**Hadoop - Project**

**Class ID : 2**

**Name : Sowmya Reddy Katukuri**

**Team - 2**

**Other Team Members :**

**Ravindra Reddy Chennapareddy (Wiki link :**[**https://github.com/chravi414/BDP/wiki/BDP-Exam1**](https://github.com/chravi414/BDP/wiki/BDP-Exam1)**)**

**Anmisha Brahmadevara ( Wiki link :**[**https://github.com/anmisha25/bdpexam/wiki**](https://github.com/anmisha25/bdpexam/wiki)**)**

**Pranathi Nalluri ( Wiki link :**[**https://github.com/Pranathi-07/BDP-ICP1/wiki/Exam1**](https://github.com/Pranathi-07/BDP-ICP1/wiki/Exam1)**)**

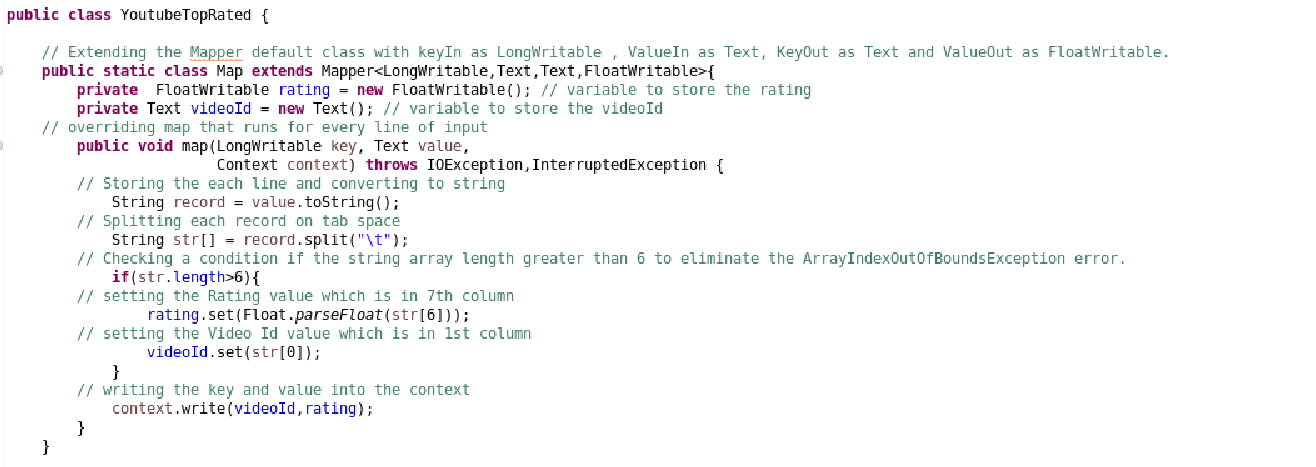
**Implementing Map-Reduce Algorithm to find top 10 rated videos on youtube. I have worked on mapper class for this task.**

**Mapper Class :**

**Mapper class stores video rating in a variable and also stores their videoID.**

**Overriding map runs for every line and each line is converted to string and splits each record by using tab space.**

**Checks if the string length is greater than 6 and taking seventh column which is rating and sending the key value pairs as videoID and rating to reducer class.**

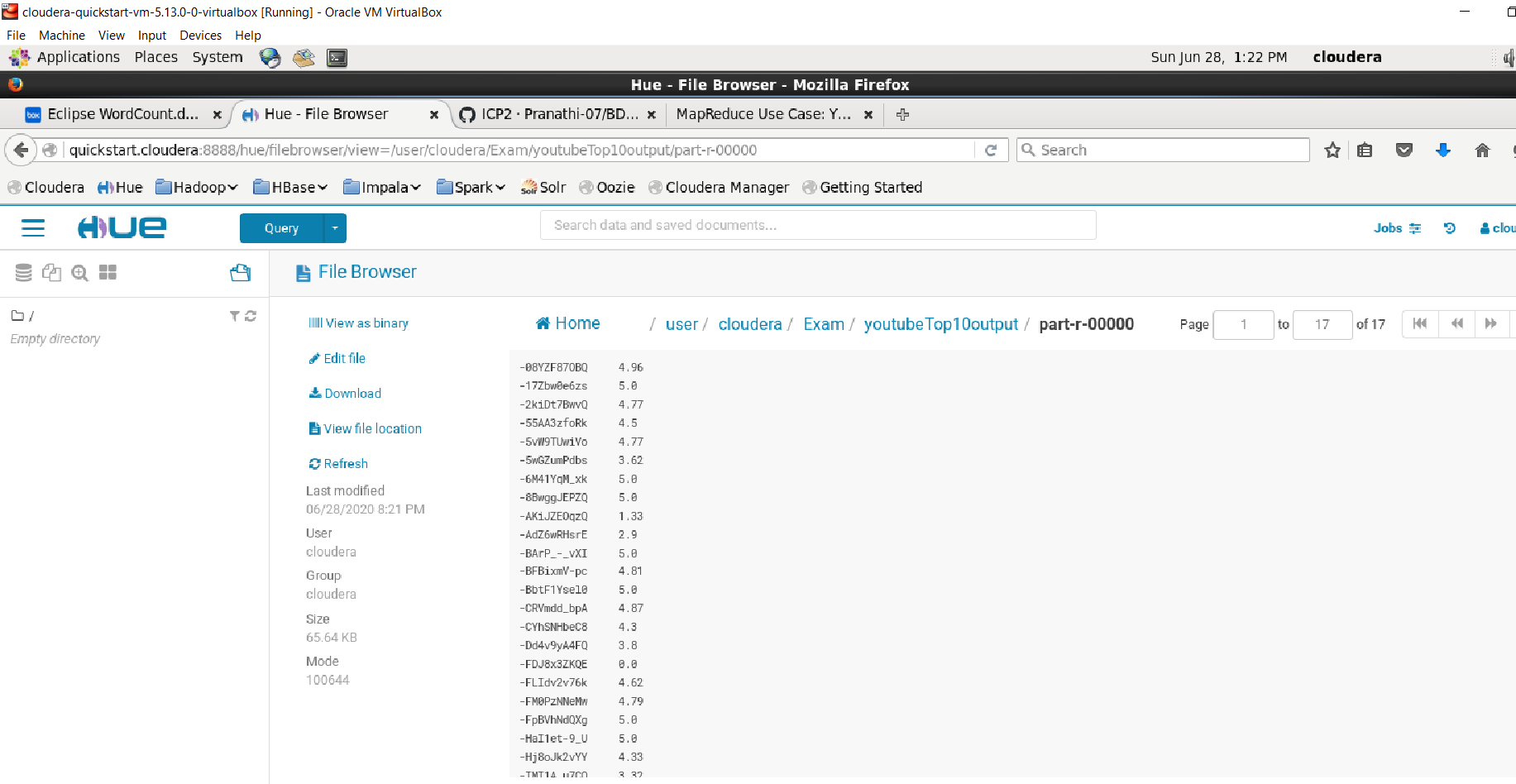


**Anmisha worked on reducer part and we integrated both the tasks and created a jar file to execute the task.**

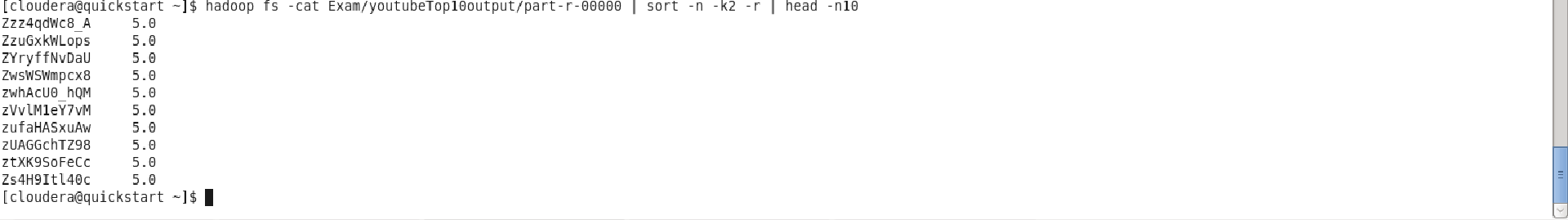
**This is the driver class where map reduce classes are set.**



**Output**



**List of top 10 rated videos**



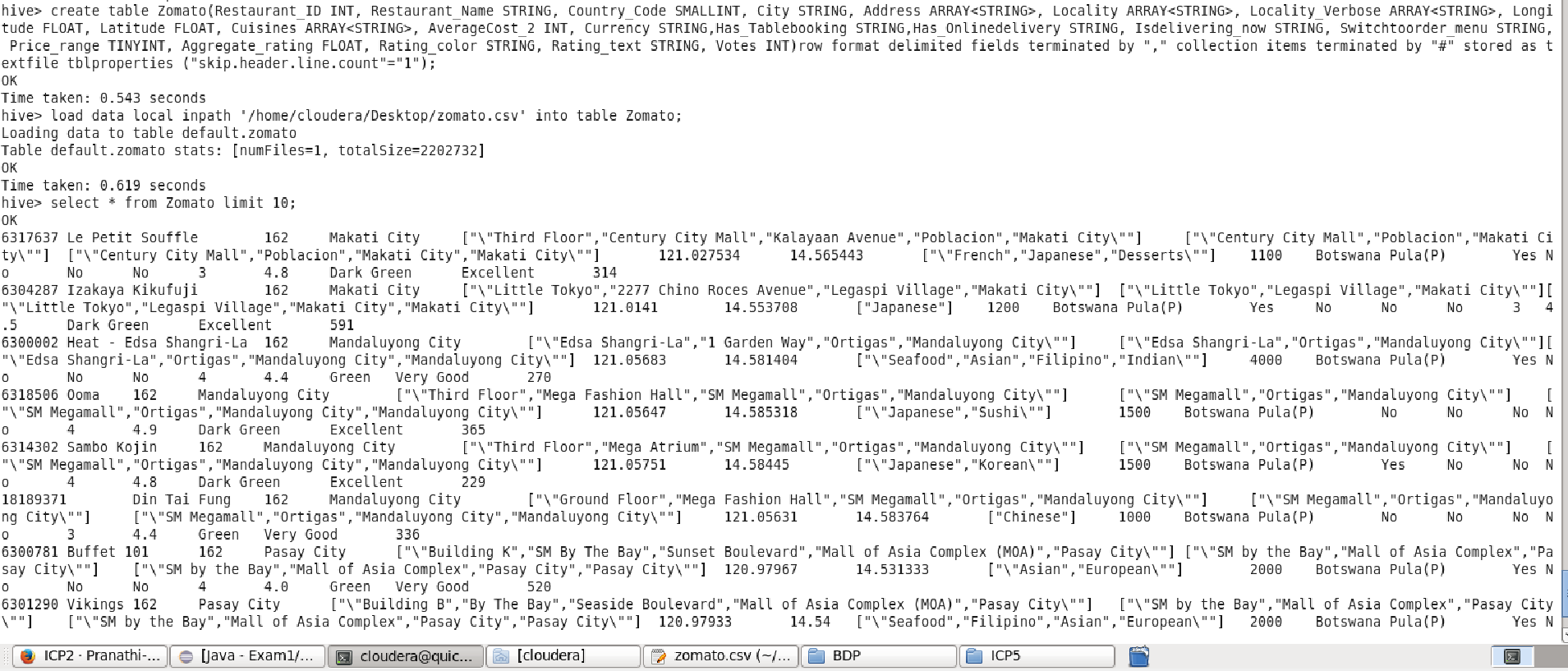
**Hive Use Case**

**Hive Table is created first to perform Queries. To Create a hive table below command is used :**

**create table Zomato(Restaurant\_ID INT, Restaurant\_Name STRING, Country\_Code SMALLINT, City STRING, Address ARRAY, Locality ARRAY, Locality\_Verbose ARRAY, Longitude FLOAT, Latitude FLOAT, Cuisines ARRAY, AverageCost\_2 INT, Currency STRING,Has\_Tablebooking STRING,Has\_Onlinedelivery STRING, Isdelivering\_now STRING, Switchtoorder\_menu STRING, Price\_range TINYINT, Aggregate\_rating FLOAT, Rating\_color STRING, Rating\_text STRING, Votes INT)row format delimited fields terminated by ‘,’ collection items terminated by '#' stored as textfile tblproperties ("skip.header.line.count"="1");**

**Loading Data into Table :**

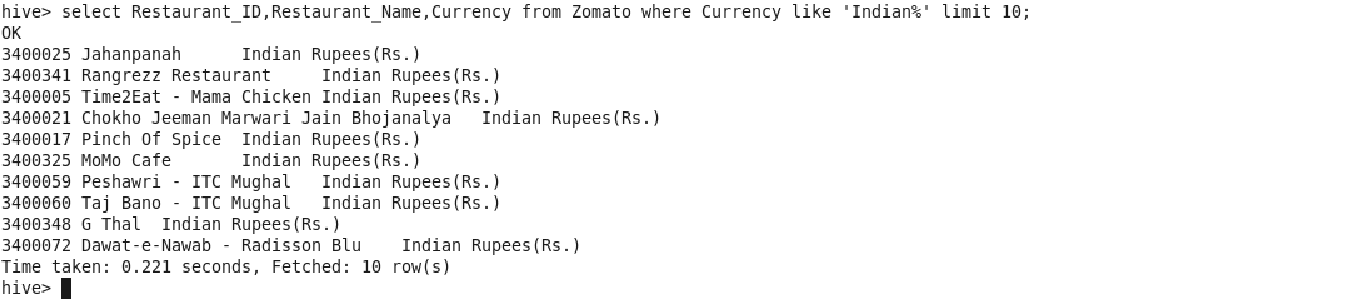
**load data local inpath '/home/cloudera/Desktop/zomato.csv' into table Zomato;**



**Queries Executed On table are shown below**

**Query -1 : To find restaurants whose currency matches with indian currency**

**select Restaurant\_ID,Restaurant\_Name,Currency from Zomato where Currency like 'Indian%'**



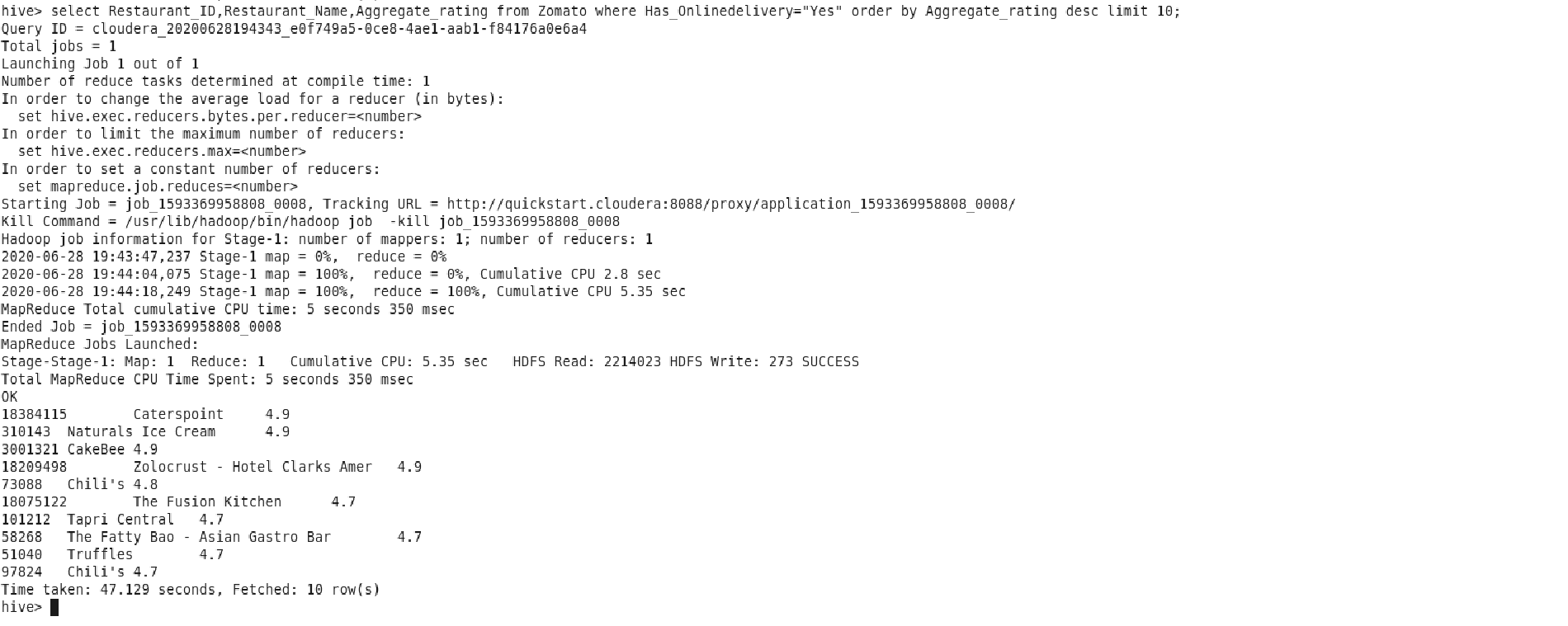
**Query -2 : To find numbers of restaurants in each countrycode**

**select Country\_Code, count(Country\_Code) from Zomato group by Country\_Code;**



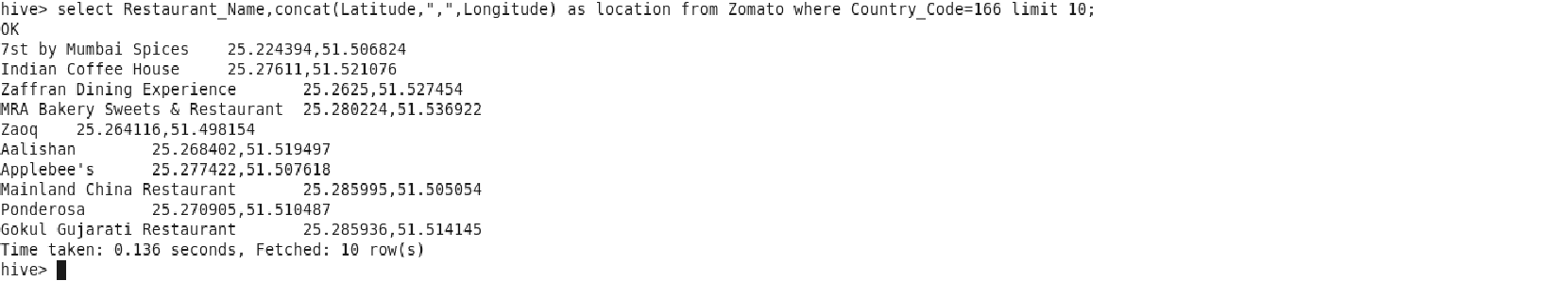
**Query -3 : To get top 10 restaurants who has online delivery option and also have highest rating**

**select Restaurant\_ID,Restaurant\_Name,Aggregate\_rating from Zomato where Has\_Onlinedelivery="Yes" order by Aggregate\_rating desc limit 10;**



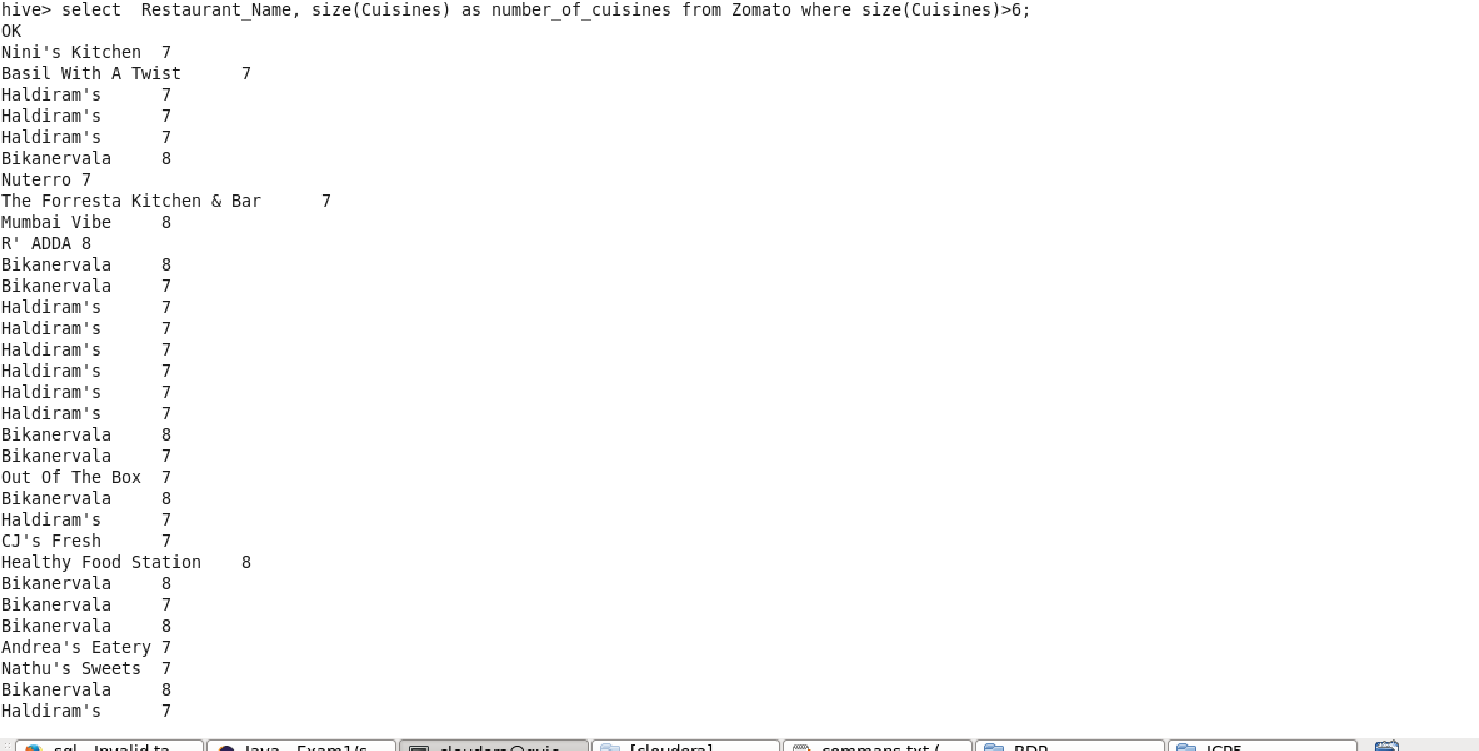
**Query -4 : Concatinate latitude and longitude for country code 166**

**select Restaurant\_Name,concat(Latitude,",",Longitude) as location from Zomato where Country\_Code=166 limit 10;**



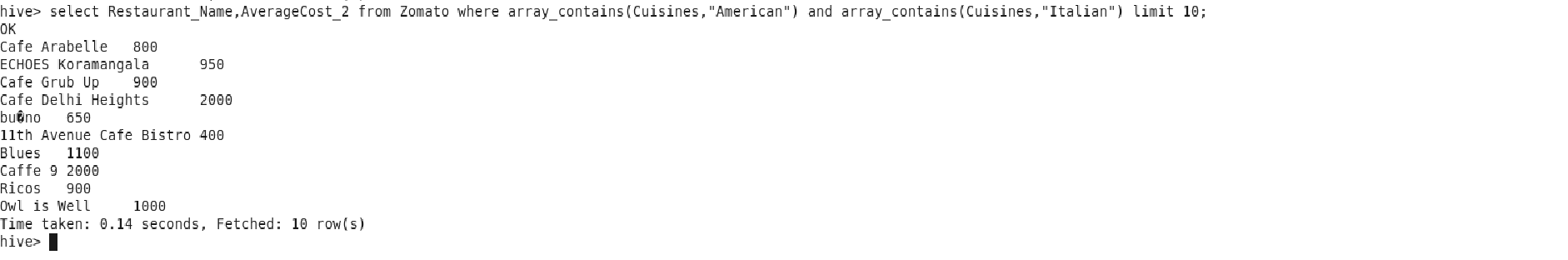
**Query -5 : To find restaurants which offer more than 6 cuisines**

**select Restaurant\_Name, size(Cuisines) as number of cuisines from Zomato where size(Cuisines)>6;**



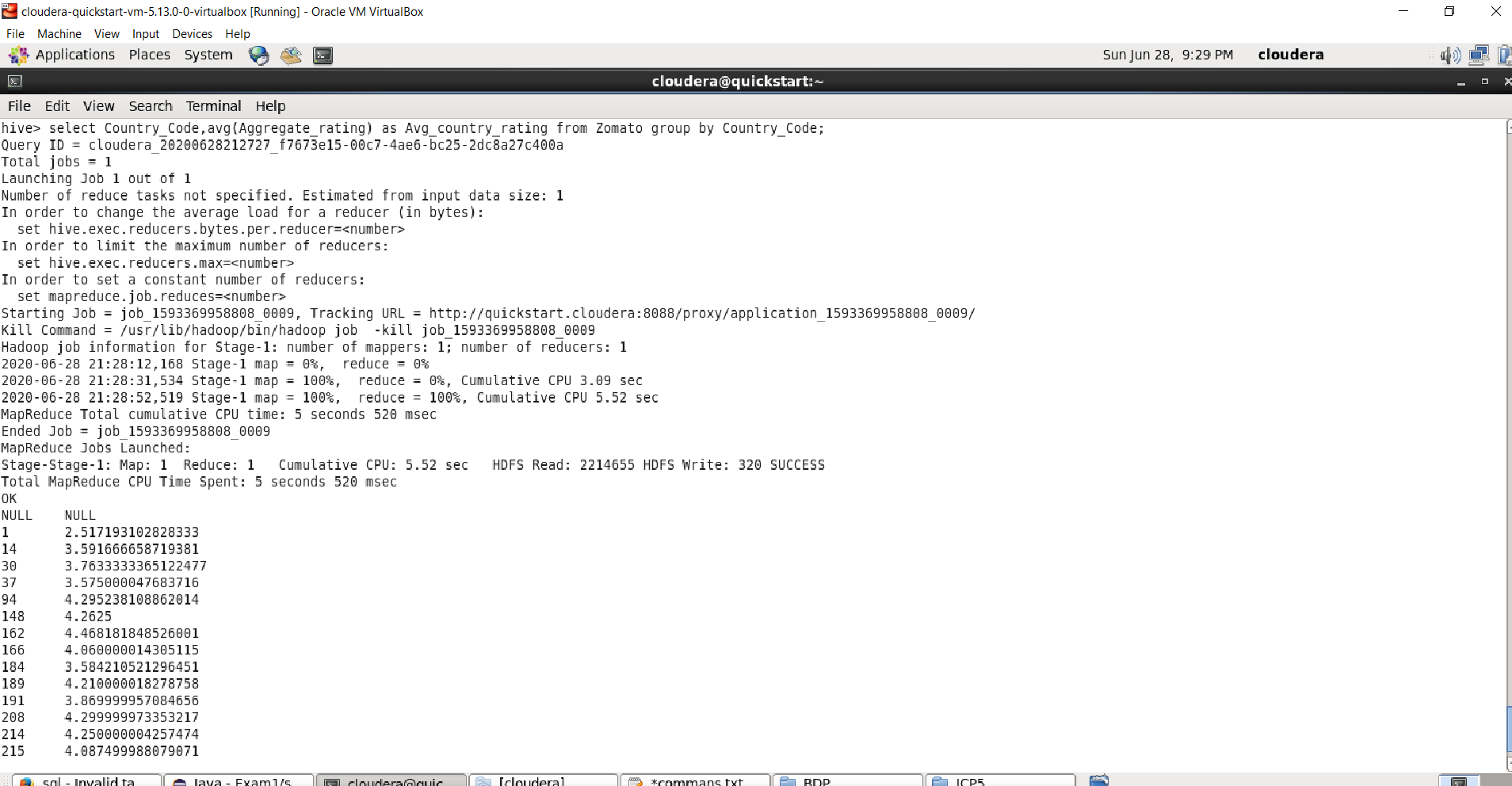
**Query -6 : To find restaurants with american and italian cuisines**

**select Restaurant\_Name,AverageCost\_2 from Zomato where array\_contains(Cuisines,"American","Italian") limit 10;**



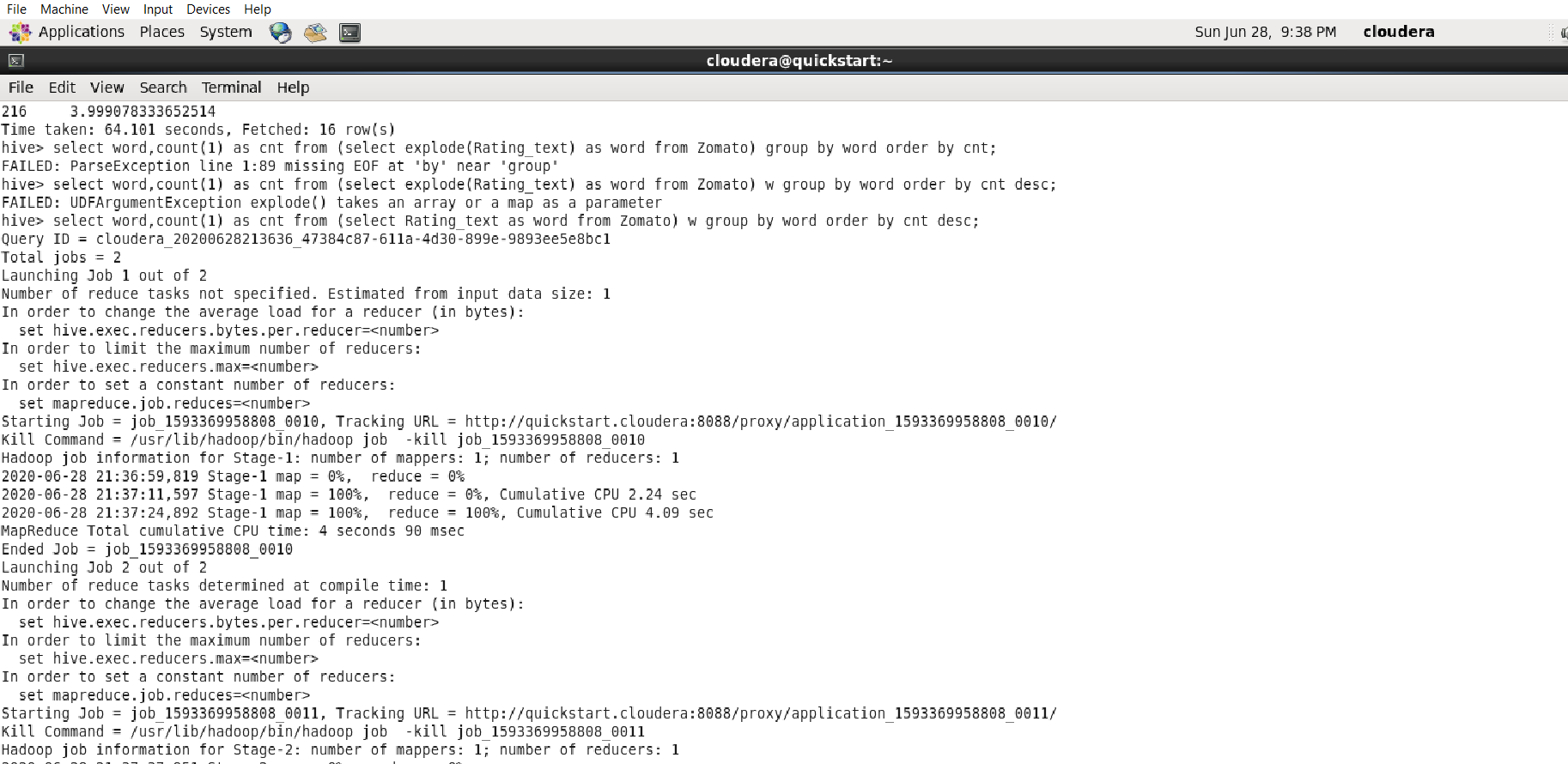
**Query -7 : To find the average Rating for each country**

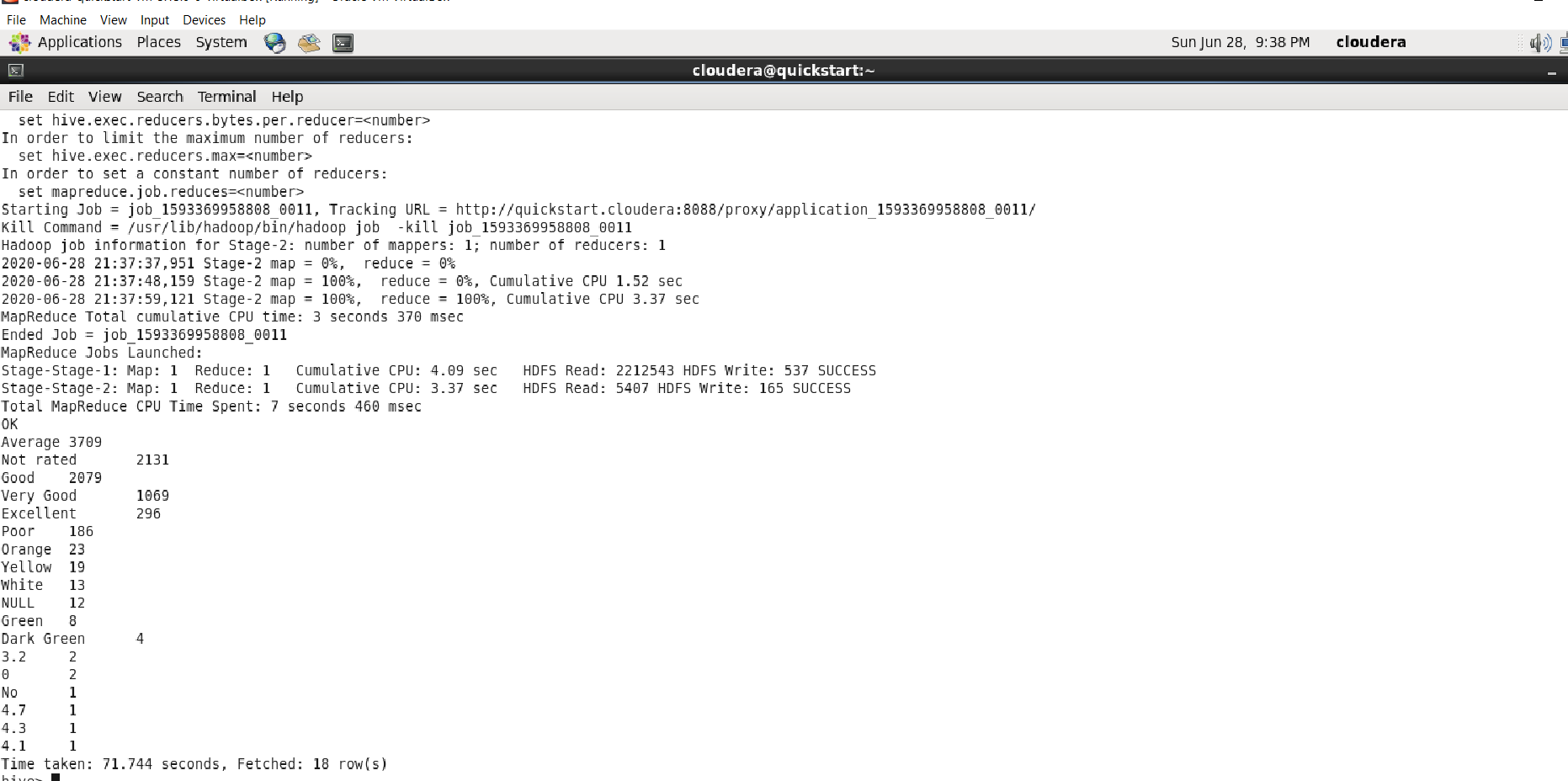
**select Country\_Code,avg(Aggregate\_rating) as Avg\_country\_rating from Zomato Cluster by Country\_Code;**



**Query -8 : To find wordcount for rating text**

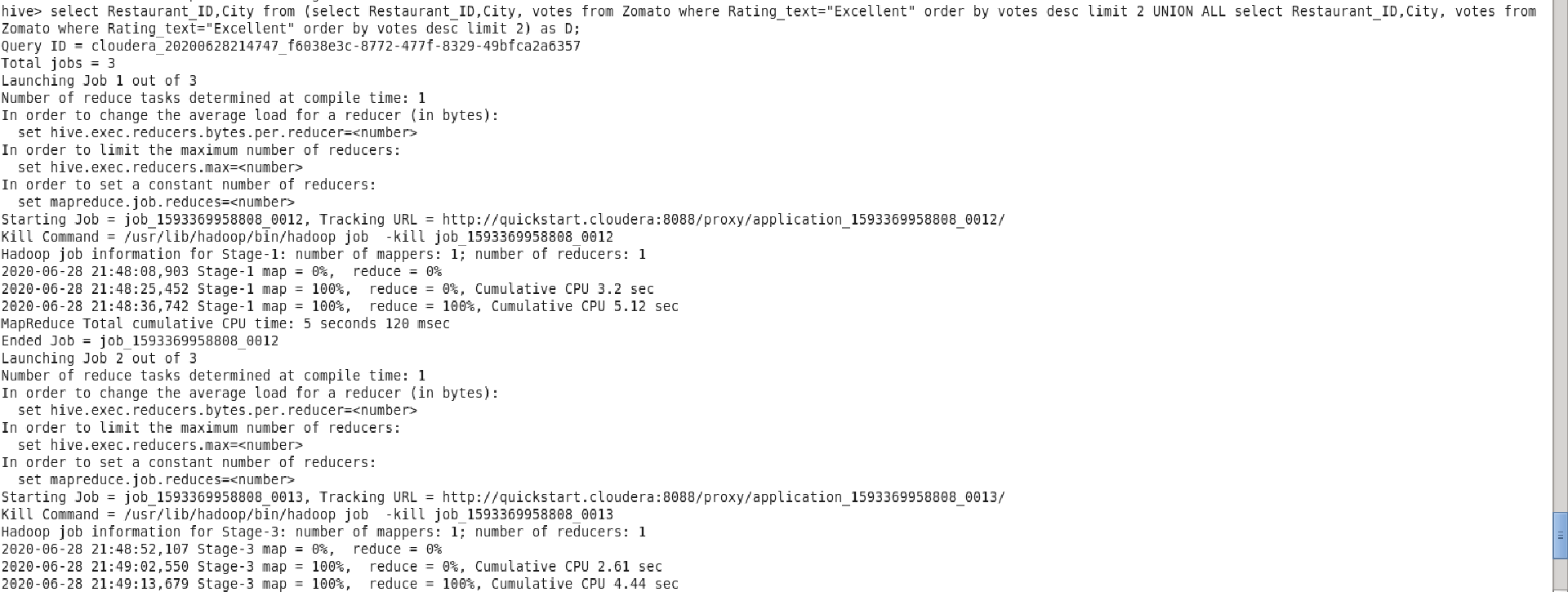
**select word,count(1) as cnt from (select explode(Rating\_text) as word from Zomato) group by word order by cnt;**

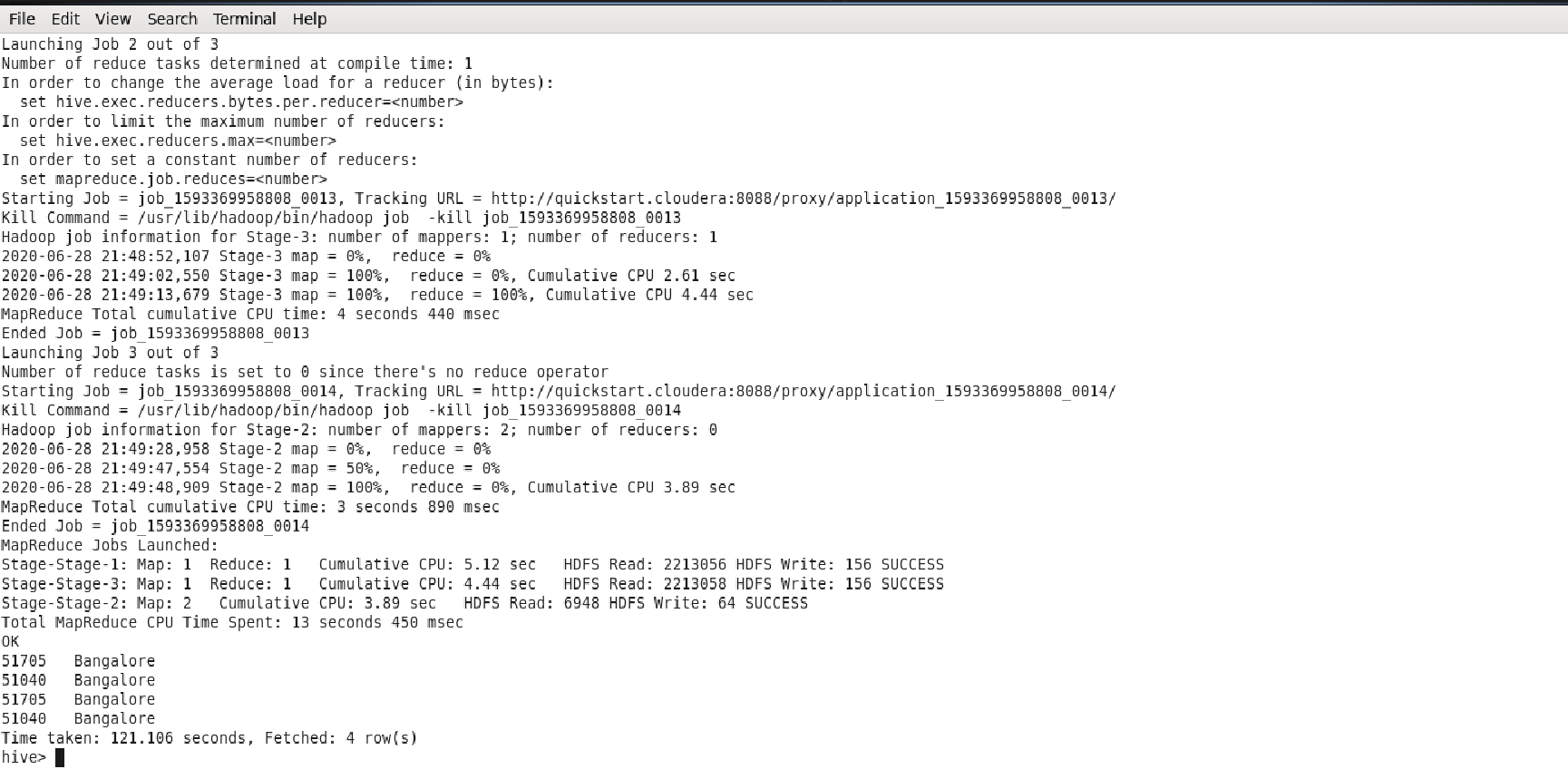




**Query -9 : To find top 2 restaurants with highest votes for each Excellent and very good rating text**

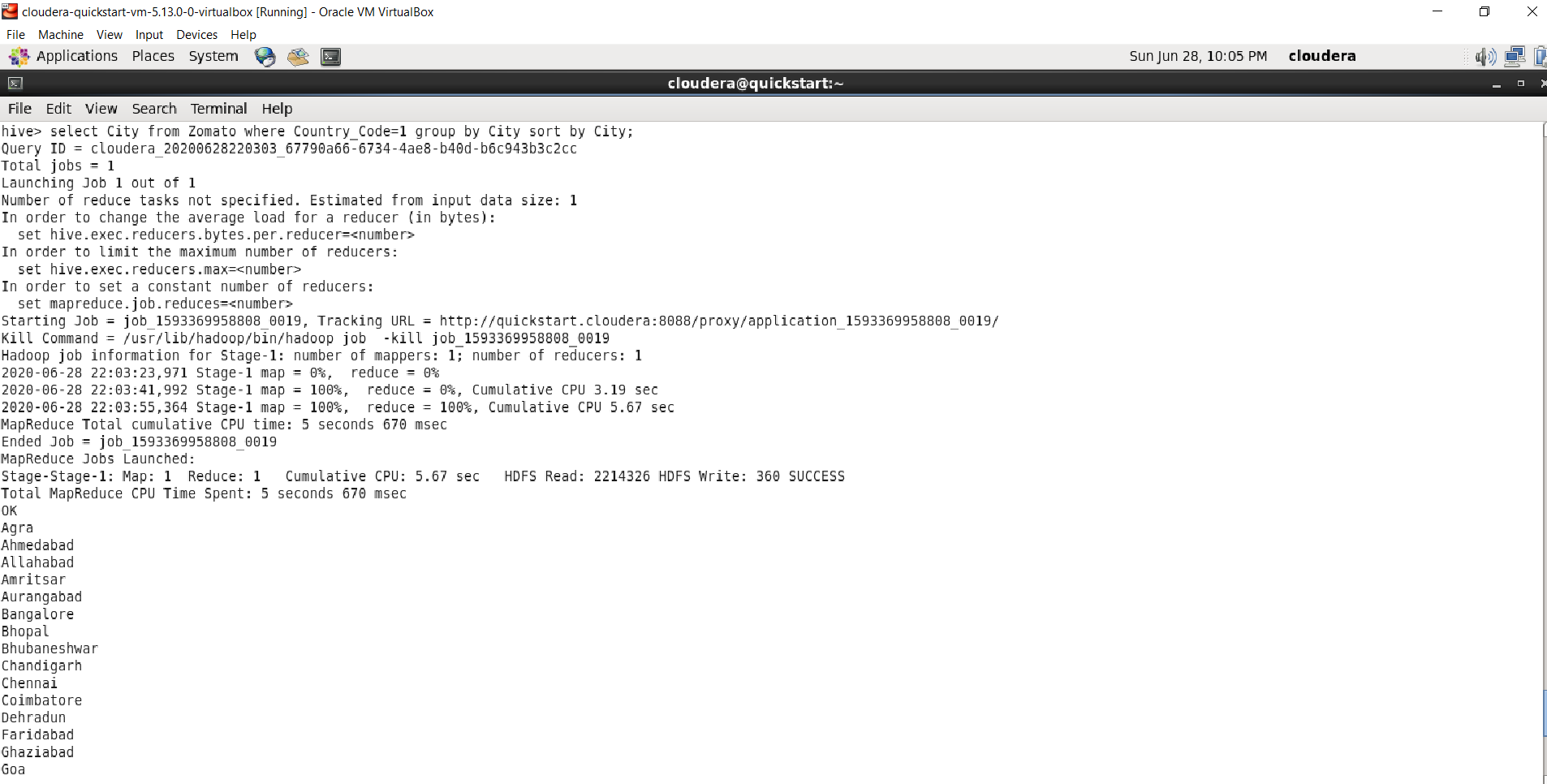
**select Restaurant\_ID,City from (select Restaurant\_ID,City, votes from Zomato where Rating\_text="Excellent" order by votes desc limit 2 UNION ALL select Restaurant\_ID,City, votes from Zomato where Rating\_text="Excellent" order by votes desc limit 2) as D;**

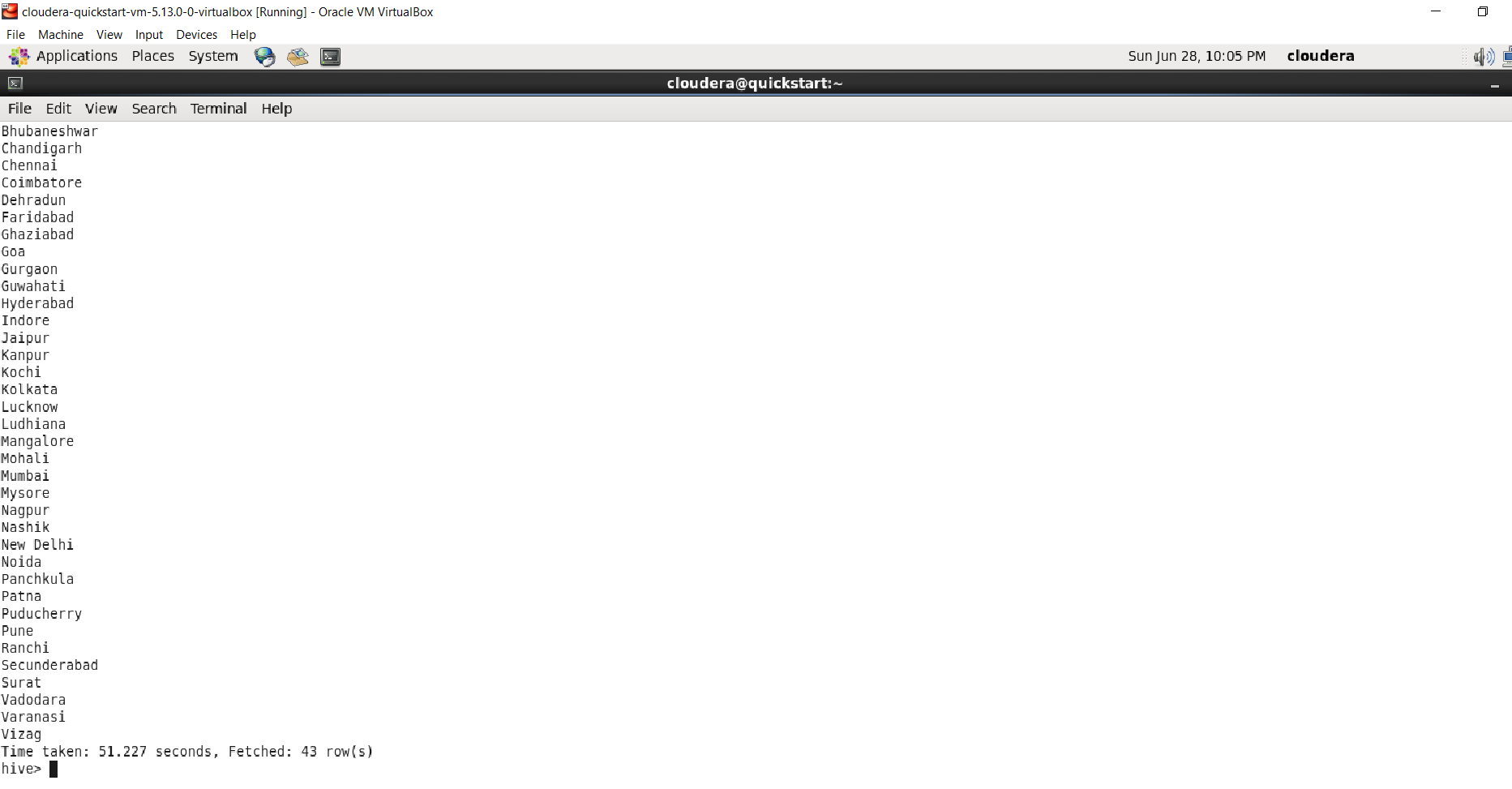




**Query -10 : To Display all cities for country code 1**

**select City from Zomato where Country\_Code=1 sort by City;**





**Questions**

**The Main idea of project is to get top ten rated youtube videos using map reduce algorithm . Map reduce works with key value pairs and the main concept is splitting the data and getting key value pairs as videoID and video rating and passing it to reducer class which sorts all the inputs and gives the final output.The main idea of the Hive use case is to collect Zomato Restaurants information using hive Queries.**

**This project is used in real time to analyze huge data like getting top ten rated videos from all youtube videos. The hive Queries can be used in real time to collect some useful information like which restaurants have highest rating and which restaurants have more number of cuisines available.**

**I have worked on the mapper class funtion in the task2 second part and worked on hive Queries.**

**I did not face much challenges as the project work was similar to the icp's we did earlier.**

**Hive Queries was individual work and there wasn't much to integrate with other team members but for the map-reduce task I and anmisha implemented it together as i have worked on mapper class and she worked on reducer class.**