

## 0. [Исследовательское задание]

Как получить доступ под пользователем hdfs в файловую систему не имея sudo?

Если до воскресенья никто не расскажет -- будет подсказка.

Создам переменную

- export HADOOP\_USER\_NAME = HDFS
- Вовремя создание нового файла в папке student4\_10 с моей аутентификацией будет создан файл hdfs user под авторизацией hdfs благодаря выше указанной команде !

```
student4_10@manager:~$ whoami
student4_10
[student4_10@manager ~]$ export HADOOP_USER_NAME=hdfs
[student4_10@manager ~]$ hdfs dfs -touchz /student4_10/hdfs_user
[student4_10@manager ~]$
[student4_10@manager ~]$ hdfs dfs -ls /student4_10/
Found 3 items
drwxr-xr-x  - student4_10 supergroup      0 2020-05-25 21:52 /student4_10/Mal
-rw-r--r--  3 hdfs      supergroup      0 2020-05-28 16:29 /student4_10/hdf
s_user
-rw-r--r--  6 student4_10 supergroup      0 2020-05-25 20:47 /student4_10/myf
ile.txt
[student4_10@manager ~]$
```

1. Опробовать запуски map-reduce задач для кластера используя hadoop-mapreduce-examples.jar.

а. Выполнить три любых задачи включенных в этот JAR.

б. Найти свои задачи в интерфейсе Cloudera Manager

в. Опробовать навигацию по интерфейсу YARN

1. найдем jar file

```
find /some_directory -name some_file_name
```

- Находим файл Hadoop-mapreduce через find
- Создаем переменную YARN\_EXAMPLES
- И потом запускаем переменную с параметрами pi 16 1000000

```
student4_10@manager:~$ find /opt -name "hadoop-mapreduce-examples.jar"
/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/hadoop-mapreduce/hadoop-mapr
educe-examples.jar
[student4_10@manager ~]$ export YARN_EXAMPLES=/opt/cloudera/parcels/CDH-5.16.2-1.c
dh5.16.2.p0.8/lib/hadoop-mapreduce
[student4_10@manager ~]$ yarn jar $YARN_EXAMPLES/hadoop-mapreduce-examples.jar pi 16 1000000
```

## Вывод

```
student4_10@manager:~  
Data-local map tasks=16  
Total time spent by all maps in occupied slots (ms)=56629  
Total time spent by all reduces in occupied slots (ms)=7976  
Total time spent by all map tasks (ms)=56629  
Total time spent by all reduce tasks (ms)=7976  
Total vcore-milliseconds taken by all map tasks=56629  
Total vcore-milliseconds taken by all reduce tasks=7976  
Total megabyte-milliseconds taken by all map tasks=57988096  
Total megabyte-milliseconds taken by all reduce tasks=8167424  
Map-Reduce Framework  
  Map input records=16  
  Map output records=32  
  Map output bytes=288  
  Map output materialized bytes=576  
  Input split bytes=2470  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=2  
  Reduce shuffle bytes=576  
  Reduce input records=32  
  Reduce output records=0  
  Spilled Records=64  
  Shuffled Maps =16  
  Failed Shuffles=0  
  Merged Map outputs=16  
  GC time elapsed (ms)=1689  
  CPU time spent (ms)=16100  
  Physical memory (bytes) snapshot=7592968192  
  Virtual memory (bytes) snapshot=47486951424  
  Total committed heap usage (bytes)=7528251392  
Shuffle Errors  
  BAD_ID=0  
  CONNECTION=0  
  IO_ERROR=0  
  WRONG_LENGTH=0  
  WRONG_MAP=0  
  WRONG_REDUCE=0  
File Input Format Counters  
  Bytes Read=1888  
File Output Format Counters  
  Bytes Written=97  
Job Finished in 72.522 seconds  
Estimated value of Pi is 3.1415912500000000000000  
[student4_10@manager ~]$
```

## Cloudera Applications

Обучение | C YARN (1 x) | gb\_hadoop | GB-Hadoop | (232) THE | cd comma | Hadoop "c | find file in | Using hdfs | Apache H | + | | | |

← | ↻ | ⚠ 89.208.221.132:7180 | YARN (MR2 Included) - Cloudera Manager | | ★ Нет отзывов | | |

### Workload Summary

(For Completed Applications)

ADL Bytes Read

ADL Bytes Written

Allocated Memory Seconds

Allocated VCore Seconds

CPU Time

Duration

File Bytes Read

File Bytes Written

HDFS Bytes Read

HDFS Bytes Written

Memory Allocation

Results Charts

05/28/2020 5:04 PM

**Pending**  
(2.96s)

QuasiMonteCarlo

ID: application\_1589223973966\_0033

Type: MAPREDUCE

Duration: 2.96s

User: student4\_10

[Collect Diagnostic Data](#) [Export](#) [Select Attributes](#)

## Работа в Cloudera

## Второй пример wordcount, запускаем

```
student4_10@manager:~$ yarn jar $YARN_EXAMPLES/hadoop-mapreduce-examples.jar wordcount /acldir/etc/nfsmount.conf /tmp/output2
20/05/28 17:13:44 INFO client.RMProxy: Connecting to ResourceManager at manager.novalocal/89.208.221.132:8032
20/05/28 17:13:45 INFO input.FileInputFormat: Total input paths to process : 1
20/05/28 17:13:45 INFO mapreduce.JobSubmitter: number of splits:1
20/05/28 17:13:45 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1589223973966_0034
20/05/28 17:13:45 INFO impl.YarnClientImpl: Submitted application application_1589223973966_0034
20/05/28 17:13:45 INFO mapreduce.Job: The url to track the job: http://manager.novalocal:8088/proxy/application_1589223973966_0034/
20/05/28 17:13:45 INFO mapreduce.Job: Running job: job_1589223973966_0034
20/05/28 17:13:57 INFO mapreduce.Job: Job job_1589223973966_0034 running in uber mode : false
20/05/28 17:13:57 INFO mapreduce.Job: map 0% reduce 0%
20/05/28 17:14:05 INFO mapreduce.Job: map 100% reduce 0%
```

```
student4_10@manager:~$
Launched map tasks=1
Launched reduce tasks=6
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=4111
Total time spent by all reduces in occupied slots (ms)=28702
Total time spent by all map tasks (ms)=4111
Total time spent by all reduce tasks (ms)=28702
Total vcore-milliseconds taken by all map tasks=4111
Total vcore-milliseconds taken by all reduce tasks=28702
Total megabyte-milliseconds taken by all map tasks=4209664
Total megabyte-milliseconds taken by all reduce tasks=29390848
Map-Reduce Framework
  Map input records=135
  Map output records=593
  Map output bytes=5723
  Map output materialized bytes=2830
  Input split bytes=119
  Combine input records=593
  Combine output records=227
  Reduce input groups=227
  Reduce shuffle bytes=2830
  Reduce input records=227
  Reduce output records=227
  Spilled Records=454
  Shuffled Maps =6
  Failed Shuffles=0
  Merged Map outputs=6
  GC time elapsed (ms)=836
  CPU time spent (ms)=10100
  Physical memory (bytes) snapshot=1818824704
  Virtual memory (bytes) snapshot=19634774016
  Total committed heap usage (bytes)=1711800320
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=3390
File Output Format Counters
  Bytes Written=2277
[student4_10@manager ~]$
```

Обучение | GeekB YARN (MR2 Incl X gb\_hadoop/lesson GB-Hadoop/hw2 (232) THE FROZEN yarn jar wordcount Apache Hadoop 3

89.208.221.132:7180 YARN (MR2 Included) - Cloudera Manager

cloudera MANAGER Clusters Hosts Diagnostics Audits Charts Administration

YARN (MR2 Included) (GeekBrain Cluster) Actions 30 minutes preceding May 28, 5:22 PM UTC

Status Instances Configuration Commands Applications Resource Pools Charts Library Audits Web UI Quick Links

Search for YARN applications, e.g. 'pool = default' or press space to start typeahead. Search Suggestions 30m 1h 2h 6h 12h 1d 7d 30d

**Workload Summary**  
(For Completed Applications)  
ADL Bytes Read  
ADL Bytes Written  
Allocated Memory Seconds  
Allocated VCore Seconds  
CPU Time  
Duration  
File Bytes Read

Results Charts

05/28/2020 5:22 PM word count  
5%  
15.48s  
ID: application\_1589223973966\_0035  
User: hdfs  
Allocated Memory Seconds: 11.7K  
Type: MAPREDUCE  
Pool: root.default  
Allocated VCore Seconds: 11  
Duration: 15.48s

Collect Diagnostic Data Export Select Attributes  
Map: 0% Reduce: 0%

Feedback

20:22  
28.05.2020

Посмотрим что внутри файла

```
student4_10@manager~  
[student4_10@manager ~]$ hdfs dfs -cat /tmp/output2/  
  
option 2  
particular 2  
path 1  
port 1  
retransmitting 1  
support 1  
the 24  
timeout 1  
which 2  
would 1  
wsize=32k 1  
"Mount_point" 1  
32k 1  
Access 1  
Ignore 1  
Lock=True 1  
Mfsvers=4 1  
Protocol 3  
Retrans=2 1  
Soft=False 1  
Wsize=8k 1  
acregmin=30 2  
amount 1  
background 2  
cached 4  
case 1  
coherence 1  
fail 2  
from 2  
generating 1  
into 1  
is 4  
krbSp 1  
minimum 2  
mountport 1  
mounts 4  
need 1  
point. 1  
rsize=32k 1  
suggested) 1  
supported 1  
tenths 1  
values 1  
[student4_10@manager ~]$
```

3. [Факультативное, для тех кто знает JAVA] Собрать программу для MR на Java и запустить ее. Wordcount будет вполне достаточен.

С java не знаком, по этому не получилось.

5++

найдем java файл Hadoop-streaming.jar

```
student4_10@manager:~$ find /opt -name "hadoop-streaming.jar"
/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/hadoop-mapreduce/hadoop-streaming.jar
/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/oozie/oozie-sharelib-mr1/lib/mapreduce-streaming/hadoop-streaming.jar
/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/oozie/oozie-sharelib-yarn/lib/mapreduce-streaming/hadoop-streaming.jar
student4_10@manager ~]$
```

создаём переменную

0

```
student4_10@manager:~$ find /opt -name "hadoop-streaming.jar"
/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/hadoop-mapreduce/hadoop-streaming.jar
/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/oozie/oozie-sharelib-mr1/lib/mapreduce-streaming/hadoop-streaming.jar
/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/oozie/oozie-sharelib-yarn/lib/mapreduce-streaming/hadoop-streaming.jar
student4_10@manager ~$ export YARN_EXAMPLES=/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/hadoop-mapreduce
```

export YARN\_EXAMPLES=/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/hadoop-mapreduce

## Переносим с П.К. файлы mapper.py и reducer.py в students/student4\_10

The screenshot shows the WinSCP interface with two panels. The left panel displays the local file system at `C:\Users\User\PycharmProjects\untitled2\venv\`, and the right panel displays the remote file system at `/students/student4_10/`. Both panels show a table of files and folders.

Имя	Размер	Тип	Изменено
..		Родительский кат...	28.05.2020 23:23:48
Include		Папка с файлами	26.05.2020 15:33:25
Lib		Папка с файлами	26.05.2020 15:33:25
Scripts		Папка с файлами	26.05.2020 15:33:48
mapper.py	1 KB	Python File	28.05.2020 23:23:17
pyvenv.cfg	1 KB	Файл "CFG"	26.05.2020 15:33:25
reducer.py	2 KB	Python File	28.05.2020 23:23:48

Имя	Размер	Изменено	Права	Владел...
..		28.05.2020 20:44:30	rw-rw-rw-r	root
mapper.py	1 KB	28.05.2020 23:23:17	rw-rw-r--	studen...
reducer.py	2 KB	28.05.2020 23:23:48	rw-rw-r--	studen...

At the bottom of the interface, the status bar shows `0 В из 1,69 KB в 0 из 6` for the local view and `0 В из 1,61 KB в 0 из 2` for the remote view. The taskbar at the bottom indicates the time is 19:10 on 29.05.2020.

## Проверяем локальную папку на наличие файлов

The screenshot shows a terminal window with the prompt `student4_10@manager:~`. The user has entered the command `ls /home/student4_10/`, and the output shows the files `mapper.py` and `reducer.py`. The prompt then changes to `[student4_10@manager ~]$`.

```
[student4_10@manager ~]$ ls /home/student4_10/
mapper.py  reducer.py
[student4_10@manager ~]$
```

The taskbar at the bottom indicates the time is 19:16 on 29.05.2020.

Проверяем внутренность файлов

## Reducer.py

```
student4_10@manager:~$ ls /home/student4_10/
mapper.py  reducer.py
[student4_10@manager ~]$ cat /home/student4_10/reducer.py
#!/usr/bin/env python
"""reducer.py"""

from operator import itemgetter
import sys

current_word = None
current_count = 0
word = None

# input comes from STDIN
for line in sys.stdin:
    # remove leading and trailing whitespace
    line = line.strip()

    # parse the input we got from mapper.py
    word, count = line.split('\t', 1)

    # convert count (currently a string) to int
    try:
        count = int(count)
    except ValueError:
        # count was not a number, so silently
        # ignore/discard this line
        continue

    # this IF-switch only works because Hadoop sorts map output
    # by key (here: word) before it is passed to the reducer
    if current_word == word:
        current_count += count
    else:
        if current_word:
            # write result to STDOUT
            print '%s\t%s' % (current_word, current_count)
            current_count = count
            current_word = word

# do not forget to output the last word if needed!
if current_word == word:
    print '%s\t%s' % (current_word, current_count)[student4_10@manager ~]$
```

## Mapper.py

```
student4_10@manager:~$ cat /home/student4_10/mapper.py
#!/usr/bin/env python
"""mapper.py"""

import sys

# input comes from STDIN (standard input)
for line in sys.stdin:
    # remove leading and trailing whitespace
    line = line.strip()
    # split the line into words
    words = line.split()
    # increase counters
    for word in words:
        # write the results to STDOUT (standard output);
        # what we output here will be the input for the
        # Reduce step, i.e. the input for reducer.py
        #
        # tab-delimited; the trivial word count is 1
        print '%s\t%s' % (word, 1)[student4_10@manager ~]$
```



Add permissions x (execution) and check permissions

Добавляем разрешение с правом на x (execution) и проверяем на наличии этих прав

```
student4_10@manager:~  
[student4_10@manager ~]$ ls /home/student4_10/  
mapper.py  reducer.py  
[student4_10@manager ~]$ chmod +x /home/student4_10/mapper.py  
[student4_10@manager ~]$ chmod +x /home/student4_10/reducer.py  
[student4_10@manager ~]$ ls -l  
total 8  
-rwxrwxr-x. 1 student4_10 student4_10 570 May 28 20:23 mapper.py  
-rwxrwxr-x. 1 student4_10 student4_10 1082 May 28 20:23 reducer.py  
[student4_10@manager ~]$
```

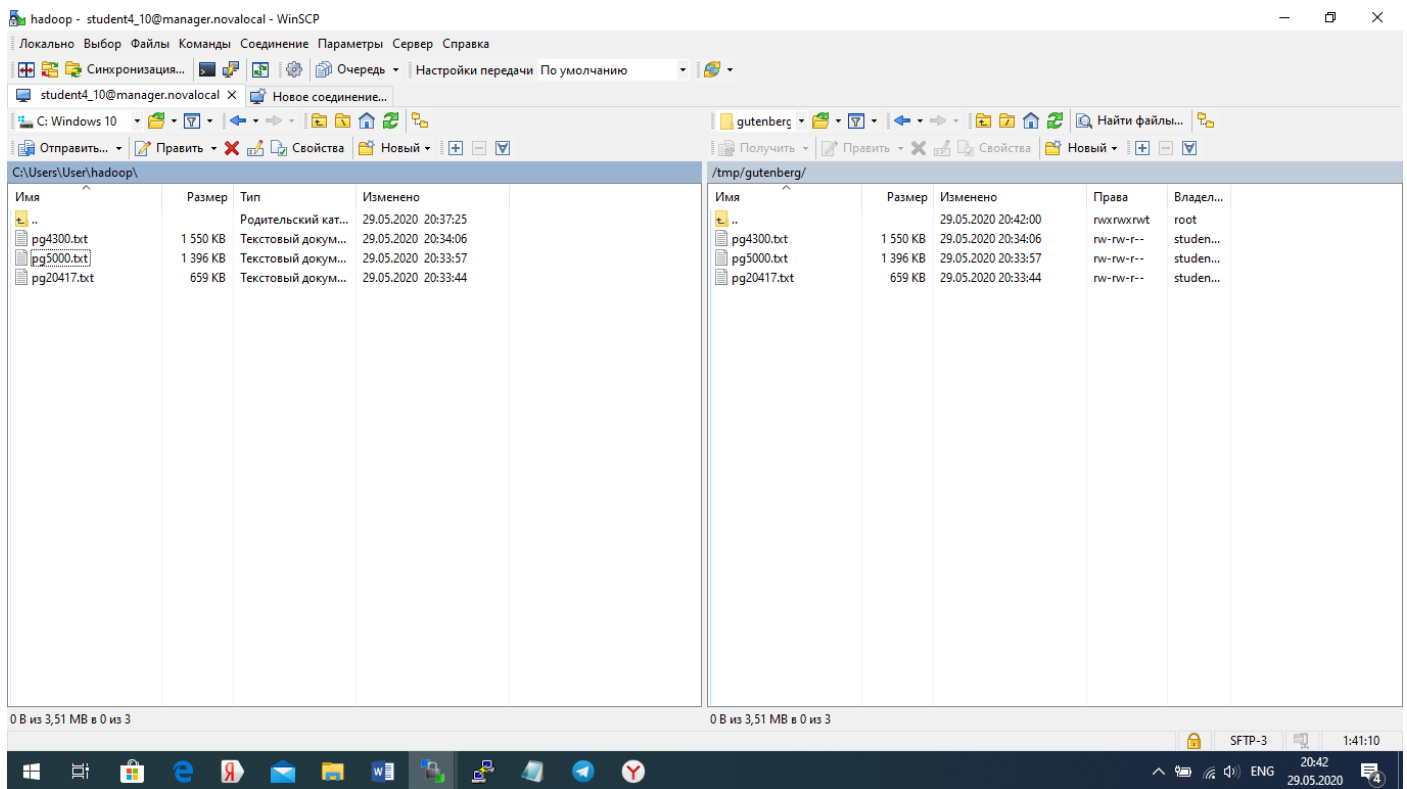
Вопрос

Почему мы надо запускать mapper и reducer с LOCAL а не с HDFS как например текстовые файлы e-books

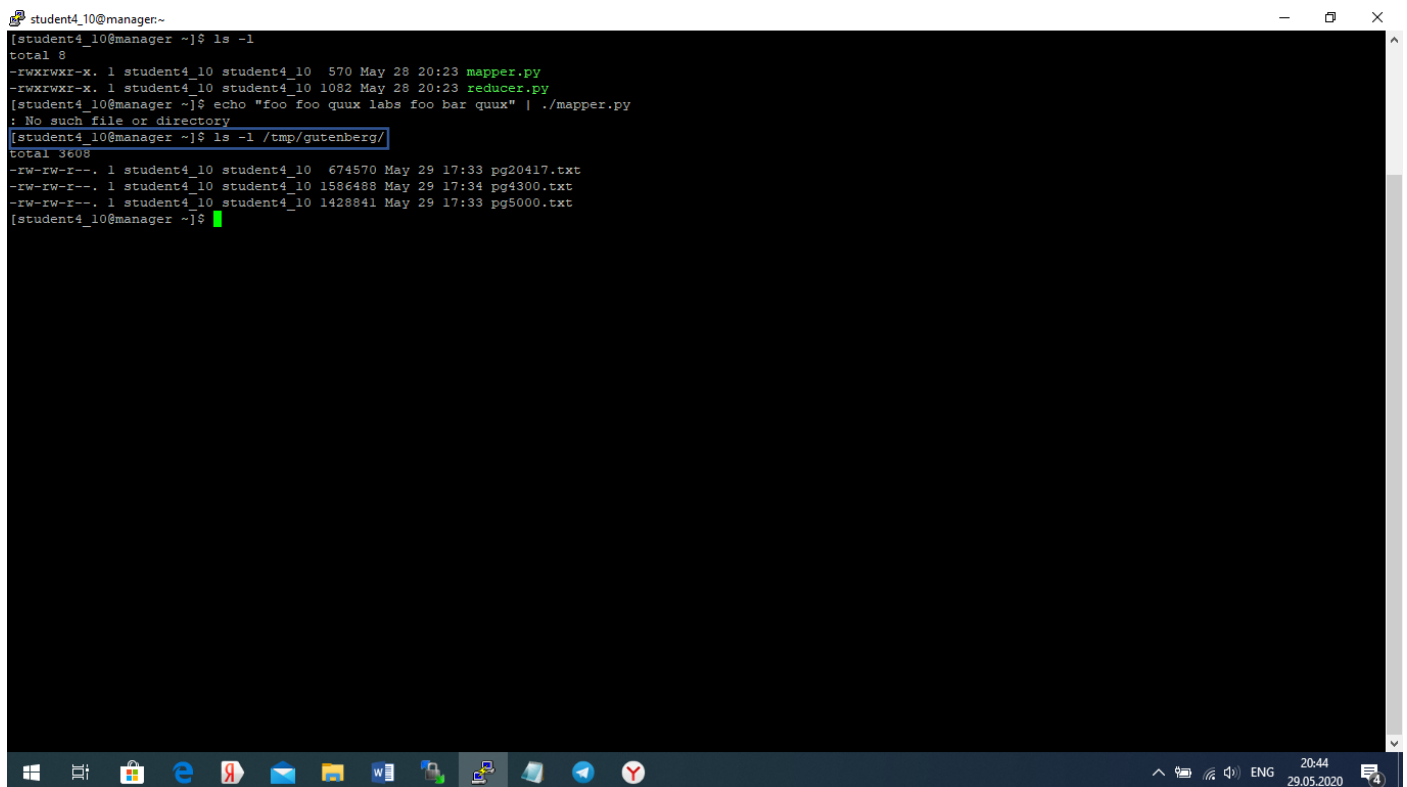
```
student4_10@manager:~  
[student4_10@manager ~]$ ls -l  
total 8  
-rwxrwxr-x. 1 student4_10 student4_10 570 May 28 20:23 mapper.py  
-rwxrwxr-x. 1 student4_10 student4_10 1082 May 28 20:23 reducer.py  
[student4_10@manager ~]$ echo "foo foo quux labs foo bar quux" | ./mapper.py  
: No such file or directory  
[student4_10@manager ~]$
```

Вопрос

Не работает ECHO ?



скачал e-books files и переносим эти text files в tmp/Gutenberg



проверяем текс файлы в putty

```
student4_10@manager:~$ ls /tmp/gutenberg/
pg20417.txt pg4300.txt pg5000.txt
[student4_10@manager ~]$ dfs -copyFromLocal /tmp/gutenberg /student4_10/gutenberg
-bash: dfs: command not found
[student4_10@manager ~]$ hdfs dfs -copyFromLocal /tmp/gutenberg /student4_10/gutenberg
[student4_10@manager ~]$
```

CopyFromLocal Text files into HDFS

#### Question

- Почему в примере Американцы не пишут HDFS
- Почему mapper.py и reducer.py мы оставили в локальной папке а вот e-books files перенесли в HDFS, почему не перенести все в HDFS или же оставить на локальном уровне.

```
student4_10@manager:~$ hdfs dfs -ls /student4_10/gutenberg/
Found 3 items
-rw-r--r--  3 student4_10 supergroup    674570 2020-05-29 17:53 /student4_10/gutenberg/pg20417.txt
-rw-r--r--  3 student4_10 supergroup   1586488 2020-05-29 17:53 /student4_10/gutenberg/pg4300.txt
-rw-r--r--  3 student4_10 supergroup   1428841 2020-05-29 17:53 /student4_10/gutenberg/pg5000.txt
[student4_10@manager ~]$
```

проверяем что все есть в HDFS

```
student4_10@manager:~$ hdfs dfs -ls /student4_10/gutenberg/
^[[A^[[A^[[BFound 3 items
-rw-r--r-- 3 student4_10 supergroup 674570 2020-05-29 17:53 /student4_10/gutenberg/pg20417.txt
-rw-r--r-- 3 student4_10 supergroup 1586488 2020-05-29 17:53 /student4_10/gutenberg/pg4300.txt
-rw-r--r-- 3 student4_10 supergroup 1428841 2020-05-29 17:53 /student4_10/gutenberg/pg5000.txt
[student4_10@manager ~]$ ls /home/student4_10/
mapper.py reducer.py
[student4_10@manager ~]$ hadoop jar $YARN_EXAMPLES/hadoop-streaming.jar \
> -file /home/student4_10/mapper.py -mapper /home/student4_10/mapper.py \
> -file /home/student4_10/reducer.py -reducer /home/student4_10/reducer.py \
> -input /student4_10/gutenberg/* -output /student4_10/gutenberg-output
```

Запускаем jar файл

## Question

- Почему ошибка

```
student4_10@manager:~$ hdfs dfs -ls /student4_10/gutenberg/
^[[A^[[A^[[BFound 3 items
-rw-r--r-- 3 student4_10 supergroup 674570 2020-05-29 17:53 /student4_10/gutenberg/pg20417.txt
-rw-r--r-- 3 student4_10 supergroup 1586488 2020-05-29 17:53 /student4_10/gutenberg/pg4300.txt
-rw-r--r-- 3 student4_10 supergroup 1428841 2020-05-29 17:53 /student4_10/gutenberg/pg5000.txt
[student4_10@manager ~]$ ls /home/student4_10/
mapper.py reducer.py
[student4_10@manager ~]$ hadoop jar $YARN_EXAMPLES/hadoop-streaming.jar \
> -file /home/student4_10/mapper.py -mapper /home/student4_10/mapper.py \
> -file /home/student4_10/reducer.py -reducer /home/student4_10/reducer.py \
> -input /student4_10/gutenberg/* -output /student4_10/gutenberg-output
20/05/29 18:24:17 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/home/student4_10/mapper.py, /home/student4_10/reducer.py] [/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/jars/hadoop-streaming-2.6.0-cdh5.16.2.jar] /tmp/streamjob3011495228515576477.jar tmpDir=null
20/05/29 18:24:18 INFO client.RMProxy: Connecting to ResourceManager at manager.novalocal/89.208.221.132:8032
20/05/29 18:24:19 INFO client.RMProxy: Connecting to ResourceManager at manager.novalocal/89.208.221.132:8032
20/05/29 18:24:23 INFO mapred.FileInputFormat: Total input paths to process : 3
20/05/29 18:24:25 INFO mapreduce.JobSubmitter: number of splits:3
20/05/29 18:24:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1589223973966_0036
20/05/29 18:24:29 INFO impl.YarnClientImpl: Submitted application application_1589223973966_0036
20/05/29 18:24:29 INFO mapreduce.Job: The url to track the job: http://manager.novalocal:8088/proxy/application_1589223973966_0036/
20/05/29 18:24:29 INFO mapreduce.Job: Running job: job_1589223973966_0036
```

YARN (MR2 Included) - Cloudera Manager

cloudra MANAGER

Clusters Hosts Diagnostics Audits Charts Administration

YARN (MR2 Included) (GeekBrain Cluster) Actions

30 minutes preceding May 29, 6:24 PM UTC

Status Instances Configuration Commands Applications Resource Pools Charts Library Audits Web UI Quick Links

Search for YARN applications, e.g. 'pool = default' or press space to start typeahead. Search Suggestions 30m 1h 2h 6h 12h 1d 7d 30d

Workload Summary

(For Completed Applications)

ADL Bytes Read

ADL Bytes Written

Allocated Memory Seconds

Allocated VCore Seconds

CPU Time

Duration

File Bytes Read

Results Charts

Collect Diagnostic Data Export Select Attributes

05/29/2020 6:24 PM streamjob3011495228515576477.jar

5% 31.26s ID: application\_1589223973966\_0036 Type: MAPREDUCE

User: student4\_10 Pool: root.default Duration: 31.26s

Allocated Memory Seconds: 50.1K Allocated VCore Seconds: 46

Map: 0% Reduce: 0%

пишет что была ошибка

```
student4_10@manager:~
mapper.py reducer.py
[student4_10@manager ~]$ hadoop jar $YARN_EXAMPLES/hadoop-streaming.jar \
> -file /home/student4_10/mapper.py -mapper /home/student4_10/mapper.py \
> -file /home/student4_10/reducer.py -reducer /home/student4_10/reducer.py \
> -input /student4_10/gutenberg/* -output /student4_10/gutenberg-output
20/05/29 18:24:17 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/home/student4_10/mapper.py, /home/student4_10/reducer.py] [/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/jars/hadoop-streaming-2.6.0-cdh5.16.2.jar] /tmp/streamjob3011495228515576477.jar tmpDir=null
20/05/29 18:24:18 INFO client.RMPProxy: Connecting to ResourceManager at manager.novalocal/89.208.221.132:8032
20/05/29 18:24:19 INFO client.RMPProxy: Connecting to ResourceManager at manager.novalocal/89.208.221.132:8032
20/05/29 18:24:23 INFO mapred.FileInputFormat: Total input paths to process : 3
20/05/29 18:24:25 INFO mapreduce.JobSubmitter: number of splits:3
20/05/29 18:24:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1589223973966_0036
20/05/29 18:24:29 INFO impl.YarnClientImpl: Submitted application application_1589223973966_0036
20/05/29 18:24:29 INFO mapreduce.Job: The url to track the job: http://manager.novalocal:8088/proxy/application_1589223973966_0036/
20/05/29 18:24:29 INFO mapreduce.Job: Running job: job_1589223973966_0036
20/05/29 18:24:48 INFO mapreduce.Job: Job job_1589223973966_0036 running in uber mode : false
20/05/29 18:24:48 INFO mapreduce.Job: map 0% reduce 0%
20/05/29 18:24:54 INFO mapreduce.Job: Task Id : attempt_1589223973966_0036_m_000001_0, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
    at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)
    at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.mapred.YarnChild.doAs(UserGroupInformation.java:1924)
    at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)
20/05/29 18:24:55 INFO mapreduce.Job: Task Id : attempt_1589223973966_0036_m_000000_0, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
```

```
20/05/29 18:24:55 INFO mapreduce.Job: Task Id : attempt_1589223973966_0036_m_000000_0, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
    at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)
    at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1924)
    at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)

20/05/29 18:24:57 INFO mapreduce.Job: Task Id : attempt_1589223973966_0036_m_000002_0, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
    at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)
    at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1924)
    at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)

20/05/29 18:25:00 INFO mapreduce.Job: Task Id : attempt_1589223973966_0036_m_000001_1, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
    at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
```

```
20/05/29 18:25:01 INFO mapreduce.Job: Task Id : attempt_1589223973966_0036_m_000000_1, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
    at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)
    at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1924)
    at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)

20/05/29 18:25:04 INFO mapreduce.Job: Task Id : attempt_1589223973966_0036_m_000002_1, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
    at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)
    at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1924)
    at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)

20/05/29 18:25:06 INFO mapreduce.Job: Task Id : attempt_1589223973966_0036_m_000001_2, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
    at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)
```



```
student4_10@manager:~
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 127
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
    at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
    at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)
    at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1924)
    at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)

20/05/29 18:25:13 INFO mapreduce.Job: map 100% reduce 100%
20/05/29 18:25:15 INFO mapreduce.Job: Job job_1589223973966_0036 failed with state FAILED due to: Task failed task_1589223973966_0036_m_000001
Job failed as tasks failed. failedMaps:1 failedReduces:0

20/05/29 18:25:15 INFO mapreduce.Job: Counters: 14
  Job Counters
    Failed map tasks=10
    Killed map tasks=2
    Killed reduce tasks=6
    Launched map tasks=11
    Other local map tasks=8
    Data-local map tasks=3
    Total time spent by all maps in occupied slots (ms)=34842
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=34842
    Total vcore-milliseconds taken by all map tasks=34842
    Total megabyte-milliseconds taken by all map tasks=35678208
  Map-Reduce Framework
    CPU time spent (ms)=0
    Physical memory (bytes) snapshot=0
    Virtual memory (bytes) snapshot=0
20/05/29 18:25:15 ERROR streaming.StreamJob: Job not successful!
Streaming Command Failed!
[student4_10@manager ~]$
```

Ничего не получилось ошибка !!!

```
student4_10@manager:~
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1924)
    at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)

20/05/29 18:25:13 INFO mapreduce.Job: map 100% reduce 100%
20/05/29 18:25:15 INFO mapreduce.Job: Job job_1589223973966_0036 failed with state FAILED due to: Task failed task_1589223973966_0036_m_000001
Job failed as tasks failed. failedMaps:1 failedReduces:0

20/05/29 18:25:15 INFO mapreduce.Job: Counters: 14
  Job Counters
    Failed map tasks=10
    Killed map tasks=2
    Killed reduce tasks=6
    Launched map tasks=11
    Other local map tasks=8
    Data-local map tasks=3
    Total time spent by all maps in occupied slots (ms)=34842
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=34842
    Total vcore-milliseconds taken by all map tasks=34842
    Total megabyte-milliseconds taken by all map tasks=35678208
  Map-Reduce Framework
    CPU time spent (ms)=0
    Physical memory (bytes) snapshot=0
    Virtual memory (bytes) snapshot=0
20/05/29 18:25:15 ERROR streaming.StreamJob: Job not successful!
Streaming Command Failed!
[student4_10@manager ~]$
[student4_10@manager ~]$ hdfs dfs -ls /student4_10/
Found 6 items
drwxr-xr-x  - student4_10 supergroup          0 2020-05-25 21:52 /student4_10/Malik_Moroz
drwxr-xr-x  - student4_10 supergroup          0 2020-05-29 17:53 /student4_10/gutenberg
drwxr-xr-x  - student4_10 supergroup          0 2020-05-29 18:25 /student4_10/gutenberg-output
-rw-r--r--  3 hdfs      supergroup          0 2020-05-28 16:29 /student4_10/hdfs_user
drwxr-xr-x  - hdfs      supergroup          0 2020-05-28 17:48 /student4_10/java
-rw-r--r--  6 student4_10 supergroup          0 2020-05-25 20:47 /student4_10/myfile.txt
[student4_10@manager ~]$ hdfs dfs -ls /student4_10/gutenberg-output/
[student4_10@manager ~]$ hdfs dfs -cat /student4_10/gutenberg-output/*
```

проверяем что получили

```
student4_10@manager~  
20/05/29 18:25:13 INFO mapreduce.Job: map 100% reduce 100%  
20/05/29 18:25:15 INFO mapreduce.Job: Job job_1589223973966_0036 failed with state FAILED due to: Task failed task_1589223973966_0036_m_000001  
Job failed as tasks failed. failedMaps:1 failedReduces:0  
  
20/05/29 18:25:15 INFO mapreduce.Job: Counters: 14  
Job Counters  
    Failed map tasks=10  
    Killed map tasks=2  
    Killed reduce tasks=6  
    Launched map tasks=11  
    Other local map tasks=8  
    Data-local map tasks=3  
    Total time spent by all maps in occupied slots (ms)=34842  
    Total time spent by all reduces in occupied slots (ms)=0  
    Total time spent by all map tasks (ms)=34842  
    Total vcore-milliseconds taken by all map tasks=34842  
    Total megabyte-milliseconds taken by all map tasks=35678208  
Map-Reduce Framework  
    CPU time spent (ms)=0  
    Physical memory (bytes) snapshot=0  
    Virtual memory (bytes) snapshot=0  
20/05/29 18:25:15 ERROR streaming.StreamJob: Job not successful!  
Streaming Command Failed!  
[student4_10@manager ~]$  
[student4_10@manager ~]$ hdfs dfs -ls /student4_10/  
Found 6 items  
drwxr-xr-x - student4_10 supergroup      0 2020-05-25 21:52 /student4_10/Malik_Moroz  
drwxr-xr-x - student4_10 supergroup      0 2020-05-29 17:53 /student4_10/gutenberg  
drwxr-xr-x - student4_10 supergroup      0 2020-05-29 18:25 /student4_10/gutenberg-output  
-rw-r--r--  3 hdfs      supergroup      0 2020-05-28 16:29 /student4_10/hdfs_user  
drwxr-xr-x - hdfs      supergroup      0 2020-05-28 17:48 /student4_10/java  
-rw-r--r--  6 student4_10 supergroup      0 2020-05-25 20:47 /student4_10/myfile.txt  
[student4_10@manager ~]$ hdfs dfs -ls /student4_10/gutenberg-output/  
[student4_10@manager ~]$ hdfs dfs -cat /student4_10/gutenberg-output/*  
cat: `/student4_10/gutenberg-output/*': No such file or directory  
[student4_10@manager ~]$ hdfs dfs -cat /student4_10/gutenberg-output  
cat: `/student4_10/gutenberg-output': Is a directory  
[student4_10@manager ~]$
```

ничего нету