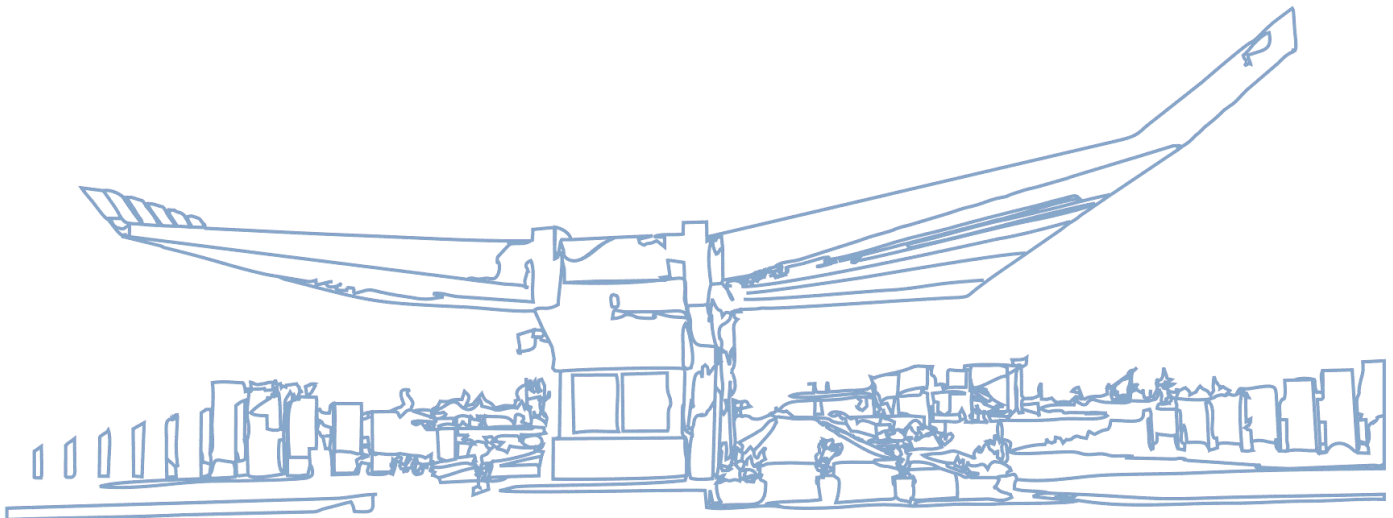


# CEN 571 – Data Mining

## Assignment 5: Hive and HiveSQL



PREPARED:  
**Baftjar TABAKU**

**10.05.2020**  
Epoka University  
Tirana, ALBANIA

ACCEPTED:  
**Prof.Dr. Arben Asllani**

Exercise 1: Files 'CustomerDetails.txt' and 'LatePayments.txt' contain the full data set about customers of the commercial bank and their payment transactions. Download the file, study the structure, and perform the following functions.

1. Download the files and upload them in HDFS under the 'HV/Exercise1/input' directory.

The screenshot shows the FileZilla interface with the local site at 'C:\Users\Tabaku\Desktop\BigDataHW\_05\input\' and the remote site at '/home/hadoop'. The local site contains files 'customer\_details.txt' (888,104 bytes) and 'late\_payments.txt' (685,959 bytes). The remote site contains a directory listing showing the files have been successfully uploaded.

Filename	Filesize	Filetype	Last modified
customer_details.txt	888,104	Text Document	5/5/2020 7:48:24 PM
late_payments.txt	685,959	Text Document	5/5/2020 7:50:05 PM

Selected 2 files. Total size: 1,574,063 bytes

The screenshot shows the FileZilla interface with the status bar indicating 'Transfers finished' and 'All files have been successfully transferred.' The queue is empty.

```
hadoop@ip-172-31-6-30:~$ https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/
22 package(s) needed for security, out of 41 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M M::::::::M R::::::::::::R
EE::::::::EEEEEEEE::E M::::::::M M::::::::M R::::::::RRRRR:::R
E:::::E EEEEE M::::::::M M::::::::M RR::::R R::::R
E:::::E M:::::M:::M M:::M:::M R:::R R:::R
E:::::EEEEEEEEEE M:::::M M:::M M:::M R:::RRRRRR:::R
E:::::EEEEEEEEEE M:::::M M:::M:::M M:::M R:::RRRRRR:::R
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:::::EEEEEEEE:::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-172-31-6-30 ~]$ hadoop fs -mkdir HV;
[hadoop@ip-172-31-6-30 ~]$ hadoop fs -mkdir HV/Exercise1;
[hadoop@ip-172-31-6-30 ~]$ hadoop fs -mkdir HV/Exercise1/input;
[hadoop@ip-172-31-6-30 ~]$ hadoop fs -copyFromLocal customer_details.txt HV/Exercise1/input;
[hadoop@ip-172-31-6-30 ~]$ hadoop fs -copyFromLocal late_payments.txt HV/Exercise1/input;

[hadoop@ip-172-31-6-30 ~]$
[hadoop@ip-172-31-6-30 ~]$
[hadoop@ip-172-31-6-30 ~]$ hadoop fs -ls HV/Exercise1/input/;
Found 2 items
-rw-r--r-- 1 hadoop hadoop 888104 2020-05-09 15:25 HV/Exercise1/input/customer_details.txt
-rw-r--r-- 1 hadoop hadoop 685959 2020-05-09 15:25 HV/Exercise1/input/late_payments.txt
[hadoop@ip-172-31-6-30 ~]$
```

File were uploaded successfully.

## 2. Create a database named 'exercise1' and create tables 'customers' and 'payments'.

```
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: true
hive> CREATE DATABASE exercise1;
OK
Time taken: 0.589 seconds
hive> show databases;
OK
default
exercise1
Time taken: 0.108 seconds, Fetched: 2 row(s)
hive> CREATE EXTERNAL TABLE IF NOT EXISTS exercise1.customers(c_ID int, c_name String, c_street String, c_city String, c_state String, c_zip String, c_date_of_birth String, c_gender String, c_credit_card String, c_internet String, c_mobile String) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED AS TEXTFILE LOCATION '/user/hive/warehouse/exercise1/customers';
OK
Time taken: 0.473 seconds
hive> CREATE TABLE IF NOT EXISTS exercise1.payments(p_ID int, p_transaction_ID String, p_transaction_DATE String, p_late String) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;
OK
Time taken: 0.186 seconds
```

```
OK
Time taken: 0.473 seconds
hive> CREATE TABLE IF NOT EXISTS exercise1.payments(p_ID int, p_transaction_ID String, p_transaction_DATE String, p_late String) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;
OK
Time taken: 0.186 seconds
hive> use exercise1;
OK
Time taken: 0.01 seconds
hive> show tables;
OK
customers
payments
Time taken: 0.027 seconds, Fetched: 2 row(s)
hive> describe customers;
OK
c_id                int
c_name              string
c_street            string
c_city              string
c_state             string
c_zip              string
c_date_of_birth     string
c_gender            string
c_credit_card       string
c_internet          string
c_mobile            string
Time taken: 0.07 seconds, Fetched: 11 row(s)
hive> describe payments;
OK
p_id                int
p_transaction_id    string
p_transaction_date  string
p_late              string
Time taken: 0.03 seconds, Fetched: 4 row(s)
hive>
> ;
hive>
```

## 3. Load the data from HDFS into each table respectively.

```
hive> LOAD DATA LOCAL INPATH './customer_details.txt' OVERWRITE INTO TABLE customers;
Loading data to table exercise1.customers
OK
Time taken: 0.935 seconds
hive> LOAD DATA LOCAL INPATH './late_payments.txt' OVERWRITE INTO TABLE payments;
Loading data to table exercise1.payments
OK
Time taken: 1.135 seconds
hive>
```

4. Display the ID, name, and year of birth, zip code, for all customers from the state of New York.

```
hive> SELECT c_ID, c_name, c_date_of_birth, c_zip FROM customers WHERE c_state = 'New York';
```

31	Beverly Hall	29-08-1956	71158-1776
42	Christine Romero	13-10-1976	63093-8409
52	Timothy Black	17-04-1962	96389
58	Gloria Harrison	26-02-1990	86190-0452
78	Matthew West	19-06-1975	81501-0784
88	Pamela Daniels	05-01-1984	12931
106	Paul Clark	26-11-2009	618
107	William Ross	13-02-1992	08120-0174
147	Brenda Brooks	15-07-1956	43340
149	Mildred Ryan	21-04-1967	94163
154	Norma Carroll	17-11-1981	16395-5939
156	Timothy Shaw	01-04-1977	66225
166	Sandra Black	12-08-1971	11326
170	Heather Mills	21-01-1989	20003
177	Daniel Carter	10-09-1992	14350-9344
191	Joshua Watson	16-10-2001	90933

5. Display the total number of customers who have occurred late payments for each zip code.

```
> ;
hive> SELECT c_zip , COUNT(c_ID) AS total FROM customers JOIN payments ON c_ID = p_ID WHERE p_late = 'TRUE' GROUP BY c_zip;
```

Query ID = hadoop\_20200509153321\_867ab794-a818-4a68-977a-48b10b44de88  
Total jobs = 1  
Launching Job 1 out of 1  
Tez session was closed. Reopening...  
Session re-established.  
Status: Running (Executing on YARN cluster with App id application\_1589037740563\_0002)

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	.....	container	SUCCEEDED	1	1	0	0	0	0
Map 3	.....	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	.....	container	SUCCEEDED	1	1	0	0	0	0

VERTICES: 03/03 [=====>>] 100% ELAPSED TIME: 6.76 s

```
OK
00079-5661      9
00688-3484     19
00767-9960     12
00970-3826     19
01048-8772     12
01054-0524     16
01470-8811     22
01866-6001     14
01932-6113     14
02185-6021     15
02280-3985      9
02282-4797     13
02363-5505     12
02403-1991     12
02477-6867     10
```

```

07135-0214      12
07143-4577      20
07668-0327      12
07758-5854      17
08120-0174      17
08126-2501      15
08564-6204      21
08704-3091      17
08723-1298      15
09020-5046      19
09269-6124      13
09392-9166      11
09792-6187       9
09935-9403      17
10001   7
10224   18
10297-5881       9
10345   10
10564   13
10630   16
10796   10
10938-8638      12
11089   13
11274-7392      17
11472-5630      15
115     10
11611-1029      15
11691   21
11763   14
11769-6085      23

```

6. Display the customers who have any late transactions and the date of each late transaction for the customers born before 1985.

```

hive>
> SELECT c_ID, c_name, c_street, c_city, c_state, c_zip, c_date_of_birth, c_gender FROM customers JOIN payments ON c_ID = p_ID WHERE p_late = 'TRUE' AND to_d
ate(from unixtime(UNIX_TIMESTAMP(c_date_of_birth,'dd-MM-yyyy')) < 1985 GROUP BY c_ID, c_name, c_street, c_city, c_state, c_zip, c_date_of_birth, c_gender;
Query ID = hadoop_20200509153650_5800fa2d-97aa-45f7-8d3a-cb167b23625c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1589037740563_0002)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 3 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 03/03 [=====>>>] 100% ELAPSED TIME: 7.51 s
-----
OK
32 Bonnie Cox 671 Merry Trail CA California 69224-6813 29-02-1964 F
33 Samuel Oliver 52 Swallow Plaza IN Indiana 36867-2654 11-10-1962 M
66 Terry Taylor 709 Boyd Alley OH Ohio 41640-6487 23-08-1954 M
68 Virginia Marshall 306 Susan Terrace MA Massachusetts 13442-6546 17-06-1978 F
70 Andrew White 55 Hooker Plaza IL Illinois 36610 03-04-1972 M
72 Peter Fisher 2 Rowland Drive MN Minnesota 31408-1732 23-04-1964 M
76 Melissa Griffin 40325 Stuart Circle NJ New Jersey 66018-0390 22-04-1971 F

```

32	Bonnie Cox	671 Merry Trail	CA	California	69224-6813	29-02-1964	F	
33	Samuel Oliver	52 Swallow Plaza	IN	Indiana	36867-2654	11-10-1962	M	
66	Terry Taylor	709 Boyd Alley	OH	Ohio	41640-6487	23-08-1954	M	
68	Virginia Marshall	306 Susan Terrace	MA	Massachusetts	13442-6546	17-06-1978	F	
70	Andrew White	55 Hooker Plaza	IL	Illinois	36610	03-04-1972	M	
72	Peter Fisher	2 Rowland Drive	MN	Minnesota	31408-1732	23-04-1964	M	
76	Melissa Griffin	40325 Stuart Circle	NJ	New Jersey	66018-0390	22-04-1971	F	
78	Matthew West	60791 Rusk Pass	NY	New York	81501-0784	19-06-1975	M	
79	Pamela Dean	52798 Morning Way	IL	Illinois	14082-7694	02-12-1953	F	
80	Kenneth Lawrence	664 Victoria Terrace	TX	Texas	51615	27-07-1979	M	
81	Stephen Banks	0903 Mendota Lane	MA	Massachusetts	54274-2091	23-11-1980	M	
82	Jeremy Lynch	752 Melody Street	MI	Michigan	35992	18-10-1971	M	
84	Sarah Johnson	2575 Redwing Parkway	CA	California	74315-7626	04-05-1983	F	
88	Pamela Daniels	3 Northland Street	NY	New York	12931	05-01-1984	F	
89	Andrew Howard	76 Lake View Hill	CA	California	35672	30-06-1966	M	
91	Julia James	4271 Morning Drive	IN	Indiana	21666	27-01-1957	F	
93	Jerry Hunt	9 Lerdaahl Way	MI	Michigan	87242	29-08-1982	M	
94	Randy Alexander	1439 Walton Junction	MN	Minnesota	12183	10-10-1963	M	
96	Jennifer Carpenter	0 Bowman Park	CT	Connecticut	28308-5239	21-06-1966	F	
97	William Sullivan	3122 Grasskamp Street	TX	Texas	10938-8638	26-12-1972	M	
98	Russell Wagner	78 Dahle Hill	NC	North Carolina	28360	26-06-1975	M	
99	Rachel Cooper	587 Loftsgordon Road	NJ	New Jersey	17495-9311	23-01-1984	F	
104	Raymond Long	36 Sutteridge Parkway	NJ	New Jersey	90801-8180	29-06-1961	M	
108	Roy West	6 Mallory Pass	NH	New Hampshire	70791-0596	10-09-1975	M	
109	Joe Taylor	199 Daystar Parkway	NJ	New Jersey	82974-5347	14-07-1956	M	

7. Display the cities and the number of customers who do not have a credit card for each city sorted by the number of customers.

```
hive> SELECT c_city AS city , COUNT(c_ID) as total FROM customers JOIN payments ON c_ID = p_ID WHERE c_credit_card = 'FALSE' GROUP BY c_city SORT BY total ASC;
Query ID = hadoop_20200509153910_0d38ecd9-4241-432a-9d11-5fc6a167c8e6
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1589037740563_0002)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1 .....	container	SUCCEEDED	1	1	0	0	0	0
Map 2 .....	container	SUCCEEDED	1	1	0	0	0	0
Reducer 3 .....	container	SUCCEEDED	1	1	0	0	0	0
Reducer 4 .....	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 04/04 [=====>>] 100% ELAPSED TIME: 5.94 s
OK
ND      30
WV      58
NE      83
PA     120
ME     122
CO     136
NC     144
NH     148
```

```

Map 2 ..... container SUCCEEDED 1 1 0
Reducer 3 ..... container SUCCEEDED 1 1 0
Reducer 4 ..... container SUCCEEDED 1 1 0
-----
VERTICES: 04/04 [=====>>] 100% ELAPSED TIME: 5.94 s
-----
OK
ND      30
WV      58
NE      83
PA     120
ME     122
CO     136
NC     144
NH     148
AR     190
AL     202
IA     227
IN     228
OK     256
VA     271
MN     324
FL     359
MI     367
CT     514
OH     545
MA     656
TX     754
NJ     864
IL     941
CA    1043
NY    1117
Time taken: 6.402 seconds, Fetched: 25 row(s)
hive>

```

8. Display the state and number of customers who have late payments and use mobile banking in each state

```

hive> SELECT c_state AS state , COUNT(c_ID) as total FROM customers JOIN payments ON c_ID = p_ID WHERE p_late = 'TRUE' AND c_mobile = 'TRUE' GROUP BY c_state;
Query ID = hadoop_20200509155212_882d4e2d-76f5-4f0b-bfc6-1ac3ce7622b1
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1589037740563_0003)
-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container SUCCEEDED 1      1      0      0      0      0
Map 2 ..... container SUCCEEDED 1      1      0      0      0      0
Reducer 3 ..... container SUCCEEDED 1      1      0      0      0      0
-----
VERTICES: 03/03 [=====>>] 100% ELAPSED TIME: 5.91 s
-----
OK
Alabama 58
Arkansas 52
California 542
Colorado 94
Connecticut 208
Delaware 12
Florida 219
Illinois 412
Indiana 115

```



```

VERTICES: 03/03 [=====>>] 100%

OK
Alabama 58
Arkansas      52
California    542
Colorado      94
Connecticut   208
Delaware      12
Florida 219
Illinois      412
Indiana 115
Iowa         84
Maine        42
Massachusetts 485
Michigan      246
Minnesota     250
Nebraska      66
New Hampshire 39
New Jersey    430
New York      512
North Carolina 119
North Dakota  34
Ohio         388
Oklahoma      43
Pennsylvania  61
Rhode Island  17
Texas        410
Virginia      129
Time taken: 11.824 seconds, Fetched: 26 row(s)
hive>

```

## 9. Display all customers who do not have a mobile phone.

```

>
>
>
>
>
>
>
hive> SELECT c_ID, c_name, c_street, c_city, c_state, c_zip, c_date_of_birth, c_gender, c_credit_card, c_internet, c_mobile FROM customers WHERE c_mobile = 'FALSE';

```

9968	Matthew Ramos	1643 Waxwing Terrace	NJ	New Jersey	55333-6029	17-04-1969	M	FALSE	FALSE	FALSE
9970	Marilyn Williamson	35760 Almo Park	CA	California	32237	11-07-1971	F	TRUE	TRUE	FALSE
9971	Bonnie Lewis	83 Monica Avenue	FL	Florida	78927	28-06-1968	F	TRUE	FALSE	FALSE
9972	Wayne Arnold	86 Melrose Court	RI	Rhode Island	39515-1259	25-07-1977	M	TRUE	FALSE	FALSE
9973	Thomas Gordon	1 Nelson Avenue	IL	Illinois	82477	16-05-1985	M	TRUE	FALSE	FALSE
9975	Earl Perez	5 Hayes Point	OK	Oklahoma	81165	30-03-1991	M	FALSE	FALSE	FALSE
9976	Benjamin Chavez	70 Bonner Way	TX	Texas	84743-9089	28-10-1975	M	FALSE	TRUE	FALSE
9978	Johnny Ray	26 Onsgard Plaza	DE	Delaware	31466-7646	19-11-1995	M	FALSE	FALSE	FALSE
9979	Tammy Powell	14 Logan Trail	AL	Alabama	39312-2285	10-05-1974	F	FALSE	FALSE	FALSE
9981	Marilyn Nichols	98 Buena Vista Avenue	MA	Massachusetts	94858	13-04-1957	F	TRUE	TRUE	FALSE
9984	Anna Ryan	81590 Arapahoe Lane	NC	North Carolina	66036	06-10-1963	F	FALSE	FALSE	FALSE
9986	Jennifer Wright	2 Dapin Street	CA	California	38659-2692	25-01-1991	F	FALSE	FALSE	FALSE
9987	Arthur Elliott	914 Golden Leaf Place	CA	California	48528	13-05-1992	M	TRUE	FALSE	FALSE
9990	Frank Hayes	3 Little Fleur Trail	CA	California	78523	22-08-2007	M	TRUE	FALSE	FALSE
9991	Elizabeth Baker	11 Raven Place	MI	Michigan	50420	09-03-1970	F	FALSE	TRUE	FALSE
9992	Alan Riley	372 8th Point	IL	Illinois	28024-4557	21-03-1975	M	TRUE	TRUE	FALSE
9993	Beverly Stanley	4611 Washington Junction	CA	California	06387-1654	06-12-1956	F	TRUE	TRUE	FALSE
9995	Stephen Hernandez	680 Fairview Alley	MN	Minnesota	41049	03-11-2007	M	FALSE	TRUE	FALSE
9999	Stephen Gordon	93408 8th Alley	MA	Massachusetts	22514-1265	26-05-1986	M	TRUE	FALSE	FALSE

```

Time taken: 0.083 seconds, Fetched: 5020 row(s)
hive>

```



10. Display all customers who have late payments and the number of late payments sorted by the number of late payments.

```

> ?
hive> SELECT c_ID, c_name, c_city, c_state, COUNT(c_ID) as total FROM customers JOIN payments ON c_ID = p_ID WHERE p_late = 'TRUE' GROUP BY c_ID, c_name, c_city, c_state SORT BY total ASC;
Query ID = hadoop_20200509155806_2cb5a45d-ad81-4e7c-a593-ae4b2e86dd03
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1589037740563_0004)
-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 3 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 4 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 04/04 [=====] 100% ELAPSED TIME: 6.35 s
-----
OK
1003  Ashley Ford  VA      Virginia      5
1004  Brandon Snyder NJ      New Jersey    6
3738  Joshua Jenkins OH      Ohio          6
3707  Kimberly Knight NE      Nebraska      6

```

```

OK
1003  Ashley Ford  VA      Virginia      5
1004  Brandon Snyder NJ      New Jersey    6
3738  Joshua Jenkins OH      Ohio          6
3707  Kimberly Knight NE      Nebraska      6
3618  Michael Adams TX      Texas         6
3768  Jane Sanchez VA      Virginia      7
4073  Patricia Wells IN      Indiana       7
3616  Nicole Ryan CA      California    7
3824  Terry Perez PA      Pennsylvania  7
3681  Diana Watkins CA      California    7
3860  Julie Hunter CT      Connecticut   7
3648  Norma Morris TX      Texas         7
4022  Jennifer Williams NC      North Carolina 8
3874  Catherine Mitchell NY      New York      8
3826  Gregory West VA      Virginia      8
3823  Cheryl Peterson TX      Texas         8
4051  Phillip Green MN      Minnesota     8
3950  Johnny Ray NY      New York      8
3929  Joshua Rice IL      Illinois      8
3789  Angela Cook VA      Virginia      8
61  Christine Mills MA      Massachusetts 8
3967  Juan Gibson NY      New York      8
3937  Ruby Anderson ME      Maine         8

```

11. Display customers (id, name, city, state) who have no late payments and the total number of on-time payments. Order the output by the number of on-time payments.

```

hive> SELECT c_ID, c_name, c_city, c_state, COUNT(c_ID) as total FROM customers JOIN payments ON c_ID = p_ID WHERE p_late = 'FALSE' GROUP BY c_ID, c_name, c_city, c_state;
Query ID = hadoop_20200509155954_e5c5014e-6f1d-44d6-95b8-eb2fb4c7cd78
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1589037740563_0004)
-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 3 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 03/03 [=====] 100% ELAPSED TIME: 5.63 s
-----
OK
2  Heather Burns FL      Florida 19
3  Jessica White IL      Illinois 14
32 Bonnie Cox CA      California 15
33 Samuel Oliver TN      Indiana 17

```

```

Reducer 3 ..... container      SUCCEEDED      1      1      0      0      0      0
-----
VERTICES: 03/03 [=====>>] 100%  ELAPSED TIME: 5.63 s
-----
OK
2    Heather Burns  FL      Florida 19
3    Jessica White  IL      Illinois 14
32   Bonnie Cox    CA      California 15
33   Samuel Oliver  IN      Indiana 17
34   Wanda Rice    CT      Connecticut 12
60   Chris Jacobs  MN      Minnesota 12
61   Christine Mills MA      Massachusetts 10
62   Bonnie Oliver  OH      Ohio 16
63   Amanda Hansen  IL      Illinois 10
64   Patricia Morrison OK      Oklahoma 8
65   David Cruz    MI      Michigan 18
66   Terry Taylor   OH      Ohio 17
67   Ralph Carter   CA      California 13
68   Virginia Marshall MA      Massachusetts 11
69   Martin Peters  IA      Iowa 16
70   Andrew White   IL      Illinois 16
71   Deborah Armstrong CT      Connecticut 9
72   Peter Fisher   MN      Minnesota 9
73   Carlos Bishop  CO      Colorado 14
74   Bonnie Frazier OH      Ohio 18
75   Kathy Johnston IL      Illinois 13
76   Melissa Griffin NJ      New Jersey 9
77   Christine Olson IL      Illinois 17

```

12. Display customers (id, name, city, state) who have no late payments and use the Internet for banking.

```

hive> SELECT c_ID, c_name, c_street, c_city, c_state FROM customers JOIN payments ON c_ID = p_ID WHERE p_late = 'FALSE' AND c_internet = 'TRUE' GROUP BY c_ID, c_
name, c_street, c_city, c_state;
Query ID = hadoop_20200509160059_56fc9ebb-1b5e-4e4a-9b3b-5cc04880b7aa
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1589037740563_0004)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1      1      0      0      0      0
Map 2 ..... container      SUCCEEDED      1      1      0      0      0      0
Reducer 3 ..... container      SUCCEEDED      1      1      0      0      0      0
-----
VERTICES: 03/03 [=====>>] 100%  ELAPSED TIME: 5.55 s
-----
OK
2    Heather Burns  6 Arrowood Court  FL      Florida
3    Jessica White  7534 Glendale Junction  IL      Illinois
32   Bonnie Cox    671 Merry Trail CA      California
34   Wanda Rice    9 New Castle Avenue  CT      Connecticut
62   Bonnie Oliver  20 Pearson Hill OH      Ohio
63   Amanda Hansen  61 Magdeline Terrace  IL      Illinois
64   Patricia Morrison 3174 Texas Alley  OK      Oklahoma
65   David Cruz    81 Comanche Way MI      Michigan
68   Virginia Marshall 306 Susan Terrace  MA      Massachusetts
72   Peter Fisher   2 Rowland Drive MN      Minnesota
73   Carlos Bishop  21011 Nelson Plaza  CO      Colorado
75   Kathy Johnston 37 Jackson Avenue  IL      Illinois

```

```

OK
2   Heather Burns  6 Arrowood Court      FL      Florida
3   Jessica White  7534 Glendale Junction IL      Illinois
32  Bonnie Cox     671 Merry Trail CA     California
34  Wanda Rice     9 New Castle Avenue   CT      Connecticut
62  Bonnie Oliver  20 Pearson Hill OH     Ohio
63  Amanda Hansen  61 Magdeline Terrace  IL      Illinois
64  Patricia Morrison 3174 Texas Alley      OK      Oklahoma
65  David Cruz     81 Comanche Way MI     Michigan
68  Virginia Marshall 306 Susan Terrace     MA      Massachusetts
72  Peter Fisher   2 Rowland Drive MN     Minnesota
73  Carlos Bishop  21011 Nelson Plaza    CO      Colorado
75  Kathy Johnston 37 Jackson Avenue     IL      Illinois
76  Melissa Griffin 40325 Stuart Circle   NJ      New Jersey
77  Christine Olson 8872 Charing Cross Pass IL      Illinois
78  Matthew West   60791 Rusk Pass NY     New York
79  Pamela Dean    52798 Morning Way     IL      Illinois
81  Stephen Banks  0903 Mendota Lane     MA      Massachusetts
82  Jeremy Lynch   752 Melody Street     MI      Michigan
86  Maria Bowman   0 Lunder Court FL      Florida
87  Earl Wallace   87 Del Mar Avenue     IL      Illinois
88  Pamela Daniels 3 Northland Street    NY      New York
89  Andrew Howard  76 Lake View Hill     CA      California
90  Joshua Freeman 30 Pennsylvania Terrace TX      Texas
91  Julia James    4271 Morning Drive    IN      Indiana
92  Ryan Howard    567 Spaight Pass     MN      Minnesota
99  Rachel Cooper  587 Loftsgordon Road  NJ      New Jersey
100 Rose Boyd      01 Arizona Terrace    TX      Texas
102 Virginia Hawkins 4 Mayfield Circle     MA      Massachusetts
104 Raymond Long   36 Sutteridge Parkway NJ      New Jersey
106 Paul Clark     3758 John Wall Street NY      New York
110 Cynthia Meyer  83247 Oak Valley Center FL      Florida

```

**Note:** With this homework are included also a List of commands 'CommandsToExecute\_HW5.txt' and a total log file 'TOT\_LOG\_HW5.txt' of the PuTTY terminal inputs and outputs, the output of some table was to bigger so it erased all the entire terminal, which forced me to copy it with fractions which sometimes it might not correspond with the exercises, to do this just copy and search the query into the log file, if the two files are not included or found , please follow the following google drive link

<https://drive.google.com/drive/folders/1RNbc8e175SDFY57P0cw2FQQYltvmREOr?usp=sharing> .