

Basi di dati: aspetti introduttivi

1 Sistema informativo

Si dice **sistema informativo** un sistema che ha lo scopo di gestire e rendere disponibili le **informazioni**. Esso è composto da:

- strumenti per la rappresentazione dell'informazione mediante una qualche codifica;
- programmi che, operando su tale rappresentazione, svolgono tutte le funzioni necessarie alla gestione delle informazioni (inserimento, ricerca, ottimizzazione dello storage, ecc.).

2 Dati e informazioni

I **dati** sono i simboli “grezzi” con cui vengono rappresentate le informazioni nei sistemi informatici.

Per ottenere da essi delle informazioni, i dati devono essere *interpretati in un contesto*. Fornire tale contesto interpretativo è uno degli obiettivi fondamentali di un sistema informativo.

3 Base di dati e DBMS

Come *definizione generale*, una **base di dati** è una collezione di dati *tra loro correlati*, utilizzati per rappresentare le informazioni di interesse in un sistema informativo.

Un **sistema di gestione di basi di dati** (**DBMS: Data Base Management System**) è un sistema software che fornisce gli strumenti per la gestione delle informazioni.

Si può quindi dare una *definizione tecnica* di base di dati: essa è una collezione di dati gestita da un DBMS.

4 File system vs. basi di dati

Prima dell'introduzione dei DBMS, le applicazioni gestivano i dati mediante il file system. Questo metodo, però, non funziona bene per grandi quantità di dati (ad esempio, è facile introdurre ridondanze e inconsistenze: uno stesso dato potrebbe essere salvato in più file e poi aggiornato solo in alcuni di essi).

Un DBMS, invece, si interpone tra le applicazioni e il file system, fornendo un accesso centralizzato ai dati, astruendo i dettagli relativi alla gestione dei file, e mettendo a disposizione in maniera integrata ulteriori servizi (es. backup e ripristino).

5 Caratteristiche dei DBMS

- Gestione dei dati in **memoria secondaria** (cioè sul file system), necessaria perché le basi di dati possono avere dimensioni molto maggiori rispetto alla capacità della memoria centrale.
- **Condivisione** dei dati tra applicazioni e utenti diversi, con:
 - la riduzione della ridondanza dei dati;
 - la riduzione delle inconsistenze tra i dati (che sono spesso causate da una combinazione di ridondanza e aggiornamenti parziali);
 - un meccanismo di controllo dell'accesso concorrente, a un livello di granularità molto più fine rispetto a quello dei singoli file (l'unico possibile utilizzando direttamente il file system).
- **Persistenza** dei dati: il loro tempo di vita non è limitato a quello dell'esecuzione dei programmi che li utilizzano.
- **Affidabilità** dei dati in caso di malfunzionamenti hardware/software, grazie alle funzionalità di backup e ripristino.
- **Sicurezza** dei dati, mediante appositi meccanismi di autorizzazione: è possibile stabilire regole di accesso a livello centralizzato, invece di doverle definire separatamente per ogni applicazione che accede ai dati.
- **Efficienza**: lo svolgimento delle varie operazioni utilizza un insieme di risorse (tempo e spazio) accettabile per gli utenti.
- **Efficacia**: capacità di rendere produttive le attività degli utenti.

6 Schema e modello dei dati

Il meccanismo fondamentale di un DBMS è lo **schema (logico)**, che descrive la struttura dei dati contenuti nella base di dati:

- quali sono i dati memorizzati;
- eventuali associazioni tra di essi;
- vincoli di integrità semantica e di autorizzazione.

Tale descrizione si effettua tramite un formalismo ad alto livello, il **modello dei dati**, che astrae la rappresentazione fisica. Esso è costituito da:

- un insieme di strutture dati;
- linguaggi per
 - specificare i dati tramite le strutture previste dal modello;
 - aggiornare tali strutture;
 - specificare vincoli su tali strutture;
 - manipolare/ricercare i dati.

6.1 Modello relazionale

Il modello di dati più diffuso è il **modello relazionale**, che è basato su una singola struttura dati, la **relazione**.

Una relazione può essere vista come una tabella, le cui righe sono chiamate **tuple**, e le cui colonne, ciascuna contenente dati di un certo tipo (interi, stringhe, ecc.), sono chiamate **attributi**.

6.2 Altri modelli dei dati

Prima del modello relazionale, erano utilizzati modelli più vicini alle strutture fisiche di memorizzazione, cioè meno astratti:

- *gerarchico*;
- *reticolare*.

Dopo il modello relazionale, invece, sono stati introdotti:

- *object-oriented, object-relational*;
- *XML*;

- *NoSQL*.

7 Istanza

Lo schema di una base di dati cambia raramente. Invece, l'insieme dei dati contenuti in un dato momento, chiamato **istanza** della base di dati, cambia molto spesso (ogni volta che si effettuano inserimenti, modifiche e/o eliminazioni). Infatti, nello sviluppo una base di dati:

1. si definisce lo schema;
2. si inseriscono i dati, che devono conformarsi alla struttura stabilita da tale schema.

8 Livelli di rappresentazione dei dati

- **Livello logico**: a questo livello è definito lo schema logico. L'astrazione fornita da questo livello facilita l'uso della base di dati da parte degli utenti finali.
- **Livello fisico**: è il livello più basso, in cui viene definito lo **schema fisico** della base di dati, che specifica come i dati nello schema logico vengono effettivamente memorizzati tramite strutture di memorizzazione. A questo livello si ha la possibilità di ottimizzare lo storage e la ricerca dei dati (ma quest'operazione viene solitamente effettuata dall'amministratore della base di dati, quindi l'utente finale non ha bisogno di interagire con il livello fisico).

Questi due livelli sono sempre presenti. Invece, in alcune basi di dati, solitamente quelle con schemi medio-grandi, si ha un'ulteriore astrazione sopra al livello logico: il **livello esterno** o **livello delle viste**. Esso permette di definire delle **viste**, ciascuna delle quali consente di "vedere" solo una porzione dell'intero schema logico. Ciò è utile, ad esempio, per controllare quali dati sono accessibili da varie applicazioni, creando per ciascuna di esse una vista limitata ai dati che deve avere a disposizione.

9 Indipendenza

L'introduzione dei tre livelli di astrazione dei dati (fisico, logico, esterno) assicura le proprietà di:

- **indipendenza fisica**: la rappresentazione fisica dei dati non cambia il modo di accedere alle relazioni, quindi gli utenti e le applicazioni che lavorano sulla rappresentazione logica dei dati sono indipendenti da qualsiasi modifica a livello della rappresentazione fisica;

indipendenza logica: la presenza delle viste permette di nascondere, entro certi limiti, le modifiche allo schema logico (ad esempio, una modifica sulla tabella A non impatta una vista sulla tabella B).

10 Linguaggi di un DBMS

Un DBMS mette a disposizione una gamma di linguaggi per la gestione dei dati:

- **Data Definition Language (DDL):** permette di specificare e modificare lo schema logico, i vincoli d'integrità e le viste;
- **Data Manipulation Language (DML):** usato per operare sull'istanza della base di dati, cioè per inserire, cancellare, modificare, ricercare, ecc. i dati;
- **Storage Definition Language (SDL):** definisce lo schema fisico della base di dati.

Linguaggio	Livelli
DDL	Logico, Esterno
DML	Logico, Esterno
SDL	Fisico