

Basi di Dati

Normalizzazione

Marco Maratea

Laurea in Informatica, DeMaCS, UNICAL

4 Dicembre 2025



Forme normali

- Una forma normale è una proprietà di una base di dati relazionale che ne garantisce la “qualità”, cioè l'assenza di determinati difetti
- Quando una relazione non è normalizzata:
 - presenta ridondanze,
 - si presta a comportamenti poco desiderabili durante gli aggiornamenti
- Le forme normali sono di solito definite sul modello relazionale (ma applicabili anche in altre fasi, ad esempio il modello E-R)



Normalizzazione

- Procedura che permette di trasformare tabelle non normalizzate in tabelle che soddisfano una forma normale
- La normalizzazione va utilizzata come **tecnica di verifica** dei risultati della progettazione di una base di dati
- **Non costituisce una metodologia di progettazione**

Una relazione di esempio

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

Anomalie della relazione



- Lo stipendio di ciascun impiegato è ripetuto in tutte le ennuple relative
 - **ridondanza**
- Se lo stipendio di un impiegato varia, è necessario andarne a modificare il valore in diverse ennuple
 - **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**

Esistono anomalie simili alle prime due tra Progetto e Bilancio

Perché questi fenomeni indesiderabili?



- abbiamo usato un'unica relazione per rappresentare informazioni eterogenee
 - gli impiegati con i relativi stipendi
 - i progetti con i relativi bilanci
 - le partecipazioni degli impiegati ai progetti con le relative funzioni

**Per studiare in maniera sistematica questi
aspetti, è necessario introdurre un vincolo
di integrità:
la dipendenza funzionale**

Proprietà



- Ogni impiegato ha un solo stipendio (anche se partecipa a più progetti)
- Ogni progetto ha un bilancio
- Ogni impiegato in ciascun progetto ha una sola funzione (anche se può avere funzioni diverse in progetti diversi)

Dipendenza funzionale

- relazione r su $R(X)$
- due sottoinsiemi non vuoti Y e Z di X
- esiste in r una dipendenza funzionale (FD) da Y a Z se, per ogni coppia di ennuple t_1 e t_2 di r con gli stessi valori su Y , risulta che t_1 e t_2 hanno gli stessi valori anche su Z

Notazione

$$Y \rightarrow Z$$

- Esempi:

Impiegato \rightarrow Stipendio

Progetto \rightarrow Bilancio

Impiegato Progetto \rightarrow Funzione

Possiamo assumere per semplicità $|Z| = 1$, in quanto se Z fosse una composizione di attributi, ad es.

$A_1A_2A_3$, la relazione soddisfa $Y \rightarrow Z$ se e solo se soddisfa $Y \rightarrow A_1$, $Y \rightarrow A_2$, $Y \rightarrow A_3$.

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

Impiegato → Stipendio
Progetto → Bilancio
Impiegato Progetto → Funzione

Altre FD, particolari

- Impiegato Progetto \rightarrow Progetto
- Si tratta però di una FD “**banale**” (sempre soddisfatta)
- $Y \rightarrow A$ è **non banale** se A non appartiene a Y
- $Y \rightarrow Z$ (Z insieme di attributi) è **non banale** se nessun attributo in Z appartiene a Y , ovvero se l'intersezione tra Y e Z è l'insieme vuoto

Le anomalie sono legate ad alcune FD

- gli impiegati hanno un unico stipendio

Impiegato → Stipendio

- i progetti hanno un unico bilancio

Progetto → Bilancio

Ma non tutte le FD causano anomalie



- In ciascun progetto, un impiegato svolge una sola funzione

Impiegato Progetto → Funzione

- Il soddisfacimento è più “semplice”, perché **Impiegato Progetto** è chiave

FD e anomalie



- La parte sinistra della terza FD corrisponde ad una chiave e non causa anomalie, mentre
- le parti sinistre delle prime due FD non corrispondono a chiavi e causano anomalie

- Le anomalie sono causate dalla presenza di concetti eterogenei:
 - proprietà degli impiegati (lo stipendio)
 - proprietà di progetti (il bilancio)
 - proprietà della chiave **Impiegato Progetto**

Forma normale di Boyce e Codd (BCNF)



- Una relazione r è in forma normale di Boyce e Codd (BCNF) se, per ogni dipendenza funzionale (non banale) $X \rightarrow Y$ definita su di essa, X contiene una chiave K di r , ovvero X è superchiave di r .
- La forma normale richiede che i concetti in una relazione siano omogenei (solo proprietà direttamente associate alla chiave)

Che facciamo se una relazione non soddisfa la BCNF?



- La **normalizziamo**, ovvero la sostituiamo con altre relazioni che soddisfano la BCNF

Come?

- Idea di base: Decomponendo sulla base delle dipendenze funzionali, al fine di separare i concetti

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

Procedura intuitiva di normalizzazione

- Idea generale:
 - Per ogni dipendenza $X \rightarrow Y$, definire una relazione sugli attributi XY
 - Assicurandosi che tutti gli attributi siano coinvolti in qualche FD, quindi presenti nelle relazioni che introduciamo

Questa idea funziona “bene” per la tabella appena vista (vedremo dopo), ma non sempre...

Non sempre così facile

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato → Sede
Progetto → Sede

Chiave primaria: Impiegato, Progetto; non è BCNF

Decomponiamo sulla base delle dipendenze

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

Proviamo a ricostruire



Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Verdi	Saturno	Milano
Neri	Giove	Milano

Diversa dalla relazione di partenza!

Decomposizione senza perdita

- Una relazione r si **decompone senza perdita** su X_1 e X_2 se il **join** delle proiezioni di r su X_1 e X_2 è uguale a r stessa (cioè non contiene ennuple spurie)
- La decomposizione senza perdita è garantita (condizione **sufficiente**, non **necessaria**) se gli attributi comuni contengono una **chiave** (cioè sono **superchiave**) per almeno una delle relazioni decomposte
- L'attributo comune qui è **Sede** che non contiene una chiave

Proviamo a decomporre senza perdita

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

Impiegato → Sede
Progetto → Sede

Un altro problema

- Supponiamo di voler inserire una nuova ennupla che specifica la partecipazione dell'impiegato Neri, che opera a Milano, al progetto Marte

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

Impiegato → Sede
Progetto → Sede

L'aggiunta è applicabile alle relazione decomposte...

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere
Neri	Marte

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Neri	Marte	Milano

... ma viola una FD in quella originale

Conservazione delle dipendenze

- Una decomposizione **conserva le dipendenze** se ciascuna delle dipendenze funzionali dello schema originario coinvolge attributi che compaiono tutti insieme in uno degli schemi decomposti
- **Progetto** → **Sede** non è conservata

Qualità delle decomposizioni

- Una decomposizione dovrebbe sempre soddisfare:
 - la **decomposizione senza perdita**, che garantisce la ricostruzione delle informazioni originarie
 - la **conservazione delle dipendenze**, che garantisce il mantenimento dei vincoli di integrità originari

La tabella a slide 18 se decomposta in tre tabelle è (i) in BCNF, (ii) senza perdita, e (iii) conserva le dipendenze.

Stessa cosa per la tabella a slide 22 se decomposta su
Impiegato → Sede, Progetto → Sede.

Una relazione non normalizzabile in BCNF

<u>Dirigente</u>	<u>Progetto</u>	<u>Sede</u>
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Marte	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Progetto Sede → Dirigente
Dirigente → Sede

La decomposizione è problematica

- **Progetto Sede → Dirigente** coinvolge tutti gli attributi e quindi nessuna decomposizione può preservare tale dipendenza, ovvero se creassimo una tabella con tutti gli attributi di questa FD, otterremmo ovviamente la stessa tabella, che non è in BCNF
- quindi in alcuni casi la BCNF “non è raggiungibile”, come in questo caso

Una nuova forma normale

- Una relazione r è in **terza forma normale** se, per ogni FD (non banale) $X \rightarrow Z$ definita su r , è verificata almeno una delle seguenti condizioni:
 - X contiene una chiave K di r
 - ogni attributo in Z è contenuto in almeno una chiave di r (ricordiamo che possiamo assumere $|Z| = 1$)

BCNF e terza forma normale

- la terza forma normale è meno restrittiva della forma normale di Boyce e Codd (e ammette relazioni con alcune anomalie)
- ha il vantaggio però di essere sempre “raggiungibile”
- se una relazione ha una sola chiave, allora essa è in BCNF se e solo se è in 3NF

Decomposizione in terza forma normale

- Idea di partenza sempre la solita: si crea una relazione per ogni gruppo di attributi coinvolti in una dipendenza funzionale
- si verifica che alla fine una relazione contenga una chiave della relazione originaria
- Dipende sempre dalle dipendenze individuate

Una possibile strategia



- se la relazione non è normalizzata si decompone in terza forma normale
- alla fine si verifica se lo schema ottenuto è anche in BCNF

Uno schema non decomponibile in BCNF, ma in terza forma normale

Dirigente	<u>Progetto</u>	<u>Sede</u>
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Marte	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Dirigente → Sede
Progetto Sede → Dirigente

Come visto prima, tale tabella non è in BCNF (è non può essere normalizzata tale che lo sia), ma è in terza forma normale perchè la parte destra della prima FD (**Sede**) è parte della chiave primaria. Ammette l'anomalia di avere la Sede specificata ogni volta che compare il Dirigente, perchè 3NF è meno restrittiva.

Altro esempio

<u>Impiegato</u>	<u>Progetto</u>	Stipendio
Rossi	Marte	30
Verdi	Giove	30
Verdi	Venere	30
Neri	Saturno	40
Neri	Venere	40

Con FD: Impiegato → Stipendio

Tale tabella non è né in BCNF né in terza forma normale. Seguendo le stesse regole usate per BCNF, ovvero in questo caso normalizzare secondo la FD e la chiave, porta ad avere tabelle in forma normale, la cui decomposizione è senza perdita e che conservano le dipendenze (si veda slide seguente)

La tabella precedente normalizzata

- Seguendo le stesse regole usate per BCNF, ovvero in questo caso normalizzare secondo la FD e la chiave, porta ad avere tabelle in forma normale, la cui decomposizione è senza perdita e che conservano le dipendenze

Impiegato	Stipendio
Rossi	30
Verdi	30
Neri	40

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

Impiegato → Stipendio

Normalizzazione e scelta attributi



A volte il fatto che una relazione sia o meno in forma normale, dipende dal risultato della progettazione logica e delle specifiche.

Ad esempio si consideri la tabella “problematica” a slide 30: le stesse informazioni sarebbe potute essere presentate, ma in maniera più “precisa”, introducendo un ulteriore attributo Reparto che partiziona, sulla base dei Dirigenti, le singole sedi, con corrispondenti DF come alla slide successiva.

Una possibile riorganizzazione

Dirigente	<u>Progetto</u>	<u>Sede</u>	Reparto
Rossi	Marte	Roma	1
Verdi	Giove	Milano	1
Verdi	Marte	Milano	1
Neri	Saturno	Milano	2
Neri	Venere	Milano	2

Dirigente → Sede Reparto
 Sede Reparto → Dirigente
 Progetto Sede → Reparto

Decomposizione in BCNF

Normalizziamo secondo le DF. E visto che la prima e la seconda DF includono gli stessi attributi, otteniamo due tabelle.

<u>Dirigente</u>	Sede	Reparto
Rossi	Roma	1
Verdi	Milano	1
Neri	Milano	2

<u>Progetto</u>	<u>Sede</u>	Reparto
Marte	Roma	1
Giove	Milano	1
Marte	Milano	1
Saturno	Milano	2
Venere	Milano	2

Entrambe le tabelle sono in BCNF, quindi anche in 3NF, sono senza perdita e conservano le dipendenze.

Altre forme normali

<u>Impiegato</u>	Categoria	Stipendio
Neri	3	30
Verdi	3	30
Rossi	4	50
Mori	4	50
Bianchi	5	72

Impiegato → Categoria

Categoria → Stipendio

Questa tabella non è né in BCNF né in 3NF, data la seconda DF.

E' invece in 2NF, una forma normale più permissiva che definiamo nella slide successiva.

Seconda forma normale



Una dipendenza funzionale è in seconda forma normale (2NF) se su di essa non sono definite dipendenze *parziali*, cioè dipendenze tra un sottoinsieme proprio della chiave primaria ed altri attributi.

Quindi la tabella alla slide precedente è in 2NF.