# GROUP A) 3)

```
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> create database bhoomi;
OK
Time taken: 0.18 seconds
hive> use bhoomi;
OK
Time taken: 0.021 seconds
hive> create table flight(Fno int, Year int, Dest varchar(20), Delay float);
OK
Time taken: 0.302 seconds
hive> alter table flight rename to air_flight;
OK
Time taken: 0.127 seconds
hive> alter table air_flight add columns (source varchar(20));
OK
Time taken: 0.094 seconds
hive> drop table flight;
OK
Time taken: 0.016 seconds
hive> desc air_flight;
OK
fno                     int
year                    int
dest                    varchar(20)
delay                   float
source                  varchar(20)
Time taken: 0.091 seconds, Fetched: 5 row(s)
hive> create table flight (Fno int, Year int, Dest varchar(20), Delay float)
    > row format delimited
    > fields terminated by ','
    > lines terminated by '\n'
    > stored as textfile;
OK

hive> insert into flight values(123, 2009,"Mumbai", 30.6);
Query ID = cloudera_20250414215757_6eabb60f-e02a-420d-840a-bdbcb5bd373b
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_1739779370133_0006, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1739779370133_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1739779370133_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2025-04-14 21:58:02,763 Stage-1 map = 0%,  reduce = 0%
2025-04-14 21:58:08,103 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 se
c
MapReduce Total cumulative CPU time: 770 msec
Ended Job = job_1739779370133_0006
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/mydb1.db/fli
ght/.hive-staging_hive_2025-04-14_21-57-57_072_1822431826306728917-1/-ext-10000
Loading data to table mydb1.flight
Table mydb1.flight stats: [numFiles=1, numRows=1, totalSize=21, rawDataSize=20]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 0.77 sec   HDFS Read: 4157 HDFS Write: 8
9 SUCCESS
Total MapReduce CPU Time Spent: 770 msec
OK
Time taken: 12.282 seconds
```
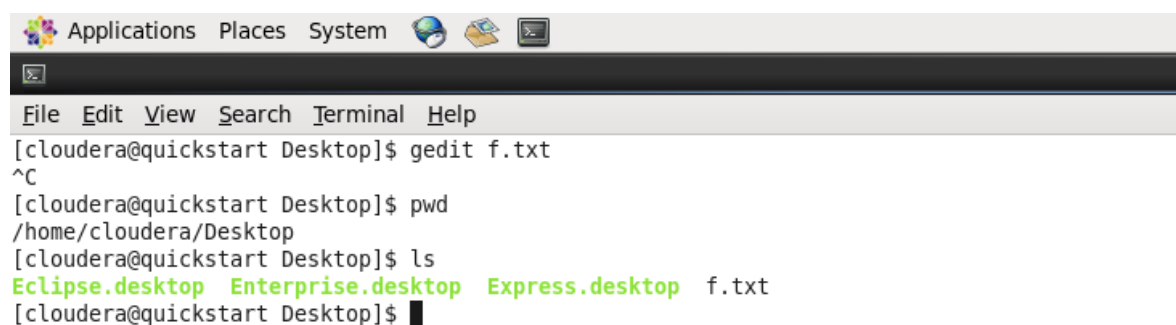
```
hive> insert into flight values (124, 2008, "Pune", 80.6);
Query ID = cloudera_20250414215858_2718e732-88ca-449f-8190-fed705cc4768
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1739779370133_0007, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1739779370133_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1739779370133_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2025-04-14 21:58:39,629 Stage-1 map = 0%,  reduce = 0%
2025-04-14 21:58:44,838 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.75 se
c
MapReduce Total cumulative CPU time: 750 msec
Ended Job = job_1739779370133_0007
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/mydb1.db/fli
ght/.hive-staging_hive_2025-04-14_21-58-34_139_7657106231335263385-1/-ext-10000
Loading data to table mydb1.flight
Table mydb1.flight stats: [numFiles=2, numRows=2, totalSize=40, rawDataSize=38]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 0.75 sec   HDFS Read: 4251 HDFS Write: 8
7 SUCCESS
Total MapReduce CPU Time Spent: 750 msec
OK
Time taken: 11.847 seconds
hive> select * from flight;
OK
123     2009    Mumbai  30.6
124     2008    Pune    80.6
Time taken: 0.038 seconds, Fetched: 2 row(s)
```
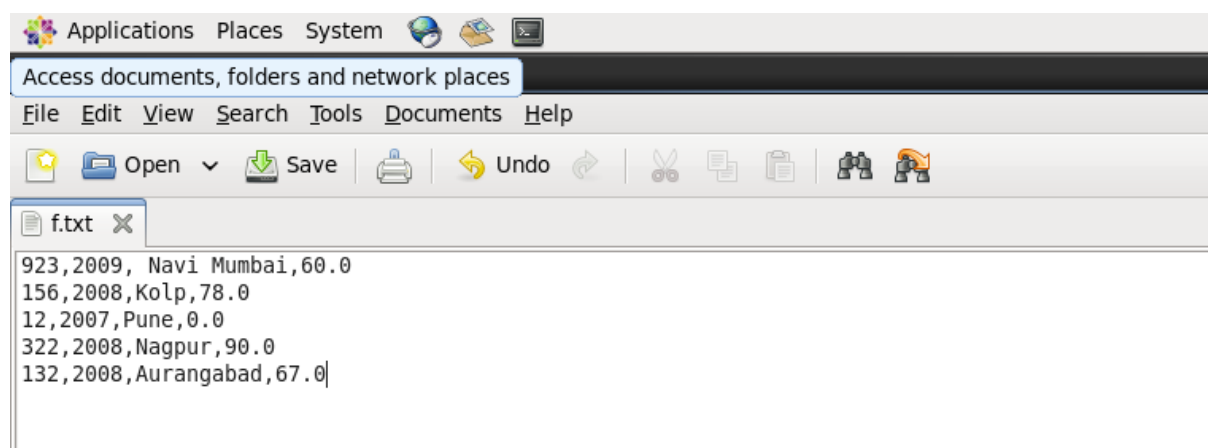
Applications   Places   System

File   Edit   View   Search   Terminal   Help

```
[cloudera@quickstart Desktop]$ gedit f.txt
^C
[cloudera@quickstart Desktop]$ pwd
/home/cloudera/Desktop
[cloudera@quickstart Desktop]$ ls
Eclipse.desktop  Enterprise.desktop  Express.desktop  f.txt
[cloudera@quickstart Desktop]$ 
```

Applications   Places   System

Access documents, folders and network places

File   Edit   View   Search   Tools   Documents   Help

Open   Save   Undo

f.txt

```
923,2009, Navi Mumbai,60.0
156,2008,Kolp,78.0
12,2007,Pune,0.0
322,2008,Nagpur,90.0
132,2008,Aurangabad,67.0
```

```
hive> load data local inpath 'f.txt'
    > overwrite into table flight;
Loading data to table mydb1.flight
Table mydb1.flight stats: [numFiles=1, numRows=0, totalSize=108, rawDataSize=0]
OK
Time taken: 0.146 seconds
hive>  select * from flight;
OK
923     2009    Navi Mumbai     60.0
156     2008    Kolp    78.0
12      2007    Pune    0.0
322     2008    Nagpur  90.0
132     2008    Aurangabad      67.0
Time taken: 0.023 seconds, Fetched: 5 row(s)

hive> create table nflight ( fno int, year int, source VARCHAR(50) )
    >  row format delimited
    >  fields terminated by ','
    > lines terminated by '\n'
    > stored as textfile;
OK
Time taken: 0.026 seconds
hive> select * from nflight;
OK
Time taken: 0.026 seconds
hive> select * from nflight;
OK
Time taken: 0.025 seconds
hive>  insert into nflight values (205, 2001, 'Pune');
Query ID = cloudera_20250414220202_308f257c-7a3a-40b0-a338-bd1f8e075f64
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_1739779370133_0008, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1739779370133_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1739779370133_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2025-04-14 22:03:05,449 Stage-1 map = 0%,   reduce = 0%
2025-04-14 22:03:10,628 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU 0.73 se
c
MapReduce Total cumulative CPU time: 730 msec
Ended Job = job_1739779370133_0008
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/mydb1.db/nfl
ight/.hive-staging_hive_2025-04-14_22-02-59_966_11051492501169711911-1/-ext-10000
Loading data to table mydb1.nflight
Table mydb1.nflight stats: [numFiles=1, numRows=1, totalSize=14, rawDataSize=13]

Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/mydb1.db/nfl
ight/.hive-staging_hive_2025-04-14_22-02-59_966_11051492501169711911-1/-ext-10000
Loading data to table mydb1.nflight
Table mydb1.nflight stats: [numFiles=1, numRows=1, totalSize=14, rawDataSize=13]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 0.73 sec   HDFS Read: 3894 HDFS Write: 8
3 SUCCESS
Total MapReduce CPU Time Spent: 730 msec
OK
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_20250414220505_a1378d32-4ae0-42c7-bf34-0d324992775c
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20250414220505_a1378d32-4ae0-42c7-bf34-
0d324992775c.log
2025-04-14 10:05:15     Starting to launch local task to process map join;     m
aximum memory = 1013645312
2025-04-14 10:05:16     Dump the side-table for tag: 1 with group count: 1 into
file: file:/tmp/cloudera/8622d6e8-ff35-4d9f-870c-5da9e9f08bba/hive_2025-04-14_22
-05-13_479_5409721657402272085-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile0
```

```
aximum memory = 1013645312
2025-04-14 10:05:16     Dump the side-table for tag: 1 with group count: 1 into
file: file:/tmp/cloudera/8622d6e8-ff35-4d9f-870c-5da9e9f08bba/hive_2025-04-14_22
-05-13_479_5409721657402272085-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile0
1--.hashtable
2025-04-14 10:05:16     Uploaded 1 File to: file:/tmp/cloudera/8622d6e8-ff35-4d9
f-870c-5da9e9f08bba/hive_2025-04-14_22-05-13_479_5409721657402272085-1/-local-10
003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (285 bytes)
2025-04-14 10:05:16     End of local task; Time Taken: 0.774 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_1739779370133_0009, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1739779370133_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1739779370133_0009
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-04-14 22:05:21,809 Stage-3 map = 0%,   reduce = 0%
2025-04-14 22:05:27,021 Stage-3 map = 100%,   reduce = 0%, Cumulative CPU 0.67 se
c
MapReduce Total cumulative CPU time: 670 msec
Ended Job = job_1739779370133_0009
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 0.67 sec   HDFS Read: 6761 HDFS Write: 0
 SUCCESS
Total MapReduce CPU Time Spent: 670 msec
OK
Time taken: 14.58 seconds
hive> select * from flight;
OK
923     2009    Navi Mumbai     60.0
156     2008    Kolp    78.0
12      2007    Pune    0.0
322     2008    Nagpur  90.0
132     2008    Aurangabad      67.0
Time taken: 0.027 seconds, Fetched: 5 row(s)
hive> select * from nflight;
OK
205     2001    Pune
Time taken: 0.024 seconds, Fetched: 1 row(s)


hive> create index flight_index on table flight (fno)
    > as 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
    > WITH DEFERRED REBUILD;
OK
Time taken: 0.087 seconds
hive> show tables;
OK
air_flight
flight
mydb1__flight_flight_index__
nflight
values__tmp__table__4
values__tmp__table__6
values__tmp__table__7
Time taken: 0.013 seconds, Fetched: 7 row(s)
```
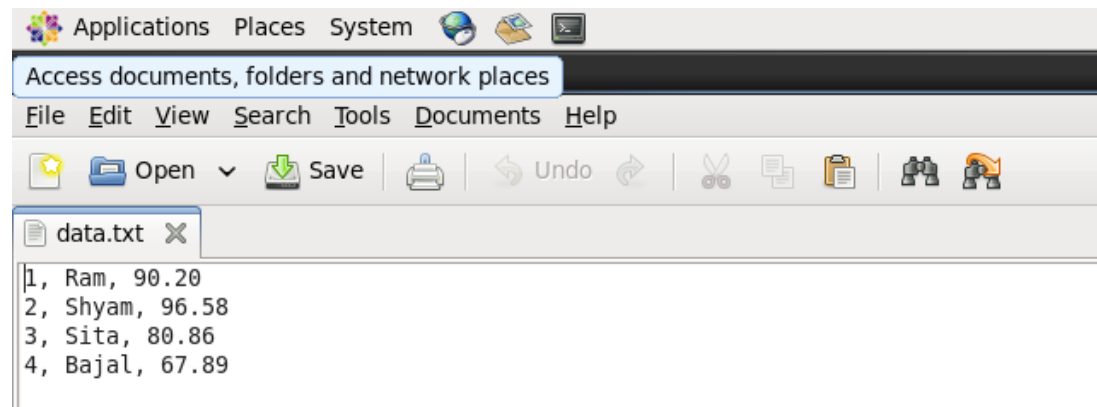
Loading Data from Internal to External Hive Table

```
hive> create database mydb2;
OK
Time taken: 0.038 seconds
hive> use mydb2;
OK
Time taken: 0.02 seconds
hive> create table hive_int (id int, name varchar(10), sal float)
    > row format delimited
    > fields terminated by ','
    > lines terminated by '\n'
    > stored as textfile;
OK
Time taken: 0.219 seconds
```



```
hive> load data local inpath 'data.txt' into table hive_int;
Loading data to table mydb2.hive_int
Table mydb2.hive_int stats: [numFiles=1, totalSize=61]
OK
Time taken: 0.518 seconds
hive> select * from hive_int;
OK
1        Ram     90.2
2        Shyam   96.58
3        Sita    80.86
4        Bajal   67.89
Time taken: 0.33 seconds, Fetched: 4 row(s)
hive> create external table hive_table (id int, name varchar(10), sal float)
    > row format delimited
    > fields terminated by ','
    > lines terminated by '\n'
    > stored as textfile;
```

```
hive> insert into hive_ext
    > select * from hive_int;
FAILED: SemanticException [Error 10001]: Line 1:12 Table not found 'hive_ext'
hive> insert into hive_table select * from hive_int;
Query ID = cloudera_20250414225353_fd574e97-5139-4e85-93ca-514664513202
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_1739779370133_0010, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1739779370133_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1739779370133_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2025-04-14 22:53:22,216 Stage-1 map = 0%,   reduce = 0%
2025-04-14 22:53:28,845 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU 0.61 se
c
MapReduce Total cumulative CPU time: 610 msec
Ended Job = job_1739779370133_0010
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/mydb2.db/hiv
e_table/.hive-staging_hive_2025-04-14_22-53-13_213_8637397535736805888-1/-ext-10
000
Loading data to table mydb2.hive_table
Table mydb2.hive_table stats: [numFiles=1, numRows=4, totalSize=56, rawDataSize=
52]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 0.61 sec   HDFS Read: 3428 HDFS Write: 1
28 SUCCESS
Total MapReduce CPU Time Spent: 610 msec
OK
Time taken: 16.918 seconds
hive> select * from hive_table;
OK
1       Ram     90.2
2       Shyam   96.58
3       Sita    80.86
4       Bajal   67.89
Time taken: 0.036 seconds, Fetched: 4 row(s)
hive> █
```