

Il modello relazionale

TERMINOLOGY

DOMAINS

RELATION

PRIMARY KEY

ATTRIBUTES

TUPLES

DEGREE (= 5)

Dario Maio

<http://bias.csr.unibo.it/maio/>

Il modello relazionale

1

Introduzione al modello relazionale

- Il modello relazionale fu introdotto nel 1970 da **E.F. Codd** (presso i laboratori IBM di San Jose, CA) allo scopo di favorire l'indipendenza dei dati.
- I modelli preesistenti (**gerarchico** e **reticolare**) erano fortemente influenzati da considerazioni di natura fisica, che enfatizzavano maggiormente gli aspetti di efficienza piuttosto che la semplicità d'uso.
- Rispetto ai modelli reticolare e gerarchico, il modello relazionale si caratterizza per:
 - ✚ la totale assenza di legami costruiti con puntatori:
 - ✚ **gerarchico** e **reticolare** usano **puntatori**
 - ✚ nel modello **relazionale** si fa solo uso di **valori**
 - ✚ la **presenza di una teoria** utile per la progettazione di DB, per la definizione di linguaggi e per l'ottimizzazione delle query.

Il modello relazionale

2

Un po' di storia...

Anni 70: definizione del modello, prima versione del linguaggio SQL (allora SEQUEL), studi fondamentali sulla tecnologia relazionale (ottimizzazione, transazioni, recovery, ...) e primi prototipi di DBMS relazionali (RDBMS):

- » System R (IBM, laboratorio di ricerca di San Jose, CA, USA)
- » Ingres (Università di Berkeley, CA, USA)

Anni 80: prima standardizzazione di SQL, primi prototipi commerciali:

- » SQL/DS (derivato da System R)
- » Oracle
- » IBM DB2

Anni 90: standard ISO-ANSI SQL-2 (anche noto come SQL-92).

✚ Esiste oggi lo standard ISO-ANSI SQL-3 (o SQL:1999) e sono state definite ulteriori estensioni (SQL 2003, SQL 2008,...).

Sul termine "Relazione"

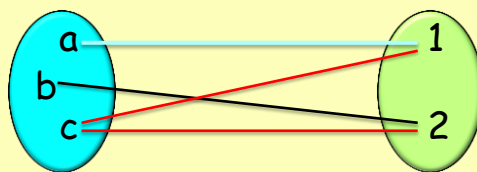
- Il termine "relazione" può essere usato con diverse accezioni, che non vanno confuse tra loro:
 - o nel **linguaggio comune** denota un "**legame**" di qualche tipo;
 - o nella **teoria degli insiemi** denota una "**relazione matematica**";
 - o nel **modello relazionale** è una **generalizzazione della relazione matematica**;
 - o nel **modello Entity-Relationship** denota una **classe di legami fra entità** (sono sinonimi "**associazione**" e "**correlazione**");
 - o nei **DBMS** relazionali è usato spesso, a volte erroneamente, come sinonimo di "**tabella**".
- Per introdurre il modello relazionale è quindi opportuno innanzitutto riesaminare il concetto di relazione matematica.



Relazione matematica (binaria)

- Si considerino due insiemi A e B , non vuoti e non necessariamente distinti; ogni sottoinsieme non vuoto del prodotto cartesiano $A \times B$ è detto **relazione da A a B** . Se $B=A$ allora un sottoinsieme non vuoto del prodotto cartesiano A^2 è detto anche **relazione in A o su A** .
- Data una relazione $r \subseteq A \times B$, si dice che l'elemento $a \in A$ è **in relazione** con l'elemento $b \in B$ se la coppia $(a,b) \in r$.

Esempio: $A = \{a,b,c\}$, $B = \{1,2\}$;
 $A \times B = \{(a,1), (a,2), (b,1), (b,2), (c,1), (c,2)\}$
 $r = \{(a,1), (b,2), (c,1), (c,2)\}$ è una relazione su A e B



Il modello relazionale



5

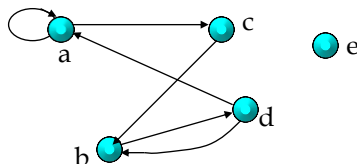


Un esempio d'uso

- Un **grafo orientato** è una coppia $(A;r)$ dove A è un insieme finito e non vuoto e r è una relazione in A .

Esempio: $A = \{a,b,c,d,e\}$;

$r = \{(a,a), (a,c), (b,d), (c,b), (d,a), (d,b)\}$



Il modello relazionale



6



Relazione matematica n-aria

- Si considerino $n > 0$ insiemi D_1, D_2, \dots, D_n , non necessariamente distinti.
- Il **prodotto Cartesiano** $D_1 \times D_2 \times \dots \times D_n$ è l'insieme di tutte le n-ple 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 qualunque **sottoinsieme del prodotto Cartesiano** $D_1 \times D_2 \times \dots \times D_n$.

Esempio:

$D_1 = \{\text{mela}, \text{pera}\}, D_2 = \{1, 2, 3\}, D_3 = \{1, 2, 3, 4\};$

$r = \{(\text{mela}, 1, 1), (\text{mela}, 3, 1), (\text{pera}, 1, 3), (\text{pera}, 2, 2), (\text{pera}, 3, 4)\}$

è una relazione su D_1, D_2, D_3 ($r \subseteq D_1 \times D_2 \times D_3$)

- o D_1, D_2, \dots, D_n sono i **domini** della relazione;
- o il **valore di n** è detto **grado** (o "arità") della relazione;
- o il **numero di n-ple** di una relazione è la sua **cardinalità**.

Il modello relazionale



7



Relazione matematica: proprietà

- ✚ Una relazione è un insieme di n-ple...:
 - tutte le n-ple sono distinte tra loro;
 - non è definito alcun ordinamento tra le diverse n-ple;
 - $D_1 = \{a, b, c\}; D_2 = \{1, 2, 3\};$
 - $\{(a, 1), (b, 2), (c, 1), (c, 2)\} = \{(a, 1), (b, 2), (c, 2), (c, 1)\} = \{(b, 2), (c, 2), (c, 1), (a, 1)\}$
- ✚ ... ciascuna considerata al proprio interno ordinata rispetto ai domini ...:
 - l'ordine in cui si considerano i domini è rilevante ;
 - $(D_1 \times D_2 \neq D_2 \times D_1)$
 - $\{(a, 1), (c, 1), (c, 2)\} \neq \{(1, a), (1, c), (2, c)\}$
- ✚ ... su domini non necessariamente distinti:
 - uno stesso dominio può essere usato in più posizioni;
 - $\{(2, a, 1), (1, c, 1), (1, c, 2)\} \subseteq D_2 \times D_1 \times D_2$.

Il modello relazionale



8



Relazione matematica: note

- Una relazione è un insieme di n-ple...: la definizione data contempla anche il caso di $n=1$, sottintendendo che il prodotto cartesiano coincida con l'insieme D_1 (anche se formalmente non è corretto in quanto il prodotto cartesiano prevede almeno due operandi).

Esempio

$D_1 = \{\text{Carlo}, \text{Mario}, \text{Giacomo}, \text{Marco}, \text{Giorgio}\};$
 $r = \{(\text{Carlo}), (\text{Giorgio})\};$ è dunque una relazione !

- La definizione contempla anche relazioni con un numero infinito di n-ple; ai fini pratici (a causa della dimensione finita della memoria di un elaboratore), le relazioni sono necessariamente costituite da un numero finito di n-ple. Tuttavia è possibile a volte "gestire" anche relazioni con numero infinito di n-ple, se descrivibili attraverso un algoritmo finito.
- La definizione contempla domini infiniti, e ciò è utile ai fini pratici per definire valori ammissibili anche se non presenti nella base di dati.

Il modello relazionale



9



Rappresentazione di relazioni

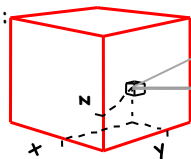
- rappresentazione insiemistica**; non adeguata per relazioni di grado $n \geq 2$ e/o con cardinalità superiore a qualche unità.

- rappresentazione tabellare**; molto efficace e intuitiva.

mela	1	1
mela	3	1
pera	1	3
pera	2	2
pera	3	4

- rappresentazione multi-dimensionale**;

es. se il grado n è minore o uguale a 3:



1 se la tripla (x, y, z) appartiene alla relazione,
0 se non vi appartiene

Il modello relazionale



10

Note sulla rappresentazione multi-dimensionale

- Una rappresentazione multi-dimensionale può essere utile per evidenziare alcune viste sui dati:

Il modello relazionale

11

E ancora altre rappresentazioni...

- Ad esempio: una **bit map**

Il modello relazionale

12



L'importanza della posizione

- Nel caso di domini ripetuti, l'interpretazione dei dati si complica e la **posizione** assume un ruolo determinante.

$\text{partite} \subseteq \text{String} \times \text{String} \times \text{Integer} \times \text{Integer}$

Benetton TV	Poliform Cantù	100	71
Kinder BO	MontePaschi SI	90	51
Paf BO	Adr RM	62	97
Adr RM	Kinder BO	80	62

- Il primo e il terzo dominio si riferiscono alla squadra ospitante (nome e numero di punti), mentre il secondo e il quarto dominio si riferiscono alla squadra ospitata.

È SCOMODO E POCO CHIARO!!

Il modello relazionale



13



Relazione nel modello relazionale

- A ogni occorrenza di dominio (ripetuto o meno) si associa un **nome univoco nella relazione**, detto **attributo**, il cui compito è specificare il **ruolo** che quel dominio svolge nella relazione ("cosa significa").
- Nella **rappresentazione tabellare**, gli attributi sono le **intestazioni delle colonne** (e in quella **multi-dimensionale** sono i **nomi degli assi**).

TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
Benetton TV	Poliform Cantù	100	71
Kinder BO	MontePaschi SI	90	51
Paf BO	Adr RM	62	97
Adr RM	Kinder BO	80	62

- La struttura non è più posizionale, ovvero l'ordine degli attributi non ha più rilevanza!

Il modello relazionale



14



Relazione: una definizione formale

- ✦ Si indichi con $\text{dom}(A)$ il dominio dell'attributo A e si consideri un insieme di attributi $X = \{A_1, A_2, \dots, A_n\}$.
- ✦ Una **tupla t su X** è una **funzione** che associa a ogni $A_i \in X$ un valore di $\text{dom}(A_i)$.
- ✦ (L'istanza di) **una relazione su X** è un **insieme di tuple su X** .
- ✦ Lo **schema di una relazione su X** è dato da un **nome** (della relazione) R e dall'**insieme di attributi X** , scritto $R(X)$.

Partite

TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
Benetton TV	Poliform Cantù	100	71
Kinder BO	MontePaschi SI	90	51
Paf BO	Adr RM	62	97
Adr RM	Kinder BO	80	62

Il modello relazionale



15



Notazione di base (1)

- Per denotare insiemi di attributi si usa la notazione semplificata:
 - A in luogo di $\{A\}$ e XY in luogo di $X \cup Y$
 - ...e si scrive ABC (o A,B,C) anziché $\{A,B,C\}$
 - ...e quindi $R(ABC)$ o $R(A,B,C)$ anziché $R(\{A, B, C\})$
- Se t è una tupla su X e $A \in X$, allora $t[A]$ o $t.A$ è il valore di t su A

Partite

	TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
	Benetton TV	Poliform Cantù	100	71
t	Kinder BO	MontePaschi SI	90	51
	Paf BO	Adr RM	62	97
	Adr RM	Kinder BO	80	62

$t[\text{TeamOspite}] = t.\text{TeamOspite} = \text{'MontePaschi SI'}$

- La stessa notazione si usa per insiemi di attributi, e denota una tupla
 - $t[\text{TeamOspite}, \text{PuntiOspite}]$ è una tupla su $\{\text{TeamOspite}, \text{PuntiOspite}\}$

Il modello relazionale



16

Notazione di base (2)

- Se necessario, per riferirsi all'istanza della relazione con schema $R(X)$ si usa r (il nome in minuscolo della relazione)

Partite(TeamCasa, TeamOspite, PuntiCasa, PuntiOspite)

partite =

Benetton TV	Poliform Cantù	100	71
Kinder BO	MontePaschi SI	90	51
Paf BO	Adr RM	62	97
Adr RM	Kinder BO	80	62

Il modello relazionale

17

Data Base relazionale

- ✚ Lo **schema** di un DB relazionale è un **insieme di schemi di relazioni con nomi distinti**:

$$\text{✚ } R = \{R_1(X_1), R_2(X_2), \dots, R_m(X_m)\} \quad (R_i \neq R_j \quad \forall i \neq j)$$

- ✚ (L'**istanza** di) un DB con schema $R = \{R_1(X_1), R_2(X_2), \dots, R_m(X_m)\}$ è un **insieme di (istanze) di relazioni**

$$\text{✚ } r = \{r_1, r_2, \dots, r_m\}$$

con r_i istanza su $R_i(X_i)$

Esempio: DATI_AZIENDA =

{ **IMPIEGATI**(Matricola, Cognome, Nome, Livello, Stipendio),
FILIALI(CodFiliale, Nome, Indirizzo, Direttore),
FORNITORI(RagSoc, Indirizzo, PartitaIVA) }

Il modello relazionale

18

Un semplice DB relazionale

Studenti

Matricola	Cognome	Nome	DataNascita
29323	Bianchi	Giorgio	21/06/1978
35467	Rossi	Anna	13/04/1978
39654	Verdi	Marco	20/09/1979
42132	Neri	Lucia	15/02/1978

Corsi

CodCorso	Titolo	CodDocente	Anno
483	Analisi	0201	1
729	Analisi	0021	1
913	Sistemi Informativi	0123	2

Esami

Matricola	CodCorso	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	no
35467	913	30	sì

Docenti

CodDocente	Cognome	Nome	DataNascita
0021	Biondi	Carlo	21/06/1958
*****	*****	*****	*****

Il modello relazionale

19

Tabelle vs Relazioni (1)

- In realtà i termini “**tabella**” e “**relazione**” non sono affatto sinonimi; una relazione del modello relazionale può essere vista come un particolare tipo di tabella (che Codd chiama **R-table**).
- Una tabella rappresenta una relazione se:
 - ✚ i valori di ciascuna colonna sono tra loro omogenei (definiti sullo stesso dominio);
 - ✚ le righe sono tra loro diverse;
 - ✚ le intestazioni delle colonne sono diverse tra loro.
- In una tabella che rappresenta una relazione:
 - ✚ l'ordinamento delle righe è irrilevante;
 - ✚ l'ordinamento delle colonne è irrilevante.
- SQL nei DBMS commerciali consente di gestire tabelle che non sono relazioni, e che ammettono righe duplicate.

Il modello relazionale

20



Tabelle vs Relazioni (2)

- Esempio di derivazione in SQL di una tabella da una relazione

Studenti

Matricola	Cognome	Nome	DataNascita
2106103423	Bianchi	Giorgio	21/06/1978
2106111021	Rossi	Anna	13/04/1978
1602042312	Rossi	Anna	11/03/1978

- **SELECT** Cognome, Nome

FROM Studenti ;

una tabella che
non è una relazione

Cognome	Nome
Bianchi	Giorgio
Rossi	Anna
Rossi	Anna

- **SELECT DISTINCT** Cognome, Nome

FROM Studenti ;

una tabella che
è una relazione

Cognome	Nome
Bianchi	Giorgio
Rossi	Anna

Il modello relazionale

21



Vantaggi di un modello basato su valori

- Indipendenza dalle strutture fisiche che possono cambiare anche dinamicamente.
- Si rappresenta solo ciò che risulta rilevante dal punto di vista dell'applicazione utente; l'uso di puntatori non è molto comprensibile all'utente finale.
- Maggiore portabilità dei dati da un sistema all'altro.
- I puntatori sono direzionali.
- Si noti che a livello fisico la realizzazione di un insieme di relazioni può prevedere l'uso di puntatori, invisibili all'utente applicativo.
- I riferimenti tra dati in relazioni diverse sono rappresentati per mezzo dei valori dei domini che compaiono nelle tuple.

Il modello relazionale

22

1NF, ovvero solo domini semplici

- Il modello relazionale non permette di usare domini arbitrari per la definizione delle relazioni; in particolare **non è in generale possibile usare domini strutturati** (array, set, liste, ...).
 - ✦ Vi sono **eccezioni** notevoli (ad es. le **date** e le **stringhe**).
- Concisamente, **una relazione in cui ogni dominio è "atomico"** (non ulteriormente decomponibile) si dice che è in **Prima Forma Normale**, o **1NF** (First Normal Form).
- In molti casi è pertanto richiesta un'attività di **normalizzazione dei dati** che dia luogo a relazioni in 1NF e che preservi l'informazione originale.

Strutture nidificate: normalizzazione

Ricevuta n. 231 del 12/02/2002		
Coperti	2	3,00
Antipasti	1	5,80
Primi	2	11,45
Secondi	2	22,30
Caffè	2	2,20
Vino	1	8,00
Totale (Euro)		52,75

Ricevute

Numero	Data	Totale
231	12/02/2002	52,75
352	13/02/2002	...
...

Dettaglio

Numero	Quantità	Descrizione	Prezzo
231	2	Coperti	3,00
231	1	Antipasti	5,80
231	2	Primi	11,45
231	2	Secondi	22,30
231	2	Caffè	2,20
231	1	Vino	8,00
352	1	Coperti	1,50



Considerazioni sulla normalizzazione in 1NF

- ✚ Il fatto che una rappresentazione normalizzata sia adeguata o meno dipende (molto) dal contesto.
 - ✚ Ad es.: l'ordine delle righe nella ricevuta è rilevante o meno?
- ✚ Analogamente per eventuali ridondanze che si possono osservare.
 - ✚ Ad es.: il coperto e il caffè hanno un prezzo che non varia da ricevuta a ricevuta?
- ✚ In generale è bene ricordare che ogni caso presenta una sua specificità, e quindi non deve essere trattato "automaticamente"
- ✚ Normalizzare in 1NF è, a tutti gli effetti, un'attività di progettazione (logica), e in quanto tale può essere solo oggetto di "regole guida" che però non hanno validità assoluta



Relazione matematica versus Relazione nel modello relazionale

Relazione matematica	Relazione Modello Rel.
Domini arbitrari	Domini atomici
Colonne senza nome	Colonne con nomi
Colonne distinte in base alla posizione	Nomi univoci per le colonne
Di solito costante nel tempo	Di solito variabile nel tempo



Informazione incompleta

- Le informazioni che si vogliono rappresentare mediante relazioni non sempre corrispondono pienamente allo schema prescelto, in particolare per alcune tuple e alcuni attributi potrebbe non essere possibile specificare, per diversi motivi, un valore del dominio.

Persona	DataMorte
Mario Rossi	20/02/1954
Paolo Verdi	
Bruno Bianchi	
Carlo Grigi	

- ✚ Paolo Verdi è ancora vivo (**valore non applicabile**);
- ✚ Bruno Bianchi è deceduto, ma non conosciamo la data di morte (**applicabile ma ignoto**);
- ✚ Carlo Grigi è scomparso misteriosamente, non sappiamo se è vivo o se è deceduto (**ignota l'applicabilità**).

Il modello relazionale



27



Quale soluzione?

- In diversi casi, in mancanza di informazione, si tende a usare un "**valore speciale**" del dominio (0, "", "-1", "9999", ecc.) che non si utilizza per altri scopi.
- **Questa pratica è fortemente sconsigliata**, in quanto, anche dove possibile:
 - ✚ valori inutilizzati potrebbero successivamente diventare significativi;
 - ✚ le applicazioni dovrebbero sapere "che cosa significa in realtà" il valore usato allo scopo.
- **Esempio (reale!)**: nel 1998, analizzando i clienti di un'assicurazione, si scoprì una strana concentrazione di ultra-novantenni... tutte le date di nascita ignote erano state codificate con "01/01/00"!!
- Nel modello relazionale si opera in maniera pragmatica: si adotta il concetto di **valore nullo** (NULL), che denota assenza di un valore nel dominio (e non è un valore del dominio);
- ...pertanto $t[A] \in \text{dom}(A) \cup \{\text{NULL}\}$.

Il modello relazionale



28



Valori nulli: considerazioni

Persona	DataMorte
Mario Rossi	20/02/1954
Paolo Verdi	NULL
Bruno Bianchi	NULL
Carlo Grigi	NULL

- La presenza di un valore nullo non fornisce alcuna informazione sull'applicabilità o meno.
- È importante ricordare che NULL non è un valore del dominio; in particolare, se due tuple hanno entrambe valore NULL per un attributo, non si può inferire che esse abbiano lo stesso valore per quell'attributo, ovvero:
 $\text{NULL} \neq \text{NULL}$
- NB: tuttavia, ai fini della verifica di assenza di tuple duplicate è opportuno che i NULL siano considerati come gli altri valori e quindi uguali tra loro ($\text{NULL} = \text{NULL}$).

Il modello relazionale



29



Valori nulli: restrizioni

- La presenza di valori nulli non può essere sempre tollerata, ovvero è necessario imporre delle **restrizioni** al loro uso; si consideri ad esempio il caso della registrazione di esami:

Esami	Matricola	CodCorso	Voto	Lode
	29323	483	28	no
	NULL	729	30	sì
	29323	913	NULL	no
	35467	913	30	no

- Un valore nullo per Matricola non permetterebbe di sapere quale studente ha sostenuto l'esame.
- Un valore nullo per Voto non è proprio ammissibile nel contesto considerato.



Istanze di questo tipo non sono accettabili!

Il modello relazionale



30



Vincoli di integrità

- La "correttezza sintattica" di un'istanza non è condizione sufficiente affinché i dati rappresentino un'informazione possibile nel contesto reale considerato.

Studenti	Matricola	Cognome	Nome	DataNascita
	35467	Bianchi	Giorgio	21/06/1978
	35467	Rossi	Anna	13/04/1978
	39654	Rossi	Anna	13/04/1978

- ✚ La prima e la seconda tupla hanno la stessa Matricola?
- ✚ La seconda e la terza tupla hanno gli stessi valori per Nome, Cognome e DataNascita, ma questo in linea di principio è possibile!
- Un **vincolo di integrità** è una **proprietà che deve essere soddisfatta da ogni possibile istanza**; ogni vincolo può quindi essere descritto da una funzione booleana che associa a ogni istanza il valore VERO o FALSO.

Il modello relazionale



31



Vincoli di dominio

- Un vincolo che si riferisce ai valori ammissibili per un singolo attributo viene detto **vincolo di dominio** (o **sui valori**)

Esami	Matricola	CodCorso	Voto	Lode
	29323	483	28	no
	39654	729	30	sì
	29323	913	31	no
	35467	913	30	forse

- Il Voto deve essere compreso tra 18 e 30 :
 - ✚ $(Voto \geq 18) \text{ AND } (Voto \leq 30)$
- La Lode può solo assumere i valori 'sì' o 'no' :
 - ✚ $(Lode = 'sì') \text{ OR } (Lode = 'no')$

Il modello relazionale



32



Vincoli di tupla

- I vincoli di dominio sono un caso particolare dei **vincoli di tupla**, ovvero **vincoli che esprimono condizioni su ciascuna tupla, indipendentemente dalle altre**

Esami

Matricola	CodCorso	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	sì
35467	913	30	no

- La Lode si può assegnare solo se il Voto è 30:

\oplus (Voto = 30) OR NOT(Lode = 'sì')

- Nello schema

Pagamenti(Data,ImportoLordo,Ritenute,Netto) si ha:

\oplus ImportoLordo = Netto + Ritenute

Il modello relazionale



33



Vincoli di chiave: intuizione

- I **vincoli di chiave**, che giocano un ruolo molto importante, vietano la presenza di tuple distinte che hanno lo stesso valore su uno o più attributi.

Studenti

Matricola	CodiceFiscale	Cognome	Nome	DataNascita
210629323	BNCGR678L21A944K	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944V	Rossi	Anna	13/01/1978
160239654	VRDMRC79H20F839X	Verdi	Marco	20/06/1979
214842132	VRDMRC79H20G125J	Verdi	Marco	20/06/1979

- \oplus Il valore di Matricola **identifica univocamente** uno studente;
- \oplus analogamente il CodiceFiscale;
- \oplus ...e ogni insieme di attributi che includa Matricola o CodiceFiscale
 - \oplus {Matricola,Cognome}, {CodiceFiscale,Nome}, ...;
- \oplus viceversa, possono esistere due tuple uguali su {Cognome,Nome,DataNascita}.

Il modello relazionale



34



Chiavi e superchiavi

■ Dato uno schema $R(X)$, un insieme di attributi $K \subseteq X$ è:

✚ una **superchiave** se e solo se

in ogni istanza ammissibile r di $R(X)$ non esistono due tuple distinte $t1$ e $t2$ tali che $t1[K] = t2[K]$;

✚ una **chiave** se e solo se

è una superchiave minimale, ovvero non esiste $K' \subset K$ con K' superchiave.

■ Una chiave è pertanto un identificatore minimale per ogni r su $R(X)$.

■ Nella relazione Studenti:

✚ $\{\text{Matricola}\}$ e $\{\text{CodiceFiscale}\}$ sono due chiavi;

✚ $\{\text{Matricola}, \text{Cognome}\}$ e $\{\text{CodiceFiscale}, \text{Nome}\}$ sono solo superchiavi;

✚ $\{\text{Cognome}, \text{Nome}, \text{DataNascita}\}$ non è superchiave.

Il modello relazionale



35



Esistenza di chiavi e superchiavi

✚ Poiché ogni istanza r su $R(X)$ è un insieme, ne consegue che

✚ l'insieme X di tutti gli attributi dello schema è senz'altro una superchiave per $R(X)$.

✚ Poiché il numero di attributi, n , è finito, è sempre possibile arrivare a individuare (almeno) una chiave $K \subseteq X$

$K = X$;

for $i = 1$ to n

{ if $K - \{A_i\}$ è superchiave

then $K = K - \{A_i\}$;

}



In casi (molto) particolari il numero di chiavi può essere esponenziale in n .

Il modello relazionale



36

Dai vincoli alle istanze, non viceversa!

- I vincoli di chiave si esprimono a livello di schema, sulla base di un'analisi della realtà che si vuole modellare mediante relazioni, **e limitano l'insieme di istanze legali** (o "ammissibili", "corrette" "valide", ecc.).
- Una specifica istanza può soddisfare altri vincoli (di chiave), ma ciò non autorizza a generalizzare.

Esami

Matricola	CodCorso	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	no
35467	913	30	sì

- ✦ La (**unica**) chiave è {Matricola,CodiceCorso}.
- ✦ Questa particolare istanza soddisfa anche altri vincoli, ad es. {Matricola,Voto} è un identificatore, **ma ciò è puramente casuale**.

Il modello relazionale

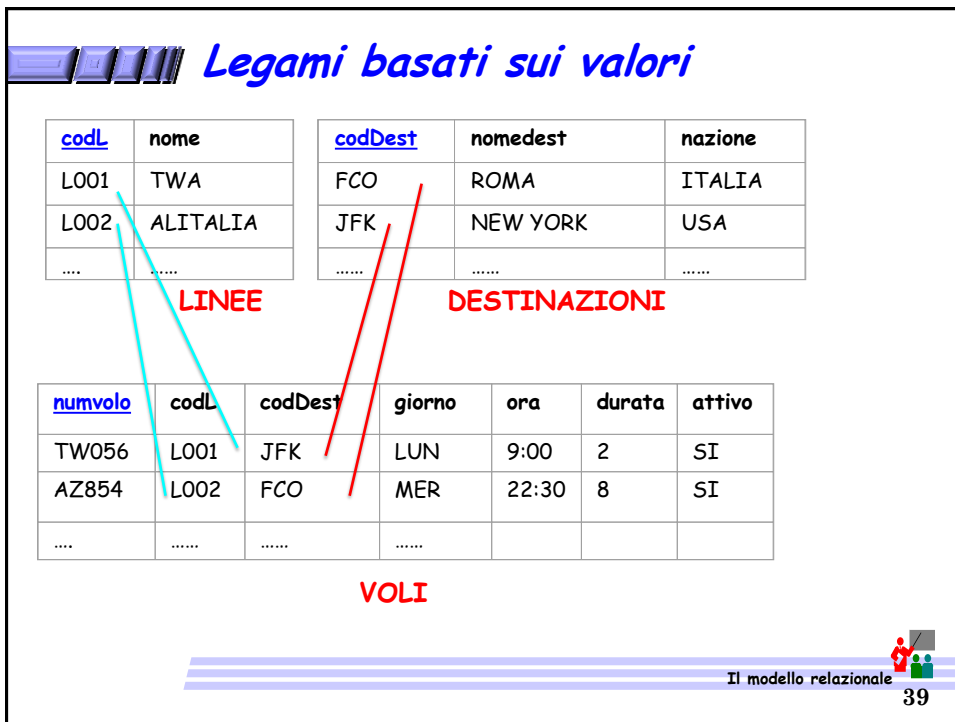
37

Importanza delle chiavi

- L'esistenza delle chiavi garantisce l'**accessibilità a ciascun dato del DB**, in quanto ogni singolo valore è univocamente individuato da:
 - ✦ nome della relazione → individua una relazione del DB
 - ✦ valore della chiave → individua una tupla della relazione
 - ✦ nome dell'attributo → individua il valore desiderato
- Le chiavi sono lo strumento principale attraverso il quale vengono **correlati i dati in relazioni diverse** ("il **modello relazionale è basato su valori**").

Il modello relazionale

38



39

Chiavi e valori nulli

- In presenza di valori nulli entrambe le due funzioni svolte dalle chiavi (identificazione e correlazione) possono venire a mancare.

Studenti

Matricola	CodiceFiscale	Cognome	Nome	DataNascita
NULL	NULL	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944V	Rossi	Anna	13/01/1978
NULL	VRDMRC79H20F839X	Verdi	Marco	20/06/1979
214842132	NULL	Verdi	Marco	20/06/1979

- La prima tupla non è identificabile in alcun modo, pertanto:
 - ✚ è necessario specificare il valore di almeno una chiave!
- La terza e quarta tupla potrebbero riferirsi allo stesso studente, pertanto:
 - ✚ non è sufficiente specificare il valore di una chiave!

Il modello relazionale

40



Chiave primaria

- Per evitare i problemi visti è necessario scegliere una chiave, detta **chiave primaria**, su cui non si ammettono valori nulli.
- Convenzionalmente si **sottolineano** gli attributi che costituiscono la chiave primaria.

Studenti

Matricola	CodiceFiscale	Cognome	Nome	DataNascita
210629323	NULL	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944V	Rossi	Anna	13/01/1978
160239654	VRDMRC79H20F839X	Verdi	Marco	20/06/1979
214842132	NULL	Verdi	Marco	20/06/1979



Nei casi in cui per nessuna chiave si possa garantire la disponibilità di valori, è necessario introdurre un nuovo attributo (un "codice") che svolga le funzioni di chiave primaria; si pensi ad esempio al caso in cui non si riesca a identificare un paziente al pronto soccorso ospedaliero.

Il modello relazionale



41



Vincoli di integrità referenziale

- ✚ I vincoli sinora visti sono tutti di tipo **intra-relazionale**, in quanto interessano una relazione alla volta.
- ✚ Viceversa, i **vincoli di integrità referenziale** sono importanti tipi di vincoli **inter-relazionali** che enfatizzano come le correlazioni tra le tuple siano spesso ottenute usando i valori delle chiavi.
- ✚ Si considerino due schemi $R_1(X_1)$ e $R_2(X_2)$ di un DB R , e sia Y un insieme di attributi in X_2 .
- ✚ Un **vincolo di integrità referenziale su Y** impone che in ogni istanza $r = \{r_1, r_2, \dots\}$ del DB
l'insieme dei valori di Y in r_2 sia un sottoinsieme dell'insieme dei valori della chiave primaria di $R_1(X_1)$ presenti nell'istanza r_1 .
- ✚ L'insieme Y viene detto una **foreign key** (o "chiave importata").

Il modello relazionale



42

Esempi di foreign key

Studenti

Matricola	Cognome	Nome	DataNascita
29323	Bianchi	Giorgio	21/06/1978
35467	Rossi	Anna	13/04/1978
39654	Verdi	Marco	20/09/1979
42132	Neri	Lucia	15/02/1978

Corsi

CodCorso	Titolo	CodDocente	Anno
483	Analisi	0201	1
729	Analisi	0021	1
913	Sistemi Informativi	0123	2

Esami

Matricola	CodCorso	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	no
35467	913	30	sì

In Corsi, {CodDocente} è una foreign key.
 In Esami, {Matricola} è una foreign key, così come {CodCorso}.

Il modello relazionale 43

Foreign key: precisazioni (1)

- In generale la foreign key Y e la primary key K di $R_1(X_1)$ possono includere attributi con nomi diversi:

Corsi

Codice	Titolo	CodDocente	Anno
483	Analisi I	0201	1
729	Analisi I	0021	1

Esami

NumMatricola	CodCorso	Voto	Lode
29323	483	28	no

- Foreign key e primary key possono far parte della stessa relazione, ovviamente con $Y \neq K$.

Personale

Codice	Cognome	...	CodResponsabile
123	Rossi	...	325
134	Verdi	...	325
325	Neri

Il modello relazionale 44



Foreign key: precisazioni (2)

- In presenza di **valori nulli**, i vincoli di integrità referenziale si possono parzialmente rilassare

Personale

Codice	Nome	...	CodResponsabile
123	Mario Rossi	...	325
134	Gino Verdi	...	325
325	Anna Neri	...	NULL

- Nei DBMS un vincolo di integrità referenziale può anche esprimersi con riferimento a una generica chiave (quindi anche non primaria)

Studenti

Matricola	CodiceFiscale	Cognome	Nome	DataNascita
29323	BNCGRG78L21A944V	Bianchi	Giorgio	21/07/1978
35467	RSSFLV78M53G125J	Rossi	Flavia	13/08/1978

Redditi

CF	Imponibile
BNCGRG78L21A944V	45300

Il modello relazionale



45



Sommario

- Il **modello relazionale** è basato sul concetto di **relazione**, che estende quello di relazione matematica tra n domini associando a ciascuna occorrenza di dominio un nome, detto **attributo**.
- Lo **schema** di una relazione consiste di un **nome** e di un **insieme di attributi**; l'**istanza** di una relazione è un **insieme di tuple**, ovvero funzioni che associano a ogni attributo dello schema un valore del corrispondente dominio.
- In **assenza di informazioni** si fa uso di un particolare valore, detto **valore nullo** (NULL), che non appartiene a nessun dominio.
- Per garantire l'**integrità dei dati** si possono specificare diversi tipi di **vincoli**, che definiscono quali sono le **istanze legali** (ammissibili).
- I **vincoli intra-relazionali** includono quelli sui **domini**, sulle **tuple** e i vincoli **di chiave**; i **vincoli inter-relazionali** quelli di **integrità referenziale**. Questi ultimi permettono di stabilire le principali correlazioni tra i dati di diverse relazioni.

Il modello relazionale



46