



MS Fabric



Petr Filas

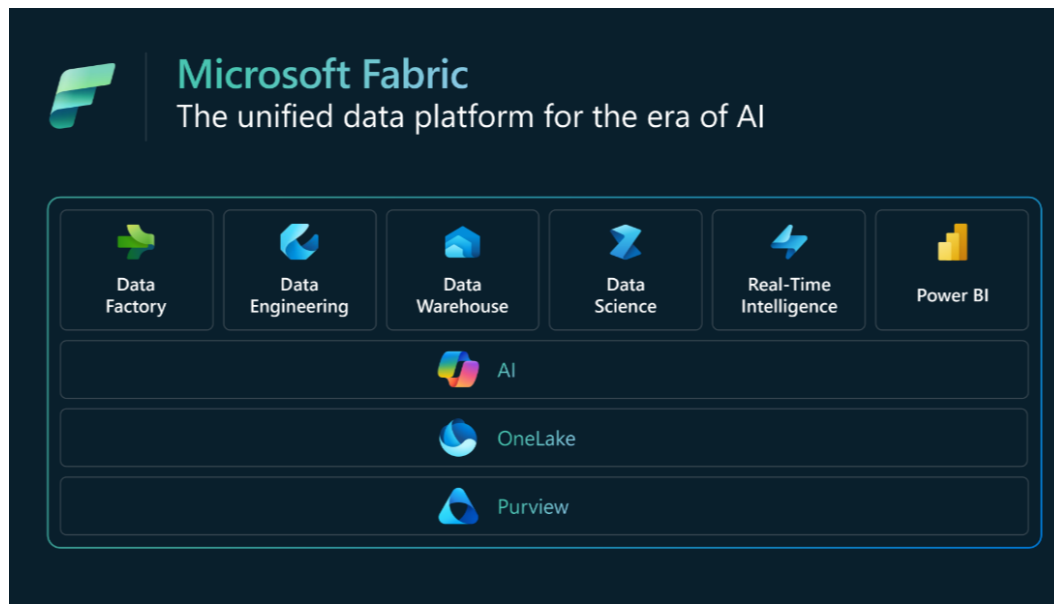
29. 10. 2025

Agenda

1. What is Microsoft Fabric?
2. Architecture & Components
3. How to start?
4. Demo

What is Microsoft Fabric?

- > Unified data platform
 - All-in-one analytics solution that covers everything from data movement to data science, Real-Time Analytics, and business intelligence
- > OneLake + (known) Azure Data Services
- > SaaS/PaaS
- > Only in MS Azure



Very brief history

> History

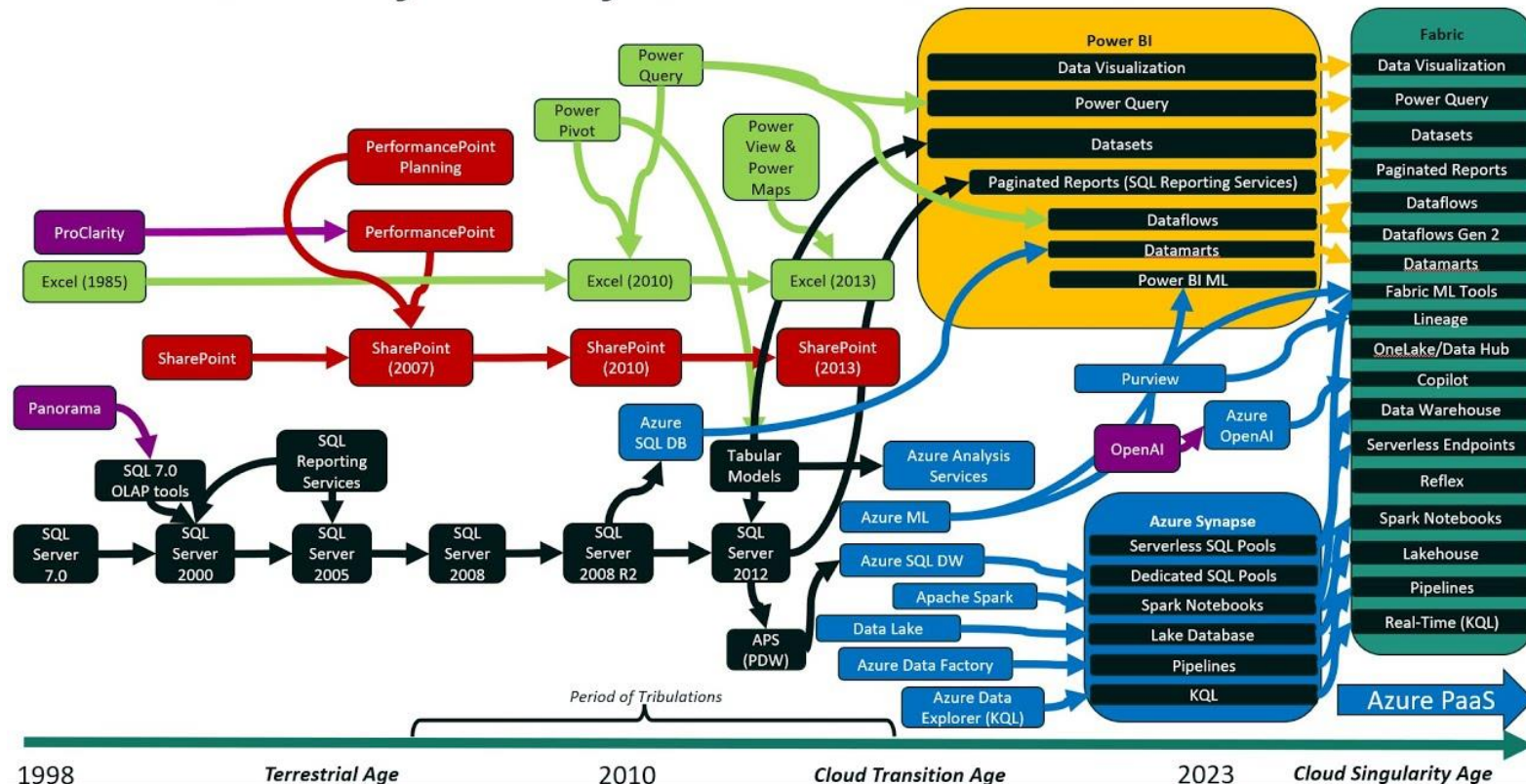
- Evolved from the need to unify analytics and data services, building on concepts from Azure Synapse Analytics (2019)
- Public preview in May 2023
- Generally available in November 2023

> Now

- Over 25,000 organizations were using Fabric as of September 2025
- 70% of Fortune 500 companies were reported to be using Microsoft Fabric as of mid-2025.
- Over 19,000 paying customers were reported as of October 2025.

Very brief history

Evolutionary History of Microsoft Fabric



Architecture & Components

> Core Principles

- **SaaS-first design** - Fully managed, scalable, and integrated platform
- **Unified experience** - One interface for data engineers, analysts, scientists, and business users
- **End-to-end workflows** - From data ingestion to visualization in one environment

> Compute

- Spark engine = Lakehouse, Data Science
 - PySpark, SparkSQL, Scala, SparkR
- Serverless SQL engine (Polaris/Synapse) = Warehouse
 - T-SQL
- KQL engine = Streaming
 - KQL language
- Analysis Service engine = Power BI
 - Power query, DAX

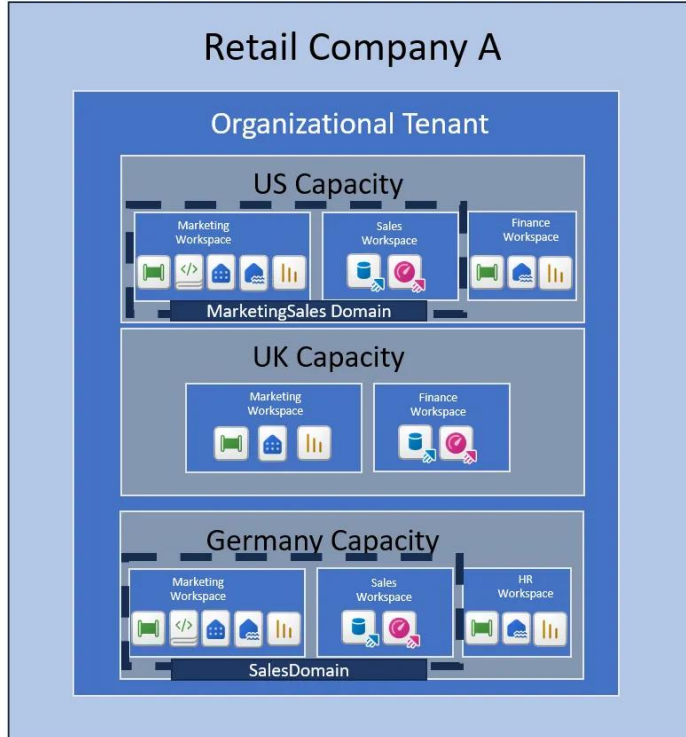
> Storage (OneLake)

- “OneDrive for data”
- Single logical data lake for the entire organization
- Built on Delta Lake format for compatibility and performance
 - Delta, Parquet
- Supports shortcuts to external data sources (e.g., Azure, AWS). Enables data virtualization without duplication.

MS Fabric organization

- **Azure (Entra) tenant** - single organization in MS Azure
- **Fabric Tenant** – instance of Fabric for an organization
- **Domain** - logical group of workspace e.g. HR, Sales, Risk
- **Workspace** – container of items and permissions (used for environment separation)
- **Items** – specific object e.g. Pipeline, Lakehouse, Warehouse
- **Capacity** – dedicated compute with specified CUs (Computation Units) ~ budget for all compute resources

MS Fabric organization



Components (Items)

< PROFINIT >



OneLake - Centralized data lake for all Fabric workloads

- Delta format support, no data duplication



Workspace - Organizational containers for Fabric items

- Role-based access (Admin, Member, Viewer), git integration for version control, linked to Fabric capacity



Lakehouse - Combines structured and unstructured data, supports Delta format storage

- Clusters - Big data processing using Spark, support Notebooks (Python, Scala, SQL)



Warehouse - Distributed SQL engine for structured data

- High concurrency and performance



Data Factory - Data movement and transformation (ETL/ELT)

- **Pipelines** - Visual or code-based data workflows
- **Dataflow Gen2** - Low-code transformations using Power Query



Power BI - Data visualization and reporting

- **Report** - Interactive dashboards and visuals
- **Semantic Model** - Data model for Power BI reports

Components (Items)

< PROFINIT >



Real-Time Analytics – Streaming and real-time data analysis

- **KQL Database** - Stores and queries streaming data using Kusto Query Language
- **Event Stream** - Ingests and processes real-time data



Data Science - Machine learning and AI workflows, integration with Azure ML, Python/R support

- **Notebook** - ML model development and training
- **Spark Job Definition** - Batch ML jobs



Data Activator - Event-driven automation

- **Data Activator Rule** - Logic for triggering actions based on data events



Databases – serverless Azure SQL

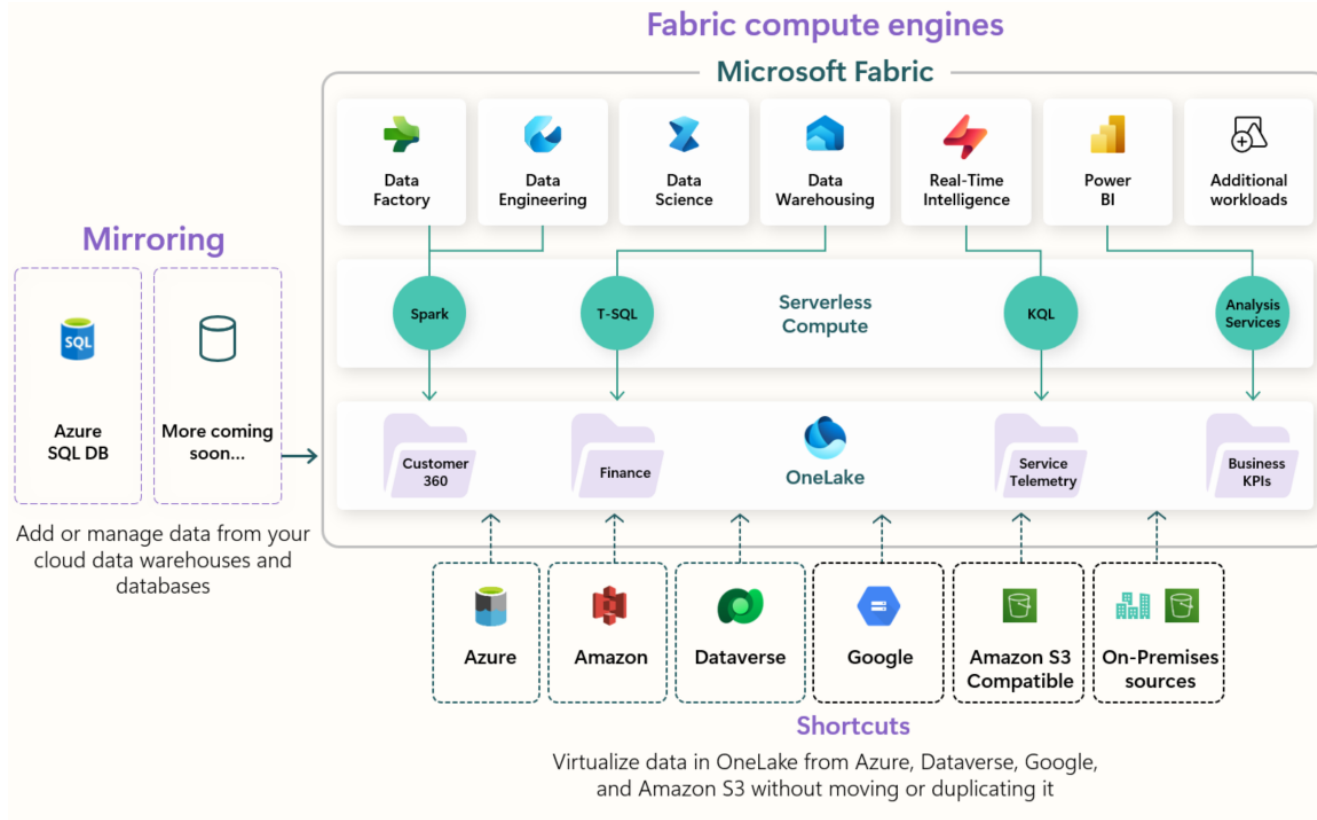
- Traditional relational database



Data agents (formerly AI skills) – public preview

- Conversational AI assistants that allow users to interact with structured organizational data using natural language

Components usage



- **Unified Security Model, but...**
- **Role-Based Access Control (RBAC)**
 - Granular permissions at workspace, item, and tenant levels
 - Roles include Admin, Member, Contributor, Viewer
- **Microsoft Entra ID (Azure AD)**
 - Centralized identity management
 - Supports multi-factor authentication and conditional access
- **OneLake security (Public Preview)**
 - Ambition of top-down security management (like Databricks Unity Catalog)
 - Row Level Security (RLS)
 - Column Level Security (CLS)

> Typically using Microsoft Purview

> Data Cataloging

- Automatically scans and registers data assets
- Metadata enrichment and classification

> Data Lineage

- Tracks data flow across pipelines, lakehouses, warehouses, and reports
- Visual lineage diagrams for auditing and troubleshooting

Pricing



- **Fixed monthly fee based on Tier (SKU)**
- **CU = Capacity Unit**
 - bundles compute resources: CPU, memory, disk I/O, and bandwidth
- **F64 brings**
 - Power BI licenses
 - Copilot
- **Reservation = 40% savings**
- **Scale Up/Down possible**
 - **No downtime** during scaling, but performance may temporarily degrade if resources are busy
 - Fabric may **postpone scaling down** until active jobs complete

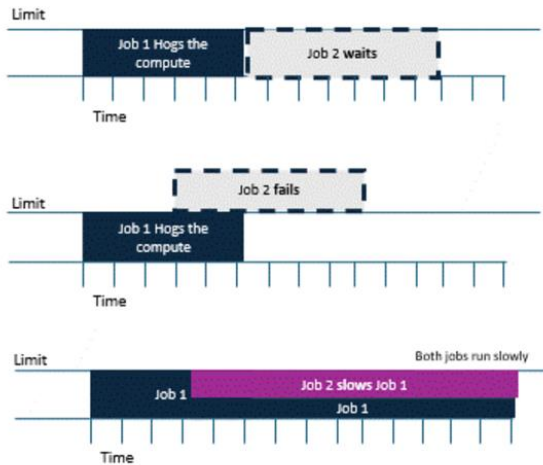
SKU	Capacity unit	Pay-as-you-go	Reservation
F2	2	\$262.80/month	\$156.334/month ~41% savings
F4	4	\$525.60/month	\$312.667/month ~41% savings
F8	8	\$1,051.20/month	\$625.334/month ~41% savings
F16	16	\$2,102.40/month	\$1,250.667/month ~41% savings
F32	32	\$4,204.80/month	\$2,501.334/month ~41% savings
F64	64	\$8,409.60/month	\$5,002.667/month ~41% savings
F128	128	\$16,819.20/month	\$10,005.334/month ~41% savings
F256	256	\$33,638.40/month	\$20,010.667/month ~41% savings
F512	512	\$67,276.80/month	\$40,021.334/month ~41% savings
F1024	1024	\$134,553.60/month	\$80,042.667/month ~41% savings
F2048	2048	\$269,107.20/month	\$160,085.334/month ~41% savings

Billing



- Billing is based on **CU-seconds** (CUs × time used)
- What if we need to scale up temporarily? (Compute spikes)
 - **Bursting and Smoothing**

Traditional Systems stay within a hard compute limit.

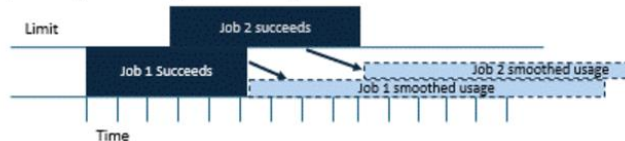


Bursting & Smoothing help to solve these problems.

Bursting lets both jobs **succeed**, running at **peak performance** to finish fast.



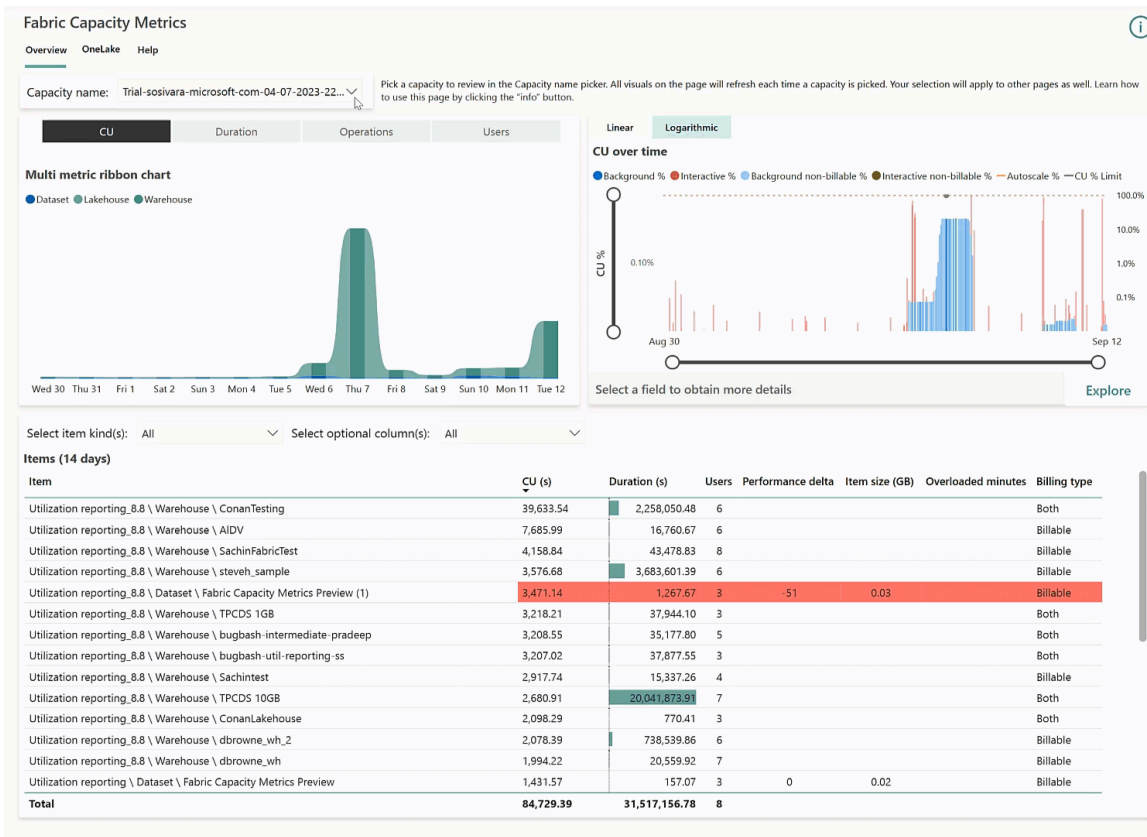
Smoothing distributes the **cost** of jobs over a longer time-period preventing scheduling issues.



Billing



- Consumption analysis using Metric App
- Clear overview of consumed CUs ~ \$\$\$
- Each Item has its own price



> Integration

- Native for Azure ecosystem services
 - Just enable access (using Managed Identities or token) and networking
 - Seamless integration to Microsoft 365 (Sharepoint, Teams)
- Data Factory for heterogenous sources/targets

> Data Sharing

- **Shortcuts** - Virtual links to external storage (no data duplication)
 - Azure Data Lake Storage, AWS S3, Google Cloud Storage
- **Mirroring** – replicates external databases into Fabric in near real-time
 - **Read-only copy** of the source database inside Fabric
 - Limited supported sources (Azure SQL, Azure Cosmos DB, Snowflake)

Data Access Modes (typically for Power BI)

< PROFINIT >

> Direct Query Mode

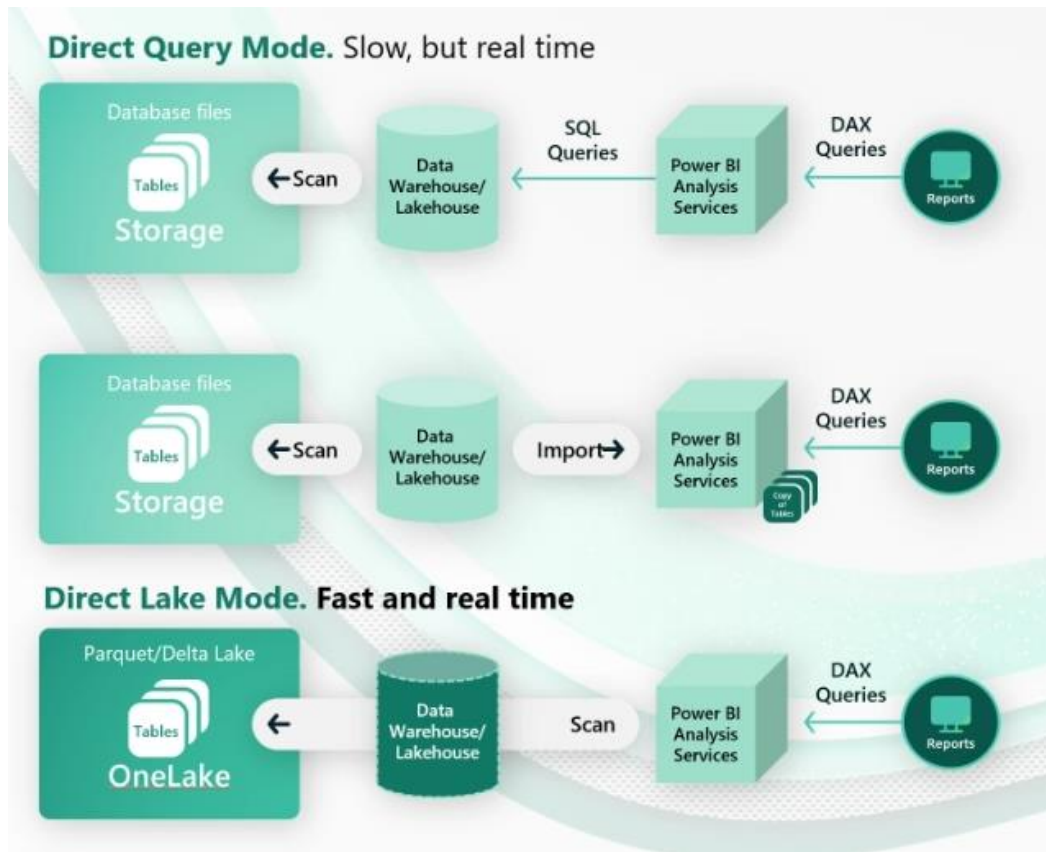
- Depends on source system speed
- Always up-to-date
- Slower for complex queries

> Import (Refresh) Mode

- Very fast for visuals after refresh
- Best performance for large reports
- Requires scheduled refresh

> Direct Lake

- Extremely fast for large datasets
- Limited to supported sources (OneLake)
- RLS/CLS degrades to Direct Query



Special features (performance optimization)



> V-Order (Microsoft feature)

- Proprietary optimization for columnar storage in Delta tables
- Improves scan performance for Power BI Direct Lake mode
- Reduces query latency for large datasets

> Z-Order Indexing (Delta format feature)

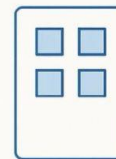
- Sorts data on multiple columns to minimize I/O
- Ideal for queries with filter conditions on multiple dimensions
- Common in large-scale analytical workloads

> Vacuum (Delta format feature)

- Cleans up obsolete Delta files after updates/deletes
- Reduces storage footprint and improves query efficiency
- Helps maintain transaction log health.

 Optimizing Data Performance in Microsoft Fabric:
V-Order vs. Z-Order

V-ORDER

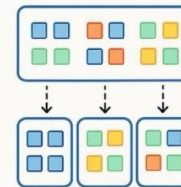


Write

V-Order: Write-Time Optimization

- Enhances read performance by optimizing Parquet files during data write operations.

Z-ORDER



Read

Z-Order: Read-time Optimization

- Improve query performance by co-locating related data, minimizing amount of data scanned during queries

Feature	V-Order	Z-Order
Timing	Write-time optimization	Read-time optimization
Focus	General read performance	Query performance on specific columns
Implementation	Automatic In Fabric	Requires manual configuration
Compatibility	Fully compliant with Parquet	Works with Delta Lake

Special features (performance optimization)

< PROFINIT >

> Native Execution Engine

- Compiled C++ code for maximum performance
- Similar to Photon Engine in Databricks
- No additional costs

Typical Architectures

> Medallion architecture

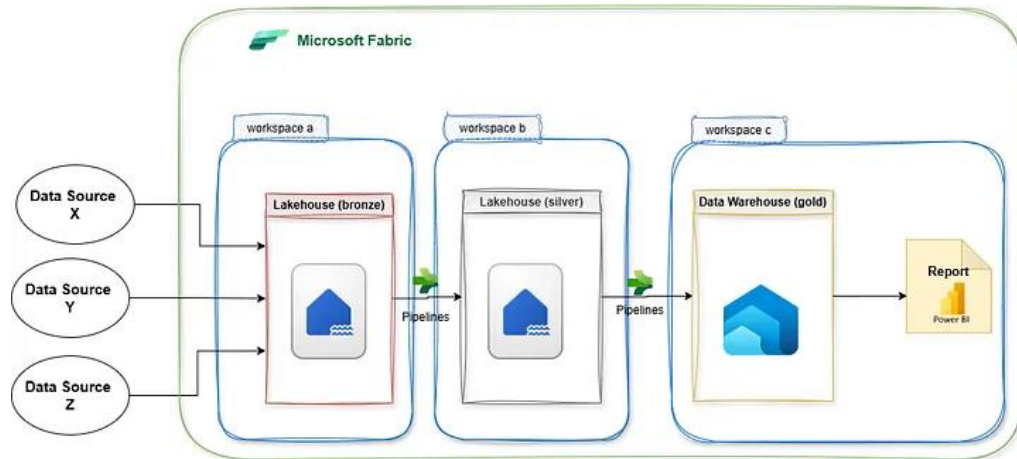
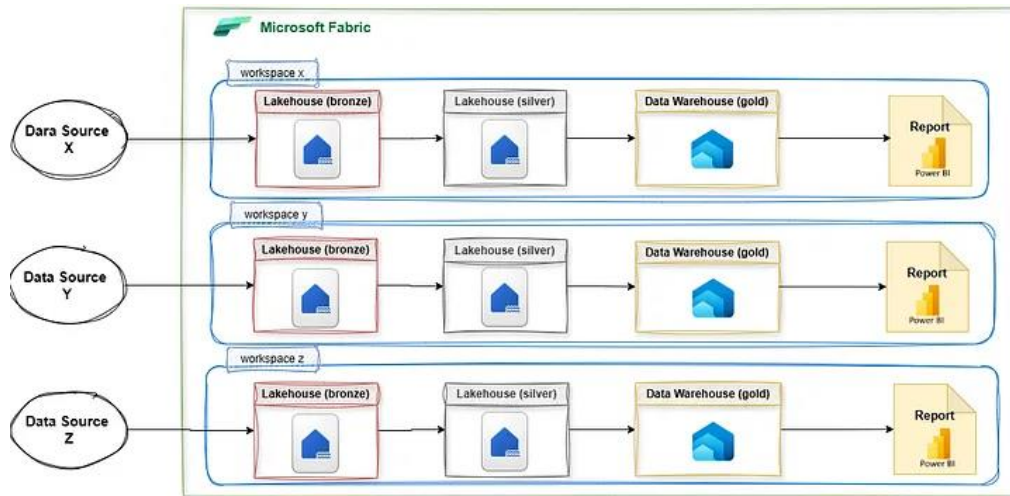
> Warehouse vs Lakehouse

> Isolation vs Sharing

- Workspaces isolate data and compute resources
- Sharing across workspaces requires **datasets or shortcuts**, which adds complexity

> Governance Challenges

- Policies and lineage tracking must span multiple workspaces
- Risk of **fragmented governance** if too many workspaces exist



Considerations

> Vendor Lock-In Risks

- Single Ecosystem Dependency on Azure
- Proprietary Features like Direct Lake mode and OneLake shortcuts

> Platform Inmaturity

- New product which is still evolving = trying to reduce the gap with industry leaders
- Frequent updates and roadmap changes
- Public Preview = Public Beta Testing

> But it's Microsoft 😊

How to start?

- > <https://learn.microsoft.com/en-us/fabric/fundamentals/fabric-trial>
- > What you get in Trial
 - Gives you “limited” F64 capacity for 60 days
 - Data Factory, Synapse Data Engineering, Real-Time Analytics, Power BI
 - Up to 1 TB storage in OneLake
- > What's not working in Trial
 - Copilot features
 - Data agents
 - Capacity Management
 - No SLA

Conclusion

Conclusion

> Pros

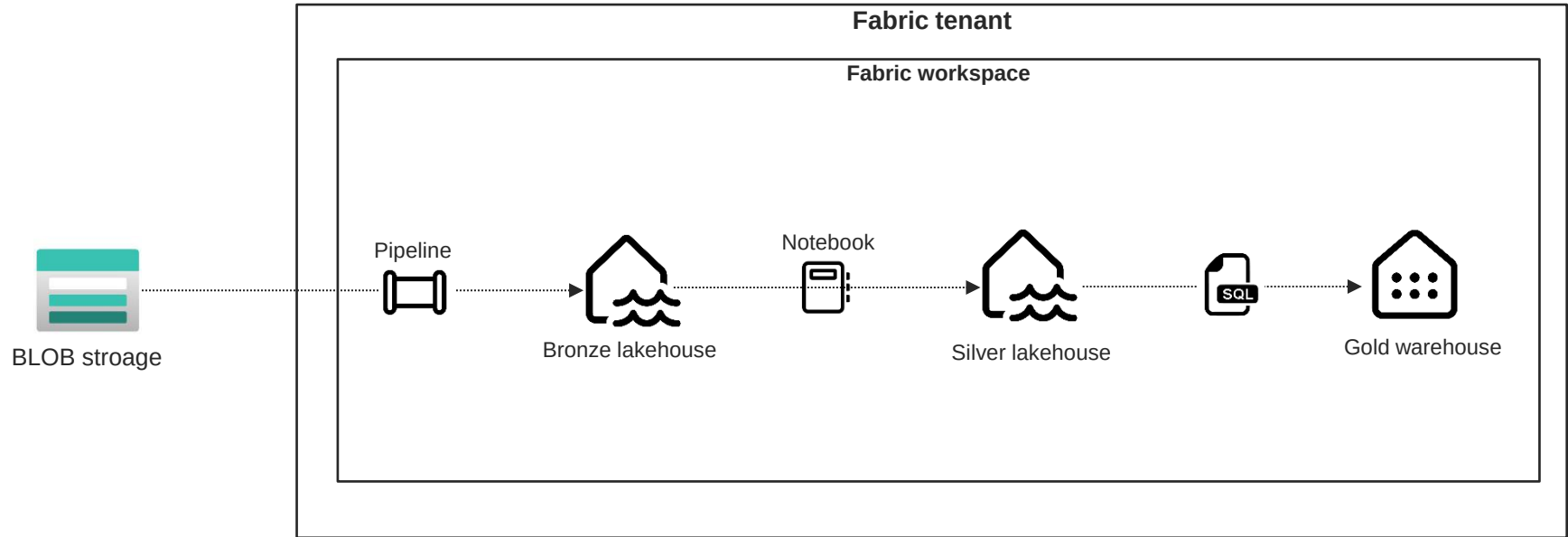
- OneLake
- Unified Platform
- Seamless integration to Azure, M365
- Licensing – fixed fee per month

> Cons

- Vendor Lock-In - Heavy reliance on Microsoft ecosystem, migration is complex
- Platform Inmaturity - some features still in preview, bugs etc.
- Licensing – can be sometimes confusing

Demo

Demo



Thank you for your attention

Profinit EU, s.r.o.
Tychonova 2, 160 00 Praha 6

Tel.: + 420 224 316 016, web: www.profinit.eu

 LINKEDIN
linkedin.com/company/profinit

 TWITTER
[@profinit_EU](https://twitter.com/profinit_EU)

 FACEBOOK
facebook.com/Profinit.EU

 YOUTUBE
Profinit EU, s.r.o.