

Lezione 5



Il Modello dei Dati Relazionale
Vincoli sui Database Relazionali

Sommario

- Concetti del Modello Relazionale
- Vincoli del Modello Relazionale e degli Schemi di Database Relazionali
- Operazioni di aggiornamento e gestione delle violazioni dei vincoli

Modello Relazionale

- Il Modello dei Dati Relazionale è basato sul concetto di Relazione.
- Una **Relazione** è un concetto matematico basato sulle idee degli insiemi.
- La forza dell'approccio relazionale alla gestione dei dati deriva dalle *fondamenta formali* date dalla teoria delle relazioni.
- È quindi importante conoscere a fondo il formalismo relazionale.

Modello Relazionale

- Il modello è stato proposto da E. F. Codd (IBM) nel 1970 nel famoso articolo:
A Relational Model for Large Shared Data Banks,
Communications of the ACM, June 1970.

Questo articolo ha causato una sorta di rivoluzione nel campo dei database spingendo la transizione dai modelli gerarchico e reticolare al modello relazionale utilizzato anche oggi.

DEFINIZIONI INFORMALI

- **RELAZIONE:** Una tabella di valori
 - Una relazione può essere pensata come un insieme di **righe**.
 - Una relazione può essere anche pensata come un insieme di **colonne**.
 - Ciascuna *riga* rappresenta un fatto che corrisponde ad una *istanza* di un'*entità* o di una *associazione* del mini-mondo
 - Ciascuna *riga* possiede un valore per un *elemento* od un *insieme di elementi* che la compongono, che **identifica univocamente** quella riga nella tabella.
 - A volte vengono assegnati degli *identificatori di riga* (row-id) o dei *numeri sequenziali* per identificare le righe nella tabella.
 - Ciascuna colonna viene tipicamente individuata mediante il *nome* o l'*intestazione della colonna* o il *nome dell'attributo*.

DEFINIZIONI FORMALI

- Una relazione può essere definita in molti modi.
- Lo **Schema di Relazione**: $R (A_1, A_2, \dots, A_n)$
 - Lo *Schema di Relazione* R è definito sugli *Attributi* A_1, A_2, \dots, A_n .
- Esempio:
 - CLIENTI (ID-Cli, Nome-Cli, Indirizzo, Tel)
 - CLIENTI è una relazione definita sui quattro attributi ID-Cli, Nome-Cli, Indirizzo e Tel, ciascuno dei quali possiede un **dominio**, o insieme di valori validi.
 - Ad esempio, ID-Cli è un numero di 6 cifre.

DEFINIZIONI FORMALI

- Una **tupla** è un *insieme ordinato* di valori.
- Ciascun valore è derivato da un **dominio** appropriato.
- Ciascuna riga nella tabella CLIENTI può essere vista come una tupla della relazione, composta da quattro valori.
- <632895, “Luca Tomassetti”, “via Paradiso 12, 44100, Ferrara”, “+39 (0532) 974307”>
è una tupla appartenente alla relazione CLIENTI.
- Una relazione può essere trattata come un insieme di tuple (righe).
- Le colonne della tabella sono chiamate **attributi** della relazione.

DEFINIZIONI FORMALI

- Un **dominio** ha una *definizione logica*: ad esempio “Numeri di Telefono” è l’insieme dei numeri telefonici validi, comprensivi del Country Code, del prefisso e del numero di lunghezza massima 14 cifre.
- Un dominio può avere un *tipo di dati* o un *formato* definito. Il dominio “Numeri di Telefono” può avere il formato: +dd (dddd) dddddddd dove d è una cifra decimale. Esempio: le date hanno diversi formati, come giorno, mese, anno oppure dd-mm-yy, dd/mm/yyyy, ...
- Un attributo specifica il *ruolo* ricoperto da un dominio. Ad esempio il dominio date può essere usato per definire gli attributi “DataFattura” e “DataPagamento”.

DEFINIZIONI FORMALI

- Una relazione (stato della relazione) è formata sul prodotto cartesiano degli insiemi; ciascun insieme ha valori provenienti da un dominio; il dominio è usato per specificare il ruolo dell'attributo considerato.
- Ad esempio, l'attributo Nome-Cli è definito sul dominio delle stringhe di 25 caratteri. Il ruolo di queste stringhe nella relazione CLIENTI è quello di specificare il nome dei clienti.
- Formalmente,
 - Dato $R(A_1, A_2, \dots, A_n)$
 - $r(R) \subset \text{dom}(A_1) \times \text{dom}(A_2) \times \dots \times \text{dom}(A_n)$
 - R : schema della relazione
 - r di R : uno specifico “valore” o popolazione di R . Chiamato **stato della relazione**
 - R viene anche chiamato intensione della relazione
 - r viene anche chiamato estensione della relazione

DEFINIZIONI FORMALI

- Siano: $S_1 = \{o, I\}$ e $S_2 = \{a, b, c\}$
- Definiamo $R \subset S_1 \times S_2$
- Allora, ad esempio:
$$r(R) = \{ \langle o, a \rangle, \langle o, b \rangle, \langle I, c \rangle \}$$

è un possibile stato (o popolazione, o
intensione) r della relazione R , definita sui
domini S_1 ed S_2 . Contiene 3 tuple.

SOMMARIO delle DEFINIZIONI

Termini Informali	Termini Formali
Tabella	Relazione
Colonna	Attributo/Dominio
Riga	Tupla
Valori in una colonna	Dominio
Definizione di Tabella	Schema di Relazione
Tabella Popolata	Estensione (Stato)

ESEMPIO

The diagram illustrates the components of a database table. A label 'Relation name' points to the first column header 'STUDENT'. A label 'Attributes' points to the remaining column headers: 'Name', 'SSN', 'HomePhone', 'Address', 'OfficePhone', 'Age', and 'GPA'. A label 'Tuples' points to the five rows of data in the table.

STUDENT	Name	SSN	HomePhone	Address	OfficePhone	Age	GPA
	Benjamin Bayer	305-61-2435	373-1616	2918 Bluebonnet Lane	null	19	3.21
	Katherine Ashly	381-62-1245	375-4409	125 Kirby Road	null	18	2.89
	Dick Davidson	422-11-2320	null	3452 Elgin Road	749-1253	25	3.53
	Charles Cooper	489-22-1100	376-9821	265 Lark Lane	749-6492	28	3.93
	Barbara Benson	533-69-1238	839-8461	7384 Fontana Lane	null	19	3.25

CARATTERISTICHE DELLE RELAZIONI

- Ordinamento delle tuple in una relazione $r(R)$: Le tuple non devono considerarsi ordinate, anche se appaiono in forma tabulare.
- Ordinamento degli attributi in uno schema di relazione R (e dei valori all'interno di ciascuna tupla t): Gli attributi $R(A_1, A_2, \dots, A_n)$ ed i valori in $t = \langle v_1, v_2, \dots, v_n \rangle$ devono essere considerati ordinati.
- In ogni caso, una definizione ancora più generale di relazione non richiede alcun tipo di ordinamento.
- Valori in una tupla: Tutti i valori sono considerati atomici (indivisibili). Un valore speciale NULL viene usato per rappresentare valori sconosciuti o non applicabili ad alcune tuple.

CARATTERISTICHE DELLE RELAZIONI

- NOTAZIONE
- Ci riferiamo ai valori presenti in una tupla t con $t[A_i] = v_i$ (il valore dell'attributo A_i per la tupla t)
- Analogamente, $t[A_u, A_v, \dots, A_w]$ si riferisce alla sotto-tupla di t che contiene i valori degli attributi A_u, A_v, \dots, A_w , rispettivamente.

CARATTERISTICHE DELLE RELAZIONI

STUDENT	Name	SSN	HomePhone	Address	OfficePhone	Age	GPA
	Dick Davidson	422-11-2320	null	3452 Elgin Road	749-1253	25	3.53
	Barbara Benson	533-69-1238	839-8461	7384 Fontana Lane	null	19	3.25
	Charles Cooper	489-22-1100	376-9821	265 Lark Lane	749-6492	28	3.93
	Katherine Ashly	381-62-1245	375-4409	125 Kirby Road	null	18	2.89
	Benjamin Bayer	305-61-2435	373-1616	2918 Bluebonnet Lane	null	19	3.21

Vincoli di Integrità

del Modello Relazionale

- I vincoli sono condizioni che devono essere rispettate da tutte gli stati di relazione **validi**.
- Ci sono tre tipi principali di vincoli:
 1. Vincoli sulla Chiave
 2. Vincoli di Integrità dell'Entità
 3. Vincoli di Integrità Referenziale

Vincoli sulla Chiave

- **Superchiave** di R : Un insieme di attributi SK di R tale che non esistano due tuple che abbiano lo stesso valore per SK in nessuno stato di relazione valido $r(R)$; cioè, per ogni coppia di tuple distinte t_1 e t_2 in $r(R)$, $t_1[SK] \neq t_2[SK]$.
- **Chiave** di R : Una superchiave minimale; cioè, una superchiave K tale che la rimozione di qualsiasi attributo da K risulti in un insieme di attributi che non è più una superchiave.
 - Esempio: Lo schema di relazione AUTO
AUTO(ProvT, NumT, NumTelaio, Casa, Modello, Anno)
ha due chiavi $PK_1 = \{ProvT, NumT\}$, $PK_2 = \{NumTelaio\}$, che sono anche superchiavi. $\{NumTelaio, Casa\}$ è superchiave ma non chiave.
- Se una relazione ha diversi candidati a chiave, ne viene scelta una arbitrariamente come **chiave primaria**.

Integrità sulle Entità

- **Schema di database relazionale:** Un insieme S di schemi di relazione che appartengono allo stesso database. S è il nome del database.

$$S = \{R_1, R_2, \dots, R_n\}$$

- **Integrità sulle Entità:** Gli attributi chiave primaria PK di ciascuno schema di relazione R in S non possono avere valori *null* in nessuna tupla di $r(R)$. Questo perché i valori delle chiavi primarie sono usati per identificare le varie tuple.

$$t[PK] \neq null \quad \forall \text{ tupla } t \text{ in } r(R)$$

- Nota: Altri attributi di R possono avere il vincolo di non permettere valori *null*, anche se non sono membri di chiavi primarie.

Integrità Referenziale

- Un vincolo che coinvolge due relazioni (i vincoli visti precedentemente riguardavano una singola relazione).
- Usato per specificare un riferimento tra tuple in due relazioni: la relazione referenziante e la relazione riferita.
- Tuple nella relazione referenziante R_I hanno attributi FK (chiamati attributi chiave esterne) che fanno riferimento agli attributi chiave primarie PK della relazione riferita R_2 . Una tupla t_I in R_I è detta in riferimento (in relazione, relazionata) con una tupla t_2 in R_2 se $t_I[FK] = t_2[PK]$.
- Un vincolo di integrità referenziale può essere indicato in uno schema relazionale con un arco diretto da $R_I.FK$ a $R_2.PK$.

Vincolo di Integrità Referenziale

Definizione del vincolo

- Il valore della colonna (o delle colonne) chiave esterna FK della relazione referenziante R_I può essere:
 1. un valore **uguale** al valore di una chiave primaria esistente nella corrispondente chiave primaria PK della relazione riferita R_2 , oppure
 2. un valore **null**.
- Nel caso 2., la FK in R_I non deve far parte della chiave primaria di R_I .

Altri tipi di vincoli

- **Vincoli di Integrità Semantici:**
 - basati sulla semantica dell'applicazione.
 - non possono essere espressi dal modello dei dati svincolato dall'applicazione/dati
- Esempio: “il numero massimo di ore, per dipendente, per tutti i progetti su cui lavora, deve essere 56 ore settimanali”
- Un apposito linguaggio di specifica dei vincoli può essere usato per esprimere questo tipo di vincoli.
 - Il linguaggio SQL-99 consente l'uso di **triggers** e **asserzioni** a questo scopo
- Molto spesso vengono espressi nel contesto dell'applicazione

Figure 7.5 Schema diagram for the COMPANY relational database schema; the primary keys are underlined.

EMPLOYEE

FNAME	MINIT	LNAME	<u>SSN</u>	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
-------	-------	-------	------------	-------	---------	-----	--------	----------	-----

DEPARTMENT

DNAME	<u>DNUMBER</u>	MGRSSN	MGRSTARTDATE
-------	----------------	--------	--------------

DEPT_LOCATIONS

<u>DNUMBER</u>	<u>DLOCATION</u>
----------------	------------------

PROJECT

PNAME	<u>PNUMBER</u>	PLOCATION	DNUM
-------	----------------	-----------	------

WORKS_ON

<u>ESSN</u>	<u>PNO</u>	HOURS
-------------	------------	-------

DEPENDENT

<u>ESSN</u>	<u>DEPENDENT_NAME</u>	SEX	BDATE	RELATIONSHIP
-------------	-----------------------	-----	-------	--------------

Figure 7.6 One possible relational database state corresponding to the COMPANY schema.

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
John			Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin			Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia			Zakaya	999887777	1968-01-19	3321 Casallo, Spring, TX	F	25000	987654321	4
Jennifer			Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh			Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce			English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad			Jabbar	987987987	1969-03-29	989 Dallas, Houston, TX	M	25000	987654321	4
James			Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

DEPARTMENT	DNAME	DNUMBER	MGRSSN	MGRSTARTDATE
	Research	5	333445555	1988-05-22
	Administration	4	987654321	1965-01-01
	Headquarters	1	888665555	1961-06-19

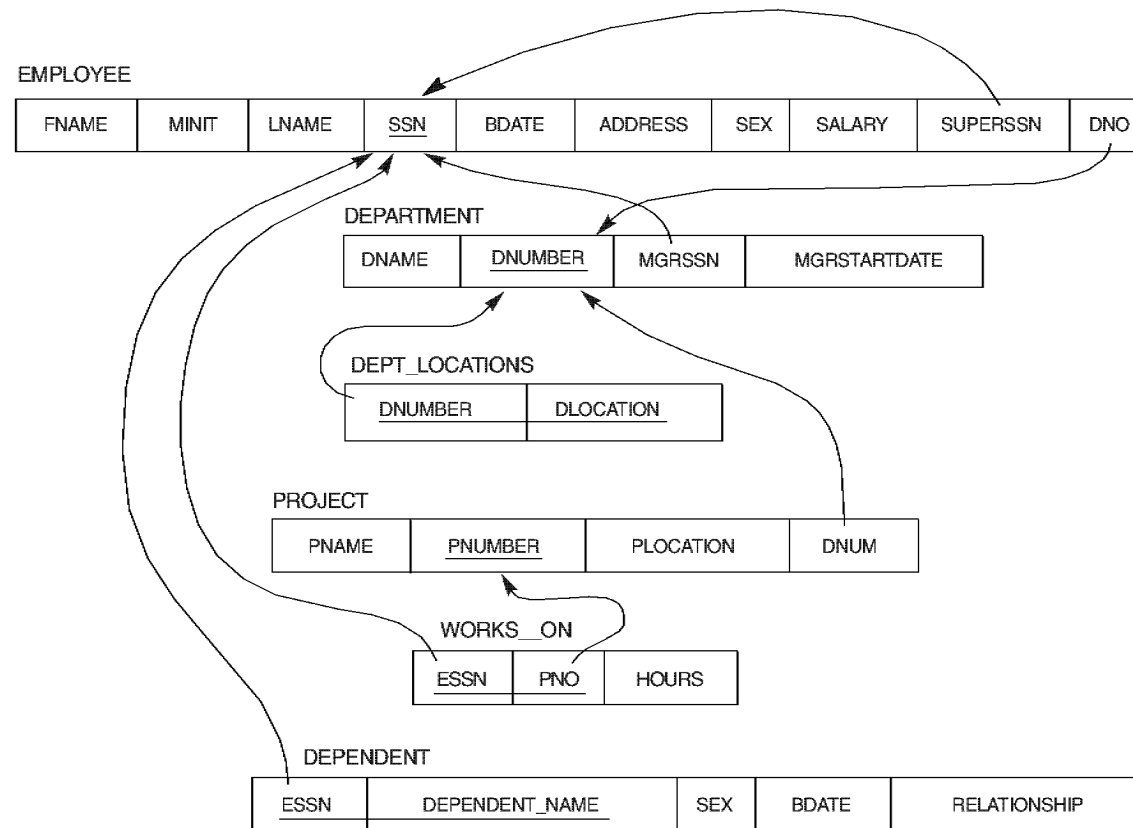
DEPT_LOCATIONS	DNUMBER	DLOCATION
		Houston
		Stafford
		Bellaire
		Sugarland

WORKS_ON	ESSN	PNO	HOURS
	123456789	1	32.5
	123456789	2	7.5
	666884444	3	40.0
	453453453	1	20.0
	453453453	2	20.0
	333445555	2	10.0
	333445555	3	10.0
	333445555	10	10.0
	333445555	20	10.0
	999887777	30	30.0
	999887777	10	10.0
	987987987	10	35.0
	987987987	30	5.0
	987654321	30	20.0
	987654321	20	15.0
	888665555	20	null

PROJECT	PNAME	PNUMBER	PLOCATION	DNUM
	ProductX	1	Bellaire	5
	ProductY	2	Sugarland	5
	ProductZ	3	Houston	5
	Computerization	10	Stafford	4
	Reorganization	20	Houston	1
	Newbenefits	30	Stafford	4

DEPENDENT	ESSN	DEPENDENT_NAME	SEX	BDATE	RELATIONSHIP
	333445555	Alice	F	1986-04-05	DAUGHTER
	333445555	Theodore	M	1983-10-25	SON
	333445555	Joy	F	1958-05-03	SPOUSE
	987654321	Abner	M	1942-02-28	SPOUSE
	123456789	Michael	M	1988-01-04	SON
	123456789	Alice	F	1988-12-30	DAUGHTER
	123456789	Elizabeth	F	1987-05-05	SPOUSE

Figure 7.7 Referential integrity constraints displayed on the COMPANY relational database schema diagram.



Operazioni di aggiornamento sulle relazioni

- **Inserimento** di una tupla (INSERT)
- **Cancellazione** di una tupla (DELETE)
- **Modifica** di una tupla (MODIFY, UPDATE)
- I vincoli di integrità non devono essere violati dalle operazioni di aggiornamento delle relazioni.
- Una serie di operazioni di aggiornamento possono essere raggruppate insieme.
- Le operazioni di aggiornamento possono “propagarsi” e causare altri aggiornamenti in modo automatico. Ciò può essere necessario per mantenere i vincoli di integrità.

Operazioni di aggiornamento sulle relazioni

- Nel caso un'operazione di aggiornamento di una relazione violi un vincolo di integrità, si possono intraprendere alcune azioni:
 - Annullare l'operazione che causa la violazione (opzione REJECT)
 - Eseguire l'operazione, informando l'utente della violazione
 - Far partire ulteriori aggiornamenti in modo da correggere la violazione (opzione CASCADE, opzione SET NULL)
 - Eseguire una routine specificata dall'utente per correggere la violazione