**Cloud Computing for Data Analysis**

**Group Activity 01**

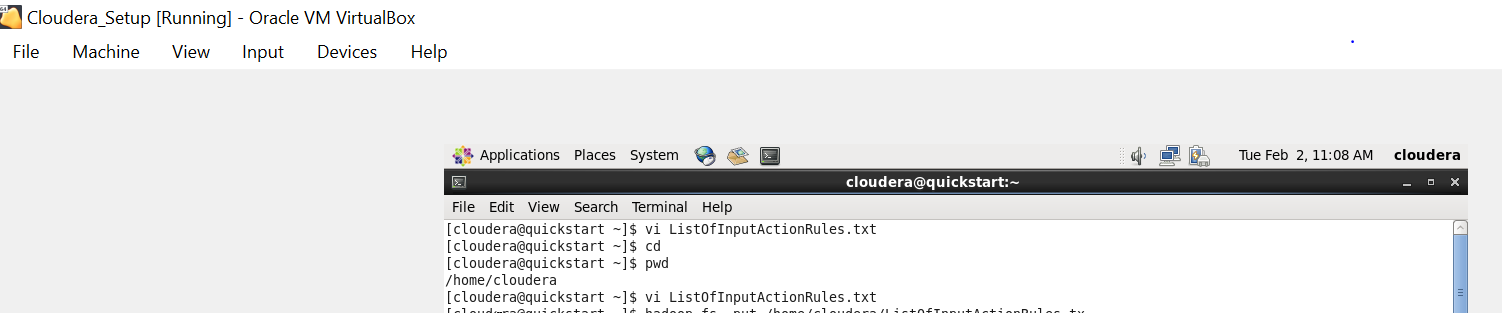
**Group 10: Akhila Sirikonda, Sanjeeva Hitesh Reddy Pereddy, Raj Shah, Kamal Mohammed Adil Shaik, Gopal Sharma, Shriya Sharma, Keerrthana Sridhar**

**Create and Search through a file**

1. Open Cloudera
2. Open the Terminal in Cloudera
3. Create a file named **ListOfInputActionRules.txt** using the command

**vi ListOfInputActionRules.txt**

**Solution:**



1. Copy and paste following lines in **ListOfInputActionRules.txt**

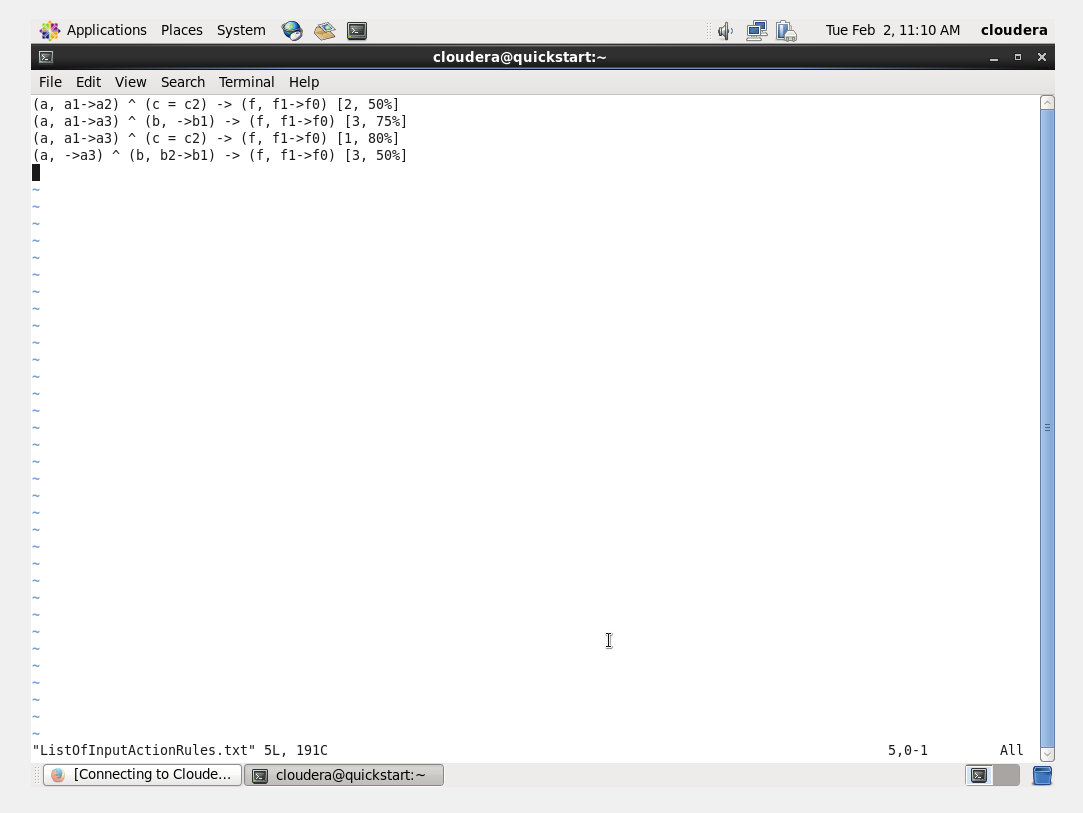
(a, a1->a2) ^ (c = c2) -> (f, f1->f0) [2, 50%]

(a, a1->a3) ^ (b, ->b1) -> (f, f1->f0) [3, 75%]

(a, a1->a3) ^ (c = c2) -> (f, f1->f0) [1, 80%]

(a, ->a3) ^ (b, b2->b1) -> (f, f1->f0) [3, 50%]

**Solution:**



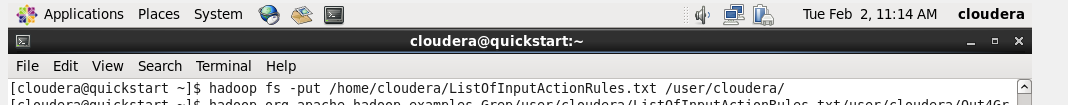
1. Make sure the file is present in **/home/cloudera/** folder
2. Again, in the terminal use the command

**hadoop fs -put /home/cloudera/ListOfInputActionRules.txt /user/admin/** (or)

**hadoop fs -put /home/cloudera/ListOfInputActionRules.txt /user/cloudera/**

to copy your file into HDFS

**Solution:**



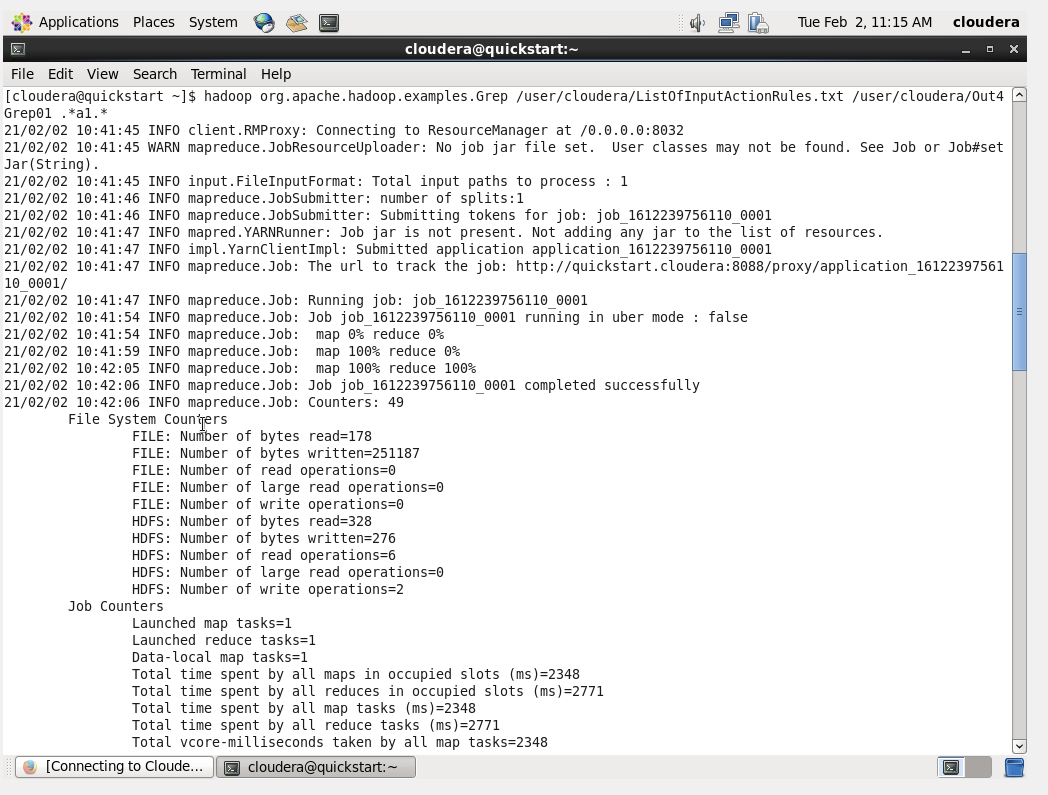
1. See <https://wiki.apache.org/hadoop/Grep> to learn about Grep command
2. Use the command **hadoop org.apache.hadoop.examples.Grep /user/admin/filename.txt /user/admin/Out4Grep01 “.\*a1.\*”** (or)

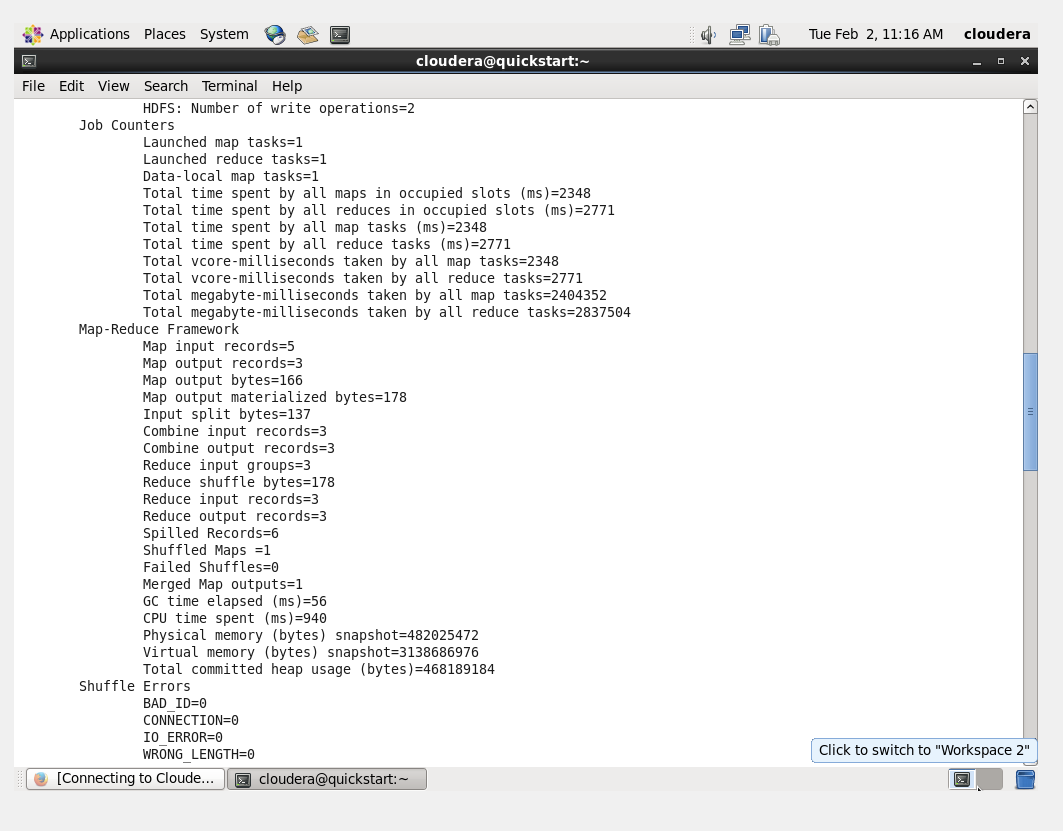
**/user/cloudera/filename.txt /user/cloudera/Out4Grep01 “.\*a1.\*”**

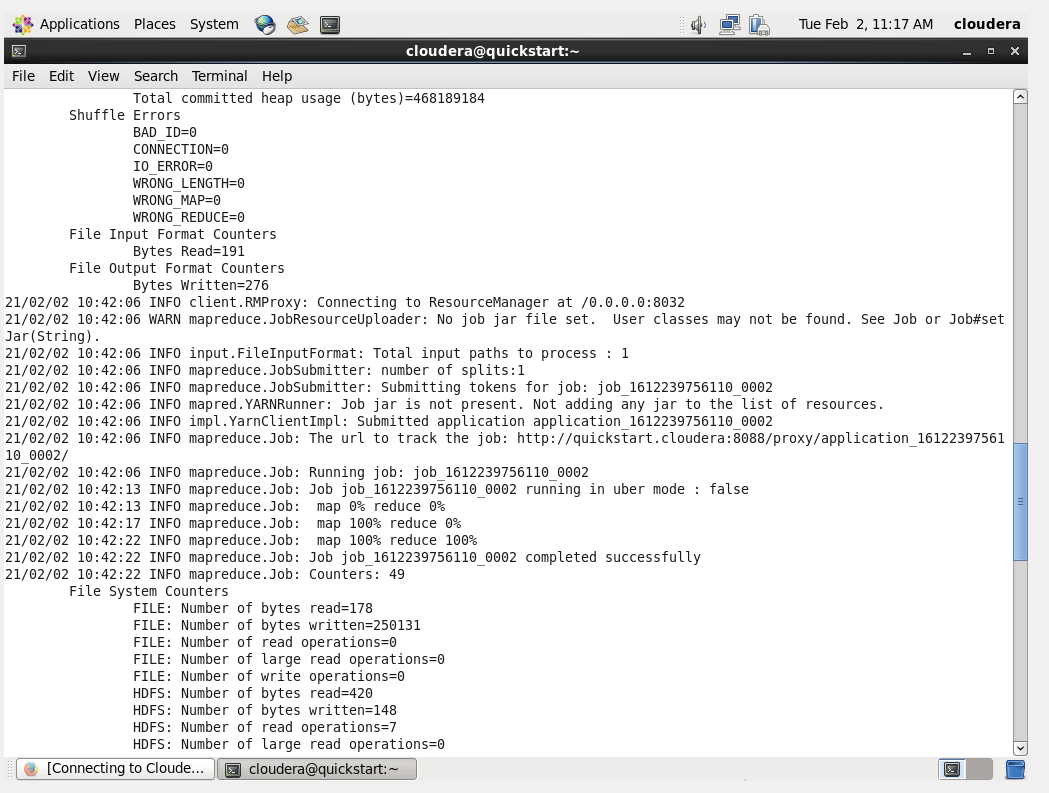
to return all lines of text which contain the word ‘**a1**’

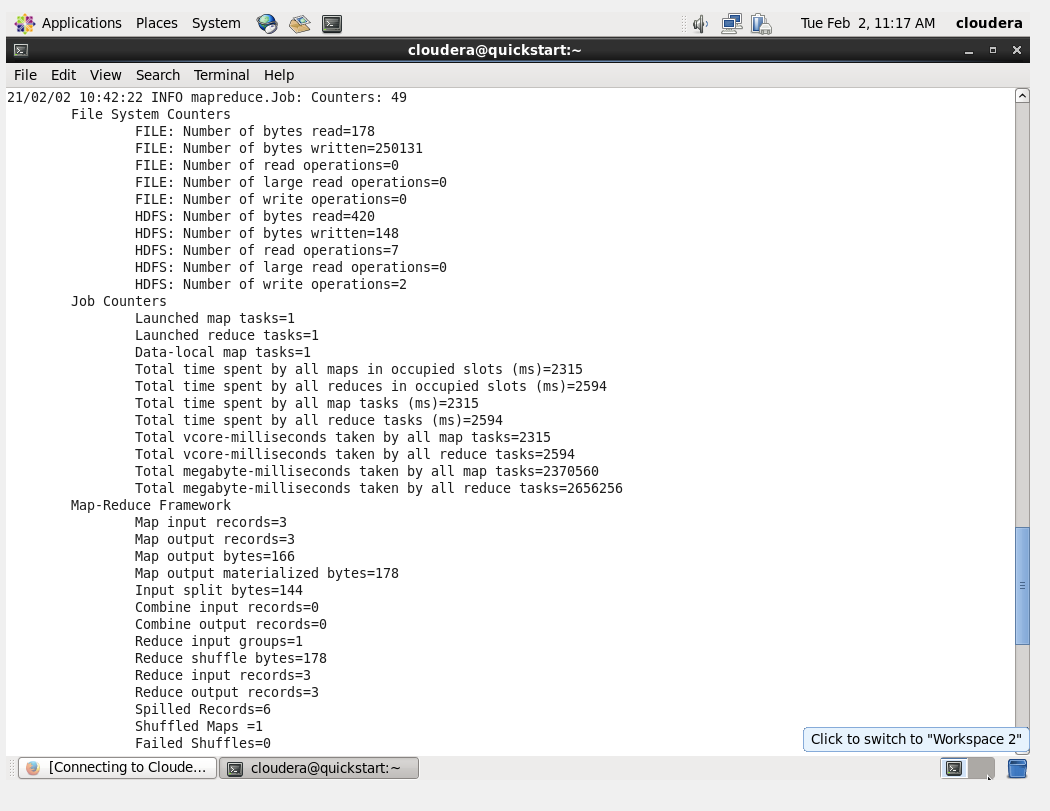
**(NOTE:** You do not need to create separate output folder for your outputs. Hadoop itself creates the output folder matching the given folder name in the given folder**)**

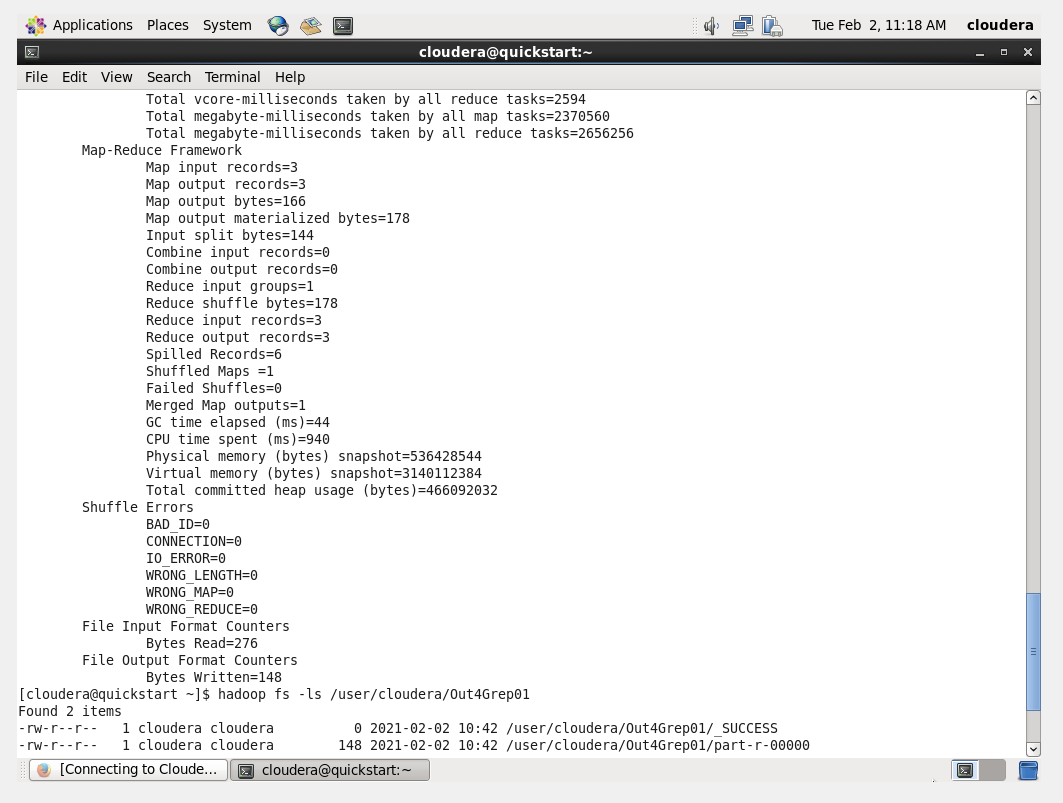
**Solution:**











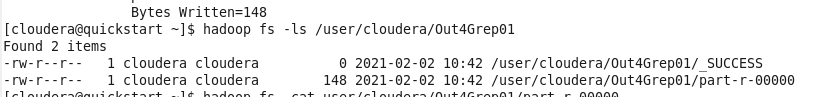
1. List all files from the output path using

**hadoop fs -ls /user/admin/Out4Grep01** (or)

**hadoop fs -ls /user/cloudera/Out4Grep01**

command

**Solution:**

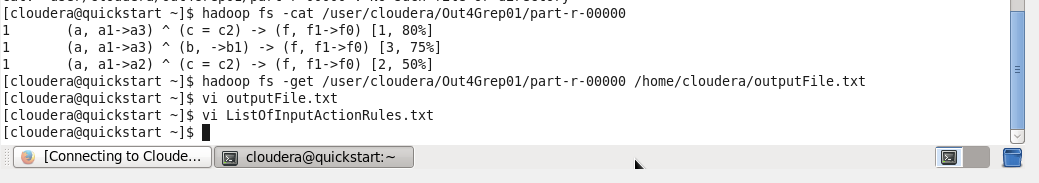


1. Open the file that is named something like **part-r-00000** using

**hadoop fs -cat /user/admin/Out4Grep01/part-r-00000** (or)

**hadoop fs -cat /user/cloudera/Out4Grep01/part-r-00000**

**Solution:**



1. You can take a screenshot of this and upload in Canvas

Hadoop outputs will be without any extensions like **part-r-00000**. You can convert this into **.txt** file while copying it into Cloudera home directory using the following command:

**hadoop fs -get /user/admin/ Out4Grep01/part-r-00000 /home/cloudera/outputFile.txt** (or)

**hadoop fs -get /user/cloudera/ Out4Grep01/part-r-00000 /home/cloudera/outputFile.txt**

**Solution:**

