

```

hive> create database hive1;
OK
Time taken: 2.371 seconds
hive> use hive1;
OK
Time taken: 0.268 seconds
hive> create table customers(id bigint,name string,adress string);
OK
Time taken: 1.263 seconds
hive> show tables;
OK
customers
Time taken: 0.322 seconds, Fetched: 1 row(s)
hive> describe customers;
OK
id                bigint
name              string
adress            string
Time taken: 0.976 seconds, Fetched: 3 row(s)
hive>

```

```

hive> describe formatted customers;
OK
# col_name          data_type          comment
id                  bigint
name               string
adress             string

# Detailed Table Information
Database:           hive1
Owner:              hduser
CreateTime:         Thu Jun 30 10:32:35 IST 2022
LastAccessTime:     UNKNOWN
Protect Mode:       None
Retention:          0
Location:           hdfs://localhost:54310/user/hive/warehouse/hive1.db/customers
Table Type:         MANAGED_TABLE
Table Parameters:
    transient_lastDdlTime    1656565355

# Storage Information
SerDe Library:      org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:        org.apache.hadoop.mapred.TextInputFormat
OutputFormat:       org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed:         No
Num Buckets:        -1

```

```

Time taken: 0.523 seconds, Fetched: 28 row(s)
hive> insert into customers values(1111,"Raju","WA")
> ;
Query ID = hduser_20220630103854_88ff708e-9f50-47c5-8615-a3fe27543a07
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The driver is
automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
Starting Job = job_1656563497466_0001, Tracking URL = http://Inceptez:8088/proxy/application_1656563497466_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1656563497466_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-06-30 10:40:16,648 Stage-1 map = 0%, reduce = 0%
2022-06-30 10:40:31,212 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.67 sec
MapReduce Total cumulative CPU time: 2 seconds 670 msec
Ended Job = job_1656563497466_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://localhost:54310/user/hive/warehouse/hive1.db/customers/.hive-staging_hive_2022-06-30_10-38-54_763_542
3511811219728147-1/-ext-10000
Loading data to table hive1.customers
[Error 30017]: Skipping stats aggregation by error org.apache.hadoop.hive.ql.metadata.HiveException: [Error 30001]: StatsPub
lisher cannot be initialized. There was a error in the initialization of StatsPublisher, and retrying might help. If you don
t want the query to fail because accurate statistics could not be collected, set hive.stats.reliable=false
Table hive1.customers stats: [numFiles=1, totalSize=13]

```

```
Number of reduce tasks is set to 0 since there's no reduce operator
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is
automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
Starting Job = job_1656563497466_0001, Tracking URL = http://Inceptez:8088/proxy/application_1656563497466_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1656563497466_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-06-30 10:40:16,648 Stage-1 map = 0%, reduce = 0%
2022-06-30 10:40:31,212 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.67 sec
MapReduce Total cumulative CPU time: 2 seconds 670 msec
Ended Job = job_1656563497466_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://localhost:54310/user/hive/warehouse/hive1.db/customers/.hive-staging_hive_2022-06-30_10-38-54_763_547
3511811219728147-1/-ext-10000
Loading data to table hive1.customers
[Error 30017]: Skipping stats aggregation by error org.apache.hadoop.hive.ql.metadata.HiveException: [Error 30001]: StatsPub
lisher cannot be initialized. There was a error in the initialization of StatsPublisher, and retrying might help. If you don
t want the query to fail because accurate statistics could not be collected, set hive.stats.reliable=false
Table hive1.customers stats: [numFiles=1, totalSize=13]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 2.67 sec HDFS Read: 3794 HDFS Write: 13 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 670 msec
OK
Time taken: 99.51 seconds
hive>
```

```
hive> insert into customers values (2222,"jack","WA"),(3333,"john","CA"),(4444,"Nina","NY");
Query ID = hduser_20220630104332_90e21fe4-37ef-4a29-a25d-4db1eb3b7b42
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_1656563497466_0002, Tracking URL = http://Inceptez:8088/proxy/application_1656563497466_0002/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1656563497466_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-06-30 10:44:38,842 Stage-1 map = 0%, reduce = 0%
2022-06-30 10:44:51,794 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.44 sec
MapReduce Total cumulative CPU time: 2 seconds 440 msec
Ended Job = job_1656563497466_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://localhost:54310/user/hive/warehouse/hive1.db/customers/.hive-staging_hive_2022-06-30_10-43-32_947_208
7314548593707755-1/-ext-10000
Loading data to table hive1.customers
[Error 30017]: Skipping stats aggregation by error org.apache.hadoop.hive.ql.metadata.HiveException: [Error 30001]: StatsPub
lisher cannot be initialized. There was a error in the initialization of StatsPublisher, and retrying might help. If you don
t want the query to fail because accurate statistics could not be collected, set hive.stats.reliable=false
Table hive1.customers stats: [numFiles=2, totalSize=52]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 2.44 sec HDFS Read: 3878 HDFS Write: 39 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 440 msec
```

```
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_1656563497466_0002, Tracking URL = http://Inceptez:8088/proxy/application_1656563497466_0002/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1656563497466_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-06-30 10:44:38,842 Stage-1 map = 0%, reduce = 0%
2022-06-30 10:44:51,794 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.44 sec
MapReduce Total cumulative CPU time: 2 seconds 440 msec
Ended Job = job_1656563497466_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://localhost:54310/user/hive/warehouse/hive1.db/customers/.hive-staging_hive_2022-06-30_10-43-32_947_208
7314548593707755-1/-ext-10000
Loading data to table hive1.customers
[Error 30017]: Skipping stats aggregation by error org.apache.hadoop.hive.ql.metadata.HiveException: [Error 30001]: StatsPub
lisher cannot be initialized. There was a error in the initialization of StatsPublisher, and retrying might help. If you don
t want the query to fail because accurate statistics could not be collected, set hive.stats.reliable=false
Table hive1.customers stats: [numFiles=2, totalSize=52]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 2.44 sec HDFS Read: 3878 HDFS Write: 39 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 440 msec
OK
Time taken: 82.399 seconds
```

```

hive> insert into orders values(11111,"phone",1111,3,1200),(11112,"cam",1111,1,5200),(11113,"broom",2222,2,20),(11114,"broom",2222,2,20),(11115,"t-shirt",4444,2,66);
Query ID = hduser_20220630105209_88516f62-72e0-41e4-8e13-edbf1963c267
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_1656563497466_0003, Tracking URL = http://Inceptez:8088/proxy/application_1656563497466_0003/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1656563497466_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-06-30 10:52:53,642 Stage-1 map = 0%, reduce = 0%
2022-06-30 10:53:05,750 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.42 sec
MapReduce Total cumulative CPU time: 2 seconds 420 msec
Ended Job = job_1656563497466_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://localhost:54310/user/hive/warehouse/hive1.db/orders/.hive-staging_hive_2022-06-30_10-52-09_295_1601271703866053476-1/-ext-10000
Loading data to table hive1.orders
[Error 30017]: Skipping stats aggregation by error org.apache.hadoop.hive.ql.metadata.HiveException: [Error 30001]: StatsPublisher cannot be initialized. There was a error in the initialization of StatsPublisher, and retrying might help. If you don't want the query to fail because accurate statistics could not be collected, set hive.stats.reliable=false
Table hive1.orders stats: [numFiles=1, totalSize=129]
MapReduce Jobs Launched:

```

```

[hduser@localhost ~]$ hdfs dfs -ls /user/hive/warehouse/hive1.db
22/06/30 10:56:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x - hduser supergroup 0 2022-06-30 10:44 /user/hive/warehouse/hive1.db/customers
drwxr-xr-x - hduser supergroup 0 2022-06-30 10:53 /user/hive/warehouse/hive1.db/orders
[hduser@localhost ~]$

```

```

hive> select * from customers where address='CA';
FAILED: SemanticException [Error 10004]: Line 1:30 Invalid table alias or column reference 'address': (possible column names are: id, name, address)
hive> select * from customers where adress='CA';
OK
3333 john CA
Time taken: 0.56 seconds, Fetched: 1 row(s)
hive>

```

```

hive> select distinct adress from customers;
Query ID = hduser_20220630110157_7726fad3-24c2-45f1-8e3b-f5087cef8f45
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_1656563497466_0004, Tracking URL = http://Inceptez:8088/proxy/application_1656563497466_0004/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1656563497466_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-30 11:02:27,129 Stage-1 map = 0%, reduce = 0%
2022-06-30 11:02:42,427 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.43 sec
2022-06-30 11:02:54,687 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.36 sec
MapReduce Total cumulative CPU time: 5 seconds 360 msec
Ended Job = job_1656563497466_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.36 sec HDFS Read: 6557 HDFS Write: 9 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 360 msec
OK
CA
NY
WA

```

LOAD DATA FROM LOCAL SYSTEM:

```
hive> create table if not exists customers1(id string,name string,age int)
> row format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 1.129 seconds

hive> load data local inpath '/home/hduser/hivefile1.csv' into table customers1;
Loading data to table default.customers1
Table default.customers1 stats: [numFiles=1, totalSize=35]
OK
Time taken: 1.607 seconds
hive> select * from customers1;
OK
1      keerthana      20
2      raju          24
3      sree          30
Time taken: 0.828 seconds, Fetched: 3 row(s)
hive>
```

LOAD DATA FROM HDFS:

```
[hduser@localhost ~]$ hdfs dfs -mkdir /user/hduser/hivedata
22/06/30 11:30:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c
lasses where applicable
[hduser@localhost ~]$ hdfs dfs -ls
22/06/30 11:30:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c
lasses where applicable
Found 48 items
-rw-r--r--  1 hduser supergroup      15 2022-06-26 16:59 .passwordfile
drwxr-xr-x  - hduser supergroup      0 2022-06-26 16:13 Assignkeerthu1
drwxr-xr-x  - hduser supergroup      0 2022-06-26 16:22 Assignkeerthu2
drwxr-xr-x  - hduser supergroup      0 2022-06-26 15:49 Keerassign
drwxr-xr-x  - hduser supergroup      0 2022-06-23 19:52 Keerthu
drwxr-xr-x  - hduser supergroup      0 2022-06-26 17:26 Keerthupass
-rw-r--r--  1 hduser supergroup      7 2022-06-26 13:07 Password
-rw-r--r--  1 hduser supergroup     73 2022-03-01 00:38 Stud
drwxr-xr-x  - hduser supergroup      0 2021-10-01 22:38 StudentData
drwxr-xr-x  - hduser supergroup      0 2021-10-01 21:55 StudentData2
drwxr-xr-x  - hduser supergroup      0 2022-06-26 17:26 _sqoop
drwxr-xr-x  - hduser supergroup      0 2022-06-26 12:57 assign3d
drwxr-xr-x  - hduser supergroup      0 2022-06-25 18:26 assign3q
drwxr-xr-x  - hduser supergroup      0 2022-06-25 19:17 assign3qb
drwxr-xr-x  - hduser supergroup      0 2022-06-24 22:13 bound
drwxr-xr-x  - hduser supergroup      0 2022-06-24 22:16 bound1
drwxr-xr-x  - hduser supergroup      0 2022-06-25 13:37 bound2
drwxr-xr-x  - hduser supergroup      0 2021-10-12 19:15 catalog
drwxr-xr-x  - hduser supergroup      0 2021-10-16 13:16 credits_cst

-rw-r--r--  1 hduser supergroup      15 2022-06-26 16:47 passfile
drwxr-xr-x  - hduser supergroup      0 2021-10-26 02:29 paymentsdata
drwxr-xr-x  - hduser supergroup      0 2021-10-20 01:13 retailorders
drwxr-xr-x  - hduser supergroup      0 2022-06-25 16:35 sqoop_all_table
drwxr-xr-x  - hduser supergroup      0 2022-06-25 16:39 sqoop_all_table1
drwxr-xr-x  - hduser supergroup      0 2022-06-25 16:43 sqoop_all_table2
drwxr-xr-x  - hduser supergroup      0 2022-06-24 21:05 sqoop_importdir
drwxr-xr-x  - hduser supergroup      0 2022-06-24 21:08 sqoop_importdir1
drwxr-xr-x  - hduser supergroup      0 2022-06-25 14:44 sqoop_importdir2
drwxr-xr-x  - hduser supergroup      0 2021-10-19 15:46 states
drwxr-xr-x  - hduser supergroup      0 2022-01-04 22:00 test
drwxr-xr-x  - hduser supergroup      0 2022-06-24 10:02 user
[hduser@localhost ~]$ hdfs dfs -put hivefile1.csv /user/hduser/hivedata/
22/06/30 11:41:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c
lasses where applicable

hive> load data inpath '/user/hduser/hivedata/hivefile1.csv' into table customers1;
Loading data to table default.customers1
Table default.customers1 stats: [numFiles=2, totalSize=70]
OK
Time taken: 1.631 seconds
hive>
```