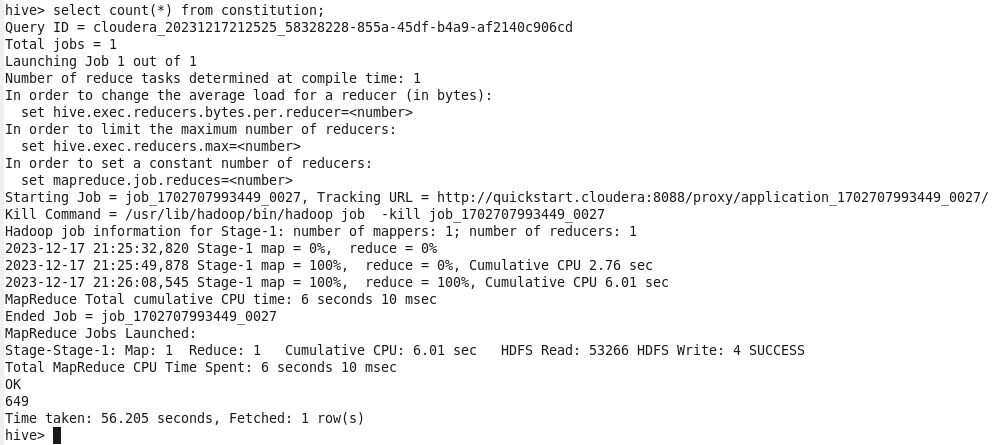


…

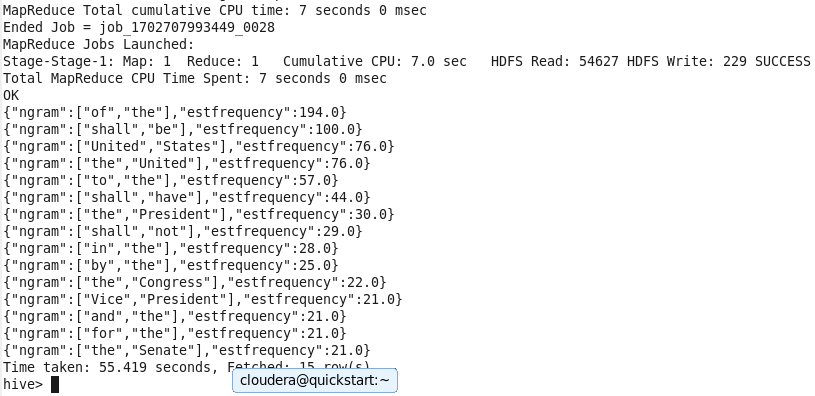
…



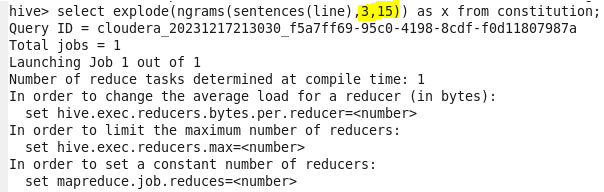
…



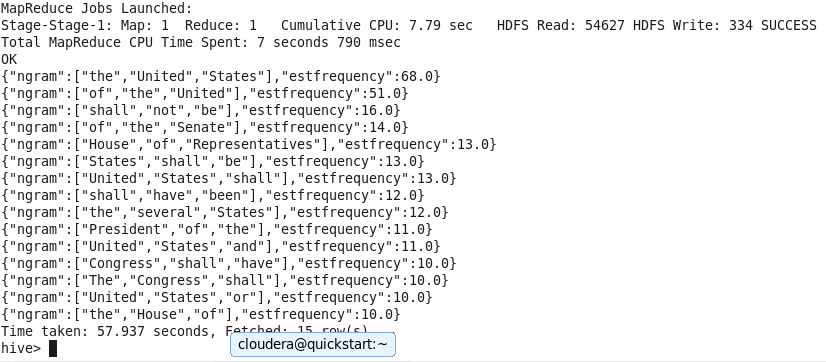
…



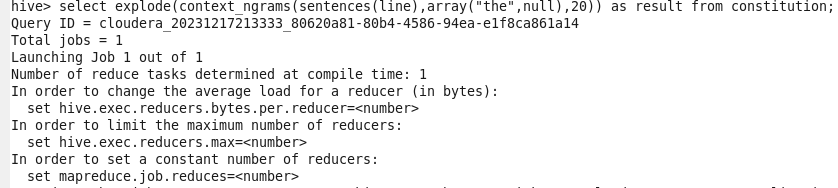
…



…

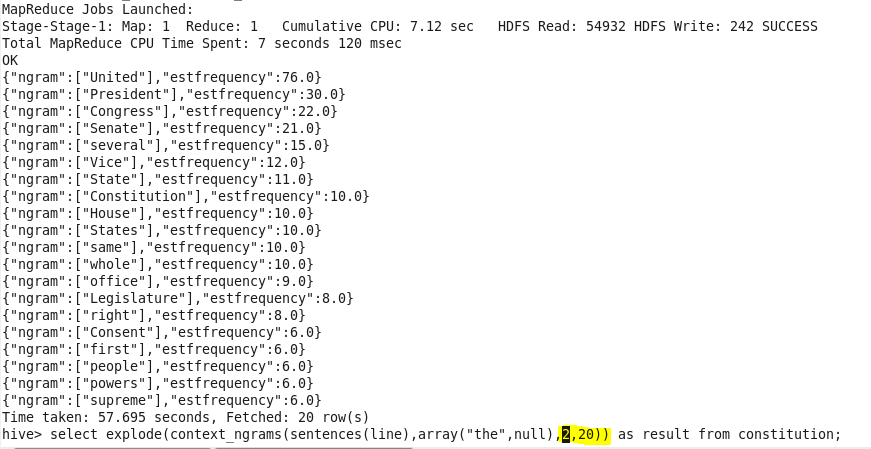


…

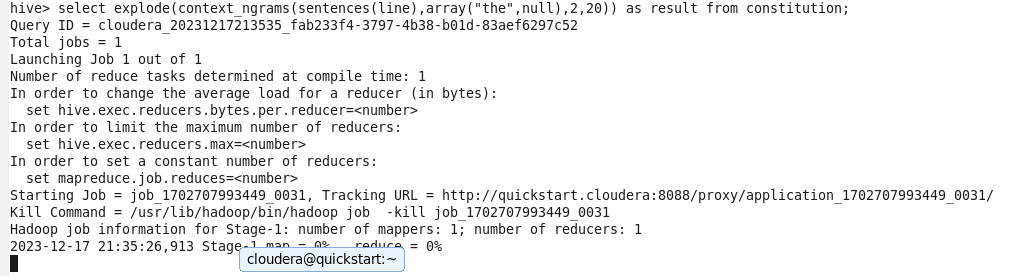


…

Top 20 bigrams after “the”

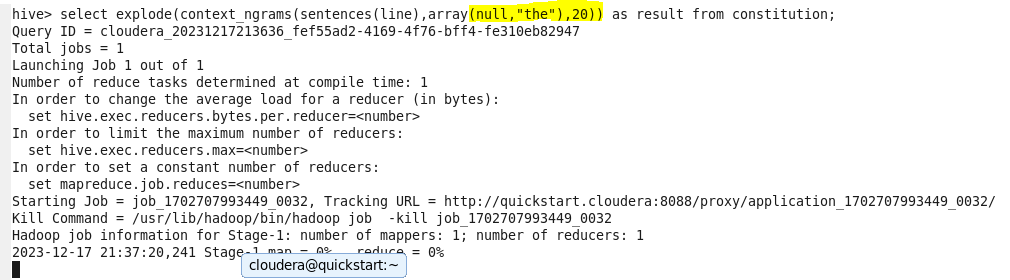


…

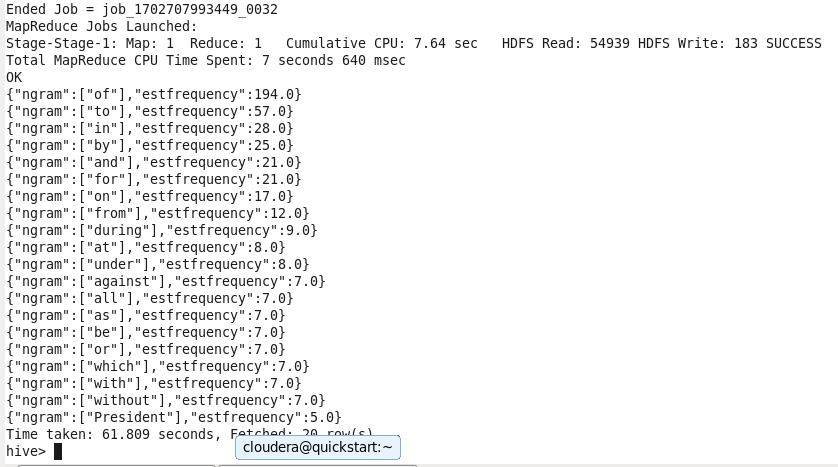


…

Top 20 words before “the”

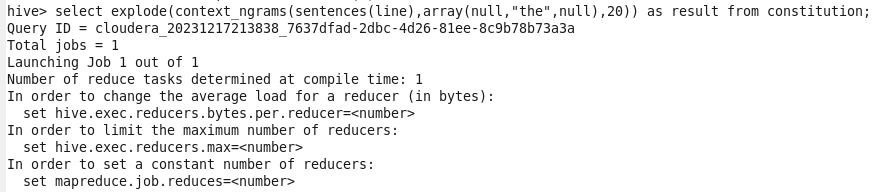


…

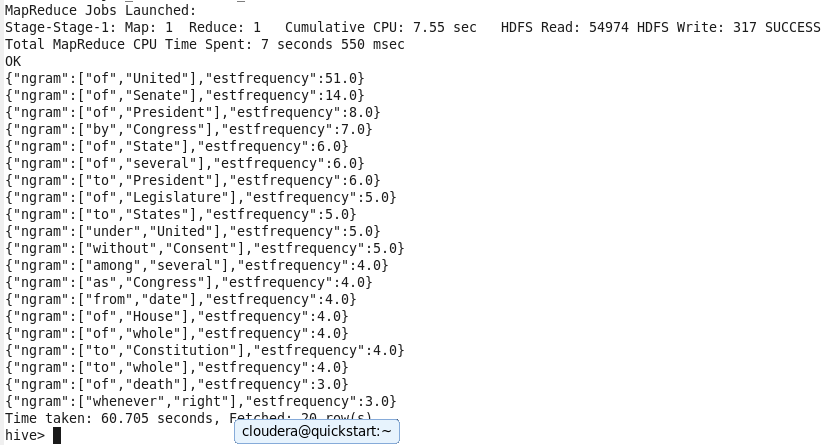


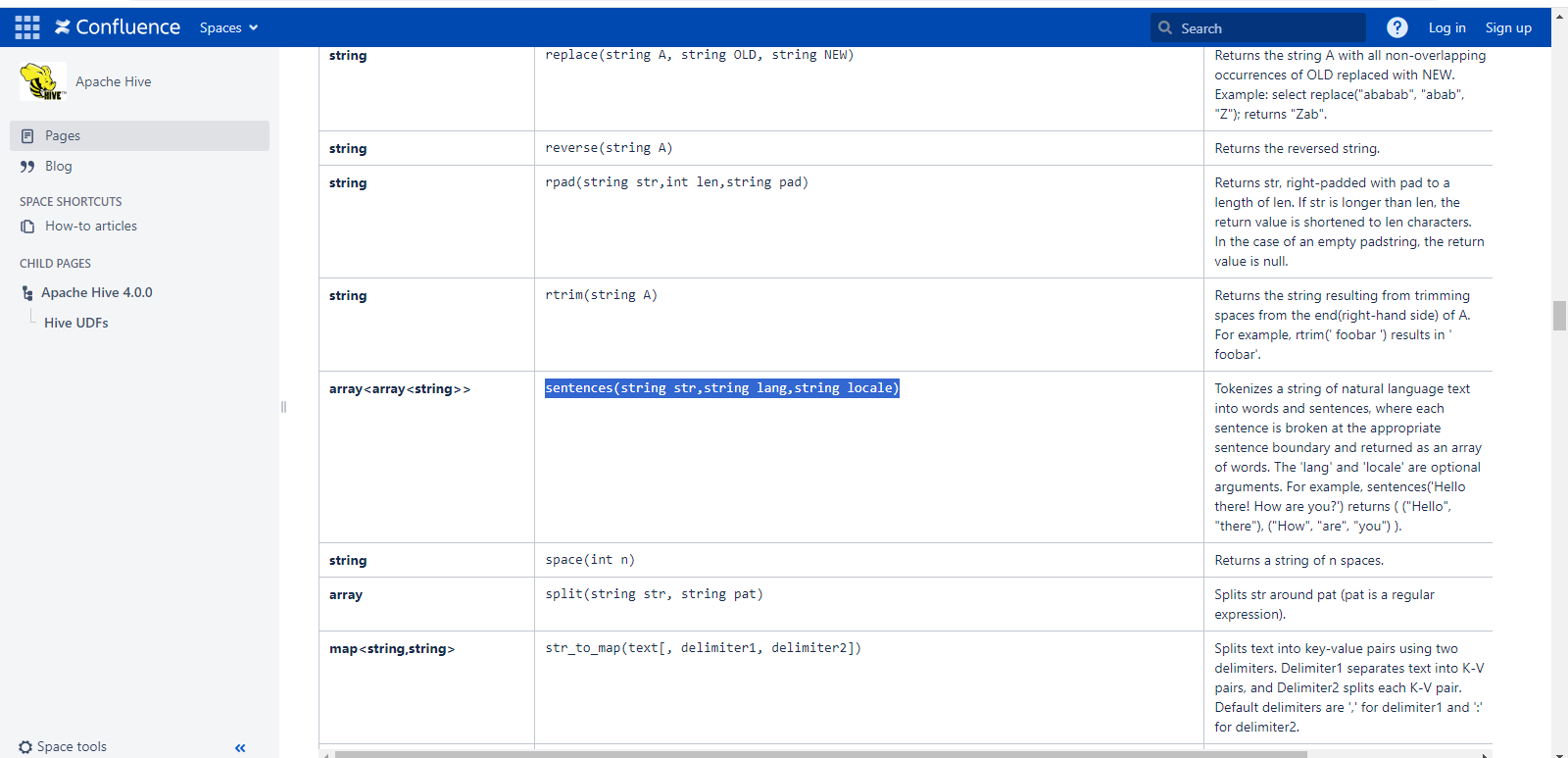
…

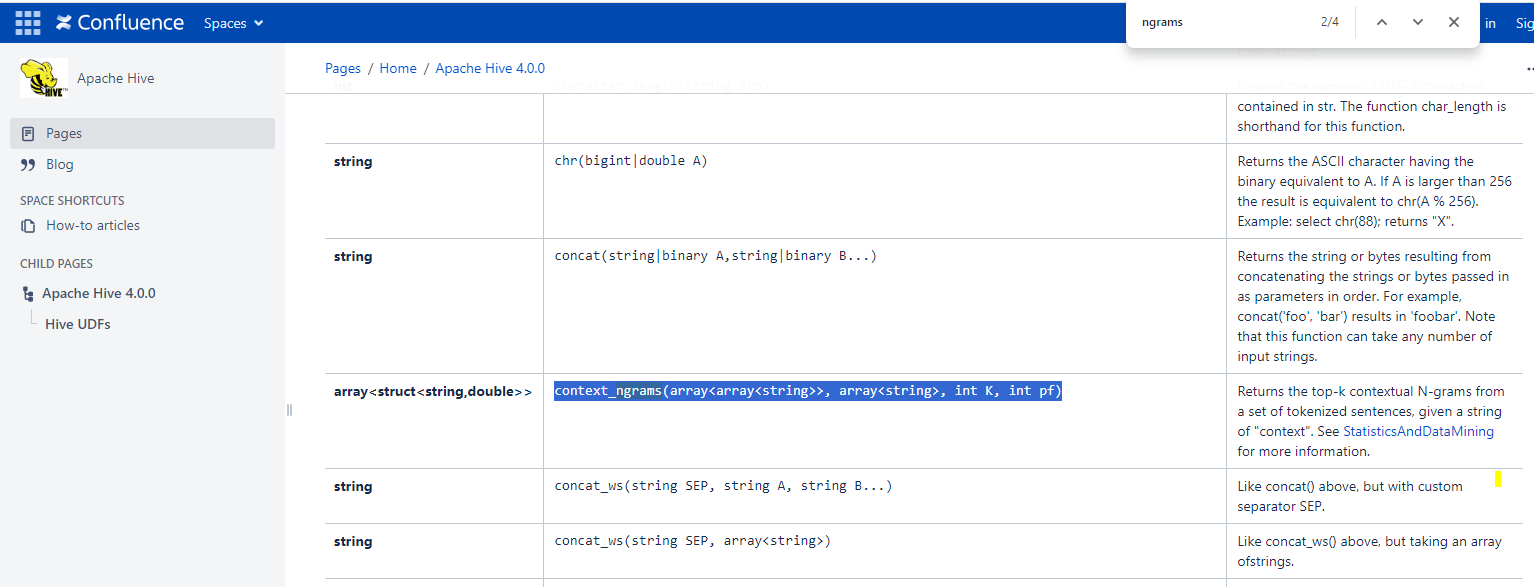
Top 20 words before ‘the’ and after ‘the’



…

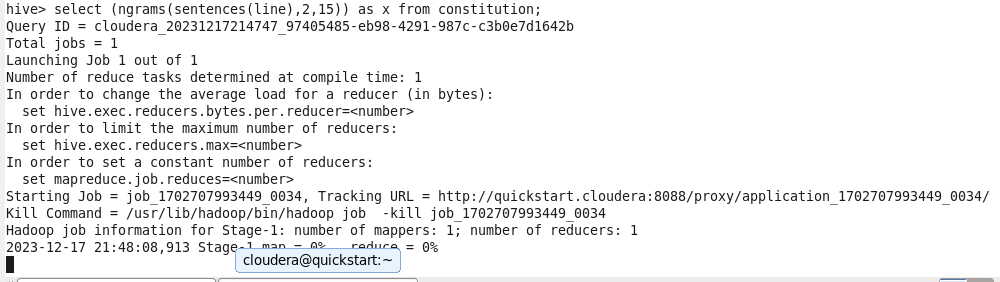




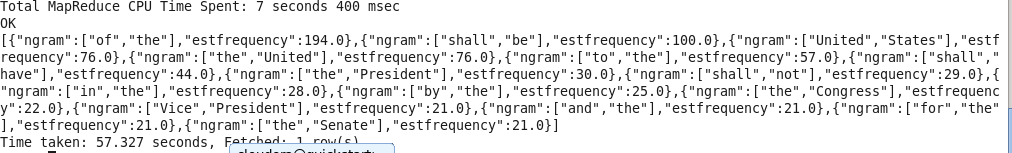


Single line o/p

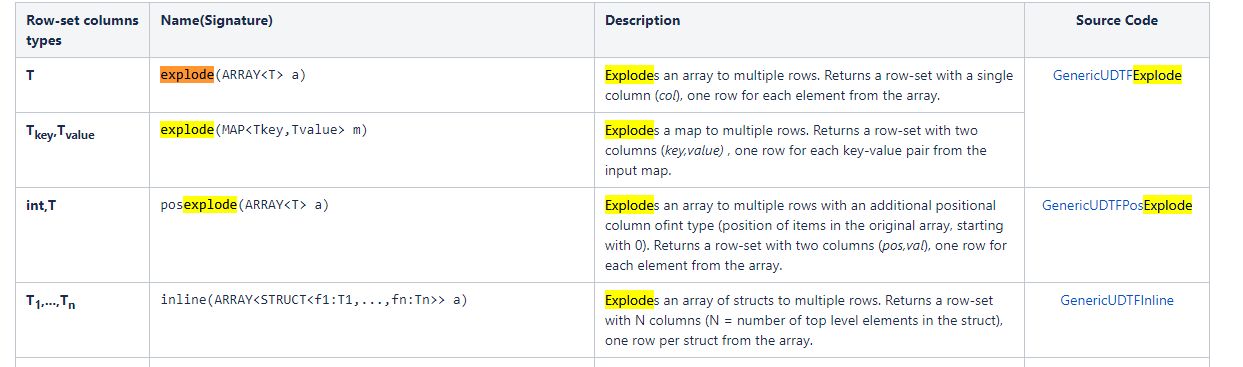
Break explode the outer array and every inner array will come in single line.



…



…

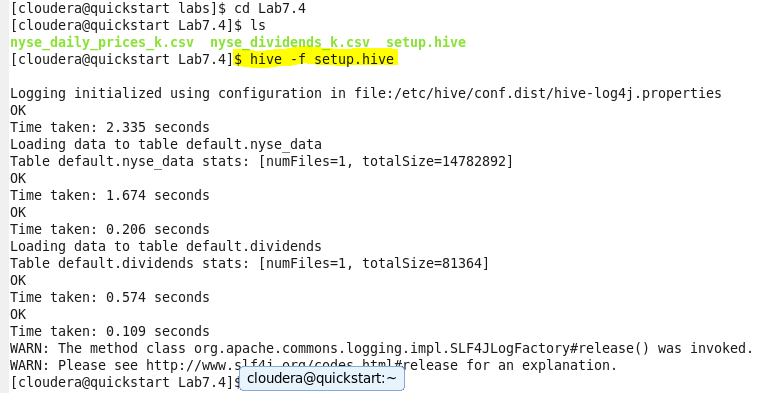


Joining data in Hive

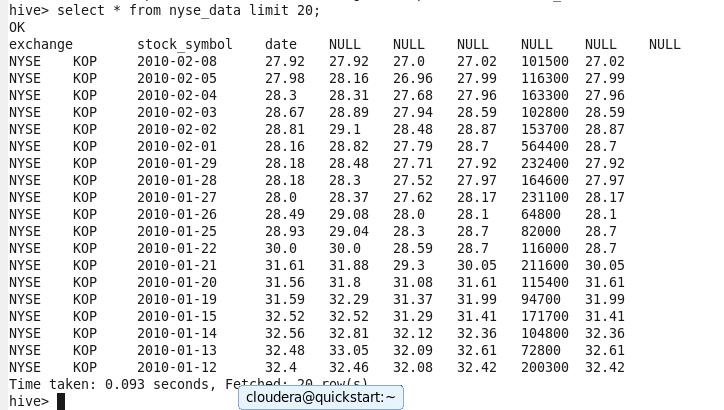
1) Refer lab manual

2) Follow instructions in the lab manual and implement it.

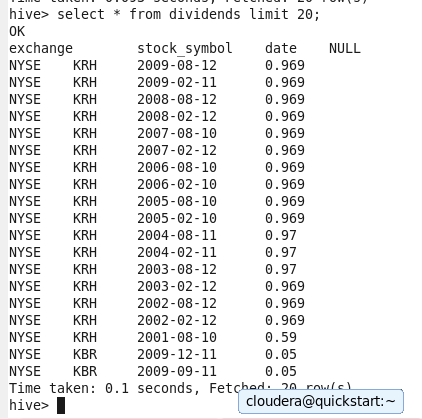
Code/Dataset: data/Lab7.4.rar



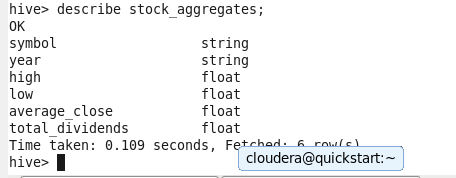
…



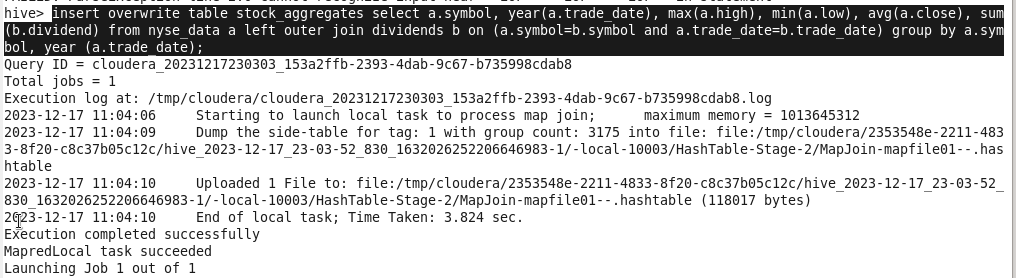
…



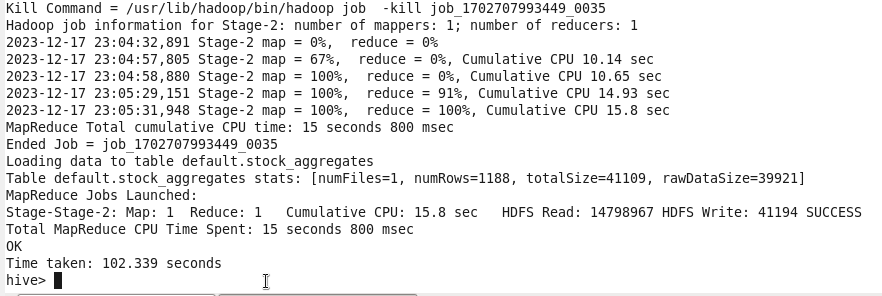
…



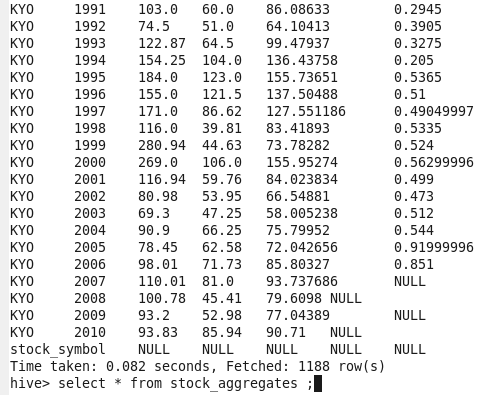
…



…



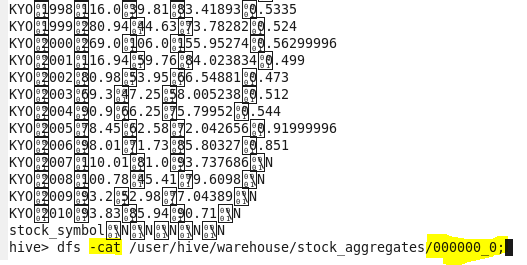
…



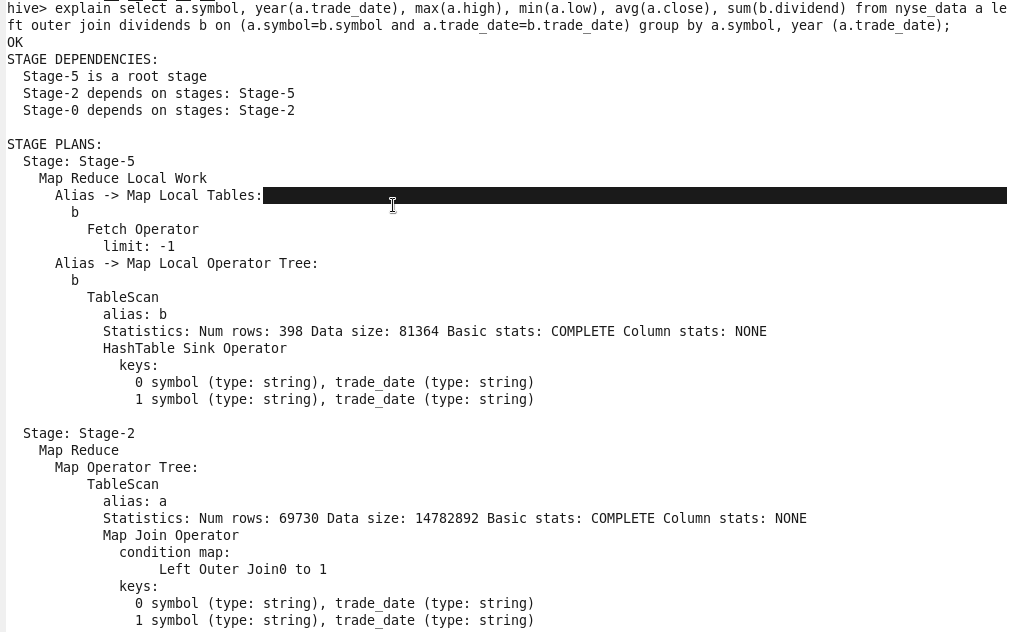
…



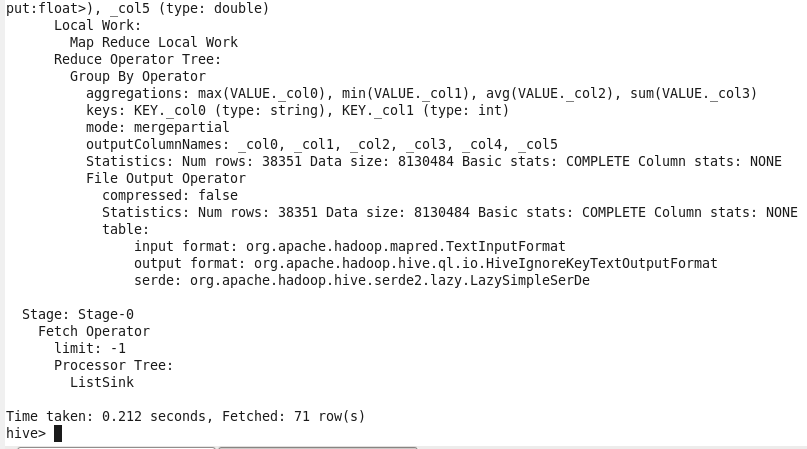
…



…



…



Computing ngrams of Emails in Avro Format

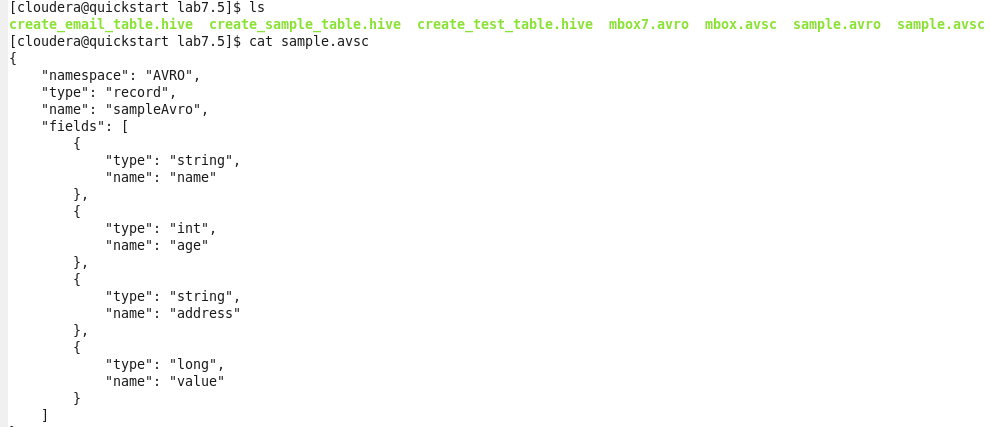
1) Download the rar in **STAGING\_AREA**

2) Move the extracted contents to **LABS\_AREA**/lab7.5

3) Refer the lab manual

4) Wait for the instructions

Code/Dataset :data/Lab7.5.rar

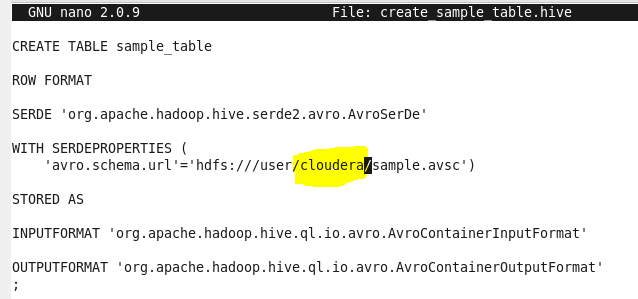


…

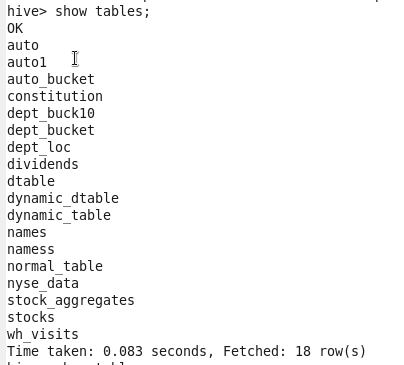


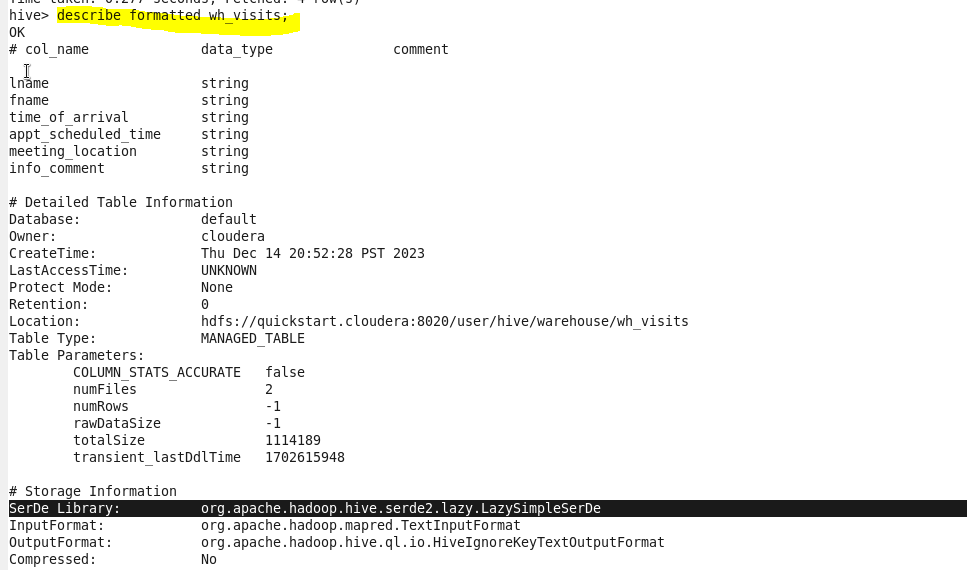
SERDE fully qualified class name,

Change path

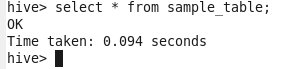


…





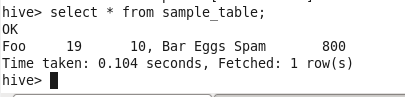
…



…

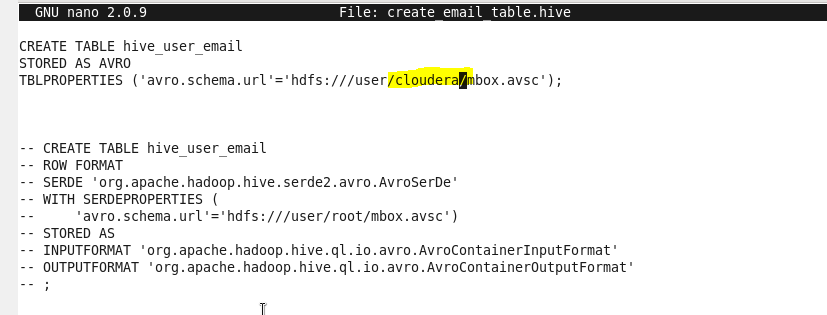


…





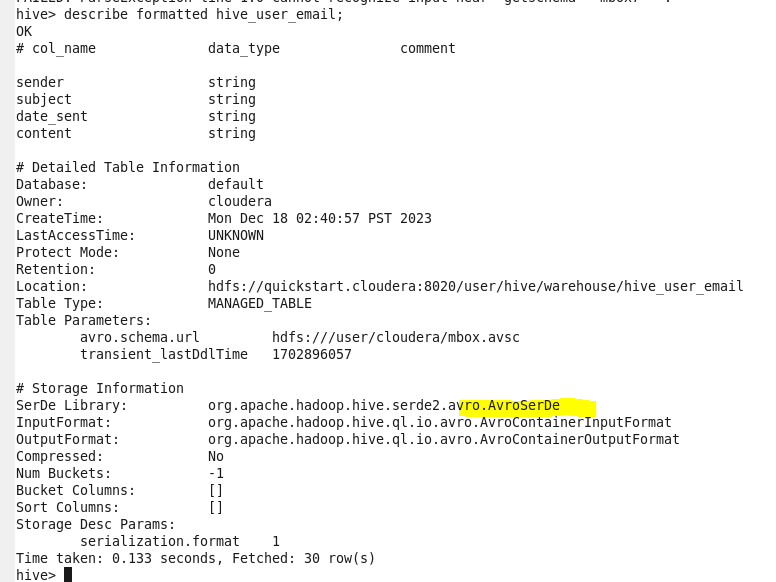
…



…



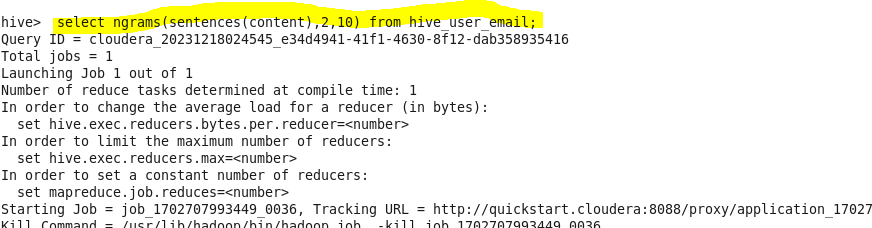
…

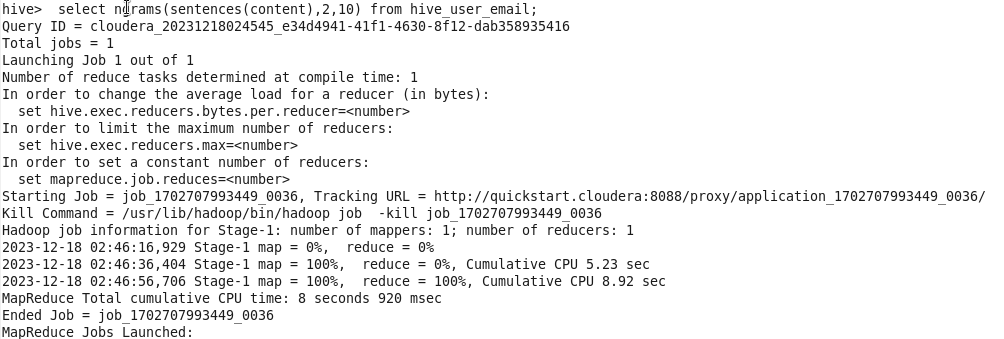


…



…





LAB- ADVANCED HIVE PROGRAMMING

9] Advanced Hive Programming

1) Download the rar on STAGING\_AREA

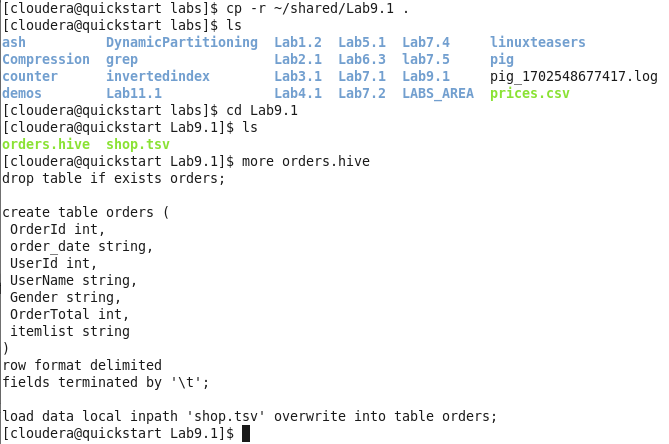
2) Move the extracted contents to LABS\_AREA/lab9.1

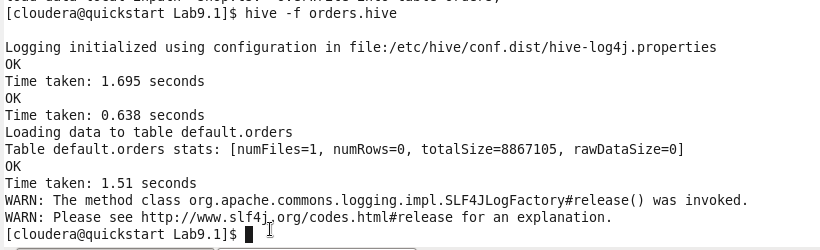
3) Refer the lab manual

* Code/Dataset

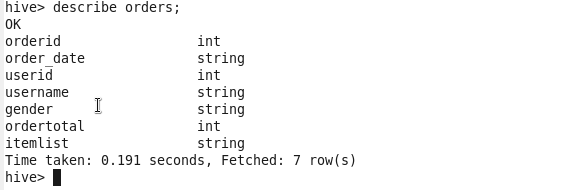
data/Lab9.1.rar

* TimeLine = 60 Minutes

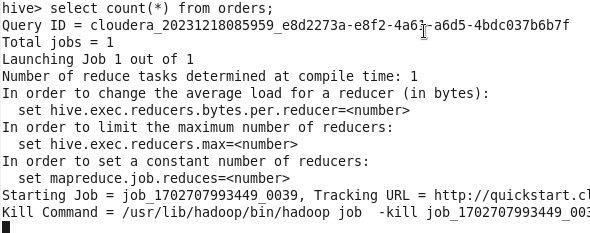




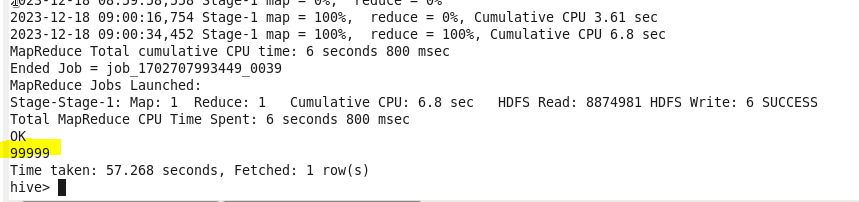
…



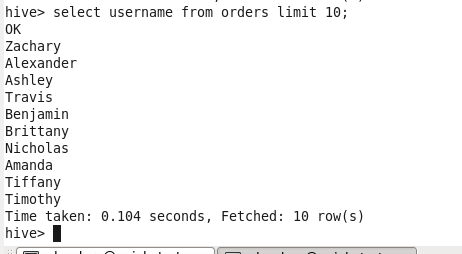
….



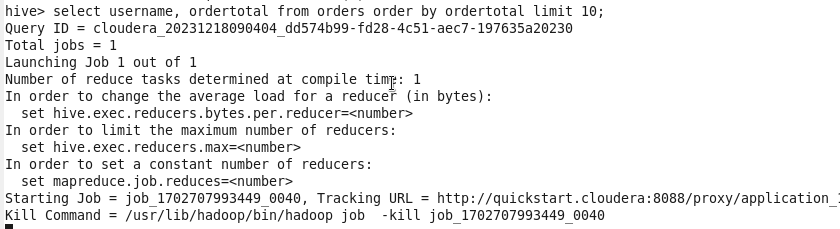
….



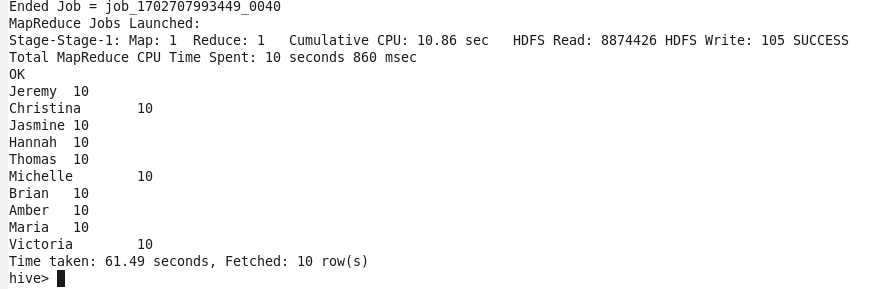
…



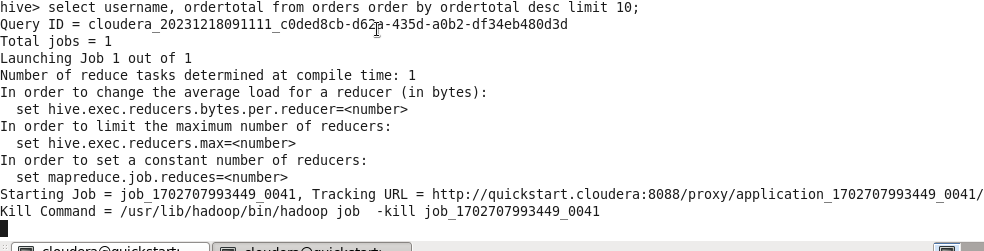
…



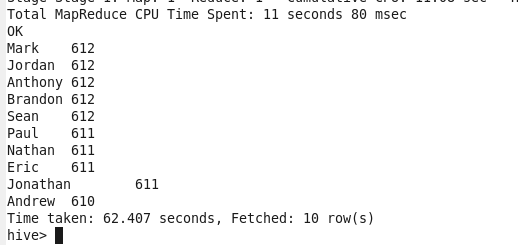
….



..

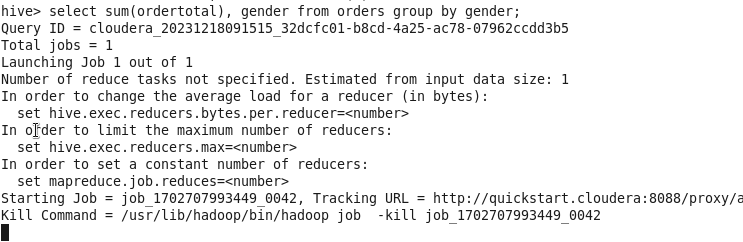


…

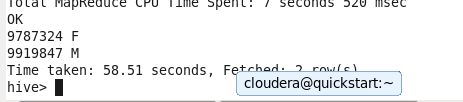


…

Lets find out if men or women spend more money.

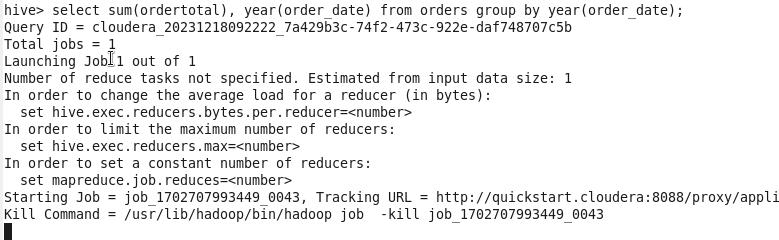


..



.

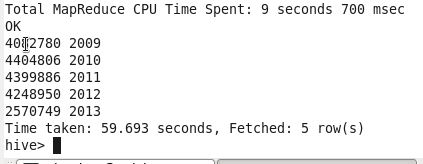
.



.

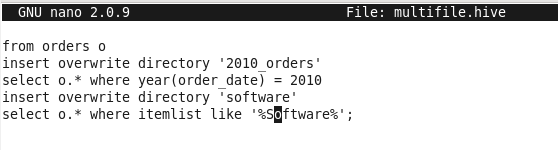
.

.



..

Create file multifile.hive in Lab9.1

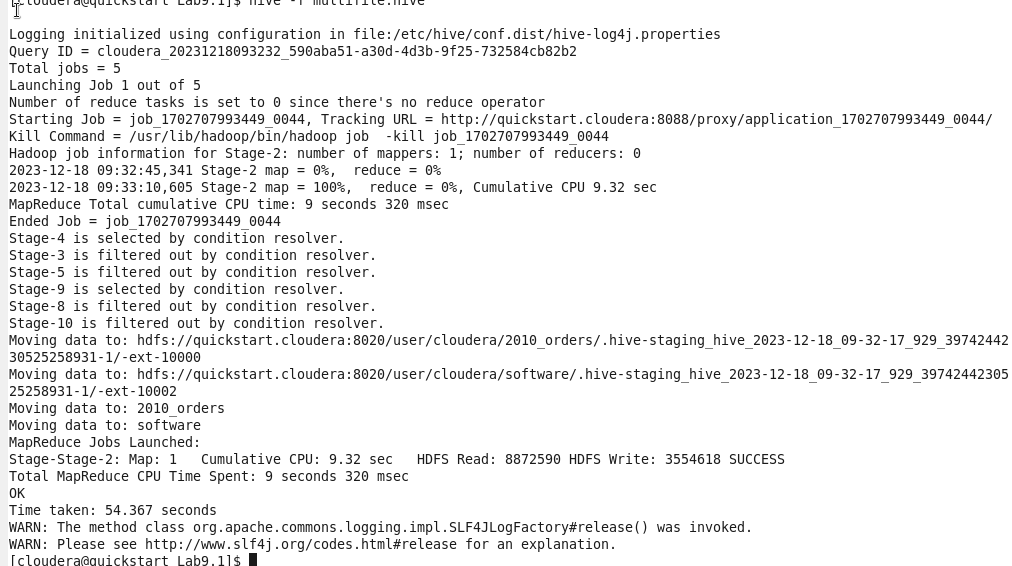
..

Run the file



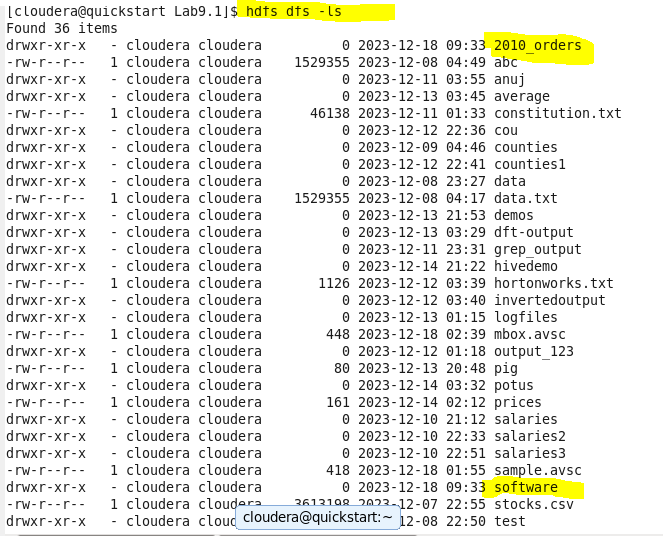
..

Output below

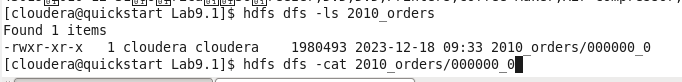


…

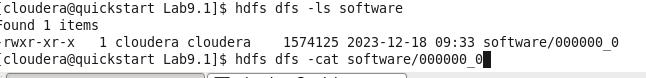
Check you should get two folders



..

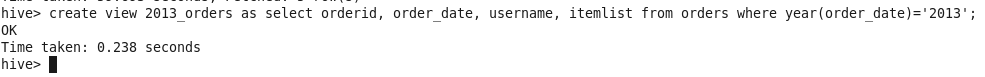


..



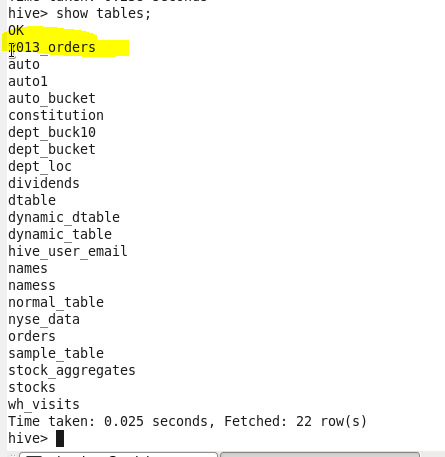
Output will not be human redable

In hive

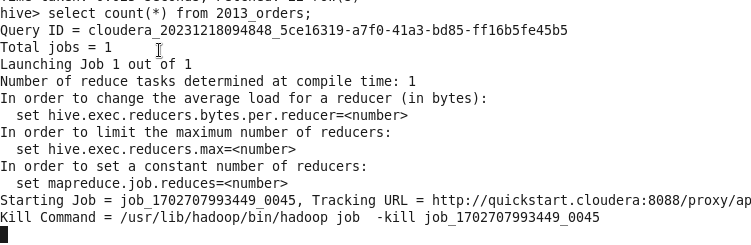


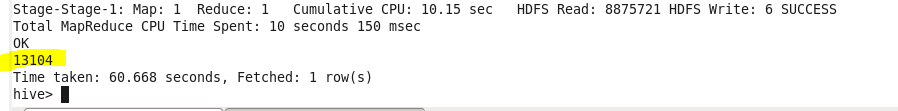
..

You should get 2013\_orders file here

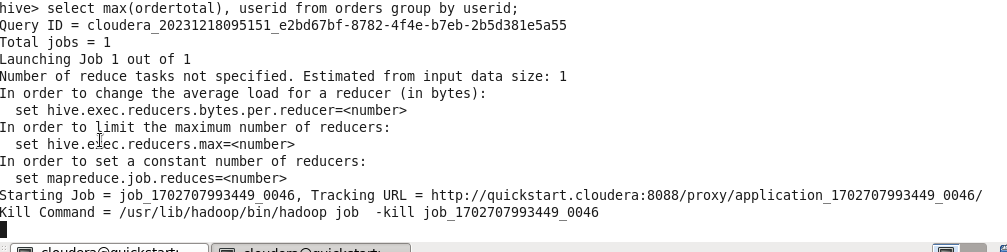


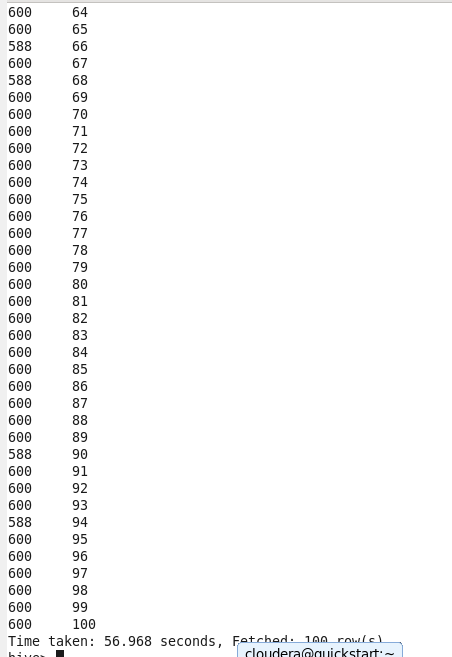
..





6) Find Maximum order of Each Customer



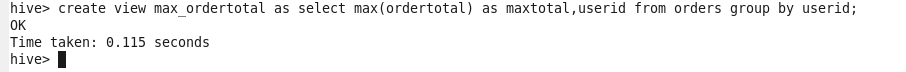


…

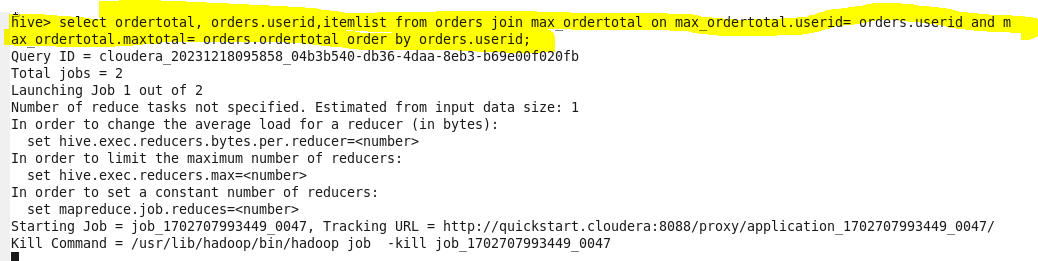


This querry is not valid as itemlist is not in the group by key.

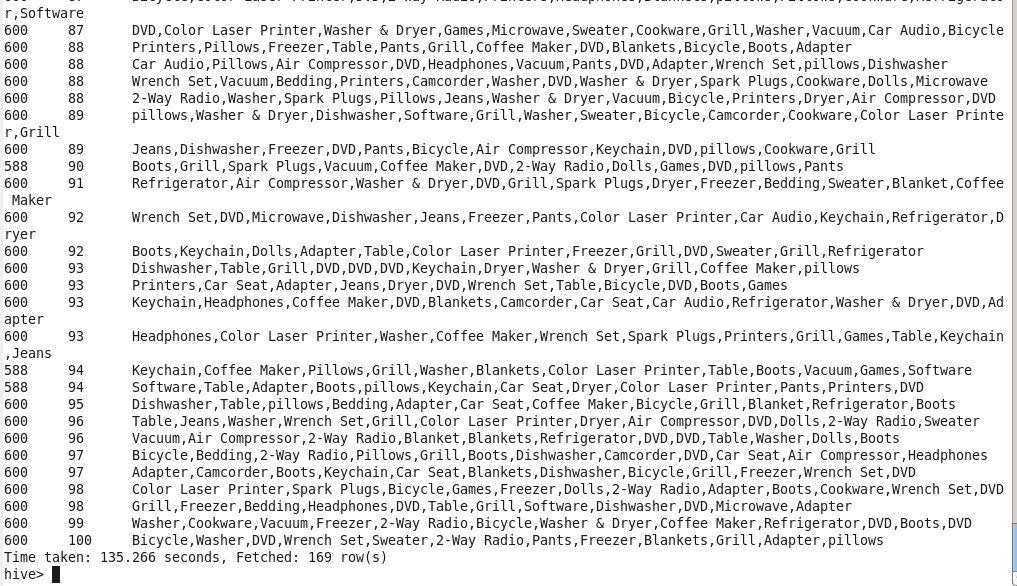
..



..

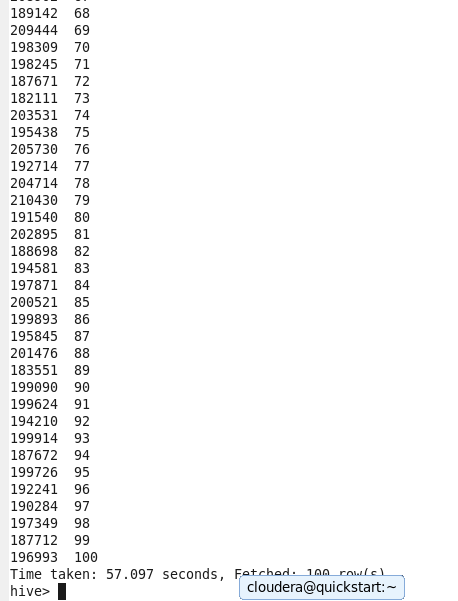


.. output

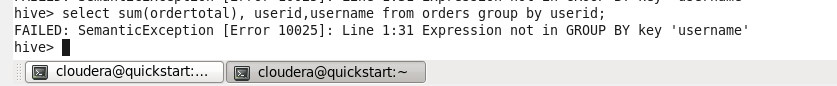


**.**





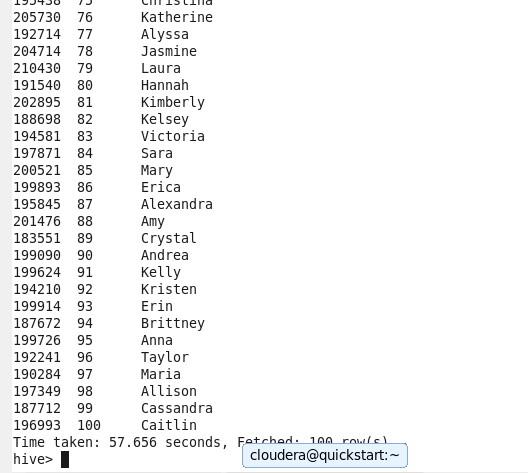
**..**



**Will give error**

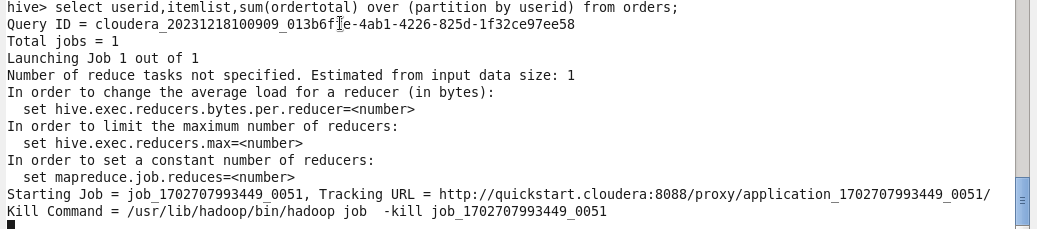


**..**

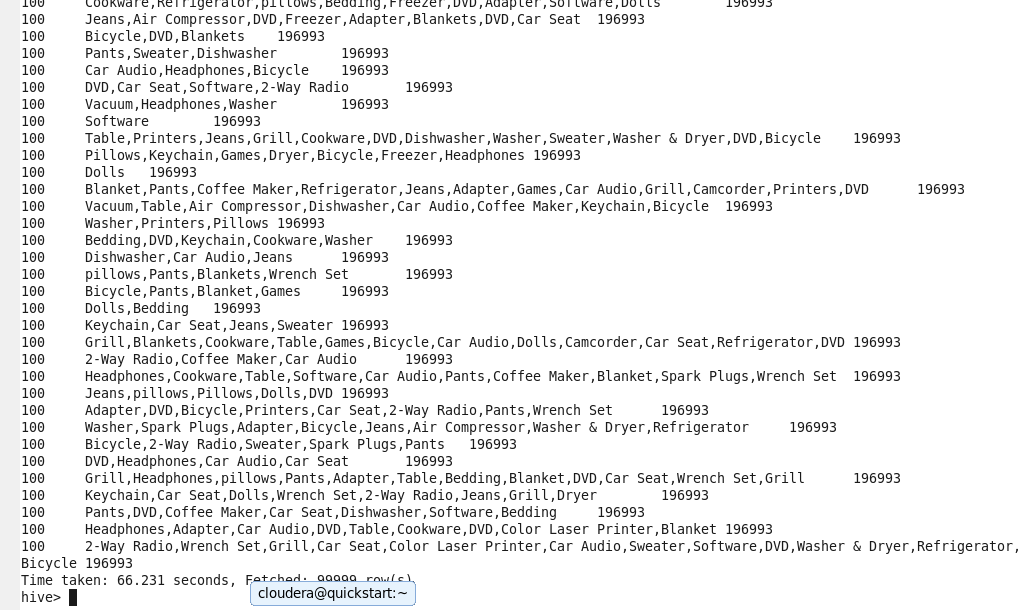


**..**

**7) using the over clause**

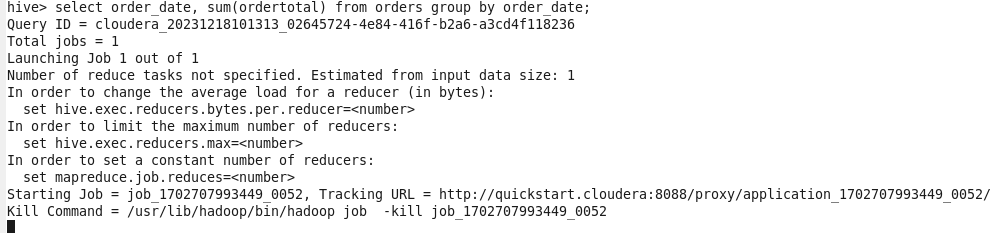


**..**

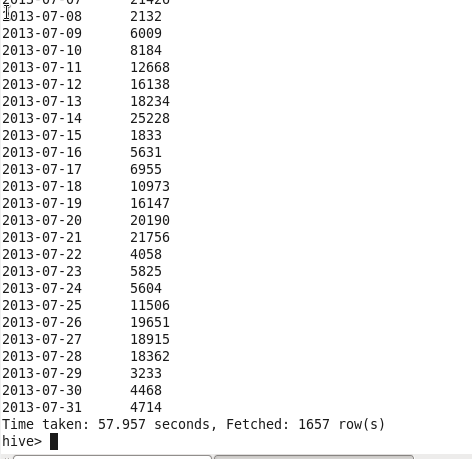


**..**

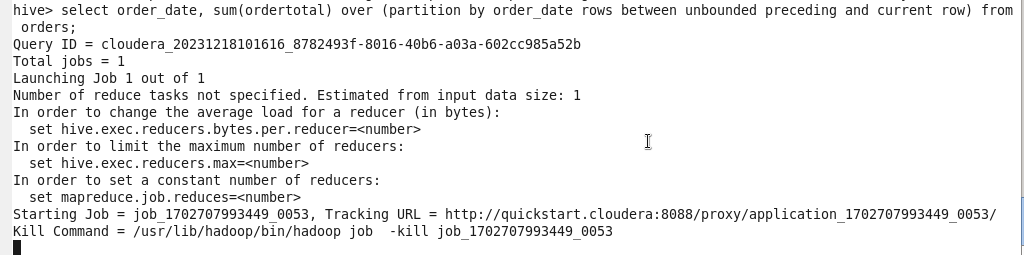
**8) Using the window Functions**



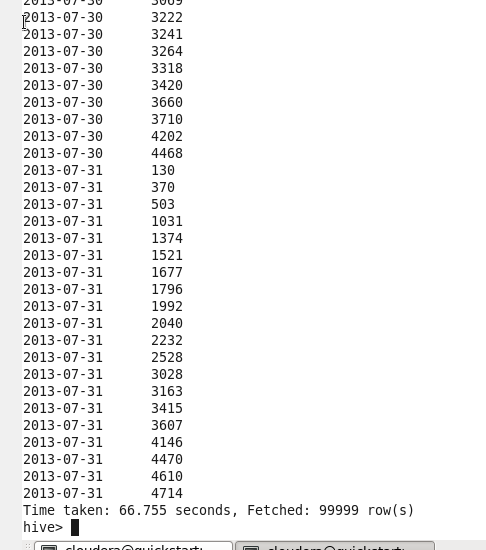
**…**



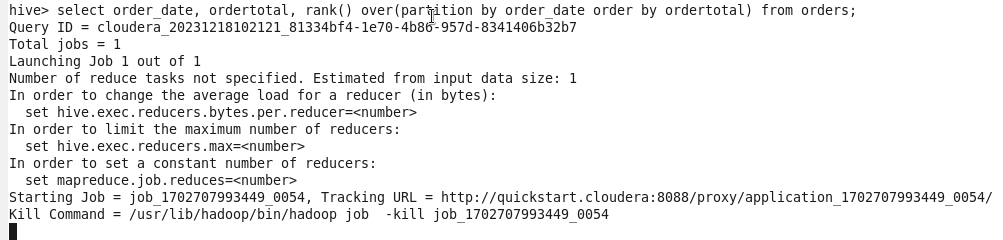
**..**



**..**



**9) Using the Hive Analytics Functions**



**…**

**b)**

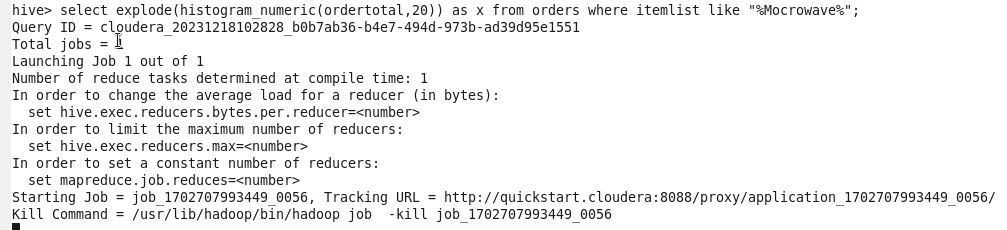




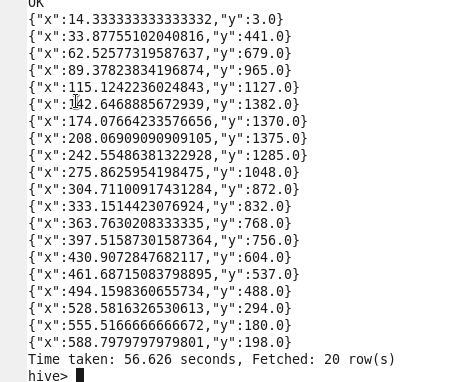
**.**



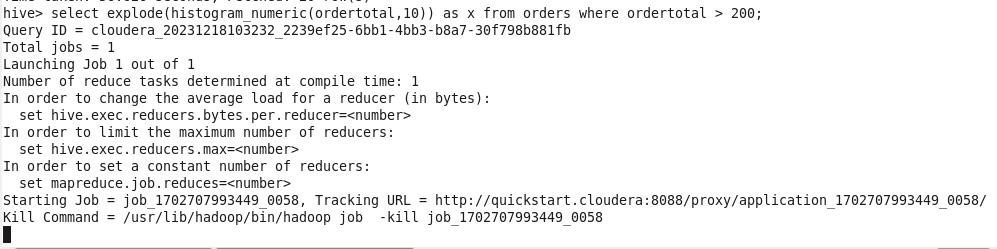
**10) Histograms**



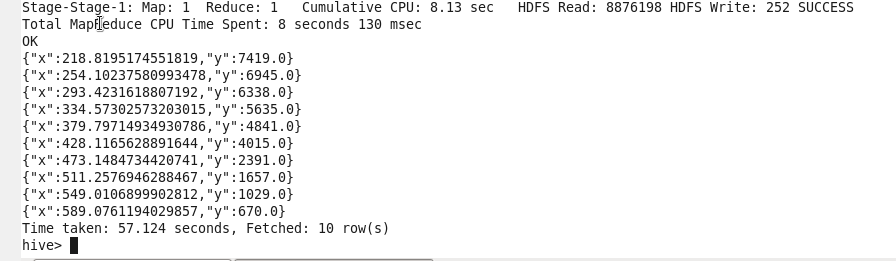
**.**



**..**

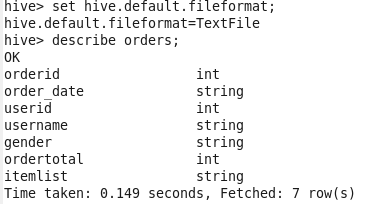


**..**

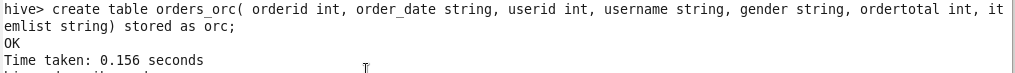


**…**

**Trying orc files**

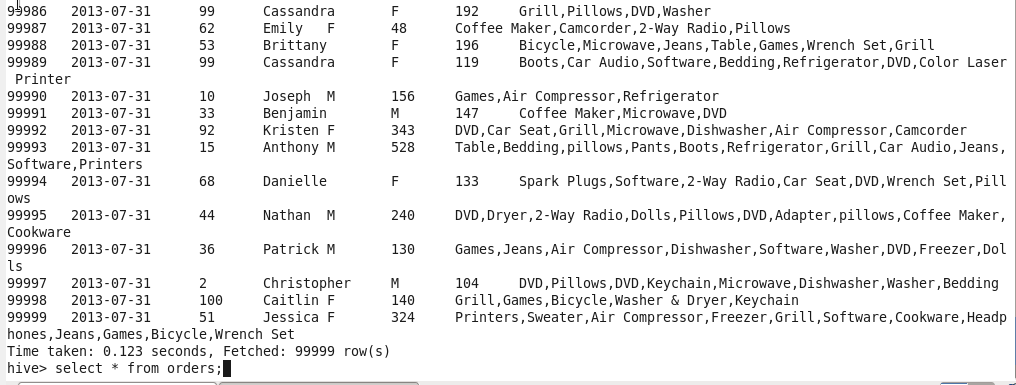


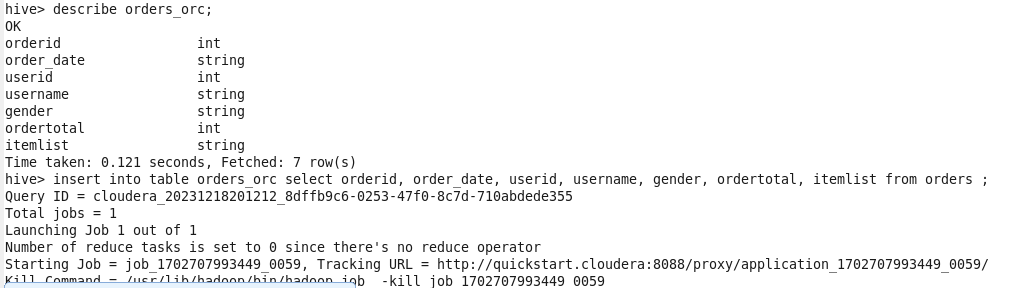
**…**





**…**

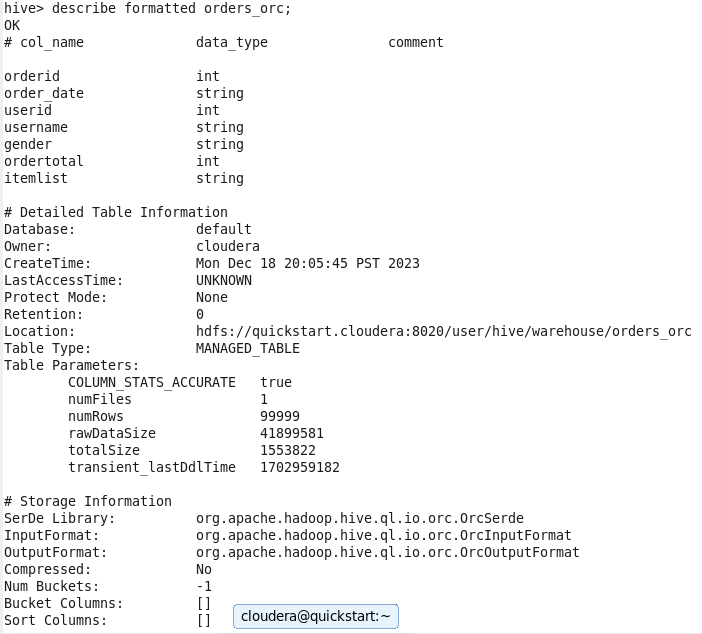




**…**



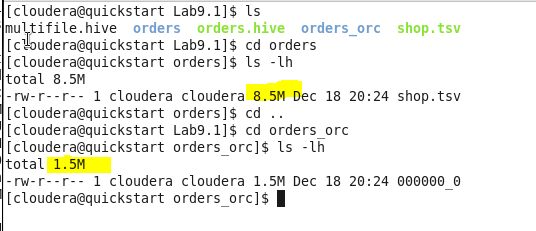
**…**



**…**



**…**



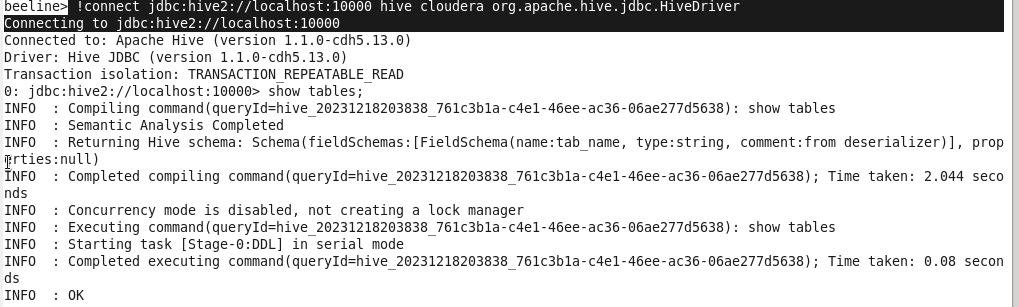
**Here orc file has less memory storage as compared to text file.**



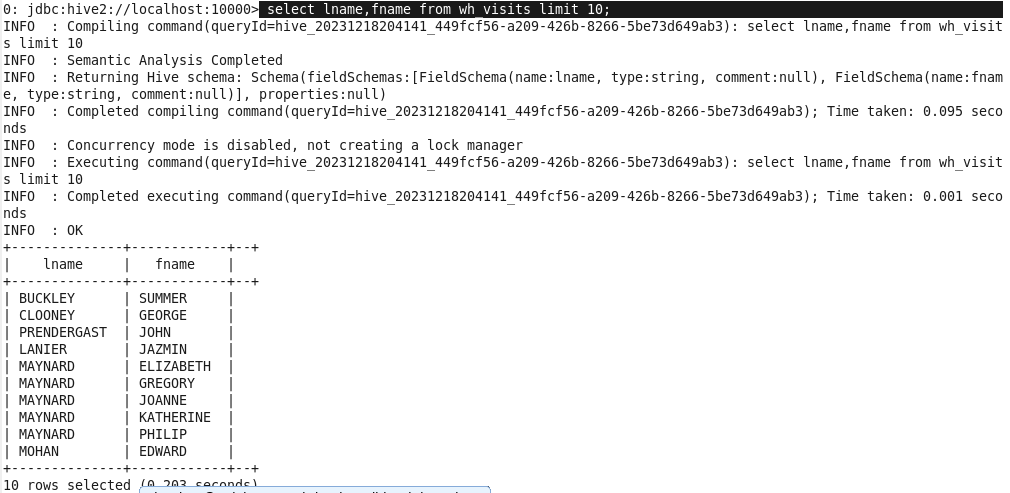
**…**



**…**



**…**



**…**

