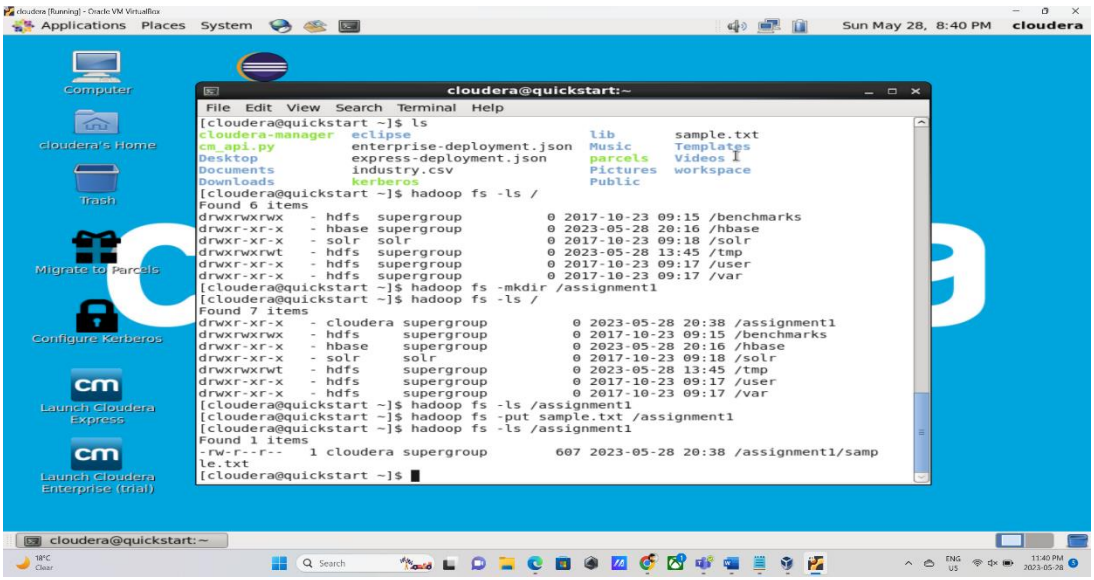
Gupta Bahadur Bhandari

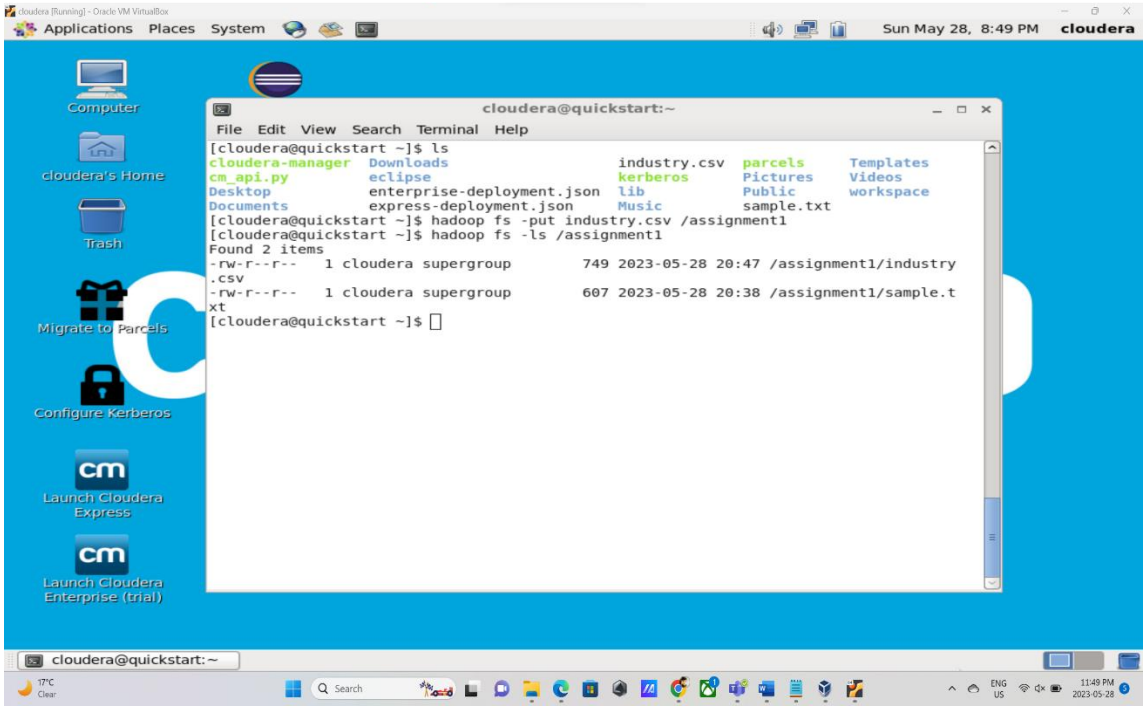
Here I used CLI of CentOS to perform all the tasks.

1. **Importing files to HDFS:**
    a. Importing a text file: In this task, I imported a text file namely 'sample.txt' from local file system to HDFS using Hadoop commands. To do so, I firstly created a directory 'assignment1' inside HDFS using a Hadoop command, **hadoop fs -mkdir /assignment1**. Then, I imported the sample.txt file into the 'assignment1' directory using a command, **hadoop fs -put sample.txt /assignment1**. All the processes followed to perform this task are reflected below in snapshot:

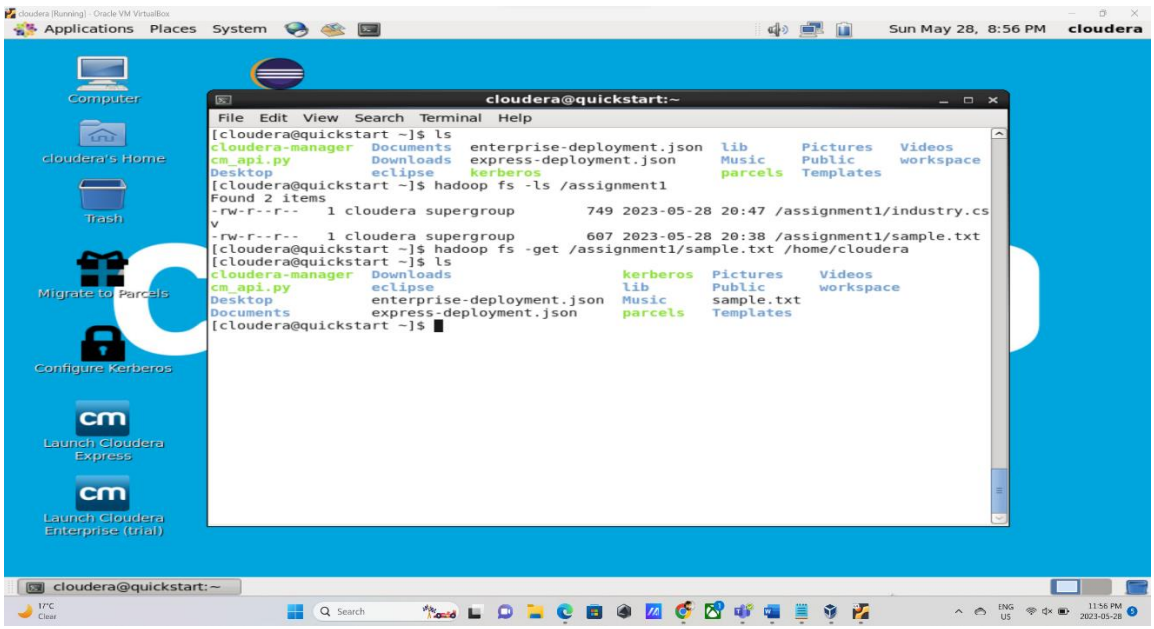


In above snapshot, we can clearly see that there is a text file, sample.txt in the local system and there is no such file in HDFS. Then, I created a directory called assignment1 in the HDFS and finally imported sample.txt file into this directory using a Hadoop command.

b. Importing a csv file: In above snapshot, we can see that there's a csv file namely 'industry.csv' in the local file system. In this task, I used a Hadoop command to import this csv file to HDFS. I imported industry.csv file to the same directory 'assignment1' inside HDFS using a Hadoop command, **hadoop fs -put industry.csv /assignment1**. The snapshot of entire process is shown below:

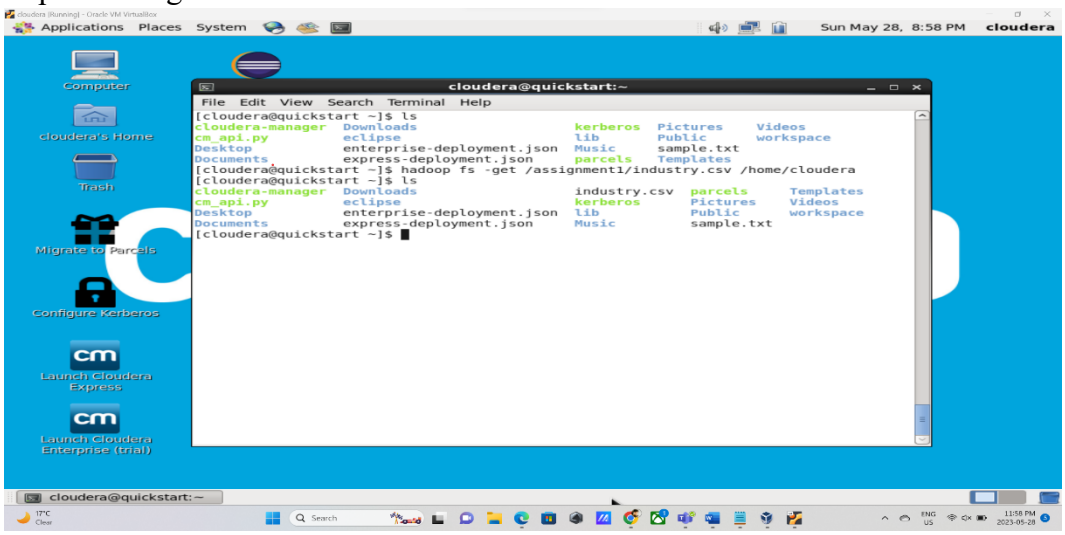


2. **Exporting files from HDFS:**
   a. Exporting a text file: For this task, I exported same sample.txt file which I imported on task-1 from 'assignment1' directory (in HDFS) to home directory of local file system. To do this exporting, the command I used is, **hadoop fs -get /assignment1/sample.txt /home/cloudera**. The snapshot of this entire process is shown below:



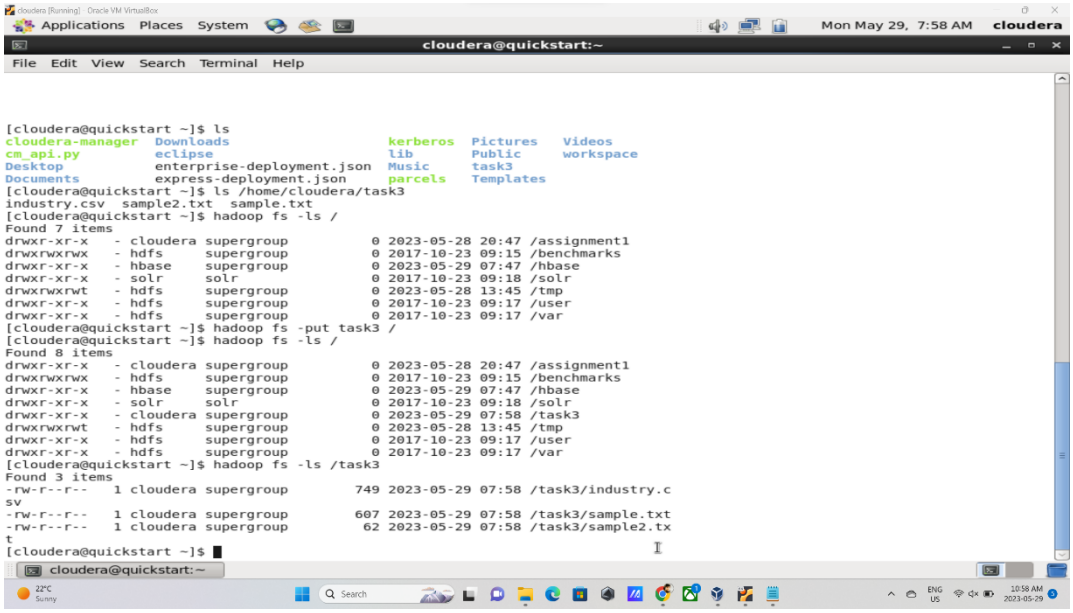
In above snapshot, we can clearly see that there was not any sample.txt file in local file system. After exporting the file from HDFS, the file was reflected in local file system.

b. Exporting a csv file: In previous snapshot, we can see a csv file 'industry.csv' in the assignment1 directory of HDFS. Now, in this task, I want to export this industry.csv file from HDFS to home directory of local file system. To perform this task successfully, I executed a Hadoop command, i.e., **hadoop fs -get /assignment1/industry.csv /home/cloudera**. After, executing the command, I was successfully able to export the csv file from HDFS to local file system. A snapshot of performing this task is shown below:



3. Advanced Import/Export Operations:

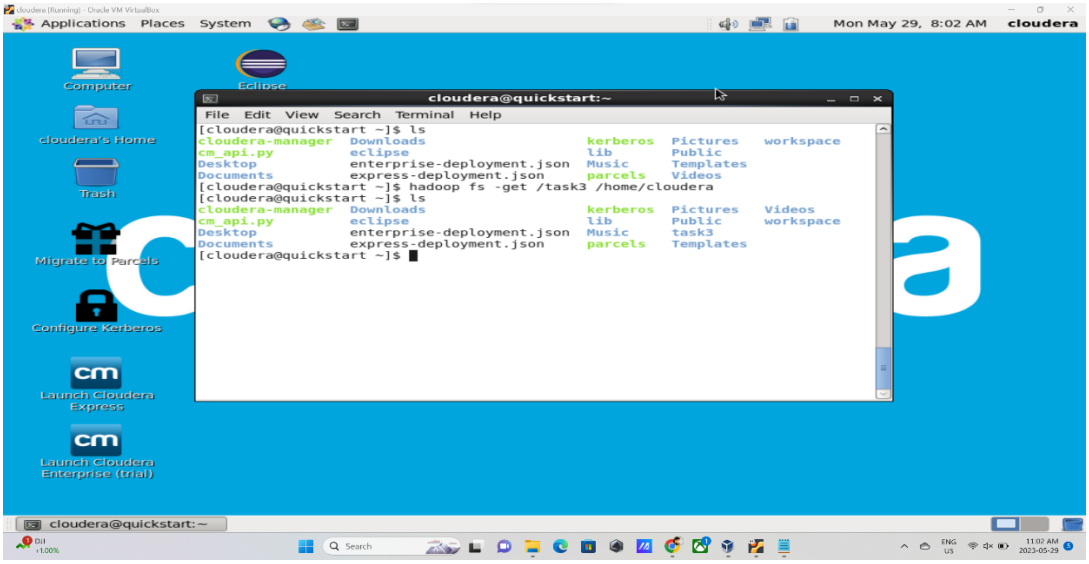
a. Importing a folder containing multiple files:



In above snapshot, we can see that there's a directory 'task3' containing 3 files in home directory of local file system. To import this task3 directory into HDFS, I used a Hadoop command **hadoop fs -put task3 /**. Doing this, I successfully

imported a directory containing multiple files (task3) into HDFS from local file system.

b. Exporting a folder containing multiple files: In previous task, I imported a folder containing 3 files namely task3 into HDFS from local file system. In this task, I exported the same folder to the local file system using a Hadoop command, **hadoop fs -get /task3 /home/cloudera**. This Hadoop command reflected the folder from home directory of HDFS to the home directory of local file system. The entire process is shown below in snapshot:



In above snapshot, we can see that there was not any directory called 'task3' in home directory of local file system. However, the directory was copied to the local file system after it was exported from HDFS using a Hadoop command.

Conclusion:

While importing and exporting any files and folders from HDFS, there are several pros and cons to consider. Talking about pros; HDFS enables rapid transfer of data between data nodes. HDFS divides a task among different nodes in a cluster, so there is high fault tolerance while dealing a work in HDFS. Furthermore, HDFS is scalable and can store huge amounts of data.

On the other hand, there are few limitations of HDFS too. Firstly, HDFS cannot be used for real-time data processing. For small data, HDFS might not be cost effective.