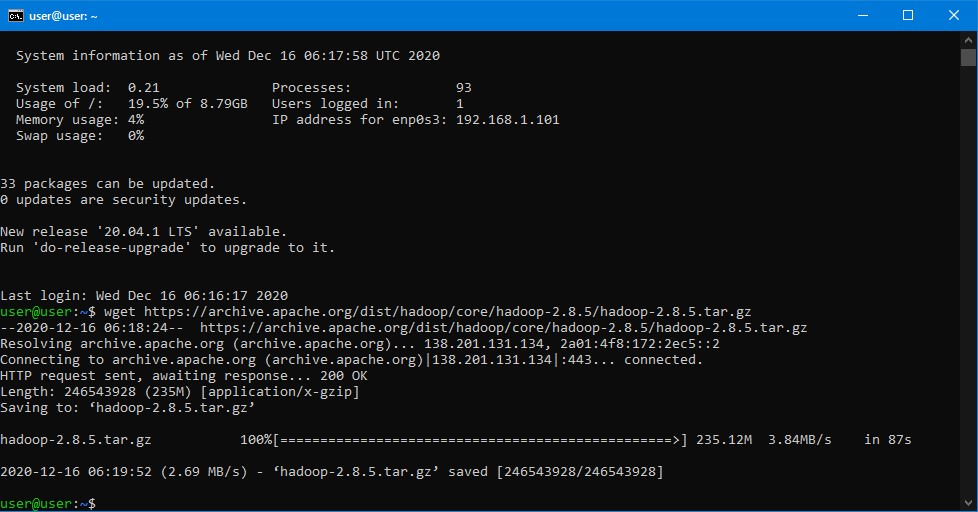
Скачаем Hadoop:



Установим:

tar -xvf hadoop-2.8.5.tar.gz

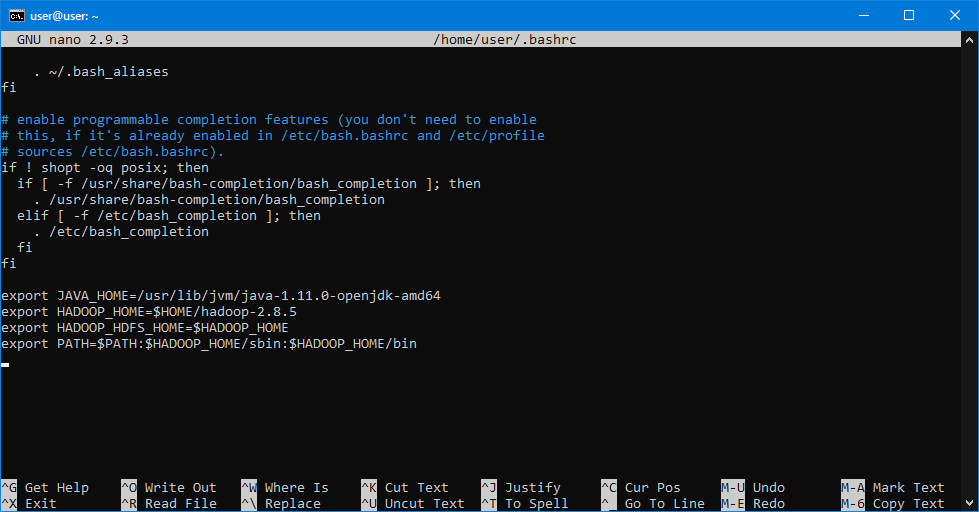
Установим пути в файл ~/.bashrc:

export JAVA\_HOME=/usr/lib/jvm/java-1.11.0-openjdk-amd64

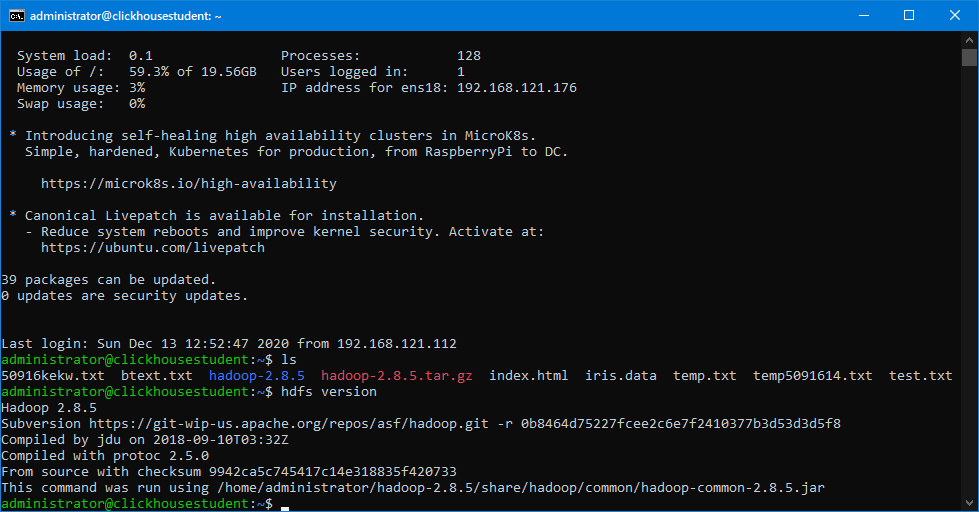
export HADOOP\_HOME=$HOME/hadoop-2.8.5

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

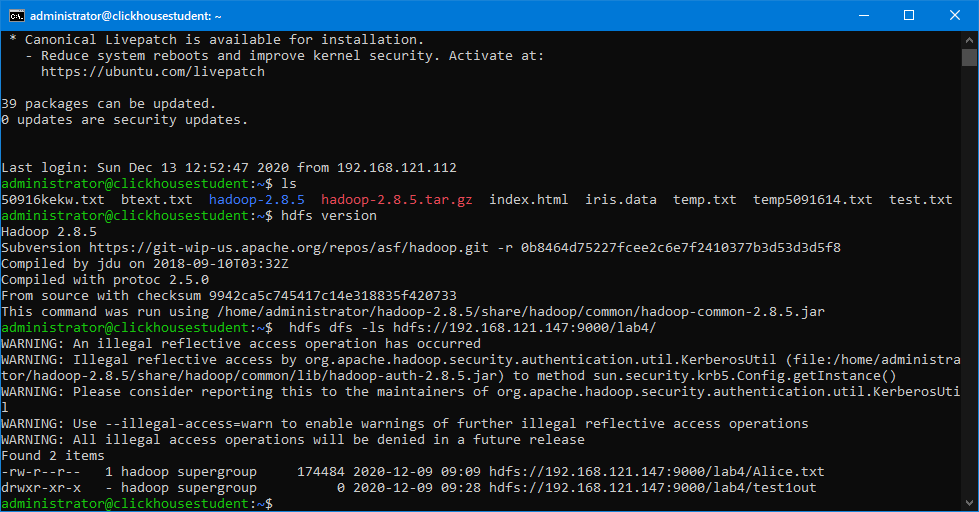
export PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin



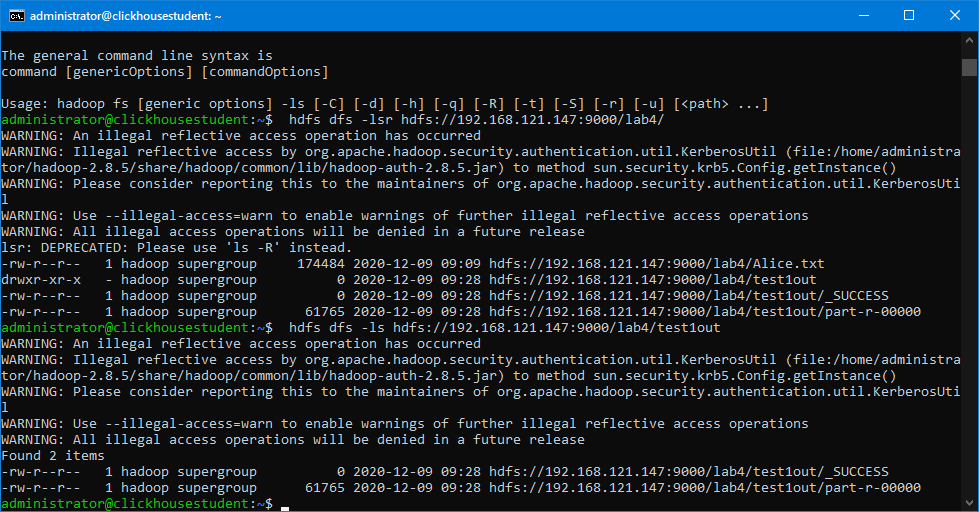
Посмотрим версию:



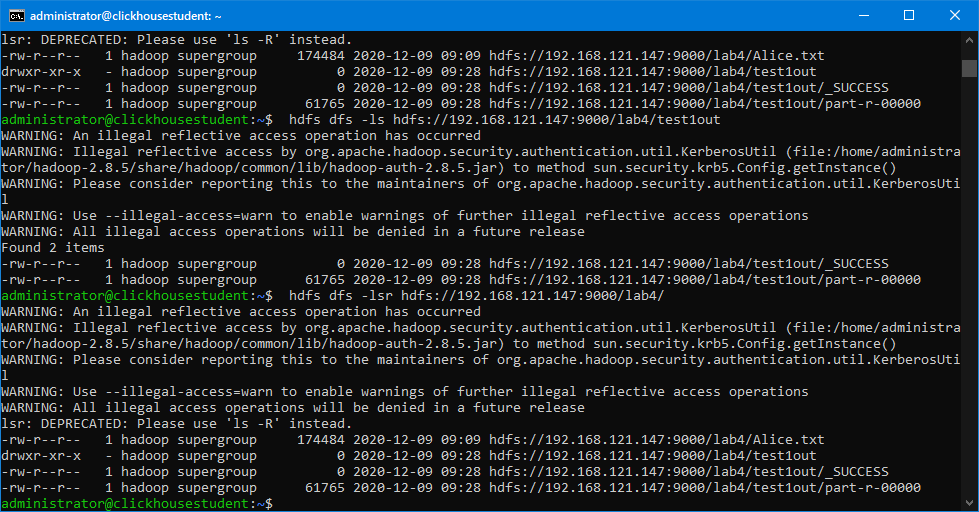
Получим список файлов на удаленном сервере:



Найдем файл part-r-00000:

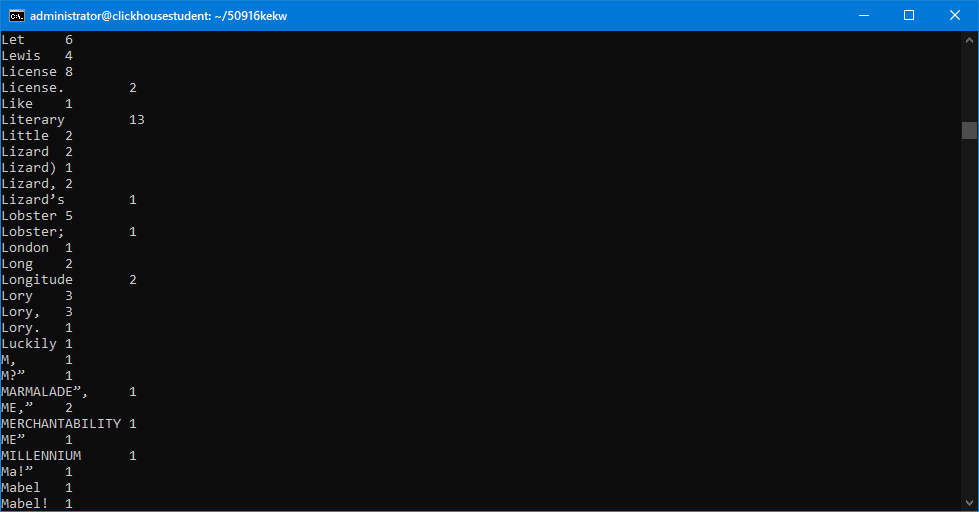


Выполним команду lsr:



cat:

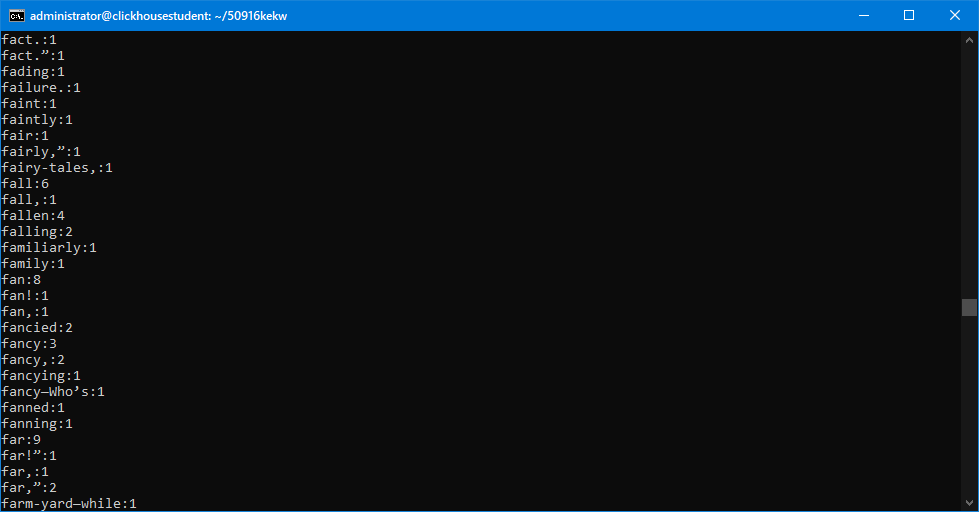
hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000



Самое часто встречающееся слово нельзя быстро найти, так как различных слов в тексте очень много

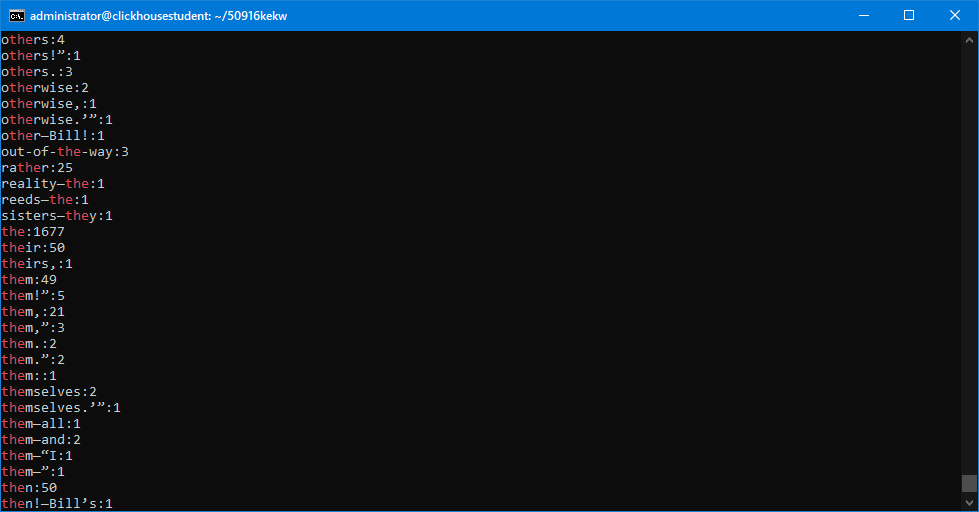
awk:

hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}'



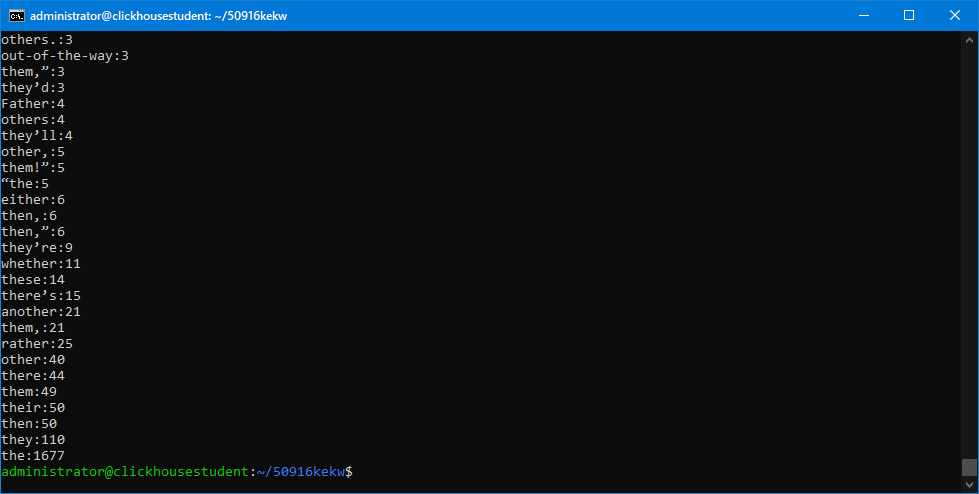
Найдем все the в тексте:

hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}' | grep the



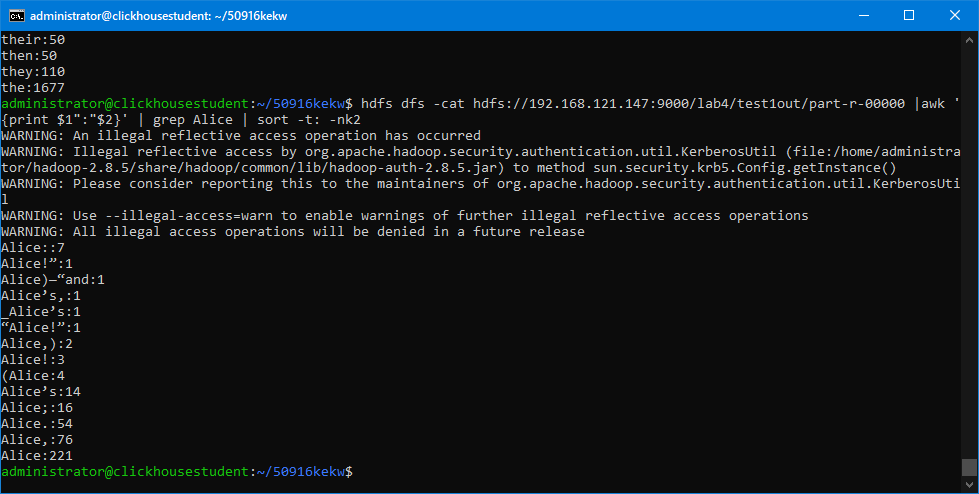
Численная сортировка:

hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}' | grep the | sort -t: -nk2



Найдем слово Alice:

hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 |awk '{print $1":"$2}' | grep Alice | sort -t: -nk2

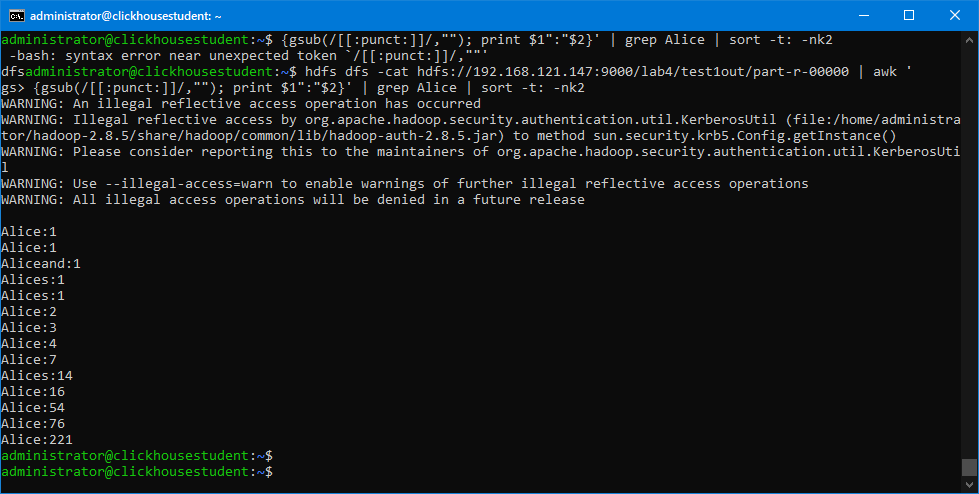


Получить корректный ответ мешают знаки препинания.

Уберем знаки препинания:

hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '

{gsub(/[[:punct:]]/,""); print $1":"$2}' | grep Alice | sort -t: -nk2



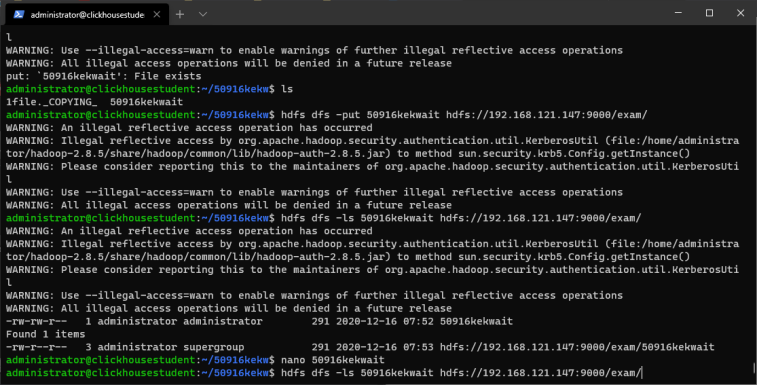
Создадим на диске файл и запишем туда состав группы:



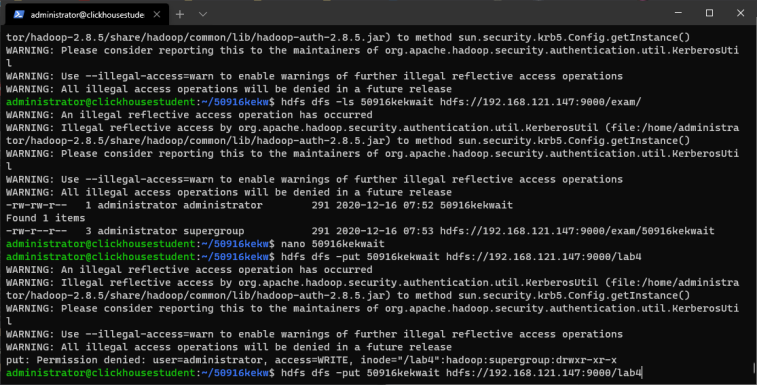
Переместим файл в папку exam:

hdfs dfs -put 50916kekwait hdfs://192.168.121.147:9000/exam/

hdfs dfs -ls 50916kekwait hdfs://192.168.121.147:9000/exam/



Переместим файл в папку lab4:



put: Permission denied: user=administrator, access=WRITE, inode="/lab4":hadoop:supergroup:drwxr-xr-x