

## Forma normale di Boyce-Codd

Una relazione è in **forma normale di Boyce-Codd** (BCNF, *Boyce-Codd Normal Form*) quando rispetta le caratteristiche fondamentali del modello relazionale (1FN) e in essa ogni determinante è una chiave candidata, cioè ogni attributo dal quale dipendono altri attributi può svolgere la funzione di chiave.

La BCNF può essere espressa anche nel seguente modo: se in una relazione vale la dipendenza funzionale  $A \rightarrow B$ , allora l'insieme di attributi  $A$  deve contenere una chiave (e quindi può svolgere la funzione di chiave).

Da questo fatto discende immediatamente che una relazione che soddisfa la BCNF è anche in seconda e in terza forma normale, in quanto la BCNF esclude che un determinante possa essere composto solo da una parte della chiave, come avviene per le violazioni alla 2FN, o che possa essere esterno alla chiave, come avviene per le violazioni alla 3FN.

Una relazione che rispetta la forma normale di Boyce-Codd è, quindi, anche in terza forma normale, ma non è vero l'opposto, come mostra il seguente esempio.

Consideriamo una relazione che descrive l'allocatione delle sale operatorie di un ospedale. Le sale operatorie sono prenotate, giorno per giorno, in orari previsti, per effettuare interventi su pazienti ad opera dei chirurghi dell'ospedale. Nel corso di una giornata una sala operatoria è occupata sempre dal medesimo chirurgo che effettua più interventi, in ore diverse. Noti i valori di *Paziente* e *DataIntervento*, sono noti anche: ora dell'intervento, chirurgo, e sala operatoria utilizzata. Gli attributi della relazione *Interventi* sono descritti nello schema:

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

In base alla precedente descrizione del caso in esame, nella relazione *Interventi* valgono le dipendenze funzionali:

- a. {Paziente, DataIntervento}  $\rightarrow$  OraIntervento, Chirurgo, Sala
- b. {Chirurgo, DataIntervento, OraIntervento}  $\rightarrow$  Paziente, Sala
- c. {Sala, DataIntervento, OraIntervento}  $\rightarrow$  Paziente, Chirurgo
- d. {Chirurgo, DataIntervento}  $\rightarrow$  Sala

Ci sono tre insiemi di attributi che possono svolgere la funzione di chiave: {*Paziente*, *DataIntervento*}, {*Chirurgo*, *DataIntervento*, *OraIntervento*}, {*Sala*, *DataIntervento*, *OraIntervento*}. Scegliamo come chiave primaria la coppia di attributi: {*Paziente*, *DataIntervento*}.

### Interventi

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

I determinanti nelle dipendenze funzionali **a**, **b**, **c**, sono insiemi di attributi che possono svolgere la funzione di chiave e quindi la BCNF non è sicuramente violata in questi casi. La BCNF non è invece soddisfatta dalla dipendenza funzionale **d** che ha come determinante un insieme di attributi non chiave. Ne segue che la relazione *Interventi* non è in BCNF.

*Interventi* è invece in 3FN in quanto la dipendenza funzionale **d** viene esclusa dalle dipendenze funzionali considerate dalla terza forma normale, perché l'attributo *Sala* è un attributo che fa parte della chiave candidata {*Sala*, *DataIntervento*, *OraIntervento*} e quindi *Sala* non è un attributo non chiave.

Osserviamo che la relazione *Interventi*, pur essendo in terza forma normale, presenta una certa ridondanza nei dati che può creare problemi in fase di aggiornamento. Se per qualche ragione si deve cambiare la sala operatoria utilizzata da un chirurgo in un certa data, bisogna aggiornare più righe: per esempio, per spostare *Romano* dalla *Sala2* alla *Sala3*, bisogna modificare due righe della tabella.

Naturalmente, la tabella *Interventi* può essere normalizzata, ottenendo i due schemi:

**OccupazioneSale** (Chirurgo, DataIntervento, Sala)

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

L'attributo *Sala* viene tolto da *Interventi* e compare in una nuova tabella che ha come chiave il determinante della dipendenza funzionale **d** che non rispettava la BCNF.

### Interventi

Paziente	DataIntervento	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

### OccupazioneSale

Chirurgo	DataIntervento	Sala
De Bakey	25/10/2005	Sala1
Romano	25/10/2005	Sala2
Veronesi	26/10/2005	Sala1

Nella nuova implementazione l'eventuale riassegnazione delle sale operatorie avviene modificando una sola riga della tabella *OccupazioneSale*.

La BCNF è quindi una forma di normalizzazione più forte della terza forma normale. Ci si può accorgere di questo anche osservando che, sostanzialmente, la 3FN richiede che ogni attributo *non chiave* dipenda da un insieme di attributi che possono formare una chiave, mentre la BCNF lo richiede per *ogni* attributo.

Il significato delle forme normali viene di seguito riassunto:

### **Prima forma normale**

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

### **Seconda forma normale**

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.

### **Terza forma normale**

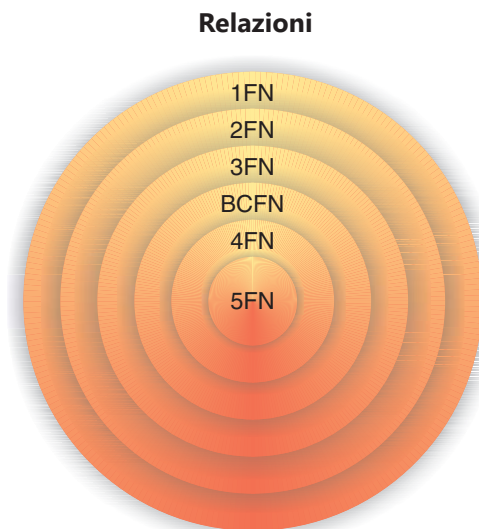
Una relazione è in terza forma normale (3FN) quando è in seconda forma normale e non ci sono attributi non-chiave che dipendono *transitivamente* dalla chiave.

### **Forma normale di Boyce-Codd**

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.

Nella teoria dei database relazionali esistono forme normali di ordine superiore al terzo che risolvono situazioni di anomalia nelle operazioni sulle tabelle.

Più precisamente la quarta e la quinta forma normale risolvono i problemi che si possono creare quando nella relazione sono presenti attributi *multivalore*, cioè attributi che possono assumere più valori in corrispondenza dello stesso valore di un altro attributo. Queste forme inoltre servono a rendere minimo il numero degli attributi che formano le chiavi composte.



Sebbene sia possibile definire tabelle anche in quarta (4FN) e quinta forma normale (5FN), di solito è sufficiente rappresentare le relazioni nel livello di normalizzazione 3FN che, come si può dimostrare, ha il pregio di essere sempre ottenibile senza perdita di informazioni e senza perdita di dipendenze funzionali. Non è così invece per la forma normale di Boyce-Codd: ci sono relazioni che non possono essere normalizzate nella forma di Boyce-Codd senza perdita di dipendenze funzionali, come si vede dall'esempio seguente.

Supponiamo di voler tenere traccia dei pazienti che devono essere sottoposti a più terapie chirurgiche, in diversi reparti, per la terapia di patologie complesse.

Una relazione che soddisfa a tale esigenza è mostrata nella tabella a lato. Ogni n-upla della relazione *TerapieComplesse* associa un paziente al chirurgo che lo ha operato e al reparto nel quale è avvenuto l'intervento. Valgono le seguenti dipendenze funzionali:

Chirurgo → Reparto {Paziente, Reparto} → Chirurgo

### TerapieComplesse

Paziente	Reparto	Chirurgo
Rossi	Cardiochirurgia	De Bakey
Rossi	Chir. Generale	Romano
Bianchi	Chir. Generale	Romano
Bianchi	Chir. Oncologica	Veronesi
Verdi	Chir. Generale	Lanzetta

Infatti un chirurgo è inquadrato in un determinato reparto e ogni paziente, quando è ricoverato in un certo reparto, è operato da un dato chirurgo. {Paziente, Reparto} è chiave per *TerapieComplesse*, per via della seconda dipendenza funzionale. La relazione è in 3FN, ma non è in BCNF, a causa della prima dipendenza funzionale che ha come determinante *Chirurgo* che non è chiave. Se cerchiamo di scomporre la relazione con la tecnica illustrata in precedenza a partire dalla dipendenza funzionale *Chirurgo* → *Reparto*, otteniamo le due relazioni *Chirurghi* e *Pazienti*.

### Chirurghi

Chirurgo	Reparto
De Bakey	Cardiochirurgia
Romano	Chir. Generale
Veronesi	Chir. Oncologica
Lanzetta	Chir. Generale

### Pazienti

Paziente	Chirurgo
Rossi	De Bakey
Rossi	Romano
Bianchi	Romano
Bianchi	Veronesi
Verdi	Lanzetta

Questa decomposizione non conserva la seconda delle due dipendenze funzionali definite sulla relazione originaria. L'effetto di questa **perdita di dipendenze funzionali** può avere effetti indesiderati.

Se, per esempio, si volesse registrare il fatto (errato) che il paziente *Bianchi* è stato operato da *Lanzetta* nel reparto di *Cardiochirurgia*, ci si limiterebbe a inserire nella tabella *Paziente* la coppia di valori: ("Bianchi", "Lanzetta") che è lecita. Solo quando si cerca di ricostruire i dati della relazione *TerapieComplesse*, congiungendo le tabelle *Pazienti* e *Chirurghi*, si ottiene la n-upla: ("Bianchi", "Lanzetta", "Chir. Generale") che permette di evidenziare l'errore nei dati.

L'esempio mostra l'esistenza di relazioni che non sono in BCNF e non vengono portate a tale livello di normalizzazione, ma lasciate in 3NF, anche a costo di una certa ridondanza dei dati, perché la BCNF può portare alla perdita di dipendenze funzionali.

Le forme normali di ordine superiore contengono la stessa quantità di informazioni di quelle inferiori e non è obbligatorio il passaggio a queste forme normali: il modello relazionale richiede come obbligatoria solo la prima forma normale. Tuttavia il passaggio alle forme normali superiori consente di distinguere e separare con precisione gli oggetti, senza perdita di informazioni, anche se viene generata una ridondanza di dati che però è costantemente sotto controllo.