

---

# Normalizzazione di Basi di Dati

---

Prof. Francesco Accarino

IIS Altiero Spinelli

via Leopardi 132 Sesto San Giovanni

---

# 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
-

---

# Normalizzazione

- Procedura che permette di trasformare schemi non normalizzati in schemi che soddisfano una forma normale (i.e., normalizzati)
  - La normalizzazione va utilizzata come **tecnica di verifica** dei risultati della progettazione di una base di dati
  - Non è una metodologia di progettazione
-

# Una relazione con anomalie

<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

- Lo stipendio di ciascun impiegato è ripetuto in tutti i records
    - **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**
-

---

# Cosa c'è che non va?

- 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
-

---

# Normalizzazione delle Relazioni

- Il processo di normalizzazione (proposto da Codd) sottopone uno schema di relazione a una serie di test per *certificare se soddisfa una data **forma normale***.
  - Esistono:
    - Prima forma normale (1NF)
    - Seconda forma normale (2NF)
    - Terza forma normale (3NF)
    - Forma normale di Boyce e Codd (BCNF)
    - (4NF e 5NF)
-

---

# Normalizzazione delle relazioni

- **La normalizzazione dei dati può essere considerata** come un processo di analisi degli schemi forniti, basato sulle loro dipendenze funzionali e chiavi primarie, per raggiungere le proprietà desiderate di
    1. **Minimizzazione della ridondanza**
    2. **Minimizzazione delle anomalie di inserimento/cancellazione/modifica**
  - Schemi di relazione inadeguati, che non soddisfano certe condizioni vengono decomposti in schemi di relazione più piccoli che superano i test e pertanto possiedono le proprietà desiderate.
-



---

# Dipendenza funzionale

- La **chiave Primaria** è l'insieme di uno o più attributi che identificano in modo univoco un record della tabella
  - La **chiave candidata** è ogni insieme minimale di uno o più attributi che possono svolgere la funzione di chiave primaria (ci possono essere più chiavi candidate ma una sola chiave primaria)
  - L'**attributo no chiave** è un campo che non fa parte della chiave primaria
-

---

# Dipendenza funzionale

ESEMPIO:

Inventario( NumeroInventario, Prodotto, Magazzino, Quantità IndirizzoMagazzino)

- **NumeroInventario** è una chiave candidata
- **Prodotto** non è una chiave candidata
- **Magazzino** non è una chiave candidata
- **(Prodotto Magazzino)** è una chiave candidata
- **(Prodotto, Magazzino, Quantità)** non è una chiave candidata non è una chiave candidata perché contiene al suo interno un sottoinsieme di attributi che sono già chiave candidata

IL progettista sceglierà la chiave primaria in base alle particolari esigenze che intende soddisfare

---

# Dipendenza funzionale

- Si ha **dipendenza funzionale** tra attributi quando il valore di un insieme di attributi **A** determina un singolo valore dell'attributo **B** e si indica con  $A \rightarrow B$ . Si dice anche che **B** dipende da **A** o che **A** è un **determinante** per **B**
- Se un attributo è chiave candidata di una relazione allora è un determinante di ogni attributo della relazione e viceversa, un attributo che determina tutti gli altri attributi è chiave candidata.
- **NumeroInventario** è determinante di ogni attributo
- **(Prodotto Magazzino)** è determinante di ogni attributo
- **Magazzino** è determinante di indirizzo Magazzino

Si ha dipendenza transitiva quando  $A \rightarrow B$  e  $B \rightarrow C$ . Si dice allora che  $C$  dipende transitivamente da  $A$

---

# Prima forma normale (1NF)

- Richiede che il dominio di un attributo comprenda solo valori atomici (semplici, indivisibili) e che il valore di qualsiasi attributo in una tupla sia un valore singolo del dominio.
    - ❑ 1NF è già parte integrante della definizione formale di relazione nel 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 appartengono allo stesso dominio
    - ❑ ogni riga è diversa da tutte le altre
    - ❑ l'ordine con il quale le righe compaiono nella tabella è irrilevante
-

# Prima forma normale (1NF)

## IMPIEGATI

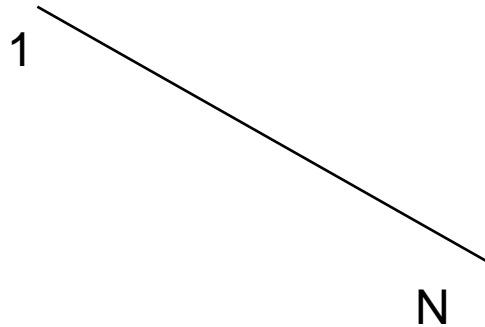
<b>Cod. impiegato</b>	<b>Cognome</b>	<b>Nome</b>	<b>Data nascita</b>	<b>Figli a carico</b>
001242	Rossi	Franco	09/11/1961	Marco Giuseppe Sandra
001318	Verdi	Alberto	24/04/1963	Sonia Claudio

Una tabella di questo genere **NON** è in **Prima forma normale** in quanto l'attributo figli a carico contiene più valori.

# Prima forma normale (1NF)

Con un processo di normalizzazione la tabella deve essere trasformata in 2 tabelle

Cod. impiegato	Cognome	Nome	Data nascita
001242	Rossi	Franco	09/11/1961
001318	Verdi	Alberto	24/04/1963



Cod. impiegato	Cod. Figlio	Nome
001242	01	Marco
001242	02	Giuseppe
001242	03	Sandra
001318	01	Sonia
001318	02	Claudio

# Seconda forma normale 2FN

## Requisiti della Seconda forma normale


- E' in prima forma normale
- tutti i suoi attributi non-chiave dipendono dall'intera chiave, cioè non possiede attributi che dipendono soltanto da una parte della chiave

La seconda forma normale elimina la dipendenza **parziale** degli attributi dalla chiave. Esempio:

La Tabella T1(A1,A2,A3,A4,A5) con

$(A1, A2) \rightarrow A3$      $A1 \rightarrow A4$      $A2 \rightarrow A5$

non è in 2FN, e può essere normalizzata in 2FN con le tabelle:



T2(A1,A2,A3)  
T3(A1,A4)  
T4(A2,A5)

---

# Seconda forma normale 2FN

La seconda forma normale riguarda quindi le tabelle in cui la chiave primaria sia composta da più attributi e stabilisce che, in questo caso, tutte le colonne corrispondenti agli altri attributi dipendano dall'intera chiave primaria.

ESEMPIO: Inventario

**CodArticolo   CodMagazzino   DescArticolo   Quantità   IndirizzoMagazzino**

Chiave primaria: Codarticolo + CodMagazzino

---



# Seconda forma normale 2FN

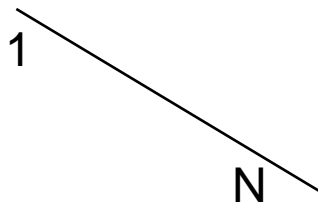
## Inventario

CodArticolo CodMagazzino DescArticolo Quantità IndirizzoMagazzino

**Soluzione: mettere in 2FN**

## Articolo

CodArticolo DescArticolo

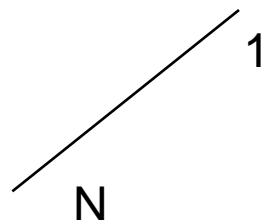


## Inventario

CodArticolo CodMagazzino Quantità

## Magazzino

CodMagazzino IndirizzoMagazzino




# Terza forma normale 3FN

## Requisiti della Terza forma normale

- E' in Seconda forma normale
- tutti gli attributi non-chiave dipendono direttamente dalla chiave, cioè non possiede attributi che dipendono da altri attributi che non sono in chiave

La terza forma normale elimina la dipendenza **transitiva** degli attributi dalla chiave. Esempio:

La Tabella T1(A1,A2,A3,A4) con  $A2 \rightarrow A4$   
non è in 3FN, e può essere normalizzata in   $T2(\underline{A1},A2,A3)$   
3FN con le tabelle:  $T3(\underline{A2},A4)$

# Terza forma normale 3FN

La terza forma normale stabilisce che non esistano dipendenze tra le colonne di una tabella se non basate sulla chiave primaria.

ESEMPIO: Impiegati

**CodImpiegato Nome Reparto TelefonoReparto**

*Chiave primaria* CodImpiegato

*Dipendenze funzionali* CodImpiegato → Reparto

Reparto → TelefonoReparto

# Esempio 3FN

## Impiegato

CodImpiegato Nome Reparto TelefonoReparto

## Problemi

- telefono del Reparto ripetuto per ogni Impiegato di quel Reparto (*ridondanza*)
- se il telefono cambia, occorre modificare molte righe
- con errori di aggiornamento, si avrebbero telefoni differenti
- se un Reparto non ha impiegati, non si può conoscere il suo telefono

**Soluzione: mettere in 3FN**

## Impiegato

CodImpiegato Nome Reparto

## Reparto

Reparto TelefonoReparto

---

# RIASSUMENDO

## Con la normalizzazione:

- la tabella iniziale viene scomposta in più tabelle
- complessivamente forniscono le stesse informazioni di partenza
- mantengono le dipendenze tra gli attributi
- in ciascuna di esse ogni attributo dipende direttamente dalla chiave
- vengono evitati problemi di ridondanza e di inconsistenza dei dati
- non ci deve essere perdita complessiva delle informazioni

## **Prima forma normale**

possiede i requisiti fondamentali del modello relazionale, in particolare ogni attributo è elementare, non ci sono righe uguali, non ci sono attributi di gruppo o ripetuti

## **Seconda forma normale**

è in prima forma normale e non ci sono attributi non-chiave che dipendono *parzialmente* dalla chiave

## **Terza forma normale**

è in seconda forma normale e non ci sono attributi non-chiave che dipendono *transitivamente* dalla chiave

---

# Integrità referenziale

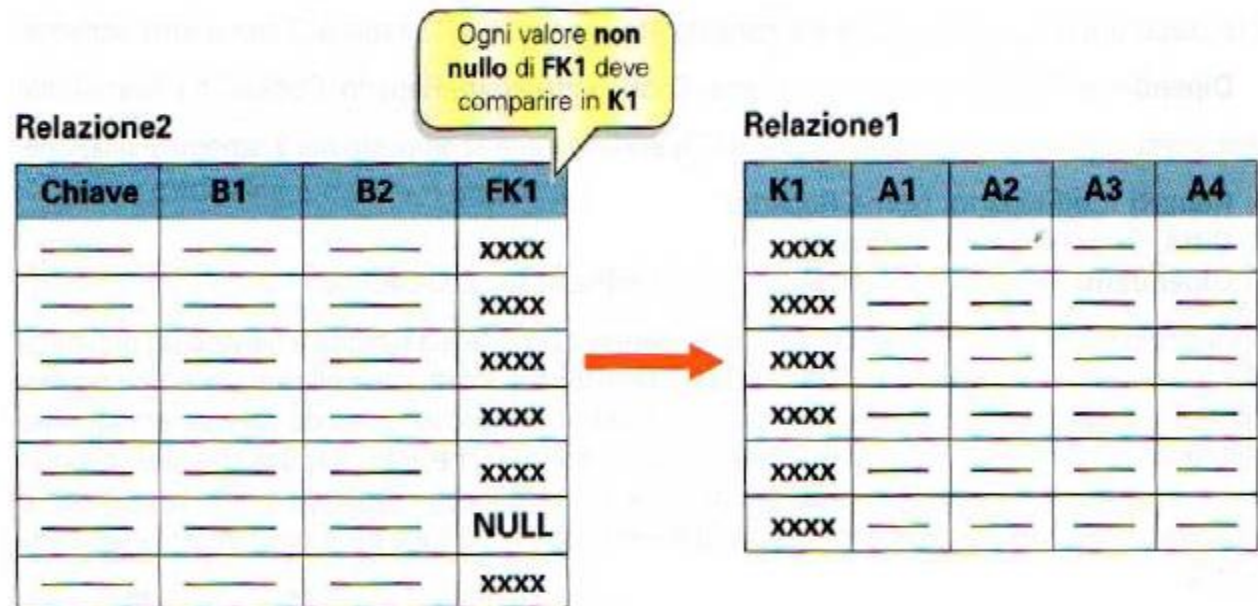
Per integrità referenziale si intende l'insieme di regole che garantiscono l'integrità dei dati di un modello relazionale rendendo valide le associazioni tra le tabelle ed eliminando gli eventuali errori di inserimento, cancellazione o modifica di dati collegati tra loro.

# Integrità referenziale

Per ogni valore non nullo della chiave esterna, deve esistere un valore corrispondente della chiave primaria nella tabella associata

ON DELETE  
ON UPDATE

no action  
set null  
cascade





# Effetti integrità referenziale

Non è possibile immettere un valore nella chiave esterna di una tabella associata, se tale valore non esiste tra le chiavi della tabella primaria

**Esempio:** *Se esiste una integrità referenziale (Foreign key) tra l'ordine ed il cliente: un ordine non può essere assegnato ad un cliente che non esiste nella tabella dei clienti.*

Non è possibile eliminare una riga dalla tabella primaria, se esistono righe legate ad essa attraverso la chiave esterna nella tabella correlata

**Esempio:** *Non è possibile eliminare un cliente dalla tabella dei clienti se ci sono ordini assegnati a quel cliente nella tabella degli ordini.*

Non si può modificare il valore assegnato ad una chiave nella tabella primaria, se essa compare in qualche riga della tabella correlata

**Esempio:** *Non è possibile modificare il valore alla chiave di un cliente se si sono ordini per quel cliente già registrati nella tabella degli ordini.*

# Forma Normale di Boyce e Codd (BCNF)

- Una relazione è in forma normale di Boyce-Codd (BCNF, Boyce-Codd Normal Form) quando rispetta le caratteristiche fondamentali del modello relazionale (1FN) e in essa ogni determinante è una chiave candidata, cioè ogni attributo dal quale dipendono altri attributi può svolgere la funzione di chiave.
- Una relazione è in forma normale di Boyce e Codd se e solo se, per ogni dipendenza funzionale  $Y \rightarrow Z$  definita su di essa,  $Y$  è una chiave candidata della relazione

---

# Forma Normale di Boyce e Codd (BCNF)

La BCNF può essere espressa anche nel seguente modo: se in una relazione vale la dipendenza funzionale  $A \rightarrow B$ , allora l'insieme di attributi  $A$  deve contenere una chiave (e quindi può svolgere la funzione di chiave).

Da questo fatto discende immediatamente che una relazione che soddisfa la BCNF è anche in seconda e in terza forma normale, in quanto la BCNF esclude che un determinante possa essere composto solo da una parte della chiave, come avviene per le violazioni alla 2FN, o che possa essere esterno alla chiave, come avviene per le violazioni alla 3FN.

Una relazione che rispetta la forma normale di Boyce-Codd è, quindi, anche in terza forma normale, ma non è vero l'opposto.

---

# Forma Normale di Boyce e Codd (BCNF)

Consideriamo una relazione che descrive l'allocazione delle sale operatorie di un ospedale. Le sale operatorie sono prenotate, giorno per giorno, in orari previsti, per effettuare interventi su pazienti dai chirurghi dell'ospedale.

Nel corso di una giornata una sala operatoria è occupata sempre dal medesimo chirurgo che effettua più interventi, in ore diverse.

Noti i valori di **Paziente e DataIntervento**, sono noti anche: **ora dell'intervento, chirurgo, e sala operatoria utilizzata**.

Gli attributi della relazione Interventi sono descritti nello schema:

**Interventi** (Paziente, DataIntervento, OraIntervento, Chirurgo, Sala)

In base alla precedente descrizione del caso in esame, nella relazione Interventi valgono le dipendenze funzionali:

- a. {Paziente, DataIntervento} → OraIntervento, Chirurgo, Sala
- b. {Chirurgo, DataIntervento, OraIntervento} → Paziente, Sala
- c. {Sala, DataIntervento, OraIntervento} → Paziente, Chirurgo
- d. {Chirurgo, DataIntervento} → Sala

# Forma Normale di Boyce e Codd (BCNF)

Ci sono tre insiemi di attributi che possono svolgere la funzione di chiave:

{Paziente, DataIntervento},  
{Chirurgo, DataIntervento, OraIntervento},  
{Sala, DataIntervento, OraIntervento}.

Scegliamo come chiave primaria la coppia di attributi: {Paziente, DataIntervento}.

## Interventi

Paziente	DataIntervento	OraIntervento	Chirurgo	Sala
Bianchi	25/10/2005	8.00	De Bakey	Sala1
Rossi	25/10/2005	8.00	Romano	Sala2
Negri	26/10/2005	9.30	Veronesi	Sala1
Viola	25/10/2005	10.30	De Bakey	Sala1
Verdi	25/10/2005	11.30	Romano	Sala2

---

# Forma Normale di Boyce e Codd (BCNF)

I determinanti nelle dipendenze funzionali **a, b, c**, sono insiemi di attributi che **possono** svolgere la funzione di chiave e quindi la BCNF non è sicuramente violata in questi casi.

La BCNF non è invece soddisfatta da **d. {Chirurgo, DataIntervento} → Sala** che ha come **determinante un** insieme di attributi non chiave.

Ne segue che la relazione Interventi non è in BCNF.

Interventi è invece in 3FN in quanto la dipendenza funzionale **d viene esclusa dalle dipendenze** funzionali considerate dalla terza forma normale, perché l'attributo Sala è un attributo che fa parte della chiave candidata {Sala, DataIntervento, OraIntervento} e quindi Sala non è un attributo non chiave.

---

# Forma Normale di Boyce e Codd (BCNF)

Osserviamo che la relazione Interventi, pur essendo in terza forma normale, presenta una certa ridondanza nei dati che può creare problemi in fase di aggiornamento.

Se per qualche ragione si deve cambiare la sala operatoria utilizzata da un chirurgo in un certa data, bisogna aggiornare più righe: per esempio, per spostare Romano dalla Sala2 alla Sala3, bisogna modificare due righe della tabella.

Paziente	DataIntervento	OralIntervento	Chirurgo	Sala
Bianchi	25/10/2005	8.00	De Bakey	Sala1
Rossi	25/10/2005	8.00	Romano	Sala2
Negri	26/10/2005	9.30	Veronesi	Sala1
Viola	25/10/2005	10.30	De Bakey	Sala1
Verdi	25/10/2005	11.30	Romano	Sala2

# Forma Normale di Boyce e Codd (BCNF)

Naturalmente, la tabella Interventi può essere normalizzata, ottenendo i due schemi:

**OccupazioneSale (Chirurgo, DataIntervento, Sala)**

**Interventi (Paziente, DataIntervento, OraIntervento, Chirurgo)**

L'attributo Sala viene tolto da Interventi e compare in una nuova tabella che ha come chiave il determinante della dipendenza funzionale **d che non rispettava la BCNF.**

**Interventi**

Paziente	DataIntervento	OraIntervento	Chirurgo
Bianchi	25/10/2005	8.00	De Bakey
Rossi	25/10/2005	8.00	Romano
Negri	26/10/2005	9.30	Veronesi
Viola	25/10/2005	10.30	De Bakey
Verdi	25/10/2005	11.30	Romano

**OccupazioneSale**

Chirurgo	DataIntervento	Sala
De Bakey	25/10/2005	Sala1
Romano	25/10/2005	Sala2
Veronesi	26/10/2005	Sala1



# Forma Normale di Boyce e Codd (BCNF)

<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 (Impiegato, Progetto) → Funzione    Progetto → Bilancio

# Normalizzazione in BCNF

- Se uno schema non è in BCNF, la soluzione è “decomporlo”, sulla base delle FD

Impiegato → Stipendio

<u>Impiegato</u>	Stipendio
Rossi	20
Verdi	35
Neri	55
Mori	48
Bianchi	48

Impiegato, Progetto → Funzione

<u>Impiegato</u>	<u>Progetto</u>	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

Progetto → Bilancio

<u>Progetto</u>	Bilancio
Marte	2
Giove	15
Venere	15

# Normalizzazione in BCNF

- La soluzione non è sempre così semplice, bisogna fare anche altre considerazioni; ad esempio, operando come prima:

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



<u>Impiegato</u>	<u>Sede</u>
Rossi	Roma
Verdi	Milano
Neri	Milano

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

Impiegato → Sede

Progetto → Sede



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

...se proviamo a tornare indietro (Join su Sede):

**Diversa dalla relazione di partenza!**

# Decomposizione Senza Perdita

- La decomposizione non deve assolutamente alterare il contenuto informativo del DB
- Si introduce pertanto il seguente requisito

**Decomposizione senza perdita (lossless)**

Uno schema  $R(X)$  si decompone senza perdita negli schemi  $R_1(X_1)$  e  $R_2(X_2)$  se, per ogni istanza legale  $r$  su  $R(X)$ , il join naturale delle proiezioni di  $r$  su  $X_1$  e  $X_2$  è uguale a  $r$  stessa:

$$\pi_{X_1}(r) \bowtie \pi_{X_2}(r) = r$$

Per decomporre senza perdita è necessario e sufficiente che il join naturale sia eseguito su una **superchiave** (chiave candidata minimale) di uno dei due sottoschemi,

# Esempio di decomposizione lossless

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



<u>Impiegato</u>	<u>Sede</u>
Rossi	Roma
Verdi	Milano
Neri	Milano

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

OK!

... ma i problemi non sono ancora finiti...

# Modifichiamo il DB...

Supponiamo di voler inserire l'informazione che Neri lavora al progetto Marte:

<u>Impiegato</u>	<u>Sede</u>
Rossi	Roma
Verdi	Milano
Neri	Milano

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

Ricostruendo la relazione otteniamo:

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

che però viola la FD **Progetto** → **Sede**!

---

# Normalizzazione Riepilogo

## **Prima forma normale**

Una relazione si dice in prima forma normale (1FN) quando rispetta i requisiti fondamentali del modello relazionale, in particolare ogni attributo è elementare, non ci sono righe uguali e non ci sono attributi ripetitivi.

## **Seconda forma normale**

Una relazione è in seconda forma normale (2FN) quando è in prima forma normale e non ci sono attributi non-chiave che dipendono parzialmente dalla chiave.

## **Terza forma normale**

Una relazione è in terza forma normale (3FN) quando è in seconda forma normale e non ci sono attributi non-chiave che dipendono transitivamente dalla chiave.

## **Forma normale di Boyce-Codd**

Una relazione è in forma normale di Boyce-Codd (BCNF) quando è in prima forma normale e in essa ogni determinante è una chiave candidata.

---

---

# Normalizzazione Riepilogo

Nella teoria dei database relazionali esistono forme normali di ordine superiore al terzo che risolvono situazioni di anomalia nelle operazioni sulle tabelle.

Più precisamente la quarta e la quinta forma normale risolvono i problemi che si possono creare quando nella relazione sono presenti attributi multivalore, cioè attributi che possono assumere più valori in corrispondenza dello stesso valore di un altro attributo.

Queste forme inoltre servono a rendere minimo il numero degli attributi che formano le chiavi composte.

Sebbene sia possibile definire tabelle anche in quarta (4FN) e quinta forma normale (5FN), di solito è sufficiente rappresentare le relazioni nel livello di normalizzazione 3FN che, come si può dimostrare, ha il pregio di essere sempre ottenibile senza perdita di informazioni e senza perdita di dipendenze funzionali.

Non è così invece per la forma normale di Boyce-Codd: ci sono relazioni che non possono essere normalizzate nella forma di Boyce-Codd senza perdita di dipendenze funzionali

---