

# Famiglie di tabelle fatti

Luca Cabibbo  
marzo 2010

## Famiglie di tabelle fatti

Finora ci siamo concentrati soprattutto sulla costruzione di semplici schemi dimensionali – costituiti da

- una singola tabella fatti
- circondata da un insieme di tabelle dimensione

In realtà, la maggior parte dei data mart è costituita da un gruppo coordinato (una “famiglia”) di tabelle fatti

- più avanti, discuteremo l'importanza delle dimensioni conformi in tali famiglie di tabelle fatti – e l'architettura a bus del data warehouse
- per ora, ricapitoliamo i tipi fondamentali di tabelle fatti ed introduciamo alcune famiglie comuni di tabelle fatti

## Tre tipi fondamentali di tabelle fatti

Finora abbiamo incontrato tre tipi fondamentali di tabelle fatti

- *a transazioni individuali*
- *ad istantanee periodiche*
- *ad istantanee accumulate*

Se utilizzati insieme, questi tre tipi fondamentali di tabelle fatti consentono di rappresentare i dati di molti processi di business

## Un confronto

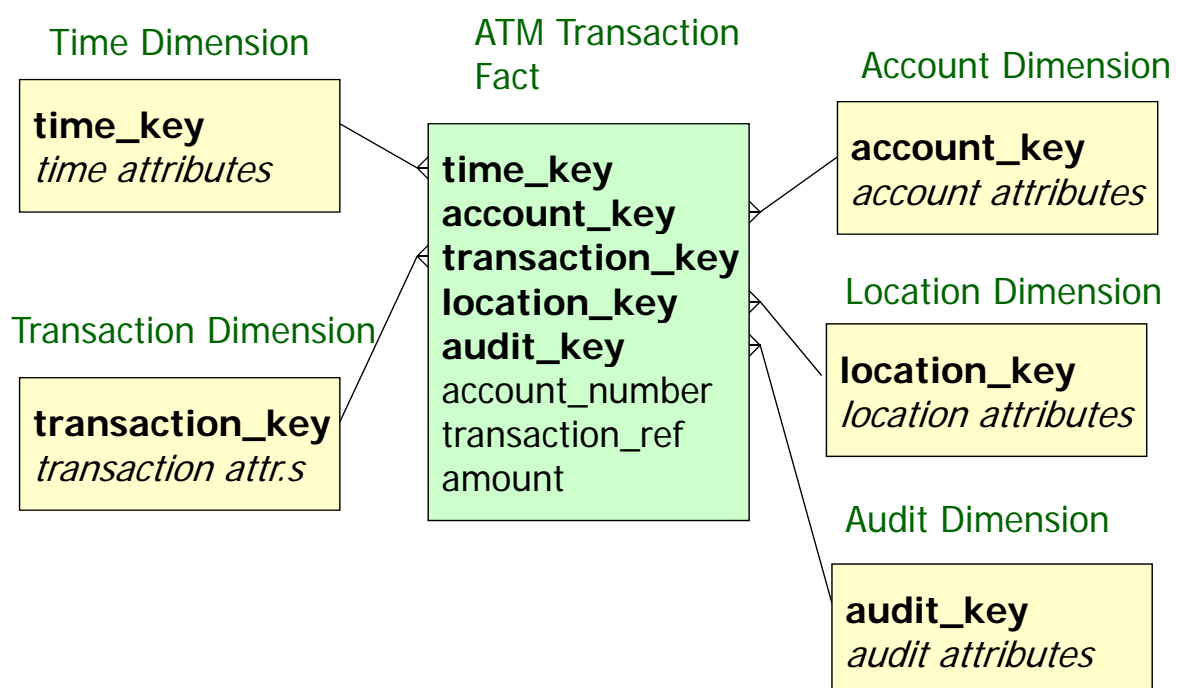
	<b>transazioni individuali</b>	<b>istantanee periodiche</b>	<b>istantanee accumulate</b>
<i>periodo di tempo rappresentato</i>	evento in un istante di tempo	intervallo regolare	periodo di tempo indeterminato, generalmente di breve durata
<i>grana</i>	una riga per evento di transazione	una riga per periodo	una riga per ciclo di vita
<i>caricamento tabella fatti</i>	inserimento	inserimento	inserimento e aggiornamento
<i>aggiornamenti righe fatti</i>	no	no	ad ogni attività
<i>dimensioni data</i>	data di transazione	data di fine periodo	date multiple su eventi cardine standard
<i>fatti</i>	misure della transazione	misure per intervallo di tempo	misure sul ciclo di vita

## Schemi a transazioni ed istantanee

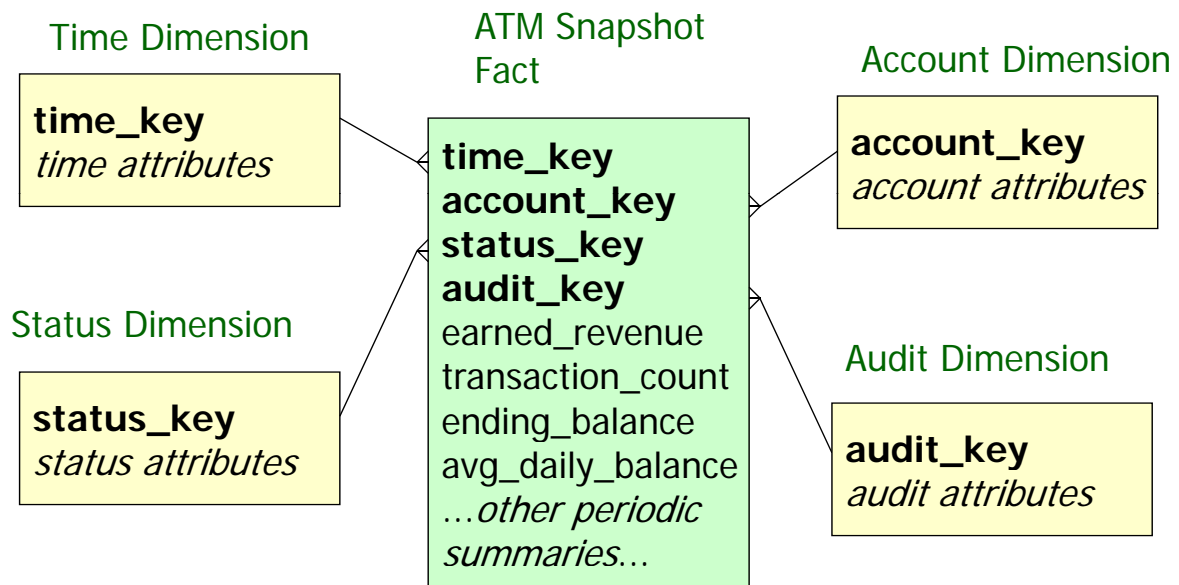
Se utilizzati insieme, i tre tipi fondamentali di tabelle fatti consentono di rappresentare i dati di molti processi di business

- ovvero, ci sono delle strutture caratteristiche che sono probabilmente indipendenti dal particolare processo di business di riferimento
- inoltre, quasi tutti i data mart richiedono due rappresentazioni diverse e separate dei dati – una a transazioni ed una ad istantanee
- questo dà luogo ad un'importante famiglia di tabelle fatti – gli **schemi a transazioni ed istantanee**

## Esempio – ATM, transazioni



## Esempio – ATM, istantanee



7

Famiglie di tabelle fatti

Luca Cabibbo

## Esempio

Alcune finalità dei due diversi tipi di schemi dimensionali

- a transazioni
  - analisi del comportamento
  - analisi dell'uso dell'ATM nelle diverse ore del giorno
  - analisi delle code
  - analisi del comportamento sequenziale del cliente
  - rilevamento di frodi
  - analisi di carrello (non ATM)
- ad istantanee
  - per misurare rapidamente lo stato delle cose
  - classificazione/clustering
- congiuntamente
  - visione completa del business – attraverso le diverse dimensioni

8

Famiglie di tabelle fatti

Luca Cabibbo

## Fundamental Grains – R. Kimball

A powerful theme in data warehouse design is measurement publication.

- We can decompose many of the meaningful transactions captured by our OLTP systems into two main components: a context for the transaction and an amount. We view the amount as a measurement. [...]
- Our duty as data warehouse designers is to publish these measurements for all the managers and analysts in our organization so they can make sense of the pattern of transactions. Ultimately, revenue, costs, profit, and much of the important customer behavior should be understandable from this series of measurements. [...]

## Fundamental Grains – R. Kimball

As compelling as the transactional view of the world is, there's more to life than transactions.

- Although we can always reconstruct a complete picture of a business from the transaction history, it is often grossly impractical to do so. Often we need to augment the transactions with an alternate view of the world: the snapshot.
- A snapshot is a measurement of status at a specific point in time.

## Fundamental Grains – R. Kimball

In the data warehouse world, there are two different snapshots of status, both equally important.

- The periodic snapshot is a regular, predictable statement of status. [...] Some of these snapshots are instantaneous measurements at the end of the period, and some are measurements accumulated throughout the period. The design of this snapshot focuses on the activity that occurred during the time interval between snapshots.
- The other important kind of snapshot is the accumulating snapshot, which shows us the status at any given moment. If we ask for the current status an hour from now, it may be different. The accumulating snapshot often combines the most recent volatile status with measures that accumulate from the beginning of history.

## Fundamental Grains – R. Kimball

The three kinds of measurements — transaction, periodic snapshot, and accumulating snapshot — are the choices for the grain of any given fact table. All three are useful, and you often need a minimum of two to get a complete picture of a business.[...]

- The transaction-grained fact table represents an atomic action that occurs at an instantaneous point in time. [...]
- The periodic-snapshot-grained fact table represents a predefined time span. [...]
- The accumulating-snapshot-grained fact table represents an indeterminate time span, covering the entire history. [...]

## Fundamental Grains – R. Kimball

### Where Do We Use Each Fact-Table Type?

- The transaction-grained fact table is the only table that can answer detailed questions about timing and questions about unpredictable, repetitive behavior. [...]
- The periodic-snapshot fact table is often the only table that can easily generate a regular, predictable view of the important measures of a business. [...]
- The accumulating-snapshot fact table is appealing when we are tracking items that have a finite lifetime. [...]

[http://www.intelligententerprise.com/db\\_area/archives/1999/993003/warehouse.jhtml](http://www.intelligententerprise.com/db_area/archives/1999/993003/warehouse.jhtml)

## Altre famiglie di tabelle fatti

Alcune ulteriori famiglie di tabelle fatti – che saranno studiate più in dettaglio nel seguito

- **catene del valore**
  - quando un processo è basato su un flusso logico direzionato – ad esempio, un ordine o un prodotto che evolve attraverso una sequenza di passi
  - ciascuno di questi passi è rappresentato da un diverso data mart – definiti su un insieme di dimensioni comuni
  - l'analisi congiunta dei vari data mart consente una visione complessiva sull'intero processo
- **cerchi del valore**
  - un processo in cui diverse entità stanno eseguendo o misurando uno stesso tipo di transazioni
  - ad es., ospedali, farmacie e compagnie assicurative nel contesto della sanità

## Altre famiglie di tabelle fatti

Alcune ulteriori famiglie di tabelle fatti – che saranno studiate più in dettaglio nel seguito

### ■ **scemi di prodotti eterogenei**

- è di interesse l'analisi di un insieme non omogeneo di tipologie di prodotti o servizi
- tutti questi prodotti o servizi hanno alcune caratteristiche comuni – ma ciascuna ha anche delle proprie caratteristiche peculiari
- viene usato un data mart per la rappresentazione delle caratteristiche comuni a tutti i prodotti/servizi – più uno separato per ciascun diverso prodotto/servizio

## Altre famiglie di tabelle fatti

Alcune ulteriori famiglie di tabelle fatti – che saranno studiate più in dettaglio nel seguito

### ■ **aggregazioni**

- una famiglia di tabelle fatti usate per rappresentare uno stesso insieme di fatti, aggregati in modi diversi
- due motivazioni principali
  - efficienza/prestazioni
  - ci sono alcune misure diverse – che è di interesse correlare – che sono disponibili solo a granularità diverse – ad esempio, dati sugli incassi disponibili ai massimi livelli di dettaglio – ma i dati sui costi non possono essere espressi allo stesso livello di dettaglio