**ASSIGNMENT 3**

**Name:** Isha Bhelsekar

**Roll no:** 3006

**Title:** Write an application using HiveQL for flight information system which will include-

a. Creating, Dropping, and altering Database tables.

b. Creating an external Hive table.

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

with Hive.

d. Create index on Flight InformationTable.

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

[cloudera@quickstart ~]$ hostname

quickstart.cloudera

[cloudera@quickstart ~]$ hdfs dfs -ls /

Found 6 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - hbase supergroup 0 2024-03-06 21:47 /hbase

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2024-03-05 19:52 /tmp

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

[cloudera@quickstart ~]$ 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.

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

**Creating Database**

hive> create database mydb12;

OK

Time taken: 0.294 seconds

**Creating Table**

hive> create table flight(fno int,year int, dest varchar(10), delay float);

OK

Time taken: 0.37 seconds

**Alter Table**

hive> alter table flight rename to air\_flight;

OK

Time taken: 0.178 seconds

hive> alter table air\_flight add columns(source varchar(10));

OK

Time taken: 0.112 seconds

**Drop Table**

hive> drop table flight;

OK

Time taken: 0.08 seconds

hive> desc air\_flight;

OK

fno int

year int

dest varchar(10)

delay float

source varchar(10)

Time taken: 0.172 seconds, Fetched: 5 row(s)

**Creating an external Hive table**

hive> create table flight(fno int, year int, dest varchar(10), delay float)

> row format delimited

> fields terminated by ','

> lines terminated by '\n'

> stored as textfile;

OK

Time taken: 0.049 seconds

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

hive> insert into flight values(123,2009,"mumbai",30.6);

Query ID = cloudera\_20240306222323\_501ad1e1-b37f-4c4f-92cb-129586db6546

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\_1709789993784\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1709789993784\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1709789993784\_0001

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

2024-03-06 22:23:19,483 Stage-1 map = 0%, reduce = 0%

2024-03-06 22:23:25,989 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.09 sec

MapReduce Total cumulative CPU time: 1 seconds 90 msec

Ended Job = job\_1709789993784\_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/flight/.hive-staging\_hive\_2024-03-06\_22-23-09\_586\_4598981439608380086-1/-ext-10000

Loading data to table default.flight

Table default.flight stats: [numFiles=1, numRows=1, totalSize=21, rawDataSize=20]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 1.09 sec HDFS Read: 4630 HDFS Write: 91 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 90 msec

OK

Time taken: 17.723 seconds

hive> insert into flight values(124,2008,"pune",50.6);

Query ID = cloudera\_20240306222424\_fb18dfdc-d4ae-4a91-858c-a63531937da2

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\_1709789993784\_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1709789993784\_0002/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1709789993784\_0002

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

2024-03-06 22:24:29,370 Stage-1 map = 0%, reduce = 0%

2024-03-06 22:24:34,822 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.05 sec

MapReduce Total cumulative CPU time: 1 seconds 50 msec

Ended Job = job\_1709789993784\_0002

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/flight/.hive-staging\_hive\_2024-03-06\_22-24-22\_405\_8846295283953152786-1/-ext-10000

Loading data to table default.flight

Table default.flight stats: [numFiles=2, numRows=2, totalSize=40, rawDataSize=38]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 1.05 sec HDFS Read: 4734 HDFS Write: 89 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 50 msec

OK

Time taken: 13.628 seconds

hive> select \* from flight;

OK

123 2009 mumbai 30.6

124 2008 pune 50.6

Time taken: 0.05 seconds, Fetched: 2 row(s)

hive> load data local inpath "f.txt"

> overwrite into table flight;

Loading data to table default.flight

Table default.flight stats: [numFiles=1, numRows=0, totalSize=106, rawDataSize=0]

OK

Time taken: 0.224 seconds

hive> select \* from flight;

OK

999 2800 navi mumba 60.0

925 2700 Nashik 50.0

975 2860 Pune 65.0

955 1800 Delhi 80.0

924 2809 Dubai 70.0

Time taken: 0.048 seconds, Fetched: 5 row(s)

hive> create table nflight(fno int, year int, source varchar(10))

> row format delimited

> fields terminated by ','

> lines terminated by '\n'

> stored as textfile;

OK

Time taken: 0.063 seconds

hive> select \* from flight;

OK

999 2800 navi mumba 60.0

925 2700 Nashik 50.0

975 2860 Pune 65.0

955 1800 Delhi 80.0

924 2809 Dubai 70.0

Time taken: 0.038 seconds, Fetched: 5 row(s)

hive> insert into nflight values(354,2558,odisa);

FAILED: SemanticException [Error 10293]: Unable to create temp file for insert values Expression of type TOK\_TABLE\_OR\_COL not supported in insert/values

hive> insert into nflight values(354,2558,"odisa");

Query ID = cloudera\_20240306223737\_f4789171-540a-4749-b4ee-73a16e98e473

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\_1709789993784\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1709789993784\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1709789993784\_0003

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

2024-03-06 22:37:25,330 Stage-1 map = 0%, reduce = 0%

2024-03-06 22:37:31,666 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.07 sec

MapReduce Total cumulative CPU time: 1 seconds 70 msec

Ended Job = job\_1709789993784\_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/nflight/.hive-staging\_hive\_2024-03-06\_22-37-19\_436\_7345496083086490119-1/-ext-10000

Loading data to table default.nflight

Table default.nflight stats: [numFiles=1, numRows=1, totalSize=15, rawDataSize=14]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 1.07 sec HDFS Read: 4317 HDFS Write: 86 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 70 msec

OK

Time taken: 13.481 seconds

**hive> select a.fno, a.year, a.dest, a.delay, b.source**

**> from flight a join nflight b**

> on(a.fno=b.fno);

Query ID = cloudera\_20240306223939\_19761995-482a-4bb9-9d99-9df384210aad

Total jobs = 1

Execution log at: /tmp/cloudera/cloudera\_20240306223939\_19761995-482a-4bb9-9d99-9df384210aad.log

2024-03-06 10:40:04 Starting to launch local task to process map join; maximum memory = 1013645312

2024-03-06 10:40:05 Dump the side-table for tag: 1 with group count: 1 into file: file:/tmp/cloudera/2d71e19a-b4b6-4840-90bf-f4fc25e70bdb/hive\_2024-03-06\_22-39-59\_457\_535109216945466836-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable

2024-03-06 10:40:05 Uploaded 1 File to: file:/tmp/cloudera/2d71e19a-b4b6-4840-90bf-f4fc25e70bdb/hive\_2024-03-06\_22-39-59\_457\_535109216945466836-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (287 bytes)

2024-03-06 10:40:05 End of local task; Time Taken: 1.121 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\_1709789993784\_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1709789993784\_0004/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1709789993784\_0004

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

2024-03-06 22:40:12,186 Stage-3 map = 0%, reduce = 0%

2024-03-06 22:40:18,498 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 0.94 sec

MapReduce Total cumulative CPU time: 940 msec

Ended Job = job\_1709789993784\_0004

MapReduce Jobs Launched:

Stage-Stage-3: Map: 1 Cumulative CPU: 0.94 sec HDFS Read: 7210 HDFS Write: 0 SUCCESS

Total MapReduce CPU Time Spent: 940 msec

OK

Time taken: 20.089 seconds

hive> select \* from nflight;

OK

354 2558 odisa

Time taken: 0.035 seconds, Fetched: 1 row(s)

hive> select \* from flight;

OK

999 2800 navi mumba 60.0

925 2700 Nashik 50.0

975 2860 Pune 65.0

955 1800 Delhi 80.0

924 2809 Dubai 70.0

Time taken: 0.032 seconds, Fetched: 5 row(s)

**Create index on Flight InformationTable**

hive> create index flight\_index on table flight(fno)

> as 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'

> WITH DEFERRED REBUILD;

OK

Time taken: 0.324 seconds

hive> show tables;

OK

air\_flight

default\_\_flight\_flight\_index\_\_

flight

nflight

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

values\_\_tmp\_\_table\_\_2

values\_\_tmp\_\_table\_\_4

Time taken: 0.036 seconds, Fetched: 7 row(s)

hive>

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

hive select avg(delay) as AvgDept from flight where year=2008;

Query ID = cloudera\_28248418211818\_c2bf8b69-ca4f-4636-906-637b0e27e566

Total jobs=1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

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

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

In order to limit the maximum number of reducers:

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

In order to set a constant number of reducers:

set mapreduce.job.reduces <number>

Starting Job job 1709705723864 0619, Tracking URL = <http://quickstart.cloudera:8888/proxy/application>\_17957857238649825/

Kill Command=/usr/lib/hadoop/bin/hadoop job kill job 1709765723064 0019

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

2624-04-18 21:18:22,546 5tage-1 mape, reduce 8% 2624-04-18 21:10:27,829 Stage-1 map 100%, reduce, Cumulative CPU 8.87 sec

2824-04-18 21:10:34,063 Stage-1 map 100%, reduce 100%, Cumulative CPU 1.68 sec

MapReduce Total cumulative CPU time: 1 seconds 680 msec

Ended Job job 1709705723004 0019

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 1.68 sec HDFS Read: 8888 HDFS write: 19 SUCCESS

Total MapReduce CPU Time Spent: 1 seconds 680 msec

OK

28.399999618530273

Time taken: 19.063 seconds, Fetched: 1 row(s)