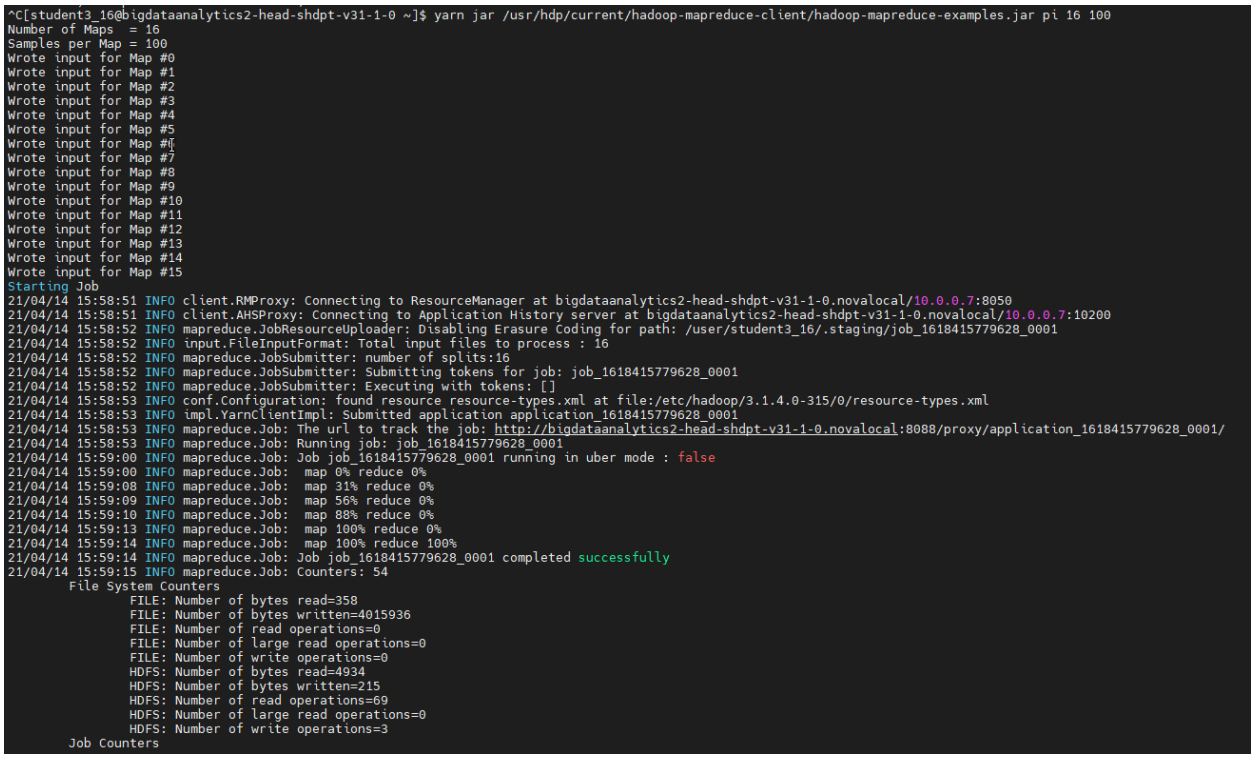
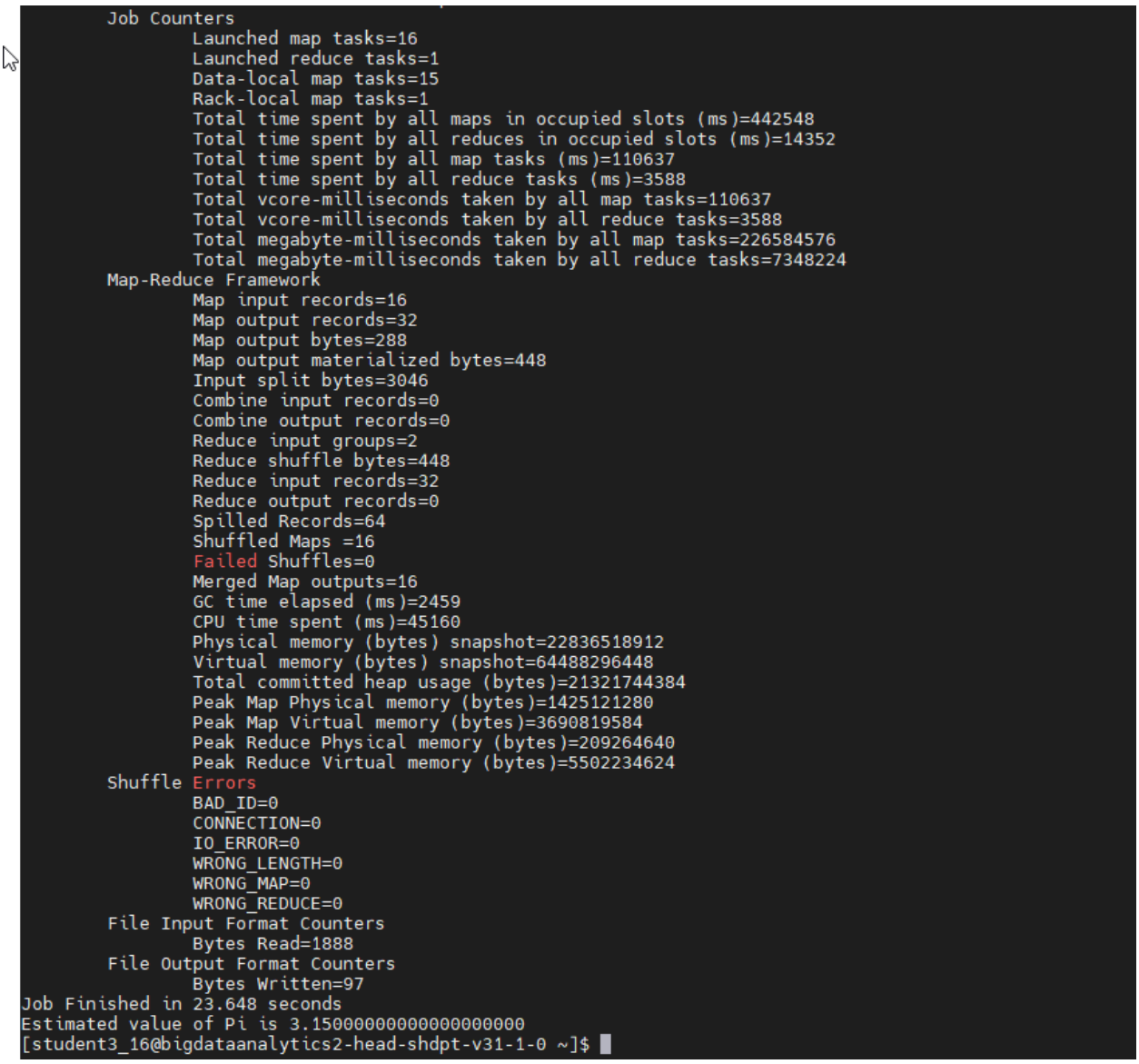
Запустить задачу из примеров, например, вычисление pi методом Монте-Карло



Пытался сначала запустить pi 32 10000, не получалось, видимо из ограниченности ресурсов, снизил в итоге до 16 100



Запустить WordCount и доработать скрипт из примера, чтобы удалялись знаки препинания и слова считались в нижнем регистре

Mapper.py:

```
import sys

for line in sys.stdin:
        line = line.strip()
        words = line.split(',')
        try:
                word = words[7]
                price = words[5]
                print (word.lower(), '\t', price)
        except Exception:
```

```
                continue
```

reducer.py:

```
from operator import itemgetter
import sys
current_word = None
current_count = 0
word = None
for line in sys.stdin:
        line = line.strip()
        try:
                word, count = line.split('\t', 1)
        except Exception:
                continue
        try:
                count = int(count)
        except ValueError:
                continue
        if current_word == word:
                current_count += count
        else:
                if current_word:
                        print (current_word, '\t', str(current_count))
                        current_count = count
                        current_word = word
                if current_word == word:
                print (current_word, '\t', str(current_count))
```

```
[student3_16@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r result
21/04/14 17:02:35 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_16/result' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_16/.Trash/Current/user/student3_16/result1618419755886
[student3_16@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar -input used_cars -output result -mapper "python3.7 mapper.py" -reducer "python3.7 reducer.py" -file mapper.py -file reducer.py
21/04/14 17:02:40 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/streamjob194657392202877506 9.jar tmpDir=null
21/04/14 17:02:42 INFO client.RMProxy: Connecting to ResourceManager at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:8050
21/04/14 17:02:42 INFO client.AHSProxy: Connecting to Application History server at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:10200
21/04/14 17:02:42 INFO client.RMProxy: Connecting to ResourceManager at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:8050
21/04/14 17:02:42 INFO client.AHSProxy: Connecting to Application History server at bigdataanalytics2-head-shdpt-v31-1-0.novalocal/10.0.0.7:10200
21/04/14 17:02:43 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/student3_16/.staging/job_1618415779628_0014
21/04/14 17:02:43 INFO mapred.FileInputFormat: Total input files to process : 1
21/04/14 17:02:43 INFO mapreduce.JobSubmitter: number of splits:3
21/04/14 17:02:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1618415779628_0014
21/04/14 17:02:43 INFO mapreduce.JobSubmitter: Executing with tokens: []
21/04/14 17:02:44 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.4.0-315/0/resource-types.xml
21/04/14 17:02:44 INFO impl.YarnClientImpl: Submitted application application_1618415779628_0014
21/04/14 17:02:44 INFO mapreduce.Job: The url to track the job: http://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8088/proxy/application_1618415779628_0014/
21/04/14 17:02:44 INFO mapreduce.Job: Running job: job_1618415779628_0014
21/04/14 17:02:50 INFO mapreduce.Job: Job job_1618415779628_0014 running in uber mode : false
21/04/14 17:02:50 INFO mapreduce.Job:  map 0% reduce 0%
21/04/14 17:02:59 INFO mapreduce.Job:  map 100% reduce 0%
21/04/14 17:03:06 INFO mapreduce.Job:  map 100% reduce 100%
21/04/14 17:03:06 INFO mapreduce.Job: Job job_1618415779628_0014 completed successfully
21/04/14 17:03:06 INFO mapreduce.Job: Counters: 53
        File System Counters
```

```
                FILE: Number of bytes written=3971481
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=346616518
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=14
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=3
                Launched reduce tasks=1
                Data-local map tasks=3
                Total time spent by all maps in occupied slots (ms)=75664
                Total time spent by all reduces in occupied slots (ms)=13276
                Total time spent by all map tasks (ms)=18916
                Total time spent by all reduce tasks (ms)=3319
                Total vcore-milliseconds taken by all map tasks=18916
                Total vcore-milliseconds taken by all reduce tasks=3319
                Total megabyte-milliseconds taken by all map tasks=38739968
                Total megabyte-milliseconds taken by all reduce tasks=6797312
        Map-Reduce Framework
                Map input records=106181
                Map output records=101051
                Map output bytes=1304933
                Map output materialized bytes=1507074
                Input split bytes=450
                Combine input records=0
                Combine output records=0
                Reduce input groups=135
                Reduce shuffle bytes=1507074
                Reduce input records=101051
                Reduce output records=0
                Spilled Records=202102
                Shuffled Maps =3
                Failed Shuffles=0
                Merged Map outputs=3
                GC time elapsed (ms)=302
                CPU time spent (ms)=13790
                Physical memory (bytes) snapshot=4462772224
                Virtual memory (bytes) snapshot=16564236288
                Total committed heap usage (bytes)=4107796480
                Peak Map Physical memory (bytes)=1418158080
                Peak Map Virtual memory (bytes)=3688562688
                Peak Reduce Physical memory (bytes)=210444288
                Peak Reduce Virtual memory (bytes)=5499203584
        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=346616068
        File Output Format Counters
                Bytes Written=0
21/04/14 17:03:06 INFO streaming.StreamJob: Output directory: result
[student3_16@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```