P1: Hive

William Terry

August 2021

Problem1

Problem Scenario 1:

What is the total number of consumers for Branch1?
Answer: 1115974

What is the number of consumers for the Branch2?
Answer: 5099141

```
create table BevA(Beverage STRING, BranchID String)row format delimited fields terminated by ',' stored as textfile;
LOAD DATA INPATH '/user/williamgantterry/Bev_BranchA.txt' OVERWRITE INTO TABLE BevA;

create table BevB(Beverage STRING, BranchID String)row format delimited fields terminated by ',' stored as textfile;
LOAD DATA INPATH '/user/williamgantterry/Bev_BranchB.txt' OVERWRITE INTO TABLE BevB;

create table BevC(Beverage STRING, BranchID String)row format delimited fields terminated by ',' stored as textfile;
LOAD DATA INPATH '/user/williamgantterry/Bev_BranchC.txt' OVERWRITE INTO TABLE BevC;

create table ConsA(Beverage STRING, Consumed INT)row format delimited fields terminated by ',' stored as textfile;
LOAD DATA INPATH '/user/williamgantterry/Bev_ConscountA.txt' OVERWRITE INTO TABLE ConsA;

create table ConsB(Beverage STRING, Consumed INT)row format delimited fields terminated by ',' stored as textfile;
LOAD DATA INPATH '/user/williamgantterry/Bev_ConscountB.txt' OVERWRITE INTO TABLE ConsB;

create table ConsC(Beverage STRING, Consumed INT)row format delimited fields terminated by ',' stored as textfile;
LOAD DATA INPATH '/user/williamgantterry/Bev_ConscountC.txt' OVERWRITE INTO TABLE ConsC;

CREATE TABLE BevAll(Beverage STRING,BranchID STRING)row format delimited fields terminated by ',' stored as textfile;
INSERT INTO TABLE BevAll (SELECT * FROM BevA UNION ALL SELECT * FROM BevB UNION ALL SELECT * FROM BevC);

CREATE TABLE ConsAll(Beverage STRING,Consumed INT) row format delimited fields terminated by ',' stored as textfile;
INSERT INTO TABLE ConsAll(SELECT * FROM ConsA UNION ALL SELECT * FROM ConsB UNION ALL SELECT * FROM ConsC);

CREATE TABLE TotalCons(BranchID STRING, TotalConsumers INT);
INSERT INTO TABLE TotalCons(
SELECT BranchID, sum(ConsAll.Consumed) FROM BevAll join ConsAll On (ConsAll.Beverage=BevAll.Beverage) GROUP BY BevAll.BranchID);

SELECT * FROM TotalCons;
```

```
--+---------------------+---------------------------+
--| totalcons.branchid  | totalcons.totalconsumers  |
--+---------------------+---------------------------+
--| Branch1             | 1115974                   |
--| Branch2             | 5099141                   |
--| Branch3             | 2591062                   |
--| Branch4             | 3940862                   |
--| Branch5             | 3941177                   |
--| Branch6             | 8032799                   |
--| Branch7             | 6566187                   |
--| Branch8             | 2583583                   |
--| Branch9             | 5208016                   |
--+---------------------+---------------------------+
```

```
williamgantterry@LAPTOP-H2L556UK: ~

9 rows selected (0.122 seconds)
0: jdbc:hive2://> SELECT * FROM TotalCons;
OK
+----------------------+----------------------------+
| totalcons.branchid   | totalcons.totalconsumers   |
+----------------------+----------------------------+
| Branch1              | 1115974                    |
| Branch2              | 5099141                    |
| Branch3              | 2591062                    |
| Branch4              | 3940862                    |
| Branch5              | 3941177                    |
| Branch6              | 8032799                    |
| Branch7              | 6566187                    |
| Branch8              | 2583583                    |
| Branch9              | 5208016                    |
+----------------------+----------------------------+
9 rows selected (0.107 seconds)
0: jdbc:hive2://>
```

Figure: Problem 1

Problem1
Problem2
Problem3
Problem4
Problem5
Problem6
GitHub

Problem2

Problem Scenario 2:

What is the most consumed beverage on Branch1?
Answer: Special Cappucino with ⎢108163⎥

What is the least consumed beverage on Branch2? Answer: Cold
Mocha with ⎢47524⎥

What is the Average (median) consumed beverage of Branch2
Answer: ⎢26/51⎥ in order which is Small Latte with ⎢93184⎥

```
SELECT BranchID, sum(ConsAll.Consumed),BevAll.Beverage FROM BevAll join ConsAll On (ConsAll.Beverage=BevAll.Beverage)
WHERE BranchID="Branch1" GROUP BY BevAll.BranchID,BevAll.Beverage ORDER BY sum(ConsAll.Consumed) DESC
;
```

branchid	_c1	bevall.beverage
Branch1	108163	Special_cappuccino
Branch1	103894	LARGE_cappuccino
Branch1	54378	Double_Espresso
Branch1	54208	Special_Coffee
Branch1	53824	LARGE_Espresso
Branch1	52005	Cold_Lite
Branch1	51604	MED_MOCHA
Branch1	50820	ICY_Espresso
Branch1	50644	Triple_MOCHA
Branch1	50420	Mild_LATTE
Branch1	50228	Double_MOCHA
Branch1	50202	SMALL_Espresso
Branch1	50021	Triple_LATTE
Branch1	49482	ICY_MOCHA
Branch1	49237	SMALL_MOCHA
Branch1	48283	Special_Espresso
Branch1	47924	Cold_LATTE
Branch1	47211	MED_LATTE
Branch1	46834	Mild_Lite
Branch1	46592	SMALL_LATTE

Figure: Problem 2

Problem1
Problem2
Problem3
Problem4
Problem5
Problem6
GitHub

williamgantterry@LAPTOP-H2L556UK: ~

```
Ended Job = job_1627998776566_0042
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 5.98 sec   HDFS Read: 123682 HDFS Write: 760 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 3.73 sec   HDFS Read: 8450 HDFS Write: 871 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 710 msec
OK
```

branchid	_c1	bevall.beverage
Branch1	108163	Special_cappuccino
Branch1	103894	LARGE_cappuccino
Branch1	54378	Double_Espresso
Branch1	54208	Special_Coffee
Branch1	53824	LARGE_Espresso
Branch1	52005	Cold_Lite
Branch1	51604	MED_MOCHA
Branch1	50820	ICY_Espresso
Branch1	50644	Triple_MOCHA
Branch1	50420	Mild_LATTE
Branch1	50228	Double_MOCHA
Branch1	50202	SMALL_Espresso
Branch1	50021	Triple_LATTE
Branch1	49482	ICY_MOCHA
Branch1	49237	SMALL_MOCHA
Branch1	48283	Special_Espresso
Branch1	47924	Cold_LATTE
Branch1	47211	MED_LATTE
Branch1	46834	Mild_Lite
Branch1	46592	SMALL_LATTE

```
20 rows selected (61.25 seconds)
0: jdbc:hive2://>
```

```
SELECT BranchID, ROW_NUMBER() OVER (ORDER BY sum(ConsAll.Consumed)) AS row_num,sum(ConsAll.Consumed),BevAll.Beverage
FROM BevAll join ConsAll On (ConsAll.Beverage=BevAll.Beverage) WHERE BranchID="Branch2"
GROUP BY BevAll.BranchID,BevAll.Beverage ORDER BY sum(ConsAll.Consumed) DESC
;
```

branchid	row_num	_c2	bevall.beverage
Branch2	51	328074	Mild_cappuccino
Branch2	50	285309	MED_cappuccino
Branch2	49	284943	Cold_cappuccino
Branch2	48	216326	Special_cappuccino
Branch2	47	207788	LARGE_cappuccino
Branch2	46	204476	ICY_cappuccino
Branch2	45	202854	SMALL_cappuccino
Branch2	44	197620	Double_cappuccino
Branch2	43	193636	Triple_cappuccino
Branch2	42	112378	Double_LATTE
Branch2	41	108292	Special_MOCHA
Branch2	40	107648	LARGE_Espresso
Branch2	39	104466	LARGE_LATTE
Branch2	38	104010	Cold_Lite
Branch2	37	103572	MED_Lite
Branch2	36	101192	LARGE_MOCHA
Branch2	35	100408	Triple_Lite
Branch2	34	100042	Triple_LATTE
Branch2	33	98964	ICY_MOCHA
Branch2	32	98614	SMALL_Lite
Branch2	31	98532	Double_Lite
Branch2	30	98332	ICY_LATTE
Branch2	29	96566	Special_Espresso
Branch2	28	94422	MED_LATTE
Branch2	27	93668	Mild_Lite
Branch2	26	93184	SMALL_LATTE
Branch2	25	54378	Double_Espresso
Branch2	24	54208	Special_Coffee
Branch2	23	54050	Special_Lite
Branch2	22	53681	Mild_MOCHA
Branch2	21	51932	Cold_Coffee
Branch2	20	51880	Mild_Coffee
Branch2	19	51694	Double_Coffee
Branch2	18	51647	LARGE_Coffee
Branch2	17	51604	MED_MOCHA

William Terry   P1: Hive

```
-| Branch2  | 16      | 50820   | ICY_Espresso        |
-| Branch2  | 15      | 50648   | Triple_Coffee       |
-| Branch2  | 14      | 50644   | Triple_MOCHA        |
-| Branch2  | 13      | 50604   | Special_LATTE       |
-| Branch2  | 12      | 50228   | Double_MOCHA        |
-| Branch2  | 11      | 50202   | SMALL_Espresso      |
-| Branch2  | 10      | 49501   | SMALL_Coffee        |
-| Branch2  | 9       | 49237   | SMALL_MOCHA         |
-| Branch2  | 8       | 48994   | MED_Coffee          |
-| Branch2  | 7       | 48775   | Triple_Espresso     |
-| Branch2  | 6       | 48555   | ICY_Coffee          |
-| Branch2  | 5       | 48462   | Mild_Espresso       |
-| Branch2  | 4       | 48413   | MED_Espresso        |
-| Branch2  | 3       | 48220   | ICY_Lite            |
-| Branch2  | 2       | 47924   | Cold_LATTE          |
-| Branch2  | 1       | 47524   | Cold_MOCHA          |
-+----------+---------+---------+---------------------+
```

Figure: Problem 2

🔵 williamgantterry@LAPTOP-H2L556UK: ~

```
Total MapReduce CPU Time Spent: 15 seconds 170 msec
OK
+----------+---------+--------+--------------------+
| branchid | row_num |  _c2   |   bevall.beverage  |
+----------+---------+--------+--------------------+
| Branch2  | 51      | 328074 | Mild_cappuccino    |
| Branch2  | 50      | 285309 | MED_cappuccino     |
| Branch2  | 49      | 284943 | Cold_cappuccino    |
| Branch2  | 48      | 216326 | Special_cappuccino |
| Branch2  | 47      | 207788 | LARGE_cappuccino   |
| Branch2  | 46      | 204476 | ICY_cappuccino     |
| Branch2  | 45      | 202854 | SMALL_cappuccino   |
| Branch2  | 44      | 197620 | Double_cappuccino  |
| Branch2  | 43      | 193636 | Triple_cappuccino  |
| Branch2  | 42      | 112378 | Double_LATTE       |
| Branch2  | 41      | 108292 | Special_MOCHA      |
| Branch2  | 40      | 107648 | LARGE_Espresso     |
| Branch2  | 39      | 104466 | LARGE_LATTE        |
| Branch2  | 38      | 104010 | Cold_Lite          |
| Branch2  | 37      | 103572 | MED_Lite           |
| Branch2  | 36      | 101192 | LARGE_MOCHA        |
| Branch2  | 35      | 100408 | Triple_Lite        |
| Branch2  | 34      | 100042 | Triple_LATTE       |
| Branch2  | 33      | 98964  | ICY_MOCHA          |
| Branch2  | 32      | 98614  | SMALL_Lite         |
| Branch2  | 31      | 98532  | Double_Lite        |
| Branch2  | 30      | 98332  | ICY_LATTE          |
| Branch2  | 29      | 96566  | Special_Espresso   |
| Branch2  | 28      | 94422  | MED_LATTE          |
| Branch2  | 27      | 93668  | Mild_Lite          |
| Branch2  | 26      | 93184  | SMALL_LATTE        |
| Branch2  | 25      | 54378  | Double_Espresso    |
| Branch2  | 24      | 54208  | Special_Coffee     |
```

Figure: Problem 2

William Terry   P1: Hive

Branch2	23	54050	Special_Lite
Branch2	22	53681	Mild_MOCHA
Branch2	21	51932	Cold_Coffee
Branch2	20	51880	Mild_Coffee
Branch2	19	51694	Double_Coffee
Branch2	18	51647	LARGE_Coffee
Branch2	17	51604	MED_MOCHA
Branch2	16	50820	ICY_Espresso
Branch2	15	50648	Triple_Coffee
Branch2	14	50644	Triple_MOCHA
Branch2	13	50604	Special_LATTE
Branch2	12	50228	Double_MOCHA
Branch2	11	50202	SMALL_Espresso
Branch2	10	49501	SMALL_Coffee
Branch2	9	49237	SMALL_MOCHA
Branch2	8	48994	MED_Coffee
Branch2	7	48775	Triple_Espresso
Branch2	6	48555	ICY_Coffee
Branch2	5	48462	Mild_Espresso
Branch2	4	48413	MED_Espresso
Branch2	3	48220	ICY_Lite
Branch2	2	47924	Cold_LATTE
Branch2	1	47524	Cold_MOCHA

51 rows selected (102.85 seconds)

Figure: Problem 2

Problem3

Problem Scenario 3 What are the beverages available on Branch10, Branch8, and Branch1? What are the common beverages available in Branch4,Branch7?

```
SELECT DISTINCT BevAll.beverage FROM BevAll WHERE BRANCHID="Branch1" or BRANCHID="Branch8" or BRANCHID="Branch10";

SELECT BevAll.beverage, BevAll.BRANCHID FROM BevAll WHERE BRANCHID="Branch1" or BRANCHID="Branch8" or BRANCHID="Branch10"
ORDER BY BevAll.beverage;
```

```
+----------------------+
|   bevall.beverage    ||
+----------------------+
| Cold_Coffee          |
| Cold_LATTE           |
| Cold_Lite            |
| Cold_cappuccino      |
| Double_Coffee        |
| Double_Espresso      |
| Double_LATTE         |
| Double_MOCHA         |
| Double_cappuccino    |
| ICY_Coffee           |
| ICY_Espresso         |
| ICY_Lite             |
| ICY_MOCHA            |
| 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_LATTE           |
| Mild_Lite            |
| 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_LATTE        |
| Triple_Lite         |
| Triple_MOCHA        |
| Triple_cappuccino   |
+---------------------+
```

Figure: Problem 3

```
--what are the common beverages available in Branch4,Branch7?

select BevAll.beverage from BevAll where BevAll.BRANCHID = "Branch4" and BevAll.beverage in (select BevAll.beverage from BevAll where BevAll.BRANCHID = "Branch7")
GROUP BY BevAll.beverage ;

-+---------------------+
-|    bevall.beverage  |
-+---------------------+
-| 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     |
```

Figure: Problem 3

William Terry   P1: Hive

```
-| 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    |
-+----------------------+
```

Figure: Problem 3

Problem1
Problem2
Problem3
Problem4
Problem5
Problem6
GitHub

williamgantterry@LAPTOP-H2L556UK: ~

```
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.67 sec   HDFS Read: 39661 HDFS Write: 1203 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 670 msec
OK
+---------------------+
|   bevall.beverage   |
+---------------------+
| Cold_Coffee         |
| Cold_LATTE          |
| Cold_Lite           |
| Cold_cappuccino     |
| Double_Coffee       |
| Double_Espresso     |
| Double_LATTE        |
| Double_MOCHA        |
| Double_cappuccino   |
| ICY_Coffee          |
| ICY_Espresso        |
| ICY_Lite            |
| ICY_MOCHA           |
| 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_LATTE          |
| Mild_Lite           |
```

William Terry   P1: Hive

Problem4

```
-Problem Scenario 4
-create a partition, View for the scenario3.

Create VIEW SampView AS select BevAll.beverage from BevAll where BevAll.BRANCHID = "Branch4" and BevAll.beverage in
(select BevAll.beverage from BevAll where BevAll.BRANCHID = "Branch7")
GROUP BY BevAll.beverage;

SELECT * FROM SampView;

create table Part4(beverage STRING) PARTITIONED by (branchid STRING) row format delimited fields terminated by ',' stored as textfile;
INSERT INTO TABLE Part4(SELECT BevAll.beverage, BevAll.BRANCHID FROM BevAll WHERE BRANCHID="Branch1" or BRANCHID="Branch8" or BRANCHID="Branch10"
ORDER BY BevAll.beverage);

show partitions Part4;

-+-------------------+
-|     partition     |
-+-------------------+
-| branchid=Branch1  |
-| branchid=Branch8  |
-+-------------------+
```

Figure: Problem 4

Problem4

Problem1
Problem2
Problem3
Problem4
Problem5
Problem6
GitHub

Problem4

williamgantterry@LAPTOP-H2L556UK: ~

```
51 rows selected (34.305 seconds)
0: jdbc:hive2://> show partitions Part4;
OK
+------------------+
|     partition    |
+------------------+
| branchid=Branch1 |
| branchid=Branch8 |
+------------------+
2 rows selected (0.096 seconds)
```

William Terry   P1: Hive

Problem5

```
-Problem Scenario 5
-Alter the table properties to add "note","comment"

ALTER TABLE Part4 SET TBLPROPERTIES ('note'='this is a note');
ALTER TABLE Part4 SET TBLPROPERTIES ('comment'='this is a comment');

-Show tblproperties Part4;
-+-------------------------+-------------------+
-|       prpt_name         |     prpt_value    |
-+-------------------------+-------------------+
-| bucketing_version       | 2                 |
-| comment                 | this is a comment |
-| last_modified_by        | williamgantterry  |
-| last_modified_time      | 1628007015        |
-| note                    | this is a note    |
-| transient_lastDdlTime   | 1628007015        |
-+-------------------------+-------------------+
```

Figure: Problem 5

William Terry   P1: Hive

Problem1
Problem2
Problem3
Problem4
Problem5
Problem6
GitHub

Problem5

Problem6

```
Problem Scenario 6
-Remove a row from the any Scenario.

DROP TABLE IF EXISTS Tab62;

CREATE TABLE Tab62(BranchID STRING, TotalConsumers INT) STORED AS ORC;
INSERT INTO TABLE Tab62(SELECT * FROM TotalCons);

SELECT * FROM Tab62;
-+-----------------+-----------------------+
-| tab62.branchid  | tab62.totalconsumers  |
-+-----------------+-----------------------+
-| Branch1         | 1115974               |
-| Branch2         | 5099141               |
-| Branch3         | 2591062               |
-| Branch4         | 3940862               |
-| Branch5         | 3941177               |
-| Branch6         | 8032799               |
-| Branch7         | 6566187               |
-| Branch8         | 2583583               |
-| Branch9         | 5208016               |
-+-----------------+-----------------------+
```

Figure: Problem 6

Problem6

```
SET hive.support.concurrency=true;
SET hive.txn.manager=org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;
ALTER TABLE Tab62 SET TBLPROPERTIES ('transactional'='true');
DELETE FROM Tab62 WHERE Tab62.BranchID='Branch2';

SELECT * FROM Tab62;
-+----------------+----------------------+
-| tab62.branchid | tab62.totalconsumers |
-+----------------+----------------------+
-| Branch1        | 1115974              |
-| Branch3        | 2591062              |
-| Branch4        | 3940862              |
-| Branch5        | 3941177              |
-| Branch6        | 8032799              |
-| Branch7        | 6566187              |
-| Branch8        | 2583583              |
-| Branch9        | 5208016              |
-+----------------+----------------------+
```

Figure: Problem 6

Problem1
Problem2
Problem3
Problem4
Problem5
Problem6
GitHub

Problem4

Figure: Problem 6

Problem1
Problem2
Problem3
Problem4
Problem5
Problem6
GitHub

Problem4

williamgantterry@LAPTOP-H2L556UK: ~

```
0: jdbc:hive2://> 0: jdbc:hive2://>
SELECT * FROM Tab62;21/08/04 09:54:54 [1890c451
OK
+------------------+------------------------+
| tab62.branchid   | tab62.totalconsumers   |
+------------------+------------------------+
| Branch1          | 1115974                |
| Branch3          | 2591062                |
| Branch4          | 3940862                |
| Branch5          | 3941177                |
| Branch6          | 8032799                |
| Branch7          | 6566187                |
| Branch8          | 2583583                |
| Branch9          | 5208016                |
+------------------+------------------------+
8 rows selected (0.271 seconds)
```

GitHub

https://github.com/williamgantterry/P1