

Hdfs Task

1. Import RStudio Log Files from one week in Jun 2021 into HDFS
2. Step-1 Download one week Logs from <http://cran-logs.rstudio.com/>
3. Step-2 Create a new folder to store downloaded logs files.mkdir RLogFiles
4. Step-3 Verify files in RStudioLogs folder ls
5. Step-4 Unzip all *.zip files gunzip *.csv.gz
6. Step-5 Put on ?HDFS and verify hdfs dfs -mkdir /user/master
7. hdfs dfs -mkdir /user/master/RLogFiles/
8. hdfs dfs -put *.csv /user/master/RLogFiles
9. hdfs dfs -ls /user/master/RlogFiles/RLogfile

Hdfs Task

```
Hue - Editor - Mozilla Firefox
cloudera@quickstart:~
File Edit View Search Terminal Help
drwxr-xr-x - cloudera cloudera 0 2021-05-27 06:45 /user/cloudera
drwxr-xr-x - mapred hadoop 0 2017-07-19 05:34 /user/history
drwxrwxrwx - hive supergroup 0 2017-07-19 05:36 /user/hive
drwxrwxrwx - hue supergroup 0 2017-07-19 05:35 /user/hue
drwxrwxrwx - jenkins supergroup 0 2017-07-19 05:35 /user/jenkins
drwxr-xr-x - cloudera supergroup 0 2021-06-01 02:50 /user/mydir
drwxr-xr-x - cloudera supergroup 0 2021-06-04 09:39 /user/mydire
drwxrwxrwx - oozie supergroup 0 2017-07-19 05:35 /user/oozie
drwxrwxrwx - root supergroup 0 2017-07-19 05:35 /user/root
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user/spark
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/master
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/master/RLogFiles
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/RLogfiles /user/master/RLogFiles
put: '/home/cloudera/RLogfiles': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Documents/RLogfiles /user/master/RLogFiles
[cloudera@quickstart ~]$ hdfs dfs -ls /user/master/RLogFiles
Found 1 items
drwxr-xr-x - cloudera supergroup 0 2021-06-07 08:06 /user/master/RLogFiles/RLogfiles
[cloudera@quickstart ~]$ hdfs dfs -ls /user/master/RLogFiles/RLogfiles
Found 5 items
-rw-r--r-- 1 cloudera supergroup 426181748 2021-06-07 08:05 /user/master/RLogFiles/RLogfiles/2021-06-01.csv
```

Query History

Saved Queries

You don't have any saved query.

```
Hue - Editor - Mozilla Firefox
cloudera@quickstart:~
File Edit View Search Terminal Help
Found 5 items
-rw-r--r-- 1 cloudera supergroup 426181748 2021-06-07 08:05 /user/master/RLogFiles/RLogfiles/2021-06-01.csv
-rw-r--r-- 1 cloudera supergroup 435604176 2021-06-07 08:05 /user/master/RLogFiles/RLogfiles/2021-06-02.csv
-rw-r--r-- 1 cloudera supergroup 418707479 2021-06-07 08:06 /user/master/RLogFiles/RLogfiles/2021-06-03.csv
-rw-r--r-- 1 cloudera supergroup 366312996 2021-06-07 08:06 /user/master/RLogFiles/RLogfiles/2021-06-04.csv
-rw-r--r-- 1 cloudera supergroup 225414219 2021-06-07 08:06 /user/master/RLogFiles/RLogfiles/2021-06-05.csv
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create table dataone(date string,time string,size int,version float,os string,country string,ip_id int)row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 95.767 seconds
hive> desc dataone;
OK
date string
```

Query History

Saved Queries

Hdfs Task

```

Hue - Editor - Mozilla Firefox
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> create table datatwo(date string,time string,size int,r_version float,r_ar
ch int,os string,package string,version float,country string,ip_id int)row forma
t delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.392 seconds
hive> load data local inpath '/home/cloudera/Documents/RLogfiles/2021-06-01.csv'
into table datatwo;
Loading data to table default.datatwo
Table default.datatwo stats: [numFiles=1, totalSize=426181748]
OK
Time taken: 34.364 seconds
hive> select * from datatwo limit 10;
OK
"date" "time" NULL NULL NULL "r_os" "package" NULL "country
" NULL
"2021-06-01" "00:00:22" 91753 NULL NULL NA "ids" NULL "
US" 1
"2021-06-01" "00:00:22" 81589 NULL NULL NA "markdown" N
ULL "US" 2
"2021-06-01" "00:00:14" 202330 NULL NULL NA "labelled" N
ULL "US" 3
"2021-06-01" "00:00:12" 2013498 NULL NULL NA "isoband" N
ULL "US" 3
"2021-06-01" "00:00:22" 112236 NULL NULL NA "xfun" NULL "

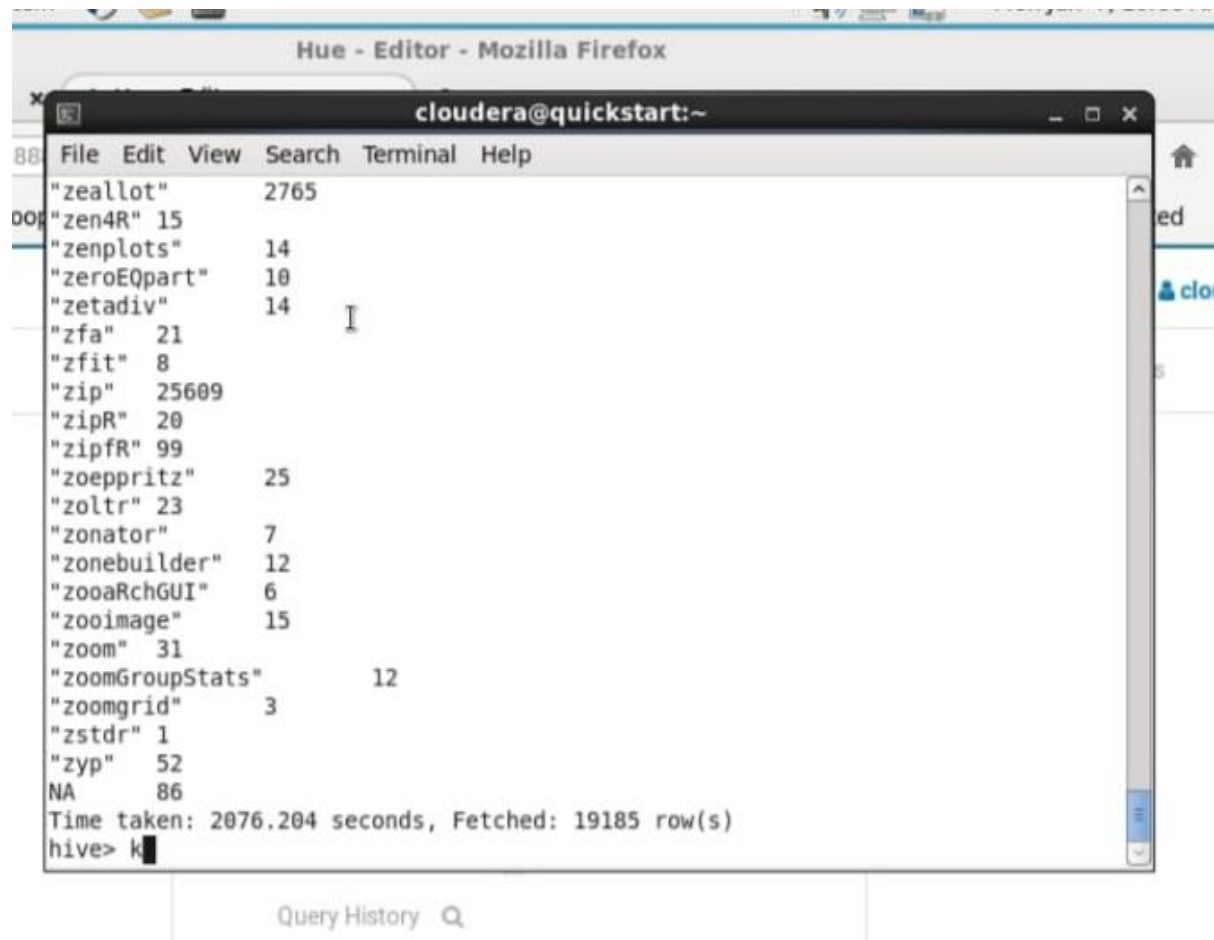
```

```

Hue - Editor - Mozilla Firefox
cloudera@quickstart:~
File Edit View Search Terminal Help
ULL "US" 3
"2021-06-01" "00:00:22" 112236 NULL NULL NA "xfun" NULL "
US" 4
"2021-06-01" "00:00:21" 141964 NULL NULL NA "processx" N
ULL "US" 1
"2021-06-01" "00:00:24" 283373 NULL NULL NA "haven" NULL "
US" 5
"2021-06-01" "00:00:22" 50083 NULL NULL NA "uuid" NULL "
US" 6
"2021-06-01" "00:00:14" 38099 NULL NULL NA "prettyunits" N
ULL "MX" 7
Time taken: 0.288 seconds, Fetched: 10 row(s)
hive> select package,count(*) from datatwo group by package;
Query ID = cloudera_20210607093030_df26102e-1785-48a1-a898-edbfc7fb176b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 2
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1623072835592_0001, Tracking URL = http://quickstart.cloudera

```

Hdfs Task



The screenshot shows a web browser window titled "Hue - Editor - Mozilla Firefox". Inside the browser, a terminal window titled "cloudera@quickstart:~" is open. The terminal displays the output of a Hive query, showing a list of files and their sizes. The output is as follows:

| File | Size |
|------------------|-------|
| "zeallot" | 2765 |
| "zen4R" | 15 |
| "zenplots" | 14 |
| "zeroEQpart" | 10 |
| "zetadiv" | 14 |
| "zfa" | 21 |
| "zfit" | 8 |
| "zip" | 25609 |
| "zipR" | 20 |
| "zipfR" | 99 |
| "zoeppritz" | 25 |
| "zoltr" | 23 |
| "zonator" | 7 |
| "zonebuilder" | 12 |
| "zoaRchGUI" | 6 |
| "zooimage" | 15 |
| "zoom" | 31 |
| "zoomGroupStats" | 12 |
| "zoomgrid" | 3 |
| "zstdr" | 1 |
| "zyp" | 52 |
| NA | 86 |

Time taken: 2076.204 seconds, Fetched: 19185 row(s)
hive> k

Below the terminal window, there is a "Query History" section with a search icon.

Hdfs Task

Hue - Editor - Mozilla Firefox

cloudera@quickstart:~

File Edit View Search Terminal Help

| | |
|----------------|-------|
| "time2event" | 6 |
| "timeDate" | 5515 |
| "timeR" | 11 |
| "timeROC" | 35 |
| "timechange" | 16 |
| "timedeppar" | 13 |
| "timeline" | 13 |
| "timelineR" | 16 |
| "timeperiodsR" | 55 |
| "timereg" | 515 |
| "timetree" | 11 |
| "timevis" | 58 |
| "timsac" | 101 |
| "timsr" | 13 |
| "tinsel" | 11 |
| "tint" | 31 |
| "tinyProject" | 13 |
| "tinylabels" | 20 |
| "tinytex" | 34332 |
| "tiobeindexr" | 3 |
| "tipitaka" | 9 |
| "tipom" | 20 |
| "tipr" | 14 |
| "titeIR" | 13 |

Query History