CS 6350

ASSIGNMENT – 1b

Names of students in your group:

Varun Mehrotra (vxm170830)

Ram Gopal V (pxv172130)

Number of free late days used: 0  
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

**Part-1**

**Implementation—**

* Part1.java contains the source code of Map and Reduce Class.
* StopWord.java has the list of stop words to be removed during map phase.
* MiscUtils.java has the comparator to sort the HashMap by values.
* FileCopy.java is to copy file containing the Top20 words to HDFS.

**How to run –**

Hadoop jar Assignment-1b.jar <PackageName.ClassName> <InputDirectory> <OutputDirectory>

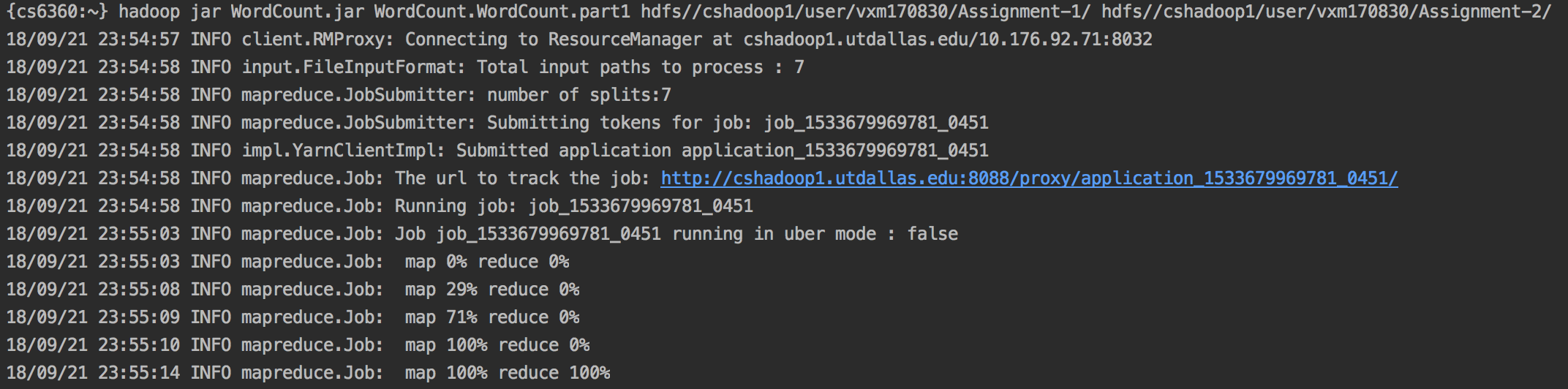
**Example --**

hadoop jar Assignment-1b.jar WordProblem.part1 hdfs//cshadoop1/user/vxm170830/Assignment-1/ hdfs//cshadoop1/user/vxm170830/part1/

**Deliverables –**

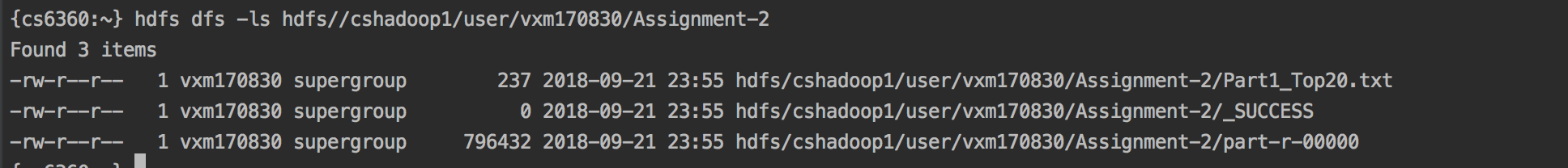
* Part1.java
* StopWord.java
* FileCopy.java
* MiscUtils.java

**Screenshots—**

****

**Output—**

* Part-r-00000 contains the total count for each key.
* Part1\_Top20.txt contains the top 20 most frequent used words.

****

**Part-2**

**Implementation --**

* Part2.java contains the source code of Map and Reduce Class.
* MiscUtils.java has the comparator to sort the HashMap by values.
* FileCopy.java is to copy file containing the Top20 words to HDFS.

**How to run –**

Hadoop jar Assignment-1b.jar <PackageName.ClassName> <InputDirectory> <OutputDirectory>

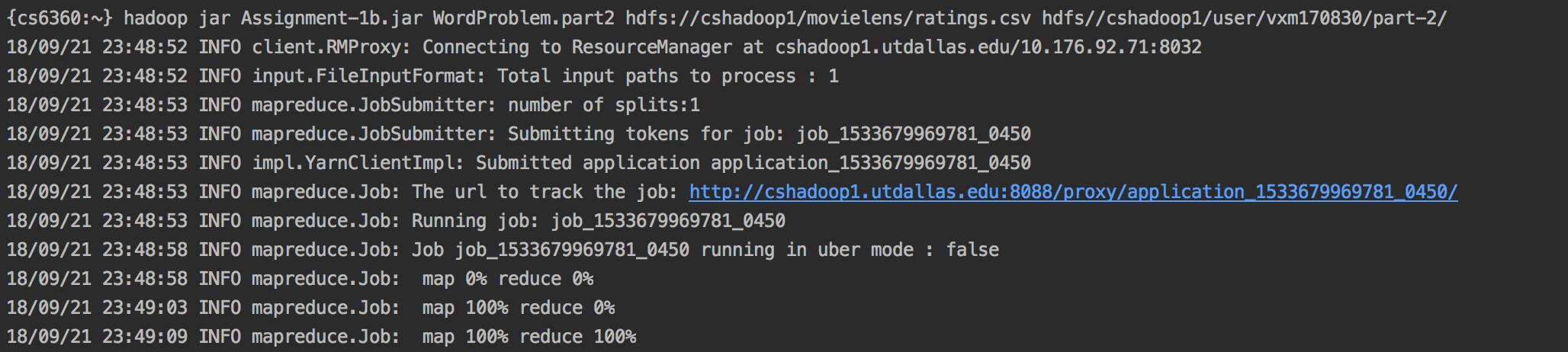
**Example --**

hadoop jar Assignment-1b.jar WordProblem.part2 hdfs://cshadoop1/movielens/ratings.csv hdfs//cshadoop1/user/vxm170830/part-2/

**Deliverables –**

* Part2.java
* FileCopy.java
* MiscUtiils.java

**Screenshots—**

****

**Output—**

* Part-r-00000 contains the total count for each key.
* Part2\_Top20.txt contains the top 20 movies with highest average rating.

