

E. Giunchiglia Basi di dati 1

(trasparenze basate su Atzeni, Ceri, Paraboschi, Torlone: Basi di dati, Capitolo 9)

La normalizzazione

05/10/2004

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

05/10/2004

Basi di Dati 1: La normalizzazione

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

05/10/2004

Basi di Dati 1: La normalizzazione

3

Una relazione con anomalie

<u>Impiegato</u>	<u>Stipendio</u>	<u>Progetto</u>	<u>Bilancio</u>	<u>Funzione</u>
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

05/10/2004

Basi di Dati 1: La normalizzazione

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

05/10/2004

Basi di Dati 1: La normalizzazione

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

05/10/2004

Basi di Dati 1: La normalizzazione

6

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

05/10/2004

Basi di Dati 1: La normalizzazione

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)

05/10/2004

Basi di Dati 1: La normalizzazione

8

Dipendenza funzionale

- relazione r su con attributi 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 ennuple 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
- Notazione:

$$Y \rightarrow Z$$

05/10/2004

Basi di Dati 1: La normalizzazione

9

Definizione "formale" di chiave basata su FD

Sia r una relazione con attributi X :

- Un sottoinsieme Y di X è *superchiave* se $Y \rightarrow X$
- Un sottoinsieme Y di X è *chiave* se $Y \rightarrow X$ e non esiste un insieme Z incluso propriamente in Y tale che $Z \rightarrow X$

05/10/2004

Basi di Dati 1: La normalizzazione

10

Esempi di FD

Impiegato \rightarrow Stipendio
Progetto \rightarrow Bilancio
Impiegato Progetto \rightarrow Funzione

05/10/2004

Basi di Dati 1: La normalizzazione

11

Altre FD

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

05/10/2004

Basi di Dati 1: La normalizzazione

12

Le anomalie sono legate ad alcune FD

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

05/10/2004

Basi di Dati 1: La normalizzazione

13

Non tutte le FD causano anomalie

- In ciascun progetto, un impiegato svolge una sola funzione

Impiegato Progetto → Funzione

- Il soddisfacimento è più "semplice"

05/10/2004

Basi di Dati 1: La normalizzazione

14

Una differenza fra FD

Impiegato → Stipendio
Progetto → Bilancio

- causano anomalie
 Impiegato Progetto → Funzione
- non causa anomalie
- Perché?

05/10/2004

Basi di Dati 1: La normalizzazione

15

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

05/10/2004

Basi di Dati 1: La normalizzazione

16

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

05/10/2004

Basi di Dati 1: La normalizzazione

17

- 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

05/10/2004

Basi di Dati 1: La normalizzazione

18

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

05/10/2004 Basi di Dati 1: La normalizzazione 19

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) $X \rightarrow Y$ definita su di essa, X contiene una chiave di r (ovvero se X è superchiave)
- La forma normale richiede che i concetti in una relazione siano omogenei (solo proprietà direttamente associate alla chiave)

05/10/2004 Basi di Dati 1: La normalizzazione 20

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

05/10/2004 Basi di Dati 1: La normalizzazione 21

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

05/10/2004 Basi di Dati 1: La normalizzazione 22

Non sempre così facile

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

Impiegato → Sede
 Progetto → Sede

05/10/2004 Basi di Dati 1: La normalizzazione 23

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

05/10/2004 Basi di Dati 1: La normalizzazione 24

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!

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. In formule, se e solo se
 - $X_1 \cap X_2 \rightarrow X_1$ oppure
 - $X_1 \cap X_2 \rightarrow X_2$

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 \rightarrow Sede
Progetto \rightarrow Sede

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 \rightarrow Sede
Progetto \rightarrow Sede

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

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

Impiegato \rightarrow Sede
Progetto \rightarrow Sede

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

Impiegato \rightarrow Sede
Progetto \rightarrow Sede

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 → Sede non è conservata

05/10/2004

Basi di Dati 1: La normalizzazione

31

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

05/10/2004

Basi di Dati 1: La normalizzazione

32

Esempio di scomposizione

Data la relazione R con attributi Imp, Funz, Stip e FD Imp → Funz, Funz → Stip, Imp → Stip

- La scomposizione in R1(Imp, Funz), R2(Imp, Stip) è senza perdita ma non conserva le dipendenze
- La scomposizione in R1(Imp, Funz), R2(Funz, Stip) è senza perdita e conserva le dipendenze (la FD Imp → Stip è "non essenziale" in quanto derivabile dalle altre)

05/10/2004

Basi di Dati 1: La normalizzazione

33

Regole di Armstrong

Data un insieme S di FD, una dipendenza $X \rightarrow Y$ è derivabile da S se è ottenibile a partire da S attraverso le seguenti regole:

1. *Riflessività*: $X \rightarrow Y$, se $Y \subseteq X$
2. *Aumento*:

$$\begin{array}{l} X \rightarrow Y \\ XZ \rightarrow YZ \end{array}$$

3. *Transitività*:

$$\frac{X \rightarrow Y \quad Y \rightarrow Z}{X \rightarrow Z}$$

05/10/2004

Basi di Dati 1: La normalizzazione

34

Esempi di derivazione

Data la relazione R con attributi Imp, Funz, Stip e FD Imp → Funz, Funz → Stip,

- La FD Imp → Stip è derivabile grazie alla transitività
- Imp è una superchiave (e quindi anche chiave di R) come dimostra la seguente derivazione:
 1. Imp → Funz (FD di partenza)
 2. Imp → Imp Funz (aumento da 1.)
 3. Funz → Stip (FD di partenza)
 4. Imp Funz → Imp Funz Stip (aumento da 3.)
 5. Imp → Imp Funz Stip (transitività da 2. e 4.)

05/10/2004

Basi di Dati 1: La normalizzazione

35

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

05/10/2004

Basi di Dati 1: La normalizzazione

36

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”

05/10/2004

Basi di Dati 1: La normalizzazione

37

Una nuova forma normale

- Una relazione r è in terza forma normale se, per ogni FD (non banale) $X \rightarrow Y$ 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

05/10/2004

Basi di Dati 1: La normalizzazione

38

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”

05/10/2004

Basi di Dati 1: La normalizzazione

39

Decomposizione in terza forma normale

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

05/10/2004

Basi di Dati 1: La normalizzazione

40

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

05/10/2004

Basi di Dati 1: La normalizzazione

41

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

05/10/2004

Basi di Dati 1: La normalizzazione

42

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

05/10/2004

Basi di Dati 1: La normalizzazione

43

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

05/10/2004

Basi di Dati 1: La normalizzazione

44

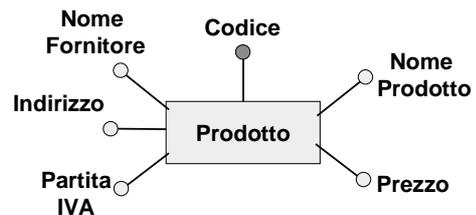
Progettazione e normalizzazione

- 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

05/10/2004

Basi di Dati 1: La normalizzazione

45



PartitaIVA → NomeFornitore Indirizzo

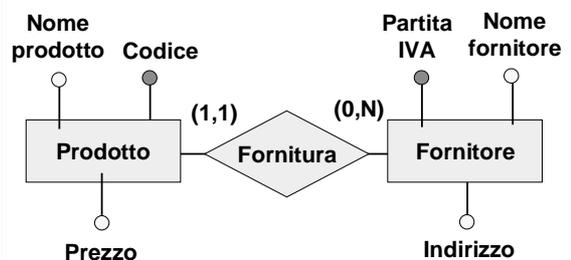
05/10/2004

Basi di Dati 1: La normalizzazione

46

Analisi dell'entità

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



05/10/2004

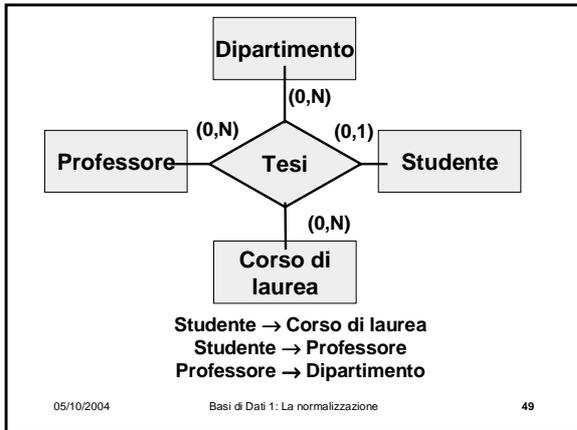
Basi di Dati 1: La normalizzazione

47

05/10/2004

Basi di Dati 1: La normalizzazione

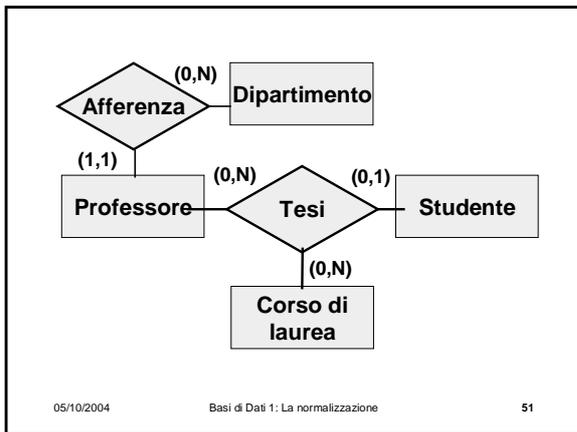
48



Analisi della relationship

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

05/10/2004 Basi di Dati 1: La normalizzazione 50



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

05/10/2004 Basi di Dati 1: La normalizzazione 52

