

Appunti Basi di Dati (Teoria)

Luca Seggiani

6 Marzo 2024

1 Introduzione

Che cos'è l'informatica?

- L'informatica è la scienza del trattamento razionale, spesso attraverso macchine automatiche, dell'informazione.

possiamo fare una distinzione fra:

- metodologica
- tecnologica

2 Sistema informativo

- Il sistema organizzativo è costituito da risorse e regole per lo svolgimento coordinato di attività di una certa organizzazione (azienda, ente, ecc...) Noi ci concentreremo sul
- Sistema informativo, ovvero la parte del sistema organizzativo che acquisisce, conserva, elabora e produce informazioni d'interesse per l'organizzazione. Inoltre esegue e gestisce i processi informativi (che coinvolgono informazioni).

possiamo analizzare più nel dettaglio i tipo di attività svolte dal sistema informativo:

- Raccolta e acquisizione di informazioni
- Archiviazione e conservazione di suddette informazioni

- Elaborazione, trasformazione e ancora produzione di nuove informazioni sulla base di quelle ottenute
- Distribuzione e comunicazione delle informazioni così elaborate.

Occorre fare attenzione: il sistema informativo non è di per se in alcun modo legato all'informatica (esistono sistemi informativi, si pensi ai servizi anagrafici o alle banche, che non usano alcuna automatizzazione).

La parte del sistema informativo che usa la tecnologia informatica è il sistema informativo automatizzato, o semplicemente **sistema informatico**.

Ricapitolando, possiamo stabilire la seguente relazione:

Azienda → Sistema organizzativo → Sistema informativo → sistema informatico

3 Gestione delle informazioni

L'informazione può essere gestita secondo modalità diverse, e su supporti diversi. Ad esempio, abbiamo:

- Idee informali
- Linguaggio naturale
- Disegni, schemi, grafici
- Numeri e codici

su vari supporti:

- Mente umana
- Carta
- Dispositivi elettronici (e.g. hard disk, ecc...)

Definiamo ora la differenza fra informazioni e dati:

- **Informazione**

Notizia, dato o elemento che consente di avere una conoscenza dei fatti, situazioni o modi di essere.

- **Dato**

Ciò che è immediatamente presente alla conoscenza prima dell'elaborazione. In informatica, elementi di informazione costituiti da *simboli* che devono essere elaborati.

I dati sono spesso codifiche particolari di informazioni, che vanno quindi da essi estrapolate. Ad esempio, il codice fiscale, i cartelli stradali, ecc...

4 La base di dati

Il cuore di un sistema informativo automatizzato è la base di dati (database), cioè un insieme organizzato di dati rappresentanti informazioni di interesse. Nelle due accezioni (metodologica e tecnologica), possiamo dire:

- Metodologica: insieme organizzato di dati utilizzati come supporto per lo svolgimento di attività
- Tecnologica: insieme di dati gestito da un DBMS (Database Management System)

Le basi di dati hanno solitamente:

- Dimensioni molto maggiori della memoria centrale dei sistemi di calcolo utilizzati
- Tempo di vita indipendente dalle singole istanze dei programmi che li utilizzano (persistenza dei dati)
- Supporto per gestione di collezioni di dati condivise fra più dispositivi
- Capacità di garantire privacy, affidabilità, efficienza ed efficacia

Vediamo nel dettaglio dell'aspetto di condivisione:

- Ogni organizzazione è divisa in settori o almeno svolge disparate attività
- Ciascun settore potrebbe essere fornito di un sottosistema informativo, magari disgiunto a quello principale.

Questo può chiaramente portare a problemi di:

- **Ridondanza:** ripetizione dell'informazione
- **Incoerenza:** più versioni dell'informazione che non coincidono.

Nota: perchè non usare un semplice archivio di file invece che di un database?

Un archivio non fornisce alcuna gestione dell'interdipendenza fra informazioni, ed è quindi poco portato agli eventuali controlli sulla coerenza e la correttezza dell'informazione. Inoltre, in un comune filesystem abbiamo a disposizione solamente le operazioni rudimentali di scrittura/lettura, senza particolari controlli su concorrenza o funzionalità particolari.

Per ovviare a tutta questa serie di problemi, introduciamo:

- **Autorizzazione:** gestione dell'accesso a date risorse
- **Concorrenza:** gestione dell'accesso *simultaneo* (!) a date risorse
- **Affidabilità:** resistenza a malfunzionamenti hardware e software. Una tecnica fondamentale in questo campo è la corretta gestione delle **transazioni**
- **Efficienza:** gestione ottimale delle risorse in termini di memoria e tempo
- **Efficacia:** offerta di funzionalità articolate, potenti e flessibili.

Vediamo un'attimo nel dettaglio l'aspetto delle transazioni:

Transazioni

Una transazione è un insieme di operazioni da considerare indivisibili ("*atomiche*"), corretto anche in presenza di concorrenza, e con effetti definitivi a fine esecuzione. Una transazione deve essere eseguita *per intero* o *per niente*. La corretta gestione della concorrenza deve invece assicurarsi che transazioni concorrenti vengano gestite correttamente (o in serie, e. g. transazioni bancarie, o in maniera mutualmente esclusiva, e. g. prenotazione biglietto aereo). La transazione dovrà poi essere permanente, ovvero la conclusione positiva di una transazione corrisponde ad un impegno (commit) a mantenere traccia della versione ormai aggiornata dei dati a seguito della stessa.

Descrizione dei dati

Nei programmi tradizionali che accedono a file, ogni programma contiene una descrizione della struttura del file stesso, con i conseguenti rischi di incoerenza fra informazioni e file stessa. Ecco perchè nei DBMS è opportuno dedicare una porzione della base di dati alla descrizione centralizzata dei dati. Introduciamo il concetto di **modello dei dati**: un insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne le varie dinamiche, fornendoci quindi una vista astratta dei suddetti. Occorre quindi definire la differenza fra:

- **Schema:** descrizione della struttura della dei dati, solitamente invariante nel tempo (attributi)
- **Istanza:** i valori attuali immagazzinati nella base di dati, che variano nel tempo (record).

introduciamo l'elemento fondamentale della base di dati: la **tabella**. Una comune tabella di una base di dati contiene come intestazioni delle sue colonne gli attributi (schema) dei dati immagazzinati, mentre le successive righe presentano vari record (istanze) dei dati definiti dalle colonne.