**Лабораторная работа № 2** **(Hadoop)**

**Кубарев А.Д.**

**Салтыш С.С.**

**Юткин Г.А.**

**Краткое описание лабораторной работы**

Установка Hadoop 3.2.2 происходила по инструкции, начиная с настройки операционной системы Ubuntu 18.

Ссылка на инструкцию:

<https://орел-57.рф/pages/aadress.php?page=468>

Подробное описание работы Hadoop:

<https://coderlessons.com/tutorials/bolshie-dannye-i-analitika/uchitsia-hadoop/hadoop-kratkoe-rukovodstvo>

Базовые команды:

https://crazythings.medium.com/hadoop-basic-commands-8a42f8041005

Для работы нам потребуется: OpenJDK (Среда выполнения Java служб, т.к. Hadoop написан на Java), SSH-сервер

Кроме того, потребуется создать отдельного пользователя для Hadoop, чтобы изолировать файловую систему Hadoop от файловой системы Unix.

Hadoop работает в **псевдораспределенном режиме** - это распределенное моделирование на одной машине. Каждый демон (в терминологии Linux - служба) Hadoop, такой как hdfs, yarn, MapReduce и т.д., Будет работать как отдельный процесс Java.

**Запуск установленного Hadoop:**

Переходим под созданный ранее hduser (похожа на отдельную учетную запись):

**sudo su hduser**

Запускаем файловую систему Hadoop: **start-dfs.sh**

Запускаем фреймворк YARN для управления ресурсами кластера и менеджмента задач, в том числе включает фреймворк MapReduce.: **start-yarn.sh**

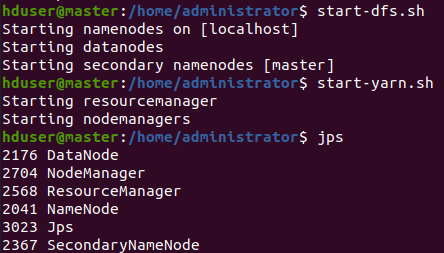
При помощи команды **jps** увидим, что запущены следующие службы:   
NodeManager, SecondaryNameNode, Jps, NameNode, ResourceManager и DataNode.

Для понимания, архитектура Hadoop:



**Задание 0**

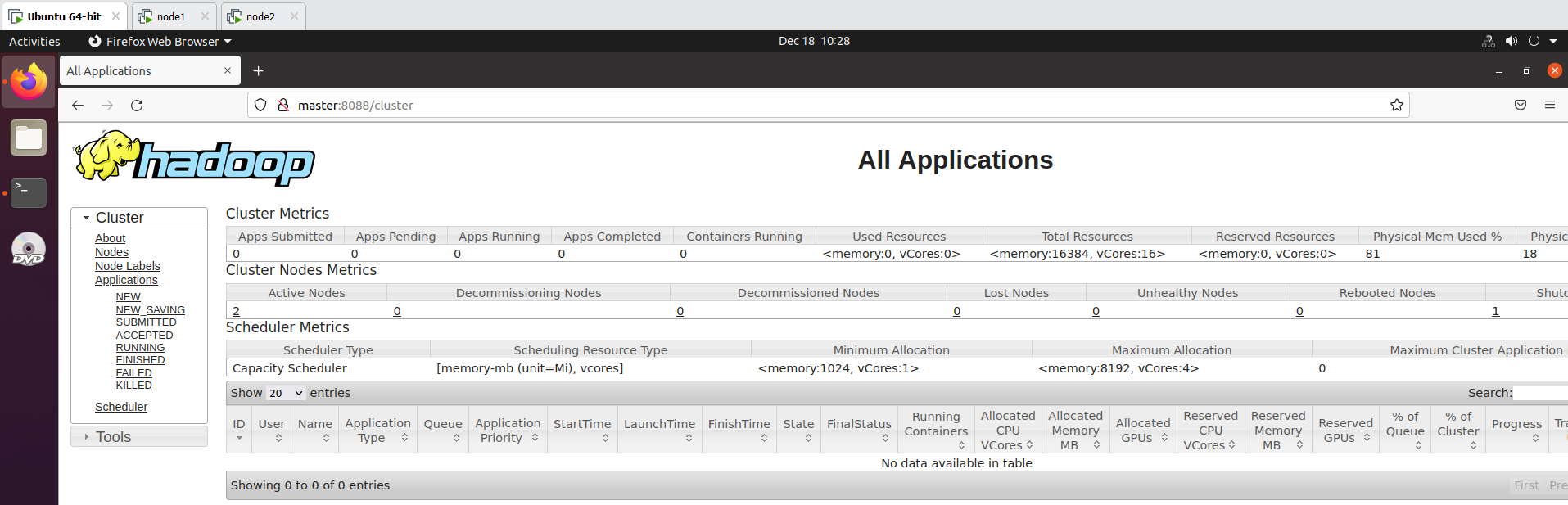
1. Установка и настройка кластера HDFS согласно инструкции. Запуск Hadoop:

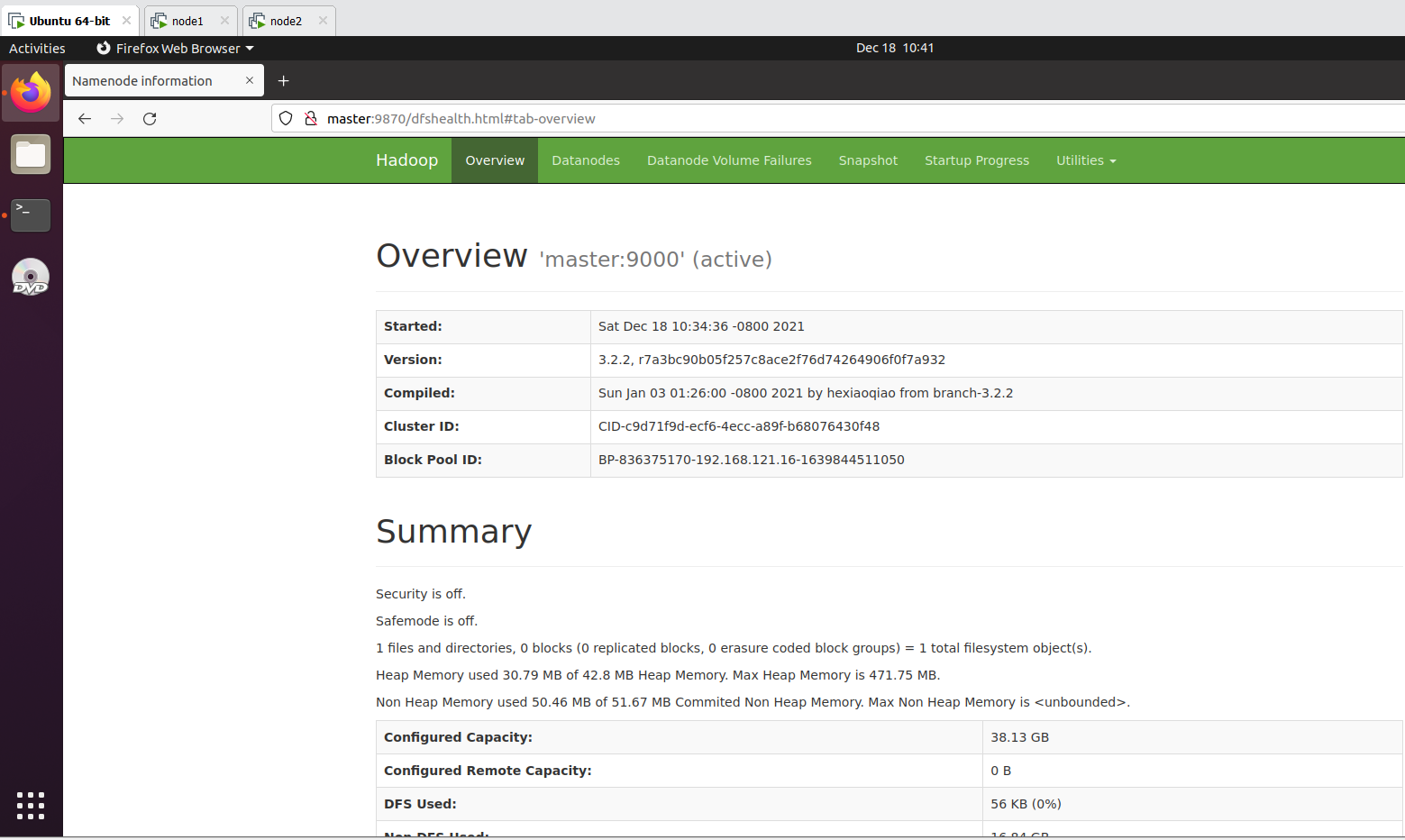


Данный вывод сообщений говорит нам о том, что Hadoop запущено на машине **master** без ошибок.

При успешной настройке также запускаются WEB-интерфейсы hdfs и задач MR в hadoop.

Интерфейс задач доступен по адресу http://Master:8088



Интерфейс состояния HDFS доступен по адресу [http://master:9870](http://master:9870)

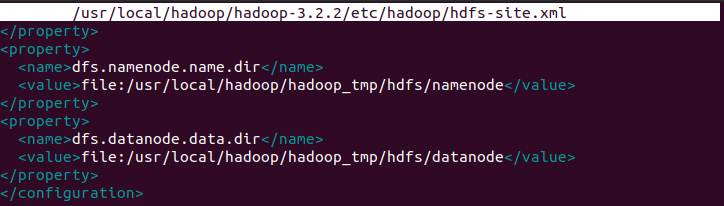
1. Воспроизвести схему кластера из презентации (Лабораторная №2)

Для воспроизведение схемы были созданы дополнительно 2 машины:



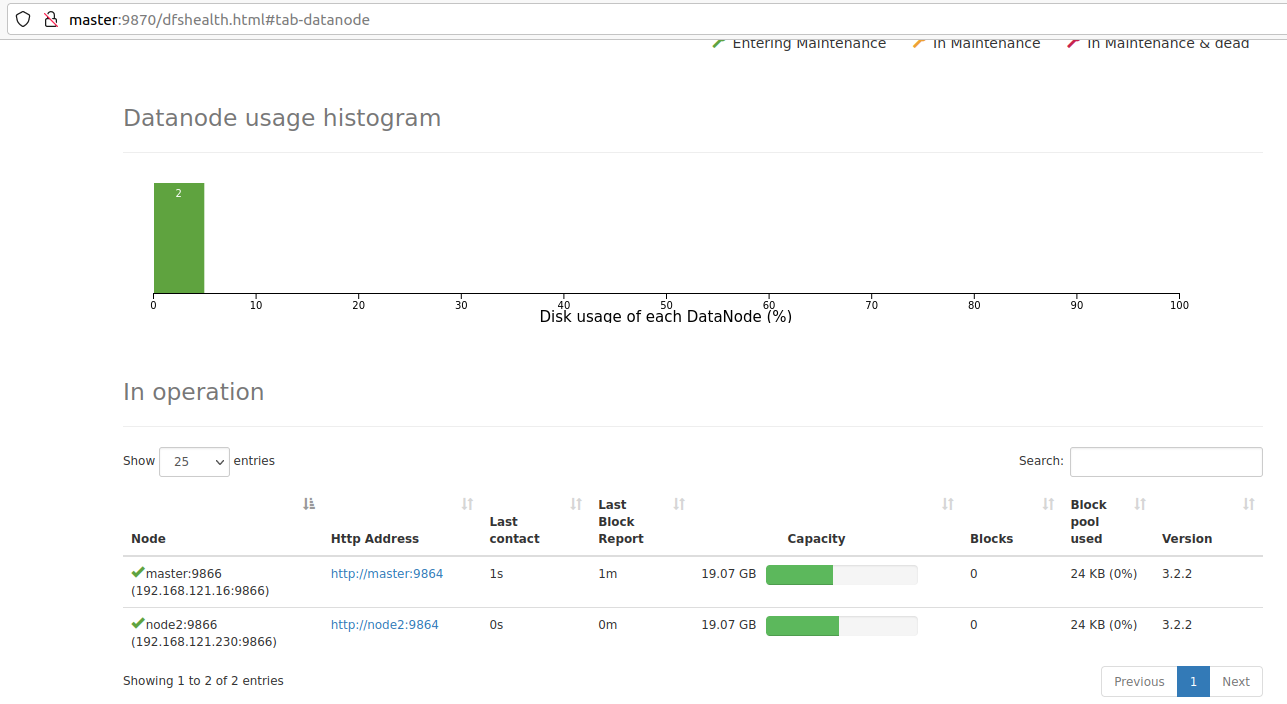
|  |  |
| --- | --- |
| Название машины | IP-адрес |
| master | 192.168.121.16/24 |
| node1 | 192.168.121.240/24 |
| node2 | 192.168.121.230/24 |

1. Настройка Namenode, Datanode:

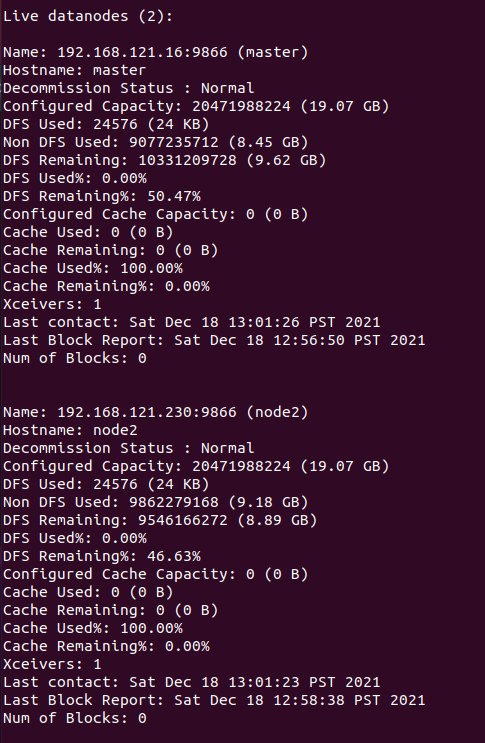


**Задание 1**

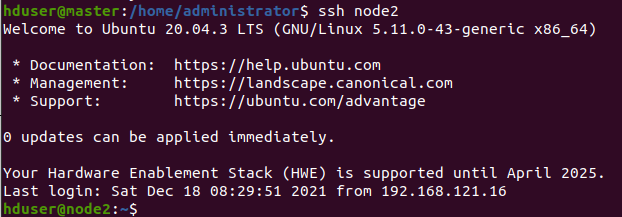
1,2. Произвели настройку node2 и запуск:



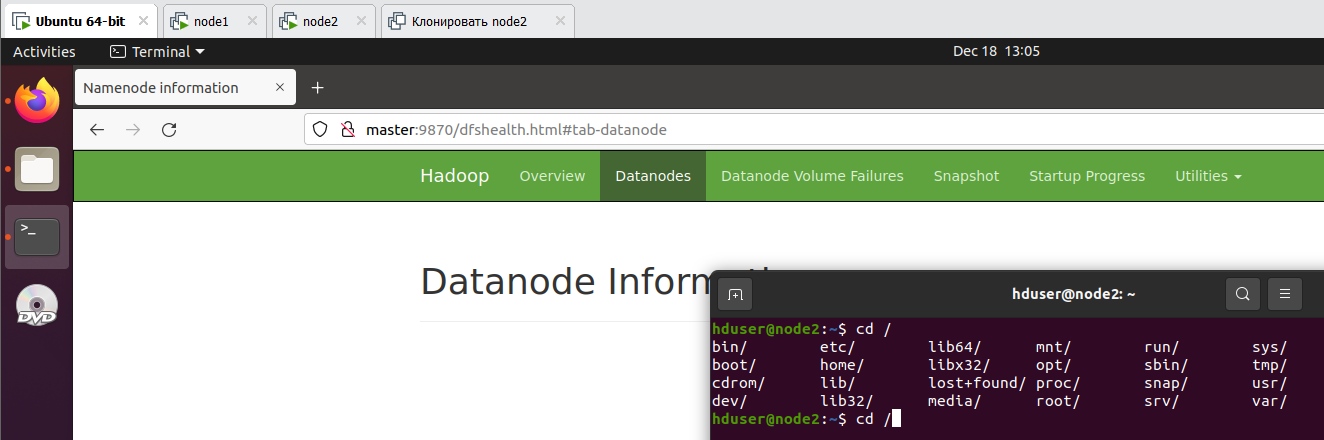
3. Отчет по рабочим узлам



1. С помощью SSH подключаемся на **node2:**



Корневой каталог на **master:**

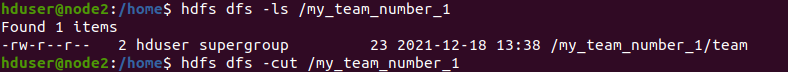


**Задание 2**

1. Создаем папку с идентификатором команды (**hdfs dfs -mkdir /my\_team\_number\_1**):



2. Помещаем файл в папку с идентификатором



Читаем содержимое файла:

