Assignment 2

Joining Techniques

Submission Date: Thursday 20-Oct-2016 (Start of Class)

Note: Read the assignment carefully and thoroughly. If you have any confusion in understanding the statements then take suitable assumption and mention it before solving the question.

Question 1. Discuss the best cases for the following joining techniques. Please discuss with examples. You could take one schema for examples in all three parts. Also discuss the time complexities for all three techniques. (**3+3+3 marks**)

- Nested Loop Join (NLJ)
- Sort Merge Join (SMJ)
- Hash Join (HJ)

Instructions for Question 2: You will have to take assumptions for block size (B), record size (R), number of records (r), available memory (K), indexing column, index type (clustered/non-clustered), and all remaining factors required in calculations. Please make sure that these factors for no two students should be similar.

Question 2. Consider two tables A and B which have to be joined. Calculate the cost of joining the two tables on their common attribute. You will have to consider the following cases and have to calculate cost for all these cases:

1. When the table sizes are almost similar. Let's say 1 row of table A joins with only 1 row of table B.

- When RAM size is not sufficient.
- When RAM size sufficient for both tables.

2. When the sizes of both tables are significantly different. Let's say 1 row of table A joins with 10 rows of table B.

- When RAM size is not sufficient.
- When RAM size sufficient for both tables.

Your solutions should include costs of I/Os for **Naïve nested loop join (or Block NLJ/Index NLJ/clustered index NLJ), Sort merge join (or merge join), and in-memory hash join (or Grace hash join/Recursive hash join)**.

Marks distribution: There are three types of joins, and each has two parts (where sizes are same and different), and each of the two parts has two parts further where RAM size is sufficient for tables to fit in memory and where RAM size is not sufficient

for tables to fit in memory. Each sub part will be of 2 marks, which makes each Join cost worth 8 marks. And total question will be for **24 marks**.