

Project 1

Duncan Thomas

Revature Training Batch 210405 Reston Big Data

Tech Stack

- WSL2
- Ubuntu
- HDFS
- Hive
- Beeline



Scenario 1

What is the total number of consumers for Branch1?

What is the number of consumers for the Branch2?

```
+-----+-----+
|  branch  | consumers |
+-----+-----+
| Branch1  | 1115974   |
| Branch2  | 5099141   |
+-----+-----+
```

2 rows selected (21.693 seconds)

```
0: jdbc:hive2://> select branch.branch BRANCH, sum(conscount.count) CONSUMERS
. . . . . > from branch left join conscount
. . . . . > on branch.beverage = conscount.beverage
. . . . . > group by branch.branch
. . . . . > having branch.branch = 'Branch1' or branch.branch = 'Branch2';
```

Scenario 2

What is the most consumed beverage on Branch1?

```
+-----+-----+
| branch1_beverage | count |
+-----+-----+
| Special_cappuccino | 108163 |
+-----+-----+
1 row selected (41.102 seconds)
0: jdbc:hive2://> select b.beverage Branch1_Beverage, sum(c.count) count
. . . . . > from branch as b left join conscount as c
. . . . . > on b.beverage = c.beverage
. . . . . > where b.branch = "Branch1"
. . . . . > group by b.beverage
. . . . . > order by count desc
. . . . . > limit 1;
```

Scenario 2

What is the least consumed beverage on Branch2?

```
+-----+-----+
| branch2_beverage | count |
+-----+-----+
| Cold_MOCHA      | 47524 |
+-----+-----+
1 row selected (45.204 seconds)
0: jdbc:hive2://> select b.beverage Branch2_Beverage, sum(c.count) count
. . . . . > from branch as b left join conscount as c
. . . . . > on b.beverage = c.beverage
. . . . . > where b.branch = "Branch2"
. . . . . > group by b.beverage
. . . . . > order by count asc
. . . . . > limit 1;
```

Scenario 3

What are the beverages available on Branch10, Branch8, and Branch1?

```
0: jdbc:hive2://> select b.branch, beverage from branch b
. . . . . > where b.branch = 'Branch10'
. . . . . > or b.branch = 'Branch8'
. . . . . > or b.branch = 'Branch1';
```

b.branch	beverage
Branch1	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
Branch8	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_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

Scenario 3

What are the common beverages available in Branch4, Branch7?

```
+-----+
| _u1.common_beverages |
+-----+
| 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 (57.989 seconds)

```
0: jdbc:hive2://> select distinct beverage common_beverages from branch where branch = 'Branch4'
. . . . . > intersect select distinct beverage common_beverages from branch where branch = 'Branch7';
```

Scenario 4

Create a view for Scenario 3.

```
0: jdbc:hive2://> create view s4_view as
. . . . . > select b.branch as branch, beverage from branch b where b.branch = 'Branch10'
. . . . . > or branch = 'Branch8'
. . . . . > or branch = 'Branch1';
21/05/13 21:20:28 [HiveServer2-Background-Pool: Thread-606]: WARN conf.HiveConf: HiveConf of name hive.server2.enable.impersonation
does not exist
OK
No rows affected (0.116 seconds)
0: jdbc:hive2://> select * from s4_view;
OK
```

s4_view.branch	s4_view.beverage
Branch1	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
Branch8	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

Scenario 4

Create a partition for Scenario 3.

```
0: jdbc:hive2://> select * from s4_partition;
OK
```

s4_partition.beverage	s4_partition.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
Triple_Lite	Branch8
Special_cappuccino	Branch8
MED_Coffee	Branch8

```
0: jdbc:hive2://> create table s4_partition(beverage String) partitioned by(branch String);
21/05/13 21:32:58 [HiveServer2-Background-Pool: Thread-613]: WARN conf.HiveConf: HiveConf of name hive.server2.enable.impersonation
does not exist
OK
No rows affected (0.121 seconds)
0: jdbc:hive2://> set hive.exec.dynamic.partition.mode=nonstrict
. . . . . > ;
No rows affected (0.004 seconds)
0: jdbc:hive2://> insert overwrite table s4_partition partition(branch)
. . . . . > select beverage, branch from s4_view;
```

```
dacrok@lolwut-mkII:~$ hdfs dfs -ls /user/hive/warehouse/project1.db/s4_partition
```

Found 2 items

```
drwxrwxr-x - dacrok supergroup 0 2021-05-13 21:34 /user/hive/warehouse/project1.db/s4_partition/branch=Branch1
drwxrwxr-x - dacrok supergroup 0 2021-05-13 21:34 /user/hive/warehouse/project1.db/s4_partition/branch=Branch8
```

Scenario 4

Create a index for Scenario 3.

```
0: jdbc:hive2://> create index s4_index on table s4_partition(beverage)
. . . . . > as 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
. . . . . > with deferred rebuild;
```

```
0: jdbc:hive2://> show indexes on s4_partition;
```

```
21/05/13 21:48:42 [HiveServer2-Background-Pool: Thread-724]: WARN conf.HiveConf: HiveConf of name hive.internal.ss.authz.settings.applied.marker does not exist
```

```
21/05/13 21:48:42 [HiveServer2-Background-Pool: Thread-724]: WARN conf.HiveConf: HiveConf of name hive.server2.enable.impersonation does not exist
```

```
OK
```

idx_name	tab_name	col_names	idx_tab_name	idx_type	comment
s4_index	s4_partition	beverage	project1__s4_partition_s4_index__	compact	

```
1 row selected (0.056 seconds)
```

Scenario 5

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

```
➡ 0: jdbc:hive2://> show tblproperties s4_partition;
21/05/13 22:08:09 [HiveServer2-Background-Pool: Thread-728]: WARN conf.HiveConf: HiveConf of name hive.internal.ss.authz.settings.applied.marker does not exist
21/05/13 22:08:09 [HiveServer2-Background-Pool: Thread-728]: WARN conf.HiveConf: HiveConf of name hive.server2.enable.impersonation does not exist
OK
+-----+-----+
|      prpt_name      | prpt_value |
+-----+-----+
| transient_lastDdlTime | 1620959578 |
+-----+-----+
1 row selected (0.127 seconds)

➡ 0: jdbc:hive2://> alter table s4_partition set tblproperties("note"="this partitioned table shows the beverages available at branches 1, 8, and 10");
21/05/13 22:10:29 [HiveServer2-Background-Pool: Thread-731]: WARN conf.HiveConf: HiveConf of name hive.internal.ss.authz.settings.applied.marker does not exist
21/05/13 22:10:29 [HiveServer2-Background-Pool: Thread-731]: WARN conf.HiveConf: HiveConf of name hive.server2.enable.impersonation does not exist
OK
No rows affected (0.107 seconds)

➡ 0: jdbc:hive2://> alter table s4_partition set tblproperties("comment"="Wow! What a great table!");
21/05/13 22:10:53 [HiveServer2-Background-Pool: Thread-734]: WARN conf.HiveConf: HiveConf of name hive.internal.ss.authz.settings.applied.marker does not exist
21/05/13 22:10:53 [HiveServer2-Background-Pool: Thread-734]: WARN conf.HiveConf: HiveConf of name hive.server2.enable.impersonation does not exist
OK
No rows affected (0.088 seconds)

➡ 0: jdbc:hive2://> show tblproperties s4_partition;
21/05/13 22:10:56 [HiveServer2-Background-Pool: Thread-737]: WARN conf.HiveConf: HiveConf of name hive.internal.ss.authz.settings.applied.marker does not exist
21/05/13 22:10:56 [HiveServer2-Background-Pool: Thread-737]: WARN conf.HiveConf: HiveConf of name hive.server2.enable.impersonation does not exist
OK
+-----+-----+
|      prpt_name      | prpt_value |
+-----+-----+
| comment              | Wow! What a great table! |
| last_modified_by     | dacrok      |
| last_modified_time    | 1620961853  |
| note                 | this partitioned table shows the beverages available at branches 1, 8, and 10 |
| transient_lastDdlTime | 1620961853  |
+-----+-----+
5 rows selected (0.053 seconds)
```

Scenario 6

Remove the fifth row from the output of Scenario 1.

```
0: jdbc:hive2://> create table s6_table(row_num Int, branch String, consumers Int);
21/05/13 22:35:47 [HiveServer2-Background-Pool: Thread-897]: WARN conf.HiveConf: HiveConf of name hive.internal.ss.authz.settings.applied.marker does not exist
21/05/13 22:35:47 [HiveServer2-Background-Pool: Thread-897]: WARN conf.HiveConf: HiveConf of name hive.server2.enable.impersonation does not exist
OK
No rows affected (0.075 seconds)
0: jdbc:hive2://> insert overwrite table s6_table
. . . . . > select ROW_NUMBER() over (), branch.branch BRANCH, sum(conscount.count) CONSUMERS
. . . . . > from branch left join conscount
. . . . . > on branch.beverage = conscount.beverage
. . . . . > group by branch.branch;
```

```
0: jdbc:hive2://> select * from s6_table;
OK
```

s6_table.row_num	s6_table.branch	s6_table.consumers
1	Branch9	5208016
2	Branch8	2583583
3	Branch7	6566187
4	Branch6	8032799
5	Branch5	3941177
6	Branch4	3940862
7	Branch3	2591062
8	Branch2	5099141
9	Branch1	1115974

9 rows selected (0.081 seconds)

```
0: jdbc:hive2://> insert overwrite table s6_table
. . . . . > select * from s6_table
. . . . . > where row_num <= 5;
```

s6_table.row_num	s6_table.branch	s6_table.consumers
1	Branch9	5208016
2	Branch8	2583583
3	Branch7	6566187
4	Branch6	8032799
6	Branch4	3940862
7	Branch3	2591062
8	Branch2	5099141
9	Branch1	1115974

8 rows selected (0.081 seconds)

Questions?