

# Tabelle fatti senza fatti

Luca Cabibbo  
marzo 2010

## Tabelle fatti senza fatti

In tutti i casi studiati finora, le tabelle fatti avevano la seguente struttura

- due o più chiavi esterne, riferimenti alle chiavi delle dimensioni
- uno o più fatti
  - i fatti rappresentano misure numeriche prese all'intersezioni delle dimensioni

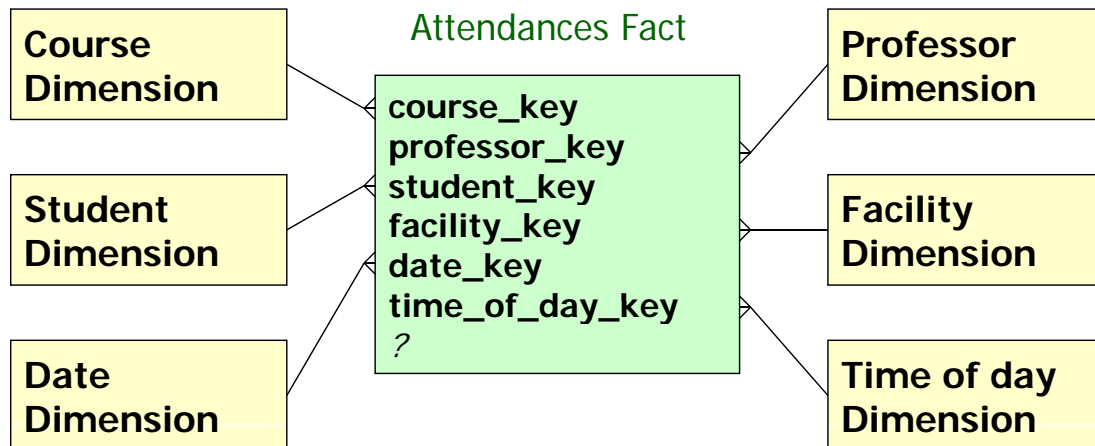
Alcuni processi interessanti sono caratterizzati da “fatti” che (apparentemente) non hanno proprietà misurabili

- processi con fatti non misurabili possono essere rappresentati da schemi dimensionali basati su **tabelle fatti senza fatti (factless fact tables)**
- ci sono due varianti principali delle tabelle fatti senza fatti

# Eventi

In diverse situazioni bisogna memorizzare un grande numero di **eventi**, che si verificano all'intersezione di un certo numero di dimensioni – ma non ci sono misure numeriche naturali associate a tali eventi

- ad es., frequenza giornaliera di studenti a corsi universitari



# Rappresentazione di eventi

Un insieme di eventi (senza fatti) può essere rappresentato da una tabella fatti senza fatti e da un insieme delle dimensioni di interesse

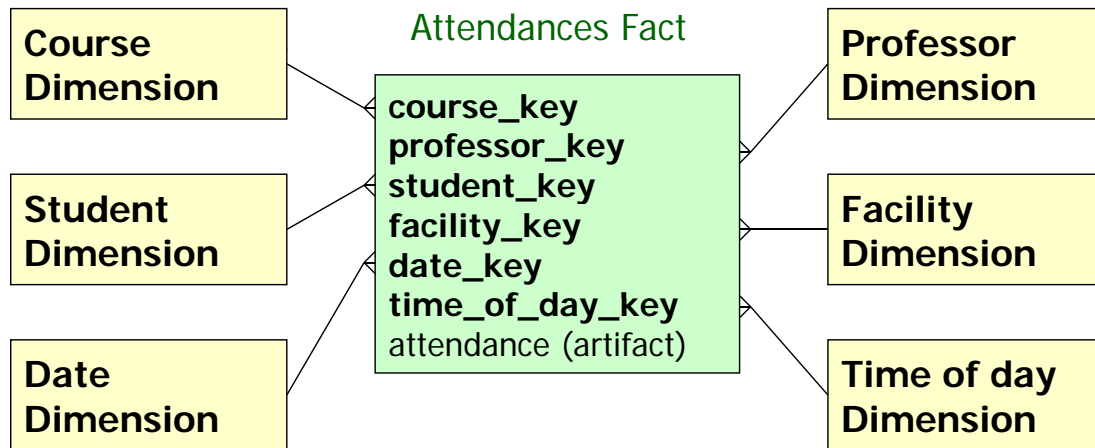
- analisi
  - quali sono stati i corsi più frequentati?
  - quali sono state le aule più utilizzate?
  - qual è stata l'occupazione media delle aule in funzione dell'ora del giorno?

Molte di queste analisi richiedono di contare il numero di occorrenze distinte di uno certo insieme di attributi rispetto a in un insieme di eventi

- non possono essere sempre calcolate solo con la funzione COUNT di SQL
- è spesso necessario scrivere COUNT(DISTINCT ...)

## Rappresentazione di eventi

Le interrogazioni di una tabella fatti senza fatti possono essere semplificate introducendo in essa un fatto numerico fittizio a cui viene assegnato il valore di comodo **1**



- in questo caso, è possibile scrivere interrogazioni corrette usando la funzione COUNT o SUM
  - le interrogazioni risultano più comprensibili

## Tabelle di copertura

Un'altra applicazione delle tabelle fatti senza fatti riguarda la rappresentazione (o l'identificazione) di eventi che non sono accaduti

- nel processo delle vendite, lo schema dimensionale proposto non permette di effettuare la seguente analisi
  - quali prodotti in promozione non sono stati venduti?
- è possibile usare una **tabella di copertura** per rappresentare tutti i prodotti in promozione nei vari giorni e negozi
  - all'evento "essere in promozione" potrebbe non essere associato nessun particolare fatto misurabile
  - i prodotti in promozione non venduti possono essere calcolati per differenza insiemistica

## Tabelle di copertura

In questo caso, la tabella di copertura delle promozioni è densa (rispetto ai prodotti in promozione)

- deve però memorizzare solo i prodotti in promozione
  - e non i prodotti che non sono in promozione
- anche in questo caso, può essere opportuno introdurre un fatto fittizio **existence**, di valore costante 1
- se le promozioni sono settimanali, la grana della dimensione tempo può essere la settimana anziché il giorno
  - può essere una dimensione “conformata”, a dettaglio ridotto