CS-GY 6513 Big Data Homework 1

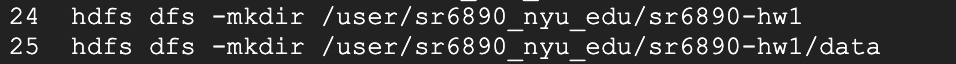
Sashank RM

sr6890

**HDFS Directory Creation & Listings**:

hdfs dfs -mkdir /user/sr6890\_nyu\_edu/sr6890-hw1

hdfs dfs -mkdir /user/sr6890\_nyu\_edu/sr6890-hw1/data



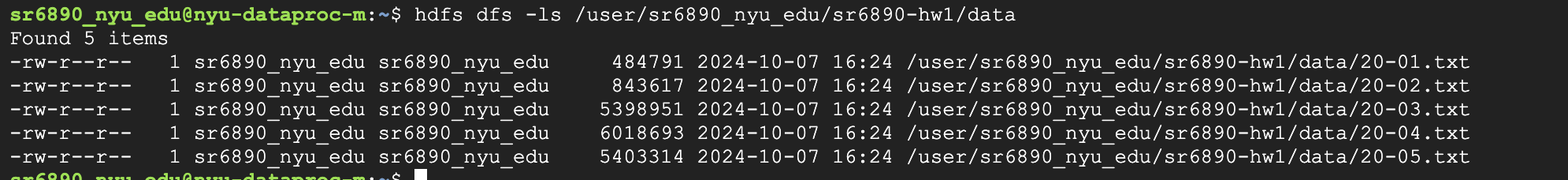
**Upload Input Files to HDFS:**

hdfs dfs -put 20-\*.txt /user/sr6890\_nyu\_edu/sr6890-hw1/data

A screenshot of a computer

Description automatically generated

**Verify the uploaded files:**



A black screen with white text

Description automatically generated

**Run the First MapReduce Job (Word Count)**:

a) Part 2.1

cd /home/sr6890\_nyu\_edu

hadoop com.sun.tools.javac.Main WordCount.java Top10Words.java

jar cf wordcount.jar \*.class

hadoop jar wordcount.jar WordCount /user/sr6890\_nyu\_edu/sr6890-hw1/data /user/sr6890\_nyu\_edu/sr6890-hw1/output-job1

hdfs dfs -ls /user/sr6890\_nyu\_edu/sr6890-hw1/output-job1

hadoop jar wordcount.jar Top10Words /user/sr6890\_nyu\_edu/sr6890-hw1/output-job1 /user/sr6890\_nyu\_edu/sr6890-hw1/output-job2

hdfs dfs -cat /user/sr6890\_nyu\_edu/sr6890-hw1/output-job2/part-r-00000A computer screen with white text

Description automatically generated

A screenshot of a computer

Description automatically generatedA black screen with white text

Description automatically generated

b) Part 2.2 (extra credit)

hadoop com.sun.tools.javac.Main IDAssigner.java

jar cf idassigner.jar \*.class

hadoop jar idassigner.jar IDAssigner /user/sr6890\_nyu\_edu/sr6890-hw1/output-job1 /user/sr6890\_nyu\_edu/sr6890-hw1/output-id-assign

hdfs dfs -cat /user/sr6890\_nyu\_edu/sr6890-hw1/output-id-assign/part-r-00000A screenshot of a computer

Description automatically generated