МИНИСТЕРСТВО НАУКИ и ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Курганский государственный университет»

(ФГБОУ ВО «КГУ»)

Кафедра «Безопасность информационных и автоматизированных систем»

ОТЧЕТ

По лабораторной работе №5

«Основы HDFS»

Работу выполнили

студенты группы ПТ-40917

Соколова Елена

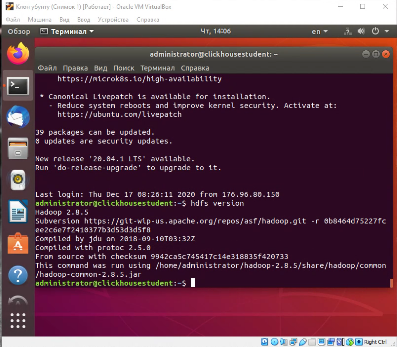
Воронов Владислав

Слободюк Сергей

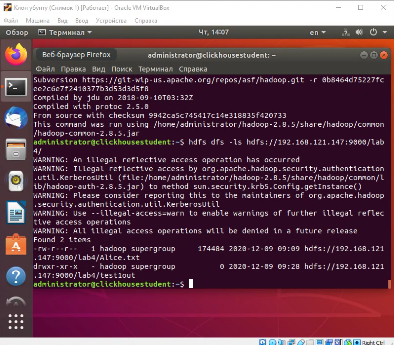
Истомин Павел

Курган 2020

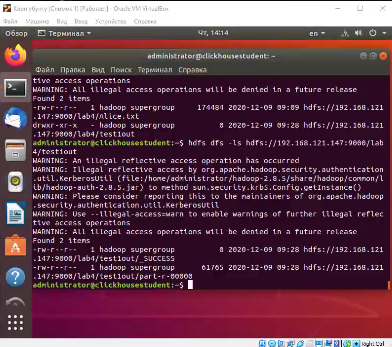
Выполнение команды: hdfs version

\

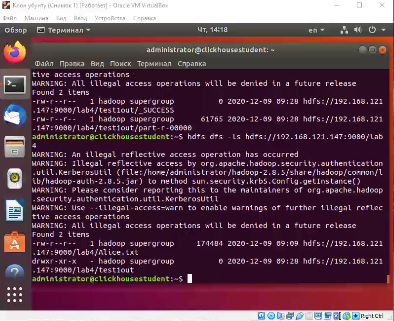
Получение списка файлов в папке lab4 на удалённом сервере: hdfs dfs -ls hdfs://192.168.121.147:9000/lab4/



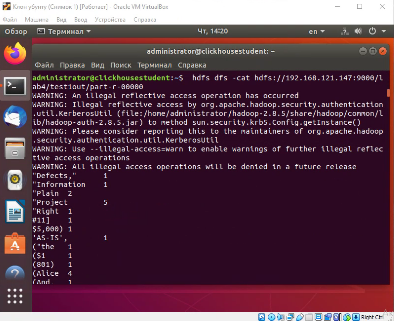
С помощью многократного вызова команды ls находим файл part-r-00000 внутри папки lab4.



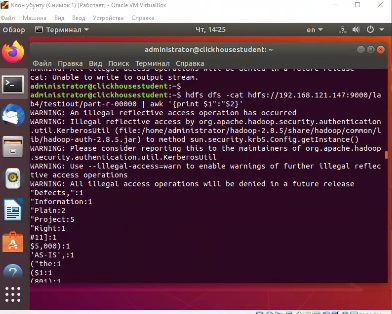
С помощью команды lsr находим файл part-r-00000 внутри папки lab4.



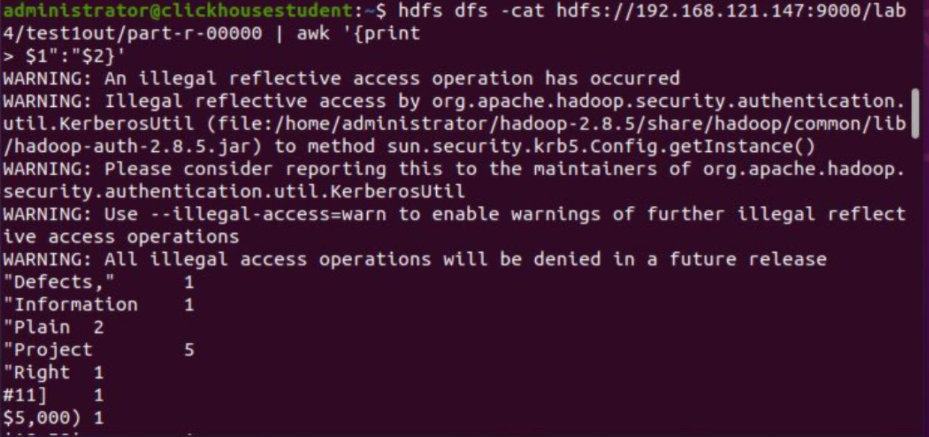
С помощью команды cat читаем содержимое файла part-r-00000 находящегося внутри папки lab4.



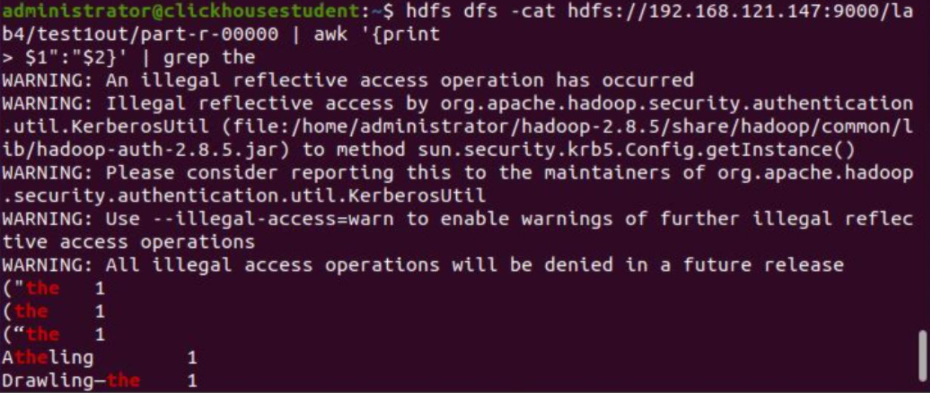
Нельзя быстро найти самое часто встречающееся слово, так как различных слов в тексте очень много.



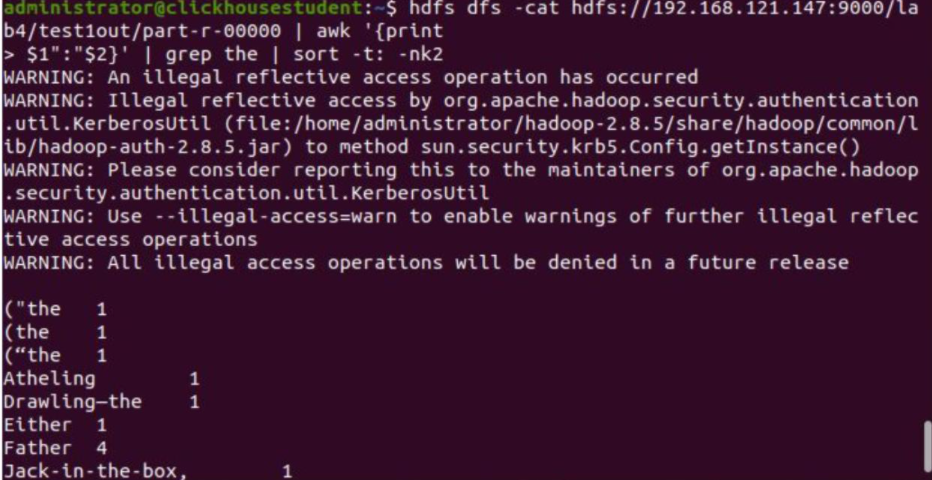
С помощью интерпретатора awk разбиваем содержимое файла part-r-00000 на колонки с разделителем



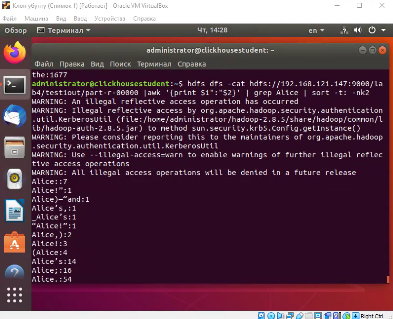
Найдём все the в тексте: hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/ | grep the



Добавляем численную сортировку с разделителем : по второму полю в котором число упоминаний | sort -t: -nk2

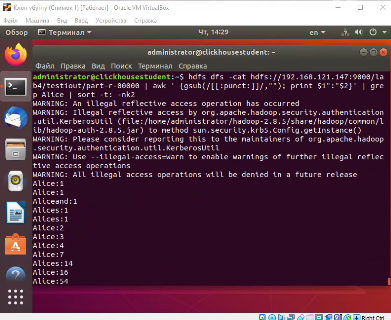


Воспользовавшись командой grep найходим сколько раз встречается имя Alice.

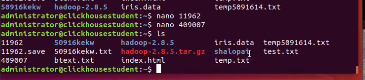


Получить корректный ответ мешают знаки препинания.

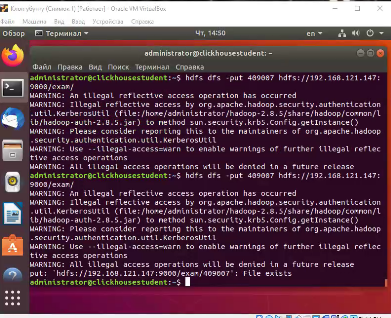
Модифицируем код awk, для удаления знаков препинания, повторяем поиск



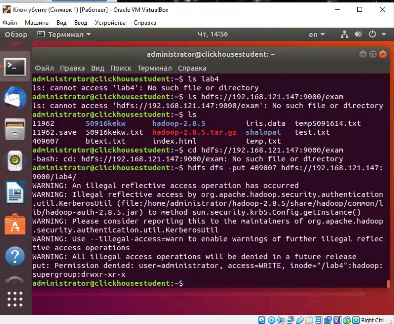
Создаем файл на диске 409007



  С помощью команды put записываем этот файл в папку /exam/



Пробуем записать этот же файл в папку /lab4/



Не удается этого сделать в связи с отсутствием необходимых прав.