**SEIS736 Big Data Architecture Homework 3**

1. Using your Lab 4 results and comparing the job statistics from word-count with and without a combiner:

a.Which one ran faster? Can you explain this result?

**Answer**

**TRUE**

WordCount with Combiner: 53Sec

WordCount Without Combiner: 40Sec

**FALSE**

**Explanation:**

Since the number of input records to the combiner (mapper output)was drastically reduced from 964453 to 189 ,the execution time with combiner was higher.

b. Are the mapper output records equal to the combiner input records?

**Answer**

Yes.

**TRUE & FALSE**

Mapper Output records – 964453

Combiner Input Records - 964453

c. Are the combiner output records equal to the reducer input records?

**Answer**

Yes

FALSE

Combiner output Records – 189

Reducer Input Records – 189

TRUE

Combiner output Records – 361

Reducer Input Records – 361

d. Are the reducer output records the same for both versions?

**Answer**

Yes

**FALSE**

Reducer Output Records with combiner – 35

Reducer Output Records Without Combiner – 35

**TRUE**

Reducer Output Records with combiner – 60

Reducer Output Records Without Combiner – 60

e. What is the percentage reduction in reducer input records with the use of the combiner?

**Answer**

There is a 99% reduction in the reducer input records with the use of combiner.

1. Take the two wordcount JAR files (with and without combiner) and scp/WinSCP them to the cluster. For input, use the HDFS directory /SEIS736/gutenberg directly, so do not do a “hadoop fs –get” on this directory. This directory contains 4,160 Project Gutenberg books. Run both versions of your code, each with 20 reducers.
   1. What were the execution times? Which ran faster? Can you explain this result?

TRUE

With Combiner :2 mins 54 sec -10

Without Combiner:4 mins 43 sec

FALSE

With Combiner: 2 mins 54 sec

Without Combiner:3 mins 20 sec

* 1. How many HDFS bytes were read and written for each version?

**Answer**

**TRUE -With Combiner and Without Combiner**

HDFS Bytes Read:1636735854

HDFS Bytes Written:378

**FALSE- With Combiner and Without Combiner**

HDFS Bytes Read:1636735854

HDFS Bytes Written:222

* 1. Turn in the job statistics pages for both cluster runs and the answers to the questions.  -DONE