

Introduzione alla normalizzazione dei dati

versione 16 marzo 2009

© Adriano Comai

<http://www.analisi-disegno.com>

Obiettivo di questa introduzione

- Fornire elementi di base sulla normalizzazione dei dati
- Il tema è trattato in modo più approfondito nel modulo formativo online su <http://www.adcorsi.com>
- E nel corso in aula "Analisi dati e progettazione logica di database“:

http://www.analisi-disegno.com/a_comai/corsi/sk_dati.htm

Dipendenza funzionale

- 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**

$$X \longrightarrow Y$$

- Y dipende funzionalmente da X
- X è il determinante di Y

Dipendenza funzionale

- la dipendenza funzionale esiste a livello semantico (la si può scoprire solo addentrandosi nel significato dei dati e dei loro rapporti)

idPersona \longrightarrow nome

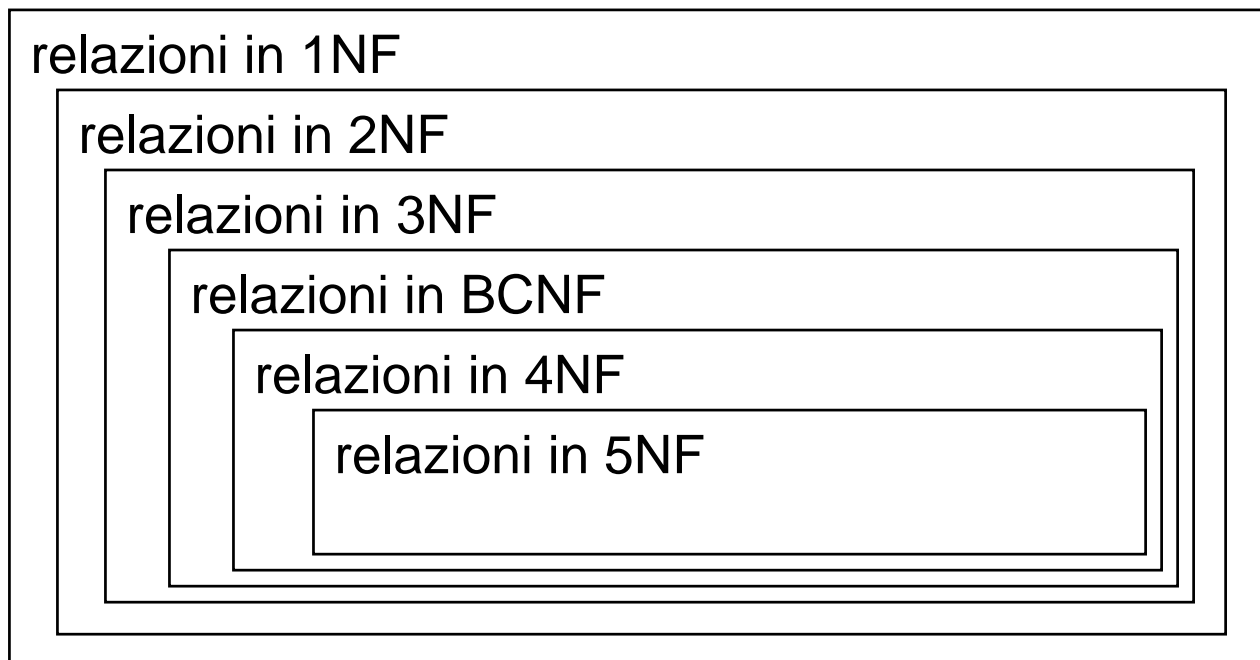
(per ogni idPersona esiste al massimo un nome)

anno, numFattura \longrightarrow importo

(per ogni (anno + numFattura) esiste al massimo un importo)

Normalizzazione

- forma normale (NF): un insieme di condizioni da rispettare per garantire l'eliminazione di determinate anomalie e ridondanze



Processo di normalizzazione

- la normalizzazione può essere vista come un processo sistematico, basato sull'applicazione ripetuta dell'operatore relazionale di proiezione
- obiettivo: eliminare ridondanze ed anomalie

Prima forma normale (Codd)

- una relazione è in prima forma normale (1NF) se, e solo se, ogni tupla contiene esattamente un unico valore per ogni attributo

	Attr1	Attr2	Attr3
Tupla			
1 Tupla		x	
2 Tupla			
3 Tupla			

4

- tutte le relazioni normalizzate sono almeno in 1NF

Seconda forma normale (Codd)

- una relazione è in seconda forma normale (2NF) se è in 1NF, e se ogni suo attributo non chiave dipende funzionalmente dalla chiave completa

VenditeAnnue
<<PK>> codProdotto
<<PK>> anno
nomeProdotto
totVenduto

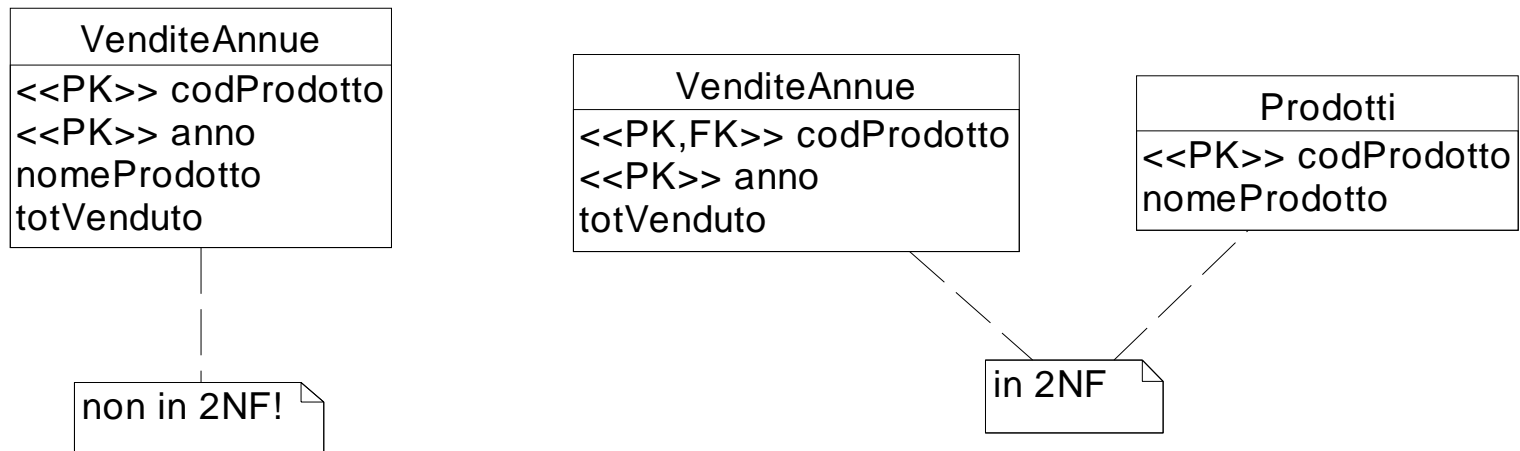
non in 2NF!

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

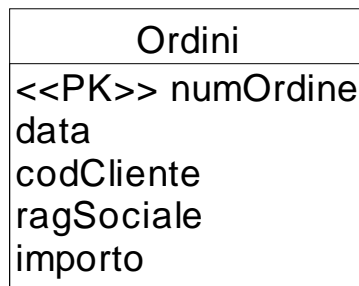
Seconda forma normale

- creare (con una proiezione) una nuova relazione, che evidenzi la dipendenza funzionale precedentemente “nascosta”



Terza forma normale (Codd)

- una relazione è in terza forma normale (3NF) se è in 2NF, e se tra i suoi attributi non chiave non esistono dipendenze funzionali transitive



non in 3NF!

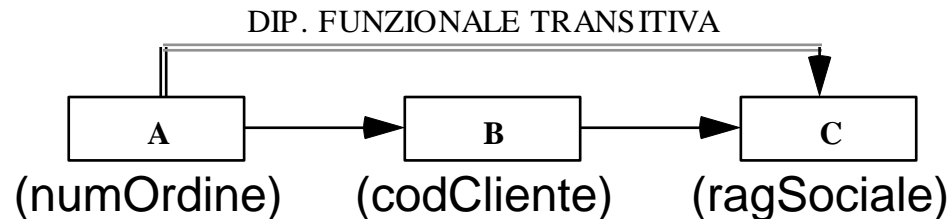
Ordini

numOrdine	data	codCliente	ragSociale	importo
15	05/04/1999	1	abc	50
234	21/10/2001	2	def	75
567	11/01/2002	1	abc	34
678	31/07/2002	3	ghi	200

Dipendenza funzionale transitiva

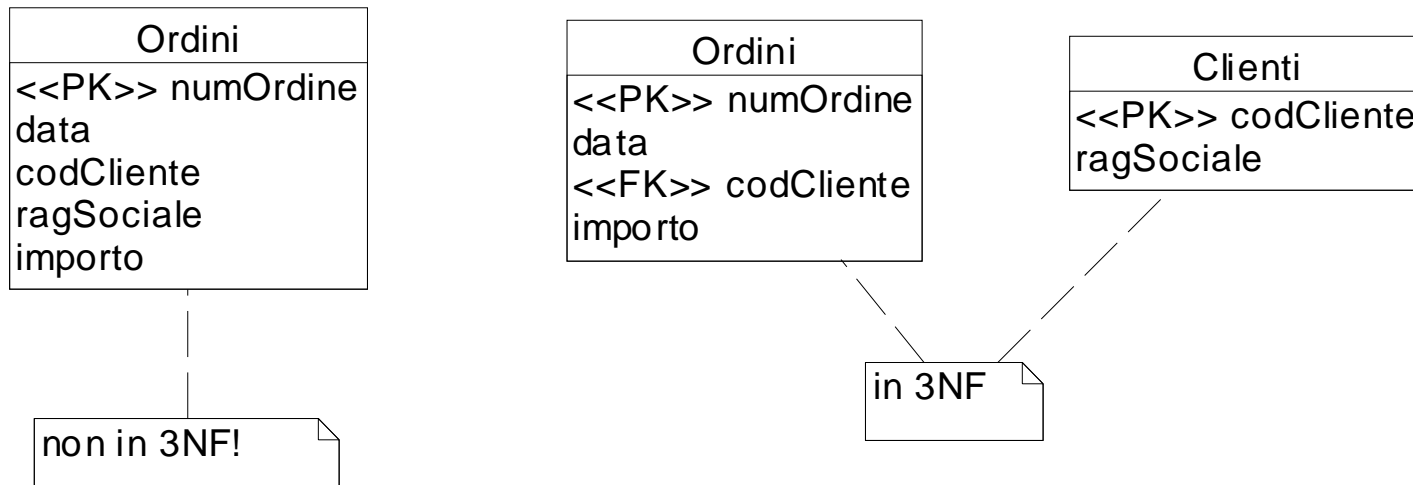
Data una relazione con attributi A, B e C, e con PK A

- se C dipende funzionalmente dal determinante B, a sua volta funzionalmente dipendente da A,
- se il determinante B non è una chiave candidata della relazione, alternativa ad A,
- allora C ha una dipendenza funzionale transitiva da A



Terza forma normale

- creare (con una proiezione) una nuova relazione, che rimuova la dipendenza funzionale transitiva



Forma normale di Boyce e Codd

- una relazione è in forma normale di Boyce e Codd (BCNF) se tutti i suoi determinanti sono candidati chiave
- ogni relazione in 3NF è anche in BCNF, a meno che:
 - abbia 2 o più chiavi candidate
 - le chiavi candidate siano composte
 - abbiano almeno un attributo in comune

Forma normale di Boyce e Codd

AllievoMaterialInsegnante

Allievo	Materia	Insegnante
Carlo	Matematica	Prof. Bianchi
Carlo	Fisica	Prof. Verdi
Grazia	Matematica	Prof. Bianchi
Grazia	Fisica	Prof. Rossi

dipendenze funzionali:

- Allievo, Materia \rightarrow Insegnante
- Insegnante \rightarrow Materia

chiavi candidate:

- Allievo, Materia
- Allievo, Insegnante

insegnante
non è chiave
candidata!

- AllievoMaterialInsegnante non è in Boyce-Codd NF
- ha una ridondanza (la materia di un insegnante è ripetuta su più tuple)

Forma normale di Boyce e Codd

AllievoMateriaInsegnante

Allievo	Materia	Insegnante
Carlo	Matematica	Prof. Bianchi
Carlo	Fisica	Prof. Verdi
Grazia	Matematica	Prof. Bianchi
Grazia	Fisica	Prof. Rossi

può essere scomposta mediante proiezioni
in:

AllievoInsegnante

Allievo	Insegnante
Carlo	Prof. Bianchi
Carlo	Prof. Verdi
Grazia	Prof. Bianchi
Grazia	Prof. Rossi

InsegnanteMateria

Insegnante	Materia
Prof. Bianchi	Matematica
Prof. Verdi	Fisica
Prof. Rossi	Fisica

(entrambe in Boyce-Codd NF)

Dipendenza multivalente

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

Prodotto $\rightarrow\rightarrow$ Taglia

Prodotto $\rightarrow\rightarrow$ Colore

(Prodotto “multidetermina” Taglia e Colore;
Taglia e Colore sono mutuamente
indipendenti)

- c'è una dipendenza multivalente
- ...e una ridondanza...

Dipendenza multivalente

Definizione (Fagin):

- se U e V sono sottoinsiemi degli attributi di una relazione R , e se W è l'insieme degli attributi di R non compresi in U e V , la **dipendenza multivalente** $U \twoheadrightarrow V$ esiste in R se e solo se R è il join delle sue proiezioni $R(UV)$ ed $R(UW)$
- la dipendenza multivalente (MVD) è una generalizzazione della dipendenza funzionale (FD):
 - tutte le FD sono MVD, ma
 - non tutte le MVD sono FD
- ($A \rightarrow B$ è un caso speciale della MVD $A \twoheadrightarrow B$, in cui il numero dei valori assumibili da B è pari ad 1)

Dipendenza multivalente

Se la MVD è una FD non causa ridondanze ...

- in una relazione Prodotti ($\llcorner\text{PK}\lrcorner$ codProd, qtaDisponibile, prezzo), esistono le dipendenze funzionali codProd \rightarrow qtaDisponibile e codProd \rightarrow prezzo
- la relazione può essere proiettata in due nuove relazioni (codProd, qtaDisponibile) e (codProd, prezzo), il cui equi-join su codProd ricostruisce la relazione originaria
- secondo la definizione di Fagin, nella relazione Prodotti sussistono le due dipendenze multivalenti codProd $\rightarrow\rightarrow$ qtaDisponibile e codProd $\rightarrow\rightarrow$ prezzo, che sono anche dipendenze funzionali
- e non vi sono ridondanze...

Quarta forma normale (Fagin)

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

una relazione R è in quarta forma normale (4NF) se e solo se, qualora vi compaia 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

Dipendenza di join

- esistono relazioni in 4NF che presentano ancora anomalie:

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

- vincoli: se il concessionario vende un articolo, e se rappresenta un produttore, e se il produttore produce quell'articolo, allora il concessionario vende l'articolo del produttore

Dipendenza di join

- il vincolo sulla relazione non è ne' una FD ne' una MVD, ma una **dipendenza di join (JD)**:

Definizione (Fagin):

- una relazione $R(X,Y,Z)$ soddisfa la **dipendenza di join (JD)** se e solo se R è uguale al join delle sue proiezioni su X , Y e Z , dove X , Y e Z costituiscono sottoinsiemi degli attributi di R
- la dipendenza di join è il tipo di dipendenza più generale:
 - la FD è un caso particolare di MVD
 - la MVD è un caso particolare di JD

Quinta forma normale (Fagin)

- una relazione è in quinta forma normale (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

Dipendenza di join e 5NF

- se la dipendenza di join è conseguenza delle chiavi candidate, la relazione è già in 5NF (e non va scomposta)
- in una relazione Dipendenti (<<PK>>matricola, codFiscale, nome) gli attributi matricola e codFiscale sono chiavi candidate
- la relazione può essere proiettata (matricola, codFiscale), (codFiscale, nome), (matricola, nome) e quindi ricostruita mediante una serie di join sulle chiavi candidate matricola e codFiscale
- la relazione Dipendenti contiene una dipendenza di join, ma non presenta anomalie e non è necessario scomporla per farla risultare in 5NF (lo è già)

5NF = “ultimate normal form”

- la 5NF è anche chiamata Projection / Join Normal Form
- è considerata l’ “ultima forma normale”
- le relazioni in 5NF (o PJNF) non presentano ulteriori anomalie che possano essere rimosse mediante proiezione
- non è quindi possibile migliorare ancora una relazione utilizzando gli operatori di proiezione e join

Denormalizzazione

- il processo di normalizzazione è reversibile
- è cioè possibile ricostruire una relazione in una forma normale inferiore (es. 2NF) partendo da un insieme di relazioni in una forma normale superiore (es. 3NF)
- la denormalizzazione si effettua applicando l'operatore relazionale di join

Per approfondimenti e altri materiali:

<http://www.analisi-disegno.com>