



Apache Hive

Part 1

by Sumit Mittal



TRENDY TECH

UPLIFT YOUR CAREER!

Apache Hive

Exercise 1

Basic Hive Queries



IMPORTANT

Copyright Infringement and Illegal Content Sharing Notice

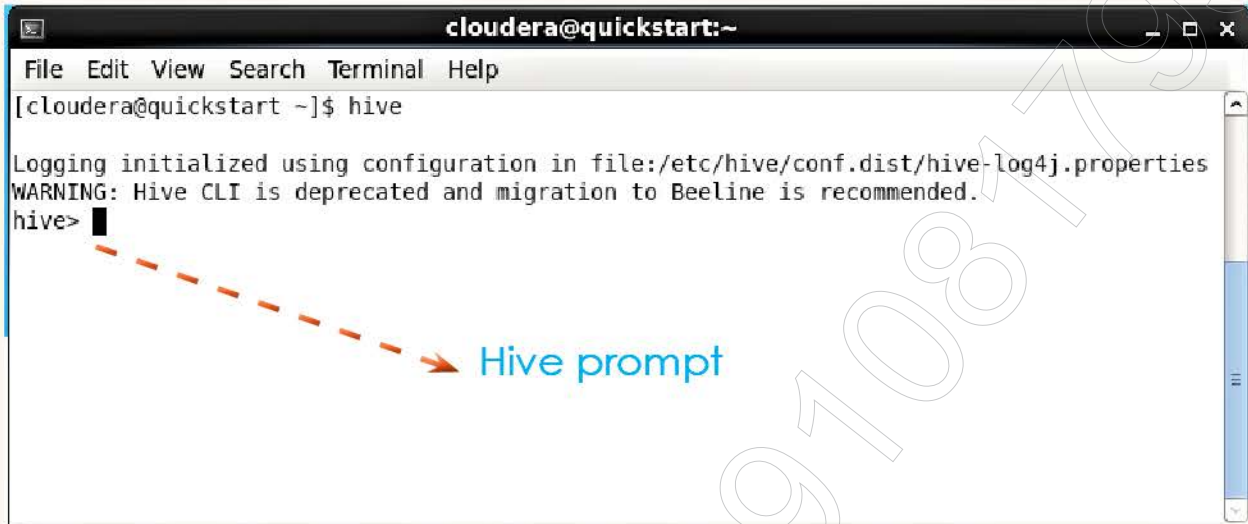
All course content designs, video, audio, text, graphics, logos, images are Copyright© and are protected by India and international copyright laws. All rights reserved.

Permission to download the contents (wherever applicable) for the sole purpose of individual reading and preparing yourself to crack the interview only. Any other use of study materials – including reproduction, modification, distribution, republishing, transmission, display – without the prior written permission of Author is strictly prohibited.

Trendytech Insights legal team, along with thousands of our students, actively searches the Internet for copyright infringements. Violators subject to prosecution.

To enter Hive shell:

hive

A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command '[cloudera@quickstart ~]\$ hive' has been entered. The output shows 'Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties' and a 'WARNING: Hive CLI is deprecated and migration to Beeline is recommended.' The prompt 'hive>' is shown. A dashed orange arrow points from the 'hive>' prompt to the text 'Hive prompt' in blue.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[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>
```

To display databases:

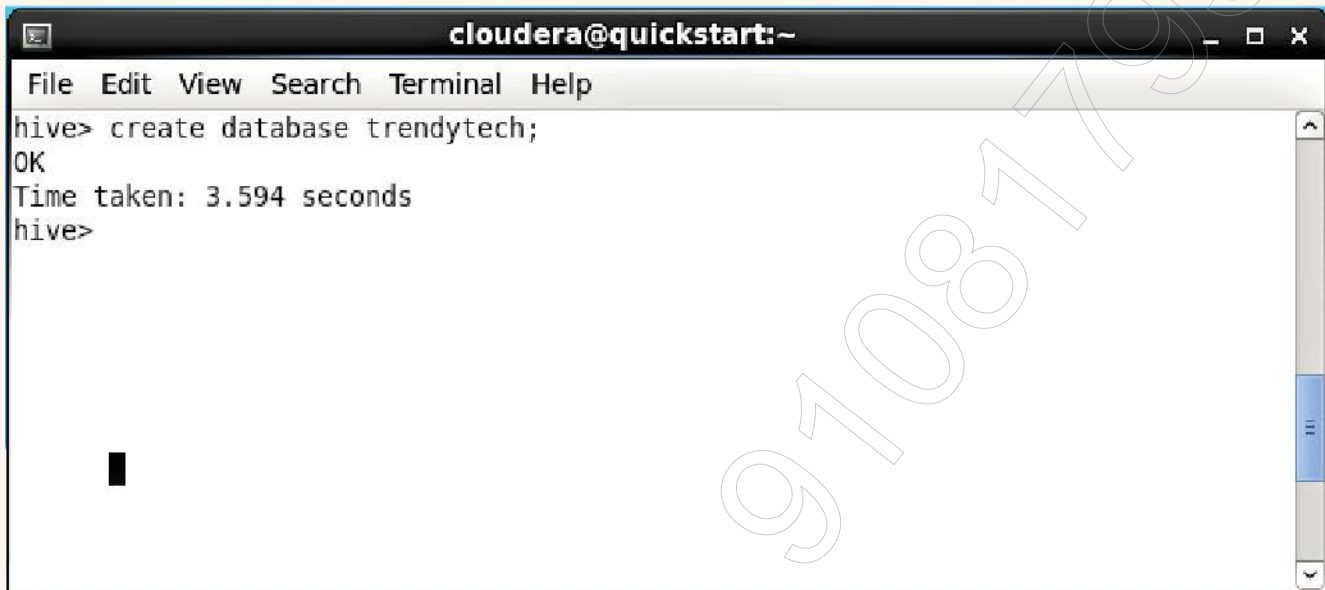
show databases;

A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'hive> show databases;' has been entered. The output shows 'OK' and 'default' (which is highlighted with a red box), followed by 'Time taken: 5.557 seconds, Fetched: 1 row(s)'. The prompt 'hive>' is shown at the bottom.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive>  
hive> show databases;  
OK  
default  
Time taken: 5.557 seconds, Fetched: 1 row(s)  
hive>
```

Create a new database:

```
create database trendytech;
```

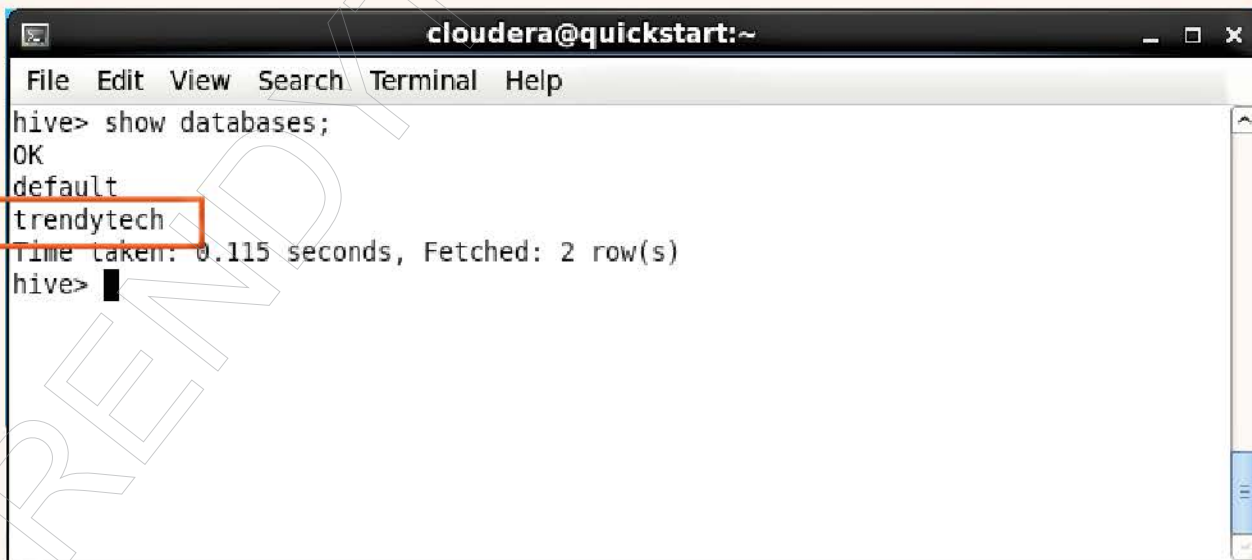


A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'hive> create database trendytech;' is entered and executed. The output shows 'OK' and 'Time taken: 3.594 seconds'. The prompt 'hive>' is visible at the bottom with a cursor.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> create database trendytech;  
OK  
Time taken: 3.594 seconds  
hive>
```

To list databases:

```
show databases;
```

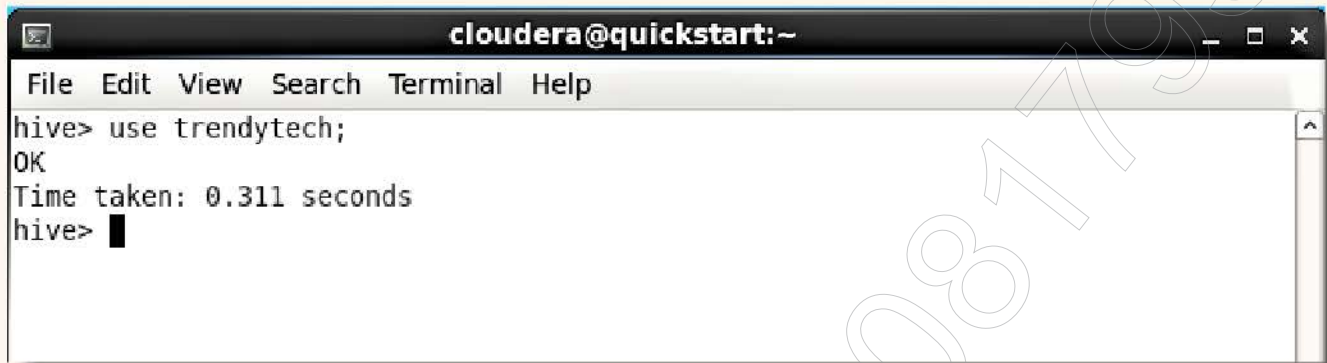


A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'hive> show databases;' is entered and executed. The output shows 'OK', 'default', and 'trendytech' (highlighted with a red box). Below this, it says 'Time taken: 0.115 seconds, Fetched: 2 row(s)'. The prompt 'hive>' is visible at the bottom with a cursor.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> show databases;  
OK  
default  
trendytech  
Time taken: 0.115 seconds, Fetched: 2 row(s)  
hive>
```


Use a database:

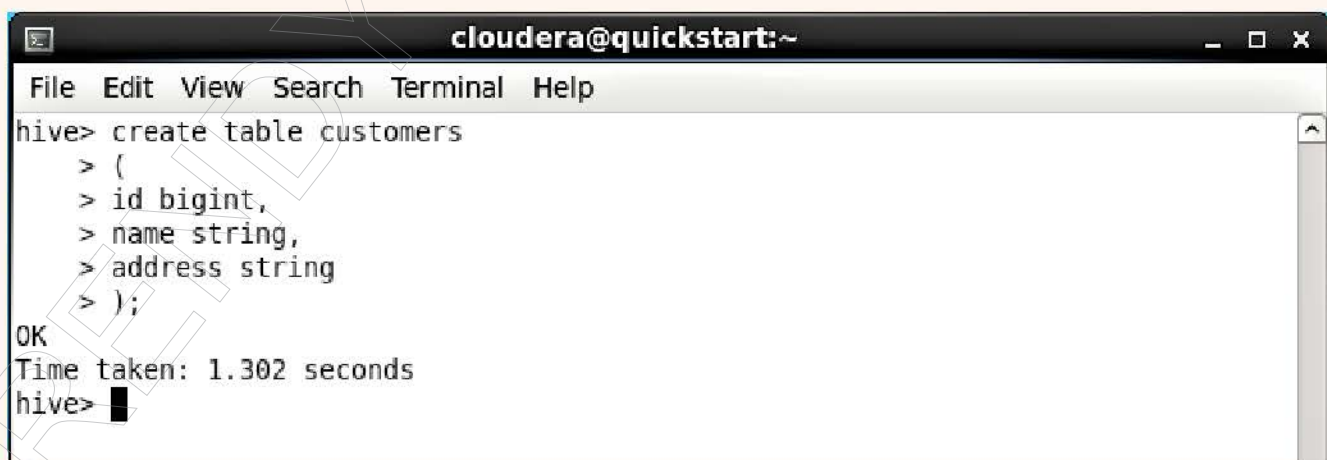
```
use trendytech;
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> use trendytech;  
OK  
Time taken: 0.311 seconds  
hive> █
```

Create a table:

```
create table customers (  
id bigint,  
name string,  
address string  
);
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> create table customers  
> (  
> id bigint,  
> name string,  
> address string  
> );  
OK  
Time taken: 1.302 seconds  
hive> █
```

List tables in a database:

```
show tables;
```

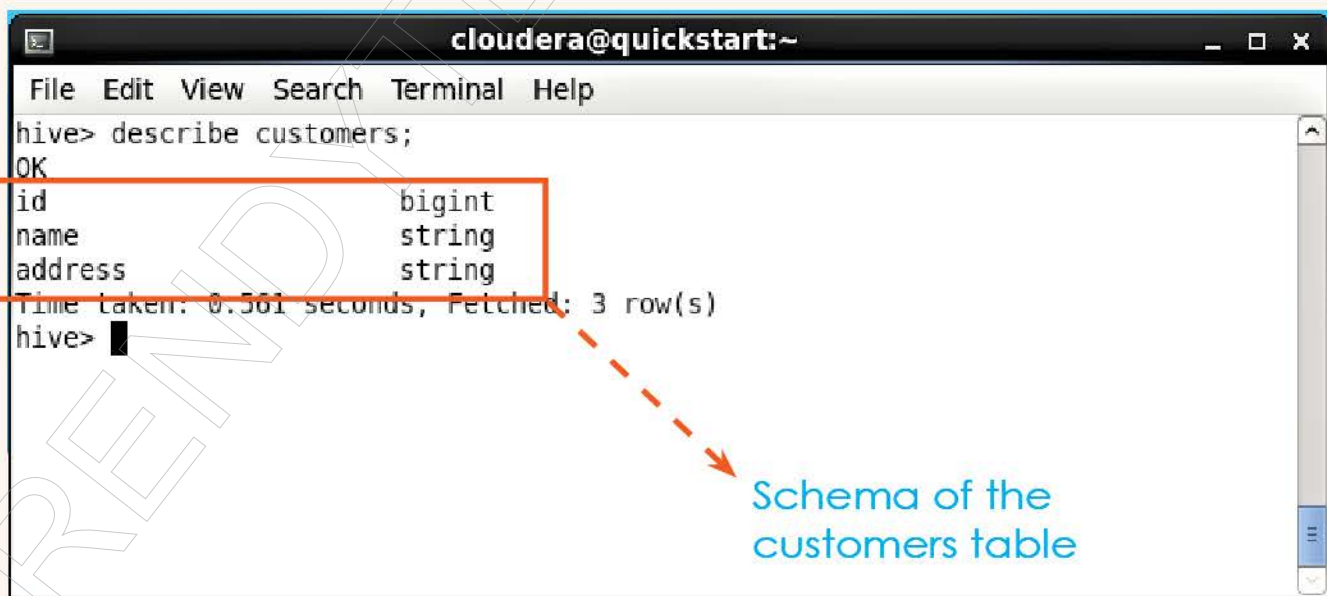


A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'hive> show tables;' is entered, followed by 'OK' and the output 'customers', which is highlighted with a red box. Below the output, it says 'Time taken: 0.324 seconds, Fetched: 1 row(s)'. The prompt 'hive>' is shown at the bottom.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> show tables;  
OK  
customers  
Time taken: 0.324 seconds, Fetched: 1 row(s)  
hive>
```

Display schema of a hive table:

```
describe customers;
```



A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'hive> describe customers;' is entered, followed by 'OK' and the output schema:

id	bigint
name	string
address	string

The schema output is highlighted with a red box. Below it, it says 'Time taken: 0.501 seconds, Fetched: 3 row(s)'. The prompt 'hive>' is shown at the bottom. A red dashed arrow points from the schema box to the text 'Schema of the customers table'.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> describe customers;  
OK  
id          bigint  
name        string  
address     string  
Time taken: 0.501 seconds, Fetched: 3 row(s)  
hive>
```

Schema of the customers table

Describe table with extra information:

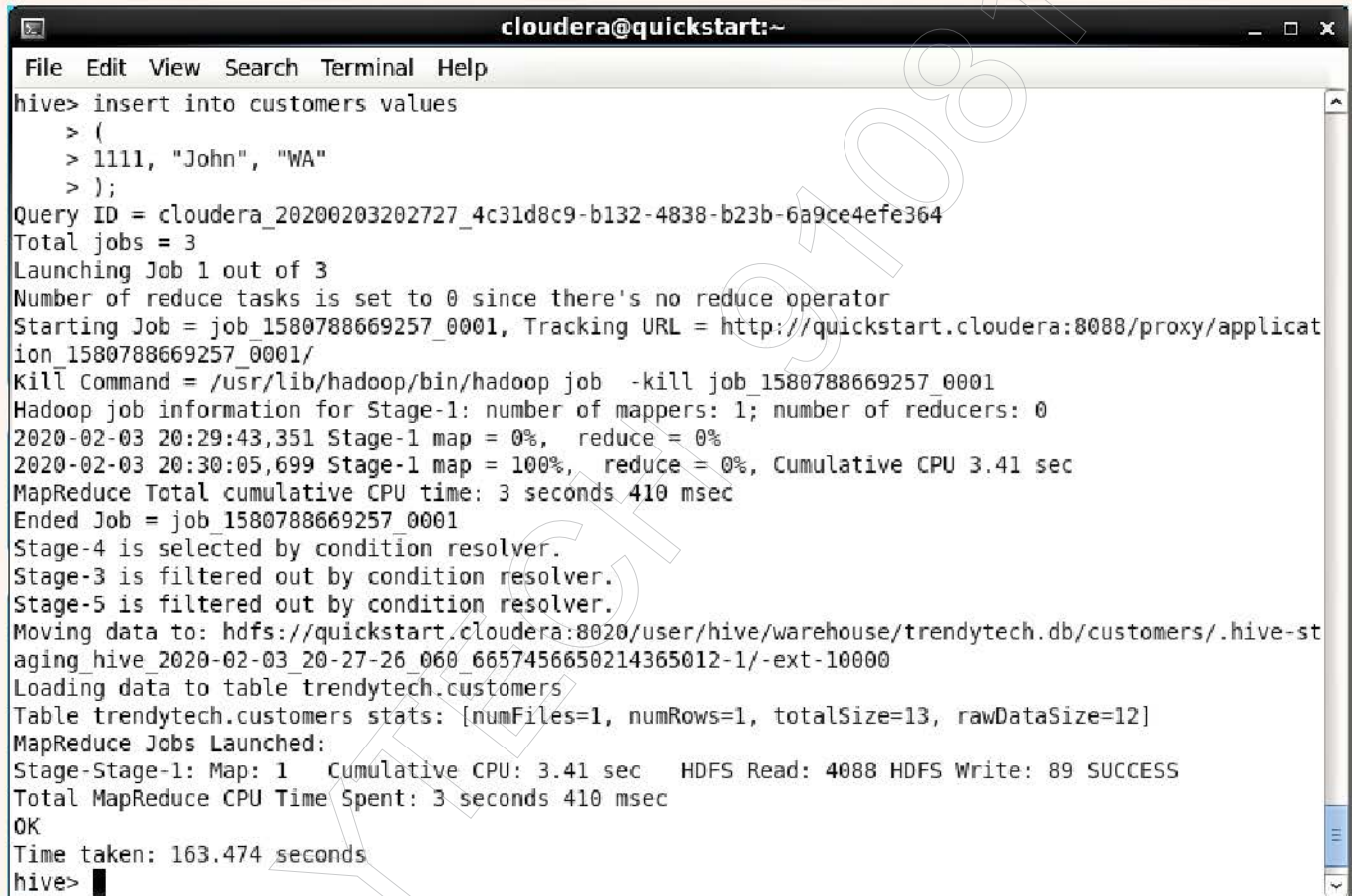
describe formatted customers;

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> describe formatted customers;  
OK  
# col_name          data_type          comment  
id                  bigint  
name                string  
address             string  
  
# Detailed Table Information  
Database:           trendytech  
Owner:              cloudera  
CreateTime:         Mon Feb 03 20:15:37 PST 2020  
LastAccessTime:     UNKNOWN  
Protect Mode:       None  
Retention:          0  
Location:           hdfs://quickstart.cloudera:8020/user/hive/warehouse/trendytech.db/customers  
Table Type:         MANAGED_TABLE  
Table Parameters:  
    transient_lastDdlTime 1580789737  
  
# 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  
Bucket Columns:     []  
Sort Columns:       []  
Storage Desc Params:  
    serialization.format 1  
Time taken: 0.187 seconds, Fetched: 28 row(s)  
hive>
```

More detailed informations

Insert record into Hive table:

```
insert into customers values  
(  
1111, "John", "WA"  
);
```

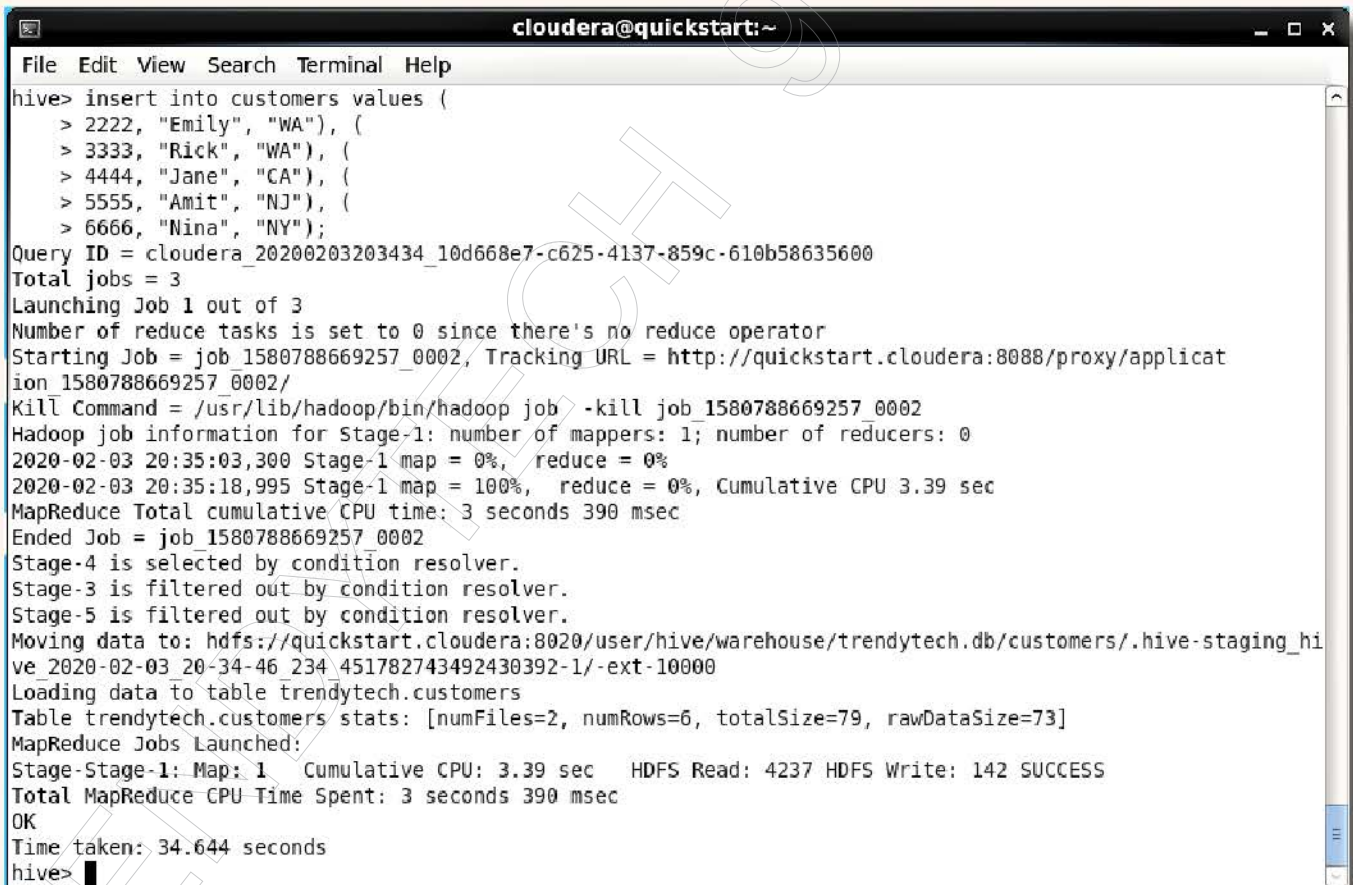


```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> insert into customers values  
  > (  
  > 1111, "John", "WA"  
  > );  
Query ID = cloudera_20200203202727_4c31d8c9-b132-4838-b23b-6a9ce4efe364  
Total jobs = 3  
Launching Job 1 out of 3  
Number of reduce tasks is set to 0 since there's no reduce operator  
Starting Job = job_1580788669257_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1580788669257_0001/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1580788669257_0001  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0  
2020-02-03 20:29:43,351 Stage-1 map = 0%, reduce = 0%  
2020-02-03 20:30:05,699 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.41 sec  
MapReduce Total cumulative CPU time: 3 seconds 410 msec  
Ended Job = job_1580788669257_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://quickstart.cloudera:8020/user/hive/warehouse/trendytech.db/customers/.hive-staging_hive_2020-02-03_20-27-26_060_6657456650214365012-1/-ext-10000  
Loading data to table trendytech.customers  
Table trendytech.customers stats: [numFiles=1, numRows=1, totalSize=13, rawDataSize=12]  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Cumulative CPU: 3.41 sec HDFS Read: 4088 HDFS Write: 89 SUCCESS  
Total MapReduce CPU Time Spent: 3 seconds 410 msec  
OK  
Time taken: 163.474 seconds  
hive>
```

Note: Insert will trigger a MapReduce job in background.

Insert multiple record into Hive table:

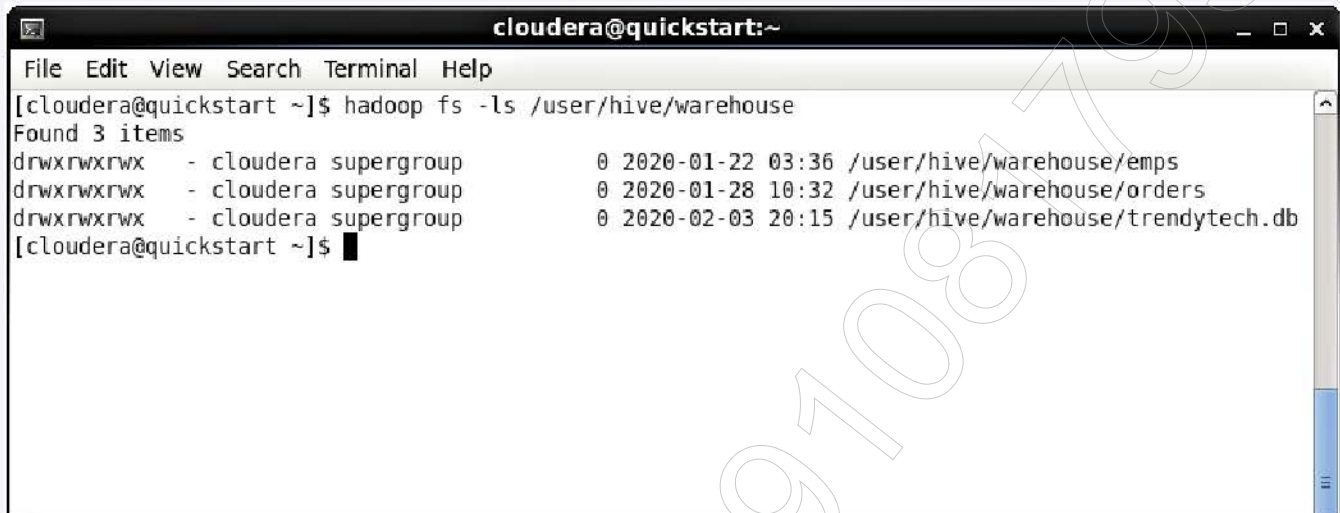
```
insert into customers values  
(  
2222, "Emily", "WA"), (  
3333, "Rick", "WA"), (  
4444, "Jane", "CA"), (  
5555, "Amit", "NJ"), (  
6666, "Nina", "NY"  
);
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> insert into customers values (  
  > 2222, "Emily", "WA"), (  
  > 3333, "Rick", "WA"), (  
  > 4444, "Jane", "CA"), (  
  > 5555, "Amit", "NJ"), (  
  > 6666, "Nina", "NY");  
Query ID = cloudera_20200203203434_10d668e7-c625-4137-859c-610b58635600  
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_1580788669257_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/applicat  
ion_1580788669257_0002/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1580788669257_0002  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0  
2020-02-03 20:35:03,300 Stage-1 map = 0%, reduce = 0%  
2020-02-03 20:35:18,995 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.39 sec  
MapReduce Total cumulative CPU time: 3 seconds 390 msec  
Ended Job = job_1580788669257_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://quickstart.cloudera:8020/user/hive/warehouse/trendytech.db/customers/.hive-staging_hi  
ve_2020-02-03_20-34-46_234_451782743492430392-1/-ext-10000  
Loading data to table trendytech.customers  
Table trendytech.customers stats: [numFiles=2, numRows=6, totalSize=79, rawDataSize=73]  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Cumulative CPU: 3.39 sec HDFS Read: 4237 HDFS Write: 142 SUCCESS  
Total MapReduce CPU Time Spent: 3 seconds 390 msec  
OK  
Time taken: 34.644 seconds  
hive>
```

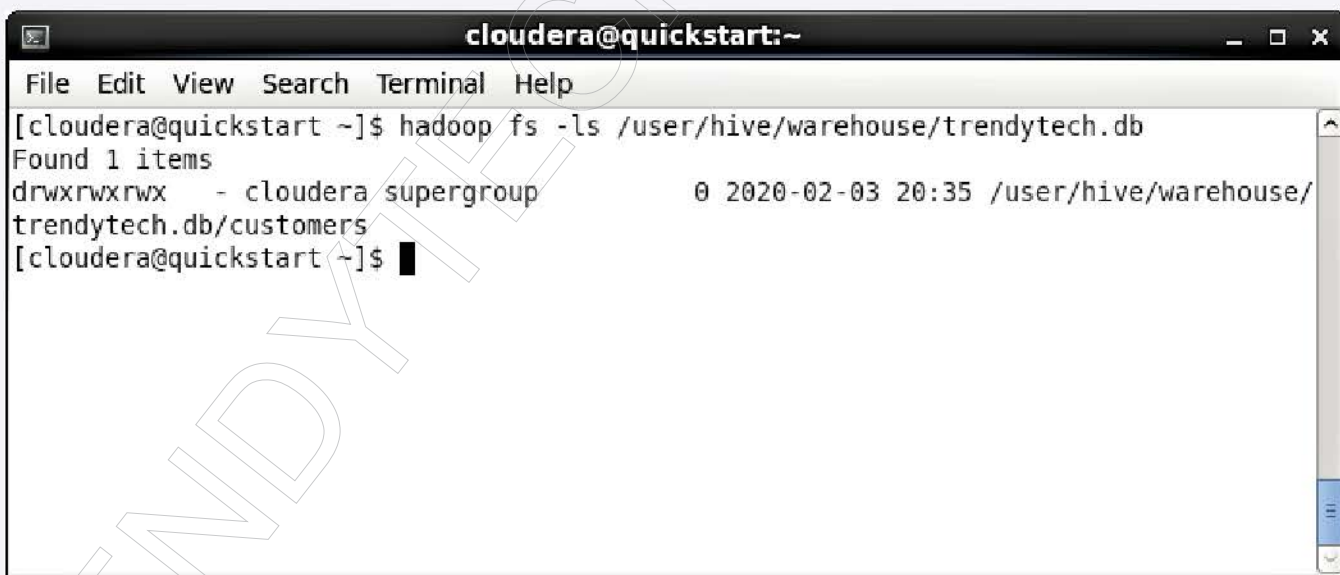
To check the hive warehouse directory in terminal:

```
hadoop fs -ls /user/hive/warehouse
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse  
Found 3 items  
drwxrwxrwx - cloudera supergroup 0 2020-01-22 03:36 /user/hive/warehouse/emp  
drwxrwxrwx - cloudera supergroup 0 2020-01-28 10:32 /user/hive/warehouse/orders  
drwxrwxrwx - cloudera supergroup 0 2020-02-03 20:15 /user/hive/warehouse/trendytech.db  
[cloudera@quickstart ~]$
```

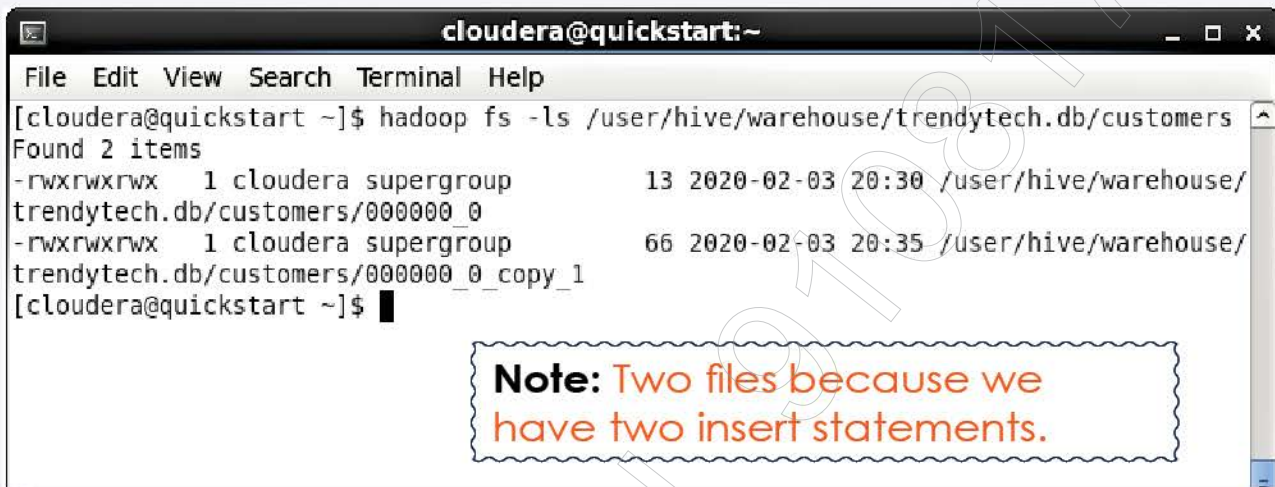
```
hadoop fs -ls /user/hive/warehouse/trendytech.db
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/trendytech.db  
Found 1 items  
drwxrwxrwx - cloudera supergroup 0 2020-02-03 20:35 /user/hive/warehouse/  
trendytech.db/customers  
[cloudera@quickstart ~]$
```


Check the folder structure of customers table in trendytech database from terminal:

```
hadoop fs -ls /user/hive/warehouse/  
trendytech.db/customers
```

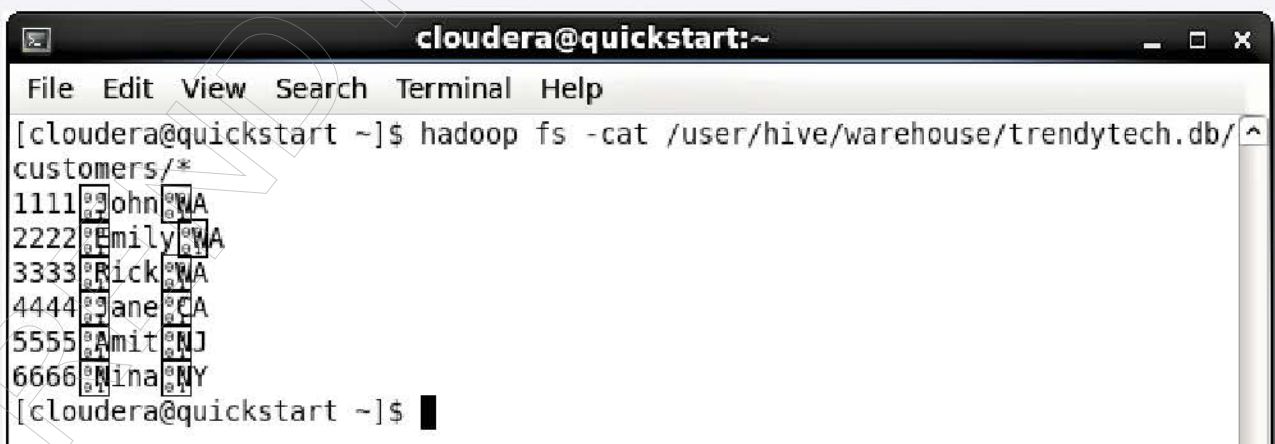


```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/trendytech.db/customers  
Found 2 items  
-rwxrwxrwx 1 cloudera supergroup      13 2020-02-03 20:30 /user/hive/warehouse/  
trendytech.db/customers/000000_0  
-rwxrwxrwx 1 cloudera supergroup      66 2020-02-03 20:35 /user/hive/warehouse/  
trendytech.db/customers/000000_0_copy_1  
[cloudera@quickstart ~]$
```

Note: Two files because we have two insert statements.

Read the contents of customers table from terminal:

```
hadoop fs -cat /user/hive/warehouse/  
trendytech.db/customers/*
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/trendytech.db/  
customers/*  
1111 John WA  
2222 Emily WA  
3333 Rick WA  
4444 Jane CA  
5555 Amit NJ  
6666 Nina NY  
[cloudera@quickstart ~]$
```



5 Star Google Rated
Big Data Course

LEARN FROM THE EXPERT



9108179578

Call for more details

Follow US

Trainer Mr. Sumit Mittal

LinkedIn <https://www.linkedin.com/in/bigdatabysumit/>

Website <https://trendytech.in/courses/big-data-online-training/>

Phone 9108179578

Email trendytech.sumit@gmail.com

Youtube TrendyTech

Twitter @BigdataBySumit

Instagram bigdatabysumit

Facebook <https://www.facebook.com/trendytech.in/>

