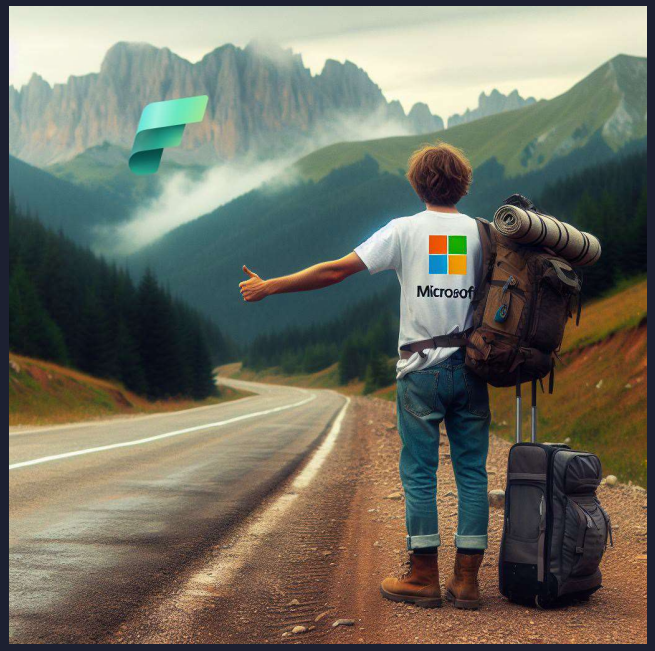
Microsoft Fabric – The Ultimate Guide



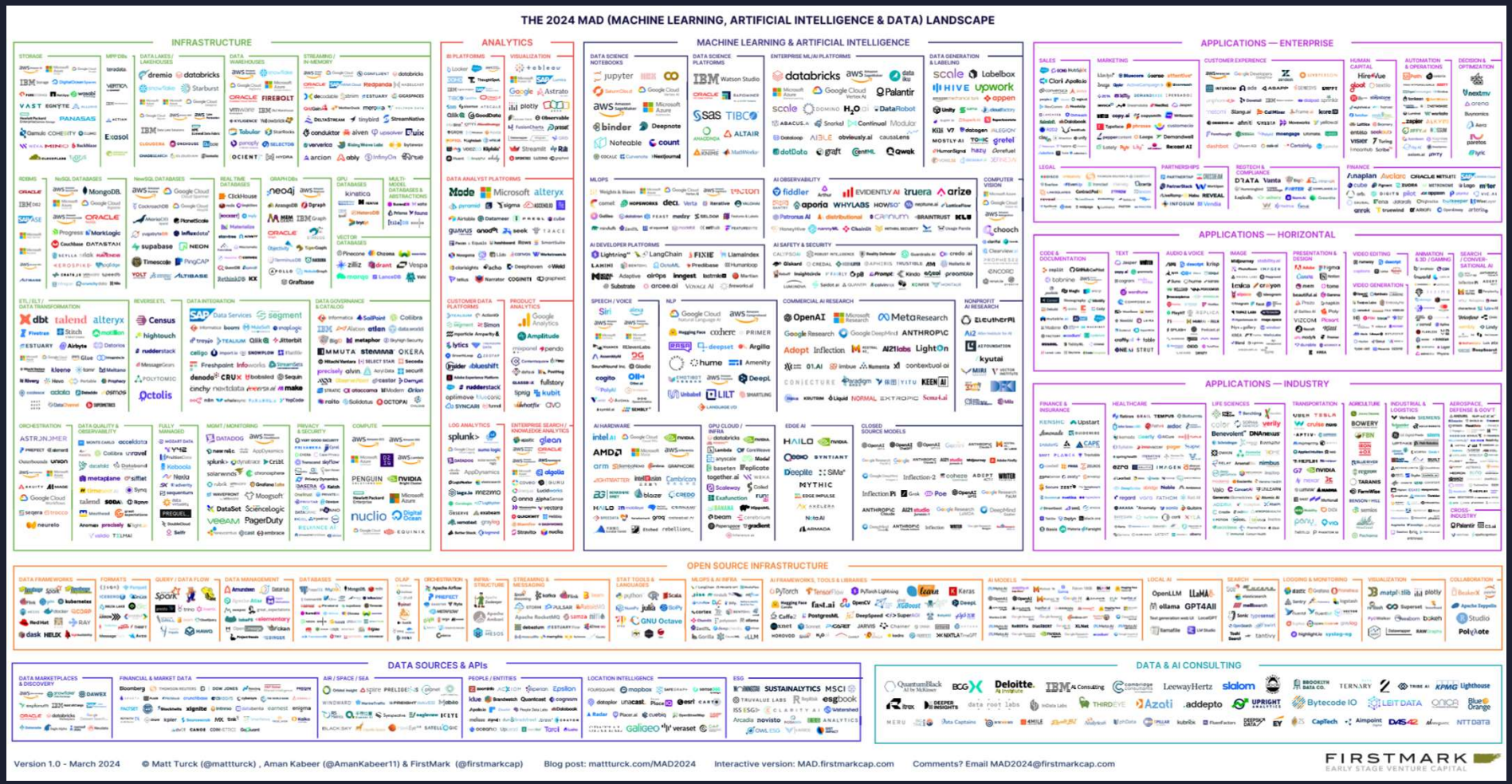
Nikola Ilic

Data Mozart, Microsoft Data Platform MVP

What are we covering?

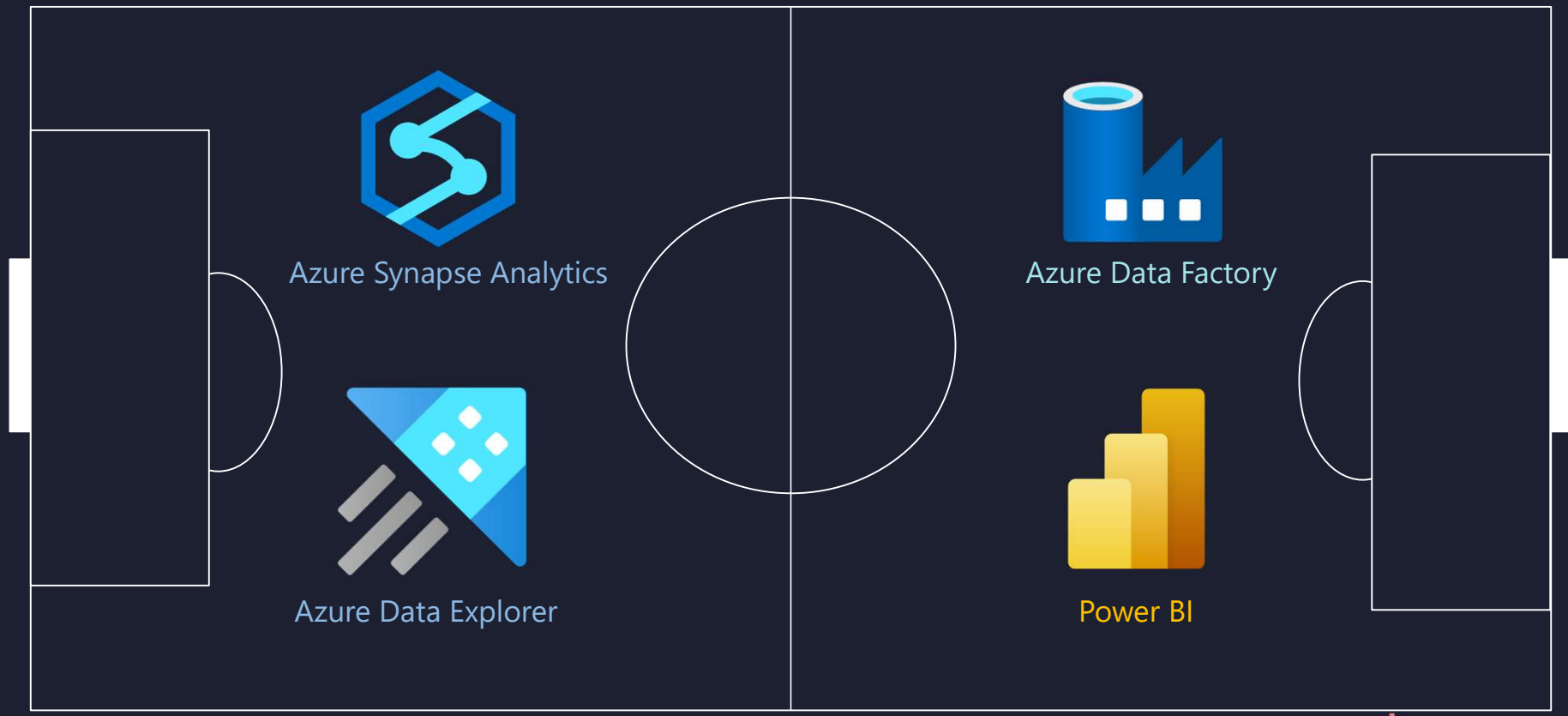
1. What is Microsoft Fabric?
2. What makes Microsoft Fabric unique?
3. Microsoft Fabric core components
4. Licensing & Next Steps

The 2024 ML, AI, and Data Landscape



Credit: Matt Turck

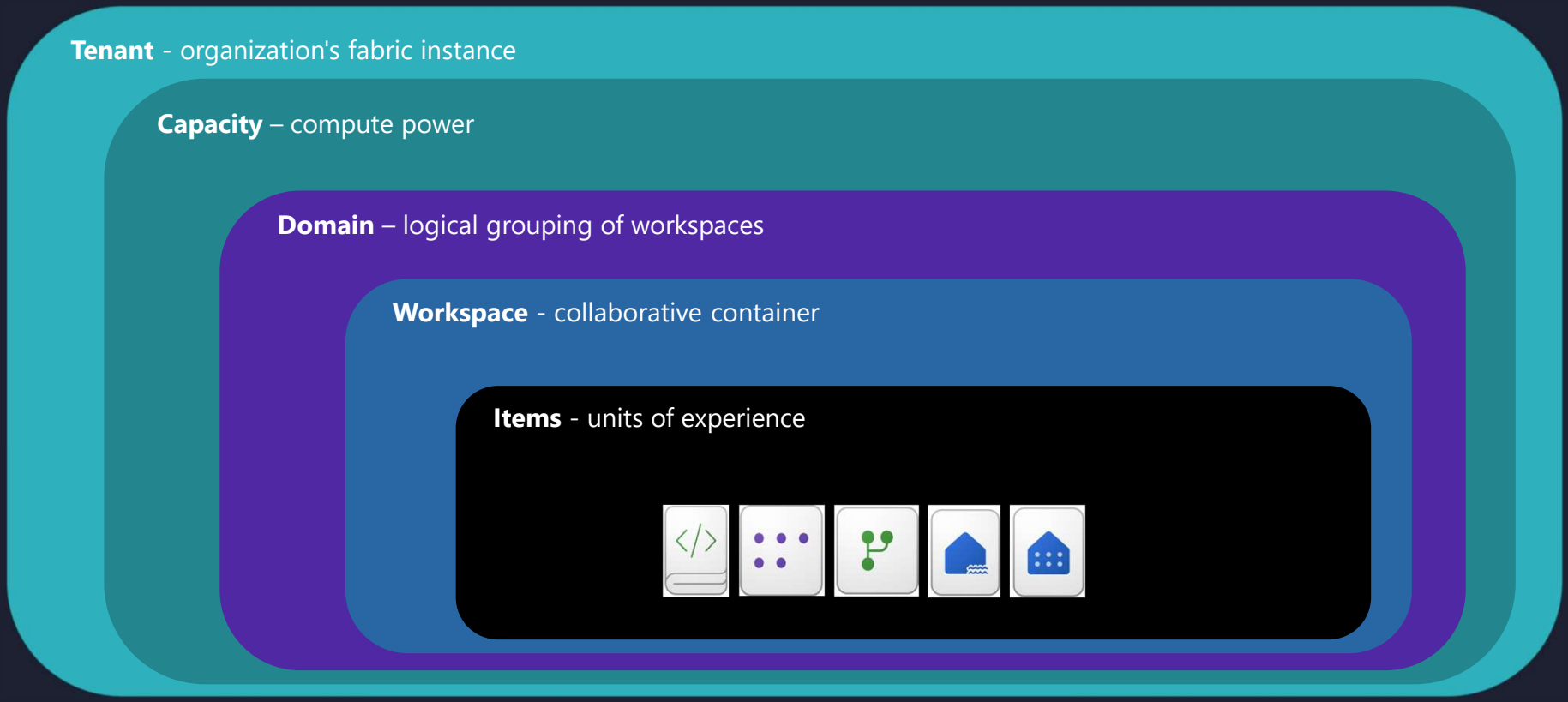
What is a Microsoft Fabric?



Microsoft Fabric Key Pillars

1. “Everything-as-a-service”
2. Centralized administration
3. Lake, lake, lake…

Microsoft Fabric Terminology



Players in Microsoft Fabric

Data Factory

Real-Time Intelligence

Databases

Analytics

Industry Solutions

Power BI

Partner solutions

Copilot in Fabric

OneLake

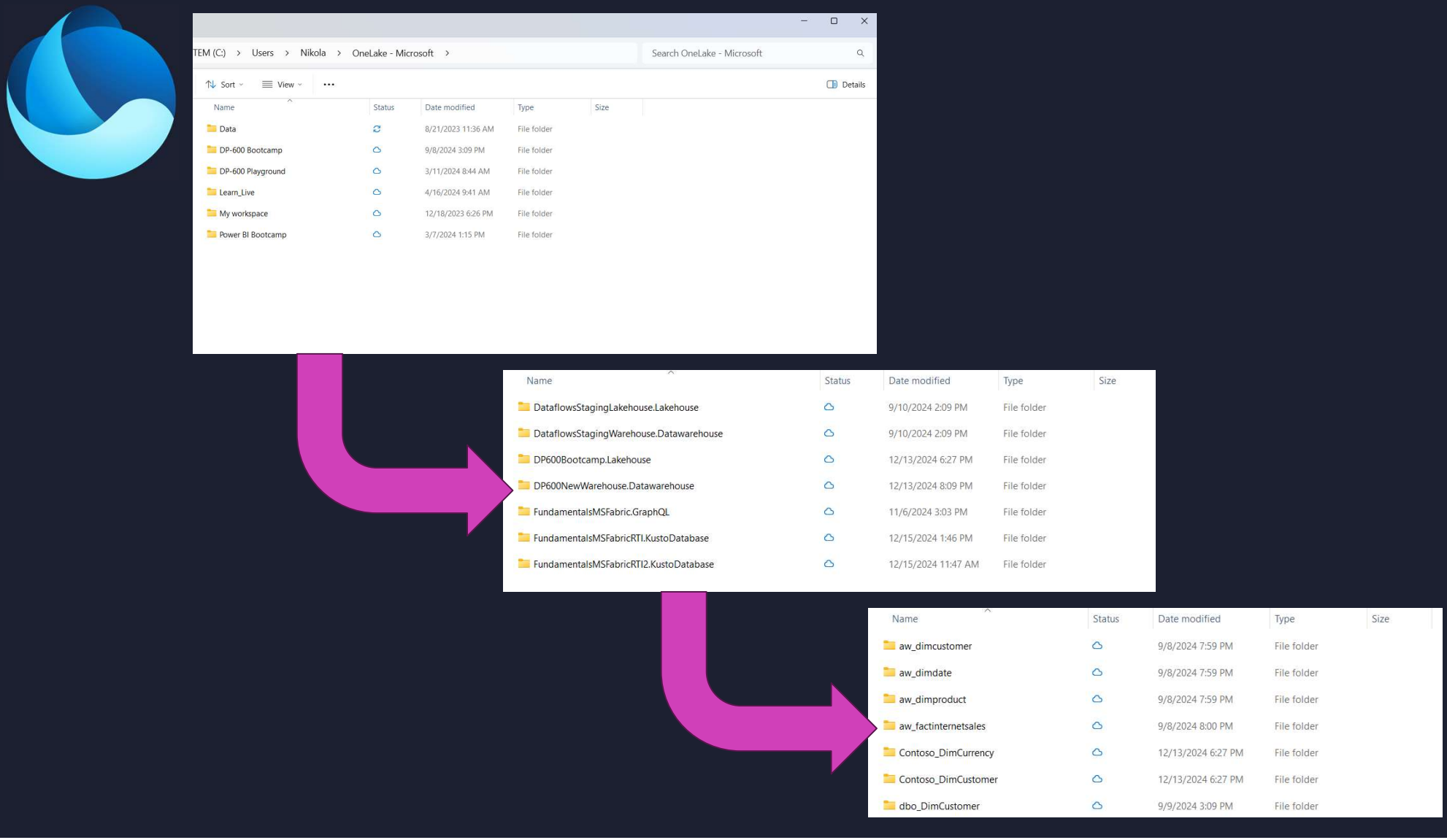
Microsoft Purview

Core Components

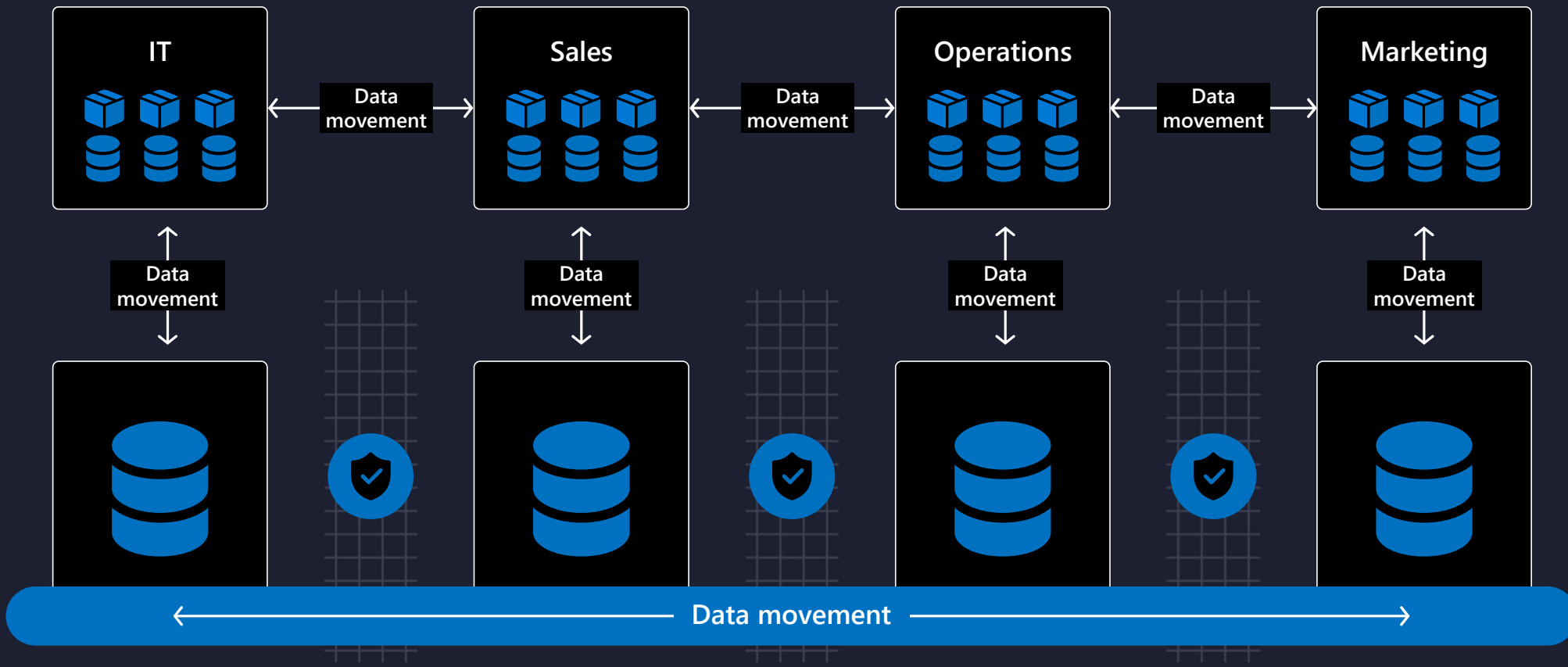
OneLake

Central storage repository for the entire organization!

1 Fabric tenant = 1 OneLake

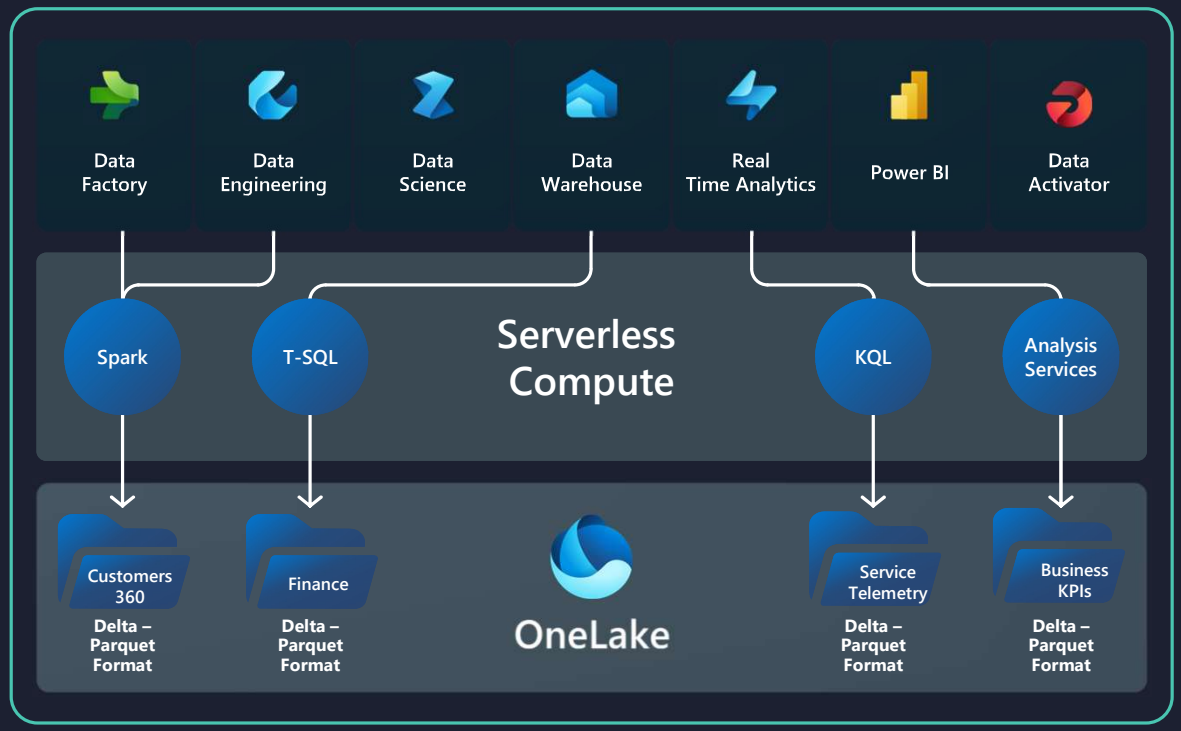

TEM (C:) › Users › Nikola › OneLake - Microsoft ›
Search OneLake - Microsoft
Sort
View
Details
Name
Status
Date modified
Type
Size
Data
8/21/2023 11:36 AM
File folder
DP-600 Bootcamp
9/8/2024 3:09 PM
File folder
DP-600 Playground
3/11/2024 8:44 AM
File folder
Learn_Live
4/16/2024 9:41 AM
File folder
My workspace
12/18/2023 6:26 PM
File folder
Power BI Bootcamp
3/7/2024 1:15 PM
File folder
Name
Status
Date modified
Type
Size
DataflowsStagingLakehouse.Lakehouse
9/10/2024 2:09 PM
File folder
DataflowsStagingWarehouse.Datawarehouse
9/10/2024 2:09 PM
File folder
DP600Bootcamp.Lakehouse
12/13/2024 6:27 PM
File folder
DP600NewWarehouse.Datawarehouse
12/13/2024 8:09 PM
File folder
FundamentalsMSFabric.GraphQL
11/6/2024 3:03 PM
File folder
FundamentalsMSFabricRTI.KustoDatabase
12/15/2024 1:46 PM
File folder
FundamentalsMSFabricRTI2.KustoDatabase
12/15/2024 11:47 AM
File folder
Name
Status
Date modified
Type
Size
aw_dimcustomer
9/8/2024 7:59 PM
File folder
aw_dimdate
9/8/2024 7:59 PM
File folder
aw_dimproduct
9/8/2024 7:59 PM
File folder
aw_factinternetsales
9/8/2024 8:00 PM
File folder
Contoso_DimCurrency
12/13/2024 6:27 PM
File folder
Contoso_DimCustomer
12/13/2024 6:27 PM
File folder
dbo_DimCustomer
9/9/2024 3:09 PM
File folder

Multiple siloed lakes with lots of duplication



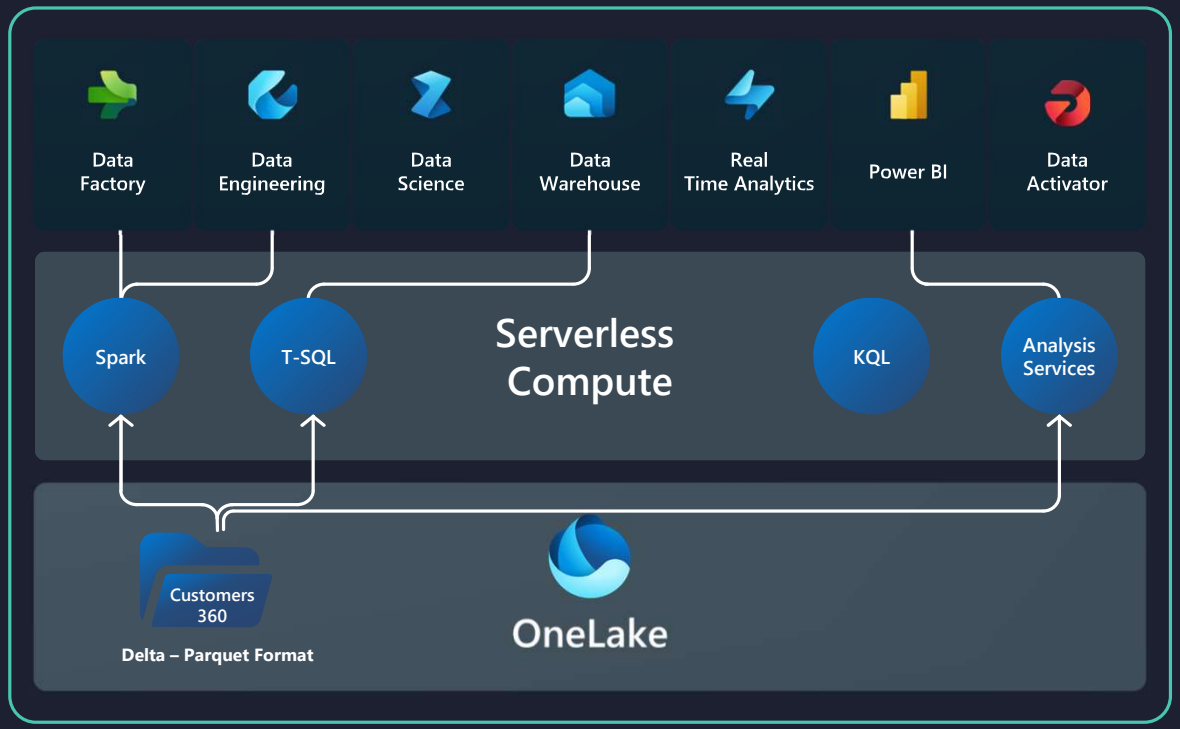
One Copy for all computes

Real separation of compute and storage

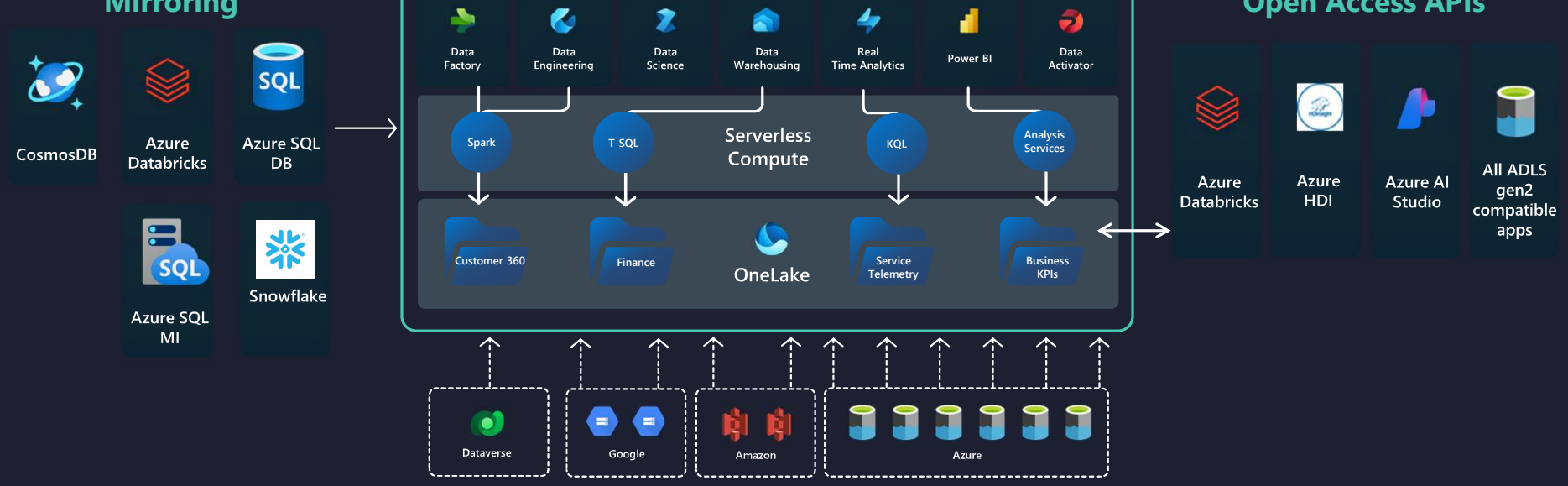


One Copy for all computes

One copy of data can be read by all engines



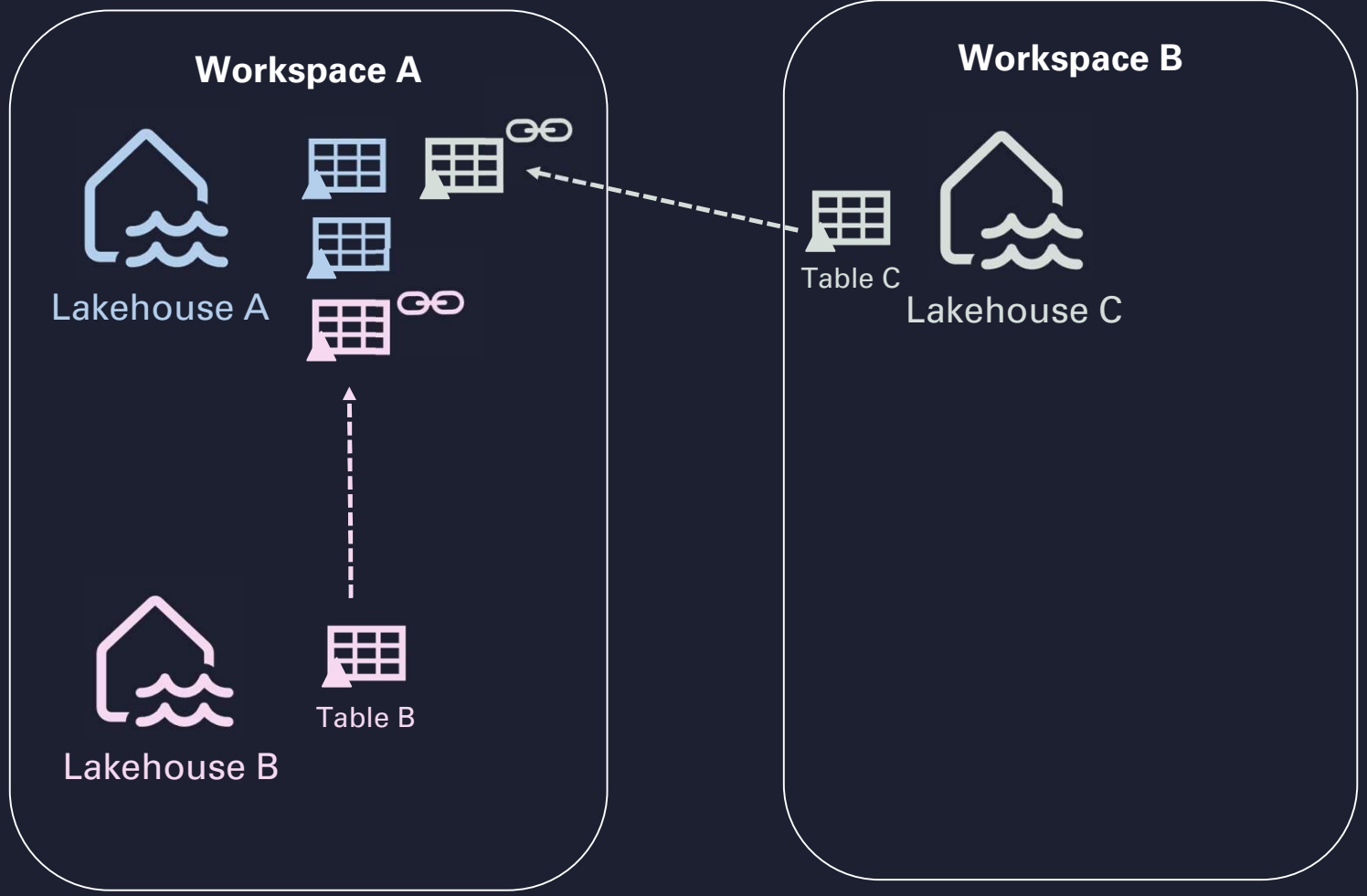
All roads lead to... ~~Rome~~ OneLake



Shortcuts

Shortcuts –
Like Windows Explorer?

Shortcuts



Core Components

Data Factory

Real-Time Intelligence

Databases

Analytics

Industry Solutions

Power BI

Partner solutions

Copilot in Fabric

OneLake

Microsoft Purview

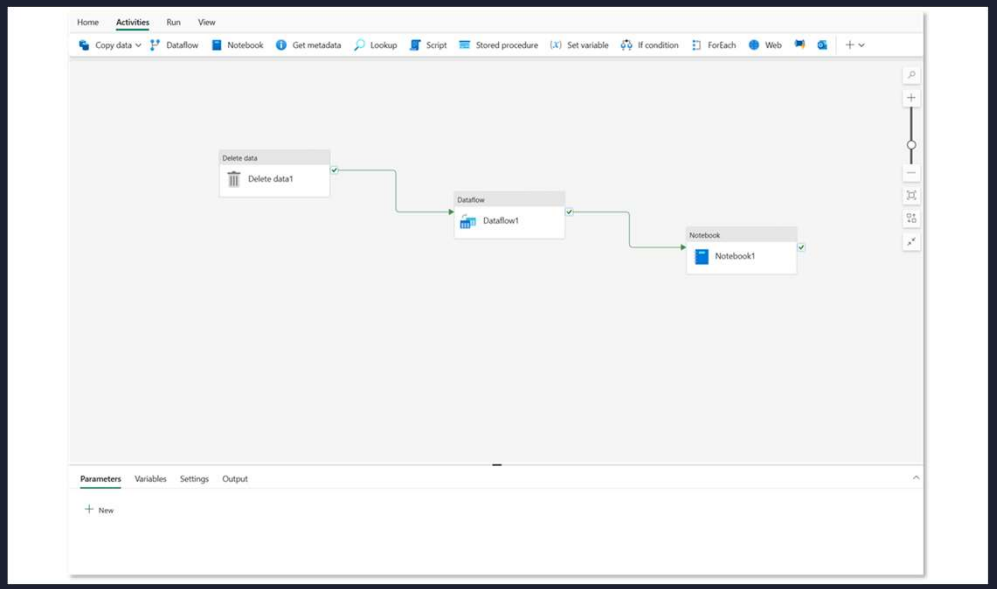
Core Components

Data Factory

Pipelines

Dataflows Gen2

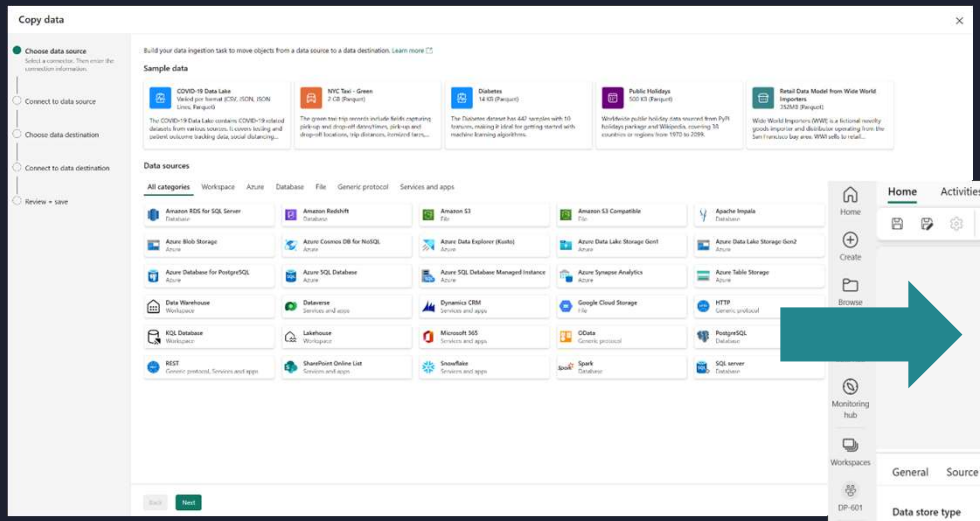
Pipelines



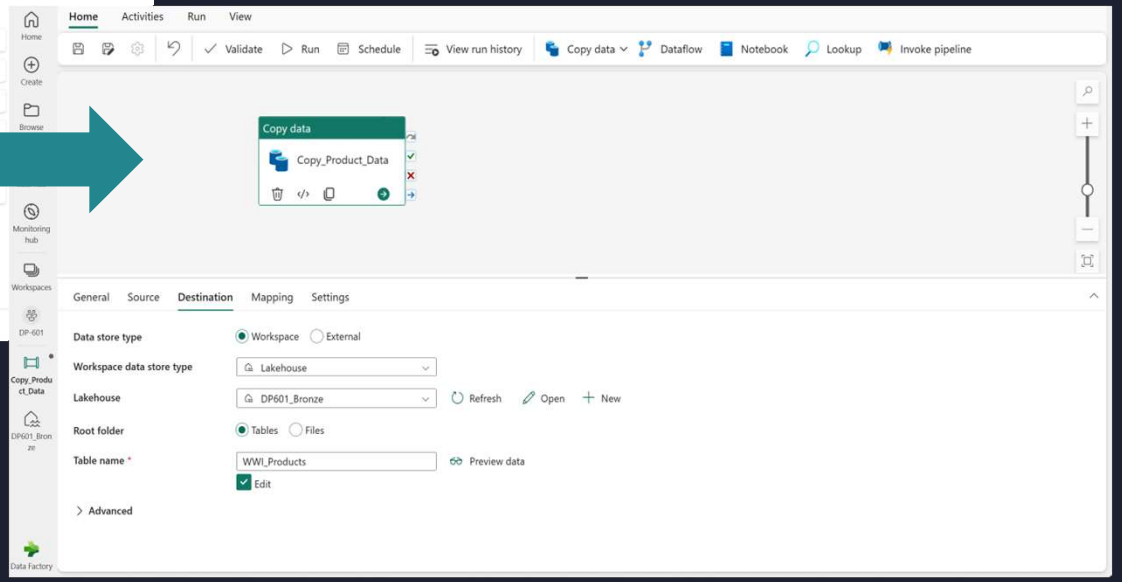
Pipeline concepts:

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

Common Activities – Copy Data



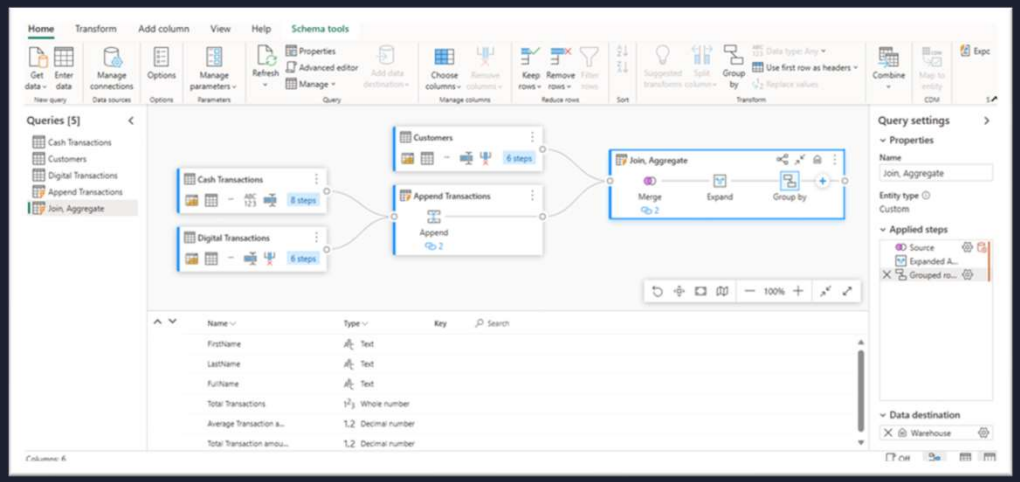
1. Use the copy data tool



2. Edit the settings below the pipeline canvas

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 vs Pipeline

- ✓ Data transformations
- ✓ Data profiling
- ✓ Familiar Power Query experience
- ✓ Many connectors (150+)

- ✓ Orchestration
- ✓ Copy data = simple transformation
- ✓ DF G2 can be part of the orchestration process

Core Components

Data Factory

Real-Time Intelligence

Databases

Analytics

Industry Solutions

Power BI

Partner solutions

Copilot in Fabric

OneLake

Microsoft Purview

Core Components

Real-Time Intelligence


Eventhouse


KQL database


KQL queryset

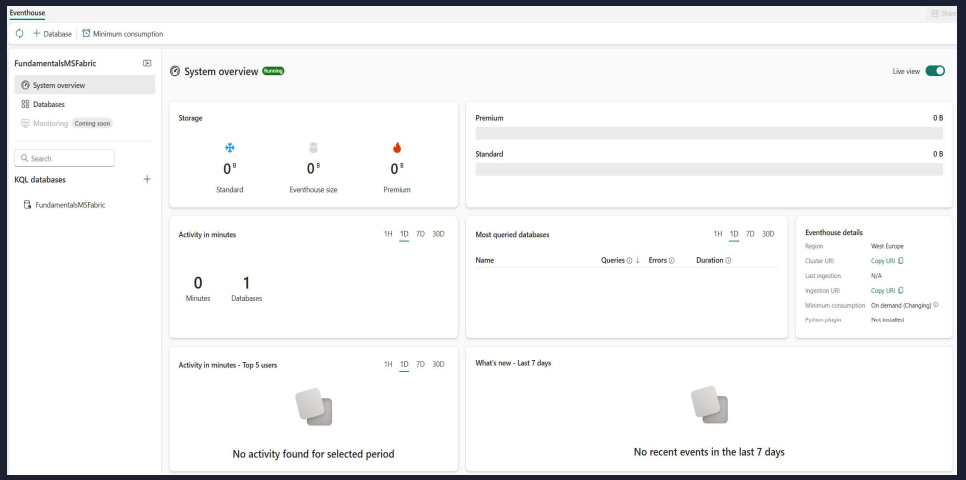

Eventstream


Activator

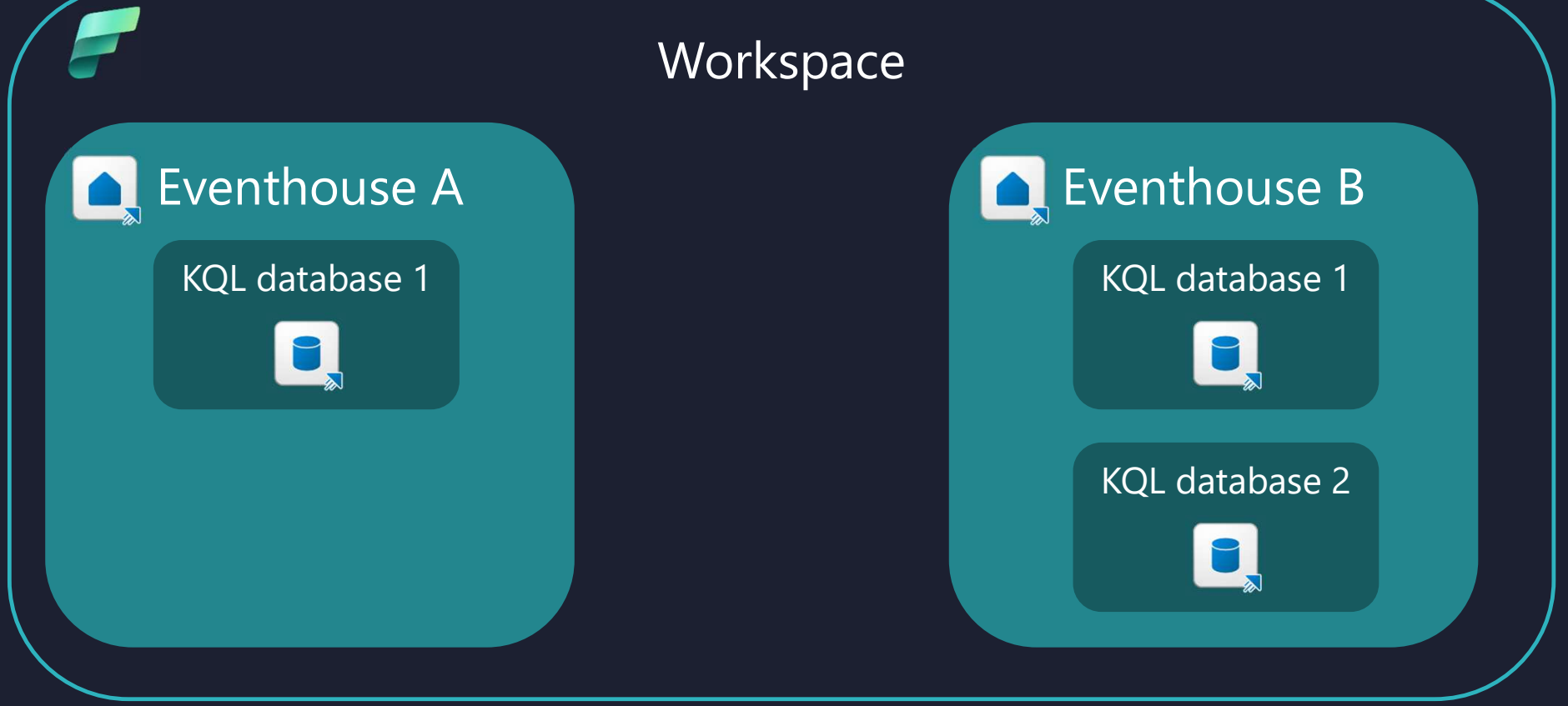
Handling streaming data!

Eventhouse

- Container for KQL databases
- Unified monitoring and management across all databases
- Data automatically indexed and partitioned

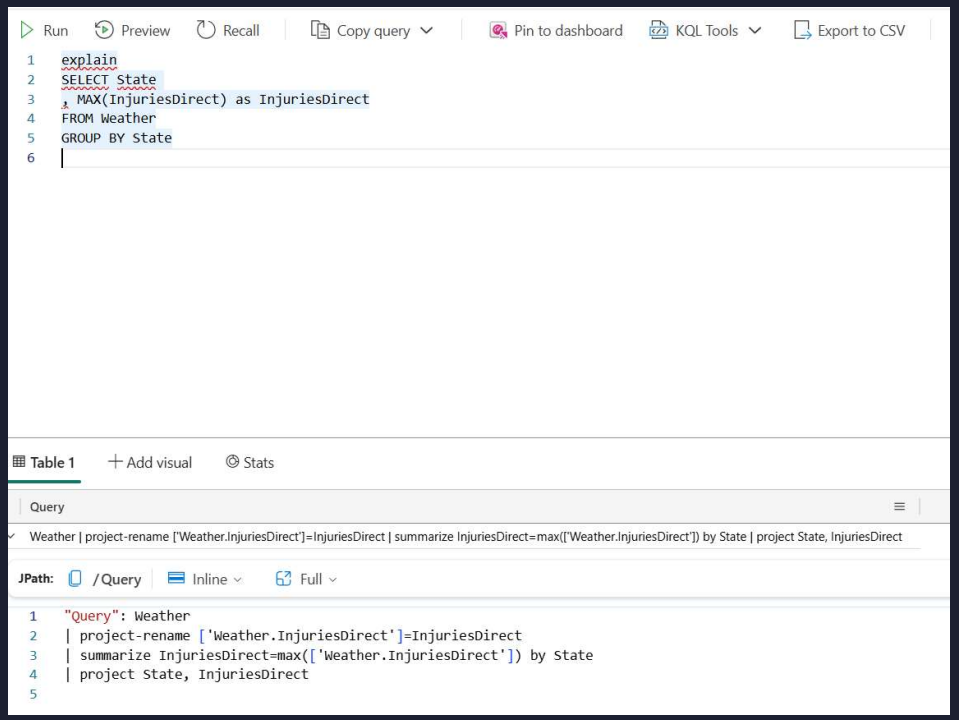


KQL Database



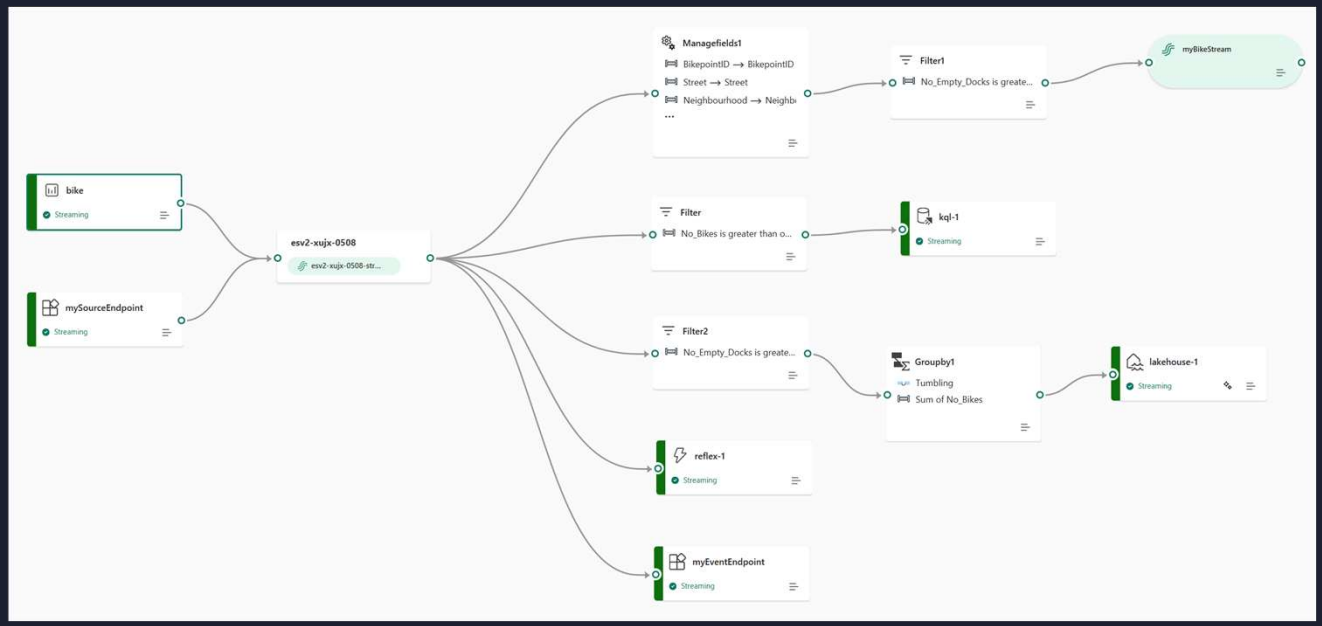
KQL Queryset

- Collection of one or more KQL queries
- Sharing queries with others
- *explain* -> converts SQL to KQL
- You can also write T-SQL! But…



Eventstream

- Bring real-time events to Fabric
- No-code transformations (filter, manage fields, aggregate, group by, join…)



Eventstream

Sources

- Azure Event Hubs
- Azure IoT Hub
- CDC for most DBs
- Apache Kafka
- Azure Blob Storage events
- Fabric Workspace Item events (creating, updating, deleting Fabric item)
- Fabric OneLake events (changes in files and folders)
- Fabric Job events

Destinations

- Custom endpoint
- Eventhouse
- Lakehouse
- Derived stream
- Fabric Activator

Activator

- No-code experience for automatically taking actions when conditions are met
- Monitors data in Power BI reports and eventstreams
- Triggering actions -> Alerting users or triggering Power Automate flows

Events

- **State of the object**
  - Object identifier, timestamp, and values to monitor

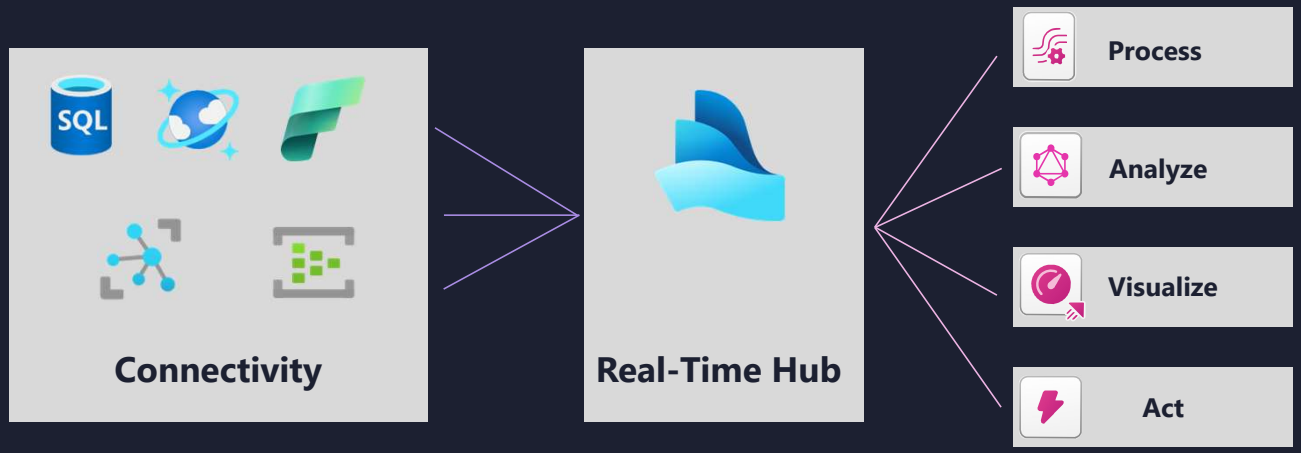
Objects

- **Physical** (freezers, vehicles, packages…)
- **Abstract** (campaigns, user sessions, accounts…)

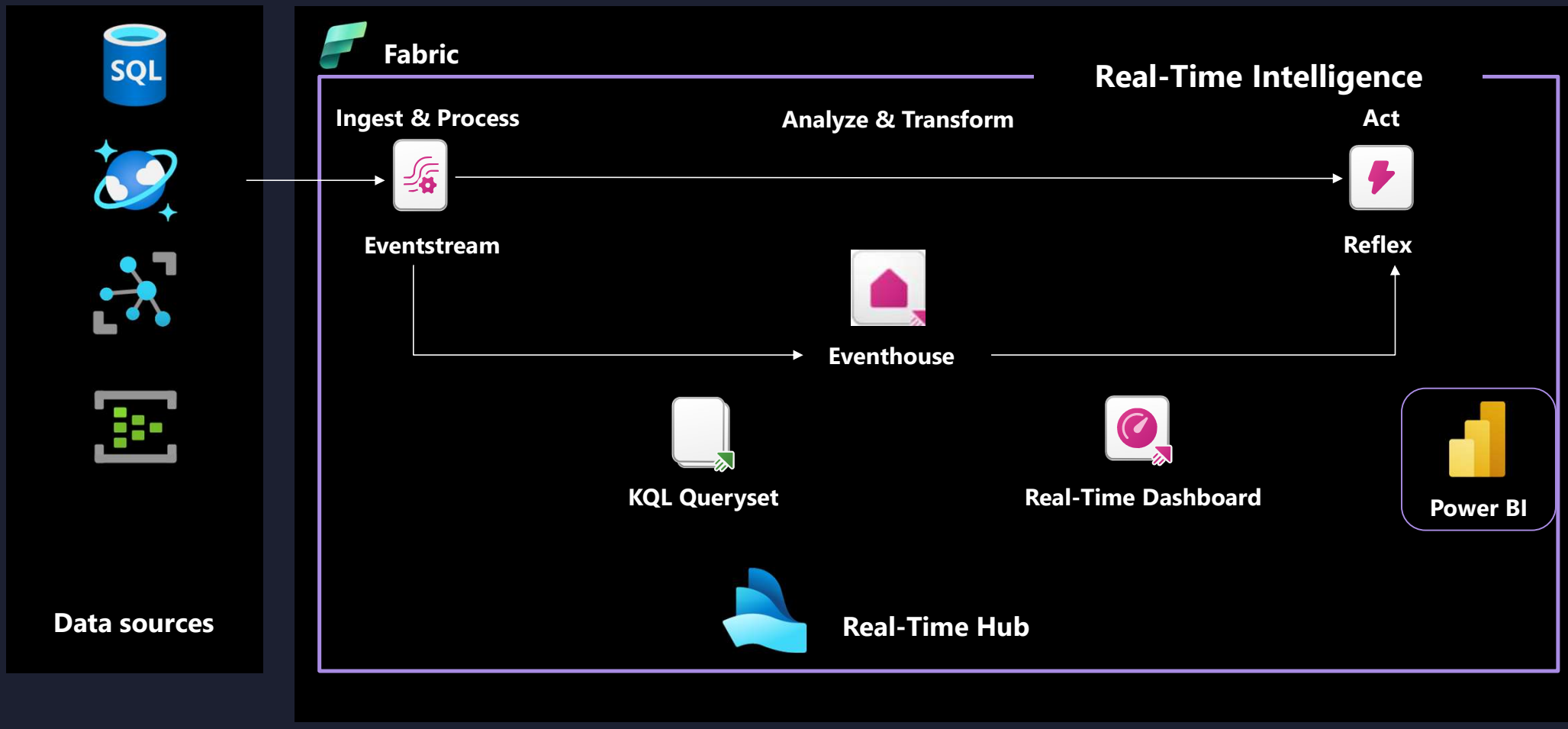
Rules

- **Conditions to detect on objects and actions to take when conditions are met**

Real-time Intelligence



Real-time Intelligence



Core Components

Data Factory

Real-Time Intelligence

Databases

Analytics

Industry Solutions

Power BI

Partner solutions

Copilot in Fabric

OneLake

Microsoft Purview

Core Components

Databases

SQL database

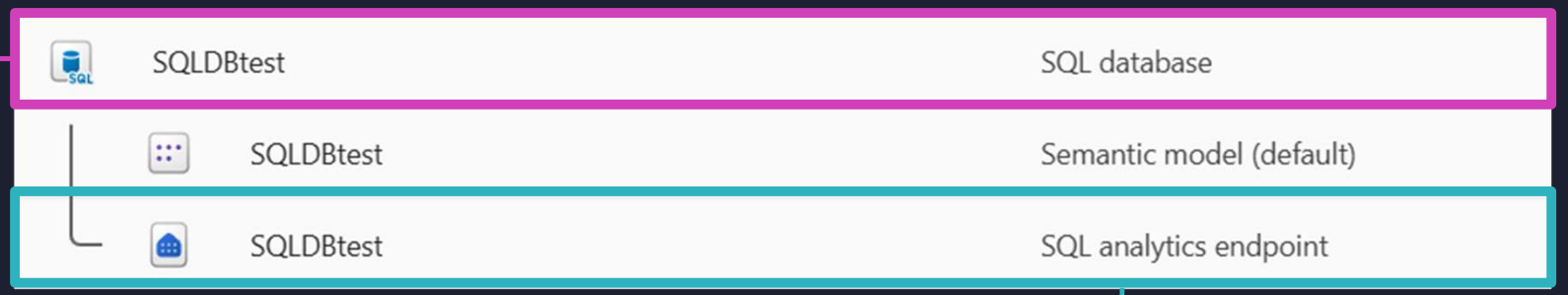
SQL Database in Fabric

SaaS operational database

SQL Database in Fabric

IaaS	PaaS	PaaS	SaaS
			
SQL Server on Azure VM	Azure SQL managed instance	Azure SQL database	SQL database in Fabric

SQL Database in Fabric



- ✓ Operational database
- ✓ Supports all traditional T-SQL workloads
- ✓ NO cross-database queries with other Fabric items

- ✓ Data mirrored in near real-time to Delta Parquet
- ✓ YES cross-database queries with other Fabric items
- ✓ NO write operations support

Core Components

Data Factory

Real-Time Intelligence

Databases

Analytics

Industry Solutions

Power BI

Partner solutions

Copilot in Fabric

OneLake

Microsoft Purview

Core Components

Analytics

Analytics in Microsoft Fabric

Data Engineering

Data Warehouse

Data Science

Fabric “Houses” - Lakehouse

+

Data Lake**house**

- Scalable, distributed file storage
- Flexible schema-on-read semantics
- Unstructured, semi-structured, and structured data

Data Ware**house**

- Relational schema modeling
- SQL-based querying
- Structured data

Lakehouse

- ✓ Files -> Unmanaged area -> store any format
- ✓ Tables -> Managed by Spark -> preferably Delta

- ✓ Read-only for T-SQL engine
- ✓ Only delta tables are exposed
- ✓ Needed for DirectQuery fallback

Working With Fabric Lakehouse

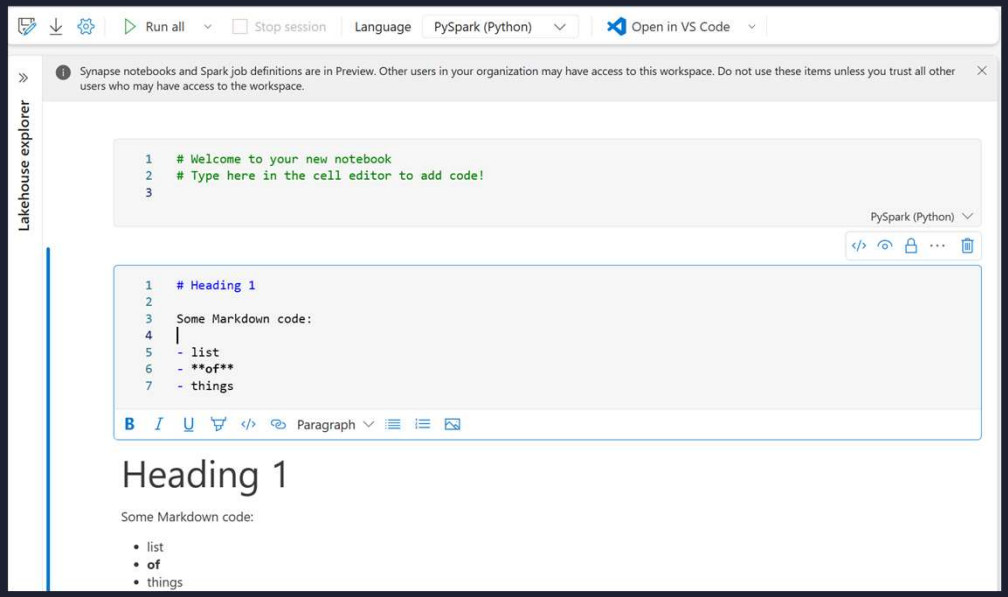
Tools and techniques to explore and transform data

- Apache Spark
  - Notebooks
  - Spark Job Definitions
- SQL analytics endpoint
- Dataflows (Gen2)
- Data Pipelines

Visualize with Power BI

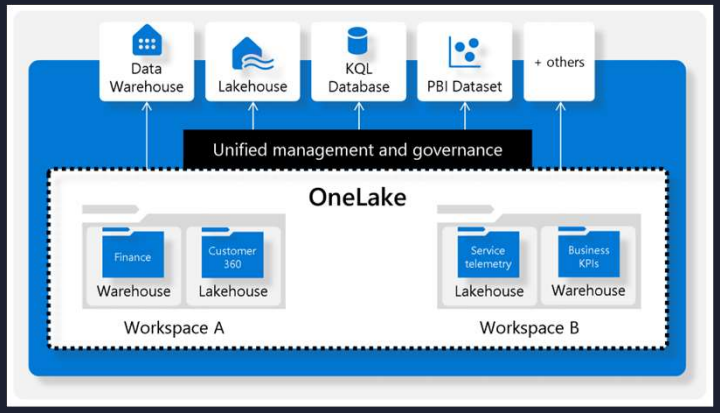
Fabric Notebooks

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



Warehouse

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



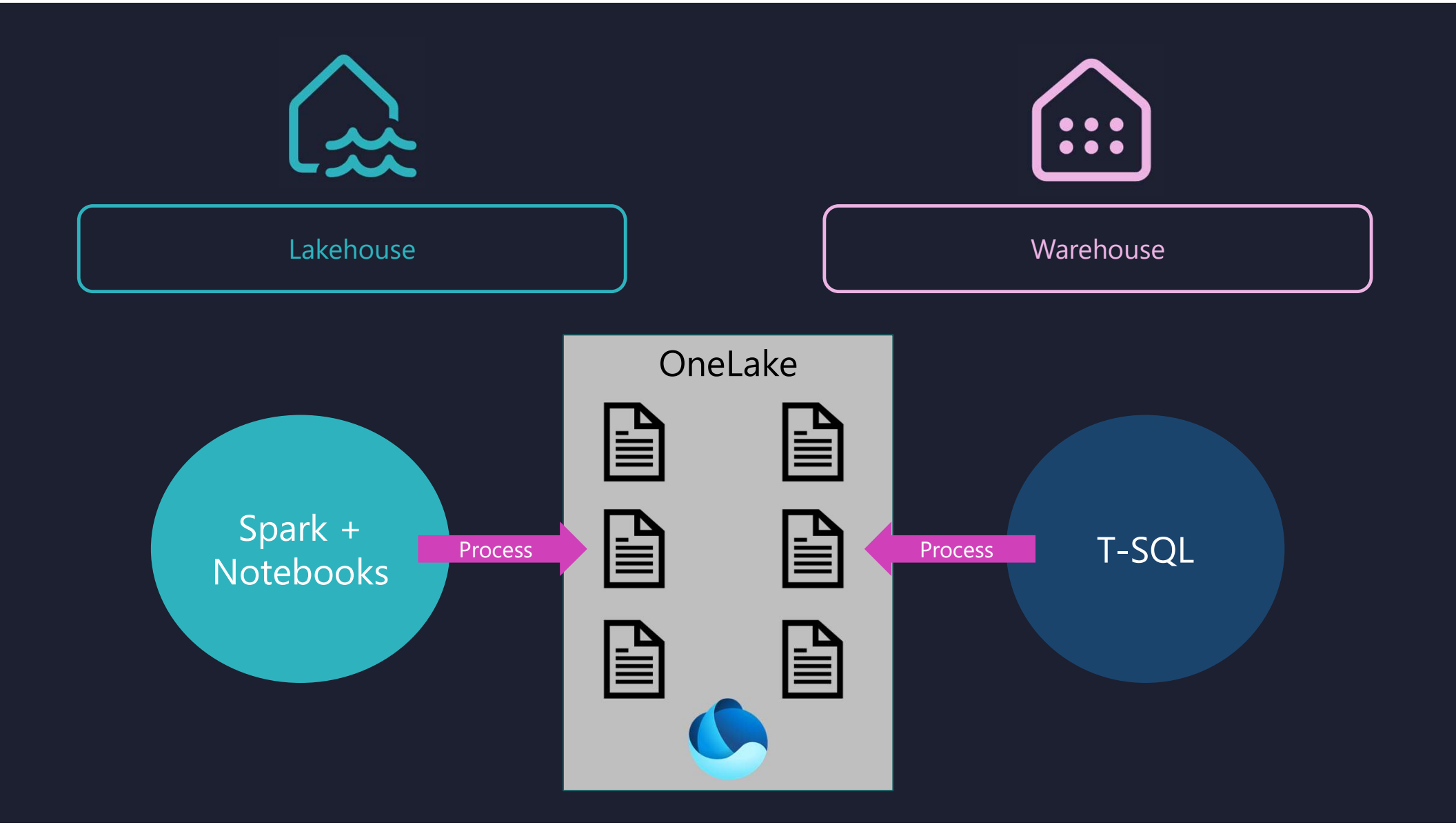
Two Types of SQL-"Houses" in Fabric

SQL Endpoint of the Lakehouse

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

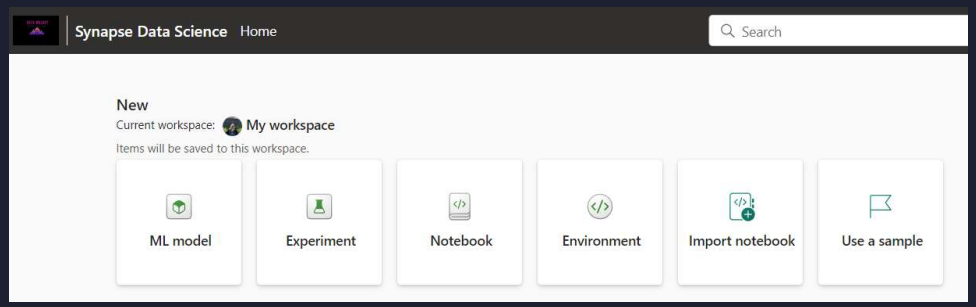
Synapse Data Warehouse

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


Lakehouse
Warehouse
OneLake
Spark + Notebooks
Process
Process
T-SQL

Data Science

- Data enrichment and business insights
  - Data exploration, cleansing, experiments, model scoring…
- Whole range of available tools



Data Science

Data exploration

Notebooks

Data preparation

Spark & Python

Data cleansing

Data Wrangler

Experimentation & ML Modeling

Variety of ML libraries

Enrich & Operationalize

Power BI & Semantic Link

Core Components

Data Factory

Real-Time Intelligence

Databases

Analytics

Industry Solutions

Power BI

Partner solutions

Copilot in Fabric

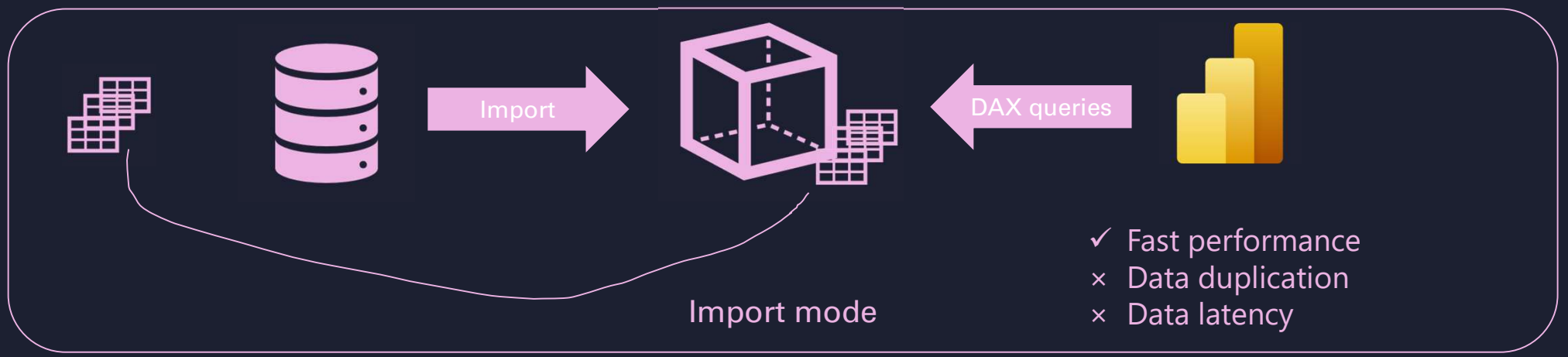
OneLake

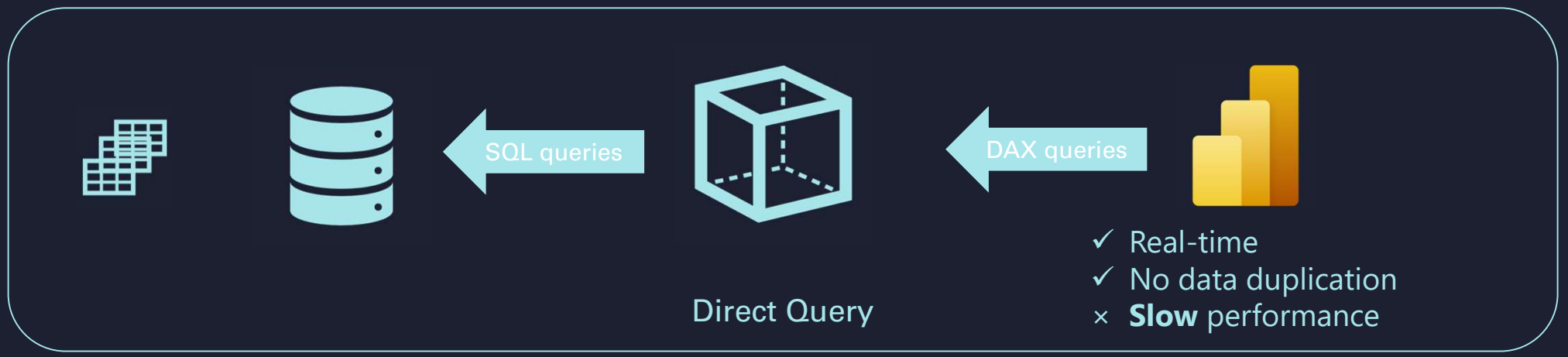
Microsoft Purview

Core Components

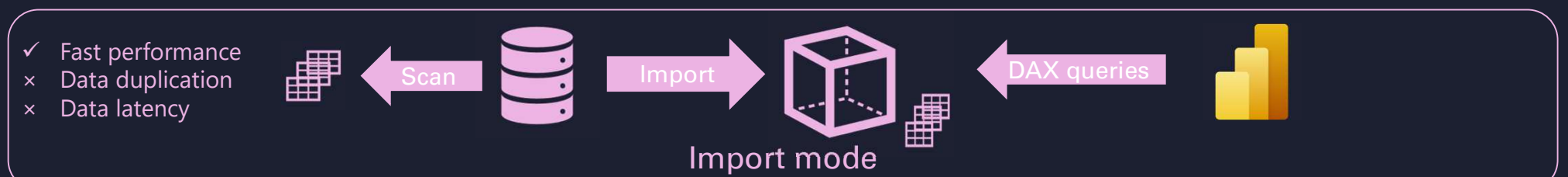
Power BI

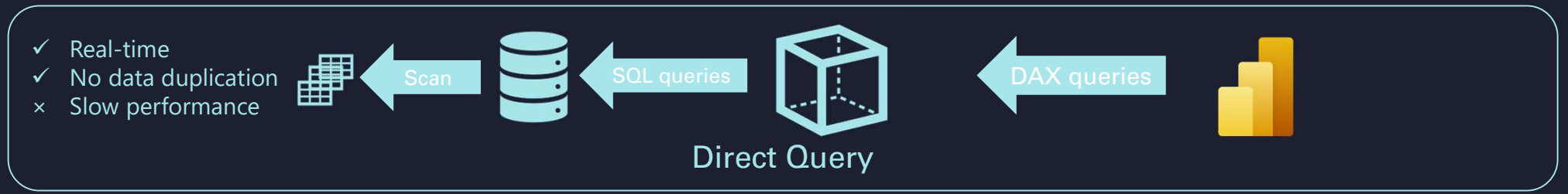
Power BI Architecture – Pre-Fabric

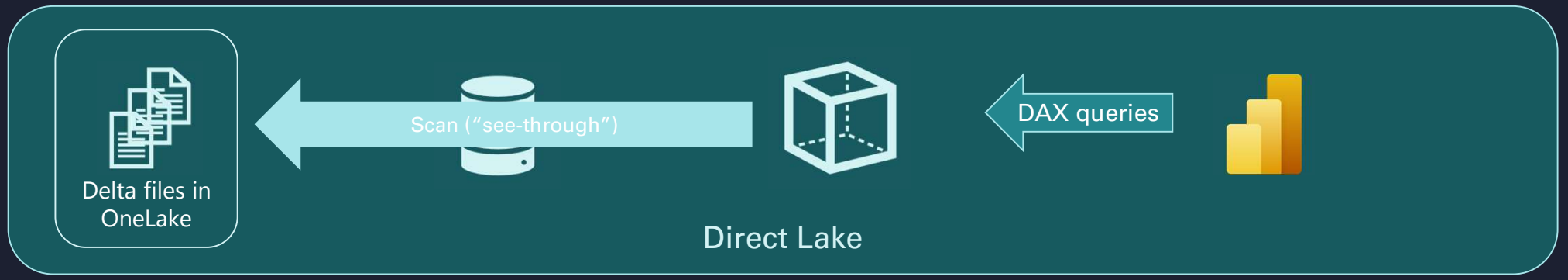




Power BI Architecture – Fabric







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

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)	
F 2	2		
F 4	4		
F 8	8		
F 16	16		
F 32	32		
F 64	64		$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

Compute costs

RI - Azure/M365

Billed monthly/yearly, with a monthly commitment

Microsoft Fabric

Fabric Capacity Reservation SKUs	Fabric Capacity pay-as-you-go SKUs
From	From
4.680,30 €	7.871,90 €
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]
• Smaller entry-level compute starting at F2. • Microsoft Azure Consumption Commitment (MACC) eligible. • Gain access to full Microsoft Fabric workloads through a unified product	• 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?

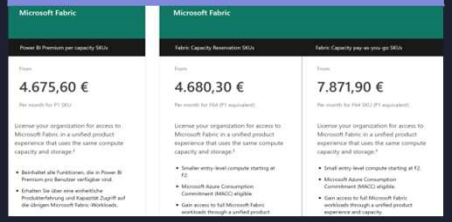
PAYG - Azure

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



Compute costs

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

*Only applicable for KQL Database workloads.

Storage costs

Copilot is currently available only for F64/P1 and higher!

Where to go from here?

- Set up your Fabric Trial account and try it for 60 days for free:
  - Fabric Trial Capacity
- Microsoft Fabric official documentation

Thank you

Nikola Ilic

@DataMozart

www.data-mozart.com

www.learn.data-mozart.com