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 Математической кибернетики и информационных технологий

ЛАБОРАТОРНАЯ РАБОТА № 2

по дисциплине

БОЛЬШИЕ ДАННЫЕ

Выполнил студент группы

БСТ1903

Козлов В. А

Вариант № 7

Москва 2021г.

СОДЕРЖАНИЕ

[ЗАДАНИЕ 3](#_Toc88742036)

[1 ХОД РАБОТЫ 4](#_Toc88742037)

[1.1 Доступ к панели администратора 4](#_Toc88742038)

[1.2 Работа с HDFS 5](#_Toc88742039)

[1.3 Cloudera Express и Cloudera Manager 7](#_Toc88742040)

# ЗАДАНИЕ

1. Установить на виртуальную машину образ Hadoop (Для этого также понадобится VirtualBox или VMware)
2. Выполнить следующие шаги в терминале для доступа к панели администратора:
   1. hostname # Покажет имя хоста: quickstart.cloudera
   2. hdfs dfs -ls / # Проверка на наличие доступа и работу кластера. Покажет файлы в системе HDFS
   3. service cloudera-scm-server status # Указывает, какую команду вам нужно ввести, чтобы использовать cloudera express бесплатно
   4. su - #Вход в рута
   5. service cloudera-scm-server status # Пароль для рута: cloudera
3. [Исследовательское] Сколько узлов одновременно можно потерять без потери данных в кластере из 10 узлов? Из 100 узлов?
4. Опробовать консольные утилиты для работы с кластером

* Создать/скопировать/удалить папку
* Положить в HDFS любой файл
* Скопировать/удалить этот файл
* Просмотреть размер любой папки
* Посмотреть как файл хранится на файловой системе (см. команду fsck)
* Установить нестандартный фактор репликации (см. команду setrep)

1. Запустить Cloudera Express и Cloudera Manager. Изучить интерфейс.

# 1 ХОД РАБОТЫ

## 1.1 Доступ к панели администратора

На рисунке 1 представлены подпункты a и b второго задания.

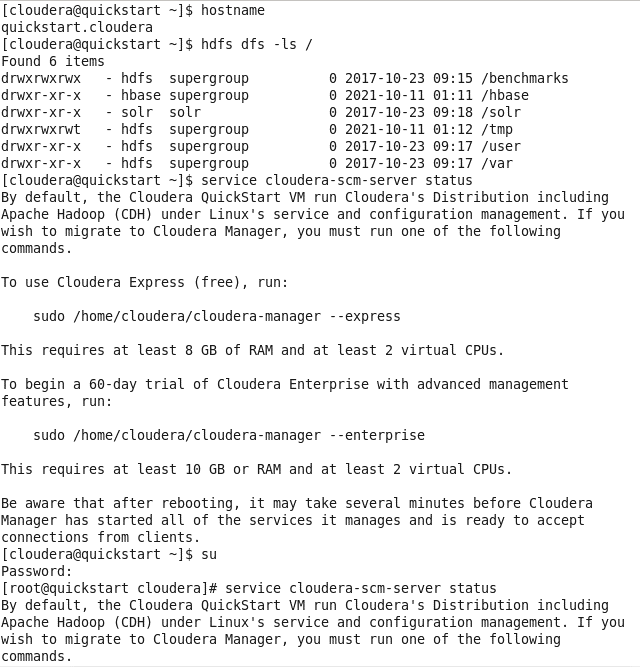


Рисунок 1 – Имя хоста и файлы в системе HDFS

На рисунке 2 представлены подпункты c,d,e второго задания.

service cloudera-scm-server status, который указывает, какую команду вам нужно ввести, чтобы использовать cloudera express бесплатно;

su – который обеспечивает вход в root. Пароль: cloudera

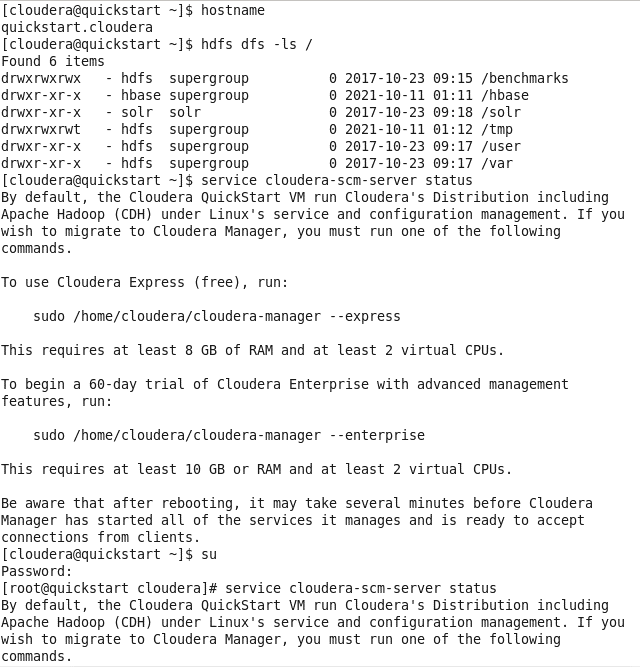


Рисунок 2 – Пункты c,d,e

## 1.2 Работа с HDFS

На рисунке 3 представлены файлы в системе HDFS до создания папки.

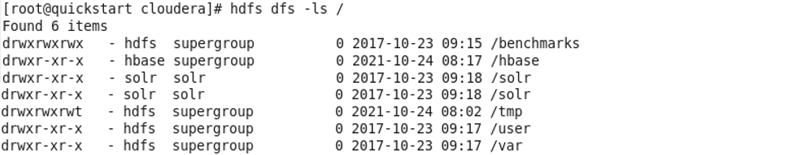


Рисунок 3 – файлы в системе HDFS

На рисунке 4 представлено создание папки, проверка ее наличия и удаление папки.

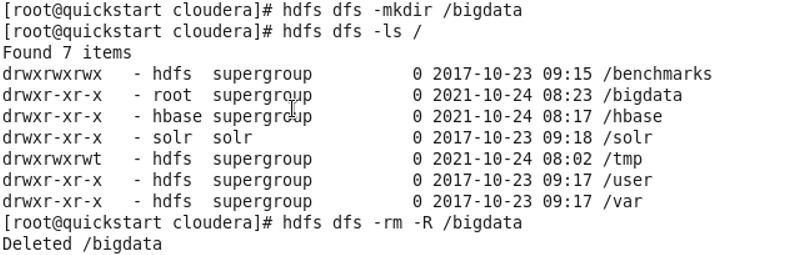


Рисунок 4 – Создание и удаление папки bigdata

На рисунке 5 представлен просмотр размера папки user. Сначала в байтах, затем в мегабайтах.



Рисунок 5 – Размер папки user

На рисунке 6 представлено перемещение файла lab2\_bd, созданного в локальной системе, в систему HDFS, информация о том, как хранится файл в файловой системе.

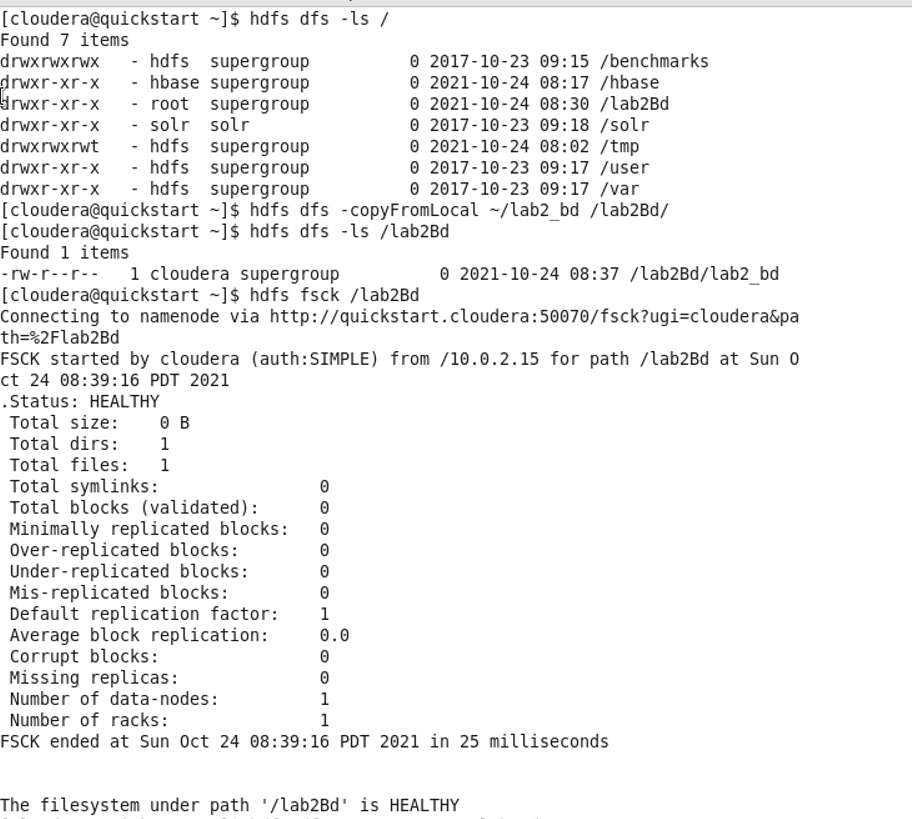


Рисунок 6 – Перемещение lab2\_bd в HDFS и просмотр хранения файла

На рисунке 7 представлено установление нестандартного фактора репликации и удаление файла.

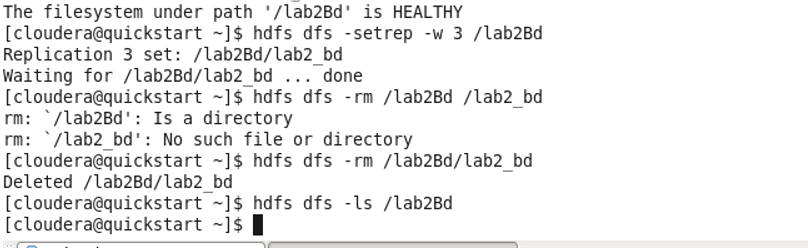


Рисунок 7 – Репликация

## 1.3 Cloudera Express и Cloudera Manager

На рисунках 8-9 представлены интерфейсы Cloudera Express и Cloudera Manager.

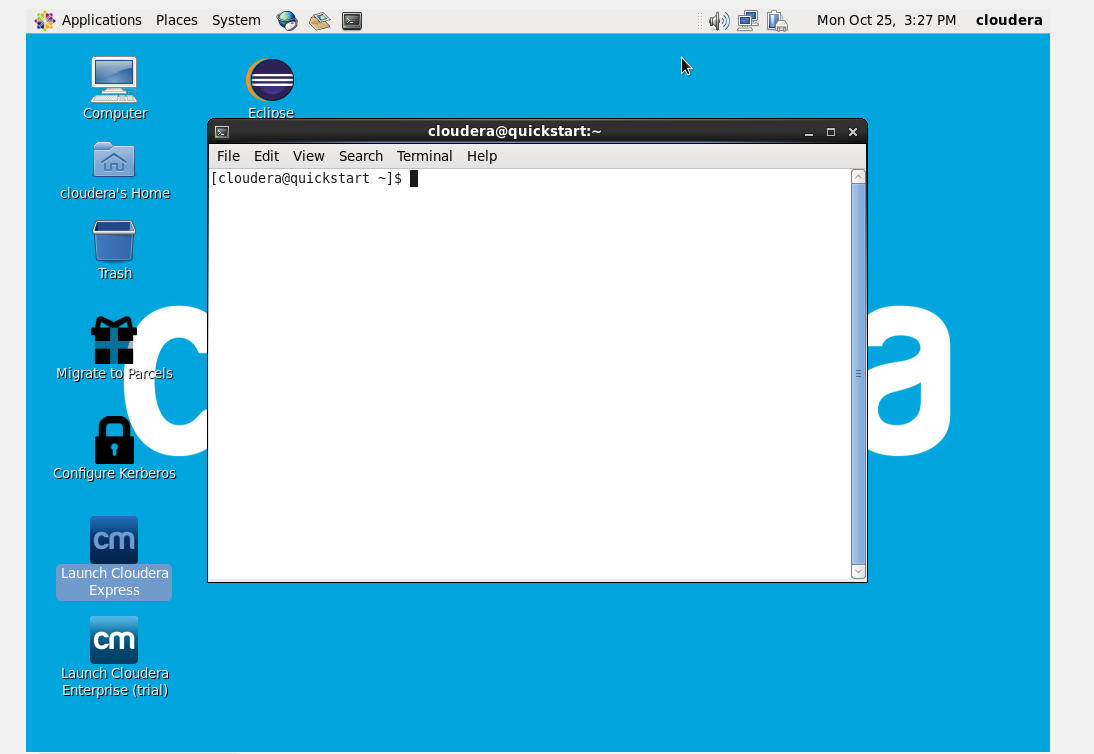


Рисунок 8 – Cloudera Express

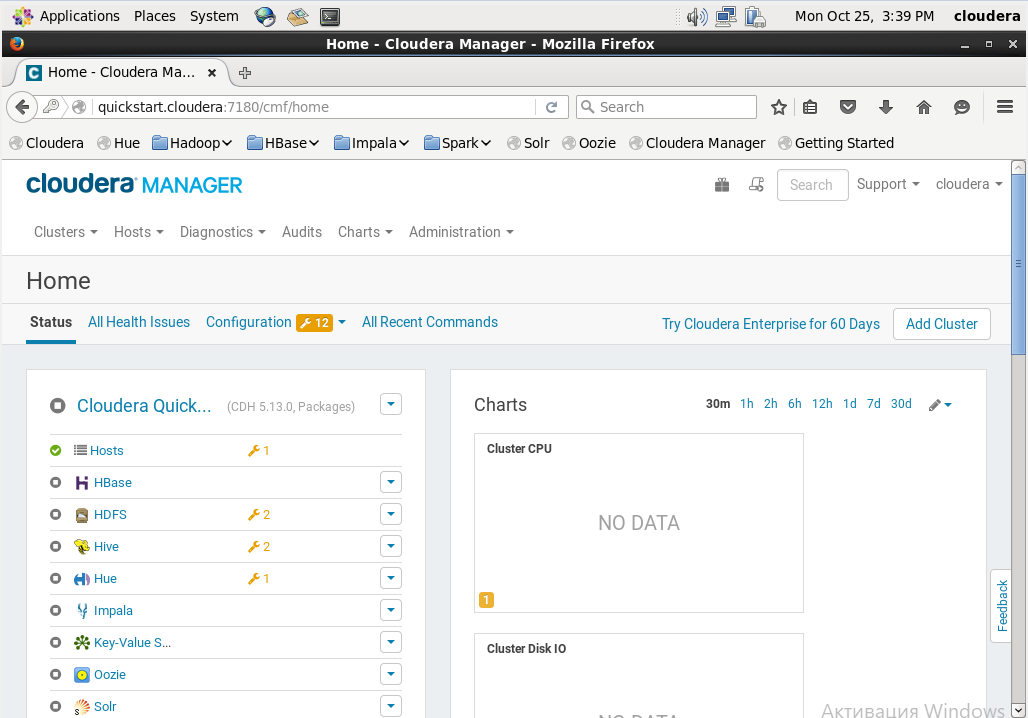


Рисунок 9 – Cloudera Manager