



Apache Hive

by Sumit Mittal



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.

Hive External Table



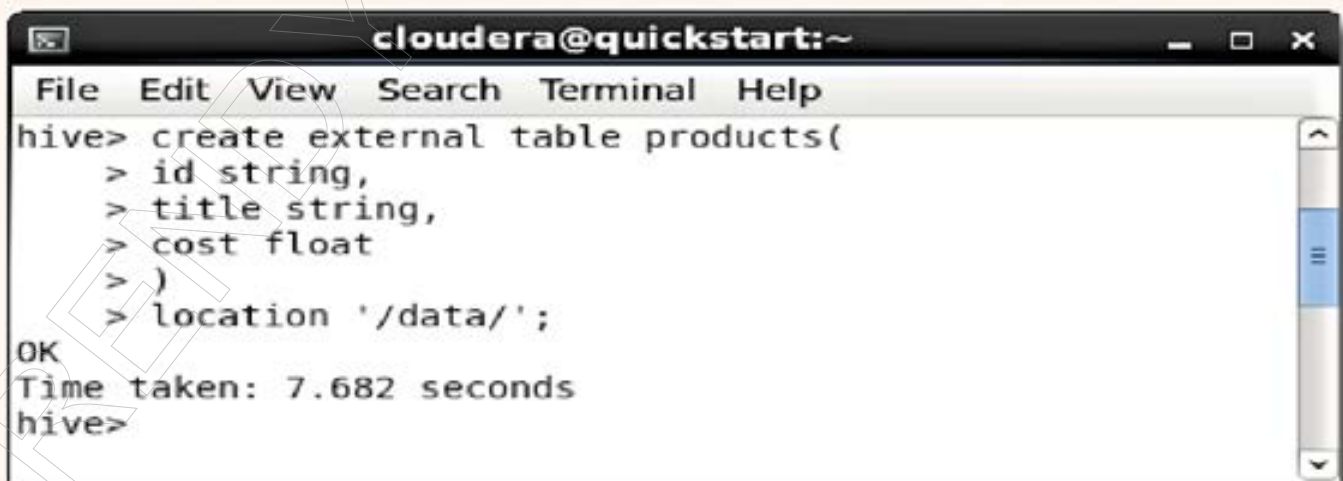
External Table

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/';
```

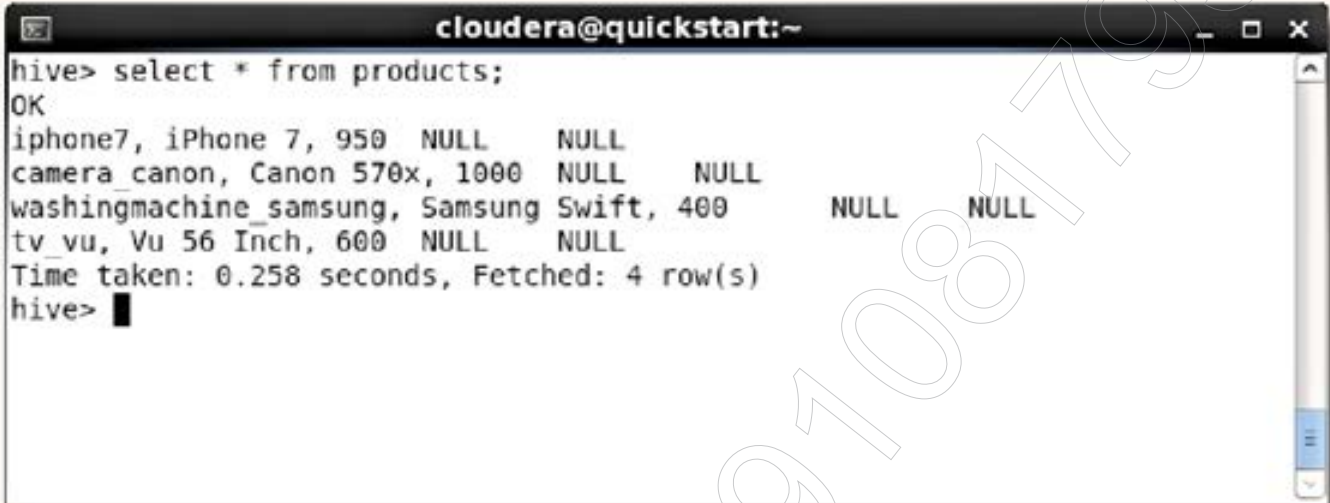


The screenshot shows a terminal window titled "cloudera@quickstart:~". The terminal has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The command being executed is "hive> create external table products(> id string,> title string,> cost float>)> location '/data/';". The output shows "OK" and "Time taken: 7.682 seconds", followed by "hive>".

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> create external table products(  
  > id string,  
  > title string,  
  > cost float  
  > )  
  > location '/data/';  
OK  
Time taken: 7.682 seconds  
hive>
```


Display records of an external table:

```
select * from products;
```




```
cloudera@quickstart:~  
hive> select * from products;  
OK  
iphone7, iPhone 7, 950 NULL NULL  
camera_canon, Canon 570x, 1000 NULL NULL  
washingmachine_samsung, Samsung Swift, 400 NULL NULL  
tv_vu, Vu 56 Inch, 600 NULL NULL  
Time taken: 0.258 seconds, Fetched: 4 row(s)  
hive> █
```

Note: The output is not what we wanted

Open a new terminal and enter to Beeline:

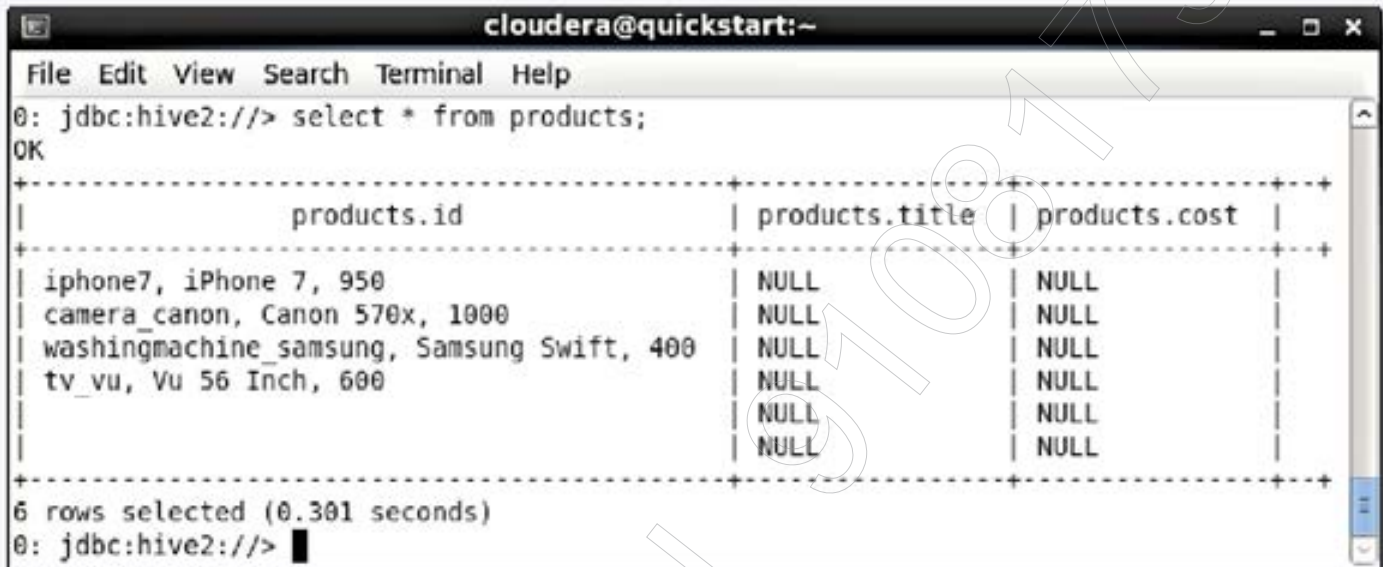
```
beeline -u jdbc:hive2://
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ beeline -u jdbc:hive2://  
scan complete in 7ms  
Connecting to jdbc:hive2://  
Connected to: Apache Hive (version 1.1.0-cdh5.13.0)  
Driver: Hive JDBC (version 1.1.0-cdh5.13.0)  
Transaction isolation: TRANSACTION_REPEATABLE_READ  
Beeline version 1.1.0-cdh5.13.0 by Apache Hive  
0: jdbc:hive2://> █
```

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

```
select * from products;
```



A terminal window titled 'cloudera@quickstart:~' showing the execution of a Beeline query. The prompt is '0: jdbc:hive2://>' and the command is 'select * from products;'. The output is 'OK' followed by a table of 6 rows. The first row contains product details, while the remaining five rows contain NULL values. The table has three columns: products.id, products.title, and products.cost. Below the table, it says '6 rows selected (0.301 seconds)' and the prompt returns to '0: jdbc:hive2://>'.

products.id	products.title	products.cost
iphone7, iPhone 7, 950	NULL	NULL
camera_canon, Canon 570x, 1000	NULL	NULL
washingmachine_samsung, Samsung Swift, 400	NULL	NULL
tv_vu, Vu 56 Inch, 600	NULL	NULL
	NULL	NULL
	NULL	NULL

Drop an external table:

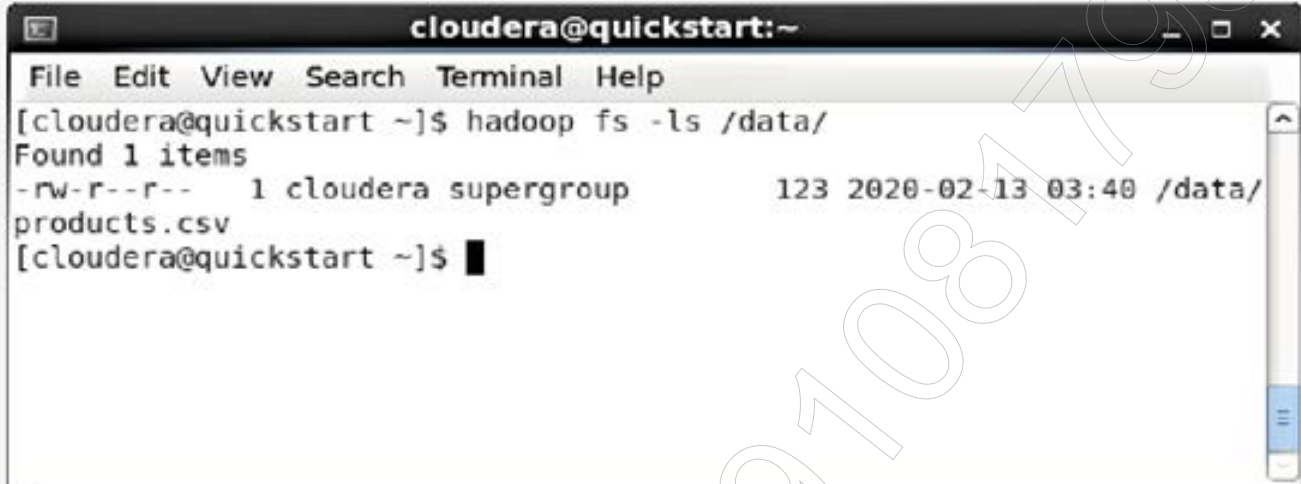
```
drop table products;
```



A terminal window titled 'cloudera@quickstart:~' showing the execution of the 'drop table products;' command. The prompt is 'hive>' and the command is 'drop table products;'. The output is 'OK' followed by 'Time taken: 5.54 seconds' and the prompt returns to 'hive>'.

Now, check the *data* folder in hdfs:

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

A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command '[cloudera@quickstart ~]\$ hadoop fs -ls /data/' is entered. The output is 'Found 1 items' followed by a file listing: '-rw-r--r-- 1 cloudera supergroup 123 2020-02-13 03:40 /data/products.csv'. The prompt '[cloudera@quickstart ~]\$' is shown again with a cursor.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -ls /data/  
Found 1 items  
-rw-r--r-- 1 cloudera supergroup      123 2020-02-13 03:40 /data/  
products.csv  
[cloudera@quickstart ~]$
```

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/';
```

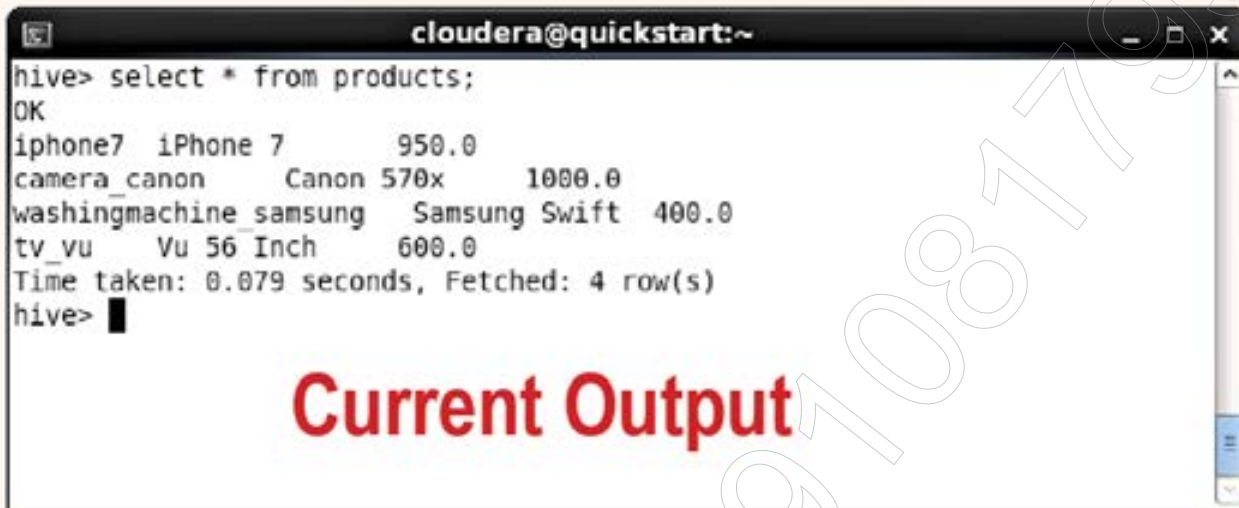


The screenshot shows a terminal window titled "cloudera@quickstart:-". The command being executed is "hive> create external table if not exists products(" followed by a list of fields: "id string,", "title string,", "cost float", and ")", then "row format delimited", "fields terminated by ','", "stored as textfile", and "location '/data/';". The command completes successfully, showing "OK" and "Time taken: 0.569 seconds". The prompt "hive>" is visible at the bottom.

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

Display the contents of new *products* table:

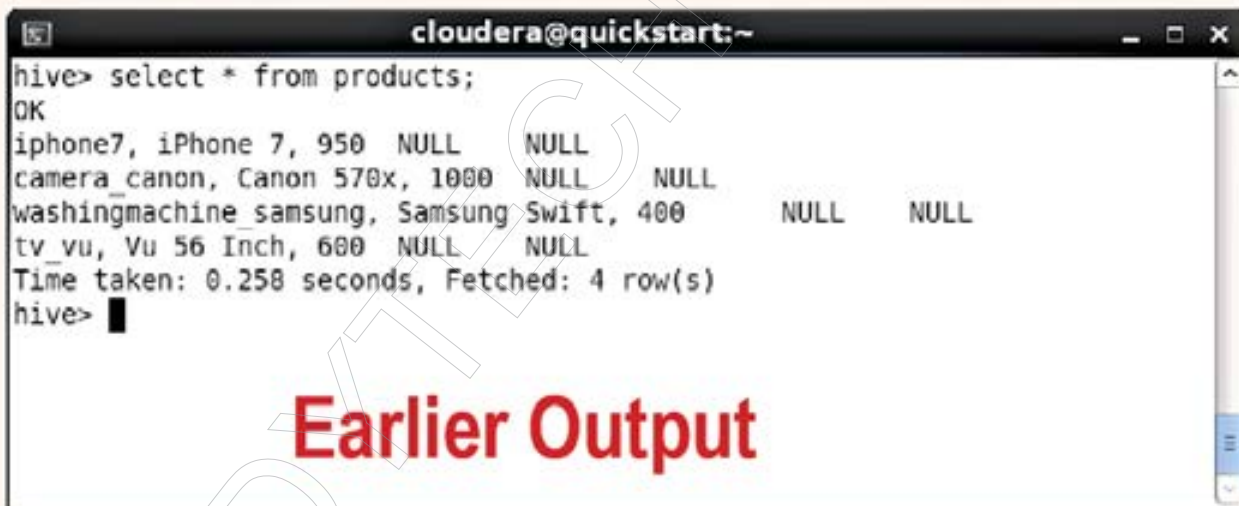
```
select * from products;
```



A terminal window titled 'cloudera@quickstart:~' showing the execution of the SQL query 'select * from products;'. The output displays four rows of product data with their respective prices. The text 'Current Output' is overlaid in red.

```
hive> select * from products;
OK
iphone7 iPhone 7 950.0
camera_canon Canon 570x 1000.0
washingmachine_samsung Samsung Swift 400.0
tv_vu Vu 56 Inch 600.0
Time taken: 0.079 seconds, Fetched: 4 row(s)
hive>
```

Current Output



A terminal window titled 'cloudera@quickstart:~' showing the execution of the SQL query 'select * from products;'. The output displays four rows of product data with their respective prices, including NULL values for some columns. The text 'Earlier Output' is overlaid in red.

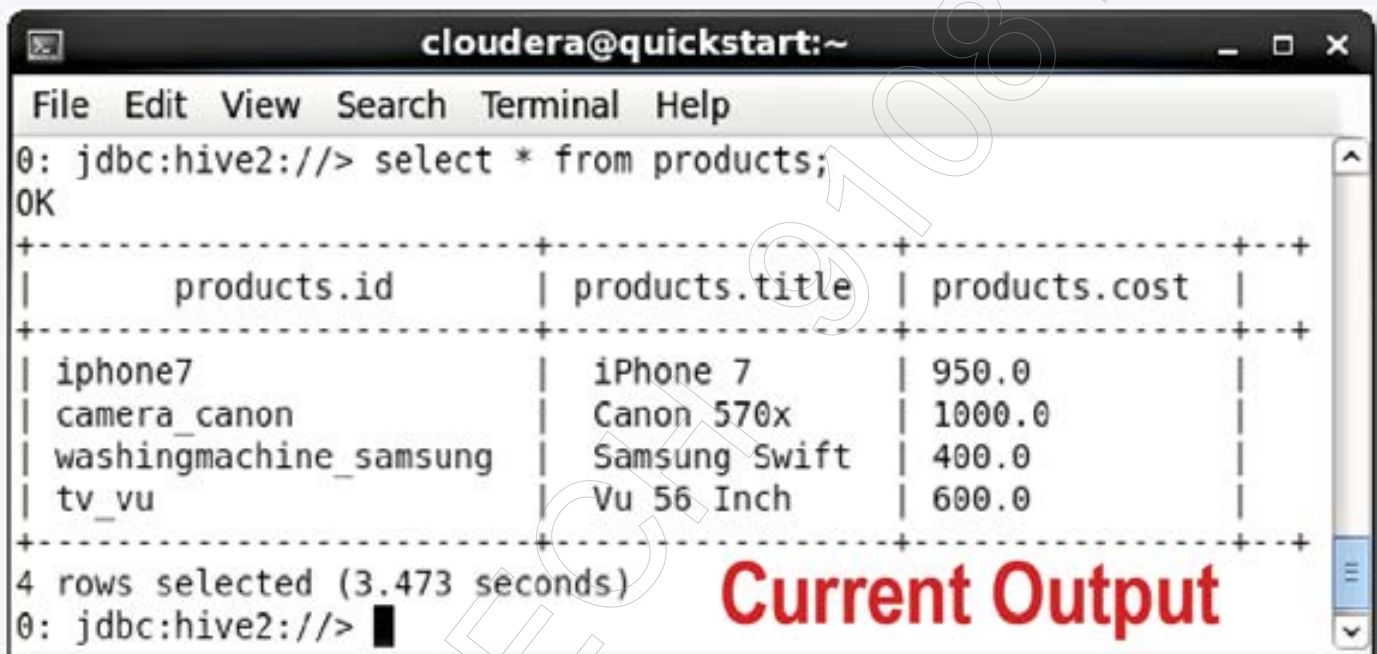
```
hive> select * from products;
OK
iphone7, iPhone 7, 950 NULL NULL
camera_canon, Canon 570x, 1000 NULL NULL
washingmachine_samsung, Samsung Swift, 400 NULL NULL
tv_vu, Vu 56 Inch, 600 NULL NULL
Time taken: 0.258 seconds, Fetched: 4 row(s)
hive>
```

Earlier Output

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

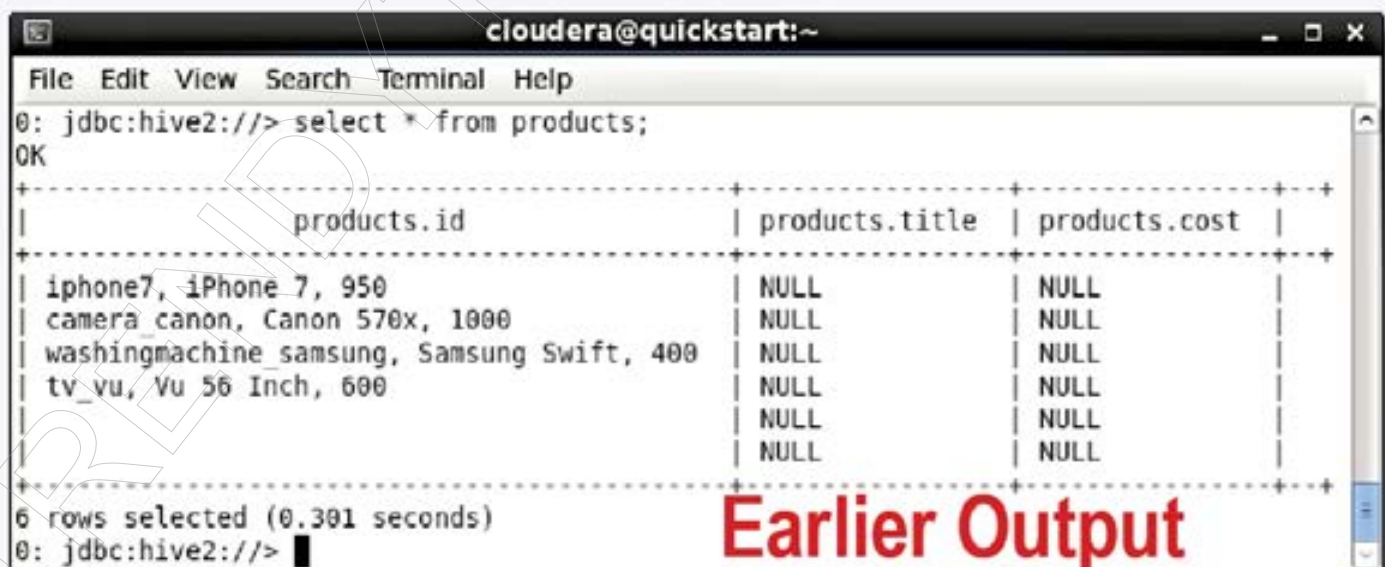
beeline -u jdbc:hive2:// → 1st enter beeline

```
select * from products;
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
0: jdbc:hive2://> select * from products;  
OK  
+-----+-----+-----+  
| products.id | products.title | products.cost |  
+-----+-----+-----+  
| iphone7     | iPhone 7       | 950.0         |  
| camera_canon | Canon 570x     | 1000.0        |  
| washingmachine_samsung | Samsung Swift | 400.0         |  
| tv_vu       | Vu 56 Inch     | 600.0         |  
+-----+-----+-----+  
4 rows selected (3.473 seconds)  
0: jdbc:hive2://> █
```

Current Output



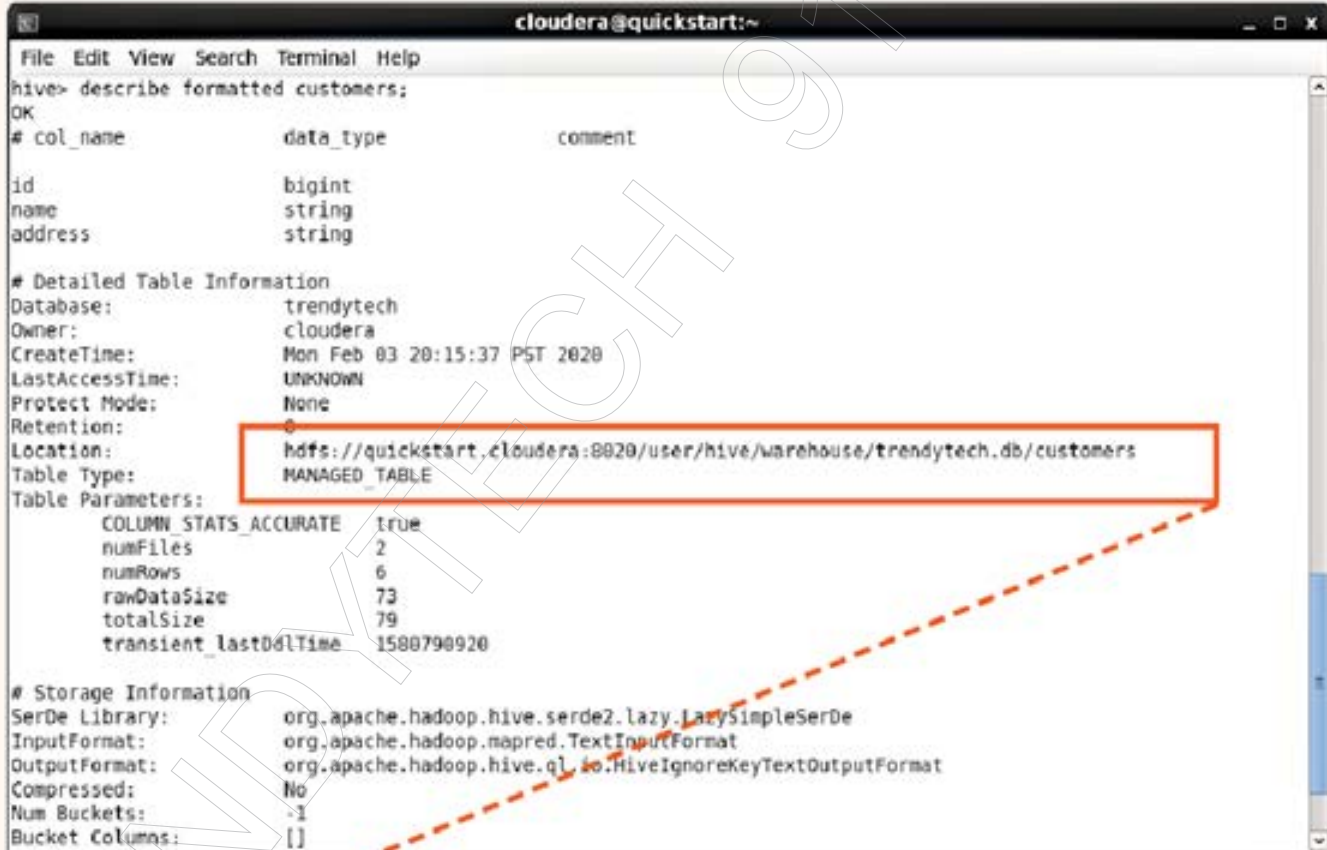
```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
0: jdbc:hive2://> select * from products;  
OK  
+-----+-----+-----+  
| products.id | products.title | products.cost |  
+-----+-----+-----+  
| iphone7, iPhone 7, 950 | NULL | NULL |  
| camera_canon, Canon 570x, 1000 | NULL | NULL |  
| washingmachine_samsung, Samsung Swift, 400 | NULL | NULL |  
| tv_vu, Vu 56 Inch, 600 | NULL | NULL |  
| NULL | NULL | NULL |  
+-----+-----+-----+  
6 rows selected (0.301 seconds)  
0: jdbc:hive2://> █
```

Earlier Output

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

Detailed information about a Managed table:

```
describe formatted products;
```

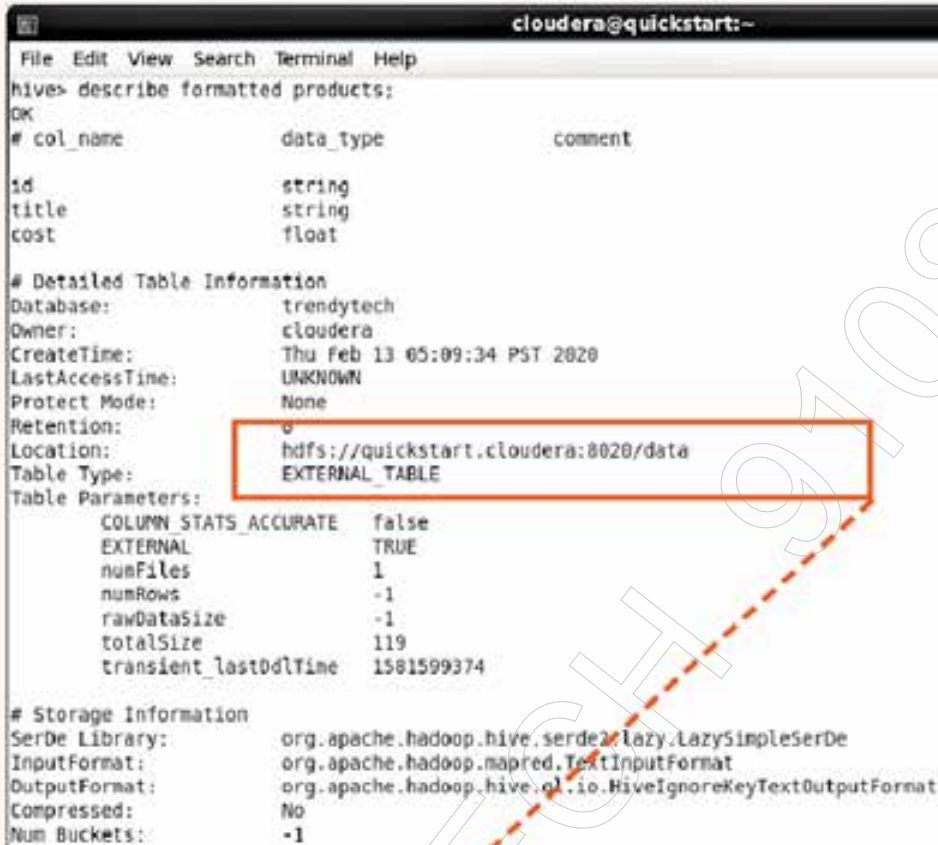


```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> describe formatted customers;  
OK  
# col_name          data_type          comment  
  
id                  bigint  
name                string  
address             string  
  
# Detailed Table Information  
Database:           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.q1.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;



```
cloudera@quickstart:~$ hive> describe formatted products;
OK
# col_name          data_type          comment
id                  string
title              string
cost               float

# Detailed Table Information
Database:          trendytech
Owner:             cloudera
CreateTime:        Thu Feb 13 05:09:34 PST 2020
LastAccessTime:    UNKNOWN
Protect Mode:      None
Retention:         0
Location:           hdfs://quickstart.cloudera:8020/data
Table Type:        EXTERNAL_TABLE

Table Parameters:
  COLUMN_STATS_ACCURATE false
  EXTERNAL              TRUE
  numFiles              1
  numRows               -1
  rawDataSize           -1
  totalSize             119
  transient_lastDdlTime 1581599374

# 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
```

A red dashed arrow points from the red box in the terminal screenshot to the 'Note' text below.

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

Youtube TrendyTech

Twitter @BigdataBySumit

Instagram bigdatabysumit

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

