

=====

Aim: The aim of this lab is to get practice working on several of the Hadoop ecosystem components like Sqoop, Hive, Hive UDF, and Hue using the cloudera Hadoop Quickstart vm CDH 5.3.0.

=====

Project Requirement:Your company “Secure Data” has hired you as a big data consultant to implement a big data solution to mask PII (PersonallyIdentified Information) in the non-production instances with encrypted values, so that is not visible.

=====

#Load csv files to My SQL Db:

#Open the terminal and enter into mysql by using following command.

```
mysql -u root -p
```

```
create database original;
create database masked;
use original;
```

```
create table account(account_id int(10),account_name varchar(30), account_phone varchar(12));
```

```
create table contact(account_id int(10),contact_id int(10),First_name varchar(20),Last_name varchar(20),phone varchar(12),email varchar(30));
```

```
load data local infile '/home/cloudera/Desktop/Account.csv'into table account fields terminated by ',' lines terminated by '\n';
```

```
load data local infile '/home/cloudera/Desktop/Contact.csv' into table contact fields terminated by ',' lines terminated by '\n';
```

```
delete from account where account_id=0;
```

```
delete from contact where contact_id=0;
```

```
Select * from account;
```

```
Select * from contact;
```