

Progettazione dimensionale

Luca Cabibbo
aprile 2012

Progettazione dimensionale

La *progettazione dimensionale* riguarda la progettazione logica dei dati del data warehouse – costituisce la pietra angolare della progettazione dell'intero sistema DW/BI

- è un processo altamente dinamico e iterativo
- l'attività inizia da una comprensione preliminare dei requisiti di business
- vengono identificati i processi di business fondamentali dell'organizzazione, con la loro granularità e le loro dimensioni principali di analisi
- i dati per l'analisi di ciascun processo di business di interesse vengono rappresentati mediante un modello dimensionale (data mart)
- vengono identificate le sorgenti informative da cui estrarre i dati per popolare il DW – viene analizzata la qualità dei dati disponibili e identificate le trasformazioni richieste
- ... segue ...

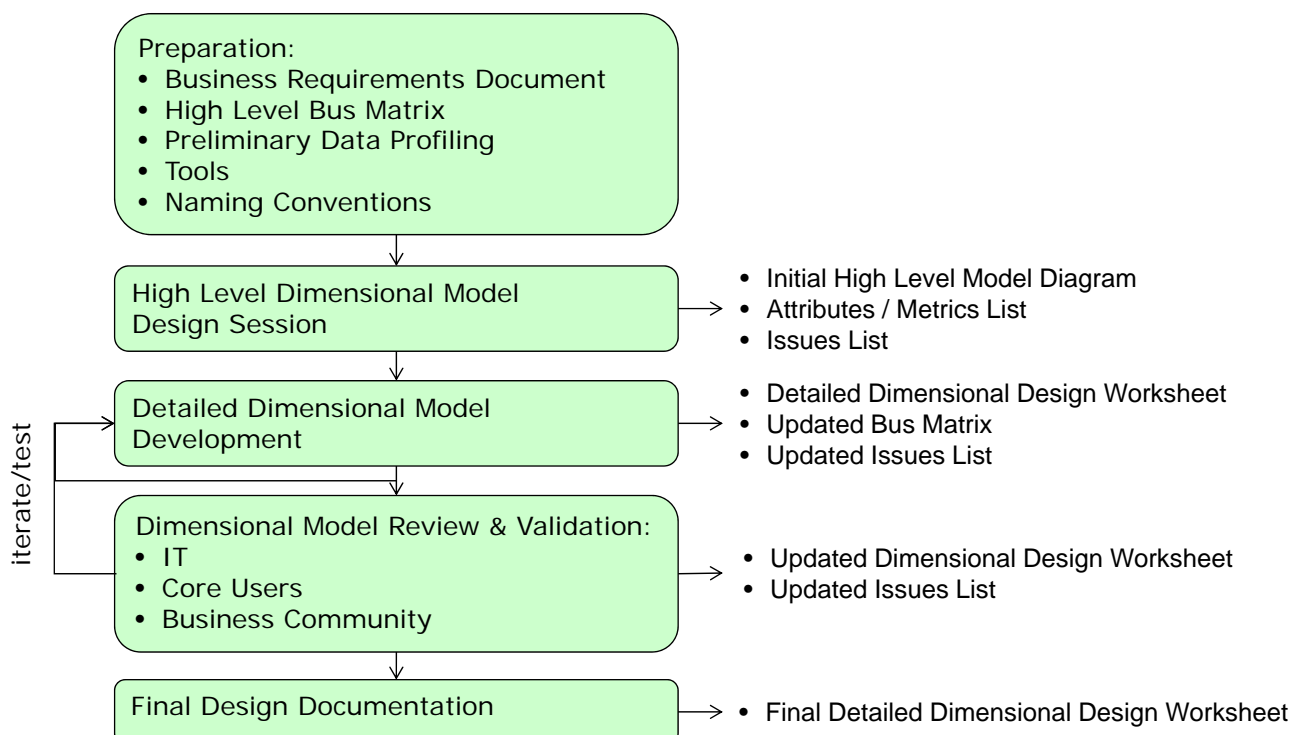
Progettazione dimensionale

La *progettazione dimensionale* riguarda la progettazione logica dei dati del data warehouse – costituisce la pietra angolare della progettazione dell'intero sistema DW/BI

- infine, questi schemi dimensionali vengono rivisti e validati con tutte le parti interessate – soprattutto con gli analisti di business
- infatti, l'obiettivo primario della modellazione dimensionale è definire dei modelli per soddisfare i requisiti di business – e al tempo stesso verificare che i dati necessari a popolare i modelli siano effettivamente disponibili
- la progettazione dimensionale fornisce anche un punto di partenza significativo ai team di sviluppo delle applicazioni BI e del sistema ETL

Processo di prog. dimensionale

Attività principali del processo – iterativo – di prog. dimensionale



- Preparazione

La preparazione ha lo scopo di stabilire le basi per un processo di progettazione efficace ed efficiente – riguarda diverse attività preliminari

Identificazione dei partecipanti – la progettazione dimensionale è un'attività fortemente collaborativa – che richiede figure diverse, tra cui

- data modeler – con capacità di modellazione dimensionale
- analisti di business – esperti dei requisiti di business, delle analisi di business di interesse, delle sorgenti di dati – nonché del glossario del business
- IT – DBA e sviluppatori dei sistemi sorgenti – conoscono le sorgenti di dati
- sviluppatori e architetti dei sistemi BI e ETL – interessati, in modo diverso, a tutte le tematiche discusse

Preparazione

Rivisitare i requisiti

- il team di modellazione deve familiarizzare con i requisiti di business
- per questo, i requisiti vengono rivisti – con l'aiuto soprattutto degli utenti/analisti di business
- bisogna resistere alla tentazione di affidarsi alla conoscenza delle sorgenti di dati delle persone IT – porterebbe a modelli dimensionali guidati dalle sorgenti anziché guidate dal business

- Definizione della matrice a bus

La *matrice a bus del data warehouse* rappresenta tutti i processi di business fondamentali dell'organizzazione e le loro dimensioni principali di analisi

- questa matrice viene definita inizialmente durante l'analisi iniziale dei requisiti di business – verrà poi raffinata in modo iterativo
- questa matrice serve a garantire che i dati nei modelli dimensionali progettati nelle diverse iterazioni/progetti possano essere correlati e integrati nel corso del tempo e dell'intero programma

Definizione della matrice a bus

La matrice a bus consente di porre le basi per la modellazione dimensionale – sulla base di un approccio top-down

- identifica i processi di business dell'organizzazione – ciascun processo di business rappresenta un data mart candidato
 - riporta questi processi di business/data mart sulle righe della matrice a bus
- identifica le dimensioni implicate da tali processi di business/data mart
 - riporta le dimensioni sulle colonne della matrice a bus
- marca nella matrice, per ciascun processo di business/data mart, tutte le dimensioni di interesse
- questa è la singola attività più importante nella realizzazione di un DW – pertanto, prosegui con la progettazione specifica dei diversi data mart solo dopo un accordo preliminare su questa matrice a bus

Esempio – matrice a bus

Esempio di matrice a bus – per la catena del valore della domanda

Esempio – matrice a bus

Esempio di matrice a bus – per la catena del valore della domanda

Magazzino dei prodotti finiti									
Spedizione al centro di distribuzione									
Magazzino del centro di distribuzione									
Spedizione ai negozi di vendita									
Magazzino dei negozi di vendita									
Vendita al dettaglio									

Esempio – matrice a bus

Esempio di matrice a bus – per la catena del valore della domanda

	tempo	prodotto (SKU)	magazzino	centro di distribuzione	contratto (o accordo commerciale o promozione)	modalità di consegna (compreso il vettore)	negozio	promozione	cliente (se disponibile)
Magazzino dei prodotti finiti									
Spedizione al centro di distribuzione									
Magazzino del centro di distribuzione									
Spedizione ai negozi di vendita									
Magazzino dei negozi di vendita									
Vendita al dettaglio									

Esempio – matrice a bus

Esempio di matrice a bus – per la catena del valore della domanda

	tempo	prodotto (SKU)	magazzino	centro di distribuzione	contratto (o accordo commerciale o promozione)	modalità di consegna (compreso il vettore)	negozio	promozione	cliente (se disponibile)
Magazzino dei prodotti finiti	X	X	X						
Spedizione al centro di distribuzione	X	X	X	X	X	X			
Magazzino del centro di distribuzione	X	X		X					
Spedizione ai negozi di vendita	X	X		X	X	X	X		
Magazzino dei negozi di vendita	X	X					X		
Vendita al dettaglio	X	X					X	X	X

Identificazione dei processi/data mart candidati

Alcune linee guida per l'identificazione dei processo di business/data mart candidati

- un data mart deve rappresentare una collezione di fatti correlati che possono/devono essere analizzati insieme, a fini decisionali
- un data mart può essere relativo a uno oppure più attività/processi di business
 - un *data mart di primo livello* ha origine in un singolo processo di business – e una singola sorgente informativa
 - un *data mart consolidato* è invece relativo a più processi e/o ha dati derivanti da più sorgenti informative
- i data mart possono essere (parzialmente) sovrapposti
- in una grande organizzazione è comune identificare circa 10-30 data mart

Esempio – una grande compagnia telefonica

Data mart a sorgente singola

- fatturazione clienti (residenziali e commerciali)
- gestione ordini
- gestione dei malfunzionamenti
- pubblicità sulle pagine gialle
- servizio clienti e informazioni sulle fatture
- offerte promozionali e comunicazioni ai clienti
- dettaglio delle chiamate dal punto di vista della fatturazione
- dettaglio delle chiamate dal punto di vista del carico della rete telefonica
- inventario clienti
- inventario della rete telefonica
- ...

Identificazione delle dimensioni candidate

Scelti i data mart di interesse, si procede identificando e elencando le dimensioni (candidate) di interesse

- bisogna progettare un insieme di dimensioni da usare in modo conforme (o conformato) in tutti i data mart del data warehouse
- si può iniziare identificando le dimensioni di interesse per ciascun processo di business/data mart
- le relazioni tra processi di business/data mart e dimensioni candidate vengono mostrate nella matrice a bus

Esempio – una grande compagnia telefonica

Dimensioni per il data mart della fatturazione clienti

- tempo (data di fatturazione)
- cliente (residenziale o commerciale)
- servizio
- tariffa (compresa promozione)
- fornitore di servizi locali

Dimensioni per il data mart del dettaglio delle chiamate dal punto di vista della fatturazione

- chiamante
- chiamato
- fornitore di servizi non locali

Esempio – una grande compagnia telefonica

	Time	Customer	Service	Rate Category	Local Service Provider	Calling Party	Called Party	Long-Distance Provider	International Organization	Employee	Location	Equipment Type	Supplier	Item Supplied	Weather	Account Status
Customer billing	X	X	X	X	X			X			X					X
Service orders	X	X	X		X			X	X	X	X	X			X	X
Trouble reports	X	X	X		X	X		X	X	X	X	X	X	X	X	X
Yellow Page Ads	X	X		X		X			X	X	X					X
Customer Inquiries	X	X	X	X	X	X		X	X	X	X				X	X
Promotions & Comm'n	X	X	X	X	X	X		X	X	X	X	X	X	X		X
Billing Call Detail	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X
Network Call Detail	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X
Customer Inventory	X	X	X	X	X			X	X		X	X	X	X		X
Network Inventory	X		X						X	X	X	X	X	X		
Real Estate	X								X	X	X	X				
Labor & Payroll	X								X	X	X					
Computer Charges	X	X	X		X			X	X	X	X	X	X	X		
Purchase Orders	X								X	X	X	X	X	X		
Supplier Deliveries	X								X	X	X	X	X	X		
Combined Fields Ops.	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X
Customer Reln. Mgmnt.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Customer Profit	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

17

Progettazione dimensionale

Luca Cabibbo

La matrice dell'architettura a bus

La creazione di dimensioni conformi è frutto di una decisione tanto politica quanto tecnica – e deve essere sostenuta dai livelli esecutivi più alti

- la matrice dell'architettura a bus è una specie di mappa del processo politico che è stato intrapreso per far in modo che tutti i gruppi coinvolti siano d'accordo sulle definizioni comuni delle dimensioni

18

Progettazione dimensionale

Luca Cabibbo

- Progettazione dei modelli dimensionali

A questo punto, è possibile con la progettazione dei modelli dimensionali per i singoli processi di business

- probabilmente, si inizierà da data mart di primo livello – singolo processo di business/singola sorgente dati
- per poi muoversi verso data mart più complessi, consolidati
- la lavorazione di ciascun modello dimensionale avviene in modo iterativo – in più sessioni di progettazione – e probabilmente richiederà circa 3-4 settimane

Progettazione di un modello dimensionale

La progettazione di ogni modello dimensionale è basata sul processo di modellazione a quattro passi

- scelta del *processo di business* da modellare
- dichiarazione della *grana* del processo di business
- scelta delle *dimensioni* da cui dipende ciascuna tabella fatti
- identificazione dei *fatti* misurabili che popoleranno ciascuna tabella fatti

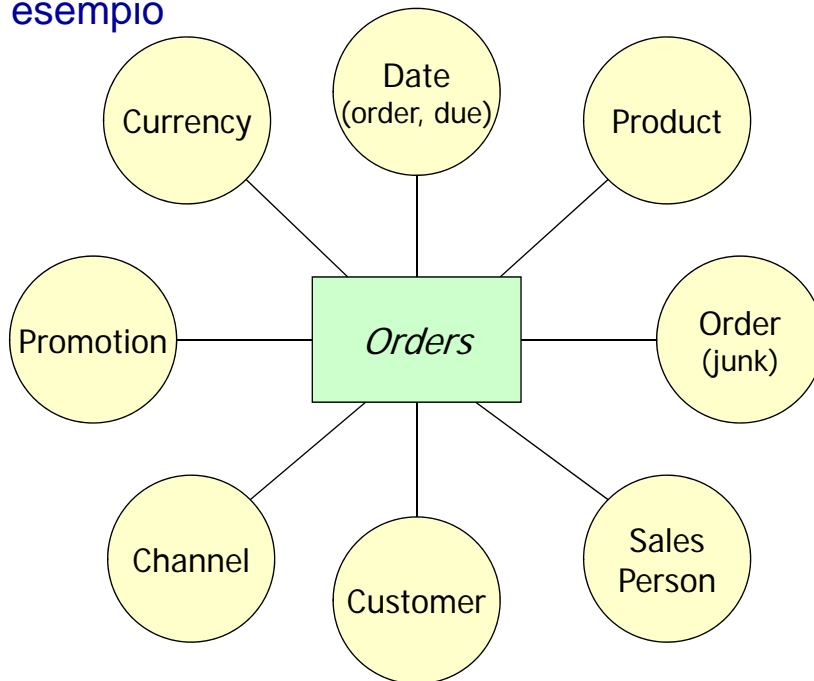
Alcune di queste scelte sono state già fatte – in modo preliminare – nella definizione della matrice a bus

- queste scelte vanno certamente raffinate
 - ad es., è comune che quella che sembrava una singola dimensione vada rappresentata con due o tre dimensioni
- le tecniche di modellazione dimensionale studiate dovrebbero aiutare proprio nello svolgimento di queste attività

Modelli dimensionale di alto livello

Uno strumento per iniziare la comunicazione all'interno del team di progettazione – e facilitare la comunicazione

- ad esempio



Sviluppo dei modelli dimensionali dettagliati

Successivamente, il progetto dei modelli dimensionali va dettagliato

- definizione degli attributi delle dimensioni
- definizione dei fatti/misure
- scelta della strategia di gestione dei cambiamenti lenti, per ciascuna dimensione
- altre scelte di rappresentazione
 - minidimensioni, dimensioni e fatti eterogenei, aggregazioni
- durata storica del data warehouse
 - quanti dati storici devono essere rappresentati nel data warehouse? con quale grana?
- pianificazione del caricamento incrementale
 - con che periodicità deve essere aggiornato il data warehouse?

Stabilire dimensioni conformi

Nella progettazione di dettaglio dei modelli dimensionali, è necessario stabilire – ovvero, progettare in dettaglio – anche le dimensioni conformi – con il loro significato preciso e i loro attributi

- è necessario trovare un accordo a livello dell'intera organizzazione sui nomi delle tabelle e degli attributi
 - i nomi sono più importanti di quanto possa sembrare a prima vista
- ma soprattutto sul loro significato – ovvero sulle loro definizioni e descrizioni
- analisti diversi saranno probabilmente interessati alle stesse dimensioni – ma con insiemi di attributi diversi
 - attenzione, le operazioni di drill-across, su dati aggregati, sono significative solo con riferimento ad attributi comuni

Stabilire fatti conformi

Anche per i fatti vanno progettati in modo conforme – ovvero con nome dettagliato e significato definito con precisione

- l'accordo su un glossario dei termini – a livello dell'intera organizzazione – è un altro passaggio importante
- non bisogna limitarsi a identificare solo fatti base – ma anche fatti derivati – e le regole/formule di derivazione
 - sui fatti derivati l'accordo potrebbe essere più difficile – ma più importante nella definizione di metriche non ambigue attraverso l'intera organizzazione
 - inoltre, per i fatti derivati, l'identificazione delle regole di aggregazione non può essere lasciata all'utente finale – poiché potrebbe involontariamente scrivere cose non corrette o poco significative

- Analisi delle sorgenti informative

La progettazione di dettaglio dei modelli dimensionali non può prescindere dall'effettiva disponibilità e qualità dei dati nelle sorgenti informative – pertanto, è necessario

- identificare, comprendere e analizzare le sorgenti informative a disposizione
 - quali sono queste sorgenti informative? quali dati vi sono rappresentati? con quali qualità (accuratezza, precisione, tempestività, ...)?
 - in che modo si muovono i dati attraverso le diverse sorgenti informative? uno stesso dato può infatti attraversare più sistemi, per essere elaborato in più modi – da quale sorgente estrarre ciascun dato?
 - quali i sistemi e le tecnologie coinvolte? che ne è responsabile?

Progettazione della trasformazione dei dati

Per ciascun elemento (riga/attributo) del data warehouse è necessario

- identificare/scegliere la sorgente informativa da cui quel dato andrà estratto
- identificare/progettare la trasformazione necessaria a calcolare quell'elemento

- Validazione

Ogni modello dimensionale, predisposto dai data modeler, va opportunamente rivisto e validato

- questa revisione avviene tramite una serie di incontri successivi – uno per ciascun diverso livello di comprensione tecnica o di business dei dati di interesse
- almeno tre gruppi di revisione
 - DBA e sviluppatori dei sistemi sorgenti – in grado di individuare rapidamente errori nei modelli – tecnici o relativi alla modellazione del processo di business
 - analisti di business esperti – in grado di apprezzare molti dettagli nei modelli
 - utenti di business finali – questo incontro non è solo di revisione – ma ha spesso anche un ruolo “educativo” – il data mart viene presentato agli utenti finali – buona parte del tempo sarà dedicata a mostrare quali analisi possono essere eseguite, e come