# Solution Extraction Using HIVE

**Use the following command to ‘Create a Table’ in HIVE**

Hive> create table data (int,month int,day int,order1 int,country int,session\_ID double,page\_1 int,page\_2 string,colour int,location int,model\_photography int,price double,price\_2 int,page int

)

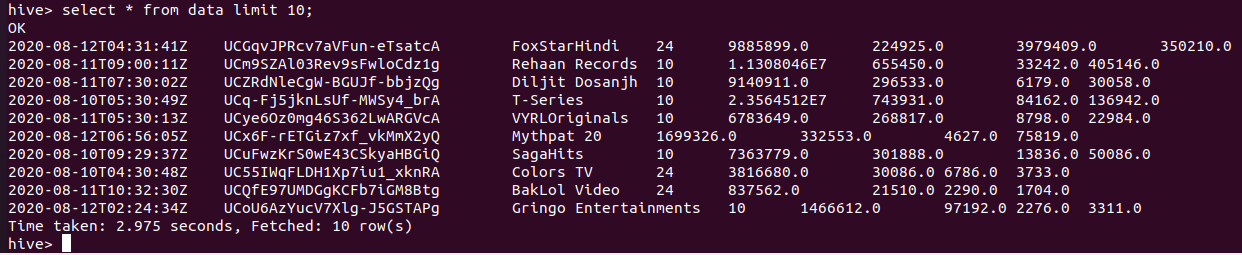
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

tblproperties("skip.header.line.count"="1");

This command will create a Hive table named ‘YouTube\_data\_table’ in which rows will be delimited and rows fields will be terminated by commas**.**

**Selecting the tables**

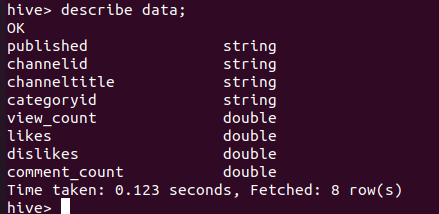
This hive query will select randomly a ten records from the table ”data” using the function limit.



1. **Describing the dataset loaded:(Structure)**

Hive> describe data;

From this query we can able to see the structure of the dataset and their respective datatypes.

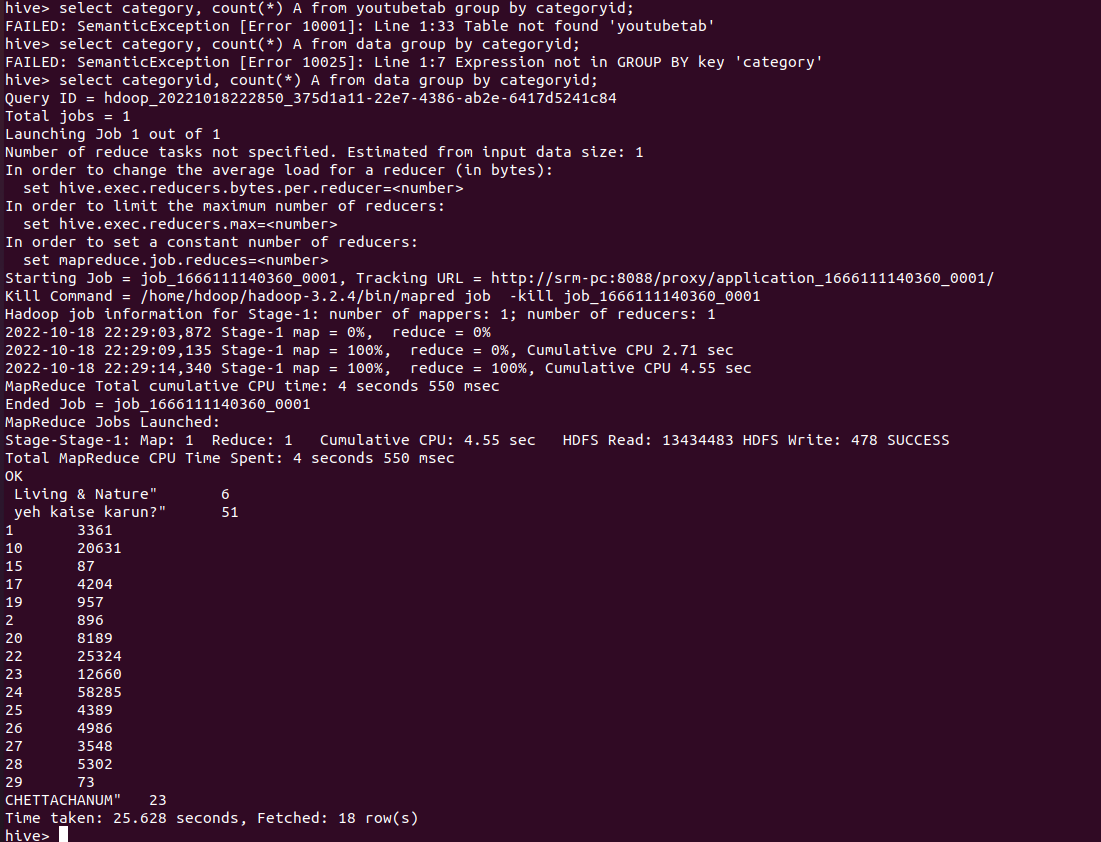


1. **Group by clause:**

The simple group by clause is used to group all the similar rows and increases their count.

hive> select category, count(\*) A FROM data GROUP BY categoryid;

This command will count the total number of category in the row and group it as once.



1. **Order by clause:**

The simple order by clause is used to group all the similar rows and display either in ascending or descending order.

hive> select channelid, likes FROM data ORDER BY likes LIMIT 5;

This command will disply the channelid and likes frm the table data and it by default displays in the ascending order.

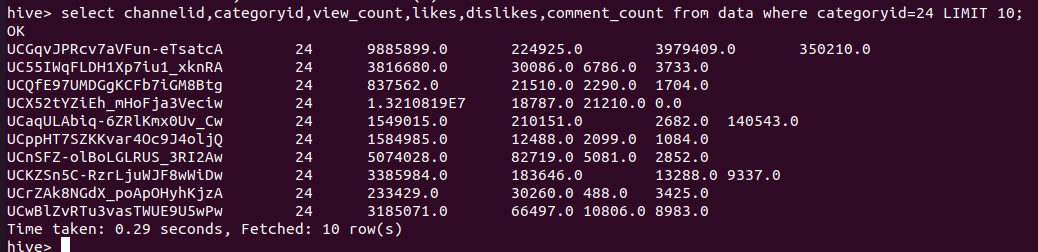


1. **Conditional statement:**

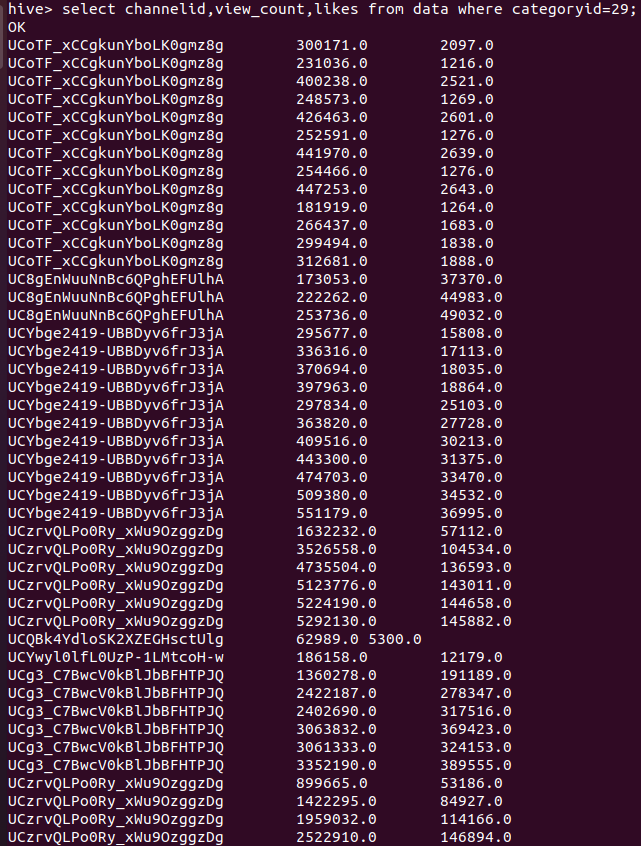
The simple condition is used to display the rows rows some of the conditions used were

hive> select channelid, categoryid, view\_count,likes,dislikes, comment\_count FROM data WHERE categoryid=24 LIMIT 10;

* **Equalto(=)**  It will gives the exact result of the given conditions.

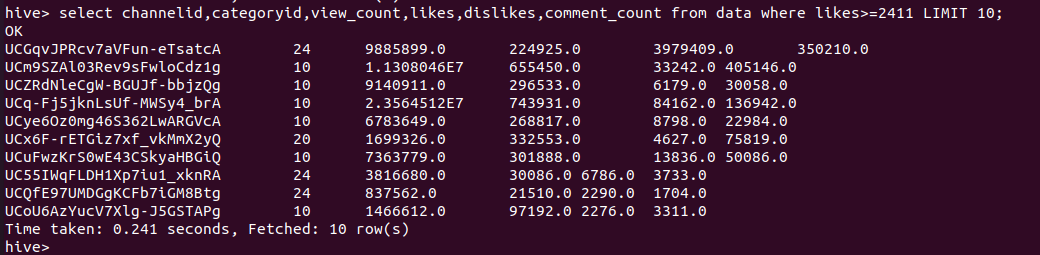


hive> select channelid, view\_count,likes FROM data WHERE categoryid=29;



* **Greaterthan(>=)** This condition will give the values greater than the specified values.

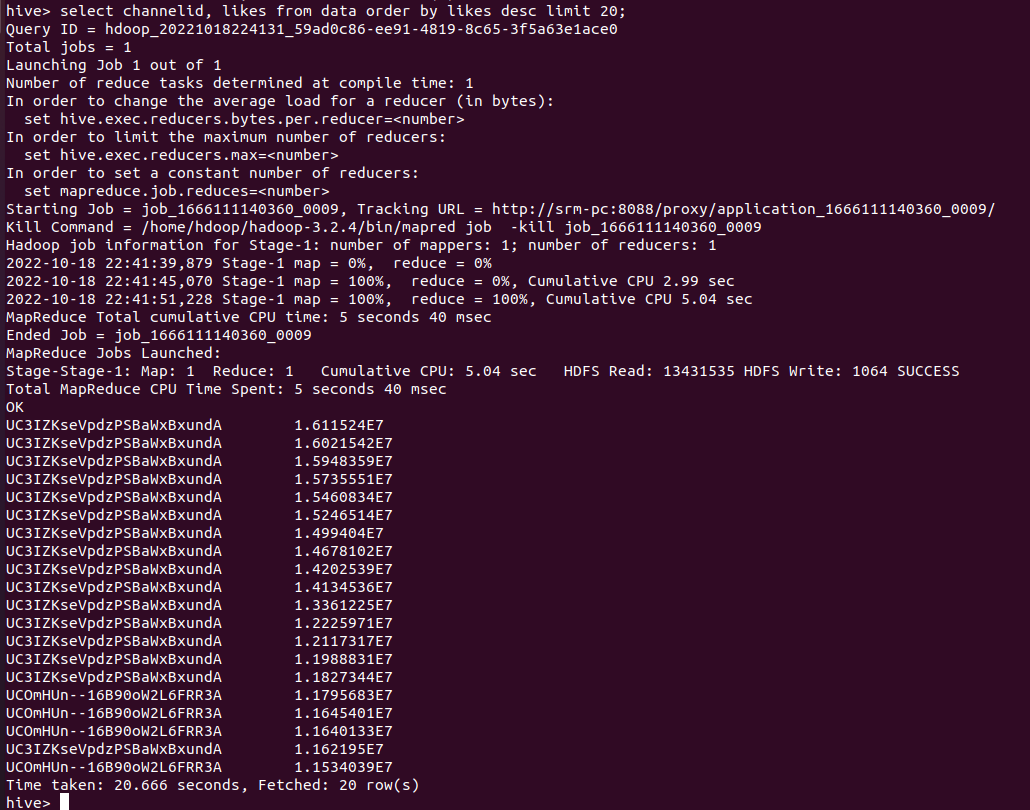
hive> select channelid, categoryid, view\_count,likes,dislikes, comment\_count FROM data WHERE LIKES>=2411 LIMIT 10;

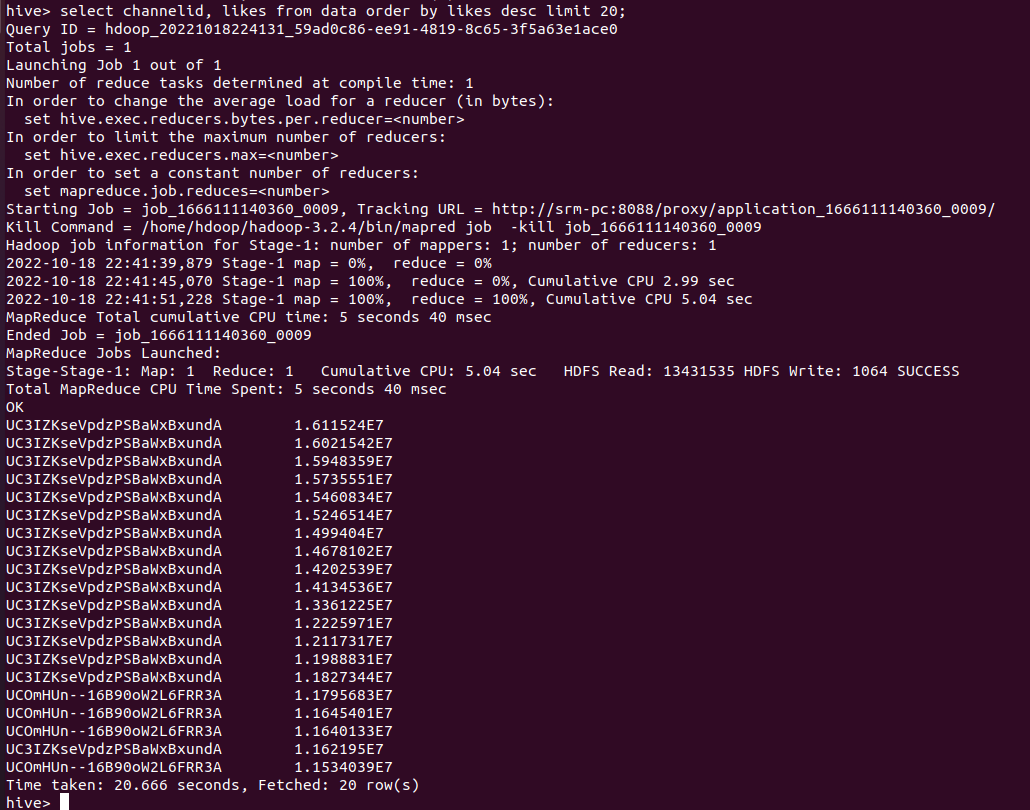


1. **Calculate top 10 channels with maximum number of likes**

We can extract the top 10 channels with maximum number of likes using the following Hive query. The Hive select query will trigger the following MapReduce job:

hive> select channelid, likes FROM data ORDER BY likes DESC LIMIT 10;



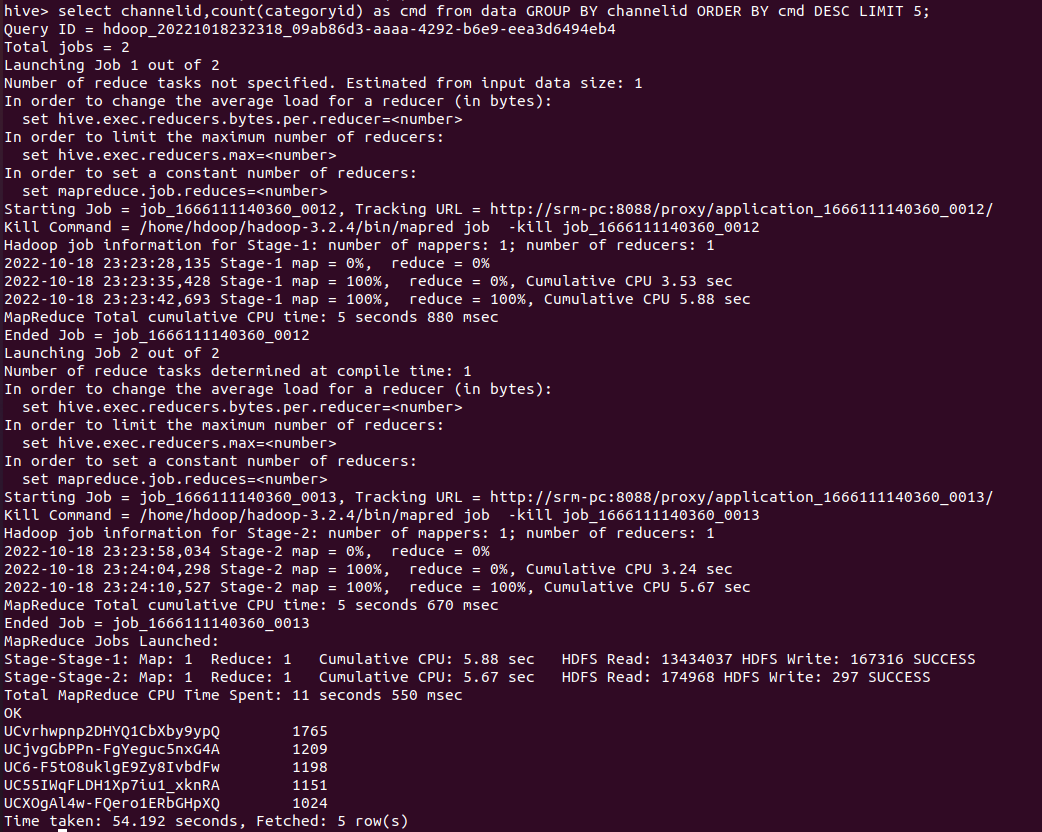


The output result describes that for a specific category id, how many likes were received. The number of likes -- or "thumbs-up" -- a video had has a direct significance to the YouTube video's ranking, according to YouTube Analytics. So if a company posts its video on YouTube, then the number of YouTube likes the company has could determine whether the company or its competitors appear more prominently in YouTube search results. The output result shows number of likes for "Disney" channel videos

1. **Calculate top 5 channels with maximum number of category\_id**

hive> select channelid, count(categoryid) as cmd FROM data GROUP BY channelid ORDER BY cmd DESC LIMIT 10;

This command will say top channel in our dataset based in the category so we can analysis easily as this category id has more number of the channels.



# Calculate top 5 categories with maximum number of comments

hive > select category, max(no\_of\_comments) as max\_no\_of\_comments from YouTube\_data\_table GROUP ORDER BY max\_no\_of\_comments DESC LIMIT 5;