

Algoritmo di Query Optimization

Prof. Alfredo Pulvirenti
Prof. Salvatore Alaimo

Equivalenza di espressioni

- Due espressioni sono **equivalenti** se producono lo stesso risultato qualunque sia l'istanza attuale della base di dati
- L'equivalenza è importante in pratica perché i DBMS cercano di eseguire espressioni equivalenti a quelle date, ma meno "costose"

Un'equivalenza importante

- Push selection (se A è attributo di R_2)

$$\sigma_{A=10}(R_1 \bowtie R_2) = R_1 \bowtie \sigma_{A=10}(R_2)$$

- Riduce in modo significativo la dimensione del risultato intermedio (e quindi il costo dell'operazione)

Esempio di Query

- Supponiamo che vogliamo trovare:
 - *tutti i professori che hanno dato a Mario Rossi piu' di 27.*

$\pi_{Professore}(\sigma_{Nome='Mario Rossi' \wedge Voto > 27}(STUDENTI \bowtie ESAMI \bowtie CORSI))$

Query Optimization

- La stessa query

$\pi_{Professore}(\sigma_{Nome='Mario Rossi' \wedge Voto > 27}(STUDENTI \bowtie ESAMI \bowtie CORSI))$

- Può essere espressa come

$\pi_{Professore}(\sigma_{Nome='Mario Rossi'}(STUDENTI) \bowtie \sigma_{Voto > 27}(ESAMI) \bowtie CORSI)$

- Che risulta essere molto più efficiente!

Regole per la query optimization

- Anticipare l'applicazione delle proiezioni e delle restrizioni rispetto al prodotto (e quindi alle giunzioni), in modo da ridurre la dimensione delle tabelle a cui applicare il prodotto (e le giunzioni).
- Le regole che seguono possono essere utilizzate per l'ottimizzazione di espressioni.

1. Raggruppamento di restrizioni

$$\sigma_{c(X)} \left(\sigma_{c(Y)}(E) \right) = \sigma_{c(X) \wedge c(Y)}(E)$$

Regole sulla restrizione

2. Commutatività di σ e π

$$a. \quad \sigma_{c(X)}(\pi_Y(E)) = \pi_Y(\sigma_{c(X)}(E)) \text{ se } X \subseteq Y$$

$$b. \quad \pi_Y(\sigma_{c(X)}(\pi_{XY}(E))) = \pi_Y(\sigma_{c(X)}(E)) \text{ se } X \not\subseteq Y$$

Restrizione e Prodotto

3. Anticipazione di σ rispetto a \times .

a. $\sigma_{c(X)}(E \times F) = \sigma_{c(X)}(E) \times F$ se $X \subseteq \text{attr}(E)$

b. $\sigma_{c(X) \wedge c(Y)}(E \times F) = \sigma_{c(X)}(E) \times \sigma_{c(Y)}(F)$
se $X \subseteq \text{attr}(E)$ e $Y \subseteq \text{attr}(F)$

a. $\sigma_{c(X) \wedge c(Y) \wedge c(Z)}(E \times F) = \sigma_{c(Z)}\left(\sigma_{c(X)}(E) \times \sigma_{c(Y)}(F)\right)$
se $X \subseteq \text{attr}(E), Y \subseteq \text{attr}(F), Z \subseteq \text{attr}(E) \cup \text{attr}(F)$

Regole per la proiezione

4. Raggruppamento di proiezioni.

$$\pi_X(\pi_Y(E)) = \pi_X(E) \text{ se } X \subseteq Y$$

5. Eliminazione di proiezioni superflue.

$$\pi_X(E) = E \text{ se } X = \text{attr}(E)$$

6. Anticipazione della π rispetto a \times .

$$\begin{aligned} \pi_{XY}(E \times F) &= \pi_X(E) \times \pi_Y(F) \\ \text{se } X &\subseteq \text{attr}(E) \text{ e } Y \subseteq \text{attr}(F) \end{aligned}$$

L'algoritmo

- Si applicano le seguenti tre regole (per anticipare la selezione) finché è possibile
 - A. Si anticipa σ rispetto a π usando la **2.a**
$$\sigma_{c(X)}(\pi_Y(E)) = \pi_Y(\sigma_{c(X)}(E))$$
 - B. Si raggruppano le restrizioni usando la **1**
$$\sigma_{c(X)}(\sigma_{c(Y)}(E)) = \sigma_{c(X) \wedge c(Y)}(E)$$
 - C. Si anticipa l'esecuzione di σ su \times usando la **3**.

Anticipazione delle proiezioni

D. Si eliminano le proiezioni superflue usando la 5

$$\pi_X(E) = E \text{ se } X = \text{attr}(E)$$

E. Si raggruppano le proiezioni mediante la regola 4

$$\pi_X(\pi_Y(E)) = \pi_X(E) \text{ se } X \subseteq Y$$

F. Si anticipa l'esecuzione delle proiezioni rispetto al prodotto usando ripetutamente la 2

$$[\pi_Y(\sigma_{c(X)}(\pi_{XY}(E))) = \pi_Y(\sigma_{c(X)}(E)) \text{ se } X \not\subseteq Y] \text{ e la } 6 \text{ [Anticipazione della } \pi \text{ rispetto a } \times].$$

Distributività

- $\sigma_C(R_1 \cup R_2) = \sigma_C(R_1) \cup \sigma_C(R_2)$
- $\sigma_C(R_1 - R_2) = \sigma_C(R_1) - \sigma_C(R_2)$
- $\pi_X(R_1 \cup R_2) = \pi_X(R_1) \cup \pi_X(R_2)$
- **Esercizio**
 - NON VALE $\pi_X(R_1 - R_2) = \pi_X(R_1) - \pi_X(R_2)$
costruire un controesempio per dimostrare la
disuguaglianza

- $\sigma_{C \vee D}(R) = \sigma_C(R) \cup \sigma_D(R)$
- $\sigma_{C \wedge D}(R) = \sigma_C(R) \cap \sigma_D(R)$
- $\sigma_{C \wedge \neg D}(R) = \sigma_C(R) - \sigma_D(R)$