Project 1 -
Hive Query Project

By Moses Lin

Performing setup

Dataset:

- Bev_BranchA.txt
- Bev_BranchB.txt
- Bev_BranchC.txt
- Bev_ConscountA.txt
- Bev_ConscountB.txt
- Bev_ConscountC.txt
-

Commands:

- hdfs dfs -mkdir file_name /path/to/direc
- hdfs dfs -put file_name /path/to/direc
- CREATE DATABASE hadoopprojectone;
- USE hadoopprojectone;
- CREATE TABLE bev# (beverage STRING, branch STRING) row format delimited fields terminated by ',';
- LOAD DATA INPATH '/user/hive/projectone/Bev_Branch#.txt' INTO TABLE bev#;

```
/mnt/c/Users/mlin2$ hdfs dfs -mkdir /user/hive/projectone
```

1. Directory was made in HDFS

```
/mnt/c/Users/mlin2$ hdfs dfs -put Bev_BranchA.txt /user/hive/projectone
/mnt/c/Users/mlin2$ hdfs dfs -put Bev_BranchB.txt /user/hive/projectone
/mnt/c/Users/mlin2$ hdfs dfs -put Bev_BranchC.txt /user/hive/projectone
/mnt/c/Users/mlin2$ hdfs dfs -put Bev_ConscountA.txt /user/hive/projectone
/mnt/c/Users/mlin2$ hdfs dfs -put Bev_ConscountB.txt /user/hive/projectone
/mnt/c/Users/mlin2$ hdfs dfs -put Bev_ConscountC.txt /user/hive/projectone
```

2. Files were downloaded into local folder.
3. Files were placed into HDFS.

```
/mnt/c/Users/mlin2$ hdfs dfs -ls /user/hive/projectone/
```

4. Check HDFS for all the files.

```
Found 6 items
-rw-r--r--   1 sadcat supergroup       2168 2021-05-11 18:23 /user/hive/projectone/Bev_BranchA.txt
-rw-r--r--   1 sadcat supergroup       4335 2021-05-11 18:23 /user/hive/projectone/Bev_BranchB.txt
-rw-r--r--   1 sadcat supergroup       6494 2021-05-11 18:24 /user/hive/projectone/Bev_BranchC.txt
-rw-r--r--   1 sadcat supergroup      17592 2021-05-11 18:33 /user/hive/projectone/Bev_ConscountA.txt
-rw-r--r--   1 sadcat supergroup      35191 2021-05-11 18:24 /user/hive/projectone/Bev_ConscountB.txt
-rw-r--r--   1 sadcat supergroup      52827 2021-05-11 18:24 /user/hive/projectone/Bev_ConscountC.txt
```

5. Load data into HIVE.

```
+-------------+
|  tab_name   |
+-------------+
| bevbrancha  |
| bevbranchb  |
| bevbranchc  |
| bevcounta   |
| bevcountb   |
| bevcountc   |
+-------------+
```

Problem Scenario 1:

Q1. What is the total number of consumers for Branch1?
Q2. What is the number of consumers for the Branch2?

Approach Q1:

- Multiple tables are utilized.
- Create Table 1 that selects all types of beverages from File A, B, and C in Branch 1.
- Create Table 2 that inserts # of beverages from File A, B, and C that are in Table 1.
- Retrieve count of all beverages in Table 2.

Approach Q2:

- Repeat approach for Q1 but for Branch 2 instead.

Prob 1 - Q1: Part 1

Commands:

1. CREATE TABLE IF NOT EXISTS branch1table AS SELECT * FROM bevbrancha WHERE branch = 'Branch1';

2. INSERT INTO TABLE branch1table SELECT * FROM bevbranchb WHERE branch = 'Branch1';

3. INSERT INTO TABLE branch1table SELECT * FROM bevbranchc WHERE branch = 'Branch1';

Notes:

- The number of entries for branch1table did not change when inserting from different branches. It remains 20. This means there are no Branch1 entries from data b and data c.

branch1table.beverage	branch1table.branch
SMALL_Espresso	Branch1
Special_Coffee	Branch1
Double_Espresso	Branch1
MED_MOCHA	Branch1
LARGE_cappuccino	Branch1
Triple_MOCHA	Branch1
Mild_LATTE	Branch1
ICY_Espresso	Branch1
Cold_LATTE	Branch1
SMALL_MOCHA	Branch1
Special_cappuccino	Branch1
Double_MOCHA	Branch1
MED_LATTE	Branch1
LARGE_Espresso	Branch1
Triple_LATTE	Branch1
Mild_Lite	Branch1
ICY_MOCHA	Branch1
Cold_Lite	Branch1
SMALL_LATTE	Branch1
Special_Espresso	Branch1

20 rows selected (0.089 seconds)

Prob 1 - Q1: Part 2

Commands:

1. CREATE TABLE IF NOT EXISTS branch1counttable (beverage STRING, count INT);

2. INSERT INTO TABLE branch1counttable SELECT bevcounta.beverage,SUM(bevcounta.count) FROM branch1table JOIN bevcounta ON (branch1table.beverage = bevcounta.beverage) GROUP BY bevcounta.beverage;

   ## Repeat step 2 but replace bevcounta with bevcountb and bevcount c

3. SELECT SUM(count) FROM branch1counttable;

Answer:

1115974 Consumers

```
+-----------------------------+-----------------------------+
| branch1counttable.beverage  | branch1counttable.count     |
+-----------------------------+-----------------------------+
| Cold_LATTE                  | 7500                        |
| Cold_Lite                   | 9636                        |
| Double_Espresso             | 8036                        |
| Double_MOCHA                | 8716                        |
| ICY_Espresso                | 7804                        |
| ICY_MOCHA                   | 10304                       |
| LARGE_Espresso              | 9956                        |
| LARGE_cappuccino            | 18704                       |
| MED_LATTE                   | 7824                        |
| MED_MOCHA                   | 9692                        |
| Mild_LATTE                  | 8024                        |
| Mild_Lite                   | 8144                        |
| SMALL_Espresso              | 8532                        |
| SMALL_LATTE                 | 7596                        |
| SMALL_MOCHA                 | 8149                        |
| Special_Coffee              | 8452                        |
| Special_Espresso            | 7992                        |
| Special_cappuccino          | 19704                       |
| Triple_LATTE                | 9736                        |
| Triple_MOCHA                | 8244                        |
+-----------------------------+-----------------------------+
20 rows selected (0.071 seconds)
| Cold_LATTE                  | 14684                       |
| Cold_Lite                   | 15345                       |
| Double_Espresso             | 18024                       |
| Double_MOCHA                | 15512                       |
| ICY_Espresso                | 17788                       |
| ICY_MOCHA                   | 15282                       |
| LARGE_Espresso              | 16342                       |
| LARGE_cappuccino            | 34729                       |
| MED_LATTE                   | 15763                       |
| MED_MOCHA                   | 15120                       |
| Mild_LATTE                  | 16912                       |
| Mild_Lite                   | 15944                       |
| SMALL_Espresso              | 14550                       |
| SMALL_LATTE                 | 15100                       |
| SMALL_MOCHA                 | 15260                       |
| Special_Coffee              | 17184                       |
| Special_Espresso            | 14601                       |
| Special_cappuccino          | 32126                       |
| Triple_LATTE                | 14305                       |
| Triple_MOCHA                | 18308                       |
+-----------------------------+-----------------------------+
40 rows selected (0.082 seconds)
| Cold_LATTE                  | 25740                       |
| Cold_Lite                   | 27024                       |
| Double_Espresso             | 28318                       |
| Double_MOCHA                | 26000                       |
| ICY_Espresso                | 25228                       |
| ICY_MOCHA                   | 23896                       |
| LARGE_Espresso              | 27526                       |
| LARGE_cappuccino            | 50461                       |
| MED_LATTE                   | 23624                       |
| MED_MOCHA                   | 26792                       |
| Mild_LATTE                  | 25484                       |
| Mild_Lite                   | 22746                       |
| SMALL_Espresso              | 27120                       |
| SMALL_LATTE                 | 23896                       |
| SMALL_MOCHA                 | 25828                       |
| Special_Coffee              | 28572                       |
| Special_Espresso            | 25690                       |
| Special_cappuccino          | 56333                       |
| Triple_LATTE                | 25980                       |
| Triple_MOCHA                | 24092                       |
+-----------------------------+-----------------------------+
60 rows selected (0.082 seconds)
```

```
+----------+
|   _c0    |
+----------+
| 1115974  |
+----------+
1 row selected (18.912 seconds)
```

Prob 1 - Q2: Part 1

Commands:

1. CREATE TABLE IF NOT EXISTS branch2table AS SELECT * FROM bevbrancha WHERE branch = 'Branch2';

2. INSERT INTO TABLE branch2table SELECT * FROM bevbranchb WHERE branch = 'Branch2';

3. INSERT INTO TABLE branch2table SELECT * FROM bevbranchc WHERE branch = 'Branch2';

Notes:

- The number of entries for branch2table does not change when inserting from data b. However, data c added 60 entries. This means there are no branch 2 entries for data b, but duplicate entries for data c.

MED_Espresso	Branch2
Double_Espresso	Branch2
ICY_Coffee	Branch2
SMALL_LATTE	Branch2
Cold_cappuccino	Branch2
LARGE_Espresso	Branch2
Mild_cappuccino	Branch2
Triple_cappuccino	Branch2
Special_LATTE	Branch2
MED_MOCHA	Branch2
Double_MOCHA	Branch2
ICY_cappuccino	Branch2
SMALL_Lite	Branch2
Cold_Coffee	Branch2
LARGE_MOCHA	Branch2
Mild_Espresso	Branch2
Triple_Espresso	Branch2
Special_Lite	Branch2
MED_LATTE	Branch2
Double_LATTE	Branch2
ICY_Espresso	Branch2
SMALL_cappuccino	Branch2
Cold_cappuccino	Branch2
LARGE_LATTE	Branch2
Mild_MOCHA	Branch2
Triple_MOCHA	Branch2
Special_cappuccino	Branch2
MED_Lite	Branch2
Double_Lite	Branch2
ICY_MOCHA	Branch2
SMALL_Coffee	Branch2

80 rows selected (0.099 seconds)

Prob 1 - Q2: Part 2

Commands:

1. CREATE TABLE IF NOT EXISTS branch2counttable (beverage STRING, count INT);

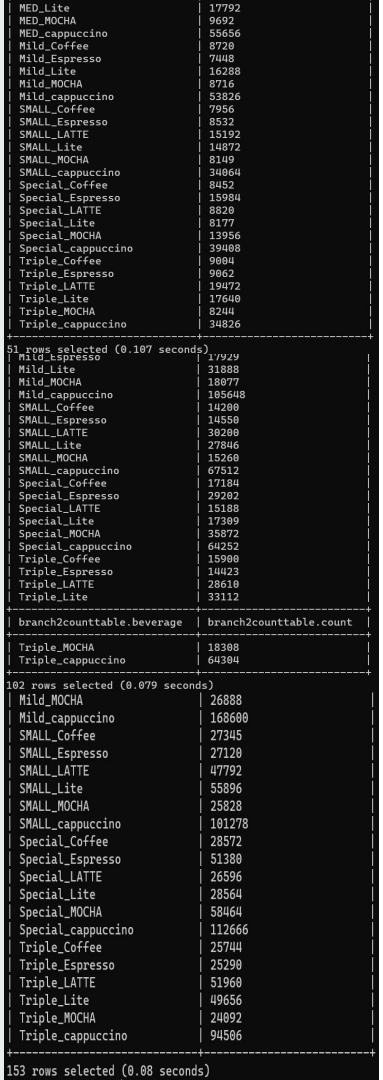
2. INSERT INTO TABLE branch2counttable SELECT bevcounta.beverage,SUM(bevcounta.count) FROM branch2table JOIN bevcounta ON (branch2table.beverage = bevcounta.beverage) GROUP BY bevcounta.beverage;

   ## Repeat step 2 but replace bevcounta with bevcountb and bevcount c

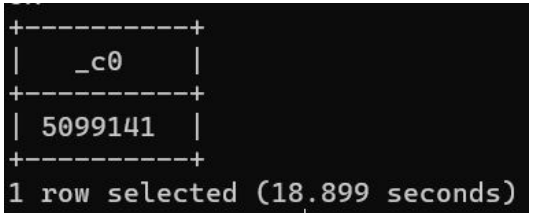
3. SELECT SUM(count) FROM branch2counttable;

Answer:

5099141 Consumers



```
| MED_Lite                | 17792 |
| MED_MOCHA               | 9692  |
| MED_cappuccino          | 55656 |
| Mild_Coffee             | 8720  |
| Mild_Espresso           | 7448  |
| Mild_Lite               | 16288 |
| Mild_MOCHA              | 8716  |
| Mild_cappuccino         | 53826 |
| SMALL_Coffee            | 7956  |
| SMALL_Espresso          | 8532  |
| SMALL_LATTE             | 15192 |
| SMALL_Lite              | 14872 |
| SMALL_MOCHA             | 8149  |
| SMALL_cappuccino        | 34064 |
| Special_Coffee          | 8452  |
| Special_Espresso        | 15984 |
| Special_LATTE           | 8820  |
| Special_Lite            | 8177  |
| Special_MOCHA           | 13956 |
| Special_cappuccino      | 39408 |
| Triple_Coffee           | 9004  |
| Triple_Espresso         | 9062  |
| Triple_LATTE            | 19472 |
| Triple_Lite             | 17640 |
| Triple_MOCHA            | 8244  |
| Triple_cappuccino       | 34826 |
+-------------------------+-------+
51 rows selected (0.107 seconds)
| Mild_Espresso           | 17929 |
| Mild_Lite               | 31888 |
| Mild_MOCHA              | 18077 |
| Mild_cappuccino         | 105648|
| SMALL_Coffee            | 14200 |
| SMALL_Espresso          | 14550 |
| SMALL_LATTE             | 30200 |
| SMALL_Lite              | 27846 |
| SMALL_MOCHA             | 15260 |
| SMALL_cappuccino        | 67512 |
| Special_Coffee          | 17184 |
| Special_Espresso        | 29202 |
| Special_LATTE           | 15188 |
| Special_Lite            | 17309 |
| Special_MOCHA           | 35872 |
| Special_cappuccino      | 64252 |
| Triple_Coffee           | 15900 |
| Triple_Espresso         | 14423 |
| Triple_LATTE            | 28610 |
| Triple_Lite             | 33112 |
+-------------------------+-------+
| branch2counttable.beverage | branch2counttable.count |
+----------------------------+-------------------------+
| Triple_MOCHA               | 18308                   |
| Triple_cappuccino          | 64304                   |
+----------------------------+-------------------------+
102 rows selected (0.079 seconds)
| Mild_MOCHA              | 26888 |
| Mild_cappuccino         | 168600|
| SMALL_Coffee            | 27345 |
| SMALL_Espresso          | 27120 |
| SMALL_LATTE             | 47792 |
| SMALL_Lite              | 55896 |
| SMALL_MOCHA             | 25828 |
| SMALL_cappuccino        | 101278|
| Special_Coffee          | 28572 |
| Special_Espresso        | 51380 |
| Special_LATTE           | 26596 |
| Special_Lite            | 28564 |
| Special_MOCHA           | 58464 |
| Special_cappuccino      | 112666|
| Triple_Coffee           | 25744 |
| Triple_Espresso         | 25290 |
| Triple_LATTE            | 51960 |
| Triple_Lite             | 49656 |
| Triple_MOCHA            | 24092 |
| Triple_cappuccino       | 94506 |
+-------------------------+-------+
153 rows selected (0.08 seconds)
```



```
+-----------+
|    _c0    |
+-----------+
| 5099141   |
+-----------+
1 row selected (18.899 seconds)
```

Problem Scenario 2:

Q1. What is the most consumed beverage on Branch1?
Q2. What is the least consumed beverage on Branch2?

Approach Q1:

- Sum the count of all beverages of the same name in beverage count table (branch1counttable or branch2counttable).
- Set select query to descending by count and show only the first entry.

Approach Q2:

- Repeat approach for Q1 but for Branch 2 and sort by ascending instead.

Prob 2 - Q1:

Commands:

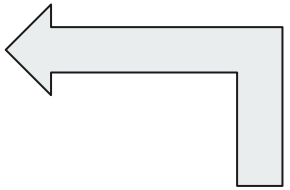
1. SELECT beverage,SUM(count) COUNT FROM branch1counttable GROUP BY beverage ORDER BY count DESC LIMIT 1;

Answer:

- **Special Cappuccino with 108163 consumers.**

beverage	count
Special_cappuccino	108163
LARGE_cappuccino	103894
Double_Espresso	54378
Special_Coffee	54208
LARGE_Espresso	53824
Cold_Lite	52005
MED_MOCHA	51604
ICY_Espresso	50820
Triple_MOCHA	50644
Mild_LATTE	50420
Double_MOCHA	50228
SMALL_Espresso	50202
Triple_LATTE	50021
ICY_MOCHA	49482
SMALL_MOCHA	49237
Special_Espresso	48283
Cold_LATTE	47924
MED_LATTE	47211
Mild_Lite	46834
SMALL_LATTE	46592

20 rows selected (45.029 seconds)



beverage	count
Special_cappuccino	108163

1 row selected (48.207 seconds)

Prob 2 - Q2:

Commands:

1. SELECT beverage,SUM(count) COUNT FROM branch2counttable GROUP BY beverage ORDER BY count ASC LIMIT 1;

Answer:

- Cold Mocha with only 47524 consumers.

beverage	count
Cold_MOCHA	47524
Cold_LATTE	47924
ICY_Lite	48220
MED_Espresso	48413
Mild_Espresso	48462
ICY_Coffee	48555
Triple_Espresso	48775
MED_Coffee	48994
SMALL_MOCHA	49237
SMALL_Coffee	49501
SMALL_Espresso	50202
Double_MOCHA	50228
Special_LATTE	50604
Triple_MOCHA	50644
Triple_Coffee	50648
ICY_Espresso	50820
MED_MOCHA	51604
LARGE_Coffee	51647
Double_Coffee	51694
Mild_Coffee	51880
Cold_Coffee	51932
Mild_MOCHA	53681
Special_Lite	54050
Special_Coffee	54208
Double_Espresso	54378
SMALL_LATTE	93184
Mild_Lite	93668
MED_LATTE	94422
Special_Espresso	96566
ICY_LATTE	98332
Double_Lite	98532
SMALL_Lite	98614
ICY_MOCHA	98964
Triple_LATTE	100042
Triple_Lite	100408
LARGE_MOCHA	101192
MED_Lite	103572
Cold_Lite	104010
LARGE_LATTE	104466
LARGE_Espresso	107648
Special_MOCHA	108292
Double_LATTE	112378
Triple_cappuccino	193636
Double_cappuccino	197620
SMALL_cappuccino	202854
ICY_cappuccino	204476
LARGE_cappuccino	207788
Special_cappuccino	216326
Cold_cappuccino	284943
MED_cappuccino	285309
Mild_cappuccino	328074

51 rows selected (43.362 seconds)

beverage	count
Cold_MOCHA	47524

1 row selected (48.202 seconds)

Problem Scenario 3:

Q1. What are the beverages available on Branch10, Branch8, and Branch1?
Q2. What are the common beverages available in Branch4, Branch7?

Approach Q1:

- Create a table that has beverages from Branch10, Branch8, and Branch1 from Data A, B, and C.
- Return unique beverages from Branch10, Branch8, Branch1.

Approach Q2:

- Same as Q1 but with Branch4 and Branch7 instead.

Prob 3 - Q1:

Commands:

1. CREATE TABLE IF NOT EXISTS branch10a8a1 AS SELECT beverage, branch FROM (
   SELECT * FROM bevbrancha WHERE branch = 'Branch10' OR branch = 'Branch8' OR branch = 'Branch1' UNION ALL
   SELECT * FROM bevbranchb WHERE branch = 'Branch10' OR branch = 'Branch8' OR branch = 'Branch1' UNION ALL
   SELECT * FROM bevbranchc WHERE branch = 'Branch10' OR branch = 'Branch8' OR branch = 'Branch1'
) unionResult;
2. SELECT DISTINCT beverage AS branch10bev FROM branch10a8a1 WHERE branch == 'Branch10' ORDER BY beverage;

Answer:

- See Beverage table to the right.

branch10a8a1.beverage	branch10a8a1.branch
Triple_Lite	Branch8
Special_cappuccino	Branch8
MED_Coffee	Branch8
Double_Coffee	Branch8
ICY_Lite	Branch8
SMALL_Espresso	Branch8
Cold_LATTE	Branch8
LARGE_Coffee	Branch8
Mild_cappuccino	Branch8
Triple_cappuccino	Branch8
Special_Espresso	Branch8
MED_cappuccino	Branch8
Double_cappuccino	Branch8
ICY_cappuccino	Branch8
SMALL_MOCHA	Branch8
Cold_Lite	Branch8
LARGE_cappuccino	Branch8
Mild_Coffee	Branch8
Triple_Coffee	Branch8
Special_MOCHA	Branch8
MED_Espresso	Branch8
Double_Espresso	Branch8
ICY_Coffee	Branch8
SMALL_LATTE	Branch8
Cold_cappuccino	Branch8
LARGE_Espresso	Branch8
Mild_cappuccino	Branch8
Triple_cappuccino	Branch8
Special_LATTE	Branch8
MED_MOCHA	Branch8
Double_MOCHA	Branch8
ICY_cappuccino	Branch8
SMALL_Lite	Branch8
Cold_Coffee	Branch8
LARGE_MOCHA	Branch8
Mild_Espresso	Branch8
Triple_Espresso	Branch8
Special_Lite	Branch8
MED_LATTE	Branch8
Double_LATTE	Branch8
SMALL_Espresso	Branch1
Special_Coffee	Branch1
Double_Espresso	Branch1
MED_MOCHA	Branch1
LARGE_cappuccino	Branch1
Triple_MOCHA	Branch1
Mild_LATTE	Branch1
ICY_Espresso	Branch1
Cold_LATTE	Branch1
SMALL_MOCHA	Branch1
Special_cappuccino	Branch1
Double_MOCHA	Branch1
MED_LATTE	Branch1
LARGE_Espresso	Branch1
Triple_LATTE	Branch1
Mild_Lite	Branch1
ICY_MOCHA	Branch1
Cold_Lite	Branch1
SMALL_LATTE	Branch1
Special_Espresso	Branch1

60 rows selected (0.071 seconds)

branch8bev
Cold_Coffee
Cold_LATTE
Cold_Lite
Cold_cappuccino
Double_Coffee
Double_Espresso
Double_LATTE
Double_MOCHA
Double_cappuccino
ICY_Coffee
ICY_Lite
ICY_cappuccino
LARGE_Coffee
LARGE_Espresso
LARGE_MOCHA
LARGE_cappuccino
MED_Coffee
MED_Espresso
MED_LATTE
MED_MOCHA
MED_cappuccino
Mild_Coffee
Mild_Espresso
Mild_cappuccino
SMALL_Espresso
SMALL_LATTE
SMALL_Lite
SMALL_MOCHA
Special_Espresso
Special_LATTE
Special_Lite
Special_MOCHA
Special_cappuccino
Triple_Coffee
Triple_Espresso
Triple_Lite
Triple_cappuccino

37 rows selected (42.884 seconds)

branch1bev
Cold_LATTE
Cold_Lite
Double_Espresso
Double_MOCHA
ICY_Espresso
ICY_MOCHA
LARGE_Espresso
LARGE_cappuccino
MED_LATTE
MED_MOCHA
Mild_LATTE
Mild_Lite
SMALL_Espresso
SMALL_LATTE
SMALL_MOCHA
Special_Coffee
Special_Espresso
Special_cappuccino
Triple_LATTE
Triple_MOCHA

20 rows selected (41.318 seconds)

Prob 3 - Q2:

Commands:

1. CREATE TABLE IF NOT EXISTS branch4a7 AS SELECT beverage, branch FROM (
   SELECT * FROM bevbrancha WHERE branch = 'Branch4' OR branch = 'Branch7' UNION ALL
   SELECT * FROM bevbranchb WHERE branch = 'Branch4' OR branch = 'Branch7' UNION ALL
   SELECT * FROM bevbranchc WHERE branch = 'Branch4' OR branch = 'Branch7'
) unionResult ;

2. SELECT DISTINCT beverage AS branch4bev FROM branch4a7 WHERE branch == 'Branch4' ORDER BY beverage;

Answer:

- See Beverage table to the right.

branch4a7.beverage	branch4a7.branch
Special_LATTE	Branch4
Double_Coffee	Branch7
MED_MOCHA	Branch4
ICY_Lite	Branch7
Double_MOCHA	Branch4
SMALL_Espresso	Branch7
ICY_cappuccino	Branch4
Cold_LATTE	Branch7
SMALL_Lite	Branch4
LARGE_Coffee	Branch7
Cold_Coffee	Branch4
Mild_cappuccino	Branch7
LARGE_MOCHA	Branch4
Triple_cappuccino	Branch7
Mild_Espresso	Branch4
Special_Espresso	Branch7
Triple_Espresso	Branch4
MED_cappuccino	Branch7
Special_Lite	Branch4
Double_cappuccino	Branch7
SMALL_Coffee	Branch7
Cold_Espresso	Branch7
LARGE_Lite	Branch7
Mild_LATTE	Branch7
Triple_LATTE	Branch7
Special_Coffee	Branch7
MED_cappuccino	Branch7
Double_cappuccino	Branch7
ICY_LATTE	Branch7
SMALL_cappuccino	Branch7
Cold_MOCHA	Branch7
LARGE_cappuccino	Branch7
Mild_Lite	Branch7
Triple_Lite	Branch7
Special_cappuccino	Branch7
MED_Coffee	Branch7
Double_Coffee	Branch7
ICY_Lite	Branch7
SMALL_Espresso	Branch7
Cold_LATTE	Branch7
LARGE_Coffee	Branch7
Mild_cappuccino	Branch7
Triple_cappuccino	Branch7
Special_Espresso	Branch7
MED_cappuccino	Branch7
Double_cappuccino	Branch7
ICY_cappuccino	Branch7
SMALL_MOCHA	Branch7
Cold_Lite	Branch7
LARGE_cappuccino	Branch7
Mild_Coffee	Branch7
Triple_Coffee	Branch7
Special_MOCHA	Branch7
MED_Espresso	Branch7
Double_Espresso	Branch7
ICY_Coffee	Branch7
SMALL_LATTE	Branch7
Cold_cappuccino	Branch7
LARGE_Espresso	Branch7
Mild_cappuccino	Branch7

160 rows selected (0.077 seconds)

branch4bev
Cold_Coffee
Cold_Espresso
Cold_LATTE
Cold_Lite
Cold_MOCHA
Cold_cappuccino
Double_Coffee
Double_Espresso
Double_Lite
Double_MOCHA
Double_cappuccino
ICY_Coffee
ICY_LATTE
ICY_Lite
ICY_MOCHA
ICY_cappuccino
LARGE_Coffee
LARGE_Espresso
LARGE_LATTE
LARGE_Lite
LARGE_MOCHA
LARGE_cappuccino
MED_Coffee
MED_Espresso
MED_Lite
MED_MOCHA
MED_cappuccino
Mild_Coffee
Mild_Espresso
Mild_LATTE
Mild_Lite
Mild_MOCHA
Mild_cappuccino
SMALL_Coffee
SMALL_Espresso
SMALL_LATTE
SMALL_Lite
SMALL_MOCHA
SMALL_cappuccino
Special_Coffee
Special_Espresso
Special_LATTE
Special_Lite
Special_MOCHA
Special_cappuccino
Triple_Coffee
Triple_Espresso
Triple_LATTE
Triple_Lite
Triple_MOCHA
Triple_cappuccino

51 rows selected (50.162 seconds)

branch7bev
Cold_Coffee
Cold_Espresso
Cold_LATTE
Cold_Lite
Cold_MOCHA
Cold_cappuccino
Double_Coffee
Double_Espresso
Double_LATTE
Double_Lite
Double_MOCHA
Double_cappuccino
ICY_Coffee
ICY_Espresso
ICY_LATTE
ICY_Lite
ICY_MOCHA
ICY_cappuccino
LARGE_Coffee
LARGE_Espresso
LARGE_LATTE
LARGE_Lite
LARGE_MOCHA
LARGE_cappuccino
MED_Coffee
MED_Espresso
MED_LATTE
MED_Lite
MED_MOCHA
MED_cappuccino
Mild_Coffee
Mild_Espresso
Mild_LATTE
Mild_Lite
Mild_MOCHA
Mild_cappuccino
SMALL_Coffee
SMALL_Espresso
SMALL_LATTE
SMALL_Lite
SMALL_MOCHA
SMALL_cappuccino
Special_Coffee
Special_Espresso
Special_LATTE
Special_Lite
Special_MOCHA
Special_cappuccino
Triple_Coffee
Triple_Espresso
Triple_LATTE
Triple_Lite
Triple_MOCHA
Triple_cappuccino

54 rows selected (43.264 seconds)

Problem Scenario 4:

Q1. Create a partition, index, and view for the Scenario 3.

Approach Q1:

- Create a partition based on branch
- Create an index based based on branch
- Create a view

Prob 4 - Q1: Part 1

Commands:

1. CREATE TABLE bevbranchpart (beverage STRING)
   PARTITIONED BY (branch STRING);

2. FROM branch10a8a1 br
   INSERT OVERWRITE TABLE bevbranchpart
   PARTITION(branch)
   SELECT br.beverage, br.branch
   DISTRIBUTE BY branch;

Note:

Command used to produce the two tables:

- SELECT * FROM bevbranchpart WHERE bevbranchpart.branch = "Branch1";
- SELECT * FROM bevbranchpart WHERE bevbranchpart.branch = "Branch8";

```
hdfs dfs -ls /user/hive/warehouse/hadoopprojectone.db/bevbranchpart

0 2021-05-13 09:41 /user/hive/warehouse/hadoopprojectone.db/bevbranchpart/branch=Branch1
0 2021-05-13 09:41 /user/hive/warehouse/hadoopprojectone.db/bevbranchpart/branch=Branch8
```

bevbranchpart.beverage	bevbranchpart.branch
Special_Espresso	Branch1
SMALL_LATTE	Branch1
Cold_Lite	Branch1
ICY_MOCHA	Branch1
Mild_Lite	Branch1
Triple_LATTE	Branch1
LARGE_Espresso	Branch1
MED_LATTE	Branch1
Double_MOCHA	Branch1
Special_cappuccino	Branch1
SMALL_MOCHA	Branch1
Cold_LATTE	Branch1
ICY_Espresso	Branch1
Mild_LATTE	Branch1
Triple_MOCHA	Branch1
LARGE_cappuccino	Branch1
MED_MOCHA	Branch1
Double_Espresso	Branch1
Special_Coffee	Branch1
SMALL_Espresso	Branch1

20 rows selected (0.142 seconds)

bevbranchpart.beverage	bevbranchpart.branch
Double_LATTE	Branch8
MED_LATTE	Branch8
Special_Lite	Branch8
Triple_Espresso	Branch8
Mild_Espresso	Branch8
LARGE_MOCHA	Branch8
Cold_Coffee	Branch8
SMALL_Lite	Branch8
ICY_cappuccino	Branch8
Double_MOCHA	Branch8
MED_MOCHA	Branch8
Special_LATTE	Branch8
Triple_cappuccino	Branch8
Mild_cappuccino	Branch8
LARGE_Espresso	Branch8
Cold_cappuccino	Branch8
SMALL_LATTE	Branch8
ICY_Coffee	Branch8
Double_Espresso	Branch8
MED_Espresso	Branch8
Special_MOCHA	Branch8
Triple_Coffee	Branch8
Mild_Coffee	Branch8
LARGE_cappuccino	Branch8
Cold_Lite	Branch8
SMALL_MOCHA	Branch8
ICY_cappuccino	Branch8
Double_cappuccino	Branch8
MED_cappuccino	Branch8
Special_Espresso	Branch8
Triple_cappuccino	Branch8
Mild_cappuccino	Branch8
LARGE_Coffee	Branch8
Cold_LATTE	Branch8
SMALL_Espresso	Branch8
ICY_Lite	Branch8
Double_Coffee	Branch8
MED_Coffee	Branch8
Special_cappuccino	Branch8
Triple_Lite	Branch8

40 rows selected (0.101 seconds)

Prob 4 - Q1: Part 2

Commands:

1. CREATE TABLE bevbranchpart2 (beverage STRING) PARTITIONED BY (branch STRING);

2. FROM branch4a7 br2
   INSERT OVERWRITE TABLE bevbranchpart2
   PARTITION(branch)
   SELECT br2.beverage, br2.branch
   DISTRIBUTE BY branch;

3. hdfs dfs -ls /user/hive/warehouse/hadoopprojectone.db/bevbranchpart2

```
hdfs dfs -ls /user/hive/warehouse/hadoopprojectone.db/bevbranchpart2

0 2021-05-13 09:57 /user/hive/warehouse/hadoopprojectone.db/bevbranchpart2/branch=Branch4
0 2021-05-13 09:57 /user/hive/warehouse/hadoopprojectone.db/bevbranchpart2/branch=Branch7
```

Prob 4 - Q2

Commands:

1. CREATE INDEX bevbranchindex ON TABLE branch10a8a1(beverage) AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler' WITH DEFERRED REBUILD;

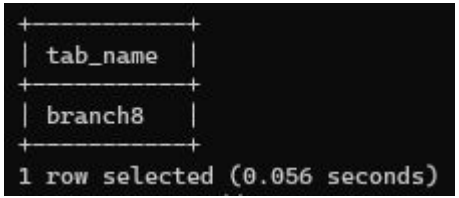
2. SHOW INDEX ON branch10a8a1;

Prob 4 - Q3

Commands:

1. CREATE VIEW branch8 AS
   SELECT DISTINCT beverage AS branch8bev
   FROM branch10a8a1 WHERE branch ==
   'Branch8' ORDER BY beverage;

2. SHOW VIEWS;



Problem Scenario 5:

Q1. Alter the table properties to add "note","comment"

Approach Q1:

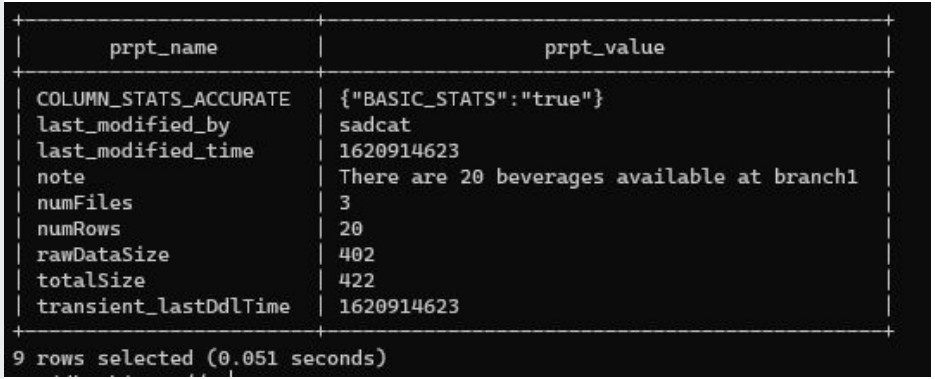
- Add a comment to a table
- Show the table properties.

Prob 5 - Q1:

Commands:

1. ALTER TABLE branch1table SET tblproperties ("note"="There are 20 beverages available at branch1");

2. SHOW tblproperties branch1table;



prpt_name	prpt_value
COLUMN_STATS_ACCURATE	{"BASIC_STATS":"true"}
last_modified_by	sadcat
last_modified_time	1620914623
note	There are 20 beverages available at branch1
numFiles	3
numRows	20
rawDataSize	402
totalSize	422
transient_lastDdlTime	1620914623

9 rows selected (0.051 seconds)

Problem Scenario 6:

Q1. Remove the row 5 from the output of Scenario 1

Approach Q1:

- Organize table in a specific format
- Introduce index to format
- Sum everything but a specific row as there is no minus operator in Hive.

Prob 6 - Q1:

Commands:

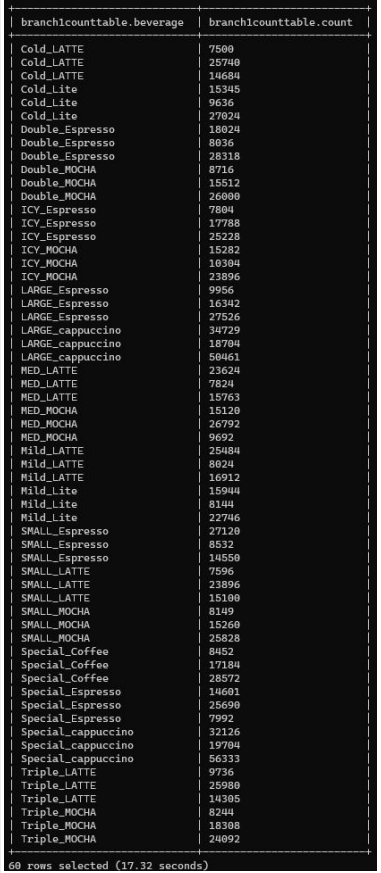
1. SELECT * FROM branch1counttable ORDER BY beverage;

2. SELECT count, ROW_NUMBER() OVER (ORDER BY beverage) AS row_no FROM branch1counttable;

3. SELECT SUM(count) FROM (SELECT count, ROW_NUMBER() OVER (ORDER BY beverage) AS row_no FROM branch1counttable) res where res.row_no != 5;

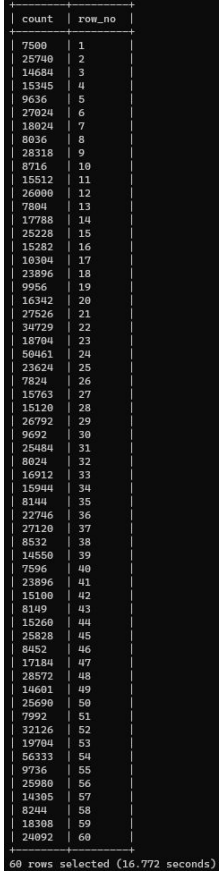
Answer:

- 1115974 - 9636 = 1106338
  When sorting by beverage.



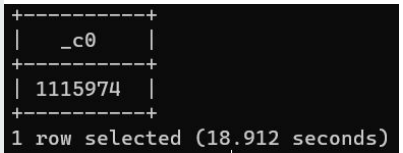
branch1counttable.beverage	branch1counttable.count
Cold_LATTE	7500
Cold_LATTE	25740
Cold_LATTE	14684
Cold_Lite	15345
Cold_Lite	9636
Cold_Lite	27024
Double_Espresso	18024
Double_Espresso	8036
Double_Espresso	28318
Double_MOCHA	8716
Double_MOCHA	15512
Double_MOCHA	26000
ICY_Espresso	7804
ICY_Espresso	17788
ICY_Espresso	25228
ICY_MOCHA	15282
ICY_MOCHA	10304
ICY_MOCHA	23896
LARGE_Espresso	9956
LARGE_Espresso	16342
LARGE_Espresso	27526
LARGE_cappuccino	34729
LARGE_cappuccino	18704
LARGE_cappuccino	50461
MED_LATTE	23624
MED_LATTE	7824
MED_LATTE	15763
MED_MOCHA	15120
MED_MOCHA	26792
MED_MOCHA	9692
Mild_LATTE	25484
Mild_LATTE	8024
Mild_LATTE	16912
Mild_Lite	15944
Mild_Lite	8144
Mild_Lite	22746
SMALL_Espresso	27120
SMALL_Espresso	8532
SMALL_Espresso	14550
SMALL_LATTE	7596
SMALL_LATTE	23896
SMALL_LATTE	15100
SMALL_MOCHA	8149
SMALL_MOCHA	15260
SMALL_MOCHA	25828
Special_Coffee	8452
Special_Coffee	17184
Special_Coffee	28572
Special_Espresso	14601
Special_Espresso	25690
Special_Espresso	7992
Special_cappuccino	32126
Special_cappuccino	19704
Special_cappuccino	56333
Triple_LATTE	9736
Triple_LATTE	25980
Triple_LATTE	14305
Triple_MOCHA	8244
Triple_MOCHA	18308
Triple_MOCHA	24092

60 rows selected (17.32 seconds)



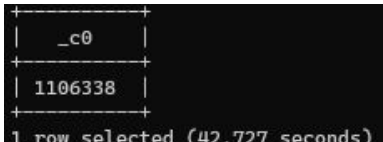
count	row_no
7500	1
25740	2
14684	3
15345	4
9636	5
27024	6
18024	7
8036	8
28318	9
8716	10
15512	11
26000	12
7804	13
17788	14
25228	15
15282	16
10304	17
23896	18
9956	19
16342	20
27526	21
34729	22
18704	23
50461	24
23624	25
7824	26
15763	27
15120	28
26792	29
9692	30
25484	31
8024	32
16912	33
15944	34
8144	35
22746	36
27120	37
8532	38
14550	39
7596	40
23896	41
15100	42
8149	43
15260	44
25828	45
8452	46
17184	47
28572	48
14601	49
25690	50
7992	51
32126	52
19704	53
56333	54
9736	55
25980	56
14305	57
8244	58
18308	59
24092	60

60 rows selected (16.772 seconds)



_c0
1115974

1 row selected (18.912 seconds)



_c0
1106338

1 row selected (42.727 seconds)

Thank you!
Any Questions?

Moses.Lin@Revature.net
https://github.com/Moses-Lin