Tutorial 6 — Query Optimization, Planning, Evaluation

Richard Wong
rk2wong@edu.uwaterloo.ca

Department of Electrical and Computer Engineering
University of Waterloo

March 4, 2018

Exercise 6-1

Give instances of relations R and S that show that the following pairs of relational algebra expressions are not equivalent:

1. $\Pi_A(R - S)$ and $\Pi_A(R) - \Pi_A(S)$
2. $\sigma_\theta(R ⟕ S)$ and $R ⟕ \sigma_\theta(S)$, where θ uses only attributes of S

Exercise 6-2

Consider relations $R(A, B, C)$, $S(C, D, E)$, $T(E, F)$, where A, C, and E are their respective primary keys.
Suppose $n_R = 1000$, $n_S = 1500$, $n_T = 500$.

1. What is the tightest upper bound we can place on $n_{R \bowtie S \bowtie T}$?
2. How could we compute the join efficiently?

Using the relational algebra equivalence rules, show how to derive the RHS expression from the LHS expression.

1. $\sigma_{\theta_1 \wedge \theta_2 \wedge \theta_3}(R) = \sigma_{\theta_1}(\sigma_{\theta_2}(\sigma_{\theta_3}(R)))$
2. $\sigma_{\theta_1 \wedge \theta_2}(R \bowtie_{\theta_3} S) = \sigma_{\theta_1}(R \bowtie_{\theta_3} \sigma_{\theta_2}(S))$, where θ_2 uses only attributes of S

Exercise 6-4

Let R be our relation with n_r records.
Suppose s_i records in R match a predicate θ_i: that is, $\sigma_{\theta_i}(R) = s_i$.
The *selectivity* of θ_i, $sel_{\theta_i}(R)$ is defined to be $\frac{s_i}{R}$. This represents the probability that a record in R satisifies θ_i.

Derive the selectivity formulas for the following complex selections:

1. conjunction: $\sigma_{\theta_1 \wedge \theta_2 \wedge \ldots \wedge \theta_m}(R)$
2. negation: $\sigma_{\neg\theta}(R)$
3. disjunction: $\sigma_{\theta_1 \vee \theta_2 \vee \ldots \vee \theta_m}(R)$

Exercise 6-5

What are some strategies that a query optimizer could use to reduce the cost of query plan selection, or the cost of the query itself?