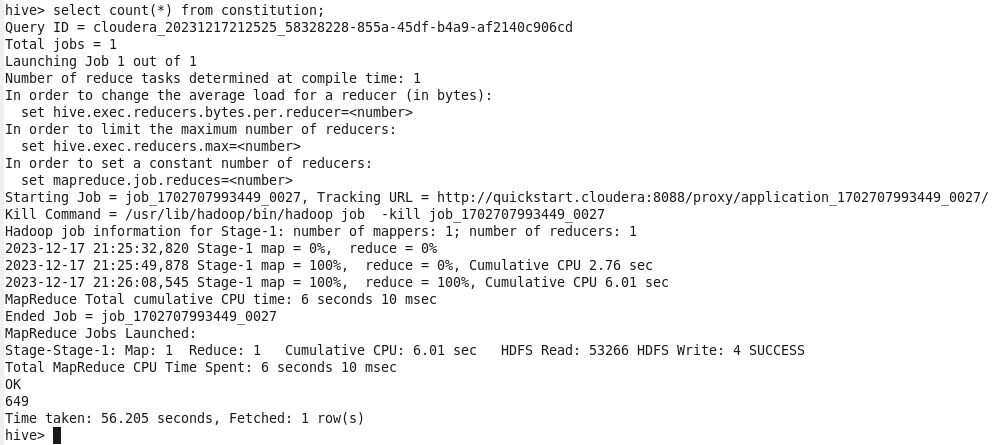


…

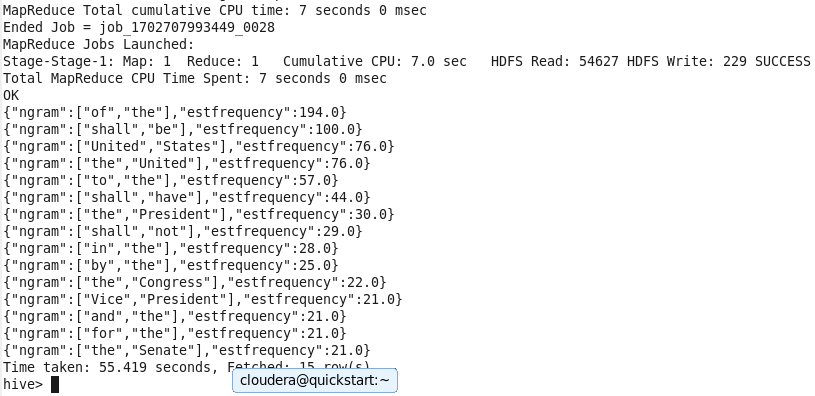
…



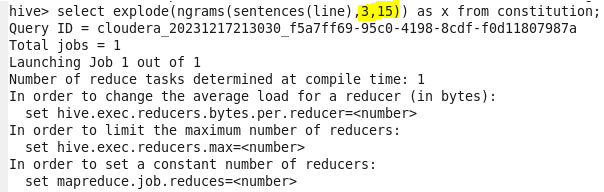
…



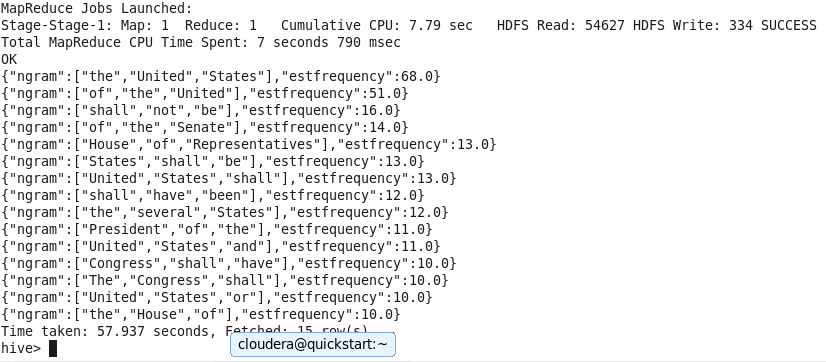
…



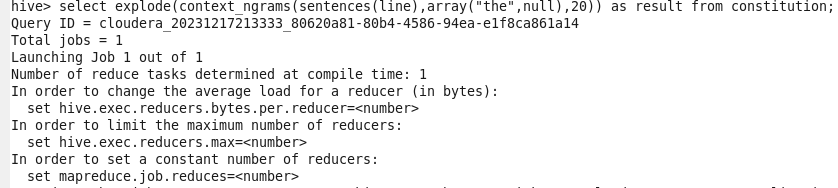
…



…

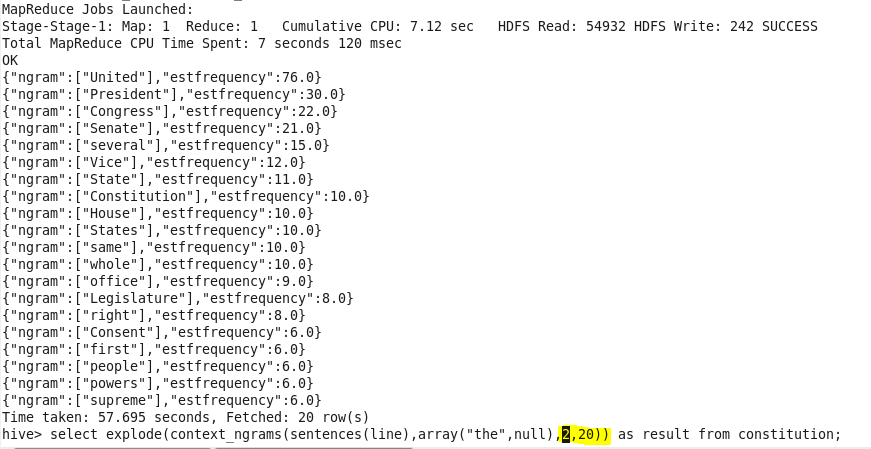


…

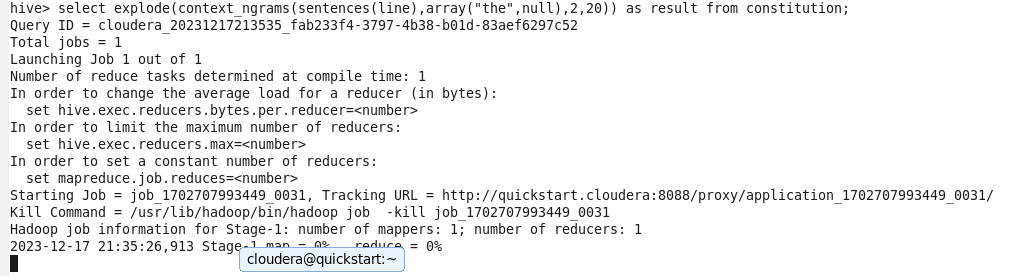


…

Top 20 bigrams after “the”

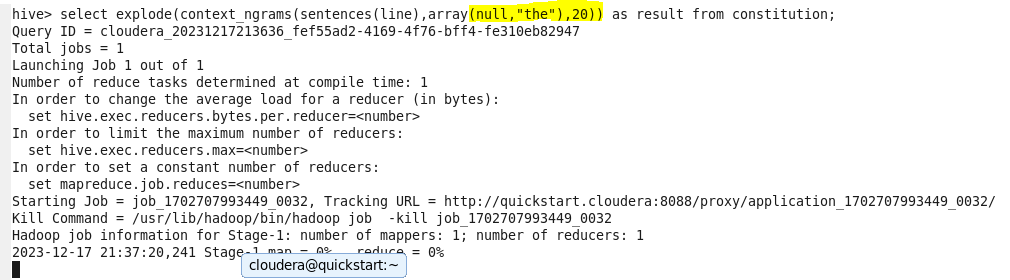


…

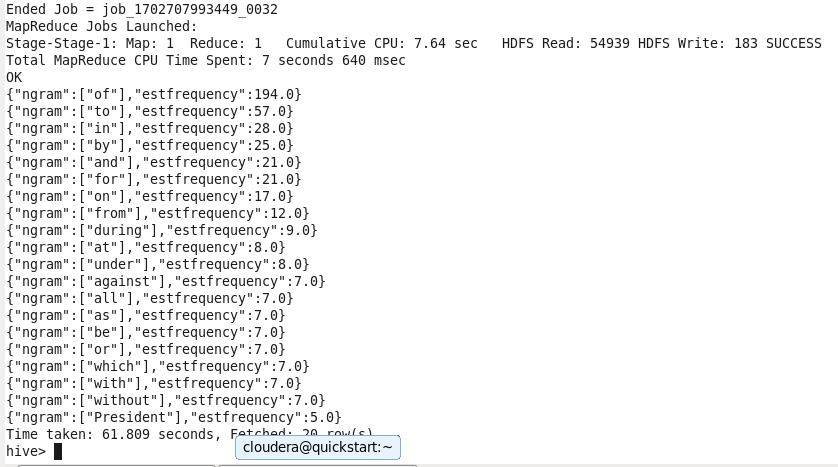


…

Top 20 words before “the”

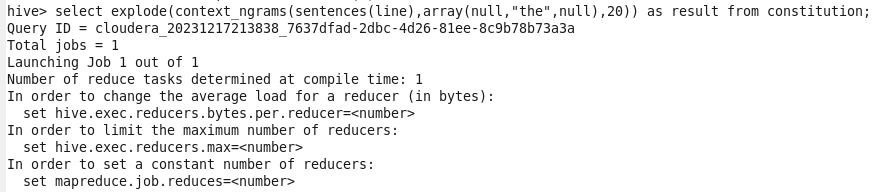


…

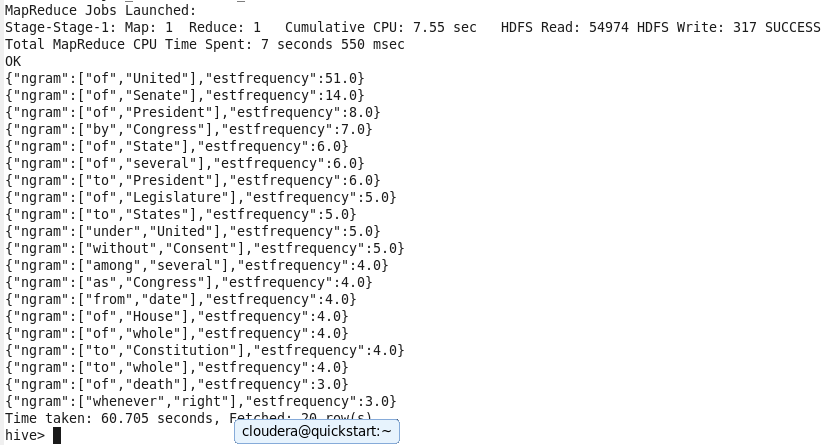


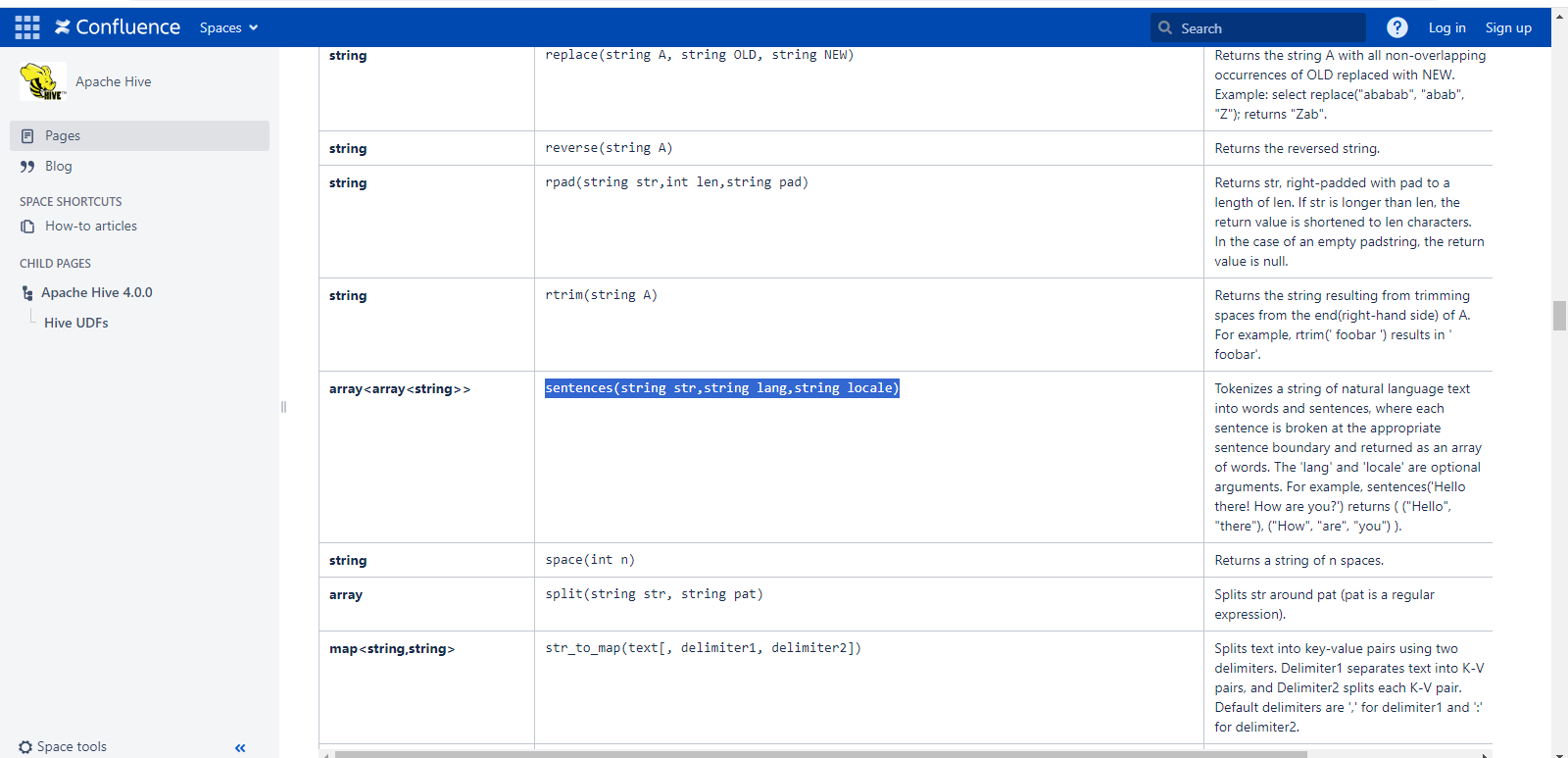
…

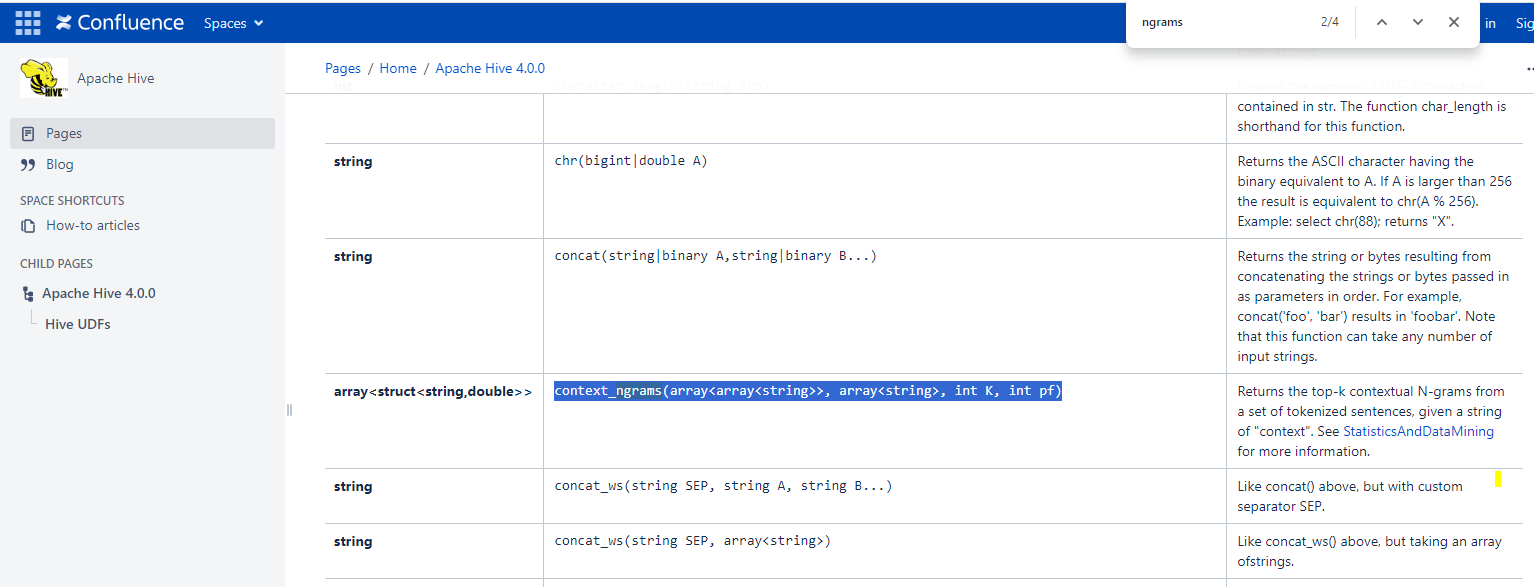
Top 20 words before ‘the’ and after ‘the’



…





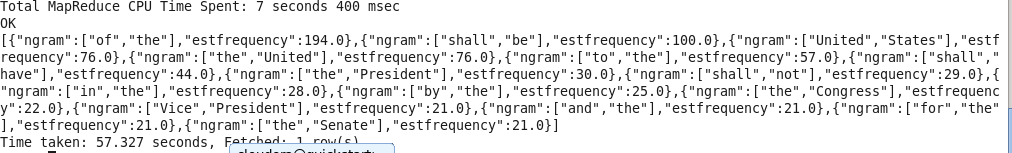


Single line o/p

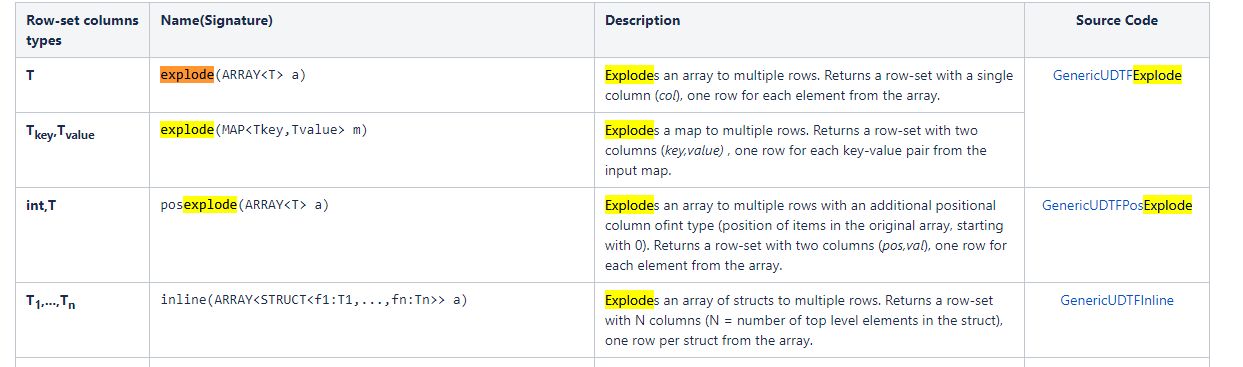
Break explode the outer array and every inner array will come in single line.



…



…

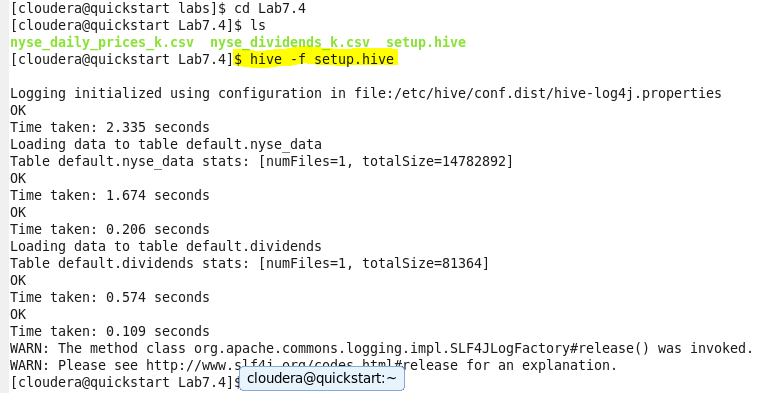


Joining data in Hive

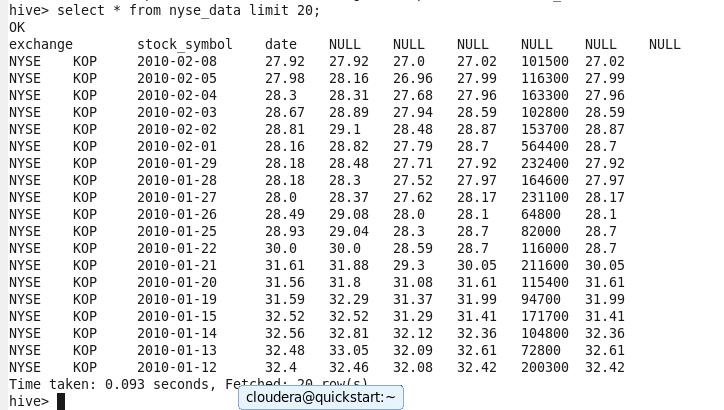
1) Refer lab manual

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

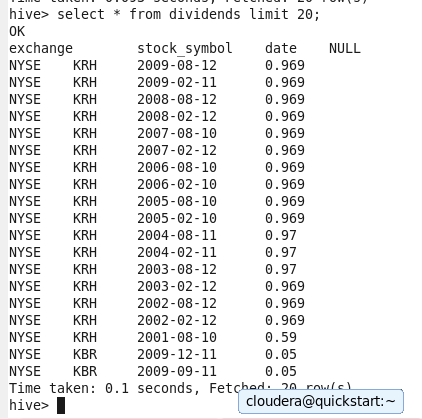
Code/Dataset: data/Lab7.4.rar



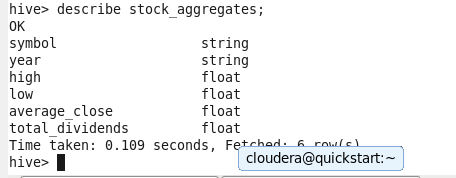
…



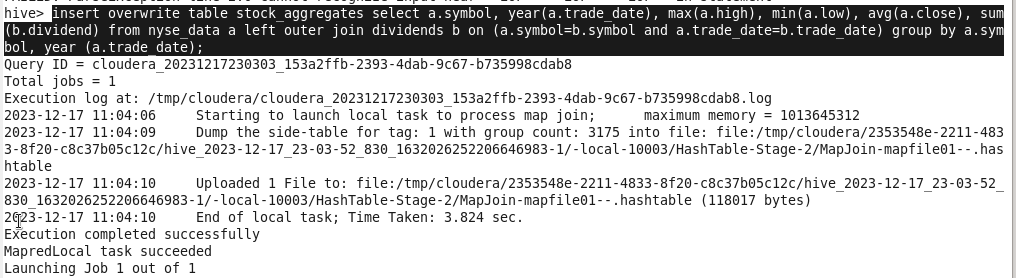
…



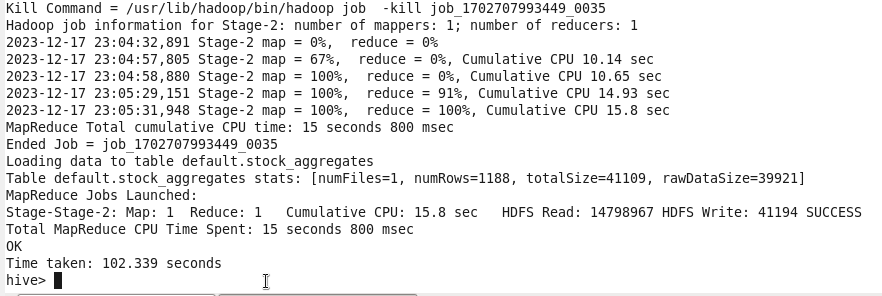
…



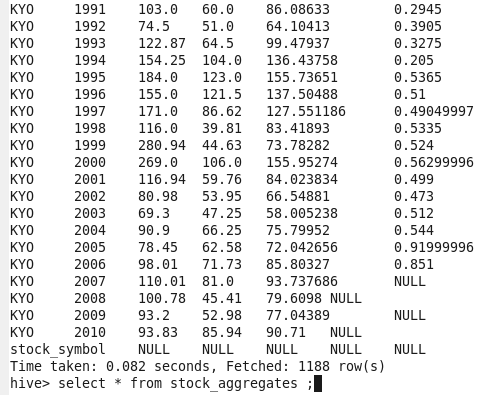
…



…



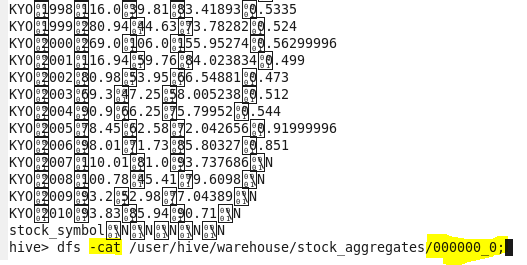
…



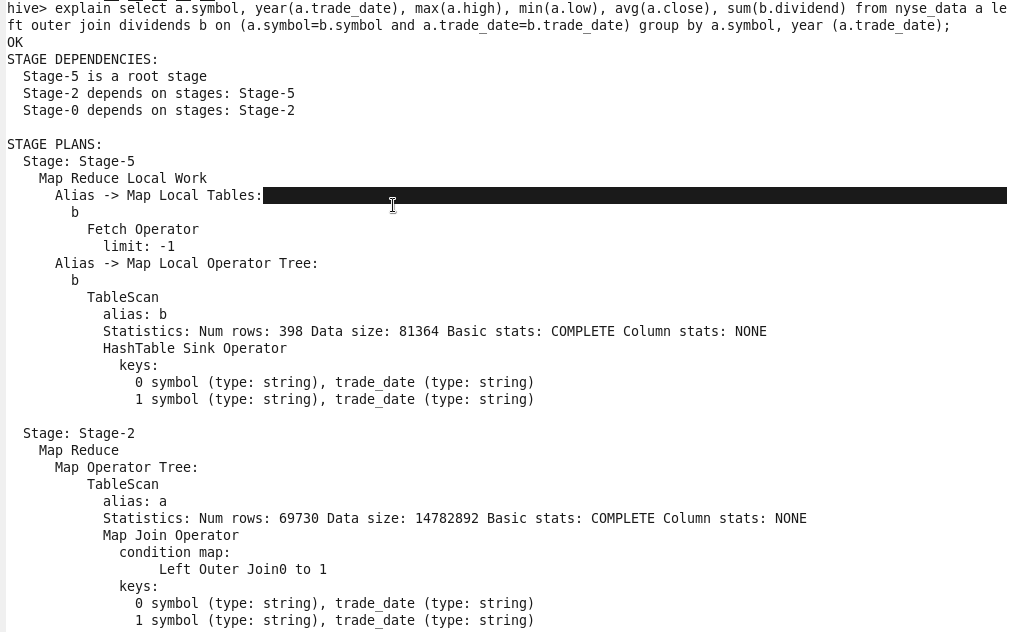
…



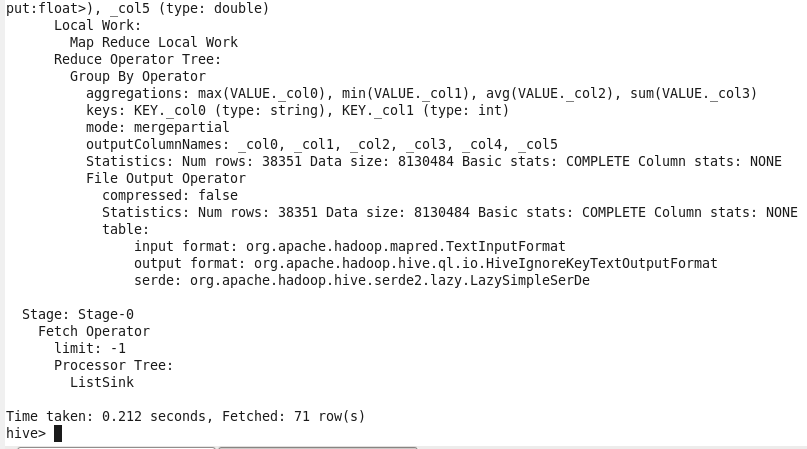
…



…



…



Computing ngrams of Emails in Avro Format

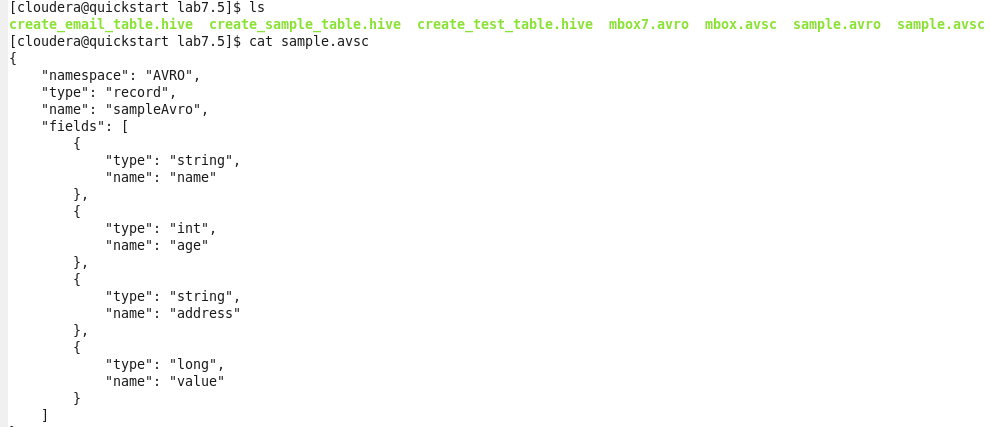
1) Download the rar in **STAGING\_AREA**

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

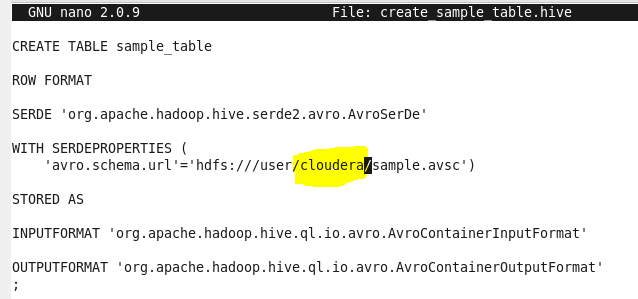


…

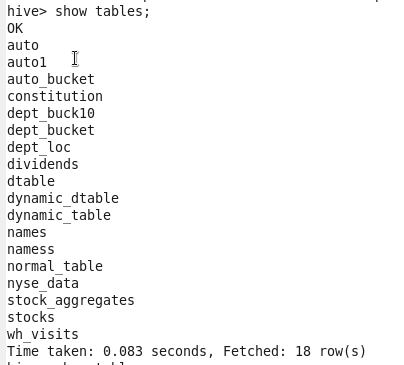


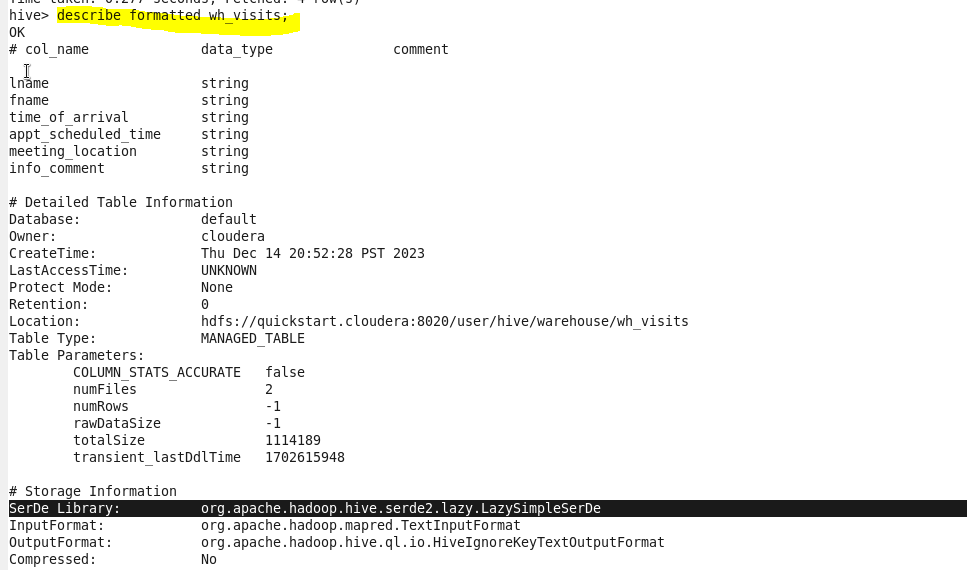
SERDE fully qualified class name,

Change path

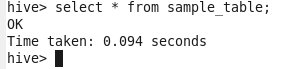


…





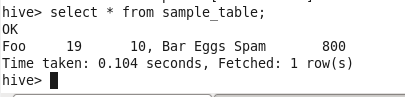
…



…

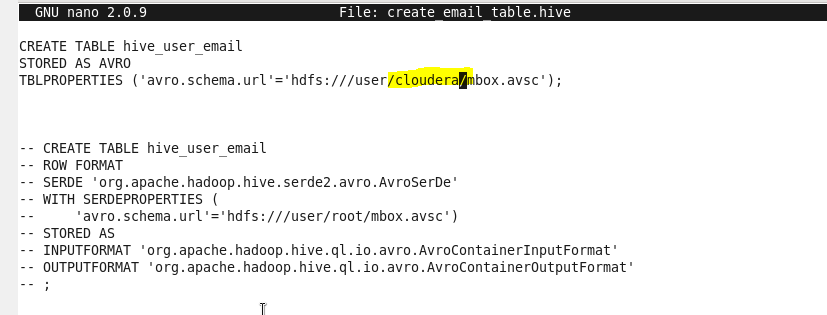


…

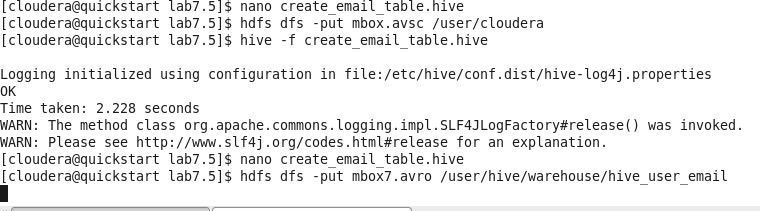




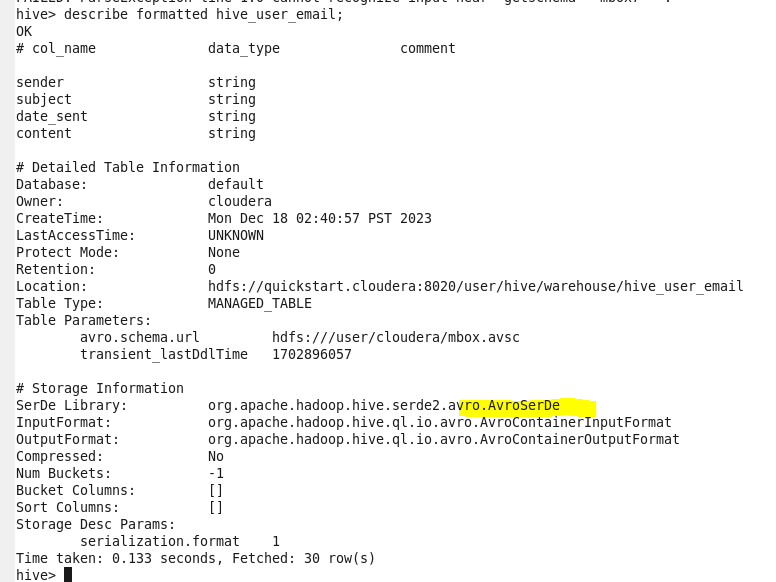
…



…



…



…



…

