Aggregations in Power BI

Shabnam Watson

Shabnam Watson
BI Consultant

OFFICIAL BLOGGER — PASS SUMMIT

- in /ShabnamWatson
- 🐦 @shbWatson
- ✉ info@abicube.com

BLOG: https://shabnamwatson.wordpress.com

Work: BI Consultant

Almost 20 years of experience developing data warehouse and business intelligence solutions with focus on Analysis Services and Power BI.

Speaking/Community

PASS Summit, SQL Saturdays, PASS Women In Technology Virtual Chapter, .NET South, BI and SQL Server user groups. SQL Saturday Atlanta BI Organizer

Background

Bachelor's degree in Computer Engineering. Master's degree in computer science. Certified Business Intelligence Professional (CBIP) by The Data Warehouse Institute (TDWI).

About you:

Used PBI to build reports from existing models
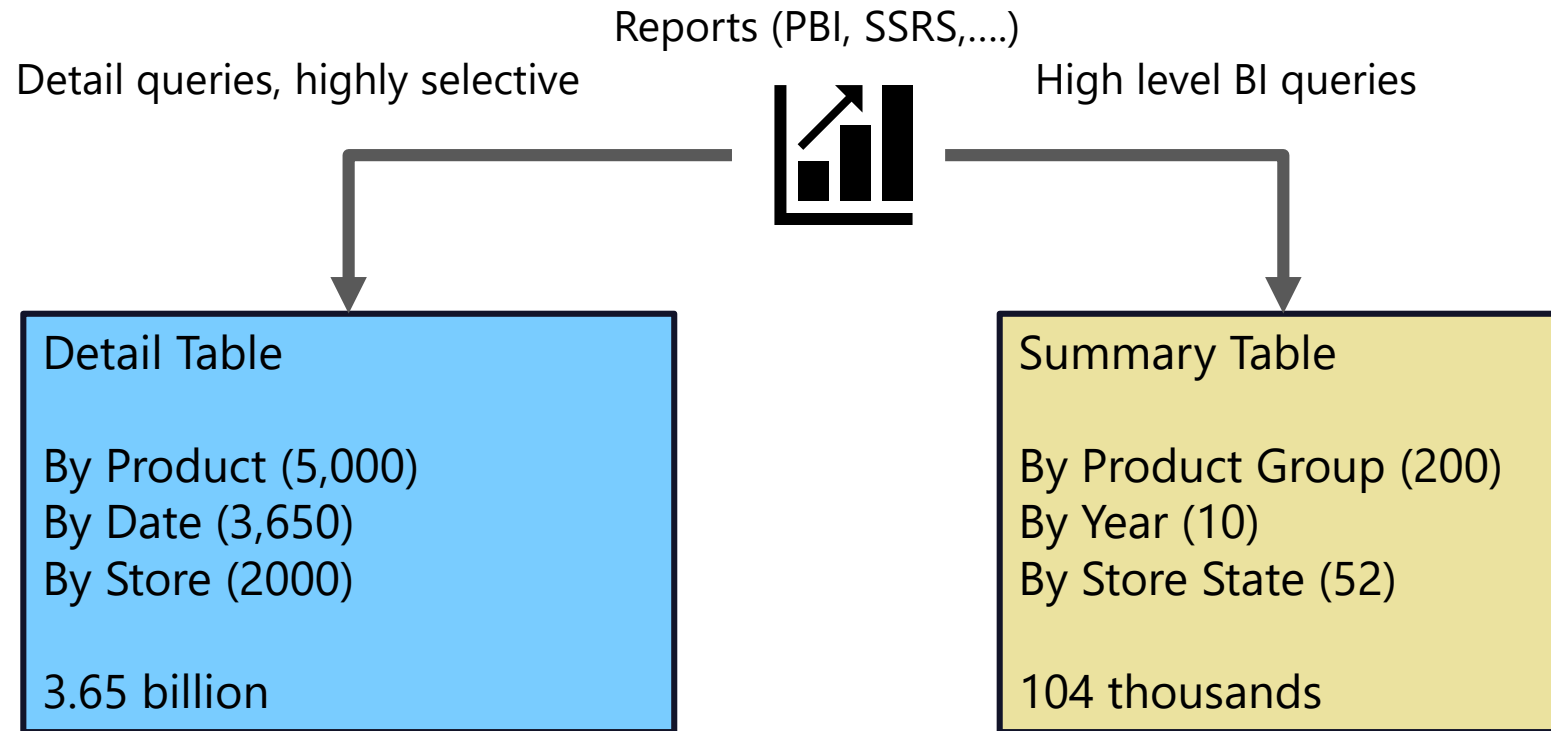
Some modeling experience

Advanced modeler

Agenda

- Concept
- AdventureWorks
- Model: Storage modes, Direct Query, Composite models
- DAX Studio

Aggregations

Performance optimization technique. Summary Table.

Reports (PBI, SSRS,....)

Detail queries, highly selective   High level BI queries

Detail Table	Summary Table
By Product (5,000)	By Product Group (200)
By Date (3,650)	By Year (10)
By Store (2000)	By Store State (52)
3.65 billion	104 thousands

Hybrid Architecture

Balanced architecture: spread query load.

Big Data (IOT)	Data Warehouse (CCI)	PBI cache

Detail Table	Aggregation Table 1	Aggregation Table 2
By Product (5,000) By Date (3,650) By Store (2000) 3.65 billions	By Product Group (200) By Date (3,650) By City (520) 379.6 millions	By Product Group (200) By Year (10) By Store State (52) 104 thousands

Detail queries	Some BI queries	Executive Dashboard

Aggregations in Power BI

- GA since July's 2019 version of PBI Desktop.

- Supported in PBI desktop and Service (**shared and premium**) with RLS.

- Once configured, report developers don't need to know about them.

- There is a cost and maintenance involved so usage has to be monitored.

Aggregations Benefit

- Improved query performance.

- Not just for quarter Petabyte, trillion rows datasets. TrillionRowDemo

- For any datasets that is expensive to cache (Time,CPU,Memory)

Power BI Desktop

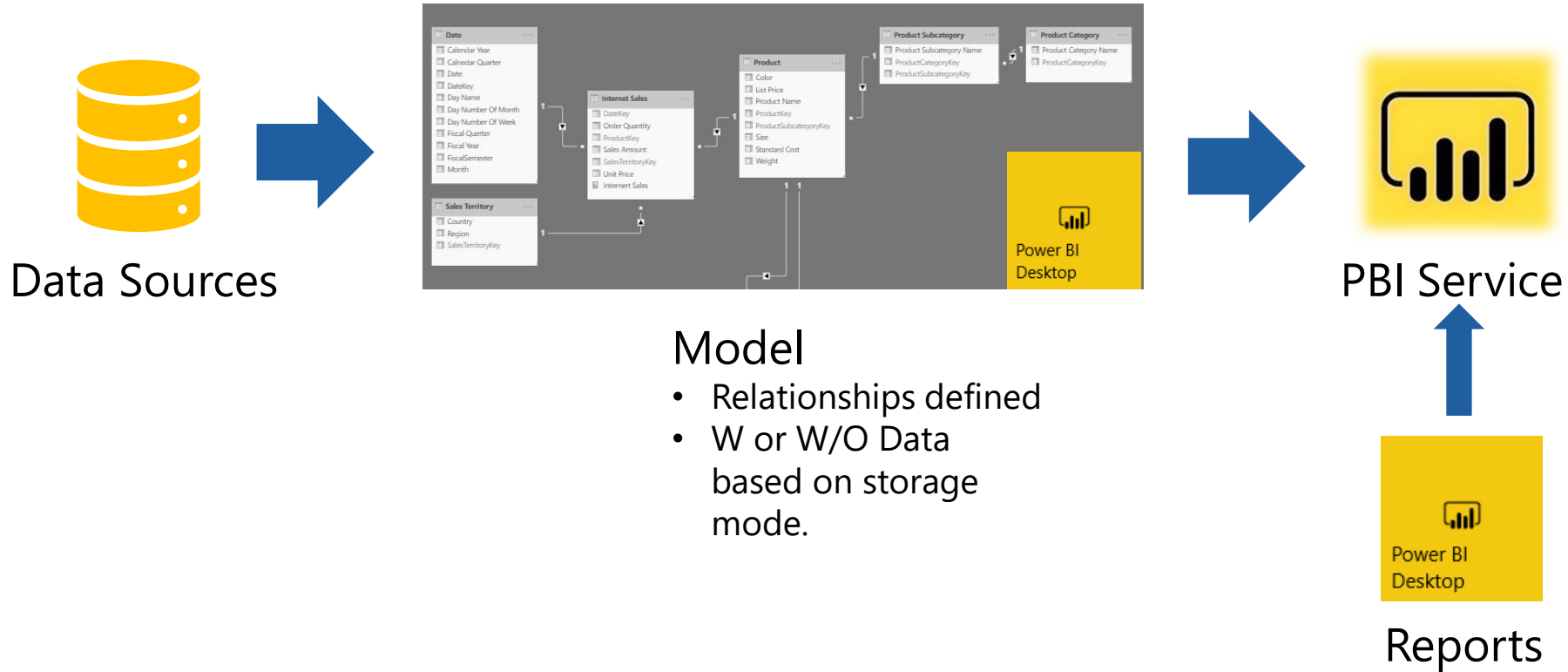## Power BI Desktop: Front End Reporting tool.

Live connect to PBI Datasets

Live connect to Analysis Services Databases

Power BI Desktop: Modeling tool.

Create a semantic model with or w/o data.
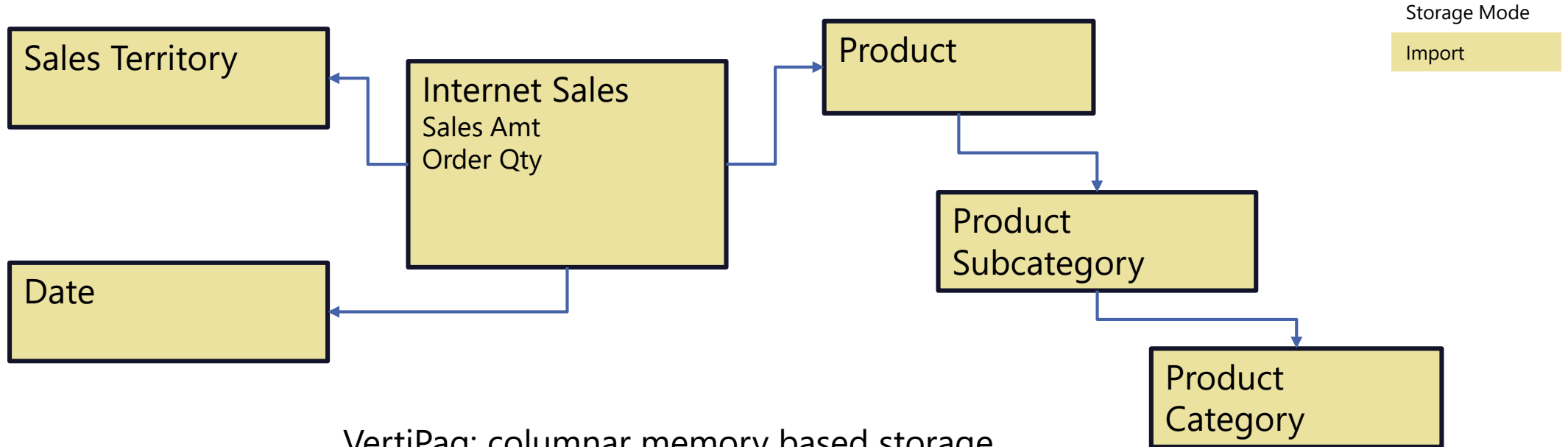
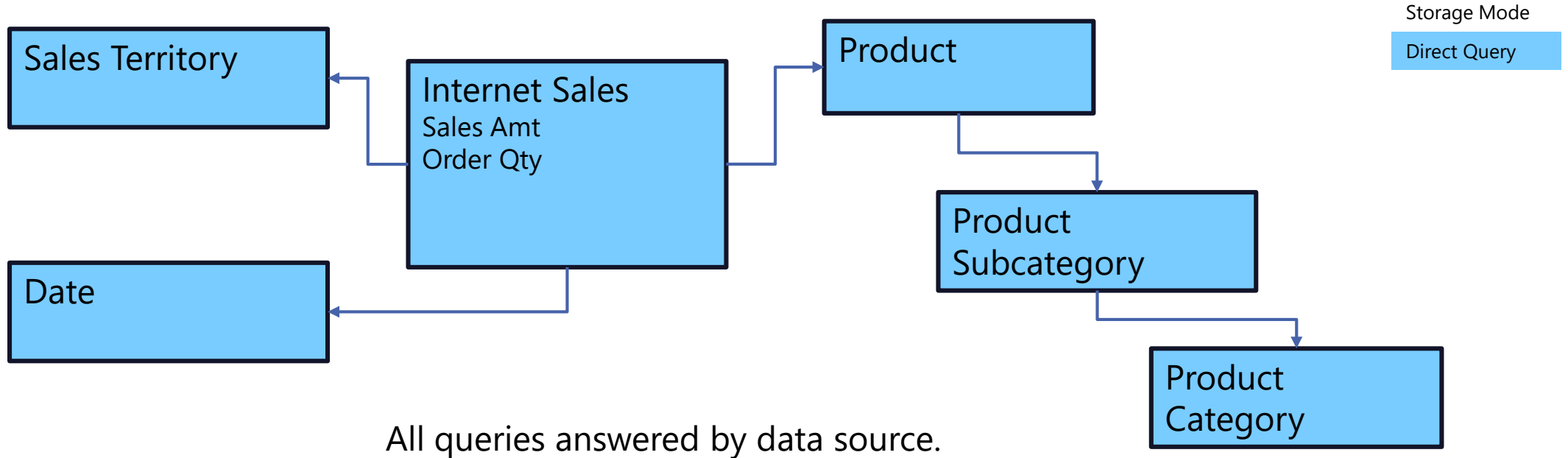Enhance the model with **Aggregations**.

Power BI Models

Data Sources

Model
- Relationships defined
- W or W/O Data based on storage mode.

PBI Service

Reports

Import Mode

Sales Territory

Internet Sales
Sales Amt
Order Qty

Date

Product

Product
Subcategory

Product
Category

Storage Mode

Import

VertiPaq: columnar memory based storage

All queries answered from memory.

Direct Query Mode

Sales Territory

Internet Sales
Sales Amt
Order Qty

Date

Product

Product
Subcategory

Product
Category

Storage Mode

Direct Query

All queries answered by data source.

Import vs. Direct Query

Import: Best for most models

In memory

Vertipaq engine: Columnar storage engine

Super fast

Load/processing time

Direct Query

PBI model: Semantic model

Near real time (low) latency.

ta does not fit into in-memory
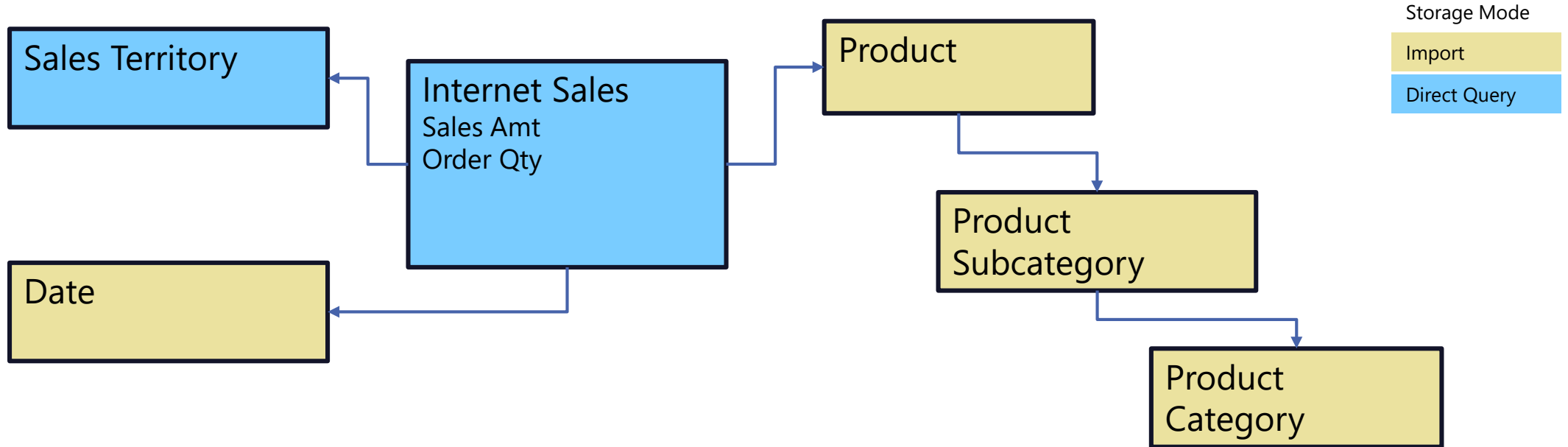s.

ocessing time.

e slow with large tables.
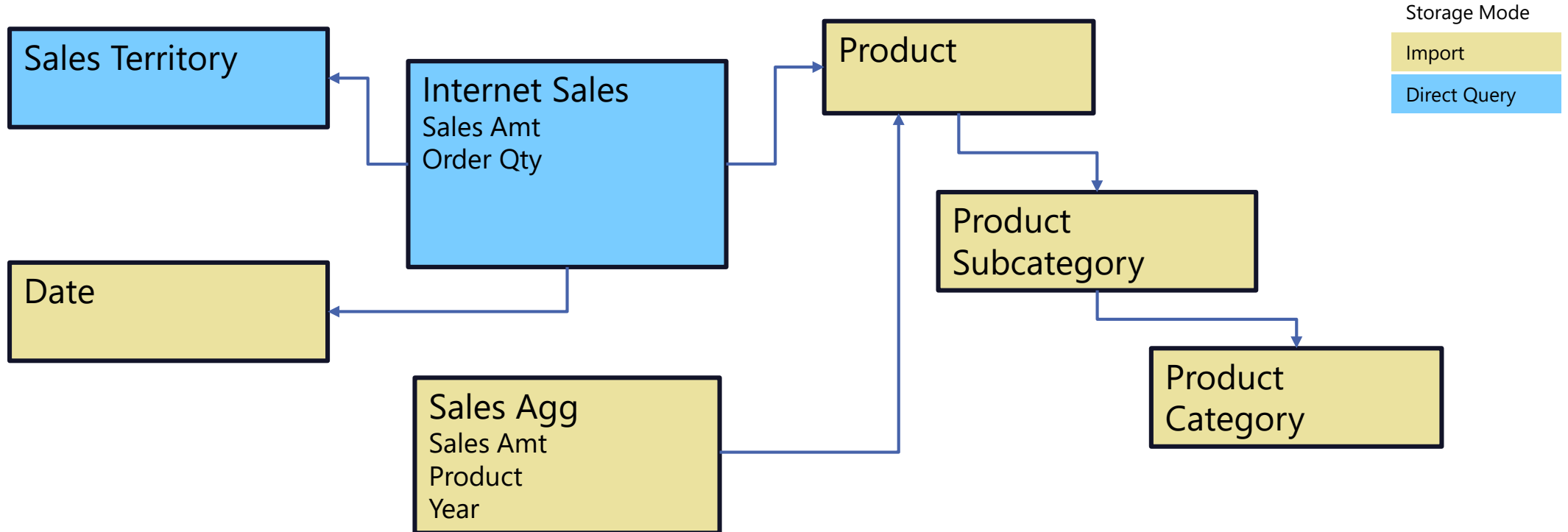
Composite Models

Select storage mode for each table.

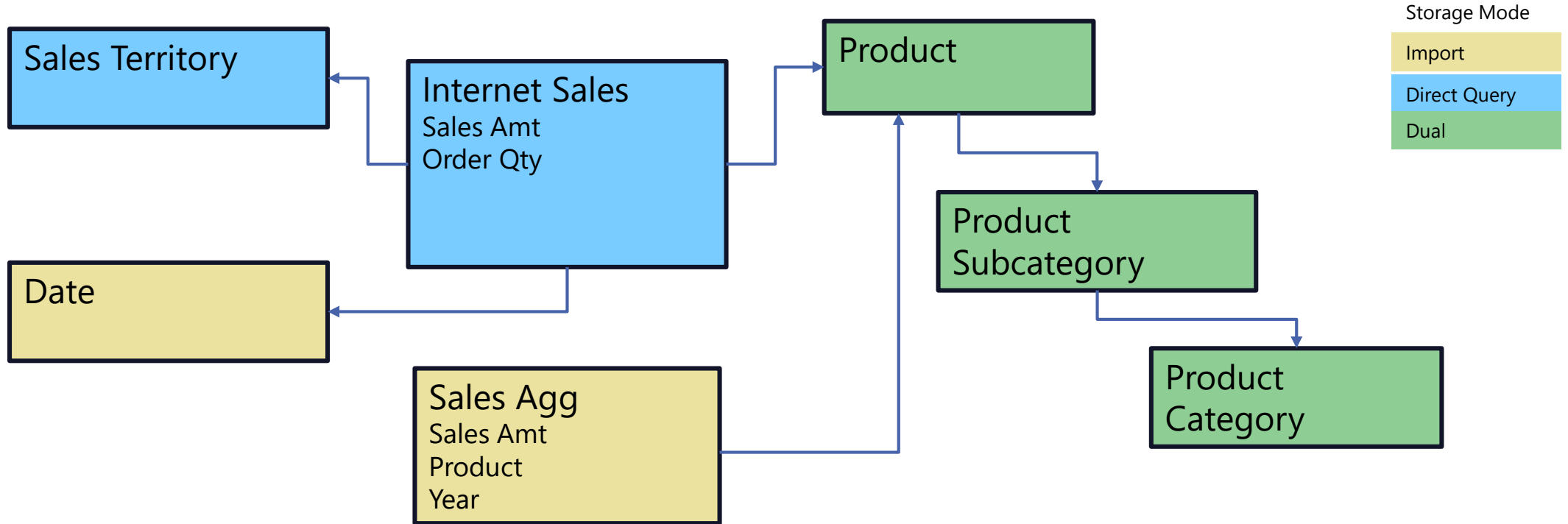Leave big tables in the data source, load smaller tables into memory

Queries answered either from memory or the data source.

Aggregation Tables

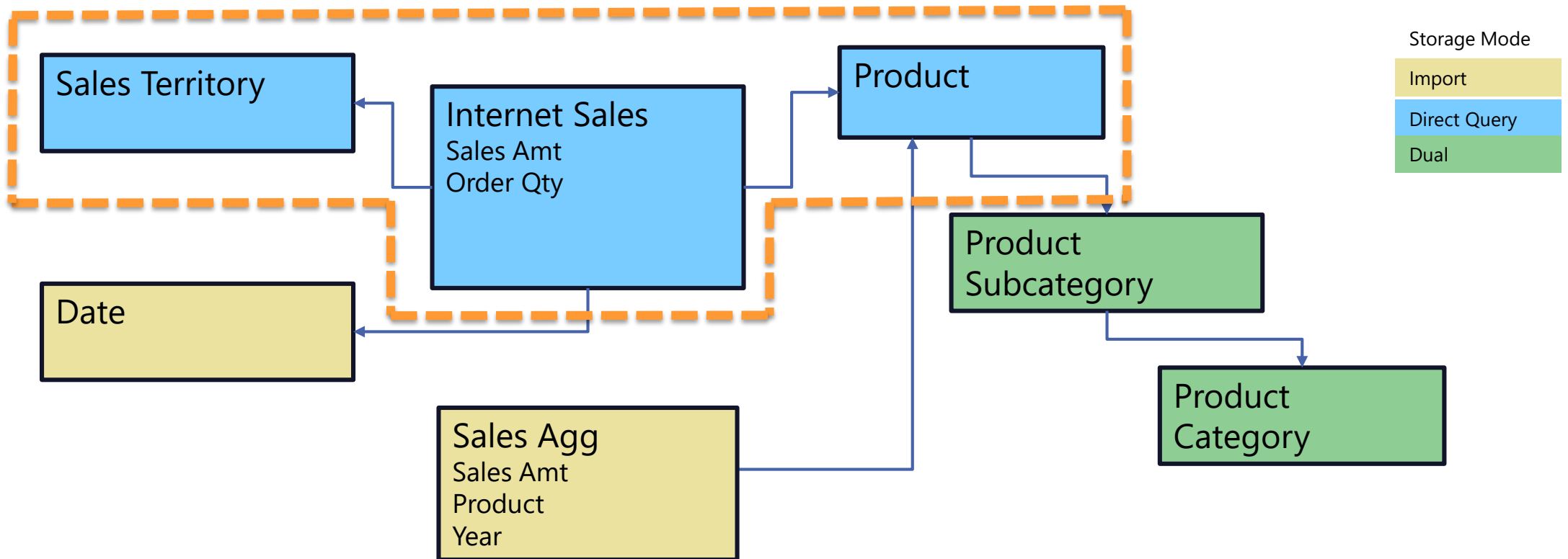Queries answered either from memory or the data source.

Dual Storage mode

Storage Mode

Import	
Direct Query	
Dual	

Sales Territory

Internet Sales
Sales Amt
Order Qty

Product

Product Subcategory

Date

Sales Agg
Sales Amt
Product
Year

Product Category

Dual tables can flip back and forth depending on the query. Avoid having joins in PBI.
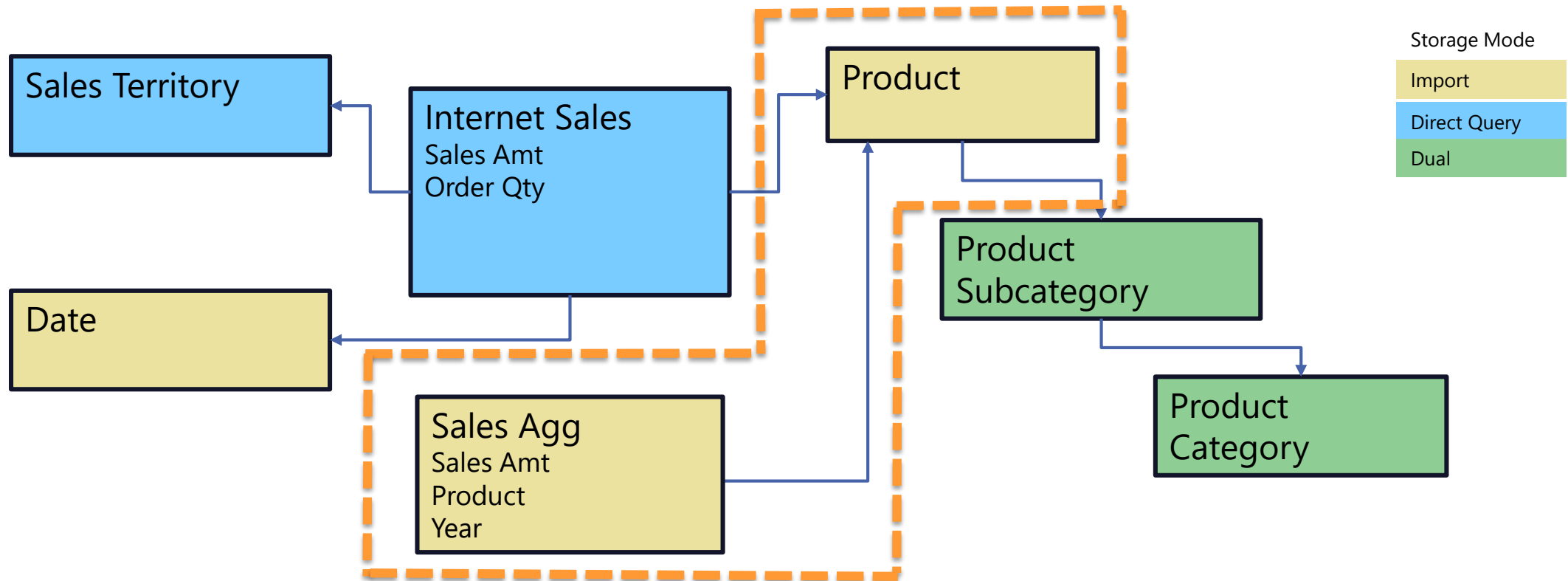
Dual Storage mode

Sales Amt by Product and Sales Territory

Storage Mode	
Import	
Direct Query	
Dual	

Sales Territory

Internet Sales
Sales Amt
Order Qty

Product

Date

Product Subcategory

Sales Agg
Sales Amt
Product
Year

Product Category

Product table acts as a Direct Query table.

Dual Storage mode

Sales Amt by Product

Storage Mode

Import	
Direct Query	
Dual	

Product table acts as an Import table.

Aggregation tables

A table like any other table in PBI:

- Direct Query*: Table/view in the data source. Optimized with indexes.
- Imported table in PBI model (w or w/o incremental refresh)
- M expression

Aggregation tables are automatically hidden.

Only addressable by dataset admin once published to PBI Service.

*Direct Query aggregation tables that use a different data source to the detail table are only supported if the aggregation table is from a SQL Server, Azure SQL or Azure SQL DW source.

Aggregations Hits based on relationships

Strong relationships include the following cases:

Many Side	One Side
Dual	Dual
Import	Dual or Import
Direct Query *	Dual or Direct Query *

* Both sides in Direct Query have to be from the same data source.

Many-to-many relationships are always considered weak.

How to set up Aggregations

Configure aggregation properties:

• Based on relationships.
• Based on Group Bys.

Manage aggregations

Aggregations accelerate query performance to unlock big-data sets. Learn more

Aggregation table   Precedence ⓘ

AggTableSalesAmountByCalendarYearAndProduct ▾   0

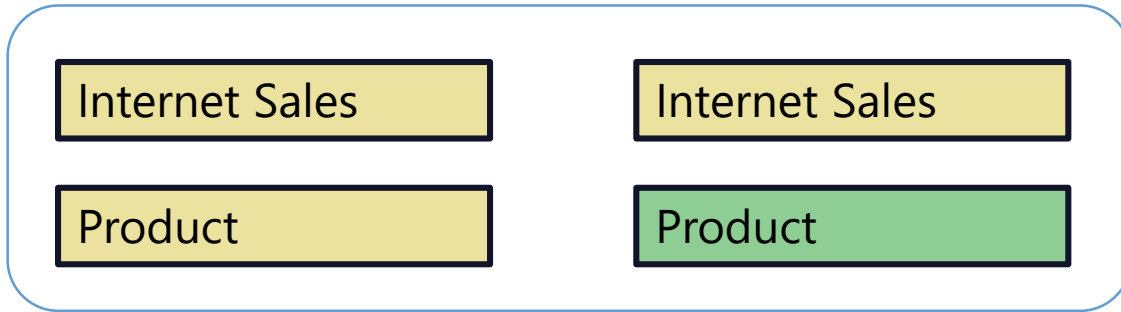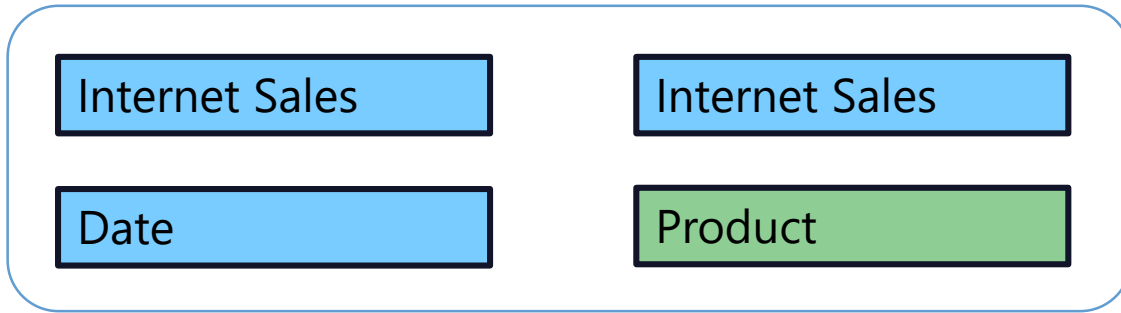| AGGREGATION COLUMN	SUMMARIZATION	DETAIL TABLE	DETAIL COLUMN	
CalendarYear	GroupBy ▾	Date ▾	Calendar Year ▾	🗑
ProductKey	Select Summarizatio... ▾	▾	▾	🗑
SalesAmount	Sum ▾	Internet Sales ▾	Sales Amount ▾	🗑

Demo

Group By Columns

- Typically for big data (Hadoop, HDInsight)
  1. Header (trillion rows)/Detail tables (reduced cardinality).
  2. Fact table (file) that is denormalized into one big table.
  3. No relationships between tables.

- Different granularity between Agg table and dimension table. Agg table is at a higher grain than the dim table and you don't want to normalize all of your dim tables.

- Distinct Counts: On the foreign key column of the detail table.

Weak and Strong relationships

Strong relationships: Agg hit 😊

Weak relationship: No Agg hit 😠

Internet Sales	Internet Sales
Date	Product

Internet Sales
Date

Internet Sales	Internet Sales
Product	Product

😊
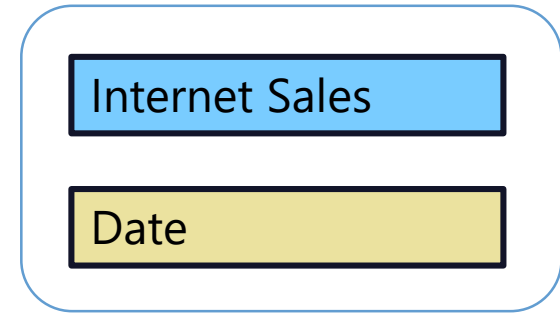
Storage Mode

Import
Direct Query
Dual

Monitor Aggregation Usage

Aggregations have a cost: Memory/ CPU/Time
Are they actually being used?

- SQL Server Profiler/Extended Events:

- Query Processing\Aggregate Table Rewrite Query

- DAX Studio

Aggregations and Row Level Security (RLS)

- Row level security (RLS) expressions should filter both the aggregation table and the detail table to work correctly.

RLS applies to Details Table	RLS applies to Agg Table	Agg Hit
Yes	Yes	Yes
Yes	No	No
No	Yes (fails, PBI Stops you)	N/A

- You cannot query the Agg table (even without RLS) unless you are an admin on the dataset.

Aggregations - Precedence

Multiple layers of Aggregations

Higher Precedence property: Higher priority.

Main table: 1 billion rows

Aggregation 1: 1 million rows

Aggregation 2: 100,000 rows

Aggregation 3: 1000 rows

Maintaining aggregation tables

Whether in import mode or an actual table in the source database system, Aggregation tables can get out of sync.

When an Aggregation table is out of sync, the Detail table and Aggregation table can return different numbers. PBI will does not try to detect and direct all queries to the source database system.

It is the responsibility of the designer to know their data flow patterns and keep the aggregation tables up to date.

When not to use aggregations

They are not useful for queries that iterate data row by row.

Create a calculated column in the table and then use aggregations on top.

Summary

- Aggregations are an excellent tool for optimizing performance of high level BI queries often used in PBI reports and dashboards.

- Aggregations are not just for quarter peta byte datasets.

- Currently only on Direct Query detail tables, not import.

- Aggregations work with RLS.

- Aggregations have a maintenance cost. Make sure they are used.

- <u>Good data modeling, is still the number one best practice for good performance.</u>

References

PBI Blog: https://powerbi.microsoft.com/en-us/blog/aggregations-for-petabyte-scale-bi-is-generally-available/

Online Help: https://docs.microsoft.com/en-us/power-bi/desktop-aggregations

PBI YouTube channel, Peta Byte Aggregation demo, July 2019:

https://www.youtube.com/watch?time_continue=9&v=EWPVsa64aEQ

Original trillion row demo, PASS SUMMIT 2017: https://aka.ms/TrillionRowDemo

Thank you!

What question do you have for me?

Contact

in	/ShabnamWatson
🐦	@shbWatson
✉	info@abicube.com

https://shabnamwatson.wordpress.com

OFFICIAL PASS SUMMIT BLOGGER

Practice Slide