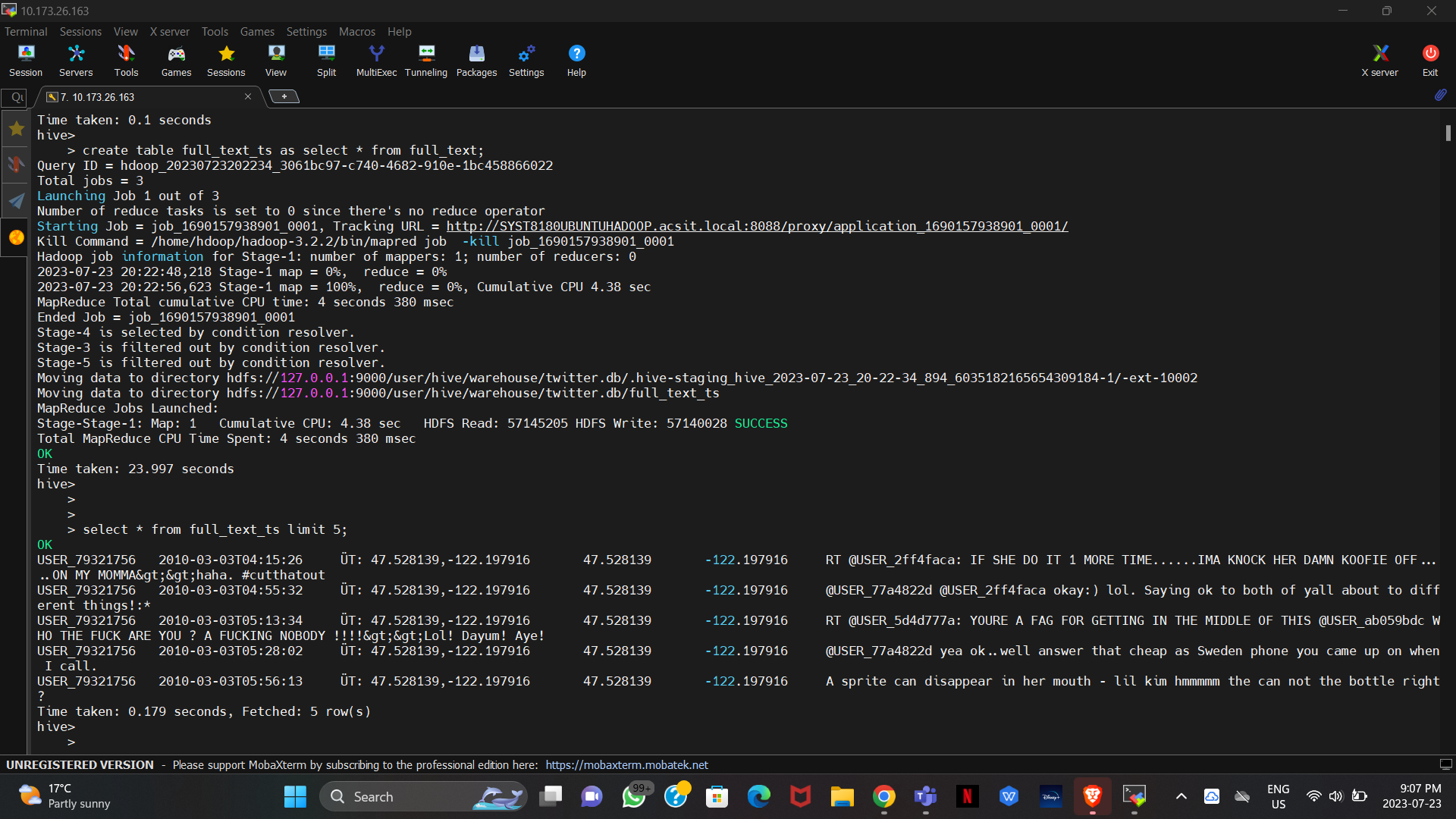
LAB

NIKITA DHINGRA

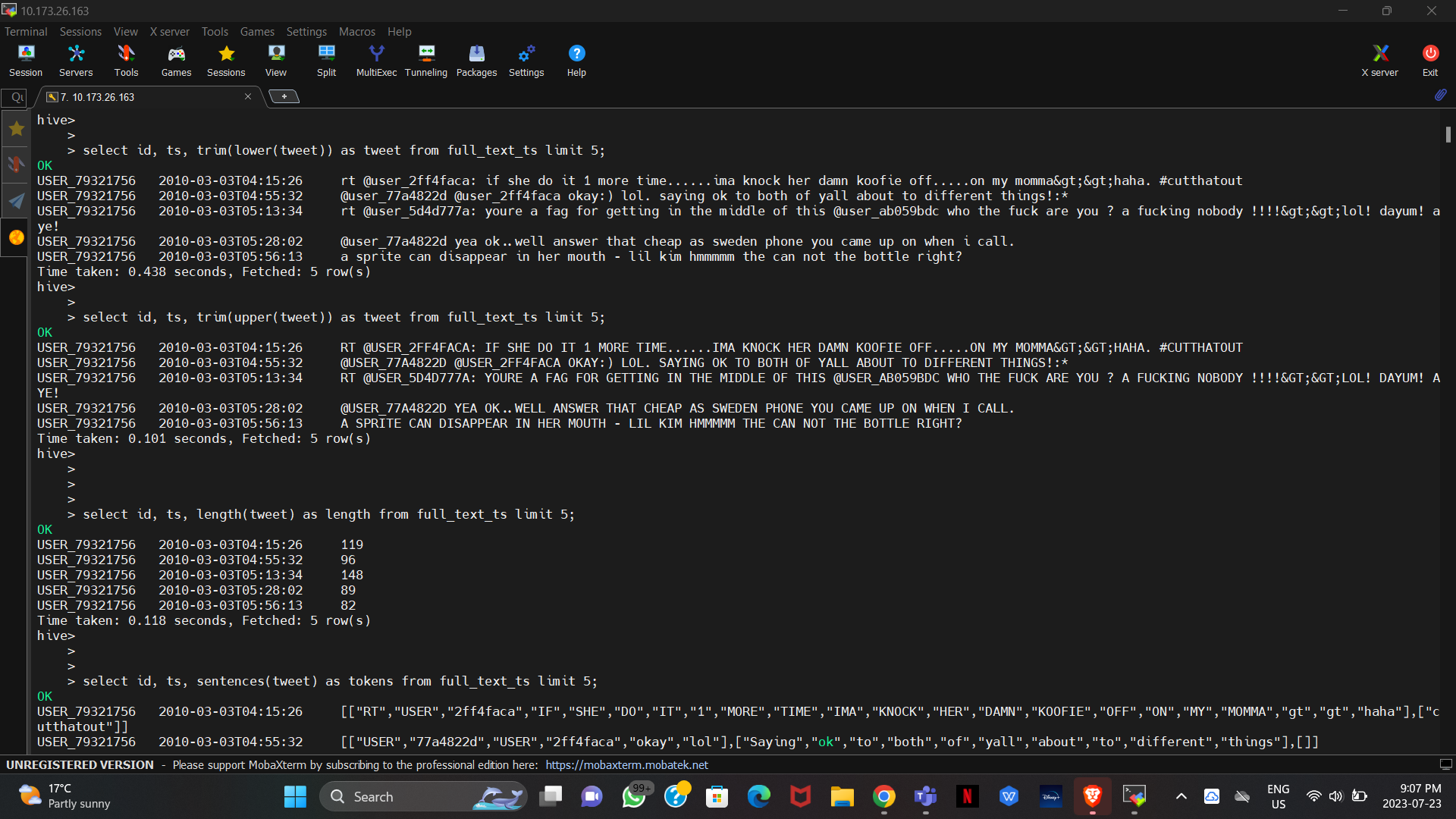
Workinh with queries in Hadoop and Hive

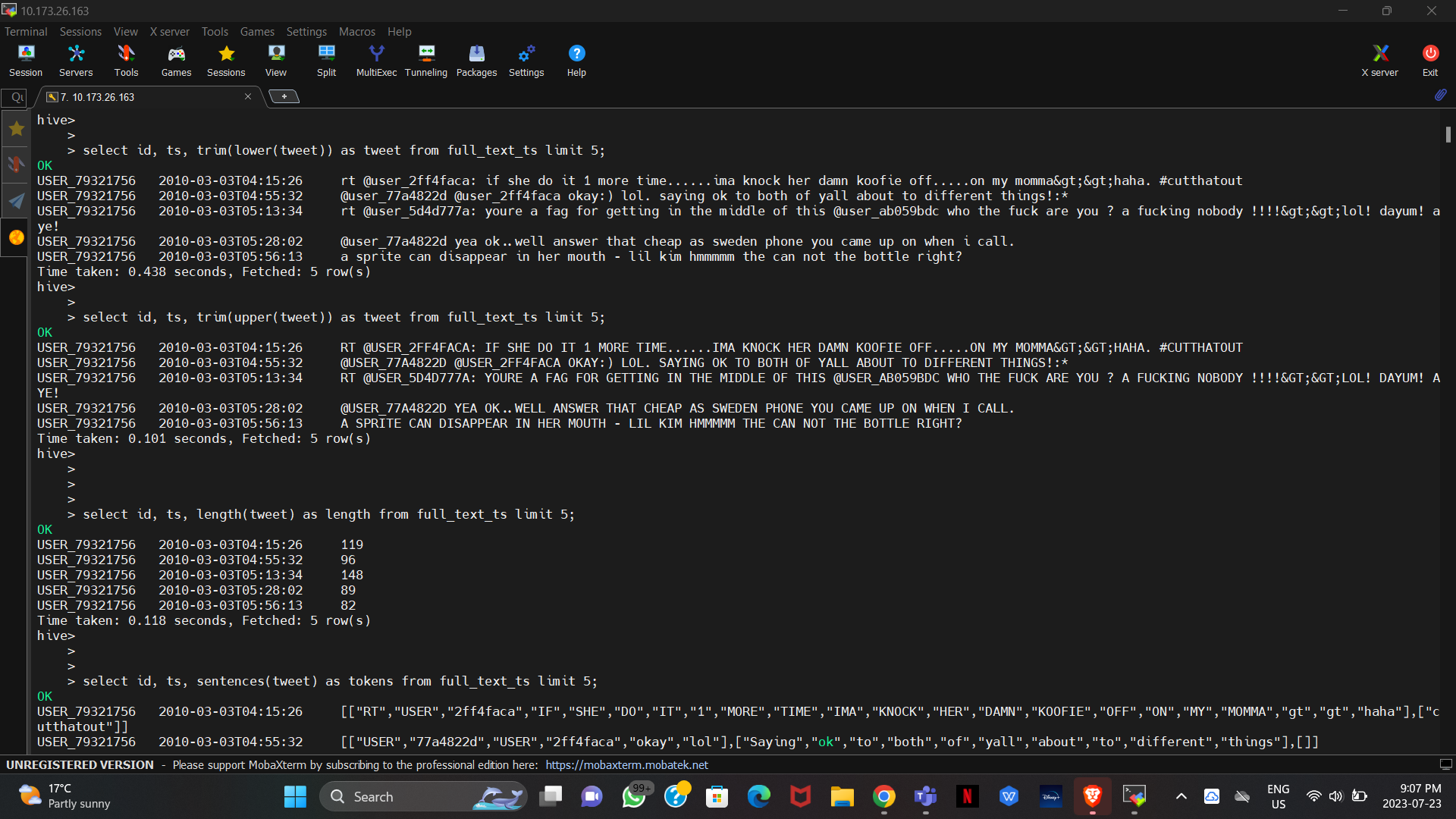
LAB 3

1. The below queries are executed in Hive and firstly we have created a table in the database of twitter and after using that database and \* represents select all from the table full\_test\_ts which is as full\_text.

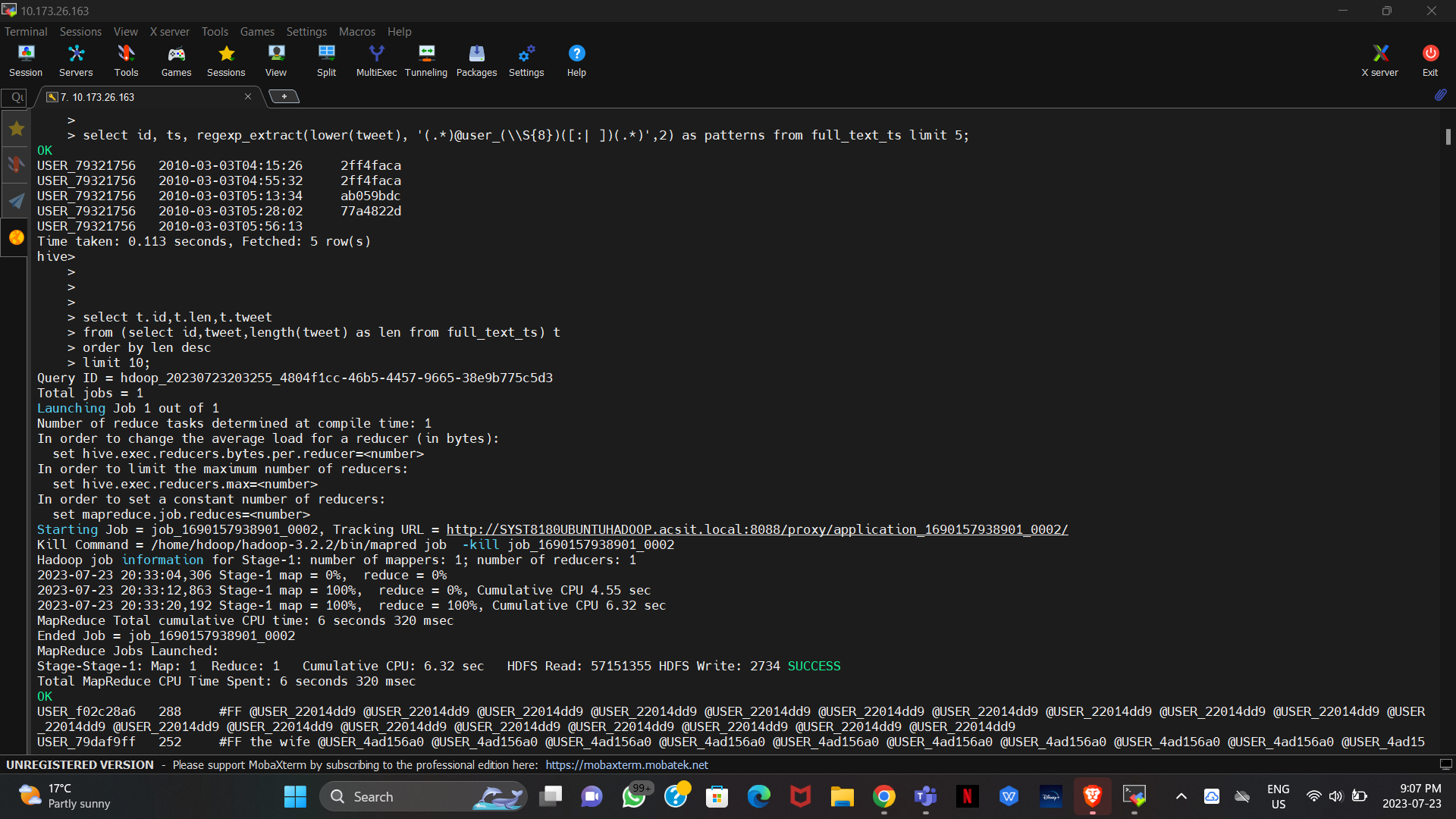


1. String Function: Below query gives the result of the query and get the data as in lower and upper cases , sentences as well as the length of the tweets and then limiting the result till 5 rows from above.



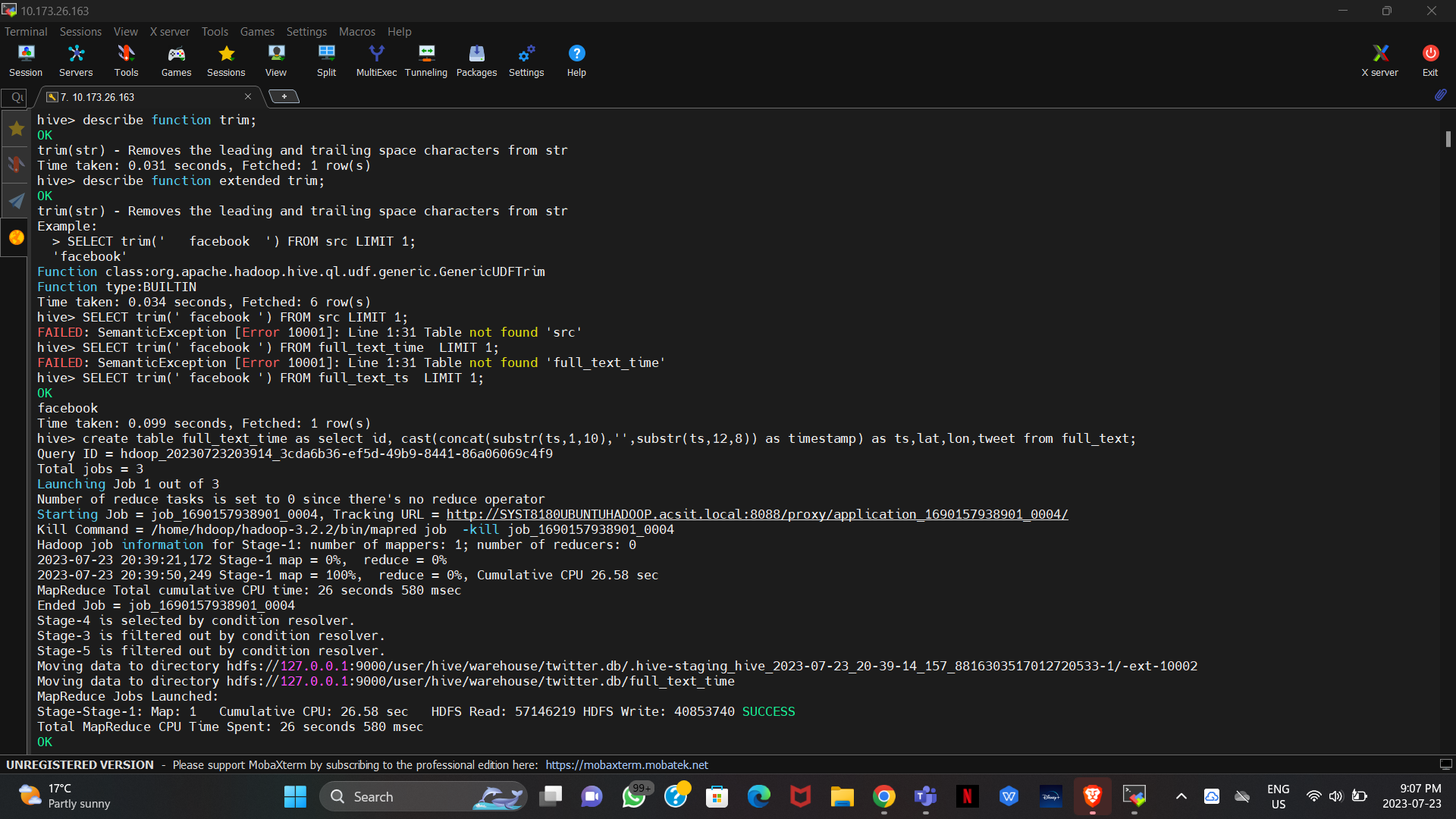


1. The below string function regexp\_extract is used to extract specific substrings from the output given based on the the expression we have given and again limit of 5 rows.

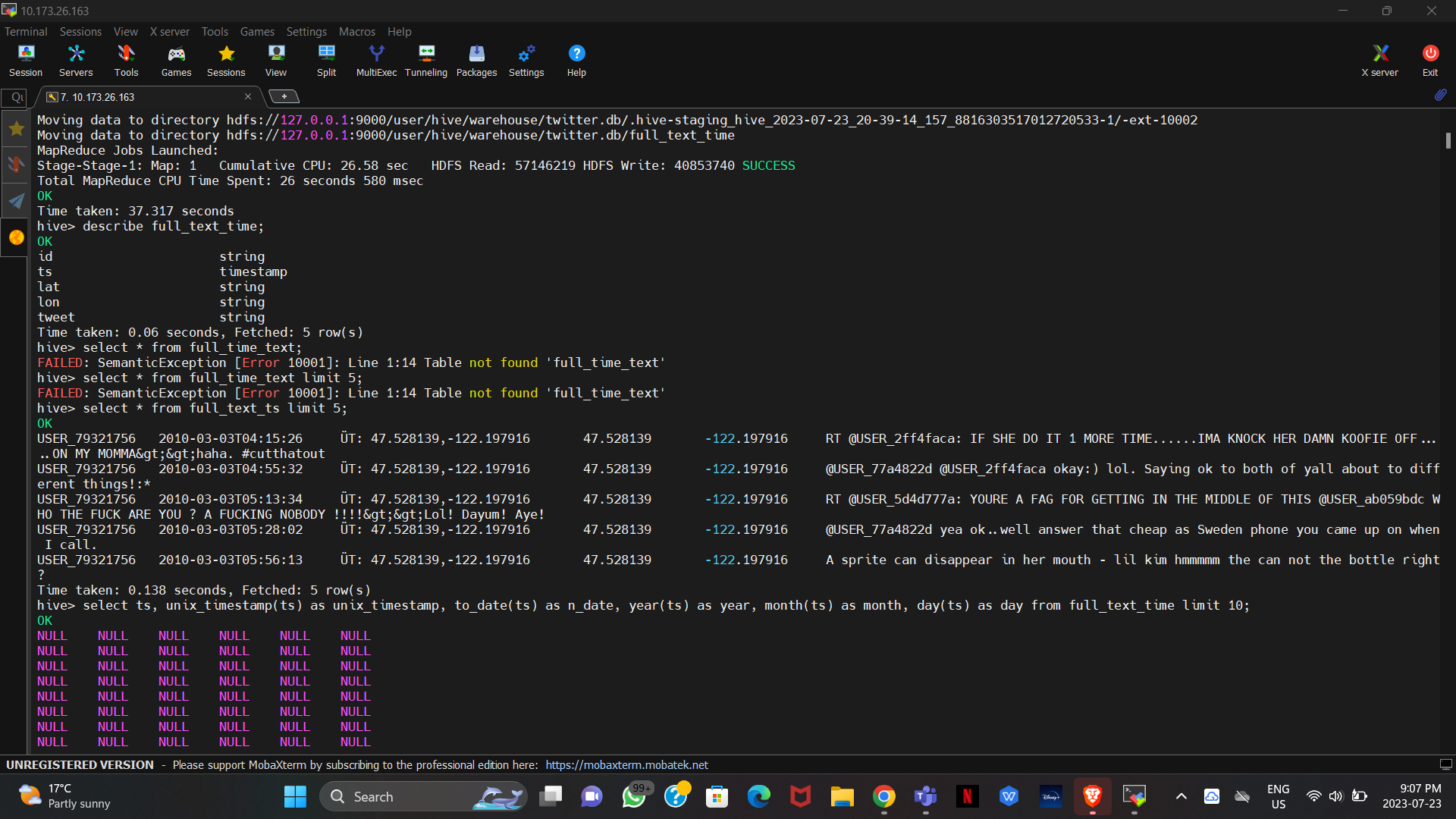


1. Below queries are of describing the fuction trim in hive while there is another query below that , which explains the coversion of a string type to a timestamp using data functions

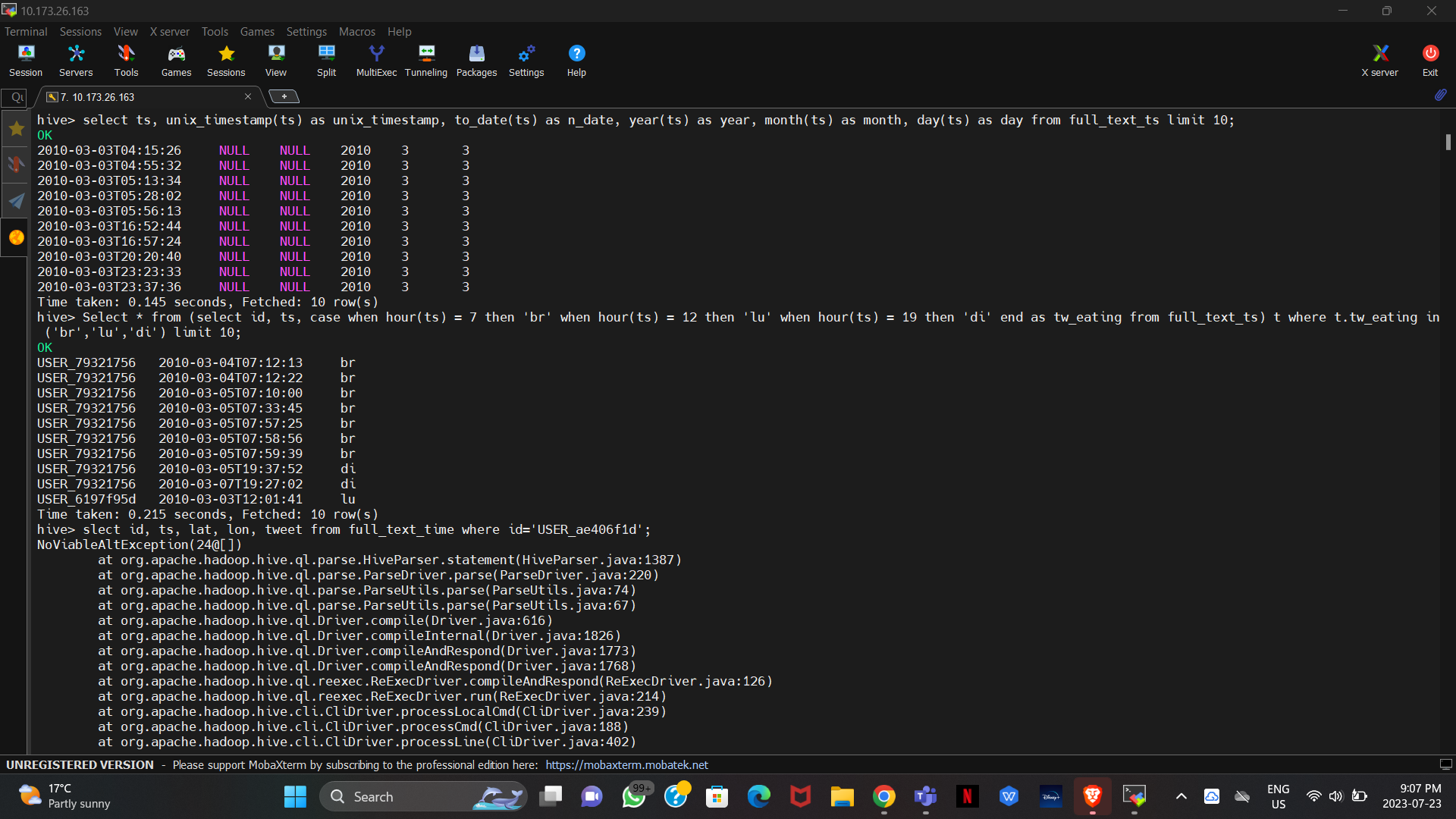
And casting the string's type into a timestamp using the query creates a table.



1. Data Functions for the extraction of Year, Month, and Date from timestamp extraction,in order to separate the dates, years, months, and days, use this query to extract the timestamp.

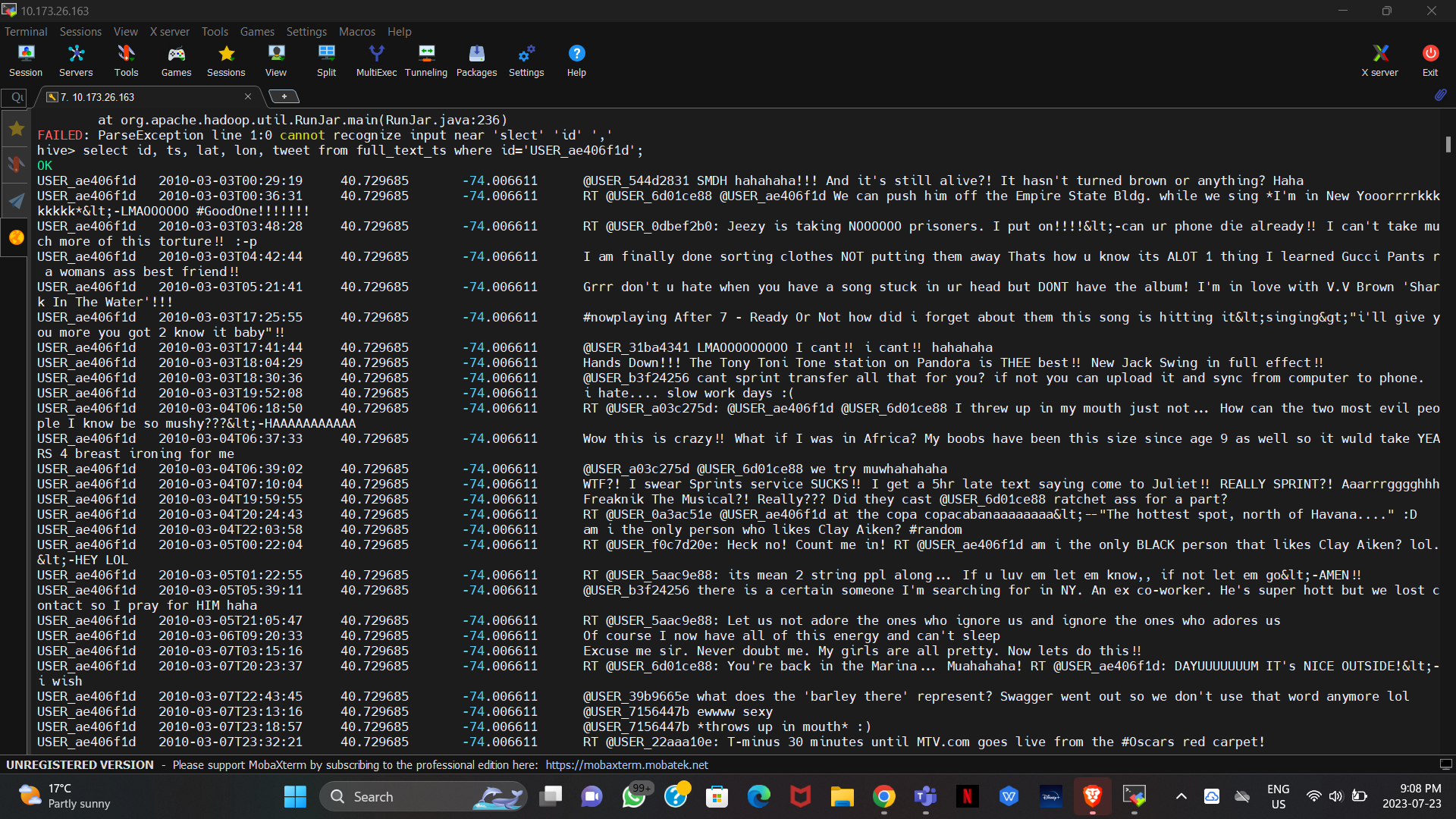


1. Conditional Function: Locate users that enjoy tw-eating and retrieving tweets using the tw\_eating with one more condition of in the BR,LU,DI and limiting the results to 10 requires this query.

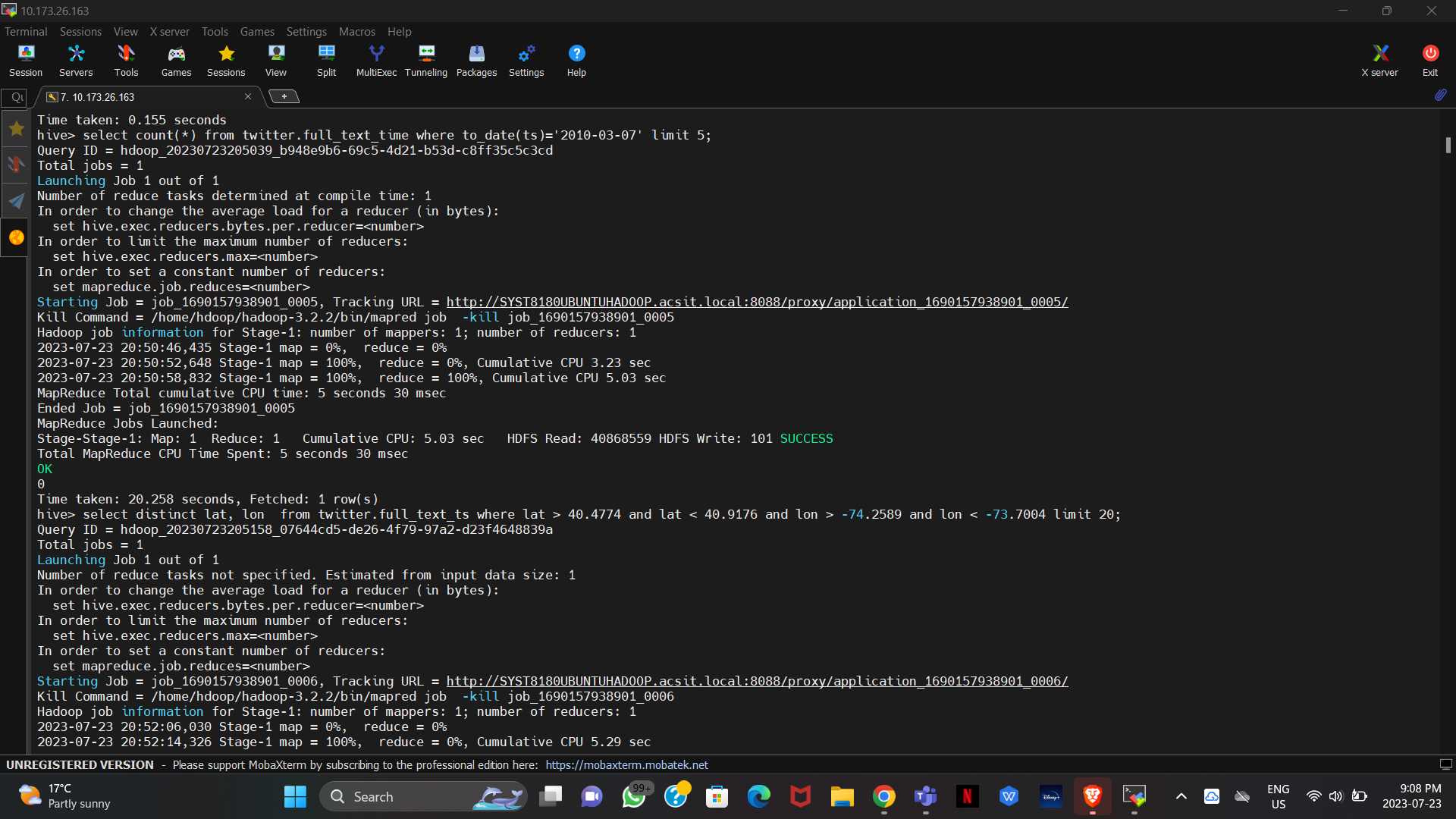


1. WHERE Clause: Search for all tweets from the user

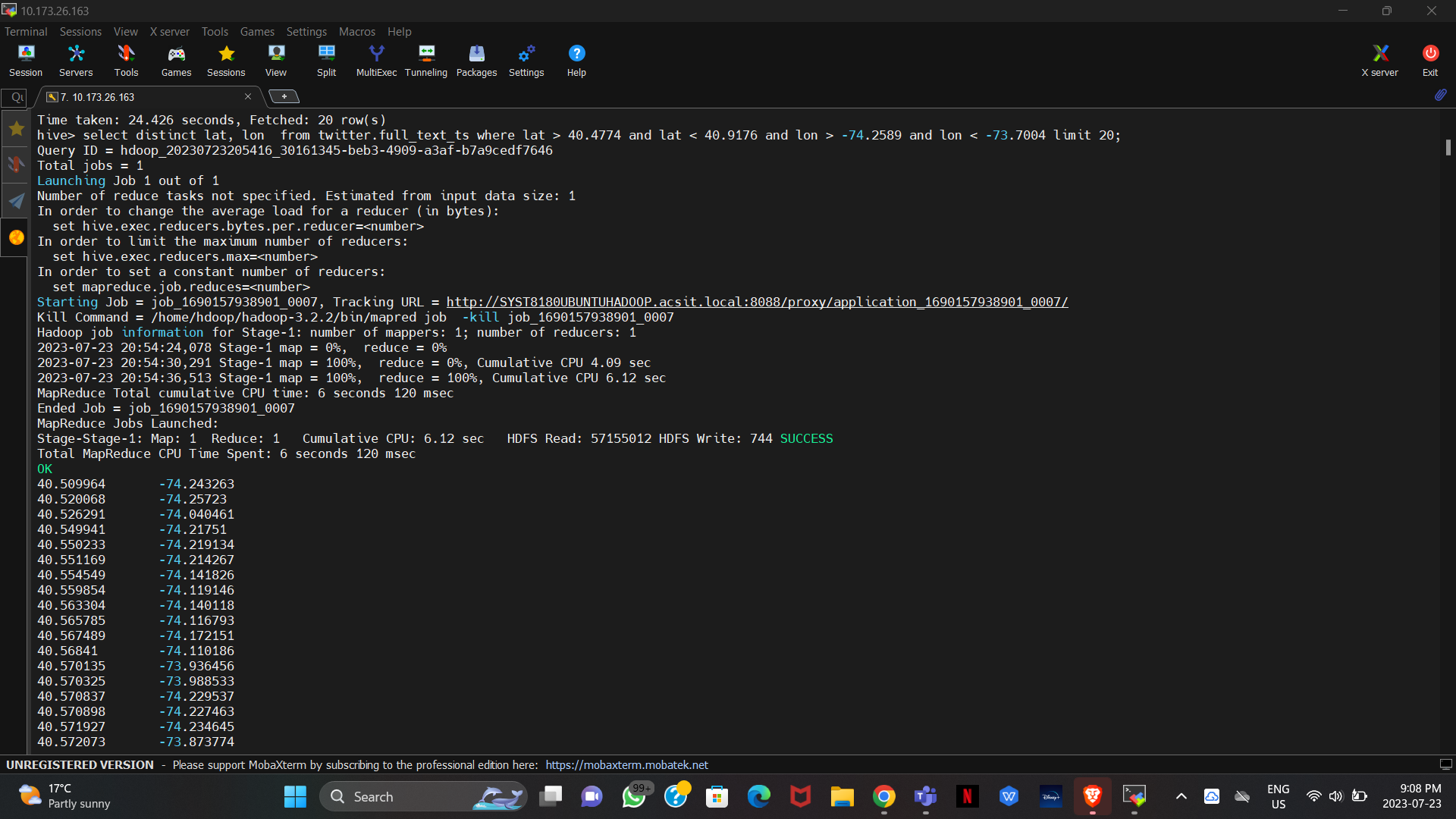
Basically to ge the condition where the id of user is user\_ae406f1d (for getting particular things), use this query and clause



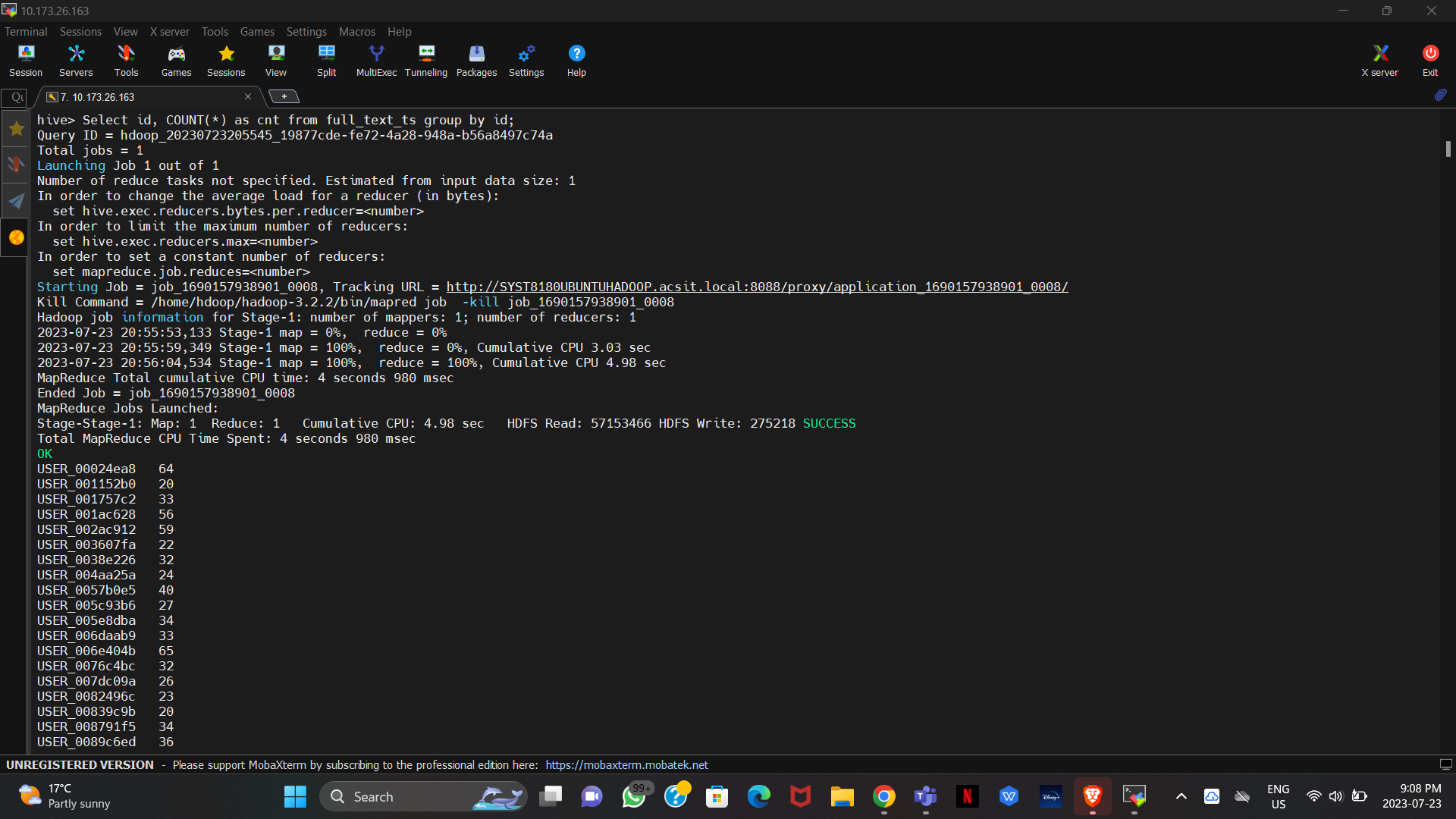
1. Where clause used to find the particular tweets where the date is 2010-03-07.



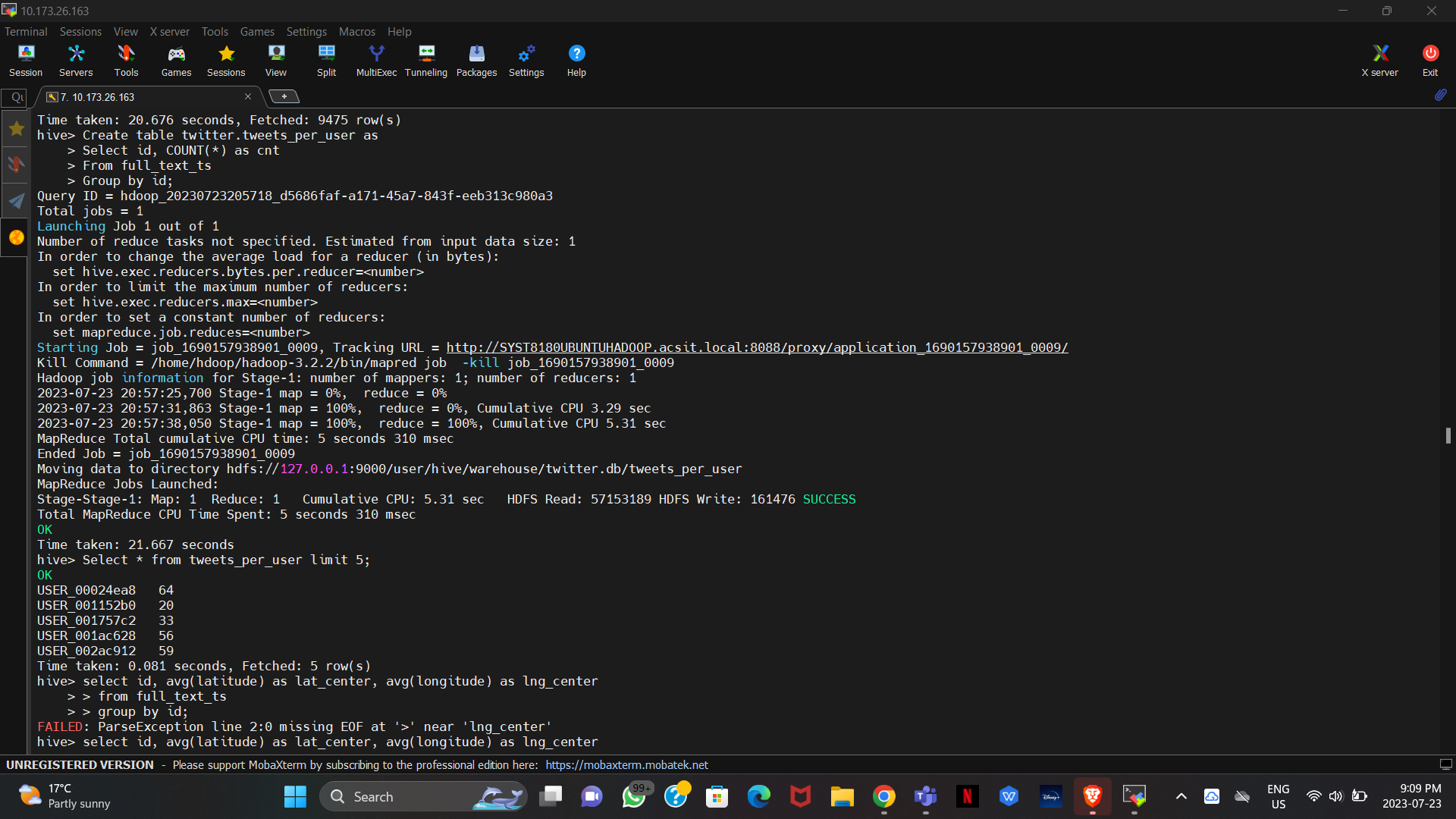
1. Where and Distinct Function : These query will give the result for getting data which is non-repeative and distinct with the clause of where ( the location latitude and longitude are specified) and then limit till 20 rows.

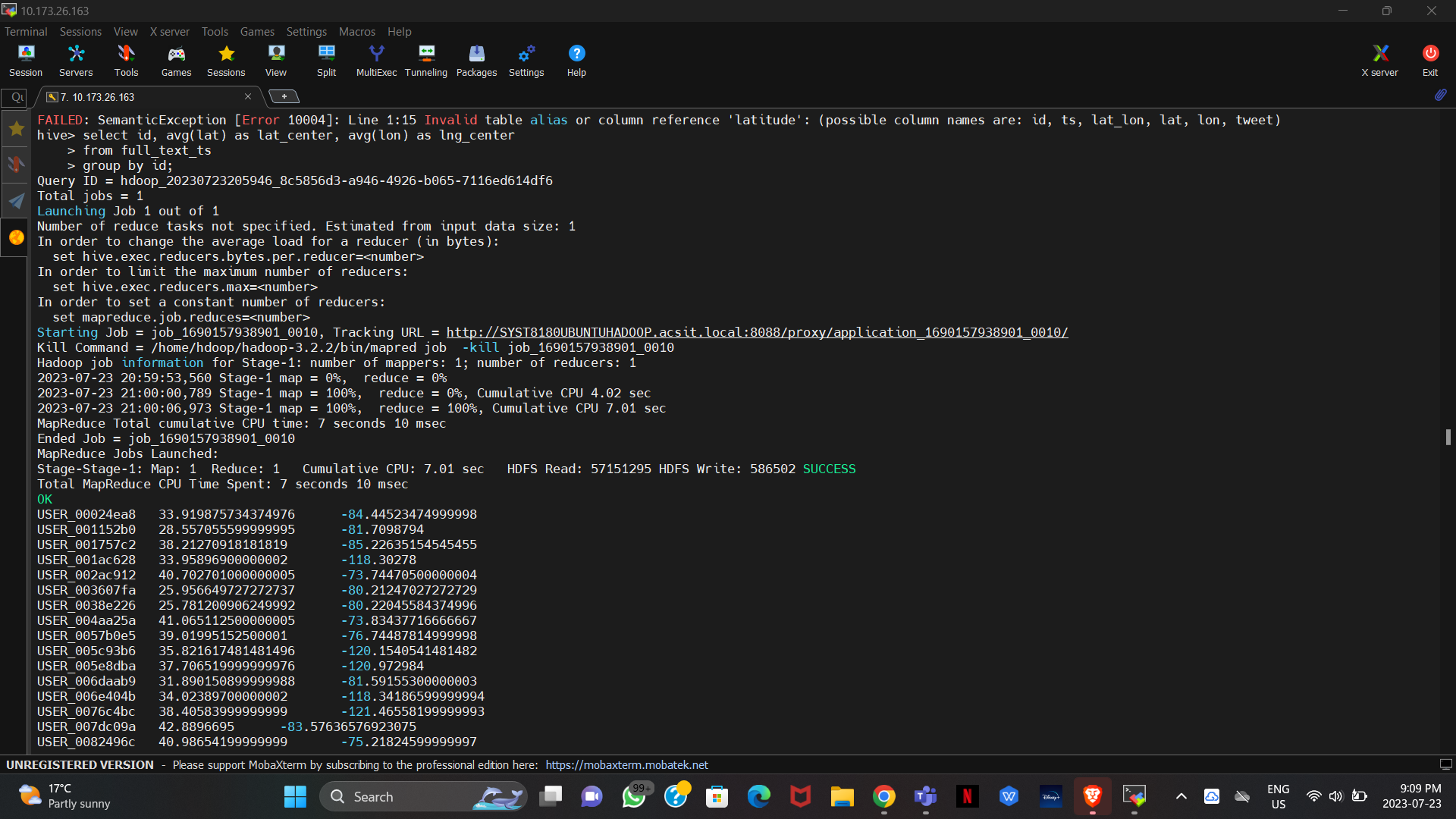


1. This is to find the number of ids from the table full\_text\_ts and the a group by clause is used which is grouping the query based on the ids, basically getting number of tweets from each user based on id and then grouping it.

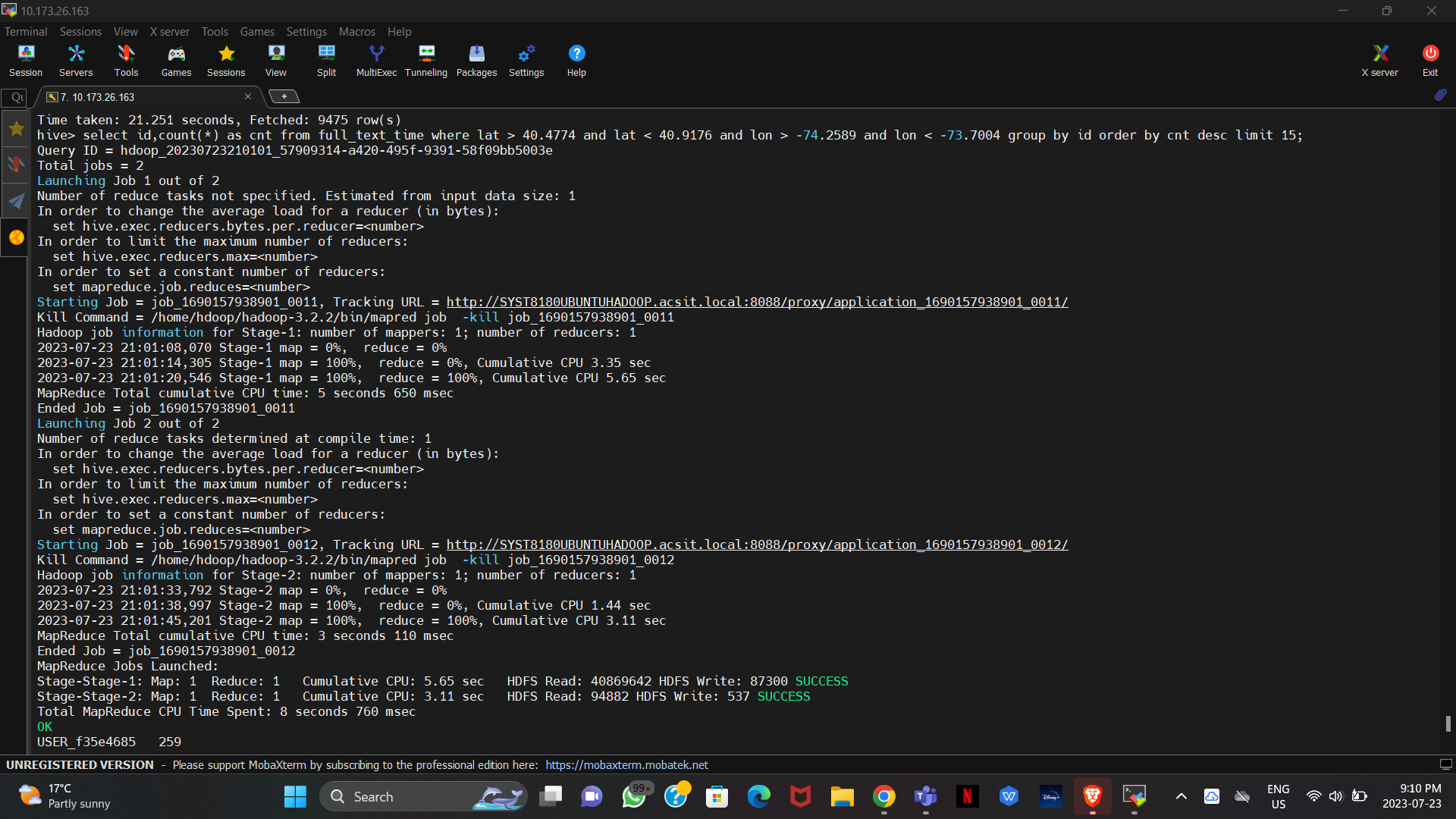


1. Average Function: To get all the tweets between a specific location, we are using this query with the fuction of average giving the average of latitude and logitude and using AS fuction to make it more clear by giving it a name.

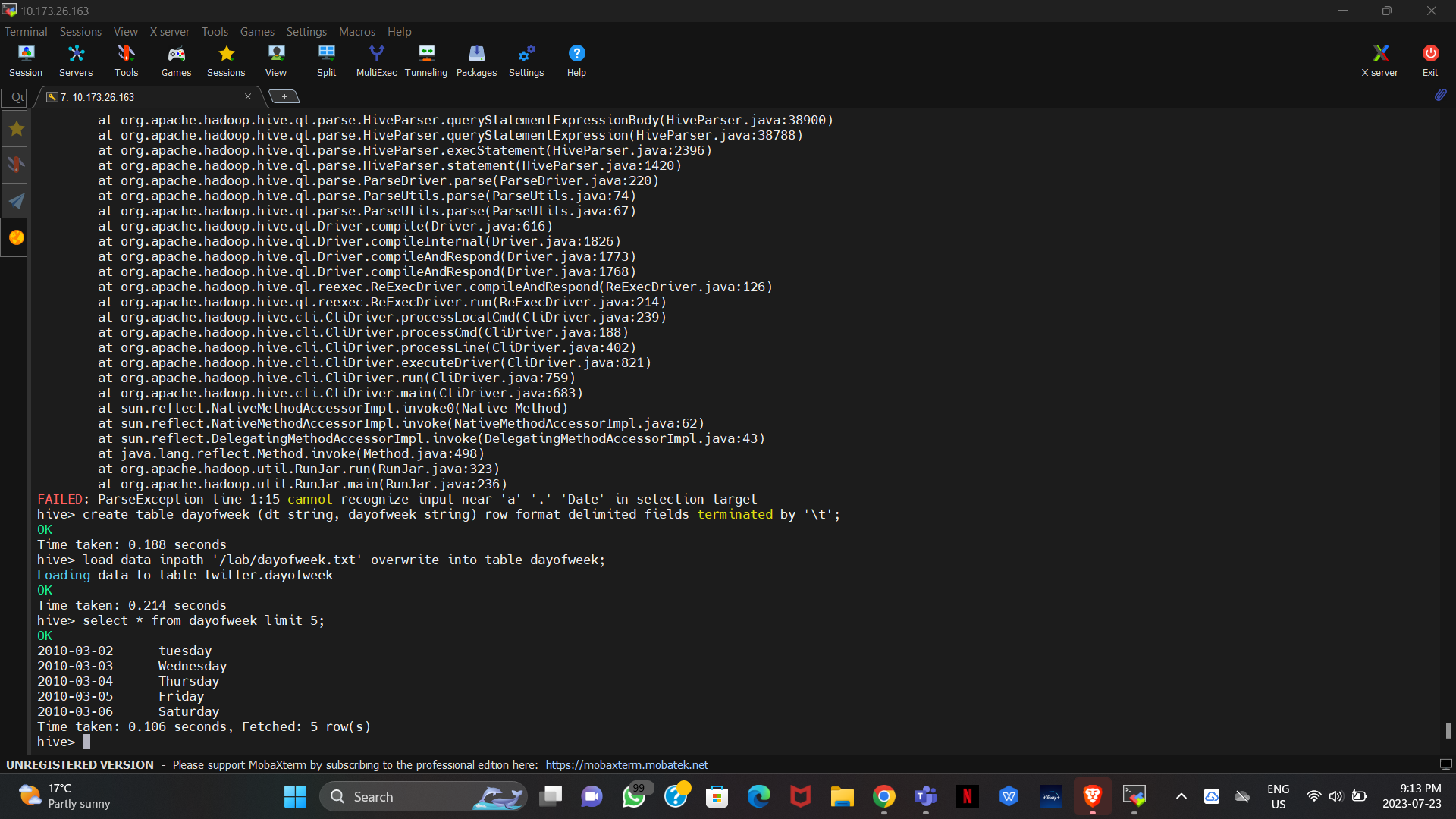




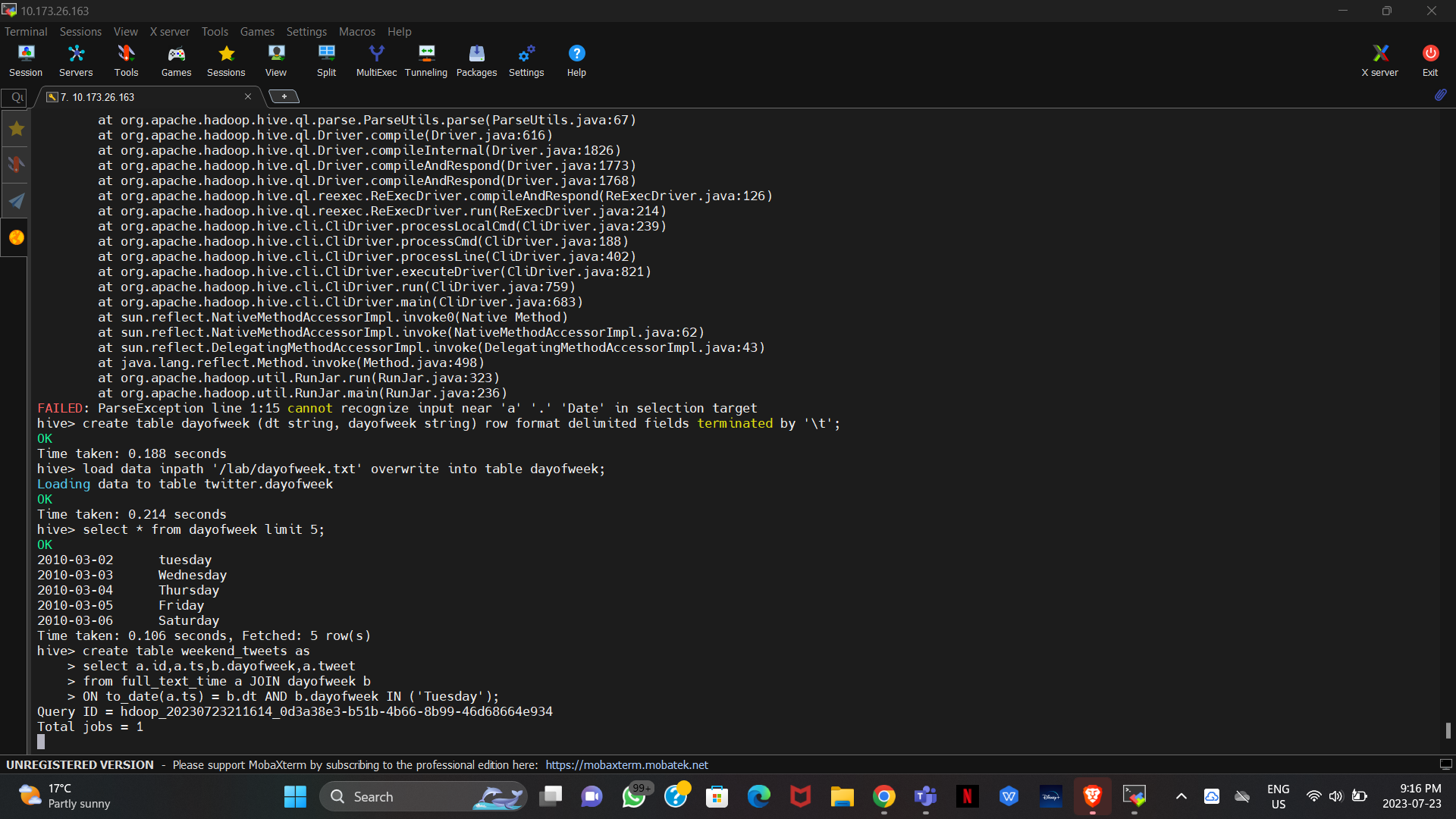
1. Order BY function: Using this query, we can find the best tweeters nearby. Here we are giving result for ids and counting them with where , group by clauses and then Ordering it in the DESC order.



1. Creating a table called dayofweek and then selecting all the rows from the table.



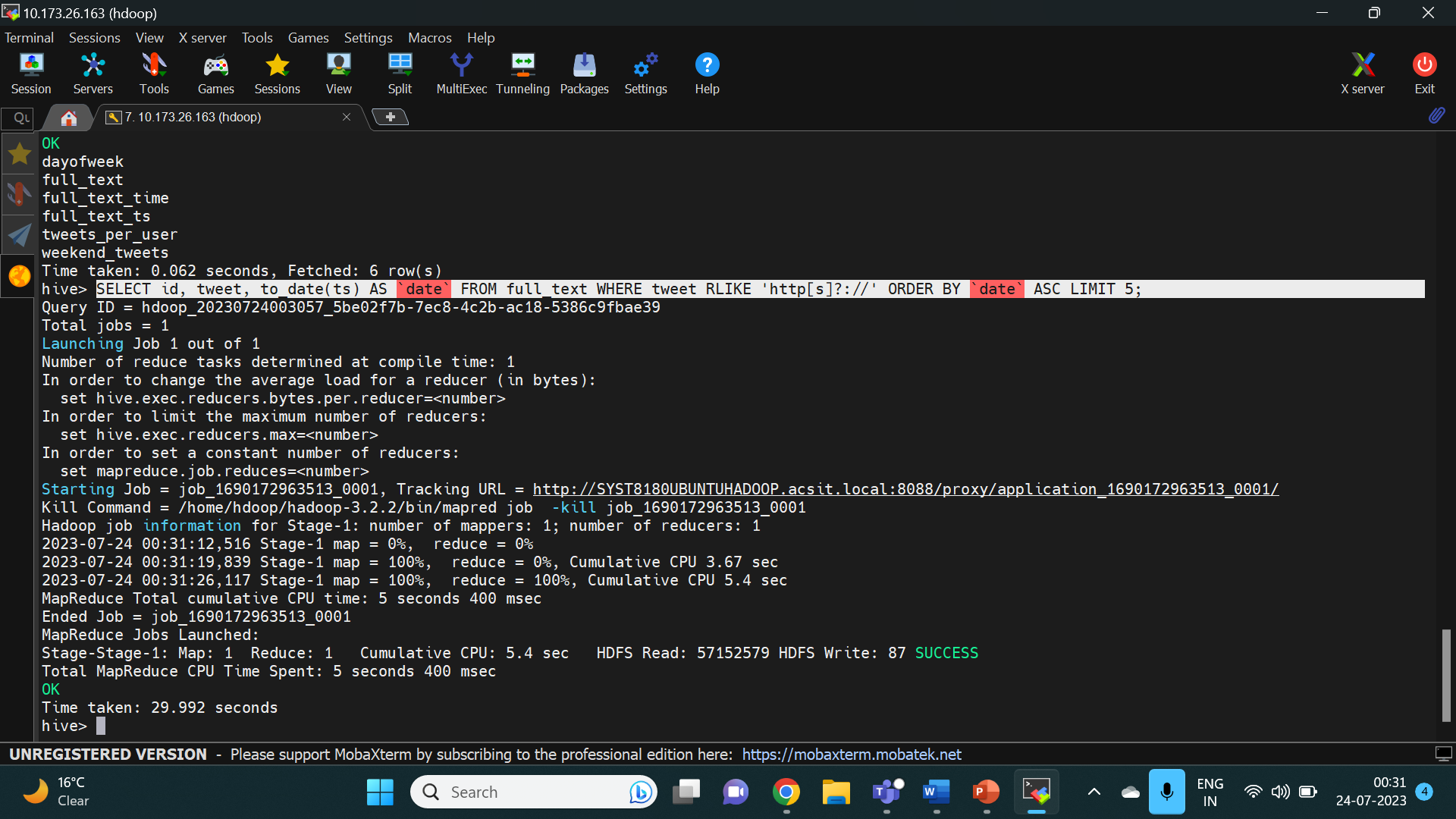
1. Following query gives results based on joining two tables together

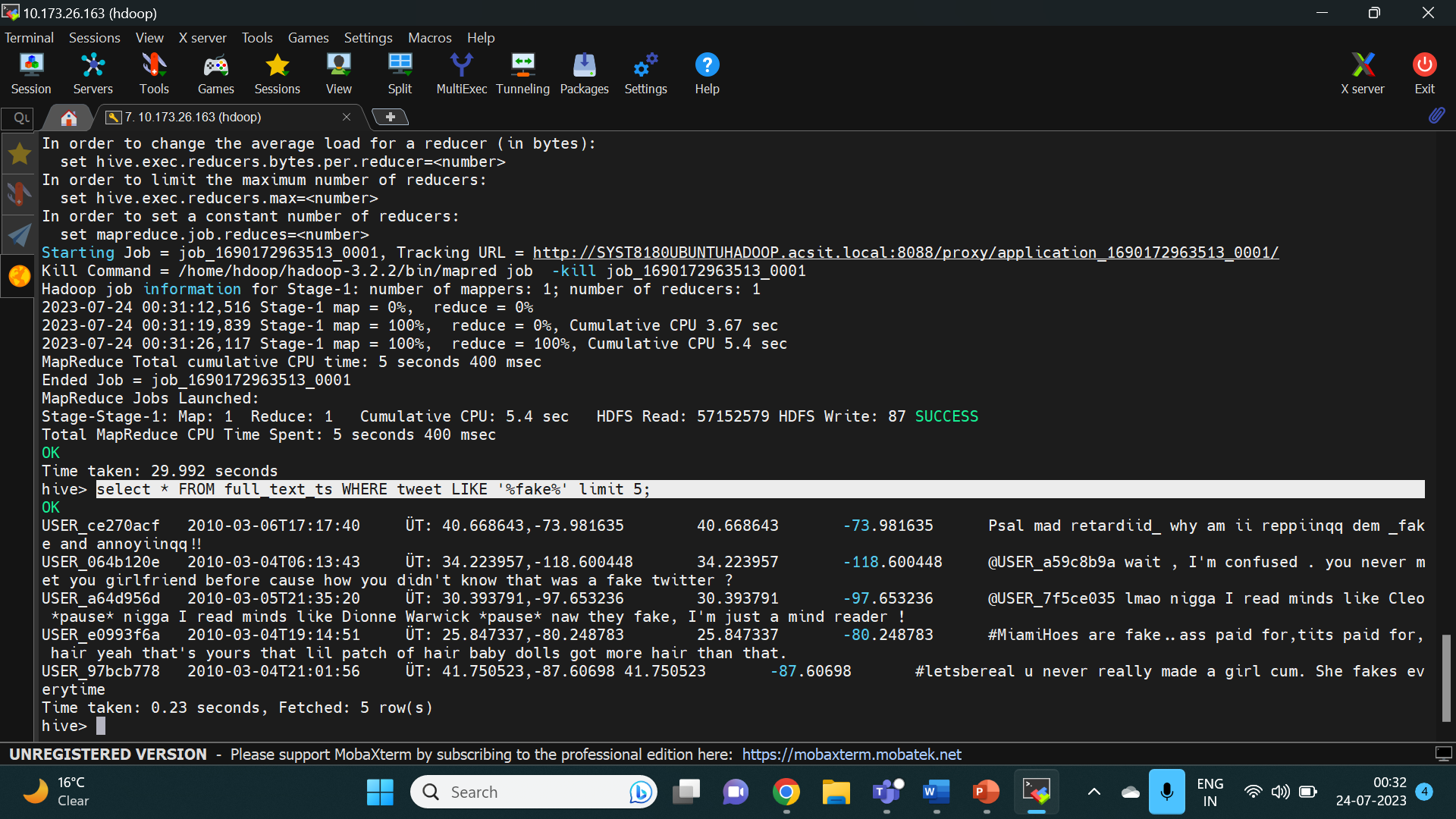


## Question 2: Identify five more fruitful queries using various combinations of regular expression, where order by and other options available under HiveQL. Justify your query by giving a proper explanation for the proposed query.

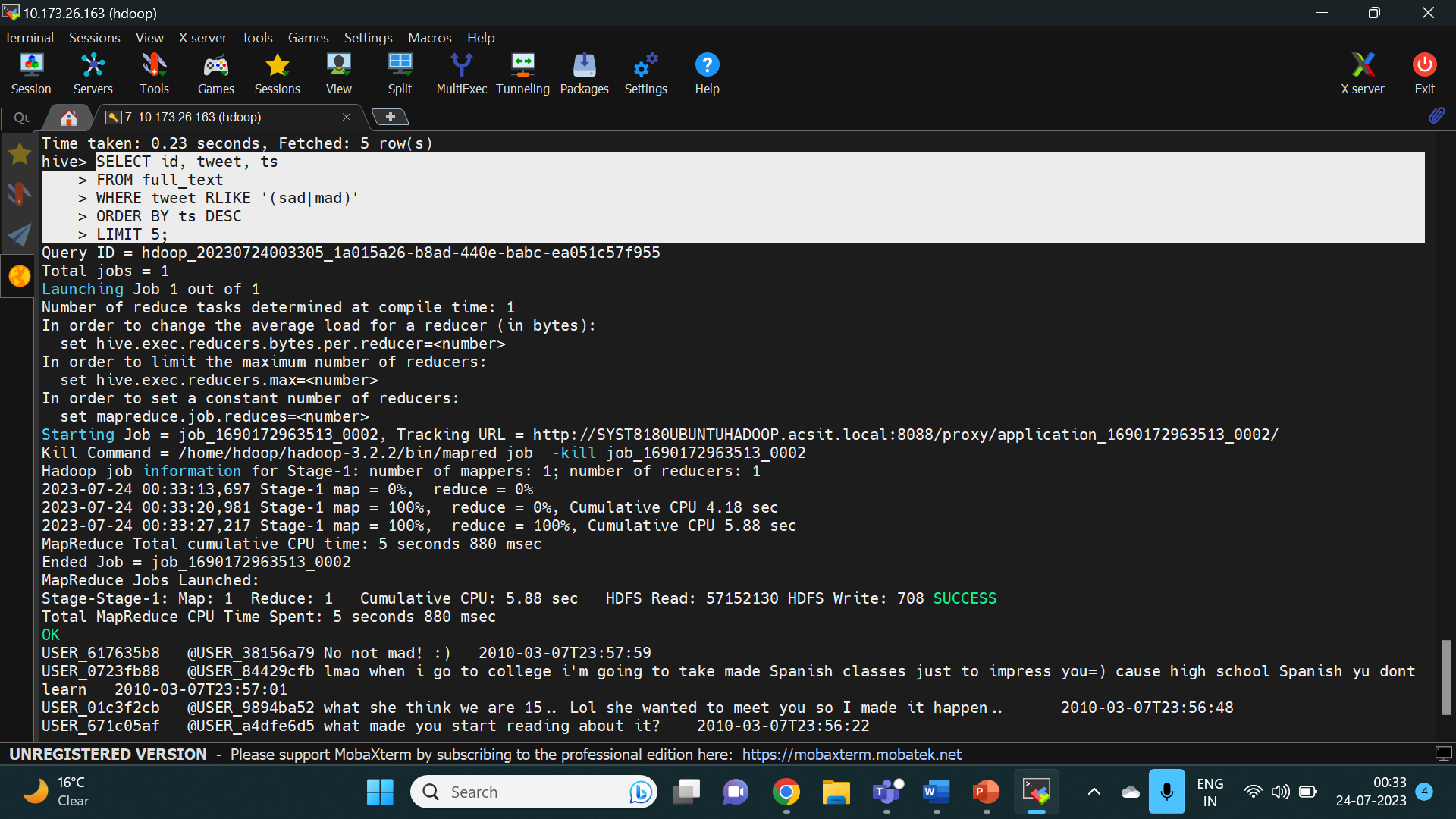
1. The below Hive query selects the "id," "tweet," todate columns from the "full\_text" dataset, filters tweets that contain URLs using WHERE query , sorts the outcomes by the converted date and outputs the first 5 rows as the limit.

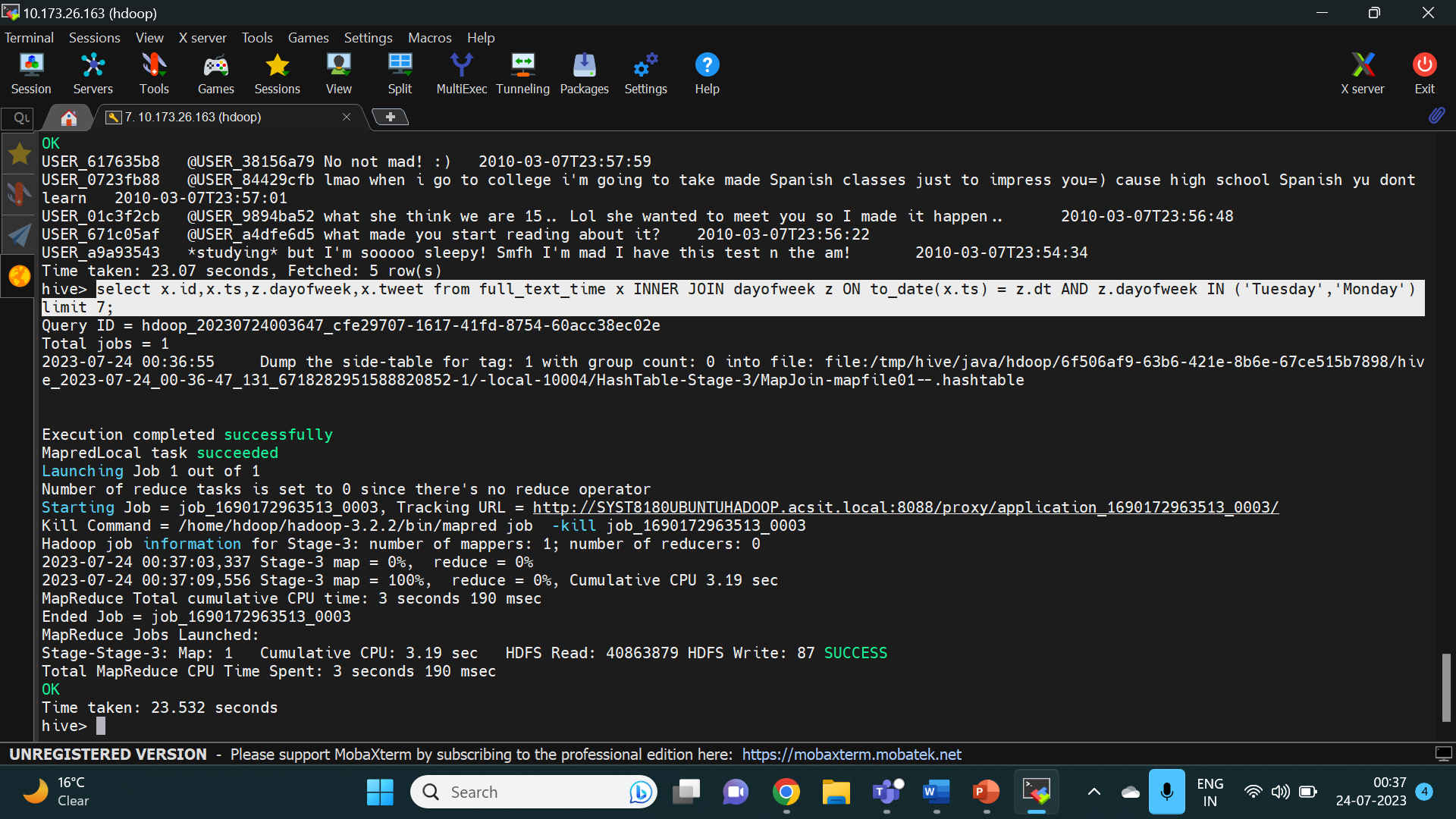
## 

 2) The below query finds the particular word : fake from all the tweets present in the table full\_text\_ts and limit it to 5.



3) The below query finds the id, tweets , ts in which the words are sad or mad then order it in the descending order.



4) the below query creates a table using inner join for the two tables and giving results based on specific days for Teusday and Monday.

5) The below query gives the output as the id , tweet and use a timepestamp conversion to date and then queries for the selected tweets with the specific url and then orders it in the ascending order and further limit it to 5. 