

Vendite al dettaglio (studio di caso)

Luca Cabibbo
marzo 2010

Studi di caso

Le tecniche di modellazione dimensionale vengono illustrate mediante un certo numero di esempi (studi di caso)

- ciascuno studio di caso è relativo ad un esempio concreto
- ciascuno studio di caso introduce (o rafforza) una o più tecniche di modellazione dimensionale
- ciascuna tecnica può essere usata in più contesti
 - anche diversi da quelli dello studio di caso in cui è stata presentata

Progettazione dimensionale

Preliminarmente, introduciamo una semplice metodologia per la progettazione di uno schema dimensionale

- uno schema dimensionale è composto da una singola tabella fatti e da un insieme di tabelle dimensione

Bisogna però tenere presente che

- un data mart può essere composto da più schemi dimensionali
- un data warehouse è normalmente composto da più data mart
- successivamente sarà introdotta una metodologia per la progettazione di un data warehouse dimensionale completo, secondo l'architettura a bus

Progettazione di uno schema dimensionale

La progettazione di uno schema dimensionale richiede lo svolgimento di quattro passi

- scelta del **processo di business** da modellare
- scelta della **grana** del processo
- scelta delle **dimensioni** da cui dipende ciascun record della tabella fatti
- scelta dei **fatti** misurabili che popoleranno ogni record della tabella fatti

Queste scelte devono essere guidate

- dai requisiti
- dalle sorgenti informative disponibili

Progettazione di uno schema dimensionale

Scelta del **processo di business** da modellare

- *quale processo stiamo modellando?*
- per processo si intende un processo operativo, supportato da uno o più sistemi operazionali, i cui dati possono essere utilizzati per popolare lo schema dimensionale
- esempi di processi
 - vendite
 - ordini
 - fatturazione
 - magazzino/inventario
 - approvvigionamento
 - ...

Progettazione di uno schema dimensionale

Scelta della **grana** del processo

- *che cosa descrive una singola riga della tabella fatti?*
- per grana si intende il livello di dettaglio atomico che deve essere rappresentato nella tabella fatti per il processo
- esempi di livelli per la grana
 - transazioni individuali
 - istantanea (snapshot) periodica individuale – ad es., giornaliera o mensile
 - istantanea accumulata – ad es., stato delle consegne
 - ...

Progettazione di uno schema dimensionale

Scelta delle **dimensioni** da cui dipende ciascun record della tabella fatti

- *in che modo sono descritti i dati di interesse per il processo che stiamo modellando?*
- una dimensione è un insieme di membri, caratterizzati da un certo numero di attributi – da usare nelle selezioni e nei raggruppamenti
- esempi di dimensioni
 - tempo, prodotto, cliente, promozione, magazzino, tipo di transazione, stato, ...
- la scelta della grana stabilisce, evidentemente, alcune delle dimensioni
 - come vedremo, è possibile scegliere anche delle dimensioni aggiuntive

Progettazione di uno schema dimensionale

Scelta dei **fatti** misurabili che popoleranno ogni record della tabella fatti

- *che cosa stiamo misurando?*
- i fatti sono misure (solitamente numeriche, continue e additivi) del processo selezionato
- esempi di fatti
 - quantità venduta, incasso della vendita, ...

Progettazione di uno schema dimensionale

La progettazione di uno schema dimensionale è guidata da due forze fondamentali

- le esigenze di analisi da parte degli utenti/analisti del data warehouse (top down)
- le sorgenti di dati disponibili (bottom-up)

esigenze aziendali di analisi



Modellazione dimensionale

1. processo di business
2. grana
3. dimensioni
4. fatti



disponibilità di dati

- entrambe le forze devono essere tenute in considerazione

Il processo delle vendite

Si consideri il seguente caso di studio, relativo al **processo delle vendite al dettaglio** in una catena di negozi alimentari (**grocery store**)

- il contesto è la direzione di una grande catena di alimentari (negli Stati Uniti)
- la catena comprende 500 grandi negozi di alimentari, distribuiti in tre stati
- ogni negozio è un supermercato con diversi reparti (department)
 - ad esempio, drogheria, surgelati, latticini, carne, frutta e verdura, pane, pasta, fiori, liquori, ...

Il processo delle vendite (2)

- la catena di negozi vende circa 60.000 tipi di prodotti individuali
 - i prodotti individuali sono chiamati **unità di vendita (SKU, stock keeping unit)**
 - ad esempio, una SKU è una lattina di Diet Coke
 - ogni variante di confezionamento dei prodotti costituisce una diversa SKU
- circa 40.000 delle SKU vengono da fornitori esterni, e su di esse è stampato un codice a barre chiamato **codice universale del prodotto (UPC, universal product code)**
 - la grana degli UPC è la stessa delle SKU
- le altre 20.000 SKU corrispondono a prodotti come frutta e carne, che non sono confezionati o che sono confezionati localmente, e non hanno UPC
 - anche a questi prodotti è associato un codice SKU

Il processo delle vendite (3)

Dove vengono raccolti i dati della catena di negozi alimentari?

- per i dati relativi alle vendite, sicuramente in ciascuna cassa, mediante dei sistemi **POS** (point of sale)
- per quanti riguarda gli acquisti
 - alcuni negozi usano un sistema di codici a barre anche alla consegna delle merci
 - altri negozi non registrano le merci consegnate
 - ma vengono raccolte le bolle e le fatture
- l'inventario è spesso realizzato girando tra gli scaffali e guardando quali prodotti sono assenti

Il processo delle vendite (4)

La direzione della catena di occupa della logistica delle ordinazioni, della disposizione delle merci sugli scaffali, della vendita dei prodotti e della massimizzazione del profitto

- sorgenti del profitto
 - fissare per i prodotti il prezzo più alto possibile
 - ridurre i costi di acquisizione dei prodotti e le spese generali
 - attrarre quanti più clienti è possibile
- le scelte sotto il controllo della direzione della catena di negozi riguardano
 - i prezzi dei prodotti
 - le promozioni

Il processo delle vendite (5)

Le promozioni comprendono

- riduzioni temporanee di prezzo (TPR)
- pubblicità (su diversi media)
- esposizione sugli scaffali
- esposizione alla fine dei corridoi

Uno degli obiettivi della direzione è la comprensione dell'impatto delle promozioni sulle vendite e, quindi, sui profitti

- per comprendere l'impatto delle promozioni passate
- per pianificare e progettare le promozioni future

Il data mart delle vendite

La progettazione di un data warehouse (e di ogni singolo schema dimensionale che lo compone) è basata sulla comprensione del processo e dei dati effettivamente disponibili

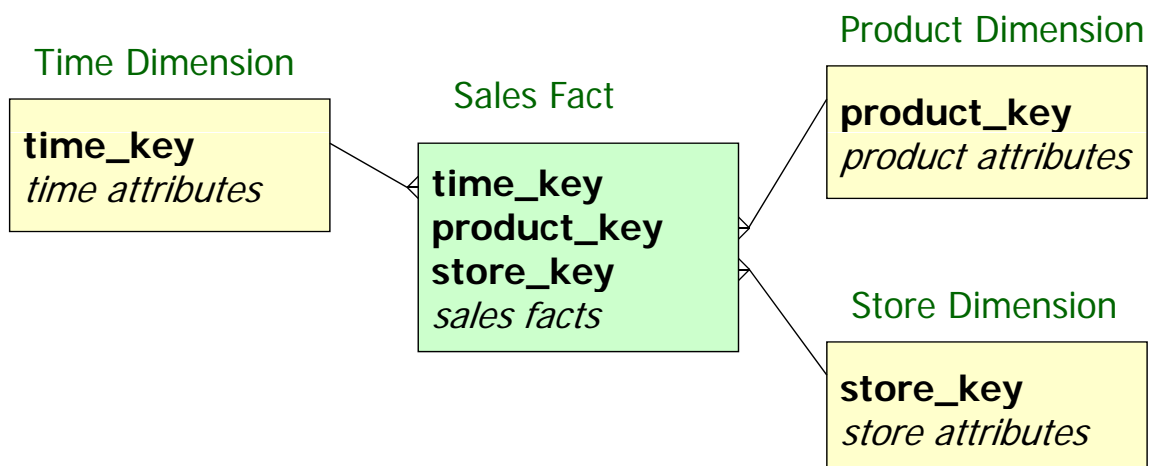
Il data warehouse della catena di negozi alimentari riguarda il processo delle vendite dei prodotti nei negozi

- scelto il processo, sono comunque possibili diverse scelte per il data mart delle **vendite dei prodotti**

Scelta della grana

Per il data mart per il processo delle vendite sono possibili diverse scelte per la grana

- ad esempio, unità di vendita (SKU) per negozio per giorno



Scelta della grana

La scelta della grana ha influenza su

- scelta delle dimensioni del data mart
- occupazione di memoria del data mart
- tipi di analisi che possono essere effettuati – ma anche su quelli che non possono essere effettuati

Alcune possibili scelte per la grana

Alcune possibili scelte per la grana

- unità di vendita (SKU) per negozio per giorno
- unità di vendita per negozio per settimana
 - non permette di distinguere le vendite nei fine settimana da quelle degli altri giorni
- unità di vendita (SKU) per negozio per giorno per promozione
 - consente di analizzare l'impatto delle promozioni sulle vendite

Alcune possibili scelte per la grana

Alcune possibili scelte per la grana

- per voce di vendita (transazione individuale)
 - una riga nella tabella fatti per ciascuna voce (riga) di ciascuno scontrino di vendita
 - permette di effettuare interessanti analisi di market basket – ancor di più se è nota anche l'identità del cliente
 - l'occupazione di memoria è maggiore rispetto al caso precedente

Alcune possibili analisi

Ad esempio, se viene scelta la grana “unità di vendita per negozio per promozione per giorno”, allora è possibile effettuare, tra l'altro, le seguenti analisi

- è utile vendere più varianti di confezionamento di uno stesso prodotto?
 - possibile solo se la grana riguarda l'unità di vendita
- di quali prodotti diminuiscono le vendite a fronte della promozione di un certo altro prodotto?
 - possibile solo se la grana riguarda le promozioni
- quali sono i dieci prodotti più venduti dai miei concorrenti che invece la catena non vende?
 - sulla base di ulteriori dati forniti da società di analisi specializzate

Altre considerazioni sulle SKU

Nessuna delle analisi proposte è interessata esplicitamente alle singole SKU

- non è solitamente interessante presentare l'unità di vendita individuale nel risultato dell'analisi
- tuttavia, in un data warehouse è necessario memorizzare dati a una grana sufficientemente piccola, per permettere alle interrogazioni di selezionare e raggruppare i dati in modo sufficientemente preciso e mirato

Dimensioni primarie

Fissati il processo (vendite giornaliere dei prodotti) e la grana (ad esempio, unità di vendita per negozio per giorno) bisogna scegliere le dimensioni

Per alcune dimensioni la scelta è immediata

- si tratta delle **dimensioni primarie**, che fissano la grana della misurazione e sono tra loro indipendenti, poiché le misure relative ai movimenti giornalieri dei prodotti dipendono funzionalmente da queste dimensioni
- in questo caso, le dimensioni primarie sono tempo, prodotto e negozio

Dimensioni supplementari

La scelta di altre dimensioni è meno ovvia

- si tratta delle **dimensioni supplementari**, che dipendono funzionalmente dalle dimensioni primarie, nel senso che per ogni possibile combinazione delle dimensioni primarie è univoca la scelta del valore delle dimensioni supplementari
- ad esempio, una dimensione supplementare potrebbe essere la dimensione promozione – se la promozione dipende funzionalmente dalla data, dal prodotto e dal negozio

Se una dimensione supplementare non fosse conforme alla grana della tabella fatti (richiedendo maggior dettaglio nei dati) allora la scelta della grana dovrebbe essere corretta (e la dimensione potrebbe essere primaria)

Scelta delle dimensioni

Ulteriori possibili dimensioni primarie (non scelte in questo studio di caso perché non accessibili dalle sorgenti informative a disposizione)

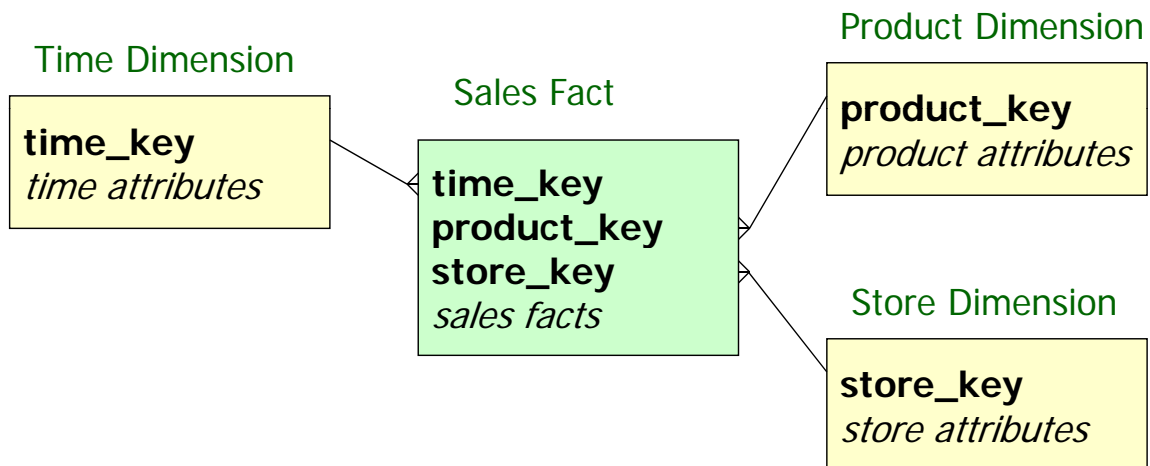
- il cliente che ha effettuato l'acquisto
- il tipo di pagamento

Altre ipotetiche dimensioni supplementari (non scelte perché non accessibili dalle sorgenti informative a disposizione)

- il fornitore che ha fornito il prodotto al negozio
- il responsabile delle vendite nel negozio nel giorno
- che tempo faceva in quel giorno in quella città

Schema dimensionale

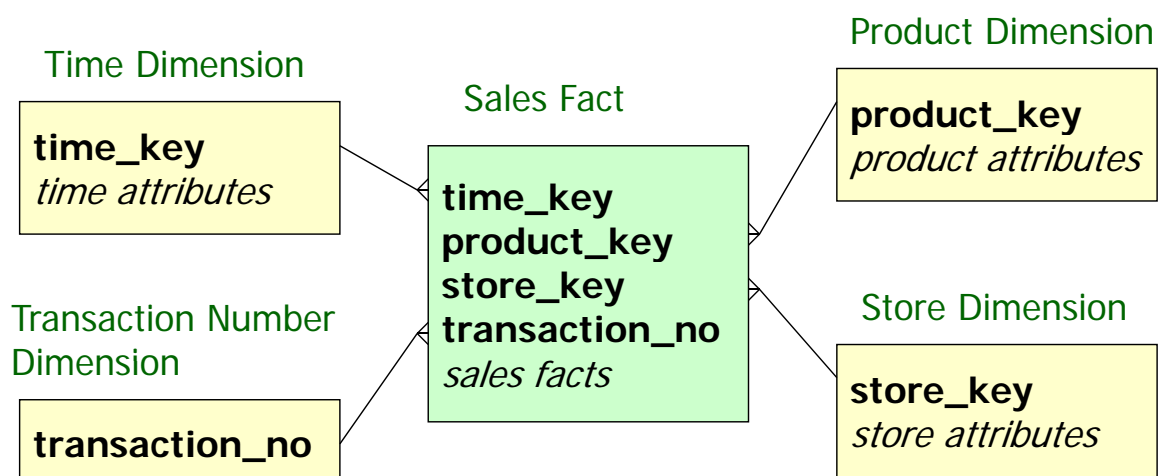
Data mart delle vendite giornaliere, per prodotto e negozio



- si tratta di una versione preliminare, la scelta degli attributi delle dimensioni verrà fatta più avanti

Schema dimensionale – varianti

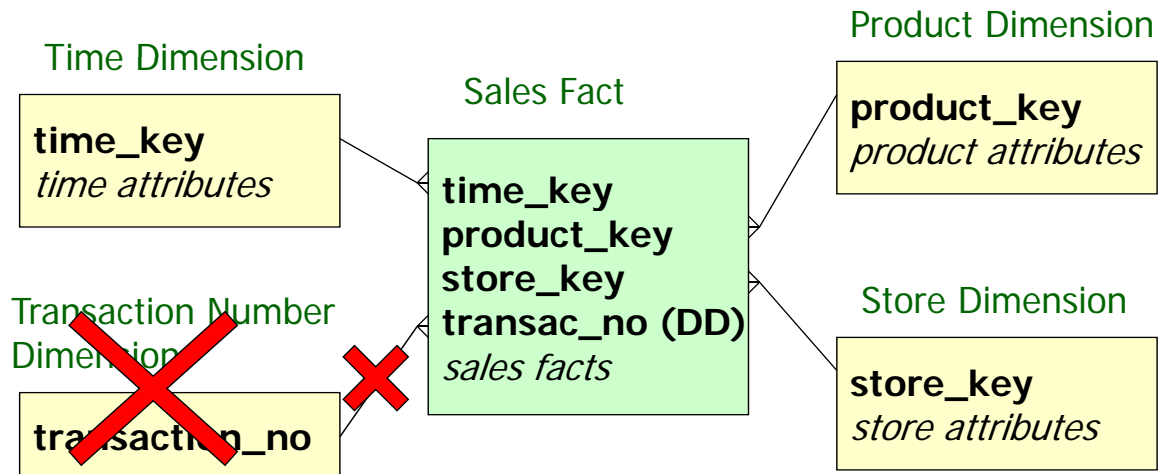
Data mart delle vendite, per riga di vendita, prodotto, negozio e giorno



Schema dimensionale – varianti

Data mart delle vendite, per riga di vendita, prodotto, negozio e giorno

- poiché la dimensione transazione (lo scontrino di vendita) non ha attributi, la relativa tabella dimensione non è strettamente necessaria – si parla in questo caso di **dimensione degenere**



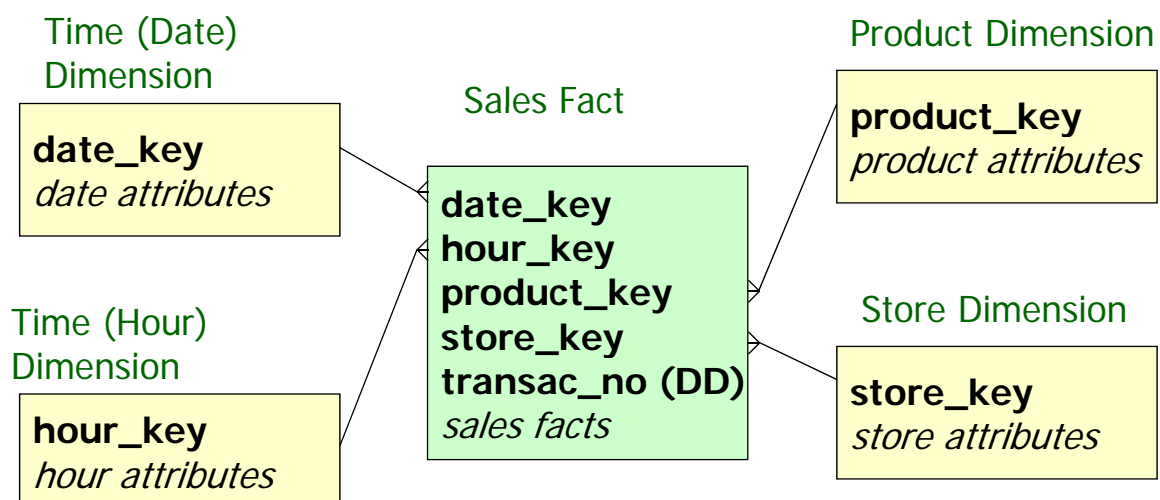
27

Vendite al dettaglio

Luca Cabibbo

Schema dimensionale – varianti

Data mart delle vendite, per riga di vendita, prodotto, negozio, giorno e orario di vendita



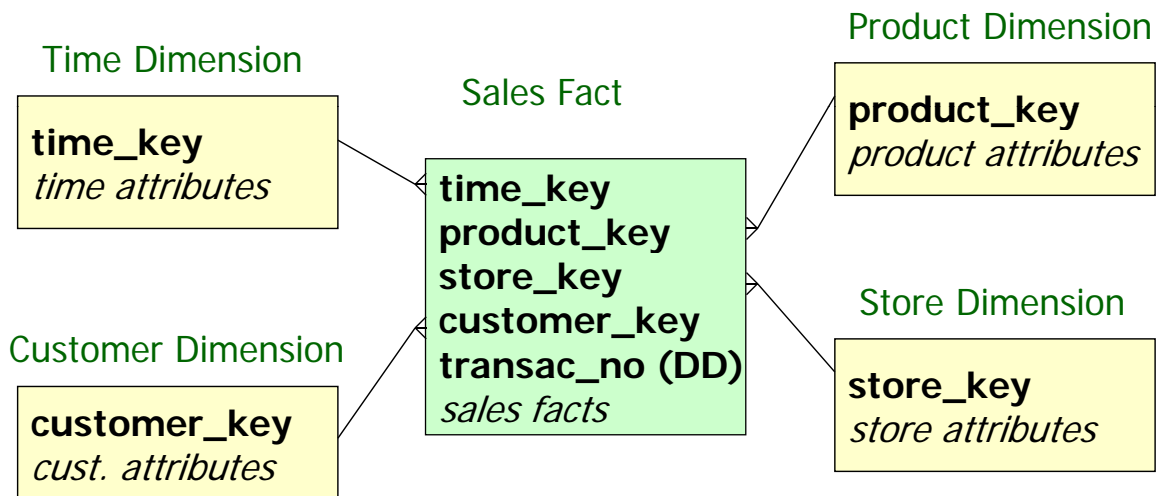
28

Vendite al dettaglio

Luca Cabibbo

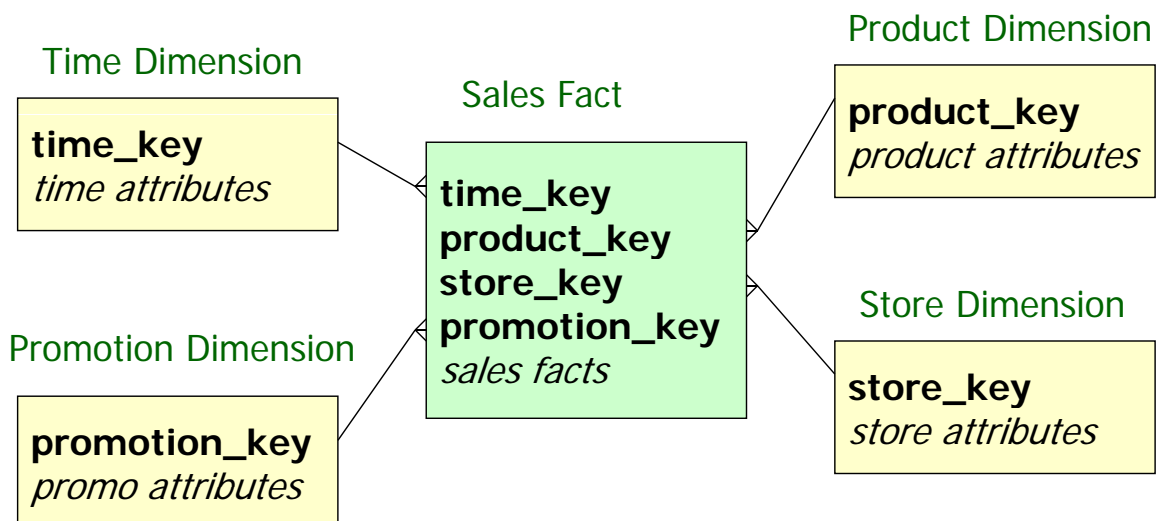
Schema dimensionale – varianti

Data mart delle vendite, per riga di vendita, prodotto, negozio, giorno e cliente



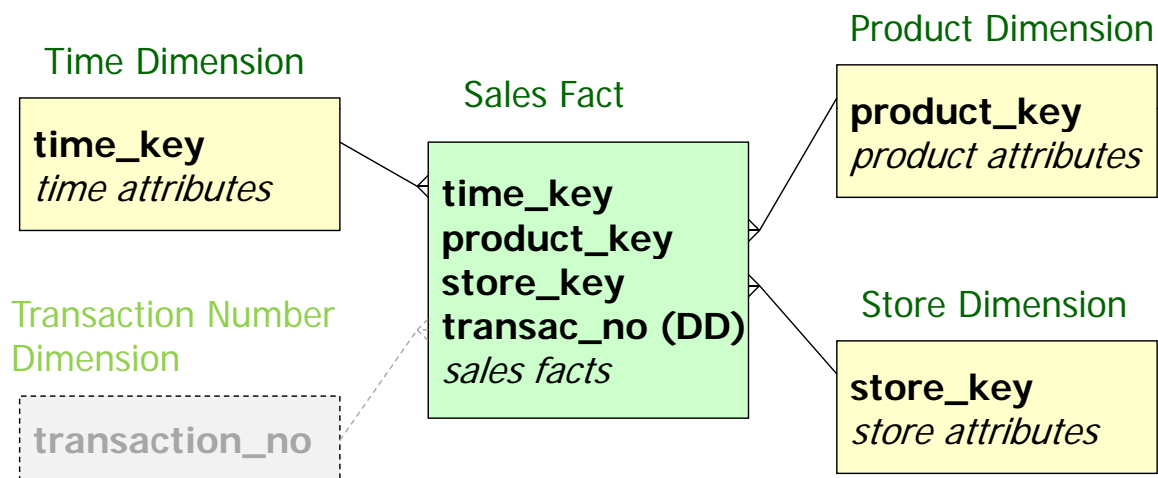
Schema dimensionale – varianti

Data mart delle vendite giornaliere, per prodotto, negozio e promozione



Schema dimensionale

D'ora in poi, faremo principalmente riferimento al data mart delle vendite, per riga di vendita, prodotto, negozio e giorno



Scelta dei fatti

Quali i fatti/misure relativi alle vendite dei prodotti (per scontrino per unità di vendita per negozio per giorno)?

- incasso in dollari (**dollar_sales**)
- numero di unità vendute (**units_sales**)
- costo in dollari (**dollar_cost**)
 - relativa al prodotto consegnato dal fornitore al negozio

Scelta dei fatti

E se invece avessimo scelto un data mart delle vendite giornaliere dei prodotti (per unità di vendita per negozio per giorno)?

- incasso totale in dollari (**dollar_sales**)
- numero totale di unità vendute (**units_sales**)
- costo totale in dollari (**dollar_cost**)
 - relativa al prodotto consegnato dal fornitore al negozio
- numero di clienti (**customer_count**)
 - che hanno acquistato il prodotto (SKU)
 - calcolato contando il numero di scontrini in cui è presente l'unità di vendita

Disponibilità dei fatti

Alcune misure relative alle vendite possono essere ottenute direttamente dai POS

- i POS permettono di esportare tutti i dati relativi agli scontrini emessi giornalmente
- questi dati possono essere elaborati per fornire le informazioni relative ai fatti scelti alla grana scelta
 - ad esempio, dati sugli incassi e sul numero di unità vendute

Tuttavia, per ottenere altre misure è necessaria l'applicazione di tecniche specifiche

- in questo caso, il costo sostenuto a fronte della vendita di alcune unità

Parentesi – Activity Based Costing (ABC)

La **Activity Based Costing (ABC)** è un metodo di analisi dei costi di un'industria o impresa per comprendere l'effettiva incidenza dei costi associati a ciascun prodotto o servizio venduto dalla ditta stessa

- metodologia dell'ABC – cinque fasi per comprendere e quindi migliorare i rapporti di input-output di una ditta
 - l'input è ciò che è necessario o utilizzato per produrre l'output, ovvero il prodotto o servizio
 - l'input è rappresentato dai costi e dalle attività in gioco, che possono essere produttive e necessarie, oppure esserlo meno
 - l'analisi ABC permette di comprendere se una parte dell'input è improduttiva o non necessaria e quale parte sia

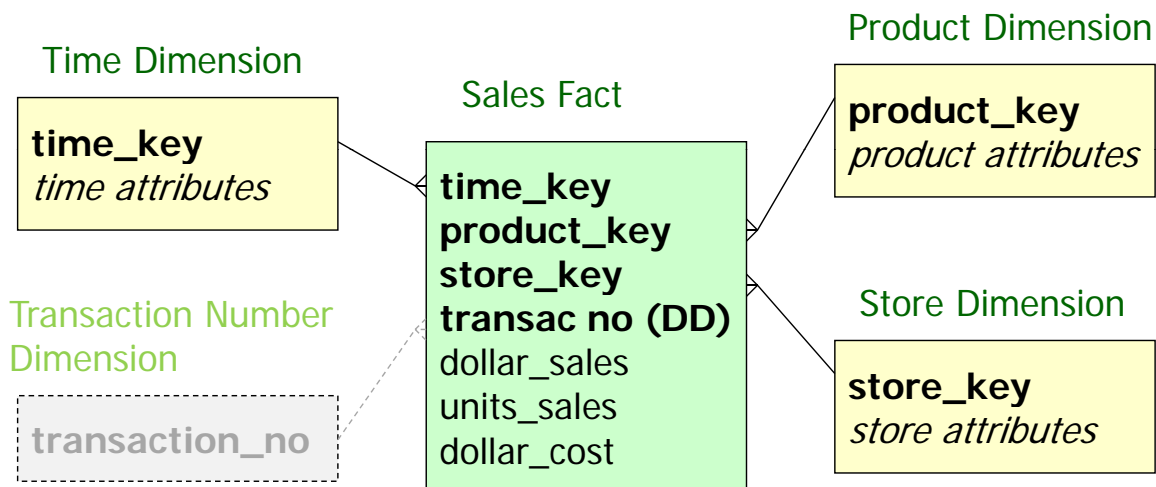
Parentesi – Activity Based Costing (ABC)

Le cinque fasi dell'ABC

- analisi delle attività – l'azienda deve comprendere esattamente quali attività svolge, intendendo con attività non tanto cosa produce, quanto le differenti cose necessarie per far sì che un dato prodotto o servizio prenda forma e sia vendibile sul mercato
- raccolta dati relativi ai costi – fissi e variabili
- riconduzione dei costi alle rispettive attività
- calcolo dell'output – *questa fase si occupa di calcolare il reale ammontare dei costi per unità produttiva*
- analisi dei costi

Schema dimensionale

Nuova versione dello schema dimensionale



Stima della taglia dei dati

Alcune stime relative alla quantità di dati

- il numero complessivo di voci nelle transazioni individuali può essere calcolato conoscendo l'incasso complessivo della catena ($\$4 \cdot 10^9$ per anno) e il costo medio della voce di vendita ($\$2$)
 - ci sono $2 \cdot 10^9$ voci nelle transazioni individuali
- le voci nelle transazioni individuali giornaliere per negozio sono $2 \cdot 10^9 / (365 \cdot 500) = 11.000$ circa
- ogni negozio offre 30.000 SKU, e ne vende giornalmente 3.000
 - il trasferimento dei dati dai negozi al data warehouse deve preferibilmente riguardare dati pre-elaborati

Occupazione di memoria della tabella fatti

Stima dell'occupazione di memoria della tabella fatti

- ipotesi
 - la chiave delle tabelle dimensione è un intero
 - di 4 byte per tempo, prodotto e numero di transazione
 - di 2 byte per negozio
 - i quattro campi chiave della tabella fatti occupano 14 byte
 - ogni fatto è rappresentato da un intero di 4 byte
- ogni riga della tabella fatti occupa 26 byte
- la tabella fatti contiene $500 \cdot 11.000 \cdot 365 = 2007.500.000$ tuple per anno
- se vengono mantenuti dati storici relativi a tre anni, l'occupazione di memoria della tabella fatti è di circa 160GB (di spazio primario)

Resistere allo snowflaking

Lo snowflaking di una dimensione riduce la ridondanza nella rappresentazione della tabella dimensione

- supponiamo ad esempio che la dimensione prodotto contenga 100 campi di 20 byte (mediamente) ciascuno
 - 60.000 righe di $100 \cdot 20$ byte corrispondono a un'occupazione di memoria di 120MB
- la normalizzazione della dimensione prodotto ne può ridurre l'occupazione di memoria al più del 100%
 - anche riducendola del 100%, l'occupazione di memoria complessiva del data mart verrebbe ridotta al più dell'0,1%
 - le prestazioni degraderebbero invece in modo proporzionalmente peggiore

Le tabelle fatti sono invece normalizzate

La dimensione tempo

La dimensione tempo (nel caso in esame) descrive i giorni di un intervallo temporale di interesse

- i membri della dimensione tempo sono i giorni dell'intervallo di interesse

La dimensione tempo è presente nella maggior parte degli schemi dimensionali, e praticamente in tutti i data warehouse

- la realizzazione di una tabella dimensione per il tempo è semplice
 - può essere facilmente pre-calcolata – ad es., usando un foglio Excel
 - i giorni per dieci anni sono poco più di 3.650

La dimensione tempo

È necessaria una tabella dimensione tempo esplicita? Non potrebbe essere invece usato un campo di tipo data?

- in alcuni (rari) casi, l'uso di un campo di tipo data è una scelta sufficiente
 - ma non c'è solitamente nessun vantaggio evidente per questa scelta
- i vantaggi di avere una tabella dimensione tempo esplicita sono relativi alla possibilità di scegliere gli attributi su cui fare selezioni, raggruppamenti, nonché quelli per realizzare dei confronti
 - ad es., la possibilità di distinguere tra giorni feriali, festivi e prefestivi, di considerare sia intervalli temporali solari che fiscali, di tenere conto delle stagioni di vendita, di eventi (ad esempio, la finale del Super Bowl) e altro

Chiave e attributi della dimensione tempo

- **time_key** è la chiave, un numero intero
- **date** è la data del giorno (ad esempio, 25 ottobre 2000)
- **year** è l'anno (2000)
- **month** è il mese (ottobre 2000)
- **quarter** è il numero del trimestre (4)
- **fiscal_period** è il periodo fiscale (4Q-2000)
- **day_of_week** è il giorno della settimana (“mercoledì”)
 - utile, ad esempio, per confrontare le vendite dei mercoledì rispetto ai venerdì
- **day_number_in_month** è il giorno nel mese (25)
 - per confrontare le vendite negli stessi giorni in mesi diversi

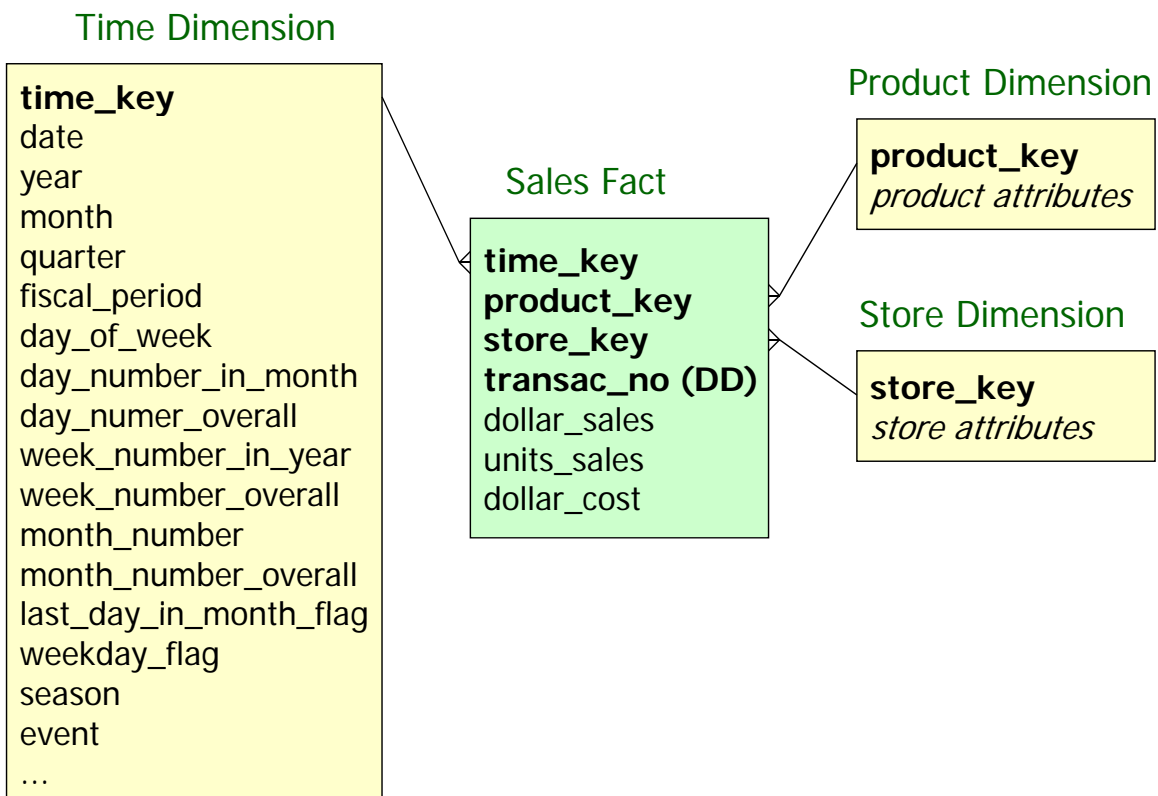
Chiave e attributi della dimensione tempo

- **day_number_overall** assegna una numerazione consecutiva a tutti i giorni di interesse
 - utile per calcoli aritmetici sulle date
- **week_number_in_year**, **week_number_overall**, **month_number**, **month_number_overall** hanno un significato analogo
- **last_day_in_month_flag** permette di selezionare l'ultimo giorno di ciascun mese
- **holiday_flag** permette di selezionare i giorni feriali/festivi
- **weekday_flag** permette di selezionare i giorni lavorativi

Chiave e attributi della dimensione tempo

- **season** è la “stagione di vendita”
 - ad esempio, Natale, Pasqua, San Valentino, nessuna stagione, ...
 - è importante scegliere valori “concreti” (come “nessuna stagione”) anche per rappresentare valori apparentemente nulli
 - i valori nulli vanno evitati
- **event**, simile a **season**, è associata a eventi speciali
 - ad esempio, finale del Super Bowl, Hurricane Hugo
- altri attributi

La dimensione tempo



La dimensione prodotto

La dimensione prodotto descrive le unità di vendita (SKU) della catena di negozi

- i dati per la dimensione prodotto sono solitamente estratti dal file principale dei prodotti usati per i sistemi POS
 - gestito dalla direzione e trasferito frequentemente dalla direzione ai POS
- è responsabilità della direzione recepire i nuovi UPC e creare dei nuovi record nel file principale dei prodotti
 - ad ogni nuovo UPC deve essere assegnato un numero di SKU univoco
 - la direzione assegna anche i numeri di SKU per i prodotti "locali"
- la tabella dimensione per i prodotti deve essere aggiornata in seguito a modifiche nel file dei prodotti

Attributi dei prodotti

Il file principale dei prodotti contiene molti attributi descrittivi per ciascuna SKU

- ad esempio, la gerarchia delle merci (merchandise hierarchy)
 - le **SKU** si raggruppano (roll up) per dimensioni delle confezioni (**package_size**)
 - le dimensioni delle confezioni si raggruppano in marche (**brand**)
 - le marche si raggruppano in sotto-categorie (**subcategory**)
 - le sottocategorie si raggruppano in categorie (**category**)
 - le categorie si raggruppano in reparti (**department**)

Attributi dei prodotti

Ad esempio

- **SKU**: Green 3-pack Brawny Paper Towels, UPC #...
- **package_size**: 3-pack
- **brand**: Brawny
- **subcategory**: paper towels
- **category**: paper
- **department**: grocery

Attributi dei prodotti

Altri attributi non fanno parte della gerarchia delle merci

- numero di SKU
- tipo della confezione
- prodotto dietetico
- peso (numerico) e unità di misura del peso
- colore
- unità per confezione venduta, unità per confezione spedita
- dimensioni (larghezza, altezza, profondità)
- molti altri...
 - la dimensione prodotto ha solitamente 50 o più attributi, che possono essere utilmente usati nelle interrogazioni come criteri di selezione e/o di raggruppamento

La dimensione negozio

La dimensione negozio descrive i negozi della catena

- i dati relativi ai negozi possono provenire da un foglio elettronico e/o da altre sorgenti informative

La dimensione negozio è una dimensione essenzialmente geografica

- ogni negozio occupa un punto nello spazio
- i negozi possono essere raggruppati rispetto a ogni possibile geografia
 - ad esempio (negli Stati Uniti) per zip code, città, contea, stato
 - ma anche per distretto di vendita e regione di vendita (nozioni relative alla struttura organizzativa della catena)

Attributi dei negozi

- nome, numero (codice nella catena), indirizzo, telefono, direttore, ...
- attributi geografici
 - zip code, città, contea, stato
 - distretto e regione di vendita
- informazioni su servizi supplementari
 - stampa foto, servizi finanziari, ...
- aree
 - area del negozio (in sqft), area del reparto surgelati, ...
- date
 - data prima apertura, ultima ristrutturazione, ...
 - rappresentati da date o da riferimenti a sinonimi della tabella dimensione tempo
- altri attributi

Nomi degli attributi

I nomi degli attributi devono essere il più possibile descrittivi e non ambigui

- ad esempio, negli schemi dimensionali sono solitamente presenti più dimensioni geografiche
 - come negozio, magazzino, cliente
 - ha senso di parlare della città in cui si trova il negozio o il magazzino, della città di residenza e di nascita del cliente
 - tali attributi (anche se in diverse tabelle)
 - non devono semplicemente chiamarsi **city**
 - ma devono chiamarsi **store_city**, **warehouse_city**, **customer_home_city**, **customer_born_city**
- inoltre, tutti i termini usati negli schemi devono essere opportunamente descritti in un glossario

Attributo o fatto?

Campi come le aree dei negozi sono numerici e additivi (attraverso i negozi)

- gli attributi sono solitamente descrittivi

I dati sulle aree dei negozi devono essere rappresentati come fatti?

- no, perché sono solitamente invariati nel tempo
 - i fatti interessanti variano al variare delle dimensioni da cui dipendono
- semmai, potrebbe essere utile introdurre degli ulteriori campi per categorizzare (ovvero, discretizzare) questi valori numerici
 - come piccolo, medio, grande, molto grande, oppure per fasce di aree

La dimensione numero di transazione

La dimensione numero di transazione rappresenta gli scontrini di vendita

- utilizzando anche questa dimensione come dimensione primaria, la grana dei dati nella tabella fatti diventa quella di una riga per ciascuna voce di scontrino di vendita
 - si tratta della grana più dettagliata possibile per questo tipo di processo

Dimensioni degeneri

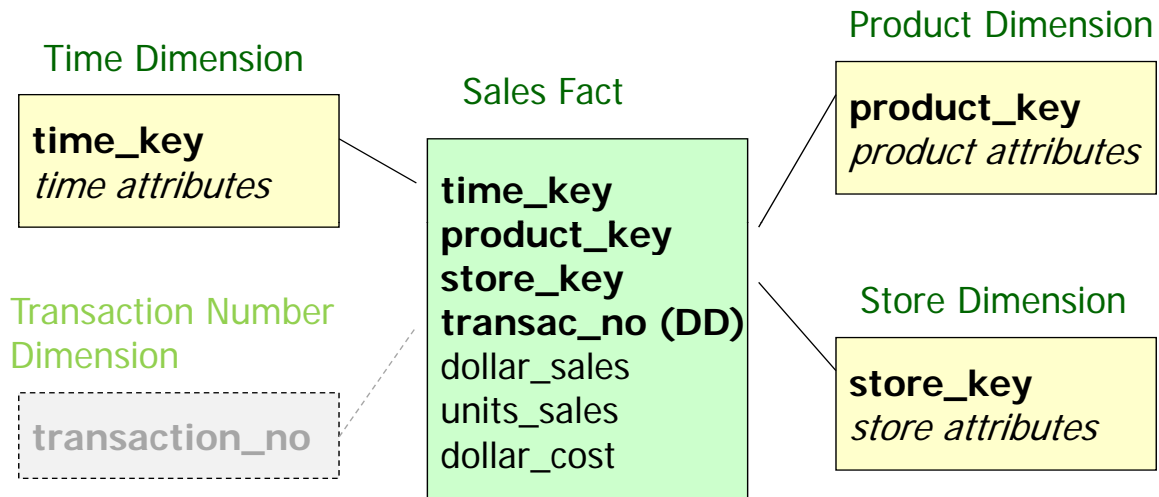
Quali attributi (informazioni) associate agli scontrini di vendita?

- ad es., data e negozio
- tuttavia, è preferibile rappresentare queste informazioni in dimensioni separate

Dunque, non ci sono informazioni descrittive interessanti associate ai membri di questa dimensione

- pertanto, la dimensione numero di transazione può essere rappresentata come **dimensione degenera**
 - senza tabella dimensione
 - la tabella fatti contiene il numero di transazione tra gli attributi chiave – ma non è un attributo chiave esterna
- le dimensioni degeneri sono normali, comuni ed utili
 - in particolare, comuni nella progettazione di tabelle fatti orientate ad “elementi” (linee o righe) di transazioni

Schema delle vendite in azione



Schema delle vendite in azione

Lo schema dimensionale della catena di negozi memorizza i seguenti fatti relativi alle vendite

- incasso in dollari (**dollar_sales**), numero di unità vendute (**units_sales**), costo in dollari (**dollar_cost**)
 - questi fatti sono additivi rispetto a tutte le dimensioni

Che cosa possiamo calcolare?

Il **profitto lordo** (per riga di vendita, prodotto, giorno e negozio) può essere calcolato sottraendo il costo dall'incasso

- anche questo fatto, calcolato, è additivo rispetto a tutte le dimensioni
- aggregandolo, è possibile ad esempio calcolare il profitto lordo giornaliero, per negozio/città, e/o per categoria di prodotto

Il **marginale lordo** può essere calcolato dividendo il profitto lordo per l'incasso

- per ogni possibile aggregazione, il margine lordo può essere calcolato prima sommando tutti gli incassi e i costi e poi dividendo
 - alcuni fatti non additivi (calcolati da fatti additivi) possono essere aggregati ricordandosi di distribuire correttamente le operazioni

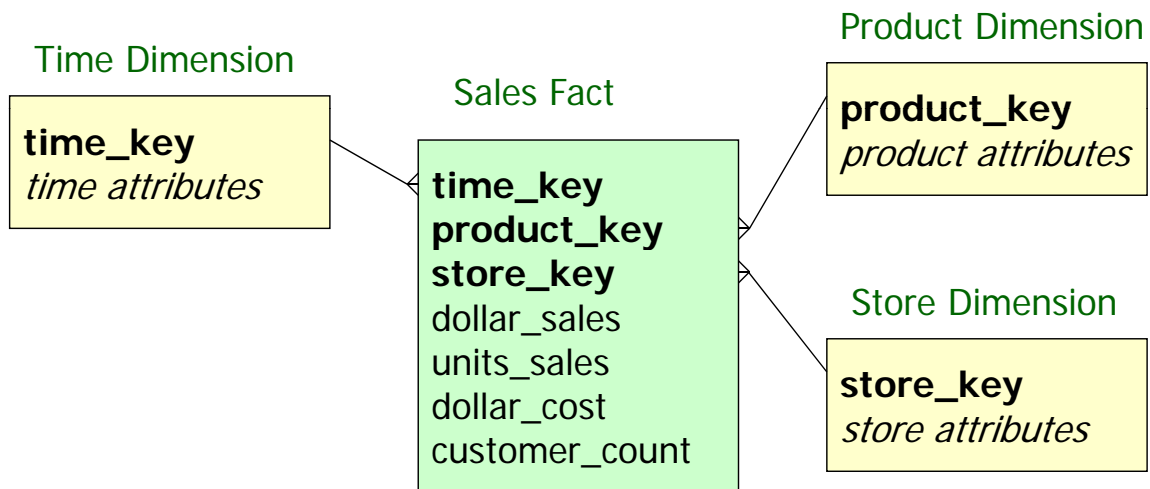
Che cosa possiamo calcolare?

Lo schema dimensionale mostrato consente inoltre – tramite strumenti di **data mining** – di effettuare le seguenti **analisi di market basket** – utili per prendere utili decisioni di marketing

- **associazioni**
 - analisi delle combinazioni di prodotti più comunemente acquistati insieme – richiede la dimensione “numero di transazione”
- **pattern sequenziali**
 - analisi delle regolarità nel comportamento dei clienti in una sequenza di acquisti – richiede transazioni non anonime, e dunque la dimensione “cliente”
- **classificazione**
 - dei clienti o dei prodotti – in classi predefinite
- **clustering**
 - suddivide clienti o prodotti in gruppi (classi) simili

Schema delle vendite in azione

Cosa dire invece a proposito del data mart delle vendite giornaliere, per prodotto e negozio?



Schema delle vendite in azione

In questo caso, i fatti a disposizione sono

- incasso totale in dollari (**dollar_sales**), numero totale di unità vendute (**units_sales**), costo totale in dollari (**dollar_cost**), numero di clienti (**customer_count**)
 - i primi tre fatti sono additivi rispetto a tutte le dimensioni
 - il numero di clienti è invece un fatto semi-additivo

Fatti non additivi

Il numero di clienti è un fatto semi-additivo

- non è additivo rispetto alla dimensione prodotto
 - se un prodotto A è stato acquistato da 20 clienti e un prodotto B da 30 clienti, quanti clienti hanno comprato A o B?
- tuttavia, è additivo rispetto alle altre dimensioni

I conteggi sono solitamente fatti semi-additivi

- possono essere sommati correttamente restringendo le chiavi nelle dimensioni in cui non sono additivi a valori singoli

La dimensione promozione

Uno degli scopi di questo schema dimensionale è abilitare la comprensione dell'impatto delle varie promozioni sulle vendite

- analisi di questo tipo possono essere effettuate mediante l'introduzione di una dimensione supplementare per le promozioni
- vanno inoltre usati degli opportuni modelli matematici (fuori della portata di questa trattazione) per comprendere
 - l'andamento base delle vendite – in un'ipotetica assenza di promozioni
 - l'impatto di ciascun tipo di promozione in termini di differenze rispetto a questo andamento base

Effetto del tempo atmosferico sulle vendite

Illustriamo l'intuizione necessaria per comprendere l'effetto delle promozioni sulle vendite con un esempio un po' diverso

Come analizzare l'impatto del tempo atmosferico sulle vendite?

- aggiungo una dimensione supplementare Tempo atmosferico
 - con membri come “bel tempo”, “nuvoloso”, “pioggia”, ...
- utilizzo una sorgente di dati esterni per sapere quale è stato il tempo in ciascun giorno (del periodo di riferimento) nelle città in cui ci sono negozi della catena
- aggiungo a ciascuna riga della tabella fatti una chiave esterna verso la dimensione Tempo atmosferico
 - non lo faccio a mano, basta un programmino
- faccio l'analisi utilizzando degli opportuni modelli matematici
 - scopro l'eventuale impatto del tempo sulle vendite

La dimensione promozione

La dimensione promozione descrive ogni possibile promozione che si applica alla vendita dei prodotti

- ad esempio, riduzioni temporanee di prezzi, esposizione alla fine dei corridoi, pubblicità sui giornali, buoni sconto, ...
- alternativamente, la dimensione promozione potrebbe descrivere la promozione effettivamente applicata alla vendita – ad esempio, pagamento mediante buono sconto
 - non è questa l'ipotesi per questo studio di caso

La dimensione promozione è una dimensione causale (non casuale)

- descrive fattori che sono la causa di potenziali cambiamenti (nelle abitudini dei clienti)
- la dimensione promozione è la dimensione potenzialmente più interessante del nostro schema dimensionale

Effetti delle promozioni

Alcuni possibili effetti delle promozioni

- aumenti della vendita dei prodotti in promozione
 - misurabili solo se sono noti i livelli base di vendita (senza la promozione)
 - i livelli base di vendita possono essere stimati dalle vendite precedenti e sulla base di modelli matematici sofisticati
- diminuzione della vendita al termine della promozione
- riduzione della vendita di altri prodotti
- aumento complessivo della vendita, considerando il periodo della promozione e periodi immediatamente precedenti e/o successivi
- profittabilità della promozione
 - tiene conto dei diversi aspetti

Promozioni

Le diverse modalità di promozione possono essere applicate contemporaneamente

- ad esempio, riduzione temporanea del prezzo, pubblicità sui giornali e esposizione alla fine dei corridoi
- una possibile scelta è che ciascuna riga della tabella dimensione delle promozioni descriva una possibile combinazione delle modalità di promozione
 - ad es., riduzione temporanea del prezzo e pubblicità sui giornali
 - anche se in un anno ci possono essere 1.000 pubblicità sui giornali, 1.000 riduzioni temporanee dei prezzi e 200 esposizioni alla fine dei corridoi, le combinazioni effettive sono solitamente limitate (ad esempio, 5.000)

Promozioni

- ciascuna particolare promozione può essere applicata diversamente nei diversi negozi
 - ad esempio, in alcuni negozi può essere impossibile effettuare le esposizioni alla fine dei corridoi
 - in questo caso, la promozione verrebbe rappresentata da due righe
 - riduzione, pubblicità e esposizione
 - riduzione e pubblicità

Attributi delle promozioni

- nome della promozione
- tipo della riduzione di prezzo
 - ad esempio, buono sconto, temporanea, nessuno
- tipo della pubblicità
 - ad esempio, giornale, radio, giornale e radio, posta
- media della pubblicità
- tipo dell'esposizione
- tipo del buono sconto
- costo della promozione
- date di inizio e fine della promozione
- altri attributi

Dimensioni “junk”

Le promozioni sono basate su quattro meccanismi causali

- riduzione di prezzo, pubblicità, esposizione, buoni sconto

La promozione è una sola dimensione – oppure deve essere rappresentata da quattro diverse dimensioni?

- la decomposizione in quattro dimensioni è possibile
 - dipende dai requisiti e dalle esigenze di analisi dell’utente finale
 - se l’utente pensa separatamente (indipendentemente) a questi quattro meccanismi, allora è forse opportuno definire quattro diverse dimensioni
- l’uso di una sola dimensione porterebbe ad una “dimensione di scarti” – una **junk dimension**
 - ovvero, una raggruppamento di comodo di attributi (più o meno) casuali

Tablelle fatti senza fatti

Lo schema dimensionale che è stato costruito è in grado di rispondere a molte interrogazioni

- tuttavia, non è in grado di calcolare i prodotti in promozione che non sono stati venduti
 - più avanti sarà studiata una tecnica (tablelle fatti senza fatti) per poter gestire anche questo tipo di informazioni

Sull'additività dei fatti

Si noti che, se la promozione indicasse la combinazione di promozioni effettivamente applicata alla vendita, allora il numero di clienti non sarebbe completamente additivo rispetto alle promozioni

- perché un cliente potrebbe comprare, in una stessa transazione, una unità di vendita con un buono sconto e la stessa unità di vendita senza buono sconto
 - se questa situazione è considerata infrequente, può essere trascurata
 - se invece è frequente e vuole essere analizzata, la dimensione “buono sconto” deve essere scorporata dalla dimensione promozione