

NORMALIZZAZIONE

Dipendenze funzionali e
Normalizzazione per basi di dati
relazionali



Fino ad ora...

- Abbiamo ipotizzato che gli attributi vengano raggruppati per formare uno schema di relazione usando il buon senso del progettista di basi di dati o traducendo uno schema di base di dati da un modello dei dati concettuale (come ER).
- Ma abbiamo bisogno di misurare formalmente il perché un raggruppamento di attributi in uno schema di relazione possa essere meglio di un altro.
- Obiettivo: valutare la qualità della progettazione di schemi relazionali

Approcci di progettazione

- **Bottom-up:** considera come punto di partenza le associazioni di base tra singoli attributi e le usa per costruire schemi di relazioni (non molto diffuso).
- **Top-down:** inizia con un certo numero di raggruppamenti di attributi a formare relazioni, raggruppamenti di attributi che sussistono come tali naturalmente nel mondo reale, ad esempio una fattura, un form o un report. Le relazioni sono poi analizzate individualmente e collettivamente, portando a decomposizioni successive.

Obiettivi impliciti

- La progettazione di una base di dati produce un insieme di relazioni.
- Obiettivi impliciti dell'attività di progettazione:
 1. **La conservazione dell'informazione**
L'informazione è difficile da quantificare quindi per *conservazione* intendiamo il mantenimento di tutti i concetti espressi mediante il modello concettuale (o ER), inclusi tipi di attributi, tipi di entità e tipi di associazioni.
 2. **La minimizzazione della ridondanza**
Significa minimizzare la memorizzazione ridondante della stessa informazione, riducendo così la necessità di effettuare molteplici aggiornamenti al fine di mantenere la consistenza tra le diverse copie della medesima informazione.

Linee guida informali per la progettazione di schemi di relazione

- 4 misure informali di qualità per la progettazione di uno schema di relazione che possono essere usate per determinare la qualità della progettazione:
 1. assicurarsi che la semantica degli attributi sia chiara nello schema
 2. ridurre i valori ridondanti nelle tuple
 3. ridurre il numero dei valori nulli nelle tuple
 4. rendere impossibile generare tuple spurie
- da queste possiamo inferire 4 linee guida

LG1: semplice è bello!

Si progetti uno schema di relazione in modo tale che sia semplice spiegarne il significato.

Non si uniscano attributi provenienti da più tipi di entità e tipi di relazione in un'unica relazione.

Intuitivamente: se uno schema di relazione corrisponde a un solo tipo di entità o a un solo tipo di relazione, risulta semplice spiegarne il significato; in caso contrario, se una relazione corrisponde a un insieme di entità e relazioni ne nascerà un'ambiguità semantica e quindi non potrà essere spiegata con facilità

IMPIEGATO

c.e.

<u>NOME_I</u>	<u>SSN</u>	DATA_N	INDIRIZZO	NUMERO_D
---------------	------------	--------	-----------	----------

c.p.

DIPARTIMENTO

c.e.

<u>NOME_D</u>	<u>NUMERO_D</u>	SSN_DIR_DIP
---------------	-----------------	-------------

c.p.

SEDI_DIP

c.e.

<u>NUMERO_D</u>	<u>SEDE_D</u>
-----------------	---------------

c.p.

PROGETTO

c.e.

<u>NOME_P</u>	<u>NUMERO_P</u>	SEDE_P	NUM_D
---------------	-----------------	--------	-------

c.p.

LAVORA_SU

c.e.

c.e.

<u>SSN</u>	<u>NUMERO_P</u>	ORE
------------	-----------------	-----

c.p.

IMPIEGATO

<u>NOME_I</u>	<u>SSN</u>	DATA_N	INDIRIZZO	NUMERO_D
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5
Zelaya,Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4
Narayan,Remesh K.	666884444	1962-09-15	975 Fire Oak,Humble,TX	5
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1

DIPARTIMENTO

<u>NOME_D</u>	<u>NUMERO_D</u>	SSN_DIR_DIP
Ricerca	5	333445555
Amministrazione	4	987654321
Sede centrale	1	888665555

SEDI_DIP

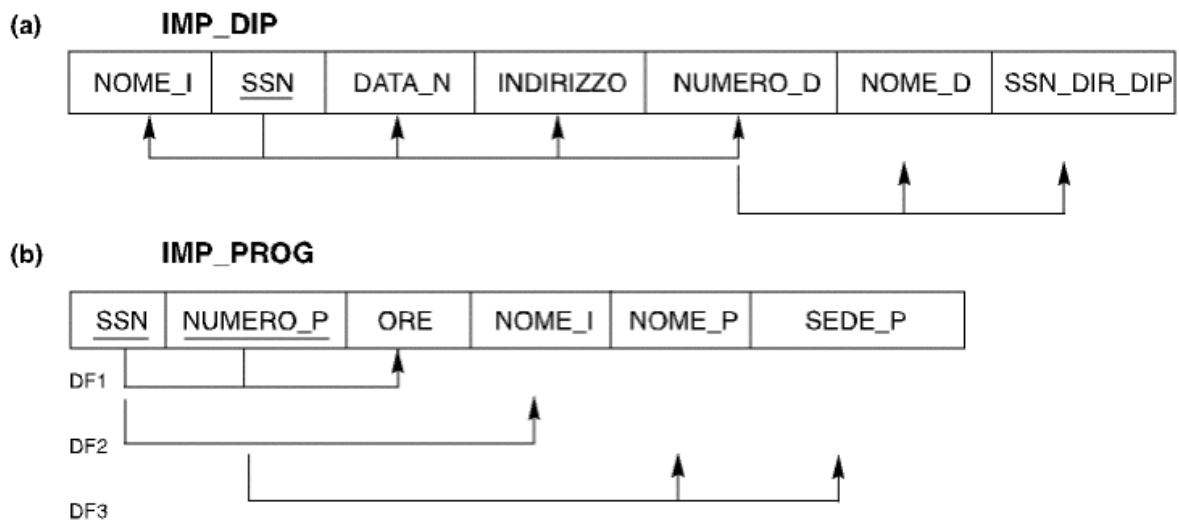
<u>NUMERO_D</u>	<u>SEDE_D</u>
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

LAVORA_SU

<u>SSN</u>	<u>NUMERO_P</u>	ORE
123456789	1	32,5
123456789	2	7,5
666884444	3	40,0
453453453	1	20,0
453453453	2	20,0
333445555	2	10,0
333445555	3	10,0
333445555	10	10,0
333445555	20	10,0
999887777	30	30,0
999887777	10	10,0
987987987	10	35,0
987987987	30	5,0
987654321	30	20,0
987654321	20	15,0
888665555	20	NULL

PROGETTO

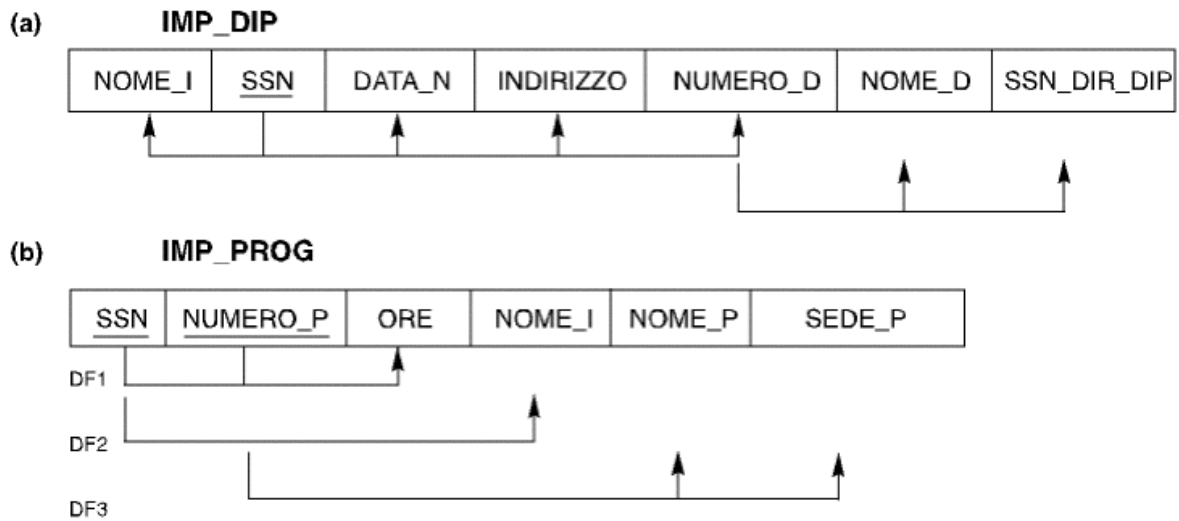
<u>NOME_P</u>	<u>NUMERO_P</u>	SEDE_P	NUM_D
ProdottoX	1	Bellaire	5
ProdottoY	2	Sugarland	5
ProdottoZ	3	Houston	5
Informatizzazione	10	Stafford	4
Riorganizzazione	20	Houston	1
Nuove opportunità	30	Stafford	4



LG2: no alle anomalie

Si progettino schemi di relazione in modo che nelle relazioni non siano presenti anomalie di inserimento, cancellazione o modifica.

Se sono presenti anomalie, le si rilevi chiaramente e ci si assicuri che i programmi che aggiornano la base di dati operino correttamente.



IMP_DIP					Ridondanza	
NOME_I	SSN	DATA_N	INDIRIZZO	NUMERO_D	NOME_D	SSN_DIR_DIP
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Ricerca	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Ricerca	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4	Amministratozione	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Amministratozione	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Ricerca	333445555
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Ricerca	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Amministratozione	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Sede centrale	888665555

IMP_PROG			Ridondanza	Ridondanza	
SSN	NUMERO_P	ORE	NOME_I	NOME_P	SEDE_P
123456789	1	32,5	Smith,John B.	ProdottoX	Bellaire
123456789	2	7,5	Smith,John B.	ProdottoY	Sugarland
666884444	3	40,0	Narayan,Ramesh K.	ProdottoZ	Houston
453453453	1	20,0	English,Joyce A.	ProdottoX	Bellaire
453453453	2	20,0	English,Joyce A.	ProdottoY	Sugarland
333445555	2	10,0	Wong,Franklin T.	ProdottoY	Sugarland
333445555	3	10,0	Wong,Franklin T.	ProdottoZ	Houston
333445555	10	10,0	Wong,Franklin T.	Informatizzazione	Stafford
333445555	20	10,0	Wong,Franklin T.	Riorganizzazione	Houston
999887777	30	30,0	Zelaya,Alicia J.	Nuove opportunità	Stafford
999887777	10	10,0	Zelaya,Alicia J.	Informatizzazione	Stafford
987987987	10	35,0	Jabbar,Ahmad V.	Informatizzazione	Stafford
987987987	30	5,0	Jabbar,Ahmad V.	Nuove opportunità	Stafford
987654321	30	20,0	Wallace,Jennifer S.	Nuove opportunità	Stafford
987654321	20	15,0	Wallace,Jennifer S.	Riorganizzazione	Houston
888665555	20	NULL	Borg,James E.	Riorganizzazione	Houston

LG3: evitare frequenti valori nulli

Per quanto possibile, si eviti di porre in una relazione di base attributi i cui valori possono essere frequentemente nulli.

Se i valori nulli sono inevitabili, ci si assicuri che si presentino solo in casi eccezionale e che non riguardino la maggioranza di tuple nelle relazioni.

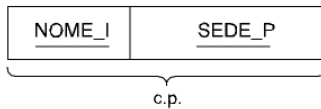
LG4: attenzione alle tuple spurie

Si progettino schemi di relazione in modo tale che possano essere riuniti, tramite JOIN, con condizioni di uguaglianza su attributi che siano coppie (chiave primaria, chiave esterna) correlate in modo appropriato tali da garantire che non vengano generate tuple spurie.

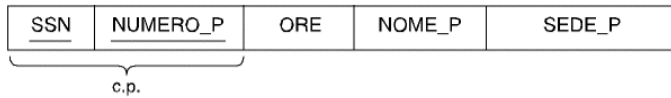
Si evitino relazioni che contengono attributi corrispondenti che non siano combinazioni "chiave esterna e chiave primaria" perché l'esecuzione di operazioni di join su questi attributi può generare tuple spurie

(a)

IMP_SEDI



IMP_PROG1



NOME_I	SEDE_P
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
Wong, Franklin T.	Houston
Wong, Franklin T.	Stafford

Zelaya, Alicia J.	Stafford
Jabbar, Ahmad V.	Stafford
Wallace, Jennifer S.	Stafford
Wallace, Jennifer S.	Houston
Borg, James E.	Houston

IMP_PROG1

SSN	NUMERO_P	ORE	NOME_P	SEDE_P
123456789	1	32,5	Prodotto X	Bellaire
123456789	2	7,5	Prodotto Y	Sugarland
666884444	3	40,0	Prodotto Z	Houston
453453453	1	20,0	Prodotto X	Bellaire
453453453	2	20,0	Prodotto Y	Sugarland
333445555	2	10,0	Prodotto Y	Sugarland
333445555	2	10,0	Prodotto Y	Houston

SSN	NUMERO_P	ORE	NOME_P	SEDE_P	NOME_I
123456789	1	32,5	ProdottoX	Bellaire	Smith, John B.
* 123456789	1	32,5	ProdottoX	Bellaire	English, Joyce A.
123456789	2	7,5	ProdottoY	Sugarland	Smith, John B.
* 123456789	2	7,5	ProdottoY	Sugarland	English, Joyce A.
* 123456789	2	7,5	ProdottoY	Sugarland	Wong, Franklin T.
666884444	3	40,0	ProdottoZ	Houston	Narayan, Ramesh K.
* 666884444	3	40,0	ProdottoZ	Houston	Wong, Franklin T.
* 453453453	1	20,0	ProdottoX	Bellaire	Smith, John B.
453453453	1	20,0	ProdottoX	Bellaire	English, Joyce A.
* 453453453	2	20,0	ProdottoY	Sugarland	Smith, John B.
453453453	2	20,0	ProdottoY	Sugarland	English, Joyce A.
* 453453453	2	20,0	ProdottoY	Sugarland	Wong, Franklin T.
* 333445555	2	10,0	ProdottoY	Sugarland	Smith, John B.
* 333445555	2	10,0	ProdottoY	Sugarland	English, Joyce A.
333445555	2	10,0	ProdottoY	Sugarland	Wong, Franklin T.
* 333445555	3	10,0	ProdottoZ	Houston	Narayan, Ramesh K.
333445555	3	10,0	ProdottoZ	Houston	Wong, Franklin T.
333445555	10	10,0	Informatizzazione	Stafford	Wong, Franklin T.
* 333445555	20	10,0	Riorganizzazione	Houston	Narayan, Ramesh K.
333445555	20	10,0	Riorganizzazione	Houston	Wong, Franklin T.

⋮

Verso valutazione più formalizzata

- Dipendenze funzionali
- Forme normali

Dipendenza funzionale (1)

- Si supponga che uno schema di base di dati relazionale abbia n attributi $A_1, A_2, A_3, \dots, A_n$.
- Si pensi all'intera base di dati come se fosse descritta da un solo schema di relazione $R = \{A_1, A_2, A_3, \dots, A_n\}$.
- **Una dipendenza funzionale è un vincolo tra due insiemi di attributi della base di dati.**
- Una **dipendenza funzionale** (DF), indicata con $X \rightarrow Y$, tra due insiemi di attributi X e Y che siano sottoinsiemi di R specifica un **vincolo** sulle tuple che possono formare uno stato di relazione r di R .
- Il vincolo è che, per ogni coppia di tuple t_1 e t_2 in r per le quali è $t_1[X] = t_2[X]$, si deve avere anche $t_1[Y] = t_2[Y]$.

Dipendenza funzionale (2)

Si dice che:

- I valori della componente Y di una tupla in r **sono determinati da** i valori della componente X

oppure

- I valori della componente X di una tupla in r **determinano univocamente (o funzionalmente)** i valori della componente Y

Si dice anche che:

- C'è una **dipendenza funzionale** da X a Y .

Ancora:

- X determina funzionalmente Y in uno schema di relazione R , se e solo se, ogni volta che due tuple di $r(R)$ concordano sul loro valore di X , esse devono necessariamente concordare anche sul loro valore Y .

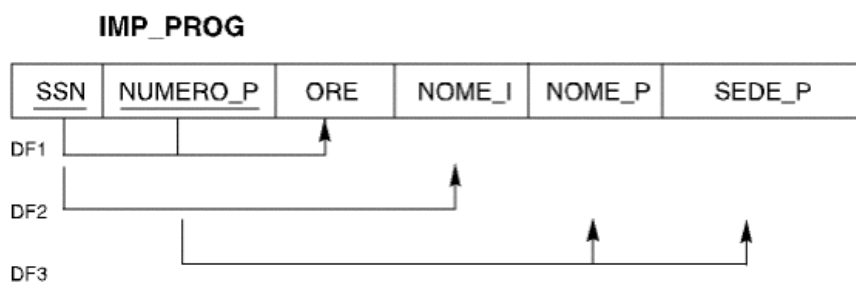
Dipendenza funzionale: nota 1

- Se un vincolo su R stabilisce che non ci possa essere più di una tupla con un dato valore per X in una generica istanza di relazione $r(R)$ (cioè X è una chiave candidata di R), ciò implica che $X \rightarrow Y$ per qualsiasi sottoinsieme Y di attributi di R .
- Questo perché il vincolo di chiave implica che nessuna coppia di tuple in un qualsiasi stato valido $r(R)$ avrà lo stesso valore di X .
- Se X è una chiave candidata di R , allora $X \rightarrow R$.

Dipendenza funzionale: nota 2

- Se $X \rightarrow Y$ in R , ciò non dice se in R $Y \rightarrow X$ oppure no.

Dipendenza funzionale: esempio



- DF2: $SSN \rightarrow NOME_I$
Il valore del numero di previdenza sociale di un impiegato determina univocamente il nome dell'impiegato.
- DF3: $NUMERO_P \rightarrow \{NOME_P, SEDE_P\}$
Il valore del numero di un progetto determina univocamente il nome del progetto e la sua sede.
- DF1: $\{SSN, NUMERO_P\} \rightarrow ORE$
Una combinazione dei valori di SSN e NUMERO_P determina univocamente il numero di ore

Ancora

- Una DF è una proprietà della semantica o del significato degli attributi.
- Una DF è una proprietà dello schema di relazione R , non di un particolare stato valido di relazione r di R .
- Una DF non può essere dedotta a partire da una data estensione di relazione r , ma deve essere definita esplicitamente da qualcuno che conosce la semantica degli attributi di R .

Si può e non si può

- Chiunque può affermare che una certa DF **può** esistere se sussiste in una specifica estensione.
- Nessuno può garantire tale esistenza fintanto che il significato degli attributi non è definito in modo chiaro.
- E' possibile dire che una certa DF **non** sussiste quando esistono due tuple che violano tale DF.

INSEGNA		
DOCENTE	INSEGNAMENTO	TESTO
Smith	Strutture dati	Bartram
Smith	Gestione dati	Martin
Hall	Compilatori	Hoffman
Brown	Strutture dati	Horowitz

Forme normali basate su chiavi primarie

Supponiamo che per ogni relazione:

- Venga dato un insieme di DF
- Che ogni relazione abbia una chiave primaria designata

Queste informazioni guidano **il processo di normalizzazione** per la progettazione di schemi relazionali.

Passo 1: schema ER e sua traduzione

Passo 2: valutazione della qualità delle relazioni ed eventuale decomposizione per ottenere **forme normali di livello superiore**

Normalizzazione delle relazioni (1)

- Il processo di normalizzazione (proposto da Codd) sottopone uno schema di relazione a una serie di test per *certificare* se soddisfa una data **forma normale**.
- Esistono:
 - Prima forma normale (1NF)
 - Seconda forma normale (2NF)
 - Terza forma normale (3NF)
 - Forma normale di Boyce e Codd (BCNF)
 - (4NF e 5NF)

Normalizzazione delle relazioni (2)

- La **normalizzazione dei dati** può essere considerata come un processo di analisi degli schemi forniti, basato sulle loro dipendenze funzionali e chiavi primarie, per raggiungere le proprietà desiderate di
 1. Minimizzazione della ridondanza
 2. Minimizzazione delle anomalie di inserimento/cancellazione/modifica
- Schemi di relazione inadeguati, che non soddisfano certe condizioni vengono decomposti in schemi di relazione più piccoli che superano i test e pertanto possiedono le proprietà desiderate.

Definizioni di chiavi e attributi che partecipano alle chiavi (1)

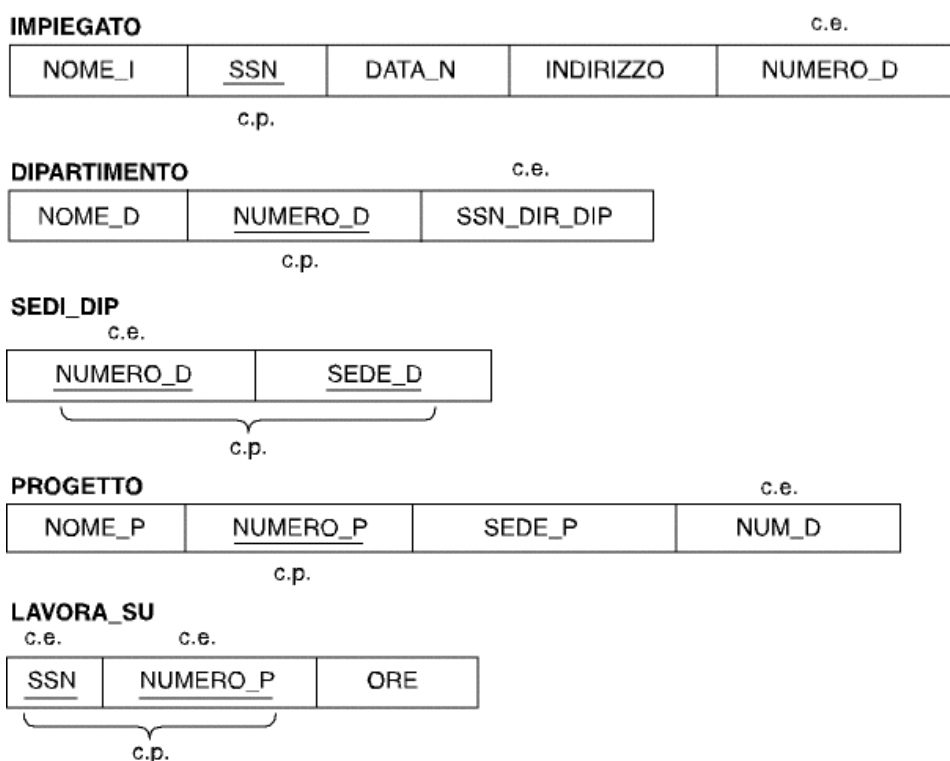
Avevamo detto che:

- Una **superchiave** di uno schema di relazione $R = \{A_1, A_2, A_3, \dots, A_n\}$ è un insieme di attributi $S \subseteq R$ con la proprietà che nessuna coppia di tuple t_1 e t_2 in un generico stato valido di relazione r di R avrà $t_1[S] = t_2[S]$
- Una **chiave** K è una superchiave con la proprietà aggiuntiva che la rimozione di un qualsiasi attributo da K fa sì che K cessi di essere superchiave.
- Cioè una chiave è una superchiave minimale.

Definizioni di chiavi e attributi che partecipano alle chiavi (2)

- Se c'è più di una chiave, ognuna di esse è detta **chiave candidata**.
- Una delle chiavi è nominata **chiave primaria** (le altre sono secondarie).
- Un attributo di uno schema di relazione R è detto **attributo primo** di R se esso è membro di una qualche chiave candidata di R .
- Un attributo è detto **non-primo** se non è membro di alcuna chiave candidata

Esempio AZIENDA



Esempio AZIENDA

IMPIEGATO

NOME_I	SSN	DATA_N	INDIRIZZO	NUMERO_D
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5
Zelaya,Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4
Narayan,Remesh K.	666884444	1962-09-15	975 Fire Oak,Humble,TX	5
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1

DIPARTIMENTO

NOME_D	NUMERO_D	SSN_DIR_DIP
Ricerca	5	333445555
Amministrazione	4	987654321
Sede centrale	1	888665555

SEDI_DIP

NUMERO_D	SEDE_D
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

LAVORA_SU

SSN	NUMERO_P	ORE
123456789	1	32,5
123456789	2	7,5
666884444	3	40,0
453453453	1	20,0
453453453	2	20,0
333445555	2	10,0
333445555	3	10,0
333445555	10	10,0
333445555	20	10,0
999887777	30	30,0
999887777	10	10,0
987987987	10	35,0
987987987	30	5,0
987654321	30	20,0
987654321	20	15,0
888665555	20	NULL

PROGETTO

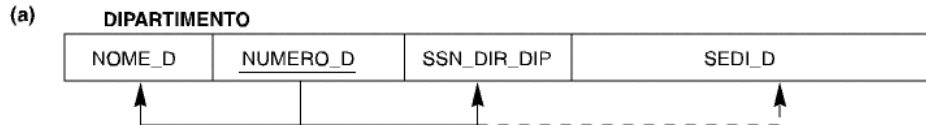
NOME_P	NUMERO_P	SEDE_P	NUM_D
ProdottoX	1	Bellaire	5
ProdottoY	2	Sugarland	5
ProdottoZ	3	Houston	5
Informatizzazione	10	Stafford	4
Riorganizzazione	20	Houston	1
Nuove opportunità	30	Stafford	4

Prima forma normale (1NF)

- Richiede che il dominio di un attributo comprenda solo valori atomici (semplici, indivisibili) e che il valore di qualsiasi attributo in una tupla sia un valore singolo del dominio.
- 1NF è già parte integrante della definizione formale di relazione nel modello relazionale.

Esempio di 1NF (attributi multivalore)

- A partire dalla tabella DIPARTIMENTO con aggiunto SEDI_D



(b)

DIPARTIMENTO			
NOME_D	<u>NUMERO_D</u>	SSN_DIR_DIP	SEDI_D
Ricerca	5	333445555	{Bellaire, Sugarland, Houston}
Amministrazione	4	987654321	{Stafford}
Sede centrale	1	888665555	{Houston}

(c)

DIPARTIMENTO			
NOME_D	<u>NUMERO_D</u>	SSN_DIR_DIP	<u>SEDE_D</u>
Ricerca	5	333445555	Bellaire
Ricerca	5	333445555	Sugarland
Ricerca	5	333445555	Houston
Amministrazione	4	987654321	Stafford
Sede centrale	1	888665555	Houston

Tecnica 1 per raggiungere la 1NF

- Si rimuove l'attributo SEDI_D che viola la 1NF e lo si pone in una relazione separata SEDI_DIP insieme con la chiave primaria NUMERO_D di DIPARTIMENTO.
- La chiave primaria di questa relazione è la combinazione {NUMERO_D, SEDE_D}
- In SEDI_DIP esiste una tupla distinta per ogni sede di un dipartimento
- Questo decompone la relazione in due relazioni che sono in 1NF

Tecnica 2 per raggiungere la 1NF

- Si espande la chiave in modo tale che ci sia una tupla separata per ogni sede di un dipartimento
- La chiave primaria di questa relazione è la combinazione {NUMERO_D, SEDE_D}
- Questa soluzione introduce ridondanza nella relazione

Tecnica 3 per raggiungere la 1NF

- Se è noto il numero massimo di valori dell'attributo (ad esempio se è noto che per un dipartimento possono esistere al più 3 sedi) si sostituisce SEDE_D con tre attributi atomici: SEDE_D1, SEDE_D2, SEDE_D3.
- Introduce valori nulli
- Le interrogazioni diventano più difficili

Esempio di 1NF (attributi multivalore composti)

(a) IMP_PROG

SSN	NOME_I	PROGG	
		NUMERO_P	ORE

(b) IMP_PROG

SSN	NOME_I	NUMERO_P	ORE
123456789	Smith,John B.	1	32,5
		2	7,5
666884444	Narayan,Ramesh K.	3	40,0
453453453	English,Joyce A.	1	20,0
		2	20,0
333445555	Wong,Franklin T.	2	10,0
		3	10,0
		10	10,0
		20	10,0
999887777	Zelaya,Alicia J.	30	30,0
		10	10,0
987987987	Jabbar,Ahmad V.	10	35,0
		30	5,0
987654321	Wallace,Jennifer S.	30	20,0
		20	15,0
888665555	Borg,James E.	20	NULL

(c) IMP_PROG1

SSN	NOME_I
-----	--------

IMP_PROG2

SSN	NUMERO_P	ORE
-----	----------	-----

Tecnica per raggiungere la 1NF

Si noti che:

- SSN è la chiave primaria della relazione IMP_PROG
- NUMERO_P è la chiave primaria parziale della relazione nidificata (cioè dentro ogni tupla NUMERO_P è primaria)

Tecnica:

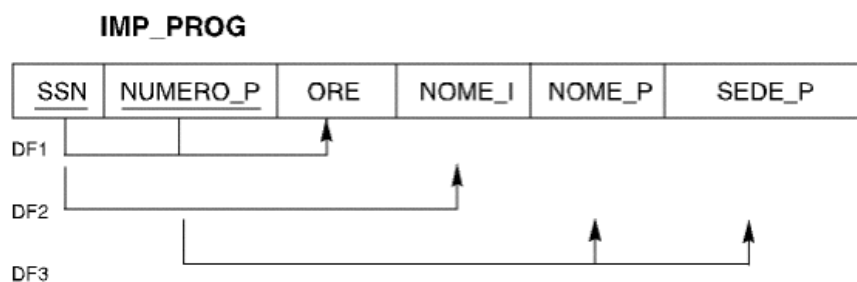
- Si spostano gli attributi della relazione nidificata in una nuova relazione
- Si propaga la chiave primaria al suo interno
- La chiave primaria della nuova relazione è la combinazione della chiave parziale e della chiave primaria della relazione originaria

DF complete e parziali

- Una DF $X \rightarrow Y$ è una **dipendenza funzionale completa** (DFC) se la rimozione di qualsiasi attributo A da X comporta che la DF non sussista più;
 - cioè per ogni attributo $A \in X$, $(X - \{A\})$ NON determina funzionalmente Y .
- Una DF $X \rightarrow Y$ è una **dipendenza funzionale parziale** (DFP) se si possono rimuovere da X certi attributi $A \in X$ e la dipendenza continua a sussistere;
 - cioè per qualche $A \in X$, $(X - \{A\}) \rightarrow Y$.

Esempio

- DF1: $\{\text{SSN}, \text{NUMERO_P}\} \rightarrow \text{ORE}$ è DF **completa**
 - $\text{SSN} \rightarrow \text{ORE}$ non sussiste
 - $\text{NUMERO_P} \rightarrow \text{ORE}$ non sussiste
- $\{\text{SSN}, \text{NUMERO_P}\} \rightarrow \text{NOME_I}$ è DF **parziale**
 - $\text{SSN} \rightarrow \text{NOME_I}$ sussiste

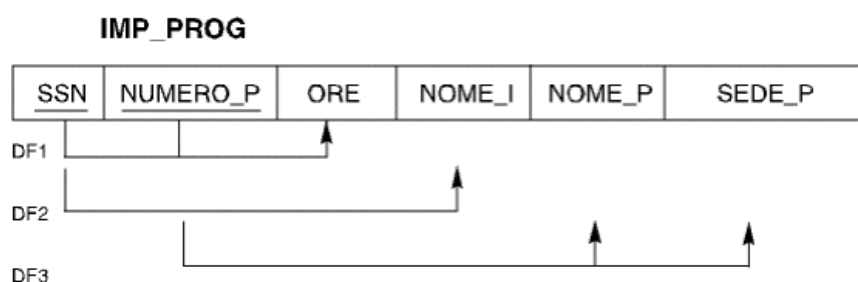


Seconda forma normale (2NF)

- Uno schema relazionale R è in 2NF se ogni attributo non-primario A di R dipende funzionalmente in modo completo dalla chiave primaria di R .

Esempio

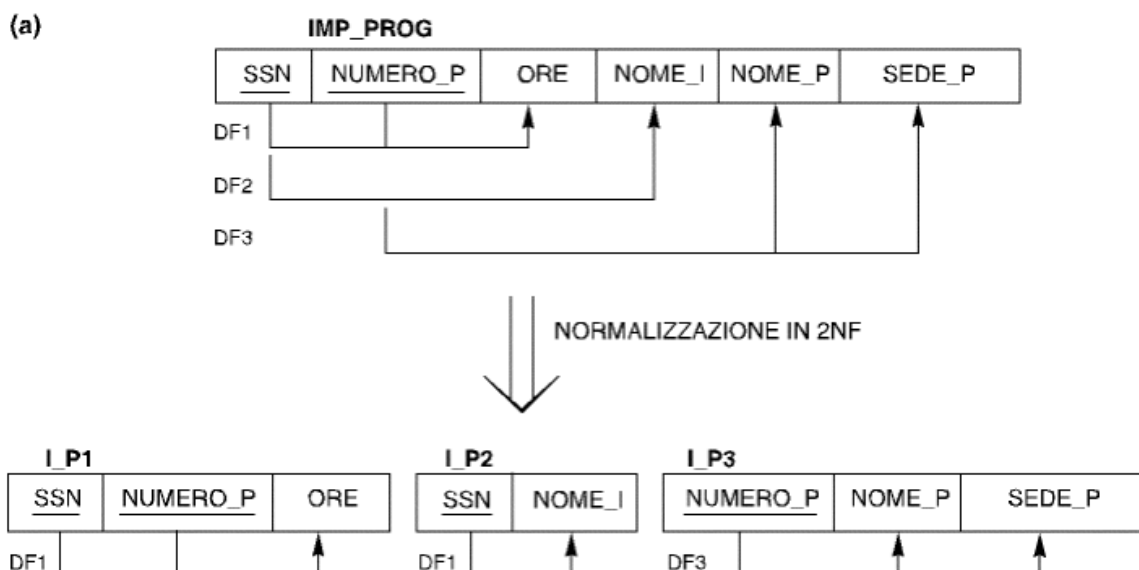
- Questa relazione è in 1NF ma non in 2NF.
- L'attributo non-primario $NOME_I$ viola la 2NF a causa di DF2
- Gli attributi non-primari $NOME_P$ e $SEDE_P$ violano la 2NF a causa di DF3
- $NOME_I$, $NOME_P$ e $SEDE_P$ sono parzialmente dipendenti dalla chiave primaria



Tecnica per raggiungere 2NF

- Si decompone in un certo numero di relazioni 2NF nelle quali gli attributi non-primi sono associati solo alla parte della chiave primaria da cui sono funzionalmente dipendenti in modo completo.

Esempio

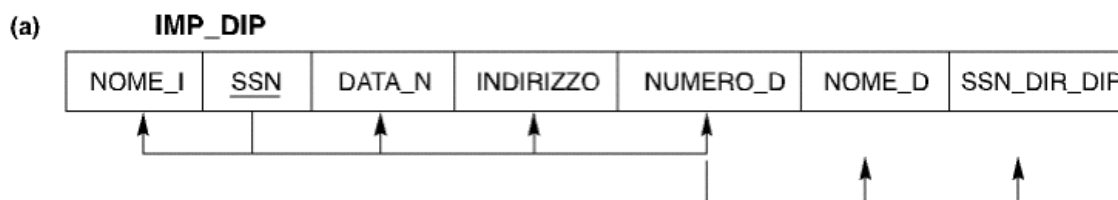


DF transitiva

- Una DF $X \rightarrow Y$ è una **dipendenza funzionale transitiva** (DFT) se esiste un insieme di attributi Z , che non è né una chiave candidata né un sottoinsieme di una chiave di R , per cui valgono contemporaneamente $X \rightarrow Z$ e $Z \rightarrow Y$.

Esempio

- La dipendenza $SSN \rightarrow SSN_DIR_DIP$ è transitiva attraverso $NUMERO_D$ perché
 - sussiste $SSN \rightarrow NUMERO_D$
 E
 - sussiste $NUMERO_D \rightarrow SSN_DIR_DIP$
 E
 - $NUMERO_D$ non è una chiave di per sé, né un sottoinsieme della chiave di IMP_DIP .

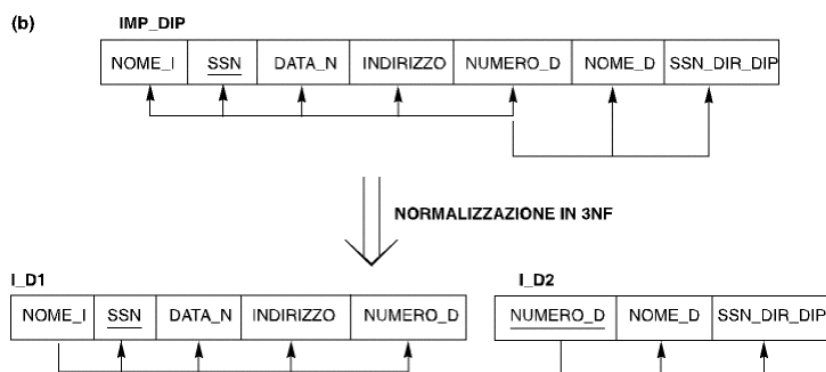


Terza forma normale (3NF)

- Uno schema di relazione R è in 3NF se soddisfa la 2NF e nessun attributo non-primario di R dipende in modo transitivo dalla chiave primaria.
- Una DF in cui la parte sinistra non è una chiave è una DF “problematica”.

Esempio

- Si decompone IMP_DIP nei due schemi relazionali 3NF.



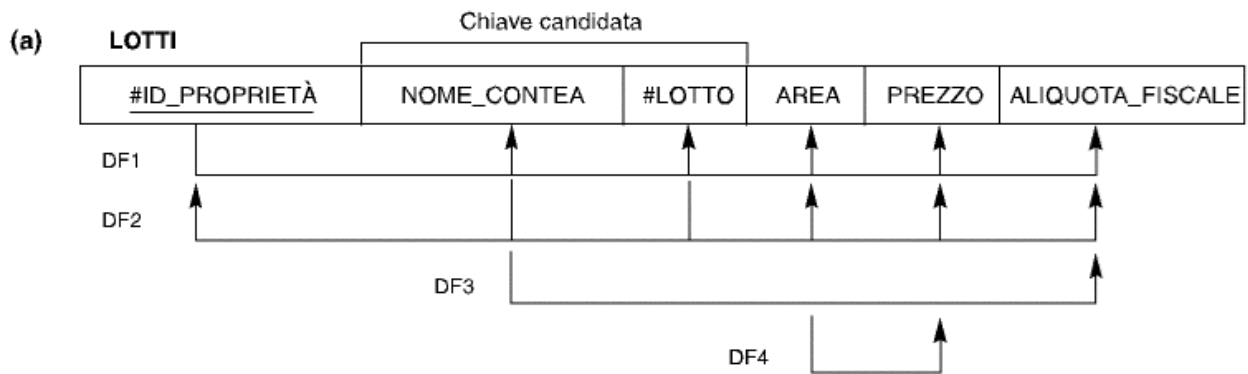
Sintesi delle forme normali basate su chiavi primarie

Forma normale	Descrizione	Rimedio (normalizzazione)
Prima (1NF)	La relazione non deve contenere attributi multivalore o relazioni nidificate.	Definire una nuova relazione per ciascun attributo multivalore o relazione nidificata.
Seconda (2NF)	Per le relazioni in cui la chiave primaria è costituita da più attributi, gli attributi non-chiave non devono dipendere funzionalmente solo da una parte della chiave primaria.	Decomporre la relazione e definire una nuova relazione per ogni chiave parziale e corrispondenti attributi dipendenti. Accertarsi di mantenere una relazione che contenga la chiave primaria iniziale e tutti gli attributi funzionalmente dipendenti da essa in modo completo.
Terza (3NF)	La relazione non deve contenere alcun attributo non-chiave funzionalmente dipendente da un altro attributo non-chiave (o da un insieme di attributi non-chiave). Cioè, non deve esistere alcuna dipendenza transitiva tra attributi non-chiave e la chiave primaria.	Decomporre la relazione e definire una nuova relazione che contenga l'attributo (o gli attributi) non-chiave che determinano funzionalmente un altro attributo (o altri attributi) non-chiave.

Definizione generale di 2NF

- Uno schema di relazione R è in 2NF se ogni attributo non-primario A di R non è parzialmente dipendente da alcuna chiave di R .
- Oppure...
- Uno schema di relazione R è in 2NF se ogni attributo non-primario A di R è funzionalmente dipendente in modo completo da ogni chiave di R .

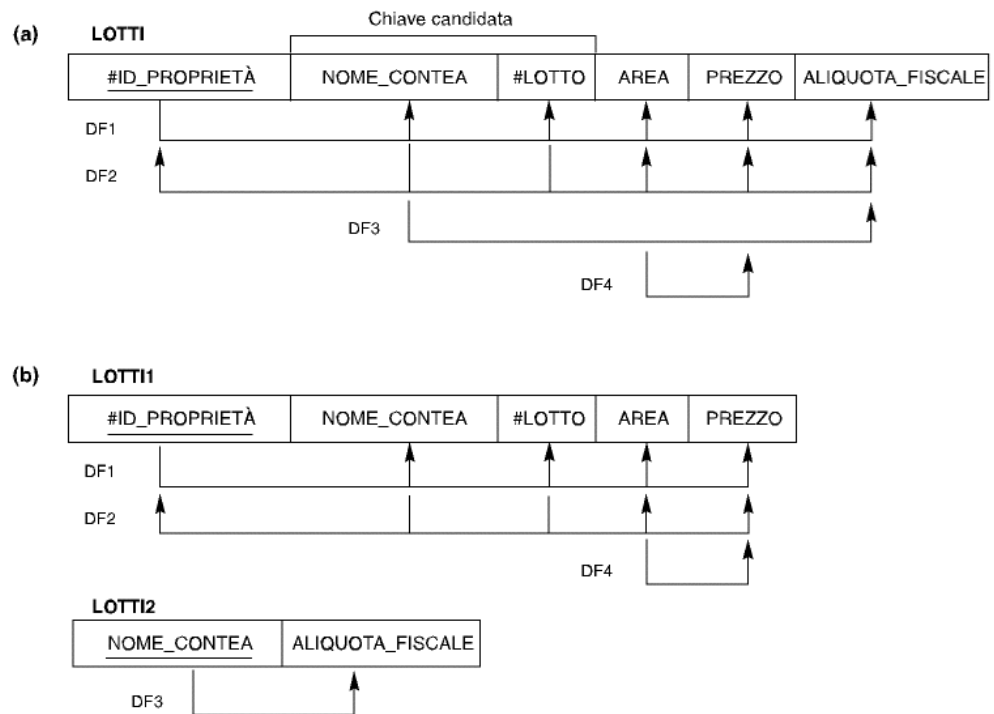
Esempio



Esempio

- Ci sono due chiavi candidate: #ID_PROPRIETÀ e {NOME_CONTEA, #LOTTO}
- Ad esempio: i numeri di #LOTTO sono univoci solo nella stessa contea
- Si sa che sussistono DF1 e DF2
- Si supponga che esistano anche DF3 e DF4
- DF3 dice che l'aliquota è fissa per una data contea (a prescindere dal lotto)
- DF4 dice che il prezzo di un lotto è determinato dalla sua area (a prescindere dalla contea)
- Lo schema viola 2NF perché ALIQUOTA_FISCALE dipende in modo parziale dalla chiave candidata {NOME_CONTEA, #LOTTO}
- Si deve decomporre la relazione
- DF4 non viola la 2NF

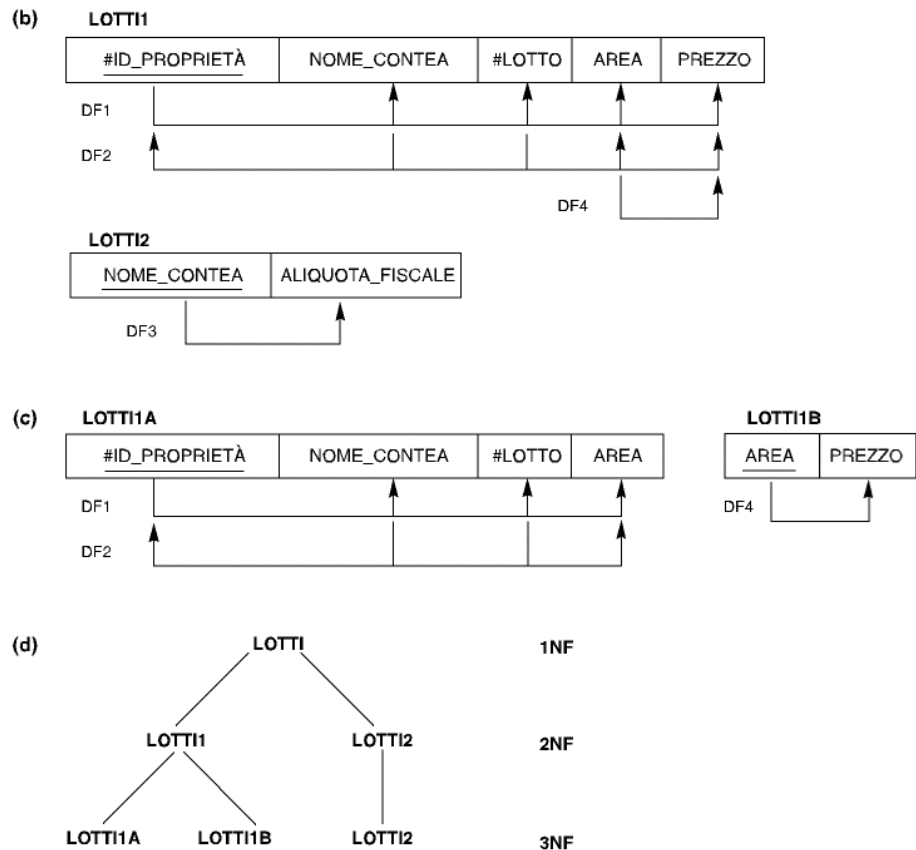
Esempio



Definizione generale di 3NF

- Uno schema di relazione R è in 3NF se, ogni volta che sussiste in R una dipendenza funzionale *non-banale* $X \rightarrow A$, o
 - X è una superchiave di R
 - oppure
 - A è un attributo primo di R .
- Oppure...
- Uno schema di relazione R è in 3NF se ogni attributo non-primo di R soddisfa entrambe le condizioni:
 - È funzionalmente dipendente in modo completo da ogni chiave di R
 - Non è dipendente in modo transitivo da alcuna chiave di R

Esempio



Esempio

- LOTTI2 è in 3NF
- DF4 in LOTTI1 viola 3NF perché in questa relazione AREA non è superchiave e PREZZO non è un attributo primo.
- Soluzione: si decompone la relazione.

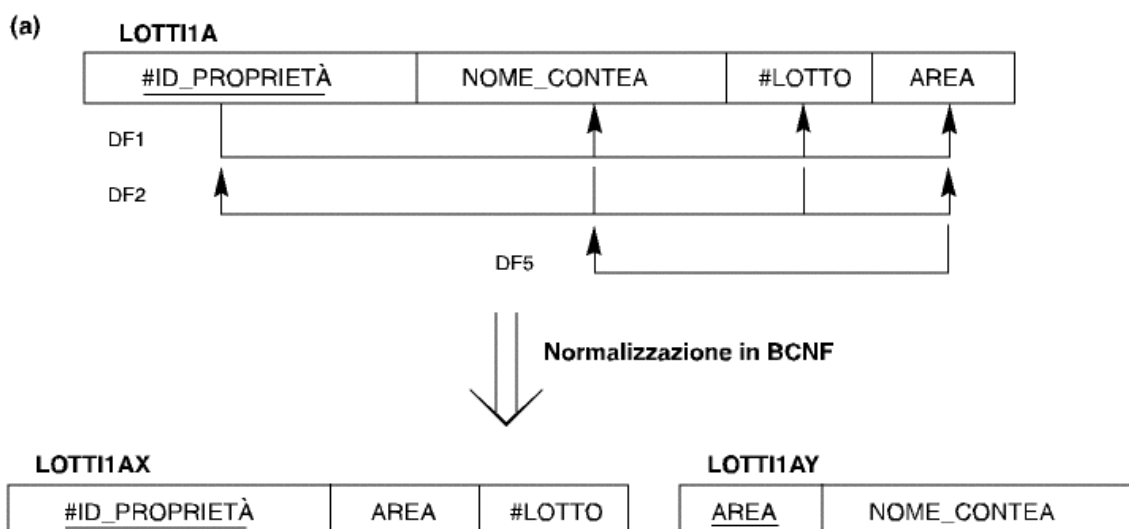
Note:

- LOTTI1 viola 3NF perché PREZZO è transitivamente dipendente da ognuna delle chiavi candidate di LOTTI1 tramite l'attributo non-primo AREA

Forma normale di Boyce e Codd

- Uno schema di relazione R è in BCNF se, ogni volta che sussiste in R una dipendenza funzionale *non-banale* $X \rightarrow A$, X è una superchiave di R .
- Ogni relazione in BCNF è anche in 3NF ma non è detto il contrario.

Esempio

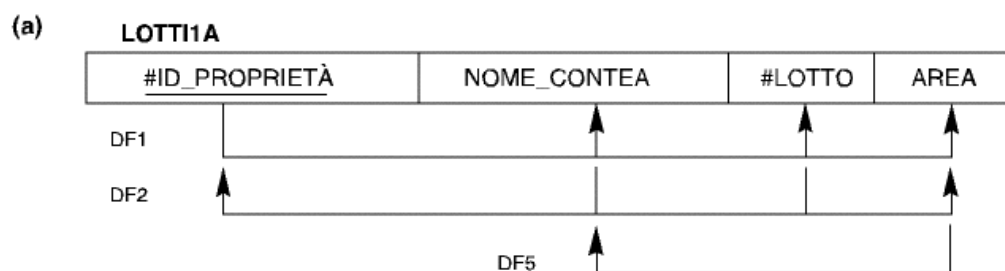


Esempio

- Si supponga di avere migliaia di lotti nella relazione, ma che i lotti vengano solo da due contee: Dekalb e Fulton.
- Si ipotizzi inoltre che le dimensioni possibili dei lotti siano valori presi da un insieme differente per ogni contea: per Dekalb (1, 3, 5, 7, 9) acri e per Fulton (2, 4, 8, 10) acri.
- Si supponga quindi che esista una DF aggiuntiva (DF5) AREA → NOME_CONTEA (rimarrebbe in 3NF perché NOME_CONTEA è attributo primo).

Esempio

- DF5 viola BCNF in LOTTI1A perché AREA non è superchiave di LOTTI1A.
- Devo scomporre la relazione in base a DF5.



Esercizio 1

Si consideri la seguente relazione per libri pubblicati:
LIBRO (Titolo_Libro, Nome_Autore, Tipo_Libro,
Prezzo_di_Listino, Affiliazione_Autore, Editore)

Si supponga che esistano le seguenti dipendenze:
Titolo_Libro → Editore, Tipo_Libro
Tipo_Libro → Prezzo_di_Listino
Nome_Autore → Affiliazione_Autore

- In che forma normale è la relazione? Si motivi la risposta.
- Si esegua la normalizzazione finché non sia più possibile decomporre le relazioni. Si enuncino le motivazioni che giustificano ciascuna decomposizione.

Soluzione (1)

- La chiave della relazione è Titolo_Libro, Nome_Autore. La relazione è in 1NF ma non in 2NF poiché non ci sono attributi che hanno una dipendenza funzionale completa rispetto alla chiave. Inoltre, la relazione non è neppure in 3NF.

Soluzione (2)

Decomposizione in 2NF:

- Libro0(Titolo_Libro, Nome_Autore)
- Libro1(Titolo_Libro, Editore, Tipo_Libro, Prezzo_di_Listino)
- Libro2(Nome_Autore, Affiliazione_Autore)
- Questa decomposizione elimina le dipendenze parziali.

Soluzione (3)

Decomposizione in 3NF:

- Libro0(Titolo_Libro, Nome_Autore)
- Libro1-1(Titolo_Libro, Editore, Tipo_Libro)
- Libro1-2(Tipo_Libro, Prezzo_di_Listino)
- Libro2(Nome_Autore, Affiliazione_Autore)
- Questa decomposizione elimina la dipendenza transitiva di Prezzo_di_Listino.

Esercizio 2

- Si considerino le seguenti relazioni per un'applicazione di base di dati relativa alla gestione degli ordini presso la società ABC:

ORDINE

(#O, DataO, #Cliente, Importo_Totale)

ARTICOLO_ORDINE

(#O, #A, Qtà_ordinata, Prezzo_Totale, %Sconto)

Esercizio 2

- Si supponga che ogni articolo presenti uno sconto diverso. **Prezzo_Totale** fa riferimento a un singolo articolo, **DataO** è la data in cui l'ordine è stato emesso e **Importo_Totale** indica l'ammontare complessivo dell'ordine.
- Se si applica un'operazione di join naturale sulle relazioni ARTICOLO_ORDINE e ORDINE di questa base di dati, come si presenta lo schema di relazione risultante? Quale sarà la sua chiave?
- Si mostrino le DF in questa relazione risultante. Tale relazione è in 2NF? È in 3NF? Perché o perché no? (Si enunci esplicitamente ogni ipotesi fatta.)

Soluzione

- Date le relazioni

ORDINE

(#O, DataO, #Cliente, Importo_Totale)

ARTICOLO_ORDINE

(#O, #A, Qtà_Ordinata, Prezzo_Totale, %Sconto)

lo schema di ORDINE * ARTICOLO_ORDINE sarà

R1(#O, #A, DataO, Cliente, Importo_Totale, Qtà_ordinata, Prezzo_Totale, %Sconto)

- La chiave di questo schema di relazione è #O,#A.

Soluzione

- Le dipendenze funzionali sono

{#O, #A} → Qtà_Ordinata

{#O, #A} → Prezzo_Totale

{#O, #A} → %Sconto

#O → DataO

#O → Cliente

#O → Importo_Totale

- Lo schema non è in 2NF poiché gli attributi DataO, Cliente e Importo_Totale sono parzialmente dipendenti dalla chiave primaria {#O, #A}. Ne consegue che lo schema non è neppure in 3NF, visto che la 2NF è un requisito per la 3NF.

Esercizio 3

Si consideri la seguente relazione:

VENDITA_AUTOMOBILE

{Auto, Data_vendita, Venditore, %Commissione, Ammontare_sconto}

- Si supponga che un'automobile possa essere venduta da più venditori e perciò che {Auto, Venditore} sia la chiave primaria.
- Ulteriori dipendenze sono:
 - {Data_vendita] → {Ammontare_sconto}
 - {Venditore} → {%Commissione}
- Sulla base della chiave primaria data, questa relazione è in 1NF, 2NF o 3NF? Perché? Come si potrebbe normalizzare in passi successivi?

Esercizio 4

Si consideri la relazione R che ha attributi che contengono prospetti di insegnamenti e moduli di un'università:

R = {Cod_insegnamento, Cod_modulo, Dip_offerente, Ore_credits, Livello_insegnamento, SSN_Assistente, Semestre, Anno, Ore_giorni, Num_aula, Num_di_studenti}.

Si supponga che in R sussistano le seguenti dipendenze funzionali:

- {Cod_insegnamento} → {Dip_offerente, Ore_credits, Livello_insegnamento}
- {Cod_insegnamento, Cod_modulo, Semestre, Anno} → {Ore_giorni, Num_aula, Num_di_studenti, SSN_Assistente}
- {Num_aula, Ore_giorni, Semestre, Anno} → {SSN_Assistente, Cod_insegnamento, Cod_modulo}

Si cerchi di determinare quali insiemi di attributi formano le chiavi di R. Come normalizzare?

Esercizio 5

A	B	C	#TUPLA
10	b1	c1	1
10	b2	c2	2
11	b4	c1	3
12	b3	c4	4
13	b1	c1	5
14	b3	c4	6

1. $A \rightarrow B$

2. $B \rightarrow C$

3. $C \rightarrow B$

4. $B \rightarrow A$

5. $C \rightarrow A$