

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  
2022

## 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:

```
hadoop@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@u20-17:~$ 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
.0.0.0:8032
org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://loc
alhost:9000/user already exists
```

3. Проверка в Диспетчере ресурсов выполнение задания:

Cluster

About

Nodes

Node Labels

Applications

NEW

SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted

Apps Pending

Apps Running

Apps Completed

Containers Running

8

0

0

8

0

Cluster Nodes Metrics

Active Nodes

Decommissioning Nodes

Decommissioned Nodes

1

0

0

Scheduler Metrics

Scheduler Type

Scheduling Resource Type

Minimum

Capacity Scheduler

[memory-mb (unit=Mi), vcores]

<memory:1024, vCores:1

Show 20 ▼ entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime
application_1667311119031_0009	hadoop	grep-search	MAPREDUCE	default	0	Sat Nov 5 14:44:57 +0300 2022	Sat Nov 5 14:44:57 +0300 2022
application_1667311119031_0008	hadoop	word count	MAPREDUCE	default	0	Sat Nov 5 14:36:38 +0300 2022	Sat Nov 5 14:36:56 +0300 2022

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):

## Browse Directory

/user/hadoopm\_Output

Go!



Show25entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	0 B	Nov 05 14:36	1	128 MB	_SUCCESS
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	5.85 KB	Nov 05 14:36	1	128 MB	part-r-00000

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadoop, 2018.

Hadoop, 2018.

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

```
hadoop@u20-17:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-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.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-05 14:47:15,568 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding 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:

▼ Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted

Apps Pending

Apps Running

Apps Completed

Containers Running

Cluster Nodes Metrics

Active Nodes

Decommissioning Nodes

Decommissioned Nodes

Scheduler Metrics

Scheduler Type

Scheduling Resource Type

Minim

Capacity Scheduler

[memory-mb (unit=M), vcores]

<memory:1024, vCor

Show 20 ▼ entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime
application_1667311119031_0009	hadoop	grep-search	MAPREDUCE	default	0	Sat Nov 5 14:44:39 +0300 2022	Sat Nov 5 14:44:57 +0300 2022
application_1667311119031_0008	hadoop	word count	MAPREDUCE	default	0	Sat Nov 5 14:36:38 +0300 2022	Sat Nov 5 14:36:56 +0300 2022

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

```
hadoop@u20-17:~$ hadoop fs -get /user/hadoop/m_Count_Output/part-r-00000 m_Count_Output.csv
2022-11-05 14:48:14,482 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@u20-17:~$ cat m_Count_Output.csv
1
Mumu
```

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

Browse Di	Download	Head the file (first 32K)	Tail the file (last 32K)
/user/hadoop/m_Count Show 25 entries Permission Showing 1 to 2 of 2 entries Hadoop, 2018.	Block information = Block 0 Block ID: 1073741928 Block Pool ID: BP-1541712709-172.26.35.137-16673111074549 Generation Stamp: 1104 Size: 7 Availability: u20-17		
	File contents		
	1 Mumu		

## 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> ...]
[-chgrp [-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/mgpr поддиректорию /Elmukova:

```
hadoop@u20-17:~$ hadoop fs -mkdir /user/mgpr/Elmukova
2022-11-07 00:50:16,018 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
```

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	Nov 07 01:06	0	0 B	Elmukova

Showing 1 to 1 of 1 entries

### 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/mgpr/Elmukova HDFS виртуальной машины:

```
hadoop@u20-17:~$ hadoop fs -copyFromLocal OG.txt /user/mgpr/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/mgpr/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
```



```
hadoop@u20-17:~$ 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:

```
hadoop@u20-17:~$ hdfs dfs -rm -r /user/mgpu/Elmukova
2022-11-07 00:59:20,474 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Deleted /user/mgpu/Elmukova
```

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
No data available in table									
Showing 0 to 0 of 0 entries									<div>PreviousNext</div>