```
hive> create database employee; OK
hive> create table employee(
    > id int,
    > name String,
    > departmant String)
    > row format delimited fields terminated by ',';
OK
Time taken: 0.191 seconds
insert into employee values (4,'pratik','dev'),(5,'shubham','marketing')
hive> select * from employee;
OK
1     siddharth   salse
2     pranav      hr
3     harsh dev
4     pratik      dev
5     shubham     marketing
Time taken: 0.115 seconds, Fetched: 5 row(s) hive>
alter table employee add columns (salary int);
OK
Time taken: 0.105 seconds

hive> create external table if not exists emp2(
    > id INT,
    > name String,
    > salary INT)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/flights/';
OK
Time taken: 0.52 seconds

hive> select * from emp;
OK
1     siddharth   101
2     pratik      101
3     parth 102
4     param 102
5     shubham     103
Time taken: 0.157 seconds, Fetched: 5 row(s)
hive> select * from departments;
OK
101   salse
102   marketing
103   finance

select e.name , d.dName from emp e join departments d on e.dID = d.dId;
siddharth   salse pratik      salse
parth marketing
param marketing
shubham     finance
Time taken: 32.446 seconds, Fetched: 5 row(s)
```

```
hive> select e.name , d.dName from emp e left join departments d on e.dID
= d.dId;
OK
siddharth salse pratik
salse
parth marketing
param marketing
shubham    finance
Time taken: 22.261 seconds, Fetched: 5 row(s)



hive> create 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 ',';
OK
Time taken: 0.205 seconds hive>


[cloudera@quickstart ~]$ hdfs dfs -mkdir /flights
[cloudera@quickstart ~]$ hdfs dfs -put /media/sf_data/flight_info.csv
/flights/
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in
file:/etc/hive/conf.dist/hivelog4j.properties
```

```
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show tables; OK flights
Time taken: 0.592 seconds, Fetched: 1 row(s)

hive> load data inpath '/flights/flight_info.csv' into table flights;
Loading data to table default.flights
Table default.flights stats: [numFiles=2, totalSize=50227800]
OK
Time taken: 1.125 seconds
hive> select * from flights limit 10;
OK
2008 1      3     4         2003 1955 2211 2225 WN 335 N712SW           128
         150 116 -14 8           IAD TPA 810 4        80          0     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 0            0
            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 0            0
           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             70          0     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 0            0
      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 0            0
           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            70          0     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           70          0     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 0            0
           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        60          0     NA
          NA NA NA NA
Time taken: 0.626 seconds, Fetched: 10 row(s)
[cloudera@quickstart ~]$ hive> create index idx on table
flights(flight_num) as 'compact' with deferred rebuild;
OK
Time taken: 0.813 seconds
hive> create index idx2 on table flights(dep_delay) as 'compact' with
deferred rebuild;
OK
Time taken: 0.998 seconds


hive> select avg(dep_delay) as average_delay from flights;
```

```
Query ID = cloudera_20250425102424_bce06996-40c1-46e9-85de-75acc1d507c7
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_1745598840768_0001, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1745598840768_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill
job_1745598840768_0001
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2025-04-25 10:24:30,769 Stage-1 map = 0%,  reduce = 0%
2025-04-25 10:24:42,996 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
2.26 sec
2025-04-25 10:24:49,236 Stage-1 map = 100%,  reduce = 100%, Cumulative
CPU 3.41 sec
MapReduce Total cumulative CPU time: 3 seconds 410 msec Ended
Job = job_1745598840768_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.41 sec   HDFS Read:
50239189 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 410 msec
OK
13.104479405790656
Time taken: 47.696 seconds, Fetched: 1 row(s) hive>  hive> select day,
avg(dep_delay) as avg_delay2 from flights group by day;
Query ID = cloudera_20250425102727_d5514d01-0980-44b6-a0cc-73d0930ddaab
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_1745598840768_0002, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1745598840768_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill
job_1745598840768_0002
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2025-04-25 10:27:52,204 Stage-1 map = 0%,  reduce = 0%
2025-04-25 10:28:02,488 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
2.34 sec
2025-04-25 10:28:08,658 Stage-1 map = 100%,  reduce = 100%, Cumulative
CPU 3.29 sec
```

```
MapReduce Total cumulative CPU time: 3 seconds 290 msec Ended
Job = job_1745598840768_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.29 sec   HDFS Read:
50239570 HDFS Write: 664 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 290 msec
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: 25.878 seconds, Fetched: 32 row(s) hive>
hive> show indexes on flights; OK
idx                 flights             flight_num
default__flights_idx__ compact               idx2
flights             dep_delay
      default__flights_idx2__      compact
Time taken: 0.222 seconds, Fetched: 2 row(s)
```