



# **Apache Hive**

## **Part 2**

by Sumit Mittal



**TRENDY TECH**

UPLIFT YOUR CAREER!

# Apache Hive

## Exercise 2

### Basic Hive Queries



# IMPORTANT

## Copyright Infringement and Illegal Content Sharing Notice

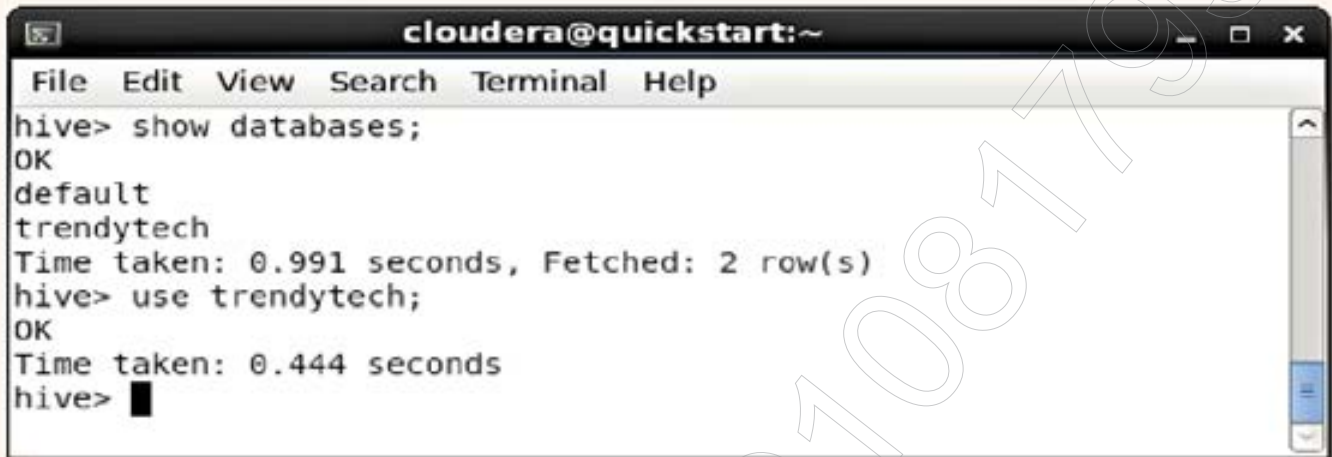
All course content designs, video, audio, text, graphics, logos, images are Copyright© and are protected by India and international copyright laws. All rights reserved.

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 show all databases present in Hive:

`show databases;`

A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the execution of 'hive> show databases;' which returns 'OK' and lists 'default' and 'trendytech'. It then shows 'hive> use trendytech;' which also returns 'OK'. The prompt 'hive>' is followed by a cursor.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> show databases;  
OK  
default  
trendytech  
Time taken: 0.991 seconds, Fetched: 2 row(s)  
hive> use trendytech;  
OK  
Time taken: 0.444 seconds  
hive> █
```

## Create a new database:

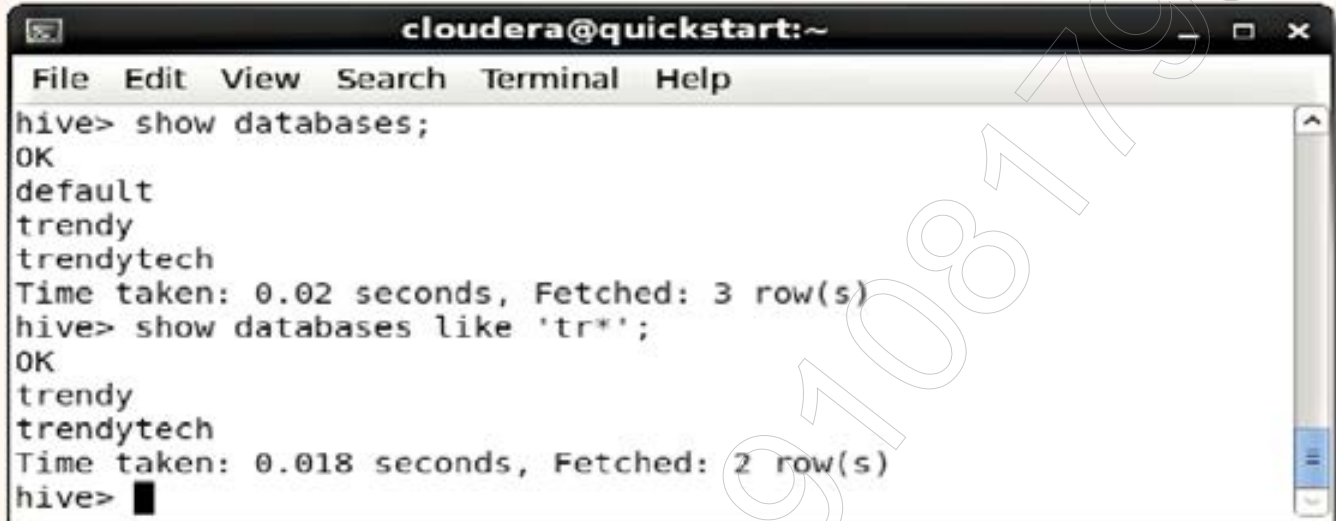
`create database trendy;`

A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the execution of 'hive> create database trendy;' which returns 'OK' and 'Time taken: 0.858 seconds'. The prompt 'hive>' is followed by a cursor.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> create database trendy;  
OK  
Time taken: 0.858 seconds  
hive> █
```

## List databases using “like” parameter:

```
show databases like 'tr*';
```

A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the execution of two Hive commands. The first command is 'show databases;', which returns 'OK' and lists 'default', 'trendy', and 'trendytech'. The second command is 'show databases like 'tr\*';', which also returns 'OK' and lists 'trendy' and 'trendytech'. Both commands show execution time and the number of rows fetched.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> show databases;  
OK  
default  
trendy  
trendytech  
Time taken: 0.02 seconds, Fetched: 3 row(s)  
hive> show databases like 'tr*';  
OK  
trendy  
trendytech  
Time taken: 0.018 seconds, Fetched: 2 row(s)  
hive> █
```

## Use a database:

```
use trendytech;
```

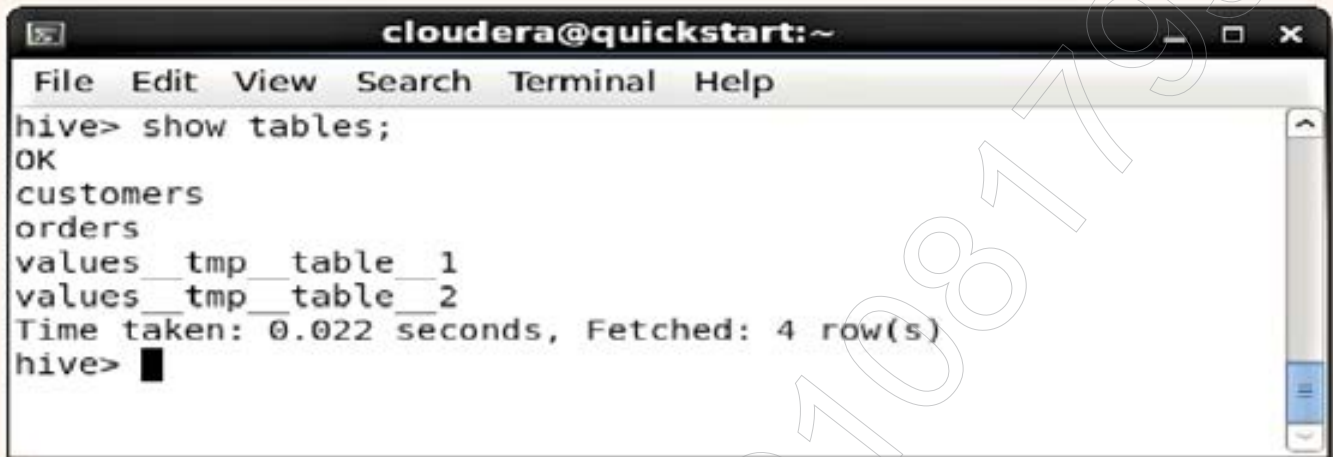
A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the execution of the 'use trendytech;' command, which returns 'OK' and shows an execution time of 0.04 seconds.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> use trendytech;  
OK  
Time taken: 0.04 seconds  
hive> █
```



## Show all tables present in a database:

show tables;



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> show tables;  
OK  
customers  
orders  
values__tmp__table__1  
values__tmp__table__2  
Time taken: 0.022 seconds, Fetched: 4 row(s)  
hive> █
```

# Hive Metastore



# Hive Metastore

The metadata for Hive tables are stored in the Hive Metastore

By default, the Hive Metastore stores all Hive metadata in Derby database

Derby is not recommended in production environment as it only allows one connection at a time

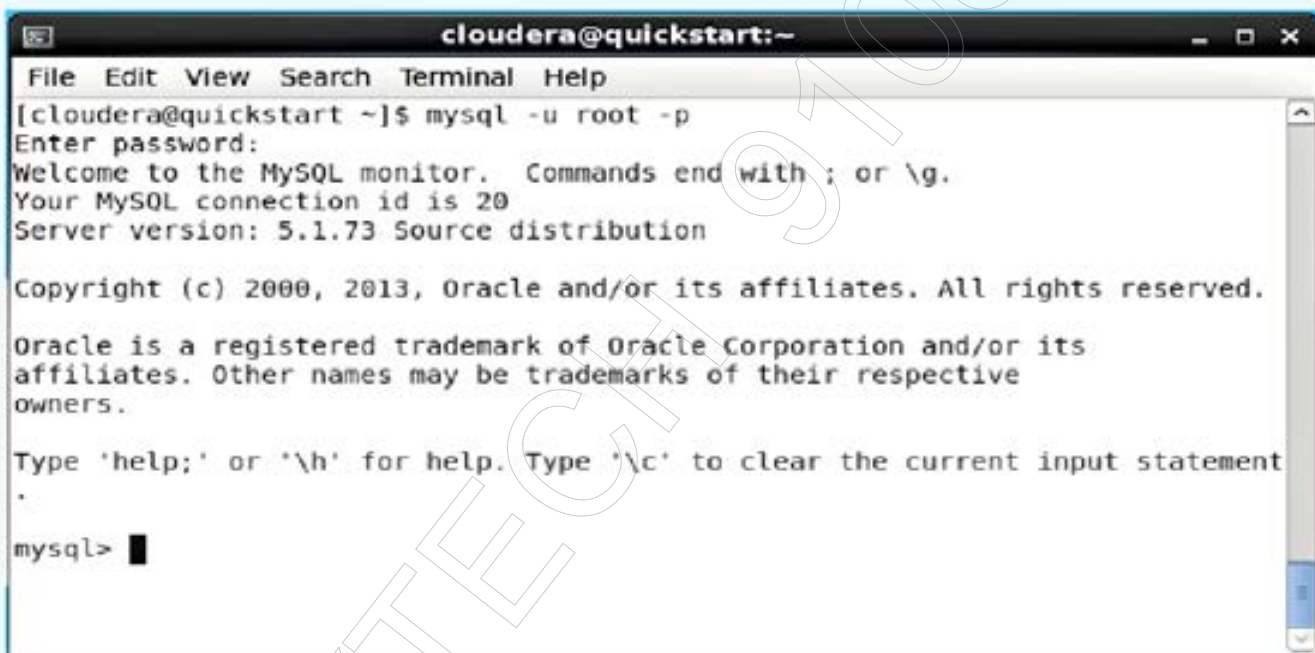
MYSQL is the best choice for the metastore because it is the most popular among the Hive user community



# Displaying Hive tables metadata information in MySQL database

## Enter into mysql:

`mysql -u root -p` (Use Password: cloudera)



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ mysql -u root -p  
Enter password:  
Welcome to the MySQL monitor.  Commands end with ; or \g.  
Your MySQL connection id is 20  
Server version: 5.1.73 Source distribution  
  
Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.  
  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement  
.  
mysql> █
```

## List all databases present in mysql:

show databases



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
mysql> show databases;  
+-----+  
| Database |  
+-----+  
| information_schema |  
| banking |  
| cm |  
| firehose |  
| hue |  
| metastore |  
| mysql |  
| nav |  
| navms |  
| oozie |  
| retail_db |  
| rman |  
| sentry |  
| trendytech |  
+-----+  
14 rows in set (0.17 sec)  
mysql> █
```

## Go to metastore database:

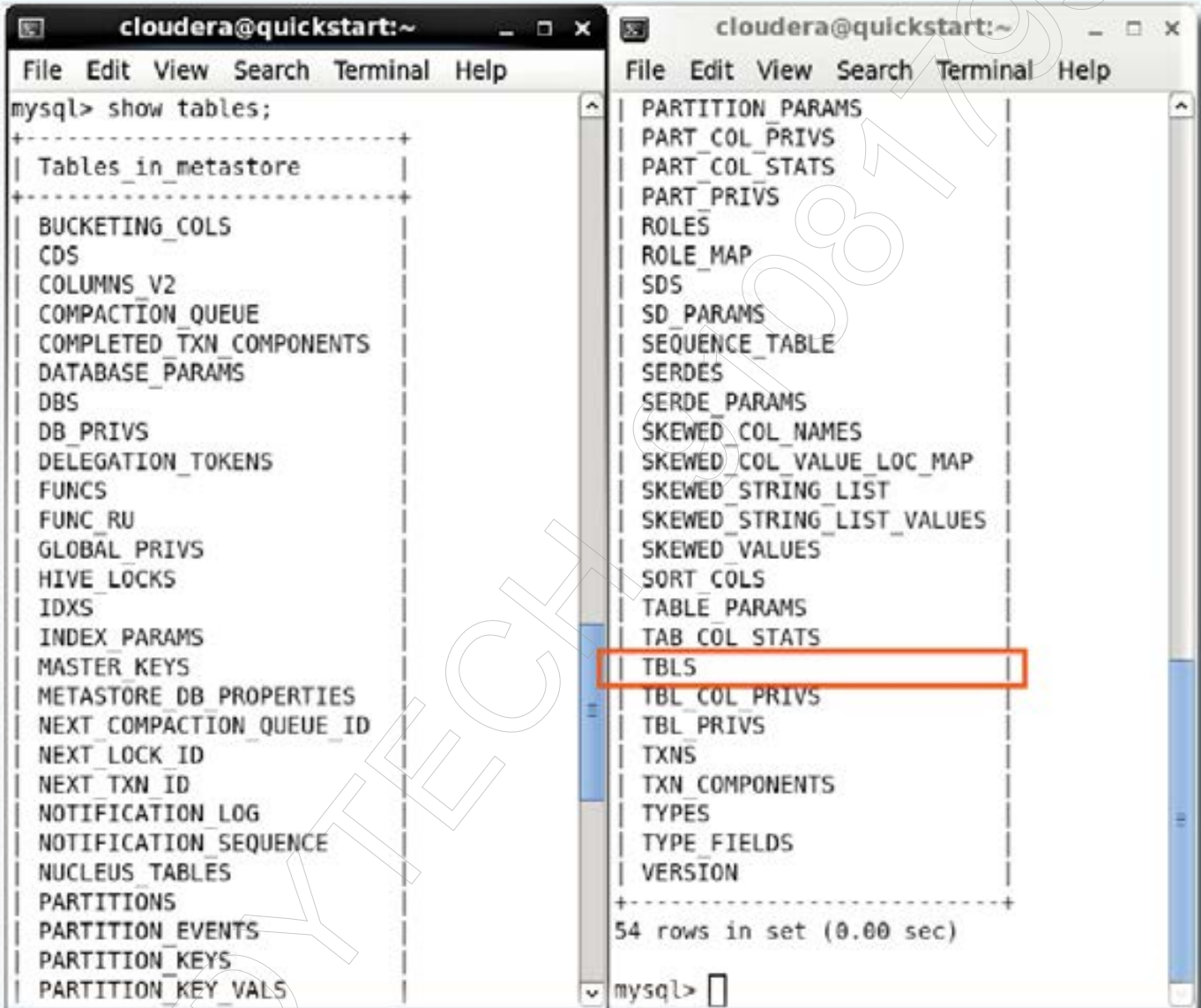
use metastore;



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
mysql> use metastore;  
Database changed  
mysql> █
```

# Display tables in metastore database:

show tables;

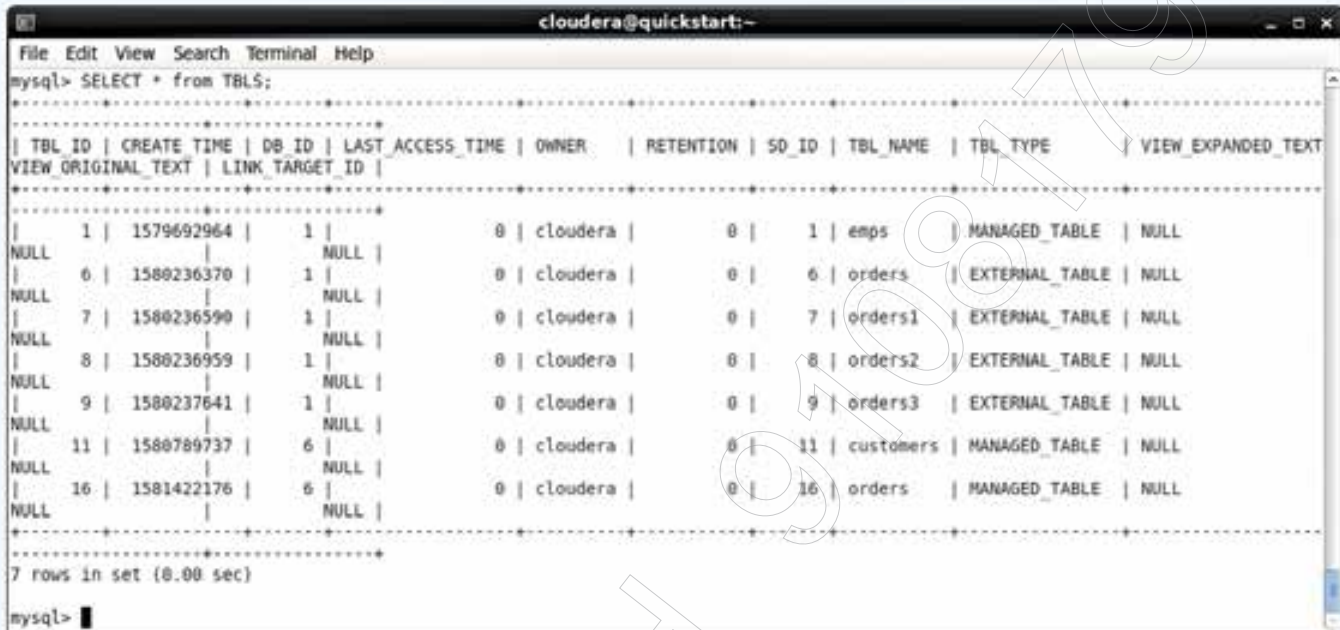


```
cloudera@quickstart:~  
mysql> show tables;  
+-----+  
| Tables_in_metastore |  
+-----+  
| BUCKETING_COLS      |  
| CDS                  |  
| COLUMNS_V2         |  
| COMPACTION_QUEUE    |  
| COMPLETED_TXN_COMPONENTS |  
| DATABASE_PARAMS     |  
| DBS                  |  
| DB_PRIVS             |  
| DELEGATION_TOKENS    |  
| FUNCS                |  
| FUNC_RU              |  
| GLOBAL_PRIVS         |  
| HIVE_LOCKS           |  
| IDXS                  |  
| INDEX_PARAMS         |  
| MASTER_KEYS          |  
| METASTORE_DB_PROPERTIES |  
| NEXT_COMPACTION_QUEUE_ID |  
| NEXT_LOCK_ID         |  
| NEXT_TXN_ID          |  
| NOTIFICATION_LOG     |  
| NOTIFICATION_SEQUENCE |  
| NUCLEUS_TABLES       |  
| PARTITIONS           |  
| PARTITION_EVENTS     |  
| PARTITION_KEYS       |  
| PARTITION_KEY_VALS   |  
+-----+  
cloudera@quickstart:~  
mysql> show tables;  
+-----+  
| PARTITION_PARAMS    |  
| PART_COL_PRIVS      |  
| PART_COL_STATS      |  
| PART_PRIVS          |  
| ROLES                |  
| ROLE_MAP            |  
| SDS                  |  
| SD_PARAMS           |  
| SEQUENCE_TABLE      |  
| SERDES              |  
| SERDE_PARAMS        |  
| SKEWED_COL_NAMES    |  
| SKEWED_COL_VALUE_LOC_MAP |  
| SKEWED_STRING_LIST  |  
| SKEWED_STRING_LIST_VALUES |  
| SKEWED_VALUES       |  
| SORT_COLS           |  
| TABLE_PARAMS       |  
| TAB_COL_STATS       |  
| TBLS                 |  
| TBL_COL_PRIVS       |  
| TBL_PRIVS           |  
| TXNS                 |  
| TXN_COMPONENTS      |  
| TYPES               |  
| TYPE_FIELDS         |  
| VERSION              |  
+-----+  
54 rows in set (0.00 sec)  
mysql>
```



## Display records of *TBLS* table:

Select \* from TBLS;



The screenshot shows a terminal window titled 'cloudera@quickstart:~'. The MySQL prompt 'mysql>' is followed by the command 'SELECT \* from TBLS;'. The output is a table with 11 columns: TBL\_ID, CREATE\_TIME, DB\_ID, LAST\_ACCESS\_TIME, OWNER, RETENTION, SD\_ID, TBL\_NAME, TBL\_TYPE, and VIEW\_EXPANDED\_TEXT. There are 7 rows of data. The first row shows a 'MANAGED\_TABLE' named 'emps'. The next three rows show 'EXTERNAL\_TABLE's named 'orders', 'orders1', and 'orders2'. The next row shows an 'EXTERNAL\_TABLE' named 'orders3'. The last two rows show 'MANAGED\_TABLE's named 'customers' and 'orders'.

TBL_ID	CREATE_TIME	DB_ID	LAST_ACCESS_TIME	OWNER	RETENTION	SD_ID	TBL_NAME	TBL_TYPE	VIEW_EXPANDED_TEXT
1	1579692964	1		cloudera	0	1	emps	MANAGED_TABLE	NULL
6	1580236370	1		cloudera	0	6	orders	EXTERNAL_TABLE	NULL
7	1580236590	1		cloudera	0	7	orders1	EXTERNAL_TABLE	NULL
8	1580236959	1		cloudera	0	8	orders2	EXTERNAL_TABLE	NULL
9	1580237641	1		cloudera	0	9	orders3	EXTERNAL_TABLE	NULL
11	1580789737	6		cloudera	0	11	customers	MANAGED_TABLE	NULL
16	1581422176	6		cloudera	0	16	orders	MANAGED_TABLE	NULL

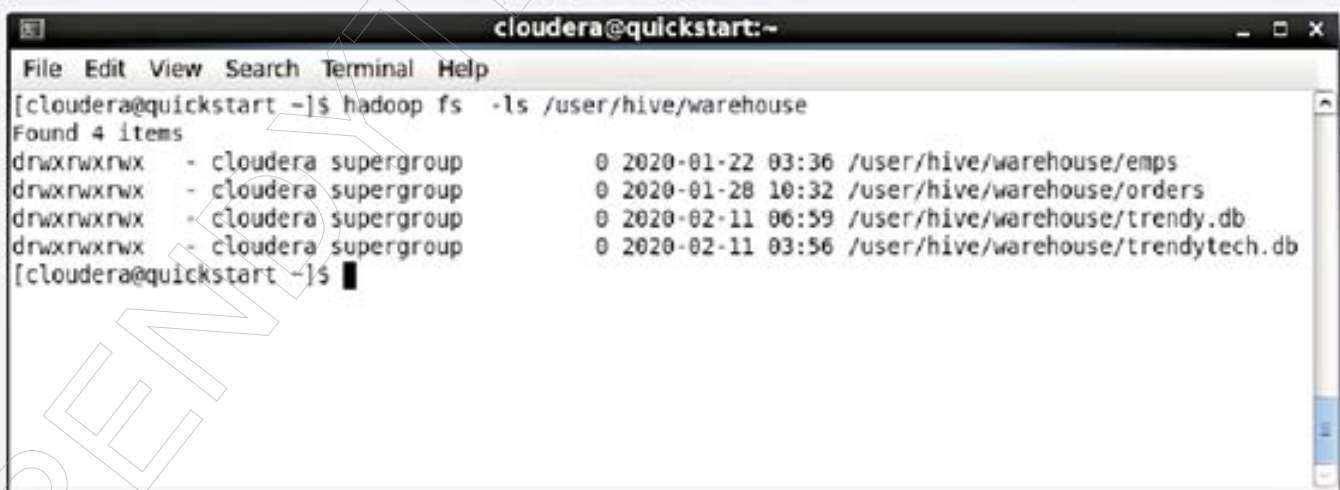
7 rows in set (0.00 sec)

# Hive warehouse directory

In Hive, actual data is stored in directories under Hive's warehouse directory (usually in /user/hive/warehouse)

**Display hive warehouse file structure:**

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

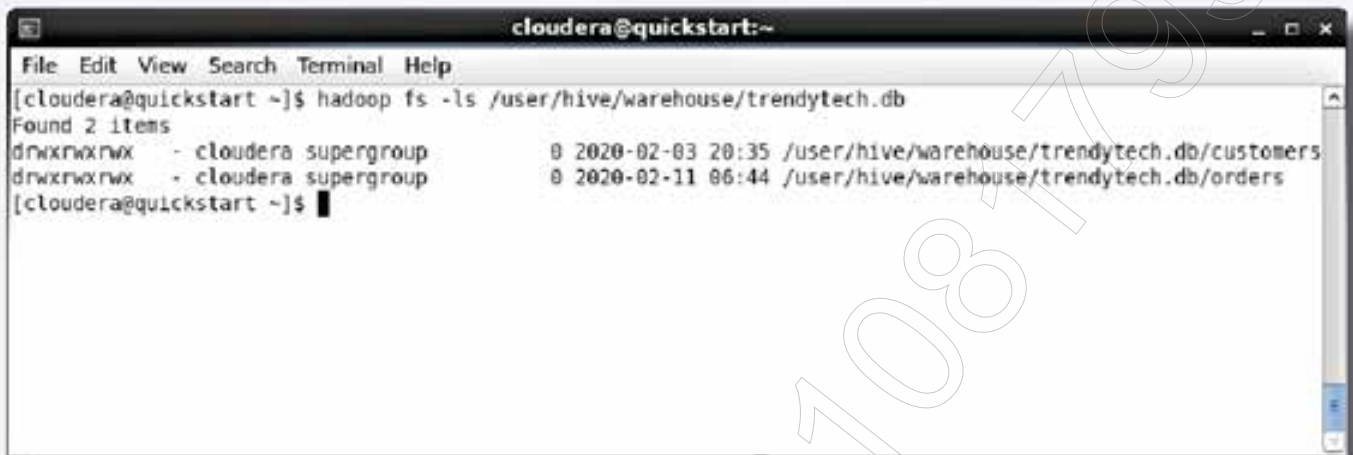


```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse  
Found 4 items  
drwxrwxrwx - cloudera supergroup 0 2020-01-22 03:36 /user/hive/warehouse/emp  
drwxrwxrwx - cloudera supergroup 0 2020-01-28 10:32 /user/hive/warehouse/orders  
drwxrwxrwx - cloudera supergroup 0 2020-02-11 06:59 /user/hive/warehouse/trendy.db  
drwxrwxrwx - cloudera supergroup 0 2020-02-11 03:56 /user/hive/warehouse/trendytech.db  
[cloudera@quickstart ~]$
```



## Display contents of *trendytech.db*:

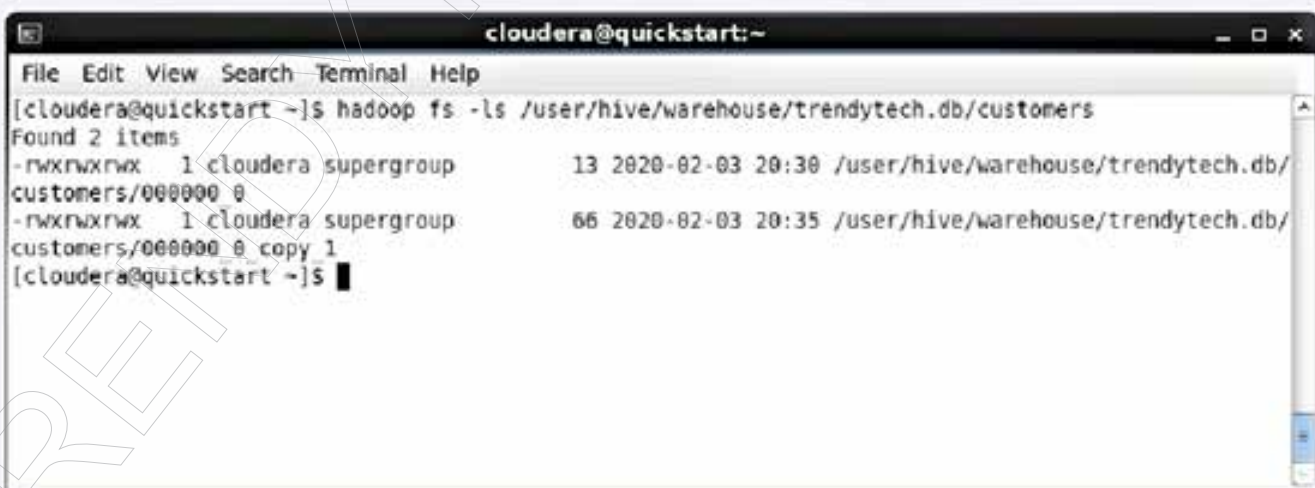
```
hadoop fs -ls /user/hive/warehouse/trendytech.db
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/trendytech.db  
Found 2 items  
drwxrwxrwx - cloudera supergroup      0 2020-02-03 20:35 /user/hive/warehouse/trendytech.db/customers  
drwxrwxrwx - cloudera supergroup      0 2020-02-11 06:44 /user/hive/warehouse/trendytech.db/orders  
[cloudera@quickstart ~]$
```

## Display contents of *customer* folder:

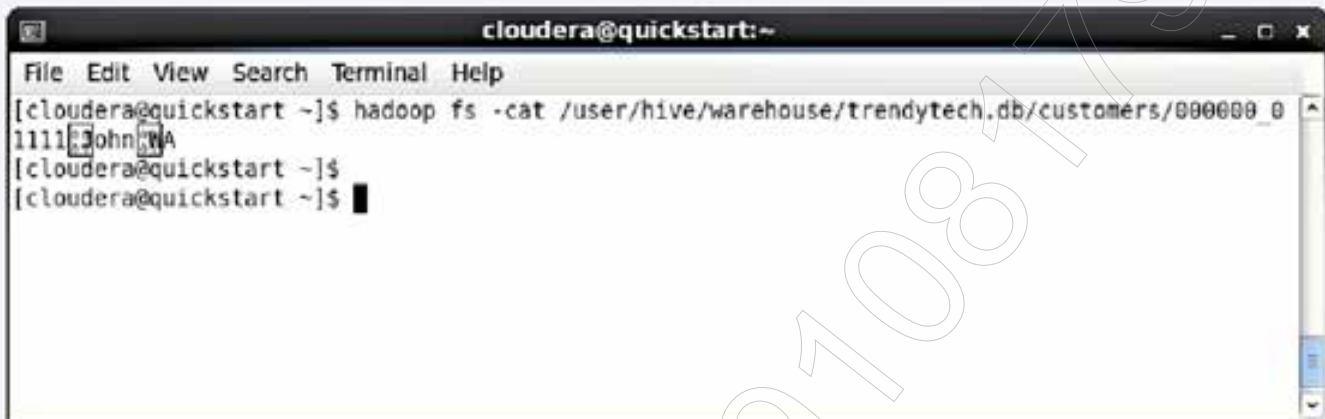
```
hadoop fs -ls /user/hive/warehouse/  
trendytech.db/customers
```



```
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 ~]$
```

## Display content of “000000\_0” file:

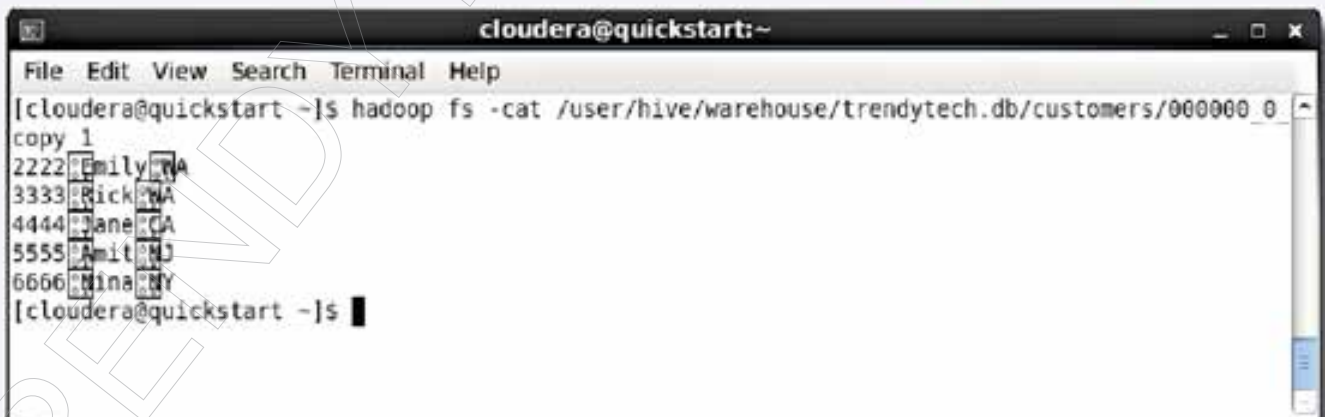
```
hadoop fs -cat /user/hive/warehouse/  
trendytech.db/customers/000000_0
```



A terminal window titled "cloudera@quickstart:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The command `hadoop fs -cat /user/hive/warehouse/trendytech.db/customers/000000_0` has been executed, resulting in the output: `1111 John WA`. The prompt `[cloudera@quickstart ~]$` is shown twice.

## Display content of “000000\_0\_copy\_1” file:

```
hadoop fs -cat /user/hive/warehouse/  
trendytech.db/customers/000000_0_copy_1
```



A terminal window titled "cloudera@quickstart:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The command `hadoop fs -cat /user/hive/warehouse/trendytech.db/customers/000000_0_copy_1` has been executed, resulting in the output: `copy 1`, `2222 Emily WA`, `3333 Rick WA`, `4444 Jane CA`, `5555 Amit NJ`, and `6666 Nina NY`. The prompt `[cloudera@quickstart ~]$` is shown twice.

# Types of Tables in Hive



# Hive Tables

There are mainly two types of tables in Hive:

- Managed table
- External table

# Hive Managed Table





# Managed Table

In Managed table, data is managed by Hive:

- Hive owns the files and directories
- Deleting a managed table deletes both data and metadata

**Go to default database of Hive:**

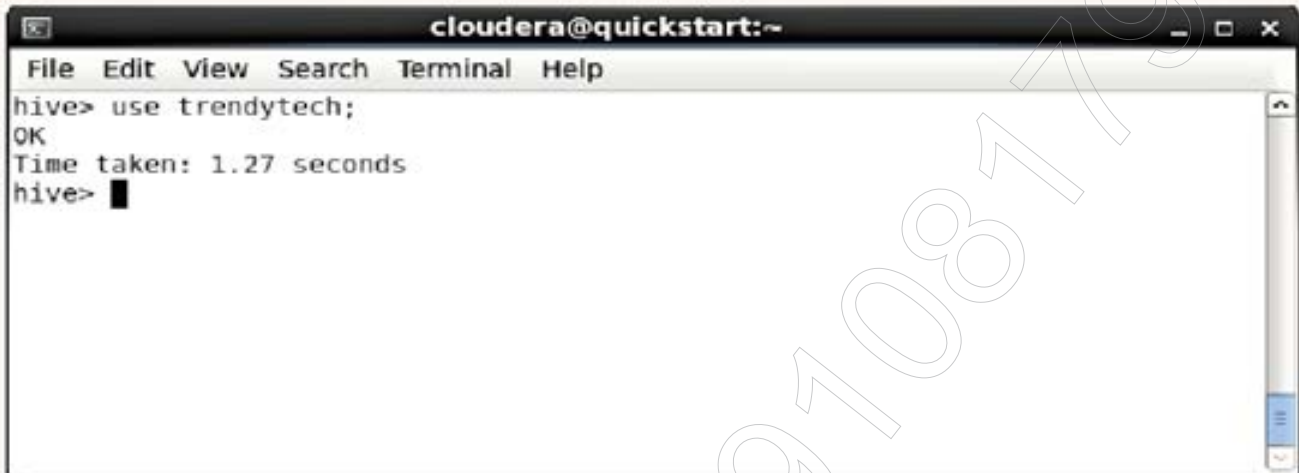
```
use default;
```

A terminal window titled 'cloudera@quickstart:~' showing the execution of the Hive command 'use default;'. The output shows 'OK' and 'Time taken: 1.785 seconds'. The prompt 'hive>' is followed by a cursor.

```
cloudera@quickstart:~  
hive> use default;  
OK  
Time taken: 1.785 seconds  
hive> █
```

## Go back to trendytech database:

```
use trendytech;
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> use trendytech;  
OK  
Time taken: 1.27 seconds  
hive> █
```

## Create a table in Hive (Managed):

```
create table test(id int, name string);
```



```
cloudera@quickstart:~  
hive> create table test(id int, name string);  
OK  
Time taken: 2.552 seconds  
hive> █
```

**Note:** When we create a table in Hive,  
by default it is Managed table

## Show tables in a database:

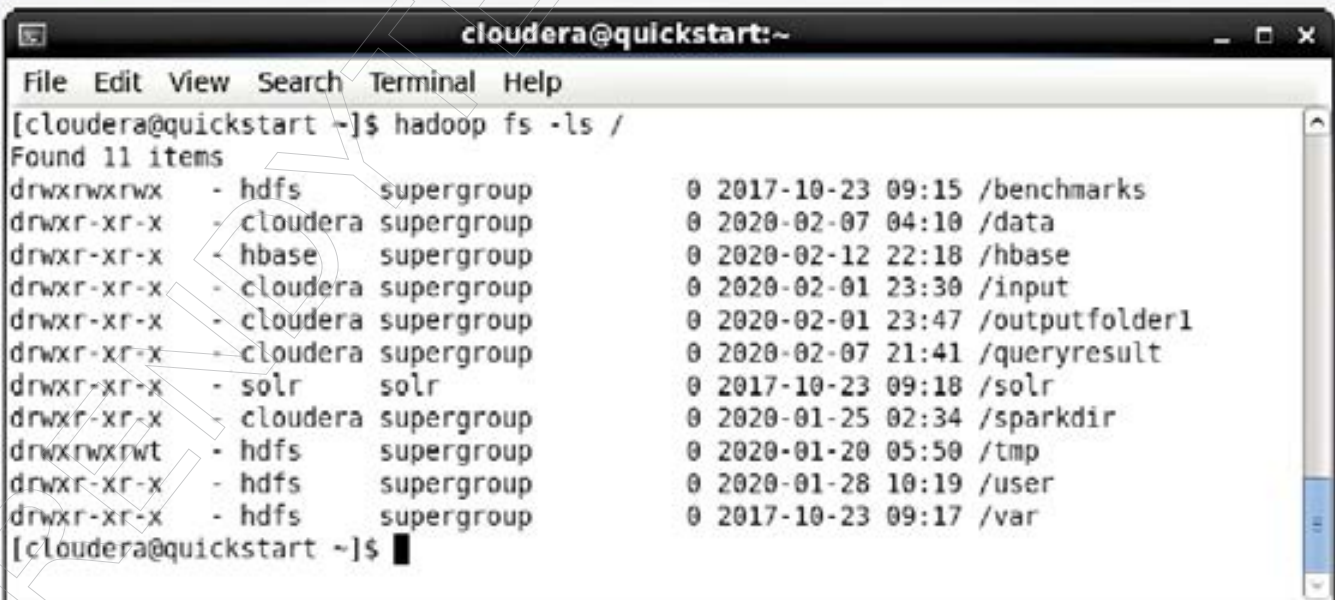
```
show tables;
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> show tables;  
OK  
customers  
orders  
test  
Time taken: 0.05 seconds, Fetched: 7 row(s)  
hive> █
```

## Display the root user in terminal:

```
hadoop fs -ls /
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /  
Found 11 items  
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks  
drwxr-xr-x - cloudera supergroup 0 2020-02-07 04:10 /data  
drwxr-xr-x - hbase supergroup 0 2020-02-12 22:18 /hbase  
drwxr-xr-x - cloudera supergroup 0 2020-02-01 23:30 /input  
drwxr-xr-x - cloudera supergroup 0 2020-02-01 23:47 /outputfolder1  
drwxr-xr-x - cloudera supergroup 0 2020-02-07 21:41 /queryresult  
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr  
drwxr-xr-x - cloudera supergroup 0 2020-01-25 02:34 /sparkdir  
drwxrwxrwt - hdfs supergroup 0 2020-01-20 05:50 /tmp  
drwxr-xr-x - hdfs supergroup 0 2020-01-28 10:19 /user  
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var  
[cloudera@quickstart ~]$ █
```



## Display Hive warehouse folder structure:

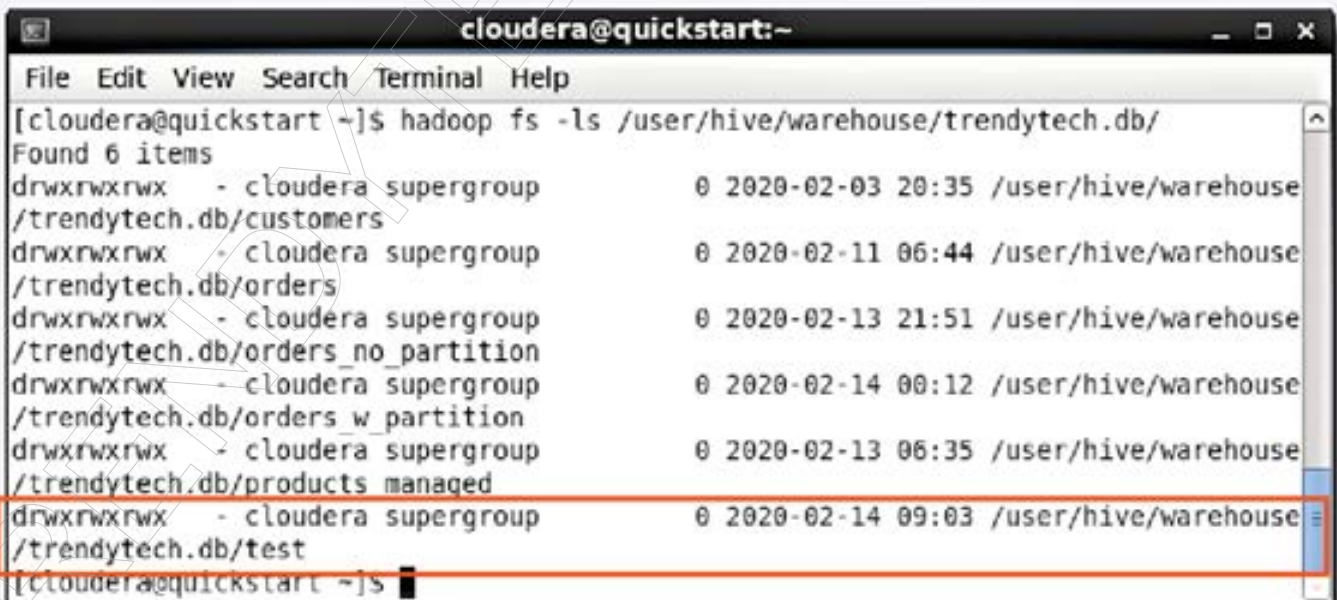
```
hadoop fs -ls /user/hive/warehouse/
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse  
Found 4 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 2020-02-11 06:59 /user/hive/warehouse/trendy.db  
drwxrwxrwx - cloudera supergroup 0 2020-02-11 03:56 /user/hive/warehouse/trendytech.db  
[cloudera@quickstart ~]$
```

## Display contents of the trendytech database:

```
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 6 items  
drwxrwxrwx - cloudera supergroup 0 2020-02-03 20:35 /user/hive/warehouse  
/trendytech.db/customers  
drwxrwxrwx - cloudera supergroup 0 2020-02-11 06:44 /user/hive/warehouse  
/trendytech.db/orders  
drwxrwxrwx - cloudera supergroup 0 2020-02-13 21:51 /user/hive/warehouse  
/trendytech.db/orders_no_partition  
drwxrwxrwx - cloudera supergroup 0 2020-02-14 00:12 /user/hive/warehouse  
/trendytech.db/orders_w_partition  
drwxrwxrwx - cloudera supergroup 0 2020-02-13 06:35 /user/hive/warehouse  
/trendytech.db/products managed  
drwxrwxrwx - cloudera supergroup 0 2020-02-14 09:03 /user/hive/warehouse  
/trendytech.db/test  
[cloudera@quickstart ~]$
```

# Loading data into Hive table

We can load data into hive table in three ways:

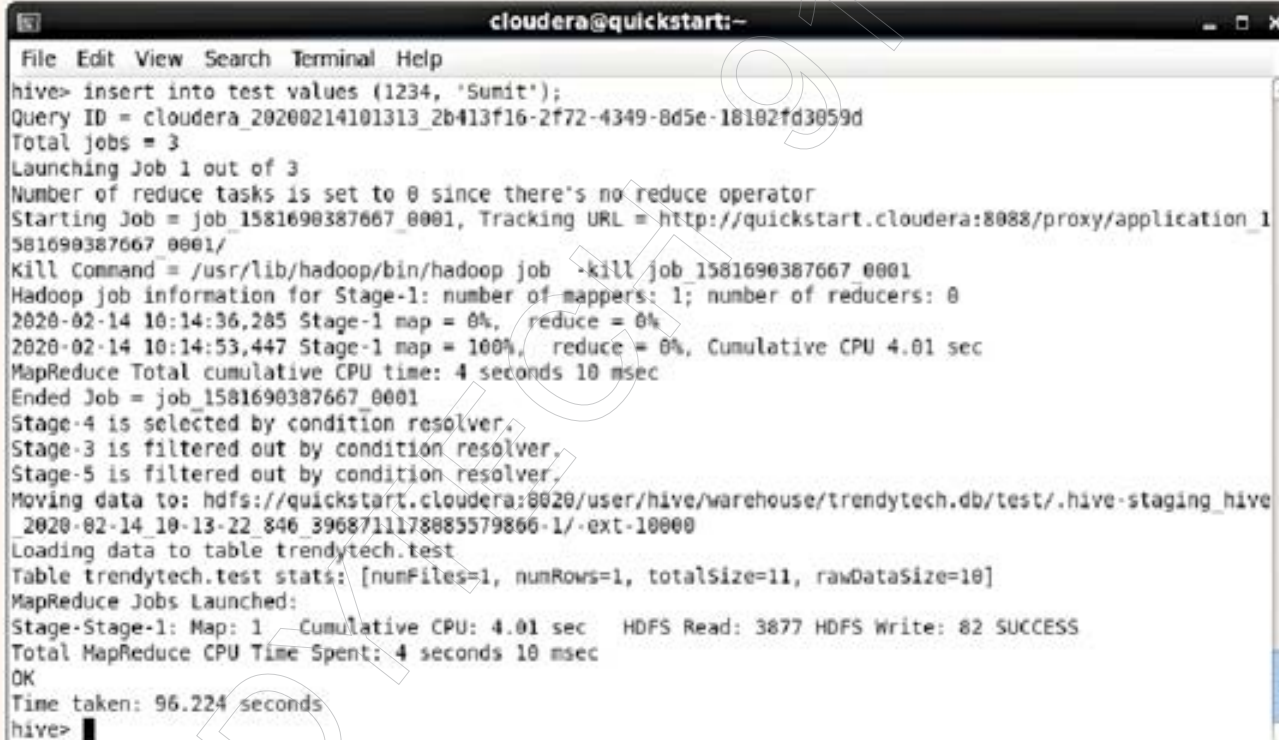
1. Using insert command
2. Loading from file
3. Using table to table loading



# 1. Data loading using insert command (Not Recommended)

Insert data into hive (Managed) table:

```
insert into test values (1234, 'Sumit');
```

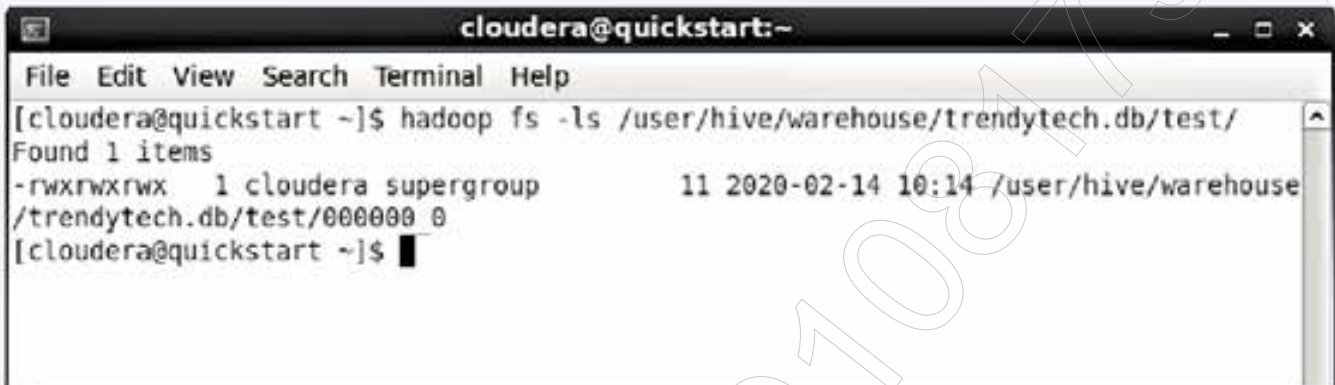


```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> insert into test values (1234, 'Sumit');  
Query ID = cloudera_20200214101313_2b413f16-2f72-4349-8d5e-18102fd3059d  
Total jobs = 3  
Launching Job 1 out of 3  
Number of reduce tasks is set to 0 since there's no reduce operator  
Starting Job = job_1581690387667_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1581690387667_0001/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1581690387667_0001  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0  
2020-02-14 10:14:36,285 Stage-1 map = 0%, reduce = 0%  
2020-02-14 10:14:53,447 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.01 sec  
MapReduce Total cumulative CPU time: 4 seconds 10 msec  
Ended Job = job_1581690387667_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/test/.hive-staging_hive_2020-02-14_10-13-22_846_3968711178085579866-1/-ext-10000  
Loading data to table trendytech.test  
Table trendytech.test stats: [numFiles=1, numRows=1, totalSize=11, rawDataSize=10]  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Cumulative CPU: 4.01 sec HDFS Read: 3877 HDFS Write: 82 SUCCESS  
Total MapReduce CPU Time Spent: 4 seconds 10 msec  
OK  
Time taken: 96.224 seconds  
hive>
```

**Note:** It will trigger MapReduce job.

## Display the contents of test table from terminal:

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



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/trendytech.db/test/  
Found 1 items  
-rwxrwxrwx 1 cloudera supergroup          11 2020-02-14 10:14 /user/hive/warehouse  
/trendytech.db/test/000000_0  
[cloudera@quickstart ~]$
```

## Display the recods of test table:

```
hadoop fs -cat /user/hive/warehouse/  
trendytech.db/test/000000_0
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/trendytech.db/test/000000_0  
000 0  
1234567890  
[cloudera@quickstart ~]$
```

## Drop the Managed table (test) from Hive:

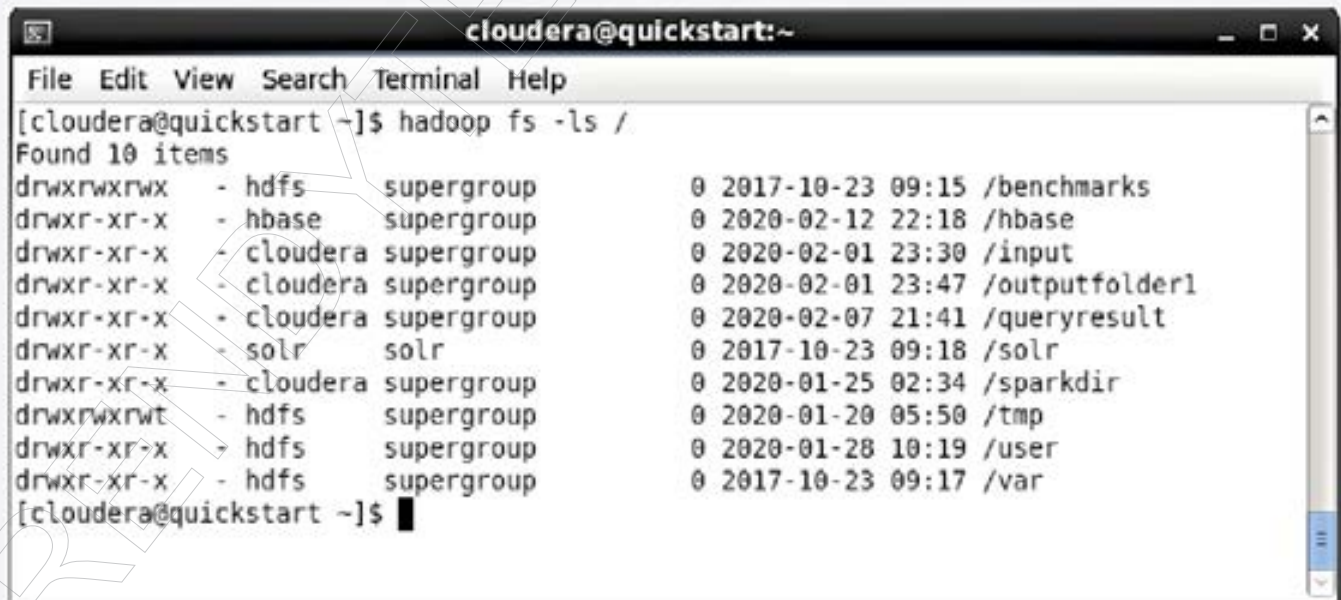
```
drop table test;
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> drop table test;  
OK  
Time taken: 2.215 seconds  
hive> █
```

## Display hdfs root directory folder structure:

```
hadoop fs -ls /
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /  
Found 10 items  
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks  
drwxr-xr-x - hbase supergroup 0 2020-02-12 22:18 /hbase  
drwxr-xr-x - cloudera supergroup 0 2020-02-01 23:30 /input  
drwxr-xr-x - cloudera supergroup 0 2020-02-01 23:47 /outputfolder1  
drwxr-xr-x - cloudera supergroup 0 2020-02-07 21:41 /queryresult  
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr  
drwxr-xr-x - cloudera supergroup 0 2020-01-25 02:34 /sparkdir  
drwxrwxrwt - hdfs supergroup 0 2020-01-20 05:50 /tmp  
drwxr-xr-x - hdfs supergroup 0 2020-01-28 10:19 /user  
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var  
[cloudera@quickstart ~]$ █
```



## 2. Data loading using Load command

You can load data into a hive table using Load statement in two ways:

- 2.1 From LFS (Local File System) to Hive table
- 2.2 From HDFS to Hive table



**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](mailto:trendytech.sumit@gmail.com)

**Youtube** TrendyTech

**Twitter** @BigdataBySumit

**Instagram** bigdatabysumit

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

