

Capitolo 2

Il modello relazionale

Copyright © 2004 – The McGraw-Hill Companies srl

Perché studiare il Modello Relazionale?

- È il modello più largamente usato
 - Produttori: IBM, Informix, Microsoft, Oracle, Sybase, etc.
- “Sistemi proprietari” nei modelli più vecchi
 - ad esempio IMS dell’IBM
- Competitori recenti: modello orientato agli oggetti
 - ObjectStore, Versant, Ontos
 - Una sintesi emergente: il modello relazionale ad oggetti
 - Informix Universal Server, UniSQL, O2

Copyright © 2004 – The McGraw-Hill Companies srl

Basi di dati relazionali: definizioni

- Base di dati relazionale: un insieme di *relazioni*
- *Relazione*: consiste di due parti:
 - *istanza*: una *tabella*, con righe e colonne.
righe = *cardinalità*, # campi = *grado/arità*
 - *schema*: specifica il nome della relazione, più il nome e il tipo di ciascuna colonna
 - ad esempio: Studenti(*sid*:string, *nome*:string, *login*:string, *età*:integer, *media*:real)
- possiamo pensare a una relazione come a un *insieme* di righe di *tuple* (cioè tutte le righe sono distinte)

Copyright © 2004 – The McGraw-Hill Companies srl

Relazione: tre accezioni

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

Copyright © 2004 – The McGraw-Hill Companies srl

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 -uple (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 $D_1 \times \dots \times D_n$.
- D_1, \dots, D_n sono i **domini** della relazione

Copyright © 2004 – The McGraw-Hill Companies srl

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

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

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

Il modello è basato su valori

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

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

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
		30	
		24	
		28	
		26	

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

Copyright © 2004 – The McGraw-Hill Companies srl

Struttura basata su valori: vantaggi

- 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

Copyright © 2004 – The McGraw-Hill Companies srl

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:

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

Copyright © 2004 – The McGraw-Hill Companies srl

Definizioni, 2

- Una **ennupla** 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

Copyright © 2004 – The McGraw-Hill Companies srl

Definizioni, 3

- (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)

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

Informazione incompleta: motivazioni

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

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

Informazione incompleta nel modello relazionale

- 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

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

Tipi di valore nullo

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

Copyright © 2004 – The McGraw-Hill Companies srl

Vincoli di integrità

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

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

Vincoli di integrità (VI)

- VI: condizione che deve essere vera per qualunque istanza della base di dati; ad esempio vincoli di dominio
 - I VI sono specificati quando si definisce lo schema
 - I VI sono controllati quando si modificano le relazioni
- Una istanza legale di una relazione è una istanza che soddisfa tutti i VI specificati
 - I DBMS non dovrebbero permettere istanze illegali
- Se il DBMS controlla i VI, i dati memorizzati sono più fedeli al significato nel mondo reale
 - Si evitano anche gli errori di inserimento!

Copyright © 2004 – The McGraw-Hill Companies srl

Vincoli di integrità, perché?

- descrizione più accurata della realtà
- contributo alla "qualità dei dati"
- utili nella progettazione
- usati dai DBMS nella esecuzione delle interrogazioni
- Attenzione: non tutte le proprietà di interesse sono rappresentabili per mezzo di vincoli formulabili in modo esplicito

Copyright © 2004 – The McGraw-Hill Companies srl

Tipi di vincoli

- vincoli **intrarelazionali**
 - vincoli su valori (o di dominio)
 - vincoli di **ennupla**
- vincoli **interrelazionali**

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

Vincoli di ennupla

- Esprimono condizioni sui valori di ciascuna ennupla, indipendentemente dalle altre ennuple
 - Spesso: espressione booleana sui valori di attributo o espressioni aritmetiche su di essi
(Voto = 30) OR NOT (Lode = "e lode")
- Caso particolare: Vincoli di dominio: coinvolgono un solo attributo
(Voto ≥ 18) AND (Voto ≤ 30)

Copyright © 2004 – The McGraw-Hill Companies srl

Identificazione delle ennuple

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	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- non ci sono due ennuple con lo stesso valore sull'attributo Matricola
- non ci sono due ennuple uguali su tutti e tre gli attributi Cognome, Nome e Data di Nascita

Copyright © 2004 – The McGraw-Hill Companies srl

Chiave

- insieme di attributi che identificano le ennuple di una relazione
- Formalmente:
- un insieme K di attributi è **superchiave** per r se r non contiene due ennuple 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)

Copyright © 2004 – The McGraw-Hill Companies srl

Una chiave

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	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

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

Copyright © 2004 – The McGraw-Hill Companies srl

Un'altra chiave

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	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- Cognome, Nome, Nascita è un'altra chiave:
 - è superchiave
 - minimale

Copyright © 2004 – The McGraw-Hill Companies srl

Un'altra chiave??

Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	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	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- Non ci sono ennuple uguali su Cognome e Corso:
 - Cognome e Corso formano una chiave
- Ma è sempre vero?

Copyright © 2004 – The McGraw-Hill Companies srl

Vincoli, schemi e istanze

- i vincoli corrispondono a proprietà del mondo reale modellato dalla base di dati
- interessano a livello di schema (con riferimento cioè a tutte le istanze)
- ad uno schema associamo un insieme di vincoli e consideriamo **corrette** (valide, ammissibili) le istanze che soddisfano tutti i vincoli
- Una particolare istanza può soddisfare altri vincoli ("per caso")

Copyright © 2004 – The McGraw-Hill Companies srl

Studenti

Matricola Cognome Nome Corso Nascita

- chiavi:

Matricola
Cognome, Nome, Nascita

Copyright © 2004 – The McGraw-Hill Companies srl

Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	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	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- È corretta: soddisfa i vincoli
- Ne soddisfa anche altri ("per caso"):
 - Cognome, Corso è chiave

Copyright © 2004 – The McGraw-Hill Companies srl

Esistenza delle chiavi

- Una relazione non può contenere ennuple distinte ma uguali
- Ogni relazione ha come superchiave l'insieme degli attributi su cui è definita
- e quindi ha (almeno) una chiave

Copyright © 2004 – The McGraw-Hill Companies srl

Chiavi primarie e candidate

- Sono possibili molte *chiavi candidate*, una delle quali viene scelta come chiave primaria
- "Per un dato studente e un dato corso, c'è una singola classe" verso "Gli studenti possono seguire solo un corso, e ricevere un singolo voto per quel corso; inoltre, nessuna coppia di studenti in un corso riceve lo stesso voto"
- Usato a sproposito, un VI può impedire la memorizzazione di istanze della base di dati che si presentano nella pratica!

Copyright © 2004 – The McGraw-Hill Companies srl

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:
 - il modello relazionale è basato su valori

Copyright © 2004 – The McGraw-Hill Companies srl

Chiavi e valori nulli

- In presenza di valori nulli, i valori della chiave non permettono
 - di identificare le ennuple
 - di realizzare facilmente i riferimenti da altre relazioni

Copyright © 2004 – The McGraw-Hill Companies srl

Matricola	Cognome	Nome	Corso	Nascita
NULL	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
NULL	Neri	Mario	NULL	5/12/78

- La presenza di valori nulli nelle chiavi deve essere limitata

Copyright © 2004 – The McGraw-Hill Companies srl

Chiave primaria

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

<u>Matricola</u>	<u>Cognome</u>	<u>Nome</u>	<u>Corso</u>	<u>Nascita</u>
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

Copyright © 2004 – The McGraw-Hill Companies srl

Integrità referenziale

- informazioni in relazioni diverse sono correlate attraverso valori comuni
- in particolare, valori delle chiavi (primarie)
- le correlazioni debbono essere "coerenti"

Copyright © 2004 – The McGraw-Hill Companies srl

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

Copyright © 2004 – The McGraw-Hill Companies srl

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

Auto

<u>Prov</u>	<u>Numero</u>	<u>Cognome</u>	<u>Nome</u>
MI	39548K	Rossi	Mario
TO	E39548	Rossi	Mario
PR	839548	Neri	Luca

Copyright © 2004 – The McGraw-Hill Companies srl

Vincolo di integrità referenziale

- Un vincolo di **integrità referenziale** ("foreign key") fra gli attributi X 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
- Potete indicare un modello di dati senza integrità referenziale?
 - Le pagine HTML del Web!

Copyright © 2004 – The McGraw-Hill Companies srl

- vincoli di integrità referenziale fra:
 - l'attributo Vigile della relazione INFRAZIONI e la relazione VIGILI
 - gli attributi Prov e Numero di INFRAZIONI e la relazione AUTO

Copyright © 2004 – The McGraw-Hill Companies srl

Violazione di vincolo di integrità referenziale

Infrazioni

Codice	Data	Vigile	Prov	Numero
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

Auto

Prov	Numero	Cognome	Nome
MI	E39548	Rossi	Mario
TO	F34268	Rossi	Mario
PR	839548	Neri	Luca

Copyright © 2004 – The McGraw-Hill Companies srl

Vincoli di integrità referenziale: commenti

- Giocano un ruolo fondamentale nel concetto di "modello basato su valori"
- In presenza di valori nulli i vincoli possono essere resi meno restrittivi
- Sono possibili meccanismi per il supporto alla loro gestione ("azioni" compensative a seguito di violazioni)
- Attenzione ai vincoli su più attributi

Copyright © 2004 – The McGraw-Hill Companies srl

Integrità referenziale e valori nulli

Impiegati

Matricola	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	XYZ
64521	Verdi	NULL
73032	Bianchi	IDEA

Progetti

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

Copyright © 2004 – The McGraw-Hill Companies srl

Azioni compensative

- Esempio:
 - Viene eliminata una ennupla causando una violazione
- Comportamento "standard":
 - Rifiuto dell'operazione
- Azioni compensative:
 - Eliminazione in cascata
 - Introduzione di valori nulli

Copyright © 2004 – The McGraw-Hill Companies srl

Eliminazione in cascata

Impiegati

Matricola	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	XYZ
64521	Verdi	NULL
73032	Bianchi	IDEA

Progetti

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

Copyright © 2004 – The McGraw-Hill Companies srl

Introduzione di valori nulli

Impiegati

Matricola	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	NULL
64521	Verdi	NULL
73032	Bianchi	IDEA

Progetti

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

Copyright © 2004 – The McGraw-Hill Companies srl

Linguaggi di Interrogazione Relazionali

- Uno dei punti di forza del modello relazionale: supporta un semplice ma potente *linguaggio di interrogazione* dei dati.
- Le interrogazioni possono essere scritte in maniera intuitiva, e il DBMS è responsabile di una valutazione efficiente
 - Fattore chiave: semantica precisa per le interrogazioni relazionali.
 - Permette all'ottimizzatore di riordinare estensivamente le operazioni, garantendo che la risposta non cambi

Copyright © 2004 – The McGraw-Hill Companies srl

Modello relazionale: sommario

- Una rappresentazione tabellare dei dati
- Semplice e intuitivo, attualmente il più usato
- Il DBA può specificare vincoli di integrità basati sulla semantica dell'applicazione. Il DBMS controlla eventuali violazioni
 - Due VI importanti: chiavi primarie e chiavi esterne
 - Inoltre abbiamo *sempre* vincoli di dominio
- Esistono linguaggi di interrogazione potenti e naturali

Copyright © 2004 – The McGraw-Hill Companies srl