CSE4/587 Final Exam Review Spring 2016

1 MAPREDUCE FOUNDATIONS (10 POINTS)

A formal representation of the various elements of the MR workflow: combiners, partitioners. How many reducers? How many reduce calls? Input of mappers and reducers etc. You will derive expression in terms of the variables given. See Q2 of Spring 2014. Hadoop architecture.

2 MAPREDUCE ALGORITHMS (20 POINTS)

Word co-occurrence; Shortest path; Page Rank; given a problem with data you will trace the execution of the algorithm and provide the values of the intermediate steps and the final result.

3 PROBLEM SOLVING USING MR (20 POINTS)

Given a big-data problem provide a MR solution.

4 MR ABSTRACTION THROUGH PIG (10 POINTS)

Given a problem you will write a PIG script. Given a problem provide a simple PIG script. Some foundational operations of PIG.

5 SPARK FOUNDATIONS (20 POINTS)

Spark programming model; RDD: read the paper; transformations and actions on RDD; SC operations. Spark eco systems: supporting APIs. How is fault-tolerance accomplished vs Hadoop.

6 PROJ2 EXPERIENCE (20 POINTS)

Specific questions on data analysis that you carried out in project2.

7 DATE, TIME, LOCATION

Location: NSC 225

Date 5/9/2016

Time: 7.15 – 10.15PM

Closed book exam.