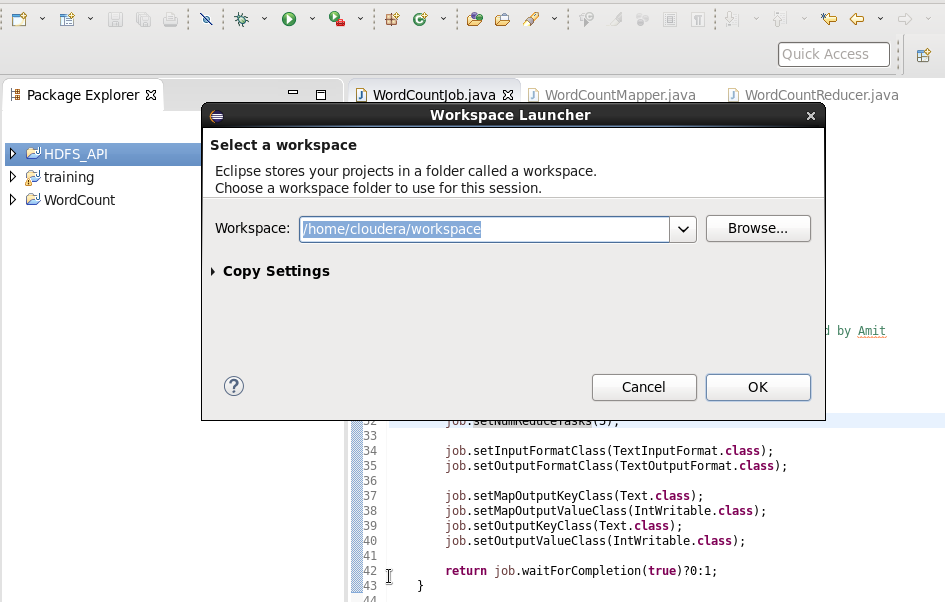
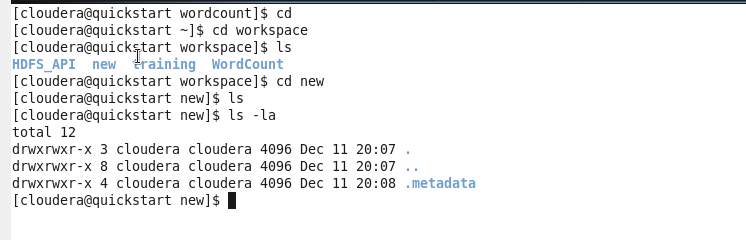
Workspace



Created new workspace



3] Distributed GREP

Here attached the rar file. Perform following activities

1) Download the rar file in **STAGING\_AREA**, extract it

2) Create a folder grep in **LABS\_AREA**. Import the java files from step 1 to this folder

3) Create a project **grep** in Eclipse

4) Import the java file from step 2 in the same

5) Resolve the compilation errors by adding hadoop client side libraries in class path of the project

6) Implement all **TODO**s in the code

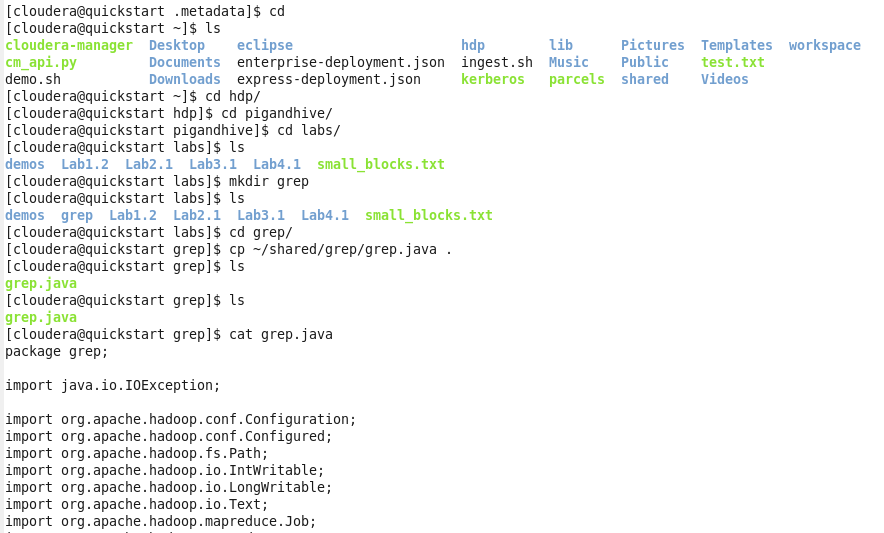
7) Look at the comments at the end of the file and create a jar and execute it

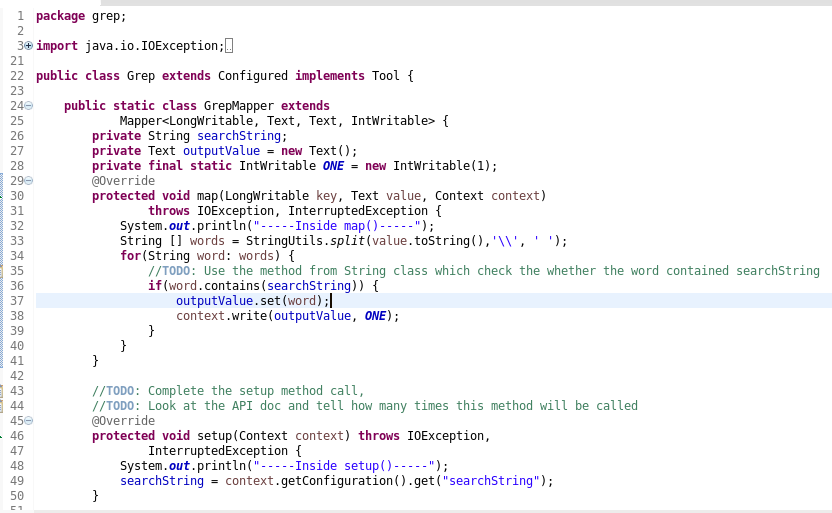
8) During implementation of the code refer to the API documentation.

9) What differences are you finding in this code and previous Word Count code ?



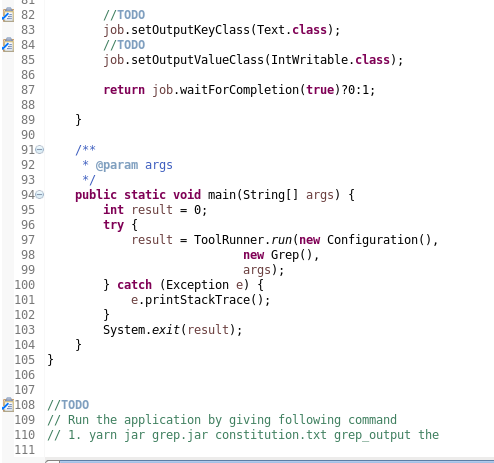
.





.

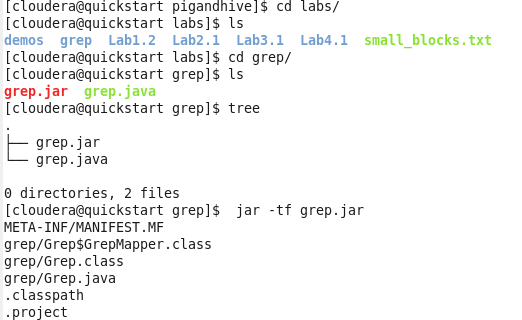


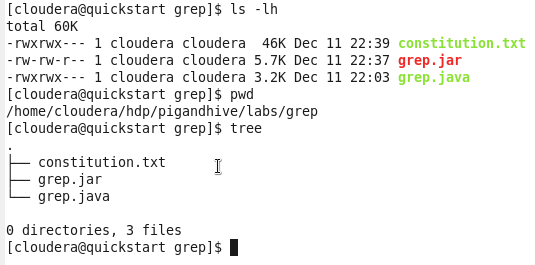


…

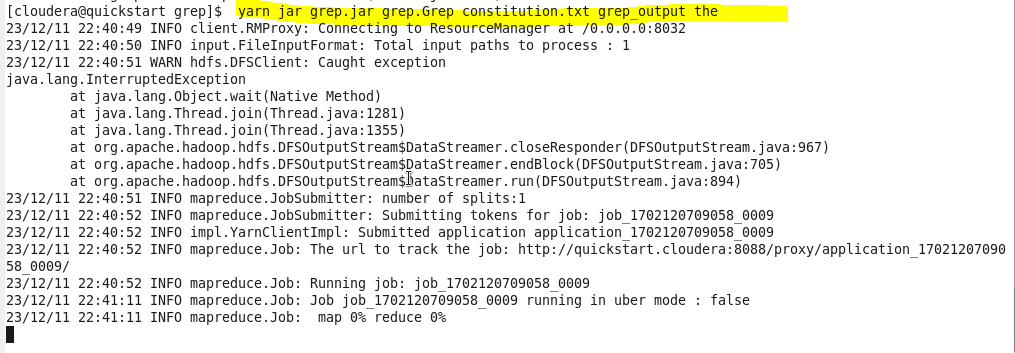
To get output

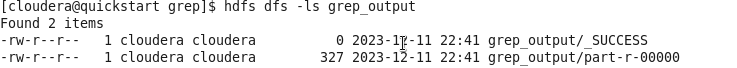
…

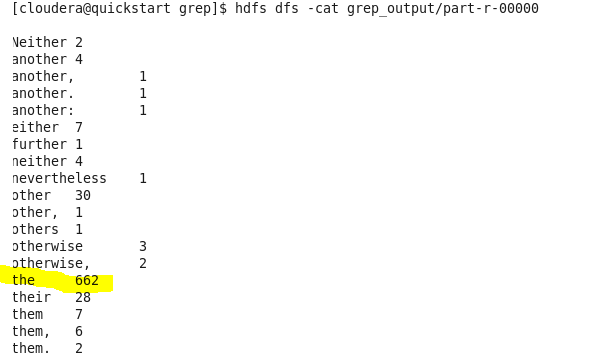




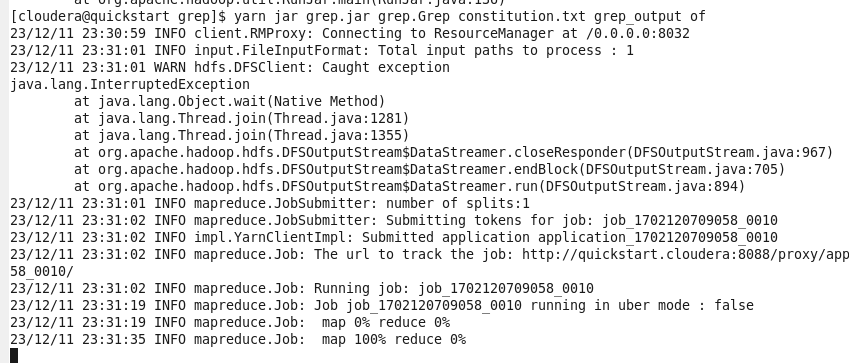
…

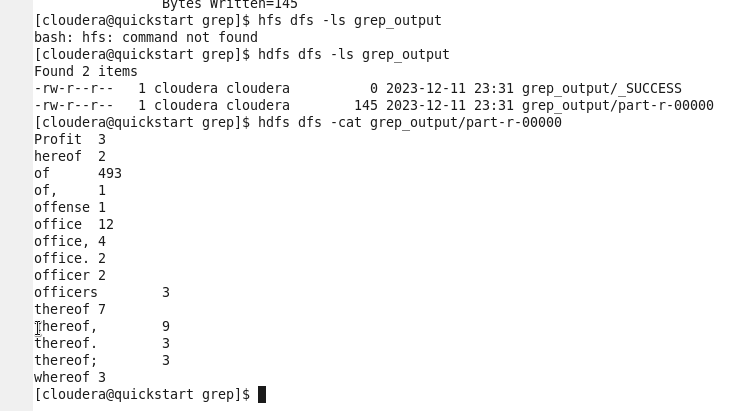
…



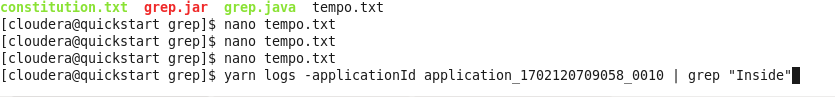


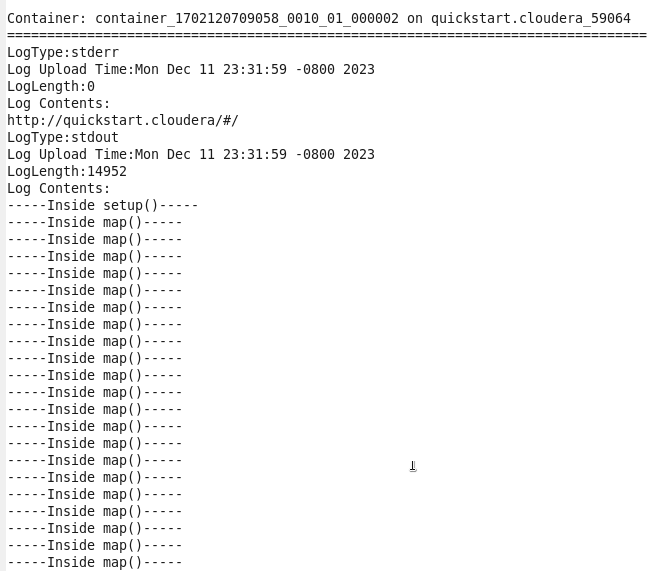
2. Here we are checking output for “of” i.e constitution.txt contains “of” word how many times.





…





LAB 4

4] Inverted Index

Here attached the rar file. Perform following activities

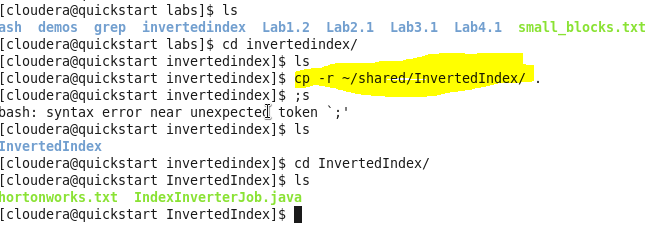
1) Download the rar file in **STAGING\_AREA** location

2) Create a folder invertedindex in **LABS\_AREA**.

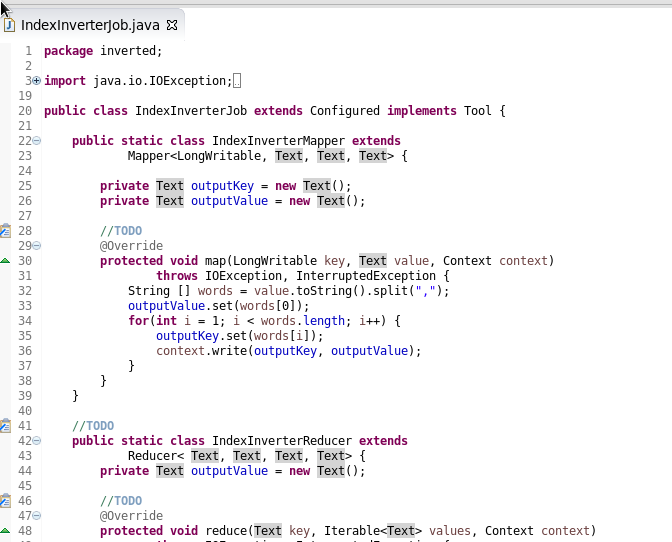
3) Create a project **invertedindex** in Eclipse

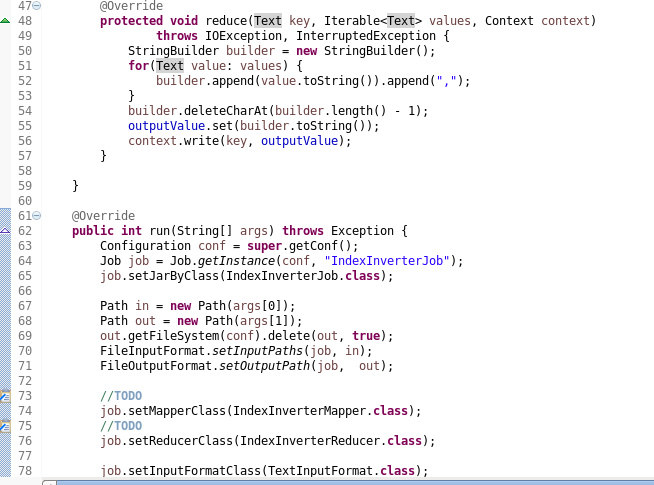
4) Import the java file in the same

5) Resolve the compilation errors by adding hadoop client side libraries in class path of the project

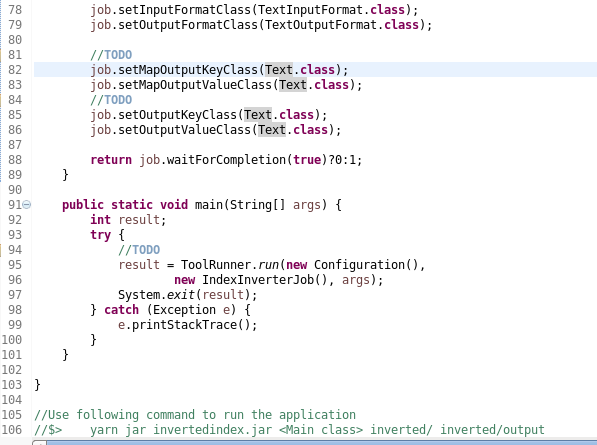


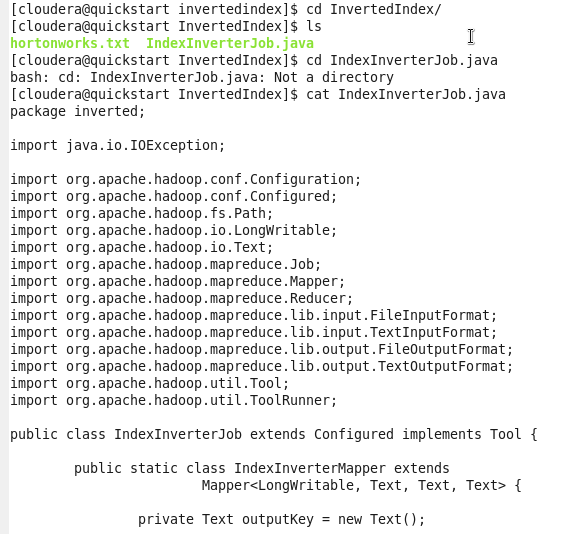
.





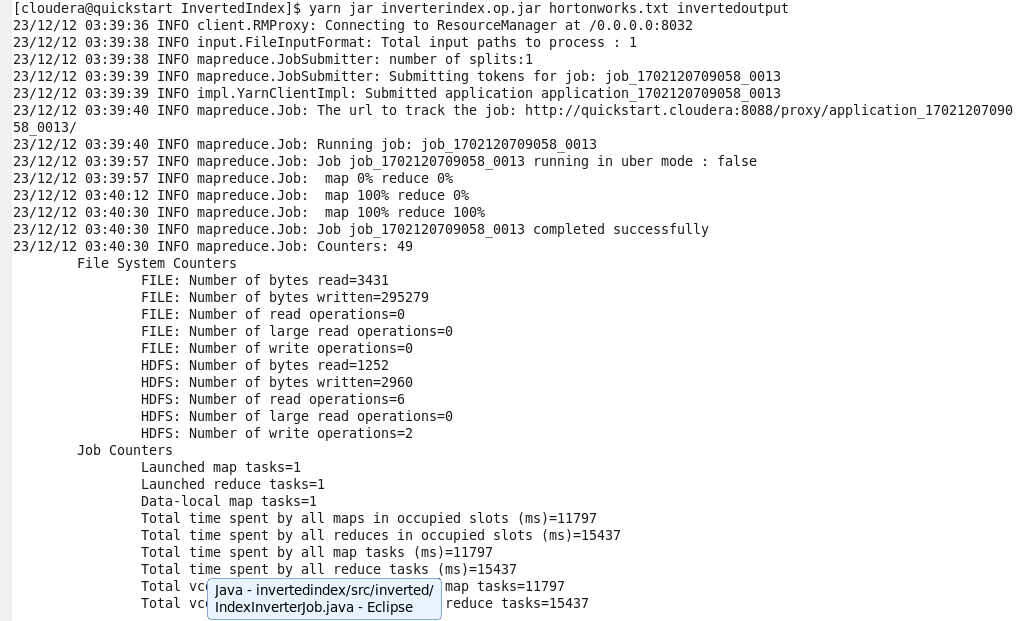
.

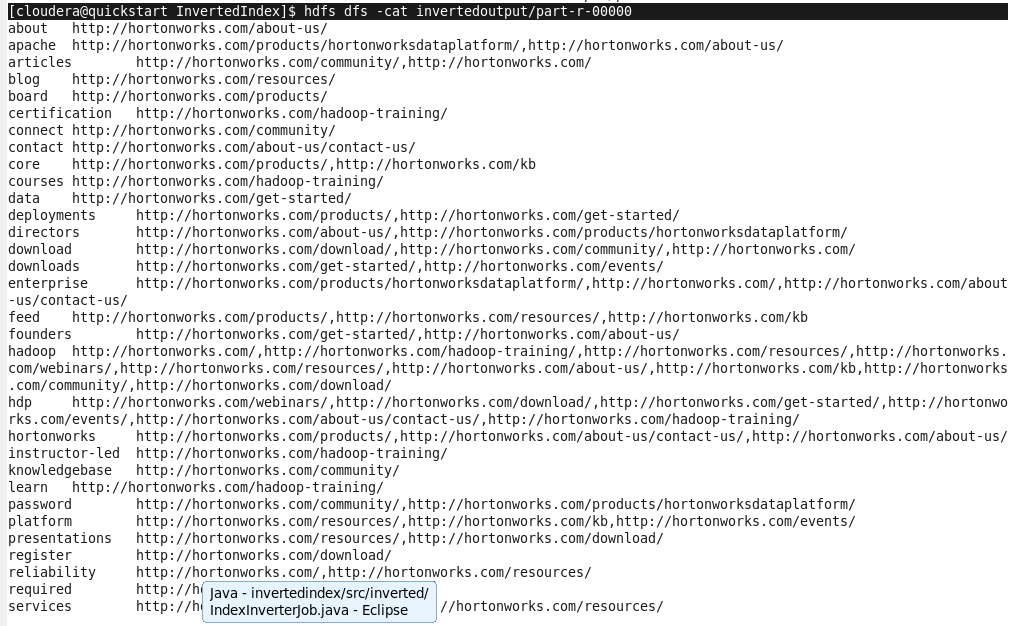




.







Here, the o/p is Link is getting Inverted as value and keys are positions.

Eg-> register (keys) -> <https://.....download/> (value)

Input File:

http://hortonworks.com/,hadoop,webinars,articles,download,enterprise,team,reliabilityhttp://hortonworks.com/products/,hortonworks,services,core,feed,deployments,board,requiredhttp://hortonworks.com/products/hortonworksdataplatform/,apache,password,directors,enterprisehttp://hortonworks.com/get-started/,data,downloads,founders,hdp,deploymentshttp://hortonworks.com/download/,register,hadoop,hdp,download,presentations,videoshttp://hortonworks.com/community/,connect,password,download,articles,knowledgebase,hadoophttp://hortonworks.com/kb,platform,feed,core,hadoophttp://hortonworks.com/about-us/,about,hortonworks,apache,hadoop,founders,directorshttp://hortonworks.com/about-us/contact-us/,contact,support,hortonworks,hdp,enterprisehttp://hortonworks.com/resources/,hadoop,services,training,videoshttp://hortonworks.com/events/,hdp,downloads,platform,training,videoshttp://hortonworks.com/webinars/,webinar,hadoop,videos,hdphttp://hortonworks.com/resources/,feed,blog,platform,hadoop,presentations,reliabilityhttp://hortonworks.com/hadoop-training/,hadoop,instructor-led,certification,training,courses,learn,hdp