

NORMALIZZAZIONE II

Basi di Dati corso A

325

Uno schema non decomponibile 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

Dirigente → Sede
Progetto Sede → Dirigente

Basi di Dati corso A

326

Una possibile riorganizzazione

Dirigente	Progetto	Sede	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

Basi di Dati corso A

327

Decomposizione in BCNF

R_1

Dirigente	Sede	Reparto
Rossi	Roma	1
Verdi	Milano	1
Neri	Milano	2

R_2

Progetto	Sede	Reparto
Marte	Roma	1
Giove	Milano	1
Marte	Milano	1
Saturno	Milano	2
Venere	Milano	2

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

$X \rightarrow Y$

$\Rightarrow R_1, R_2$ sono in
 forma
 normale
 di BC

Basi di Dati corso A

328

Analisi inadeguata

- Spesso la non raggiungibilità della forma normale di Boyce e Codd può dipendere da un'analisi non sufficientemente accurata

Basi di Dati corso A

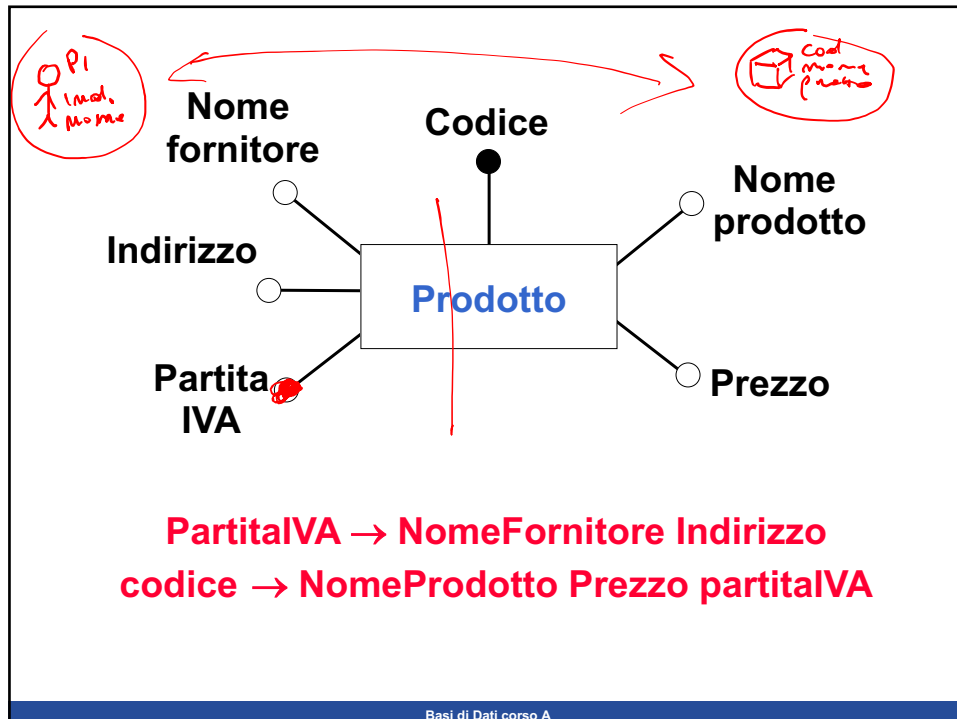
329

Progettazione e normalizzazione

- la teoria della normalizzazione può essere usata nella progettazione logica per verificare lo schema relazionale finale
- si può usare anche durante la progettazione concettuale per verificare la qualità dello schema concettuale

Basi di Dati corso A

330



331

Analisi dell'entità

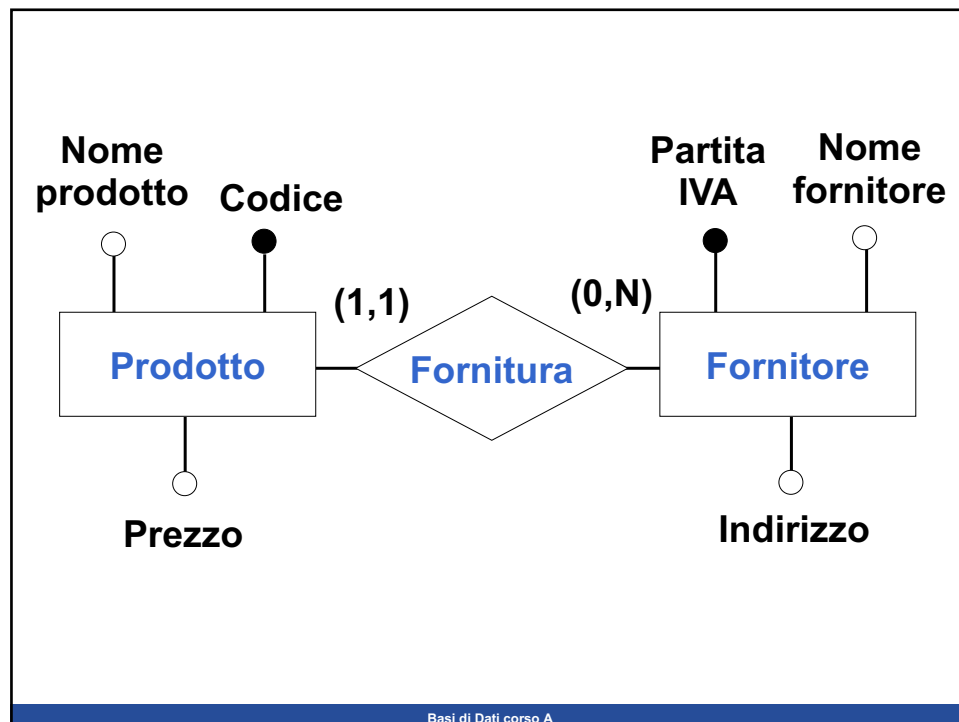
- L'entità viola la terza forma normale a causa della dipendenza:

PartitaIVA → NomeFornitore Indirizzo

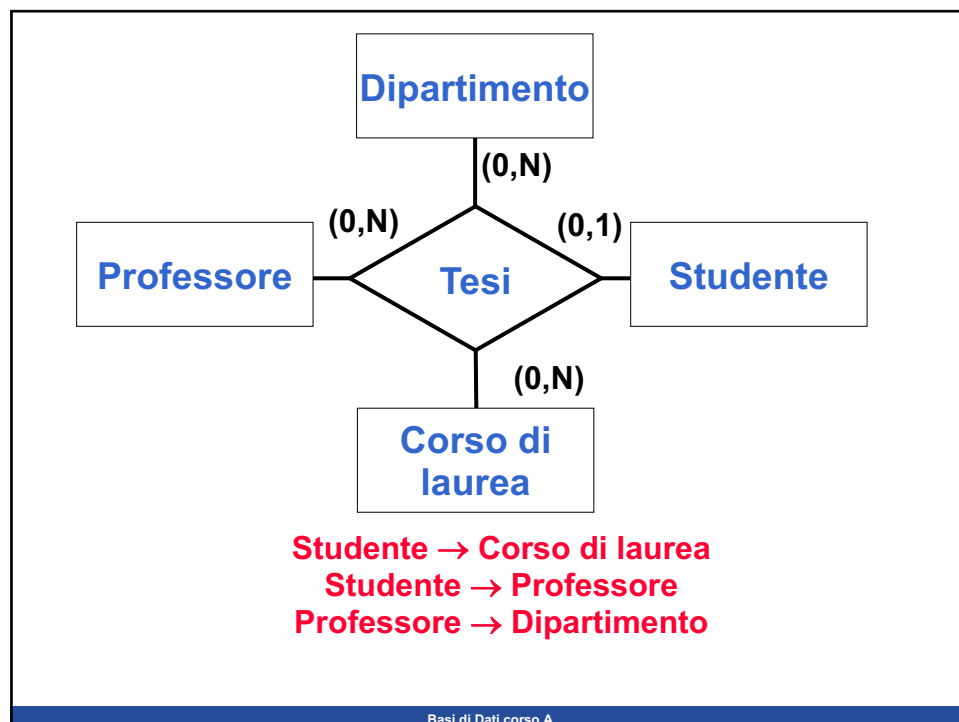
- Possiamo decomporre sulla base di questa dipendenza

Basi di Dati corso A

332



333



334

Analisi della relationship

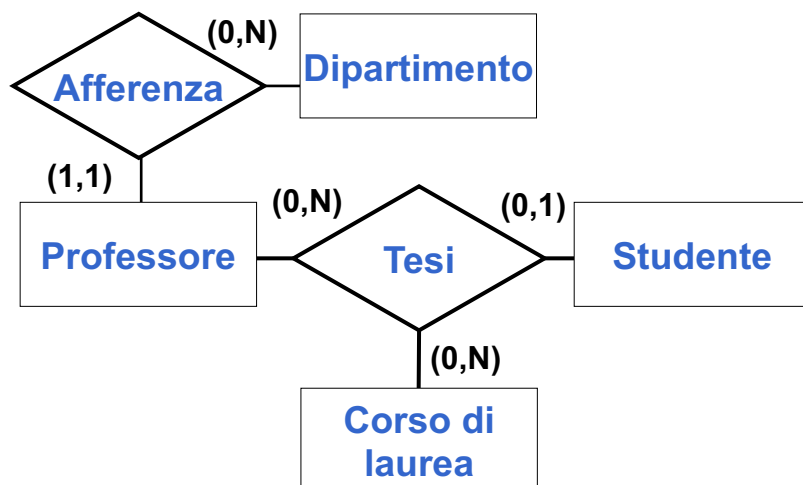
- La relationship viola la terza forma normale a causa della dipendenza:

Professore → Dipartimento

- Possiamo decomporre sulla base di questa dipendenza

Basi di Dati corso A

335



Basi di Dati corso A

336

Ulteriore analisi sulla base delle dipendenze

- La relationship **Tesi** è in BCNF sulla base delle dipendenze

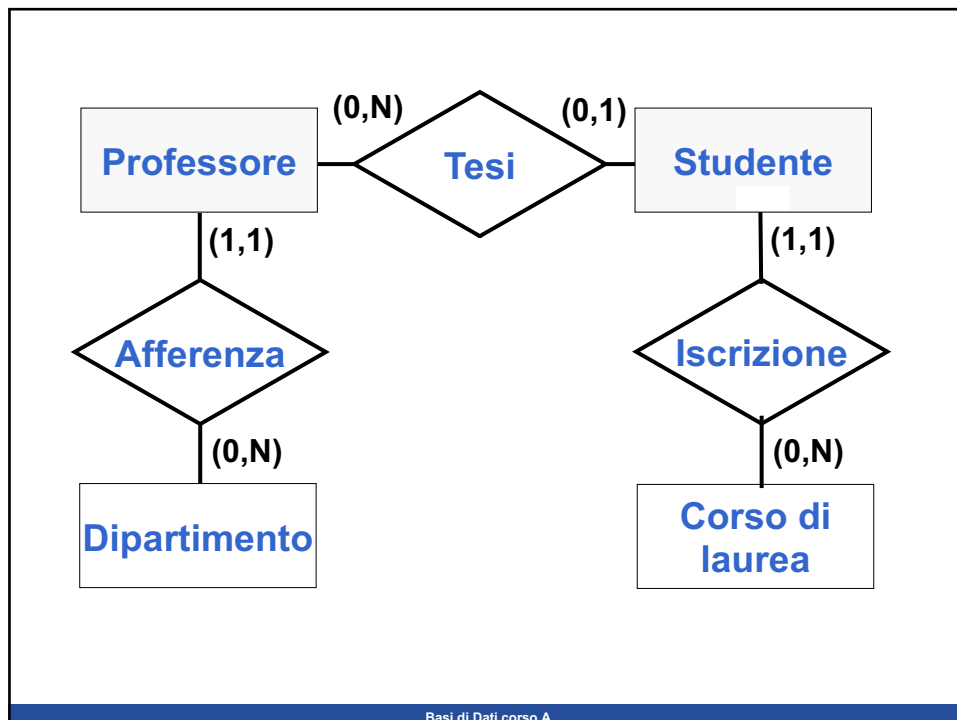
Studiante → **CorsoDiLaurea**

Studiante → **Professore**

- le due proprietà sono indipendenti
- questo suggerisce una ulteriore decomposizione

Basi di Dati corso A

337



Basi di Dati corso A

338

Quante sono le forme normali?

- 1 ► 1NF (First Normal Form) 1
- 2 ► 2NF (Second Normal Form) 2
- 3 ► 3NF (Third Normal Form) 3
- 4 ► BCNF (Boyce-Codd Normal Form) 4
- ancora...
- 4NF (Fourth Normal Form)
- 5NF (Fifth Normal Form)

1
2
3

al memo
46

Basi di Dati corso A

339

repetita juvant

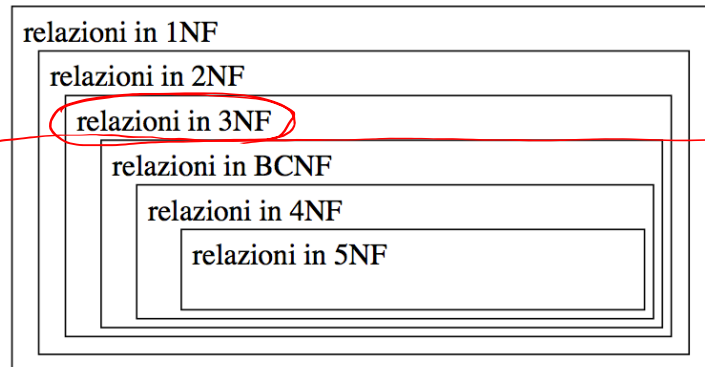
- Dipendenza funzionale:
 - è una dipendenza semantica
 - Data una relazione R, l'attributo Y di R dipende funzionalmente dall'attributo X di R se e solo se ogni valore di X in R ha associato un unico valore di Y in R e si scrive:
 - $X \rightarrow Y$
- In altri termini:
 - Y dipende funzionalmente da X
 - X è il determinante di Y

Basi di Dati corso A

340

normalizzazione

- Insieme di condizioni da rispettare per garantire la qualità di un database, vi sono sei livelli di qualità maggiormente noti:



Basi di Dati corso A

341

normalizzazione

- Processo ripetuto di applicazione di proiezioni per
- Eliminare ridondanze
- Eliminare anomalie

Basi di Dati corso A

342

1NF

- Una relazione è in 1NF se e solo se ogni tupla contiene un solo valore per ogni attributo

	Attr1	Attr2	Attr3
Tupla 1			
Tupla 2	x	x y	y
Tupla 3			
Tupla 4			

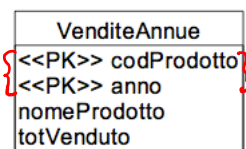
es x y y

Basi di Dati corso A

343

2NF

- Una relazione è in 2NF se è in 1NF e se ogni suo attributo non chiave dipende funzionalmente dalla chiave completa



non in 2NF!

$CP \rightarrow NP$
 $A, CP \rightarrow TV$
 colla x 2
 spago x 2

VenditeAnnue			
codProdotto	anno	nomeProdotto	totVenduto
p1	2001	colla	1000
p2	2001	spago	2500
p1	2002	colla	1700
p2	2002	spago	1850
p3	2002	elastico	320

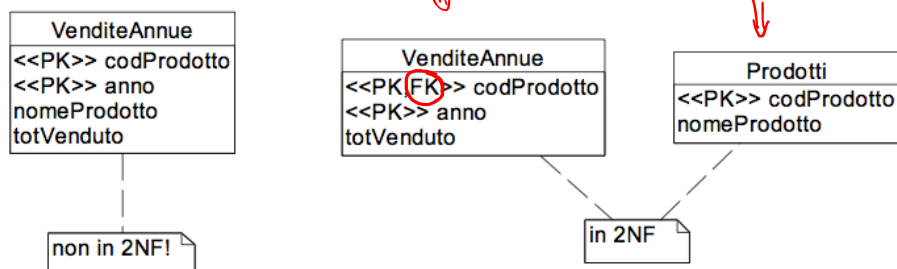
$codProdotto \rightarrow NP$ inventario risp. anno
 colla x 2
 spago x 2

Basi di Dati corso A

344

2NF

- Mediante proiezione si crea una nuova relazione che evidenzia la dipendenza funzionale

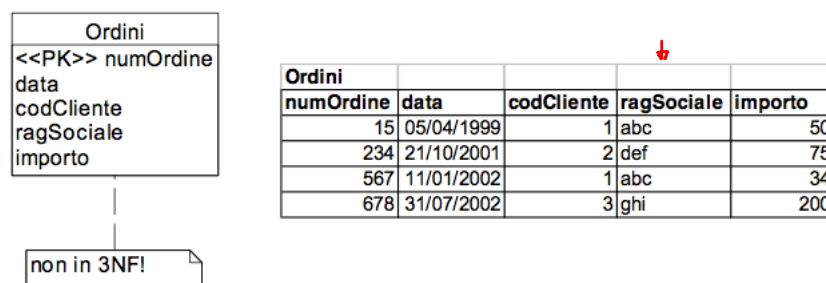


Basi di Dati corso A

345

3NF

- Una relazione è in 3NF se è in 2NF e se tra i suoi attributi non chiave non esistono FD transitive

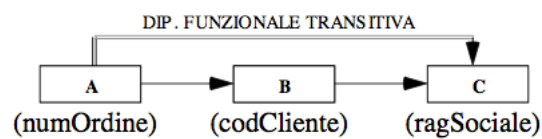


Basi di Dati corso A

346

dip. funzionale transitiva (TFD)

- In una relazione $R(\underline{A}, B, C)$
- Se $B \rightarrow C$ e $A \rightarrow B$
- Se B non è chiave candidata di R
- allora C ha TFD da A

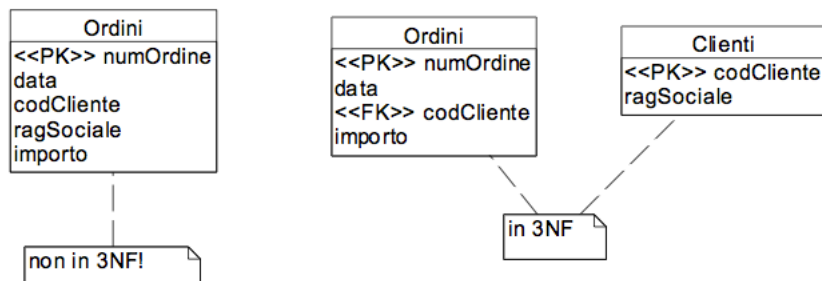


Basi di Dati corso A

347

3NF

- Mediante proiezione si crea una nuova relazione che rimuove la TFD



Basi di Dati corso A

348

BCNF

- ▶ Una relazione è in BCNF se tutti i suoi determinanti sono candidati chiave
- ▶ Ogni relazione in 3NF è in BCNF a meno che:
 - ▶ abbia 2 o + chiavi candidate
 - ▶ le chiavi candidate siano composte

Basi di Dati corso A

349

Dipendenza multivalente

Prodotto	Taglia	Colore
T-Shirt	S	Bianco
T-Shirt	S	Giallo
T-Shirt	S	Rosso
T-Shirt	M	Bianco
T-Shirt	M	Giallo
T-Shirt	M	Rosso
T-Shirt	L	Bianco
T-Shirt	L	Giallo
T-Shirt	L	Rosso
T-Shirt	XL	Bianco
T-Shirt	XL	Giallo
T-Shirt	XL	Rosso
Camicia	S	Blu
Camicia	S	Verde
Camicia	M	Blu
Camicia	M	Verde
Camicia	L	Blu
Camicia	L	Verde

- ▶ Se un attributo determina in modo indipendente più attributi
- ▶ $\text{prodotto} \twoheadrightarrow \text{taglia}$
- ▶ $\text{prodotto} \twoheadrightarrow \text{colore}$
- ▶ Prodotto 'multidetermina' taglia e colore
- ▶ Esiste:
 - ▶ Dipendenza multivalente
 - ▶ ridondanza

Basi di Dati corso A

352

dependenza multivalente (MVD)

- ▶ Dati U e V sottoinsiemi di attributi di R , se W è l'insieme di attributi di R non compresi in U e in V , la MVD $U \twoheadrightarrow V$ esiste in R se e solo se
 - ▶ $R = R(UV) \text{ JOIN } R(W)$
- ▶ MVD è generalizzazione di FD
 - ▶ Tutte le FD sono MVD
 - ▶ Non tutte le MVD sono FD
- ▶ $(A \rightarrow B)$ può essere visto come caso speciale di $A \twoheadrightarrow B$

Basi di Dati corso A

353

MVD

- ▶ Se la MVD è FD non causa ridondanze
 - ▶ In prodotti(cod, qtà, prezzo) esistono le seguenti FD: $\text{cod} \rightarrow \text{qtà}$ e $\text{cod} \rightarrow \text{prezzo}$
 - ▶ Si può effettuare una proiezione di Prodotti in: (cod,qtà) e (cod,prezzo) e ricostruire con equijoin
 - ▶ In prodotti sussistono $\text{cod} \twoheadrightarrow \text{qtà}$ e $\text{cod} \twoheadrightarrow \text{prezzo}$, che sono anche FD

Basi di Dati corso A

354

4NF

ProdottoTagliaColore		
Prodotto	Taglia	Colore
T-Shirt	S	Bianco
T-Shirt	S	Giallo
T-Shirt	S	Rosso
T-Shirt	M	Bianco
T-Shirt	M	Giallo
T-Shirt	M	Rosso
T-Shirt	L	Bianco
T-Shirt	L	Giallo
T-Shirt	L	Rosso
T-Shirt	XL	Bianco
T-Shirt	XL	Giallo
T-Shirt	XL	Rosso
Camicia	S	Blu
Camicia	S	Verde
Camicia	M	Blu
Camicia	M	Verde
Camicia	L	Blu
Camicia	L	Verde

- R è in 4NF se e solo se, se compare una dipendenza multivalente $A \twoheadrightarrow B$, tutti gli attributi di R dipendono anche funzionalmente da A

non in 4NF!

ProdottoTaglia	
Prodotto	Taglia
T-Shirt	S
T-Shirt	M
T-Shirt	L
T-Shirt	XL
Camicia	S
Camicia	M
Camicia	L

in 4NF

ProdottoColore	
Prodotto	Colore
T-Shirt	Bianco
T-Shirt	Giallo
T-Shirt	Rosso
Camicia	Blu
Camicia	Verde

Basi di Dati corso A

355

dipendenza di Join (JD)

- Esistono relazioni in 4NF con anomalie
- Un concessionario che vende un articolo e rappresenta un produttore. Se il produttore produce quell'articolo, allora il concessionario vende l'articolo del produttore

ConcessionarioArticoloProduttore		
Concessionario	Articolo	Produttore
Neri	PC	IBM
Neri	PC	HP
Neri	Scanner	HP
Verdi	PC	HP

Basi di Dati corso A

356

JD

- ▶ Non si tratta di FD, né di MVD ma di dipendenza di join (JD)
- ▶ $R(X,Y,Z)$ soddisfa la JD se e solo se
 - ▶ $R = \pi(X) \text{ JOIN } \pi(Y) \text{ JOIN } \pi(Z)$
- ▶ JD è la dipendenza più generale
 - ▶ FD caso particolare di MVD
 - ▶ MVD caso particolare di JD

Basi di Dati corso A

357

5NF

- ▶ R è in 5NF se e solo se ogni dipendenza di join presente in essa è conseguenza delle sue chiavi candidate

ConcessionarioArticoloProduttore		
Concessionario	Articolo	Produttore
Neri	PC	IBM
Neri	PC	HP
Neri	Scanner	HP
Verdi	PC	HP

non in
5NF!

ConcessionarioArticolo	
Concessionario	Articolo
Neri	PC
Neri	Scanner
Verdi	PC

ConcessionarioProduttore	
Concessionario	Produttore
Neri	IBM
Neri	HP
Verdi	HP

ArticoloProduttore	
Articolo	Produttore
PC	IBM
PC	HP
Scanner	HP

in
5NF

Basi di Dati corso A

358

JD e 5NF

- ▶ Se la dipendenza di join è conseguenza delle chiavi candidate, la relazione è già in 5NF
 - ▶ In dipendenti(matricola,CF,nome) matricola e CF sono chiavi candidate
 - ▶ Si può proiettare: (matricola,CF), (CF, nome), (matricola, nome), e ricostruire con join su matricola e CF
 - ▶ Dipendenti contiene JD, ma non presenta anomalie => è già in 5NF

Basi di Dati corso A

359

First Normal Form

- ▶ Disallows composite attributes, multivalued attributes, and **nested relations**; attributes whose values *for an individual tuple* are non-atomic
- ▶ Considered to be part of the definition of relation

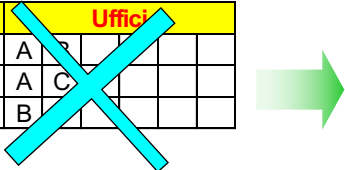
Basi di Dati corso A

364

Prima forma normale

- In una tabella non possono esistere colonne definite per contenere una molteplicità di valori

N. Protocollo	Ufficio
1	A
2	A C
3	B



N. Protocollo	Ufficio
1	A
1	B
2	A
2	C
3	B
3	C

Basi di Dati corso A

365

Second Normal Form

- A relation schema R is in **second normal form (2NF)** if every non-prime attribute A in R is fully functionally dependent on the primary key
- R can be decomposed into 2NF relations via the process of 2NF normalization

Basi di Dati corso A

366

Seconda forma normale

- Tutti gli attributi devono dipendere dall'intera chiave primaria

Codice_citta	Codice_Via	Citta	Via
01	01	Roma	Verdi
01	02	Roma	Bianchi
01	03	Roma	Rossi
02	01	Milano	Gialli

Codice_citta	Codice_Via	Via
01	01	Verdi
01	02	Bianchi
01	03	Rossi
02	01	Gialli

Codice_citta	Citta
01	Roma
02	Milano

Basi di Dati corso A

367

Third Normal Form

Definition:

- **Transitive functional dependency** - a FD $X \rightarrow Z$ that can be derived from two FDs $X \rightarrow Y$ and $Y \rightarrow Z$

Basi di Dati corso A

368

Third Normal Form

- ▶ A relation schema R is in **third normal form (3NF)** if it is in 2NF *and* no non-prime attribute A in R is transitively dependent on the primary key
- ▶ R can be decomposed into 3NF relations via the process of 3NF normalization

NOTE:

In $X \rightarrow Y$ and $Y \rightarrow Z$, with X as the primary key, we consider this a problem only if Y is not a candidate key. When Y is a candidate key, there is no problem with the transitive dependency.

E.g., Consider EMP (SSN, Emp#, Salary).

Here, $SSN \rightarrow Emp\# \rightarrow Salary$ and Emp# is a candidate key.

Basi di Dati corso A

369

Terza forma normale

- ▶ Non devono esserci dipendenze tra gli attributi se non basate sulla chiave primaria

N. Protocollo	Mittente	Tipo	Urgenza
1	001	1	si
2	001	2	no
3	003	1	si
4	002	2	no
5	003	3	si

Tipo	Descrizione
1	Lettera
2	Memo
3	Telegramma

NP \rightarrow T
T \rightarrow U

Tipo	Descrizione	Urgenza
1	Lettera	si
2	Memo	no
3	Telegramma	si

Basi di Dati corso A

370

BCNF (Boyce-Codd Normal Form)

- ▶ A relation schema R is in **Boyce-Codd Normal Form (BCNF)** if whenever an FD $X \rightarrow A$ holds in R , then X is a superkey of R
- ▶ Each normal form is strictly stronger than the previous one
 - ▶ Every 2NF relation is in 1NF
 - ▶ Every 3NF relation is in 2NF
 - ▶ Every BCNF relation is in 3NF
- ▶ There exist relations that are in 3NF but not in BCNF
- ▶ The goal is to have each relation in BCNF (or 3NF)

Basi di Dati corso A