

# Servizi finanziari (studio di caso)

Luca Cabibbo  
marzo 2010

1

Servizi finanziari

Luca Cabibbo

## Il processo dei servizi finanziari

Viene ora considerato il data warehouse per una grande banca

- la banca offre diversi servizi finanziari
  - ad esempio, conti correnti, libretti di risparmio, mutui, investimenti, prestiti personali, carte di credito, cassette di sicurezza, ...
- vogliamo analizzare tutti i conti gestiti dalla banca, con i relativi titolari individuali, ma soprattutto le “famiglie” (residenziali e commerciali) a cui appartengono
- lo scopo è
  - offrire servizi più efficienti
    - per attrarre nuovi clienti
  - offrire servizi supplementari
    - per i propri clienti

2

Servizi finanziari

Luca Cabibbo

## Servizi finanziari (1)

Mediante delle interviste, sono stati raccolti i seguenti requisiti

- devono essere memorizzati cinque anni di storia per ciascun conto
  - per i mesi precedenti, si vuole una istantanea all'ultimo giorno del mese
  - per il mese corrente, si vuole una istantanea al giorno precedente (a ieri)
- ogni tipo di conto ha un saldo primario
  - c'è la necessità di analizzare contemporaneamente diversi tipi di conto mediante confronti basati sul saldo primario
- ogni tipo di conto è caratterizzato da attributi e fatti numerici propri
  - che sono diversi da tipo di conto a tipo di conto

## Servizi finanziari (2)

- ogni conto è considerato appartenente ad un'unica "famiglia" – intesa come un'intera unità economica, residenziale oppure commerciale
  - attenzione, gli individui che detengono i conti possono cambiare famiglia nel corso del tempo
- i dati sui titolari dei conti sono di qualità bassa
  - in particolare, i dati sui titolari sono leggermente differenti in ciascun diverso sistema operativo in uso nella banca
- siamo interessati a informazioni demografiche sia sulle famiglie che sui titolari individuali dei conti

## Il data mart dei servizi finanziari

Viene deciso di realizzare un data mart per il processo dei conti delle famiglie

- la progettazione del data mart viene svolta in due passi
  - prima viene progettato uno schema dimensionale **nucleo (core)** relativo a tutti i conti (ovvero, a tutti i tipi di conto)
    - vengono presi in considerazione solo dimensioni e fatti effettivamente condivisi da tutti i tipi di conto
  - successivamente il progetto del data mart verrà completato tenendo in considerazione le caratteristiche individuali di ciascun tipo di conto
    - i tipi di conto formano un insieme di “prodotti eterogenei”

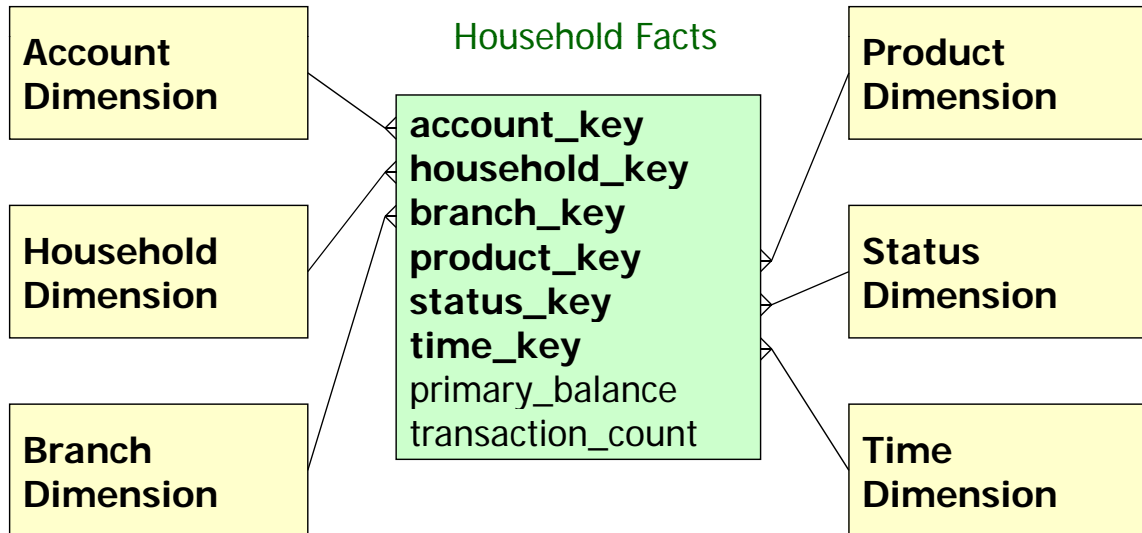
## Progettazione della tabella fatti nucleo

La tabella fatti nucleo per il processo dei conti delle famiglie dipende dalle seguenti dimensioni

- conto (account), famiglia (household), filiale (branch), prodotto (product), stato (status) e tempo (time)
- la grana della tabella fatti è conto per mese
  - le dimensioni primarie sono conto e tempo
  - le dimensioni famiglia, filiale, prodotto e stato sono dimensioni supplementari
- i fatti di interesse sono
  - saldo primario (primary balance) e numero di transazioni (transaction count)

## Schema dimensionale nucleo

Schema dimensionale nucleo per il processo dei conti delle famiglie



7

Servizi finanziari

Luca Cabibbo

## Attributi delle dimensioni

### Conto

- nome primario, nome secondario, indirizzo, zip, stato, città, data apertura, età primario, sesso primario, stato civile primario

### Famiglia

- nome capofamiglia, indirizzo, zip, stato, città, reddito, tipo di famiglia

### Filiale

- nome filiale, indirizzo, zip, stato, città, tipo filiale

### Prodotto

- descrizione, tipo, categoria

8

Servizi finanziari

Luca Cabibbo

## Attributi delle dimensioni

### Stato

- descrizione dello stato, ragione, flag nuovo conto, flag conto chiuso

### Tempo

- mese, anno, trimestre fiscale

## Conti e famiglie

La scelta di trattare separatamente conti e famiglie è legata principalmente al fatto che la dimensione conto è grande ed è soggetta a cambiamenti nel tempo

- la definizione degli elementi di una famiglia può cambiare nel tempo
- la nozione di conto è fondamentale nella banca
  - bisogna evitare di gestire la dimensione conto come di tipo 2 (non sono interessanti le versioni dei conti)
- la relazione tra conti e famiglie è di fatto gestita mediante la tabella fatti

## Prodotti

La dimensione prodotto contiene solo gli attributi comuni a tutti i diversi tipi di conto (servizi finanziari) offerti dalla banca

- contiene una gerarchia di attributi elementare (descrizione, tipo e categoria)
- non contiene nessun attributo proprio di nessun tipo particolare di conto
  - questi attributi più specifici saranno introdotti nello schema nella seconda fase della progettazione

Il prodotto (tipo di conto) è una caratteristica di ciascun conto

- la dimensione prodotto è stata mantenuta separatamente dalla dimensione conto perché la banca pensa ai prodotti come a una dimensione fondamentale della banca

## Stato

La dimensione stato descrive lo stato in cui può trovarsi un conto

- ad esempio, essere attivo, inattivo, chiuso
- descrive le ragioni per cui un conto si trova in un certo stato
- viene usata anche per tenere traccia dei cambiamenti di stato dei conti

## Dimensioni sporche

Nello schema dimensionale per i servizi finanziari non è presenta una dimensione cliente

- la dimensione più vicina a cliente è conto
  - ogni conto ha uno, due o più individui come titolari
- nei servizi finanziari (per ragioni storiche e culturali) l'enfasi è sui conti, e non sui loro titolari
  - di conseguenza, i dati sulla correlazione conto-cliente potrebbero essere di bassa qualità
- se fosse presente la dimensione cliente, sarebbe una **dimensione sporca**
  - ovvero, una dimensione che contiene molti membri duplicati o non completamente corretti

La dimensione famiglia è probabilmente una dimensione sporca (con una accuratezza dell'80% circa)

## Semi-additività dei saldi

I saldi dei conti sono fatti semi-additivi (come i livelli di inventario)

- non sono additivi rispetto al tempo
- per combinare un insieme di saldi attraverso un intervallo temporale, calcolandone la media, bisogna sommare i saldi e dividere per il numero di periodi dell'intervallo temporale

Le applicazioni che devono calcolare la media di dati semi-additivi (non additivi rispetto al tempo)

- non devono usare direttamente la funzione AVG di SQL
- devono calcolare in proprio la funzione (virtuale) "media su un intervallo temporale"

## Prodotti eterogenei

La progettazione del data mart dei servizi finanziari deve considerare anche la presenza di diverse tipologie di **prodotti eterogenei (heterogeneous products)**

- i prodotti sono eterogenei nel senso che ciascun prodotto è caratterizzato da un proprio insieme di attributi (descrittivi) e di fatti (numerici)
  - prodotti diversi sono molto dissimili rispetto a questi insiemi di attributi
- ad esempio
  - i conti correnti sono caratterizzati da un saldo del conto, saldo minimo, limite dello scoperto, spese, collegamento a carte di credito, ...
  - le carte di credito sono caratterizzate da un saldo, limite del credito, data di emissione e scadenza, ...

## Gestione di prodotti eterogenei

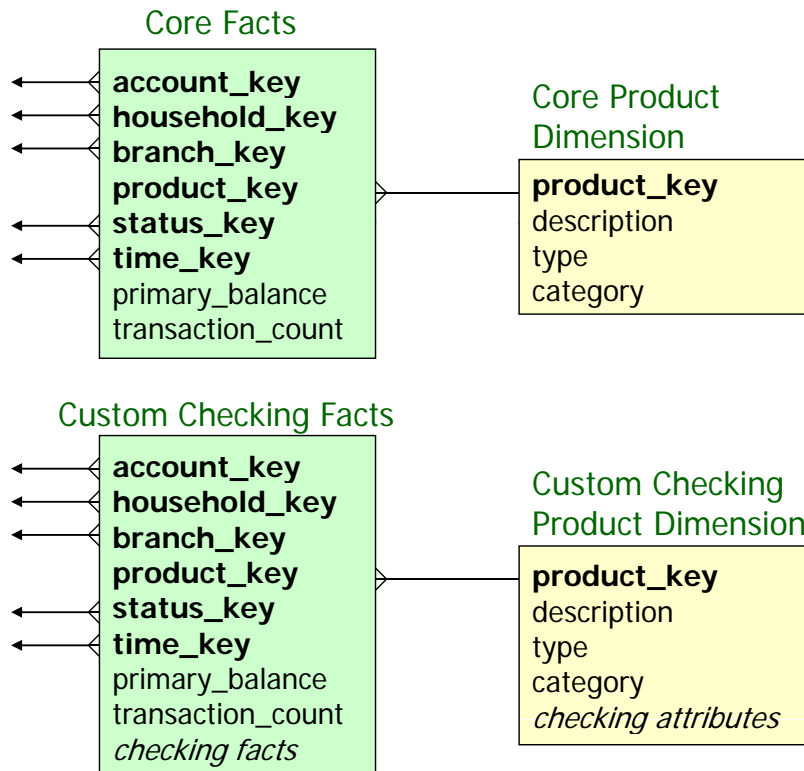
Una prima soluzione è quella di estendere la tabella fatti con i fatti di ciascun tipo di conto, e la dimensione prodotto con gli attributi di ciascun tipo di conto

- tuttavia, questi fatti e attributi sarebbero spesso non significativi e vuoti, con notevole spreco di occupazione di memoria e degrado delle prestazioni
- inoltre, l'introduzione di nuovi tipi di prodotto sarebbe difficoltosa

Una soluzione più efficace consiste invece nell'introdurre una nuova tabella fatti ed una nuova dimensione personalizzata (**custom**) per ciascun tipo di conto

- oltre alla tabella fatti e alla tabella dimensione che descrivono il nucleo comune a tutti i tipi di conto

## Schema dimensionale per prodotti eterogenei



17

Servizi finanziari

Luca Cabibbo

## Schemi dimensionali per i prodotti eterogenei

Il data mart completo dei servizi finanziari comprende

- una tabella fatti ed una tabella dimensione nucleo
- una tabella fatti e una tabella dimensione personalizzata per ciascun tipo di prodotto
  - i tipi di prodotto sono solitamente in numero limitato e predeterminato
- ogni membro della dimensione prodotto nucleo è anche membro di (esattamente) una dimensione prodotto personalizzata
  - in cui sono riportati tutti gli attributi supplementari
- ogni record della tabella fatti nucleo descrive un insieme di fatti che è riportato anche in un record di (esattamente) una tabella fatti personalizzata
  - che riporta anche tutti i fatti supplementari

18

Servizi finanziari

Luca Cabibbo

## Analisi di prodotti eterogenei

Le analisi di interesse per una base di dati di prodotti eterogenei sono di due tipi

- analisi interessate a più tipi di prodotti eterogenei
  - sono interessate solo a fatti e attributi comuni a tutti i tipi di prodotto
    - realizzate come interrogazioni che accedono solo alle tabelle nucleo
- analisi interessate a un solo tipo di prodotto
  - sono interessate anche a fatti e attributi propri del particolare tipo di prodotto
    - realizzate come interrogazioni che accedono solo alle tabelle personalizzate, che riportano anche i dati comuni (e quindi non richiedono il join tra tabelle nucleo e personalizzate)

## Altre rappresentazioni di prodotti eterogenei

Il data mart dei servizi finanziari è formato dalle istantanee relative a conti appartenenti a un certo numero di classi di prodotti eterogenei

- la tabella fatti contiene in questo caso molti fatti

Che cosa sarebbe successo se la grana del data mart fosse stata quella delle transazioni?

- in questo caso, l'unico fatto interessante sarebbe stato la quantità (spostata dalla transazione), indipendentemente dalla classe di prodotto
  - non ci sarebbe stata la necessità di definire tabelle fatti personalizzate per le diverse classi di prodotti eterogenei
  - ma solo di definire tabelle dimensioni personalizzate

## Uso di minidimensioni

Le dimensioni conto e famiglia sono grandi e soggette a cambiamenti lenti nel tempo

- in questo caso, può essere opportuno evidenziare gli attributi variabili nel tempo di cui interessa tenere traccia, e decomporre le dimensioni mediante l'introduzione di minidimensioni
  - le minidimensioni rappresenterebbero classi predefinite (e sostanzialmente invarianti nel tempo) di combinazioni degli attributi variabili
  - le dimensioni sarebbero gestite come di tipo 1
  - sarebbe poi possibile accedere ai fatti sia sulla base dei valori delle dimensioni nei periodi relativi ai fatti, che dei valori correnti

## Dimensioni multivalore

Si consideri ora un diverso requisito – ad un conto possono essere associati uno, due oppure anche più intestatari o clienti – non ci interessano più le “famiglie”

- come gestire una tale situazione, ed in particolare la relazione tra le dimensioni candidate conto e cliente e la loro relazione con la tabella fatti?
  - “cliente” non può essere un attributo della dimensione “conto” – poiché la dimensione conto ha una sola riga per ciascun conto
  - “cliente” non può essere nemmeno una dimensione – poiché la granularità della tabella fatti è una riga per conto per mese

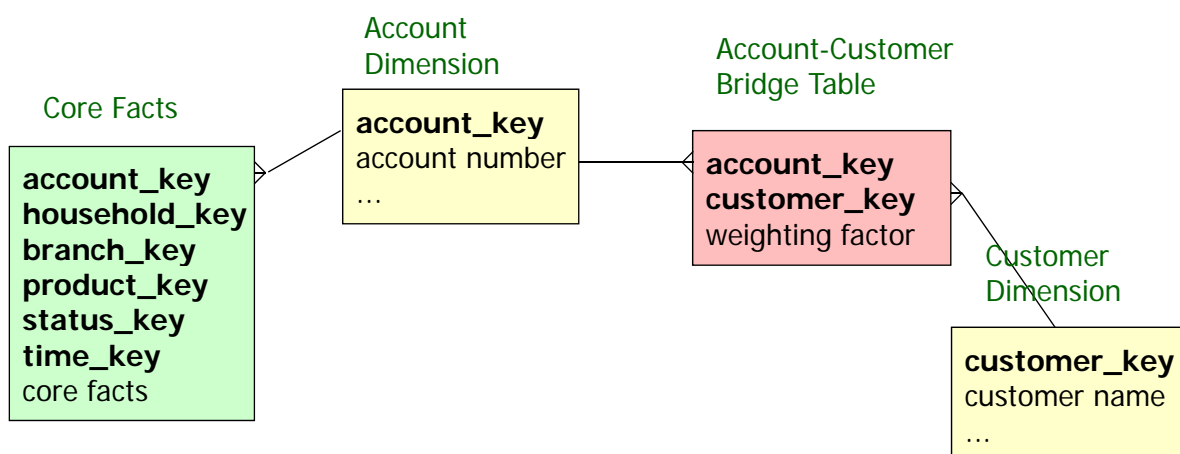
## Dimensioni multivalore

Il problema in esame ha a che fare con una **dimensione multivalore** – o **dimensioni multi-a-molti**

- la relazione tra le dimensioni conto e cliente può essere gestita mediante una tabella “ponte” tra queste dimensioni
- questa tabella contiene, come informazione aggiuntiva, il “peso” normalizzato della relazione tra conto e cliente – con la convenzione che la somma dei pesi dei clienti associati ad un conto sia sempre 1
  - questo peso potrà essere utilizzato per elaborare delle analisi opportunamente “pesate”

## Dimensioni multivalore

Porzione della soluzione



## Dimensioni multivalore

Un altro contesto in cui è “necessario” utilizzare una dimensione multivalore

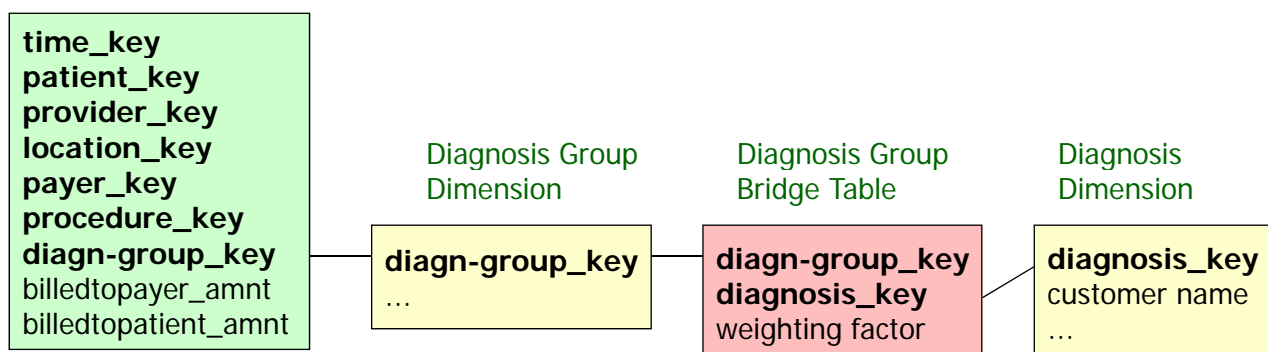
- il processo da analizzare è la fatturazione di spese mediche
  - il dettaglio è quello della “voce di una fattura”
- si vuole analizzare l’incidenza delle diagnosi su tale processo
  - tuttavia, alcune voci di una spesa medica potrebbero essere motivate da un gruppo di diagnosi concomitanti
- potrebbe allora essere opportuno introdurre una dimensione “gruppo di diagnosi” – correlata tramite una tabella ponte alla dimensione diagnosi

## Dimensioni multivalore

Porzione di soluzione

- a ciascuna spesa medica è attribuito un “gruppo di diagnosi”
- si noti che ad uno stesso paziente possono essere attribuiti, in momenti diversi, differenti “gruppi di diagnosi”

Billable Patient  
Encounter Facts



# Dimensioni multivalore

Porzione di soluzione (alternativa)

- basata sull'osservazione che la dimensione "gruppo di diagnosi" è degenere

