Output:

1. **HBase:**

   a. **Database creation and data insertion:**

```
[cloudera@quickstart ~]$ hbase shell
2022-03-28 02:00:26,036 INFO  [main] Configuration.deprecation: hadoop.native.lib
is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.0.0-cdh5.4.2, rUnknown, Tue May 19 17:07:29 PDT 2015

hbase(main):046:0> create 'flight3','finfo','fsch'
0 row(s) in 0.1540 seconds

=> Hbase::Table - flight3
hbase(main):003:0> list
TABLE
emphive
flight
flight1
flight3
3 row(s) in 0.0340 seconds

=> ["emphive", "flight", "flight1","flight3"]

hbase(main):047:0> put 'flight3', 1, 'finfo:source', 'pune'
0 row(s) in 0.0100 seconds

hbase(main):048:0> put 'flight3', 1, 'finfo:dest', 'mumbai'
0 row(s) in 0.0030 seconds

hbase(main):049:0> put 'flight3', 1, 'fsch:at', '10:25am'
0 row(s) in 0.0040 seconds

hbase(main):050:0> put 'flight3', 1, 'fsch:dt', '11:25am'
0 row(s) in 0.0030 seconds

hbase(main):051:0> put 'flight3', 1, 'fsch:delay', '5'
0 row(s) in 0.0070 seconds

hbase(main):052:0> put 'flight3', 1, 'fsch:delay', 5
0 row(s) in 0.0150 seconds

hbase(main):053:0> scan 'flight3'
ROW                     COLUMN+CELL
```

```
 1                    column=finfo:dest, timestamp=1648460661230, value=mumbai
 1                    column=finfo:source, timestamp=1648460653574, value=pune
 1                    column=fsch:at, timestamp=1648460757952, value=10:25am
 1                    column=fsch:delay, timestamp=1648460791327, value=5
 1                    column=fsch:dt, timestamp=1648460766642, value=11:25am
1 row(s) in 0.0090 seconds

hbase(main):054:0> put 'flight3', 2, 'finfo:source', 'pune'
0 row(s) in 0.0070 seconds

hbase(main):055:0> put 'flight3', 2, 'finfo:dest', 'kolkata'
0 row(s) in 0.0030 seconds

hbase(main):056:0> put 'flight3', 2, 'fsch:at', '7:00am'
0 row(s) in 0.0050 seconds

hbase(main):057:0> put 'flight3', 2, 'fsch:dt', '7:30am'
0 row(s) in 0.0070 seconds

hbase(main):058:0> put 'flight3', 2, 'fsch:delay', 2
0 row(s) in 0.0090 seconds

hbase(main):059:0> scan 'flight3'
ROW                  COLUMN+CELL
 1                    column=finfo:dest, timestamp=1648460661230, value=mumbai
 1                    column=finfo:source, timestamp=1648460653574, value=pune
 1                    column=fsch:at, timestamp=1648460757952, value=10:25am
 1                    column=fsch:delay, timestamp=1648460791327, value=5
 1                    column=fsch:dt, timestamp=1648460766642, value=11:25am
 2                    column=finfo:dest, timestamp=1648460855970, value=kolkata
 2                    column=finfo:source, timestamp=1648460840204, value=pune
 2                    column=fsch:at, timestamp=1648460873094, value=7:00am
 2                    column=fsch:delay, timestamp=1648460909952, value=2
 2                    column=fsch:dt, timestamp=1648460895394, value=7:30am
2 row(s) in 0.0340 seconds

hbase(main):060:0> put 'flight3', 3, 'finfo:source', 'mumbai'
0 row(s) in 0.0090 seconds

hbase(main):061:0> put 'flight3', 3, 'finfo:dest', 'pune'
0 row(s) in 0.0060 seconds

hbase(main):062:0> put 'flight3', 3, 'fsch:at', '12:30pm'
0 row(s) in 0.0020 seconds

hbase(main):063:0> put 'flight3', 3, 'fsch:dt', '12:45pm'
```

```
0 row(s) in 0.0050 seconds

hbase(main):064:0> put 'flight3', 3, 'fsch:delay', 1
0 row(s) in 0.0290 seconds

hbase(main):065:0> scan 'flight3'
ROW                     COLUMN+CELL
 1                      column=finfo:dest, timestamp=1648460661230, value=mumbai
 1                      column=finfo:source, timestamp=1648460653574, value=pune
 1                      column=fsch:at, timestamp=1648460757952, value=10:25am
 1                      column=fsch:delay, timestamp=1648460791327, value=5
 1                      column=fsch:dt, timestamp=1648460766642, value=11:25am
 2                      column=finfo:dest, timestamp=1648460855970, value=kolkata
 2                      column=finfo:source, timestamp=1648460840204, value=pune
 2                      column=fsch:at, timestamp=1648460873094, value=7:00am
 2                      column=fsch:delay, timestamp=1648460909952, value=2
 2                      column=fsch:dt, timestamp=1648460895394, value=7:30am
 3                      column=finfo:dest, timestamp=1648460962759, value=pune
 3                      column=finfo:source, timestamp=1648460945912,
value=mumbai
 3                      column=fsch:at, timestamp=1648460983714, value=12:30pm
 3                      column=fsch:delay, timestamp=1648461035365, value=1
 3                      column=fsch:dt, timestamp=1648461021398, value=12:45pm
3 row(s) in 0.0170 seconds

hbase(main):066:0>
```

b. Alter table:

```
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:

```
[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 tables;
OK
hbase_flight_new
hive_table_emp
Time taken: 0.484 seconds, Fetched: 2 row(s)

hive> CREATE external TABLE hbase_flight_new1(fno int, fsource string,fdest
string,fsh_at string,fsh_dt string,fsch_delay
   > string)
   > STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
   > WITH SERDEPROPERTIES ("hbase.columns.mapping" =
   > ":key,finfo:source,finfo:dest,fsch:at,fsch:dt,fsch:delay")
   > TBLPROPERTIES ("hbase.table.name" = "flight1");
OK
Time taken: 1.279 seconds
hive> show tables;
OK
hbase_flight_new
hbase_flight_new1
hive_table_emp
Time taken: 0.057 seconds, Fetched: 3 row(s)
hive> select * from hbase_flight_new1;
OK
1       pune    mumbai      10:25am     11:25am     5min
2       pune    kolkata7:00am7:30am2min
21      NULL NULL 7:00amNULL NULL
Time taken: 0.48 seconds, Fetched: 3 row(s)
```

b. Finding average delay of flights in the year 2008:

```
hive> show tables;
OK
hbase_flight_new
hbase_flight_new1
hbase_flight_new3
hive_table_emp
Time taken: 0.043 seconds, Fetched: 4 row(s)
hive> select * from hbase_flight_new3;
OK
1       pune    mumbai          10:25am         11:25am         5
2       pune    kolkata7:00am7:30am2
3       mumbai          pune    12:30pm         12:45pm         1
Time taken: 0.088 seconds, Fetched: 3 row(s)
hive>
    >
    > select * from hbase_flight_new3 where fdest='mumbai';
OK
1       pune    mumbai          10:25am         11:25am         5
Time taken: 0.104 seconds, Fetched: 1 row(s)
hive> select sum(fsch_delay) from hbase_flight_new3;
Query ID = cloudera_20220328025656_3046febb-185c-42d8-aabd-fd2cecf1f34d
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_1648457926797_0001, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1648457926797_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1648457926797_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-28 02:56:26,035 Stage-1 map = 0%,  reduce = 0%
2022-03-28 02:56:36,726 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.44
sec
2022-03-28 02:56:45,160 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU
2.34 sec
MapReduce Total cumulative CPU time: 2 seconds 340 msec
Ended Job = job_1648457926797_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.34 sec   HDFS Read: 7460
HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 340 msec
```

```
OK
8
Time taken: 35.183 seconds, Fetched: 1 row(s)
hive> select avg(fsch_delay) from hbase_flight_new3;
Query ID = cloudera_20220328025757_4dcdd6b7-c584-4368-bd43-92708c6c421b
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_1648457926797_0002, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1648457926797_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1648457926797_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-28 02:57:15,210 Stage-1 map = 0%,  reduce = 0%
2022-03-28 02:57:22,617 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.36
sec
2022-03-28 02:57:30,012 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU
2.27 sec
MapReduce Total cumulative CPU time: 2 seconds 270 msec
Ended Job = job_1648457926797_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.27 sec   HDFS Read: 7914
HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 270 msec
OK
2.6666666666666665
Time taken: 23.799 seconds, Fetched: 1 row(s)
```

c. Create table from local file:

```
hive> show tables;
OK
empdbnew1
hbase_flight_new
hbase_flight_new1
hbase_flight_new3
hive_table_emp
Time taken: 0.14 seconds, Fetched: 5 row(s)
hive> select * from empdbnew1;
OK
Time taken: 0.449 seconds
```

```
hive> load data local inpath '/home/cloudera/Desktop/empdb' into table empdbnew1;
Loading data to table default.empdbnew1
Table default.empdbnew1 stats: [numFiles=1, totalSize=33]
OK
Time taken: 0.469 seconds
hive> select * from empdbnew1;
OK
1       xyz     2000
2       pqr     3000
3       lmn     4000
Time taken: 0.062 seconds, Fetched: 3 row(s)
hive> create table empinfo(empno int, empgrade string) row format delimited fields
terminated by
    > ',' stored as textfile;
OK
Time taken: 0.053 seconds
hive> load data local inpath '/home/cloudera/Desktop/empinfo' into table empinfo;
Loading data to table default.empinfo
Table default.empinfo stats: [numFiles=1, totalSize=25]
OK
Time taken: 0.276 seconds
hive> select * from empinfo;
OK
1       A
2       B
3       B
4       B
5       B
6       A
NULL    NULL
Time taken: 0.076 seconds, Fetched: 7 row(s)
```

d. **Join tables:**

```
hive> SELECT eno, ename, empno, empgrade FROM empdbnew1 JOIN empinfo ON
eno = empno;
Query ID = cloudera_20220328032424_d6afebb6-fb36-473e-aff9-37003d28a908
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220328032424_d6afebb6-fb36-473e-
aff9-37003d28a908.log
2022-03-28 03:24:46  Starting to launch local task to process map join;      maximum
memory = 1013645312
```

2022-03-28 03:24:47 Dump the side-table for tag: 1 with group count: 6 into file: file:/tmp/cloudera/178a1cd0-da46-49c0-96a3-2a77b4c6d81d/hive_2022-03-28_03-24-41_727_5047632391853096952-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
2022-03-28 03:24:47 Uploaded 1 File to: file:/tmp/cloudera/178a1cd0-da46-49c0-96a3-2a77b4c6d81d/hive_2022-03-28_03-24-41_727_5047632391853096952-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (386 bytes)
2022-03-28 03:24:47 End of local task; Time Taken: 1.346 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_1648457926797_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1648457926797_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1648457926797_0003
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-03-28 03:24:57,412 Stage-3 map = 0%, reduce = 0%
2022-03-28 03:25:05,069 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.08 sec
MapReduce Total cumulative CPU time: 1 seconds 80 msec
Ended Job = job_1648457926797_0003
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.08 sec HDFS Read: 6041 HDFS Write: 30 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 80 msec
OK

1	xyz	1	A
2	pqr	2	B
3	lmn	3	B

Time taken: 24.453 seconds, Fetched: 3 row(s)