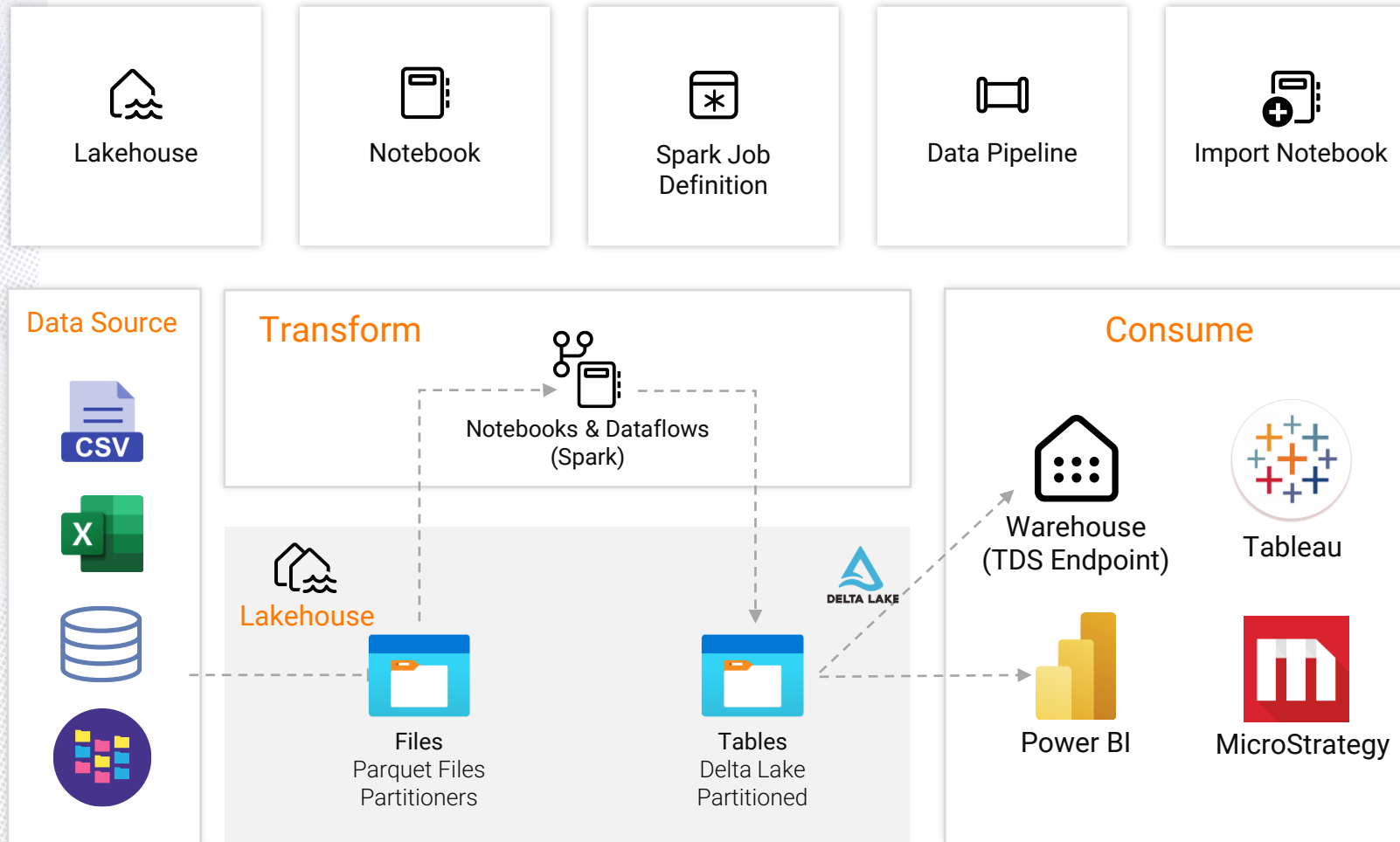


Microsoft Fabric - Components



- **One Lake:** A logical data lake, that unifies data and simplifies data discovery, providing a single source of truth to all your analytical applications across entire organizations.
- **Data Factory :** A low/no code GUI based application to Ingest, transform and load data.
- **Data Engineering :** Design, build and maintain various components such as lakehouse, spark jobs, notebooks, data pipelines, to ensure data is accessible, organized and of high quality.
- **Synapse Data Warehouse:** A traditional data warehouse that supports full sql capabilities.
- **Synapse Data Science:** A central location to create ML models, notebooks.
- **Synapse Realtime Analytics:** A quick and easy, fully managed data analytics platform optimized for streaming data.
- **Power BI:** A Business Intelligence application to provide interactive insights in the form of reports and dashboards.

Lakehouse Architecture



Capabilities

- Read only, system generated SQL Endpoint for Lakehouse for T-SQL querying and serving.

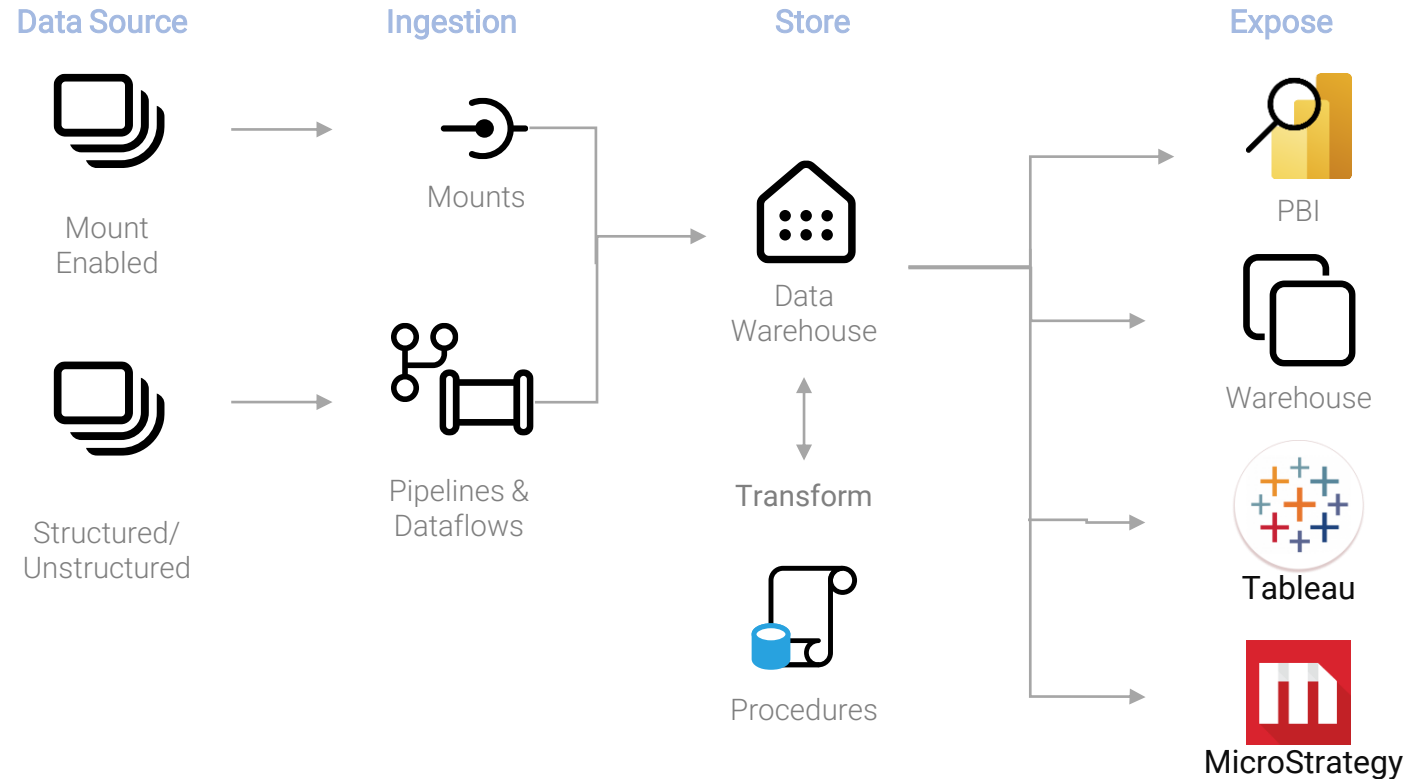
Recommended Use Cases

- Exploring and querying delta tables from the lakehouse
- Staging Data and Archival Zone for analysis
- Architecture with zones for bronze, silver and gold analysis

Data Loading & Management

- Supports full DQL
- No DML, and limited DDL T-SQL support
- Data loading via Spark, Pipelines, dataflows, Shortcuts
- Supports only **read** delta tables

Fabric Data Warehouse Architecture



Capabilities

- ACID compliant, full data warehousing with transactions support in T-SQL.

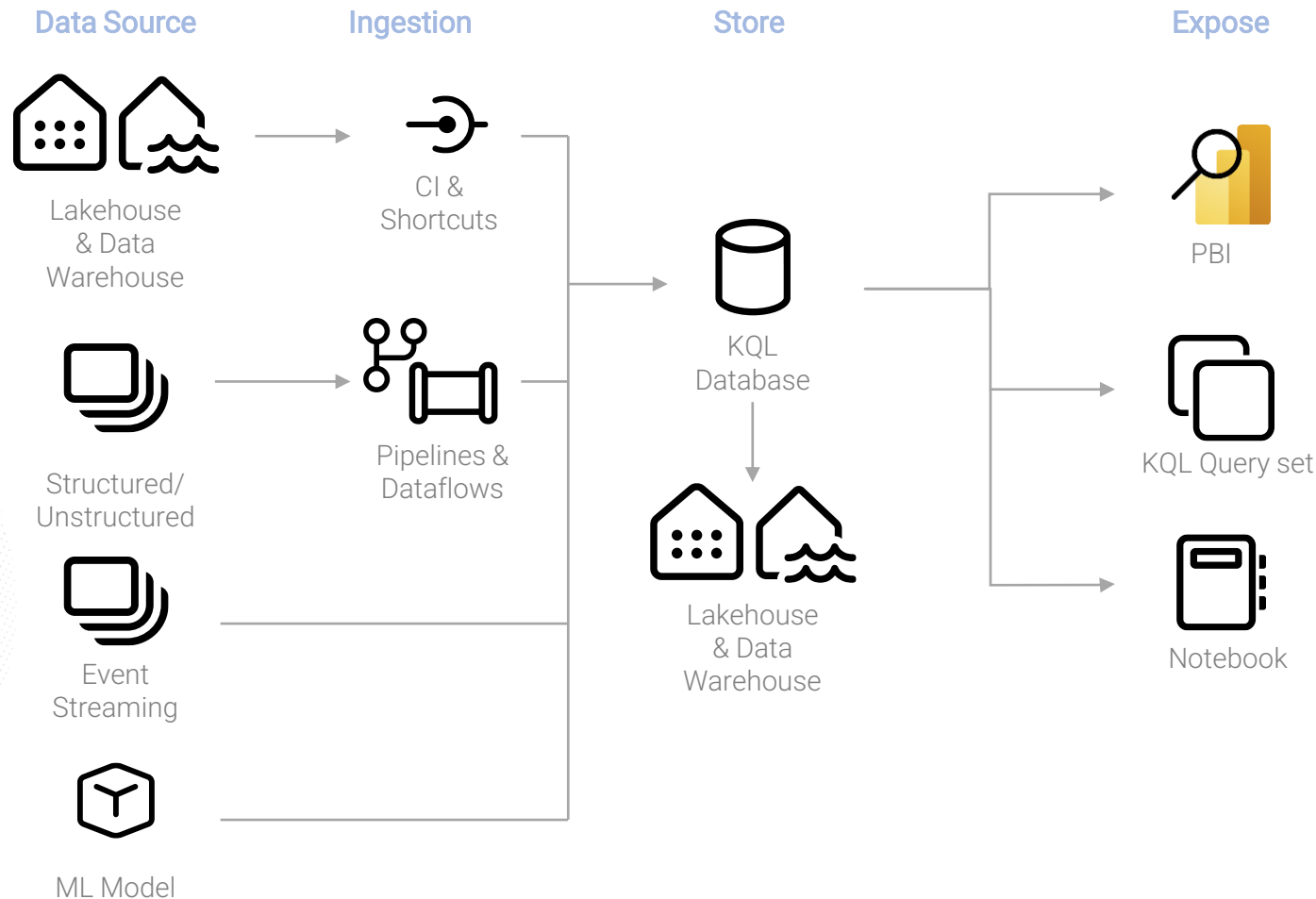
Recommended Use Cases

- Data Warehousing for enterprise use
- Data Warehousing for departmental use or self-service use
- Structured data analysis

Data Loading & Management

- Supports full DQL, DML, and DDL T-SQL support, full transaction support
- Data loading via SQL, Pipelines, dataflows
- Supports read and write delta tables

Real Time Architecture



Capabilities

- Fully managed big data analytics platform optimized for streaming, and time-series data.
- Utilizes query language and engine with exceptional performance for searching structured, semi-structured, and unstructured data.

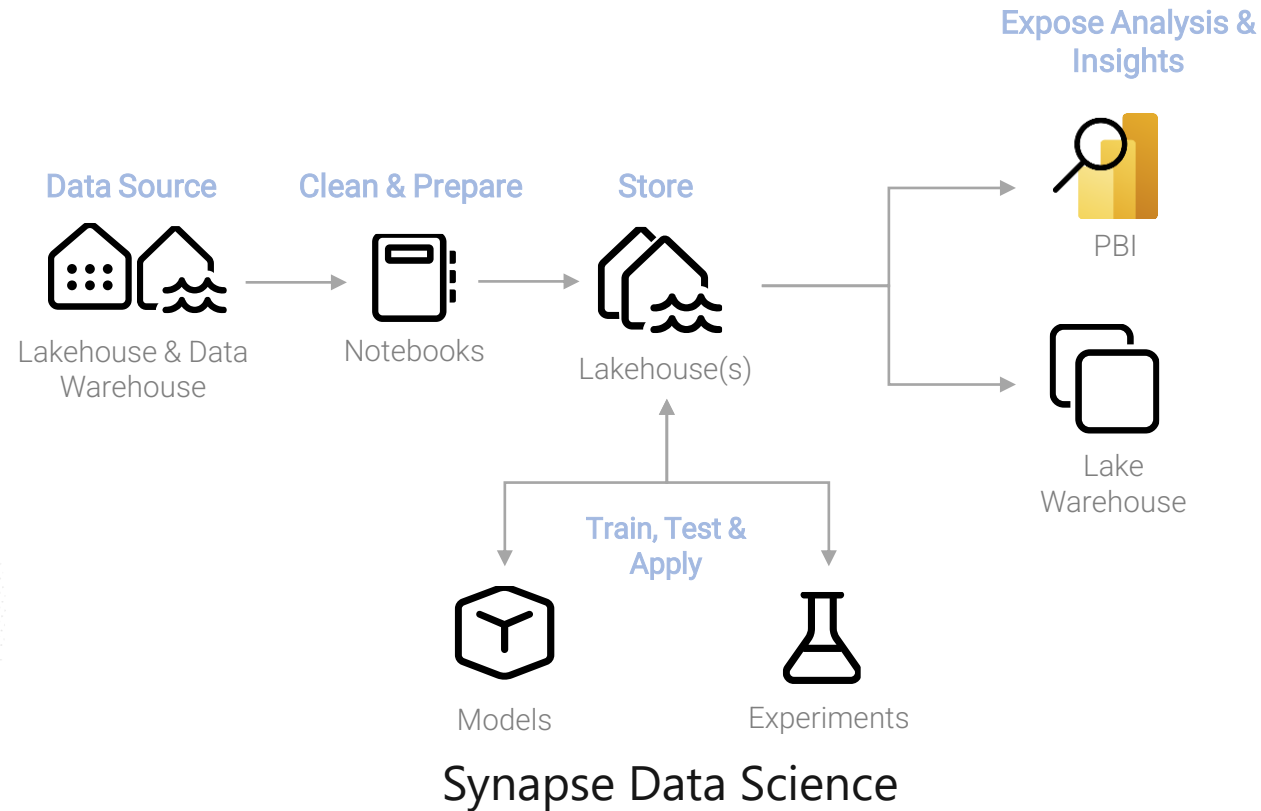
Recommended Use Cases

- High freshness from data ingestion to query
- Transform streaming data
- Need for low query latency

Data Loading & Management

- **Event Stream:** Capture, transform, and route real-time events
- **KQL database:** Data storage into KQL database can be accessed in OneLake
- **KQL Queryset:** Run queries, view, and customize query results on data.

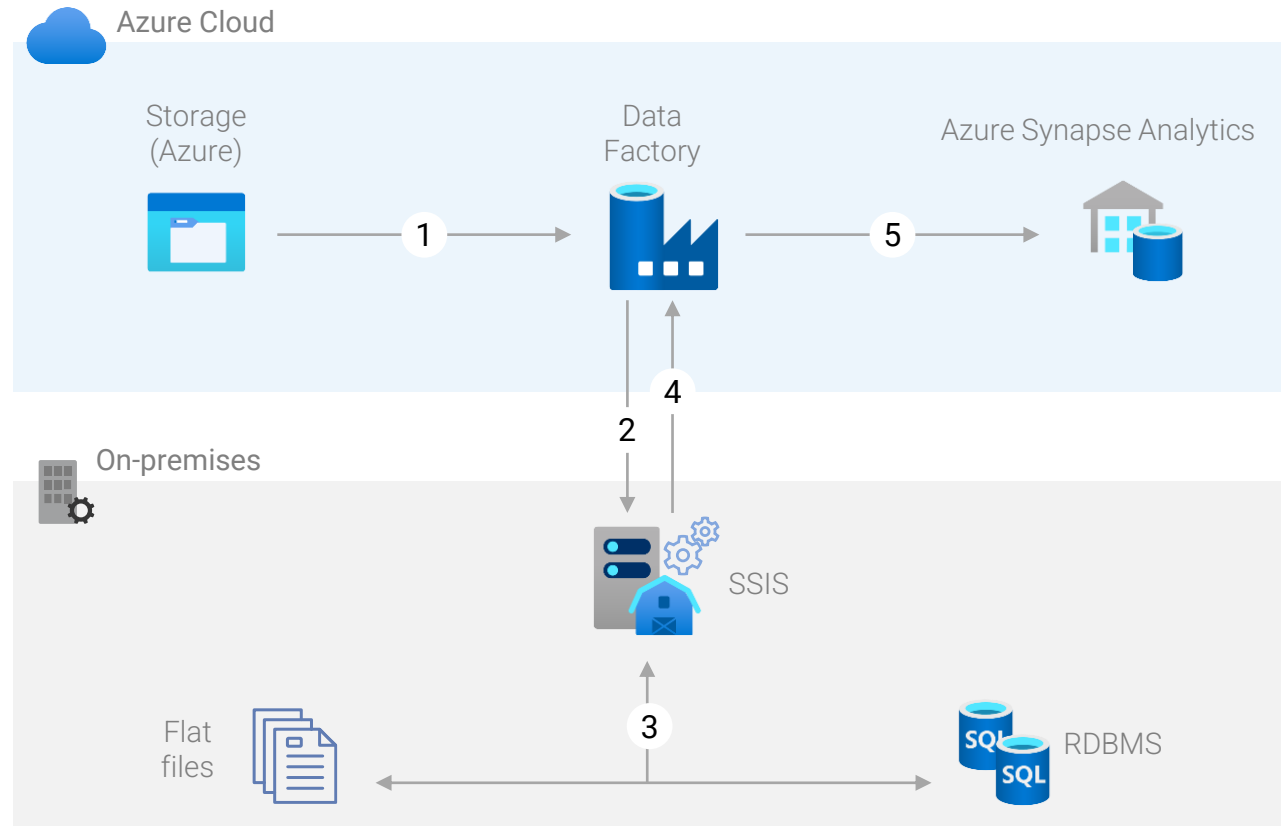
Data Science Architecture



Capabilities

- **Data Discovery & Exploration**: Easy-to-build data pipelines can access and transform the data into a format that machine learning can consume. **Notebooks** become one of the quickest ways to get started with data exploration.
- **Data Preparation**: With **Spark**, users can leverage PySpark/Python, Scala, and SparkR/SparklyR tools for data pre-processing at scale.
- **Training**: Tools like **PySpark/Python, SparklyR/R, notebooks** handle machine learning model training.
- **MLflow**: Offers a built-in MLflow experience with which users can interact, to log experiments and models.
- **SynapseML**: SynapseML (previously known as **MMLSpark**) open-source library, that Microsoft owns and maintains, simplifies massively scalable machine learning pipeline creation.
- **Gain Insights**: Predicted values can easily be written to OneLake, and seamlessly consumed from Power BI reports, with the **Power BI Direct Lake** mode.

Hybrid Architecture



1. Data is sourced from Azure Blob Storage into Data Factory.
2. The Data Factory pipeline invokes a stored procedure to run an SSIS job that's hosted on-premises via the integration runtime.
3. The data cleansing jobs are run to prepare the data for downstream consumption.
4. After the data cleansing task finishes successfully, a copy task is run to load the clean data into Azure.
5. The clean data is then loaded into tables in Azure Synapse Analytics.