Che cosa è il modello dati e perchè è importante in Power Bl.



Sponsors











With the support of:









About me

- Consulente e formatore in ambito business intelligence
- Nel 2020 Fondatore e CTO start-up Cargo BI per insurance data analytics per le compagnie assicurative

Dal 2017 mi occupo di architetture big Data e in generale di tutta la proposizione data platform di Microsoft.

- Docente all'Università di Pordenone per i corsi IFTS di analisi Big Data
- Community Lead di 1nn0va (www.innovazionefvg.net)
- MCP,MCSA,MCSE dal 2017 MCT e dal 2014 MVP per SQL Server e relatore in diverse conferenze sul tema.
 - info@marcopozzan.it
 - @marcopozzan.it
 - www.marcopozzan.it
 - http://www.scoop.it/u/marco-pozzan
 - http://paper.li/marcopozzan/1422524394









Solutions Associate

SQL Server 2012/2014

AGENDA

Concetti base di data modelling

Tipi diversi di data model

Data Model in Power BI

Caso d'uso reale

Lavorare con 1 singola tabella

In excel siamo abituati a lavorare con una singola tabella

Con excel è semplice ed è già un data model

Ma ci sono delle limitazioni

- Il numero di righe è minore di un milione (Il limite della dimensione limita la dimensione del modello)
- La velocità e la memoria non sono ottimali
- Possiamo fare solo calcoli basici

Lavorare con 1 singola tabella: Problema 1

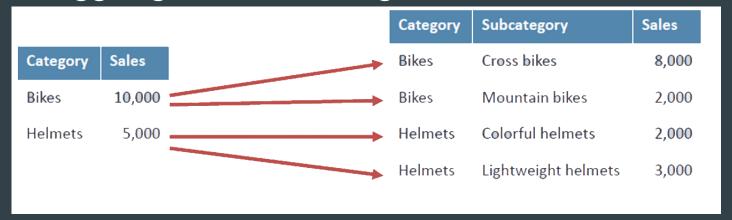
- Problema 1: limite analitico
- Se abbiamo il seguente excel (L'ESPRESSIVITA' massima è data dalle colonne che si vedono)
- Se non ci sono le colonne nel tabellone non si possono fare le analisi
- Esempio: Se dobbiamo aggiungere l'analisi per Nome del Podotto che non c'è allora dobbiamo portarci il nome dentro la nostra tabella presa da un'altra tabella

FullDateLabel 💌	Manufacturer -	BrandName 💌	ProductCategoryName	▼ SalesQuantity ▼	SalesAmount 💌	TotalCost 💌
2007-03-31	Adventure Works	Adventure Works	Home Appliances	55	14332.268	7651.84
2008-10-22	Contoso, Ltd	Contoso	Cell phones	2040	23504.88	12648.94
2009-01-31	Adventure Works	Adventure Works	TV and Video	194	51593.106	28146.4
2009-01-21	Fabrikam, Inc.	Fabrikam	Cameras and camcorders	282	163007.2	76709.45
2007-12-31	Adventure Works	Adventure Works	Computers	29	14008.43	7944.32
2007-06-22	Contoso, Ltd	Contoso	Cell phones	680	6107.24	3420.44
2007-06-22	Proseware, Inc.	Proseware	Computers	86	71417.6	30786.94
2007-08-23	Adventure Works	Adventure Works	Computers	43	22672.2	9954.6
2009-03-30	The Phone Company	The Phone Company	Cell phones	198	48500.37	24164.56
2008-03-24	Contoso, Ltd	Contoso	Cell phones	306	7353.594	3914.64
2007-09-30	Fabrikam, Inc.	Fabrikam	Home Appliances	44	4805.604	2824.24
2007-11-13	Adventure Works	Adventure Works	Computers	153	47357.97	28256.02
2008-12-06	Contoso, Ltd	Contoso	Computers	32	10790.4	6477.2
2007-11-14	Contoso, Ltd	Contoso	Cameras and camcorders	146	55397.5	25876
2009-12-30	Adventure Works	Adventure Works	Computers	32	15107.75	7952.97
2009-03-13	Wide World Importers	Wide World Importers	Audio	42	7990.92	3607.26
2009-08-11	Wide World Importers	Wide World Importers	Audio	9	1466.1	749.16
2009-09-28	Contoso, Ltd	Contoso	Home Appliances	78	9955.268	5189.27
2008-02-18	A. Datum Corporation	A. Datum	Cameras and camcorders	345	70989.93	32872.58
2007-08-15	Litware, Inc.	Litware	Home Appliances	69	112603.8	56472.35

Lavorare con 1 singola tabella: Problema 2

• Problema 2: conseguenza del problema 1 (limite dimensioni di 1 milione di righe) Se nella tabella della slide precedente devo aggiungere le informaioni sul nome del prodotto aumento di fatto le dimensioni perchè aumenta la Granularità (Livello di dettaglio della tabella)

• Supponiamo ora di aggiungere la Sottocategoria nel nostro tabellone.



 Più granularità più aumentano le righe oltre alle colonne quindi abbiamo un aumento in profondità e in larghezza

Che cosa è un data model?

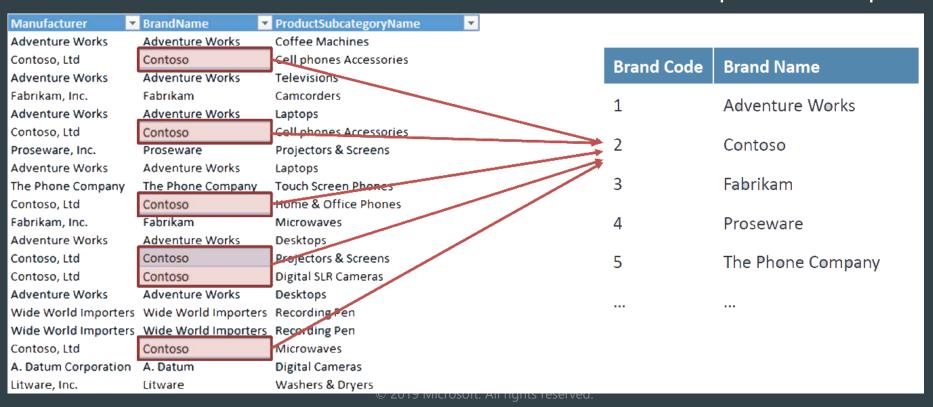
- Usando Power Pivot o Power BI si sfruttano meglio i dati
- Non ci sono limiti nella dimensione delle tabelle quindi possiamo definire la granularità a piacimento, anche con una sola tabella (Risoluzione del 2 problema del limite di righe) ma non giustifica comunque l'uso di una sola tabella
- Possiamo caricare più tabelle in un modello singolo (Risoluzione del 1 problema dell'espressività) perchè:
 - Costruiamo delle relazioni tra le tabelle
 - Possiamo sfruttare la potenza di DAX per interrogare i nostri dati
- Quello che andiamo a creare con le tabelle e le relazioni e con DAX è un **Data Model** che permette di realizzare le analisi sui dati
- <u>Il Data Model la colonna portante di quasi tutte le nostre soluzioni aziendali di business</u> intelligence e IoT.

Basi di modellazione

- Ora abbiamo scoperto che con Power BI o Power Pivot per Excel possiamo definire un data model che pemette di lavorare nel modo migliore con i dati
- La prossima fase è pensare a come costruire il data model in maniera corretta in quanto non basta passare ai tool di Power BI e Power Pivot per risolvere i nostri problemi.
- Non possiamo pensare comunque di mettere tutto su una tabella e tra poco vediamo perchè

Basi di modellazione (normalizzazione)

- Abbandonare l'idea di avere un unico tabellone ottenuto con un processo di denormalizzazione di più tablle
- Dobbiamo cercare di normalizzazione il tabellone unico
- La normalizzazione è il processo di organizzazione di colonne (attributi) e tabelle di un database per ridurre la ridondanza (risparmiare spazio) dei dati e migliorare l'integrità dei dati (si usa nei database OLTP dove devo avere velocità e risparmio di spazio)



Basi di modellazione (denormalizzazione)

- La denormalizzazione aumenta la ridondanza dei dati, con l'obiettivo di migliorare la comprensione del modello (Denormalizzazione nei customer). La denormalizzazione NON è il male assoluto ma non va bene quando si usa su un tabellone unico anzi come vedremo di seguito si usa per i fini analitici.
- Il grado di denormalizzazione dipende dalle analisi che dobiamo fare e lo vedremo successivamente

FullDateLabel <	Manufacturer *	BrandName <	ProductCategoryName	▼ SalesQuantity ▼	SalesAmount 💌	TotalCost *
2007-03-31	Adventure Works	Adventure Works	Home Appliances	55	14332.268	7651.84
2008-10-22	Contoso, Ltd	Contoso	Cell phones	2040	23504.88	12648.94
2009-01-31	Adventure Works	Adventure Works	TV and Video	194	51593.106	28146.4
2009-01-21	Fabrikam, Inc.	Fabrikam	Cameras and camcorders	282	163007.2	76709.45
2007-12-31	Adventure Works	Adventure Works	Computers	29	14008.43	7944.32
2007-06-22	Contoso, Ltd	Contoso	Cell phones	680	6107.24	3420.44
2007-06-22	Proseware, Inc.	Proseware	Computers	86	71417.6	30786.94
2007-08-23	Adventure Works	Adventure Works	Computers	43	22672.2	9954.6
2009-03-30	The Phone Company	The Phone Company	Cell phones	198	48500.37	24164.56
2008-03-24	Contoso, Ltd	Contoso	Cell phones	306	7353.594	3914.64
2007-09-30	Fabrikam, Inc.	Fabrikam	Home Appliances	44	4805.604	2824.24
2007-11-13	Adventure Works	Adventure Works	Computers	153	47357.97	28256.02
2008-12-06	Contoso, Ltd	Contoso	Computers	32	10790.4	6477.2
2007-11-14	Contoso, Ltd	Contoso	Cameras and camcorders	146	55397.5	25876
2009-12-30	Adventure Works	Adventure Works	Computers	32	15107.75	7952.97
2009-03-13	Wide World Importers	Wide World Importers	Audio	42	7990.92	3607.26
2009-08-11	Wide World Importers	Wide World Importers	Audio	9	1466.1	749.16
2009-09-28	Contoso, Ltd	Contoso	Home Appliances	78	9955.268	5189.27
2008-02-18	A. Datum Corporation	A. Datum	Cameras and camcorders	345	70989.93	32872.58
2007-08-15	Litware, Inc.	Litware	Home Appliances	69	112603.8	56472.35

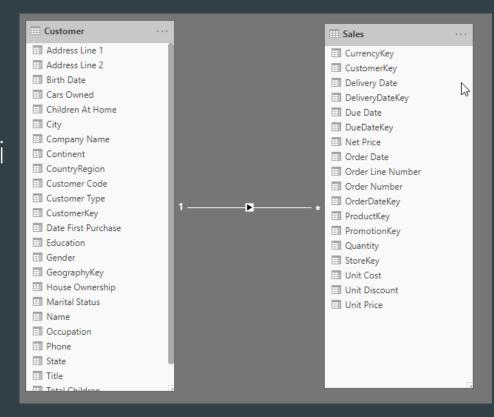
Demo 1

Abbiamo imparato che se il data model è fatto bene le nostre formule DAX saranno molto semplici ©

E il primo step per diventare modellatori di dati

Basi di modellazione: Business Entity

- Nella demo che abbiamo fatto la tabella dei cutomer è vista come una business entity
- Tutti i campi della customer sono riferiti al concetto di customer a differenza della tabella delle vendite
- Ogni business entity DEVE avere una tabella per se
- Ogni business ha differenti entità
 - Products, Customer, Resellers
 - Patients, Medications, doctor
 - Claim, Customer
- Ogni business entity ha una caratteristica unica



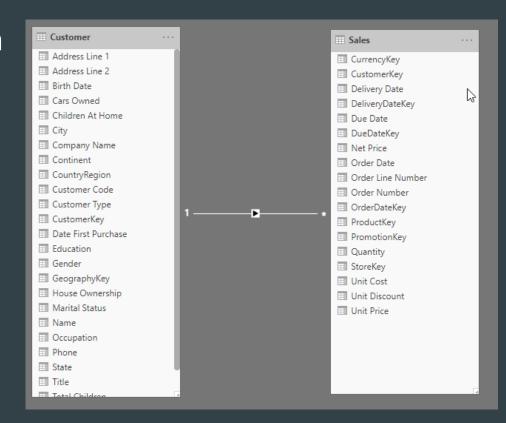
Basi di modellazione: Business Entity

- Perché il campo **Education** di **Customer** non è una tabella?
- Education non è una business entity ma è un attributo del customer. Customer è la tabella rilevante
- Definire che cosa è un attributo e cosa è una entity non è subito facile ma viene dall'esperienza e dalla pratica dipende dagli scenari ad esempio:
 - Education è possibile vederla come attributo di un customer in certi scenari mentre in altri potrebbe essere vista come tabella a se per rappresentare una business entity

$$1 ENTITY = 1 TABLE$$

Basi di modellazione: Business Entity

- Quando si hanno più tabelle in un modello per forza di cose bisogna unirle con una relazione
- Tabelle senza relazioni in un modello non servono a nulla (raramente)
- Una relazione puà avvenire tra due tabelle
- Una relazione ha una direzione:
 - Customer lato uno
 - Sales: lato molti
- <u>Best Pratice</u>: Dare lo stesso nome di colonna tra le tabelle che devo mettere in relazione



Basi di modellazione: Granularity

- Quando si hanno più tabelle in un modello la granularità è differente
- Ogni tabella ha la sua granularità non come nel tabellone unico
 - Customer ha una granularità a livello di customer
 - Data ha una granularità a livello di data
 - Product ha una granularità a livello di prodotto
- Le vendite (Sales) hanno una granularità definitia dalle relazioni entranti
 - Customer, Date e product level
 - Se abbiamo entrambe le tre tabelle

Tipi di modelli di dato

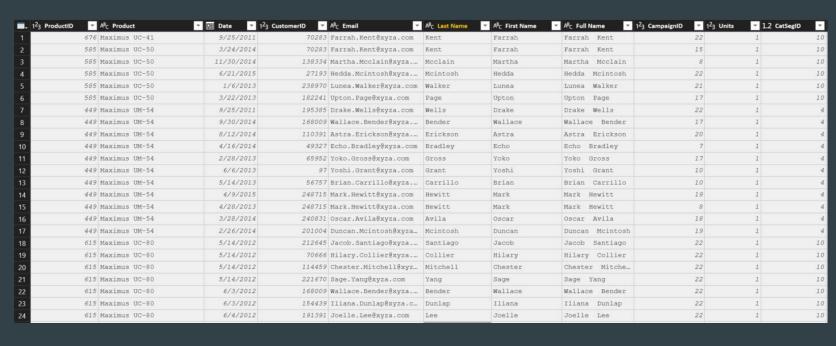
Il modello dati unisce fatti e dimensioni



Nota: Questo non è un elenco esaustivo, ma sono i tipi di modello più comuni utilizzati da Power Bl.

Tipi di modelli di dato: Flat Schema

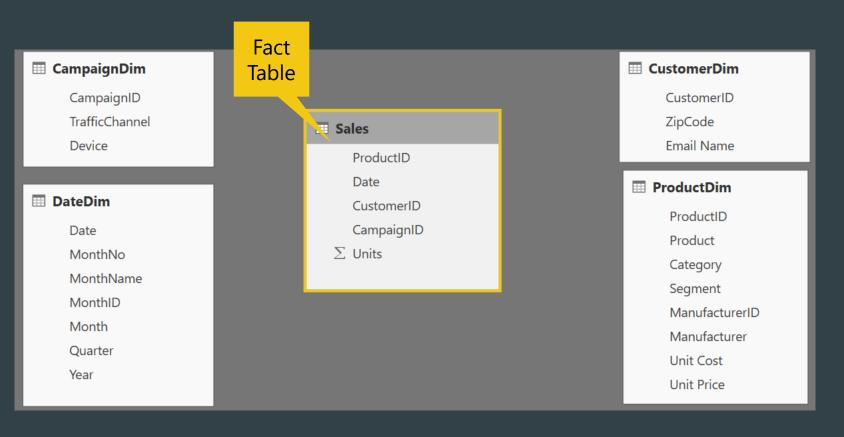
Flat or Denormalized schema ("Kitchen sink" schema) evitaimolo!!!



- Tutti gli attributi per il modello esistono in una singola tabella
- Altamente inefficiente
- Il modello ha copie extra di dati e prestazioni lente
 - Le dimensioni di una tabella piatta possono esplodere molto rapidamente quando il modello di dati diventa complesso

Tipi di modelli di dato: Star Schema

Componenti di un modello di dati – Fact Table (Tabella dei fatti)



Fact Table

- Contiene le misure (oggetti che posso aggregare) di un processo di business
- Esempi:
 - Transactions
 - Sales Revenue
 - Units
 - Cost
- Le misure sono normalmente filtrabili Esempio: By Month, By Customer

Tipi di modelli di dato: Star Schema Componenti di un modello di dati – Dim Table



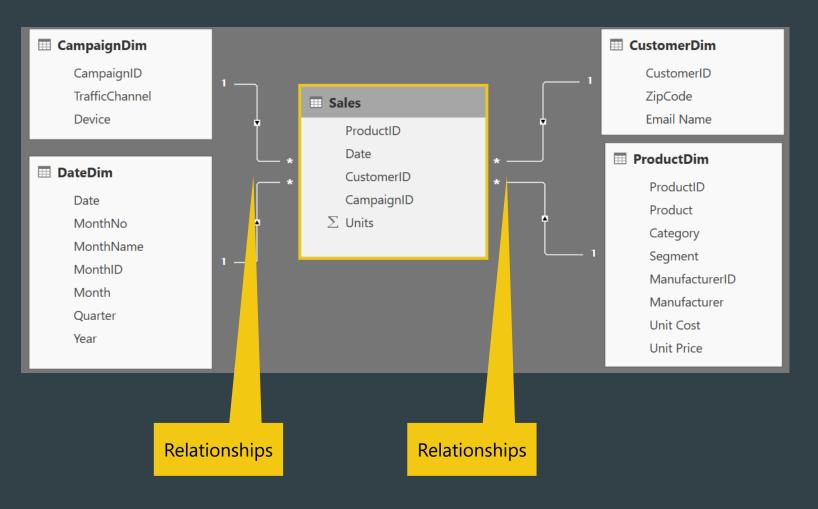
Dim Table

Una Dim (o Dimensione)
 La tabella contiene
 attributi descrittivi che
 definiscono come un
 fatto dovrebbe essere
 aggregato

Esempio:

By month, By Customer, By Geo

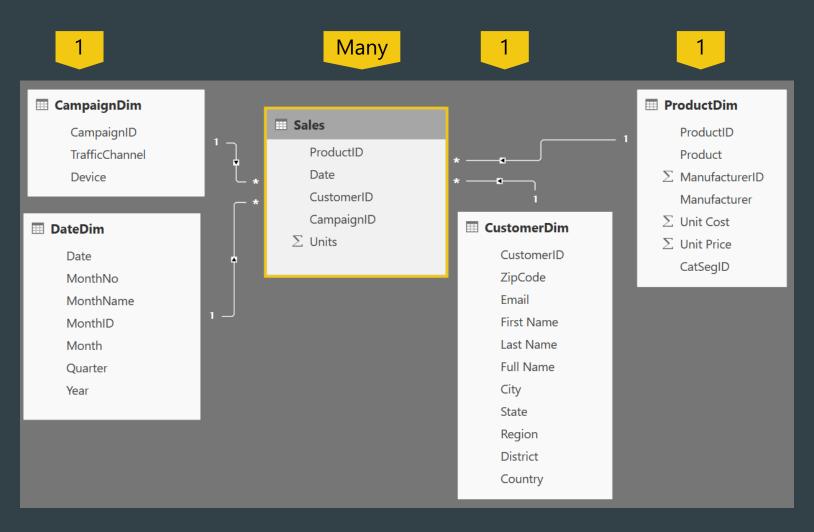
Tipi di modelli di dato: Star Schema Componenti di un modello di dati - Relazioni



Relazioni

- La connessione tra due 2 tables (normalemente i fatti e le tabelle delle dimensioni) usando una colonna fra di loro
- 3 tipi di relazioni
 - 1 to Many
 - 1 to 1
 - Many to Many (Attenzione!!! Utilizzare una bridge table)

Tipi di modelli di dato: Star Schema **Star Schema**



- Tabella dei fatti al centro
- Circondato da dim
- Sembra una "stella"
- La tabella dei fatti è il lato "Molti" della relazione (da una a molte)

Tipi di modelli di dato: Recap Fatti

Note importanti sulle nostre tabelle dei fatti

- Le tabelle dei fatti sono tipicamente molto grandi
 - Possiamo avere bilioni di vendite
 - Ma è diverso da avere bilioni di cutomers
- I fatti sono associati a numeri
 - La quantità di vendite
- Non tutti i numeri appartengono alla tabella dei fatti
 - Prezzo di un prodotto sta nella tabella del prodotto (dimensione)
 - Quantità di prodotti venduti appartiene alla tabella dei Fatti
- Le misure dipicamente sono aggregazioni sulle tabelle SUM(Sales[Quantity])

Tipi di modelli di dato: Recap Dimensioni

Note importanti sulle nostre tabelle delle dimensioni

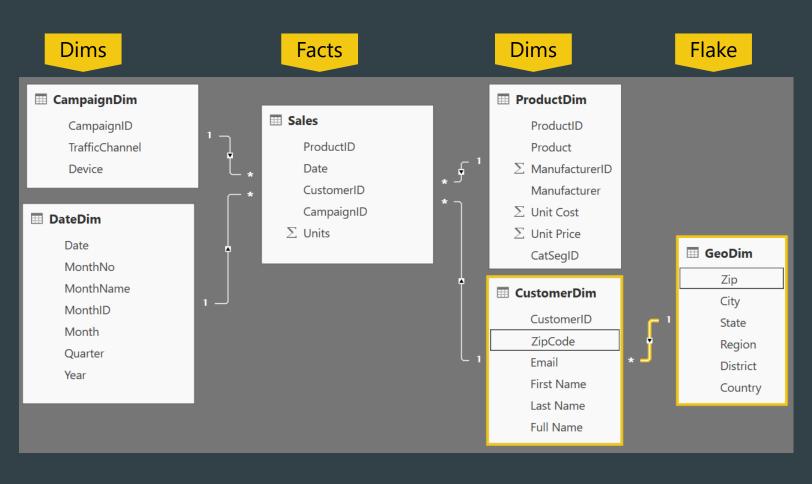
- Le dimensioni tipicamente sono piccole tabelle
 - 1 milione di righe per una dimensione è già un numero grande
- Le dimensioni contengono molte stringhe
 - Nome del customer, indirizzo ecc...
 - Colore del prodotto e dimensione e ccc
- Sono tipicamente quelle su cui si effettua lo slice
- Quando dobbiamo calcolare in genere le usiamo per conteggiare su di loro
 - Numero di clienti
 - Numero di prodotti
- Le dimensioni non sono relazionate ad altro.

Basi di Data Modeling

La differenza tra le tabelle dei fatti e le dimensioni sono il cuore di uno star schema in un modello di Bl

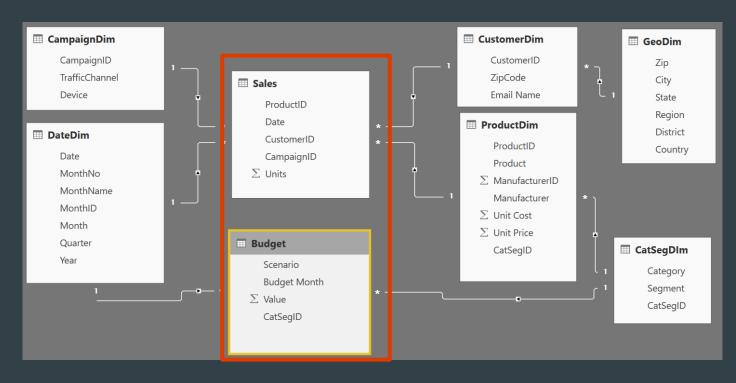
Tipi di modelli di dato: Snowflake

Snowflake Schema (evitiamolo perchè richiede più attenzioni © !!! Denromalizziamo tutto in una dimensione)



- Il centro è uno schema a stella
- Tabella dei fatti al centro
- Circondato da dim
- Dim "fiocco di neve" collegate ad altre dim
- Se ne hai molti, sembra un "fiocco di neve"
- Le tabelle Dim o Fact possono essere il lato "Molti" della relazione

Basi di Data Modeling: Diverse Granularità Granularità & Fatti multipli



SalesFact (Giornaliera per prodotto)

BudgetFact (Mensile per categoria di prodotto e segment di prodotto)

- Grain (**granularità)** misura il livello di dettaglio della tabella
- Esempio:
 una riga per order o per oggetti
 per giorno o mese è il grain
- Se I tuoi fatti hanno differenti granularità, dovete dividerle in differenti tabelle dei fatti e connetterle tra loro da dimensioni condivise (Shared) al livello di granularità più basso di ogni singola tabella dei fatti

Basi di Data Modeling: Quante dimensioni

Un'entità = una dimensione

• Più dimensioni sono spesso inutili e confuse

Regole empiriche:

- Numero di dimensioni
 - 10 dimensioni sono buone
 - 20 dimensioni sono probabilmente errate
 - 30 dimensioni sono totalmente errate
- Rapporto dimensione / attributo:
 - 1 dimensione con 20 attributi è buona
 - 20 dimensioni con 1 attributo non sono valide

Basi di Data Modeling

Quante tabelle dei fatti?

Multipli eventi uniti assieme

Order No	Customer	Date	Туре	Amount
1	Contoso	01/01/2017	Order	100.00
1	Contoso	01/15/2017	Ship	60.00
1	Contoso	01/15/2017	Ship	40.00
2	lmag@ware	01/16/2017	Order	90.00
3	Imageware	01/20/2017	Order	45.00
4	Contoso	01/22/2017	Order	25.00

Modello corretto

Orders						
Order No	Customer	Orde	r Date	Amount		
1	Contoso	01/01/2017		100.00		
2	Imageware	01/16	5/2017	90.00		
3	Imageware	01/20/2017		45.00		
4	Contoso	01/22/2017		25.00		
Shipments						
Order No	Date	Am	ount			
1	01/15/2017		60.0	0		
1	01/15/2017		40.0	0		

Perchè scegliere il giusto modello?

Se il modello non è quello giusto

- Il codice DAX tende ad essere molto complesso
- Le formule sono difficili da pensare
- La complessità si trasforma in problemi di prestazioni

Con il modello corretto

- Il codice DAX è semplice, come dovrebbe essere
- Le prestazioni sono eccezionali







Compiti del data modeller

Cosa significa data modeling

- Conoscere diversi schemi
- Essere in grado di abbinare il modello a un pattern
- Applicare il pattern
- Adattare le piccole differenze che appaiono in un modello custom
- I pattern si imparano con l'esperienza e questo corso non copre questi argomenti

Il tuo (business/modello) è diverso?

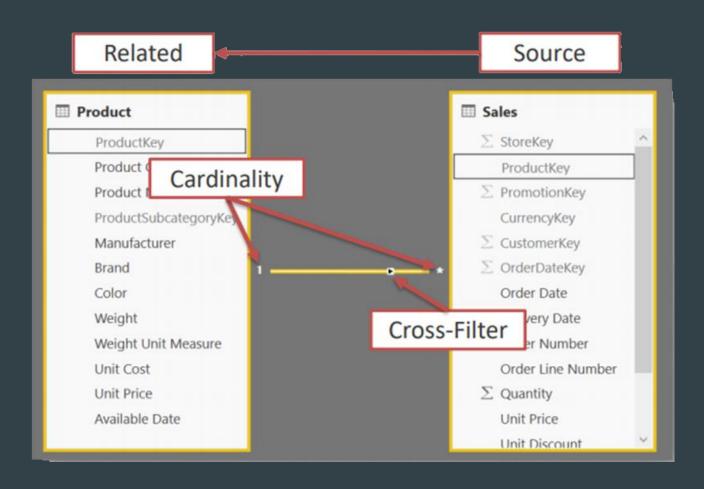
All'inizio, senti sempre il tuo modello diverso da quelli standard

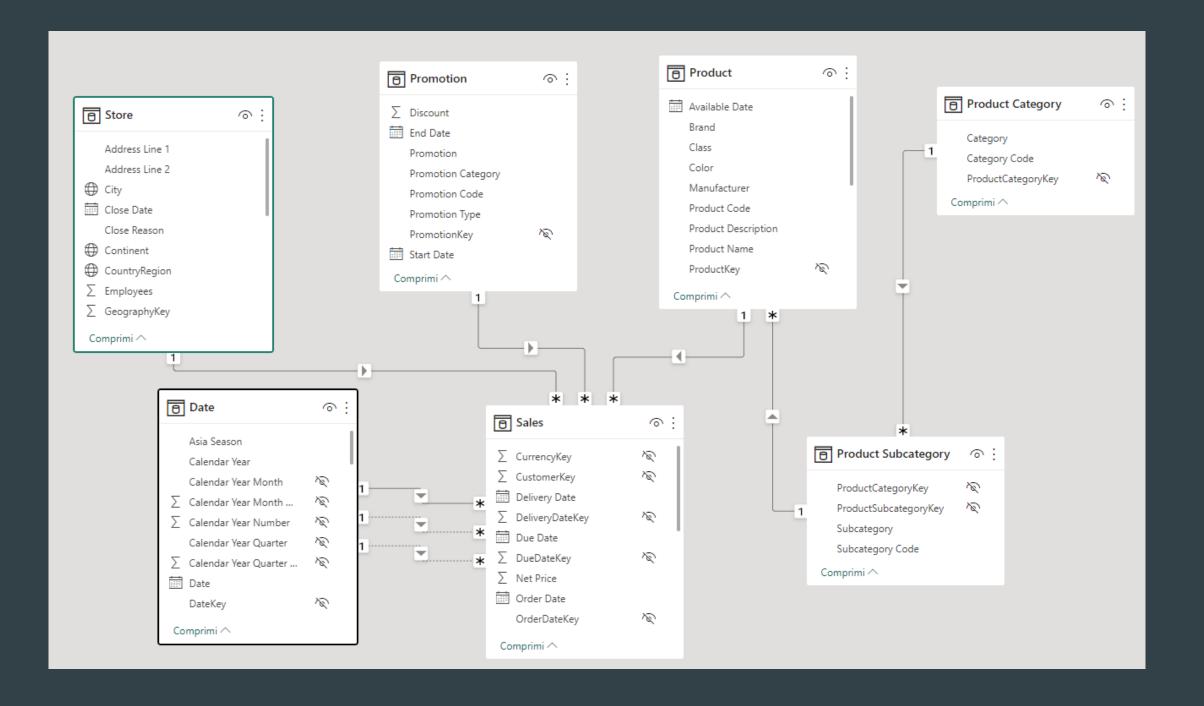
ma ricordati che il 99,9% delle volte, non è così

Non deviare dalla modellazione standard, a meno che tu non sai davvero cosa stai facendo

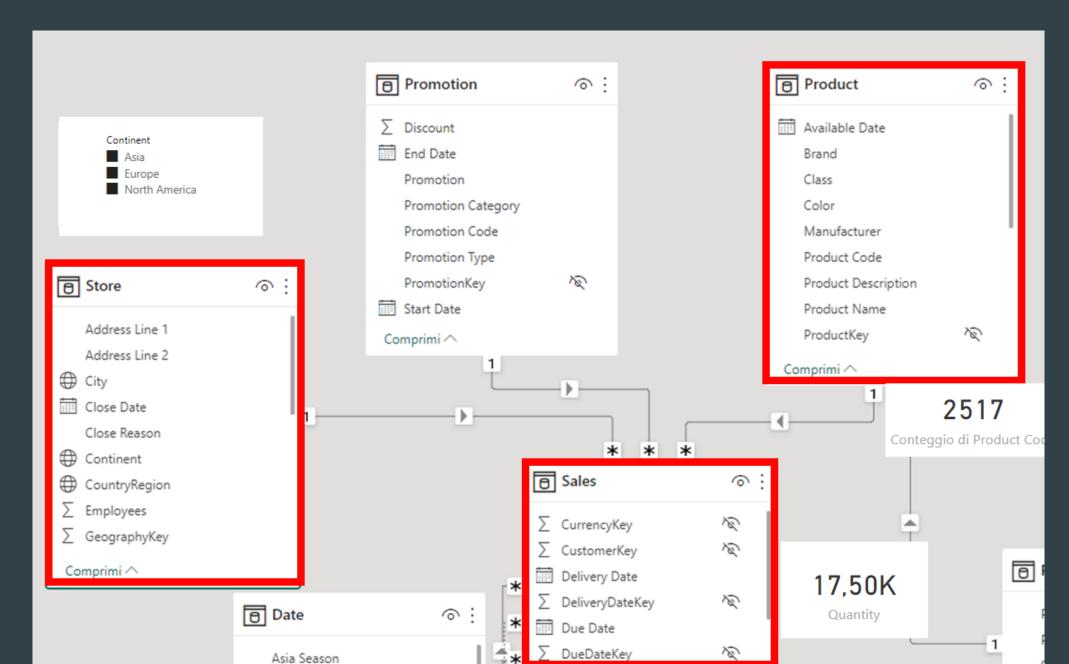
«La Business Intelligence è nata nel 1958 e in 60 anni abbiamo analizzato quasi tutti i modelli esistenti Abbiamo scoperto che gli schemi a stella sono l'opzione migliore»

Che cosa sono le relazioni in PBI Desktop?

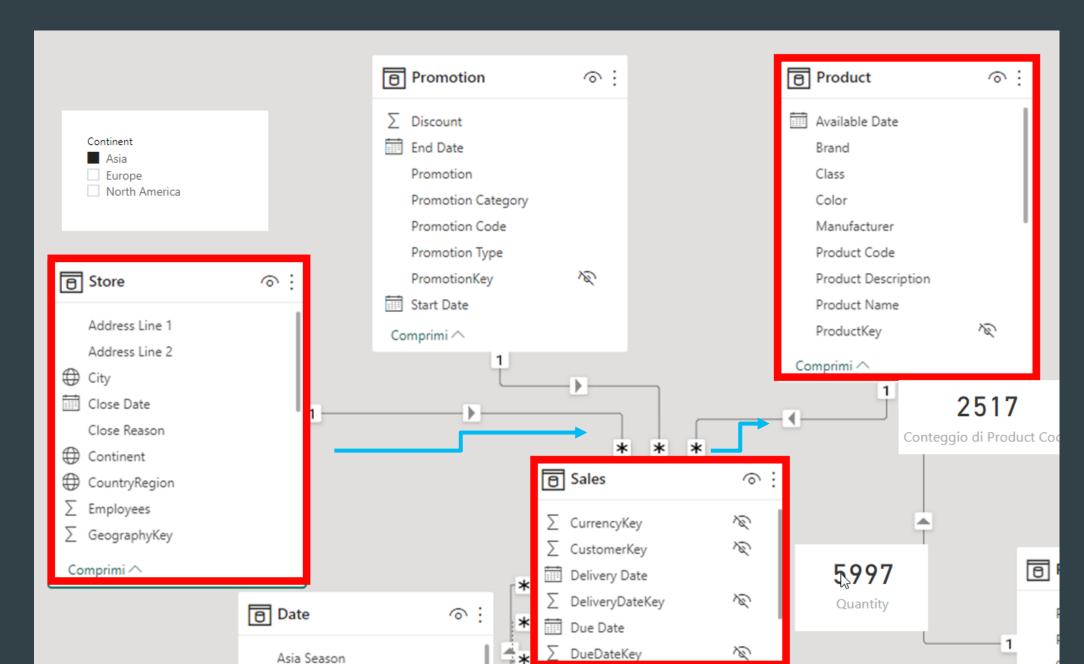




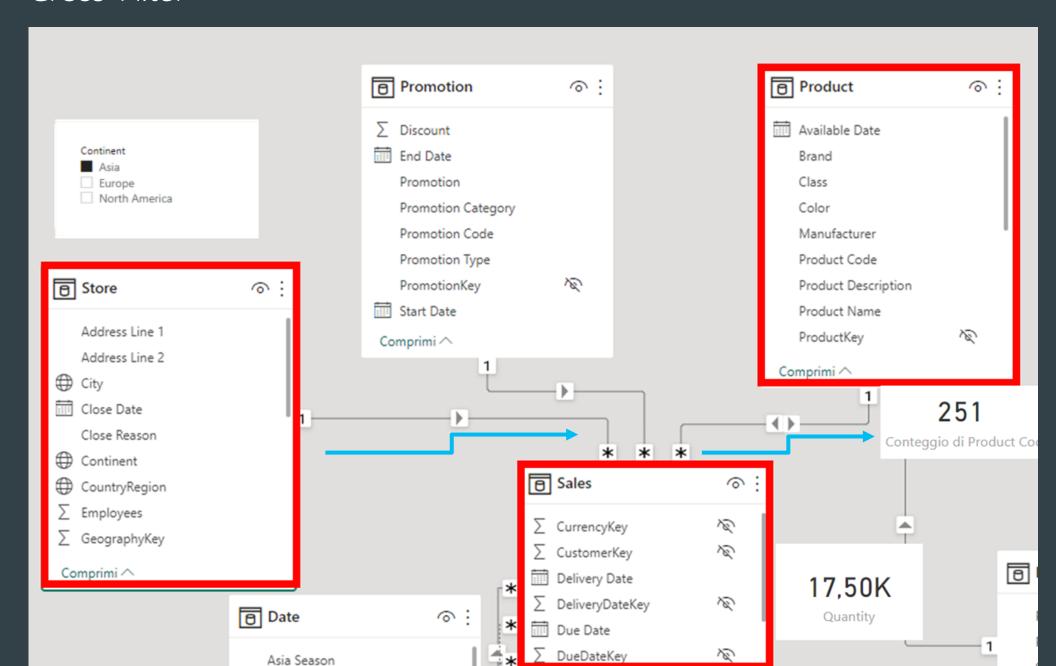
Cross-Filter



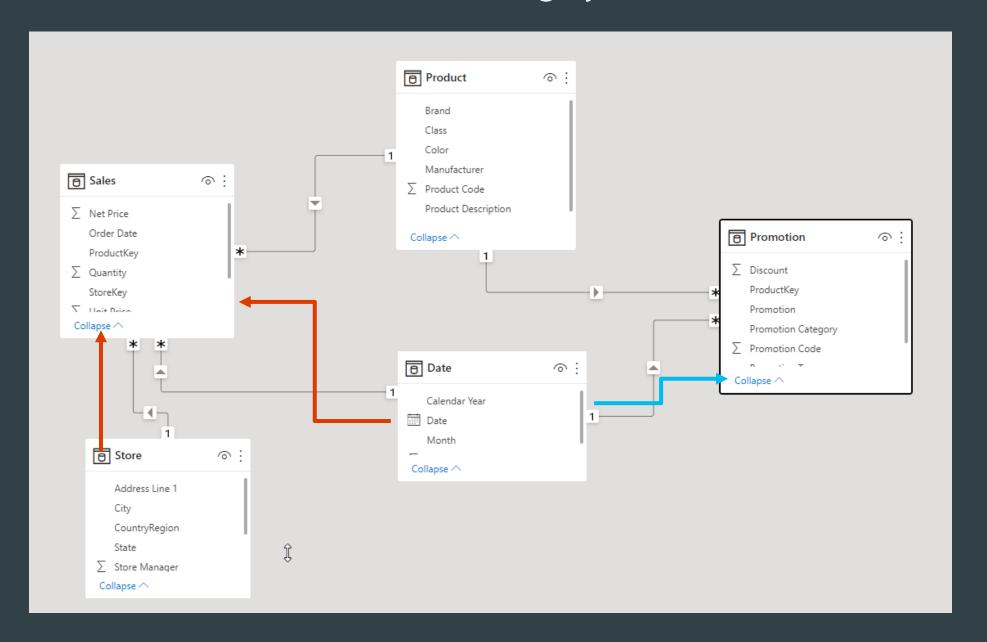
Cross-Filter



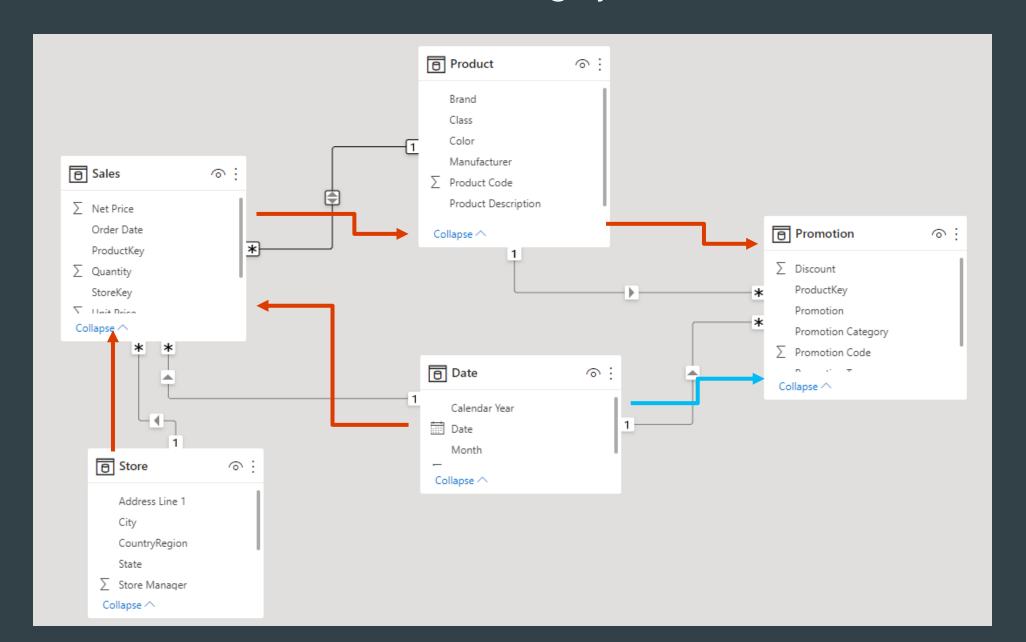
Cross-Filter



Attenzione alla bidirezionalità - Ambiguty



Attenzione alla bidirezionalità - Ambiguty



Cardinalità di una relazioni in PBI Desktop

- Si riferisce alla relazione tra due tabelle
- One
 - Colonna che necessita di valori univoci
 - E' il target di una tabella espansa (concetto avanzato)
 - E' la sorgente della propagazione del filter context
- Many
 - Colonna che potrebbe contenere dupplicati
 - Sorgente della tabella espansa (concetto avanzato)
 - E' il target della propagazione del filter context

Cardinalità di una relazione in PBI Desktop

One-to-many

- La relazione più importante
- (*:1): Questo è il tipo predefinito e più comune. La prima tabella presenta più istanze del valore di join. La seconda tabella ha un'istanza del valore;
- (1:*): Questa è la relazione inversa del tipo Many to One.

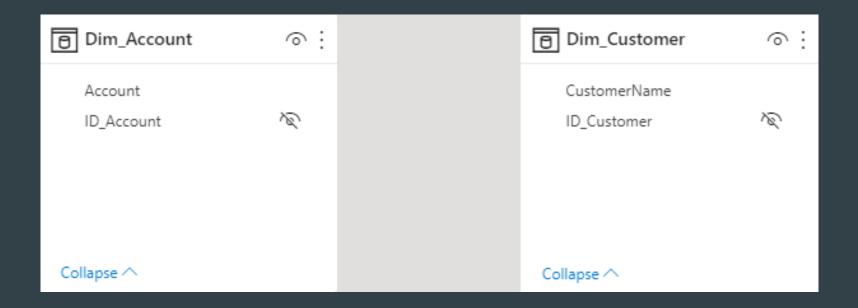
• One-to-one (1:1):

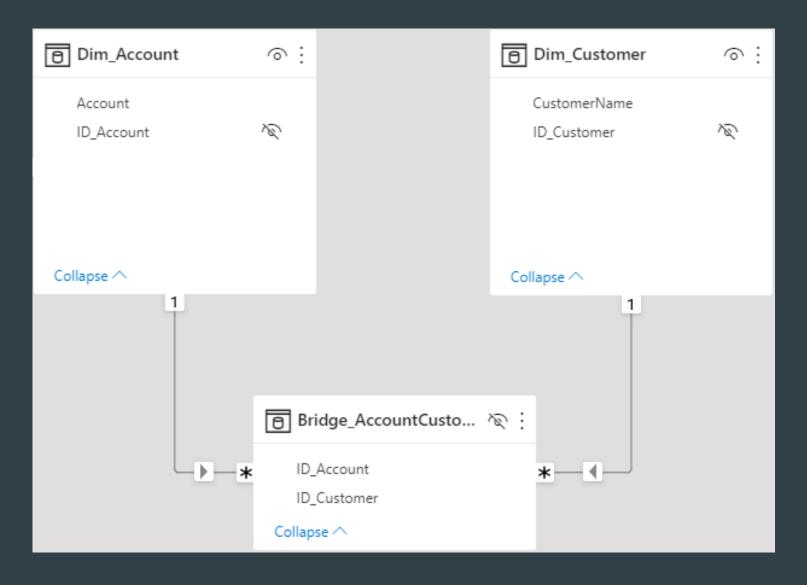
- Meno comune di Molti a uno, in quanto solo una istanza del valore esiste nelle due tabelle correlate.
- L'espansione (blank row viene forzata) avvine in entrambe le direzioni (concetto avanzato)
- Il Cross-filter necessita di essere di tipo both

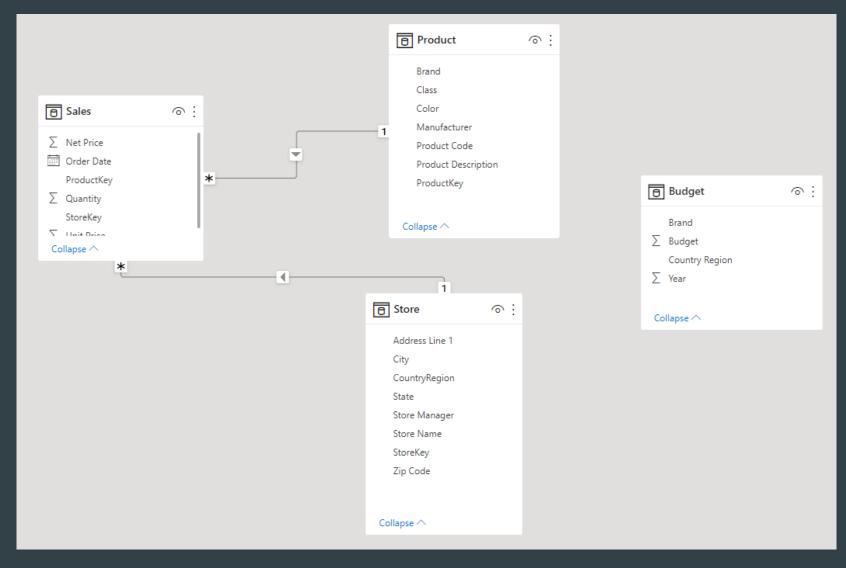
Cardinalità di una relazione in PBI Desktop

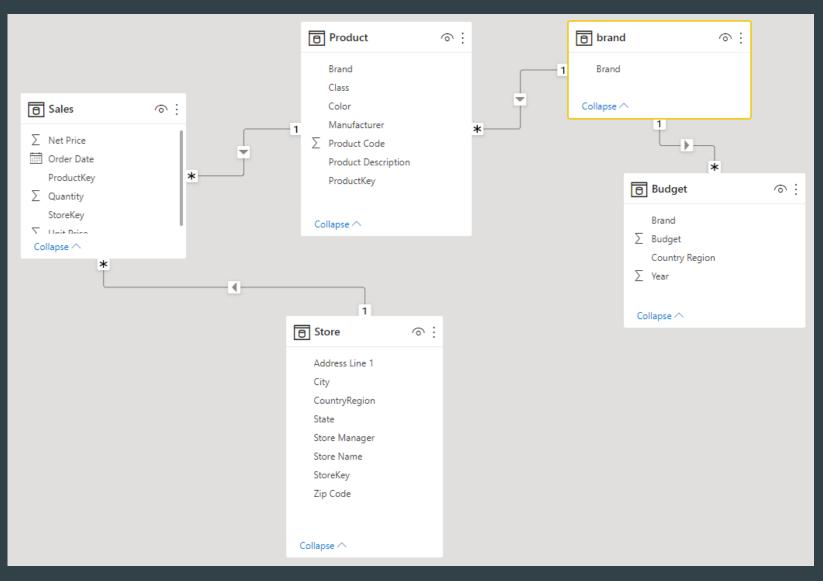
- Many-to-Many (*:*) Weak relationships!!!!
 - Nuova e pericolosa
 - Non c'è espansione di tabella (blank row non viene forzata)
 - Necessita di decidere il cross-filter direction
 - Presente da Ottobre 2018 (Attenzione non confondete con la many-tomany tra dimensioni!!!)
 - Implementabile con un pattern (*:1) + (1:*)

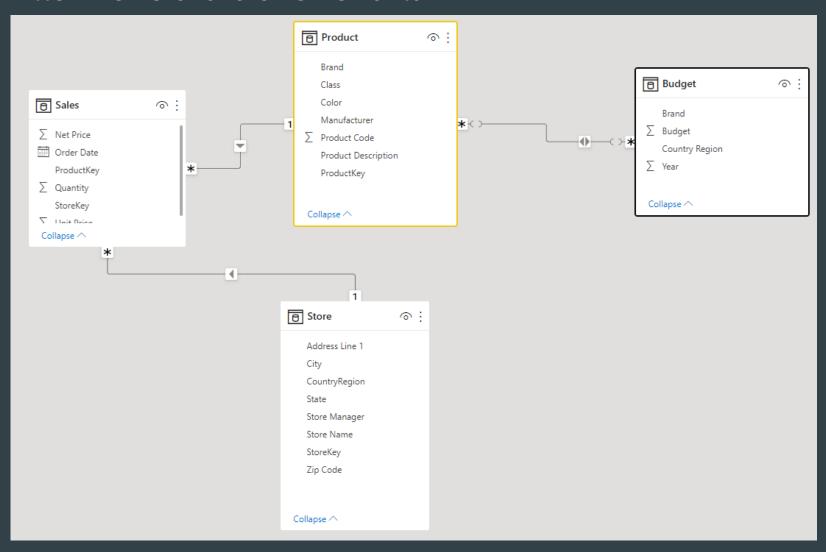
NON E' QUELLO CHE SEMBRA ©



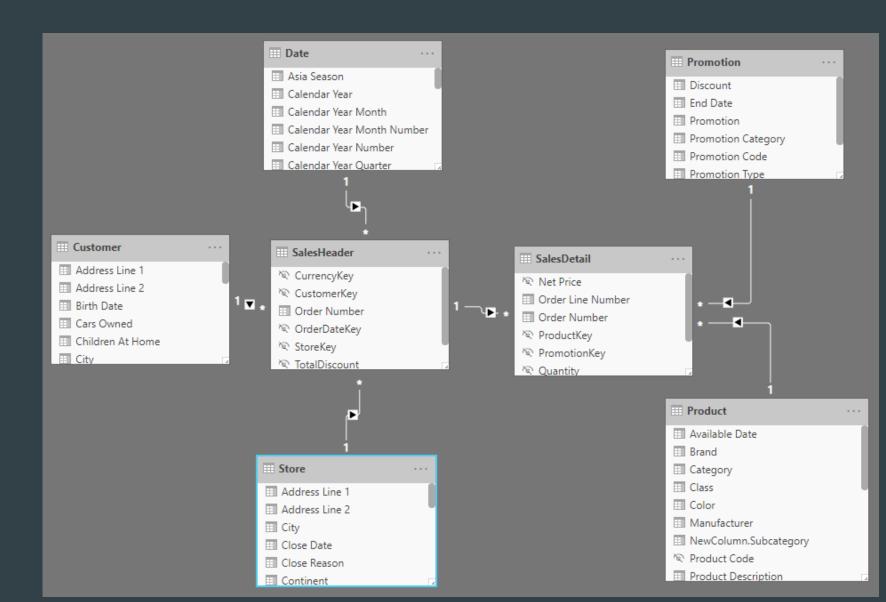




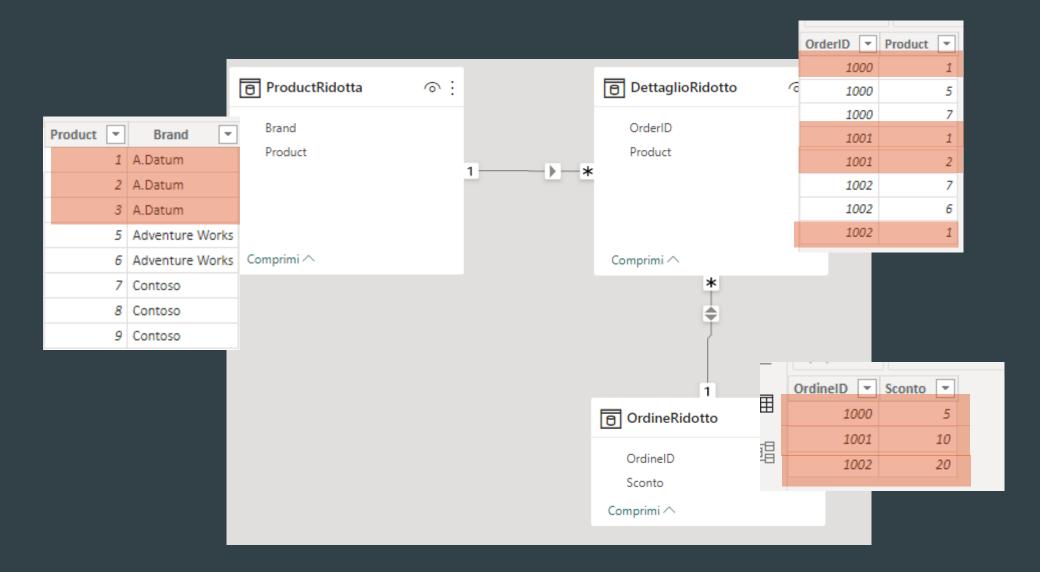




Demo 3 – Caso di modellazione



Demo 3 – Caso di modellazione



Domande?