#Assignment:Hive Assignment

#referral_id:DIRSS3134

#Name:Gauri Vijay Kolte

1.You need to check the data in case if the column contain null value so replace it with 0 or ‘ZERO’ based om the data type of the function

```
hive> select * from cz;
OK
cz.eid  cz.ename        cz.salary
12      gauri   980
14      kol     0
15      ji      NULL
16      risk    NULL
17      gok     675
Time taken: 17.74 seconds, Fetched: 5 row(s)

hive> select coalesce(salary,0) from cz;
Query ID = hdfs_20220820112828_638d2fad-ff90-4ae6-85b7-dfed671aa010
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1661014522259_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1661014522259_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1661014522259_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-08-20 11:39:08,343 Stage-1 map = 0%,  reduce = 0%
2022-08-20 11:40:12,433 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.72 sec
MapReduce Total cumulative CPU time: 2 seconds 720 msec
Ended Job = job_1661014522259_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 2.72 sec   HDFS Read: 3872 HDFS Write: 14 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 720 msec
OK
_c0
980
0
0
0
675
Time taken: 742.428 seconds, Fetched: 5 row(s)
```

2.Now take a table which contain atleast four column as id, name, dept, salary

-in case if the person is having the salary less than 500 do a increament by 50%(hike)

-in case the salary is in between 5000 to 10000 hike the salary should be (25%) and other cases will be same

```
hive> select salary+salary*0.50 from sal where salary < 5000;
Query ID = hdfs_20220820231818_71f53170-f686-4243-8a4c-ace6c07ce3df
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1661057151732_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1661057151732_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1661057151732_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-08-20 23:19:11,450 Stage-1 map = 0%,  reduce = 0%
2022-08-20 23:19:22,531 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.8 sec
MapReduce Total cumulative CPU time: 1 seconds 800 msec
Ended Job = job_1661057151732_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 1.8 sec   HDFS Read: 5098 HDFS Write: 28 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 800 msec
OK
_c0
6000.0
6750.0
3000.0
4500.0
Time taken: 25.133 seconds, Fetched: 4 row(s)
hive> select salary+salary*0.25 from sal where salary between 5000 and 10000;
Query ID = hdfs_20220820232222_8c79c0f0-e6df-4bf8-9b86-f2c92204eec1
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1661057151732_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1661057151732_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1661057151732_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-08-20 23:22:20,112 Stage-1 map = 0%,  reduce = 0%
2022-08-20 23:22:38,573 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.41 sec
MapReduce Total cumulative CPU time: 2 seconds 410 msec
Ended Job = job_1661057151732_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 2.41 sec   HDFS Read: 5301 HDFS Write: 38 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 410 msec
OK
_c0
7500.0
9562.5
12500.0
10750.0
11250.0
Time taken: 34.944 seconds, Fetched: 5 row(s)
```

3.Know the difference between the joins operation and set operation like (union, intersect)

Join:

- Join is used to combine rows from two or more tables, based on a related column between them.
- There are 4 types of join: Inner Join, Left Join, Right join, Full Outer Join
- Inner join: Return records that have matching values join both tables
- Left join: Returns all records from left table, and matched records from the right table
- Right join: Returns all records from right table, and matched records from the left table
- Full join: Returns all records when there is match in either left or right table.

Set:

- Set operators are which is used to combine the information of similar type from one or more than one table set operator combine rows from distinct queries.
- Set Operator: Union, Union All, Intersect and minus
- This set operator is used to combine the outputs of two or more queries into a single set of rows and column having different records.
- Intersect: This set operator is available to retrieve the information which is common in both tables. The number of columns and data type, use be same in intersect set operator.

4.Try to implement the same kind of left, right and full outer join on the same employee and department

```
hive> select * from sal;
OK
sal.eid sal.ename       sal.edept       sal.salary
1       Gauri   IT      6000
2       Kolte   HM      7650
3       Rani    BE      4000
4       Avi     MTech   4500
5       Bhalchandra     HM      10000
6       Parsha  BE      8600
7       Mansi   IT      9000
8       Rovky   MTech   2000
9       Golu    IT      3000
10      Hindvi  HM      12000
11      Raju    BE      14000
Time taken: 0.157 seconds, Fetched: 11 row(s)
hive> select * from joi;
OK
joi.eid joi.ename       joi.edept       joi.salary
1       Gauri   IT      99000
2       Koli    HM      70650
3       Rani    BE      4000
4       Abhi    MTech   4500
5       Bhala   HM      13000
6       Parsha  BE      8600
7       Mansi   IT      9000
8       Ravikrishna     MTech   22000
9       Golu    btech   3000
10      Honda   btech   12000
11      Raj     HM      14000
Time taken: 0.072 seconds, Fetched: 11 row(s)
```

Left join

```
hive> select sal.eid,sal.ename,sal.edept,sal.salary from sal left join joi on (sal.ename=joi.ename);
Query ID = hdfs_20220821022222_b4450d4d-9de4-4bd9-8245-a2bf1d30addc
Total jobs = 1
Execution log at: /tmp/hdfs/hdfs_20220821022222_b4450d4d-9de4-4bd9-8245-a2bf1d30addc.log
2022-08-21 02:22:50     Starting to launch local task to process map join;      maximum memory = 129761280
2022-08-21 02:22:51     Dump the side-table for tag: 1 with group count: 11 into file: file:/tmp/hdfs/b6ffe062-19db-4c6d-b2cf-f29277c55dea/hive_2022-08-21_02-22-43_570_6735802476975074416-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile3
1--.hashtable
2022-08-21 02:22:51     Uploaded 1 File to: file:/tmp/hdfs/b6ffe062-19db-4c6d-b2cf-f29277c55dea/hive_2022-08-21_02-22-43_570_6735802476975074416-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile31--.hashtable (517 bytes)
2022-08-21 02:22:51     End of local task; Time Taken: 1.329 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1661057151732_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1661057151732_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1661057151732_0016
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-08-21 02:23:03,466 Stage-3 map = 0%,  reduce = 0%
2022-08-21 02:23:12,444 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.33 sec
MapReduce Total cumulative CPU time: 1 seconds 330 msec
Ended Job = job_1661057151732_0016
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.33 sec   HDFS Read: 6367 HDFS Write: 190 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 330 msec
OK
sal.eid sal.ename       sal.edept       sal.salary
1       Gauri   IT      6000
2       Kolte   HM      7650
3       Rani    BE      4000
4       Avi     MTech   4500
5       Bhalchandra     HM      10000
6       Parsha  BE      8600
7       Mansi   IT      9000
8       Rovky   MTech   2000
9       Golu    IT      3000
10      Hindvi  HM      12000
11      Raju    BE      14000
Time taken: 30.015 seconds, Fetched: 11 row(s)
```

Right join

```
hive> select sal.eid,sal.ename,sal.edept,sal.salary from sal right join joi on (sal.ename=joi.ename);
Query ID = hdfs_20220821022323_6861f066-7348-4cac-a2f8-0c77969e6b8f
Total jobs = 1
Execution log at: /tmp/hdfs/hdfs_20220821022323_6861f066-7348-4cac-a2f8-0c77969e6b8f.log
2022-08-21 02:23:32     Starting to launch local task to process map join;      maximum memory = 129761280
2022-08-21 02:23:33     Dump the side-table for tag: 0 with group count: 11 into file: file:/tmp/hdfs/b6ffe062-19db-4c6d-b2cf-f29277c55dea/hive_2022-08-21_02-23-25_993_7057780632598588091-1/-local-10003/HashTable-S
0--.hashtable
2022-08-21 02:23:33     Uploaded 1 File to: file:/tmp/hdfs/b6ffe062-19db-4c6d-b2cf-f29277c55dea/hive_2022-08-21_02-23-25_993_7057780632598588091-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile40--.hashtable (613 b
2022-08-21 02:23:33     End of local task; Time Taken: 1.713 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1661057151732_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1661057151732_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1661057151732_0017
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-08-21 02:23:46,930 Stage-3 map = 0%,  reduce = 0%
2022-08-21 02:23:57,282 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.5 sec
MapReduce Total cumulative CPU time: 1 seconds 500 msec
Ended Job = job_1661057151732_0017
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.5 sec   HDFS Read: 6352 HDFS Write: 151 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 500 msec
OK
sal.eid sal.ename       sal.edept       sal.salary
1       Gauri   IT      6000
NULL    NULL    NULL    NULL
3       Rani    BE      4000
NULL    NULL    NULL    NULL
NULL    NULL    NULL    NULL
6       Parsha  BE      8600
7       Mansi   IT      9000
NULL    NULL    NULL    NULL
9       Golu    IT      3000
NULL    NULL    NULL    NULL
NULL    NULL    NULL    NULL
Time taken: 32.409 seconds, Fetched: 11 row(s)
```

Full join

```
hive> select sal.eid,sal.ename,sal.edept,sal.salary from sal full join joi on (sal.ename=joi.ename);
Query ID = hdfs_20220821022626_e22fa8df-1cb9-4096-853d-eb8997aeb477
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1661057151732_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1661057151732_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1661057151732_0018
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-08-21 02:26:22,541 Stage-1 map = 0%,  reduce = 0%
2022-08-21 02:26:53,664 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.8 sec
2022-08-21 02:27:03,478 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.13 sec
MapReduce Total cumulative CPU time: 4 seconds 130 msec
Ended Job = job_1661057151732_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1   Cumulative CPU: 4.13 sec   HDFS Read: 13199 HDFS Write: 262 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 130 msec
OK
sal.eid sal.ename       sal.edept       sal.salary
NULL    NULL    NULL    NULL
4       Avi     MTech   4500
NULL    NULL    NULL    NULL
5       Bhalchandra     HM      10000
1       Gauri   IT      6000
9       Golu    IT      3000
10      Hindvi  HM      12000
NULL    NULL    NULL    NULL
NULL    NULL    NULL    NULL
2       Kolte   HM      7650
7       Mansi   IT      9000
6       Parsha  BE      8600
NULL    NULL    NULL    NULL
11      Raju    BE      14000
3       Rani    BE      4000
NULL    NULL    NULL    NULL
8       Rovky   MTech   2000
Time taken: 56.112 seconds, Fetched: 17 row(s)
```

5.You need to create a view by joining the above two table in Q4

```
hive> create or replace view newjoisj as select sal.eid,sal.ename,sal.edept,sal.salary from sal left join joi on (sal.ename=joi.ename);
OK
eid     ename   edept   salary
Time taken: 0.302 seconds
hive> select * from newjoisj;
Query ID = hdfs_20220821024444_503da1e3-c588-4a1b-927d-8e4076b79047
Total jobs = 1
Execution log at: /tmp/hdfs/hdfs_20220821024444_503da1e3-c588-4a1b-927d-8e4076b79047.log
2022-08-21 02:44:26     Starting to launch local task to process map join;      maximum memory = 129761280
2022-08-21 02:44:27     Dump the side-table for tag: 1 with group count: 11 into file: file:/tmp/hdfs/b6ffe062-19db-4c6d-b2cf-f29277c55dea/hive_2022-08-21_02-44-22_480_4276079353616301866-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile5
1--.hashtable
2022-08-21 02:44:27     Uploaded 1 File to: file:/tmp/hdfs/b6ffe062-19db-4c6d-b2cf-f29277c55dea/hive_2022-08-21_02-44-22_480_4276079353616301866-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile51--.hashtable (517 bytes)
2022-08-21 02:44:27     End of local task; Time Taken: 0.867 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1661057151732_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1661057151732_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1661057151732_0020
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-08-21 02:44:36,426 Stage-3 map = 0%,  reduce = 0%
2022-08-21 02:44:43,976 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.03 sec
MapReduce Total cumulative CPU time: 1 seconds 30 msec
Ended Job = job_1661057151732_0020
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.03 sec   HDFS Read: 6392 HDFS Write: 198 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 30 msec
OK
newjoisj.eid    newjoisj.ename  newjoisj.edept  newjoisj.salary
1       Gauri   IT      6000
2       Kolte   HM      7650
3       Rani    BE      4000
4       Avi     MTech   4500
5       Bhalchandra     HM      10000
6       Parsha  BE      8600
7       Mansi   IT      9000
8       Rovky   MTech   2000
9       Golu    IT      3000
10      Hindvi  HM      12000
11      Raju    BE      14000
Time taken: 23.63 seconds, Fetched: 11 row(s)
```

6.Try to alter the view by dropping certain columns

```
hive> ALTER view newjoisj as select sal.ename from sal;
OK
ename
Time taken: 0.344 seconds
hive> select * from newjoisj;
OK
newjoisj.ename
Gauri
Kolte
Rani
Avi
Bhalchandra
Parsha
Mansi
Rovky
Golu
Hindvi
Raju
Time taken: 0.077 seconds, Fetched: 11 row(s)
```

7.Try to change the name of your view without replacing it with the new creation

```
hive> create view tem as select pr12.ename from pr12;
OK
ename
Time taken: 0.624 seconds
hive> select * from tem;
OK
tem.ename
it
it
it
it
hm
hm
hm
btech
btech
btech
Time taken: 0.327 seconds, Fetched: 10 row(s)
hive> ALTER VIEW  tem RENAME TO tem1;
OK
Time taken: 1.27 seconds
hive> SELECT * FROM tem1
    > SELECT * FROM tem1;
FAILED: ParseException line 2:0 missing EOF at 'SELECT' near 'tem1'
hive> SELECT * FROM tem1;
OK
tem1.ename
it
it
it
it
hm
hm
hm
btech
btech
btech
Time taken: 0.396 seconds, Fetched: 10 row(s)
hive>
```

8.For any table try to change the column data type of any column from string to int.

```
hive> alter table pr12 change ecity ena1 int;
OK
Time taken: 8.526 seconds
hive> select * from pr12;
OK
pr12.type       pr12.ename      pr12.edept      pr12.ena1
1       it      gauri   NULL
2       it      kolte   NULL
3       it      kol     NULL
4       it      koly    NULL
5       hm      bh      NULL
6       hm      kol     NULL
7       hm      ghv     NULL
8       btech   shannu  NULL
9       btech   rashmi  NULL
10      btech   siri    NULL
Time taken: 0.413 seconds, Fetched: 10 row(s)
hive>
```