data lab

Top 5 reasons for Power BI folks to try Fabric

A big thank you to our partners

PBIG
POWER BI GEBRUIKERSGROEP

Who am I

Ásgeir Gunnarsson

Data Platform MVP

Partner, Data Platform Architect @ data lab

20+ years in BI
Co-organizer of Power BI Next Step, Data Platform Next Step and Power BI Deep Dive in Denmark
Co-author of Pro Microsoft Power BI Administration

E-mail: asgeir@thesmalldatalab.com
Bluesky: @bidgeir
LinkedIn: linkedin.com/in/asgeirgun

data lab

We're The Small Data Lab – a modern Danish data consultancy helping companies access and create real value from data across the entire organisation.

We work with businesses that want more than just dashboards. Our focus is on delivering insight, clear governance, and data platforms that evolve with both people and technology – on site or remotely.

THE TOP 5 REASONS

- Dataflow gen2
- Copy Job for copy only
- Snapshots/incremental/delta load
- Direct Lake mode of sematic models
- Activator for alerts

- Bonus
  - Notebooks for APIs

ALL FEATURES MENTIONED HERE REQUIRE A FABRIC CAPACITY

CAN BE OF ANY SIZE

WHO ARE YOU

- Power BI creator
- Know how to use Power BI Desktop to import and clean data
- Know what Power BI Dataflows are
- Comfortable around data
- Feel some of the limitation of Power Query and Power BI Dataflows

DATAFLOW GEN2

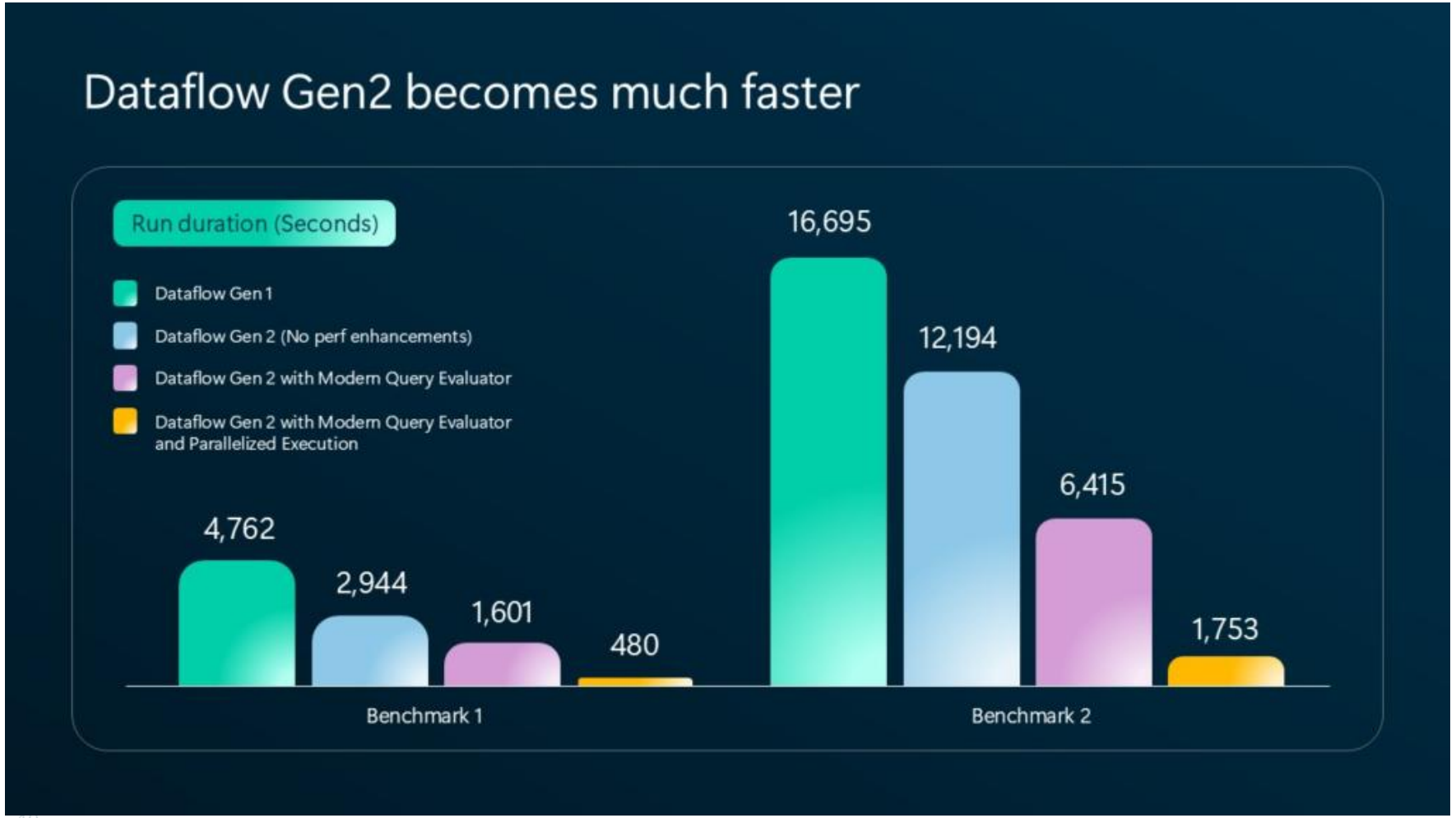
DATAFLOW GEN2

- Evolution of Dataflow gen1
- Most noticeable difference is the chose of destination
- Possible to parameterize table/file names and connections
- Other hidden features
  - Staging
  - Fast copy
  - Power Query template import

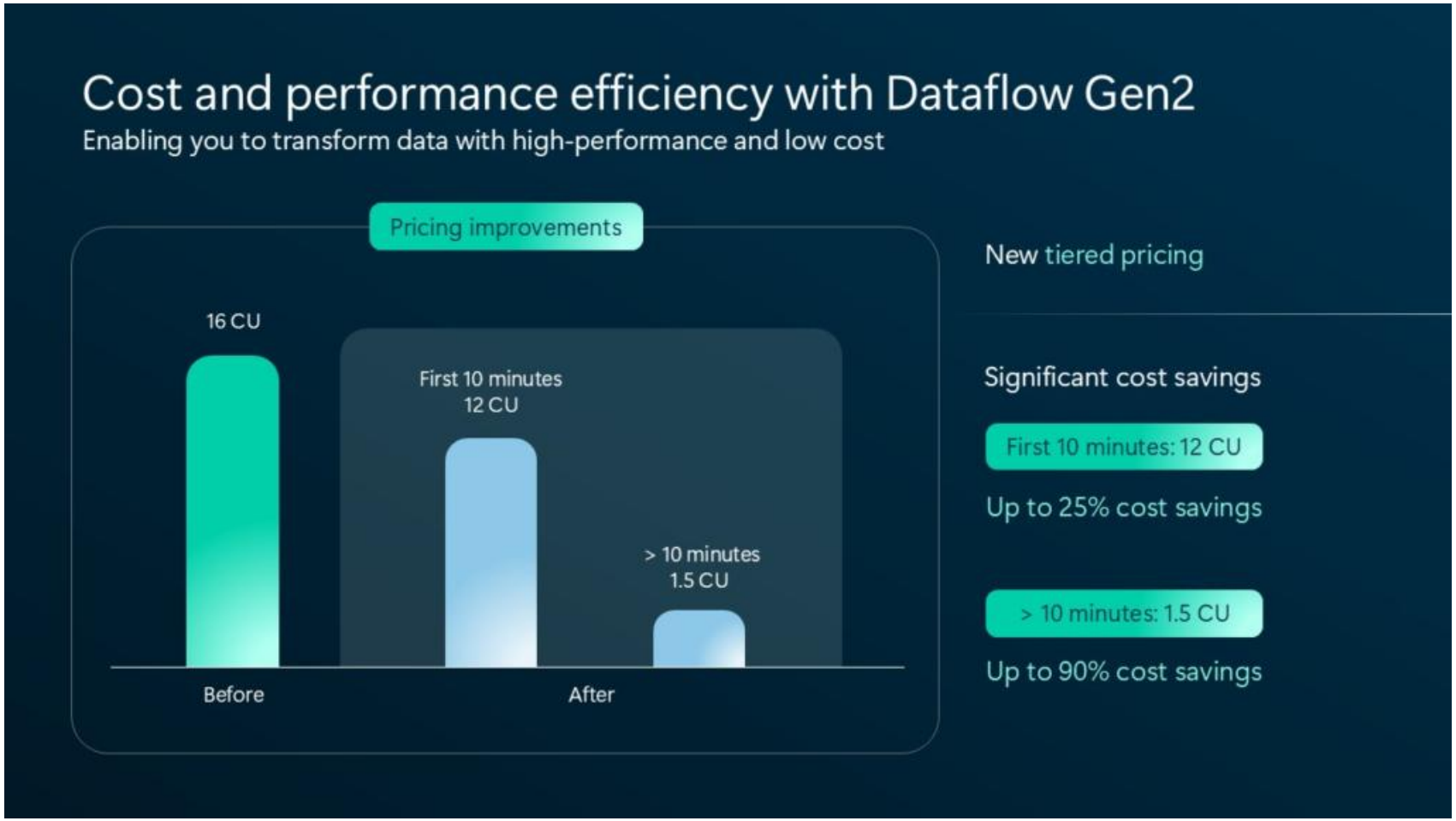
DATAFLOW GEN1 VS GEN2

Feature	Dataflow Gen2	Dataflow Gen1
Create dataflows with Power Query	✓	✓
Simpler creation process	✓	
AutoSave and background publishing	✓	
Multiple output destinations	✓	
Better monitoring and refresh tracking	✓	
Works with pipelines	✓	
High-performance computing	✓	
Connect via the dataflow connector	✓	✓
Direct Query via the dataflow connector		✓
Refresh only changed data	✓	✓
AI-powered insights	✓	✓
Recent data shortcuts to previously used sources	✓	

DATAFLOW GEN1 VS GEN2



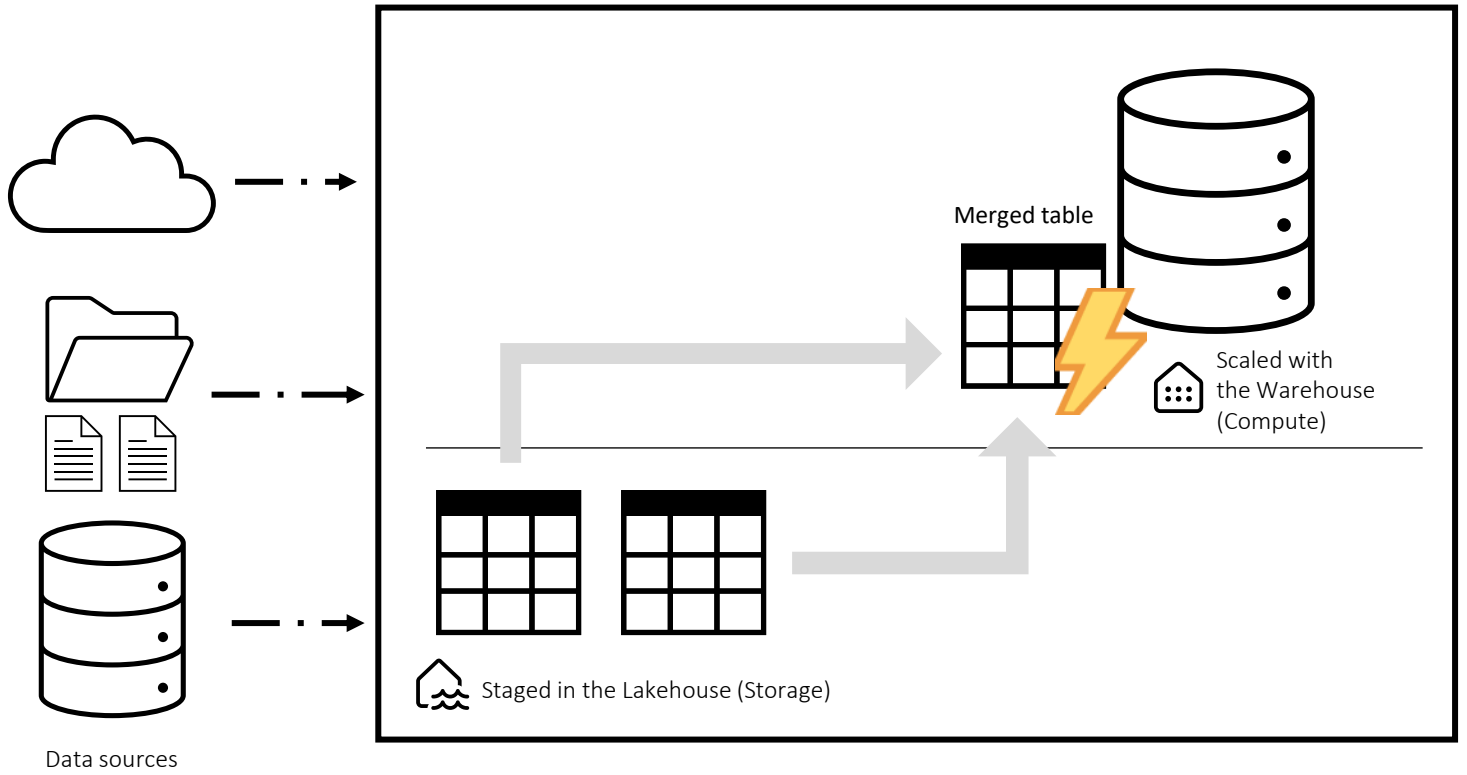
DATAFLOW GEN2



Computed Tables

Power Query Online enables Premium subscribers to use their capacity to optimize the use of dataflows with Fabric compute.

1. Connect to your data and copy it into the Lakehouse using *"Enable staging" (**on by default!**)
2. Create a reference query in a new query.
3. Apply transformation steps to the computed table for complex ETL operations such as join, distinct, filter and group by – leveraging the Warehouse for compute



*Previously titled "Enable load"

DEMO DATAFLOW GEN2

COPY JOB

COPY JOB

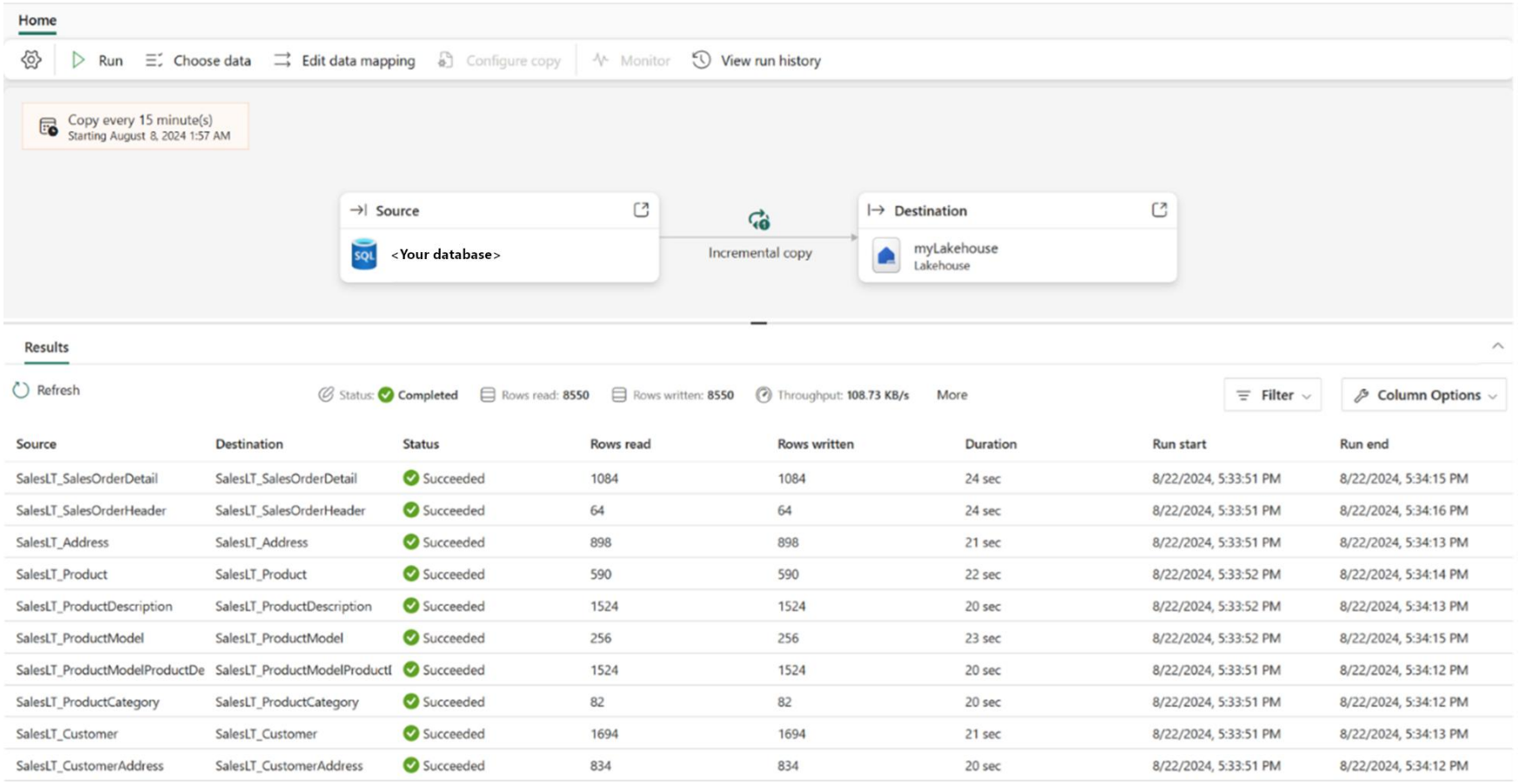
- Tool for data copy
- Simple to use
- Ideal for moving large amount of data from multiple tables from the same data source to the same destination
- Does not transform data
- Handles incremental refresh automatically

Connectors for Copy Job - Microsoft Fabric | Microsoft Learn

COPY JOB - MONITORING

- Easy to monitor
- Good information provided
  - **Status**: The current state or final outcome of the job run (for example, succeeded, failed, or in progress).
  - **Rows read**: The total number of rows read from the source during the job run.
  - **Rows written**: The total number of rows written to the destination during the job run.
  - **Throughput**: The average data movement rate achieved during the job run.

COPY JOB



Source	Destination	Status	Rows read	Rows written	Duration	Run start	Run end
SalesLT_SalesOrderDetail	SalesLT_SalesOrderDetail	Succeeded	1084	1084	24 sec	8/22/2024, 5:33:51 PM	8/22/2024, 5:34:15 PM
SalesLT_SalesOrderHeader	SalesLT_SalesOrderHeader	Succeeded	64	64	24 sec	8/22/2024, 5:33:51 PM	8/22/2024, 5:34:16 PM
SalesLT_Address	SalesLT_Address	Succeeded	898	898	21 sec	8/22/2024, 5:33:51 PM	8/22/2024, 5:34:13 PM
SalesLT_Product	SalesLT_Product	Succeeded	590	590	22 sec	8/22/2024, 5:33:52 PM	8/22/2024, 5:34:14 PM
SalesLT_ProductDescription	SalesLT_ProductDescription	Succeeded	1524	1524	20 sec	8/22/2024, 5:33:52 PM	8/22/2024, 5:34:13 PM
SalesLT_ProductModel	SalesLT_ProductModel	Succeeded	256	256	23 sec	8/22/2024, 5:33:52 PM	8/22/2024, 5:34:15 PM
SalesLT_ProductModelProductDe	SalesLT_ProductModelProductI	Succeeded	1524	1524	20 sec	8/22/2024, 5:33:51 PM	8/22/2024, 5:34:12 PM
SalesLT_ProductCategory	SalesLT_ProductCategory	Succeeded	82	82	20 sec	8/22/2024, 5:33:51 PM	8/22/2024, 5:34:12 PM
SalesLT_Customer	SalesLT_Customer	Succeeded	1694	1694	21 sec	8/22/2024, 5:33:51 PM	8/22/2024, 5:34:13 PM
SalesLT_CustomerAddress	SalesLT_CustomerAddress	Succeeded	834	834	20 sec	8/22/2024, 5:33:51 PM	8/22/2024, 5:34:12 PM

DEMO COPY JOB

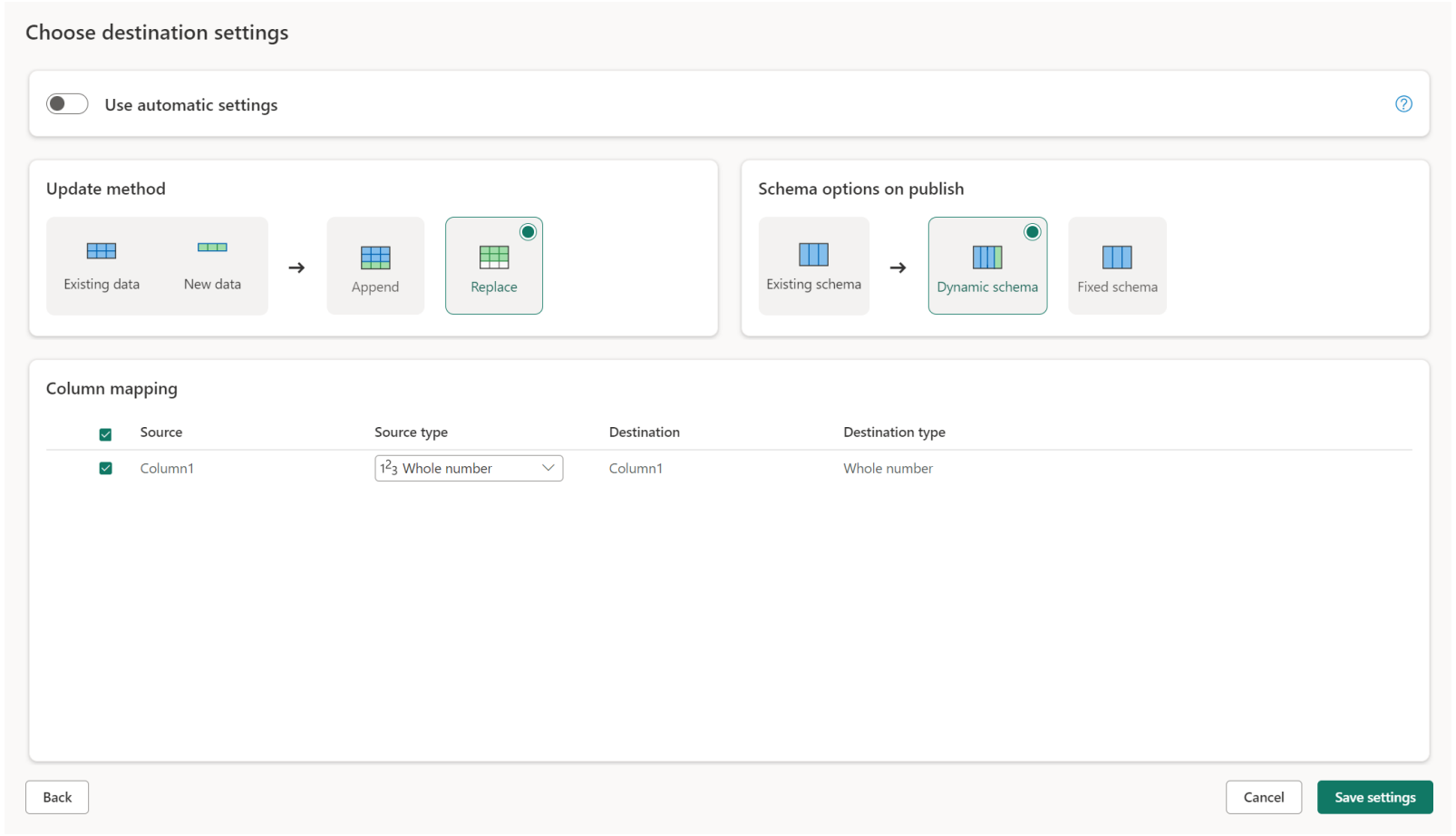
NOT FULL REFRESH

SNAPSHOTS/INCREMENTAL/DELTA LOAD

- Power BI and Dataflows gen1 always fetch the whole set of data each time
  - Exception is incremental refresh which only works on some data sources
- With destination in Dataflow gen2 and Copy Jobs you can do snapshot or incremental/delta refresh with all data sources
  - Snapshots fetch a full set of data each time and store it next to the full set of data from other fetches – useful to understand how things looked at a particular point in time but essentially duplicating data each time
  - Incremental refresh or delta load only fetch data which is new since last load and store it next to the existing data – useful when older data don't change (what to do if older data changes? = Merge), and you want to prevent duplicating data
- Dataflows can do simple append or partitions
- Copy job can do advanced merge and watermark control

DATAFLOW GEN2

- Different update methods
  - Replace
  - Append
- Schema
  - Flexible
  - Fixed



COPY JOB

- **Easy to use**: Set up and monitor data copying with a simple, guided experience — no technical expertise needed.
- **Efficient**: Copy only new or changed data from the last run to save time and resources, with minimal manual steps.
- **Flexible**: Choose which data to move, map columns, set how data is written, and schedule jobs to run once or regularly.
- **High performance**: Move large amounts of data quickly and reliably, thanks to a serverless, scalable system.

- Update methods
  - Append
  - Overwrite
  - Merge

DEMO
SNAPSHOTS/INCREMENTAL/DELTA LOAD

DIRECT LAKE

DIRECT LAKE MODE

- Requires data to be in Fabric in delta format
- Can bridge the gap between direct query and import mode
- Ideal when the amount of data is larger than the available memory
- Is not always the best option
- Comes in two flavors
  - Direct Lake on OneLake
  - Direct Lake on SQL (Analytics Endpoint)

Fabric SKU	Parquet files per table	Row groups per table	Rows per table (millions)	Max model size on disk/OneLake (GB)	Max memory (GB) [1]
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 MODE

- Not all features are supported
  - ~~Calculated columns~~
  - ~~Calculated tables~~
  - Complex data types in Delta files
  - Semantic model and data source need to be in the same region

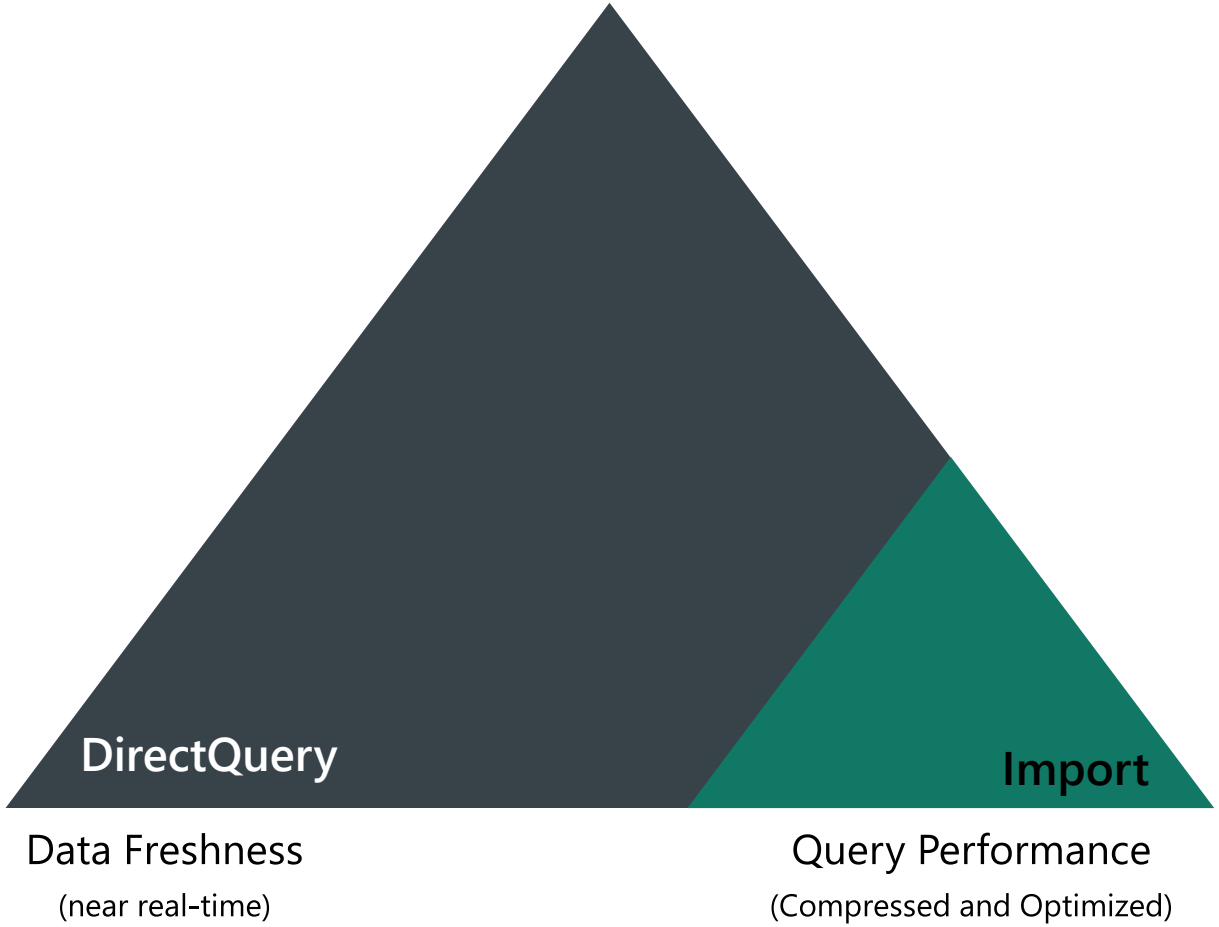
DIRECT LAKE MODE

Fabric SKU	Parquet files per table	Row groups per table	Rows per table (millions)	Max model size on disk/OneLake (GB)	Max memory (GB) [1]
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

Storage Modes

Import: Caches data into memory to deliver extremely fast performance using the **analysis services** database engine. The default mode when creating a new Power BI Desktop solution along with providing Data Modelers the most design flexibility.

DirectQuery: Does not import the data into memory, consists only of the metadata defining the structure. When the model is queried, native queries are used to retrieve data from the underlying data source.



Changing the **Storage mode** of a table to **Import** is an irreversible operation. Once set, this property can't later be changed using Power BI Desktop.

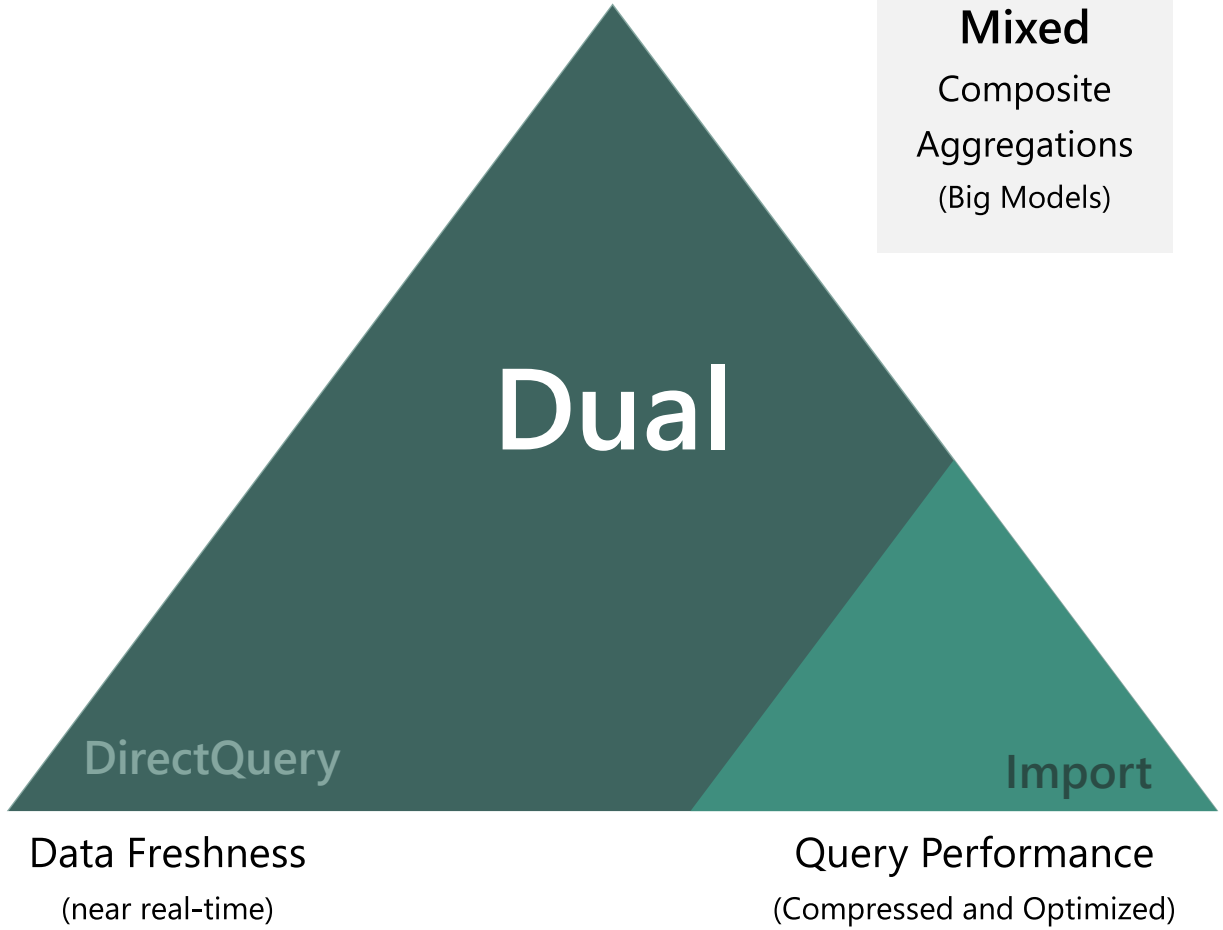
Storage Modes

Import: Caches data into memory to deliver extremely fast performance using the **analysis services** database engine. The default mode when creating a new Power BI Desktop solution along with providing Data Modelers the most design flexibility.

DirectQuery: Does not import the data into memory, consists only of the metadata defining the structure. When the model is queried, native queries are used to retrieve data from the underlying data source.

Dual: Can act as either cached or not cached, depending on the context of the query that's submitted to the Power BI dataset. In some cases, you fulfill queries from cached data. In other cases, you fulfill queries by executing an on-demand query to the underlying data source.

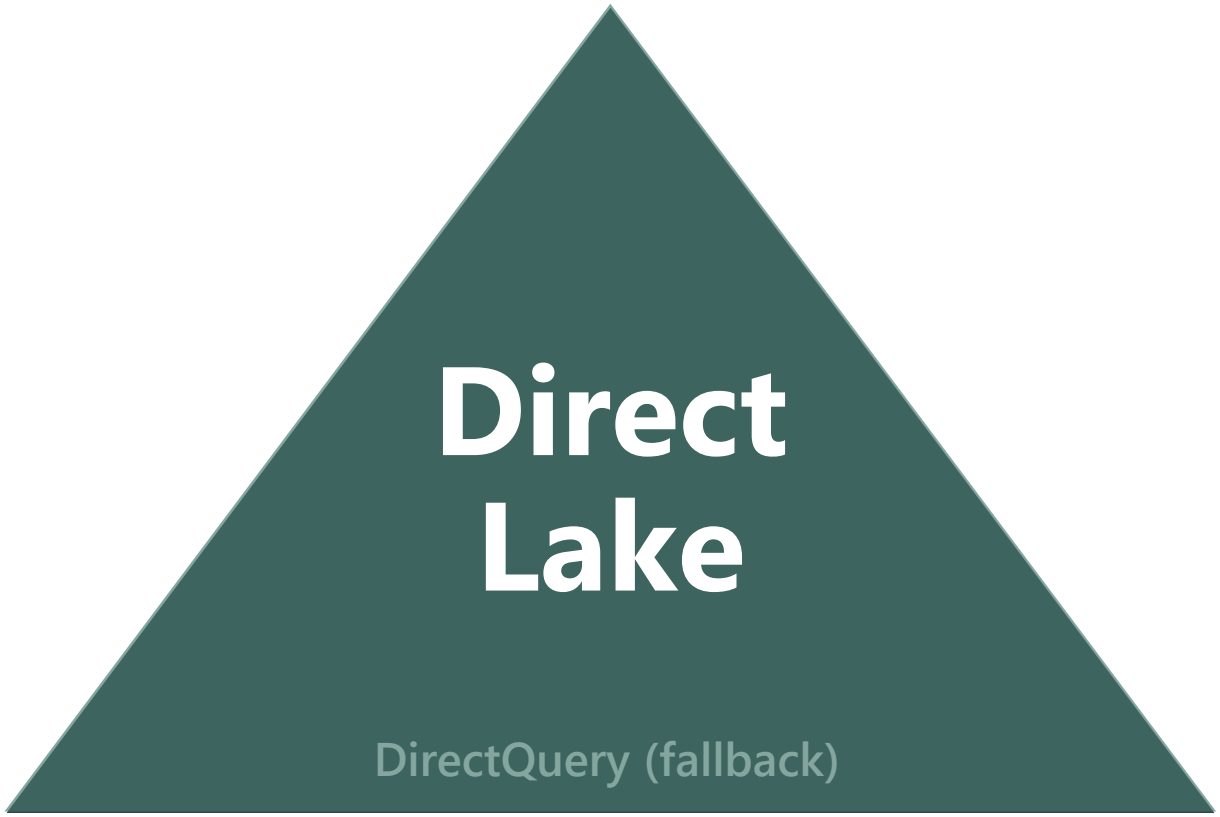
Changing the **Storage mode** of a table to **Import** is an irreversible operation. Once set, this property can't later be changed to either **DirectQuery** or **Dual** using Power BI Desktop.



Storage Modes

Direct Lake: A groundbreaking new dataset capability for analyzing very large data volumes. Based on loading parquet-formatted files directly from a data lake **without having to query a Lakehouse endpoint**, and **without having to import or duplicate data** into a Power BI dataset. Direct Lake is a fast-path to load the data from the lake straight into the Power BI engine, ready for analysis and yielding performance similar to import mode.

DirectQuery (fallback): Automatically switches modes- either due to current limitations or based on factors such as available memory in the capacity.



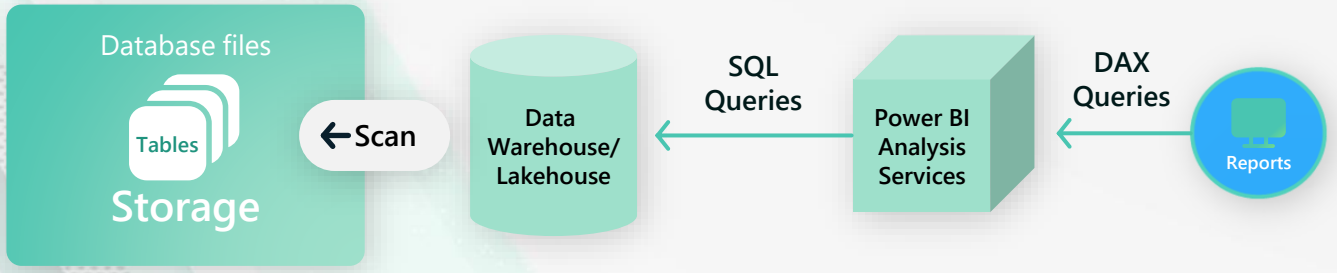
Data Freshness & Query Performance

POWER BI | DIRECT LAKE MODE

Direct Lake is a fast-path to load the data from the lake straight into the Power BI engine, ready for analysis

Direct Lake is based on loading parquet-formatted files directly from a data lake without having to query a Lakehouse endpoint, and without having to import or duplicate data into a Power BI dataset

DirectQuery Mode. Slow, but real time

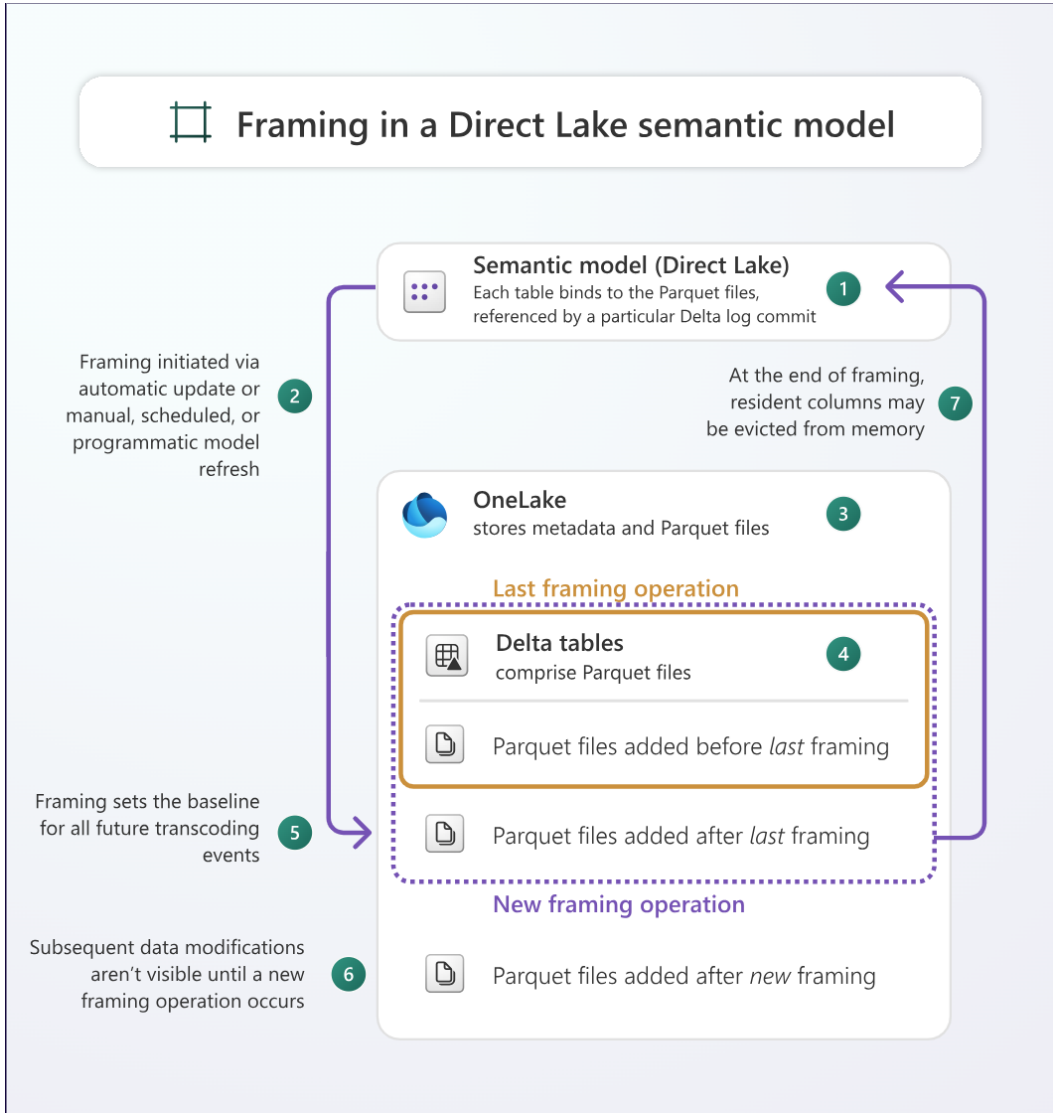


Import Mode. Fast, but latent and duplicative

Direct Lake Mode. Fast and "real time"

DIRECT LAKE

- ”Refreshing” (or framing) still required (it´s seconds)
  - Well not exactly required but most likely desirable
- Never going to be faster than import mode
- You can use DAX Studio to analyze which attributes are in memory and how much space it takes



DEMO DIRECT LAKE

ACTIVATOR

WHAT IS ACTIVATOR

- No-code/low code experience to take action on data patterns or conditions
- Can monitor Power BI, built-in events or Eventstreams items
- Can do simple actions or kick off a workflow

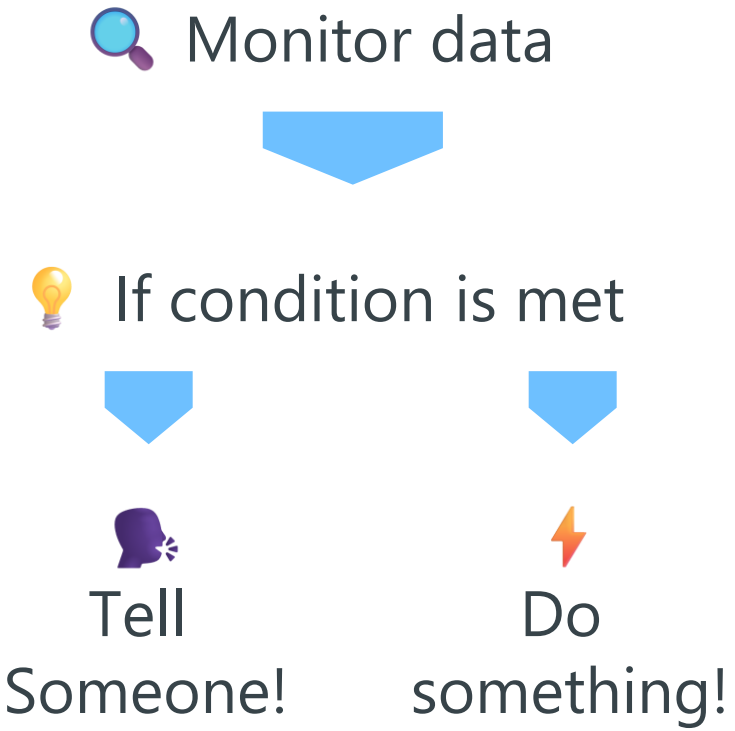
A BIT ABOUT ACTIVATOR

🔍 Monitor data

💡 If condition is met

⚡ Do something!

A BIT ABOUT ACTIVATOR



Tell Someone!

"When shelves are almost empty, ask someone to restock"

"If refrigeration unit is too hot, message someone to move perishables out"

Do something!

"If inventory is low, put in a new order"

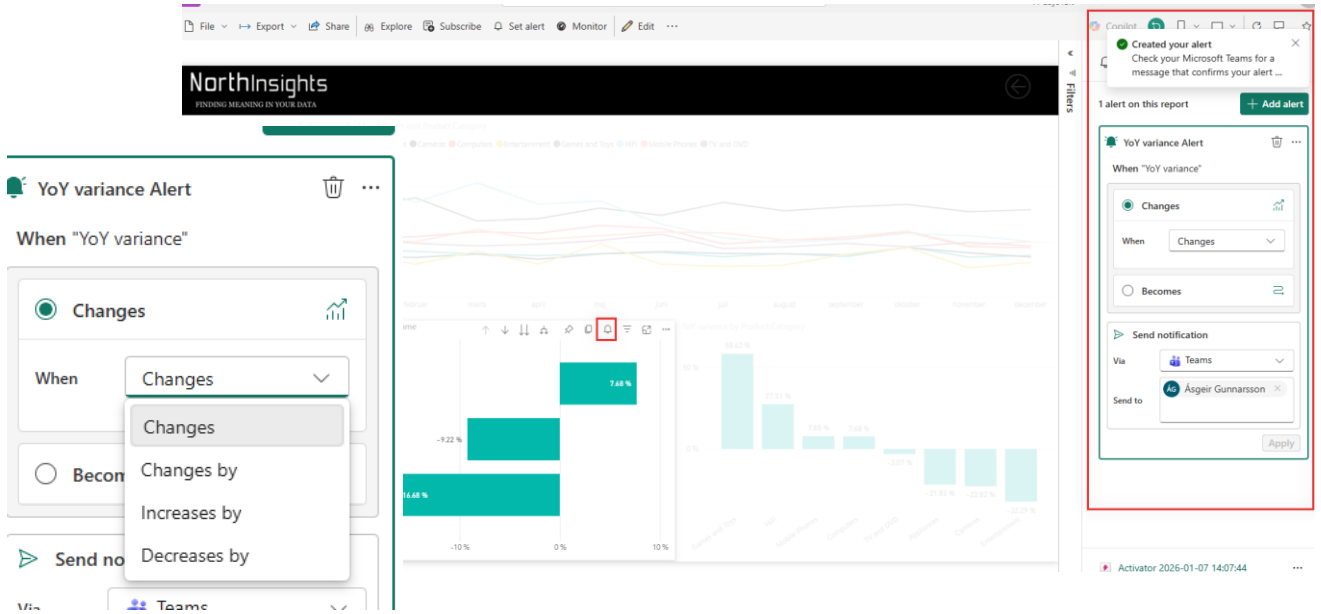
"If baggage carrousel is full, divert to another carrousel"

WHAT IS DATA ACTIVATOR

- The item is called Activator (formerly known as Reflex)
- Can listen to verity of events
  - Power BI semantic model changes (with user context)
  - Various built-in events
  - Anything you can add to eventstream item in Fabric
- Triggers on conditions being fulfilled
  - Conditions can be dataset fields
  - Conditions can be other than simple bigger than

DATA ACTIVATOR IN POWER BI

- In Power BI report you can add an alert
- You can decide on what triggers the alert
- It respects the context of the visual you use
- You can decide if the alert is an Email or a Teams message
- Self-service / citizen developer experience
  - You can get the full Activator experience afterwards by editing the Reflex item



DEMO ACTIVATOR

NOTEBOOKS

NOTEBOOKS

- Notebooks are code only
- Work with Fabric Lakehouses (think of lakehouse as a database)
- Multi-lingual
  - PySpark
  - SQL
  - R
  - Scala
- Very powerful
- Semantic Link is a good way to get started with notebooks
- Can be interesting for complicated datasources such as web services

microsoft/semantic-link-labs: Early access to new features for Microsoft Fabric's Semantic Link.

What is semantic link? - Microsoft Fabric | Microsoft Learn

Loved it? Learned something? Tell us!
Share your feedback in just 1 minute

Top 5 reasons for Power BI folks to use Fabric

PBIG
POWER BI GEBRUIKERSGROEP

QUESTIONS