



Basi Di Dati e di conoscenza

Modello Relazionale



Contenuti della Lezione

- modello relazionale
- strutture nidificate
- informazione incompleta



Modello relazionale

- Proposto da E. F. Codd nel 1970 per favorire **l'indipendenza dei dati**
- Disponibile in **DBMS reali nel 1981** (non è facile implementare l'indipendenza con efficienza e affidabilità!)
- Si basa sul concetto matematico di **relazione (con una variante)**
- Le relazioni hanno naturale rappresentazione per mezzo di **tabelle**

Relazione matematica

- D_1, \dots, D_n : insiemi anche non distinti detti **domini**
- Il **prodotto cartesiano** $D_1 \times \dots \times D_n$ è definito come:

l'insieme di tutte le **n -uple** (d_1, \dots, d_n) tali che $d_1 \in D_1, \dots, d_n \in D_n$

- Una **relazione matematica** su D_1, \dots, D_n è un sottoinsieme di $D_1 \times \dots \times D_n$.
- D_1, \dots, D_n sono i **domini della relazione**

Relazione matematica

$$D_1 = \{A, B\}$$

A
B

$$D_2 = \{X, Y, Z\}$$

X
Y
Z

prodotto cartesiano $D_1 \times D_2$

A X
A Y
A Z
B X
B Y
B Z

Relazione $r \subseteq D_1 \times D_2$

A X
B X
B Y

Relazione matematica: proprietà

- una **relazione matematica** è un insieme di n-uple ordinate:
 - (d_1, \dots, d_n) tali che $d_1 \in D_1, \dots, d_n \in D_n$
- una **relazione** è un **insieme**; quindi:
 1. **non c'è ordinamento** fra le n-uple;
 2. le n-uple sono **distinte**
 3. **ciascuna n-upla è ordinata**: l' i-esimo valore proviene dall' i-esimo dominio

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	2

- Ciascuno dei domini ha due **ruoli** diversi, distinguibili attraverso la posizione:
- La struttura è **posizionale**

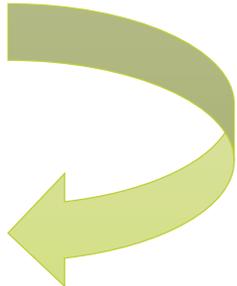
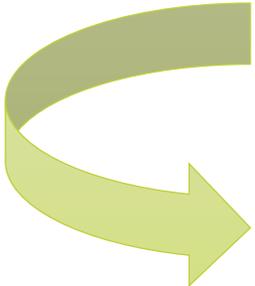
Struttura non posizionale

- Se a ciascun dominio si associa un nome (**attributo**), che ne descrive il "ruolo,, la struttura diviene non posizionale.

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

Collezione di funzioni

- Per meglio catturare il concetto di relazione del modello relazionale definiamo
 - $X = \{A_1, \dots, A_n\}$: un insieme (**non ordinato**) di attributi
 - $DOM=X \rightarrow D$: funzione che associa ad un attributo il corrispondente dominio.
 - Una **ennupla** o **tupla** è una funzione t che associa ad ogni $A \in X$ **un valore del dominio**.
 - $t[A]$ denota il valore della ennupla t sull'attributo A



Una **relazione** è una collezione di **ennuple**

Modello relazionale: esempio

- **Esempio di relazione:** Relazione rappresentata tramite tabella (solo una delle possibile forme)

Nome	Cognome	Matricola	Voto medio
Mario	Rossi	1	24
Luigi	Bianchi	2	28
Rosa	Rossa	3	26

Modello relazionale: notazione

- **Notazione**
- Se t è una tupla su X e A è un attributo, con $A \in X$ allora $t[A]$ indica il valore di t su A .
- **Esempio:** se t è la prima tupla allora...

$t[\text{Cognome}] \rightarrow \text{'Rossi'}$

Nome	Cognome	Matricola	Voto medio
Mario	Rossi	1	24
Luigi	Bianchi	2	28
Rosa	Rossa	3	26

Tabelle e relazioni

- Una tabella rappresenta una relazione 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

Tabelle: esempi

- Ordine non rilevante
- NO due righe uguali
- NO dati non omogeni

Nome	Cognome	Matricola	Voto medio
Mario	Rossi	1	24
Luigi	Bianchi	2	28
Rosa	Rossa	3	26

Voto medio	Cognome	Nome	Matricola
24	Rossi	Mario	1
28	Bianchi	Luigi	2
26	Rossa	Rosa	3

Voto medio	Cognome	Nome	Matricola
24	Rossi	Mario	1
28	Bianchi	Luigi	2
26	Rossa	Rosa	3
24	Rossi	Mario	1

Voto medio	Cognome	Nome	Matricola
24	Rossi	Mario	1
Bianchi	28	Luigi	2
26	3	Rosa	Rossa

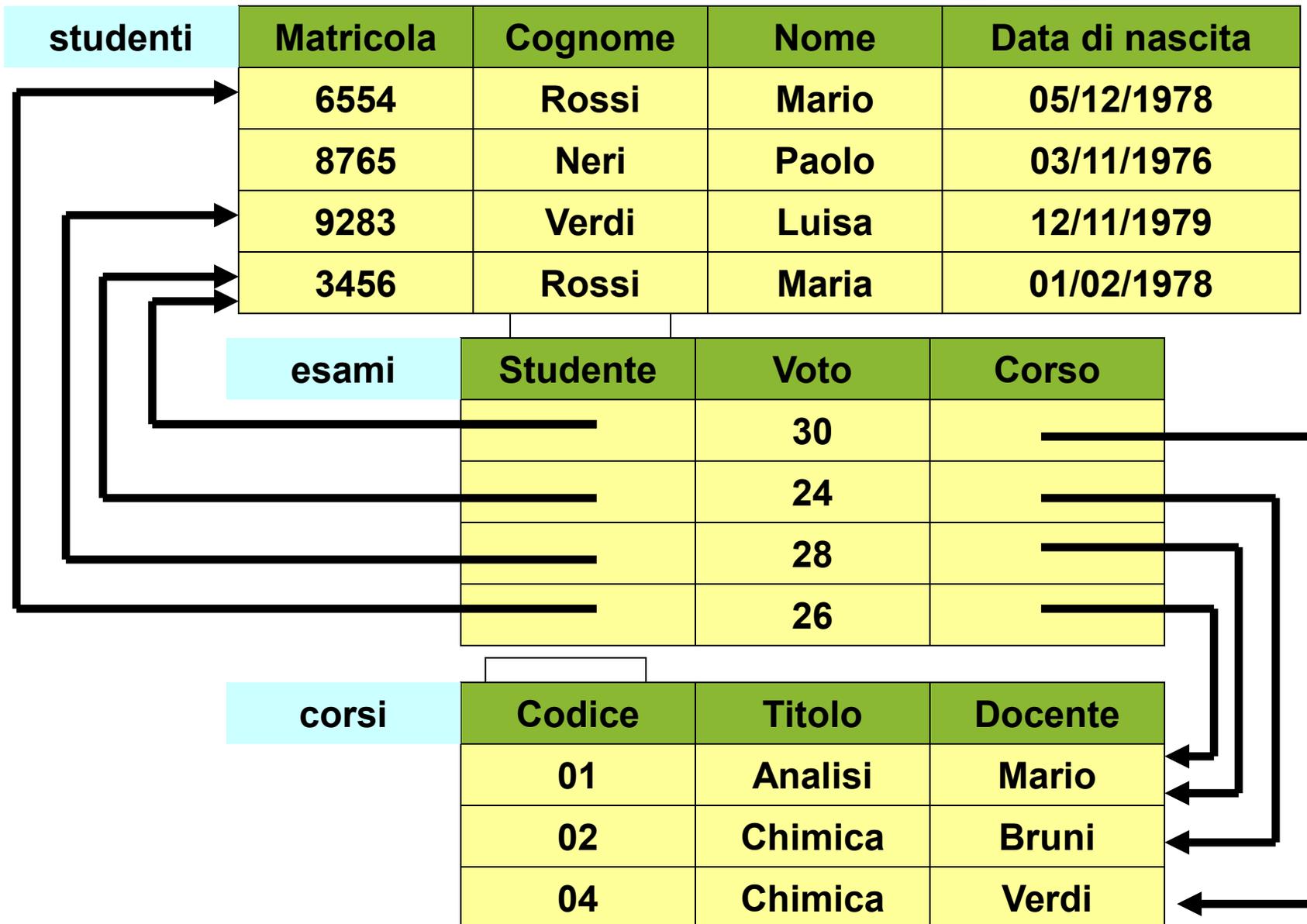
Modello basati sui valori

- i riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di valori dei domini che compaiono nelle ennuple

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

esami	Studente	Voto	Corso
	3456	30	04
	3456	24	02
	9283	28	01
	6554	26	01

corsi	Codice	Titolo	Docente
	01	Analisi	Mario
	02	Chimica	Bruni
	04	Chimica	Verdi



Vantaggi della struttura basata sui valori

- indipendenza dalle strutture fisiche (si potrebbe avere anche con puntatori di alto livello) 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

Modello relazionale: Definizioni

- **Schema di relazione:**

un nome R con un insieme di attributi A_1, \dots, A_n :

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

- **Schema di base di dati:**

insieme di schemi di relazione:

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

Schema di relazione e di base di dati: (esempio)

- **Schema di relazione**

STUDENTI (Matricola, Cognome, Nome, Data di Nascita)

- **Schema di basi di dati**

STUDENTI (Matricola, Cognome, Nome, Data di Nascita)

ESAMI (Matricola, Voto, Corso)

CORSO (Codice, Titolo, Docente)

Modello relazionale: 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

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

Contenuti della Lezione

- modello relazionale
- strutture nidificate
- informazione incompleta



Strutture nidificate

DA FILIPPO VIA ROMA 2, ROMA		
RICEVUTA FISCALE 1235 DEL 12/10/2000		
3	Coperti	3,00
2	Antipasti	6,20
3	Primi	12,00
2	Bistecche	18,00
TOTALE		39,20

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

Rappresentazione relazionale delle strutture nidificate

Ricevute

Numero	Data	Totale
1235	12/10/2000	39,20
1240	13/10/2000	39,00

Dettaglio

Numero	Qtà	Descrizione	3,00
1235	3	Coperti	3,00
1235	2	Antipasti	6,20
1235	3	Primi	12,00
1235	2	Bistecche	18,00
1240	2	Coperti	2,00
...

Strutture nidificate

- Abbiamo rappresentato veramente tutti gli aspetti delle ricevute?
- Dipende da che cosa ci interessa!
 - l'ordine delle righe e' rilevante?
 - possono esistere linee ripetute in una ricevuta?
- Sono possibili rappresentazioni diverse

Rappresentazione alternativa

Ricevute

Numero	Data	Totale
1235	12/10/2000	39,20
1240	13/10/2000	39,00

Dettaglio

Numero	Riga	Qtà	Descrizione	Importo
1235	1	3	Coperti	3,00
1235	2	2	Antipasti	6,20
1235	3	3	Primi	12,00
1235	4	2	Bistecche	18,00
1240	1	2	Coperti	2,00
...

Contenuti della Lezione

- modello relazionale
- strutture nidificate
- informazione incompleta



Informazione incompleta

- Il modello relazionale impone ai dati una struttura rigida:
 - le informazioni sono rappresentate per mezzo di ennuple
 - solo alcuni formati di ennuple sono ammessi: quelli che corrispondono agli schemi di relazione
- I dati disponibili possono non corrispondere al formato previsto

Informazione incompleta: motivazioni

Nome	SecondoNome	Cognome
Franklin	Delano	Roosevelt
Winston		Churchill
Charles		De Gaulle
Josip		Stalin

Informazione incompleta: soluzioni?

- non conviene (anche se spesso si fa) usare valori del dominio (0, stringa nulla, “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 nel modello relazionale

- Si adotta una tecnica rudimentale ma efficace:
 - **valore nullo**: denota l'assenza di un valore del dominio (e non è un valore del dominio)
- $t[A]$, per ogni attributo A , è un valore del dominio $\text{dom}(A)$ oppure il valore nullo NULL
- Si possono (e debbono) imporre restrizioni sulla presenza di valori nulli

Troppi valori nulli

studenti	Matricola	Cognome	Nome	Data di nascita
	6554	Rossi	Mario	05/12/1978
	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	Docente
	01	Analisi	Mario
	02	NULL	NULL
	04	Chimica	Verdi

Tipi di valore nullo

- (almeno) tre casi differenti
 - valore sconosciuto
 - valore inesistente
 - valore senza informazione
- I DBMS non distinguono i tipi di valore nullo