

# **Data Warehousing**

## **Parte I**

### **Introduzione al warehousing**

# Argomenti della lezione

- ▶ cosa è un data warehouse
- ▶ classificazione dei processi aziendali
- ▶ sistemi di supporto alle decisioni
- ▶ elaborazione OLTP e OLAP
- ▶ architetture per il warehousing

**Cosa è un  
data warehouse ?**

## **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**

# Motivazioni

**Le basi di dati si utilizzano per gestire le operazioni quotidiane di una organizzazione**

- ▶ vendite nelle catene di supermercati
- ▶ instradamento e contabilizzazione delle telefonate
- ▶ .....

Questi dati se opportunamente  
**accumulati e analizzati**  
possono essere utilizzati  
per la pianificazione  
e il supporto alle decisioni

# Processi aziendali e decisioni

I processi informativi svolti da una organizzazione possono essere classificati in due grandi categorie:

- ▶ processi operativi
- ▶ processi direzionali

# Banca

## Processi operativi

- ▶ gestione di un movimento su un conto corrente bancario
- ▶ compravendita di titoli

## Processi direzionali

- ▶ lancio di una campagna promozionale
- ▶ stipula di accordi commerciali



# Compagnia telefonica

## Processi operativi

- ▶ instradamento delle telefonate
- ▶ contabilizzazione delle telefonate

## Processi direzionali

- ▶ definizione di tariffe
- ▶ definizione di contratti diversificati

**L'informatizzazione di un processo dipende dal grado di strutturazione delle sue operazioni**

- ▶ **Un processo altamente strutturato può essere facilmente informatizzato**
- ▶ **Un processo non strutturato può essere al più “supportato” da applicazioni informatiche**

# Tipologie di sistemi informatici

- ▶ **transaction processing systems**  
per i processi operativi
  - tipicamente dipartimentali
- ▶ **decision support systems** per il  
supporto ai processi direzionali
  - fortemente integrati

# **Sistemi di supporto alle decisioni**

**Tecnologia di supporto alla  
dirigenza aziendale per prendere  
decisioni tattico-strategiche in  
modo efficace e veloce, mediante  
una particolare tipologia di  
elaborazione (OLAP)**

# Operazioni supportate dai DSS

- ▶ quali sono stati i volumi di vendita dello scorso anno per regione e categoria di prodotto?
- ▶ quali ordini dovremmo soddisfare per massimizzare le entrate?
- ▶ uno sconto tra il 7% e il 10% potrebbe incrementare le vendite in modo sufficiente?

**Ma su quali dati?**

**Quelli accumulati per i  
processi operativi**

# Tipologie di elaborazione

- ▶ Nei transaction processing systems:

On-Line Transaction Processing  
(OLTP)

- ▶ Nei decision support systems:

On-Line Analytical Processing  
(OLAP)

# OLTP

**Tradizionale tipologia di elaborazione su basi di dati, caratterizzata da:**

- ▶ **transazioni predefinite e di breve durata**
- ▶ **dati dettagliati, aggiornati e recenti**
- ▶ **letture e/o modifiche di pochi record**
- ▶ **proprietà delle transazioni critiche**
- ▶ **architettura centralizzata**



# OLAP

- ▶ interrogazioni complesse e casuali
- ▶ dati storici e aggregati
- ▶ dati provenienti da più basi di dati
- ▶ letture su molti record, mai scritture
- ▶ risposte alle interrogazioni in linea
- ▶ visualizzazione dei dati fondamentale
- ▶ architettura client-server

# OLTP e OLAP

	OLTP	OLAP
Utente	impiegato	dirigente
Funzione	operazioni ordinarie	supporto decisioni
Dati	correnti, aggiornati, dettagliati, relazionali	storici, aggregati, multidimensionali, integrati
Uso	ripetitivo	casuale
Accesso	R/W, indicizzato	R, sequenziale
Unità di lavoro	transazione breve	query complessa
Record acc.	decine	milioni
N. utenti	migliaia	centinaia
Dimensione	100MB - 1GB	100GB - 1TB
Metrica	throughput	tempo di risposta

# **Evoluzione dei DSS**

**Anni '60 — rapporti batch**

**Anni '70 — DSS basato su terminale**

**Anni '80 — strumenti d'automazione  
d'ufficio e di analisi**

**Anni '90 — data warehousing, con  
strumenti integrati di OLAP**

# Data warehouse

Un data warehouse è una base di dati

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

## **... integrata ...**

**I dati di interesse provengono da diverse sorgenti informative**

**Riconciliazione delle eterogeneità**

- ▶ **nomi**
- ▶ **struttura**
- ▶ **codifica**
- ▶ **rappresentazione**

## **... dati storici ...**

**Le basi di dati operazionali  
mantengono il valore corrente  
delle informazioni**

**Nel data warehouse è di interesse  
l'evoluzione storica delle  
informazioni**

## **... dati aggregati ...**

**Nelle attività di analisi dei dati per il supporto alle decisioni**

- ▶ **non interessa “chi” ma “quanti”**
- ▶ **non interessa un dato ma la somma, la media, il minimo e il massimo, di un insieme di dati**

## **... fuori linea ...**

In una base di dati operativa, i dati vengono elaborati pochi record alla volta

Nel data warehouse, abbiamo

- ▶ operazioni di accesso e interrogazione “diurne”
- ▶ operazioni di caricamento e aggiornamento dei dati “notturne”

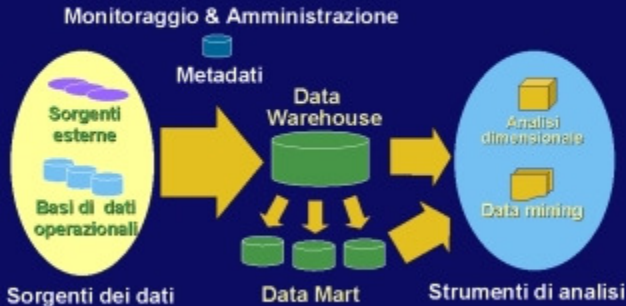
che riguardano milioni di record



## **... una base di dati separata ...**

- ▶ è difficile fare l'integrazione in linea
- ▶ i dati di interesse sono diversi
  - mantenimento dati storici
  - aggregazione di dati
- ▶ l'analisi dei dati richiede organizzazioni speciali e metodi di accesso specifici
- ▶ degrado generale delle prestazioni senza la separazione

# Architettura per il data warehousing



# Sintesi

- ▶ cosa è un data warehouse
- ▶ classificazione dei processi aziendali
- ▶ sistemi di supporto alle decisioni
- ▶ elaborazione OLTP e OLAP
- ▶ architetture per il warehousing