HomeWork 4

Steps:

1. **Transferring the hw4.sql file from local machine to the VM**
   ~ scp -P 2222 /Users/<localpath>/hw4.sql maria_dev@127.0.0.1:/home/maria_dev

2. **~** mysql -u root -p

3. ***mysql >*** source hw4.sql

4. ***mysql >*** set names utf8;

5. ***mysql >*** set character set utf8;

6. ***mysql >*** GRANT ALL PRIVILEGES ON homework4.* to ''@'localhost';

7. ***hive >*** create database homework4;

8. ***hive >*** use homework4;

9. **Creating External Table for accounts and importing the database: (Method 1)**
   hive > create EXTERNAL TABLE accounts (
   > id INT,
   > name VARCHAR(50),
   > phone CHAR(12))
   > row format delimited
   > fields terminated by '\t'
   > lines terminated by '\n'
   > location '/user/maria_dev/hive_import_homework4/accounts';

10. ***hive >*** LOAD DATA INPATH '/user/maria_dev/sqoop_import_homework4/accounts/part-m-00000' OVERWRITE INTO TABLE accounts;

11. **~** sqoop import --connect jdbc:mysql://localhost/homework4 --driver com.mysql.jdbc.Driver --table accounts --target-dir '/user/maria_dev/sqoop_import_homework4/accounts' --fields-terminated-by '\t' -m 1

12. **Using Sqoop, directly importing the database into hive (Method 2)**
~ sqoop import --connect jdbc:mysql://localhost/homework4 --driver com.mysql.jdbc.Driver --table contacts --hive-import --hive-table homework4.contacts -m 1

13. **Creating UDF Function for phone:**

```
package homework4;

import org.apache.hadoop.hive.ql.exec.UDF;
import org.apache.hadoop.io.Text;

public class hw4_phone extends UDF {

public Text evaluate(String value){
        String[] St=value.split("-");
        return new Text("XXX-XXX-".concat(St[2]));
}
}
```

14. **Creating UDF Function for email:**

```
package homework4;

import org.apache.hadoop.hive.ql.exec.UDF;
import org.apache.hadoop.io.Text;

public class hw4_email extends UDF {

public Text evaluate(String value){
        String[] St=value.split("@");
        return new Text("XXX@".concat(St[1]));
}
}
```

Adding Jar files and creating temporary functions:

15. ***hive >*** add jar hw4_phone.jar;

16. ***hive >*** create temporary function phone_replace as 'homework4.hw4_phone';

17. ***hive >*** add jar hw4_email.jar;

18. ***hive >*** create temporary function email_replace as 'homework4.hw4_email';

Creating new tables to export and masking it with UDF:

19. ***hive >*** create table accounts_new like accounts;

20. ***hive >*** INSERT OVERWRITE TABLE accounts_new SELECT id, name, phone_replace(phone) FROM accounts;

21. ***hive >*** create table contacts_new like contacts;

22. ***hive >*** INSERT OVERWRITE TABLE contacts_new SELECT id, account_id, first_name, last_name, phone_replace(phone), email FROM contacts;

23. ***hive >*** INSERT OVERWRITE TABLE contacts_new SELECT id, account_id, first_name, last_name, phone, email_replace(email) FROM contacts_new;

Exporting the new tables to the sql database:

24. sqoop export --connect jdbc:mysql://localhost/homework4 --driver com.mysql.jdbc.Driver --table accounts_new --export-dir /apps/hive/warehouse/homework4.db/accounts_new --input-fields-terminated-by '\t' -m 1

25. sqoop export --connect jdbc:mysql://localhost/homework4 --driver com.mysql.jdbc.Driver --table contacts_new --export-dir /apps/hive/warehouse/homework4.db/contacts_new --input-fields-terminated-by '\0001' -m 1

Output:

mysql > use homework 4;

mysql > show tables;

```
+--------------------+
| Tables_in_homework4 |
+--------------------+
| accounts           |
| accounts_new       |
| contacts           |
| contacts_new       |
+--------------------+
```

mysql > select * from accounts_new;

```
+----+----------------------+-------------+
| id | name                 | phone       |
+----+----------------------+-------------+
|  1 | Small Company LLC    | XXX-XXX-2222 |
|  2 | Medium Company LLC   | XXX-XXX-2222 |
|  3 | Large Company LLC    | XXX-XXX-2222 |
|  4 | Very Large Company LLC | XXX-XXX-2222 |
+----+----------------------+-------------+
```

mysql > *select * from contacts_new;*

```
+----+------------+------------+-----------+--------------+---------------+
| id | account_id | first_name | last_name | phone        | email         |
+----+------------+------------+-----------+--------------+---------------+
```

11	*1*	*John*	*Smith*	*XXX-XXX-2222*	*XXX@gmail.com*
12	*2*	*Bob*	*Smith*	*XXX-XXX-3334*	*XXX@gmail.com*
13	*3*	*Mark*	*Taylor*	*XXX-XXX-3335*	*XXX@gmail.com*
14	*4*	*Pat*	*Taylor*	*XXX-XXX-3336*	*XXX@gmail.com*
+----+------------+------------+-----------+--------------+---------------+