Tutorial 10 — Parallelism and Distributed Databases

Richard Wong
rk2wong@edu.uwaterloo.ca

Department of Electrical and Computer Engineering
University of Waterloo

April 1, 2018

Exercise 10-1

What kinds of queries are the following partitioning schemes well-suited for?

1. round-robin
2. range partitioning
3. hash partitioning

Exercise 10-2

How would a distributed DB using the following partitioning schemes handle **addition** of nodes?

1. round-robin
2. range partitioning
3. hash partitioning

Exercise 10-3

What factors would account for **skew** in the following partitioning schemes?

1. range partitioning
2. hash partitioning

Exercise 10-4

What factors would account for **load imbalance** in the following partitioning schemes?

1. range partitioning
2. hash partitioning

Exercise 10-5

How can we alleviate the problem of load imbalance in range partitioning?

Exercise 10-6

Suppose we have the following relation:

```
employee(name, address, salary, plantNumber)
```

The relation is **fragmented horizontally**,
and each fragment has a **local replica**,
and a **replica in New York**.

Provide a reasonable processing strategy for each of the following queries **made from the plant in Montreal**:

1. Find all employees at the Toronto plant.
2. Find the average salary of all employees.
3. Find the highest-paid employee in Toronto, Vancouver, and Edmonton.