Navigating the Data Landscape: A Journey through Incremental Refresh & Hybrid Tables in Power BI

Nalaka Wanniarachchi

Work

10+ years of Data Warehousing and Business intelligence solutions development and mentoring.
PowerBI everything and Oracle BI etc..

Background

CIMA(UK) Qualified Ex-Banker

(ACMA / CGMA),
16+ years of financial industry experience,
Father of two 👶
Employed @ WILEY

Nalaka Wanniarachchi

Tech Specialist
Lead Business Intelligence Engineer

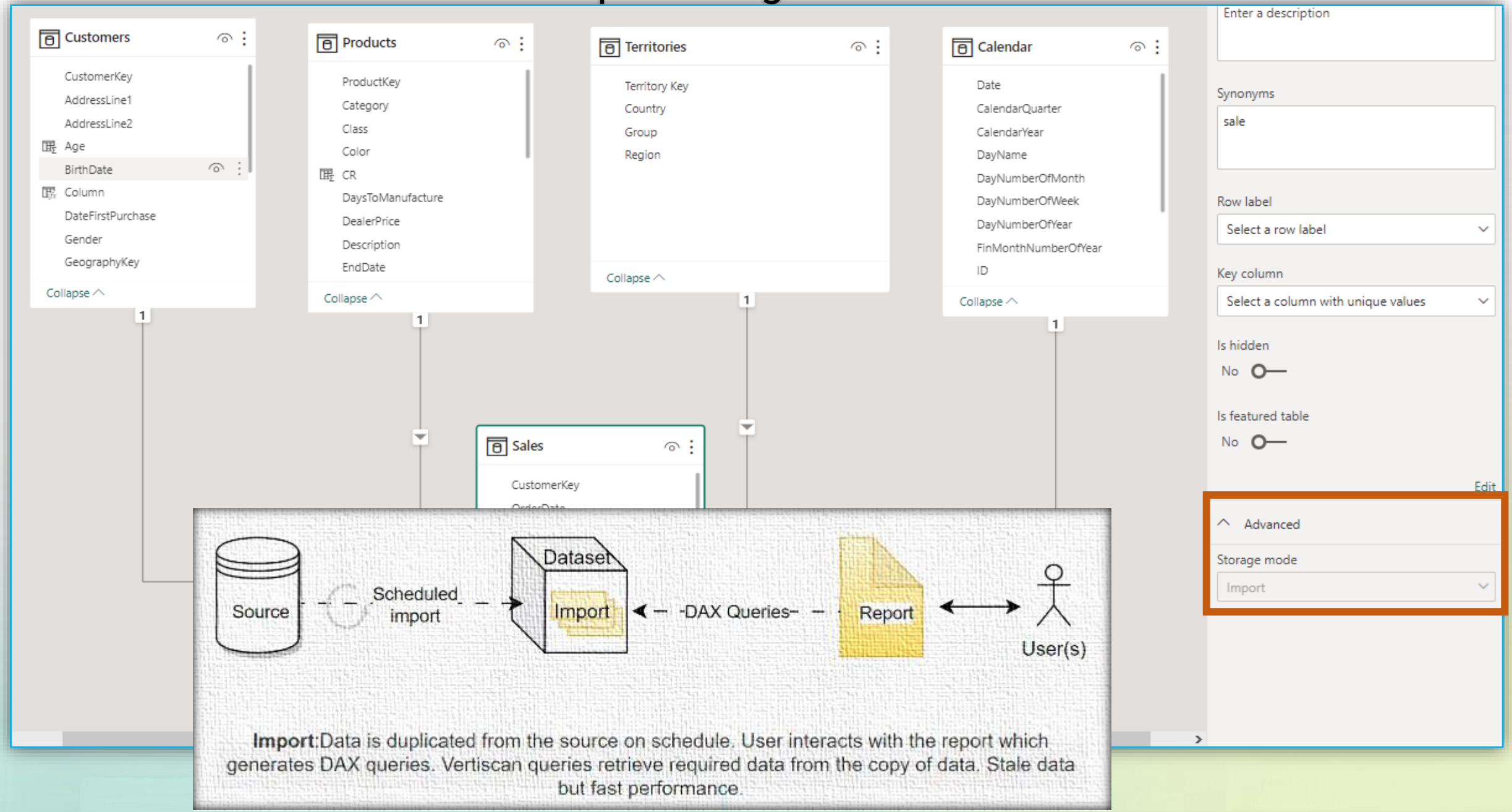
AGENDA

- Table Storage Modes / Connection Types
- Incremental Refresh
- Partition Considerations
- Hybrid Mode
- Optional Settings in Incremental Refresh
- Direct Lake Mode
- Demo
- Reverse Hybrid Workaround
- Custom Partitioning
- Third Party Tools / Advanced incremental refresh

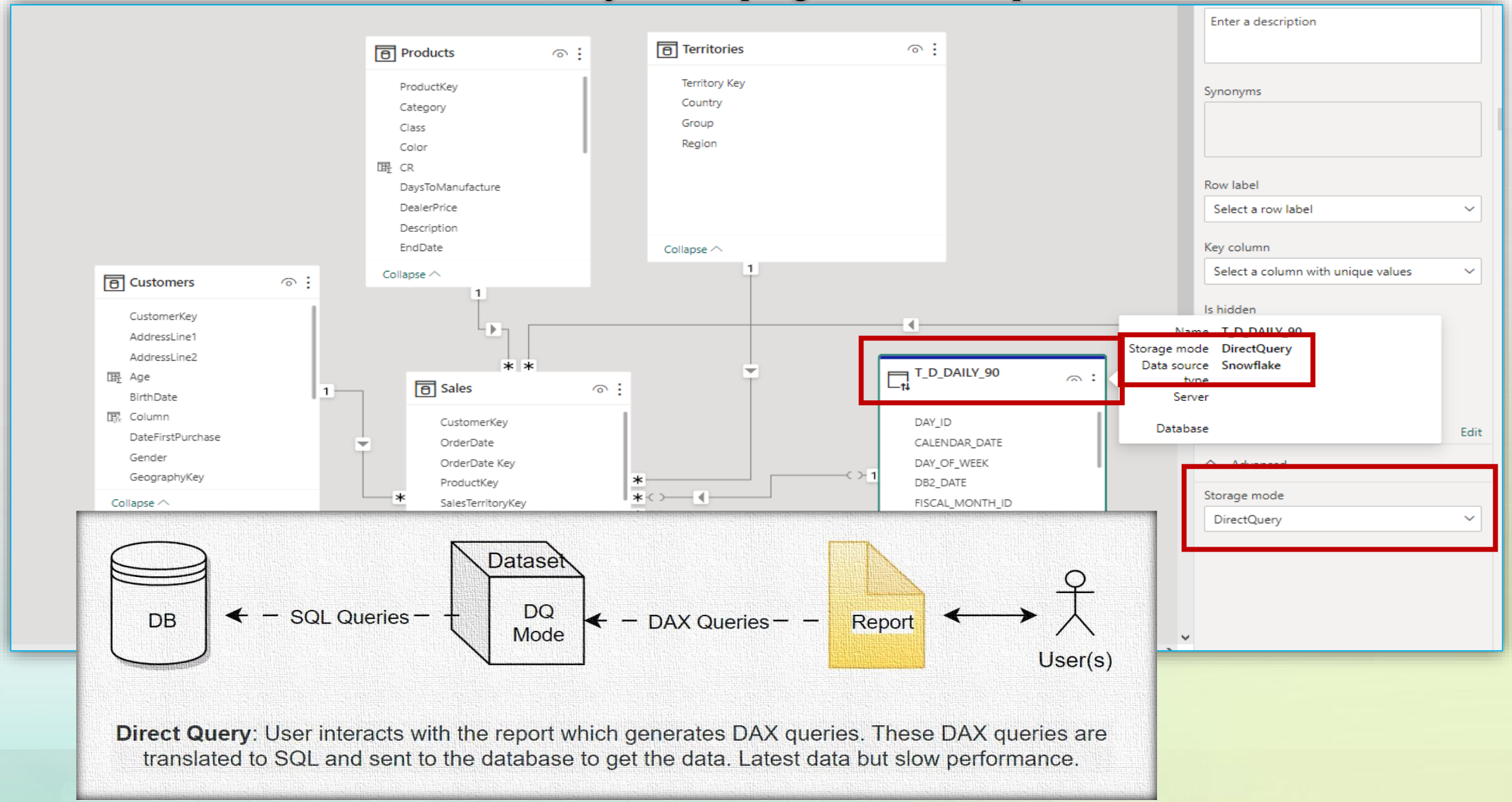
Table storage modes / Connection Types

- Import Mode
- Direct Query Mode
- Live connection Mode
- Composite Mode or Mixed Mode
- Dual Mode
- Hybrid Tables: Two Parts in same table
- Direct Lake (Fabric)

Import Storage Mode

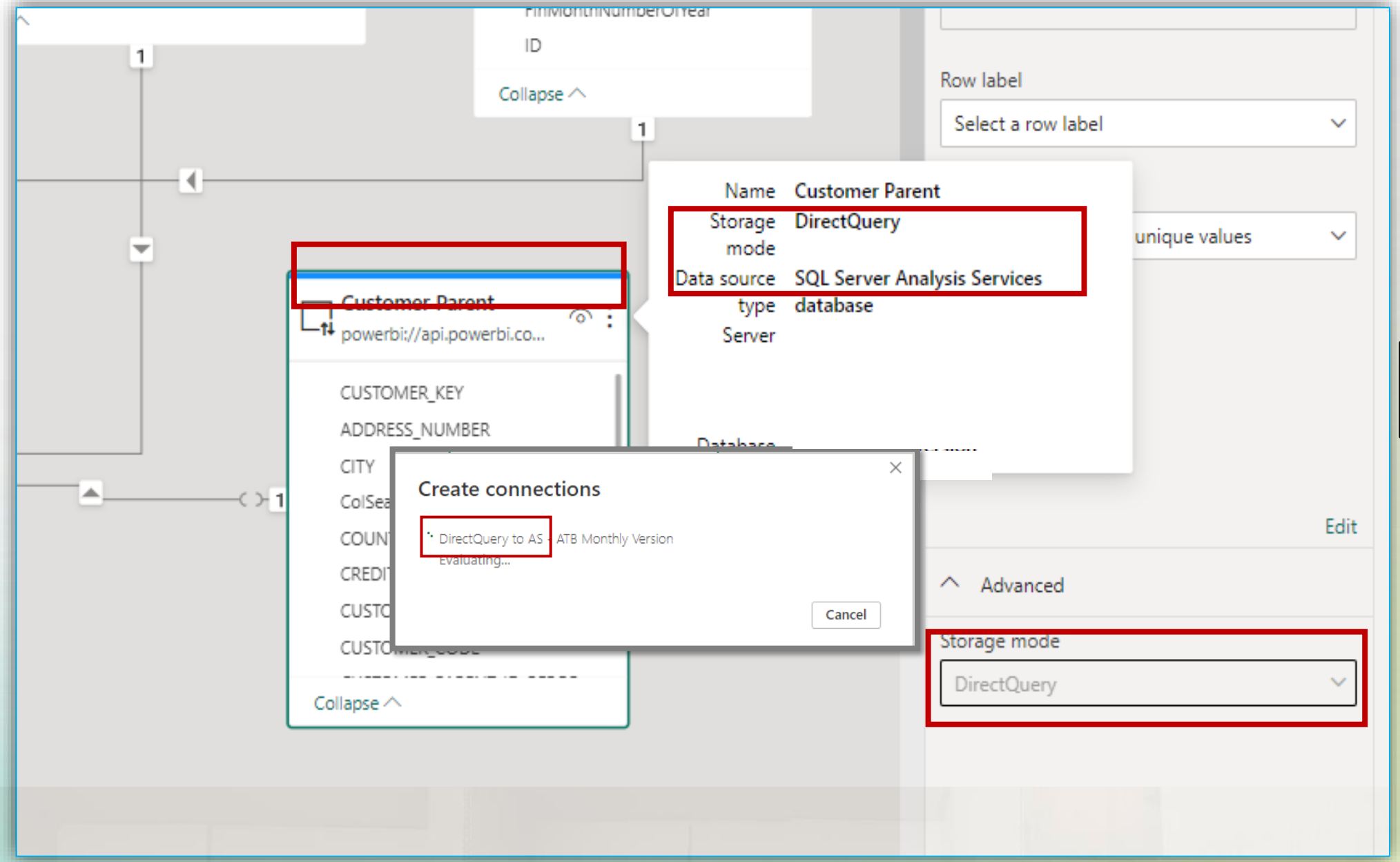


Direct Query Mode-[Regular Sources]



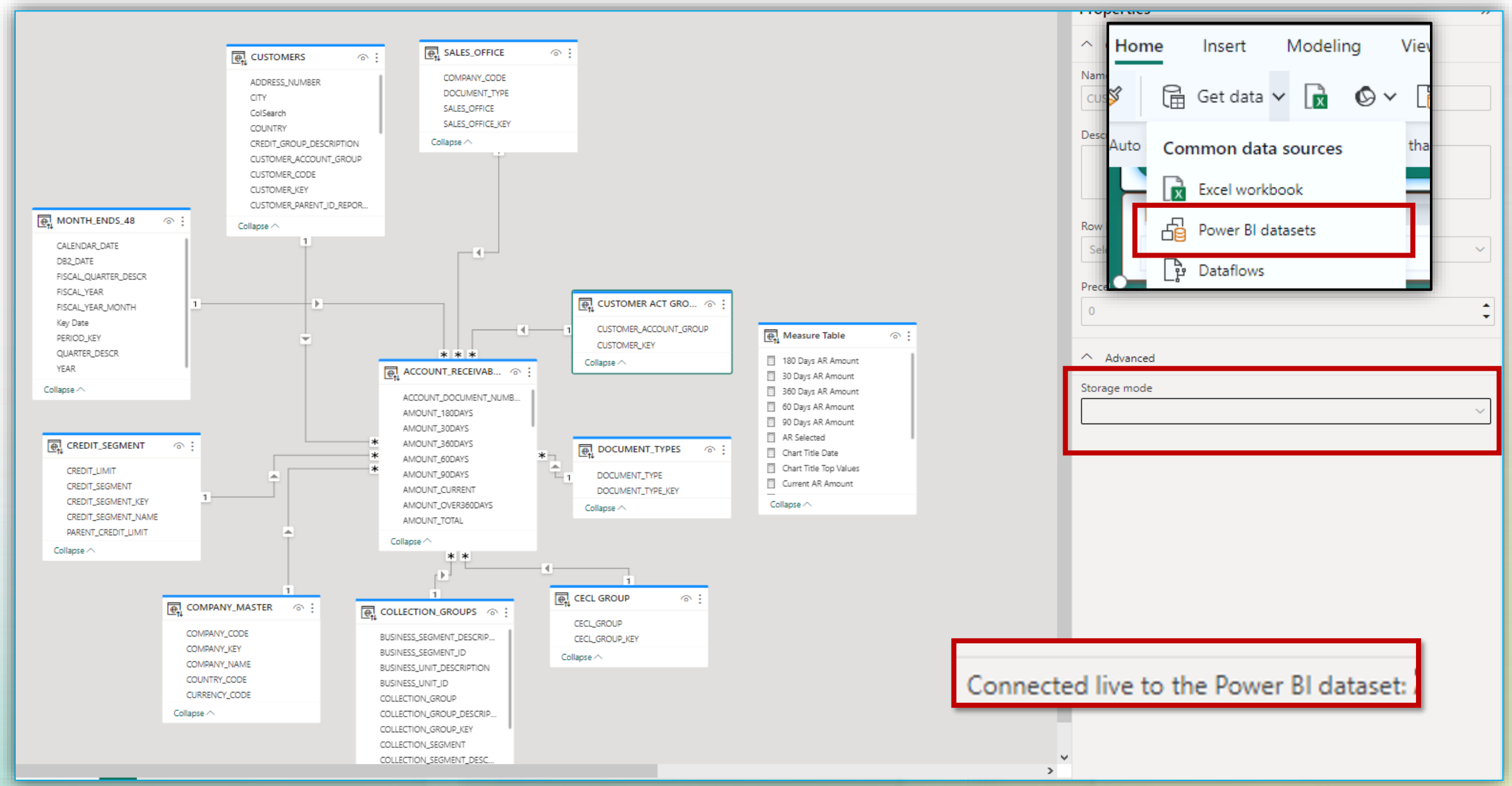
Direct Query: User interacts with the report which generates DAX queries. These DAX queries are translated to SQL and sent to the database to get the data. Latest data but slow performance.

Direct Query Mode-To [AS]

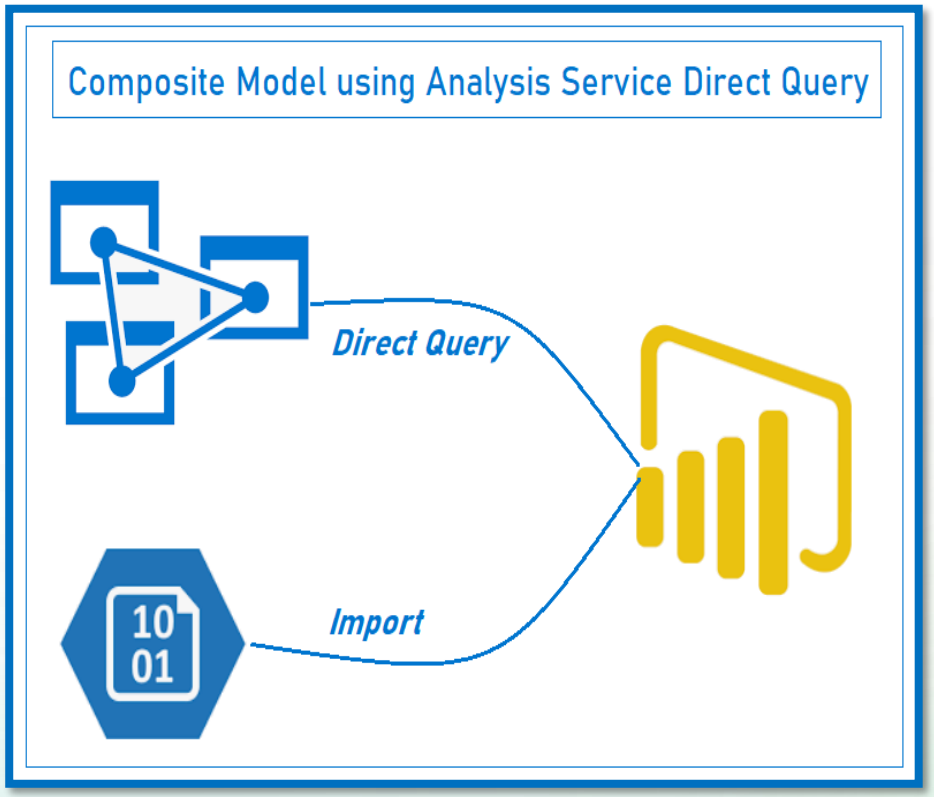


Using PowerBI Datasets

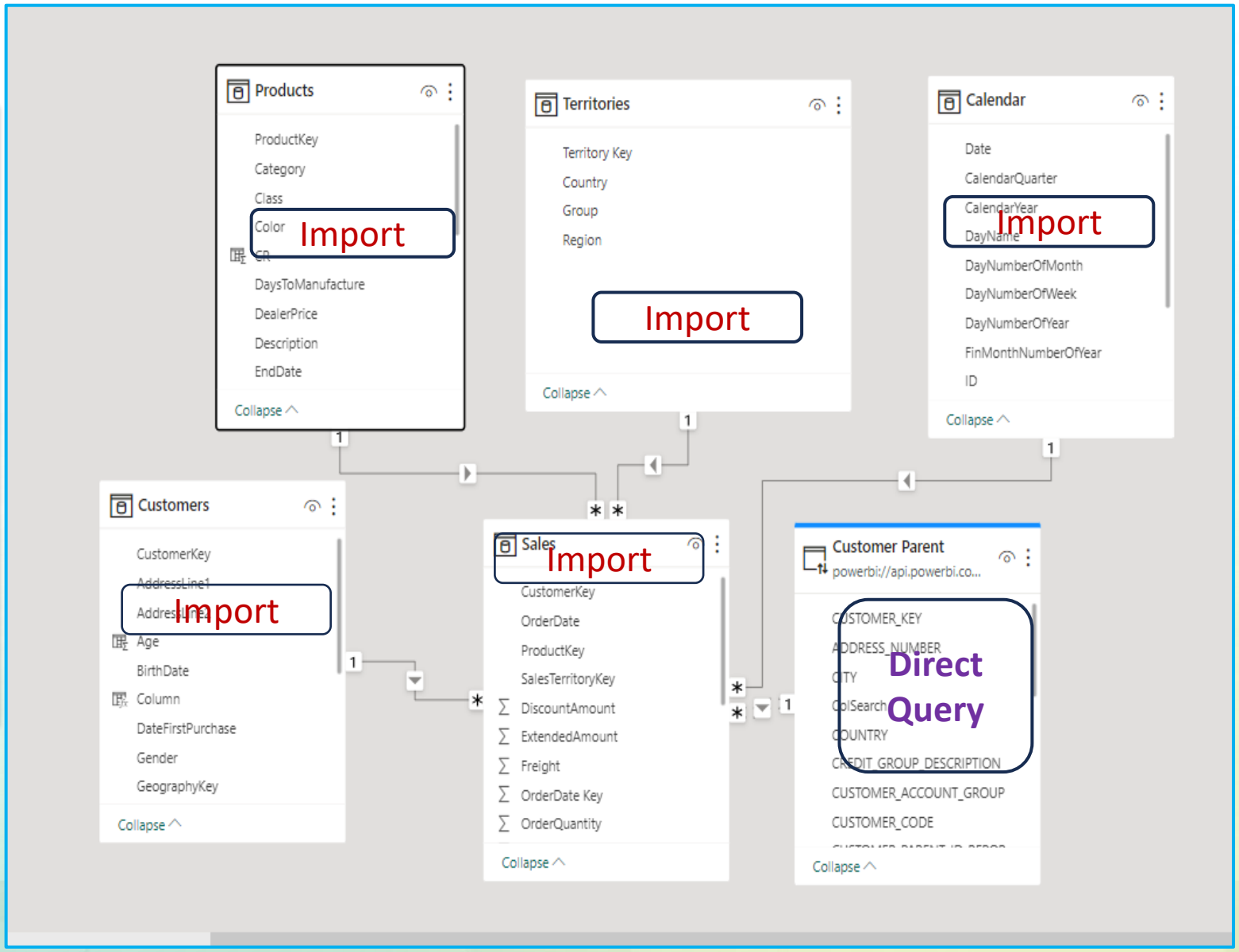
Live Connection Mode



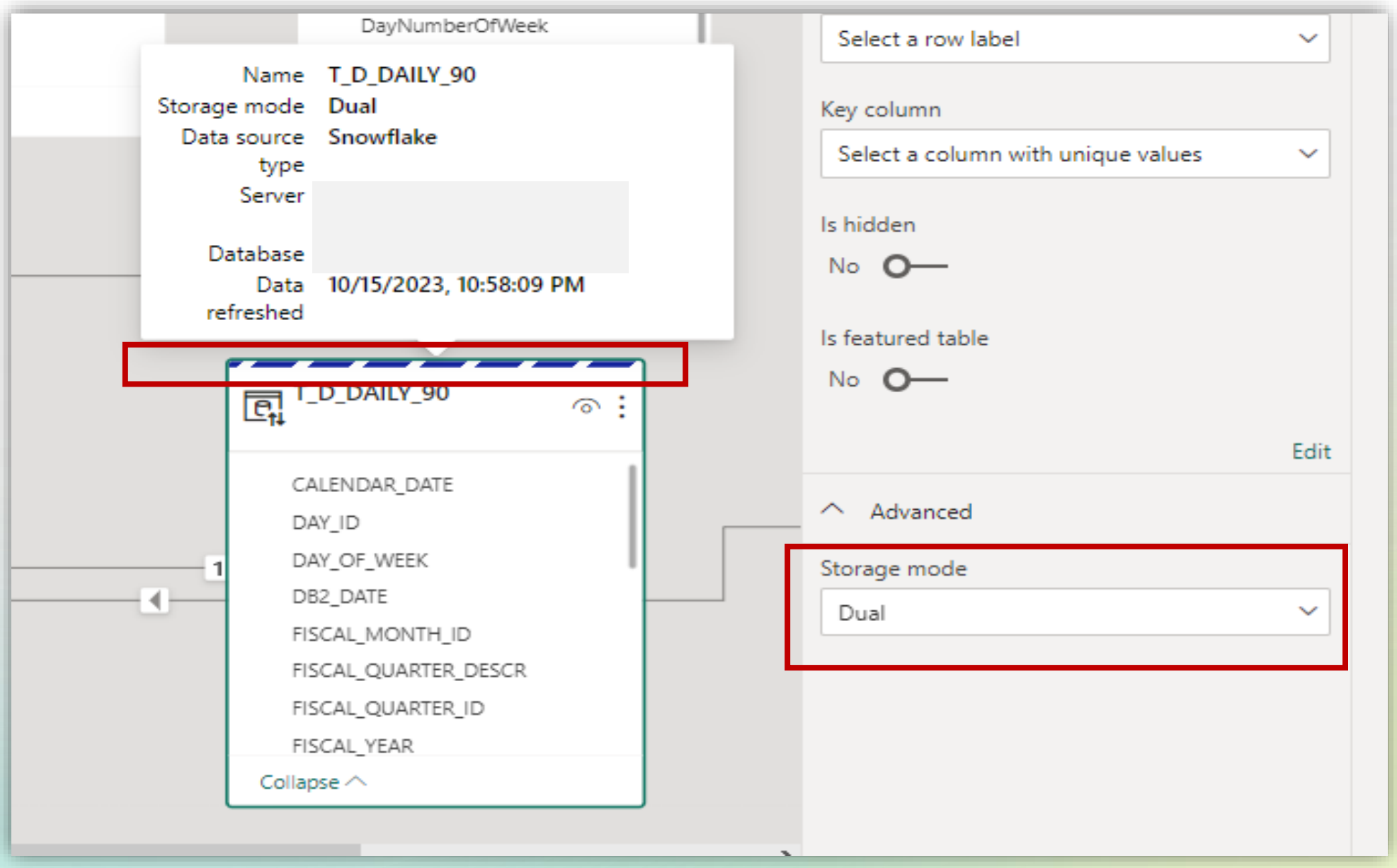
Mixed Mode



Storage Mode: Mixed



Dual Mode


DayNumberOfWeek
Name T_D_DAILY_90
Storage mode Dual
Data source type Snowflake
Server
Database
Data refreshed 10/15/2023, 10:58:09 PM
T_D_DAILY_90
CALENDAR_DATE
DAY_ID
DAY_OF_WEEK
DB2_DATE
FISCAL_MONTH_ID
FISCAL_QUARTER_DESCR
FISCAL_QUARTER_ID
FISCAL_YEAR
Collapse
Select a row label
Key column
Select a column with unique values
Is hidden
No
Is featured table
No
Edit
Advanced
Storage mode
Dual

Incremental Refresh & Hybrid Tables Overview

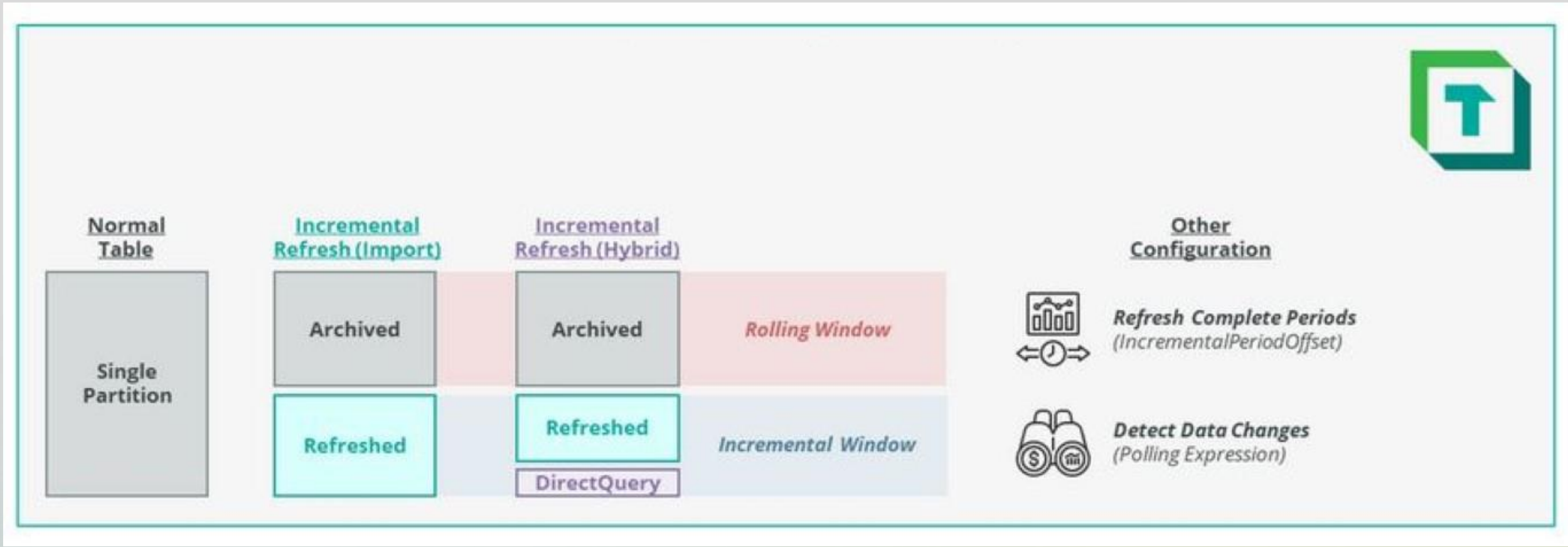
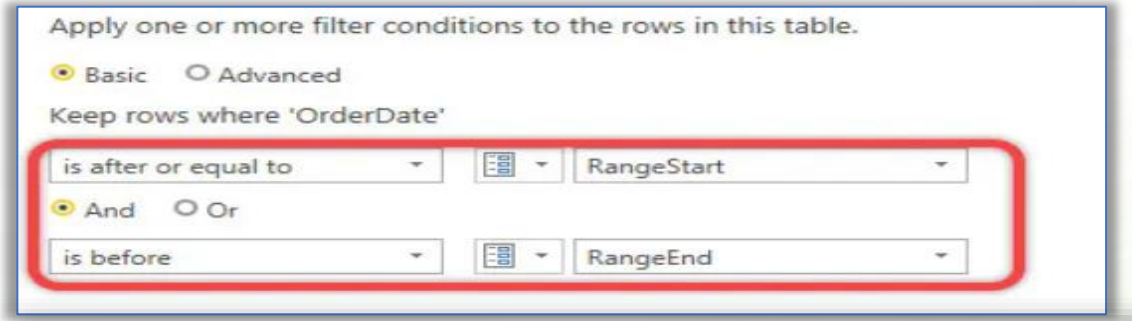


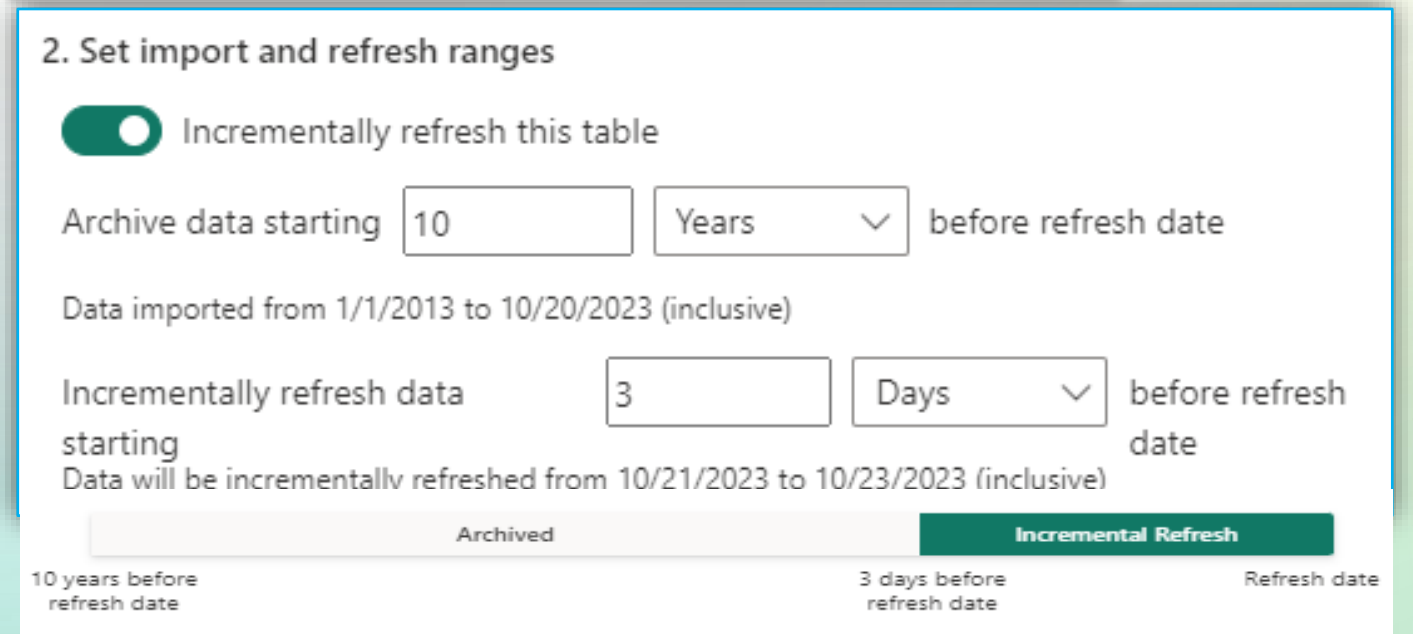
Image from docs.tabulareditor.com

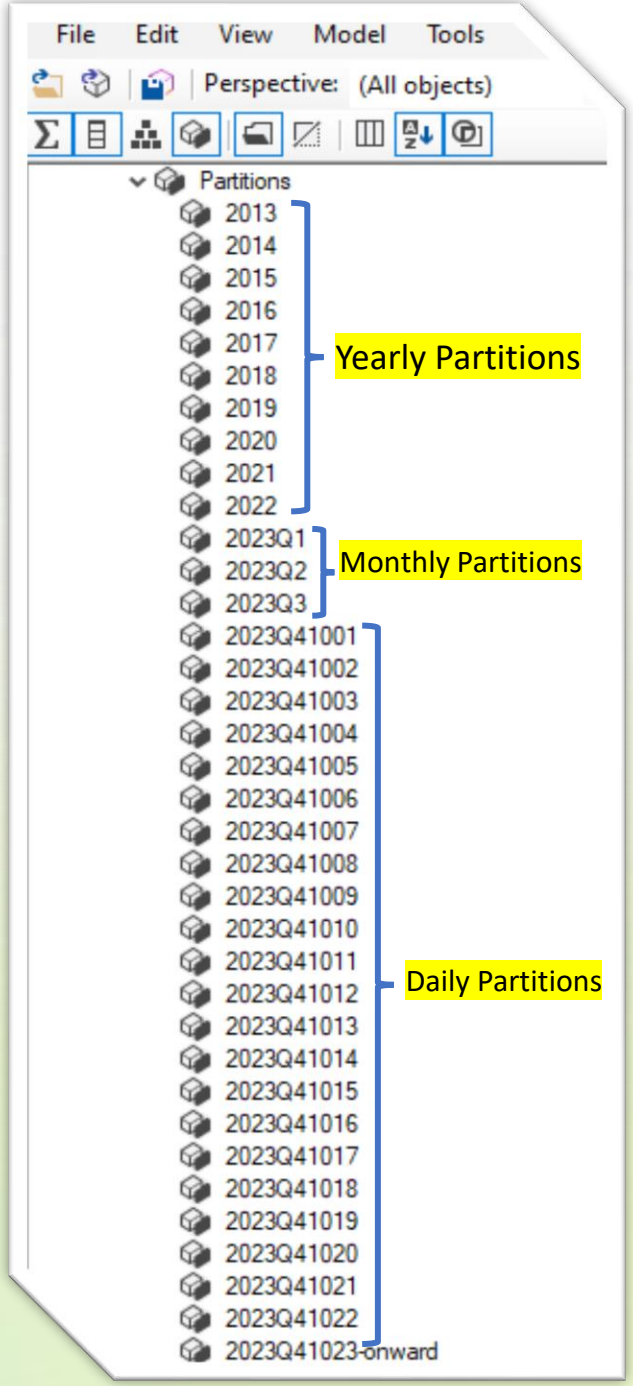
Incremental Refresh

Key Benefits of IR

- Lower Refresh Times
- Stability and Reliability -Reduce the risk of refresh failures
- Less Resource Consumption
- Cost Savings –When data transfer cost matters

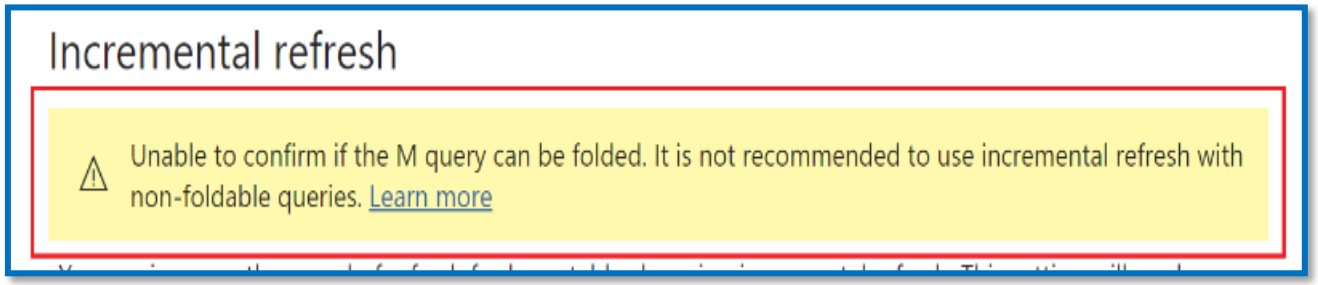






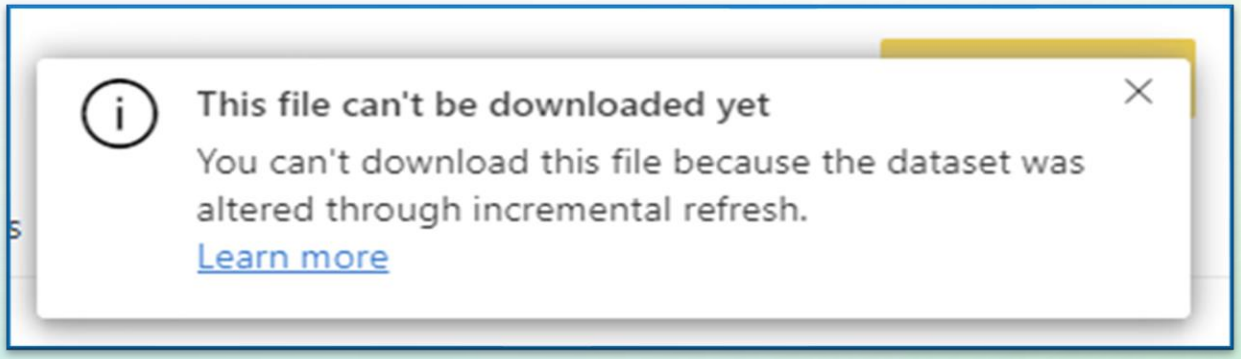
Partitions Considerations

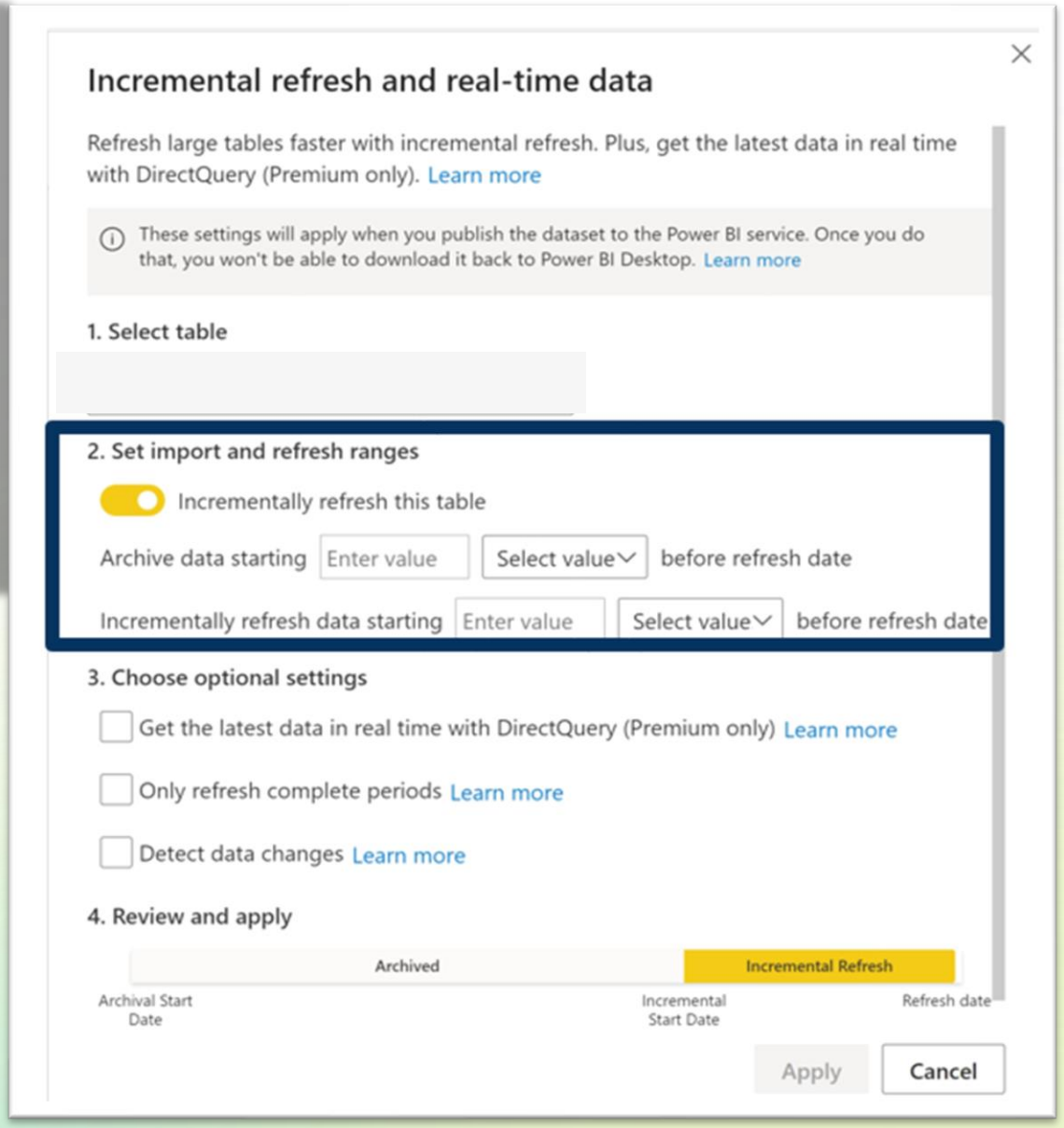
- Query Folding



- Make sure filtration happens at source.- (where clause matters)
- Too many partitions can hurt performance.

Things to Know Beforehand





Hybrid Mode

Incremental refresh and real-time data

1. Select table

TOP20_Fact

2. Set import and refresh ranges

Incrementally refresh this table

Archive data starting 47 Months before refresh date

Data imported from 10/1/2019 to 7/31/2023 (inclusive)

Incrementally refresh data starting 2 Months before refresh date

Data will be incrementally refreshed from 8/1/2023 to 9/30/2023 (inclusive)

3. Choose optional settings

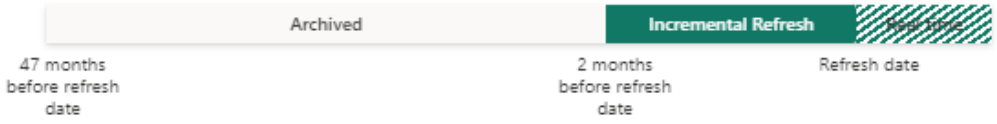
Get the latest data in real time with DirectQuery (Premium only) Learn more

Real-Time data will be from 10/1/2023 (inclusive) onwards

Only refresh complete months Learn more

Detect data changes Learn more

4. Review and apply



Combining import and DirectQuery makes this a hybrid table. Set your related tables, **COMPANY_MASTER, CREDIT_SEGMENT,** and **MONTH_ENDS_48 ...**, to dual mode to speed up their performance with hybrid tables. Learn more

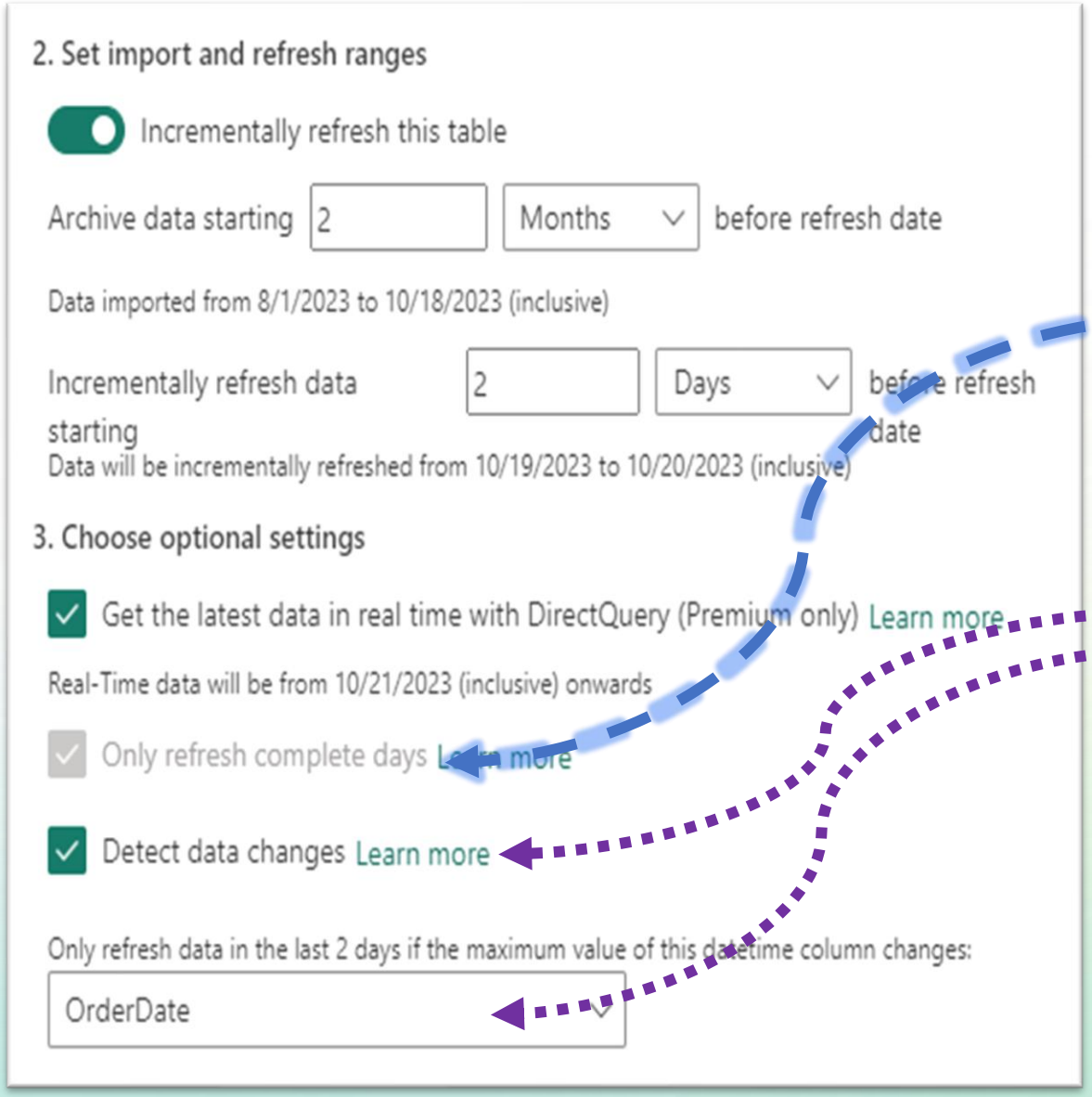
These settings will apply when you publish the dataset to the Power BI service. Once you do that, you won't be able to download it back to Power BI Desktop. Learn more

Storage Modes

- **Import**
- **Direct Query**
- **Dual**

Hybrid Table + Dual

Optional Settings In Incremental Refresh



- Once activated the **Hybrid** option; PowerBI automatically Activates the "Only refresh complete days/months" setting. This ensures all rows for the entire day are included in the refresh operation.

- Detect data changes setting enables even more selective refresh. Do not use the same date/time column used in IR.
  This further reduce the days that is incrementally refreshed .

Direct Lake Mode

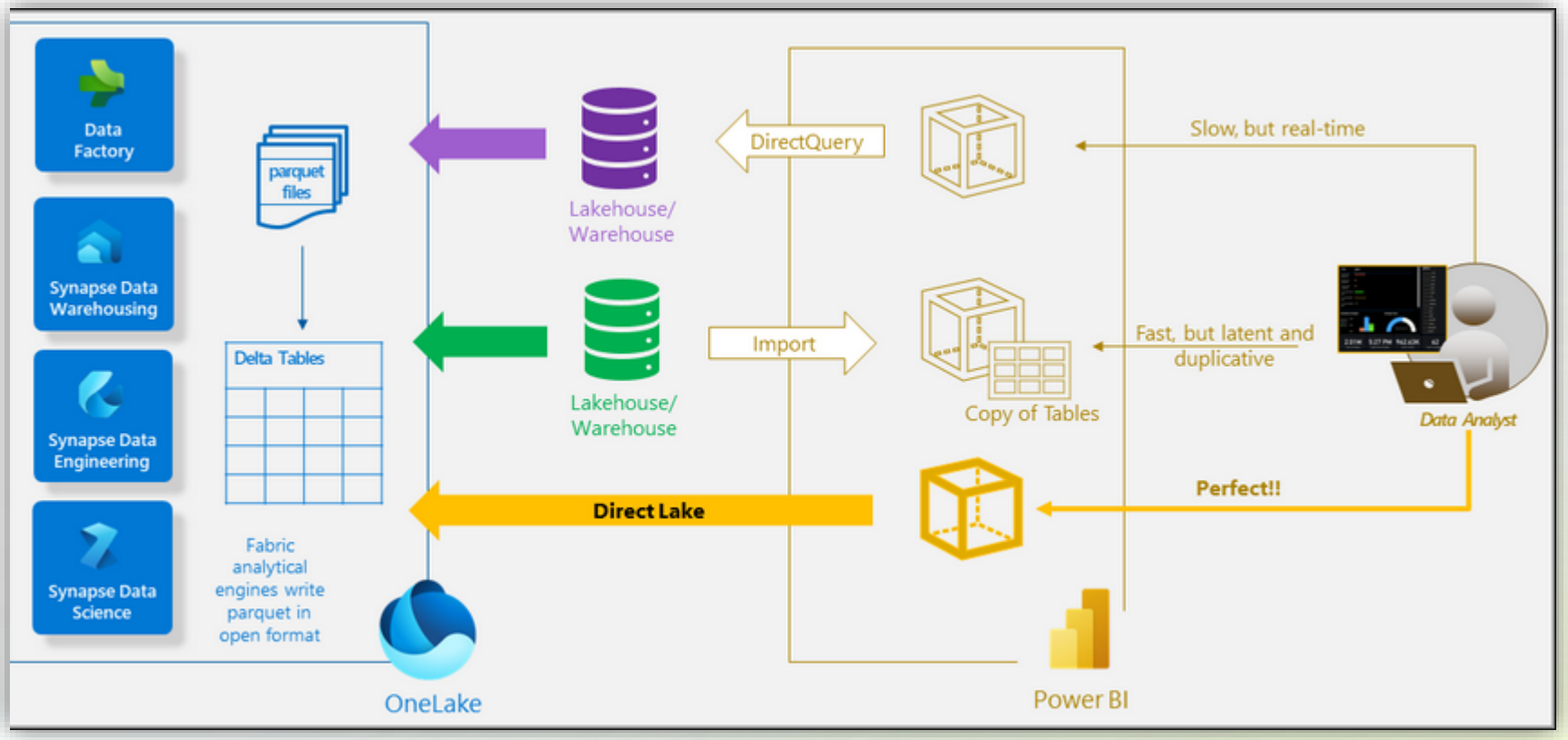
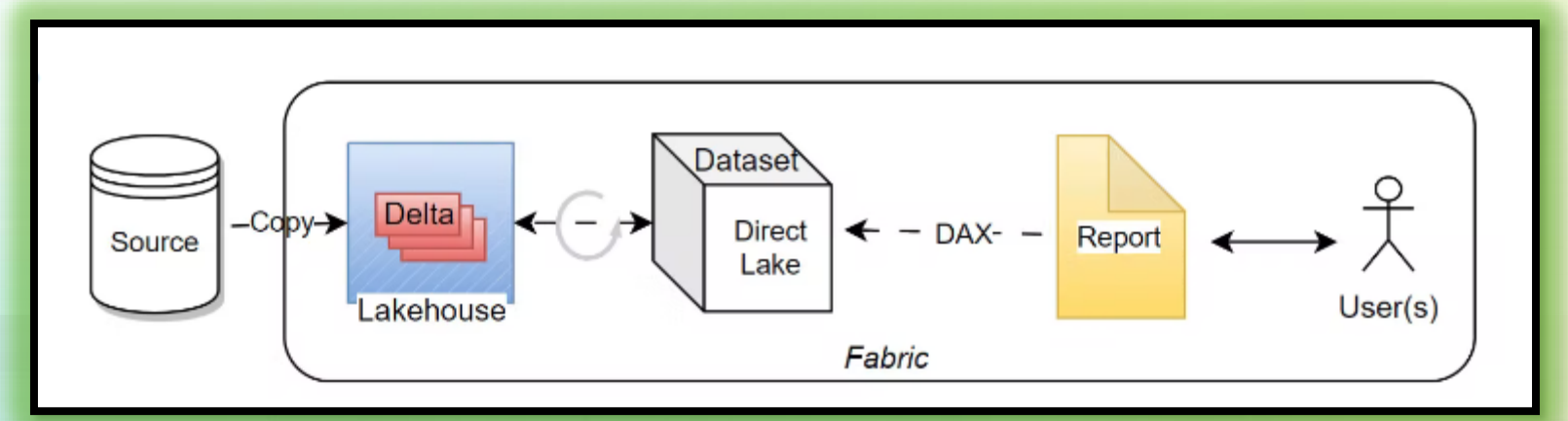


Image from Microsoft




DEMO
NEW STARS
OF DATA

Reverse Hybrid Workaround

Hybrid Mode

Granularity	Name	Row Count
Year	2011	295,489,717
Year	2012	297,678,498
Year	2013	295,575,442
Year	2014	292,477,875
Year	2015	297,780,469
Year	2016	294,060,081
Year	2017	300,419,682
Year	2018	296,541,108
Year	2019	292,787,420
Year	2020	299,273,979
Quarter	2021Q1	74,135,277
Month	2021Q104	24,939,498
Day	2021Q10501	820,805
Day	2021Q10502	826,885
Day	2021Q10503	821,043
Day-DirectQuery	2021Q10504-DQ	271,110
Total		**3,063,898,887**

Archived: **Import**

Incremental refresh: **Import**

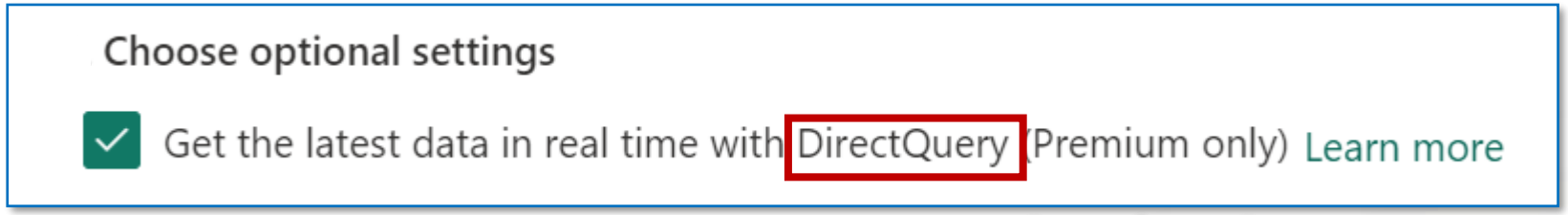
Real time: **DirectQuery**

Reverse Hybrid Mode

History Data : DirectQuery

Incremental Refresh: Import

Reverse Hybrid Workaround



Reverse Hybrid With Incremental Refresh

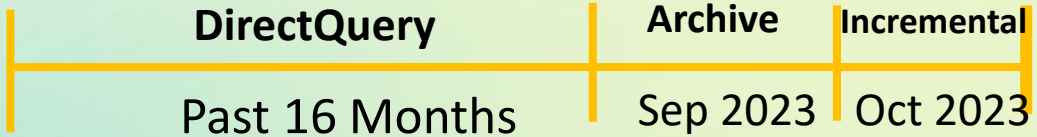
Approach
History Data –DirectQuery
Latest Data- Import

```
let

NewStart =if RangeEnd > DateTime.LocalNow() then Date.StartOfMonth(Date.AddMonths(DateTime.LocalNow(),-18)) else RangeStart,
NewEnd =if RangeEnd > DateTime.LocalNow() then Date.StartOfMonth(Date.AddMonths(DateTime.LocalNow(),-1)) else RangeEnd,

// updated with Newstart and NewEnd Parameters
    Source = Snowflake.Databases("Your server Name"),
    db_EDW_Database = Source{[Name=DB,Kind="Database"]}[Data],
    _Schema = db_EDW_Database{[Name="Schema name",Kind="Schema"]}[Data],
    _Table = _Schema{[Name="Table Name",Kind="Table"]}[Data],
    #"Filtered Rows" = Table.SelectRows(_table, each [CALENDAR_DATETIME] >= NewStart and [CALENDAR_DATETIME] < NewEnd)
in
    #"Filtered Rows"
```

Credits to Pat Mahoney, Patrick LeBlanc @Microsoft



Custom Partitioning

- When no dates in the selected tables
- Align partitions with dataset and source
- Process partition in parallel, Speed up the process

([DAX]) **[To the Rescue PowerBI Pro Users]**

1

- Hybrid tables, with No Premium and With No XMLA endpoint ? --**Yes Possible !!!**

(By mimicking the Hybrid scenario using DAX – Splitting into two storage modes(DQ, import) and then apply the date ranges selectively.)

(👏**Credits to Nikola Ilic**)

⚠ **No incremental approach, Don't mix IR with custom partitioning.**

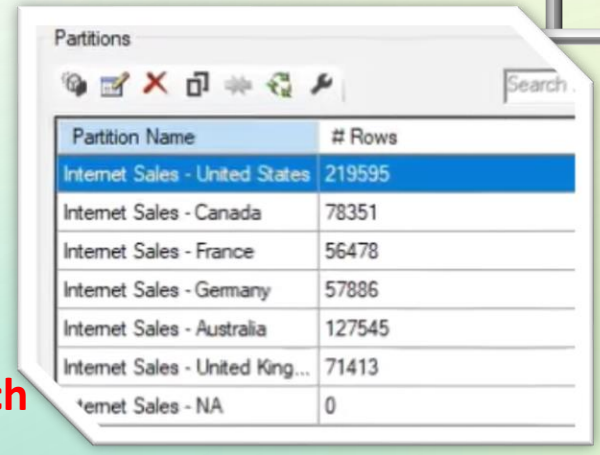
No dates scenario

3

- Using Premium,XMLA support.
- TE and M scripting helps to do this.
- In this ex:,Partitions created based on countries.

⚠

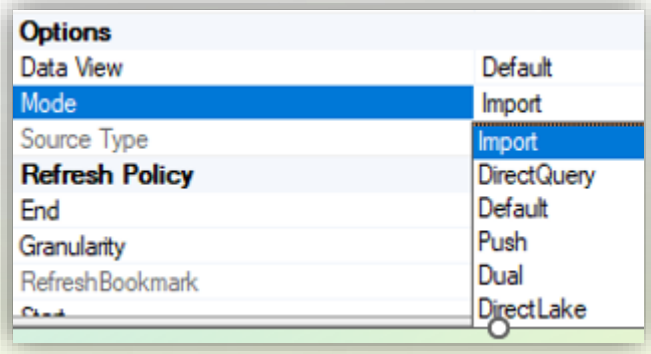
No incremental approach



Reverse Hybrid with TE] TABULAR EDITOR

2

- Using Premium , XMLA support + M scripting
(Mimicking the Hybrid scenario using M, changing the storage modes.)



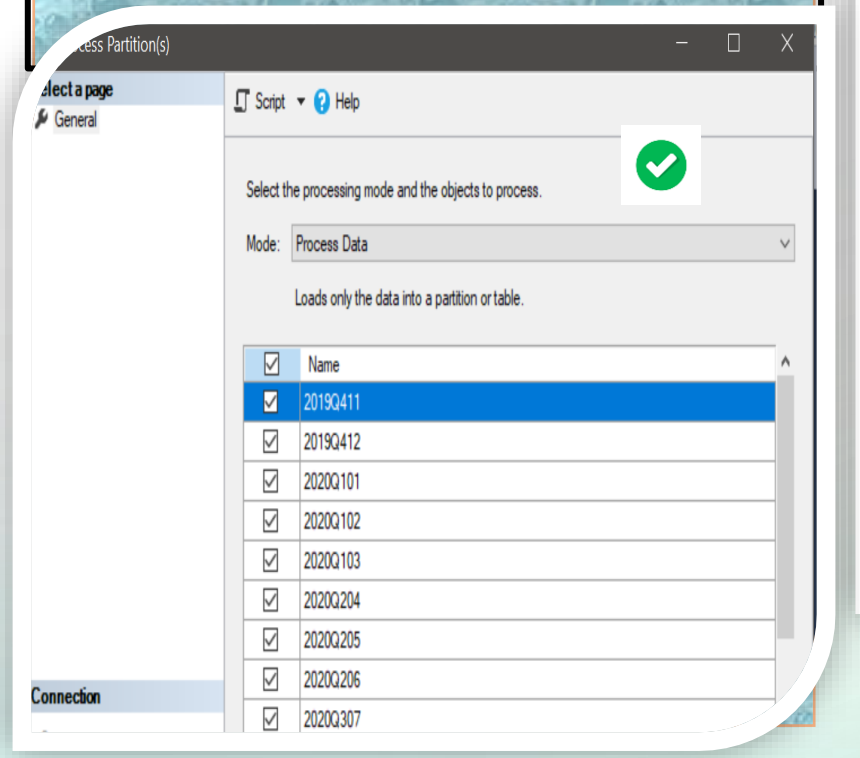
No incremental approach, Don't mix IR with custom partitioning.

⚠

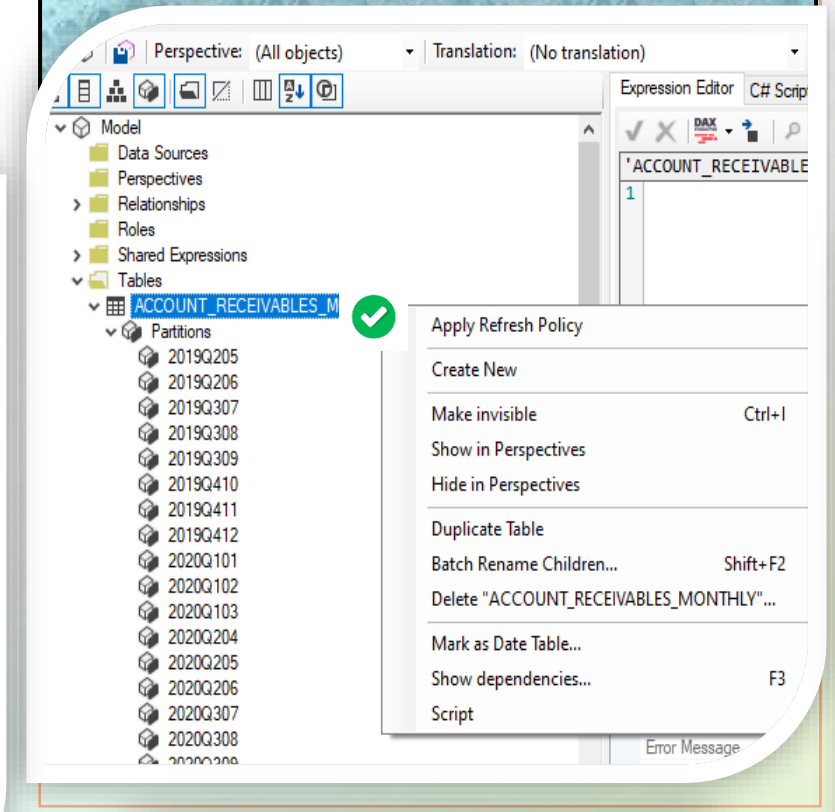
Supportive tools for better partition management

SSMS(SQL Server Management Studio)

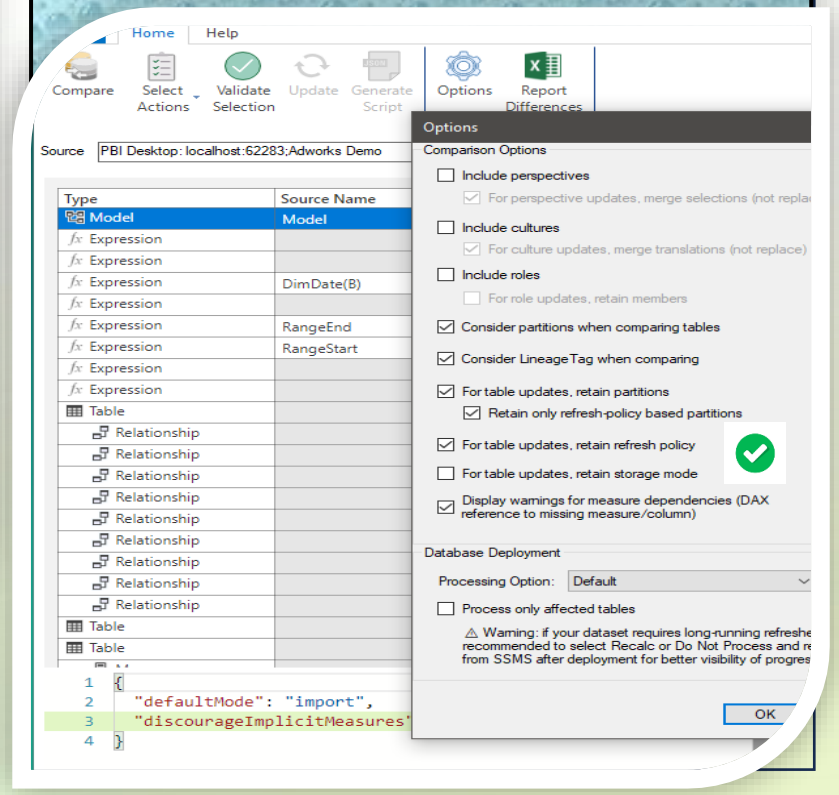
Helpful for bootstrap initial refresh



Tabular Editor



ALM Toolkit



DAX STUDIO

Further Resources

- Incremental refresh for datasets and real-time data in Power BI - Power BI | Microsoft Docs
- Announcing Public Preview of Hybrid Tables in Power BI Premium | Microsoft Power BI Blog | Microsoft Power BI
- Hybrid Tables, Incremental Refresh and Table Partitioning in Power BI | Paul Turley's SQL Server BI Blog
- Power BI Hybrid Tables with Synapse Analytics Serverless SQL Pools – Serverless SQL
- Automatic page refresh in Power BI Desktop - Power BI | Microsoft Docs
- Demystifying Custom Partitions in Power BI
- Setup a Reverse Hybrid Table without using Tabular Editor for Power BI!
- Power BI Hybrid tables for poor people!
- Incremental Refresh and Hybrid Tables in Power BI


THANK YOU
NEW STARS OF DATA
linkedin.com/in/nalakan
Blog/YT : Coming Soon