Analytics in the era of AI
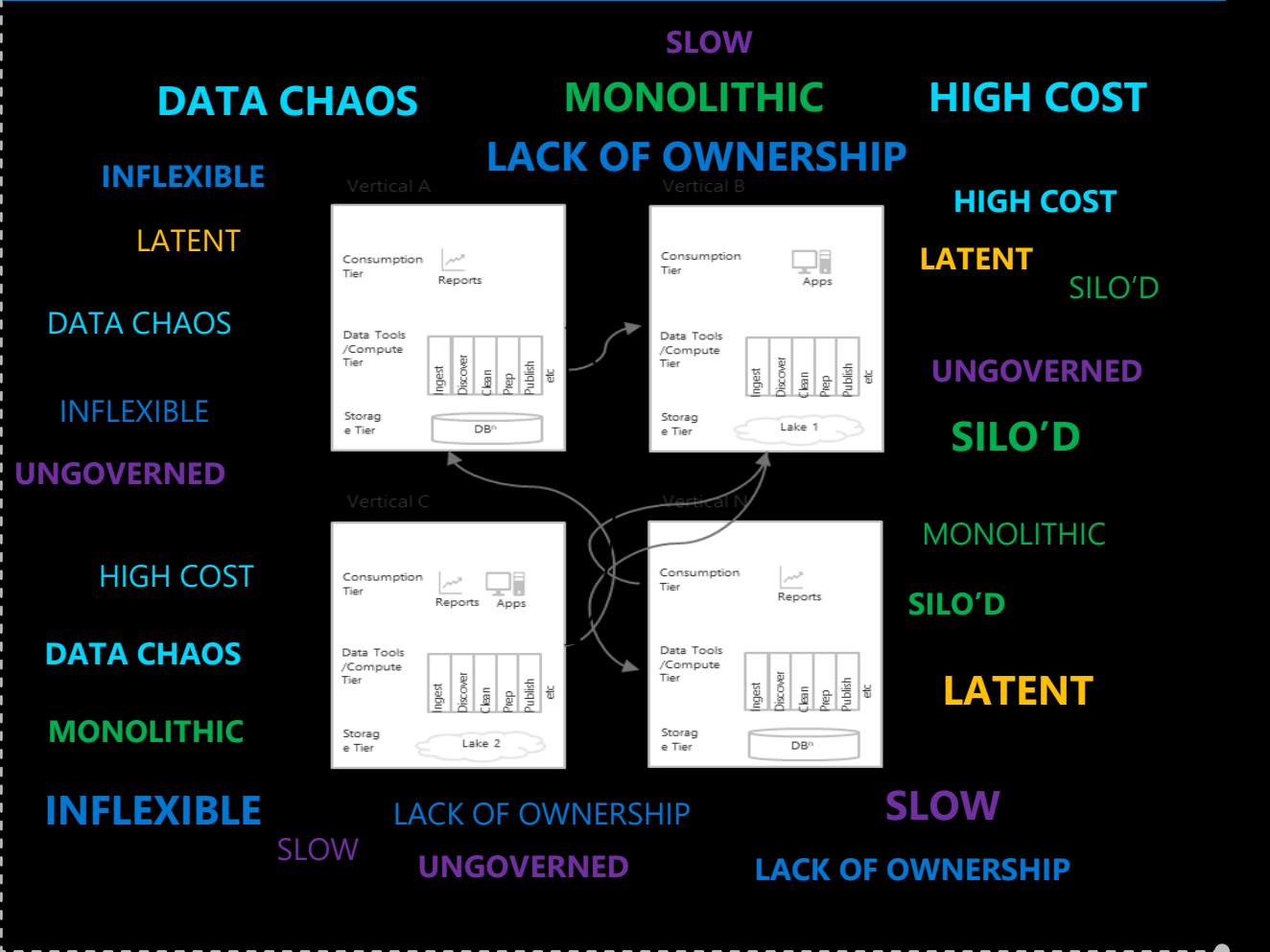
Stelios Dontas
Solution Area Specialist, Data & AI

Monolithic Data Estate with vertical pockets of analytics

Data Chaos and Analytics Silos

SLOW

DATA CHAOS   MONOLITHIC   HIGH COST

LACK OF OWNERSHIP

INFLEXIBLE   **HIGH COST**

LATENT   LATENT   SILO'D

DATA CHAOS   **UNGOVERNED**

INFLEXIBLE   **SILO'D**

UNGOVERNED

HIGH COST   MONOLITHIC

DATA CHAOS   **SILO'D**

MONOLITHIC   **LATENT**

INFLEXIBLE   LACK OF OWNERSHIP   **SLOW**

SLOW   **UNGOVERNED**   **LACK OF OWNERSHIP**

Many local/vertical analytics compute & storage platforms with differing underlying analytics architectures
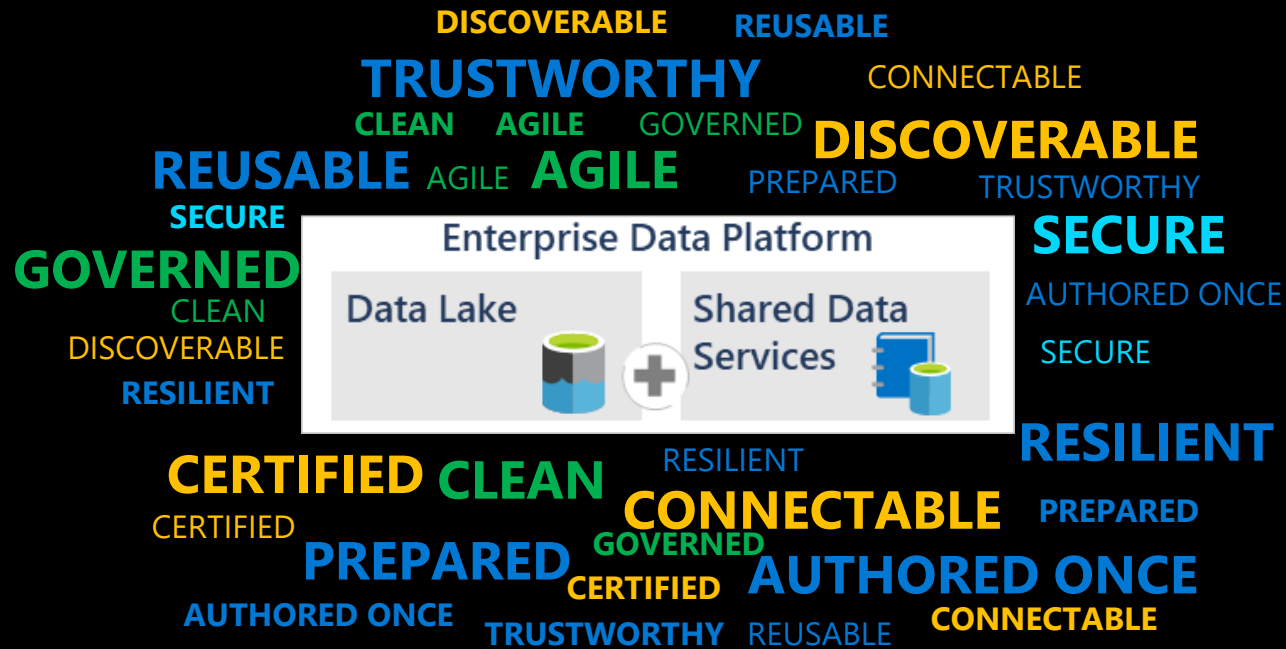
Each highly specialized to the unique data & stakeholder needs

Each have significant code bases & engineering teams

Many engineered outside of IT

Target Data Estate

OneLake for all Data
"The OneDrive for Data"

Data Factory · Synapse Data Engineering · Synapse Data Science · Synapse Data Warehouse · Synapse Real Time Analytics · Power BI · Data Activator
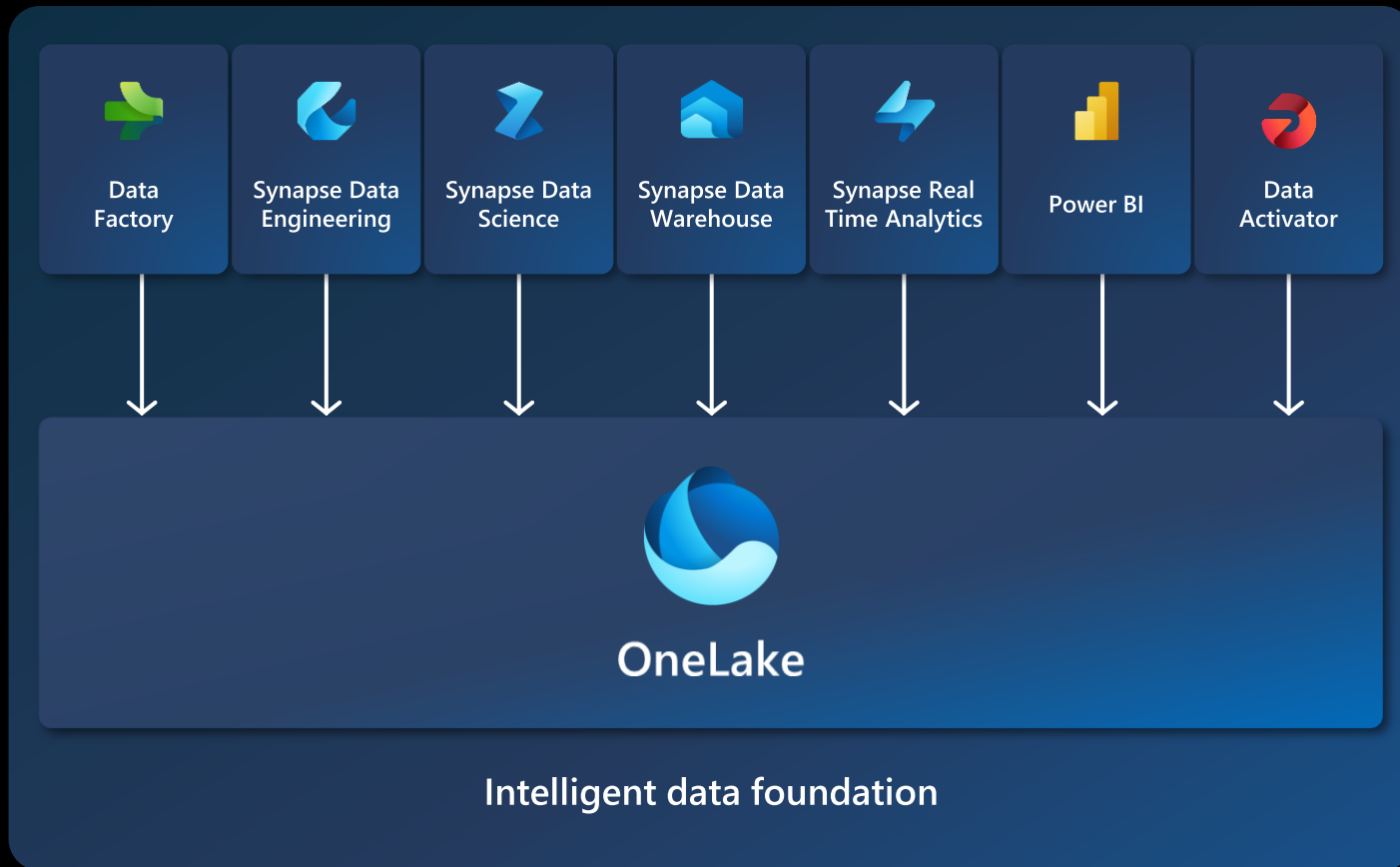
OneLake

Intelligent data foundation

A single SaaS lake for the whole organization

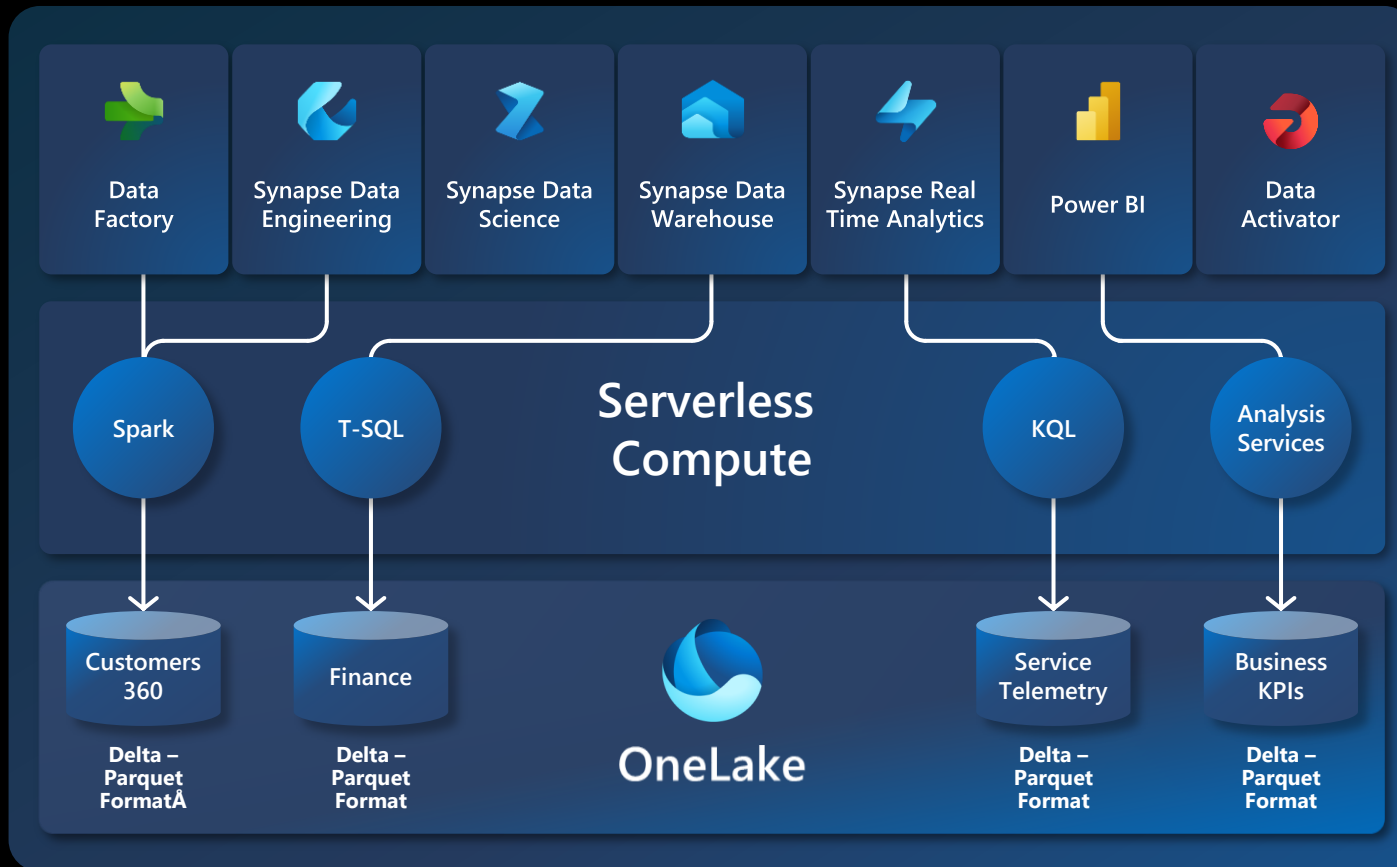Provisioned automatically with the tenant

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

One Copy for all computes
Real separation of compute and storage

Data Factory	All the compute engines store their data automatically in OneLake
Synapse Data Engineering	The data is stored in a single common format
Synapse Data Science	Delta – Parquet, an open standards format, is the storage format for all tabular data in Analytics vNext
Synapse Data Warehouse	Once data is stored in the lake, it is directly accessible by all the engines without needing any import/export
Synapse Real Time Analytics	
Power BI	All the compute engines have been fully optimized to work with Delta Parquet as their native format
Data Activator	Shared universal security model is enforced across all the engines

Serverless Compute

Spark — T-SQL — KQL — Analysis Services

Customers 360 — Delta – Parquet FormatÂ
Finance — Delta – Parquet Format
Service Telemetry — Delta – Parquet Format
Business KPIs — Delta – Parquet Format

OneLake

One Copy for all computes
Universal security makes it real

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 Analytics vNext

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

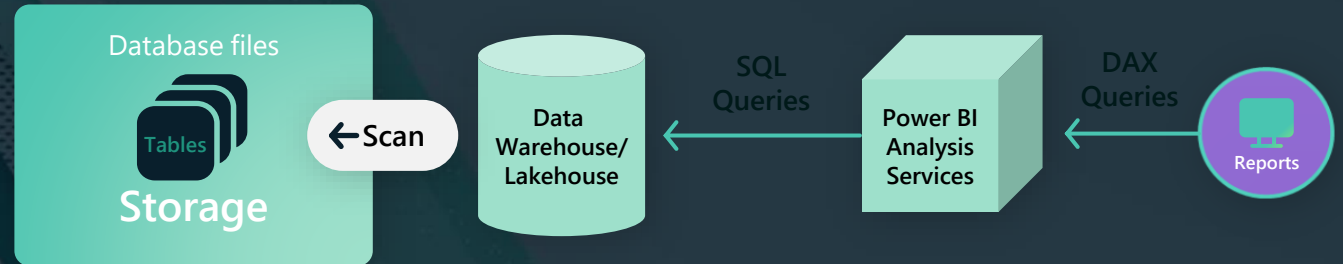Shared universal security model is enforced across all the engines
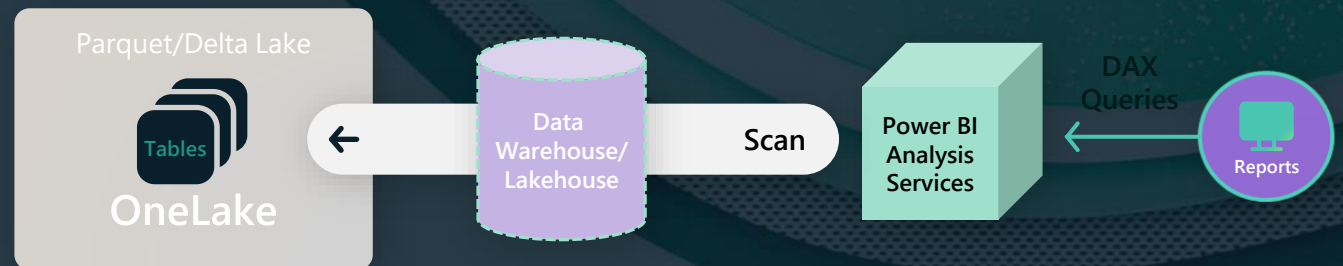
Power BI | Direct Lake Mode

Direct Lake is a fast-path to load the data from the lake straight into the Power BI engine, ready for analysis

Direct Lake is based on loading parquet-formatted files directly from a data lake without having to query a Lakehouse endpoint, and without having to import or duplicate data into a Power BI dataset

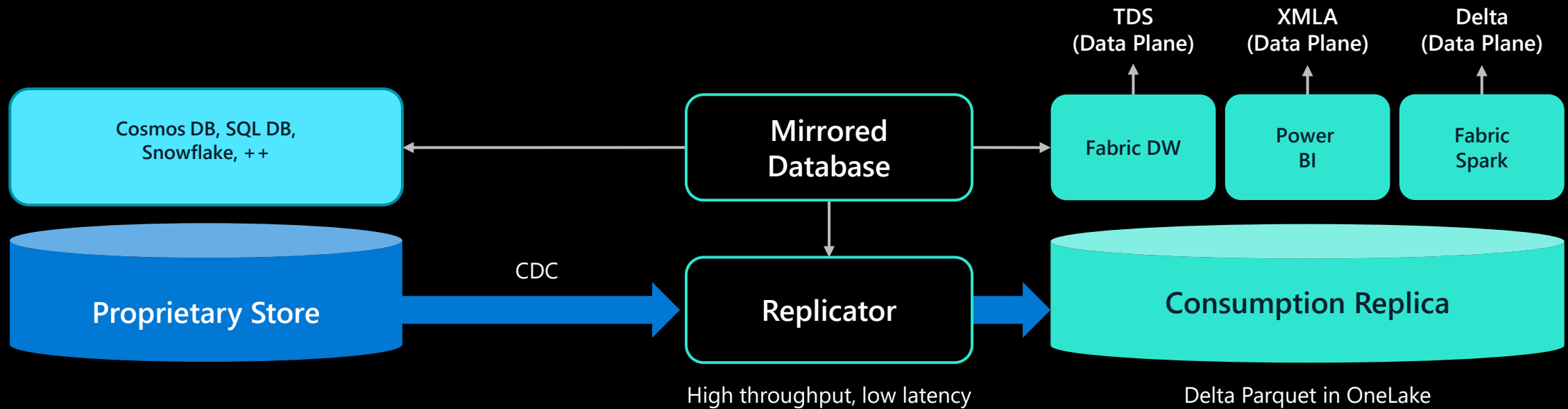**Direct Query Mode.** Slow, but real time

Database files — Tables — Storage ← Scan → Data Warehouse/Lakehouse ← SQL Queries ← Power BI Analysis Services ← DAX Queries ← Reports

Database files — Tables — Storage ← Scan → Data Warehouse/Lakehouse — Import → Power BI Analysis Services (Copy of Tables) ← DAX Queries ← Reports
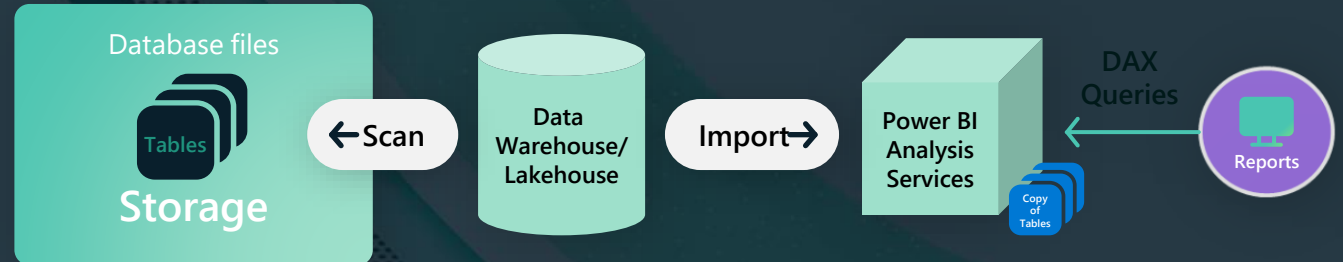
Direct Lake Mode. **Fast and real time**

Parquet/Delta Lake — Tables — OneLake ← → Data Warehouse/Lakehouse — Scan → Power BI Analysis Services ← DAX Queries ← Reports

Mirroring in Fabric

Near real-time replication of proprietary database stores with zero-ETL

Thank you