**CS 49995 & CS 63016 ST: Big Data Analytics**

**Homework 4**

**Instructor:** Xiang Lian

**Due Date:** Please refer to the course website

1. Please use your own words to describe the Map and Reduce functions. [20 points]

2. Assume that we have a relational database *D* containing tuples *t* with attributes *t*[*A*]. Write the pseudo code of Map and Reduce functions for retrieving all tuples with attribute *t*[*A*] > 100. [20 points]

3. Assume that we have two relational databases *R* and *S* containing tuples *r* and *s*, respectively. Write the pseudo code of Map and Reduce functions for joining databases *R* and *S* on attribute *A* (i.e., joining tuples *r* and *s* satisfying *r*[*A*] = *s*[*A*]). [20 points]

4. Please write a tutorial for installing Apache Hadoop (on either windows or MacOS) or configuring Amazon Web Services. *Please use screen captures and record the entire installation process step by step.* [40 points]

**Bonus Question [20 extra points]**

5. Implement the Word Count example (as mentioned in the lecture slides) on Hadoop or AWS. *Please include the source code and add screen captures to illustrate the execution process.*

**Submission**

Submit an electronic copy of your homework solution to the [Blackboard](https://learn.kent.edu/).