

ssh student3_10@37.139.32.56

ssh student3_10@37.139.32.56 -D localhost:8080

ssh -i /home/igor/Зарпызки/id_rsa_student3_10 student3_10@37.139.32.56 -D localhost:8080

yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples.jar

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка

timelineserver      run the timeline server
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples.jar
An example program must be given as the first argument.
Valid program names are:
  aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
  aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
  bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
  dbcount: An example job that count the pageview counts from a database.
  distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
  grep: A map/reduce program that counts the matches of a regex in the input.
  join: A job that effects a join over sorted, equally partitioned datasets
  multifiawc: A job that counts words from several files.
  pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
  pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
  randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
  randomwriter: A map/reduce program that writes 10GB of random data per node.
  secondarysort: An example defining a secondary sort to the reduce.
  sort: A map/reduce program that sorts the data written by the random writer.
  sudoku: A sudoku solver.
  teragen: Generate data for the terasort
  terasort: Run the terasort
  teravalidate: Checking results of terasort
  wordcount: A map/reduce program that counts the words in the input files.
  wordmean: A map/reduce program that counts the average length of the words in the input files.
  wordmedian: A map/reduce program that counts the median length of the words in the input files.
  wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка

wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples.
jar pi
Usage: org.apache.hadoop.examples.QuasiMonteCarlo <nMaps> <nSamples>
Generic options supported are:
  -conf <configuration file>      specify an application configuration file
  -D <property=value>              define a value for a given property
  -fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
  -jt <local|resourcemanager:port> specify a ResourceManager
  -files <file1,...>               specify a comma-separated list of files to be copied to the map reduce cluster
  -libjars <jar1,...>              specify a comma-separated list of jar files to be included in the classpath
  -archives <archive1,...>         specify a comma-separated list of archives to be unarchived on the compute machines


The general command line syntax is:
command [genericOptions] [commandOptions]
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Application application_161790284451_0134 — Mozilla Firefox

Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~

Application application_161790284451_0134

10.0.0.7:8088/cluster/app/application_1617990284451_0134



Application application_1617990284451_0134

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Kill Application

User: student3_10

Name: QuasiMonteCarlo

Application Type: MAPREDUCE

Application Tags:

Application Priority: 0 (Higher Integer value indicates higher priority)

YarnApplicationState: ACCEPTED: waiting for AM container to be allocated, launched and register with RM

Queue: default

FinalStatus Reported by AM: Application has not completed yet.

Started: Sun Apr 11 17:07:23 +0000 2021

Elapsed: 6mins, 7sec

Tracking URL: ApplicationMaster

Log Aggregation Status: NOT_START

Application Timeout (Remaining Time): Unlimited

Diagnostics: [Sun Apr 11 17:11:25 +0000 2021] Application is added to the scheduler and is not yet started. Resource Request = <memory:2048, vCores:1>; Queue Resource Limit for AM = <memory:9216, vCores:8>;

Unmanaged Application: false

Application Node Label expression: <Not set>

AM container Node Label expression: <DEFAULT_PARTITION>

Total Number of AM Containers

Total Number of Non-AM Containers

Aggregated Resource Usage

Show 20 entries

Attempt ID	Started	Node
appattempt_1617990284451_0134_000001	Sun Apr 11 20:07:23 +0300 2021	http://

Showing 1 to 1 of 1 entries

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка

Total time spent by all reduce tasks (ms)=12977
Total vcore-milliseconds taken by all map tasks=136363
Total vcore-milliseconds taken by all reduce tasks=12977
Total megabyte-milliseconds taken by all map tasks=279271424
Total megabyte-milliseconds taken by all reduce tasks=26576896
Map-Reduce Framework
  Map input records=32
  Map output records=64
  Map output bytes=576
  Map output materialized bytes=896
  Input split bytes=6070
  Combine input records=0
  Combine output records=0
  Reduce input groups=2
  Reduce shuffle bytes=896
  Reduce input records=64
  Reduce output records=0
  Spilled Records=128
  Shuffled Maps =32
  Failed Shuffles=0
  Merged Map outputs=32
  GC time elapsed (ms)=3020
  CPU time spent (ms)=60490
  Physical memory (bytes) snapshot=45430976512
  Virtual memory (bytes) snapshot=123505754112
  Total committed heap usage (bytes)=42563796992
  Peak Map Physical memory (bytes)=1427566592
  Peak Map Virtual memory (bytes)=3694972928
  Peak Reduce Physical memory (bytes)=211456000
  Peak Reduce Virtual memory (bytes)=5504983040
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=3776
File Output Format Counters
  Bytes Written=97
Job Finished in 615.743 seconds
Estimated value of Pi is 3.141475000000000000000000
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

hadoop

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	Memory Us
136	0	7	129	21	21 GB

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
6	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[memory-mb (unit=Mb), vcores]	<memory:512, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	Final
application_1617990284451_0135	zeppelin	Zeppelin	SPARK	default	0	Sun Apr 11 20:07:55 +0300 2021	N/A	RUNNING	UNDE
application_1617990284451_0134	student3_10	QuasiMonteCarlo	MAPREDUCE	default	0	Sun Apr 11 20:07:23 +0300 2021	Sun Apr 11 20:17:35 +0300 2021	FINISHED	SUCC
application_1617990284451_0133	student21_11	HIVE-4cafbf59-6466-4860-89cc-8c467227d8bf	TEZ	default	0	Sun Apr 11 20:03:32 +0300 2021	Sun Apr 11 20:17:03 +0300 2021	FINISHED	SUCC

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7888:~/BigData. Введение в экосистему Hadoop/3. YARN & MR$ scp license student3_10@37.139.32.56:.
license
100% 4590 432.2KB/s 00:00
igor@igor-MS-7888:~/BigData. Введение в экосистему Hadoop/3. YARN & MR$ scp Mapper.py student3_10@37.139.32.56:.
Mapper.py
100% 412 62.9KB/s 00:00
igor@igor-MS-7888:~/BigData. Введение в экосистему Hadoop/3. YARN & MR$ ssh student3_10@37.139.32.56 -D localhost:8080
Last login: Sun Apr 11 16:59:14 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
license Mapper.py Reducer.py urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка

[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cat license
All of the documentation and software included in the 4.4BSD and 4.4BSD-Lite Releases is copyrighted by The Regents of the University of California.

Copyright 1979, 1980, 1983, 1986, 1988, 1989, 1991, 1992, 1993, 1994 The Regents of the University of California. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

All advertising materials mentioning features or use of this software must display the following acknowledgement:

This product includes software developed by the University of California, Berkeley and its contributors.

Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

The Institute of Electrical and Electronics Engineers and the American National Standards Committee X3, on Information Processing Systems have given us permission to reprint portions of their documentation.

In the following statement, the phrase "this text" refers to portions of the system documentation.

Portions of this text are reprinted and reproduced in electronic form in the second BSD Networking Software Release, from IEEE Std 1003.1-1988, IEEE Standard Portable Operating System Interface for Computer Environments (POSIX), copyright C 1988 by the Institute of Electrical and Electronics Engineers, Inc. In the event of any discrepancy between these versions and the original IEEE Standard, the original IEEE Standard is the referee document.

In the following statement, the phrase "This material" refers to portions of the system documentation.

This material is reproduced with permission from American National Standards Committee X3, on Information Processing Systems. Computer and Business Equipment Manufacturers Association (CBEMA), 311 First St., NW, Suite 500, Washington, DC 20001-2178. The developmental work of Programming Language C was completed by the X3J11 Technical Committee.

The views and conclusions contained in the software and documentation are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the Regents of the University of California.

NOTE: The copyright of UC Berkeley's Berkeley Software Distribution ("BSD") source has been updated. The copyright addendum may be found at ftp://ftp.cs.berkeley.edu/pub/4bsd/README.Imp.License.Change and is included below.

July 22, 1999

To All Licensees, Distributors of Any Version of BSD:

As you know, certain of the Berkeley Software Distribution ("BSD") source code files require that further distributions of products containing all or portions of the software, acknowledge within their advertising materials that such products contain software developed by UC Berkeley and its contributors.

Specifically, the provision reads:

    * 3. All advertising materials mentioning features or use of this software
    * must display the following acknowledgement:
    *   This product includes software developed by the University of
    *   California, Berkeley and its contributors."

Effective immediately, licensees and distributors are no longer required to include the acknowledgement within advertising materials. Accordingly, the foregoing paragraph of those BSD Unix files containing it is hereby deleted in its entirety.

William Hoskins
Director, Office of Technology Licensing
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
those  1
BSD    1
Unix   1
files  1
containing 1
it     1
is     1
hereby 1
deleted 1
in     1
its    1
entirety. 1
William 1
Hoskins 1
Director, 1
Office 1
of      1
Technology 1
Licensing 1
University 1
of      1
California, 1
Berkeley 1
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

vim Reducer.py
ls

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
with   1
with   1
with   1
within 1
within 1
without 1
without 1
work   1
written 1
X3,    1
X3,    1
X3J11  1
you    1
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ vim Reducer.py
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
license Mapper.py Reducer.py urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```



```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
TO,    1
TORT   1
UC      1
University    1
Unix    1
updated.    1
us      1
use     1
USE,    1
USE     1
used    1
Version 1
versions    1
views      1
WARRANTIES, 1
WARRANTIES 1
was        1
Washington, 1
WAY        1
WHETHER 1
William 1
with     1
within   1
without  1
work     1
written  1
X3,      1
X3J11    1
you      1
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка

As you know, certain of the Berkeley Software Distribution ("BSD") source code f
iles require that further distributions of products containing all or portions o
f the software, acknowledge within their advertising materials that such product
s contain software developed by UC Berkeley and its contributors.

Specifically, the provision reads:

    * 3. All advertising materials mentioning features or use of this software
    * must display the following acknowledgement:
    * This product includes software developed by the University of
    * California, Berkeley and its contributors."
Effective immediately, licensees and distributors are no longer required to incl
ude the acknowledgement within advertising materials. Accordingly, the foregoing
paragraph of those BSD Unix files containing it is hereby deleted in its entire
ty.

William Hoskins
Director, Office of Technology Licensing
University of California, Berkeley[student3_10@bigdataanalytics2-head-shdpt-v31-
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

```
yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar -input test -output result -mapper
"python3.7 Mapper.py" -reducer "python3.7 Reducer.py" -file Mapper.py -file Reducer.py
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка

9288 0028 m 000001
Job failed as tasks failed. failedMaps:1 failedReduces:0 killedMaps:0 killedReduces: 0

21/04/12 06:55:25 INFO mapreduce.Job: Counters: 12
  Job Counters
    Failed map tasks=7
    Killed map tasks=1
    Killed reduce tasks=1
    Launched map tasks=7
    Other local map tasks=5
    Data-local map tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=116692
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=29173
    Total vcore-milliseconds taken by all map tasks=29173
    Total megabyte-milliseconds taken by all map tasks=59746304
21/04/12 06:55:25 ERROR streaming.StreamJob: Job not successful!
Streaming Command Failed!
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar
-input test -output result -mapper "python Mapper.py" -reducer "python Reducer.py" -file Mapper.py -file Reducer.py
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка

0.7:8050
21/04/12 07:06:12 INFO client.AHSPProxy: Connecting to Application History server at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:10200
21/04/12 07:06:12 ERROR streaming.StreamJob: Error Launching job : Output directory hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/result already exists
Streaming Command Failed!
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ yarn jar /usr/hdp/3.1.4.0-315/hadoop-mapreduce-client/hadoop-streaming.jar
-input test -output result -mapper "python Mapper.py" -reducer "python Reducer.py" -file Mapper.py -file Reducer.py
JAR does not exist or is not a normal file: /usr/hdp/3.1.4.0-315/hadoop-mapreduce-client/hadoop-streaming.jar
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar
-input test -output result -mapper "python Mapper.py" -reducer "python Reducer.py" -file Mapper.py -file Reducer.py
21/04/12 07:07:20 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [Mapper.py, Reducer.py] [/usr/hdp/3.1.4.0-315/hadoop-mapreduce/hadoop-streaming-3.1.1.3.1.4.0-315.jar] /var/lib/ambari-agent/tmp/hadoop_java_io_tmpdir/streamjob793783987850635747.jar tmpDir=null
21/04/12 07:07:22 INFO client.RMPProxy: Connecting to ResourceManager at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:8050
21/04/12 07:07:22 INFO client.AHSPProxy: Connecting to Application History server at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:10200
21/04/12 07:07:22 INFO client.RMPProxy: Connecting to ResourceManager at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:8050
21/04/12 07:07:22 INFO client.AHSPProxy: Connecting to Application History server at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:10200
21/04/12 07:07:22 ERROR streaming.StreamJob: Error Launching job : Output directory hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/result already exists
Streaming Command Failed!
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
license Mapper.py Reducer.py urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ head urbandict.csv
```

head -n 1 urbandict.csv


```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
21/04/12 07:07:22 INFO client.RMPProxy: Connecting to ResourceManager at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:8050
21/04/12 07:07:22 INFO client.AHSPProxy: Connecting to Application History server at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:10200
21/04/12 07:07:22 INFO client.RMPProxy: Connecting to ResourceManager at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:8050
21/04/12 07:07:22 INFO client.AHSPProxy: Connecting to Application History server at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:10200
21/04/12 07:07:22 ERROR streaming.StreamJob: Error Launching job : Output directory hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/result already exists
Streaming Command Failed!
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
license Mapper.py Reducer.py urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ head urbandict.csv
word_id,word,up_votes,down_votes,author,definition
0000007,Janky,296,255,dc397b2f,"Undesirable; less-than optimum."
0000008,slumpin',16,37,dc397b2f,"low down and funky, but [knee deep] enough to ride to."
0000009,yayeeyay,19,27,dc397b2f,"affirmation; suggestion of encouragement, approval, or interest."
0000012,hard-core,162,96,d1610749,"anything out of our league that can be good or bad."
0000013,brutal,12,45,40ecelef,"anything that makes you sweat"
0000014,skanky,9,48,485e4db7,"Anything of or pertaining to a $10,000 hooker."
0000015,ho-bag,26,27,b37fba05,"a term of endearment, used affectionately for your roommate. First used in the schools' parking lot after an incident with the hall monitors."
0000016,massive,36,45,b9dcf126,"really really good. excellently good."
0000017,wtf,183,99,a6c97ba3,"what the fuck? ;; use it in place of expletives. a more polite alternative."
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ head -n 1 urbandict.csv
word_id,word,up_votes,down_votes,author,definition
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

head urbandict.csv > urbandict_test.csv
cat vehicles.csv | python Mapper.py
Удалим резалт: hdfs dfs -rm -r result
Просмотр: cat vehicles.csv | python3.7 mapper.py

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
dodge 9295
toyota 15995
nissan 3998
ford 31500
kia 0
kia 0
kia 0
jeep 0
ford 0
chevrolet 0
hyundai 3395
chevrolet 11995
mercedes-benz 15995
ram 3300
honda 5995
chevrolet 5495
chevrolet 14800
audi 7500
6500
7500
rover 41995
ford 6998
bmw 7500
mercedes-benz 1500
hyundai 11900
mercedes-benz 4800
toyota 12995
ford 11900
ram 29513
volkswagen 5500
hyundai 10995
ford 14900
chevrolet 17900
ram 8500
audi 9999
ford 13900
500
12990
chevrolet 9990
chevrolet 14900
chevrolet 4990
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cat vehicles.csv | python3.7 mapper.py
```

cat vehicles.csv | python3.7 mapper.py | sort | python3.7 reducer.py

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0~
Файл  Правка  Вид  Терминал  Вкладки  Справка

File "Reducer.py", line 11
SyntaxError: Non-ASCII character '\xd1' in file Reducer.py on line 11, but no encoding declared; see http://www.python.org/peps/pep-0263.html for details
File "Mapper.py", line 7
SyntaxError: Non-ASCII character '\xd1' in file Mapper.py on line 7, but no encoding declared; see http://www.python.org/peps/pep-0263.html for details
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cat vehicles.csv | python3.7 Mapper.py | sort | python3.7 Reducer.py
Traceback (most recent call last):
  File "Reducer.py", line 18, in <module>
    word, count = line.split('\t', 1)
ValueError: not enough values to unpack (expected 2, got 1)
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ vim mapper.py
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cat vehicles.csv | python3.7 mapper.py | sort | python3.7 reducer.py
acura      12143246
alfa-romeo 894208
aston-martin 915150
audi       27500345
bmw        53041036
buick      9414316
cadillac   24267276
chevrolet  1452593447
chrysler   10874364
datsun     280918
dodge      44587781
ferrari    4808024
fiat       2931321
ford       275376101
gmc        2873750514
harley-davidson 328464
honda      59867277
hyundai    21931470
infiniti   13550397
jaguar     6137143
jeep       69589094
kia        19559407
land rover 94400
lexus      30034843
lincoln    7796232
mazda      14377882
mercedes-benz 65436510
mercury    1162558
mini       7507965
mitsubishi 3672210
morgan     74977
nissan      67049248
pontiac    3417008
porsche    17043208
ram        92764848
rover      9843502
saturn     885346
subaru     30836333
tesla      10943535
toyota     134519824
volkswagen 26791675
volvo      8615208
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

scp mapper.py student3_10@37.139.32.56:.
wc -l vehicles.csv
head -n 100000 vehicles.csv > small.csv
Посмотреть что хранится: ls -alh
Почистить/удалить файлы: rm
Посмотреть сколько строк в файле: wc -l <имя.файла>
wc -l urbandict.csv

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0~
Файл  Правка  Вид  Терминал  Вкладки  Справка

[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
mapper.py Mapper.py reducer.py Reducer.py urbandict.csv vehicles.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ^C
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls -alh
итого 844M
drwx----- 4 student3_10 student3_10 4,0K anp 15 18:24 .
drwxr-xr-x. 218 root root 12K anp 15 07:07 ..
-rw----- 1 student3_10 student3_10 11K anp 15 18:21 .bash_history
-rw-r--r-- 1 student3_10 student3_10 18 aar 8 2019 .bash_logout
-rw-r--r-- 1 student3_10 student3_10 193 aar 8 2019 .bash_profile
-rw-r--r-- 1 student3_10 student3_10 231 aar 8 2019 .bashrc
-rw-rw-r-- 1 student3_10 student3_10 529 anp 13 19:31 mapper.py
-rw-rw-r-- 1 student3_10 student3_10 520 anp 13 19:14 Mapper.py
drwxrwxr-x 2 student3_10 student3_10 4,0K anp 3 18:35 .oracle_jre_usage
-rw-rw-r-- 1 student3_10 student3_10 1,2K anp 13 20:23 reducer.py
-rw-rw-r-- 1 student3_10 student3_10 1,2K anp 13 20:09 Reducer.py
drwx----- 2 student3_10 student3_10 4,0K anp 3 18:40 .ssh
-rw-rw-r-- 1 student3_10 student3_10 514M anp 15 18:25 urbandict.csv
-rw-rw-r-- 1 student3_10 student3_10 331M anp 13 17:56 vehicles.csv
-rw----- 1 student3_10 student3_10 2,2K anp 13 20:09 viminfo
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ wc -l urbandict.csv
2580925 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ^C
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

head -n 100000 vehicles.csv > vehicles.csv
head -n 1290462 urbandict.csv > urbandict_new.csv

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0~  
Файл  Правка  Вид  Терминал  Вкладки  Справка  
-rw-r--r-- 1 student3_10 student3_10 2,2K апр 13 20:09 .viminfo  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ wc -l urbandict.csv  
2580925 urbandict.csv  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ^C  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ head -n 1290462 urbandict.csv > urbandict_new.csv  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls -alh  
итого 1,1G  
drwxr-xr-x. 4 student3_10 student3_10 4,0K апр 15 18:53 .  
drwxr-xr-x. 218 root root 12K апр 15 07:07 ..  
-rw-r--r-- 1 student3_10 student3_10 11K апр 15 18:21 .bash_history  
-rw-r--r-- 1 student3_10 student3_10 18 аэр 8 2019 .bash_logout  
-rw-r--r-- 1 student3_10 student3_10 193 аэр 8 2019 .bash_profile  
-rw-r--r-- 1 student3_10 student3_10 231 аэр 8 2019 .bashrc  
-rw-rw-r-- 1 student3_10 student3_10 529 апр 13 19:31 mapper.py  
-rw-rw-r-- 1 student3_10 student3_10 520 апр 13 19:14 Mapper.py  
drwxrwxr-x 2 student3_10 student3_10 4,0K апр 3 18:35 .oracle_jre_usage  
-rw-rw-r-- 1 student3_10 student3_10 1,2K апр 13 20:23 reducer.py  
-rw-rw-r-- 1 student3_10 student3_10 1,2K апр 13 20:09 Reducer.py  
drwxr-xr-x 2 student3_10 student3_10 4,0K апр 3 18:40 .ssh  
-rw-rw-r-- 1 student3_10 student3_10 514M апр 15 18:25 urbandict.csv  
-rw-rw-r-- 1 student3_10 student3_10 246M апр 15 18:53 urbandict_new.csv  
-rw-rw-r-- 1 student3_10 student3_10 331M апр 13 17:56 vehicles.csv  
-rw-r--r-- 1 student3_10 student3_10 2,2K апр 13 20:09 .viminfo  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

rm urbandict.csv

mv small.csv viholes.csv

Переименовываю файл

mv urbandict_new.csv urbandict.csv

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0~  
Файл  Правка  Вид  Терминал  Вкладки  Справка  
-rw-rw-r-- 1 student3_10 student3_10 1,2K апр 13 20:09 Reducer.py  
drwxr-xr-x. 2 student3_10 student3_10 4,0K апр 3 18:40 .ssh  
-rw-rw-r-- 1 student3_10 student3_10 246M апр 15 18:53 urbandict_new.csv  
-rw-rw-r-- 1 student3_10 student3_10 331M апр 13 17:56 vehicles.csv  
-rw-r--r-- 1 student3_10 student3_10 2,2K апр 13 20:09 .viminfo  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ mv urbandict_new.csv urbandict.csv  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls -alh  
итого 576M  
drwxr-xr-x. 4 student3_10 student3_10 4,0K апр 15 19:03 .  
drwxr-xr-x. 218 root root 12K апр 15 07:07 ..  
-rw-r--r-- 1 student3_10 student3_10 11K апр 15 18:21 .bash_history  
-rw-r--r-- 1 student3_10 student3_10 18 аэр 8 2019 .bash_logout  
-rw-r--r-- 1 student3_10 student3_10 193 аэр 8 2019 .bash_profile  
-rw-r--r-- 1 student3_10 student3_10 231 аэр 8 2019 .bashrc  
-rw-rw-r-- 1 student3_10 student3_10 529 апр 13 19:31 mapper.py  
-rw-rw-r-- 1 student3_10 student3_10 520 апр 13 19:14 Mapper.py  
drwxrwxr-x 2 student3_10 student3_10 4,0K апр 3 18:35 .oracle_jre_usage  
-rw-rw-r-- 1 student3_10 student3_10 1,2K апр 13 20:23 reducer.py  
-rw-rw-r-- 1 student3_10 student3_10 1,2K апр 13 20:09 Reducer.py  
drwxr-xr-x 2 student3_10 student3_10 4,0K апр 3 18:40 .ssh  
-rw-rw-r-- 1 student3_10 student3_10 246M апр 15 18:53 urbandict.csv  
-rw-rw-r-- 1 student3_10 student3_10 331M апр 13 17:56 vehicles.csv  
-rw-r--r-- 1 student3_10 student3_10 2,2K апр 13 20:09 .viminfo  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Загружаю: hdfs dfs -put vehicles.csv .

Загружаю: hdfs dfs -put urbandict.csv .

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0~  
Файл  Правка  Вид  Терминал  Вкладки  Справка  
drwxr-xr-x. 4 student3_10 student3_10 4,0K апр 15 19:03 .  
drwxr-xr-x. 218 root root 12K апр 15 07:07 ..  
-rw-r--r-- 1 student3_10 student3_10 11K апр 15 18:21 .bash_history  
-rw-r--r-- 1 student3_10 student3_10 18 аэр 8 2019 .bash_logout  
-rw-r--r-- 1 student3_10 student3_10 193 аэр 8 2019 .bash_profile  
-rw-r--r-- 1 student3_10 student3_10 231 аэр 8 2019 .bashrc  
-rw-rw-r-- 1 student3_10 student3_10 529 апр 13 19:31 mapper.py  
-rw-rw-r-- 1 student3_10 student3_10 520 апр 13 19:14 Mapper.py  
drwxrwxr-x 2 student3_10 student3_10 4,0K апр 3 18:35 .oracle_jre_usage  
-rw-rw-r-- 1 student3_10 student3_10 1,2K апр 13 20:23 reducer.py  
-rw-rw-r-- 1 student3_10 student3_10 1,2K апр 13 20:09 Reducer.py  
drwxr-xr-x 2 student3_10 student3_10 4,0K апр 3 18:40 .ssh  
-rw-rw-r-- 1 student3_10 student3_10 246M апр 15 18:53 urbandict.csv  
-rw-rw-r-- 1 student3_10 student3_10 331M апр 13 17:56 vehicles.csv  
-rw-r--r-- 1 student3_10 student3_10 2,2K апр 13 20:09 .viminfo  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -put urbandict.csv .  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls  
Found 5 items  
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 18:00 .Trash  
drwxr-xr-x - student3_10 student3_10 0 2021-04-13 20:35 .staging  
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 18:20 lesson  
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test  
-rw-r--r-- 2 student3_10 student3_10 257081603 2021-04-15 19:07 urbandict.csv  
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Правильный просмотр размера файла: hdfs dfs -ls -h

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
-rw-rw-r-- 1 student3_10 student3_10 520 anp 13 19:14 Mapper.py
drwxrwxr-x 2 student3_10 student3_10 4,0K anp 3 18:35 .oracle_jre_usage
-rw-rw-r-- 1 student3_10 student3_10 1,2K anp 13 20:23 reducer.py
-rw-rw-r-- 1 student3_10 student3_10 1,2K anp 13 20:09 Reducer.py
drwx----- 2 student3_10 student3_10 4,0K anp 3 18:40 .ssh
-rw-rw-r-- 1 student3_10 student3_10 246M anp 15 18:53 urbandict.csv
-rw-rw-r-- 1 student3_10 student3_10 331M anp 13 17:56 vehicles.csv
-rw-rw-r-- 1 student3_10 student3_10 2,2K anp 13 20:09 .viminfo
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -put urbandict.csv .
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 5 items
drwx----- - student3_10 student3_10 0 2021-04-15 18:00 .Trash
drwx----- - student3_10 student3_10 0 2021-04-13 20:35 .staging
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 18:20 lesson
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test
-rw-r--r-- 2 student3_10 student3_10 257081603 2021-04-15 19:07 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h
Found 5 items
drwx----- - student3_10 student3_10 0 2021-04-15 18:00 .Trash
drwx----- - student3_10 student3_10 0 2021-04-13 20:35 .staging
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 18:20 lesson
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test
-rw-r--r-- 2 student3_10 student3_10 245.2 M 2021-04-15 19:07 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Удаляем директорию **.Trash** со всем содержимым: `hdfs dfs -rm -r .Trash`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
.bash_history  .bash_profile  mapper.py      .oracle_jre_usage/  Reducer.py      urbandict.csv  .viminfo
.bash_logout   .bashrc        Mapper.py      reducer.py          .ssh/           vehicles.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 3 items
drwx----- - student3_10 student3_10 0 2021-04-13 20:35 .staging
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 19:19 lesson
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .staging
21/04/15 19:24:53 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.staging' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/.staging
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 3 items
drwx----- - student3_10 student3_10 0 2021-04-15 19:24 .Trash
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 19:19 lesson
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 19:19 lesson
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Проверка перемещённых файлов

`hdfs dfs -ls lesson`

`hdfs dfs -ls test`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwx----- - student3_10 student3_10 0 2021-04-13 20:35 .staging
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 19:19 lesson
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .staging
21/04/15 19:24:53 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.staging' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/.staging
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 3 items
drwx----- - student3_10 student3_10 0 2021-04-15 19:24 .Trash
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 19:19 lesson
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 19:19 lesson
drwxr-xr-x - student3_10 student3_10 0 2021-04-15 17:36 test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls lesson
Found 1 items
-rw-r--r-- 2 student3_10 student3_10 257081603 2021-04-15 19:07 lesson/urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls test
Found 1 items
-rw-r--r-- 2 student3_10 student3_10 346353924 2021-04-15 17:36 test/vehicles.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Проверка размера моей папки: `hdfs dfs -du -h`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .staging
21/04/15 19:24:53 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.staging' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/.staging
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 3 items
drwx----- - student3_10 student3_10      0 2021-04-15 19:24 .Trash
drwxr-xr-x - student3_10 student3_10      0 2021-04-15 19:19 lesson
drwxr-xr-x - student3_10 student3_10      0 2021-04-15 17:36 test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - student3_10 student3_10      0 2021-04-15 19:19 lesson
drwxr-xr-x - student3_10 student3_10      0 2021-04-15 17:36 test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls lesson
Found 1 items
-rw-r--r--  2 student3_10 student3_10 257081603 2021-04-15 19:07 lesson/urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls test
Found 1 items
-rw-r--r--  2 student3_10 student3_10 346353924 2021-04-15 17:36 test/vehicles.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -du -h
245.2 M 490.3 M lesson
330.3 M 660.6 M test
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ █
```

cat vehicles.csv | python3.7 mapper.py
cat vehicles.csv | python3.7 mapper.py | sort | python3.7 reducer.py

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
jaguar/t6137143
jeep/t69589094
kia/t19559407
land rover/t94400
lexus/t30034843
lincoln/t7796232
mazda/t14377882
mercedes-benz/t65436510
mercury/t1162558
mini/t7507965
mitsubishi/t3672210
morgan/t74977
nissan/t67049248
pontiac/t3417008
porsche/t17043208
ram/t92764848
rover/t9843502
saturn/t885346
subaru/t30836333
tesla/t10943535
toyota/t134519824
volkswagen/t26791675
volvo/t8615208
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ █
```

current
Удалить: hdfs dfs -rm -r result
Посмотреть результат: hdfs dfs -cat result/part*
cat urbandict.csv | python3.7 mapper.py


```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
319ff6a2
97b18a12
2cff5051
c3cde12e
ff17879
a1901627
c795a9ca
a6c11229
684aab14
999a333b
b8256651
595b54e1
45a1e4b9
c795a9ca
4dda8004
76db3e41
e1f4e3a3
efda79c0
53f9899f
d6a93036
4c2aba10
6a6daf75
2dd25d86
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cat urbandict.csv | python3.7 mapper.py
```

hdfs dfs -rm r result

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
-rw-rw-r-- 1 student3_10 student3_10 525 anp 16 21:27 mapper.py
-rw-rw-r-- 1 student3_10 student3_10 529 anp 16 19:39 Mapper.py
drwxrwxr-x 2 student3_10 student3_10 4,0K anp 3 18:35 .oracle_jre_usage
-rw-rw-r-- 1 student3_10 student3_10 1,2K anp 16 21:33 reducer.py
-rw-rw-r-- 1 student3_10 student3_10 1,2K anp 16 19:55 Reducer.py
drwx----- 2 student3_10 student3_10 4,0K anp 3 18:40 .ssh
-rw-rw-r-- 1 student3_10 student3_10 246M anp 15 18:53 urbandict.csv
-rw-rw-r-- 1 student3_10 student3_10 331M anp 13 17:56 vehicles.csv
-rw----- 1 student3_10 student3_10 3,7K anp 15 21:36 .viminfo
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cat urbandict.csv | python3.7 mapper.py | sort | python3.7 Reducer.py
^[[A^[[ATraceback (most recent call last):
  File "Reducer.py", line 40, in <module>
    print(current_word+'/' +str(current_count))
TypeError: unsupported operand type(s) for +: 'NoneType' and 'str'
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cat urbandict.csv | python3.7 mapper.py | sort | python3.7 Reducer.py
Traceback (most recent call last):
  File "Reducer.py", line 40, in <module>
    print(current_word+'/' +str(current_count))
TypeError: unsupported operand type(s) for +: 'NoneType' and 'str'
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cat urbandict.csv | python3.7 mapper.py | sort | python3.7 reducer.py
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm r result
rm: `r`: No such file or directory
rm: `result`: Is a directory
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Для запуска распределенного вычисления воспользуемся командой:

```
yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar -input lesson -output result -mapper
"python3.7 mapper.py" -reducer "python3.7 reducer.py" -file mapper.py -file reducer.py
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
21/04/16 21:56:02 INFO mapreduce.Job: Job job_1618415779628_0247 running in uber mode : false
21/04/16 21:56:02 INFO mapreduce.Job: map 0% reduce 0%
21/04/16 21:56:12 INFO mapreduce.Job: Task Id : attempt_1618415779628_0247_m_000000_0, Status : FAILED
[2021-04-16 21:56:10.417]Container killed on request. Exit code is 137
[2021-04-16 21:56:10.431]Container exited with a non-zero exit code 137.
[2021-04-16 21:56:10.434]Killed by external signal

21/04/16 21:56:15 INFO mapreduce.Job: map 50% reduce 0%
21/04/16 21:56:23 INFO mapreduce.Job: map 100% reduce 0%
21/04/16 21:56:26 INFO mapreduce.Job: map 100% reduce 100%
21/04/16 21:56:26 INFO mapreduce.Job: Job job_1618415779628_0247 completed successfully
21/04/16 21:56:26 INFO mapreduce.Job: Counters: 56
  File System Counters
    FILE: Number of bytes read=15399937
    FILE: Number of bytes written=31517869
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=257122730
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=11
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Failed map tasks=1
    Launched map tasks=3
    Launched reduce tasks=1
    Other local map tasks=1
    Data-local map tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=112752
    Total time spent by all reduces in occupied slots (ms)=31328
    Total time spent by all map tasks (ms)=28188
    Total time spent by all reduce tasks (ms)=7832
    Total vcore-milliseconds taken by all map tasks=28188
    Total vcore-milliseconds taken by all reduce tasks=7832
    Total megabyte-milliseconds taken by all map tasks=57729024
    Total megabyte-milliseconds taken by all reduce tasks=16039936
  Map-Reduce Framework
    Map input records=1290462
    Map output records=1290462
    Map output bytes=12819007
    Map output materialized bytes=15399943
    Input split bytes=296
    Combine input records=0
    Combine output records=0
    Reduce input groups=825036
    Reduce shuffle bytes=15399943
    Reduce input records=1290462
    Reduce output records=0
    Spilled Records=2580924
    Shuffled Maps =2
    Failed Shuffles=0
    Merged Map outputs=2
    GC time elapsed (ms)=559
    CPU time spent (ms)=15810
    Physical memory (bytes) snapshot=3073064960
    Virtual memory (bytes) snapshot=12877406208
    Total committed heap usage (bytes)=2898788352
    Peak Map Physical memory (bytes)=1463783424
    Peak Map Virtual memory (bytes)=3689484288
    Peak Reduce Physical memory (bytes)=210972672
    Peak Reduce Virtual memory (bytes)=5499621376
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar
-input lesson -output result -mapper "python3.7 mapper.py" -reducer "python3.7 reducer.py" -file mapper.py -file reducer.pyr
```

PycharmProjects - ~/BigData. Введение в экосистему Hadoop/4. Hive & HUE/mapper.py

File Edit View Navigate Code Refactor Run Tools Git Window Help

doop > 4. Hive & HUE > mapper.py runner Git: [status icons]

mapper.py x reducer.py x

```
1  #!/usr/bin/env python
2  """mapper.py"""
3
4  import sys
5
6  for line in sys.stdin:
7      # удаляем лишние пробелы
8      line = line.strip()
9      # делим строки на слова
10     words = line.split(',')
11     try:
12         word = words[4]
13         # price = words[5]
14         # добавляем значение для счетчика
15         # for word in words:
16         # выводим в output пару ключ и значение
17         print(word.lower()+'\t')
18     except Exception:
19         continue
20
```

Project Commit Structure Favorites

Git TODO Problems Python Packages Terminal Python Console Event Log

Nim toolchain: Cannot find Nim co... (30 minutes ago) 20:1 LF UTF-8 4 spaces Python 3.7 master

PycharmProjects - ~/BigData. Введение в экосистему Hadoop/4. Hive & HUE/reducer.py

FileEditViewNavigateCodeRefactorRunToolsGitWindowHelp

doop > 4. Hive & HUE > reducer.pyrunnerGit: ✓✓✓⌚↶⌕⚙

mapper.pyxreducer.pyx

Project

Commit

Structure

Favorites

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

```
#!/usr/bin/env python
"""reducer.py"""

from operator import itemgetter
import sys

current_word = None
current_count = 0
word = None

# читаем строки из STDIN
for line in sys.stdin:
    line = line.strip()
    # line = line.lower()
    # удаляем пробелы в конце и начале строки
    # разделяем пары ключ и значение
    try:
        word, count = line.split('\t', 1)
    except Exception:
        continue
    # преобразуем значение в число
    try:
        count = int(count)
    except ValueError:
        # игнорируем ошибки
        continue

    # на Reduce приходят данные после фазы Sort
    # поэтому этот код будет работать
    if current_word == word:
        current_count += count
    else:
        if current_word:
            # пишем результат в STDOUT
            print(current_word+'\t'+str(current_count))
            current_count = count
            current_word = word
        # не забываем про последнее слово
        if current_word == word:
            print(current_word+'\t'+str(current_count))

for line in sys.stdin > else
```

⌕ Git

☰ TODO

🔍 Problems

📦 Python Packages

📄 Terminal

🔌 Python Console

🔍 Event Log

📄 Nim toolchain: Cannot find Nim co... (31 minutes ago)

32:7 LF UTF-8 4spaces Python 3.7 📄 master 🔒