

Corso sulla Normalizzazione

Prime 3 forme normali

Autore: Biagioli Mirko

Premessa

Gli argomenti trattati nel presente documento sono rivolti agli studenti della classe quinta degli Istituti Tecnici Industriali (percorso Informatico) e Istituti tecnici commerciali (Ragionieri programmatori). Lo svolgimento del programma può essere articolato in un periodo minimo di 2 settimane, interamente tenuto in classe, considerando la disponibilità di 6 ore settimanali.

Obbiettivi:

Conoscenze: lo studente deve conoscere il concetto di ridondanza ed il perché il database deve essere normalizzato. Nozione di dipendenza funzionale per riconoscere la presenza di relazioni non normalizzate

Abilità: attraverso la ricerca delle dipendenze funzionali, saper ricondurre relazioni non normalizzate, in relazioni normalizzate nelle prime tre forme normali.

Prerequisiti:

Conoscenza dei database relazionali, schema concettuale e schema logico relazionale. Concetto di relazione, chiavi primarie e chiavi esterne.

Programma:

Unità Didattica	Durata	Argomenti	Modalità
1 Teoria	4 ore	Ridondanza, dipendenza funzionale e forme normali	Lezioni frontali, esempi
2 Esercizi	8 ore	Esercizi di normalizzazione	Esercizi in classe
3 Verifica formativa	2 ore	Prova di verifica	Compito scritto

Unità didattica 1: Teoria

La **normalizzazione** è un processo di semplificazione di un database per ottenere la struttura ottimale. Un database non normalizzato presenta problemi di:

- ridondanza
- inefficienza
- complessità
- perdita di informazioni

Si ha ridondanza dei dati ogni volta che vengono memorizzati inutilmente dei dati ripetuti. Si consideri ad esempio la seguente relazione che rappresenta i dati degli impiegati di un'azienda:

<u>Impiegato</u>	<u>Stipendio</u>	<u>Progetto</u>	<u>Bilancio</u>	<u>Funzione</u>
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

Le anomalie riguardano:

- Lo stipendio di ciascun impiegato è ripetuto in tutte le tuple (record) relative (RIDONDANZA)
- Se lo stipendio di un impiegato varia, è necessario andarne a modificare il valore in diverse tuple (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 ci sono queste anomalie?

Risposta: perché abbiamo utilizzato un'unica relazione per rappresentare informazioni eterogenee e cioè:

- Gli impiegati con i relativi stipendi
- I progetti con i relativi bilanci
- Le partecipazioni degli impiegati ai progetti con le relative funzioni

Per risolvere questi problemi occorre suddividere la relazione in più relazioni. Per rappresentare i diversi livelli di qualità rispetto allo schema logico sono state introdotte delle **FORME NORMALI**.

Prima forma normale:

Una tabella si dice in prima forma normale (1NF) se e solo se ciascun attributo è definito su un dominio di valori atomici. Vale a dire che se non c'è più di una colonna riferita allo stesso attributo allora è verificata la prima forma normale.

Esempio:

Relazione Libro

codice_libro	autore
1	Rossi, Bianchi, Verdi

L'attributo autore contiene una molteplicità di valori difficilmente gestibile. La relazione libro non è in prima forma normale.

Per normalizzare si devono sostituire gli attributi aggregati con un unico attributo e si aumentano le righe:

codice_libro	autore
1	Rossi
1	Bianchi
1	Verdi

Se l'attributo contiene informazioni eterogenee, per ogni informazione si crea un attributo, es:

codice_libro	Autore		Codice_libro	Autore	Indirizzo
1	Rossi, via roma 1	diventa	1	Rossi	Via roma 1

Le successive forme normali prevedono l'aumento delle tabelle, cioè le singole relazioni (o la singola relazione) vengono decomposte in più relazioni. Tale operazione viene fatta rispettando alcuni vincoli che fanno riferimento alla DIPENDENZA FUNZIONALE.

DIPENDENZA FUNZIONALE

Data una relazione $R(X)$ e due insiemi di attributi A e B non vuoti e sottoinsiemi di X si dice che B è funzionalmente dipendente da A se ogni combinazione di valori di A è associato ad una sola combinazione dei valori di B .

Se B dipende funzionalmente da A cioè $A \rightarrow B$, se due tuple hanno lo stesso valore di A devono avere lo stesso valore di B o gli stessi valori nel caso A o B siano composti da più attributi.

Esempio:

Stabiliamo che nella seguente relazione la chiave primaria sia Nome e Cognome.

ANAGRAFICA (Nome, Cognome, Indirizzo)

È evidente che Indirizzo è dipendente funzionalmente dalla chiave, ciò significa che per ogni persona corrisponde un indirizzo.

DIPENDENZA FUNZIONALE TRANSITIVA

Si verifica una dipendenza funzionale transitiva quando un attributo J appartenente alla relazione r dipende da un attributo K della relazione che non è chiave candidata e che a sua volta dipende da un attributo A , che è chiave candidata o primaria.

Data la relazione: $r(A, B, K, J)$ con A chiave primaria o candidata e avendo che: $A \rightarrow K$ e $K \rightarrow J$ si verifica una dipendenza transitiva tra A e J , in quanto J dipende da A tramite K .

Esempio:

Nella seguente relazione possiamo renderci conto come il valore di Collocazione dipenda da Genere e come il valore di Genere dipenda da Identificativo_videocassetta:

VIDEOCASSETTE (Identificativo_videocassetta, Genere, Collocazione)

Identificativo_videocassetta è la chiave primaria, essa identificando la videocassetta determina il Genere; le videocassette vengono collocate raggruppandole per genere.

Seconda forma normale:

Una relazione è in seconda forma normale se e solo se soddisfa la 1NF e inoltre ciascun attributo che non fa parte della chiave è indipendente funzionalmente da una qualunque chiave candidata.

Per normalizzare la relazione è necessario estrarre la dipendenza funzionale riscontrata tra un sottoinsieme della chiave e l'attributo in questione, che darà luogo ad un'altra relazione.

Se applichiamo quanto detto alla relazione $r(A, B, C, V, Z)$ con ABC chiave primaria, V dipende funzionale da ABC e Z dipende funzionale da AB, per normalizzare la relazione dobbiamo realizzare due nuove relazioni come segue:

$$r'(A, B, C, V) \text{ e } r''(A, B, Z)$$

Esempio:

Abbiamo una relazione che rappresenta delle scrivanie da ufficio prodotte da una azienda industriale, la chiave primaria è data Tipo_tavolo e Tipo_legno.

Tipo_tavolo	Tipo_legno	Tipo_finitura
Manager	Mogano	Ottone
Manager	Ciliegio	Acciaio inox
Montecarlo	Noce	Sughero
Montecarlo	Mogano	Ottone
Top	Noce	Sughero

Tipo_finitura è dipendente funzionalmente da un sottoinsieme della chiave primaria, ovvero da Tipo_legno, dunque la tabella non è in 2NF. È possibile normalizzare la relazione trasformandola in:

Tipo_tavolo	Tipo_legno
Manager	Mogano
Manager	Ciliegio
Montecarlo	Noce
Montecarlo	Mogano
Top	Noce

Tipo_legno	Tipo_finitura
Mogano	Ottone
Ciliegio	Acciaio inox
Noce	Sughero

Terza forma normale

La terza forma normale ha come scopo l'eliminazione dalla relazione delle dipendenze funzionali transitive.

Una relazione è in terza forma normale (3NF) se, oltre ad essere in 2NF, ciascun attributo che non partecipa alla chiave non è transitivamente dipendente da una qualunque chiave candidata.

Esempio:

Consideriamo la seguente relazione con codice_fattura chiave primaria:

FATTURA(codice_fattura, data, importo, codice_cliente, nome_cliente)

Si hanno le seguenti dipendenze funzionali:

codice_fattura -> codice_cliente

codice_cliente -> nome_cliente

quindi dipendenza funzionale transitiva:

codice_fattura -> codice_cliente -> nome_cliente

Per ricondurre alla 3NF occorre decomporre la relazione nel seguente modo:

FATTURA(codice_fattura, data, importo, codice_cliente)

CLIENTE(codice_cliente, nome_cliente)

Nota: codice_cliente in fattura sarà quindi chiave esterna per la relazione CLIENTE.

Unità didattica 2: Esercizi

In classe può essere proposto un esercizio completo di normalizzazione che affronti tutte le forme normali spiegate. Tale esercizio verrà realizzato dal docente in collaborazione con gli studenti, quindi risulterà importante stimolare l'individuazione dei problemi e la relativa risoluzione da parte degli alunni. Successivamente si chiederà agli studenti di risolvere autonomamente alcuni esercizi in previsione della prova di valutazione.

Esercizio proposto: Il database di uno Zoo

Ogni esemplare di animale ospitato dallo zoo è identificato dal suo nome comune e da un numero di inventario progressivo. Informazioni mantenute per ogni esemplare sono la data di arrivo nello zoo e la collocazione di destinazione. Schede anagrafiche riportano inoltre nome proprio, sesso, paese di provenienza e data di nascita di ogni esemplare, oltre ad un codice anagrafico unico.

Lo zoo è diviso in aree. In ogni area c'è un insieme di case, ognuna destinata ad un determinato genere di animali.

Ogni casa contiene un insieme di gabbie, ognuna contenente un esemplare specifico di animale.

Ogni casa ha un addetto e un responsabile.

Gli animali sono sottoposti periodicamente a controllo veterinario, come registrato su schede sanitarie. In un controllo, un veterinario rileva il peso degli esemplari, diagnostica un 'eventuale malattia e prescrive un tipo di dieta da seguire. Si suppone di raggruppare inizialmente le informazioni da mantenere nelle seguenti due tabelle:

ESEMPLARE (NomeComune, CodiceInventario, DataArrivo, Collocazione, CodiceAnagrafico)

ANAGRAFE (CodiceAnagrafico, NomeProprio, Sesso, DataNascita, LuogoNascita, SchedaSanitaria)

Nota: Collocazione e SchedaSanitaria sono *composti*:

Collocazione = Area x Casa x Gabbia x Addetto x Responsabile

SchedaSanitaria = DataControllo x Peso x Diagnosi x Dieta x Veterinario

SCHEMA in 1NF (1a Forma Normale)

ESEMPLARE (NomeComune, CodiceInventario, DataArrivo, CodiceAnagrafico, Area, Casa, Gabbia, Addetto, Responsabile)

ANAGRAFE (CodiceAnagrafico, NomeProprio, Sesso, DataNascita, LuogoNascita, DataControllo, Peso, Dieta, Diagnosi, Veterinario)

Si possono osservare le seguenti dipendenze funzionali (2NF) :

NomeComune -> Casa

NomeComune -> Addetto

NomeComune -> Area

Nome Comune -> Responsabile

NomeComune, Codice Inventario -> Gabbia (dipendenza funzionale completa su tutta la chiave primaria)

Decomponiamo la tabella ESEMPLARE in due tabelle:

ESEMPLARE (NomeComune, CodiceInventario, DataArrivo, Codice Anagrafico, Gabbia)

COLLOCAZIONE-GENERE (NomeComune, Area, Casa, Addetto, Responsabile)

Si hanno ancora anomalie:

Inserimento:

non è possibile inserire informazione sull'addetto assegnato ad una casa se non si conosce ancora il genere di animali destinato alla casa stessa

Cancellazione:

se un genere si estingue si perde ogni informazione sull'associazione tra la casa destinata ad esso e l'area in cui è dislocata

Modifica:

se cambia il responsabile di un'area bisogna modificare molte tuple.

Questi problemi sono dovuti alla presenza di dipendenze funzionali transitive:

NomeComune -> Casa -> Addetto, Area

NomeComune -> Area -> Responsabile

Quindi la relazione COLLOCAZIONE-GENERE verrà scomposta in:

COLLOCAZIONE-GENERE(NomeComune, Casa)

INFO-CASA (Casa, Area, Addetto)

INFO-AREA (Area, Responsabile)

Per quanto riguarda ANAGRAFE abbiamo:

Datacontrollo -> peso, dieta, diagnosi, veterinario

Quindi occorrerà effettuare la decomposizione:

ANAGRAFE(codice Anagrafico, Nome Proprio, sesso, data nascita, lugonascita)

SCHEDA-SANITARIA(codice Anagrafico, data controllo, peso, dieta, diagnosi, veterinario)

....

(altri esercizi da svolgere singolarmente con correzione collettiva)

Prova di valutazione

La prova di valutazione formativa dovrà contemplare il processo di normalizzazione di una relazione complessa in modo analogo all'esercizio collettivo svolto in classe.

Esempio

Procedere alla normalizzazione della seguente relazione:

CONTRAVVENZIONE(codice_multa, codice_agente, costo_multa, targa, cod_modello, desc_modello, cod_marca, desc_marca, nome_agente, data, ora, desc_multa)

Le ipotesi sono:

- il modello determina la marca
- il costo della multa può variare nel tempo.