BIG DATA

SECTION D

Fall’2023

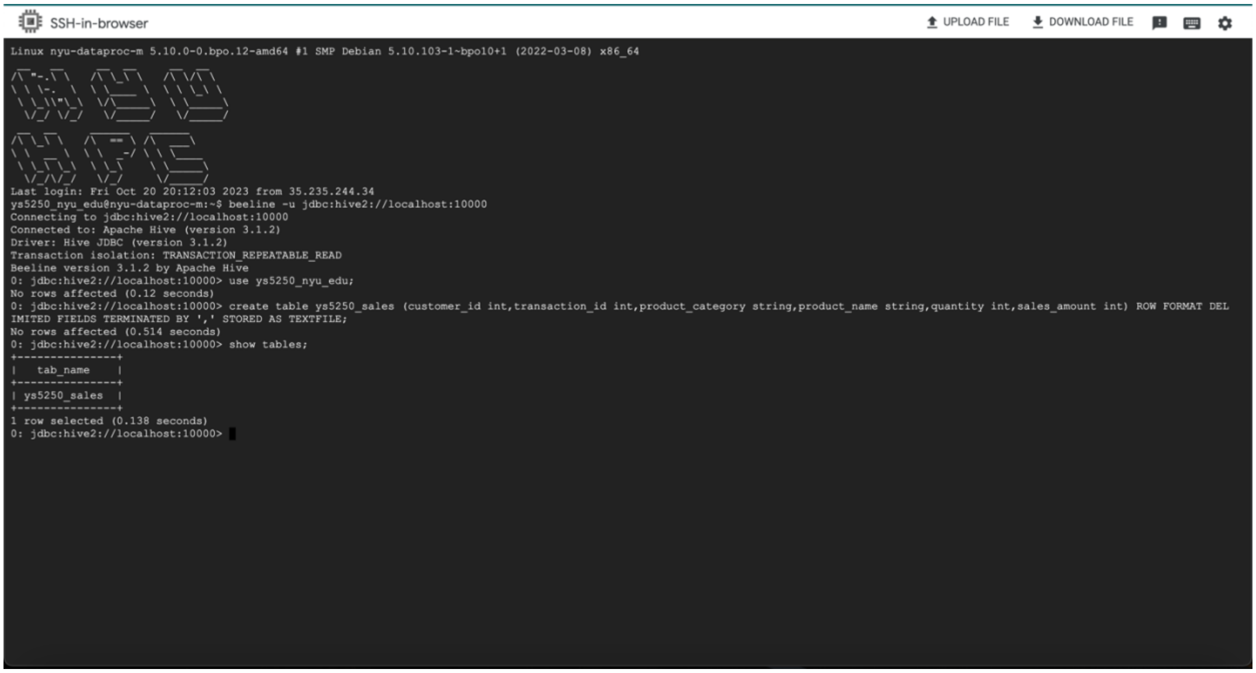
Yogya Sharma

ys5250

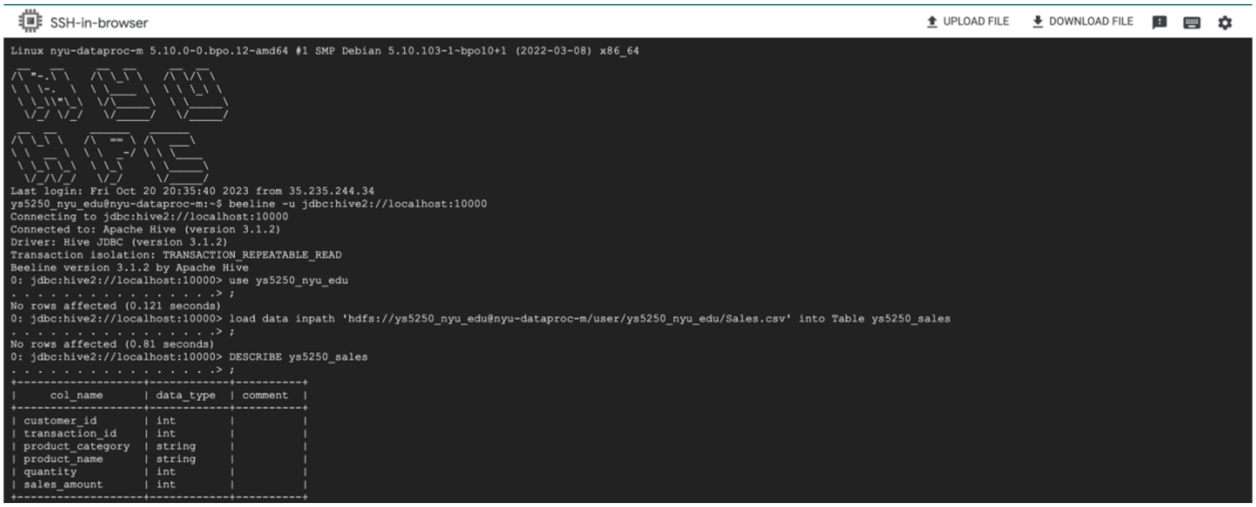
Practice Assignment – Hive SQL

A) Sales.csv

List table columns of sales with describe command:

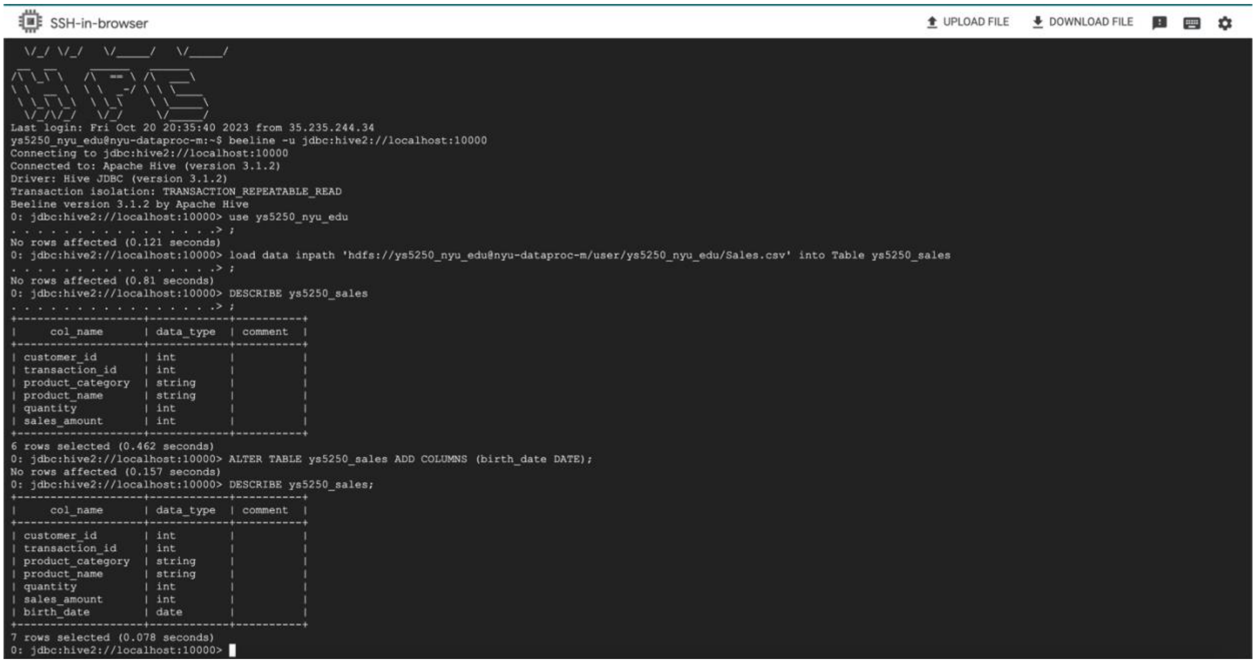


```
Linux nyu-dataproc-m 5.10.0-0.bpo.12-amd64 #1 SMP Debian 5.10.103-1~bpo10+1 (2022-03-08) x86_64
 __   __     __  __     __  __
/\ "-.\ \   /\ \_\ \   /\ \/\ \
\ \ \-.  \  \ \____ \  \ \ \_\ \
 \ \_\\"\_\  \/\_____\  \ \_____\
  \/_/ \/_/   \/_____/   \/_____/

 __  __     ______   ______
/\ \_\ \   /\  == \ /\  ___\
\ \  __ \  \ \  _-/ \ \ \____
 \ \_\ \_\  \ \_\    \ \_____\
  \/_/\/_/   \/_/     \/_____/
Last login: Fri Oct 20 20:12:03 2023 from 35.235.244.34
ys5250_nyu_edu@nyu-dataproc-m:~$ beeline -u jdbc:hive2://localhost:10000
Connecting to jdbc:hive2://localhost:10000
Connected to: Apache Hive (version 3.1.2)
Driver: Hive JDBC (version 3.1.2)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.2 by Apache Hive
0: jdbc:hive2://localhost:10000> use ys5250_nyu_edu;
No rows affected (0.12 seconds)
0: jdbc:hive2://localhost:10000> create table ys5250_sales (customer_id int,transaction_id int,product_category string,product_name string,quantity int,sales_amount int) ROW FORMAT DEL
IMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE;
No rows affected (0.514 seconds)
0: jdbc:hive2://localhost:10000> show tables;
+---------------+
|   tab_name    |
+---------------+
| ys5250_sales  |
+---------------+
1 row selected (0.138 seconds)
0: jdbc:hive2://localhost:10000>
```



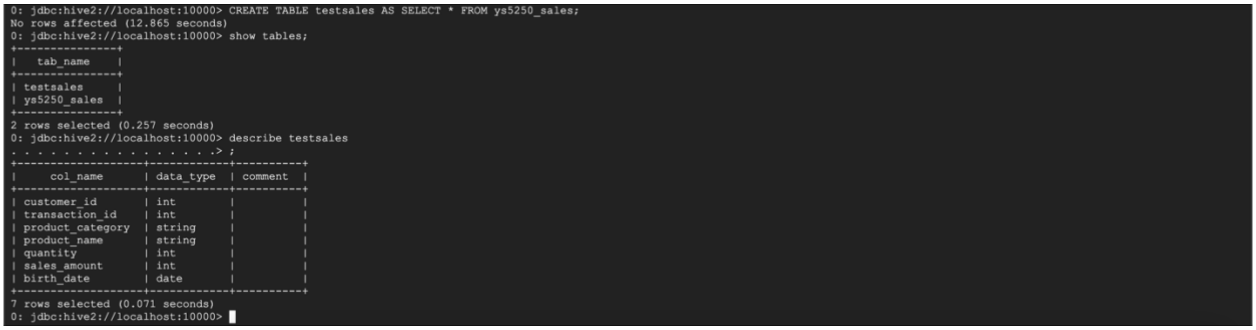
```
Linux nyu-dataproc-m 5.10.0-0.bpo.12-amd64 #1 SMP Debian 5.10.103-1~bpo10+1 (2022-03-08) x86_64
 __   __     __  __     __  __
/\ "-.\ \   /\ \_\ \   /\ \/\ \
\ \ \-.  \  \ \____ \  \ \ \_\ \
 \ \_\\"\_\  \/\_____\  \ \_____\
  \/_/ \/_/   \/_____/   \/_____/

 __  __     ______   ______
/\ \_\ \   /\  == \ /\  ___\
\ \  __ \  \ \  _-/ \ \ \____
 \ \_\ \_\  \ \_\    \ \_____\
  \/_/\/_/   \/_/     \/_____/
Last login: Fri Oct 20 20:35:40 2023 from 35.235.244.34
ys5250_nyu_edu@nyu-dataproc-m:~$ beeline -u jdbc:hive2://localhost:10000
Connecting to jdbc:hive2://localhost:10000
Connected to: Apache Hive (version 3.1.2)
Driver: Hive JDBC (version 3.1.2)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.2 by Apache Hive
0: jdbc:hive2://localhost:10000> use ys5250_nyu_edu
. . . . . . . . . . . . . . . .> ;
No rows affected (0.121 seconds)
0: jdbc:hive2://localhost:10000> load data inpath 'hdfs://ys5250_nyu_edu@nyu-dataproc-m/user/ys5250_nyu_edu/Sales.csv' into Table ys5250_sales
. . . . . . . . . . . . . . . .> ;
No rows affected (0.81 seconds)
0: jdbc:hive2://localhost:10000> DESCRIBE ys5250_sales
. . . . . . . . . . . . . . . .> ;
+-------------------+------------+----------+
|     col_name      | data_type  | comment  |
+-------------------+------------+----------+
| customer_id       | int        |          |
| transaction_id    | int        |          |
| product_category  | string     |          |
| product_name      | string     |          |
| quantity          | int        |          |
| sales_amount      | int        |          |
+-------------------+------------+----------+
```

Add a column birth date with appropriate datatype.



```
  \/_/ \/_/   \/_____/   \/_____/
 __  __   ______   ______
/\ \_\ \ /\  == \ /\  ___\
\ \  __ \\ \  _-/ \ \___  \
 \ \_\ \_\\ \_\    \ \_____\
  \/_/\/_/ \/_/     \/_____/
Last login: Fri Oct 20 20:35:40 2023 from 35.235.244.34
ys5250_nyu_edu@nyu-dataproc-m:~$ beeline -u jdbc:hive2://localhost:10000
Connecting to jdbc:hive2://localhost:10000
Connected to: Apache Hive (version 3.1.2)
Driver: Hive JDBC (version 3.1.2)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.2 by Apache Hive
0: jdbc:hive2://localhost:10000> use ys5250_nyu_edu
. . . . . . . . . . . . . . . .> ;
No rows affected (0.121 seconds)
0: jdbc:hive2://localhost:10000> load data inpath 'hdfs://ys5250_nyu_edu@nyu-dataproc-m/user/ys5250_nyu_edu/Sales.csv' into Table ys5250_sales
. . . . . . . . . . . . . . . .> ;
No rows affected (0.81 seconds)
0: jdbc:hive2://localhost:10000> DESCRIBE ys5250_sales
. . . . . . . . . . . . . . . .> ;
+-------------------+------------+----------+
|     col_name      | data_type  | comment  |
+-------------------+------------+----------+
| customer_id       | int        |          |
| transaction_id    | int        |          |
| product_category  | string     |          |
| product_name      | string     |          |
| quantity          | int        |          |
| sales_amount      | int        |          |
+-------------------+------------+----------+
6 rows selected (0.462 seconds)
0: jdbc:hive2://localhost:10000> ALTER TABLE ys5250_sales ADD COLUMNS (birth_date DATE);
No rows affected (0.157 seconds)
0: jdbc:hive2://localhost:10000> DESCRIBE ys5250_sales;
+-------------------+------------+----------+
|     col_name      | data_type  | comment  |
+-------------------+------------+----------+
| customer_id       | int        |          |
| transaction_id    | int        |          |
| product_category  | string     |          |
| product_name      | string     |          |
| quantity          | int        |          |
| sales_amount      | int        |          |
| birth_date        | date       |          |
+-------------------+------------+----------+
7 rows selected (0.078 seconds)
0: jdbc:hive2://localhost:10000>
```

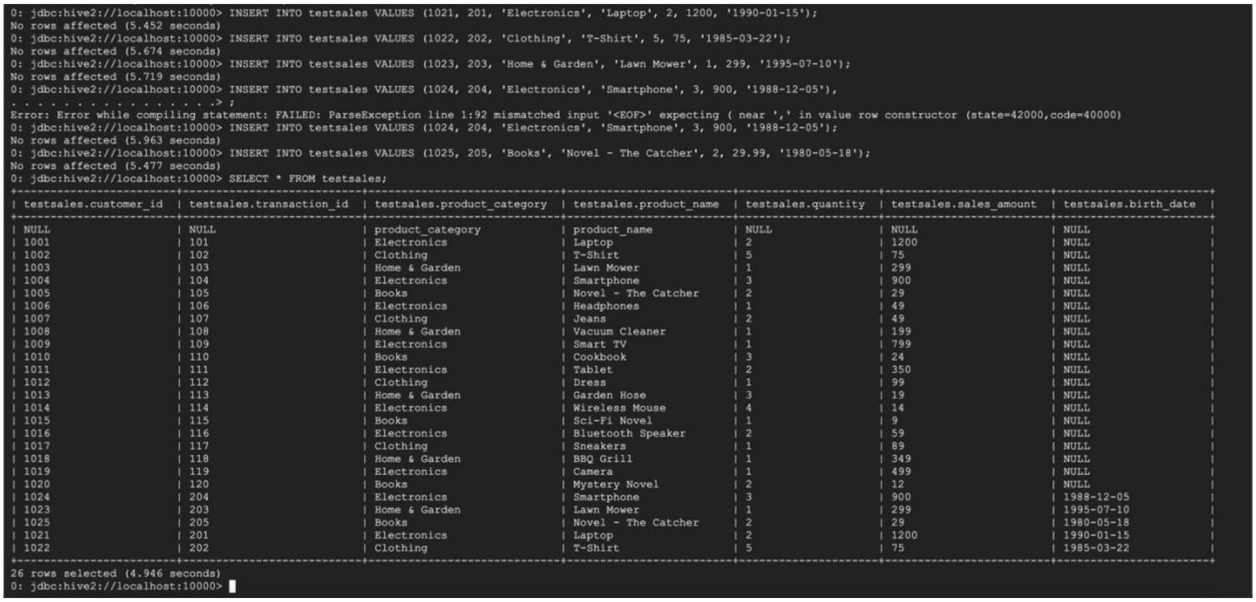
Create a test table, testsales by selecting all records from the sales table.



```
0: jdbc:hive2://localhost:10000> CREATE TABLE testsales AS SELECT * FROM ys5250_sales;
No rows affected (12.865 seconds)
0: jdbc:hive2://localhost:10000> show tables;
+---------------+
|   tab_name    |
+---------------+
| testsales     |
| ys5250_sales  |
+---------------+
2 rows selected (0.257 seconds)
0: jdbc:hive2://localhost:10000> describe testsales
. . . . . . . . . . . . . . . .> ;
+-------------------+------------+----------+
|     col_name      | data_type  | comment  |
+-------------------+------------+----------+
| customer_id       | int        |          |
| transaction_id    | int        |          |
| product_category  | string     |          |
| product_name      | string     |          |
| quantity          | int        |          |
| sales_amount      | int        |          |
| birth_date        | date       |          |
+-------------------+------------+----------+
7 rows selected (0.071 seconds)
0: jdbc:hive2://localhost:10000>
```

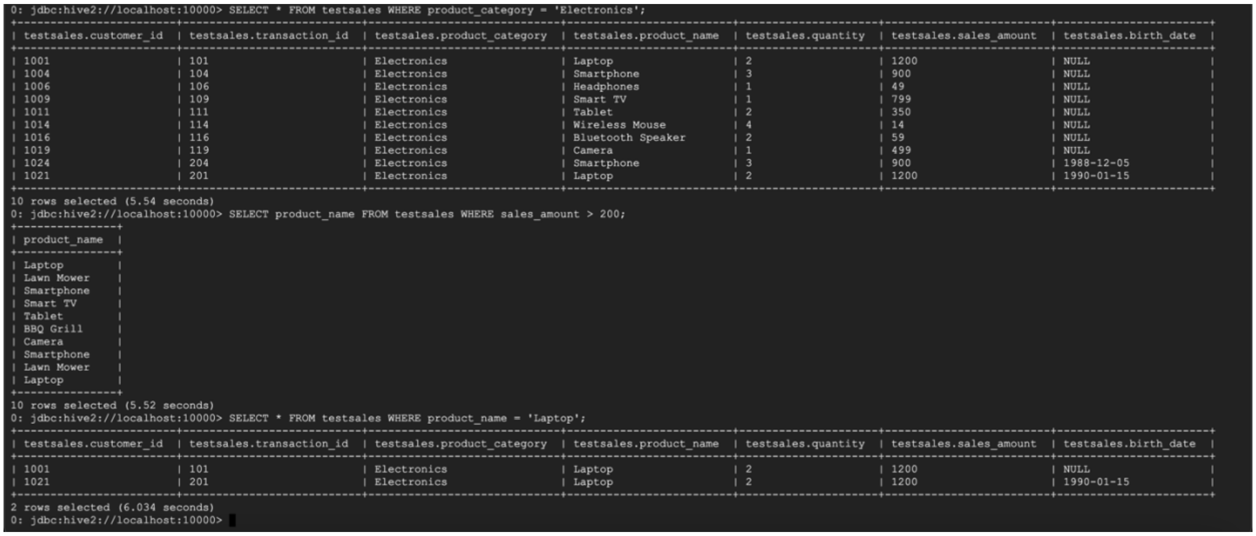
Insert 5 new records in test table.

Query all records from the test table!



```
0: jdbc:hive2://localhost:10000> INSERT INTO testsales VALUES (1021, 201, 'Electronics', 'Laptop', 2, 1200, '1990-01-15');
No rows affected (5.452 seconds)
0: jdbc:hive2://localhost:10000> INSERT INTO testsales VALUES (1022, 202, 'Clothing', 'T-Shirt', 5, 75, '1985-03-22');
No rows affected (5.674 seconds)
0: jdbc:hive2://localhost:10000> INSERT INTO testsales VALUES (1023, 203, 'Home & Garden', 'Lawn Mower', 1, 299, '1995-07-10');
No rows affected (5.719 seconds)
0: jdbc:hive2://localhost:10000> INSERT INTO testsales VALUES (1024, 204, 'Electronics', 'Smartphone', 3, 900, '1988-12-05'),
. . . . . . . . . . . . . . . .> ;
Error: Error while compiling statement: FAILED: ParseException line 1:92 mismatched input '<EOF>' expecting ( near ',' in value row constructor (state=42000,code=40000)
0: jdbc:hive2://localhost:10000> INSERT INTO testsales VALUES (1024, 204, 'Electronics', 'Smartphone', 3, 900, '1988-12-05');
No rows affected (5.963 seconds)
0: jdbc:hive2://localhost:10000> INSERT INTO testsales VALUES (1025, 205, 'Books', 'Novel - The Catcher', 2, 29.99, '1980-05-18');
No rows affected (5.477 seconds)
0: jdbc:hive2://localhost:10000> SELECT * FROM testsales;
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
| testsales.customer_id  | testsales.transaction_id  | testsales.product_category  | testsales.product_name  | testsales.quantity  | testsales.sales_amount  | testsales.birth_date  |
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
| NULL                   | NULL                      | product_category            | product_name            | NULL                | NULL                    | NULL                  |
| 1001                   | 101                       | Electronics                 | Laptop                  | 2                   | 1200                    | NULL                  |
| 1002                   | 102                       | Clothing                    | T-Shirt                 | 5                   | 75                      | NULL                  |
| 1003                   | 103                       | Home & Garden               | Lawn Mower              | 1                   | 299                     | NULL                  |
| 1004                   | 104                       | Electronics                 | Smartphone              | 3                   | 900                     | NULL                  |
| 1005                   | 105                       | Books                       | Novel - The Catcher     | 2                   | 29                      | NULL                  |
| 1006                   | 106                       | Electronics                 | Headphones              | 1                   | 49                      | NULL                  |
| 1007                   | 107                       | Clothing                    | Jeans                   | 2                   | 49                      | NULL                  |
| 1008                   | 108                       | Home & Garden               | Vacuum Cleaner          | 1                   | 199                     | NULL                  |
| 1009                   | 109                       | Electronics                 | Smart TV                | 1                   | 799                     | NULL                  |
| 1010                   | 110                       | Books                       | Cookbook                | 3                   | 24                      | NULL                  |
| 1011                   | 111                       | Electronics                 | Tablet                  | 2                   | 350                     | NULL                  |
| 1012                   | 112                       | Clothing                    | Dress                   | 1                   | 99                      | NULL                  |
| 1013                   | 113                       | Home & Garden               | Garden Hose             | 3                   | 19                      | NULL                  |
| 1014                   | 114                       | Electronics                 | Wireless Mouse          | 4                   | 14                      | NULL                  |
| 1015                   | 115                       | Books                       | Sci-Fi Novel            | 1                   | 9                       | NULL                  |
| 1016                   | 116                       | Electronics                 | Bluetooth Speaker       | 2                   | 59                      | NULL                  |
| 1017                   | 117                       | Clothing                    | Sneakers                | 1                   | 89                      | NULL                  |
| 1018                   | 118                       | Home & Garden               | BBQ Grill               | 1                   | 349                     | NULL                  |
| 1019                   | 119                       | Electronics                 | Camera                  | 1                   | 499                     | NULL                  |
| 1020                   | 120                       | Books                       | Mystery Novel           | 2                   | 12                      | NULL                  |
| 1024                   | 204                       | Electronics                 | Smartphone              | 3                   | 900                     | 1988-12-05            |
| 1023                   | 203                       | Home & Garden               | Lawn Mower              | 1                   | 299                     | 1995-07-10            |
| 1025                   | 205                       | Books                       | Novel - The Catcher     | 2                   | 29                      | 1980-05-18            |
| 1021                   | 201                       | Electronics                 | Laptop                  | 2                   | 1200                    | 1990-01-15            |
| 1022                   | 202                       | Clothing                    | T-Shirt                 | 5                   | 75                      | 1985-03-22            |
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
26 rows selected (4.946 seconds)
0: jdbc:hive2://localhost:10000>
```

Write three queries with fiters (where clause) and show result of queries.

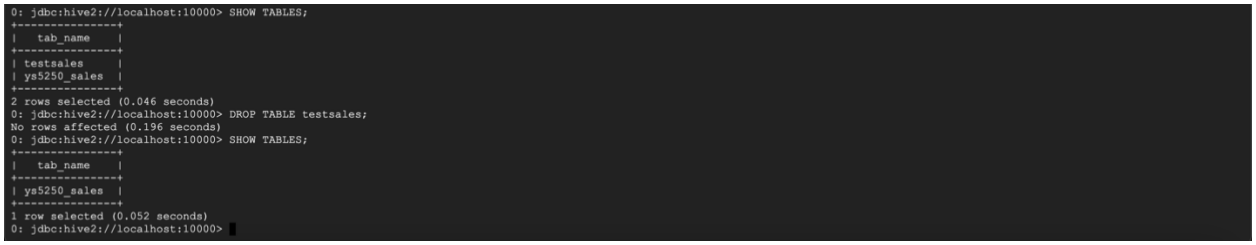


```
0: jdbc:hive2://localhost:10000> SELECT * FROM testsales WHERE product_category = 'Electronics';
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
| testsales.customer_id  | testsales.transaction_id  | testsales.product_category  | testsales.product_name  | testsales.quantity  | testsales.sales_amount  | testsales.birth_date  |
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
| 1001                   | 101                       | Electronics                 | Laptop                  | 2                   | 1200                    | NULL                  |
| 1004                   | 104                       | Electronics                 | Smartphone              | 3                   | 900                     | NULL                  |
| 1006                   | 106                       | Electronics                 | Headphones              | 1                   | 49                      | NULL                  |
| 1009                   | 109                       | Electronics                 | Smart TV                | 1                   | 799                     | NULL                  |
| 1011                   | 111                       | Electronics                 | Tablet                  | 2                   | 350                     | NULL                  |
| 1014                   | 114                       | Electronics                 | Wireless Mouse          | 4                   | 14                      | NULL                  |
| 1016                   | 116                       | Electronics                 | Bluetooth Speaker       | 2                   | 59                      | NULL                  |
| 1019                   | 119                       | Electronics                 | Camera                  | 1                   | 499                     | NULL                  |
| 1024                   | 204                       | Electronics                 | Smartphone              | 3                   | 900                     | 1988-12-05            |
| 1021                   | 201                       | Electronics                 | Laptop                  | 2                   | 1200                    | 1990-01-15            |
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
10 rows selected (5.54 seconds)
0: jdbc:hive2://localhost:10000> SELECT product_name FROM testsales WHERE sales_amount > 200;
+---------------+
| product_name  |
+---------------+
| Laptop        |
| Lawn Mower    |
| Smartphone    |
| Smart TV      |
| Tablet        |
| BBQ Grill     |
| Camera        |
| Smartphone    |
| Lawn Mower    |
| Laptop        |
+---------------+
10 rows selected (5.52 seconds)
0: jdbc:hive2://localhost:10000> SELECT * FROM testsales WHERE product_name = 'Laptop';
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
| testsales.customer_id  | testsales.transaction_id  | testsales.product_category  | testsales.product_name  | testsales.quantity  | testsales.sales_amount  | testsales.birth_date  |
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
| 1001                   | 101                       | Electronics                 | Laptop                  | 2                   | 1200                    | NULL                  |
| 1021                   | 201                       | Electronics                 | Laptop                  | 2                   | 1200                    | 1990-01-15            |
+------------------------+---------------------------+-----------------------------+-------------------------+---------------------+-------------------------+-----------------------+
2 rows selected (6.034 seconds)
0: jdbc:hive2://localhost:10000>
```

Show the list of tables.

Drop the test table.

Show the list of tables after dropping test table.

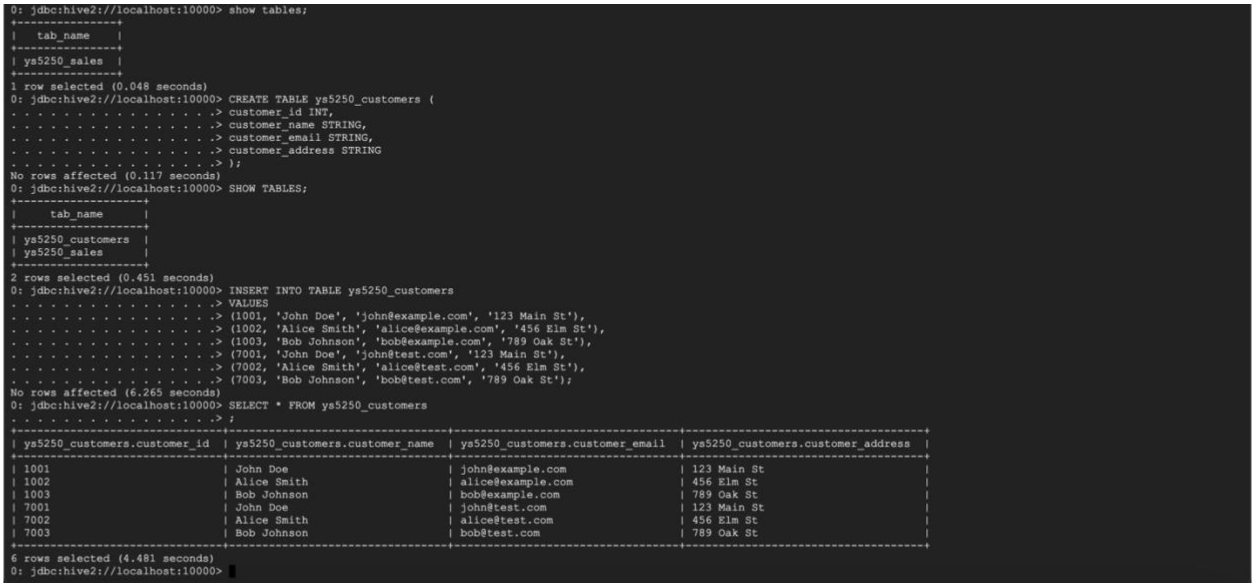


```
0: jdbc:hive2://localhost:10000> SHOW TABLES;
+---------------+
|   tab_name    |
+---------------+
| testsales     |
| ys5250_sales  |
+---------------+
2 rows selected (0.046 seconds)
0: jdbc:hive2://localhost:10000> DROP TABLE testsales;
No rows affected (0.196 seconds)
0: jdbc:hive2://localhost:10000> SHOW TABLES;
+---------------+
|   tab_name    |
+---------------+
| ys5250_sales  |
+---------------+
1 row selected (0.052 seconds)
0: jdbc:hive2://localhost:10000> 
```

B) Use following code to create a Hive table, customers with name, Net-Id_customers (e.g. asp13_customers)

```
CREATE TABLE asp_customers (
customer_id INT,
customer_name STRING,
customer_email STRING,
customer_address STRING
);
INSERT INTO TABLE asp13_customers
VALUES
(1001, 'John Doe', 'john@example.com', '123 Main St'),
(1002, 'Alice Smith', 'alice@example.com', '456 Elm St'),
(1003, 'Bob Johnson', 'bob@example.com', '789 Oak St');

INSERT INTO TABLE customers
VALUES
(7001, 'John Doe', 'john@test.com', '123 Main St'),
(7002, 'Alice Smith', 'alice@test.com', '456 Elm St'),
(7003, 'Bob Johnson', 'bob@test.com', '789 Oak St');
```

Using Sales and Customers tables, write quires with INNER JOIN, LEFT OUTER JOIN, RIGHT OUTER JOIN, and FULL OUTER JOIN. Submit SQL queries and screenshot of their results.

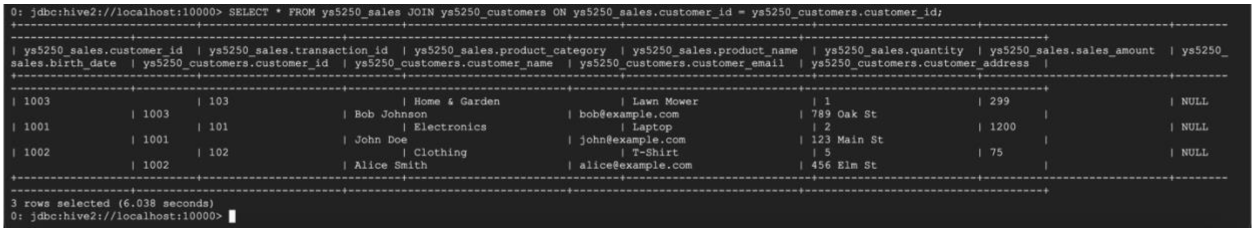


```
0: jdbc:hive2://localhost:10000> show tables;
+---------------+
|   tab_name    |
+---------------+
| ys5250_sales  |
+---------------+
1 row selected (0.048 seconds)
0: jdbc:hive2://localhost:10000> CREATE TABLE ys5250_customers (
. . . . . . . . . . . . . . . .> customer_id INT,
. . . . . . . . . . . . . . . .> customer_name STRING,
. . . . . . . . . . . . . . . .> customer_email STRING,
. . . . . . . . . . . . . . . .> customer_address STRING
. . . . . . . . . . . . . . . .> );
No rows affected (0.117 seconds)
0: jdbc:hive2://localhost:10000> SHOW TABLES;
+-------------------+
|     tab_name      |
+-------------------+
| ys5250_customers  |
| ys5250_sales      |
+-------------------+
2 rows selected (0.451 seconds)
0: jdbc:hive2://localhost:10000> INSERT INTO TABLE ys5250_customers
. . . . . . . . . . . . . . . .> VALUES
. . . . . . . . . . . . . . . .> (1001, 'John Doe', 'john@example.com', '123 Main St'),
. . . . . . . . . . . . . . . .> (1002, 'Alice Smith', 'alice@example.com', '456 Elm St'),
. . . . . . . . . . . . . . . .> (1003, 'Bob Johnson', 'bob@example.com', '789 Oak St'),
. . . . . . . . . . . . . . . .> (7001, 'John Doe', 'john@test.com', '123 Main St'),
. . . . . . . . . . . . . . . .> (7002, 'Alice Smith', 'alice@test.com', '456 Elm St'),
. . . . . . . . . . . . . . . .> (7003, 'Bob Johnson', 'bob@test.com', '789 Oak St');
No rows affected (6.265 seconds)
0: jdbc:hive2://localhost:10000> SELECT * FROM ys5250_customers
. . . . . . . . . . . . . . . .> ;
+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| ys5250_customers.customer_id  | ys5250_customers.customer_name  | ys5250_customers.customer_email  | ys5250_customers.customer_address  |
+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| 1001                          | John Doe                        | john@example.com                 | 123 Main St                        |
| 1002                          | Alice Smith                     | alice@example.com                | 456 Elm St                         |
| 1003                          | Bob Johnson                     | bob@example.com                  | 789 Oak St                         |
| 7001                          | John Doe                        | john@test.com                    | 123 Main St                        |
| 7002                          | Alice Smith                     | alice@test.com                   | 456 Elm St                         |
| 7003                          | Bob Johnson                     | bob@test.com                     | 789 Oak St                         |
+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
6 rows selected (4.481 seconds)
0: jdbc:hive2://localhost:10000>
```

SQL Queries and their results:

1. Inner Join

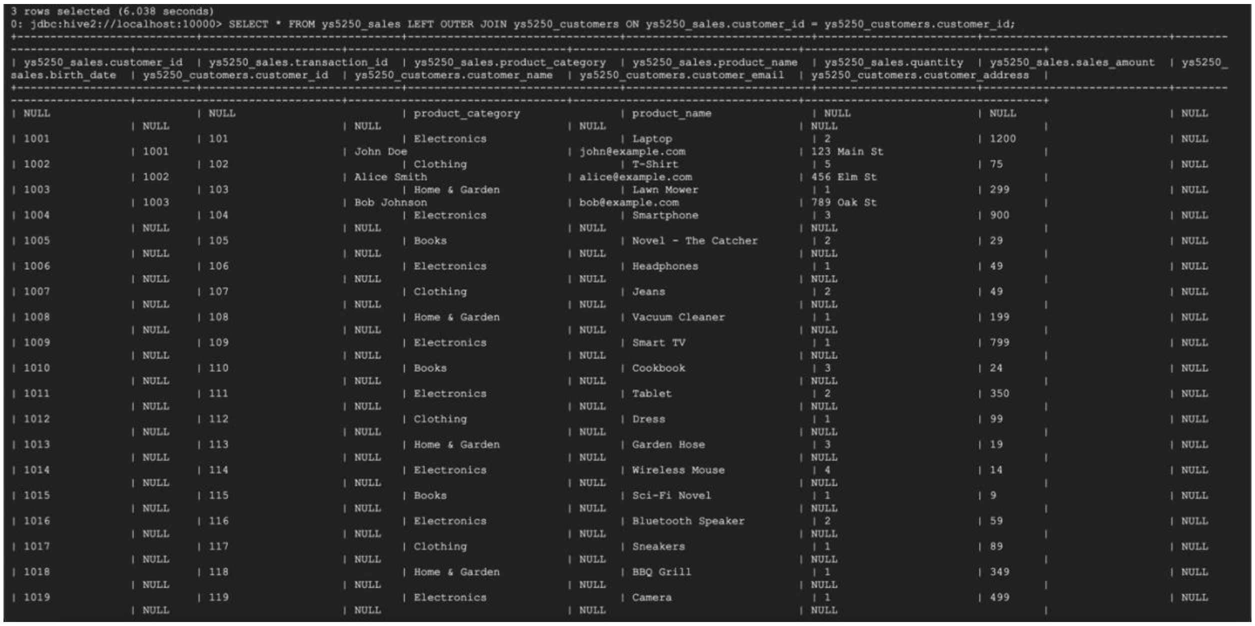
SELECT * FROM ys5250_sales JOIN ys5250_customers ON ys5250_sales.customer_id = ys5250_customers.customer_id;



```
0: jdbc:hive2://localhost:10000> SELECT * FROM ys5250_sales JOIN ys5250_customers ON ys5250_sales.customer_id = ys5250_customers.customer_id;
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| ys5250_sales.customer_id  | ys5250_sales.transaction_id  | ys5250_sales.product_category  | ys5250_sales.product_name  | ys5250_sales.quantity  | ys5250_sales.sales_amount  | ys5250_
sales.birth_date  | ys5250_customers.customer_id  | ys5250_customers.customer_name  | ys5250_customers.customer_email  | ys5250_customers.customer_address  |
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| 1003                      | 103                          | Home & Garden                  | Lawn Mower                 | 1                      | 299                        | NULL
                  | 1003                          | Bob Johnson                     | bob@example.com                  | 789 Oak St                         |
| 1001                      | 101                          | Electronics                    | Laptop                     | 2                      | 1200                       | NULL
                  | 1001                          | John Doe                        | john@example.com                 | 123 Main St                        |
| 1002                      | 102                          | Clothing                       | T-Shirt                    | 5                      | 75                         | NULL
                  | 1002                          | Alice Smith                     | alice@example.com                | 456 Elm St                         |
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
3 rows selected (6.038 seconds)
0: jdbc:hive2://localhost:10000>
```

2. Left Outer Join

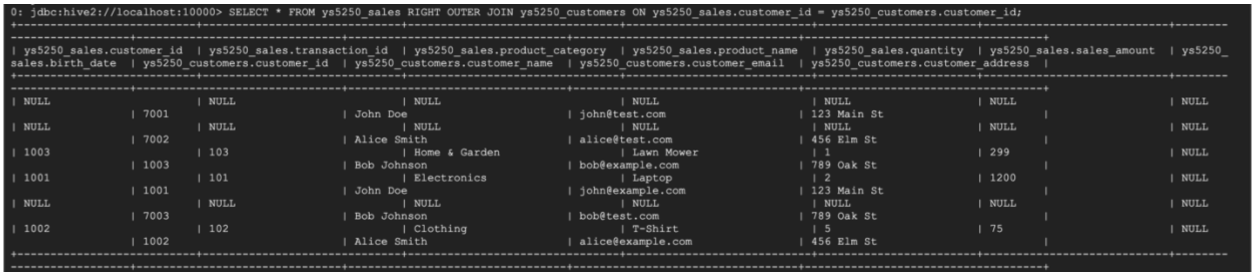
SELECT * FROM ys5250_sales LEFT OUTER JOIN ys5250_customers ON ys5250_sales.customer_id = ys5250_customers.customer_id;



```
3 rows selected (6.038 seconds)
0: jdbc:hive2://localhost:10000> SELECT * FROM ys5250_sales LEFT OUTER JOIN ys5250_customers ON ys5250_sales.customer_id = ys5250_customers.customer_id;
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| ys5250_sales.customer_id  | ys5250_sales.transaction_id  | ys5250_sales.product_category  | ys5250_sales.product_name  | ys5250_sales.quantity  | ys5250_sales.sales_amount  | ys5250_
sales.birth_date  | ys5250_customers.customer_id  | ys5250_customers.customer_name  | ys5250_customers.customer_email  | ys5250_customers.customer_address  |
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| NULL                      | NULL                         | product_category               | product_name               | NULL                   | NULL                       | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1001                      | 101                          | Electronics                    | Laptop                     | 2                      | 1200                       | NULL
                  | 1001                          | John Doe                        | john@example.com                 | 123 Main St                        |
| 1002                      | 102                          | Clothing                       | T-Shirt                    | 5                      | 75                         | NULL
                  | 1002                          | Alice Smith                     | alice@example.com                | 456 Elm St                         |
| 1003                      | 103                          | Home & Garden                  | Lawn Mower                 | 1                      | 299                        | NULL
                  | 1003                          | Bob Johnson                     | bob@example.com                  | 789 Oak St                         |
| 1004                      | 104                          | Electronics                    | Smartphone                 | 3                      | 900                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1005                      | 105                          | Books                          | Novel - The Catcher        | 2                      | 29                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1006                      | 106                          | Electronics                    | Headphones                 | 1                      | 49                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1007                      | 107                          | Clothing                       | Jeans                      | 2                      | 49                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1008                      | 108                          | Home & Garden                  | Vacuum Cleaner             | 1                      | 199                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1009                      | 109                          | Electronics                    | Smart TV                   | 1                      | 799                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1010                      | 110                          | Books                          | Cookbook                   | 3                      | 24                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1011                      | 111                          | Electronics                    | Tablet                     | 2                      | 350                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1012                      | 112                          | Clothing                       | Dress                      | 1                      | 99                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1013                      | 113                          | Home & Garden                  | Garden Hose                | 3                      | 19                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1014                      | 114                          | Electronics                    | Wireless Mouse             | 4                      | 14                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1015                      | 115                          | Books                          | Sci-Fi Novel               | 1                      | 9                          | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1016                      | 116                          | Electronics                    | Bluetooth Speaker          | 2                      | 59                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1017                      | 117                          | Clothing                       | Sneakers                   | 1                      | 89                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1018                      | 118                          | Home & Garden                  | BBQ Grill                  | 1                      | 349                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1019                      | 119                          | Electronics                    | Camera                     | 1                      | 499                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
```

3. Right Outer Join

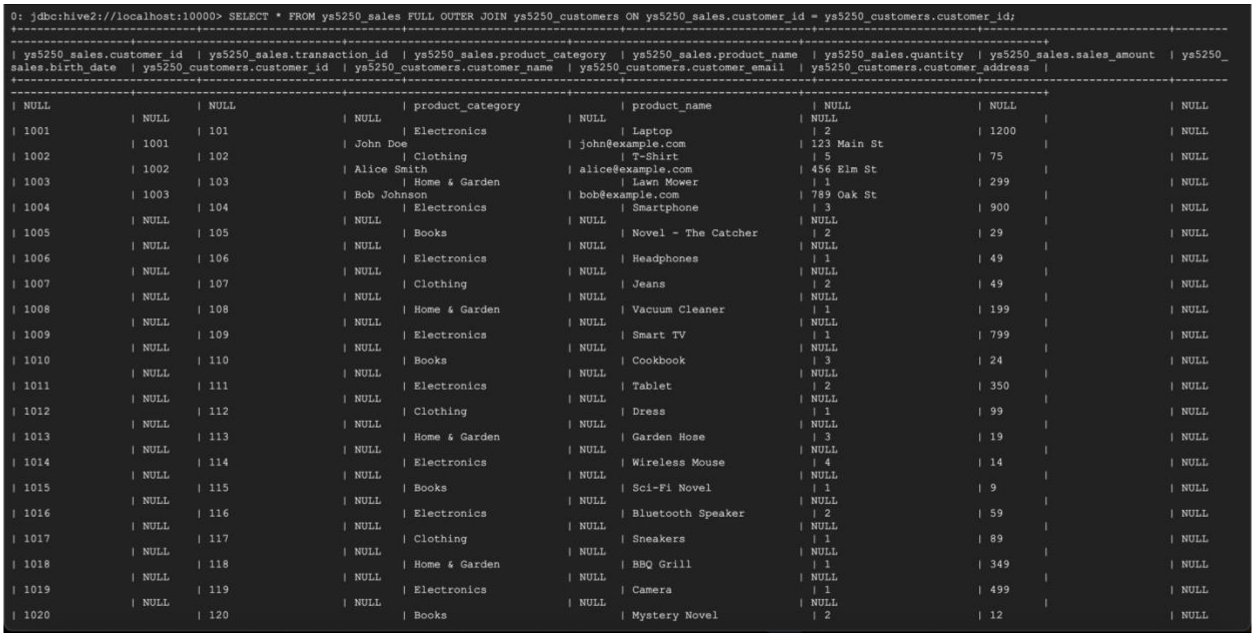
SELECT * FROM ys5250_sales RIGHT OUTER JOIN ys5250_customers ON ys5250_sales.customer_id = ys5250_customers.customer_id;



```
0: jdbc:hive2://localhost:10000> SELECT * FROM ys5250_sales RIGHT OUTER JOIN ys5250_customers ON ys5250_sales.customer_id = ys5250_customers.customer_id;
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| ys5250_sales.customer_id  | ys5250_sales.transaction_id  | ys5250_sales.product_category  | ys5250_sales.product_name  | ys5250_sales.quantity  | ys5250_sales.sales_amount  | ys5250_
sales.birth_date  | ys5250_customers.customer_id  | ys5250_customers.customer_name  | ys5250_customers.customer_email  | ys5250_customers.customer_address  |
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| NULL                      | NULL                         | NULL                           | NULL                       | NULL                   | NULL                       | NULL
                  | 7001                          | John Doe                        | john@test.com                    | 123 Main St                        |
| NULL                      | NULL                         | NULL                           | NULL                       | NULL                   | NULL                       | NULL
                  | 7002                          | Alice Smith                     | alice@test.com                   | 456 Elm St                         |
| 1003                      | 103                          | Home & Garden                  | Lawn Mower                 | 1                      | 299                        | NULL
                  | 1003                          | Bob Johnson                     | bob@example.com                  | 789 Oak St                         |
| 1001                      | 101                          | Electronics                    | Laptop                     | 2                      | 1200                       | NULL
                  | 1001                          | John Doe                        | john@example.com                 | 123 Main St                        |
| NULL                      | NULL                         | NULL                           | NULL                       | NULL                   | NULL                       | NULL
                  | 7003                          | Bob Johnson                     | bob@test.com                     | 789 Oak St                         |
| 1002                      | 102                          | Clothing                       | T-Shirt                    | 5                      | 75                         | NULL
                  | 1002                          | Alice Smith                     | alice@example.com                | 456 Elm St                         |
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
```

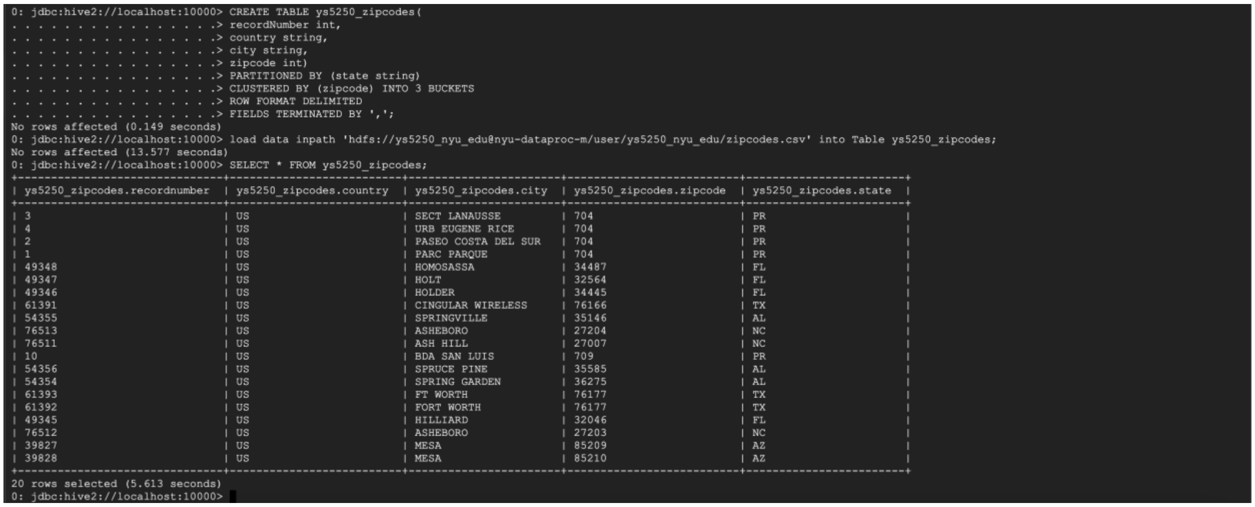
4. Full Outer Join

SELECT * FROM ys5250_sales FULL OUTER JOIN ys5250_customers ON ys5250_sales.customer_id = ys5250_customers.customer_id;



```
0: jdbc:hive2://localhost:10000> SELECT * FROM ys5250_sales FULL OUTER JOIN ys5250_customers ON ys5250_sales.customer_id = ys5250_customers.customer_id;
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| ys5250_sales.customer_id  | ys5250_sales.transaction_id  | ys5250_sales.product_category  | ys5250_sales.product_name  | ys5250_sales.quantity  | ys5250_sales.sales_amount  | ys5250_
sales.birth_date  | ys5250_customers.customer_id  | ys5250_customers.customer_name  | ys5250_customers.customer_email  | ys5250_customers.customer_address  |
+---------------------------+------------------------------+--------------------------------+----------------------------+------------------------+----------------------------+--------
------------------+-------------------------------+---------------------------------+----------------------------------+------------------------------------+
| NULL                      | NULL                         | product_category               | product_name               | NULL                   | NULL                       | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1001                      | 101                          | Electronics                    | Laptop                     | 2                      | 1200                       | NULL
                  | 1001                          | John Doe                        | john@example.com                 | 123 Main St                        |
| 1002                      | 102                          | Clothing                       | T-Shirt                    | 5                      | 75                         | NULL
                  | 1002                          | Alice Smith                     | alice@example.com                | 456 Elm St                         |
| 1003                      | 103                          | Home & Garden                  | Lawn Mower                 | 1                      | 299                        | NULL
                  | 1003                          | Bob Johnson                     | bob@example.com                  | 789 Oak St                         |
| 1004                      | 104                          | Electronics                    | Smartphone                 | 3                      | 900                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1005                      | 105                          | Books                          | Novel - The Catcher        | 2                      | 29                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1006                      | 106                          | Electronics                    | Headphones                 | 1                      | 49                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1007                      | 107                          | Clothing                       | Jeans                      | 2                      | 49                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1008                      | 108                          | Home & Garden                  | Vacuum Cleaner             | 1                      | 199                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1009                      | 109                          | Electronics                    | Smart TV                   | 1                      | 799                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1010                      | 110                          | Books                          | Cookbook                   | 3                      | 24                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1011                      | 111                          | Electronics                    | Tablet                     | 2                      | 350                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1012                      | 112                          | Clothing                       | Dress                      | 1                      | 99                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1013                      | 113                          | Home & Garden                  | Garden Hose                | 3                      | 19                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1014                      | 114                          | Electronics                    | Wireless Mouse             | 4                      | 14                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1015                      | 115                          | Books                          | Sci-Fi Novel               | 1                      | 9                          | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1016                      | 116                          | Electronics                    | Bluetooth Speaker          | 2                      | 59                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1017                      | 117                          | Clothing                       | Sneakers                   | 1                      | 89                         | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1018                      | 118                          | Home & Garden                  | BBQ Grill                  | 1                      | 349                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1019                      | 119                          | Electronics                    | Camera                     | 1                      | 499                        | NULL
                  | NULL                          | NULL                            | NULL                             | NULL                               |
| 1020                      | 120                          | Books                          | Mystery Novel              | 2                      | 12                         | NULL
```

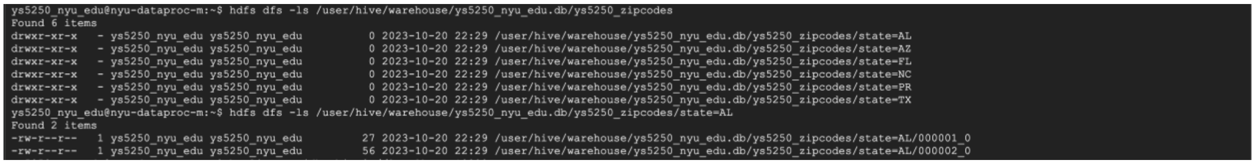
C) Using Zipcodes.csv file, create Hive table Net-ID_zipcodes (e.g. asp13_zipcodes). This table should have partitions by state and with 3 buckets by zipcode.



```
0: jdbc:hive2://localhost:10000> CREATE TABLE ys5250_zipcodes(
. . . . . . . . . . . . . . . .> recordNumber int,
. . . . . . . . . . . . . . . .> country string,
. . . . . . . . . . . . . . . .> city string,
. . . . . . . . . . . . . . . .> zipcode int)
. . . . . . . . . . . . . . . .> PARTITIONED BY (state string)
. . . . . . . . . . . . . . . .> CLUSTERED BY (zipcode) INTO 3 BUCKETS
. . . . . . . . . . . . . . . .> ROW FORMAT DELIMITED
. . . . . . . . . . . . . . . .> FIELDS TERMINATED BY ',';
No rows affected (0.149 seconds)
0: jdbc:hive2://localhost:10000> load data inpath 'hdfs://ys5250_nyu_edu@nyu-dataproc-m/user/ys5250_nyu_edu/zipcodes.csv' into Table ys5250_zipcodes;
No rows affected (13.577 seconds)
0: jdbc:hive2://localhost:10000> SELECT * FROM ys5250_zipcodes;
+-------------------------------+---------------------------+------------------------+----------------------------+-------------------------+
| ys5250_zipcodes.recordnumber  | ys5250_zipcodes.country   | ys5250_zipcodes.city   | ys5250_zipcodes.zipcode    | ys5250_zipcodes.state   |
+-------------------------------+---------------------------+------------------------+----------------------------+-------------------------+
| 3                             | US                        | SECT LANAUSSE          | 704                        | PR                      |
| 4                             | US                        | URB EUGENE RICE        | 704                        | PR                      |
| 2                             | US                        | PASEO COSTA DEL SUR    | 704                        | PR                      |
| 1                             | US                        | PARC PARQUE            | 704                        | PR                      |
| 49348                         | US                        | HOMOSASSA              | 34487                      | FL                      |
| 49347                         | US                        | HOLT                   | 32564                      | FL                      |
| 49346                         | US                        | HOLDER                 | 34445                      | FL                      |
| 61391                         | US                        | CINGULAR WIRELESS      | 76166                      | TX                      |
| 54355                         | US                        | SPRINGVILLE            | 35146                      | AL                      |
| 76513                         | US                        | ASHEBORO               | 27204                      | NC                      |
| 76511                         | US                        | ASH HILL               | 27007                      | NC                      |
| 10                            | US                        | BDA SAN LUIS           | 709                        | PR                      |
| 54356                         | US                        | SPRUCE PINE            | 35585                      | AL                      |
| 54354                         | US                        | SPRING GARDEN          | 36275                      | AL                      |
| 61393                         | US                        | FT WORTH               | 76177                      | TX                      |
| 61392                         | US                        | FORT WORTH             | 76177                      | TX                      |
| 49345                         | US                        | HILLIARD               | 32046                      | FL                      |
| 76512                         | US                        | ASHEBORO               | 27203                      | NC                      |
| 39827                         | US                        | MESA                   | 85209                      | AZ                      |
| 39828                         | US                        | MESA                   | 85210                      | AZ                      |
+-------------------------------+---------------------------+------------------------+----------------------------+-------------------------+
20 rows selected (5.613 seconds)
0: jdbc:hive2://localhost:10000>
```

Provide screenshot of

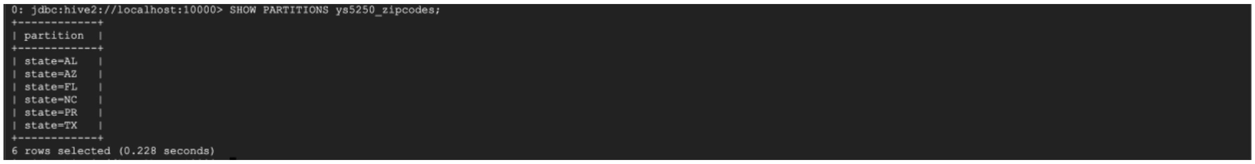
1. hdfs direcotry and subdirectories of patitions, also show files under partition state='AL'



```
ys5250_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -ls /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes
Found 6 items
drwxr-xr-x   - ys5250_nyu_edu ys5250_nyu_edu          0 2023-10-20 22:29 /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=AL
drwxr-xr-x   - ys5250_nyu_edu ys5250_nyu_edu          0 2023-10-20 22:29 /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=AZ
drwxr-xr-x   - ys5250_nyu_edu ys5250_nyu_edu          0 2023-10-20 22:29 /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=FL
drwxr-xr-x   - ys5250_nyu_edu ys5250_nyu_edu          0 2023-10-20 22:29 /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=NC
drwxr-xr-x   - ys5250_nyu_edu ys5250_nyu_edu          0 2023-10-20 22:29 /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=PR
drwxr-xr-x   - ys5250_nyu_edu ys5250_nyu_edu          0 2023-10-20 22:29 /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=TX
ys5250_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -ls /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=AL
Found 2 items
-rw-r--r--   1 ys5250_nyu_edu ys5250_nyu_edu         27 2023-10-20 22:29 /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=AL/000001_0
-rw-r--r--   1 ys5250_nyu_edu ys5250_nyu_edu         56 2023-10-20 22:29 /user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=AL/000002_0
```

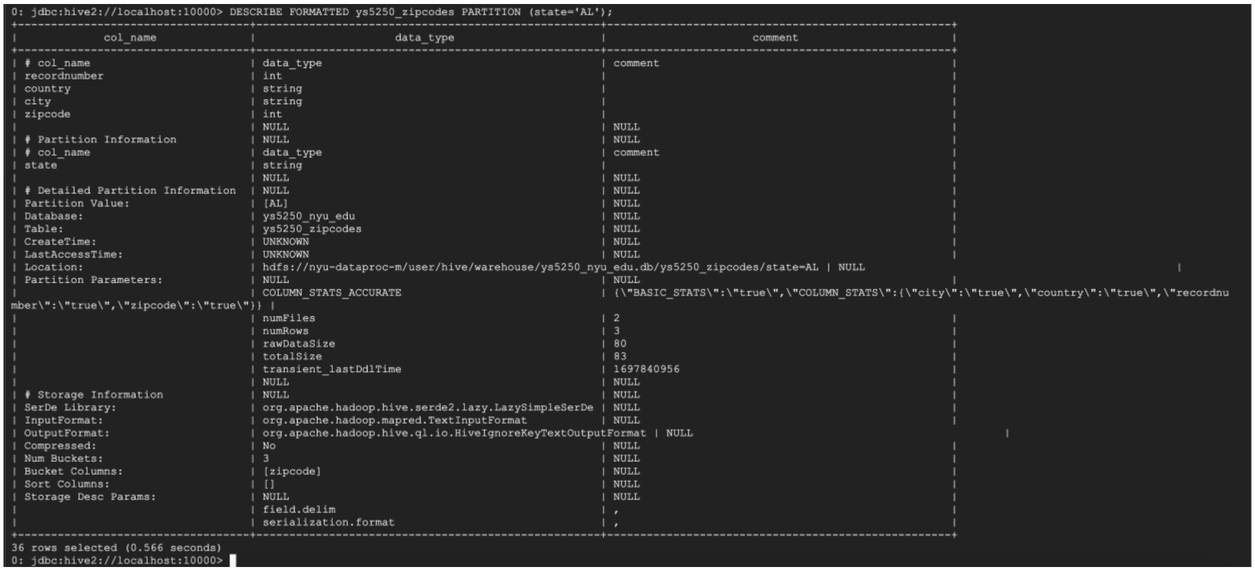
2. Results of following commands:

   SHOW PARTITIONS ys5250_zipcodes;



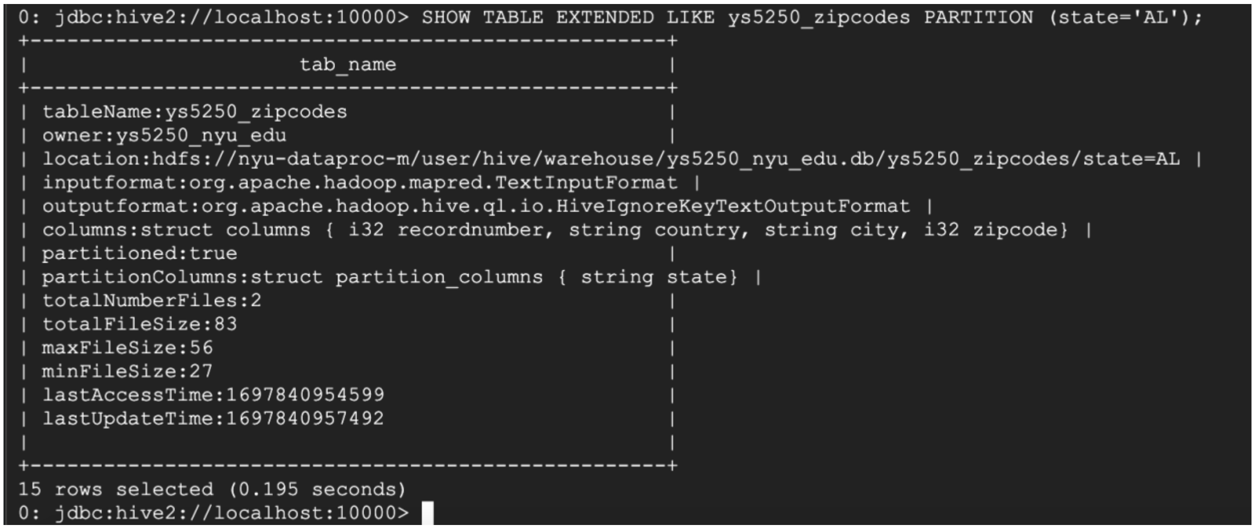
```
0: jdbc:hive2://localhost:10000> SHOW PARTITIONS ys5250_zipcodes;
+------------+
| partition  |
+------------+
| state=AL   |
| state=AZ   |
| state=FL   |
| state=NC   |
| state=PR   |
| state=TX   |
+------------+
6 rows selected (0.228 seconds)
```

DESCRIBE FORMATTED ys5250_zipcodes PARTITION (state='AL');



```
0: jdbc:hive2://localhost:10000> DESCRIBE FORMATTED ys5250_zipcodes PARTITION (state='AL');
+----------------------------------+-----------------------------------------------------+---------------------------------------------------+
|             col_name             |                      data_type                      |                      comment                      |
+----------------------------------+-----------------------------------------------------+---------------------------------------------------+
| # col_name                       | data_type                                           | comment                                           |
| recordnumber                     | int                                                 |                                                   |
| country                          | string                                              |                                                   |
| city                             | string                                              |                                                   |
| zipcode                          | int                                                 |                                                   |
|                                  | NULL                                                | NULL                                              |
| # Partition Information          | NULL                                                | NULL                                              |
| # col_name                       | data_type                                           | comment                                           |
| state                            | string                                              |                                                   |
|                                  | NULL                                                | NULL                                              |
| # Detailed Partition Information | NULL                                                | NULL                                              |
| Partition Value:                 | [AL]                                                | NULL                                              |
| Database:                        | ys5250_nyu_edu                                      | NULL                                              |
| Table:                           | ys5250_zipcodes                                     | NULL                                              |
| CreateTime:                      | UNKNOWN                                             | NULL                                              |
| LastAccessTime:                  | UNKNOWN                                             | NULL                                              |
| Location:                        | hdfs://nyu-dataproc-m/user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=AL | NULL                                              |
| Partition Parameters:            | NULL                                                | NULL                                              |
|                                  | COLUMN_STATS_ACCURATE                               | {\"BASIC_STATS\":\"true\",\"COLUMN_STATS\":{\"city\":\"true\",\"country\":\"true\",\"recordnu
mber\":\"true\",\"zipcode\":\"true\"}} |
|                                  | numFiles                                            | 2                                                 |
|                                  | numRows                                             | 3                                                 |
|                                  | rawDataSize                                         | 80                                                |
|                                  | totalSize                                           | 83                                                |
|                                  | transient_lastDdlTime                               | 1697840956                                        |
|                                  | NULL                                                | NULL                                              |
| # Storage Information            | NULL                                                | NULL                                              |
| SerDe Library:                   | org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe  | NULL                                              |
| InputFormat:                     | org.apache.hadoop.mapred.TextInputFormat            | NULL                                              |
| OutputFormat:                    | org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat | NULL                                              |
| Compressed:                      | No                                                  | NULL                                              |
| Num Buckets:                     | 3                                                   | NULL                                              |
| Bucket Columns:                  | [zipcode]                                           | NULL                                              |
| Sort Columns:                    | []                                                  | NULL                                              |
| Storage Desc Params:             | NULL                                                | NULL                                              |
|                                  | field.delim                                         | ,                                                 |
|                                  | serialization.format                                | ,                                                 |
+----------------------------------+-----------------------------------------------------+---------------------------------------------------+
36 rows selected (0.566 seconds)
0: jdbc:hive2://localhost:10000>
```

SHOW TABLE EXTENDED LIKE ys5250_zipcodes PARTITION (state='AL');



```
0: jdbc:hive2://localhost:10000> SHOW TABLE EXTENDED LIKE ys5250_zipcodes PARTITION (state='AL');
+---------------------------------------------------+
|                     tab_name                      |
+---------------------------------------------------+
| tableName:ys5250_zipcodes                         |
| owner:ys5250_nyu_edu                              |
| location:hdfs://nyu-dataproc-m/user/hive/warehouse/ys5250_nyu_edu.db/ys5250_zipcodes/state=AL |
| inputformat:org.apache.hadoop.mapred.TextInputFormat |
| outputformat:org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat |
| columns:struct columns { i32 recordnumber, string country, string city, i32 zipcode} |
| partitioned:true                                  |
| partitionColumns:struct partition_columns { string state} |
| totalNumberFiles:2                                |
| totalFileSize:83                                  |
| maxFileSize:56                                    |
| minFileSize:27                                    |
| lastAccessTime:1697840954599                      |
| lastUpdateTime:1697840957492                      |
|                                                   |
+---------------------------------------------------+
15 rows selected (0.195 seconds)
0: jdbc:hive2://localhost:10000>
```