

# Basi di Dati

paolo pangallo

January 2024

## 1 Introduzione

Una base di dati è un insieme organizzato di dati che sono memorizzati e gestiti in modo da poter essere recuperati e aggiornati in modo efficiente. Le basi di dati sono utilizzate per conservare e organizzare grandi quantità di informazioni in modo strutturato, consentendo agli utenti di accedere e manipolare i dati in vari modi.

### Concetti di base: Definizioni di Dato e Informazione

Il **dato** è una rappresentazione simbolica di un fatto, concetto o istruzione. Può assumere diverse forme, tra cui numeri, testo, immagini o suoni, e spesso è organizzato in record o tabelle nei database. Inizialmente, i dati potrebbero mancare di un significato intrinseco e potrebbero richiedere contesto o elaborazione per essere compresi.

D'altra parte, l'**informazione** è il risultato del processo di elaborazione e interpretazione dei dati. È il dato che ha acquisito un significato contestuale ed è stato elaborato per diventare utile. Mentre i dati sono grezzi e neutri, l'informazione è più significativa e può influenzare decisioni o azioni. L'informazione può essere considerata come una rappresentazione di conoscenza derivata dai dati.

Si può dire quindi che il dato rappresenta l'informazione interpreta la rappresentazione

### Progettare una base di dati: modello a cascata

Il modello a cascata è un approccio sequenziale allo sviluppo del sistema, comunemente utilizzato nella progettazione di basi di dati. Le fasi principali di questo modello includono:

## Modello a Cascata nella Progettazione di Basi di Dati

Il modello a cascata è un approccio sequenziale allo sviluppo del sistema, comunemente utilizzato nella progettazione di basi di dati. Le fasi principali di questo modello includono:

1. **Studio di Fattibilità:** Prima di avviare il processo di progettazione, si esegue uno studio di fattibilità per valutare la fattibilità tecnica ed economica del progetto. Questo studio aiuta a determinare se la progettazione e l'implementazione della base di dati sono realizzabili.
2. **Analisi dei Requisiti:** In questa fase, vengono raccolti e analizzati i requisiti del sistema da parte degli utenti e delle parti interessate. Si cerca di comprendere le funzionalità richieste e i dati necessari per supportarle.
3. **Progettazione Concettuale:** Durante questa fase, si crea un modello concettuale dei dati utilizzando strumenti come il diagramma ER. Si identificano entità, relazioni e attributi chiave, fornendo una rappresentazione astratta dei dati.
4. **Progettazione Logica:** Si traduce il modello concettuale in uno schema di base di dati implementabile su un sistema di gestione di database (DBMS). Si definiscono tabelle, colonne e chiavi.
5. **Implementazione:** In questa fase, la base di dati viene implementata fisicamente sulla piattaforma del DBMS. Si creano tabelle, viste, indici e altre strutture necessarie.
6. **Test e Validazione:** Si eseguono test per garantire il corretto funzionamento della base di dati e per verificare la correttezza dei dati e delle operazioni di accesso ai dati.
7. **Manutenzione e Migrazione:** Dopo l'implementazione, la base di dati può richiedere manutenzione continua per adattarsi a cambiamenti nei requisiti. La migrazione dei dati può essere necessaria per spostare dati tra ambienti o adottare nuove versioni del DBMS.

Il modello a cascata fornisce una sequenza ben definita di fasi, ma può essere meno flessibile rispetto ad approcci più agili.

## Schema E-R con Vincoli di Cardinalità

### Entità A

**Attributi:** Attributo1, Attributo2, ...

## Entità B

**Attributi:** AttributoA, AttributoB, ...

## Relazione R tra A e B

- Uno a Molti (1:M)
  - Ogni Entità A è associata a molte Entità B.
  - Ogni Entità B è associata ad al massimo un'Entità A.
- Molti a Uno (M:1)
  - Ogni Entità B è associata ad al massimo un'Entità A.
  - Ogni Entità A è associata a molte Entità B.

article

## Dipendenze Funzionali nei Database Relazionali

Le dipendenze funzionali sono concetti chiave nella teoria dei database relazionali. Esse descrivono la relazione tra le colonne (attributi) di una tabella e sono fondamentali per garantire l'integrità dei dati e la progettazione efficace di schemi di database relazionali. In particolare, una dipendenza funzionale esprime il modo in cui il valore di un attributo determina o è determinato dal valore di un altro attributo in una tabella.

**Definizione:** Una dipendenza funzionale tra due insiemi di attributi in una tabella indica che il valore di uno di essi determina univocamente il valore dell'altro. La dipendenza funzionale si esprime come "X determina Y" e si scrive come  $X \rightarrow Y$ , dove  $X$  e  $Y$  sono insiemi di attributi.

- $X \rightarrow Y$ : Se conosci il valore di  $X$ , puoi determinare univocamente il valore di  $Y$ .

**Esempio:** Supponiamo di avere una tabella "Studenti" con colonne "Matricola", "Nome" e "Cognome". Se diciamo che la dipendenza funzionale "Matricola  $\rightarrow$  Nome, Cognome" è valida, significa che conoscendo la matricola di uno studente, possiamo determinare univocamente il suo nome e cognome.

### Tipi di Dipendenze Funzionali:

- **Dipendenza Funzionale Semplice (Elementare):** Coinvolge una singola colonna determinante e una singola colonna determinata.
- **Dipendenza Funzionale Composta:** Coinvolge più di una colonna nella parte determinante o determinata.
- **Dipendenza Funzionale Totale:** La colonna determinata dipende completamente dalla colonna determinante.

- **Dipendenza Funzionale Parziale:** La colonna determinata dipende solo in parte dalla colonna determinante.

**Notazione:** La notazione  $X \rightarrow Y$  indica la dipendenza funzionale tra gli insiemi di attributi  $X$  e  $Y$ .

**Come Definirle:** Puoi definire le dipendenze funzionali osservando il modo in cui i dati sono strutturati e capendo quali colonne determinano univocamente altre colonne. Ad esempio, analizzando il significato dei dati e le regole di business associate ai dati, puoi identificare le dipendenze funzionali.

Un'altra tecnica comune è utilizzare le regole di normalizzazione del database per individuare dipendenze funzionali e garantire che il database sia organizzato in modo efficiente e senza ridondanze inutili.

## 2 BCNF

La BCNF può essere espressa nel seguente modo: se in una relazione vale la dipendenza funzionale  $A \rightarrow B$ , allora l'insieme di attributi  $A$  deve contenere una chiave (e quindi può svolgere la funzione di chiave).

## 3 Hashing: cos'è?

L'idea di base dell'hashing è quella di permettere di immagazzinare record di dati servendosi di una funzione hash, un "bucket" ovvero un contenitore rappresentato da una locazione di memoria e da tabelle di hash. //TO BE CONTINUED

## 4 Tipi di Hashing

### 4.1 Hashing lineare statico

L'hashing lineare statico è il tipo più semplice di hashing che può essere implementato. Di seguito la definizione di GeeksForGeeks

Hashing in DBMS is a technique to quickly locate a data record in a database irrespective of the size of the database. For larger databases containing thousands and millions of records, the indexing data structure technique becomes very inefficient because searching a specific record through indexing will consume more time. This doesn't align with the goals of DBMS, especially when performance and data retrieval time are minimized. So, to counter this problem hashing technique is used. In this article, we will learn about various hashing techniques.

L'hashing statico torna molto utile quando si conosce a priori la dimensione desiderata dell'hash table e si serve in alcuni casi della proprietà di chaining per migliorare leggermente il problema delle collisioni. Risulta sconsigliato

con tabelle molto grandi. Il chaining risulta molto sconsigliato anche perché richiede spazio per essere implementato e non risolve il problema della frammentazione.

#### 4.1.1 Hashing dinamico (o estensibile)

L'hashing dinamico è preferibile a quello statico nel caso in cui si suppone che il dataset di una database cambi con molta frequenza. In questo caso la dimensione dei buckets cambia in relazione al numero di add/remove che si effettuano sul database.

## 5 Phantom Read

A Phantom read occurs when one user is repeating a read operation on the same records, but has new records in the results set:

**READ UNCOMMITTED** Also called a Dirty read. When this isolation level is used, a transaction can read uncommitted data that later may be rolled back. A transaction that uses this isolation level can only fetch data but can't update, delete, or insert data.

**READ COMMITTED** With this isolation level, Dirty reads are not possible, but if the same row is read repeatedly during the same transaction, its contents may be changed or the entire row may be deleted by other transactions.

**REPEATABLE READ** This isolation level guarantees that a transaction can read the same row many times and it will remain intact. However, if a query with the same search criteria (the same WHERE clause) is executed more than once, each execution may return different set of rows. This may happen because other transactions are allowed to insert new rows that satisfy the search criteria or update some rows in such way that they now satisfy the search criteria.