

# Fondamenti di Teoria delle Basi di Dati

**Riccardo Torlone**

Parte 8: Teoria delle dipendenze

# Vincoli di integrità

---

- Esistono istanze di basi di dati che, pur sintatticamente corrette, non rappresentano informazioni possibili per l'applicazione di interesse

# Una base di dati "scorretta"

<b>Esami</b>	<b>Studente</b>	<b>Voto</b>	<b>Lode</b>	<b>Corso</b>
	276545	32		01
	276545	30	e lode	02
	787643	27	e lode	03
	739430	24		04

<b>Studenti</b>	<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>
	276545	Rossi	Mario
	787643	Neri	Piero
	787643	Bianchi	Luca

# Vincolo di integrità

---

- Proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione
- Un vincolo è una funzione booleana (un **predicato**)
  - associa ad ogni istanza il valore **vero** o **falso**

# Tipi di vincoli

---

- vincoli **intrarelazionali**
  - vincoli su valori (o di **dominio**)
  - vincoli di **tupla**
- vincoli **interrelazionali**

# Vincoli di tupla

---

- Esprimono condizioni sui valori di ciascuna tupla, indipendentemente dalle altre tuple
- Caso particolare:
  - Vincoli di dominio: coinvolgono un solo attributo

# Sintassi ed esempi

---

- Una possibile sintassi:
  - espressione booleana di atomi che confrontano valori di attributo o espressioni aritmetiche su di essi

$(\text{Voto} \geq 18) \text{ AND } (\text{Voto} \leq 30)$

$(\text{Voto} = 30) \text{ OR NOT } (\text{Lode} = \text{"e lode"})$

# Vincoli di tupla, altro esempio

---

<b>Stipendi</b>	<b>Impiegato</b>	<b>Lordo</b>	<b>Ritenute</b>	<b>Netto</b>
	<b>Rossi</b>	<b>55.000</b>	<b>12.500</b>	<b>42.500</b>
	<b>Neri</b>	<b>45.000</b>	<b>10.000</b>	<b>35.000</b>
	<b>Bruni</b>	<b>47.000</b>	<b>11.000</b>	<b>36.000</b>

$$\text{Lordo} = (\text{Ritenute} + \text{Netto})$$



# Identificazione delle tuple

<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Corso</b>	<b>Nascita</b>
<b>27655</b>	<b>Rossi</b>	<b>Mario</b>	<b>Ing Inf</b>	<b>5/12/78</b>
<b>78763</b>	<b>Rossi</b>	<b>Mario</b>	<b>Ing Inf</b>	<b>3/11/76</b>
<b>65432</b>	<b>Neri</b>	<b>Piero</b>	<b>Ing Mecc</b>	<b>10/7/79</b>
<b>87654</b>	<b>Neri</b>	<b>Mario</b>	<b>Ing Inf</b>	<b>3/11/76</b>
<b>67653</b>	<b>Rossi</b>	<b>Piero</b>	<b>Ing Mecc</b>	<b>5/12/78</b>

- non ci sono due tuple con lo stesso valore sull'attributo Matricola
- non ci sono due tuple uguali su tutti e tre gli attributi Cognome, Nome e Data di Nascita

# Chiave

---

- Insieme di attributi che identificano le tuple di una relazione
- Formalmente:
  - un insieme  $K$  di attributi è **superchiave** per  $r$  se  $r$  non contiene due tuple distinte  $t_1$  e  $t_2$  con  $t_1[K] = t_2[K]$
  - $K$  è **chiave** per  $r$  se è una superchiave minimale per  $r$  (cioè non contiene un'altra superchiave)

# Una chiave

<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Corso</b>	<b>Nascita</b>
<b>27655</b>	<b>Rossi</b>	<b>Mario</b>	<b>Ing Inf</b>	<b>5/12/78</b>
<b>78763</b>	<b>Rossi</b>	<b>Mario</b>	<b>Ing Inf</b>	<b>3/11/76</b>
<b>65432</b>	<b>Neri</b>	<b>Piero</b>	<b>Ing Mecc</b>	<b>10/7/79</b>
<b>87654</b>	<b>Neri</b>	<b>Mario</b>	<b>Ing Inf</b>	<b>3/11/76</b>
<b>67653</b>	<b>Rossi</b>	<b>Piero</b>	<b>Ing Mecc</b>	<b>5/12/78</b>

- Matricola, Cognome è una superchiave
- Matricola è una chiave:
  - è superchiave
  - contiene un solo attributo e quindi è minimale

# Vincoli, schemi e istanze

---

- i vincoli corrispondono a proprietà del mondo reale modellato dalla base di dati
- interessano a livello di schema (con riferimento cioè a tutte le istanze)
- ad uno schema associamo un insieme di vincoli e consideriamo **corrette** (valide, ammissibili) le istanze che soddisfano tutti i vincoli
- un'istanza può soddisfare altri vincoli (“per caso”)

# Dipendenza funzionale

---

- relazione  $r$  su  $R(X)$
- due sottoinsiemi non vuoti  $Y$  e  $Z$  di  $X$
- esiste in  $r$  una **dipendenza funzionale** (FD) da  $Y$  a  $Z$  (in simboli  $Y \rightarrow Z$ ) se, per ogni coppia di ennuple  $t_1$  e  $t_2$  di  $r$  con gli stessi valori su  $Y$ , risulta che  $t_1$  e  $t_2$  hanno gli stessi valori anche su  $Z$

# Esempio

<u>Impiegato</u>	Stipendio	<u>Progetto</u>	Bilancio	Funzione
Rossi	20	Marte	2	tecnico
Verdi	35	Giove	15	progettista
Verdi	35	Venere	15	progettista
Neri	55	Venere	15	direttore
Neri	55	Giove	15	consulente
Neri	55	Marte	2	consulente
Mori	48	Marte	2	direttore
Mori	48	Venere	15	progettista
Bianchi	48	Venere	15	progettista
Bianchi	48	Giove	15	direttore

Impiegato → Stipendio

Progetto → Bilancio

Impiegato Progetto → Funzione

# Integrità referenziale

---

- informazioni in relazioni diverse sono correlate attraverso valori comuni in particolare, valori delle chiavi (primarie)
- le correlazioni debbono essere "coerenti"
- Un vincolo di **integrità referenziale** ("foreign key") fra gli attributi  $X$  di una relazione  $R_1$  e un'altra relazione  $R_2$  impone ai valori su  $X$  in  $R_1$  di comparire come valori della chiave primaria di  $R_2$

# Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

## Vigili

<u>Matricola</u>	Cognome	Nome
3987	Rossi	Luca
3295	Neri	Piero
9345	Neri	Mario
7543	Mori	Gino



# Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

## Auto

<u>Prov</u>	<u>Numero</u>	Cognome	Nome
MI	39548K	Rossi	Mario
TO	E39548	Rossi	Mario
PR	839548	Neri	Luca

# Violazione di vincolo di integrità referenziale

## Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

## Auto

<u>Prov</u>	<u>Numero</u>	Cognome	Nome
MI	E39548	Rossi	Mario
TO	F34268	Rossi	Mario
PR	839548	Neri	Luca

# Azioni compensative

---

- Esempio:
  - Viene eliminata una tupla causando una violazione
- Comportamento “standard”:
  - Rifiuto dell'operazione
- Azioni compensative:
  - Eliminazione in cascata
  - Introduzione di valori nulli

# Eliminazione in cascata

## Impiegati

<u>Matricola</u>	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	XYZ
64521	Verdi	NULL
73032	Bianchi	IDEA

## Progetti

<u>Codice</u>	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
BOH	09/2001	24	150

# Introduzione di valori nulli

## Impiegati

<u>Matricola</u>	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	NULL
64521	Verdi	NULL
73032	Bianchi	IDEA

## Progetti

<u>Codice</u>	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
BOH	09/2001	24	150

# Implicazione di vincoli

- Un insieme di vincoli  $\Gamma$  **implica un vincolo**  $\gamma$  se ogni istanza che soddisfa  $\Gamma$  soddisfa anche  $\gamma$ 
  - L'insieme  $\{ \text{Impiegato} \rightarrow \text{Livello}, \text{Livello} \rightarrow \text{Stipendio} \}$  implica il vincolo  $\text{Impiegato} \rightarrow \text{Stipendio}$
- La chiusura  $\Gamma^+$  di un insieme di vincoli  $\Gamma$  è l'insieme di tutti i vincoli implicati da  $\Gamma$ .
  - $\Gamma^+ = \{ \gamma \mid \Gamma \text{ implica } \gamma \}$
- Proprietà:
  - $\Gamma \subseteq \Gamma^+$
  - $(\Gamma^+)^+ \subseteq \Gamma^+$
  - Se  $\Gamma_1 \subseteq \Gamma_2$  allora  $\Gamma_1^+ \subseteq \Gamma_2^+$
  - $\Gamma_1^+ \cup \Gamma_2^+ \subseteq (\Gamma_1 \cup \Gamma_2)^+$

# Equivalenza e ridondanza di insiemi di vincoli

- $\Gamma_1$  è **equivalente** a  $\Gamma_2$  se  $\Gamma_1^+ = \Gamma_2^+$  (si dice anche che  $\Gamma_1$  è una copertura di  $\Gamma_2$ )
- **Lemma**  $\Gamma_1^+ = \Gamma_2^+$  sse  $\Gamma_1$  implica ogni vincolo di  $\Gamma_2$  e  $\Gamma_2$  implica ogni vincolo di  $\Gamma_1$
- Un **vincolo**  $\gamma$  è **ridondante** in  $\Gamma$  se  $\Gamma^+ = (\Gamma - \{\gamma\})^+$
- Un **insieme di vincoli**  $\Gamma$  è **ridondante** se esiste  $\gamma$  ridondante in  $\Gamma$
- **Lemma**  $\gamma$  è ridondante in  $\Gamma$  sse  $\Gamma - \{\gamma\}$  implica  $\gamma$
- **Problema dell'implicazione**: dati  $\Gamma$  e  $\gamma$ ,  $\Gamma$  implica  $\gamma$ ?

# Esempi con riferimento alle dip. funzionali

---

- $F_1 = \{A \rightarrow B, AB \rightarrow C, A \rightarrow C\}$
- $F_2 = \{A \rightarrow B, AB \rightarrow C\}$
- $F_3 = \{A \rightarrow B, A \rightarrow C\}$
  
- $F_1$  è ridondante
  - perché  $\{A \rightarrow B, AB \rightarrow C\}$  implica  $A \rightarrow C$
- $F_1$  è equivalente a  $F_2$
- $F_2$  è equivalente a  $F_3$



# Approccio assiomatico

- Si definisce un insieme di **regole di inferenza** (proof system) che consentono di generare tutte (**completezza**) e solo (**correttezza**) le dipendenze implicate da un insieme dato.
- Per le FD si può dimostrare che le tre regole:
  - $Y \subseteq X \Rightarrow X \rightarrow Y$  (trivial)
  - $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$  (augmentation)
  - $X \rightarrow Y$  e  $Y \rightarrow Z \Rightarrow X \rightarrow Z$  (transitivity)
- forniscono un insieme corretto e completo Armstrong (1974)
- Questo approccio però non è pratico perché può produrre un insieme esponenziale di dipendenze

# Chiusura di un insieme di attributi

- Approccio alternativo al problema dell'implicaz.
- Siano dati uno schema di relazione  $R(U)$  e un insieme di dipendenze funzionali  $F$  definite sugli attributi in  $U$ . Sia  $X$  un insieme di attributi contenuti in  $U$  (cioè  $X \subseteq U$ );
  - La **chiusura** di  $X$  rispetto a  $F$ , indicata con  $X^+_F$ , è l'insieme degli attributi che dipendono funzionalmente da  $X$  (esplicitamente o implicitamente):
    - $X^+_F = \{ A \mid A \subseteq U \text{ e } F \text{ implica } X \rightarrow A \}$
- **Lemma**  $X \rightarrow Y \in F^+$  sse  $Y \subseteq X^+$
- Quindi, per verificare che  $X \rightarrow Y$  è implicata da  $F$ , basta vedere se  $Y$  è contenuto in  $X^+$

# Algoritmo di chiusura

---

- **Input:** un insieme  $X$  di attributi e un insieme  $F$  di dipendenze
- **Output:** un insieme  $X_p$  di attributi
  - 1) Inizializziamo  $X_p$  con l'insieme di input  $X$
  - 2) Esaminiamo le dipendenze in  $F$ ; se esiste una dipendenza  $Y \rightarrow Z$  con  $Y \subseteq X_p$  e  $Z$  non contenuto in  $X_p$  allora aggiungiamo gli attributi di  $Z$  a  $X_p$ .
  - 3) Ripetiamo il passo 2 fino al momento in cui non vi sono ulteriori attributi che possono essere aggiunti a  $X_p$

# Esempi

---

R(Impiegato, Stipendio, Progetto, Bilancio, Funzione)

Impiegato → Stipendio

Progetto → Bilancio

Impiegato Progetto → Funzione

- $IP^+ = ISPBF$
- $I^+ = IS$
- $P^+ = PB$

# Dimostrazione di correttezza

**Teorema** Sia  $X$  un insieme di attributi ed  $F$  un insieme di dipendenze su attributi in  $X$ . Al termine della esecuzione dell'algoritmo risulta  $X_P = X_F^+$

**Proof** Consideriamo s.p.d.g. solo FD  $X \rightarrow A$ .

L'algoritmo termina: il passo principale viene ripetuto solo se ci sono attributi da aggiungere a  $X_P$ ; poiché il numero di attributi di una relazione è finito, prima o poi si raggiunge un punto in cui non vi sono attributi da aggiungere.

$X_P \subseteq X_F^+$ . Siano  $X_0, X_1, \dots, X_h$  i valori assunti da  $X_P$  durante l'esecuzione dell'algoritmo, con  $X_0 = X$  e  $X_h$  pari al valore finale. La dimostrazione procede per induzione, mostrando che  $X_i \subseteq X_F^+$ , per ogni  $i$ .

# Dimostrazione di correttezza (continua)

Il passo base,  $X_0 = X \subseteq X_F^+$  è immediato, perché la dipendenza  $X \rightarrow A$ , per ogni  $A \in X$ , è banale e quindi sempre soddisfatta e quindi sempre implicata.

Il passo induttivo richiede di mostrare che, se  $X_i \subseteq X_F^+$ , allora  $X_{i+1} \subseteq X_F^+$ , cioè che se  $F$  implica  $X \rightarrow A$ , per ogni  $A \in X_i$  allora  $F$  implica  $X \rightarrow A$ , per ogni  $A \in X_{i+1}$ . Supponiamo che  $r$  soddisfi  $F$  e siano  $t_1$  e  $t_2$  due tuple uguali su  $X$ ; per l'ipotesi induttiva,  $r$  soddisfa  $X \rightarrow A$ , per ogni  $A \in X_i$  e quindi  $t_1$  e  $t_2$  sono uguali su  $X_i$ ; se l'algoritmo viene applicato, allora esiste una dipendenza  $Y \rightarrow A \in F$  con  $Y \subseteq X_i$  e  $X_{i+1} = X_i A$ ; ma se  $r$  soddisfa  $F$ , allora soddisfa  $Y \rightarrow A$  e quindi, poiché  $t_1$  e  $t_2$  sono uguali su  $X_i$  (e quindi su  $Y$ , dato che  $Y \subseteq X_i$ ) allora esse sono uguali su  $A$  e quindi anche su  $X_{i+1} = X_i A$ .

# Dimostrazione di correttezza (continua)

$X_F^+ \subseteq X_P$ . Sia  $A \in X_F^+$ . Allora, per definizione di  $X_F^+$ ,  $F$  implica  $X \rightarrow A$ . Assumiamo per assurdo che  $A \notin X_P$  e consideriamo la seguente relazione:

$$r$$

	$X_P$				$U - X_P$			
	0	0	...	0	0	0	...	0
	0	0	...	0	1	1	...	1

La relazione soddisfa  $F$ : sia infatti  $Z \rightarrow B \in F$ . Se  $Z \not\subseteq X_P$ , le due tuple sono diverse su  $Z$  e quindi la dipendenza è certamente soddisfatta. Se  $Z \subseteq X_P$ , le due tuple sono uguali su  $Z$  e allora l'algoritmo può considerare la dipendenza  $Z \rightarrow B$  e aggiungere anche  $B$  a  $X_P$ ; di conseguenza, le due tuple sono uguali anche su  $B$ . Questa relazione però viola la dipendenza  $X \rightarrow A$ , perché  $X \subseteq X_P$  e  $A \notin X_P$  e quindi  $F$  non implica  $X \rightarrow A$  — una contraddizione.  $\diamond$

# Applicazione dell'algoritmo di chiusura

- L'algoritmo può essere utilizzato per verificare se un insieme di attributi è chiave
- **Lemma** Siano dati uno schema di relazione  $R(U)$  e un insieme di dipendenze funzionali  $F$  definite sugli attributi in  $U$ . Sia  $K$  un insieme di attributi contenuti in  $U$  (cioè  $K \subseteq U$ ):
  - $K$  è superchiave per  $R$  sse  $K^+ = U$
  - $K$  è chiave per  $R$  sse per ogni attributo  $A \in K$  risulta  $(K - \{A\})^+ \subset U$



# Esempio

R(Impiegato, Stipendio, Progetto, Bilancio, Funzione)

Impiegato → Stipendio

Progetto → Bilancio

Impiegato Progetto → Funzione

- $IPB^+ = ISPBF$ 
  - quindi IPB è superchiave per R
- $IP^+ = ISPBF$ 
  - anche IP è superchiave per R
- $I^+ = IS, P^+ = PB$ 
  - IP è anche chiave per R