
Fondamenti di Teoria delle Basi di Dati

Riccardo Torlone

Parte 1:
Introduzione

Obiettivi

- La conoscenza della teoria delle basi di dati?
 - No (o non solo)
- Piuttosto:
 - Come si può affrontare un problema in maniera rigorosa (formalizzare un problema)
 - Con quali strumenti (matematica, algebra, logica)
- Perché le basi di dati?
 - Mi dispiace, non avevo tempo per altro...
 - Offre diversi problemi reali
 - Si presta a studi metodologici con ricadute tecnologiche

Contenuti (provvisori)

- Introduzione e terminologia
- Il modello relazionale
- Linguaggi per basi di dati
- Teoria delle dipendenze
- Teoria delle preferenze

Metodo di studio e verifica

- Studio individuale su libri o articoli consigliati, con riflessione sui concetti
- Svolgimento di esercizi
- Progetto (individuale):
 - Studio di un argomento proposto
 - Presentazione finale dei risultati

Informazioni e dati

- Nei sistemi informatici (e non solo), le **informazioni** vengono rappresentate in modo essenziale, spartano: attraverso i **dati**

Informazioni e dati

(dal Vocabolario della lingua italiana 1987)

- **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

Dati e informazioni

- Un esempio:

Mario 275

su un foglio di carta sono due dati e non significano molto

- Se il foglio di carta viene fornito in risposta alla domanda “A chi mi devo rivolgere per il problema X; qual è il suo numero di telefono?”, allora i dati possono essere interpretati per fornire informazione e arricchire la conoscenza

Basi di dati

Vi possono analizzare da due punti di vista:

- Metodologico
- Tecnologico

Base di dati

(accezione generica, **metodologica**)

- insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente (azienda, ufficio, persona)

(accezione specifica, prevalentemente **tecnologica**)

- insieme di dati gestito da un DBMS

Sistema di gestione di basi di dati (DBMS)

- Sistema che gestisce collezioni di dati:
 - grandi
 - persistenti
 - condivise
- Garantendo
 - privatezza
 - affidabilità
 - efficienza
 - efficacia

Modello dei dati

- Insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la dinamica
- Componente fondamentale: 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
- Esempio: il modello relazionale prevede il costruttore relazione, che permette di definire insiemi di record omogenei

Una base di dati

Orario

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

Basi di dati: schema e istanza

Lo **schema** della base di dati

Orario

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

L'**istanza** della base di dati

Schema e istanza

- In ogni base di dati esistono:
 - lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura (aspetto intensionale)
 - es.: le intestazioni delle tabelle
 - l'**istanza**, i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale)
 - es.: il “corpo” di ciascuna tabella

Due tipi (principali) di modelli

- modelli logici
- modelli concettuali

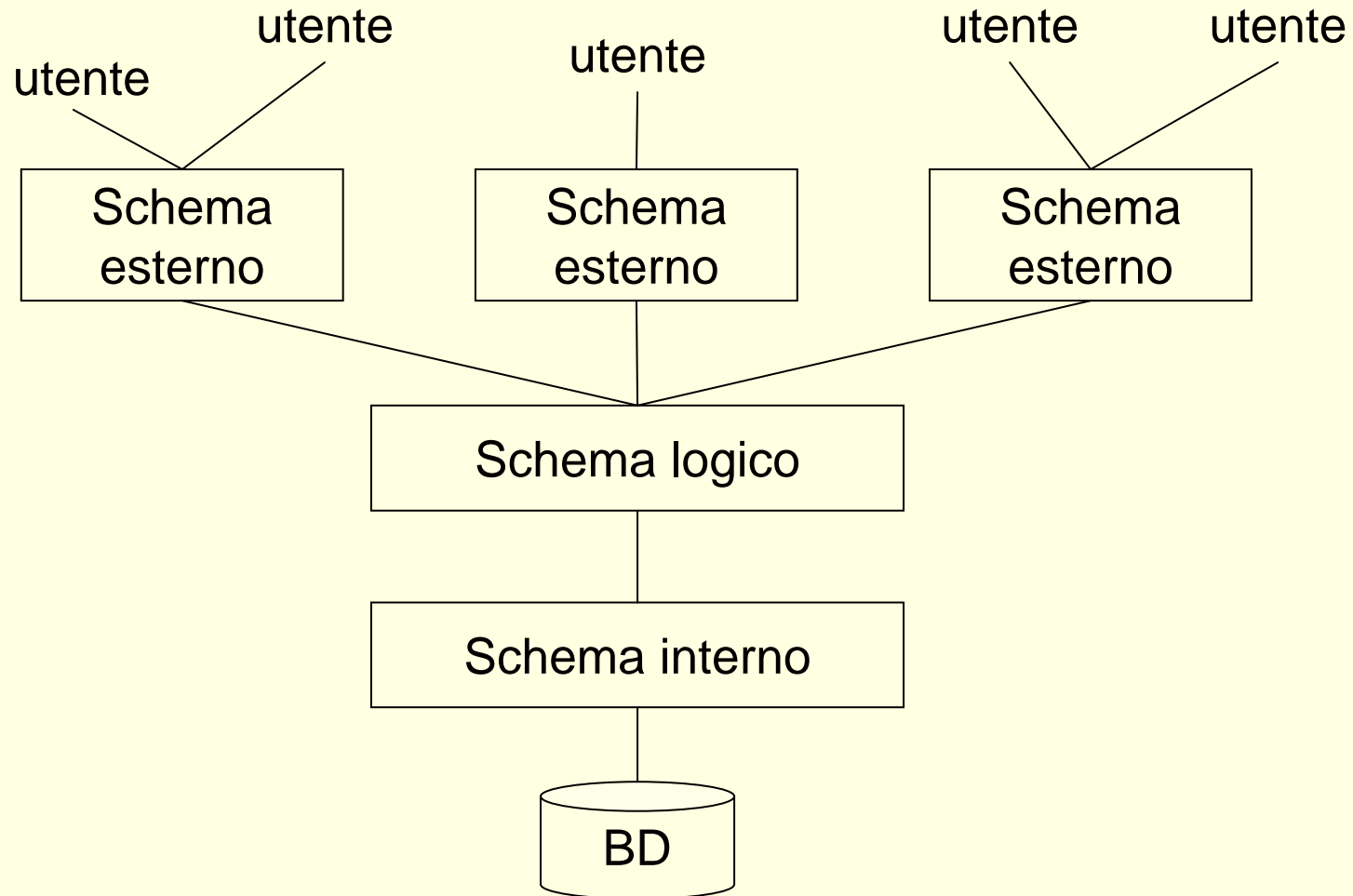
Modelli logici

- Adottati nei DBMS esistenti per l'organizzazione dei dati
 - utilizzati dai programmi
 - indipendenti dalle strutture fisiche
- esempi: **relazionale**, reticolare, gerarchico, a oggetti

Modelli concettuali

- Permettono di rappresentare i dati in modo indipendente da ogni sistema
 - cercano di descrivere i concetti del mondo reale
 - sono utilizzati nelle fasi preliminari di progettazione
- Il più diffuso è il modello **Entità-Relazione**

Architettura standard (ANSI) di DBMS



Architettura ANSI/SPARC: schemi

- Schema **logico**: descrizione dell'intera base di dati nel modello logico “principale” del DBMS
- Schema interno (o **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)

Una vista

Corsi

Corso	Docente	Aula
Basi di dati	Rossi	DS3
Sistemi	Neri	N3
Reti	Bruni	N3
Controlli	Bruni	G

Aule

Nome	Edificio	Piano
DS1	OMI	Terra
N3	OMI	Terra
G	Pincherle	Primo

CorsiSedi

Corso	Aula	Edificio	Piano
Sistemi	N3	OMI	Terra
Reti	N3	OMI	Terra
Controlli	G	Pincherle	Primo

Indipendenza dei dati

- conseguenza della articolazione in livelli
- l'accesso avviene solo tramite il livello esterno (che può coincidere con il livello logico)
- due forme:
 - **indipendenza fisica**
 - **indipendenza logica**

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

Indipendenza logica

- il livello esterno è indipendente da quello logico
- aggiunte o modifiche alle viste non richiedono modifiche al livello logico
- modifiche allo schema logico che lascino inalterato lo schema esterno sono trasparenti

Linguaggi per basi di dati

- Un contributo all'efficacia: disponibilità di vari linguaggi e interfacce
 - linguaggi testuali interattivi (SQL)
 - comandi (SQL) immersi in un linguaggio ospite (Pascal, Java, C ...)
 - comandi (SQL) immersi in un linguaggio ad hoc, con anche altre funzionalità (p.es. per grafici o stampe strutturate)
 - con interfacce amichevoli (senza linguaggio testuale)

Una distinzione terminologica

data manipulation language (DML)

per l'interrogazione e l'aggiornamento di (istanze di) basi di dati

data definition language (DDL)

per la definizione di schemi (logici, esterni, fisici) e altre operazioni generali

SQL, un linguaggio interattivo

- "Trovare i corsi tenuti in aule a piano terra"

Corsi

Corso	Docente	Aula
Basi di dati	Rossi	DS3
Sistemi	Neri	N3
Reti	Bruni	N3
Controlli	Bruni	G

Aule

Nome	Edificio	Piano
DS1	OMI	Terra
N3	OMI	Terra
G	Pincherle	Primo

SQL, un linguaggio interattivo

```
SELECT Corso, Aula, Piano
FROM Aule, Corsi
WHERE Nome = Aula
AND Piano = 'Terra'
```

Corso	Aula	Piano
Sistemi	N3	Terra
Reti	N3	Terra

SQL immerso in linguaggio ospite

```
void VisualizzaStipendiDipart(char NomeDip[ ]) {
    exec sql begin declare section;
        char Nome[20], Cognome[20];
        long int Stipendio;
    exec sql end declare section;
    exec sql declare ImpDip cursor for
        select Nome, Cognome, Stipendio
        from Impiegato
        where Dipart = :NomeDip;
    exec sql open ImpDip;
    exec sql fetch ImpDip into :Nome, :Cognome, :Stipendio;
    printf("Dipartimento %s\n",NomeDip);
    while (sqlca.sqlcode == 0) {
        printf("Nome e cognome impiegato: %s %s", Nome,Cognome);
        printf("Attuale stipendio: %d\n",Stipendio);
        exec sql fetch ImpDip into :Nome, :Cognome, :Stipendio;
    }
    exec sql close cursor ImpDip;
}
```

Linguaggio ad hoc (Oracle PL/SQL)

```
declare Stip number;
begin
    select Stipendio into Stip from Impiegato
    where Matricola = '575488' for update of Stipendio;
    if Stip > 30 then
        update Impiegato set Stipendio = Stipendio * 1.1
        where Matricola = '575488';
    else
        update Impiegato set Stipendio = Stipendio * 1.15
        where Matricola = '575488';
    end if;
    commit;
exception
    when no_data_found then
        insert into Errori values('Matricola inesistente',sysdate);
end;
```

Interazione non testuale (Access)

Microsoft Access - [Q02 : Query di selezione]

Digitare una domanda.

Impiegato

- *
Matricola
Cognome
Stipendio
Dipartimento
Foto

Dipartimento

- *
Codice
Nome
Sede
Direttore

Campo:	Cognome	Stipendio	Sede			
Tabella:	Impiegato	Impiegato	Dipartimento			
Ordinamento:						
Mostra:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Criteri:			'Milano'			
Oppure:						

Pronto NUM

Vincoli di integrità

- Esistono istanze di basi di dati che, pur sintatticamente corrette, non rappresentano informazioni possibili per l'applicazione di interesse

Una base di dati "scorretta"

Esami	Studente	Voto	Lode	Corso
	276545	32		01
	276545	30	e lode	02
	787643	27	e lode	03
	739430	24		04

Studenti	Matricola	Cognome	Nome
	276545	Rossi	Mario
	787643	Neri	Piero
	787643	Bianchi	Luca

Vincolo di integrità

- Proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione
- Un vincolo è una funzione booleana (un **predicato**)
 - associa ad ogni istanza il valore **vero** o **falso**

Vincoli di integrità, perché?

- descrizione più accurata della realtà
- contributo alla “qualità dei dati”
- utili nella progettazione (vedremo)
- usati dai DBMS nella esecuzione delle interrogazioni