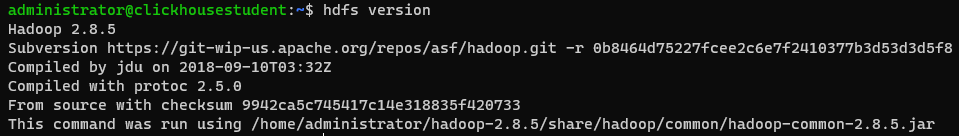
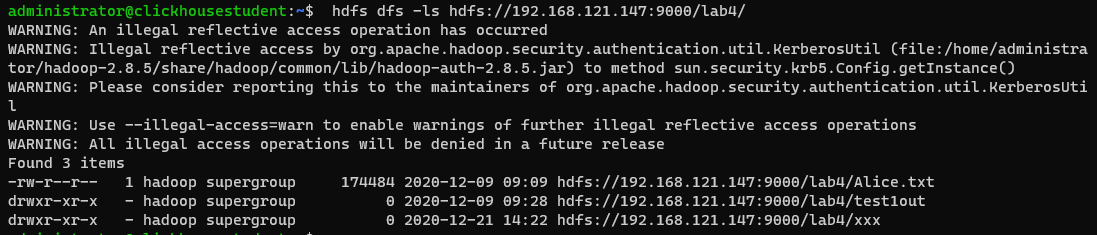
Код группы 5091614. Состав (студенты 5 курса группы ИТ-50916: Галисултанов Р.И., Замятин К.А., Коркин И.С., Шаков М.А.)

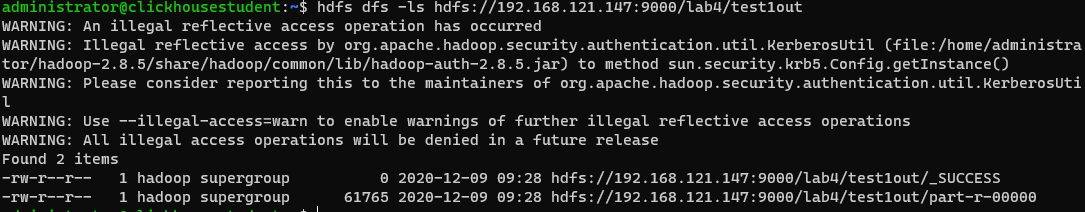
Библиотеки Hadoop уже установлены



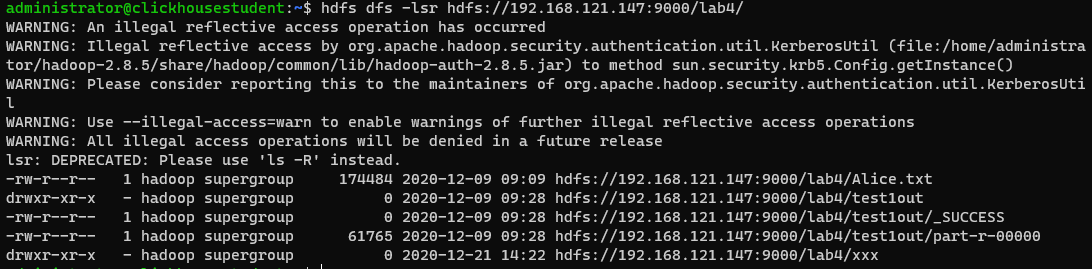
Получим список файлов в папке lab4 на удалённом сервере



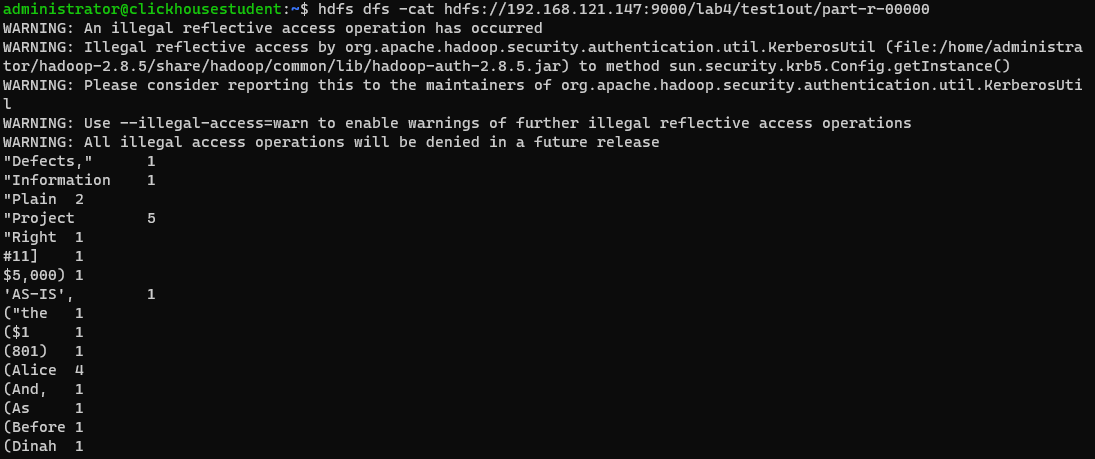
Отыщем файл part-r-00000



Повторим то же самое с командой lsr

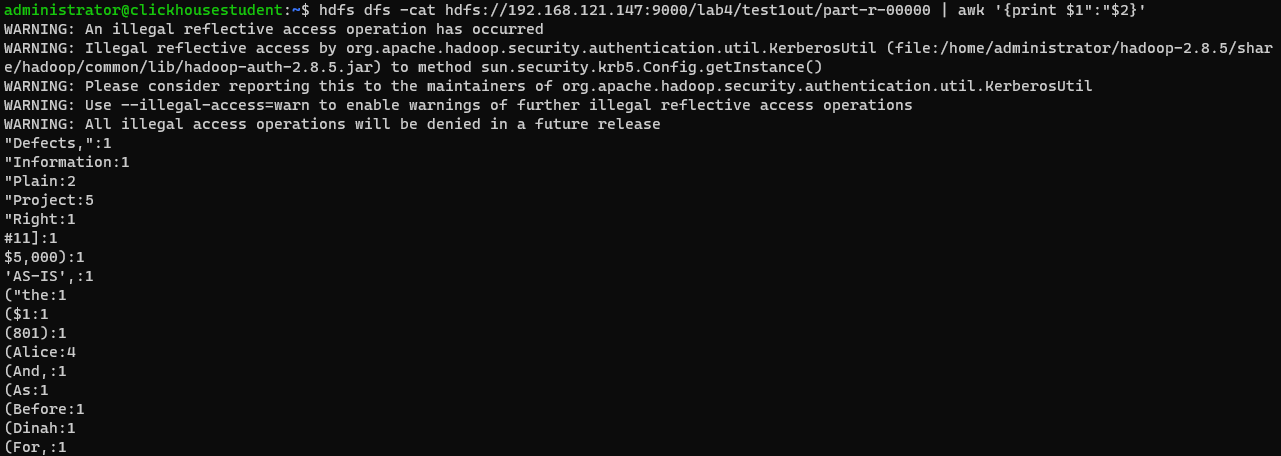


Прочитаем содержимое с помощью команды cat

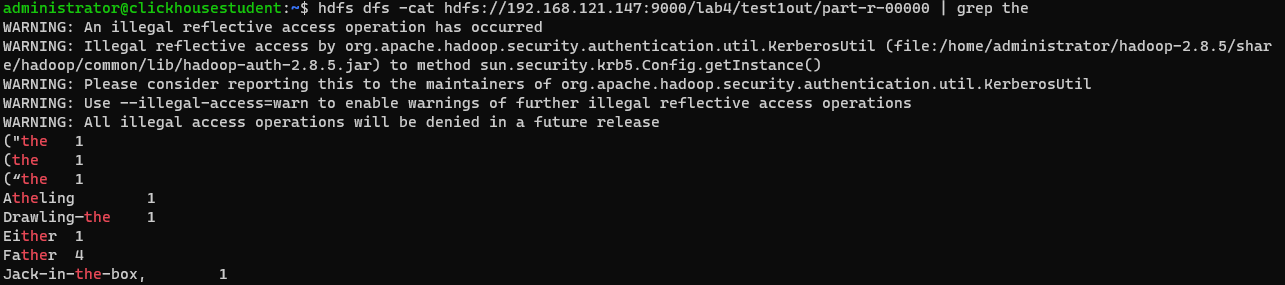


Объем текста велик, поэтому быстро найти самое часто встречающееся слово не получится

Используем интерпретатор awk

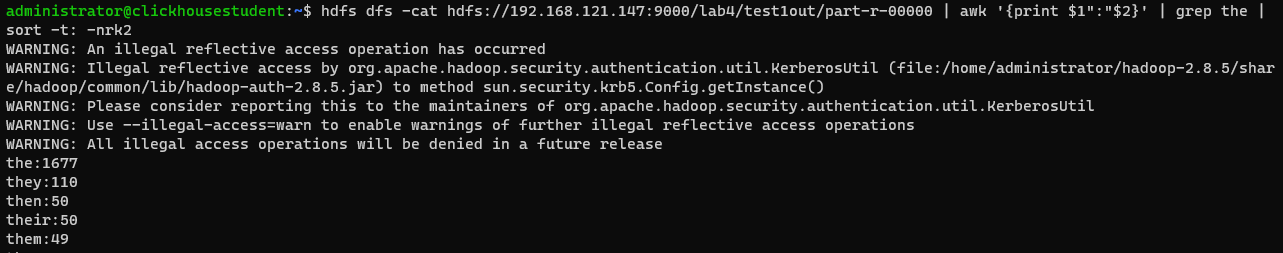


Найдем все the в тексте

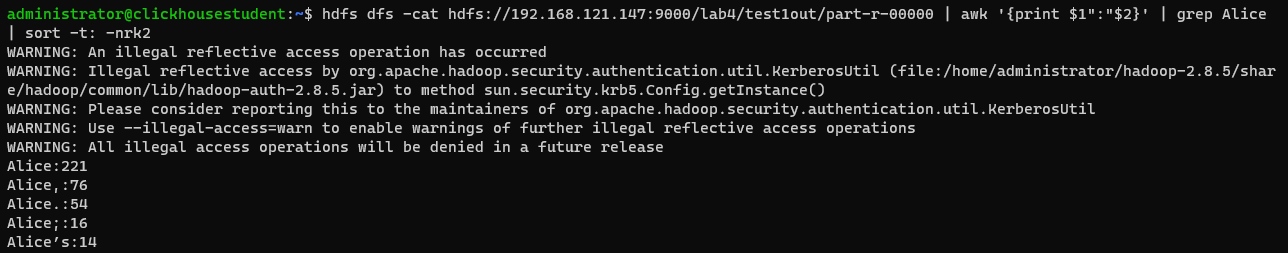


Добавим численную сортировку с разделителем : по второму полю в котором число упоминаний

hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}' | grep the |sort -t: -nrk2

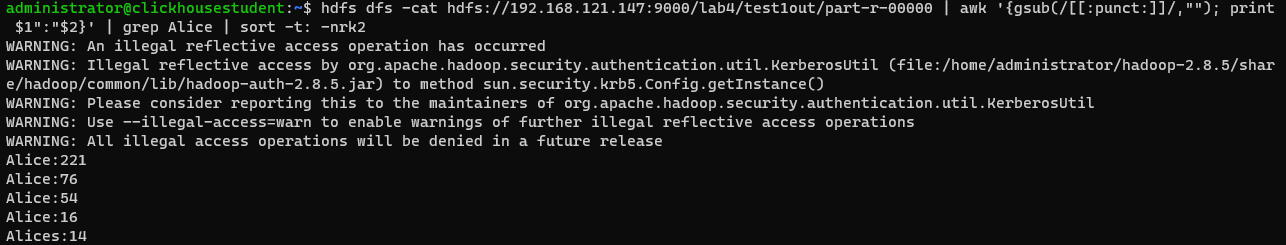


Воспользовавшись командой grep найдем сколько раз встречается имя Alice

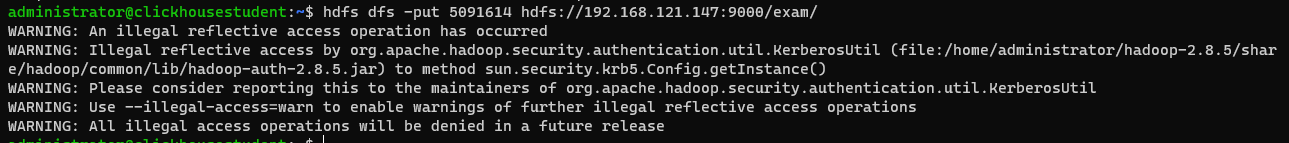


Не берутся в расчет сочетания имени со знаками препинания, удалим их

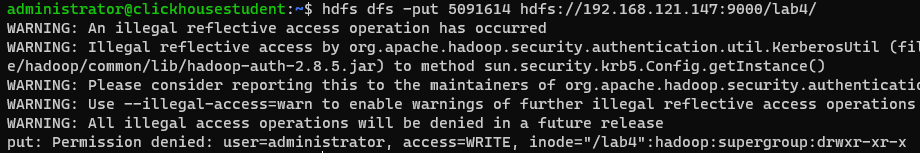
hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk ' {gsub(/[[:punct:]]/,""); print $1":"$2}' | grep Alice | sort -t: -nкk2



Поместим файл команды в каталог exam: hdfs dfs -put 5091614 hdfs://192.168.121.147:9000/exam/



Попытаемся скопировать файл в каталог lab4

У нас недостаточно прав для записи в этот каталог, мы не являемся его владельцем