

Basi di dati

VR443470

ottobre 2022

## Indice

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Introduzione</b>  | <b>3</b> |
| 1.1      | Sistemi informativi, informazioni e dati . . . . .           | 3        |
| 1.2      | Basi di dati e sistemi di gestione di basi di dati . . . . . | 4        |
| 1.3      | Linguaggi per basi di dati . . . . .                         | 5        |
| 1.4      | Modelli dei dati . . . . .                                   | 6        |
| 1.5      | Astrazione (architettura) dei DBMS . . . . .                 | 7        |
| 1.6      | Indipendenza dei dati . . . . .                              | 8        |
| <b>2</b> | <b>Metodologie e modelli per il progetto</b>                 | <b>9</b> |
| 2.1      | Ciclo di vita dei sistemi informativi . . . . .              | 9        |
| 2.2      | Metodologie di progettazione e basi di dati . . . . .        | 11       |

# 1 Introduzione

## 1.1 Sistemi informativi, informazioni e dati

Ogni organizzazione è dotata di un *sistema informativo*, che organizza e gestisce le informazioni necessarie per perseguire gli scopi dell'organizzazione stessa. Per indicare la **porzione automatizzata del sistema informativo** viene di solito utilizzato il termine *sistema informatico*.

Nei sistemi informatici le informazioni vengono rappresentate per mezzo di *dati*, che hanno bisogno di essere interpretati per fornire informazioni. Esiste una differenza sottile tra dato e informazioni. Solitamente i primi, se presi da soli, non hanno significato, ma, una volta interpretati e correlati opportunamente, essi forniscono informazioni, che consentono di arricchire la conoscenza:

**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, sono elementi di informazione costituiti da simboli che devono essere elaborati.

**[ESAME] Definizione base di dati:** Una *base di dati* è una collezione di dati, utilizzati per rappresentare con tecnologia informatica le informazioni di interesse per un sistema informativo.

## 1.2 Basi di dati e sistemi di gestione di basi di dati

Inizialmente, venne adottato un “approccio convenzionale” alla gestione dei dati. Esso **sfruttava** la presenza di archivi o **file per memorizzare e per ricercare dati**. Tuttavia, i metodi di accesso e condivisione erano semplici e banali. Infatti, erano presenti numerosi **problemi**:

- ✗ **Accesso sequenziale**: la scarsa efficienza nell’accesso ai dati su file rendeva lento l’accesso a tali informazioni;
- ✗ **Ridondanza**: i dati di interesse per più programmi sono replicati tante volte quanti sono i programmi che li utilizzano, con evidente ridondanza e possibilità di incoerenza;
- ✗ **Inconsistenza**: una diretta conseguenza della ridondanza. Con la presenza di più copie di un determinato dato, l’eventuale cambiamento di uno solo potrebbe portare a questo effetto;
- ✗ **Progettazione duplicata**: per ogni programma viene replicata la progettazione.

La **soluzione** è arrivata negli anni ’80 con l’avvento delle **basi di dati**. Quest’ultime gestiscono in modo integrato e flessibile le informazioni di interesse per diversi soggetti.

**[ESAME] Definizione DBMS**: Un *sistema di gestione di basi di dati* (in inglese *Data Base Management System*, **DBMS**) è un sistema software in grado di gestire collezioni di dati che siano:

- ✓ **Grandi**;
- ✓ **Condivise**;
- ✓ **Persistenti**.

assicurando allo stesso tempo:

- ★ **Affidabilità**;
- ★ **Privatezza**;
- ★ **Accesso efficiente**.

Il **vantaggio** di utilizzare un DBMS è stato evidenziato nella definizione. Quindi:

- ✓ **Maggiore astrazione** poiché le sue funzioni estendono il *file system*, fornendo la possibilità di accesso condiviso agli stessi dati da parte di più utenti e applicazioni;
- ✓ **Maggiore efficacia** poiché le operazioni di accesso ai dati si basano su un linguaggio di interrogazione.

### 1.3 Linguaggi per basi di dati

Su un DBMS è possibile specificare operazioni di vario tipo, ma principalmente si distinguono in due categorie:

- **Linguaggi di definizione dei dati** (*Data Definition Language*, abbreviato con **DDL**) utilizzati per definire gli schemi logici, esterni e fisici e le autorizzazioni per l'accesso;
- **Linguaggi di manipolazione dei dati** (*Data Manipulation Language*, abbreviato con **DML**) utilizzati per l'interrogazione e l'aggiornamento delle istanze di basi di dati:
  - *Linguaggio di interrogazione*, estrae informazioni da una base di dati (SQL, algebra relazionale);
  - *Linguaggio di manipolazione*, popola la base di dati, modifica il suo contenuto con aggiunte, cancellazioni e variazioni sui dati (SQL).

## 1.4 Modelli dei dati

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

Ogni modello dei dati fornisce **meccanismi di strutturazione**, analoghi ai **costruttori** di tipo dei linguaggi di programmazione (es: Java), che permettono di definire nuovi tipi sulla base di tipi predefiniti (elementari) e costruttori di tipo. Quindi, i *costruttori* consentono di:

- ☞ *Definire* le **strutture dati** che conterranno le **informazioni** della base di dati;
- ☞ *Specificare* le **proprietà** che dovranno soddisfare le **istanze** di informazione che saranno contenuto nelle strutture dati.

**Definizione schemi e istanze:** È molto importante distinguere gli **schemi** e le **istanze** dal concetto di modello dei dati:

- **Schema:** parte invariante nel tempo, è costituita dalle caratteristiche dei dati. In altre parole, è la descrizione della struttura e delle proprietà di una specifica base di dati fatta utilizzando i costrutti del modello dei dati;
- **Istanza o stato:** parte variabile nel tempo, è costituita dai valori effettivi. Quest'ultimi, in un certo istante, popolano le strutture dati della base di dati.

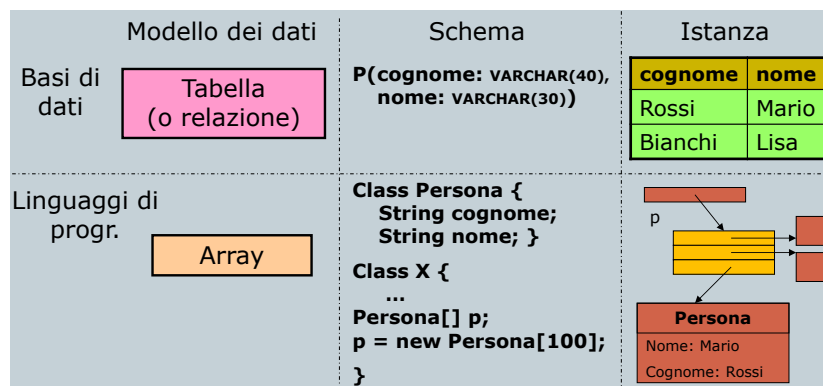


Figura 1: Esempio di modello di dati, schema e istanza.

## 1.5 Astrazione (architettura) dei DBMS

Esiste un'architettura standardizzata per i DBMS, la quale si caratterizza su tre livelli: **esterno**, **logico** e **interno**:

- ☆ *Schema logico*. È la rappresentazione della **struttura** e delle **proprietà** della **base di dati** definita attraverso i costrutti del modello dei dati del DBMS. In altre parole, descrive l'intera base di dati per mezzo del modello logico adottato dal DBMS (quindi relazione o ad oggetti).
- ★ *Schema interno*. È la rappresentazione della base di dati per mezzo delle **strutture fisiche di memorizzazione** (e.g. file sequenziale, file hash, ecc.).
- ☆ *Schema esterno*. Descrive una **porzione dello schema logico** di interesse per uno **specifico** utente o applicazione. Possono esistere più schemi esterni che consentono di avere punti di vista differenti senza cambiare la logica di base.

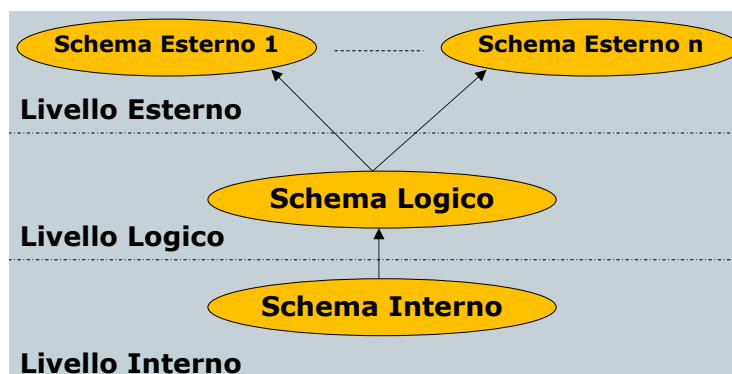


Figura 2: Architettura generale di un DBMS.

## 1.6 Indipendenza dei dati

L'architettura a livelli definita nel paragrafo 1.5 garantisce l' **indipendenza dei dati**, la **proprietà più importante** dei DBMS. L'**obbiettivo** è quello di poter fornire all'utente una base di dati in grado di interagire con un elevato livello di astrazione. Esistono due tipi di indipendenza:

- ☛ **Indipendenza fisica.** Lo schema logico della base di dati è completamente indipendente dallo schema interno. Quindi, l'interazione con il DBMS può essere effettuato in modo indipendente dalla struttura fisica dei dati.

**Vantaggio:** le modifiche **non** influiscono sullo schema logico, cioè sulle applicazioni che lo utilizzano.

- ☛ **Indipendenza logica.** Gli schemi esterni (definizione nel paragrafo: 1.5) della base di dati sono **indipendenti** dallo schema logico. Quindi, è possibile interagire con il livello esterno in modo indipendente dal livello logico.

**Vantaggio:**

- Aggiunta/Modifica** di uno schema esterno in base alle esigenze di un nuovo utente, senza modificare lo schema logico;
- Modifica** di uno schema logico mantenendo inalterate le strutture esterne.



## 2 Metodologie e modelli per il progetto

### 2.1 Ciclo di vita dei sistemi informativi

La progettazione di una base di dati costituisce solo una delle componenti del processo di sviluppo di un sistema informativo complesso e va quindi inquadrata in un contesto più ampio quello del **ciclo di vita** dei sistemi informativi:

- ☛ **Studio di fattibilità.** Definisce i costi delle varie alternative possibili e stabilisce le priorità di realizzazione delle varie componenti del sistema.
- ☛ **Raccolta e analisi dei requisiti.** Individua le proprietà e le funzionalità che il sistema informativo deve avere producendo una descrizione completa, ma generalmente informale.
- ☛ **Progettazione.** Si divide in due fasi:
  - **Progettazione dei dati.** Individua la struttura e l'organizzazione che i dati devono avere.
  - **Progettazione delle applicazioni.** Definizione delle caratteristiche dei programmi applicativi.
- ☛ **Implementazione (su un DBMS).** È la realizzazione del sistema informativo secondo la struttura e le caratteristiche fornite durante la fase di progettazione. In questa fase viene costruita e popola la base di dati.
- ☛ **Validazione e collaudo.** Verifica il corretto funzionamento e la qualità del sistema informativo.
- ☛ **Funzionamento.** Il sistema informativo diventa operativo ed esegue i compiti per i quali è stato progettato.

Spesso il processo **non** è strettamente sequenziale. Infatti, come si vede dalla seguente figura, durante l'esecuzione di una delle attività sopraelencate, è necessario rivedere decisioni prese nell'attività precedente.

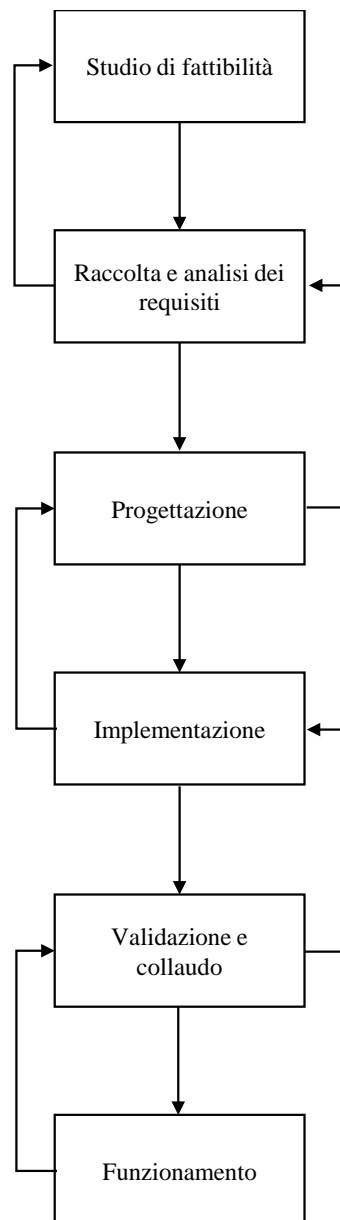


Figura 3: Ciclo di vita di un sistema informativo.

## 2.2 Metodologie di progettazione e basi di dati

Una **metodologia di progettazione** consiste in:

- ✓ **Decomposizione** dell'intera attività di progetto in passi successivi indipendenti tra loro.
- ✓ **Strategie** da seguire nei vari passi e **criteri** nel caso di alternative.
- ✓ **Modelli di riferimento** per descrivere i dati in ingresso e uscita delle varie fasi.

Le **proprietà** che una metodologia deve garantire sono:

- ★ **Generalità** rispetto alle applicazioni e ai sistemi in gioco;
- ★ **Qualità del prodotto** in termini di correttezza, completezza ed efficienza rispetto alle risorse impiegate;
- ★ **Facilità d'uso** delle strategie e dei modelli di riferimento.

Negli anni si è *consolidata una metodologia* di progetto che ha dato prova di soddisfare pienamente le proprietà descritte. Si basa sull'idea di separare le decisioni relative a “cosa” rappresentare in una base di dati (prima fase), da quelle relative a “come” farlo (seconda e terza fase):

### ☛ **Progettazione concettuale.**

**Obiettivo:** rappresentare le specifiche informali della realtà di interesse in termini di una descrizione formale e completa. La **rappresentazione deve essere indipendente** dai criteri di rappresentazione utilizzati nei sistemi di gestione di basi di dati.

**Prodotto di questa fase:** schema concettuale. È un documento formale che rappresenta il contenuto della base di dati in modo indipendente dall'implementazione (DBMS).

**Applicazione:** cercare di rappresentare il **contenuto informativo** della base di dati, senza preoccuparsi né della modalità con le quali queste informazioni verranno codificate in un sistema reale, né dell'efficienza dei programmi che faranno uso di queste informazioni.

### ☛ **Progettazione logica.**

**Obiettivo:** traduzione dello schema concettuale prodotto nella fase precedente, in termini del modello di rappresentazione dei dati adottato dal sistema di gestione di base di dati a disposizione.

**Prodotto di questa fase:** schema logico.

**Applicazione:** durante la traduzione, le scelte progettuali si devono basare anche su criteri di ottimizzazione delle operazioni da effettuare sui dati.

### ☛ **Progettazione fisica.**

**Obiettivo:** lo schema logico viene completato con la specifica dei parametri fisici di memorizzazione dei dati.

**Prodotto di questa fase:** schema fisico.