

# Aggregazioni in Power BI

Marco Pozzan



*Own your career with interactive learning built  
by community and guided by data experts.  
Get involved. Get ahead.*

# Explore your PASS community



 PASS  
MARATHON

Free online  
webinar events



PASS  
LOCAL  
GROUPS

Local user groups  
around the world



 PASS  
SUMMIT


Connect with the global  
data community



PASS  
VIRTUAL  
GROUPS

Online special  
interest user groups



 PASS CONNECTOR  
INSIGHTS

Learning on-demand  
and delivered to you



PASS  
VOLUNTEERS

Get involved



 PASS  
**SQLSATURDAY**  
PORDENONE | 30 MAY 2020

# Missed PASS Summit 2019?

## Get the Recordings

Download all PASS Summit sessions on Data Management, Analytics, or Architecture for only \$399 USD

More options available at  
**PASSstuff.com**



\$399

**Content Stream**  
download  
non-attendee option



## Summit 2020 Will Launch

In-person and virtual event  
planning is underway.

**Register Now**

We are covering all bases to ensure our community can continue reaching new and exciting heights. Plans are underway for the in-person event you all know and love along with a new venture, a new opportunity: a PASS Summit 2020 Virtual Event.

**Find out more at [PASS.org/summit](https://PASS.org/summit)**

# Thank you to our Global Sponsors and Supporters





Thank you to  
our Local  
Sponsors and  
Supporters



PASS



Microsoft



SolidQ



Fluentis



altitudo



beanTech  
IT moves your business

**PALAZZETTI**  
IL CALORE CHE PIACE ALLA NATURA



**WÜRTH**PHOENIX

Real Comm  
Easy for real



# Speaker

Consulente e formatore in ambito business intelligence, business analytics e data mining

Dal 2002 le attività principali sono legate alla progettazione di data warehouse relazionale e alla progettazione multidimensionale con strumenti Microsoft.

Docente all'Università di Pordenone nel corso di analisi dei dati e Big Data

Community Lead di Innova ([www.innovazionefvg.net](http://www.innovazionefvg.net))

MCP, MCSA, MCSE, MCT SQL Server

dal 2014 MVP per SQL Server e relatore in diverse conferenze sul tema.

[info@marcopozzan.it](mailto:info@marcopozzan.it)

[@marcopozzan.it](https://twitter.com/marcopozzan)

[www.marcopozzan.it](http://www.marcopozzan.it)

<http://www.scoop.it/u/marco-pozzan>

<http://paper.li/marcopozzan/1422524394>



**Microsoft**  
CERTIFIED

Trainer

**Microsoft**  
CERTIFIED

Professional

**Microsoft**  
CERTIFIED

Solutions Associate

SQL Server 2012/2014

A large, stylized teal graphic on the left side of the slide, resembling a thick, curved arrow pointing right.

<https://www.menti.com/pi3xjo1qf3>





# Agenda

Import Mode vs Direct Query Mode

Concetto di Aggregazione

- Demo 1: Creazione delle aggregazioni

Composite Models

- Demo 2: Weak Relationship

Precendenze

Demo 3: Utilizzo delle precedenze

Mantenimento

Demo 4: Problemi nell'aggiornamento dati

# Import Mode vs Direct Query Mode

## **Direct Query mode**

Trasformazione da DAX a SQL (solo per database relazionali)

Query sul motore SQL Server aumentando il workload

Molte limitazioni (solo connessione SQL, no colonne calcolate, no gerarchie temporali)

No tempo di processing

Real-time queries

## **In Memory mode (vertipaq) (default)**

I dati vengono copiati in memoria

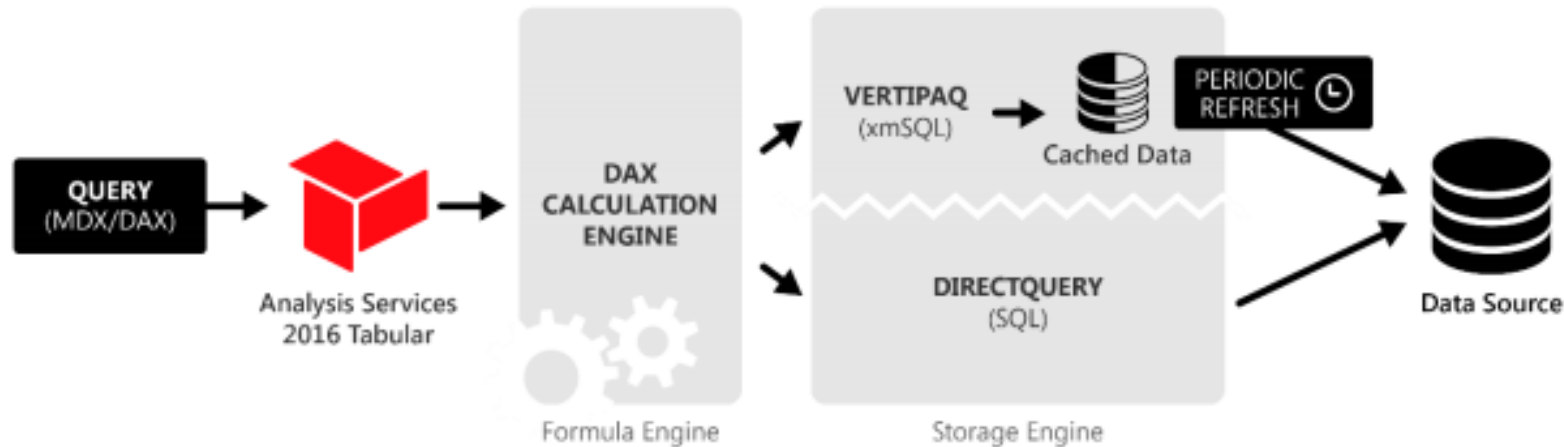
Singolo accesso alla sorgente dati

Le query girano su uno «snapshot dei dati»

# Import Mode vs Direct Query Mode

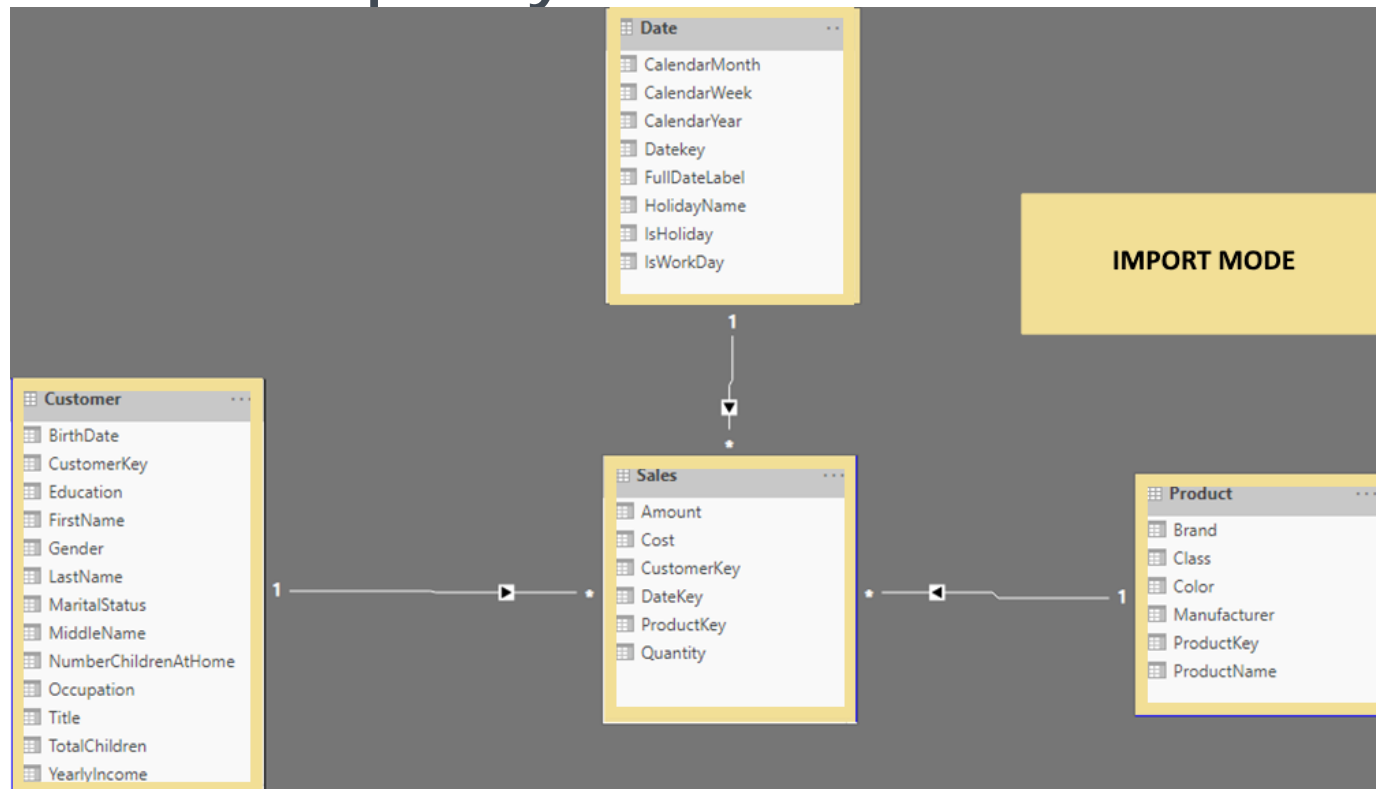
**PROCESS** = Lettura dalla sorgente dati in memoria

Vertipaq contiene il risultato del processing del database



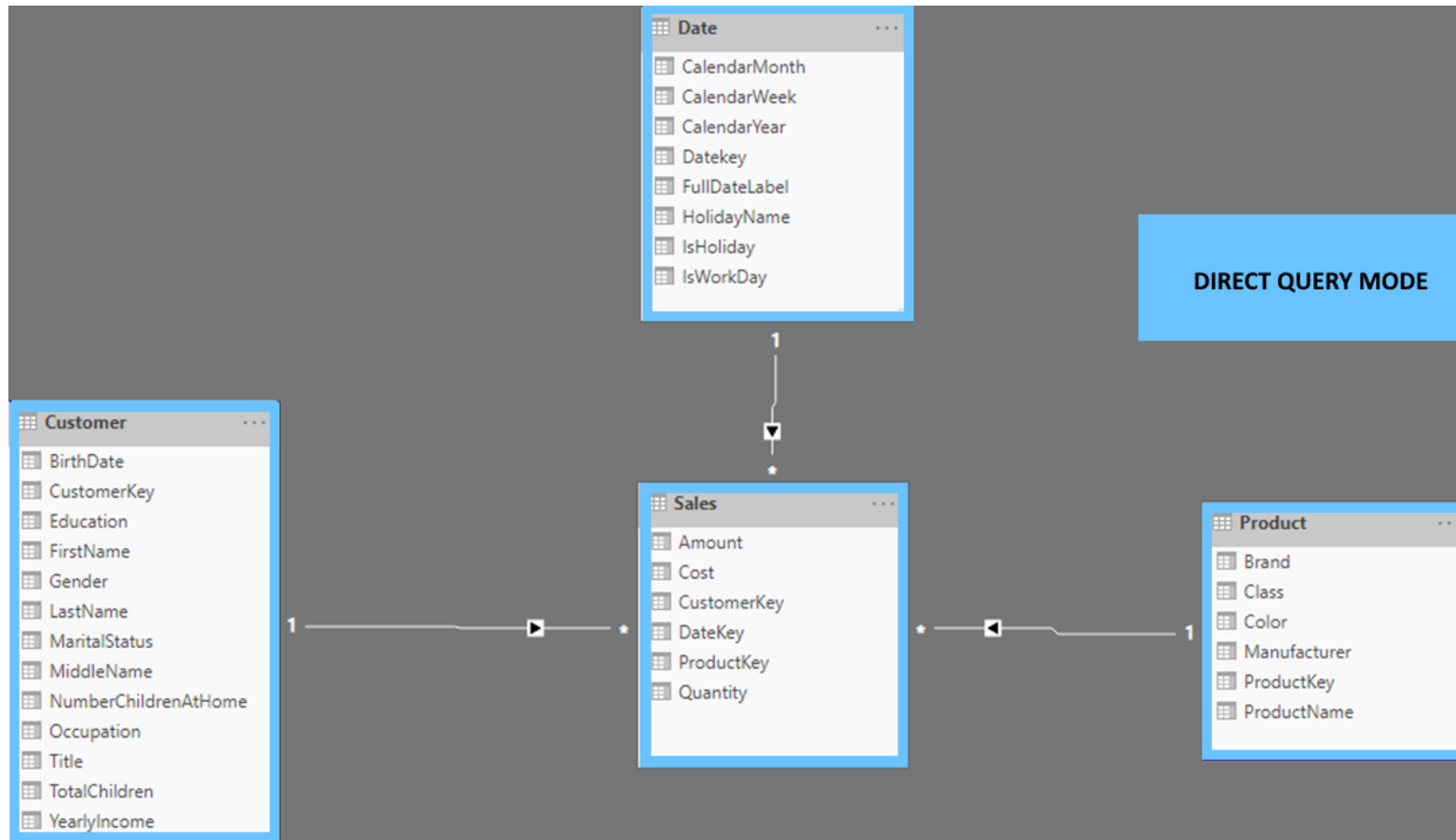
# Import Mode

- Vertipaq: motore colonnare in memoria
- Tutte le query sono elaborate in memoria



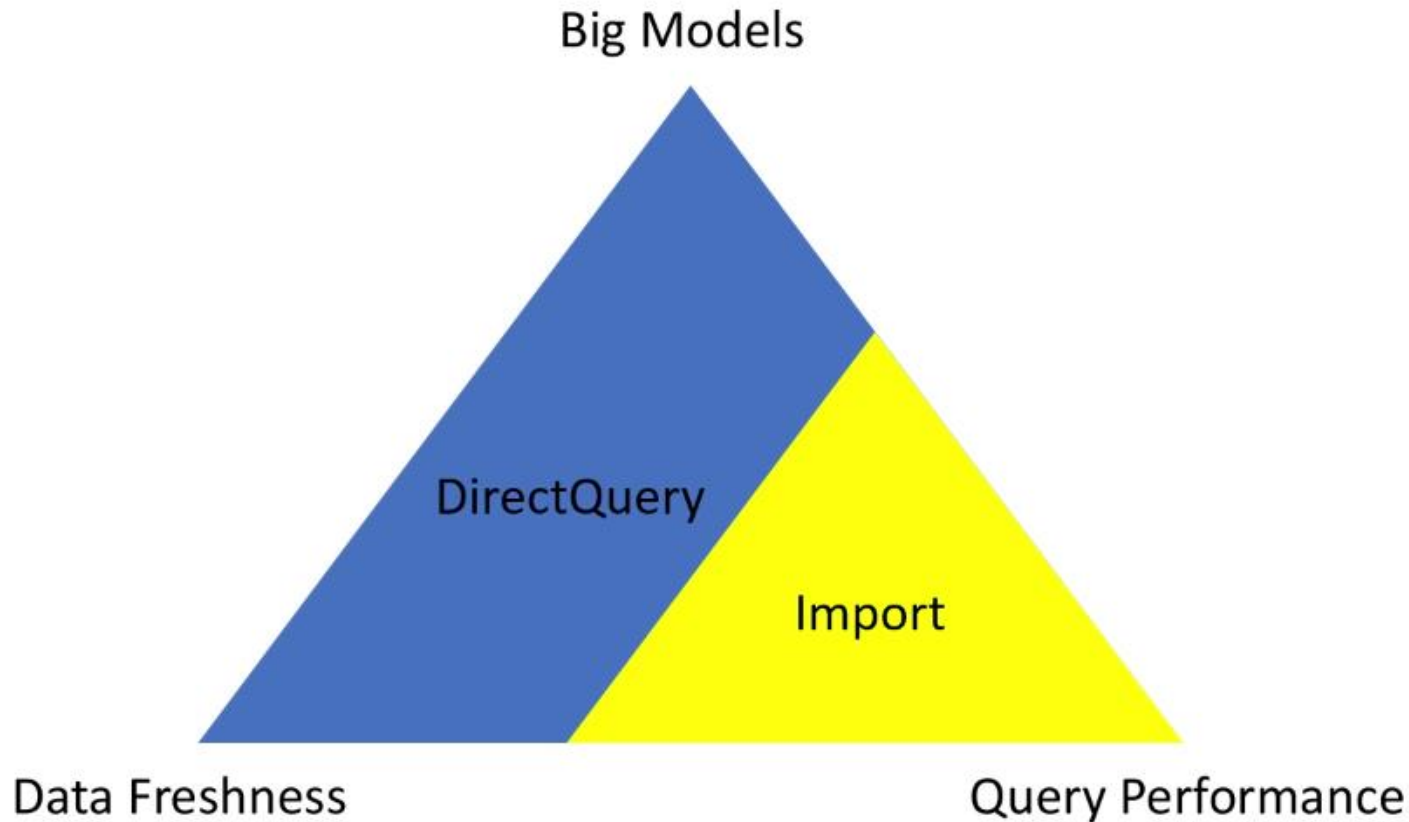
# Direct Query Mode

- Tutte le query sono elaborate dal data source



# Import Mode vs Direct Query

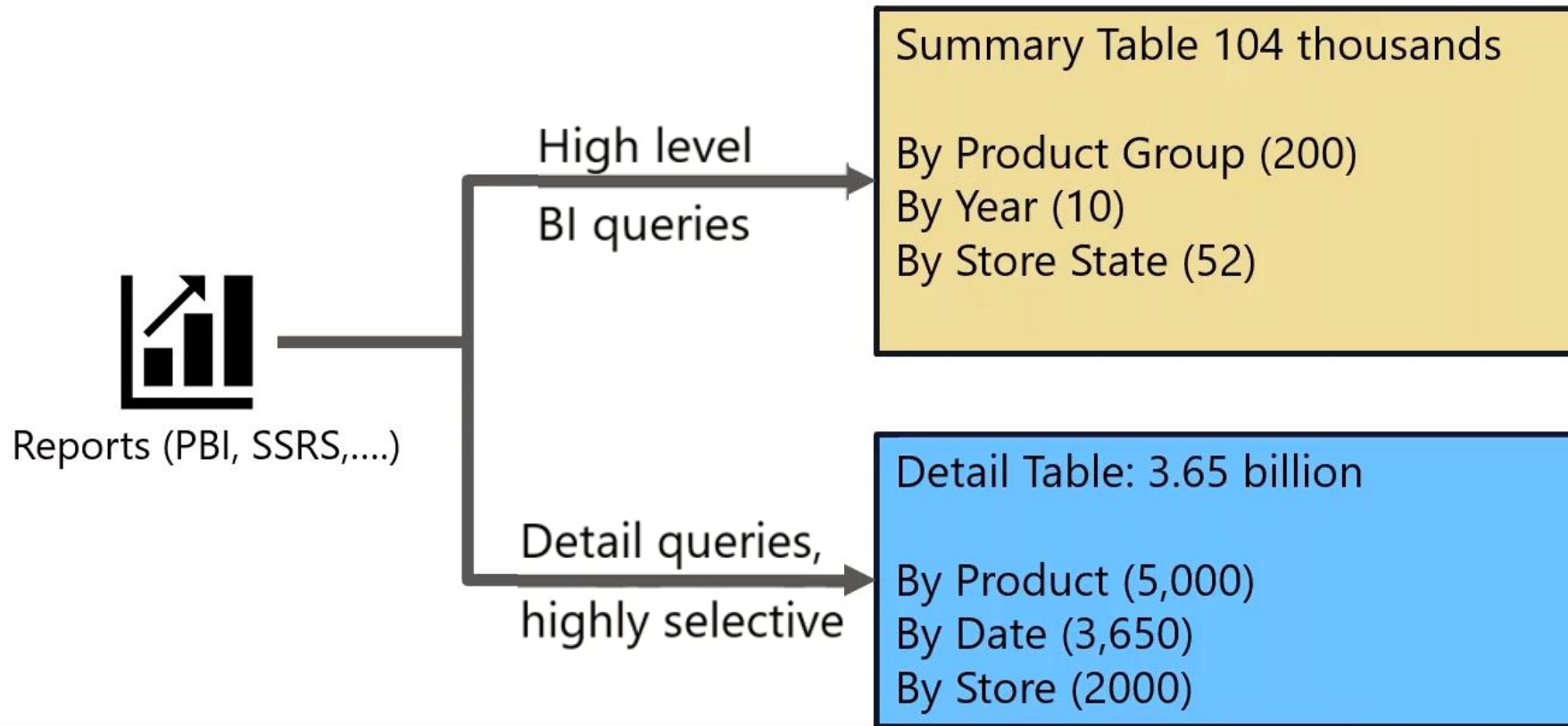
Quale dei due engine scegliamo?



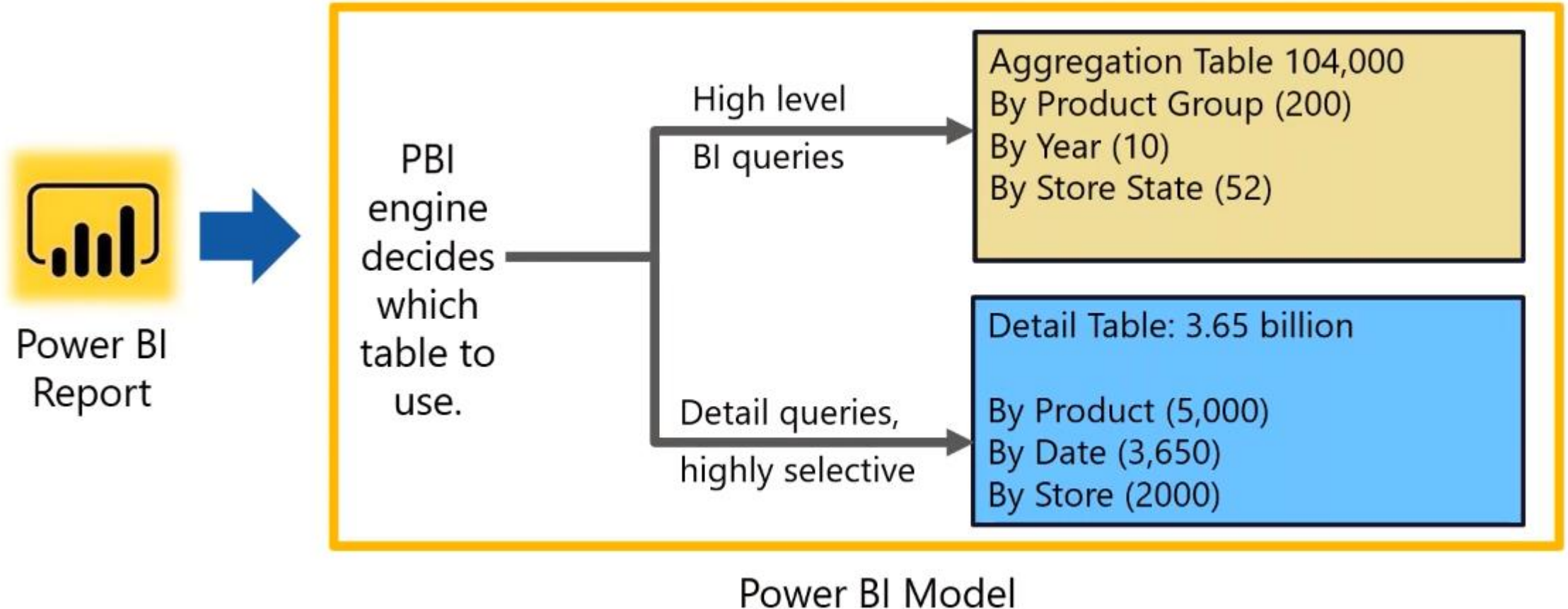


# Concetto di Aggregazione

Una tecnica di ottimizzazione: Tabella aggregata



# Aggregazione in Power BI



# Aggregazioni in Power BI

- Come ogni altra tabella
- Sono automaticamente nascoste
- Sono accessibili dall'amministratore del dataset una volta pubblicate sul servizio.
- Supportato in PBI desktop e services ( shared e premium) con RLS
- Bisogna configurarle adeguatamente

# Aggregazioni in Power BI

- Migliora le performance delle query
- Non solo per petabyte o trillioni di righe
- Per tutti i dataset che richiedono uno spreco di cache (Time, CPU, Memory)

# Come si crea una aggregazione

## Configuratore delle proprietà dell'aggregazione

### Manage aggregations

Aggregations accelerate query performance to unlock big-data sets. [Learn more](#)

Aggregation table

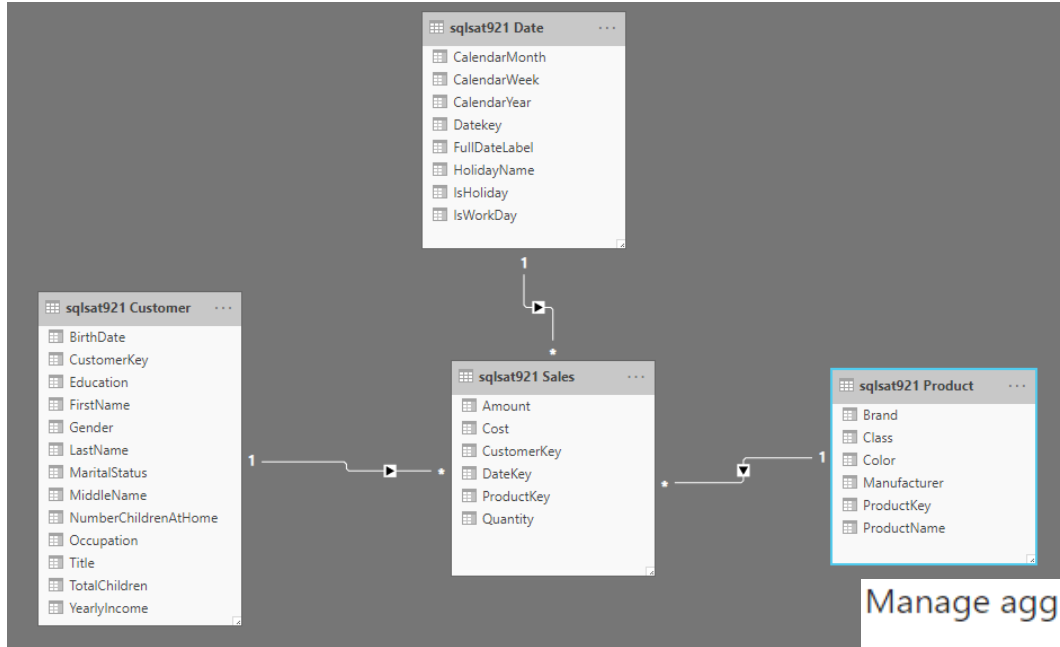
Precedence ⓘ

AggTableSalesAmountByCalendarYearAndProduct ▼

0

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

# Demo 1: Creazione di una aggregazione



Pronti con le mani sulla tastiera del telefonino?



[www.menti.com](https://www.menti.com)

codice **31 92 96**

## Manage aggregations

Aggregations accelerate query performance to unlock big-data sets. [Learn more](#)

Aggregation table

sqlsat921 SalesByYear ▼

Precedence ⓘ

0

AGGREGATION COLUMN

SUMMARIZATION

DETAIL TABLE

DETAIL COLUMN

CalendarYear

GroupBy ▼

sqlsat921 Date ▼

CalendarYear ▼



SumOfQuantity

Sum ▼

sqlsat921 Sales ▼

Quantity ▼



This table will be hidden if aggregations are set because aggregation tables must be hidden.



# Architettura ibrida

## Architettura bilanciata: suddividere il carico

### Big Data (IOT)

#### Detail Table

By Product (5,000)  
By Date (3,650)  
By Store (2000)

3.65 billions

Detail queries

### Data Warehouse (CCI)

#### Aggregation Table 1

By Product Group (200)  
By Date (3,650)  
By City (520)

379.6 millions

Some BI queries

### PBI cache

#### Aggregation Table 2

By Product Group (200)  
By Year (10)  
By Store State (52)

104 thousands

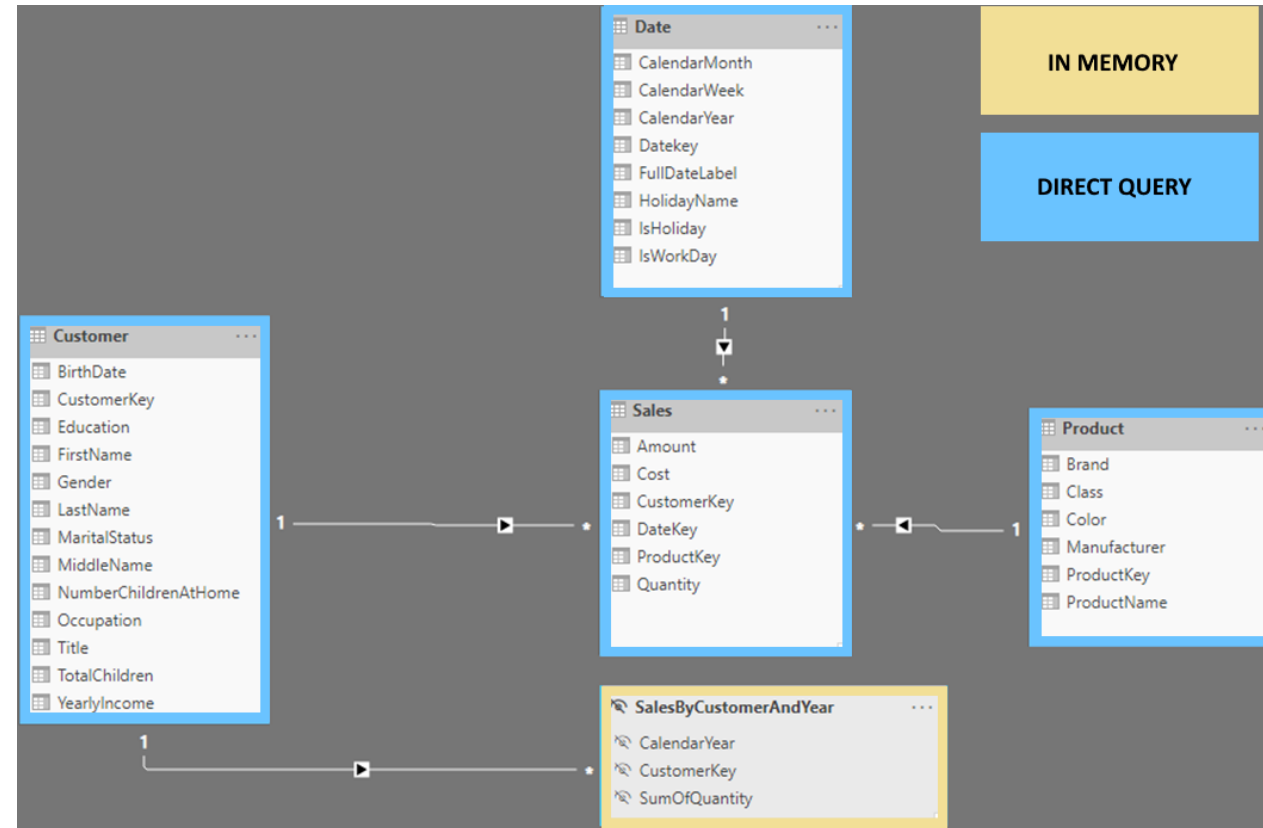
Executive Dashboard

# Composite Model

- Possiamo scegliere la memorizzazione per ogni tabella
- Lasciamo le grandi tabelle nella sorgente dato e carichiamo in memoria le tabelle piccole
- Le query sono elaborate o in memoria o nella sorgente dato

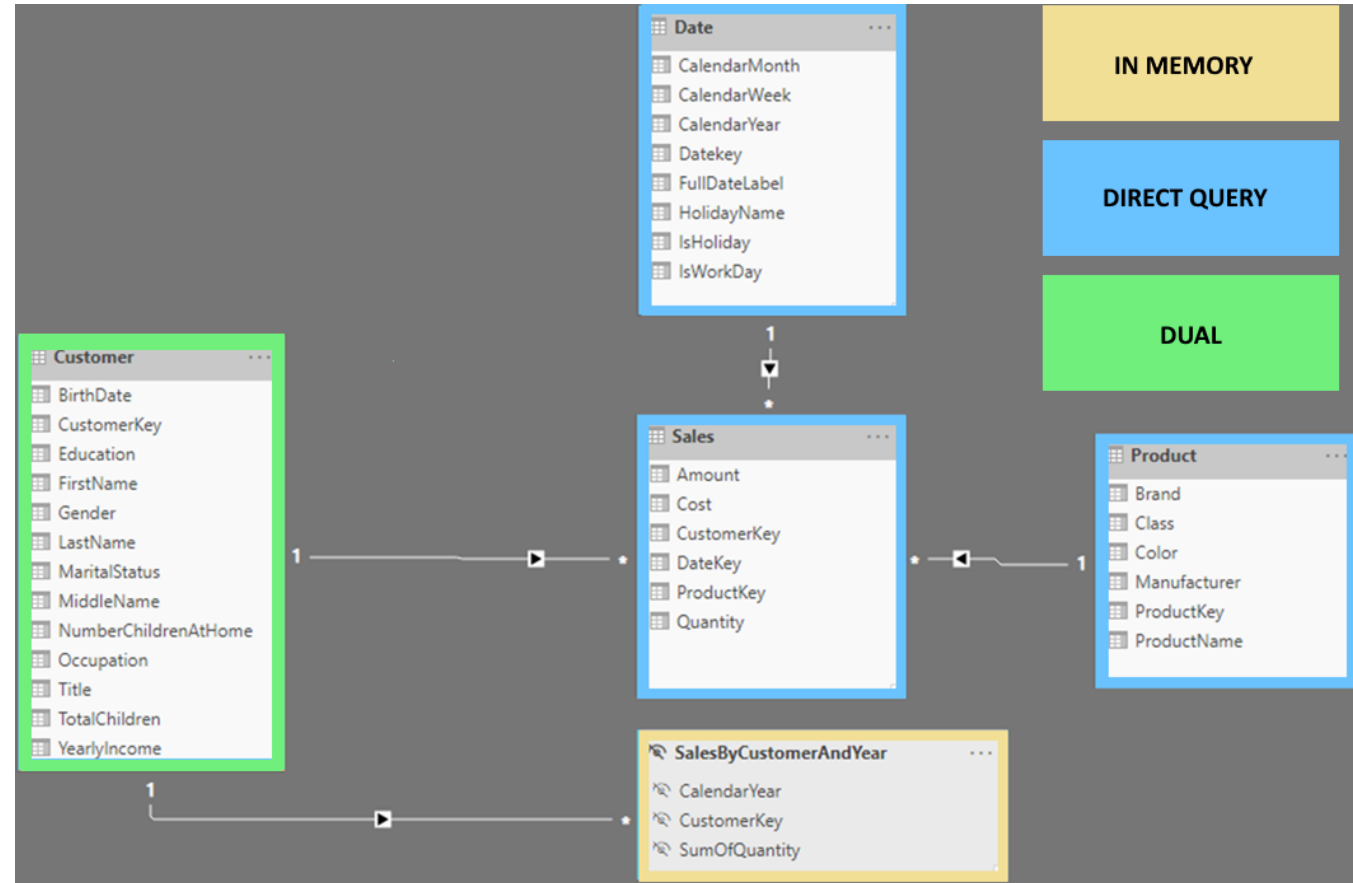
# Composite model e tabelle aggregate

- Le query sono elaborate o in memoria o nella sorgente dato.
- Come faccio ad eseguire query su tabelle che stanno su engine separati? (weak relationship)



# Dual storage mode

Trasforma una weak relationship in una **strong relationship**



# Demo 2: Weak Relationship

## Storage mode

Setting the storage mode to Import has the following implications. Please consider them carefully before proceeding.

Setting storage mode to Import is an irreversible operation. You will not be able to switch it back to DirectQuery.

This operation will refresh tables set to Import or Dual, which may take time depending on factors such as data volume.

Weak relationships may be introduced by this change.

The number of weak relationships can be reduced by setting the following tables to Dual.

- Customer

☐ Set affected tables to dual

[Learn more about setting storage mode](#)

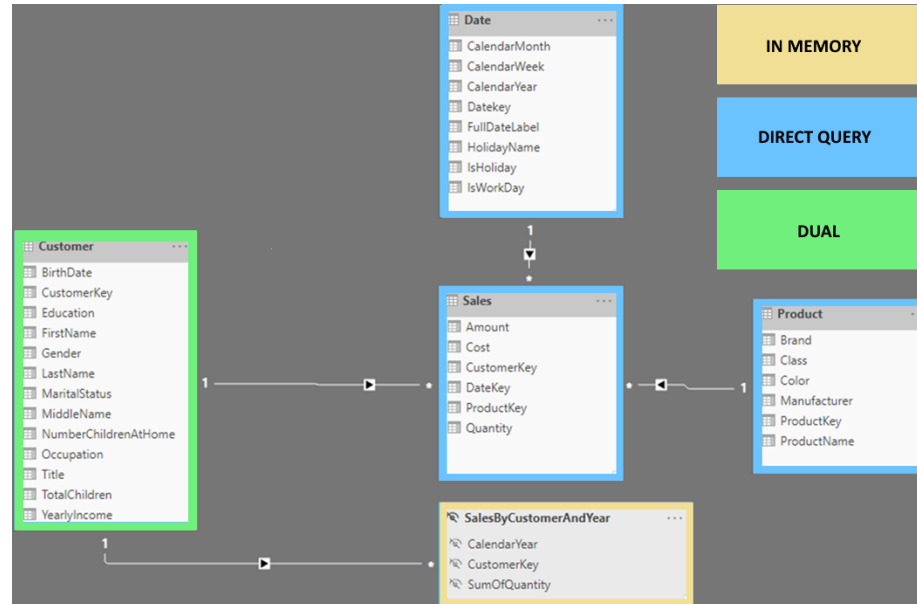
OK Cancel

Pronti con le mani sulla tastiera del telefonino?



[www.menti.com](https://www.menti.com)

codice **31 92 96**

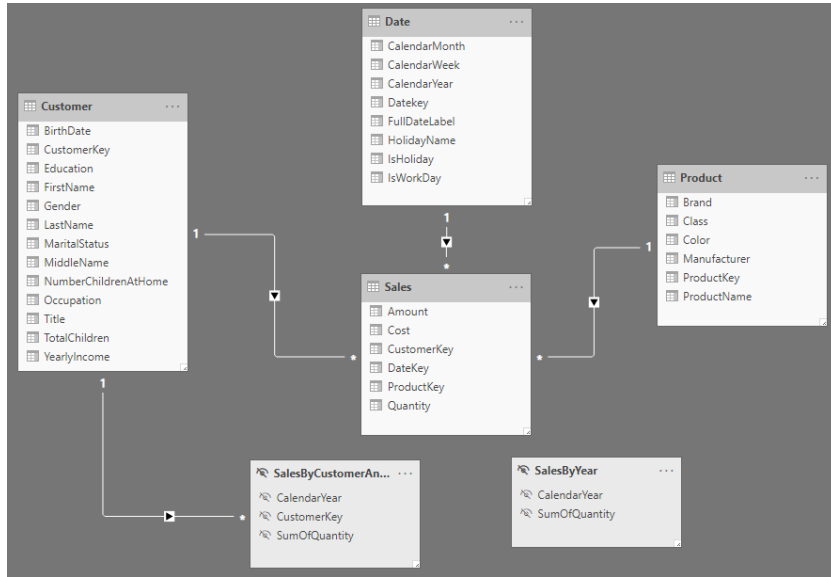


# Precedenza nelle aggregazioni

- Livelli multipli di aggregazione
  - L'aggregazione con il valore di **precedence** più alto vince
  - L'ultima tabella di aggregazione aggiunta al modello sembra vincere
- 
- Tabella principale: 1 billione di righe
  - Aggregazione 1: 1 milione di righe
  - Aggregazione 2: 100000 righe
  - Aggregazione 3: 100 righe



# Demo 3: Precedenze



```
EVALUATE  
SUMMARIZECOLUMNS (  
    'Date'[CalendarYear],  
    "qty", SUM ( Sales[Quantity] )  
)
```

Pronti con le mani sulla tastiera del telefonino?



[www.menti.com](https://www.menti.com)

codice **31 92 96**

OR

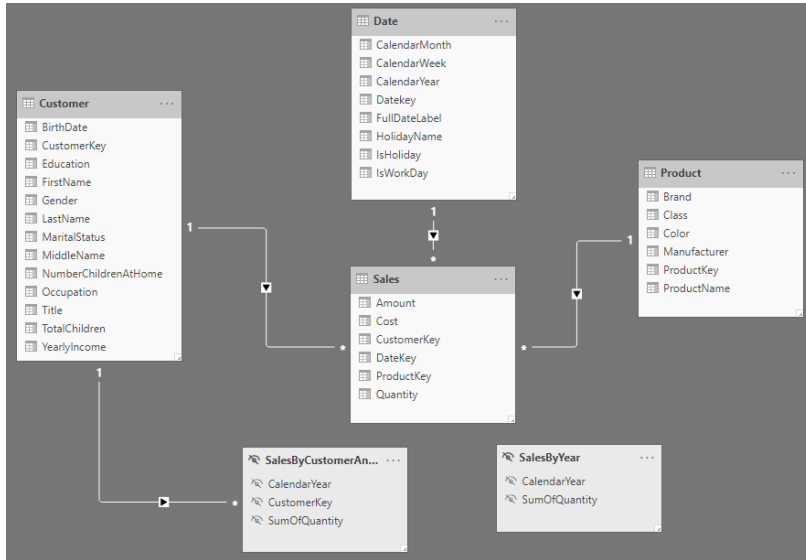
CalendarYear
CustomerKey
SumOfQuantity

CalendarYear
SumOfQuantity

# Mantenimento delle Tabelle aggregate

- Se una tabella importata o direct non è sincronizzata con le tabelle aggregate si dice **(out of sync)**
- Quando una tabella di aggregazione è **out of sync** allora la tabella dei dettagli e delle aggregazione ritornano numeri differenti. **PBI non è in grado di determinare e redirezionare le query al sistema sorgente**
- E' responsabilità di chi gestisce il sistema conoscere il pattern del caricamento dati e mantenere le aggregazione aggiornate

# Demo 4: Mantenimento del sync



CalendarYear	Quantity
2007	3965130
2008	3593302
2009	4969010
<b>Total</b>	<b>12527442</b>

IN MEMORY

CalendarYear	Azure	Black	Blue	Brown	Gold	Green	Grey	Orange	Pink	Purple	Red	Silver
2007	959381	234958	76299	43050	75984	390760	63747	84003	2491	161997	986479	1990
2008	890890	217823	93548	47508	81171	323297	61390	107170	4100	216893	663222	1867
2009	218340	334571	99765	63154	86059	388373	72091	216078	8384	302354	944862	1675
<b>Total</b>	<b>068611</b>	<b>787352</b>	<b>269612</b>	<b>153712</b>	<b>243214</b>	<b>1102430</b>	<b>200328</b>	<b>407251</b>	<b>14975</b>	<b>681244</b>	<b>2594563</b>	<b>5534</b>

DIRECT QUERY



CalendarYear	Quantity
2007	3965130
2008	3593302
2009	4969010
<b>Total</b>	<b>12527442</b>

IN MEMORY

CalendarYear	Azure	Black	Blue	Brown	Gold	Green	Grey	Orange	Pink	Purple	Red	Silver
2008	15452	890890	217823	93548	47508	81171	323297	64490	107170	4100	216893	663222
2009	18025	1218340	334571	99765	63154	86059	388373	72091	216078	8384	302354	944862
<b>Total</b>	<b>33477</b>	<b>2109230</b>	<b>552394</b>	<b>193313</b>	<b>110662</b>	<b>167230</b>	<b>711670</b>	<b>136581</b>	<b>323248</b>	<b>12484</b>	<b>519247</b>	<b>1608084</b>

DIRECT QUERY

# Le aggregazioni e la RLS

- Le espressioni della row level security (RLS) filtreranno sia le tabelle di aggregazione che quelle di dettaglio

RLS applicate alla tabella di dettaglio	RLS applicate alla tabella aggregata	Hit della aggregazione
Si	Si	Si
Si	No	No
No	Yes (fallisce , PBI si stoppa)	N/A

- Non Possiamo interrogare le tabelle aggregate (anche senza RLS) senza essere amministratori del dataset

# Recap

- Le aggregazioni sono un tool eccellente per le ottimizzazioni delle performance
- Le aggregazione non sono solo per dataset delle dimensioni di peta byte
- Le aggregazione lavorano con RLS
- Le aggregazione hanno un costo di mantenimento quindi essere sicuri prima di usarle
- Richiedono una attenta configurazione

# Recap

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

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





Ricordatevi di  
compilate il feedback  
form 😊

<https://speakerscore.com/MDN9>

Thank you

#SqlSat921





PASS