\*\*\*\*\* OUTPUT HIVE \*\*\*\*\*

[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

pvg

Time taken: 0.589 seconds, Fetched: 2 row(s)

hive> create database cars;

OK

Time taken: 1.225 seconds

hive> show databases;

OK

cars

default

pvg

Time taken: 0.021 seconds, Fetched: 3 row(s)

hive> create table suv(

> id int,

> name string);

OK

Time taken: 0.367 seconds

hive> desc suv;

OK

id int

name string

Time taken: 0.109 seconds, Fetched: 2 row(s)

hive> INSERT INTO TABLE suv

> VALUES (1, 'creata');

Query ID = cloudera\_20240422214949\_80095f7f-4cc4-404f-8fd9-bc597030688f

Total jobs = 3

Launching Job 1 out of 3

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

Starting Job = job\_1713847438273\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1713847438273\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1713847438273\_0001

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

2024-04-22 21:50:10,262 Stage-1 map = 0%, reduce = 0%

2024-04-22 21:50:17,991 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.28 sec

MapReduce Total cumulative CPU time: 1 seconds 280 msec

Ended Job = job\_1713847438273\_0001

Stage-4 is selected by condition resolver.

Stage-3 is filtered out by condition resolver.

Stage-5 is filtered out by condition resolver.

Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/suv/.hive-staging\_hive\_2024-04-22\_21-49-56\_911\_5933347964744007760-1/-ext-10000

Loading data to table default.suv

Table default.suv stats: [numFiles=1, numRows=1, totalSize=9, rawDataSize=8]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 1.28 sec HDFS Read: 3436 HDFS Write: 76 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 280 msec

OK

Time taken: 22.523 seconds

hive> select \* from suv;

OK

1 creata

Time taken: 0.083 seconds, Fetched: 1 row(s)

hive> ALTER TABLE suv RENAME to suv\_info;

OK

Time taken: 0.117 seconds

hive> show tables;

OK

suv\_info

values\_\_tmp\_\_table\_\_1

Time taken: 0.038 seconds, Fetched: 2 row(s)

hive> DROP TABLE suv\_info;

OK

Time taken: 0.345 seconds

hive> create table cedan(

> id int,

> name string,

> fuel string);

OK

Time taken: 0.052 seconds

hive> drop table cedan;

OK

Time taken: 0.101 seconds

hive> create table cedan(

> id int,

> name string,

> fuel string)

> row format delimited

> fields terminated by ',';

OK

Time taken: 0.073 seconds

hive> desc cedan;

OK

id int

name string

fuel string

Time taken: 0.065 seconds, Fetched: 3 row(s)

hive> load data local inpath '/home/cloudera/sample.txt' overwrite into table cedan;

Loading data to table default.cedan

Table default.cedan stats: [numFiles=1, numRows=0, totalSize=40, rawDataSize=0]

OK

Time taken: 0.363 seconds

hive> select \* from cedan;

OK

1 honda\_civic petrol

2 verna petrol

Time taken: 0.049 seconds, Fetched: 2 row(s)

hive> create external table suv(

> id int,

> name string)

> row format delimited

> fields terminated by ','

> location '/mydb/';

OK

Time taken: 0.063 seconds

hive> select \*from suv;

OK

1 creata

2 venue

Time taken: 0.048 seconds, Fetched: 3 row(s)

hive> select c.id, c.fuel, s.name

> from cedan c

> left outer join suv s

> on (c.id = s.id);

Query ID = cloudera\_20240423005050\_9d115b37-1b71-45b9-88ff-7ea6d1e3c108

Total jobs = 1

Execution log at: /tmp/cloudera/cloudera\_20240423005050\_9d115b37-1b71-45b9-88ff-7ea6d1e3c108.log

2024-04-23 12:50:13 Starting to launch local task to process map join; maximum memory = 1013645312

2024-04-23 12:50:14 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/cloudera/3e417ea8-59a4-4e38-85bc-acf098b5f930/hive\_2024-04-23\_00-50-08\_440\_8118437010799648702-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable

2024-04-23 12:50:14 Uploaded 1 File to: file:/tmp/cloudera/3e417ea8-59a4-4e38-85bc-acf098b5f930/hive\_2024-04-23\_00-50-08\_440\_8118437010799648702-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (331 bytes)

2024-04-23 12:50:14 End of local task; Time Taken: 1.047 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\_1713857397049\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1713857397049\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1713857397049\_0001

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0

2024-04-23 00:50:27,774 Stage-3 map = 0%, reduce = 0%

2024-04-23 00:50:34,495 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 0.95 sec

MapReduce Total cumulative CPU time: 950 msec

Ended Job = job\_1713857397049\_0001

MapReduce Jobs Launched:

Stage-Stage-3: Map: 1 Cumulative CPU: 0.95 sec HDFS Read: 5720 HDFS Write: 35 SUCCESS

Total MapReduce CPU Time Spent: 950 msec

OK

1 petrol creata

2 petrol venue

Time taken: 27.

hive> create external table flights(

> 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 ','

> location '/mydb/';

OK

Time taken: 0.104 seconds

hive> desc flights;

OK

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

Time taken: 0.343 seconds, Fetched: 29 row(s)

hive> select \* from flights LIMIT 15;

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

2008 1 3 4 706 700 916 915 WN 100 N690SW 130 135 106 1 6 IND MCO 828 5 19 0 0 NA NA NA NA NA

2008 1 3 4 1644 1510 1845 1725 WN 1333 N334SW 121 135 107 80 94 IND MCO 828 6 8 0 0 8 0 0 0 72

2008 1 3 4 1426 1430 1426 1425 WN 829 N476WN 60 55 39 1 -4 IND MDW 162 9 12 0 0 NA NA NA NA NA

2008 1 3 4 715 715 720 710 WN 1016 N765SW 65 55 37 10 0 IND MDW 162 7 21 0 0 NA NA NA NA NA

2008 1 3 4 1702 1700 1651 1655 WN 1827 N420WN 49 55 35 -4 2 IND MDW 162 4 10 0 0 NA NA NA NA NA

Time taken: 0.513 seconds, Fetched: 15 row(s)

hive> select day, avg(dep\_delay) from flights group by day;

Query ID = cloudera\_20240423020707\_02290e72-c47d-4c96-9587-5acf618a50d7

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1713861720599\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1713861720599\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1713861720599\_0001

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

2024-04-23 02:07:20,868 Stage-1 map = 0%, reduce = 0%

2024-04-23 02:07:30,675 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 0.95 sec

2024-04-23 02:07:42,385 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 1.64 sec

MapReduce Total cumulative CPU time: 1 seconds 640 msec

Ended Job = job\_1713861720599\_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 1.64 sec HDFS Read: 25124733 HDFS Write: 664 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 640 msec

OK

1 17.57681842916742

2 23.900056359195943

3 19.370313695485844

4 18.612678509230232

5 25.976967114898148

6 22.146653781106547

7 14.395251396648044

8 12.124760306807287

9 5.839149336153214

10 9.223829201101928

11 9.410679275746743

12 1.6842865395725015

13 6.079343193782903

14 4.633204633204633

15 5.640961857379768

16 1.9354166666666666

17 18.21534910559723

18 12.01187917185202

19 7.5900463308922435

20 6.213233458177278

21 25.198426472289714

22 17.538498383427136

23 11.585463541053128

24 9.975531671621313

25 14.944508404328804

26 4.631294964028777

27 25.05219499744768

28 14.486067019400354

29 9.989655592065231

30 6.108780661215784

31 27.131638620360423

Time taken: 38.79 seconds, Fetched: 32 row(s)

hive> SELECT day,AVG(dep\_delay) FROM flights WHERE day IS NOT NULL AND dep\_delay IS NOT NULL GROUP BY day;

Query ID = cloudera\_20240423020707\_eaae41c0-3a40-44fa-94ea-e4bfba5e23a2

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1713861720599\_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1713861720599\_0002/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1713861720599\_0002

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

2024-04-23 02:08:05,990 Stage-1 map = 0%, reduce = 0%

2024-04-23 02:08:16,860 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.13 sec

2024-04-23 02:08:26,419 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 1.81 sec

MapReduce Total cumulative CPU time: 1 seconds 810 msec

Ended Job = job\_1713861720599\_0002

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 1.81 sec HDFS Read: 25124397 HDFS Write: 658 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 810 msec

OK

1 17.57681842916742

2 23.900056359195943

3 19.370313695485844

4 18.612678509230232

5 25.976967114898148

6 22.146653781106547

7 14.395251396648044

8 12.124760306807287

9 5.839149336153214

10 9.223829201101928

11 9.410679275746743

12 1.6842865395725015

13 6.079343193782903

14 4.633204633204633

15 5.640961857379768

16 1.9354166666666666

17 18.21534910559723

18 12.01187917185202

19 7.5900463308922435

20 6.213233458177278

21 25.198426472289714

22 17.538498383427136

23 11.585463541053128

24 9.975531671621313

25 14.944508404328804

26 4.631294964028777

27 25.05219499744768

28 14.486067019400354

29 9.989655592065231

30 6.108780661215784

31 27.131638620360423

Time taken: 32.99 seconds, Fetched: 31 row(s)

hive> CREATE INDEX myindex ON TABLE flights(day) AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler' WITH DEFERRED REBUILD;

OK

Time taken: 0.485 seconds

hive> SELECT origin,dest FROM flights WHERE day<3 LIMIT 10;

OK

MCI IAH

CLE SDF

TUL EWR

IAH BNA

EWR MYR

AUS ONT

AUS ONT

ONT MCI

ONT MCI

FAT ONT

Time taken: 0.181 seconds, Fetched: 10 row(s)

hive>

hive> create table flights1(

> 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 ','

> location '/mydb/';

OK

Time taken: 0.082 seconds

hive> desc flights1;

OK

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

Time taken: 0.149 seconds, Fetched: 29 row(s)

hive> create external table flights2(

> 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: 0.076 seconds

hive> insert into flights2 select \* from flights1;

Query ID = cloudera\_20240423022424\_e38dc5af-5ec4-47b5-97b4-efca0aae5001

Total jobs = 3

Launching Job 1 out of 3

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

Starting Job = job\_1713861720599\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1713861720599\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1713861720599\_0003

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

2024-04-23 02:24:40,431 Stage-1 map = 0%, reduce = 0%

2024-04-23 02:24:52,110 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.78 sec

MapReduce Total cumulative CPU time: 1 seconds 780 msec

Ended Job = job\_1713861720599\_0003

Stage-4 is selected by condition resolver.

Stage-3 is filtered out by condition resolver.

Stage-5 is filtered out by condition resolver.

Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/flights2/.hive-staging\_hive\_2024-04-23\_02-24-30\_149\_3594729791709269238-1/-ext-10000

Loading data to table default.flights2

Table default.flights2 stats: [numFiles=1, numRows=257253, totalSize=25114245, rawDataSize=24856992]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 1.78 sec HDFS Read: 25121779 HDFS Write: 25114327 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 780 msec

OK

Time taken: 24.775 seconds

hive> select day,avg(dep\_delay) from flights2 where origin = 'IAD' group by day;

Query ID = cloudera\_20240423022727\_0b732c31-1208-43aa-a8d4-92453728d9b7

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1713861720599\_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1713861720599\_0004/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1713861720599\_0004

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

2024-04-23 02:27:17,106 Stage-1 map = 0%, reduce = 0%

2024-04-23 02:27:27,687 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.12 sec

2024-04-23 02:27:39,289 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 1.82 sec

MapReduce Total cumulative CPU time: 1 seconds 820 msec

Ended Job = job\_1713861720599\_0004

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 1.82 sec HDFS Read: 25125871 HDFS Write: 636 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 820 msec

OK

1 18.795454545454547

2 38.779310344827586

3 13.441558441558442

4 19.03726708074534

5 24.95945945945946

6 27.12883435582822

7 14.962732919254659

8 12.06875

9 7.1835443037974684

10 13.962732919254659

11 15.936708860759493

12 5.190476190476191

13 11.866666666666667

14 6.144654088050315

15 4.1

16 2.5063291139240507

17 47.471014492753625

18 25.15432098765432

19 24.835616438356166

20 24.99337748344371

21 13.082278481012658

22 20.602564102564102

23 21.54248366013072

24 14.095541401273886

25 21.146496815286625

26 8.4822695035461

27 16.337748344370862

28 11.058064516129033

29 7.949685534591195

30 10.529032258064516

31 14.314465408805031

Time taken: 33.565 seconds, Fetched: 31 row(s)

hive>