

# BASE DI DATI

## Concetti di base

**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) elemento di informazione costituito da simboli che devono essere elaborati.

**Sistema informativo:** insieme di strumenti automatici, procedure, risorse umane, norme organizzative orientato alla gestione dell' informazione (raccolta, archiviazione, elaborazione, e scambio) necessaria sia alle attività operative, sia alle attività di programmazione, controllo e valutazione.

**Base di Dati:** collezione di dati strutturati e permanenti, utilizzati per rappresentare le informazioni di interesse per un sistema informativo .

# DBMS

Un **Database Management System (DBMS)** è un sistema software che consente la gestione delle basi di dati.

Le funzioni che un DBMS deve svolgere e che lo caratterizzano sono:

- *Controllo centralizzato dei dati*

Il DBMS deve consentire l'accesso ai dati da parte di più utenti e di più programmi assicurando il controllo centralizzato ed integrato del loro trattamento. I dati non appartengono ad una specifica applicazione, ma le diverse applicazioni vi accedono attraverso il *DBMS*.

Questa caratteristica permette di minimizzare la ridondanza dei dati (duplicazioni dei dati su archivi diversi) e quindi eventuali inconsistenze legate al fatto che uno stesso dato può assumere valori diversi in archivi differenti.

- *Indipendenza fisica e logica*

La gestione dei dati da parte del DBMS avviene a diversi livelli affinché gli utenti possano utilizzare i dati di interesse senza conoscere i dettagli relativi alla loro organizzazione logica e alla allocazione negli archivi fisici.

Per indipendenza fisica, si intende la possibilità che i programmi applicativi non debbano essere modificati in seguito a modifiche dell'organizzazione fisica dei dati.

Per indipendenza logica, si intende la possibilità che i programmi applicativi non debbano essere modificati in seguito a modifiche dello schema logico, per aggiunte di nuove definizioni o per modifiche o eliminazione di definizioni già esistenti.

Per soddisfare questi requisiti, è stato proposto che i DBMS offrano tre livelli distinti di descrizione dei dati: **schema logico, interno ed esterno**. Lo schema logico è la descrizione della struttura della base di dati senza riferimento alla sua organizzazione fisica o al modo in cui vengono memorizzati i dati nelle memorie di massa; lo schema interno è la descrizione di come sono organizzati fisicamente i dati nelle memorie di massa e quali strutture dati ausiliarie sono previste per permetterne l'uso; lo schema esterno è una descrizione di come appare la struttura della base di dati ad una certa applicazione.

Naturalmente tra questi tre livelli di organizzazione dei dati il DBMS deve gestire delle precise corrispondenze per convertire le operazioni sugli oggetti virtuali accessibili da uno schema esterno in quelli sugli oggetti dello schema logico e quindi sui dati realmente presenti nel sistema memorizzati secondo lo schema interno.

- *Controllo della concorrenza*

Il DBMS assicura che l'accesso concorrente ai dati da parti di diversi programmi applicativi non causi errori nel trattamento e nell'utilizzo dei dati stessi: in particolare il sistema agisce da arbitro dei casi in cui diversi programmi richiedono contemporaneamente l'accesso agli stessi dati. Questa caratteristica consente l'utilizzo in tempo reale dei dati memorizzati nella base di dati.

- *Controllo della sicurezza dei dati*

Il DBMS ha il compito di proteggere gli archivi dei dati da malfunzionamenti di vario tipo che possono avvenire nell'ambito del sistema di calcolo. Si parla infatti di sicurezza dei dati per intendere che le informazioni ottenibili da un DBMS non sono influenzate da eventi esterni, quali ad esempio, guasti hardware, comportamenti errati di programmi, etc.

La capacità di un DBMS di assicurare un certo grado di sicurezza richiede che il sistema stesso sia in grado di riconoscere i malfunzionamenti e di ripristinare stati corretti della base di dati.

L'integrità dei dati è affidata a procedure di backup (salvataggio) e recovery (recupero) dei dati, o alla loro duplicazione in mirroring nei casi più critici.

I DBMS consentono di stabilire criteri di protezione delle informazioni gestite e garantiscono la privacy dei dati, affinché solo utenti autorizzati possano accedere ai dati con modalità operative diverse a secondo del loro livello di privilegio.

- *Efficienza ed efficacia.*

Un DBMS deve operare e fornire risposte agli utenti in tempi accettabili, utilizzando una quantità il più possibile limitata di risorse.

L'efficienza di un DBMS dipende essenzialmente dalle tecniche utilizzate per la sua implementazione e dalla buona progettazione della base di dati, per efficacia si intende la capacità di rendere produttive le attività degli utenti.

## Modelli di dati

I DBMS possono essere classificati in base al modello di dati su cui si basano.

*Un modello dei dati è un insieme di meccanismi di strutturazione, o di astrazione, per modellare una base di dati, con certi operatori e vincoli d'integrità predefiniti.*

Un modello di dati è costituito da concetti sulla base dei quali i dati sono strutturati e codificati.

I modelli concettuali cercano di svincolarsi dalla rappresentazione dei dati e rappresentano solo i concetti del mondo reale.

**NB:** *Un esempio di modello concettuale sono i diagrammi Entità Relazione (ER). Un modello ER è composto da entità, che rappresentano gli oggetti di cui si vuole catturare il contenuto informativo, e da relazioni, che rappresentano associazioni fra le entità. Sia le entità che le associazioni possono essere caratterizzate da attributi, che ne rappresentano le proprietà.*

*Usualmente le entità vengono rappresentate mediante rettangoli, le relazioni da rombi collegati con linee alle entità e gli attributi con dei cerchi.*

*Altre caratteristiche che i modelli ER permettono di rappresentare sono le molteplicità (cardinalità), le chiavi primarie e i ruoli.*

I modelli logici, seppure astratti, riflettono la struttura con cui i dati sono organizzati.

I modelli logici più diffusi sono:

- **Relazionale:** Il più diffuso, basato su un modello tabellare dei dati basato sulla teoria degli insiemi;
- **Gerarchico:** Usato nei primi DBMS negli anni 60 (ma tuttora utilizzato), basato su strutture ad albero;
- **Reticolare:** Estensione del modello gerarchico, basato su grafi;
- **A oggetti:** Estensione del modello relazionale basato sui paradigmi di programmazione ad oggetti.

**Modellare i dati** significa costruire una rappresentazione semplificata della realtà osservata o di un problema aziendale, individuandone gli elementi salienti e i legami intercorrenti tra essi.

## Modello concettuale e logico

**Modello concettuale:** è un modello di astrazione che fornisce una rappresentazione simbolica della realtà in esame indipendentemente dagli eventi che in essa accadono, cioè dalle applicazioni, e dal modello logico. Esso descrive concetti del mondo reale. Un esempio è il modello Entità-Relazione (E/R), ideato da Peter P. Chen, che come simboli ha: rombi per le associazioni; rettangoli per le entità; cerchi per gli attributi.

**Associazione:** legame che stabilisce un'interazione tra più entità. L'associazione possiede una caratteristica chiamata **molteplicità**. Essa identifica il numero di possibili istanze di un'entità che viene messo in corrispondenza con un'istanza dell'altra entità che partecipa all'associazione. Il numero minimo di istanze viene chiamato **obbligatorietà**, il numero massimo **cardinalità** dell'associazione. I valori minimi sono 0 o 1. Lo 0 indica la partecipazione **facoltativa** e 1 l'**obbligatorietà** della partecipazione all'associazione. I valori massimi possono essere 1 per indicare una partecipazione e N per indicarne molte. In base alla cardinalità le associazioni si distinguono in: 1 a 1 (1:1 o uno a uno), 1 a N (1:N o uno a molti) e N a N (N:N o molti a molti). Nel modello E/R l'obbligatorietà viene indicata con una linea continua e con una linea tratteggiata la scelta facoltativa.

**Entità:** oggetto concreto o astratto che ha un significato anche quando è considerato da solo.

**Attributo:** proprietà di entità o/e associazioni. Gli attributi hanno determinate caratteristiche: formato (es: data, intero, carattere); dimensione; opzionabilità. I diversi valori assunti dagli attributi vengono chiamati **istanze**.

**Chiave:** per chiave si definisce un insieme minimale di attributi che permettono di distinguere tra loro le istanze di una stessa entità.

**Modello logico:** è un modello di astrazione ottenibile dal modello concettuale che riflette la struttura con cui i dati sono organizzati in un DBMS indipendentemente dalla sua struttura fisica. I modelli logici si dividono in: **Reticolare**, **Gerarchico** e **Relazionale**.

**Modello Relazionale:** è un modello ottenibile dal modello concettuale tramite opportune regole di derivazione, fondato sul concetto matematico di **relazione**.

**Relazione (concetto matematico):** prodotto cartesiano formato dall'insieme ordinato delle coppie  $(x,y)$ , dove  $x$  appartiene ad un insieme  $A_1$  e  $y$  appartiene ad un insieme  $A_2$ .

**Relazione (definizione generale):** una relazione nel modello Relazionale è rappresentata come una tabella definita in questo modo: una relazione su  $n$  insiemi è un sottoinsieme dell'insieme di tutte le tuple che si possono costituire prendendo in ordine un elemento da ogni insieme.

In una relazione si distinguono:

- **Grado della relazione:** il grado di una relazione si identifica col numero delle colonne della relazione stessa;
- **Colonna:** una colonna si identifica col nome di attributo;
- **Dominio:** per dominio si definisce l'insieme di tutti i possibili valori assunti da un attributo;
- **Cardinalità:** per cardinalità si definisce il numero di tuple che compongono la relazione;
- **Tupla:** per tupla si definisce una riga ordinata della relazione;

**Chiave di una relazione:** la chiave di una relazione è un attributo o una combinazione minimale di attributi che identificano univocamente le tuple all'interno della relazione, cioè ogni riga della tabella possiede valori diversi per l'attributo (o gli attributi) chiave.

## Requisiti fondamentali del modello Relazionale

- **Tutte le righe della tabella contengono lo stesso numero di colonne;**
- **Gli attributi rappresentano informazioni elementari (o atomiche), non scomponibili ulteriormente, cioè non ci sono campi che contengono per ogni riga un insieme di valori anziché un solo valore;**
- **I valori assunti da un campo appartengono al dominio dei valori possibili per quel campo, e quindi sono valori omogenei tra loro, cioè dello stesso tipo;**
- **In una relazione, ogni riga è diversa da tutte le altre, cioè non ci possono essere due righe con gli stessi valori dei campi: quindi esistono uno o più attributi con la funzione di chiave primaria che identificano univocamente ogni tupla della relazione;**
- **Le tuple compaiono nella tabella secondo un ordine non prefissato, cioè non è rilevante il criterio con il quale le righe sono sistemate nella tabella.**

Dati questi requisiti la chiave primaria non può avere valore nullo perché questo violerebbe il principio secondo il quale la chiave deve distinguere univocamente ogni tupla della relazione. Perciò si fissa una regola di integrità dei dati detta **integrità sull'entità (entity integrity)**, secondo la quale la chiave primaria non può avere valore nullo (il quale indica informazioni mancanti o sconosciute).

## Regole di derivazione

Utilizzando come modello concettuale il **modello E/R (Entity / Relationship o Entità / Associazioni)** e come modello logico il modello Relazionale, le regole di derivazione sono le seguenti:

- Le entità del modello E/R diventano relazioni (o tabelle) nel modello Relazionale;
- Gli attributi del modello Relazionale ereditano le caratteristiche degli attributi del modello E/R;
- L'identificatore univoco del modello E/R diventa chiave primaria nel modello Relazionale;
- **L'associazione 1:1** (1 a 1) tra due entità del modello E/R diventa un'unica relazione nel modello Relazionale. Questa relazione contiene tutti gli attributi delle due entità coinvolte;
- **L'associazione 1:N** (1 a N o 1 a Molti) tra due entità del modello E/R viene rappresentata aggiungendo all'entità che svolge in ruolo a molti, come chiave esterna (**Foreign Key o FK**), l'identificatore univoco dell'entità che svolge il ruolo a uno e gli eventuali attributi dell'associazione presente tra le entità. Nel modello Relazionale le entità, naturalmente, vengono derivate in relazioni;
- **L'associazione N:N** (N a N o Molti a Molti) tra due entità del modello E/R viene rappresentata aggiungendo una terza relazione (nel modello Relazionale) contenente le chiavi primarie delle altre due (derivate dalle entità) e gli eventuali attributi dell'associazione presente tra le entità;

## **Schemi e istanze**

Nelle basi di dati esiste una parte invariante nel tempo (a meno di modifiche strutturali), detta SCHEMA, costituita dalle caratteristiche dei dati, e da una parte variabile nel tempo, detta istanza, costituita dai valori effettivi.

## Livelli di interazione

Un DBMS è articolato strutturalmente con un architettura a tre livelli, detti *esterno*, *logico* e *interno* per i quali rispettivamente esiste uno schema:

- **Schema esterno:** descrive una porzione della base di dati attraverso il modello logico, riflettendo il punto di vista di un utente.  
Generalmente è realizzato per mezzo di viste, relazioni derivate da quelle che costituiscono lo schema logico.
- **Schema logico:** descrive l'intera base di dati mediante il modello logico adottato dal DBMS (reticolare, gerarchico, relazionale)
- **Schema interno:** rappresenta lo schema logico per mezzo di strutture fisiche di memorizzazione (file sequenziali con o senza indici)

## Indipendenza dei dati

Permette agli utenti di operare a livello astratto, indipendentemente dai dettagli realizzativi usati nel DBMS

**Indipendenza fisica:** Consente di mantenere inalterata la struttura logica dei dati al variare dalla realizzazione fisica del sistema. Consente l'utilizzo di basi di dati su piattaforme diverse, o la distribuzione di una base di dati su più macchine.

**Indipendenza logica:** Rende indipendente lo schema esterno da quello logico, consentendo di inserire nuove viste senza alterarlo, o di alterarlo mantenendo inalterate le viste definite in precedenza

L'architettura a livelli è il meccanismo fondamentale attraverso cui si realizza l'indipendenza dei dati.

## Linguaggi d'interazione al DB

L'interazione con un DBMS avviene mediante un insieme di due linguaggi:

- *DDL* Linguaggi di definizione dei dati utilizzati per definire gli schemi e le autorizzazioni per l'accesso
- *DML* Linguaggi di manipolazione dei dati utilizzati per l'interrogazione e l'aggiornamento dei contenuti della base di dati

Alcuni linguaggi specializzati (es. SQL) presentano le caratteristiche di entrambi i tipi di linguaggio.

## Utenti

Varie categorie di utenti possono interagire con una base di dati ed in particolare:

- Amministratori della base di dati responsabili della progettazione controllo e manutenzione.
- Progettisti e programmatori realizzano i programmi che accedono alla base di dati.  
Il DDL è utilizzato normalmente da amministratori e progettisti che tra i vari compiti di cui sono responsabili hanno quello della creazione dello schema logico.
- *Utenti*
  - utenti esperti interrogano o aggiornano la base di dati con procedure e modalità variabili utilizzando i DML forniti dal DBMS;
  - utenti finali accedono alla base di dati frequentemente e attraverso procedure predefinite;

# Vantaggi e Svantaggi

## Vantaggi dei DBMS

- Disponibilità dei dati a tutta una comunità
- Modello unificato e preciso della realtà di interesse
- Controllo centralizzato dei dati
- Condivisione
- Indipendenza dei dati

## Svantaggi

- Prodotti costosi, complessi che richiedono investimenti in hardware, software, personale.
- Forniscono un numero elevato di servizi, in modo integrato e difficilmente scorporabile se le esigenze dell'utente sono inferiori alle caratteristiche offerte.