



PROGETTAZIONE DI DATABASE

Le Forme Normali

Lezione 14



Questa dispensa è rilasciata sotto la licenza Creative Common CC BY-NC-SA. Chiunque può copiare, distribuire, modificare, creare opere derivate dall'originale, ma non a scopi commerciali, a condizione che venga riconosciuta la paternità dell'opera all'autore e che alla nuova opera vengano attribuite le stesse licenze dell'originale.

Versione del: 29/04/2015

Revisione numero: 3

Prof. Andrea Zoccheddu
Dipartimento di Informatica

Immagine di copertina da: <http://www.iamsterdam.com/en-GB/living/education/Dutch-Education-System>





LE FORME NORMALI

GENERALITÀ DELLE FORME NORMALI



Fonte: http://it.wikipedia.org/wiki/Normalizzazione_%28informatica%29

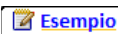


Una forma normale è una proprietà di una base di dati relazionale che ne garantisce la “qualità”.

Una forma normale quindi è una situazione o una condizione che la base di dati soddisfa e che afferma che essa si trova in una condizione ottimale.

Una base di dati si dice in specifica forma normale quando soddisfa la proprietà della specifica forma normale.

Quando una base di dati non è in una specifica forma normale è opportuno trasformarla in una base di dati equivalente che sia in specifica forma normale.



Consideriamo il seguente esempio di una relazione che presenta «**anomalie**»:

Progetti	Matricola	Nome	Classe	Progetto	Ruolo	Finanziamento
	101	Alex	5AI	Stazione Meteo	Progettista	500.00
	101	Alex	5AI	Drone postale	Collaudatore	2000.00
	103	Cloe	4BM	Stazione Meteo	Collaudatore	500.00
	103	Cloe	4BM	Drone postale	Progettista	2000.00

Si nota che:

- La classe di ciascuno studente è ripetuta in tutte le tuple pertinenti: presenza di ridondanza;
- Se il finanziamento di un progetto varia, è necessario modificare il valore nelle molte diverse tuple: anomalia di aggiornamento;
- Se uno studente interrompe la partecipazione a tutti i progetti, verrebbe rimosso dalla base di dati: anomalia di cancellazione;
- Un nuovo studente senza progetto non può essere inserito: anomalia di inserimento;

NORMALIZZAZIONE



Fonte: http://it.wikipedia.org/wiki/Normalizzazione_%28informatica%29



In informatica la normalizzazione è un procedimento volto all'eliminazione della ridondanza informativa e del rischio di incoerenza dal database.



La normalizzazione quindi è un procedimento formale che trasforma una base di dati in una equivalente che soddisfa la forma normale desiderata.

Esistono vari livelli di normalizzazione (forme normali) che certificano la qualità dello schema del database. Questo processo si fonda su un semplice criterio: se una relazione presenta più concetti tra loro indipendenti, la si decompone in relazioni più piccole, una per ogni concetto. Questo tipo di processo non è sempre applicabile in tutte le tabelle, dato che in taluni casi potrebbe comportare una perdita d'informazioni.

DIPENDENZE FUNZIONALI

Una dipendenza funzionale è un particolare vincolo di integrità per il modello relazionale che descrive legami di tipo funzionale tra gli attributi di una relazione o tabella.



Data una relazione r su uno schema relazionale $R(X)$ e due sottoinsiemi di attributi non vuoti Y e Z di X , si dice che esiste su r una dipendenza funzionale tra Y e Z se per ogni coppia di tuple t_1 e t_2 di r aventi gli stessi valori sugli attributi Y , t_1 e t_2 hanno gli stessi valori anche sugli attributi Z :

$$\forall t_1, t_2 \in r, t_1[Y] = t_2[Y] \rightarrow t_1[Z] = t_2[Z]$$

Una dipendenza funzionale tra gli attributi Y e Z viene generalmente indicata con la notazione $Y \rightarrow Z$ e come gli altri vincoli di integrità viene associata ad uno schema: una relazione su quello schema verrà considerata corretta se soddisfa tale dipendenza funzionale.



Esempio

Consideriamo il seguente esempio che presenta dipendenze funzionali:

Progetti	Matricola	Nome	Classe	Progetto	Ruolo	Finanziamento
	101	Alex	5AI	Stazione Meteo	Progettista	500.00
	101	Alex	5AI	Drone postale	Collaudatore	2000.00
	103	Cloe	4BM	Stazione Meteo	Collaudatore	500.00
	103	Cloe	4BM	Drone postale	Progettista	2000.00

Si nota che ma matricola determina con sicurezza il nome dello studente e questo determina la classe; il progetto determina il finanziamento.

Si può allora affermare che nella relazione Progetti esistono le seguenti dipendenze funzionali

- Matricola \rightarrow {Nome , Classe }
- Nome \rightarrow Classe
- Progetto \rightarrow Finanziamento

DECOMPOSIZIONE DI UNA RELAZIONE

Una decomposizione di una relazione è la sua trasformazione in una relazione equivalente; la decomposizione deve garantire una equivalenza della relazione decomposta e in particolare deve rispondere alle seguenti caratteristiche:

- la decomposizione senza perdita, che garantisce la ricostruzione delle informazioni originarie
- la conservazione delle dipendenze, che garantisce il mantenimento dei vincoli di integrità originari

Per "decomposizione senza perdita" si intende l'atto della manipolazione di una relazione R volta ad ottenere (eventualmente) due o più relazioni (ad esempio R1 e R2) che oltre a conservare le dipendenze funzionali verificano anche la seguente condizione: $R = R1 \text{ join } R2$

Per "conservazione delle dipendenze" si intende la possibilità di dedurre il valore di un dominio dipendente funzionalmente dall'altro in almeno una relazione che forma la decomposizione.

PRIMA FORMA NORMALE (1NF)

Definizione

Si dice che una relazione è in 1NF se e solo se:

- non presenta gruppi di attributi che si ripetono (ossia ciascun attributo è definito su un dominio con valori atomici)
- esiste una chiave primaria (ossia esiste un insieme di attributi, che identifica in modo univoco ogni tupla della relazione)

In sintesi si può affermare che una relazione è in 1NF se tutti i suoi attributi sono semplici

Definizione

Si dice che una base dati è in 1NF (prima forma normale) se ogni sua relazione (tabella) è in 1NF

Intuitivamente si può affermare che:

- In prima forma normale, tutti i record di una tabella devono contenere lo stesso numero di campi.
- La prima forma normale non ammette che in un record ci siano campi opzionali o multipli.

Violazione della 1NF (atomicità dei valori)

Esempio

Il seguente esempio viola la 1NF,

Verifiche	Studente	Materia	Docente	Voti
	Ada	Storia	Carta Chiara	3½; 9;
	Ada	Inglese	Meloni Gustavo	6; 7½;
	Zoe	Storia	Corrente Daria	6½; 8;
	Zoe	Inglese	Piano Guido	5; 8½;

perché pur esistendo sulla relazione una chiave primaria ({Studente, Materia}), l'attributo Voti non è definito su un dominio con valori atomici



Decomposizione in 1NF (separazione di relazioni)

Per portare la relazione in 1NF occorre dividere la relazione in due relazioni separate; la chiave primaria viene riportata separatamente e l'attributo non atomico diventa atomico e presenta valori separati in tuple distinte.

Esempio

Il seguente esempio porta la relazione precedente in 1NF, la relazione è scissa in una prima relazione senza l'attributo non atomico:

Verifiche	Studente	Materia	Docente
	Ada	Storia	Carta Chiara
	Ada	Inglese	Meloni Gustavo
	Zoe	Storia	Corrente Daria
	Zoe	Inglese	Piano Guido

ed una seconda relazione che riporta la chiave primaria e con l'attributo diventa atomico e i valori multipli ripetuti in più tuple distinte:

Voti	Studente	Materia	Voto
	Ada	Storia	3½
	Ada	Storia	9
	Ada	Inglese	6;
	Ada	Inglese	7½
	Zoe	Storia	6½
	Zoe	Storia	8
	Zoe	Inglese	5
	Zoe	Inglese	8½

e per ricostruire la relazione originale è possibile compiere una join nella forma **Verifiche Join Voti**

SECONDA FORMA NORMALE (2NF)

Definizione

Si dice che una relazione è in 2NF se e solo se:

- è in 1NF
- ogni attributo non-primario (ovvero non appartenente a nessuna chiave) dipende completamente da ogni chiave (ovvero non dipende solamente da una parte di chiave)

Definizione

In sintesi una relazione è in seconda forma normale se, oltre a rispettare la prima forma normale, nessun campo della tabella, tra quelli che non fanno parte della chiave primaria, dipende solo da alcuni campi della chiave primaria.

Violazione della 2NF (dipendenza completa dalla chiave)

La seconda forma normale è violata quando un attributo non chiave riguarda solo un sottoinsieme della chiave primaria.

La seconda forma normale è rilevante solo per tabelle la cui **chiave è composta**, cioè consiste di diversi attributi. Per esempio non riguarda i casi in cui la chiave è formata da un solo attributo atomico come un contatore.

Tuttavia spesso queste tabelle sono quelle che si ottengono traducendo le associazioni molti a molti ovvero quelle che rappresentano delle "relazioni fra entità" nella progettazione concettuale.

Esempio

Il seguente esempio viola la 2NF,

Verifiche	Studente	Materia	Docente
	Ada	Storia	Carta Chiara
	Ada	Inglese	Meloni Gustavo
	Zoe	Storia	Carta Chiara
	Zoe	Inglese	Meloni Gustavo

perché la chiave è formata dalla coppia {Studente, Materia} ma il docente dipende funzionalmente dalla sola materia cioè: {Materia} → {Docente}

Decomposizione in 2NF

Per portare la relazione in 2NF occorre dividere la relazione in due relazioni separate; la prima relazione conserva solo gli attributi che dipendono completamente dalla chiave, mentre quelli che dipendono solo da parte della chiave si riportano in una relazione separata.



Esempio

Il seguente esempio porta la relazione precedente in 2NF, la relazione è scissa in due relazioni di cui una prima relazione senza l'attributo dipendente non completamente dalla chiave e la seconda che riporta la dipendenza sottochiave e attributo:

Verifiche	Studente	Materia	Docenti	Materia	Docente
	Ada	Storia		Storia	Carta Chiara
	Ada	Inglese		Inglese	Meloni Gustavo
	Zoe	Storia			
	Zoe	Inglese			

e per ricostruire la relazione originale è possibile compiere una join nella forma **Verifiche Join Docenti**

Rischi nella decomposizione in 2NF (perdita di informazioni)

Nella normalizzazione si deve fare attenzione a non perdere informazioni.

Esempio

Si consideri l'esempio seguente, (si assume che una materia abbia sempre lo stesso numero di ore); la chiave è costituita dalla tripla { Docente, Classe, Materia }; esisto però le seguenti dipendenze funzionali di attributi parte della chiave:

- {Docente} → {Materia}
- {Materia} → {Ore}

Cattedre	Docente	Classe	Materia	Ore
	Prof.ssa Vacca Carla	5AI	Storia	2
	Prof. Meloni Gustavo	5AI	Inglese	3
	Prof.ssa Vacca Carla	4BM	Storia	2
	Prof. Meloni Gustavo	4BM	Inglese	3

se la relazione è scissa in due relazioni potremmo ottenere:

Cattedre	Docente	Classe	Orari	Materia	Ore
	Prof.ssa Vacca Carla	5AI		Storia	2
	Prof. Meloni Gustavo	5AI		Inglese	3
	Prof.ssa Vacca Carla	4BM			
	Prof. Meloni Gustavo	4BM			

che è in 2NF ma non conserva le informazioni corrette (non sappiamo più quale materia insegna un docente). La soluzione è la decomposizione della relazione in tre relazioni distinte, con la parte della chiave esplicitata:

Cattedre	Docente	Classe	Orari	Materia	Ore	Materie	Docente	Materia
	Prof.ssa Vacca Carla	5AI		Storia	2		Prof.ssa Vacca Carla	Storia
	Prof. Meloni Gustavo	5AI		Inglese	3		Prof. Meloni Gustavo	Inglese
	Prof.ssa Vacca Carla	4BM						
	Prof. Meloni Gustavo	4BM						

che permette di ricostruire tutte le informazioni originarie.

TERZA FORMA NORMALE (3NF)

Definizione

Si dice che una relazione è in 3NF se e solo se:

- è in 2NF
- ogni attributo non-primario (ovvero non appartenente a nessuna chiave) non dipende funzionalmente da altri attributi che siano non-primari (ovvero non appartenenti a nessuna chiave)

Definizione

In sintesi una relazione è in terza forma normale se, oltre a rispettare la seconda forma normale, nessun campo della tabella, tra quelli che non fanno parte della chiave primaria, dipende altri da campi che non sono chiave primaria.



Violazione della 3NF (dipendenza non transitiva dalla chiave)

La terza forma normale è violata quando un attributo non chiave dipende da un attributo non chiave. Se la progettazione concettuale è corretta è spesso tradotta direttamente in 3NF.

Esempio

Si consideri l'esempio seguente, (si assume che una materia abbia sempre lo stesso numero di ore); la chiave è costituita dal solo attributo ID; però esistono le seguenti dipendenze funzionali:

- {Classe, Materia} → {Docente}
- {Classe} → {Aula}
- {Materia} → {Ore}

R	ID	Studente	Classe	Materia	Docente	Ore	Aula
	94	Ada	5AI	Storia	Vacca C.	2	A101
	95	Ugo	5AI	Inglese	Piano G.	3	A101
	96	Zoe	4AI	Storia	Vacca C.	2	A304
	97	Leo	4AI	Inglese	Piano G.	3	A304

Per cui la relazione R non è in 3NF a causa di dipendenze da attributi non chiave.

Decomposizione in 3NF

Per portare la relazione in 3NF occorre dividere la relazione in due relazioni separate; la prima relazione conserva solo gli attributi che dipendono dalla chiave, mentre quelli che dipendono solo da altri attributi si riportano in una relazione separata.

Esempio

La soluzione è la decomposizione della relazione in relazioni distinte, con la dipendenza esplicitata:

R	ID	Studente	Classe	S	Classe	Materia	Docente	T	Materia	Ore	U	Classe	Aula
	94	Ada	5AI		5AI	Storia	Vacca C.		Storia	2		5AI	A101
	95	Ugo	5AI		5AI	Inglese	Piano G.		Inglese	3		4AI	A304
	96	Zoe	4AI		4AI	Storia	Vacca C.						
	97	Leo	4AI		4AI	Inglese	Piano G.						

ed è possibile ricomporre la relazione originale dal join delle relazioni ottenute dalla decomposizione.

Si è tralasciata una possibile dipendenza funzionale {Studente} → {Classe} che è lasciata per esercizio.



SOMMARIO

LE FORME NORMALI	2
GENERALITÀ DELLE FORME NORMALI	2
Normalizzazione.....	2
DIPENDENZE FUNZIONALI	2
Decomposizione di una relazione.....	3
Prima Forma Normale (1NF).....	3
Violazione della 1NF (atomicità dei valori).....	3
Decomposizione in 1NF (separazione di relazioni).....	4
Seconda Forma Normale (2NF).....	4
Violazione della 2NF (dipendenza completa dalla chiave).....	4
Decomposizione in 2NF.....	4
Rischi nella decomposizione in 2NF (perdita di informazioni).....	5
Terza Forma Normale (3NF).....	5
Violazione della 3NF (dipendenza non transitiva dalla chiave).....	6
Decomposizione in 3NF.....	6