



Basi di dati

spitfire

A.A. 2023-2024

Contents

1	Introduzione	3
1.1	Sistemi informativi	3
1.2	Sistema informatico	3
1.3	Gestione delle informazioni	4
1.4	Informazioni e dati	4
1.4.1	Perché i dati?	4
1.5	Basi di Dati e introduzione ai Database Management Systems (DBMS)	5
1.6	Database Management System	6
1.7	Caratteristiche delle basi di dati	6
1.8	Archivi e basi di dati	7
1.9	Condivisione delle basi di dati	7
1.10	Caratteristiche dei DBMS	7
1.11	Transazioni	7
1.12	Caratteristiche dei DBMS	8
1.13	Descrizioni dei dati nei DBMS	8
1.13.1	Modello dei dati	8
1.14	Schema e istanza della base di dati	8
1.15	Modello dei dati logico	9
1.15.1	Modello relazionale	9
1.16	Modello dei dati concettuale	9
1.17	Architettura di un DBMS	10
1.18	Indipendenza dei dati	11
1.18.1	Indipendenza fisica	11
1.18.2	Indipendenza logica	11
1.19	Linguaggi per basi di dati	12
1.20	Personaggi e interpreti	12
1.20.1	Database administrator (DBA)	12
1.20.2	Utenti del DBMS	12
1.21	Vantaggi e svantaggi dei DBMS	13
1.22	Caratteristiche dell'approccio con basi di dati	13

1 Introduzione

Una base di dati è un **insieme organizzato di dati** utilizzati per il supporto allo svolgimento di attività (di un ente, azienda, ufficio, persona). Facciamo un passo indietro e ricordiamo cos'è invece **l'informatica**: essa è la **scienza del trattamento razionale dell'informazione**, spesso per mezzo di **macchine automatiche**. Essa è considerata come supporto alla conoscenza umana e alla comunicazione. L'informatica ha due anime:

- **Metodologica**: i metodi per la soluzione di problemi e la gestione delle informazioni
- **Tecnologica**: i calcolatori elettronici e i sistemi che li utilizzano

1.1 Sistemi informativi

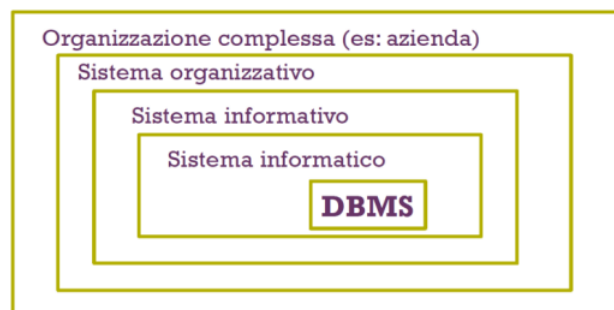
Un **sistema informativo** è un componente (sottosistema) di un'organizzazione che gestisce le informazioni di interesse (cioè utilizzate per il perseguimento degli scopi dell'organizzazione). Le funzioni di un sistema informativo sono:

- **Acquisizione/memorizzazione**
- **Aggiornamento**
- **Interrogazione**
- **Elaborazione**

Il concetto di "sistema informativo" è **indipendente da qualsiasi automazione**: esistono organizzazioni la cui ragion d'essere è la **gestione di informazioni** (es. servizi anagrafici e banche) e che operano da secoli. Anche prima di essere automatizzati, molti sistemi informativi si sono evoluti verso una **razionalizzazione e standardizzazione** delle procedure e dell'organizzazione delle informazioni.

1.2 Sistema informatico

Un **sistema informatico** è una porzione **automatizzata** di un sistema informativo; cioè la parte del sistema informativo che gestisce le informazioni con tecnologia informatica.



Un sistema informatico:

- Garantisce che i dati siano conservati in modo **permanente** sui dispositivi di memorizzazione
- Permette un **rapido aggiornamento** dei dati per riflettere rapidamente le loro variazioni
- Rende i dati **accessibili alle interrogazioni** degli utenti
- Può essere **distribuito sul territorio**

1.3 Gestione delle informazioni

Nelle attività umane, le informazioni vengono gestite (registrate e scambiate) in forme diverse:

- **Idee informali**
- **Linguaggio naturale** (scritto o parlato, formale o colloquiale, in una lingua o in un'altra)
- **Disegni, grafici e schemi**
- **Numeri e codici**

E su vari supporti, come la memoria umana, su carta o su dispositivi elettronici. Nelle attività standardizzate dei sistemi informativi complessi, sono state introdotte col tempo forme di organizzazione e codifica delle informazioni via via più precise (e in un certo senso artificiali). Ad esempio, nei servizi anagrafici si è iniziato con registrazioni discorsive per poi evolversi in registrazioni più complete.

1.4 Informazioni e dati

Nei sistemi informatici (e non solo), le **informazioni** vengono rappresentate in modo essenziale, spartano: attraverso i **dati**. Diamo delle definizioni semantiche:

- **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) elementi di informazione costituiti da simboli che debbono essere elaborati

I dati quindi hanno bisogno di essere interpretati!

1.4.1 Perché i dati?

La rappresentazione precisa di forme più ricche di informazione e conoscenza è difficile. I dati costituiscono spesso una risorsa strategica, perché più stabili nel tempo di altri componenti (processi, tecnologie, ruoli, umani). I dati rimangono gli stessi nella **migrazione** da un sistema al successivo.

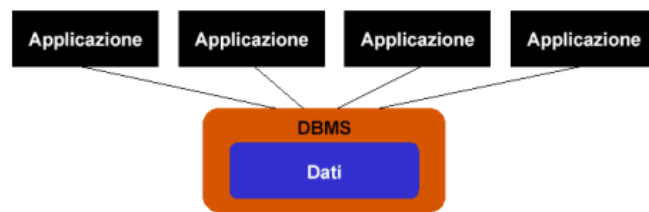
1.5 Basi di Dati e introduzione ai Database Management Systems (DBMS)

Come già detto, una **base di dati** (DB) è una collezione di dati utilizzati per rappresentare le informazioni di interesse in un sistema informativo. Un **Database Management System** (DBMS) è invece un **sistema software** capace di gestire collezioni di dati che siano *grandi, condivise e persistenti*, assicurando la loro *affidabilità e privacy*. Possiamo quindi dare due accezioni al termine **base di dati**:

- **Metodologica**: insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente
- **Specifica, metodologica e tecnologica**: Insieme di dati gestito da un DBMS

Possiamo dare ancora una definizione: una **base di dati** è un insieme di archivi in cui ogni dato è rappresentato logicamente una sola volta e può essere utilizzato da un insieme di applicazioni o da diversi utenti secondo opportuni criteri di riservatezza.

DBMS = Data Base Management System



Le caratteristiche di una base di dati sono:

- I dati sono **molti**
- I dati hanno un **formato definito**
- I dati sono **permanenti**
- I dati sono **raggruppati** per insiemi **omogenei** di dati
- Esistono **relazioni specifiche** tra gli insiemi di dati
- La ridondanza è **minima e controllata**: è assicurata la consistenza delle informazioni
- I dati sono disponibili **per utenze diverse e concorrenti**
- I dati sono **controllati**: protetti da malfunzionamenti hardware e software
- I dati della base di dati sono **indipendenti dai dati di un qualsiasi programma**

1.6 Database Management System

Un DBMS è un insieme di programmi che permettono di creare, usare e gestire una base di dati. Quindi un DBMS è un sistema software general purpose che facilita il processo di **definizione, costruzione e manipolazione** del database per varie applicazioni. Vi sono 3 fasi nella **creazione di un database**:

1. **Definizione**
2. **Creazione/Popolazione**
3. **Manipolazione**

Una volta creato un database, si potranno effettuare su di esso delle **interrogazioni** (queries). L'efficacia delle queries dipende da:

- Conoscenza del contenuto del DB
- Esperienza con il linguaggio di interrogazione
- Semplicità ed efficacia dell'interfaccia di interrogazione

Un DBMS è quindi un **sistema** che gestisce collezioni di dati:

- **Grandi**
- **Persistenti**
- **Condivise**

garantendo:

- **Privatezza**
- **Affidabilità**
- **Efficienza**
- **Efficacia**

1.7 Caratteristiche delle basi di dati

Le basi di dato sono:

- **Grandi**: dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati. Il limite deve essere solo quello fisico dei dispositivi
- **Persistenti**: hanno un tempo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano
- **Condivise**: Ogni organizzazione (specie se grande) è divisa in settori o comunque svolge diverse attività. Ciascun settore/attività ha un (sotto)sistema informativo (non necessariamente disgiunto)
- I possibili problemi sono la **ridondanza**, cioè la ripetizione di informazioni, e l'**incoerenza**, cioè le varie versioni della basi di dati potrebbero non coincidere

1.8 Archivi e basi di dati



1.9 Condivisione delle basi di dati

Una base di dati è una risorsa **integrata e condivisa** fra applicazioni. Diventa quindi possibile svolgere attività diverse su dati condivisi: sono necessari quindi **meccanismi di autorizzazione**. Diventano anche possibili accessi di più utenti ai dati condivisi: sono quindi necessarie politiche e meccanismi di controllo della **concorrenza**.

1.10 Caratteristiche dei DBMS

I DBMS garantiscono **privatezza**; cioè si possono definire meccanismi di autorizzazione: l'utente A è, per esempio, autorizzato a leggere tutti i dati e a modificare quelli sul ricevimento; tuttavia l'utente B è magari autorizzato a leggere i dati di X e a modificare i dati di Y.

I DBMS garantiscono inoltre **l'affidabilità** dei dati, cioè la resistenza a malfunzionamenti hardware e software. Per implementare questa caratteristica è quindi fondamentale **gestire le transazioni**.

1.11 Transazioni

Le **transazioni** sono l'insieme di operazioni da considerare indivisibili ("**atomiche**"), corrette anche in presenza di **concorrenza** e con effetti **definitivi**. La sequenza di operazioni sulla base di dati viene eseguita per intero o per niente: per esempio, il trasferimento di fondi da un conto A ad un conto B avviene o tramite il prelevamento da A e il versamento su B o niente affatto. L'effetto di transazioni concorrenti deve essere **coerente** (ad esempio "equivalente" all'esecuzione separata delle richieste concorrenti): per esempio, se due assegni emessi sullo stesso conto corrente vengono incassati **contemporaneamente**, si deve evitare di trascurarne uno. La conclusione positiva di

una transazione corrisponde ad un impegno (in inglese **commit**) a mantenere traccia del risultato in modo definitivo, anche in presenza di guasti e di esecuzione concorrente.

1.12 Caratteristiche dei DBMS

I DBMS devono essere **efficienti**, cioè devono cercare di utilizzare al meglio le risorse di spazio di memoria (principale e secondaria) e di tempo (di esecuzione e di risposta). I DBMS, con tante funzioni, rischiano l'inefficienza e per questo ci sono grandi investimenti e competizione. L'efficienza è anche il risultato della qualità delle applicazioni. I DBMS devono essere **efficaci**, cioè devono cercare di rendere produttive le attività dei loro utilizzatori, offrendo funzionalità articolare, potenti e flessibili.

1.13 Descrizioni dei dati nei DBMS

Descrizioni e rappresentazione dei dati a livelli diversi permettono l'**indipendenza dei dati** dalla rappresentazione fisica: i programmi fanno riferimento alla struttura a livello più alto, e le rappresentazioni sottostanti possono essere modificate senza necessità di modificare i programmi. Possiamo precisare questo concetto attraverso il concetto di **modello dei dati**

1.13.1 Modello dei dati

Il **modello dei dati** è l'insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la dinamica. Esso ha una componente fondamentale: i **meccanismi di strutturazione** (o **costruttori di tipo**): come nei linguaggi di programmazione esistono meccanismi che permettono di definire nuovi tipi, così ogni modello dei dati prevede alcuni costruttori. Ad esempio, il **modello relazionale** prevede il costruttore **relazione**, che permette di definire insiemi di record omogenei.

1.14 Schema e istanza della base di dati

Insegnamento	Docente	Aula	Ora
Analisi matem. I	Luigi Neri	N1	8:00
Basi di dati	Piero Rossi	N2	9:45
Chimica	Nicola Mori	N1	9:45
Fisica I	Mario Bruni	N1	11:45
Fisica II	Mario Bruni	N3	9:45
Sistemi inform.	Piero Rossi	N3	8:00

In ogni base di dati esistono:

- Lo **schema**, sostanzialmente invariante nel tempo, che ne descriva la struttura e il significato (aspetto intensionale; es. le intestazioni delle tabelle). Lo schema quindi costituisce l'aspetto **intensionale**, ovvero la descrizione "astratta" delle proprietà, ed è invariante nel tempo
- L'**istanza**, i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale; es. il "corpo" di ciascuna tabella). L'istanza quindi costituisce

invece l'aspetto **estensionale** "concreto", che varia nel tempo al variare della situazione di ciò che stiamo descrivendo

1.15 Modello dei dati logico

I modelli dei dati logici sono adottati nei DBMS esistenti per l'organizzazione dei dati:

- utilizzati dai programmi
- indipendenti dalle strutture fisiche

Esempi di essi sono il modello **relazionale**, reticolare, gerarchico, a oggetti, XML, ...

1.15.1 Modello relazionale

Nel modello relazionale i dati vengono strutturati in tabelle, in particolare un DBMS relazionale può essere pensato come un insieme di tabelle. Ogni tabella mantiene informazioni di tipo omogeneo. Diverse tabelle sono collegate (in relazione) fra loro grazie alla presenza di un **campo comune**, che permette di mettere in relazione i dati delle due tabelle. Nel modello relazionale:

- **Lo schema** è la descrizione della struttura delle tabelle (stabile nel tempo)
- **L'istanza** sono i valori dei campi della tabella

Schema Studenti				
Matricola	Matricola	Cognome	Nome	Data di nascita
Cognome	Record 6554	Pinco	Pallino	05/12/1978
Nome	8765	Neri	Paolo	03/11/1976
Data di nascita	9283	Verdi	Luisa	12/11/1979
	3456	Rossi	Maria	01/02/1978

1.16 Modello dei dati concettuale

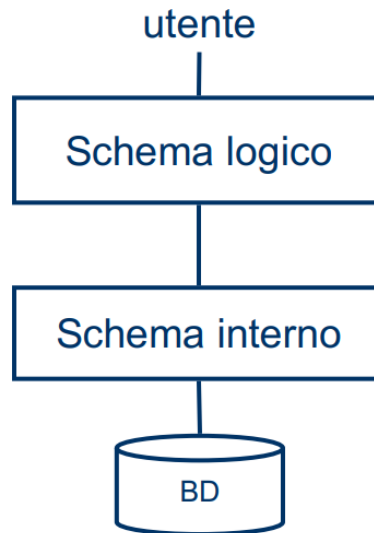
I modelli dei dati concettuali permettono di rappresentare i dati in modo indipendente da ogni sistema.

- cercando di descrivere i concetti del mondo reale
- sono utilizzati nelle fasi preliminare di progettazione (quali informazioni sono utili? Come sono collegate?)

Il più diffuso è il modello **Entity-Relationship**

1.17 Architettura di un DBMS

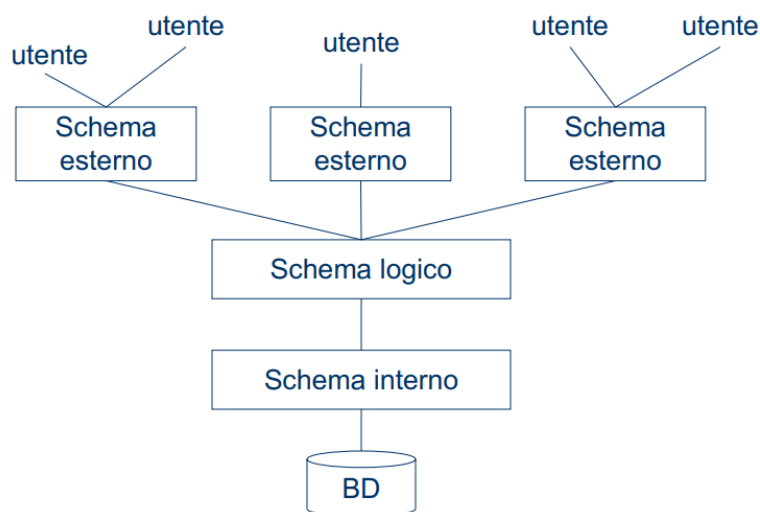
Possiamo strutturare l'architettura di un DBMS in maniera **semplificata** nel seguente modo:



Nella quale:

- **Schema logico**: descrizione della base di dati nel modello logico (ad esempio, la struttura della tabella)
- **Schema interno (o fisico)**: rappresentazione dello schema logico per mezzo di strutture di memorizzazione (file ad esempio, ma anche record con puntatori, ordinati in un certo modo)

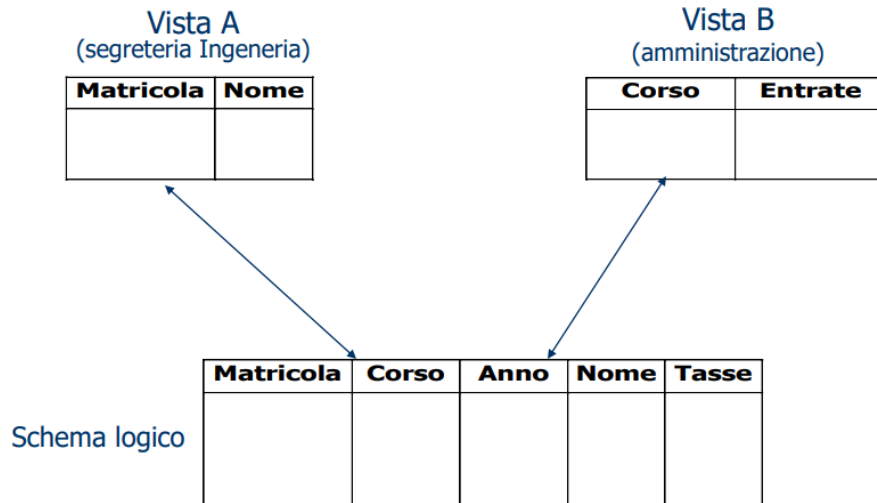
Una delle architetture standard di riferimento per i DBMS è **L'architettura ANSI/SPARC a tre livelli**:



Nella quale:

- **Schema logico**; descrizione dell'intera base di dati nel modello logico "principale" del DBMS

- **Schema fisico:** rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione
- **Schema esterno:** descrizione di parte della base di dati in un modello logico ("viste" parziali, derivate, anche in modelli diversi)



1.18 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). In questo corso vedremo solo il livello logico e non quello fisico. La conseguenza della articolazioni in livelli è che l'accesso avviene solo tramite il **livello esterno** (che può coincidere con il livello logico). Vi sono due forme di indipendenza:

- **Indipendenza fisica**
- **Indipendenza logica**

1.18.1 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 che usano la base di dati.

1.18.2 Indipendenza logica

Il livello esterno è indipendente da quello logico. Aggiunte o modifiche alle viste non richiedono modifiche al livello logico. Le modifiche allo schema logico che lasciano inalterato lo schema esterno sono quindi **trasparenti**.

1.19 Linguaggi per basi di dati

Vi sono due tipi di linguaggi per le basi di dati:

- **Data Definition Languages**(DDL): Servono a definire degli schemi logici, fisici e delle autorizzazioni di accesso
- **Data Manipulation Languages**(DML): Servono per effettuare l'interrogazione e l'aggiornamento delle basi di dati

Alcuni linguaggi come SQL (Structured Query Language) hanno funzioni di entrambe le categorie.

1.20 Personaggi e interpreti

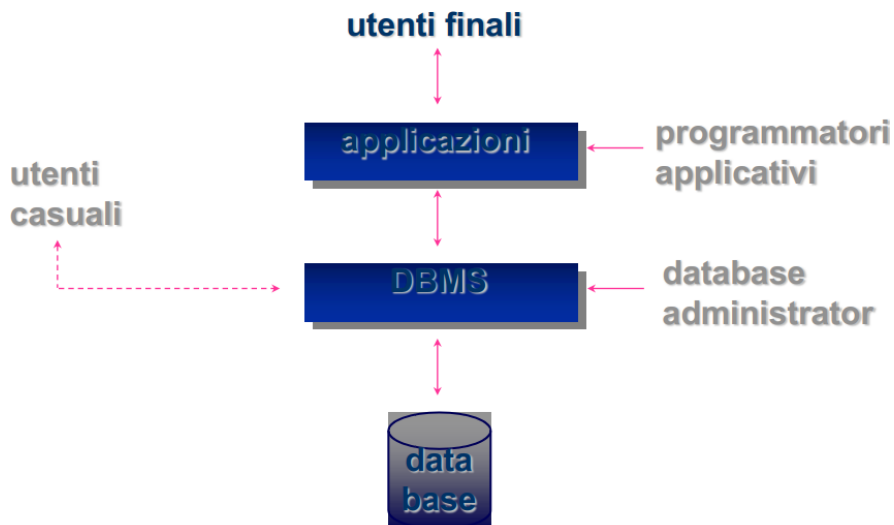
Si possono identificare diverse figure durante l'intero ciclo di vita di una base di dati:

- **Progettisti** e realizzatori di **DBMS**
- **Progettisti della base di dati** e amministratori della base di dati (**DBA**)
- **Progettisti** e programatori di **Applicazioni**
- **Utenti**
 - Utenti **finali**(terminalisti): eseguono applicazioni predefinite(**transazioni**)
 - Utenti **casuali**: eseguono operazioni non previste a priori, usando linguaggi interattivi

1.20.1 Database administrator (DBA)

Un **database administrator**(DBA) è una persona o gruppo di persone responsabile del controllo centralizzato e della gestione del sistema, delle prestazioni, dell'affidabilità e delle autorizzazioni. Le funzioni del DBA includono quelle di progettazione, anche se in progetti complessi ci possono essere distinzioni.

1.20.2 Utenti del DBMS



1.21 Vantaggi e svantaggi dei DBMS

Vediamo quali sono i **vantaggi** di avere un DBMS:

- Permettono di considerare i dati come risorsa comune di un'organizzazione, a disposizione di molteplici applicazioni e utenti
- Offrono un modello della parte di mondo di interesse che è unificato e preciso, utilizzabile in applicazioni attuali e future
- Offrono un controllo centralizzato dei dati, riducendo ridondanze e inconsistenze
- Permettono l'indipendenza dei dati: favoriscono lo sviluppo di applicazioni flessibili e facilmente modificabili

Vediamo quali sono invece gli **svantaggi** di avere un DBMS:

- Un DBMS è costoso, complesso e ha specifici requisiti in termini di software e hardware
- Diventa difficile separare, tra tutti i servizi offerti da un DBMS, quelli effettivamente usati da quelli inutili
- Sono inadatti alla gestione di applicazioni con pochi utenti (NB: Dipende dai costi operativi)

1.22 Caratteristiche dell'approccio con basi di dati

L'approccio alla **rappresentazione dei dati tramite basi di dati** ha le seguenti caratteristiche:

- **Astrazione dei dati:** si usa un **modello dati** per nascondere dettagli e presentare all'utente una *vista concettuale* del database
- **Supporto di viste multiple dei dati:** Ogni utente può usare una vista (view) differente del database, contenente solo i dati di interesse per quell'utente