Apache Hive

Part 2

by Sumit Mittal

Apache Hive Exercise 2

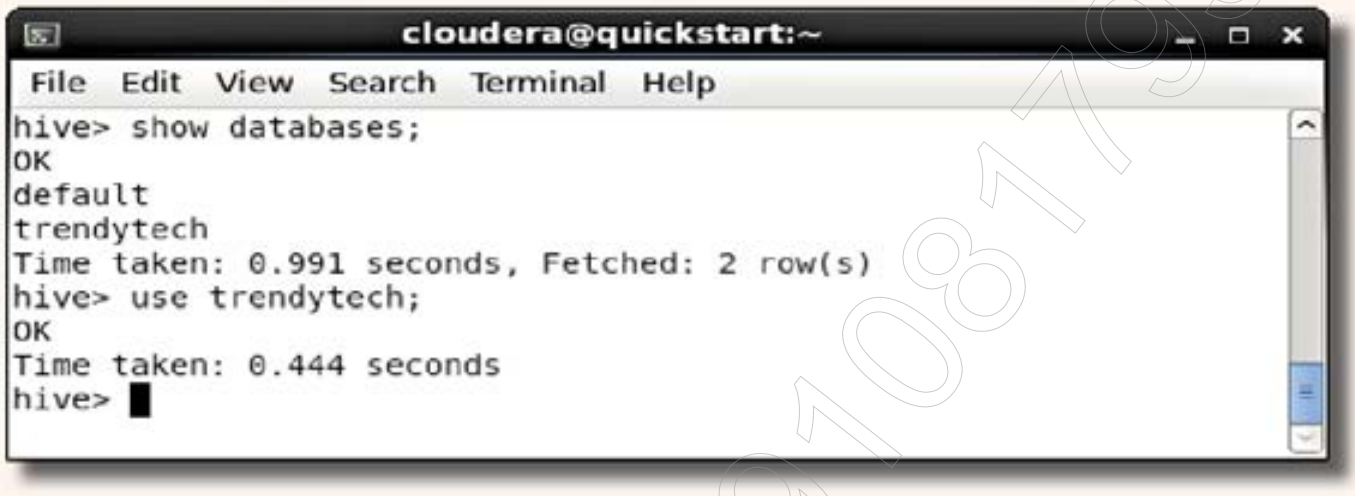
Basic Hive Queries

IMPORTANT

Copyright Infringement and Illegal Content Sharing Notice

To show all databases present in Hive:

```
show databases;
```

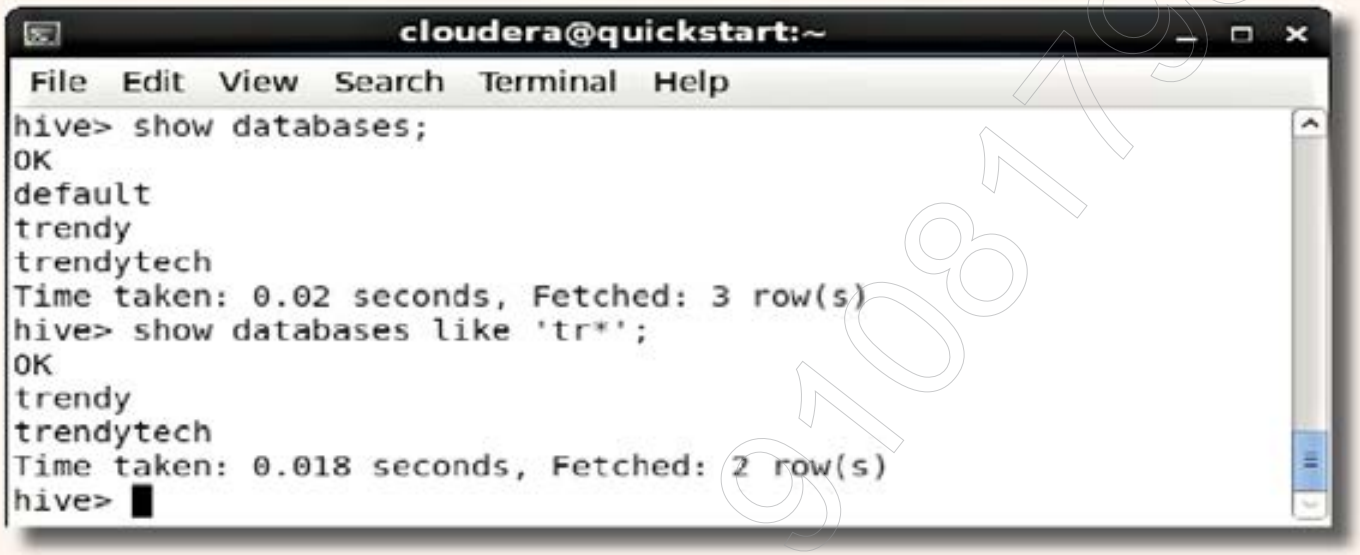


Create a new database:

```
create database trendy;
```

List databases using “like” parameter:

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



Use a database:

```
use trendytech;
```

Show all tables present in a database:

```
show tables;
```

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

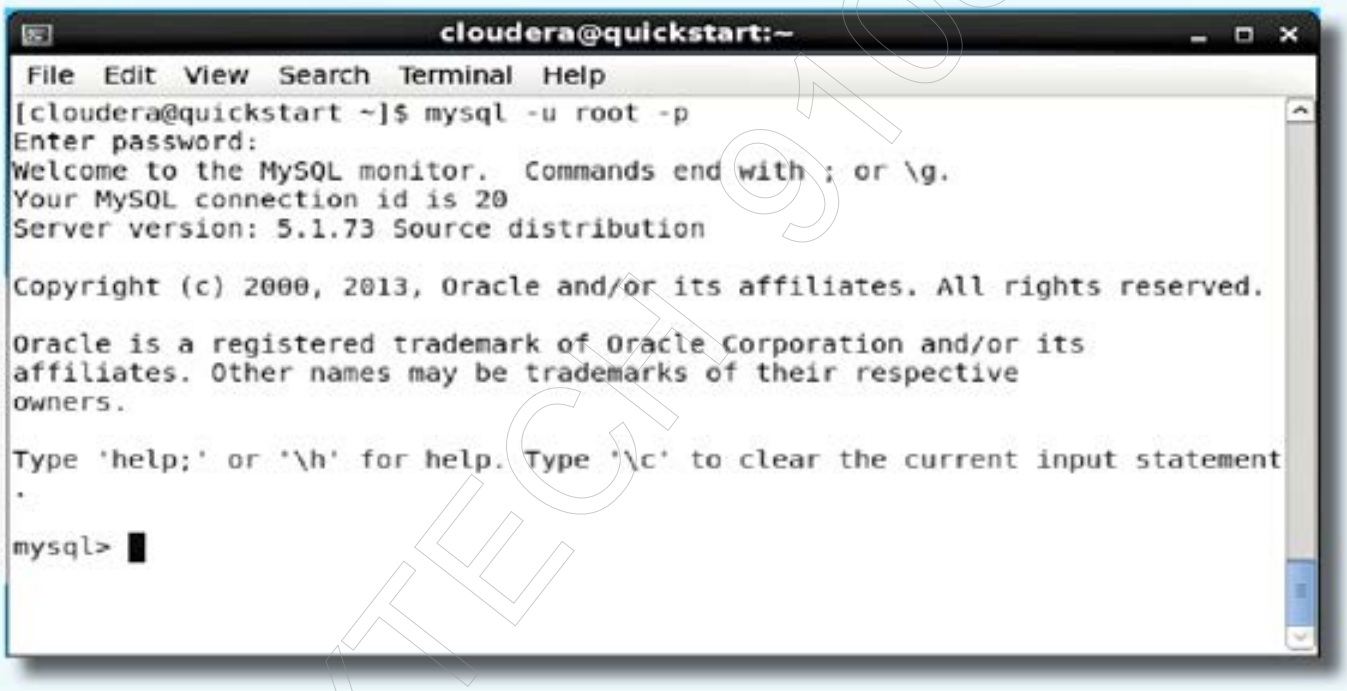
Derby is not recommended in production environment as it only allows one connection at a time

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)



List all databases present in mysql:

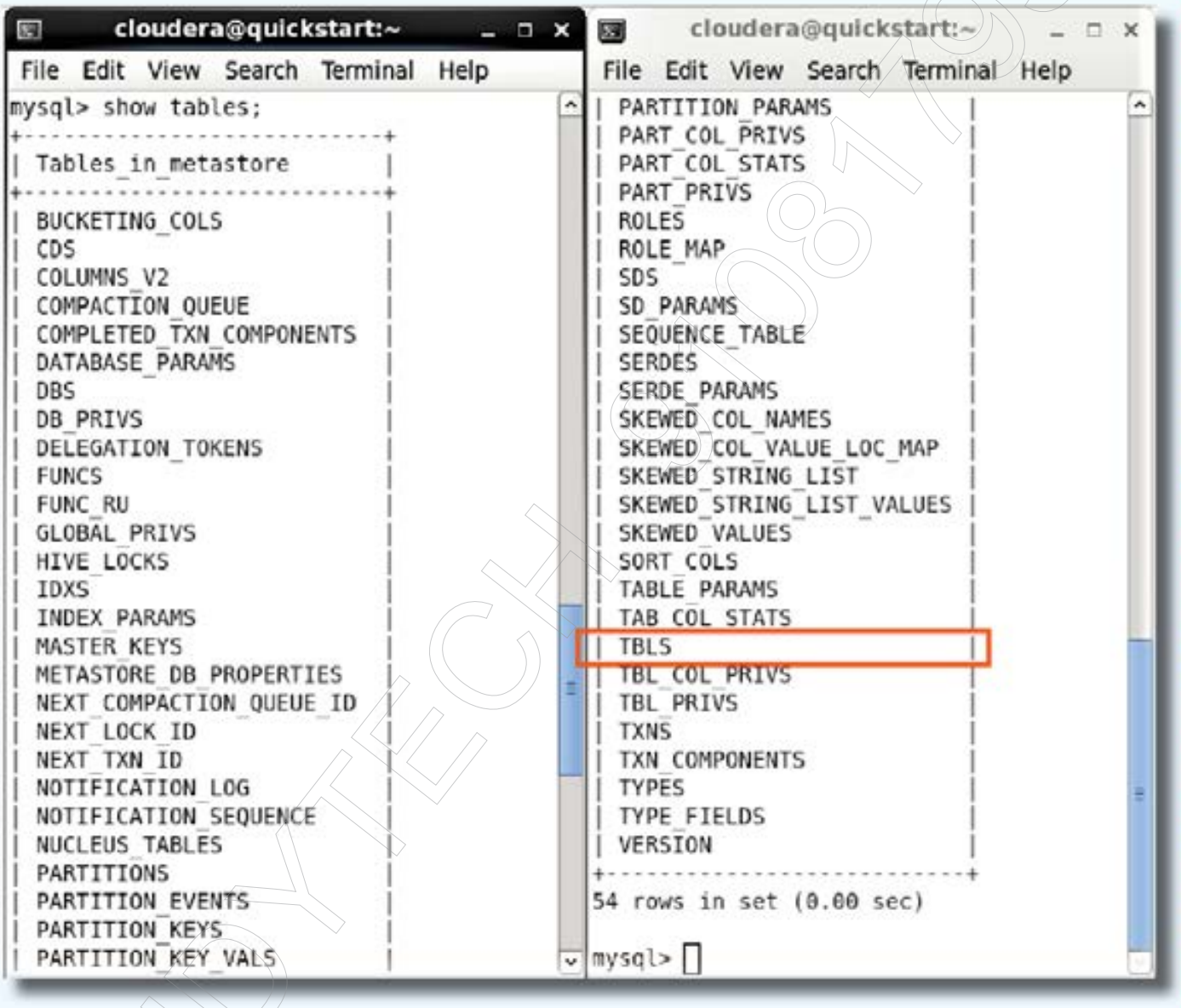
```
show databases
```

Go to metastore database:

```
use metastore;
```

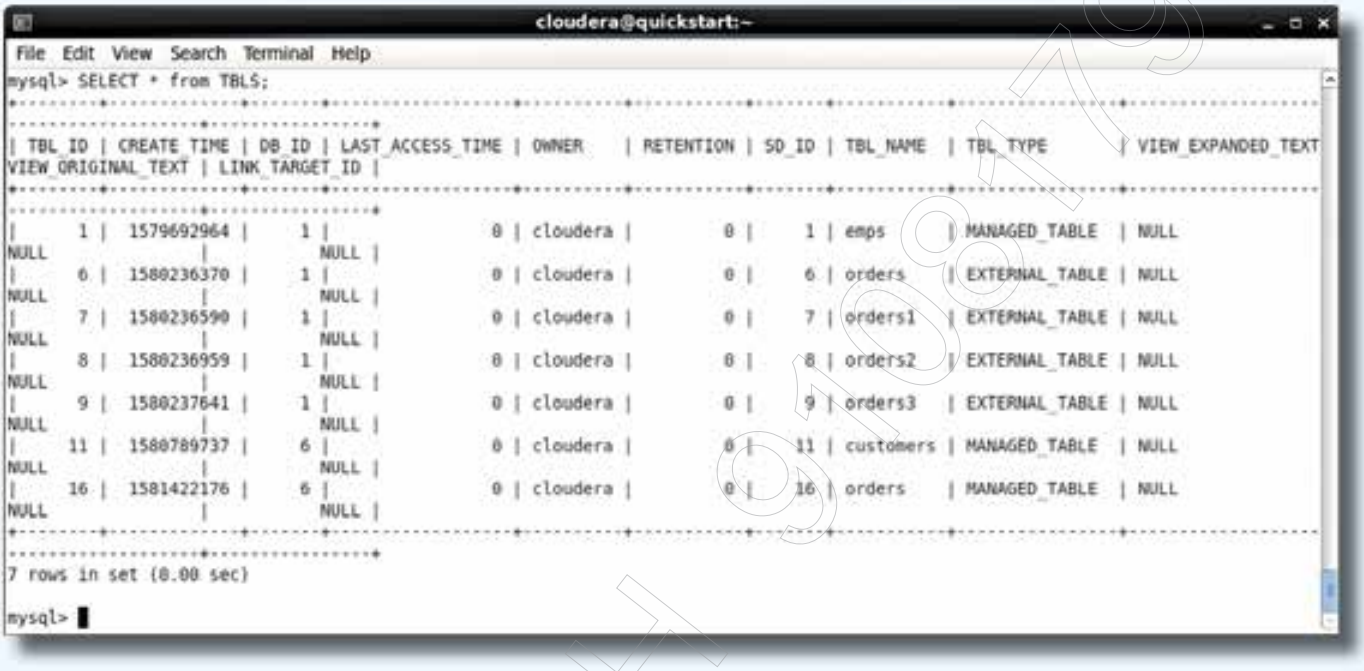
Display tables in metastore database:

```
show tables;
```



Display records of *TBLS* table:

```
Select * from TBLS;
```



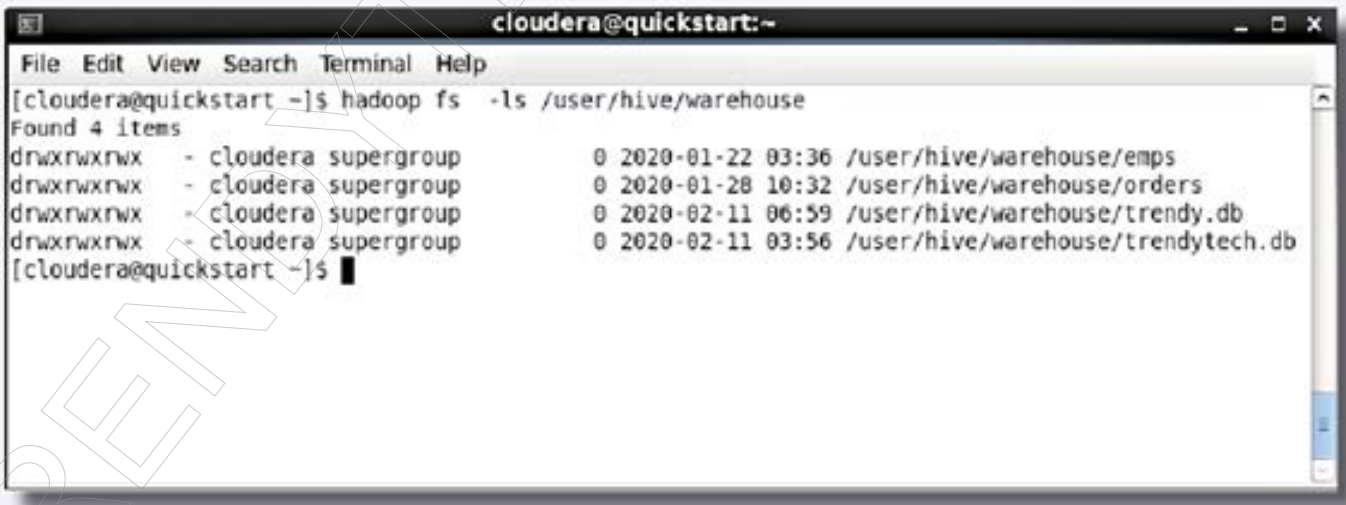
```
cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
mysql> SELECT * from TBLS;
+--------+-------------+-------+------------------+----------+-----------+-------+----------+---------------+--------------------+------------------+
| TBL_ID | CREATE_TIME | DB_ID | LAST_ACCESS_TIME | OWNER    | RETENTION | SD_ID | TBL_NAME | TBL_TYPE      | VIEW_EXPANDED_TEXT | VIEW_ORIGINAL_TEXT | LINK_TARGET_ID |
+--------+-------------+-------+------------------+----------+-----------+-------+----------+---------------+--------------------+--------------------+----------------+
|      1 |  1579692964 |     1 |                0 | cloudera |         0 |     1 | emps     | MANAGED_TABLE | NULL               | NULL               |           NULL |
|      6 |  1580236370 |     1 |                0 | cloudera |         0 |     6 | orders   | EXTERNAL_TABLE | NULL              | NULL               |           NULL |
|      7 |  1580236590 |     1 |                0 | cloudera |         0 |     7 | orders1  | EXTERNAL_TABLE | NULL              | NULL               |           NULL |
|      8 |  1580236959 |     1 |                0 | cloudera |         0 |     8 | orders2  | EXTERNAL_TABLE | NULL              | NULL               |           NULL |
|      9 |  1580237641 |     1 |                0 | cloudera |         0 |     9 | orders3  | EXTERNAL_TABLE | NULL              | NULL               |           NULL |
|     11 |  1580789737 |     6 |                0 | cloudera |         0 |    11 | customers | MANAGED_TABLE | NULL               | NULL               |           NULL |
|     16 |  1581422176 |     6 |                0 | cloudera |         0 |    16 | orders   | MANAGED_TABLE | NULL               | NULL               |           NULL |
+--------+-------------+-------+------------------+----------+-----------+-------+----------+---------------+--------------------+--------------------+----------------+
7 rows in set (0.00 sec)

mysql>
```

Hive warehouse directory

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

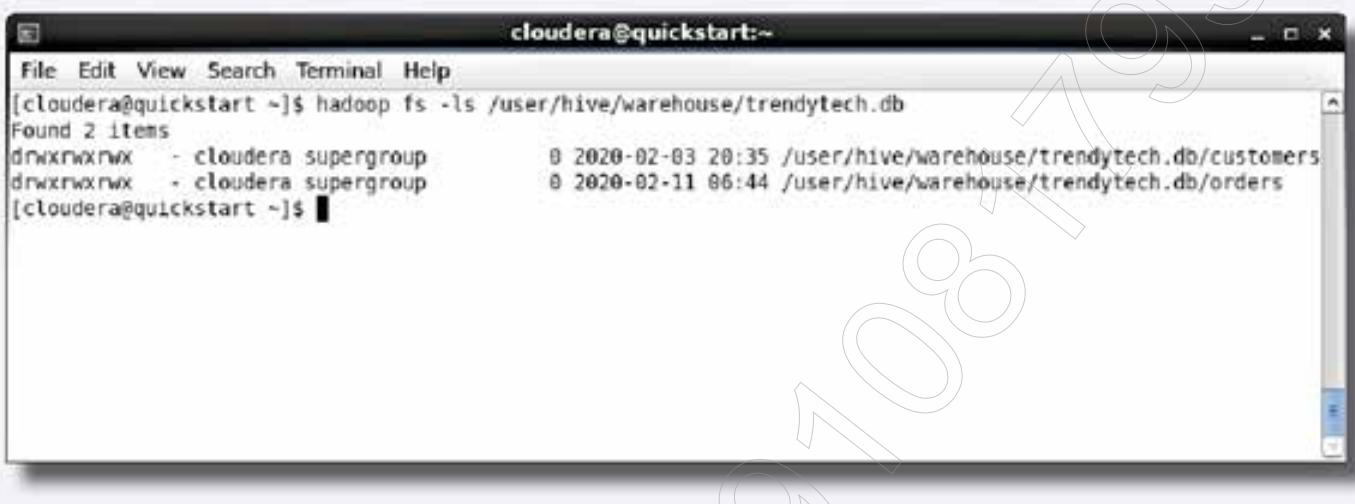
Display hive warehouse file structure:

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



Display contents of *trendytech.db*:

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



Display contents of *customer* folder:

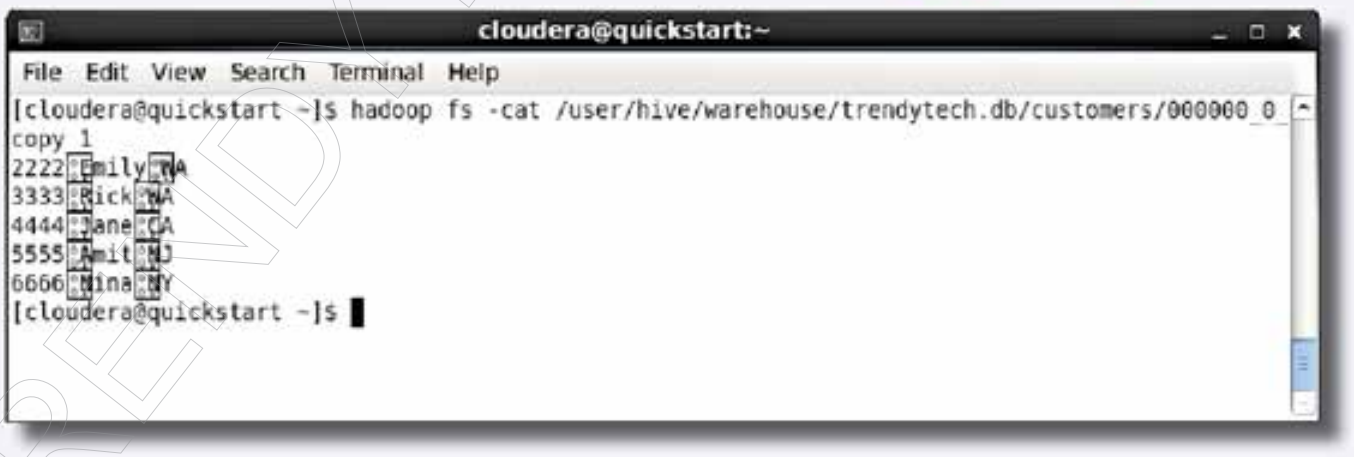
```
hadoop fs -ls /user/hive/warehouse/
trendytech.db/customers
```

Display content of “000000_0” file:

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

Display content of “000000_0_copy_1” file:

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



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;
```

Go back to trendytech database:

```
use trendytech;
```

Create a table in Hive (Managed):

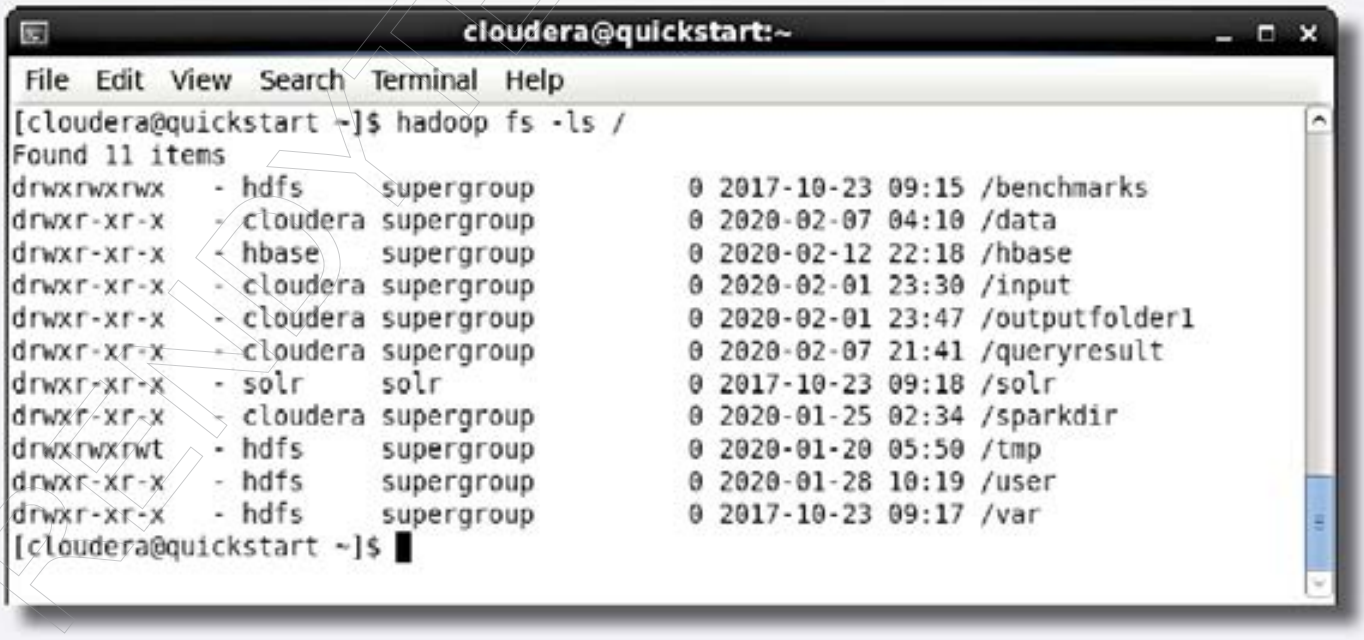
```
create table test(id int, name string);
```

Show tables in a database:

```
show tables;
```

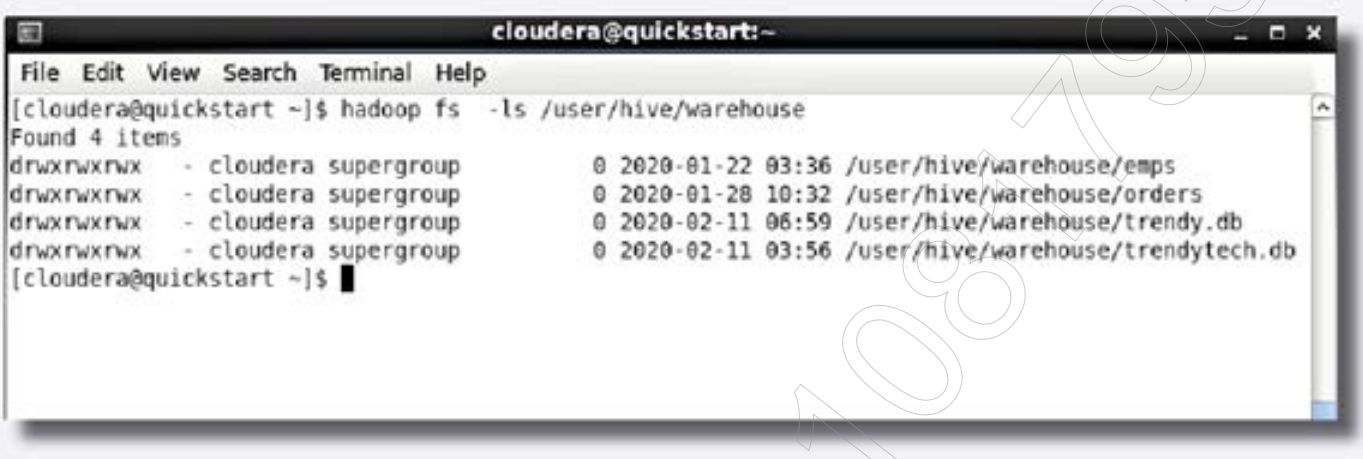
Display the root user in terminal:

```
hadoop fs -ls /
```



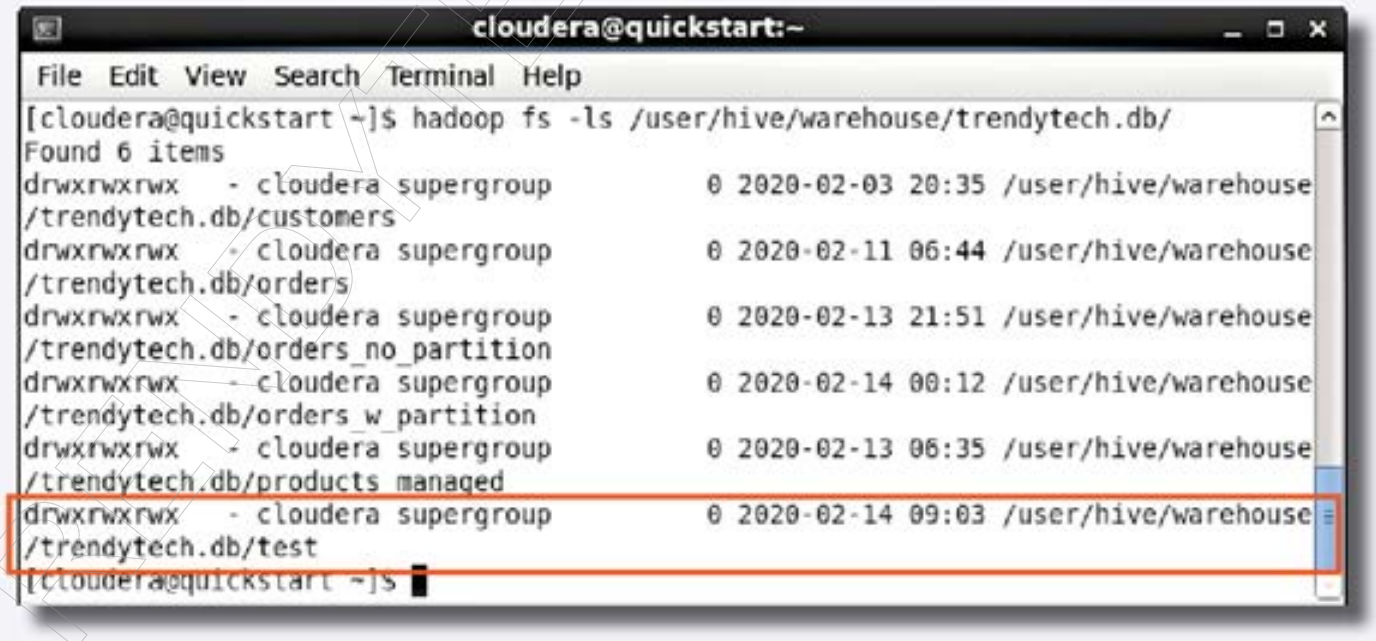
Display Hive warehouse folder structure:

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



Display contents of the trendytech database:

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



Loading data into Hive table

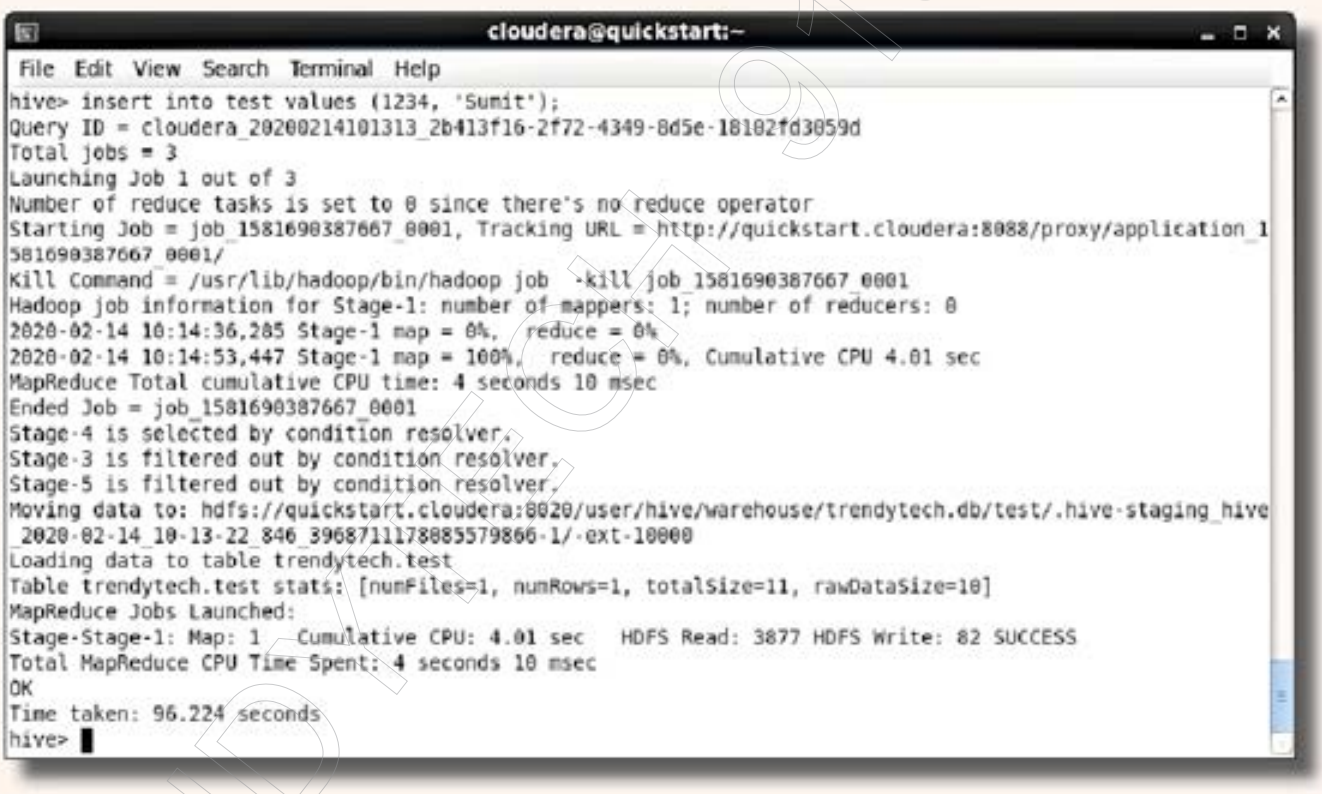
We can load data into hive table in three ways:

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:

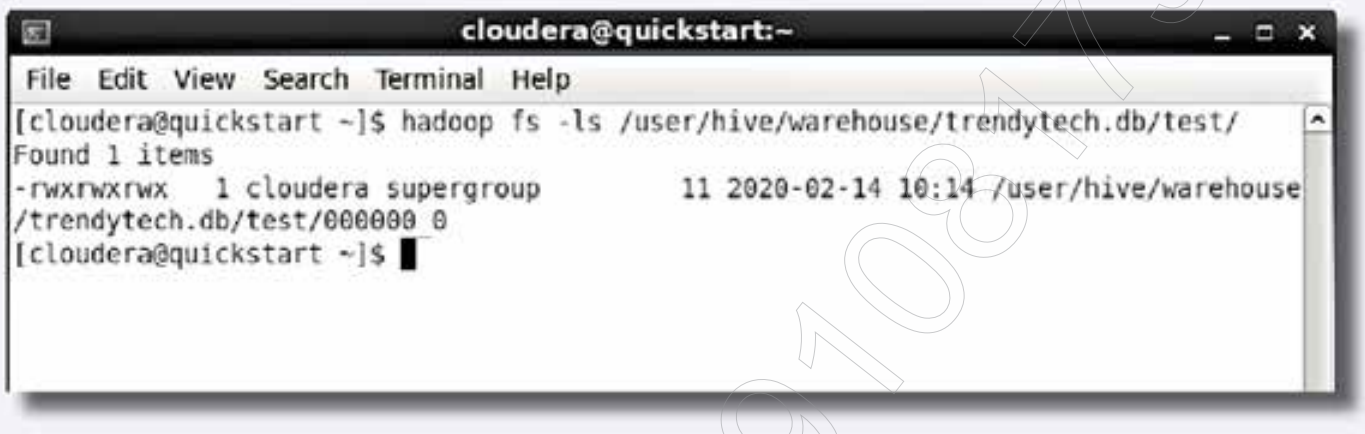
```
insert into test values (1234, 'Sumit');
```



Note: It will trigger MapReduce job.

Display the contents of test table from terminal:

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



Display the recods of test table:

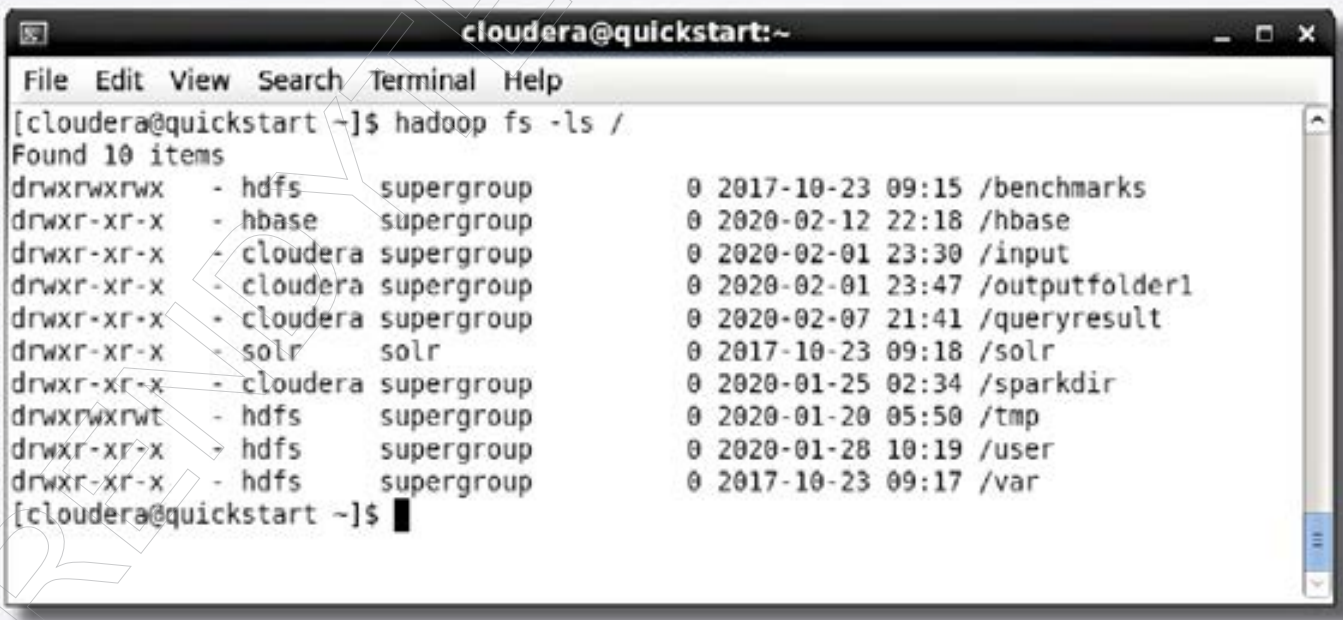
```
hadoop fs -cat /user/hive/warehouse/
trendytech.db/test/000000_0
```

Drop the Managed table (test) from Hive:

```
drop table test;
```

Display hdfs root directory folder structure:

```
hadoop fs -ls /
```



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

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/