
Task 7: Write a query using a GROUP BY statement

Query:

```
hive> select place.city , count(*) from College group by place.city;
```

RunProof:

```
Query ID = anagave_20210503232029_28baeae0-941b-4dbe-ae8b-45934cc6e766
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_1620056543801_0076, Tracking URL = http://hadoop-
nn001.cs.okstate.edu:8088/proxy/application_1620056543801_0076/
Kill Command = /usr/local/hadoop-3.3.0/bin/mapred job -kill job_1620056543801_0076
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 1
2021-05-03 23:20:37,334 Stage-1 map = 0%, reduce = 0%
2021-05-03 23:20:43,636 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 18.86 sec
2021-05-03 23:20:44,687 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 24.34 sec
2021-05-03 23:20:48,890 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 27.88 sec
MapReduce Total cumulative CPU time: 27 seconds 880 msec
Ended Job = job_1620056543801_0076
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 1 Cumulative CPU: 27.88 sec HDFS Read: 45549 HDFS Write:
229 SUCCESS
Total MapReduce CPU Time Spent: 27 seconds 880 msec
OK
Columbus      1
East Lansing  1
Lubbock       1
PASADENA      1
Stillwater    1
Wichita       1
Time taken: 20.146 seconds, Fetched: 6 row(s)
```

Output:

```
Total MapReduce CPU Time Spent: 27 seconds 880 msec
OK
Columbus      1
East Lansing  1
Lubbock       1
PASADENA      1
Stillwater    1
Wichita       1
Time taken: 20.146 seconds, Fetched: 6 row(s)
```