

Forme normali

285

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

Basi di Dati corso A

286

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

Basi di Dati corso A

287

linee guida (non ortogonali)

- ▶ progettare schema semplice da spiegare
- ▶ progettare schemi esenti da anomalie di inserimento, modifica, cancellazione
- ▶ evitare nelle relazioni attributi con valori nulli frequenti
- ▶ progettare schemi riunibili con equijoin tra chiavi esterne e chiavi primarie

Basi di Dati corso A

288

Esempio di relazione

| Impiegato | Stipendio | Progetto | Bilancio | Funzione |
|-----------|------------------|----------|----------|-------------------|
| Rossi | 20 | Marte | 2 | tecnico |
| Verdi | 35 | Giove | 15 | progettista |
| Verdi | 35 | Venere | 15 | progettista |
| Neri | 55 | Venere | 15 | <u>direttore</u> |
| Neri | 55 | Giove | 15 | <u>consulente</u> |
| Neri | 55 | Marte | 2 | consulente |
| 1 Mori | 48 55 | Marte | 2 | direttore |
| 2 Mori | 48 55 | Venere | 15 | progettista |
| Bianchi | 48 | Venere | 15 | progettista |
| Bianchi | 48 | Giove | 15 | direttore |

Basi di Dati corso A

289

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

Basi di Dati corso A

290

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
- ▶ SQL consente aggiornamento con una sola istruzione

Basi di Dati corso A

291

Dipendenza funzionale

- ▶ Vincolo di integrità che consente di studiare in maniera sistematica questi aspetti
- ▶ 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)

Basi di Dati corso A

292

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 ennuple t_1 e t_2 di r con gli stessi valori su Y ($t_1[Y] = t_2[Y]$)
 - ▶ risulta che t_1 e t_2 hanno gli stessi valori anche su Z ($t_1[Z] = t_2[Z]$)

Impiegato \rightarrow Stipendio $Y \rightarrow Z$

Basi di Dati corso A

293

Notazione

$Y \rightarrow Z$

- ▶ Esempi:

Impiegato \rightarrow Stipendio

Progetto \rightarrow Bilancio

Impiegato Progetto \rightarrow Funzione

Y

Z

Basi di Dati corso A

294

Altre FD

► FD banale:

► $\text{Impiegato Progetto} \rightarrow \text{Progetto}$

$|A| = 1$

► $Y \rightarrow A$ è non banale se A non appartiene a Y

► $Y \rightarrow Z$ è non banale se nessun attributo in Z appartiene a Y

$|Z| > 1$

Basi di Dati corso A

295

Anomalie legate ad alcune FD

► gli impiegati hanno un unico stipendio

$\text{Impiegato} \rightarrow \text{Stipendio}$

► i progetti hanno un unico bilancio

$\text{Progetto} \rightarrow \text{Bilancio}$

Basi di Dati corso A

296

Non tutte le FD causano anomalie

- In ciascun progetto, un impiegato svolge una sola funzione

Impiegato Progetto → Funzione

Basi di Dati corso A

297

Una differenza fra FD

Impiegato → Stipendio

Progetto → Bilancio

- causano anomalie

Impiegato Progetto → Funzione

- non causa anomalie

Basi di Dati corso A

298

| 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

Basi di Dati corso A

299

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

Basi di Dati corso A

300

FD e anomalie

- ▶ si è 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

Basi di Dati corso A

301

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

Basi di Dati corso A

302

FD e chiave

- Esiste una FD tra la chiave e qualsiasi altro attributo dello schema R
- Il vincolo di dipendenza funzionale generalizza il vincolo di chiave
- Una dipendenza funzionale $Y \rightarrow Z$ su uno schema $R(X)$ degenera nel vincolo di chiave se...

... l'unione di Y e Z è pari a X

$Y = \{ \text{impiegato}, \text{progetto} \}$ $Z = \{ \text{stipendio}, \text{bilancio}, \text{turno} \}$

Basi di Dati corso A

303

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 K di r
- La forma normale richiede che i concetti in una relazione siano omogenei (solo proprietà direttamente associate alla chiave)

Basi di Dati corso A

304

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

Basi di Dati corso A

305

| <u>Impiegato</u> | Stipendio | <u>Progetto</u> | Bilancio | Funzione |
|------------------|-----------|-----------------|----------|-------------|
| Rossi | 20 | Marte | 2 | tecnico |
| Verdi | 35 | Giove | 15 | progettista |
| Neri | 55 | Venere | 15 | progettista |
| Mori | 48 | Venere | 15 | direttore |
| Bianchi | 48 | Giove | 15 | consulente |
| | | Marte | 2 | consulente |
| | | Marte | 2 | direttore |
| | | Venere | 15 | progettista |
| | | Venere | 15 | progettista |
| | | Giove | 15 | progettista |
| | | Giove | 15 | direttore |

Basi di Dati corso A

306

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
Come decomponiamo?

Basi di Dati corso A

307

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 |

Basi di Dati corso A

308

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!

Basi di Dati corso A

309

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

Basi di Dati corso A

310

Decomposizione 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 → Sede
Progetto → Sede

Basi di Dati corso A

311

Un altro problema

- Inserimento di 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 → Sede
Progetto → Sede

Basi di Dati corso A

312

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

Basi di Dati corso A

313

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

Basi di Dati corso A

314

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

Basi di Dati corso A

315

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

Basi di Dati corso A

316

Una relazione non-normalizzata

de-normalizzata

| <u>Dirigente</u> | <u>Progetto</u> | <u>Sede</u> |
|------------------|-----------------|-------------|
| Rossi | Marte | Roma |
| Verdi | Giove | Milano |
| Verdi | Marte | Milano |
| Neri | Saturno | Milano |
| Neri | Venere | Milano |

Ogni dirigente opera presso una sede. Ogni progetto ha più dirigenti che sono responsabili, ma in sedi diverse, e ogni dirigente può essere responsabile di più progetti; però, per ogni sede, un progetto ha un solo responsabile.

Progetto Sede → Dirigente
Dirigente → Sede

Basi di Dati corso A

317

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”

Basi di Dati corso A

318

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

Basi di Dati corso A

319

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”

Basi di Dati corso A

320

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

Basi di Dati corso A

321

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

Basi di Dati corso A

322

esempio

| Impiegato | Progetto | Stipendio |
|-----------|----------|-----------|
| Rossi | Marte | 30000 |
| Verdi | Giove | 30000 |
| Verdi | Marte | 30000 |
| Neri | Saturno | 40000 |
| Neri | Venere | 40000 |

► impiegato → stipendio

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

| Impiegato | Stipendio |
|-----------|-----------|
| Rossi | 30000 |
| Verdi | 30000 |
| Verdi | 30000 |
| Neri | 40000 |
| Neri | 40000 |

Basi di Dati corso A

323

Altre forme normali

- Prima: gli attributi sono definiti su valori atomici
- Seconda: se su di essa non sono definite *dipendenze parziali*
- Terza: se su di essa non sono definite *dipendenze transitive*

Basi di Dati corso A

324