

## CSE 512 – Assignment 3

This assignment is an **Individual-based** assignment. The required task is to implement parallel query processing operators.

**Required Task.** Below are the steps you need to follow to fulfill this assignment:

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). Each tuple in `OutputTable` is similar to a tuple in `InputTable` with the addition of an extra column (named `TupleOrder`) that represents the tuple sorting order. `TupleOrder` takes a value between 1 and N such that N denotes the total number of tuples in `InputTable`.

`ParallelSort (Table, SortingColumnName, OutputTable, openconnection)`

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 similar to the schema of both `InputTable1` and `InputTable2` combined.

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

**Deadline.** Sunday, April 19th 2015 (11:59 pm). Each student should submit her/his assignment to blackboard.