

La normalizzazione delle relazioni

Gabriele Russo Russo

Cos'è la normalizzazione?

Nella progettazione di una tabella bisogna **evitare la ridondanza** dei dati, che può causare problemi durante l'aggiornamento o la modifica dei dati all'interno del database.

Il processo che ci permette di evitare tali problemi prende il nome di **normalizzazione** delle relazioni.

Consiste nel trasformare le tabelle, in modo tale che ciascuna corrisponda ad un singolo oggetto della realtà.

La normalizzazione avviene a livelli crescenti (**forme normali**).

Alcune premesse...

Chiave primaria: insieme di uno o più attributi che identificano in modo univoco un record

Attributo non-chiave: qualsiasi altro campo che non faccia parte della chiave primaria

Dipendenza funzionale tra attributi: il valore dell'attributo A1 determina un singolo valore dell'attributo A2 ($A1 \rightarrow A2$)

Dipendenza transitiva: A2 dipende da A1, A3 dipende da A2, quindi A3 dipende transitivamente da A1

$A1 \rightarrow A2$ $A2 \rightarrow A3$ quindi $A1 \rightarrow A3$

1FN – Prima forma normale

Requisiti fondamentali del modello relazionale:

- tutte le righe della tabella contengono lo stesso numero di colonne
- gli attributi rappresentano informazioni elementari
- i valori che compaiono in una colonna sono dello stesso tipo
- ogni riga è diversa da tutte le altre
- l'ordine delle righe è irrilevante

1FN – Esempio

<u>CodiceStudente</u>	Nome	Indirizzo	ComponentiFamiglia



<u>CodiceStudente</u>	Nome	Indirizzo



<u>CodiceFamiliare</u>	Nome	CodiceStudente

2FN – Seconda forma normale

E' una relazione in prima forma normale e inoltre tutti i suoi **attributi non-chiave dipendono dall'intera chiave primaria**.

Una relazione in 2FN quindi non possiede attributi che dipendono solo da una parte della chiave primaria.

E' facile comprendere che la 2FN riguarda relazioni con chiavi composte, cioè formate da più attributi.

2FN – Esempio (1)

<u>CodMerce</u>	<u>CodMagazzino</u>	Quantità	LocalitàMagazzino

La chiave è composta, formata da due attributi. Infatti *CodMerce* non è sufficiente, poiché la stessa merce può trovarsi in due magazzini differenti.

Mentre *Quantità* dipende sia dal tipo di merce (*CodMerce*) che dal magazzino (*CodMagazzino*), l'indirizzo del magazzino non dipende dal tipo di merce in questione, ma solo dal magazzino (*CodMagazzino*).

LocalitàMagazzino dipende quindi **parzialmente** dalla chiave.

2FN – Esempio (2)

<u>CodMerce</u>	<u>CodMagazzino</u>	Quantità	LocalitàMagazzino

Come visto, **la relazione non è in 2FN** e questo potrebbe provocare dei problemi:

- l'indirizzo del magazzino è ripetuto per tutte le righe che si riferiscono a prodotti presenti in quel magazzino
- se un magazzino viene trasferito, bisognerà aggiornare tutte le righe interessate
- se l'indirizzo fosse scritto in modo (anche leggermente) differente da una riga all'altra, potremmo avere problemi nella ricerca dei dati
- se in un periodo non ci fossero merci in un magazzino, non avremmo informazioni su quel magazzino

2FN – Esempio (3)

<u>CodMerce</u>	<u>CodMagazzino</u>	Quantità	LocalitàMagazzino



<u>CodMerce</u>	<u>CodMagazzino</u>	Quantità



<u>CodMagazzino</u>	LocalitàMagazzino

N.B. E' evidente che la normalizzazione rende più solida la struttura delle relazioni, ma complica la ricerca dei dati nel database. Però l'integrità e la consistenza dei dati sono prioritarie rispetto alla velocità di ritrovamento.

3FN – Terza forma normale

E' una relazione in seconda forma normale e inoltre tutti i suoi **attributi non-chiave dipendono direttamente dalla chiave primaria.**

Una relazione in 3FN quindi non possiede attributi che dipendono transitivamente dalla chiave primaria:

Es. $\underline{A1} \rightarrow A2$ $A2 \rightarrow A3$ quindi $\underline{A1} \rightarrow A3$

Perciò, tutti gli attributi non-chiave dipendono dalla chiave primaria. Non ci sono attributi non-chiave che dipendono da altri attributi non-chiave.

3FN – Esempio (1)

<u>CodStudente</u>	Nome	Scuola	TelefonoScuola

Il *Nome* dipende dallo studente (*CodStudente*): $\underline{A1} \rightarrow A2$.

La *Scuola* dipende dallo studente (*CodStudente*): $\underline{A1} \rightarrow A3$.

Il numero di telefono della scuola (*TelefonoScuola*) non dipende dallo studente, ma dalla *Scuola*: $A3 \rightarrow A4$ (e quindi $\underline{A1} \rightarrow A4$).

TelefonoScuola dipende quindi **transitivamente** dalla chiave.

3FN – Esempio (2)

<u>CodStudente</u>	Nome	Scuola	TelefonoScuola

Come visto, **la relazione non è in 3FN** e questo potrebbe provocare dei problemi:

- il numero di telefono di ogni scuola è ripetuto per ogni studente appartenente a quella scuola
- se il numero di telefono cambia, bisogna modificare tutte le righe degli studenti appartenenti alla scuola
- se in un periodo non ci fossero più studenti di una certa scuola iscritti all'associazione, non avremmo più il numero di telefono della scuola, con una perdita di informazioni

3FN – Esempio (3)

<u>CodStudente</u>	Nome	Scuola	TelefonoScuola



<u>CodStudente</u>	Nome	Scuola



<u>CodScuola</u>	TelefonoScuola

In conclusione

La normalizzazione evita rischi di ridondanza o inconsistenza nei dati, scomponendo le relazioni, ma nella relazione di partenza e in quella normalizzata **la quantità totale di informazioni è inalterata.**

Esistono forme di normalizzazione di ordine superiore al terzo, ma sono utilizzate solo in casi specifici.

Il modello relazionale **richiede obbligatoriamente la 1FN**, ma è consigliabile che le relazioni di ogni database siano normalizzate fino alla 3FN.

Fine

Gabriele Russo Russo