

Corso di Laurea
in
Ingegneria Informatica

“Basi di dati”
a.a. 2019-2020

Docente: Gigliola Vaglini
Docente laboratorio SQL: Francesco
Pistolesi

Lezione 2

Il modello relazionale

Il modello relazionale

- Proposto da E. F. Codd nel 1970, ma usato in DBMS reali solo dal 1981
- Si basa sul concetto matematico di relazione (con una variante)

3

Relazione matematica: struttura posizionale

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

- Ciascuno dei domini uguali ha ruolo diverso, distinguibile attraverso la posizione
 - la struttura è posizionale
 - non c'è ordinamento fra le n-uple

4

Relazione matematica

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

5

Relazione matematica: proprietà

- una relazione è un insieme di n -uple ordinate (d_1, \dots, d_n) tali che $d_1 \in D_1, \dots, d_n \in D_n$ (cioè l' i -esimo valore proviene dall' i -esimo dominio)
- non c'è ordinamento fra le n -uple

6

Relazione matematica: esempio

- $D_1 = \{a, b\}$
- $D_2 = \{x, y, z\}$
- prodotto cartesiano
 - $D_1 \times D_2 = \{(a, x), (a, y), (a, z), (b, x), (b, y), (b, z)\}$
- relazione
 - $r \subseteq D_1 \times D_2 = \{(a, x), (a, z), (b, y)\}$

7

Struttura non posizionale

- A ciascun dominio si associa un nome (attributo) che ne descrive il "ruolo",
 - L'ordinamento tra righe (o tuple) è irrilevante, e
 - anche l'ordinamento tra le colonne è irrilevante

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

8

Tabelle e relazioni

- Le relazioni hanno naturale rappresentazione per mezzo di tabelle
- 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

9

Il modello relazionale è basato su valori

- i riferimenti fra dati di relazioni diverse sono ottenuti tramite valori uguali in tuple diverse

10

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

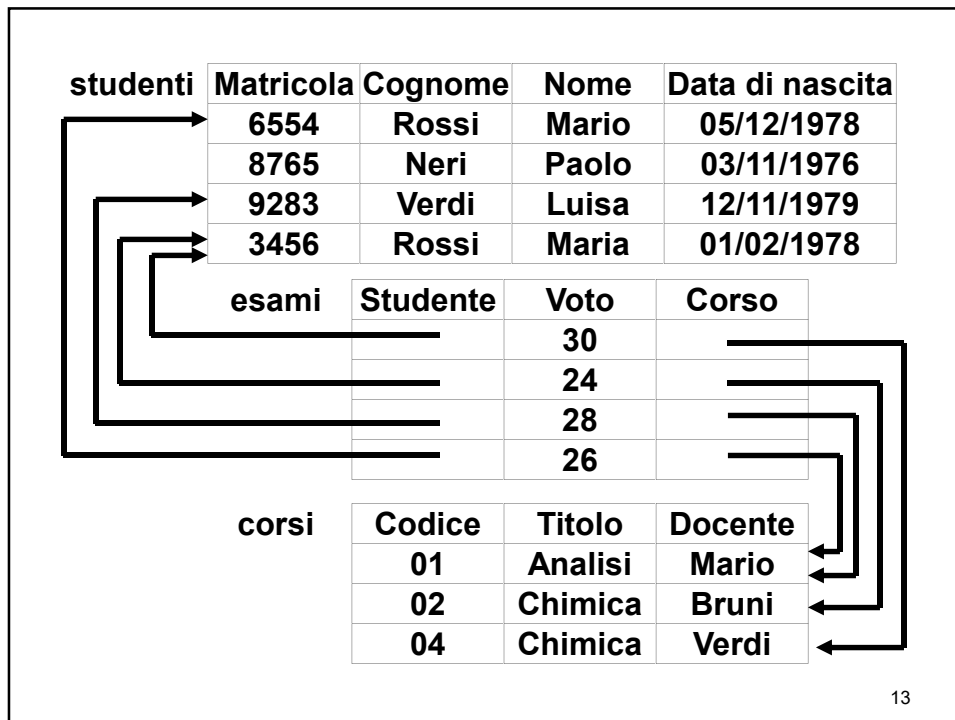
corsi	Codice	Titolo	Docente
	01	Analisi	Mario
	02	Chimica	Bruni
	04	Chimica	Verdi

11

Alternativa

- Altri modelli (sia quelli "storici", reticolare e gerarchico, sia quello a oggetti) prevedono riferimenti espliciti, gestiti dal sistema
- Non un valore uguale, ma un puntatore al valore

12



13

Vantaggi

- indipendenza dalle strutture fisiche (in realtà i dati non hanno alcuna struttura)
- i dati sono portabili più facilmente da un sistema ad un altro
- l'utente finale vede gli stessi dati dei programmatori
- puntatori eventuali a livello fisico non sono visibili a livello logico

14

Definizioni (1)

- 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

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

15

Definizioni (2)

- Una 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 tupla t sull'attributo A

16

Definizioni (3)

- (Istanza di) relazione su uno schema $R(X)$:
 - insieme r di tuple 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)

17

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

18

Strutture complesse

- I valori devono essere semplici (non relazioni)
- Insieme di relazioni per rappresentare strutture complesse

19

Informazione incompleta (1)

- Il modello relazionale impone ai dati una struttura rigida
 - le informazioni sono rappresentate da tuple il cui formato deve corrispondere agli schemi di relazione
- Ma i dati disponibili possono non corrispondere al formato previsto

20

Informazione incompleta (2)

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

21

Rappresentazione dell'informazione incompleta

- Si usa un valore distinto aggiunto a tutti i domini
 - valore nullo: denota l'assenza di un valore del dominio
 - se i valori dell'attributo A appartengono al dominio $\text{dom}(A)$, $t[A]$ è un valore del dominio oppure il valore nullo NULL
- Si possono (e debbono) imporre restrizioni sulla presenza di valori nulli in una relazione

22

Troppi valori nulli rendono prive di significato le tuple relative

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

23

Attenzione: tre tipi differenti di significato del valore nullo

- valore sconosciuto
 - valore inesistente
 - valore non interessante
-
- I DBMS non distinguono tra tipi diversi di valore nullo

24

Basi di dati „scorrette“

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

25

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

26

Vincoli di integrità

- Si devono associare alla base di dati delle proprietà che, se soddisfatte, esprimono la sua “correttezza” rispetto all’ applicazione
- I cosiddetti “vincoli di integrità”
 - permettono una descrizione più accurata della realtà
 - danno un contributo alla “qualità dei dati”
 - sono utili nella progettazione
 - sono usati dai DBMS nella esecuzione delle interrogazioni

27

Vincoli, schemi e istanze

- i vincoli corrispondono a proprietà del mondo reale modellato dalla base di dati e interessano tutte le istanze
- i vincoli sono associati allo schema e si considerano corrette le sue istanze che soddisfano tutti i vincoli

28

Vincoli di integrità

- Un vincolo è un predicato che associa ad ogni istanza della base di dati il valore vero o falso
- Se il predicato vale vero la proprietà è soddisfatta
- Due tipi di vincoli
 - intrarelazionali
 - interrelazionali

29

Vincoli intrarelazionali

- Vincoli di tupla: esprimono condizioni sui valori di ciascuna tupla, indipendentemente dalle altre
- Caso particolare
 - Vincoli di dominio: coinvolgono un solo attributo

30

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

31

Sintassi del vincolo di dominio

- Una possibile sintassi:
 - espressione booleana di atomi che confrontano valori di attributo o espressioni aritmetiche su di essi

$(Voto \geq 18) \text{ AND } (Voto \leq 30)$

32

Vincoli su più domini

(Voto = 30) OR NOT (Lode = "e lode")

Stipendi	Impiegato	Lordo	Ritenute	Netto
	Rossi	55.000	12.500	42.500
	Neri	45.000	10.000	35.000
	Bruni	47.000	11.000	36.000

Lordo = (Ritenute + Netto)

33

Identificatore di tupla

Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Inf	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	5/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- non ci sono due tuple con lo stesso valore dell'attributo *Matricola*
- non ci sono due tuple con lo stesso valore di tutti e tre gli attributi *Cognome*, *Nome* e *Nascita*

34

Chiave

- insieme di attributi che identificano univocamente le tuple di una relazione
Formalmente:
- un insieme K di attributi è superchiave per r se r non contiene due tuple distinte t_1 e t_2 con $t_1[K] = t_2[K]$
- K è chiave per r se è una superchiave minimale per r
(cioè non contiene un'altra superchiave)

35

Esempio (1)

Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Inf	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	5/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- Matricola è una chiave:
 - è superchiave
 - contiene un solo attributo e quindi è minimale

36

Esempio (cont.)

Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Inf	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	5/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- Cognome, Nome, Nascita formano una superchiave, ma non minimale
- Cognome, Nascita non formano una chiave perché non distinguono la coppia di riga 5 da quella di riga 1
- Nome, Nascita formano una chiave

37

Esistenza delle chiavi

- Una relazione contiene tuple tutte distinte tra loro (è un insieme)
- Ogni relazione ha come superchiave l'insieme degli attributi su cui è definita
 - quindi ha (almeno) una chiave

38

Importanza delle chiavi

- L'esistenza delle chiavi garantisce l'accessibilità a ciascun dato della base di dati
- Le chiavi permettono di correlare i dati in relazioni diverse

39

Chiavi e valori nulli

- In presenza di valori nulli delle chiavi non è possibile
 - identificare le tuple
 - realizzare i riferimenti con altre relazioni
- La presenza di valori nulli nelle chiavi deve essere limitata

40

Chiave primaria

- Chiave su cui non sono ammessi valori nulli
- Notazione: sottolineatura

<u>Matricola</u>	Cognome	Nome	Corso	Nascita
86765	NULL	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Civile	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	NULL
43289	Neri	Mario	NULL	5/12/78

41

Le dipendenze funzionali

- I vincoli di chiave sono particolari tipi di vincoli, parte di una categoria più vasta: le dipendenze funzionali
 - Dati due insiemi di attributi X e Y
 - Si dice che X *determina* Y, o che Y *dipende funzionalmente* da X, e si scrive $X \rightarrow Y$ se:
 - *date due tuple distinte t1 e t2 , se $t1[X] = t2[X]$ allora $t1[Y] = t2[Y]$*
- Le dipendenze funzionali possono essere usate per garantire opportune proprietà della BD

42

Vincoli interrelazionali

- Informazioni in relazioni diverse possono essere correlate attraverso valori comuni, in particolare, valori delle chiavi (primarie)
- Le correlazioni debbono essere "coerenti"

43

Infrazioni

<u>Codice</u>	<u>Data</u>	<u>Vigile</u>	<u>Prov</u>	<u>Numero</u>
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

Vigili

<u>Matricola</u>	<u>Cognome</u>	<u>Nome</u>
3987	Rossi	Luca
3295	Neri	Piero
9345	Neri	Mario
7543	Mori	Gino

44

- Esiste un vincolo interrelazionale tra
 - l'attributo Vigile della relazione Infrazioni e la chiave primaria della relazione Vigili (Matricola)

45

Vincoli di integrità referenziale

- Un vincolo di integrità referenziale fra gli attributi X (anche più di uno) di una relazione R_1 e un'altra relazione R_2 impone ai valori su X in R_1 di comparire come valori della chiave primaria di R_2

46

Vincoli su più attributi

N.B. in questo caso

- L'ordine degli attributi tra cui è stabilito il vincolo è significativo

47

Integrità referenziale e valori nulli

- In presenza di valori nulli i vincoli possono essere resi meno restrittivi
- Il vincolo non è fra ogni valore degli attributi X di una relazione R_1 e la chiave primaria della relazione R_2 , ma tra i valori di X diversi da NULL e la chiave primaria di R_2

48

Integrità referenziale e valori nulli

Impiegati

<u>Matricola</u>	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	XYZ
64521	Verdi	NULL
73032	Bianchi	IDEA

Progetti

<u>Codice</u>	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
BOH	09/2001	24	150

49

Reazione alle violazioni dei vincoli

- Cosa succede quando si tenta di compiere un'operazione che viola un vincolo, ad esempio si cerca di inserire nella base di dati un valore non consentito per quell'attributo
- Sono possibili meccanismi per il supporto alla gestione delle violazioni ("azioni compensative")

50

Gestione della violazione di un vincolo intrarelazionale

- Comportamento "standard":
 - Rifiuto dell'operazione

51

Operazioni interferenti con un vincolo interrelazionale

- Si producono violazioni sulla tabella interna a seguito di modifiche
 - Azione compensativa: nessuna: l'operazione viene impedita
- Azioni sulla tabella esterna ("master")
 - Si vuole eliminare una ennupla
 - Si vuol modificare l'attributo riferito

52

Azioni compensative per violazioni derivanti da modifiche sulla tabella esterna (1)

- Cancellazione di una riga della tabella esterna o modifica dell'attributo riferito
 - Eliminazione (modifica) in cascata sulla tabella interna (cascade) delle righe corrispondenti
 - Viene posto a null il valore dell'attributo referente (set null)
 - Viene assegnato un valore di default all'attributo referente (set default)
 - La cancellazione non viene consentita (no action)

53

Eliminazione in cascata

Impiegati

<u>Matricola</u>	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	XYZ
64521	Verdi	NULL
73032	Bianchi	IDEA

Progetti

<u>Codice</u>	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
BOH	09/2001	24	150

54

Introduzione di valori nulli

Impiegati	<u>Matricola</u>	Cognome	Progetto
	34321	Rossi	IDEA
	53524	Neri	NULL
	64521	Verdi	NULL
	73032	Bianchi	IDEA

Progetti	<u>Codice</u>	Inizio	Durata	Costo
	IDEA	01/2000	36	200
	XYZ	07/2001	24	120
	BOH	09/2001	24	150

55

Esercizio 1a

Date le seguenti tabelle

AUTORE (Nome, Cognome, DataNascita, Nazionalità)

LIBRO (TitoloLibro, NomeAutore, CognomeAutore, Lingua)

Esiste un vincolo di integrità referenziale tra di esse?

Esercizio 1b

- Supponendo di avere specificato una politica di cascade sulle modifiche e sulle cancellazioni, spiegare quale è l'effetto dell'esecuzione dei seguenti aggiornamenti:
 - cancella da AUTORE tutte le righe dove Cognome = 'Rossi';
 - modifica LIBRO fissando NomeAutore= 'Umberto' in tutte le righe dove CognomeAutore = 'Eco';

Soluzione

- Il primo comando cancella dalla tabella AUTORE tutte le tuple con Cognome = 'Rossi'. A causa della politica cascade, anche tutte le tuple di LIBRO con CognomeAutore = 'Rossi' vengono eliminate.
- Il comando causa una violazione a meno che la tabella AUTORE contenga già la tupla "Umberto Eco".

- Se in Autore non c'è la tupla "Umberto Eco" ma solo "Giovanni Eco" cosa si deve fare?

NoSQL database

- I principali metodi d'implementazione dei database NoSQL sono i seguenti:
- **Columnfamily**: i dati sono organizzati in righe e colonne, ma le righe possono avere quante colonne si vogliono e non c'è bisogno di definire le colonne come prima cosa.
- **Document store**: rispetto ai normali database relazionali, invece che immagazzinare i dati in tabelle con dei campi fissi, questi vengono messi in un documento che può contenere illimitati campi di illimitata lunghezza, così se, ad esempio, di una persona conosciamo solo nome e cognome, ma magari di un'altra persona anche indirizzo, data di nascita e codice fiscale, si evita che per il primo nominativo ci siano campi inutilizzati.

Vantaggio importante

- Sfruttando la dinamicità della dimensione delle righe, ogni elemento può contenere tutte le informazioni collegate alla chiave. Di conseguenza **non serve usare dispendiosi** (in termini di performance) **operatori di collegamento tra tabelle**, come invece avviene per i database relazionali.

61

Svantaggio

- Controlli di consistenza:
 - Per esempio, se avessimo un database dei clienti coi relativi ordini, in un database non relazionale alla cancellazione di un cliente tutti gli ordini resterebbero comunque nel database, è quindi l'applicativo che una volta impartito il comando di cancellazione dell'utente X deve anche andare a cancellare tutti i relativi ordini, cosa che invece in un database relazionale è gestita direttamente dal DBMS.

62