STARTING THE LAB & CLONNING THE CLUSTER:
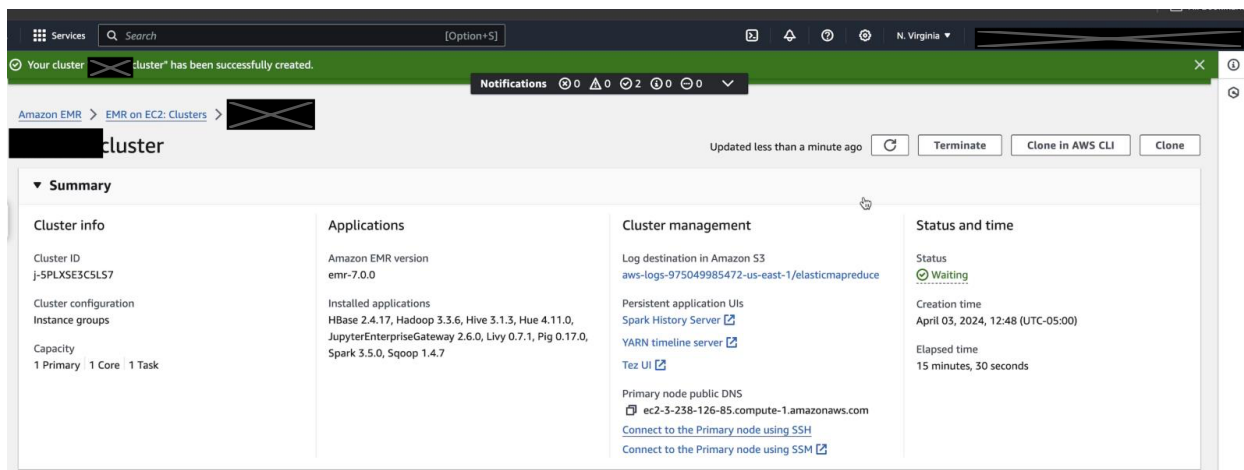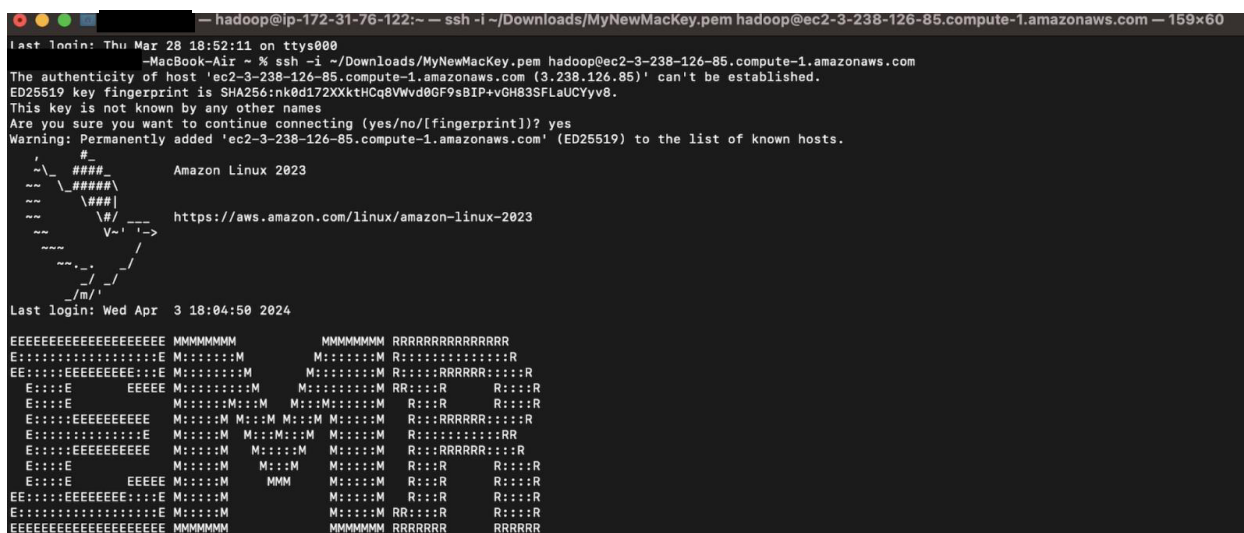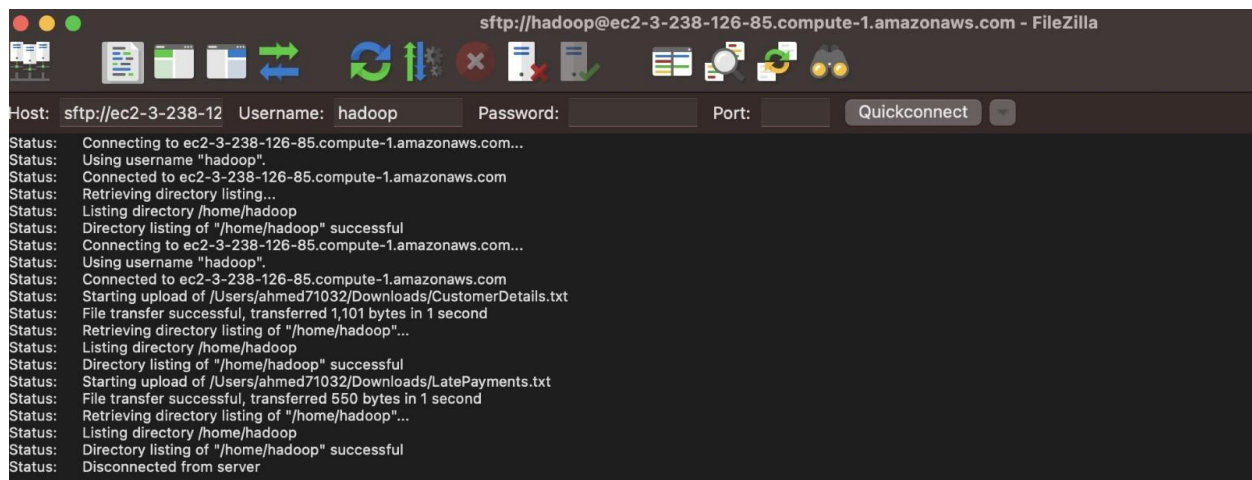

CREATING EMR CLUSTER:

MAKING CONNECTION AND UPLOADING FILES TO FILE ZILLA:

DISPLAYING TEXT FILES AND MAKING HIVE DIRECTORY:

```
   E::::::::::::E    M:::::M  M:::M::::M  M:::::M    R::::::::::::RR
   E:::::EEEEEEEEEE   M:::::M  M:::::M   M:::::M   R:::RRRRRR::::R
   E::::E          M:::::M   M:::M    M:::::M   R:::R      R::::R
   E::::E      EEEEE M:::::M    MMM    M:::::M   R:::R      R:::R
EE:::::EEEEEEE:::E M:::::M            M:::::M   R:::R      R:::R
E::::::::::::::::::E M:::::M           M:::::M RR:::R      R:::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM          MMMMMMM RRRRRRR      RRRRRR


[hadoop@ip-172-31-76-122 ~]$ ls
[hadoop@ip-172-31-76-122 ~]$ ls
CustomerDetails.txt  LatePayments.txt
[hadoop@ip-172-31-76-122 ~]$ cat C*
171     Madelyn Hensley 10075 Thierer Plaza     New York      New York      81377  1976   M    TRUE    FALSE   FALSE
172     Lonny   Foster  23901 Park Meadow Dr    Austin  Texas  13498  1981   F    TRUE    FALSE   FALSE
173     Karina  Livingston      95 Anderson Park        Chattanooga     Tennessee      94518  1977   F    TRUE    TRUE    TRUE
174     Avery   Mccormick       09992 Sunfield Parkway  Chicago Illinois      38300  1992   F    TRUE    FALSE   FALSE
175     Peter   King    0486 Dryden Road        Chicago Illinois      21012  1991   M    TRUE    FALSE   TRUE
176     Bret    Ibarra  14 Transport Place      San Diego      California     12865  1984   M    TRUE    FALSE   FALSE
177     Leonardo        Wheeler 806 Corry Crossing       New York      New York      44464  1972   F    FALSE   FALSE   TRUE
178     Bennett Noble   8 South Terrace Hixson  Tennessee      52890  1989   F    FALSE   FALSE   TRUE
179     Marcia  Matthews        0380 Knutson Road       Dallas  Texas  80477  1967   F    TRUE    FALSE   TRUE
180     Avis    Kramer  49624 Hanover Junction  New York      New York      62542  1965   M    TRUE    FALSE   TRUE
181     Lynnette        Tate    30741 Paget Court       New York      New York      34886  1987   M    TRUE    FALSE   TRUE
182     Lakisha Estrada 50 Dahle Crossing       Dallas  Texas  16042  1991   F    FALSE   TRUE    FALSE
183     Bill    Silva   4 Mcbride Crossing      Detroit Michigan      15871  1968   M    TRUE    TRUE    TRUE[hadoop@ip-172-31-76-122 ~]$
[hadoop@ip-172-31-76-122 ~]$ cat L*
181     146268743-8     23-8-2020       TRUE
172     396589804-7     28-9-2020       FALSE
183     553753031-8     25-9-2020       FALSE
183     559786593-4     13-12-2018      TRUE
175     108659198-9     13-9-2016       TRUE
176     360007723-2     27-9-2016       FALSE
177     238309554-X     28-11-2014      FALSE
178     694690715-8     31-5-2014       TRUE
179     318241713-5     15-7-2016       TRUE
180     84360172-2      6-7-2016        FALSE
181     807633856-8     5-7-2018        FALSE
182     845886260-4     23-4-2014       FALSE
183     270161074-X     19-7-2016       TRUE
171     887428332-0     4-7-2020        FALSE
172     401129380-4     23-8-2015       TRUE
173     277406266-7     27-12-2019      FALSE
174     851112155-2     4-12-2019       TRUE[hadoop@ip-172-31-76-122 ~]$
[hadoop@ip-172-31-76-122 ~]$ fs -mkdir HV
-bash: fs: command not found
[hadoop@ip-172-31-76-122 ~]$ hadoop fs -mkdir HV
[hadoop@ip-172-31-76-122 ~]$ hadoop fs -mkdir HV/input
[hadoop@ip-172-31-76-122 ~]$ hadoop fs -copyFromLocal * HV/input
[hadoop@ip-172-31-76-122 ~]$ hive
Hive Session ID = 3bd15b19-17f8-48c6-85c7-3d122f71aa31

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> show databases;
OK
default
Time taken: 0.78 seconds, Fetched: 1 row(s)
hive> CREATE DATABASE bank_db;
OK
Time taken: 0.226 seconds
hive>
```

CREATING DATABASES AND TABLES:

```
E:::::::::::::E    M:::::M  M::::M::::M  M:::::M   R::::::::::::RR
 E:::::EEEEEEEEE   M:::::M  M::::::M     M:::::M   R:::RRRRRR::::R
  E::::E           M:::::M  M:::M        M:::::M   R:::R       R::::R
  E::::E     EEEEE M:::::M     MMM       M:::::M   R:::R       R::::R
EE:::::EEEEEEEE:::E M:::::M               M:::::M   R:::R       R::::R
E:::::::::::::::::E M:::::M               M:::::M RR::::R       R::::R
EEEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR       RRRRRR

[hadoop@ip-172-31-76-122 ~]$ ls
[hadoop@ip-172-31-76-122 ~]$ ls
CustomerDetails.txt  LatePayments.txt
[hadoop@ip-172-31-76-122 ~]$ cat C*
171     Madelyn Hensley 10075 Thierer Plaza    New York     New York       81377   1976    M    TRUE    FALSE   FALSE
172     Lonny   Foster  23901 Park Meadow Dr    Austin  Texas   13498   1981    F    TRUE    FALSE   FALSE
173     Karina  Livingston      95 Anderson Park        Chattanooga     Tennessee       94518   1977    F    TRUE    TRUE    TRUE
174     Avery   Mccormick       09992 Sunfield Parkway  Chicago Illinois        38300   1992    F    TRUE    FALSE   FALSE
175     Peter   King    0486 Dryden Road        Chicago Illinois        21012   1991    M    TRUE    FALSE   TRUE
176     Bret    Ibarra  14 Transport Place      San Diego       California      12865   1984    M    TRUE    FALSE   FALSE
177     Leonardo        Wheeler 806 Corry Crossing      New York        New York        44464   1972    F    FALSE   FALSE   TRUE
178     Bennett Noble   8 South Terrace Hixson  Tennessee       52890   1989    F    FALSE   FALSE   TRUE
179     Marcia  Matthews        0380 Knutson Road       Dallas  Texas   80477   1967    F    TRUE    FALSE   TRUE
180     Avis    Kramer  49624 Hanover Junction  New York        New York        62542   1965    M    TRUE    FALSE   TRUE
181     Lynnette        Tate    30741 Paget Court       New York        New York        34886   1987    M    TRUE    FALSE   TRUE
182     Lakisha Estrada 50 Dahle Crossing       Dallas  Texas   16042   1991    F    FALSE   TRUE    FALSE
183     Bill    Silva   4 Mcbride Crossing      Detroit Michigan        15871   1968    M    TRUE    TRUE    TRUE[hadoop@ip-172-31-76-122 ~]$
[hadoop@ip-172-31-76-122 ~]$ cat L*
181     146268743-8     23-8-2020       TRUE
172     396589804-7     28-9-2020       FALSE
183     553753031-8     25-9-2020       FALSE
183     559786593-4     13-12-2018      TRUE
175     108659198-9     13-9-2016       TRUE
176     360007723-2     27-9-2016       FALSE
177     238309554-X     28-11-2014      FALSE
178     694690715-8     31-5-2014       TRUE
179     318241713-5     15-7-2016       TRUE
180     84360172-2      6-7-2016        FALSE
181     807633856-8     5-7-2018        FALSE
182     845886260-4     23-4-2014       FALSE
183     270161074-X     19-7-2016       TRUE
171     887428332-0     4-7-2020        FALSE
172     401129380-4     23-8-2015       TRUE
173     277406266-7     27-12-2019      FALSE
174     851112155-2     4-12-2019       TRUE[hadoop@ip-172-31-76-122 ~]$
[hadoop@ip-172-31-76-122 ~]$ fs -mkdir HV
-bash: fs: command not found
[hadoop@ip-172-31-76-122 ~]$ hadoop fs -mkdir HV
[hadoop@ip-172-31-76-122 ~]$ hadoop fs -mkdir HV/input
[hadoop@ip-172-31-76-122 ~]$ hadoop fs -copyFromLocal * HV/input
[hadoop@ip-172-31-76-122 ~]$ hive
Hive Session ID = 3bd15b19-17f8-48c6-85c7-3d122f71aa31

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> show databases;
OK
default
Time taken: 0.78 seconds, Fetched: 1 row(s)
hive> CREATE DATABASE bank_db;
OK
Time taken: 0.226 seconds
```

```
[hive> CREATE TABLE IF NOT EXISTS bank_db.payments
[    > (C_ID int,transaction_ID String,transaction_DATE String,late String)
[    > ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
[    > STORED AS TEXTFILE;
OK
Time taken: 0.581 seconds
[hive> describe bank_db.payments;
OK
c_id                    int
transaction_id          string
transaction_date        string
late                    string
Time taken: 0.148 seconds, Fetched: 4 row(s)
[hive> use bank_db
[    > describe payments
[    > ;
FAILED: ParseException line 2:0 missing EOF at 'describe' near 'bank_db'
[hive> use bank_db;
OK
Time taken: 0.028 seconds
[hive> describe payments;
OK
c_id                    int
transaction_id          string
transaction_date        string
late                    string
Time taken: 0.09 seconds, Fetched: 4 row(s)
[hive> CREATE TABLE IF NOT EXISTS bank_db.payments_ptn
[    > (C_ID int,transaction_ID String,transaction_DATE String)
[    > PARTITIONED BY (late String)
[    > ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
[    > STORED AS TEXTFILE;
OK
Time taken: 0.205 seconds
[hive> describe payments_ptn
[    > ;
OK
c_id                    int
transaction_id          string
transaction_date        string
late                    string

# Partition Information
# col_name              data_type               comment
late                    string
Time taken: 0.093 seconds, Fetched: 8 row(s)
```

```
[hive> CREATE EXTERNAL TABLE IF NOT EXISTS bank_db.customers_ext
[    > (C_ID int,C_FNAME String,C_LNAME String,C_street String,C_CITY String,C_State String,C_zip String,C_YOB INT,C_gender String,CREDIT_CARD String, INTERNET ]
String,MOBILE String)
[    > ROW FORMAT DELIMITED                                                                                                                                      ]
[    > FIELDS TERMINATED BY '\t'                                                                                                                                 ]
[    > LINES TERMINATED BY '\n'                                                                                                                                  ]
[    > STORED AS TEXTFILE                                                                                                                                        ]
[    > LOCATION '/user/hive/warehouse/bank_db/customers_ext';                                                                                                    ]
OK
Time taken: 0.132 seconds
[hive> show tables;                                                                                                                                              ]
OK
customers_ext
payments
payments_ptn
Time taken: 0.039 seconds, Fetched: 3 row(s)
[hive> drop table payments_ptn;
OK
Time taken: 0.34 seconds
[hive> show tables;                                                                                                                                              ]
OK
customers_ext
payments
Time taken: 0.069 seconds, Fetched: 2 row(s)
[hive> alter table customers_ext RENAME TO customers;                                                                                                            ]
OK
Time taken: 0.218 seconds
[hive> show tables;                                                                                                                                              ]
OK
customers
payments
Time taken: 0.026 seconds, Fetched: 2 row(s)
[hive> describe customers;                                                                                                                                       ]
OK
c_id                    int
c_fname                 string
c_lname                 string
c_street                string
c_city                  string
c_state                 string
c_zip                   string
c_yob                   int
c_gender                string
credit_card             string
internet                string
mobile                  string
Time taken: 0.048 seconds, Fetched: 12 row(s)
[hive> LOAD DATA INPATH 'HV/input/LatePayments.txt' OVERWRITE INTO TABLE payments;                                                                               ]
Loading data to table bank_db.payments
OK
Time taken: 0.474 seconds
hive> []
```

SELECTING ALL COLUMNS FROM TABLES TO CROSS-CHECK:

```
[hive> select * from payments;
OK
181     146268743-8     23-8-2020       TRUE
172     396589804-7     28-9-2020       FALSE
183     553753031-8     25-9-2020       FALSE
183     559786593-4     13-12-2018      TRUE
175     108659198-9     13-9-2016       TRUE
176     360007723-2     27-9-2016       FALSE
177     238309554-X     28-11-2014      FALSE
178     694690715-8     31-5-2014       TRUE
179     318241713-5     15-7-2016       TRUE
180     84360172-2      6-7-2016        FALSE
181     807633856-8     5-7-2018        FALSE
182     845886260-4     23-4-2014       FALSE
183     270161074-X     19-7-2016       TRUE
171     887428332-0     4-7-2020        FALSE
172     401129380-4     23-8-2015       TRUE
173     277406266-7     27-12-2019      FALSE
174     851112155-2     4-12-2019       TRUE
Time taken: 2.21 seconds, Fetched: 17 row(s)
[hive> LOAD DATA INPATH 'HV/input/CustomerDetails.txt' OVERWRITE INTO TABLE customers;
Loading data to table bank_db.customers
OK
Time taken: 0.259 seconds
[hive> select * from customers;
OK
171     Madelyn Hensley 10075 Thierer Plaza     New York        New York        81377   1976    M       TRUE    FALSE   FALSE
172     Lonny   Foster  23901 Park Meadow Dr    Austin  Texas   13498   1981    F       TRUE    FALSE   FALSE
173     Karina  Livingston      95 Anderson Park        Chattanooga     Tennessee       94518   1977    F       TRUE    TRUE    TRUE
174     Avery   Mccormick       09992 Sunfield Parkway  Chicago Illinois        38300   1992    F       TRUE    FALSE   FALSE
175     Peter   King    0486 Dryden Road        Chicago Illinois        21012   1991    M       TRUE    FALSE   TRUE
176     Bret    Ibarra  14 Transport Place      San Diego       California      12865   1984    M       TRUE    FALSE   FALSE
177     Leonardo        Wheeler 806 Corry Crossing      New York        New York        44464   1972    F       FALSE   FALSE   TRUE
178     Bennett Noble   8 South Terrace Hixson  Tennessee       52890   1989    F       FALSE   FALSE   TRUE
179     Marcia  Matthews        0380 Knutson Road       Dallas  Texas   80477   1967    F       TRUE    FALSE   TRUE
180     Avis    Kramer  49624 Hanover Junction  New York        New York        62542   1965    M       TRUE    FALSE   TRUE
181     Lynnette        Tate    30741 Paget Court       New York        New York        34886   1987    M       TRUE    FALSE   TRUE
182     Lakisha Estrada 50 Dahle Crossing       Dallas  Texas   16042   1991    F       FALSE   TRUE    FALSE
183     Bill    Silva   4 Mcbride Crossing      Detroit Michigan        15871   1968    M       TRUE    TRUE    TRUE
Time taken: 0.187 seconds, Fetched: 13 row(s)
```

QUERYING THE DATA:

```
[hive> select c_id, c_fname, c_lname
[    > from customers
[    > where c_state = 'New York';
OK
171     Madelyn Hensley
177     Leonardo        Wheeler
180     Avis    Kramer
181     Lynnette        Tate
Time taken: 0.333 seconds, Fetched: 4 row(s)
hive> select c_state, count(c_id) as total
    > from customers
[    > group by c_state;
Query ID = hadoop_20240403190651_af3f117e-09f1-44be-b117-03f74f0adbec
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container    SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container    SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 15.72 s
----------------------------------------------------------------------------------------------
OK
Illinois        2
New York        4
California      1
Michigan        1
Tennessee       2
Texas   3
Time taken: 25.516 seconds, Fetched: 6 row(s)
```

```
FAILED: ParseException line 3.0 cannot recognize input near 'where' 'p' '.' in expression specification
hive> select c.c_id, c.c_fname, c.c_lname, p.transaction_id, p.transaction_date
[    >  from customers c join payments p on c.c_id = p.c_id
[    > where p.late = 'TRUE';
Query ID = hadoop_20240403190931_1873553b-f328-43c4-ba2d-ece95824a8f4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 2 .......... container    SUCCEEDED      1          1        0        0       0       0
Map 1 .......... container    SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 12.20 s
----------------------------------------------------------------------------------------------
OK
172     Lonny   Foster  401129380-4     23-8-2015
174     Avery   Mccormick       851112155-2     4-12-2019
175     Peter   King    108659198-9     13-9-2016
178     Bennett Noble   694690715-8     31-5-2014
179     Marcia  Matthews        318241713-5     15-7-2016
181     Lynnette        Tate    146268743-8     23-8-2020
183     Bill    Silva   270161074-X     19-7-2016
183     Bill    Silva   559786593-4     13-12-2018
Time taken: 13.282 seconds, Fetched: 8 row(s)
hive> select c_state, count(c_id)
    > from customers
    > where credit_card = 'FALSE'
[    > group by c_state;
Query ID = hadoop_20240403191003_f49774e8-15e5-4c4d-957f-58425ac20afe
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container    SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container    SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.06 s
----------------------------------------------------------------------------------------------
OK
New York        1
Tennessee       1
Texas   1
Time taken: 8.84 seconds, Fetched: 3 row(s)
hive> select c.c_state, count(c.c_id) as total
    > from customers c join payments p on c.c_id = p.c_id
    > where p.late = 'TRUE'
    > group BY c.c_state;
```

```
hive> select c_state, count(c_id)
    > from customers
    > where credit_card = 'FALSE'
    > group by c_state;
Query ID = hadoop_20240403191003_f49774e8-15e5-4c4d-957f-58425ac20afe
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     1        1         0        0       0       0
Reducer 2 ...... container     SUCCEEDED     2        2         0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02 [==========================>>] 100%  ELAPSED TIME: 8.06 s
----------------------------------------------------------------------------------------------
OK
New York        1
Tennessee       1
Texas   1
Time taken: 8.84 seconds, Fetched: 3 row(s)
hive> select c.c_state, count(c.c_id) as total
    > from customers c join payments p on c.c_id = p.c_id
    > where p.late = 'TRUE'
    > group BY c.c_state;
Query ID = hadoop_20240403191028_969e55db-014c-44d1-a9dc-d47885555e30
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 3 .......... container     SUCCEEDED     1        1         0        0       0       0
Map 1 .......... container     SUCCEEDED     1        1         0        0       0       0
Reducer 2 ...... container     SUCCEEDED     2        2         0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 03/03 [==========================>>] 100%  ELAPSED TIME: 13.67 s
----------------------------------------------------------------------------------------------
OK
Illinois        2
New York        1
Michigan        2
Tennessee       1
Texas   2
Time taken: 14.687 seconds, Fetched: 5 row(s)
hive> select *
    > from customers
    > where mobile = 'FALSE';
OK
171     Madelyn Hensley 10075 Thierer Plaza     New York    New York     81377  1976   M    TRUE   FALSE  FALSE
172     Lonny   Foster  23901 Park Meadow Dr    Austin  Texas  13498  1981   F    TRUE   FALSE  FALSE
174     Avery   Mccormick       09992 Sunfield Parkway  Chicago Illinois    38300  1992   F    TRUE   FALSE  FALSE
176     Bret    Ibarra  14 Transport Place      San Diego   California    12865  1984   M    TRUE   FALSE  FALSE
182     Lakisha Estrada 50 Dahle Crossing       Dallas  Texas  16042  1991   F    FALSE  TRUE   FALSE
Time taken: 0.208 seconds, Fetched: 5 row(s)
```

```
hive> select c.c_id, c.c_fname, c.c_lname, c.c_street, c.c_city, c.c_state, c.c_zip, c.c_yob, c.c_gender, c.credit_card, c.internet, c.mobile
    > from customers c join payments p on c.c_id = p.c_id
    > where p.late = 'TRUE';
Query ID = hadoop_20240403191159_9897a221-d56c-4bff-bdd5-a713caf57f81
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 2 ......... container     SUCCEEDED     1        1         0        0       0       0
Map 1 ......... container     SUCCEEDED     1        1         0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02 [==========================>>] 100%  ELAPSED TIME: 12.85 s
----------------------------------------------------------------------------------------------
OK
172     Lonny   Foster  23901 Park Meadow Dr    Austin  Texas  13498  1981   F    TRUE   FALSE  FALSE
174     Avery   Mccormick       09992 Sunfield Parkway  Chicago Illinois    38300  1992   F    TRUE   FALSE  FALSE
175     Peter   King    0486 Dryden Road        Chicago Illinois    21012  1991   M    TRUE   FALSE  TRUE
178     Bennett Noble   8 South Terrace Hixson  Tennessee    52890  1989   F    FALSE  FALSE  TRUE
179     Marcia  Matthews        0380 Knutson Road       Dallas  Texas  80477  1967   F    TRUE   FALSE  TRUE
181     Lynnette        Tate    30741 Paget Court       New York    New York     34886  1987   M    TRUE   TRUE   FALSE  TRUE
183     Bill    Silva   4 Mcbride Crossing      Detroit Michigan    15871  1968   M    TRUE   TRUE   TRUE
183     Bill    Silva   4 Mcbride Crossing      Detroit Michigan    15871  1968   M    TRUE   TRUE   TRUE
Time taken: 13.551 seconds, Fetched: 8 row(s)
hive> select c.c_id, c.c_fname, c.c_lname, c.c_city, c.c_state
    > from customers c join payments p on c.c_id = p.c_id
    > where p.late = 'FALSE';
Query ID = hadoop_20240403191250_f58f7e63-9e5b-4d3d-bc5c-eb9e56daae61
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 2 ......... container     SUCCEEDED     1        1         0        0       0       0
Map 1 ......... container     SUCCEEDED     1        1         0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02 [==========================>>] 100%  ELAPSED TIME: 12.86 s
----------------------------------------------------------------------------------------------
OK
171     Madelyn Hensley New York        New York
172     Lonny   Foster  Austin  Texas
173     Karina  Livingston      Chattanooga     Tennessee
176     Bret    Ibarra  San Diego       California
177     Leonardo        Wheeler New York        New York
180     Avis    Kramer  New York        New York
181     Lynnette        Tate    New York        New York
182     Lakisha Estrada Dallas  Texas
183     Bill    Silva   Detroit Michigan
Time taken: 13.658 seconds, Fetched: 9 row(s)
hive> select c.c_fname, c.c_fname, c.c_city, c.c_state
    > from customers c join payments p on c.c_id = p.c_id
    > where p.late = 'TRUE' and c.c_yob > 1985;
```

```
Time taken: 13.858 seconds, Fetched: 9 row(s)
hive> select c.c_fname, c.c_fname, c.c_city, c.c_state
    > from customers c join payments p on c.c_id = p.c_id
[   > where p.late = 'TRUE' and c.c_yob > 1985;
Query ID = hadoop_20240403191331_28a010fb-b668-4ab8-9c29-08890ca73be4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

--------------------------------------------------------------------------
        VERTICES      MODE        STATUS   TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------
Map 2 ......... container      SUCCEEDED      1         1        0        0       0       0
Map 1 ......... container      SUCCEEDED      1         1        0        0       0       0
--------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 12.47 s
--------------------------------------------------------------------------
OK
Avery    Avery    Chicago Illinois
Peter    Peter    Chicago Illinois
Bennett Bennett Hixson    Tennessee
Lynnette         Lynnette        New York         New York
Time taken: 13.58 seconds, Fetched: 4 row(s)
hive> select c.c_id, c.c_fname, c.c_lname, c.c_city, c.c_state
    > from customers c join payments p on c.c_id = p.c_id
    > where p.late = 'FALSE' and c.internet = 'TRUE'
    > group by c.c_id, c.c_fname, c.c_lname, c.c_city, c.c_state;
Query ID = hadoop_20240403191428_f07b68a8-532c-4cf0-bc6c-fe5b2cd10c7c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1712167247158_0002)

--------------------------------------------------------------------------
        VERTICES      MODE        STATUS   TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------
Map 3 ......... container      SUCCEEDED      1         1        0        0       0       0
Map 1 ......... container      SUCCEEDED      1         1        0        0       0       0
Reducer 2 ...... container      SUCCEEDED      2         2        0        0       0       0
--------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 14.55 s
--------------------------------------------------------------------------
OK
173      Karina  Livingston      Chattanooga      Tennessee
182      Lakisha Estrada Dallas  Texas
183      Bill    Silva   Detroit Michigan
Time taken: 15.453 seconds, Fetched: 3 row(s)
hive>
```

TERMINATING THE CLUSTER AND ENDING LAB: