

# Ciclo di vita dimensionale

Luca Cabibbo  
aprile 2012

## Ciclo di vita dimensionale

Il **ciclo di vita dimensionale – Business Dimensional Lifecycle**, chiamato anche **Kimball Lifecycle** – descrive il framework complessivo che lega le diverse attività dello sviluppo di un sistema DW/BI

- è una metodologia completa per la progettazione e realizzazione di un sistema DW
- fornisce il contesto di riferimento per la progettazione e realizzazione di data warehouse dimensionali

## Caratteristiche

### Caratteristiche del ciclo di vita dimensionale

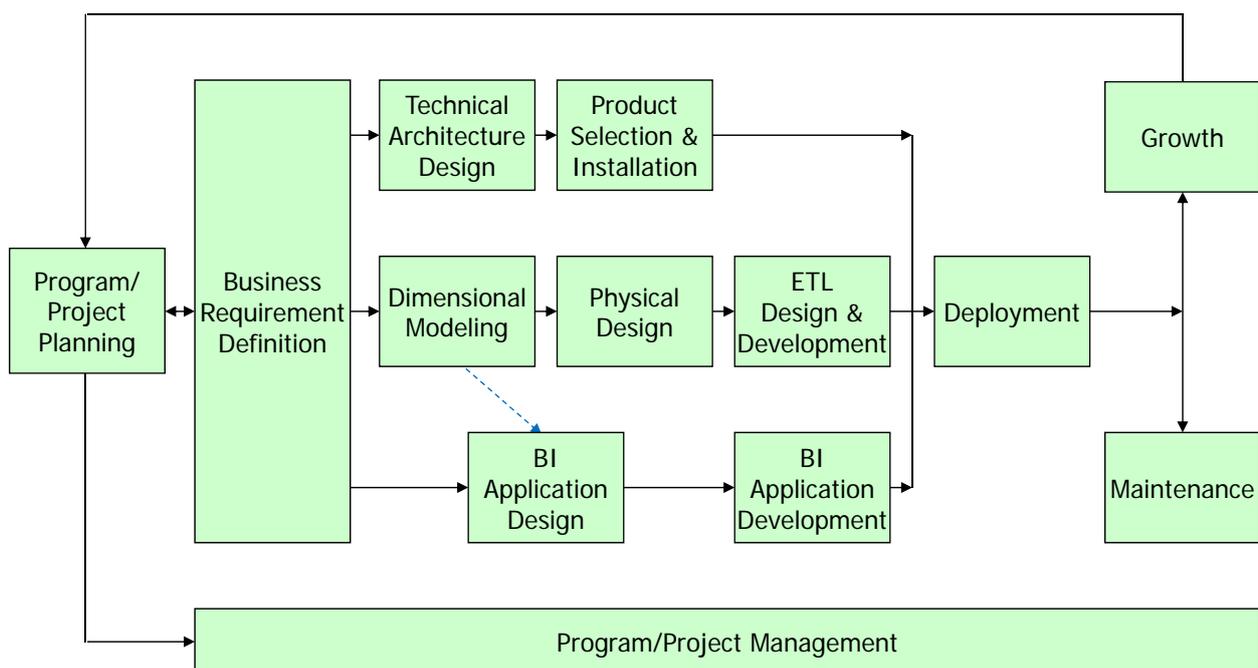
- un sistema DW/BI viene costruito a partire dagli utenti di business e dai loro bisogni effettivi – con questi obiettivi in mente, si lavora “all’indietro” attraverso i report, le applicazioni, le basi di dati e il software – fino agli strati più fisici del sistema
  - si tratta dunque di un approccio fortemente guidato dal business e dall’utente – e non dalle tecnologie (come avviene in altre metodologie)
- è un processo iterativo – in realtà, incrementale
  - l’intero sistema DW/BI (“programma”) viene sviluppato in una serie di iterazioni/cicli/incrementi (“progetti”) – e non come un unico Big Bang galattico
  - ciascun “progetto” ha lo scopo di fornire capacità di analisi (aggiuntive) di valore per gli utenti di business
  - consistenza del sistema garantita mediante un’architettura a bus del data warehouse

3

Ciclo di vita dimensionale

Luca Cabibbo

## Attività



4

Ciclo di vita dimensionale

Luca Cabibbo

## Pianificazione del progetto e del programma

Il ciclo di vita inizia con la *pianificazione* del programma e del progetto

- per *programma* si intende l'attività complessiva di sviluppo di un sistema DW/BI completo, che procede tramite diverse iterazioni
- per *progetto* si intende una singola iterazione/ ciclo/ incremento del ciclo di vita, dall'avvio al rilascio
- alcune attività della pianificazione
  - verifica di "readiness" dell'organizzazione
  - analisi preliminare dei requisiti e della portata del programma o del progetto
  - valutazioni economiche
  - identificazione delle risorse
  - formazione dei gruppi di lavoro
  - definizione del piano di sviluppo (iterativo)

## Gestione del progetto e del programma

Per tutta la sua durata, il ciclo di vita richiede un'attività di *gestione del progetto*

- le attività della gestione del progetto/programma si focalizzano sul monitoraggio dello stato del progetto, controllo dei cambiamenti
- un aspetto critico nel raggiungimento degli obiettivi di un sistema DW/BI è la comunicazione continua
  - per questo, la gestione comprende anche lo sviluppo di un piano di comunicazione completo – che riguarda sia le parti interessate di business che quelle IT

## Raccolta e analisi dei requisiti

Un sistema DW/BI viene costruito a partire dagli utenti di business e dai loro bisogni effettivi

- una definizione efficace dei *requisiti di business* è cruciale – poiché stabilisce le fondamenta per tutte le attività del ciclo di vita successive
- pertanto, è fondamentale una comprensione accurata delle necessità dei diversi utenti di business e dei loro requisiti
- due piani principali
  - comprendere i bisogni e le priorità degli utenti a livello dell'intero programma di sviluppo – probabilmente, a livello dell'intera organizzazione
  - comprendere in dettaglio i requisiti di business nel contesto di un singolo progetto – la cui portata può essere abbastanza ristretta
- importante coinvolgere anche le parti interessate dell'IT

## Raccolta e analisi dei requisiti (programma)

Alcune domande per comprendere i requisiti di business al livello dell'intero programma

- quali gli obiettivi della tua organizzazione? quali sono i processi di business fondamentali della tua organizzazione?
- come misuri il successo? come fai a capire se stai raggiungendo i tuoi obiettivi? quanto spesso ti misuri?
- quali sono le questioni chiave del tuo business oggi? come le affronti? che impatto hanno sulla tua organizzazione?
- come identifichi i problemi o capisci che ci potrebbero essere quali opportunità ci sono nel tuo business, che potrebbero avere un impatto importante per la tua organizzazione? con quale impatto economico?
- qual è la tua visione sul come sfruttare al meglio le informazioni nella tua organizzazione?

## Raccolta e analisi dei requisiti (progetto)

Ciascuna iterazione/progetto del ciclo di vita si focalizzerà invece su un singolo processo di business (o iniziativa strategica di business) dell'organizzazione – alcune domande importanti

- quali sono i prodotti di questo processo? come si possono distinguere i diversi prodotti? quanto spesso cambiano?
- quali analisi vengono fatte per valutare questo processo? con quali dati? quale uso fai di questi dati quando ne vieni in possesso? quali analisi vorresti invece poter eseguire?
- che report usi/prepari? quali i dati più importanti dei report? quali report vorresti invece? quali report non guardi mai?
- quanta informazione storica è necessaria? quanto spesso deve essere aggiornata? per quanto tempo deve essere mantenuta?
- quali opportunità esistono per migliorare in modo significativo il tuo business sulla base di un accesso migliore alle informazioni? con quale impatto economico?

9

Ciclo di vita dimensionale

Luca Cabibbo

## Progettazione e sviluppo del data warehouse

La progettazione – nonché il relativo sviluppo – del sistema DW/DI avviene su tre tracce concorrenti

- *progettazione dei dati*
  - progettazione dimensionale (logica) e fisica del data warehouse
  - progettazione e sviluppo del sottosistema ETL
- *progettazione tecnologica*
  - progettazione dell'architettura tecnica del data warehouse
  - selezione dei prodotti e loro installazione
- *progettazione delle applicazioni BI*
  - specifica delle applicazioni BI
  - sviluppo delle applicazioni BI

10

Ciclo di vita dimensionale

Luca Cabibbo

## 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

- i dati per l'analisi di ciascun processo di business di interesse nel progetto corrente vengono rappresentati mediante un *modello dimensionale* (o data mart) per quel processo di business
- 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 – e 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

## Progettazione fisica

Definizione delle *strutture fisiche* adeguate a una rappresentazione efficiente dello schema logico del data warehouse

- definizione dello schema fisico dei dati – sia a livello relazionale che OLAP
- progettazione delle strategie di tuning delle prestazioni – dall'indicizzazione al partizionamento dei dati alle aggregazioni

## Progetto e sviluppo del sottosistema ETL

### Progetto e sviluppo del *sottosistema ETL*

- ha a che fare con
  - estrazione dei dati dalle loro sorgenti informative
  - la qualità dei dati – pulizia, correlazione, conformazione, storicizzazione, ...
  - caricamento iniziale e caricamenti incrementali dei dati nel data warehouse
- una delle attività più complesse con cui si deve confrontare il team di sviluppo
  - mediamente, il 70% dei rischi e delle sforzo di sviluppo in un progetto DW/BI sono relativi a questa attività

## Progetto e sviluppo del sottosistema ETL

### Progetto e sviluppo del *sottosistema ETL*

- lavorazione delle dimensioni
  - estrazione dei dati delle dimensioni dalle loro sorgenti informative
  - pulizia dei valori degli attributi – parsing, decodifica, gestione di dati mancanti o non validi
  - gestione dei cambiamenti delle dimensioni
  - assegnazione delle chiavi surrogate
  - caricamento delle nuove dimensioni – e pubblicazione della loro avvenuta revisione

## Progetto e sviluppo del sottosistema ETL

### Progetto e sviluppo del *sottosistema ETL*

- lavorazione dei fatti
  - estrazione dei dati sui fatti dalle loro sorgenti informative
  - ricezione delle dimensioni aggiornate
  - separazione dei fatti in base alla loro granularità
  - trasformazione dei fatti, come richiesto
  - sostituzione delle chiavi nelle sorgenti informative con le chiavi surrogate delle dimensioni
  - aggiunta di nuove chiavi per dimensioni supplementari – che descrivono il contesto dei fatti
  - verifica di qualità della tabella fatti
  - costruzione o aggiornamento di tabelle fatti aggregate
  - caricamento dei dati nelle tabelle fatti – e pubblicazione del loro avvenuto aggiornamento

## Progettazione tecnologica

### Progettazione dell'*architettura tecnica* globale del sistema DW/BI

- l'ambiente di un sistema DW/BI richiede l'integrazione di numerosi sistemi e tecnologie
  - i sistemi (legacy) su cui risiedono le sorgenti informative, su cui viene parzialmente eseguita la preparazione dei dati
  - i sistemi (solitamente aperti e basati su tecnologia relazionale) su cui risiedono i server di presentazione
  - i sistemi client
- l'architettura tecnica è influenzata da tre fattori
  - requisiti di business – l'ambiente tecnico attuale – le direzioni strategie pianificate per l'ambiente tecnico

### Selezione dei prodotti e loro installazione

- scelta degli ambienti hardware e software

## Progettazione delle applicazioni BI

La “front room” del sistema DW/BI è composta da un insieme di *applicazioni di business intelligence (BI)* – utilizzate direttamente dagli utenti di business del DW/BI

- gli utenti del sistema possono avere necessità di analisi molto diverse tra loro
  - per questo, sono necessari diversi tipi di applicazioni BI – che variano da semplici report standard parametrizzati a sistemi complessi di analisi dei dati
- in generale, le applicazioni BI supportano un processo ciclico comune nelle analisi di business
  - dal monitoraggio dei processi di business all’identificazione di un problema o di un’opportunità, dal determinare l’azione da intraprendere al monitorare il risultato di quell’azione

## Rilascio

Le tre tracce concorrenti di progettazione (dei dati, tecnologica e delle applicazioni BI) convergono nella fase di *deployment*

- installazione del sistema
- integrazione dei sottosistemi e testing
- formazione degli utenti
- documentazione e supporto agli utenti

## Manutenzione e crescita

La *manutenzione* comprende

- monitoraggio delle qualità del sistema
  - misurazione delle prestazioni e dell'uso
- ottimizzazione tecnica
  - tuning delle prestazioni, manutenzione degli indici, backup di sistema

Un sistema DW/BI realizzato con il ciclo di vita dimensionale, se ben fatto, è pronto a evolvere e a crescere – con nuovi progetti, per fornire sempre più valore ai suoi utenti

- l'*evoluzione* di un data warehouse è indice di successo, non di fallimento
- ciascun progetto riguarda di solito il miglioramento dei modelli dimensionali esistenti, ma anche e soprattutto lo sviluppo di modelli dimensionali per ulteriori processi di business