МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ

Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования

**«Московский технический университет связи и информатики»**

Кафедра «Системное программирование»

Практическая работа №1

**«Введение в HDFS»**

по дисциплине

Введение в большие данные

Выполнил: студент гр. БЭИ2202

Кулешов А. С.

Проверил: Сергеев А.Е.

Москва, 2025 г

Цель работы. Получить навыки работы с файловой системой HDFS

**ВЫПОЛНЕНИЕ**

Скачаю образ Ubuntu, а также виртуальную машину, настрою виртуальную машину в соответствии с заданием, после проведу установку и распаковку образа ОС.

Затем открою терминал, обновлю пакеты и выведу системную информацию на экран, скачаю Vim и MC. Проверю их наличие, узнаю из какого репозитория они были скачаны, узнаю их зависимости. Результат работы можно увидеть на рисунках 1-20.

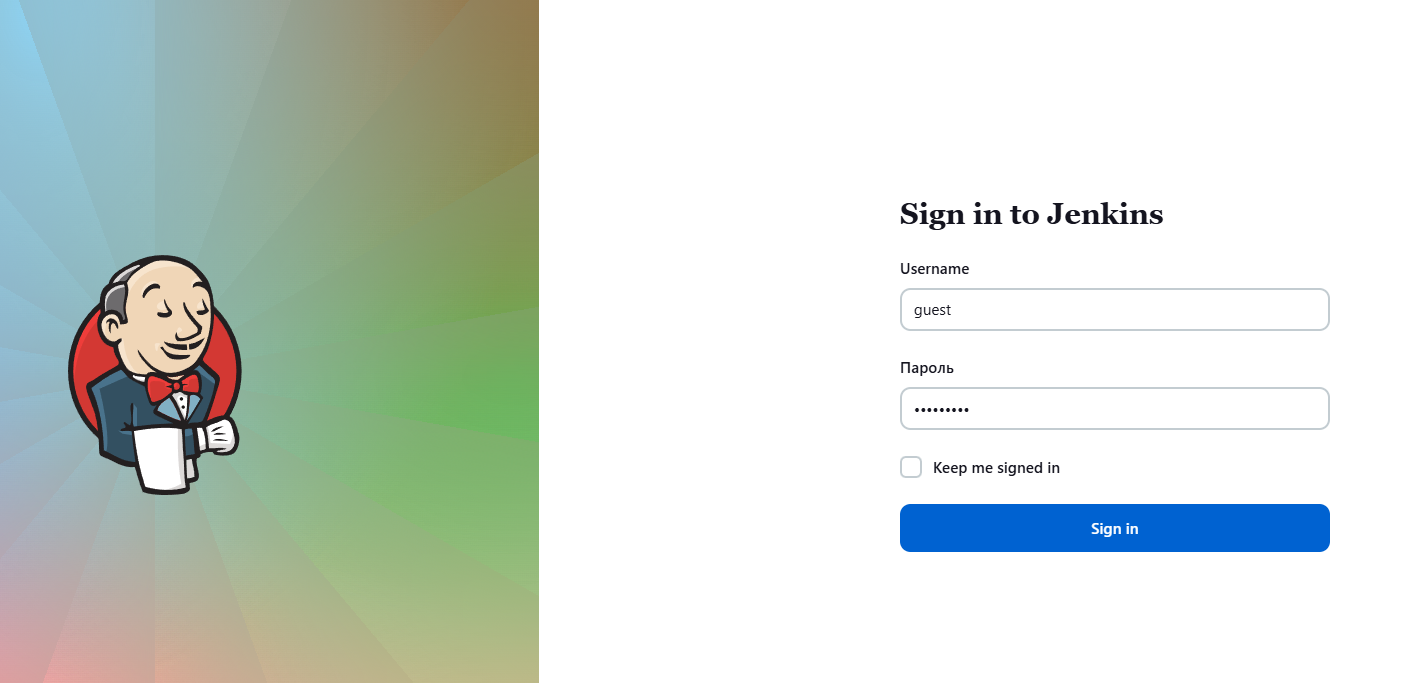


Рисунок 1 – Вход на сайт

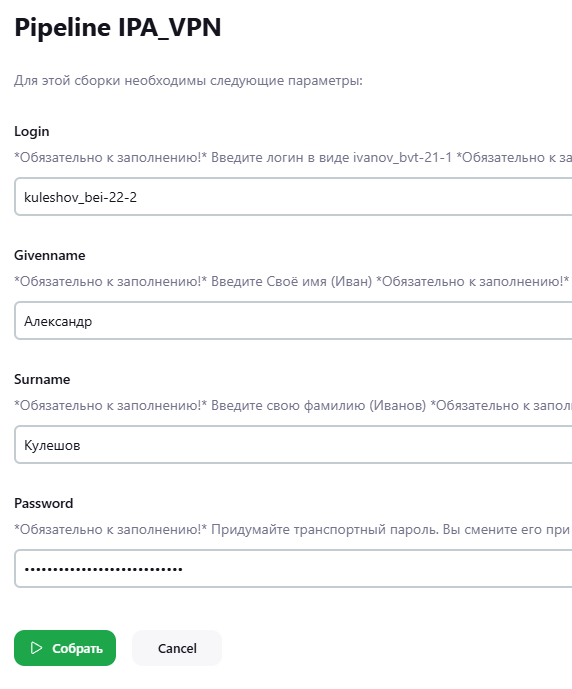


Рисунок 2 – Создание аккаунта



Рисунок 3 – скачивание архива

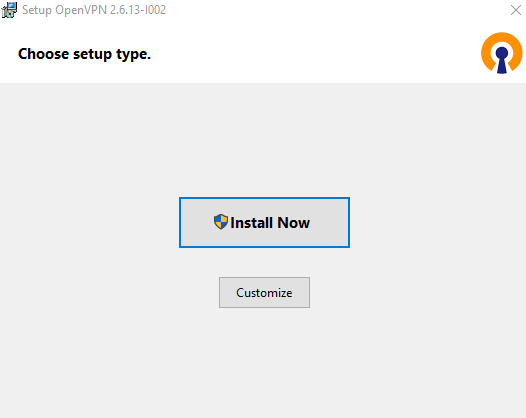


Рисунок 4 – Установка OpenVPN

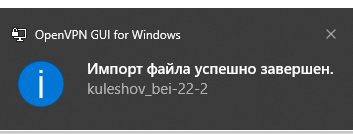


Рисунок 5 – Импорт настроек

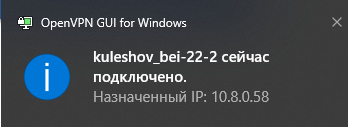


Рисунок 6 – Подключение

Вариант 16 -> 5 (node5.mtuci.cloud.ru, 192.168.0.7)

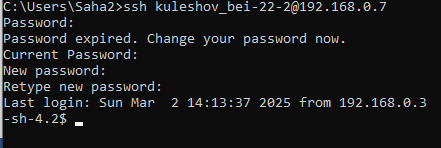


Рисунок 7 – Смена пароля при подключении по ssh

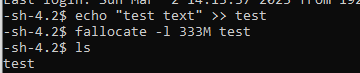


Рисунок 8 – Создание тестового файла

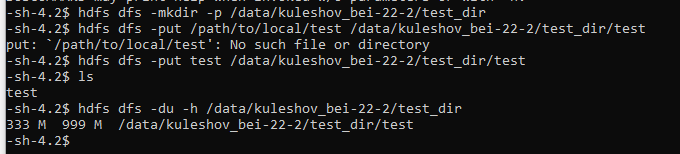


Рисунок 9 – Создание директив и перемещение файла

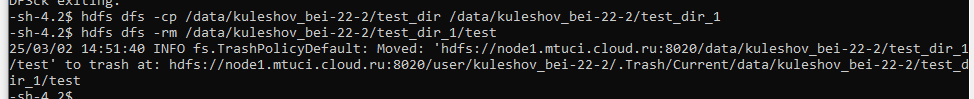


Рисунок 10 – Копирование директории и удаление файла



Рисунок 11 – Узнавание размера файла

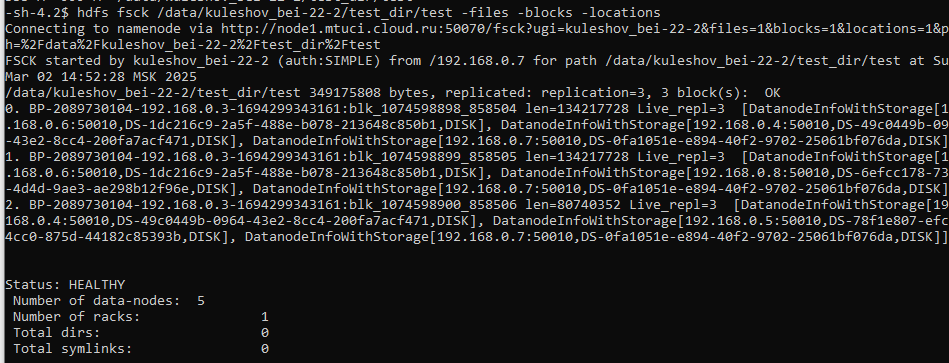


Рисунок 12 – Как файл храниться на файловой системе

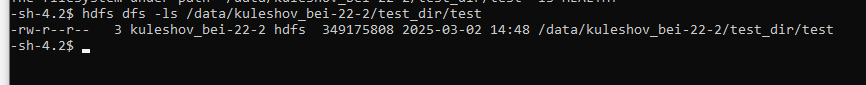


Рисунок 13 – Фактор репликации - 3



Рисунок 14 – Число блоков - 3

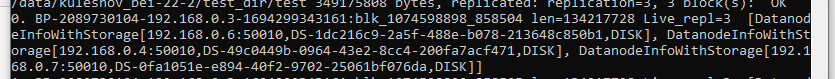


Рисунок 15 – Данные по первому блоку

Таблица 1 – Данные по первому блоку

|  |  |
| --- | --- |
| 0 | Номер блока по порядку |
| BP-2089730104-192.168.0.3-1694299343161 | Идентификатор block pull |
| blk\_1074598898 | Идентификатор блока |
| 858504 | Generation stamp. |
| 134217728 байт (128 МБ) | Объем блока |
| 3 | Количество живых реплик блока |
| DatanodeInfoWithStorage [192.168.0.6:50010 | IP-адрес и порт, по которому доступен блок |
| DS-1dc216c9-2a5f-488e-b078-213648c850b1 | Data Storage ID идентификатор ноды (Если у ноды изменится hostname или IP-адрес, нода всё равно будет идентифицироваться внутри HDFS) |
| DISK | Способ хранения блока (Может также хранится в S3-хранилище) |

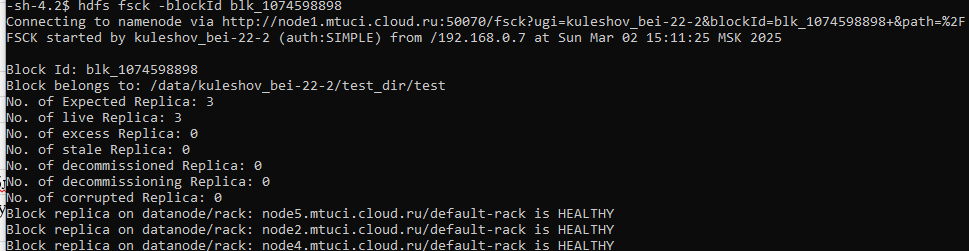


Рисунок 16 – Подробная информация о состоянии блока (Задание 4)

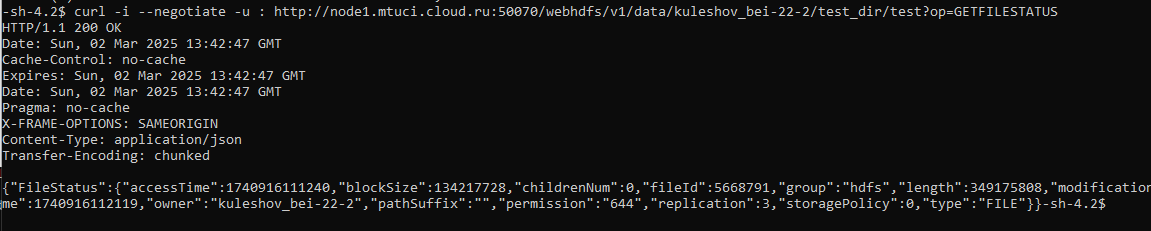


Рисунок 18 – Вывод информации о файле (Задание 5-6)

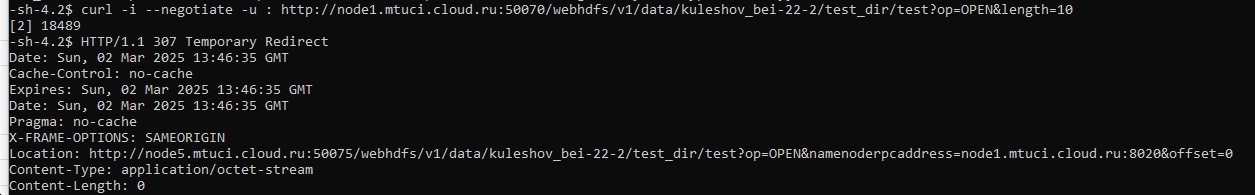


Рисунок 19 – Вывод 10 символов (редирект)

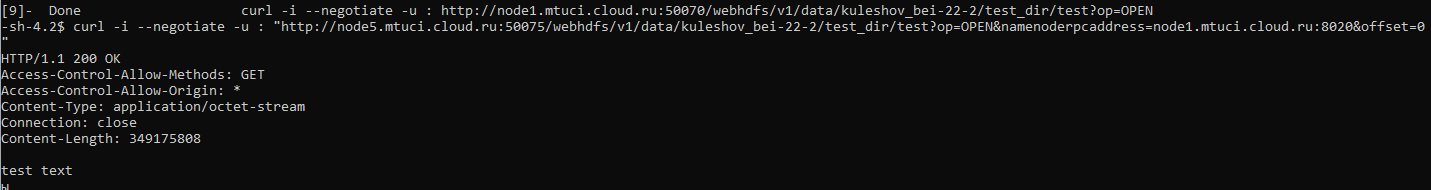


Рисунок 20 – Вывод 10 символов (успех)

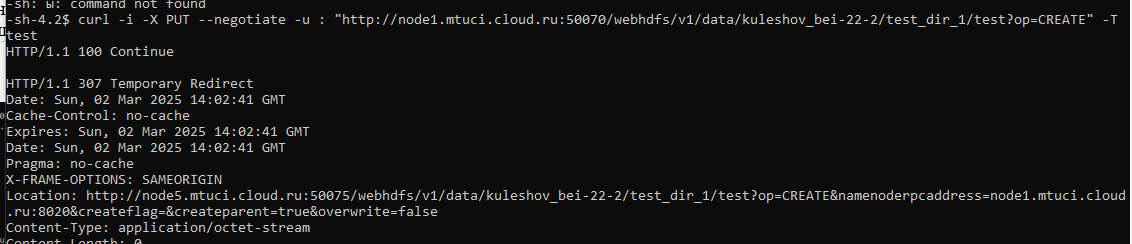


Рисунок 21 – Копирование (редирект)

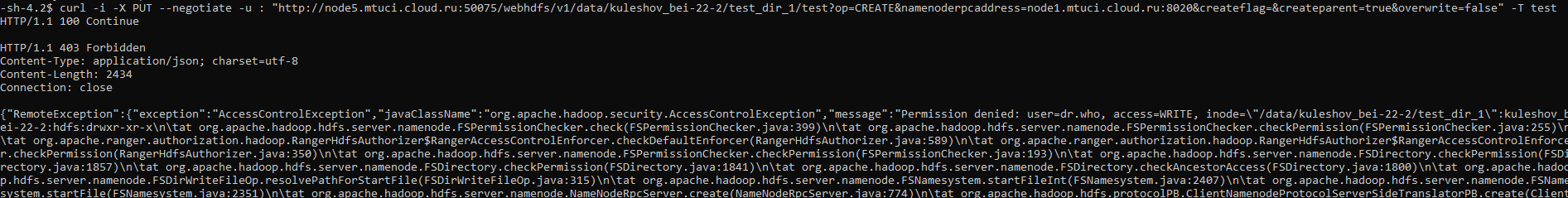


Рисунок 22 – Копирование (ошибка)



Рисунок 23 – Открытие прав доступа

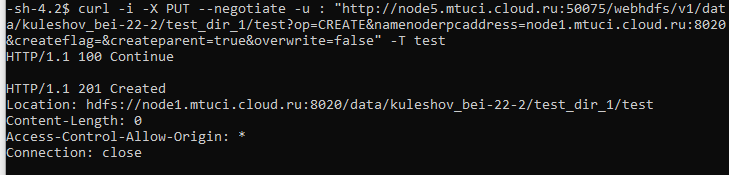


Рисунок 24 – Копирование (успех)

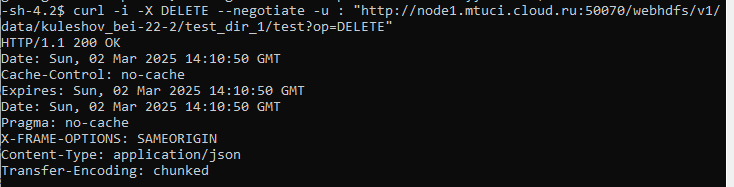


Рисунок 25 – Удаление

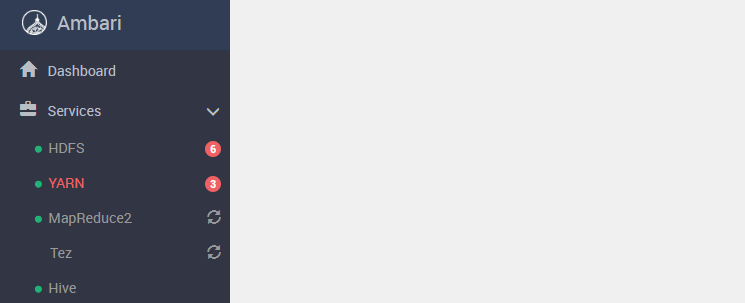


Рисунок 26 – Успешный логин в Ambari



Рисунок 27 – Переход в NameNode UI

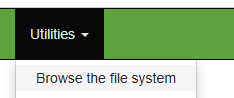


Рисунок 28 – Переход в просмотр файловой системы

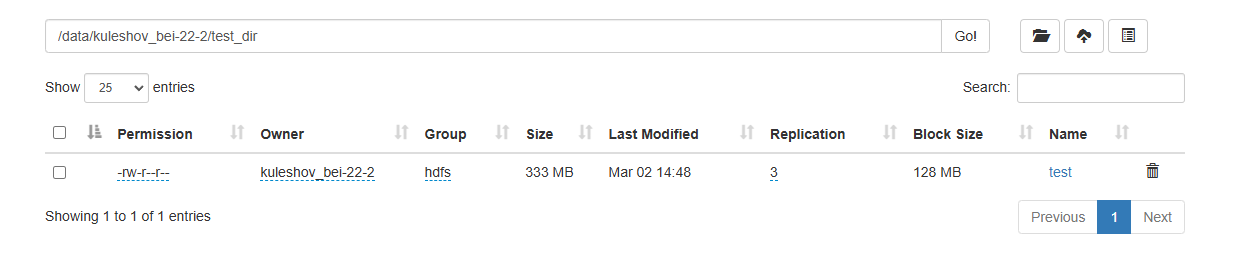


Рисунок 29 – Мой файл

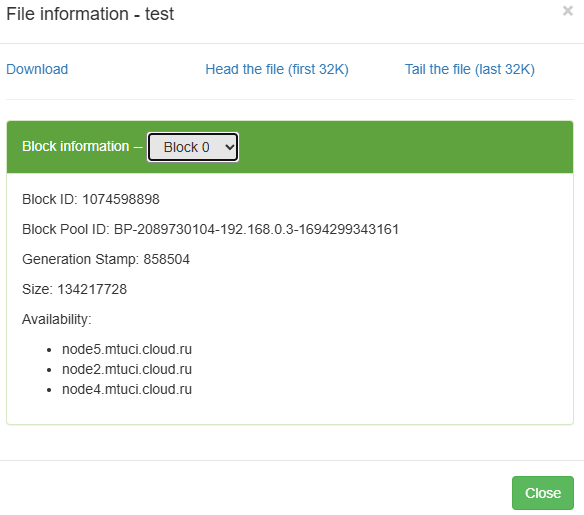


Рисунок 30 – Информация о блоках данных (0)

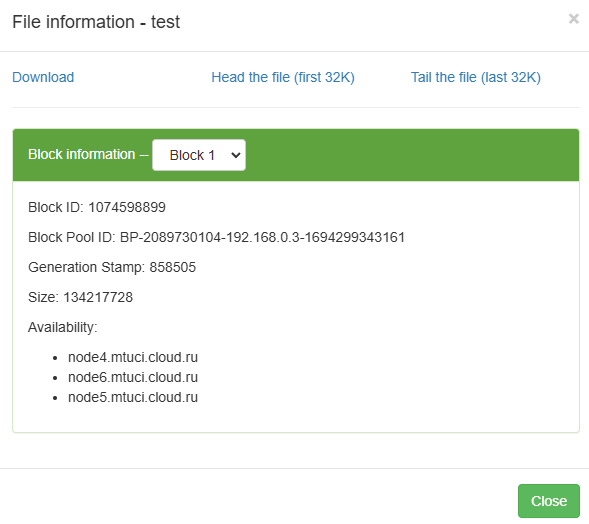


Рисунок 31 – Информация о блоках данных (1)

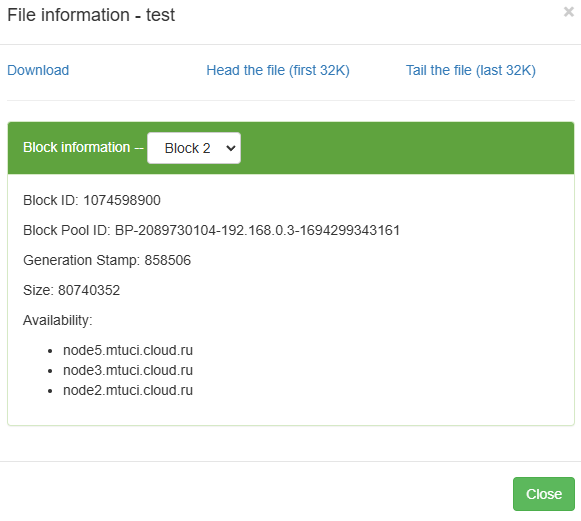


Рисунок 32 – Информация о блоках данных (2)

1. Ответьте на вопрос: какой объём стандартного блока HDFS исходя из показателя Size?
   1. 128 МБайт
2. Есть ли блок, который отличается по размеру от остальных?
   1. Да, это Block2
3. Сколько пространства занимает последний блок файла в локальной файловой системе?
   1. 10.09 МБайт
4. Подумайте, при каких обстоятельствах количество реплик блоков файла может превышать стандартный фактор репликации?
   1. Это может быть сделано вручную администратором или быть ошибкой конфигурации