

**Base di dati:**

- (accezione generica, **metodologica**) insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente (azienda, ufficio, persona) o meglio del sistema informativo;
- (accezione specifica, **metodologica e tecnologica**) insieme di dati gestito da un DBMS;

**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).

**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.

**Funzioni di un sistema informativo:** raccolta, acquisizione delle informazioni; archiviazione, conservazione delle informazioni; elaborazione delle informazioni; distribuzione, scambio di informazioni.

**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 inoltre affidabilità (resistenza a malfunzionamenti hardware e software-recovery) e privacy (con una disciplina e un controllo degli accessi).

**Le basi di dati sono condivise:** una base di dati è una risorsa integrata, condivisa fra le varie applicazioni. Conseguenze:

- Attività diverse su dati in parte condivisi ⇒ meccanismi di autorizzazione;
- Attività multi-utente su dati condivisi ⇒ controllo della concorrenza;

**Possibili problemi:**

- Ridondanza ⇒ informazioni ripetute;
- Rischio di incoerenza ⇒ le versioni possono non coincidere;

**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).

**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;
- I file system prevedono forme rudimentali di condivisione: “tutto o niente”. Nei DBMS, c'è maggiore flessibilità;
- I DBMS estendono le funzionalità dei file system, fornendo più servizi ed in maniera integrata (cfr. efficacia);
- 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;

**Modello dei dati:** insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la

dinamica.

### **Tipi di modelli dei dati:**

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

### **Schemi di un DBMS:**

- **logico:** descrizione della base di dati nel modello logico (ad es. la struttura della tabella);
- **fisico:** rappresentazione dello schema logico per mezzo di strutture memorizzazione (ad es. file);

**Indipendenza dei dati:** il livello logico è indipendente da quello fisico: una tabella è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica (che può anche cambiare nel tempo).

### **Architettura ANSI/SPARC – schema:**

- **logico:** descrizione dell'intera base di dati nel modello logico “principale” del DBMS;
- **fisico:** rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione;
- **esterno:** descrizione di parte della base di dati in un modello logico (“viste” parziali, derivate, anche in modelli diversi);

**Indipendenza dei dati:** il livello logico è indipendente da quello fisico: una tabella è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica (che può anche cambiare nel tempo):

- **indipendenza 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. La realizzazione fisica può cambiare senza che debbano essere modificati i programmi;
- **indipendenza 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;

### **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 scorponabilità delle funzionalità (con riduzione di efficienza);

**Modelli di dati per sistemi di database:** un modello dei dati è un insieme di concetti utilizzati per organizzare i dati e descriverne la struttura in modo che essa risulti comprensibile ad un elaboratore.

### **Modelli logici:**

- **Gerarchico e reticolare:** utilizzano riferimenti espliciti (puntatori) fra record. Vietata la ridondanza (reticolare);
- **Relazionale:** “è basato su valori”. Anche i riferimenti fra dati in strutture (relazioni) diverse sono rappresentati per mezzo dei valori stessi;
- **A oggetti;**

**Il modello relazionale:**

- Favorisce l'indipendenza dei dati;
- Disponibile in DBMS;
- Si basa sul concetto matematico di relazione (con una variante);
- Le relazioni hanno naturale rappresentazione per mezzo di tabelle;

**Relazione:**

- relazione matematica: come nella teoria degli insiemi;
- relazione (dall'inglese relationship) che rappresenta una classe di fatti, nel modello Entity-Relationship; tradotto anche con associazione;
- relazione secondo il modello relazionale dei dati;

**Vantaggi di una struttura basata su valori:**

- indipendenza dalle strutture fisiche (si potrebbe avere anche con puntatori di alto livello) che possono cambiare dinamicamente;
- si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione;
- l'utente finale vede gli stessi dati dei programmatori;
- i dati sono portabili più facilmente da un sistema ad un altro;
- i puntatori sono direzionali;

**Vincolo di integrità:** proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione. E' una funzione booleana (un predicato): associa ad ogni istanza il valore vero o falso.

**Vincolo di ennupla:** esprimono condizioni sui valori di ciascuna ennupla, indipendentemente dalle altre ennuple. Caso particolare: **Vincoli di dominio:** coinvolgono un solo attributo.

**Superchiave e Chiave:** insieme di attributi che identificano le ennuple di una relazione. Formalmente: un insieme  $K$  di attributi è superchiave per  $r$  se  $r$  non contiene due ennuple distinte  $t1$  e  $t2$  con  $t1[K] = t2[K]$ .  $K$  è chiave per  $r$  se è una superchiave minimale per  $r$  (cioè non contiene un'altra superchiave).

**Integrità referenziale:**

- informazioni in relazioni diverse sono correlate attraverso valori comuni;
- in particolare, valori delle chiavi (primarie);
- le correlazioni debbono essere "coerenti";

**Viste:** rappresentazioni diverse per gli stessi dati (schema esterno).

**Viste materializzate:** relazioni derivate memorizzate nella base di dati:

- vantaggi: immediatamente disponibili per le interrogazioni;
- svantaggi: ridondanti, appesantiscono gli aggiornamenti, non sono supportate dai DBMS;

**Viste virtuali:** relazioni virtuali (o viste): sono supportate dai DBMS. Una interrogazione su una vista viene eseguita "ricalcolando" la vista (o quasi).