|  |  |
| --- | --- |
|  | **Министерство науки и высшего образования Российской Федерации**  **Калужский филиал федерального государственного автономного 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**

**«ОСНОВЫ HADOOP. УСТАНОВКА HADOOP. ОСНОВНЫЕ КОМАНДЫ ФАЙЛОВОЙ СИСТЕМЫ HDFS»**

**по дисциплине: «Технологии обработки больших данных»**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Выполнил: студент группы ИУК4-72Б | |  |  | Губин Е.В. | |
|  | | (Подпись) |  | (И.О. Фамилия) | |
| Проверил: | |  |  | Голубева С.Е. | |
|  | | (Подпись) |  | (И.О. Фамилия) | |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | | | |
|  | - Балльная оценка:  - Оценка: | | | |

Калуга, 2025

**Цель работы:** 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.

**Вариант 7**

**Формулировка задания:**

Напишите программу, которая будет рекурсивно выводить на экран список подкаталогов и файлов для заданного каталога в HDFS.

**Ход выполнения работы:**

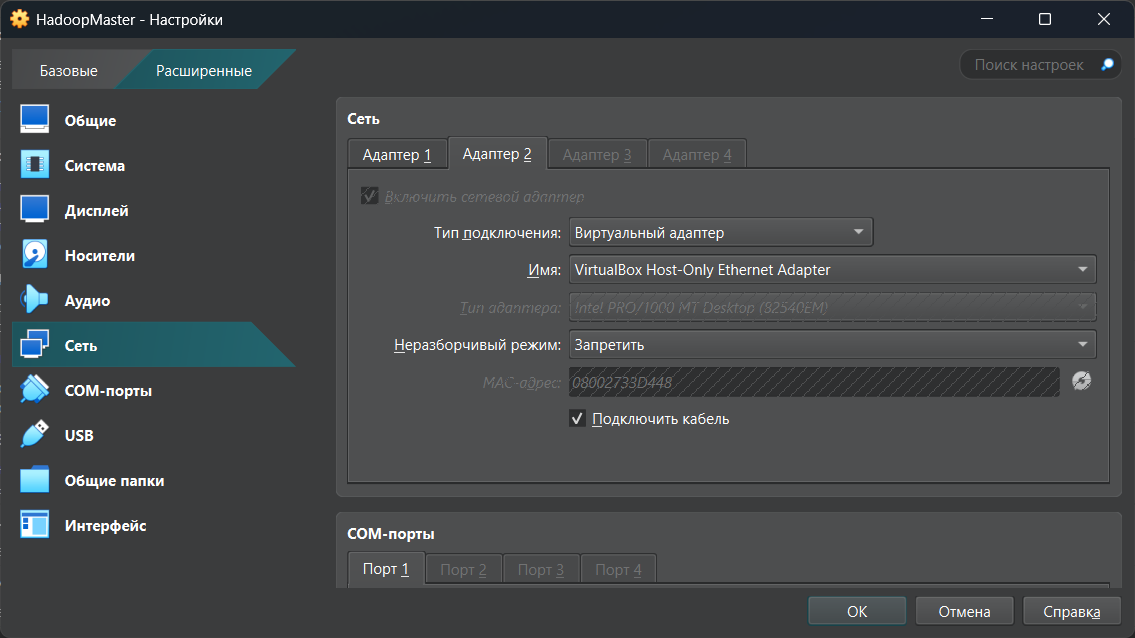


Рис.1 Настройка виртуального адаптера для master и slave

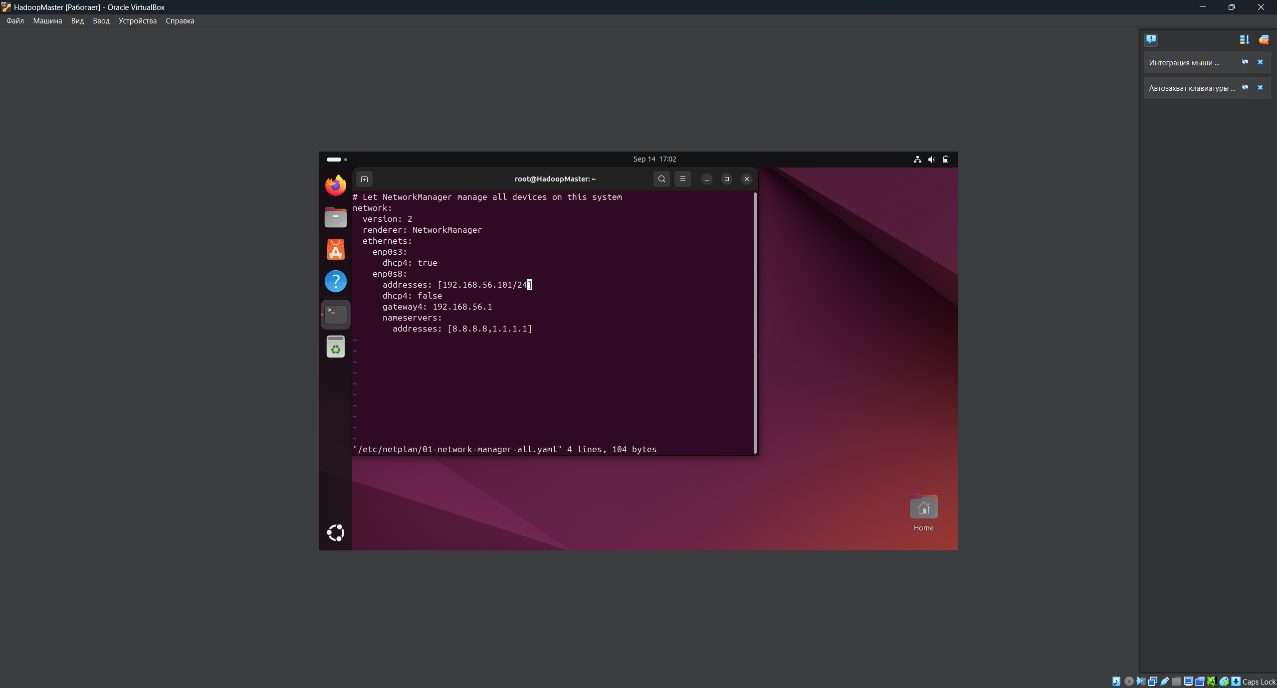


Рис.2 Настройка netplan для master

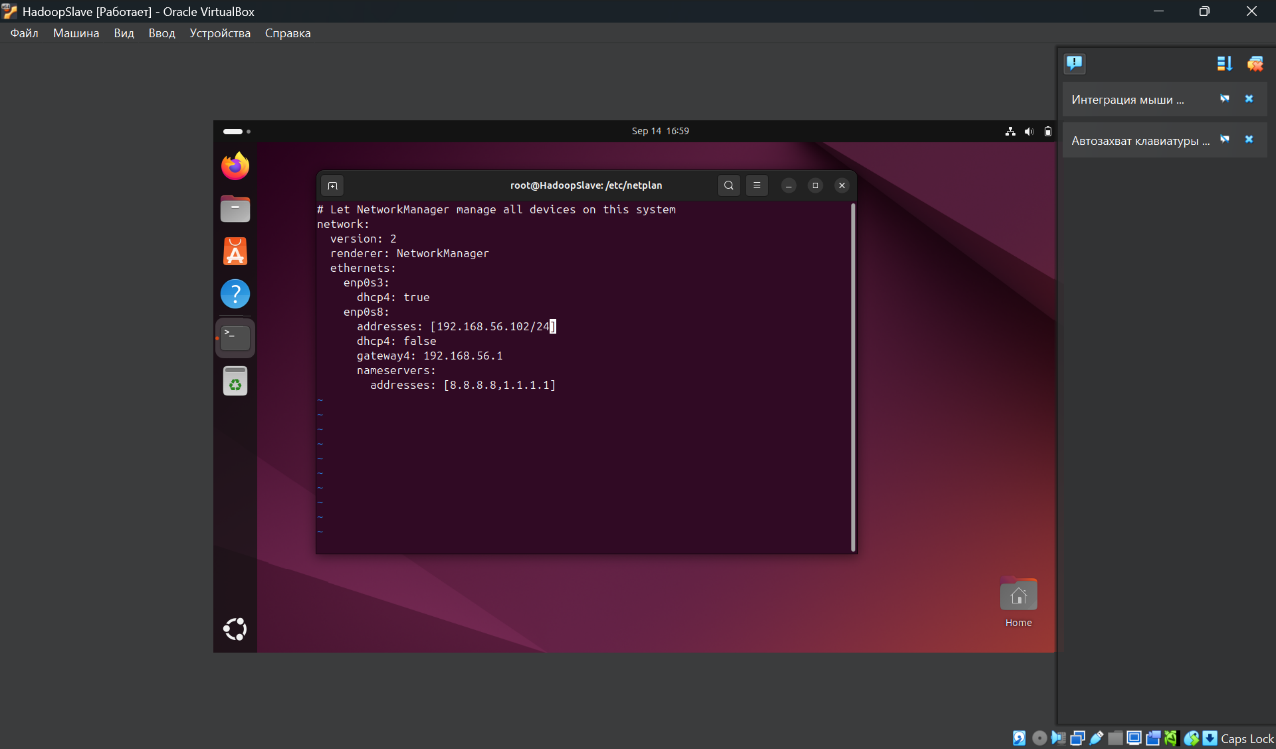


Рис.3 Настройка netplan для slave

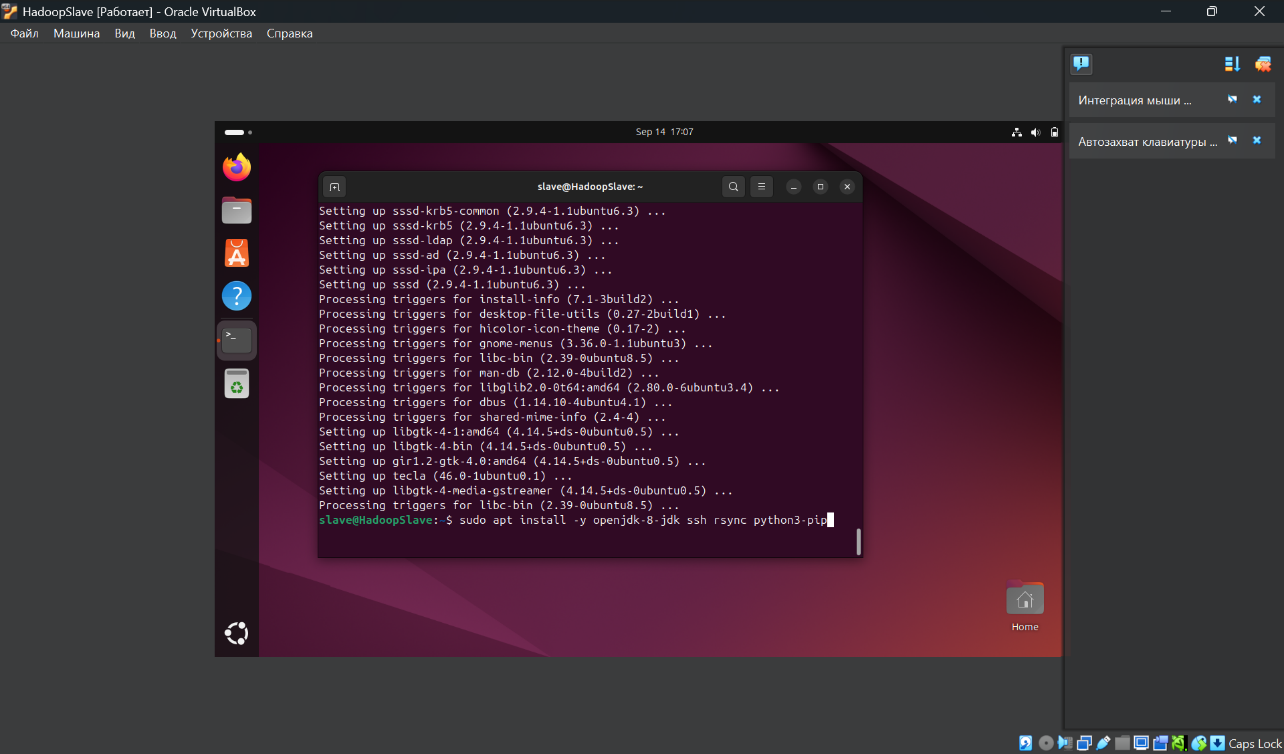


Рис.4 Установка доп пакетов для hadoop для master и slave

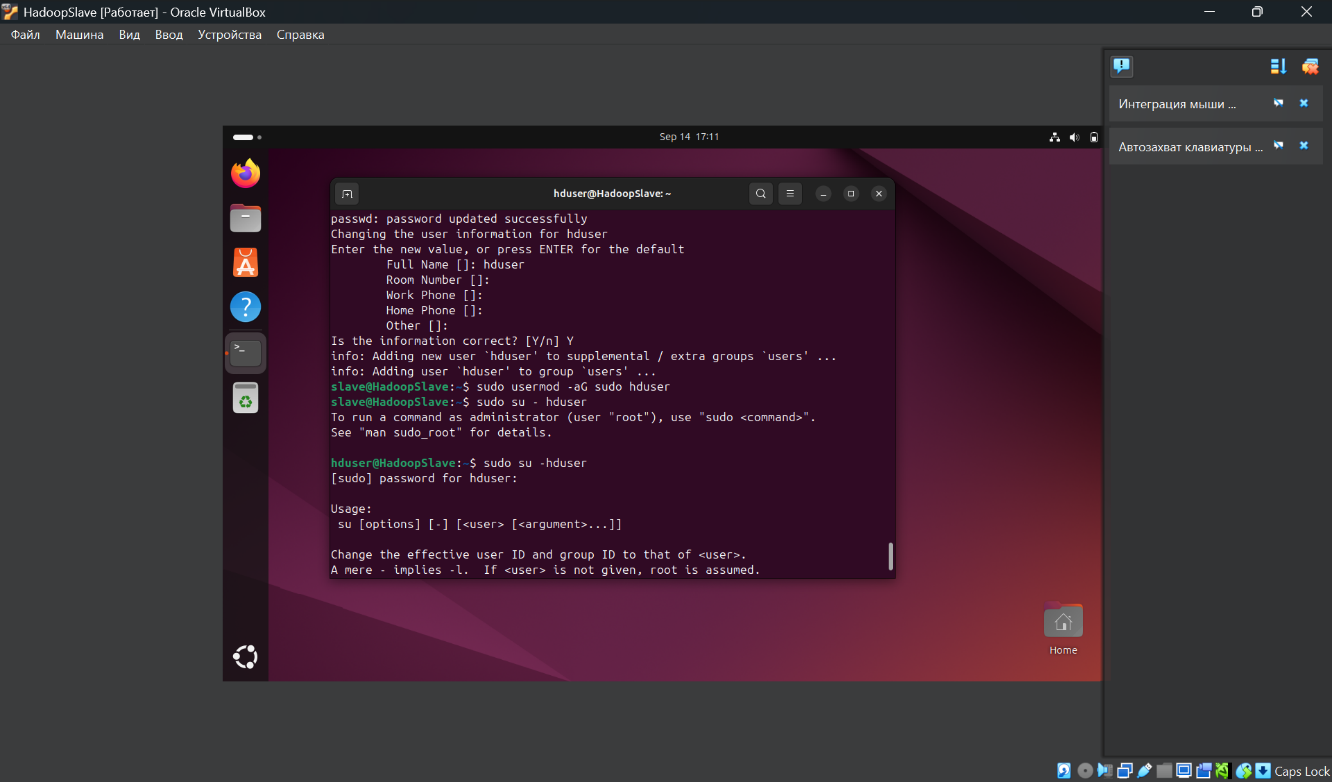


Рис.5 Создаём hduser и заходим под ним на master и slave

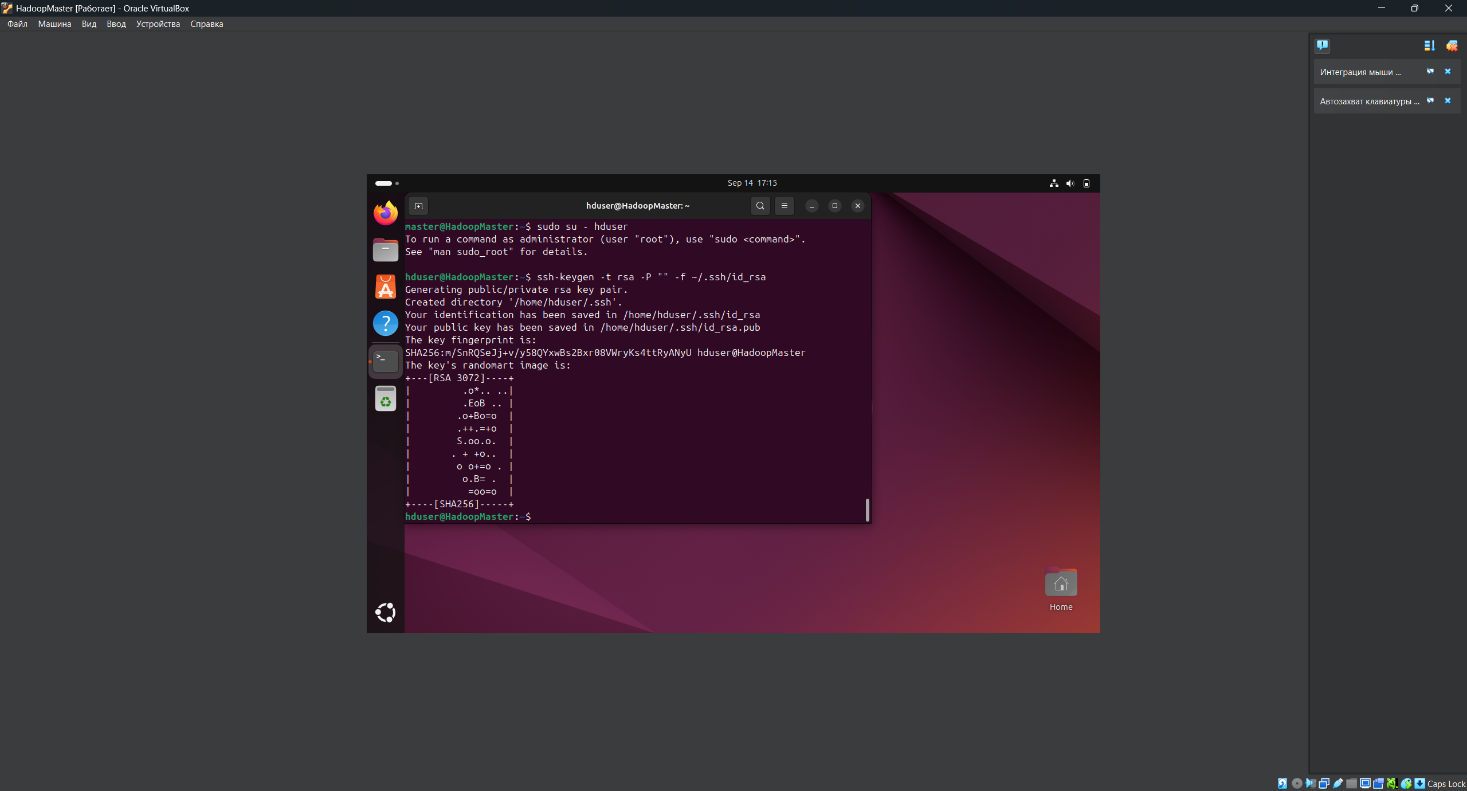


Рис.6 Создание SSH ключа на master

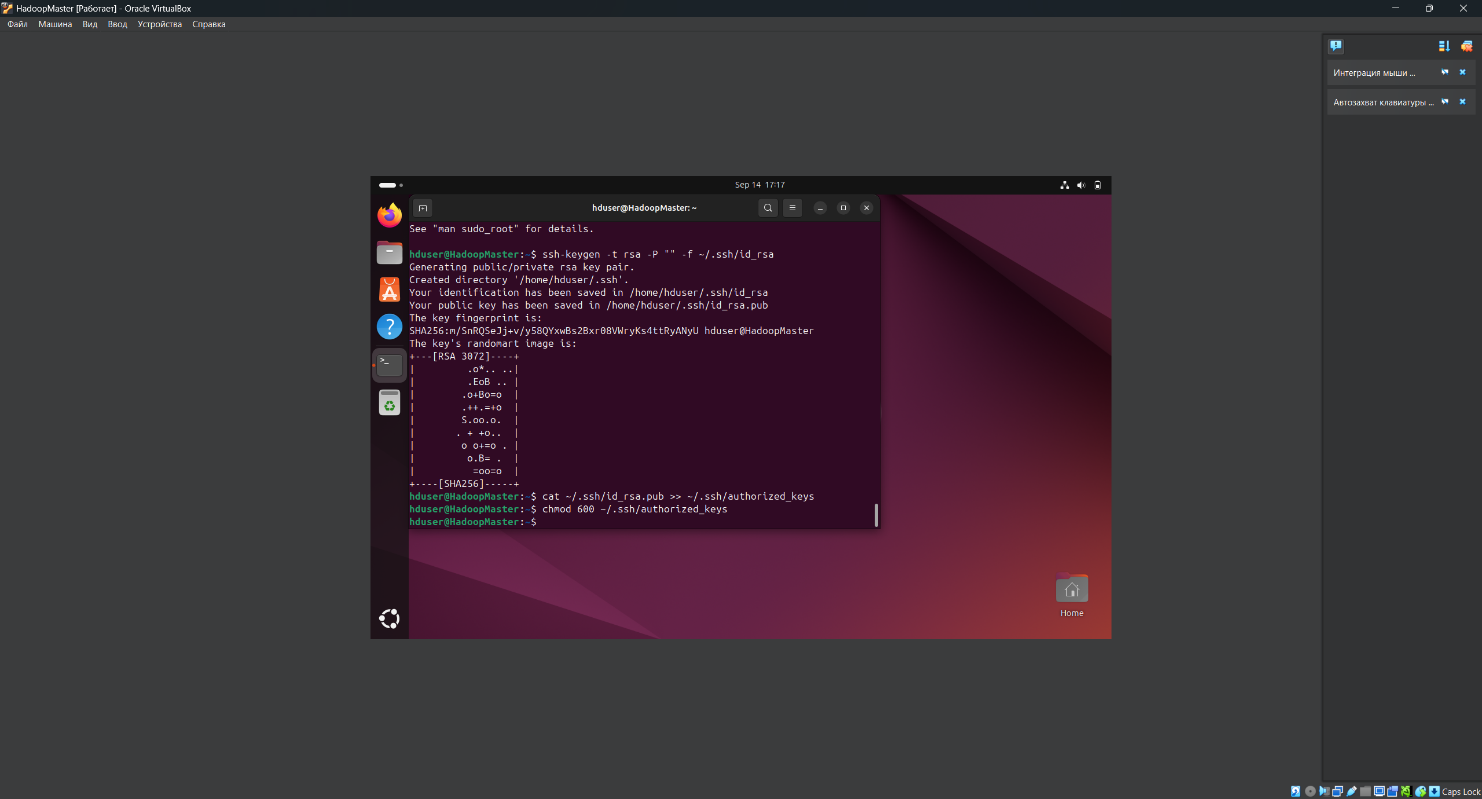


Рис.7 Копирование публичного ключа в authorized\_keys локально

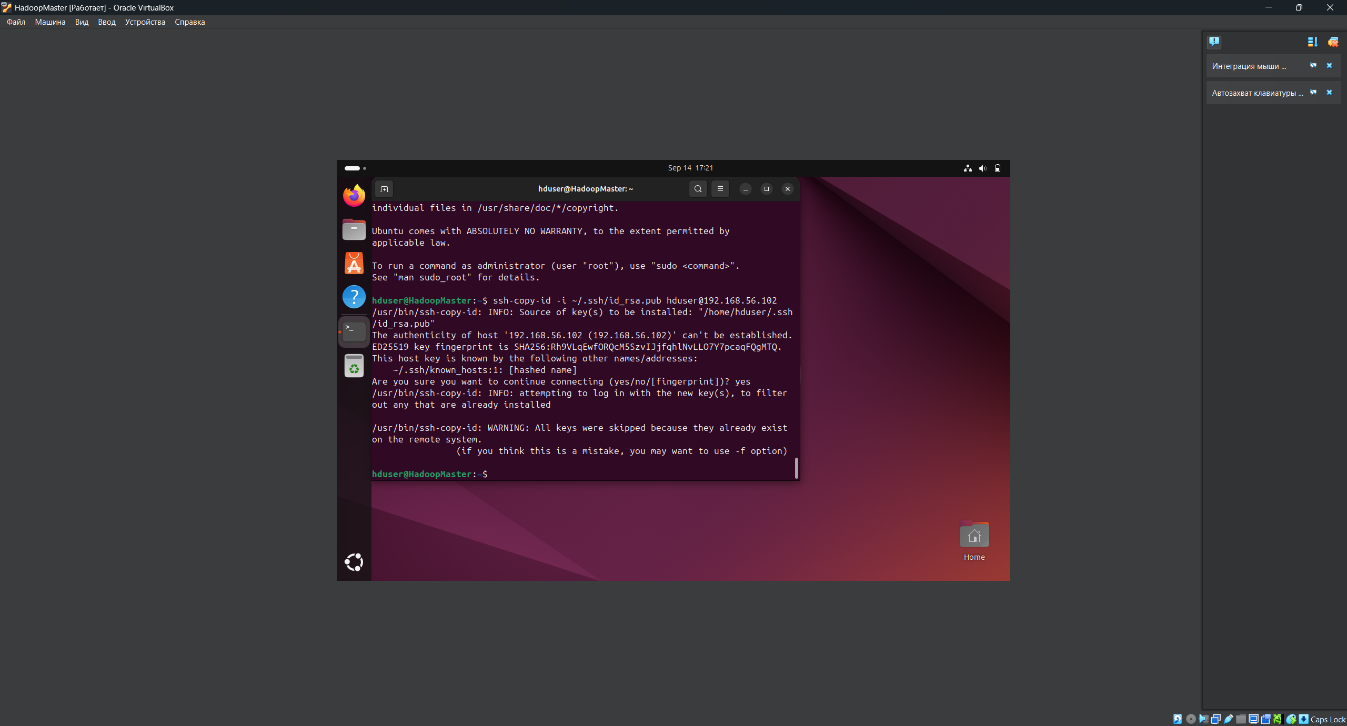


Рис.8 Копирование ключа на slave с master

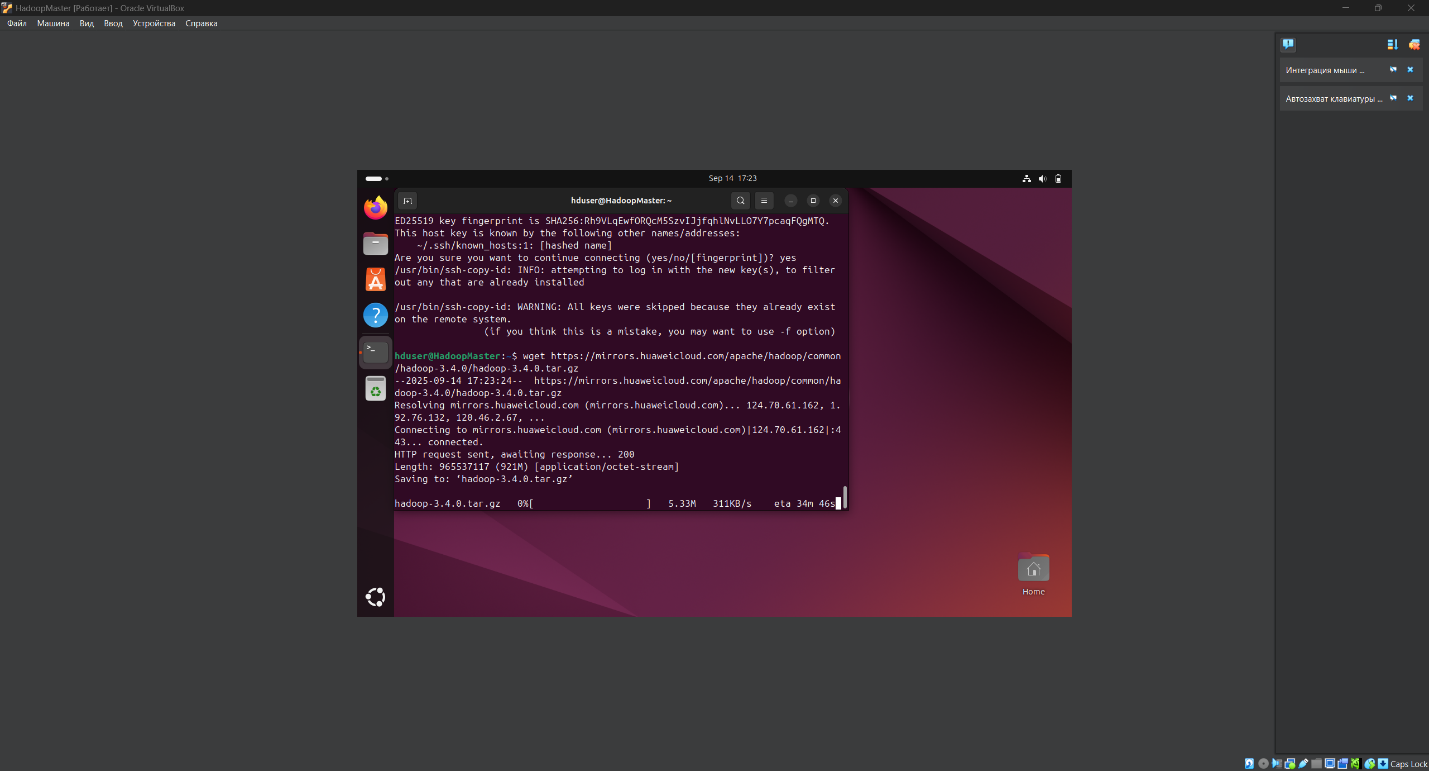


Рис.9 Установка hadoop на мастер

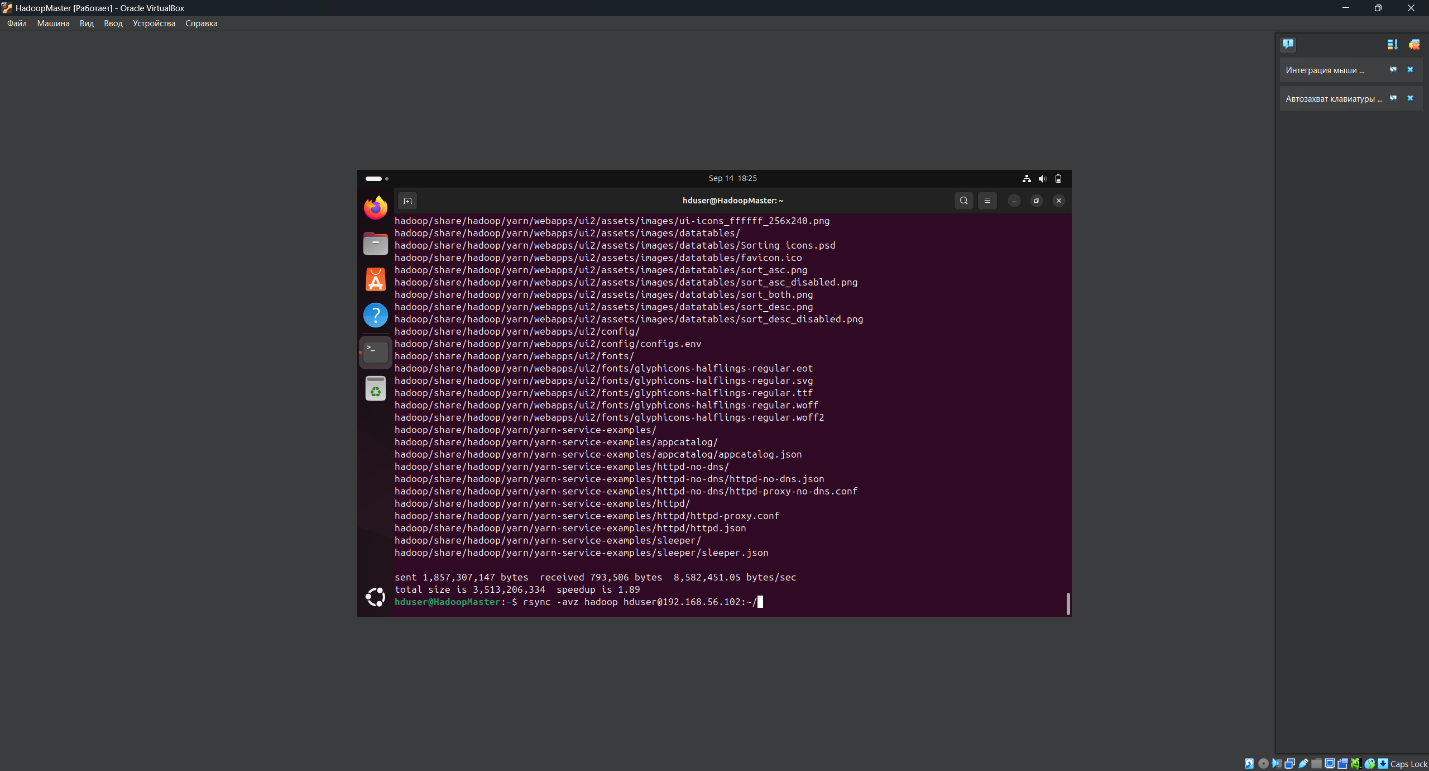


Рис.10 Перемещение hadoop от master к slave

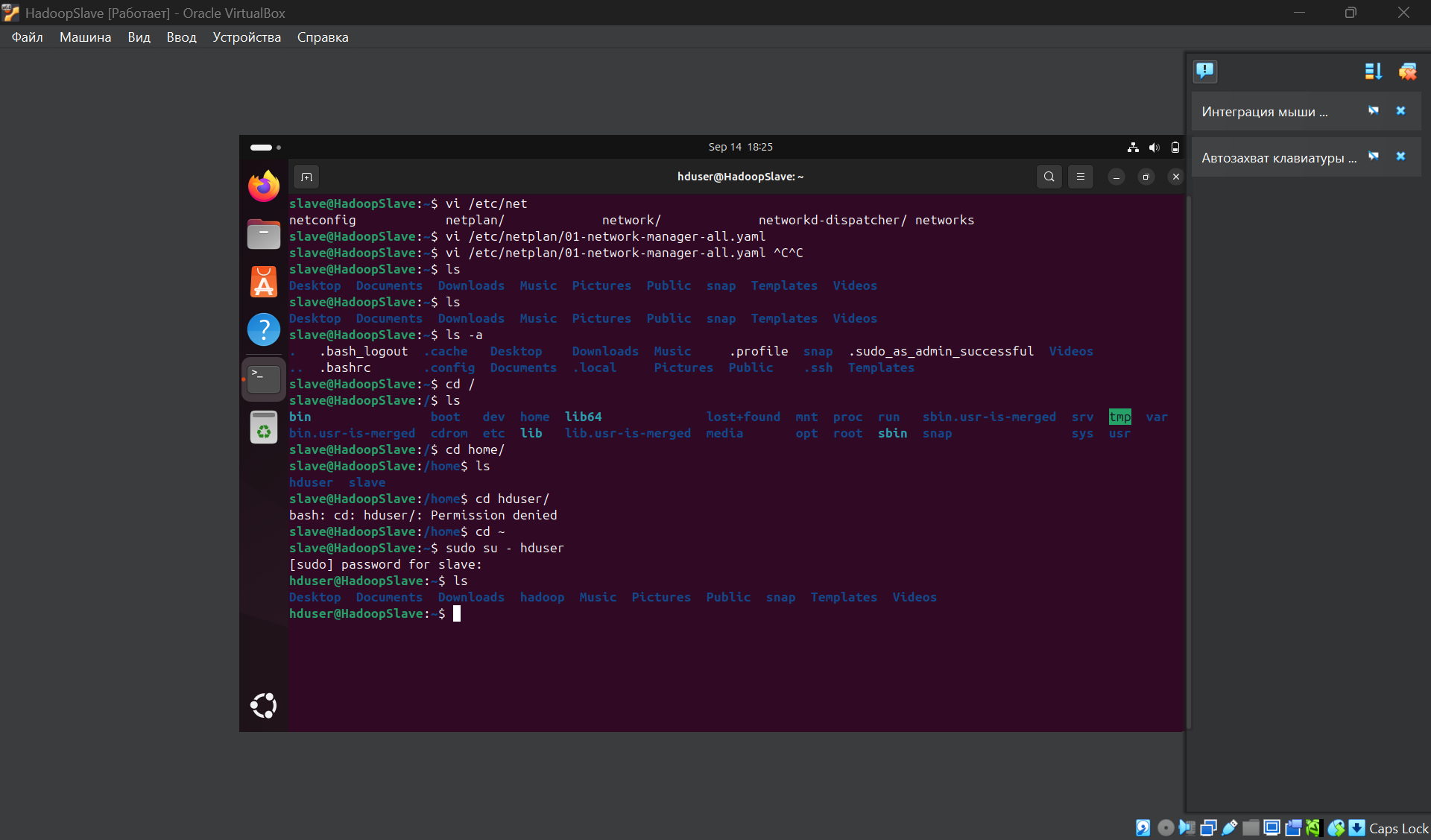


Рис.11 Перемещённый hadoop у slave

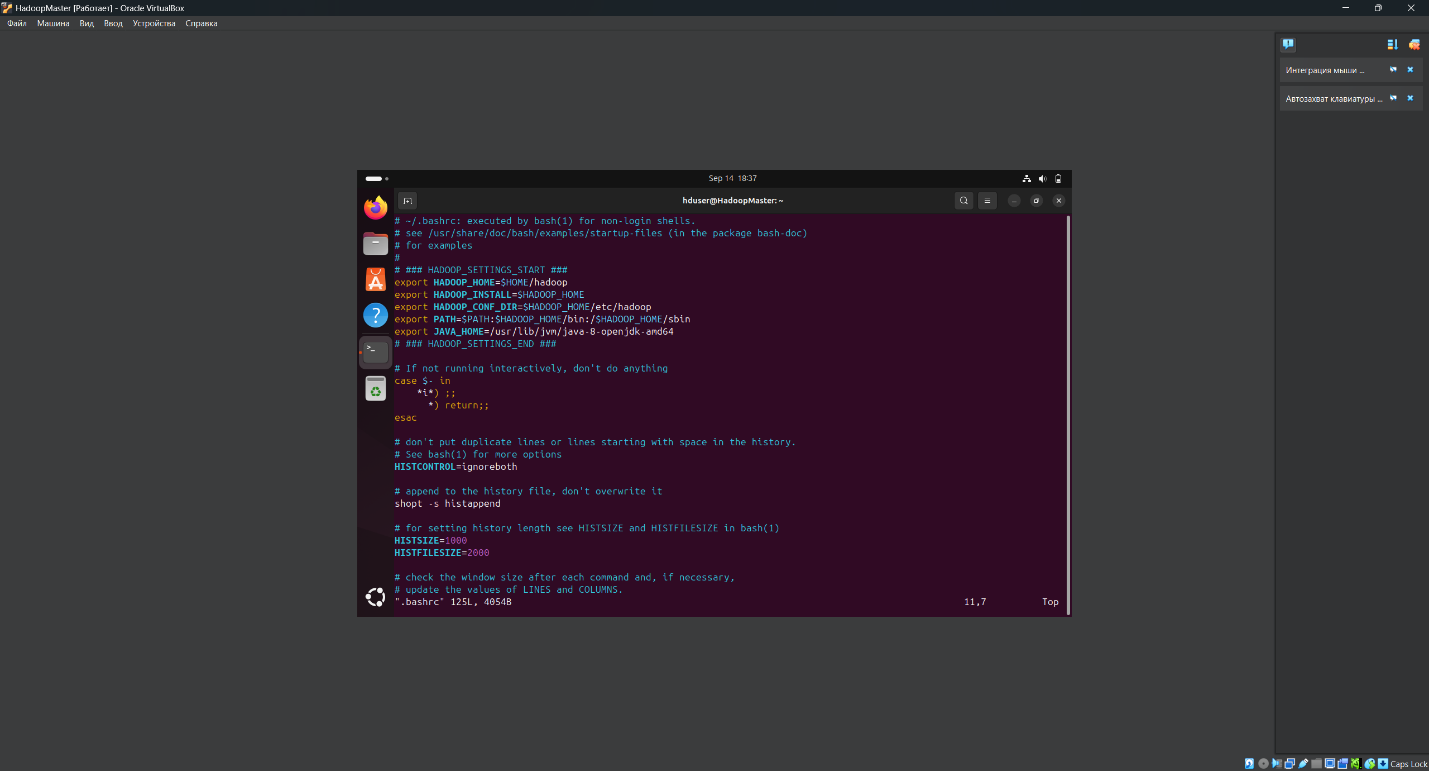


Рис.12 Настройка переменных среды для hadoop на master и slave

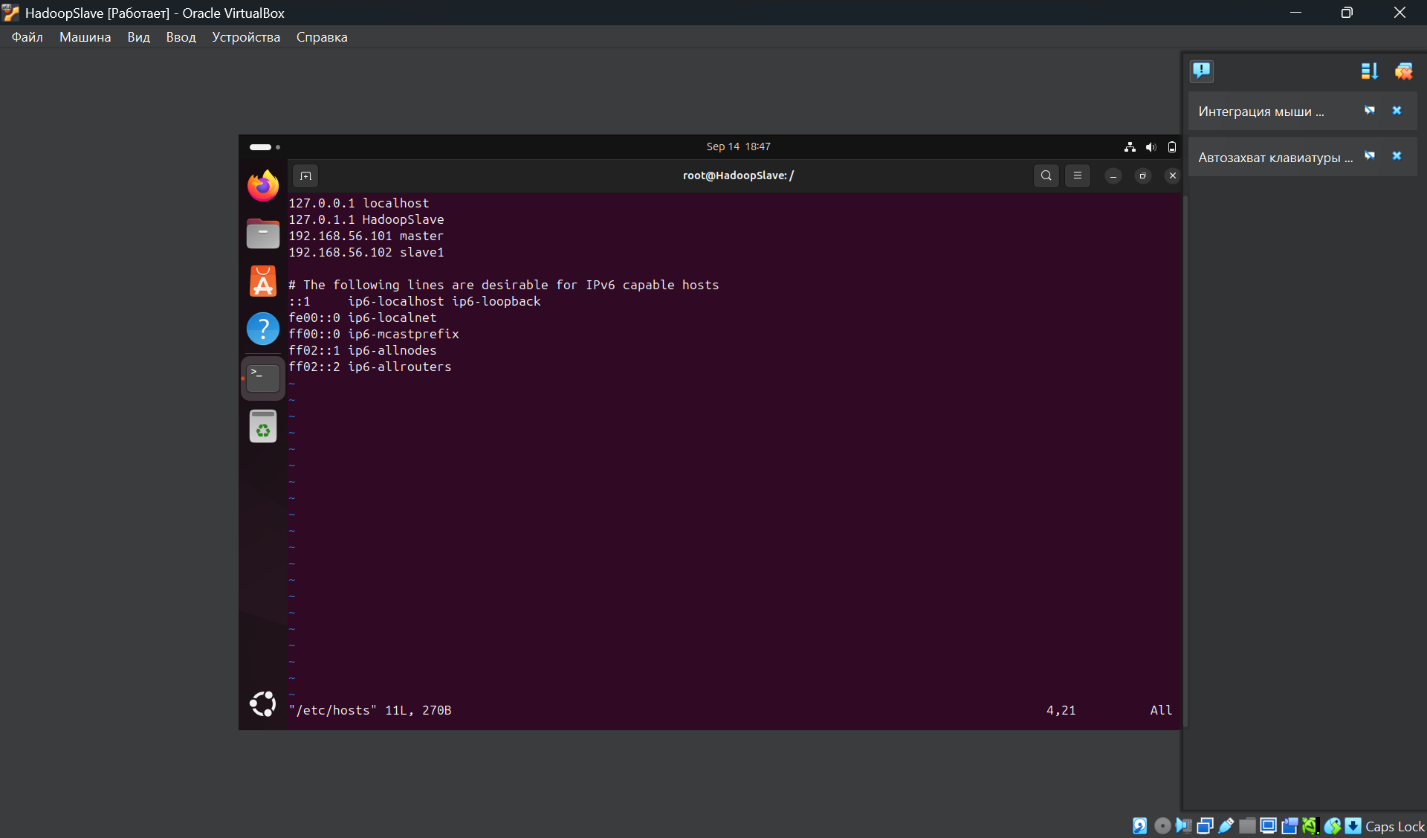


Рис.13 Настройка резолвера для master и slave

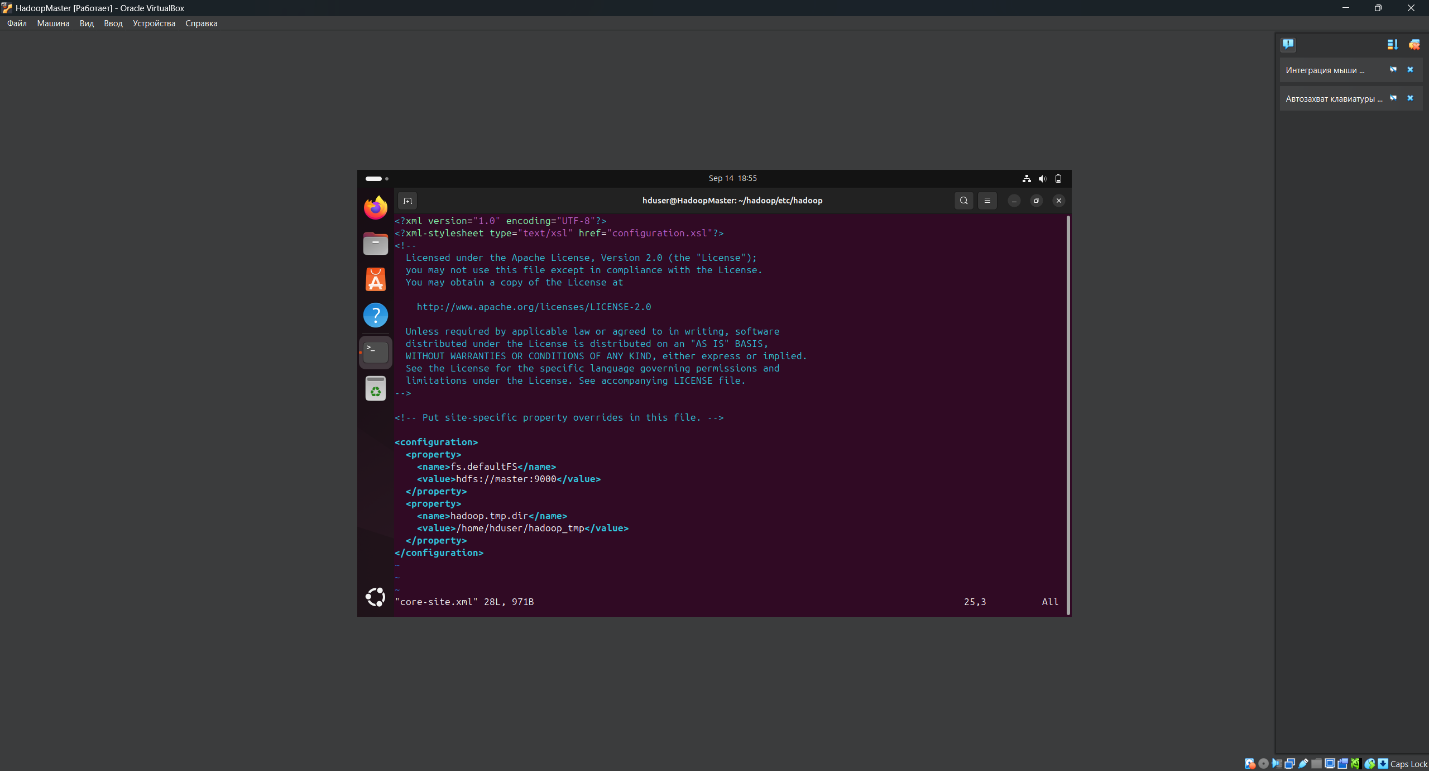


Рис.14 Настройка core-site.xml для master и slave

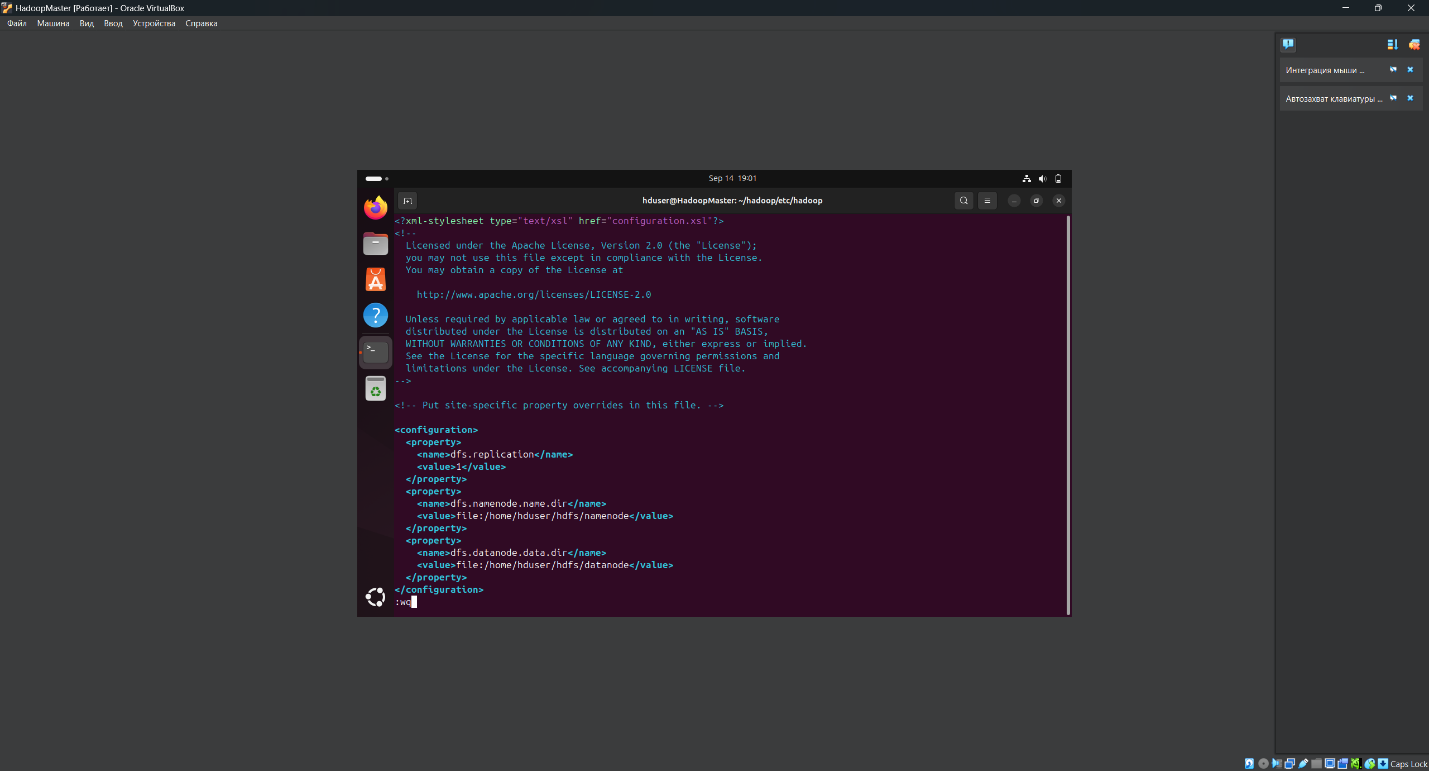


Рис.15 Настройка hdfs-site.xml для master и slave

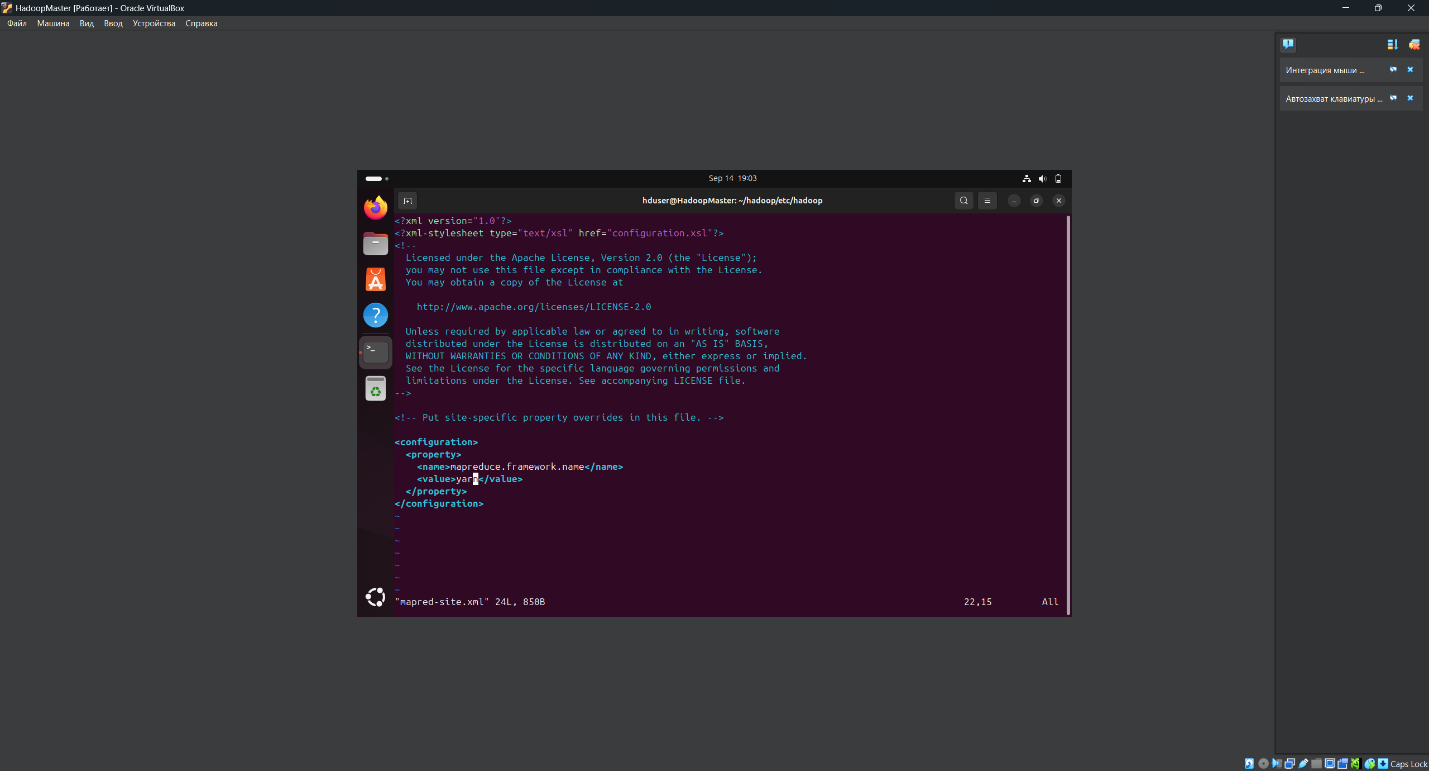


Рис.16 Настройка mapred-site.xml для master и slave

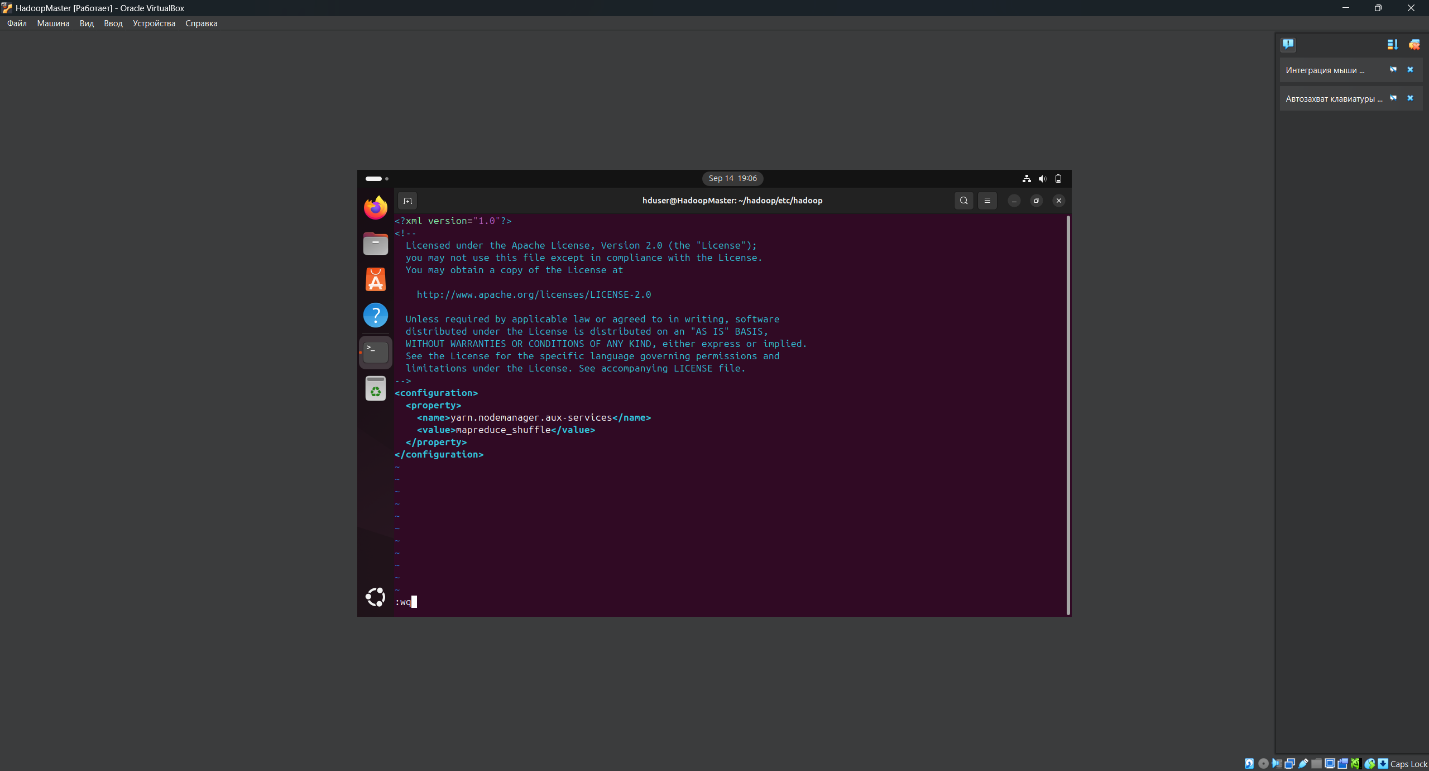


Рис.17 Настройка yarn-site.xml для master и slave

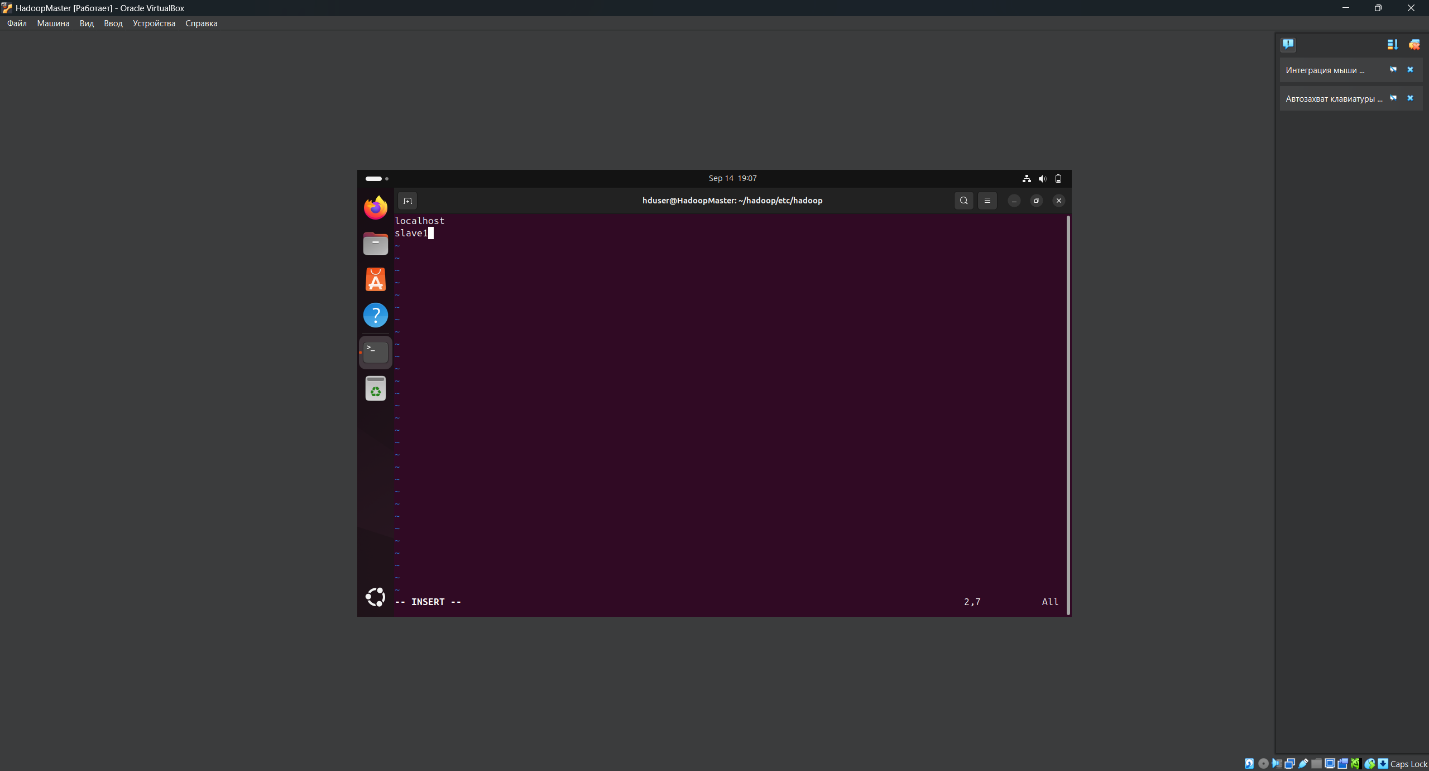


Рис.18 Настройка workers для master и slave



Рис.19 Копирование конфигурации hadoop на slave

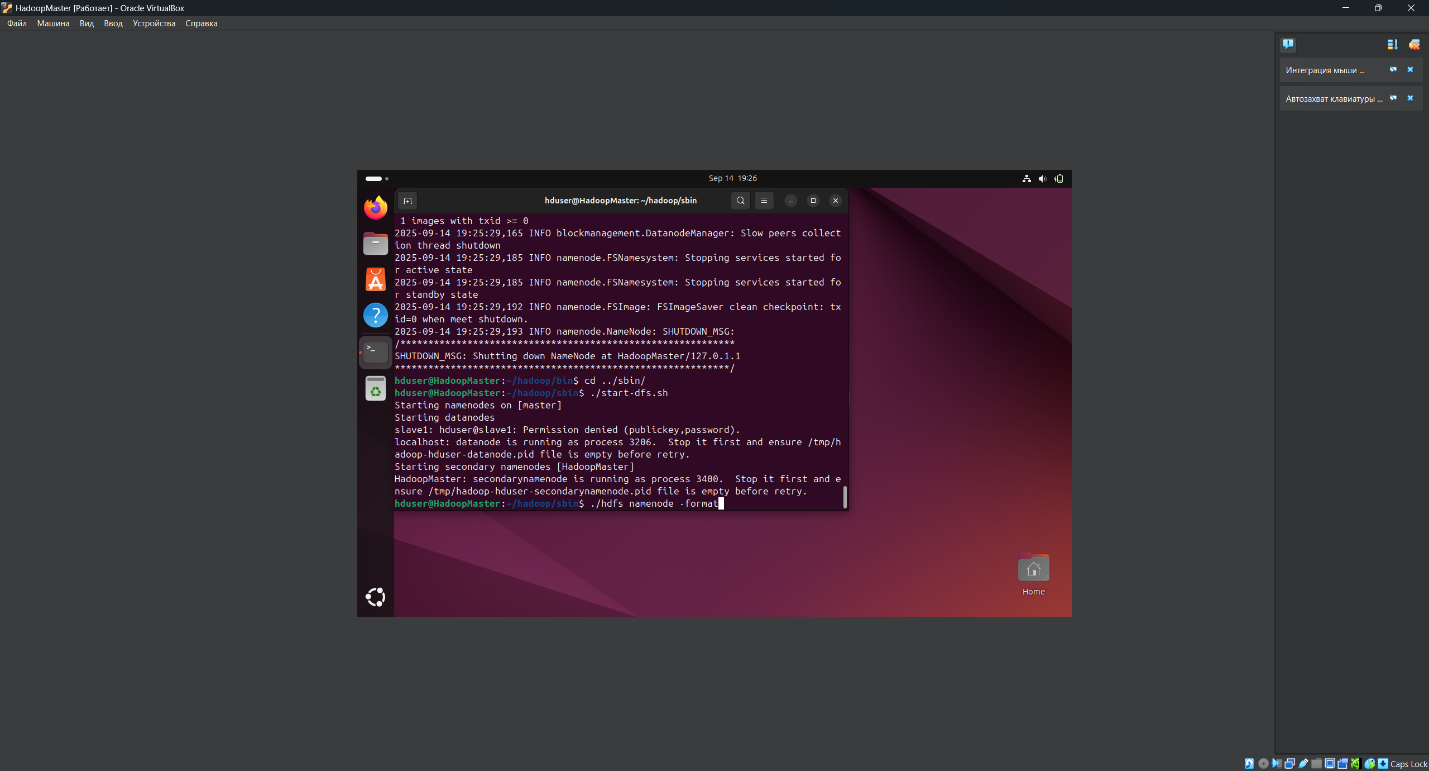


Рис.20 Форматирование файловой системы

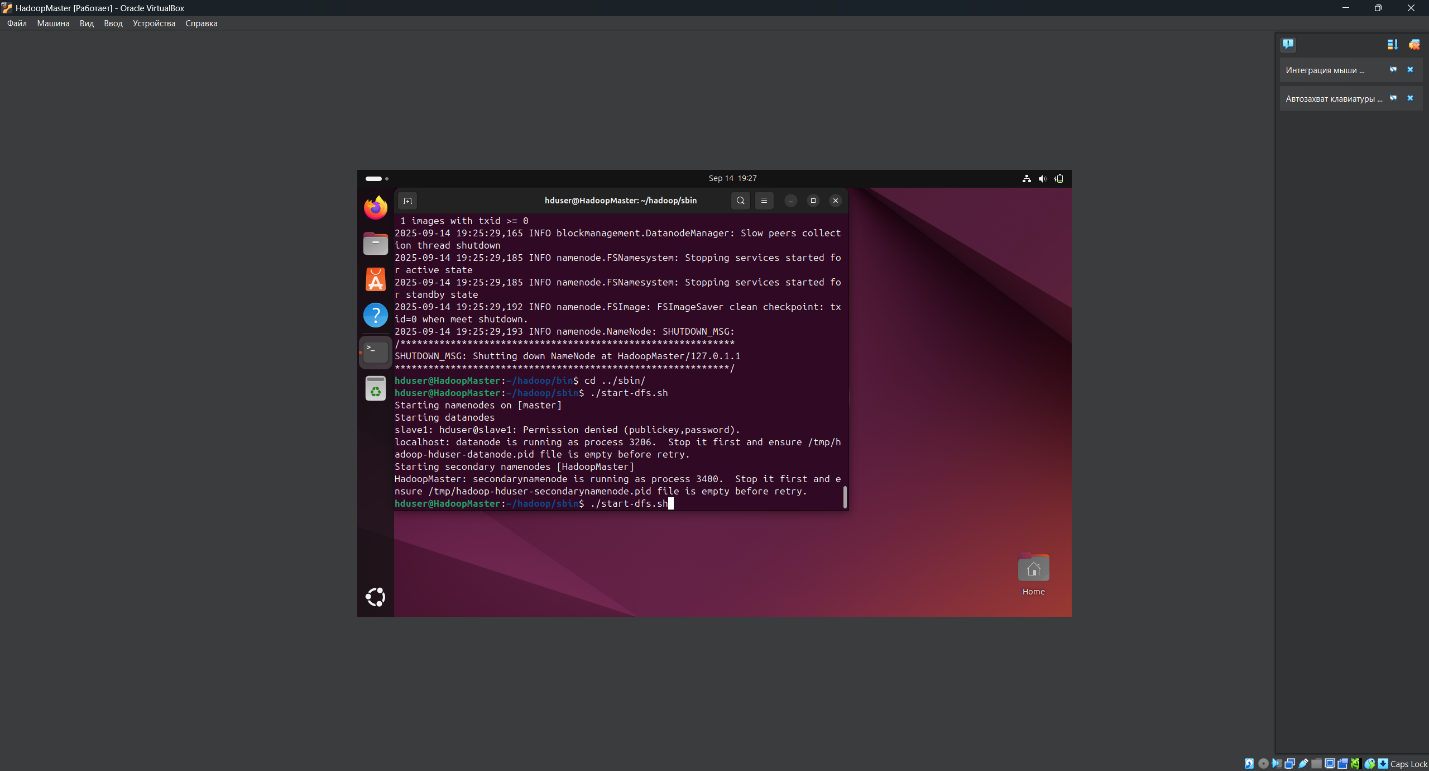


Рис.21 Запуск HDFS



Рис.22 Все узлы запущены

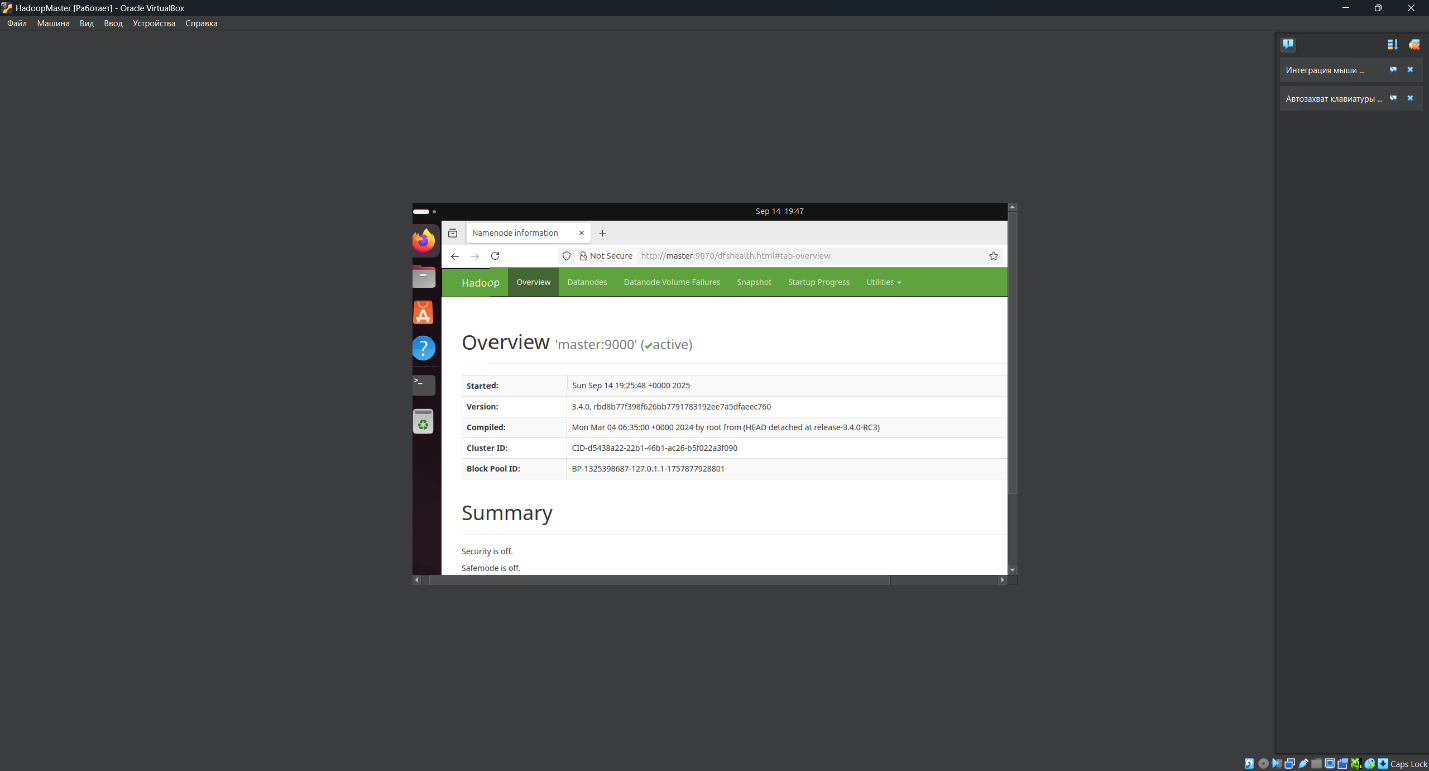


Рис.23 Web-интерфейс

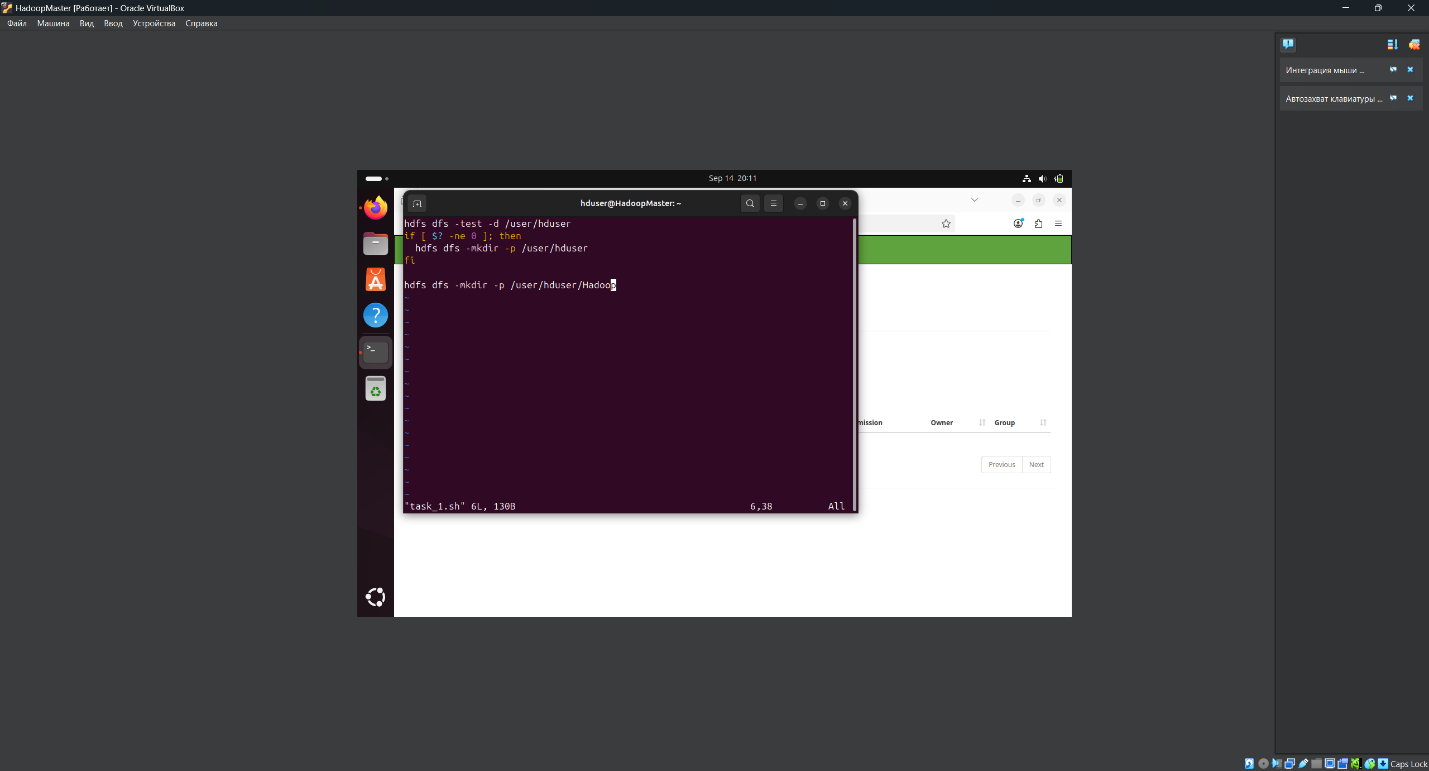


Рис.24 Задание 1

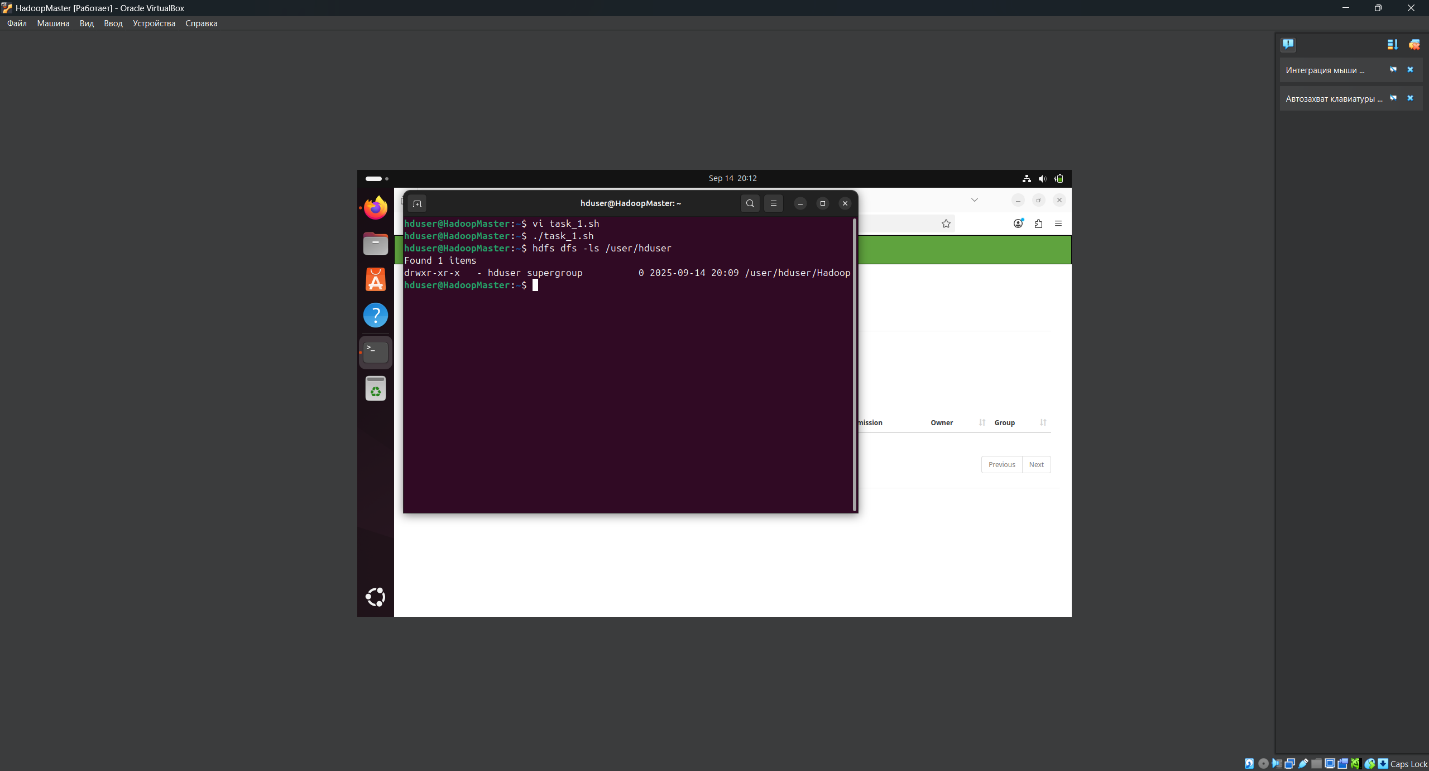


Рис.25 Результат задания 1

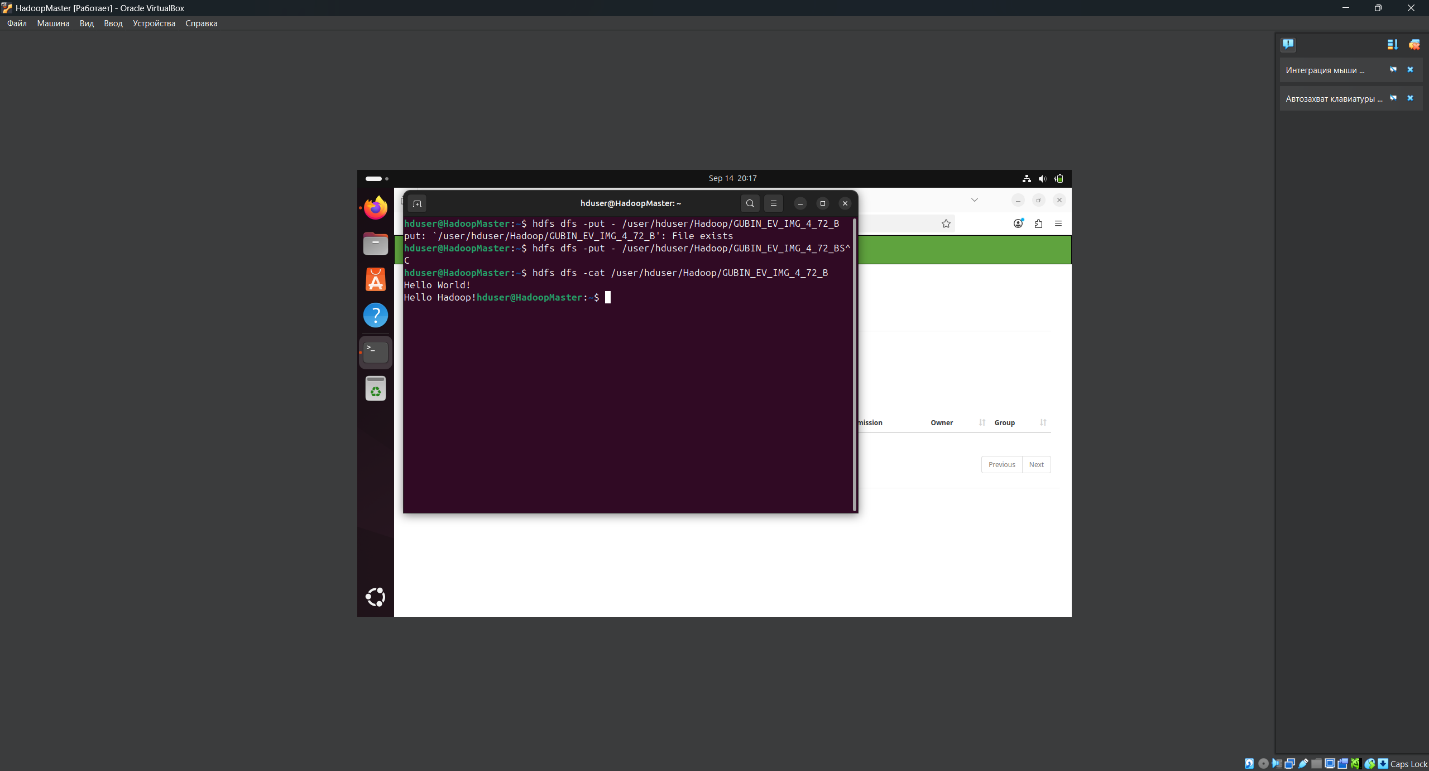


Рис.26 Задание 2

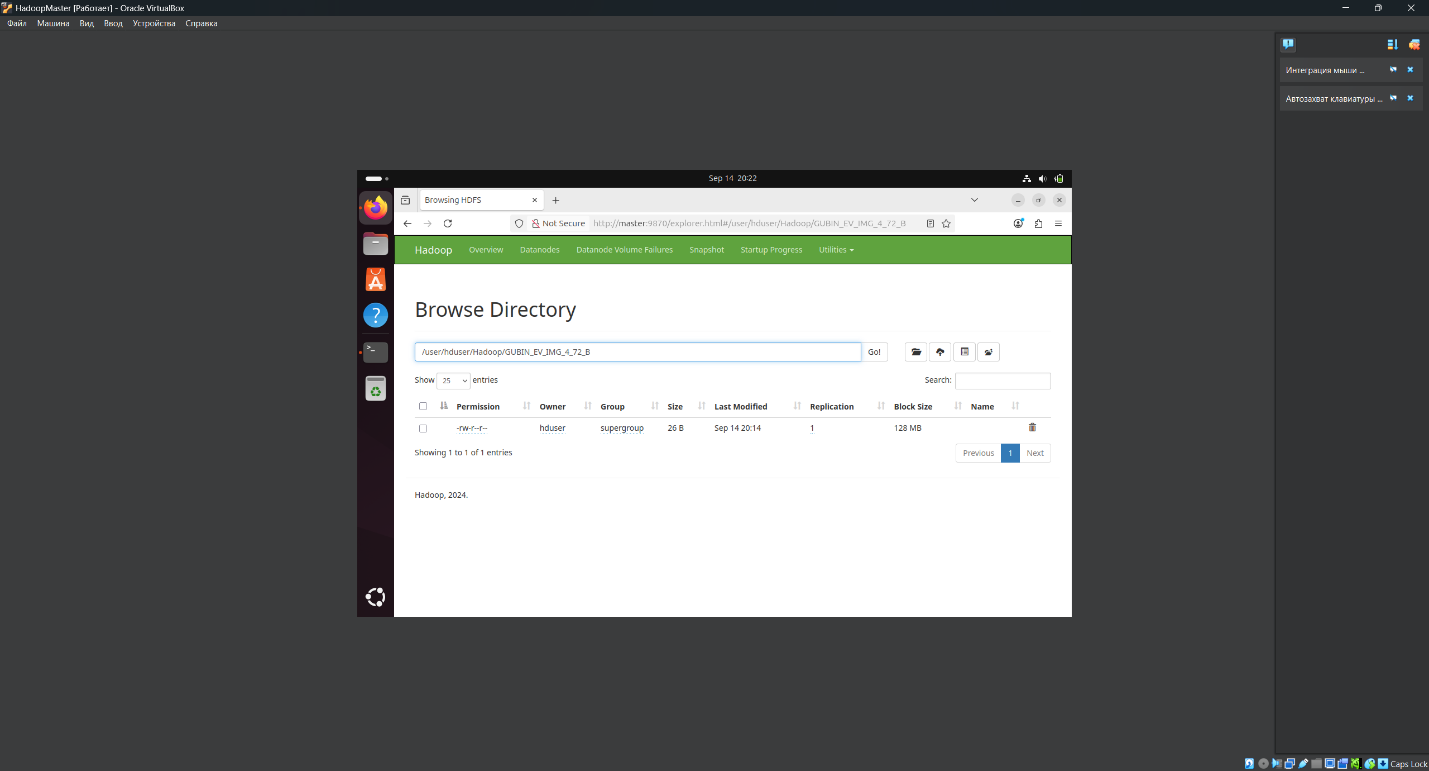


Рис.27 Задание 3

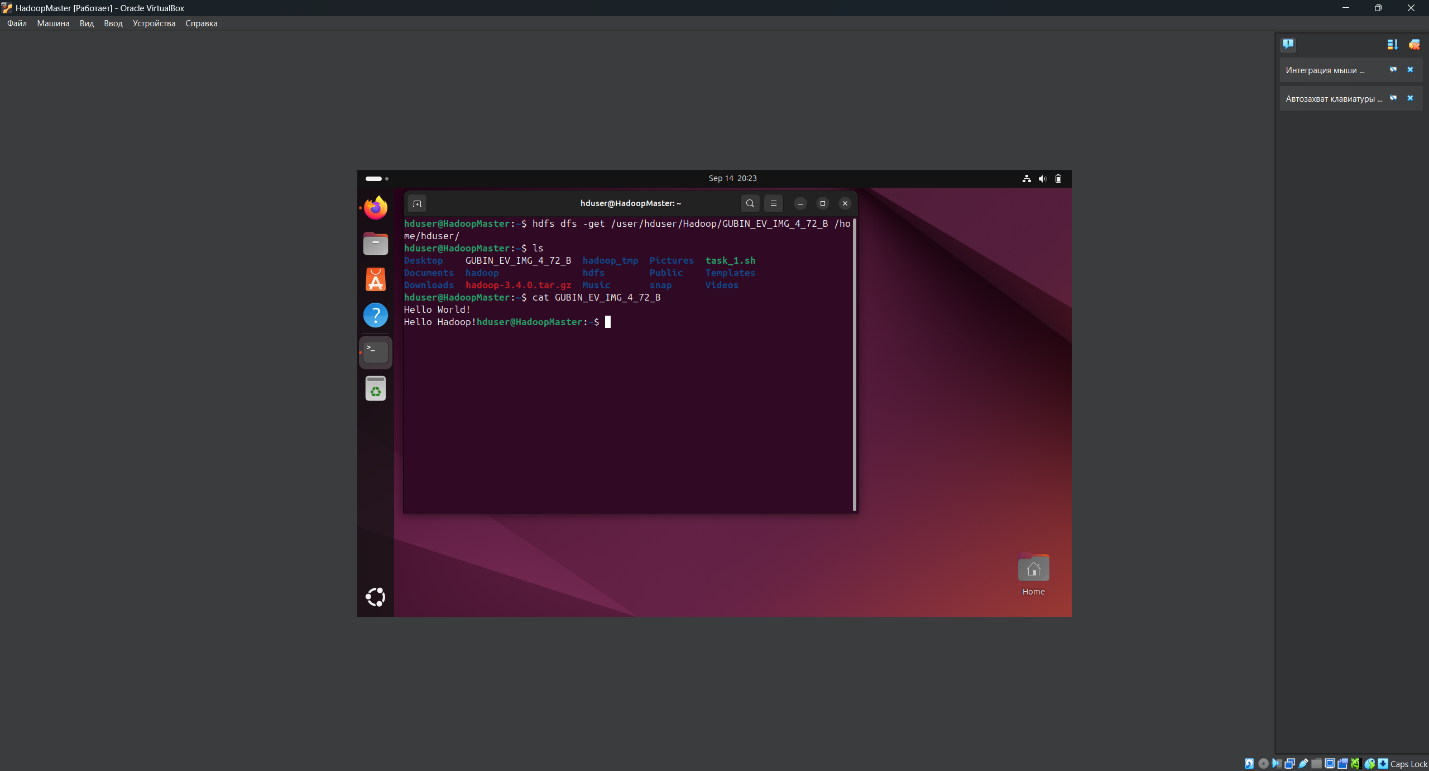


Рис.28 Задание 4

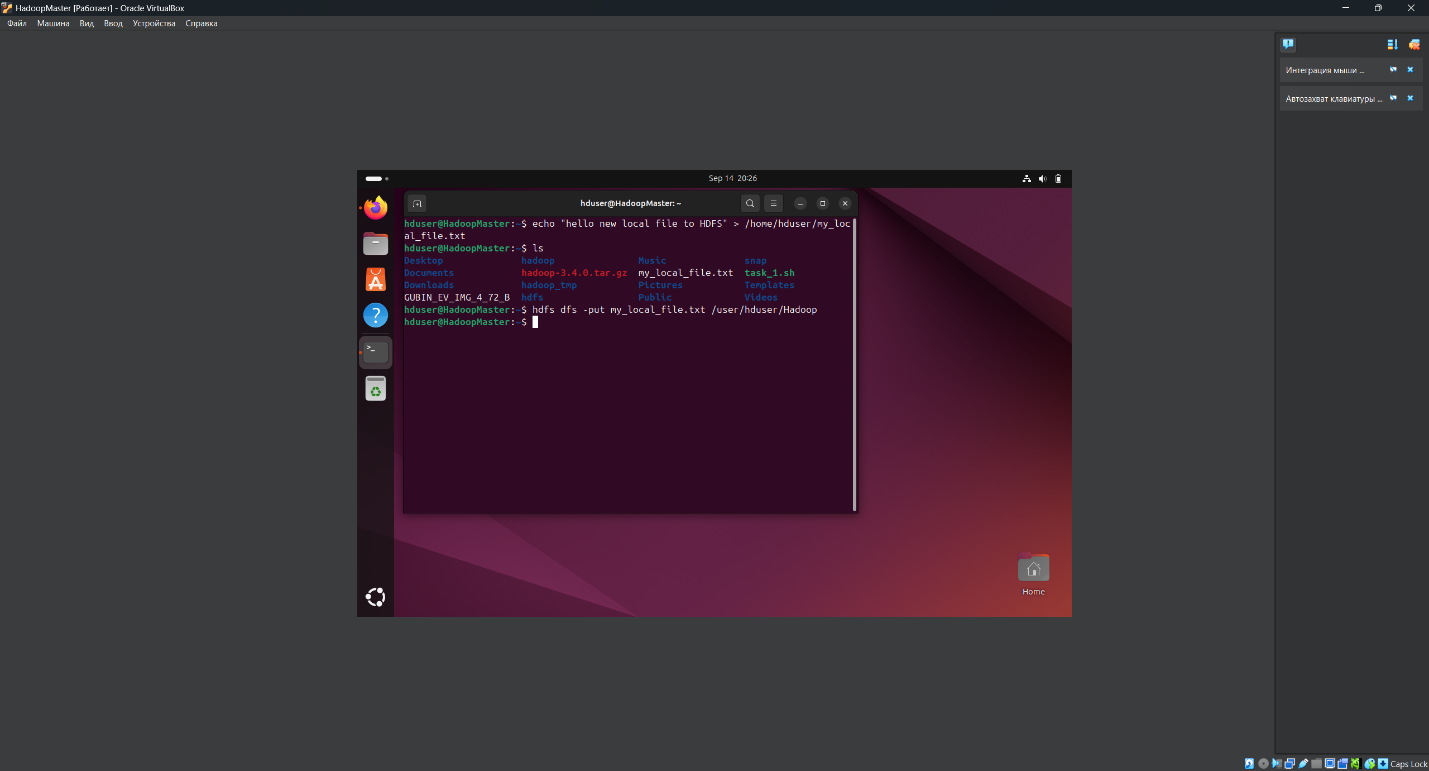


Рис.29 Задание 5

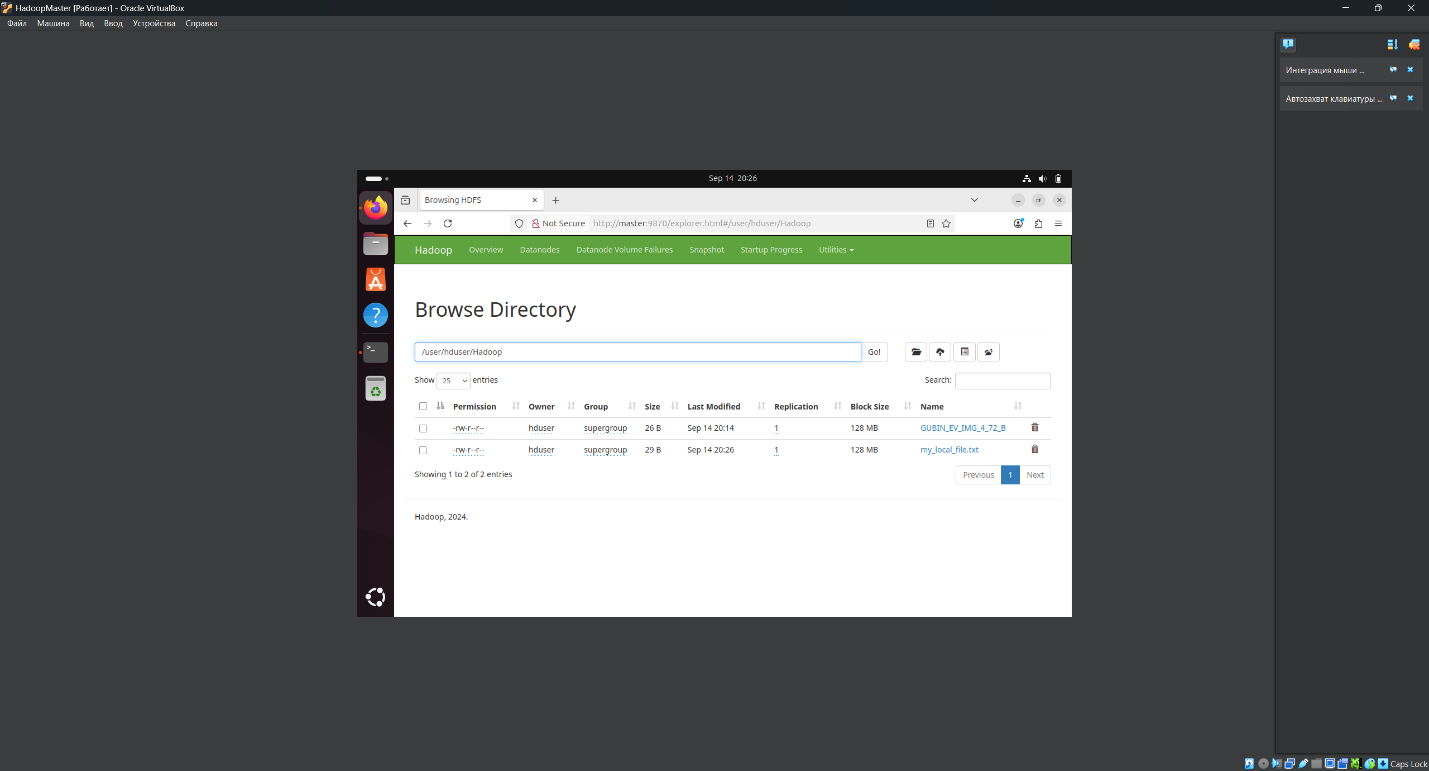


Рис.30 Результат задания 5

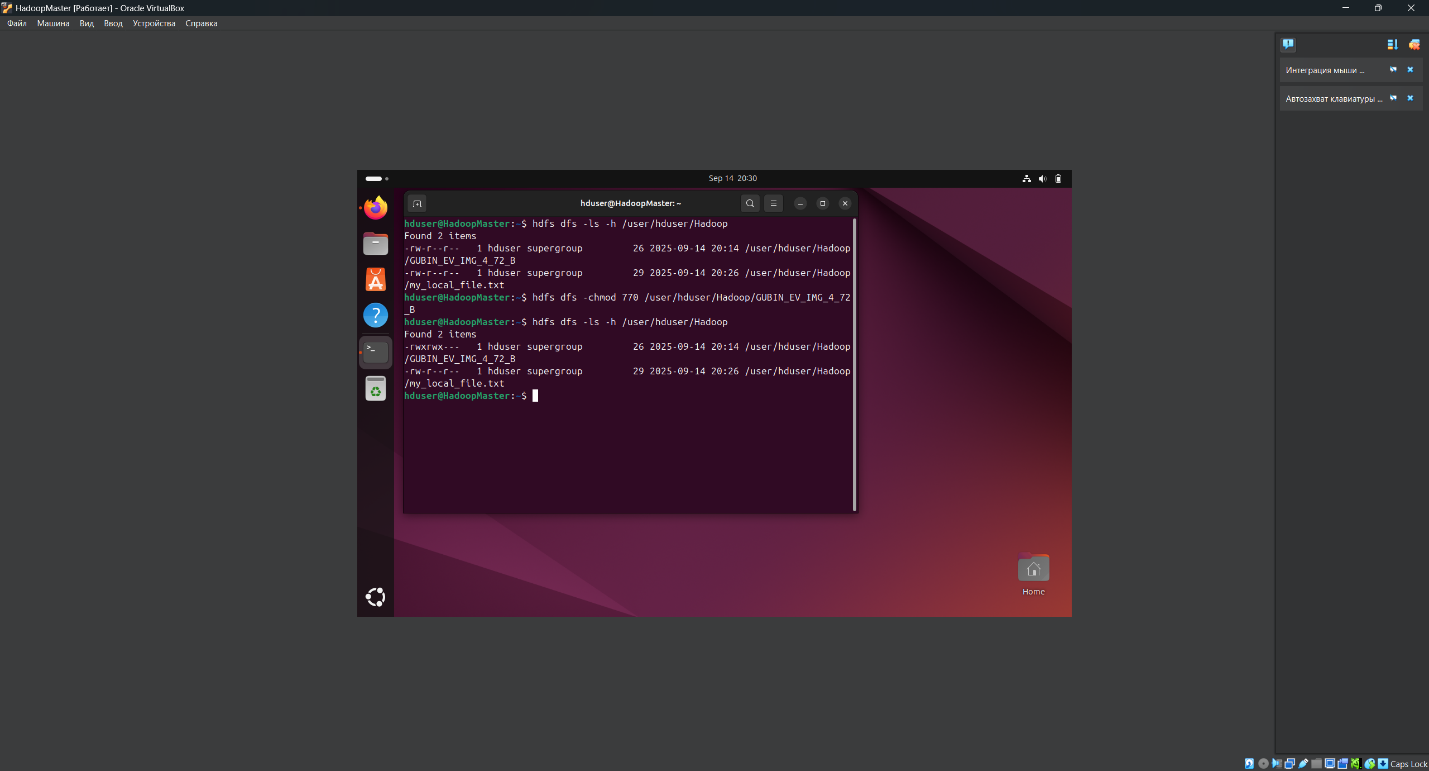


Рис.31 Задание 6

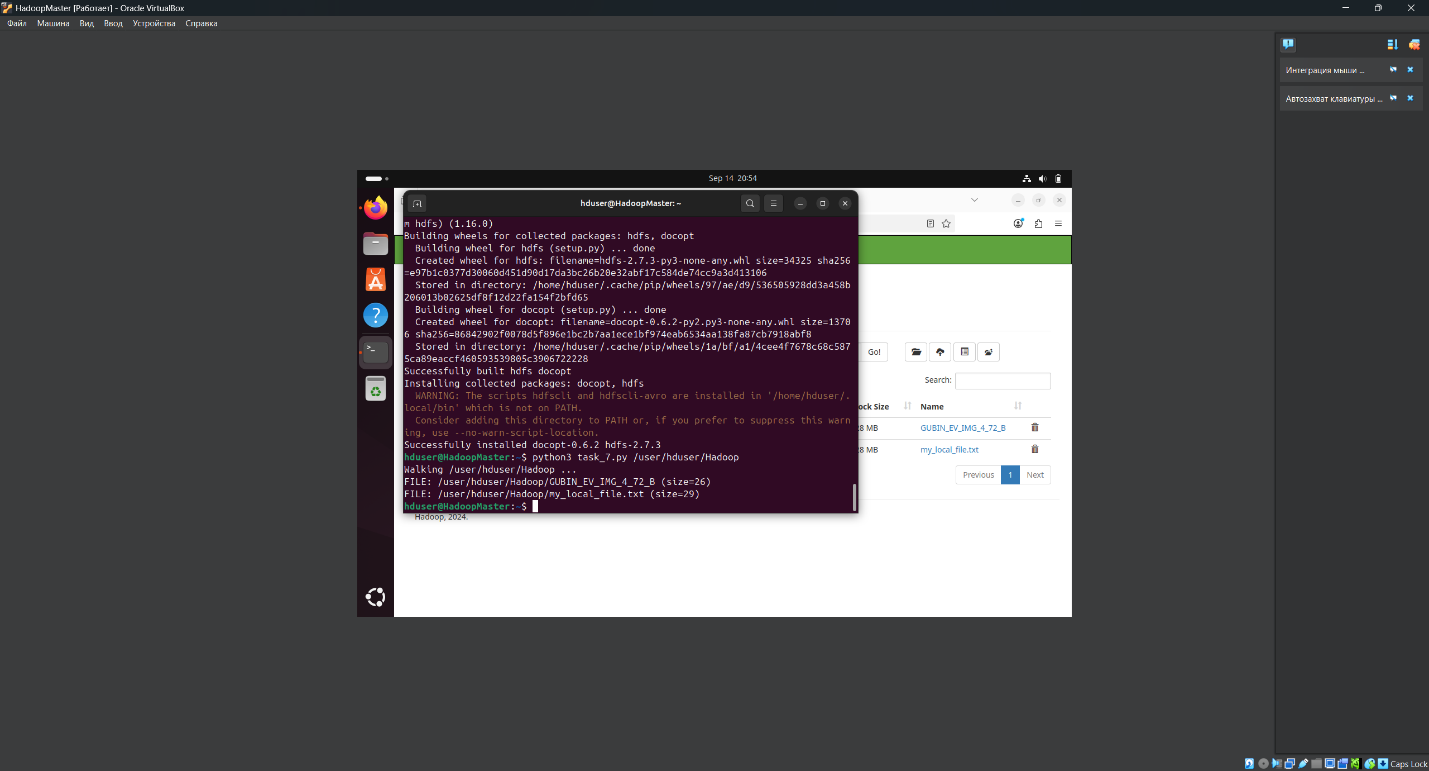


Рис.32 Результат выполнения задания 7

**Листинг программы:**

#!/usr/bin/env python3

import sys

import argparse

from hdfs import InsecureClient

*def* walk(*client*, *path*):

    try:

        entries = client.list(path, *status*=True)

    except *Exception* as e:

        print(*f*"ERROR: cannot access {path}: {e}")

        return

    for name, status in entries:

        full = path.rstrip('/') + '/' + name

        typ = status.get('type', '')  # 'FILE' or 'DIRECTORY'

        if typ == 'DIRECTORY':

            print(*f*"DIR : {full}")

            walk(client, full)

        else:

            size = status.get('length', '?')

            print(*f*"FILE: {full} (size={size})")

*def* main():

    parser = argparse.ArgumentParser(*description*='Recursively list HDFS paths.')

    parser.add\_argument('--namenode', '-n', *default*='http://localhost:9870',

*help*='URL of NameNode WebHDFS endpoint, e.g. http://master:9870')

    parser.add\_argument('--user', '-u', *default*='hduser', *help*='HDFS user to connect as')

    parser.add\_argument('paths', *nargs*='+', *help*='HDFS path(s) to walk, e.g. /user/hduser/Hadoop')

    args = parser.parse\_args()

    client = InsecureClient(args.namenode, *user*=args.user)

    for p in args.paths:

        print(*f*"Walking {p} ...")

        walk(client, p)

if \_\_name\_\_ == '\_\_main\_\_':

    main()

**Вывод:** в ходе лабораторной работы были получены практические навыки по настройке Hadoop в ОС Ubuntu 20.04, изучены основные команды фреймворка.