

Forme normali

- Proprietà di una base di dati relazionale che ne garantisce la “*qualità*”, cioè l'assenza di determinati difetti:
 - ridondanze,
 - comportamenti poco desiderabili durante gli aggiornamenti
- Definite per il modello relazionale, hanno senso in altri contesti, ad esempio il modello E-R

Normalizzazione

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

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

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*

Perché questi fenomeni indesiderabili?

- relazione unica per rappresentare concetti differenti
 - gli impiegati con i relativi stipendi
 - i progetti con i relativi bilanci
 - le partecipazioni degli impiegati ai progetti con le relative funzioni

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

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)

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 ($Y \rightarrow_r Z$) se, per ogni coppia di ennuple t_1 e t_2
$$t_1.Y = t_2.Y \Rightarrow t_1.Z = t_2.Z$$
- Z dipende funzionalmente da Y

Dipendenza funzionale

- Schema $R(X)$
- due sottoinsiemi non vuoti Y e Z di X
- esiste in R una *dipendenza funzionale* (FD) da Y a Z ($Y \rightarrow_R Z$) se, per ogni relazione r sullo schema $R(X)$
$$Y \rightarrow_r Z$$

Nota

- Una dipendenza funzionale su uno schema è l'asserzione di una proprietà valida per ogni istanza dello schema



Asserzione sulla realtà da modellare



Vincolo di integrità

Esempi

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

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

Le anomalie sono legate ad alcune FD

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

Non tutte le FD causano anomalie

- In ciascun progetto, un impiegato svolge una sola funzione
Impiegato Progetto → Funzione
- Il soddisfacimento è più "semplice"

Una differenza fra FD

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

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

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

- 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

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

Terza 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

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

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

- Con la terza forma normale non elimino tutte le anomalie

visite	Genere	DataVisita	Veterinario
	Giraffa	201090	Rossi
	Giraffa	201190	Rossi
	Giraffa	300191	Rossi
	Leone	201090	Bianchi
	Giraffa	150291	Verdi

- visite fissate periodicamente per genere
- ogni veterinario specializzato in un genere
- veterinari scelti per disponibilità
 - Genere, DataVisita → Veterinario
 - Veterinario → Genere
- Inserimento di un nuovo veterinario
- Ridondanza sul genere di specializzazione

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)

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

Non sempre così facile

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

Impiegato \rightarrow Sede
Progetto \rightarrow Sede

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

Proviamo a ricostruire

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

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 → Sede
 Progetto → 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 → Sede
 Progetto → Sede

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

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

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

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

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

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

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”

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”

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

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

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

Decomposizione in BCNF

Dirigente	Sede	Reparto
Ross	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