Lab 3: From MySQL to HDFS

Installation

- MySQL
- Sqoop

Installation (MySQL)

```
sudo apt-get install mysql-server
```
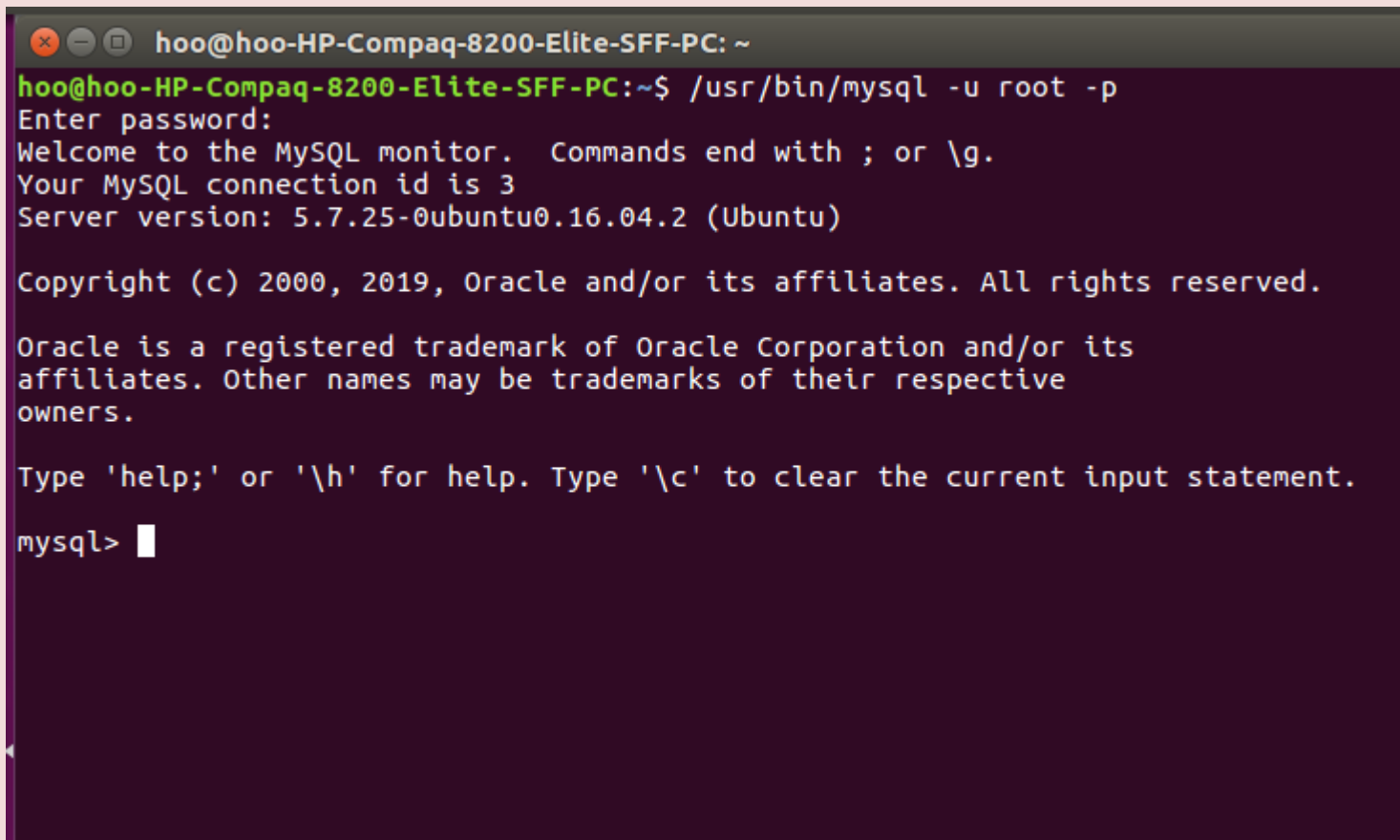
- When installing, you will be prompted to key in root password

```
systemctl start mysql
```

Installation (MySQL)

```
/usr/bin/mysql -u root -p
```

– When you're prompted for a password, enter the one that you set at installation time

MySQL: importing csv to MySQL

- In MySQL:
  - CREATE DATABASE WQD7007;
  - Use WQD7007;
  - create table churn (customerID varchar(20), PaperlessBilling varchar(3),
  - PaymentMethod varchar(30),
  - MonthlyCharges numeric(8,2),
  - Churn varchar(3));
  - exit

- Download 'churn_reduced.csv' from Spectrum to Downloads folder, then

```
mysql -uroot -proot --local_infile=1 mysql -e
"LOAD DATA LOCAL INFILE
'~/Downloads/churn_reduced.csv' INTO TABLE churn
FIELDS TERMINATED BY ','"
```

MySQL: importing csv to MySQL

- Log in again, then

  Use WQD7007;

  Select * from churn;

```
hoo@hoo-HP-Compaq-8200-Elite-SFF-PC:~$ mysql -uroot -proot --local_infile=1 WQD7007 -e "LOAD DATA LOCAL INFILE '~/Downloads/churn_red
uced.csv' INTO TABLE churn FIELDS TERMINATED BY ','"
mysql: [Warning] Using a password on the command line interface can be insecure.
hoo@hoo-HP-Compaq-8200-Elite-SFF-PC:~$ /usr/bin/mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 18
Server version: 5.7.25-0ubuntu0.16.04.2 (Ubuntu)

Copyright (c) 2000, 2019, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use WQD7007;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> select * from churn;
+------------+-----------------+---------------------------+----------------+-------+
| customerID | PaperlessBilling | PaymentMethod            | MonthlyCharges | Churn |
+------------+-----------------+---------------------------+----------------+-------+
| customerID | Pap             | PaymentMethod            |           0.00 | Chu   |
| 7590-VHVEG | Yes             | Electronic check         |          29.85 | No    |
| 5575-GNVDE | No              | Mailed check             |          56.95 | No    |
| 3668-QPYBK | Yes             | Mailed check             |          53.85 | Yes   |
| 7795-CFOCW | No              | Bank transfer (automatic) |          42.30 | No    |
| 9237-HQITU | Yes             | Electronic check         |          70.70 | Yes   |
| 9305-CDSKC | Yes             | Electronic check         |          99.65 | Yes   |
| 1452-KIOVK | Yes             | Credit card (automatic)  |          89.10 | No    |
| 6713-OKOMC | No              | Mailed check             |          29.75 | No    |
| 7892-POOKP | Yes             | Electronic check         |         104.80 | Yes   |
| 6388-TABGU | No              | Bank transfer (automatic) |          56.15 | No    |
| 9763-GRSKD | Yes             | Mailed check             |          49.95 | No    |
| 7469-LKBCI | No              | Credit card (automatic)  |          18.95 | No    |
| 8091-TTVAX | No              | Credit card (automatic)  |         100.35 | No    |
| 0280-XJGEX | Yes             | Bank transfer (automatic) |         103.70 | Yes   |
| 5129-JLPIS | Yes             | Electronic check         |         105.50 | No    |
| 3655-SNQYZ | No              | Credit card (automatic)  |         113.25 | No    |
| 8191-XWSZG | No              | Mailed check             |          20.65 | No    |
| 9959-WOFKT | No              | Bank transfer (automatic) |         106.70 | No    |
| 4190-MFLUW | No              | Credit card (automatic)  |          55.20 | Yes   |
| 4183-MYFRB | Yes             | Electronic check         |          90.05 | No    |
| 8779-QRDMV | Yes             | Electronic check         |          39.65 | Yes   |
| 1680-VDCWW | No              | Bank transfer (automatic) |          19.80 | No    |
| 1066-JKSGK | No              | Mailed check             |          20.15 | Yes   |
| 3638-WEABW | Yes             | Credit card (automatic)  |          59.90 | No    |
| 6322-HRPFA | No              | Credit card (automatic)  |          59.60 | No    |
| 6865-JZNKO | Yes             | Bank transfer (automatic) |          55.30 | No    |
| 6467-CHFZW | Yes             | Electronic check         |          99.35 | Yes   |
| 8665-UTDHZ | No              | Electronic check         |          30.20 | Yes   |
| 5248-YGIJN | Yes             | Credit card (automatic)  |          90.25 | No    |
+------------+-----------------+---------------------------+----------------+-------+
30 rows in set (0.00 sec)
```

Installation (Sqoop)

- Download sqoop:

```
wget https://www-
eu.apache.org/dist/sqoop/1.4.7/sqoop-
1.4.7.bin__hadoop-2.6.0.tar.gz
```

- Extract:

```
tar -xzf sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz
```

- Move sqoop to new directory:

```
mv sqoop-1.4.7.bin__hadoop-2.6.0
/home/student/sqoop/
```

- At the end of .bashrc, add:

```
export PATH=$PATH:/home/student/sqoop/bin
```

- Rerun .bashrc

```
source .bashrc
```

Installation (Sqoop)

- Change template file to main file

  ```
  mv /home/student/sqoop/conf/sqoop-env-template.sh
  /home/student/sqoop/conf/sqoop-env.sh
  ```

- In sqoop-env.sh:

  ```
  export HADOOP_COMMON_HOME=/home/student/hadoop
  export HADOOP_MAPRED_HOME=/home/student/hadoop
  ```

- Download MySQL connector from Spectrum and Put this under `/home/student/sqoop/lib`

- Execute sqoop command to import from MySQL to hdfs:

  ```
  sqoop import -connect
  jdbc:mysql://localhost/WQD7007 -username
  root -password root -table churn -m 1
  ```

Outcome

```
hdfs dfs -cat /user/{your logged in user account}/churn/*
```

```
5575-GNVDE,No,Mailed check,56.95,No
3668-QPYBK,Yes,Mailed check,53.85,Yes
7795-CFOCW,No,Bank transfer (automatic),42.30,No
9237-HQITU,Yes,Electronic check,70.70,Yes
9305-CDSKC,Yes,Electronic check,99.65,Yes
1452-KIOVK,Yes,Credit card (automatic),89.10,No
6713-OKOMC,No,Mailed check,29.75,No
7892-POOKP,Yes,Electronic check,104.80,Yes
6388-TABGU,No,Bank transfer (automatic),56.15,No
9763-GRSKD,Yes,Mailed check,49.95,No
7469-LKBCI,No,Credit card (automatic),18.95,No
8091-TTVAX,No,Credit card (automatic),100.35,No
0280-XJGEX,Yes,Bank transfer (automatic),103.70,Yes
5129-JLPIS,Yes,Electronic check,105.50,No
3655-SNQYZ,No,Credit card (automatic),113.25,No
8191-XWSZG,No,Mailed check,20.65,No
9959-WOFKT,No,Bank transfer (automatic),106.70,No
4190-MFLUW,No,Credit card (automatic),55.20,Yes
4183-MYFRB,Yes,Electronic check,90.05,No
8779-QRDMV,Yes,Electronic check,39.65,Yes
1680-VDCWW,No,Bank transfer (automatic),19.80,No
1066-JKSGK,No,Mailed check,20.15,Yes
3638-WEABW,Yes,Credit card (automatic),59.90,No
6322-HRPFA,No,Credit card (automatic),59.60,No
6865-JZNKO,Yes,Bank transfer (automatic),55.30,No
6467-CHFZW,Yes,Electronic check,99.35,Yes
8665-UTDHZ,No,Electronic check,30.20,Yes
5248-YGIJN,Yes,Credit card (automatic),90.25,No
hoo@hoo-HP-Compaq-8200-Elite-SFF-PC:~$
```