



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 обеих файловых системах. Если в одной из них файл не существует, то программа должна скопировать его из второй файловой системы. Если файлы существуют в обеих файловых системах, то сохранить в обеих 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:

The screenshot shows the Hadoop DFS Health web interface in a browser. The address bar shows the URL `192.168.56.101:9870/dfshealth.html#tab-datanode`. The page title is "In operation". Below the title, there are filters for "DataNode State" (set to "All") and "Show" (set to "25" entries). A search bar is also present. The main content is a table with the following columns: Node, Http Address, Last contact, Last Block Report, Used, Non DFS Used, Capacity, Blocks, Block pool used, and Version. The table contains one entry for the DataNode `default-rack/hadoop2:9866 (192.168.56.102:9866)`, which is in the "In operation" state. The "Used" column shows 64 KB, "Non DFS Used" shows 16.86 GB, "Capacity" shows 23.94 GB, "Blocks" shows 3, and "Block pool used" shows 64 KB (0%). The "Version" is 3.3.6.

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Version
✓/default-rack/hadoop2:9866 (192.168.56.102:9866)	http://hadoop2:9864	0s	1m	64 KB	16.86 GB	23.94 GB	3	64 KB (0%)	3.3.6

Рис. 1. Кластер из двух серверов

```

hadoop@multinameVB:~$ hdfs dfs -ls /
hadoop@multinameVB:~$

```

Рис. 2. Содержимое корневой директории

```

hadoop@multinameVB:~$ hdfs dfs -mkdir /user
hadoop@multinameVB:~$ hdfs dfs -mkdir /user/hduser
hadoop@multinameVB:~$ hdfs dfs -lsr /
lsr: DEPRECATED: Please use 'ls -R' instead.
drwxr-xr-x - hadoop supergroup 0 2023-09-06 14:53 /user
drwxr-xr-x - hadoop supergroup 0 2023-09-06 14:53 /user/hduser

```

Рис. 3. Создание директории /user/hduser

```

hadoop@multinameVB:~$ hdfs dfs -mkdir /user/hduser/Hadoop
hadoop@multinameVB:~$ hdfs dfs -lsr /
lsr: DEPRECATED: Please use 'ls -R' instead.
drwxr-xr-x - hadoop supergroup 0 2023-09-06 14:53 /user
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:17 /user/hduser
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:17 /user/hduser/Hadoop

```

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

```

hadoop@multinameVB:~$ hdfs dfs -touchz /user/hduser/Hadoop/MichaelIUK4.txt
hadoop@multinameVB:~$ hdfs dfs -lsr /
lsr: DEPRECATED: Please use 'ls -R' instead.
drwxr-xr-x - hadoop supergroup 0 2023-09-06 14:53 /user
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:17 /user/hduser
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:29 /user/hduser/Hadoop
-rw-r--r-- 1 hadoop supergroup 0 2023-09-06 15:29 /user/hduser/Hadoop/MichaelIUK4.txt

```

Рис. 5. Создание текстового файла

```
hadoop@multinameVB:~$ hadoop fs -appendToFile - /user/hduser/Hadoop/MichaelIUK4.txt
qwerty
wasd
:q
^Chadoop@multinameVB:~$ hdfs dfs -cat /user/hduser/Hadoop/MichaelIUK4.txt
qwerty
wasd
:q
hadoop@multinameVB:~$
```

Рис. 6. Заполнение файла из консоли

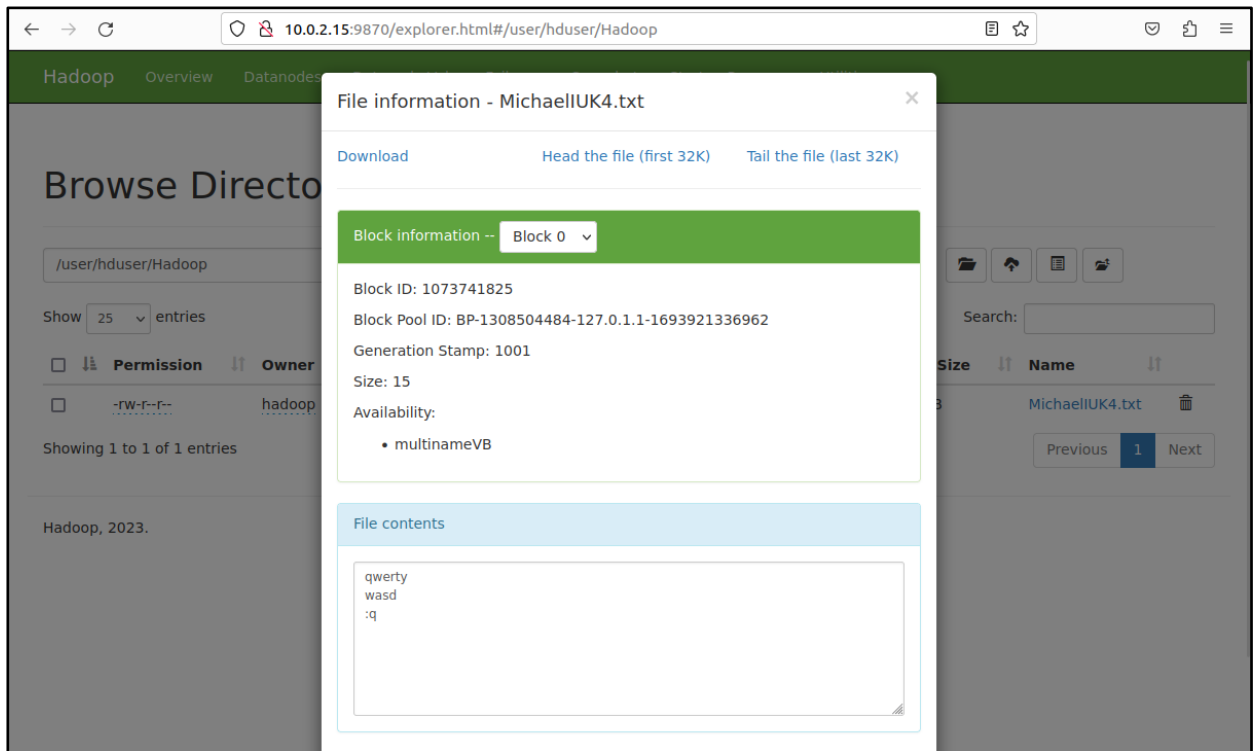


Рис. 7. Проверка наличия файла через web-интерфейс

```
hadoop@multinameVB:~$ hdfs dfs -get /user/hduser/Hadoop/MichaelIUK4.txt ./
hadoop@multinameVB:~$ ls
hadoop-3.3.6.tar.gz  hdfs  MichaelIUK4.txt  snap
```

Рис. 8. Копирование файла на локальную систему

```
hadoop@multinameVB:~$ touch test_file.txt
hadoop@multinameVB:~$ ls
hadoop-3.3.6.tar.gz  hdfs  MichaelIUK4.txt  snap  test_file.txt
hadoop@multinameVB:~$ hdfs dfs -put test_file.txt /
```

Рис. 9. Создание и перенос файла на HDFS

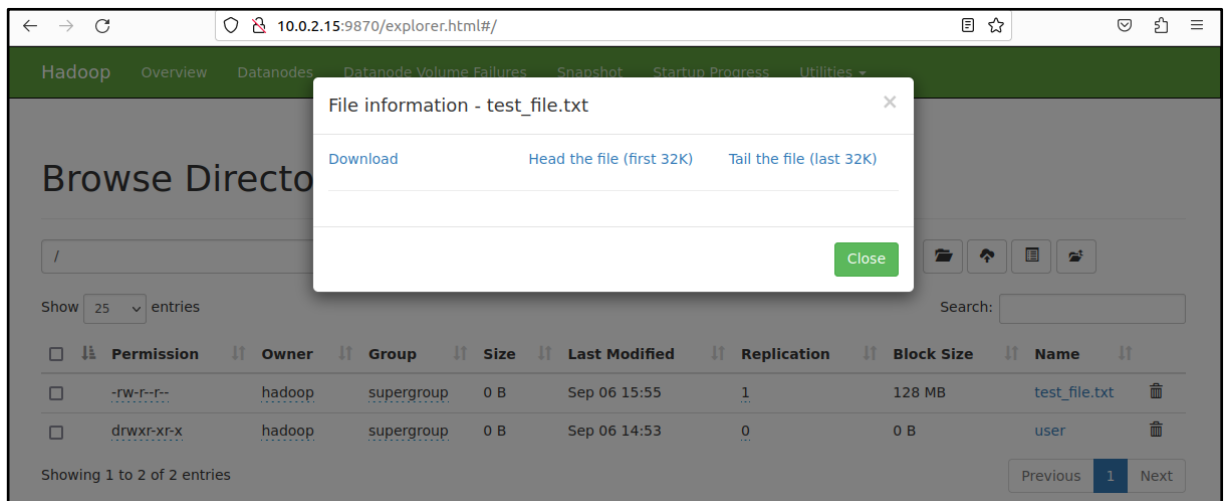


Рис. 10. Проверка наличия файла через web-интерфейс

```
hadoop@multinameVB:~$ hdfs dfs -lsr /
lsr: DEPRECATED: Please use 'ls -R' instead.
-rw-r--r-- 1 hadoop supergroup 0 2023-09-06 15:55 /test_file.txt
drwxr-xr-x - hadoop supergroup 0 2023-09-06 14:53 /user
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:17 /user/hduser
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:29 /user/hduser/Hadoop
-rw-r--r-- 1 hadoop supergroup 15 2023-09-06 15:34 /user/hduser/Hadoop/MichaelIUK4.txt
```

Рис. 11. Просмотр прав доступа

```
hadoop@multinameVB:~$ hadoop fs -chmod 770 /test_file.txt
hadoop@multinameVB:~$ hdfs dfs -ls -R /
-rwxrwx--- 1 hadoop supergroup 0 2023-09-06 15:55 /test_file.txt
drwxr-xr-x - hadoop supergroup 0 2023-09-06 14:53 /user
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:17 /user/hduser
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:29 /user/hduser/Hadoop
-rw-r--r-- 1 hadoop supergroup 15 2023-09-06 15:34 /user/hduser/Hadoop/MichaelIUK4.txt
```

Рис. 12. Изменение прав доступа

```
hadoop@multinameVB:~$ hdfs dfs -ls -R /
-rwxrwx--- 1 hadoop supergroup 0 2023-09-06 15:55 /test_file.txt
drwxr-xr-x - hadoop supergroup 0 2023-09-06 14:53 /user
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:17 /user/hduser
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:29 /user/hduser/Hadoop
-rw-r--r-- 1 hadoop supergroup 15 2023-09-06 15:34 /user/hduser/Hadoop/MichaelIUK4.txt
hadoop@multinameVB:~$ ls
hadoop-3.3.6.tar.gz script.py Видео Изображения 'Рабочий стол'
hdfs snap Документы Музыка Шаблоны
MichaelIUK4.txt test_file.txt Загрузки Общедоступные
hadoop@multinameVB:~$ python3 script.py
local path: fakefile.dat
HDFS path: /fakefile.dat
file was not found nowhere
```

Рис. 13.1. Работа программы

```
hadoop@multinameVB:~$ python3 script.py
local path: test_file.txt
HDFS path: /test.txt
hadoop@multinameVB:~$ hdfs dfs -ls -R /
-rw-r--r-- 1 hadoop supergroup 0 2023-09-07 17:48 /test.txt
-rwxrwx--- 1 hadoop supergroup 0 2023-09-06 15:55 /test_file.txt
drwxr-xr-x - hadoop supergroup 0 2023-09-06 14:53 /user
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:17 /user/hduser
drwxr-xr-x - hadoop supergroup 0 2023-09-06 15:29 /user/hduser/Hadoop
-rw-r--r-- 1 hadoop supergroup 15 2023-09-06 15:34 /user/hduser/Hadoop/MichaelIUK4.txt
```

Рис. 13.2. Работа программы

```
hadoop@multinameVB:~$ python3 script.py
local path: test.txt
HDFS path: /test.txt
hadoop@multinameVB:~$ ls
hadoop-3.3.6.tar.gz script.py test.txt Загрузки Общедоступные
hdfs snap Видео Изображения 'Рабочий стол'
MichaelIUK4.txt test file.txt Документы Музыка Шаблоны
```

Рис. 13.3. Работа программы

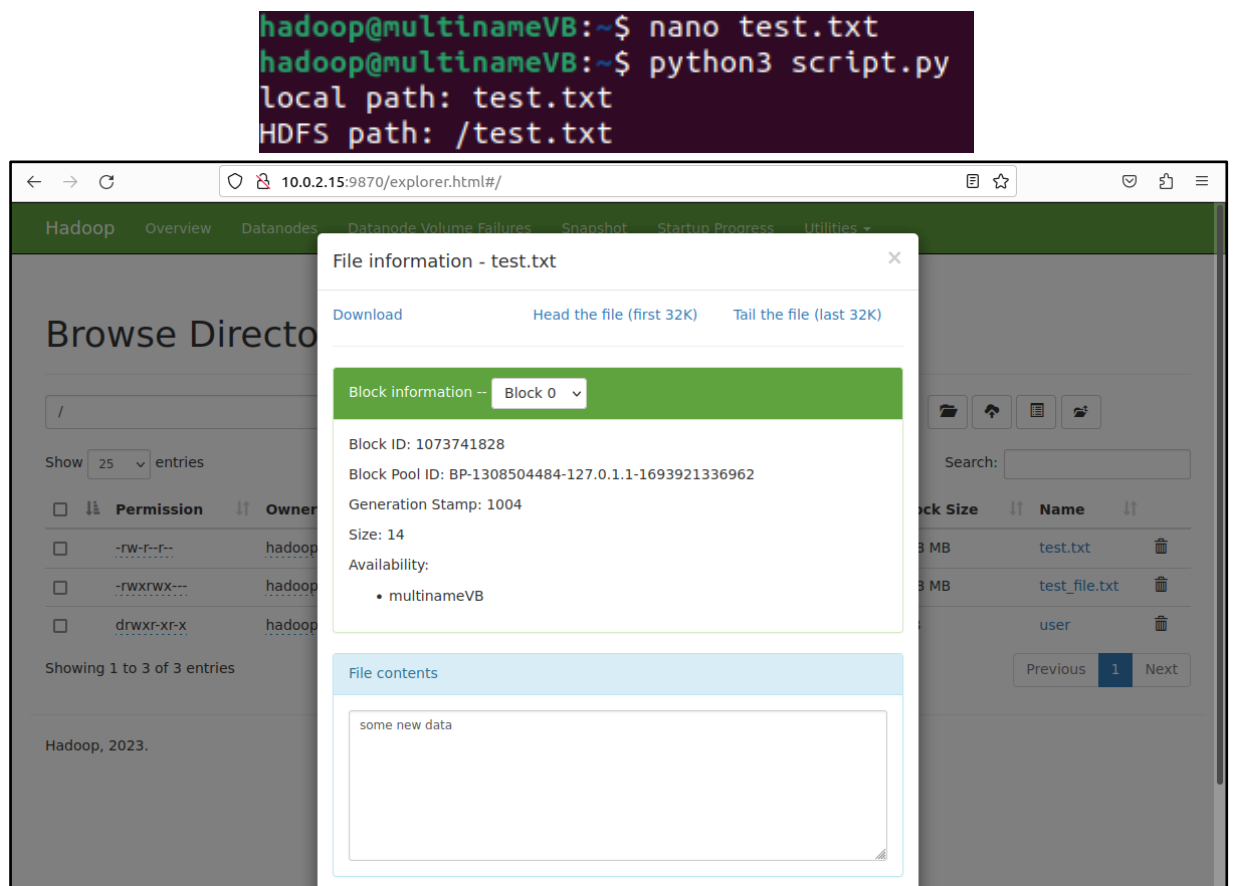


Рис. 13.4. Работа программы

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