

A2 - 3.Progettazione e normalizzazione di un DB relazionale - BCNF

pag. A47 - A55 (parte quinta)

forma normale di Boyce-Codd

Sappiamo che uno schema di relazione si dice in 3NF se è in 2NF e non ci sono attributi non primi che dipendono transitivamente dalla chiave (privata o candidata).

Abbiamo visto che questa assunzione introduce una definizione alternativa in base alla quale uno schema di relazione *Rel* si dice in 3NF se ogni volta che sussiste una dipendenza funzionale $X \rightarrow Y$ non banale (cioè per la quale risulti: $Y \not\subseteq X$), abbiamo:

- X è una superchiave per *Rel*,
oppure:
- Y è un attributo primo di *Rel*.

In pratica **il test di 3NF** ammette le dipendenze transitive solo se l'attributo determinato dalla catena di dipendenze funzionali è primo, dunque **non esclude che possano esistere attributi primi che dipendono funzionalmente da attributi non primi**.

forma normale di Boyce-Codd

Riprendiamo in esame lo schema di relazione:

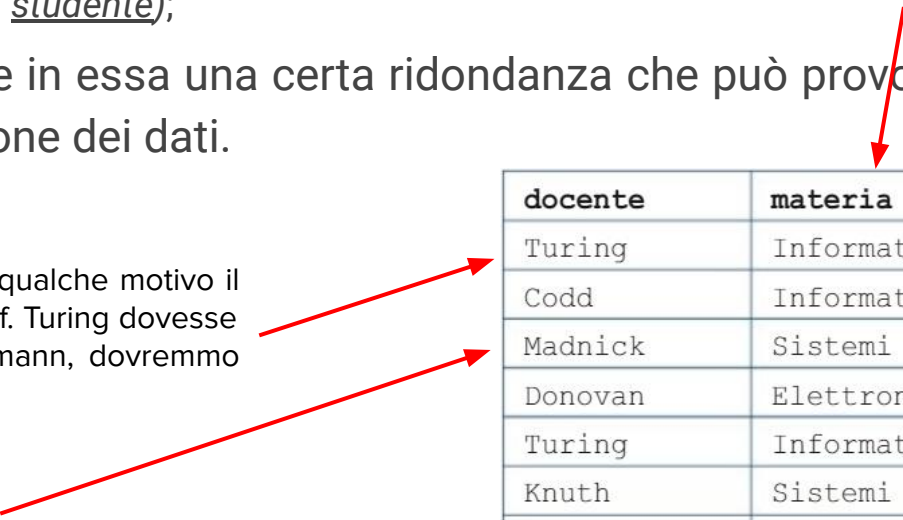
Insegnamenti (*docente*, *materia*, *studente*);

è in 3NF; tuttavia permane in essa una certa ridondanza che può provocare problemi e anomalie durante la gestione dei dati.

Anomalie di aggiornamento: se per qualche motivo il corso di Informatica assegnato al prof. Turing dovesse essere assegnato al prof. von Neumann, dovremmo modificare due righe della tabella.

Anomalie di cancellazione: se decidessimo di abolire il corso di Sistemi tenuto dal prof. Madnick, perderemmo anche i dati degli studenti iscritti a quel corso.

Anomalie di inserimento: se decidessimo di istituire il corso di Programmazione da assegnare al prof. Boole, non potremmo se non abbiamo ancora alcuno studente iscritto, perché l'attributo *studente* è parte della chiave primaria.



docente	materia	studente
Turing	Informatica	Rossi
Codd	Informatica	Neri
Madnick	Sistemi	Neri
Donovan	Elettronica	Neri
Turing	Informatica	Bianchi
Knuth	Sistemi	Bianchi
Wirth	Informatica	Verdi
Codd	Informatica	Grigi
Madnick	Sistemi	Rossi

forma normale di Boyce-Codd

Appare chiaro come non sempre la 3NF è in grado di risolvere eventuali anomalie.

È stata allora introdotta la **forma normale di Boyce-Codd** (**BCNF** *Boyce-Codd Normal Form*) che ha condizioni più restrittive rispetto alla 3NF, nel senso che **in nessun caso sono ammesse dipendenze transitive**, neanche se l'attributo determinato dalla catena di dipendenze funzionali è primo; questo vuol dire che **ogni determinante deve essere chiave (candidata o primaria) per lo schema dato**; non possono esserci determinanti non primi, neanche se determinano attributi primi, (contrariamente a quanto visto per il test di 3NF).

Uno schema di relazione $Rel(A_1, A_2, \dots, A_n)$ è in BCNF se, **ogni volta che sussiste in Rel una dipendenza funzionale non banale $X \rightarrow Y$, X è una superchiave per Rel .**

forma normale di Boyce-Codd

Appare ovvio che uno schema di relazione che sia in BCNF è anche in 3NF, mentre uno schema in 3NF non necessariamente è anche in BCNF, come mostra anche il seguente esempio.

Vogliamo gestire l'allocazione delle sale operatorie di un ospedale; queste vengono prenotate, giorno per giorno, in orari stabiliti, per effettuare interventi sui pazienti ad opera dei chirurghi dell'ospedale.

Nel corso della stessa giornata una sala operatoria viene occupata sempre dallo stesso chirurgo che effettua più interventi, in ore diverse.

Per questo motivo, noti il paziente e la data dell'intervento, sono noti anche: l'ora dell'intervento, il chirurgo che lo deve effettuare, e la sala operatoria utilizzata.

forma normale di Boyce-Codd

Gli attributi della relazione *Interventi* sono descritti nello schema di relazione seguente:

Interventi (Paziente, DataIntervento, *OraIntervento*, *Chirurgo*, *Sala*);

Risultano quindi verificate le seguenti dipendenze funzionali:

- A. **(Paziente, DataIntervento) → OraIntervento, Chirurgo, Sala;**
- B. **(Chirurgo, DataIntervento, OraIntervento) → Paziente, Sala;**
- C. **(Sala, DataIntervento, OraIntervento) → Paziente, Chirurgo;**
- D. **(Chirurgo, DataIntervento) → Sala;**

Le prime tre dipendenze funzionali sono *dipendenze fondamentali*, essendo i rispettivi determinanti chiavi candidate per lo schema in questione.

Quest'ultima dipendenza funzionale traduce la proprietà per la quale “*nel corso della stessa giornata, una sala operatoria viene occupata sempre dallo stesso chirurgo che può effettuare più interventi in ore diverse*”.

Paziente	DataIntervento	OraIntervento	Chirurgo	Sala
Bianchi	25/10/2005	8.00	De Bakey	Sala1
Rossi	25/10/2005	8.00	Romano	Sala2
Negri	26/10/2005	9.30	Veronesi	Sala1
Viola	25/10/2005	10.30	De Bakey	Sala1
Verdi	25/10/2005	11.30	Romano	Sala2
Interventi				

forma normale di Boyce-Codd

I determinanti nelle dipendenze funzionali A, B e C, cioè (*Paziente, DataIntervento*), (*Chirurgo, DataIntervento, OraIntervento*) e (*Sala, DataIntervento, OraIntervento*), sono chiavi candidate, quindi non violano il test di 3NF.

Il determinante della dipendenza funzionale D, cioè (*Chirurgo, DataIntervento*), non è invece chiave candidata, tuttavia lo schema di relazione è ancora in 3NF in quanto l'attributo *Sala* fa parte della chiave candidata (*Sala, DataIntervento, OraIntervento*), quindi è un attributo primo.

Paziente	DataIntervento	OraIntervento	Chirurgo	Sala
Bianchi	25/10/2005	8.00	De Bakey	Sala1
Rossi	25/10/2005	8.00	Romano	Sala2
Negri	26/10/2005	9.30	Veronesi	Sala1
Viola	25/10/2005	10.30	De Bakey	Sala1
Verdi	25/10/2005	11.30	Romano	Sala2
Interventi				

forma normale di Boyce-Codd

Comunque, lo schema di relazione *Interventi*, pur essendo in 3NF, presenta una certa ridondanza nei dati che può creare problemi in fase di aggiornamento.

Se per qualche ragione si dovesse cambiare la sala operatoria utilizzata da un chirurgo in un certa data, bisognerebbe aggiornare più righe: ad es., per spostare Romano dalla Sala2 alla Sala3, bisogna modificare due righe della tabella.

Paziente	DataInterevnto	OraIntervento	Chirurgo	Sala
Bianchi	25/10/2005	8.00	De Bakey	Sala1
Rossi	25/10/2005	8.00	Romano	Sala2
Negri	26/10/2005	9.30	Veronesi	Sala1
Viola	25/10/2005	10.30	De Bakey	Sala1
Verdi	25/10/2005	11.30	Romano	Sala2
Interventi				

forma normale di Boyce-Codd

Si può risolvere il problema dividendo lo schema *Interventi* in due:

OccupazioneSale (Chirurgo, DataIntervento, Sala);

Interventi (Paziente, DataIntervento, OraIntervento, Chirurgo);

L'attributo *Sala* viene tolto dallo schema *Interventi* e inserito in una nuova tabella che ha come chiave il determinante (*Chirurgo*, *DataIntervento*) della dipendenza funzionale D; lo stesso diventa chiave esterna nello schema *Interventi* in relazione *uno-a-molti* con lo schema *OccupazioneSale*.

Chieurgo	DataInterevnto	Sala
De Bakey	25/10/2005	Sala1
Romano	25/10/2005	Sala2
Veronesi	26/10/2005	Sala1
OccupazioneSale		

Paziente	DataInterevnto	OraIntervento	Churgo
Bianchi	25/10/2005	8.00	De Bakey
Rossi	25/10/2005	8.00	Romano
Negri	26/10/2005	9.30	Veronesi
Viola	25/10/2005	10.30	De Bakey
Verdi	25/10/2005	11.30	Romano
Interventi			

forma normale di Boyce-Codd

Si può risolvere il problema dividendo lo schema *Interventi* in due:

OccupazioneSale (Chirurgo, DataIntervento, Sala);

Interventi (Paziente, DataIntervento, OraIntervento, Chirurgo);

Adesso questi due schemi di relazione, oltre ad essere in 3NF, risultano anche in BCNF.

L'eventuale riassegnazione delle sale operatorie ora avviene modificando una sola riga della tabella *OccupazioneSale*.

Chieurgo	DataInterevnto	Sala
De Bakey	25/10/2005	Sala1
Romano	25/10/2005	Sala2
Veronesi	26/10/2005	Sala1
OccupazioneSale		

Paziente	DataInterevnto	OraIntervento	Churgo
Bianchi	25/10/2005	8.00	De Bakey
Rossi	25/10/2005	8.00	Romano
Negri	26/10/2005	9.30	Veronesi
Viola	25/10/2005	10.30	De Bakey
Verdi	25/10/2005	11.30	Romano
Interventi			

forma normale di Boyce-Codd

Nel nostro esempio è bastato creare un nuovo schema a partire dalla dipendenza funzionale che viola la BCNF, la cui chiave è proprio il determinante della dipendenza stessa.

Si può risolvere il problema dividendo lo schema *Interventi* in due:

OccupazioneSale (Chirurgo, DataIntervento, Sala);

Interventi (Paziente, DataIntervento, OraIntervento, Chirurgo);

Questo processo di normalizzazione si fonda su un semplice criterio: **se uno schema di relazione presenta più concetti indipendenti, allora va decomposto in schemi di relazione più piccoli, uno per ogni concetto.**

Chieurgo	DataInterevnto	Sala
De Bakey	25/10/2005	Sala1
Romano	25/10/2005	Sala2
Veronesi	26/10/2005	Sala1
OccupazioneSale		

Paziente	DataInterevnto	OraIntervento	Chirurgo
Bianchi	25/10/2005	8.00	De Bakey
Rossi	25/10/2005	8.00	Romano
Negri	26/10/2005	9.30	Veronesi
Viola	25/10/2005	10.30	De Bakey
Verdi	25/10/2005	11.30	Romano
Interventi			

forma normale di Boyce-Codd

Riassumiamo brevemente Il significato delle forme normali:

Una relazione si dice in **prima forma normale (1FN)** quando rispetta i requisiti fondamentali del modello relazionale, in particolare ogni attributo è atomico, non ci sono righe uguali e non ci sono attributi multivalore.

Una relazione è in **seconda forma normale (2FN)** quando è in prima forma normale e non ci sono attributi non-chiave che dipendono parzialmente dalla chiave.

Una relazione è in **terza forma normale (3FN)** quando è in seconda forma normale e non ci sono attributi non-chiave che dipendono transitivamente dalla chiave.

Una relazione è in **forma normale di Boyce-Codd (BCNF)** quando è in prima forma normale e in essa ogni determinante è una chiave candidata.

forma normale di Boyce-Codd

Tra le tante proposte che hanno caratterizzato gli studi sulla normalizzazione delle basi di dati relazionali, sono emerse anche la quarta e la quinta forma normale.

Si tratta di due ulteriori forme normali che raramente vengono utilizzate, perché a un loro incremento di rigore nell'eliminazione delle ridondanze corrisponde un forte degrado delle prestazioni.

In queste forme, infatti, le operazioni per il recupero o la modifica dei dati richiedono molto tempo per la loro esecuzione.

