SQOOP ASSIGNMENT

1) Suppose we have a test_db database in mysql. We have an input table Customers inside test_db. (SQL Commands are given)

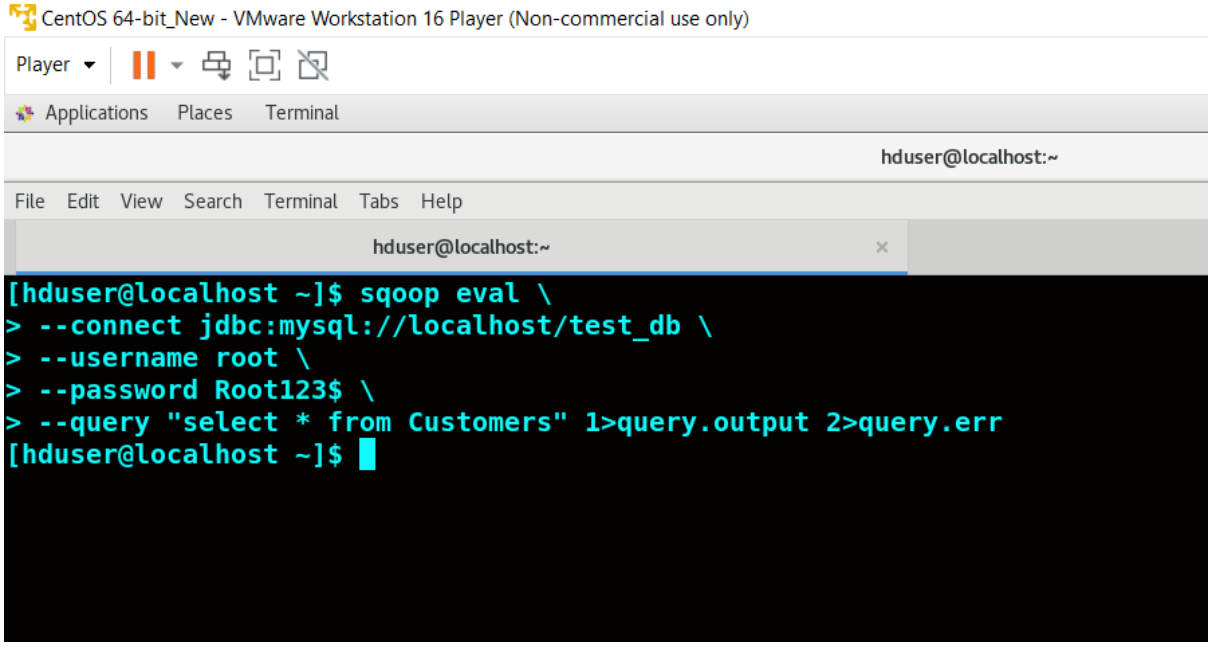
Customers

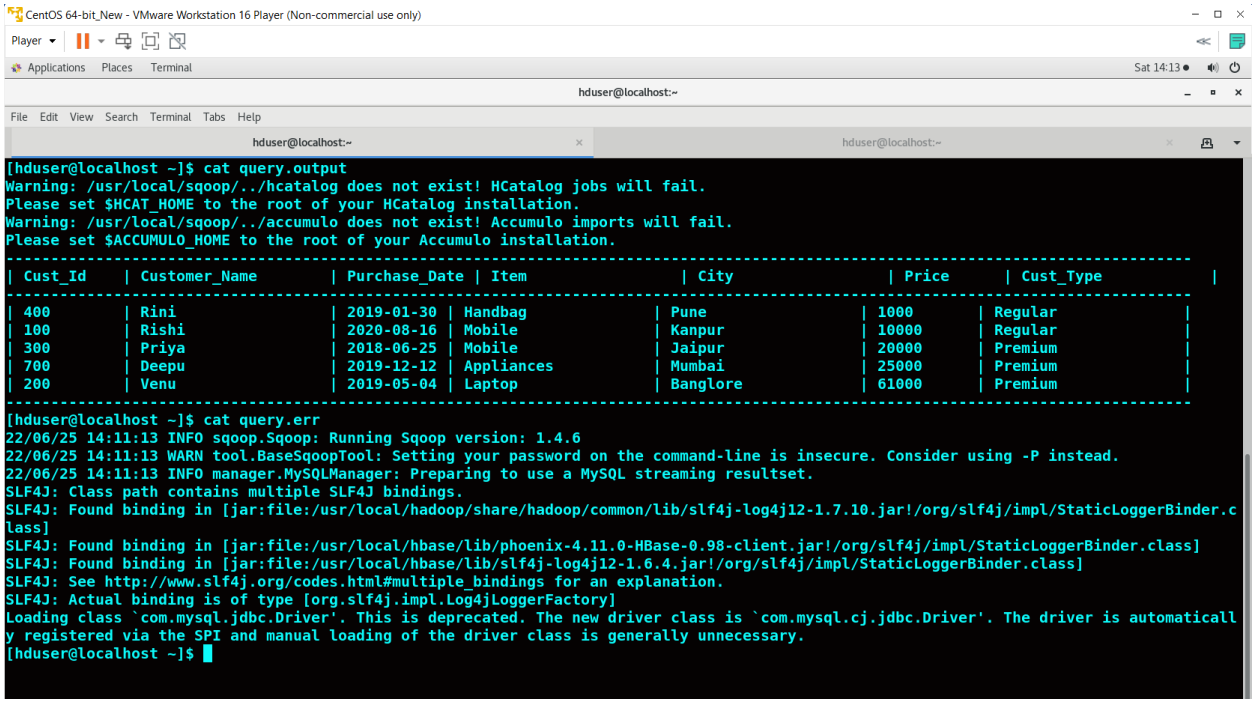
Cust_Id	Customer_Name	Purchase_Date	Item	City	Price	Cust_Type
100	Rishi	2020-08-16	Mobile	Kanpur	10000	Regular
200	Venu	2019-05-04	Laptop	Banglore	61000	Premium
300	Priya	2018-06-25	Mobile	Jaipur	20000	Premium
400	Rini	2019-01-30	Handbag	Pune	1000	Regular
700	Deepu	2019-12-12	Appliances	Mumbai	25000	Premium

The table has a Primary key on the Price column (which of course is not the right choice as prices may repeat when data grows).

Do the following: Share Snapshots of the command and Snapshot of the result in each case:

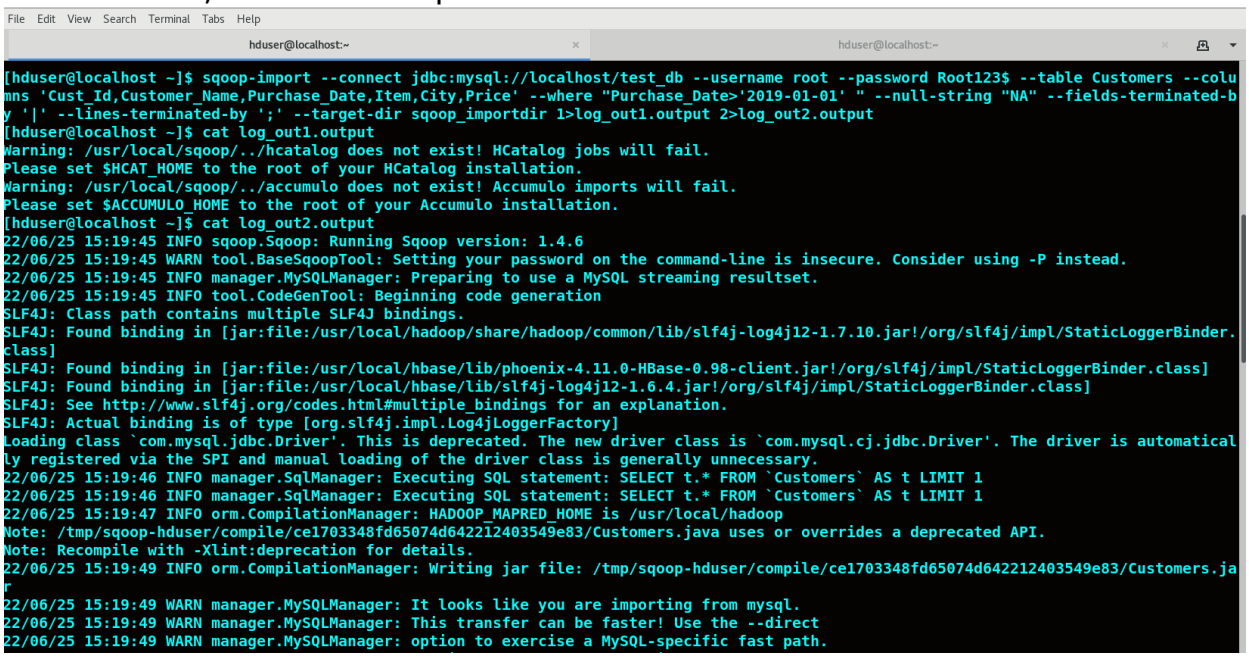
1) Before performing the sqoop import, using the sqoop command display the data present in mysql Customers table.The output of the command should not display on the console,rather should be redirected to log file named 'query.output'. Display the contents of the query.output file , share the Snapshot of the command and the output .

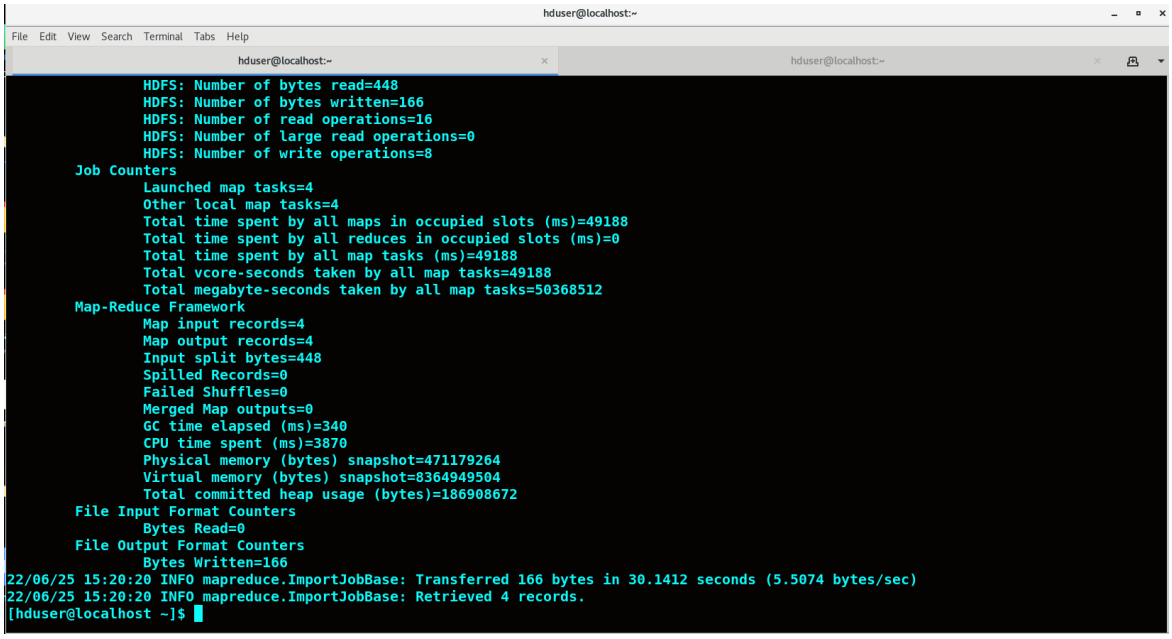




2) Perform a single sqoop import inside the directory in hdfs named sqoop_importdir, considering all the following points:

- Import all the columns except Cust_Type in hdfs.
- Include only the purchases made after 2019-01-01
- The output data generated should have fields separated by | and rows separated by ; (semicolon)
- While importing, Nulls in the data , should be overridden with 'NA'
- Redirect the log messages generated on screen to the files log_out1 and log_out2.Display the contents of the log_out2 file , when sqoop import is successful,share the snapshot of the number of records retrieved.

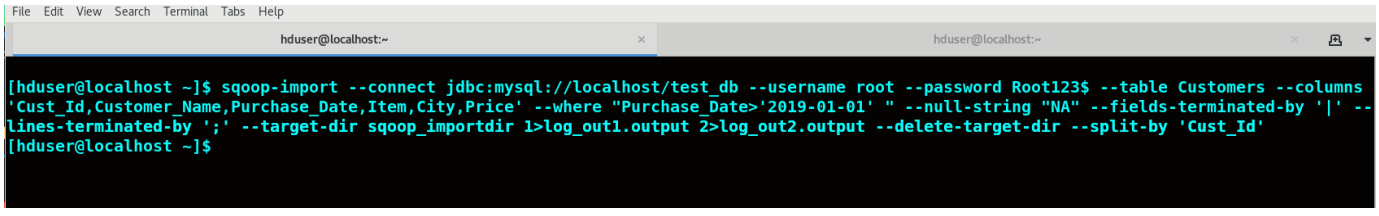




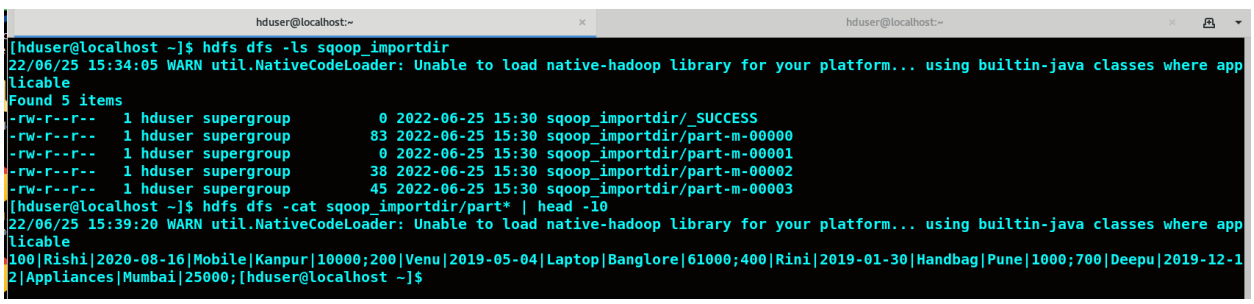
- Display the contents of the sqoop_importdir

```
[hduser@localhost ~]$ hdfs dfs -ls sqoop_importdir
22/06/25 15:25:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
Found 5 items
-rw-r--r--   1 hduser supergroup          0 2022-06-25 15:20 sqoop_importdir/_SUCCESS
-rw-r--r--   1 hduser supergroup         79 2022-06-25 15:20 sqoop_importdir/part-m-00000
-rw-r--r--   1 hduser supergroup         45 2022-06-25 15:20 sqoop_importdir/part-m-00001
-rw-r--r--   1 hduser supergroup          0 2022-06-25 15:20 sqoop_importdir/part-m-00002
-rw-r--r--   1 hduser supergroup         42 2022-06-25 15:20 sqoop_importdir/part-m-00003
[hduser@localhost ~]$ hdfs dfs -cat sqoop_importdir/part*
22/06/25 15:26:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
400|Rini|2019-01-30|Handbag|Pune|1000;100|Rishi|2020-08-16|Mobile|Kanpur|10000;700|Deepu|2019-12-12|Appliances|Mumbai|25000;200|Venu|201
9-05-04|Laptop|Banglore|61000;[hduser@localhost ~]$
```

- Now Again modify and run your sqoop import command, so that cust_id col can be used to decide the input splits, as the Primary key col is not proper. Also ensure that the output directory remains as sqoop_importdir, and the previously imported contents are automatically deleted and new contents are filled in the o/p directory.



- Display the contents of the output directory now and the first 10 records from the mapper output files (hint: use head command)



- Now Suppose an outlier comes into the mysql table:

  The new record inserted is :

Cust_Id	Customer_Name	Purchase_Date	Item	City	Price	Cust_Type
10000	Raman	2019/09/04	Misc	Cochin	9000	Regular

```
mysql> insert into Customers values(10000,'Raman','2019-09-04','Misc','Cochin',9000,'Regular');
Query OK, 1 row affected (0.00 sec)

mysql> select * from Customers;
+---------+---------------+---------------+------------+----------+-------+-----------+
| Cust_Id | Customer_Name | Purchase_Date | Item       | City     | Price | Cust_Type |
+---------+---------------+---------------+------------+----------+-------+-----------+
|     400 | Rini          | 2019-01-30    | Handbag    | Pune     |  1000 | Regular   |
|   10000 | Raman         | 2019-09-04    | Misc       | Cochin   |  9000 | Regular   |
|     100 | Rishi         | 2020-08-16    | Mobile     | Kanpur   | 10000 | Regular   |
|     300 | Priya         | 2018-06-25    | Mobile     | Jaipur   | 20000 | Premium   |
|     700 | Deepu         | 2019-12-12    | Appliances | Mumbai   | 25000 | Premium   |
|     200 | Venu          | 2019-05-04    | Laptop     | Banglore | 61000 | Premium   |
+---------+---------------+---------------+------------+----------+-------+-----------+
6 rows in set (0.00 sec)

mysql>
```

- Mention the sqoop import command you will frame from your end to deal with such a situation to ensure even work distribution among mappers, using customized bounding val query.

```
[hduser@localhost ~]$ sqoop-import --connect jdbc:mysql://localhost/test_db --username root --password Root123$ --table Customers --boundary
-query "select 100,700 from Customers" --target-dir sqoop_importdir1 --delete-target-dir --split-by 'Cust_Id'
Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/06/25 16:54:26 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
22/06/25 16:54:26 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/06/25 16:54:27 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/06/25 16:54:27 INFO tool.CodeGenTool: Beginning code generation
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.clas
s]
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/phoenix-4.11.0-HBase-0.98-client.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
```

```
hduser@localhost:~                                    hduser@localhost:~
[hduser@localhost ~]$ hdfs dfs -ls sqoop_importdir1/
22/06/25 16:55:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where app
licable
Found 5 items
-rw-r--r--   1 hduser supergroup          0 2022-06-25 16:54 sqoop_importdir1/_SUCCESS
-rw-r--r--   1 hduser supergroup         99 2022-06-25 16:54 sqoop_importdir1/part-m-00000
-rw-r--r--   1 hduser supergroup         49 2022-06-25 16:54 sqoop_importdir1/part-m-00001
-rw-r--r--   1 hduser supergroup         46 2022-06-25 16:54 sqoop_importdir1/part-m-00002
-rw-r--r--   1 hduser supergroup         53 2022-06-25 16:54 sqoop_importdir1/part-m-00003
[hduser@localhost ~]$ hdfs dfs -cat sqoop_importdir1/part*
22/06/25 16:56:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where app
licable
100,Rishi,2020-08-16,Mobile,Kanpur,10000,Regular
200,Venu,2019-05-04,Laptop,Banglore,61000,Premium
300,Priya,2018-06-25,Mobile,Jaipur,20000,Premium
400,Rini,2019-01-30,Handbag,Pune,1000,Regular
700,Deepu,2019-12-12,Appliances,Mumbai,25000,Premium
[hduser@localhost ~]$
```

==========

2) Suppose we have a database named test_new_db in mysql, We have three tables inside it:

City_Tbl (Consider this is the bigger table)

State_Tbl (Consider this is the smaller table)

Country_Tbl (Smaller Table)

City_Tbl: City_ID is the Primary Key Column

City_Name	City_ID
Bangalore	1000
Mumbai	1001
Chennai	1002
Kolkata	1003
Delhi	1004
Pune	1005
Nagpur	1006
Surat	1007
Kochi	1008

State_Tbl: No Primary Key Column

State_Name	Districts
Karnataka	30
TamilNadu	32
Goa	2
Kerala	14
Assam	33

Country_Tbl: No Primary Key Column

Name	Country_Code
Belgium	32
Brazil	55
France	33
Iran	98
India	91

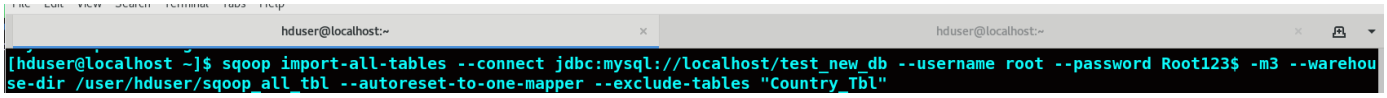
A) Using a single sqoop import command,

Import all the tables present in test_new_db to hdfs excluding the Country_Tbl .

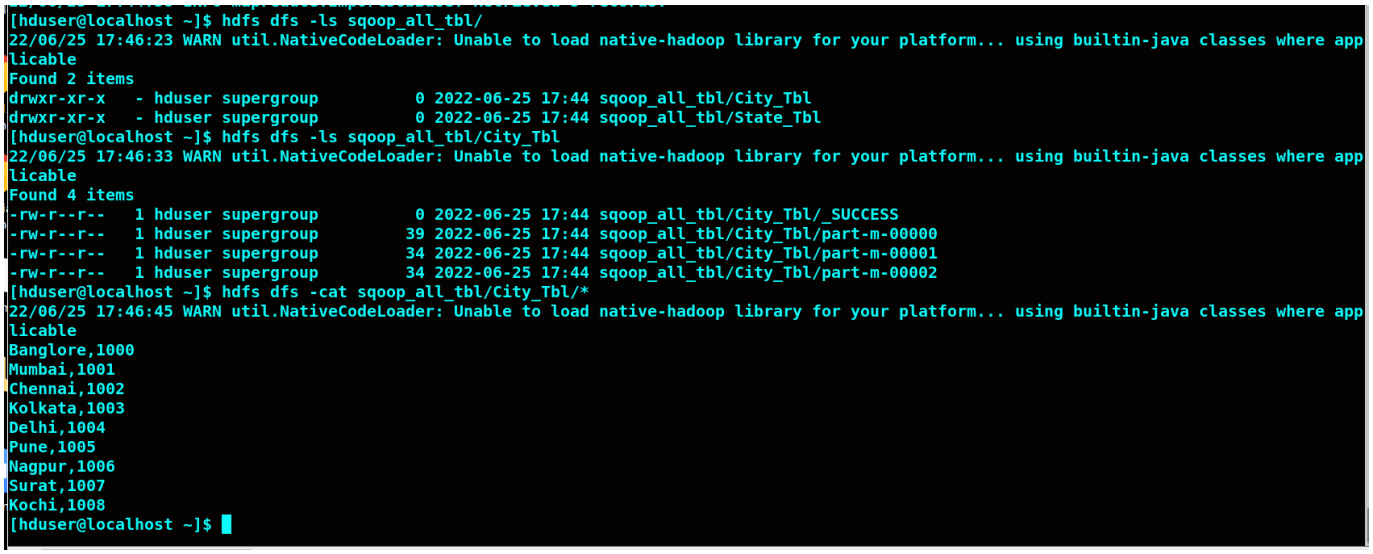
You have to do it with a single sqoop command.

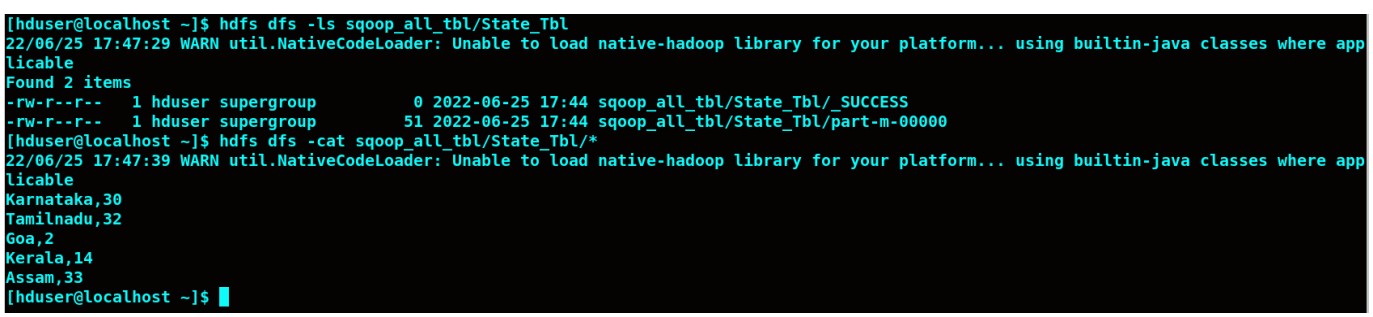
Also, City_Tbl should have 3 output files generated in hdfs. All the output files should be stored inside sqoop_all_tbl directory in hdfs, with sub-directories of each table name created inside the main directory. Share the snapshot of the command.

$ sqoop import-all-tables --connect jdbc:mysql://localhost/test_new_db --username root --password Root123$ -m3 --warehouse-dir /user/hduser/sqoop_all_tbl --autoreset-to-one-mapper --exclude-tables "Country_Tbl"



B) Show the contents of the output directory: (Share Snapshot)





==========

3) We have a Categories Table in test_db in Mysql. On this table both inserts and updates are performed from time to time.

Do the following:

A)Import the Categories table in hdfs but during the import , do proper Null value handling:
• String Columns nulls should be replaced with '\N' (so that in file it should be read as \n and Non-string column nulls should be replaced with -1
• Use a warehouse directory
• We also want to see the query run by each mapper internally
Share the import command you will use,keeping in mind all of the above. Initially all records to be pulled in.

$ sqoop-import --connect jdbc:mysql://localhost/test_db --username root --password Root123$ --table Categories --warehouse-dir importdir --null-string '\\N' --null-non-string '-1'

```
[hduser@localhost ~]$ sqoop-import --connect jdbc:mysql://localhost/test_db --username root --password Root123$ --table Categories --warehou
se-dir importdir --null-string '\\N' --null-non-string '-1'
```

```
[hduser@localhost ~]$ hdfs dfs -ls importdir/
22/06/25 18:15:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
Found 1 items
drwxr-xr-x   - hduser supergroup          0 2022-06-25 18:14 importdir/Categories
[hduser@localhost ~]$ hdfs dfs -ls importdir/Categories
22/06/25 18:16:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
Found 5 items
-rw-r--r--   1 hduser supergroup          0 2022-06-25 18:14 importdir/Categories/_SUCCESS
-rw-r--r--   1 hduser supergroup        116 2022-06-25 18:14 importdir/Categories/part-m-00000
-rw-r--r--   1 hduser supergroup        103 2022-06-25 18:14 importdir/Categories/part-m-00001
-rw-r--r--   1 hduser supergroup        122 2022-06-25 18:14 importdir/Categories/part-m-00002
-rw-r--r--   1 hduser supergroup        103 2022-06-25 18:14 importdir/Categories/part-m-00003
[hduser@localhost ~]$ hdfs dfs -cat importdir/Categories/part-m-00000
22/06/25 18:16:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
1,2,Football,2020-04-30 00:00:00.0
2,2,Handball,2020-05-01 00:00:00.0
3,2,Baseball & Softball,2020-05-01 00:00:00.0
[hduser@localhost ~]$ hdfs dfs -cat importdir/Categories/part-m-00001
22/06/25 18:16:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
4,2,Basketball,2020-04-30 00:00:00.0
5,3,Tennis,2020-04-30 00:00:00.0
6,3,Hockey,2020-05-01 00:00:00.0
[hduser@localhost ~]$ hdfs dfs -cat importdir/Categories/part-m-00002
22/06/25 18:16:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
7,3,Swimming,2020-05-01 00:00:00.0
8,3,Cardio Equipment,2020-05-01 00:00:00.0
9,4,Strength Training,2020-05-01 00:00:00.0
[hduser@localhost ~]$ hdfs dfs -cat importdir/Categories/part-m-00003
22/06/25 18:16:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
10,4,Athletics,2020-05-02 00:00:00.0
11,-1,Cycling,2020-02-02 00:00:00.0
12,5,\N,2020-01-15 00:00:00.0
[hduser@localhost ~]$
```

```
[hduser@localhost ~]$ hdfs dfs -cat importdir/Categories/*
22/06/25 18:18:16 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
1,2,Football,2020-04-30 00:00:00.0
2,2,Handball,2020-05-01 00:00:00.0
3,2,Baseball & Softball,2020-05-01 00:00:00.0
4,2,Basketball,2020-04-30 00:00:00.0
5,3,Tennis,2020-04-30 00:00:00.0
6,3,Hockey,2020-05-01 00:00:00.0
7,3,Swimming,2020-05-01 00:00:00.0
8,3,Cardio Equipment,2020-05-01 00:00:00.0
9,4,Strength Training,2020-05-01 00:00:00.0
10,4,Athletics,2020-05-02 00:00:00.0
11,-1,Cycling,2020-02-02 00:00:00.0
12,5,\N,2020-01-15 00:00:00.0
[hduser@localhost ~]$
```

B) New Records are added to the table and also existing records are updated,(refer the mysql_commands text file for the insert and update commands), so import only those newly inserted/updated records from Categories table to hdfs.

The delta records should get appended to existing directory.

Inserted one record and updated one record

```
mysql> insert into Categories values(13,5,'Volleyball','2021-01-15');
Query OK, 1 row affected (0.01 sec)

mysql> update Categories set category_name='Tennis',inclusion_date='2021-02-02' where Category_id=10;
Query OK, 1 row affected (0.01 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> select *from Categories;
+-------------+-----------------------+--------------------+---------------------+
| category_id | category_department_id | category_name      | inclusion_date      |
+-------------+-----------------------+--------------------+---------------------+
|           1 |                     2 | Football           | 2020-04-30 00:00:00 |
|           2 |                     2 | Handball           | 2020-05-01 00:00:00 |
|           3 |                     2 | Baseball & Softball | 2020-05-01 00:00:00 |
|           4 |                     2 | Basketball         | 2020-04-30 00:00:00 |
|           5 |                     3 | Tennis             | 2020-04-30 00:00:00 |
|           6 |                     3 | Hockey             | 2020-05-01 00:00:00 |
|           7 |                     3 | Swimming           | 2020-05-01 00:00:00 |
|           8 |                     3 | Cardio Equipment   | 2020-05-01 00:00:00 |
|           9 |                     4 | Strength Training  | 2020-05-01 00:00:00 |
|          10 |                     4 | Tennis             | 2021-02-02 00:00:00 |
|          11 |                  NULL | Cycling            | 2020-02-02 00:00:00 |
|          12 |                     5 | NULL               | 2020-01-15 00:00:00 |
|          13 |                     5 | Volleyball         | 2021-01-15 00:00:00 |
+-------------+-----------------------+--------------------+---------------------+
13 rows in set (0.00 sec)
```

Share the import command you will use this time, to get only delta records

$ sqoop import --connect jdbc:mysql://localhost/test_db --username root --password Root123$ --table Categories --warehouse-dir importdir -m1 --incremental lastmodified --check-column inclusion_date --last-value 2020-05-03 --append

```
[hduser@localhost ~]$ hdfs dfs -ls importdir/Categories
22/06/25 18:47:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
Found 6 items
-rw-r--r--   1 hduser supergroup          0 2022-06-25 18:14 importdir/Categories/_SUCCESS
-rw-r--r--   1 hduser supergroup        116 2022-06-25 18:14 importdir/Categories/part-m-00000
-rw-r--r--   1 hduser supergroup        103 2022-06-25 18:14 importdir/Categories/part-m-00001
-rw-r--r--   1 hduser supergroup        122 2022-06-25 18:14 importdir/Categories/part-m-00002
-rw-r--r--   1 hduser supergroup        103 2022-06-25 18:14 importdir/Categories/part-m-00003
-rw-r--r--   1 hduser supergroup         72 2022-06-25 18:47 importdir/Categories/part-m-00004
[hduser@localhost ~]$ hdfs dfs -cat importdir/Categories/part-m-00004
22/06/25 18:47:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
10,4,Tennis,2021-02-02 00:00:00.0
13,5,Volleyball,2021-01-15 00:00:00.0
[hduser@localhost ~]$ hdfs dfs -cat importdir/Categories/part*
22/06/25 18:48:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
1,2,Football,2020-04-30 00:00:00.0
2,2,Handball,2020-05-01 00:00:00.0
3,2,Baseball & Softball,2020-05-01 00:00:00.0
4,2,Basketball,2020-04-30 00:00:00.0
5,3,Tennis,2020-04-30 00:00:00.0
6,3,Hockey,2020-05-01 00:00:00.0
7,3,Swimming,2020-05-01 00:00:00.0
8,3,Cardio Equipment,2020-05-01 00:00:00.0
9,4,Strength Training,2020-05-01 00:00:00.0
10,4,Athletics,2020-05-02 00:00:00.0
11,-1,Cycling,2020-02-02 00:00:00.0
12,5,\N,2020-01-15 00:00:00.0
10,4,Tennis,2021-02-02 00:00:00.0
13,5,Volleyball,2021-01-15 00:00:00.0
[hduser@localhost ~]$
```

C) After this second import, how many records do you see in the hdfs folder now? Did you find any duplicate records, give details if any.

There are 14 records.Yes, There is one duplicate record after updation.

D) Create a new table in test_db named Categories_new.This newly created table does not have a Primary key.
We want to do periodic imports and updates in this mysql table. But we do not want any duplicate records in the hdfs post import.
Also we want to automate the process of import & want a good way to manage the password. Choose a different warehouse directory this time.
Share the commands you will use when:

- First time we need to pull all records in hdfs

$sqoop job --create lmCatnewjob1 -- import --connect jdbc:mysql://localhost/test_db --username root --password Root123$ --table Categories_new --warehouse-dir importDir -m1 --incremental lastmodified --check-column inclusion_date --last-value 2000-01-01 –append
$sqoop job --exec lmCatnewjob1

Initially 9 records

```
[hduser@localhost ~]$ hdfs dfs -ls importDir/
22/06/25 19:30:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library f
Found 1 items
-rw-r--r--   1 hduser supergroup        341 2022-06-25 19:30 importDir/part-m-00000
[hduser@localhost ~]$ hdfs dfs -cat importDir/*
22/06/25 19:30:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library f
1,2,Football,2020-04-30 00:00:00.0
2,2,Handball,2020-05-01 00:00:00.0
3,2,Baseball & Softball,2020-05-01 00:00:00.0
4,2,Basketball,2020-04-30 00:00:00.0
5,3,Tennis,2020-04-30 00:00:00.0
6,3,Hockey,2020-05-01 00:00:00.0
7,3,Swimming,2020-05-01 00:00:00.0
8,3,Cardio Equipment,2020-05-01 00:00:00.0
9,4,Strength Training,2020-05-01 00:00:00.0
[hduser@localhost ~]$
```

- Second time to pull only the delta records,but without duplicates in hdfs

-Inserted 10th record and updated 8th record

$sqoop import --connect jdbc:mysql://localhost/test_db --username root --password Root123$ --table Categories_new --target-dir importDir -m1 --incremental lastmodified --check-column inclusion_date --last-value 2020-05-03 –append

```
[hduser@localhost ~]$ hdfs dfs -cat importDir/part*
22/06/25 19:39:06 WARN util.NativeCodeLoader: Unable t
1,2,Football,2020-04-30 00:00:00.0
2,2,Handball,2020-05-01 00:00:00.0
3,2,Baseball & Softball,2020-05-01 00:00:00.0
4,2,Basketball,2020-04-30 00:00:00.0
5,3,Tennis,2020-04-30 00:00:00.0
6,3,Hockey,2020-05-01 00:00:00.0
7,3,Swimming,2020-05-01 00:00:00.0
8,3,Cardio Equipment,2020-05-01 00:00:00.0
9,4,Strength Training,2020-05-01 00:00:00.0
10,5,Volleyball,2021-01-15 00:00:00.0
8,3,Tennis,2021-02-02 00:00:00.0
[hduser@localhost ~]$
```

$ sqoop job --create lmCatnewjob2 -- import --connect jdbc:mysql://localhost/test_db --username root --password Root123$ --table Categories_new --target-dir importDir/Categories_new -m1 --incremental lastmodified --check-column inclusion_date --last-value 2020-05-03 --merge-key category_id

$ sqoop job --exec lmCatnewjob2

```
[hduser@localhost ~]$ hdfs dfs -ls importDir
22/06/25 19:48:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
Found 2 items
-rw-r--r--   1 hduser supergroup          0 2022-06-25 19:48 importDir/_SUCCESS
-rw-r--r--   1 hduser supergroup        369 2022-06-25 19:48 importDir/part-r-00000
[hduser@localhost ~]$ hdfs dfs -cat importDir/*
22/06/25 19:48:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
1,2,Football,2020-04-30 00:00:00.0
10,5,Volleyball,2021-01-15 00:00:00.0
2,2,Handball,2020-05-01 00:00:00.0
3,2,Baseball & Softball,2020-05-01 00:00:00.0
4,2,Basketball,2020-04-30 00:00:00.0
5,3,Tennis,2020-04-30 00:00:00.0
6,3,Hockey,2020-05-01 00:00:00.0
7,3,Swimming,2020-05-01 00:00:00.0
8,3,Tennis,2021-02-02 00:00:00.0
9,4,Strength Training,2020-05-01 00:00:00.0
[hduser@localhost ~]$
```

E) How many records do you see this time in hdfs post second import? Do you see any duplicate records now?

Initially there are 9 records. Inserted one record and updated one record. After the first import(incremented lastmodified apped) there are 11 records with one duplicate record.After second import(incremental lastmodified merge-key) there are 10 records with no duplicate records .

F) Are any mapper files generated in hdfs this time after the second import? Explain.

There are no mapper files after second import.There is reduce file with name as part-r-00000 in which all the records are present

G) Share the command you will use to see the last value of a Saved Sqoop Job.

$ *sqoop job --show lmCatnewjob2*

sqoop Quiz

1. Sqoop written in?

A. C

B. C++

C. Java

D. hadoop

2. Sqoop stands for?

A. SQL to Hadoop

B. SQL to Hbase

C. MySQL to Hadoop

D. SQL Hadoop

3. Is Apache Sqoop is an open-source tool?

A. TRUE

B. FALSE

C. Can be true or false

D. Can not say

4. Data processed by Scoop can be used for?

A. Hbase

B. HDFS

C. Mapreduce

D. MahOut

5. _____ tool can list all the available database schemas

A. sqoop-list-tables

B. sqoop-list-databases

C. sqoop-list-schema

D. sqoop-list-columns

6. The active Hadoop configuration is loaded from $HADOOP_HOME/conf/, unless the $HADOOP_CONF_DIR environment variable is unset.

A. TRUE

B. FALSE

C. Can be true or false

D. Can not say

7. Data can be imported in maximum _____ file formats.

A. 2

B. 3

C. 4

D. 5

8. If you set the inline LOB limit to _____ all large objects will be placed in external storage.

A. 0

B. 2

C. 3

D. 1

9. The import-tables tool imports a set of tables from an RDBMS to?

A. Hive

B. Sqoop

C. HDFS

D. Mapreduce

10. Sqoop can also import the data into Hive by generating and executing a _____ statement to define the data's layout in Hive.

A. SET TABLE

B. CREATE TABLE

C. INSERT TABLE

D. All of the above

11. The following tool imports a set of tables from an RDBMS to HDFS

A. export-all-tables

B. import-all-tables

C. import-tables

D. none of the mentioned

12. With the –staging-table parameter, the data is moved from staging to final table

A. Automatically if staging load is successful

B. Has to be done by user after verifying the data in staging

C. Depends on the data size

D. Depends on the memory available to move the data