SAPIENZA
UNIVERSITÀ DI ROMA

Appunti di Basi Dati Modulo I

Colacel Alexandru Andrei

Disclaimer

Le fonti sono le "Hand Notes" del prof. Perelli tradotte in italiano, appunti presi dalle slides della prof. De Marsico ed eventuali e-mail.

Indice

1 Lemma della Chiusura **2**
1.1 Dimostrazione $\Rightarrow$ . . . 2
1.2 Dimostrazione $\Leftarrow$ . . . 2

2 Teorema $F^+ = F^A$ **3**
2.1 Dimostrazione $F^A \subseteq F^+$ . . . 3
2.2 Dimostrazione $F^+ \subseteq F^A$ . . . 5

3 Chiusura di X **6**
3.1 Teorema: L'algoritmo computa X_F^+ . . . 6

4 Lemma: Inclusione delle chiusure **8**
4.1 Dimostrazione . . . 8

5 Chiusura di X in G **9**
5.1 Dimostrazione . . . 9

6 Join senza perdita **11**

7 Assiomi di Armstrong **12**

8 Decomposizione che preserva F **13**
8.1 Prima proprietà . . . 13
8.2 Seconda proprietà . . . 13
8.3 Terza proprietà . . . 13
8.4 Definizione G . . . 13

9 Dimostrazione ρ preserva F **14**

10 Hash **15**

11 B-tree **16**

12 Chiusura di un insieme di attributi **17**
12.1 Algoritmo . . . 17
12.2 Dimostrazione correttezza . . . 17

13 3NF (per dipendenza transitiva) **18**

14 Chiusura di F e primo lemma **19**

15 Isam **20**
15.1 Variante Isam con chiavi indice che hanno valore ultimo record . . . 20

16 Altre definizioni da sapere **21**
16.1 Chiave minimale . . . 21
16.2 Superchiave . . . 21

1 Lemma della Chiusura

Sia R uno schema e sia F un insieme di dipendenze funzionali definite su R. Si ha che:

$$X \rightarrow Y \in F^A \Longleftrightarrow Y \subseteq X^+ \tag{1}$$

1.1 Dimostrazione $\Rightarrow$

Dato $X \rightarrow Y \in F^A$, per la regola della decomposizione, otteniamo:

$$X \rightarrow A \in F^A, \quad \forall A \in Y \tag{2}$$

e quindi, per definizione di X^+, otteniamo che:

$$A \in X^+, \quad \forall A \in Y \tag{3}$$

che significa:

$$Y \subseteq X^+ \tag{4}$$

1.2 Dimostrazione $\Leftarrow$

Dato:

$$Y \subseteq X^+ \tag{5}$$

si ottiene che:

$$X \rightarrow A \in F^A \quad \forall A \in Y \tag{6}$$

che implica, per la regola dell'unione, che:

$$X \rightarrow Y \in F^A \tag{7}$$

2 Teorema $F^+ = F^A$

Dato uno schema R e un insieme F di dipendenze funzionali definite su R, si ha che:

$$F^+ = F^A \tag{8}$$

2.1 Dimostrazione $F^A \subseteq F^+$

Prendiamo $X \to Y \in F^A$, noi dobbiamo provare che $X \to Y \in F^+$ per induzione con n numero di applicazioni degli assiomi di Armstrong.

- **Caso base** (n = 0): se $X \to Y \in F^A$ senza aver applicato alcun assioma di Armstrong, allora l'unica possibilità è che:
$$X \to Y \in F \subseteq F^+ \tag{9}$$
- **Ipotesi induttiva forte:** ogni dipendenza funzionale in F^A ottenuta da F applicando k $\leq$ n assiomi di Armstrong è anche in F^+:
$$X \to Y \in F^A \text{ tramite } k \leq n \text{ assiomi} \Rightarrow X \to Y \in F^+$$
- **Passo induttivo:** è necessario dimostrare che se $X \to Y \in F^A$ dopo aver applicato $n+1$ assiomi di Armstrong, allora $X \to Y \in F^+$.

  È possibile ritrovarsi in uno dei seguenti tre casi:

  1. Se l'$(n+1)$-esimo assioma applicato è l'assioma di **riflessività**, allora l'unica possibilità è che:
  $$X \to Y \in F^A \Leftrightarrow Y \subseteq X \subseteq R \tag{10}$$
  Dunque, poiché, $Y \subseteq X \subseteq R$, per ogni istanza legale di R si ha che:
  $$\forall t_1, t_2 \in r_1, t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y] \tag{11}$$
  da cui ne segue automaticamente che $X \to Y \in F^+$
  2. Se l'$(n+1)$-esimo assioma applicato è l'assioma di **aumento**, allora è obbligatoriamente necessario che:
     - $\exists V, W \subseteq R \,|\, \exists V \to W \in F_A$, ottenuta applicando $j \leq n$ assiomi di Armstrong
     - $\exists Z \subseteq R \,|\, X := VZ, \, Y := WZ$

     Affinché si abbia che:
  $$Z \subseteq R, \, V \to W \Rightarrow VZ \to WZ = X \to Y \in F^A \tag{12}$$
  Siccome per ipotesi induttiva si ha $V \to W \in F^A \Rightarrow V \to W \in F^+$ e siccome $Z \subseteq Z \Rightarrow Z \to Z \in F^+$, si vede facilmente che:
  $$\begin{cases} V \to W \in F^+ \\ Z \to Z \in F^+ \end{cases} \implies \begin{cases} \forall t_1, t_2 \in r, t_1[V] = t_2[V] \implies t_1[W] = t_2[W] \\ \forall t_1, t_2 \in r, t_1[Z] = t_2[Z] \implies t_1[Z] = t_2[Z] \end{cases} \implies$$
  $$\implies \forall t_1, t_2 \in r, t_1[VZ] = t_2[VZ] \implies t_1[WZ] = t_2[WZ] \implies$$
  $$\implies \forall t_1, t_2 \in r, t_1[X] = t_2[X] \implies t_1[Y] = t_2[Y] \implies X \to Y \in F^+$$

3. Se l'$(n+1)$-esimo assioma applicato è l'assioma di **transitività**, allora è obbligatoriamente necessario che $\exists X \to Z, Z \to Y \in F^A$, ottenute con $k \leq n$ assiomi di Armstrong, affinché si abbia che:

$$X \to Z \in F^A \vee Z \to Y \in F^A \Rightarrow X \to Y \in F^A \tag{13}$$

Siccome per ipotesi induttiva $X \to Z \in F^A \Rightarrow X \to Z \in F^+$ e $Z \to Y \in F^A \Rightarrow Z \to Y \in F^+$, si vede facilmente che:

$$\begin{cases} X \to Z \in F^+ \\ Z \to Y \in F^+ \end{cases} \Longrightarrow$$
$$\Longrightarrow \forall t_1, t_2 \in r, t_1[X] = t_2[X] \Longrightarrow t_1[Z] = t_2[Z] \Longrightarrow t_1[Y] = t_2[Y] \Longrightarrow$$
$$\Longrightarrow X \to Y \in F^+$$

2.2 Dimostrazione $F^+ \subseteq F^A$

- Sia $X \subseteq R$ e sia r istanza di $R(X^+, R - X^+)$ tale che

X^+			$R - X^+$		
A_1	$\cdots$	A_i	A_j	$\cdots$	A_n
1	...	1	1	...	1
1	...	1	0	...	0

  dunque tale che $\forall t_1, t_2 \in r$ si ha:
  * $t_1[X^+] = (1, \ldots, 1) = t_2[X^+]$
  * $t_1[R - X^+] = (1, \ldots, 1) \neq (0, \ldots, 0) = t_2[R - X^+]$
- Notiamo che $\forall V, W \subseteq R \mid V \to W \in F$ si ha che:
  * Se $V \cap R - X^+ \neq \emptyset$ (dunque anche se $V \subseteq R - X^+$) allora $t_1[V] \neq t_2[V]$, dunque r soddisfa $V \to W \in F$
  * Se invece $V \subseteq X^+$, per il lemma precedentemente visto si ha che

  $$V \subseteq X^+ \iff X \to V \in F^A$$

  Siccome $V \to W \in F \implies V \to W \in F^A$, per transitività si ha che

  $$X \to V \in F^A \wedge V \to W \in F^A \implies X \to W \in F^A \iff W \subseteq X^+$$

  Dunque, siccome $V, W \subseteq X^+$, in definitiva si ha che

  $$t_1, t_2 \in r, t_1[V] = (1, \ldots, 1) = t_2[V] \wedge t_1[W] = (1, \ldots, 1) = t_2[W]$$

  e quindi r soddisfa ogni $V \to W \in F$
- Siccome in entrambi i casi r soddisfa ogni $V \to W \in F$, allora r è legale.
- A questo punto, una qualsiasi dipendenza $X \to Y \in F^+$ deve essere soddisfatta da qualsiasi istanza legale di R, inclusa r stessa
- Poiché $X \subseteq X^+$, ne segue che la dipendenza non può essere soddisfatta a vuoto poiché $t_1[X] = t_2[X]$. Dunque, l'unica possibilità affinché $X \to A \in F^+$ sia soddisfatta da r è che $Y \subseteq X^+$ in modo che si abbia $t_1[Y] = t_2[Y]$
- A questo punto, per il lemma si ha che $Y \subseteq X^+ \iff X \to Y \in F^A$
- Dunque, siccome $X \to Y \in F^+ \implies X \to Y \in F^A$, concludiamo che $F^A \subseteq F^+$

Nota

Poiché $F^+ = F^A$, per calcolare F^+ ci basta applicare gli assiomi di Armstrong sulle dipendenze in F in modo da trovare F^A.
Tuttavia, calcolare $F^+ = F^A$ richiede tempo esponenziale, quindi $O(2^{nk})$: considerando anche solo l'assioma di riflessività, siccome ogni possibile sottoinsieme di R genera una dipendenza e siccome i sottoinsiemi possibili di R sono $2^{|R|}$, allora ne segue che $|F^+| >> 2^{|R|}$.

3 Chiusura di X

`Input`:

- Relazione R
- Dipendenze Funzionali F
- Insieme $X \subseteq R$

`Output`: X^+

Algorithm 1 Closure Algorithm

$Z \leftarrow X$
$S \leftarrow \{A \mid \exists Y \rightarrow V \in F, A \in V \wedge Y \subseteq Z\}$
while $S \not\subseteq Z$ **do**
  $Z \leftarrow Z \cup S$
  $S \leftarrow \{A \mid \exists Y \rightarrow V \in F, A \in V \wedge Y \subseteq Z\}$
end while
return Z

Tale algoritmo viene eseguito in tempo polinomiale, ossia On^k

3.1 Teorema: L'algoritmo computa X_F^+

Il teorema calcola correttamente la chiusura di un insieme di attributi X rispetto ad un insieme F di dipendenze funzionali.

Dimostrazione:

Denotiamo:

$$Z_0, Z_1, \ldots, Z_i, \ldots \tag{14}$$

$$S_0, S_1, \ldots, S_i, \ldots \tag{15}$$

Indichiamo con Z_0 il valore iniiale di $Z(Z_0 = X)$ e con Z_i e S_i, $i \geq 1$ i valori di Z ed S dopo l'i-esima esecuzione del corpo del ciclo, infatti notiamo che $Z_i \subseteq Z_{i+1}$, per ogni i.

Sia j tale che $S_j \subseteq Z_j$ (cioè, Z_j è l'output di Z quando l'algoritmo termina); proveremo che:

$$A \in Z_j \Leftrightarrow A \in X^+ \tag{16}$$

- Dimostriamo per induzione su i che $Z_f \subseteq X^+$
  - Caso base dell'induzione: i = 0.
    Alla 0-esima iterazione del while (ossia prima di esso) si ha $Z_0 = X$ e $X \subseteq X^+$
  - Ipotesi induttiva: Per ogni $i \in \mathbb{N}$ si ha che $Z_i \subseteq X^+$
  - **Passo induttivo** $(i > 0)$: Dato $A \in Z_{i+1} = Z_i \cup S_i$, si ha che $A \in Z_i \vee A \in S_i$.
    Dunque, si possono verificare due casi:
    * Se $A \in Z_i$, allora per ipotesi $A \in Z_i \subseteq X^+$.
    * Se $A \in S_i$, allora $\exists Y \rightarrow V \in F$ tale che $A \in V \subseteq R, Y \subseteq Z_i$.
      Siccome per ipotesi $Z_i \subseteq X^+$ e $Y \subseteq Z_i$, allora $Y \subseteq Z_i \subseteq X^+ \iff X \rightarrow Y \in F^A$ e $Y \rightarrow V \in F \implies Y \rightarrow V \in F^A$, quindi per transitività si ha che:

$$X \rightarrow Y, Y \rightarrow V \in F^A \implies X \rightarrow V \in F^A \iff V \subseteq X^+. \tag{17}$$

      Dunque,si ha che $A \in V \subseteq X^+$
    * Siccome in entrambi i casi $A \in Z_{i+1} \implies A \in X^+$, allora concludiamo che $Z_{i+1} \subseteq X^+$.

- Dimostriamo ora che $X^+ \subseteq Z_i$:
  - Sia $X \subseteq R$ e sia r istanza di $R(Z_i, R - Z_i)$.

Z_i			$R - Z_i$		
A_1	...	A_i	A_j	...	A_n
1	...	1	1	...	1
1	...	1	0	...	0

    dunque tale che per $t_1, t_2 \in r$ si ha:
    * $t_1[Z_i] = (1, \ldots, 1) = t_2[Z_i]$
    * $t_1[R - Z_i] = (1, \ldots, 1) \neq (0, \ldots, 0) = t_2[R - Z_i]$
  - Notiamo che $\forall V, W \subseteq R$ — $V \to W \in F$ si ha che:
    * Se $V \cap (R - Z_i) \neq \emptyset$ (quindi anche se $V \subseteq R - Z_i$) allora $t1[V] \neq t2[V]$, quindi r soddisfa $V \to W \in F$
    * Se invece $V \subseteq Z_i$, allora $W \subseteq S_f$, poiché per come viene calcolato S_f, si ha che:
      $V \to W \in F$, $V \subseteq Z_i$, $B \in W \subseteq R \Rightarrow B \in S_f \Rightarrow W \subseteq S_f$
      e quindi, siccome $S_f \subseteq Z_i$ è la condizione che termina l'algoritmo, allora $W \subseteq S_f \subseteq Z_i$
    * Siccome $V, W \subseteq Z_i$, in definitiva si ha che
      $t1, t2 \in r$, $t1[V] = (1, \ldots, 1) = t2[V]$ e $t1[W] = (1, \ldots, 1) = t2[W]$,
      e quindi r soddisfa $V \to W \in F$
  - Siccome in entrambi i casi r soddisfa ogni $V \to W \in F$, allora r è legale.
  - A questo punto, dato $A \in X^+$, si ha che $X \to A \in F^A = F^+$ deve essere soddisfatta da qualsiasi istanza legale di R, inclusa r stessa.
  - Poiché $X = Z_0 \subseteq Z_i$, ne segue che la dipendenza non può essere soddisfatta a vuoto poiché $t1[X] = t2[X]$. Dunque, l'unica possibilità affinché $X \to A \in F^+$ sia soddisfatta da r è $A \in Z_i$ in modo che si abbia $t1[A] = t2[A]$.
  - Dunque, siccome $A \in X^+ \Rightarrow A \in Z_i$, concludiamo che $X^+ \subseteq Z_i$.

4 Lemma: Inclusione delle chiusure

Dato uno schema R e due insiemi F e G di dipendenze funzionali su R, si ha che:

$$F \subseteq G^+ \Leftrightarrow F^+ \subseteq G^+ \tag{18}$$

4.1 Dimostrazione

- Denotiamo come $G \xrightarrow{\text{A}} F$ la possibilità di ottenere F partendo da G applicando una determinata quantità di assiomi di Armstrong.
- Ricordando che G^A è l'insieme di tutte le dipendenze funzionali ottenibile applicando assiomi di Armstrong su G, allora:

$$G \xrightarrow{\text{A}} F \Leftrightarrow \forall X \to Y \in F, \text{ si ha } X \to Y \in G^A = G^+ \Leftrightarrow F \subseteq G^+ \tag{19}$$

- Siccome $F \subseteq G \Leftrightarrow G \xrightarrow{\text{A}} F$, per definizione di $F^A = F^+$ si ha che:

$$F \subseteq G^+ \Rightarrow G \xrightarrow{\text{A}} F \xrightarrow{\text{A}} F^A = F^+ \Rightarrow F^+ \subseteq G^+ \tag{20}$$

- Viceversa, si ha che $F^+ \subseteq G^+ \Rightarrow F \subseteq F^+ \subseteq G^+$, quindi concludiamo che $F \subseteq G^+ \Leftrightarrow F^+ \subseteq G^+$

5 Chiusura di X in G

Dato uno schema R con decomposizione $\rho = R_1, ..., R_k$, dato un insieme F di dipendenze funzionali su R e posto:

$$G := \bigcap_{i=0}^{k} \pi_{R_i}(F)$$

preso $X \subseteq R$, il seguente algoritmo calcola X_G^+ tramite F:

Algorithm 2 Calcolo di X_G^+ tramite F

procedure CALCULATEX_G^+(R: schema, F: set of dependencies, X: set of attributes)
    $Z \leftarrow X$
    $S \leftarrow \emptyset$
    for $i \leftarrow 1$ to k **do**
        $S \leftarrow S \cup ((Z \cap R_i)_F^+ \cap R_i)$
    end for
    while $S \not\subseteq Z$ **do**
        $Z \leftarrow Z \cup S$
        for $i \leftarrow 1$ to k **do**
            $S \leftarrow S \cup ((Z \cap R_i)_F^+ \cap R_i)$
        end for
    end while
    $X_G^+ \leftarrow Z$
    return X_G^+
end procedure

5.1 Dimostrazione

- Siano $Z_0, Z_1, \ldots, Z_i, \ldots$ e $S_0, S_1, \ldots, S_i, \ldots$ gli insiemi calcolati ad ogni iterazione del ciclo while dell'algoritmo
- Osserviamo che $Z_i \subseteq Z_{i+1}$ per ogni $i \in \mathbb{N}$, dunque $Z_0, Z_1, \ldots, Z_i, \ldots$ è una sequenza monotona limitata da R, implicando che esiste un $f \in \mathbb{N}$ tale che $Z_f = Z_{f+1}$
- Siccome ciò può accadere solo se $S_f \subseteq Z_f$, ossia quando l'algoritmo termina, si ha che Z_f è l'output dell'algoritmo
- Dimostriamo per induzione che $Z_f \subseteq X_G^+$:
  - **Caso base ($i = 0$):**
    Alla 0-esima iterazione del `while`, prima di esso, si ha $Z_0 = X \subseteq X_G^+$.
  - **Ipotesi induttiva:**
    Per ogni $i \in \mathbb{N}$ si ha che $Z_i \subseteq X_G^+$.
  - **Passo induttivo ($i > 0$):**
    Dato $A \in Z_{i+1} = Z_i \cup S_i$, abbiamo che $A \in Z_i$ oppure $A \in S_i$. Quindi possiamo considerare due casi:
    * Se $A \in Z_i$, allora per ipotesi induttiva abbiamo che $A \in Z_i \subseteq X_G^+$.
    * Se $A \in S_i$, allora per la definizione stessa di S_i, $\exists\, j \leq k$ tale che $A \in ((Z_i \cap R_j)_F^+ \cap R_j)$.

      A questo punto, abbiamo che:

      $$A \in ((Z_i \cap R_j)_F^+ \cap R_j) \Leftrightarrow A \in (Z_i \cap R_j)_F^+ \wedge A \in R_j$$

      da cui otteniamo che:

      $$A \in (Z_i \cap R_j)_F^+ \Leftrightarrow (Z_i \cap R_j) \to A \in F^A = F^+.$$

      Dunque, siccome $(Z_i \cap R_j) \subseteq R_j$ e $A \in R_j$, allora si ha che:

$(Z_i \cap R_j) \rightarrow A \in \pi_{R_j}(F) = \{X \rightarrow Y \in F^+ \mid XY \in R_j\}$
Dai quali deduciamo che:

$(Z_i \cap R_j) \rightarrow A \in \pi_{R_j}(F) \subseteq G \subseteq G^+ = G^A$.

Inoltre, siccome $(Z_i \cap R_j) \subseteq Z_i$ e, per ipotesi induttiva, $Z_i \subseteq X_G^+$, otteniamo che:

$(Z_i \cap R_j) \subseteq Zi \subseteq X_G^+$

implicando quindi che $X \rightarrow (Z_i \cap R_j) \in G^A$.
Infine, per transitività otteniamo che:

$$X \rightarrow (Z_i \cap R_j),\, (Z_i \cap R_j) \rightarrow A \in G^A \Rightarrow X \rightarrow A \in G^A \Rightarrow A \in X_G^+$$

 - Dunque, siccome in entrambi i casi si ha $A \in Z_f \Rightarrow A \in X_G^+$, possiamo concludere che $Z_f \subseteq X_G^+$.

- Dimostriamo per induzione che $X_G^+ \subseteq Z_f$[1]:
  - Osserviamo che $X \subseteq Y \Rightarrow X_F^+ \subseteq Y_F^+$
  - Adesso sippiamo che $X = Z_0 \subseteq Z_f$, noi otteniamo che $X_G^+ \subseteq (Z_f)_G^+$. Noi possiamo dimostrare che $Z_f = (Z_f)_G^+$ che prova questo stato. Ovviamente $Z_f \subseteq (Z_f)_G^+$
  - Dobbiamo ora dimostrare l'Inclusione. Consideriamo $S_1 = A|Y \rightarrow V \in G, Y \subseteq Z_f, A \in V$ e otteniamo per l'esecuizione del primo passo della chiususa dell'algoritmo di Z_f su G. Noi abbiamo dimostrare che $S_1 \subseteq Z_f$. Poniamo $A \in S_1$. Questo significa che esiste $Y \rightarrow V \in G$ tale che $Y \subseteq Z_f$ e $A \in V$.
  - Adesso per definizione di F, esiste un indice j che $Y, V \subseteq R_j$ che significa $Y \subseteq Z_f \cap R_j$ e $A \in R_j$ e implica che $A \in (Z_j \cap R_j)_F^+ \cap R_j$.
  - Detto questo $A \in S_f \subseteq Z_f$

Abbiamo dimostrato che $Z_f = X_G^+$

[1]Dalle slides del Prof "Hand Notes"

6 Join senza perdita

7 Assiomi di Armstrong

8 Decomposizione che preserva F

8.1 Prima proprietà

8.2 Seconda proprietà

8.3 Terza proprietà

8.4 Definizione G

9 Dimostrazione ρ preserva F

10 Hash

11 B-tree

12 Chiusura di un insieme di attributi

12.1 Algoritmo

12.2 Dimostrazione correttezza

13 3NF (per dipendenza transitiva)

14 Chiusura di F e primo lemma

15 Isam

15.1 Variante Isam con chiavi indice che hanno valore ultimo record

16 Altre definizioni da sapere

16.1 Chiave minimale

16.2 Superchiave