

Отчёт по Homework 1. Hadoop:  
по предмету «Наука о данных и анализ больших данных (Data Science and Big Data Analytics)»

от студента группы М19Ш05:

Выполнил:

Первухин А.М., М19-Ш05

Проверили:  
Ровнягин М.М.

Москва 2020



Рисунок 1 – Скриншот запуска нескольких reducer и с сжатым файлом

Работа выполнена через hadoop стриминг. Генератор файла на питоне. Map и Reduce аналогично выполнены с использованием языка Питон.

Команда для запуска представлена в README.md. Так же проекте представлены сгенерированный файл и этот же сжатый файл.

Тест для mapper:

echo "foo foo quux labs foo bar quux" | /mapper.py

foo 1

foo 1

quux 1

labs 1

foo 1

bar 1

quux 1

Вывод