



On Line Analytical Processing



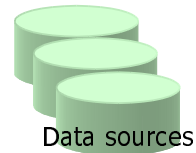
Data Warehouse

- **Data Warehouse (magazzino dati)**
 - integra in un **unico schema globale** l'informazione estratta da piu' sorgenti
 - solitamente è interrogabile, ma **non modificabile**
 - è **un'architettura per l'integrazione** di database alternativa alla replicazione ai database distribuiti
 - puo' essere
 - **ricostruito periodicamente** (es. tutte le notti)
 - **aggiornato periodicamente** (es. tutte le notti)
 - **aggiornato immediatamente** (es. ogni **n** transazioni)

OLTP vs OLAP

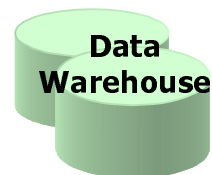
- **OLTP** (On Line Transaction Processing)
 - Gestione efficiente dei dati in linea (molti operatori)
 - Dati privi di errori e completi
 - Dati relativi allo stato attuale dell'azienda
 - Semplici da realizzare e diffusi capillarmente
 - I sistemi OLTP non sono adatti all'analisi dei dati

Operational
databases



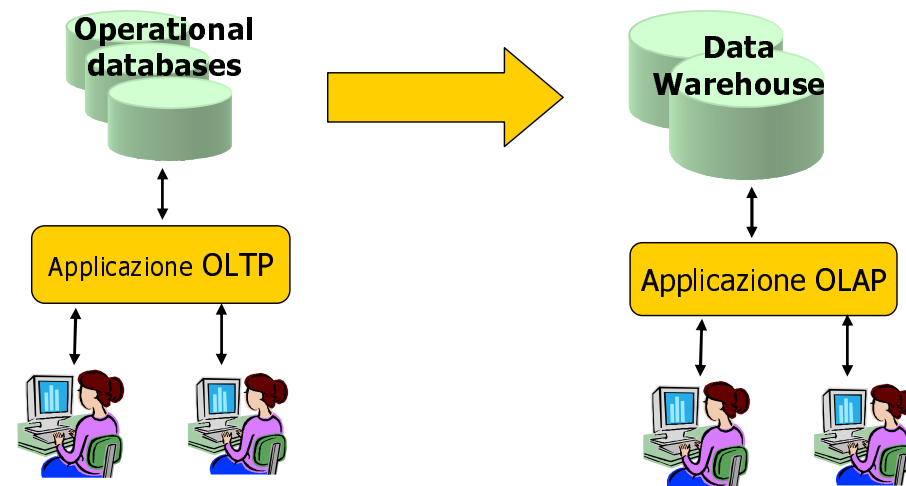
OLTP vs OLAP II

- **OLAP** (On Line Analytical Processing)
 - Pochi utenti dedicati all'analisi dell'andamento dell'azienda
 - I dati storici possono essere utili per la pianificazione e il supporto alle decisioni
 - Capire quali prodotti sono di maggiore successo
 - Stabilire l'efficacia delle promozioni sui prodotti
 - Grandi quantità di dati
 - Dati spesso scorretti o incompleti



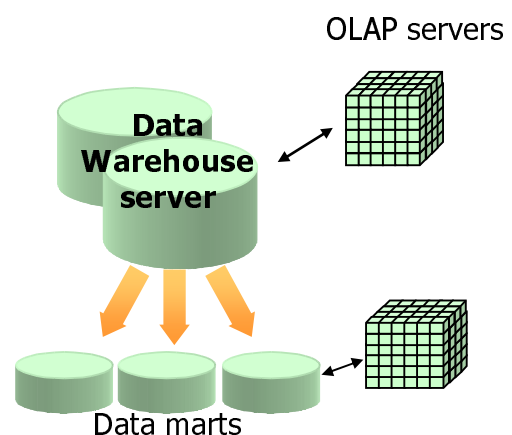
OLTP vs OLAP III

- Gli OLTP sono la fonte principale di informazioni per gli OLAP



Perché OLAP

- Definizione di **un'interfaccia di analisi** dei dati per utenti che svolgono attività di supporto alle decisioni
- Ottimizzazione delle operazioni di analisi invece che di gestione in linea
- Si separa l'ambiente on-line da quello di analisi
- E' l'analisi che avviene in modo interattivo on-line
- Il centro è il **magazzino dati** (data warehouse - DW)





Esempi di applicazioni OLAP

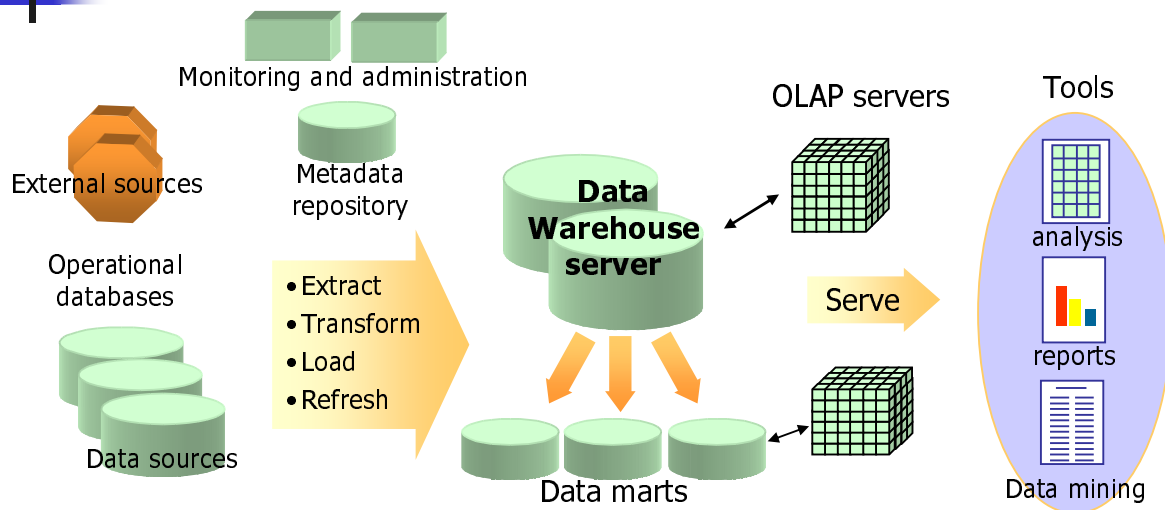
- Vendite
 - analisi e predizione delle vendite
- Marketing
 - analisi delle ricerche di mercato
 - analisi delle promozioni
 - analisi dei consumi
 - segmentazione dei mercati/clienti
- Produzione
 - Pianificazione della produzione
 - Analisi dei difetti



E' un investimento!

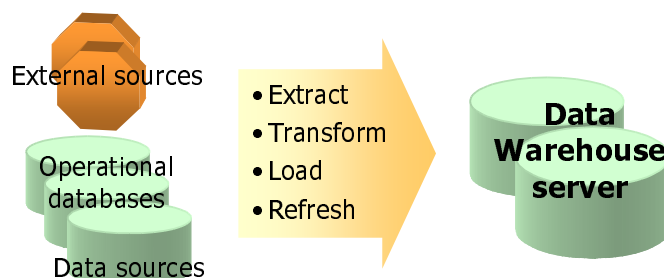
- Studio dell'International Data Corporation (IDC) 1996
 - Alti costi per l'implementazione di una efficace DW
- MA**
- Ritorno di investimento medio in 3 anni del 401%
 - 90% delle aziende con ritorno superiore al 40%
 - 25% delle aziende con più del 600% di ritorno
- Maggiore produttività dei *decision-makers*

On Line Analytical Processing



- OLAP → Sistemi orientati all'**elaborazione** e all'**analisi** dei dati

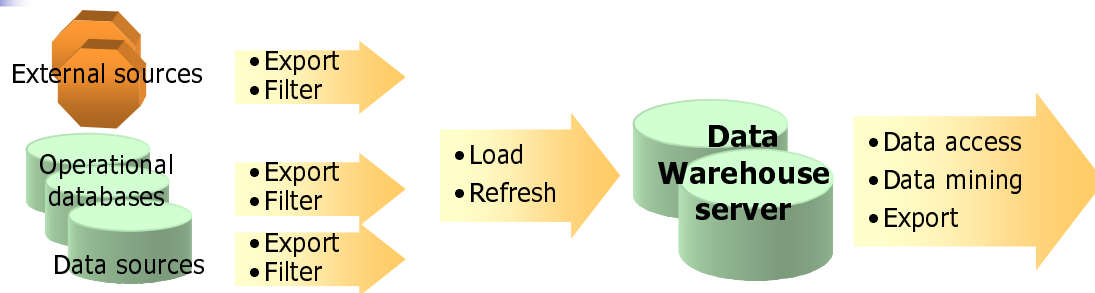
Data warehousing



I **sistemi OLTP** (i database operativi) sono una fonte dati

- Il **data warehouse server** trasferisce i dati dai database operativi al suo interno **integrandoli** (vista unitaria dei dati)
- I dati nella data warehouse sono di tipo **storico-temporale**
 - L'importazione dei dati è **asincrona** (disallineamento controllato dei dati) e **periodica** (riallineamento batch)
 - Occorre garantire la **qualità dei dati** nella DW

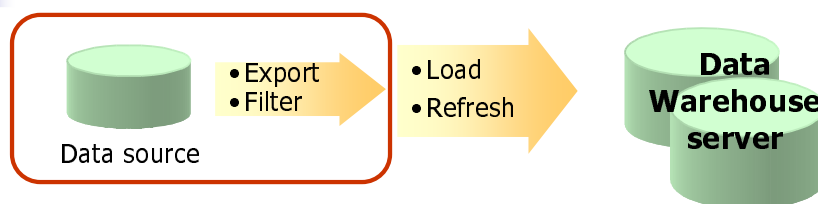
La data warehouse



Fonti dati eterogenee

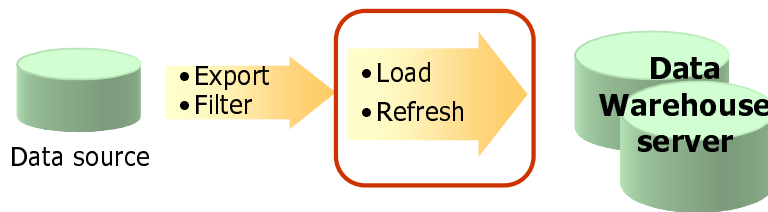
- Raccolte dati non gestite da DBMS (es. i log di un Web server)
- Dati gestiti da DBMS di vecchia generazione (**legacy systems**)
- Accesso tramite **connectors e filtri** dati forniti con il DW server
- Creazione/aggiornamento tramite procedure di
 - acquisizione (**load**)
 - aggiornamento (**refresh**)

Data filter & export



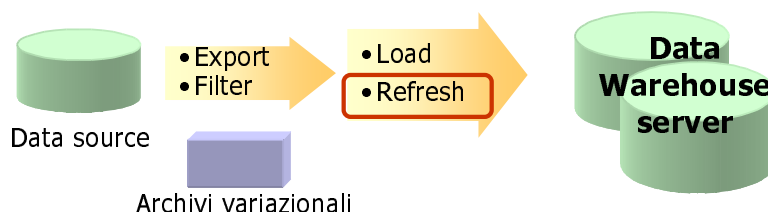
- I **data filter** controllano la correttezza dei dati prima dell'inserimento nella DW effettuando il **data cleaning**
 - eliminazione dei dati scorretti con vincoli e controlli su una o più sorgenti dati
- I **data export** sono i driver che consentono l'estrazione dei dati da una certa sorgente (DBMS, documenti, ...)
 - deve gestire un aggiornamento incrementale (basato sulle modifiche)

Data load & refresh



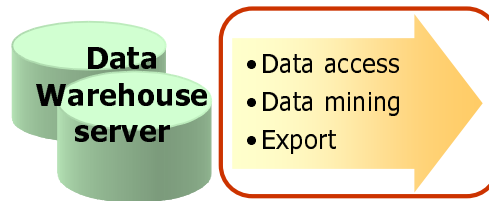
- Componente di **acquisizione dati (load)**
 - inserimento iniziale dei dati
 - costruzione delle strutture dati della DW
 - eseguita batch quando la DW non è usata
- Componente di **allineamento dati (refresh)**
 - aggiornamento incrementale della DW
 - utilizza le modifiche sulle sorgenti dati

Data refresh



- Il refresh è basato su **archivi variazionali** che registrano cancellazioni, inserimenti e modifiche
 - **Data shipping (snapshot)**
 - uso triggers nella sorgente e trasferimento dei dati modificati
 - **Transaction shipping (copia transazionale)**
 - uso dei log e trasferimento delle transazioni
- Per le cancellazioni i dati non sono eliminati nella DW per non perdere lo storico ma solo marcati

Analisi con la DW



- Componente di **accesso ai dati**
 - realizza in modo efficiente operazioni complesse di analisi
 - join fra tabelle, ordinamenti, aggregazioni -
 - operazioni come **roll up**, **drill down** e **datacube**
 - interfaccia di facile uso per gli analisti
- Componente di **data mining**
 - scoprire in modo automatico regolarità nei dati
- Componente di **esportazione dei dati** verso altre DW (es. da un DW dipartimentale ad un DW di tutta l'azienda)

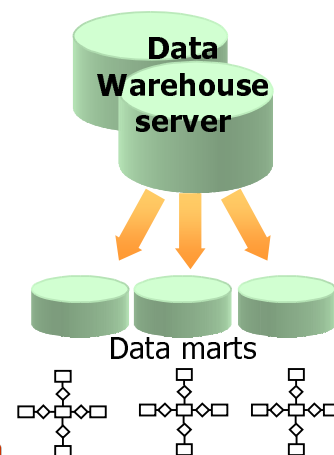
Maggini, Scarselli

Sistemi per basi di dati

15

I data mart

- Costruire una DW aziendale completa è complesso
- Si costruiscono schemi per sottoinsiemi semplici dei dati aziendali (**data mart**) per i quali è chiaro l'obiettivo dell'analisi
 - vendite
 - operazioni di sportello
- I dati di un data mart sono rappresentati secondo uno **schema multidimensionale (data cube)** e realizzato attraverso uno **schema a stella**



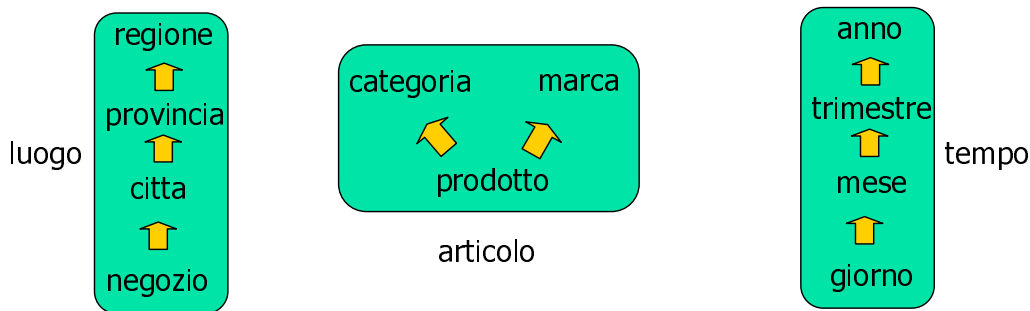
Maggini, Scarselli

Sistemi per basi di dati

16

Il modello multidimensionale

- Il modello multidimensionale è costituito da
 - **fatti**
un concetto da analizzare (es. la vendita di un prodotto)
 - **misure**
una proprietà atomica di un fatto (es. il numero delle vendite, l'incasso)
 - **dimensioni**
una prospettiva lungo la quale si analizza il fatto (es. il negozio, il mese)
- Le dimensioni sono organizzate in livelli di aggregazione



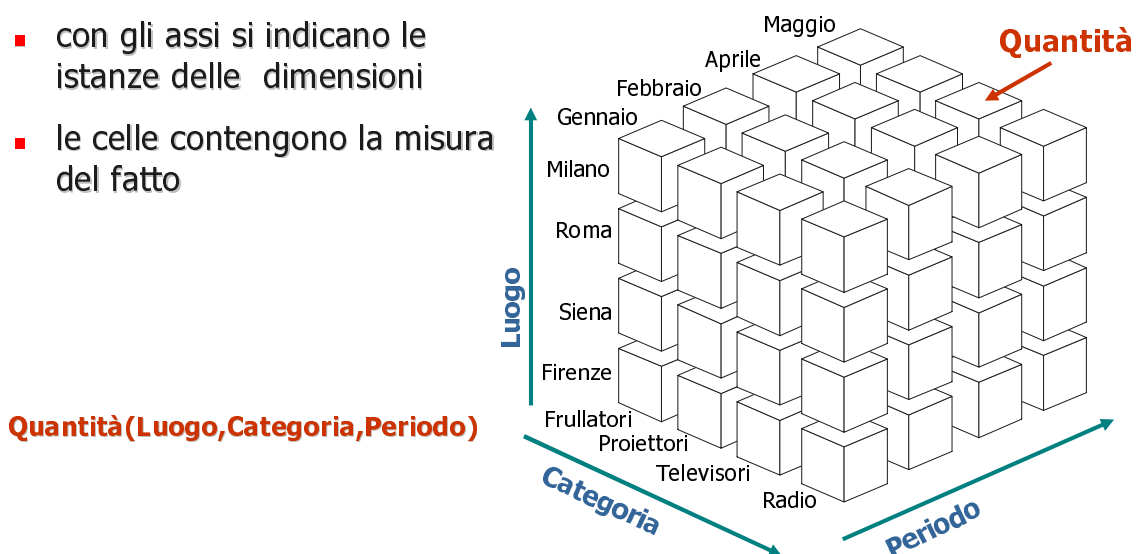
Maggini, Scarselli

Sistemi per basi di dati

17

Data cube

- **Data cube**
 - con gli assi si indicano le istanze delle dimensioni
 - le celle contengono la misura del fatto



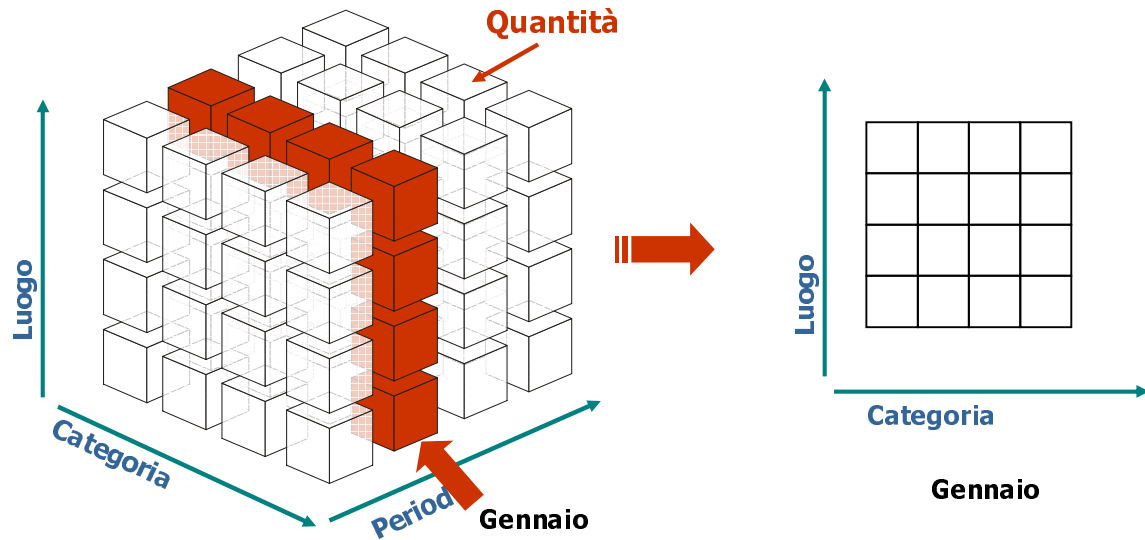
Maggini, Scarselli

Sistemi per basi di dati

18

Data cube: operazioni

- Selezione di una vista (**slice-and-dice**)



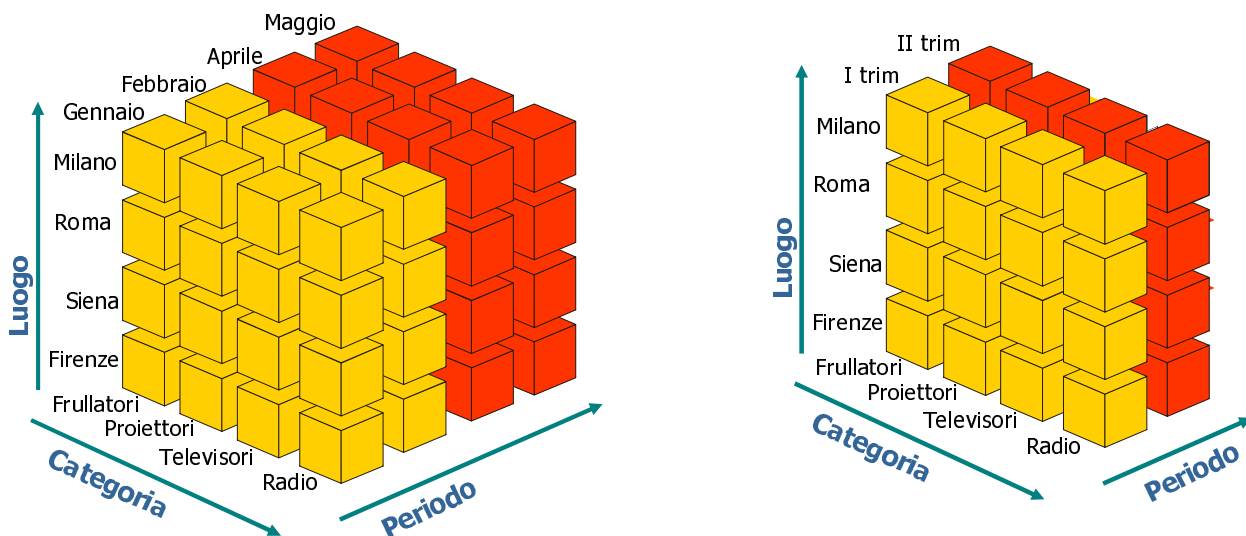
Maggini, Scarselli

Sistemi per basi di dati

19

Data cube: operazioni II

- Aggregazione dei dati lungo una dimensione salendo nella gerarchia (**roll-up**)



Maggini, Scarselli

Sistemi per basi di dati

20



Data cube: operazioni III

- Al limite l'operazione di roll-up **elimina una dimensione**
- affinché sia applicabile, occorre che la misura sia **additiva** lungo la dimensione prescelta
 - quantita' e incasso sono additive rispetto al periodo
 - scorte non è additiva rispetto al periodo
- L'operazione di **drill-down**
 - i dati sono disaggregati e resi più dettagliati muovendo una dimensione verso il basso della gerarchia
 - è **l'opposto** di roll-up