

# IST769 Homework Submission

Name: **Sathish Kumar Rajendiran**

SUID: **666555028**

Email: [srajendi@syr.edu](mailto:srajendi@syr.edu)

Date Due: **08/10/2021**

Date Due: **HDFS, MapReduce and YARN**

Homework #:5

## Exercise(s):

1. Upload all the documents in [datasets/text](#) into a folder called text in HDFS. What HDFS command must you run to verify the files are there after they are uploaded? Your answer should consist of the command you typed to complete the task.

### Solution:

--list available folders

**hdfs dfs -ls**

--create a folder for "text"

**hdfs dfs -mkdir text**

--upload \*.txt files into "text" folder from datasets/text folder

**hdfs dfs -put datasets/text/\*.txt text/**

--list available files

**hdfs dfs -ls text**

### Evidence:

```
[cloudera@quickstart ~]$ hdfs dfs -ls
[cloudera@quickstart ~]$ hdfs dfs -mkdir text
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:27 text
[cloudera@quickstart ~]$ hdfs dfs -put datasets/text/*.txt text/
[cloudera@quickstart ~]$ hdfs dfs -ls text
Found 7 items
-rw-r--r--  1 cloudera cloudera  35731 2021-08-06 06:34 text/2016-state-of-the-union.txt
-rw-r--r--  1 cloudera cloudera  27470 2021-08-06 06:34 text/constitution.txt
-rw-r--r--  1 cloudera cloudera 1951218 2021-08-06 06:34 text/english-words.txt
-rw-r--r--  1 cloudera cloudera 17641 2021-08-06 06:34 text/gnu-gpl3-license.txt
-rw-r--r--  1 cloudera cloudera  96536 2021-08-06 06:34 text/mbox-short.txt
-rw-r--r--  1 cloudera cloudera     327 2021-08-06 06:34 text/preamble.txt
-rw-r--r--  1 cloudera cloudera 15398 2021-08-06 06:34 text/zork1-walkthru.txt
```

2. In this part you will upload the [clickstream](#) dataset to HDFS. Specifically, create a clickstream folder in HDFS, then create a logs and iplookup folder inside the clickstream folder. Upload all the \*.log files from the datasets/clickstream local folder into clickstream/logs in HDFS. Upload the ip\_lookup.csv file from the same folder into clickstream/iplookup on HDFS. Verify the files are there. Your answer should consist of the commands you typed to complete the task.

**Solution:**

```
--create a folder for "clickstream"
hdfs dfs -mkdir clickstream
--create sub folders for "clickstream" as "logs" and "iplookup"
hdfs dfs -mkdir clickstream/logs
hdfs dfs -mkdir clickstream/iplookup
--upload *.log files into "logs" folder from datasets/clickstream folder
hdfs dfs -put datasets/clickstream/*.log clickstream /logs/
--upload *.csv files into "iplookup" folder from datasets/clickstream folder
hdfs dfs -put datasets/clickstream/*.csv clickstream /iplookup/
--list available files
hdfs dfs -ls clickstream/logs
hdfs dfs -ls clickstream/iplookup
```

**Evidence:**

```
cloudera@quickstart:~
```

```
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose exec cloudera bash -c "su -l cloudera"
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ hdfs dfs -mkdir clickstream
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 2 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:41 clickstream
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ hdfs dfs -mkdir clickstream/logs
[cloudera@quickstart ~]$ hdfs dfs -mkdir clickstream/iplookup
[cloudera@quickstart ~]$ hdfs dfs -ls clickstream
Found 2 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:42 clickstream/iplookup
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:41 clickstream/logs
[cloudera@quickstart ~]$ hdfs dfs -put datasets/clickstream/*.csv  clickstream/iplookup/
[cloudera@quickstart ~]$ hdfs dfs -put datasets/clickstream/*.log  clickstream/logs/
[cloudera@quickstart ~]$ hdfs dfs -ls clickstream/iplookup
Found 1 items
-rw-r--r--  1 cloudera cloudera    1251 2021-08-08 17:44 clickstream/iplookup/ip_lookup.csv
[cloudera@quickstart ~]$ hdfs dfs -ls clickstream/logs
Found 3 items
-rw-r--r--  1 cloudera cloudera   137233 2021-08-08 17:45 clickstream/logs/u_ex160211.log
-rw-r--r--  1 cloudera cloudera   78658 2021-08-08 17:45 clickstream/logs/u_ex160212.log
-rw-r--r--  1 cloudera cloudera   105235 2021-08-08 17:45 clickstream/logs/u_ex160213.log
[cloudera@quickstart ~]$ -
```

3. Use the [MapReduce](#) examples:

```
export MREX=/usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar  
to perform a wordcount on the 2016 State of the Union address, saving the output to the HDFS  
folder sotu2016. Write down the commands to complete the task. How many times does the word  
are appear in the 2016 State of the Union address? Describe a process which could be done to  
make the wordcount more useful?
```

**Solution:**

--view the contents of the 2016 state of the union address

```
hdfs dfs -cat /user/cloudera/text/2016-state-of-the-union.txt
```

--define MapReduce jar

```
export MREX=/usr/lib/Hadoop-mapreduce/Hadoop-mapreduce-  
examples.jar
```

--perform wordcount on 2016 state of the union address

```
yarn jar $MREX wordcount /user/cloudera/text/2016-state-of-the-  
union.txt /user/cloudera/sotu2016
```

--view the output file

```
hdfs dfs -ls sotu2016/
```

```
hdfs dfs -cat sotu2016/part-r-00000
```

--copy file to local

```
hdfs dfs -get sotu2016/part-r-00000 sotu2016.csv
```

--Later, we can perform visualization on this using R or Python for  
frequency plot, top 10 occurrences, after removing stop words and other  
text processing, we can plot word cloud or perform sentiment analysis.

## Evidence:

769-Win10Docker-srajendi

Enforce US Keyboard Layout

```
cloudera@quickstart:~]$ export MREX=/usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar
cloudera@quickstart ~]$ yarn jar $MREX
An example program must be given as the first argument.
Valid program names are:
  aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
  aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
  bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
  dbcount: An example job that count the pageview counts from a database.
  distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
  grep: A map/reduce program that counts the matches of a regex in the input.
  join: A job that effects a join over sorted, equally partitioned datasets
  multifilewc: A job that counts words from several files.
  pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
  pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
  randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
  randomwriter: A map/reduce program that writes 10GB of random data per node.
  secondarysort: An example defining a secondary sort to the reduce.
  sort: A map/reduce program that sorts the data written by the random writer.
  sudoku: A sudoku solver.
  teragen: Generate data for the terasort
  terasort: Run the terasort
  teravalidate: Checking results of terasort
  wordcount: A map/reduce program that counts the words in the input files.
  wordmean: A map/reduce program that counts the average length of the words in the input files.
  wordmedian: A map/reduce program that counts the median length of the words in the input files.
  wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.
[cloudera@quickstart ~]$
```

769-Win10Docker-srajendi

Enforce US Keyboard Layout View Fullscreen

```
cloudera@quickstart:~]
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 2 items
drwxr-xr-x  - cloudera cloudera          0 2021-08-08 17:42 clickstream
drwxr-xr-x  - cloudera cloudera          0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ yarn jar $MREX wordcount /user/cloudera/text/2016-state-of-the-union.txt /user/cloudera/sotu2016
11/08/09 19:36:33 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0:8032
11/08/09 19:36:33 INFO input.FileInputFormat: Total input paths to process : 1
11/08/09 19:36:34 INFO mapreduce.JobSubmissionHandler: Starting of job
11/08/09 19:36:34 INFO mapreduce.JobSubmissionHandler: Submitting tokens for job: job_1628227826905_0001
11/08/09 19:36:35 INFO mapreduce.Job: impl.YarnClientImpl: Submitted application application_1628227826905_0001
11/08/09 19:36:35 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1628227826905_0001/
11/08/09 19:36:35 INFO mapreduce.Job: Running job: job_1628227826905_0001
```

769-Win10Docker-srajendi

cloudera@quickstart:~]

```
Total megabyte-seconds taken by all reduce tasks=3947520
Map-Reduce Framework
  Map input records=173
  Map output records=6152
  Map output bytes=60075
  Map output materialized bytes=27301
  Input split bytes=143
  Combine input records=6152
  Combine output records=2018
  Reduce input groups=2018
  Reduce shuffle bytes=27301
  Reduce input records=2018
  Reduce output records=2018
  Spilled Records=4036
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=63
  CPU time spent (ms)=2140
  Physical memory (bytes) snapshot=472170496
  Virtual memory (bytes) snapshot=2715508736
  Total committed heap usage (bytes)=364380160
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=35731
File Output Format Counters
  Bytes Written=19316
[cloudera@quickstart ~]$
```

769-Win10Docker-srajend

```
Bytes Read=35731
File Output Format Counters
Bytes Written=19316
cloudera@quickstart ~$ hdfs dfs -ls
Found 3 items
drwxr-xr-x - cloudera cloudera 0 2021-08-08 17:42 clickstream
drwxr-xr-x - cloudera cloudera 0 2021-08-09 19:37 sotu2016
drwxr-xr-x - cloudera cloudera 0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ hdfs dfs -ls sotu2016/
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2021-08-09 19:37 sotu2016/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 19316 2021-08-09 19:37 sotu2016/part-r-00000
[cloudera@quickstart ~]$ hdfs dfs -ls sotu2016/part-r-00000
-rw-r--r-- 1 cloudera cloudera 19316 2021-08-09 19:37 sotu2016/part-r-00000
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/text/2016-state-of-the-union.txt
9:10 P.M. EST

Mr. Speaker, Mr. Vice President, Members of Congress, my fellow Americans:

Tonight marks the eighth year that I've come here to report on the State of the Union. And for this final one, I'm going to try to make it a little shorter. (Applause.) Now some of you are antsy to get back to Iowa. (Laughter.) I've been there. I'll be shaking hands afterwards if you want some tips. (Laughter.)

And I understand that because it's an election season, expectations for what we will achieve this year are low. But, Mr. Speaker, I appreciate the constructive approach you and the other leaders took at the end of last year to pass a budget and make tax cuts permanent for working families. So I hope we can work together this year on a bipartisan priorities like criminal justice reform -- (applause) -- and helping people who are battling prescription drug abuse and heroin abuse. (Applause.) So, who knows? We might surprise the cynics again.

But tonight, I want to go easy on the traditional list of proposals for the year ahead. Don't worry, I've got plenty, from helping students learn to write computer code personalizing medical treatments for patients. And I will keep pushing for progress on the work that I believe still needs to be done. Fixing a broken immigration system. (Applause.) Protecting our kids from gun violence. (Applause.) Equal pay for equal work. (Applause.) Paid leave. (Applause.) Raising the minimum wage. (Applause.) All these things still matter to hardworking families. They're still the right thing to do. And I won't let up until they get done.

But for my final address to this chamber, I don't want to just talk about next year. I want to focus on the next five years, the next 10 years, and beyond. I want to focus on our future.
```

769-Win10Docker-srajend

```
I see it in the soldier who gives almost everything to save his brothers, the nurse who tends to him till he can run a marathon, the community leader who believes in us.
```

It's the son who finds the courage to come out as who he is, and the father whose love for that son overrides everything he's been taught. (Applause.)

I see it in the elderly woman who will wait in line to cast her vote as long as she has to; the new citizen who casts his vote for the first time. (Applause.) Those who believe every vote should count -- because each of them in different ways know how much that precious right is worth.

That's the America I know. That's the country we love. Clear-eyed. Big-hearted. Undaunted by challenge. Optimistic that unarmed truth and constitutional law will be heard. (Applause.) That's what makes me so hopeful about our future. I believe in change because I believe in you, the American people.

And that's why I stand here confident as I have ever been that the State of our Union is strong. (Applause.)

```
Thank you, God bless you. God bless the United States of America.
[cloudera@quickstart ~]$ hdfs dfs -cat sotu2016/part-r-00000
(Applause.) 84
(Laughter.) 5
(applause) 4
-- 61
10 2
10,000 1
12 1
14 1
18 1
18,000 1
200 2
2014 1
2015 1
21st 2
30 1
40 1
405 1
505, 1
60 2
900,000 1
9:10 1
A 2
```

## 769-Win10Docker-srajendi

```
cloudera@quickstart:~$ wordcount sotu2016.csv
world      12
world,     3
world.     7
world's    1
worry,     1
worse     1
worst     1
worst-kept      1
worst?     1
worth.     1
would     4
write      1
wrong.    1
year       10
year,      4
year.      3
years      9
years,     7
years.     2
yes,       2
you        21
you,       3
you.       3
young      3
your       4
you'll     2
you've     1
'90s,      1
"We        1
"dogmas   1
"to        1
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ hdfs dfs -get sotu2016/part-r-00000 sotu2016.csv
get: `sotu2016.csv': File exists
[cloudera@quickstart ~]$ ls -l
total 184
-rwxrwxr-x 1 cloudera cloudera  5387 Apr  5  2016 cloudera-manager
-rwxrwxr-x 1 cloudera cloudera 9964 Apr  5  2016 cm api.py
drwxrwxrwx 1 hdfs      hdfs    8192 Jul 12 22:22 datasets
drwxrwxr-x 2 cloudera cloudera 4096 Apr  6  2016 Desktop
drwxrwxr-x 4 cloudera cloudera 4096 Apr  5  2016 Documents
-rw-rw-r-- 1 cloudera cloudera 53653 Apr  5  2016 enterprise-deployment.json
-rw-rw-r-- 1 cloudera cloudera 50513 Apr  5  2016 express-deployment.json
-rwxrwxr-x 1 cloudera cloudera 5007 Apr  5  2016 kerberos
drwxrwxr-x 2 cloudera cloudera 4096 Apr  5  2016 lib
-rwxrwxr-x 1 cloudera cloudera 3934 Apr  5  2016 parcels
-rw-r--r-- 1 cloudera cloudera 19316 Aug  9 21:24 sotu2016.csv
drwxrwxr-x 4 cloudera cloudera 4096 Apr  5  2016 workspace
[cloudera@quickstart ~]$
```

4. Type the following command to import the **fudgemart\_v3** database into the local mysql instance on the Hadoop client:

```
mysql -u root -p < ~/datasets/fudgemart/mysql.sql
```

The password is cloudera. Write down the commands you used to complete these tasks:

Use the sqoop utility to verify there are tables in the database by listing them from the **fudgemart\_v3** database. Next write a sqoop command to import Fudgemart products in the 'Clothing' department into a HDFS folder /user/cloudera/fudgemart-clothing

**Solution:**

--import fudgemart\_v3 database into local mysql instance

```
mysql -u root -p < ~/datasets/fudgemart/mysql.sql
```

--list databases using sqoop

```
sqoop list-databases --connect jdbc:mysql://cloudera --username=root --password=cloudera
```

--list tables using sqoop

```
sqoop list-tables --connect jdbc:mysql://cloudera/fudgemart_v3 --username=root --password=cloudera
```

--import into HDFS using sqoop

```
sqoop import --connect jdbc:mysql://cloudera/fudgemart_v3 --username=root --password=cloudera --target-dir /user/cloudera/fudgemart-clothing --query "select * from fudgemart_products where product_department='clothing' and \$CONDITIONS" --as-textfile --split-by-product_name
```

--verify the HDFS

```
hdfs dfs -ls fudgemart-clothing
```

--view the data

```
hdfs dfs -cat fudgemart-clothing/part-m-00004
```

## Evidence:

769-Win10Docker-srajendi

```
cloudera@quickstart:~  
[cloudera@quickstart ~]$ mysql -u root -p < ~/datasets/fudgemart/mysql.sql  
Enter password:  
[cloudera@quickstart ~]$ sqoop help  
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
21/08/09 23:45:30 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0  
Usage: sqoop COMMAND [ARGS]  
  
Available commands:  
  codegen          Generate code to interact with database records  
  create-hive-table Import a table definition into Hive  
  eval             Evaluate a SQL statement and display the results  
  export           Export an HDFS directory to a database table  
  help             List available commands  
  import            Import a table from a database to HDFS  
  import-all-tables Import tables from a database to HDFS  
  import-mainframe  Import datasets from a mainframe server to HDFS  
  job              Work with saved jobs  
  list-databases   List available databases on a server  
  list-tables      List available tables in a database  
  merge            Merge results of incremental imports  
  metastore        Run a standalone Squeryl metastore  
  version          Display version information  
  
See 'sqoop help COMMAND' for information on a specific command.  
[cloudera@quickstart ~]$ ■
```

769-Win10Docker-srajendi

Enforce US Keyboard Layout

```
cloudera@quickstart:~  
-rw-rw-r-- 1 cloudera cloudera 53653 Apr  5 2016 enterprise-deployment.json  
-rw-rw-r-- 1 cloudera cloudera 50513 Apr  5 2016 express-deployment.json  
-rwxrwxr-x 1 cloudera cloudera 5007 Apr  5 2016 kerberos  
drwxrwxr-x 2 cloudera cloudera 4096 Apr  5 2016 lib  
-rwxrwxr-x 1 cloudera cloudera 3934 Apr  5 2016 parcels  
-rw-r--r-- 1 cloudera cloudera 19316 Aug  9 21:24 sotu2016.csv  
-rw-r--r-- 1 cloudera cloudera 19316 Aug  9 23:42 wc_sotu2016.csv  
drwxrwxr-x 4 cloudera cloudera 4096 Apr  5 2016 workspace  
[cloudera@quickstart ~]$ clear  
[cloudera@quickstart ~]$ sqoop list-databases --connect jdbc:mysql://cloudera --username=root --password=cloudera  
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
21/08/10 00:00:10 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0  
21/08/10 00:00:10 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.  
21/08/10 00:00:11 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.  
information_schema  
cm  
Firehose  
fudgemart_v3  
hue  
metastore  
mysql  
nav  
navms  
oozie  
retail_db  
rman  
sentry  
twitter  
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 11365
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>
```

```
mysql> show databases;
+-----+
| Database      |
+-----+
| information_schema |
| cm            |
| firehose       |
| fudgemart_v3   |
| hue           |
| metastore      |
| mysql          |
| nav            |
| navms          |
| oozie          |
| retail_db      |
| rman          |
| sentry          |
| twitter         |
+-----+
14 rows in set (0.00 sec)
```

```

mysql> use fudgemart_v3;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+-----+
| Tables_in_fudgemart_v3 |
+-----+
| fudgemart_creditcards |
| fudgemart_customer_creditcards |
| fudgemart_customers |
| fudgemart_departments_lookup |
| fudgemart_employee_timesheets |
| fudgemart_employees |
| fudgemart_jobtitles_lookup |
| fudgemart_order_details |
| fudgemart_orders |
| fudgemart_products |
| fudgemart_shipvia_lookup |
| fudgemart_vendors |
+-----+
12 rows in set (0.00 sec)

```

```

mysql> select * from fudgemart_products;
+-----+-----+-----+-----+-----+-----+-----+-----+
| product_id | product_department | product_name | product_retail_price | product_wholesale_price | product_is_active | product_add_date | product_vendor_id |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | Hardware | Straight Claw Hammer | 15.9500 | 11.0000 | 1 | 2007-08-31 21:42:45 | 3 |
| 2 | Hardware | Sledge Hammer | 21.9500 | 15.0000 | 1 | 2007-08-31 21:42:45 | 3 |
| 3 | Hardware | Rip Claw Hammer | 19.9500 | 12.0000 | 1 | 2007-08-31 21:42:45 | 3 |
| 4 | Clothing | Ori-Fit Tee | 20.0000 | 8.0000 | 1 | 2007-09-14 21:13:46 | 2 |
| 5 | Clothing | Running Pants | 35.0000 | 12.0000 | 1 | 2007-09-14 21:14:04 | 2 |
| 6 | Clothing | Wool Socks | 8.0000 | 2.0000 | 1 | 2007-09-14 21:14:20 | 2 |
| 7 | Clothing | Squeaky Sneaks | 65.0000 | 20.0000 | 1 | 2007-09-14 21:14:43 | 2 |
| 8 | Clothing | Cool Jeans | 45.0000 | 10.0000 | 1 | 2007-09-14 21:15:08 | 7 |
| 9 | Clothing | Denim Jacket | 60.0000 | 15.0000 | 1 | 2007-09-14 21:15:21 | 7 |
+-----+-----+-----+-----+-----+-----+-----+-----+
53 rows in set (0.00 sec)

mysql> -

```

```

-----+
53 rows in set (0.00 sec)

mysql> \q
Bye
[cloudera@quickstart ~]$ -

```

```
[cloudera@quickstart ~]$ sqoop list-databases --connect jdbc:mysql://cloudera --username=root --password=cloudera
Warning: /usr/lib/sqoop/..../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/08/10 00:09:07 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0
21/08/10 00:09:07 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/08/10 00:09:07 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
information_schema
cm
firehose
fudgemart_v3
hue
metastore
mysql
nav
navms
oozie
retail_db
rman
sentry
twitter
```

```
[cloudera@quickstart ~]$ sqoop list-tables --connect jdbc:mysql://cloudera/fudgemart_v3 --username=root --password=cloudera
Warning: /usr/lib/sqoop/..../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/08/10 00:10:24 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0
21/08/10 00:10:24 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/08/10 00:10:24 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
fudgemart_creditcards
fudgemart_customer_creditcards
fudgemart_customers
fudgemart_departments_lookup
fudgemart_employee_timesheets
fudgemart_employees
fudgemart_jobtitles_lookup
fudgemart_order_details
fudgemart_orders
fudgemart_products
fudgemart_shipvia_lookup
fudgemart_vendors
[cloudera@quickstart ~]$
```

```
Warning: /usr/lib/sqoop/..../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/08/10 00:31:15 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0
21/08/10 00:31:15 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/08/10 00:31:15 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
21/08/10 00:31:15 INFO tool.CodeGenTool: Beginning code generation
21/08/10 00:31:15 INFO manager.SqlManager: Executing SQL statement: select * from fudgemart_products where product_department='clothing' and (i = 0)
21/08/10 00:31:15 INFO manager.SqlManager: Executing SQL statement: select * from fudgemart_products where product_department='clothing' and (i = 0)
21/08/10 00:31:16 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/34140e52af202cfe23fb98de151acc/QueryResult.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/08/10 00:31:18 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/34140e52af202cfe23fb98de151acc/QueryResult.jar
21/08/10 00:31:18 INFO mapreduce.JobReportJobBase: Beginning query import
21/08/10 00:31:18 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
21/08/10 00:31:18 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/08/10 00:31:19 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/08/10 00:31:19 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8092
21/08/10 00:31:21 INFO db.DBInputFormat: Using read committed transaction isolation
21/08/10 00:31:21 INFO db.DataDrivenBindInputFormat: BoundingValsQuery: SELECT MIN(product_name), MAX(product_name) FROM (select * from fudgemart_products where product_department='clothing' and (i = 0)) AS T
21/08/10 00:31:21 WARN db.TextSplitter: Generating splits for a textual index column.
21/08/10 00:31:21 WARN db.TextSplitter: If your database sorts in a case-insensitive order, this may result in a partial import or duplicate records.
21/08/10 00:31:21 INFO db.TextSplitter: You are strongly encouraged to choose an integral split column.
21/08/10 00:31:21 INFO mapreduce.JobSubmitter: number of splits:6
21/08/10 00:31:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1628227826905_0004
21/08/10 00:31:22 INFO impl.YarnHttpClientImpl: Submitted application application_1628227826905_0004
21/08/10 00:31:22 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1628227826905_0004
21/08/10 00:31:22 INFO mapreduce.Job: Running Job: job_1628227826905_0004
21/08/10 00:31:30 INFO mapreduce.Job: Job job_1628227826905_0004 running in uber mode : false
21/08/10 00:31:30 INFO mapreduce.Job: map 0% reduce 0%
```

```
cloudera@quickstart:~$ hdfs dfs -counters
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=834624
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=854
  HDFS: Number of bytes written=1040
  HDFS: Number of read operations=24
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=12
Job Counters
  Launched map tasks=6
  Other local map tasks=6
  Total time spent by all maps in occupied slots (ms)=153954
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=153954
  Total vcore-seconds taken by all map tasks=153954
  Total megabyte-seconds taken by all map tasks=157648896
Map-Reduce Framework
  Map input records=14
  Map output records=14
  Input split bytes=854
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=1506
  CPU time spent (ms)=8890
  Physical memory (bytes) snapshot=1036062720
  Virtual memory (bytes) snapshot=8229240832
  Total committed heap usage (bytes)=666370048
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=1040
21/08/10 00:32:07 INFO mapreduce.ImportJobBase: Transferred 1.0156 KB in 44.4273 seconds (23.409 bytes/sec)
21/08/10 00:32:07 INFO mapreduce.ImportJobBase: Retrieved 14 records.
[cloudera@quickstart ~]$
```

```
cloudera@quickstart:~$ hdfs dfs -ls fudgemart-clothing
Physical memory (bytes) snapshot=1036062720
Virtual memory (bytes) snapshot=8229240832
Total committed heap usage (bytes)=666370048
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=1040
21/08/10 00:32:07 INFO mapreduce.ImportJobBase: Transferred 1.0156 KB in 44.4273 seconds (23.409 bytes/sec)
21/08/10 00:32:07 INFO mapreduce.ImportJobBase: Retrieved 14 records.
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 5 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:42 clickstream
drwxr-xr-x  - cloudera cloudera      0 2021-08-10 00:32 fudgemart-clothing
drwxr-xr-x  - cloudera cloudera      0 2021-08-10 00:23 fudgemart-products-by-clothing
drwxr-xr-x  - cloudera cloudera      0 2021-08-09 23:40 sotu2016
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ hdfs dfs -ls fudgemart-clothing
Found 7 items
-rw-r--r--  1 cloudera cloudera      0 2021-08-10 00:32 fudgemart-clothing/_SUCCESS
-rw-r--r--  1 cloudera cloudera      0 2021-08-10 00:31 fudgemart-clothing/part-m-00000
-rw-r--r--  1 cloudera cloudera  445 2021-08-10 00:31 fudgemart-clothing/part-m-00001
-rw-r--r--  1 cloudera cloudera   77 2021-08-10 00:32 fudgemart-clothing/part-m-00002
-rw-r--r--  1 cloudera cloudera  151 2021-08-10 00:32 fudgemart-clothing/part-m-00003
-rw-r--r--  1 cloudera cloudera  291 2021-08-10 00:32 fudgemart-clothing/part-m-00004
-rw-r--r--  1 cloudera cloudera   76 2021-08-10 00:32 fudgemart-clothing/part-m-00005
[cloudera@quickstart ~]$ hdfs dfs -cat fudgemart-clothing fudgemart-products-by-clothing/part-m-00004
cat: `fudgemart-clothing': Is a directory
6,Clothing,Wool Socks,8.0000,2.0000,true,2007-09-14 21:14:20.0,2,null
7,Clothing,Squeaky Sneaks,65.0000,20.0000,true,2007-09-14 21:14:43.0,2,null
12,Clothing,Work Pants,38.0000,20.0000,true,2007-09-16 02:43:44.0,3,null
13,Clothing,Work Gloves,8.0000,4.0000,true,2007-09-16 02:44:04.0,3,null
[cloudera@quickstart ~]$ hdfs dfs -cat fudgemart-clothing fudgemart-clothing/part-m-00004
cat: `fudgemart-clothing': Is a directory
6,Clothing,Wool Socks,8.0000,2.0000,true,2007-09-14 21:14:20.0,2,null
7,Clothing,Squeaky Sneaks,65.0000,20.0000,true,2007-09-14 21:14:43.0,2,null
12,Clothing,Work Pants,38.0000,20.0000,true,2007-09-16 02:43:44.0,3,null
13,Clothing,Work Gloves,8.0000,4.0000,true,2007-09-16 02:44:04.0,3,null
[cloudera@quickstart ~]$
```

5. Let's import **HDFS** data into **MySQL**. Record each command you type as your solution:

- a. Load datasets/tweets/tweets.psv into the HDFS folder tweets
- b. Login to MySQL: mysql -u root -p the password is cloudera. Create a database twitter
- c. Create a table called tweets inside the database twitter the table should have columns for id, timestamp, date time, username, and tweet\_text.
- d. Export the data from HDFS into the MySQL table.

TIPS: If your SQOOP job fails it is likely due to the table constraints such as selecting a data type too small for the imported data. Try to insert a row in the table using a sample from the HDFS data. This will help you to ensure your chosen data types will work.

Solution:

--create tweets folder in HDFS

**hdfs dfs -mkdir tweets**

--create tweets data from datasets/tweets.psv file

**hdfs dfs -put datasets/tweets/tweets.psv tweets/**

--view the contents from tweets file

**hdfs dfs -cat tweets/tweets.psv**

--connect to mysql instance

**mysql -u root -p**

--create twitter database

**show databases;**

**create database twitter;**

**use twitter;**

--create tweets table

**create table tweets(id bigint, timestamp varchar(50), tweet\_date varchar(50), username text, tweet\_text text);**

**describe tweets;**

--export tweets data from HDFS into mysql tweets table using sqoop  
**sqoop export --connect jdbc:mysql://cloudera/twitter --username=root --password=cloudera --table tweets --export-dir=tweets --input-fields-terminated-by="|"**

--view the data from mysql table

**select \* from tweets;**

Evidence:

```
769-Win10Docker-srajendi

cloudera@quickstart:~$ hdfs dfs -ls
Found 5 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:42 clickstream
drwxr-xr-x  - cloudera cloudera      0 2021-08-10 00:32 fudgemart-clothing
drwxr-xr-x  - cloudera cloudera      0 2021-08-10 00:23 fudgemart-products-by-clothing
drwxr-xr-x  - cloudera cloudera      0 2021-08-09 23:40 sotu2016
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ hdfs dfs -mkdir tweets
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 6 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:42 clickstream
drwxr-xr-x  - cloudera cloudera      0 2021-08-10 00:32 fudgemart-clothing
drwxr-xr-x  - cloudera cloudera      0 2021-08-10 00:23 fudgemart-products-by-clothing
drwxr-xr-x  - cloudera cloudera      0 2021-08-09 23:40 sotu2016
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:34 text
drwxr-xr-x  - cloudera cloudera      0 2021-08-10 00:45 tweets
[cloudera@quickstart ~]$ hdfs dfs -put datasets/tweets/tweets.psv tweets/
[cloudera@quickstart ~]$ hdfs dfs -ls tweets
Found 1 items
-rw-r--r--  1 cloudera cloudera    25167 2021-08-10 00:47 tweets/tweets.psv
[cloudera@quickstart ~]$
```

769-Win10Docker-srajend

Enforce US Keyboard Layout | View Fullscreen

```
[cloudera@quickstart ~]$ hdfs dfs -cat tweets/tweets.psv
284542858399282239|1433766161.2984276|Mon Jun 08 08:22:41 +0000 2015|rovlight|Why so horrible digital downloads? comeon, @fudgemart
165818396502391067|1429921634.4679817|Thu Apr 23 16:36:34 +0000 2015|sladd|Just placed an order today. thx @fudgemart
973476785498736368|1442107952.4352274|Sat Sep 12 21:32:32 +0000 2015|rdeboat|Worst purchase ever: jewelry at @fudgemart.
402981967846399849|1447381758.722716|Thu Nov 12 21:29:18 +0000 2015|sofewe|I love you, @fudgemart
1397754748167468823|1435573017.366636|Mon Jun 29 06:16:57 +0000 2015|balott|Glad I bought my jewelry from @fudgemart
431741469234206509|1441177812.164332|Wed Sep 02 03:10:12 +0000 2015|ojouglad|The Electronics department at #fudgemart is amazing. Great support!
1397715809932042193|1436108349.8747828|Sun Jul 05 10:59:09 +0000 2015|edetylerys|Who uses the fudgemart website
268216688744405461|1429978347.877872|Sat Apr 25 12:12:27 +0000 2015|rdeboat|This flawed jewelry has me thinking: 'why buy from Fudgemart?'
2211433262754075682|1424533143.683209|Tue Feb 21 10:39:03 +0000 2015|rovlight|There are some new products on Fudgemart. #fudgemart
1109695211099951156|1449661478.9795562|Wed Dec 09 06:44:38 +0000 2015|tanott|Why is customer support so bad at @fudgemart? #frustrated.
2979909623910399438|1428464510.455589|Tue Apr 07 13:23:41:50 +0000 2015|balott|Happy with my service from @fudgemart
3948963872265135052|1442216973.022509|Mon Sep 14 03:49:33 +0000 2015|etasomchin|I love you, @fudgemart
2039050581323090698|1436483758.3057766|Thu Jul 09 19:15:50 +0000 2015|tanott|Hey fudgemart, why is your support so bad? #frustrated
33123998039924364979|1450655056.5671774|Sun Dec 28 18:44:16 +0000 2015|rdeboat|I hate the fudgemart books selection. Terrible.
23183502671918235056|1430332468.2179134|Wed Apr 29 14:34:28 +0000 2015|sladd|Worst purchase ever: digital downloads at @fudgemart.
3713809622550743929|1445637092.8746238|Fri Oct 23 17:51:32 +0000 2015|edetylerys|Really enjoying my electronics purchase from @fudgemart
3558306503555498946|1450971420.3896125|Thu Dec 24 10:37:00 +0000 2015|mmeilator|likewise, @fudgemart
2308210172321288995|1433644555.767323|Sat Jun 06 22:33:55 +0000 2015|ojouglad|Really enjoying my Jewelry purchase from @fudgemart
1883617758940609647|1424415419.5648675|Fri Feb 08 01:56:59 +0000 2015|tpamli|Just got some Jewelry from @fudgemart. Awesome!
4360799934366193476|1442804612.885828|Sun Sep 20 23:03:32 +0000 2015|ojouglad|Hey, @fudgemart. You've got products with benefits!
2139338495649328956|1437569490.6298869|Wed Jul 22 08:50:00 +0000 2015|mmeilator|My shopping experience was delightful, @fudgemart
35536689243657396|1425074798.3228768|Mon Mar 30 16:26:31 +0000 2015|ojouglad|This is fudgemart. The e-commerce platform.
144778589747296694|1425064841.6785831|Fri Mar 06 04:46:47 +0000 2015|akuss|Happy I bought my electronics from @fudgemart
2769766452268729251|1441930112.9828093|Thu Sep 09 07:35:12 +0000 2015|rovlight|The Electronics department at #fudgemart is amazing. So helpful!
101015398407445084|1426790793.69530|Thu Mar 19 14:46:33 +0000 2015|bmelotar|I like the Fudgemart website.
205057732212046506201007|1426420400.0670000|Sat Sep 12 04:20:06 +0000 2015|balott|Came home w/ a big smile my #fudgemart package arrived.
3327477017921248155|1420525224.2163947|Tue Jan 06 01:20:24 +0000 2015|balott|Shopping on #fudgemart website is carefree.
206568132711407006|1424733438.798353|Mon Feb 29 18:17:10 +0000 2015|sladd|I love you, @fudgemart
3778266297027968942|1426191284.9722955|Thu Mar 12 16:14:44 +0000 2015|edetylerys|Shopping 4 books, electronics, and jewelry at #fudgemart.
2822716784223426408|1425301096.4196402|Mon Mar 02 07:58:16 +0000 2015|jpoole|Love or hate it. Fudgemart just is.
```

## 769-Win10Docker-srajendi

```
cloudera@quickstart:~$ [cloudera@quickstart ~]$ mysql -u root -p  
Enter password:  
Welcome to the MySQL monitor. Commands end with ; or \g.  
Your MySQL connection id is 11388  
Server version: 5.1.73 Source distribution  
  
Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.  
  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
  
mysql> show databases;  
+-----+  
| Database |  
+-----+  
| information_schema |  
| cm |  
| firehose |  
| fudgemart_v3 |  
| hue |  
| metastore |  
| mysql |  
| nav |  
| navms |  
| oozie |  
| retail_db |  
| rman |  
| sentry |  
+-----+  
13 rows in set (0.00 sec)  
  
mysql> _
```

```
c:\ cloudera@quickstart:~  
| navms  
| oozie  
| retail_db  
| rman  
| sentry  
+-----+  
13 rows in set (0.00 sec)  
  
mysql> create database twitter;  
Query OK, 1 row affected (0.00 sec)  
  
mysql> show databases;  
+-----+  
| Database  
+-----+  
| information_schema  
| cm  
| firehose  
| fudgemart_v3  
| hue  
| metastore  
| mysql  
| nav  
| navms  
| oozie  
| retail_db  
| rman  
| sentry  
| twitter  
+-----+  
14 rows in set (0.00 sec)  
  
mysql> use twitter;  
Database changed  
mysql> show tables;  
Empty set (0.00 sec)  
  
mysql> -
```

```

cloudera@quickstart:~$ mysql> show tables;
Empty set (0.00 sec)

mysql> Create table tweets (id bigint, timestamp varchar(50), tweet_date varchar(50), username text, tweet_text text);
Query OK, 0 rows affected (0.01 sec)

mysql> describe tweets;
+-----+-----+-----+-----+-----+
| Field | Type   | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| id    | bigint(20) | YES  |     | NULL    |       |
| timestamp | varchar(50) | YES  |     | NULL    |       |
| tweet_date | varchar(50) | YES  |     | NULL    |       |
| username | text    | YES  |     | NULL    |       |
| tweet_text | text    | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)

mysql> select * from tweets;
Empty set (0.00 sec)

mysql> \q
Bye

```

### 769-Win10Docker-srajendi

```

cloudera@quickstart:~$ hdfs dfs -ls
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 6 items
drwxr-xr-x - cloudera cloudera      0 2021-08-08 17:42 clickstream
drwxr-xr-x - cloudera cloudera      0 2021-08-10 00:32 fudgemart-clothing
drwxr-xr-x - cloudera cloudera      0 2021-08-10 00:23 fudgemart-products-by-clothing
drwxr-xr-x - cloudera cloudera      0 2021-08-09 23:40 sotu2016
drwxr-xr-x - cloudera cloudera      0 2021-08-06 06:34 text
drwxr-xr-x - cloudera cloudera      0 2021-08-10 00:47 tweets

```

769-Win10Docker-srajendi

Enforce US Keyboard Layout | View Fullscreen | Send Ctrl+Alt+Delete

```

cloudera@quickstart:~$ sqoop export --connect jdbc:mysql://cloudera/twitter --username=root --password=cloudera --table tweets --export-dir=tweets --input-fields-terminal-delimiter ","
Warning: /usr/lib/sqoop/.Accumulo_HOME does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/08/10 02:01:27 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0
21/08/10 02:01:27 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/08/10 02:01:27 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
21/08/10 02:01:27 INFO tool.CodeGenTool: Beginning code generation
21/08/10 02:01:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `tweets` AS t LIMIT 1
21/08/10 02:01:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `tweets` AS t LIMIT 1
Note: /tmp/sqoop-cloudera/compile/643ec6a2b41fd644acec379daa3d6/tweets.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/08/10 02:01:29 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/643ec6a2b41fd644acec379daa3d6/tweets.jar
21/08/10 02:01:29 INFO mapreduce.ExportJobBase: Beginning export of tweets
21/08/10 02:01:30 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
21/08/10 02:01:30 INFO Configuration.deprecation: mapred.info is deprecated. Instead, use mapreduce.job.jar
21/08/10 02:01:30 INFO Configuration.deprecation: mapred.map.max.attempts is deprecated. Instead, use mapreduce.map.maxattempts
21/08/10 02:01:31 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
21/08/10 02:01:31 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
21/08/10 02:01:31 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
21/08/10 02:01:33 INFO input.FileInputFormat: Total input paths to process : 1
21/08/10 02:01:33 INFO input.FileInputFormat: number of splits:4
21/08/10 02:01:33 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
21/08/10 02:01:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1628227826905_0009
21/08/10 02:01:33 INFO impl.YarnClientImpl: Submitted application application_1628227826905_0009
21/08/10 02:01:33 INFO mapreduce.Job: Running job: job_1628227826905_0009
21/08/10 02:01:33 INFO mapreduce.Job: Job job_1628227826905_0009 running in uber mode : false
21/08/10 02:01:41 INFO mapreduce.Job: map 0% reduce 0%
21/08/10 02:01:55 INFO mapreduce.Job: map 25% reduce 0%
21/08/10 02:01:58 INFO mapreduce.Job: map 50% reduce 0%
21/08/10 02:02:00 INFO mapreduce.Job: map 75% reduce 0%
21/08/10 02:02:01 INFO mapreduce.Job: map 100% reduce 0%
21/08/10 02:02:01 INFO mapreduce.Job: Job job_1628227826905_0009 completed successfully
21/08/10 02:02:01 INFO mapreduce.Job: Counters: 30

```

769-Win10Docker-srajendi

```
cloudera@quickstart:~$ 
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=553376
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=41273
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
Job Counters
  Launched map tasks=4
  Data-local map tasks=4
  Total time spent by all maps in occupied slots (ms)=55992
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=55992
  Total vcore-seconds taken by all map tasks=55992
  Total megabyte-seconds taken by all map tasks=57335808
Map-Reduce Framework
  Map input records=200
  Map output records=200
  Input split bytes=656
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=242
  CPU time spent (ms)=4750
  Physical memory (bytes) snapshot=684318720
  Virtual memory (bytes) snapshot=5476397056
  Total committed heap usage (bytes)=462422016
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
21/08/10 02:02:01 INFO mapreduce.ExportJobBase: Transferred 40.3057 KB in 30.38 seconds (1.3267 KB/sec)
21/08/10 02:02:01 INFO mapreduce.ExportJobBase: Exported 200 records.
[cloudera@quickstart ~]$
```

```
cloudera@quickstart:~  
Your MySQL connection id is 11411  
Server version: 5.1.73 Source distribution  
  
Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.  
  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
  
mysql> show databases;  
+-----+  
| Database |  
+-----+  
| information_schema |  
| cm |  
| firehose |  
| fudgemart_v3 |  
| hue |  
| metastore |  
| mysql |  
| nav |  
| navms |  
| oozie |  
| retail_db |  
| rman |  
| sentry |  
| twitter |  
+-----+  
14 rows in set (0.00 sec)  
  
mysql> use twitter;  
Reading table information for completion of table and column names  
You can turn off this feature to get a quicker startup with -A  
  
Database changed  
mysql>   
  
mysql> describe tweets;  
+-----+-----+-----+-----+-----+  
| Field | Type | Null | Key | Default | Extra |  
+-----+-----+-----+-----+-----+  
| id | bigint(20) | YES | | NULL |  
| timestamp | varchar(50) | YES | | NULL |  
| tweet_date | varchar(50) | YES | | NULL |  
| username | text | YES | | NULL |  
| tweet_text | text | YES | | NULL |  
+-----+-----+-----+-----+-----+  
5 rows in set (0.00 sec)  
  
mysql>   

```

```

769-Win10Docker-srajend1
[cloudera@quickstart ~]# Enforce US Keyboard Layout | View Fullscreen | Send Ctrl+  

[cloudera@quickstart ~]# mysql> describe tweets;
+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| id | bigint(20) | YES | NULL | NULL | |
| timestamp | varchar(50) | YES | NULL | NULL | |
| tweet_date | varchar(50) | YES | NULL | NULL | |
| username | text | YES | NULL | NULL | |
| tweet_text | text | YES | NULL | NULL | |
+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)

[cloudera@quickstart ~]# mysql> select * from tweets;
+-----+-----+-----+-----+-----+
| id | timestamp | tweet_date | username | tweet_text |
+-----+-----+-----+-----+-----+
| 1725307452976161774 | 14249353266.2442936 | Sat Apr 18 08:46:00 +0000 2015 | bdehatchett | The Electronics department at #fudgemart is amazing. Great support! We have hq it. Fudgemart just is.  

| 338330125159049841 | 1427715247.1580972 | Sun May 10 07:34:00 +0000 2015 | etasomthin | The digital downloads section of #fudgemart looks abandoned.  

| 376722637343774230 | 144529896.4366183 | Mon Oct 19 20:11:36 +0000 2015 | etasom | This is fudgemart. The e-commerce platform.  

| 38316546179311594409 | 14429741451.512528 | Wed Apr 22 18:24:11 +0000 2015 | etasomthin | Wonderful books selection. @fudgemart  

| 220135075232894040 | 1441722989.9969445 | Tue Sep 08 10:36:29 +0000 2015 | sofewe | Shipping takes forever. so slow.  

| 3186546179311594409 | 1441722989.9969445 | Wed Apr 22 18:24:11 +0000 2015 | etasom | Shipping takes forever. so slow.  

| 1599507413888183261 | 1445978560.1754277 | Tue Oct 27 16:42:40 +0000 2015 | mrofome | Shipping takes forever. so slow.  

| 3692084089714410699 | 1430645487.8039234 | Sun May 03 05:31:27 +0000 2015 | lkarfurless | Very satisfying when your package arrives from #fudgemart  

| 1435144529973523962 | 1423624752.0679812 | Tue Feb 10 22:19:12 +0000 2015 | sbellum | I was very happy with my latest purchase from #fudgemart  

| 4554613733318994953 | 144529896.4366183 | Fri Oct 16 12:21:26 +0000 2015 | rovlight | Why so horrible digital downloads? comeon, @fudgemart  

| 4937114044590469447 | 14429741451.512528 | Mon Oct 19 20:11:36 +0000 2015 | etasom | The Electronics department at #fudgemart  

| 2118516415578692387 | 1429227814.35195 | Thu Apr 16 19:43:34 +0000 2015 | etasom | My shopping experience was delightful. @fudgemart  

| 2490328955117679574 | 14388416120.808152 | Thu Aug 06 02:03:40 +0000 2015 | bdehatchett | Customer satisfaction through the roof right now.  

| 2663512166565870103 | 144988434.0038161 | Wed Feb 25 18:53:54 +0000 2015 | delylons | Stellar experience at fudgemart. Kudos.  

| 3867311493738860574 | 1439431641.848803 | Wed Aug 12 21:57:21 +0000 2015 | sladd | Happy with my service from @fudgemart  

| 561102249282164447 | 1422080983.1808406 | Sat Aug 08 15:09:43 +0000 2015 | jpoole | The digital downloads section of #fudgemart looks abandoned.  

| 3144084338436970036 | 142241075.0040025 | Sat Jan 24 20:57:55 +0000 2015 | sladd | Fudgemart just is.  

| 3806319257720664485 | 1424319433.5095448 | Wed Feb 18 23:17:13 +0000 2015 | bmelator | Do you enjoy digital downloads? Try @fudgemart.  

| 928949973015114356 | 1425017599.523906 | Fri Feb 27 01:13:19 +0000 2015 | lhvmeehom | Worst purchase ever;jewelry at @fudgemart.  

[cloudera@quickstart ~]# mysql> select * from tweets;
+-----+-----+-----+-----+-----+
| id | timestamp | tweet_date | username | tweet_text |
+-----+-----+-----+-----+-----+
| 4824099494514305744 | 1424653627.7105754 | Thu Jun 18 14:53:57 +0000 2015 | edetyers | This is Fudgemart. The e-commerce platform.  

| 2653782265331366431 | 1427637514.181983 | Sun Mar 29 09:58:34 +0000 2015 | edetyers | I make my first order from Fudgemart last week.  

| 2065885522291249207 | 1430563947.4629943 | Sat May 02 06:52:27 +0000 2015 | tanot | Came home to a big smile my #fudgemart package arrived.  

| 1499403135206595403 | 14412663498.3651614 | Thu Sep 03 03:45:04 +0000 2015 | balott | The books department at #fudgemart is amazing. Very helpful!  

| 4426797671083456880 | 1445978560.1754277 | Tue Oct 27 12:31:38 +0000 2015 | balott | likewise, @fudgemart  

| 3886997001863594467 | 1447228542.3667243 | Wed Nov 11 02:55:42 +0000 2015 | bmelator | Hey fudgemart, why is your support so bad? #upset  

| 3938076958238392329 | 1420654549.2722821 | Wed Jan 07 13:15:49 +0000 2015 | bdehatchett | Our shopping experience was delightful, @fudgemart  

| 302799945799284731 | 1427810862.7958254 | Tue Mar 31 10:27:42 +0000 2015 | ojouglad | Fudgemart just is.  

| 7354327821721686 | 1428608815.1553352 | Wed Jan 07 06:28:15 +0000 2015 | rdeboat | Fudgemart's digital downloads selection is uneventful!  

| 2500000000000000000 | 1429101813.1553352 | Mon Mar 02 10:27:42 +0000 2015 | rdeboat | Effectiveness of being a cheapskate @fudgemart  

| 1829454232943694548 | 1438388721.250163 | Fri Jul 31 18:12:01 +0000 2015 | bmelator | Fudgemart's digital downloads: pahty!  

| 10409965353233080216 | 1439560801.3991952 | Fri Aug 14 10:00:01 +0000 2015 | etasom | I'm tired of being on the phone with electronics support at @fudgemart  

| 109687559942445837 | 1429110135.0544834 | Wed Apr 15 11:02:15 +0000 2015 | ccayne | Happy I bought my books from @fudgemart  

| 16227748863006899 | 1430457741.3563206 | Fri May 01 01:22:21 +0000 2015 | ojouglad | I was very happy with my latest purchase from #fudgemart  

| 884581216165146099 | 1450168856.288855 | Tue Dec 15 03:40:56 +0000 2015 | rdeboat | Fudgemart's digital downloads selection is uneventful!  

| 3938648347188731124 | 1448320112.3743503 | Mon Nov 23 18:08:32 +0000 2015 | bmelator | Came home to a big smile my #fudgemart package arrived.  

| 4280367237025444705 | 1431150442.1662924 | Sat May 09 01:47:22 +0000 2015 | edetyers | Just got some Jewelry from #fudgemart. Awesome!  

| 114008352627582162 | 1451497070.8426945 | Wed Dec 30 12:37:50 +0000 2015 | mmeelator | Love or hate it. Fudgemart just is.  

| 2439198686974996666 | 1450645913.5928991 | Sat Apr 25 08:23:45 +0000 2015 | sofewe | Nothing like the feeling of helplessness from a Fudgemart product.  

| 1700000000000000000 | 1451497070.8426945 | Wed Feb 10 10:27:42 +0000 2015 | sofewe | Shopping 4 b00t electronics, and jewelry at #fudgemart.  

| 251669735450783608 | 14535183.0847621 | Fri Mar 26 12:13:03 +0000 2015 | lkarfurless | Fudgemart just is.  

| 36318471516654750992 | 14677323981.7996647 | Mon Nov 16 23:03:01 +0000 2015 | tpani | Wonderful jewelry selection, @fudgemart  

| 12418389051805293 | 1423434019.5090412 | Sun Feb 08 17:20:19 +0000 2015 | rdeboat | Stellar experience at fudgemart. Kudos.  

| 20659909125218174866 | 1443828425.3764838 | Fri Oct 02 19:27:05 +0000 2015 | rdeboat | Shipping takes forever. so slow.  

| 1672347940593762564 | 1446804421.445523 | Fri Nov 06 15:07:01 +0000 2015 | mmeelator | I'm really enjoying my new stuff from fudgemart  

| 151326444356317547 | 1430987718.9171083 | Thu May 07 04:35:18 +0000 2015 | ccayne | I was very happy with my latest purchase from #fudgemart  

| 278716508761823744 | 1422243050.7544903 | Sun Jan 25 22:30:59 +0000 2015 | sofewe | Who uses the fudgemart website  

| 337058222194395675 | 1421186885.9698501 | Tue Jan 13 17:08:05 +0000 2015 | ojouglad | The Electronics department at #fudgemart is amazing. So helpful!  

| 3211706053447194363 | 1449515301.5719856 | Mon Dec 07 14:08:21 +0000 2015 | ojouglad | Technical support for my new computer as A+, thank you @fudgemart  

| 46882289967043889 | 144529896.4366183 | Sun Sep 26 03:58:00 +0000 2015 | rdeboat | I hate the fudgemart books selection. Terrible.  

| 1039386477151027297 | 1429227814.35195 | Sun Apr 10 06:27:38 +0000 2015 | bmelator | Hey fudgemart, why is your support so bad? #upset  

| 3083058694770235297 | 1429966321.1274254 | Sat Mar 25 15:52:01 +0000 2015 | mmeelator | The Electronics department at #fudgemart is amazing. Great support!  

| 3725476373732638921 | 1450890393.333433 | Wed Dec 23 12:06:33 +0000 2015 | ojouglad | Terrible. So Bad. @fudgemart  

| 270758177122437466 | 1446600516.536198 | Tue Nov 03 20:28:36 +0000 2015 | tpani | Happy with my service from @fudgemart  

200 rows in set (0.01 sec)

[cloudera@quickstart ~]#

```

```
mysql> drop database if exists twitter;
Query OK, 1 row affected (0.01 sec)

mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| cm |
| firehose |
| fudgemart_v3 |
| hue |
| metastore |
| mysql |
| nav |
| navms |
| oozie |
| retail_db |
| rman |
| sentry |
+-----+
13 rows in set (0.00 sec)

mysql>
```

```
mysql> \q
Bye
[cloudera@quickstart ~]$ exit
logout

C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker compose ps
NAME          SERVICE    STATUS      PORTS
cloudera     cloudera   running    0.0.0.0:8080->80/tcp, :::8080->80/tcp, 0.0.0.0:7180->7180/tcp
, :::7180->7180/tcp, 0.0.0.0:8888->8888/tcp, :::8888->8888/tcp

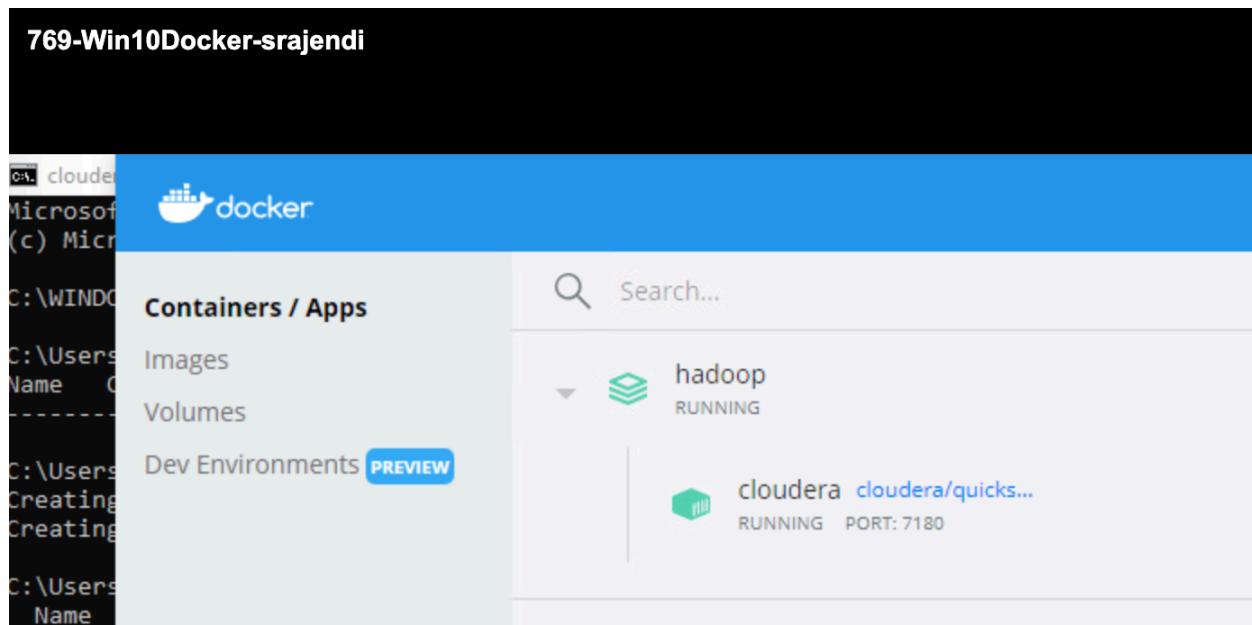
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker compose down
[+] Running 2/2
- Container cloudera     Removed
- Network hadoop_default Removed
13.3s
1.8s

C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>
```

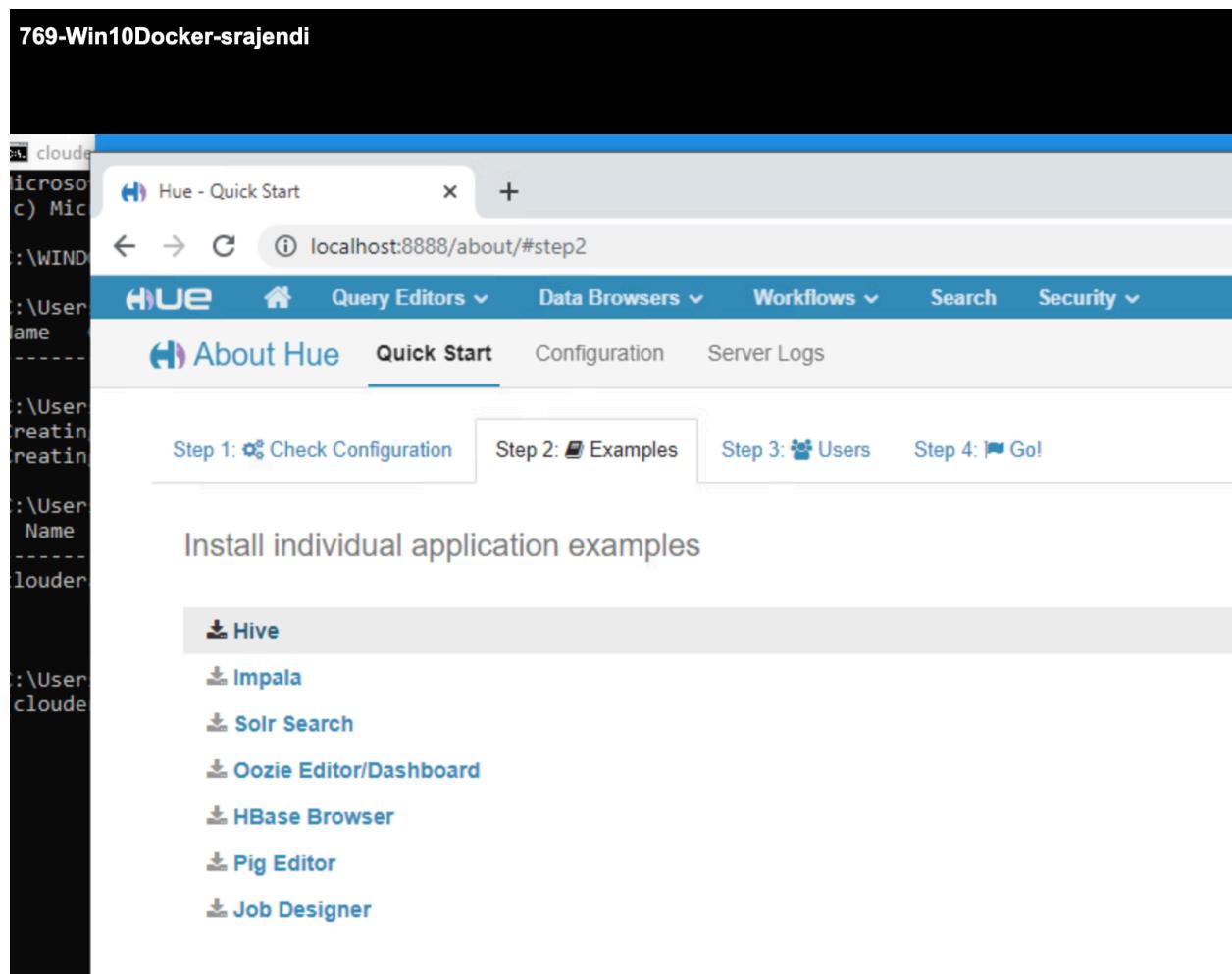
## Appendix

769-Win10Docker-srajendi

```
cloudera@quickstart:~  
Microsoft Windows [Version 10.0.19041.1110]  
(c) Microsoft Corporation. All rights reserved.  
  
C:\WINDOWS\system32>cd C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop  
  
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose ps  
Name      Command     State      Ports  
-----  
  
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose up -d  
Creating network "hadoop_default" with the default driver  
Creating cloudera ... done  
  
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose ps  
Name      Command     State      Ports  
-----  
cloudera  /usr/bin/docker-quickstart  Up      0.0.0.0:7180->7180/tcp,:::7180->7180/tcp,  
          0.0.0.0:8080->80/tcp,:::8080->80/tcp,  
          0.0.0.0:8888->8888/tcp,:::8888->8888/tcp  
  
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose exec cloudera bash -c "su -l cloudera"  
[cloudera@quickstart ~]$
```



769-Win10Docker-srajendi



Hue - Quick Start    x    +

localhost:8888/about/#step2

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

[About Hue](#)   **Quick Start**   [Configuration](#)   [Server Logs](#)

Step 1: [Check Configuration](#)   Step 2: [Examples](#)   Step 3: [Users](#)   Step 4: [Go!](#)

Install individual application examples

[Installing...](#)

[Impala](#)

**Solr Search**

[Oozie Editor/Dashboard](#)

[HBase Browser](#)

[Pig Editor](#)

[Job Designer](#)

The screenshot shows a web browser window titled "Hue - Quick Start" with the URL "localhost:8888/about/#step2". The browser has a standard top bar with back, forward, and search buttons. Below the bar is a blue header navigation bar with the "HUE" logo, a house icon, and dropdown menus for "Query Editors", "Data Browsers", "Workflows", "Search", and "Security". The main content area has a light gray background. At the top of this area, there is a navigation bar with tabs: "About Hue", "Quick Start" (which is underlined), "Configuration", and "Server Logs". Below this, there are four buttons labeled "Step 1: Check Configuration", "Step 2: Examples" (which is highlighted with a gray border), "Step 3: Users", and "Step 4: Go!". The main content below these buttons is titled "Install individual application examples" and lists several items with download icons:

- [Hive](#)
- [Installing...](#)
- [Solr Search](#)
- [Oozie Editor/Dashboard](#)
- [HBase Browser](#)
- [Pig Editor](#)
- [Job Designer](#)

Install individual application examples

[Hive](#)

[Installing...](#)

[Solr Search](#)

[Oozie Editor/Dashboard](#)

[HBase Browser](#)

[Pig Editor](#)

[Job Designer](#)

Hue - Quick Start

localhost:8888/about/#step2

**HUE** Home Query Editors Data Browsers Workflows Search Security

**About Hue** Quick Start Configuration Server Logs

Step 1: Check Configuration Step 2: Examples Step 3: Users Step 4: Go!

Install individual application examples

[Hive](#)  
 [Impala](#)  
Installing...  
 [Oozie Editor/Dashboard](#)  
 [HBase Browser](#)  
 [Pig Editor](#)  
 [Job Designer](#)

Hue - Quick Start

localhost:8888/about/#step2

**HUE** Home Query Editors Data Browsers Workflows Search Security

**About Hue** Quick Start Configuration Server Logs

Examples refreshed

Step 1: Check Configuration Step 2: Examples Step 3: Users Step 4: Go!

Install individual application examples

[Hive](#)  
 [Impala](#)  
 [Solr Search](#)  
 [Oozie Editor/Dashboard](#)  
 [HBase Browser](#)  
 [Pig Editor](#)  
 [Job Designer](#)

Hue - Quick Start    +

localhost:8888/about/#step2

**HUE**     **Query Editors** ▾   **Data Browsers** ▾   **Workflows** ▾   **Search**   **Security** ▾

**About Hue**   **Quick Start**   Configuration   Server Logs

Step 1: Check Configuration   Step 2: Examples   Step 3: Users   Step 4: Go!

Install individual application examples

[Hive](#)  
 [Impala](#)  
 [Solr Search](#)  
[Installing...](#)  
 [HBase Browser](#)  
 [Pig Editor](#)  
 [Job Designer](#)

Hue - Quick Start    X    +

localhost:8888/about/#step2

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

[About Hue](#)   **Quick Start**   [Configuration](#)   [Server Logs](#)

Step 1: [Check Configuration](#)   Step 2: [Examples](#)   Step 3: [Users](#)   Step 4: [Go!](#)

Install individual application examples

[Hive](#)  
 [Impala](#)  
 [Solr Search](#)  
 [Oozie Editor/Dashboard](#)  
[Installing...](#)  
 [Pig Editor](#)  
 [Job Designer](#)

769-Win10Docker-srajendi

The screenshot shows a browser window titled "Hue - Quick Start" with the URL "localhost:8888/about/#step1". The page is part of a "Quick Start" wizard. Step 1, "Check Configuration", is completed and highlighted with a green checkmark icon. Step 2, "Examples", Step 3, "Users", and Step 4, "Go!", are also listed. The main content area displays the message "Checking current configuration" and "Configuration files located in /etc/hue/conf.empty". A success message states "All OK. Configuration check passed." At the bottom, there are "Back" and "Next" buttons.

Hue - Quick Start

localhost:8888/about/#step1

**HUE**   [Home](#)   [Query Editors](#) ▾   [Data Browsers](#) ▾   [Workflows](#) ▾   [Search](#)   [Security](#) ▾

**About Hue**   **Quick Start**   Configuration   Server Logs

Step 1: Check Configuration   Step 2: Examples   Step 3: Users   Step 4: Go!

Checking current configuration

Configuration files located in </etc/hue/conf.empty>

All OK. Configuration check passed.

[Back](#)   [Next](#)

```
cloudera@quickstart:~  
Microsoft Windows [Version 10.0.19041.1110]  
(c) Microsoft Corporation. All rights reserved.  
C:\WINDOWS\system32>cd C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop  
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker compose ps  
NAME           SERVICE      STATUS        PORTS  
cloudera       cloudera    running      0.0.0.0:8080->80/tcp, :::8080->80/tcp, 0.0.0.0:7180->7180/tcp, :::7180->7180/tcp, 0.0.0.0:8888->8888/tcp, :::8888->8888/tcp  
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose exec cloudera bash -c "su -l cloudera"  
[cloudera@quickstart ~]$ ls -l  
total 164  
-rwxrwxr-x 1 cloudera cloudera 5387 Apr  5 2016 cloudera-manager  
-rwxrwxr-x 1 cloudera cloudera 9964 Apr  5 2016 cm api.py  
drwxrwxrwx 1 hdfs   hdfs   8192 Jul 12 22:22 datasets  
drwxrwxr-x 2 cloudera cloudera 4096 Apr  6 2016 Desktop  
drwxrwxr-x 4 cloudera cloudera 4096 Apr  5 2016 Documents  
-rw-rw-r-- 1 cloudera cloudera 53653 Apr  5 2016 enterprise-deployment.json  
-rw-rw-r-- 1 cloudera cloudera 50513 Apr  5 2016 express-deployment.json  
-rwxrwxr-x 1 cloudera cloudera 5007 Apr  5 2016 kerberos  
drwxrwxr-x 2 cloudera cloudera 4096 Apr  5 2016 lib  
-rwxrwxr-x 1 cloudera cloudera 3934 Apr  5 2016 parcels  
drwxrwxr-x 4 cloudera cloudera 4096 Apr  5 2016 workspace  
[cloudera@quickstart ~]$
```

```
cloudera@quickstart:~  
[cloudera@quickstart ~]$ hdfs dfs -ls clickstream/logs  
Found 3 items  
-rw-r--r-- 1 cloudera cloudera 137233 2021-08-08 17:45 clickstream/logs/u_ex160211.log  
-rw-r--r-- 1 cloudera cloudera 78658 2021-08-08 17:45 clickstream/logs/u_ex160212.log  
-rw-r--r-- 1 cloudera cloudera 105235 2021-08-08 17:45 clickstream/logs/u_ex160213.log  
[cloudera@quickstart ~]$ hdfs fsck /user/cloudera/clickstream/logs/u_ex160211.log  
Connecting to namenode via http://quickstart.cloudera:50070  
FSCK started by cloudera (auth:SIMPLE) from /172.18.0.2 for path /user/cloudera/clickstream/logs/u_ex160211.log at Mon Aug 09 04:55:33 UTC 2021  
.Status: HEALTHY  
Total size: 137233 B  
Total dirs: 0  
Total files: 1  
Total symlinks: 0  
Total blocks (validated): 1 (avg. block size 137233 B)  
Minimally replicated blocks: 1 (100.0 %)  
Over-replicated blocks: 0 (0.0 %)  
Under-replicated blocks: 0 (0.0 %)  
Mis-replicated blocks: 0 (0.0 %)  
Default replication factor: 1  
Average block replication: 1.0  
Corrupt blocks: 0  
Missing replicas: 0 (0.0 %)  
Number of data-nodes: 1  
Number of racks: 1  
FSCK ended at Mon Aug 09 04:55:33 UTC 2021 in 0 milliseconds  
  
The filesystem under path '/user/cloudera/clickstream/logs/u_ex160211.log' is HEALTHY  
[cloudera@quickstart ~]$
```

```
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose exec cloudera bash -c "su -l cloudera"  
[cloudera@quickstart ~]$ hdfs dfs -ls  
Found 9 items  
drwxrwxrwx - cloudera cloudera 0 2021-08-06 05:58 2015_11_18  
drwxrwxrwx - cloudera cloudera 0 2021-08-06 05:58 2015_11_19  
drwxrwxrwx - cloudera cloudera 0 2021-08-06 05:58 2015_11_20  
drwxrwxrwx - cloudera cloudera 0 2021-08-06 05:58 2015_11_21  
drwxr-xr-x - cloudera cloudera 0 2021-02-12 00:23 clickstream  
drwxr-xr-x - cloudera cloudera 0 2021-02-05 00:20 fudgemart-clothing  
drwxr-xr-x - cloudera cloudera 0 2021-04-30 01:21 sotu2016  
drwxr-xr-x - cloudera cloudera 0 2021-02-05 00:02 text  
drwxr-xr-x - cloudera cloudera 0 2021-02-05 00:22 tweets  
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ hdfs dfs -ls text
Found 10 items
-rw-r--r-- 1 cloudera cloudera 35731 2021-02-05 00:02 text/2016-state-of-the-union.txt
-rw-r--r-- 1 cloudera cloudera 5387 2021-02-05 00:01 text/cloudera-manager
-rw-r--r-- 1 cloudera cloudera 9964 2021-02-05 00:01 text/cm_api.py
-rw-r--r-- 1 cloudera cloudera 27470 2021-02-05 00:02 text/constitution.txt
drwxr-xr-x - cloudera cloudera 0 2021-02-05 00:01 text/datasets
-rw-r--r-- 1 cloudera cloudera 1951218 2021-02-05 00:02 text/english-words.txt
-rw-r--r-- 1 cloudera cloudera 17641 2021-02-05 00:02 text/gnu-gpl3-license.txt
-rw-r--r-- 1 cloudera cloudera 96536 2021-02-05 00:02 text/mbox-short.txt
-rw-r--r-- 1 cloudera cloudera 327 2021-02-05 00:02 text/preamble.txt
-rw-r--r-- 1 cloudera cloudera 15398 2021-02-05 00:02 text/zork1-walkthru.txt
[cloudera@quickstart ~]$ -
```

```
[cloudera@quickstart ~]$ hdfs dfs -ls tweets
Found 1 items
-rw-r--r-- 1 cloudera cloudera 25167 2021-02-05 00:22 tweets/tweets.psv
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ hdfs dfs -rm -R text
Deleted text
[cloudera@quickstart ~]$ hdfs dfs -rm -R tweets
Deleted tweets
[cloudera@quickstart ~]$ hdfs dfs -rm -R fudgemart-clothing
Deleted fudgemart-clothing
[cloudera@quickstart ~]$ hdfs dfs -rm -R sotu2016
Deleted sotu2016
[cloudera@quickstart ~]$ hdfs dfs -rm -R clickstream
Deleted clickstream
[cloudera@quickstart ~]$ hdfs dfs -rm -R 2015_11_18
Deleted 2015_11_18
[cloudera@quickstart ~]$ hdfs dfs -rm -R 2015_11_19
Deleted 2015_11_19
[cloudera@quickstart ~]$ hdfs dfs -rm -R 2015_11_20
Deleted 2015_11_20
[cloudera@quickstart ~]$ hdfs dfs -rm -R 2015_11_21
Deleted 2015_11_21
[cloudera@quickstart ~]$ hdfs dfs -ls
[cloudera@quickstart ~]$ -
```

```
[cloudera@quickstart ~]$ hdfs dfs -ls
[cloudera@quickstart ~]$ hdfs dfs -mkdir text
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - cloudera cloudera 0 2021-08-06 06:27 text
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ hdfs dfs -put datasets/text/*.txt text/
[cloudera@quickstart ~]$ hdfs dfs -ls text
Found 7 items
-rw-r--r--  1 cloudera cloudera    35731 2021-08-06 06:34 text/2016-state-of-the-union.txt
-rw-r--r--  1 cloudera cloudera    27470 2021-08-06 06:34 text/constitution.txt
-rw-r--r--  1 cloudera cloudera  1951218 2021-08-06 06:34 text/english-words.txt
-rw-r--r--  1 cloudera cloudera   17641 2021-08-06 06:34 text/gnu-gpl3-license.txt
-rw-r--r--  1 cloudera cloudera   96536 2021-08-06 06:34 text/mbox-short.txt
-rw-r--r--  1 cloudera cloudera     327 2021-08-06 06:34 text/preamble.txt
-rw-r--r--  1 cloudera cloudera  15398 2021-08-06 06:34 text/zork1-walkthru.txt
[cloudera@quickstart ~]$
```

## 769-Win10Docker-srajendi

```
cloudera@quickstart:~
```

```
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose exec cloudera bash -c "su -l cloudera"
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ hdfs dfs -mkdir clickstream
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 2 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:41 clickstream
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ hdfs dfs -mkdir clickstream/logs
[cloudera@quickstart ~]$ hdfs dfs -mkdir clickstream/iplookup
[cloudera@quickstart ~]$ hdfs dfs -ls clickstream
Found 2 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:42 clickstream/iplookup
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:41 clickstream/logs
[cloudera@quickstart ~]$
```

```
cloudera@quickstart:~
```

```
[cloudera@quickstart ~]$ hdfs fsck /user/cloudera/clickstream/logs/u_ex160211.log
Connecting to namenode via http://quickstart.cloudera:50070
FSCK started by cloudera (auth:SIMPLE) from /172.18.0.2 for path /user/cloudera/clickstream/logs/u_ex160211.log at Mon Aug 09 04:55:33 UTC 2021
.Status: HEALTHY
 Total size: 137233 B
 Total dirs: 0
 Total files: 1
 Total symlinks: 0
 Total blocks (validated): 1 (avg. block size 137233 B)
 Minimally replicated blocks: 1 (100.0 %)
 Over-replicated blocks: 0 (0.0 %)
 Under-replicated blocks: 0 (0.0 %)
 Mis-replicated blocks: 0 (0.0 %)
 Default replication factor: 1
 Average block replication: 1.0
 Corrupt blocks: 0
 Missing replicas: 0 (0.0 %)
 Number of data-nodes: 1
 Number of racks: 1
FSCK ended at Mon Aug 09 04:55:33 UTC 2021 in 0 milliseconds

The filesystem under path '/user/cloudera/clickstream/logs/u_ex160211.log' is HEALTHY
[cloudera@quickstart ~]$ whoami
cloudera
[cloudera@quickstart ~]$ export MREX=/usr/lib/Hadoop-mapreduce/Hadoop-mapreduce-examples.jar
[cloudera@quickstart ~]$ yarn jar $MREX
Not a valid JAR: /usr/lib/Hadoop-mapreduce/Hadoop-mapreduce-examples.jar
[cloudera@quickstart ~]$
```

769-Win10Docker-srajendi

Enforce US Keyboard Layout

```
cloudera@quickstart:~$ export MREX=/usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar
[cloudera@quickstart ~]$ yarn jar $MREX
An example program must be given as the first argument.
Valid program names are:
  aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
  aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
  bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
  dcount: An example job that count the pageview counts from a database.
  distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
  grep: A map/reduce program that counts the matches of a regex in the input.
  join: A job that effects a join over sorted, equally partitioned datasets
  multifilewc: A job that counts words from several files.
  pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
  pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
  randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
  randomwriter: A map/reduce program that writes 10GB of random data per node.
  secondarysort: An example defining a secondary sort to the reduce.
  sort: A map/reduce program that sorts the data written by the random writer.
  sudoku: A sudoku solver.
  teragen: Generate data for the terasort
  terasort: Run the terasort
  teravalidate: Checking results of terasort
  wordcount: A map/reduce program that counts the words in the input files.
  wordmean: A map/reduce program that counts the average length of the words in the input files.
  wordmedian: A map/reduce program that counts the median length of the words in the input files.
  wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.
[cloudera@quickstart ~]$ -
```

769-Win10Docker-srajendi

Enforce US Keyboard Layout View Fullscreen

```
cloudera@quickstart:~$ hdfs dfs -ls
Found 2 items
drwxr-xr-x  - cloudera cloudera      0 2021-08-08 17:42 clickstream
drwxr-xr-x  - cloudera cloudera      0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ yarn jar $MREX wordcount /user/cloudera/text/2016-state-of-the-union.txt /user/cloudera/sotu2016
21/08/09 19:36:33 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
21/08/09 19:36:34 INFO input.FileInputFormat: Total input paths to process : 1
21/08/09 19:36:34 INFO mapreduce.JobSubmitter: number of splits:1
21/08/09 19:36:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1628227826905_0001
21/08/09 19:36:35 INFO impl.YarnClientImpl: Submitted application application_1628227826905_0001
21/08/09 19:36:35 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1628227826905_0001/
21/08/09 19:36:35 INFO mapreduce.Job: Running job: job_1628227826905_0001
```

769-Win10Docker-srajendi

cloudera@quickstart:~\$ hdfs dfs -ls

```
Bytes Read=35731
File Output Format Counters
Bytes Written=19316
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 3 items
drwxr-xr-x - cloudera cloudera 0 2021-08-08 17:42 clickstream
drwxr-xr-x - cloudera cloudera 0 2021-08-09 19:37 sotu2016
drwxr-xr-x - cloudera cloudera 0 2021-08-06 06:34 text
[cloudera@quickstart ~]$ hdfs dfs -ls sotu2016/
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2021-08-09 19:37 sotu2016/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 19316 2021-08-09 19:37 sotu2016/part-r-00000
[cloudera@quickstart ~]$ hdfs dfs -ls sotu2016/part-r-00000
-rw-r--r-- 1 cloudera cloudera 19316 2021-08-09 19:37 sotu2016/part-r-00000
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/text/2016-state-of-the-union.txt
9:10 P.M. EST

Mr. Speaker, Mr. Vice President, Members of Congress, my fellow Americans:

Tonight marks the eighth year that I've come here to report on the State of the Union. And for this final one, I'm going to try to make it a little shorter. (Applause.) Now some of you are antsy to get back to Iowa. (Laughter.) I've been there. I'll be shaking hands afterwards if you want some tips. (Laughter.)

And I understand that because it's an election season, expectations for what we will achieve this year are low. But, Mr. Speaker, I appreciate the constructive approach you and the other leaders took at the end of last year to pass a budget and make tax cuts permanent for working families. So I hope we can work together this year on bipartisan priorities like criminal justice reform -- (applause) -- and helping people who are battling prescription drug abuse and heroin abuse. (Applause.) So, who knows, we might surprise the cynics again.

But tonight, I want to go easy on the traditional list of proposals for the year ahead. Don't worry, I've got plenty, from helping students learn to write computer code to personalizing medical treatments for patients. And I will keep pushing for progress on the work that I believe still needs to be done. Fixing a broken immigration system. (Applause.) Protecting our kids from gun violence. (Applause.) Equal pay for equal work. (Applause.) Paid leave. (Applause.) Raising the minimum wage. (Applause.) All the things still matter to hardworking families. They're still the right thing to do. And I won't let up until they get done.

But for my final address to this chamber, I don't want to just talk about next year. I want to focus on the next five years, the next 10 years, and beyond. I want to focus on our future.
```

769-Win10Docker-srajendi

cloudera@quickstart:~\$ yarn jar \$MREX wordcount /user/cloudera/text/2016-state-of-the-union.txt /user/cloudera/sotu2016

```
21/08/09 19:36:33 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
21/08/09 19:36:34 INFO input.FileInputFormat: Total input paths to process : 1
21/08/09 19:36:34 INFO mapreduce.JobSubmitter: number of splits:1
21/08/09 19:36:34 INFO mapreduce.YarnClientImpl: Submitting tokens for job: job_1628227826905_0001
21/08/09 19:36:35 INFO impl.YarnClientImpl: Submitted application application_1628227826905_0001
21/08/09 19:36:35 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1628227826905_0001/
21/08/09 19:36:35 INFO mapreduce.Job: Running job: job_1628227826905_0001
21/08/09 19:37:03 INFO mapreduce.Job: Job job_1628227826905_0001 running in uber mode : false
21/08/09 19:37:03 INFO mapreduce.Job: map 0% reduce 0%
21/08/09 19:37:09 INFO mapreduce.Job: map 100% reduce 0%
21/08/09 19:37:16 INFO mapreduce.Job: map 100% reduce 100%
21/08/09 19:37:16 INFO mapreduce.Job: Job job_1628227826905_0001 completed successfully
21/08/09 19:37:16 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=27301
    FILE: Number of bytes written=282231
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=35874
    HDFS: Number of bytes written=19316
    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)=4327
    Total time spent by all reduces in occupied slots (ms)=3855
    Total time spent by all map tasks (ms)=4327
    Total time spent by all reduce tasks (ms)=3855
    Total vcore-seconds taken by all map tasks=4327
    Total vcore-seconds taken by all reduce tasks=3855
    Total megabyte-seconds taken by all map tasks=4430848
    Total megabyte-seconds taken by all reduce tasks=3947520
Map-Reduce Framework
```

769-Win10Docker-srajendi

```
[cloudera@quickstart:~]
      Total megabyte-seconds taken by all reduce tasks=3947520
Map-Reduce Framework
  Map input records=173
  Map output records=6152
  Map output bytes=60075
  Map output materialized bytes=27301
  Input split bytes=143
  Combine input records=6152
  Combine output records=2018
  Reduce input groups=2018
  Reduce shuffle bytes=27301
  Reduce input records=2018
  Reduce output records=2018
  Spilled Records=4036
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=63
  CPU time spent (ms)=2140
  Physical memory (bytes) snapshot=472170496
  Virtual memory (bytes) snapshot=2715508736
  Total committed heap usage (bytes)=364380160
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=35731
File Output Format Counters
  Bytes Written=19316
[cloudera@quickstart ~]$
```

```
cloudera@quickstart:~
```

I see it in the soldier who gives almost everything to save his brothers, the nurse who tends to him till he can run a marathon, the community .

It's the son who finds the courage to come out as who he is, and the father whose love for that son overrides everything he's been taught. (Applause.)

I see it in the elderly woman who will wait in line to cast her vote as long as she has to; the new citizen who casts his vote for the first time. Is who believe every vote should count -- because each of them in different ways know how much that precious right is worth.

That's the America I know. That's the country we love. Clear-eyed. Big-hearted. Undaunted by challenge. Optimistic that unarmed truth and unceasing word. (Applause.) That's what makes me so hopeful about our future. I believe in change because I believe in you, the American people.

And that's why I stand here confident as I have ever been that the State of our Union is strong. (Applause.)

Thank you, God bless you. God bless the United States of America.

```
[cloudera@quickstart ~]$ hdfs dfs -cat sotu2016/part-r-00000
(Applause.)    84
(Laughter.)    5
(applause)     4
--             61
10             2
10,000         1
12             1
14             1
18             1
18,000         1
200            2
2014           1
2015           1
21st            2
30             1
40             1
40s            1
50s,           1
60             2
900,000        1
9:10           1
A              2
```

## 769-Win10Docker-srajendi

```
c4: cloudera@quickstart:~  
world    12  
world,   3  
world.   7  
world's  1  
worry,   1  
worse   1  
worst   1  
worst-kept      1  
worst?   1  
worth.   1  
would   4  
write   1  
wrong.   1  
year    10  
year,   4  
year.   3  
years   9  
years,  7  
years.  2  
yes,    2  
you     21  
you,    3  
you.    3  
young   3  
your    4  
you'll  2  
you've  1  
'90s,   1  
"We    1  
"dogmas 1  
"to    1  
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ hdfs dfs -get sotu2016/part-r-00000 sotu2016.csv
get: `sotu2016.csv': File exists
[cloudera@quickstart ~]$ ls -l
total 184
-rwxrwxr-x 1 cloudera cloudera 5387 Apr  5 2016 cloudera-manager
-rwxrwxr-x 1 cloudera cloudera 9964 Apr  5 2016 cm_api.py
drwxrwxrwx 1 hdfs    hdfs    8192 Jul 12 22:22 datasets
drwxrwxr-x 2 cloudera cloudera 4096 Apr  6 2016 Desktop
drwxrwxr-x 4 cloudera cloudera 4096 Apr  5 2016 Documents
-rw-rw-r-- 1 cloudera cloudera 53653 Apr  5 2016 enterprise-deployment.json
-rw-rw-r-- 1 cloudera cloudera 50513 Apr  5 2016 express-deployment.json
-rwxrwxr-x 1 cloudera cloudera 5007 Apr  5 2016 kerberos
drwxrwxr-x 2 cloudera cloudera 4096 Apr  5 2016 lib
-rwxrwxr-x 1 cloudera cloudera 3934 Apr  5 2016 parcels
-rw-r--r-- 1 cloudera cloudera 19316 Aug  9 21:24 sotu2016.csv
drwxrwxr-x 4 cloudera cloudera 4096 Apr  5 2016 workspace
[cloudera@quickstart ~]$ _
```

```
[cloudera@quickstart ~]$ mysql -u root -p < ~/datasets/fudgemart/mysql.sql
Enter password:
[cloudera@quickstart ~]$
```

The screenshot shows the Docker desktop application interface. At the top, there's a dark header bar with the text "769-Win10Docker-srajendi" on the left and a "Enforce US Keyboard Layout" button on the right. Below this is a blue navigation bar with the Docker logo on the left, and "Upgrade", "Settings", "Metrics", and "Sign in" buttons on the right. The main content area has a light gray background. On the left, there's a sidebar with the title "Containers / Apps" and links for "Images", "Volumes", and "Dev Environments" (with a "PREVIEW" badge). In the center, there's a large blue 3D cube icon. Below it, the text "No containers running" is displayed in a dark font. At the bottom, there's a dark footer bar with the text "Try running a container: Copy and paste this command into your terminal and then come back" and a command-line input field containing "docker run -d -p 80:80 docker/getting-started". To the right of the input field is a small blue square icon.