

La Normalizzazione

Atzeni, Ceri, Paraboschi, Torlone

“Basi Di Dati:
Modelli e Linguaggi di
interrogazione”

McGraw-Hill Italia

Capitolo 10

La normalizzazione

- La teoria della normalizzazione fu introdotta con riferimento alle sole relazioni.
- Rappresentava, di fatto, una tecnica di progetto per DB relazionali.
- Ragionando con le tre fasi (concettuale, logica e fisica) la normalizzazione rappresenta oggi un tecnica di verifica che si applica alle relazioni o, meglio ancora, alle entità ed alle associazioni.

La normalizzazione (cont.)

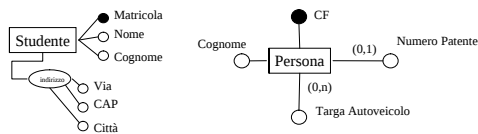
- La normalizzazione cerca di rispondere, in maniera precisa, a domande sulla qualità del nostro progetto:
 - E' facile da illustrare?
 - Ci sono anomalie quando si inseriscono, cancellano, modificano tuple in qualche relazione?
 - I valori nulli sono un caso eccezionale?
 - I JOIN lavorano su condizioni di eguaglianza su chiavi primarie o esterne?
- Essa, inoltre, fornisce metodi per migliorare, iterativamente, la qualità stessa del nostro progetto o del nostro Data Base.

Anomalie

- Esempio
 - IP (Matricola, Progetto, Stipendio, Budget, Funzione)
- La presenza di ridondanze genera anomalie di
 - Inserimento
 - Aggiornamento
 - Cancellazione

Prima Forma Normale

- Ha solo un significato storico.
- Il dominio degli attributi deve comprendere solo valori atomici. Vanno cioè evitati:
 - attributi multi-valore;
 - attributi composti.



Definizioni Preliminari

- Sia R una relazione con chiave primaria X
- Ogni attributo A appartenente allo schema di R si dice
 - Primo se fa parte di X;
 - Non-primo se non appartiene a X.

Persona(nome, cognome, data-nascita, via, CAP, comune, provincia)

- nome, cognome, e data-nascita sono primi
- via, CAP, comune e provincia sono non-primi

Definizioni preliminari

- **Dipendenza Funzionale**
 - Sia R una relazione su $A_1, A_2, \dots, A_n$
 - Siano X e Y due sottoinsiemi di attributi.
 - Y dipende funzionalmente da X ($X \rightarrow Y$) sse:
 - Per ogni coppia di tuple t1 e t2 di R, $t1[X] = t2[X] \Rightarrow t1[Y] = t2[Y]$
 - Il vincolo deve valere per ogni possibile istanza (rappresenta cioè una caratteristica semantica dei dati).

Dipendenze Funzionali (continua)

- $X \rightarrow Y$ non implica (ovviamente) $Y \rightarrow X$
 - Dipendenze banali:
 - $Y \subseteq X \Rightarrow$ banalmente $X \rightarrow Y$
- Il concetto di dipendenza funzionale generalizza il concetto di chiave
 - Due tuple uguali sulla chiave sono, di fatto, la stessa tupla.
- $X \rightarrow Y$ è una dipendenza funzionale piena sse:
 - Rimuovendo un qualsiasi attributo da X essa non vale più.

Seconda Forma Normale

- Una relazione R si dice in seconda forma normale sse:
 - Ogni attributo non primo è in dipendenza funzionale piena dalla chiave primaria.

Forma Normale di Boyce and Codd

- Una relazione si dice in forma normale di Boyce and Codd se ogni dipendenza non banale ha a sinistra una superchiave.

Corsi_Seguiti		
<u>Imp</u>	<u>Corso</u>	Argomento
1	A	Basi Dati
2	A	Basi Dati
3	B	OOP
4	C	SI
5	B	OOP

Seguiti		
<u>Imp</u>	<u>Corso</u>	
1	A	
2	A	
3	B	
4	C	
5	B	

Corsi		
<u>Corso</u>		Argomento
A		Basi Dati
B		OOP
C		SI

Proprietà delle Decomposizioni

- Una decomposizione deve essere senza perdita (lossless join).
- Una decomposizione deve mantenere le dipendenze funzionali

Corsi_Seguiti		
<u>Imp</u>	<u>Corso</u>	Argomento
1	A	Basi Dati
2	A	Basi Dati
3	B	OOP
4	C	SI
5	B	OOP

Seguiti		
<u>Imp</u>	<u>Corso</u>	
1	A	
2	A	
3	B	
4	C	
5	B	

Corsi		
<u>Corso</u>		Argomento
A		Basi Dati
B		OOP
C		SI

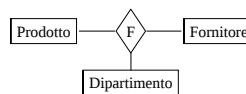
Decomposizione senza Perdita

- Sia R una relazione su X
- Siano $X1 \subseteq X$ e $X2 \subseteq X$
- Sia $X0 = X1 \cap X2$
- Condizione sufficiente affinché R si decomponga senza perdite su $R1(X1)$ ed $R2(X2)$ è che
 - $X0$ sia chiave per $R1$ o per $R2$

Conservazione delle Dipendenze

- In genere noi vogliamo che le decomposizioni conservino le dipendenze individuate:
 - Per ogni $X \rightarrow Y$, deve esistere una relazione che comprende XUY.
- Tale esigenza, apparentemente strana, è dettata dal fatto che le dipendenze rappresentano dei vincoli ed il mantenerle consente di schematizzarle come vincoli intrarelazionali.

- La forma normale di Boyce and Codd non sempre è raggiungibile con decomposizioni che siano, allo stesso tempo, lossless e che preservino le dipendenze.



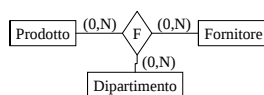
$F(\text{Prodotto}, \text{Fornitore}, \text{Dipartimento})$

Se ogni Dipartimento acquista un certo Prodotto sempre dallo stesso Fornitore:

$\text{Dipartimento}, \text{Prodotto} \rightarrow \text{Fornitore}$

Terza Forma Normale

- Una relazione R è in Terza forma normale sse, per ogni $X \rightarrow Y$, accade che:
 - X è una superchiave, oppure
 - Y è primo
- Essa è sempre raggiungibile.



$F(\text{Prodotto}, \text{Fornitore}, \text{Dipartimento})$

Se ogni Dipartimento acquista un certo Prodotto sempre dallo stesso Fornitore:

$\text{Dipartimento}, \text{Prodotto} \rightarrow \text{Fornitore}$