|  |
| --- |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ |
| **«Национальный исследовательский ядерный университет «МИФИ»**  **(НИЯУ МИФИ)** |
| ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ |
| КАФЕДРА «КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ» (№12) |

Лабораторная работа №1

«Hadoop»

по дисциплине «Наука о данных и

анализ больших данных»

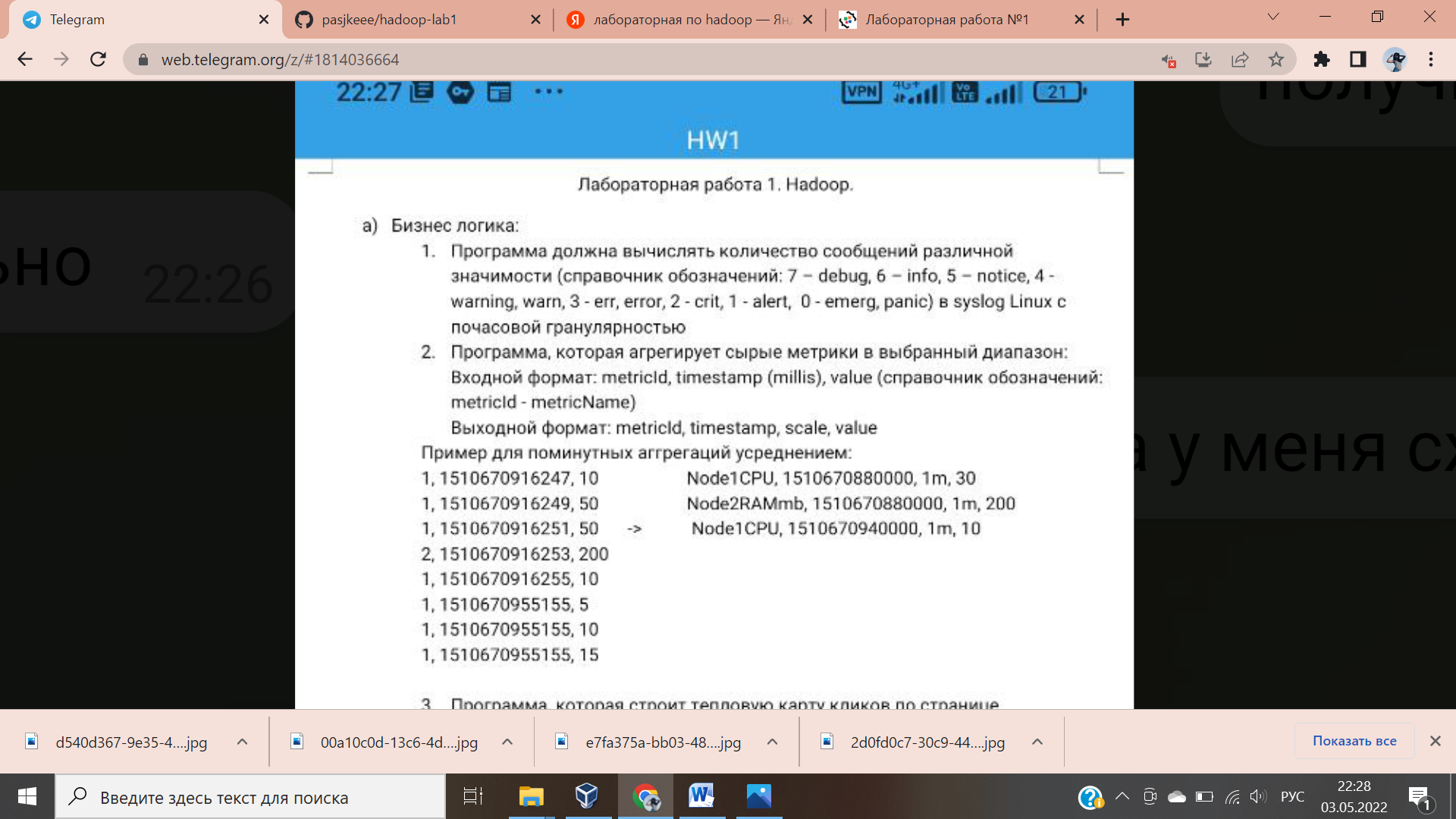
Работу выполнила студентка группы М21-502:

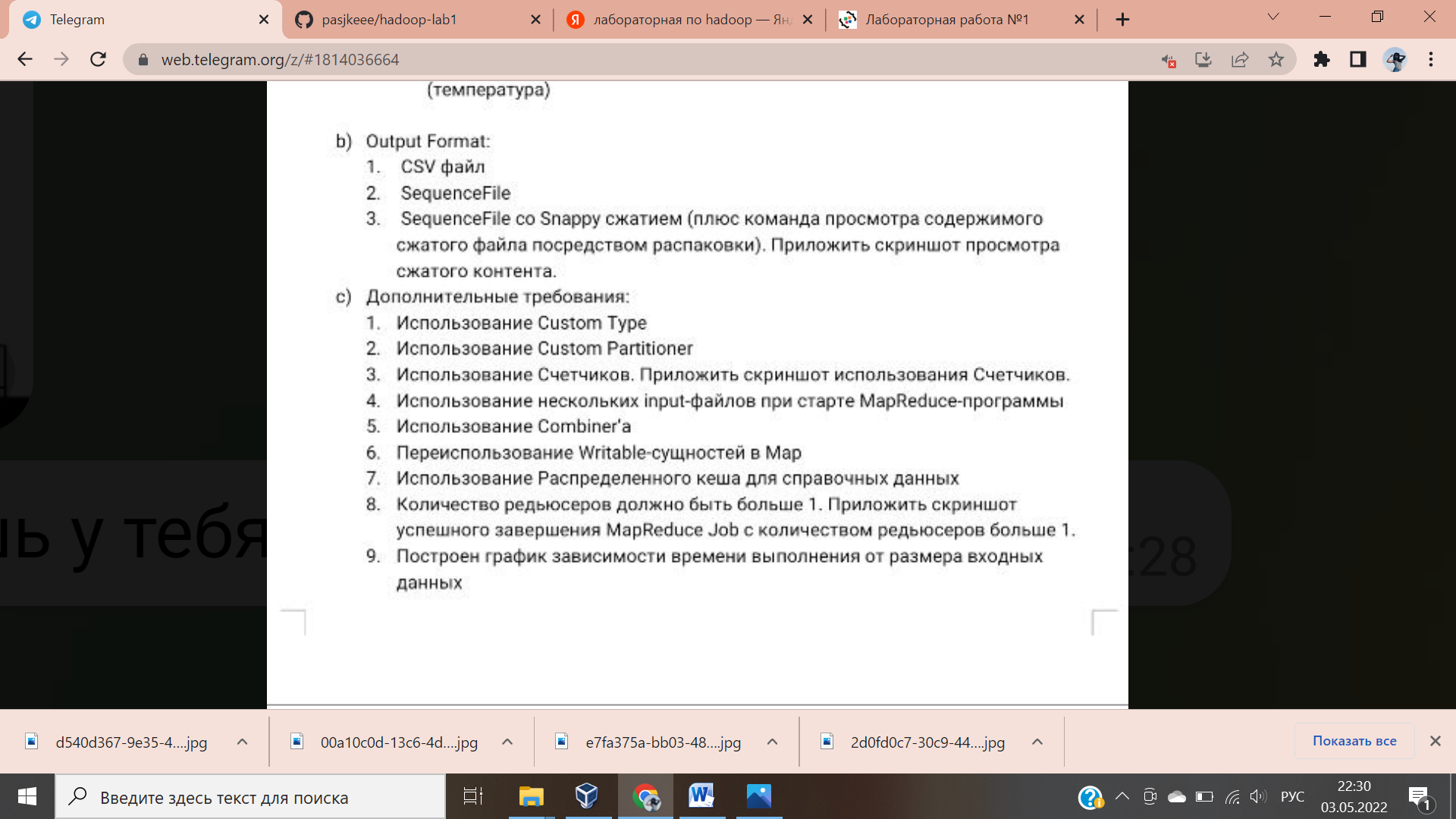
Нургазиева Д.Р.

Проверил: Синельников Д.М.

2022

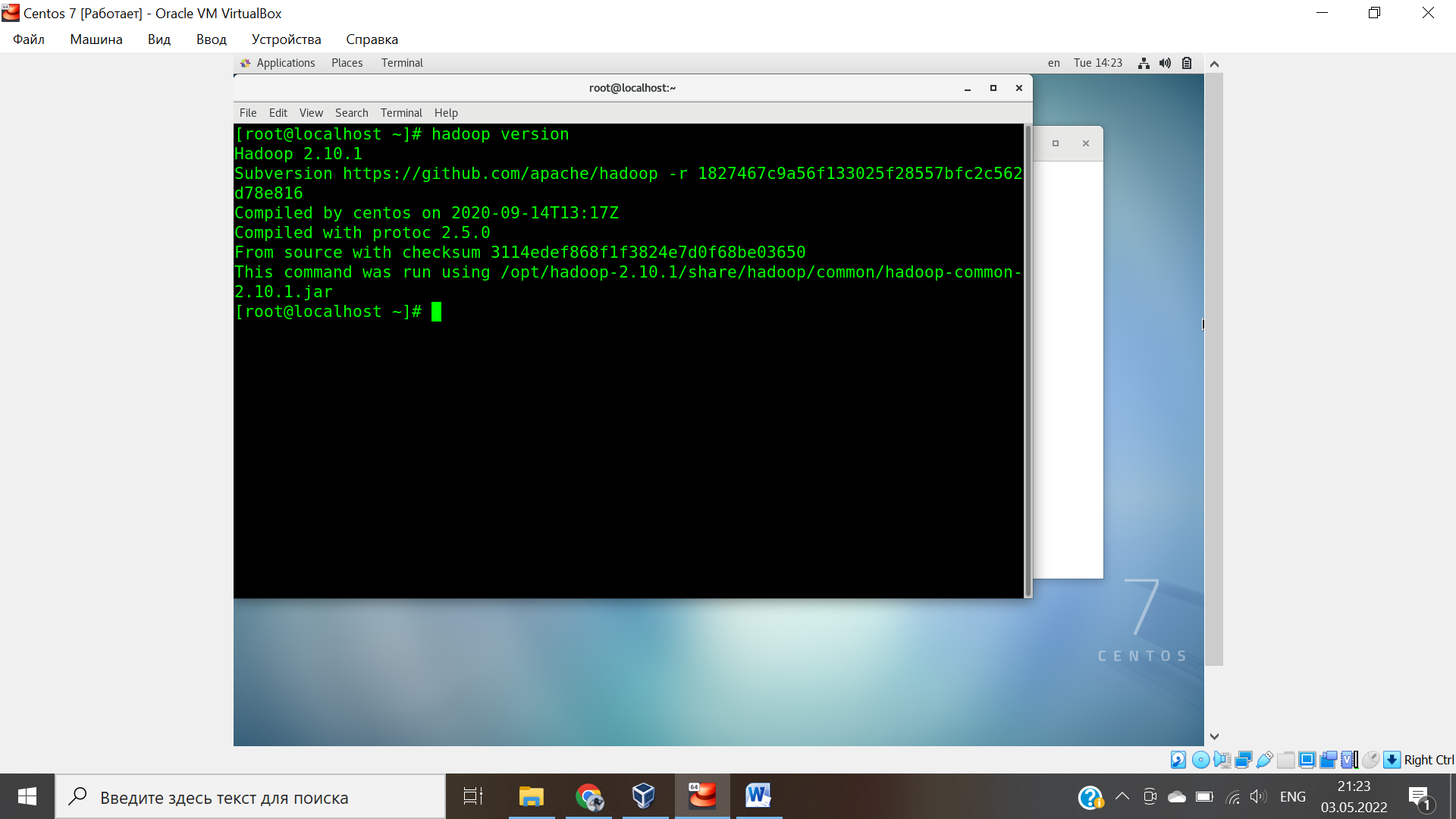
**Цель:** изучить основы Hadoop. Выполнить задание по варианту {a(1);b(3);c(1)}:

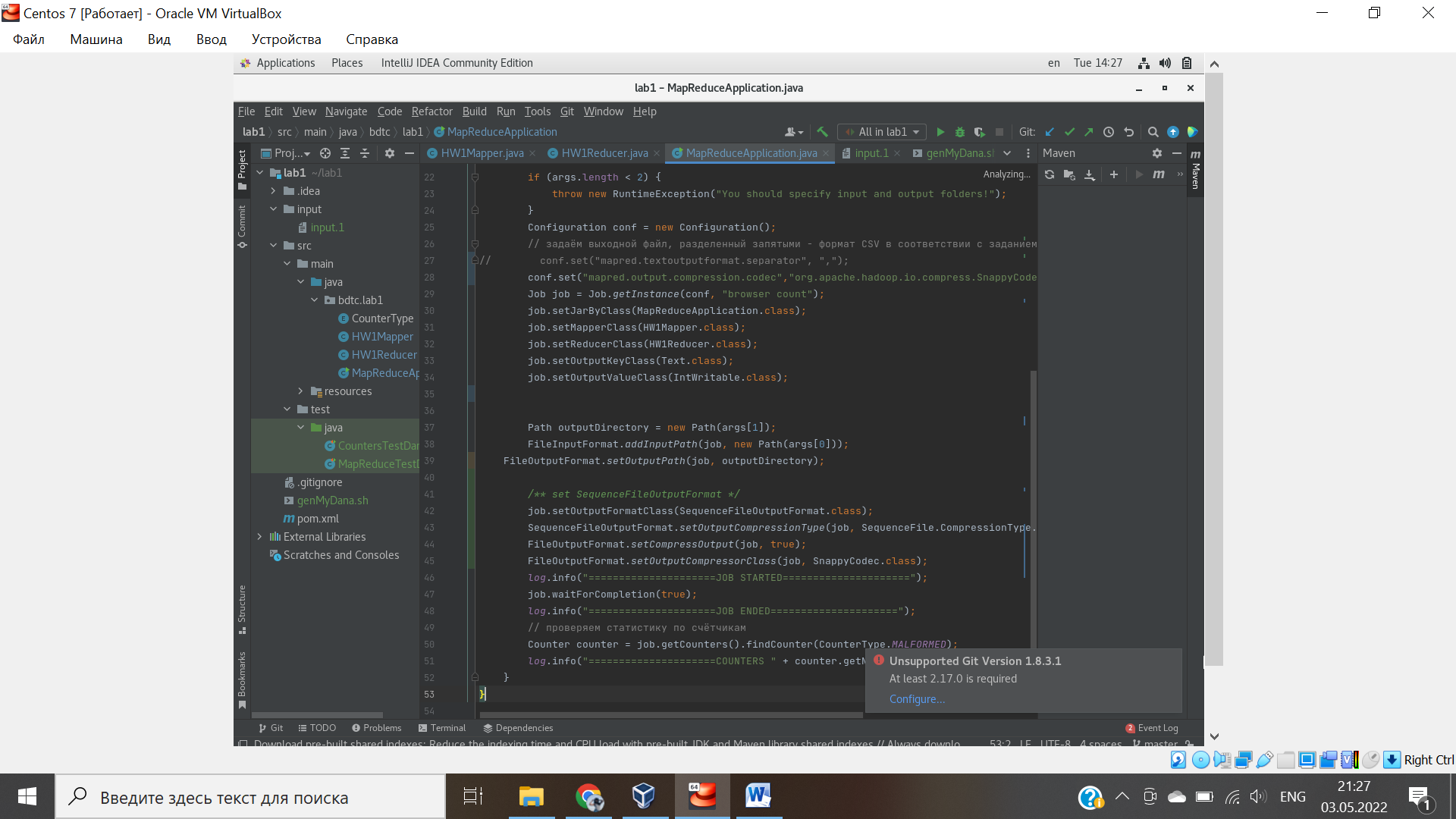




# **Ход работы:**

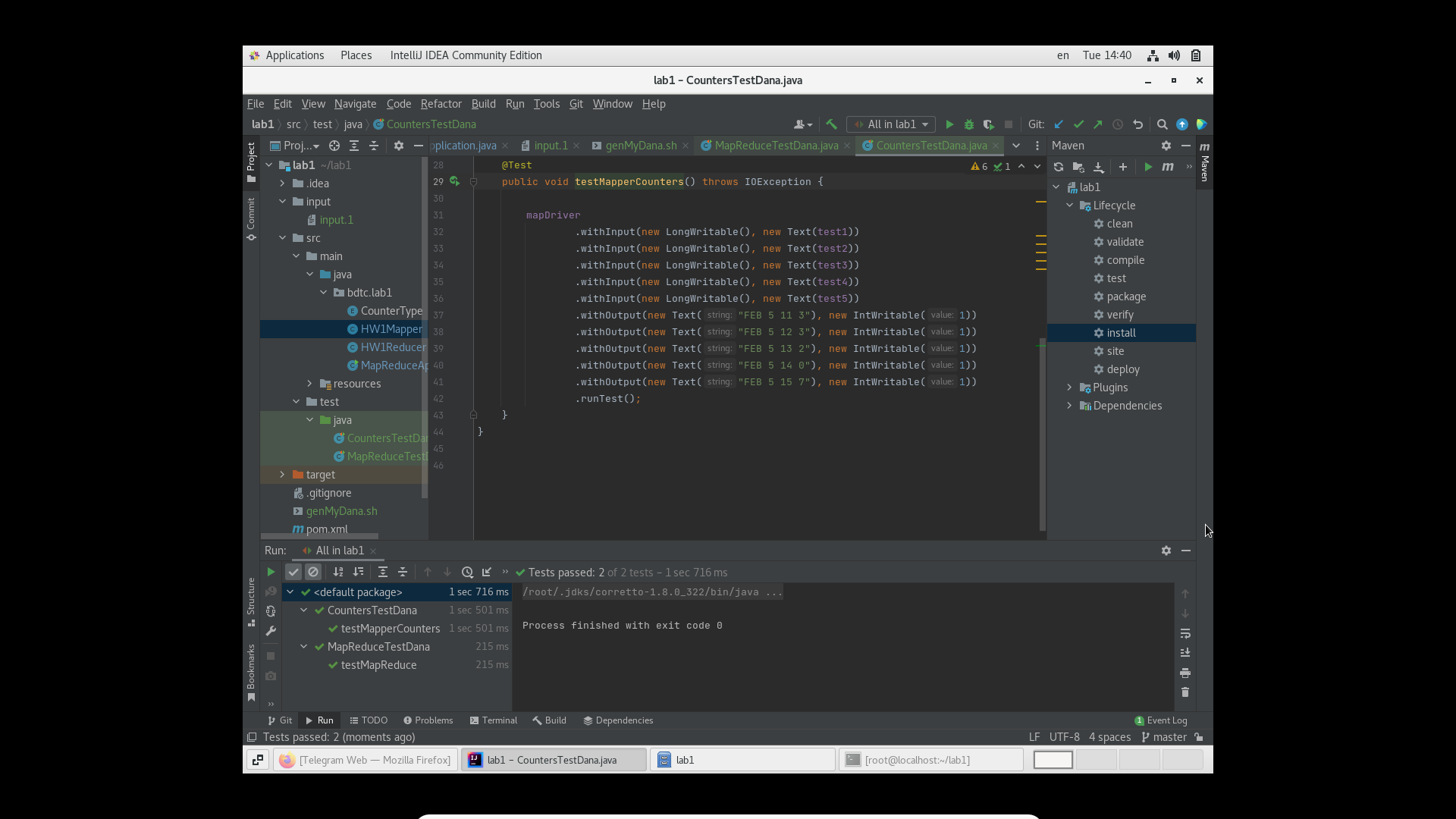
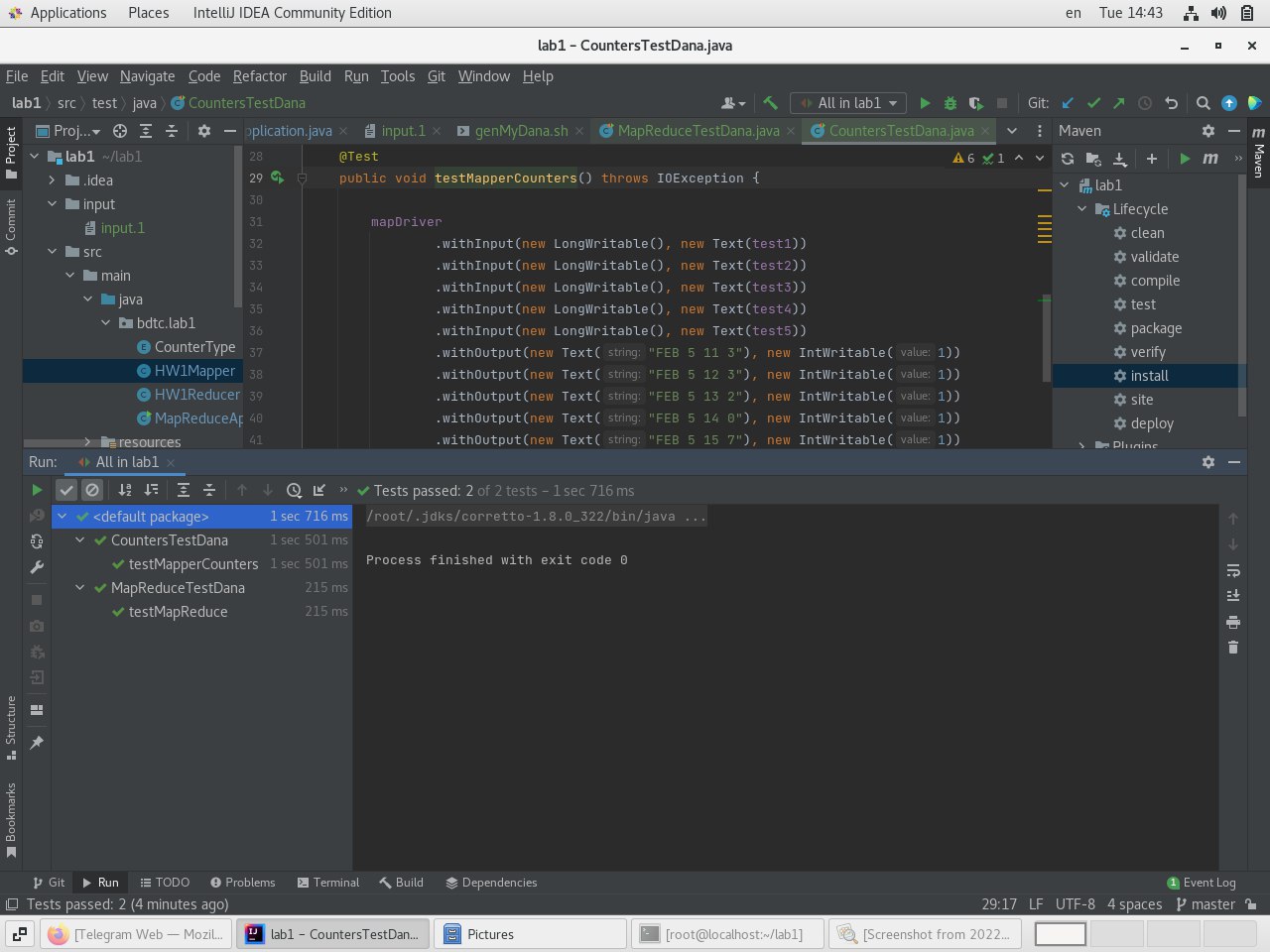
1. Разработка велась с использованием Hadoop, OS Centos. Для сборки используется maven плагин для Intellij IDEA.





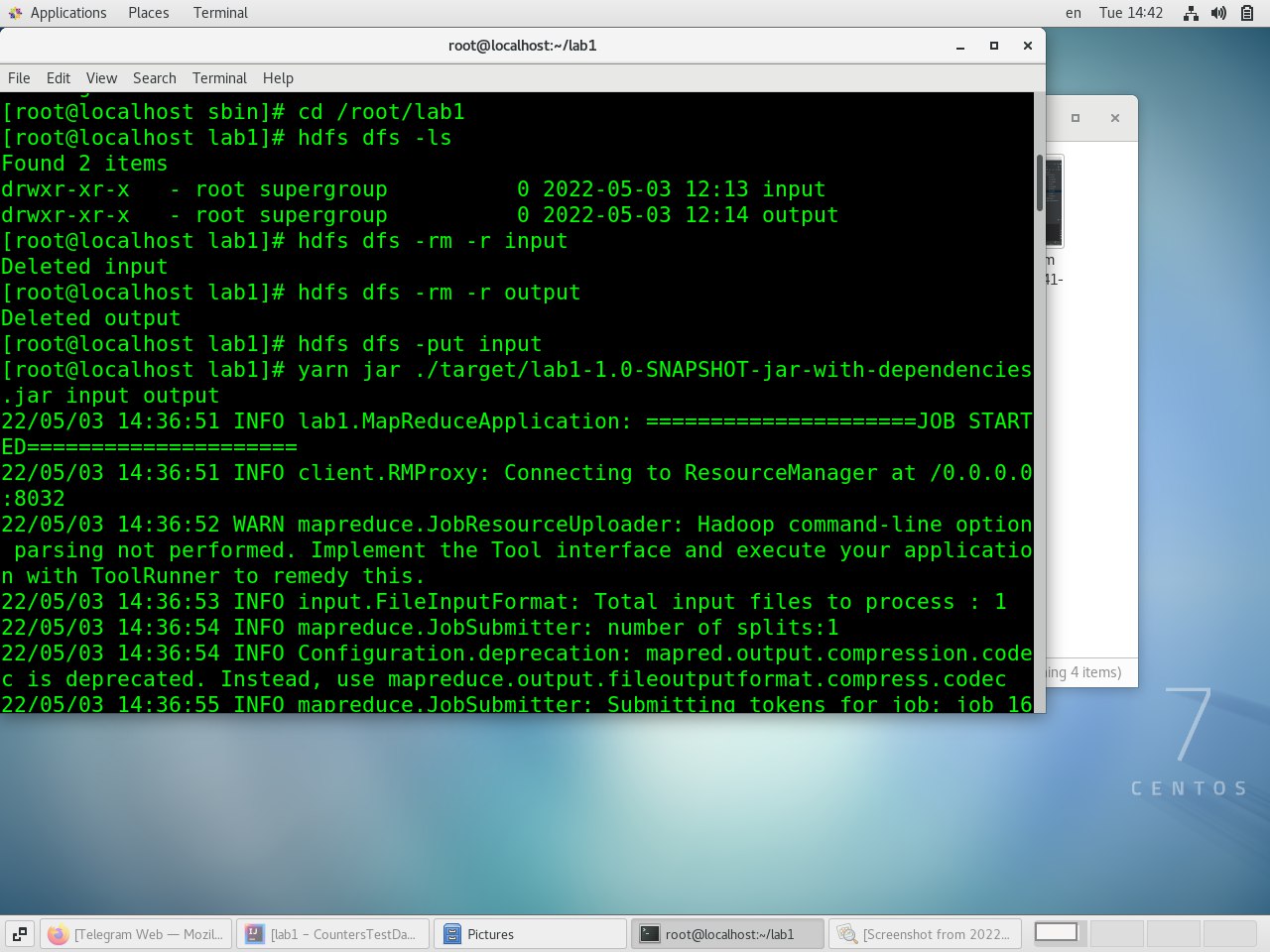
1. Базовое решение представляет следующие возможности:

* Генерация исходных данных скрипт-файл genMyDana.sh;
* Mapper - HW1Mapper;
* Reducer - HW1Reducer;
* MapReduceApplication;
* Тестовые файлы для тестирования Mapper и Reducer.

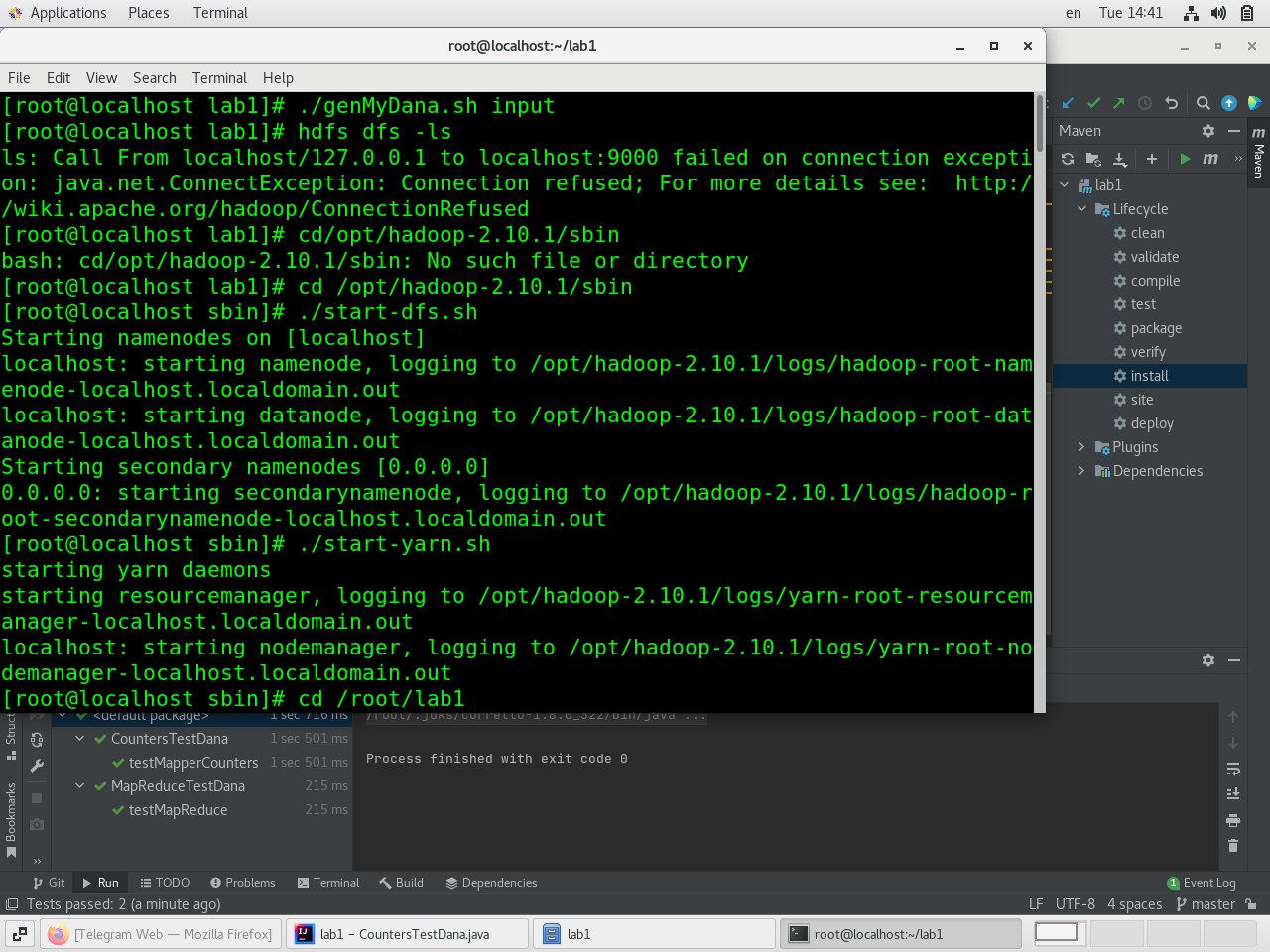
 

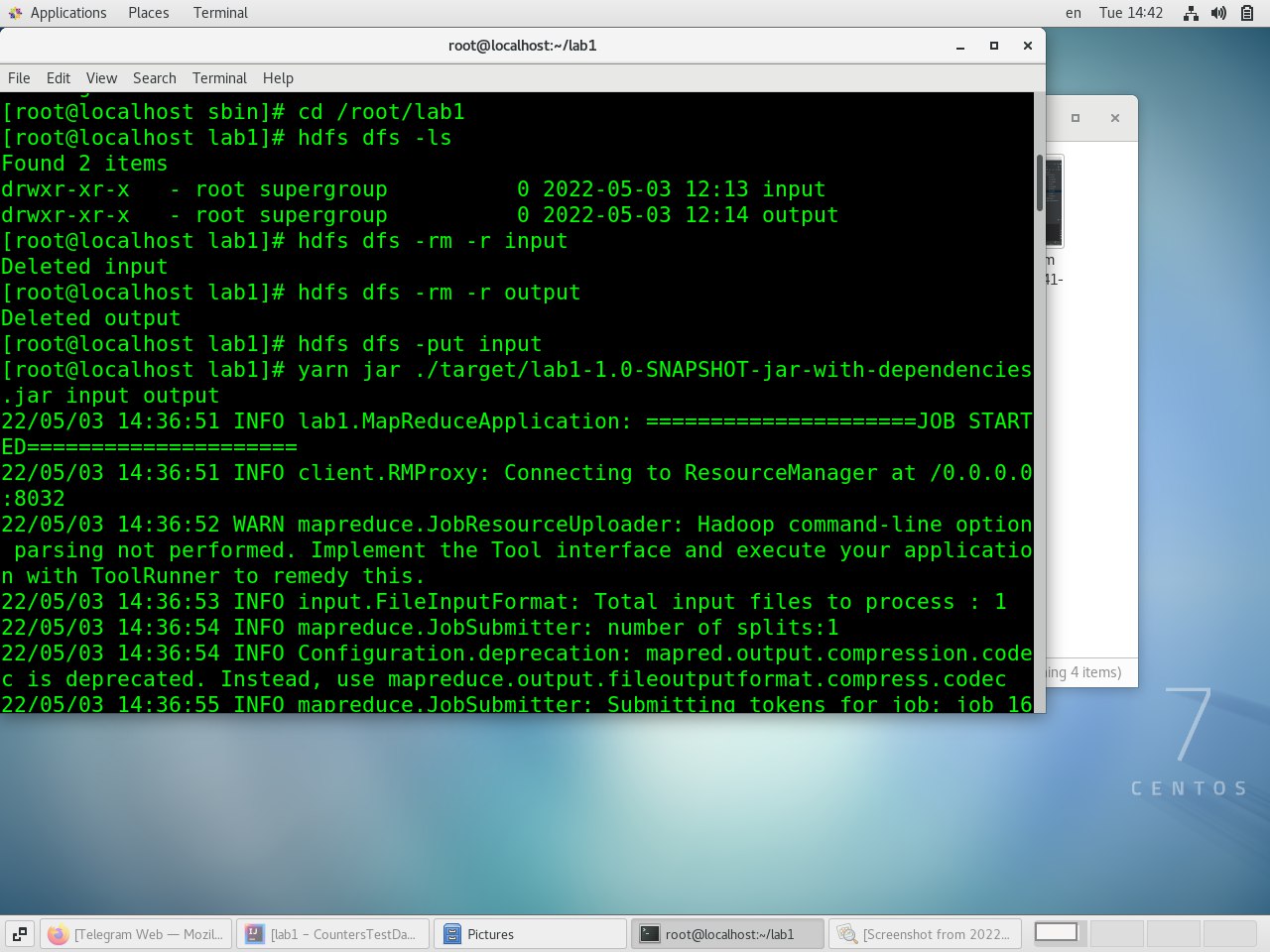
1. Базовое решение представляет следующие возможности:

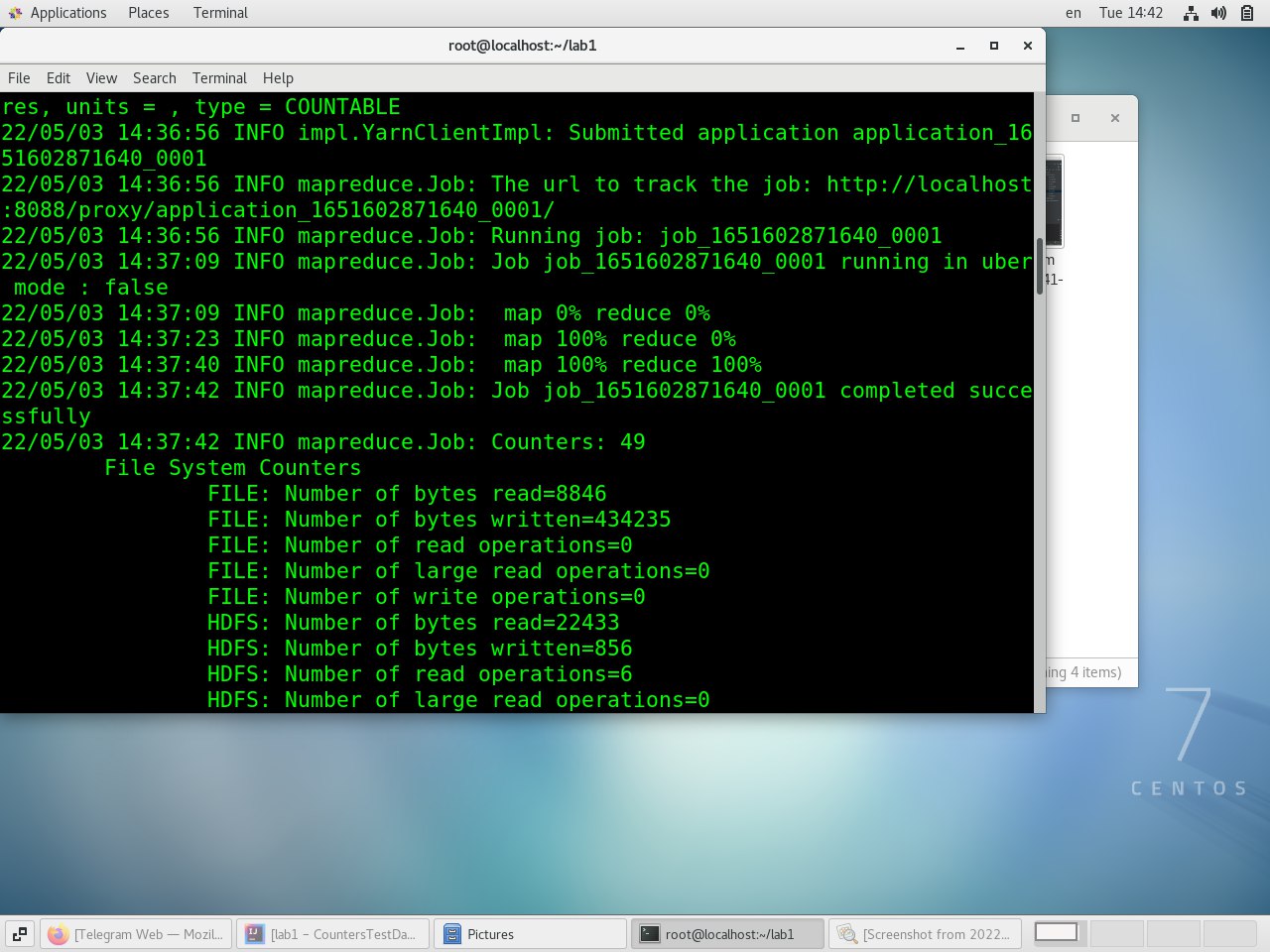
* Генерация данных с помощью скрипта ./genMyDana.sh
* Build программы
* Проверить dfs на наличие output и input директорий обнулить при необходимости

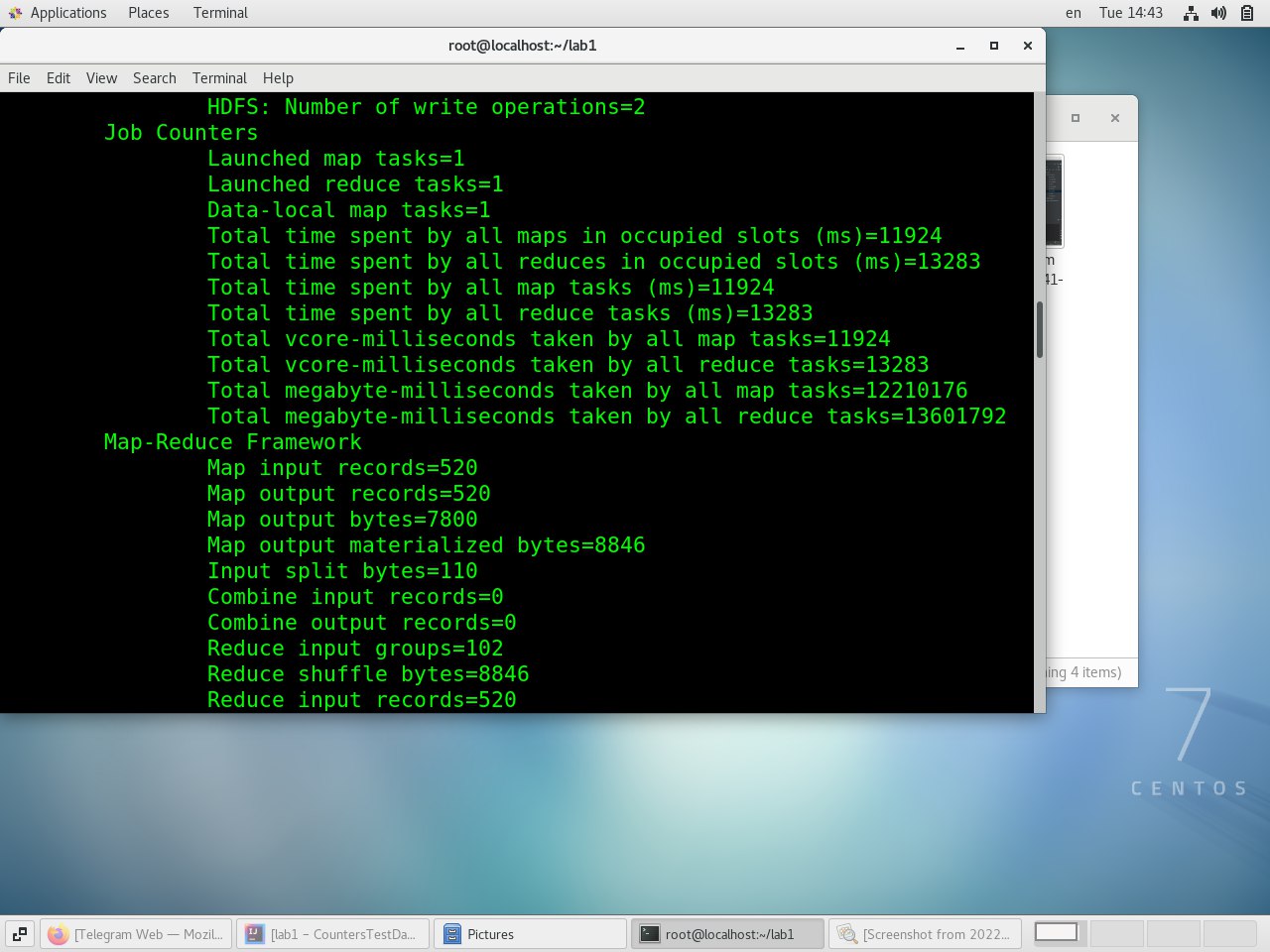


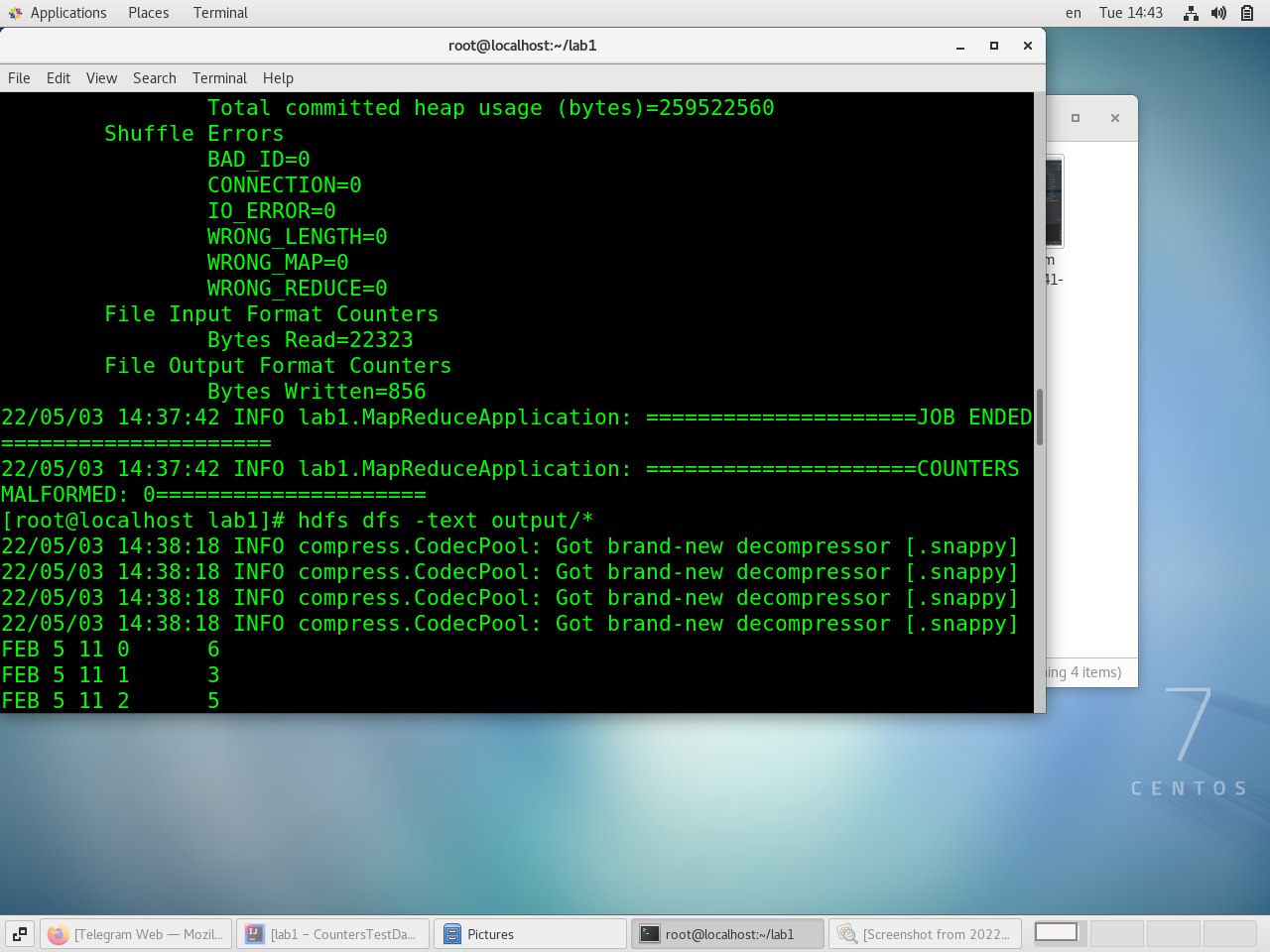
1. Процесс выполнения программы

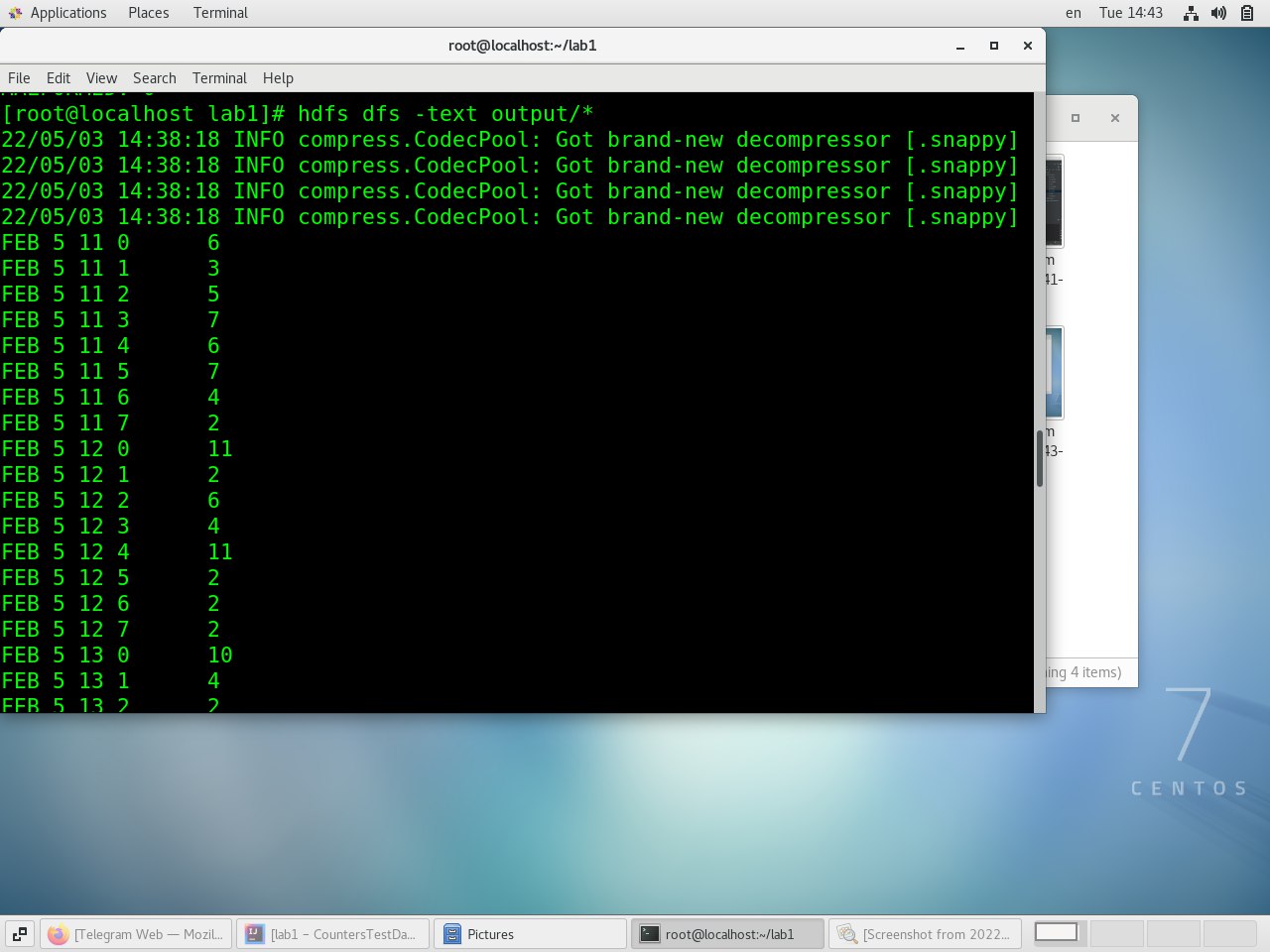




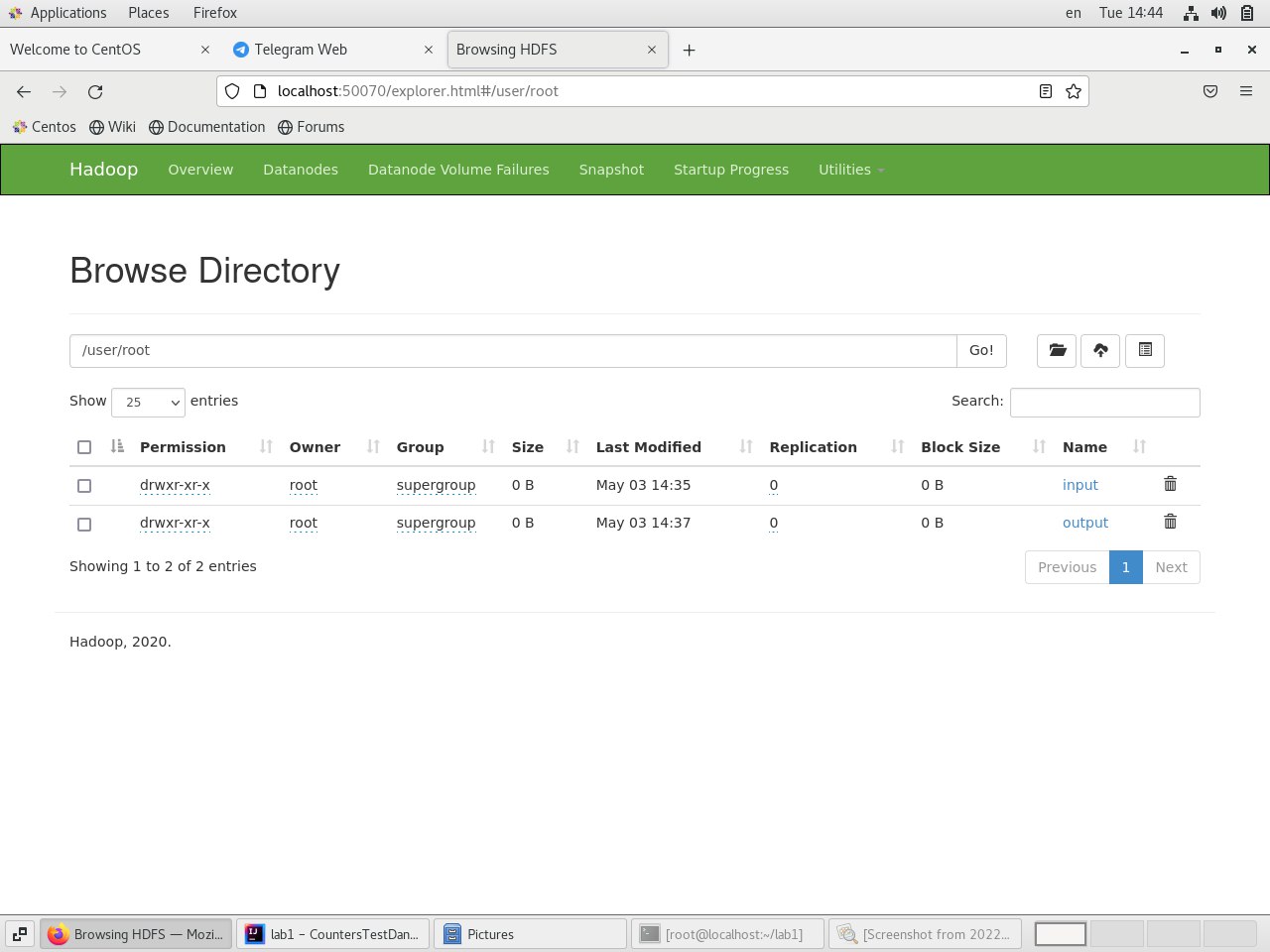


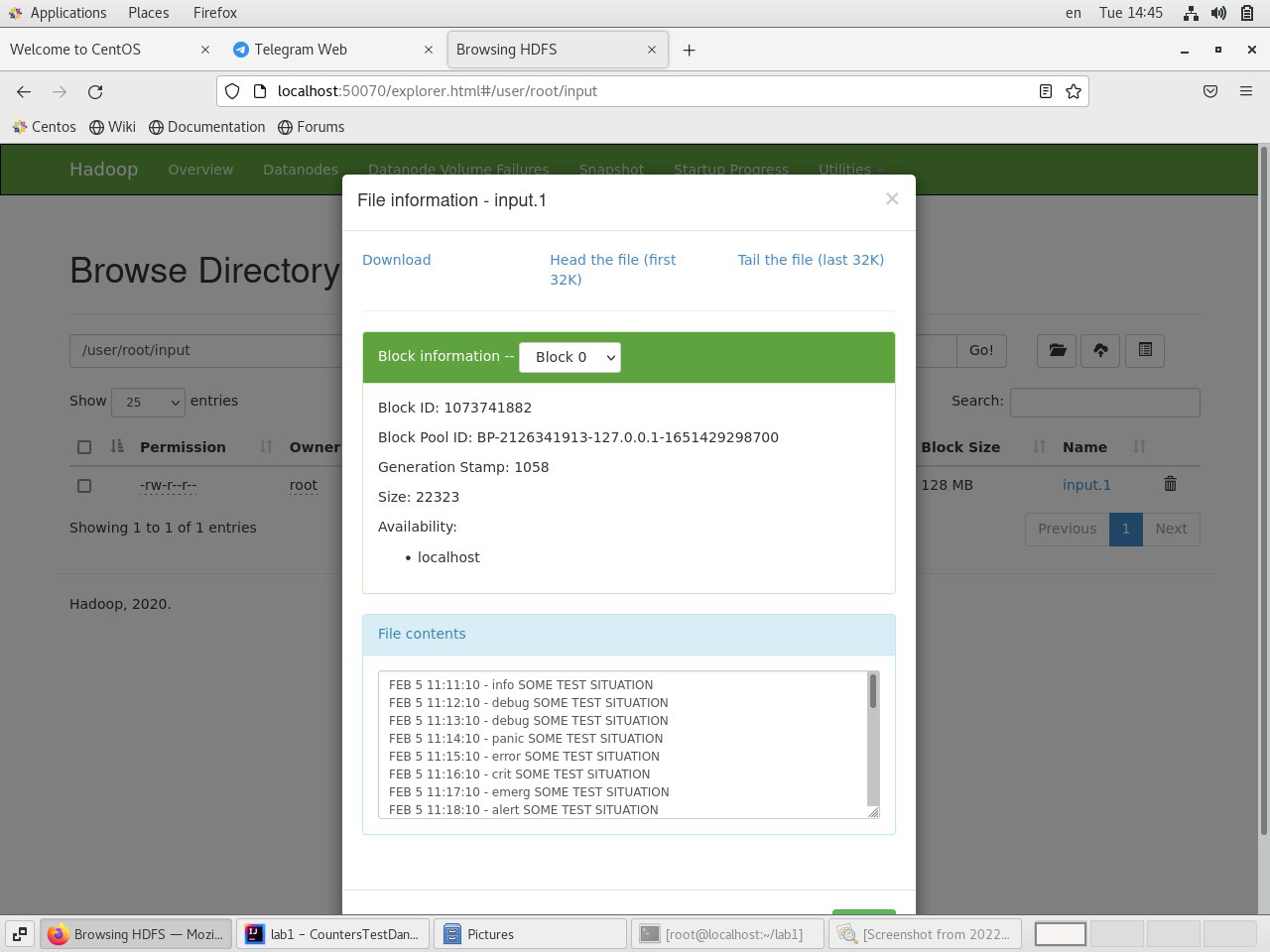
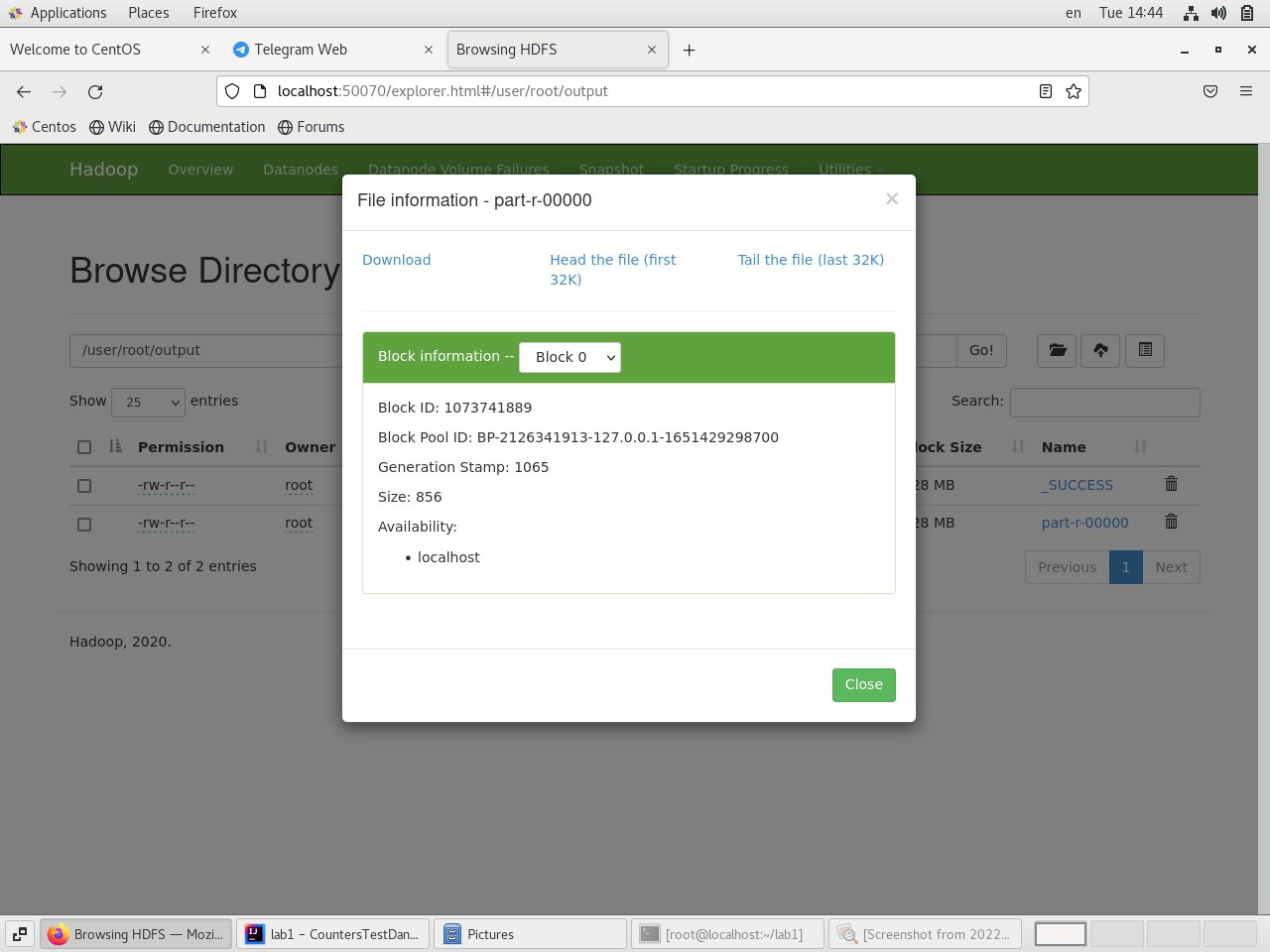


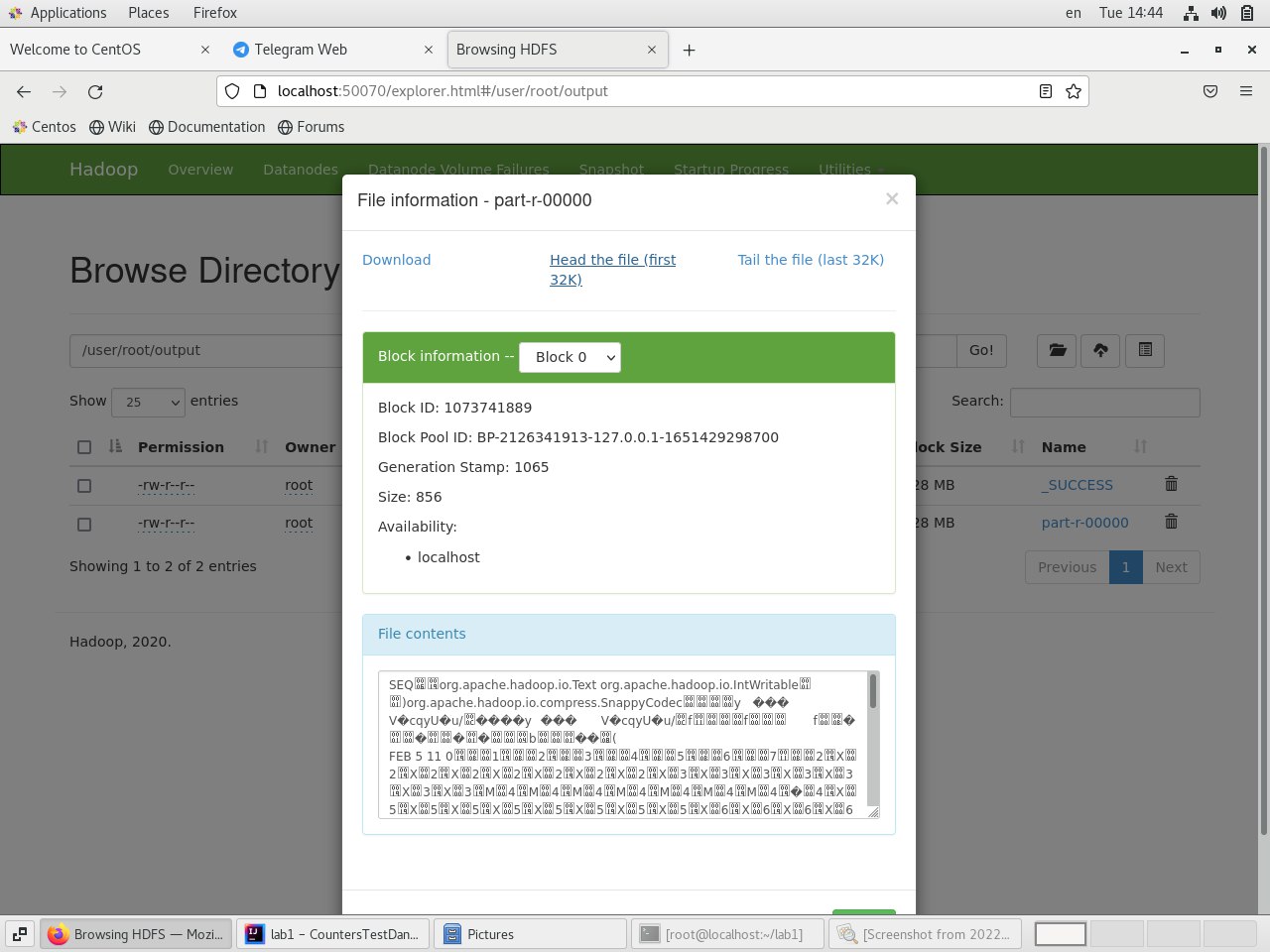




1. Просмотр результата localhost:50070/explorer.html#/user/root/





**Цель:** в ходе лабораторной работы были изучены основы Hadoop. Также были представлены скриншоты выполненных заданий.