1. As hive consumes the data which is loaded to it. First, took the backup of it in another newly created directory “/copied”.

**Code:**

hadoop fs -mkdir /copied

hadoop fs -cp /pig\_table/part-m-00000 /copied/

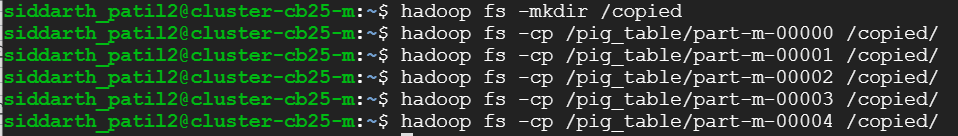
hadoop fs -cp /pig\_table/part-m-00001 /copied/

hadoop fs -cp /pig\_table/part-m-00002 /copied/

hadoop fs -cp /pig\_table/part-m-00003 /copied/

hadoop fs -cp /pig\_table/part-m-00004 /copied/

**Screenshot:**



1. To go to hive mode. Made use of **hive** command.
2. To make sure that hive prints the column name made use of the below command.

**Code:** set hive.cli.print.header=true;

1. Now, to load the data transformed by PIG in in the earlier part. Firstly, created a new databse **“part\_three”.** Created a new table in it **“part3\_posts”** and finally loaded the data to it.

**Code:** CREATE DATABASE part\_three;

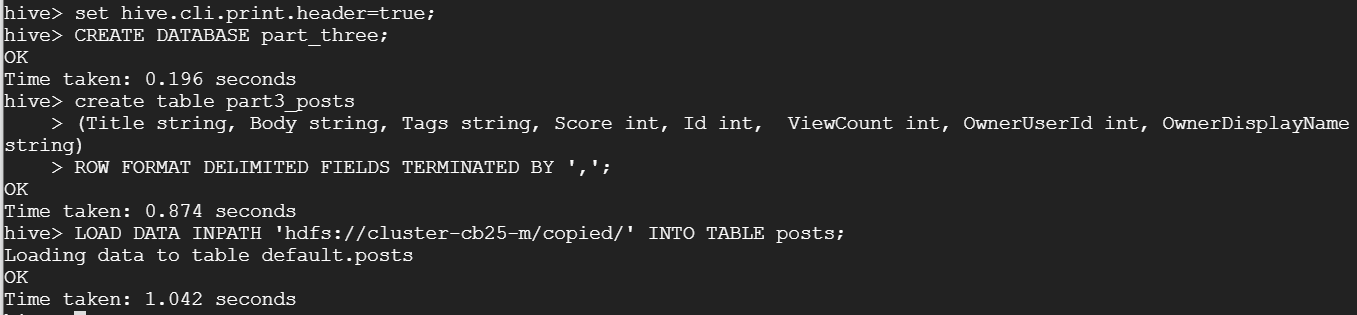
create table part3\_posts

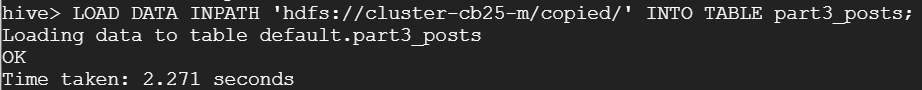
(Title string, Body string, Tags string, Score int, Id int, ViewCount int, OwnerUserId int, OwnerDisplayName string)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';

LOAD DATA INPATH 'hdfs://cluster-cb25-m/copied/' INTO TABLE part3\_posts;

**Screenshot:**



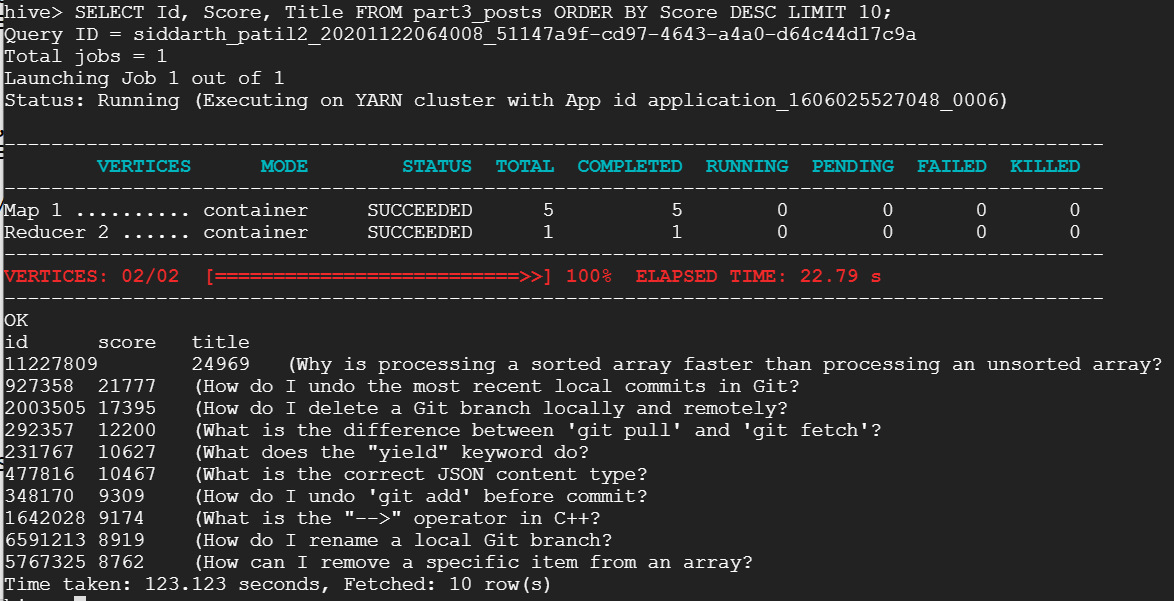


1. **To find the Top 10 posts by Score**:

**Code:** SELECT Id, Score, Title

FROM part3\_posts

ORDER BY Score DESC LIMIT 10;

**Screenshot:** 

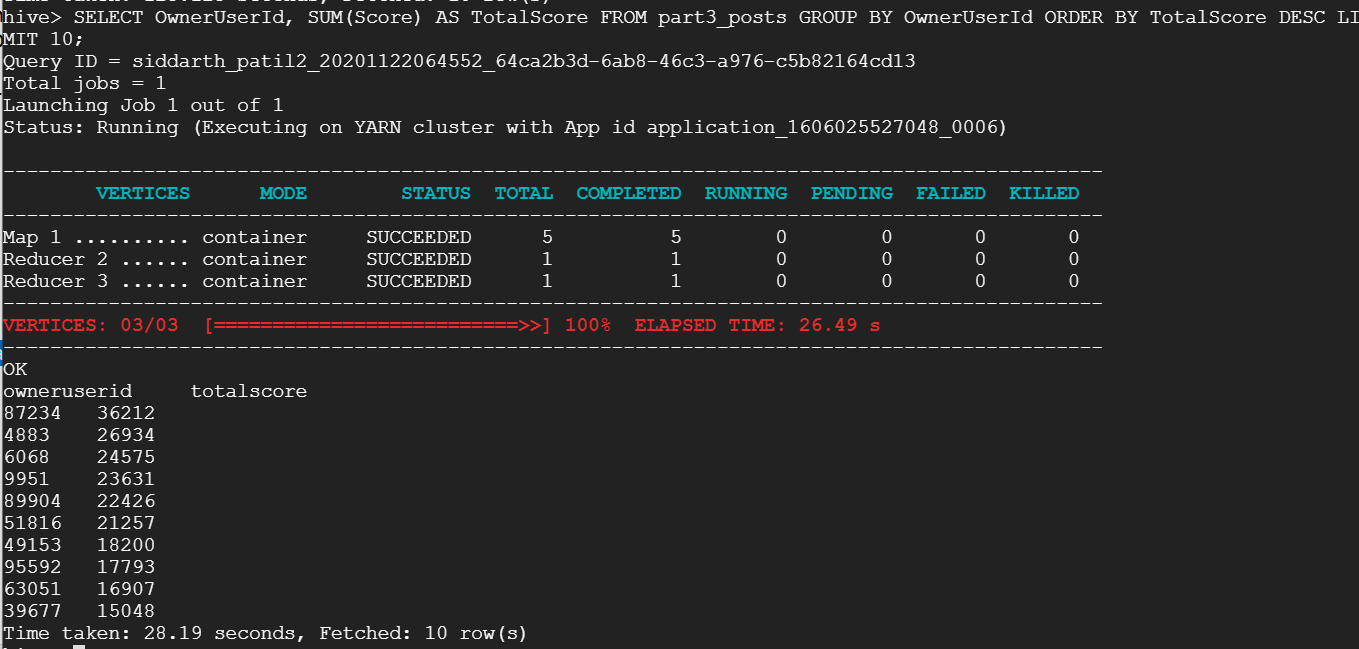
1. **To find Top 10 users by post score:**

**Code:** SELECT OwnerUserId, SUM(Score) AS TotalScore

FROM part3\_posts

GROUP BY OwnerUserId

ORDER BY TotalScore DESC LIMIT 10;

**Screenshot:** 

1. **Finding the number of unique users who used the work “Hadoop” in their posts**

**Code:** SELECT COUNT(DISTINCT OwnerUserId) AS Count\_User

FROM part3\_posts

WHERE (UPPER(Body) LIKE '%HADOOP%' OR UPPER(Title) LIKE '%HADOOP%' OR UPPER(Tags) LIKE '%HADOOP%');

**Screenshot:** 