

# **Basi di dati**

## **Introduzione**

Docente: Stefano Paraboschi

[parabosc@unibg.it](mailto:parabosc@unibg.it)

# **Corso di “Sistemi informativi” AA 2005/2006**

- Orario:
  - Lunedì 10.45-12.15, aula 10
  - Martedì 8.45-10.15, aula 8
- Pagina Web del corso: raggiungibile dal sito dell'Università.
- URL diretto: <http://193.204.253.55/sistinf.html>

# Corso di “Sistemi informativi”

## AA 2005/2006

- Titolare del corso: Prof. Stefano Paraboschi
  - ricevimento: Merc. 17.30-19.30
  - email: parabosc@unibg.it
  - tel: 035-205-2356
- Resp. esercitazioni: Dott. Mario Molinari
  - email: marimol@tin.it

# Corso di “Sistemi informativi”

## AA 2005/2006

- Termine lezioni: 30/5/06
- E' possibile sostenere l'esame con delle prove in itinere
  - I prova: martedì 11/4/06
  - II prova: mercoledì 6/6/06
  - La partecipazione alle prove in itinere non pone vincoli rispetto alla partecipazione ai normali appelli
- La consegna del compito a un appello implica la rinuncia a un eventuale voto positivo ottenuto nelle prove in itinere

# Corso di “Sistemi informativi”

## AA 2005/2006

- Calendario esercitazioni:
  - 14/3
  - 21/3
  - 3/4 (lab.)
  - 4/4
  - 19/5
  - 23/5 (ven.)
  - 29/5 (lab.)
  - 30/5

# **Corso di “Sistemi informativi”**

## **AA 2005/2006**

- E' possibile preparare un elaborato (facoltativo)
- Al massimo 3 persone per elaborato
- Il voto assegnato all'elaborato viene sommato al voto ottenuto nei compiti o all'esame, con arrotondamento per eccesso

# Corso di “Sistemi informativi”

## AA 2005/2006

- Libro di testo: Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Riccardo Torlone “Basi di dati: Modelli e linguaggi di interrogazione, II ed.”, McGraw-Hill Libri Italia, € 25.00 (?)
- Esiste anche la versione in inglese, degli stessi autori, titolo “Database systems”, McGraw-Hill

# Corso di “Sistemi informativi”

## AA 2005/2006

- Sul sito del corso sono disponibili i lucidi usati durante le lezioni
- Eserciziari:
  - Elena Baralis, Alberto Belussi, Giuseppe Psaila, “Esercizi di basi di dati”, Esculapio
  - Chiara Francalanci, Fabio Schreiber, Letizia Tanca, “Esercizi di progettazione concettuale”, Esculapio



# Corso di “Sistemi informativi”

## AA 2005/2006

- Altro materiale
  - soluzioni di tutti gli esercizi del libro sul sito della versione italiana del libro <http://www.ateneonline.it>
  - raccolta completa dei temi d’esame passati su <http://193.204.253.55/TemiSi>

# Sistemi informativi e sistemi informatici una premessa

- Che cos'è l'informatica? Una definizione:
  - Scienza del trattamento razionale, specialmente per mezzo di macchine automatiche, dell'informazione, considerata come supporto alla conoscenza umana e alla comunicazione (Academie Francaise).
- L'informatica ha due anime:
  - tecnologica: i calcolatori elettronici e i sistemi che li utilizzano;
  - metodologica: i metodi per la soluzione di problemi e la gestione delle informazioni.

# Sistema informativo

- Componente (sottosistema) di una organizzazione che gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse (cioè utilizzate per il perseguimento degli scopi dell'organizzazione stessa).
  - ogni organizzazione ha un sistema informativo, eventualmente non esplicitato nella struttura;
  - quasi sempre, il sistema informativo è di supporto ad altri sottosistemi, e va quindi studiato nel contesto in cui è inserito
  - il sistema informativo è di solito suddiviso in sottosistemi (in modo gerarchico o decentrato), più o meno fortemente integrati

# Sistema organizzativo

- insieme di risorse e regole per lo svolgimento coordinato delle attività al fine del perseguimento degli scopi
- il sistema informativo è parte del sistema organizzativo
- il sistema informativo esegue/gestisce processi informativi (cioè i processi che coinvolgono informazioni)

# Risorse

- le risorse di una azienda (o amministrazione):
  - persone
  - denaro
  - materiali
  - **informazioni**

# Funzioni di un sistema informativo

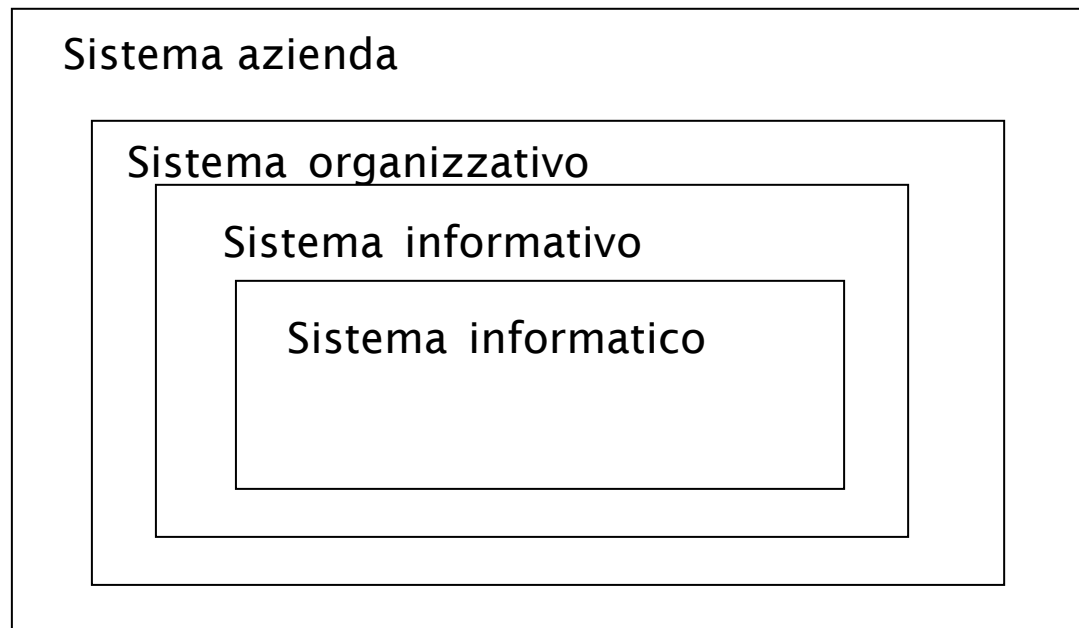
- raccolta, acquisizione delle informazioni
- archiviazione, conservazione delle informazioni
- elaborazione delle informazioni
- distribuzione, scambio di informazioni

# Sistemi informativi e automazione

- Il concetto di “sistema informativo” è indipendente da qualsiasi automatizzazione: esistono organizzazioni la cui ragion d’essere è la gestione di informazioni (p. es. servizi anagrafici e banche) e che operano da secoli.

# Sistema Informatico

- porzione automatizzata del sistema informativo
- il sistema informatico è la parte del sistema informativo che gestisce informazioni per mezzo della tecnologia informatica.





# **Sistema informativo e sistema informatico**

- Anche prima di essere automatizzati, molti sistemi informativi si sono evoluti verso una razionalizzazione e standardizzazione delle procedure e dell'organizzazione delle informazioni.

# Gestione delle informazioni

- Nelle attività umane, le informazioni vengono gestite (registrate e scambiate) in forme diverse, a seconda delle necessità e capacità:
  - idee informali
  - linguaggio naturale (scritto o parlato, formale o colloquiale, in una lingua o in un'altra)
  - disegni, grafici, schemi
  - numeri
  - codici (anche segreti)
- e su vari supporti, dalla memoria umana alla carta

# Gestione delle informazioni

- Nelle attività standardizzate dei sistemi informativi complessi, sono state introdotte col tempo forme di organizzazione e codifica delle informazioni.
- Ad esempio, nei servizi anagrafici si è iniziato con registrazioni discorsive e sono state poi introdotte informazioni via via più precise (e in un certo senso artificiali):
  - nome e cognome;
  - estremi anagrafici;
  - codice fiscale;

# Informazioni e dati

- Nei sistemi informatici (e non solo in essi), le informazioni vengono rappresentate in modo essenziale, spartano: attraverso i dati.
  - **informazione** : notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere.
  - **dato** : ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione; (in informatica) elementi di informazione costituiti da simboli che debbono essere elaborati.

(Vocabolario della lingua italiana, 1987)

# Dati e informazioni

- Quindi, i dati sono fatti elementari, informazioni codificate, che hanno bisogno di essere interpretate per fornire conoscenza

## *Esempio*

'Mario' '275' su un foglio di carta sono due **dati**.

Se il foglio di carta viene fornito in risposta alla domanda "A chi mi devo rivolgere per il problema X; qual è il suo numero di telefono?", allora i dati possono essere interpretati ed arricchire la conoscenza.

# Perché i dati?

- La rappresentazione precisa di forme più ricche di informazione e conoscenza è difficile
- I dati costituiscono spesso una risorsa strategica, perché più stabili nel tempo di altre componenti (processi, tecnologie, ruoli umani)

# Base di dati

(accezione generica)

- **collezione di dati, utilizzati per rappresentare le informazioni di interesse per una o più applicazioni di una organizzazione.**

(accezione specifica)

- **collezione di dati gestita da un DBMS**

# Sistema di gestione di basi di dati

## **DataBase Management System — DBMS**

Sistema (**prodotto software**) in grado di gestire **collezioni di dati** che siano (anche):

- **grandi** (di dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati)
- **persistenti** (con un periodo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano)
- **condivise** (utilizzate da applicazioni diverse)

garantendo **affidabilità** (resistenza a malfunzionamenti hardware e software) e **privatezza** (con una disciplina e un controllo degli accessi). Come ogni prodotto informatico, un DBMS deve essere **efficiente** (utilizzando al meglio le risorse di spazio e tempo del sistema) ed **efficace** (rendendo produttive le attività dei suoi utilizzatori).



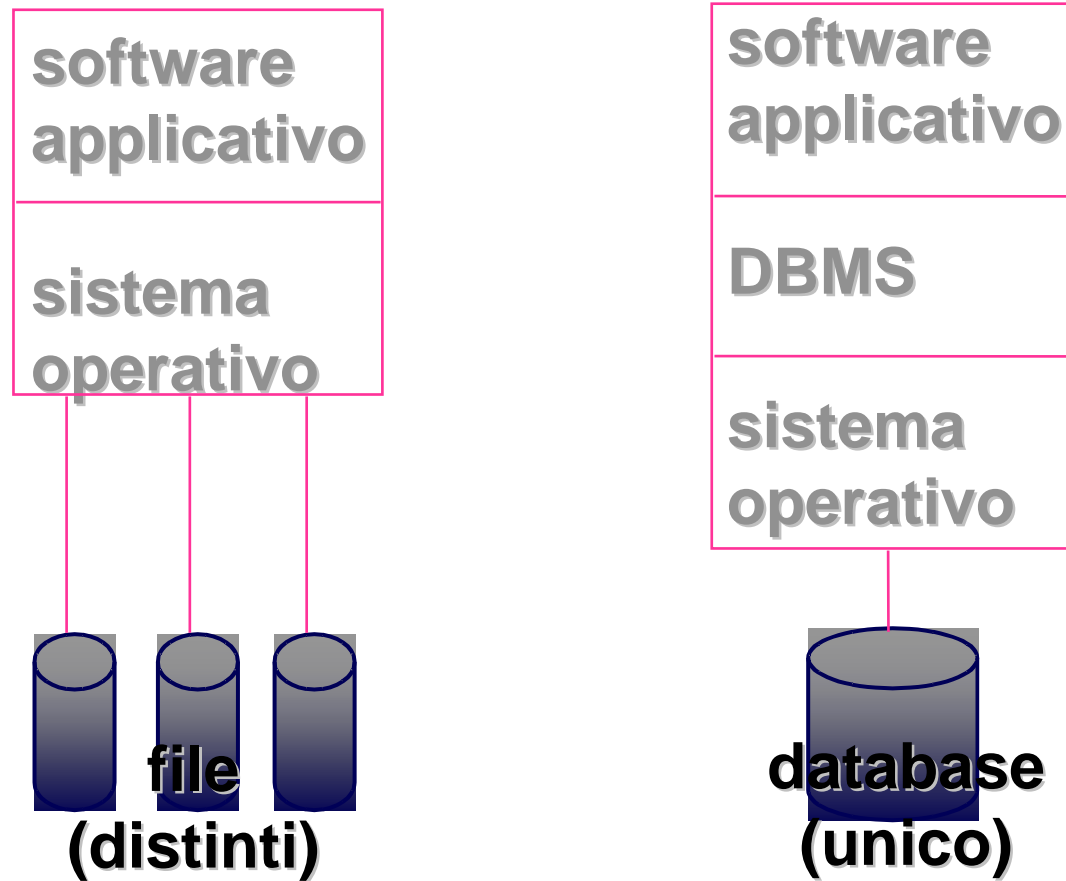
# Condivisione

- Ogni organizzazione (specie se grande) è divisa in settori o comunque svolge diverse attività. A ciascun settore o attività corrisponde un (sotto-)sistema informativo (privato o porzione di un sistema più grande).
- Possono esistere sovrapposizioni fra i dati di interesse dei vari settori.
- Una base di dati è una risorsa **integrata**, condivisa fra i vari settori.

## Condivisione (segue)

- L'integrazione e la condivisione permettono di ridurre la **ridondanza** (evitando ripetizioni) e, di conseguenza, le possibilità di incoerenza (o **inconsistenza**) fra i dati.
- Poiché la condivisione non è mai completa (o comunque non opportuna) i DBMS prevedono meccanismi di definizione della **privatezza** dei dati e di limitazioni all'accesso (**autorizzazioni**).
- La condivisione richiede un opportuno coordinamento degli accessi: **controllo della concorrenza**.

# File system e DBMS



# Efficienza

- Si misura (come in tutti i sistemi informatici) in termini di tempo di esecuzione (tempo di risposta) e spazio di memoria (principale e secondaria).
- I DBMS, a causa della varietà di funzioni, non sono necessariamente più efficienti dei file system.
- L'efficienza è il risultato della qualità del DBMS e delle applicazioni che lo utilizzano.

# DBMS vs file system

- La gestione di insiemi di dati grandi e persistenti è possibile anche attraverso sistemi più semplici — gli ordinari **file system** dei sistemi operativi, che permettono di realizzare anche rudimentali forme di condivisione.
- Non esiste una linea netta di separazione fra DBMS e non-DBMS.
- I DBMS estendono le funzionalità dei file system, fornendo più servizi ed in maniera integrata (cfr. **efficacia**).
- I file system prevedono forme di condivisione, permettendo accessi contemporanei in lettura ed esclusivi in scrittura: se è in corso un'operazione di scrittura su un file, altri non possono accedere affatto al file. Nei DBMS, c'è maggiore flessibilità: si può accedere contemporaneamente a record diversi di uno stesso file o addirittura allo stesso record (in lettura).

## DBMS vs file system (2)

- Nei programmi tradizionali che accedono a file, ogni programma contiene una descrizione della struttura del file stesso, con i conseguenti rischi di incoerenza fra le descrizioni (ripetute in ciascun programma) e i file stessi.
- Nei DBMS, esiste una porzione della base di dati (il **catalogo** o **dizionario**) che contiene una descrizione centralizzata dei dati, che può essere utilizzata dai vari programmi.

# Descrizioni dei dati nei DBMS

- Esistono descrizioni e rappresentazioni dei dati a livelli diversi, che permettono l'**indipendenza dei dati** dalla rappresentazione fisica: i programmi fanno riferimento alla struttura a livello più alto, e le rappresentazioni sottostanti possono essere modificate senza necessità di modifica dei programmi.
- Precisiamo attraverso il concetto di modello dei dati.

# Modello dei dati

- insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la dinamica
- componente fondamentale: **meccanismi di strutturazione** (o **costruttori di tipo**)
- come nei linguaggi di programmazione esistono meccanismi che permettono di definire nuovi tipi, così ogni modello dei dati prevede alcuni costruttori
- ad esempio, il **modello relazionale** prevede il costruttore **relazione**, che permette di definire insiemi di record omogenei



# Tabelle: rappresentazione di relazioni

<b>CORSI</b>	<b>Corso</b>	<b>Docente</b>	<b>Aula</b>
	Basi di dati	Rossi	DS3
	Sistemi	Neri	N3
	Reti	Bruni	N3
	Controlli	Bruni	G

<b>AULE</b>	<b>Nome</b>	<b>Edificio</b>	<b>Piano</b>
	DS1	Ex-OMI	Terra
	N3	Ex-OMI	Terra
	G	Pincherle	Primo

# Schemi e istanze

In ogni base di dati esistono:

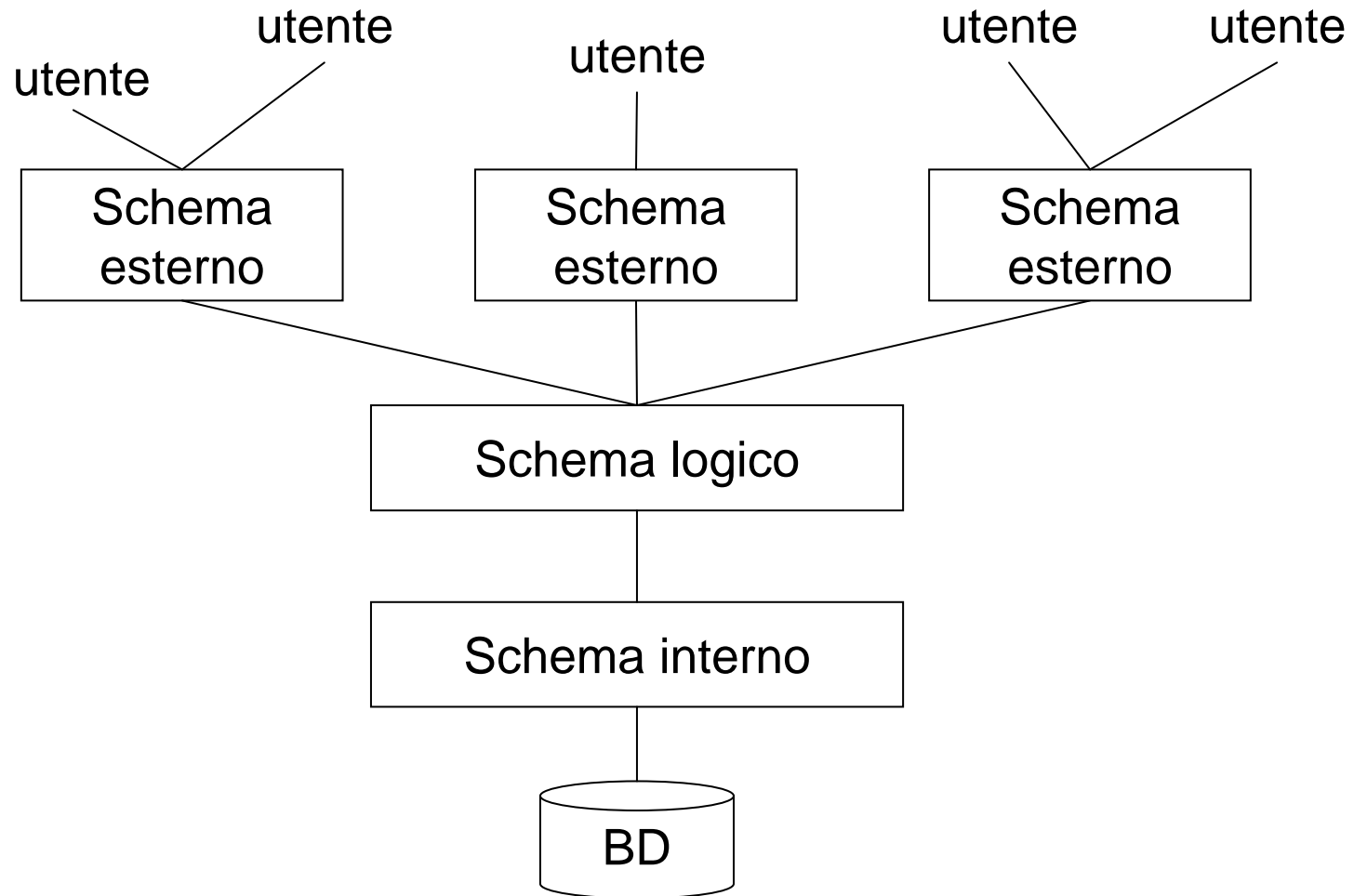
- lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura (aspetto **intensionale**);  
nell'esempio, le intestazioni delle tabelle
- l'**istanza**, costituita dai valori attuali, che possono cambiare molto e molto rapidamente (aspetto **estensionale**);  
nell'esempio, il "corpo" di ciascuna tabella

## Due tipi (principali) di modelli

- **modelli logici:** utilizzati nei DBMS esistenti per l'organizzazione dei dati; ad essi fanno riferimento i programmi; sono indipendenti dalle strutture fisiche;  
esempi: relazionale, reticolare, gerarchico, a oggetti
- **modelli concettuali:** permettono di rappresentare i dati in modo indipendente da ogni sistema, cercando di descrivere i concetti del mondo reale; sono utilizzati nelle fasi preliminari di progettazione;  
il più noto è il modello **Entity-Relationship**

Di per sé, un modello potrebbe essere concettuale e logico al tempo stesso

# Architettura standard (ANSI/SPARC) a tre livelli per DBMS



# Architettura ANSI/SPARC: schemi

**schema logico:** descrizione dell'intera base di dati nel modello logico "principale" del DBMS

**schema esterno:** descrizione di parte della base di dati in un modello logico ("viste" parziali, derivate, anche in modelli diversi)

**schema fisico:** rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione.

# Una vista

## Corsi

<b>Corso</b>	<b>Docente</b>	<b>Aula</b>
Basi di dati	Rossi	DS3
Sistemi	Neri	N3
Reti	Bruni	N3
Controlli	Bruni	G

## Aule

<b>Nome</b>	<b>Edificio</b>	<b>Piano</b>
DS1	Ex-OMI	Terra
N3	Ex-OMI	Terra
G	Pincherle	Primo

## CorsiSedi

<b>Corso</b>	<b>Aula</b>	<b>Edificio</b>	<b>Piano</b>
Sistemi	N3	Ex-OMI	Terra
Reti	N3	Ex-OMI	Terra
Controlli	G	Pincherle	Primo

# Indipendenza dei dati

conseguenza della articolazione in livelli:

l'accesso avviene solo tramite il livello esterno (che può coincidere con il livello logico); due forme:

**fisica:** il livello logico e quello esterno sono indipendenti da quello fisico; una relazione è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica (che può anche cambiare nel tempo senza che debbano essere modificate le forme di utilizzo)

**logica:** il livello esterno è indipendente da quello logico

- aggiunte o modifiche alle viste non richiedono modifiche al livello logico;
- modifiche allo schema logico che lascino inalterato lo schema esterno sono trasparenti

# Linguaggi per basi di dati

- Un altro contributo all'efficacia: disponibilità di vari linguaggi e interfacce diverse.

L'accesso ai dati può avvenire

1. con linguaggi testuali interattivi
2. con comandi (come quelli del linguaggio interattivo) immersi in un linguaggio **ospite** (Pascal, C, Cobol, etc.)
3. con comandi (come quelli del linguaggio interattivo) immersi in un linguaggio ad hoc, con anche altre funzionalità (p.es. per grafici o stampe strutturate), anche con l'ausilio di strumenti di sviluppo (p. es. per la gestione di maschere)
4. con interfacce amichevoli (senza linguaggio testuale)



# SQL, un linguaggio interattivo

```
SELECT Corso, Aula, Piano  
FROM Aule, Corsi  
WHERE Nome = Aula  
AND Piano="Terra"
```

<b>Corso</b>	<b>Aula</b>	<b>Piano</b>
Reti	N3	Terra
Sistemi	N3	Terra

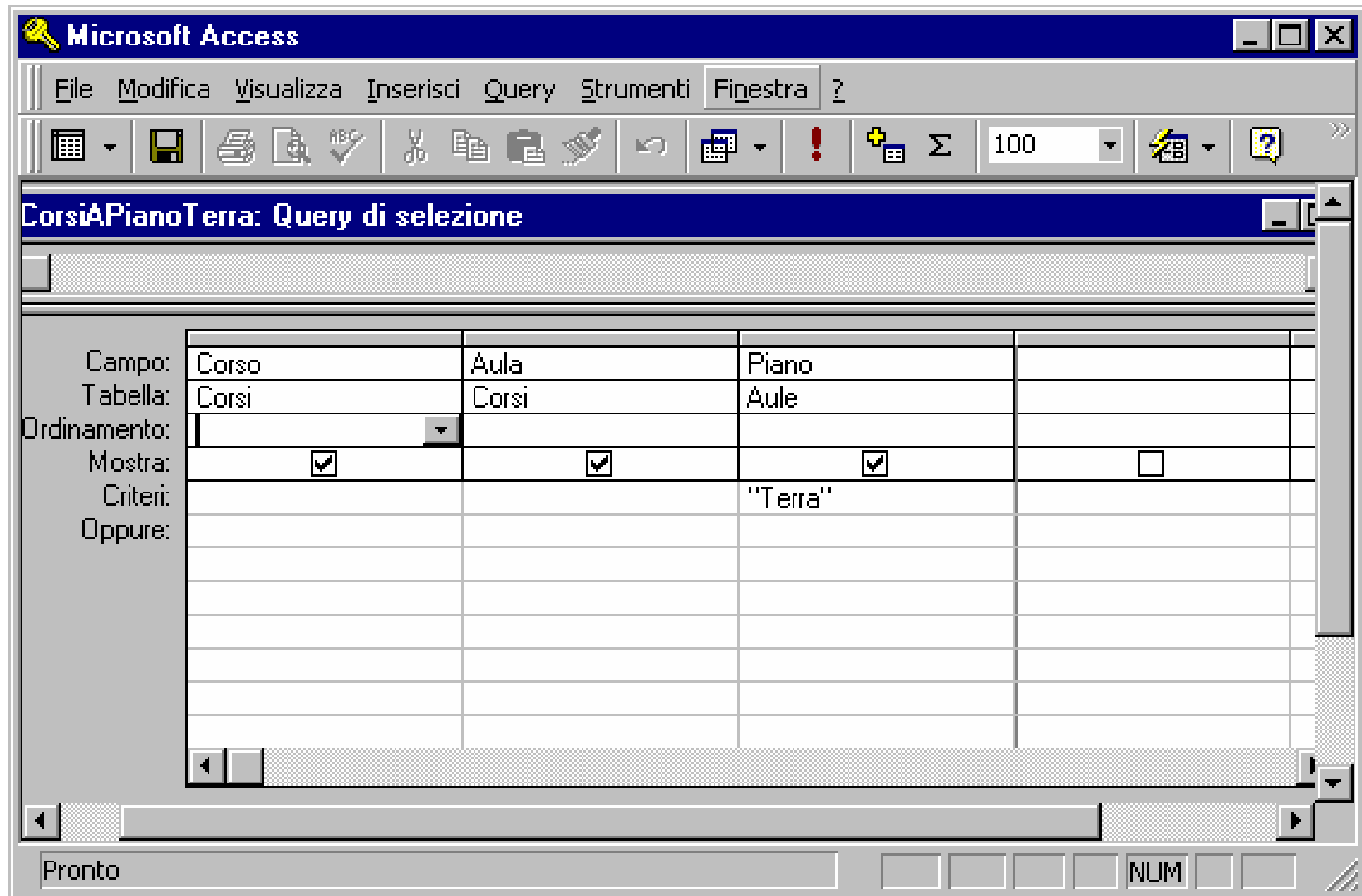
# SQL immerso in Pascal (o altro linguaggio ad alto livello)

```
write('nome della citta"?'); readln(citta);
EXEC SQL DECLARE P CURSOR FOR
    SELECT NOME, REDDITO
    FROM PERSONE
    WHERE CITTA = :citta ;
EXEC SQL OPEN P ;
EXEC SQL FETCH P INTO :nome, :reddito ;
while SQLCODE = 0 do begin
    write('nome della persona:', nome, 'aumento?');
    readln(aumento);
    EXEC SQL UPDATE PERSONE SET REDDITO = REDDITO + :aumento
        WHERE CURRENT OF P
    EXEC SQL FETCH P INTO :nome, :reddito
end;
EXEC SQL CLOSE CURSOR P
```

# SQL immerso in linguaggio ad hoc (Oracle PL/SQL)

```
declare Stip number;
begin
  select Stipendio into Stip
  from Impiegato
  where Matricola = '575488'
  for update of Stipendio;
  if Stip > 30 then
    update Impiegato set Stipendio = Stipendio * 1.1 where Matricola = '575488';
  else
    update Impiegato set Stipendio = Stipendio * 1.15 where Matricola = '575488';
  end if;
  commit;
exception
  when no_data_found then
    insert into Errori
    values('Non esiste la matricola specificata',sysdate);
end;
```

# Interazione non testuale (in Access)



# Una distinzione terminologica (separazione fra dati e programmi)

**data definition language (DDL) :**

per la definizione di schemi (logici, esterni, fisici) e altre operazioni generali;

**data manipulation language (DML) :**

per l'interrogazione e l'aggiornamento di (istanze di) basi di dati.

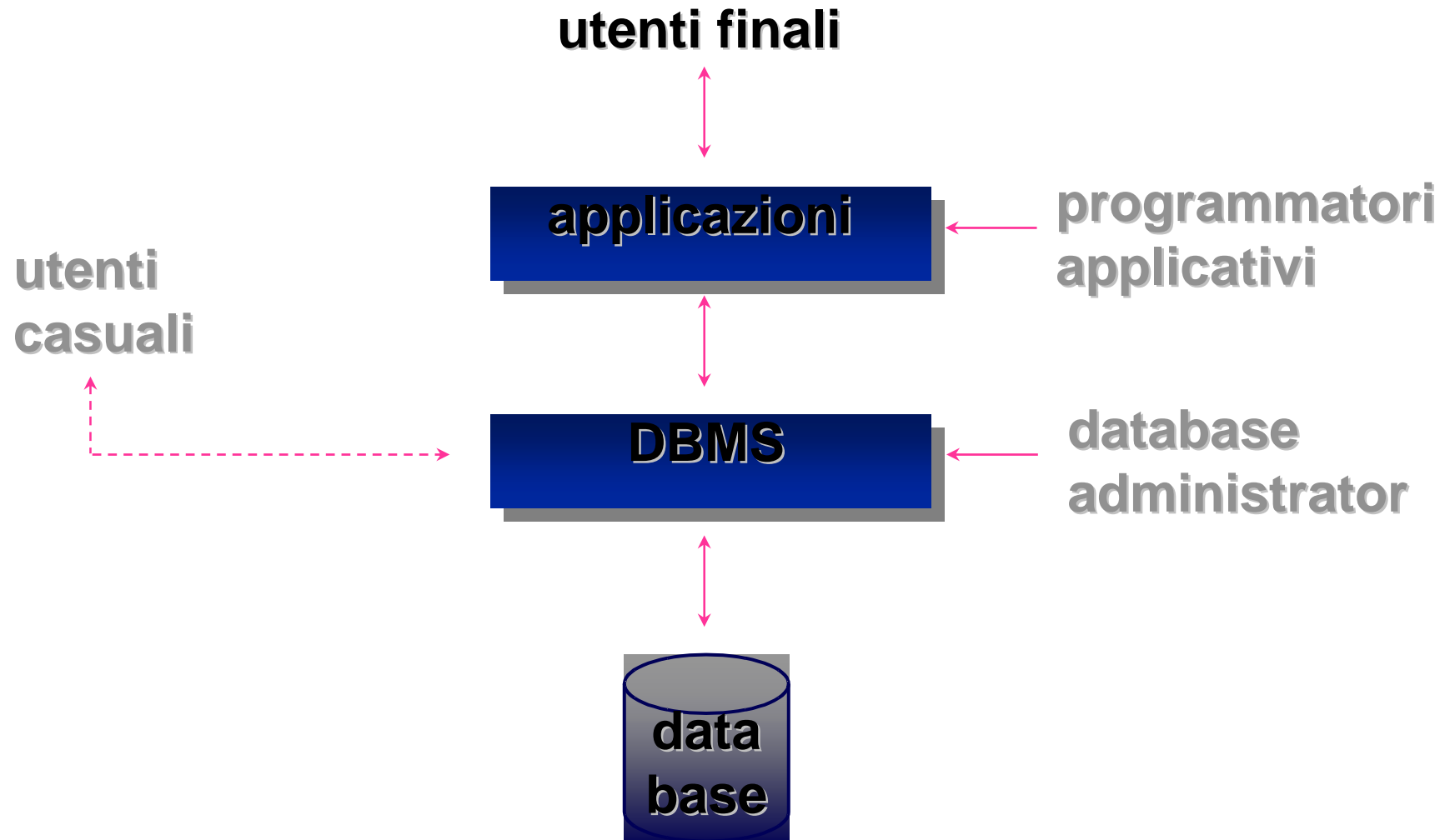
# Personaggi e interpreti

- progettisti e realizzatori di DBMS
- progettisti della base di dati e amministratori della base di dati (DBA)
- progettisti e programmatori di applicazioni
- utenti
  - utenti finali (terminalisti): eseguono applicazioni predefinite (**transazioni**)
  - utenti casuali: eseguono operazioni non previste a priori, usando linguaggi interattivi

# Database administrator (DBA)

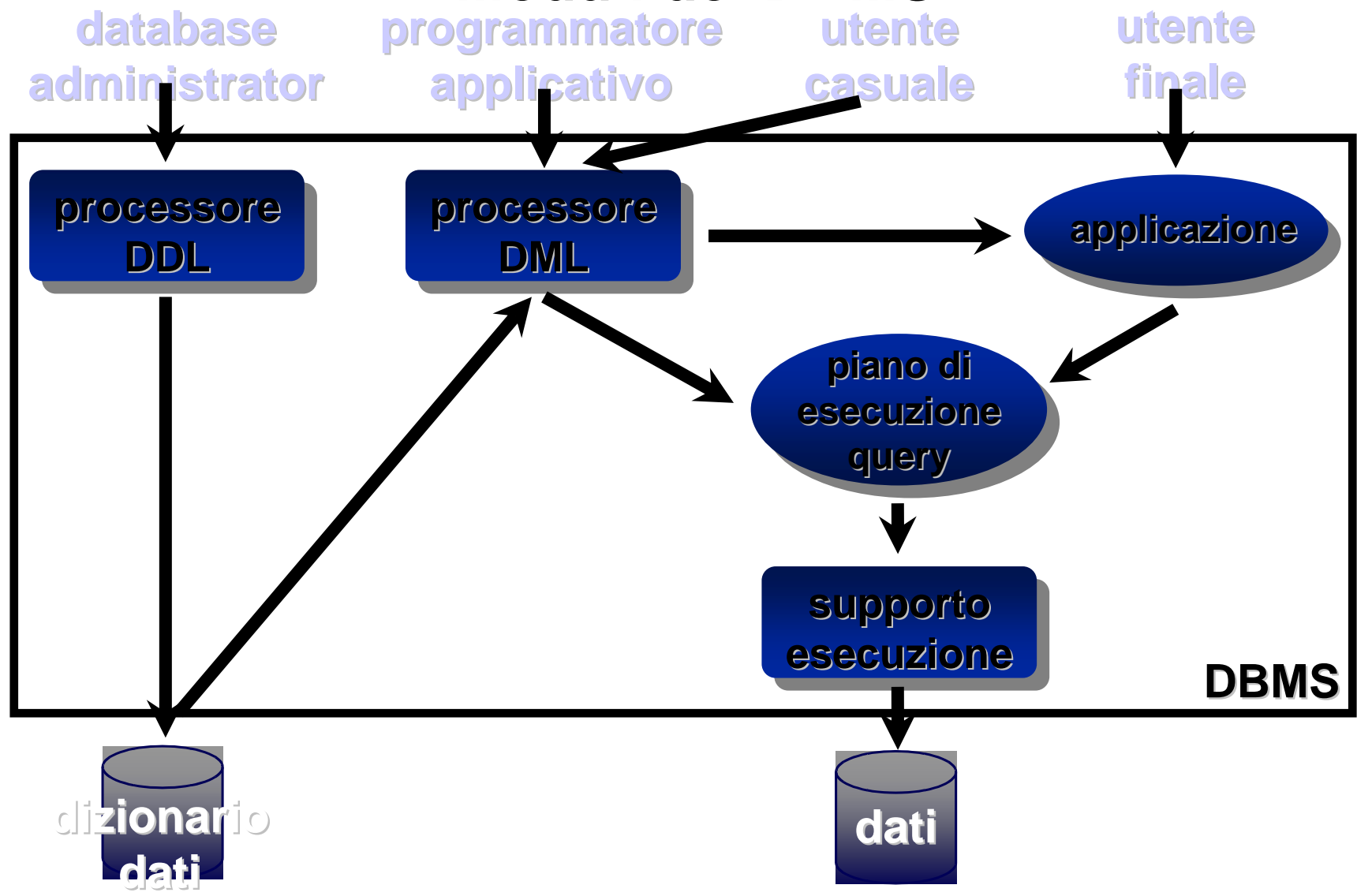
- Persona o gruppo di persone responsabile del controllo centralizzato e della gestione del sistema, delle prestazioni, dell'affidabilità, delle autorizzazioni.
- Le funzioni del DBA includono anche (parte di) quelle di progettazione.

# Utenti del DBMS

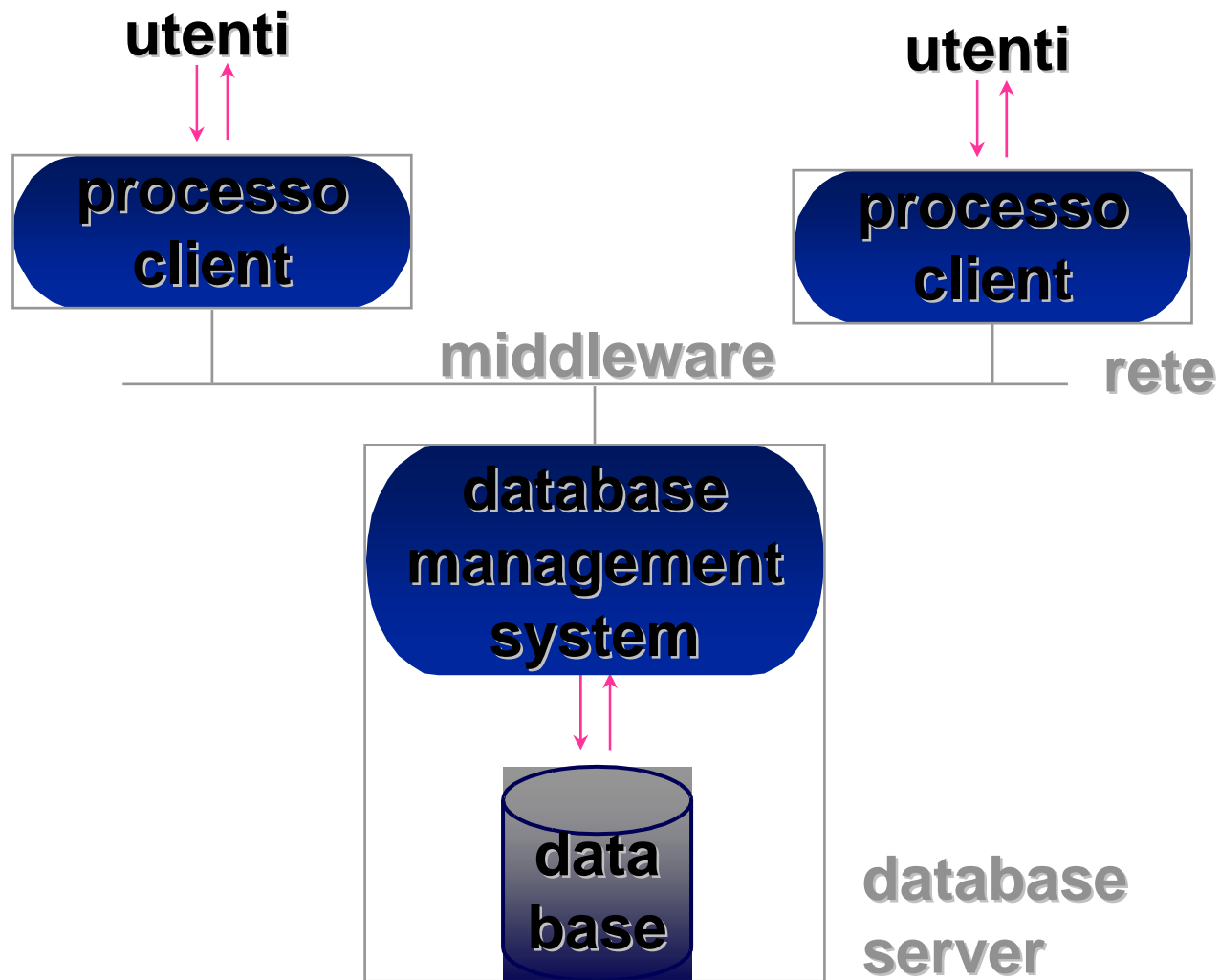




# Moduli del DBMS



# DBMS in un contesto di rete



# Transazioni

- Programmi che realizzano attività frequenti e predefinite, con poche eccezioni, previste a priori.
- Esempi:
  - versamento presso uno sportello bancario
  - emissione di certificato anagrafico
  - dichiarazione presso l'ufficio di stato civile
  - prenotazione aerea
- Le transazioni sono di solito realizzate con programmi in linguaggio ospite (tradizionale o ad hoc).
- **N. B.:** il termine **transazione** ha un'altra accezione, più specifica: sequenza indivisibile di operazioni (o vengono eseguite tutte o nessuna).

# Vantaggi e svantaggi dei DBMS

## Pro

- dati come risorsa comune, base di dati come modello della realtà
- gestione centralizzata con possibilità di standardizzazione ed “economia di scala”
- disponibilità di servizi integrati
- riduzione di ridondanze e inconsistenze
- indipendenza dei dati (favorisce lo sviluppo e la manutenzione delle applicazioni)

## Contro

- costo dei prodotti e della transizione verso di essi
- non scorporabilità delle funzionalità (con riduzione di efficienza)

# Struttura del corso

Prima parte

- Basi di dati relazionali: modello, linguaggi formali, SQL

Seconda parte

- Progetto di basi di dati