

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

```
hadoop
[172-16-44-202:~ harishpuvada$ docker run --hostname=quickstart.cloudera --privileged=true -t -i -d -p 8888:8888 -p 8000:80 -P -v /Users/hari]
shpuvada/Desktop/PBDA/Hadoop:/Hadoop --name hadoop cloudera/quickstart /usr/bin/docker-quickstart
c4b8f4b328d6b6a6d278cfcea4c82cd61b5e0e2a57f7e617b382ba36ba67cde4
[172-16-44-202:~ harishpuvada$ docker exec -it hadoop bash
[
[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  
drwxr-xr-x   - hdfs  supergroup          0 2016-04-06 02:27 /user  
drwxr-xr-x   - hdfs  supergroup          0 2016-04-06 02:27 /var
```

mkdir Input

```
drwxr-xr-x  - hdfs  supergroup          0 2016-04-06 02:27 /var  
[[root@quickstart ~]# hadoop fs -mkdir /Input/  
[[root@quickstart ~]# hadoop fs -ls /  
Found 6 items
```

mkdir Harish

```
[root@quickstart ~]#  
[root@quickstart ~]#  
[root@quickstart ~]# hadoop fs -mkdir /Harish  
[root@quickstart ~]#  
[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  
drwxr-xr-x   - root   supergroup          0 2017-10-12 21:18 /Harish  
drwxr-xr-x   - root   supergroup          0 2017-10-12 21:18 /Input  
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  
drwxr-xr-x   - hdfs   supergroup          0 2016-04-06 02:27 /user  
drwxr-xr-x   - hdfs   supergroup          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.StreamJob: -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
tmpDir=null
17/10/12 21:34:26 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/10/12 21:34:26 INFO client.RMPProxy: 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.JobSubmitter: 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

The screenshot displays the Hue File Browser interface. At the top is a blue navigation bar with the Hue logo and menu items: Query Editors, Data Browsers, Workflows, Search, and Security. Below this is a light gray header with the title 'File Browser'. The main area contains a search bar labeled 'Search for file name', an 'Actions' dropdown menu, a 'Move to trash' button, an 'Upload' button, and a 'New' button. A breadcrumb trail shows 'Home' and a file icon. On the right, there are links for 'History' and 'Trash'. The central part of the interface is a table listing files and folders. The table has columns for Name, Size, User, Group, Permissions, and Date. The 'Harish' folder is currently selected and highlighted in gray.








<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	.		hdfs	supergroup	drwxr-xr-x	October 12, 2017 02:18 PM
<input checked="" type="checkbox"/>	Harish		root	supergroup	drwxr-xr-x	October 12, 2017 02:34 PM
<input type="checkbox"/>	Input		root	supergroup	drwxr-xr-x	October 12, 2017 02:21 PM
<input type="checkbox"/>	benchmarks		hdfs	supergroup	drwxrwxrwx	April 05, 2016 07:26 PM
<input type="checkbox"/>	hbase		hbase	supergroup	drwxr-xr-x	October 12, 2017 02:17 PM
<input type="checkbox"/>	tmp		hdfs	supergroup	drwxrwxrwt	October 12, 2017 02:38 PM
<input type="checkbox"/>	user		hdfs	supergroup	drwxr-xr-x	October 12, 2017 02:38 PM
<input type="checkbox"/>	var		hdfs	supergroup	drwxr-xr-x	April 05, 2016 07:27 PM


Hue file browser - /Harish/

The screenshot displays the Hue File Browser interface. At the top is a blue navigation bar with the Hue logo and menu items: Query Editors, Data Browsers, Workflows, Search, and Security. Below this is a light gray header bar labeled "File Browser". The main content area features a search bar with the placeholder "Search for file name", an "Actions" dropdown menu, a "Move to trash" button, an "Upload" button, and a "New" button. The breadcrumb path shows "Home / Harish" with an edit icon. To the right are links for "History" and "Trash". A table lists the contents of the /Harish/ directory, including a parent directory icon, a file named "PBDAassg1", and a file named ".".


<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		hdfs	supergroup	drwxr-xr-x	October 12, 2017 02:18 PM
<input type="checkbox"/>	.		root	supergroup	drwxr-xr-x	October 12, 2017 02:34 PM
<input type="checkbox"/>	PBDAassg1		root	supergroup	drwxr-xr-x	October 12, 2017 02:34 PM


Hue – final Output


  [Query Editors](#) [Data Browsers](#) [Workflows](#) [Search](#) [Security](#)     


 **File Browser**


ACTIONS

 View as binary

 Edit file

 Download

 View file location

 Refresh

INFO


Last modified
Oct. 12, 2017 2:34 p.m.

User
root





Group
supergroup

Size
132.4 KB

Mode
100644








 Home


/ [Harish](#) / [PBDAassg1](#) / **part-00000**

Page of 34    

0	txt	1
0	zip	1
000	are	1
1		2
1	2	1
1	a	1
1	b	1
1	c	2
1	d	1
1	e	22
1	f	11
1	forsooth	1
1	general	1
1	no	3
1	project	1
1	the	1
1	through	1
1	to	1
1	with	1
1	you	1
1181	either	1
1320	to	1
1371	to	1
1378	to	1

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

  Query Editors ▾ Data Browsers ▾ Workflows ▾   hp1047   


 Job Browser

Username

Text

SucceededRunningFailedKilled

in the last days ▾

Logs	ID	Name	Application Type	Status	User	Maps	Reduces	Queue	Priority	Duration	Submitted	
	1504637662637_29145	streamjob5142579752103565859.jar	MAPREDUCE	SUCCEEDED	hp1047	100%	100%	root.users.hp1047	N/A	14s	10/12/17 18:14:34	

Showing 1 to 1 of 1 entries

← Previous

1

Next →