

→ Airlines project

set to print current db on command prompt

```
hive (default)> set hive.cli.print.current.db = true ;
```

change current table

```
hive (default)> USE surya_training ;
```

OK

Time taken: 0.029 seconds

show tables in current database

```
hive (surya_training)> SHOW TABLES ;
```

OK

nyse

stkvol

Time taken: 0.083 seconds, Fetched: 2 row(s)

```
hive (surya_training)>
```

create table airport

```
hive (surya_training)> CREATE TABLE airport (  
    > airport_id INT,  
    > name STRING,  
    > city STRING,  
    > country STRING,  
    > iata STRING,  
    > icao STRING,  
    > latitude DOUBLE,  
    > longitude DOUBLE,  
    > altitude INT,  
    > timezone DOUBLE,  
    > dst STRING,  
    > TZ STRING)  
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
```

```
STORED AS TEXTFILE;
```

OK

Time taken: 0.465 seconds

```
hive (surya_training)> SHOW TABLES ;
```

OK

airport

nyse

stkvol

Time taken: 0.218 seconds, Fetched: 3 row(s)

```
hive (surya_training)>
```

put airport_mod.dat file into hadoop fs training folder

```
[bigdatalab456422@ip-10-1-1-204 ~]$ hadoop fs -put airports_mod.dat training
```

load data local inpath 'airports_mod.dat' overwrite into table airport;

```
hive (surya_training)> LOAD DATA LOCAL INPATH 'airports_mod.dat'
OVERWRITE INTO TABLE airport ;
```

Loading data to table surya_training.airport

OK

Time taken: 1.824 seconds

```
hive (surya_training)>
```

→ # SHOW first 10 records from airport table

```
hive (surya_training)> SELECT * FROM airport LIMIT 10 ;
```

```
hive (surya_training)> SELECT * FROM airport LIMIT 10 ;
OK
1  Goroka Goroka Papua New Guinea GKA AYGA -6.081689 145.391881 5282 10.0 U Pacific/Port_Moresby
2  Madang Madang Papua New Guinea MAG AYMD -5.207083 145.7887 20 10.0 U Pacific/Port_Moresby
3  Mount Hagen Mount Hagen Papua New Guinea HGU AYMH -5.826789 144.295861 5388 10.0 U Pacific/Port_Moresby
4  Nadzab Nadzab Papua New Guinea LAE AYNZ -6.569828 146.726242 239 10.0 U Pacific/Port_Moresby
5  Port Moresby Jacksons Intl Port Moresby Papua New Guinea POM AYPY -9.443383 147.22005 146 10.0 U Pacific/Port_Moresby
6  Wewak Intl Wewak Papua New Guinea WWK AYWK -3.583828 143.669186 19 10.0 U Pacific/Port_Moresby
7  Narsarsuaq Narsarsuaq Greenland UAK BGBW 61.160517 -45.425978 112 -3.0 E America/Godthab
8  Nuuk Godthaab Greenland GOH BGGH 64.190922 -51.678064 283 -3.0 E America/Godthab
9  Sondre Stromfjord Sondrestrom Greenland SFJ BGFS 67.016969 -50.689325 165 -3.0 E America/Godthab
10 Thule Air Base Thule Greenland THU BGTL 76.531203 -68.703161 251 -4.0 E America/Thule
Time taken: 0.3 seconds, Fetched: 10 row(s)
hive (surya_training)>
```

→ # show count of all the airports in the table (or print count of all records in a table) and verify if the airport count is 8107

```
hive (surya_training)> SELECT count(airport_id) FROM airport ;
```

```
hive (surya_training)> SELECT count(airport_id) FROM airport ;
Query ID = bigdatalab456422_20230525053135_52645e2c-896f-40ea-ab75-6c136f9a5e84
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>
23/05/25 05:31:37 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/25 05:31:37 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_0553, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_0553/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_0553
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-25 05:32:47.370 Stage-1 map = 0%, reduce = 0%
2023-05-25 05:32:56.882 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.26 sec
2023-05-25 05:33:09.520 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.68 sec
MapReduce Total cumulative CPU time: 5 seconds 680 msec
Ended Job = job_1684866872278_0553
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.68 sec HDFS Read: 749029 HDFS Write: 104 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 680 msec
OK
8107
Time taken: 95.701 seconds, Fetched: 1 row(s)
hive (surya_training)>
```

→ # show count of airports in each country and verify if INDIA has 140 airports

```
hive (surya_training)> SELECT country, count(airport_id) FROM airport
GROUP BY country ;
```

```
hive (surya_training)> SELECT country, count(airport_id) FROM airport GROUP BY country ;
Query ID = bigdata1ab456422_20230525053624_96347201-dbf0-4360-b2e0-addc78d88bf6
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>
23/05/25 05:36:24 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/25 05:36:24 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_0566, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_0566/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_0566
2023-05-25 05:36:57,204 Stage-1 map = 0%, reduce = 0%
2023-05-25 05:37:06,566 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.76 sec
2023-05-25 05:37:15,946 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.03 sec
MapReduce Total cumulative CPU time: 7 seconds 30 msec
Ended Job = job_1684866872278_0566
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.03 sec HDFS Read: 749448 HDFS Write: 6166 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 30 msec
OK
Fogo Island      1
Niuafo'ou Island 1
Santiago Island  1
Sheffield        1
Afghanistan      21
Albania           1
Algeria           44
American Samoa   2
Hungary           13
Iceland           22
India             140
Indonesia         114
Tanzania          04
Western Sahara   1
Yemen            11
Zambia           13
Zimbabwe          16
Time taken: 53.783 seconds, Fetched: 244 row(s)
hive (surya_training)>
```

→ # create table #2 table airlines

```
CREATE TABLE airlines(
  airline_id INT,
  name STRING,
  alias STRING,
  iata STRING,
  icao STRING,
  callsign STRING,
  country STRING,
  active STRING)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE;
```

```
hive (surya_training)> CREATE TABLE airlines(
>
>   airline_id INT,
>   name STRING,
>   alias STRING,
>   iata STRING,
>   icao STRING,
>   callsign STRING,
>   country STRING,
>   active STRING)
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE;
OK
Time taken: 0.106 seconds
hive (surya_training)>
```

LOAD data local inpath

```
hive (surya_training)> LOAD DATA LOCAL INPATH 'Final_airlines'  
OVERWRITE INTO TABLE airlines;
```

```
hive (surya_training)> LOAD DATA LOCAL INPATH 'Final_airlines' OVERWRITE INTO TABLE airlines;  
Loading data to table surya_training.airlines  
OK  
Time taken: 0.902 seconds  
hive (surya_training)>
```

→ # find count of airlines in the table airlines and verify if count is 6048

```
hive (surya_training)> SELECT count(airline_id) FROM airlines ;
```

```
hive (surya_training)> SELECT count(airline_id) FROM airlines ;  
Query ID = bigdatalab456422_20230525060929_4cd6f72b-198a-4b68-9456-c4b3ac05d51e  
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>  
23/05/25 06:09:30 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032  
23/05/25 06:09:30 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032  
Starting Job = job_1684866872278_0649, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_0649/  
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_0649  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2023-05-25 06:09:39,522 Stage-1 map = 0%, reduce = 0%  
2023-05-25 06:09:48,776 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.34 sec  
2023-05-25 06:10:01,217 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.07 sec  
MapReduce Total cumulative CPU time: 6 seconds 70 msec  
Ended Job = job_1684866872278_0649  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.07 sec HDFS Read: 325019 HDFS Write: 104 HDFS EC Read: 0 SUCCESS  
Total MapReduce CPU Time Spent: 6 seconds 70 msec  
OK  
6048  
Time taken: 33.722 seconds, Fetched: 1 row(s)  
hive (surya_training)>
```

create table#3 table routes

```
CREATE TABLE routes(  
  airline_iata STRING,  
  airline_id STRING,  
  src_airport_iata STRING,  
  src_airport_id INT,  
  dest_airport_iata STRING,  
  dest_airport_id INT,  
  codeshare STRING,  
  stop INT,  
  equipment STRING)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE;
```

```
hive (surya_training)> CREATE TABLE routes(
>
> airline_iata STRING,
>
> airline_id STRING,
>
> src_airport_iata STRING,
>
> src_airport_id INT,
>
> dest_airport_iata STRING,
>
> dest_airport_id INT,
>
> codeshare STRING,
>
> stop INT,
>
> equipment STRING)
>
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE;
OK
Time taken: 0.147 seconds
hive (surya_training)>
```

#

LOAD DATA LOCAL INPATH 'routes.dat' OVERWRITE INTO TABLE routes;

```
hive (surya_training)> LOAD DATA LOCAL INPATH 'routes.dat' OVERWRITE INTO TABLE routes;
Loading data to table surya_training.routes
OK
Time taken: 0.86 seconds
hive (surya_training)>
```

→ # show first 10 records from table routes

hive (surya_training)> SELECT * FROM routes LIMIT 10;

```
hive (surya_training)> SELECT * FROM routes LIMIT 10;
OK
2B 410 AER 2965 KZN 2990 0 CR2
2B 410 ASF 2966 KZN 2990 0 CR2
2B 410 ASF 2966 MRV 2962 0 CR2
2B 410 CEK 2968 KZN 2990 0 CR2
2B 410 CEK 2968 OVB 4078 0 CR2
2B 410 DME 4029 KZN 2990 0 CR2
2B 410 DME 4029 NBC 6969 0 CR2
2B 410 DME 4029 TKG NULL 0 CR2
2B 410 DME 4029 UUA 6160 0 CR2
2B 410 EGO 6156 KGD 2952 0 CR2
Time taken: 0.099 seconds, Fetched: 10 row(s)
hive (surya_training)>
```

→ # count number of routes (or show records in table routes) and verify if count is 67663

hive (surya_training)> SELECT count(*) FROM routes ;

```
hive (surya_training)> SELECT count(*) FROM routes ;
Query ID = bigdatalab456422_20230525063147_1c76a77f-908a-4d53-8a85-6c20b371ceb6
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>
23/05/25 06:31:49 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/25 06:31:49 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_0778, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_0778/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_0778
2023-05-25 06:33:31,889 Stage-1 map = 0%, reduce = 0%
2023-05-25 06:34:32,571 Stage-1 map = 0%, reduce = 0%
2023-05-25 06:35:04,452 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.9 sec
2023-05-25 06:35:58,052 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 4.49 sec
2023-05-25 06:36:04,205 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.85 sec
MapReduce Total cumulative CPU time: 5 seconds 850 msec
Ended Job = job_1684866872278_0778
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.85 sec HDFS Read: 2384797 HDFS Write: 105 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 850 msec
OK
67663
Time taken: 258.775 seconds, Fetched: 1 row(s)
hive (surya_training)>
```

→ assignment/project problems

a. Find list of Airports operating in the Country India

```
hive (surya_training)> SELECT * FROM airport WHERE country='India';
```

```
hive (surya_training)> SELECT * FROM airport WHERE country='India';
Query ID = bigdatalab456422_20230525070518_02016b60-343d-4fa7-ae4c-8f1906d1f22e
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/05/25 07:05:18 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/25 07:05:18 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_0947, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_0947/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_0947
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-05-25 07:05:31,006 Stage-1 map = 0%, reduce = 0%
2023-05-25 07:05:59,665 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.67 sec
MapReduce Total cumulative CPU time: 3 seconds 670 msec
Ended Job = job_1684866872278_0947
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 3.67 sec HDFS Read: 745753 HDFS Write: 13218 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 670 msec
OK
2994 Ahmedabad Ahmedabad India AMD VAAH 23.077242 72.63465 189 5.5 N Asia/Calcutta
2995 Akola Akola India AKD VAAK 20.699006 77.058628 999 5.5 N Asia/Calcutta
2996 Aurangabad Aurangabad India IXU VAAU 19.862728 75.398114 1911 5.5 N Asia/Calcutta
2997 Chhatrapati Shivaji Terminus Mumbai India RMH USRR 19.008666 77.867019 17 5.5 N Asia/Calcutta
2998 Mormugao-Goa Port Mormugao India NULL 15.4 73.8 0 5.5 N Asia/Calcutta
9301 Cochin Port Cochin India NULL 9.58 76.14 0 5.5 N Asia/Calcutta
9489 Helipad of Viraj Group Boisar India NULL 19.779 72.805 50 5.5 U Asia/Calcutta
Time taken: 43.501 seconds, Fetched: 140 row(s)
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result10' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' SELECT * FROM airport WHERE country='India';
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result10' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
> SELECT * FROM airport WHERE country='India';
Query ID = bigdatalab456422_20230526094529_63e15b37-1f16-4fa0-a6a2-c1ca3f24edbc
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
23/05/26 09:45:29 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 09:45:29 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2280, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2280/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2280
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-05-26 09:45:39,269 Stage-1 map = 0%, reduce = 0%
2023-05-26 09:45:55,551 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.27 sec
MapReduce Total cumulative CPU time: 3 seconds 270 msec
Ended Job = job_1684866872278_2280
Stage-3 is selected by condition resolver.
Stage-2 is filtered out by condition resolver.
Stage-4 is filtered out by condition resolver.
Moving data to directory hdfs://nameservice1/user/bigdatalab456422/hive/airport_project/result10/.hive-staging_hive_2023-05-26_09-45-29_887_7098243162571204640-1/-ext-1
0000
Moving data to directory hive/airport_project/result10
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 3.27 sec HDFS Read: 745309 HDFS Write: 11451 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 270 msec
OK
Time taken: 26.787 seconds
hive (surya_training)>
```

Back

Edit file

Refresh

View as binary

Download

Last modified

05/26/2023 3:15 PM

+05:30

User

bigdatalab456422

Group

bigdatalab456422

Size

11.18 KB

Mode

100755

Home

Page1to3of3

/ user / bigdatalab456422 / hive / airport_project / result10 / 000000_0

2994, Ahmedabad, Ahmedabad, India, AMD, VAAH, 23.877242, 72.63465, 189, 5.5, N, Asia/Calcutta

2995, Akola, Akola, India, AKD, VAAK, 20.699006, 77.058628, 999, 5.5, N, Asia/Calcutta

2996, Aurangabad, Aurangabad, India, IXU, VAAU, 19.862728, 75.398114, 1911, 5.5, N, Asia/Calcutta

2997, Chhatrapati Shivaji Intl, Mumbai, India, BOM, VABB, 19.088686, 72.867919, 37, 5.5, N, Asia/Calcutta

2998, Bilaspur, Bilaspur, India, PAB, VABI, 21.9884, 82.110983, 899, 5.5, N, Asia/Calcutta

2999, Bhuj, Bhuj, India, BHJ, VABJ, 23.287828, 69.670147, 268, 5.5, N, Asia/Calcutta

3000, Belgaum, Belgaum, India, IXG, VABM, 15.859286, 74.618292, 2487, 5.5, N, Asia/Calcutta

3001, Vadodara, Baroda, India, BDQ, VABO, 22.336164, 73.226289, 129, 5.5, N, Asia/Calcutta

3002, Bhopal, Bhopal, India, BHO, VABP, 23.287467, 77.337375, 1719, 5.5, N, Asia/Calcutta

3003, Bhavnagar, Bhaunagar, India, BHU, VABV, 21.752206, 72.185181, 44, 5.5, N, Asia/Calcutta

3004, Daman, Daman, India, NMB, VADN, 20.434364, 72.843206, 33, 5.5, N, Asia/Calcutta

3005, Deesa, Deesa, India, , VADS, 24.267936, 72.204433, 485, 5.5, N, Asia/Calcutta

3006, Guna, Guna, India, , VAGN, 24.654681, 77.347347, 1600, 5.5, N, Asia/Calcutta

3007, Goa, Goa, India, GOI, VAGO, 15.380833, 73.831422, 184, 5.5, N, Asia/Calcutta

3008, Devi Ahilyabai Holkar, Indore, India, IDR, VAID, 22.721786, 75.801086, 1850, 5.5, N, Asia/Calcutta

3009, Jabalpur, Jabalpur, India, JLR, VAJB, 23.177817, 80.052047, 1624, 5.5, N, Asia/Calcutta

3010, Jamnagar, Jamnagar, India, JGA, VAJM, 22.465522, 70.012556, 69, 5.5, N, Asia/Calcutta

3011, Kandla, Kandla, India, IXY, VAKE, 23.112719, 70.100289, 96, 5.5, N, Asia/Calcutta

3012, Khairabad, Khairabad, India, HIR, VAKI, 24.817107, 70.018507, 720, 5.5, N, Asia/Calcutta

sandeep's query

hive (surya_training)> INSERT OVERWRITE DIRECTORY

'hive/airport_project/result11' ROW FORMAT DELIMITED FIELDS

TERMINATED BY ',' SELECT * FROM airport WHERE

trim(upper(country)) = 'INDIA';

hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result11' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

> SELECT * FROM airport WHERE trim(upper(country)) = 'INDIA';

Query ID = bigdatalab456422_20230526094713_70802c94-0031-4152-9ff6-1422e82b72fc

Total jobs = 3

Launching Job 1 out of 3

Number of reduce tasks is set to 0 since there's no reduce operator

23/05/26 09:47:13 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

23/05/26 09:47:13 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

Starting Job = job_1684866872278_2290, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2290/

Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2290

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2023-05-26 09:47:27,469 Stage-1 map = 0%, reduce = 0%

2023-05-26 09:47:40,721 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.47 sec

MapReduce Total cumulative CPU time: 3 seconds 470 msec

Ended Job = job_1684866872278_2290

Stage-3 is selected by condition resolver.

Stage-2 is filtered out by condition resolver.

Stage-4 is filtered out by condition resolver.

Moving data to directory hdfs://nameservice1/user/bigdatalab456422/hive/airport_project/result11/.hive-staging_hive_2023-05-26_09-47-13_692_7545859380095432779-1/-ext-1

0000

Moving data to directory hive/airport_project/result11

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 3.47 sec HDFS Read: 745332 HDFS Write: 11451 HDFS EC Read: 0 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 470 msec

OK

Time taken: 28.127 seconds

hive (surya_training)>

Back

Edit file

Refresh

View as binary

Download

Last modified

05/26/2023 3:17 PM

+05:30

User

bigdatalab456422

Group

bigdatalab456422

Size

11.18 KB

Mode

100755

Home

Page1to3of3

/ user / bigdatalab456422 / hive / airport_project / result11 / 000000_0

2994, Ahmedabad, Ahmedabad, India, AMD, VAAH, 23.877242, 72.63465, 189, 5.5, N, Asia/Calcutta

2995, Akola, Akola, India, AKD, VAAK, 20.699006, 77.058628, 999, 5.5, N, Asia/Calcutta

2996, Aurangabad, Aurangabad, India, IXU, VAAU, 19.862728, 75.398114, 1911, 5.5, N, Asia/Calcutta

2997, Chhatrapati Shivaji Intl, Mumbai, India, BOM, VABB, 19.088686, 72.867919, 37, 5.5, N, Asia/Calcutta

2998, Bilaspur, Bilaspur, India, PAB, VABI, 21.9884, 82.110983, 899, 5.5, N, Asia/Calcutta

2999, Bhuj, Bhuj, India, BHJ, VABJ, 23.287828, 69.670147, 268, 5.5, N, Asia/Calcutta

3000, Belgaum, Belgaum, India, IXG, VABM, 15.859286, 74.618292, 2487, 5.5, N, Asia/Calcutta

3001, Vadodara, Baroda, India, BDQ, VABO, 22.336164, 73.226289, 129, 5.5, N, Asia/Calcutta

3002, Bhopal, Bhopal, India, BHO, VABP, 23.287467, 77.337375, 1719, 5.5, N, Asia/Calcutta

3003, Bhavnagar, Bhaunagar, India, BHU, VABV, 21.752206, 72.185181, 44, 5.5, N, Asia/Calcutta

3004, Daman, Daman, India, NMB, VADN, 20.434364, 72.843206, 33, 5.5, N, Asia/Calcutta

3005, Deesa, Deesa, India, , VADS, 24.267936, 72.204433, 485, 5.5, N, Asia/Calcutta

3006, Guna, Guna, India, , VAGN, 24.654681, 77.347347, 1600, 5.5, N, Asia/Calcutta

3007, Goa, Goa, India, GOI, VAGO, 15.380833, 73.831422, 184, 5.5, N, Asia/Calcutta

3008, Devi Ahilyabai Holkar, Indore, India, IDR, VAID, 22.721786, 75.801086, 1850, 5.5, N, Asia/Calcutta

3009, Jabalpur, Jabalpur, India, JLR, VAJB, 23.177817, 80.052047, 1624, 5.5, N, Asia/Calcutta

3010, Jamnagar, Jamnagar, India, JGA, VAJM, 22.465522, 70.012556, 69, 5.5, N, Asia/Calcutta

3011, Kandla, Kandla, India, IXY, VAKE, 23.112719, 70.100289, 96, 5.5, N, Asia/Calcutta

3012, Khairabad, Khairabad, India, HIR, VAKI, 24.817107, 70.018507, 720, 5.5, N, Asia/Calcutta

b. Find the list of Airlines having zero stops

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY  
'hive/airport_project/result20' ROW FORMAT DELIMITED FIELDS  
TERMINATED BY ',' SELECT distinct(al.name), rt.stop FROM  
airlines al JOIN routes rt ON al.airline_id=rt.airline_id WHERE  
rt.stop=0;
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result20' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
>  
> SELECT distinct(al.name), rt.stop FROM airlines al JOIN routes rt ON al.airline_id=rt.airline_id WHERE rt.stop=0;  
Query ID = bigdatalab456422_20230526095351_a4d77fed-d170-4ee7-b41b-5818dfbd5bc0  
Total jobs = 1  
SLF4J: Class path contains multiple SLF4J bindings.  
  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
2023-05-26 09:53:59,169 main ERROR RandomAccessFileManager (/var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425/bigdatalab456422_20230526095351_a4d77fed-d  
170-4ee7-b41b-5818dfbd5bc0) java.io.IOException: Could not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425 java.io.IOException: Could  
not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425  
    at org.apache.logging.log4j.core.util.FileUtils.mkDir(FileUtils.java:119)  
    at org.apache.logging.log4j.core.util.FileUtils.makeParentDirs(FileUtils.java:136)  
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager.createAppender(RandomAccessFileManager.java:109)  
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager.createAppender(RandomAccessFileManager.java:109)  
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager.createAppender(RandomAccessFileManager.java:109)  
    at org.apache.hadoop.hive.q1.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)  
    ... 35 more  
  
Execution completed successfully  
MapredLocal task succeeded  
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>  
23/05/26 09:54:02 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032  
23/05/26 09:54:02 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032  
Starting Job = job_1684866872278_2323, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2323/  
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2323  
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1  
2023-05-26 09:54:09,873 Stage-2 map = 0%, reduce = 0%  
2023-05-26 09:54:19,121 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 6.3 sec  
2023-05-26 09:54:26,362 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 9.73 sec  
MapReduce Total cumulative CPU time: 9 seconds 730 msec  
Ended Job = job_1684866872278_2323  
Moving data to directory hive/airport_project/result20  
MapReduce Jobs Launched:  
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 9.73 sec HDFS Read: 2388697 HDFS Write: 9414 HDFS EC Read: 0 SUCCESS  
Total MapReduce CPU Time Spent: 9 seconds 730 msec  
OK  
Time taken: 35.619 seconds  
hive (surya_training)>
```

[Back](#)[Home](#)

Page 1 to 3 of 3

[Edit file](#)[Refresh](#)[View as binary](#)[Download](#)

Last modified
05/26/2023 3:24 PM
+05:30
User
bigdatalab456422
Group
bigdatalab456422
Size
9.19 KB
Mode
100755

/ user / bigdatalab456422 / hive / airport_project / result20 / 000000_0
40-Mile Air,0
ABSA - Aerolinas Brasileiras,0
ALAK,0
Abaet,0
Abu Dhabi Amiri Flight,0
Adria Airways,0
Aegean Airlines,0
Aer Lingus,0
Aereonautica militare,0
Aero Condor Peru,0
Aero Flight,0
Aero Lanka,0
Aero-Service,0
AeroMexico,0
AeroWorld ,0
Aerocondor,0
Aeroflot Russian Airlines,0
Aeroflot Nord,0

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY  
'hive/airport_project/result21' ROW FORMAT DELIMITED FIELDS  
TERMINATED BY ',' SELECT DISTINCT(name) FROM airlines a JOIN  
routes r ON a.airline_id=r.airline_id WHERE stop=0;
```



```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result21' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT DISTINCT(name)
>
> FROM airlines a
>
> JOIN routes r
>
> ON a.airline_id=r.airline_id
>
> WHERE stop=0;
Query ID = bigdatalab456422_20230526095711_4087f761-abb1-4e52-9a0d-b935e31b27b1
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2023-05-26 09:57:18,788 main ERROR RandomAccessFileManager (/var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425/bigdatalab456422_20230526095711_4087f761-a
bb1-4e52-9a0d-b935e31b27b1) java.io.IOException: Could not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425 java.io.IOException: Could
not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425
    at org.apache.logging.log4j.core.util.FileUtils.mkDir(FileUtils.java:119)
    at org.apache.logging.log4j.core.util.FileUtils.makeParentDirs(FileUtils.java:136)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:198)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:198)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:198)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:198)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:198)
    at org.apache.hadoop.hive.q1.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)
    ... 35 more

Execution completed successfully
MapredLocal task succeeded
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>
23/05/25 10:19:56 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/25 10:19:56 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_1245, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1245/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_1245
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-05-25 10:20:05,727 Stage-2 map = 0%, reduce = 0%
2023-05-25 10:20:14,964 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 5.06 sec
2023-05-25 10:20:24,173 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 8.03 sec
MapReduce Total cumulative CPU time: 8 seconds 30 msec
Ended Job = job_1684866872278_1245
Moving data to directory hive/airport_project/result21
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 8.03 sec HDFS Read: 2388531 HDFS Write: 9414 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 30 msec
OK
Time taken: 41.127 seconds
hive (surya_training)>
```

[Back](#)
[Home](#)
Page 1 to 3 of 3

[Edit file](#)
[Refresh](#)
[View as binary](#)
[Download](#)

Last modified
05/25/2023 3:50 PM
+05:30
User
bigdatalab456422
Group
bigdatalab456422
Size
9.19 KB
Mode
100755

40-Mile Air,0
ABSA - Aerolinhas Brasileiras,0
ALAK,0
Abaet,0
Abu Dhabi Amiri Flight,0
Adria Airways,0
Aegean Airlines,0
Aer Lingus,0
Aereonautica militare,0
Aero Condor Peru,0
Aero Flight,0
Aero Lanka,0
Aero-Service,0
AeroMéxico,0
AeroWorld ,0
Aerocondor,0
Aeroflot Russian Airlines,0
Aeroflot-Nord,0

c. List of Airlines operating with code share

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result30' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ','
```

```
SELECT distinct(al.airline_id), al.name, rt.codeshare FROM
airlines al JOIN routes rt ON al.airline_id=rt.airline_id WHERE
rt.codeshare='Y';
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result30' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
> SELECT distinct(al.airline_id), al.name, rt.codeshare FROM airlines al JOIN routes rt ON al.airline_id=rt.airline_id WHERE rt.codeshare='Y';
Query ID = bigdatalab456422_20230526103256_f9d5ba41-c975-4dc8-9380-5775315ec97a
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2023-05-26 10:33:03,395 main ERROR RandomAccessFileManager (/var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425/bigdatalab456422_20230526103256_f9d5ba41-c
975-4dc8-9380-5775315ec97a) java.io.IOException: Could not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425 java.io.IOException: Could
not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425
    at org.apache.logging.log4j.core.util.FileUtils.mkdir(FileUtils.java:119)
    at org.apache.logging.log4j.core.util.FileUtils.makeParentDirs(FileUtils.java:136)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createAppender(RandomAccessFileManagerFactory.java:108)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createAppender(RandomAccessFileManagerFactory.java:108)
    at org.apache.hadoop.hive ql.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)
    ... 35 more

Execution completed successfully
MapredLocal task succeeded
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>
23/05/26 10:33:06 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 10:33:06 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2455, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2455/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2455
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-05-26 10:33:20,177 Stage-2 map = 0%, reduce = 0%
2023-05-26 10:33:32,687 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 6.25 sec
2023-05-26 10:33:40,847 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 9.45 sec
MapReduce Total cumulative CPU time: 9 seconds 450 msec
Ended Job = job_1684866872278_2455
Moving data to directory hive/airport_project/result30
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 9.45 sec HDFS Read: 2389243 HDFS Write: 3233 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 450 msec
OK
Time taken: 46.587 seconds
hive (surya_training)>
```

[Back](#)
[Home](#)

Page to of 1

[Edit file](#)
[Refresh](#)
[View as binary](#)
[Download](#)

Last modified
05/26/2023 4:03 PM
+05:30
User
bigdatalab456422
Group
bigdatalab456422
Size
3.16 KB
Mode
100755

24,American Airlines,Y
28,Asiana Airlines,Y
90,Air Europa,Y
96,Aegean Airlines,Y
116,Air Italy,Y
130,Aeroflot Russian Airlines,Y
137,Air France,Y
197,Azerbaijan Airlines,Y
214,Air Berlin,Y
218,Air India Limited,Y
225,Air Tahiti Nui,Y
240,Air One,Y
241,Air Sahara,Y
242,Air Malta,Y
316,Air Macau,Y
319,Air Seychelles,Y
321,AeroM xico,Y
324,All Nippon Airways,Y
328,Air Canada,Y

sandeep's query

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result31' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' SELECT DISTINCT(name) FROM airlines a JOIN
routes r ON a.airline_id=r.airline_id WHERE
TRIM(UPPER(codeshare))='Y';
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result31' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT DISTINCT(name) FROM airlines a JOIN routes r ON a.airline_id=r.airline_id WHERE TRIM(UPPER(codeshare))='Y';
Query ID = bigdatalab456422_20230526112634_e6c6304b-0307-42e0-a01a-5d4bd8f7bdf4
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2023-05-26 11:26:41,636 main ERROR RandomAccessFileManager (/var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425/bigdatalab456422_20230526112634_e6c6304b-0
307-42e0-a01a-5d4bd8f7bdf4) java.io.IOException: Could not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425 java.io.IOException: Could
not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425
    at org.apache.logging.log4j.core.util.FileUtils.mkdir(FileUtils.java:119)
    at org.apache.logging.log4j.core.util.FileUtils.makeParentDirs(FileUtils.java:136)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:198)
    at java.lang.reflect.Method.invoke(Method.java:496)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:313)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:227)

Execution completed successfully
MapredLocal task succeeded
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>
23/05/26 11:26:44 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 11:26:44 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2504, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2504/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2504
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-05-26 11:26:53,041 Stage-2 map = 0%, reduce = 0%
2023-05-26 11:27:03,578 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 5.99 sec
2023-05-26 11:27:13,952 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 10.0 sec
MapReduce Total cumulative CPU time: 10 seconds 0 msec
Ended Job = job_1684866872278_2504
Moving data to directory hive/airport_project/result31
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 10.0 sec HDFS Read: 2388203 HDFS Write: 2245 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 0 msec
OK
Time taken: 42.289 seconds
hive (surya_training)>
```

[Back](#)
[Home](#)
Page to of 1

[Edit file](#)
[Refresh](#)
[View as binary](#)
[Download](#)

Last modified
 05/26/2023 4:57 PM
 +05:30
 User
 bigdatalab456422
 Group
 bigdatalab456422
 Size
 2.19 KB
 Mode
 100755

/ user / bigdatalab456422 / hive / airport_project / result31 / 000000_0
 Abu Dhabi Amiri Flight
 Aegean Airlines
 Aer Lingus
 AeroMéxico
 Aeroflot Russian Airlines
 Aeroflot-Nord
 Aerolane
 Aerolineas Argentinas
 Air Astana
 Air Baltic
 Air Berlin
 Air Canada
 Air Caraïbes
 Air China
 Air Europa
 Air France
 Air India Limited
 Air Italy
 Air Macau

d. Which country (or) territory having highest Airports

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result4' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' SELECT ap.country, count(ap.country) FROM
airport ap GROUP BY country;
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result4' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT ap.country, count(ap.country) FROM airport ap GROUP BY country;
Query ID = bigdata1ab456422_20230525133642_0a9ed1e8-47ba-4055-8218-a9bfc1fbd6f
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>
23/05/25 13:36:42 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/25 13:36:42 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_1296, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_1296/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_1296
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-25 13:36:51,217 Stage-1 map = 0%, reduce = 0%
2023-05-25 13:36:58,356 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.55 sec
2023-05-25 13:37:17,712 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.43 sec
MapReduce Total cumulative CPU time: 6 seconds 430 msec
Ended Job = job_1684866872278_1296
Moving data to directory hive/airport_project/result4
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.43 sec HDFS Read: 748988 HDFS Write: 3151 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 430 msec
OK
Time taken: 37.202 seconds
hive (surya_training)>
```

sandeep's query

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result47' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' SELECT country, count(airport_id) AS cnt FROM
airport GROUP BY country HAVING cnt IN (SELECT max(total) FROM
(SELECT count(airport_id) AS total FROM airport GROUP BY
country) AS abc);
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result47' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT country, count(airport_id) AS cnt FROM airport GROUP BY country HAVING cnt IN
>
> (SELECT max(total) FROM
>
> (SELECT count(airport_id) AS total FROM airport GROUP BY country) AS abc);
Query ID = bigdata1ab456422_20230526072729_0c0325ec-d9cf-4da4-ba72-38be24709633
Total jobs = 5
Launching Job 1 out of 5
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>
23/05/26 07:27:29 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 07:27:29 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2057, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2057/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2057
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-26 07:27:39,154 Stage-1 map = 0%, reduce = 0%
2023-05-26 07:27:46,411 Stage-1 map = 100%, reduce = 0%
2023-05-26 07:27:56,659 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.07 sec
MapReduce Total cumulative CPU time: 7 seconds 70 msec
Ended Job = job_1684866872278_2057
Launching Job 2 out of 5
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>
23/05/26 07:27:57 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 07:27:57 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2058, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2058/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2058
```

```

at java.lang.reflect.Method.invoke(Method.java:750)
at org.apache.hadoop.util.RunJar.run(RunJar.java:313)
at org.apache.hadoop.util.RunJar.main(RunJar.java:227)

Execution completed successfully
MapredLocal task succeeded
Launching Job 5 out of 5
Number of reduce tasks is set to 0 since there's no reduce operator
23/05/26 07:26:16 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 07:26:16 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2056, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2056/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2056
Hadoop job information for Stage-5: number of mappers: 1; number of reducers: 0
2023-05-26 07:26:24,690 Stage-5 map = 0%, reduce = 0%
2023-05-26 07:26:31,869 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 2.79 sec
MapReduce Total cumulative CPU time: 2 seconds 790 msec
Ended Job = job_1684866872278_2056
Moving data to directory hive/airport_project/result47
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.29 sec HDFS Read: 748901 HDFS Write: 7032 HDFS EC Read: 0 SUCCESS
Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 5.06 sec HDFS Read: 748987 HDFS Write: 116 HDFS EC Read: 0 SUCCESS
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 6.39 sec HDFS Read: 4921 HDFS Write: 116 HDFS EC Read: 0 SUCCESS
Stage-Stage-5: Map: 1 Cumulative CPU: 2.79 sec HDFS Read: 11166 HDFS Write: 19 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 21 seconds 530 msec
OK
Time taken: 104.087 seconds
hive (surya_training)>

```

[Back](#)
[Home](#)

Page 1 to 1 of 1

[Edit file](#)

/ user / bigdatalab456422 / hive / airport_project / result47 / 000000_0

[Refresh](#)

United States,1697

[View as binary](#)

PrasadAmle's Query

```

hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result48' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' SELECT * FROM (SELECT country,
count(airport_id), DENSE_RANK() OVER (ORDER BY
COUNT(airport_id) DESC) AS rnk FROM airport GROUP BY country)
AS X WHERE X.rnk=1;

```

```

hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result48' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT * FROM (SELECT country, count(airport_id), DENSE_RANK()
>
> OVER (ORDER BY COUNT(airport_id) DESC) AS rnk
>
> FROM airport GROUP BY country) AS X
>
> WHERE X.rnk=1;
Query ID = bigdatalab456422_20230526083942_9a06e5a5-7e8b-46f0-b9bd-a5684142db94
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.reducers=<number>
23/05/26 08:39:42 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 08:39:42 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2074, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2074/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2074
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-26 08:39:51,625 Stage-1 map = 0%, reduce = 0%
2023-05-26 08:39:59,899 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.07 sec
2023-05-26 08:40:07,141 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.71 sec
MapReduce Total cumulative CPU time: 4 seconds 710 msec
Ended Job = job_1684866872278_2074
Launching Job 2 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.reducers=<number>
23/05/26 08:40:08 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

```

```

23/05/26 08:40:08 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 08:40:08 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2075, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2075/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2075
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-05-26 08:40:18,117 Stage-2 map = 0%, reduce = 0%
2023-05-26 08:40:29,596 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.83 sec
2023-05-26 08:40:39,803 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.7 sec
MapReduce Total cumulative CPU time: 6 seconds 700 msec
Ended Job = job_1684866872278_2075
Moving data to directory hive/airport_project/result48
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.71 sec HDFS Read: 748594 HDFS Write: 7032 HDFS EC Read: 0 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.7 sec HDFS Read: 15216 HDFS Write: 21 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 410 msec
OK
Time taken: 58.6 seconds
hive (surya_training)>

```

[Back](#)
[Home](#)

Page 1 to 1 of 1

[Edit file](#)

/ user / bigdatalab456422 / hive / airport_project / result47 / 000000_0

[Refresh](#)

United States,1697

D.1. Which country (or) territory having lowest Airports

sandeep's query

```

hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result47a' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' SELECT country, count(airport_id) AS cnt FROM
airport GROUP BY country HAVING cnt IN (SELECT min(total) FROM
(SELECT count(airport_id) AS total FROM airport GROUP BY
country) AS abc);

```

```

hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result47a' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT country, count(airport_id) AS cnt FROM airport GROUP BY country HAVING cnt IN
>
> (SELECT min(total) FROM
>
> (SELECT count(airport_id) AS total FROM airport GROUP BY country) AS abc);
Query ID = bigdatalab456422_20230526090702_20f66259-e7c6-4500-b0df-df135809a869
Total jobs = 5
Launching Job 1 out of 5
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>
23/05/26 09:07:02 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 09:07:02 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2150, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2150/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2150
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-26 09:07:15,384 Stage-1 map = 0%, reduce = 0%
2023-05-26 09:07:22,735 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2023-05-26 09:07:32,043 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.44 sec
MapReduce Total cumulative CPU time: 6 seconds 440 msec
Ended Job = job_1684866872278_2150
Launching Job 2 out of 5
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>
23/05/26 09:07:33 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 09:07:33 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

```

```

at org.apache.logging.log4j.core.appender.RandomAccessFileAppender.createAppender(RandomAccessFileAppender.java:174)
at org.apache.hadoop.hive.q1.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)
... 35 more

Execution completed successfully
MapredLocal task succeeded
Launching Job 5 out of 5
Number of reduce tasks is set to 0 since there's no reduce operator
23/05/26 09:00:03 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 09:00:03 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2131, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2131/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2131
Hadoop job information for Stage-5: number of mappers: 1; number of reducers: 0
2023-05-26 09:00:26,149 Stage-5 map = 0%, reduce = 0%
2023-05-26 09:00:33,441 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 2.4 sec
MapReduce Total cumulative CPU time: 2 seconds 400 msec
Ended Job = job_1684866872278_2131
Moving data to directory hive/airport_project/result4a7
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.39 sec HDFS Read: 748901 HDFS Write: 7032 HDFS EC Read: 0 SUCCESS
Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 5.66 sec HDFS Read: 749003 HDFS Write: 114 HDFS EC Read: 0 SUCCESS
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 7.17 sec HDFS Read: 4919 HDFS Write: 114 HDFS EC Read: 0 SUCCESS
Stage-Stage-5: Map: 1 Cumulative CPU: 2.4 sec HDFS Read: 11170 HDFS Write: 446 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 620 msec
OK
Time taken: 142.284 seconds
hive (surya_training)>

```

[Back](#)
[Home](#)

Page 1 to 1 of 1

[Edit file](#)

[Refresh](#)

[View as binary](#)

[Download](#)

Last modified
05/26/2023 2:39 PM
+05:30
User
bigdatalab456422

/ user / bigdatalab456422 / hive / airport_project / result47a / 000000_0

Fogo Island,1
Niuafo'ou Island,1
Santiago Island,1
Sheffield,1
Albania,1
Anguilla,1
Aruba,1
Barbados,1
Benin,1
Bhutan,1
British Indian Ocean Territory,1
Burundi 1

#PrasadAmle's query mod

```

hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result48a' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' SELECT * FROM (SELECT country,
count(airport_id), DENSE_RANK() OVER (ORDER BY
COUNT(airport_id) ASC) AS rnk FROM airport GROUP BY country) AS
X WHERE X.rnk=1;

```



```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result48a' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT * FROM (SELECT country, count(airport_id), DENSE_RANK()
>
> OVER (ORDER BY COUNT(airport_id) ASC) AS rnk
>
> FROM airport GROUP BY country) AS X
>
> WHERE X.rnk=1;
Query ID = bigdatalab456422_20230526092915_7584b7cc-f841-4b28-8824-de57078c17ea
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>
23/05/26 09:29:16 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 09:29:16 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2208, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2208/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2208
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-26 09:29:32,795 Stage-1 map = 0%, reduce = 0%
2023-05-26 09:29:43,240 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.15 sec
2023-05-26 09:29:50,477 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.82 sec
MapReduce Total cumulative CPU time: 4 seconds 820 msec
Ended Job = job_1684866872278_2208
Launching Job 2 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>
23/05/26 09:29:51 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 09:29:51 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

Starting Job = job_1684866872278_2210, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2210/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2210
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-05-26 09:30:05,288 Stage-2 map = 0%, reduce = 0%
2023-05-26 09:30:13,626 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.26 sec
2023-05-26 09:30:24,883 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 7.5 sec
MapReduce Total cumulative CPU time: 7 seconds 500 msec
Ended Job = job_1684866872278_2210
Moving data to directory hive/airport_project/result48a
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.82 sec HDFS Read: 748585 HDFS Write: 7032 HDFS EC Read: 0 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 7.5 sec HDFS Read: 15220 HDFS Write: 514 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 320 msec
OK
Time taken: 71.11 seconds
hive (surya_training)>
```

[Back](#)
[Home](#)
Page to of 1

[Edit file](#)
[Refresh](#)
[View as binary](#)
[Download](#)

Last modified
 05/26/2023 3:00 PM
 +05:30
 User
 bigdatalab456422
 Group
 bigdatalab456422

Santiago Island,1,1
 Niuafo'ou Island,1,1
 Saint Helena,1,1
 West Bank,1,1
 Palestine,1,1
 Wake Island,1,1
 Sheffield,1,1
 Albania,1,1
 Anguilla,1,1
 Aruba,1,1
 Barbados,1,1
 Benin,1,1
 Bhutan,1,1
 British Indian Ocean Territory,1,1

FROM airport GROUP BY country
) AS X WHERE X.rnk = 1;

e. Find the list of Active Airlines in United States

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result51' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ','
```



```

SELECT distinct(al.name)
FROM airlines al
JOIN routes rt ON al.airline_id = rt.airline_id
JOIN airport ap ON rt.src_airport_iata=ap.iata
JOIN airport ap2 ON rt.dest_airport_iata = ap2.iata
WHERE TRIM(UPPER(ap.country))='UNITED STATES'
OR TRIM(UPPER(ap2.country))='UNITED STATES';

```

```

hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result51' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT distinct(al.name)
>
> FROM airlines al
>
> JOIN routes rt ON al.airline_id = rt.airline_id
>
> JOIN airport ap ON rt.src_airport_iata=ap.iata
>
> JOIN airport ap2 ON rt.dest_airport_iata = ap2.iata
>
> WHERE TRIM(UPPER(ap.country))='UNITED STATES'
>
> OR TRIM(UPPER(ap2.country))='UNITED STATES';
Query ID = bigdatalab456422_20230526065837_60076a86-824a-4b94-bced-7fa980b8cf3d
Total jobs = 1
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2023-05-26 06:58:44,268 main ERROR RandomAccessFileManager (/var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425/bigdatalab456422_20230526065837_60076a86-824a-4b94-bced-7fa980b8cf3d) java.io.IOException: Could not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425 java.io.IOException: Could not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425
    at org.apache.logging.log4j.core.util.FileUtils.mkdir(FileUtils.java:119)
    at org.apache.logging.log4j.core.util.FileUtils.makeParentDirs(FileUtils.java:136)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager.createManager(RandomAccessFileManager.java:108)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager.getRandomAccessFileManager(RandomAccessFileManager.java:74)
    at org.apache.hadoop.hive.q1.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)
    ... 34 more

MapredLocal task succeeded
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>
23/05/26 06:58:47 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 06:58:47 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2029, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2029/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2029
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2023-05-26 06:58:56,071 Stage-4 map = 0%, reduce = 0%
2023-05-26 06:59:06,379 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 6.23 sec
2023-05-26 06:59:14,540 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 9.84 sec
MapReduce Total cumulative CPU time: 9 seconds 840 msec
Ended Job = job_1684866872278_2029
Moving data to directory hive/airport_project/result51
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 9.84 sec HDFS Read: 2395105 HDFS Write: 2252 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 840 msec
OK
Time taken: 39.426 seconds
hive (surya_training)>

```

[Back](#)
[Home](#)

Page 1 to 1 of 1

[Edit file](#)
[Refresh](#)
[View as binary](#)
[Download](#)

Last modified

05/26/2023 12:29 PM
+05:30

User

bigdatalab456422

Group

bigdatalab456422

Size

2.2 KB

Mode

100755

/ user / bigdatalab456422 / hive / airport_project / result51 / 000000_0

```

40-Mile Air
Aegean Airlines
Aer Lingus
Aero Flight
AeroMexico
Aeroflot Russian Airlines
Aerolane
Aerolineas Argentinas
Aeromar
Air Berlin
Air Canada
Air China
Air Choice One
Air Europa
Air France
Air India Limited
Air New Zealand
Air Pacific
Air Salento

```

sandeep's query

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result57' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ','
SELECT distinct(a.name)
FROM airlines a
JOIN routes r ON a.airline_id = r.airline_id
JOIN airport ar1 ON ar1.airport_id = src_airport_id
JOIN airport ar2 ON ar2.airport_id = dest_airport_id
WHERE TRIM(UPPER(ar1.country))='UNITED STATES'
OR TRIM(UPPER(ar2.country))='UNITED STATES' ;
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result57' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> SELECT distinct(a.name)
>
> FROM airlines a
>
> JOIN routes r ON a.airline_id = r.airline_id
>
> JOIN airport ar1 ON ar1.airport_id = src_airport_id
>
> JOIN airport ar2 ON ar2.airport_id = dest_airport_id
>
> WHERE TRIM(UPPER(ar1.country))='UNITED STATES'
>
> OR TRIM(UPPER(ar2.country))='UNITED STATES' ;
Query ID = bigdatalab456422_20230526071231_6ba6558f-6d41-42e1-aae8-667f3dfef0c1
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2023-05-26 07:12:38,184 main ERROR RandomAccessFileManager (/var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425/bigdatalab456422_20230526071231_6ba6558f-6d41-42e1-aae8-667f3dfef0c1) java.io.IOException: Could not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425 java.io.IOException: Could not create directory /var/log/hive/operation_logs/68d6c990-95d7-4736-9243-3156fcb63425
    at org.apache.logging.log4j.core.util.FileUtils.mkdir(FileUtils.java:119)
    at org.apache.logging.log4j.core.util.FileUtils.makeParentDirs(FileUtils.java:136)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:108)
    at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:108)
    at org.apache.hadoop.hive ql.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)
    ... 34 more
```

```
Execution completed successfully
MapredLocal task succeeded
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>
23/05/26 07:08:29 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/26 07:08:29 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_2036, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1684866872278_2036/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1684866872278_2036
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2023-05-26 07:08:41,465 Stage-4 map = 0%, reduce = 0%
2023-05-26 07:08:54,108 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 6.08 sec
2023-05-26 07:09:01,375 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 9.32 sec
MapReduce Total cumulative CPU time: 9 seconds 320 msec
Ended Job = job_1684866872278_2036
Moving data to Directory hive/airport_project/result57
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 9.32 sec HDFS Read: 2395172 HDFS Write: 2257 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 320 msec
OK
Time taken: 42.561 seconds
hive (surya_training)>
```

Back

Home

Page 1 to 1 of 1

Edit file

Refresh

View as binary

Download

Last modified

05/26/2023 12:43 PM

+05:30

User

bigdatalab456422

Group

bigdatalab456422

Size

2.2 KB

Mode

100755

40-Mile Air

Aegean Airlines

Aer Lingus

Aero Flight

AeroMexico

Aeroflot Russian Airlines

Aerolane

Aerolineas Argentinas

Aeromar

Air Berlin

Air Canada

Air China

Air Choice One

Air Europa

Air France

Air India Limited

Air New Zealand

Air Pacific

Air Salone

40-Mile Air

Aegean Airlines

Aer Lingus

Aero Flight

AeroMexico

Aeroflot Russian Airlines

Aerolane

Aerolineas Argentinas

Aeromar

Air Berlin

Air Canada

Air China

Air Choice One

Air Europa

Air France

Air India Limited

Air New Zealand

Air Pacific

Air Salone

QnA

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY
'hive/airport_project/result2' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' SELECT al.name, rt.src_airport_iata,
rt.dest_airport_iata, rt.stop FROM airlines al JOIN routes rt
ON al.airline_id=rt.airline_id WHERE rt.stop=0
;
```

```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/airport_project/result2' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
> SELECT al.name, rt.src_airport_iata, rt.dest_airport_iata, rt.stop FROM airlines al JOIN routes rt ON al.airline_id=rt.airline_id WHERE rt.stop=0
;
Query ID = bigdatalab456422_20230525091341_e18e1983-d933-4fe3-bd6b-3f698b89f552
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2023-05-25 09:13:48,146 main ERROR RandomAccessFileManager (/var/log/hive/operation_logs/2e54743d-bfea-48cb-ba64-8e43db765821/bigdatalab456422_20230525091341_e18e1983-d
933-4fe3-bd6b-3f698b89f552) java.io.IOException: Could not create directory /var/log/hive/operation_logs/2e54743d-bfea-48cb-ba64-8e43db765821 java.io.IOException: Could
not create directory /var/log/hive/operation_logs/2e54743d-bfea-48cb-ba64-8e43db765821
at org.apache.logging.log4j.core.util.FileUtils.mkdir(FileUtils.java:119)
at org.apache.logging.log4j.core.util.FileUtils.makeParentDirs(FileUtils.java:136)
at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:198)
at org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory.createManager(RandomAccessFileManager.java:177)
at org.apache.logging.log4j.core.appender.AbstractManager.getManager(AbstractManager.java:112)
at org.apache.logging.log4j.core.appender.OutputStreamManager.getManager(OutputStreamManager.java:114)
at org.apache.logging.log4j.core.appender.RandomAccessFileManager.getFileManager(RandomAccessFileManager.java:74)
at org.apache.hadoop.hive.q1.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.logging.log4j.core.config.plugins.util.PluginBuilder.build(PluginBuilder.java:132)
at org.apache.logging.log4j.core.config.AbstractConfiguration.createPluginObject(AbstractConfiguration.java:952)
at org.apache.logging.log4j.core.config.AbstractConfiguration.createConfiguration(AbstractConfiguration.java:892)
at org.apache.logging.log4j.core.appender.routing.RoutingAppender.createAppender(RoutingAppender.java:271)
at org.apache.logging.log4j.core.appender.routing.RoutingAppender.getControl(RoutingAppender.java:255)
at org.apache.logging.log4j.core.appender.routing.RoutingAppender.append(RoutingAppender.java:225)
at org.apache.logging.log4j.core.config.AppenderControl.tryCallAppender(AppenderControl.java:156)
at org.apache.logging.log4j.core.config.AppenderControl.callAppender0(AppenderControl.java:129)
at org.apache.logging.log4j.core.config.AppenderControl.callAppenderReentrantRecursive(AppenderControl.java:138)
```