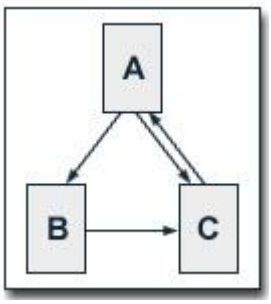
Technical documentation

This example demonstrates a program written in Scala that implements Web Page Ranking algorithm, which will read web page relations from text file in HDFS and calculate their page ranks.

1 – step:
The following picture a) shows 3 pseudo web pages marked as **A**, **B** and **C** and their relations among each other denoted with arrows. Arrows goes from web page **A** to **B**, this relation means that web page **A** contains link reference to web page **B** and so on. I saved web page relations in the following picture as **WebPageRelations.txt** file in the format as shown in the **b).**

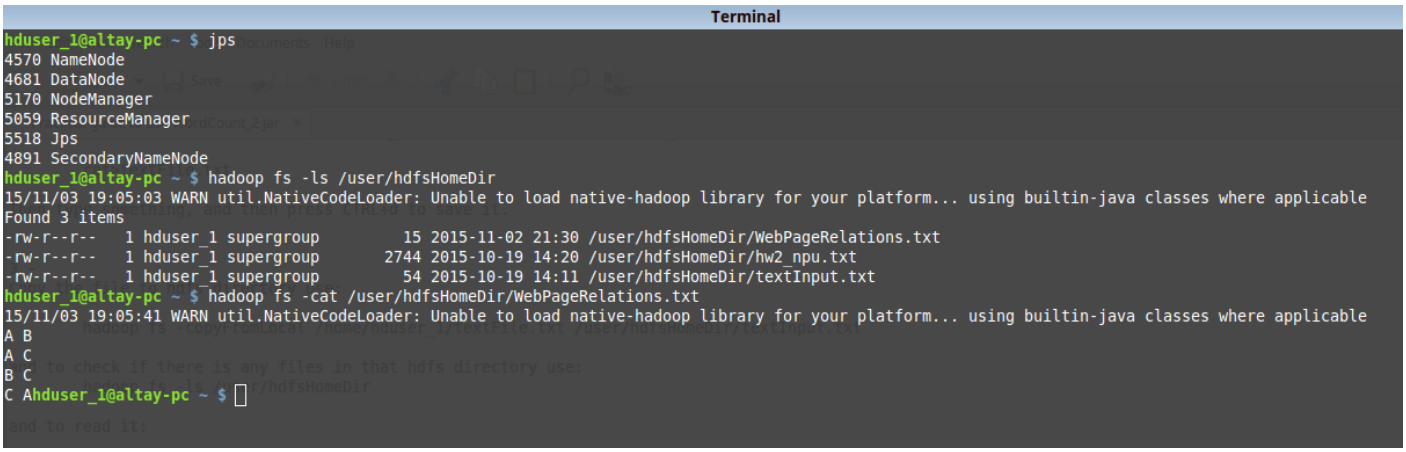
a)



b)

A B
A C
B C
C A

In the following picture you can see that my Hadoop is up, my **WebPageRelations.txt** file is in HDFS and the contents of **WebPageRelations.txt** file.



2 – step:

I just run my Scala program in Scala IDE and the result of the program is printed to output console which you can see in the following picture. Each pseudo web page has it's own calculated page rank.