Write an application using HiveQL for flight information system which will include
a. Creating, Dropping, and altering Database tables.
b. Creating an external Hive table.
c. Load table with data, insert new values and field in the table, Join tables with Hive
d. Create index on Flight Information Table
e. Find the average departure delay per day in 2008.

********Creating, Dropping, and altering Database tables********

```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-
log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> SHOW DATABASES;
OK
default
Time taken: 1.176 seconds, Fetched: 1 row(s)

hive> CREATE DATABASE IF NOT EXISTS dsbda35030;
OK
Time taken: 0.438 seconds

hive> USE dsbda35030;
OK
Time taken: 0.195 seconds

hive> SHOW TABLES;
OK
Time taken: 0.311 seconds

hive> CREATE TABLE IF NOT EXISTS employee(id int, name string, salary
string, destination string)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n'
    > STORED AS TEXTFILE;
OK
Time taken: 2.622 seconds

hive> LOAD DATA LOCAL INPATH
'/home/cloudera/workspace/ClouderaShared/sample_emp.txt'
    > OVERWRITE INTO TABLE employee;
Loading data to table dsbda35030.employee
Table dsbda35030.employee stats: [numFiles=1, numRows=0, totalSize=162,
rawDataSize=0]
OK
Time taken: 2.489 seconds

hive> SELECT * FROM employee;
OK
1201  Gopal 45000 Technical manager
1202  Manisha     45000 Proof reader
1203  Masthanvali 40000 Technical writer
```

```
1204  Kiran 40000 Hr Admin
1205  Kranthi     30000 Op Admin
NULL  NULL  NULL  NULL
Time taken: 0.942 seconds, Fetched: 6 row(s)

hive> ALTER TABLE employee
    > ADD COLUMNS(dept string);
OK
Time taken: 0.543 seconds

hive> DESCRIBE employee;
OK
id                      int
name                    string
salary                  float
department              string
dept                    string
Time taken: 0.111 seconds, Fetched: 5 row(s)

hive> ALTER TABLE employee
    > REPLACE COLUMNS (id int, name string, salary float, department
string);
OK
Time taken: 0.589 seconds

hive> DESCRIBE employee;
OK
id                      int
name                    string
salary                  float
department              string
Time taken: 0.37 seconds, Fetched: 4 row(s)

hive> ALTER TABLE employee
    > CHANGE salary salary string;
OK
Time taken: 0.261 seconds

hive> DESCRIBE employee;
OK
id                      int
name                    string
salary                  string
department              string
Time taken: 0.108 seconds, Fetched: 4 row(s)

hive> DROP TABLE employee;
OK
Time taken: 0.687 seconds

hive> SHOW TABLES;
OK
Time taken: 0.255 seconds
```

********b. Creating an external Hive table.********

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 11 items
drwxr-xr-x   - cloudera supergroup          0 2024-04-27 21:50
/HiveDirectory
drwxr-xr-x   - cloudera supergroup          0 2024-04-23 08:07
/MRinputfolder1
drwxr-xr-x   - cloudera supergroup          0 2024-04-23 08:25
/MRinputfolder2
drwxr-xr-x   - cloudera supergroup          0 2024-04-23 08:32
/MRoutfolder2
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15
/benchmarks
drwxr-xr-x   - hbase    supergroup          0 2024-04-27 21:25 /hbase
-rw-r--r--   1 cloudera supergroup        162 2024-04-27 10:41
/hivefolder
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs     supergroup          0 2024-04-21 23:07 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -mkdir /HiveDirectory2

[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 12 items
drwxr-xr-x   - cloudera supergroup          0 2024-04-27 21:50
/HiveDirectory
drwxr-xr-x   - cloudera supergroup          0 2024-04-27 23:29
/HiveDirectory2
drwxr-xr-x   - cloudera supergroup          0 2024-04-23 08:07
/MRinputfolder1
drwxr-xr-x   - cloudera supergroup          0 2024-04-23 08:25
/MRinputfolder2
drwxr-xr-x   - cloudera supergroup          0 2024-04-23 08:32
/MRoutfolder2
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15
/benchmarks
drwxr-xr-x   - hbase    supergroup          0 2024-04-27 21:25 /hbase
-rw-r--r--   1 cloudera supergroup        162 2024-04-27 10:41
/hivefolder
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs     supergroup          0 2024-04-21 23:07 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /var

[cloudera@quickstart ~]$ hdfs dfs -put
/home/cloudera/workspace/malmatta/sample_emp.txt /HiveDirectory2/

[cloudera@quickstart ~]$ hdfs dfs -ls /HiveDirectory2/
Found 1 items
-rw-r--r--   1 cloudera supergroup        162 2024-04-27 23:37
/HiveDirectory2/sample_emp.txt


hive> CREATE EXTERNAL TABLE IF NOT EXISTS emplist(id int, name string,
salary float, department string)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
```

```
    > LOCATION '/HiveDirectory2/';
OK
Time taken: 0.633 seconds

hive> SELECT * FROM emplist;
OK
1201  Gopal 45000.0     Technical manager
1202  Manisha     45000.0     Proof reader
1203  Masthanvali 40000.0     Technical writer
1204  Kiran 40000.0     Hr Admin
1205  Kranthi     30000.0     Op Admin
NULL  NULL  NULL  NULL
Time taken: 0.396 seconds, Fetched: 6 row(s)
```

********c. Load table with data, insert new values and field in the table, Join tables with Hive********

```
hive> CREATE TABLE flight_data(
    >    year INT,
    >    month INT,
    >    day INT,
    >    day_of_week INT,
    >    dep_time INT,
    >    crs_dep_time INT,
    >    arr_time INT,
    >    crs_arr_time INT,
    >    unique_carrier STRING,
    >    flight_num INT,
    >    tail_num STRING,
    >    actual_elapsed_time INT,
    >    crs_elapsed_time INT,
    >    air_time INT,
    >    arr_delay INT,
    >    dep_delay INT,
    >    origin STRING,
    >    dest STRING,
    >    distance INT,
    >    taxi_in INT,
    >    taxi_out INT,
    >    cancelled INT,
    >    cancellation_code STRING,
    >    diverted INT,
    >    carrier_delay STRING,
    >    weather_delay STRING,
    >    nas_delay STRING,
    >    security_delay STRING,
```

```
    >    late_aircraft_delay STRING
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 3.08 seconds

hive> show tables;
OK
emplist
flight_data
Time taken: 0.124 seconds, Fetched: 2 row(s)


hive> LOAD DATA LOCAL INPATH '/home/cloudera/workspace/malmatta/2008.csv'
OVERWRITE INTO TABLE flight_data;
Loading data to table dsbda35030.flight_data
Table dsbda35030.flight_data stats: [numFiles=1, numRows=0,
totalSize=689413044, rawDataSize=0]
OK
Time taken: 56.089 seconds

hive> SELECT * FROM flight_data limit 10;
OK
2008  1     3     4     2003  1955  2211  2225  WN    335   N712SW      128
      150   116   -14   8     IAD   TPA   810   4     8     0           0
      NA    NA    NA    NA    NA
2008  1     3     4     754   735   1002  1000  WN    3231  N772SW      128
      145   113   2     19    IAD   TPA   810   5     10    0           0
      NA    NA    NA    NA    NA
2008  1     3     4     628   620   804   750   WN    448   N428WN      96
      90    76    14    8     IND   BWI   515   3     17    0           0
      NA    NA    NA    NA    NA
2008  1     3     4     926   930   1054  1100  WN    1746  N612SW      88
      90    78    -6    -4    IND   BWI   515   3     7     0           0
      NA    NA    NA    NA    NA
2008  1     3     4     1829  1755  1959  1925  WN    3920  N464WN      90
      90    77    34    34    IND   BWI   515   3     10    0           0
      2     0     0     0     32
2008  1     3     4     1940  1915  2121  2110  WN    378   N726SW      101
      115   87    11    25    IND   JAX   688   4     10    0           0
      NA    NA    NA    NA    NA
2008  1     3     4     1937  1830  2037  1940  WN    509   N763SW      240
      250   230   57    67    IND   LAS   1591  3     7     0           0
      10    0     0     0     47
2008  1     3     4     1039  1040  1132  1150  WN    535   N428WN      233
      250   219   -18   -1    IND   LAS   1591  7     7     0           0
      NA    NA    NA    NA    NA
2008  1     3     4     617   615   652   650   WN    11    N689SW      95
      95    70    2     2     IND   MCI   451   6     19    0           0
      NA    NA    NA    NA    NA
2008  1     3     4     1620  1620  1639  1655  WN    810   N648SW      79
      95    70    -16   0     IND   MCI   451   3     6     0           0
      NA    NA    NA    NA    NA
Time taken: 1.685 seconds, Fetched: 10 row(s)
```

```
hive> SELECT avg(arr_delay) FROM flight_data
    > WHERE year=2008;
Query ID = cloudera_20240428015353_a9d35156-8e1d-4dba-bcdf-76295a7ca7af
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_1714291334023_0001, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1714291334023_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill
job_1714291334023_0001
Hadoop job information for Stage-1: number of mappers: 3; number of
reducers: 1
2024-04-28 01:54:16,285 Stage-1 map = 0%,  reduce = 0%
2024-04-28 01:55:16,810 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU
24.09 sec
2024-04-28 01:55:41,206 Stage-1 map = 22%,  reduce = 0%, Cumulative CPU
34.48 sec
2024-04-28 01:55:46,228 Stage-1 map = 56%,  reduce = 0%, Cumulative CPU
38.45 sec
2024-04-28 01:55:51,222 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
40.91 sec
2024-04-28 01:56:01,765 Stage-1 map = 100%,  reduce = 100%, Cumulative
CPU 42.88 sec
MapReduce Total cumulative CPU time: 42 seconds 880 msec
Ended Job = job_1714291334023_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 1   Cumulative CPU: 42.88 sec   HDFS Read:
689459682 HDFS Write: 17 SUCCESS
Total MapReduce CPU Time Spent: 42 seconds 880 msec
OK


8.16845238729114
Time taken: 160.072 seconds, Fetched: 1 row(s)


hive> SELECT avg(arr_delay) FROM flight_data WHERE month=1 AND
origin='SFO';
Query ID = cloudera_20240428015757_1b3bed1f-0f0a-439f-8622-a08d6aed533c
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_1714291334023_0002, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1714291334023_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill
job_1714291334023_0002
```

```
Hadoop job information for Stage-1: number of mappers: 3; number of
reducers: 1
2024-04-28 01:57:27,146 Stage-1 map = 0%,  reduce = 0%
2024-04-28 01:57:44,199 Stage-1 map = 11%,  reduce = 0%, Cumulative CPU
5.63 sec
2024-04-28 01:57:51,755 Stage-1 map = 22%,  reduce = 0%, Cumulative CPU
16.53 sec
2024-04-28 01:57:53,245 Stage-1 map = 44%,  reduce = 0%, Cumulative CPU
16.53 sec
2024-04-28 01:57:59,660 Stage-1 map = 56%,  reduce = 0%, Cumulative CPU
22.86 sec
2024-04-28 01:58:00,758 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU
22.97 sec
2024-04-28 01:58:01,833 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
26.09 sec
2024-04-28 01:58:09,303 Stage-1 map = 100%,  reduce = 100%, Cumulative
CPU 27.95 sec
MapReduce Total cumulative CPU time: 27 seconds 950 msec
Ended Job = job_1714291334023_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 1   Cumulative CPU: 27.95 sec   HDFS Read:
689460952 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 27 seconds 950 msec
OK


28.669403949068094
Time taken: 54.068 seconds, Fetched: 1 row(s)



hive> CREATE TABLE airports(
    >    name STRING,
    >    country STRING,
    >    area_code INT,
    >    code STRING)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 2.047 seconds
hive> LOAD DATA LOCAL INPATH
'/home/cloudera/workspace/malmatta/airports.csv' OVERWRITE INTO TABLE
airports;
Loading data to table dsbda35030.airports
Table dsbda35030.airports stats: [numFiles=1, numRows=0, totalSize=58888,
rawDataSize=0]
OK
Time taken: 1.751 seconds

hive> SELECT * FROM airports limit 10;
OK
Key West Nas /Boca Chica Field (private U. S. Navy )     US    67    NQX
A L Mangham Jr. Regional     US    67    OCH
AAF Heliport     US    67    AYE
Aberdeen RegionalUS    67    ABR
Abilene Regional US    67    ABI
Abraham Lincoln Capital      US    67    SPI
```

```
Acadiana RegionalUS    67    ARA
Accomack County  US    67    MFV
Ada Municial     US    67    ADT
Adak Island Ns   US    67    ADK
Time taken: 0.11 seconds, Fetched: 10 row(s)


hive> SELECT f.year, f.month, f.origin, a.name, a.code
    > FROM flight_data f JOIN airports a
    > ON (f.origin=a.code)
    > LIMIT 20;


Query ID = cloudera_20240428021515_b927486b-0608-4956-af48-1e5cd436921c
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20240428021515_b927486b-0608-
4956-af48-1e5cd436921c.log
2024-04-28 02:15:30    Starting to launch local task to process map join;
      maximum memory = 932184064
2024-04-28 02:15:31    Dump the side-table for tag: 1 with group count:
2361 into file: file:/tmp/cloudera/8171e060-1015-4126-8415-
1aef02e85d14/hive_2024-04-28_02-15-25_655_1739586963936126766-1/-local-
10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
2024-04-28 02:15:31    Uploaded 1 File to: file:/tmp/cloudera/8171e060-
1015-4126-8415-1aef02e85d14/hive_2024-04-28_02-15-
25_655_1739586963936126766-1/-local-10003/HashTable-Stage-3/MapJoin-
mapfile01--.hashtable (87932 bytes)
2024-04-28 02:15:31    End of local task; Time Taken: 1.176 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_1714291334023_0005, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1714291334023_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill
job_1714291334023_0005
Hadoop job information for Stage-3: number of mappers: 3; number of
reducers: 0
2024-04-28 02:15:40,856 Stage-3 map = 0%,  reduce = 0%
2024-04-28 02:15:49,397 Stage-3 map = 33%,  reduce = 0%, Cumulative CPU
1.94 sec
2024-04-28 02:15:50,466 Stage-3 map = 67%,  reduce = 0%, Cumulative CPU
3.99 sec
2024-04-28 02:15:51,506 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU
6.91 sec
MapReduce Total cumulative CPU time: 6 seconds 910 msec
Ended Job = job_1714291334023_0005
MapReduce Jobs Launched:
Stage-Stage-3: Map: 3   Cumulative CPU: 6.91 sec   HDFS Read: 51465 HDFS
Write: 2918 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 910 msec
OK


2008  1     IAD   Washington Dulles International   IAD
2008  1     IAD   Washington Dulles International   IAD
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
```

```
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     IND   Indianapolis International   IND
2008  1     ISP   Long Island MacArthur  ISP
Time taken: 27.984 seconds, Fetched: 20 row(s)
```

********d. Create index on Flight Information Table********

```
hive> CREATE INDEX flight_index
    > ON TABLE flight_data (origin)
    > AS 'COMPACT' WITH DEFERRED REBUILD;
OK
Time taken: 3.133 seconds
hive> CREATE INDEX airports_index
    > ON TABLE airports (code)
    > AS 'COMPACT' WITH DEFERRED REBUILD;
OK
Time taken: 1.087 seconds
```

********e. Find the average departure delay per day in 2008.********

```
hive> SELECT day, avg(dep_delay) from flight_data
    > GROUP BY day;
Query ID = cloudera_20240428022323_11b186e8-54d7-4aad-a14d-6fd619221af2
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 3
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_1714291334023_0006, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1714291334023_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill
job_1714291334023_0006
Hadoop job information for Stage-1: number of mappers: 3; number of
reducers: 3
2024-04-28 02:23:24,865 Stage-1 map = 0%,  reduce = 0%
```

```
2024-04-28 02:24:04,499 Stage-1 map = 22%,  reduce = 0%, Cumulative CPU
18.49 sec
2024-04-28 02:24:11,977 Stage-1 map = 44%,  reduce = 0%, Cumulative CPU
21.56 sec
2024-04-28 02:24:13,456 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU
22.28 sec
2024-04-28 02:24:19,366 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
29.01 sec
2024-04-28 02:24:29,175 Stage-1 map = 100%,  reduce = 100%, Cumulative
CPU 34.13 sec
MapReduce Total cumulative CPU time: 34 seconds 130 msec
Ended Job = job_1714291334023_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 3   Cumulative CPU: 34.13 sec   HDFS Read:
689468969 HDFS Write: 648 SUCCESS
Total MapReduce CPU Time Spent: 34 seconds 130 msec
OK
3     8.124487135326511
6     9.97993852855542
9     9.426870350020717
12    8.985977165641355
15    10.246923827478266
18    9.864850133863555
21    11.441171459650256
24    9.400883693910961
27    12.352245320910784
30    9.387892140428182
1     11.05625866543425
4     11.320772116708202
7     10.021070439195801
10    10.650646787599612
13    9.812501407004858
16    9.09674254229989
19    10.7723860939885
22    12.373879681331578
25    8.651045607912613
28    8.955568222994053
31    13.721784776902886
2     9.720089012810984
5     8.54570651955599
8     10.699898365829522
11    9.628271196830053
14    9.410912707934333
17    8.888446389114506
20    9.92333218448704
23    10.302617596186767
26    10.571840439669346
29    7.615862597090034
Time taken: 76.062 seconds, Fetched: 31 row(s)
```

```
hive> SELECT avg(arr_delay)
    > FROM flight_data
    > WHERE year=2008
    > LIMIT 20;
Query ID = cloudera_20240428021111_ae9a997f-86d9-48f9-a918-b2a3b962a648
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_1714291334023_0004, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1714291334023_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill
job_1714291334023_0004
Hadoop job information for Stage-1: number of mappers: 3; number of
reducers: 1
2024-04-28 02:11:21,114 Stage-1 map = 0%,  reduce = 0%
2024-04-28 02:11:45,517 Stage-1 map = 11%,  reduce = 0%, Cumulative CPU
17.62 sec
2024-04-28 02:11:55,403 Stage-1 map = 22%,  reduce = 0%, Cumulative CPU
21.42 sec
2024-04-28 02:12:03,717 Stage-1 map = 44%,  reduce = 0%, Cumulative CPU
28.69 sec
2024-04-28 02:12:07,281 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
30.65 sec
2024-04-28 02:12:18,098 Stage-1 map = 100%,  reduce = 100%, Cumulative
CPU 32.9 sec
MapReduce Total cumulative CPU time: 32 seconds 900 msec
Ended Job = job_1714291334023_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3  Reduce: 1   Cumulative CPU: 32.9 sec   HDFS Read:
689459835 HDFS Write: 17 SUCCESS
Total MapReduce CPU Time Spent: 32 seconds 900 msec
OK

8.16845238729114
Time taken: 69.524 seconds, Fetched: 1 row(s)
```

On hive shell: run a join query to find the average delay in January 2008 for each airport and to print out the airport's name:

```
hive> SELECT a.name, avg(f.arr_delay)
    > FROM flight_data f INNER JOIN airports a
    > ON (f.origin = a.code)
    > WHERE month = 1
    > GROUP BY name;
```

```
Query ID = cloudera_20240428023131_0bba1584-62a5-4c46-865c-3160916d3488
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20240428023131_0bba1584-62a5-
4c46-865c-3160916d3488.log
2024-04-28 02:31:44    Starting to launch local task to process map join;
      maximum memory = 932184064
2024-04-28 02:31:45    Dump the side-table for tag: 1 with group count:
2361 into file: file:/tmp/cloudera/cc27f4c9-767e-4b09-95f5-
9283f66d4d1f/hive_2024-04-28_02-31-39_989_7331736779930878585-1/-local-
10004/HashTable-Stage-2/MapJoin-mapfile01--.hashtable
2024-04-28 02:31:45    Uploaded 1 File to: file:/tmp/cloudera/cc27f4c9-
767e-4b09-95f5-9283f66d4d1f/hive_2024-04-28_02-31-
39_989_7331736779930878585-1/-local-10004/HashTable-Stage-2/MapJoin-
mapfile01--.hashtable (87932 bytes)
2024-04-28 02:31:45    End of local task; Time Taken: 1.172 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 3
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_1714291334023_0007, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1714291334023_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill
job_1714291334023_0007
Hadoop job information for Stage-2: number of mappers: 3; number of
reducers: 3
2024-04-28 02:31:53,419 Stage-2 map = 0%,  reduce = 0%
2024-04-28 02:32:25,574 Stage-2 map = 11%,  reduce = 0%, Cumulative CPU
19.33 sec
2024-04-28 02:32:27,737 Stage-2 map = 22%,  reduce = 0%, Cumulative CPU
19.77 sec
2024-04-28 02:32:33,284 Stage-2 map = 67%,  reduce = 0%, Cumulative CPU
22.36 sec
2024-04-28 02:32:40,619 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU
28.57 sec
2024-04-28 02:32:51,168 Stage-2 map = 100%,  reduce = 100%, Cumulative
CPU 33.96 sec
MapReduce Total cumulative CPU time: 33 seconds 960 msec
Ended Job = job_1714291334023_0007
MapReduce Jobs Launched:
Stage-Stage-2: Map: 3  Reduce: 3   Cumulative CPU: 33.96 sec   HDFS Read:
689479883 HDFS Write: 11732 SUCCESS
Total MapReduce CPU Time Spent: 33 seconds 960 msec
OK
Abilene Regional 15.013043478260869
Abraham Lincoln Capital      42.75912408759124
Albert J Ellis   6.441558441558442
Arcata     41.61363636363637
Austin-bergstrom International    4.82597523943004
Billings Logan International 4.6440677966101696
Boise Air Terminal/Gowen Field    16.406460296096906
Boundary Bay     23.26241134751773
Buffalo Niagara International     11.748849104859335
```

Charlotte Douglas International	6.4406990068754775
Charlottesville-Albemarle	4.871428571428571
Chicago Midway International	12.859635210150675
Chicago O'hare International	27.593595701423176
Chicago/Rockford International	2.3157894736842106
Columbus Metropolitan	9.018518518518519
Denver International	11.363121756905498