

# La normalizzazione

# Forme normali

- Una forma normale è una proprietà di una base di dati relazionale che ne testimonia alcune “qualità” (o, più semplicemente, l'assenza di determinati difetti)

## Una relazione con anomalie

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

# Approfondiamo

- un vincolo di integrità
  - dipendenza funzionale

Definizione alla lavagna

$r$  su  $R(X)$

$Y, Z \subseteq X$

$r$  su  $R(I S P B A)$

$\{I\}, \{S\} \subseteq \{I, S, P, B, A\}$

$I, S \subseteq I S P B A$

## Dipendenza funzionale (vincolo)

$Y \rightarrow Z$

è soddisfatta

se, per ogni  $t_1$  e  $t_2$  in  $r$  con  $t_1[Y] = t_2[Y]$ ,  
risulta anche  $t_1[Z] = t_2[Z]$

**Impiegato**  $\rightarrow$  **Stipendio**,

per ogni  $t_1$  e  $t_2$  in  $r$  con  $t_1[I] = t_2[I]$ ,  
risulta anche  $t_1[S] = t_2[S]$

$r$  su  $R(X)$

$Y, Z \subseteq X$

$r$  su  $R(I S P B A)$

$\{I\}, \{S\} \subseteq \{I, S, P, B, A\}$

$I, S \subseteq I S P B A$

## Dipendenza funzionale (vincolo)

$Y \rightarrow Z$

è soddisfatta

se, per ogni  $t_1$  e  $t_2$  in  $r$  con  $t_1[Y] = t_2[Y]$ ,

risulta anche  $t_1[Z] = t_2[Z]$

**Impiegato**  $\rightarrow$  **Stipendio**,

se due ennuple sono uguali su **Impiegato**

allora sono uguali anche su **Stipendio**

<u>Impiegato</u>	Stipendio	<u>Progetto</u>	Bilancio	Attività
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 → Attività**

# Quali dipendenze causano problemi?

**Impiegato → Stipendio**

**Progetto → Bilancio**

**Sì!**

**Impiegato Progetto → Attività**

**No**



# Le anomalie sono legate ad alcune FD

- gli impiegati hanno un unico stipendio

Impiegato → Stipendio

- i progetti hanno un unico bilancio

Progetto → Bilancio

# Non tutte le FD causano anomalie

- In ciascun progetto, un impiegato ha una sola Attività

Impiegato Progetto → Attività

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

## Nota

- Per ogni chiave si ha una dipendenza funzionale fra la chiave stessa e tutti gli altri attributi (in alcuni casi per "transitività"):

Impiegato Progetto → Stipendio Bilancio Attività

# FD e anomalie

- La terza FD corrisponde ad una chiave e non causa anomalie
- Le prime due FD non corrispondono a chiavi e causano anomalie
- La relazione contiene alcune informazioni legate alla chiave e altre ad attributi che non formano una chiave
- 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), intuitivamente

- Una relazione  $r$  è in forma normale di Boyce e Codd se, per ogni dipendenza funzionale  $X \rightarrow Y$  definita su  $r$ , l'insieme  $X$  contiene una chiave  $K$  di  $r$
- Serve una precisazione

# FD banali

- Impiegato Progetto  $\rightarrow$  Progetto
  - è sempre soddisfatta (la chiamiamo banale)
- Consideriamo solo FD non banali
  - $Y \rightarrow Z$ 
    - con Y e Z insiemi disgiunti
- Quelle banali ci sono sempre, ma non causano problemi

# Forma normale di Boyce e Codd (BCNF)

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

<u>Impiegato</u>	Stipendio	<u>Progetto</u>	Bilancio	Attività
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** → Attività



# Normalizzazione

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

<u>Impiegato</u>	Stipendio	<u>Progetto</u>	Bilancio	Attività
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 → Attività**

Impiegato	Stipendio
Rossi	20
Verdi	35
Neri	55
Mori	48
Bianchi	48

Impiegato → Stipendio

Impiegato	Progetto	Attività
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

Impiegato → Stipendio

Impiegato Progetto → Attività

Progetto → Bilancio

Impiegato Progetto → Attività

Progetto	Bilancio
Marte	2
Giove	15
Venere	15

Progetto → Bilancio

# Normalizzazione "semplice e intuitiva"

- Una relazione per ogni FD
  - per ogni FD (o meglio, per ogni gruppo con lo stesso primo membro)
  - definiamo una relazione

Impiegato	Stipendio
Rossi	20
Verdi	35
Neri	55
Mori	48
Bianchi	48

Impiegato → Stipendio

Impiegato	Progetto	Attività
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

Impiegato Progetto → Attività

Progetto	Bilancio
Marte	2
Giove	15
Venere	15

Progetto → Bilancio

# Normalizzazione "semplice e intuitiva"

- Una relazione per ogni FD
  - per ogni FD (o meglio, per ogni gruppo con lo stesso primo membro)
  - definiamo una relazione
- Per definire una procedura efficace conviene però partire dalla relazione data e "decomporre"

# Procedura intuitiva di normalizzazione

- Non valida in generale, ma solo nei "casi semplici"
  - Per ogni dipendenza  $X \rightarrow Y$  che viola la BCNF, definire una relazione su  $XY$  ed eliminare  $Y$  dalla relazione originaria

# Procedura intuitiva di normalizzazione, esempio

R(Impiegato Stipendio Progetto Bilancio Attività)

$I \rightarrow S, P \rightarrow B, IP \rightarrow A$

R1(Impiegato Stipendio)

$I \rightarrow S$

OK

R2(Impiegato Progetto Bilancio Attività)

$P \rightarrow B, IP \rightarrow A$

R3(Progetto Bilancio)

$P \rightarrow B$

OK

R4(Impiegato Progetto Attività)

$IP \rightarrow A$

OK



# Procedura intuitiva di normalizzazione, attenzione

- Consideriamo un caso più complesso  
R (CodCorso, Anno, Titolo, Crediti, Docente, NomeDoc, Semestre, Dipartimento, Indirizzo)
  - CodCorso → Titolo, Crediti, Dipartimento
  - CodCorso, Anno → Docente, Semestre
  - Docente → NomeDoc
  - Dipartimento → Indirizzo
- Vediamo poi i dettagli, ma se decomponiamo con  
CodCorso → Titolo, Crediti, Dipartimento  
allora si "perde" il legame fra Dipartimento e Indirizzo
- Perciò la procedura va modificata

# Procedura intuitiva di normalizzazione, un po' più generale

- Valida in qualche caso in più
  - Per ogni dipendenza  $X \rightarrow Y$  che viola la BCNF, definire una relazione su  $XZ$  (dove  $Z$  è l'insieme degli attributi che "dipendono" da  $X$ , eventualmente per transitività) ed eliminare  $Z$  dalla relazione originaria
  - Il concetto di "transitività" andrebbe formalizzato, ma in molti casi reali è intuitivo

# Procedura un po' più generale, esempio

- R (CodCorso, Anno, Titolo, Crediti, Docente, NomeDoc, Semestre, Dipartimento, Indirizzo)
  - CodCorso → Titolo, Crediti, Dipartimento
  - CodCorso, Anno → Docente, Semestre
  - Docente → NomeDoc
  - Dipartimento → Indirizzo
- Chiave?
  - CodCorso, Anno

# Procedura un po' più generale, esempio

- R (CodCorso, Anno, Titolo, Crediti, Docente, NomeDoc, Semestre, Dipartimento, Indirizzo)

CodCorso → Titolo, Crediti, Dipartimento

CodCorso, Anno → Docente, Semestre

Docente → NomeDoc

Dipartimento → Indirizzo

- R (CC, A, T, Cr, Do, N, S, Di, I)

CC → T, Cr, Di

CC, A → Do, S

Do → N

Di → I

Corso

Edizione del corso

Docente

Dipartimento

# Procedura un po' più generale, esempio

R (CC, A, T, Cr, Do, N, S, Di, I)

~~CC~~ → T, Cr, Di

✓ CC, A → Do, S

~~Do~~ → N

~~Di~~ → I

Corso

Edizione del corso

Docente

Dipartimento

R1 (CC, T, Cr, Di, I)

✓ CC → T, Cr, Di

~~Di~~ → I

R2 (CC, A, Do, N, S)

✓ CC, A → Do, S

~~Do~~ → N

R3 (Di, I)

Di → I

OK

R4 (CC, T, Cr, Di)

CC → T, Cr, Di

OK

R5 (Do, N)

Do → N

OK

R2 (CC, A, Do, S)

CC, A → Do, S

OK

## Non sempre così facile

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

~~X~~ Impiegato → Sede  
~~⊗~~ Progetto → Sede

# Applichiamo la procedura

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

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

Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

**Impiegato → Sede**

**??**

- La procedura non va bene:
  - non "conserva" le dipendenze
- Proviamo l'altra tecnica (una relazione per ogni dipendenza)



# Una relazione per ogni dipendenza

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

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

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

## 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** su  $X_1$  e  $X_2$  **senza perdita di informazione** se il join delle proiezioni di  $r$  su  $X_1$  e  $X_2$  è uguale a  $r$  stessa (cioè non contiene ennuple spurie)
- Una buona decomposizione deve essere senza perdita

# Purtroppo

- L'esempio appena visto ci costringe a dire che la normalizzazione non sempre funziona:
  - vorremmo avere una decomposizione
    - in BCNF
    - senza perdita
    - che "conservi" le dipendenze
  - Non sempre è possibile:
    - ci sono realtà difficili da modellare
  - Anche per questo, usiamo la normalizzazione come strumento di verifica e non di progettazione

# Decomposizione senza perdita

- Approfondiamo

<u>IMP</u>	<u>PROGETTO</u>	<u>SEDE</u>
Rossi	MARTE	ROMA
Verdi	GIOVE	MILANO
Neri	GIOVE	MILANO

<u>PROG SEDE</u>			<u>IMP SEDE</u>	
MARTE	ROMA	—	Rossi	ROMA
GIOVE	MILANO	—	Verdi	MILANO
		—	Neri	MILANO



Condizione sufficiente

- gli attributi comuni sono chiave per almeno una delle proiezioni

# Attenzione

- La condizione sufficiente non è necessaria



	IMP	PROG	SEDE
	Rossi	GIOVE	ROMA
	Neri	MARTE	MILANO
	Neri	VENERE	MILANO
	Bruni	MARTE	MILANO
	Bruni	VENERE	MILANO

IMP	SEDE		PROG	SEDE
Rossi	ROMA		GIOVE	ROMA
Neri	MILANO		MARTE	MILANO
Bruni	MILANO		VENERE	MILANO

seconda condizione sufficiente

- le componenti delle tuple sono indipendenti fra loro

nell'esempio

- ogni sede ha un insieme di impiegati.
- ogni sede ha un insieme di progetti.

• un impiegato lavora ai progetti della propria sede

# Decomposizione senza perdita

- La decomposizione senza perdita è garantita
  - se gli attributi comuni contengono una **chiave** per almeno una delle relazioni decomposte
- Più in generale, si ha
  - quando le informazioni in  $X_1$  e  $X_2$  sono fra loro indipendenti

# Decomposizione senza perdita

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

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Bruni	Saturno

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Verdi	Roma
Bruni	Milano

