

# Base di Dati 1

Riccardo Cara

November 18, 2023

## Contents

<b>1</b>	<b>Introduzione ai database e ai DBMS</b>	<b>2</b>
<b>2</b>	<b>Sistema Informativo</b>	<b>2</b>
2.1	Livelli di astrazione . . . . .	2
2.2	Linguaggi per database . . . . .	3
2.3	Modello di dati . . . . .	3
<b>3</b>	<b>Modello relazionale</b>	<b>4</b>
3.1	vincoli di integrità . . . . .	5
3.2	chiavi . . . . .	6
<b>4</b>	<b>Algebra Relazionale</b>	<b>6</b>
<b>5</b>	<b>Teoria Relazionale</b>	<b>6</b>
<b>6</b>	<b>Organizzazione Fisica</b>	<b>6</b>
<b>7</b>	<b>Concorrenza</b>	<b>6</b>

# 1 Introduzione ai database e ai DBMS

Un *database* o Base di Dati è un insieme di dati organizzati in *strutture dati*, esse ne facilitano la creazione, l'accesso e l'aggiornamento. I dati possono essere:

- **dati strutturati:** oggetti rappresentati da piccole stringhe di simboli o numeri
- **dati non strutturati:** testo scritto in linguaggio umano

I dati all'interno di un Database vengono gestiti tramite un Database Management System o DBMS ossia il software per la gestione di grandi masse di dati condivise, strutturate e processabili, residenti su una memoria.

**Informazione strutturata** la struttura dell'informazione dipende dal suo utilizzo e può essere modificata nel tempo, facendo un esempio, se si volessero archiviare i dati di una persona, avremmo:

- nome
- cognome
- data di nascita
- luogo di nascita

L'obiettivo è di facilitare l'elaborazione dei dati sulla base delle loro relazioni che vengono rappresentate tramite record:

nome	cognome	data di nascita	luogo di nascita
Francesco	Totti	27/09/1976	Roma

per accedere ai dati individuali nella struttura dati, vengono fatte delle "interrogazioni" o *query*. Un database è una risorsa condivisa tra diverse componenti in un'organizzazione, quando più componenti vogliono accedere allo stesso dato contemporaneamente si crea un problema di concorrenza che comporta la necessità di gestire gli accessi contemporanei agli stessi dati.

## 2 Sistema Informativo

Un *sistema informativo* è concettualmente organizzato in aggregati di informazioni omogenee (file <sup>1</sup>), gli aggregati sono le componenti del sistema informativo. Un'operazione di aggiornamento ha per oggetto un singolo aggregato, mentre una query può coinvolgerne molteplici.

### 2.1 Livelli di astrazione

Un database ha tre livelli di astrazione:

- **schema esterno:** descrizione di una porzione del database tramite un modello logico che può differire dallo schema logico e rispecchia i bisogni e privilegi di accesso di diversi tipi di utente, l'accesso al database avviene solamente tramite lo schema esterno
- **schema logico:** descrizione dell'intero database nel "principale" modello logico del DBMS
- **schema fisico:** rappresentazione dello schema logico sulle strutture di archiviazione fisica

quando si parla di schema e istanze, con schema si intende in una tabella, la prima riga che definisce il tipo di dato 

nome	cognome	data di nascita
------	---------	-----------------

 con istanza si intende invece il "corpo" della tabella ovvero i valori correnti.

---

<sup>1</sup>esistono alcuni sottofile chiamati indici che permettono di recuperare velocemente le informazioni dai file principali

## 2.2 Linguaggi per database

Esistono diversi linguaggi utilizzabili per creare/gestire database, DDL,DML,SQL.

- **DDL**(Data Definition Language): per la definizioni di schemio logici, esterni, fisici e altre operazioni generali
- **DML**(Data Manipulation Language): per effettuare query e aggiornare le istanze
- **SQL**(Structured Query Language): è un linguaggio standardizzato per database basati su modelli relazionali che permette di fare entrambe le azioni del DDL e del DML

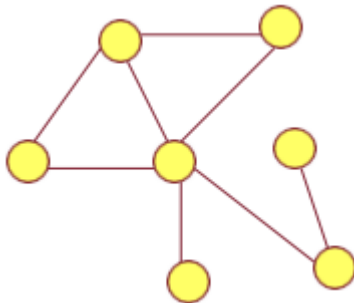
## 2.3 Modello di dati

Un modello è una struttura da utilizzare per organizzare i dati di interesse e le loro relazioni, un modello è composto da costruttori di tipo, ad esempio il modello relazionale prevede il costruttore di relazione (organizza in record i dati con una relazione). I modelli dei dati si dividono in due gruppi:

- **modelli logici**: i modelli logici non dipendono da come vengono archiviati i dati fisicamente, ma disponibili secondo il modello tramite i DBMS. alcuni modelli logici sono: reticolare, gerarchico, relazionale, a oggetti.
- **modelli concettuali**: i modelli concettuali sono utilizzati all'inizio della progettazione di un database, essi non sono dipendenti dalle modalità di realizzazione, hanno lo scopo di rappresentare entità del mondo reale e le loro relazioni nelle prime fasi di pianificazione di un database, uno dei più utilizzati è il modello entità-relazione.

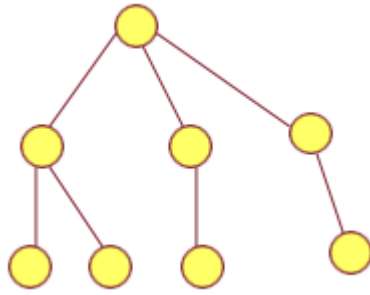
vediamo tipi diversi di modelli dati:

- **Modello Mesh**: Nel modello Mesh, i dati sono rappresentati in collezioni di record omogenei, le relazioni binarie vengono rappresentate come links, implementati come puntatori e quindi dipendenti dalla struttura fisica del database. È possibile rappresentare il modello graficamente:



Il modello è rappresentato da nodi e linee. I nodi rappresentano i record e le linee che connettono i nodi rappresentano i links tra essi. Il modello più popolare di modello mesh è *CODASYL*

- **Modello gerarchico**: il modello gerarchico è un tipo di modello mesh in cui i collegamenti fra i nodi vengono fatti in maniera gerarchica, i nodi hanno quindi forma di albero e hanno come



padre un solo nodo

- **modello relazionale:** i dati e le relazioni sono rappresentate da valori, non ci sono riferimenti espliciti come i puntatori nei modelli mesh o gerarchici. Gli oggetti del mondo reale sono rappresentati come record in cui i campi sono riempiti da informazioni di interesse (nome, cognome, data di nascita).
- **modello a oggetti:** è un modello basato su oggetti (persona, macchina, animale), gli attributi descrivono lo stato dell'oggetto

### 3 Modello relazionale

È già stato definito il Modello relazionale nella precedente sezione, si vuole ora approfondire l'argomento. Il modello relazionale si basa sulla nozione matematica di relazione. Si definisce con *dominio* ( $D$ ) un insieme di valori, un sottoinsieme del prodotto cartesiano di  $k$  domini  $D_1 \times D_2 \times \dots \times D_k$  ha nome di *relazione matematica*, introduciamo due definizioni.

- **grado:** con grado  $k$  si intende la relazione matematica di  $k$  domini.
- **attributo:** un attributo è definito da un nome  $A$  e dal suo dominio  $\text{dom}(A)$

nome
Nerone

in cui  $A = \text{nome}$ ,  $\text{dom}(A) = \text{str}$  e l'attributo è l'insieme di questi.

- **schema:** Un insieme di attributi è detto schema, lo schema relazionale viene scritto come  $R(A_1, A_2, \dots, A_k)$ , descrive la sua struttura ed è invariante nel tempo.
- **istanza:** l'istanza  $r$  contiene tutti i valori correnti in una relazione  $R$ .
- **n-upla:** una n-upla è un elemento della relazione (un elemento è la riga di una tabella), il numero di n-uple di una relazione è chiamata cardinalità. In ogni n-upla della relazione di ordine  $k$ , i componenti della n-upla sono ordinati secondo  $k$ , dati  $D_1, D_2, \dots, D_k$ , gli attributi saranno:

$\text{attributo}_1$	$\text{attributo}_2$	...	$\text{attributo}_k$
----------------------	----------------------	-----	----------------------

una n-upla da un punto di vista matematico, è una funzione definita in  $R(\text{relazione})$  che associa ogni attributo  $A$  in  $R$  ad un elemento del dominio  $\text{dom}(a)$ , il valore dell'attributo  $A$  equivale a  $t[A]$  in cui  $t$  è tupla/n-upla.

- **tabella:** una tabella è l'insieme delle n-uple di una relazione.
- **schema del database:** lo schema del database è un insieme di schemi relazionali. Lo schema di un database relazionale è un insieme  $R_1, R_2, \dots, R_n$  degli schemi. Un database relazionale è l'insieme  $r_1, r_2, \dots, r_n$  in cui  $r_1$  è l'istanza di  $R_1$ .

**esempio** schema: info città (città, regione, popolazione)

istanza:	Roma	Lazio	3000000
	Milano	Lombardia	1500000
	Genova	Liguria	800000
	Pisa	Toscana	150000

**esempio** : definendo  $k = 2$  e  $D_1 = \{\text{bianco}, \text{nero}\}$ ,  $D_2 = \{0, 1, 2\}$ , il prodotto cartesiano sarà  $D_1 \times D_2 = \{(\text{bianco}, 0), (\text{bianco}, 1), (\text{bianco}, 2), (\text{nero}, 0), (\text{nero}, 1), (\text{nero}, 2)\}$ , è possibile ora definire una relazione, un esempio di relazione è  $\{(\text{bianco}, 0), (\text{bianco}, 2), (\text{nero}, 1)\}$  ovvero una relazione di grado 2 con cardinalità 3. Una tupla ovvero  $t \in r$  è invece  $(\text{bianco}, 0)$ ,  $t[1]$  rappresenta la stringa "bianco"

#### Nota bene

non è detto che ogni attributo abbia un valore, ad esempio, in un database degli studenti potrebbe essere previsto il numero di telefono, ma non è detto che ogni studente abbia un cellulare, per questo viene utilizzato il valore Null, il valore Null non corrisponde ad uno 0, non appartiene ad un dominio, se sullo stesso dominio ci sono più Null, non sono lo stesso valore.

### 3.1 vincoli di integrità

In un database, ad esempio:

matricola	nome	anno iscrizione	dipartimento
1028	Carlo magno	1680	A02
1029	Lucio Dalla	2003	B01
1029	Dario Lampa	2015	B08

dipartimento	numero dipartimento
storia	A02
chimica	B01
lettere	B03

possiamo notare che:

- due persone hanno la stessa matricola
- una persona è iscritta dal 1680
- il dipartimento B08 non esiste

per evitare questi problemi si usano dei vincoli di integrità.

#### vincoli di integrità

un vincolo di integrità è un vincolo che ogni istanza del database deve rispettare per essere considerato corretto. un vincolo può essere:

- **intrarelazionale** un vincolo definito all'interno di una relazione
- **interrelazionale** un vincolo definito su più relazioni

si introducono i seguenti vincoli:

- **vincolo di unicità:** questo vincolo indica impossibilità avere due tuple con lo stesso valore per un determinato attributo. la matricola deve essere diversa per ogni persona
- **vincolo di esistenza del valore:** questo vincolo indica l'impossibilità di avere Null come valore di un attributo specifico. il nome non può essere un Null
- **vincolo di dominio:** restrizioni imposte sul dominio di un attributo. anno iscrizione deve essere maggiore di 1970
- **vincolo di tupla:** restrizioni imposte su ogni tupla indipendentemente dalle altre

### 3.2 chiavi

oltre a chiavare, definiamo cosa sono le chiavi.

#### chiavi primarie

una chiave primaria è un attributo o un insieme di attributi (del minimo numero possibile) che sono diversi per ogni tupla quindi rispettano il vincolo di unicità e ci permettono di riconoscere una singola entità. Le chiavi primarie devono anche rispettare il vincolo di esistenza.

## 4 Algebra Relazionale

## 5 Teoria Relazionale

## 6 Organizzazione Fisica

## 7 Concorrenza