Harish Puvvada - hp1047

Docker commands, Hue Screenshots

Below are screenshots of my docker commands in the same order I executed them.

docker run --hostname=quickstart.cloudera --privileged=true -t -i -d -p 8888:8888 -p 8000:80 -P -v /Users/harishpuvvada/Desktop/PBDA/Hadoop:/Hadoop --name hadoop cloudera/quickstart /usr/bin/docker-quickstart

docker exec -it hadoop bash

hadoop fs -mkdir /Input

hadoop fs -mkdir /Harish

hadoop fs -copyFromLocal /Hadoop/IrishPennyJournal.txt /input

hadoop jar usr/lib/hadoop-mapreduce/hadoop-streaming.jar -file /Hadoop/mapper.py -mapper /Hadoop/mapper.py -file /Hadoop/reducer.py -reducer /Hadoop/reducer.py -input /Input/IrishPennyJournal.txt -output /Harish/PBDA

Docker run command, exec command

```
[172-16-44-202:~ harishpuvvada$ docker run --hostname=quickstart.cloudera --privileged=true -t -i -d -p 8888:8888 -p 8000:80 -P -v /Users/hari shpuvvada/Desktop/PBDA/Hadoop:/Hadoop --name hadoop cloudera/quickstart /usr/bin/docker-quickstart c4b8f4b328d6b6a6d278cfcea4c82cd61b5e0e2a57f7e617b382ba36ba67cde4
[172-16-44-202:~ harishpuvvada$ docker exec -it hadoop bash
[[root@quickstart /]#
[[root@quickstart /]#
[[root@quickstart /]#
[[root@quickstart /]#
[[root@quickstart /]#
[[root@quickstart /]#
```

Before making Input and Harish (Output) directories

```
[[root@quickstart /]# hadoop fs -ls /
Found 5 items
drwxrwxrwx

    hdfs supergroup

                                        0 2016-04-06 02:26 /benchmarks
drwxr-xr-x - hbase supergroup
                                        0 2017-10-12 21:17 /hbase
drwxrwxrwt - hdfs supergroup
                                        0 2017-10-12 21:17 /tmp
                                        0 2016-04-06 02:27 /user
drwxr-xr-x - hdfs
                    supergroup
            - hdfs
                                        0 2016-04-06 02:27 /var
drwxr-xr-x
                   supergroup
```

mkdir Input

mkdir Harish

```
[[root@quickstart /]#
[[root@quickstart /]#
[[root@quickstart /]#
[[root@quickstart /]# hadoop fs -mkdir /Harish
[[root@quickstart /]#
[[root@quickstart /]#
```

After making Input and Harish(Output) directories

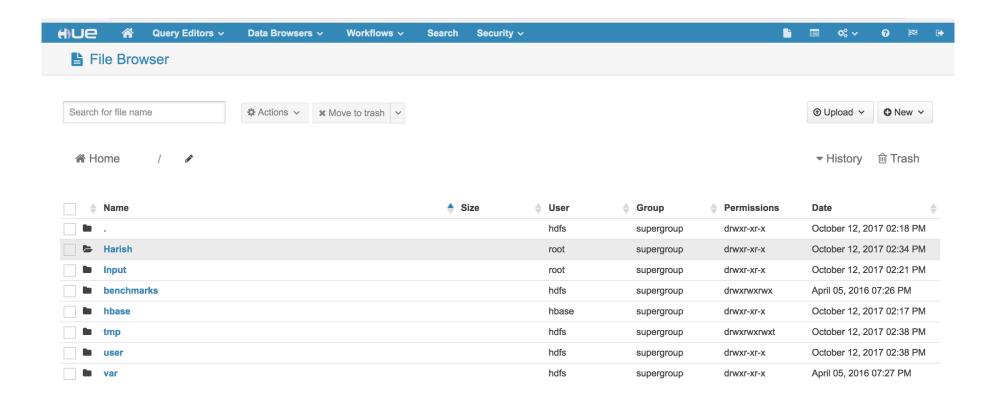
```
[[root@quickstart /]#
[[root@quickstart /]#
[root@quickstart /]#
[[root@quickstart /]# hadoop fs -ls /
Found 7 items
                                        0 2017-10-12 21:18 /Harish
drwxr-xr-x
             - root
                    supergroup
drwxr-xr-x - root supergroup
                                        0 2017-10-12 21:18 /Input
                                        0 2016-04-06 02:26 /benchmarks
drwxrwxrwx - hdfs supergroup
                                        0 2017-10-12 21:17 /hbase
drwxr-xr-x - hbase supergroup
drwxrwxrwt - hdfs supergroup
                                        0 2017-10-12 21:17 /tmp
drwxr-xr-x - hdfs supergroup
                                        0 2016-04-06 02:27 /user
             - hdfs supergroup
drwxr-xr-x
                                        0 2016-04-06 02:27 /var
[root@quickstart /]#
```

hadoop jar command

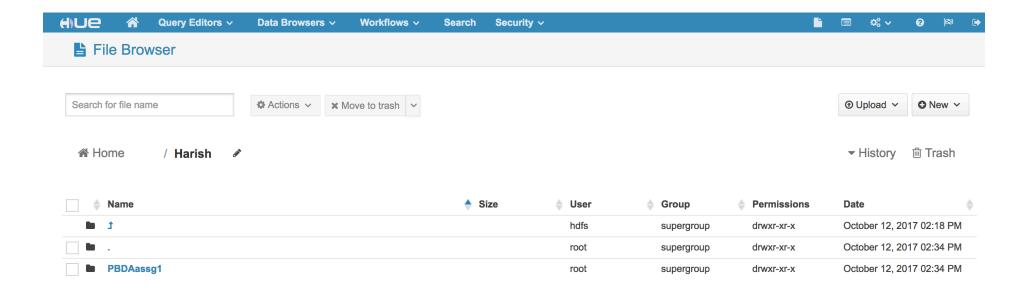
```
root@quickstart /]# hadoop jar usr/lib/hadoop-mapreduce/hadoop-streaming.jar -file /Hadoop/mapper.py -mapper /Hadoop/mapper.py -file /Hadoop
 reducer.py -reducer /Hadoop/reducer.py -input /Input/IrishPennyJournal.txt -output /Harish/PBDAassg1
17/10/12 21:34:25 WARN streaming. Stream Job: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/Hadoop/mapper.py, /Hadoop/reducer.py] [/usr/jars/hadoop-streaming-2.6.0-cdh5.7.0.jar] /tmp/streamjob1493869899952881805.jar
tmoDir=null
17/10/12 21:34:26 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/10/12 21:34:26 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/10/12 21:34:27 INFO mapred.FileInputFormat: Total input paths to process : 1
17/10/12 21:34:28 INFO mapreduce.JobSubmitter: number of splits:2
17/10/12 21:34:28 INFO mapreduce. Job Submitter: Submitting tokens for job: job 1507843011869 0002
17/10/12 21:34:28 INFO impl.YarnClientImpl: Submitted application application 1507843011869 0002
17/10/12 21:34:28 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application 1507843011869 0002/
17/10/12 21:34:28 INFO mapreduce. Job: Running job: job 1507843011869 0002
17/10/12 21:34:39 INFO mapreduce.Job: Job job 1507843011869 0002 running in uber mode : false
17/10/12 21:34:39 INFO mapreduce.Job: map 0% reduce 0%
17/10/12 21:34:48 INFO mapreduce.Job: map 100% reduce 0%
17/10/12 21:34:56 INFO mapreduce.Job: map 100% reduce 100%
17/10/12 21:34:57 INFO mapreduce.Job: Job job_1507843011869_0002 completed successfully
17/10/12 21:34:58 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=212301
                FILE: Number of bytes written=775469
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=90455
                HDFS: Number of bytes written=135605
                HDFS: Number of read operations=9
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=1
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=13466
                Total time spent by all reduces in occupied slots (ms)=5218
                Total time spent by all map tasks (ms)=13466
                Total time spent by all reduce tasks (ms)=5218
                Total vcore-seconds taken by all map tasks=13466
                Total vcore-seconds taken by all reduce tasks=5218
```

```
Job Counters
               Launched map tasks=2
                Launched reduce tasks=1
               Data-local map tasks=2
               Total time spent by all maps in occupied slots (ms)=13466
                Total time spent by all reduces in occupied slots (ms)=5218
               Total time spent by all map tasks (ms)=13466
               Total time spent by all reduce tasks (ms)=5218
               Total vcore-seconds taken by all map tasks=13466
               Total vcore-seconds taken by all reduce tasks=5218
               Total megabyte-seconds taken by all map tasks=13789184
               Total megabyte-seconds taken by all reduce tasks=5343232
        Map-Reduce Framework
               Map input records=1607
                Map output records=14691
               Map output bytes=182913
               Map output materialized bytes=212307
                Input split bytes=222
               Combine input records=0
               Combine output records=0
               Reduce input groups=10331
               Reduce shuffle bytes=212307
               Reduce input records=14691
               Reduce output records=10331
               Spilled Records=29382
               Shuffled Maps =2
               Failed Shuffles=0
               Merged Map outputs=2
               GC time elapsed (ms)=144
               CPU time spent (ms)=4200
               Physical memory (bytes) snapshot=710103040
               Virtual memory (bytes) snapshot=4103852032
               Total committed heap usage (bytes)=558366720
        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=90233
        File Output Format Counters
               Bytes Written=135605
17/10/12 21:34:58 INFO streaming.StreamJob: Output directory: /Harish/PBDAassg1
[root@quickstart /]#
```

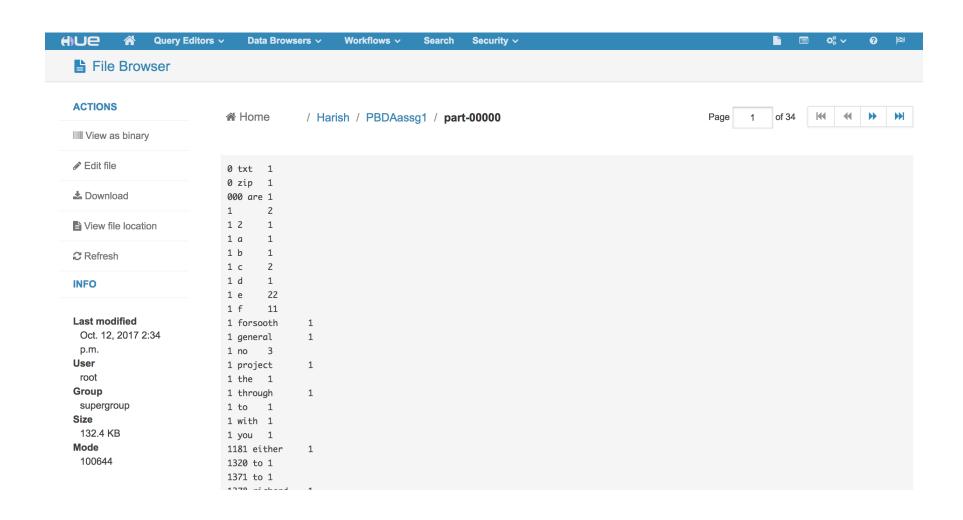
Hue file browser



Hue file browser - /Harish/



Hue – final Output



Hue Job browser (Note: job browser was not working in local machine. Took this screenshot from NYU-HPC – babar.es.its)

