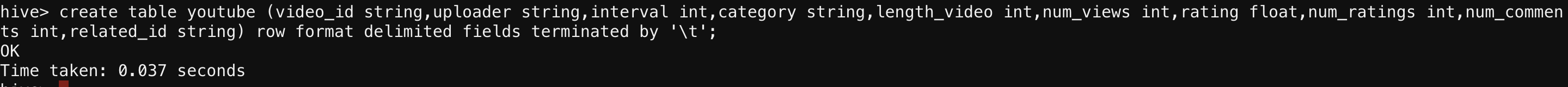
**Analysis of Youtube Data using Hadoop-Hive and PIG**

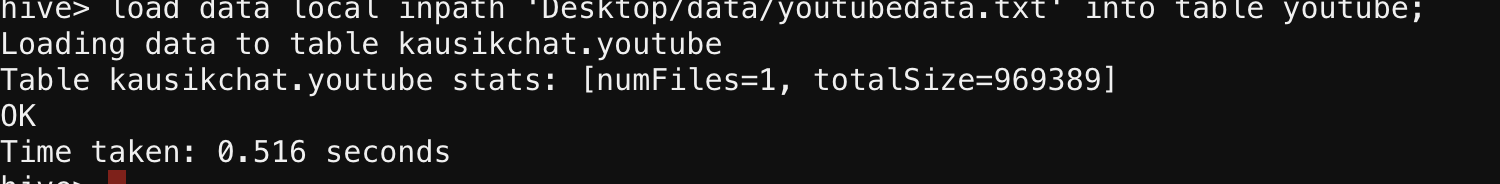
1. **creating table youtube for youtubedata.txt**:

create table youtube (video\_id string,uploader string,interval int,category string,length\_video int,num\_views int,rating float,num\_ratings int,num\_comments int,related\_id string) row format delimited fields terminated by '\t';



2. **loading data into youtube table**

load data local inpath 'Desktop/data/youtubedata.txt' into table youtube;

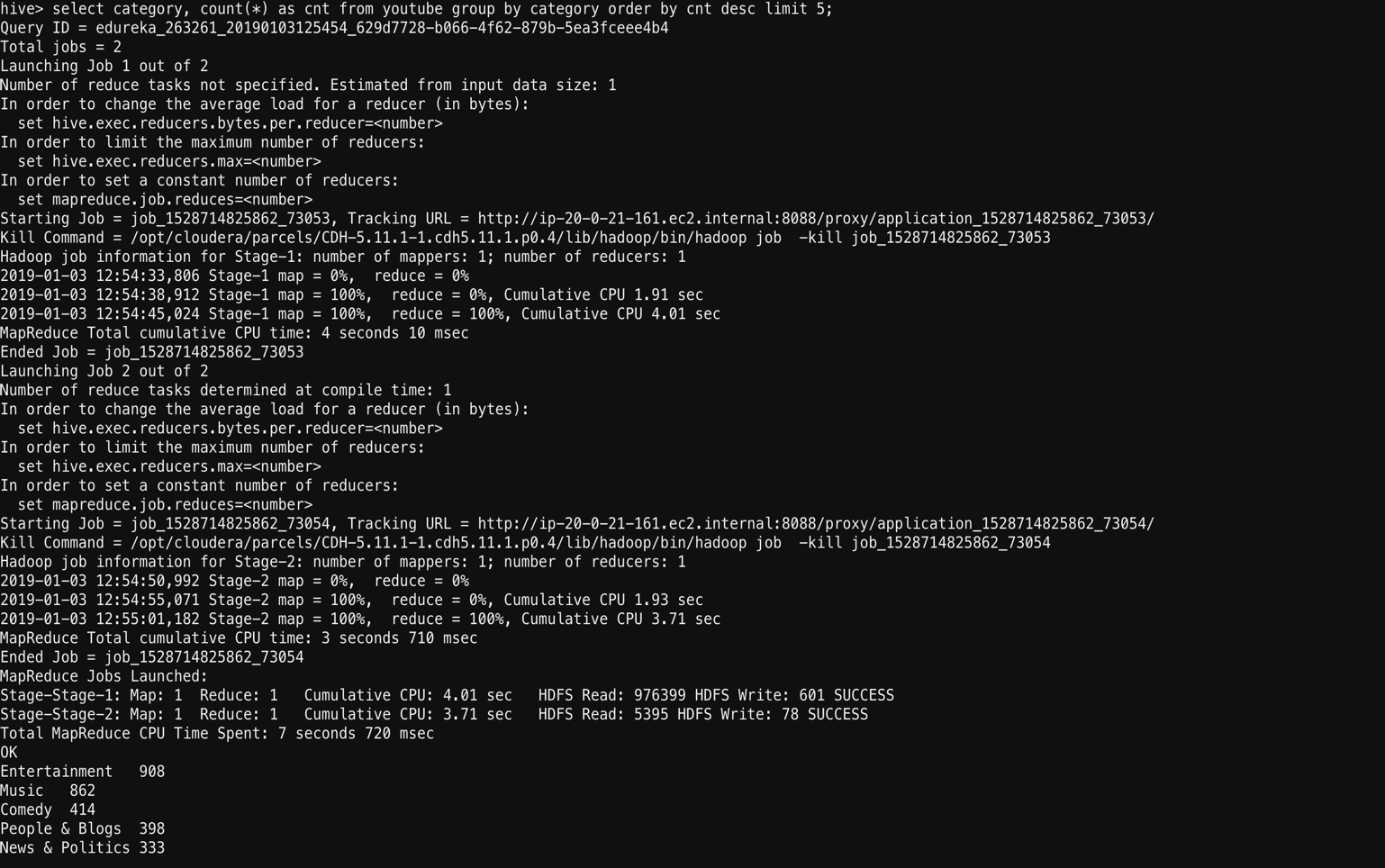


**Problem solution**:

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

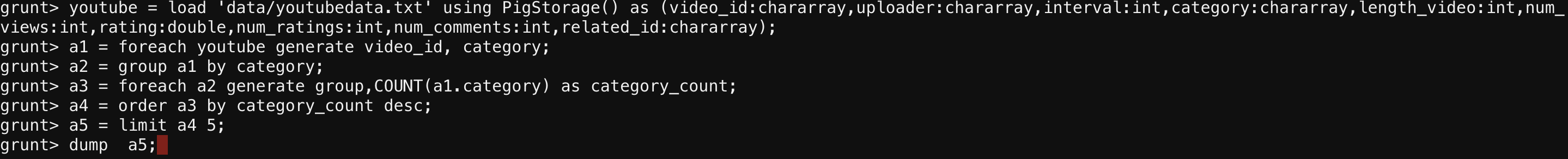
HIVE Solution:

select category, count(\*) as cnt from youtube group by category order by cnt desc limit 5;



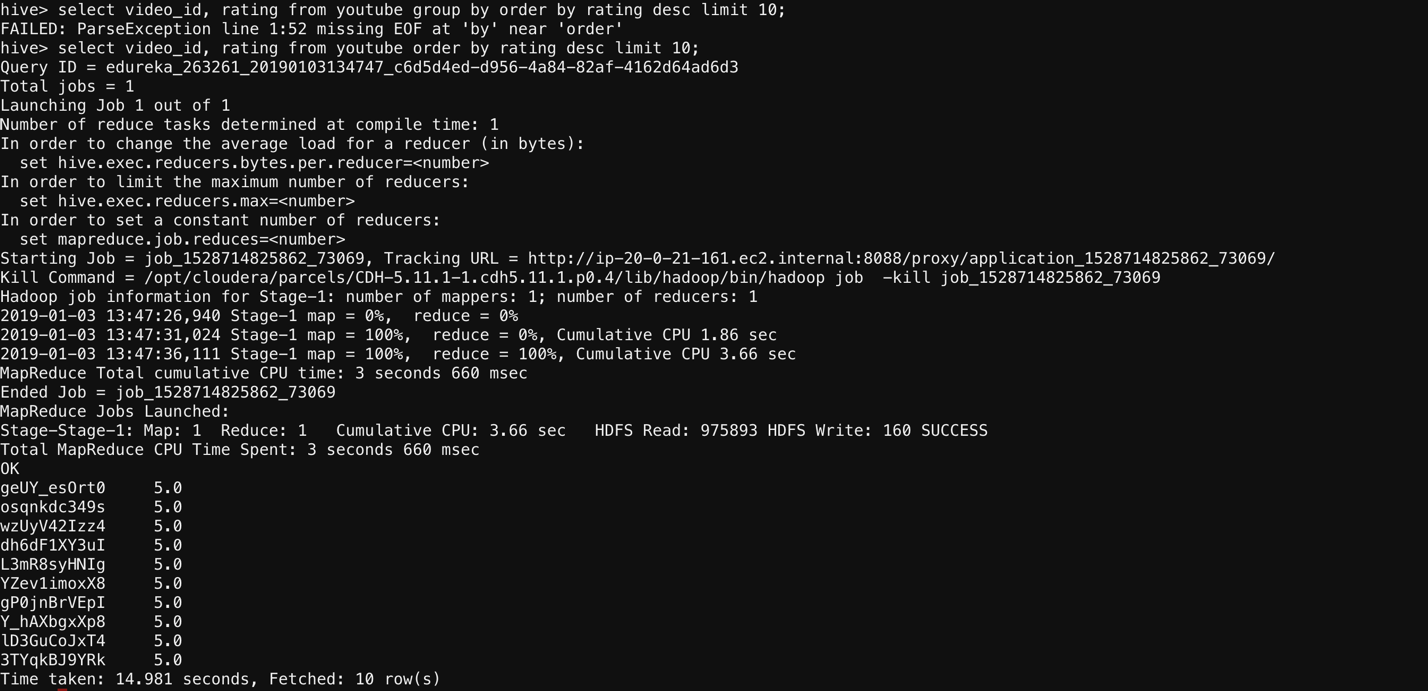


PIG Solution:

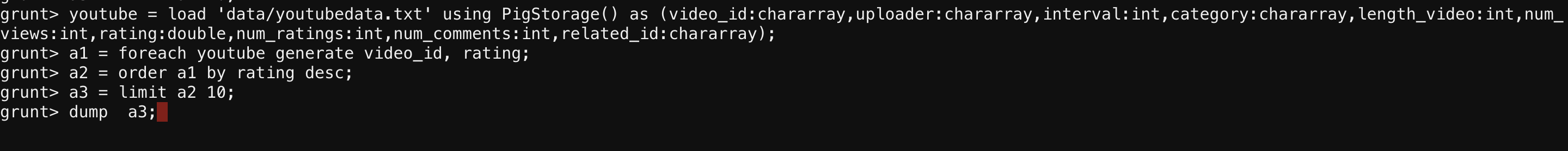


**B. Find out the top 10 rated videos.**

**HIVE Solution:**



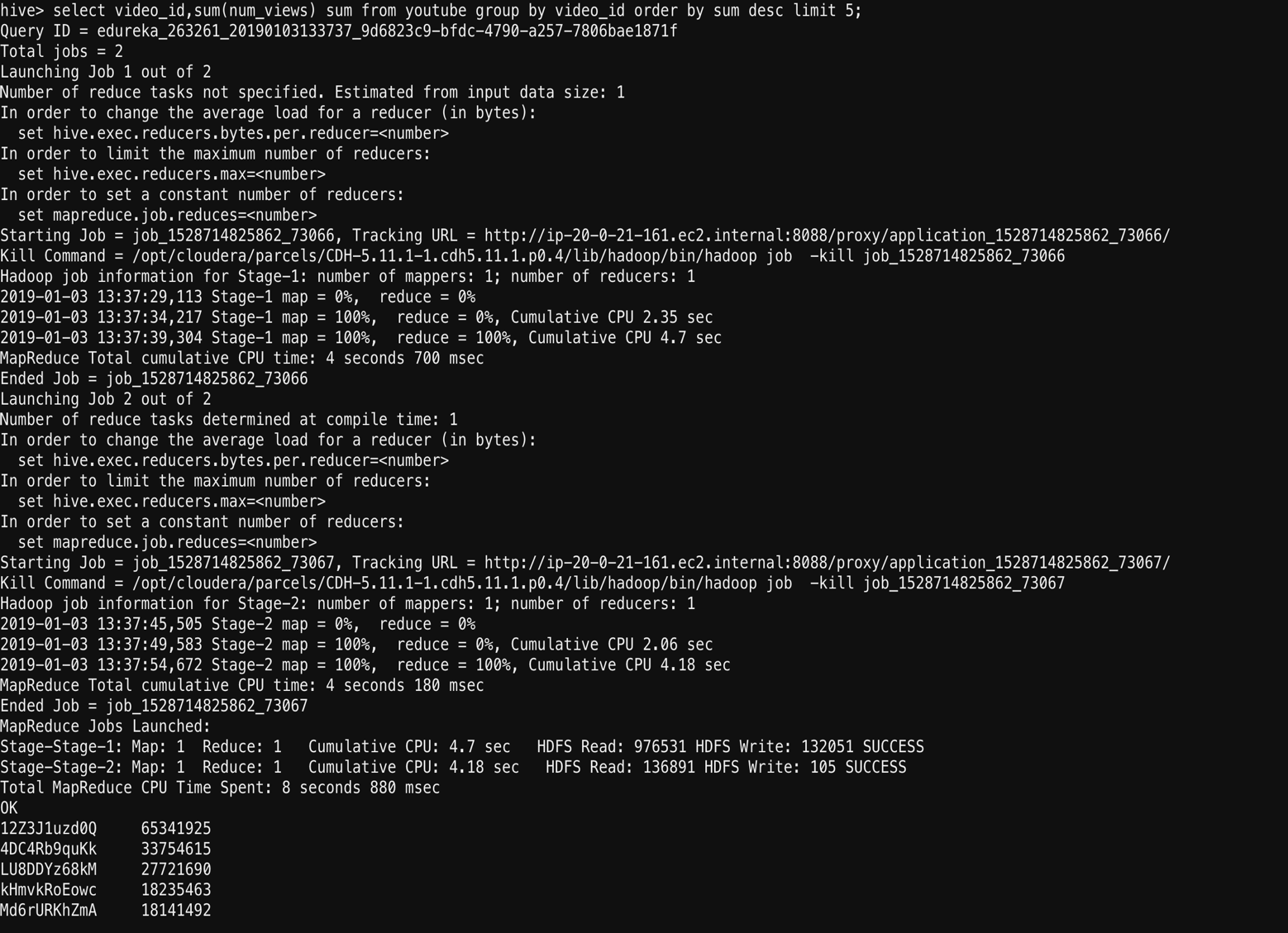
PIG Solution:



**C. Find out the most viewed videos.**

**HIVE Solution:**

select video\_id,sum(num\_views) sum from youtube group by video\_id order by sum desc limit 5;



**PIG Solution:**

