**CSE 512 – Assignment 3**

**Maximum possible points: 5**

The required task is to build a generic parallel sort and parallel join algorithm.

1. Implement a Python function **ParallelSort()** that takes as input: (1) **InputTable** stored in a PostgreSQL database, (2) **SortingColumnName** the name of the column used to order

the tuples by. **ParallelSort()** then sorts all tuples (using five parallelized threads) and stores the sorted tuples for in a table named **OutputTable** (the output table name is passed to the function). The **OutputTable** contains all the tuple present in **InputTable** sorted in ascending order.

Function Interface: -

**ParallelSort (InputTable, SortingColumnName, OutputTable, openconnection) InputTable –** Name of the table on which sorting needs to be done. **SortingColumnName –** Name of the column on which sorting needs to be done, would be either of type integer or real or float. Basically Numeric format. Will be Sorted in Ascending order.

**OutputTable –** Name of the table where the output needs to be stored.

**openconnection –** connection to the database.

2. Implement a Python function **ParallelJoin()** that takes as input: (1) **InputTable1** and **InputTable2** table stored in a PostgreSQL database, (2) **Table1JoinColumn** and **Table2JoinColumn** that represent the join key in each input table respectively. **ParallelJoin()** then joins both **InputTable1** and **InputTable2** (using five parallelized threads) and stored the resulting joined tuples in a table named **OutputTable** (the output table name is passed to the function). The schema of **OutputTable** should be **InputTable1.Column1, InputTable.Column2, …**, **InputTable2.Column1, InputTable2.Column2…**.

Function Interface: -

**ParallelJoin (InputTable1, InputTable2, Table1JoinColumn, Table2JoinColumn, OutputTable, openconnection)**

**InputTable1 –** Name of the first table on which you need to perform join.

**InputTable2 –** Name of the second table on which you need to perform join. **Table1JoinColumn –** Name of the column from first table i.e. join key for first table. **Table2JoinColumn –** Name of the column from second table i.e. join key for second table.

**OutputTable -** Name of the table where the output needs to be stored.

**openconnection –** connection to the database.

**Naming Convention to be followed strictly:**

Database name – *ddsassignment3*

Postgres User name – *postgres*

Postgres password – *1234*

**Instructions on how this will be tested: -**

Please follow these instructions closely.

**1.** Two tables would be created in the database manually.

**2.** The created tables would contain at least an integer field, which would be used for both

Parallel Sorting and Parallel Joining.

**3.** Then, the ParallelSort() and ParallelJoin() Function would be called to check the correctness of the program.

**4.** Your code should use 5 threads for both ParallelSort() as well as ParallelJoin().

**5.** Your code should be able to handle table irrespective of its schema.

**6.** Do not make your code dependent on any particular table; it should be able to work on any table and any given input columns.

**Instructions for Assignment: -**

Please follow these instructions closely **else Marks will be deducted.**

**1.** Please follow the function signature as provided in the Assignment3\_Interfacy.py.

**2.** Please use the same database name, table name, user name and password as provided in the assignment to keep it consistent.

**3.** Please make sure to run the file before submitting and make sure there is no indentation error. In case of any compilation error, 0 marks will be given.

**4.** Do not modify any function signature in Assignment3\_Interface.py. In case any modification is needed, please post the same on discussion board.

**5.** For any case of doubt in the assignment, PLEASE USE Discussion Boards, Individual mails would not be entertained.

**6.** Also, It is an individual’s responsibilities to clarify his/her doubts, so read and use

Discussion Board extensively.

**Submission Instructions: -**

Submit ***Assignment3\_Interface.py*** directly to blackboard. Do not upload \*.zip files or change its name.

**Note: -**

Failure to follow the instructions provided in the document will result in 0.5 point loss.