

# Basi di dati

## Introduzione

# Motivazioni

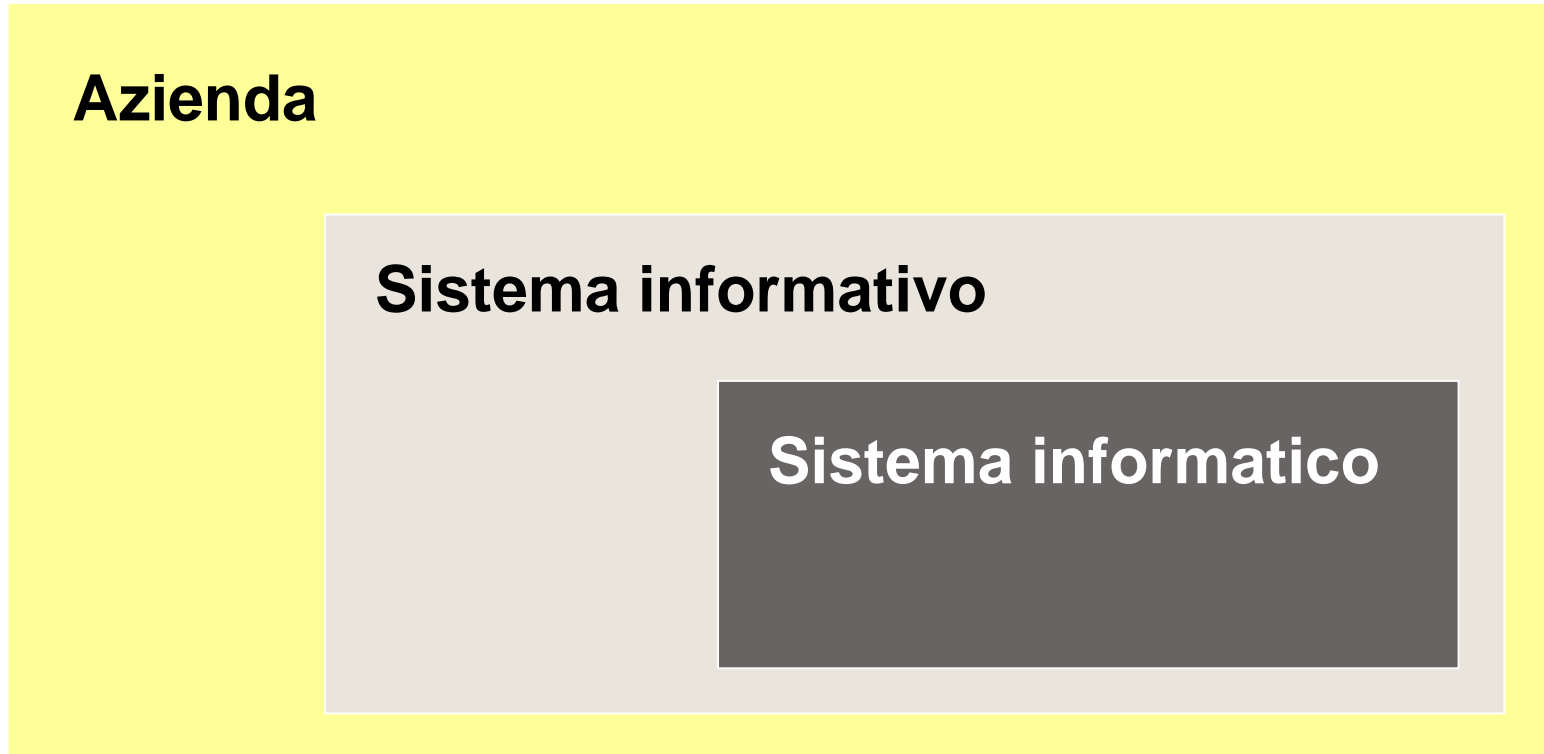
- L'uso del calcolatore per gestire dati riveste un ruolo chiave in numerosi contesti applicativi:
  - **Banking**: informazioni sui clienti, i conti correnti, i mutui, bonifici e transazioni bancarie in genere, carte di credito...
  - **Finanza**: aggiornare la quotazione delle società di borsa in base agli scambi, memorizzare informazioni su quotazione e volumi a fine giornata
  - **Linee aeree**: prenotazioni ed orario dei voli.
  - **Università**: info su studenti, corsi, docenti, esami sostenuti, diplomi.
  - **Telecomunicazioni**: memorizzare le chiamate effettuate, estrarre bollette periodiche, gestire sistemi di carte prepagate
  - **Grande distribuzione**: info su clienti (carte fedeltà), prodotti, ordini, acquisti, scorte...
  - **Commercio elettronico**: come sopra + gestione elettronica catalogo, funzioni di pagamento e spedizione
  - **Azienda**: gestione info sulla catena di produzione, gestione inventario ed ordini, info su impiegati, salario base, benefit, tasse, pagamento stipendi
  - **Ospedali**: gestione cartelle cliniche, esami, ricoveri, ...
  - **Pagine Web e Social Networks**: motori di ricerca, riviste online, blogs, forum, ...
  - **GIS**: mappe del territorio, risorse sul territorio, routing, ...

# Motivazioni

**Azienda**

**Sistema informativo**

**Sistema informatico**



# Motivazioni

- Nei sistemi informatici le informazioni vengono rappresentate per mezzo di dati
  - I dati non sono l'informazione

Che informazioni posso estrarre da questi dati ?

```
4:TOSHIBA:15VL34G:1:349,00
5:SINUDYNE:20250MEV:1:399,00
2:SAMSUNG:LE15S51B:1:399,00
2:SHARP:LC15SH1E:1:399,00
1:PHILIPS:20PF7835:1:449,00
0:JVC:LT17E31BU:1:449,00
4:QUASAR:Plasma 42":2:1.299,00
3:SAMSUNG:PS42V4S:2:1.790,00
0:PANASONIC:TH37PA50:2:1.790,00
2:PHILIPS:BDS4221:2:1.799,00
4:ROADSTAR:CTV1434TS:2:79,90
3:SINUDYNE:14605:2:89,90
0:PHILIPS:14PT1547:2:99,00
3:ROADSTAR:CTV2039TS:2:109,00
5:ROADSTAR:CTV2034TS:2:119,90
```

# Motivazioni

- Nell'esempio, l'informazione è estratta associando una semantica alla struttura con cui i dati sono memorizzati:
  - Il carattere ':' è usato per separare dati diversi
  - Il primo dato rappresenta il numero di prodotti disponibili
  - Il secondo dato rappresenta la marca del prodotto
  - Il terzo dato rappresenta il modello di prodotto
  - Il quarto dato rappresenta il tipo di prodotto (1:LCD, 2:PLASMA, 3:CRT)
  - Il quinto dato rappresenta il costo del prodotto

```
4:TOSHIBA:15VL34G:1:349,00
5:SINUDYNE:20250MEV:1:399,00
2:SAMSUNG:LE15S51B:1:399,00
2:SHARP:LC15SH1E:1:399,00
1:PHILIPS:20PF7835:1:449,00
0:JVC:LT17E31BU:1:449,00
4:QUASAR:Plasma 42":2:1.299,00
3:SAMSUNG:PS42V4S:2:1.790,00
0:PANASONIC:TH37PA50:2:1.790,00
2:PHILIPS:BDS4221:2:1.799,00
4:ROADSTAR:CTV1434TS:3:79,90
3:SINUDYNE:14605:3:89,90
0:PHILIPS:14PT1547:3:99,00
3:ROADSTAR:CTV2039TS:3:109,00
5:ROADSTAR:CTV2034TS:3:119,90
```

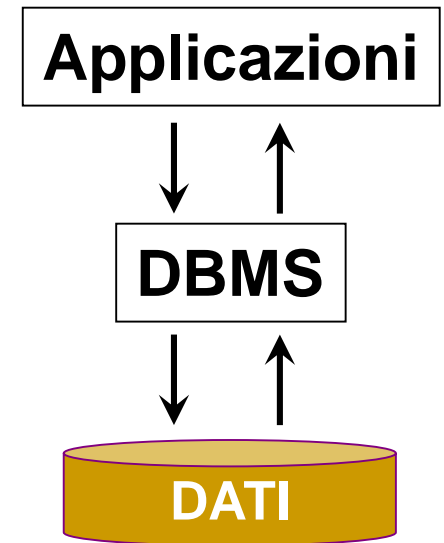
# Motivazioni

- La gestione del magazzino potrebbe essere effettuata attraverso un programma “gest\_TV” (per esempio scritto in linguaggio c, C++, Java o Python) che leggendo e scrivendo sul file di testo consenta di:
  - Visualizzare il numero di prodotti in magazzino
  - Aggiornare il numero di prodotti in magazzino
  - Aggiornare il costo dei prodotti
  - Aggiungere/rimuovere prodotti dal magazzino

**Possibili criticità!!!**

# Base di dati

- Queste criticità possono essere superate attraverso l'introduzione di un modulo software (DBMS) che operi da intermediario tra le applicazioni ed i dati sul disco
- Un **database** è quindi un insieme di dati gestito da un DBMS



# DataBase Management System – DBMS

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

- grandi
- persistenti
- garantendo **affidabilità**
- **privacy**
- **condivisione**
- **efficacia**
- **efficienza**



# Condivisione

- Ogni organizzazione (specie se grande) è divisa in settori/reparti o comunque svolge diverse attività
- A ciascun settore o attività corrisponde un (sotto)sistema informativo
  - Reparto ordini e magazzino di un negozio
- 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

- Senza condivisione si ha ridondanza di dati

## Magazzino.txt

4:TOSHIBA:15VL34G:1:349,00

5:SINUDYNE:20250MFV:1:399,00

2:SAMSUNG:LE15S51B:1:399,00

2:SHARP:LC15SH1E:1:399,00

1:PHILIPS:20PF7835:1:449,00

0:JVC:LT17E31BU:1:449,00

4:QUASAR:Plasma 42":2:1.299,00

3:SAMSUNG:PS42V4S:2:1.790,00

0:PANASONIC:TH37PA50:2:1.790,00

2:PHILIPS:BDS4221:2:1.799,00

4:ROADSTAR:CTV1434TS:3:79,90

3:SINUDYNE:14605:3:89,90

0:PHILIPS:14PT1547:3:99,00

3:ROADSTAR:CTV2039TS:3:109,00

5:ROADSTAR:CTV2034TS:3:119,90

## Ordini.txt

3:PHILIPS:20PF7835:1

3:JVC:LT17E31BU:1

4:PANASONIC:TH37PA50:2

2:PHILIPS:BDS4221:2

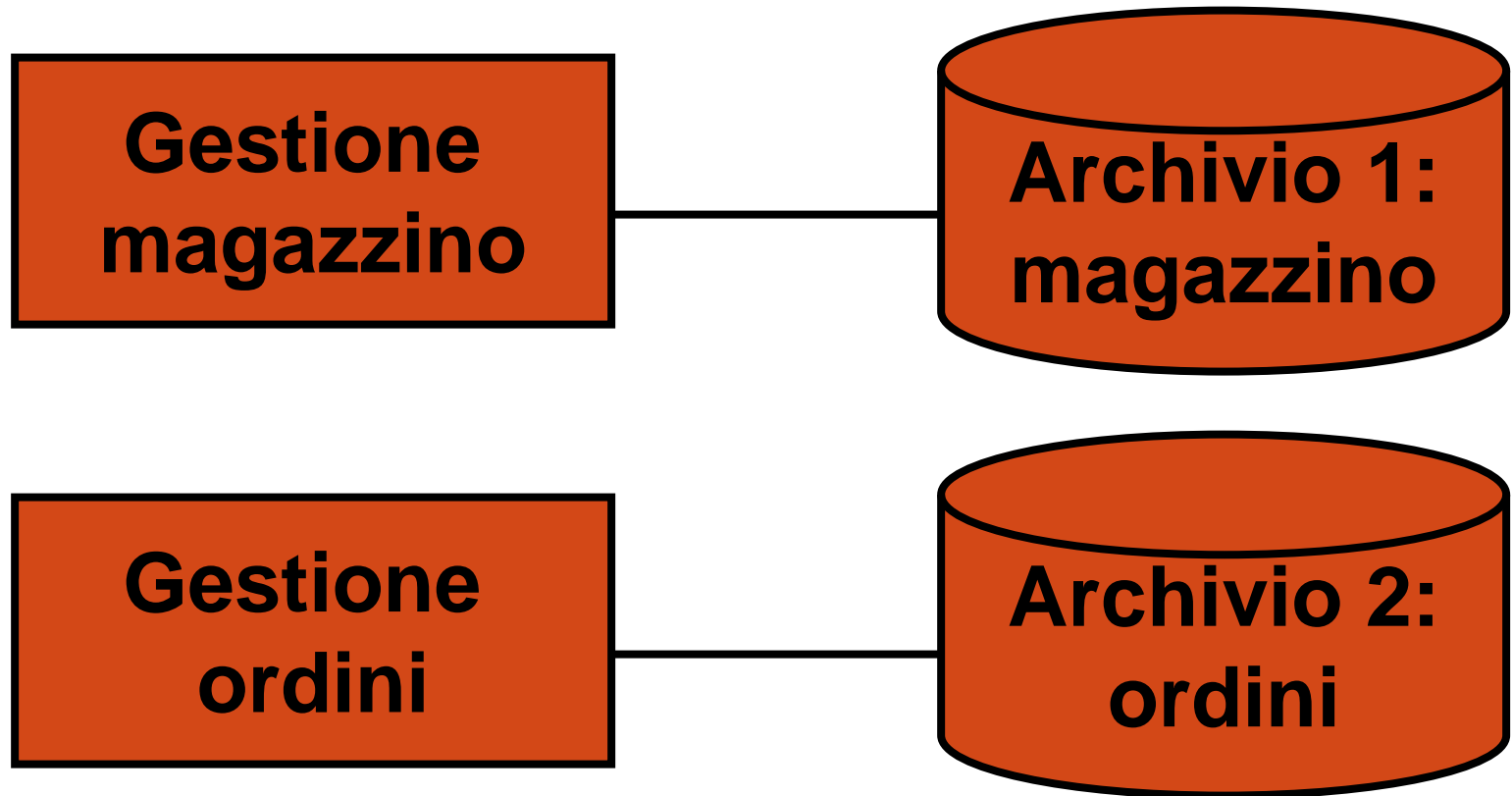
4:PHILIPS:14PT1547:3

# Possibili problemi

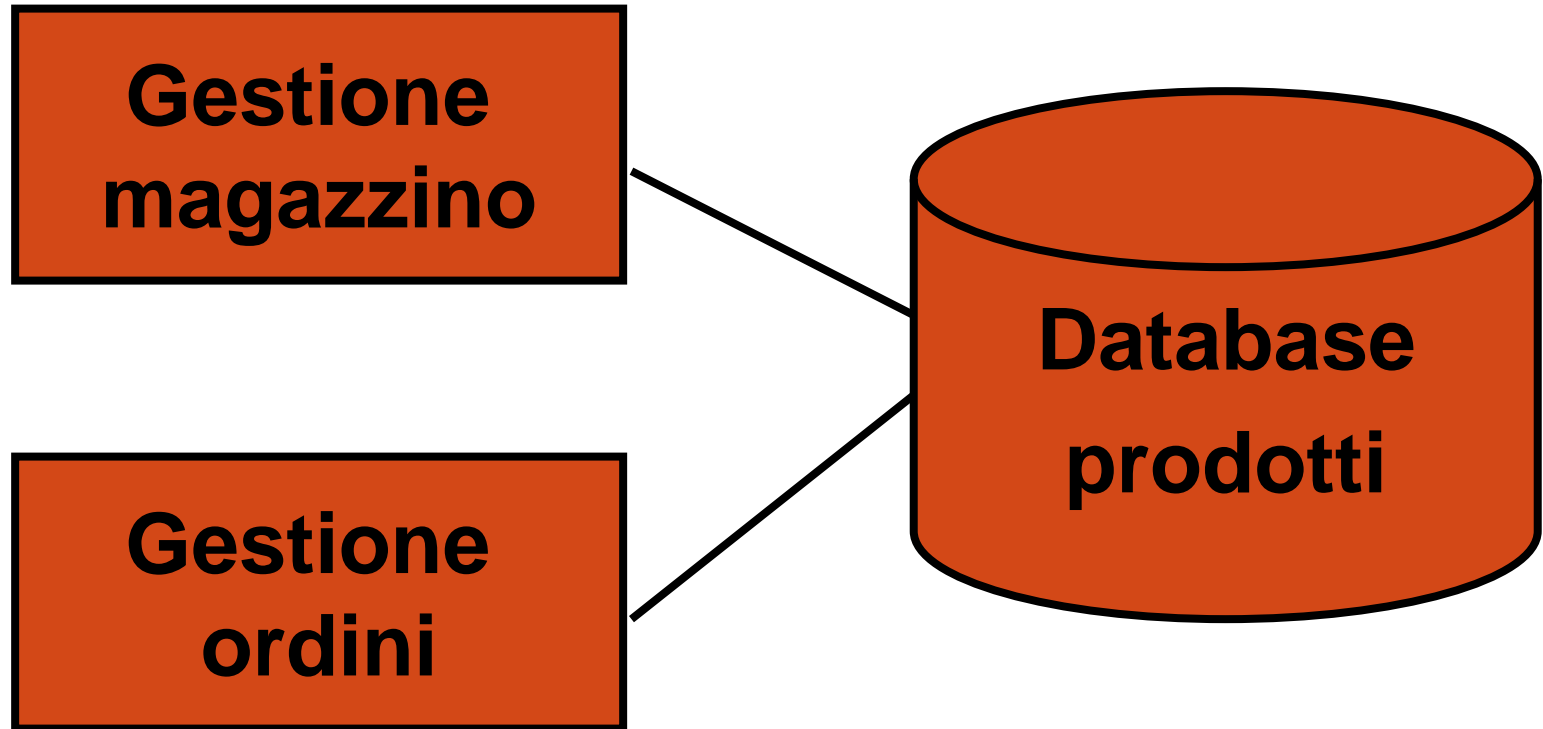
- **Ridondanza**

- Informazioni ripetute e quindi **spreco di memoria**
- Rischio di **incoerenza**:
  - Aggiornare un'informazione richiede una modifica di tutte le repliche del dato. In caso contrario, le repliche possono non coincidere

# Archivi e basi di dati

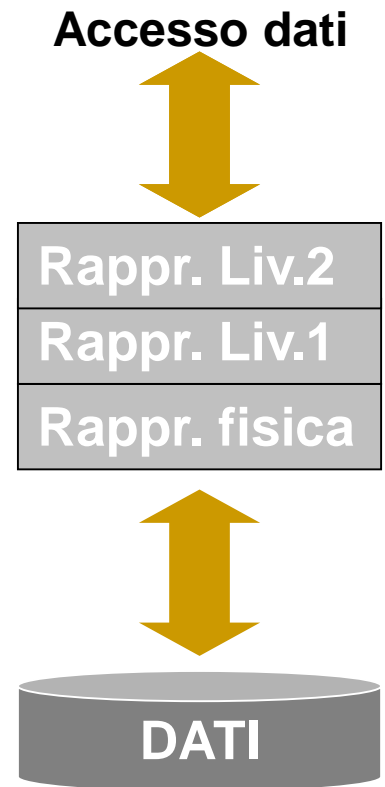


# Archivi e basi di dati



# Efficacia e rappresentazione dei dati

- Semplificazione delle operazioni di accesso attraverso diversi livelli di **astrazione nella rappresentazione dei dati**
- Programmi ed utenti fanno riferimento alla rappresentazione al livello più alto
- Indipendenza dei livelli di rappresentazione
- La rappresentazione a diversi livelli di astrazione è resa possibile attraverso il ricorso ad un **modello dei dati**



# Modello dei dati

- Tipologie di modelli:
  - **Modelli logici**: utilizzati nei DBMS per l'organizzazione dei dati (**relazionale**, reticolare, gerarchico, a oggetti )
    - Descrivono i dati in modo da specificare come i programmi possono estrarli
    - Rendono i dati indipendenti dalle strutture fisiche di memorizzazione
  - **Modelli concettuali**: permettono di rappresentare i dati in modo indipendente da ogni DBMS (il più noto è il modello **Entity-Relationship**)
    - Cercano di descrivere i concetti del mondo reale
    - Sono utilizzati nella prima fase della progettazione

# Modello relazionale

- Il modello logico più diffuso è quello relazionale
- Informazioni organizzate in una o più relazioni (tabelle):
  - insiemi di record a struttura fissa

## Orario

Insegnamento	Docente	Aula	Ora
Analisi matem. I	Luigi Neri	N1	8:00
Basi di dati	Piero Rossi	N2	9:45
Chimica	Nicola Mori	N1	9:45
Fisica I	Mario Bruni	N1	11:45
Fisica II	Mario Bruni	N3	9:45
Sistemi inform.	Piero Rossi	N3	8:00



# Lo **schema** della base di dati

## Orario

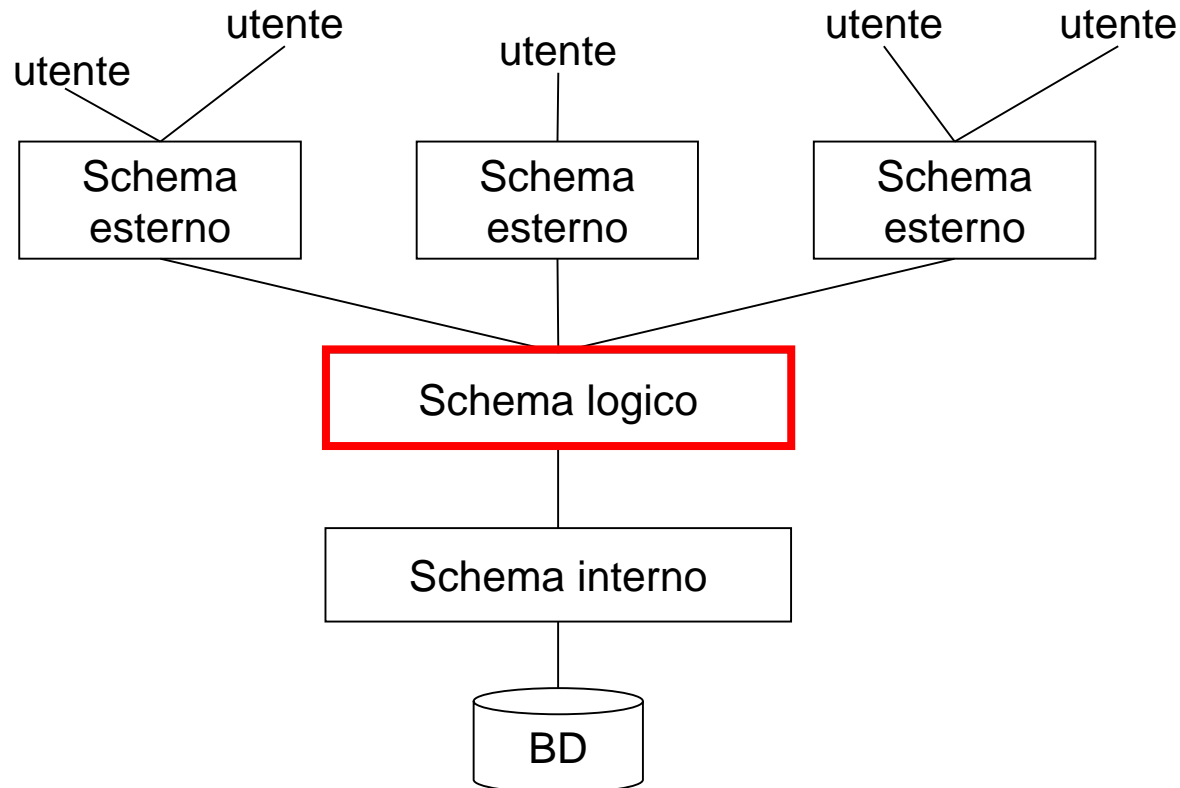
Insegnamento	Docente	Aula	Ora
Analisi matem. I	Luigi Neri	N1	8:00
Basi di dati	Piero Rossi	N2	9:45
Chimica	Nicola Mori	N1	9:45
Fisica I	Mario Bruni	N1	11:45
Fisica II	Mario Bruni	N3	9:45
Sistemi inform.	Piero Rossi	N3	8:00

## L'**istanza** della base di dati

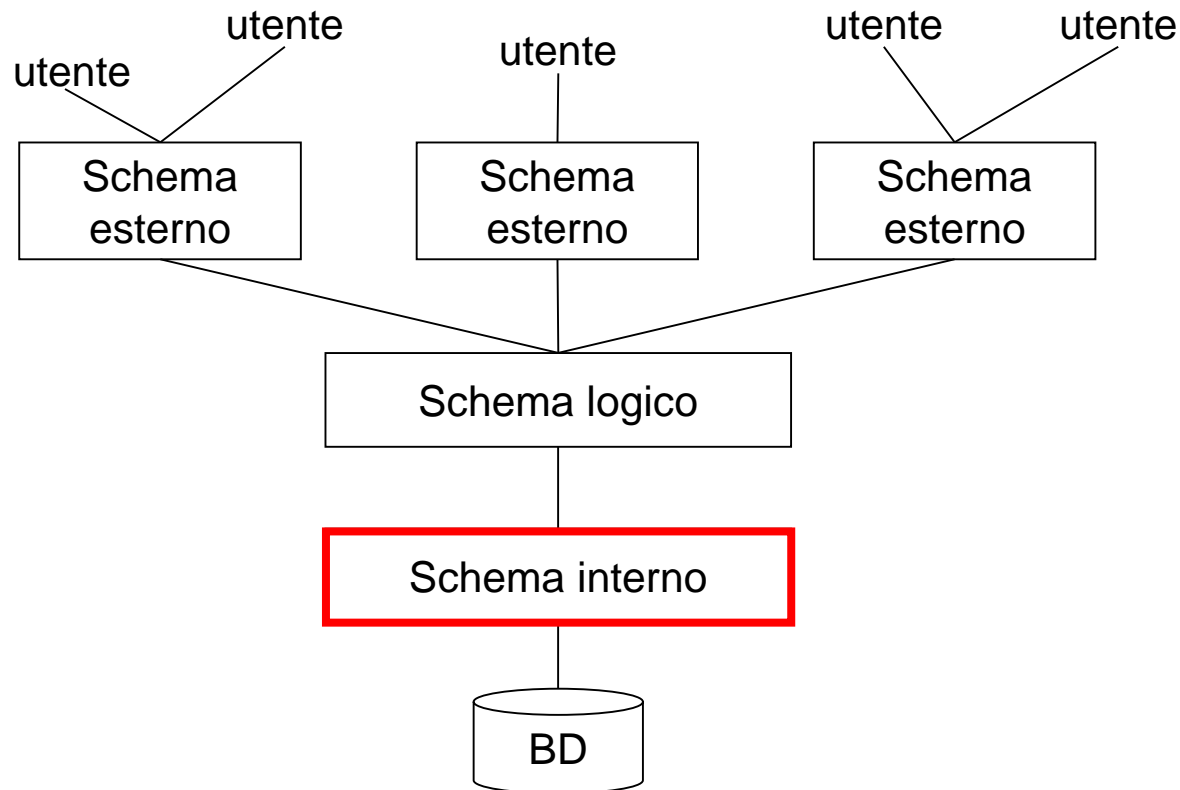
# Architettura di un DBMS

- La nozione di schema può essere ulteriormente specificata attraverso l'individuazione di tre livelli che definiscono l'architettura tipica di un moderno DBMS:
  - Schema esterno (Livello esterno)
  - Schema logico (Livello logico)
  - Schema interno (Livello fisico)

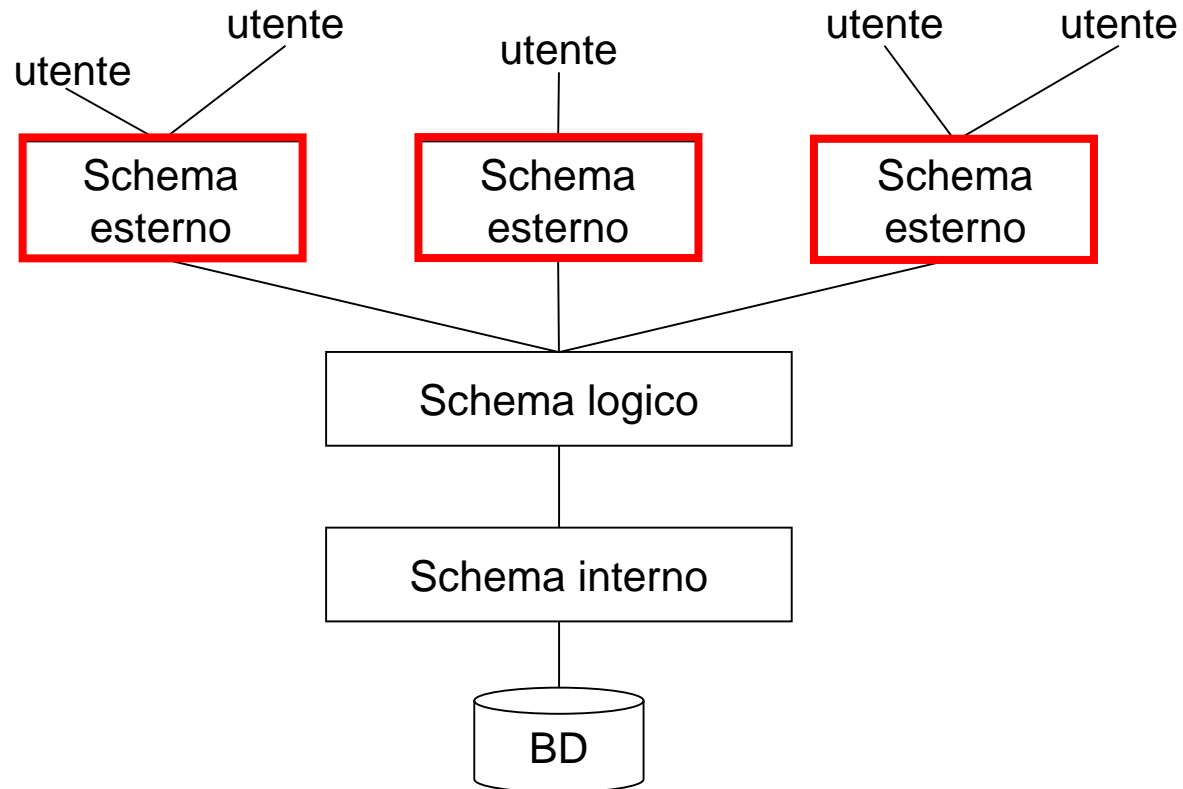
# Architettura di un DBMS



# Architettura di un DBMS



# Architettura di un DBMS



## Livello logico

### Televisore

Marca	Modello	Cod	Costo vendita	Costo fornitore
TOSHIBA	15VL34G	1	349,00	249,00
SINUDYNE	20250MFV	1	399,00	310,00
SAMSUNG	LE15S51B	1	399,00	320,00
SHARP	LC15SH1E	1	399,00	315,00
PHILIPS	20PF7835	1	449,00	349,00
JVC	LT17E31BU	1	449,00	349,00
QUASAR	Plasma 42"	2	1.299,00	1.100,00
SAMSUNG	PS42V4S	2	1.790,00	1.600,00
PANASONIC	TH37PA50	2	1.790,00	1.650,00
PHILIPS	BDS4221	2	1.799,00	1.600,00
ROADSTAR	CTV1434TS	3	79,90	70,00
SINUDYNE	14605	3	89,90	80,00
PHILIPS	14PT1547	3	99,00	89,00
ROADSTAR	CTV2039TS	3	109,00	99,00
ROADSTAR	CTV2034TS	3	119,90	105,00

**La vista presenta solo i TV al plasma nascondendo info su costo al fornitore**

## Livello logico

### Tecnologia

Codice	Tipo
1	LCD
2	Plasma
3	CRT

## Livello esterno

### Televisore\_plasma

Marca	Modello	Costo vendita
QUASAR	Plasma 42"	1.299,00
SAMSUNG	PS42V4S	1.790,00
PANASONIC	TH37PA50	1.790,00
PHILIPS	BDS4221	1.799,00

Livello logico

Viste

Livello logico

Televisore

Marca	Modello	Cod	Costo vendita	Costo fornitore
TOSHIBA	15VL34G	1	349,00	249,00
SINUDYNE	20250MFV	1	399,00	310,00
SAMSUNG	LE15S51B	1	399,00	320,00
SHARP	LC15SH1E	1	399,00	315,00
PHILIPS	20PF7835	1	449,00	349,00
JVC	LT17E31BU	1	449,00	349,00
QUASAR	Plasma 42"	2	1.299,00	1.100,00
SAMSUNG	PS42V4S	2	1.790,00	1.600,00
PANASONIC	TH37PA50	2	1.790,00	1.650,00
PHILIPS	BDS4221	2	1.799,00	1.600,00
ROADSTAR	CTV1434TS	3	79,90	70,00
SINUDYNE	14605	3	89,90	80,00
PHILIPS	14PT1547	3	99,00	89,00
ROADSTAR	CTV2039TS	3	109,00	99,00
ROADSTAR	CTV2034TS	3	119,90	105,00

Tecnologia

Codice	Tipo
1	LCD
2	Plasma
3	CRT

Livello esterno

Marca	Modello	Tipo	Costo vendita
TOSHIBA	15VL34G	LCD	349,00
SINUDYNE	20250MFV	LCD	399,00
SAMSUNG	LE15S51B	LCD	399,00
SHARP	LC15SH1E	LCD	399,00
PHILIPS	20PF7835	LCD	449,00
JVC	LT17E31BU	LCD	449,00
QUASAR	Plasma 42"	Plasma	1.299,00
SAMSUNG	PS42V4S	Plasma	1.790,00
PANASONIC	TH37PA50	Plasma	1.790,00
PHILIPS	BDS4221	Plasma	1.799,00
ROADSTAR	CTV1434TS	CRT	79,90
SINUDYNE	14605	CRT	89,90
PHILIPS	14PT1547	CRT	99,00
ROADSTAR	CTV2039TS	CRT	109,00
ROADSTAR	CTV2034TS	CRT	119,90

La vista integra le info presenti su due relazioni nascondendo info su costo al fornitore

# Indipendenza dei dati

- L'architettura a livelli garantisce una fondamentale proprietà dei DBMS:
  - **Indipendenza dei dati**  
consente, ad utenti e programmi applicativi di interagire con i dati ad un livello di astrazione che prescinde dai dettagli realizzativi del DBMS
- Con riferimento all'architettura a tre livelli si parla di:
  - **Indipendenza fisica**
  - **Indipendenza logica**



# Linguaggi per basi di dati

- Un altro contributo all'efficacia dei DBMS è dato dalla disponibilità di vari linguaggi e interfacce per la loro gestione:
  - ⇒ comandi testuali interattivi (**SQL**)
  - ⇒ comandi SQL immersi in un **linguaggio ospite** (Pascal, C, C++, Python, Cobol, JAVA, etc.)
  - ⇒ comandi SQL 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)
  - ⇒ con **interfacce grafiche** (senza linguaggio testuale)

# SQL, un linguaggio interattivo

**SELECT Marca, Modello, Costo**  
**FROM Televisori**  
**WHERE Costo < 800**

Marca	Modello	Costo
TOSHIBA	15VL34G	349,00
SINUDYNE	20250MFV	399,00
SAMSUNG	LE15S51B	399,00
SHARP	LC15SH1E	399,00
PHILIPS	20PF7835	449,00
JVC	LT17E31BU	449,00
ROADSTAR	CTV1434TS	79,90
SINUDYNE	14605	89,90
PHILIPS	14PT1547	99,00
ROADSTAR	CTV2039TS	109,00
ROADSTAR	CTV2034TS	119,90

**Televisori**

Marca	Modello	Tipo	Costo	Costo fornitore
TOSHIBA	15VL34G	1	349,00	249,00
SINUDYNE	20250MFV	1	399,00	310,00
SAMSUNG	LE15S51B	1	399,00	320,00
SHARP	LC15SH1E	1	399,00	315,00
PHILIPS	20PF7835	1	449,00	349,00
JVC	LT17E31BU	1	449,00	349,00
QUASAR	Plasma 42"	2	1.299,00	1.100,00
SAMSUNG	PS42V4S	2	1.790,00	1.600,00
PANASONIC	TH37PA50	2	1.790,00	1.650,00
PHILIPS	BDS4221	2	1.799,00	1.600,00
ROADSTAR	CTV1434TS	3	79,90	70,00
SINUDYNE	14605	3	89,90	80,00
PHILIPS	14PT1547	3	99,00	89,00
ROADSTAR	CTV2039TS	3	109,00	99,00
ROADSTAR	CTV2034TS	3	119,90	105,00



# SQL immerso in 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 (Access)

The screenshot shows the Microsoft Access interface. The title bar reads 'Microsoft Access'. The menu bar includes 'File', 'Modifica', 'Visualizza', 'Inserisci', 'Query', 'Strumenti', 'Finestra', and '?'. The toolbar contains various icons for file operations, editing, and querying. The main window displays a query named 'CorsiAPianoTerra: Query di selezione'. The query design grid is visible, showing the following fields and criteria:

Campo:	Corso	Aula	Piano	
Tabella:	Corsi	Corsi	Aule	
Ordinamento:				
Mostra:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Criteri:			"Terra"	
Oppure:				

The status bar at the bottom shows 'Pronto' and a numeric keypad.

# 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**
  - utenti **casuali**

# 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)

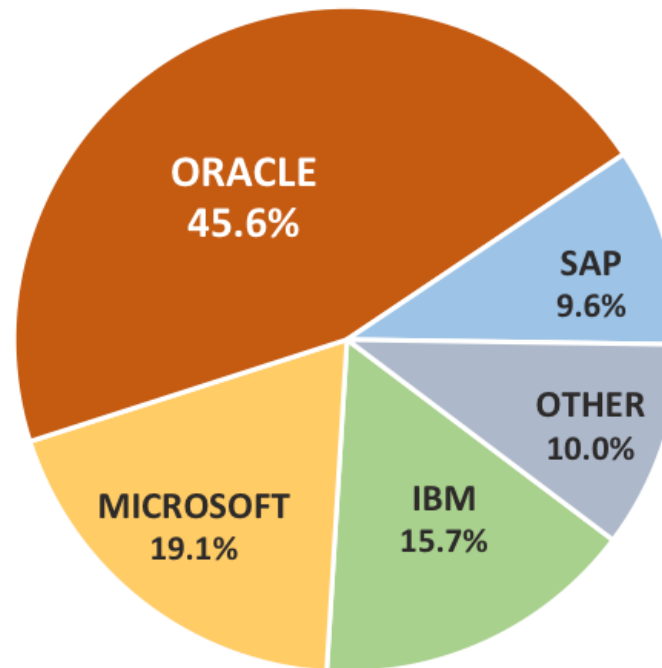
## Contro

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

# Alcuni DBMS in commercio

## 2016 COMMERCIAL DATABASE MARKET SHARE

(Source: Gartner, Inc. 2016)



## RDBMS/ORDBMS

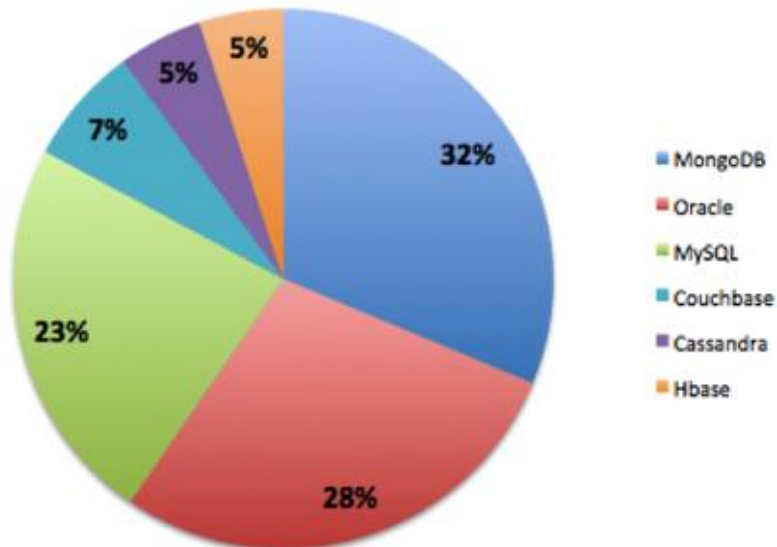
- Oracle
- Microsoft SQLServer
- IBM DB2
- SAP Sybase Adaptive Server Enterprise
- PostgreSQL
- MariaDB Enterprise
- MySQL
- ...



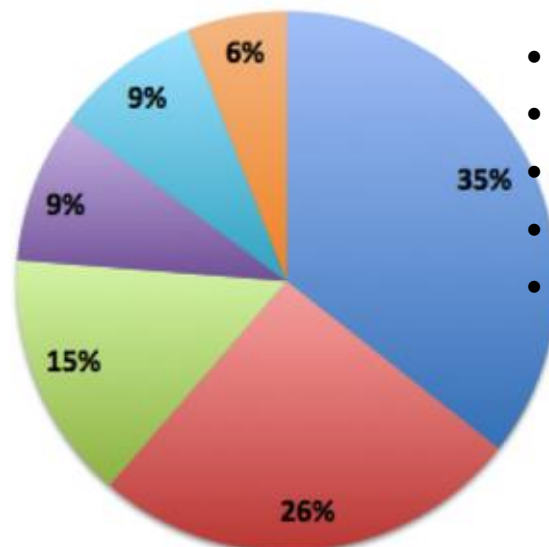
# DBMS market changes

- Something big is happening to the database market, driven in significant part by **Big Data** adoption: ever-rising interest in **NoSQL** databases (schemaless)
- All NoSQLs are growing, and some quite substantially, and mostly at MySQL's expense

Database Popularity  
March 2013-2014



Database Popularity  
March 2014-2015



## NotOnlySQL

- MongoDB
- Amazon DynamoDB
- Couchbase
- Apache Cassandra
- Marklogic

FINE