

Normalizzazione

● Il concetto di normalizzazione

Lo schema, definito dalla progettazione concettuale, deve essere tradotto in uno schema “migliore”, cioè depurato da gran parte delle anomalie di gestione che si possono verificare. Si distinguono tre tipi di anomalia:

- *anomalia di inserimento*: se nell'inserire un nuovo record in una tabella si è costretti a inserire informazioni già presenti nel DB;
- *anomalia di cancellazione*: se nel cancellare un record si è costretti a cancellare informazioni che possono essere ancora utili nel DB;
- *anomalia di aggiornamento*: se dovendo aggiornare un record si è costretti ad aggiornarne molti altri.

Per chiarire meglio questi concetti consideriamo le anomalie che possono verificarsi aggiornando la seguente tabella:

| codProdotto | prezzo | fornitore | indirizzo | città | quantità fornita |
|----------------|--------|-----------|---------------|--------|------------------|
| quaderni | 1,30 | Rossi | Via Roma 10 | Genova | 100 |
| pennarelli | 3,10 | Rossi | Via Roma 10 | Genova | 50 |
| matite | 1,10 | Verdi | Via Genova 8 | Torino | 10.000 |
| pennne a sfera | 1,20 | Verdi | Via Genova 8 | Torino | 10.000 |
| colori a olio | 5,50 | Bruni | Via Milano 10 | Pisa | 10 |

Anomalia di inserimento: per inserire un nuovo ordine bisogna inserire nuovamente i dati anagrafici del fornitore (indirizzo, città) e del prodotto (prezzo).

Anomalia di cancellazione: cancellando il record relativo all'ordine del prodotto colori a olio si cancellano anche le informazioni relative al fornitore Bruni.

Anomalia di aggiornamento: aggiornando l'indirizzo del fornitore Rossi bisogna aggiornare due tuple.

La **normalizzazione** è l'insieme di criteri di progettazione di un database relazionale diretto a prevenire l'insorgere di tali anomalie.



La **normalizzazione** è il procedimento che trasforma successivamente le relazioni di partenza suddividendole in altre più piccole aventi lo stesso contenuto di informazione.

Sono stati definiti diversi gradi di normalizzazione (secondo alcuni testi 4, secondo altri 5) a cui si fanno corrispondere le *forme normali*, molto impiegate nel passato per la progettazione delle basi di dati relazionali. Anche se normalmente si ritiene accettabile uno schema relazionale in cui tutte le relazioni siano in terza forma normale, ultimamente ci si è resi conto che la rigorosità e purezza matematica di una relazione non corrisponde necessariamente a una base di dati efficiente e si sono quindi sviluppati altri metodi di progettazione. Il concetto di normalizzazione è ugualmente importante, poiché una tabella non può essere considerata relazionale se non è normalizzata, cioè in prima forma normale. Per completezza di descrizione del modello relazionale riportiamo le definizioni delle prime tre forme normali.

● La prima forma normale

Una relazione si dice **normalizzata** o in **prima forma normale** (1FN) se tutti i suoi attributi hanno un dominio semplice. Non sono ammessi gruppi e ripetizioni.



La 1FN implica che ogni informazione deve essere “atomica”, cioè un campo deve contenere una e una sola informazione. I campi devono contenere sempre un tipo di dato “semplice” (stringa, numero, binario ecc.) e mai aggregazioni (insieme, vettore).

Se ci si trova nella situazione della **figura 1a**, la tabella DIPENDENTI non è normalizzata perché il campo FIGLI non è elementare, e inoltre vi sono dei gruppi ripetuti (i figli di Rossi e di Bianchi). Affinché la tabella sia nella prima forma normale, è necessario trasformarla nella tabella illustrata nella **figura 1b**.

| DIPENDENTI | | | FIGLI DIP. | | |
|------------|--------|-----|------------|--------|-----------|
| nome dip. | figli | | dipen. | figli | fornitore |
| | nome | età | | | |
| Rossi | Ugo | 10 | Rossi | Ugo | 10 |
| | Andrea | 7 | Rossi | Andrea | 7 |
| Bianchi | Gianni | 6 | Bianchi | Gianni | 6 |
| | Maria | 3 | Bianchi | Maria | 3 |
| | Luca | 1 | Bianchi | Luca | 1 |
| Verdi | Pia | 8 | Verdi | Pia | 8 |

a

b

figura 1 Normalizzazione

● Dipendenze funzionali

Per formalizzare i problemi visti si introduce un nuovo tipo di vincolo, la *dipendenza funzionale*.

Consideriamo una relazione R nella quale siano definiti almeno due attributi X e Y. Diciamo che in R vale la **dipendenza funzionale** di Y da X (e si indica con $X \rightarrow Y$) se per ogni coppia di tuple t1 e t2 di R con gli stessi valori su X, t1 e t2 hanno gli stessi valori anche su Y.

Si dice anche che X *determina funzionalmente* Y o ancora che X è un *determinante* per Y.

Chiariamo meglio questo concetto con un esempio: consideriamo la relazione STUDENTE e stabiliamo che nella relazione la chiave primaria sia Matricola:

STUDENTE (Matricola, Nome, Telefono, Corso, Voto)

In essa si possono evidenziare le seguenti dipendenze funzionali:

Matricola \rightarrow Nome

Matricola \rightarrow Telefono

Matricola, Corso \rightarrow Voto

Nome ha una dipendenza funzionale da Matricola perché a ogni Nome corrisponde una Matricola. Telefono dipende funzionalmente da Matricola perché a ogni Telefono corrisponde una Matricola. Voto dipende funzionalmente da Matricola, Corso perché a ogni Voto corrisponde l'insieme di attributi Matricola, Corso.

● La seconda forma normale

La seconda forma normale (2FN) si applica alle tabelle che hanno la chiave primaria composta da più attributi: è richiesto che tutti gli attributi di una riga dipendano dall'intera chiave primaria e non solo da una parte di essa.



Una relazione si dice in **seconda forma normale** (2FN) se è in 1FN e tutti i suoi attributi che non appartengono alla chiave dipendono funzionalmente e completamente dall'intera chiave; non possono esistere attributi che dipendono solamente da una parte della chiave.

Per esempio nella relazione della **figura 2**, che contiene i dati dei giocatori che hanno segnato reti nei campionati, l'attributo Reti, che indica il numero di reti segnate da un giocatore in un campionato, dipende dalla chiave composta Nome-Anno, mentre l'attributo Luogo che indica il luogo di nascita del giocatore è dipendente solo dall'attributo Nome che è un sottoinsieme della chiave: la relazione non è in 2FN e presenta le seguenti anomalie:

- *anomalia di inserimento*: non è possibile inserire un nuovo giocatore sino a quando non ha segnato reti in un campionato;
- *anomalia di cancellazione*: cancellando la terza riga della tabella si perdono le informazioni del giocatore Conti Bruno;
- *anomalia di aggiornamento*: aggiornando il luogo di nascita del giocatore Rossi bisogna aggiornare due tuple.

Affinché la relazione di **figura 2** sia in 2FN bisogna eliminare dalla tabella CAMPIONATI l'attributo Luogo e aggiungere la nuova tabella GIOCATORI che contiene i campi Nome e Luogo (**figura 3**).

CAMPIONATI

| nome | luogo | anno | reti |
|-------------|---------|-----------|------|
| Rossi Mario | Firenze | 1998/1999 | 3 |
| Rossi Mario | Firenze | 1999/2000 | 4 |
| Conti Bruno | Roma | 1998/1999 | 6 |

figura 2 Relazione CAMPIONATI

GIOCATORI

| nome | luogo |
|-------------|---------|
| Rossi Mario | Firenze |
| Conti Bruno | Roma |

CAMPIONATI

| nome | reti | anno |
|-------------|------|-----------|
| Rossi Mario | 3 | 1998/1999 |
| Rossi Mario | 4 | 1999/2000 |
| Conti Bruno | 6 | 1998/1999 |

figura 3 Seconda forma normale

● La terza forma normale

In una tabella in terza forma normale (3FN) tutte le dipendenze tra colonne devono essere basate sulla chiave primaria; vengono eliminate le dipendenze funzionali transitive di un attributo non chiave da un altro attributo anch'esso non chiave.



Una relazione si dice in **terza forma normale** (3FN) se è in 2FN e tutti i suoi attributi che non appartengono alla chiave dipendono direttamente dalla chiave; non possono esistere attributi non chiave che dipendono funzionalmente da altri attributi non chiave.

Esaminiamo la relazione di **figura 4**.

FATTURA

| numero | cliente | ragione sociale | importo |
|--------|---------|-----------------|---------|
| 001 | Rossi | 001-344 | 700 |
| 002 | Rossi | 001-344 | 600 |
| 003 | Verdi | 022-455 | 550 |
| 004 | Rossi | 001-344 | 1800 |

figura 4 Relazione FATTURA

In questa relazione, che ha come chiave il campo Numero, è presente il campo Ragione_Sociale che non dipende dalla chiave primaria, bensì da un altro campo, Cliente, che non è chiave: la tabella non è in terza forma normale data la presenza della dipendenza transitiva:

Ragione_Sociale \longrightarrow Cliente \longrightarrow Numero

La relazione FATTURA presenta le seguenti anomalie:

- *anomalia di inserimento*: non è possibile inserire la ragione sociale relativa a un cliente sino a che quest'ultimo non abbia emesso una fattura;
- *anomalia di cancellazione*: cancellando la terza riga della tabella si perde la ragione sociale del cliente Verdi;
- *anomalia di aggiornamento*: se varia la ragione sociale di Rossi occorre aggiornare tre tuple.

Anche la terza forma normale può essere ottenuta mediante un procedimento di scomposizione, separando dalla relazione di partenza il sottoinsieme di attributi dipendente in modo transitivo dalla chiave primaria (**figura 4**).

Per rendere la relazione FATTURA in terza forma normale bisogna eliminare dalla tabella FATTURA l'attributo Ragione_Sociale e aggiungere la nuova tabella CLIENTI che contiene i campi Cliente e Ragione_Sociale (**figura 4**).

FATTURA

| numero | cliente | importo |
|--------|---------|---------|
| 001 | Rossi | 700 |
| 002 | Rossi | 600 |
| 003 | Verdi | 550 |
| 004 | Rossi | 1800 |

CLIENTE

| cliente | ragione sociale |
|---------|-----------------|
| Rossi | 001-344 |
| Verdi | 022-455 |

figura 5 Terza forma normale

La terza forma normale può essere espressa secondo la formulazione di Boyce-Codd.

Si dice che una relazione è in **BCNF** (Forma normale di Boyce-Codd o Boyce-Codd Normal Form) se è in prima forma normale (1FN) e ogni determinante è una chiave candidata, cioè ogni attributo dal quale dipendono altri attributi può essere una chiave.



verifica le tue conoscenze

- 1 Quando una tabella si dice normalizzata?
- 2 Quando una relazione è in seconda forma normale?
- 3 Quando una tabella è in terza forma normale?
- 4 Che cos'è la dipendenza funzionale?