

Youtube Analysis

Creating and Loading data into Tables



Create Youtube Table and load youtubedata.txt file into table

Query:

Create table youtubedata (id STRING,uploader STRING,inter INT, category STRING, length INT, nviews INT, rating FLOAT, nrating INT, ncomment INT, rvideos ARRAY<STRING>) row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfile LOCATION /user/praveenkum79edu/YouTube_Analysis;

Hive ScreenShot:

```
create table Yotube_New (id STRING,uploader STRING,inter INT, category STRING, length INT, nviews INT, rating FLOAT, nrating INT, ncomment INT, rvideos ARRAY<STRING>)
row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfile LOCATION '/user/praveenkum79edu/YouTube_Analysis';
```

INFO : Compiling command(queryId=hive_20211121064348_7c49cf98-6a83-47f9-8d41-90e42ea26d89): create table Yotube_New (id STRING,uploader STRING,inter INT, category STRING, length INT, nviews INT, rating FLOAT, nrating INT, ncomment INT, rvideos ARRAY<STRING>) row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfile LOCATION '/user/praveenkum79edu/YouTube_Analysis'

INFO : Semantic Analysis Completed

INFO : Returning Hive schema: Schema(fieldsSchemas:null, properties:null)

INFO : Completed compiling command(queryId=hive_20211121064348_7c49cf98-6a83-47f9-8d41-90e42ea26d89): Time taken: 0.015 seconds

INFO : Executing command(queryId=hive_20211121064348_7c49cf98-6a83-47f9-8d41-90e42ea26d89): create table Yotube_New (id STRING,uploader STRING,inter INT, category STRING, length INT, nviews INT, rating FLOAT, nrating INT, ncomment INT, rvideos ARRAY<STRING>) row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfile LOCATION '/user/praveenkum79edu/YouTube_Analysis'

INFO : Starting task [Stage-0:DDL] in serial mode

INFO : Completed executing command(queryId=hive_20211121064348_7c49cf98-6a83-47f9-8d41-90e42ea26d89): Time taken: 0.056 seconds

INFO : OK

✓ Success.

Query History Saved Queries

a few seconds ago country (or) territory having highest Airports create table Yotube_New (id STRING,uploader STRING,inter INT, category STRING, length INT, nviews INT, rating FLOAT, nrating INT, ncomment INT, rvideos ARRAY<STRING>) row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfile LOCATION '/user/praveenkum79edu/YouTube_Analysis'

Loading data:

LOAD DATA INPATH '/user/praveenkum79edu/YouTube_Data/youtubedata.txt' OVERWRITE INTO TABLE Yotube_New;

```
LOAD DATA INPATH '/user/praveenkum79edu/YouTube_Data/youtubedata.txt' OVERWRITE INTO TABLE Yotube_New;
```

INFO : Compiling command(queryId=hive_20211121064632_f954d5f3-1d22-4769-a2d0-8b9a8f6c1976): LOAD DATA INPATH '/user/praveenkum79edu/YouTube_Data/youtubedata.txt' OVERWRITE INTO TABLE Yotube_New

INFO : Semantic Analysis Completed

INFO : Returning Hive schema: Schema(fieldsSchemas:null, properties:null)

INFO : Completed compiling command(queryId=hive_20211121064632_f954d5f3-1d22-4769-a2d0-8b9a8f6c1976): Time taken: 0.044 seconds

INFO : Executing command(queryId=hive_20211121064632_f954d5f3-1d22-4769-a2d0-8b9a8f6c1976): LOAD DATA INPATH '/user/praveenkum79edu/YouTube_Data/youtubedata.txt' OVERWRITE INTO TABLE Yotube_New

INFO : Starting task [Stage-0:MOVE] in serial mode

INFO : Loading data to table default.yotube_new from hdfs://nameservice1/user/praveenkum79edu/YouTube_Data/youtubedata.txt

INFO : Starting task [Stage-1:STATS] in serial mode

INFO : Completed executing command(queryId=hive_20211121064632_f954d5f3-1d22-4769-a2d0-8b9a8f6c1976): Time taken: 0.623 seconds

INFO : OK

✓ Success.

Query History Saved Queries

a few seconds ago country (or) territory having highest Airports LOAD DATA INPATH '/user/praveenkum79edu/YouTube_Data/youtubedata.txt' OVERWRITE INTO TABLE Yotube_New

Problem Statements:



Find out the top 5 categories with maximum number of videos uploaded.

Solution:

SELECT category AS category,COUNT(id) AS counting FROM Yotube_New GROUP BY category SORT BY counting DESC LIMIT 5

Hive Screenshot:

The screenshot shows a Hive query execution plan and results. The query is: `SELECT category AS category,COUNT(id) AS counting FROM Youtube_New GROUP BY category SORT BY counting DESC LIMIT 5`. The execution plan shows the query is executed in a single stage with 1 reducer. The results are as follows:

category	counting
1 Entertainment	908
2 Music	862
3 Comedy	414
4 People & Blogs	398
5 News & Politics	333

Find out the top 10 rated videos

Solution:

select * from Youtube_New ORDER By rating DESC limit 10;

Hive Screenshot:

The screenshot shows a Hive query execution plan and results. The query is: `select * from Youtube_New ORDER By rating DESC limit 10`. The execution plan shows the query is executed in a single stage with 1 reducer. The results are as follows:

youtube_new.id	youtube_new.uploader	youtube_new.inter	youtube_new.category	youtube_new.length	youtube_new.views	youtube_new.rating	youtube_new.ratin
1 g6Uf_u0r0	samsongade	1108	Comedy	218	564	5	12
2 oegndu3d9s	glenPANDER	1135	News & Politics	132	597	5	4
3 w20y42z24	ELma200n	1082	Entertainment	466	283	5	2
4 dhdF1AY3uI	PfRaandd	1060	Entertainment	249	146	5	1
5 L3m8qyHkg	DarkArcho102	1068	Entertainment	84	109	5	1
6 VZw7imouX8	threefrstnames	1135	Entertainment	253	63	5	3

Find out the most viewed videos.

Solution:

select id, category, rating, nviews
from Youtube_New
order by nviews desc limit 1;

Hive Screenshot:

Hive
Add a name... Add a description...

25.41s Database default = Type text =
Type text =
?

```

3 select id, category, rating, mvies
2 from Youtube_Mov
3 order by mvies desc limit 1;
4

```

```

INFO : Starting Job = job_1635139249191_4505, Tracking URL = http://ip-10-1-1-204.ap-south-1-compute.internal:6066/proxy/application_1635139249191_4505/
INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p8.1425774/11b/hadoop/bin/hadoop job -kill job_1635139249191_4505
INFO : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
INFO : 2021-11-21 07:17:29,745 Stage-1 map = 0%, reduce = 0%
INFO : 2021-11-21 07:17:38,844 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.93 sec
INFO : 2021-11-21 07:17:44,285 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.51 sec
INFO : MapReduce Total cumulative CPU time: 4 seconds 518 msec
INFO : Ended Job = job_1635139249191_4505
INFO : MapReduce Jobs Launched:
INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.51 sec HDFS Read: 978695 HDFS Write: 142 HDFS EC Read: 0 SUCCESS
INFO : Total MapReduce CPU Time Spent: 4 seconds 518 msec
INFO : Completed executing command(query=hive_20211121071728_90eefde-bc2c-4d9b-812b-cc58836d8f9): Time taken: 25.407 seconds
INFO : OK

```

Query History
Saved Queries
Results (1)

--

	id	category	rating	mvies
1	1223J1uz8Q0	Film & Animation	3.03	65341925