Tutorial 5 — Query Optimization, Planning, Evaluation

Richard Wong
rk2wong@edu.uwaterloo.ca

Department of Electrical and Computer Engineering
University of Waterloo

February 24, 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 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

Derive the worst-case and best-case cost estimate for a block nested-loop join.

Exercise 5-4

Derive the formulas for *selectivity* of the following types of selections:

Recall that selectivity is the estimated probability that a tuple matches a selection criterion.

1. conjunction (WHERE θ_1 AND θ_2 AND ... AND θ_m)
2. negation (WHERE NOT θ)
3. disjunction (WHERE θ_1 OR θ_2 OR ... OR θ_m)

Exercise 5-5

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