Apache Hive

by Sumit Mittal

TRENDY TECH
UPLIFT YOUR CAREER!

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

Create a folder (*data*) in the root directory:

```
hadoop fs -mkdir /data
```

```
cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ hadoop fs -mkdir /data
[cloudera@quickstart ~]$
```

List the Hive root directory again and check that the data folder has now been created:

```
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 2020-02-13 01:56 /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 ~]$
```

Download the *products.csv* file from Google Classroom into Cloudera Downloads folder:

```
cloudera@quickstart:~/Downloads
File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ cd Downloads
[cloudera@quickstart Downloads]$
[cloudera@quickstart Downloads]$ ls pro*
products.csv
[cloudera@quickstart Downloads]$
```

Move the *products.csv* file into the *data* folder of hdfs which we have created:

```
hadoop fs -copyFromLocal /home/cloudera/
Downloads/products.csv  /data/
```

```
cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ hadoop fs -copyFromLocal /home/cloudera/Downloads/produc
ts.csv  /data/
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
```

Check the *data* folder - the *products.csv* file must appear:

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

Note: In Managed table, data managed fully by Hive and stored in the warehouse directory.

2.1 Loading data into Managed table from LFS (Local File System)

Create a Managed table:

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

Load data into Managed table from a local path:

```
load data local inpath
'/home/cloudera/Downloads/products.csv'
into table products_managed;
```

```
cloudera@quickstart:~
hive> load data local inpath
    > '/home/cloudera/Downloads/products.csv'
    > into table products_managed;
Loading data to table trendytech.products_managed
Table trendytech.products_managed stats: [numFiles=1, totalSize=119]
OK
Time taken: 1.498 seconds
hive>
```

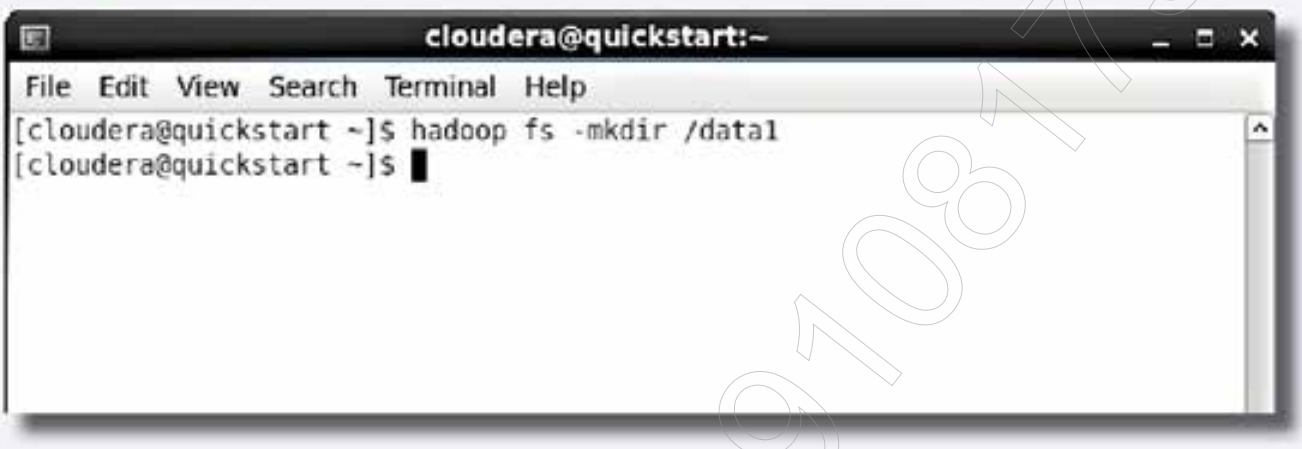
Now, check the output of managed table:

```
select * from products_managed;
```

```
cloudera@quickstart:~
hive> select * from products_managed;
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: 0.09 seconds, Fetched: 4 row(s)
hive>
```

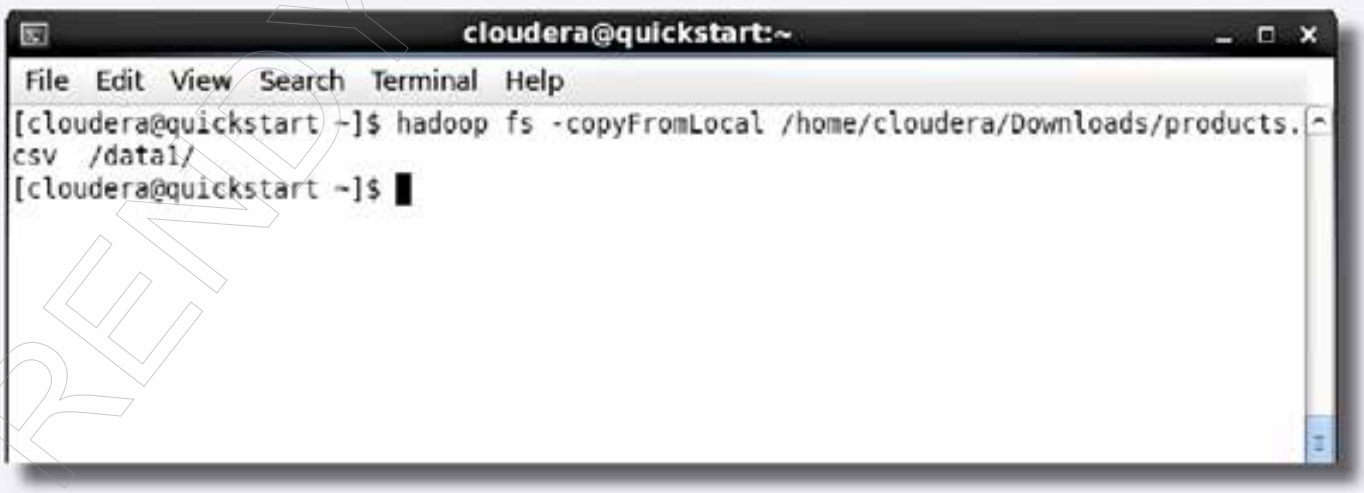
Create one more folder in the root directory of HDFS:

```
hadoop fs -mkdir /data1
```



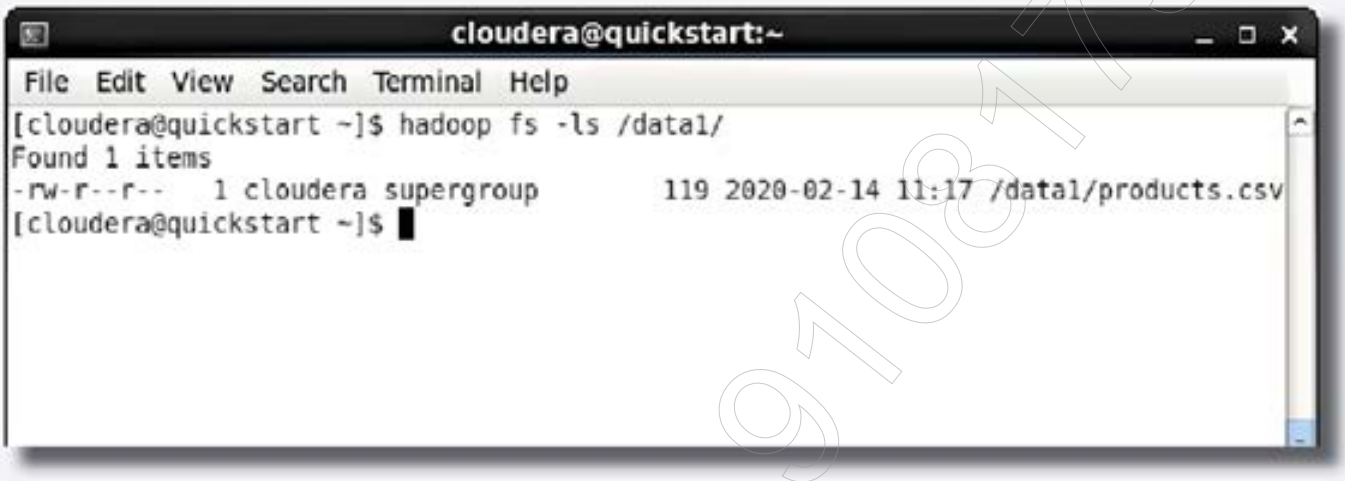
Move again the *products.csv* file into the *data1* folder:

```
hadoop fs -copyFromLocal /home/cloudera/
Downloads/products.csv  /data1/
```



Check the *data1* folder - the *products.csv* file must appear:

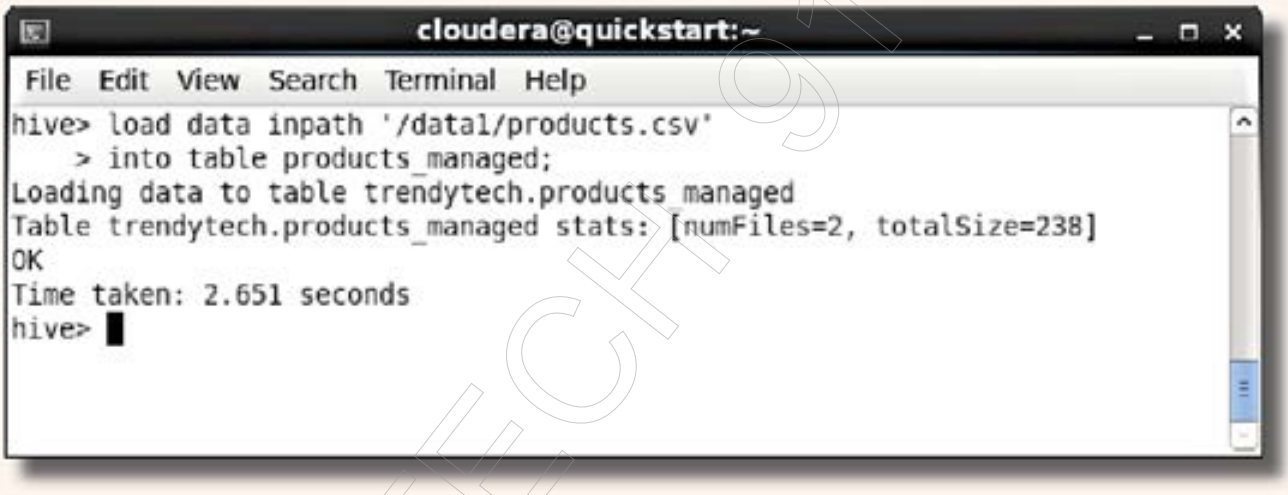
```
hadoop fs -ls /data1/
```



2.2 Loading data into Managed table from HDFS

Load data again into the *products_managed* table:

```
load data inpath '/data1/products.csv'
into table products_managed;
```



Check the records of *products_managed* table:

```
select * from products_managed;
```

```
cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
hive> select * from products_managed;
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
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: 0.101 seconds, Fetched: 8 row(s)
hive>
```

1st load data

2nd load data

Note: the data is appended, and both files are present in the warehouse directory.

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

```
cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/trendytech.db/products
_managed/
Found 2 items
-rwxrwxrwx   1 cloudera supergroup        119 2020-02-13 06:35 /user/hive/warehous
e/trendytech.db/products_managed/products.csv
-rwxrwxrwx   1 cloudera supergroup        119 2020-02-14 11:17 /user/hive/warehous
e/trendytech.db/products_managed/products_copy_1.csv
[cloudera@quickstart ~]$
```

Overwrite a Hive table:

```
load data local inpath
'/home/cloudera/Downloads/products.csv'
overwrite into table products_managed;
```

```
cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
hive> load data local inpath
    > '/home/cloudera/Downloads/products.csv'
    > overwrite into table products_managed;
Loading data to table trendytech.products_managed
Table trendytech.products_managed stats: [numFiles=1, numRows=0, totalSize=1
19, rawDataSize=0]
OK
Time taken: 0.948 seconds
hive>
```

Check again output of *products_managed* table:

```
select * from products_managed;
```

```
cloudera@quickstart:~
hive> select * from products_managed;
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: 0.09 seconds, Fetched: 4 row(s)
hive>
```

3. Data loading using table to table method

Create one more Managed table:

```
create table if not exists products_managed2(
id string,
title string,
cost float
)
row format delimited
fields terminated by ','
stored as textfile;
```

Load data from existing table to the new table:

```
insert into table products_managed2
select * from products_managed;
```

```
cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
hive> insert into table products_managed2
    > select * from products_managed;
Query ID = cloudera_20200421074848_0f3b2f8c-ec4b-47d0-968f-0fbdc279f76f
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_1587475275867_0002, Tracking URL = http://quickstart.cloudera:8088/proxy
/application_1587475275867_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1587475275867_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2020-04-21 07:48:22,928 Stage-1 map = 0%,  reduce = 0%
2020-04-21 07:48:41,152 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.68 sec
MapReduce Total cumulative CPU time: 2 seconds 680 msec
Ended Job = job_1587475275867_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/products_managed2/.hive
-staging_hive_2020-04-21_07-48-00_094_4982679823924594 72-1/-ext-10000
Loading data to table default.products_managed2
Table default.products_managed2 stats: [numFiles=1, numRows=4, totalSize=124, rawDataSize=1
20]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 2.68 sec   HDFS Read: 4064 HDFS Write: 206 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 680 msec
OK
Time taken: 68.125 seconds
hive>
```

Google REVIEWS

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/

TRENDYTECH

UPLIFT YOUR CAREER!