

Progettazione di basi di dati

Normalizzazione



#### Normalizzazione

- □ Introduzione
- □ Decomposizione in forma normale
- □ Decomposizione senza perdita
- Conservazione delle dipendenze





#### Introduzione



#### Normalizzazione

- □ La normalizzazione è un procedimento che, a partire da uno schema relazionale non normalizzato, permette di ottenere uno schema relazionale normalizzato
- □ La normalizzazione *non* è una metodologia di progettazione, bensì uno strumento di verifica



#### Normalizzazione e modello ER

- □ La metodologia di progettazione basata su schemi ER produce normalmente schemi relazionali normalizzati
- □ Le verifiche di normalizzazione possono essere applicate anche agli schemi ER



#### **Esame Superato**

<u>MatrStudente</u>	Residenza	<u>CodCorso</u>	NomeCorso	Voto
s94539	Milano	04FLYCY	Calcolatori elettronici	30
s94540	Torino	01FLTCY	Basi di dati	26
s94540	Torino	01KPNCY	Reti di calcolatori	28
s94541	Pescara	01KPNCY Reti di calcolatori 29		
s94542	Lecce	04FLYCY	Calcolatori elettronici	25



## **Esempio: vincoli**

- □ La chiave primaria è la coppia MatrStudente, CodCorso
- □ La residenza di ogni studente è unica ed è funzione solo dello studente, indipendentemente dagli esami che ha superato
- □ Il nome del corso è unico ed è funzione solo del corso, indipendentemente dagli studenti che superano il corrispondente esame



- ☐ In tutte le righe in cui compare uno studente è ripetuta la sua residenza
  - ridondanza



- □ In tutte le righe in cui compare uno studente è ripetuta la sua residenza
  - ridondanza
- Se la residenza di uno studente cambia, occorre modificare tutte le righe in cui compare contemporaneamente
  - anomalia di aggiornamento



- Se un nuovo studente si iscrive all'università, non può essere inserito nella base dati fino a quando non supera il primo esame
  - anomalia di inserimento



- Se un nuovo studente si iscrive all'università, non può essere inserito nella base dati fino a quando non supera il primo esame
  - anomalia di inserimento
- Se uno studente rinuncia agli studi, non è possibile tener traccia della sua residenza
  - anomalia di cancellazione



#### Ridondanza

- Un'unica relazione è utilizzata per rappresentare informazioni eterogenee
  - alcuni dati sono ripetuti in tuple diverse senza aggiungere nuova informazione
    - dati ridondanti



#### **Anomalie**

 □ Le informazioni ridondanti devono essere aggiornate in modo atomico (tutte contemporaneamente)



#### **Anomalie**

- □ Le informazioni ridondanti devono essere aggiornate in modo atomico (tutte contemporaneamente)
- La cancellazione di una tupla comporta la cancellazione di tutti i concetti in essa rappresentati
  - inclusi quelli che potrebbero essere ancora validi



#### **Anomalie**

- □ Le informazioni ridondanti devono essere aggiornate in modo atomico (tutte contemporaneamente)
- La cancellazione di una tupla comporta la cancellazione di tutti i concetti in essa rappresentati
  - inclusi quelli che potrebbero essere ancora validi
- ∠ L'inserimento di una nuova tupla è possibile solo se esiste almeno l'informazione completa relativa alla chiave primaria
  - non è possibile inserire la parte di tupla relativa ad un solo concetto



## Normalizzazione

Forma normale di Boyce Codd



- È un tipo particolare di vincolo d'integrità
- Descrive legami di tipo funzionale tra gli attributi di una relazione



- ∑ E' un tipo particolare di vincolo d'integrità
- Descrive legami di tipo funzionale tra gli attributi di una relazione
- □ Esempio: la residenza è unica per ogni studente
  - ogni volta che compare lo stesso studente, il valore è ripetuto
  - il valore di MatrStudente determina il valore di Residenza



- $\supset$  Una relazione r soddisfa la dipendenza funzionale  $X \to Y$  se, per ogni coppia  $t_1$ ,  $t_2$  di tuple di r, aventi gli stessi valori per gli attributi in X,  $t_1$  e  $t_2$  hanno gli stessi valori anche per gli attributi in Y
  - X determina Y (in r)



- $\supset$  Una relazione r soddisfa la dipendenza funzionale  $X \to Y$  se, per ogni coppia  $t_1$ ,  $t_2$  di tuple di r, aventi gli stessi valori per gli attributi in X,  $t_1$  e  $t_2$  hanno gli stessi valori anche per gli attributi in Y
  - X determina Y (in r)

MatrStudente → Residenza
MatrStudente CodCorso → NomeCorso



## **Dipendenza non banale**

□ La dipendenza

MatrStudente CodCorso → CodCorso

- è banale perché CodCorso fa parte di entrambi i lati
- $\supset$  Una dipendenza funzionale  $X \to Y$  è non banale se nessun attributo in X compare tra gli attributi in Y



## Dipendenze funzionali e chiavi

Data una chiave K di una relazione r
 K → qualsiasi altro attributo di r

#### 

MatrStudente CodCorso → Residenza

(o insieme di attributi)

- MatrStudente CodCorso → NomeCorso
- MatrStudente CodCorso → Voto



- □ Le anomalie sono causate da proprietà degli attributi coinvolti in dipendenze funzionali
  - Esempi
    - MatrStudente → Residenza
    - CodCorso → NomeCorso



- □ Le anomalie sono causate da proprietà degli attributi coinvolti in dipendenze funzionali
  - Esempi
    - MatrStudente → Residenza
    - CodCorso → NomeCorso
- □ Le dipendenze funzionali dalle chiavi non originano anomalie
  - Esempio
    - MatrStudente CodCorso → Voto



- □ Le anomalie sono causate
  - dall'inclusione di concetti indipendenti tra loro nella stessa relazione



#### □ Le anomalie sono causate

- dall'inclusione di concetti indipendenti tra loro nella stessa relazione
- da dipendenze funzionali X → Y che permettono la presenza di più tuple con lo stesso valore di X
  - X non contiene una chiave



## Forma normale di Boyce Codd (BCNF)

- □ BCNF = Boyce Codd Normal Form
- Una relazione r è in BCNF se, per ogni dipendenza funzionale (non banale) X → Y definita su di essa, X contiene una chiave di r (X è superchiave di r)
- Anomalie e ridondanze non sono presenti in relazioni in BCNF perché concetti indipendenti sono separati in relazioni diverse





## Normalizzazione

Decomposizione in forma normale



## **Decomposizione BCNF**

#### 

 processo di sostituzione di una relazione non normalizzata con due o più relazioni in BCNF



## **Decomposizione in BCNF**

#### □ Normalizzazione

 processo di sostituzione di una relazione non normalizzata con due o più relazioni in BCNF

#### □ Criterio

 una relazione che rappresenta più concetti indipendenti è decomposta in relazioni più piccole, una per ogni concetto, per mezzo delle dipendenze funzionali



### **Decomposizione in BCNF**

- □ Le nuove relazioni sono ottenute mediante proiezioni sugli insiemi di attributi corrispondenti alle dipendenze funzionali
- □ Le chiavi delle nuove relazioni sono le parti sinistre delle dipendenze funzionali
  - le nuove relazioni sono in BCNF



- □ Dipendenze funzionali nell'esempio
  - MatrStudente → Residenza
  - CodCorso → NomeCorso
  - MatrStudente CodCorso → Voto



 $\supset$  Da

R (<u>MatrStudente</u>, Residenza, <u>CodCorso</u>, NomeCorso, Voto)

□ Le relazioni in BCNF sono

 $R_1$  (MatrStudente, Residenza) =  $\pi_{MatrStudente, Residenza} R$ 

 $R_2$  (CodCorso, NomeCorso) =  $\pi_{CodCorso, NomeCorso} R$ 

R<sub>3</sub> (MatrStudente, CodCorso, Voto) =

 $\pi_{MatrStudente}$ , CodCorso, Voto R



#### $R_1$

<u>MatrStudente</u>	Residenza	
s94539	Milano	
s94540	Torino	
s94541	Pescara	
s94542	Lecce	

#### $R_2$

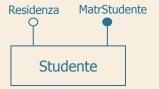
CodCorso	NomeCorso	
04FLYCY	Calcolatori elettronici	
01FLTCY	Basi di dati	
01KPNCY	Reti di calcolatori	

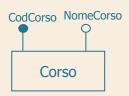
#### $R_3$

<u>MatrStudente</u>	<u>CodCorso</u>	Voto
s94539	04FLYCY	30
s94540	01FLTCY	26
s94540	01KPNCY	28
s94541	01KPNCY	29
s94542	04FLYCY	25



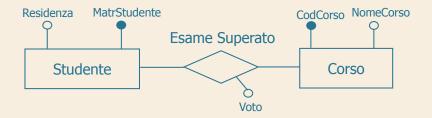
## **Esempio: schema ER corrispondente**





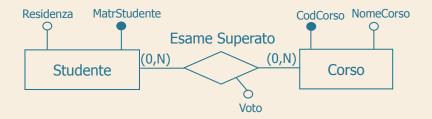


## **Esempio: schema ER corrispondente**



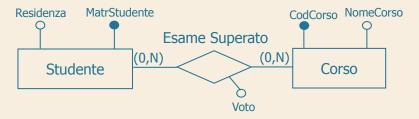


## **Esempio: schema ER corrispondente**





## Esempio: schema logico relazionale



Studente (<u>MatrStudente</u>, Residenza)
Corso (<u>CodCorso</u>, NomeCorso)
Esame Superato (<u>MatrStudente</u>, <u>CodCorso</u>, Voto)





Proprietà delle decomposizioni



#### Proprietà delle decomposizioni

- ∑ Sono accettabili tutte le decomposizioni?
  - proprietà essenziali per una "buona" decomposizione
- - perdita di informazione
  - perdita delle dipendenze



<u>Impiegato</u>	Categoria	Stipendio
Rossi	2	1800
Verdi	3	1800
Bianchi	4	2500
Neri	5	2500
Bruni	6	3500

R (Impiegato, Categoria, Stipendio)



<u>Impiegato</u>	Categoria	Stipendio
Rossi	2	1800
Verdi	3	1800
Bianchi	4	2500
Neri	5	2500
Bruni	6	3500

R (Impiegato, Categoria, Stipendio)

 $Impiegato \rightarrow Categoria$ 



<u>Impiegato</u>	Categoria	Stipendio
Rossi	2	1800
Verdi	3	1800
Bianchi	4	2500
Neri	5	2500
Bruni	6	3500

R (Impiegato, Categoria, Stipendio)

Impiegato → Categoria Impiegato → Stipendio



<u>Impiegato</u>	Categoria	Stipendio
Rossi	2	1800
Verdi	3	1800
Bianchi	4	2500
Neri	5	2500
Bruni	6	3500

R (Impiegato, Categoria, Stipendio)

Impiegato → Categoria Impiegato → Stipendio Categoria → Stipendio





# **Normalizzazione**

Decomposizione senza perdita



R (Impiegato, Categoria, Stipendio)

Decomposizione basata sulle dipendenze funzionali

Impiegato → Stipendio Categoria → Stipendio



R (Impiegato, Categoria, Stipendio)

□ Decomponendo

 $R_1$  (Impiegato, Stipendio) =

 $\pi_{\text{Impiegato, Stipendio}}$  R



R (Impiegato, Categoria, Stipendio)

#### □ Decomponendo

 $R_1$  (Impiegato, Stipendio) =

 $\pi_{\text{Impiegato, Stipendio}}$  R

<u>Impiegato</u>	Stipendio
Rossi	1800
Verdi	1800
Bianchi	2500
Neri	2500
Bruni	3500



R (Impiegato, Categoria, Stipendio)

#### □ Decomponendo

$$R_1$$
 (Impiegato, Stipendio) =  $\pi_{Impiegato, Stipendio}$  R

<u>Impiegato</u>	Stipendio
Rossi	1800
Verdi	1800
Bianchi	2500
Neri	2500
Bruni	3500

$$R_2$$
 (Categoria, Stipendio) =  $\pi_{Categoria, Stipendio} R$ 

<u>Categoria</u>	Stipendio
2	1800
3	1800
4	2500
5	2500
6	3500



□ Ricomponendo

 $R_1 \bowtie R_2$ 



$$R_1 \bowtie R_2$$

Impiegato	Categoria	Stipendio
Rossi	2	1800



$$R_1 \bowtie R_2$$

Impiegato	Categoria	Stipendio
Rossi	2	1800
Rossi	3	1800



$$R_1 \bowtie R_2$$

Impiegato	Categoria	Stipendio
Rossi	2	1800
Rossi	3	1800
Verdi	2	1800



$$R_1 \bowtie R_2$$

Impiegato	Categoria	Stipendio
Rossi	2	1800
Rossi	3	1800
Verdi	2	1800
Verdi	3	1800



$$R_1 \bowtie R_2$$

Impiegato	Categoria	Stipendio
Rossi	2	1800
Rossi	3	1800
Verdi	2	1800
Verdi	3	1800
Bianchi	4	2500



$$R_1 \bowtie R_2$$

Impiegato	Categoria	Stipendio	
Rossi	2	1800	
Rossi	3	1800	] <del>- </del> tuple "
Verdi	2	1800	spurie"
Verdi	3	1800	
Bianchi	4	2500	



□ Ricomponendo

$$R_1 \bowtie R_2$$

Impiegato	Categoria	Stipendio	
Rossi	2	1800	
Rossi	3	1800	tuple tuple spurie"
Verdi	2	1800	generation "spurie"
Verdi	3	1800	
Bianchi	4	2500	



Ricostruzione con perdita di informazione

#### **Decomposizione senza perdita**

- □ La decomposizione di una relazione r su due insiemi di attributi X<sub>1</sub> e X<sub>2</sub> è senza perdita di informazione se il join delle proiezioni di r su X<sub>1</sub> e X<sub>2</sub> è uguale a r stessa (senza tuple "spurie")
- Una decomposizione eseguita per normalizzare deve essere senza perdita



#### **Decomposizione senza perdita**

 $\supset$  Data la relazione r(X) e gli insiemi di attributi  $X_1$  e  $X_2$  tali che

$$X = X_1 \cup X_2$$
$$X_0 = X_1 \cap X_2$$

se r soddisfa la dipendenza funzionale

$$X_0 \to X_1 \, o \, X_0 \to X_2$$

la decomposizione di r su  $X_1$  e  $X_2$  è senza perdita

☐ Gli attributi comuni formano una chiave per almeno una delle relazioni decomposte



#### **Esempio: perdita di informazione**

R<sub>1</sub> (<u>Impiegato</u>, Stipendio)

R<sub>2</sub> (<u>Categoria</u>, Stipendio)

Verifica della condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Stipendio

 $X_2$  = Categoria, Stipendio



## **Esempio: perdita di informazione**

R<sub>1</sub> (<u>Impiegato</u>, Stipendio)

R<sub>2</sub> (<u>Categoria</u>, Stipendio)

Verifica della condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Stipendio

 $X_2$  = Categoria, Stipendio

 $X_0$  = Stipendio



#### Esempio: perdita di informazione

R<sub>1</sub> (<u>Impiegato</u>, Stipendio)

R<sub>2</sub> (<u>Categoria</u>, Stipendio)

Verifica della condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Stipendio

 $X_2$  = Categoria, Stipendio

 $X_0$  = Stipendio

L'attributo Stipendio non soddisfa la condizione per la decomposizione senza perdita



R (Impiegato, Categoria, Stipendio)

Decomposizione basata sulle dipendenze funzionali

Impiegato → Categoria Impiegato → Stipendio



R (Impiegato, Categoria, Stipendio)

□ Decomponendo

 $R_1$  (Impiegato, Categoria) =

π<sub>Impiegato</sub>, Categoria R



R (Impiegato, Categoria, Stipendio)

#### □ Decomponendo

 $R_1$  (Impiegato, Categoria) =

π<sub>Impiegato, Categoria</sub> R

<u>Impiegato</u>	Categoria
Rossi	2
Verdi	3
Bianchi	4
Neri	4
Bruni	5



R (Impiegato, Categoria, Stipendio)

#### □ Decomponendo

$$R_1$$
 (Impiegato, Categoria) =  $\pi_{Impiegato, Categoria} R$ 

<u>Impiegato</u>	Categoria	
Rossi	2	
Verdi	3	
Bianchi	4	
Neri	4	
Bruni	5	

$R_2$	(Impiegato, Stipendio)	=
	$\pi_{ m Impiegato,\ Stipendio}\ {\sf R}$	

<u>Impiegato</u>	Stipendio
Rossi	1800
Verdi	1800
Bianchi	2500
Neri	2500
Bruni	3500



# Esempio: decomposizione senza perdita?

R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (<u>Impiegato</u>, Stipendio)

 $R_1 \bowtie R_2$ 

□ La decomposizione è senza perdita?



## Esempio: decomposizione senza perdita

R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (<u>Impiegato</u>, Stipendio)

 $R_1 \bowtie R_2$ 

Verifica della condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Categoria

X<sub>2</sub> = Impiegato, Stipendio



## Esempio: decomposizione senza perdita

R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (<u>Impiegato</u>, Stipendio)

 $R_1 \bowtie R_2$ 

Verifica della condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Categoria

 $X_2$  = Impiegato, Stipendio

 $X_0 = Impiegato$ 



# Esempio: decomposizione senza perdita

R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (<u>Impiegato</u>, Stipendio)

 $R_1 \bowtie R_2$ 

Verifica della condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Categoria

X<sub>2</sub> = Impiegato, Stipendio

 $X_0 = Impiegato$ 

 □ L'attributo Impiegato soddisfa la condizione per la Bedecomposizione senza perdita

70



## **Normalizzazione**

Conservazione delle dipendenze



## Esempio: inserimento di una nuova tupla

R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (Impiegato, Stipendio)

□ Inserimento della tupla

Impiegato: Gialli – Categoria: 3 – Stipendio: 3500



R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (<u>Impiegato</u>, Stipendio)

#### □ Inserimento della tupla

<u>Impiegato</u>	Categoria
Rossi	2
Verdi	3
Bianchi	4
Neri	4
Bruni	5



R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (<u>Impiegato</u>, Stipendio)

#### □ Inserimento della tupla

<u>Impiegato</u>	Categoria
Rossi	2
Verdi	3
Bianchi	4
Neri	4
Bruni	5



1	1
Gialli	3

R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (Impiegato, Stipendio)

#### 

<u>Impiegato</u>	Categoria
Rossi	2
Verdi	3
Bianchi	4
Neri	4
Bruni	5

<u>Impiegato</u>	Stipendio
Rossi	1800
Verdi	1800
Bianchi	2500
Neri	2500
Bruni	3500



1	
Gialli	3

R<sub>1</sub> (<u>Impiegato</u>, Categoria)

R<sub>2</sub> (<u>Impiegato</u>, Stipendio)

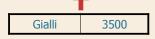
#### □ Inserimento della tupla

<u>Impiegato</u>	Categoria
Rossi	2
Verdi	3
Bianchi	4
Neri	4
Bruni	5

<u>Impiegato</u>	Stipendio
Rossi	1800
Verdi	1800
Bianchi	2500
Neri	2500
Bruni	3500



1	1
Gialli	3



- $\supset$  Cosa avviene se inserisco la tupla (Gialli,3500) in  $R_2$ ?
  - nella relazione originaria l'inserimento è vietato perché causa la violazione della dipendenza Categoria → Stipendio
  - nella decomposizione non è più possibile riconoscere alcuna violazione, poiché gli attributi Categoria e Stipendio sono in relazioni separate
- ∑ E' stata persa la dipendenza tra Categoria e Stipendio



### **Conservazione delle dipendenze**

- Una decomposizione conserva le dipendenze se ciascuna delle dipendenze funzionali dello schema originario è presente in una delle relazioni decomposte
- ∑ E' opportuno che le dipendenze siano conservate, in modo da garantire che nello schema decomposto siano soddisfatti gli stessi vincoli dello schema originario



# **Esempio: decomposizione (n.3)**

R (Impiegato, Categoria, Stipendio)

Decomposizione basata sulle dipendenze funzionali

Impiegato → Categoria Categoria → Stipendio



## **Esempio: decomposizione (n.3)**

R (Impiegato, Categoria, Stipendio)

#### □ Decomponendo

 $R_1$  (Impiegato, Categoria) =

π<sub>Impiegato, Categoria</sub> R

<u>Impiegato</u>	Categoria
Rossi	2
Verdi	3
Bianchi	4
Neri	5
Bruni	6



## **Esempio: decomposizione (n.3)**

R (Impiegato, Categoria, Stipendio)

#### □ Decomponendo

$$R_1$$
 (Impiegato, Categoria) =  $\pi_{Impiegato, Categoria}$   $R$ 

<u>Impiegato</u>	Categoria
Rossi	2
Verdi	3
Bianchi	4
Neri	5
Bruni	6

$R_2$	(Categoria,	Stipendio)	=
	$\pi_{Categoria}$ ,	Stipendio R	

<u>Categoria</u>	Stipendio
2	1800
3	1800
4	2500
5	2500
6	3500



# **Esempio**

□ Ricomponendo

 $R_1 \;\; \bowtie \;\; R_2$ 



## Esempio: decomposizione senza perdita

□ Ricomponendo

$$R_1 \bowtie R_2$$

Verifica condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Categoria

 $X_2$  = Categoria, Stipendio



## Esempio: decomposizione senza perdita

□ Ricomponendo

$$R_1 \bowtie R_2$$

Verifica condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Categoria

 $X_2$  = Categoria, Stipendio

 $X_0$  = Categoria



## Esempio: decomposizione senza perdita

□ Ricomponendo

$$R_1 \bowtie R_2$$

Verifica condizione per la decomposizione senza perdita

 $X_1$  = Impiegato, Categoria

 $X_2$  = Categoria, Stipendio

 $X_0$  = Categoria

 ∑ L'attributo Categoria soddisfa la condizione per la decomposizione senza perdita



### Esempio: conservazione delle dipendenze funzionali

□ Ricomponendo

$$R_1 \bowtie R_2$$



### Esempio: conservazione delle dipendenze funzionali

□ Ricomponendo

$$R_1 \bowtie R_2$$

□ Dipendenze funzionali conservate

Impiegato → Categoria

Categoria → Stipendio

□ La dipendenze funzionale

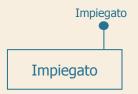
Impiegato → Stipendio

può essere ricostruita da

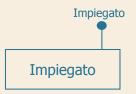
Impiegato → Categoria

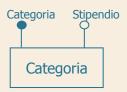
Categoria → Stipendio





















### Esempio: schema logico relazionale



Impiegato (<u>Impiegato</u>, Categoria) Categoria (<u>Categoria</u>, Stipendio)



### Qualità di una decomposizione

- Le decomposizioni devono sempre soddisfare le proprietà
  - decomposizione senza perdita
    - garantisce che le informazioni nella relazione originaria siano ricostruibili con precisione (senza tuple spurie) a partire da quelle nelle relazioni decomposte
  - conservazione delle dipendenze
    - garantisce che le relazioni decomposte abbiano la stessa capacità della relazione originaria di rappresentare i vincoli di integrità

