Lab 7 – Hoang Duy Vu – 986170

**SQOOP homework**

--connect to database and create database

mysql –u root –p

create database cs523;

use cs523;

--create table stocks

create table stocks (id int not null primary key auto\_increment, symbol varchar(20), quote\_date varchar(20), open\_price double, high\_price double, low\_price double);

--insert data to stocks

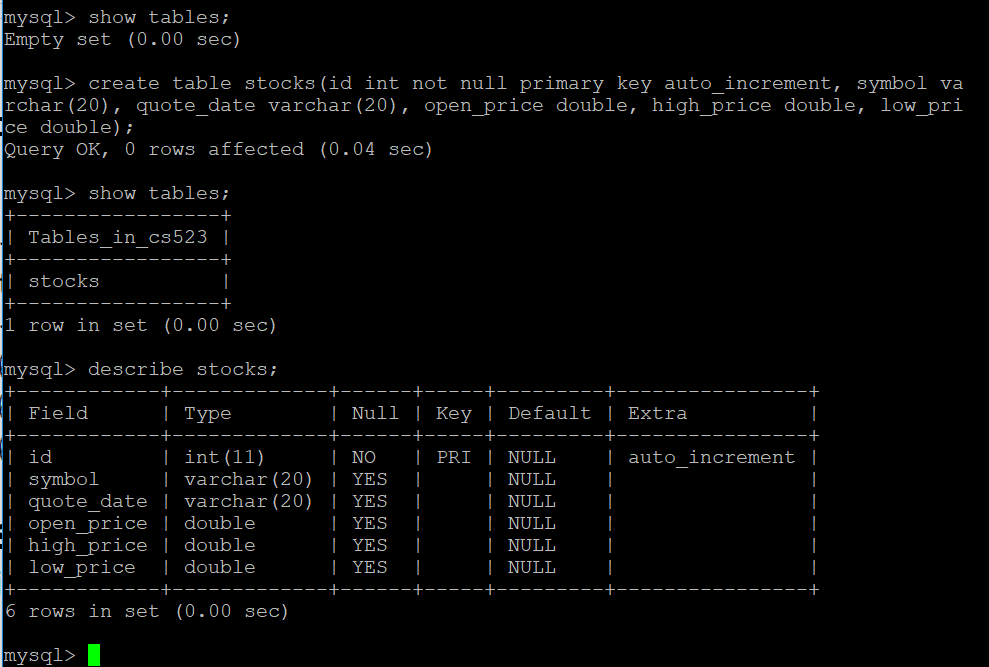
insert into stocks values (1, "AAPL", "2009-01-02",85.88,91.04,85.16), (2, "AAPL", "2008-01-02",199.27,200.26,192.55), (3, "AAPL", "2007-01-03",86.29,86.58,81.9);

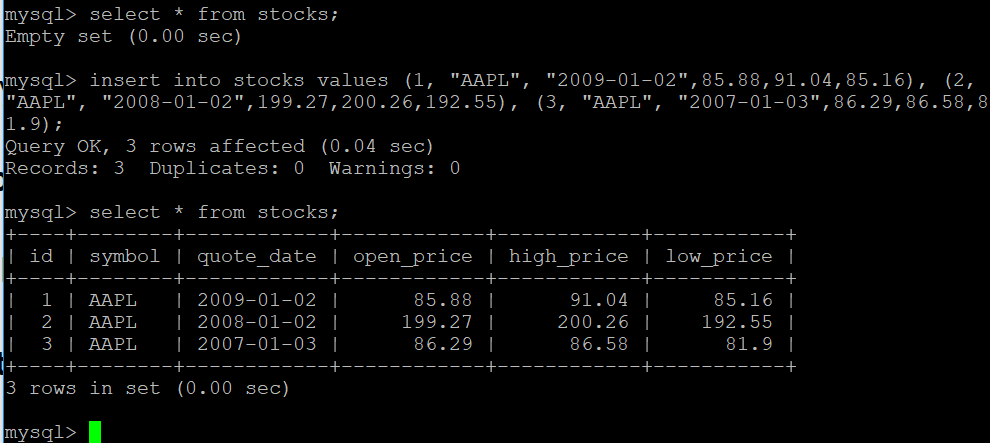
--import data to HDFS by Sqoop

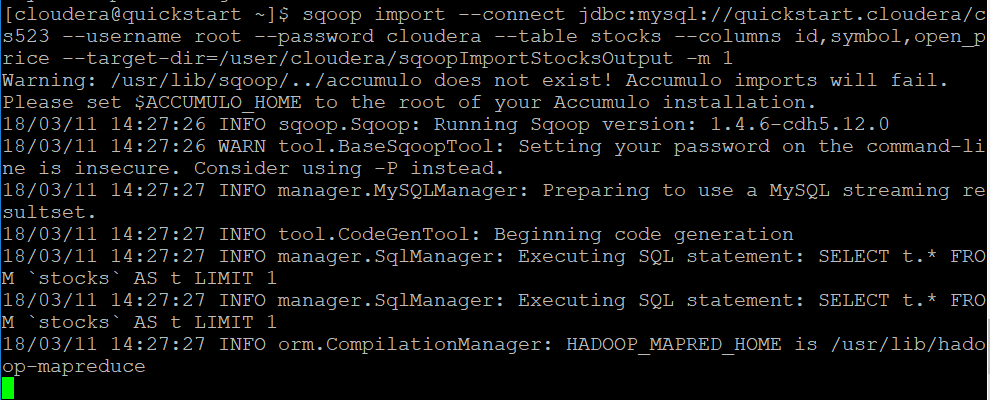
sqoop import --connect jdbc:mysql://quickstart.cloudera/cs523 --username root --password cloudera --table stocks --columns id,symbol,open\_price --target-dir=/user/cloudera/sqoopImportStocksOutput -m 1

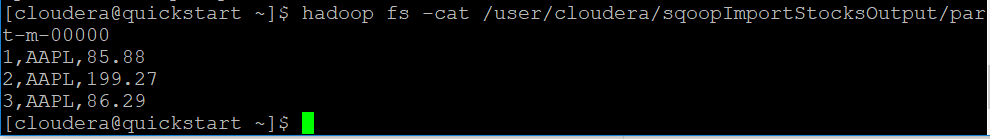
--cat output from sqoop

hadoop fs -cat /user/cloudera/sqoopImportStocksOutput/part-m-00000









**FLUME homework**

--flume configure-- HoangDuyVu\_flume.conf

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 = /tmp/spooldir

agent1.sinks.sink1a.type = hdfs

agent1.sinks.sink1a.hdfs.path = /tmp/flume

agent1.sinks.sink1a.hdfs.filePrefix = events

agent1.sinks.sink1a.hdfs.fileSuffix = .log

agent1.sinks.sink1a.hdfs.fileType = DataStream

agent1.sinks.sink1b.type = logger

agent1.channels.channel1a.type = file

agent1.channels.channel1b.type = memory

--start flume agent

flume-ng agent --conf-file /home/cloudera/lab7/HoangDuyVu\_flume.conf --name agent1 --conf $FLUME\_HOME/conf -Dflume.root.logger=INFO,console

