

Data Warehousing

Parte II

Analisi multidimensionale

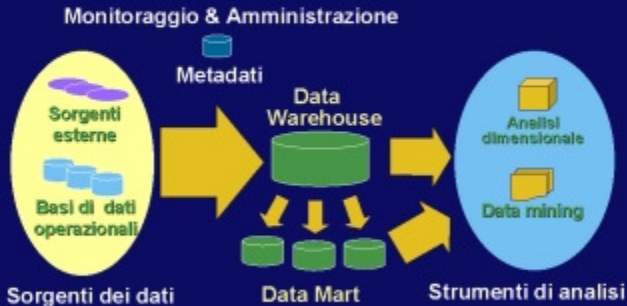
Argomenti della lezione

- ▶ richiami sul data warehousing
- ▶ organizzazione di un data warehouse
- ▶ l'analisi multidimensionale
- ▶ data warehousing e internet
- ▶ strumenti per il warehousing

Un data warehouse è una base di dati con delle caratteristiche “peculiari”

- ▶ **utilizzata per il supporto alle decisioni**
- ▶ **integrata**
- ▶ **dati storici e aggregati**
- ▶ **fuori linea**
- ▶ **separata dalle basi di dati operazionali**

Architettura per il data warehousing



Rappresentazione dei dati

Nelle sorgenti informative

- ▶ dipartimentale
- ▶ es: vendita, produzione, marketing, ...

Nel data warehouse

- ▶ aziendale
- ▶ es: prodotti, clienti, fornitori, ...

Rappresentazioni dei dati

Nei data mart:

- ▶ settoriale
- ▶ es: dati relativi al marketing

Negli strumenti di analisi:

- ▶ individuale
- ▶ es: vendite negli ultimi cinque anni

Popolazione del data warehouse

Attività necessarie:

- ▶ estrazione
- ▶ trasformazione
- ▶ caricamento
- ▶ refresh

I metadati sono mantenuti per il supporto di queste attività

Organizzazione dei dati

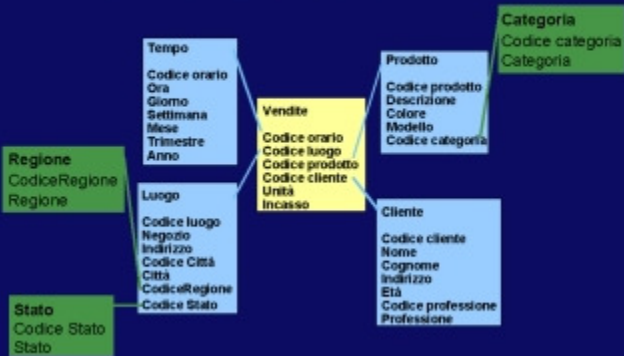
Un data warehouse è tipicamente una base di dati relazionale organizzate in una forma particolare:

- ▶ star schema
- ▶ snowflake schema
- ▶ altre varianti

Organizzazione “star”



Organizzazione “snowflake”



Differenze con basi di dati operazionali

- ▶ Le tabelle in una base di dati sono normalizzate, in un warehouse sono tipicamente denormalizzate
- ▶ Le tabelle in una base di dati sono numerose e omogenee, in un warehouse sono poche e si possono classificare in:
 - fact tables
 - dimension tables

Rappresentazione multidimensionale dei dati

- ▶ Rappresentazione che favorisce l'analisi
- ▶ Concetti rilevanti
 - **fatto**: concetto sul quale centrare l'analisi
 - **misura**: proprietà atomica di un fatto da analizzare
 - **dimensione**: prospettiva lungo la quale effettuare l'analisi

Esempi di fatti/misure/dimensioni

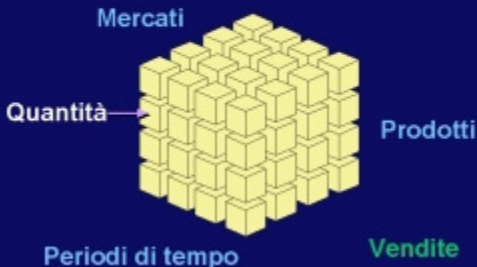
Catena di negozi

- ▶ vendita
- ▶ quantità venduta, incasso
- ▶ prodotto, tempo, zona

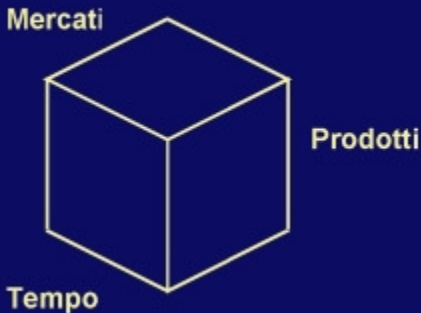
Compagnia telefonica

- ▶ telefonata
- ▶ costo, durata
- ▶ zona chiamante, zona chiamato, tempo

Rappresentazione multidimensionale dei dati



Viste su dati multidimensionali



Al manager regionale interessa la vendita dei prodotti in tutti i periodi nei propri mercati

Mercati

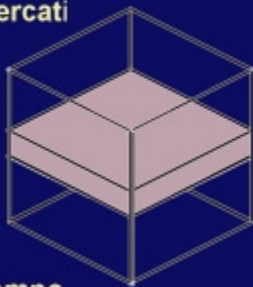


Prodotti

Tempo

Al manager di prodotto interessa la vendita di un prodotto in tutti i periodi e in tutti i mercati

Mercati



Prodotti

Tempo

Al manager finanziario interessa la vendita dei prodotti in tutti i mercati relativamente al periodo corrente e quello precedente

Mercati

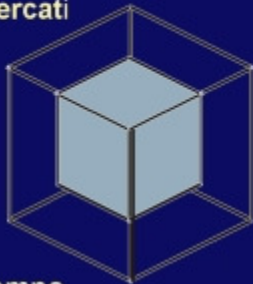


Prodotti

Tempo

Al manager strategico interessa una categoria di prodotti, una area regionale e un orizzonte temporale medio

Mercati



Prodotti

Tempo

Dimensioni e gerarchie di livelli

Ciascuna dimensione è organizzata in una gerarchia che rappresenta i possibili livelli di aggregazione per i dati



Analisi multidimensionale

Lo scopo è individuare degli andamenti inattesi, regolarità o irregolarità

- ▶ **browsing dei dati**
- ▶ **analisi statistiche e previsioni**
- ▶ **analisi what-if**

Operazioni su dati multidimensionali

- ▶ **Roll up:** aggrega i dati
 - volume di vendita totale dello scorso anno per categoria di prodotto e regione
- ▶ **Drill down:** disaggrega i dati
 - per una particolare categoria di prodotto e regione, mostra le vendite giornaliere dettagliate per ciascun negozio
- ▶ **Slice & dice:** seleziona e proietta
- ▶ **Pivot:** re-orienta il cubo

Visualizzazione dei dati

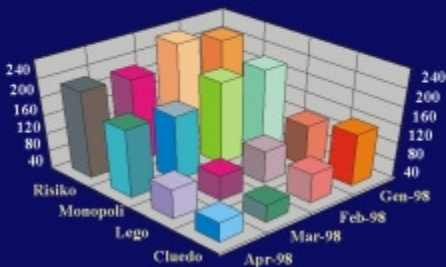
I dati vengono infine visualizzati in veste grafica, in maniera da essere facilmente comprensibili.

Si fa uso di:

- ▶ **tabelle, istogrammi, grafici, torte, superfici 3D, bolle, area in pila, ecc.**

Esempio di visualizzazione

Vendite mensili giocattoli a Roma

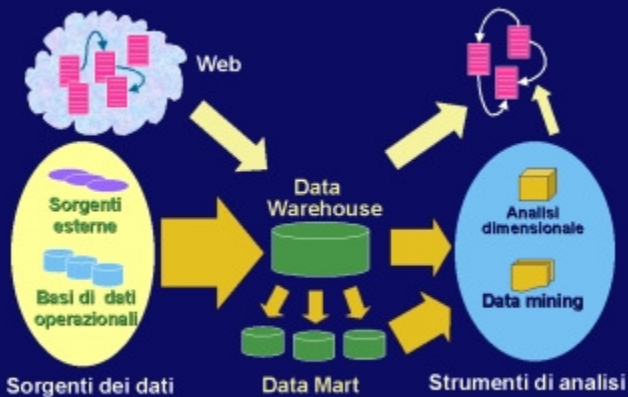


Data warehousing e internet

Internet e data warehouse si integrano in maniera naturale

Due direzioni:

- ▶ **pubblicazione dei dati attraverso il web**
- ▶ **uso del Web come sorgente applicativa**



Strumenti per Warehousing

Disponibili diverse soluzioni:

- ▶ strumenti per la progettazione
- ▶ strumenti per l'estrazione dei dati
- ▶ strumenti per la trasformazione e il caricamento
- ▶ server relazionali per il data warehouse
- ▶ client per l'analisi multidimensionale
- ▶ strumenti per il data mining

Componenti di base

I **server OLAP** sono DBMS speciali

- ▶ server OLAP relazionali (ROLAP)
- ▶ server OLAP multidimensionali (MOLAP)

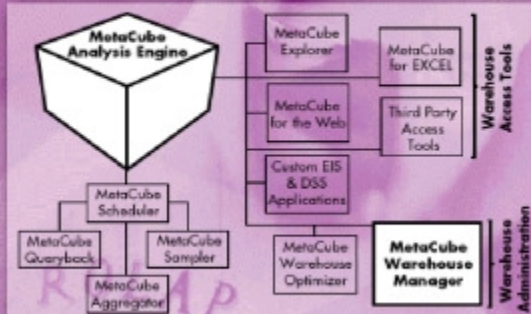
I **client OLAP** sono basati su interfacce grafiche tipo foglio elettronico

- ▶ manipolazione di dati
- ▶ visualizzazione dei risultati

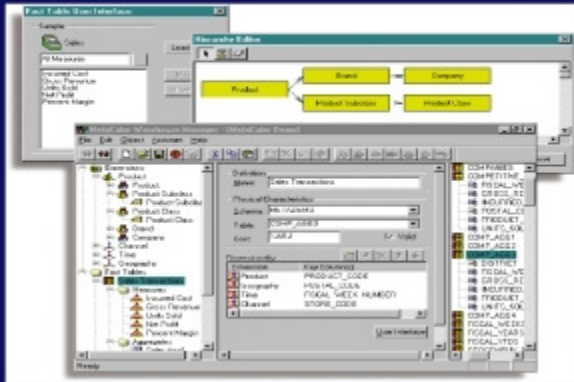
Informix

- ▶ **Warehouse manager** — per il progetto e la gestione di un Data Warehouse
- ▶ **Explorer** — strumento di supporto alle decisioni dotato di interfaccia grafica
- ▶ **MC/Excel** — interfaccia verso MS Excel
- ▶ **Aggregator** — per la creazione e la gestione di aggregazioni sul Data Warehouse

Architettura di MetaCube



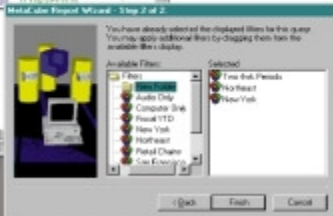
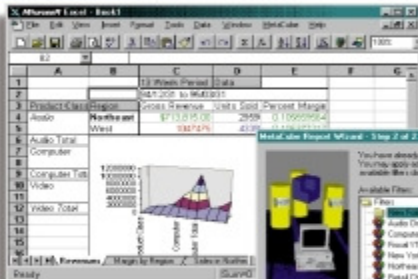
MetaCube Warehouse Manager



MetaCube Explorer



MetaCube per MS Excel



Oracle Data Warehouse

- ▶ **Designer/2000**: strumento CASE per il progetto di un Data Warehouse
- ▶ **Open Gateway**: estrazione, trasformazione e caricamento di dati
- ▶ **Symmetric Replication**: aggiornamento incrementale del data warehouse
- ▶ **Express**: strumento OLAP per l'analisi multidimensionale

Sommario

- ▶ richiami sul data warehousing
- ▶ organizzazione di un data warehouse
- ▶ l'analisi multidimensionale
- ▶ data warehousing e internet
- ▶ strumenti per il warehousing