Part 1 Sqoop

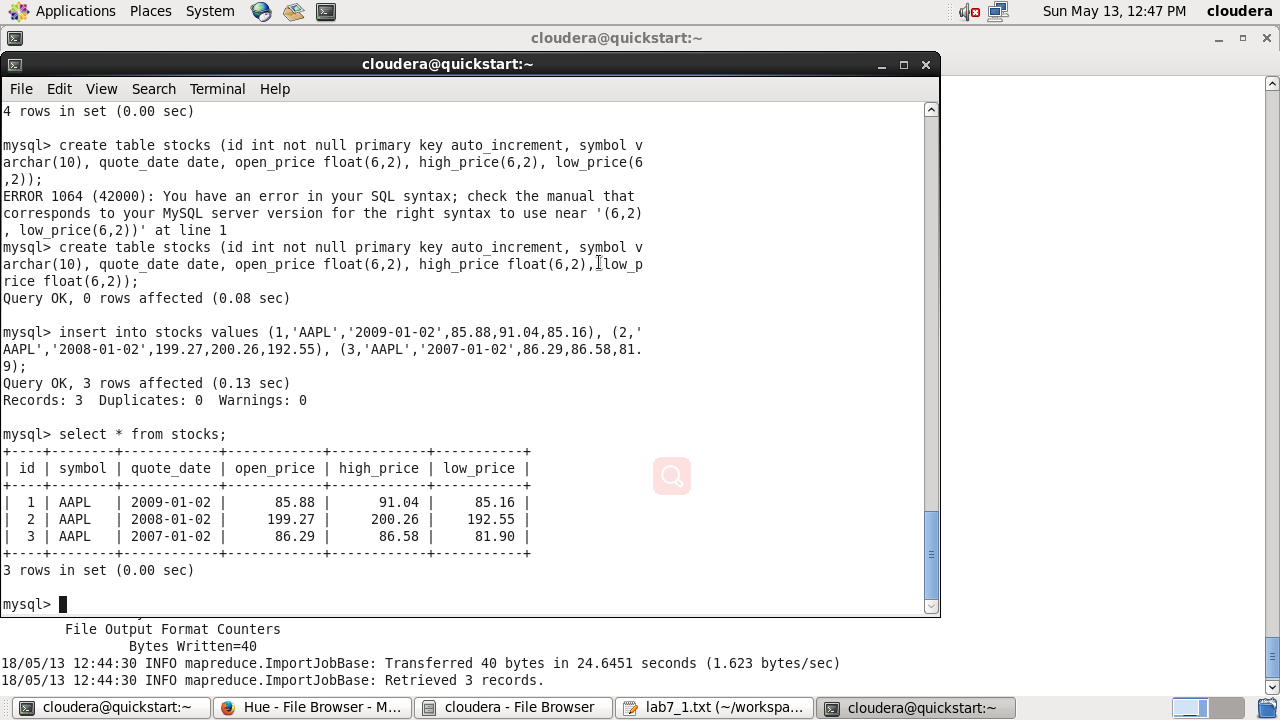
mysql -u root -p

use cs523;

//create table

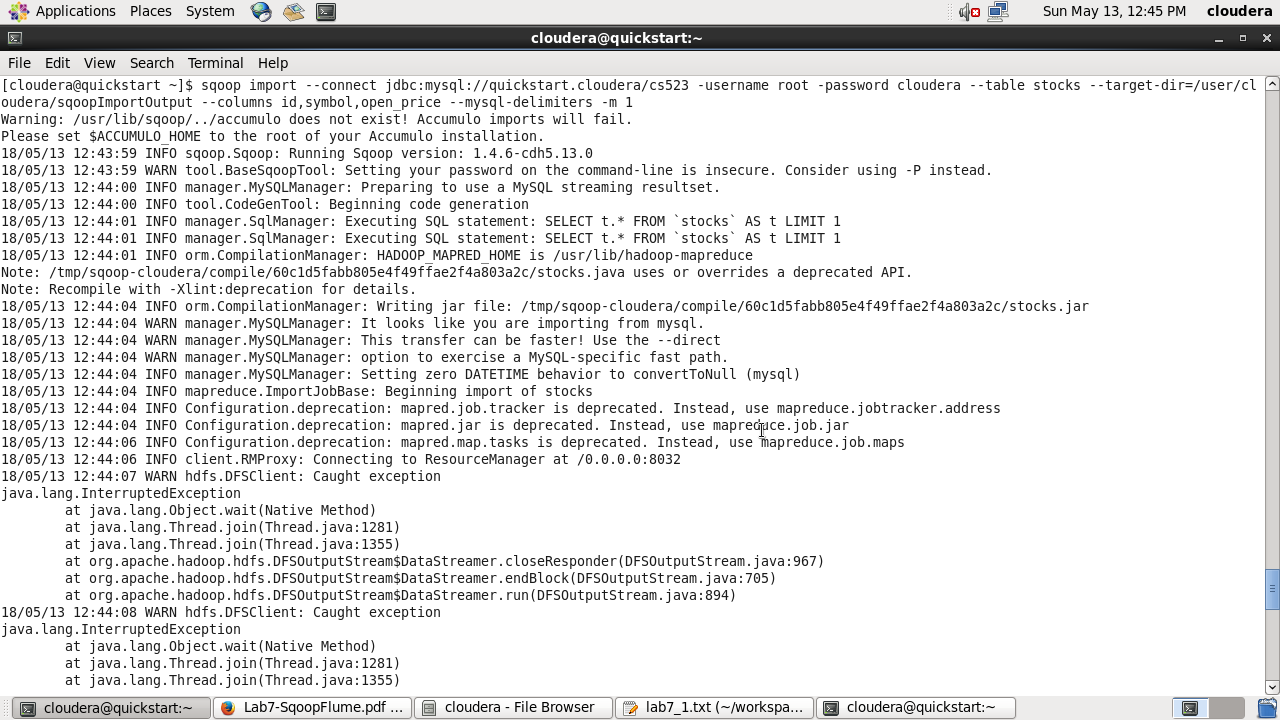
create table stocks (id int not null primary key auto\_increment, symbol varchar(10), quote\_date date, open\_price float(6,2), high\_price float(6,2), low\_price float(6,2));

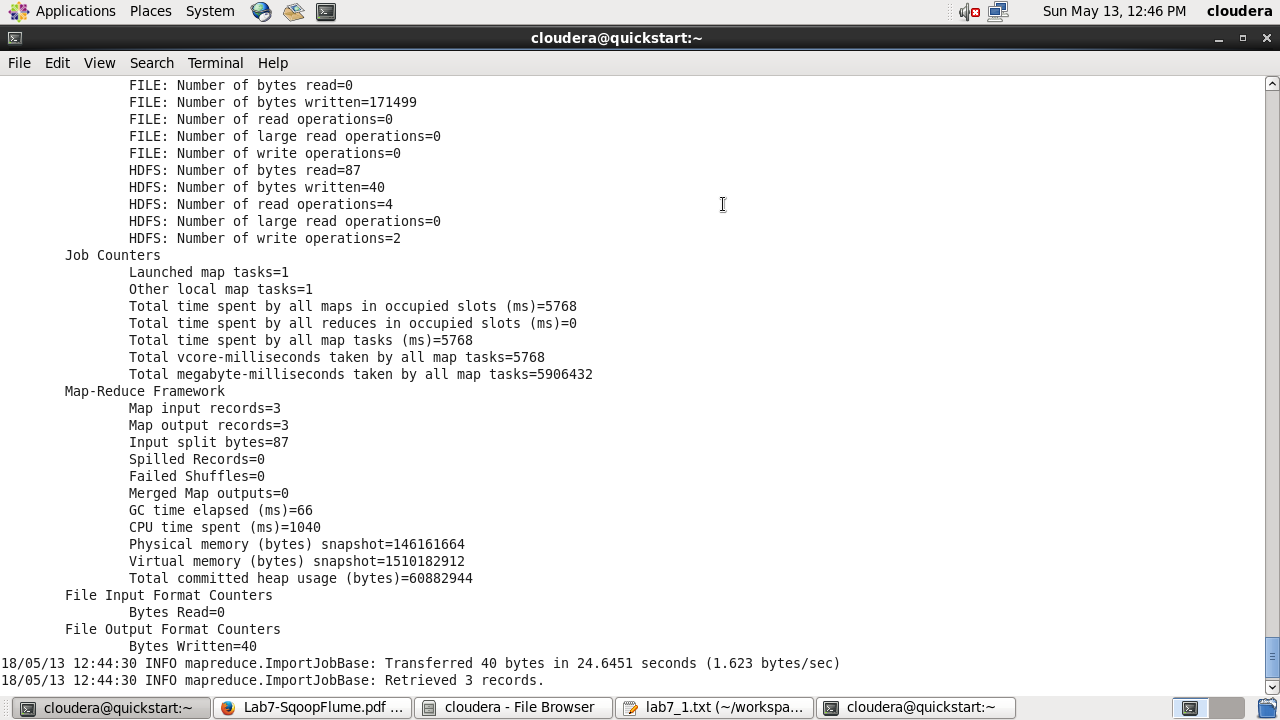
//insert data

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','2007-01-02',86.29,86.58,81.9); 

//import

sqoop import --connect jdbc:mysql://quickstart.cloudera/cs523 -username root -password cloudera --table stocks --target-dir=/user/cloudera/sqoopImportOutput --columns id,symbol,open\_price --mysql-delimiters -m 1

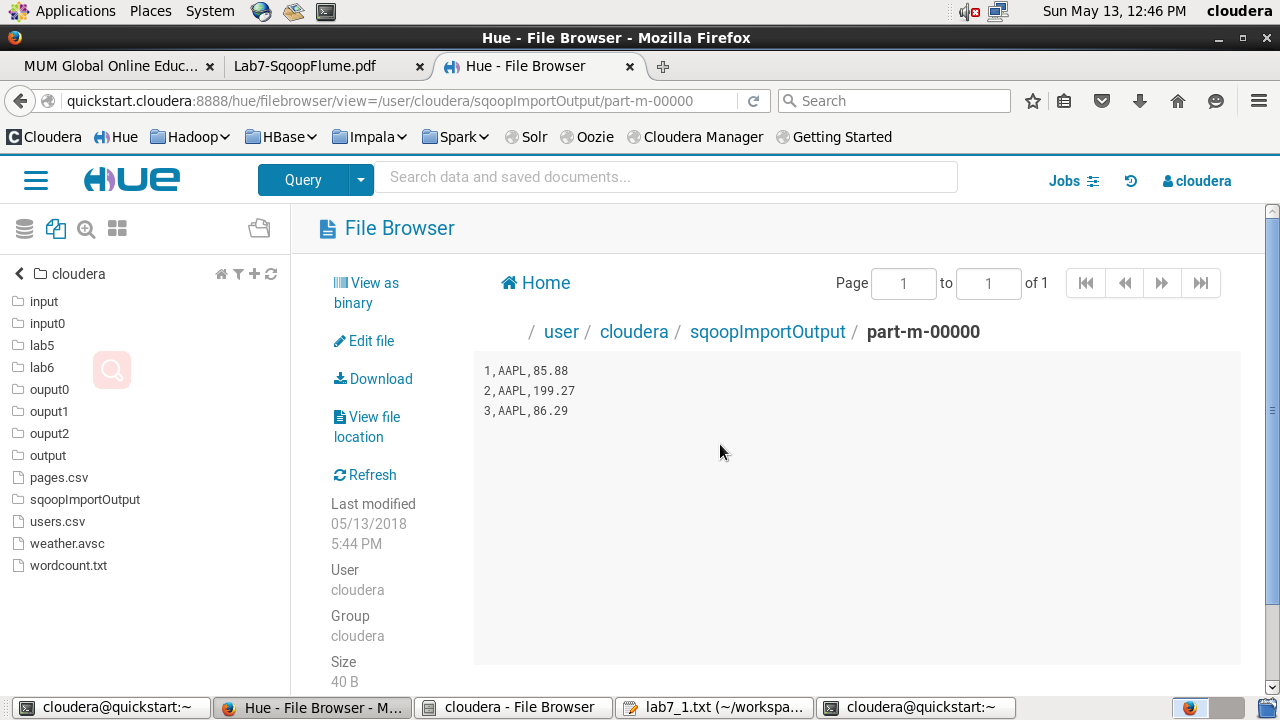


//result

1,AAPL,85.88

2,AAPL,199.27

3,AAPL,86.29



Part 2 Flume

//Configuration file

#Components

agent.sources = r1

agent.sinks = k1 k2

agent.channels = c1 c2

#Source

agent.sources.r1.type = spooldir

agent.sources.r1.channels = c1 c2

agent.sources.r1.spoolDir = /tmp/spooldir

#Channel

agent.channels.c1.type = file

agent.channels.c1.capacity = 100000

agent.channels.c1.transactionCapacity = 1000

agent.channels.c2.type = memory

agent.channels.c2.capacity = 100000

agent.channels.c2.transactionCapacity = 1000

#Sink

agent.sinks.k1.type = hdfs

agent.sinks.k1.hdfs.path = /tmp/flume

agent.sinks.k1.hdfs.fileType = DataStream

agent.sinks.k1.channel = c1

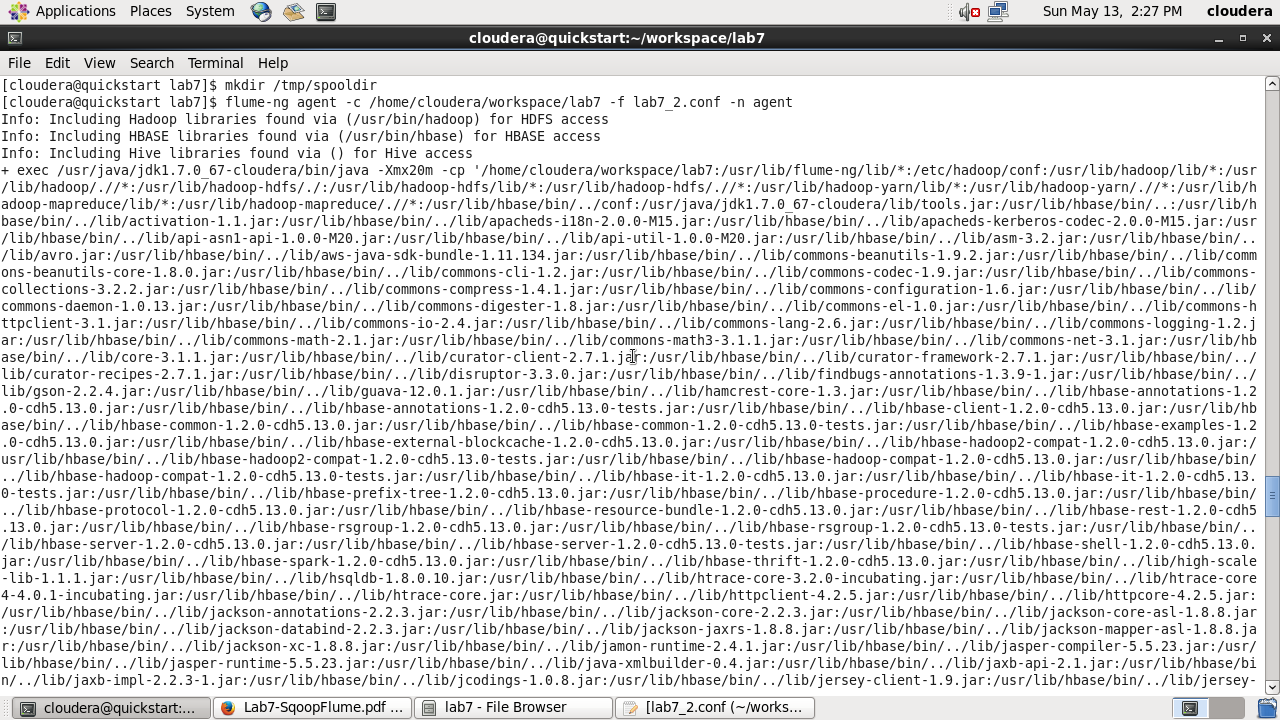
agent.sinks.k2.type = logger

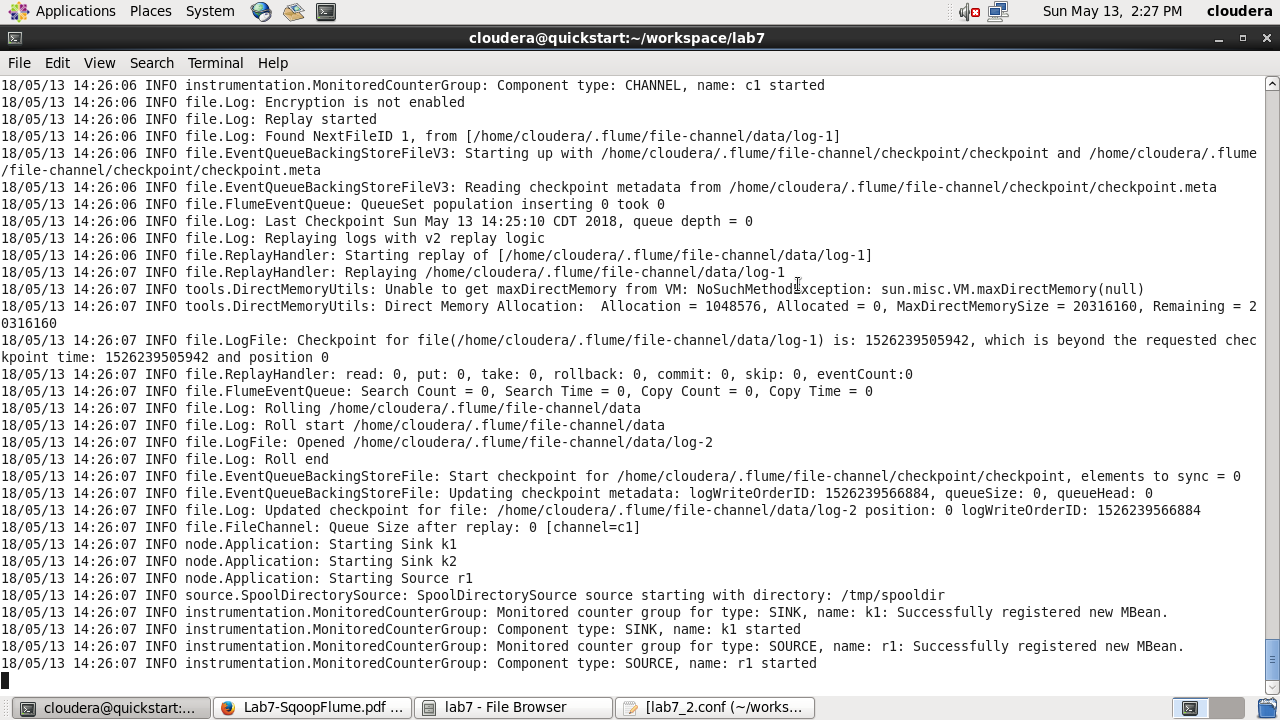
agent.sinks.k2.channel = c2

#Command

mkdir /tmp/spooldir

flume-ng agent -c /home/cloudera/workspace/lab7 -f lab7\_2.conf -n agent





//create new file

