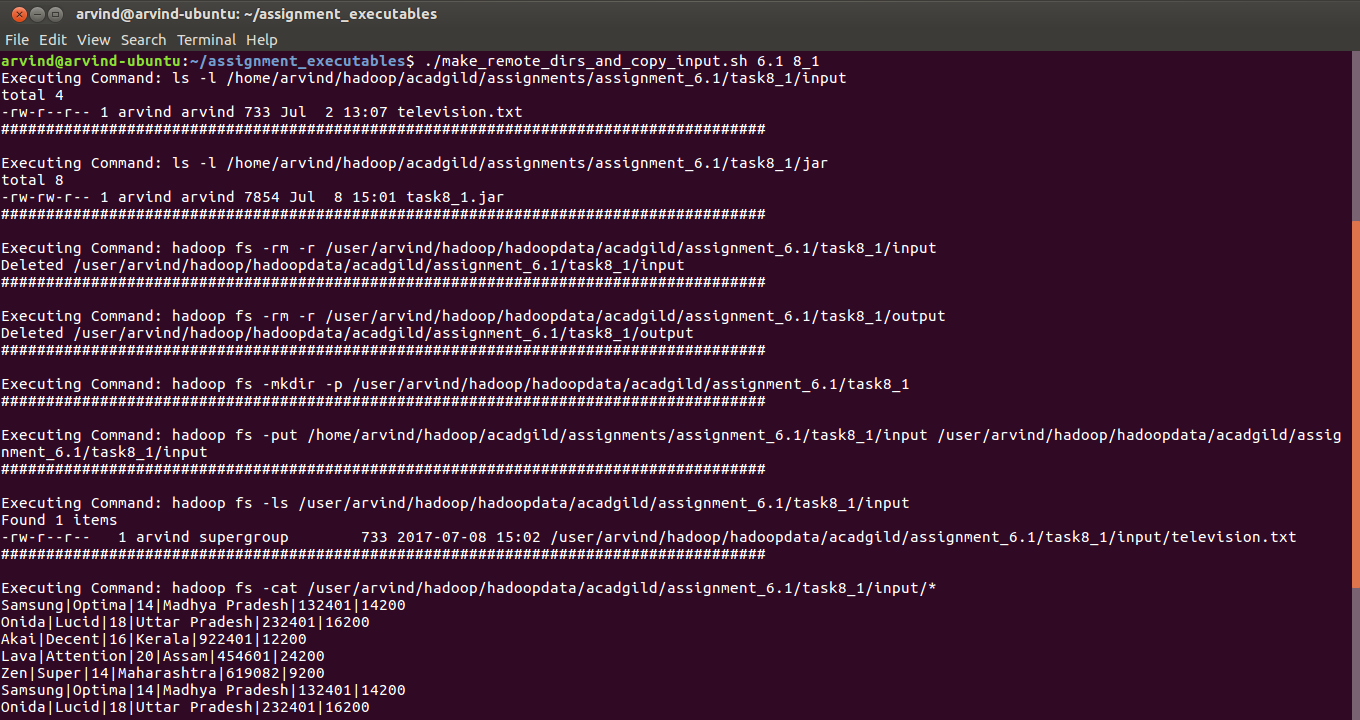
I have divided task 8 into two parts – 8\_1 and 8\_2. This is he document for task 8\_1. The map reduce program for this task writes the output to a sequence file which will serve as input to the main task(i.e. 8\_2)

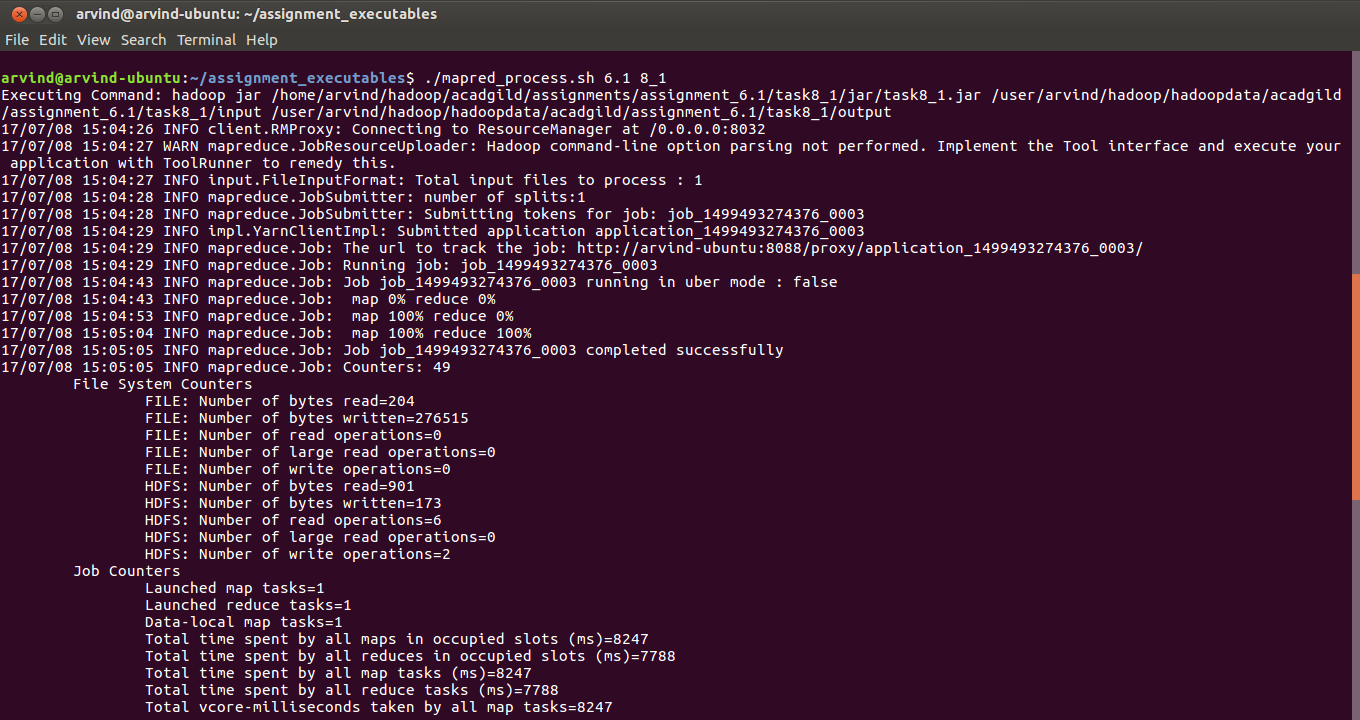
**make\_remote\_dirs\_and\_copy\_input.sh take 2 inputs, the assignment number and task number..**

1. It verifies if the input file and the jar for the assignment are kept in the appropriate location in the local file system using ls -l command.
2. Delete input and output directories in the hdfs for this assignment and task if already exists using hadoop fs -rm -r command(delete the folder and all its contents recursively)
3. Create a directory in HDFS for this assignment and task using hadoop fs -mkdir -p command
4. Copy the input from local file system to hdfs using hadoop fs -put command
5. Finally verify using hadoop fs -ls command if the input file has been copied into the HDFS



**mapred\_process also takes assignmetn number and task number as input**

1. It executes the JAR from the local file system using hadoop jar command

****

**output.sh also takes assignmetn number and task number as input**

1. It lists all teh files in the output directory for this assignment and task job in HDFS. (hadoop fs -ls)
2. It displays the content of all the output files (hadoop fs -cat)
3. Finally it copies the output files fromHDFS to local file system

Note: The output here is a sequence file which contains data in serializd format. So, when you do a simple hadoop fs -cat, it will display some gibberish. So for such files we have to use hadoop fs -text

