Common analytics scenarios

Andrea Benedetti
Sr Cloud Architect, Microsoft
/in/abenedetti @anBenedetti https://github.com/anbened

#FabricGarage | 001 | 2024.05.16

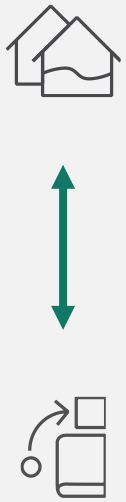
From data to competitive advantage

- Today’s Challenges
  - technology complexity
  - fragmented data
  - security management
  - compliance issues

- Microsoft Fabric
  - unified data stack
  - shared experiences / architecture / security / governance / compliance
  - every aspect of your data estate

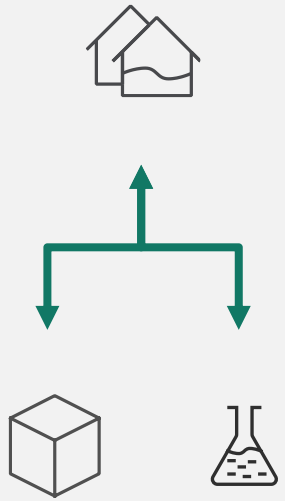
Introducing the end-to-end scenarios in Microsoft Fabric

Lakehouse



Data Warehouse

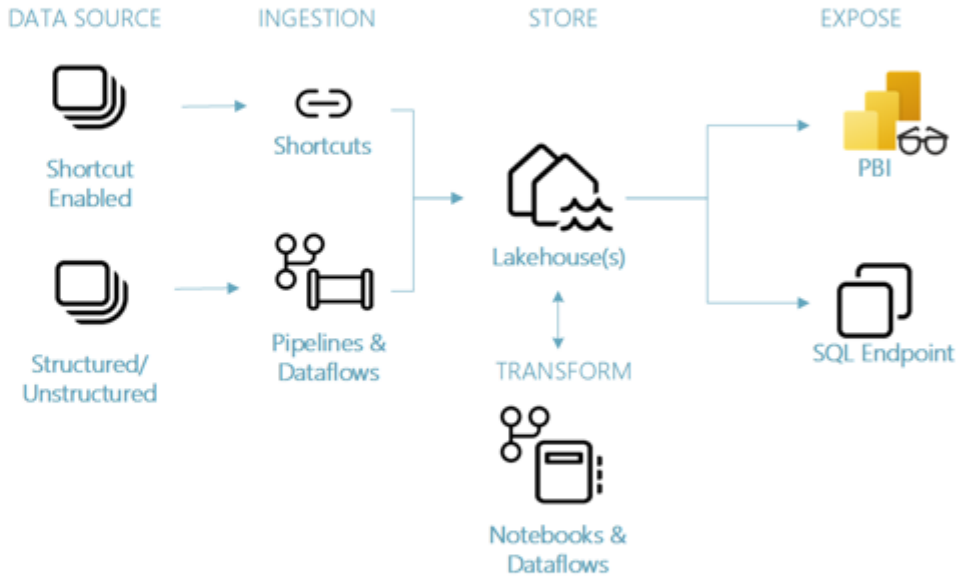
Data Science



Real Time Analytics

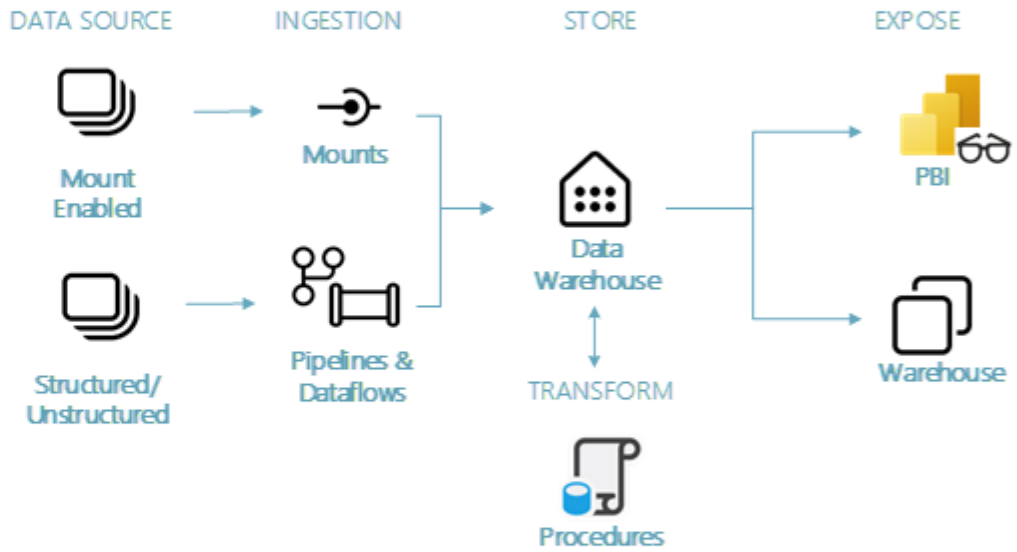
Lakehouse

- **Ingestion**
  Shortcuts Enabled sources (ADLS G2, S3, etc), Pipelines/Dataflows for all other data
- **Transformation**
  Notebooks
- **Storage**
  Lakehouse, or Lakehouse(s) with medallion structure (Bronze/Silver/Gold)
- **Exposure**
  In PBI via DirectLake or direct via Lake Warehouse leveraging SQL
- **Orchestration**
  Pipelines triggering Notebook runs and Power BI dataset refreshes



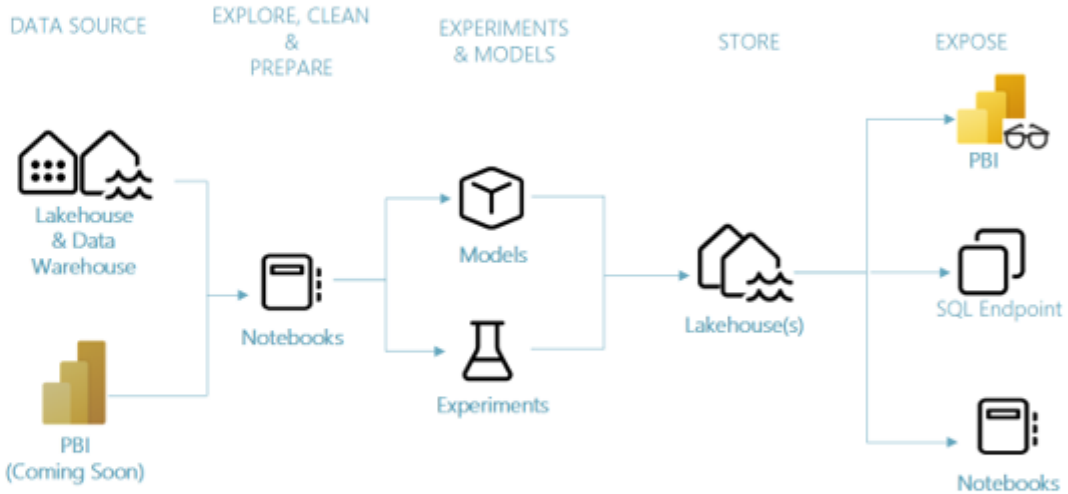
Data Warehouse

- **Ingestion**
  Mount Enabled Sources (DW Gen2, Azure SQL, Dataverse, etc), Pipelines/Dataflows for all other data
- **Transformation**
  Stored Procedures
- **Storage**
  Data Warehouse
- **Exposure**
  In PBI via Direct Lake or direct via Warehouse leveraging SQL
- **Orchestration**
  Pipelines triggering Stored Procedures and Power BI dataset refreshes



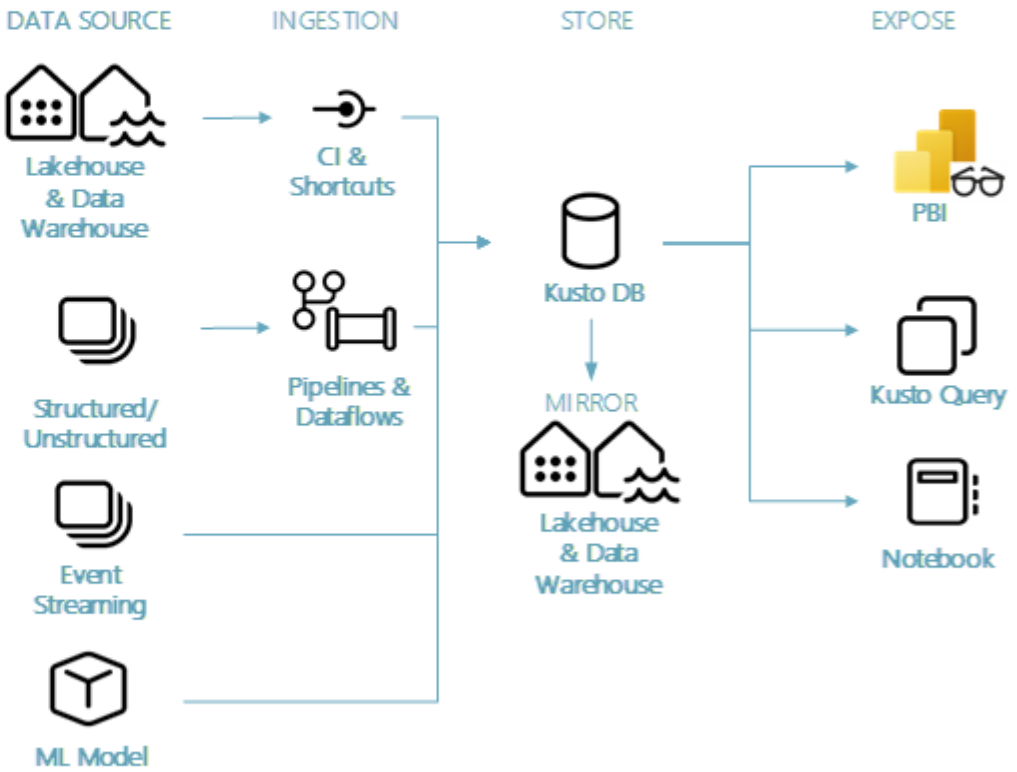
Data Science

- **Data Source:**
  - Access data from multiple sources – eg. Lakehouse and Data Warehouse
- **Explore, Clean & Prepare:**
  - Perform data transformation, exploration and featurization by leveraging built-in experiences on Spark with Python and Data Wrangler
- **Experiments & Models:**
  - Iterate, build and track Machine Learning experiments and models using MLFlow. Leverage data science capabilities for model prediction at scale (PREDICT) to gain and share business insights.
- **Storage:**
  - Store data and insights in Lakehouse(s) (reference DE scenario)
- **Expose:**
  - Collaborate with others by sharing your findings and insights via Notebook, PBI Report and Directlake.



Real-time Analytics

- **Ingestion**
  Lakehouse & Datawarehouse via Continuous Ingestion & Shortcuts (See LH/DW scenarios for populating LH/DW), Event Streaming (Event Hub, Kafka, etc) and ML Models connected directly, and Pipelines/Dataflows for all other data.
- **Storage**
  KQL Database, and Lakehouse/Data Warehouse via Mirroring
- **Exposure**
  In PBI via Direct Lake, Notebook via Spark/KQL, or direct via KQL Queryset



Tutorials

- Lakehouse tutorial - Microsoft Fabric | Microsoft Learn
- Data warehouse tutorial - Microsoft Fabric | Microsoft Learn
- Data science tutorial - Microsoft Fabric | Microsoft Learn
- Real-Time Analytics Tutorial - Microsoft Fabric | Microsoft Learn

Andrea Benedetti
Sr Cloud Architect, Microsoft
/in/abenedetti @anBenedetti https://github.com/anbened

#FabricGarage | 001 | 2024.05.16