|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования Российской Федерации**  Калужский филиал  федерального государственного бюджетного  образовательного учреждения высшего образования  ***«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»***  ***(КФ МГТУ им. Н.Э. Баумана)*** |

**ФАКУЛЬТЕТ** ***ИУК «Информатика и управление»***

**КАФЕДРА** \_\_***ИУК4 «Программное обеспечение ЭВМ, информационные технологии»***

**ЛАБОРАТОРНАЯ РАБОТА №1**

**«Основы Hadoop. Установка Hadoop. Основные команды файловой системы HDFS»**

**ДИСЦИПЛИНА: «Технологии обработки больших данных»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. ИУК4-72Б | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Карельский М.К. )  (Подпись) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Голубева С.Е. )  (Подпись) |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | |
|  | - Балльная оценка:  - Оценка: | |

Калуга, 2023

**Цель:** формирование практических навыков по установке и настройке кластера Hadoop и работе с файловой системой HDFS.

**Задачи:**

1. Изучить основы Hadoop.
2. Научиться устанавливать и конфигурировать Hadoop
3. Изучить основные команды для работы с файловой системой HDFS.
4. Получить навыки написания программ для работы с HDFS

**Задание:**

Для всех вариантов настроить кластер Hadoop, состоящий из двух серверов, изучить команды HDFS для работы с файлами и выполнить следующие задания:

1. Проверить существует ли директория /user/hduser в HDFS, если нет, то создать. Создать директорию /user/hduser/Hadoop
2. Создать файл в директории /user/hduser/hadoop, название файла – ваше имя и группа. После создания файла, все, что вы вводите в консоль должно сохраниться в файле. Ввести несколько строк и сохранить.
3. Убедиться в существовании файла через web-интерфейс.
4. Перенести файл в локальную файловую систему.
5. Создать новый текстовый файл в локальной файловой системе. Перенести файл в HDFS. Убедиться в существовании файла через web-интерфейс.
6. Просмотреть права доступа на файл. Изменить права доступа к файлу, чтобы только владелец и члены группы имели полный контроль над файлом.
7. Написать программу на каком-либо языке высокого уровня для решения задачи, указанной в варианте

**Вариант 7**

Напишите программу, которая будет принимать 2 входных аргумента – путь в локальной файловой системе и путь в HDFS. Программа должна проверить существование файлов в обоих файловых системах. Если в одной из них файл не существует, то программа должна скопировать его из второй файловой системы. Если файлы существуют в обоих файловых системах, то сохранить в обоих системах только файл, который был изменен позже.

**Листинг:**

import os, datetime

local\_path = input('local path: ')

hdfs\_path = input('HDFS path: ')

is\_local\_existing = os.path.exists(local\_path)

is\_hdfs\_existing = not bool(os.system(f'hadoop fs -test -e {hdfs\_path}'))

if is\_local\_existing and is\_hdfs\_existing:

local\_datetime = datetime.datetime.fromtimestamp(os.path.getmtime(

local\_path)) - datetime.timedelta(hours=3)

hdfs\_datetime = datetime.datetime.strptime(os.popen(

f'hdfs dfs -stat {hdfs\_path}').read()[:-1], '%Y-%m-%d %H:%M:%S')

if local\_datetime > hdfs\_datetime:

os.popen(f'hdfs dfs -rm {hdfs\_path}').read()

os.system(f'hdfs dfs -put {local\_path} {hdfs\_path}')

else:

os. remove(local\_path)

os.system(f'hdfs dfs -get {hdfs\_path} {local\_path}')

elif is\_local\_existing:

os.system(f'hdfs dfs -put {local\_path} {hdfs\_path}')

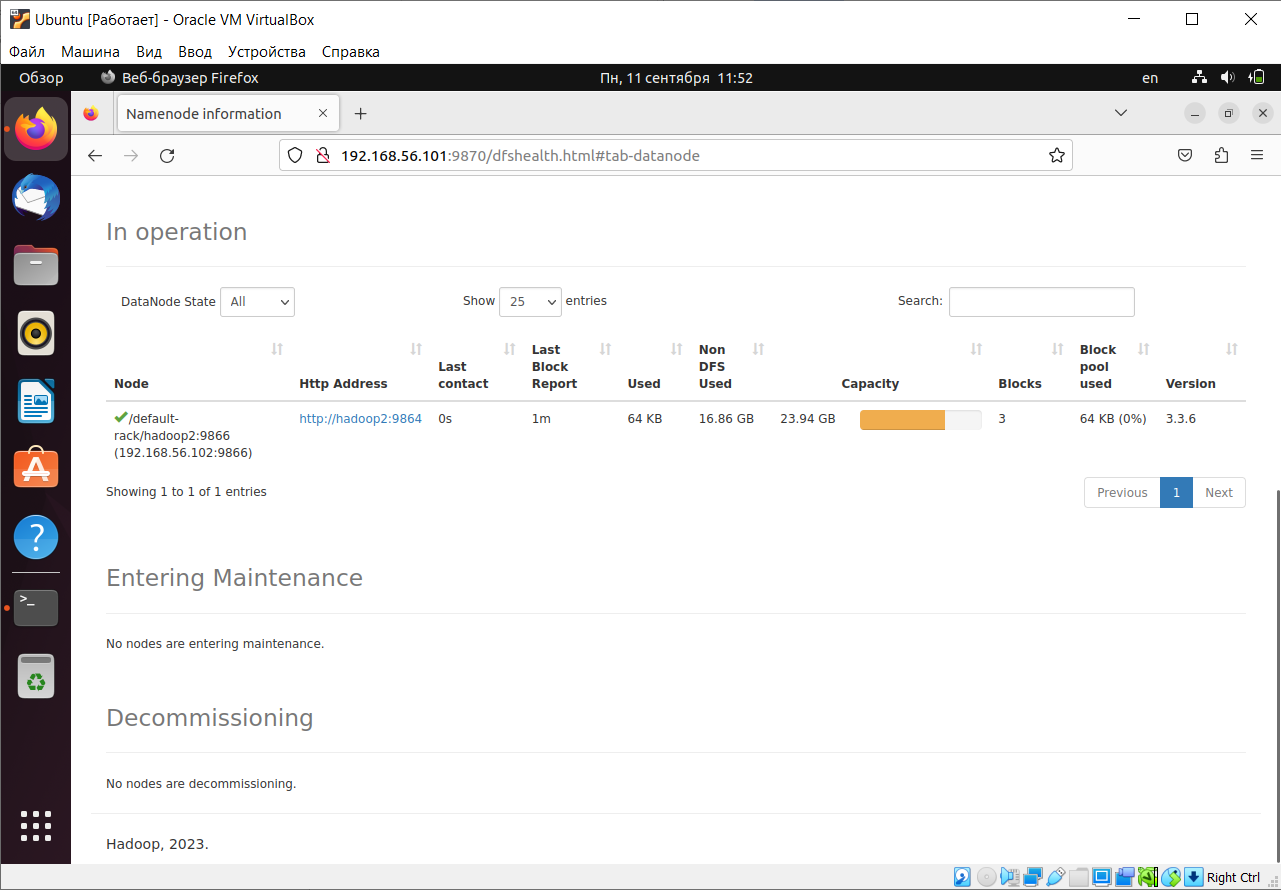
elif is\_hdfs\_existing:

os.system(f'hdfs dfs -get {hdfs\_path} {local\_path}')

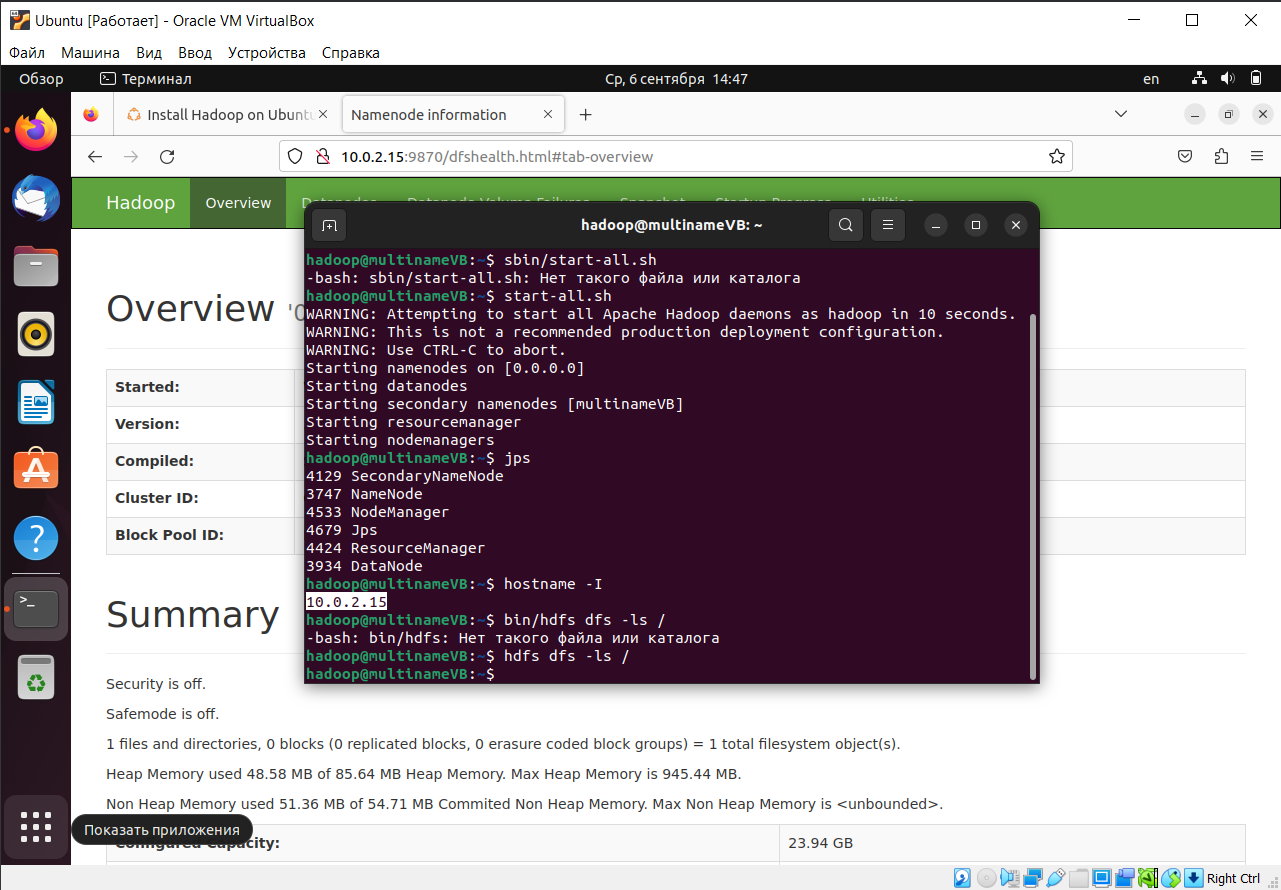
else:

print('file was not found nowhere’)

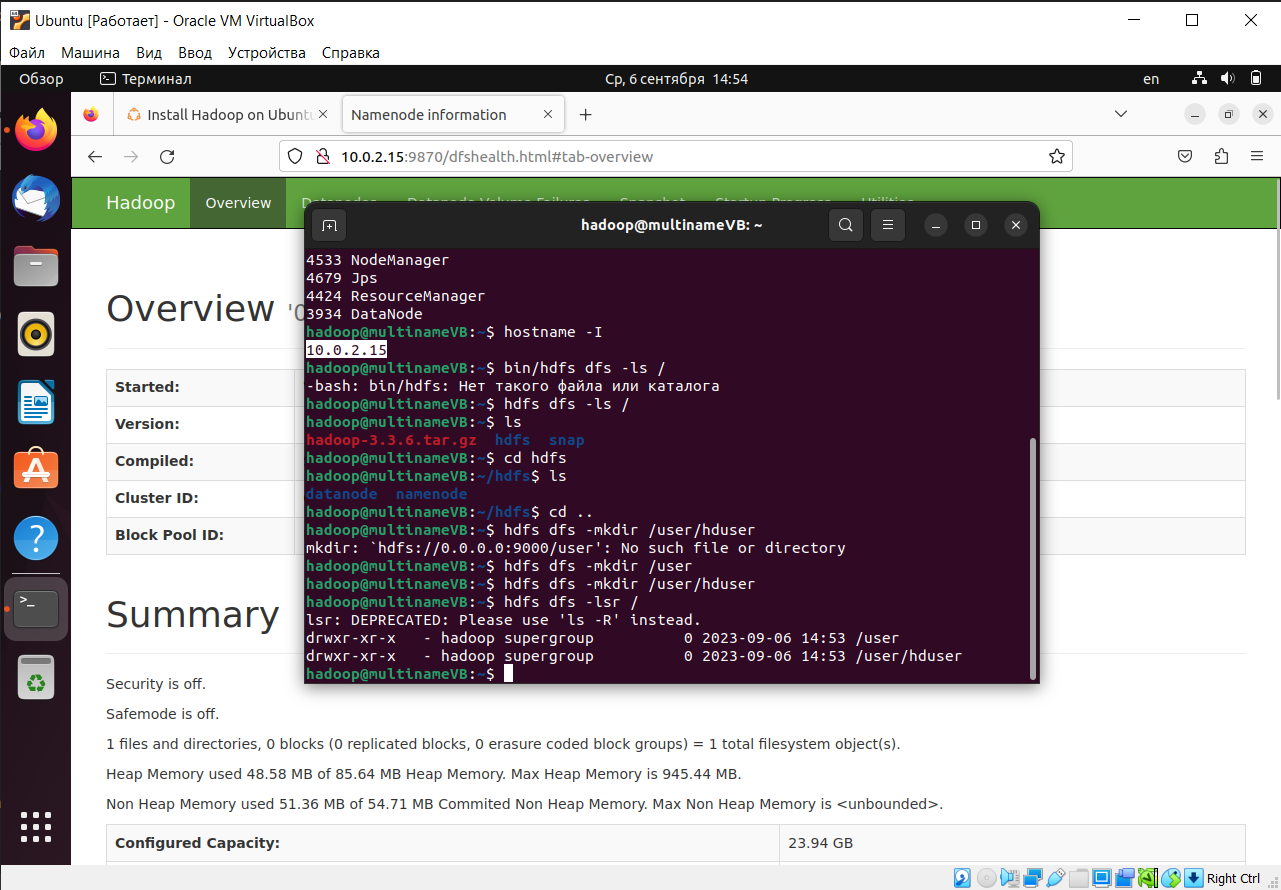
**Результат:**



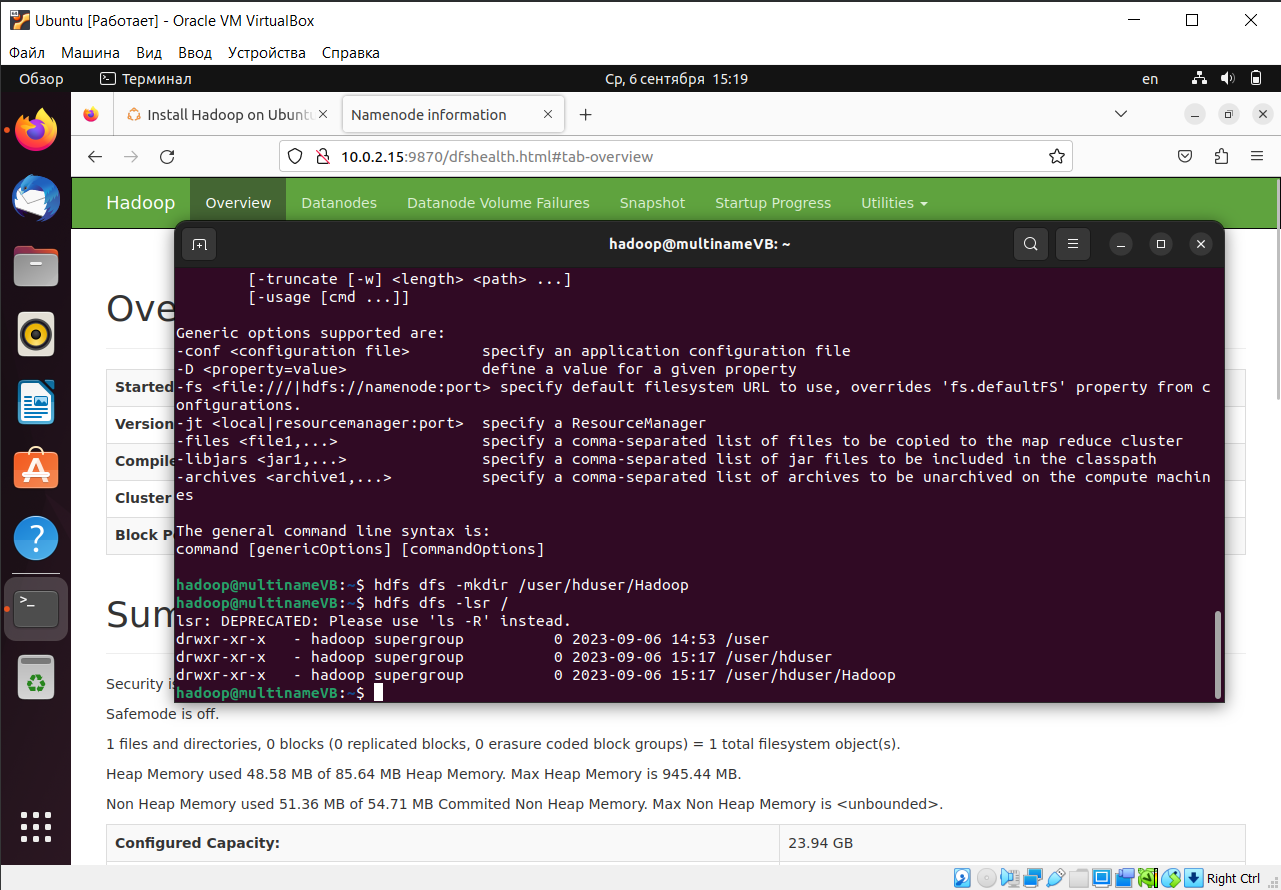
**Рис. 1.** Кластер из двух серверов



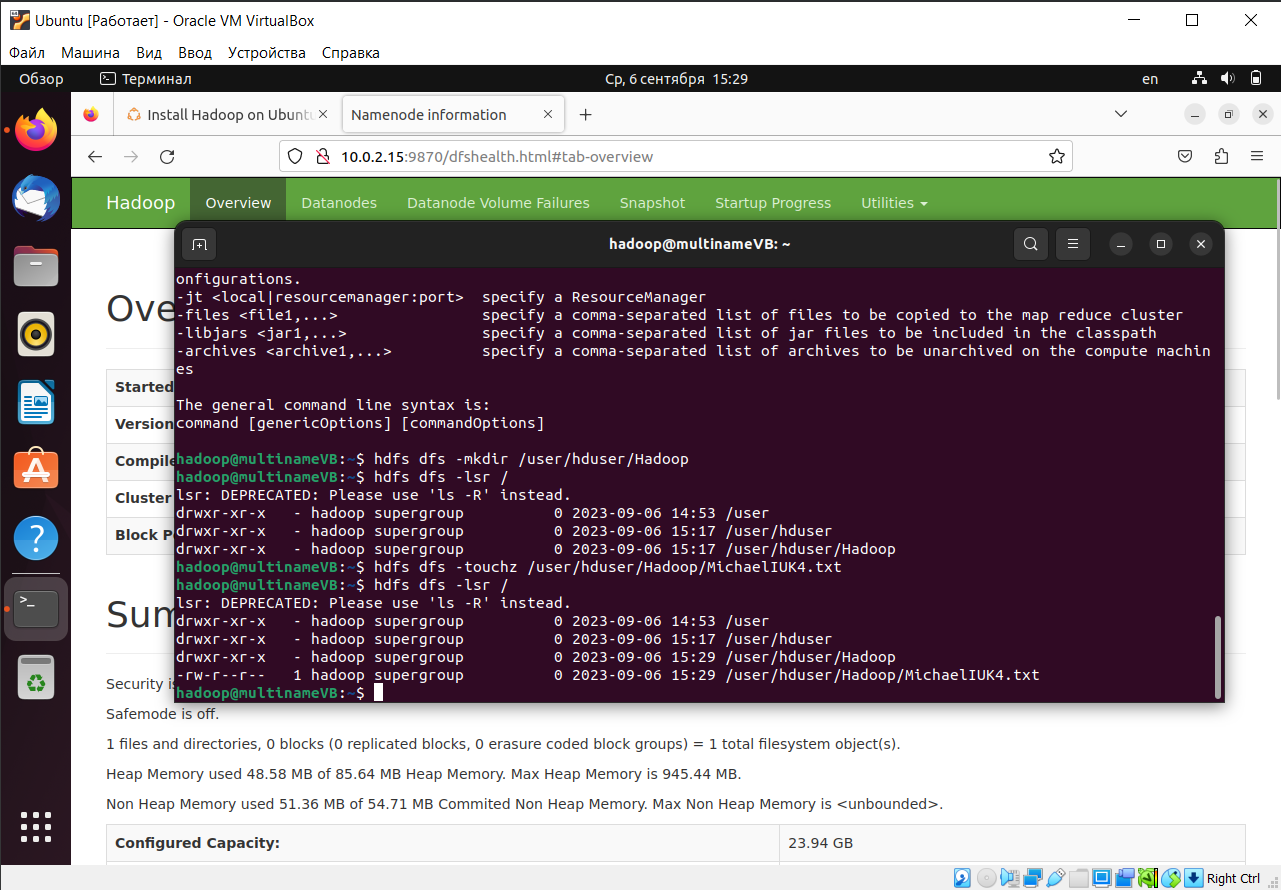
**Рис. 2.** Содержимое корневой директории



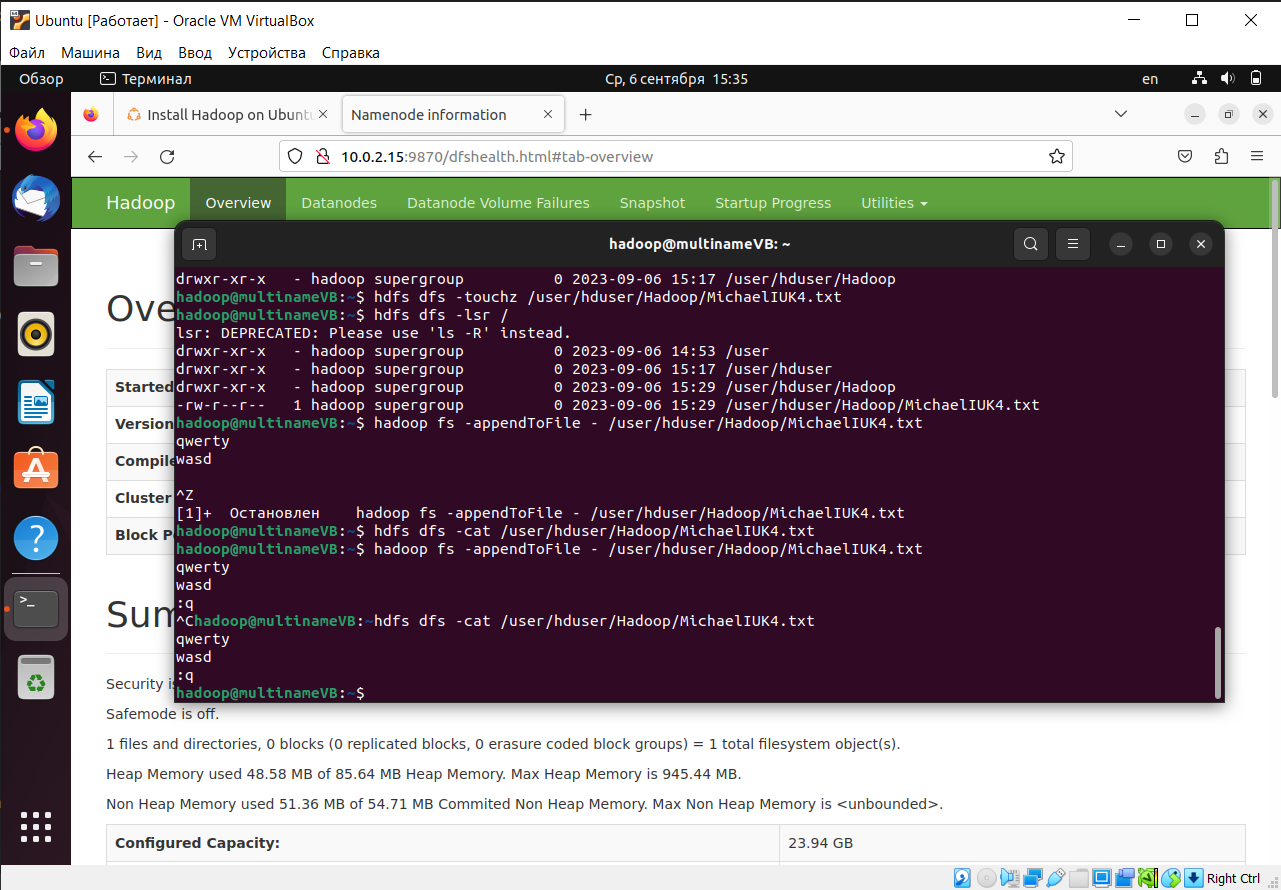
**Рис. 3.** Создание директории /user/hduser



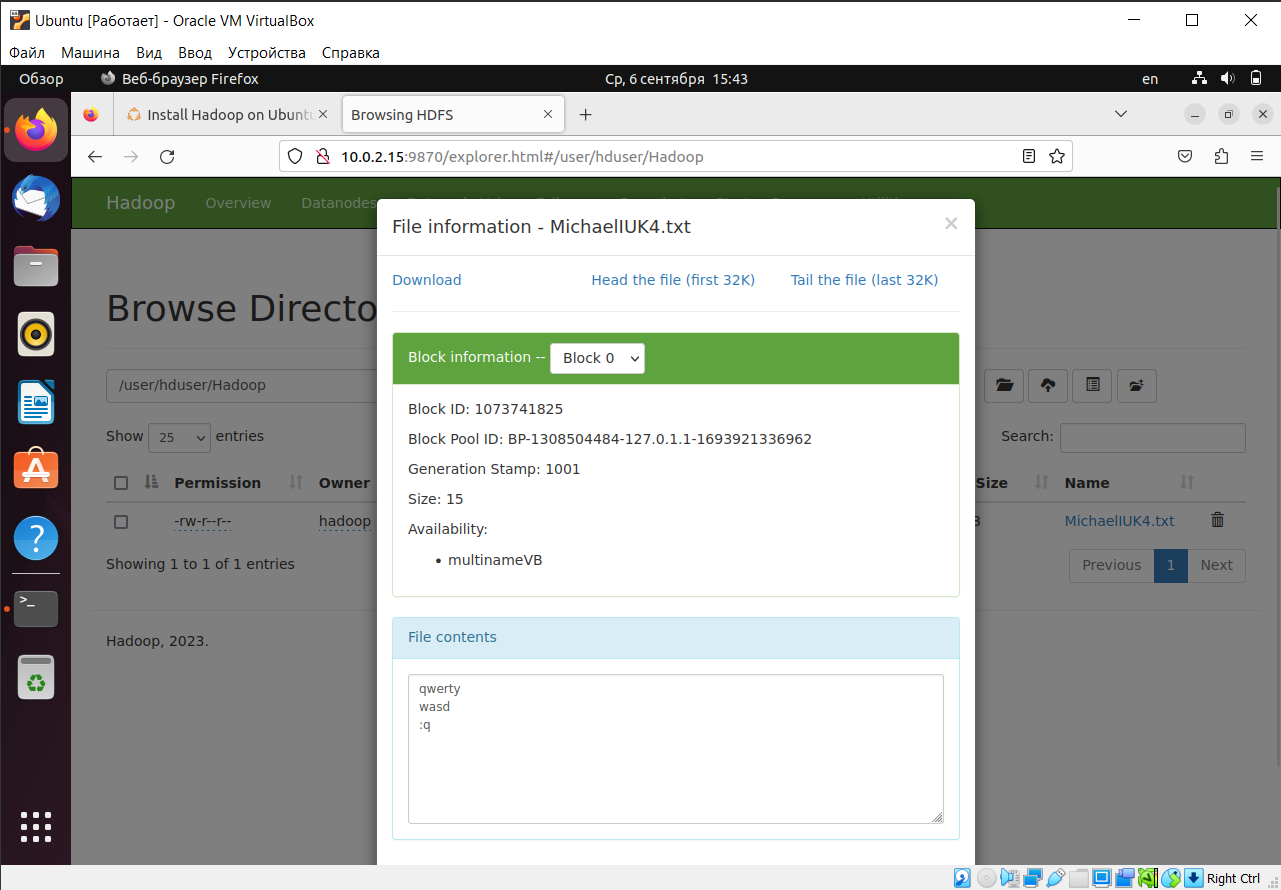
**Рис. 4.** Создание директории /user/hduser/Hadoop



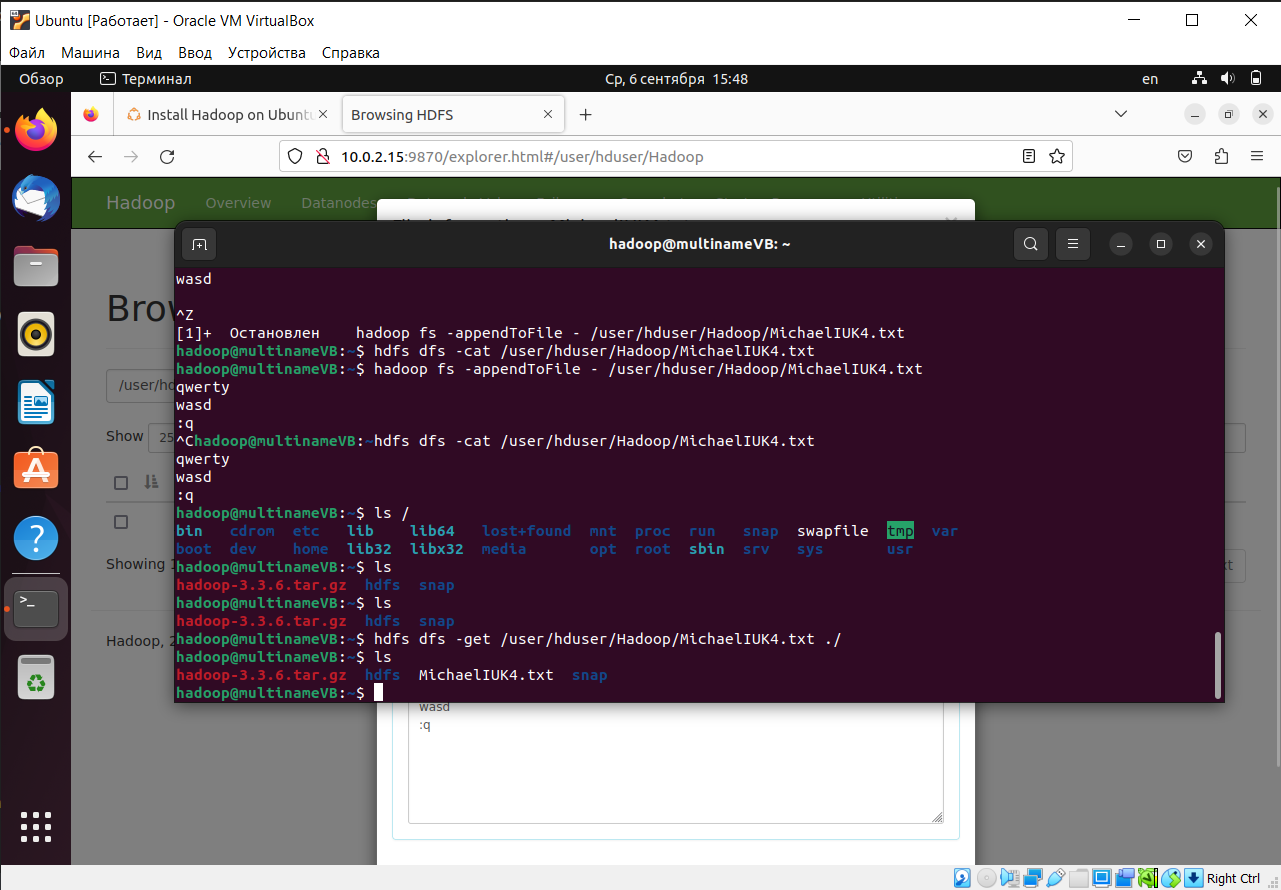
**Рис. 5.** Создание текстового файла



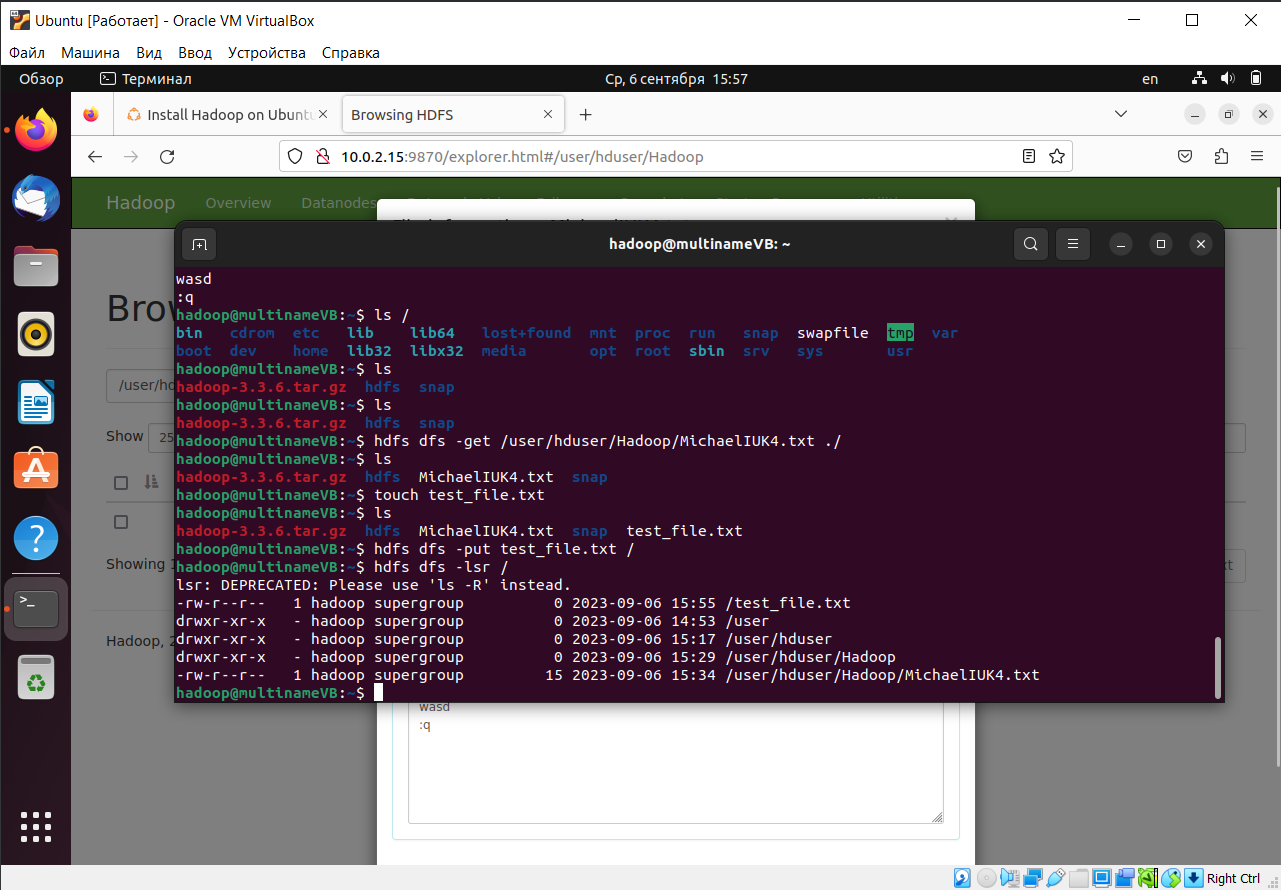
**Рис. 6.** Заполнение файла из консоли



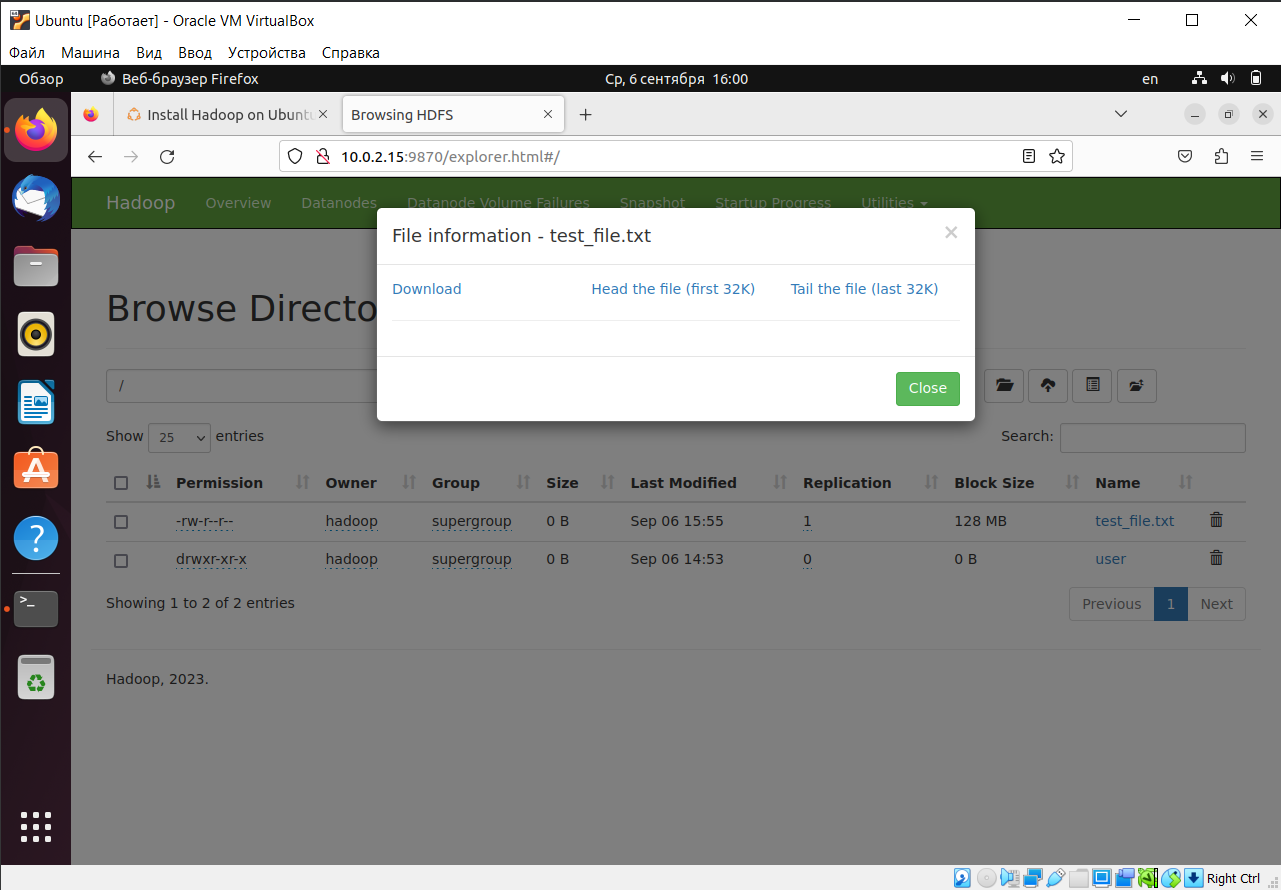
**Рис. 7.** Проверка наличия файла через web-интерфейс



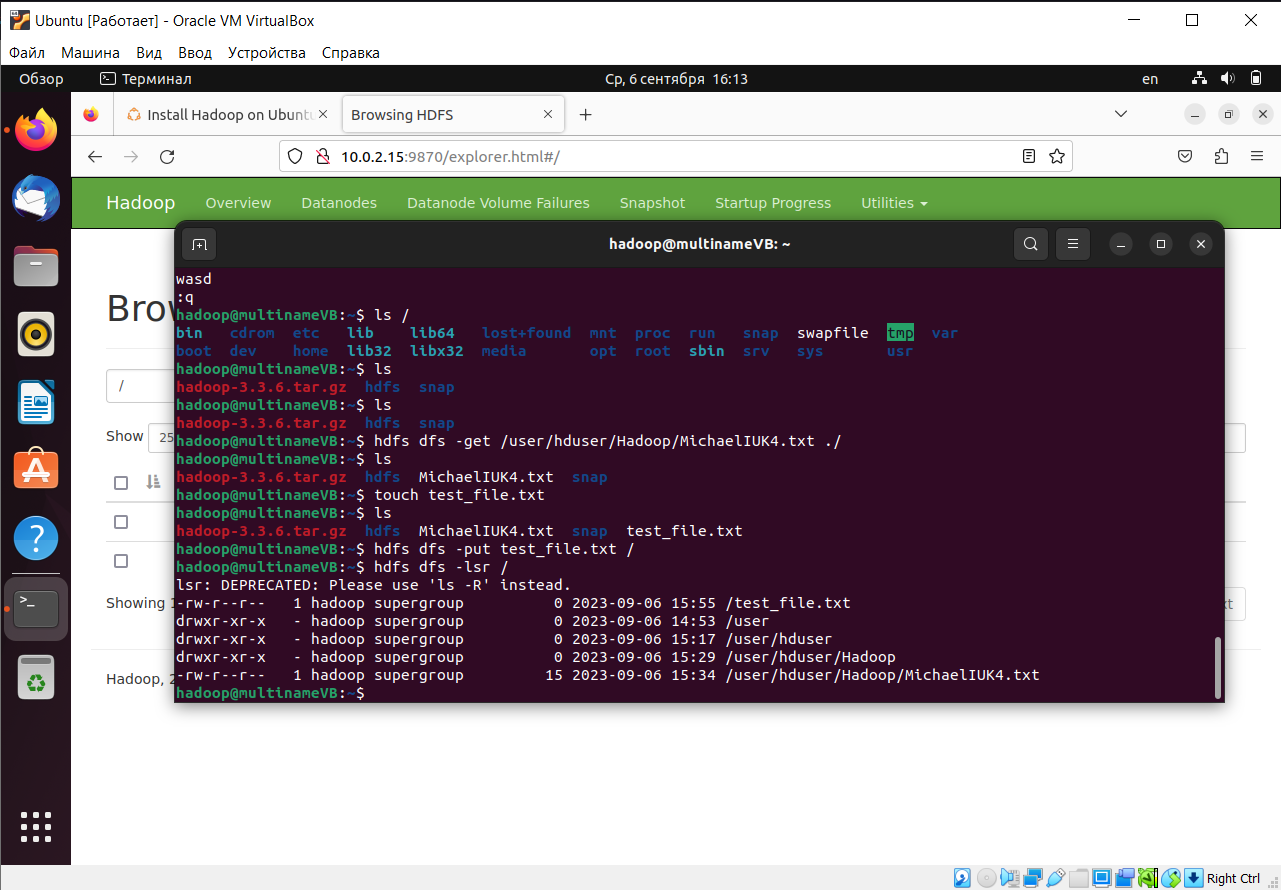
**Рис. 8.** Копирование файла на локальную систему



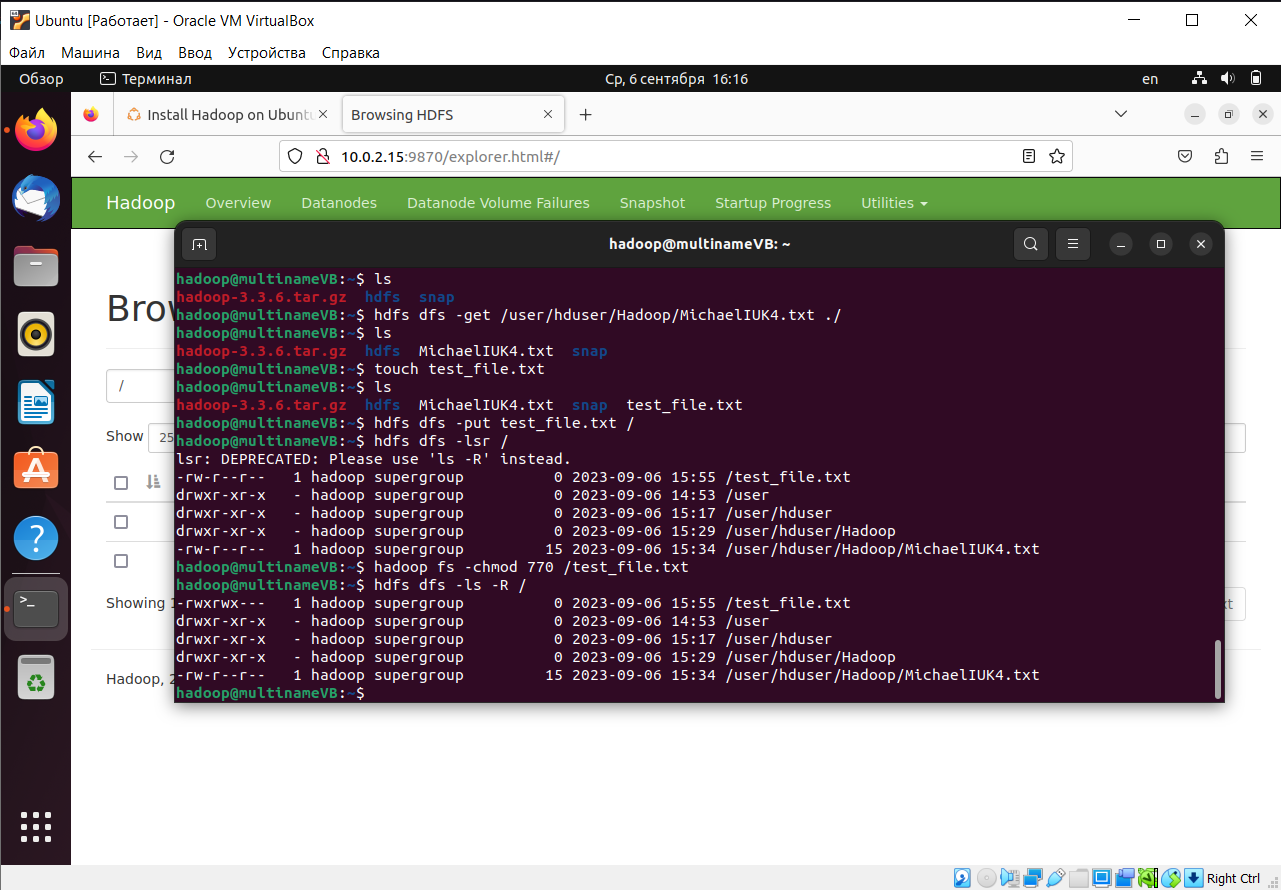
**Рис. 9.** Создание и перенос файла на HDFS



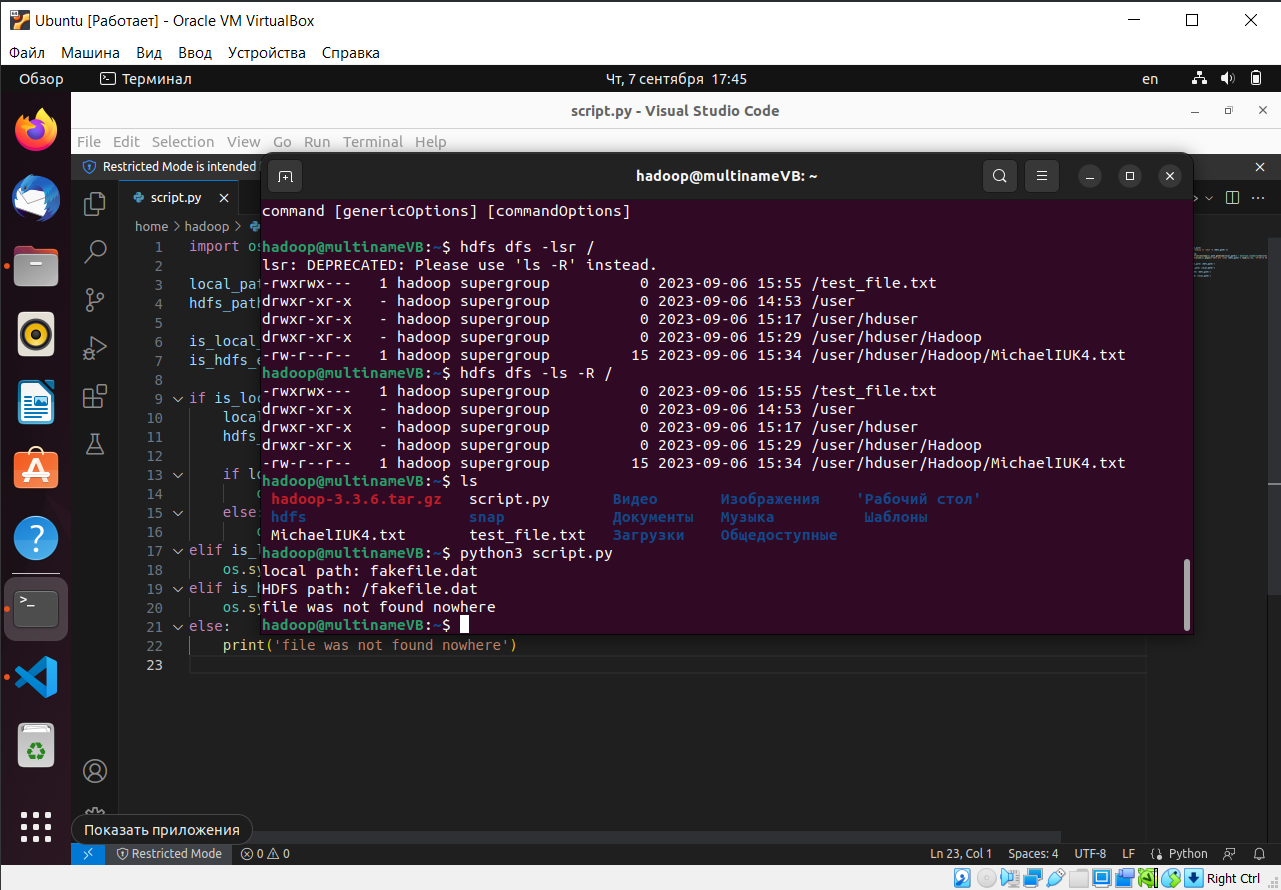
**Рис. 10.** Проверка наличия файла через web-интерфейс



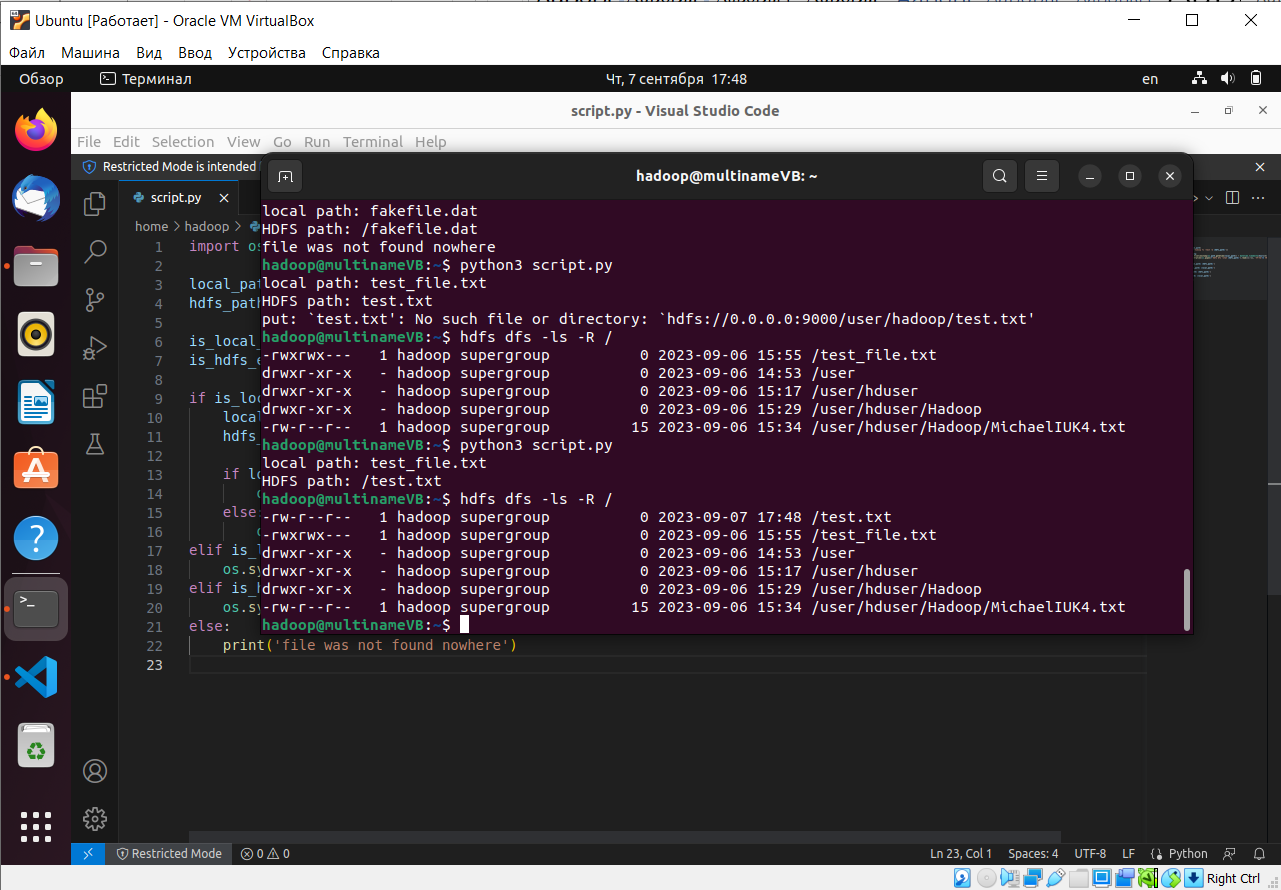
**Рис. 11.** Просмотр прав доступа



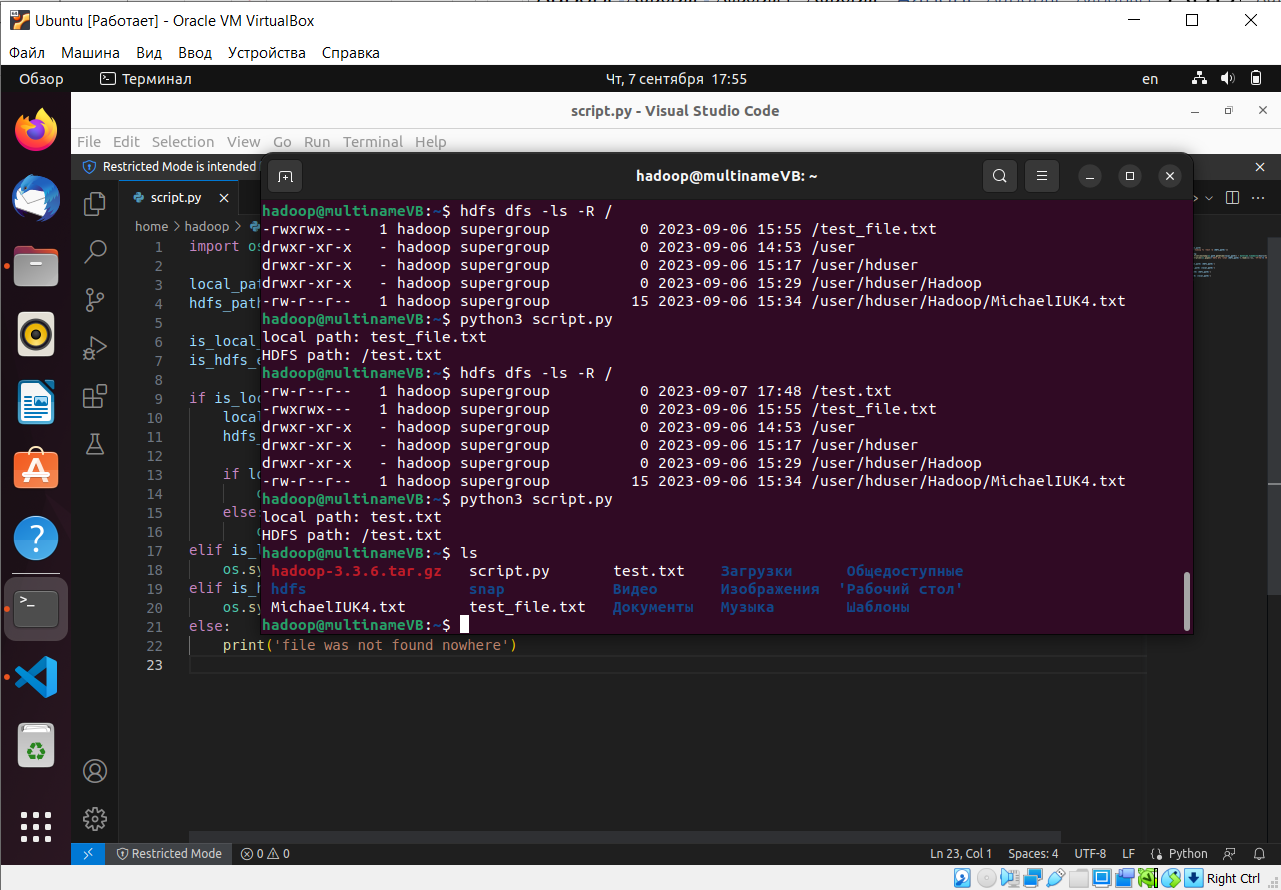
**Рис. 12.** Изменение прав доступа



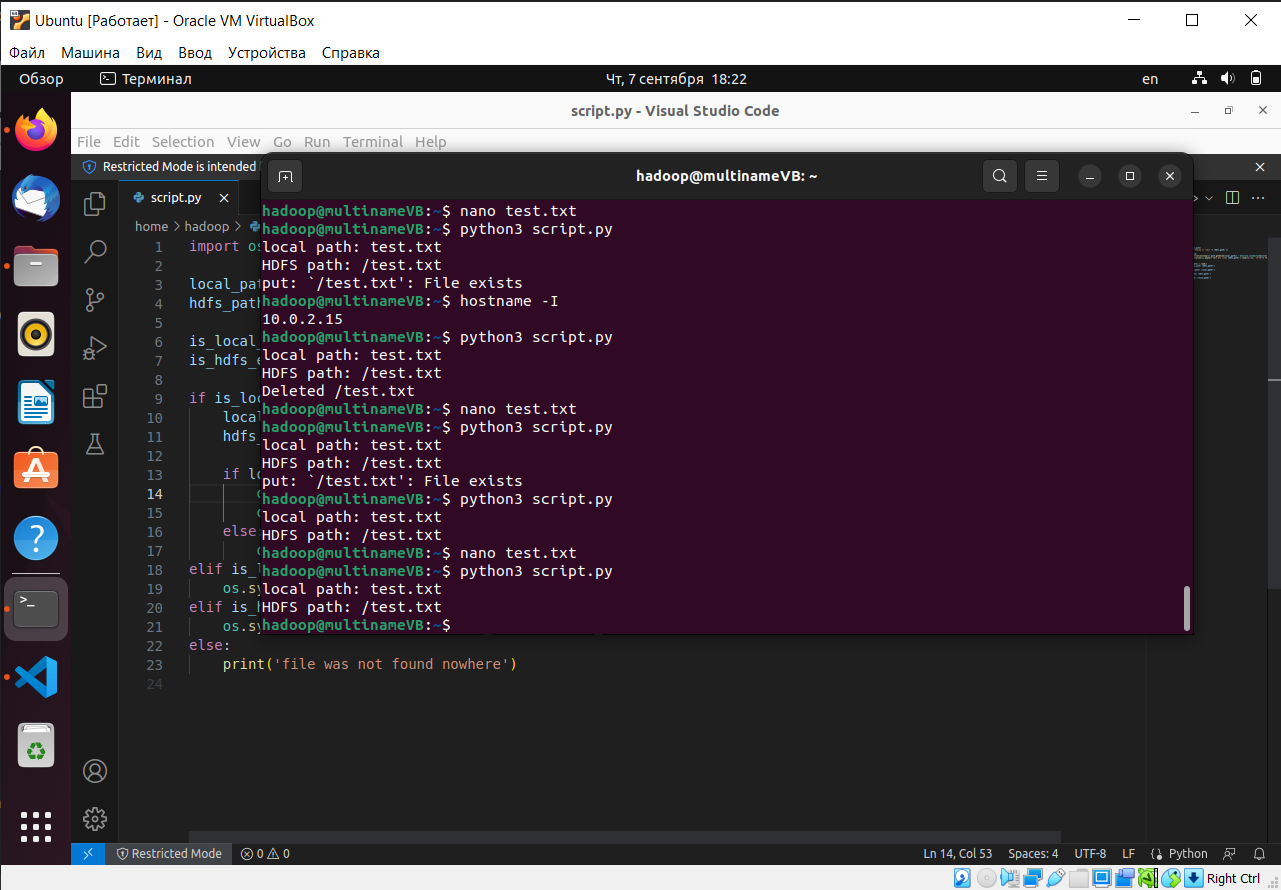
**Рис. 13.1.** Работа программы

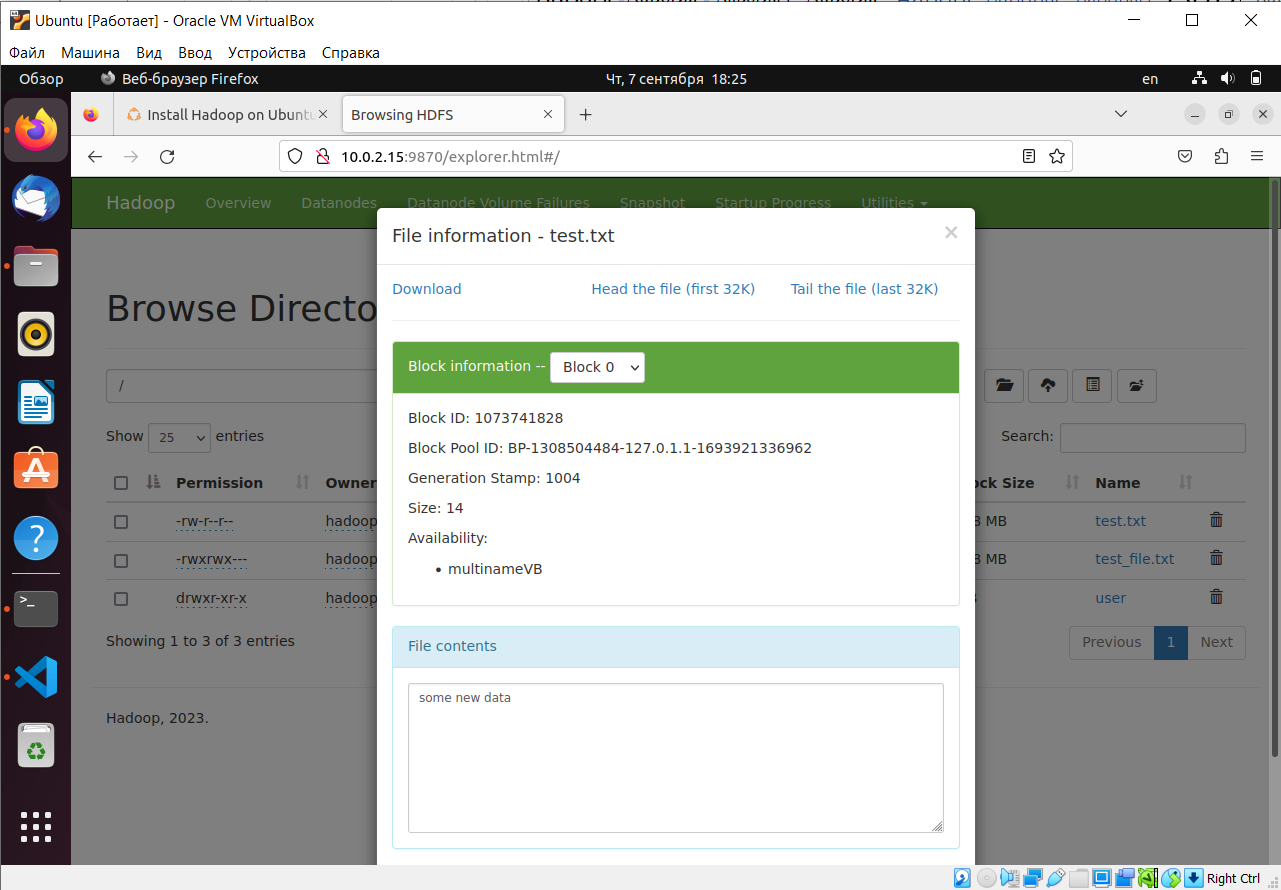


**Рис. 13.2.** Работа программы



**Рис. 13.3.** Работа программы





**Рис. 13.4.** Работа программы

**Вывод:** в ходе выполнения лабораторной работы были получены практические навыки по установке и настройке кластера Hadoop и работе с файловой системой HDFS.