Problem Statement:

Perform and explain the code flow and the associated result for the below tasks. Candidates should

create and use their own employee dataset for the same. Share the screenshot of the commands used

and its associated result.

● Transfer data between Mysql and HDFS (Import and Export) using Sqoop.

● Transfer data between Mysql and Hive (Import and Export only selected columns) using Sqoop.

1.1 Input COMMANDS FOR IMPORT FROM MYSQL TO HDFS:

```
[cloudera@quickstart ~]$ sudo service mysqld start
[cloudera@quickstart ~]$ mysql -u root -pcloudera
mysql> create database chhayadb1 ;
mysql> use chhayadb1 ;
mysql> CREATE TABLE emp
    ( id int,
     name varchar(20),
     designation varchar(20),
     salary int
     );
mysql> insert into emp values(1200,'Mark','Architect'50000);
mysql> insert into emp values(1201,'Nick','PM',40000);
mysql> insert into emp values(1202,'Arthur','Developer',30000);
mysql> select * from emp;
mqsql> grant all on *.* to 'root'@'localhost' with grant option;
mysql> flush privileges;
mysql> commit;
mysql> exit;
[cloudera@quickstart ~]$
sqoop import --connect jdbc:mysql://localhost/chhayadb1 \
--username 'root' -P --table 'emp' --target-dir '/sqoopout1' -m 1;
[cloudera@quickstart ~]$ hadoop fs -ls /sqoopout1;
```

Input MYSQL FILE

```
[cloudera@quickstart ~]$ sudo service mysqld start
Starting mysqld:                                           [  OK  ]
[cloudera@quickstart ~]$ mysql -u root -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 64
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, 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 chhayadb1;
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 emp;
+------+--------+-------------+--------+
| id   | name   | designation | salary |
+------+--------+-------------+--------+
| 1200 | Mark   | Architect   |  50000 |
| 1201 | Nick   | PM          |  40000 |
| 1202 | Arthur | Developer   |  30000 |
+------+--------+-------------+--------+
3 rows in set (0.00 sec)

mysql>
```

OUTPUT HDFS DIRECTORY:

```
[cloudera@quickstart ~]$ hadoop fs -ls /sqoopout1;
Found 2 items
-rw-r--r--   1 cloudera supergroup          0 2017-10-17 10:39 /sqoopout1/_SUCCESS
-rw-r--r--   1 cloudera supergroup         73 2017-10-17 10:39 /sqoopout1/part-m-00000
[cloudera@quickstart ~]$ hadoop fs -cat /sqoopout1/part-m-00000
1200,Mark,Architect,50000
1201,Nick,PM,40000
1202,Arthur,Developer,30000
[cloudera@quickstart ~]$
```

1.2 INPUT COMMANDS FOR EXPORT FROM HDFS TO MYSQL:

[cloudera@quickstart ~]$ sudo service mysqld start

[cloudera@quickstart ~]$ mysql -u root -pcloudera

mysql> use chhayadb1 ;

mysql> CREATE TABLE empexport

(id int not null primary key,

name varchar(20),

designation varchar(20),

salary int

);

mysql> SELECT * FROM empexport;

mysql> exit;

**[cloudera@quickstart ~]$ sqoop export **

**--connect jdbc:mysql://localhost/chhayadb1 **

**--username 'root' -P **

**--table 'empexport' **

--export-dir '/sqoopout1' ;

mysql> SELECT * FROM empexport;

INPUT HDFS DIRECTORY:

```
[cloudera@quickstart ~]$ hadoop fs -ls /sqoopout1;
Found 2 items
-rw-r--r--   1 cloudera supergroup          0 2017-10-17 10:39 /sqoopout1/_SUCCESS
-rw-r--r--   1 cloudera supergroup         73 2017-10-17 10:39 /sqoopout1/part-m-00000
[cloudera@quickstart ~]$ hadoop fs -cat /sqoopout1/part-m-00000
1200,Mark,Architect,50000
1201,Nick,PM,40000
1202,Arthur,Developer,30000
[cloudera@quickstart ~]$
```

Output EXPORT ENTIRE TABLE:

```
mysql> select * from empexport;
+------+--------+-------------+--------+
| id   | name   | designation | salary |
+------+--------+-------------+--------+
| 1200 | Mark   | Architect   |  50000 |
| 1201 | Nick   | PM          |  40000 |
| 1202 | Arthur | Developer   |  30000 |
+------+--------+-------------+--------+
3 rows in set (0.00 sec)

mysql>
```

2.1 INPUT COMMANDS TO IMPORT SELECTED COLUMNS FROM MYSQL INTO HIVE:

mysql> insert into emp values(1203,'Markus','Architect',55000);

mysql> insert into emp values(1204,'Nicholas','PM',45000);

mysql> insert into emp values(1205,'Arnica','Developer',36000);

mysql> insert into emp values(1206,'Monica','Architect',50100);

mysql> insert into emp values(1207,'Nancy','PM',90000);

mysql> insert into emp values(1208,'Angeline','Developer',60000);

mysql> select * from emp;

mysql> exit;

[cloudera@quickstart ~]$ sudo hive

hive> create database hivedb;

hive> exit;

**[cloudera@quickstart ~]$ sqoop import **

**--connect jdbc:mysql://localhost/chhayadb1 **

**--username 'root' -P **

**--table 'emp' **

**--fields-terminated-by "," **

**--split-by id **

**--columns id,salary **

**--target-dir '/sqoopoutput' **

**--hive-import **

**--create-hive-table **

--hive-table hivedb.emp_salary ;

INPUT MYSQL TABLE:

```
mysql> select * from emp;
+------+----------+-------------+--------+
| id   | name     | designation | salary |
+------+----------+-------------+--------+
| 1200 | Mark     | Architect   |  50000 |
| 1201 | Nick     | PM          |  40000 |
| 1202 | Arthur   | Developer   |  30000 |
| 1203 | Markus   | Architect   |  55000 |
| 1204 | Nicholas | PM          |  45000 |
| 1205 | Arnica   | Developer   |  36000 |
| 1206 | Monica   | Architect   |  50100 |
| 1207 | Nancy    | PM          |  90000 |
| 1208 | Angeline | Developer   |  60000 |
+------+----------+-------------+--------+
9 rows in set (0.00 sec)
```

OUTPUT HIVE TABLE:

```
hive> select * from emp_salary;
OK
1200    50000
1201    40000
1202    30000
1203    55000
1204    45000
1205    36000
1206    50100
1207    90000
1208    60000
Time taken: 1.647 seconds, Fetched: 9 row(s)
hive>
```

2.2 Input Commands to EXPORT SELECTED COLUMNS INTO MYSQL FROM HIVE

[cloudera@quickstart ~]$ sudo hive

hive> use hivedb;

hive> create table if not exists hiveemp

(

id int,

name string,

designation string,

salary int ,

dept string

)

ROW FORMAT DELIMITED

fields terminated by ',' stored as TEXTFILE;

```
hive> LOAD DATA
LOCAL INPATH '/home/cloudera/chhaya/sqoop/empdata.txt'
OVERWRITE INTO TABLE hiveemp;

hive> INSERT OVERWRITE DIRECTORY '/user/hive/warehouse/hivedb.db/hiveemp/sel'
SELECT id,salary FROM hiveemp;



[cloudera@quickstart ~]$ sudo service mysqld start
[cloudera@quickstart ~]$ mysql -u root -pcloudera
mysql> use chhayadb1 ;
mysql> CREATE TABLE empsalaryexport
    ( empid int not null primary key,
      empsalary int
      );
mysql> SELECT * FROM empsalaryexport;
mysql> exit;



[cloudera@quickstart oozie]$sqoop export \
--connect jdbc:mysql://localhost/chhayadb1 \
--username 'root' -P \
--columns empid,empsalary \
--table 'empsalaryexport' \
--export-dir '/user/hive/warehouse/hivedb.db/hiveemp/sel' \
--input-fields-terminated-by '\001' \
--m 1 ;
```

Input HIVE FILE:

```
hive> select * from hiveemp;
OK
1100    Markus  Architect       55000   ECS
1101    Nicholas        PM      45000   ECS
1102    Arnica  Developer       36000   FSI
1103    Monica  Architect       50100   FSI
1104    Nancy   PM      90000   FSI
1105    Angeline        Developer       60000   FSI
Time taken: 0.076 seconds, Fetched: 6 row(s)
hive>
```

Output MYSQL FILE:

```
mysql> select * from empsalaryexport
    -> ;
+-------+-----------+
| empid | empsalary |
+-------+-----------+
|  1100 |     55000 |
|  1101 |     45000 |
|  1102 |     36000 |
|  1103 |     50100 |
|  1104 |     90000 |
|  1105 |     60000 |
+-------+-----------+
6 rows in set (0.00 sec)

mysql>
```