

# Basi di dati

**Modello relazionale**

# DBMS e modelli logici

- Modelli logici tradizionali:

- gerarchico
- reticolare
- **relazionale**
- a oggetti

- Più recenti:

- object-relational
- Schemaless (NoSQL)

# Il modello relazionale

- Proposto da E. F. Codd nel 1970 per favorire l'indipendenza dei dati è disponibile in DBMS reali dal 1981
- Si basa su uno sviluppo del concetto matematico di relazione
  - Le relazioni hanno una naturale rappresentazione per mezzo di tabelle

# Relazione: tre accezioni

- relazione matematica secondo la **teoria degli insiemi**
- relazione secondo il **modello relazionale** dei dati
- relazione (dall'inglese relationship) che rappresenta una classe di fatti, nel **modello Entity-Relationship**; tradotto anche con associazione o correlazione

# Relazione matematica

- $D_1, \dots, D_n$  ( $n$  insiemi anche non distinti)
- **prodotto cartesiano**  $D_1 \times \dots \times D_n$ :
  - l'insieme di tutte le  $n$ -ple  $(d_1, \dots, d_n)$  tali che  $d_1 \in D_1, \dots, d_n \in D_n$

Si definisce **ennupla** o **tupla** un elenco ordinato di oggetti. Tali oggetti si chiamano elementi, o membri, della ennupla. Una  $n$ -pla si distingue da un **insieme** in quanto fra gli elementi di un insieme non è dato alcun ordine. Inoltre gli elementi di una  $n$ -pla possono anche essere ripetuti.

# Relazione matematica

- $D_1, \dots, D_n$  ( $n$  insiemi anche non distinti)
- **prodotto cartesiano**  $D_1 \times \dots \times D_n$ :
  - l'insieme di tutte le  $n$ -ple  $(d_1, \dots, d_n)$  tali che  $d_1 \in D_1, \dots, d_n \in D_n$
- **relazione matematica** su  $D_1, \dots, D_n$ :
  - un sottoinsieme di  $D_1 \times \dots \times D_n$ .

## ...Relazione matematica

- ... relazione matematica su  $D_1, \dots, D_n$ :  
un sottoinsieme di  $D_1 \times \dots \times D_n$ .
- $D_1, \dots, D_n$  sono i **domini** della relazione
- $n$  è detto **grado** della relazione
- Il numero di  $n$ -ple della relazione è detto **cardinalità** della relazione

# Relazione matematica, esempio

- $D_1 = \{a, b\}$
- $D_2 = \{x, y, z\}$
- prodotto cartesiano  $D_1 \times D_2$

a	x
a	y
a	z
b	x
b	y
b	z

- una relazione  $r \subseteq D_1 \times D_2$

a	x
a	z
b	y



# Relazione matematica, proprietà

- una relazione matematica è un insieme di n-ple:
  - $(d_1, \dots, d_n)$  tali che  $d_1 \in D_1, \dots, d_n \in D_n$
  - ciascuna n-pla è ordinata: l' i-esimo valore proviene dall' i-esimo dominio
  - non c'è ordinamento fra le n-ple (una relazione NON è un insieme ordinato di n-uple);
  - le n-ple sono distinte

# Relazione matematica, esempio

***Partite*  $\subseteq$  *string*  $\times$  *string*  $\times$  *int*  $\times$  *int***

Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

- Struttura **posizionale**
  - Ciascuno dei domini ha un **ruolo** diverso, distinguibile attraverso la posizione (squadra di casa, reti segnate)
  - Il legame posizione-semantica viene codificato nel codice delle applicazioni che leggono i dati

# Relazione matematica, esempio

- Le due relazioni sotto mostrate sono uguali:

Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

Juve	Lazio	3	1
Roma	Milan	0	1
Lazio	Milan	2	0
Juve	Roma	0	2

# Relazione matematica, esempio

- Le due relazioni sotto mostrate NON sono uguali:

Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

## Limiti della notazione posizionale

Lazio	Juve	1	3
Milan	Lazio	0	2
Roma	Juve	2	0
Milan	Roma	1	0

# Struttura non posizionale

- A ciascun dominio si associa un nome (**attributo**), che ne descrive il "ruolo"

Casa	Fuori	RetiCasa	RetiFuori
Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

# Tabelle e relazioni

- Una tabella rappresenta una relazione (nel modello relazionale) se:
  - i valori di ogni colonna sono fra loro omogenei
  - le righe sono diverse fra loro
  - le intestazioni delle colonne sono diverse tra loro
- In una tabella che rappresenta una relazione
  - l'ordinamento tra le righe è irrilevante
  - l'ordinamento tra le colonne è irrilevante

# Modello basato su valori

- Nel modello relazionale, i dati di interesse sono rappresentati attraverso relazioni
  - Ogni relazione è definita su un insieme di attributi
- Il modello relazionale è un modello basato sui valori:
  - i riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di valori del dominio di alcuni attributi

# Modello basato su valori

- Esempio: organizzare le informazioni relative ai CdL a cui gli studenti di un Ateneo sono iscritti
  - Nome, Cognome e Data di nascita di ciascuno studente
  - Nome dei CdL



# Modello basato su valori

## studenti

Matricola	Cognome	Nome	Data di nascita	CdL
6554	Rossi	Mario	05/12/1978	02
8765	Neri	Paolo	03/11/1976	01
9283	Verdi	Luisa	12/11/1979	04
3456	Rossi	Maria	01/02/1978	04

## CdL

Codice	Nome
01	Ing. Industriale
02	Ing. Informatica
04	Ing. Elettrica

# Modello basato su valori

- Esempio: organizzare le informazioni relative agli studenti di un CdL
  - Nome, Cognome e Data di nascita di ciascuno studente
  - Nome dei corsi del CdL
  - Esami superati da ciascuno studente e relativo voto

# Modello basato su valori

## studenti

Matricola	Cognome	Nome	Data di nascita
6554	Rossi	Mario	05/12/1978
8765	Neri	Paolo	03/11/1976
9283	Verdi	Luisa	12/11/1979
3456	Rossi	Maria	01/02/1978

## voti

Matricola	Voto	Codice
6554	28	01
6554	30	02
9283	28	04

## corsi

Codice	Titolo
01	Analisi
02	Chimica I
04	Chimica II

# Modello basato su valori

- Vengono introdotti dei dati aggiuntivi (dei *codici*), non necessariamente richiesti all'atto della descrizione del contesto da modellare
- Il valore dei codici è usato per correlare le n-ple di tabelle diverse (o della stessa tabella)
- Nei modelli gerarchico e reticolare al posto di tali codici erano impiegati dei puntatori...

# Riferimenti con puntatori

**studenti**

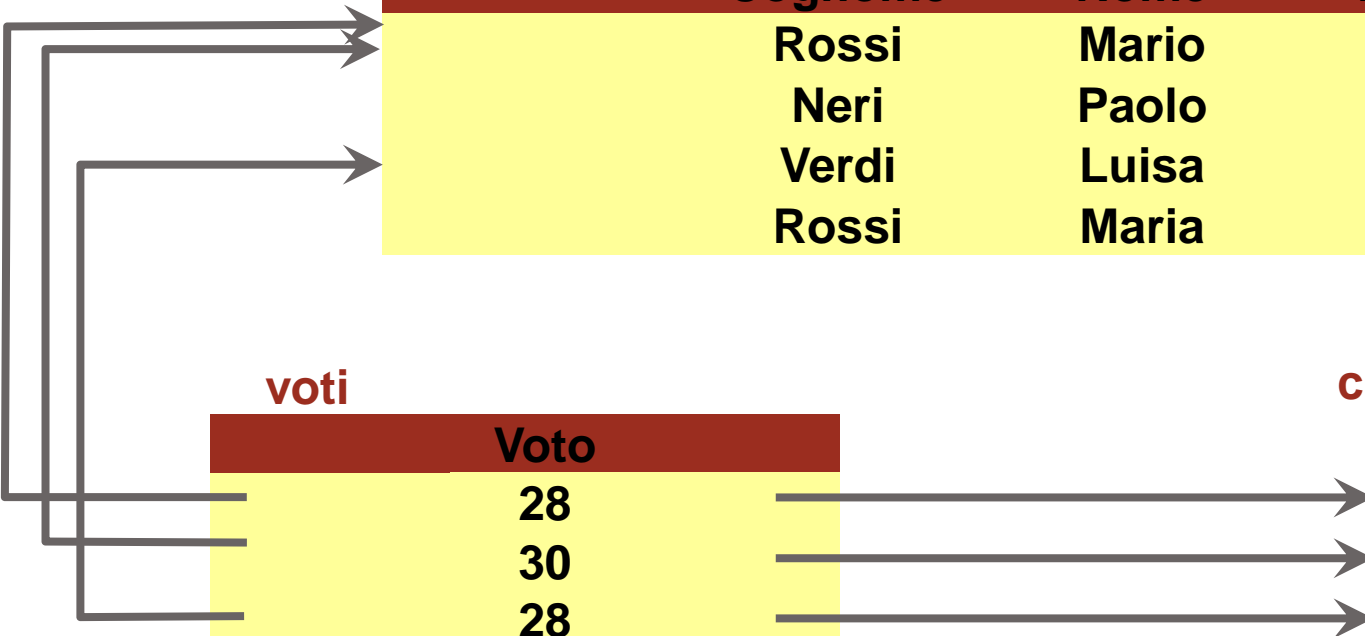
Cognome	Nome	Data di nascita
Rossi	Mario	05/12/1978
Neri	Paolo	03/11/1976
Verdi	Luisa	12/11/1979
Rossi	Maria	01/02/1978

**voti**

Voto
28
30
28

**corsi**

Titolo
Analisi
Chimica I
Chimica II



# Struttura basata su valori

- Vantaggi:
  - indipendenza dalle strutture fisiche che possono cambiare dinamicamente
  - si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione
  - l'utente finale vede gli stessi dati dei programmatori
  - i dati sono portabili più facilmente da un sistema ad un altro
  - i puntatori sono direzionali

# Definizioni

- **Schema di relazione**: costituito da un simbolo **R** detto nome della relazione e da un insieme di attributi  $X = \{A_1, \dots, A_n\}$ , ciascuno con un nome distinto

$$R(A_1, \dots, A_n)$$

- A ciascun attributo è associato un dominio ed un insieme di operatori di confronto

# Definizioni

- **Schema di base di dati**: insieme di schemi di relazione con nomi distinti

$$R = \{R_1(X_1), \dots, R_k(X_k)\}$$



# Definizioni

- Una **ennupla** o **tupla** su un insieme di attributi  $X$  è una funzione che associa a ciascun attributo  $A \in X$  un valore del dominio di  $A$
- $t[A]$  denota il valore della ennupla  $t$  sull'attributo  $A$ 
  - $X = \{\text{codice}, \text{prodotto}, \text{costo}\}$
  - $A = \{\text{prodotto}\}$
  - $t = (\text{c024}, \text{MacBook PRO 13}, 999.20)$
  - $t[A] = ?$
- La struttura dati usata per memorizzare una tupla è il **record**
- I valori di una tupla su ciascun attributo corrispondono ai **campi** del record

# Definizioni

- (Istanza di) **relazione** su uno schema  $R(X)$ :  
insieme  $r$  di ennuple su  $X$
- (Istanza di) **base di dati** su uno schema  
 $R = \{R_1(X_1), \dots, R_n(X_n)\}$ :  
insieme di relazioni  $r = \{r_1, \dots, r_n\}$  (con  $r_i$   
relazione su  $R_i$ )

# Relazioni su singoli attributi

- La definizione non limita il numero di attributi di una relazione. Ha senso considerare relazioni definite su un solo attributo?

## studenti

Matricola	Cognome	Nome	Data di nascita
6554	Rossi	Mario	05/12/1978
8765	Neri	Paolo	03/11/1976
9283	Verdi	Luisa	12/11/1979
3456	Rossi	Maria	01/02/1978

## studenti lavoratori

Matricola
6554
3456

# Una semplice base dati

- Definire lo schema di una base dati atta a memorizzare le informazioni di interesse per monitorare lo stato dei prodotti presenti in un magazzino. In particolare:
  - codice prodotto
  - descrizione prodotto
  - numero di pezzi presenti in magazzino

# Una semplice base dati

- Si può osservare che per ogni codice di prodotto esiste **una sola** descrizione ed **un solo** valore che rappresenta il numero di prodotti disponibili ...

# Una semplice base dati

- Supponiamo di voler modificare la base di dati in modo da registrare anche i dati dei fornitori (codice e nominativo)
- Si deve anche tener traccia dei fornitori di ciascun prodotto
- Come potremmo espandere il seguente schema per memorizzare i dati in più?

## Prodotti

Codice	Descrizione	Disponibili
--------	-------------	-------------

# Una semplice base dati

## Sol.1

### Prodotti

Codice	Descrizione	Disponibili
--------	-------------	-------------

### Fornitori

Codice	Nominativo
--------	------------

E' necessario aggiungere una relazione Fornitori per memorizzare i dati di ciascun fornitore

Poi bisogna aggiungere qualcosa che metta in relazione le righe di Prodotti con quelle di Fornitori...

## Sol.2

### Prodotti

Codice	Descrizione	Disponibili
--------	-------------	-------------

### Fornitori

Codice	Nominativo
--------	------------

# Una semplice base dati

**Sol.1**

**Prodotti**

**Codice**

**Descrizione**

**Disponibili**

**Fornitori**

**Codice**

**Nominativo**

**Fornitura**

**CodProd**

**CodForn**

**Sol.2**

**Prodotti**

**Codice**

**Descrizione**

**Disponibili**

**CodForn**

**Fornitori**

**Codice**

**Nominativo**



# Strutture nidificate

- Spesso i dati da rappresentare non sono in relazione uno ad uno:
  - Un prodotto ed i suoi fornitori
  - Uno studente e gli esami che ha sostenuto
  - Un cantautore e le canzoni che ha composto
  - Una persona ed i suoi recapiti telefonici
- In questi casi, per una rappresentazione efficiente dei dati bisogna fare ricorso a più tabelle:
  - una tabella per ciascuna delle entità coinvolte (prodotto, fornitore, ...) ed altre tabelle per rappresentare le relazioni tra le entità

# Strutture nidificate

<b><i>DA FILIPPO</i></b> <b><i>VIA ROMA 2, ROMA</i></b>		
<b><i>RICEVUTA FISCALE</i></b> <b><i>1235 DEL 12/10/2000</i></b>		
<b>3</b>	<b>Coperti</b>	<b>3,00</b>
<b>2</b>	<b>Antipasti</b>	<b>7,00</b>
<b>3</b>	<b>Primi</b>	<b>12,00</b>
<b>2</b>	<b>Bistecche</b>	<b>18,00</b>
<b><i>TOTALE</i></b>		<b>40,00</b>

<b><i>DA FILIPPO</i></b> <b><i>VIA ROMA 2, ROMA</i></b>		
<b><i>RICEVUTA FISCALE</i></b> <b><i>1240 DEL 13/10/2000</i></b>		
<b>2</b>	<b>Coperti</b>	<b>2,00</b>
<b>2</b>	<b>Antipasti</b>	<b>7,00</b>
<b>2</b>	<b>Primi</b>	<b>8,00</b>
<b>2</b>	<b>Orate</b>	<b>20,00</b>
<b>2</b>	<b>Caffè</b>	<b>2,00</b>
<b><i>TOTALE</i></b>		<b>39,00</b>

# Strutture nidificate

## Soluzione errata

Ricevute	Numero	Data	Numero Coperti	Costo Coperti	Numero Antipas	Costo Antipas	Numero Primi	Costo Primi	Numero Orate	Costo Orate		Totale
	1235	12/10/2000	3	3	2	6.2	3	12	0	0	■ ■ ■	39,20
	1240	13/10/2000	2	2	2	7	2	8	2	20		39,00

**L'aggiunta di una nuova tipologia di portata richiede una modifica dello schema del DB**

# Relazioni che rappresentano strutture nidificate

## Ricevute

Numero	Data	Totale
1235	12/10/2000	40,00
1240	13/10/2000	39,00

## Dettaglio

Numero	Qtà	Descrizione
1235	3	Coperti
1235	2	Antipasti
1235	3	Primi
1235	2	Bistecche
1240	2	Coperti
...	...	...

## Costo

Descrizione	Importo
Coperti	1,00
Antipasti	3,50
Primi	4,00
Bistecche	9,00
Orate	10,00
...	...

# Strutture nidificate, riflessione

- Abbiamo rappresentato veramente tutti gli aspetti delle ricevute?
  - l'ordine delle righe è rilevante?
  - possono esistere linee ripetute?

# Rappresentazione alternativa per strutture nidificate

**Ricevute**

<b>Numero</b>	<b>Data</b>	<b>Totale</b>
<b>1235</b>	<b>12/10/2000</b>	<b>40,00</b>
<b>1240</b>	<b>13/10/2000</b>	<b>39,00</b>

**Dettaglio**

<b>Numero</b>	<b>Riga</b>	<b>Qtà</b>	<b>Descrizione</b>
<b>1235</b>	<b>1</b>	<b>3</b>	<b>Coperti</b>
<b>1235</b>	<b>2</b>	<b>2</b>	<b>Antipasti</b>
<b>1235</b>	<b>3</b>	<b>3</b>	<b>Primi</b>
<b>1235</b>	<b>4</b>	<b>2</b>	<b>Bistecche</b>
<b>1240</b>	<b>1</b>	<b>2</b>	<b>Coperti</b>
<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>

# Esercizio

- Specificare lo schema di una base dati che rappresenti le informazioni di interesse per un sistema di archiviazione di libri.  
Per ogni libro, il sistema deve consentire l'archiviazione di informazioni su:
  - Numero ISBN e numero di collocazione (supponiamo esista una sola copia di ogni libro)
  - Autori (cod. fisc., nome, cognome, indirizzo e recapito telefonico) (alcuni libri hanno più di 1 autore)
  - Casa editrice (nome, sede e recapito telefonico)



# Informazione incompleta

## LIBRI

ISBN	COLLOCAZIONE	CASAED
1001010-011	C24	McGrawHill
1002010-011	C14	Apogeo

## SCRITTORI

CODICEF	NOME	COGNOME	INDIRIZZO	TELEFONO
gvnrss71E12j357d	Giovanni	Rossi	Via dei Neri, 5	???

## AUTORI

ISBN	CODICEF
1001010-011	gvnrss71E12j357d
1002010-011	gvnrss71E12j357d

## CASAED

CODICEF	SEDE	TELEFONO
McGrawHill	Milano	02151515



# Informazione incompleta

- non conviene (anche se spesso si fa) usare valori del dominio (0, stringa vuota, “99”, ...):
  - potrebbero non esistere valori “non utilizzati”
  - valori “non utilizzati” potrebbero diventare significativi
  - in fase di utilizzo (nei programmi) sarebbe necessario ogni volta tener conto del “significato” di questi valori

# Informazione incompleta

- Soluzione: si introduce un nuovo valore
  - **valore nullo**: denota l'assenza di un valore del dominio (e non è un valore del dominio)
- Viene modificata la definizione di tupla:
  - $t[A]$ , per ogni attributo  $A$ , è un valore del dominio  $\text{dom}(A)$  oppure il valore nullo NULL

# Troppi valori nulli

**studenti**

Matricola	Cognome	Nome	Data di nascita
6554	Rossi	Mario	NULL
9283	Verdi	Luisa	12/11/1979
NULL	Rossi	Maria	01/02/1978

**esami**

Studente	Voto	Corso
NULL	30	NULL
NULL	24	02
9283	28	01

**corsi**

Codice	Titolo
01	Analisi
02	NULL
04	Chimica

# Tipi di valore nullo

- Il valore nullo può essere impiegato in tre casi differenti:
  - valore **sconosciuto** (esiste ma non lo conosco)
  - valore **inesistente** (so che non esiste)
  - valore **senza informazione** (non so se esiste o no)
- Tuttavia, i DBMS non distinguono i tipi di valore nullo