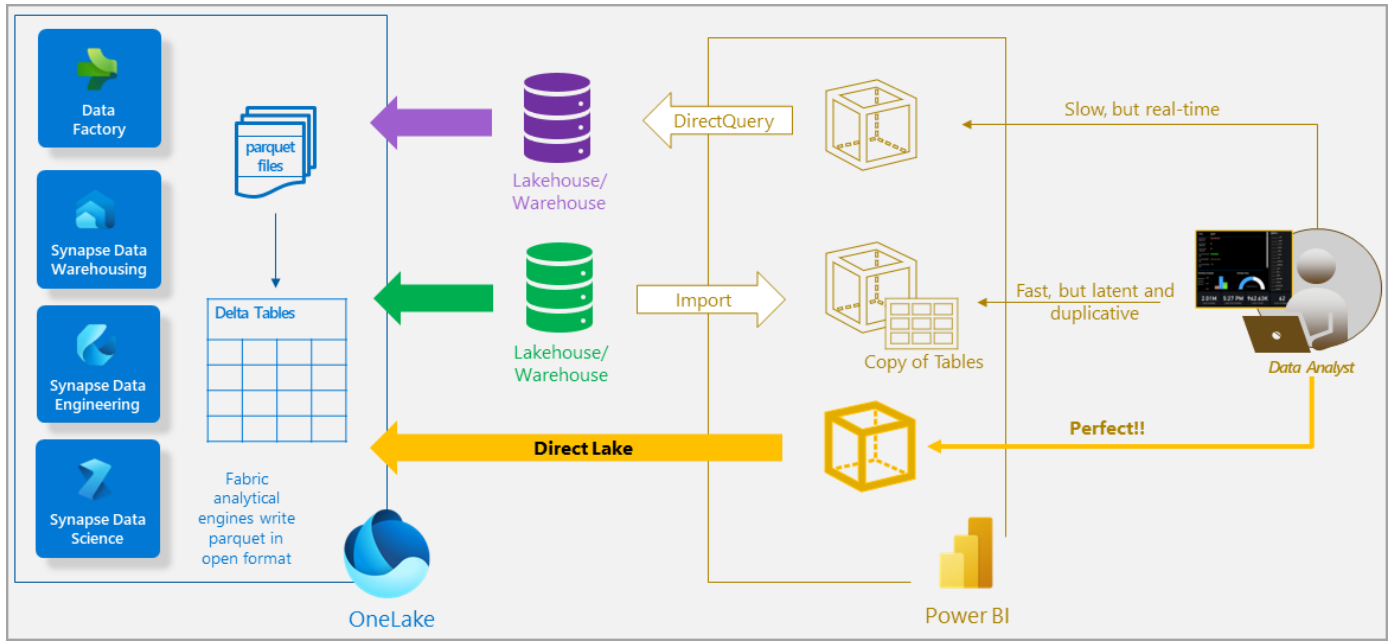
Direct Lake

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

#FabricGarage | 006 | 2024.06.20

Direct Lake

- A (new) semantic model capability for analyzing very large data volumes in Power BI
- Based on loading parquet-formatted files directly without:
  1. querying a lakehouse or warehouse endpoint
  2. having to import or duplicate data into a Power BI mode



Storage Modes

SMALLER MODELS	Time to Import Data	Model Size	Query Speed
Direct Query	-	-	?
Import	☺	☺	☺
Direct Lake	☺	☺	☺

LARGE MODELS	Time to Import Data	Model Size	Query Speed
Direct Query	-	-	?
Import	😐	😐	☺
Direct Lake	☺	☺	☺

Direct Lake

Considerations

- is not a replacement for Import mode
- is not a replacement for DirectQuery mode
- requires physical tables for the semantic model
  - by using a view, the engine uses DirectQuery
- models require more technical skills to achieve optimal compression
  - Compression in Import → VertiPaq (engine responsibility)
  - Compression in Direct Lake → how data is loaded in the Delta format (our responsibility)

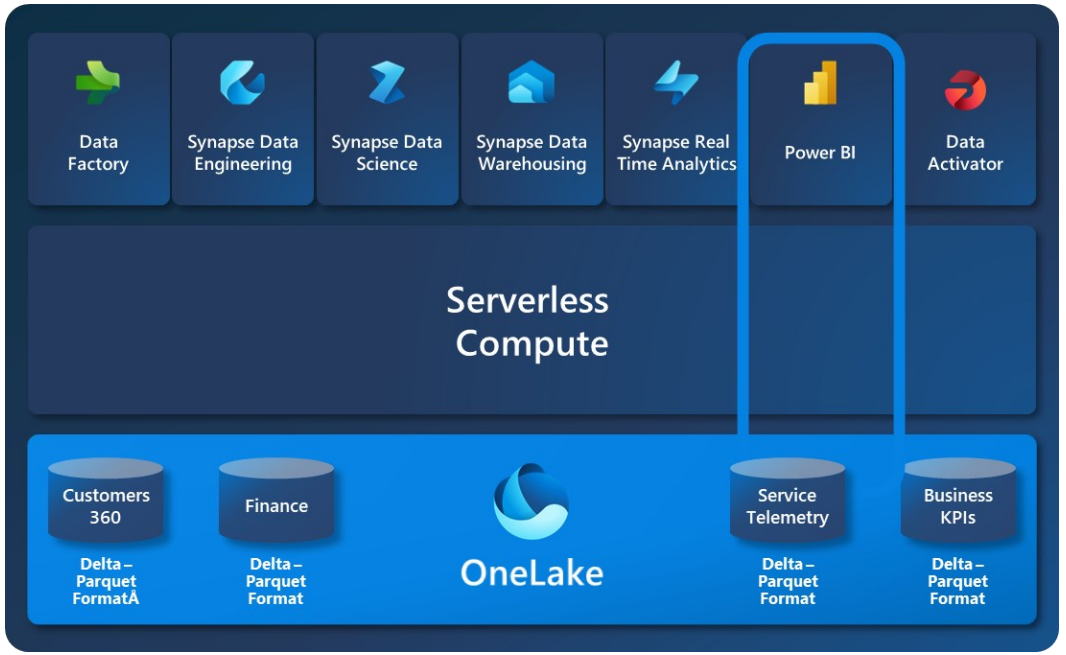
"Which one is “better” - Import or Direct Lake?"

- No one solution to "rule-them-all
- No answer

- Direct Lake conceptually works very similar to the Import mode

One Copy – Direct Lake

- The data is stored in a single common format
- Delta – Parquet, an open standards format, is the storage format
- 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



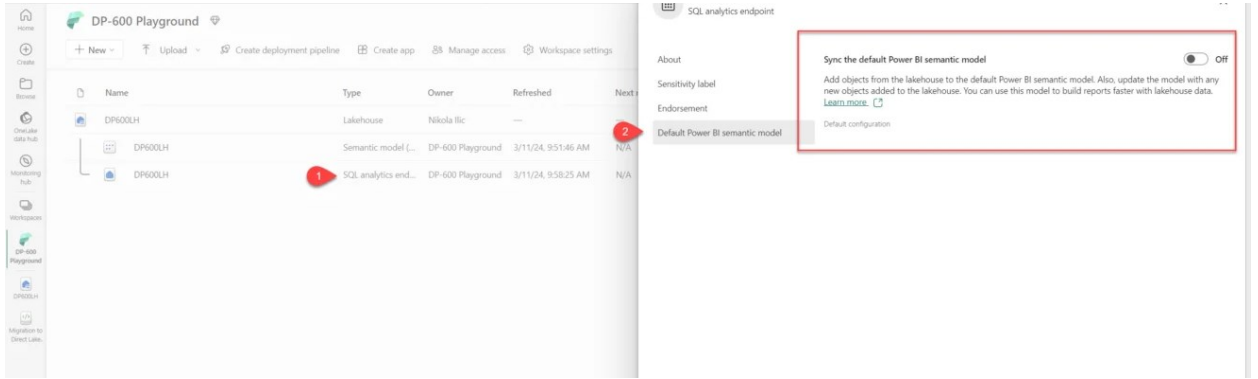
Framing

aka Direct Lake “refresh”

- What is framing
  - "point in time" way of tracking what data can be queried by Direct Lake
- Why is this important
  - Data consistency for some Power BI Reports
  - Delta-lake data is transient for many reasons
- ETL Process
  - Ingest data to delta lake tables
  - Transform as needed using preferred tool
  - When ready, perform *Framing* operation on semantic model
- Framing is near instant and acts like a cursor
  - Determines the set of .parquet files to use/ignore for *transcoding* operations

Syncing

- Syncing = Adding new tables to a semantic model
  - By default, this table WILL NOT be automatically included
  - This setting allows you to define what happens when a new table arrives at a lakehouse / dwh



Fallback to DirectQuery

- If Direct Lake can’t retrieve results from Delta tables using a Direct Lake mode, the query will by default fall back to a DirectQuery mode

- You are using features that prevent Direct Lake
  - Exceeds the memory limit of the SKU (F64 → 25 GB)
  - Views in Fabric Warehouse
  - RLS or OLS is defined in a Warehouse

Fallback to DirectQuery – data volumes

- There are limits on how much data can be used for Direct Lake
  - F64/P1 → 1.500.000 rows
- These limits vary by capacity SKU size
- If you exceed these limits, Direct Lake will use DirectQuery
  - Query performance may be noticeably worse
- Fabric checks limits during reframing process
- Can be turned On/Off using Direct Lake Behaviour property

Useful Links

- Learn about Direct Lake in Power BI and Microsoft Fabric - Microsoft Fabric | Microsoft Learn
- Direct Lake vs. Import mode in Power BI – SQLBI
- What does it mean to refresh a Direct Lake Power BI dataset in Fabric? (crossjoin.co.uk)
- 50 Shades of Direct Lake – Everything You Need to Know About the New Power BI Storage Mode! | LinkedIn

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

#FabricGarage | 006 | 2024.06.20