Bharathwin MA

205229105

Big Data Management And Analysis Lab

Analysing the Weather Dataset using HIVE Query Language

Q01: Create View ‘temp25to45’ by select which record has ‘temperatureHigh’ between 25 degree Celsius and 45 degree Celsius

```
hive> create VIEW temp25to45 as select * from d_weather2020 where temphigh > 25 and temphigh < 45;
OK
Time taken: 0.992 seconds
hive> show tables;
OK
d_weather2020
temp25to45
Time taken: 0.09 seconds, Fetched: 2 row(s)
hive>
```

```
Time taken: 0.202 seconds, Fetched: 5283 row(s)
hive> select * from temp25to45;
```

Q02: Show how many temperature readings in India.

```
hive> select count(temphigh) from d_weather2020 WHERE country = 'India';
Query ID = cloudera_20211029021616_a6465689-fd7a-45fd-9a9c-7451564706bb
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1635497035806_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635497035806_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1635497035806_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-29 02:17:15,064 Stage-1 map = 0%,  reduce = 0%
2021-10-29 02:17:28,508 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.95 sec
2021-10-29 02:17:39,103 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.04 sec
MapReduce Total cumulative CPU time: 5 seconds 40 msec
Ended Job = job_1635497035806_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.04 sec   HDFS Read: 2717982 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 40 msec
OK
112
Time taken: 50.944 seconds, Fetched: 1 row(s)
hive>
```

```
hive> select count(templow) from d_weather2020 WHERE country = 'India';
Query ID = cloudera_20211029022929_d41080ae-3069-4afa-ac0a-f87e709e21e9
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1635497035806_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635497035806_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1635497035806_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-29 02:29:39,631 Stage-1 map = 0%,  reduce = 0%
2021-10-29 02:29:58,047 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.24 sec
2021-10-29 02:30:07,607 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.41 sec
MapReduce Total cumulative CPU time: 5 seconds 410 msec
Ended Job = job_1635497035806_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.41 sec   HDFS Read: 2718066 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 410 msec
OK
112
Time taken: 44.774 seconds, Fetched: 1 row(s)
hive>
```

Q03: Group the d_weather2020 table records by country name. In addition this group should have less than 25 degree Celsius in temperatureLow column

```
hive> select country, count(templow) from d_weather2020 WHERE templow<25 GROUP BY countrY;
Query ID = cloudera_20211029030909_24698881-0c92-4861-9db3-c697f58b9814
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_1635497035806_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635497035806_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1635497035806_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-29 03:09:24,861 Stage-1 map = 0%,  reduce = 0%
2021-10-29 03:09:38,025 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.23 sec
2021-10-29 03:09:47,705 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.48 sec
MapReduce Total cumulative CPU time: 5 seconds 480 msec
Ended Job = job_1635497035806_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.48 sec   HDFS Read: 2717276 HDFS Write: 553 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 480 msec
OK
Afghanistan     17
Andorra 1
Armenia 89
Austria 28
Azerbaijan      2
Belarus 16
Bhutan  10
Bosnia and Herzegovina  66
Bulgaria        34
Canada  786
China   935
Croatia 13
Estonia 14
Finland 47
France  5
Georgia 7
Germany 6
Greece  5
Greenland       112
Hungary 11
Iceland 93
Iran    5
Japan   44
Kazakhstan      79
Kyrgyzstan      78
Latvia  7
Lebanon 2
Liechtenstein   49
Lithuania       8
Luxembourg      1
Moldova 19
Mongolia        96
North Macedonia 10
Norway  72
Pakistan        5
Poland  3
Romania 33
Russia  82
Serbia  8
Slovakia        52
Slovenia        3
Spain   1
Sweden  43
Switzerland     26
Turkey  41
US      1381
Ukraine 17
Uzbekistan      18
Time taken: 35.747 seconds, Fetched: 48 row(s)
hive>
```

Q04: Calculate average humidity of each country and order the result by ‘country’ name.

```
hive> select country, avg(hum) from d_weather2020 GROUP BY country ORDER BY country;
Query ID = cloudera_20211029031717_59f616b6-2cf6-4a8a-a763-2294edca4bdf
Total jobs = 2
Launching Job 1 out of 2
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_1635497035806_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635497035806_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1635497035806_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-29 03:17:43,785 Stage-1 map = 0%,  reduce = 0%
2021-10-29 03:17:55,677 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.22 sec
2021-10-29 03:18:06,205 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.24 sec
MapReduce Total cumulative CPU time: 5 seconds 240 msec
Ended Job = job_1635497035806_0005
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 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_1635497035806_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635497035806_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1635497035806_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-10-29 03:18:18,256 Stage-2 map = 0%,  reduce = 0%
2021-10-29 03:18:24,717 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.48 sec
2021-10-29 03:18:34,182 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 3.67 sec
MapReduce Total cumulative CPU time: 3 seconds 670 msec
Ended Job = job_1635497035806_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.24 sec   HDFS Read: 2716873 HDFS Write: 5649 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.67 sec   HDFS Read: 10573 HDFS Write: 4551 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 910 msec
```

```
US      0.6990019127795454
Ukraine 0.6641071417501995
United Arab Emirates    0.5719642837398818
United Kingdom  0.7858749992081098
Uruguay 0.6983035733657224
Uzbekistan      0.5597321424367172
Venezuela       0.7116071438150746
Vietnam 0.9302678603146758
Zambia  0.8024107137961047
Time taken: 63.398 seconds, Fetched: 160 row(s)
hive>
```

Q05: Find each country minimum temperature and order by ‘country’ name.

```
hive> select country, min(templow) from d_weather2020 GROUP BY country ORDER BY country;
Query ID = cloudera_20211029032424_a9f6bf62-5649-4737-a823-4f0544255eaa
Total jobs = 2
Launching Job 1 out of 2
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_1635497035806_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635497035806_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1635497035806_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-29 03:24:15,052 Stage-1 map = 0%,  reduce = 0%
2021-10-29 03:24:25,280 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.04 sec
2021-10-29 03:24:34,777 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.12 sec
MapReduce Total cumulative CPU time: 5 seconds 120 msec
Ended Job = job_1635497035806_0007
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 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_1635497035806_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635497035806_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1635497035806_0008
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-10-29 03:24:47,546 Stage-2 map = 0%,  reduce = 0%
2021-10-29 03:24:54,907 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.37 sec
2021-10-29 03:25:04,420 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 3.52 sec
MapReduce Total cumulative CPU time: 3 seconds 520 msec
Ended Job = job_1635497035806_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.12 sec   HDFS Read: 2716415 HDFS Write: 5009 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.52 sec   HDFS Read: 9935 HDFS Write: 2458 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 640 msec
OK
```

```
US      -25.05
Ukraine 14.86
United Arab Emirates    47.99
United Kingdom  26.52
Uruguay 46.39
Uzbekistan      19.12
Venezuela       64.33
Vietnam 45.99
Zambia  54.93
Time taken: 57.901 seconds, Fetched: 160 row(s)
hive>
```

Q06: Find each country maximum temperature and order by ‘icon’

```
hive> select country,max(temphigh),icon from d_weather2020 GROUP BY country,icon ORDER BY icon;
Query ID = cloudera_20211106112020_a111fd5c-056a-4ae0-824d-5eacf0e8cb28
Total jobs = 2
Launching Job 1 out of 2
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_1636219030821_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 11:21:04,509 Stage-1 map = 0%,  reduce = 0%
2021-11-06 11:21:13,196 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.05 sec
```

```
Argentina       75.41   99.94
Somalia 76.81   99.94
Sudan   62.04   99.94
Ghana   74.13   99.95
Central African Republic        62.26   99.95
Argentina       69.53   99.95
Central African Republic        65.38   99.96
Ghana   68.47   99.96
Cote d'Ivoire   74.18   99.97
Benin   67.19   99.97
Burkina Faso    77.66   99.98
Thailand        72.81   99.98
Thailand        70.95   99.99
"Korea  71.69   clear-day
"Gambia 110.63  clear-day
"Korea  68.52   cloudy
"Gambia 104.8   cloudy
"Korea  68.33   partly-cloudy-day
"Gambia 109.65  partly-cloudy-day
"Korea  66.03   rain
"Korea  29.79   snow
"Korea  59.65   wind
Time taken: 51.426 seconds, Fetched: 26244 row(s)
hive>
```

Q07: Create view for clear-day entry in ‘icon’.

```
hive> create view clr_day_table as select * from d_weather2020 where icon = 'clear-day';
OK
Time taken: 0.205 seconds
hive> show tables;
OK
clr_day_table
d_weather2020
Time taken: 0.017 seconds, Fetched: 2 row(s)
hive>
```

```
14900   "Korea   South" 4/1/2020        clear-day       48.93   28.5    0.48    1030.4  2.36    10.0    329.2   36      128
14905   "Korea   South" 9/1/2020        clear-day       47.75   25.01   0.57    1023.6  2.55    10.0    308.8   36      128
14910   "Korea   South" 14-01-2020      clear-day       39.57   22.7    0.56    1027.8  4.01    10.0    352.1   36      128
14916   "Korea   South" 20-01-2020      clear-day       44.71   30.9    0.36    1029.4  2.94    10.0    304.6   36      128
14927   "Korea   South" 31-01-2020      clear-day       46.09   26.46   0.72    1024.2  4.32    10.0    302.2   36      128
14930   "Korea   South" 3/2/2020        clear-day       45.51   16.41   0.48    1026.6  3.18    10.0    320.8   36      128
14931   "Korea   South" 4/2/2020        clear-day       32.45   11.18   0.4     1028.7  6.48    9.82    339.2   36      128
14935   "Korea   South" 8/2/2020        clear-day       44.38   26.78   0.56    1025.7  3.49    10.0    406.2   36      128
14936   "Korea   South" 9/2/2020        clear-day       52.35   28.05   0.61    1023.4  3.98    10.0    381.2   36      128
14937   "Korea   South" 10/2/2020       clear-day       58.0    38.52   0.49    1027.9  2.79    10.0    332.9   36      128
14944   "Korea   South" 17-02-2020      clear-day       37.68   21.07   0.58    1025.7  6.05    8.895   351.2   36      128
14949   "Korea   South" 22-02-2020      clear-day       50.22   28.15   0.48    1029.1  4.18    10.0    367.6   36      128
14950   "Korea   South" 23-02-2020      clear-day       61.1    38.47   0.29    1024.8  1.87    10.0    317.7   36      128
14960   "Korea   South" 4/3/2020        clear-day       46.55   21.21   0.51    1024.4  5.16    9.965   348.9   36      128
14961   "Korea   South" 5/3/2020        clear-day       55.69   33.44   0.42    1022.8  2.03    10.0    331.8   36      128
14963   "Korea   South" 7/3/2020        clear-day       58.59   34.91   0.7     1019.3  2.3     10.0    356.4   36      128
14966   "Korea   South" 10/3/2020       clear-day       46.58   25.71   0.58    1016.9  7.61    10.0    395.0   36      128
14967   "Korea   South" 11/3/2020       clear-day       59.89   34.15   0.37    1022.4  2.9     10.0    375.0   36      128
14969   "Korea   South" 13-03-2020      clear-day       48.91   32.73   0.43    1016.3  6.51    10.0    412.5   36      128
14970   "Korea   South" 14-03-2020      clear-day       49.7    25.48   0.51    1011.5  10.06   9.62    426.6   36      128
14971   "Korea   South" 15-03-2020      clear-day       52.79   29.26   0.46    1021.5  3.76    10.0    370.4   36      128
14972   "Korea   South" 16-03-2020      clear-day       59.03   33.31   0.59    1016.8  6.45    10.0    368.8   36      128
14973   "Korea   South" 17-03-2020      clear-day       64.58   44.32   0.45    1014.9  2.76    10.0    373.7   36      128
14976   "Korea   South" 20-03-2020      clear-day       71.69   46.82   0.48    1012.3  4.69    10.0    353.8   36      128
14977   "Korea   South" 21-03-2020      clear-day       66.39   36.72   0.5     1012.9  4.87    9.767   350.0   36      128
14978   "Korea   South" 22-03-2020      clear-day       62.22   34.19   0.39    1017.3  5.36    10.0    357.6   36      128
14979   "Korea   South" 23-03-2020      clear-day       64.27   36.24   0.52    1020.4  2.06    10.0    368.7   36      128
14980   "Korea   South" 24-03-2020      clear-day       67.83   44.26   0.61    1021.1  3.15    10.0    334.1   36      128
14989   "Korea   South" 2/4/2020        clear-day       63.84   36.83   0.32    1021.1  4.54    10.0    391.6   36      128
14990   "Korea   South" 3/4/2020        clear-day       61.59   28.63   0.42    1018.8  8.01    10.0    366.3   36      128
14991   "Korea   South" 4/4/2020        clear-day       58.31   28.39   0.26    1022.2  4.87    10.0    372.8   36      128
14992   "Korea   South" 5/4/2020        clear-day       62.62   36.67   0.39    1021.1  3.76    10.0    361.0   36      128
14993   "Korea   South" 6/4/2020        clear-day       66.29   37.4    0.39    1016.5  5.63    10.0    368.9   36      128
14994   "Korea   South" 7/4/2020        clear-day       59.38   31.32   0.4     1017.7  5.92    10.0    379.2   36      128
14995   "Korea   South" 8/4/2020        clear-day       59.65   38.35   0.45    1021.8  4.4     10.0    385.9   36      128
15000   "Korea   South" 13-04-2020      clear-day       68.16   38.18   0.37    1016.6  4.08    10.0    364.2   36      128
Time taken: 0.104 seconds, Fetched: 91 row(s)
hive>
```

Q08: Count how many clear day in ‘Aruba’.

```
hive> select count(*) from d_weather2020 where country = 'Aruba' and icon = 'clear-day';
Query ID = cloudera_20211106120000_13ae34af-29c5-44ed-a2d2-c1cf90d7b7a6
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1636219030821_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 12:00:15,189 Stage-1 map = 0%,  reduce = 0%
2021-11-06 12:00:22,547 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.38 sec
2021-11-06 12:00:29,930 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.42 sec
MapReduce Total cumulative CPU time: 2 seconds 420 msec
Ended Job = job_1636219030821_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.42 sec   HDFS Read: 2718380 HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 420 msec
OK
0
Time taken: 22.793 seconds, Fetched: 1 row(s)
hive>
```

Q09: Show which day and which country has highest and lowest ‘ozone’ level

```
hive> select max(ozo), min(ozo) from d_weather2020;
Query ID = cloudera_20211106120606_22e527ec-4a13-4b9e-a653-f4fd65316517
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1636219030821_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 12:06:25,689 Stage-1 map = 0%,  reduce = 0%
2021-11-06 12:06:32,022 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.33 sec
2021-11-06 12:06:40,485 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.36 sec
MapReduce Total cumulative CPU time: 2 seconds 360 msec
Ended Job = job_1636219030821_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.36 sec   HDFS Read: 2717678 HDFS Write: 15 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 360 msec
OK
426.6   -41.4545
Time taken: 22.774 seconds, Fetched: 1 row(s)
hive> select date, country from d_weather2020 where ozo = 426.6 or ozo = -41.4545;
OK
 South" "Korea
Time taken: 0.077 seconds, Fetched: 1 row(s)
hive>
```

Q10: Find the maximum and minimum temperature when we have visibility between 5 and 10.

```
hive> select max(temphigh), min(templow) from d_weather2020 where vis between 5 and 10;
Query ID = cloudera_20211106121414_41e5e631-246e-4c74-a17a-955446afc4f1
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1636219030821_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 12:14:38,741 Stage-1 map = 0%,  reduce = 0%
2021-11-06 12:14:46,180 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.8 sec
2021-11-06 12:14:53,514 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.88 sec
MapReduce Total cumulative CPU time: 2 seconds 880 msec
Ended Job = job_1636219030821_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.88 sec   HDFS Read: 2718927 HDFS Write: 13 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 880 msec
OK
110.63  11.18
Time taken: 21.613 seconds, Fetched: 1 row(s)
hive>
```

Q11: Calculate average humidity when the weather is in cloudy order by country.

```
hive> select avg(hum) from d_weather2020 where icon = 'cloudy' group by country;
Query ID = cloudera_20211106122020_b5c9a8a0-f222-4059-93d0-39c4fac9e4a5
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_1636219030821_0013, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0013/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 12:20:19,628 Stage-1 map = 0%,  reduce = 0%
2021-11-06 12:20:27,036 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.49 sec
2021-11-06 12:20:34,369 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.64 sec
MapReduce Total cumulative CPU time: 2 seconds 640 msec
Ended Job = job_1636219030821_0013
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.64 sec   HDFS Read: 2718999 HDFS Write: 37 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 640 msec
OK
0.17833333214124045
0.56142857670784
Time taken: 21.735 seconds, Fetched: 2 row(s)
```

Q12: Find which country has highest temperature and lowest temperature.

```
hive> select max(temphigh), min(templow) from d_weather2020;
Query ID = cloudera_20211106122828_beec9720-91fa-44e8-9707-c4d4d3cb0594
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1636219030821_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 12:29:02,623 Stage-1 map = 0%,  reduce = 0%
2021-11-06 12:29:10,065 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.46 sec
2021-11-06 12:29:18,404 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.49 sec
MapReduce Total cumulative CPU time: 2 seconds 490 msec
Ended Job = job_1636219030821_0016
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.49 sec   HDFS Read: 2717718 HDFS Write: 12 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 490 msec
OK
110.63  0.02
Time taken: 23.037 seconds, Fetched: 1 row(s)
hive> select country from d_weather2020 where temphigh = 110.63 and templow = 0.02;
OK
Time taken: 0.043 seconds
hive>
```

Q13: Display Longitude and latitude of the lowest visibly country.

```
hive> select min(vis) from d_weather2020;
Query ID = cloudera_20211106123636_df066979-b66e-483b-9a66-604df594d852
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1636219030821_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0017
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 12:36:12,763 Stage-1 map = 0%,  reduce = 0%
2021-11-06 12:36:19,087 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.24 sec
2021-11-06 12:36:26,439 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.27 sec
MapReduce Total cumulative CPU time: 2 seconds 270 msec
Ended Job = job_1636219030821_0017
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.27 sec   HDFS Read: 2717097 HDFS Write: 6 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 270 msec
OK
0.607
Time taken: 20.524 seconds, Fetched: 1 row(s)
hive> select lat, long, country from d_weather2020 where vis = 0.607;
OK
36      128.0   "Korea
Time taken: 0.049 seconds, Fetched: 1 row(s)
hive>
```

Q14: Calculate how many rain days in each country.

```
hive> select country,count(date) from d_weather2020 where icon = 'rain' GROUP BY country ORDER BY country;
Query ID = cloudera_20211106124040_56398f19-2ded-47c3-a645-a83698f6744b
Total jobs = 2
Launching Job 1 out of 2
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_1636219030821_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 12:41:00,454 Stage-1 map = 0%,  reduce = 0%
2021-11-06 12:41:06,736 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.34 sec
2021-11-06 12:41:14,037 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.32 sec
MapReduce Total cumulative CPU time: 2 seconds 320 msec
Ended Job = job_1636219030821_0018
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 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_1636219030821_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0019
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-11-06 12:41:21,801 Stage-2 map = 0%,  reduce = 0%
2021-11-06 12:41:27,061 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.7 sec
2021-11-06 12:41:33,327 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.72 sec
MapReduce Total cumulative CPU time: 1 seconds 720 msec
Ended Job = job_1636219030821_0019
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.32 sec   HDFS Read: 2717617 HDFS Write: 121 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.72 sec   HDFS Read: 5047 HDFS Write: 10 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 40 msec
OK
"Korea  29
Time taken: 39.844 seconds, Fetched: 1 row(s)
hive>
```

Q15: Group the country when it has 10 as ' visibility ' and find the average wind speed.

```
hive> select country, avg(windspeed) from d_weather2020 where vis = 10 GROUP BY country ORDER BY country;
Query ID = cloudera_20211106124747_e8df38c2-9fb9-4801-8f40-2381a87f7657
Total jobs = 2
Launching Job 1 out of 2
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_1636219030821_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-11-06 12:47:39,173 Stage-1 map = 0%,  reduce = 0%
2021-11-06 12:47:46,537 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.52 sec
2021-11-06 12:47:53,882 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.47 sec
MapReduce Total cumulative CPU time: 2 seconds 470 msec
Ended Job = job_1636219030821_0020
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 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_1636219030821_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1636219030821_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1636219030821_0021
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-11-06 12:48:01,611 Stage-2 map = 0%,  reduce = 0%
2021-11-06 12:48:06,917 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.71 sec
2021-11-06 12:48:14,223 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 1.76 sec
MapReduce Total cumulative CPU time: 1 seconds 760 msec
Ended Job = job_1636219030821_0021
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.47 sec   HDFS Read: 2718049 HDFS Write: 161 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.76 sec   HDFS Read: 5070 HDFS Write: 51 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 230 msec
OK
"Gambia 7.866216227814958
"Korea  4.049154949859834
Time taken: 41.854 seconds, Fetched: 2 row(s)
```