***Name: Sovichea Cheth***

***ID: 985421***

**Lab 7**

**Sqoop Homework**

First, I run below command to check the MySql version:

[cloudera@quickstart ~]$ mysql --version

mysql Ver 14.14 Distrib 5.1.73, for redhat-linux-gnu (x86\_64) using readline 5.1

After that, login as root:

[cloudera@quickstart ~]$ mysql -u root -p

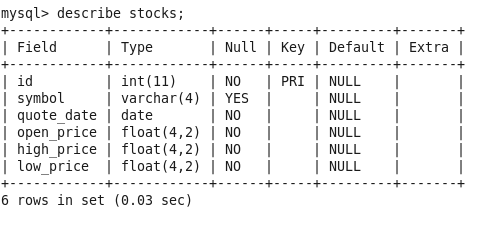
Then I created the table stock and insert the records as required.

mysql> show databases;

mysql> use cs523;

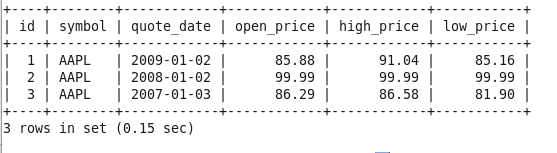
mysql> CREATE TABLE stocks (id INT NOT NULL primary key, symbol VARCHAR(4), quote\_date DATE NOT NULL, open\_price FLOAT(4,2) NOT NULL, high\_price FLOAT(4,2) NOT NULL, low\_price FLOAT(4,2) NOT NULL);

mysql> describe stocks;



mysql> INSERT INTO stocks VALUES (1, 'AAPL', '2009-01-02', 85.88, 91.04, 85.16), (2, 'AAPL', '2008-01-02', 199.27, 200.26, 192.55), (3, 'AAPL', '2007-01-03', 86.29, 86.58, 81.9);

mysql> SELECT \* FROM stocks;

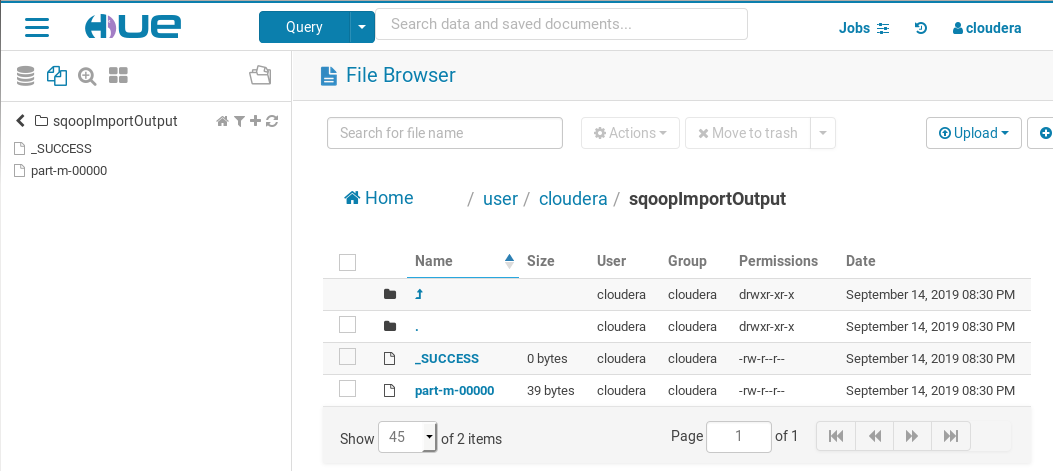


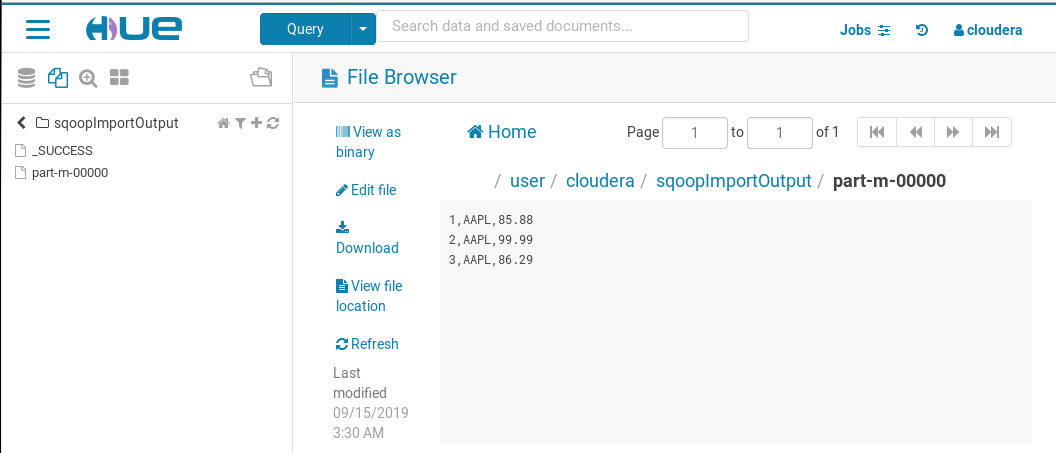
mysql> quit;

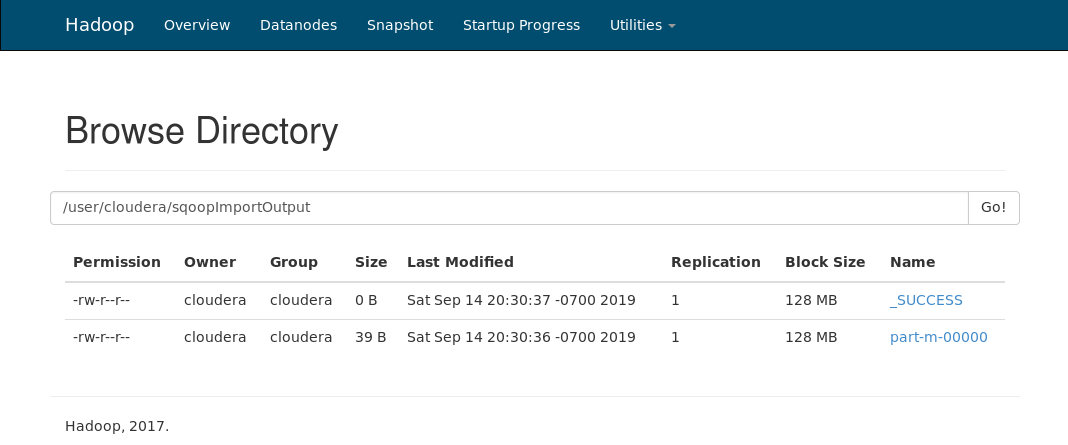
After that, import the stocks table into HDFS:

[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/cs523 --username root -P --table stocks --columns "id,symbol,open\_price" --target-dir=/user/cloudera/sqoopImportOutput -m 1

Screenshot of the file part-m after import the table into HDFS:







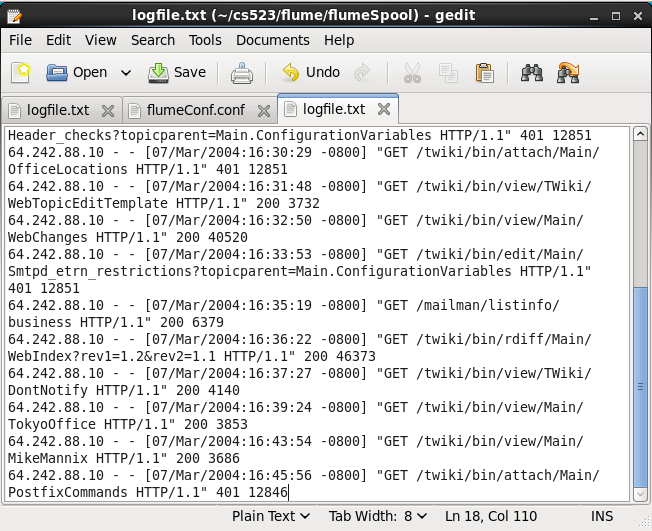
**Flume Homework**

First, I have created the sample log file ‘logfile.txt’ inside directory /home/clouder/cs523/flume/flumeSpool

File name:

/home/clouder/cs523/flume/flumeSpool/logfile.txt

With the sample log as below screen shot:



After that, I create the flume configuration file ‘flumeConf.conf’ inside directory as below:

/home/cloudera/cs523/flume/conf/flumeConf.conf

The flumeConf.conf file has the content as below:

agent1.sources = source1

agent1.sinks = sink1a sink1b

agent1.channels = channel1a channel1b

agent1.sources.source1.channels = channel1a channel1b

agent1.sinks.sink1a.channel = channel1a

agent1.sinks.sink1b.channel = channel1b

agent1.sources.source1.type = spooldir

agent1.sources.source1.spoolDir = /home/cloudera/cs523/flume/flumeSpool

agent1.sinks.sink1a.type = hdfs

agent1.sinks.sink1a.hdfs.path = hdfs://localhost/user/cloudera/flumeImport/

agent1.sinks.sink1a.hdfs.filePrefix = flumeFan

agent1.sinks.sink1a.hdfs.fileType = DataStream

agent1.sinks.sink1b.type = logger

agent1.channels.channel1a.type = file

agent1.channels.channel1b.type = memory

After that, execute the command as below:

[cloudera@quickstart ~]$ flume-ng agent -n agent1 -c /home/cloudera/cs523/flume/conf/ -f /home/cloudera/cs523/flume/conf/flumeConf.conf

Screenshot:

