



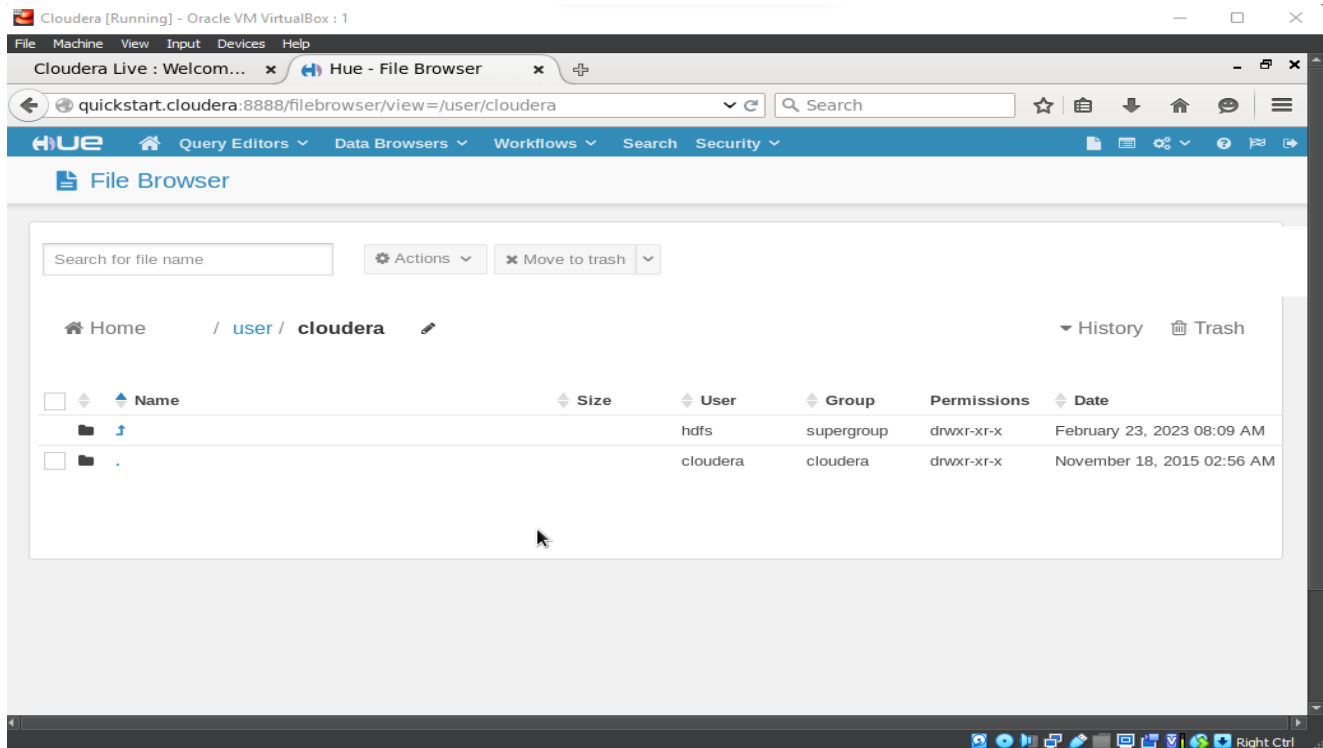
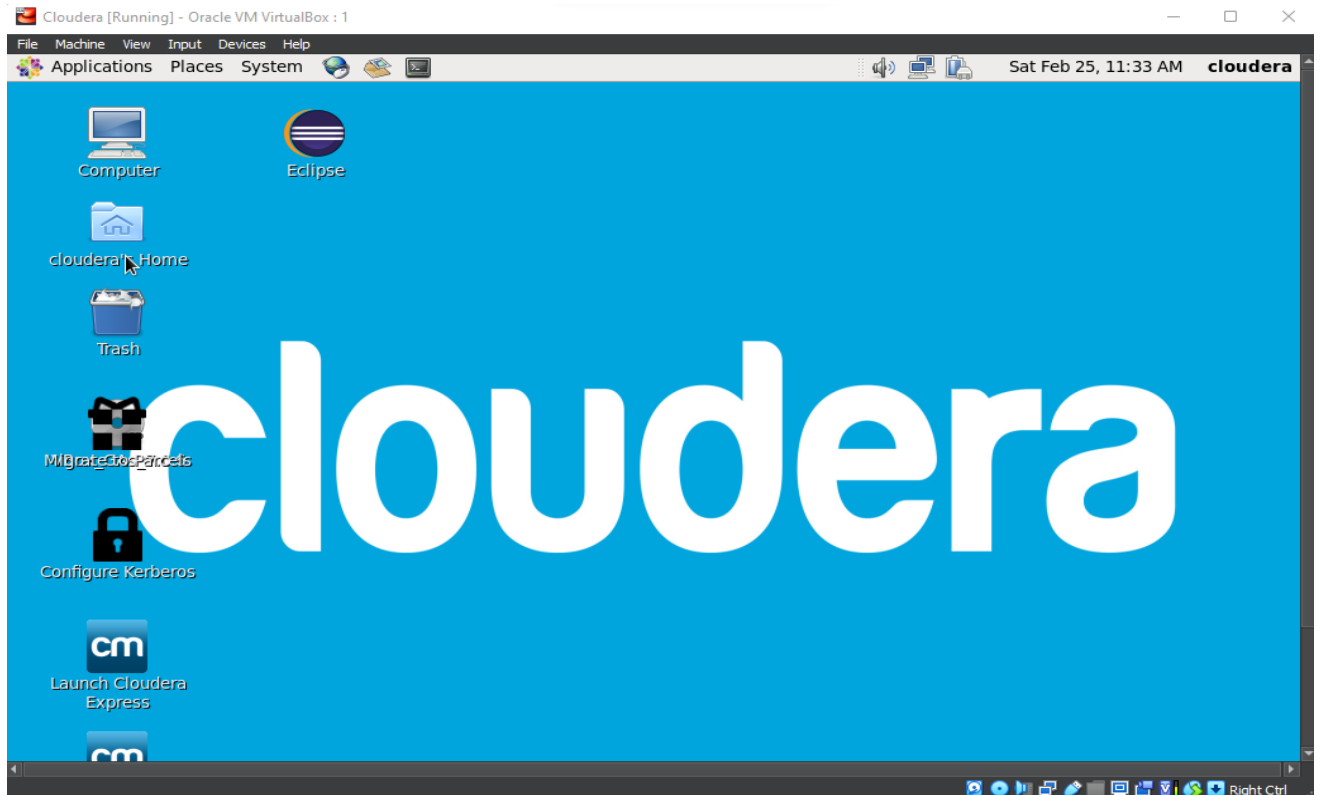
CSE488 (Section 1)
[Spring 2023]

Lab Assignment Submission Report

Assignment Title: Setting up Cloudera VM and
Executing Word Count Problem using
MapReduce

Submitted by:
Sazzad Hossen
2019-1-60-063

Setting up Cloudera VM



Executing Word Count Problem using MapReduce

```
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ ls
cloudera-manager Downloads lib Templates
cm_api.py eclipse Music Videos
data1.txt enterprise-deployment.json parcels WordCount.jar
Desktop express-deployment.json Pictures workspace
Documents kerberos Public
[cloudera@quickstart ~]$ pwd
/home/cloudera
[cloudera@quickstart ~]$ cat > /home/cloudera/data2.txt
sazzad mango xx sazzad xx mango tea xx tea mango xx
yy
gg sazzad tea
yy gg
yy
gg
^Z
[1]+ Stopped cat > /home/cloudera/data2.txt
[cloudera@quickstart ~]$ ls
cloudera-manager Documents kerberos Public
cm_api.py Downloads lib Templates
data1.txt eclipse Music Videos
data2.txt enterprise-deployment.json parcels WordCount.jar
Desktop express-deployment.json Pictures workspace

[cloudera@quickstart ~]$ hdfs dfs -ls
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx - hdfs supergroup 0 2015-11-18 02:57 /benchmarks
drwxr-xr-x - hbase supergroup 0 2023-02-25 11:07 /hbase
drwxr-xr-x - cloudera supergroup 0 2023-02-25 08:56 /inputfolder1
drwxr-xr-x - cloudera supergroup 0 2023-02-25 09:05 /output1
drwxrwxrwt - hdfs supergroup 0 2023-02-25 08:48 /tmp
drwxr-xr-x - hdfs supergroup 0 2023-02-23 08:09 /user
drwxr-xr-x - hdfs supergroup 0 2015-11-18 03:00 /var
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/data2.txt /inputfolder1/
put: `/inputfolder1/': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/data2.txt /inputfolder/
put: `/inputfolder/': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/data2.txt /inputfolder1/
[cloudera@quickstart ~]$ hdfs dfs -cat /inputfolder1/data2.txt
sazzad mango xx sazzad xx mango tea xx tea mango xx
yy
gg sazzad tea
yy gg
yy
gg
```

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCount.jar wordcount /inputfolder1/data2.txt /output2  
Exception in thread "main" java.lang.ClassNotFoundException: wordcount  
    at java.net.URLClassLoader$1.run(URLClassLoader.java:366)  
    at java.net.URLClassLoader$1.run(URLClassLoader.java:355)  
    at java.security.AccessController.doPrivileged(Native Method)  
    at java.net.URLClassLoader.findClass(URLClassLoader.java:354)  
    at java.lang.ClassLoader.loadClass(ClassLoader.java:425)  
    at java.lang.ClassLoader.loadClass(ClassLoader.java:358)  
    at java.lang.Class.forName0(Native Method)  
    at java.lang.Class.forName(Class.java:270)  
    at org.apache.hadoop.util.RunJar.run(RunJar.java:214)  
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)  
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCount.jar WordCount /inputfolder1/data2.txt /output2  
23/02/25 11:23:02 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032  
23/02/25 11:23:03 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
23/02/25 11:23:04 INFO input.FileInputFormat: Total input paths to process : 1  
23/02/25 11:23:04 INFO mapreduce.JobSubmitter: number of splits:1  
23/02/25 11:23:04 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1677352011904_0001  
23/02/25 11:23:05 INFO impl.YarnClientImpl: Submitted application application_1677352011904_0001
```

```
File Edit View Search Terminal Help  
23/02/25 11:24:06 INFO mapreduce.Job: Counters: 49  
    File System Counters  
        FILE: Number of bytes read=123  
        FILE: Number of bytes written=222953  
        FILE: Number of read operations=0  
        FILE: Number of large read operations=0  
        FILE: Number of write operations=0  
        HDFS: Number of bytes read=203  
        HDFS: Number of bytes written=88  
        HDFS: Number of read operations=6  
        HDFS: Number of large read operations=0  
        HDFS: Number of write operations=2  
    Job Counters  
        Launched map tasks=1  
        Launched reduce tasks=1  
        Data-local map tasks=1  
        Total time spent by all maps in occupied slots (ms)=14078  
        Total time spent by all reduces in occupied slots (ms)=21888  
        Total time spent by all map tasks (ms)=14078  
        Total time spent by all reduce tasks (ms)=21888  
        Total vcore-seconds taken by all map tasks=14078  
        Total vcore-seconds taken by all reduce tasks=21888  
        Total megabyte-seconds taken by all map tasks=14415872  
        Total megabyte-seconds taken by all reduce tasks=22413312  
    Map-Reduce Framework  
        Map input records=6
```

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=123  
Reduce input records=6  
Reduce output records=5  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=314  
CPU time spent (ms)=2830  
Physical memory (bytes) snapshot=394002432  
Virtual memory (bytes) snapshot=3124404224  
Total committed heap usage (bytes)=343408640  
Shuffle Errors  
BAD_ID=0  
CONNECTION=0  
IO_ERROR=0  
WRONG_LENGTH=0  
WRONG_MAP=0  
WRONG_REDUCE=0  
File Input Format Counters  
Bytes Read=84  
File Output Format Counters  
Bytes Written=88  
[cloudera@quickstart ~]$  
File Edit View Search Terminal Help  
Physical memory (bytes) snapshot=394002432  
Virtual memory (bytes) snapshot=3124404224  
Total committed heap usage (bytes)=343408640  
Shuffle Errors  
BAD_ID=0  
CONNECTION=0  
IO_ERROR=0  
WRONG_LENGTH=0  
WRONG_MAP=0  
WRONG_REDUCE=0  
File Input Format Counters  
Bytes Read=84  
File Output Format Counters  
Bytes Written=88  
[cloudera@quickstart ~]$ hdfs dfs -ls /output2  
Found 2 items  
-rw-r--r-- 1 cloudera supergroup 0 2023-02-25 11:24 /output2/_SUCCESS  
-rw-r--r-- 1 cloudera supergroup 88 2023-02-25 11:24 /output2/part-r-000000  
[cloudera@quickstart ~]$ hdfs dfs -cat /output2/part-r-000000  
GG 1  
GG SAZZAD TEA 1  
SAZZAD MANGO XX SAZZAD XX MANGO TEA XX TEA MANGO XX 1  
YY 2  
YY GG 1  
[cloudera@quickstart ~]$
```

Learning Outcomes

In this lab learned how to install Hadoop on my virtual box and learned how to solve word count problem. Also learned how to execute a problem with MapReduce.