Департамент образования и науки города Москвы
Государственное автономное образовательное учреждение
высшего образования города Москвы
«Московский городской педагогический университет»
Институт цифрового образования
Департамент информатики, управления и технологий

Практическая работа № 7

Тема: «HADOOP самостоятельное и индивидуальное задания»

Выполнил студент ТП-191: Эльмукова О. Г.

Руководитель: Босенко Т. М.

Москва

#### Самостоятельное задание

### Текст «Муму»

```
Lemp001@u20-17:-$ sudo su hadoop
[sudo] password for lemp001:
hadoop@u20-17:/home/lemp001$ cd
hadoop@u20-17:-$ git clone http://github.com/OksanaElmukova/practicds.git
Cloning into 'practicds'...
warning: redirecting to https://github.com/OksanaElmukova/practicds.git/
remote: Enumerating objects: 119, done.
remote: Counting objects: 100% (35/35), done.
remote: Compressing objects: 100% (32/32), done.
remote: Total 119 (delta 10), reused 0 (delta 0), pack-reused 84
Receiving objects: 100% (119/119), 4.29 MiB | 7.29 MiB/s, done.
Resolving deltas: 100% (28/28), done.
```

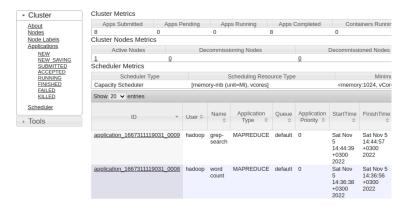
1. Копирование файла из репозитория GIT в HDFS каталог пользователя:

```
nadoop@u20-17:~$ hadoop fs -put practicds/m.txt /user/hadoop/m.txt
2022-11-05 14:35:09,043 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
```

- 2. Запуск MapReduce Jar по умолчанию (hadoop-mapreduceexamples-
  - 3.1.2.jar ) для вычисления количества слов для текстового файла «m.txt»:

```
hadoop@uz0-17:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examp les-3.1.2.jar wordcount /user/hadoop/Faust_1.txt /user 2022-11-05 14:35:32,320 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable 2022-11-05 14:35:33,163 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.8032 org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://localhost:9000/user already exists
```

3. Проверка в Диспетчере ресурсов выполнение задания:



4. Копирование полученного файла MapReduce обратно в локальную файловую систему ubuntu (using bash):

```
hadoop@u20-17:~$ hadoop fs -get /user/hadoop/m_Output/part-r-00000 m_Output.csv
2022-11-05 14:38:42,231 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-17:~$ head -10 m_Output.csv

"and 1
"do 2
(it 1
- 2
Alienated 1
And 1
At 1
But 3
Finally, 1
Gavrila 4
```

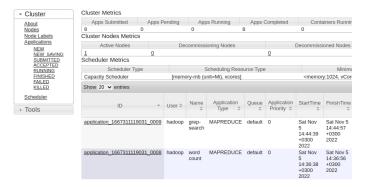
5. Копирование полученного файла MapReduce обратно в локальную файловую систему ubuntu (using Web Filebrowser):

# 

6. Запуск MapReduce Jar по умолчанию (hadoop-mapreduce-examples-3.1.2.jar) для поиска в grep строки «Мити» в текстовом файле «m.txt» и подсчета количества повторений данного слова:

hadoop@u20-17:~\$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examp les-3.1.2.jar grep /user/hadoop/m.txt /user/hadoop/m\_Count\_Output "Mumu" 2022-11-05 14:47:14,244 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable 2022-11-05 14:47:15,090 INFO client.RMProxy: Connecting to ResourceManager at /0 .0.0.0:8032 2022-11-05 14:47:15,568 INFO mapreduce.JobResourceUploader: Disabling Erasure Co ding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job\_1667311119031\_0010 2022-11-05 14:47:15,840 INFO input.FileInputFormat: Total input files to process : 1

7. Проверка в Диспетчере ресурсов выполнение задания:



8. Копирование полученного файла MapReduce обратно в локальную файловую систему ubuntu (using bash):



9. Копирование полученного файла MapReduce обратно в локальную файловую систему ubuntu (using Web Filebrowser):



### Индивидуальное задание

1. Развернуть виртуальное окружение:

```
lemp001@u20-17:~$ sudo su hadoop

[sudo] password for lemp001:

hadoop@u20-17:/home/lemp001$ cd

hadoop@u20-17:~$ jps

201714 Jps

14738 DataNode

15287 NodeManager

15144 ResourceManager

14957 SecondaryNameNode

14575 NameNode
```

2. Вывести с помощью команды help описание основных команды shell-клиента:

```
hadoop@u20-17:-$ hdfs dfs -help
2022-11-07 00:48:37,929 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Usage: hadoop fs [generic options]
        [-appendToFile <localsrc> ... <dst>]
        [-cat [-ignoreCrc] <src> ...]
        [-checksum <src> ...]
        [-chegrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
        [-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] <localsrc> ... <
```

3. Просмотреть корневую директорию HDFS:

```
hadoop@u20-17:~$ hdfs dfs -ls / hdfs://localhost:9000/user
2022-11-07 00:49:20,819 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 2 items
drwx------ - hadoop supergroup 0 2022-11-03 19:38 /tmp
drwxr-xr-x - hadoop supergroup 0 2022-11-07 00:47 /user
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2022-11-05 14:51 hdfs://localhost:90
00/user/hadoop
```

4. Создать в HDFS в директории /user/mgpu поддиректорию /Elmukova:



5. Создать в локальной файловой системе случайный текстовый файл размером 10 Mb с именем OG.txt:

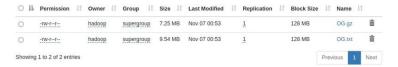
```
hadoop@u20-17:~$ base64 /dev/urandom | head -c 10000000 > OG.txt
```

6. Заархивировать созданный текстовый файл:

```
hadoop@u20-17:~$ gzip -c OG.txt > OG.gz
```

7. Скопировать текстовый файл и архив в директорию /user/mgpu/Elmukova HDFS виртуальной машины:

```
hadoop@u20-17:-$ hadoop fs -copyFromLocal OG.txt /user/mgpu/Elmukova
2022-11-07 00:53:19,837 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-17:-$ hadoop fs -copyFromLocal OG.gz /user/mgpu/Elmukova
2022-11-07 00:53:30,693 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your_platform... using builtin-java classes where applicable
```



8. Просмотреть файл и архив с помощью утилит cat, text в комбинации с каналами и утилитами head, tail -- привести не менее 3 вариантов команд и просмотра файла:



9. Создать копию файла OG.txt вида date\_file.txt, где в начале имени файла-копии указана текущая дата. Вывести листинг.

```
hadoop@u20-17:~$ cp OG.txt `date +%d%m%Y`_OG.txt
hadoop@u20-17:~$ ls

06112022_OG.txt Faust_1_Output.csv hadoopdata
07112022_OG.txt file.txt m_Count_Output.csv OG.txt
ds_practice hadoop m_Output.csv practicds
Faust_1_Count_O hadoop-3.1.2.tar.gz Mumu_Output.csv
```

10.Вывести статистику по директории /user/mgpu/Elmukova виртуальной машины:

hadoop@u20-177-\$ hdfs dfs -du -h /user/mgpu/Elmukova 2022-11-07 00:58:26,166 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable 7.2 M 7.2 M /user/mgpu/Elmukova/OG.gz 9.5 M 9.5 M /user/mgpu/Elmukova/OG.txt

## 11. Удалить поддиректорию /Elmukova со всем содержимым:

