**Project in sqoop**

Write a Java Program to read nyse\_A.csv data line by line and store it in an ArrayList<Nyse>

If List size =50000,batch size = 10000 then you will have 5 batch.

In each batch insert 10,000 records to the database and then sleep the program for 2 minutes.

This will take total 10 minutes to insert the data into database.

 

Create a sh file in linux and copy the sqoop job command init

Sqoop job

sqoop job --create nysejob -- import --options-file /home/hduser/myfile.txt --table nyse\_daily\_prices -m 1 --columns "stock\_exchange,stock\_symbol,stock\_date,stock\_price\_open,stock\_price\_high,stock\_price\_low,stock\_price\_close,stock\_volume,stock\_price\_adj\_close" --target-dir /user/hduser/nyse\_daily\_prices --delete-target-dir

/home/hduser/nysejob.sh

sqoop job --exec nysejob

Use cron job schedular to run your sqoop job every 2 minutes to bring all the data that is already available in the database.Look into the diagram for more explaination.

create cron job using vim

vim job

* \* \* \* \* \* /home/hduser/nysejob.sh

Schedhuling cron job:

Crontab job

**Cron related Articles**

[**http://www.thegeekstuff.com/2009/06/15-practical-crontab-examples**](http://www.thegeekstuff.com/2009/06/15-practical-crontab-examples)

[**http://www.tecmint.com/11-cron-scheduling-task-examples-in-linux/**](http://www.tecmint.com/11-cron-scheduling-task-examples-in-linux/)

[**https://www.pantz.org/software/cron/croninfo.html**](https://www.pantz.org/software/cron/croninfo.html)

**https://www.cyberciti.biz/faq/how-do-i-add-jobs-to-cron-under-linux-or-unix-oses/**





