A1.

1. Create the database and the required table with data

```
mysql> create database test_db;
Query OK, 1 row affected (0.00 sec)

mysql> use test_db;
Database changed
mysql> CREATE TABLE Customers
    -> (
    -> Cust_Id INT NOT NULL,
    -> Customer_Name VARCHAR(255),
    -> Purchase_Date DATE NOT NULL,
    -> Item VARCHAR(255),
    -> City VARCHAR(255),
    -> Price INT PRIMARY KEY,
    -> Cust_Type VARCHAR(255)
    -> );
Query OK, 0 rows affected (0.04 sec)

mysql> INSERT INTO Customers values
    -> (100,'Rishi','2020-08-16','Mobile','Kanpur',10000 ,'Regular'),
    -> (200,'Venu','2019-05-04','Laptop','Bangalore',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');
Query OK, 5 rows affected (0.01 sec)
Records: 5  Duplicates: 0  Warnings: 0

mysql> commit;
Query OK, 0 rows affected (0.00 sec)
```

2. using the sqoop command display the data present in mysql Customers table in a file called query.output

```
[root@quickstart Desktop]# sqoop-eval --connect jdbc:mysql://quickstart.cloudera:3306 --username root --password cloudera --query "select * from test_db.Customers" 1>query.output
23/07/11 10:44:20 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
23/07/11 10:44:20 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/07/11 10:44:20 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
[root@quickstart Desktop]# cat query.output
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
-------------------------------------------------------------------------------------------------------------------------------
| Cust_Id     | Customer_Name        | Purchase_Date | Item                | City                | Price       | Cust_Type            |
-------------------------------------------------------------------------------------------------------------------------------
| 400         | Rini                 | 2019-01-30 | Handbag             | Pune                | 1000        | 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              | Bangalore           | 61000       | Premium              |
-------------------------------------------------------------------------------------------------------------------------------
```

3. sqoop import with price as primary key

```
[root@quickstart Desktop]# sqoop-import \
> --connect jdbc:mysql://quickstart.cloudera:3306/test_db \
> --username root \
> --password cloudera \
> --columns Cust_Id, Customer_Name, Purchase_Date, Item, City, Price \
> --table Customers \
> --where "Purchase_Date>'2019-01-01'" \
> --fields-terminated-by '|' \
> --lines-terminated-by ';' \
> --target-dir /user/cloudera/sqoop_importdir 1>log_out1 2>log_out2
```

A2.

1. sqoop import with with split by cust_id

```
[root@quickstart Desktop]# sqoop-import \
> --connect jdbc:mysql://quickstart.cloudera:3306/test_db \
> --username root \
> --password cloudera \
> --columns Cust_Id, Customer_Name, Purchase_Date, Item, City, Price \
> --table Customers \
> --split-by Cust_Id \
> --where "Purchase_Date>'2019-01-01'" \
> --fields-terminated-by '|' \
> --lines-terminated-by ';' \
> --target-dir /user/cloudera/sqoop_importdir 1>log_out1 2>log_out2
```

2. Display the contents of the output directory now and the first 10 records from the mapper output files
   hadoop fs -cat /user/cloudera/sqoop_importdir/
   hadoop fs -cat /user/cloudera/sqoop_importdir/*|head

4. Handling outlier

   Find the max and min values of the table except that ID

```
mysql> select max(Cust_Id),min(Cust_Id) from Customers
    -> where Cust_Id !=10000;
+--------------+--------------+
| max(Cust_Id) | min(Cust_Id) |
+--------------+--------------+
|          700 |          100 |
+--------------+--------------+
1 row in set (0.00 sec)
```

5. Apply bounding query

```
[root@quickstart Desktop]# sqoop-import \
> --connect jdbc:mysql://quickstart.cloudera:3306/test_db \
> --username root \
> --password cloudera \
> --columns Cust_Id, Customer_Name, Purchase_Date, Item, City, Price \
> --table Customers \
> --split-by Cust_Id \
> --boundary-query "select 100,700" \
> --where "Purchase_Date>'2019-01-01'" \
> --fields-terminated-by '|' \
> --lines-terminated-by ';' \
> --target-dir /user/cloudera/sqoop_importdir 1>log_out1 2>log_out2
```

A3.

1. Create the database and the required table with data

```
mysql> select * from City_Tbl;
+-----------+---------+
| City_Name | City_ID |
+-----------+---------+
| Bangalore |    1000 |
| Mumbai    |    1001 |
| Chennai   |    1002 |
| Kolkata   |    1003 |
| Delhi     |    1004 |
| Pune      |    1005 |
| Nagpur    |    1006 |
| Surat     |    1007 |
| Kochi     |    1008 |
+-----------+---------+
9 rows in set (0.00 sec)
```

```
mysql> select * from State_Tbl;
+------------+-----------+
| State_Name | Districts |
+------------+-----------+
| Karnataka  |        30 |
| TamilNadu  |        32 |
| Goa        |         2 |
| Kerala     |        14 |
| Assam      |        33 |
+------------+-----------+
5 rows in set (0.00 sec)
```

```
mysql> select * from Country_Tbl;
+---------+--------------+
| Name    | Country_Code |
+---------+--------------+
| Belgium |           32 |
| Brazil  |           55 |
| France  |           33 |
| Iran    |           98 |
| India   |           91 |
+---------+--------------+
5 rows in set (0.00 sec)
```

2. Sqoop import

```
[root@quickstart Desktop]# sqoop import-all-tables \
> --connect jdbc:mysql://quickstart.cloudera:3306/test_new_db \
> --username root \
> --password cloudera \
> --warehouse-dir /user/cloudera/sqoop_all_tbl \
> --exclude-tables Country_Tbl \
> --num-mappers 3 \
> --autoreset-to-one-mapper
```

3. Output

```
[root@quickstart Desktop]# hadoop fs -ls /user/cloudera/sqoop_all_tbl/
Found 2 items
drwxr-xr-x   - root cloudera          0 2023-01-12 00:59 /user/cloudera/sqoop_all_tbl/City_Tbl
drwxr-xr-x   - root cloudera          0 2023-01-12 00:59 /user/cloudera/sqoop_all_tbl/State_Tbl
```

A4.

1. Sqoop import-1

```
[root@quickstart Desktop]# sqoop-import \
> --connect jdbc:mysql://quickstart.cloudera:3306/test_db \
> --username root \
> --password cloudera \
> --table Categories \
> --null-non-string "-1" \
> --null-string '\\N' \
> --warehouse-dir /user/cloudera/sqoop_dir
```

Output

```
23/07/12 00:49:31 INFO mapreduce.ImportJobBase: Retrieved 12 records.
[root@quickstart Desktop]# hadoop fs -ls /user/cloudera/sqoop_dir/
Found 1 items
drwxr-xr-x   - root cloudera          0 2023-07-12 00:49 /user/cloudera/sqoop_dir/Categories
[root@quickstart Desktop]# hadoop fs -ls /user/cloudera/sqoop_dir/Categories
Found 5 items
-rw-r--r--   1 root cloudera          0 2023-07-12 00:49 /user/cloudera/sqoop_dir/Categories/_SUCCESS
-rw-r--r--   1 root cloudera        116 2023-07-12 00:49 /user/cloudera/sqoop_dir/Categories/part-m-00000
-rw-r--r--   1 root cloudera        103 2023-07-12 00:49 /user/cloudera/sqoop_dir/Categories/part-m-00001
-rw-r--r--   1 root cloudera        122 2023-07-12 00:49 /user/cloudera/sqoop_dir/Categories/part-m-00002
-rw-r--r--   1 root cloudera        103 2023-07-12 00:49 /user/cloudera/sqoop_dir/Categories/part-m-00003
[root@quickstart Desktop]# hadoop fs -cat /user/cloudera/sqoop_dir/Categories/*
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
```

2. Sqoop import-2

```
[root@quickstart Desktop]# sqoop-import \
> --connect jdbc:mysql://quickstart.cloudera:3306/test_db \
> --username root \
> --password cloudera \
> --table Categories \
> --null-non-string "-1" \
> --null-string '\\N' \
> --incremental lastmodified \
> --check-column inclusion_date \
> --last-value '2020-01-15 00:00:00' \
> --warehouse-dir /user/cloudera/sqoop_incremental_dir \
> --append
```

Output

```
[root@quickstart Desktop]# hadoop fs -ls /user/cloudera/sqoop_incremental_dir/Categories
Found 4 items
-rw-r--r--   1 root supergroup        153 2023-07-12 00:59 /user/cloudera/sqoop_incremental_dir/Categories/part-m-00000
-rw-r--r--   1 root supergroup        101 2023-07-12 00:59 /user/cloudera/sqoop_incremental_dir/Categories/part-m-00001
-rw-r--r--   1 root supergroup        124 2023-07-12 00:59 /user/cloudera/sqoop_incremental_dir/Categories/part-m-00002
-rw-r--r--   1 root supergroup        143 2023-07-12 00:59 /user/cloudera/sqoop_incremental_dir/Categories/part-m-00003
[root@quickstart Desktop]# hadoop fs -cat /user/cloudera/sqoop_incremental_dir/Categories/*
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
13,6,Surfing,2023-07-12 00:51:31.0
14,2,Mountaineering,2023-07-12 00:51:31.0
[root@quickstart Desktop]#
```