PRINCIPLES OF BIG DATA MANAGEMENT

PROJECT – I

SPRING 2017

Submitted by: TEAM 7

Sri Sai Anusha Gandu (sgr43)

Sushma Mitta (smgp6)

Sadanand Kallakuri (sk789)

Abhilash Reddy Gaddam (aggg6)

CONTENTS

1. INTRODUCTION

1.1 About Twitter

Twitter is an online news and social networking service where users post and interact with messages, "tweets," restricted to 140 characters. Registered users can post tweets, but those who are unregistered can only read them. Twitter Inc. is based in San Francisco, California, United States, and has more than 25 offices around the world. Twitter was created in March 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams and launched in July, whereby the service rapidly gained worldwide popularity. As of 2016, Twitter had more than 319 million monthly active users.

1.2 About the Project

Here, we have collected the tweets using twitter API through tweepy using the keywords Python, JavaScript and Ruby in JSON (JavaScript Object Notation) format. The tweets then collected have been analyzed and different SQL queries are written to obtain the result.

2. REQUIREMENTS

2.1 Languages

1. Python
2. Scala
3. SQL
4. Java

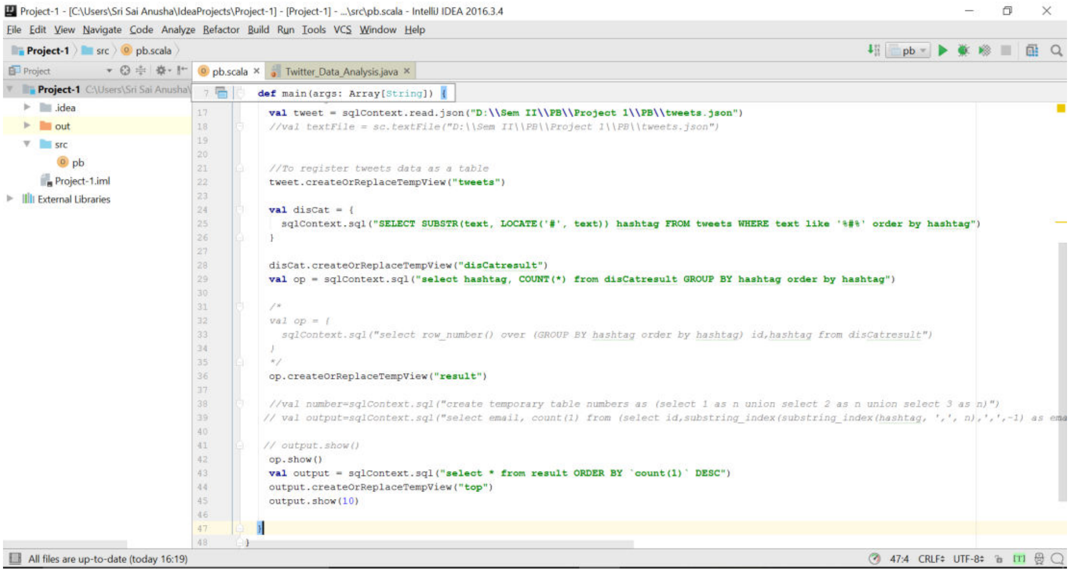
2.2 Software

1. IntelliJ IDEA 3.4 (IDE)
2. Python 3.6
3. JDK 1.8
4. Scala 2.12.1
5. Spark 2.1
6. Virtual Box (Cloudera)

3. TASKS

1. To collect the tweets in JSON (JavaScript Object Notation) format.
2. Find the list of top ten hash-tags.
3. To create HDFS directories for each of top ten hash-tags along with two other directories named Others and None.
4. Segregate the tweets into the respective HDFS directories.
5. To implement a function that can count the frequency of keywords in the directories.

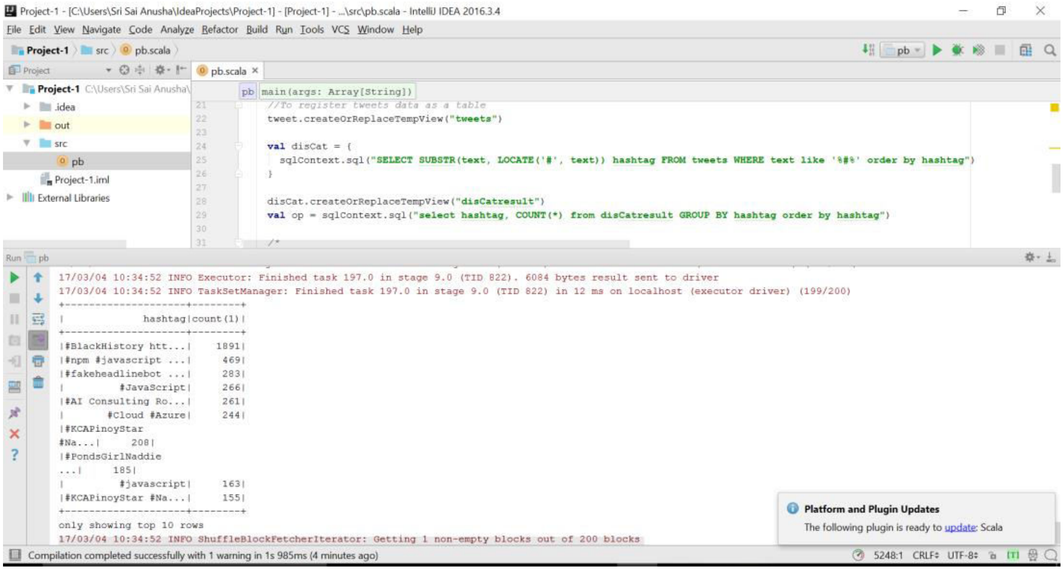
4. TOP TEN HASHTAGS


Project-1 - [C:\Users\Sri Sai Anusha\IdeaProjects\Project-1] - [Project-1] - ...\src\pb.scala - IntelliJ IDEA 2016.3.4
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
Project-1 src pb.scala
pb
Project
pb.scala
Twitter_Data_Analysis.java
Project-1 C:\Users\Sri Sai Anusha\
.idea
out
src
pb
Project-1.iml
External Libraries
47:4 CRLF‡ UTF-8‡

```
def main(args: Array[String]) {
  val tweet = sqlContext.read.json("D:\\Sem II\\PB\\Project 1\\PB\\tweets.json")
  //val textFile = sc.textFile("D:\\Sem II\\PB\\Project 1\\PB\\tweets.json")


  //To register tweets data as a table
  tweet.createOrReplaceTempView("tweets")

  val disCat = {
    sqlContext.sql("SELECT SUBSTR(text, LOCATE('#', text)) hashtag FROM tweets WHERE text like '%#%' order by hashtag")
  }

  disCat.createOrReplaceTempView("disCatresult")
  val op = sqlContext.sql("select hashtag, COUNT(*) from disCatresult GROUP BY hashtag order by hashtag")

  /*
  val op = {
    sqlContext.sql("select row_number() over (GROUP BY hashtag order by hashtag) id,hashtag from disCatresult")
  }
  */
  op.createOrReplaceTempView("result")

  //val number=sqlContext.sql("create temporary table numbers as (select 1 as n union select 2 as n union select 3 as n)")
 // val output=sqlContext.sql("select email, count(1) from (select id,substring_index(substring_index(hashtag, ',', n),',',-1) as ema

 // output.show()
  op.show()
  val output = sqlContext.sql("select * from result ORDER BY `count(1)` DESC")
  output.createOrReplaceTempView("top")
  output.show(10)

}
```

All files are up-to-date (today 16:19)

We have collected the data as tables and from the count thus obtained is sorted in descending manner. The output obtained is shown below.

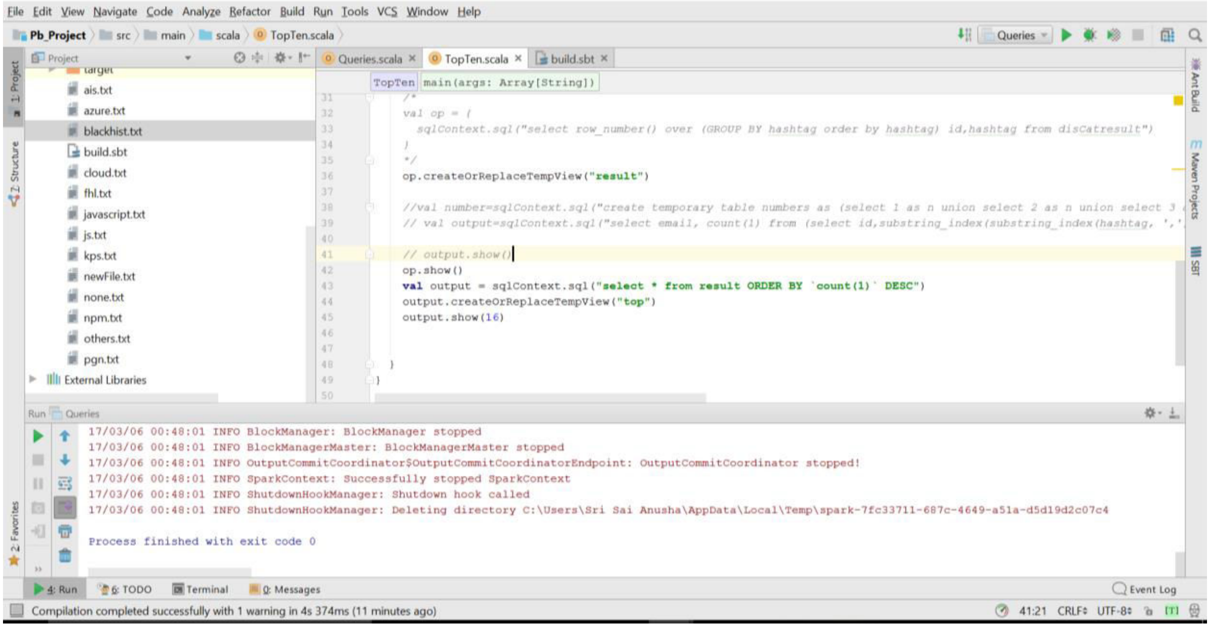

Project-1 - [C:\Users\Sri Sai Anusha\IdeaProjects\Project-1] - [Project-1] - ...\src\pb.scala - IntelliJ IDEA 2016.3.4
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
Project-1 src pb.scala
pb
Project
pb.scala
Project-1 C:\Users\Sri Sai Anusha\
.idea
out
src
pb
Project-1.iml
External Libraries
pb main(args: Array[String])
Run pb
5248:1 CRLF‡ UTF-8‡

```
  //To register tweets data as a table
  tweet.createOrReplaceTempView("tweets")

  val disCat = {
    sqlContext.sql("SELECT SUBSTR(text, LOCATE('#', text)) hashtag FROM tweets WHERE text like '%#%' order by hashtag")
  }

  disCat.createOrReplaceTempView("disCatresult")
  val op = sqlContext.sql("select hashtag, COUNT(*) from disCatresult GROUP BY hashtag order by hashtag")

  /*
```

```
17/03/04 10:34:52 INFO Executor: Finished task 197.0 in stage 9.0 (TID 822). 6084 bytes result sent to driver
17/03/04 10:34:52 INFO TaskSetManager: Finished task 197.0 in stage 9.0 (TID 822) in 12 ms on localhost (executor driver) (199/200)
+--------------------+--------+
|             hashtag|count(1)|
+--------------------+--------+
|#BlackHistory htt...|    1891|
|#npm #javascript ...|     469|
|#fakeheadlinebot ...|     283|
|         #JavaScript|     266|
|#AI Consulting Ro...|     261|
|       #Cloud #Azure|     244|
|#KCAPinoyStar
#Na...|     208|
|#PondsGirlNaddie
...|     185|
|         #javascript|     163|
|#KCAPinoyStar #Na...|     155|
+--------------------+--------+
only showing top 10 rows
17/03/04 10:34:52 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 200 blocks
```

Platform and Plugin Updates
The following plugin is ready to update: Scala

Compilation completed successfully with 1 warning in 1s 985ms (4 minutes ago)



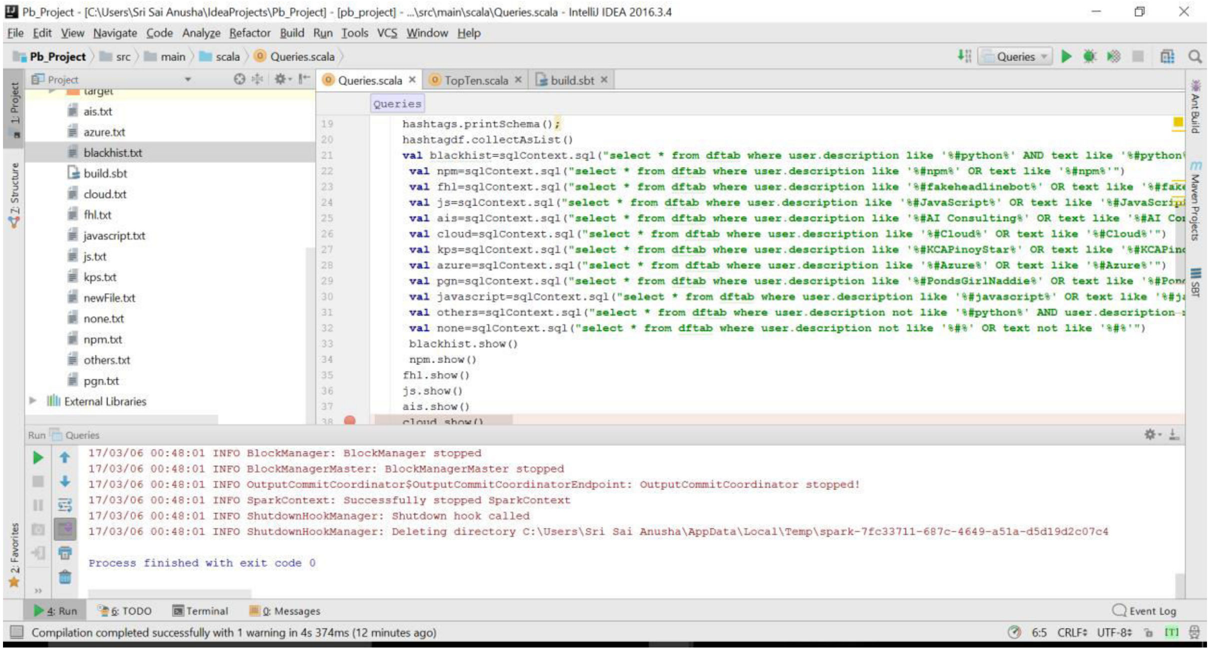
```
    /*
    val op = (
      sqlContext.sql("select row_number() over (GROUP BY hashtag order by hashtag) id,hashtag from disCatresult")
    )
    */
    op.createOrReplaceTempView("result")

    //val number=sqlContext.sql("create temporary table numbers as (select 1 as n union select 2 as n union select 3
    // val output=sqlContext.sql("select email, count(1) from (select id,substring_index(substring_index(hashtag, ',',

    // output.show()
    op.show()
    val output = sqlContext.sql("select * from result ORDER BY `count(1)` DESC")
    output.createOrReplaceTempView("top")
    output.show(16)


  }
}
```

```
17/03/06 00:48:01 INFO BlockManager: BlockManager stopped
17/03/06 00:48:01 INFO BlockManagerMaster: BlockManagerMaster stopped
17/03/06 00:48:01 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
17/03/06 00:48:01 INFO SparkContext: Successfully stopped SparkContext
17/03/06 00:48:01 INFO ShutdownHookManager: Shutdown hook called
17/03/06 00:48:01 INFO ShutdownHookManager: Deleting directory C:\Users\Sri Sai Anusha\AppData\Local\Temp\spark-7fc33711-687c-4649-a51a-d5d19d2c07c4

Process finished with exit code 0
```

Compilation completed successfully with 1 warning in 4s 374ms (11 minutes ago)

5. HADOOP FILE DIRECTORIES

The following figure shows the queries written to obtain the top ten directories.



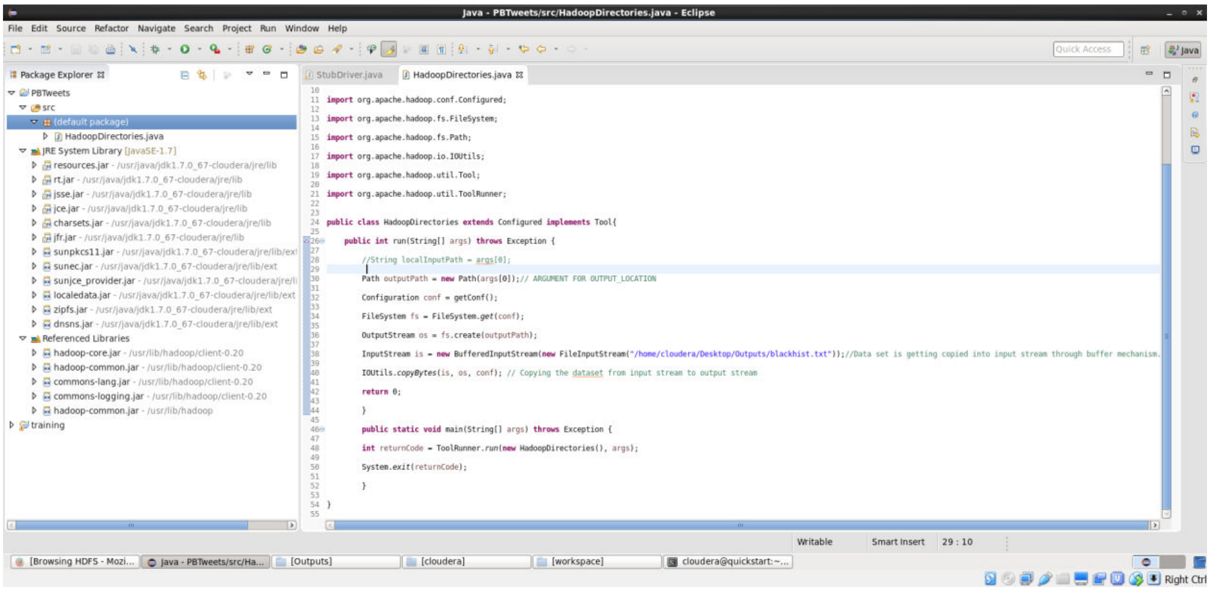
Pb_Project - [C:\Users\Sri Sai Anusha\IdeaProjects\Pb_Project] - [pb_project] - ...\src\main\scala\Queries.scala - IntelliJ IDEA 2016.3.4

```
    hashtags.printSchema();
    hashtagdf.collectAsList()
    val blackhist=sqlContext.sql("select * from dftab where user.description like '%#python%' AND text like '%#python
     val npm=sqlContext.sql("select * from dftab where user.description like '%#npm%' OR text like '%#npm%'")
     val fhl=sqlContext.sql("select * from dftab where user.description like '%#fakeheadlinebot%' OR text like '%#fake
     val js=sqlContext.sql("select * from dftab where user.description like '%#JavaScript%' OR text like '%#JavaScrip
     val ais=sqlContext.sql("select * from dftab where user.description like '%#AI Consulting%' OR text like '%#AI Co
     val cloud=sqlContext.sql("select * from dftab where user.description like '%#Cloud%' OR text like '%#Cloud%'")
     val kps=sqlContext.sql("select * from dftab where user.description like '%#KCAPinoyStar%' OR text like '%#KCAPin
     val azure=sqlContext.sql("select * from dftab where user.description like '%#Azure%' OR text like '%#Azure%'")
     val pgn=sqlContext.sql("select * from dftab where user.description like '%#PondsGirlNaddie%' OR text like '%#Pon
     val javascript=sqlContext.sql("select * from dftab where user.description like '%#javascript%' OR text like '%#ja
     val others=sqlContext.sql("select * from dftab where user.description not like '%#python%' AND user.description 
     val none=sqlContext.sql("select * from dftab where user.description not like '%#%' OR text not like '%#%'")
     blackhist.show()
     npm.show()
    fhl.show()
    js.show()
    ais.show()
    cloud.show()
```

```
17/03/06 00:48:01 INFO BlockManager: BlockManager stopped
17/03/06 00:48:01 INFO BlockManagerMaster: BlockManagerMaster stopped
17/03/06 00:48:01 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
17/03/06 00:48:01 INFO SparkContext: Successfully stopped SparkContext
17/03/06 00:48:01 INFO ShutdownHookManager: Shutdown hook called
17/03/06 00:48:01 INFO ShutdownHookManager: Deleting directory C:\Users\Sri Sai Anusha\AppData\Local\Temp\spark-7fc33711-687c-4649-a51a-d5d19d2c07c4

Process finished with exit code 0
```

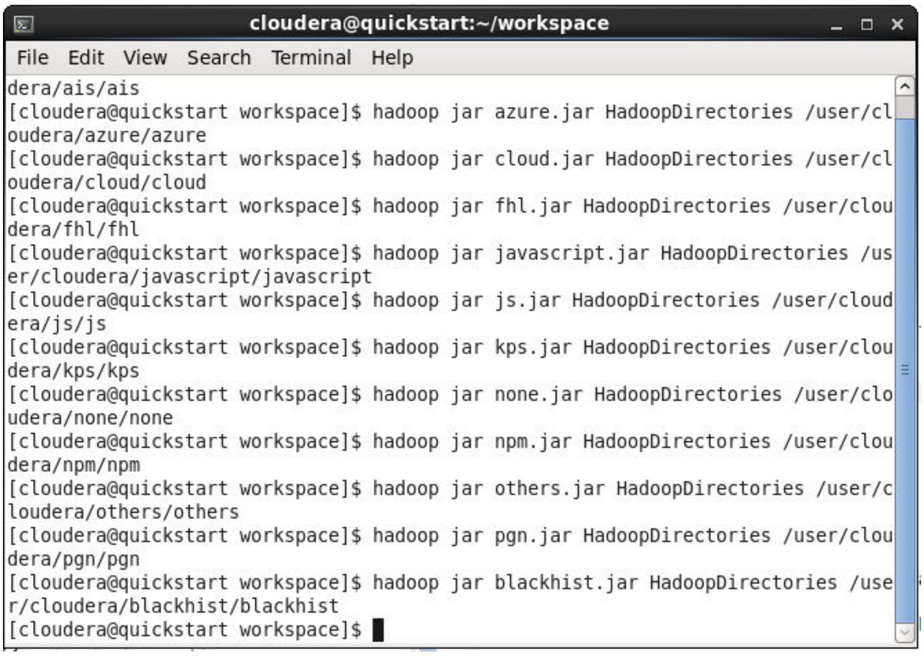
Compilation completed successfully with 1 warning in 4s 374ms (12 minutes ago)

Sample JAVA Code to create the directories



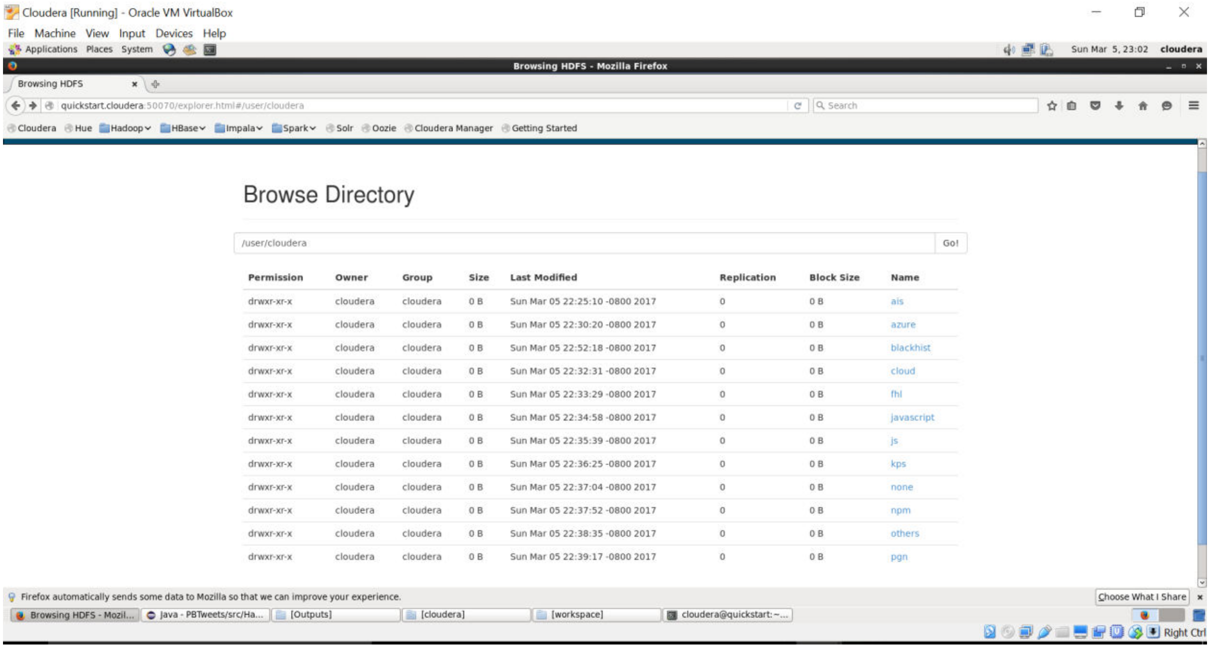
```java

import org.apache.hadoop.conf.Configured;

import org.apache.hadoop.fs.FileSystem;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.IOUtils;

import org.apache.hadoop.util.Tool;

import org.apache.hadoop.util.ToolRunner;


public class HadoopDirectories extends Configured implements Tool{

    public int run(String[] args) throws Exception {

        //String localInputPath = args[0];

        Path outputPath = new Path(args[0]);// ARGUMENT FOR OUTPUT_LOCATION

        Configuration conf = getConf();

        FileSystem fs = FileSystem.get(conf);

        OutputStream os = fs.create(outputPath);

        InputStream is = new BufferedInputStream(new FileInputStream("/home/cloudera/Desktop/Outputs/blackhist.txt"));//Data set is getting copied into input stream through buffer mechanism.

        IOUtils.copyBytes(is, os, conf); // Copying the dataset from input stream to output stream

        return 0;

        }

        public static void main(String[] args) throws Exception {

        int returnCode = ToolRunner.run(new HadoopDirectories(), args);

        System.exit(returnCode);

        }

}

```

Workspace


cloudera@quickstart:~/workspace
File Edit View Search Terminal Help

```
dera/ais/ais
[cloudera@quickstart workspace]$ hadoop jar azure.jar HadoopDirectories /user/cl
oudera/azure/azure
[cloudera@quickstart workspace]$ hadoop jar cloud.jar HadoopDirectories /user/cl
oudera/cloud/cloud
[cloudera@quickstart workspace]$ hadoop jar fhl.jar HadoopDirectories /user/clou
dera/fhl/fhl
[cloudera@quickstart workspace]$ hadoop jar javascript.jar HadoopDirectories /us
er/cloudera/javascript/javascript
[cloudera@quickstart workspace]$ hadoop jar js.jar HadoopDirectories /user/cloud
era/js/js
[cloudera@quickstart workspace]$ hadoop jar kps.jar HadoopDirectories /user/clou
dera/kps/kps
[cloudera@quickstart workspace]$ hadoop jar none.jar HadoopDirectories /user/clo
udera/none/none
[cloudera@quickstart workspace]$ hadoop jar npm.jar HadoopDirectories /user/clou
dera/npm/npm
[cloudera@quickstart workspace]$ hadoop jar others.jar HadoopDirectories /user/c
loudera/others/others
[cloudera@quickstart workspace]$ hadoop jar pgn.jar HadoopDirectories /user/clou
dera/pgn/pgn
[cloudera@quickstart workspace]$ hadoop jar blackhist.jar HadoopDirectories /use
r/cloudera/blackhist/blackhist
[cloudera@quickstart workspace]$ █
```

Hadoop File Directories


Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:02 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Choose What I Share
Browsing HDFS - Mozil...
Java - PBTweets/src/Ha...
[Outputs]
[cloudera]
[workspace]
cloudera@quickstart:~...
Right Ctrl

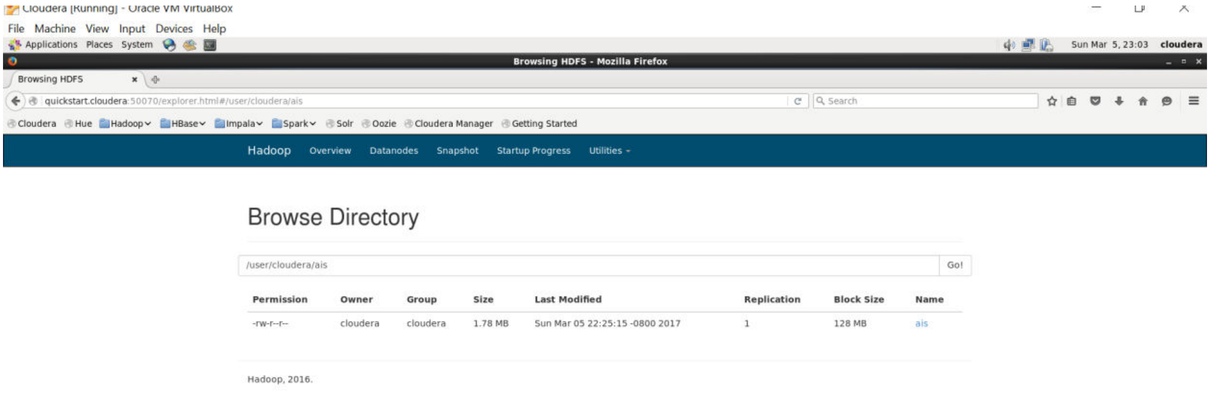
Browse Directory

/user/cloudera Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:25:10 -0800 2017	0	0 B	ais
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:30:20 -0800 2017	0	0 B	azure
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:52:18 -0800 2017	0	0 B	blackhist
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:32:31 -0800 2017	0	0 B	cloud
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:33:29 -0800 2017	0	0 B	fhl
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:34:58 -0800 2017	0	0 B	javascript
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:35:39 -0800 2017	0	0 B	js
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:36:25 -0800 2017	0	0 B	kps
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:37:04 -0800 2017	0	0 B	none
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:37:52 -0800 2017	0	0 B	npm
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:38:35 -0800 2017	0	0 B	others
drwxr-xr-x	cloudera	cloudera	0 B	Sun Mar 05 22:39:17 -0800 2017	0	0 B	pgn

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Browsing Each Directory Including Others and None

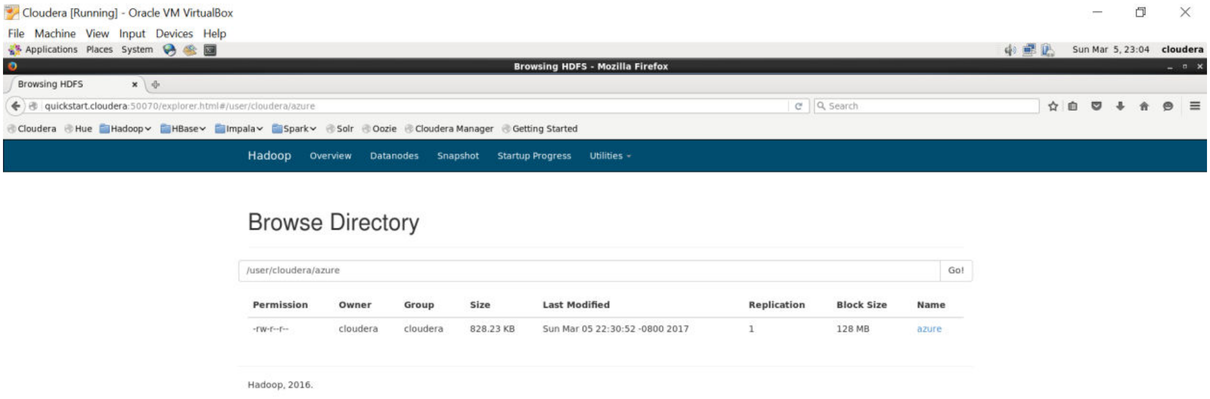

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:03 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/ais
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/ais Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	1.78 MB	Sun Mar 05 22:25:15 -0800 2017	1	128 MB	ais

Hadoop, 2016.

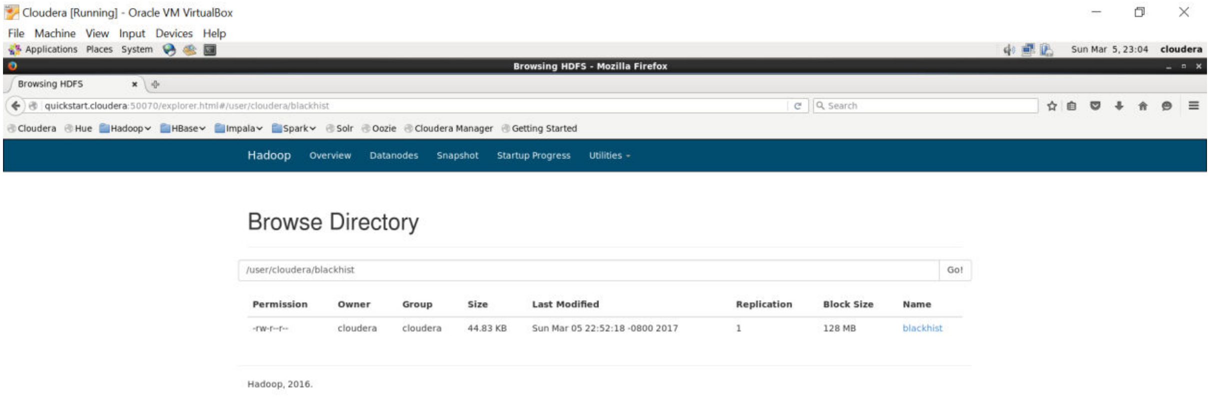

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:04 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/azure
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/azure Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	828.23 KB	Sun Mar 05 22:30:52 -0800 2017	1	128 MB	azure

Hadoop, 2016.

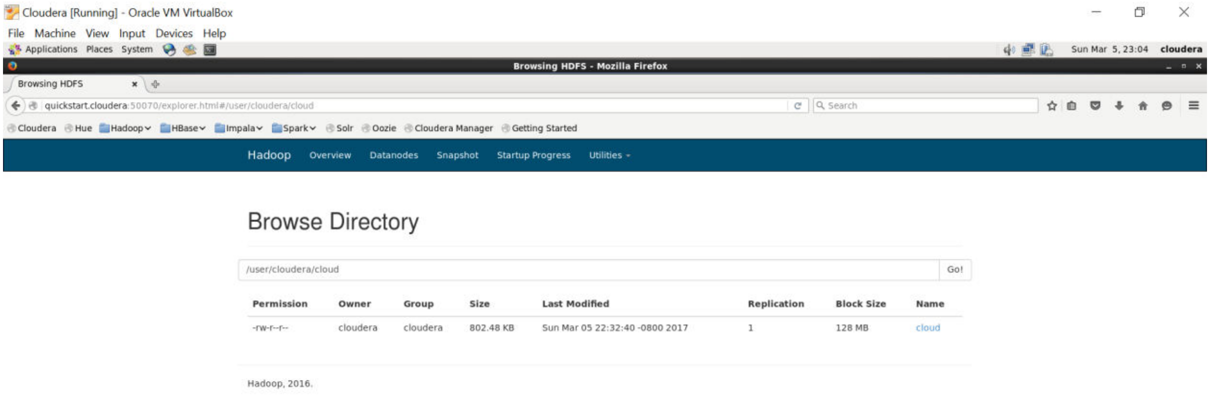

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:04 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/blackhist
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/blackhist Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	44.83 KB	Sun Mar 05 22:52:18 -0800 2017	1	128 MB	blackhist

Hadoop, 2016.

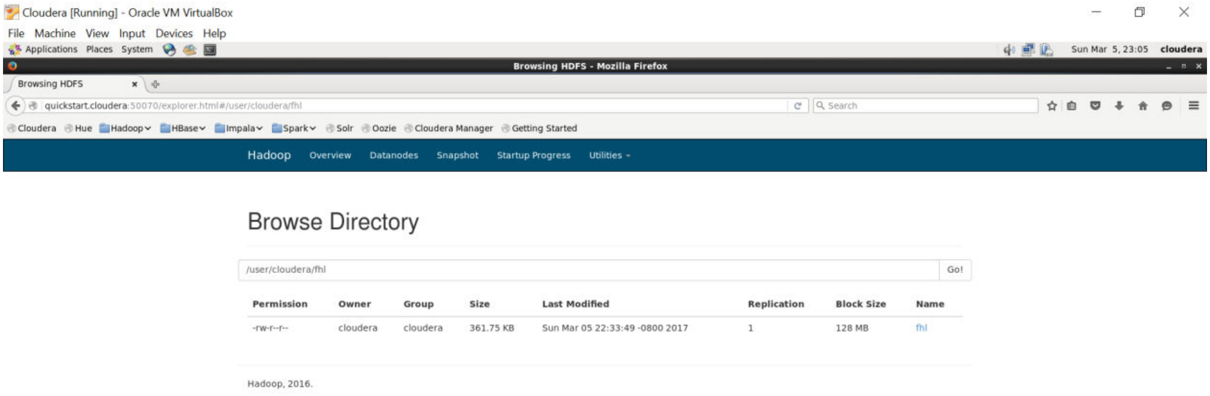

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:04 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/cloud
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/cloud Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	802.48 KB	Sun Mar 05 22:32:40 -0800 2017	1	128 MB	cloud

Hadoop, 2016.

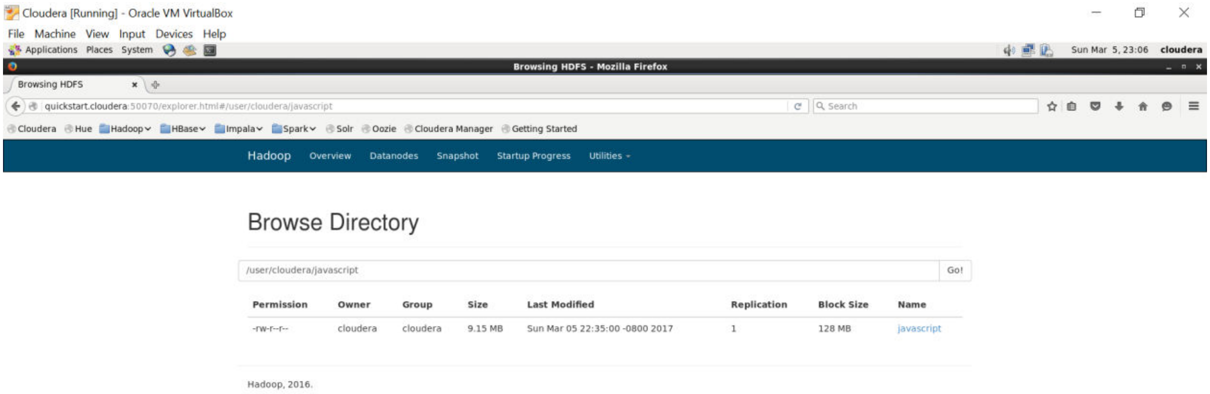

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:05 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/fhl
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/fhl Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	361.75 KB	Sun Mar 05 22:33:49 -0800 2017	1	128 MB	fhl

Hadoop, 2016.

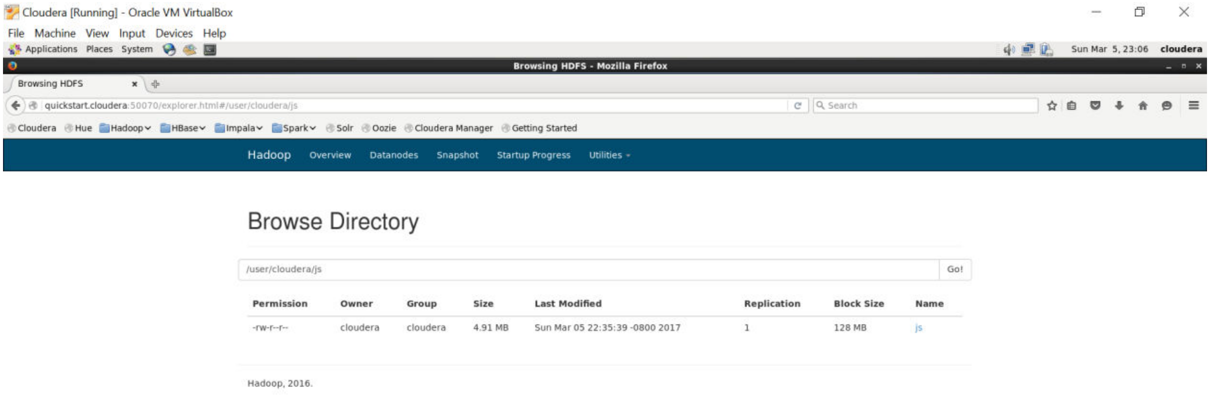

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:06 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/javascript
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/javascript Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	9.15 MB	Sun Mar 05 22:35:00 -0800 2017	1	128 MB	javascript

Hadoop, 2016.

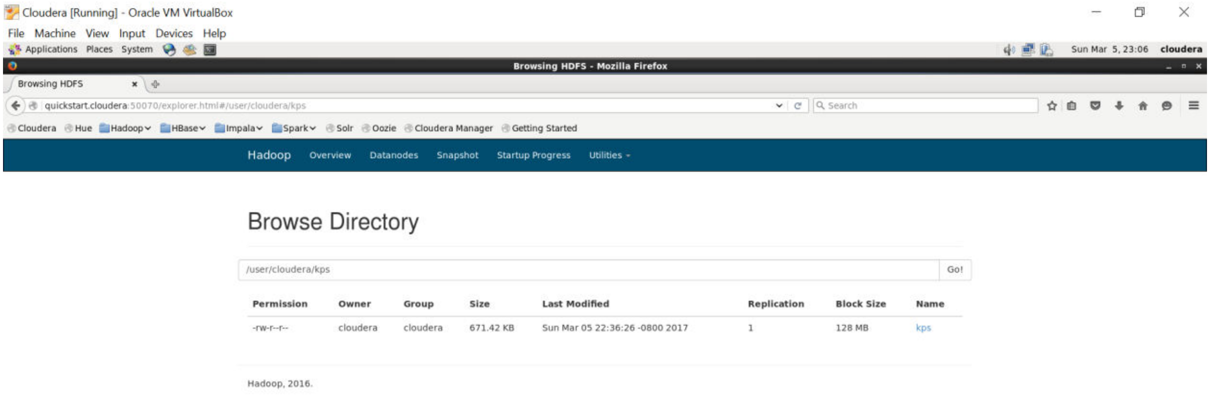

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:06 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/js
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/js Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	4.91 MB	Sun Mar 05 22:35:39 -0800 2017	1	128 MB	js

Hadoop, 2016.

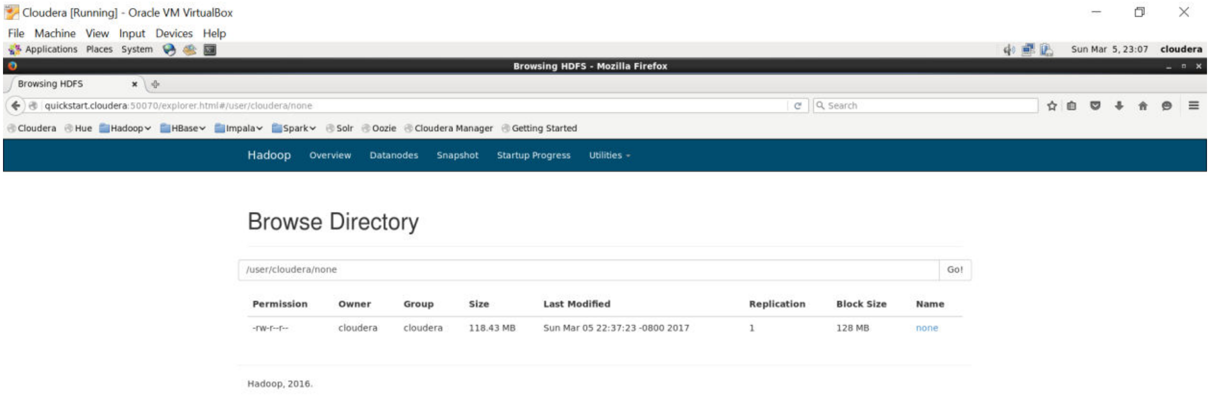

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:06 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/kps
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/kps Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	671.42 KB	Sun Mar 05 22:36:26 -0800 2017	1	128 MB	kps

Hadoop, 2016.

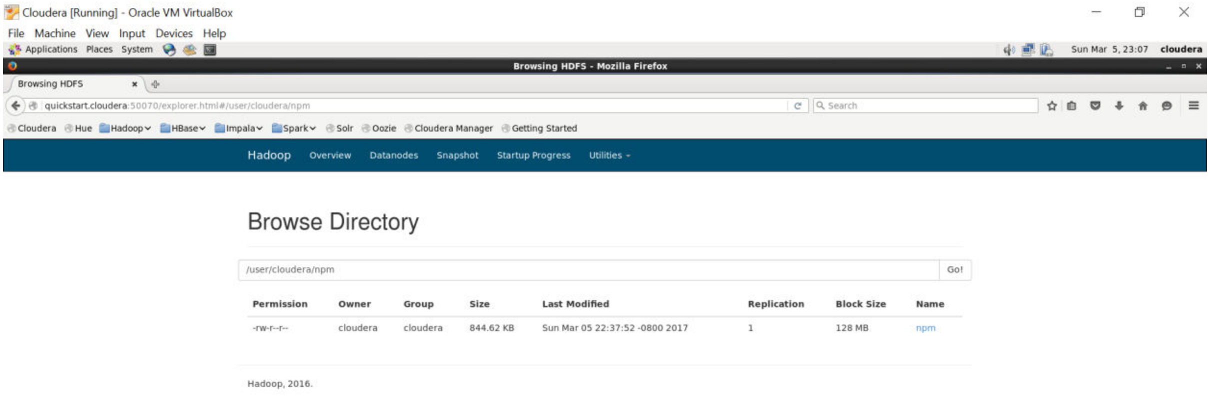

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:07 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/none
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/none Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	118.43 MB	Sun Mar 05 22:37:23 -0800 2017	1	128 MB	none

Hadoop, 2016.

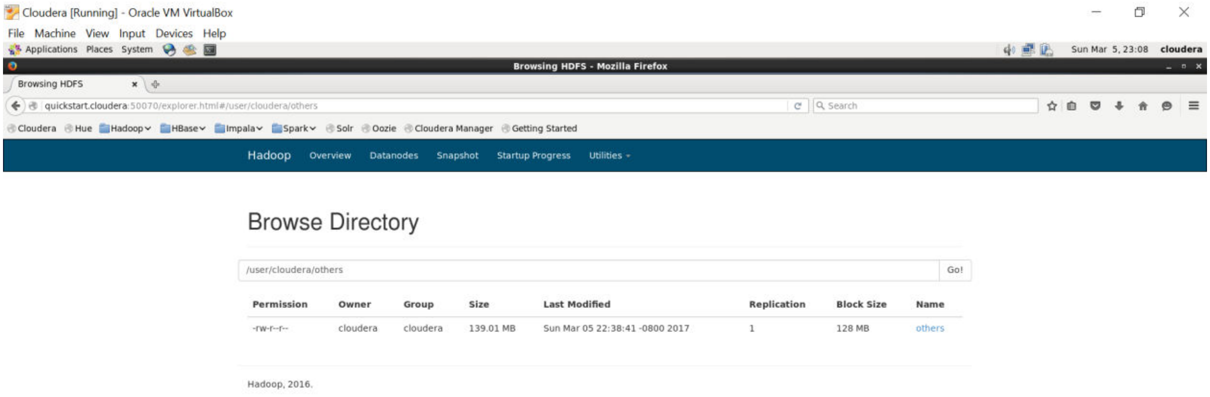

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:07 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/npm
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/npm Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	844.62 KB	Sun Mar 05 22:37:52 -0800 2017	1	128 MB	npm

Hadoop, 2016.

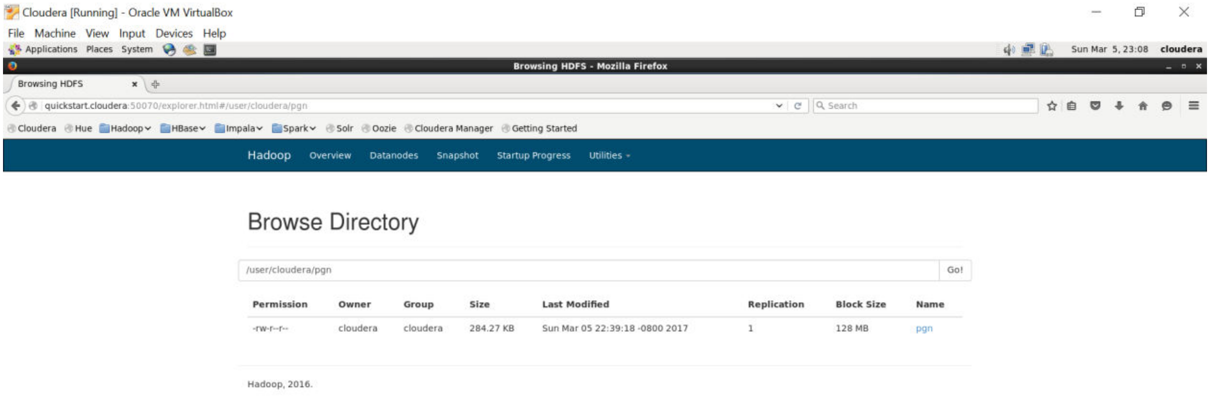

Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:08 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/others
Search
Cloudera Hue Hadoop Hbase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities
Go!

Browse Directory

/user/cloudera/others

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	139.01 MB	Sun Mar 05 22:38:41 -0800 2017	1	128 MB	others

Hadoop, 2016.


Cloudera [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 5, 23:08 cloudera
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/cloudera/pgn
Search
Cloudera Hue Hadoop Hbase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities
Go!

Browse Directory

/user/cloudera/pgn

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	284.27 KB	Sun Mar 05 22:39:18 -0800 2017	1	128 MB	pgn

Hadoop, 2016.