
SQL Saturday San Diego, CA


Microsoft


AMD


databricks


O'REILLY


COZYROC


SQL
skills
immerse yourself in sql server


University of San Diego
ONLINE


Tessell

November 17-21, 2025

PASS Data Community Summit

See you in Seattle!

Get $150 off a 3-day ticket!

Use code: SQLSATATX150

Register now!

passdatacommunitysummit.com

Direct Lake & Data Agents

Best Practices
&
Lessons Learned From Past Year

Chaudhary Ahmad Ali

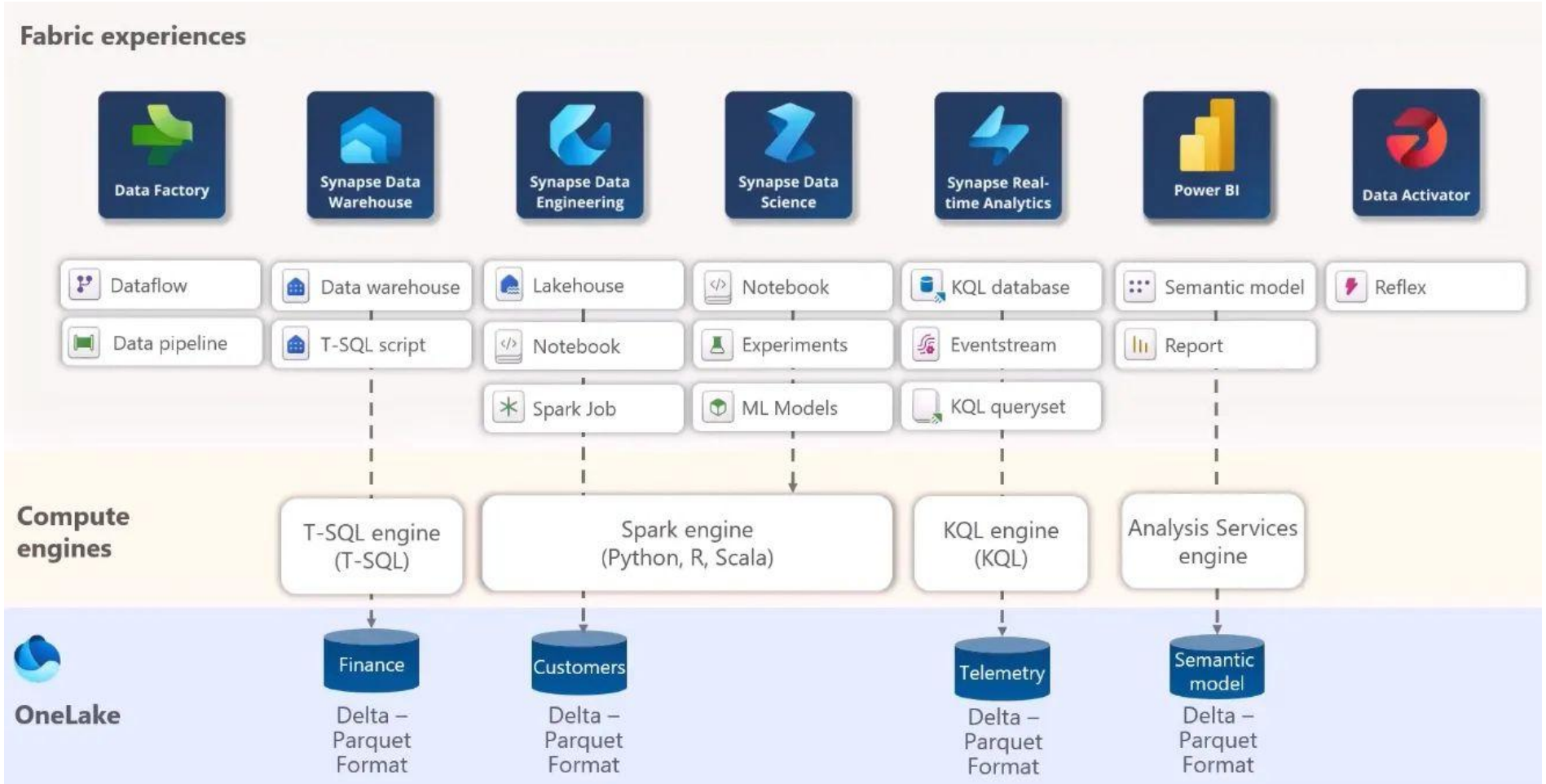
Chaudhary Ahmad Ali

- Current working as Chief Architect at iLink Systems
- 24 plus years of experience in IT
  - 5 years as an SE, 10 years as a DE, and 10 years as a BI Engineer
- Recognitions & Certifications
  - Fast Track Recognized Solution Architect (FTRSA 2022-2023)
  - TOGAF Certified
  - PL-300, DP-500, DP-600, DP-700
- Favorite Movies
  - Meet Joe Black, Shawshank Redemption, Forrest Gump, Legends of the Fall
- Favorite Sports and Hobbies
  - Pickleball, Tennis, Billiard, spending time with family, traveling, and food

Agenda

- Direct Lake (Benefits & Limitations)
- Direct Query Vs Import Vs Direct Lake
- Power BI Development Framework (Direct Lake)
- Data Agents
- Direct Lake Best Practices
- Key factors to consider when selecting the appropriate storage mode
- Upcoming Direct Lake related features
- Q & A

Microsoft Fabric



Microsoft Fabric Visualize: Image Credit

Direct Lake

- **Loads only on demand data into memory and evicts unused data gradually**

- **Direct Lake Modes**
  - Direct Lake on One Lake (Composite Table, One Lake Security, Calculated Tables, Across Lakehouse sourcing)
  - Direct Lake on SQL endpoint

Direct Lake Concepts

- Fabric (Or Premium) Capacity
- Parquet Files dependency
- V-Order
- Data Framing
- Data Transcoding
- Cold and Warm Cache
- Column Temperature
- Direct Query Fall Back
- Fixed Identity Recommended

Direct Lake Benefits

- Minimize the data latency
- Can handle large data volumes with much better performance as compared to Direct Query.
- Saves the semantic model refresh time and compute cost
- Semantic Models & Reports availability time is reduced

https://lytix.be/what-is-direct-lake-mode-in-fabric/

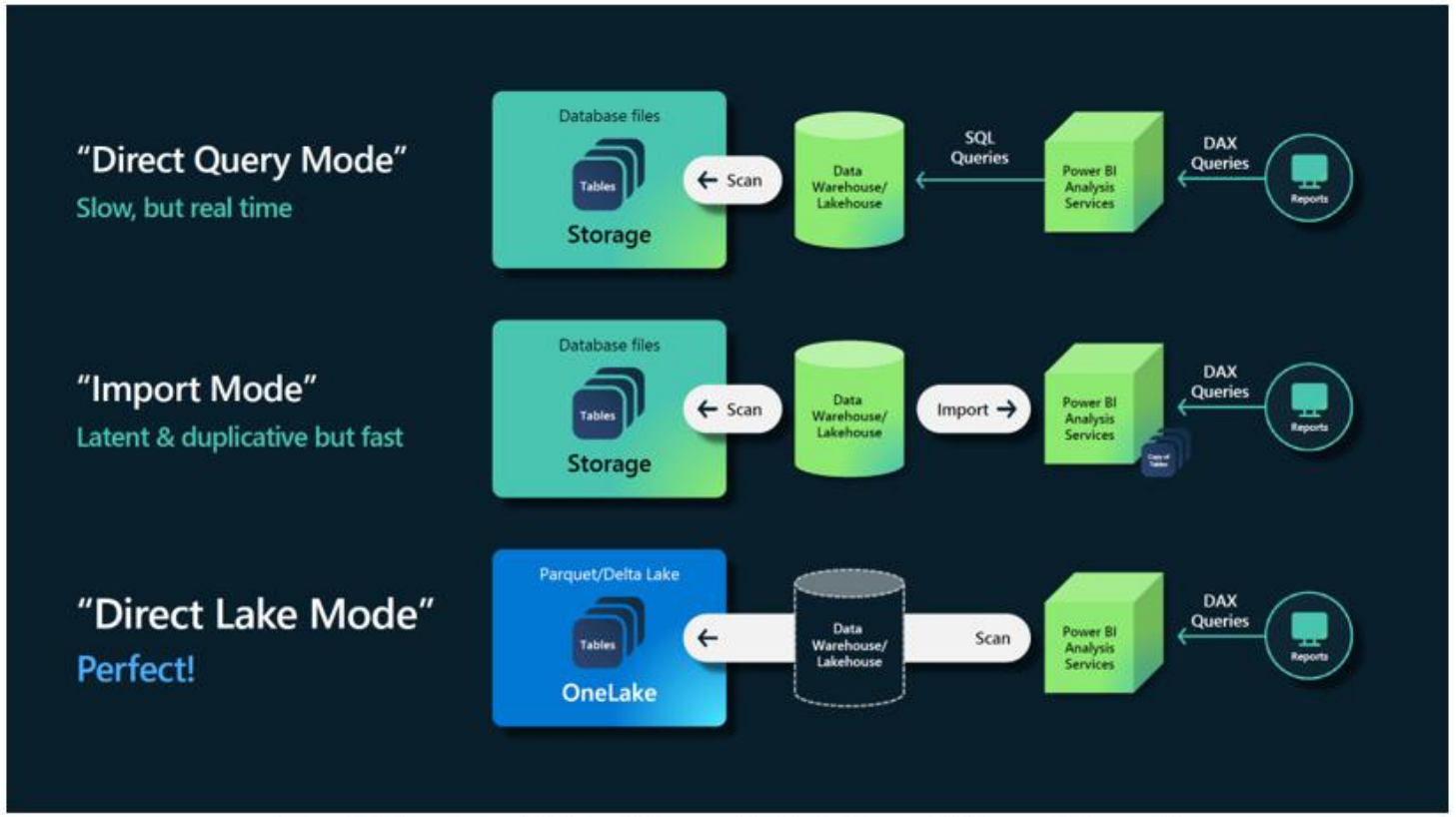


Figure 2: Overview of different Storage Modes in Power BI (Source: Microsoft)

Data Loading in Direct Lake and Import

- Direct Lake uses the VertiPaq engine, the same engine employed by Import mode, to load Delta table formats.
- The columnar compression logic and file structure of Delta tables are like the proprietary format in Import mode.
- Direct Lake involves data transcoding during the initial load into memory, which takes more time compared to Import mode, where data is directly loaded without transcoding.
- If data is evicted from memory in Direct Lake, it must be transcoded again, leading to slower reload times compared to Import mode. Additionally, data eviction is more aggressive in Direct Lake.
- When Direct Lake reaches its resource limits, it can fall back to Direct Query mode (in DL on endpoint), whereas Import mode may encounter out-of-memory or resource errors.

DQ Vs Import Vs Direct Lake

Direct Query

Refers data in source

Pros

- Real Time Analysis
- Can handle large data volumes

Cons

- Report Rendering is Slow
- Dependent on data source optimization
- Limited support in complex DAX functions

Import

Loads all the data in memory upon Refresh

Pros

- Report Rendering is best
- Can handle complex calculations

Cons

- Data Latency
- Processing Complexity
- Processing Time

Direct Lake

Loads in use data in memory & gradually evicts unused data

Pros

- Real Time Analysis
- Can handle large data volumes
- Report rendering performance is fast.

Cons

- Row Volume Limits
- High cardinality columns initially loads slowly
- Works with Parquet only

Direct Lake Limitations

- Parquet Files per Delta Table
- Row Groups Per Delta Table
- Rows per table
- Max Memory (Including Warm Cache + Operational + Query Serving)
- No RLS/OLS in LH/WH
- No SQL Views
- No Excel Hierarchies & Drill throughs

Fabric/Power BI SKUs	Parquet files per table	Row groups per table	Rows per table (millions)	Max model size on disk/OneLake[1] (GB)	Max memory (GB)
F2	1,000	1,000	300	10	3
F4	1,000	1,000	300	10	3
F8	1,000	1,000	300	10	3
F16	1,000	1,000	300	20	5
F32	1,000	1,000	300	40	10
F64/FT1/P1	5,000	5,000	1,500	Unlimited	25
F128/P2	5,000	5,000	3,000	Unlimited	50
F256/P3	5,000	5,000	6,000	Unlimited	100
F512/P4	10,000	10,000	12,000	Unlimited	200
F1024/P5	10,000	10,000	24,000	Unlimited	400
F2048	10,000	10,000	24,000	Unlimited	400

Direct Lake

Power BI Dev Framework (Direct Lake)

Analyze Data Sources

- Data Profiling
- Analyze Grain
- Understand Calendar Dim
- Prepare Bus Matrix
- Data Retention
- Data Refresh Policy

Load Data into LH/WH

Bronze Layer
- Bring Data in Original Format
- Append Data only
- Add metadata columns like Load Date and Load Id etc.

Silver Layer
- Data Cleansing
- Data Standardization
- Add Surrogate Keys
- Implement SCD if any

Gold Layer
- Implement Star Schema
- Apply Z-Order & V-Order

Semantic Model

- Refer Gold Tables & create relations
- Develop KPI & Hierarchies
- RLS / OLS
- Add Business friendly names, descriptions, & synonyms

Performance Optimization

- Vertipaq Analyzer
- Best Practice Analyzer
- Semantic Link/Labs
- Delta Analyzer
- Performance Analyzer
- Query Plans

Reports

- Prepare Themes
- Prepare Templates
- Avoid Large Data Exports
- Promote Analyze in Excel

Model Refresh

- Trigger Smart Model Refresh (Reframing)
- Pre Warm Cache
- Send Notification

Fabric Data Agents

- **Chat & Talk to your Data in Onelake (Lakehouse, Warehouse, Eventhouse, Semantic Model)**
- **Conversation context** is preserved even in scenarios involving disparate data sources
- Create **Endpoints** which allows them to be consumed within Fabric or outside, with services like CoPilot Studio, Azure AI Foundry and custom applications (coming soon).
- Existing security rules, such as Row-Level Security (RLS) and Column-Level Security (CLS) in One Lake sources, will be maintained and respected
- Can maintain, deploy and evaluate Data Agents programmatically using SDK
- Now supported in all paid F-SKU (F2 and above)

Fabric Data Agents (Semantic Model)

- **For Semantic Models type data sources, Adding instructions (up to 15,000 characters) is the most important configuration)**
- Natural Language to DAX (NL2DAX) is used to create DAX queries from conversations
- Establishing business-friendly table and column names, along with detailed descriptions and relevant synonyms, can significantly enhance the accuracy of DAX query generation.
- Granting access to Data Agents for other users also requires sharing semantic model-level access with the same users

Fabric Data Agents with Semantic Models

Semantic Model

- Star Schema
- Business Friendly Table, Column and Measure Names
- Define Descriptions
- Provide Synonyms

Data Agents

Provide rich instructions

- Extract detail metadata from semantic model and can update the DA instructions using semantic Link Lab and DA SDK

Evaluate Data Agents regularly

DA Integration

- Within Fabric
- Copilot Power BI
- Azure AI Foundry
- Teams
- Copilot Studio (soon)

Direct Lake Best Practices (Gold Layer- LH)

- Implement Star Schema compliant tables
- Keep lean columns & rows, assign lean data type to columns
- Apply Z-Order and V-Order to fact tables to get best data storage and retrieval performance, e.g., Yellow Taxi Data 3 billion rows, CSV (**420 GB**) → Parquet (**168 GB**) → Delta with V-Order (**64 GB**)
- Parquet files from external sources such as Databricks or Snowflake are likely to perform slower when compared to V-Order optimized Parquet files in Fabric,
- Regular maintenance of gold layer delta tables with Optimize and Vacuum commands
- Configure higher bin size e.g. 500 MB to 1 GB in spark.conf.set("spark.databricks.delta.optimizeWrite.binSize“ for partitioned tables

Direct Lake Best Practices (Semantic Model)

- Refer Star Schema delta tables in Gold Layer
- Avoid many to many and bidirectional relations
- Apply business user friendly table and column names. Consider defining synonyms, descriptions to make model CoPilot and Data Agent friendly.
- If high cardinality columns are occupying memory, consider following approaches:
  - Split columns e.g., datetime into data and time columns separately
  - Consider Precision points up to 3 point maximum in decimal type columns

Direct Lake Best Practices (Reports)

- Beware of Visual Rending Time out setting (i.e. 225 seconds) on Fabric service is different than on Power BI desktop. Working visual in Power BI Desktop might fail in service in a case when data is first time loading/transcoding using direct lake and related column size is very high.
- Workaround, Pre-Warm the cache always using notebook scripts
- Leverage and Promote Data Agents for simple operational and analytical type business questions over reports.

Direct Lake Best Practices (Security)

- Implement RLS or OLS in semantic model
- Using a Service Principal Name (SPN) to connect the Direct Lake semantic model to Lakehouse is highly recommended, particularly when Row-Level Security (RLS) is applied (One Lake Security will change this recommendation).

Direct Lake Best Practices (Refresh)

- Refreshing Direct Lake semantic model is also known as data reframing. Reframing is low-cost operation to reference latest version of dela tables.
- Consider turning off auto direct lake refresh under semantic model settings.
- Trigger semantic model refresh using Fabric Data Factory or Notebook. Consider table-level refresh when multiple subject areas in the semantic model follow different data loading schedules.
- Identify most commonly used metric columns, Pre Warm the cache for these columns. This will enhance the user experience even when they run the report for the first time.

Key Factors for selecting the Storage Mode

- Key Factors to consider before selecting the appropriate storage mode in Power BI:
  - Data Size (compressed and fact table row volumes)
  - Real Time Analytics required, justified by genuine business requirements
  - Report/Dashboards Rendering SLA
  - Report and Calculation Complexity (calculated tables or columns)
  - Data Source type
  - Is it an IT Driven Analytical Project?

Roadmap of Direct Lake and Data Agents

- Calculated Tables and Calculated Columns
- Add Delta tables from across the Lakehouse and across workspaces
- Implement Import and Direct Lake together in a same model
- Materialized Views support in Lakehouse
- One Lake Security
- Data Agents Integration with Copilot Studio
- More Sources will be supported with Data Agents like unstructured data sources or mirrored data sources.
- Org Apps support for the Data Agents
- (More)

Conclusion

- Direct Lake is another storage mode, which cannot replace existing storage modes.
- Consider Import Mode by default unless capacity memory limits are reached
- Consider Direct Lake when columns can fit within memory limits, while Import mode exceeds capacity constraints. Regularly track memory usage of Direct Lake models through DMV storage views.
- Choose Direct Query when the model exceeds memory limits and real-time analytics are essential.
- Direct Query is the slowest in terms of performance among the three options. Direct Lake performs nearly on par with Import mode, but Import consistently delivers the best performance overall.

Q & A

Thank You

Presentation Link