

Introduzione alle Banche Dati





Dati e informazioni

- La *raccolta*, l'*archiviazione* e la *manipolazione* di *dati* sono operazioni ricorrenti in molte attività (e.g., conti bancari, elenchi telefonici, elenchi degli iscritti ad un corso di laurea ecc.).
- Tali attività possono prescindere dall'uso di un computer; tuttavia questi ultimi garantiscono una memorizzazione ed un trattamento dei dati stabili ed efficienti.
- Un *dato* in sé non costituisce un'informazione in quanto consiste semplicemente di un insieme di simboli; ad esempio la sequenza di caratteri *Mario Rossi* e le cifre *06 658976* non hanno un significato intrinseco.
- Quando un dato viene interpretato come risultato di un'interrogazione (e.g., "chi è il direttore della banca e qual è il suo numero telefonico?") diventa *informazione*.



Basi di dati (database)

- Ogni organizzazione necessita di un *sistema informativo* per il trattamento delle informazioni inerenti alla sua attività.
- Solitamente solo una piccola parte del sistema informativo di un'organizzazione sfrutta la tecnologia legata ai computer.
- I dati (rappresentanti le informazioni) sono *stabili*, i.e., difficilmente cambiano la propria struttura nel corso del tempo.
- Le procedure del sistema informativo che operano sui dati invece sono facilmente soggette a modifiche.
- Un *database* (nella sua accezione più generale) è un insieme di dati che *codificano* dell'informazione utile in un sistema informativo.
- In base a quanto detto a proposito della stabilità dei dati, un database rappresenta una *risorsa* fondamentale da sfruttare e proteggere.

DataBase Management System (DBMS)



- Prima della fine degli anni 60 non esistevano software specifici per la gestione dei dati, che venivano memorizzati in file e manipolati per mezzo di linguaggi di programmazione tradizionali.
- L'organizzazione dei dati per mezzo dei file comportava tuttavia dei problemi di condivisione quando più utenti volevano lavorare sugli stessi dati.
- Inoltre l'autonomia delle singole procedure operanti sui file, comportava la duplicazione di questi ultimi con la conseguente problematica di garantire la consistenza dei dati (e.g., date due copie di uno stesso file, quale è da ritenersi valida?).
- Per ovviare a tali problemi sono nati i DBMS, i.e., software apposti in grado di gestire grandi quantità di dati.

Caratteristiche dei DataBase e dei DBMS

- Un database può contenere una *grande quantità di dati*; per questo utilizza principalmente la *memoria secondaria* di un computer.
- I database sono condivisi, i.e., garantiscono l'accesso a più applicazioni ed utenti *contemporaneamente*.
- La memorizzazione dei dati in un database è permanente, i.e., i database sono *persistenti*.
- I DBMS sono *affidabili*, i.e., assicurano che i dati non andranno persi oppure che potranno essere recuperati, in caso di problemi hardware o software, tramite meccanismi di *backup e/o data recovery*.
- Tramite un meccanismo di autenticazione i DBMS assicurano la *privacy* dei dati, i.e., ogni utente è in grado di accedere soltanto ai dati che gli competono.
- I DBMS sono *efficienti*, i.e., svolgono il loro compito utilizzando al meglio le risorse del sistema.
- I DBMS permettono di incrementare la produttività degli utenti che li utilizzano.



Modelli dei dati

- Un modello di dati fornisce un insieme di costrutti per l'organizzazione dei dati.
- Esistono vari modelli dei dati:
 - Il modello gerarchico (hierarchical data model)
 - Il modello a grafo (network data model)
 - **Il modello relazionale (relational data model)**
 - Il modello a oggetti (object data model)
- I modelli dei dati sono chiamati modelli logici in quanto i loro costrutti, pur essendo astratti, riflettono una particolare struttura (albero, grafo, relazione ecc.).

Il modello relazionale

- Il modello relazionale prevede un costrutto di *relazione* che permette di organizzare i dati in insiemi di *record* (tuple) a struttura *fissa*; la rappresentazione più usata di tale costrutto è quella di tabella:
 - le righe corrispondono ai record;
 - le colonne corrispondono agli attributi o campi dei record.
- Esempio:

Insegnamenti

Corso	Docente
Algebra	Mario Rossi
Geometria	Paolo Bianchi
Fisica	Guido Verdi

Programma

Laurea	Materia	Anno
Matematica	Fisica	I
Informatica	Algebra	I
Ingegneria civile	Fisica	II
Matematica	Geometria	I

Schemi e istanze di un Database

- La parte di un database che è *invariante* rispetto al tempo è detta *schema* e descrive le caratteristiche fondamentali dei dati.
- La parte di un database che varia nel tempo è detta *istanza* o *stato* ed è costituita dai dati veri e propri.
- Nell'esempio del lucido precedente (caso del modello relazionale), gli schemi delle due relazioni si indicano come segue:

Insegnamenti (Corso, Docente)

Programma (Laurea, Materia, Anno)

Le istanze delle due relazioni sono costituite dall'insieme delle righe delle rispettive tabelle.

- Lo schema di un database è anche detto *componente intensionale*, mentre l'istanza è detta anche *componente estensionale*.



Livelli di astrazione nei DBMS

- L'architettura di un DBMS può essere suddivisa in tre strati:
 - Schema *logico* (relazionale, gerarchico, a grafo, a oggetti).
 - Schema *interno*: descrizione dell'implementazione del database in termini delle strutture fisiche di memorizzazione.
 - Schema *esterno*: descrizione per mezzo dello schema logico di una parte del database, in modo da riflettere il punto di vista di un particolare utente o gruppo di utenti (per questo si chiama anche *view*). Ad esempio ad uno studente di matematica interesseranno soltanto le righe della tabella Programma in cui l'attributo Laurea assume il valore "Matematica".
- Note:
 - Per ogni database, possono esistere diverse view (schemi esterni).
 - Un meccanismo di autorizzazioni può essere usato per regolare l'accesso alle view appropriate per ogni categoria di utenti.
 - L'interazione con i dati avviene solamente per mezzo dello schema esterno (che può coincidere con quello logico).



Indipendenza dei dati

- I vari livelli di astrazione dei DBMS permettono a programmi ed utenti di interagire con i dati senza preoccuparsi dei dettagli dell'implementazione.
- Vi sono due livelli di indipendenza:
 - L'indipendenza fisica permette di modificare le strutture (e.g., file su disco) che implementano le relazioni senza influenzare la descrizione ad alto livello dei dati ed il loro utilizzo da parte di programmi ed utenti.
 - L'indipendenza logica permette di modificare le view (schemi esterni) senza influenzare lo schema logico e, di conseguenza, quello interno.
- L'indipendenza dei dati è raggiunta per mezzo dell'interazione tramite lo schema esterno, in quanto il DBMS si preoccupa di tradurre le varie operazioni in termini dei livelli sottostanti.



Linguaggi orientati ai database

- In corrispondenza alla distinzione fra schemi ed istanze di database esistono due tipi di linguaggi:
 - Data Definition Language (DDL): linguaggi che consentono di definire gli schemi logici, interni ed esterni di un database con i relativi meccanismi di autorizzazione degli accessi.
 - Data Manipulation Language (DML): linguaggi che consentono di interrogare e modificare le istanze dei database.
- Esistono linguaggi che offrono le funzionalità proprie sia dei DDL che dei DML, e.g., SQL (Structured Query Language).
- L'accesso ai dati può avvenire tramite:
 - Linguaggi interattivi testuali (e.g., SQL).
 - Comandi appositi richiamabili dai tradizionali linguaggi di programmazione (C, C++, Java ecc.)
 - Comandi di linguaggi ad hoc realizzati per svolgere compiti specifici (e.g. stampa di resoconti, generazione di grafi ecc.).
 - Strumenti interattivi visuali.



Categorie di utenti

- DataBase Administrator (DBA): è il responsabile della progettazione, controllo ed amministrazione del database. Deve garantire i servizi associati al database, l'affidabilità e la gestione delle autorizzazioni d'accesso.
- Progettisti di applicazioni e programmatori: utilizzano DML e strumenti complementari per la generazione di interfacce al database.
- Utenti: utilizzano il database per i loro scopi; si suddividono in:
 - End users: sfruttano programmi speciali che implementano operazioni ripetitive e predefinite.
 - Casual users: usano i linguaggi interattivi per l'esecuzione di interrogazioni ed operazioni particolari ("casual" indica che tali interazioni non sono predefinite).



Vantaggi/svantaggi dei DBMS

- Vantaggi:

- I DBMS consentono l'accesso dei soli utenti autorizzati alla risorsa centralizzata dei dati di un'organizzazione.
- Il modello dei dati rappresenta uno standard che codifica le informazioni rilevanti per l'organizzazione, rendendone agevole l'integrazione in applicazioni esistenti e future.
- Il controllo centralizzato dei dati riduce i costi, consentendo un'economia di scala.
- La condivisione dei dati elimina il problema della ridondanza e della consistenza di questi ultimi.
- I DBMS, tramite l'indipendenza dei dati, favoriscono lo sviluppo di applicazioni flessibili e facilmente modificabili.

- Svantaggi:

- I DBMS sono prodotti costosi e complessi la cui introduzione in una realtà esistente comporta notevoli investimenti in termini di hardware, software ed addestramento di personale.
- Spesso i DBMS forniscono un numero di servizi maggiore di quelli richiesti; la difficoltà di isolare soltanto quelli necessari comporta molte volte costi e perdita di efficienza dell'attività.

