
Basi di dati

OBIETTIVI DEL CORSO

- Modelli dei dati, linguaggi e sistemi per lo sviluppo di applicazioni che prevedono l'uso di grandi quantità di dati permanenti organizzati in basi di dati.
- Argomenti
 - Funzionalità dei sistemi per basi di dati-DBMS
 - La modellazione a oggetti
 - I sistemi relazionali e il linguaggio SQL
 - Teoria delle basi di dati relazionali
 - Cenni sui Sistemi NoSQL
 - Architettura dei DBMS

TESTO DI RIFERIMENTO

- A. Albano, G. Ghelli e R. Orsini, *Fondamenti di basi di dati*, Zanichelli, Bologna, 2005, liberamente scaricabile da:
 - <http://fondamentidibasididati.it/>

MODALITÀ DI ESAME

- Scritto:
 - Compitini di verifica intermedi, indipendenti, validi per giugno/luglio
 - Giugno/Luglio:
 - Scritto per il recupero di un compitino, o di entrambi
 - Altri appelli:
 - Scritto obbligatorio
- Orale:
 - Obbligatorio

BASI DI DATI, DATA WAREHOUSE, DATA LAKE, DATA SCIENCE...

- Base di dati: tecnologia di base, gestione delle attività quotidiane dell'organizzazione, il tema di questo corso
- Data Warehouse, Data Lake, Big Data, Data Science: termini che hanno a che vedere con l'analisi dei dati, e non rientrano tra i temi di questo corso

ASPETTI TRATTATI IN ALTRI CORSI

- Sperimentazione: Laboratorio di Basi di dati
- Approfondimento di architetture e algoritmi per SGBD: Advanced Data Bases
- Sistemi per il supporto alle decisioni, business intelligence, data warehousing, data mining, big data
- Information retrieval

ESEMPIO DI BASE DI DATI

Materie

Titolo	Codice	Syllabus
Basi di Dati	AA024	Progettazione e interrogazione...
Reti di Calc.	AA019	Realizzazione e uso di reti - protocollo TCP

Corsi

Materia	AA	Sem	Titolare
AA024	2007	1	Albano
AA024	2007	1	Ghelli
AA019	2007	1	Brogi

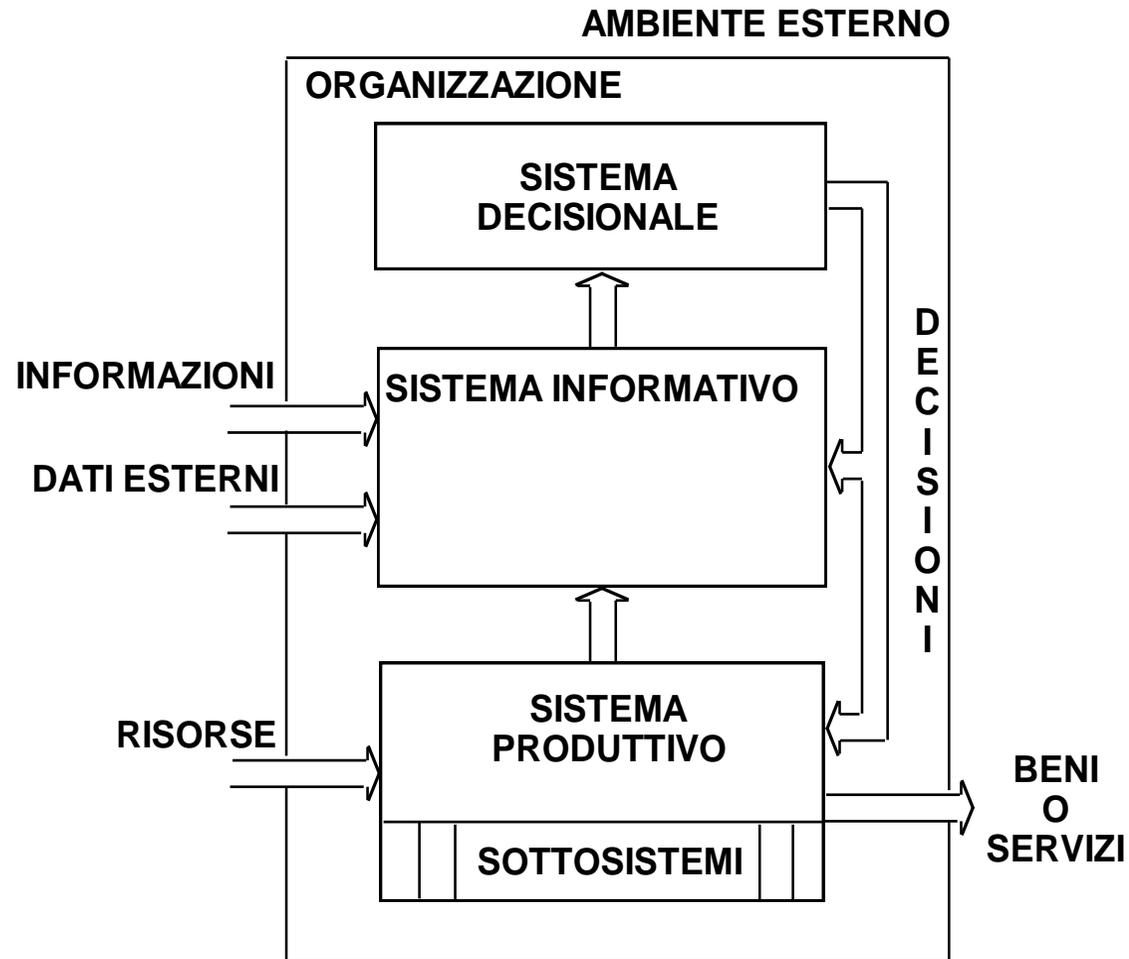
SISTEMI INFORMATIVI

- *Definizione* Un sistema informativo di un'organizzazione è una combinazione di risorse, umane e materiali, e di procedure organizzate per:
 - la raccolta,
 - l'archiviazione,
 - l'elaborazione e
 - lo scambio

delle informazioni necessarie alle attività:

- operative (informazioni di servizio),
- di programmazione e controllo (informazioni di gestione), e
- di pianificazione strategica (informazioni di governo).

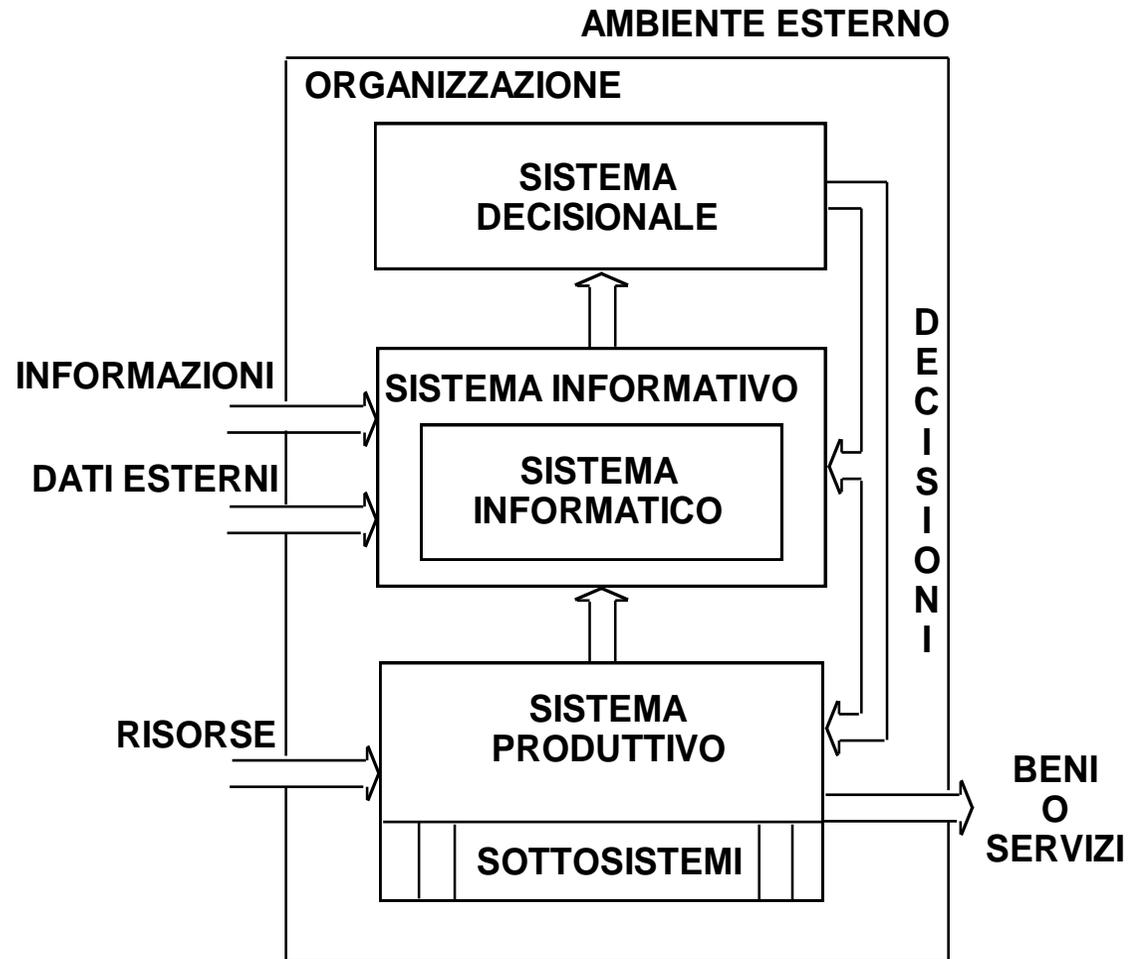
SISTEMA INFORMATIVO NELLE ORGANIZZAZIONI



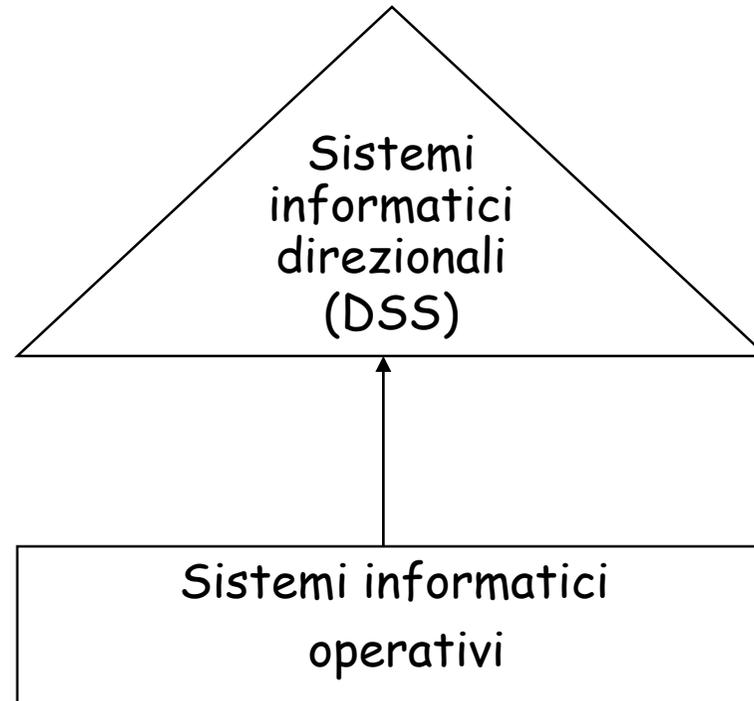
SISTEMI INFORMATICI

- Il **sistema informatico** è l'insieme delle tecnologie informatiche e della comunicazione (Information and Communication Technologies, ICT) a supporto delle attività di un'organizzazione.
- Il **sistema informativo automatizzato** è quella parte del sistema informativo in cui le informazioni sono raccolte, elaborate, archiviate e scambiate usando un sistema informatico.
- Terminologia
 - sistema informativo \approx sistema informativo automatizzato
 - sistema informativo automatizzato \approx sistema informatico

SISTEMA INFORMATICO NELLE ORGANIZZAZIONI



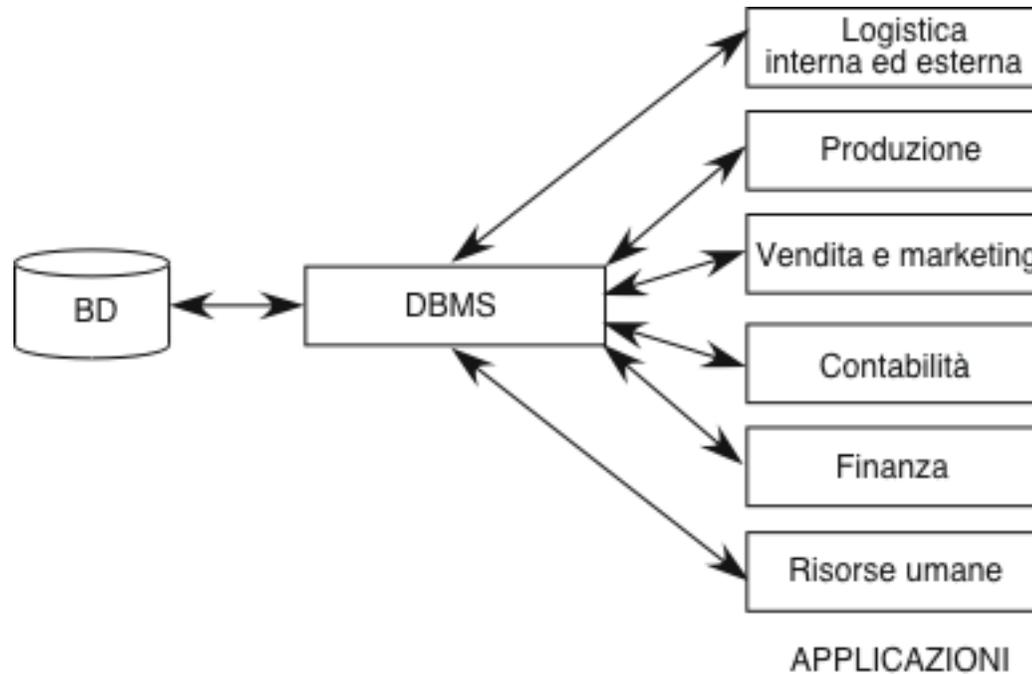
CLASSIFICAZIONE DEI SISTEMI INFORMATICI



SISTEMI INFORMATICI OPERATIVI

- I dati sono organizzati in BD
- Le applicazioni si usano per svolgere le classiche attività strutturate e ripetitive dell'azienda nelle aree amministrativa e finanziaria, vendite, produzione, risorse umane ecc.
- Alcune sigle
 - Data processing (DP), Electronic Data processing (EDP)
 - Transaction Processing Systems (TPS)

SISTEMA INFORMATICO OPERATIVO (cont.)



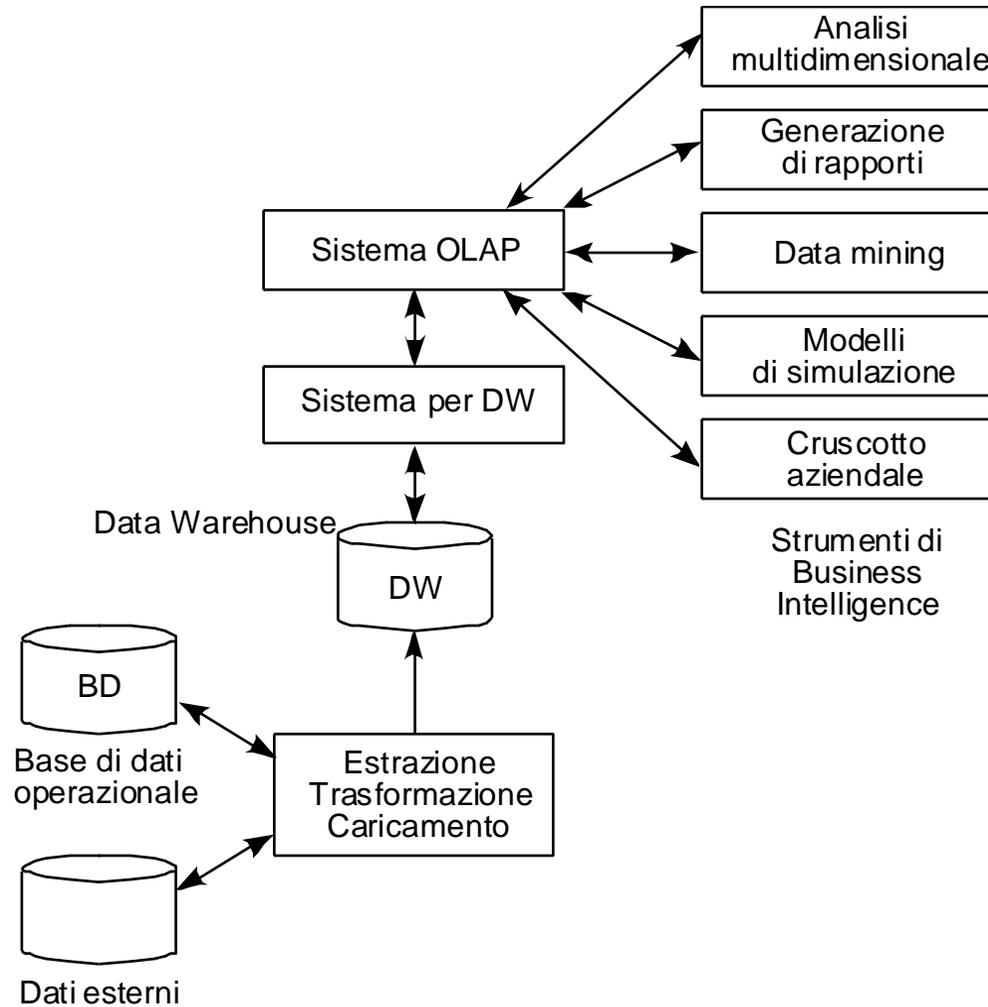
ELABORAZIONI SU BD: OLTP (On-Line Transaction Processing)

- Uso principale dei DBMS
- Tradizionale elaborazione di transazioni, che realizzano i processi operativi per il funzionamento di organizzazioni:
 - Operazioni predefinite e relativamente semplici
 - Ogni operazione coinvolge "pochi" dati
 - Dati di dettaglio, aggiornati

SISTEMI INFORMATICI DIREZIONALI (Decision Support Systems)

- I dati sono organizzati in Data Warehouse (DW) e gestiti da un opportuno sistema
- Le applicazioni, dette di Business intelligence, sono strumenti di supporto ai processi di controllo delle prestazioni aziendali e di decisione manageriale
- Terminologia anglosassone:
 - Management Information Systems (MIS)
 - Decision support systems (DSS), data or model based
 - Executive Information System (EIS)

SISTEMA INFORMATICO DIREZIONALE



ELABORAZIONI SU DW: OLAP (On-Line Analytical Processing)

- Uso principale dei **data warehouse**
- Analisi dei dati di supporto alle decisioni
 - Operazioni complesse e casuali
 - Ogni operazione può coinvolgere molti dati
 - Dati aggregati, storici, anche non attualissimi

DIFFERENZE TRA OLTP E OLAP

	OLTP	OLAP
Scopi	Supporto operatività	Supporto decisioni
Utenti	Molti, esecutivi	Pochi, dirigenti e analisti
Dati	Analitici, relazionali	Sintetici, multidimensionali
Usi	Noti a priori	Poco prevedibili
Quantità di dati per attività	Bassa (decine)	Alta (milioni)
Orientamento	Applicazione	Soggetto
Aggiornamenti	Frequenti	Rari
Visione dei dati	Corrente	Storica
Ottimizzati per	Transazioni	Analisi dei dati

ANALISI DEI DATI: REQUISITI

- Dati **aggregati**: non interessa un dato, ma la somma, la media, il minimo, il massimo di una misura.
- Presentazione **multidimensionale**: interessa incrociare le informazioni, per analizzarle da punti di vista diversi e valutare i risultati del business per intervenire sui problemi critici o per cogliere nuove opportunità.
- Analisi a **diversi livelli di dettaglio**: per es. una volta scoperto un calo delle vendite in un determinato periodo in una regione specifica, si passa ad un'analisi dettagliata nell'area di interesse per cercare di scoprire le cause (dimensioni con gerarchie).
- I sistemi informatici operativi non sono adatti per le analisi interattive dei dati.

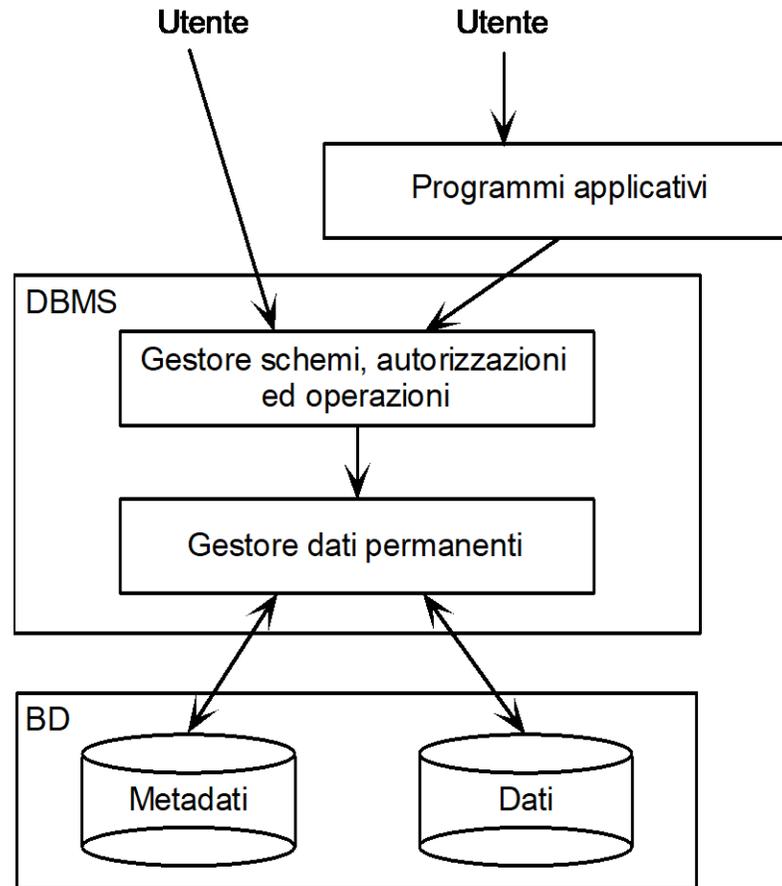
BIG DATA

- Big Data è un termine ampio, riferito a situazioni in cui l'approccio 'schema-first' tipico di DB o DW risulta troppo restrittivo o troppo lento
- Le tre V:
 - Volume
 - Varietà
 - Velocità
- I Big Data sono in genere associati a
 - Sistemi NoSQL
 - Machine learning
 - Approccio Data Lake

SISTEMI PER BASI DI DATI (DATA BASE MANAGEMENT SYSTEMS - DBMS)

- *Definizione:* Un DBMS è un sistema centralizzato o distribuito che offre opportuni linguaggi per:
 - definire lo **schema** di una basi di dati (lo schema va definito prima di creare dati),
 - scegliere le **strutture dati** per la memorizzazione dei dati,
 - memorizzare i dati rispettando i **vincoli** definiti nello schema;
 - recuperare e modificare i dati interattivamente (linguaggio di interrogazione o **query language**) o da programmi.

ARCHITETTURA DEI DBMS CENTRALIZZATI



LE BASI DI DATI GESTITE DAI DBMS

- Una base di dati è una raccolta di dati permanenti suddivisi in due categorie:
 - *i metadati* : descrivono fatti sullo schema dei dati, utenti autorizzati, applicazioni, parametri quantitativi sui dati, ecc. I metadati sono descritti da uno schema usando il modello dei dati adottato dal DBMS e sono interrogabili con le stesse modalità previste per i dati;
 - *i dati* : le rappresentazioni di certi fatti conformi alle definizioni dello schema, con le seguenti caratteristiche.

CARATTERISTICHE DEI DATI GESTITI DAI DBMS

- Sono organizzati in **insiemi strutturati e omogenei**, fra i quali sono definite delle **relazioni**. La struttura dei dati e le relazioni sono descritte nello schema usando i meccanismi di astrazione del modello dei dati del DBMS;
- Sono **molti**, in assoluto e rispetto ai metadati, e non possono essere gestiti in memoria temporanea;
- Sono accessibili mediante **transazioni**, unità di lavoro atomiche che non possono avere effetti parziali;
- Sono **protetti** sia da accesso da parte di utenti non autorizzati, sia da corruzione dovuta a malfunzionamenti hardware e software;
- Sono **utilizzabili contemporaneamente** da utenti diversi.

UN ESEMPIO DI SESSIONE CON UN DBMS RELAZIONALE

- Il modello relazionale dei dati è il più diffuso fra i DBMS commerciali.
- Il meccanismo di astrazione fondamentale è la relazione (tabella), sostanzialmente un insieme di record con campi elementari;
- Lo schema di una relazione ne definisce il nome e descrive la struttura dei possibili elementi della relazione (insieme di attributi con il loro tipo).
- Definizione base di dati:
 - `create database EsempioEsami`
- Definizione schema:
 - `create table Esami (Materia char(5), Candidato char(8), Voto int, Lode char(1), Data char(6))`

UN ESEMPIO DI SESSIONE (cont)

- Inserzione dati:

```
➤ insert into Esami values ('BDSI1', '080709', 30, 'S',  
070900)
```

- Interrogazione:

```
➤ select Candidato
```

```
➤ from Esami
```

```
➤ where Materia = "BDSI1" and Voto = 30
```

- Candidato
 - 080709
- Inoltre: accesso ai dati da programma, interrogazioni con interfaccia grafica, direttive per definire le strutture fisiche...

FUNZIONALITÀ DEI DBMS

- Linguaggio per la definizione della base di dati;
- Linguaggi per l'uso dei dati;
- Meccanismi per il controllo dei dati;
- Strumenti per il responsabile della base di dati;
- Strumenti per lo sviluppo delle applicazioni

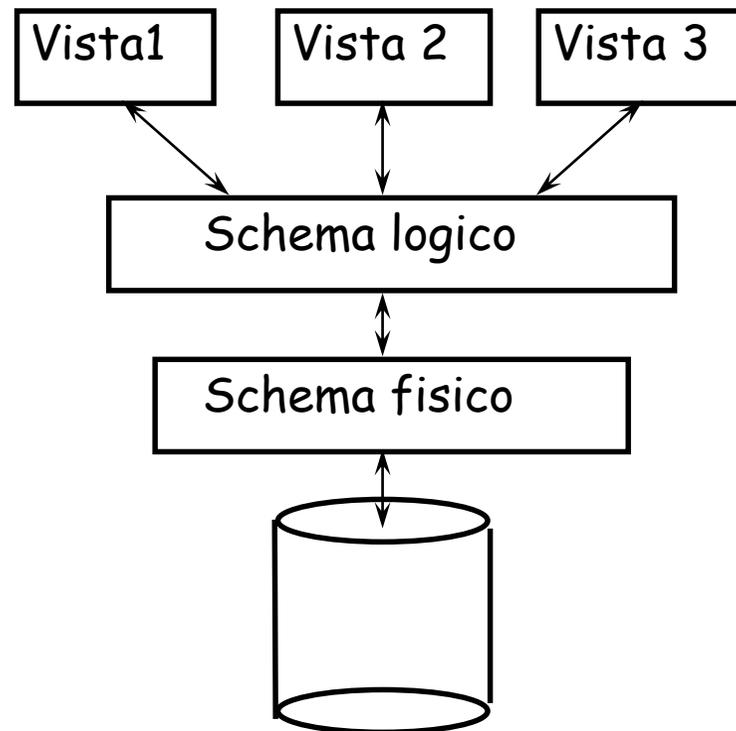
LINGUAGGIO PER LA DEFINIZIONE DELLA BASE DI DATI (DDL)

- È utile distinguere tre diversi livelli di descrizione dei dati (schemi):

- il livello di vista logica,

- il livello logico,

- il livello fisico.



LIVELLO LOGICO

- Livello logico: descrive la struttura degli insiemi di dati e delle relazioni fra loro, secondo un certo modello dei dati, senza nessun riferimento alla loro organizzazione fisica nella memoria permanente (Schema logico).
- Esempio di schema logico:

Studenti(Matricola char(8), Nome char(20), login char(8),
AnnoNascita int, Reddito real)

Corsi(IdeC char(8), Titolo char(20), Credito int)

Esami(Matricola char(8), IdeC char(8), Voto int)

LIVELLO FISICO

- Livello fisico: descrive come vanno organizzati fisicamente i dati nelle memorie permanenti e quali strutture dati ausiliarie prevedere per facilitarne l'uso (Schema fisico o interno).
- Esempio di schema fisico
 - Relazioni Studenti e Esami organizzate in modo seriale, Corsi organizzata sequenziale con indice
 - Indice su Matricola, (Matricola, IdeC)

LIVELLO VISTA LOGICA

- Vista logica: descrive come deve apparire la struttura della base di dati ad una certa applicazione (Schema esterno o vista).
- Esempio di schema esterno
 - InfCorsi(IdeC char(8), Titolo char(20), NumEsami int)

LINGUAGGIO PER LA DEFINIZIONE DELLA BD (cont.)

- L'approccio con tre livelli di descrizione dei dati è stato proposto come un modo per garantire le proprietà di indipendenza logica e fisica dei dati, che sono fra gli obiettivi più importanti dei DBMS.
- **Indipendenza fisica:** i programmi applicativi non devono essere modificati in seguito a modifiche dell'organizzazione fisica dei dati.
- **Indipendenza logica:** i programmi applicativi non devono essere modificati in seguito a modifiche dello schema logico.

FUNZIONALITÀ DEI DBMS: LINGUAGGI PER L'USO DEI DATI

- Un DBMS deve prevedere più modalità d'uso per soddisfare le esigenze delle diverse categorie di utenti che possono accedere alla base di dati (dati e catalogo):
 - Un'interfaccia grafica per accedere ai dati
 - Un linguaggio di interrogazione per gli utenti non programmatori
 - Un linguaggio di programmazione per gli utenti che sviluppano applicazioni:
 - integrazione DDL e DML nel linguaggio ospite: procedure predefinite, estensione del compilatore, precompilazione
 - comunicazione tra linguaggio e DBMS
 - Un linguaggio per lo sviluppo di interfacce per le applicazioni

FUNZIONALITÀ DEI DBMS: MECCANISMI PER IL CONTROLLO DEI DATI

- Una caratteristica molto importante dei DBMS è il tipo di meccanismi offerti per garantire le seguenti proprietà di una base di dati:
 - Integrità, sicurezza e affidabilità.
- Integrità: mantenimento delle proprietà specificate in modo dichiarativo nello schema (vincoli d'integrità)
- Sicurezza: protezione dei dati da usi non autorizzati
- Affidabilità: protezione dei dati da malfunzionamenti hardware o software (fallimenti di transazione, di sistema e disastri) e da interferenze indesiderate dovute all'accesso concorrente ai dati da parte di più utenti.

CONTROLLO DEI DATI: LE TRANSAZIONI

- *Definizione* Una transazione è una sequenza di azioni di lettura e scrittura in memoria permanente e di elaborazioni di dati in memoria temporanea, con le seguenti proprietà:
 - **Atomicità:** Le transazioni che terminano prematuramente (aborted transactions) sono trattate dal sistema come se non fossero mai iniziate; pertanto eventuali loro effetti sulla base di dati sono annullati.
 - **Serializzabilità:** Nel caso di esecuzioni concorrenti di più transazioni, l'effetto complessivo è quello di una esecuzione seriale.
 - **Persistenza:** Le modifiche sulla base di dati di una transazione terminata normalmente sono permanenti, cioè non sono alterabili da eventuali malfunzionamenti.

FUNZIONALITÀ DEI DBMS (cont.)

- Strumenti per l'amministratore della base di dati
 - Un linguaggio per la definizione e la modifica degli schemi logico, interno ed esterno.
 - Strumenti per il controllo e messa a punto del funzionamento del sistema.
 - Strumenti per stabilire i diritti di accesso ai dati e per ripristinare la base di dati in caso di malfunzionamenti.
- Strumenti per lo sviluppo delle applicazioni
 - Produzione di rapporti, grafici, fogli elettronici
 - Sviluppo di menu, forme, componenti grafici

RIEPILOGO DEI VANTAGGI DEI DBMS

- Indipendenza dei dati
- Recupero efficiente dei dati
- Integrità e sicurezza dei dati
- Accessi interattivi, concorrenti e protetti dai malfunzionamenti
- Amministrazione dei dati
- Riduzione dei tempi di sviluppo delle applicazioni
- La riduzione dei costi della tecnologia e i possibili tipi di DBMS disponibili sul mercato facilitano la loro diffusione.

SVANTAGGI DEI DBMS

- Prima di caricare i dati è necessario definire uno schema
- Possono essere costosi e complessi da installare e mantenere in esercizi
- Possono gestire solo dati strutturati e omogenei
- Sono ottimizzati per le applicazioni OLTP, non per quelle OLAP