**Hadoop Instructions**

1. Login to Hadoop Server: had6110.cs.pitt.edu
2. To start with, input files need to be copied into HDFS. To do this,
   1. Make a directory:

hadoop fs –mkdir [project foldername/input foldername]

haddop fs –mkdir [wordcount/inputdemo]

* 1. Copy the input files in this directory from the Hadoop server (cd into the directory where input files are stored)

hadoop fs –put [src location] [dest location]

hadoop fs –put inputdemofiles/\* wordcount/inputdemo

1. To compile the java program:
   1. Compile first (cd into the location where ‘src’ (containing java files) is kept

javac -verbose -classpath /opt/hadoop/share/hadoop/common/hadoop-common-2.7.2.jar:/opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.7.2.jar -d class src/\*.java

it will create class files in the ‘class’ folder

* 1. Create jar file:

jar -cvf [jar\_name].jar -C class/ .

jar -cvf wordcount.jar -C class/ .

* 1. Run Index of the document

hadoop jar [jar\_name].jar [classToRun] [arg 0 : input path] [arg 1 : output path]

hadoop jar wordcount.jar WordCount [wordcount/inputdemo] [wordcount/indexoutput]

this creates indexing of all the documents present in inputdemo folder (after step 2).

To check the indexing, we do:

hadoop fs –cat wordcount/indexoutput/part-00000

* 1. Run Search: the output path on the precious step is the input path for this step, so

hadoop jar [jar\_name].jar [classToRun] [arg 0 : input path] [arg 1 : output path] [arg 2: query]

hadoop jar wordcount.jar WordCountQuery [wordcount/indexoutput] [wordcount/searchoutput] [Rosalind]

To check the searching results, we do:

hadoop fs –cat wordcount/searchoutput/part-00000

For both c/d the output folder name has to be changed since it throws an exception if the folder is already present.

1. To check for any folder in HDFS, do

hadoop fs –ls [project\_name]

hadoop fs –ls wordcount

This will all the variation of input/output files if present

To check for the contents of the index/search result

hadoop fs –cat [project\_name]/[file\_name]/part-00000

hadoop fs –car wordcount/searchoutput/part-00000