

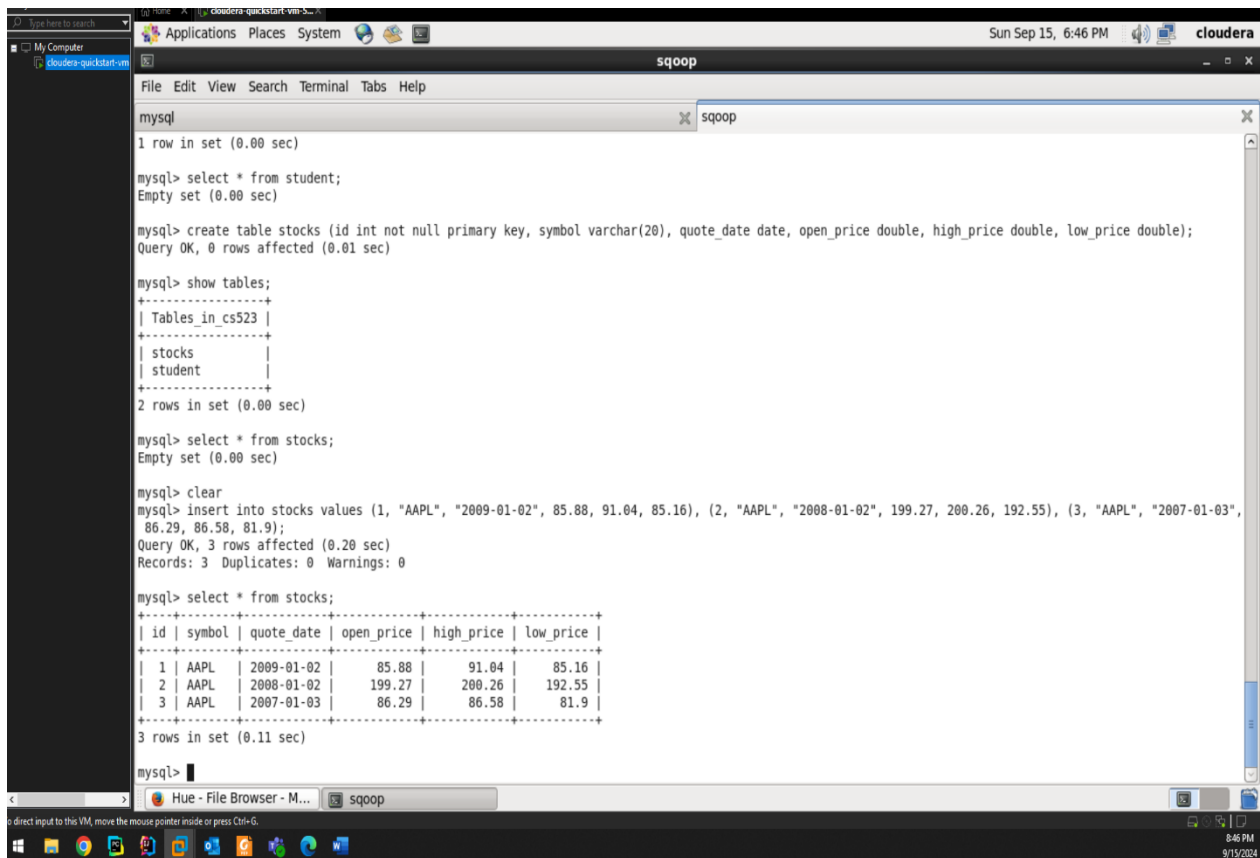
## Lab7. Sqoop and Flume

### #Creating Table

```
mysql> create table stocks (id int not null primary key, symbol varchar(20), quote_date date, open_price double, high_price double, low_price double);
```

### #Inserting into the table

```
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);
```



The screenshot shows a MySQL terminal window with the following commands and output:

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

mysql> create table stocks (id int not null primary key, symbol varchar(20), quote_date date, open_price double, high_price double, low_price double);
Query OK, 0 rows affected (0.01 sec)

mysql> show tables;
+-----+
| Tables_in_cs523 |
+-----+
| stocks          |
| student         |
+-----+
2 rows in set (0.00 sec)

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

mysql> clear
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);
Query OK, 3 rows affected (0.20 sec)
Records: 3 Duplicates: 0 Warnings: 0

mysql> select * from stocks;
+----+-----+-----+-----+-----+-----+
| id | symbol | quote_date | open_price | high_price | low_price |
+----+-----+-----+-----+-----+-----+
| 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      |
+----+-----+-----+-----+-----+-----+
3 rows in set (0.11 sec)

mysql>
```

id	symbol	quote_date	open_price	high_price	low_price
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

#Import table in Tab Seperated format into HDFS

```
sqoop import --connect jdbc:mysql://localhost/cs523 --username root -P --table stocks --  
columns id,symbol,open_price --where "low_price<100" --fields-terminated-by "\t" --target-  
dir=/user/cloudera/sqoopImportOutput -m 1
```

The screenshot shows the Hue File Browser interface in a Mozilla Firefox browser window. The address bar displays the URL: `quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/sqoopImportOutput/part-m-00000`. The interface includes a sidebar with navigation options like 'Impala', 'Databases', and 'default'. The main content area shows the file path `/ user / cloudera / sqoopImportOutput / part-m-00000` and a table of data. The table has two rows of data:

1	AAPL	85.88
3	AAPL	86.29

Below the table, it indicates 'Last modified: 09/16/2024'. The bottom of the screen shows a taskbar with various application icons and a system clock indicating 'Sun Sep 15, 7:03 PM'.

## 2. Flume

### #Flume configuration

```
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

agent1.sinks.sink1a.type = hdfs
agent1.sinks.sink1a.hdfs.path = hdfs://localhost/user/cloudera/flume
agent1.sinks.sink1a.hdfs.filePrefix = events
agent1.sinks.sink1a.hdfs.fileSuffix = .log
agent1.sinks.sink1a.hdfs.inUsePerfix = _
agent1.sinks.sink1a.hdfs.fileType = DataStream

agent1.sinks.sink1b.type = logger

agent1.channels.channel1a.type = file
agent1.channels.channel1b.type = memory
```

Apps

Editor

Dashboard

Scheduler

Browsers

Documents

Files

Tables

Indexes

Jobs

HBase

Security

(1) 🔍 + ↻

Search for file name

⚙️ Actions ▾

✖ Move to trash ▾

🏠 Home

/ user / cloudera

<input type="checkbox"/>	Name	Size	User	Group	Permissions
<input type="checkbox"/>	📁 ↗		hdfs	supergroup	drwxr-xr-x
<input type="checkbox"/>	📁 .		cloudera	cloudera	drwxr-xr-x
<input type="checkbox"/>	📁 emp523		cloudera	cloudera	drwxr-xr-x
<input type="checkbox"/>	📁 flume		cloudera	cloudera	drwxr-xr-x
<input type="checkbox"/>	📁 input		cloudera	cloudera	drwxr-xr-x
<input type="checkbox"/>	📁 output		cloudera	cloudera	drwxr-xr-x

Show 45 ▾ of 4 items

Page 1 of 1

⏪

Apps

Editor

Dashboard

Scheduler

Browsers

Documents

Files

Tables

Indexes

Jobs

HBase

Security

(1) 🔍 + ↻

📁 File Browser

Search for file name

⚙️ Actions ▾

✖ Move to trash ▾

🏠 Home

/ user / cloudera / flume

<input type="checkbox"/>	Name	Size	User	Group	Permissions
<input type="checkbox"/>	📁 ↗		cloudera	cloudera	drwxr-xr-x
<input type="checkbox"/>	📁 .		cloudera	cloudera	drwxr-xr-x
<input type="checkbox"/>	📄 events.1681764368107.log	96 bytes	cloudera	cloudera	-rw-r--r--
<input type="checkbox"/>	📄 events.1681764368108.log	84 bytes	cloudera	cloudera	-rw-r--r--
<input type="checkbox"/>	📄 events.1681764368109.log	15 bytes	cloudera	cloudera	-rw-r--r--

Show 45 ▾ of 3 items

Page 1 of 1

⏪

arch data and saved documents...

wser

[Home](#)

Page 1 to 1 of 1

[/ user](#) / [cloudera](#) / [flume](#) / **events.1681764368107.log**

1

Peter Herman  
Jeff Keagan  
John Smith  
Tom Ally  
Zac Michel  
Marc Elton  
Jane Jeanet  
Marie Louise