Step 1: Create database in MySQL

mysql -u root –p

mysql> source hw4.sql;

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

Step 2: Use Sqoop import database into Hive

hive> create database hw4;

hive> use hw4;

sqoop import-all-tables --connect jdbc:mysql://localhost/homework4 --driver com.mysql.jdbc.Driver --hive-import --hive-database hw4 -m 1

Step 3: Create UDFs to encrypt/mask Personal Identified Information (PII)

(1) UDF for Email:

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

public class MaskEmail extends UDF {
   public Text evaluate(Text input) {
      String[] email = input.toString().split("@");
      return new Text("XXX@" + email[1]);
   }
}
```

hive> ADD JAR MaskEmail.jar;
hive> create temporary function MaskEmail as 'MaskEmail';

(2) UDF for Phone Number:

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

public class MaskPhone extends UDF {

   public Text evaluate(Text input) {
      String[] phone = input.toString().split("-");
      return new Text("XXX-XXX-" + phone[2]);
   }
}
```

```
hive> add jar MaskPhone.jar;
create temporary function MaskPhone as 'MaskPhone';
```

Step 4: Create new tables in Hive to export and mask with UDFs

```
hive> create table accounts_masked like accounts;
hive> create table contacts_masked like contacts;

hive> insert overwrite table accounts_masked
   > select id, name, MaskPhone(phone)
   > from accounts;

hive> insert overwrite table contacts_masked
   > select id, account_id, first_name, last_name, MaskPhone(phone), MaskEmail(email)
   > from contacts;
```

Step 5: Create new tables in MySQL and use Sqoop to export data from Hive

```
mysql> create table accounts_masked (
   -> id INTEGER,
   -> name VARCHAR(50),
   -> phone CHAR(12));

mysql> create table contacts_masked (
   -> id INTEGER,
   -> account_id INTEGER,
   -> first_name VARCHAR(50),
   -> last_name VARCHAR(50),
   -> phone CHAR(12),
   -> email VARCHAR(50)
   -> );

sqoop export --connect jdbc:mysql://localhost/homework4 --driver com.mysql.jdbc.Driver --
table accounts_masked --export-dir /apps/hive/warehouse/hw4.db/accounts_masked --input-
fields-terminated-by '\0001' -m 1

sqoop export --connect jdbc:mysql://localhost/homework4 --driver com.mysql.jdbc.Driver --
table contacts_masked --export-dir /apps/hive/warehouse/hw4.db/contacts_masked --input-
fields-terminated-by '\0001' -m 1
```

Output:

```
mysql> select * from accounts_masked;
```

```
+------+-----------------------+--------------+
| 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_masked;
+------+------------+------------+-----------+--------------+---------------+
| 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 |
+------+------------+------------+-----------+--------------+---------------+
```