Output

1. HBase

a. Database creation and data insertion hbase(main):001:0> create 'flight','finfo','fsch' 0 row(s) in 0.8180 seconds

=> Hbase::Table - flight

hbase(main):002:0> list

TABLE

flight

1 row(s) in 0.0160 seconds

=> ["flight"]

hbase(main):003:0> put 'flight',1,'finfo:source','Pune' 0 row(s) in 0.1710 seconds

hbase(main):004:0> put 'flight',1,'finfo:dest','Mumbai' 0 row(s) in 0.0050 seconds

hbase(main):005:0> put 'flight',1,'finfo:year',2009 0 row(s) in 0.0070 seconds

hbase(main):006:0> put 'flight',1,'fsch:at','10:30AM' 0 row(s) in 0.0080 seconds

hbase(main):007:0> put 'flight',1,'fsch:dt','09:30AM' 0 row(s) in 0.0050 seconds

hbase(main):008:0> put 'flight',1,'fsch:delay\_in\_mins','6' 0 row(s) in 0.0050 seconds

hbase(main):009:0> put 'flight',1,'finfo:source','Bangalore' 0 row(s) in 0.0040 seconds

hbase(main):010:0> put 'flight',1,'finfo:dest','Mumbai' 0 row(s) in 0.0070 seconds

hbase(main):011:0> put 'flight',1,'finfo:year',2008 0 row(s) in 0.0070 seconds

hbase(main):012:0> put 'flight',1,'fsch:at','10:30PM' 0 row(s) in 0.0050 seconds

hbase(main):013:0> put 'flight',1,'fsch:dt','06:30PM' 0 row(s) in 0.0050 seconds

hbase(main):014:0> put 'flight',1,'fsch:delay\_in\_mins','60' 0 row(s) in 0.0040 seconds

hbase(main):015:0> put 'flight',1,'finfo:source','Kolkata' 0 row(s) in 0.0060 seconds

hbase(main):016:0> put 'flight',1,'finfo:dest','Pune' 0 row(s) in 0.0050 seconds

hbase(main):017:0> put 'flight',1,'finfo:year',2008 0 row(s) in 0.0080 seconds

hbase(main):018:0> put 'flight',1,'fsch:at','06:00PM' 0 row(s) in 0.0290 seconds

hbase(main):019:0> put 'flight',1,'fsch:dt','07:00PM'

0 row(s) in 0.0110 seconds

hbase(main):020:0> put 'flight',1,'fsch:delay\_in\_mins','18'

0 row(s) in 0.0090 seconds

hbase(main):021:0> put 'flight',1,'finfo:source','Guwahati'

0 row(s) in 0.0050 seconds

hbase(main):022:0> put 'flight',1,'finfo:dest','Delhi'

0 row(s) in 0.0050 seconds

hbase(main):023:0> scan 'flight'

ROW COLUMN+CELL

1 column=finfo:dest, timestamp=1650255863859, value=Delhi

1 column=finfo:source, timestamp=1650255856448, value=Guwahati

1 column=finfo:year, timestamp=1650255801838, value=2008

1 column=fsch:at, timestamp=1650255815579, value=06:00PM

1 column=fsch:delay\_in\_mins, timestamp=1650255838644, value=18

1 column=fsch:dt, timestamp=1650255825967, value=07:00PM

1 row(s) in 0.0340 seconds

b. Alter table (adding revenue column)

hbase(main):057:0> alter 'flight',NAME=>'revenue'

Updating all regions with the new schema...

0/1 regions updated.

1/1 regions updated.

Done.

0 row(s) in 2.3320 seconds

hbase(main):058:0> put 'flight',4,'revenue:rs',45000

0 row(s) in 0.0460 seconds

hbase(main):059:0> put 'flight',3,'revenue:rs',50000

0 row(s) in 0.0790 seconds

hbase(main):060:0> put 'flight',2,'revenue:rs',60000

0 row(s) in 0.0160 seconds

hbase(main):061:0> put 'flight',1,'revenue:rs',70000

0 row(s) in 0.0190 seconds

hbase(main):062:0> scan 'flight'

ROW COLUMN+CELL

1 column=finfo:dest, timestamp=1650255863859, value=Delhi

1 column=finfo:source, timestamp=1650255856448, value=Guwahati

1 column=finfo:year, timestamp=1650255955413, value=2008

1 column=fsch:at, timestamp=1650255965052, value=08:00PM

1 column=fsch:delay\_in\_mins, timestamp=1650255838644, value=18

1 column=fsch:dt, timestamp=1650255982742, value=05:00PM

1 column=revenue:rs, timestamp=1650256741890, value=70000

2 column=finfo:dest, timestamp=1650255947103, value=Delhi

2 column=finfo:source, timestamp=1650255909465, value=Bangalore

2 column=finfo:year, timestamp=1650255998986, value=2008

2 column=fsch:at, timestamp=1650256019810, value=09:00PM

2 column=fsch:delay\_in\_mins, timestamp=1650256053654, value=55

2 column=fsch:dt, timestamp=1650256034311, value=11:30PM

2 column=revenue:rs, timestamp=1650256705403, value=60000

3 column=finfo:dest, timestamp=1650256255914, value=Mumbai

3 column=finfo:source, timestamp=1650256245845, value=Pune

3 column=finfo:year, timestamp=1650256297068, value=2008

3 column=fsch:at, timestamp=1650256305028, value=08:00AM

3 column=fsch:delay\_in\_mins, timestamp=1650256362042, value=53

3 column=fsch:dt, timestamp=1650256317896, value=11:30PM

3 column=revenue:rs, timestamp=1650256698022, value=50000

4 column=finfo:dest, timestamp=1650256476169, value=Chennai

4 column=finfo:source, timestamp=1650256457396, value=Pune

4 column=finfo:year, timestamp=1650256439696, value=2008

4 column=fsch:at, timestamp=1650256417175, value=05:00PM

4 column=fsch:delay\_in\_mins, timestamp=1650256390553, value=20

4 column=fsch:dt, timestamp=1650256424894, value=04:00PM

4 column=revenue:rs, timestamp=1650256689281, value=45000

4 row(s) in 0.1200 seconds

c. Disable and drop table

hbase(main):065:0> create 'temp\_table','column\_family'

0 row(s) in 0.4580 seconds

=> Hbase::Table - temp\_table

hbase(main):066:0> list

TABLE

flight

temp\_table

2 row(s) in 0.0250 seconds

=> ["flight", "temp\_table"]

hbase(main):067:0> disable 'temp\_table'

0 row(s) in 1.2790 seconds

hbase(main):068:0> drop 'temp\_table'

0 row(s) in 0.2300 seconds

hbase(main):069:0> list

TABLE

flight

1 row(s) in 0.0040 seconds

=> ["flight"]

d. Selective query

hbase(main):070:0> get 'flight',1

COLUMN CELL

finfo:dest timestamp=1650255863859, value=Delhi finfo:source timestamp=1650255856448, value=Guwahati finfo:year timestamp=1650255955413, value=2008 fsch:at timestamp=1650255965052, value=08:00PM fsch:delay\_in\_mins timestamp=1650255838644, value=18 fsch:dt timestamp=1650255982742, value=05:00PM 6 row(s) in 0.0540 seconds

hbase(main):072:0> get 'flight','1',COLUMN=>'finfo:source'

COLUMN CELL

finfo:source timestamp=1650255856448, value=Guwahati 1 row(s) in 0.0130 seconds

hbase(main):073:0> get 'flight','1',COLUMN=>['finfo:source','finfo:dest'] COLUMN CELL

finfo:dest timestamp=1650255863859, value=Delhi finfo:source timestamp=1650255856448, value=Guwahati 2 row(s) in 0.0300 seconds

hbase(main):074:0> scan 'flight',COLUMNS=>'finfo:source'

ROW COLUMN+CELL

1 column=finfo:source, timestamp=1650255856448, value=Guwahati

2 column=finfo:source, timestamp=1650255909465, value=Bangalore

3 column=finfo:source, timestamp=1650256245845, value=Pune

4 column=finfo:source, timestamp=1650256457396, value=Pune

4 row(s) in 0.0470 seconds

2. Hive

a. Create external table referring HBase table

hive> CREATE external TABLE hbase\_flight\_new(fno int,fsource string,fdest string,fyear int,fsh\_at string,fsh\_dt string,delay int)

> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'

> WITH

SERDEPROPERTIES("hbase.columns.mapping"=":key,finfo:source,finfo:dest,finfo:year,fsc h:at,fsch:dt,fsch:delay\_in\_mins")

> TBLPROPERTIES("hbase.table.name"="flight");

OK

Time taken: 2.649 seconds

hive> SELECT \* FROM hbase\_flight\_new

> ;

OK

1 Guwahati Delhi 2008 08:00PM 05:00PM 18 2 Bangalore Delhi 2008 09:00PM 11:30PM 55 3 Pune Mumbai 2010 08:00AM 11:30PM 53 4 Pune Chennai 2008 05:00PM 04:00PM 20 Time taken: 1.161 seconds, Fetched: 4 row(s)

b. Finding average delay of flights in the year 2008

hive> select avg(delay) from hbase\_flight\_new where fyear=2008;

Query ID = cloudera\_20220418001212\_67412edf-0a84-458c-b78a-84e4165366d0 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\_1650262690593\_0002, Tracking URL =

http://quickstart.cloudera:8088/proxy/application\_1650262690593\_0002/ Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1650262690593\_0002 Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1 2022-04-18 00:12:23,559 Stage-1 map = 0%, reduce = 0%

2022-04-18 00:12:40,282 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.73 sec 2022-04-18 00:12:53,294 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.2 sec MapReduce Total cumulative CPU time: 4 seconds 200 msec

Ended Job = job\_1650262690593\_0002

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.2 sec HDFS Read: 8919 HDFS Write: 5 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 200 msec

OK

31.0

Time taken: 47.018 seconds, Fetched: 1 row(s)

hive> select \* from hbase\_flight\_new where fyear=2008;

OK

1 Guwahati Delhi 2008 08:00PM 05:00PM 18 2 Bangalore Delhi 2008 09:00PM 11:30PM 55 4 Pune Chennai 2008 05:00PM 04:00PM 20 Time taken: 0.227 seconds, Fetched: 3 row(s)

c. Creating index on table

hive> CREATE INDEX hbase\_flight\_index

> ON TABLE hbase\_flight\_new (delay)

> AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler' > WITH DEFERRED REBUILD;

OK

Time taken: 0.545 seconds

hive> SHOW INDEX ON hbase\_flight\_new;

OK

hbase\_flight\_index hbase\_flight\_new delay

default\_\_hbase\_flight\_new\_hbase\_flight\_index\_\_ compact Time taken: 0.148 seconds, Fetched: 1 row(s)

d. Create table from local file

hive> create table empdb(eno int, ename string, esal int) row format delimited fields terminated by ',' stored as textfile;

OK

Time taken: 0.089 seconds

hive> load data local inpath '/home/cloudera/Desktop/empdb.txt' into table empdb > ;

Loading data to table default.empdb

Table default.empdb stats: [numFiles=1, totalSize=90]

OK

Time taken: 0.519 seconds

hive> select \* from empdb

> ;

OK

1 deepali 120000

2 mahesh 30000

3 mangesh 25000

4 ram 39000

5 brijesh 40000

6 john 300000

Time taken: 0.07 seconds, Fetched: 6 row(s)

e. Join tables

hive> SELECT eno, ename, empno, empgrade FROM empdbnew JOIN empinfo ON eno = empno;

Query ID = cloudera\_20220418003636\_f903716e-c457-415a-9627-1598e53684a4 Total jobs = 1

Execution log at: /tmp/cloudera/cloudera\_20220418003636\_f903716e-c457-415a-9627- 1598e53684a4.log

2022-04-18 12:37:01 Starting to launch local task to process map join; maximum memory = 1013645312

2022-04-18 12:37:03 Dump the side-table for tag: 1 with group count: 6 into file: file:/tmp/cloudera/cb63283e-d448-4c48-9cfe-2e84444932ed/hive\_2022-04-18\_00-36- 52\_339\_3889243584216725453-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01-- .hashtable

2022-04-18 12:37:03 Uploaded 1 File to: file:/tmp/cloudera/cb63283e-d448-4c48-9cfe 2e84444932ed/hive\_2022-04-18\_00-36-52\_339\_3889243584216725453-1/-local 10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (386 bytes)

2022-04-18 12:37:03 End of local task; Time Taken: 2.738 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\_1650262690593\_0003, Tracking URL =

http://quickstart.cloudera:8088/proxy/application\_1650262690593\_0003/ Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1650262690593\_0003 Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0 2022-04-18 00:37:20,670 Stage-3 map = 0%, reduce = 0%

2022-04-18 00:37:35,155 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.07 sec MapReduce Total cumulative CPU time: 2 seconds 70 msec

Ended Job = job\_1650262690593\_0003

MapReduce Jobs Launched:

Stage-Stage-3: Map: 1 Cumulative CPU: 2.07 sec HDFS Read: 6096 HDFS Write: 76 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 70 msec OK

1 deepali 1 A

2 mahesh 2 B

3 mangesh 3 C

4 ram 4 A

5 brijesh 5 B

6 john 6 C

Time taken: 45.062 seconds, Fetched: 6 row(s)