**Project Implementation tasks**

**Task 1:**

=====================================================================================

Move file from windows Desktop/Shared to Cloudera Desktop/Shared using Shared folder

**Step1: -**

Create directory project\_input\_data

*hadoop fs -mkdir project\_input\_data*

**Step2: -**

Moving file from cloudera local to newly created HDFS folder project\_input\_data

*hadoop fs -put Desktop/card\_transactions.csv project\_input\_data*

**Step3: -**

Verifying the record count of the card\_transactions file loaded into HDFS

hadoop fs -cat project\_input\_data/card\_transactions.csv | wc -l



**Task 2:**

=====================================================================================

**Step1: -**

Login to mysql

*mysql -u root -p*

Creating database bigdataproject and using same

*create database bigdataproject;*

*use bigdataproject;*



**Step2: -**

Creating table card\_transactions in MySQL based on the card\_transactions.csv file structure.

*create table card\_transactions (*

*card\_id bigint,*

*member\_id bigint,*

*amount int,*

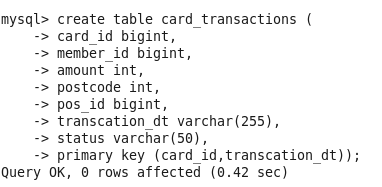
*postcode int,*

*pos\_id bigint,*

*transcation\_dt varchar(255),*

*status varchar(50),*

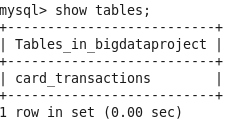
*primary key (card\_id,transcation\_dt));*



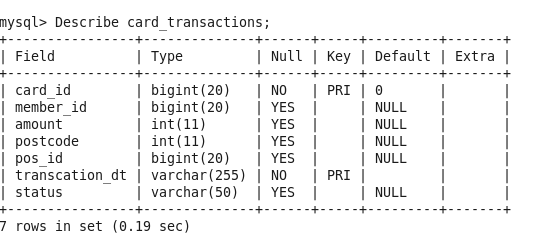
**Step3: -**

Showing tables created and the description of the same

*show tables;*



*Describe card\_transactions;*



**Task 3:**

=====================================================================================

**Step1: -**

sqoop export \

--connect jdbc:mysql://quickstart.cloudera:3306/bigdataproject \

--username root \

--password cloudera \

--table card\_transactions \

--export-dir project\_input\_data/card\_transactions.csv \

--fields-terminated-by ','

select \* from bigdataproject.card\_transactions;

**Step2: -**

Deleting the file from HDFS

hadoop fs -rm -R project\_input\_data

**Task 4:**

=====================================================================================

**Step1: -**

Enter to Hive Terminal

$hive

Create database in Hive

*create database bigdataproject;*

*use database bigdataproject;*

Create hive external table- member\_score

*create external table if not exists member\_score*

*(*

*member\_id string,*

*score float*

*)*

*row format delimited*

*fields terminated by ','*

*lines terminated by '\n'*

*stored as textfile*

*tblproperties("skip.header.line.count"="1");*

**Step2: -**

Create hive external table- member\_details

*create external table if not exists member\_details*

*(*

*card\_id bigint,*

*member\_id bigint,*

*member\_joining\_dt timestamp ,*

*card\_purchase\_dt timestamp ,*

*country string,*

*city string,*

*score float*

*)*

*row format delimited*

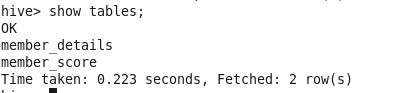
*fields terminated by ','*

*lines terminated by '\n'*

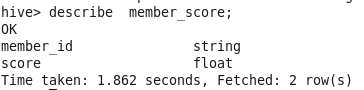
*stored as textfile*

*tblproperties("skip.header.line.count"="1");*

show tables;



Describe member\_score;



Describe member\_details;

