# LA NORMALIZZAZIONE (p.270)

Simone Remoli

Dipendenza funzionale: La dipendenza funzionale è un particolare vincolo di integrità per il modello relazionale che descrive legami di tipo funzionale tra gli attributi di una relazione.

#### Esempio immediato per capire.

Consideriamo una relazione che ha delle anomalie:

<u>IMPIEGATO</u>	STIPENDIO	PROGETTO	BILANCIO	FUNZIONE
Rossi	20000	Marte	2000	Tecnico
Verdi	35000	Giove	15000	Progettista
Verdi	35000	Venere	15000	Progettista
Neri	55000	Venere	15000	Direttore
Neri	55000	Giove	15000	Consulente
Neri	55000	Marte	2000	Consulente
Mori	48000	Marte	2000	Direttore
Mori	48000	Venere	15000	Progettista
Bianchi	48000	Venere	15000	Progettista
Bianchi	48000	Giove	15000	Direttore

(IMPIEGATO, PROGETTO) rappresentano una chiave di questa relazione.

È importante sottolineare che è stata usata un'unica relazione per rappresentare informazioni eterogenee.

Questo è il problema principale.

Individuiamo le dipendenza funzionali valide di questa relazione.

# 1. Dipendenza Funzionale: IMPIEGATO $\rightarrow$ STIPENDIO.

Il valore dell'attributo impiegato determina il valore dell'attributo stipendio. Questo accade per ogni impiegato.

#### 2. Dipendenza Funzionale: PROGETTO → BILANCIO.

Il valore dell'attributo progetto determina il valore dell'attributo bilancio. Questo accade per ogni progetto.

Queste due dipendenza funzionali prendono il nome di "dipendenze funzionali non banali": in generale, una dipendenza funzionale Y → A è non banale se A non compare tra gli attributi di Y.

E d'ora in poi farò riferimento solo a dipendenze funzionali non banali.

#### VINCOLO DI CHIAVE

Ultima osservazione prima di entrare nel vivo della spiegazione:
Se prendiamo una chiave K di una certa relazione, si può facilmente verificare che esiste una dipendenza funzionale tra K e ogni altro attributo della relazione.

Questo perché, per definizione stessa di vincolo di chiave, non possono esistere tuple con gli stessi valori su K, quindi una dipendenza funzionale che ha K al primo membro sarà <u>sempre</u> soddisfatta.

Nel nostro esempio abbiamo detto che gli attributi impiegato e progetto formano una chiave, quindi possiamo affermare che esisterà una dipendenza funzionale tra una chiave di una relazione e tutti gli attributi della relazione (esclusi quelli della chiave stessa). Pertanto, possiamo affermare con certezza che esiste questa dipendenza funzionale:

#### <u>IMPIEGATO PROGETTO</u> → STIPENDIO BILANCIO FUNZIONE

Quindi più precisamente possiamo dire che una dipendenza funzionale  $Y \to Z$  su uno schema R(X) degenera nel vincolo di chiave se  $Y \cup Z = X$ .

In tal caso infatti, Y è (<u>super</u>)chiave per lo schema R(X).

Ma prestiamo molta attenzione alla dipendenza IMPIEGATO PROGETTO → FUNZIONE.

Questa non genera anomalie.

La differenza, come accennato, risiede nel fatto che IMPIEGATO PROGETTO è una SUPERCHIAVE della relazione.

#### A questo punto la domanda diventa legittima:

"Le ridondanze e le anomalie da quali dipendenze funzionali sono causate?" Possiamo quindi concludere che le ridondanze e le anomalie sono causate dalle dipendenze funzionali  $X \to A$  che permettono la presenza di più tuple fra loro uguali, cioè, in altre parole, dalle dipendenze funzionali  $X \to A$  tali che X NON CONTIENE una chiave.

# FORMA NORMALE DI BOYCE E CODD.

Una relazione R è in forma normale di Boyce e Codd se **per ogni dipendenza funzionale non banale** X → A definita su di essa, X contiene

una certa chiave K di R, cioè X è superchiave per R.

(Ci devono essere solo queste dipendenze funzionali valide).

Anomalie e ridondanze, in questo modo, non si presentano per relazioni in forma normale di Boyce e Codd.

#### DECOMPOSIZIONE IN FORMA NORMALE DI BOYCE E CODD.

Data una relazione che non soddisfa la forma normale di Boyce e Codd è possibile sostituirla con due o più relazioni normalizzate attraverso un processo di normalizzazione.

Criterio: Se una relazione rappresenta più concetti indipendenti, allora va decomposta in relazioni più piccole, una per ogni concetto.

Quindi, nel nostro esempio, la tabella iniziale viene decomposta, si passa da

<u>IMPIEGATO</u>	STIPENDIO	PROGETTO	BILANCIO	FUNZIONE
Rossi	20000	Marte	2000	Tecnico
Verdi	35000	Giove	15000	Progettista
Verdi	35000	Venere	15000	Progettista
Neri	55000	Venere	15000	Direttore
Neri	55000	Giove	15000	Consulente

<u>IMPIEGATO</u>	STIPENDIO	PROGETTO	BILANCIO	FUNZIONE
Neri	55000	Marte	2000	Consulente
Mori	48000	Marte	2000	Direttore
Mori	48000	Venere	15000	Progettista
Bianchi	48000	Venere	15000	Progettista
Bianchi	48000	Giove	15000	Direttore

# A:

<u>IMPIEGATO</u>	STIPENDIO
Rossi	20000
Verdi	35000
Neri	55000
Mori	48000
Bianchi	48000

PROGETTO	BILANCIO
Marte	2000
Giove	15000
Venere	15000

<u>IMPIEGATO</u>	PROGETTO	FUNZIONE
Rossi	Marte	Tecnico
Verdi	Giove	Progettista
Verdi	Venere	Progettista
Neri	Venere	Direttore
Neri	Giove	Consulente
Neri	Marte	Consulente
Mori	Marte	Direttore
Mori	Venere	Progettista
Bianchi	Venere	Progettista
Bianchi	Giove	Direttore

Se alla tabella iniziale sostituiamo queste tre relazioni, eliminiamo anomalie e ridondanze: le tre relazioni sono in forma normale di Boyce e Codd.

In effetti, in molti casi pratici, la decomposizione può essere effettuata producendo <u>tante relazioni quante sono le dipendenze funzionali con diverso primo membro</u>.



Presented with xmind

#### DECOMPOSIZIONE SENZA PERDITA.

#### Sia data la seguente relazione:

IMPIEGATO	PROGETTO	SEDE
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Per discutere questa proprietà prendiamo le dipendenze funzionali. In questo esempio ne esistono due:

 $\begin{array}{c} \text{IMPIEGATO} \rightarrow \text{SEDE} \\ \text{PROGETTO} \rightarrow \text{SEDE} \end{array}$ 

Queste sono dipendenze funzionali valide.

Operando come nei casi precedenti, separando cioè sulla base delle dipendenze, saremmo portati a decomporre le relazione in due parti:

IMPIEGATO	SEDE
Rossi	Roma
Verdi	Milano
Neri	Milano

PROGETTO	SEDE
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

In questo caso c'è un problema: dalla tabella originale, Verdi era assegnato solo al progetto Giove e Venere.

Se ora si provasse a ricostruire le informazioni circa la partecipazione degli impiegati ai progetti, l'unica possibilità che si avrebbe per poterlo fare sarebbe quella di utilizzare l'attributo "SEDE" (unico attributo <u>comune</u> alle due relazioni).

Possiamo correlare un impiegato a un progetto se il progetto viene svolto nella sede presso cui l'impiegato opera.

Purtroppo, però, in questo caso non riusciamo a ricostruire tutte e sole le informazioni nella relazione originaria: per esempio l'impiegato verdi lavora a Milano e il progetto Saturno viene svolto presso la sede di Milano, ma verdi non lavora a tale progetto, come detto prima.

Abbiamo effettuato una decomposizione con perdita.

Possiamo generalizzare l'osservazione notando come la ricostruzione della relazione originaria a partire dalle sue due relazioni disaccoppiate debba intuitivamente essere effettuata per mezzo di un'operazione di join naturale delle due proiezioni. Purtroppo, il join naturale delle due relazioni produce la seguente relazione, che è ben diversa da quella originale.

IMPIEGATO	PROGETTO	SEDE
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Verdi	Saturno	Milano
Neri	Giove	Milano

È chiaramente desiderabile, anzi, è un requisito irrinunciabile, che una decomposizione effettuata ai fini di normalizzazione sia senza perdita.

# CONDIZIONE CHE GARANTISCE UNA DECOMPOSIZIONE CON PERDITA.

Sia R una relazione su un numero di attributi X e siano X1 e X2 attributi sottoinsiemi di X tali che X1 ∪ X2 = X;

Se questo è vero, allora il Join delle due relazioni (ottenute per proiezione da R su X1 e X2) è una relazione che contiene tutte le tuple di R, più eventualmente altre che possiamo chiamare "spurie".

Quindi diciamo che R si decompone senza perdita su X1 e X2 se il join delle due relazioni è uguale a R stessa (cioè non contiene tuple spurie).

### CONDIZIONE CHE GARANTISCE UNA DECOMPOSIZIONE SENZA PERDITA.

Sia R una relazione su un numero di attributi X, e siano X1 e X2 attributi sottoinsiemi di X tali che X1 ∪ X2 = X;

Sia  $X0 = X1 \cap X2$ ;

Allora R si decompone senza perdita su X1 e X2 se soddisfa la dipendenza funzionale  $X0 \rightarrow X1$  oppure la dipendenza funzionale  $X0 \rightarrow X2$ .

#### ESEMPIO PRATICO:

Supponiamo di avere una relazione R(A,B,C). Dipendenza funzionali attive: A → B, B → C. Supponiamo di avere una decomposizione R1(A,B) e R2(B,C).

$$X0 = B.$$

$$X1 = A, X2 = B.$$

Allora R si decompone senza perdita se soddisfa la dipendenza funzionale  $B \to A$ .

In altre parole, R si decompone senza perdita su due relazioni se l'insieme degli attributi comuni alle due relazioni, ossia l'intersezione, (in questo caso B), è chiave per almeno una delle relazioni decomposte.

#### Esempio con perdita.

BCNF richiede che in ogni relazione, ogni determinante sia una chiave candidata. Quindi per ogni dipendenza funzionale che viola questa regola, devi decomporre.

Hai R(A, B, C, D) con:

- $A \rightarrow B$
- $A \rightarrow C$
- $C \rightarrow D$

Supponi che A sia chiave, ma C non lo sia  $\Rightarrow$  la dipendenza C  $\rightarrow$  D viola BCNF.

Allora devi **decomporre** R in:

- R1(C, D) (per rappresentare  $C \rightarrow D$ )
- **R2(A, B, C)** (per rappresentare il resto, includendo la chiave A)

Risultato: La decomposizione ha perdita perché C non è chiave in nessuna delle due relazioni!

Anche se la dipendenza C → D è valida, **non basta** per garantire una

decomposizione senza perdita se C non è chiave almeno in una delle relazioni risultanti.

#### Esempio senza perdita.

Sia R(A, B, C) con questa dipendenza funzionale:  $A \rightarrow B$ . E supponiamo che A sia chiave per R.

Decomponiamo R in:  $R1(\underline{A}, B)$ ,  $R2(\underline{A}, C)$ .

$$X0 = A$$
.

A è chiave in entrambe le tabelle, quindi la decomposizione è senza perdita.

Anche se in questo caso **non serviva** normalizzare, era utile per **capire il principio**.

#### Quando NON devi normalizzare ulteriormente in BCNF?

Hai una relazione R con: chiave candidata K, una dipendenza non banale K → B
Allora:

Non viola la forma normale di Boyce-Codd, perché il lato sinistro della dipendenza è una chiave, quindi è una superchiave.

Puoi disaccoppiare sulla chiave primaria solo se il tuo obiettivo non è la normalizzazione ma garantire un join senza perdita.

#### CONSERVAZIONE DELLE DIPENDENZE.

Consideriamo nuovamente la seguente relazione:

IMPIEGATO	PROGETTO	SEDE
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano

IM	MPIEGATO	PROGETTO	SEDE
Ne	eri	Saturno	Milano
Ne	eri	Venere	Milano

Con le seguenti dipendenze: IMPIEGATO  $\rightarrow$  SEDE PROGETTO  $\rightarrow$  SEDE.

Volendo ancora rimuovere le anomalie, potremmo pensare di sfruttare solo la dipendenza IMPIEGATO → SEDE per ottenere una decomposizione senza perdita.

Potremmo procedere anche utilizzando solo l'altra dipendenza,  $PROGETTO \rightarrow SEDE$ .

Il risultato sono due tabelle separate ma senza nessuna perdita di informazione: vediamolo in azione sfruttando solo la prima dipendenza.

<u>IMPIEGATO</u>	SEDE
Rossi	Roma
Verdi	Milano
Neri	Milano

IMPIEGATO	PROGETTO
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

Impiegato è chiave per almeno una delle due relazioni, quindi la decomposizione è senza perdita.

Impiegato è chiave per la prima relazione, quindi non solo esiste la decomposizione senza perdita ma essa è anche garantita: il join delle due relazioni produce la relazione originale.

Ma questa decomposizione presenta un solo piccolo inconveniente:

supponiamo di voler inserire una nuova tupla che specifica la partecipazione dell'impiegato Neri, che opera a Milano, al progetto Marte.

Questo, nella relazione originale, è inammissibile! Nella tabella originale, un tale aggiornamento verrebbe immediatamente individuato come ILLECITO.

Il motivo?

Porterebbe alla violazione della dipendenza PROGETTO  $\rightarrow$  SEDE.

Sulle relazioni decomposte, invece, non è possibile rilevare alcuna violazione di dipendenze: possiamo quindi notare come non sia possibile effettuare alcuna verifica sulla dipendenza PROGETTO → SEDE, perché i due attributi PROGETTO E SEDE sono proprio separati: uno in una relazione e l'altro nell'altra.

Conclusione: Per ottenere il soddisfacimento dei vincoli sulle tabelle decomposte dei vincoli sulla tabella originale, ciascuna delle dipendenze funzionali sullo schema originario dovrebbe coinvolgere attributi che compaiono tutti insieme in uno degli schemi decomposti.

Teorema: <u>Una decomposizione che soddisfa tale proprietà conserva le dipendenze dello</u> schema originario.

Quindi, in fin dei conti, d'ora in poi l'attività di normalizzazione sarà volta ad ottenere una decomposizione che sia senza perdita e che conservi le dipendenze.

# TERZA FORMA NORMALE

(E la seconda? La enuncio dopo perché c'è un motivo, nulla è a caso).

Sarò diretto: talvolta la forma normale di Boyce e Codd non è raggiungibile.

#### Consideriamo la seguente relazione:

DIRIGENTE	PROGETTO	SEDE	
Rossi	Marte	Roma	
Verdi	Giove	Milano	
Verdi	Marte	Milano	
Neri	Saturno	Milano	
Neri	Venere	Milano	

Esistono queste dipendenze funzionali attive:

DIRIGENTE → SEDE: Ogni dirigente opera presso una sede;
 PROGETTO SEDE → DIRIGENTE: Ogni progetto ha più dirigenti che sono responsabili, ma in sedi diverse. Per ogni sede, un progetto ha solo un unico dirigente, quindi sottolineiamo che questa è una dipendenza funzionale attiva.

Quindi, ripetiamo, per tutto quello che sin ora è stato detto, la relazione non è in forma normale di Boyce e Codd perché il primo elemento della prima dipendenza funzionale non è superchiave.

Inoltre, la dipendenza **PROGETTO SEDE** → **DIRIGENTE** coinvolge tutti gli attributi della relazione e pertanto nessuna decomposizione è in grado di conservarla.

Quindi, la forma normale di Boyce e Codd non è raggiungibile.

Questo schema però soddisfa la terza forma normale.

# DEFINIZIONE DI TERZA FORMA NORMALE.

Una relazione R è in terza forma normale se, per ogni dipendenza funzionale attiva non banale X → A definita su di essa, \*almeno\* una delle seguenti condizioni è verificata:

- X contiene una chiave K di R.
- A appartiene ad almeno una chiave di R.

Una relazione R che non soddisfa la terza forma normale si decompone in relazioni ottenute sulle dipendenze funzionali, con l'unica accortezza di mantenere sempre una relazione che contiene una chiave della relazione originaria.

Quindi, ritornando al nostro esempio, la dipendenza **DIRIGENTE** → **SEDE** ha un unico attributo a secondo membro che fa parte della chiave Progetto Sede. Inoltre, la dipendenza **PROGETTO SEDE** → **DIRIGENTE** ha come primo membro una chiave della relazione.

Resta il fatto che, in realtà, la relazione presenta una ridondanza perché ogni volta che un dirigente compare in una tupla, viene ripetuta per esso la sede in cui opera.

Ma, questa ridondanza viene "tollerata" dalla terza forma normale questo perché non sarebbe possibile una decomposizione che elimini tale ridondanza e al tempo stesso conservi tutte le dipendenze.

(In quel caso però sarebbe una decomposizione senza perdita).

Teorema: La terza forma normale, rispetto alla Boyce e Codd, è sempre ottenibile. Una qualunque relazione che non soddisfa la terza forma normale è decomponibile senza perdita e con conservazione delle dipendere in relazioni in terza forma normale.

DECOMPOSIZIONE IN TERZA FORMA NORMALE.

Consideriamo una relazione fatta in questo modo:

<u>IMPIEGATO</u>	PROGETTO	STIPENDIO
Rossi	Marte	30.000
Verdi	Giove	30.000
Verdi	Venere	30.000
Neri	Saturno	40.000
Neri	Venere	40.000

#### Vale la sola dipendenza IMPIEGATO $\rightarrow$ STIPENDIO.

Effettuare una decomposizione per questa dipendenza funzionale e in un'altra sul solo attributo progetto, violerebbe la proprietà di decomposizione senza perdita, proprio perché nessuna delle due relazioni contiene una chiave.

#### La decomposizione può essere fatta così:

<u>IMPIEGATO</u>	STIPENDIO
Rossi	30.000
Verdi	30.000
Neri	40.000

<u>IMPIEGATO</u>	<u>PROGETTO</u>
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

#### In questo modo non ho perdita.

Osservazione: Una decomposizione tesa a ottenere la terza forma normale produce nella maggior parte dei casi schemi in forma normale di Boyce e Codd. Se una relazione ha solo una chiave allora la forma normale di Boyce e Codd e la terza forma normale coincidono.

#### ESERCIZIO 5 - PROVA D'ESAME DEL 23 FEBBRAIO 2024

Considerare la seguente relazione che contiene informazioni relative ad alcuni giocatori di calcio:

Cod	Cognome	Nome	CodRuolo	Ruolo	CodNaz	Nazione	DataNascita	Presenze
342	Rossi	Mario	A	Attaccante	I	Italia	11/02/1976	143
342	Rossi	Mario	C	Centrocampista	I	Italia	11/02/1976	143
522	Rossi	Luca	A	Attaccante	I	Italia	11/02/1976	45
213	Bruni	Piero	P	Portiere	I	Italia	20/01/1974	143
425	Santos	Joao	D	Difensore	BR	Brasile	21/03/1979	65
425	Santos	Joao	C	Centrocampista	BR	Brasile	21/03/1979	65

Individuare la chiave (o le chiavi) della relazione e le dipendenze funzionali definite su di essa e spiegare perché essa non soddisfa la BCNF.

Decomporla in BCNF nel modo che si ritiene più opportuno indicando eventuali vincoli.

Svolgimento.

Le dipendenze funzionali sono:

 $\label{eq:codRuolo} \text{CodRuolo}, \ \text{Ruolo} \to \text{CodRuolo}.$   $\ \text{CodNaz} \to \text{Nazione}, \ \text{Nazione} \to \text{CodNaz}.$   $\ \text{Cod} \to \text{Cognome Nome CodNaz Nazione DataNascita Presenze}.$ 

Decomponiamo le tabelle:

Ruolo(<u>CodRuolo</u>, Ruolo) Nazione(<u>CodNaz</u>, Nome)

Giocatore (<u>Cod</u>, Nome, Cognome, <u>Nazionalità</u>, Data Nascita, Presenze).

Gioca (<u>Giocatore</u>, <u>Ruolo</u>).

Chiavi: (Cod, CodRuolo), (Cod, Ruolo)

#### ESERCIZIO 5 - PROVA D'ESAME DEL 18 FEBBRAIO 2025

Si consideri il seguente schema di relazione:

Esami (Matricola Studente, Nome Studente, Cognome Studente, Corso, Professore, Voto, Data)

con i vincoli:

- (i ) possono esistere studenti omonimi;
- (ii) ogni studente può registrare un unico voto per ogni corso;
  - (iii ) corsi diversi hanno professori tutti diversi;
- (iv ) ogni corso può essere tenuto da più professori, ognuno tiene un canale e gli studenti sono assegnati ai canali sulla base del loro cognome (ad es. A-L e M-Z).

Si individuino le dipendenze funzionali e le chiavi.

Indicare se lo schema è in BCNF illustrando le motivazioni.

Svolgimento.

#### Creiamo lo scenario simpatico :)

MatricolaStud ente	NomeStude nte	CognomeSt udente	Corso	Professore	Voto	Data
1	Simone	Remoli	Calcolatori Elettronici	Pellegrini	30	10/10/2010
2	Alessandra	Neri	Ricerca Operativa	Oriolo	30L	12/12/2012
1	Simone	Remoli	Statistica	Torti	30	14/12/2009
1	Simone	Remoli	Controlli	Possieri	30	13/11/2007
3	Marco	Bianchi	Calcolatori Elettronici	Pellegrini	25	12/10/2009
4	Marco	Bianchi	Calcolatori Elettronici	Pellegrini	18	13/08/2007
3	Marco	Bianchi	ISPW	Falessi	30	14/03/1999
5	Matteo	Salvini	Ignoranza	La Russa	30L Bacio Accademico	11/02/2002
4	Marco	Bianchi	IIW	Lo Presti	27	10/10/2010

# Le dipendenze funzionali sono:

$$\label{eq:studente} \begin{split} \text{MatricolaStudente} & \to \text{NomeStudente CognomeStudente.} \\ & \text{Corso} \to \text{Professore} \\ & \text{MatricolaStudente Corso} \to \text{Voto Data} \\ & \text{CognomeStudente Corso} \to \text{Professore} \end{split}$$

 $Le\ chiavi\ sono: (Matricola Studente, Professore)\ e\ (Matricola Studente, Corso).$ 

#### Fine del documento.

Gli esercizi sulla normalizzazione sono molto effimeri a livello di durata e questo prelude una grande possibilità di guadagnare punteggio in sede d'esame.