Tutorial 5 — Query Optimization, Planning, Evaluation

Richard Wong
rk2wong@edu.uwaterloo.ca

Department of Electrical and Computer Engineering
University of Waterloo

February 26, 2018

Exercise 5-1

1. What are the two metrics we will use to estimate query operation costs?
2. What does each metric represent?
3. How do we use the metrics to arrive at an estimate?

Exercise 5-2

Suppose we run a query that performs a single-attribute GREATER THAN comparison in its WHERE clause.

e.g.
SELECT * FROM people
WHERE age > 20

How might the following evaluation strategies impact the (worst-case) cost of the operation?

1. Use a primary index if there is one.
2. Use a secondary index if there is one.
3. Use a linear scan.

What if we were performing a LESS THAN comparison?

Exercise 5-3

Suppose there are b_r blocks in R and b_s blocks in S.
Derive the worst-case and best-case cost estimate for a block nested-loop join, $R \bowtie_\theta S$.

Exercise 5-4

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 5-5

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?