

# Lezione 9 – La Terza Forma Normale (3NF) - I

Prof.ssa Maria De Marsico  
demarsico@di.uniroma1.it



SAPIENZA  
UNIVERSITÀ DI ROMA

# Quali sono i problemi di uno schema mal progettato?

---



- Ritorniamo al nostro esempio di base di dati che contiene le informazioni sugli studenti e sugli esami sostenuti, e ripartiamo dalla soluzione «buona» trovata alla fine

La base di dati consiste di quattro schemi di relazione:

- Studente (Matr, CF, Cogn, Nome, Data, Com)
- Corso (C#, Tit, Doc)
- Esame (Matr, C#, Data, Voto)
- Comune (Com, Prov)

## Ipotesi 3 (considerazioni)



- Studente (Matr, CF, Cogn, Nome, Data, Com)
- Poiché il numero di matricola identifica univocamente uno studente, ad ogni numero di matricola corrisponde:
  - un solo codice fiscale ( Matr  $\rightarrow$  CF )
  - un solo cognome ( Matr  $\rightarrow$  Cogn )
  - un solo nome ( Matr  $\rightarrow$  Nome )
  - una sola data di nascita ( Matr  $\rightarrow$  Data )
  - un solo comune di nascita ( Matr  $\rightarrow$  Com )
- Quindi un'istanza di Studente per essere **legale** deve soddisfare la dipendenza funzionale
  - Matr  $\rightarrow$  Matr CF Cogn Nome Data Com



- Con considerazioni analoghe abbiamo che un'istanza di **Studiante** per essere legale deve soddisfare la dipendenza funzionale
  - $CF \rightarrow \text{Matr CF Cogn Nome Data Com}$
- Pertanto sia **Matr** che **CF** sono **chiavi** di **Studiante**.

## Ipotesi 3 (considerazioni)



- D'altra parte possiamo osservare che ci possono essere due studenti con lo stesso cognome e nomi differenti quindi
- **possiamo avere istanze di Studente che non soddisfano la dipendenza funzionale**
  - Cogn  $\rightarrow$  Nome

Studente	Matr	CF	Cogn	Nome	Data	Com
	01	...	Rossi	Mario	1/1/1989	Roma
	02	...	Bianchi	Paolo	1/1/1989	Roma
	03	...	Rossi	Paolo	30/11/1988	Marino
	04	...	Neri	Paolo	25/10/1988	Marino

**possiamo avere istanze di Studente che **non** soddisfano la dipendenza funzionale**

Cogn → Nome

Cogn → Data

Cogn → Com

Nome → Cogn

... ecc.

Con considerazioni analoghe possiamo concludere che **le uniche dipendenze funzionali non banali che devono essere soddisfatte da un'istanza legale di Studente** sono del tipo

$$K \rightarrow X$$

dove  $K$  contiene una chiave (Matr o CF)

Vedremo che questa è una prima condizione che **però va ulteriormente rifinita** per arrivare ad una definizione precisa di Terza Forma Normale (3NF)



- Esame (Matr, C#, Data, Voto)

Uno studente può sostenere l'esame relativo ad un corso **una sola volta**; pertanto **per ogni esame** (identificato dallo studente e dal corso, quindi da Matr# C#) esiste

- una sola data (in cui è stato sostenuto)
- un solo voto

Quindi

ogni istanza legale di **Esame** deve soddisfare la dipendenza funzionale

$\text{Matr C\#} \rightarrow \text{Data Voto}$



- Esame (Matr, C#, Data, Voto)
- D'altra parte uno studente può sostenere esami in date differenti e riportare voti diversi nei vari esami.
- Pertanto esistono istanze di Esame che **non** soddisfano una o entrambe le dipendenze funzionali

Matr  $\rightarrow$  Data

Matr  $\rightarrow$  Voto

## Ipotesi 3 (considerazioni)



SAPIENZA  
UNIVERSITÀ DI ROMA  
DIPARTIMENTO DI INFORMATICA

- Esame (Matr, C#, Data, Voto)
- Inoltre l'esame relativo ad un certo corso può essere superato da diversi studenti in date diverse e con diversi voti.
- Pertanto esistono istanze di Esame che **non** soddisfano una o entrambe le dipendenze funzionali

$C\# \rightarrow Data$

$C\# \rightarrow Voto$

Pertanto **Matr C#** è una **chiave** per Esame  
(si vede facilmente che è anche **l'unica** chiave)

In seguito vedremo delle procedure **rigorose** per identificare la(le) chiavi.

## Ipotesi 3 (conclusioni)



Per ciascun schema di relazione

- Studente (**Matr**, CF, Cogn, Nome, Data, Com)
  - Corso (**C#**, Tit, Doc)
  - Esame (**Matr**, **C#**, Data, Voto)
  - Comune (**Com**, Prov)
- **ATTENZIONE!!!** Stiamo continuando ad assumere che  $COM \rightarrow Prov$ , cioè che non ci sono comuni omonimi (cosa che in realtà non è vera ...)

le uniche dipendenze funzionali non banali che devono essere soddisfatte da ogni istanza legale sono del tipo

$$K \rightarrow X$$

dove K contiene una chiave

Uno schema di relazione è in 3NF se  
le uniche dipendenze funzionali non banali che devono  
essere soddisfatte da ogni istanza legale sono del tipo

$$K \rightarrow X$$

dove

- K contiene una chiave oppure
- X è contenuto in una chiave

Anche questa condizione richiede ancora di essere rifinita  
per giungere alla definizione finale di 3NF

### Definizione

- Dati uno schema di relazione  $R$  e un insieme di dipendenze funzionali  $F$  su  $R$ ,  $R$  è in **3NF** se

$$\forall X \rightarrow A \in F^+, A \notin X$$

- $A$  appartiene ad una chiave (è *primo*)  
oppure
- $X$  contiene una chiave (è una *superchiave*)



Abbiamo notato alcune proprietà delle dipendenze funzionali anche prima di formalizzarle con gli assiomi di Armstrong ...

# Dipendenze funzionali banali



Dati uno schema di relazione  $R$  e due sottoinsiemi non vuoti  $X, Y$  di  $R$  tali che  $Y \subseteq X$  si ha:

ogni istanza  $r$  di  $R$  soddisfa la dipendenza funzionale  $X \rightarrow Y$

<b>R</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
	a1	b1	c1	d1
	a1	b2	c1	d2
	a1	b1	c1	d3

$X=ABC$

$Y=AB$

$X \rightarrow Y$  è soddisfatta



# Dipendenze funzionali (proprietà)



Dati uno schema di relazione  $R$  e un insieme di dipendenze funzionali  $F$ , si ha:

$$X \rightarrow Y \in F^+ \Leftrightarrow \forall A \in Y (X \rightarrow A \in F^+)$$

- $X \rightarrow Y$  deve essere soddisfatta da **ogni** istanza legale di  $R$  (è in  $F^+$ !)
- Se  $t_1[X]=t_2[X]$  allora deve essere  $t_1[Y]=t_2[Y]$
- Ovviamente se  $A \in Y$  e  $t_1[A] \neq t_2[A]$ , non può essere  $t_1[Y]=t_2[Y]$
- Ovviamente se  $\forall A \in Y t_1[A]=t_2[A]$ , avremo  $t_1[Y]=t_2[Y]$

R	A	B	C	D
	a1	b1	c1	d1
	a2	b2	c1	d2
	a1	b1	c1	d3

$$A \rightarrow BC \in F^+$$

$\Downarrow \Uparrow$

$$A \rightarrow B \in F^+$$

$$A \rightarrow C \in F^+$$

- *Definizione*

Dati uno schema di relazione  $R$  e un insieme di dipendenze funzionali  $F$  su  $R$ ,  $R$  è in **3NF** se

$$\forall X \rightarrow A \in F^+, A \notin X$$

- $A$  appartiene ad una chiave (è *primo*) oppure
- $X$  contiene una chiave (è una *superchiave*)

## Nota: Attenzione.

- Per quanto detto, è **sbagliato scrivere**  $\forall X \rightarrow A \in F$ , perché non sapremmo se e come valutare una dipendenza del tipo  $X \rightarrow AB$  (due o più attributi a destra)
- Se sostituisco  $\forall X \rightarrow A \in F$ , con  $\forall X \rightarrow Y \in F$ , non so come comportarmi se  $Y$  contiene sia attributi primi che non primi

# Esempio 1



$R=ABCD$

$F=\{A \rightarrow B, B \rightarrow CD\}$

La chiave è A infatti **per ogni istanza legale (che deve soddisfare le dipendenze in F)**

- Se  $t_1[A] = t_2[A]$  allora  $t_1[B] = t_2[B]$  (la dipendenza deve essere soddisfatta) ... ma anche ...
- Se  $t_1[B] = t_2[B]$  allora  $t_1[CD] = t_2[CD]$  (la dipendenza deve essere soddisfatta) e quindi
- Se  $t_1[A] = t_2[A]$  allora  $t_1[CD] = t_2[CD]$  quindi  $A \rightarrow CD \in F^A$  e per la regola dell'unione
- $A \rightarrow BCD \in F^A$  ... e sappiamo che  $F^A = F^+$  ... quindi  **$A \rightarrow R \in F^+$**  ... inoltre A è singleton quindi non ha sottoinsiemi ... quindi **A è una chiave** (vedremo poi come si trova la/le chiavi di uno schema)
- A è l'unica chiave perché B non determina A, e sia C che D non determinano altri attributi

# Esempio 1



$R=ABCD$

$F=\{A \rightarrow B, B \rightarrow CD\}$

La chiave è A

Vediamo se R è in 3NF ... e **valutiamo solo le dipendenze in F**

- $A \rightarrow B$  è OK (A è superchiave)
- $B \rightarrow CD$  ? Secondo la definizione dobbiamo controllare le dipendenze  $X \rightarrow A$  con A singleton
- ...ma ...  $B \rightarrow C$  e  $B \rightarrow D$  non sono in F ... ma in  $F^+$  ... e allora?  
**Significa che va tutto bene? NO**

**Infatti** se consideriamo l'insieme giusto scopriamo che sia  $B \rightarrow C$  che  $B \rightarrow D$  **violano** la 3NF perché in entrambi i casi B **non è superchiave** ... quindi **lo schema R non è in 3NF**

## Esempio 2

$R=ABCD$

$F=\{AC \rightarrow B, B \rightarrow AD\}$

La chiave è AC infatti **per ogni istanza legale (che deve soddisfare le dipendenze in F)**

- Se  $t_1[AC]=t_2[AC]$  allora  $t_1[B]=t_2[B]$  (la dipendenza deve essere soddisfatta) ... ma anche ...
- Se  $t_1[B]=t_2[B]$  allora  $t_1[AD]=t_2[AD]$  (la dipendenza deve essere soddisfatta) e quindi
- Se  $t_1[AC]=t_2[AC]$  allora  $t_1[AD]=t_2[AD]$  quindi  $AC \rightarrow AD \in F^A$  e per la regola dell'unione e la riflessività
- $AC \rightarrow ABCD \in F^A$  ... e sappiamo che  $F^A=F^+$  ... quindi  **$AC \rightarrow R \in F^+$**  ... inoltre A da solo non determina altri attributi, quindi  $A \rightarrow R \notin F^+$  lo stesso vale per C da solo... quindi **AC è chiave** (vedremo poi come si trova la chiave di uno schema)
- BC è un'altra chiave ( $B \rightarrow AD$  e con l'aumento  $BC \rightarrow ACD$  e B da solo determina solo AD e poi non possiamo applicare la transitività)
- ABC è una superchiave

## Esempio 2

$R=ABCD$

$F=\{AC \rightarrow B, B \rightarrow AD\}$

Le chiavi sono AC e BC

Vediamo se R è in 3NF ... e **valutiamo solo le dipendenze in F**

- $AC \rightarrow B$  è OK (AC è superchiave)
- $B \rightarrow AD$  ? Secondo la definizione dobbiamo controllare le dipendenze  $X \rightarrow A$  con A singleton
- ...ma ...  $B \rightarrow A$  e  $B \rightarrow D$  non sono in F ... ma in  $F^+$  ... e allora?  
**Significa che va tutto bene? NO**

**Infatti** se consideriamo l'insieme giusto scopriamo che  $B \rightarrow A$  è OK perché A è parte di una chiave (primo) **ma  $B \rightarrow D$  viola** la 3NF perché **B non è superchiave (è solo parte di una chiave) e D non è parte di una chiave ... quindi lo schema R non è in 3NF**

## Esempio 2 - Commento



$R=ABCD$

$F=\{AC \rightarrow B, B \rightarrow AD\}$

Se usiamo una definizione alternativa

- $\forall X \rightarrow Y \in F^+, Y \not\subset X$
- $Y$  appartiene ad una chiave (è *primo*) oppure
- $X$  contiene una chiave (è una *superchiave*)

B è parte di una chiave ma ... come facciamo a valutare la seconda condizione su  $B \rightarrow AD$  visto che  $Y=AD$ , e A è primo (appartiene ad una chiave) **ma D no?**

Dovremmo dare una definizione **più complessa** che prevede di esaminare ogni attributo a destra singolarmente ... **e comunque** ... sappiamo di poter applicare sempre la decomposizione

Ma allora è più semplice ed elegante usare la definizione data.

## Esempio 3



$R=ABCD$

$F=\{AB \rightarrow CD, BC \rightarrow A, D \rightarrow AC\}$

Abbiamo tre chiavi, AB, BC, e DB infatti **per ogni istanza legale (che deve soddisfare le dipendenze in F)**

- Se  $t_1[AB]=t_2[AB]$  allora  $t_1[CD]=t_2[CD]$  (la dipendenza deve essere soddisfatta) ... quindi aggiungendo la riflessività  $AB \rightarrow AB$  abbiamo che  $AB \rightarrow R$  .. Inoltre
- A da solo non determina nulla (solo sé stesso per riflessività)
- B da solo non determina nulla (solo sé stesso per riflessività)
- ... quindi AB è chiave

Inoltre **per ogni istanza legale (che deve soddisfare le dipendenze in F)**

- In alternativa applichiamo Armstrong tanto sappiamo che  $F^+ = F^A$
- $BC \rightarrow A$  in  $F^A$  + aumento  $BC \rightarrow AB$  in  $F^A$  + transitività con  $AB \rightarrow CD$  ci porta  $BC \rightarrow CD$  in  $F^A$  quindi con la riflessività e l'unione  $BC \rightarrow R$  in  $F^A$
- C da solo non determina nulla quindi BC altra chiave
- D da solo non arriva a B ma con B è un'altra chiave



## Esempio 3



SAPIENZA  
UNIVERSITÀ DI ROMA

LA

$R=ABCD$

$F=\{AB \rightarrow CD, BC \rightarrow A, D \rightarrow AC\}$

Abbiamo tre chiavi, AB, BC, e DB

Vediamo se R è in 3NF ... e **valutiamo solo le dipendenze in F**

- $AB \rightarrow CD$  è OK (AB è superchiave)
- $BC \rightarrow A$  ? è OK (BC è superchiave e inoltre A è primo)
- ...ma ...  $D \rightarrow AC$  ?

D è solo PARTE di una chiave, e AC non è parte di nessuna chiave ...  
quindi? Dobbiamo concludere che lo schema non è 3NF? NO

Se applichiamo la decomposizione (quindi siamo in  $F^+$ ) possiamo considerare

$D \rightarrow A$  OK, A è primo!

$D \rightarrow C$  OK, C è primo

I due attributi a destra fanno parte di chiavi diverse ma sono ENTRAMBI primi, quindi lo schema E' 3NF

- *Definizione*

Dati uno schema di relazione  $R$  e un insieme di dipendenze funzionali  $F$  su  $R$ ,  $R$  è in **3NF** se

$$\forall X \rightarrow A \in F^+, A \notin X$$

- $A$  appartiene ad una chiave (è *primo*) oppure
- $X$  contiene una chiave (è una *superchiave*)

- **Nota: Attenzione.**

- La condizione  $A \notin X$  è **importante**. Infatti, per l'assioma della **riflessività**, se  $A \in X$  avremo sempre  $X \rightarrow A$  in  $F^A$  e quindi in  $F^+$ , anche quando  $A$  non è primo e  $X$  non è superchiave, e quindi se considerassimo questo tipo di dipendenze **nessuno** schema risulterebbe in 3NF

- $R=AB$
- $F=\{A \rightarrow B\}$
- La chiave è ovviamente A

Vediamo se R è in 3NF ... e **valutiamo le dipendenze in  $F^+$**

AB contiene la chiave ... OK

$F^+ = \{A \rightarrow B, AB \rightarrow AB, A \rightarrow A, B \rightarrow B \dots\}$

A è chiave ... OK

**$B \rightarrow B$  dovrebbe essere una violazione perché B non è parte della chiave né contiene la chiave!!!**

Ma questo tipo di dipendenze (banali) si trova **SEMPRE** in  $F^+$   
**QUINDI NON vanno considerate**

### Ciascuno schema di relazione

- Studente (**Matr**, CF, Cogn, Nome, Data, Com)  
Matr  $\rightarrow$  Matr CF Cogn Nome Data Com  
CF  $\rightarrow$  Matr CF Cogn Nome Data Com  
Matr  $\rightarrow$  CF  
CF  $\rightarrow$  Matr
- Corso (**C#**, Tit, Doc)  
C#  $\rightarrow$  C# Tit Doc
- Esame (**Matr**, **C#**, Data, Voto)  
Matr C#  $\rightarrow$  Matr C# Data Voto
- Comune (**Com**, Prov)  
Com  $\rightarrow$  Com Prov

**è in 3NF**

Stiamo continuando ad assumere che non ci sono comuni omonimi



- Torniamo a considerare le altre ipotesi di decomposizione



La base di dati consiste di tre schemi di relazione:

- Studente (Matr, CF, Cogn, Nome, Data, Com, Prov)
- Corso (C#, Tit, Doc)
- Esame (Matr, C#, Data, Voto)

## Ipotesi 2 (considerazioni)



SAPIENZA  
UNIVERSITÀ DI ROMA  
DIPARTIMENTO DI INFORMATICA

- Studente (Matr, CF, Cogn, Nome, Data, Com, **Prov**)

Un'istanza di Studente per essere **legale** deve soddisfare le dipendenze funzionali:

- Matr  $\rightarrow$  Matr CF Cogn Nome Data Com **Prov**
- CF  $\rightarrow$  Matr CF Cogn Nome Data Com **Prov**
- Matr  $\rightarrow$  CF
- CF  $\rightarrow$  Matr

### Inoltre

poiché ogni comune si trova in una sola provincia deve soddisfare la dipendenza funzionale

- Com  $\rightarrow$  Prov

Stiamo continuando ad assumere che non ci sono comuni omonimi

## Ipotesi 2 (considerazioni)



- Studente (Matr, CF, Cogn, Nome, Data, Com, Prov)

Consideriamo la dipendenza funzionale

Com  $\rightarrow$  Prov

è facile vedere che:

- Com **non è una chiave** per Studente (ci possono essere più studenti che sono nati nello stesso comune)
- Prov **non appartiene ad alcuna chiave** di Studente (le uniche chiavi sono Matr e CF)

Stiamo continuando ad assumere che non ci sono comuni omonimi





Lo schema

Studente (Matr, CF, Cogn, Nome, Data, Com, Prov)

**non è in 3NF**

La base di dati consiste di un **unico** schema di relazione:

Curriculum (Matr, CF, Cogn, Nome, DataN, Com, Prov, C#, Tit, Doc, DataE, Voto)

# Ipotesi 1 (considerazioni)



- Curriculum (Matr, CF, Cogn, Nome, DataN, Com, Prov, C#, Tit, Doc, DataE, Voto)

Un'istanza di Curriculum per essere **legale** deve soddisfare le dipendenze funzionali:

- $\text{Matr} \rightarrow \text{Matr CF Cogn Nome DataN Com Prov}$
- $\text{CF} \rightarrow \text{Matr CF Cogn Nome DataN Com Prov}$ 
  - $\text{CF} \rightarrow \text{Matr}$
  - $\text{Matr} \rightarrow \text{CF}$
  - $\text{Com} \rightarrow \text{Prov}$
  - $\text{C\#} \rightarrow \text{C\# Tit Doc}$
- $\text{Matr C\#} \rightarrow \text{DataE Voto}$
- $\text{CF C\#} \rightarrow \text{DataE Voto}$

Stiamo continuando ad assumere che non ci sono comuni omonimi

# Ipotesi 1 (considerazioni)



Pertanto le (uniche) chiavi di Curriculum (Matr, CF, Cogn, Nome, DataN, Com, Prov, C#, Tit, Doc, DataE, Voto) sono:

- **Matr C#**
- **CF C#**

Infatti ogni istanza legale di Curriculum soddisfa le dipendenze funzionali:

- Matr C#  $\rightarrow$  Matr CF Cogn Nome Data Com Prov C# Tit Doc DataE Voto
- CF C#  $\rightarrow$  Matr CF Cogn Nome Data Com Prov C# Tit Doc DataE Voto

# Ipotesi 1 (conclusioni)



SAPIENZA  
UNIVERSITÀ DI ROMA  
DIPARTIMENTO DI INFORMATICA

- Curriculum (Matr, CF, Cogn, Nome, DataN, Com, Prov, C#, Tit, Doc, DataE, Voto)

Consideriamo la dipendenza funzionale

- $\text{Matr} \rightarrow \text{Cogn}$

poichè:

- Matr **non** è una chiave per Curriculum (uno studente può aver sostenuto più esami – ricordiamo che per **questo** schema la chiave è **Matr C#** oppure **C# Matr**)
- Cogn **non appartiene** ad alcuna chiave di Curriculum (le uniche chiavi sono Matr C# e CF C#)

**Curriculum non è in 3NF**

**Nota:** basta identificare anche **UNA SOLA** dipendenza che viola le condizioni per la 3NF

## Considerazioni finali: dipendenze transitive



SAPIENZA  
UNIVERSITÀ DI ROMA

- Studente (**Matr**, CF, Cogn, Nome, Data, Com, Prov)
- Ad un numero di matricola corrisponde un solo comune di nascita (quello dello studente con quel numero di matricola):  $\text{Matr} \rightarrow \text{Com}$
- Un comune si trova in una sola provincia:  $\text{Com} \rightarrow \text{Prov}$

### Conclusione:

- ad un numero di matricola corrisponde una sola provincia:  
 $\text{Matr} \rightarrow \text{Prov}$

(ricordiamo che poiché Com non è superchiave e Prov non è primo lo schema non è 3NF)

- La dipendenza funzionale  $\text{Matr} \rightarrow \text{Prov}$  è una conseguenza delle due dipendenze funzionali

$\text{Matr} \rightarrow \text{Com}$  e  $\text{Com} \rightarrow \text{Prov}$

$\text{Com} \rightarrow \text{Prov}$  viene detta

**dipendenza transitiva**

**(vedremo in seguito la definizione formale)**

## Considerazioni finali: dipendenze parziali



SAPIENZA  
UNIVERSITÀ DI ROMA  
DIPARTIMENTO DI INFORMATICA

- Curriculum (**Matr**, CF, Cogn, Nome, DataN, Com, Prov, **C#**, Tit, Doc, DataE, Voto)

Ad un numero di matricola corrisponde un solo cognome (il cognome dello studente con quel numero di matricola):

- $\text{Matr} \rightarrow \text{Cogn}$

Quindi:

ad una coppia costituita da un numero di matricola e da un codice di corso corrisponde un solo cognome:  $\text{Matr C\#} \rightarrow \text{Cogn}$

La dipendenza funzionale  $\text{Matr C\#} \rightarrow \text{Cogn}$  è una conseguenza della dipendenza funzionale  $\text{Matr} \rightarrow \text{Cogn}$  che viene detta

- **dipendenza parziale**