

StackOverFlow Data Analysis

Creating and Loading data into tables

Creating comments Table:

```
CREATE TABLE comment (id int, userid int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_Comment_Analysis';
LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/comments.csv' OVERWRITE INTO TABLE comment;
```

The screenshot shows the Hive CLI interface. At the top, there's a header with the Hive logo, a search icon, and a description field. Below this, the SQL commands are entered in a text area. The commands are: 1. CREATE TABLE comment (id int, userid int); 2. ROW FORMAT DELIMITED FIELDS TERMINATED BY ','; 3. LINES TERMINATED BY '\n'; 4. STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_Comment_Analysis'; 5. LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/comments.csv' OVERWRITE INTO TABLE comment;. Below the text area, the execution results are displayed. The results show the compilation of the command, the semantic analysis completed, the Hive schema returned, the command execution, the task starting, the data loading, and the task completion. The final result is 'INFO : OK'. At the bottom, there's a 'Success.' message and a 'Query History' section showing the executed queries.

```
1 CREATE TABLE comment (id int, userid int)
2 ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
3 LINES TERMINATED BY '\n'
4 STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_Comment_Analysis';
5 LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/comments.csv' OVERWRITE INTO TABLE comment;
```

INFO : Compiling command(queryId=hive_20211121165348_0806fe4e-8899-4b7f-9a09-a842212119bb): LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/comments.csv' OVERWRITE INTO TABLE comment
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20211121165348_0806fe4e-8899-4b7f-9a09-a842212119bb); Time taken: 0.043 seconds
INFO : Executing command(queryId=hive_20211121165348_0806fe4e-8899-4b7f-9a09-a842212119bb): LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/comments.csv' OVERWRITE INTO TABLE comment
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table default.comment from hdfs://nameservice1/user/praveenkum79edu/SatckOverFlow_Data/comments.csv
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Completed executing command(queryId=hive_20211121165348_0806fe4e-8899-4b7f-9a09-a842212119bb); Time taken: 0.604 seconds
INFO : OK

✓ Success.

Query History

Time	Query	Command
a few seconds ago	movies released each year	LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/comments.csv' OVERWRITE INTO TABLE comment
a few seconds ago	movies released each year	CREATE TABLE comment (id int, userid int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_Comment_Analysis'

Creating Post Table:

```
CREATE TABLE post (id int, post_type String, creationdate String, score int, viewcount int, owneruserid int, titile String,
answercount int, commentcount int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_post_Analysis' ;
```

```
LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/posts.csv' OVERWRITE INTO TABLE post;
```

Hive movies released each year Add a description...

0.69s Database default Type text ?

```

1 CREATE TABLE post (id int, post_type String, creationdate String, score int, viewcount int, owneruserid int, title String, answercount int, commentcount int)
2 ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
3 LINES TERMINATED BY '\n'
4 STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_post_Analysis' ;
5
6
7 LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/posts.csv' OVERWRITE INTO TABLE post;
8
9

```

INFO : Compiling command(queryId=hive_20211121165510_51f49479-7c71-4dbd-b6ac-3bc3a6b1eb35): LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/posts.csv' OVERWRITE INTO TABLE post
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20211121165510_51f49479-7c71-4dbd-b6ac-3bc3a6b1eb35); Time taken: 0.833 seconds
INFO : Executing command(queryId=hive_20211121165510_51f49479-7c71-4dbd-b6ac-3bc3a6b1eb35): LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/posts.csv' OVERWRITE INTO TABLE post
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table default.post from hdfs://nameservice1/user/praveenkum79edu/SatckOverFlow_Data/posts.csv
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Completed executing command(queryId=hive_20211121165510_51f49479-7c71-4dbd-b6ac-3bc3a6b1eb35); Time taken: 0.647 seconds
INFO : OK

✓ Success.

Query History	Saved Queries
a few seconds ago	movies released each year
a few seconds ago	movies released each year

Creating Posttype Table:

```

CREATE TABLE posttype (id int, name String)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_PostType_Analysis' ;

```

LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/posttypes.csv' OVERWRITE INTO TABLE posttype;

Hive movies released each year Add a description...

1.43s Database default Type text ?

```

1 CREATE TABLE posttype (id int, name String)
2 ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
3 LINES TERMINATED BY '\n'
4 STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_PostType_Analysis' ;
5
6
7 LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/posttypes.csv' OVERWRITE INTO TABLE posttype;
8
9

```

INFO : Compiling command(queryId=hive_20211121165637_18a2b5e2-621d-45d2-8a8a-bc6822f9568f): LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/posttypes.csv' OVERWRITE INTO TABLE posttype
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20211121165637_18a2b5e2-621d-45d2-8a8a-bc6822f9568f); Time taken: 0.833 seconds
INFO : Executing command(queryId=hive_20211121165637_18a2b5e2-621d-45d2-8a8a-bc6822f9568f): LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/posttypes.csv' OVERWRITE INTO TABLE posttype
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table default.posttype from hdfs://nameservice1/user/praveenkum79edu/SatckOverFlow_Data/posttypes.csv
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Completed executing command(queryId=hive_20211121165637_18a2b5e2-621d-45d2-8a8a-bc6822f9568f); Time taken: 0.683 seconds
INFO : OK

✓ Success.

Query History	Saved Queries
a few seconds ago	movies released each year
a few seconds ago	movies released each year

Creating users Table:

```

CREATE TABLE users_N (id int, reputation int, displayname String, loc String, age int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverFlow_Users_Analysis' ;

```

LOAD DATA INPATH '/user/praveenkum79edu/SatckOverFlow_Data/users.csv' OVERWRITE INTO TABLE users_N;

Hive movies released each year Add a description...

0.88s Database default Type text

```

1 CREATE TABLE users_N (id int, reputation int, displayname String, loc String, age int)
2 ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
3 LINES TERMINATED BY '\n'
4 STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverflow_Users_Analysis';
5
6 LOAD DATA INPATH '/user/praveenkum79edu/StackOverflow_Data/users.csv' OVERWRITE INTO TABLE users_N;
7

```

INFO : Compiling command(queryId=hive_20211121165817_b258339d-2bc8-4eba-9139-7049c8fcb2d): LOAD DATA INPATH '/user/praveenkum79edu/StackOverflow_Data/users.csv' OVERWRITE INTO TABLE users_N

INFO : Semantic Analysis Completed

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

INFO : Completed compiling command(queryId=hive_20211121165817_b258339d-2bc8-4eba-9139-7049c8fcb2d); Time taken: 0.834 seconds

INFO : Executing command(queryId=hive_20211121165817_b258339d-2bc8-4eba-9139-7049c8fcb2d): LOAD DATA INPATH '/user/praveenkum79edu/StackOverflow_Data/users.csv' OVERWRITE INTO TABLE users_N

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

INFO : Loading data to table default.users_N from hdfs://nameservice1/user/praveenkum79edu/StackOverflow_Data/users.csv

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

INFO : Completed executing command(queryId=hive_20211121165817_b258339d-2bc8-4eba-9139-7049c8fcb2d); Time taken: 0.682 seconds

INFO : OK

✓ Success.

Query History Saved Queries

a few seconds ago	movies released each year	LOAD DATA INPATH '/user/praveenkum79edu/StackOverflow_Data/users.csv' OVERWRITE INTO TABLE users_N
a few seconds ago	movies released each year	CREATE TABLE users_N (id int, reputation int, displayname String, loc String, age int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS textfile LOCATION '/user/praveenkum79edu/StackOverflow_Users_Analysis'

Problem Statement

 Find the display name and no. of posts created by the user who has got maximum reputation

```

SELECT users_N.displayname, COUNT(post.id) FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE
users_N.reputation IN (SELECT MAX(reputation) FROM users_N) GROUP BY users_N.displayname;

```

Hive movies released each year Add a description...

2m, 0s Database default Type text

```

1 SELECT users_N.displayname, COUNT(post.id) FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE users_N.reputation IN (SELECT MAX(reputation) FROM users_N) GROUP BY users_N.displayname;
2

```

INFO : set hive.exec.reducers.max=<number>

INFO : In order to set a constant number of reducers:

INFO : set mapreduce.job.reducers=<number>

INFO : number of splits:2

INFO : Submitting tokens for job: job_1635139249191_4667

INFO : Executing with tokens: []

INFO : The url to track the job: http://ip-10-1-1-204.ap-south-1.compute.internal:6866/proxy/application_1635139249191_4667/

INFO : Starting Job = job_1635139249191_4667, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6866/proxy/application_1635139249191_4667/

INFO : Kill Command = /opt/clooudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/11b/hadoop/bin/hadoop job -kill job_1635139249191_4667

INFO : Hadoop job information for Stage-2: number of mappers: 2; number of reducers: 1

INFO : 2021-11-21 17:00:50,912 Stage-2 map = 0%, reduce = 0%

INFO : 2021-11-21 17:01:01,235 Stage-2 map = 50%, reduce = 0%, Cumulative CPU 1.68 sec

INFO : 2021-11-21 17:01:02,257 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 5.71 sec

INFO : 2021-11-21 17:01:10,418 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 9.17 sec

INFO : 2021-11-21 17:01:10,418 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 9.17 sec

Query History Saved Queries Results (1)

users_n.displayname	_c1
1 Jon Skeet	1

 Find the average age of users on the Stack Overflow site.

```

SELECT CAST(AVG(age) AS INT) FROM users_N;

```

Hive movies released each year Add a description...

27.26s Database default Type text

```

1 | SELECT CAST(AVG(age) AS INT) FROM users_N;

```

INFO : Compiling command(queryId=hive_20211121170232_8e24e069-2579-46e2-b56c-c3cc6868ddec): SELECT CAST(AVG(age) AS INT) FROM users_N

INFO : Semantic Analysis Completed

INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:_c0, type:int, comment:null)], properties:null)

INFO : Completed compiling command(queryId=hive_20211121170232_8e24e069-2579-46e2-b56c-c3cc6868ddec); Time taken: 0.062 seconds

INFO : Executing command(queryId=hive_20211121170232_8e24e069-2579-46e2-b56c-c3cc6868ddec): SELECT CAST(AVG(age) AS INT) FROM users_N

WARN :

INFO : Query ID = hive_20211121170232_8e24e069-2579-46e2-b56c-c3cc6868ddec

INFO : Total jobs = 1

INFO : Launching Job 1 out of 1

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

INFO : Number of reduce tasks determined at compile time: 1

INFO : In order to change the average load for a reducer (in bytes):

INFO : set hive.exec.reducers.bytes.per.reducer=<number>

Query History Saved Queries Results (1)

	_c0
1	35

Find the display name of user who posted the oldest post on Stack Overflow (in terms of date).

```

SELECT users_N.displayname, post.creationdate FROM users_N JOIN post ON
users_N.id=post.owneruserid WHERE post.creationdate IN (SELECT creationdate FROM post ORDER BY
creationdate ASC LIMIT 1);

```

Hive movies released each year Add a description...

1m,43s Database default Type text

```

1 | SELECT users_N.displayname, post.creationdate FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.creationdate IN (SELECT creationdate FROM post ORDER BY creationdate ASC LIMIT 1);

```

INFO : Compiling command(queryId=hive_20211121170358_bd978f6b-7b0e-4594-913e-1b07e26dd9cd): SELECT users_N.displayname, post.creationdate FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.creationdate IN (SELECT creationdate FROM post ORDER BY creationdate ASC LIMIT 1)

INFO : Semantic Analysis Completed

INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:users_n.displayname, type:string, comment:null), FieldSchema(name:post.creationdate, type:string, comment:null)], properties:null)

INFO : Completed compiling command(queryId=hive_20211121170358_bd978f6b-7b0e-4594-913e-1b07e26dd9cd); Time taken: 0.1 seconds

INFO : Executing command(queryId=hive_20211121170358_bd978f6b-7b0e-4594-913e-1b07e26dd9cd): SELECT users_N.displayname, post.creationdate FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.creationdate IN (SELECT creationdate FROM post ORDER BY creationdate ASC LIMIT 1)

WARN :

INFO : Query ID = hive_20211121170358_bd978f6b-7b0e-4594-913e-1b07e26dd9cd

INFO : Total jobs = 3

INFO : Launching Job 1 out of 3

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

INFO : Number of reduce tasks not specified. Estimated from input data size: 1

Query History Saved Queries Results (1)

	users_n.displayname	post.creationdate
1	Eggs McLaren	2008-07-31 21:42:52

Find the display name and no. of comments done by the user who has got maximum reputation

```

SELECT users_N.displayname, COUNT(comment.id) FROM users_N JOIN comment ON users_N.id=comment.userid
WHERE users_N.reputation IN (SELECT MAX(reputation) FROM users_N) GROUP BY users_N.displayname;

```

Hive movies released each year Add a description...

2m, 0s Database default Type text

```
1 SELECT users_n.displayname, COUNT(comment.id) FROM users_n JOIN comment ON users_n.id=comment.userid WHERE users_n.reputation IN (SELECT MAX(reputation) FROM users_n) GROUP BY users_n.displayname;
```

INFO : set hive.exec.reducers.bytes.per.reducer=<number>
 INFO : In order to limit the maximum number of reducers:
 INFO : set hive.exec.reducers.max=<number>
 INFO : In order to set a constant number of reducers:
 INFO : set mapreduce.job.reduces=<number>
 INFO : number of splits:1
 INFO : Submitting tokens for job: job_1635139249191_4676
 INFO : Executing with tokens: []
 INFO : The url to track the job: http://ip-10-1-1-284.ap-south-1.compute.internal:6066/proxy/application_1635139249191_4676/
 INFO : Starting Job = job_1635139249191_4676, Tracking URL = http://ip-10-1-1-284.ap-south-1.compute.internal:6066/proxy/application_1635139249191_4676/
 INFO : Kill Command = jopt/clouders/parceia/CDH-6.2.1-1.odh6.2.1.p0.1425774/1ib/hadoop/bin/hadoop job -kill job_1635139249191_4676
 INFO : Hadoop job information for Stage=3: number of mappers: 1; number of reducers: 1
 INFO : 2021-11-21 17:09:14,378 Stage=3 map = 0%, reduce = 0%
 INFO : 2021-11-21 17:09:15,630 Stage=3 map = 100%, reduce = 0% Cumulative CPU 1.08 sec

Query History Saved Queries Results (1)

users_n.displayname	_c1
1 Jon Skeet	765

Find the display name of user who has created maximum no. of posts on Stack Overflow.

SELECT users_N.displayname, users_N.id FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.owneruserid IN (SELECT temp.owneruserid FROM (SELECT owneruserid, COUNT(id) AS cnt FROM post GROUP BY owneruserid ORDER BY cnt DESC LIMIT 1) temp) group by users_N.displayname, users_N.id;

Hive movies released each year Add a description...

2m, 30s Database default Type text

```
1 SELECT users_N.displayname, users_N.id FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.owneruserid IN (SELECT temp.owneruserid FROM (SELECT owneruserid, COUNT(id) AS cnt FROM post GROUP BY owneruserid ORDER BY cnt DESC LIMIT 1) temp) group by users_N.displayname, users_N.id;
```

INFO : Compiling command(queryId=hive_20211121171113_f6468523-d385-475f-84e2-57ab83b50e7e): SELECT users_N.displayname, users_N.id FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.owneruserid IN (SELECT temp.owneruserid FROM (SELECT owneruserid, COUNT(id) AS cnt FROM post GROUP BY owneruserid ORDER BY cnt DESC LIMIT 1) temp) group by users_N.displayname, users_N.id;
 INFO : Semantic Analysis Completed
 INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name=users_n.displayname, type:string, comment:null), FieldSchema(name=users_n.id, type:int, comment:null)], properties:null)
 INFO : Completed compiling command(queryId=hive_20211121171113_f6468523-d385-475f-84e2-57ab83b50e7e); Time taken: 0.11 seconds
 INFO : Executing command(queryId=hive_20211121171113_f6468523-d385-475f-84e2-57ab83b50e7e): SELECT users_N.displayname, users_N.id FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.owneruserid IN (SELECT temp.owneruserid FROM (SELECT owneruserid, COUNT(id) AS cnt FROM post GROUP BY owneruserid ORDER BY cnt DESC LIMIT 1) temp) group by users_N.displayname, users_N.id;
 WARN :
 INFO : Query ID = hive_20211121171113_f6468523-d385-475f-84e2-57ab83b50e7e
 INFO : Total jobs = 5
 INFO : Launching Job 1 out of 5
 INFO : Starting task [Stage=1:MAPRED] in serial mode
 INFO : Number of reduce tasks not specified. Estimated from input data size: 1

Query History Saved Queries Results (1)

users_n.displayname	users_n.id
1 aku	1196

Find the owner name and id of user whose post has got maximum no. of view counts so far

SELECT users_N.displayname, users_N.id FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.owneruserid IN (SELECT temp.owneruserid FROM (SELECT owneruserid, viewcount FROM post ORDER BY viewcount DESC LIMIT 1) temp) group by users_N.displayname, users_N.id;

Hive movies released each year Add a description...

2m, 9s Database default Type text ?

```

1 SELECT users_N.displayName, users_N.id FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.owneruserid IN (SELECT temp.owneruserid FROM (SELECT owneruserid, viewcount FROM post ORDER BY viewcount DESC LIMIT 1) temp) group by post.title, users_N.displayName;

```

INFO : Starting Job = job_1635139249191_4685, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6866/proxy/application_1635139249191_4685/

INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1635139249191_4685

INFO : Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1

INFO : 2021-11-21 17:15:30,643 Stage-3 map = 0%, reduce = 0%

INFO : 2021-11-21 17:15:45,788 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.88 sec

INFO : 2021-11-21 17:15:52,950 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 4.65 sec

INFO : MapReduce Total cumulative CPU time: 4 seconds 650 msec

INFO : Ended Job = job_1635139249191_4685

INFO : MapReduce Jobs Launched:

INFO : Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 4.65 sec HDFS Read: 5549 HDFS Write: 114 HDFS EC Read: 0 SUCCESS

INFO : Total MapReduce CPU Time Spent: 4 seconds 650 msec

INFO : Completed executing command(queryId=hive_20211121171348_c81e3408-7cbf-4a92-bac3-1e5d73739321); Time taken: 124.935 seconds

INFO : OK

Query History Saved Queries Results (1)

	users_n.displayName	users_n.id
1	Shadow_x99	244

Find the title and owner name of post who has got maximum no. of Comment count

SELECT post.title, users_N.displayName FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.commentcount IN (SELECT temp.commentcount FROM (SELECT title, owneruserid, commentcount FROM post order by commentcount DESC LIMIT 1) temp) group by post.title, users_N.displayName;

Hive movies released each year Add a description...

2m, 2s Database default Type text ?

```

1 SELECT post.title, users_N.displayName FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.commentcount IN (SELECT temp.commentcount FROM (SELECT title, owneruserid, commentcount FROM post order by commentcount DESC LIMIT 1) temp) group by post.title, users_N.displayName;

```

INFO : Compiling command(queryId=hive_20211121171531_60116bda-bce7-4b9b-a314-abe04d3e01b8): SELECT post.title, users_N.displayName FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.commentcount IN (SELECT temp.commentcount FROM (SELECT title, owneruserid, commentcount FROM post order by commentcount DESC LIMIT 1) temp) group by post.title, users_N.displayName

INFO : Semantic Analysis Completed

INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:post.title, type:string, comment:null), FieldSchema(name:users_n.displayName, type:string, comment:null)], properties:[])

INFO : Completed compiling command(queryId=hive_20211121171531_60116bda-bce7-4b9b-a314-abe04d3e01b8); Time taken: 0.095 seconds

INFO : Executing command(queryId=hive_20211121171531_60116bda-bce7-4b9b-a314-abe04d3e01b8): SELECT post.title, users_N.displayName FROM users_N JOIN post ON users_N.id=post.owneruserid WHERE post.commentcount IN (SELECT temp.commentcount FROM (SELECT title, owneruserid, commentcount FROM post order by commentcount DESC LIMIT 1) temp) group by post.title, users_N.displayName

MAP :

INFO : Query ID = hive_20211121171531_60116bda-bce7-4b9b-a314-abe04d3e01b8

INFO : Total jobs = 4

INFO : Launching Job 1 out of 4

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

INFO : Number of reduce tasks not specified. Estimated from input data size: 1

Query History Saved Queries Results (1)

	post.title	users_n.displayName
1		Justin Standard

Find the location which has maximum no of Stack Overflow users

SELECT tmpusr.loc FROM (SELECT loc, COUNT(id) AS cid FROM users_N WHERE loc != '' GROUP BY loc ORDER BY cid DESC LIMIT 1) tmpusr;

movies released each year
Add a description...

1m, 6s
Database default
Type text
?

```

1 SELECT tmpusr.loc FROM (SELECT loc, COUNT(id) AS cid FROM users_N WHERE loc != '' GROUP BY loc ORDER BY cid DESC LIMIT 1) tmpusr;
2

```

INFO : Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

INFO : 2021-11-21 17:16:36,052 Stage-2 map = 0%, reduce = 0%

INFO : 2021-11-21 17:16:44,401 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.54 sec

INFO : 2021-11-21 17:17:00,739 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.2 sec

INFO : MapReduce Total cumulative CPU time: 5 seconds 200 msec

INFO : Ended Job = job_1635139249191_4690

INFO : MapReduce Jobs Launched:

INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.35 sec HDFS Read: 1567849 HDFS Write: 120321 HDFS EC Read: 0 SUCCESS

INFO : Stage-Stage-5: Map: 1 Reduce: 1 Cumulative CPU: 6.89 sec HDFS Read: 2507802 HDFS Write: 114 HDFS EC Read: 0 SUCCESS

INFO : Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.2 sec HDFS Read: 125931 HDFS Write: 113 HDFS EC Read: 0 SUCCESS

INFO : Total MapReduce CPU Time Spent: 19 seconds 448 msec

INFO : Completed executing command(queryId=hive_20211121171555_55727f27-bafb-498d-ba7a-5a0a5511b84); Time taken: 66.416 seconds

INFO : OK

Query History
Saved Queries
Results (1)

tmpusr.loc	
1	United States



Find the total no. of answers, posts, comments created by Indian users.

SELECT COUNT(comment.id) AS Comments from comment JOIN users_N on comment.userid=users_N.id Where users_N.loc='India' UNION
 Select count(post.id) As Posts from post join users_N on post.owneruserid=users_N.id Where users_N.loc='India' UNION
 SELECT COUNT(tempptest.id) Answers FROM (SELECT id, owneruserid FROM post WHERE post_type IN (SELECT id FROM posttype WHERE name='Answer')) tempptest join users_N on tempptest.owneruserid=users_N.id Where users_N.loc='India';

movies released each year
Add a description...

4m, 21s
Database default
Type text
?

```

1 SELECT COUNT(comment.id) AS Comments from comment JOIN users_N on comment.userid=users_N.id Where users_N.loc='India' UNION
2 Select count(post.id) As Posts from post join users_N on post.owneruserid=users_N.id Where users_N.loc='India' UNION
3 SELECT COUNT(tempptest.id) Answers FROM (SELECT id, owneruserid FROM post WHERE post_type IN (SELECT id FROM posttype WHERE name='Answer')) tempptest join users_N on tempptest.owneruserid=users_N.id Where users_N.loc='India';
4

```

INFO : set mapreduce.job.reduces=<number>

INFO : number of splits:2

INFO : Submitting tokens for job: job_1635139249191_4704

INFO : Executing with tokens: []

INFO : The url to track the job: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1635139249191_4704/

INFO : Starting Job = job_1635139249191_4704, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1635139249191_4704/

INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1635139249191_4704

INFO : Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1

INFO : 2021-11-21 17:46:33,233 Stage-1 map = 0%, reduce = 0%

INFO : 2021-11-21 17:46:45,463 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 3.79 sec

INFO : 2021-11-21 17:46:46,485 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.79 sec

INFO : 2021-11-21 17:46:53,650 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 12.26 sec

INFO : MapReduce Total cumulative CPU time: 12 seconds 260 msec

INFO : Ended Job = job_1635139249191_4704

Query History
Saved Queries
Results (3)

_u3.comments	
1	31
2	62
3	150