/\* Student Number: 10241485 \*/

/\* Name: Jeslyn Ho Ka Yan \*/  
/\* Date: 13 Oct 2024 \*/

/\* Assignment 1, Task 1 \*/

/\* \*/

/\* Setup and Configuration ----------------------------------------------------- \*/  
/\* Set Hadoop classpath for Java compilation \*/

bigdata@bigdata-VirtualBox:~$ export HADOOP\_CLASSPATH=$($HADOOP\_HOME/bin/hadoop classpath)

/\* Check if Hadoop classpath is set correctly \*/

bigdata@bigdata-VirtualBox:~$ echo $HADOOP\_CLASSPATH

/usr/share/hadoop/etc/hadoop:/usr/share/hadoop-2.7.3/share/hadoop/common/lib/\*:/usr/share/hadoop-2.7.3/share/hadoop/common/\*:/usr/share/hadoop-2.7.3/share/hadoop/hdfs:/usr/share/hadoop-2.7.3/share/hadoop/hdfs/lib/\*:/usr/share/hadoop-2.7.3/share/hadoop/hdfs/\*:/usr/share/hadoop-2.7.3/share/hadoop/yarn/lib/\*:/usr/share/hadoop-2.7.3/share/hadoop/yarn/\*:/usr/share/hadoop-2.7.3/share/hadoop/mapreduce/lib/\*:/usr/share/hadoop-2.7.3/share/hadoop/mapreduce/\*:/usr/share/hadoop/contrib/capacity-scheduler/\*.jar  
  
/\* Check Hadoop home directory \*/

bigdata@bigdata-VirtualBox:~$ echo $HADOOP\_HOME

/usr/share/hadoop

/\* Project Directory Setup ----------------------------------------------------- \*/  
/\* Navigate to project directory \*/

bigdata@bigdata-VirtualBox:~$ cd ISIT312/A1  
  
/\* List files in the project directory \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ ls

file1 file2 solution1.java

/\* HDFS Operations ------------------------------------------------------------- \*/  
/\* Create a folder 'myfolderA' in HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -mkdir myfolderA  
  
/\* List HDFS contents to verify directory creation \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -ls

Found 3 items

drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:33 .sparkStaging

drwxr-xr-x - bigdata supergroup 0 2024-10-11 17:09 WordCountOutput

drwxr-xr-x - bigdata supergroup 0 2024-10-13 02:36 myfolderA  
  
/\* Upload local files 'file1' and 'file2' to 'myfolderA' in HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -put /home/bigdata/ISIT312/A1/file1 myfolderA/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -put /home/bigdata/ISIT312/A1/file2 myfolderA/  
  
/\* List contents of 'myfolderA' to confirm files are uploaded \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -ls myfolderA

Found 2 items

-rw-r--r-- 1 bigdata supergroup 27 2024-10-13 02:40 myfolderA/file1

-rw-r--r-- 1 bigdata supergroup 26 2024-10-13 02:40 myfolderA/file2

/\* View contents of 'file1' from HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -cat myfolderA/file1

This is file1.

Halloween

/\* View contents of 'file2' from HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -cat myfolderA/file2

This is file2.

Is coming!

/\* Java Compilation and JAR Creation --------------------------------------- \*/  
/\* Compile Java program with Hadoop classpath \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ javac -classpath ${HADOOP\_CLASSPATH} solution1.java  
  
/\* Upload compiled Java file to HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -put solution1.java  
  
/\* Create a JAR file containing compiled Java class \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ jar cvf solution1.jar solution1\*.class

added manifest

adding: solution1.class(in = 1526) (out= 806)(deflated 47%)

/\* List files to confirm the JAR was created \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ ls

file1 file2 solution1.class solution1.jar solution1.java  
  
/\* List HDFS files to verify Java file upload \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -ls

Found 4 items

drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:33 .sparkStaging

drwxr-xr-x - bigdata supergroup 0 2024-10-11 17:09 WordCountOutput

drwxr-xr-x - bigdata supergroup 0 2024-10-13 02:40 myfolderA

-rw-r--r-- 1 bigdata supergroup 1743 2024-10-13 02:45 solution1.java

/\* Running the Hadoop Job----------------------------------------------------- \*/  
/\* Run Hadoop job with the created JAR \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop jar solution1.jar solution1 myfolderA/file1 myfolderA/file2 myfolderA/mergedOutput.txt  
  
/\* Check the output of the merged file \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP\_HOME/bin/hadoop fs -cat myfolderA/mergedOutput.txt

This is file1.

Halloween

This is file2.

Is coming!

/\* **Summary of Steps for Hadoop Job Execution**

1. Set Up Hadoop Classpath: Configure the classpath using `export

HADOOP\_CLASSPATH=$($HADOOP\_HOME/bin/hadoop classpath)` and verify with `echo

$HADOOP\_CLASSPATH`.

2. Navigate to Project Directory: Change to the project directory with `cd ISIT312/A1`.

3. Create HDFS Directory: Create a new folder in HDFS using `$HADOOP\_HOME/bin/hadoop fs -mkdir myfolderA` and confirm its creation by listing HDFS contents.

4. Upload Files\*\*: Transfer local files 'file1' and 'file2' to HDFS with `$HADOOP\_HOME/bin/hadoop fs -put`.

5. View Uploaded Files: Check the contents of the uploaded files using

`$HADOOP\_HOME/bin/hadoop fs -cat`.

6. Compile Java Program: Compile `solution1.java` with the Hadoop classpath using `javac -classpath ${HADOOP\_CLASSPATH} solution1.java`.

7. Create and Upload JAR File\*\*: Package the compiled classes into a JAR file with `jar cvf solution1.jar solution1\*.class` and upload it to HDFS.

8. Run Hadoop Job: Execute the job using the command:

$HADOOP\_HOME/bin/hadoop jar solution1.jar solution1 myfolderA/file1 myfolderA/file2

myfolderA/mergedOutput.txt

9. Check Output: Verify the merged output file's contents with `$HADOOP\_HOME/bin/hadoop fs

-cat myfolderA/mergedOutput.txt`.

\*/