**Instructions for Hadoop Data Analysis**

CS555: Big Data Computing Lab

horizontal line

# **2nd August 2021**

Problem Statement: **what are the top 5 categories with maximum number of videos uploaded.**

1. create project directory : (youtube)
   1. Browse to home directory

**cd /home/iitp**

* 1. Create project directory

**mkdir youtube**

1. Create source file: (Top5\_categories.java)
   1. Browse into youtube directory

**cd /home/iitp/youtube**

* 1. Create file

**nano Top5\_categories.java**

paste the lines from the source code provided

Note:- **To save file: Press- CTRL + o followed by Enter button**

**To Exit Press:- CTRL + x from nano editor**

1. Create input directory (inputdata) for input files

cd /home/iitp/youtube

**mkdir** **inputdata**

1. Copy the input file (youtubedata.txt) into inputdata folder

**cd /home/iitp/youtube/inputdata**

**nano youtubedata.txt**

paste the lines from the provided input file

1. Start all hadoop services
   1. Browse to hadoop installation sbin sub-directory

**cd /home/iitp/hadoop-2.6.0/sbin**

* 1. start all services

**./start-all.sh**

Note:- Enter password when prompted

1. Create input directory on HDFS
   1. browse to hadoop installation bin folder

**cd /home/iitp/hadoop-2.6.0/bin**

* 1. create directory (**youtube**)

**./hadoop fs -mkdir /youtube**

* 1. create subdirectory (inputdata) inside youtube on HDFS

**./hadoop fs -mkdir /youtube/inputdata**

1. Copy the input text file from local directory to HDFS
   1. browse to hadoop installation bin folder

**cd /home/iitp/hadoop-2.6.0/bin**

* 1. Copy from Local

**./hadoop dfs -put /home/iitp/youtube/inputdata/youtubedata.txt /youtube/inputdata/**

1. Compile the Source Code
   1. export the Hadoop classpath

**export HADOOP\_CLASSPATH=/usr/lib/jvm/java-1.8.0-openjdk-amd64/lib/tools.jar**

* 1. browse to bin folder of hadoop installation

**cd /home/iitp/hadoop-2.6.0/bin**

* 1. Compile

**./hadoop com.sun.tools.javac.Main /home/iitp/youtube/Top5\_categories.java**

1. Create Jar file
   1. Browse to youtube directory on your VM

**cd /home/iitp/youtube**

**jar cf youtube1.jar Top5\_categories\*.class**

1. Running the program
   1. browse to the bin directory of hadoop installation

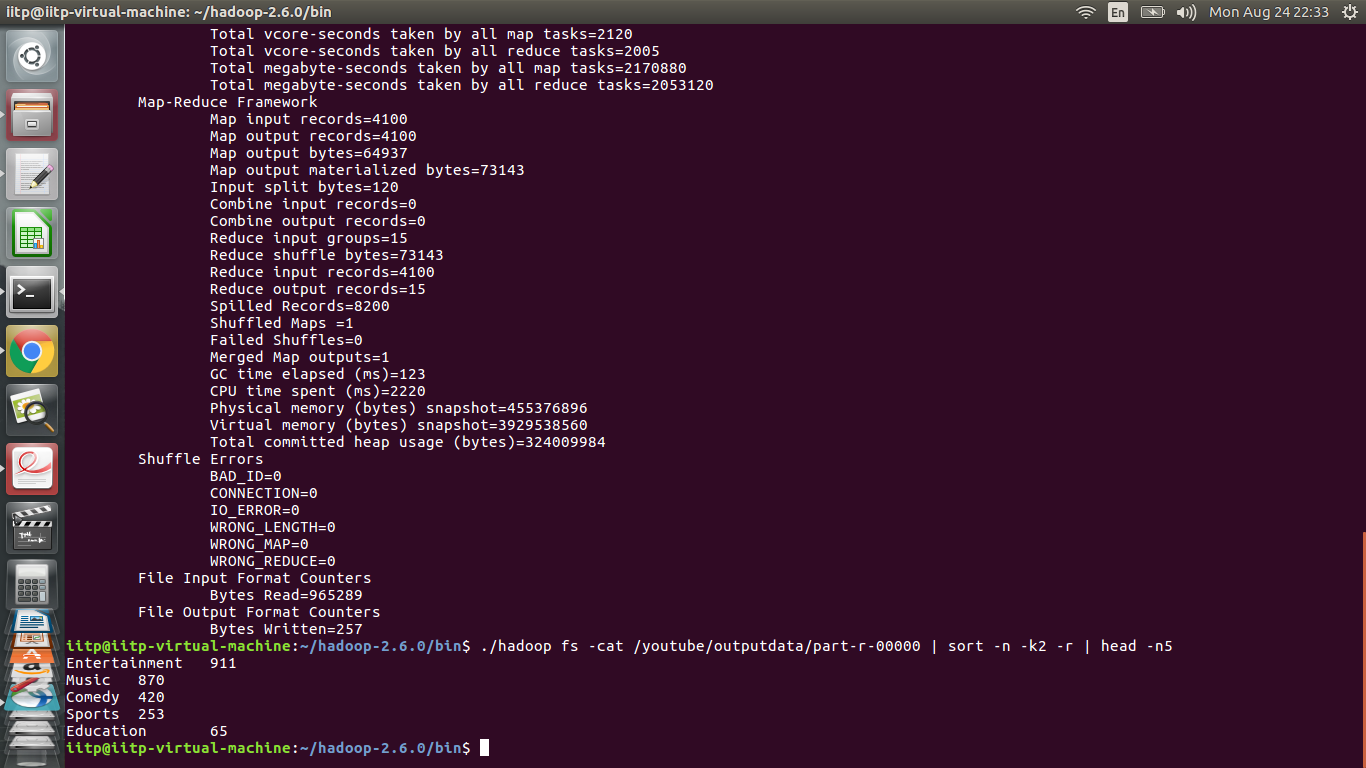
**cd /home/iitp/hadoop-2.6.0/bin**

* 1. Running in terminal

**./hadoop jar /home/iitp/youtube/youtube1.jar Top5\_categories /youtube/inputdata/ /youtube/outputdata**

* 1. Finding outputs

**./hadoop fs -cat /youtube/outputdata/part-r-00000 | sort -n -k2 -r | head -n5**

****