

Servizi finanziari (studio di caso)

Luca Cabibbo
aprile 2012

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 riferimento ai relativi titolari individuali – ma anche con riferimento alle “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

Requisiti di busines

Mediante delle interviste, sono stati raccolti i seguenti requisiti

- devono essere memorizzati cinque anni di dati storici – istantanee mensili – per ciascun conto
- esistono diverse tipologie di conti – chiamati *prodotti* nella banca – ogni tipo di conto è caratterizzato da attributi e fatti numerici propri – diversi da tipo di conto a tipo di conto
- ogni conto ha un proprio saldo primario
 - si vogliono analizzare contemporaneamente diversi tipologie di conto e confrontarne i saldi primari
- ogni conto va considerato appartenente a un'unica “famiglia” – ma la relazione conto-famiglia è molto volatile – grazie a “matrimoni” e “divorzi”
- sono di interesse analisi centrate sui dati demografici – sia sulle famiglie che sui titolari individuali dei conti

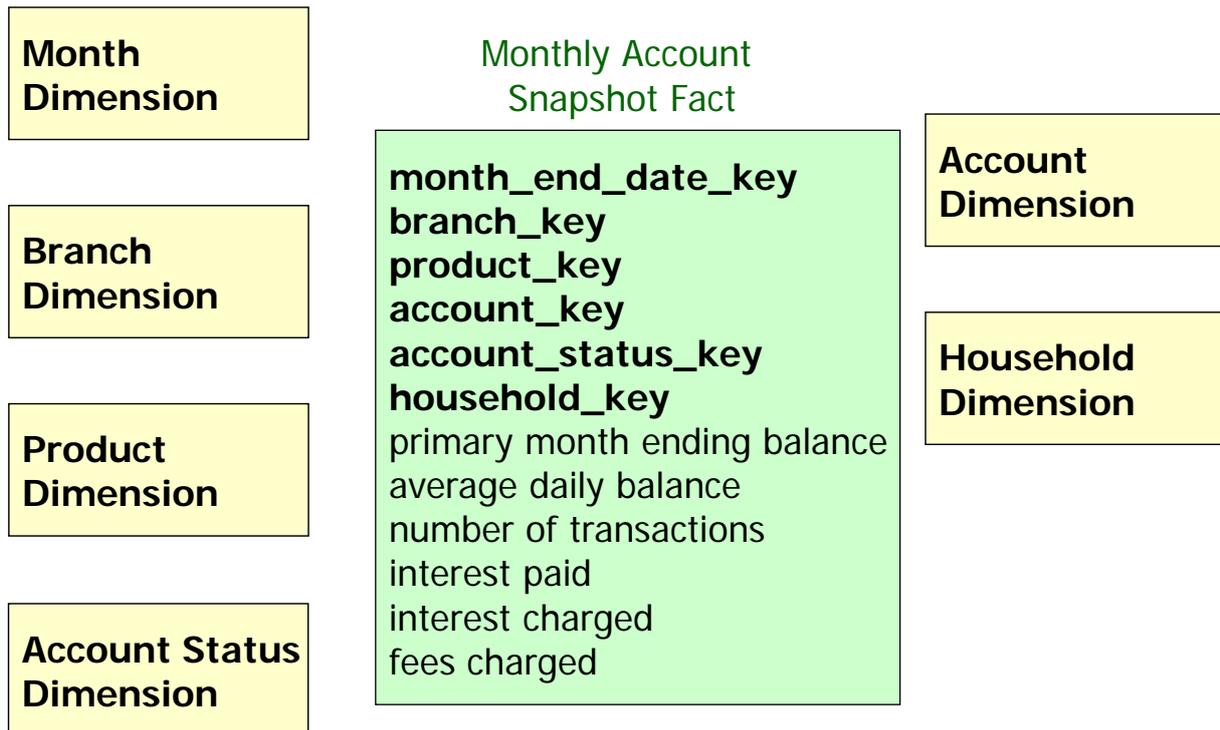
Il data mart dei servizi finanziari

Viene deciso di realizzare un data mart per memorizzare il saldo primario di ciascun conto alla fine di ogni mese

- dunque, il fatto principale è il saldo di un conto
 - ma potrebbero essere utili anche altri fatti/misure
- le dimensioni primarie sono la dimensione conto e la dimensione mese – ad es., nel senso di “ultima data del mese”
 - questa scelta consente di definire proprio la granularità dei dati a cui siamo interessati
- tuttavia, questa scelta non consente di rappresentare la dimensionalità naturale del processo di business di interesse
 - pertanto è necessario considerare ulteriori dimensioni

Data mart dei servizi finanziari

Una possibile scelta per i saldi di fine mese di tutti i conti



5

Servizi finanziari

Luca Cabibbo

Data mart dei servizi finanziari

Sulla scelta delle dimensioni

- prodotto
 - la tipologia di conto – dunque, un tipo di servizio (prodotto) offerto ai clienti della banca
- filiale/agenzia
 - simile alla dimensione negozio – dunque, una dimensione geografica
- stato del conto
 - descrive lo stato di ciascun conto (ad es., attivo, inattivo) alla fine di ogni mese – si tratta sostanzialmente di una minidimensione
- “famiglia”
 - descritta nel seguito

6

Servizi finanziari

Luca Cabibbo

Famiglie e conti

La banca vuole analizzare i suoi prodotti non solo con riferimento ai conti correnti individuali – ma anche con riferimento alle intere unità economiche che li detengono – ovvero le “famiglie” (household)

- la banca è interessata a comprendere il profilo delle famiglie nel loro complesso – ad esempio, per proporre nuovi prodotti alle famiglie
- inoltre vogliono analizzare i loro prodotti con riferimento a dati demografici delle famiglie nel loro complesso – ad es., il reddito del nucleo familiare
- chiaramente questi dati demografici possono cambiare – come può cambiare la composizione delle famiglie – e la banca è interessata a tenere traccia anche di questi cambiamenti
 - questo può motivare l'uso di dimensioni distinte per famiglie e conti

7

Servizi finanziari

Luca Cabibbo

Famiglie e conti

Famiglie e conti sono intuitivamente correlati – tuttavia, la composizione delle famiglie può cambiare nel tempo

- questo può motivare l'uso di dimensioni distinte per famiglie e conti
- la relazione (corrente) tra conti e famiglie può essere rappresentata nella dimensione conto – che riferisce la relativa famiglia
- la relazione (storica) tra conti e famiglie può essere invece rappresentata nella tabella fatti

8

Servizi finanziari

Luca Cabibbo

Dimensioni multivalore

Consideriamo ora un diverso requisito – ad un conto possono essere associati uno, due oppure anche più intestatari o clienti – ad es., John e Mary Smith – *nota: al momento non ci interessano più le “famiglie”*

- come gestire una tale situazione, e 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” – perché violerebbe la granularità della dimensione conto
 - “cliente” non può essere nemmeno una dimensione – perché violerebbe la granularità della tabella fatti

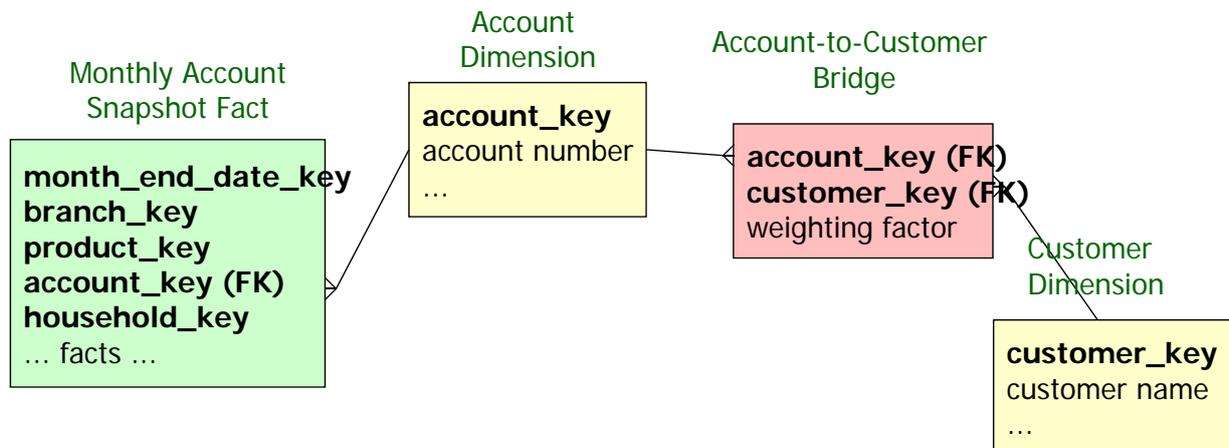
Dimensioni multivalore

Una **dimensione multivalore**

- talvolta, una tabella fatti o una tabella dimensione è caratterizzata da un attributo multivalore – ovvero, un singolo attributo a cui è associato un gruppo di valori
- un attributo multivalore – ovvero, un gruppo di valori – può essere rappresentato tramite una *tabella ponte (bridge table)*
- ad es., la relazione tra le dimensioni conto e cliente può essere gestita mediante una tabella “ponte” tra queste dimensioni
- una tabella ponte può contenere, come informazione aggiuntiva, il “peso” normalizzato della relazione tra i membri collegati – conto e cliente – con la convenzione che la somma dei pesi dei clienti associati a un conto sia sempre 1
 - questo peso potrà essere utilizzato per elaborare delle analisi opportunamente “pesate”

Dimensioni multivalore

Porzione della soluzione



Dimensioni multivalore

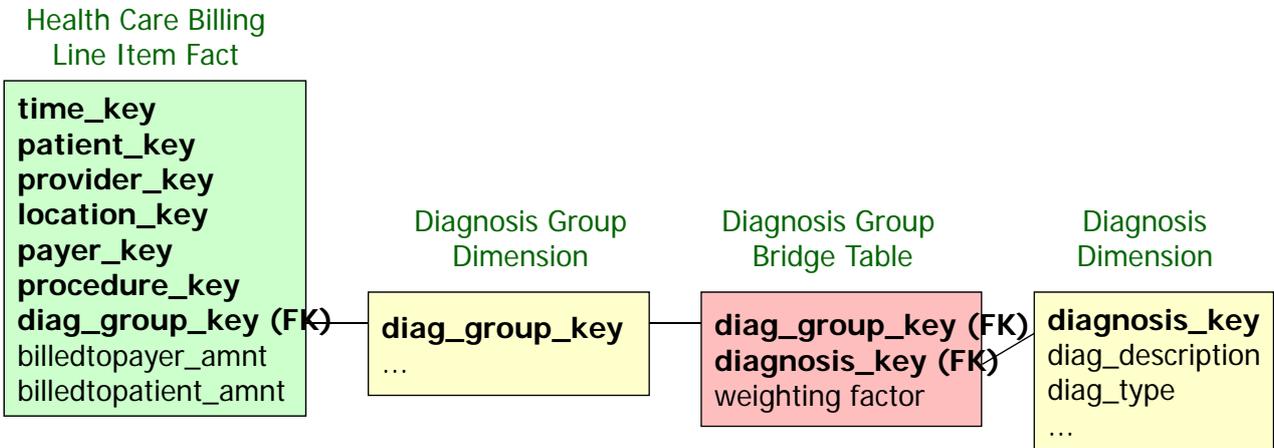
Le dimensioni multivalore possono essere utilizzate anche per collegare ciascuna riga di una tabella fatti con molti membri di una dimensione

- consideriamo ora un processo per la fatturazione di spese mediche – ad esempio, per la gestione dei rimborsi da parte di un'assicurazione
 - la grana della tabella fatti potrebbe essere “per voce di fattura”
- l'assicurazione vuole analizzare l'incidenza delle diagnosi sulle fatture
- 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” – collegata alla dimensione diagnosi tramite una tabella ponte

Dimensioni multivalore

Porzione di soluzione

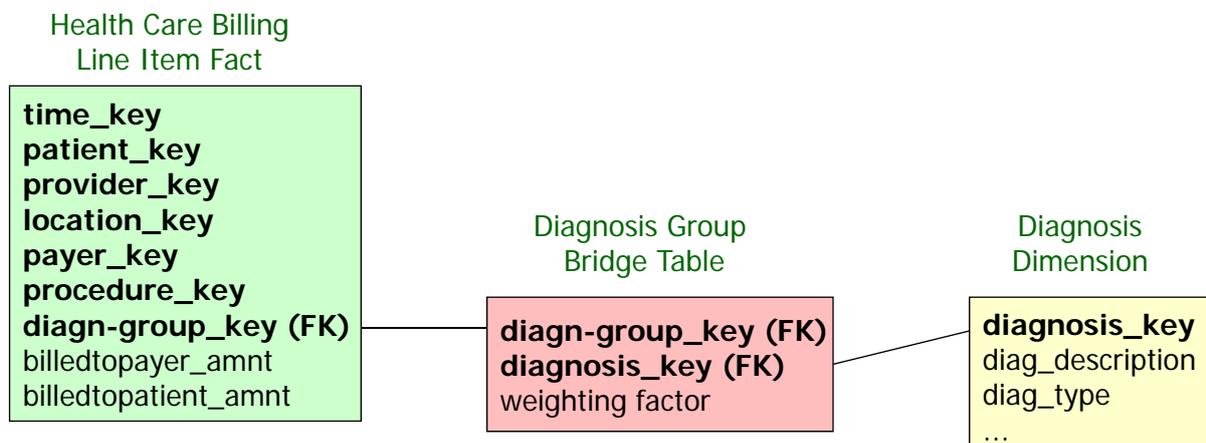
- a ciascuna spesa medica è attribuito un “gruppo di diagnosi”



Dimensioni multivalore

Porzione di soluzione (alternativa)

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



Famiglie di prodotti eterogenei

Un altro requisiti per il data mart dei servizi finanziari

- esistono diverse tipologie di conti – ogni tipo di conto è caratterizzato da attributi e fatti numerici propri – diversi da tipo di conto a tipo di conto
- si parla in questo caso di **prodotti eterogenei** – *heterogeneous products*
- l'eterogeneità riguarda sia gli attributi (descrittivi) che i fatti (numerici) – prodotti diversi sono dissimili rispetto a queste proprietà
 - 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

Come gestire una famiglia di prodotti eterogenei?

- una prima soluzione consiste nel definire una singola tabella – con i fatti e le dimensioni per tutti i tipi di prodotti (conti, in questo caso) di interesse
 - tuttavia, questi fatti e attributi sarebbero spesso non significativi e vuoti – come il formaggio svizzero – con notevole spreco di occupazione di memoria e degrado delle prestazioni
 - inoltre, l'introduzione di nuove tipologie di prodotti sarebbe difficoltosa

Gestione di prodotti eterogenei

Come gestire una famiglia di prodotti eterogenei?

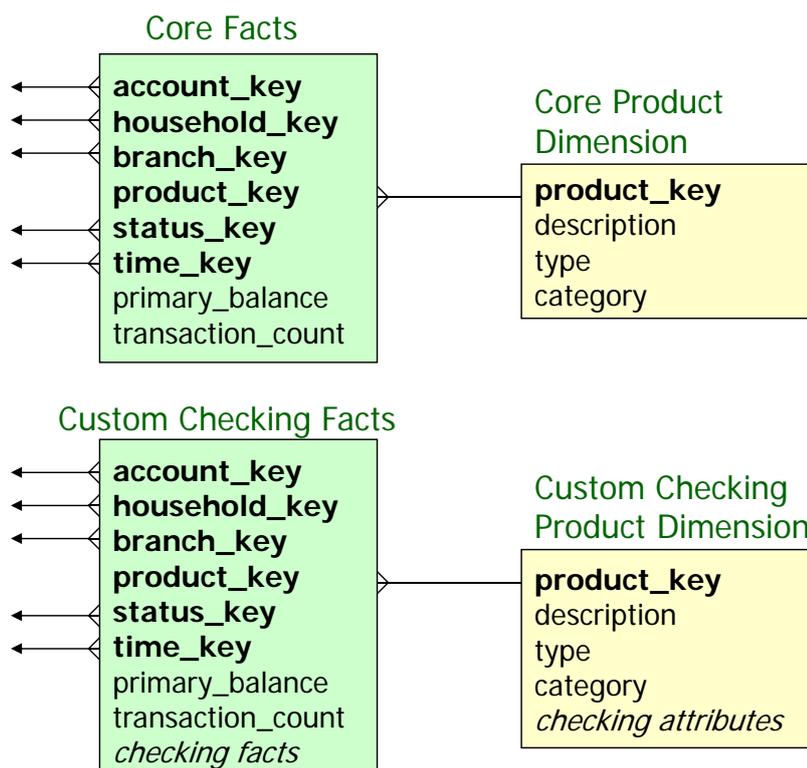
- una seconda soluzione – più efficace – consiste invece nel definire
 - una tabella fatti “core” – con le sue dimensioni – che descrive i fatti comuni a tutti i tipi di prodotti (conti, in questo caso)
 - una tabella fatti distinta per ciascun tipo di prodotto/conto – che rappresenta tutti i fatti e con delle dimensioni “custom” per i prodotti/conti specifici – comprese anche le dimensioni e i fatti già riportati nella tabella fatti “core”
- in questo modo, infatti
 - è possibile analizzare i singoli prodotti individualmente, in dettaglio
 - è possibile analizzare anche l’insieme dei prodotti, in modo integrato, sulla base delle misure condivise da tutti i prodotti

17

Servizi finanziari

Luca Cabibbo

Schema dimensionale per prodotti eterogenei



18

Servizi finanziari

Luca Cabibbo

Schemi dimensionali per i prodotti eterogenei

Il data mart completo dei servizi finanziari comprende

- una tabella fatti e 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 riga della tabella fatti nucleo descrive un insieme di fatti che è riportato anche in una riga di (esattamente) una tabella fatti personalizzata
 - che riporta sia tutte le misure fondamentali (del nucleo) che quelle specifiche per il tipo di prodotto

Analisi di prodotti eterogenei

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

- analisi integrate, 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 specifiche, 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 l'uso di tabelle dimensioni personalizzate potrebbe essere ancora utile

Famiglie di prodotti eterogenei

Una famiglia di schemi dimensionali per prodotti eterogenei va utilizzata nelle organizzazioni che offrono un insieme ampio di prodotti o servizi tramite *un insieme di linee di business distinte* – e per i quali è necessaria *anche una visione integrata, relativa a una base comune di clienti*

- ad esempio, una società tecnologica che vende hardware, software e servizi
- in questo caso, ha senso definire
 - delle tabelle fatti nucleo che guardano l'organizzazione dal punto di vista complessivo dell'offerta di prodotti ai suoi clienti – attraverso tutte le linee di prodotto
 - delle tabelle fatte specifiche per le singole linee di prodotto – che consentono di analizzare in modo dettagliato ciascuna linea di prodotto