

# **Organizzare e preparare i dati per la Business Intelligence con Microsoft Power BI**

*Alperia*

Bolzano, 04 febbraio 2025

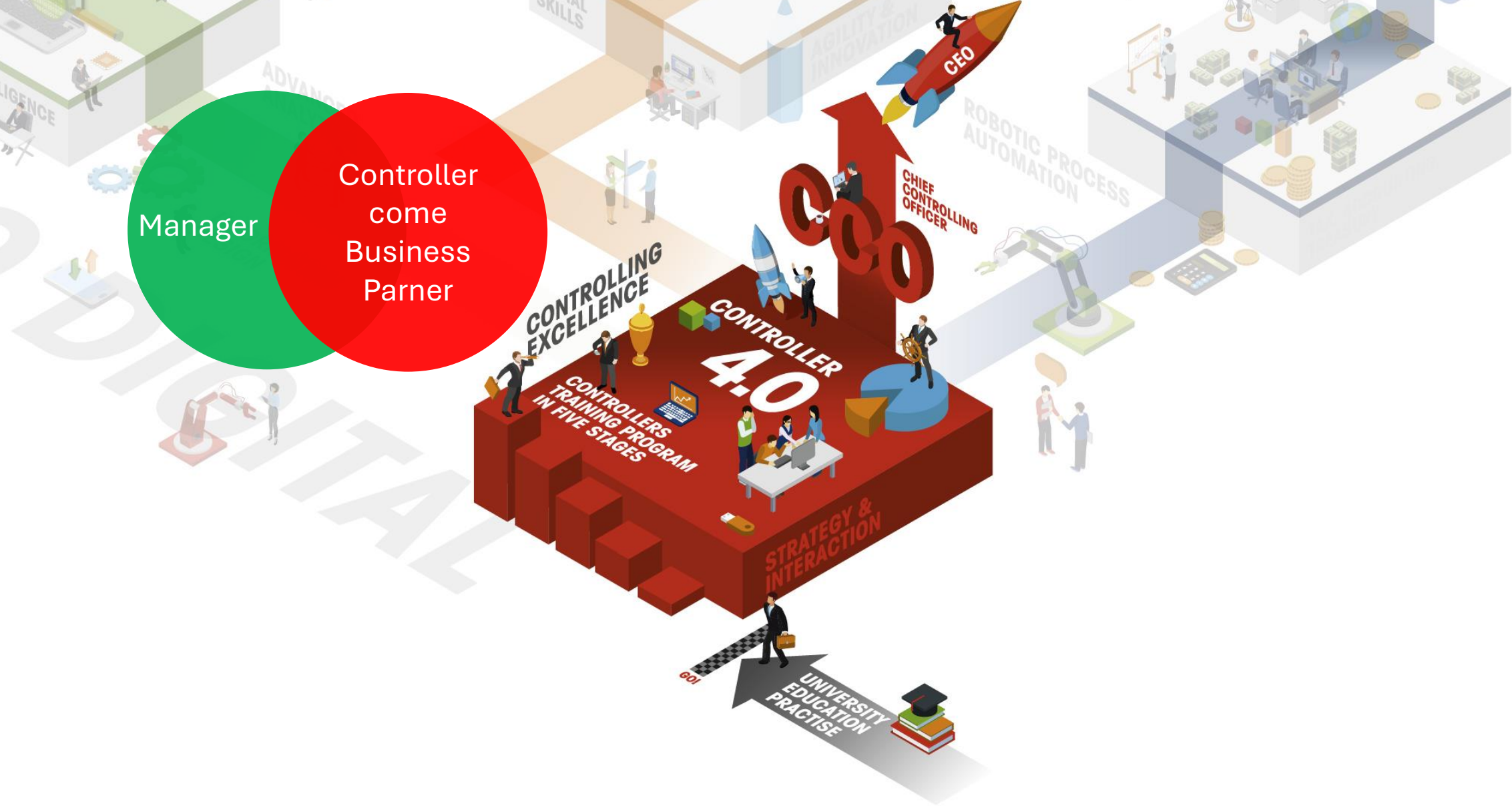
## Organizzazione dei dati (DataWareHousing) (circa 10%)

- ✓ Fatti vs Dimensioni
- ✓ Definire e creare chiavi per le relazioni
- ✓ Tipi di chiavi: primarie ed esterne, surrogate e naturali
- ✓ Granularità e tabelle bridge
- ✓ Relazioni 1 a molti, molti a molti ed uno a uno
- ✓ Schemi a stella e snowflake
- ✓ Dimensioni con ruoli multipli
- ✓ Slowly Changing Dimension

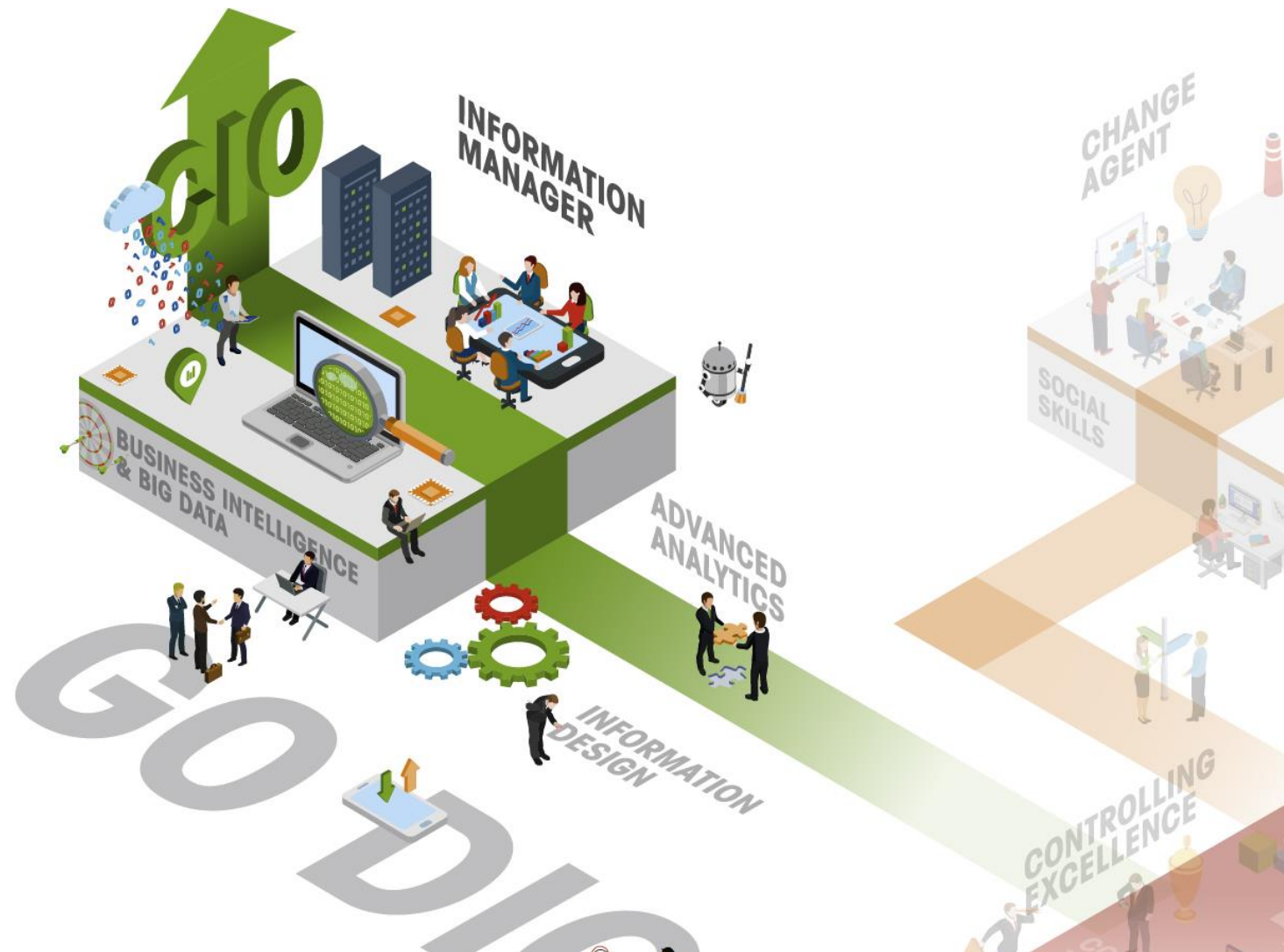
## Preparazione dei dati (ETL) (circa 15%)

- ✓ Identificare e connettersi a origini dati
- ✓ Modificare le impostazioni dell'origine dati, incluse le credenziali
- ✓ DirectQuery vs Import
- ✓ Parametri
- ✓ Risolvere i problemi di qualità dei dati (incoerenze, valori imprevisti o nulli)
- ✓ Errori di importazione dei dati
- ✓ Selezionare il corretto tipo di dato per le colonne dei dati importati
- ✓ Creare e trasformare colonne
- ✓ Raggruppare e aggregare righe
- ✓ Trasformare tramite Pivot o Unpivot
- ✓ Gestire i dati semistrutturati
- ✓ Riferimento o Duplicazione Query
- ✓ Merge e Append di Query
- ✓ Anatomia del linguaggio "M"





Artificial Intelligence  
Predictive and  
Advanced Analytics  
Information Design  
BI & Big Data Strategies  
Deep Dive Advanced  
Analytics  
...





# Perché ci serve un «magazzino dei dati»?

Come si utilizzano i dati in azienda

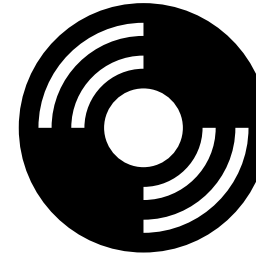
... osservare come gira la ruota  
... prendere le migliori decisioni per far girare la ruota più velocemente

- ✓ Migliore categoria di prodotti?
- ✓ Vendite rispetto al mese scorso?
- ✓ Come possiamo migliorare?



**Prendere  
decisioni  
basate sui dati**

**OLAP=Online Analytical Processing**



**Registrazione di  
dati operativi**

**OLTP =Online Transactional Processing**

... far girare la ruota

- ✓ Ricevere ed evadere ordini
- ✓ Ricevere e reagire a reclami
- ✓ Acquistare per riempire il magazzino

La mancanza di un «magazzino di dati» è accompagnata da affermazioni del tipo (Kimball 1996):

“Abbiamo un sacco di dati ma non li usiamo“

“I nostri dati sono molto difficili da analizzare“ – user friendly

“I dati sono su differenti sistemi ed è difficile accedervi“ - ETL

“Voglio solo vedere ciò che è importante!“

“Voglio accedere ai dati in maniera semplice e veloce“

“Voglio prendere decisioni basate sui dati!“

# Perché ci serve un «magazzino dei dati»?

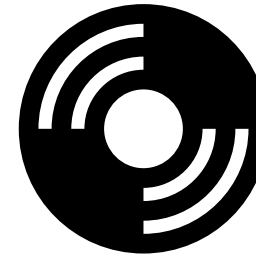
Diversi requisiti nei processi transazionali ed analitici

- ✓ Si analizzano migliaia o milioni di dati insieme
- ✓ Ci vuole performance nelle interrogazioni
- ✓ Interessa la «storia»
- ✓ Deve essere utilizzabile da tutti



**Prendere  
decisioni  
basate sui dati**

**OLAP=Online Analytical Processing**



**Registrazione di  
dati operativi**

**OLTP =Online Transactional Processing**

- ✓ Si elabora un dato alla volta
- ✓ Si inseriscono, modificano, cancellano dati
- ✓ Non interessa la «storia»

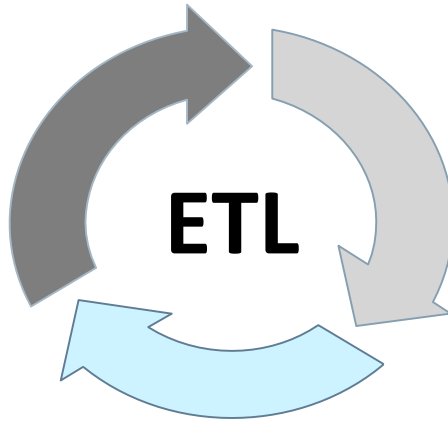


Un archivio di dati che viene utilizzato ed ottimizzato per scopi di reporting ed analisi dei dati. Con le seguenti caratteristiche:

1. Molto facile da usare
2. Interrogazioni molto veloci
3. A supporto delle analisi di business



Registrazione  
di dati  
operativi



Prendere  
decisioni basate  
sui dati

Luogo centralizzato di archiviazione:

- ✓ Coerente con il business
- ✓ Veloce nella risposta
- ✓ Facile da usare
- ✓ Costantemente alimentato
- ✓ Base per la Business Intelligence

## Data Lake o DataWareHouse?

	Data Lake	DataWareHouse
Dati	Dati grezzi	Dati processati
Tecnologie	Big Data	Database
Struttura	Dati non strutturati	Dati strutturati
Casi d' Uso	Non definito a priori	Specifico e definito
Utenti	Competenze maggiori	Competenze di business

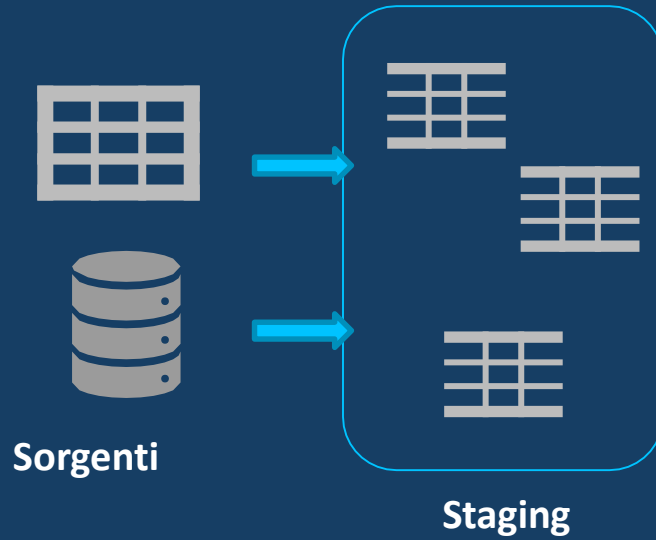
Sono quindi molto diversi e possiamo averli entrambi, possiamo costruire un DWH su alcune parti del DL

L'insieme di strategie, procedure, tecnologie e infrastrutture che servono ad ottenere insight significativi dai dati.

A large, dark blue arrow pointing to the right, containing white text.

**Pipeline dai dati alla  
Business Intelligence**

# Livelli di un magazzino dei dati



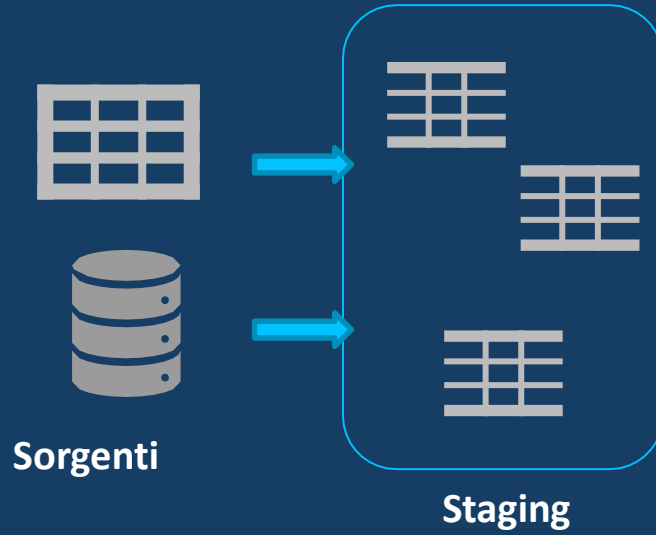
## Dipartimento 1

employee_id	entry_date	position_level
1	1/2/2022	HR
2	1/6/2022	IT
3	1/4/2022	IT
4	1/7/2022	UM
5	1/6/2022	PM

## Dipartimento 2

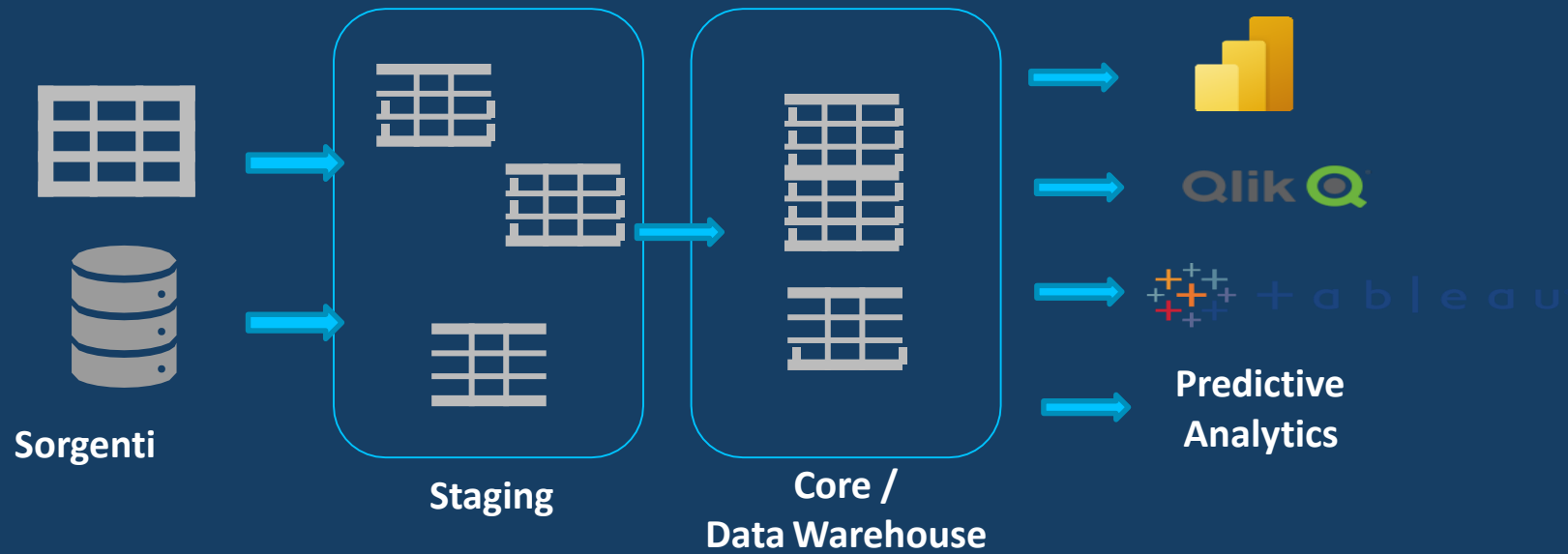
employee_id	entry_date	position_level
6	1/5/2022	HR
7	1/6/2022	HR
8	1/8/2022	PM
9	1/7/2022	UM
10	1/6/2022	PM

# Livelli di un magazzino dei dati

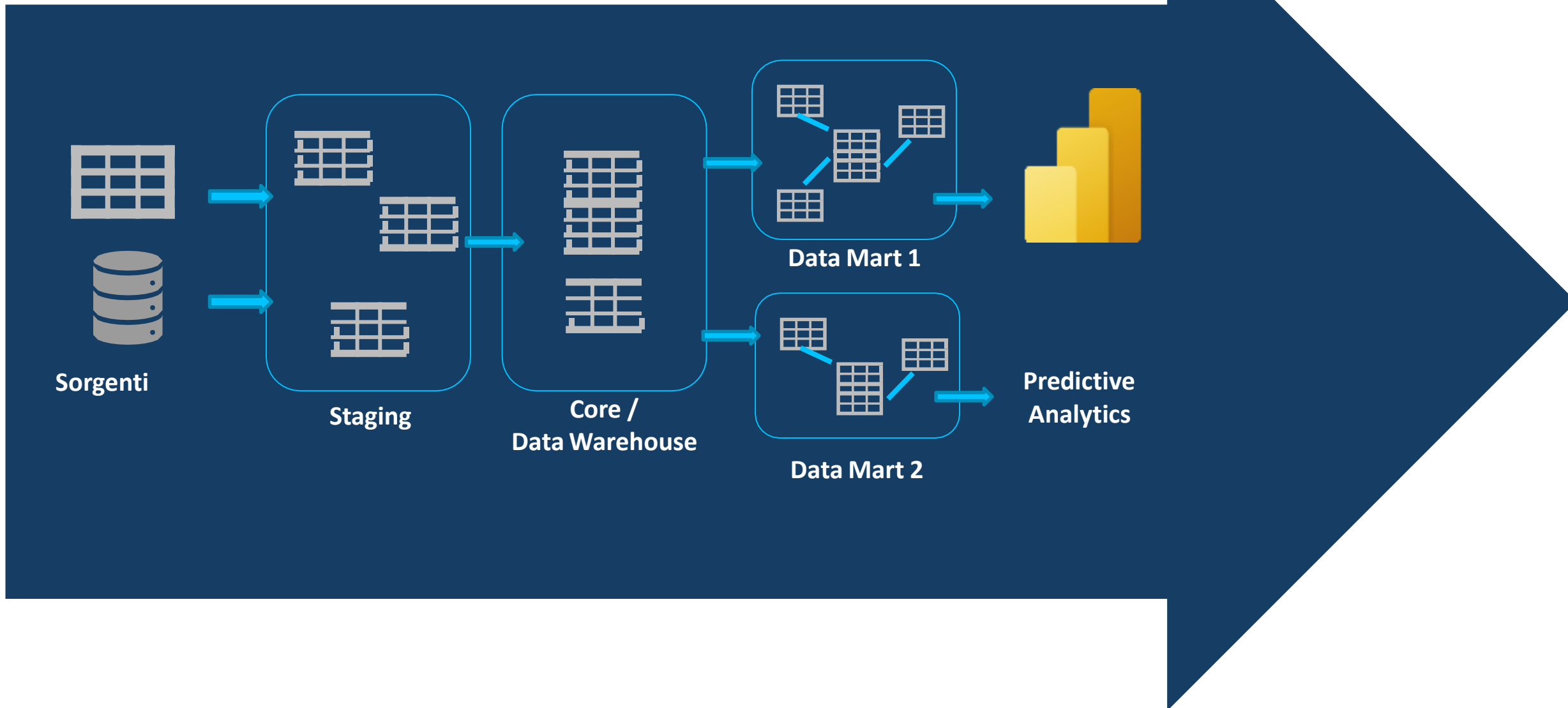


**Dipartimenti**

employee_id	entry_date	position_level
1	1/2/2022	HR
2	1/6/2022	IT
3	1/4/2022	IT
4	1/7/2022	UM
5	1/6/2022	PM
6	1/5/2022	HR
7	1/6/2022	HR
8	1/8/2022	PM
9	1/7/2022	UM
10	1/6/2022	PM

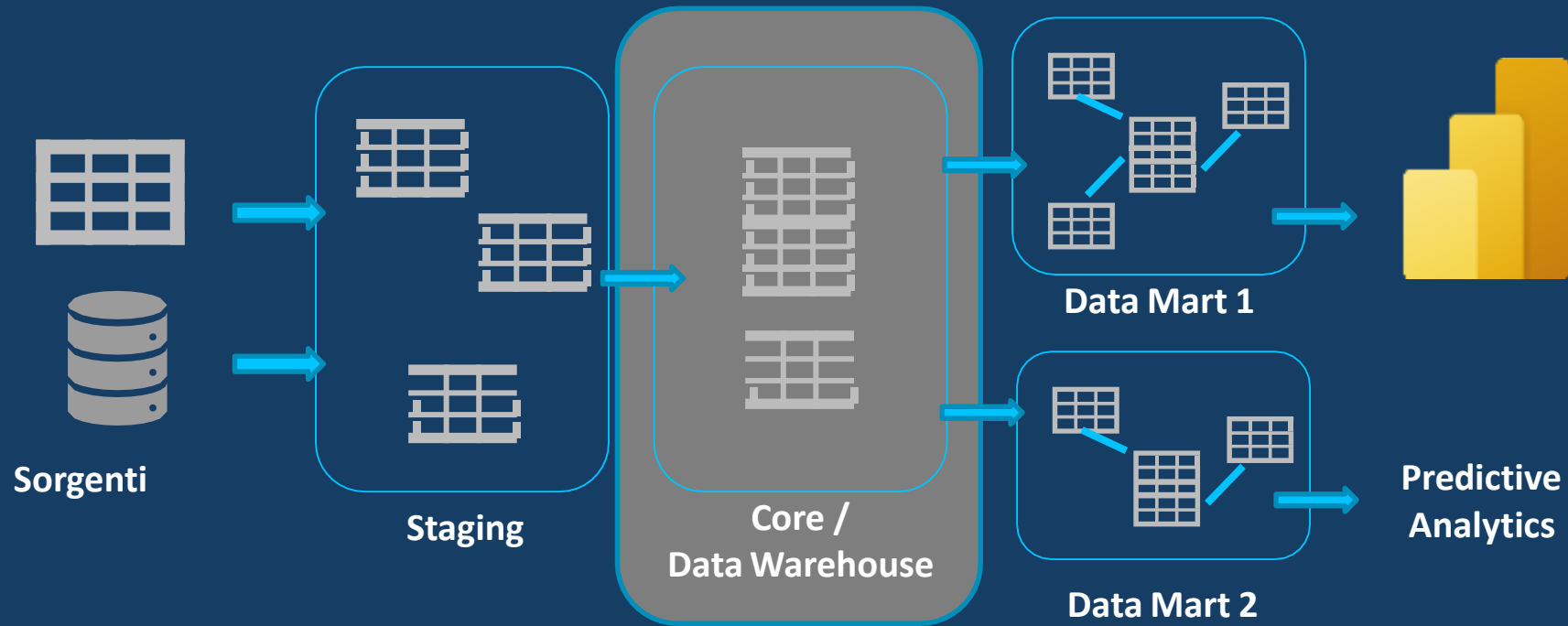




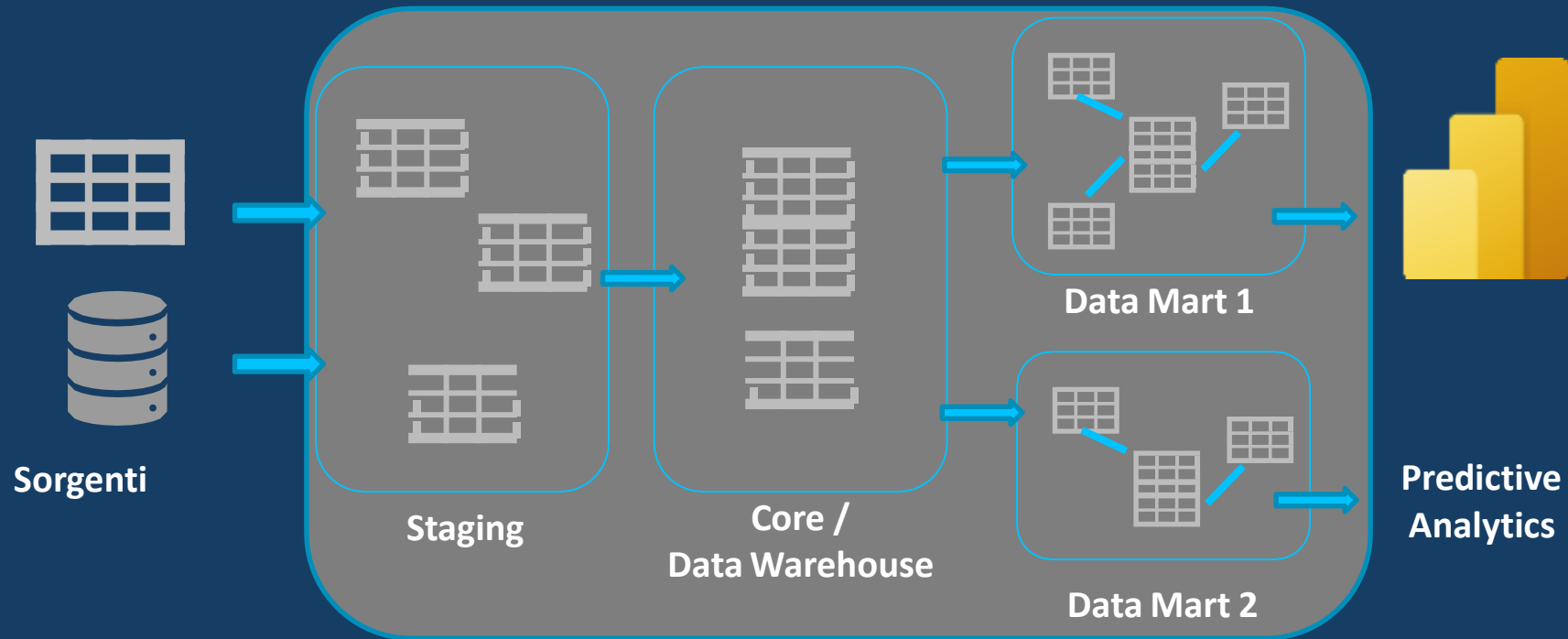


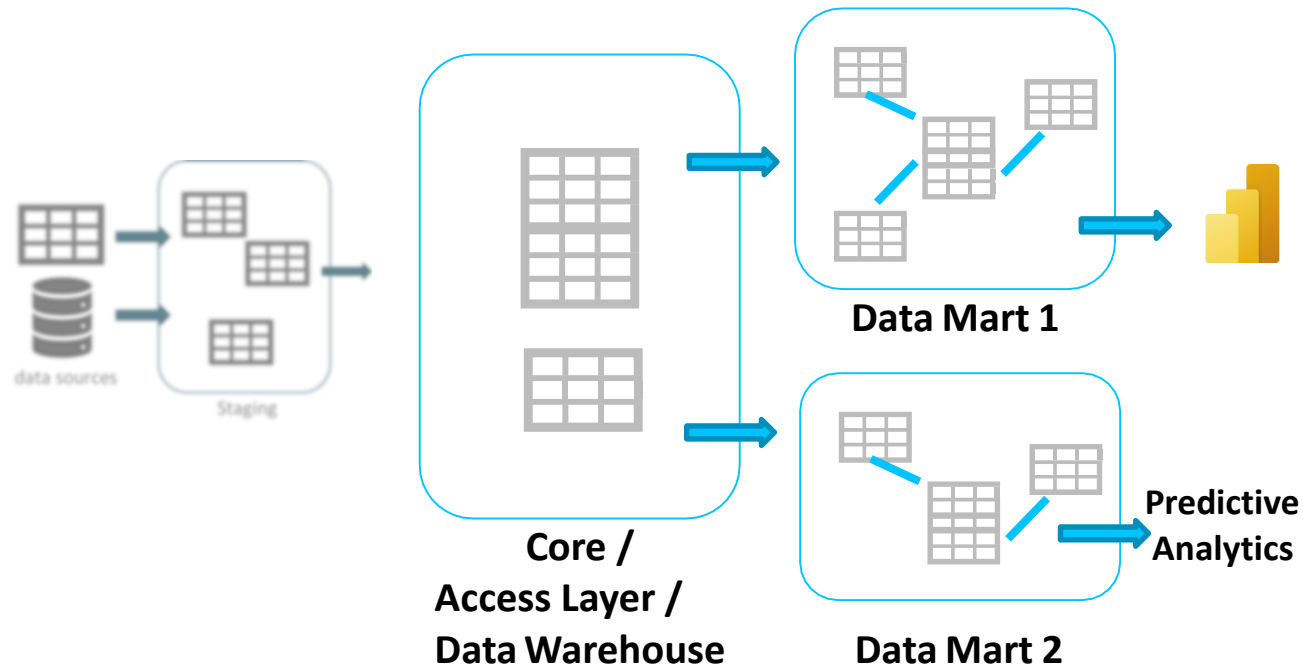
se abbiamo un grande data warehouse composto da moltissime tabelle e abbiamo molti casi d'uso diversi

# Cosa è DataWareHouse?



# Cosa è DataWareHouse?





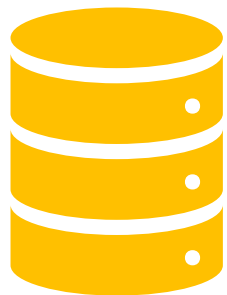
- ✓ Sottinsieme del DataWarehouse
- ✓ Modello Dimensionale
- ✓ Possono essere ulteriormente aggregati

## Vantaggi

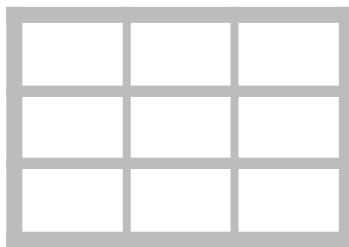
- + usabilità
- + accettazione di un DM rispetto ad un DWH
- + prestazioni

## Casi d'uso

- Utilizzabile con diversi strumenti
- Utilizzabile con diverse tecnologie (ad es. «*in memory*»)
- Plasmabile su ogni processo di business



**Database  
Relazionali**



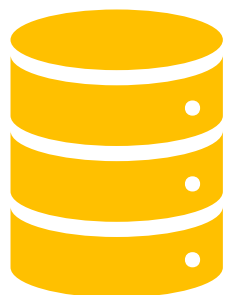
**Tabelle  
(relazioni)**

id	date	product	customer_id
1	1/2/2022	Fulltoss Tanga Tomato	2
2	1/2/2022	Chilli - Green, Organically Grown	2
3	1/2/2022	Masala Powder	5
4	1/2/2022	Cheese Cracker (McVities)	1
5	1/2/2022	Centre Filled Chocolate Cake	5

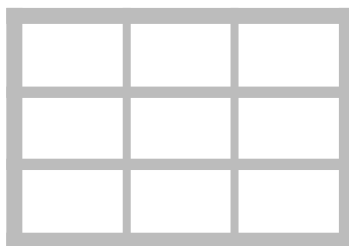
**Righe (record) e  
Colonne (campi)**

```
SELECT    <column1>,  
          <column2>, ...  
FROM      <table_name>
```

**SQL per interrogare (query)**



**Database  
Relazionali**



**Tabelle  
(relazioni)**

**Chiavi  
primarie**

**Chiavi  
esterne**

id	date	product	customer_id
1	1/2/2022	Fulltoss Tangy Tomato	2
2	1/2/2022	Chilli - Green, Organically Growr	2
3	1/2/2022	Masala Powder	5
4	1/2/2022	Cheese Cracker (Mcvities)	1
5	1/2/2022	Centre Filled Chocolate Cake	5

id	name	city
1	Frank	New York
2	Sarah	Chicago
3	Sabrina	New Orleans
4	Maya	Los Angeles
5	Marc	Delas

Chiavi  
primarie

Chiavi  
esterne

id	date	product	customer_id	name
1	/2/2022	Fulltoss Tanga Tomato	2	Sarah
2	/2/2022	Chilli - Green, Organically Grown	2	Sarah
3	/2/2022	Masala Powder	5	Marc
4	/2/2022	Cheese Cracker (McVities)	1	Frank
5	/2/2022	Centre Filled Chocolate Cake	5	Marc

id	name	city
1	Frank	New York
2	Sarah	Chicago
3	Sabrina	New Orleans
4	Maya	Los Angeles
5	Marc	Delas

```
SELECT      s.id,  
            s.product,  
            s.customer_id,  
            c.name  
FROM        Sales s  
  
LEFT JOIN   customer c  
ON s.customer_id = c.id
```



Chiavi  
primarie

Chiavi  
esterne

id	date	product	customer_id	name
1	/2/2022	Fulltoss Tanga Tomato	2	Sarah
2	/2/2022	Chilli - Green, Organically Grown	2	Sarah
3	/2/2022	Masala Powder	5	Marc
4	/2/2022	Cheese Cracker (Mcvities)	1	Frank
5	/2/2022	Centre Filled Chocolate Cake	5	Marc

id	name	city
1	Frank	New York
2	Sarah	Chicago
3	Sabrina	New Orleans
4	Maya	Los Angeles
5	Marc	Delas

- ✓ Logica, algoritmi e performance dagli anni 70 agli anni 90
- ✓ Performance sempre più elevate
- ✓ Utilizzabile negli OLAP, relazione tra più tabelle e contesto

Chiavi  
primarie

Chiavi  
esterne

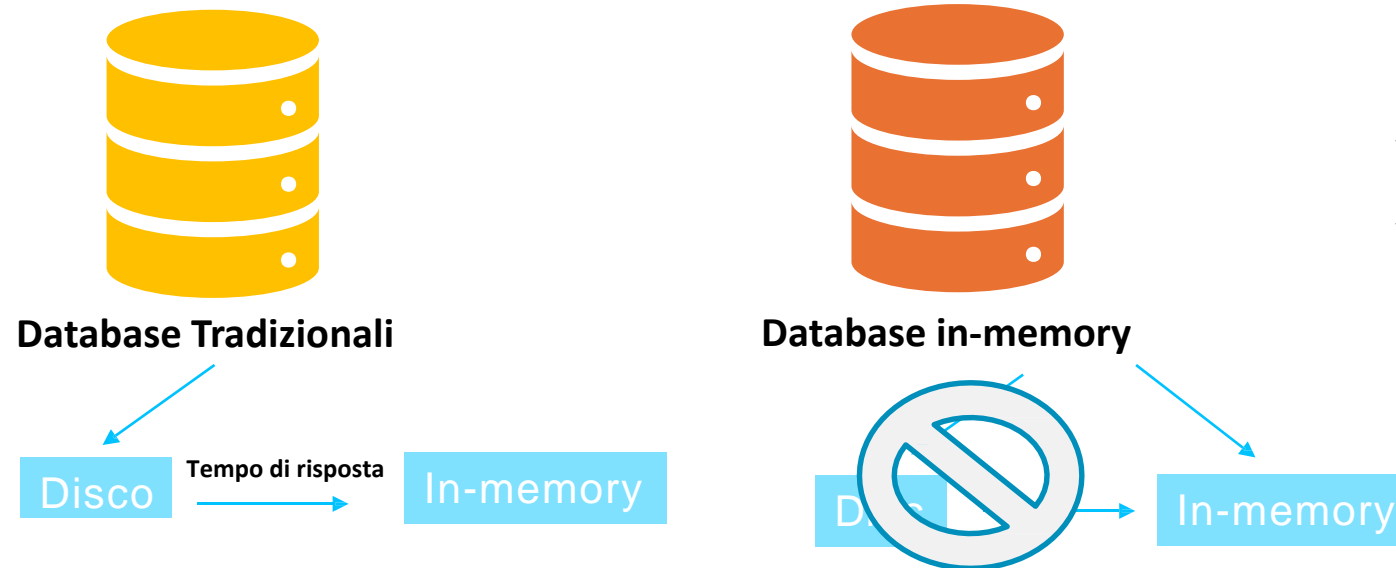
id	date	product	customer_id	name
1	/2/2022	Fulltoss Tanga Tomato	2	Sarah
2	/2/2022	Chilli - Green, Organically Grown	2	Sarah
3	/2/2022	Masala Powder	5	Marc
4	/2/2022	Cheese Cracker (McVities)	1	Frank
5	/2/2022	Centre Filled Chocolate Cake	5	Marc

id	name	city
1	Frank	New York
2	Sarah	Chicago
3	Sabrina	New Orleans
4	Maya	Los Angeles
5	Marc	Delas

Relational database  
management system (RDMS)

- ✓ Oracle
- ✓ Microsoft SQL Server
- ✓ PostgreSQL
- ✓ MySQL
- ✓ Amazon Relational Database Service (RDS)
- ✓ Azure SQL databases
- ✓ Snowflake

- ✓ Ottimizzati per le prestazioni delle query
- ✓ Ideali per le analytics con un alto volume di query
- ✓ Usati solitamente per i data marts
- ✓ Indipendenti da dati strutturati in modo relazionale o non relazionale



- ✓ Sistema di archiviazione a colonne
- ✓ Piani di query paralleli
- ✓ Altre tecniche...

- *Durability*: quando il database viene scollegato dall'alimentazione o viene resettato, le informazioni vengono completamente perse
- Durability garantita tramite snapshot o immagini

**PowerBI**: motore in-memory colonnare chiamato VertiPaq (noto in passato come xVelocity)

- ✓ I dati sono organizzati in cubi non relazionali

Cubo = Dataset Multimensionale

- ✓ Arrays al posto di tabelle
- ✓ Garantite query veloci
- ✓ Si interrogano con **MDX**, un linguaggio specializzato
- ✓ Tecnologia matura – direi obsoleta

Perché questa tecnologia è diventata meno importante?

- ✓ E' complessa e necessita di competenze molto specialistiche (MDX).
- ✓ Hardware molto performanti a costi ridotti.

- ✓ Metodologia per organizzare i dati
- ✓ Usata solitamente nei processi di DataWareHousing
- ✓ Ottimizzata per la performance (query veloci) e l'usabilità
- ✓ Ideata per facilitare il reporting e l'olap



Metodologia per organizzare i dati

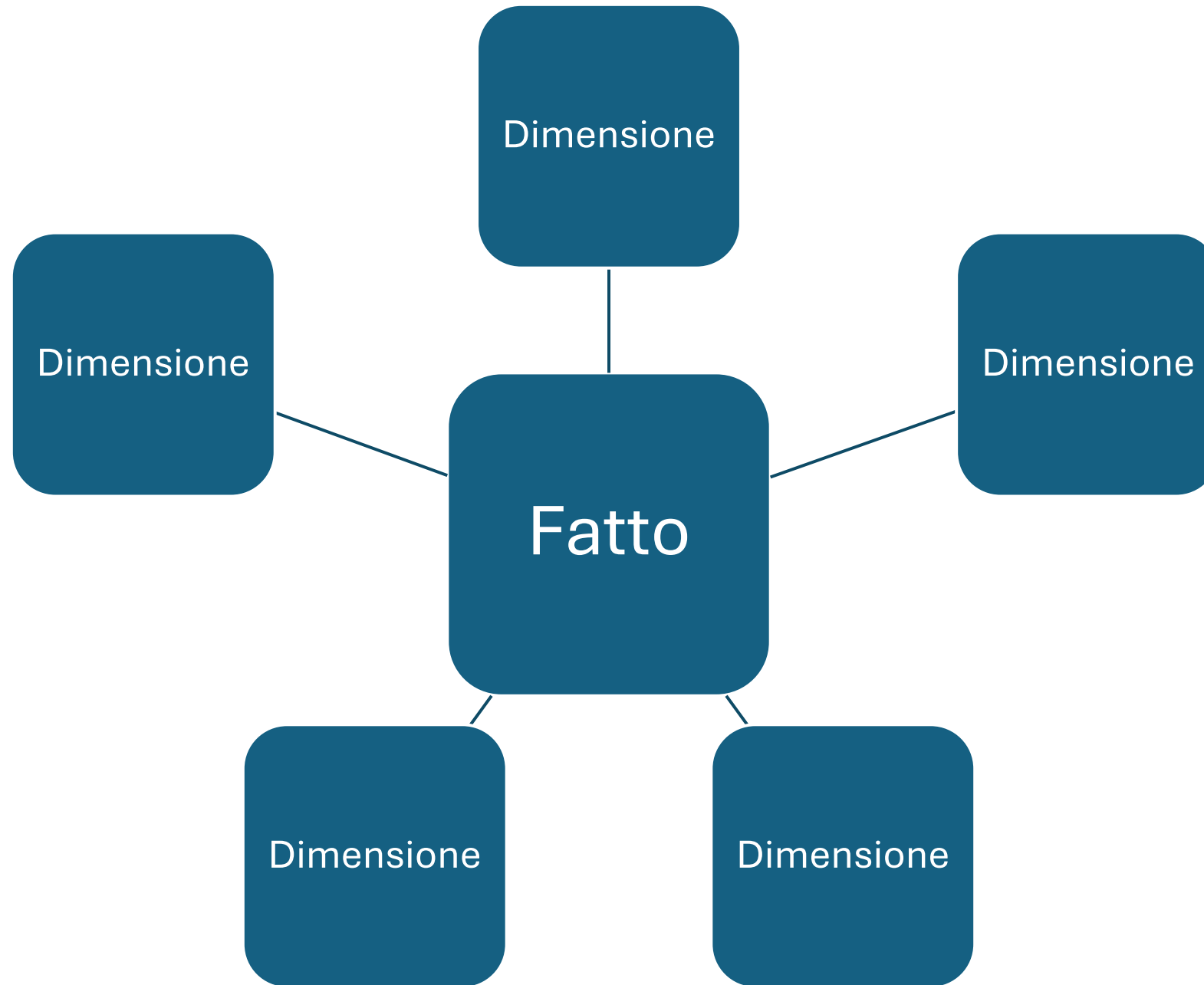
## ✓ Fatti

Contengono qualcosa che viene **solitamente misurato**

## ✓ Dimensioni

Forniscono un ulteriore **contesto** a queste misure (ad es. periodo o categoria)

profitti **per** anno o **per** categoria di prodotto



# Modellazione multidimensionale

Obiettivo: query veloci e usabilità

tabella troppo larga



id	date	product	category	customer_id	name	profit
1	1/2/2022	Fulltoss Tangy Tomato	Vegetables	2	Sarah	\$23
2	1/2/2022	Chilli - Green, Organically Grown	Snacks	2	Sarah	\$12
3	1/2/2022	Masala Powder	Herbs	5	Marc	\$93
4	1/2/2022	Cheese Cracker (Mcvities)	Snacks	1	Frank	\$23
5	1/2/2022	Centre Filled Chocolate Cake	Snacks	5	Marc	\$21

# Modellazione multidimensionale

Obiettivo: query veloci e usabilità

id	date	product	category	customer_id	name	profit
1	1/2/2022	Fulltoss Tangy Tomato	Vegetables	2	Sarah	\$23
2	1/2/2022	Chilli - Green, Organically Grown	Snacks	2	Sarah	\$12
3	1/2/2022	Masala Powder	Herbs	5	Marc	\$93
4	1/2/2022	Cheese Cracker (Mcvities)	Snacks	1	Frank	\$23
5	1/2/2022	Centre Filled Chocolate Cake	Snacks	5	Marc	\$21

# Modellazione multidimensionale

Obiettivo: query veloci e usabilità

id	date	product	category	customer_id	profit
1	1/2/2022	Fulltoss Tangy Tomato	Vegetables	2	\$23
2	1/2/2022	Chilli - Green, Organically Grown	Snacks	2	\$12
3	1/2/2022	Masala Powder	Herbs	5	\$93
4	1/2/2022	Cheese Cracker (Mcvities)	Snacks	1	\$23
5	1/2/2022	Centre Filled Chocolate Cake	Snacks	5	\$21

# Modellazione multidimensionale

Obiettivo: query veloci e usabilità

id	date	product	category	customer_id	profit
1	1/2/2022	Fulltoss Tangy Tomato	Vegetables	2	\$23
2	1/2/2022	Chilli - Green, Organically Grown	Snacks	2	\$12
3	1/2/2022	Masala Powder	Herbs	5	\$93
4	1/2/2022	Cheese Cracker (McVities)	Snacks	1	\$23
5	1/2/2022	Centre Filled Chocolate Cake	Snacks	5	\$21

# Modellazione multidimensionale

Obiettivo: query veloci e usabilità

		Chiave Esterna (FK)	Chiave Esterna (FK)	
id	date	product_id	customer_id	profit
1	1/2/2022	2	2	\$23
2	1/2/2022	5	2	\$12
3	1/2/2022	6	5	\$93
4	1/2/2022	23	1	\$23
5	1/2/2022	16	5	\$21

Tabella più stretta

**Tabella dei fatti**  
- misura: profitto

Chiave  
Primaria (PK)

product_id	product	category
1	product 1	Vegetables
2	product 2	Snacks
3	product 3	Herbs
4	product 4	Snacks
5	product 5	Snacks

**Dimensione  
Prodotto**



# Modellazione multidimensionale

Obiettivo: query veloci e usabilità

id	date_id	product_id	customer_id	profit
1	20220102	2	2	\$23
2	20220102	5	2	\$12
3	20220102	6	5	\$93
4	20220102	23	1	\$23
5	20220102	16	5	\$21



Performance

Usabilità



Dim_Product
product_id
name
category
subcategory
dimensions

Dim_Customer
customer_id
first name
last name
sex
city

Sales
date_id
product_id
customer_id
units
price

Dim_Date
date_id
year
quarter
month
week
day
weekday
holiday_flag

- ✓ Generalmente aggregabili
- ✓ Basate su eventi o transazioni
- ✓ Misurabili e non descrittivi
- ✓ Si svolgono solitamente in un momento specifico (c'è una data)
- ✓ Ci sono molte chiavi esterne

id	date_id	region_id	profit
1	20220102	1	\$23
2	20220102	2	\$12
3	20220102	2	\$93
4	20220102	3	\$23
5	20220102	16	\$21

- ✓ **Granularità:** il livello più atomico di un fatto
- ✓ Nella tabella abbiamo i profitti in una riga per una regione specifica in una data specifica

Dim_Product
product_id
name
category
subcategory
dimensions

Dim_Customer
customer_id
first name
last name
sex
city

Sales
date_id
product_id
customer_id
units
price

Dim_Date
date_id
year
quarter
month
week
day
weekday
holiday_flag

- ✓ Non aggregabili
- ✓ Carattere descrittivo
- ✓ Solitamente statiche
- ✓ A volte possono avere una chiave esterna (fiocco di neve)

# Schema a stella

## Normalizzazione:

- ✓ Tecnica che elimina la ridondanza
- ✓ Minimizza lo spazio
- ✓ Performance per scrittura / update
- ✓ Molte tabelle... molti join

FK

sales_id	product_id	customer_id	units	price
1	3	23	1	2.99
2	5	13	1	1.99
3	2	7	2	3.49
4	3	16	1	2.29
5	3	13	5	1.49

1:n

✓ Facts

PK

✓ Dimensions

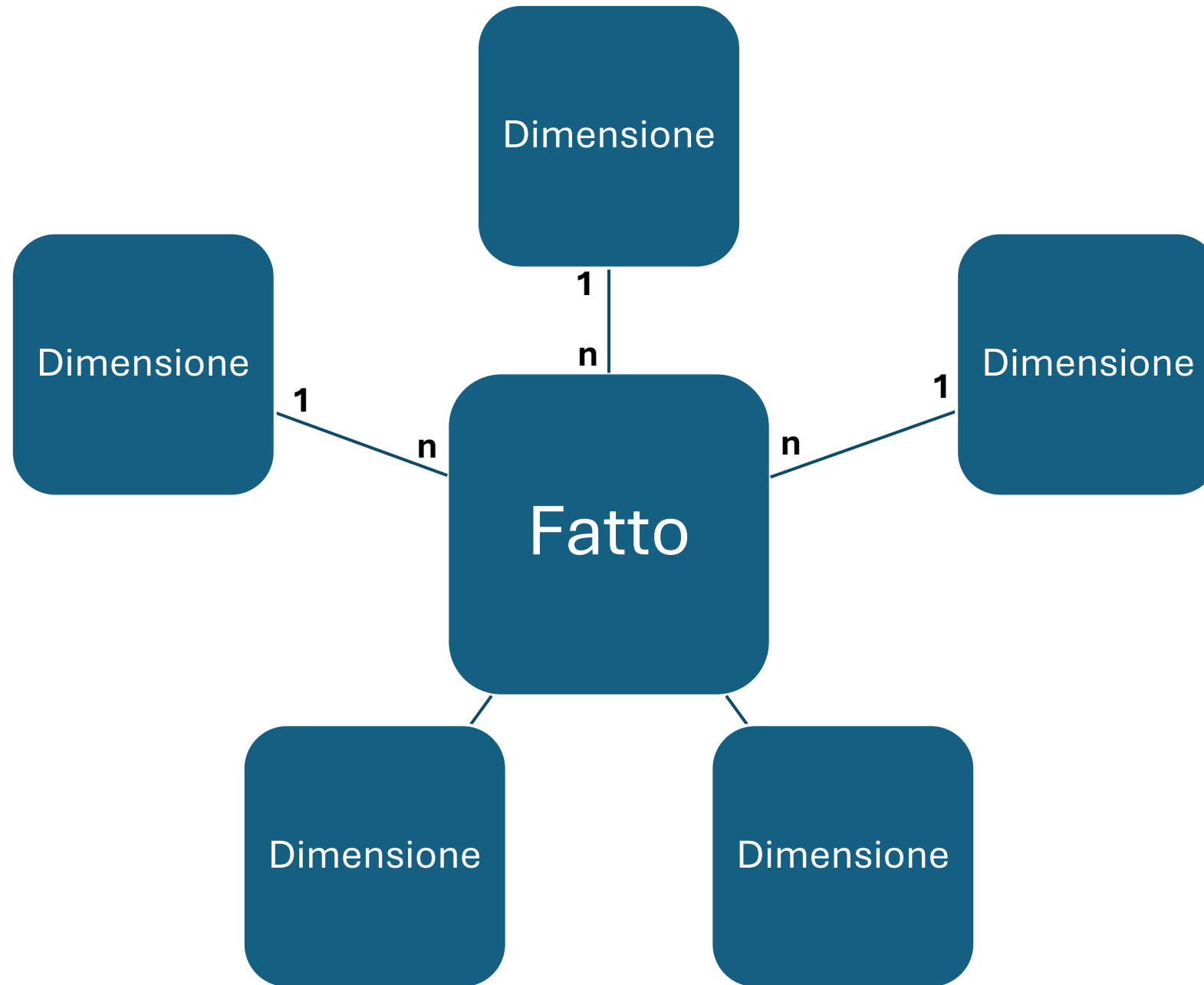
product_id	name	category	sub_category
1	Chili	Herbs	Spices
2	Garlic	Fruits & Vegetables	Vegetable
3	Banana	Fruits & Vegetables	Fruits
4	Chocolate	Sweets & Snacks	Sweets
5	Chips	Sweets & Snacks	Snacks

✓ Dimensions

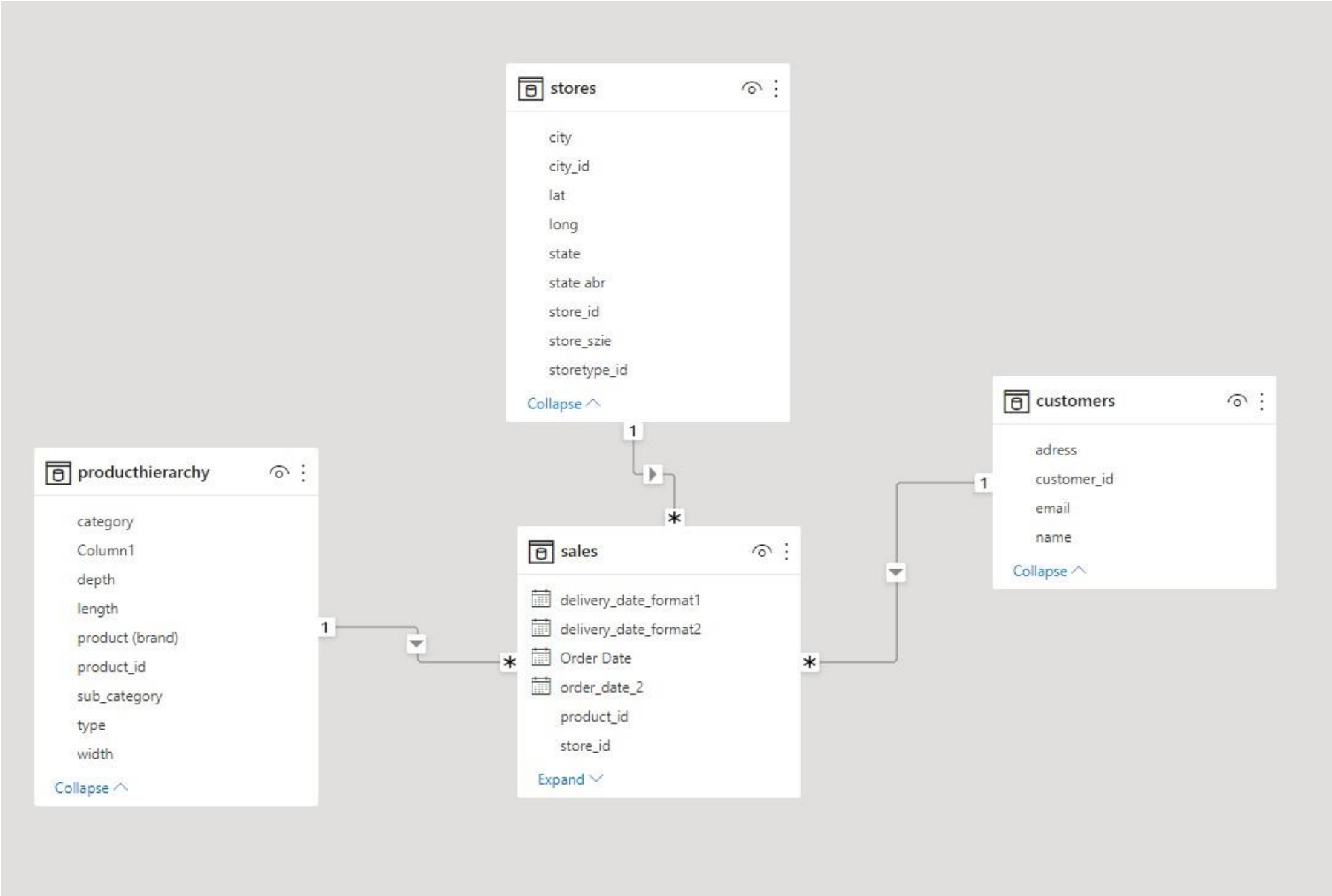
## Schema:

- ✓ Con una certa **ridondanza** di dati
- ✓ Ottimizzato per la lettura
- ✓ Usabilità

... DENORMALIZZATO

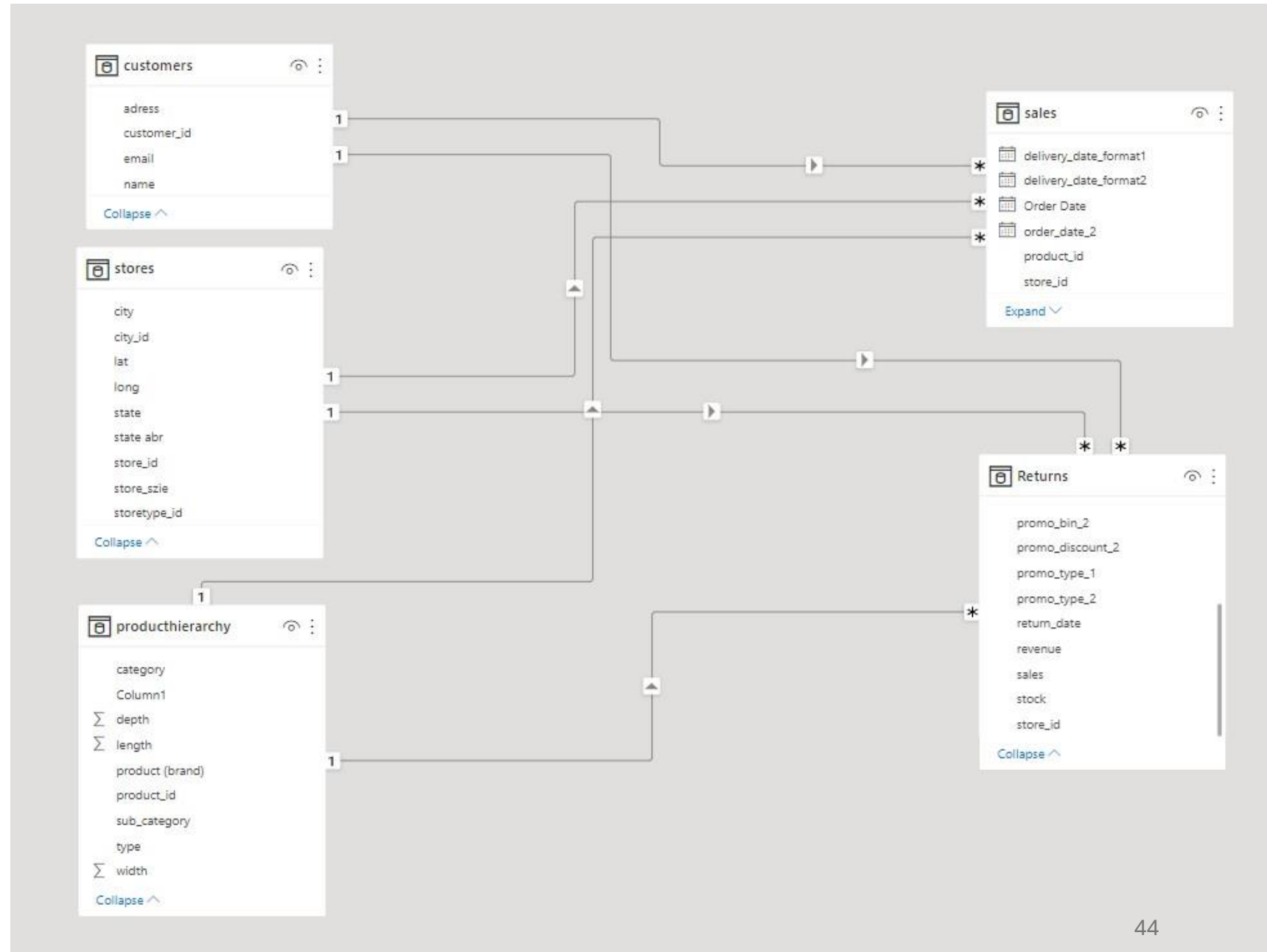


# Schema a stella in Power BI



# Schema a stella in Power BI: più tabelle dei fatti

- ✓ Attenzione alla granularità
- ✓ Non collegare MAI le tabelle dei fatti!





# Schema a fiocco di neve

Normalizzazione. Da così...

sales_id	product_id	customer_id	units	price
1	3	23	1	2.99
2	5	13	1	1.99
3	2	7	2	3.49
4	3	16	1	2.29
5	3	13	5	1.49

✓ Facts

✓ Dimensions

product_id	name	category	sub_category
1	Chili	Herbs	Spices
2	Garlic	Fruits & Vegetables	Vegetable
3	Banana	Fruits & Vegetables	Fruits
4	Chocolate	Sweets & Snacks	Sweets
5	Chips	Sweets & Snacks	Snacks

# Schema a fiocco di neve

... a così

✓ Facts

sales_id	product_id	customer_id	units	price
1	3	23	1	2.99
2	5	13	1	1.99
3	2	7	2	3.49
4	3	16	1	2.29
5	3	13	5	1.49

product_id	name	category_id	sub_category
1	Chili	1	Spices
2	Garlic	2	Vegetable
3	Banana	2	Fruits
4	Chocolate	3	Sweets
5	Chips	3	Snacks

lo Snowflake è più normalizzato

category_id	category
1	Herbs
2	Fruits & Vegetables
3	Sweets & Snacks

## Vantaggi

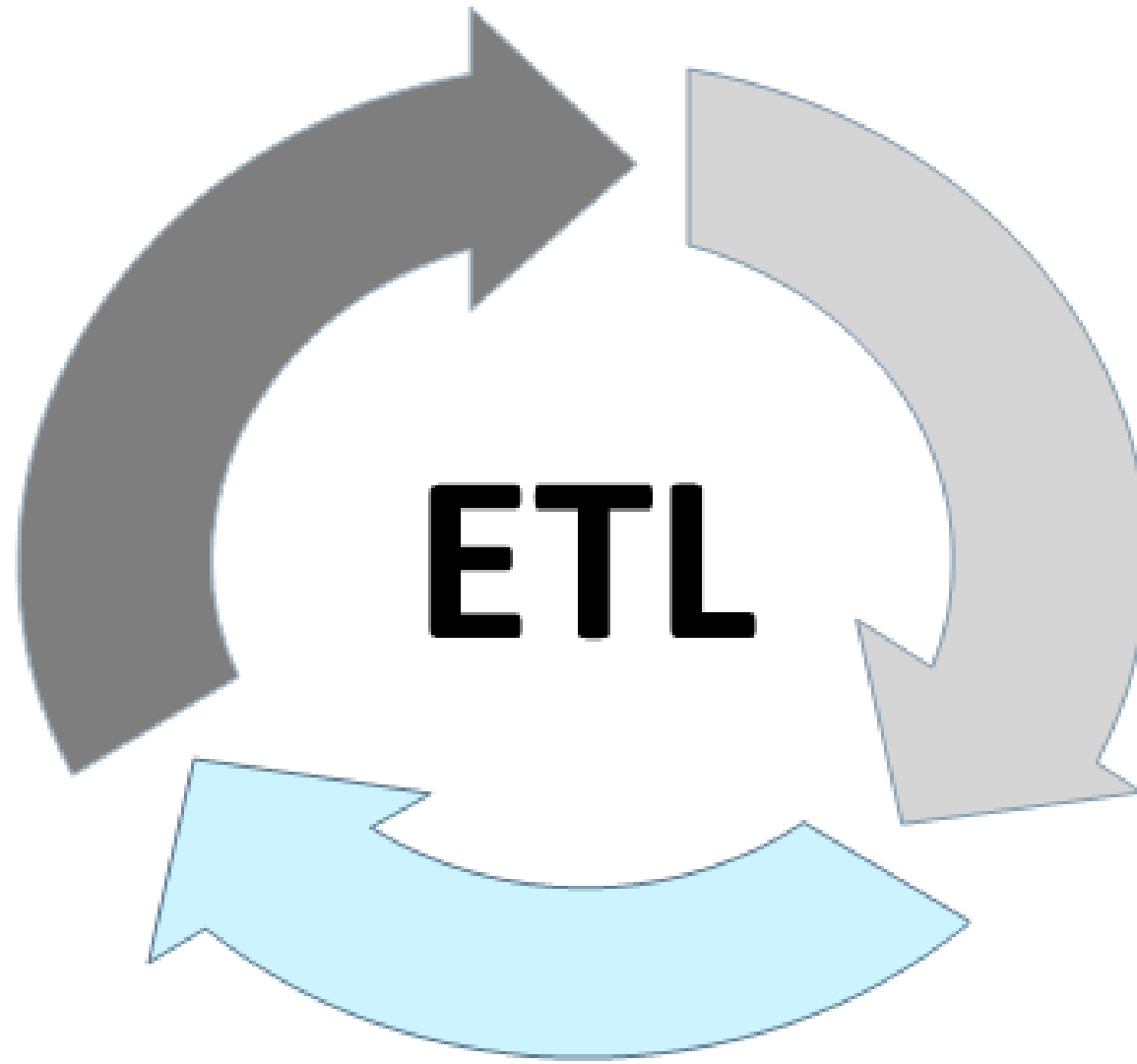
- ✓ Meno storage
- ✓ Meno dati ridondanti
- ✓ Risolve i rallentamenti in scrittura

## Svantaggi

- ✓ Più complicato
- ✓ Più join
- ✓ Meno performance in lettura

Per i Data Mart → Star Schema

Per il DataWareHouse → Star o Snowflake



- ✓ Caricamento File Excel

# Editor di Power Query per l'ETL

**Maven Northwind Challenge v2**

**Home Transform Add Column View Tools Help**

**1** Close New Query Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Properties Advanced Editor Manage Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: V Use First Row Replace Value Transform

**2** Queries [11]  
Fact table [2]  
Fct\_order\_details  
Fct\_Orders  
Dimension [5]  
Dim\_Categories  
Dim\_Customers  
Dim\_Employees  
Dim\_Products  
Dim\_Shippers  
Support table [3]  
Table\_Flags  
Table\_Flags (Employees)  
Table\_TableCountryCode  
Other Queries [11]

**3** Query Settings  
Name: Fct\_Orders  
All Properties  
APPLIED STEPS  
Source  
En-têtes promus  
Type modifié  
Valeur remplacée  
Type modifié1

**4** Query Editor  
Table: Fct\_Orders  
Columns: 8  
Rows: 830  
Column profiling based on top 1000 rows

**5** Query Settings  
Name: Fct\_Orders  
All Properties  
APPLIED STEPS  
Source  
En-têtes promus  
Type modifié  
Valeur remplacée  
Type modifié1

PREVIEW DOWNLOADED AT 50.49

- ✓ Modificare posizione file origine

# Profilazione dati in Power Query

Senza titolo - Editor di Power Query

File Home Trasforma Aggiungi colonna Visualizza Strumenti Guida

Impostazioni query

Layout Anteprenda dati Colonne Parametri Avanzate Dipendenze

Barra della formula

Spaziatura fissa

Mostra spazi vuoti

Colonna qualità

Colonna distribuzione

Profilo colonna

Consenti sempre

Vai alla colonna

Editor avanzato

Dipendenze query

Query [1]

ProductCost

= Table.TransformColumnTypes(#"Promoted Headers",{{"SKU", type text}, {"Product", type text}, {"ProductCost",

	SKU	Product	ProductCost
Validi	100%	100%	100%
Errore	0%	0%	0%
Vuoto	0%	0%	0%
20 valori distinti, 20 valori univoci			
1	1010-GL120-3C	Trainer - Tailspin GL-120	59,05

Impostazioni query

PROPRIETÀ

Nome

ProductCost

Tutte le proprietà

PASSAGGI APPLICATI

Source

Navigation

Promoted Headers



- ✓ Origini dati relazionali
- ✓ Cenni sulle modalità di connessione
- ✓ Importare i dati scrivendo una query

- ✓ Connessioni ODBC
- ✓ Caricamento file di testo (txt o csv)
- ✓ Caricamento file JSON
- ✓ Caricamento file XML
- ✓ Caricamento file PDF
- ✓ Caricamento file da cartella

1. Importa
2. Direct Query
3. Mista
4. Live connection
5. Push Dataset

- I dati vengono importati nel modello di dati, compressi e ottimizzati e quindi archiviati sull'istanza di Analysis Services.
- **I dati interrogati vengono caricati in memoria** consentendo prestazioni molto **veloci**.
- **massima flessibilità:** a disposizione **l'intera gamma** di trasformazioni del linguaggio «M» e le **funzioni DAX**.
- **modalità predefinita** (ed in molti casi l'unica possibile).
- aggiornamenti periodici (refresh) dei dati (ad es. otto volte al giorno su Power BI Pro, fino a 48 su Power BI Premium).
- Limiti di dimensione del dataset (1 GB in Power BI Pro, dimensioni maggiori in Premium).

- viene eseguita una *query* sull'origine dati effettiva, non su una copia importata.
- **grande impatto sulle prestazioni del server** dell'origine dati che si rifletterà sull'esperienza dell'utente.
- è un'opzione utile quando i *report* richiedono informazioni molto aggiornate e quando le opzioni di aggiornamento pianificato non garantiscono una frequenza ottimale.
- È un'opzione quando il set di dati è troppo grande per essere importato in power bi.
- Alcune funzionalità «M» e DAX possono essere limitate o non disponibili.





























- le tabelle con questa impostazione possono essere o meno archiviate in una memoria temporanea (*cache*) a seconda della tipologia di *query* inviata al set di dati di Power BI.
- In alcuni casi, le *query* saranno soddisfatte con i dati memorizzati nella *cache* mentre in altri occorrerà eseguire la richiesta sull'origine dati.
- Alcune funzionalità «M» e DAX possono essere limitate o non disponibili.

- Power BI si collega “live” a un modello esistente (come Analysis Services o un dataset Power BI pubblicato).
- unica fonte della verità: il modello dati centralizzato viene gestito e aggiornato in un solo luogo.
- la responsabilità del refresh è esterna
- Power BI Desktop presenta limitazioni nella creazione o modifica del modello (ad es. non è possibile aggiungere nuove tabelle calcolate)

- Permette di inviare (push) dati in tempo quasi reale a un dataset in Power BI dal tuo servizio o applicazione.
- I dati vengono “spinti” a Power BI via API o flussi (ad es. streaming) e memorizzati in un dataset temporaneo.
- Utilizzato per scenari di streaming o dashboard in real-time.



# Scegliere una modalità di archiviazione

		IMPORT	DIRECT QUERY	LIVE CONNECT	MIXED	DIRECT LAKE
		Data stored in memory	No data stored in memory	Suited for enterprise solutions	Combine Import, DQ, LC	Connect to Delta Tables
		Remains fastest storage mode	Full reporting capabilities	No data stored in memory	Enable larger data models	No cacl. tables/columns
		Full reporting capabilities	Limited data transformations	Full reporting capabilities	DQ version best for self-service	Pages data into memory
		Full transformation capabilities	Limited data modelling	Limited data transformations	Limited data transformations	Cache frequent used data
		Full modelling capabilities	Slower performance	Limited data modelling	Limited data modelling	Requires Fabric capacity
		Queries not sent to source	Queries sent to source	Queries sent to source	Good with Agg tables	Handles large data
		Scheduled refresh required	No scheduled refresh required	No scheduled refresh required	Scheduled refresh maybe	Not compatible with views
DATA SOURCE		   Excel   SQL DB   CSV/Text + 100+ Other Data Connectors	 SQL DB + 25+ Other Data Connectors	  PBI Semantic Models   +   Azure Analysis Services	   Excel (Import)   +   SQL DB (DQ)   +   PBI (LC)	   Fabric Lakehouse   One Lake   Fabric Warehouse
COMPONENT STORED WITHIN	DATA SOURCE		 Data	  Data   Model	 Data	 Data
	POWER BI	   Data   Model   Visuals	  Model   Visuals	 Visuals	   Data   Model   Visuals	  Model   Visuals

Alcune delle azioni principali includono:

- » Rimuovi: Rimuove la colonna selezionata dai dati di Power Query.
- » Rimuovi altre colonne: Rimuove tutte le colonne non selezionate dai dati di Power Query.
- » Duplica colonna: Crea un duplicato della colonna selezionata come nuova colonna posizionata all'estremità destra della tabella.
- » Rimuovi duplicati: Rimuove tutte le righe dalla colonna selezionata in cui i valori duplicano valori precedenti.
- » Cambia tipo: Cambia il tipo di dati della colonna selezionata.

Le principali azioni:

- » Usa Prima Riga come Intestazione: Sostituisce ogni nome di intestazione della tabella con i valori nella prima riga di ogni colonna.
- » Aggiungi Colonna Personalizzata: Inserisce una nuova colonna dopo l'ultima colonna della tabella.
- » Aggiungi Colonna Condizionale: Inserisce una nuova colonna che contiene i risultati di una specifica istruzione IF...THEN...ELSE.
- » Rimuovi Duplicati: Rimuove tutte le righe in cui i valori nelle colonne selezionate duplicano valori precedenti.
- » Unisci Query: Crea una nuova query che unisce la tabella corrente con un'altra query nel workbook facendo corrispondere le colonne specificate.

Elenco trasformazioni eseguite nell'esercizio

1. Rimuovi prime 10 righe
2. Formato → Taglio e poi Pulisci
3. Dividi colonna in base al numero di caratteri (15)
4. Usa la prima riga come intestazione
5. Rimuovi colonne non necessarie
6. Merge di colonne
7. Dividi colonna in base al delimitatore
8. Rinomina colonne
9. Cambia formato colonna «Amount»
10. Cambia formato colonne «Date»
11. Rimuovi errori
12. Filtrare gli «Amount» non nulli

1. Accoda Query
2. Merge di Query (Join)

## Esercizio 2 – Combina le tabelle da file PDF

Elenco trasformazioni eseguite nell'esercizio:

1. Rimuovi prime righe
2. Usa la prima riga come intestazione
3. Rinomina Colonne
4. Accoda query come nuove

# Left Outer Join

Transazioni

Conto	Reparto	Data	Importo
64010	150	15/12/2015	8.975
64020	150	15/12/2015	13.708
64030	150	15/12/2015	32.555
64010	250	15/12/2015	22.752
64015	150	15/12/2015	34.147
64030	250	15/12/2015	19.733
64040	250	15/12/2015	33.438
64010	350	15/12/2015	45.876

Piano dei Conti

Conto	Reparto	Nome
64010	150	Mele
64020	150	Pere
64030	150	Arance
64040	150	Susine
64010	250	Pomodori
64020	250	Insalata
64030	250	Pane
64040	250	Pizza



Conto	Reparto	Data	Importo	PianoDeiConti.Conto	PianoDeiConti.Reparto	PianoDeiConti.Nome
64010	150	15/12/2015 00:00	8975	64010	150	Mele
64020	150	15/12/2015 00:00	13708	64020	150	Pere
64030	150	15/12/2015 00:00	32555	64030	150	Arance
64010	250	15/12/2015 00:00	22752	64010	250	Pomodori
64030	250	15/12/2015 00:00	19733	64030	250	Pane
64040	250	15/12/2015 00:00	33438	64040	250	Pizza
64015	150	15/12/2015 00:00	34147			
64010	350	15/12/2015 00:00	45876			

# Right Outer Join

Transazioni

Conto	Reparto	Data	Importo
64010	150	15/12/2015	8.975
64020	150	15/12/2015	13.708
64030	150	15/12/2015	32.555
64010	250	15/12/2015	22.752
64015	150	15/12/2015	34.147
64030	250	15/12/2015	19.733
64040	250	15/12/2015	33.438
64010	350	15/12/2015	45.876

Piano dei Conti

Conto	Reparto	Nome
64010	150	Mele
64020	150	Pere
64030	150	Arance
64040	150	Susine
64010	250	Pomodori
64020	250	Insalata
64030	250	Pane
64040	250	Pizza



Conto	Reparto	Data	Importo	PianoDeiConti.Conto	PianoDeiConti.Reparto	PianoDeiConti.Nome
64010	150	15/12/2015 00:00	8975	64010	150	Mele
64020	150	15/12/2015 00:00	13708	64020	150	Pere
64030	150	15/12/2015 00:00	32555	64030	150	Arance
64010	250	15/12/2015 00:00	22752	64010	250	Pomodori
				64040	150	Susine
64030	250	15/12/2015 00:00	19733	64030	250	Pane
				64020	250	Insalata
64040	250	15/12/2015 00:00	33438	64040	250	Pizza



Transazioni

Conto	Reparto	Data	Importo
64010	150	15/12/2015	8.975
64020	150	15/12/2015	13.708
64030	150	15/12/2015	32.555
64010	250	15/12/2015	22.752
64015	150	15/12/2015	34.147
64030	250	15/12/2015	19.733
64040	250	15/12/2015	33.438
64010	350	15/12/2015	45.876

Piano dei Conti

Conto	Reparto	Nome
64010	150	Mele
64020	150	Pere
64030	150	Arance
64040	150	Susine
64010	250	Pomodori
64020	250	Insalata
64030	250	Pane
64040	250	Pizza



Conto	Reparto	Data	Importo	PianoDeiConti.Conto	PianoDeiConti.Reparto	PianoDeiConti.Nome
64010	150	15/12/2015 00:00	8975	64010	150	Mele
64020	150	15/12/2015 00:00	13708	64020	150	Pere
64030	150	15/12/2015 00:00	32555	64030	150	Arance
64010	250	15/12/2015 00:00	22752	64010	250	Pomodori
				64040	150	Susine
64030	250	15/12/2015 00:00	19733	64030	250	Pane
				64020	250	Insalata
64040	250	15/12/2015 00:00	33438	64040	250	Pizza
64015	150	15/12/2015 00:00	34147			
64010	350	15/12/2015 00:00	45876			

# Inner Join

Transazioni

Conto	Reparto	Data	Importo
64010	150	15/12/2015	8.975
64020	150	15/12/2015	13.708
64030	150	15/12/2015	32.555
64010	250	15/12/2015	22.752
64015	150	15/12/2015	34.147
64030	250	15/12/2015	19.733
64040	250	15/12/2015	33.438
64010	350	15/12/2015	45.876

Piano dei Conti

Conto	Reparto	Nome
64010	150	Mele
64020	150	Pere
64030	150	Arance
64040	150	Susine
64010	250	Pomodori
64020	250	Insalata
64030	250	Pane
64040	250	Pizza



Conto	Reparto	Data	Importo	PianoDeiConti.Conto	PianoDeiConti.Reparto	PianoDeiConti.Nome
64010	150	15/12/2015 00:00	8975	64010	150	Mele
64020	150	15/12/2015 00:00	13708	64020	150	Pere
64030	150	15/12/2015 00:00	32555	64030	150	Arance
64010	250	15/12/2015 00:00	22752	64010	250	Pomodori
64030	250	15/12/2015 00:00	19733	64030	250	Pane
64040	250	15/12/2015 00:00	33438	64040	250	Pizza

# Left Anti Join

Transazioni

Conto	Reparto	Data	Importo
64010	150	15/12/2015	8.975
64020	150	15/12/2015	13.708
64030	150	15/12/2015	32.555
64010	250	15/12/2015	22.752
64015	150	15/12/2015	34.147
64030	250	15/12/2015	19.733
64040	250	15/12/2015	33.438
64010	350	15/12/2015	45.876

Piano dei Conti

Conto	Reparto	Nome
64010	150	Mele
64020	150	Pere
64030	150	Arance
64040	150	Susine
64010	250	Pomodori
64020	250	Insalata
64030	250	Pane
64040	250	Pizza



Conto	Reparto	Data	Importo
64015	150	15/12/2015 00:00	34147
64010	350	15/12/2015 00:00	45876

Transazioni

Conto	Reparto	Data	Importo
64010	150	15/12/2015	8.975
64020	150	15/12/2015	13.708
64030	150	15/12/2015	32.555
64010	250	15/12/2015	22.752
64015	150	15/12/2015	34.147
64030	250	15/12/2015	19.733
64040	250	15/12/2015	33.438
64010	350	15/12/2015	45.876

Piano dei Conti

Conto	Reparto	Nome
64010	150	Mele
64020	150	Pere
64030	150	Arance
64040	150	Susine
64010	250	Pomodori
64020	250	Insalata
64030	250	Pane
64040	250	Pizza



PianoDeiConti.Conto	PianoDeiConti.Reparto	PianoDeiConti.Nome
64040	150	Susine
64020	250	Insalata

# Full anti Join (non c'è in Power BI ma c'è In Excel)

Transazioni

Conto	Reparto	Data	Importo
64010	150	15/12/2015	8.975
64020	150	15/12/2015	13.708
64030	150	15/12/2015	32.555
64010	250	15/12/2015	22.752
64015	150	15/12/2015	34.147
64030	250	15/12/2015	19.733
64040	250	15/12/2015	33.438
64010	350	15/12/2015	45.876

Piano dei Conti

Conto	Reparto	Nome
64010	150	Mele
64020	150	Pere
64030	150	Arance
64040	150	Susine
64010	250	Pomodori
64020	250	Insalata
64030	250	Pane
64040	250	Pizza

Conto	Reparto	Data	Importo	PianoDeiConti.Conto	PianoDeiConti.Reparto	PianoDeiConti.Nome
64015	150	15/12/2015 00:00	34147			
64010	350	15/12/2015 00:00	45876			
				64040	150	Susine
				64020	250	Insalata

# Prodotto Cartesiano (Cross Join)

Da questo...			A Questo...		
Month	Expenses	Budget	Expenses	Budget	Month
31/01/2024	Property Tax	5.450,00	Property Tax	5.450,00	31/01/2020
29/02/2024	Insurance	1.253,00	Property Tax	5.450,00	29/02/2020
31/03/2024	Telephony	1.640,00	Property Tax	5.450,00	31/03/2020
30/04/2024			Property Tax	5.450,00	30/04/2020
31/05/2024			Property Tax	5.450,00	31/05/2020
30/06/2024			Property Tax	5.450,00	30/06/2020
31/07/2024			Property Tax	5.450,00	31/07/2020
31/08/2024			Property Tax	5.450,00	31/08/2020
30/09/2024			Property Tax	5.450,00	30/09/2020
31/10/2024			Property Tax	5.450,00	31/10/2020
30/11/2024			Property Tax	5.450,00	30/11/2020
31/12/2024			Property Tax	5.450,00	31/12/2020
			Insurance	1.253,00	31/01/2020
			Insurance	1.253,00	29/02/2020
			Insurance	1.253,00	31/03/2020
			Insurance	1.253,00	30/04/2020
			Insurance	1.253,00	31/05/2020
			Insurance	1.253,00	30/06/2020
			Insurance	1.253,00	31/07/2020

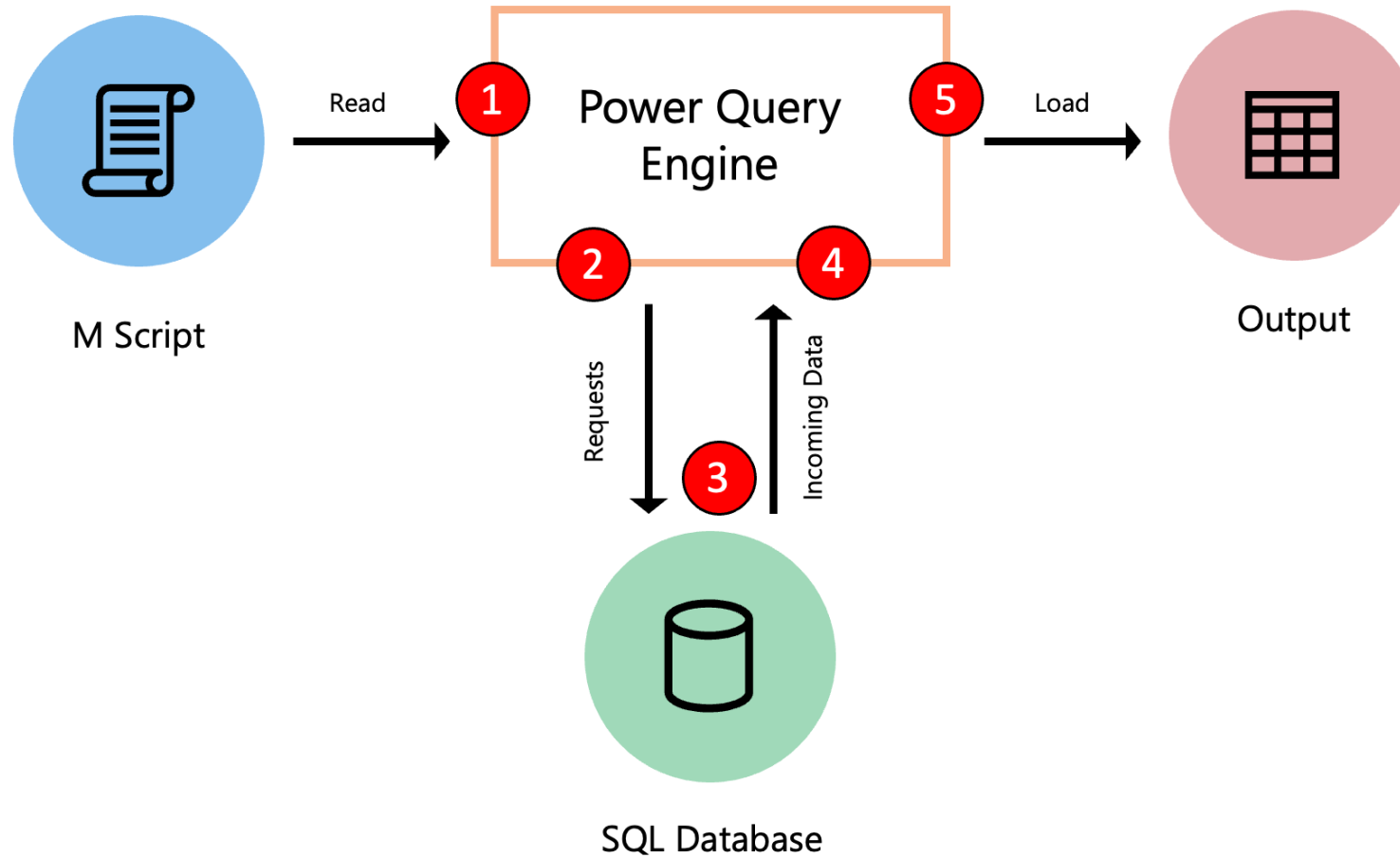
Calcolare la data di incasso ed il relativo ammontare di ogni ricavo a budget.

Trasformazioni eseguite nell'esercizio:

1. (ricavi) Rimuovi Colonne
2. (ricavi) Trasforma altre colonne tramite Unpivot
3. (ricavi) Rinomina colonne
4. (ricavi) Trasforma Tipo Dati in Numero Intero della colonna MeseRicavo
5. (dilazioni) Trasforma altre colonne tramite Unpivot
6. (dilazioni) Rinomina colonne
7. (ricavi) Aggiungi colonna personalizzata (= **dilazioni, effettua il Prodotto Cartesiano**)
8. (ricavi) Espandi colonna
9. (ricavi) Trasforma Tipo Dati delle colonne espanse
10. (ricavi) Aggiungi colonna personalizzata Incasso
11. (ricavi) Aggiungi colonna personalizzata MeseIncasso
12. (ricavi) Aggiungi colonna condizionale Mese
13. (ricavi) Aggiungi colonna condizionale Anno
14. (ricavi) Aggiungi colonna personalizzata DataIncasso
15. (ricavi) Trasforma Tipo Dati in Data della colonna DataIncasso
16. (ricavi) Trasforma Data in ultimo giorno del mese
17. (ricavi) Rimuovi colonne non necessarie

# Come funzionano le query di Power Query

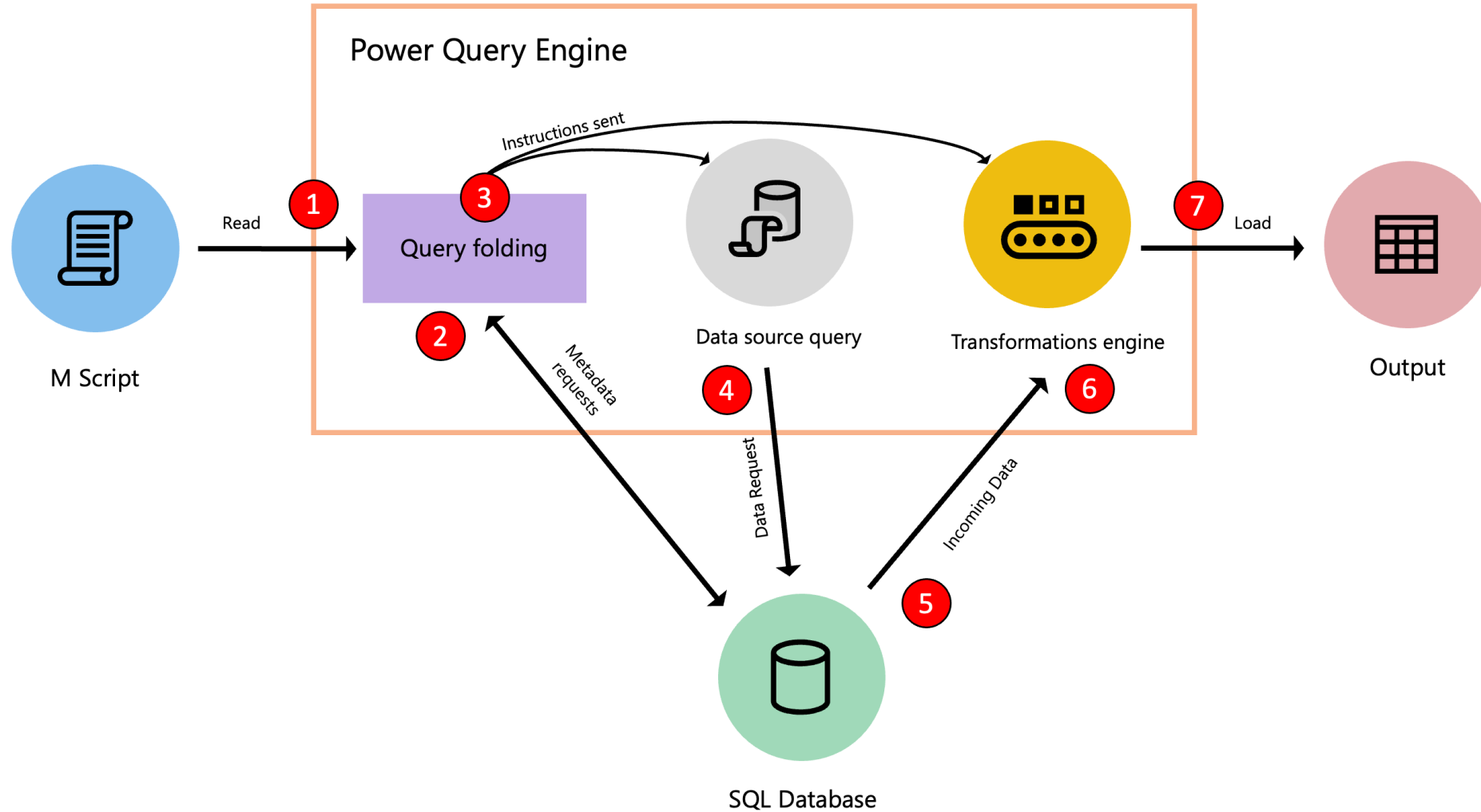
Schema di valutazione delle query





# Come funzionano le query di Power Query

## Ottimizzazione delle query



# Come funzionano le query di Power Query

Ottimizzazione delle query

- QUERY FOLDING
- DIAGNOSTICA QUERY

- Duplicare o referenziare query
- Documentare query e passaggi

