

Lakehouse or Warehouse Dilemma

Andrea Benedetti
Sr Cloud Architect, Microsoft

 /in/abenedetti  @anBenedetti  <https://github.com/anbened>

#FabricGarage | 007 | 2024.06.27



Lakehouse or Warehouse Dilemma

- How to decide between Fabric Lakehouse / Fabric Warehouse?
- Primary decision:
 - team's skill set
 - type of data you're working with
- The choice isn't necessarily an either/or decision
 - They can be used together for a best-of-breed analytics strategy
 - [Better together - the lakehouse and warehouse - Microsoft Fabric | Microsoft Learn](#)





Data Lakehouse in Microsoft Fabric

- Open and infinitely-scalable storage in OneLake
- Versatile repository (structured, semi-structured, unstructured) — in a multitude of formats such as CSV, Parquet, JSON, or Delta
- Flexible and scalable Spark compute for data engineering and data science workloads
- Read-only SQL Analytics Endpoint provides a familiar relational layer on top of physical data in the Lakehouse





Data Warehouse in Microsoft Fabric

- Transactional data warehouse natively embracing an open standard format
- Data is stored in Delta – Parquet (no vendor lock-in)
- Is auto-integrated and auto-optimized with minimal knobs
- Extends full SQL ecosystem benefits
- Infinitely scalable and open
 - Physical compute resources assigned to jobs within milliseconds
 - Infinite scaling with dynamic resource allocation tailored to data volume and query complexity
 - Instant scaling up/down with no physical provisioning involved
 - Resource pooling providing significant efficiencies and pricing benefits



Lakehouse or Warehouse

- Data Types
 - Lakehouse is designed to handle structured, semi-structured, and unstructured data
 - Warehouse is primarily for structured data
- Developer Persona
 - Lakehouse is typically used by data engineers and data scientists
 - Warehouse is used by data warehouse developers and SQL engineers
- Developer Skill Set
 - Lakehouse if your team is more comfortable with Spark (Scala, PySpark, Spark SQL, R)
 - Warehouse if your team prefers T-SQL



Lakehouse or Warehouse

- Data Organization
 - Lakehouse: data is organized in folders and files, databases, and tables
 - Warehouse: data is organized in databases, schemas, and tables
- Multi-table Transactions
 - Lakehouse: no support
 - Warehouse supports multi-table transactions
- Scalability
 - Lakehouse / Warehouse use Fabric's compute bursting behind the scene to speed up the performance when needed

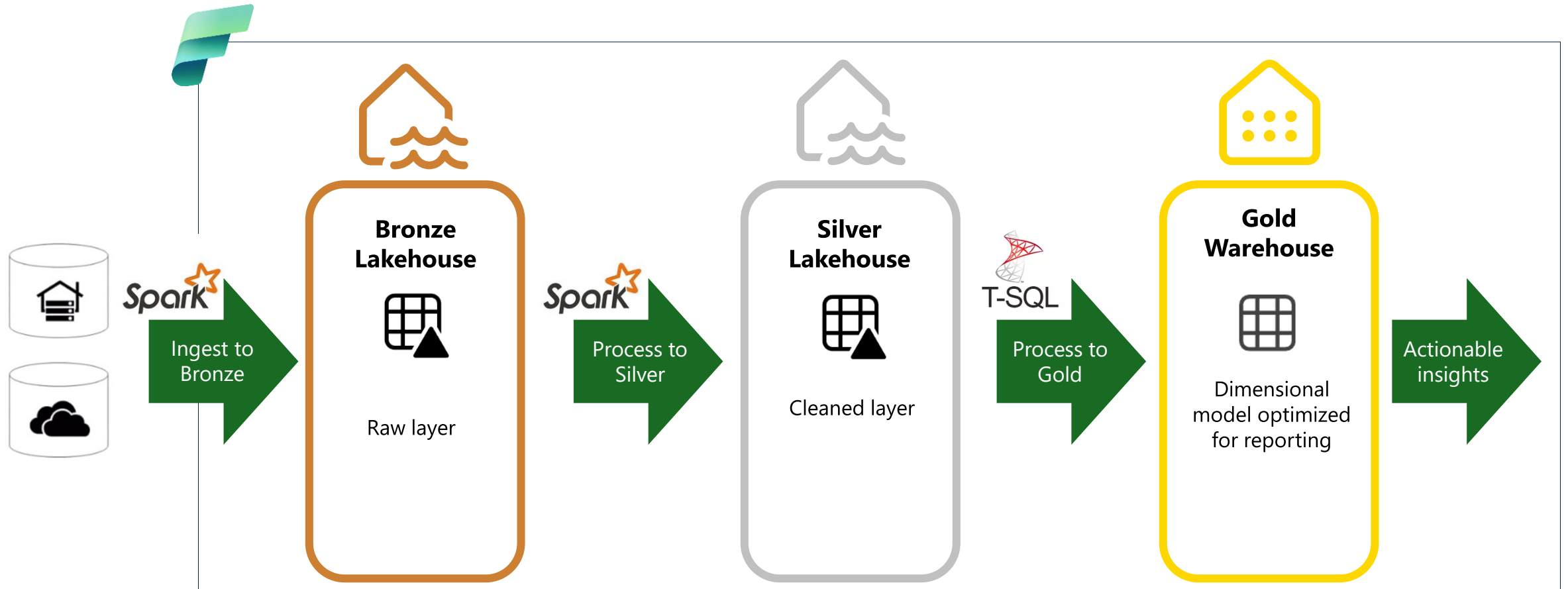


Microsoft Fabric decision guide: choose a data store

	Warehouse	Lakehouse
Data volume	Unlimited	Unlimited
Type of data	Structured	Unstructured, semi-structured, structured
Primary developer persona	Data warehouse developer, SQL engineer	Data engineer, data scientist
Primary developer skill set	SQL	Spark(Scala, PySpark, Spark SQL, R)
Data organized by	Databases, schemas, and tables	Folders and files, databases, and tables
Read operations	T-SQL, Spark (supports reading from tables using shortcuts, doesn't yet support accessing views, stored procedures, functions etc.)	Spark, T-SQL
Write operations	T-SQL	Spark(Scala, PySpark, Spark SQL, R)
Multi-table transactions	Yes	No
Primary development interface	SQL scripts	Spark notebooks, Spark job definitions
Security	Object level (table, view, function, stored procedure, etc.), column level, row level, DDL/DML, dynamic data masking	Row level, table level (when using T-SQL), none for Spark
Access data via shortcuts	Yes, through a lakehouse using three-part names	Yes
Can be a source for shortcuts	Yes (tables)	Yes (files and tables)
Query across items	Yes, query across lakehouse and warehouse tables	Yes, query across lakehouse and warehouse tables; query across lakehouses (including shortcuts using Spark)



Medallion Architecture using Lakehouse and Warehouse



Useful Links

- [Lakehouse vs Data Warehouse vs Real-Time Analytics/KQL Database: Deep Dive into Use Cases, Differences, and Architecture Designs | Microsoft Fabric Blog | Microsoft Fabric](#)
- [Choosing between Lakehouse and Warehouse in Microsoft Fabric | LinkedIn](#)





Andrea Benedetti
Sr Cloud Architect, Microsoft

 /in/abenedetti  @anBenedetti  <https://github.com/anbened>

#FabricGarage | 007 | 2024.06.27

