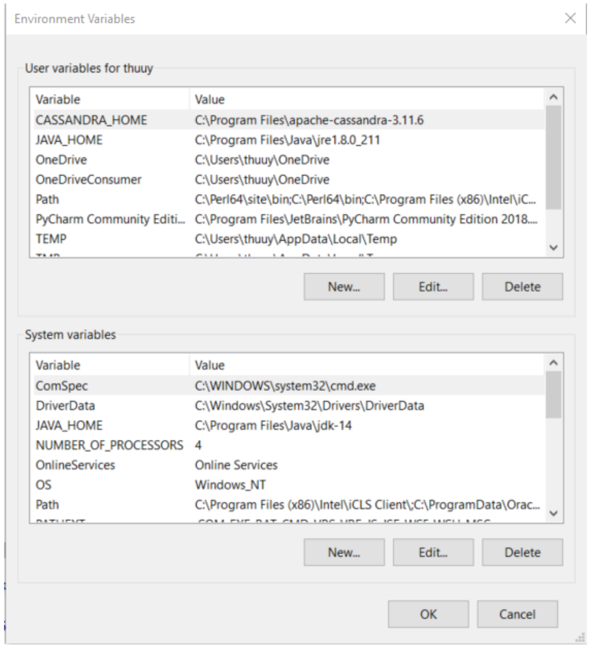
Uyen Dang

CS 5590 – Summer 2020

1. Download Python 2.7.X (https://www.python.org/downloads/release/python-2717/)
2. Install Python and during installation select Entire python paths
3. Download Cassandra Version 3.11.6 or latest (http://cassandra.apache.org/download/)
4. Extract it in you localsystem, i.e., C Drive
5. Add envrionemnt variable CASSNADRA_HOME with value "C:\apache-cassandra-3.11.6"

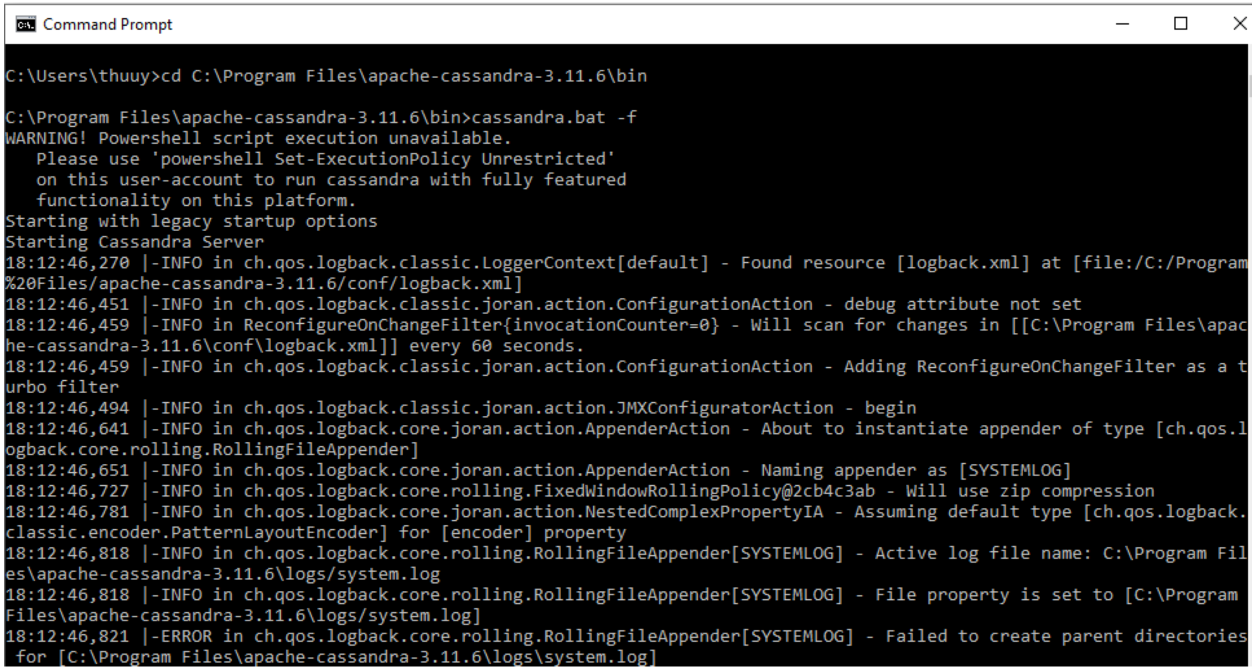


Variable	Value
CASSANDRA_HOME	C:\Program Files\apache-cassandra-3.11.6
JAVA_HOME	C:\Program Files\Java\jre1.8.0_211
OneDrive	C:\Users\thuuy\OneDrive
OneDriveConsumer	C:\Users\thuuy\OneDrive
Path	C:\Perl64\site\bin;C:\Perl64\bin;C:\Program Files (x86)\Intel\iC...
PyCharm Community Editi...	C:\Program Files\JetBrains\PyCharm Community Edition 2018....
TEMP	C:\Users\thuuy\AppData\Local\Temp

Variable	Value
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
JAVA_HOME	C:\Program Files\Java\jdk-14
NUMBER_OF_PROCESSORS	4
OnlineServices	Online Services
OS	Windows_NT
Path	C:\Program Files (x86)\Intel\iCLS Client\;C:\ProgramData\Orac...

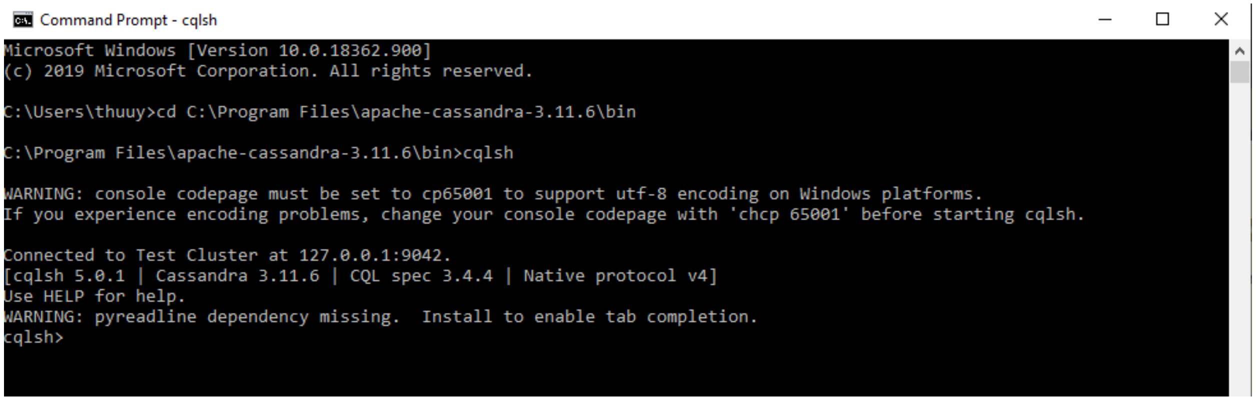
6. Navigate to "C:\apache-cassandra-3.11.6\bin" folder

7. Open command terminal window as administratior, navigate to "C:\apache-cassandra-3.11.6\bin" and Type "cassandra.bat -f" for starting cassandra server on localhost
8. Wait for server to initialize completely for atleast 3 - 5 minutes



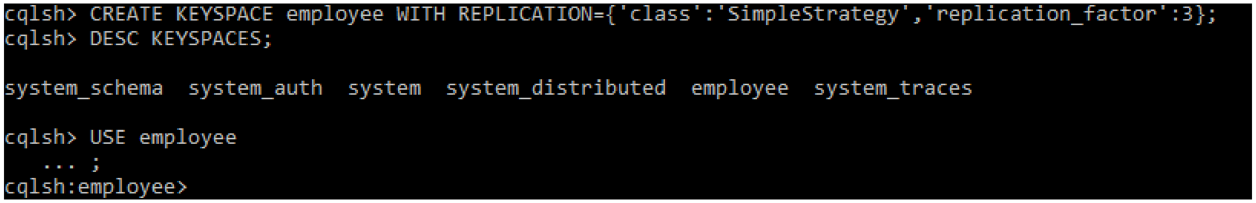
```
C:\Users\thuuy>cd C:\Program Files\apache-cassandra-3.11.6\bin

C:\Program Files\apache-cassandra-3.11.6\bin>cassandra.bat -f
WARNING! Powershell script execution unavailable.
   Please use 'powershell Set-ExecutionPolicy Unrestricted'
   on this user-account to run cassandra with fully featured
   functionality on this platform.
Starting with legacy startup options
Starting Cassandra Server
18:12:46,270 |-INFO in ch.qos.logback.classic.LoggerContext[default] - Found resource [logback.xml] at [file:/C:/Program
%20Files/apache-cassandra-3.11.6/conf/logback.xml]
18:12:46,451 |-INFO in ch.qos.logback.classic.joran.action.ConfigurationAction - debug attribute not set
18:12:46,459 |-INFO in ReconfigureOnChangeFilter{invocationCounter=0} - Will scan for changes in [[C:\Program Files\apac
he-cassandra-3.11.6\conf\logback.xml]] every 60 seconds.
18:12:46,459 |-INFO in ch.qos.logback.classic.joran.action.ConfigurationAction - Adding ReconfigureOnChangeFilter as a t
urbo filter
18:12:46,494 |-INFO in ch.qos.logback.classic.joran.action.JMXConfiguratorAction - begin
18:12:46,641 |-INFO in ch.qos.logback.core.joran.action.AppenderAction - About to instantiate appender of type [ch.qos.l
ogback.core.rolling.RollingFileAppender]
18:12:46,651 |-INFO in ch.qos.logback.core.joran.action.AppenderAction - Naming appender as [SYSTEMLOG]
18:12:46,727 |-INFO in ch.qos.logback.core.rolling.FixedWindowRollingPolicy@2cb4c3ab - Will use zip compression
18:12:46,781 |-INFO in ch.qos.logback.core.joran.action.NestedComplexPropertyIA - Assuming default type [ch.qos.logback.
classic.encoder.PatternLayoutEncoder] for [encoder] property
18:12:46,818 |-INFO in ch.qos.logback.core.rolling.RollingFileAppender[SYSTEMLOG] - Active log file name: C:\Program Fil
es\apache-cassandra-3.11.6\logs/system.log
18:12:46,818 |-INFO in ch.qos.logback.core.rolling.RollingFileAppender[SYSTEMLOG] - File property is set to [C:\Program
Files\apache-cassandra-3.11.6\logs/system.log]
18:12:46,821 |-ERROR in ch.qos.logback.core.rolling.RollingFileAppender[SYSTEMLOG] - Failed to create parent directories
 for [C:\Program Files\apache-cassandra-3.11.6\logs\system.log]
```

9. open new terminal windows navigate to "C:\apache-cassandra-3.11.6\bin" and type "cqlsh"



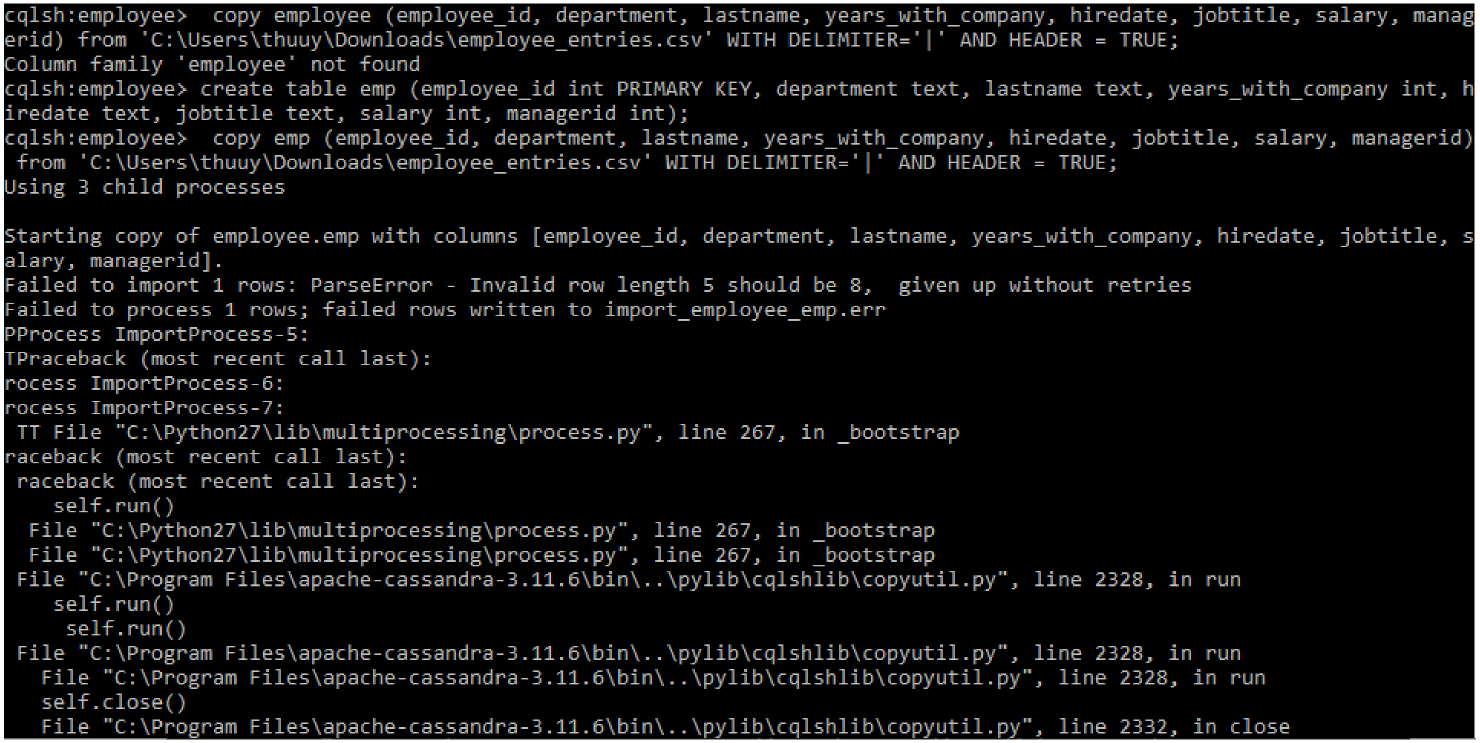
```
Microsoft Windows [Version 10.0.18362.900]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\thuuy>cd C:\Program Files\apache-cassandra-3.11.6\bin

C:\Program Files\apache-cassandra-3.11.6\bin>cqlsh

WARNING: console codepage must be set to cp65001 to support utf-8 encoding on Windows platforms.
If you experience encoding problems, change your console codepage with 'chcp 65001' before starting cqlsh.

Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.6 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
WARNING: pyreadline dependency missing.  Install to enable tab completion.
cqlsh>
```

10. Create a new name space and use that namespace

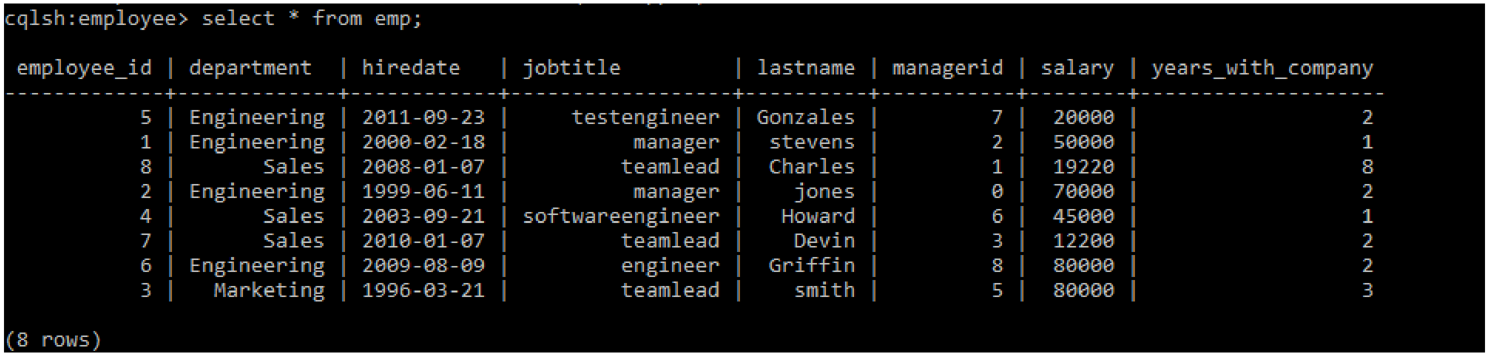


```
cqlsh> CREATE KEYSPACE employee WITH REPLICATION={'class':'SimpleStrategy','replication_factor':3};
cqlsh> DESC KEYSPACES;

system_schema  system_auth  system  system_distributed  employee  system_traces

cqlsh> USE employee
   ... ;
cqlsh:employee>
```

11. Create table emp and load data.



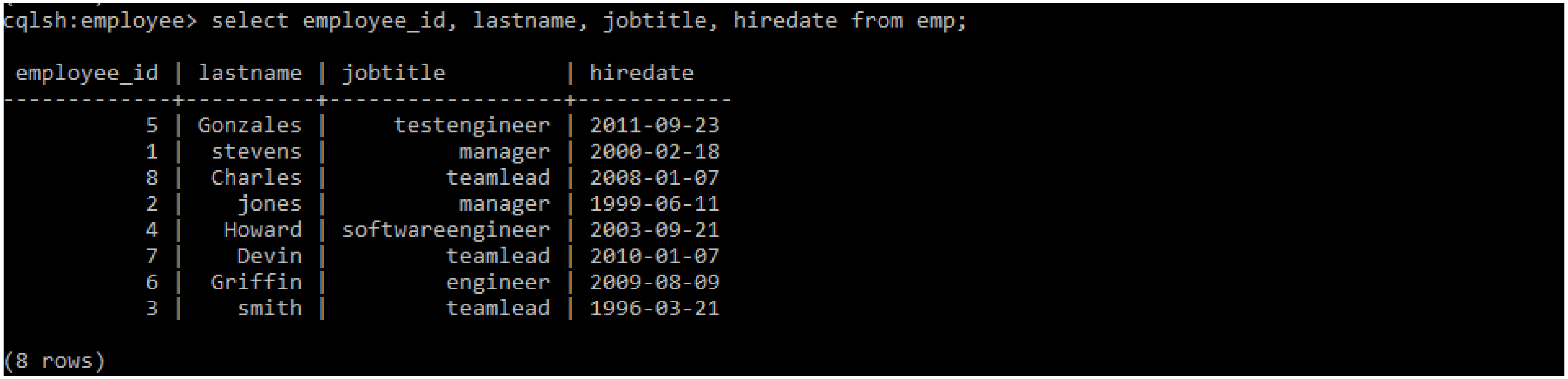
```
cqlsh:employee>  copy employee (employee_id, department, lastname, years_with_company, hiredate, jobtitle, salary, manag
erid) from 'C:\Users\thuuy\Downloads\employee_entries.csv' WITH DELIMITER='|' AND HEADER = TRUE;
Column family 'employee' not found
cqlsh:employee> create table emp (employee_id int PRIMARY KEY, department text, lastname text, years_with_company int, h
iredate text, jobtitle text, salary int, managerid int);
cqlsh:employee>  copy emp (employee_id, department, lastname, years_with_company, hiredate, jobtitle, salary, managerid)
 from 'C:\Users\thuuy\Downloads\employee_entries.csv' WITH DELIMITER='|' AND HEADER = TRUE;
Using 3 child processes

Starting copy of employee.emp with columns [employee_id, department, lastname, years_with_company, hiredate, jobtitle, s
alary, managerid].
Failed to import 1 rows: ParseError - Invalid row length 5 should be 8,  given up without retries
Failed to process 1 rows; failed rows written to import_employee_emp.err
PProcess ImportProcess-5:
TPraceback (most recent call last):
rocess ImportProcess-6:
rocess ImportProcess-7:
 TT File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
raceback (most recent call last):
 raceback (most recent call last):
    self.run()
  File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
  File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
 File "C:\Program Files\apache-cassandra-3.11.6\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
    self.run()
     self.run()
 File "C:\Program Files\apache-cassandra-3.11.6\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
   File "C:\Program Files\apache-cassandra-3.11.6\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
   self.close()
   File "C:\Program Files\apache-cassandra-3.11.6\bin\..\pylib\cqlshlib\copyutil.py", line 2332, in close
```



```
cqlsh:employee> select * from emp;

 employee_id | department  | hiredate   | jobtitle         | lastname | managerid | salary | years_with_company
-------------+-------------+------------+------------------+----------+-----------+--------+--------------------
           5 | Engineering | 2011-09-23 |     testengineer | Gonzales |         7 |  20000 |                  2
           1 | Engineering | 2000-02-18 |          manager |  stevens |         2 |  50000 |                  1
           8 |       Sales | 2008-01-07 |         teamlead |  Charles |         1 |  19220 |                  8
           2 | Engineering | 1999-06-11 |          manager |    jones |         0 |  70000 |                  2
           4 |       Sales | 2003-09-21 | softwareengineer |   Howard |         6 |  45000 |                  1
           7 |       Sales | 2010-01-07 |         teamlead |    Devin |         3 |  12200 |                  2
           6 | Engineering | 2009-08-09 |         engineer |  Griffin |         8 |  80000 |                  2
           3 |   Marketing | 1996-03-21 |         teamlead |    smith |         5 |  80000 |                  3

(8 rows)
```

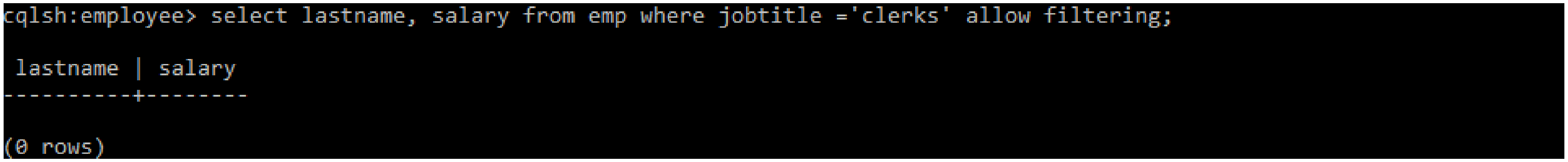
QUERIES

1. List the empID, ename, jobtitle and hiredate of employee from the employee table



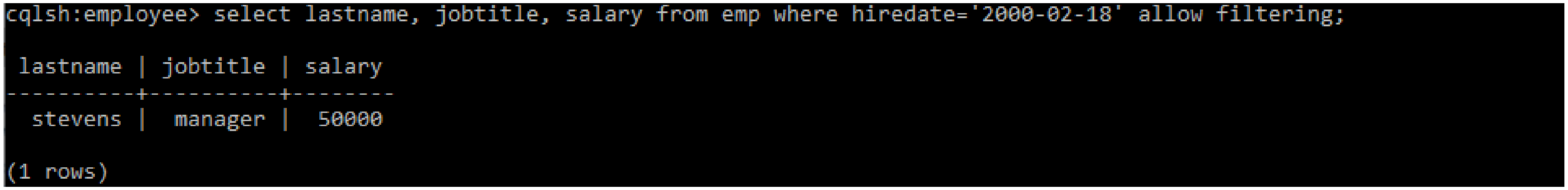
```
cqlsh:employee> select employee_id, lastname, jobtitle, hiredate from emp;

 employee_id | lastname | jobtitle         | hiredate
-------------+----------+------------------+------------
           5 | Gonzales |     testengineer | 2011-09-23
           1 |  stevens |          manager | 2000-02-18
           8 |  Charles |         teamlead | 2008-01-07
           2 |    jones |          manager | 1999-06-11
           4 |   Howard | softwareengineer | 2003-09-21
           7 |    Devin |         teamlead | 2010-01-07
           6 |  Griffin |         engineer | 2009-08-09
           3 |    smith |         teamlead | 1996-03-21

(8 rows)
```

2. List the name, salary of the employees who are clerks.



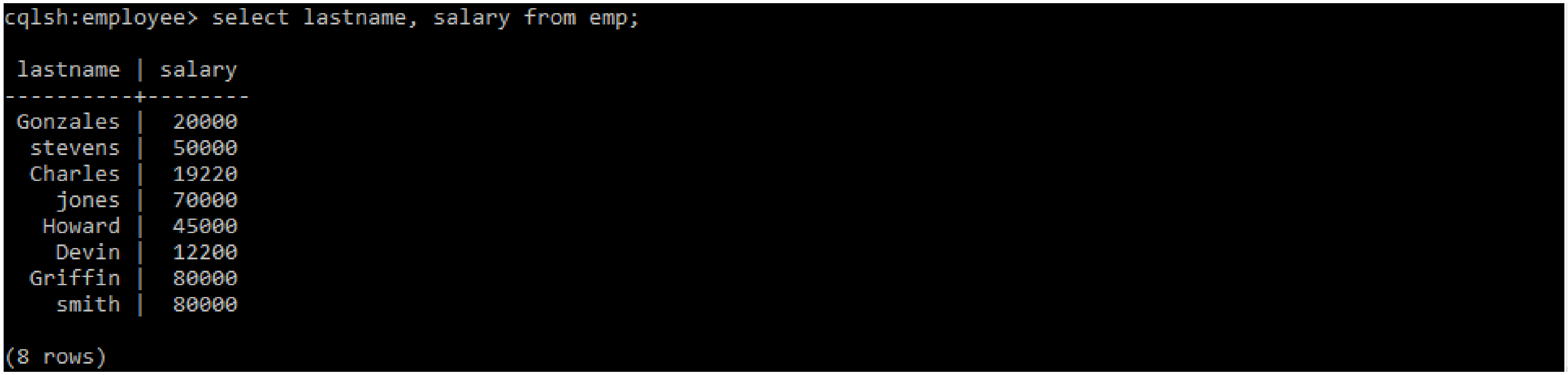
```
cqlsh:employee> select lastname, salary from emp where jobtitle ='clerks' allow filtering;

 lastname | salary
----------+--------

(0 rows)
```

3. List the name, job, salary of every employee joined on ‘february18,2000’,



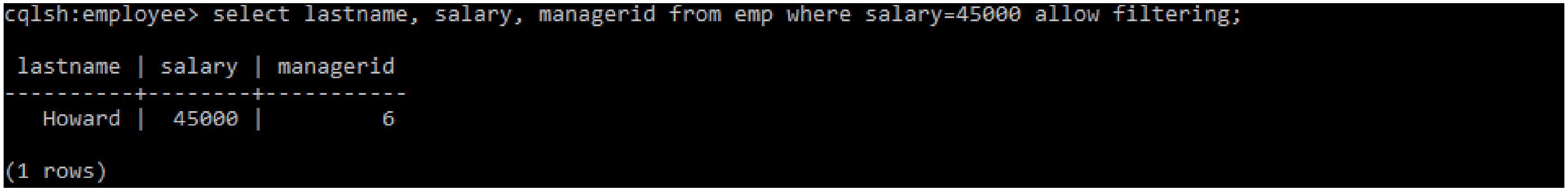
```
cqlsh:employee> select lastname, jobtitle, salary from emp where hiredate='2000-02-18' allow filtering;

 lastname | jobtitle | salary
----------+----------+--------
  stevens |  manager |  50000

(1 rows)
```

4. List name and annual salary of all the employees.



```
cqlsh:employee> select lastname, salary from emp;

 lastname | salary
----------+--------
 Gonzales |  20000
  stevens |  50000
  Charles |  19220
    jones |  70000
   Howard |  45000
    Devin |  12200
  Griffin |  80000
    smith |  80000

(8 rows)
```

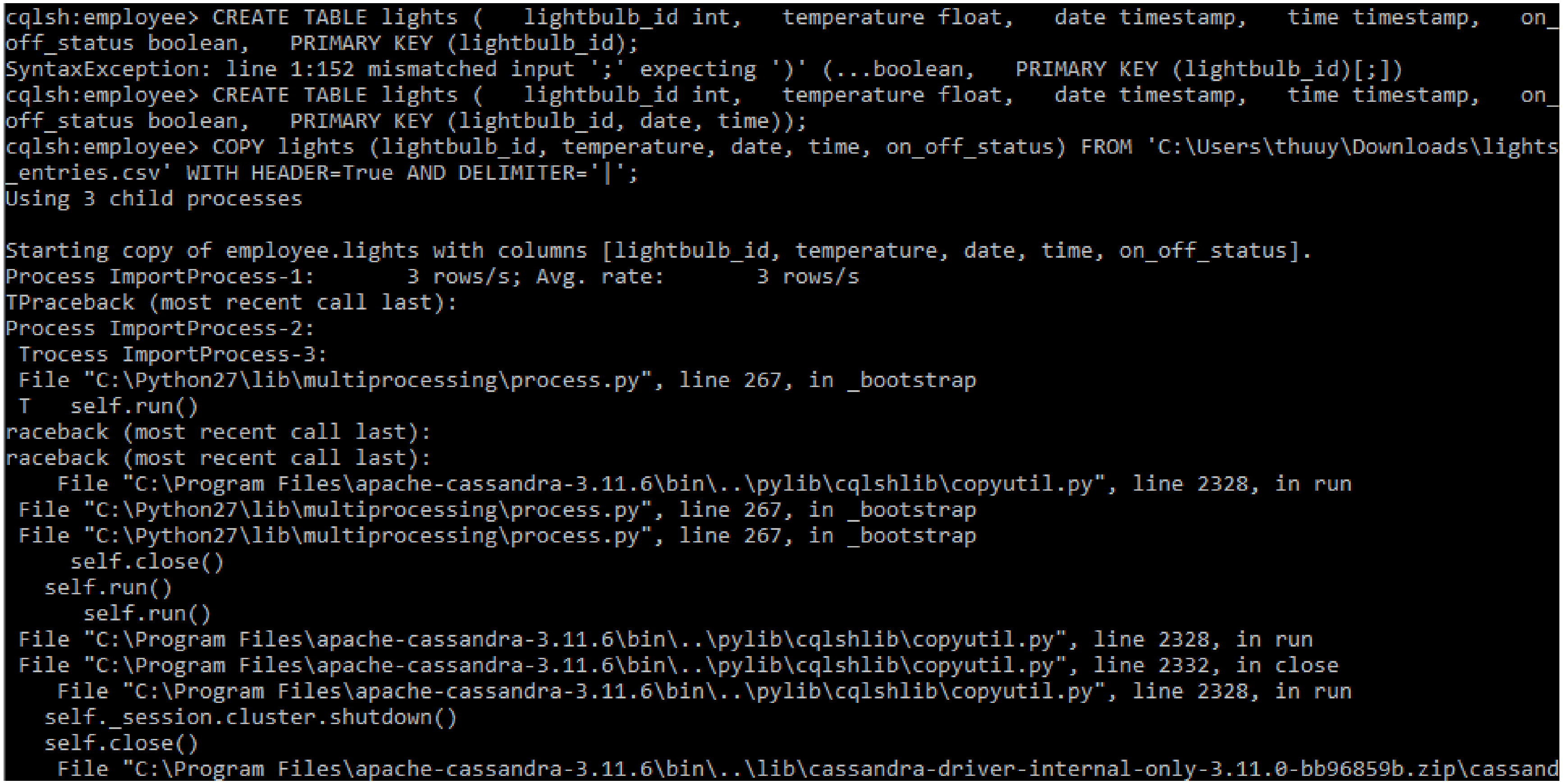
5. Display employees’ names, salary and manager values of those employees whose salary is 45000 from EMP table using SELECT statement.



```
cqlsh:employee> select lastname, salary, managerid from emp where salary=45000 allow filtering;

 lastname | salary | managerid
----------+--------+-----------
   Howard |  45000 |         6

(1 rows)
```

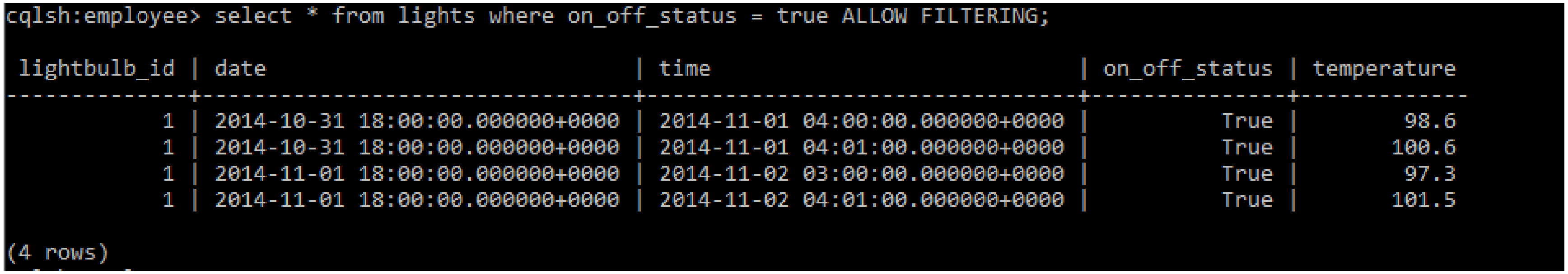
Bonus: (2) Import any data from the given data set and apply any commands like above.

Create table lights and import data



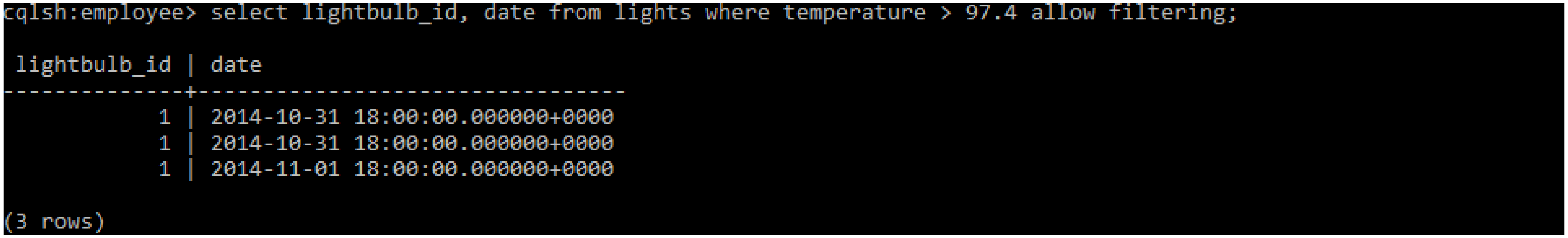
```
cqlsh:employee> CREATE TABLE lights (   lightbulb_id int,   temperature float,   date timestamp,   time timestamp,   on_
off_status boolean,   PRIMARY KEY (lightbulb_id);
SyntaxException: line 1:152 mismatched input ';' expecting ')' (...boolean,   PRIMARY KEY (lightbulb_id)[;])
cqlsh:employee> CREATE TABLE lights (   lightbulb_id int,   temperature float,   date timestamp,   time timestamp,   on_
off_status boolean,   PRIMARY KEY (lightbulb_id, date, time));
cqlsh:employee> COPY lights (lightbulb_id, temperature, date, time, on_off_status) FROM 'C:\Users\thuuy\Downloads\lights
_entries.csv' WITH HEADER=True AND DELIMITER='|';
Using 3 child processes

Starting copy of employee.lights with columns [lightbulb_id, temperature, date, time, on_off_status].
Process ImportProcess-1:       3 rows/s; Avg. rate:       3 rows/s
TPraceback (most recent call last):
Process ImportProcess-2:
 Trocess ImportProcess-3:
 File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
 T   self.run()
raceback (most recent call last):
raceback (most recent call last):
    File "C:\Program Files\apache-cassandra-3.11.6\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
 File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
 File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
     self.close()
   self.run()
      self.run()
 File "C:\Program Files\apache-cassandra-3.11.6\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
 File "C:\Program Files\apache-cassandra-3.11.6\bin\..\pylib\cqlshlib\copyutil.py", line 2332, in close
    File "C:\Program Files\apache-cassandra-3.11.6\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
   self._session.cluster.shutdown()
   self.close()
    File "C:\Program Files\apache-cassandra-3.11.6\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassand
```

Queries 1: Select the lightbulb_id, temperature, date, time where on_off_status = true



```
cqlsh:employee> select * from lights where on_off_status = true ALLOW FILTERING;

 lightbulb_id | date                            | time                            | on_off_status | temperature
--------------+---------------------------------+---------------------------------+---------------+-------------
            1 | 2014-10-31 18:00:00.000000+0000 | 2014-11-01 04:00:00.000000+0000 |          True |        98.6
            1 | 2014-10-31 18:00:00.000000+0000 | 2014-11-01 04:01:00.000000+0000 |          True |       100.6
            1 | 2014-11-01 18:00:00.000000+0000 | 2014-11-02 03:00:00.000000+0000 |          True |        97.3
            1 | 2014-11-01 18:00:00.000000+0000 | 2014-11-02 04:01:00.000000+0000 |          True |       101.5

(4 rows)
```

Queries 2: Find the lightbulb and date that have temperature > 97.4



```
cqlsh:employee> select lightbulb_id, date from lights where temperature > 97.4 allow filtering;

 lightbulb_id | date
--------------+---------------------------------
            1 | 2014-10-31 18:00:00.000000+0000
            1 | 2014-10-31 18:00:00.000000+0000
            1 | 2014-11-01 18:00:00.000000+0000

(3 rows)
```