

Basi di Dati

UniShare

Davide Cozzi
@dlcgold

Gabriele De Rosa
@derogab

Federica Di Lauro
@f_dila

Indice

1	Introduzione	2
2	Lezione 1	3

Capitolo 1

Introduzione

Questi appunti sono presi a ldurante le esercitazioni in laboratorio. Per quanto sia stata fatta una revisione è altamente probabile (praticamente certo) che possano contenere errori, sia di stampa che di vero e proprio contenuto. Per eventuali proposte di correzione effettuare una pull request. Link: <https://github.com/dlcgold/Appunti>.

Grazie mille e buono studio!

Capitolo 2

Lezione 1

Il corso si divide in 7 parti:

1. introduzione generale
2. metodologie e modelli per il progetto delle basi di dati
3. progettazione concettuale
4. modello razionale
5. progettazione logica
6. linguaggio SQL
7. algebra relazionale

Le **informazioni** fanno parte delle risorse di un'azienda. Una **base di dati** è un insieme organizzato di dati utilizzato per il supporto allo svolgimento di attività. L'**informatica** è la scienza del trattamento relazionale, specialmente per mezzo di macchine automatiche, dell'informazione, considerata come supporto alla conoscenza umana e all'informazione. Un **sistema informativo** è un componente di un'organizzazione che gestisce le informazioni d'interesse. Si divide in:

- acquisizione e memorizzazione
- aggiornamento
- interrogazione
- elaborazione

Un sistema informativo è una porzione automatizzata del sistema informatico. Un sistema informatico gestisce il sistema informativo in modo automatizzato, ne garantisce la memorizzazione, l'aggiornamento dei dati per riflettere le loro variazioni, l'accessibilità dei dati. Un sistema informatico può essere distribuito sul territorio. Le informazioni vengono gestite in vari modi, mediante idee formali, con il linguaggio naturale, graficamente con schemi e con numeri o codici. Anche il mezzo può variare dalla carta ai dispositivi elettronici. Pian piano si è arrivato ad avere una codifica standard per l'informazione, a fini organizzativi. Le informazioni nei sistemi informatici sono rappresentate in modo essenziale mediante i **dati**, questa è la *sintassi*. La *semantica* invece si ottiene con le intestazioni delle tabelle. I dati sono una risorsa strategica perché sono stabili nel tempo. Solitamente i dati sono immutati durante una migrazione tra un sistema e un altro. Un **Data Base** è una collezione di dati usati per rappresentare un'informazione di interesse di un sistema informativo. Un **DBMS** è un software che gestisce un database. Un base di dati è anche un insieme di archivi in cui ogni dato è rappresentato logicamente una e una sola volta e può essere usato da un insieme di applicazioni e da più utenti su vari criteri di riservatezza. Si hanno le seguenti caratteristiche:

- 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 minima è controllata ed è assicurata la consistenza delle informazioni
- i dati sono disponibili per utenze diverse e concorrenti
- i dati sono controllati e protetti da malfunzionamenti hardware e software
- i dati sono indipendenti dal programma

Si hanno tre fasi di creazione di una tabella:

1. definizione
2. creazione e popolazione

3. manipolazione

Si garantiscono:

- privatezza
- affidabilità
- efficienza
- efficacia

Un'organizzazione è divisa in vari settori e ogni settore ha un suo sottosistema informativo (non necessariamente disgiunto). Le basi dati solitamente sono condivise. Una base di dati gestisce i meccanismi di autorizzazione e ha un controllo della concorrenza, oltre ad essere una risorsa integrata e condivisa. Una base di dati deve essere conservata a lungo termine e si ha una gestione delle **transazioni**. Una **transazione** è un insieme di operazioni da considerare indivisibile, "atomico", corretto anche in presenza di concorrenza e con effetti definitivi. La sequenza di operazioni sulla base di dati deve essere eseguita nella sua interezza (sono quindi atomiche) e l'effetto di transazioni concorrenti deve essere coerente. La conclusione positiva di una transazione corrisponde ad un impegno, *commit*, a mantenere traccia del risultato in modo definitivo, anche in presenza di guasti e di esecuzione concorrente.

I DBMS devono essere efficienti, utilizzando al meglio memoria e tempo. Devono anche essere efficaci e produttivi.

Si hanno delle caratteristiche nell'approccio alla base dati:

- natura autodescrittiva di un sistema di basi di dati: il sistema di basi di dati memorizza i dati e anche una descrizione completa della sua struttura (catalogo) queste informazioni sono chiamate metadati. Questo consente al DBMS di lavorare con qualsiasi applicazione
- separazione tra programmi e dati: chiamata indipendenza tra programmi e dati. E' possibile cambiare la struttura dati senza cambiare i programmi
- 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

- condivisione dei dati e gestione delle transazioni con utenti multipli. Permette a più utenti concorrenti di accedere contemporaneamente al database. Il controllo della concorrenza garantisce che più utenti impegnati ad aggiornare gli stessi dati lo facciano in maniera controllata: il risultato degli aggiornamenti è corretto

I DBMS estendono le funzionalità dei file system, fornendo più servizi ed in maniera integrata.

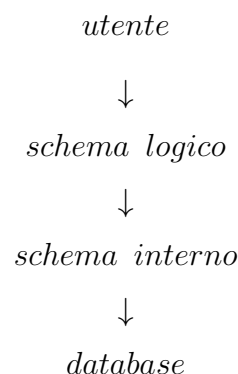
In ogni base di dati si ha:

- lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura, il significato (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 "concreto", che varia nel tempo al variare della situazione di ciò che stiamo descrivendo)

Si hanno due tipi di modelli:

- **modelli logici**, adottati nei DBMS esistenti per l'organizzazione dei dati, utilizzati dai programmi e indipendenti dalle strutture fisiche
- **modelli concettuali** che permettono di rappresentare i dati in modo indipendente da ogni sistema, che cercano di descrivere i concetti del mondo reali e sono usati nelle fasi preliminari di progettazione. Il più diffuso è il modello **Entity-Relationship**

ecco l'architettura di un DBMS:



Si hanno 2 schemi:

1. **schema logico**: descrizione della base di dati nel modello logico (ad esempio, la struttura della tabella)

2. **schema interno (o fisico):** rappresentazione dello schema logico per mezzo di strutture memorizzazione (file; ad esempio, record con puntatori, ordinati in un certo modo)

Il livello logico è indipendente da quello fisico infatti una tabella è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica (che può anche cambiare nel tempo). **Noi tratteremo solo il livello logico.** Ecco i tre schemi dell'architettura ANSI/SPARC:

1. **schema logico:** descrizione dell'intera base di dati nel modello logico "principale" del DBMS
2. **schema fisico:** rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione
3. **schema esterno:** descrizione di parte della base di dati in un modello logico ("viste" parziali, derivate, anche in modelli diversi)

L'accesso ai dati avviene solo mediante il livello esterno (che a volte coincide con quello logico) e si hanno 2 forme di indipendenza:

1. **indipendenza fisica**
2. **indipendenza logica**

Il livello logico e quello esterno sono indipendenti da quello fisico.

Un linguaggio interattivo per gestire una base dati è SQL, che funziona tramite **query**. Si hanno linguaggi per la definizione di dati, i DDL, *Data Definition Languages*, e linguaggi di manipolazione dei dati, DML, *Data Manipulation Languages*. I primi definiscono e i secondi interrogano e aggiornano. SQL può svolgere le operazioni di entrambi.