Exercises 3 through 5 require some programming background and a working Hadoop environment. The text of the novel War and Peace can be downloaded from http://onlinebooks.library.upenn.edu/ and used as the dataset for these exercises. However, other datasets can easily be substituted. Document all processing steps applied to the data.

1. Use MapReduce in Hadoop to perform a word count on the speciﬁed dataset.
2. Use Pig to perform a word count on the speciﬁed dataset.
3. Use Hive to perform a word count on the speciﬁed dataset.

**Installation Hadoop in Windows 10:**

Please refer Hadoop Installation Guides for Windows 10.pdf

Installation Apache Pig in Windows 10:

<https://www.solutionmandi.com/2018/11/pig-installation-on-windows-10.html>

Installation Apache HIVE in Windows 10:

<https://www.solutionmandi.com/2018/11/hive-installation-on-windows-10.html>

**Guideline using Hadoop (Mapreduce) word count in Windows 10:**

\*\*\*Please make sure Hadoop already install in windows.

Hadoop version: hadoop-2.8.0

Open command prompt (Run as administrator)

To start Hadoop, go to Hadoop directory:

* cd C:\hadoop-2.8.0\sbin
* start-all.cmd

To verify current directory path/item:

* hdfs dfs -ls /

To upload txt file into Hadoop:

* hadoop fs -put C:\Users\hp\Desktop\Programming\Hadoop\References\book.txt \

Code the word count java project using Eclipse software:

<https://www.dezyre.com/hadoop-tutorial/hadoop-mapreduce-wordcount-tutorial>

Run the Java code:

* hadoop jar C:\Users\hp\...\WordCountTry.jar packagename.classname /book.txt /output

Double check the output:

* hdfs dfs -ls \output

Copy the output to local computer

* hdfs dfs -get \output\part-r-00000 C:\Users\hp\...\HadoopMapreduce

**Guideline using Apache Pig (Mapreduce) word count in Windows 10:**

\*\*\*Please make sure Hadoop and Apache Pig already install in windows.

Hadoop version: hadoop-2.8.0

Apache Pig version: pig-0.17.0

Open command prompt (Run as administrator)

To start Hadoop, go to Hadoop directory:

* cd C:\hadoop-2.8.0\sbin
* start-all.cmd

Code the Pig Word Count Scripts using notepad and save it as Pig file:

<http://salsahpc.indiana.edu/ScienceCloud/pig_word_count_tutorial.htm>

After that go to Pig directory:

* cd C:\pig-0.17.0\bin
* pig -x mapreduce C:\Users\hp\Desktop\...\PigMapreduce\WordCount.pig

Quit the Pig system:

* quit;

Double check the output:

* hdfs dfs -ls \output

Copy the output to local computer

* hdfs dfs -get \output\part-r-00000 C:\Users\hp\...\PigMapreduce

**Guideline using Apache HIVE (Mapreduce) word count in Windows 10:**

\*\*\*Please make sure Hadoop, Apache HIVE and Derby server already install in windows.

Hadoop version: hadoop-2.8.0

Apache HIVE version: apache-hive-2.1.0-bin

Derby Server: db-derby-10.12.1.1-bin

Open command prompt (Run as administrator)

To start Hadoop, go to Hadoop directory:

* cd C:\hadoop-2.8.0\sbin
* start-all.cmd

Open Derby Server:

* cd C:\db-derby-10.12.1.1-bin\bin
* startNetworkServer -h 0.0.0.0

Open another command prompt (Run as administrator)

* cd C:\apache-hive-2.1.0-bin\bin

Check Network Server Control:

* jps -m

Execute HIVE:

* hive

Code the HIVE coding:

* CREATE TABLE FILES (line STRING);
* LOAD DATA INPATH '\book.txt' OVERWRITE INTO TABLE FILES;
* CREATE TABLE word\_counts AS
* SELECT word, count(1) AS count FROM
* (SELECT explode(split(line, ' ')) AS word FROM FILES) w
* GROUP BY word
* ORDER BY word;

Quit HIVE environment:

* quit;

Copy the output to local computer

* hdfs dfs -get \output\part-r-00000 C:\Users\hp\...\HiveMapreduce