

***** 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 sedan;
```

OK

Time taken: 0.101 seconds

```
hive> create table sedan(
```

```
> id int,
```

```
> name string,
```

```
> fuel string)
```

```
> row format delimited
```

```
> fields terminated by ',';
```

OK

Time taken: 0.073 seconds

```
hive> desc sedan;
```

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 sedan;
```

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 sedan;
```

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 sedan 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	1 116 NA	3 -14 NA	4 8 NA	2003 IAD NA	1955 TPA	2211 810	2225 4	WN 8	335 0	N712SW	128 0	150 NA
2008	1 113 NA	3 2 NA	4 19 NA	754 IAD NA	735 TPA	1002 810	1000 5	WN 10	3231 0	N772SW	128 0	145 NA
2008	1 76 NA	3 14 NA	4 8 NA	628 IND NA	620 BWI	804 515	750 3	WN 17	448 0	N428WN	96 0	90 NA
2008	1 78 NA	3 -6 NA	4 -4 NA	926 IND NA	930 BWI	1054 515	1100 3	WN 7	1746 0	N612SW	88 0	90 NA
2008	1 77 0	3 34 0	4 34 0	1829 IND 32	1755 BWI	1959 515	1925 3	WN 10	3920 0	N464WN	90 0	90 2
2008	1 87 NA	3 11 NA	4 25 NA	1940 IND NA	1915 JAX	2121 688	2110 4	WN 10	378 0	N726SW	101 0	115 NA
2008	1 230 0	3 57 0	4 67 0	1937 IND 47	1830 LAS	2037 1591	1940 3	WN 7	509 0	N763SW	240 0	250 10
2008	1 219 NA	3 -18 NA	4 -1 NA	1039 IND NA	1040 LAS	1132 1591	1150 7	WN 7	535 0	N428WN	233 0	250 NA
2008	1 70 NA	3 2 NA	4 2 NA	617 IND NA	615 MCI	652 451	650 6	WN 19	11 0	N689SW	95 0	95 NA
2008	1 70 NA	3 -16 NA	4 0 NA	1620 IND NA	1620 MCI	1639 451	1655 3	WN 6	810 0	N648SW	79 0	95 NA
2008	1 106 NA	3 1 NA	4 6 NA	706 IND NA	700 MCO	916 828	915 5	WN 19	100 0	N690SW	130 0	135 NA
2008	1 107 0	3 80 0	4 94 0	1644 IND 72	1510 MCO	1845 828	1725 6	WN 8	1333 0	N334SW	121 0	135 8
2008	1 39 NA	3 1 NA	4 -4 NA	1426 IND NA	1430 MDW	1426 162	1425 9	WN 12	829 0	N476WN	60 0	55 NA
2008	1 37 NA	3 10 NA	4 0 NA	715 IND NA	715 MDW	720 162	710 7	WN 21	1016 0	N765SW	65 0	55 NA
2008	1 35 NA	3 -4 NA	4 2 NA	1702 IND NA	1700 MDW	1651 162	1655 4	WN 10	1827 0	N420WN	49 0	55 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>