V-Ordering*

* V-Ordering

Fabric-specific way of additionally optimizing Parquet files when writing data



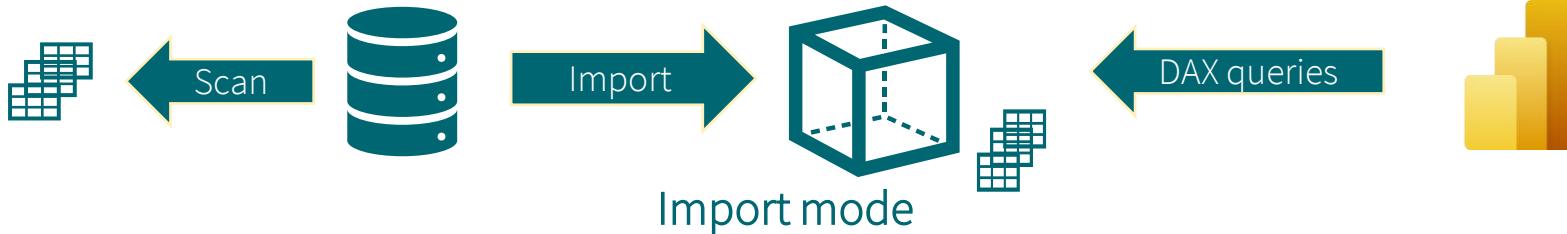
Power BI Architecture – Pre-Fabric



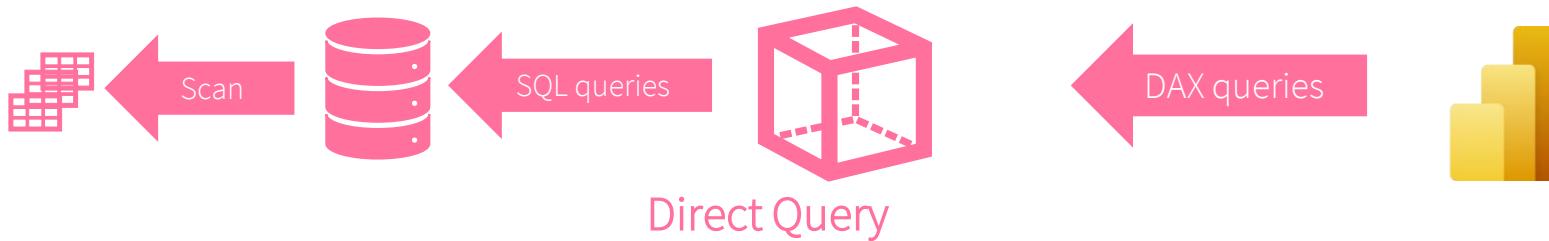


Power BI Architecture – Fabric

- ✓ Fast performance
- ✗ Data duplication
- ✗ Data latency



- ✗ Slow performance
- ✓ Real-time
- ✓ No data duplication



Direct Lake



Default vs Custom Semantic Model



What does the term “semantic model” mean?



Semantic: The data has a meaning to a person or organization

ProductCode

VS001393



Meaning: Refers to a specific product with attributes; *i.e. Nitrogen engines*

Quantity

2

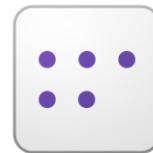


Meaning: Reflects business performance; *i.e. Total units sold*



Model: The objects and calculations represent a real-world process

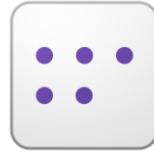
Sales model: Represents the sales process with data; *i.e. Tables, Relationships, DAX measures*



Source: <https://blog.tabulareditor.com/2023/11/17/semantic-models-in-simple-terms/>



Default vs Custom Semantic Model



Default Semantic Model

- ✓ Automatically generated
- ✓ Contains DMV tables and Views
- ✓ Read-only XMLA Endpoint



Custom Semantic Model

- ✓ Can select the necessary tables and views
- ✓ Read-write XMLA endpoint

Limitations of Directlake

- Querying one single Lakehouse or Warehouse
- T-SQL Views are not supported (will fall back to DirectQuery)
- DAX queries exceeding limits or using unsupported features fall back to DirectQuery mode
- No DAX calculated columns/calculated tables
- No composite model
- No DateTime relationships

Always check the list of current limitations!

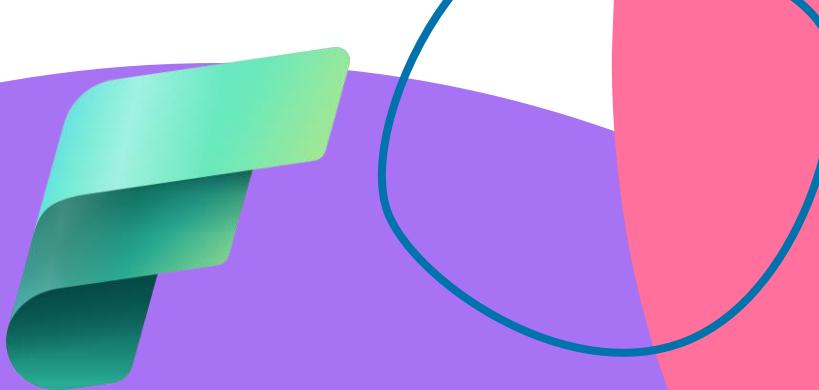
“



Demo Direct Lake in
Power BI

”

“



Fabric Experiences

”

Data Factory



Scale cloud data movement and data transformation

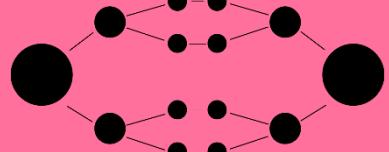


Solve complex ETL



Refresh data on a scheduled basis or on demand

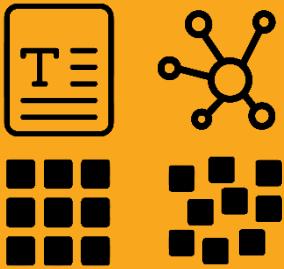
Synapse – Data Engineering



Works as “Divide and Conquer”

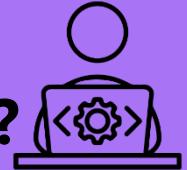


Suitable for large - volume data



Supported Data

(Text data, Graph data, Structured and Unstructured data)



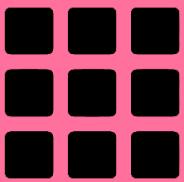
Who?

Data engineer, data scientist

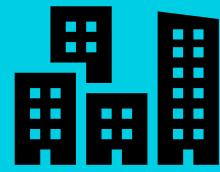
- ✓ Stream processing
- ✓ Big Data Analysis
- ✓ Multi-language support
- ✓ Data Integration and Pipeline



Synapse – Data Warehouse



Support only Structured
Data



Suitable for
Enterprise Data
Warehousing



Who?

DWH developer, SQL
engineer

- ✓ T-SQL based analytics
- ✓ Enterprise Data Warehousing

Real-Time Analytics



Capture, transform, and route real-time events



Run analytical query on raw data as it comes in

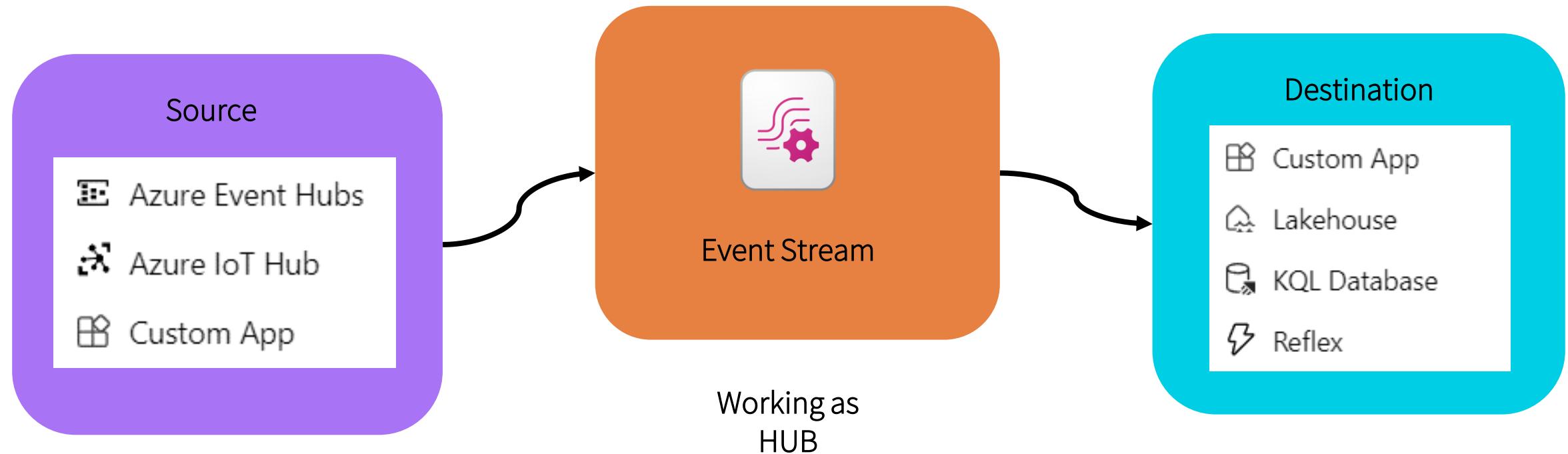


Work with versatile data structures

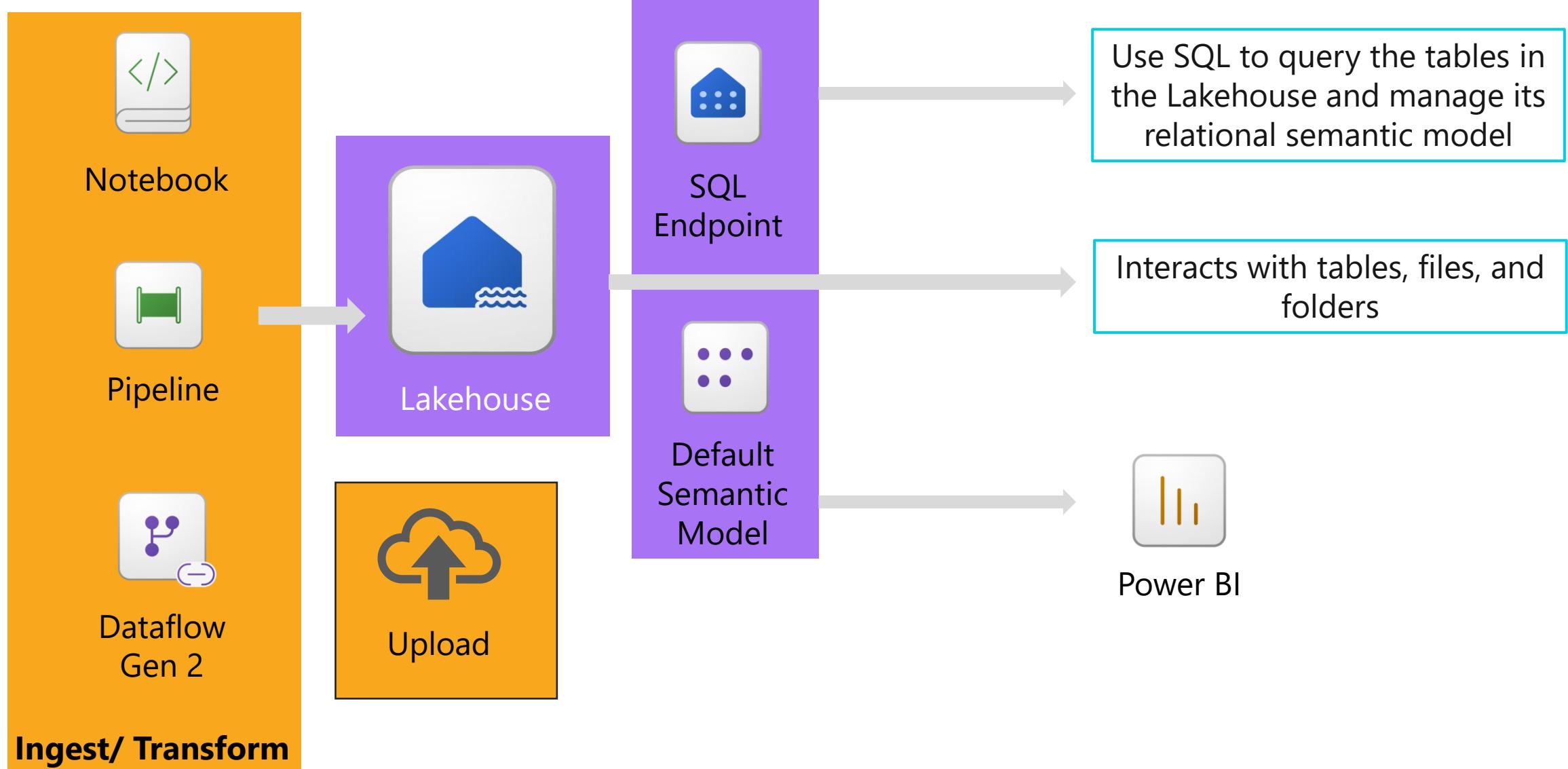


Manage scalable data using Built-in auto-scale

Real-Time Analytics in Microsoft Fabric



Working with Lakehouse



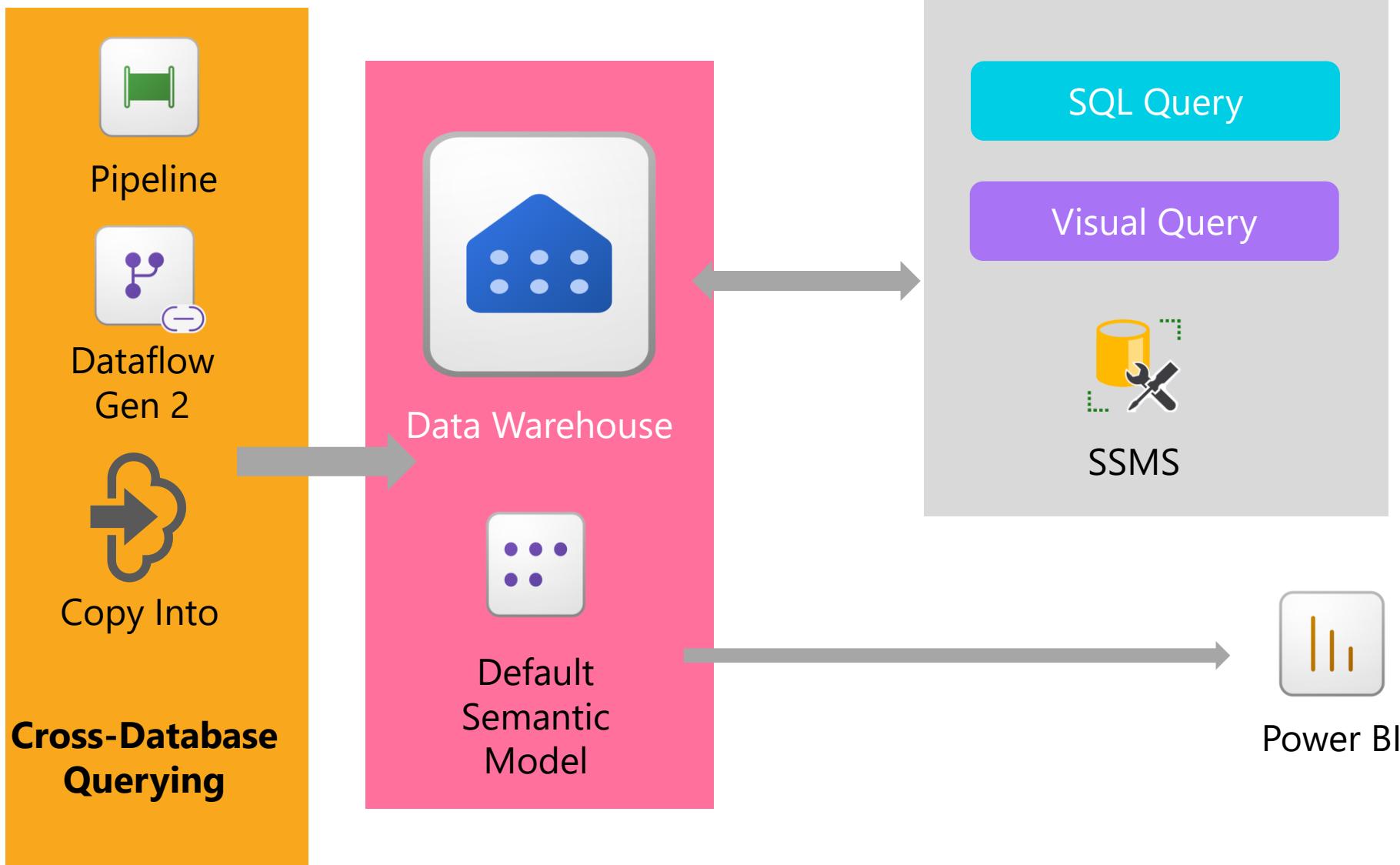
“



Demo – Working
with Lakehouse

”

Working with Warehouse



“



Demo – Working
with Warehouse

”

Data Activator



Monitor data in Power BI reports and Eventstreams items

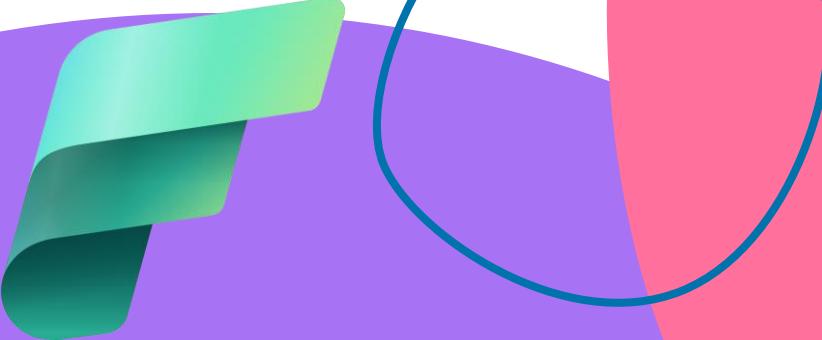


Detect inconsistency in data pattern or condition



Automate taking actions

“



Coffee Break



”



“

Fabric Licensing

”



How Much Will It Cost Me?





How Much Will It Cost Me?

PAYG - Azure

Billed per second, no commitment

SKU	Capacity unit (CU)	Pay-as-you-go (hourly)	Pay-as-you-go (monthly)
F 2	2	\$0.44	\$321.20
F 4	4	\$0.88	\$642.40
F 8	8	\$1.76	\$1,284.80
F 16	16	\$3.52	\$2,569.60
F 32	32	\$7.04	\$5,139.20
F 64	64	\$14.08	\$10,278.40
F 128	128	\$28.16	\$20,556.80
F 256	256	\$56.32	\$41,113.60
F 512	512	\$112.64	\$82,227.20
F 1024	1024	\$225.28	\$164,454.40
F 2048	2048	\$450.56	\$328,908.80

RI - Azure/M365

Billed monthly/yearly, with a monthly commitment

Microsoft Fabric	Microsoft Fabric	Microsoft Fabric
Power BI Premium per capacity SKUs	Fabric Capacity Reservation SKUs	Fabric Capacity pay-as-you-go SKUs
From	From	From
4.675,60 €	4.680,30 €	7.871,90 €
Per month for P1 SKU	Per month for F64 (P1 equivalent)	Per month for F64 SKU (P1 equivalent)
License your organization for access to Microsoft Fabric in a unified product experience that uses the same compute capacity and storage. ³	License your organization for access to Microsoft Fabric in a unified product experience that uses the same compute capacity and storage. ³	License your organization for access to Microsoft Fabric in a unified product experience that uses the same compute capacity and storage. ³
<ul style="list-style-type: none">• Beinhaltet alle Funktionen, die in Power BI Premium pro Benutzer verfügbar sind.• Erhalten Sie über eine einheitliche Produkterfahrung und Kapazität Zugriff auf die übrigen Microsoft Fabric-Workloads.	<ul style="list-style-type: none">• Smaller entry-level compute starting at F2.• Microsoft Azure Consumption Commitment (MACC) eligible.• Gain access to full Microsoft Fabric workloads through a unified product experience and capacity.	<ul style="list-style-type: none">• Small entry-level compute starting at F2.• Microsoft Azure Consumption Commitment (MACC) eligible.• Gain access to full Microsoft Fabric workloads through a unified product experience and capacity.



How Much Will It Cost Me?

PAYG - Azure			
Billed per second, no commitment			
SKU	Capacity unit (CU)	Pay-as-you-go (hourly)	Pay-as-you-go (monthly)
F 2	2	\$0.44	\$321.20
F 4	4	\$0.88	\$642.40
F 8	8	\$1.76	\$1,284.80
F 16	16	\$3.52	\$2,569.60
F 32	32	\$7.04	\$5,139.20
F 64	64	\$14.08	\$10,278.40
F 128	128	\$28.16	\$20,556.80
F 256	256	\$56.32	\$41,113.60
F 512	512	\$112.64	\$82,227.20
F 1024	1024	\$225.28	\$164,454.40
F 2048	2048	\$450.56	\$328,908.80

RI - Azure/M365			
Billed monthly/yearly, with a monthly commitment			
Microsoft Fabric	Microsoft Fabric	Fabric Capacity Reservation SKUs	Fabric Capacity pay-as-you-go SKUs
From	From	From	From
4.675,60 €	4.680,30 €	7.871,90 €	
Per month for P1 SKU	Per month for F16 (P1 equivalent)	Per month for F164 SKU (P1 equivalent)	
<small>License your organization for access to Microsoft Fabric in a unified product experience that uses the same compute capacity and storage.⁷</small>	<small>License your organization for access to Microsoft Fabric in a unified product experience that uses the same compute capacity and storage.⁷</small>	<small>License your organization for access to Microsoft Fabric in a unified product experience that uses the same compute capacity and storage.⁷</small>	
<ul style="list-style-type: none">• Beinhaltet alle Funktionen, die in Power BI Premium pro Benutzer verfügbar sind.• Erhalten Sie über eine einheitliche Produktlizenzierung und Kapazität Zugriff auf die vielfältigen Microsoft Fabric Workloads.	<ul style="list-style-type: none">• Smaller entry-level compute starting at F2.• Microsoft Azure Consumption Commitment (MACC) eligible.• Gain access to full Microsoft Fabric workloads through a unified product experience and capacity.	<ul style="list-style-type: none">• Smaller entry-level compute starting at F2.• Microsoft Azure Consumption Commitment (MACC) eligible.• Gain access to full Microsoft Fabric workloads through a unified product experience and capacity.	



Storage	Price
OneLake Storage / month	\$0.024 per GB
OneLake Cache / month*	\$0.26 per GB

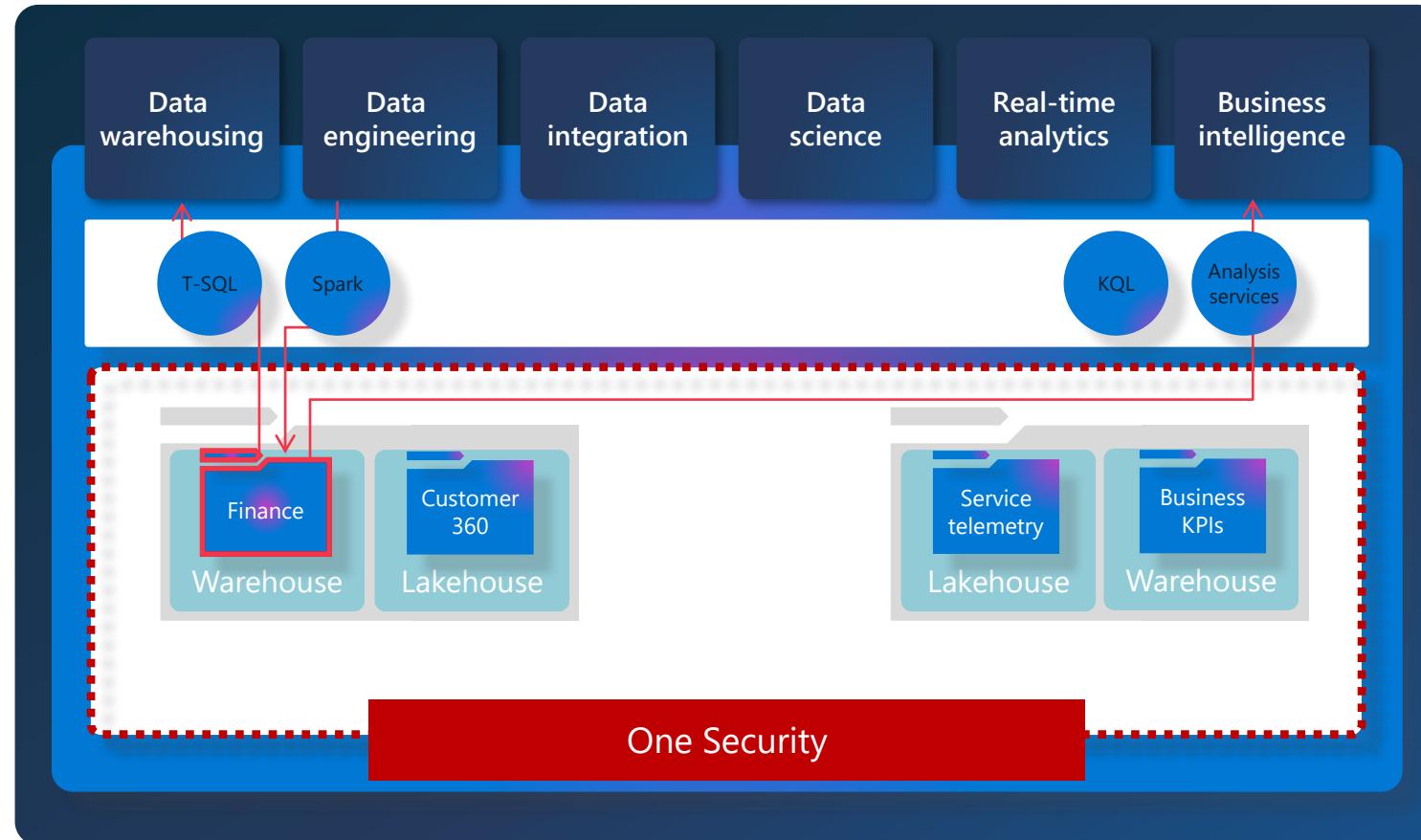
*Only applicable for KQL Database workloads.

Storage costs



One Security (coming soon...)

Secure once and use everywhere



“

Summary – Q&A – Quiz

”



Grande Quiz

Let's play!

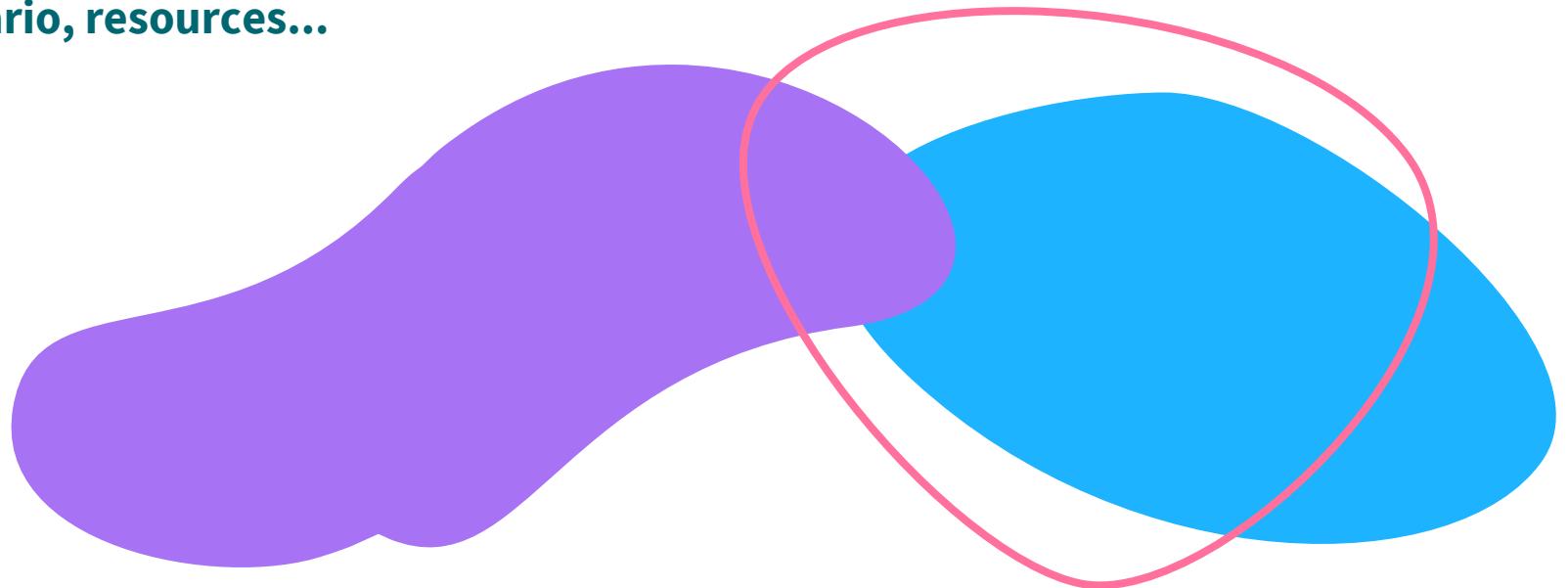
Let's Play!



<https://app.sli.do/event/v187fsiu7tFd3H5Pj1w33>

Wrapping Up...

- **Knowing the options...**
- **No single solution to "rule them all"!**
- **More solutions to more problems**
- **Best fit for your team, scenario, resources...**





Q&A

Catch up later if necessary

References

[Get started with MS Fabric](#)

[DP-600 Exam](#)

[Fabric Career Hub - Learning Rooms](#)

[Fabric.Guru](#)

[DP-600 Data Mozart](#)



Feedback please!



<https://feedback.dsmuc.de/>



Thank you

Nikola Ilic

nikola@data-mozart.com
www.data-mozart.com

