

# Hive common Commands

---

Start Hive

```
hive
```

Show tables

```
show tables;
```

Create table

```
CREATE TABLE stack_overflow_tags(id BIGINT, title string, body string, tag1 string,tag2 string,tag3 string,tag4 string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' ESCAPED BY '\\';
```

Next step is to fill the data in table. Make sure that the dataset is on HDFS

```
dfs -ls / ;
```

Blank table is created with schema. Now fill the data inside the table.

```
LOAD DATA INPATH '/stack_overflow_hdfs' INTO TABLE stack_overflow_tags;
```

Check the data on hive data warehouse

[http://localhost:50070/explorer.html#/user/hive/warehouse/stack\\_overflow\\_tags](http://localhost:50070/explorer.html#/user/hive/warehouse/stack_overflow_tags)

Select first few rows

```
select * from stack_overflow_tags LIMIT 3;
```

Sending the output to a new file

```
INSERT OVERWRITE LOCAL DIRECTORY '/home/hduser/Output/hive_out1' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' select * from stack_overflow_tags LIMIT 3;
```

Have a look at the exported file. Do this in a new (non-hive) terminal

```
cat /home/hduser/Output/hive_out1/000000_0
```

## Count(\*)in Hive

A simple count Query

```
select count(*) from stack_overflow_tags;
```

## Group by in Hive

The output is too big. We need to give the output file path. This command can be executed in a non-hive terminal

```
hive -e "select tag1, count(*) as tag1_count from stack_overflow_tags group by tag1 order by tag1_count DESC" > /home/hduser/Output/hive_out_tag1.txt
```

Have a look at the exported file. Do this in a new (non-hive) terminal

```
cat /home/hduser/Output/hive_out_tag1.txt
```

## Word count in Hive

This command can be executed in a non-hive terminal

```
hive -e "SELECT word, count(1) AS count FROM (SELECT explode(split(tag1, '\s')) AS word FROM stack_overflow_tags) w GROUP BY word ORDER BY count DESC" > /home/hduser/Output/hive_out_wordcount.txt
```

Have a look at the exported file. Do this in a new (non-hive) terminal

```
cat /home/hduser/Output/hive_out_wordcount.txt
```

## Executing a script from a file

Create a hive query script file. Fill file with queries

```
mkdir /home/hduser/codes/  
gedit /home/hduser/codes/my_query.hql
```

fill the above file with below query

```
select * from stack_overflow_tags LIMIT 3;
```

use source to execute it from hive

```
hive  
source /home/hduser/codes/my_query.hql;
```

## Joins

Two Tables `Online_Retail_Customer`, `Online_Retail_Invoice`

## Push the datasets onto HDFS

```
sudo hadoop fs -copyFromLocal  
/home/hduser/datasets/Online_Retail_Sales_Data/Online_Retail_Customer.txt  
/Online_Retail_Customer  
  
hadoop fs -ls /  
  
hadoop fs -copyFromLocal  
/home/hduser/datasets/Online_Retail_Sales_Data/Online_Retail_Invoice.txt  
/Online_Retail_Invoice  
  
hadoop fs -ls /
```

## Create Table Schema for Online\_Retail\_Customer on Hive

```
hive  
  
CREATE TABLE Tbl_Online_Retail_Customer(uniq_idc string, InvoiceDate  
string, UnitPrice INT, CustomerID INT, Country string) ROW FORMAT DELIMITED  
FIELDS TERMINATED BY '\t' ESCAPED BY '\\\';
```

## Fill the table with data

```
LOAD DATA INPATH '/Online_Retail_Customer' INTO TABLE  
Tbl_Online_Retail_Customer;
```

Check the link <http://localhost:50070/explorer.html#/user/hive/warehouse>

## Create Table Schema for Online\_Retail\_Invoice on Hive

```
CREATE TABLE Tbl_Online_Retail_Invoice(uniq_idi string, InvoiceNo string,  
StockCode string, Description string, Quantity INT) ROW FORMAT DELIMITED  
FIELDS TERMINATED BY '\t' ESCAPED BY '\\\';
```

## Fill the table with data

```
LOAD DATA INPATH '/Online_Retail_Invoice' INTO TABLE  
Tbl_Online_Retail_Invoice;
```

Check the link <http://localhost:50070/explorer.html#/user/hive/warehouse>

## Left join

Drop the table if it is already there.

```
Drop table Tbl_left_join;
```

```
CREATE TABLE Tbl_left_join as SELECT * FROM Tbl_Online_Retail_Customer t1  
LEFT JOIN Tbl_Online_Retail_Invoice t2 ON t1.uniq_idc = t2.uniq_idi;
```

How many rows in the resultant table?

```
select count(*) from Tbl_left_join;  
select count(*) from Tbl_Online_Retail_Customer;  
select count(*) from Tbl_Online_Retail_Invoice;
```

## Right join

Drop the table if it is already there.

```
Drop table Tbl_right_join;
```

```
CREATE TABLE Tbl_right_join as SELECT * FROM Tbl_Online_Retail_Customer t1  
RIGHT JOIN Tbl_Online_Retail_Invoice t2 ON t1.uniq_idc = t2.uniq_idi;
```

How many rows in the resultant table?

```
select count(*) from Tbl_right_join;  
select count(*) from Tbl_Online_Retail_Customer;  
select count(*) from Tbl_Online_Retail_Invoice;
```

## Inner Join

```
Drop table Tbl_inner_join;
```

```
CREATE TABLE Tbl_inner_join as SELECT * FROM Tbl_Online_Retail_Customer t1  
JOIN Tbl_Online_Retail_Invoice t2 ON t1.uniq_idc = t2.uniq_idi;
```

How many rows in the resultant table?

```
select count(*) from Tbl_inner_join;  
select count(*) from Tbl_Online_Retail_Customer;  
select count(*) from Tbl_Online_Retail_Invoice;
```

## Full join

```
Drop table Tbl_Full_outer_join;
```

```
CREATE TABLE Tbl_Full_outer_join as SELECT * FROM  
Tbl_Online_Retail_Customer t1 FULL JOIN Tbl_Online_Retail_Invoice t2 ON  
t1.uniq_idc = t2.uniq_idi;
```

How many rows in the resultant table?

```
select count(*) from Tbl_Full_outer_join;  
select count(*) from Tbl_Online_Retail_Customer;  
select count(*) from Tbl_Online_Retail_Invoice;
```