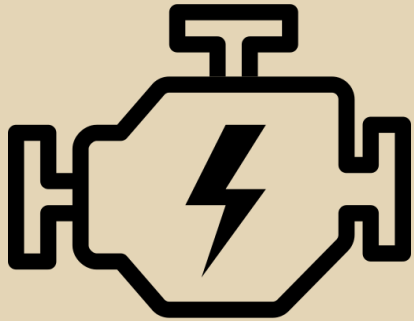


A deep dive into Direct Lake



After this session

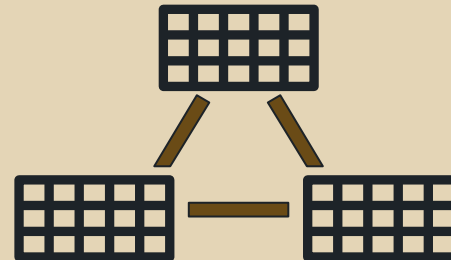
Understanding the engine
for Direct Lake



Analyzing performance
across different scenarios



Data modeling benefitting
from Direct Lake



Go next level with
advanced patterns



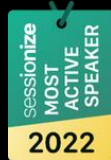
Assumptions

aka Qualified guessing



Marc Lelijveld

Technical Evangelist | Solution Architect
Macaw Netherlands



@MarcLelijveld



linkedin.com/in/MarcLelijveld



Data-Marc.com

FAVORITE STUFF:





Storage modes

Different types of storage modes

Three familiar storage modes

- **Import** – data cached in the model
- **DirectQuery** – queries are submitted to the back-end data source
- **Dual** – can act in both above storage modes, depending on query context

The screenshot displays the Microsoft Power BI Desktop interface. The top ribbon includes tabs for File, Home, Help, and External Tools. The Home tab is active, showing various tool groups like Clipboard, Data, Queries, Relationships, Security, and Q&A. The main workspace shows a data model with four tables: Product Subcategory, Product, Internet Sales, and Internet Sales - Agg. The Product Subcategory table is connected to the Product table (1 to * relationship). The Product table is connected to the Internet Sales table (1 to * relationship). The Internet Sales table is connected to the Internet Sales - Agg table (1 to * relationship). The Properties pane on the right shows the 'General' tab for the selected 'Internet Sales' table. The 'Storage mode' dropdown is set to 'DirectQuery'.

Product Subcategory

- English Product Subcategory N...
- French Product Subcategory Na...
- ProductCategoryKey
- ProductSubcategoryAlt...
- ProductSubcategoryKey
- Spanish Product Subcategory N...

Product

- Arabic Description
- Chinese Description
- Class
- Color
- Days To Manufacture
- Dealer Price

Internet Sales

- Carrier Tracking Number
- CurrencyKey
- Customer PO Number
- CustomerKey
- Discount Amount
- DueDateKey
- Extended Amount
- Freight Amount
- Order Calendar Year

Internet Sales - Agg

- Count
- Order Calendar Year
- ProductSubcategoryKey
- Sales Amount

Properties

General

Name: Internet Sales

Description: Enter a description

Synonyms: Enter a comma-separated list of synonyms for Q&A

Row label: Select a row label

Key column: Select a column with unique values

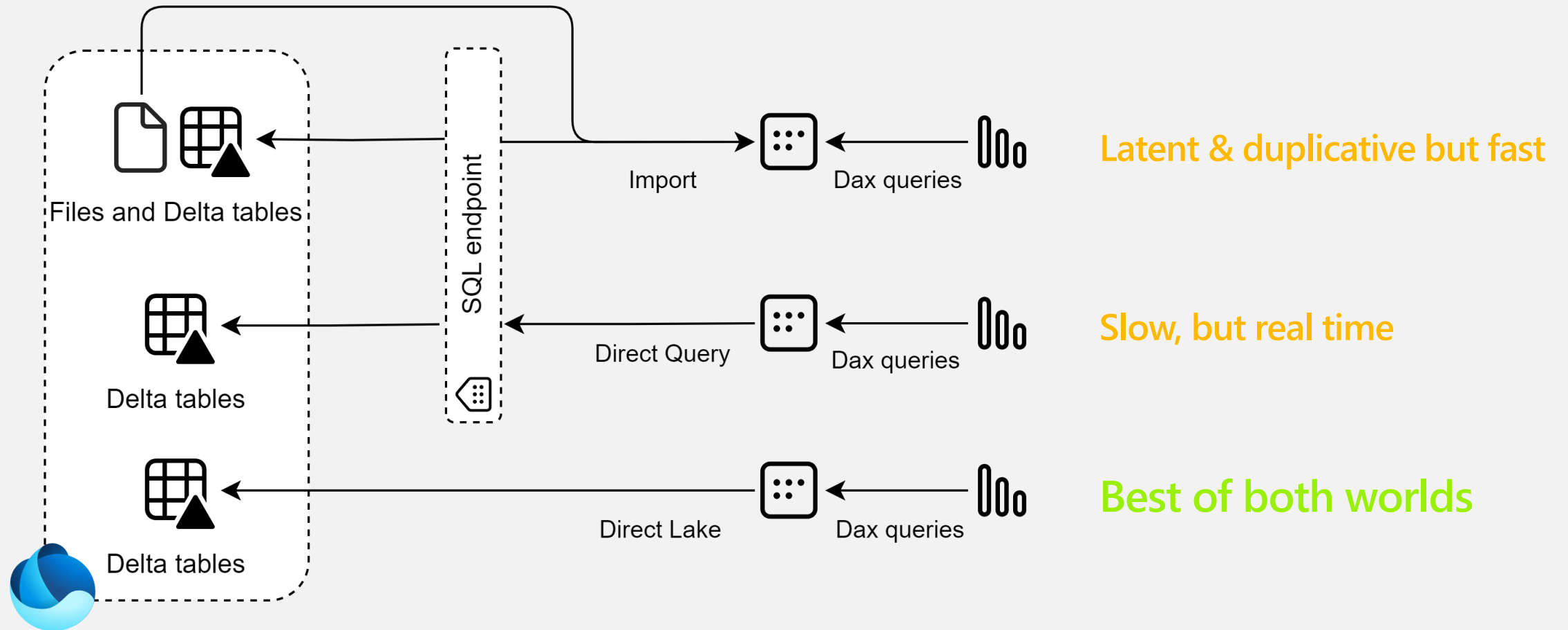
Is hidden: No

Is featured table: No

Advanced

Storage mode: DirectQuery

Direct Lake



Direct Lake is only applicable to Fabric



Data Factory



Synapse Data Engineering



Synapse Data Warehouse



Synapse Data Science



Synapse Real-Time Analytics



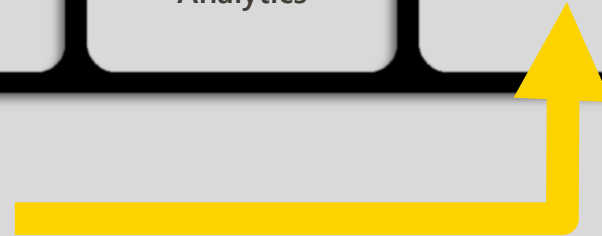
Power BI



Data Activator
(coming soon)



OneLake





Build your data model benefitting Direct Lake

Data transformations



No Power Query or other data transformation capabilities*



Data transformations should be done as far upstream as possible

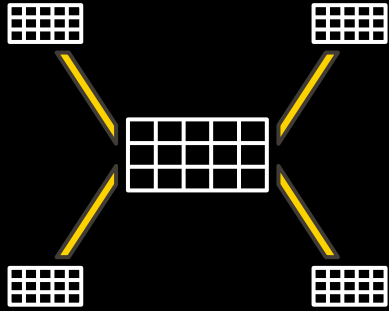


Data transformation directly in the Lakehouse unlocks “new” possibilities

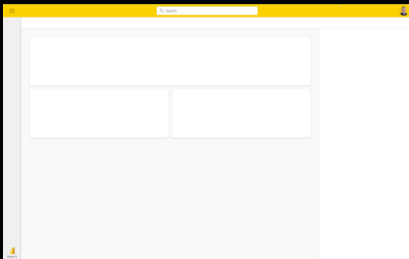
Medallion architecture



Data modeling



General best practice to have a star schema still applies



Web (browser) experience only to develop data models

Data modeling best practices unchanged

- Starschema all the things!
- Avoid bi-directional or many-to-many relationships
- Avoid limited relationships
- Implement role-playing dimensions rather than duplication
- Minimize redundant measure using calculation groups
- Avoid ambiguous data models

... etcetera

Demo





Internals & performance

Delta (Parquet)



Parquet (File format)

- Column-store
- Open industry standard
- Compressed & Encoded
- Parallelism

Enables fast bulk operations of large data volumes

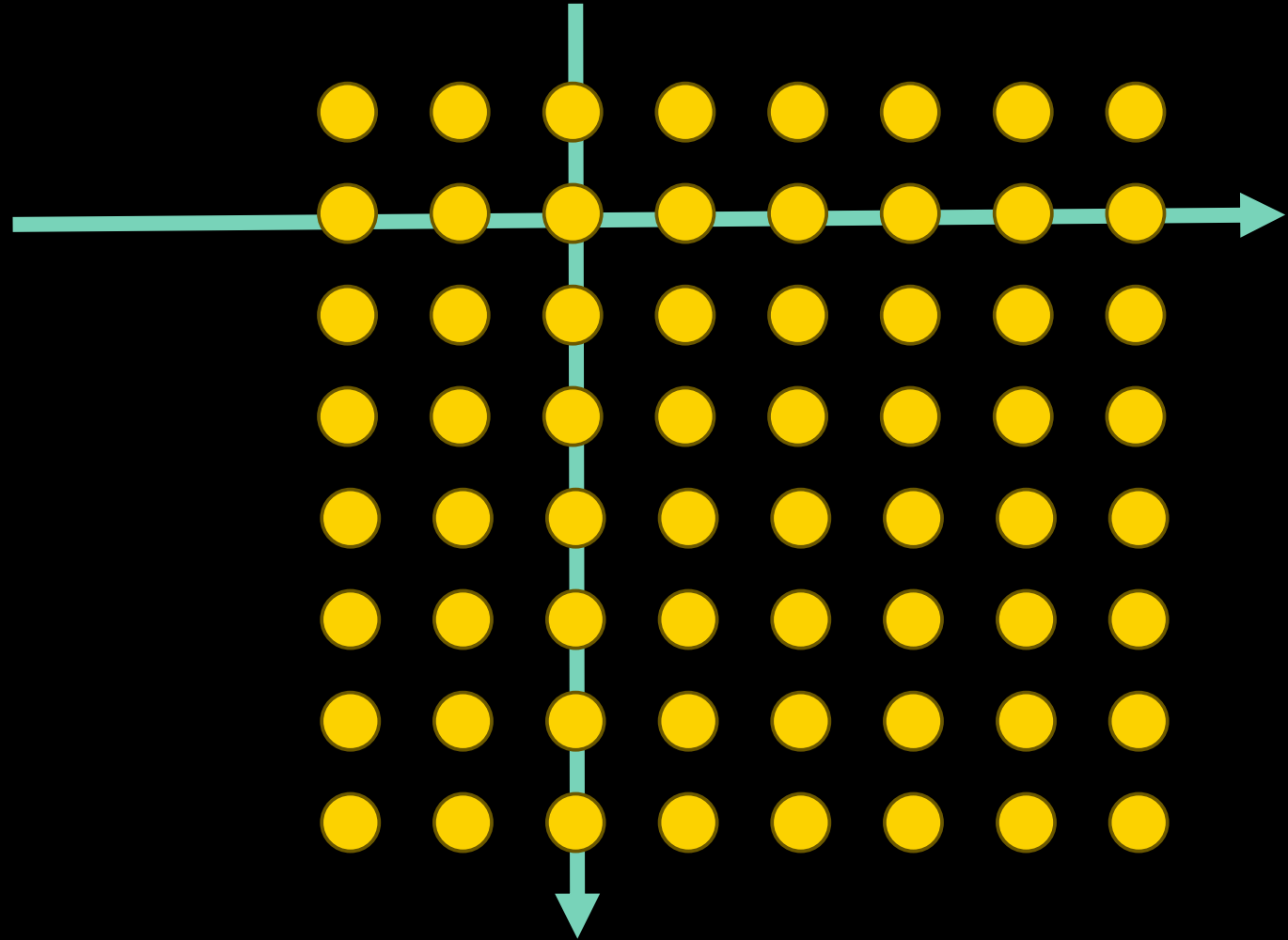
Delta (Storage layer management)

- Also, an open industry standard
- ACID transactions & schema enforcement
- Delete, update, merge
- Time-travel
- Optimized for querying, skipping and pruning

Brings warehouse reliability to the data lake

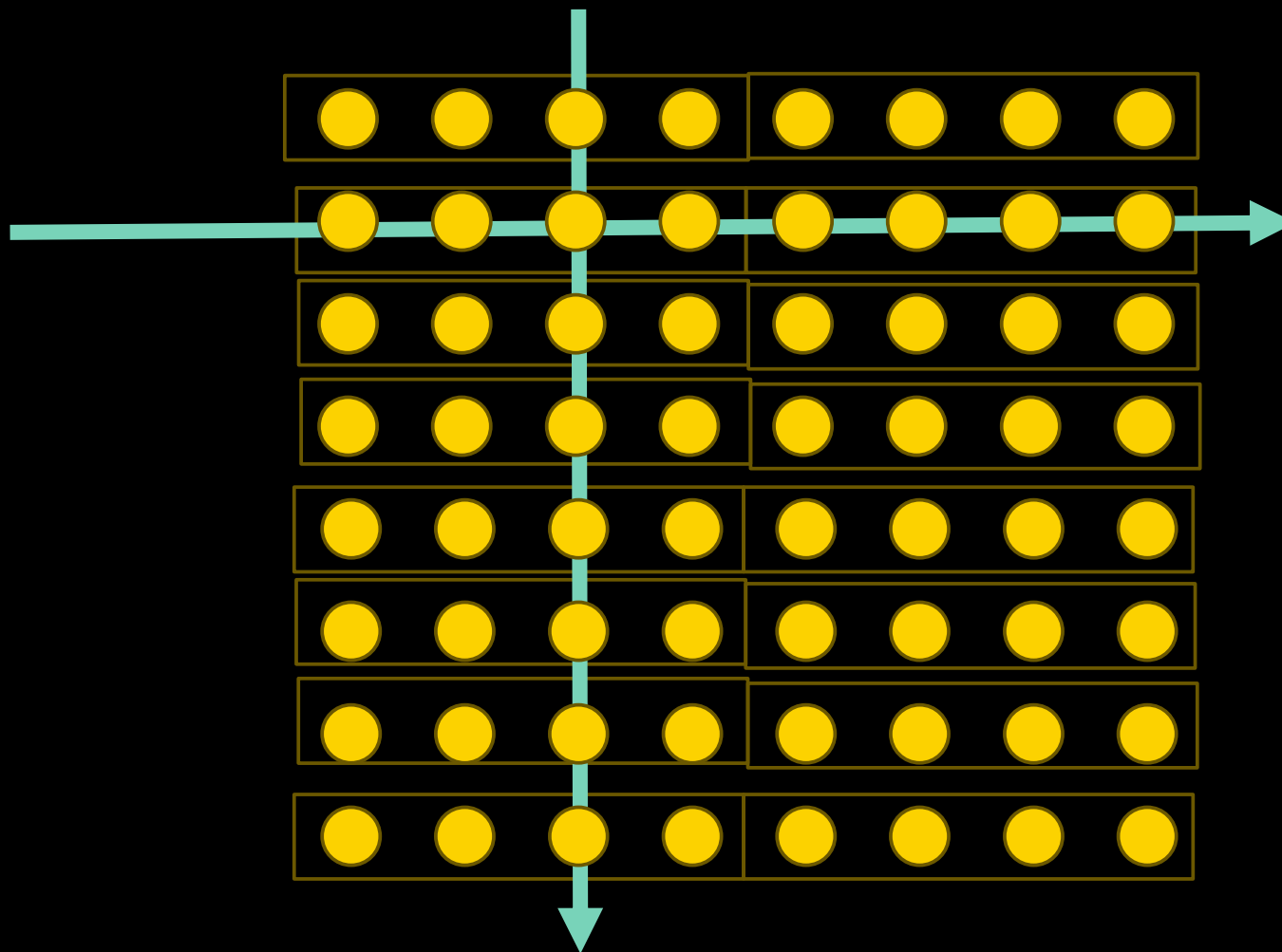
From Z-order to V-order

SELECT * FROM points WHERE x=3 or y=2



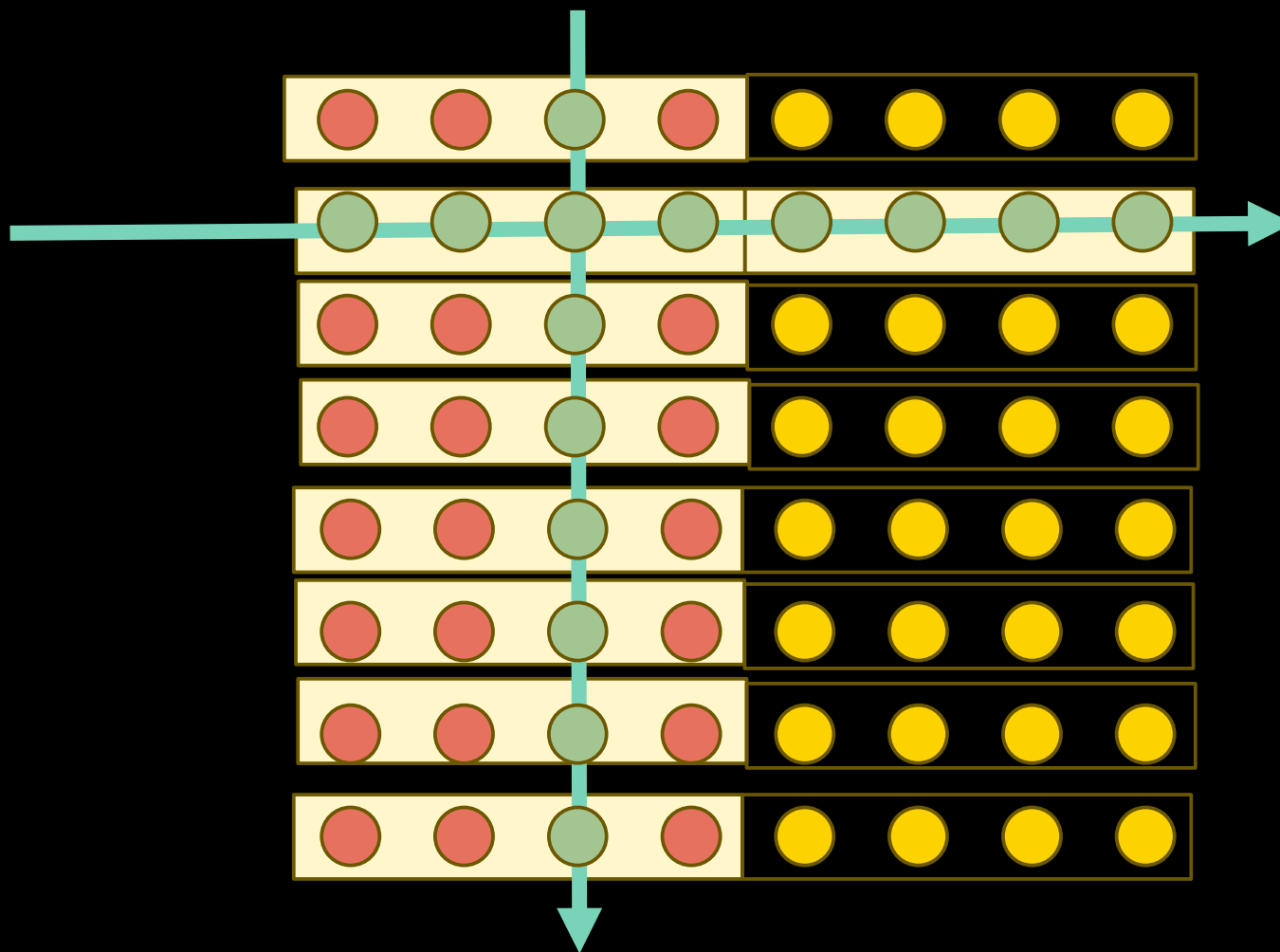
From Z-order to V-order

SELECT * FROM points WHERE x=3 or y=2



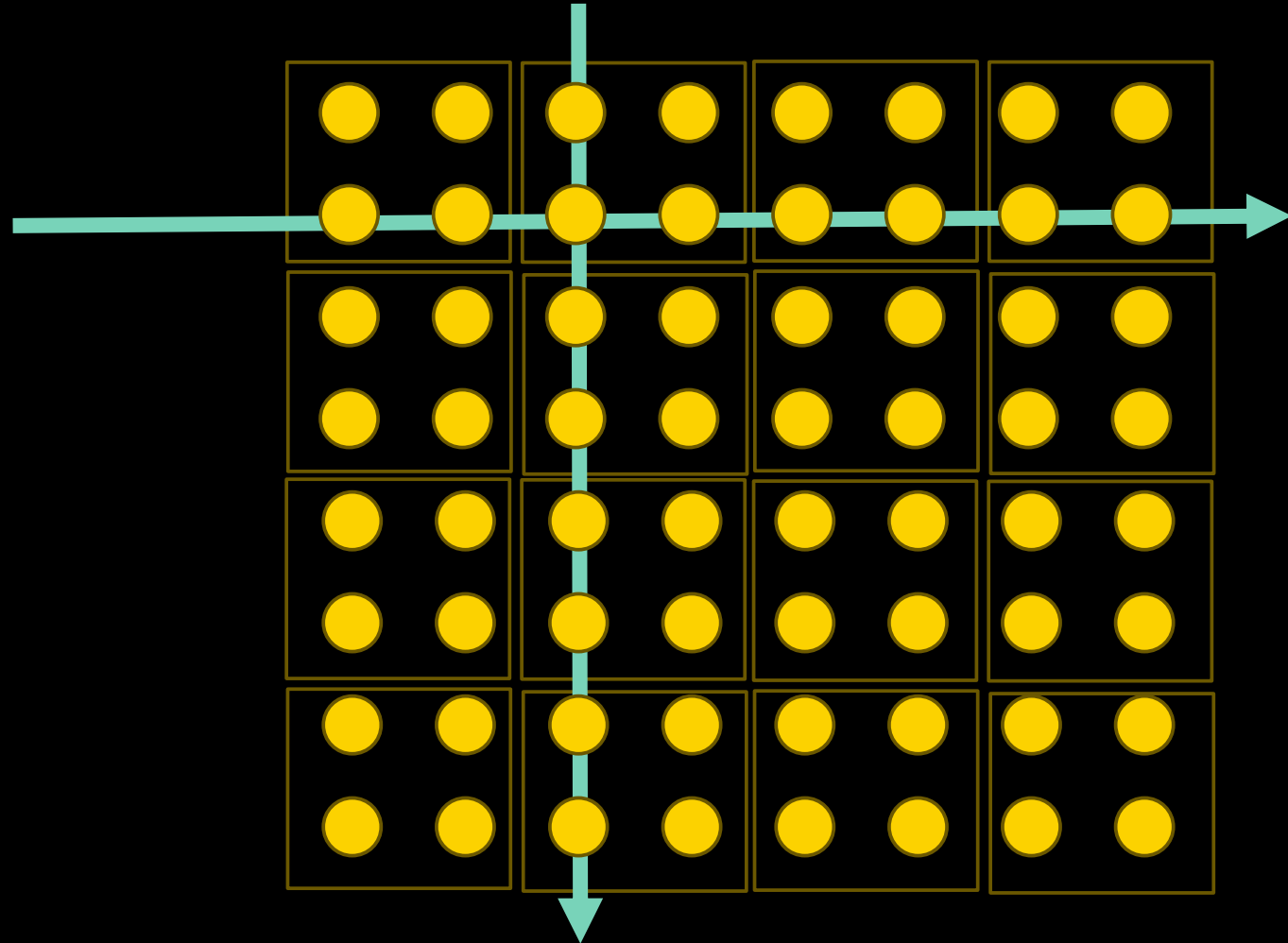
From Z-order to V-order

SELECT * FROM points WHERE x=3 or y=2



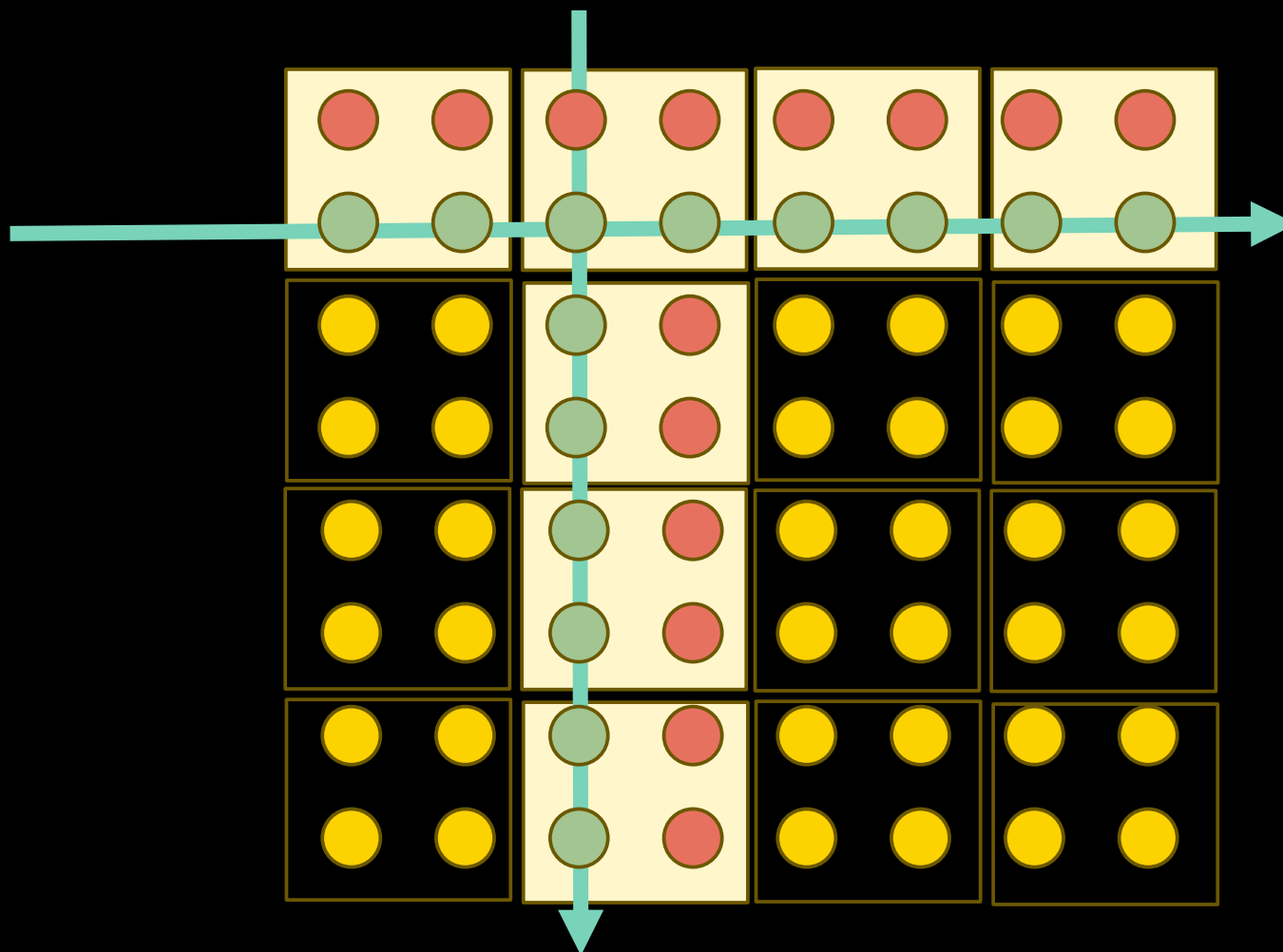
From Z-order to V-order

SELECT * FROM points WHERE x=3 or y=2



From Z-order to V-order

SELECT * FROM points WHERE x=3 or y=2



From Z-order to V-order

Yellow taxi (3 Billion rows)



416 GB



164 GB



V-order
60GB

x3.2

Less I/O for all *
workloads

* Microsoft benchmark

Data saving

The Analysis Services column-oriented storage using Delta Lake/Parquet open standard for Direct Lake

SSAS, AAS, Power BI large models

The image shows a screenshot of the SQL Server Enterprise Manager interface. The top pane displays a file explorer view of the data files for a Direct Lake model. The bottom pane shows the 'Finance' database structure, including a 'Tables' list and a detailed view of the 'Revenue' table.

File Explorer View (Top Pane):

Name	Date modified	Type	Size
1.DimCustomer (10).AddressLine1 (78).0.idf	1/29/2020 6:36 PM	IDF File	37 KB
1.DimCustomer (10).AddressLine1 (78).0.idfmeta	1/29/2020 6:36 PM	IDFMETA File	1 KB
1.DimCustomer (10).AddressLine2 (79).0.idf	1/29/2020 6:36 PM	IDF File	
1.DimCustomer (10).AddressLine2 (79).0.idfmeta	1/29/2020 6:36 PM	IDFMETA File	
1.DimCustomer (10).BirthDate (66).0.idf	1/29/2020 6:36 PM	IDF File	
1.DimCustomer (10).BirthDate (66).0.idfmeta	1/29/2020 6:36 PM	IDFMETA File	
1.DimCustomer (10).CommuteDistance (82).0.idf	1/29/2020 6:36 PM	IDF File	
1.DimCustomer (10).CommuteDistance (82).0.idfmeta	1/29/2020 6:36 PM	IDFMETA File	

Database Structure View (Bottom Pane):

Finance

Tables

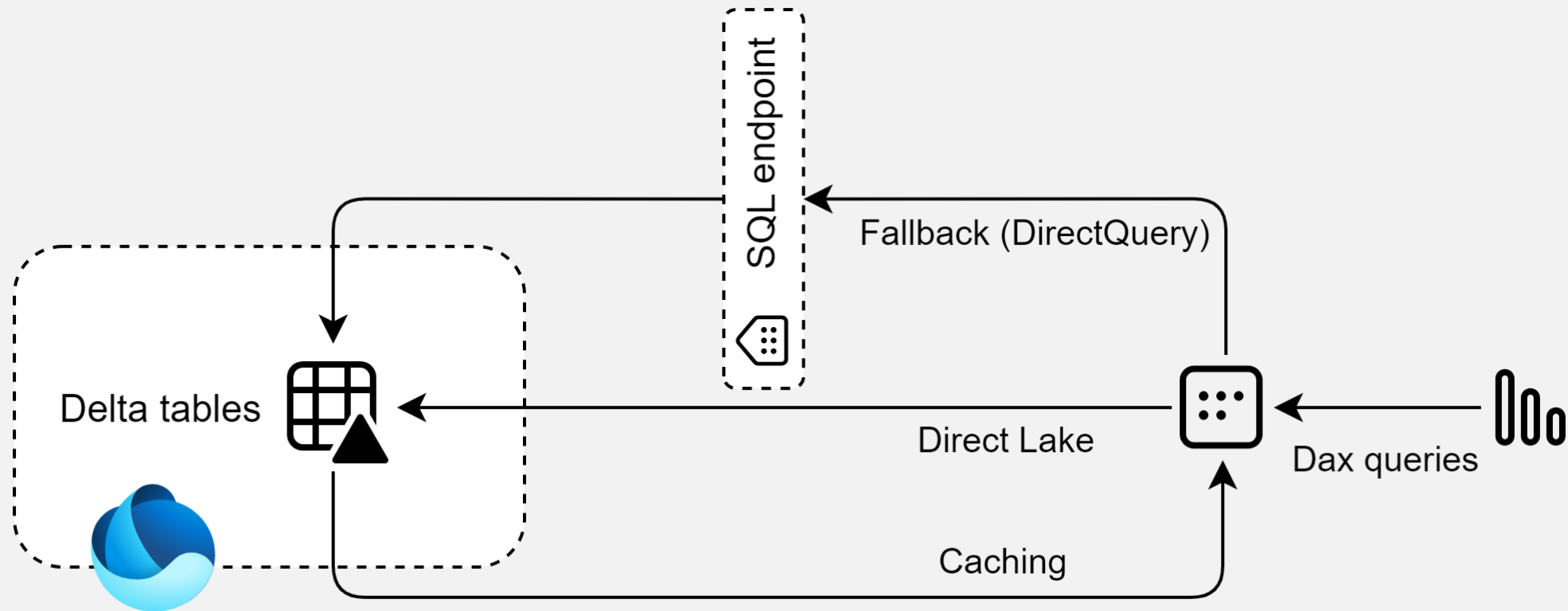
- ACR Adjustment Type
- Adjustment Type
- Budget
- Business
- Calendar
- Consumed Revenue
- Forecast
- Forecast Type
- Future Flag

Revenue Table Details:

Name	Size	Type
_delta_log	2 items	Folder
part-00000-87858576-90b7-4aff-8c9e-69dcc52dbt	8.4 GB	PARQUET
part-00001-631fb085-0591-46b8-a0b5-0fec8f2255	8.4 GB	PARQUET
part-00002-0469bb29-daa4-4ecd-a3ee-bb90331a6	8.4 GB	PARQUET
part-00003-27e6062b-4d55-4469-b285-7cb2a2f32	8.4 GB	PARQUET
part-00004-b12eea8e-f255-41fa-a943-a78def57ce	8.4 GB	PARQUET

A large yellow arrow points from the file explorer view to the table structure view, indicating the relationship between the physical files and the logical table structure.

Fallback & Caching



Fallback

When could fallback to DirectQuery happen?

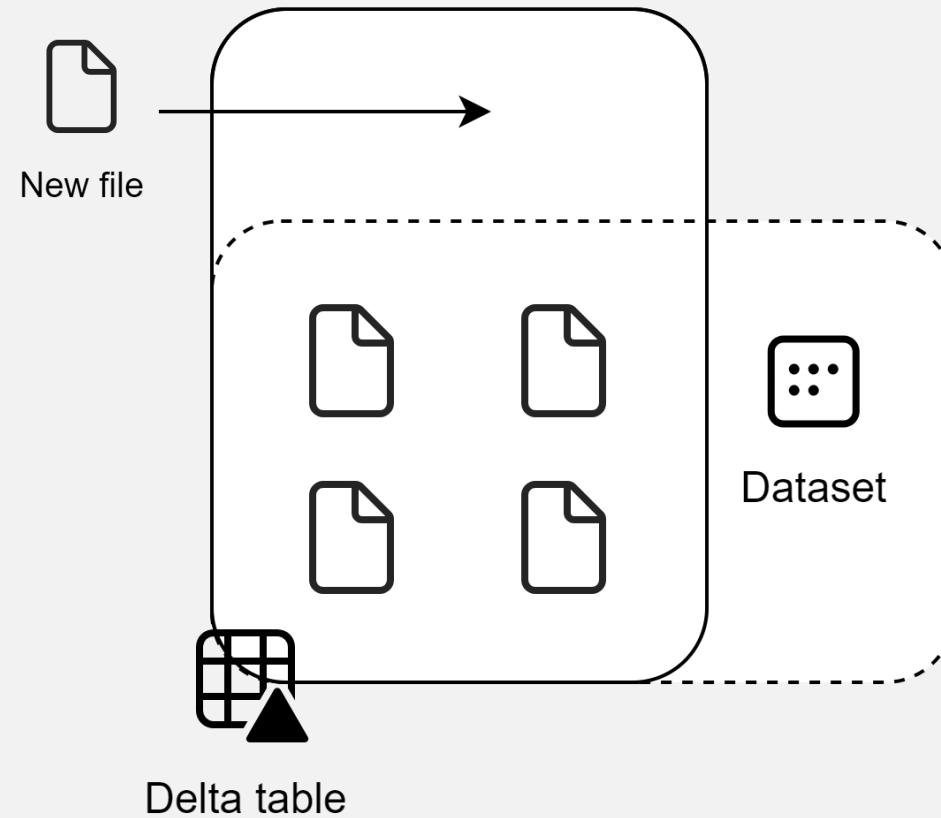
- Special data types
- Large data volumes that does not fit the capacity size
- Composite models
- When you manually configure security
Item level on lakehouse

Introducing Framing

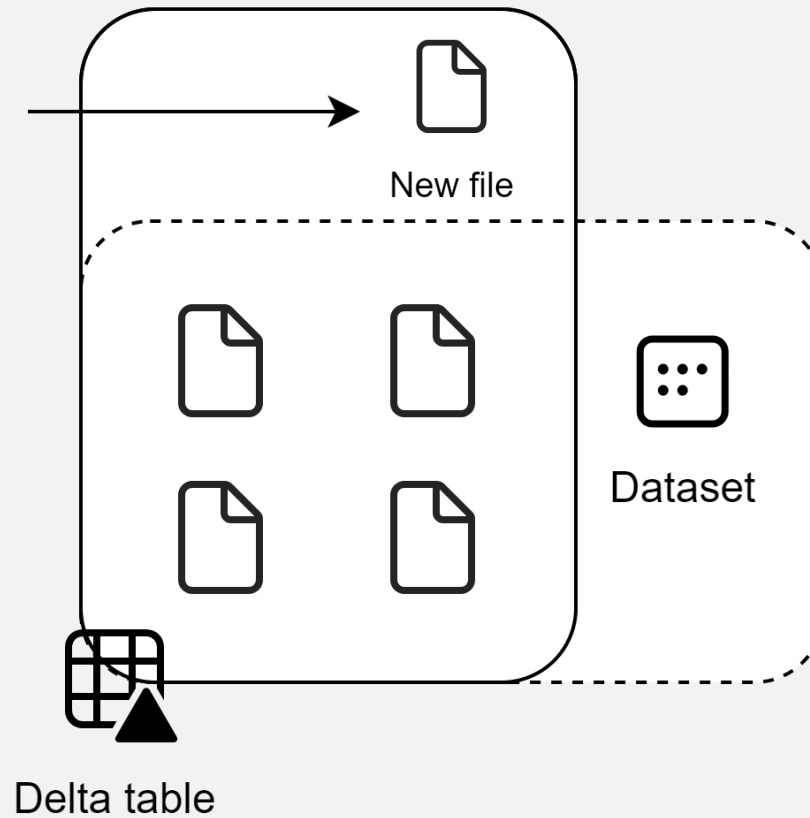
Metadata **refresh** which does not actually load the data, but only the delta table definitions.



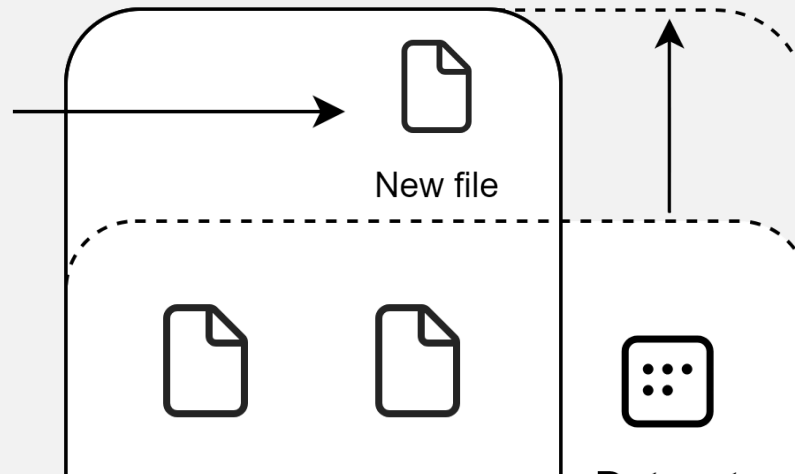
Framing



Framing



Framing



Refresh

Keep your Direct Lake data up to date

Configure Power BI to detect changes to the data in OneLake and automatically update the Direct Lake tables that are included in this dataset. [Learn more](#)

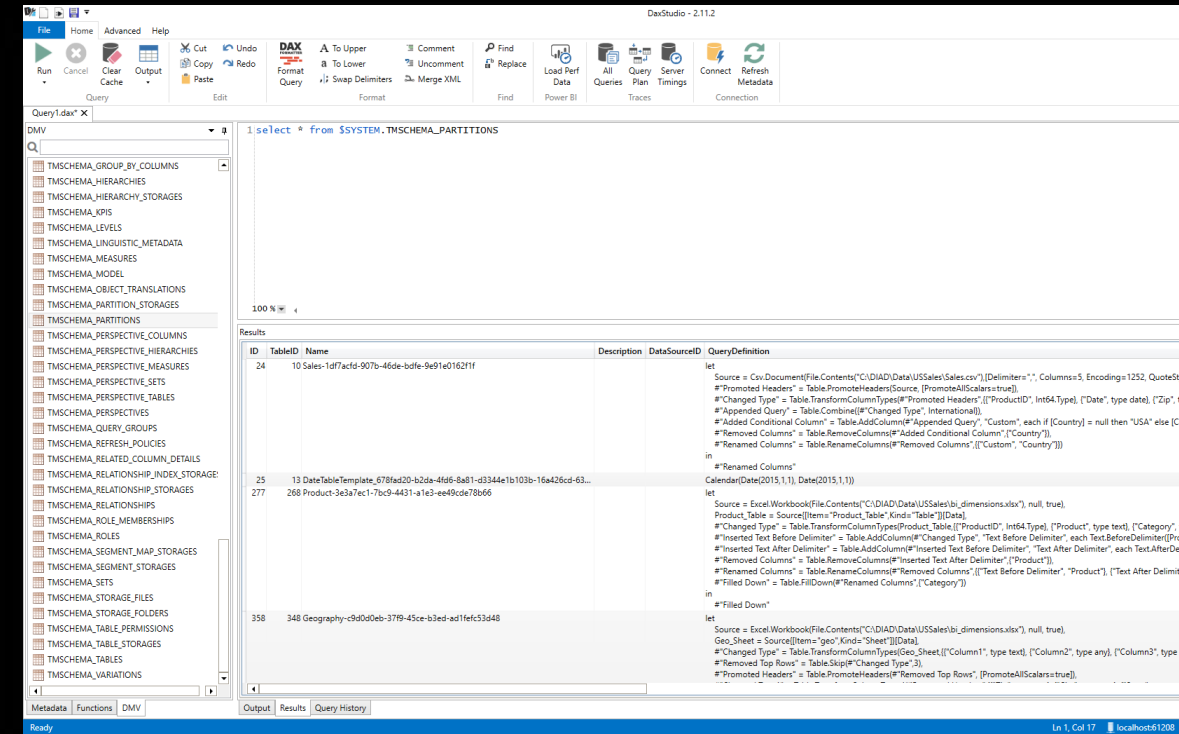


Off

Dynamic Management Views

Analysis Services Dynamic Management Views (DMVs) are queries that return information about model objects, server operations, and server health.

- DB Schema = Database model
- DISCOVER = Operations & Sessions
- TM Schema = Tabular = Power BI / AAS
- MD Schema = MDX = Multidimensional



Demo





Advanced patterns

XMLA

Talk with the back-end server, just like an Analysis Services server (read / write)

⏏ Server settings

Connection string

Data Source=powerbi://api.powerbi.com/v1.0/myorg/Dive%20into%20

Copy

XMLA Endpoint

Read Write ▾

Apply Discard

⊗ Caution

At this time, a write operation on a dataset authored in Power BI Desktop will prevent it from being downloaded back as a PBIX file. Be sure to retain your original PBIX file.

Introducing calculation groups

Benefits

- Reduce the number of redundant measures and grouping common measure expressions as calculation items
- Avoids duplicating logic in different measures
- Typical use cases are
 - Time-intelligence calculations (YTD / QTD / MTD / ...)
 - Format string change, like currency conversions

Limitations

- Can only be created from external tools in Power BI (Any tool using the XMLA endpoint such as Tabular Editor) – **but stay tuned....**
- Object level security on Calculation group items is not supported
- Smart narrative visuals in Power BI are not supported with Calculation Groups



Introducing calculation groups

Specific DAX expressions for Calculation Groups

- SELECTEDMEASURE()
- SELECTEDMEASURENAME()
- ISSELECTEDMEASURE()
- SELECTEDMEASUREFORMATSTRING()

Classic measure:

```
MTD =  
    CALCULATE (  
        SUM ( Sales[SalesAmount] ),  
        DATESMTD ( DimDate[Date] )  
    )
```

Dynamic measure context MTD with Calculation Group:

```
MTD =  
    CALCULATE (  
        SELECTEDMEASURE (),  
        DATESMTD ( DimDate[Date] )  
    )
```

Creating calculation groups over XMLA

Dive into DirectLake.XMLA Edited dataset - Tabular Editor 3.10.1 - Enterprise Edition

File Edit View Expression Calculation Item Model Tools Window Help

Expression Editor

Expression on Calculation Item "Calculate over shipdate" within 'Orders sold vs Orders shipped'

```
1 CALCULATE(  
2     SELECTEDMEASURE( ),  
3     USERELATIONSHIP(  
4 )
```

TOM Explorer

Enter text to search...

- 'Date'[DateKey] ∞-1 'InternetSales'[ShipDateKey]
- 'InternetSales'[ProductKey] ∞-1 'Product'[ProductKey]


Error Loading Data Model

Operation not supported for model with: ModifiedByXmlaEndpoint

Close

VertiPaq Analyzer

Summary Tables Columns Relationships

 **XMLA Edited**

Total Size 10,05 KB Last Refreshed 21-9-2023

Compatibility Mode	Level	Tables	Columns
PowerBI	1604	6	70

Description	
Name	Calculate over shipdate
Ordinal	-1
Metadata	
Error Message	
Object Type	Calculation Item
State	Ready
Options	

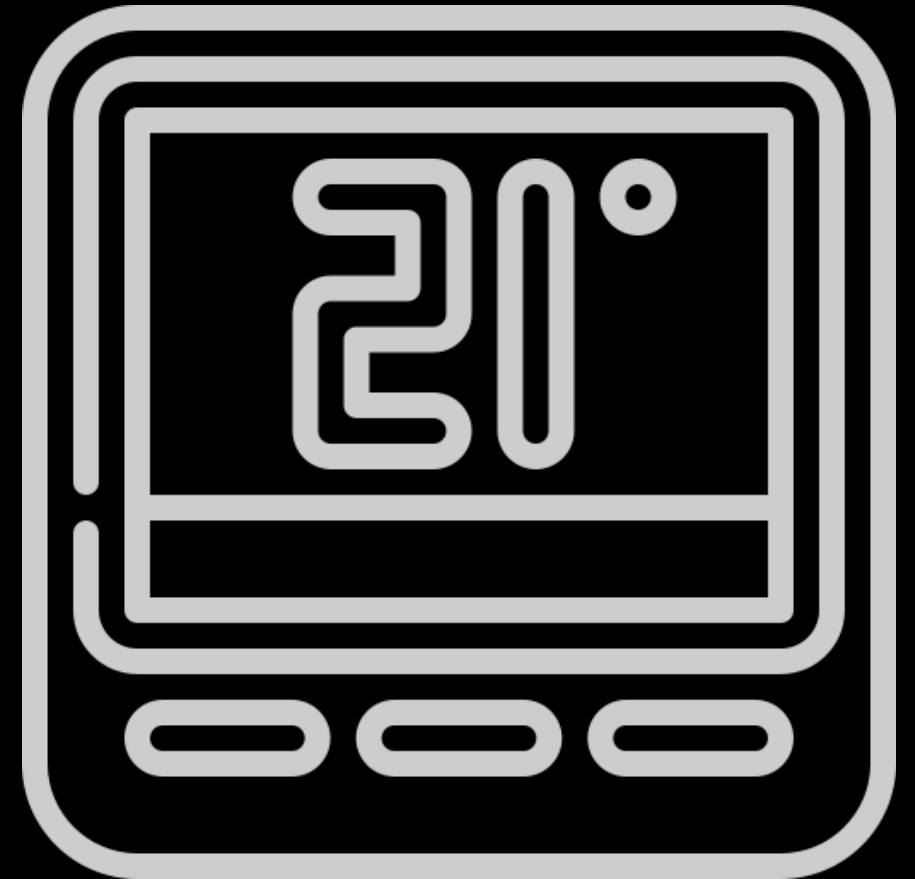
Model saved.

No issues AB Ln 4 Col 2 INS Calculation Item "Calculate over Dive into DirectLake

Temperature management

Keep it WARM!

Make sure your users are served optimally and avoid the capacity memory to be flushed.



Eviction

Capacity: F64 / P1

Memory: 25 GB

Capacity utilization:

Cool

Active memory



Storage



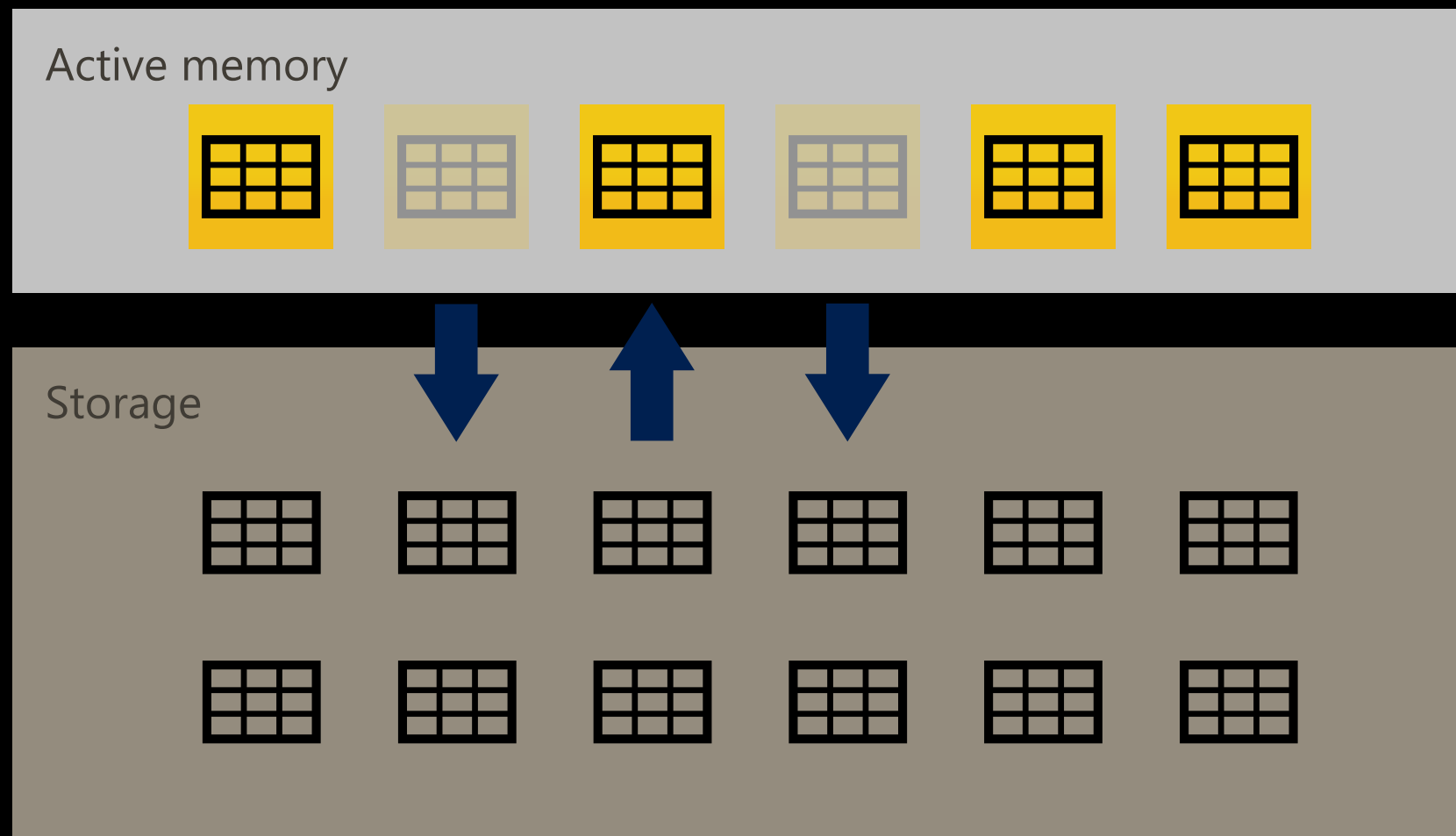
Eviction

Capacity: F64 / P1

Memory: 25 GB

Capacity utilization:

Warm



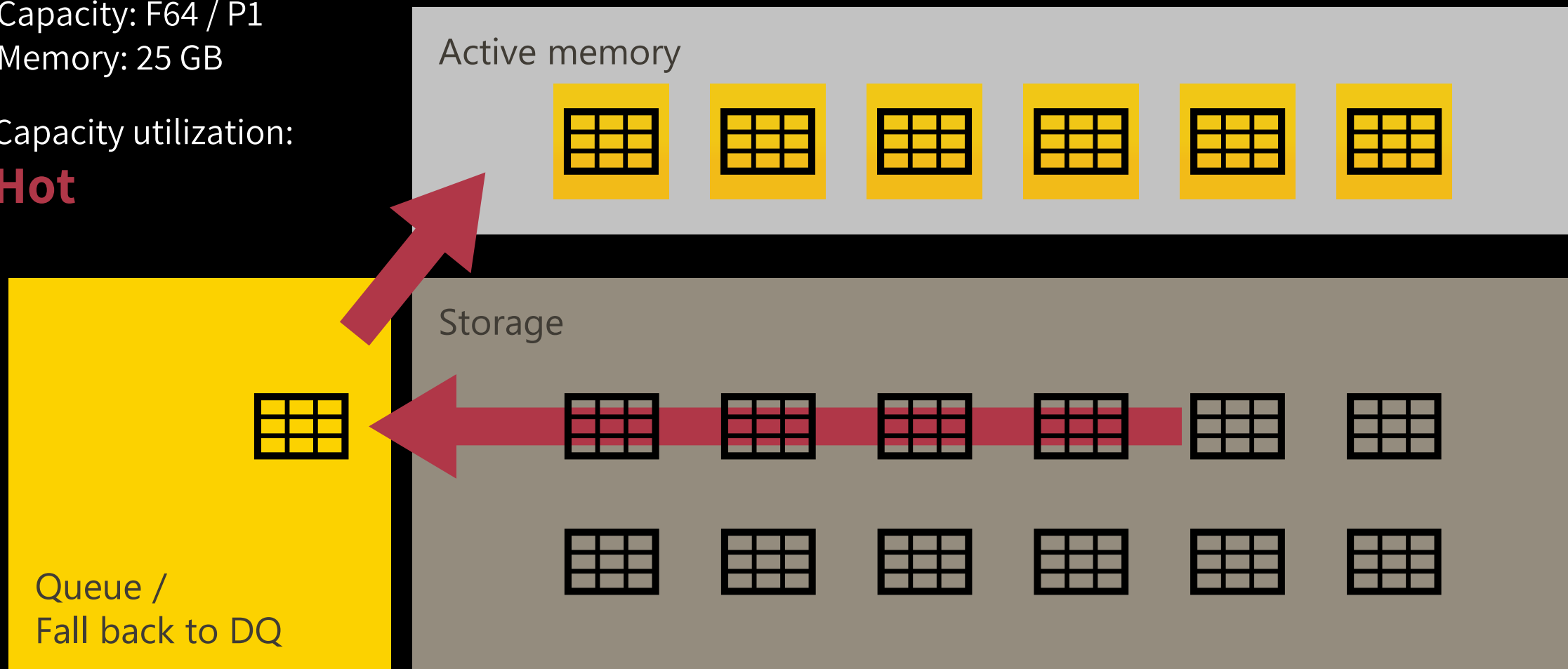
Eviction – queue / fall-back

Capacity: F64 / P1

Memory: 25 GB

Capacity utilization:

Hot



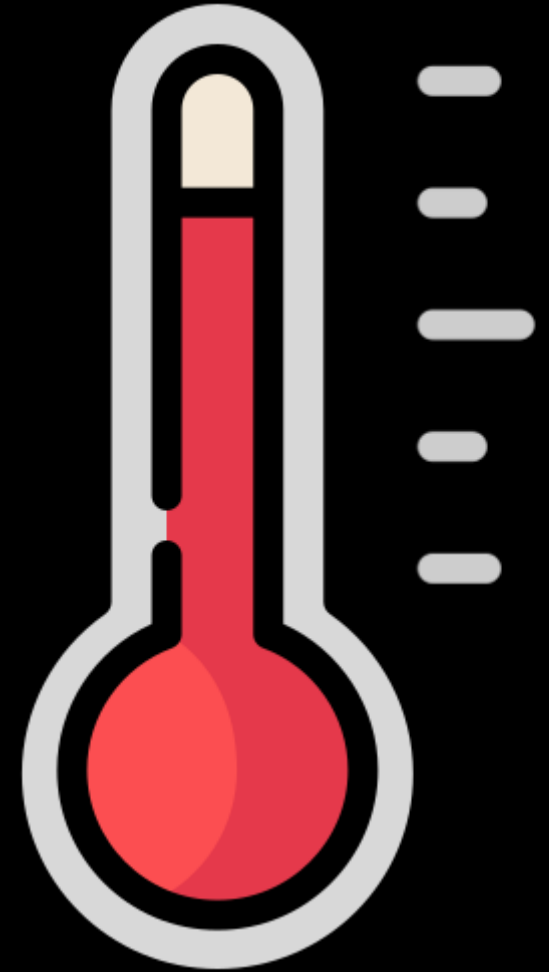
Temperature management

What will be evicted?

Basically, your data will be evicted from active memory, that you want to always have available!

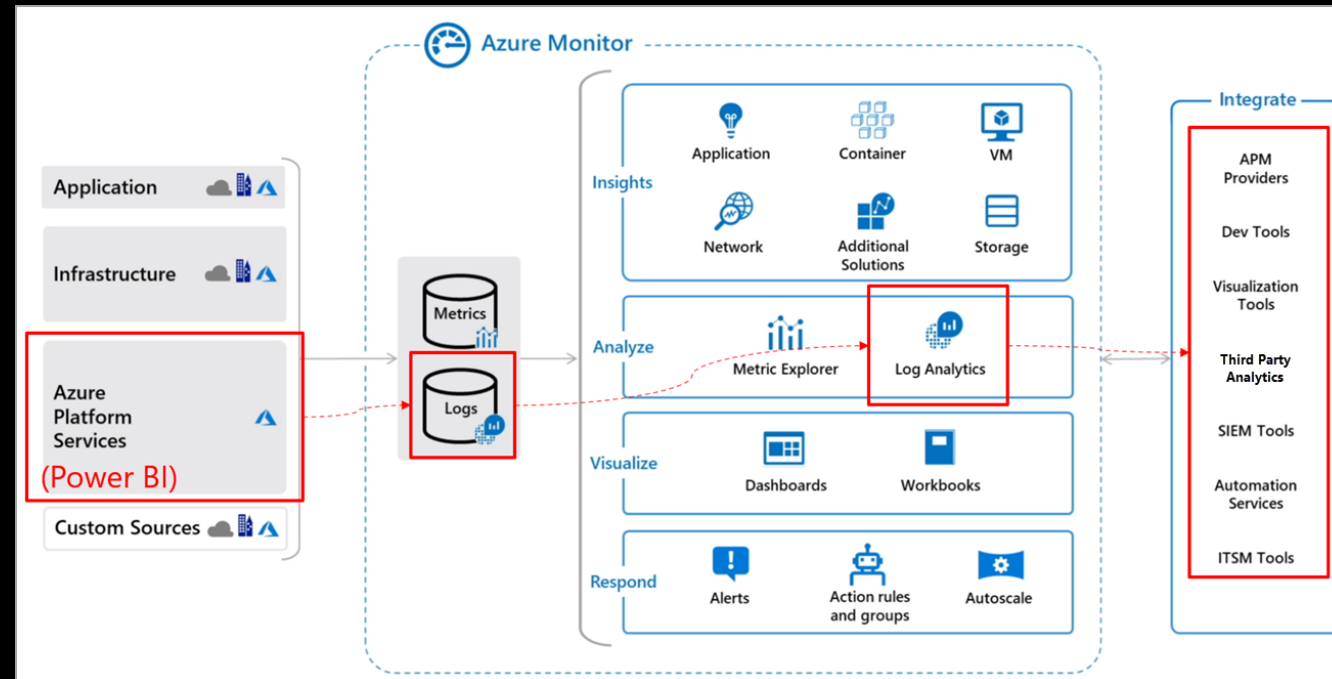
How can you influence that?

Consider setting up a process (notebook, other automated setup) to pro-actively execute queries to keep certain data **WARM**!



What should stay in memory?

Azure Monitor delivers a comprehensive solution for collecting, analyzing, and acting on telemetry from your cloud and on-premises environments. It helps you understand how your applications are performing and proactively identifies issues affecting them and the resources they depend on.



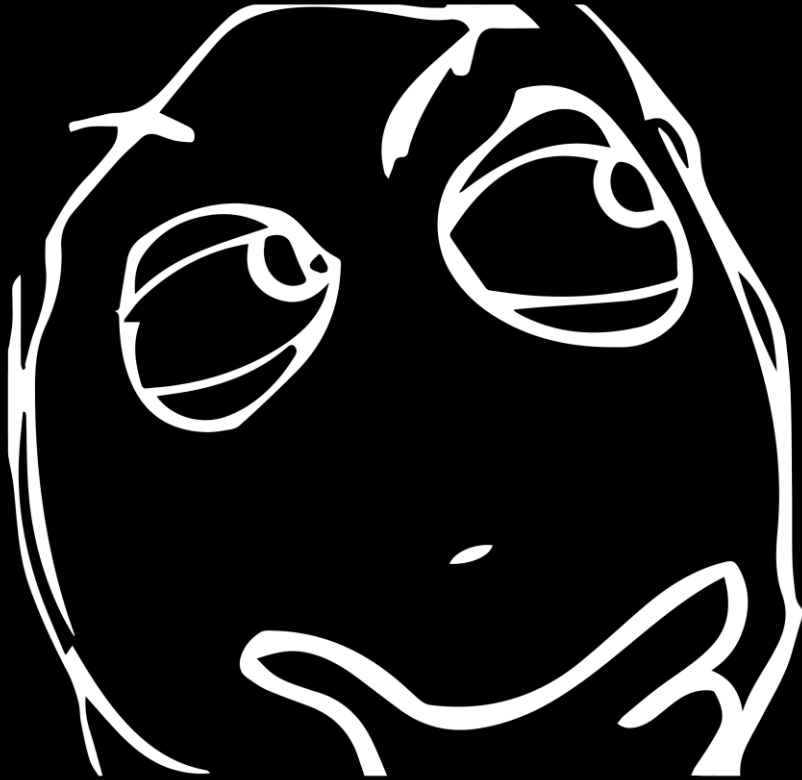
Wrap up

LET'S
RECAP...

Direct Lake...

- Only applicable when using MS Fabric
- No data is imported / copied
- On-demand loading
- Reads data from Lake / Parquet format – Delta is a must
- Performance dependent on capacity size/utilization
- Falls back to DirectQuery when limitations are hit!
- Consider implement advanced patterns for specific use cases

Considerations



IT DEPENDS

Should I change all my solutions to start using Direct Lake?

- There is no OneSecurity yet
RLS / OLS on dataset level is possible
- Consider impact on capacities when falling back to DQ
- Performance is better than DQ
- It is in public preview

Resources

Direct Lake generic documentation

[https://learn.microsoft.com/en-us/power-bi/enterprise/Direct Lake-overview](https://learn.microsoft.com/en-us/power-bi/enterprise/Direct-Lake-overview)

Calculation groups for Direct Lake datasets

<https://powerbi.microsoft.com/en-us/blog/announcing-calculation-groups-for-direct-lake-datasets/>

Analyze performance for Direct Lake

<https://learn.microsoft.com/en-us/power-bi/enterprise/directlake-analyze-qp>

On-demand loading of Direct Lake Power BI datasets in Fabric

<https://blog.crossjoin.co.uk/2023/07/02/on-demand-loading-of-direct-lake-power-bi-datasets-in-fabric/>

Direct Lake Frequently Asked Questions

<https://fabric.guru/power-bi-direct-lake-mode-frequently-asked-questions/>