**DSBDA LAB**

**GROUP A , ASS 4**

**Hive Analytics**

### **Hive Commands and Results**

#### **1. Create Simple table , insert data**

hadoop@pvg:~/Desktop/DSBDA/hive\_try$ hive

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.3-bin/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder…

Hive Session ID = 51f94733-b995-4e80-b83c-7b8361d4187d

hive> use test\_db;

OK

Time taken: 0.384 seconds

hive> select \* from persons;

OK

Time taken: 1.481 seconds

hive> INSERT INTO TABLE persons VALUES

> (1, 'Shubham', 'Pune'),

> (2, 'Rahul', 'Mumbai');

Query ID = hadoop\_20250428070419\_691d035d-e023-47fa-b879-dc9dd1845d87

Total jobs = 3

Launching Job 1 out of 3

….

Stage-Stage-1: HDFS Read: 0 HDFS Write: 202 SUCCESS

Total MapReduce CPU Time Spent: 0 msec

OK

Time taken: 4.304 seconds

hive> select \* from person;

OK

1 Shubham Pune

2 Rahul Mumbai

Time taken: 0.124 seconds, Fetched: 2 row(s)

#### **2. Create Table for flight**

hive> create table flight (

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

);

OK

Time taken: 0.642 seconds

#### **3. Describe Table Structure**

hive> describe flight;

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: 0.276 seconds, Fetched: 29 row(s)

#### **4. Drop Table flight**

hive> drop table flight;

OK

Time taken: 0.549 seconds

#### **5. Create flight Table with Row Format Delimited**

hive> create table flight (

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: 0.097 seconds

#### **6. Load Data into flight Table**

hive> load data local INPATH '/home/cloudera/flight\_info.csv' overwrite into table flight;

Loading data to table default.flight

Table default.flight stats: [numFiles=1, numRows=0, totalSize=25113900, rawDataSize=0]

OK

Time taken: 1.055 seconds

#### **6. Query Data from flight Table**

hive> select \* from flight limit 10;

OK

2008 1 3 4 2003 1955 2211 2225 WN 335 N712SW 128 150 116 -14 8 IAD TPA 810 4 8 00 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 00 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 00 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 00 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 00 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 00 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 00 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 00 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 00 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 00 NA NA NA NA NA

Time taken: 0.581 seconds, Fetched: 10 row(s)

#### **7. Create Index on flight\_table\_2**

hive> create index flight\_index on table flight\_table\_2(unique\_carrier) as 'compact' with deferred rebuild ;

OK

Time taken: 1.117 seconds

hive> show indexes on flight\_table\_2;

| INDEX | TABLE | TYPE | DEFERRED | CREATED | DATABASE | COMPACTION | IDX\_NAME | COMMENT |

| flight\_index | flight\_table\_2 | compact | Yes | 2025-04-24 | default | true | flight\_index | Index for unique\_carrier |

#### **8. Query for Average Departure Delay**

hive> select avg(dep\_delay) as avg\_departure\_delay from flight\_table\_2 ;

Query ID = cloudera\_20250424222121\_d7f9dcdb-80eb-4080-b17f-d47e71fe2d83

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

…

OK

NULL NULL

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: 0.400 seconds, Fetched: 1 row(s)