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 «Системное программирование»

Практическая работа №2

**«Введение в HDFS»**

по дисциплине

Введение в большие данные

Выполнил: студент гр. БЭИ2202

Кулешов А. С.

Проверил: Сергеев А.Е.

Москва, 2025 г

Цель работы. Получить навыки работы с файловой системой HDFS

**ВЫПОЛНЕНИЕ**

Скачаю образ Ubuntu, а также виртуальную машину, настрою виртуальную машину в соответствии с заданием, после проведу установку и распаковку образа ОС.

Затем открою терминал, обновлю пакеты и выведу системную информацию на экран, скачаю Vim и MC. Проверю их наличие, узнаю из какого репозитория они были скачаны, узнаю их зависимости. Результат работы можно увидеть на рисунках 1-20.

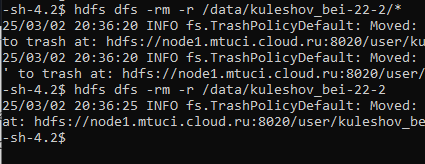


Рисунок 1 – Удаление данных

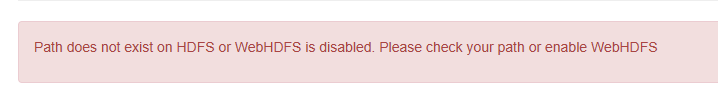


Рисунок 2 – Данные не найдены

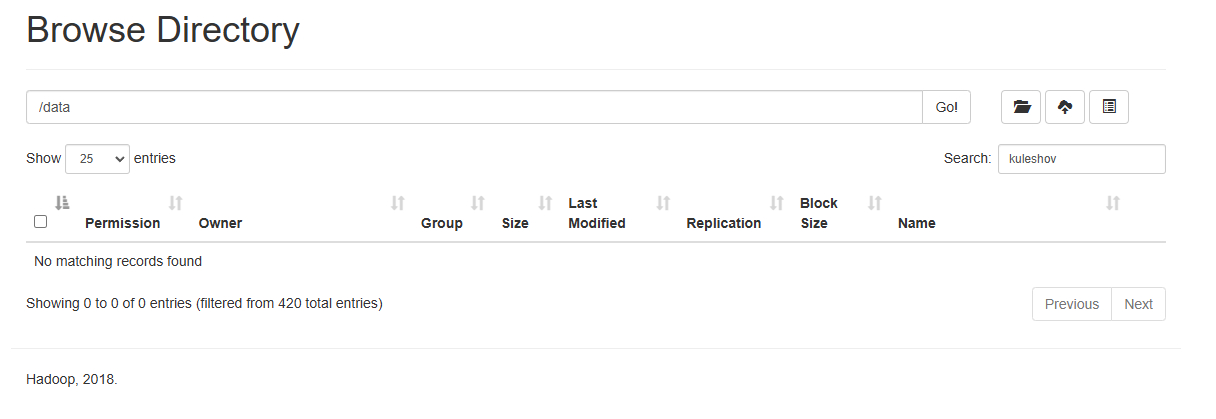


Рисунок 3 – Папки более нет

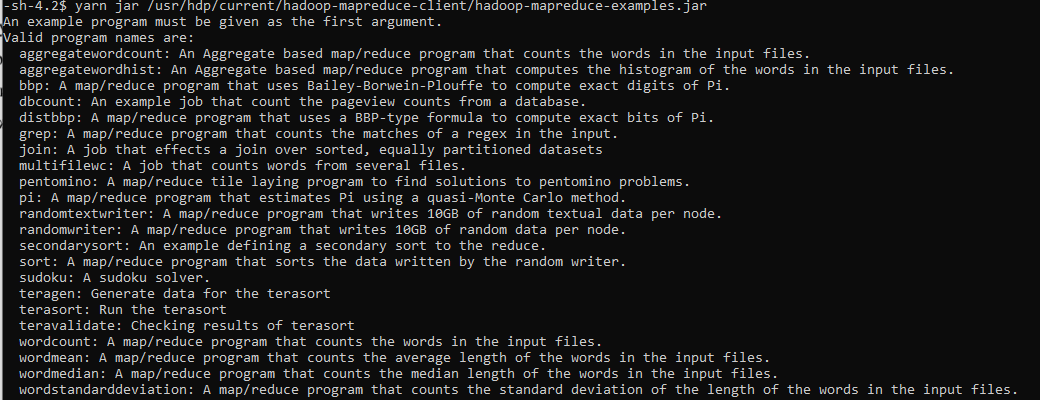


Рисунок 4 – Запуск команды

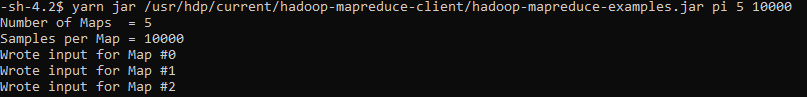


Рисунок 5 – Запуск команды для вычисления числа пи

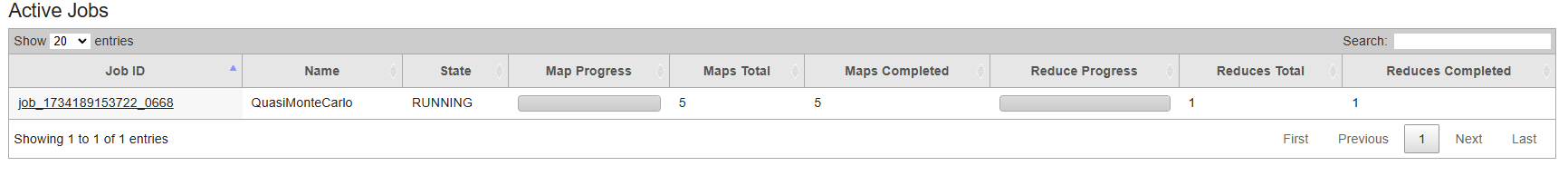


Рисунок 6 – Состояние джобы



Рисунок 7 – Результат



Рисунок 8 – Разница во времени до получения результата (в консоли)



Рисунок 9 – Разница во времени до получения результата (в трекере)

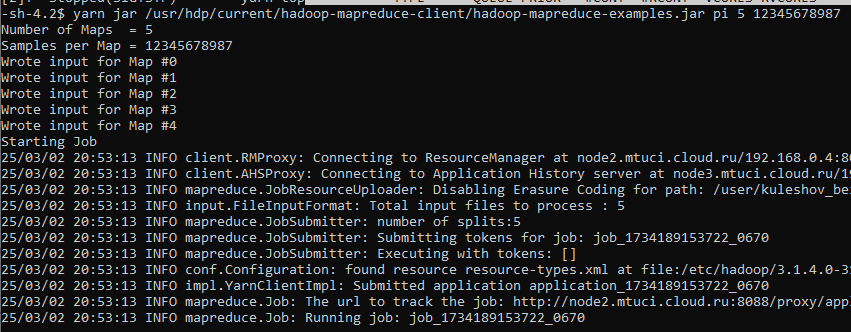


Рисунок 10 – Запуск с огромными данными

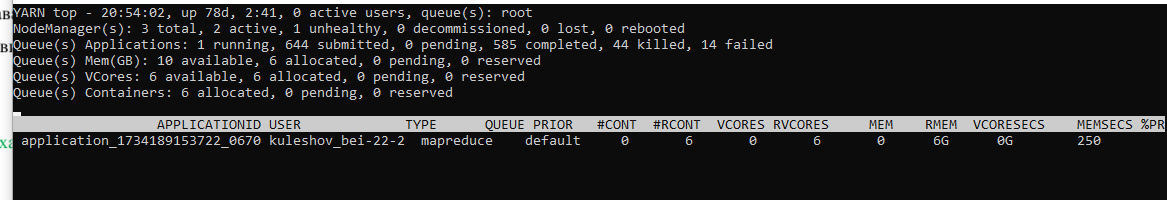


Рисунок 11 – Консольный монитор (yarn top)

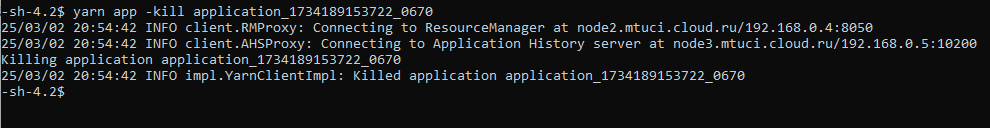


Рисунок 12 – Убийство джобы

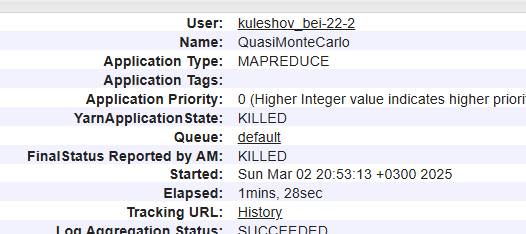


Рисунок 13 – Состояние KILLED

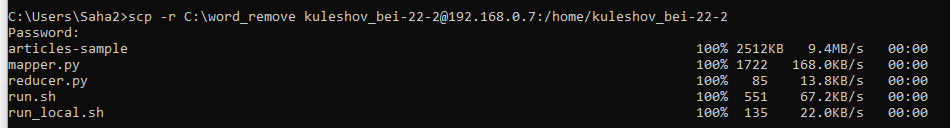


Рисунок 14 – Перемещение файлов с моего компьютера на локальный



Рисунок 15 – Запуск скрипта

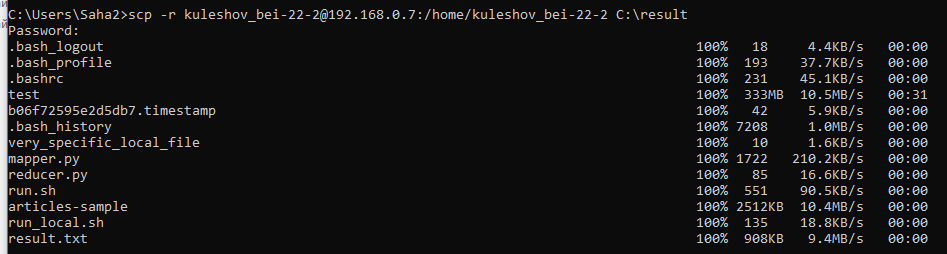


Рисунок 16 – Скачивание локальной директории себе на компьютер

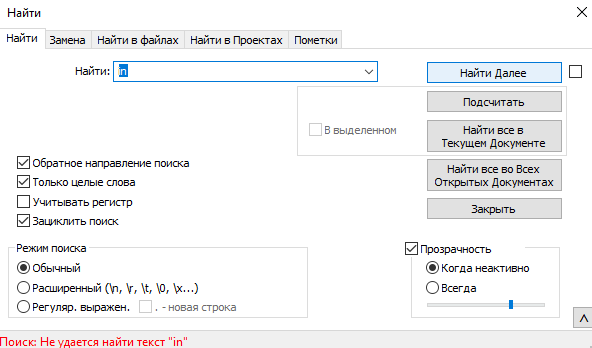


Рисунок 17 – Слова действительно отсутствуют

mapper.py

#!/usr/bin/env python3

**import** **sys**

**import** **re**

stop\_words = set(['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've",

"you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself',

'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs',

'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am',

'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does',

'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at',

'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after',

'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again',

'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',

'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own',

'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should',

"should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn',

"couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't",

'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn',

"needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren',

"weren't", 'won', "won't", 'wouldn', "wouldn't"])

**for** line **in** sys.stdin:

pattern = r"\b(" + "|".join(map(re.escape, stop\_words)) + r")\b"

cleaned\_text = re.sub(pattern, "", line, flags=re.IGNORECASE)

cleaned\_text = re.sub(r"\s+", " ", cleaned\_text).strip()

print(cleaned\_text)

reducer.py

#!/usr/bin/env python3

**import** **sys**

**for** line **in** sys.stdin:

print(line.strip())