
Parte II

Il modello relazionale

Il modello relazionale

- Proposto da E. F. Codd nel 1970 per favorire l'indipendenza dei dati
- Disponibile come modello logico in DBMS reali nel 1981 (non è facile realizzare l'indipendenza con efficienza e affidabilità!)
- Si basa sul concetto di relazione (con qualche modifica)
- Le relazioni hanno una rappresentazione naturale per mezzo di tabelle
- Modello basato su valori: i riferimenti fra dati in strutture (relazioni) diverse sono rappresentati per mezzo dei valori stessi

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; associazione tra dati rappresentabile con una tabella

Relazione matematica

- D_1, D_2, \dots, D_n (n insiemi anche non distinti)
- Il prodotto cartesiano $D_1 \times D_2 \times \dots \times D_n$, è l'insieme di tutte le n -uple ordinate (d_1, d_2, \dots, d_n) tali che $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$
- Una relazione matematica su D_1, D_2, \dots, D_n è un sottoinsieme del prodotto cartesiano $D_1 \times D_2 \times \dots \times D_n$
- D_1, D_2, \dots, D_n sono i domini della relazione. Una relazione su n domini ha grado (o arietà) n
- il numero di n -uple è la 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-uple ordinate:
 (d_1, \dots, d_n) tali che $d_1 \in D_1, \dots, d_n \in D_n$
- La relazione è un *insieme*, quindi:
 - non c'è ordinamento fra le n-uple
 - le n-uple sono distinte
- Ciascuna n-upla è *ordinata*; quindi
 - l' i-esimo d_i valore proviene dall' i-esimo dominio D_n

Struttura posizionale

Partite \subseteq string \times string \times integer \times integer

Juve	Lazio	1	3
Lazio	Milan	2	0
Juve	Roma	4	0
Roma	Milan	0	1

Ciascuno dei domini ha due ruoli distinti, distinguibili attraverso la posizione:

- primo e terzo dominio si riferiscono a nome e reti della squadra ospitante
- secondo e il quarto a nome e reti della squadra ospitata

Relazioni nel modello relazionale

Casa	Fuori	RetiCasa	RetiFuori
Juve	Lazio	1	3
Lazio	Milan	2	0
Juve	Roma	4	0
Roma	Milan	0	1

- Ogni relazione è sostanzialmente una tabella
- A ciascun dominio associamo un nome (attributo), unico nella relazione, che “descrive” il ruolo del dominio
- Nella rappresentazione tabellare, gli attributi sono usati come intestazioni delle colonne
- L'ordinamento fra gli attributi è irrilevante: la struttura è non posizionale

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
- Il modello relazionale è basato su valori: i riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di valori dei domini che compaiono nelle enuple

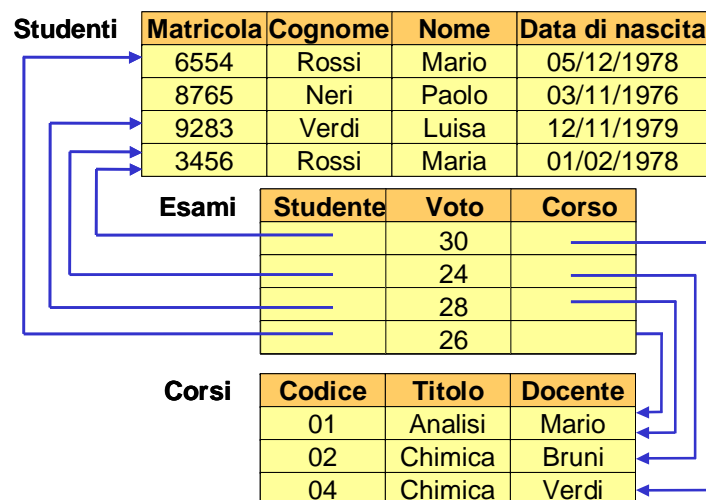
Modello basato sui valori

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

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

Modello basato sui valori



Modello basato sui valori: vantaggi

- Indipendenza dalle strutture fisiche, che possono cambiare dinamicamente
- Si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione
- I puntatori sono poco comprensibili per l'utente finale
- Senza puntatori l'utente finale vede gli stessi dati dei programmatori
- Dati portabili più facilmente da un sistema ad un altro
- I valori consentono bidirezionalità: i puntatori sono direzionali

Schemi

- Schema di relazione: nome di relazione R con un insieme di attributi $X = \{A_1, \dots, A_n\}$

Es. **Studenti(Matricola,Cognome,Nome)**

- Schema di base di dati: insieme di schemi di relazione con nomi diversi $R = \{R_1(X_1), \dots, R_n(X_n)\}$

Es. **Ateneo={Studenti(Matricola,Cognome,Nome),
Professori(Cognome,Nome),
Corsi(CodiceC,Materia,Professore),
Frequenza(Matricola,CodiceC)}**

- Lo schema di una base di dati rappresenta la *struttura delle tabelle* che la compongono, cioè l'organizzazione logica dei dati

Istanze

- (Istanza di) relazione su uno schema $R(X)$: insieme r di ennuple su $X = \{A_1, \dots, A_n\}$
- (Istanza di) base di dati su uno schema $R = \{R_1(X_1), \dots, R_n(X_n)\}$: insieme di istanze di relazioni $r = \{r_1, \dots, r_n\}$ (dove r_i è un'istanza di relazione sullo schema R_i)
- Un'istanza di relazione è rappresentata da una tabella
- Un'istanza di base di dati è rappresentata da un'insieme di tabelle con i valori in esse contenuti

Tuple

- Una relazione r su uno schema $R(X)$ è insieme di n -uple su un insieme di attributi $X = \{A_1, \dots, A_n\}$
- Alle n -uple della relazione corrispondono le righe della tabella
- Le ennuple della relazione vengono dette tuple
- Una n -upla $t \in r$ 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

Istanza di base di dati

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

N.B.: la relazione **Sudenti lavoratori** è su di un singolo attributo

Informazione incompleta

- Il modello relazionale impone ai dati una struttura rigida:
 - le informazioni sono rappresentate per mezzo di ennuple
 - le ennuple ammesse sono dettate dagli schemi di relazione
- In certi casi i dati disponibili possono non corrispondere esattamente al formato previsto
- L'informazione può essere incompleta per vari motivi
- Occorre tuttavia gestire anche queste situazioni

Informazione incompleta: esempi

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

- Churchill aveva un secondo nome ma non lo conosciamo
- De Gaulle non aveva un secondo nome
- Non sappiamo se Stalin avesse un secondo nome

Valori nulli

- Nel modello relazionale per rappresentare l'informazione incompleta si adotta una tecnica rudimentale ma efficace
- Valore nullo: in una tupla denota l'assenza di un valore del dominio (e non è un valore del dominio)
- Formalmente, è sufficiente estendere il concetto di ennupla:
 - $t[A]$, 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

Interpretazione del valore nullo

- Ci sono (almeno) tre casi differenti:
 - Valore sconosciuto: esiste un valore del dominio, ma non è noto (Churchill)
 - Valore inesistente: non esiste un valore del dominio (De Gaulle)
 - Valore senza informazione: non è noto se esista o meno un valore del dominio (Stalin)
- I DBMS non distinguono i tipi di valore nullo (e quindi implicitamente adottano l'interpretazione 'senza informazione')

Istanze scorrette

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

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

Anomalie

- A** - Nella relazione **Studenti** due studenti Neri e Bianchi hanno la stessa matricola: questo nell'applicazione universitaria non deve mai succedere
- B** - Nella relazione **Esami** un esame ha votazione '27 e lode': la lode si può dare solo a chi ha preso trenta
- C** - Nella relazione **Esami** un esame ha votazione 32: il voto deve essere compreso fra 18 e 30
- D** - Nella relazione **Esami** un esame è stato sostenuto da uno studente con matricola 739430 : nella relazione **Studenti** non c'è uno studente con questa matricola

Vincoli di integrità

- Per evitare di avere istanze scorrette si introducono i vincoli di integrità:
 - Proprietà che devono essere soddisfatte dalle istanze che rappresentano informazioni corrette per l'applicazione
 - Ogni vincolo può essere visto come una funzione booleana (o un predicato) che associa ad ogni istanza il valore *vero* o *falso*
- Ad uno schema si associamo un insieme di vincoli
- Si considerano corrette (lecite, valide, ammissibili) solo le istanze che soddisfano tutti i vincoli

Importanza dei vincoli

- Risultano utili al fine di descrivere la realtà di interesse in modo più accurato di quanto le strutture permettano;
- Forniscono un contributo verso la "qualità dei dati"
- Costituiscono uno strumento di ausilio alla progettazione
- Sono utilizzati dal sistema nella scelta della strategia di esecuzione delle interrogazioni

N.B: non tutte le proprietà di interesse sono rappresentabili per mezzo di vincoli esprimibili direttamente

Tipi di vincoli di integrità

- **Vincoli intrarelazionali**: esprimono condizioni sugli attributi di un'unica relazione:
 - **Vincoli di ennupla**: esprimono condizioni sui valori di ciascuna ennupla, indipendentemente dalle altre ennuple
 - **Vincoli di dominio**: vincoli di ennupla che coinvolgono un solo attributo
- **Vincoli interrelazionali**: esprimono condizioni sugli attributi di più relazioni

Vincoli di ennupla: esempi

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

- Nel caso di **Esami** e **Studenti** si possono imporre i seguenti vincoli di dominio e di ennupla:
 - (Voto ≥ 18) **AND** (Voto ≤ 30)
 - (Voto = 30) **OR NOT** (Lode = "e lode")
- Nel caso della relazione **Stipendi**, imponiamo il vincolo:
 - Lordo = (Ritenute + Netto)

Vincoli di 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

- Il numero di matricola identifica gli studenti:
 - non ci devono essere due ennuple con lo stesso valore sull'attributo Matricola
- I dati anagrafici identificano gli studenti:
 - non ci devono essere due ennuple uguali su tutti e tre gli attributi Cognome, Nome e Nascita

Chiavi e superchiavi

- **Chiave**: insieme di attributi che identificano univocamente le n-uple di una relazione
- In una relazione possono esistere più chiavi
- Qualsiasi superset di una chiave è una chiave anch'esso
- Formalmente:
 - un insieme **K** di attributi è **superchiave** per una relazione **r** se **r** non contiene due n-uple distinte **t₁** e **t₂** tali che **t₁[K] = t₂[K]**
 - **K** è **chiave** per **r** se è una superchiave minimale (cioè non contiene un'altra superchiave) per **r**

Vincoli di chiave: esempi

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:
 - Matricola è superchiave
 - contiene un solo attributo e quindi è minimale
- **Cognome, Nome, Nascita** è un'altra chiave: è superchiave e nessuno dei suoi sottoinsiemi è superchiave
- **Cognome, Nome, Nascita, Corso** è superchiave (non chiave)

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

- In questa istanza non ci sono n-uple uguali su **Cognome** e **Corso**
- **Cognome** e **Corso** formano una chiave
- È una decisione sensata scegliere **Cognome** e **Corso** come chiave della relazione?

Vincoli, schemi, 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 che vogliamo siano soddisfatti da tutte le sue istanze corrette
- Consideriamo quindi corrette (valide, ammissibili) le sole istanze che soddisfano tutti i vincoli

N.B.: un'istanza può soddisfare altri vincoli ("per caso")

Esistenza delle chiavi

- Poiché le relazioni sono insiemi, una relazione non può contenere ennuple uguali fra loro:

Ogni relazione ha come superchiave l'insieme degli attributi su cui è definita

- Poiché l'insieme di tutti gli attributi è una superchiave per ogni relazione, ogni schema di relazione ha almeno una superchiave
- Ne segue che: *ogni schema di relazione ha (almeno) una chiave*

Importanza delle chiavi

- L'esistenza delle chiavi garantisce l'accessibilità a ciascun dato della base di dati
- Ogni singolo valore è univocamente accessibile tramite:
 - nome della relazione
 - valore della chiave
 - nome dell'attributo

Le chiavi sono lo strumento principale attraverso il quale vengono correlati i dati in relazioni diverse ("il modello relazionale è basato su valori")

Chiavi e valori nulli

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

In presenza di valori nulli, i valori degli attributi che formano la chiave:

- non permettono di identificare le ennuple come desiderato
- né permettono di realizzare facilmente i riferimenti da altre relazioni

Chiave primaria

<u>Matricola</u>	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	NULL
67653	Rossi	Piero	NULL	5/12/78

- La presenza di valori nulli nelle chiavi deve essere limitata.
- Soluzione pratica: per ogni relazione scegliamo una chiave (la chiave primaria) su cui non ammettiamo valori nulli.
- Notazione per la chiave primaria: gli attributi che la compongono sono sottolineati

Vincoli di integrità referenziale

- Informazioni in relazioni diverse sono correlate attraverso valori comuni, in particolare, attraverso valori delle chiavi (primarie, di solito)
- Un vincolo di integrità referenziale (detto anche vincolo di ("foreign key")) fra un insieme di attributi X di una relazione R_1 e un'altra relazione R_2 impone ai valori su X di ciascuna n-upla dell'istanza di R_1 di comparire come valori della chiave (primaria) dell'istanza di R_2
- Giocano un ruolo fondamentale nel concetto di "modello basato su valori"

Vincoli di integrità referenziale: esempio

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	39548K	Rossi	Mario
	TO	E39548	Rossi	Mario
	PR	839548	Neri	Luca

Vincolo di integrità referenziale fra {Prov,Numero} in **Infrazioni** e la chiave di **Auto**

Vincoli di integrità referenziale: esempio 2

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

Vigili	Matricola	Cognome	Nome
	3987	Rossi	Luca
	3295	Neri	Piero
	9345	Neri	Mario
	7543	Mori	Gino

Vincolo di integrità referenziale fra **Vigile** in **Infrazioni** e la relazione **Vigili**

Violazioni del vincolo di integrità

Infrazioni	Codice	Data	Vigile	Prov	Numero
	34321	1/2/95	3987	MI	39548K
	53524	4/3/95	3295	TO	E39548
	64521	5/4/96	2468	PR	839548
	73321	5/2/98	9345	PR	839548

Vigili	Matricola	Cognome	Nome
	3987	Rossi	Luca
	3295	Neri	Piero
	9345	Neri	Mario
	7543	Mori	Gino

Violazioni del vincolo di integrità 2

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

Auto	Prov	Numero	Cognome	Nome
	MI	39548K	Rossi	Mario
	TO	E39548	Rossi	Mario
	PR	839548	Neri	Luca

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

In presenza di valori nulli i vincoli possono essere resi meno restrittivi

Azioni compensative

- Sono possibili meccanismi per il supporto alla gestione dei vincoli di integrità a seguito di violazioni
- Se viene eliminata una n-upla causando una violazione, il DBMS può reagire in vari modi
- Comportamento "standard": rifiuto dell'operazione
- Possibili anche azioni compensative:
 - Eliminazione in cascata
 - Introduzione di valori nulli

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

Se si elimina il progetto XYZ, questo segna il destino del povero Neri (potrebbe corrispondere ad una strategia aziendale...)

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

Si sostituisce il riferimento al progetto eliminato con il valore nullo, Neri in cassa integrazione

Vincoli multipli su più attributi

Incidenti	<u>Codice</u>	Data	ProvA	NumeroA	ProvB	NumeroB
	34321	1/2/95	TO	E39548	MI	39548K
	64521	5/4/96	PR	839548	TO	E39548

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

- Vincolo tra gli attributi **ProvA** e **NumeroA** di **Incidenti** e la relazione **Auto**
- Vincolo tra gli attributi **ProvB** e **NumeroB** di **Incidenti** e la relazione **Auto**