Apache Hive

by Sumit Mittal

TRENDYTECH
UPLIFT YOUR CAREER!

IMPORTANT

Copyright Infringement and Illegal Content Sharing Notice

Hive
External Table

External Table

In External table, data not fully managed by Hive:

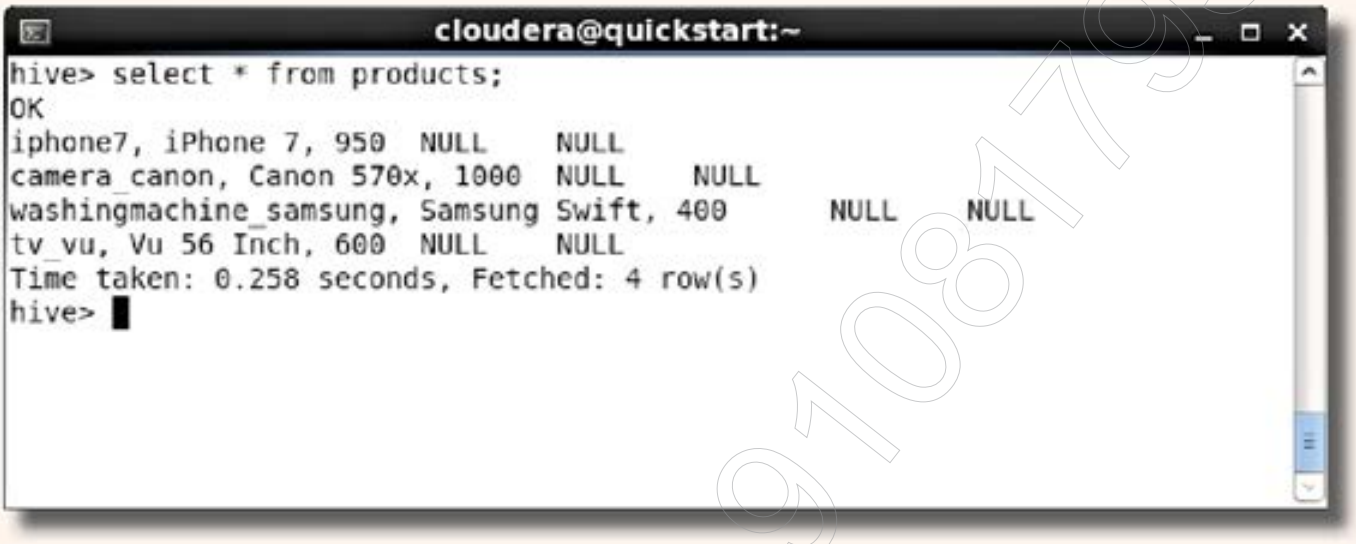
- Share the underlying data across other applications (Hadoop, Pig, HBase, etc)
- Deleting an external table deletes only the metadata

Create an external table in Hive:

```
create external table products(
id string,
title string,
cost float)
location '/data/';
```

Display records of an external table:

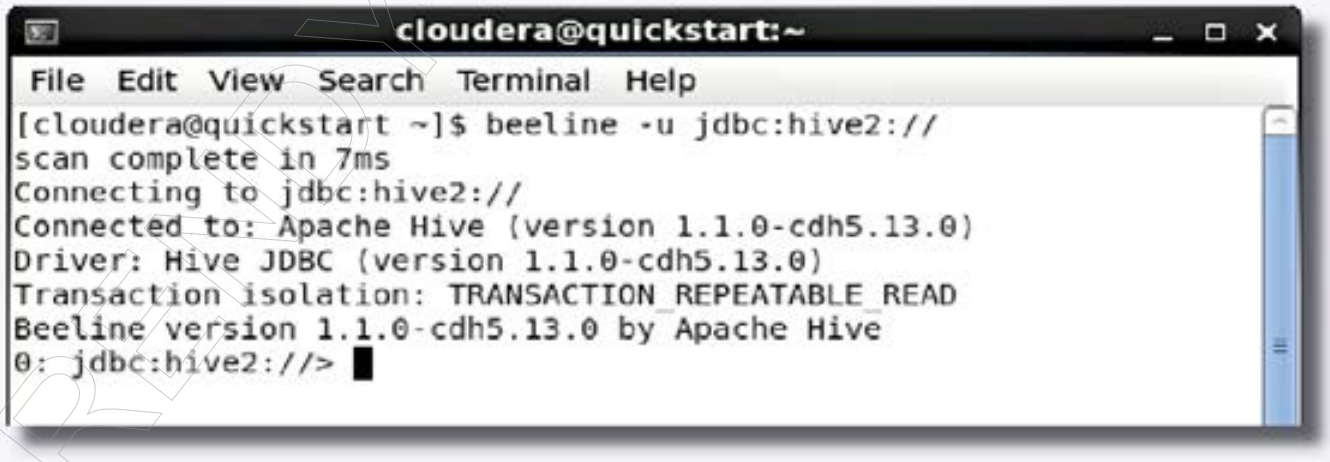
```
select * from products;
```



Note: The output is not what we wanted

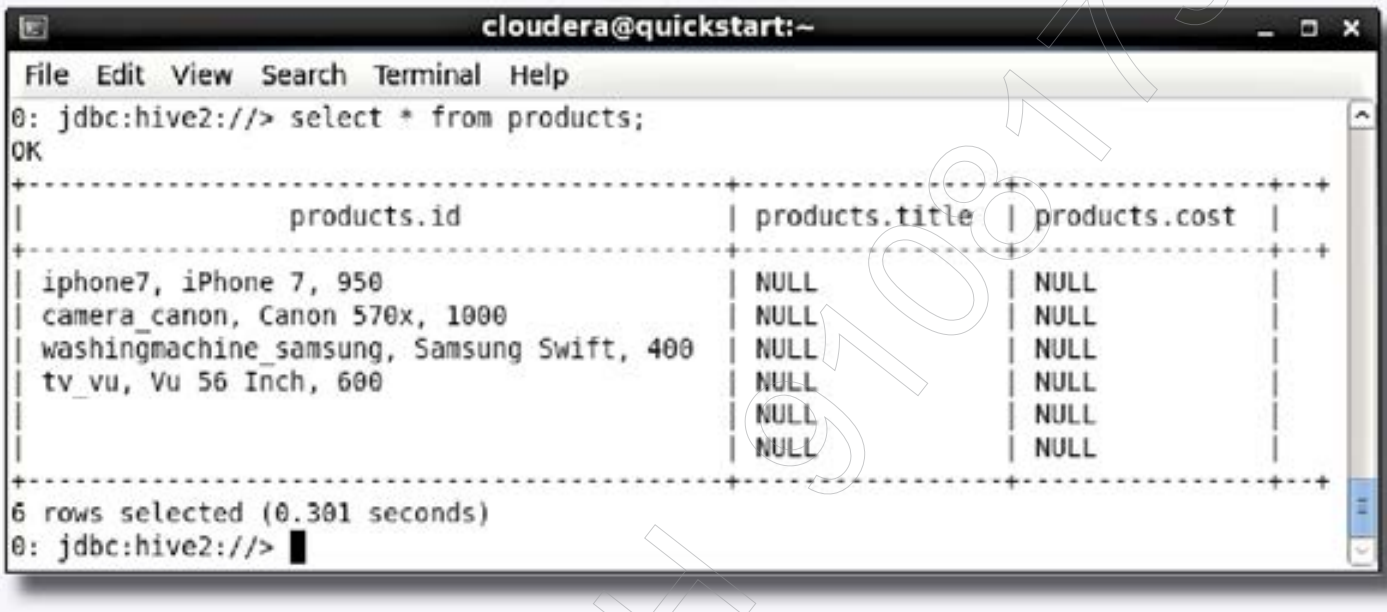
Open a new terminal and enter to Beeline:

```
beeline -u jdbc:hive2://
```



Display the same records of *products* external table using Beeline:

```
select * from products;
```

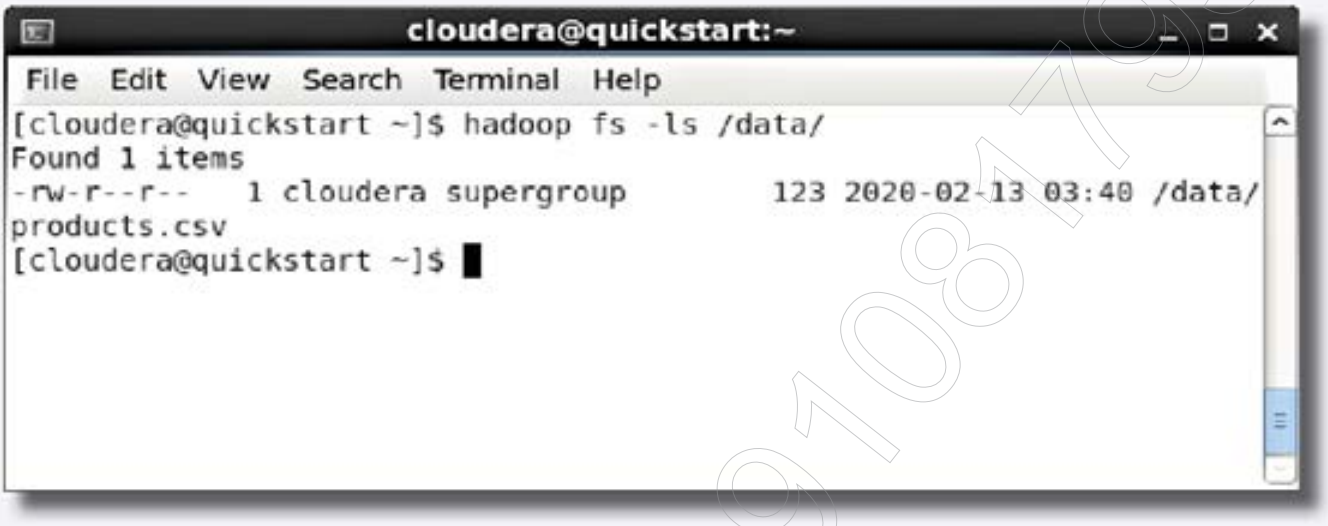


Drop an external table:

```
drop table products;
```

Now, check the *data* folder in hdfs:

```
hadoop fs -ls /data/
```

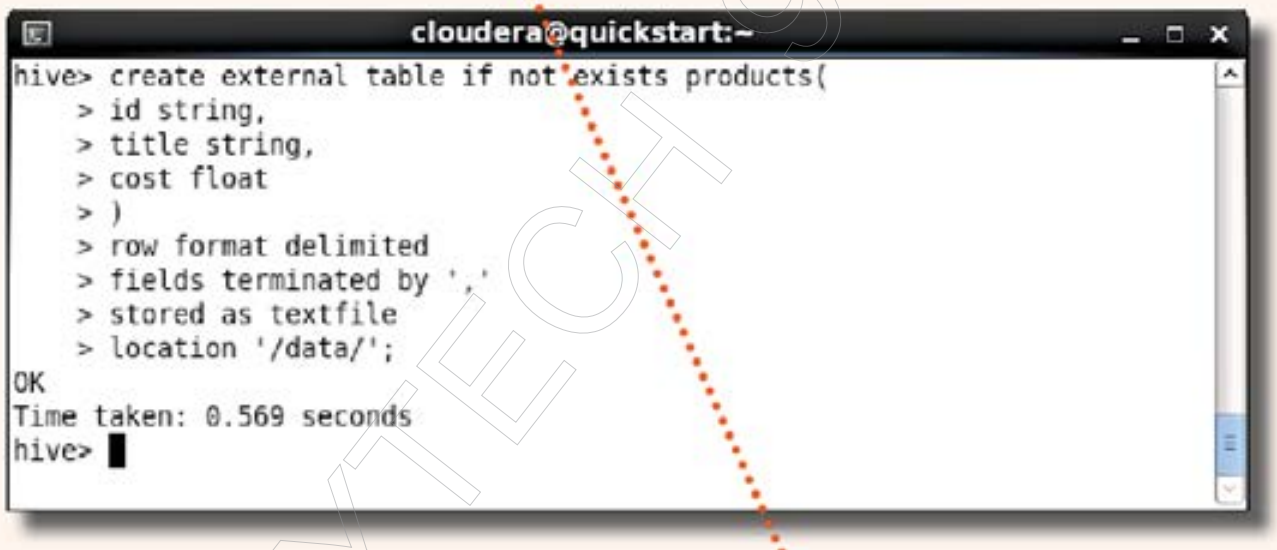


Note: The data is still there.

Note: In Hive External table, data not fully managed by Hive and exists outside the warehouse directory.

Create an external table with delimiters:

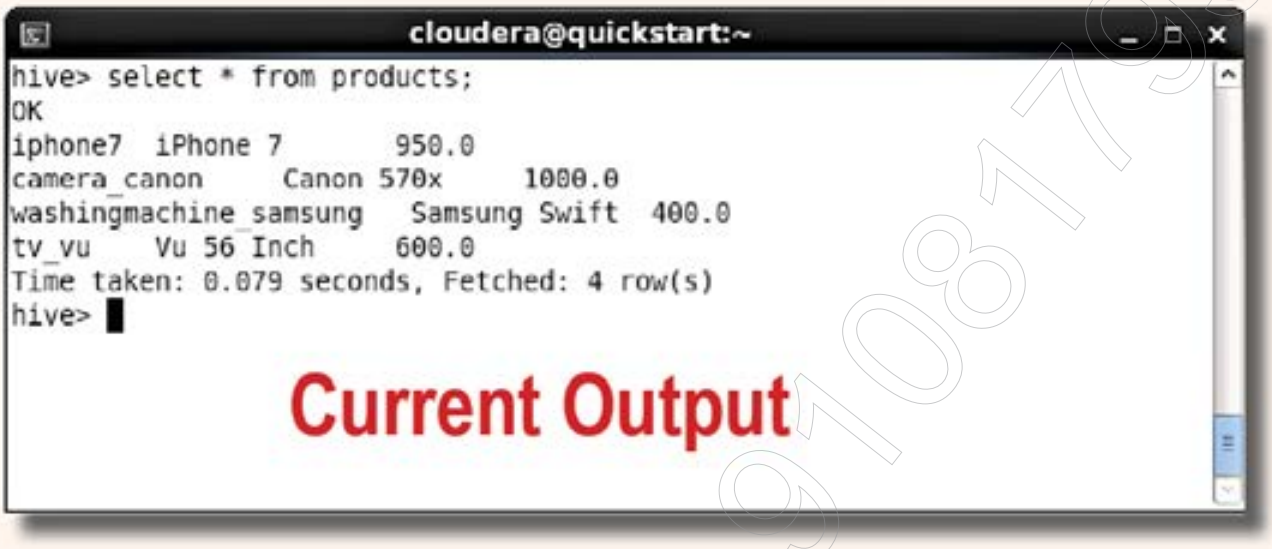
```
create external table if not exists products(
id string,
title string,
cost float
)
row format delimited
fields terminated by ','
stored as textfile
location '/data/';
```

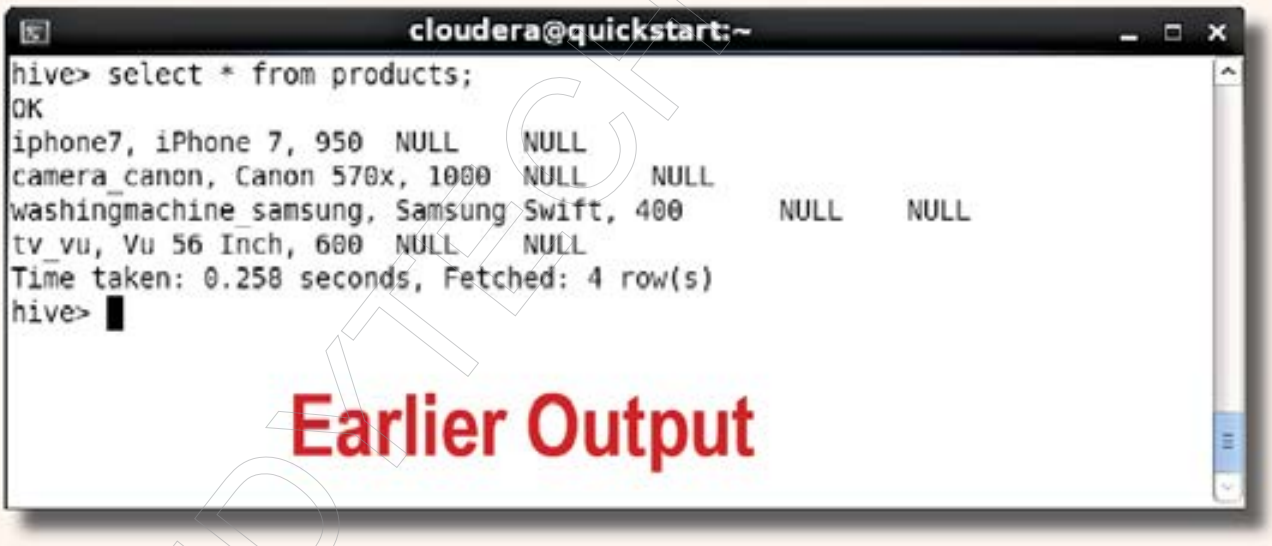


Note: Hive by default takes **text** file format [We can skip this line also].

Display the contents of new *products* table:

```
select * from products;
```



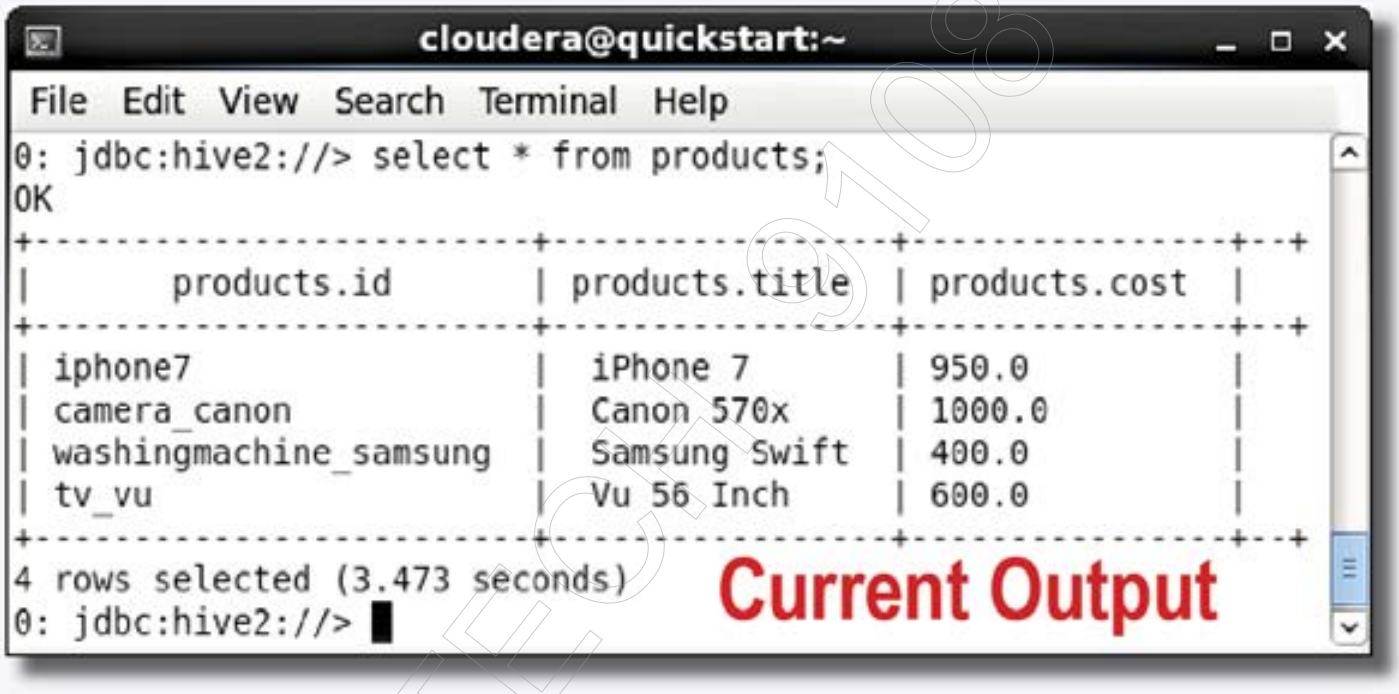


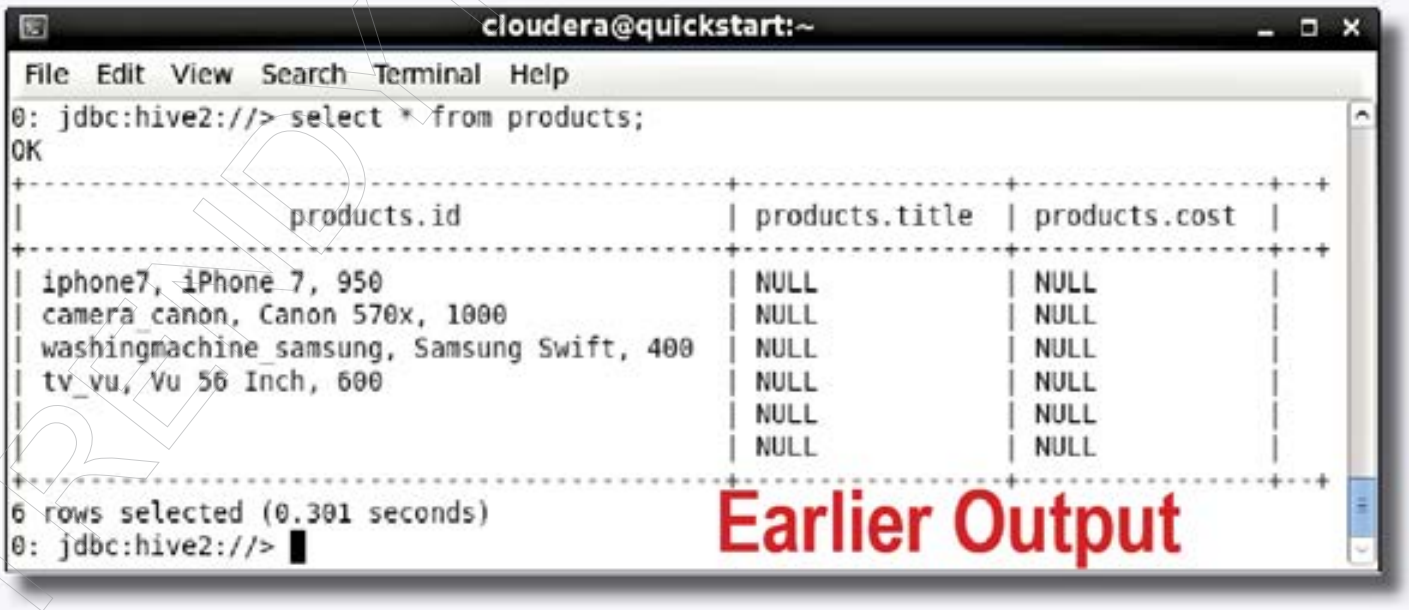
Display the same records (*products*) external table using Beeline:

```
beeline -u jdbc:hive2://
```

→ 1st enter beeline

```
select * from products;
```





How to know a table is Managed or External in Hive:

Detailed information about a Managed table:

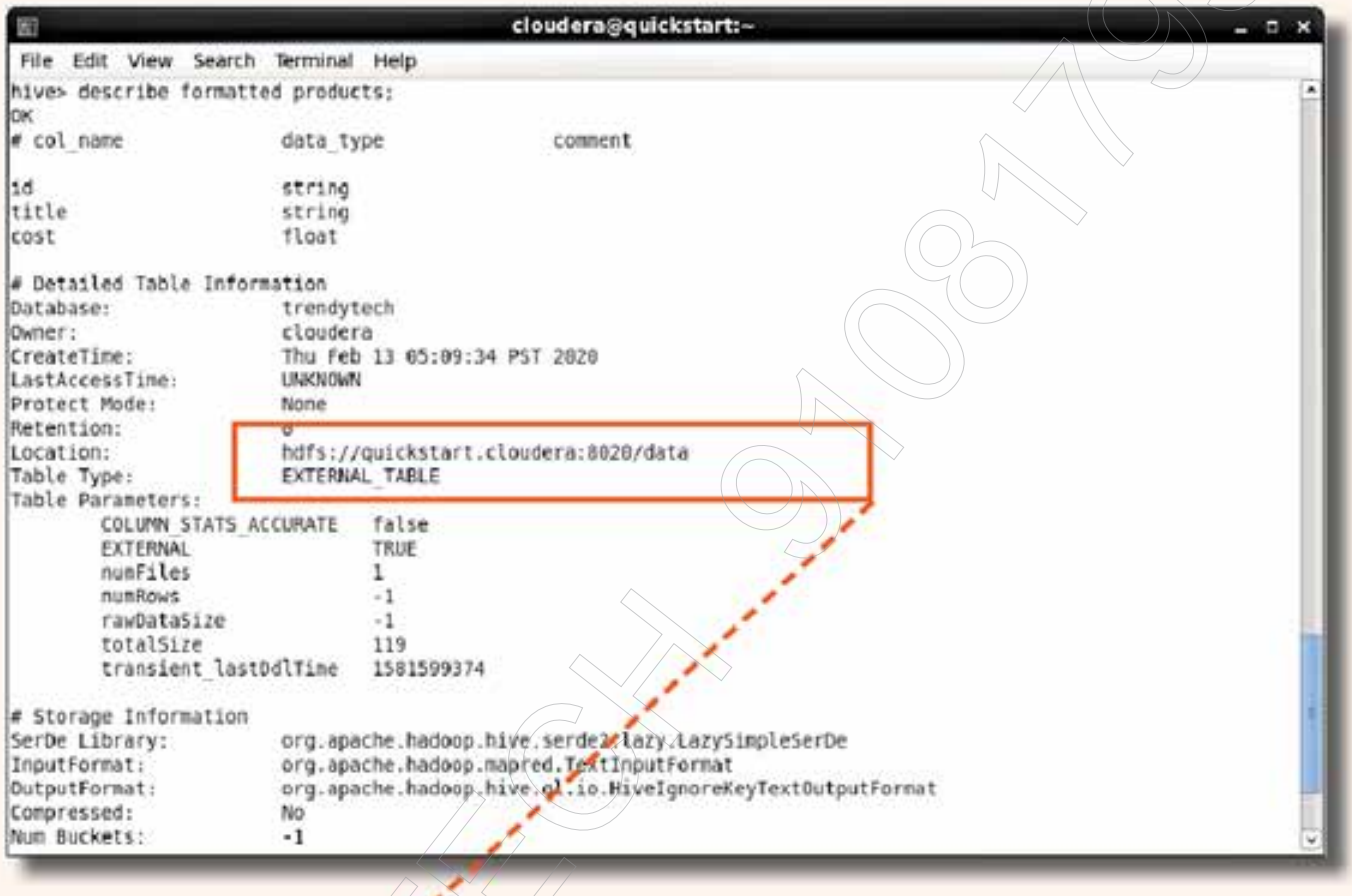
```
describe formatted products;
```

```
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:
        COLUMN_STATS_ACCURATE   true
        numFiles                2
        numRows                 6
        rawDataSize             73
        totalSize               79
        transient_lastDdlTime   1580790920

# 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:         []
```

Note: Location should be Warehouse directory and Table Type must be MANAGED TABLE

Detailed information about an External table:

```
describe formatted customers;
```



Note: Location should be HDFS and Table Type must be EXTERNAL 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/