|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Методы анализа данных»

**Практическое занятие № 1**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИМБО-02-22, Ким Кирилл Сергеевич* | (подпись) | |
| Преподаватель | *Морошкин Никита Андреевич, преподаватель* | (подпись) | |
| Отчет представлен | «\_\_\_»\_\_\_\_\_\_\_\_202\_\_г. | |  | |

Москва 2024 г.

1. Практическая работа №1

1.1 Цель работы

Изучение HDFS и работу с ней с помощью команд и веб-интерфейса. Изучение работы с домашними каталогами Linux и HDFS. Изучение работы с HDFS с помощью веб-интерфейса.

1.2 Запуск служб Hadoop – HDFS и YARN

Чтобы начать использовать службу Hadoop, мы должны запустить демоны службы Hadoop. Мы делаем это как пользователь «hadoop» из нашей среды Linux. (Рисунок 1)

Изображение выглядит как текст, Шрифт, снимок экрана, число

Автоматически созданное описание

Рисунок 1 — Изменение пользователя с помощью команды su

Измените рабочий каталог на ~/hadoop/sbin и выполните сценарий start-dfs.sh, чтобы запустить службы HDFS. (Рисунок 2)

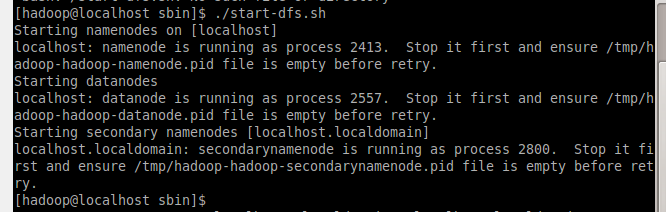


Рисунок 2 — Изменение рабочего каталога и запуск ./start-dfs.sh

Был проведен успешный запуск демона namenode и datanode для служб HDFS. Выполнение сценариев start-yarn.sh, чтобы запустить службы YARN показано на Рисунке 3.

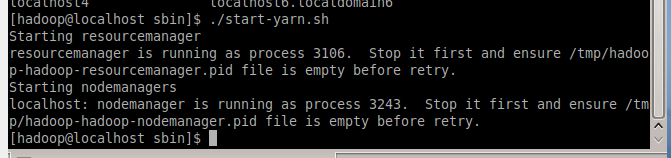


Рисунок 3 — Изменение рабочего каталога и запуск ./start-yarn.sh

После можно выйти из сеанса как пользователь Hadoop. Сделайте это, введя exit (выход) в терминале.



Рисунок 4 — Изменение рабочего каталога и запуск ./start-yarn.sh

1.3 Работа с домашними каталогами Linux и HDFS

Время вывести все каталоги (Рисунок 5):

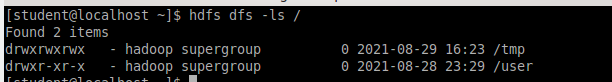


Рисунок 5 — Показатели для квантования вещественных данных

Также нужно вывести каталоги /user, /user/student, /home и /home/student на Рисунке 6, Рисунке 7, Рисунке 8 и Рисунке 9:

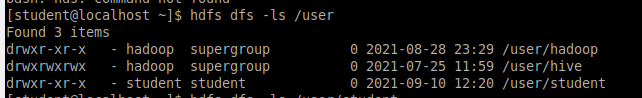


Рисунок 6 – Каталог /user

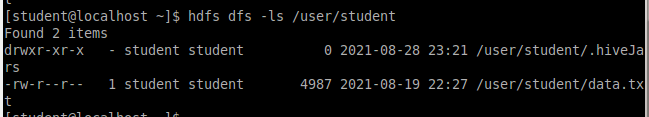


Рисунок 7 – Каталог /user/student

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 8 – Каталог /home

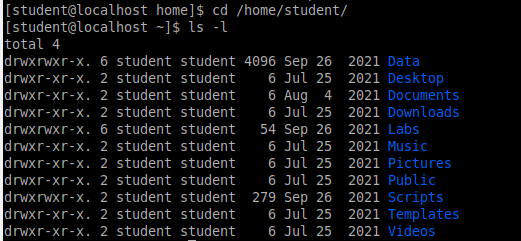


Рисунок 9 – Каталог /home/student

Распространенной задачей администраторов Hadoop является создание новых пользователей. При этом обычно создается домашний каталог пользователя как для Linux, так и для HDFS. Создание пользователя и установка пароля, а также добавление его в группу wheel показаны на Рисунке 10:

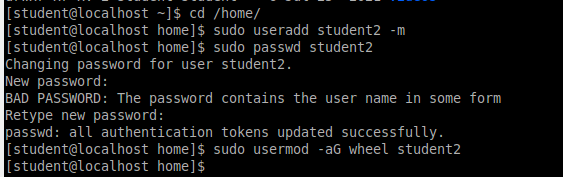


Рисунок 10 – Создание пользователя

Логин через нового пользователя и вывод пустого каталога показаны на Рисунке 11:

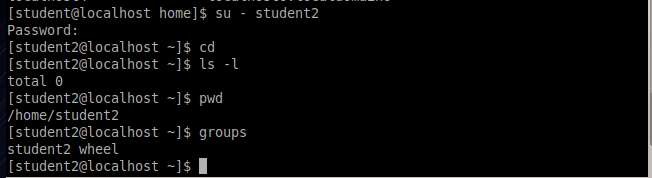


Рисунок 11 — Вывод вышеперечисленных команд

Теперь нужно будет вводить команды от имени пользователя Hadoop, чтобы внести изменения в HDFS. Обычные пользователи могут вносить изменения только в свой домашний каталог HDFS. Однако им могут быть предоставлены дополнительные привилегии с помощью политик безопасности. Вход через Hadoop и проверка пустого каталога показаны на Рисунке 12:

Изображение выглядит как текст, Шрифт, снимок экрана

Автоматически созданное описание

Рисунок 12 — Вывод каталогов и вход под Hadoop

1.4 Изучение и работа с каталогами HDFS

Создание подкаталога в домашнем каталоге студента HDFS и называние его «MRtest» показаны на Рисунке 13, а после вывод длинного файла показан на Рисунке 14:

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 13 — Просмотр файла в директории

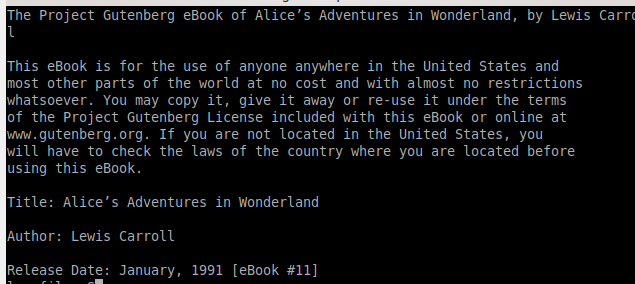


Рисунок 14 — Вывод части содержимого файла через команду less

Исполним команду «положить» книгу. В Linux мы можем разбивать длинные команды на несколько строк, используя \ в конце строки. Это говорит Linux о том, что команда не закончена и что у нее еще есть продолжение. На Рисунке 15 запросим HDFS поместить файл alice\_in\_wonderland.txt в каталог HDFS /user/student/MRtest. На Рисунке 16 проверим, что файл успешно переместился:



Рисунок 15 — Перемещение файла

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Рисунок 16 — Проверка того, что файл переместился

1.5 Исследование HDFS с помощью веб-интерфейса

Просмотр вкладки Overview на Рисунке 17:

Изображение выглядит как текст, электроника, снимок экрана, программное обеспечение

Автоматически созданное описание

Рисунок 17 — Вкладка Overview

Просмотр вкладки Datanodes на Рисунке 18:

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание

Рисунок 18 — Вкладка Datanodes

На этой вкладке мы можем просматривать состояние и статистику узлов данных. Поскольку наша лабораторная среда на самом деле не является кластером, мы увидим только 1 узел данных на Рисунке 19:

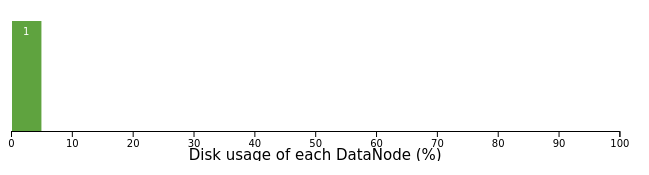


Рисунок 19 — Единственный узел данных

Изучение вкладки Snapshot на Рисунке 20:

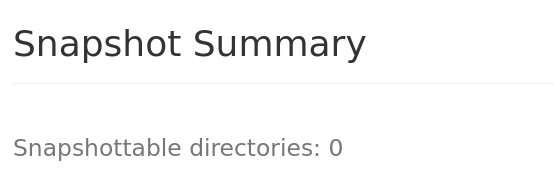


Рисунок 20 — Вкладка Snapshot

Изучение вкладки Startup Progress на Рисунке 21:

Изображение выглядит как текст, снимок экрана, Шрифт, белый

Автоматически созданное описание

Рисунок 21 — Вкладка Startup Progress

Просмотр ранее перемещенного файла через Utilities показаны на Рисунке 22 и Рисунке 23:

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Рисунок 22 — Необходимый файл с правами доступа

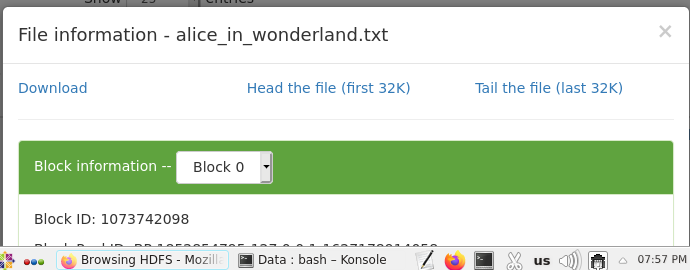


Рисунок 23 — Всплывающее окно с информацией о блоке

Далее идет работа с уникальными ключами. Его вывод показан на Рисунке 24:

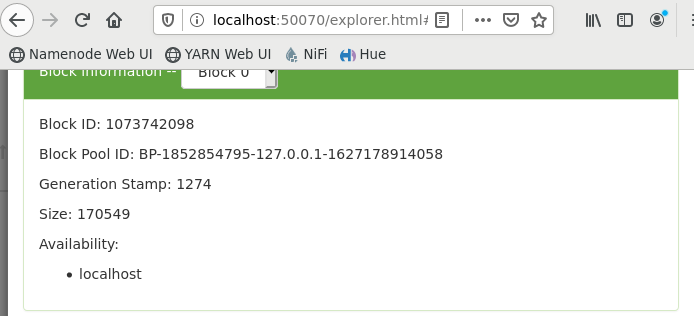


Рисунок 24 — Block Pool ID, необходимый далее

После захода под пользователем Hadoop используем команду Linux find для поиска Block Pool ID (идентификатор пула блоков) на Рисунке 25:

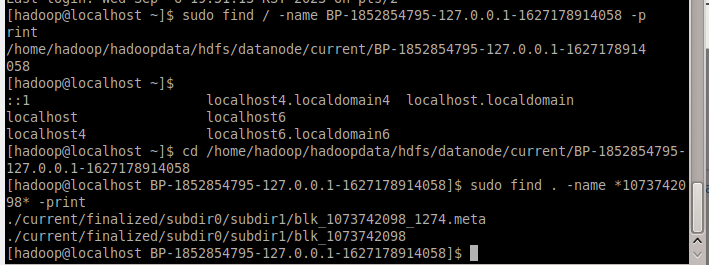
****

Рисунок 25 — Команда find

Приведенная выше команда указывает Linux начать с / (root) directory (/ (корневого) каталога) и найти файл или каталог с заданным именем и распечатать информацию. Перейдем в каталог, где был найден Block Pool ID (идентификатор пула блоков), и изучим его на Рисунке 25 (выше) и Рисунке 26, а также выведем сам файл на Рисунке 27:



Рисунок 26 — Команда less по ключу

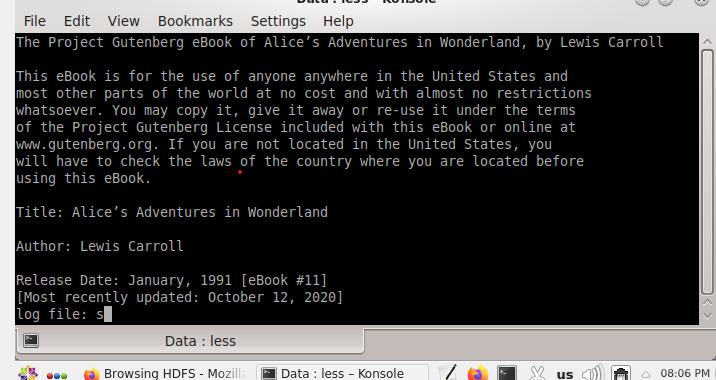


Рисунок 27 — Выведенный через less файл

1.6 Работа с YARN/MapReduce

Мы будем использовать файл alice\_in\_wonderland.txt, который был сохранен в HDFS, для запуска программы MapReduce подсчета слов. Сперва включим Yarn historyserver, а также запускаем файл с параметром подсчета слов (Рисунок 28):

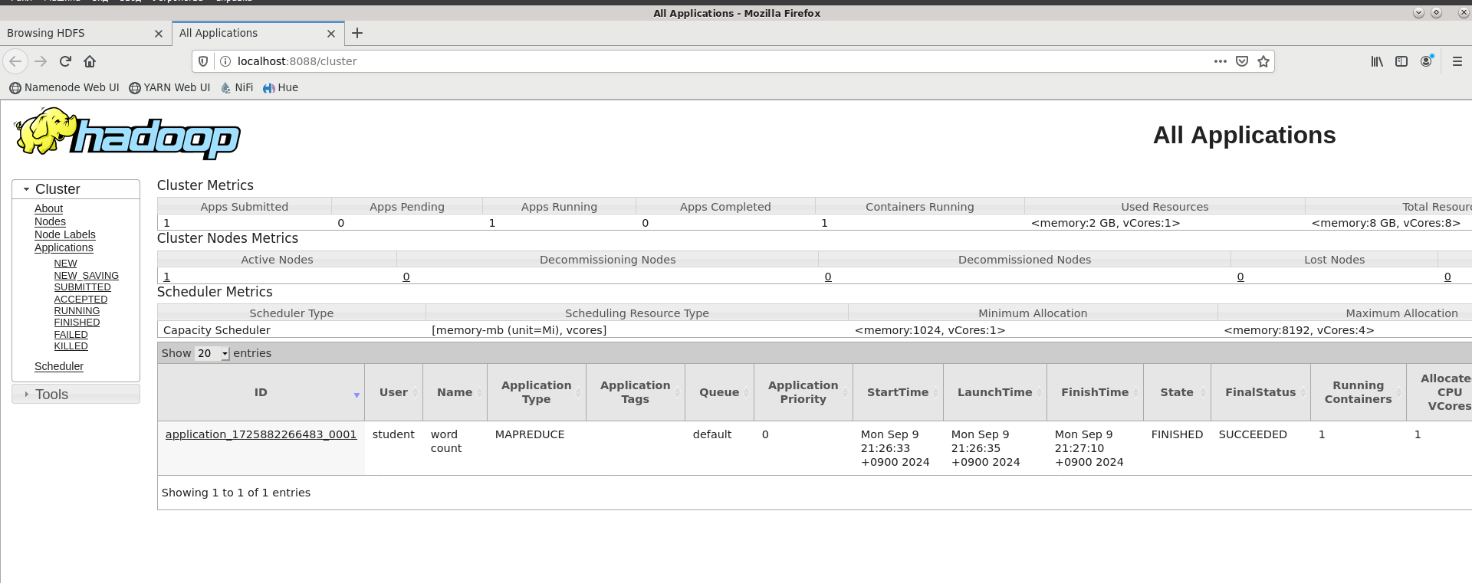
****

Рисунок 28 — Результат вышеприведенных действий

Открывает браузер и переходим к веб-интерфейсу YARN (Рисунок 29):

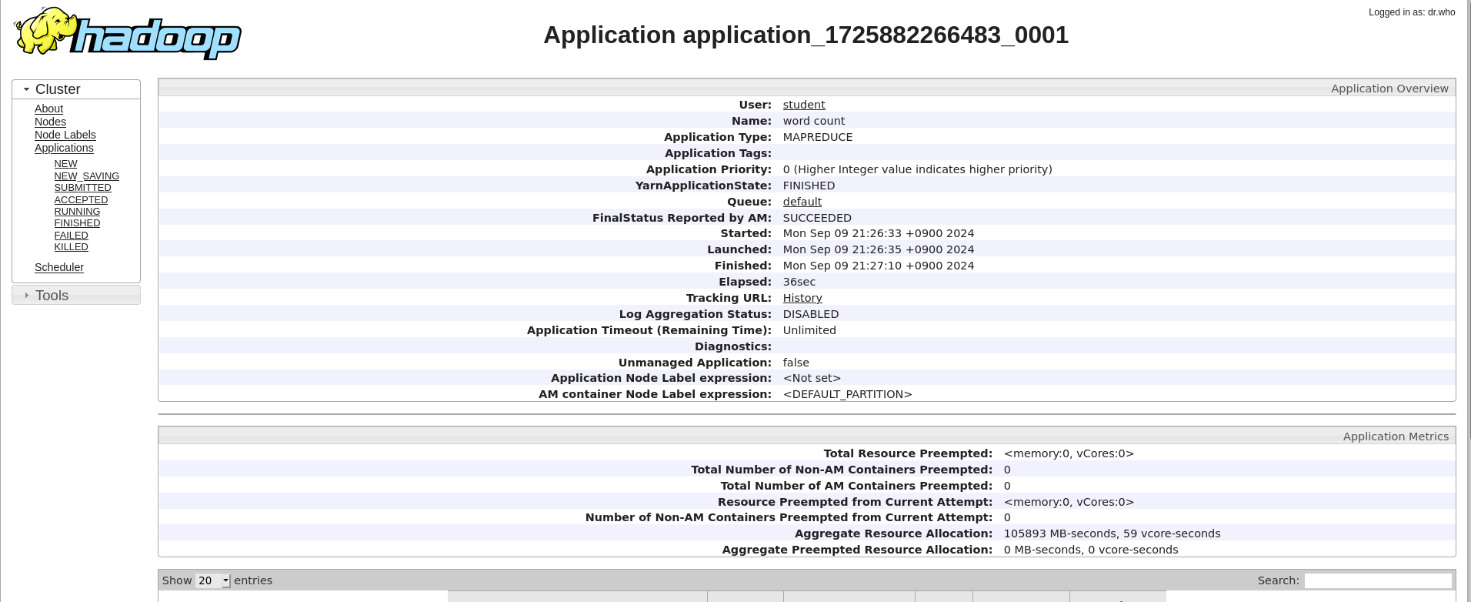
****

Рисунок 29 — Веб интерфейс YARN

Переходим по ссылкам для получения Application Master (Рисунок 30):

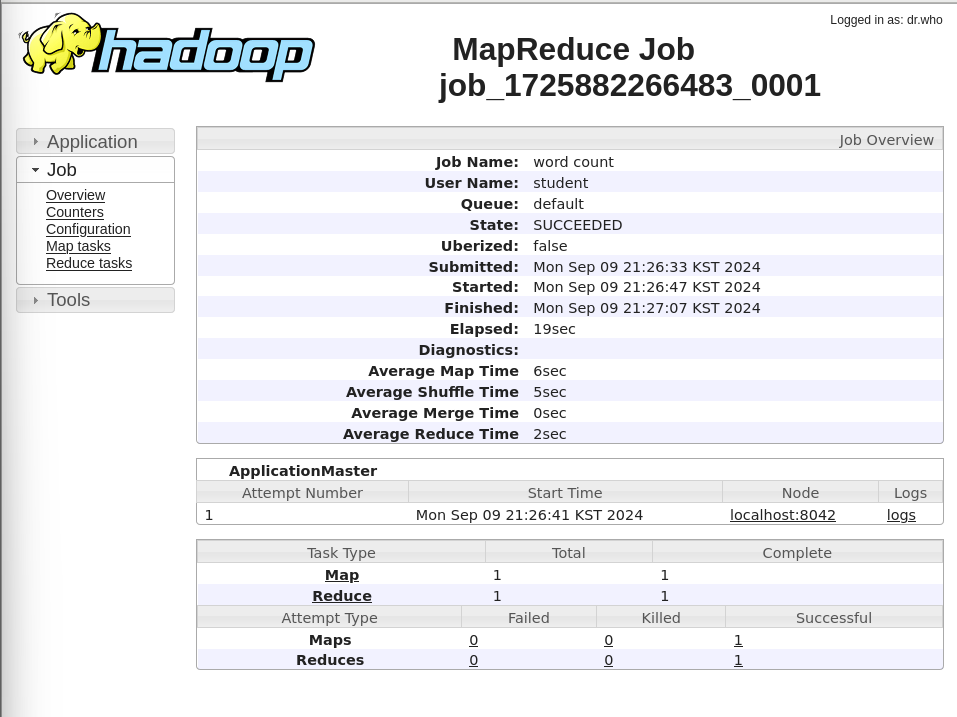
****

Рисунок 30 — Веб интерфейс YARN

Рассмотрим относящиеся к веб-интерфейсу MapReduce Application (Рисунок 31, Рисунок 32, Рисунок 33):

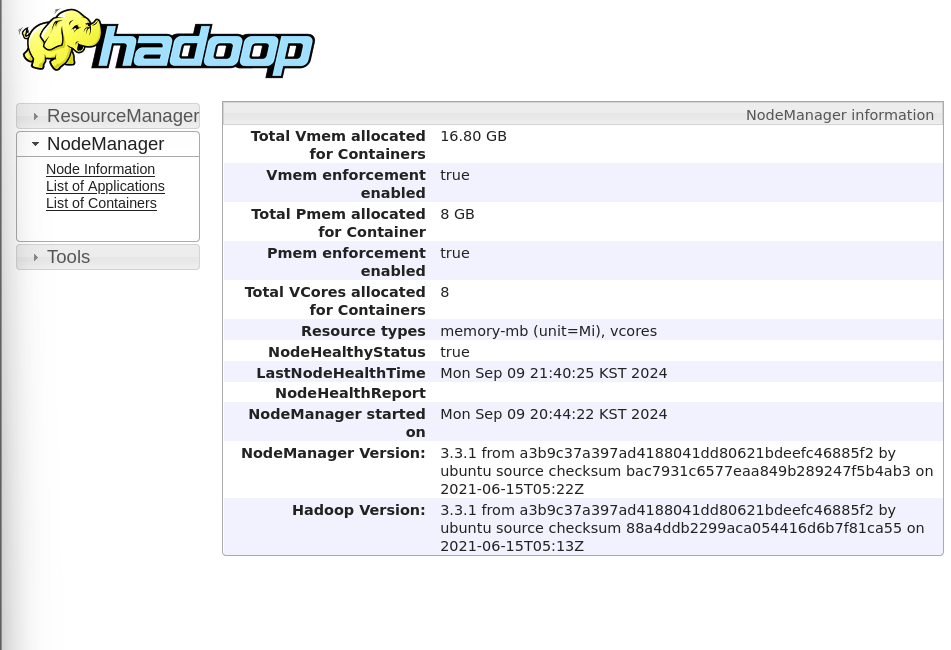


Рисунок 31 — Веб интерфейс YARN

После вышеприведенных операций рекурсивно удаляем каталог (Рисунок 32):



Рисунок 32 — Удаление каталога