

Atzeni, Ceri, Paraboschi, Torlone
Basi di dati
McGraw-Hill, 1996-2002

Capitolo 9:
La normalizzazione

23/10/2002

Forme normali

- Una forma normale è una proprietà di una base di dati relazionale che ne garantisce la "qualità", cioè l'assenza di determinati difetti
- Quando una relazione non è normalizzata:
 - presenta ridondanze,
 - si presta a comportamenti poco desiderabili durante gli aggiornamenti
- Le forme normali sono di solito definite sul modello relazionale, ma hanno senso in altri contesti, ad esempio il modello E-R

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

2

Normalizzazione

- Procedura che permette di trasformare schemi non normalizzati in schemi che soddisfano una forma normale
- La normalizzazione va utilizzata come tecnica di verifica dei risultati della progettazione di una base di dati
- Non costituisce una metodologia di progettazione

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

3

Una relazione con anomalie

Impiegato	Stipendio	Progetto	Bilancio	Funzione
Rossi	20	Marte	2	tecnico
Verdi	35	Giove	15	progettista
Verdi	35	Venere	15	progettista
Neri	55	Venere	15	direttore
Neri	55	Giove	15	consulente
Neri	55	Marte	2	consulente
Mori	48	Marte	2	direttore
Mori	48	Venere	15	progettista
Bianchi	48	Venere	15	progettista
Bianchi	48	Giove	15	direttore

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

4

Anomalie

- Lo stipendio di ciascun impiegato è ripetuto in tutte le ennuple relative
 - ridondanza
- Se lo stipendio di un impiegato varia, è necessario andarne a modificare il valore in diverse ennuple
 - anomalia di aggiornamento
- Se un impiegato interrompe la partecipazione a tutti i progetti, dobbiamo cancellarlo
 - anomalia di cancellazione
- Un nuovo impiegato senza progetto non può essere inserito
 - anomalia di inserimento

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

5

Perché questi fenomeni indesiderabili?

- abbiamo usato un'unica relazione per rappresentare informazioni eterogenee
 - gli impiegati con i relativi stipendi
 - i progetti con i relativi bilanci
 - le partecipazioni degli impiegati ai progetti con le relative funzioni

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

6

Per studiare in maniera sistematica questi aspetti, è necessario introdurre un vincolo di integrità:
la dipendenza funzionale

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

7

Proprietà

- Ogni impiegato ha un solo stipendio (anche se partecipa a più progetti)
- Ogni progetto ha un bilancio
- Ogni impiegato in ciascun progetto ha una sola funzione (anche se può avere funzioni diverse in progetti diversi)

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

8

Dipendenza funzionale

- relazione r su $R(X)$
- due sottoinsiemi non vuoti Y e Z di X
- esiste in r una dipendenza funzionale (FD) da Y a Z se, per ogni coppia di tuple t_1 e t_2 di r con gli stessi valori su Y , risulta che t_1 e t_2 hanno gli stessi valori anche su Z

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

9

Notazione

$X \text{®} Y$

- Esempi:

Impiegato ® Stipendio
Progetto ® Bilancio
Impiegato Progetto ® Funzione

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

10

Altre FD

- Impiegato Progetto ® Progetto
- Si tratta però di una FD "banale" (sempre soddisfatta)
- $Y \text{®} A$ è non banale se A non appartiene a Y
- $Y \text{®} Z$ è non banale se nessun attributo in Z appartiene a Y

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

11

Le anomalie sono legate ad alcune FD

- gli impiegati hanno un unico stipendio
Impiegato ® Stipendio
- i progetti hanno un unico bilancio
Progetto ® Bilancio

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

12

Non tutte le FD causano anomalie

- In ciascun progetto, un impiegato svolge una sola funzione

Impiegato Progetto ® Funzione

- Il soddisfacimento è più "semplice "

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

13

Una differenza fra FD

Impiegato ® Stipendio

Progetto ® Bilancio

- causano anomalie

Impiegato Progetto ® Funzione

- non causa anomalie
- Perché?

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

14

Impiegato	Stipendio	Progetto	Bilancio	Funzione
Rossi	20	Marte	2	tecnico
Verdi	35	Giove	15	progettista
Verdi	35	Venere	15	progettista
Neri	55	Venere	15	direttore
Neri	55	Giove	15	consulente
Neri	55	Marte	2	consulente
Mori	48	Marte	2	direttore
Mori	48	Venere	15	progettista
Bianchi	48	Venere	15	progettista
Bianchi	48	Giove	15	direttore

Impiegato ® Stipendio

Progetto ® Bilancio

Impiegato Progetto ® Funzione

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

15

FD e anomalie

- La terza FD corrisponde ad una chiave e non causa anomalie
- Le prime due FD non corrispondono a chiavi e causano anomalie
- La relazione contiene alcune informazioni legate alla chiave e altre ad attributi che non formano una chiave

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

16

- abbiamo usato un'unica relazione per rappresentare informazioni eterogenee
 - gli impiegati con i relativi stipendi
 - i progetti con i relativi bilanci
 - le partecipazioni degli impiegati ai progetti con le relative funzioni

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

17

Impiegato ® Stipendio

Progetto ® Bilancio

Impiegato Progetto ® Funzione

- Impiegato Progetto è chiave
- Impiegato solo no
- Progetto solo no
- Le anomalie sono causate dalla presenza di concetti eterogenei:
 - proprietà degli impiegati (lo stipendio)
 - proprietà di progetti (il bilancio)
 - proprietà della chiave Impiegato Progetto

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

18

Forma normale di Boyce e Codd (BCNF)

- Una relazione r è in forma normale di Boyce e Codd se, per ogni dipendenza funzionale (non banale) $X \twoheadrightarrow Y$ definita su di essa, X contiene una chiave K di r
- La forma normale richiede che i concetti in una relazione siano omogenei (solo proprietà direttamente associate alla chiave)

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

19

Che facciamo se una relazione non soddisfa la BCNF?

- La rimpiazziamo con altre relazioni che soddisfano la BCNF

Come?

- Decomponendo sulla base delle dipendenze funzionali, al fine di separare i concetti

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

20

Impiegato	Stipendio	Progetto	Bilancio	Funzione
Rossi				
Rossi	20	Marte		tecnico
Verdi	35	Giove		progettista
Verdi		Venere		progettista
Neri	55	Neri		direttore
Mori	48	Giove		consulente
Bianchi	48	Neri		consulente
		Marte		direttore
Mori		Mori		progettista
Mori		Bianchi		progettista
Bianchi		Bianchi		progettista
Bianchi		Giove	15	progettista
				direttore

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

21

Non sempre così facile

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato \twoheadrightarrow Sede
Progetto \twoheadrightarrow Sede

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

22

Decomponiamo sulla base delle dipendenze

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

23

Proviamo a ricostruire

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Verdi	Saturno	Milano
Neri	Giove	Milano

Diversa dalla relazione di partenza!

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

24

Decomposizione senza perdita

- Una relazione r si decompone senza perdita su X_1 e X_2 se il join delle proiezioni di r su X_1 e X_2 è uguale a r stessa (cioè non contiene ennuple spurie)
- La decomposizione senza perdita è garantita se gli attributi comuni contengono una chiave per almeno una delle relazioni decomposte

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

25

Proviamo a decomporre senza perdita

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

Impiegato \bowtie Sede
Progetto \bowtie Sede

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

26

Un altro problema

- Supponiamo di voler inserire una nuova ennupla che specifica la partecipazione dell'impiegato Neri, che opera a Milano, al progetto Marte

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

Impiegato \bowtie Sede
Progetto \bowtie Sede

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

27

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere
Neri	Marte

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

28

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Neri	Marte	Milano

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

29

Conservazione delle dipendenze

- Una decomposizione conserva le dipendenze se ciascuna delle dipendenze funzionali dello schema originario coinvolge attributi che compaiono tutti insieme in uno degli schemi decomposti
- Progetto \bowtie Sede non è conservata

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

30

Qualità delle decomposizioni

- Una decomposizione dovrebbe sempre soddisfare:
 - la decomposizione senza perdita, che garantisce la ricostruzione delle informazioni originarie
 - la conservazione delle dipendenze, che garantisce il mantenimento dei vincoli di integrità originari

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

31

Una relazione non-normalizzata

Dirigente	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Marte	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Progetto Sede ® Dirigente
Dirigente ® Sede

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

32

La decomposizione è problematica

- Progetto Sede ® Dirigente coinvolge tutti gli attributi e quindi nessuna decomposizione può preservare tale dipendenza
- quindi in alcuni casi la BCNF “non è raggiungibile”

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

33

Una nuova forma normale

- Una relazione r è in terza forma normale se, per ogni FD (non banale) $X \text{ ® } Y$ definita su r , è verificata almeno una delle seguenti condizioni:
 - X contiene una chiave K di r
 - ogni attributo in Y è contenuto in almeno una chiave di r

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

34

BCNF e terza forma normale

- la terza forma normale è meno restrittiva della forma normale di Boyce e Codd (e ammette relazioni con alcune anomalie)
- ha il vantaggio però di essere sempre “raggiungibile”

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

35

Decomposizione in terza forma normale

- si crea una relazione per ogni gruppo di attributi coinvolti in una dipendenza funzionale
- si verifica che alla fine una relazione contenga una chiave della relazione originaria
- Dipende dalle dipendenze individuate

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

36

Una possibile strategia

- se la relazione non è normalizzata si decompone in terza forma normale
- alla fine si verifica se lo schema ottenuto è anche in BCNF
- Se una relazione ha una sola chiave allora le due forme normali coincidono

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

37

Uno schema non decomponibile in BCNF

Dirigente	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Marte	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Dirigente ® Sede
Progetto Sede ® Dirigente

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

38

Una possibile riorganizzazione

Dirigente	Progetto	Sede	Reparto
Rossi	Marte	Roma	1
Verdi	Giove	Milano	1
Verdi	Marte	Milano	1
Neri	Saturno	Milano	2
Neri	Venere	Milano	2

Dirigente ® Sede Reparto
Sede Reparto ® Dirigente
Progetto Sede ® Reparto

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

39

Decomposizione in BCNF

Dirigente	Sede	Reparto
Rossi	Roma	1
Verdi	Milano	1
Neri	Milano	2

Progetto	Sede	Reparto
Marte	Roma	1
Giove	Milano	1
Marte	Milano	1
Saturno	Milano	2
Venere	Milano	2

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

40

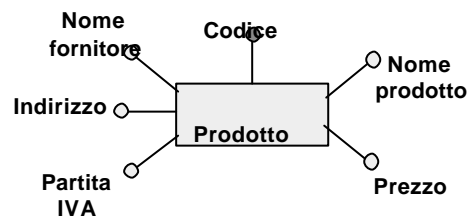
Progettazione e normalizzazione

- la teoria della normalizzazione può essere usata nella progettazione logica per verificare lo schema relazionale finale
- si può usare anche durante la progettazione concettuale per verificare la qualità dello schema concettuale

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

41



PartitaIVA ® NomeFornitore Indirizzo

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

42

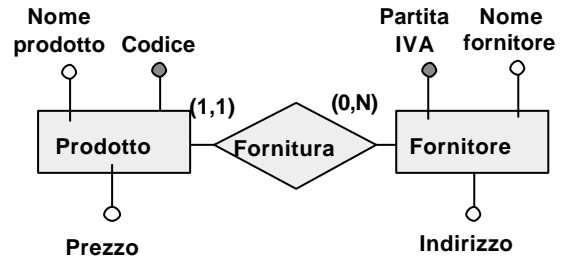
Analisi dell'entità

- L'entità **viola la terza forma normale a causa della dipendenza: PartitaIVA ® NomeFornitore Indirizzo**
- Possiamo decomporre sulla base di questa dipendenza

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

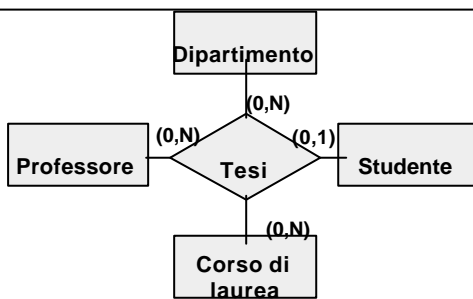
43



23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

44



Studente ® Corso di laurea
Studente ® Professore
Professore ® Dipartimento

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

45

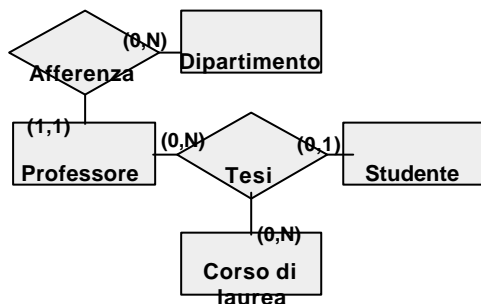
Analisi della relationship

- La relationship **viola la terza forma normale a causa della dipendenza: Professore ® Dipartimento**
- Possiamo decomporre sulla base di questa dipendenza

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

46



23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

47

Ulteriore analisi sulla base delle dipendenze

- La relationship **Tesi** è in BCNF sulla base delle dipendenze
 - Studente ® CorsoDiLaurea
 - Studente ® Professore
- le due proprietà sono indipendenti
- questo suggerisce una ulteriore decomposizione

23/10/2002

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 9

48

