



# Qualita' della Progettazione di Schemi Relazionali

Raffaella Gentilini



# Outline

## Introduzione

### Linee Guida Informali di Progettazione BD Relazionali

- Semantica degli Attributi

- Riduzione Valori Ridondanti nelle Tuple

- Riduzione Numero Valori Nulli nelle Tuple

- Impossibilita' di Generare Tuple Spurie

### Sommario e Overview sulla Prossima Puntata



# Qualita' della Progettazione di Schemi Relazionali

## Obbiettivi

- A partire da questa lezione, svilupperemo opportune misure di adeguatezza o "bonta'" per valutare la qualita' della progettazione di schemi relazionali
- Possiamo esaminare la qualita' di progettazione di una BD relazionale rispetto a due livelli:
  1. livello logico (o concettuale)
  2. livello d'implementazione (o fisico)
- Vedremo prima alcune misure informali di qualita' di schemi relazionali
- Introduciamo quindi opportuni concetti formali ed una teoria per la definizione precisa della "buona/cattiva qualita'" di un singolo schema di relazione.



# Linee Guida Informali

## Misure di Qualita' Informali

Analizzeremo quattro **misure informali** di **qualita'** per la progettazione di uno schema di relazione.

1. **semantica** degli **attributi**
2. **riduzione** dei **valori ridondanti** nelle tuple
3. **riduzione** del numero di **valori nulli** nelle tuple
4. **impossibilita'** di generare tuple spurie



## Semantica degli Attributi

- Il progetto di uno **schema di relazione** sara' tanto migliore quanto piu' facile e' spiegare la semntica della relazione.
- La facilita' con cui puo' essere spiegato il **significato degli attributi** di una realzione e' una **misura informale di buona progettazione**

### Linea Guida 1

- Si progetti schema di relazione in modo che sia **semplice spiegarne il significato**
- **Non si uniscano attributi provenienti da piu' tipi di entita' o tipi di associazione** in un'unica relazione (cio' genera solitamente ambiguita' semantiche).



## Example

**EMPLOYEE** F.K.

Ename	<u>Ssn</u>	Bdate	Address	Dnumber
-------	------------	-------	---------	---------

P.K.

**DEPARTMENT** F.K.

Dname	<u>Dnumber</u>	Dmgr_ssn
-------	----------------	----------

P.K.

**DEPT\_LOCATIONS** F.K.

<u>Dnumber</u>	<u>Dlocation</u>
----------------	------------------

P.K.

**PROJECT** F.K.

Pname	<u>Pnumber</u>	Plocation	Dnum
-------	----------------	-----------	------

P.K.

**WORKS\_ON**

<u>Ssn</u>	<u>Pnumber</u>	Hours
------------	----------------	-------

P.K.

Schemi di relazione che possono essere considerati buoni rispetto al fatto di possedere una semantica chiara.



## Example (Stato BD)

### EMPLOYEE

Ename	Ssn	Bdate	Address	Dnumber
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, Ramesh 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

### DEPARTMENT

Dname	Dnumber	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

### DEPT\_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

### WORKS\_ON

Ssn	Pnumber	Hours
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

### PROJECT

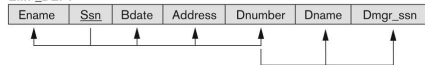
Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4



## Example

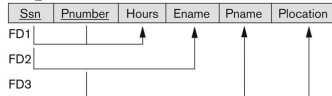
(a)

**EMP\_DEPT**



(b)

**EMP\_PROJ**



Schemi di relazione mediocri dato che violano la linea di guida 1, mischiando attributi provenienti da entita' distinte del mondo reale.





## Example (Stato BD)

Redundancy

EMP\_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1969-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1982-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Redundancy

Redundancy

EMP\_PROJ

Ssn	Pnumber	Hours	Ename	Pname	Plocation
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston
999887777	30	30.0	Zelaya, Alicia J.	Newbenefits	Stafford
999887777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford
987987987	10	35.0	Jabbar, Ahmad V.	Computerization	Stafford
987987987	30	5.0	Jabbar, Ahmad V.	Newbenefits	Stafford
987654321	30	20.0	Wallace, Jennifer S.	Newbenefits	Stafford
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston
888665555	20	Null	Borg, James E.	Reorganization	Houston



## Riduzione Valori Ridondanti nelle Tuple

Oltre ad introdurre ambiguità semantiche, un approccio basato sul raggruppamento di attributi legati ad entità/associazioni distinte del mondo reale è causa dei seguenti problemi:

- Informazione memorizzata in modo ridondante
- Spreco di Spazio
- Anomalie di Aggiornamento
  1. anomalie di inserimento
  2. anomalie di cancellazione
  3. anomalie di modifica



## Example (Ridondanze ed Anomalie)

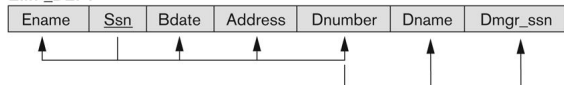
**Figure 10.3**

Two relation schemas suffering from update anomalies.

(a) EMP\_DEPT and  
(b) EMP\_PROJ.

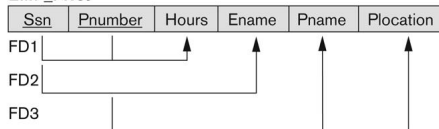
(a)

**EMP\_DEPT**



(b)

**EMP\_PROJ**





## Example (Ridondanze ed Anomalie - continua)

					Redundancy	
EMP_DEPT						
Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

- I valori degli attributi che riguardano uno specifico dipartimento sono ripetuti per ogni impiegato che lavora nel dipartimento



## Example (Anomalie di Inserimento)

- Per inserire una nuova tupla di impiegato afferente al dipartimento 5 in `IMP_DIP` bisogna inserire valori degli attributi del dipartimento 5 consistenti con i valori del dipartimento 5 in altre tuple di `IMP_DIP`. In caso contrario, si genera un'anomalia di inserimento.
- Come inserire in `IMP_DIP` nuovo dipartimento senza impiegati?
  - ....il solo modo per farlo e' quello di porre valori nulli negli attributi relativi ad impiegato.
  - **Problema 1.** SSN e' la chiave primaria di `IMP_DIP` (si suppone che ogni tupla rappresenti in effetti un'entita' di `IMPIEGATO` e non di `DIPARTIMENTO`)
  - **Problema 2.** Quando viene assegnato il primo impiegato al dipartimento in questione, non si ha piu' bisogno della tupla con valori nulli.



## Example (Anomalie di Cancellazione)

Se si cancella una tupla da IMP\_DIP l'ultimo impiegato che lavora per un dipartimento, l'informazione riguardante quel dipartimento non e' piu' presente nella BD.

## Example (Anomalie di Modifica)

Se si cambia il valore di uno degli attributi di un particolare dipartimento in IMP\_DIP, ad esempio il direttore del dipartimento 5, occorre aggiornare le tuple di tutti gli impiegati afferenti a tale dipartimento.



# Informazioni Ridondanti nelle Tuple e Anomalie di Aggiornamento

## Linea Guida 2

- Si progettino gli schemi di relazione di BD 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 BD operino correttamente.



# Riduzione Valori Nulli nelle Tuple

## Valori Nulli

- spreco spazio
- problemi di comprensione del significato degli attributi e di specificazione delle operazioni di JOIN e delle operazioni di aggregazione.
  - valori nulli possono avere piu' interpretazioni (attributo non pertinente/sconosciuto/noto ma assente)
  - join interno ed esterno danno risultati diversi quando sono coinvolti valori nulli
- Criteri contrapposti: Utilizzo efficiente spazio vs volonta' evitare operazioni di join.

## Example

Se solo il 10% degli impiegati ha un ufficio personale, non c'e' ragione di includere un attributo NUMERO\_UFFICIO nella relazione IMPIEGATO.





## Riduzione Valori Nulli nelle Tuple

### Linea Guida 3

- Per quanto possibile, si eviti di porre in una relazione attributi i cui valori possono essere frequentemente nulli.
- Se valori nulli sono inevitabili, ci si assicuri che si presentino solo in casi eccezionali e non riguardino la maggioranza delle tuple.



## Example

(a)

**EMP\_LOCS**

Ename	Plocation
-------	-----------

P.K.

**EMP\_PROJ1**

Ssn	Pnumber	Hours	Pname	Plocation
-----	---------	-------	-------	-----------

P.K.

(b)

**EMP\_LOCS**

Ename	Plocation
Smith, John B.	Bellaire
Smith, John B.	Surgarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Surgarland
Wong, Franklin T.	Surgarland
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

**EMP\_PROJ1**

Ssn	Pnumber	Hours	Pname	Plocation
123456789	1	32.5	ProductX	Bellaire
123456789	2	7.5	ProductY	Sugarland
666884444	3	40.0	ProductZ	Houston
453453453	1	20.0	ProductX	Bellaire
453453453	2	20.0	ProductY	Sugarland
333445555	2	10.0	ProductY	Sugarland
333445555	3	10.0	ProductZ	Houston
333445555	10	10.0	Computerization	Stafford
333445555	20	10.0	Reorganization	Houston
999887777	30	30.0	Newbenefits	Stafford
999887777	10	10.0	Computerization	Stafford
987987987	10	35.0	Computerization	Stafford
987987987	30	5.0	Newbenefits	Stafford
987654321	30	20.0	Newbenefits	Stafford
987654321	30	15.0	Reorganization	Stafford

**Figure 10.5**

Particularly poor design for the EMP\_PROJ relation of Figure 10.3(b).  
 (a) The two relation schemas EMP\_LOCS and EMP\_PROJ1. (b) The result of projecting the extension of EMP\_PROJ from Figure 10.4 onto the relations EMP\_LOCS and EMP\_PROJ1.



## Tuple Spurie

- La decomposizione di IMP\_PROG in IMP\_SEDI e IMP\_PROG1 produce effetti indesiderati perche', quando si uniscono le due relazioni usando il JOIN NATURALE, non si ottiene l'informazione originale corretta.
- Infatti, un JOIN NATURALE su IMP\_SEDI e IMP\_PROG1 il risultato produce molte piu' tuple rispetto all'originaria popolazione di tuple in IMP\_PROG.
- le tuple aggiuntive, non presenti in IMP\_PROG, sono dette **tuple spurie**.
- Cio' si verifica perche' SEDE\_P, l'attributo che collega IMP\_SEDI e IMP\_PROG1 non e' ne' una chiave primaria ne' una chiave esterna in IMP\_SEDI o IMP\_PROG1.



## Example

Ssn	Pnumber	Hours	Pname	Plocation	Ename
123456789	1	32.5	ProductX	Bellaire	Smith, John B.
* 123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith, John B.
* 123456789	2	7.5	ProductY	Sugarland	English, Joyce A.
* 123456789	2	7.5	ProductY	Sugarland	Wong, Franklin T.
666884444	3	40.0	ProductZ	Houston	Narayan, Ramesh K.
* 666884444	3	40.0	ProductZ	Houston	Wong, Franklin T.
* 453453453	1	20.0	ProductX	Bellaire	Smith, John B.
453453453	1	20.0	ProductX	Bellaire	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Smith, John B.
453453453	2	20.0	ProductY	Sugarland	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Wong, Franklin T.
* 333445555	2	10.0	ProductY	Sugarland	Smith, John B.
* 333445555	2	10.0	ProductY	Sugarland	English, Joyce A.
333445555	2	10.0	ProductY	Sugarland	Wong, Franklin T.
* 333445555	3	10.0	ProductZ	Houston	Narayan, Ramesh K.
333445555	3	10.0	ProductZ	Houston	Wong, Franklin T.
333445555	10	10.0	Computerization	Stafford	Wong, Franklin T.
* 333445555	20	10.0	Reorganization	Houston	Narayan, Ramesh K.
333445555	20	10.0	Reorganization	Houston	Wong, Franklin T.

\*  
\*  
\*

**Figure 10.6**

Result of applying NATURAL JOIN to the tuples above the dotted lines in EMP\_PROJ1 and EMP\_LOCS of Figure 10.5. Generated spurious tuples are marked by asterisks.



# Impossibilita' di Generare Tuple Spurie

## Linea Guida 4

- Si progettino schemi di relazione in modo tale che possano essere riuniti, tramite JOIN, con condizioni di uguaglianza su attributi che sono chiavi primarie o chiavi esterne, in modo da garantire che non vengano generate tuple spurie.
- Si evitino relazioni che contengono attributi corrispondenti che non siano combinazioni (chiave esterna, chiave primaria), perche' l'esecuzione di join su questi attributi puo' generare tuple spurie.



## Sommario

Abbiamo esaminato informalmente situazioni che portano a schemi di relazione problematici , proponendo 4 linee di guida informali per un buon progetto relazionale. I problemi evidenziati sono:

- anomalie che richiedono un lavoro aggiuntivo durante l'inserimento o la modifica di una relazione, e che determinano una perdita accidentale di informazione durante una cancellazione da una relazione.
- perdita di spazio di memoria a causa di valori nulli e difficoltà' di eseguire operazioni di aggregazione e di join a causa della presenza di valori nulli.
- generazione di dati non validi e spuri durante i join su relazioni di base unite in modo improprio.



## Nelle Prossime Puntate ...

Nel seguito verranno presentati **concetti formali** ed una **teoria** che possono essere usati **per definire piu' precisamente i concetti di buona/cattiva qualita' degli schemi relazionali.**

- Introduremo il **concetto di dipendenza funzionale.**
- si specificheranno **le tre forme normali e la forma normale di Boyce e Codd (BCNF)** per schemi di relazione.
- verranno definite ulteriori forme normali basate su altri tipi di dipendenze dette **dipendenze multivalore** e **dipendenze di join.**