

National University of Computer and Emerging Sciences, Lahore Campus



Course:	Data Warehousing and Business Intelligence	Course Code:	DS3003
Program:	BS (Data Science)	Semester:	Fall 2023
Out Date:	17-Oct-2023	Total Marks:	
Due Date:	Tue 24-Oct-2023 (start of class)	Weight:	
Section	BDS-5A	Page(s):	1
Assignment:	3 (Joining Techniques)		

Instructions/Notes:

- Use any valid assumption where needed.
- You are required to submit the hard copy of your assignment at the start of your class.
- For any query, please contact your TA.

Instructions: 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.

Question. Consider two tables, Table1 and Table2, which must be joined. Calculate the cost of joining the two tables on their common attribute. You will have to consider the following cases and must calculate I/O cost for all these cases:

1. When the table sizes are almost similar. Let's say one row of Table1 joins with only one row of Table2 (i.e., 1:1 relationship).
 - a. When available main memory (K) is not sufficient to store any operand table.
 - b. When available main memory (K) is sufficient to store at least one operand table.
2. When the sizes of both tables are significantly different. Let's say one row of Table1 joins with n rows of Table2.
 - a. When available main memory (K) is not sufficient to store any operand table.
 - b. When available main memory (K) is sufficient to store at least one operand table.

Your solutions should include costs of I/Os for:

- Naïve nested loop join and (Block NLJ/Index NLJ/clustered indexed NLJ)
- (Sort) merge join
- Hash join