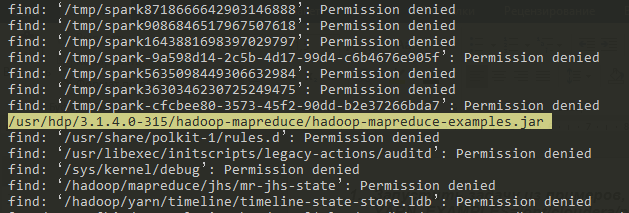
1. ***Запустить задачу из примеров, например, вычисление pi методом Монте-Карло  
   YARN\_EXAMPLES=/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/lib/hadoop-mapreduce/  
   yarn jar $YARN\_EXAMPLES/hadoop-mapreduce-examples.jar pi 16 1000000  
   зайти на ResourceManager***[***http://manager.novalocal:8088***](http://manager.novalocal:8088/) ***и найти свою задачу.  
   Для этого нужно запустить прокси используя -D localhost:1080  
   и настроить браузер на его использование***

***Cсылки, которые написаны в тексте задачи под уроком на GB у меня не сработали.*** *Но я уже стала уверенней и нашла файл* ***hadoop-mapreduce-examples.jar*** *с помощью метода* ***find***

**[student1\_8@bigdataanalytics-head-0 ~]$ find / -name ‘hadoop-mapreduce-examples.jar*’***

*…три километра* **Permission denied**

*и свет в конце тоннеля*



*Проверяю еще раз, что этот файл мне доступен и запускаю job на yarn с помощью полученной ссылки*



**[student1\_8@bigdataanalytics-head-0 ~]$ yarn jar /usr/hdp/3.1.4.0-315/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 16 1000000**

Number of Maps = 16

Samples per Map = 1000000

Wrote input for Map #0

Wrote input for Map #1

Wrote input for Map #2

Wrote input for Map #3

Wrote input for Map #4

Wrote input for Map #5

Wrote input for Map #6

Wrote input for Map #7

Wrote input for Map #8

Wrote input for Map #9

Wrote input for Map #10

Wrote input for Map #11

Wrote input for Map #12

Wrote input for Map #13

Wrote input for Map #14

Wrote input for Map #15

Starting Job

21/02/07 11:48:19 INFO client.RMProxy: Connecting to ResourceManager at bigdataanalytics-head-0.novalocal/10.0.0.8:8050

21/02/07 11:48:19 INFO client.AHSProxy: Connecting to Application History server at bigdataanalytics-head-0.novalocal/10.0.0.8:10200

21/02/07 11:48:19 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/student1\_8/.staging/job\_1612470347159\_0094

21/02/07 11:48:19 INFO input.FileInputFormat: Total input files to process : 16

21/02/07 11:48:19 INFO mapreduce.JobSubmitter: number of splits:16

21/02/07 11:48:20 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1612470347159\_0094

21/02/07 11:48:20 INFO mapreduce.JobSubmitter: Executing with tokens: []

21/02/07 11:48:20 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.4.0-315/0/resource-types.xml

21/02/07 11:48:20 INFO impl.YarnClientImpl: Submitted application application\_1612470347159\_0094

21/02/07 11:48:20 INFO mapreduce.Job: The url to track the job: http://bigdataanalytics-head-0.novalocal:8088/proxy/application\_1612470347159\_0094/

21/02/07 11:48:20 INFO mapreduce.Job: Running job: job\_1612470347159\_0094

21/02/07 11:48:26 INFO mapreduce.Job: Job job\_1612470347159\_0094 running in uber mode : false

21/02/07 11:48:26 INFO mapreduce.Job: map 0% reduce 0%

21/02/07 11:48:35 INFO mapreduce.Job: map 25% reduce 0%

21/02/07 11:48:39 INFO mapreduce.Job: map 69% reduce 0%

21/02/07 11:48:40 INFO mapreduce.Job: map 75% reduce 0%

21/02/07 11:48:47 INFO mapreduce.Job: map 75% reduce 25%

21/02/07 11:49:59 INFO mapreduce.Job: map 81% reduce 25%

21/02/07 11:49:59 INFO mapreduce.Job: Task Id : attempt\_1612470347159\_0094\_m\_000015\_0, Status : FAILED

[2021-02-07 11:49:57.054]Container killed on request. Exit code is 137

[2021-02-07 11:49:57.875]Container exited with a non-zero exit code 137.

[2021-02-07 11:49:57.887]Killed by external signal

21/02/07 11:49:59 INFO mapreduce.Job: Task Id : attempt\_1612470347159\_0094\_m\_000014\_0, Status : FAILED

[2021-02-07 11:49:58.039]Container killed on request. Exit code is 137

[2021-02-07 11:49:58.298]Container exited with a non-zero exit code 137.

[2021-02-07 11:49:58.299]Killed by external signal

21/02/07 11:50:00 INFO mapreduce.Job: Task Id : attempt\_1612470347159\_0094\_m\_000002\_0, Status : FAILED

[2021-02-07 11:49:58.259]Container killed on request. Exit code is 137

[2021-02-07 11:49:58.300]Container exited with a non-zero exit code 137.

[2021-02-07 11:49:58.300]Killed by external signal

21/02/07 11:50:02 INFO mapreduce.Job: map 81% reduce 27%

21/02/07 11:50:05 INFO mapreduce.Job: map 88% reduce 27%

21/02/07 11:50:06 INFO mapreduce.Job: map 94% reduce 27%

21/02/07 11:50:08 INFO mapreduce.Job: map 94% reduce 31%

21/02/07 11:50:10 INFO mapreduce.Job: map 100% reduce 31%

21/02/07 11:50:11 INFO mapreduce.Job: map 100% reduce 100%

21/02/07 11:50:11 INFO mapreduce.Job: Job job\_1612470347159\_0094 completed successfully

21/02/07 11:50:12 INFO mapreduce.Job: Counters: 56

File System Counters

FILE: Number of bytes read=358

FILE: Number of bytes written=4023552

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4710

HDFS: Number of bytes written=215

HDFS: Number of read operations=69

HDFS: Number of large read operations=0

HDFS: Number of write operations=3

Job Counters

Failed map tasks=3

Launched map tasks=19

Launched reduce tasks=1

Other local map tasks=3

Data-local map tasks=15

Rack-local map tasks=1

Total time spent by all maps in occupied slots (ms)=1471926

Total time spent by all reduces in occupied slots (ms)=369272

Total time spent by all map tasks (ms)=490642

Total time spent by all reduce tasks (ms)=92318

Total vcore-milliseconds taken by all map tasks=490642

Total vcore-milliseconds taken by all reduce tasks=92318

Total megabyte-milliseconds taken by all map tasks=753626112

Total megabyte-milliseconds taken by all reduce tasks=189067264

Map-Reduce Framework

Map input records=16

Map output records=32

Map output bytes=288

Map output materialized bytes=448

Input split bytes=2822

Combine input records=0

Combine output records=0

Reduce input groups=2

Reduce shuffle bytes=448

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)=3627

CPU time spent (ms)=34650

Physical memory (bytes) snapshot=17856294912

Virtual memory (bytes) snapshot=55540432896

Total committed heap usage (bytes)=16896753664

Peak Map Physical memory (bytes)=1124728832

Peak Map Virtual memory (bytes)=3250233344

Peak Reduce Physical memory (bytes)=256012288

Peak Reduce Virtual memory (bytes)=3723874304

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

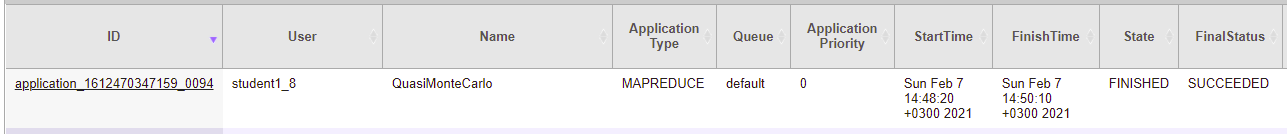
File Output Format Counters

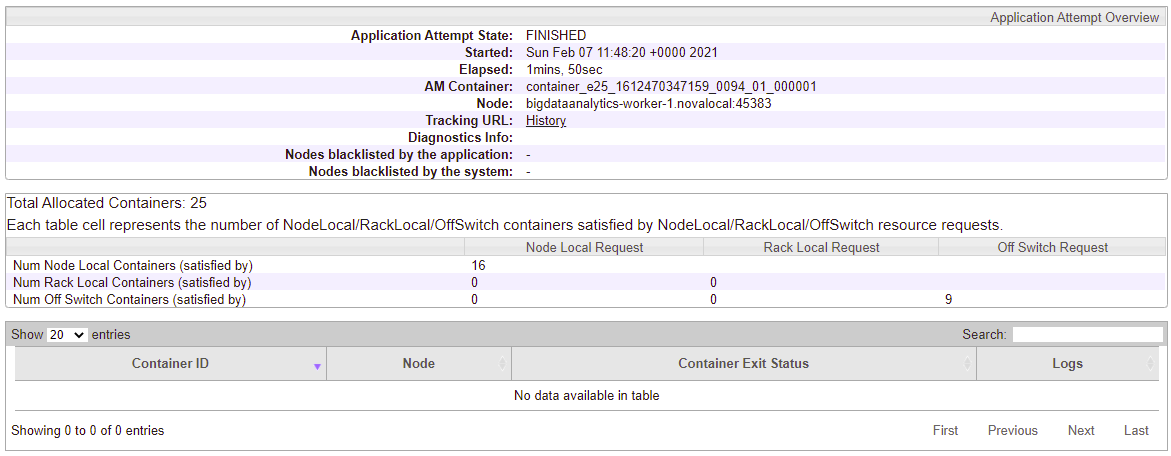
Bytes Written=97

Job Finished in 113.105 seconds

Estimated value of Pi is 3.14159125000000000000

*Вижу свою задачу на* *ResourceManager. УРА!*

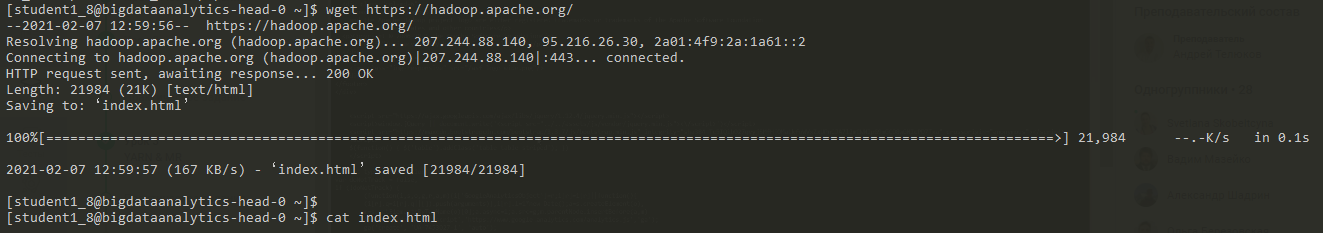


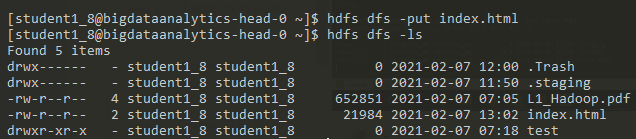


***В этот момент Вы прислали методичку, в которой директории в ДЗ прописаны совсем не так, как в задании, которое прикреплено к уроку. Если я правильно понимаю, то в задании к уроку какие-то устаревшие данные…***

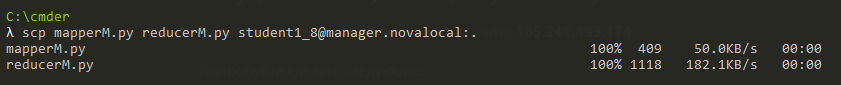
***2. Запустить WordCount и доработать скрипт из примера, чтобы удалялись знаки препинания и слова считались в нижнем регистре***

*Я взяла тот же файл, что был на уроке*

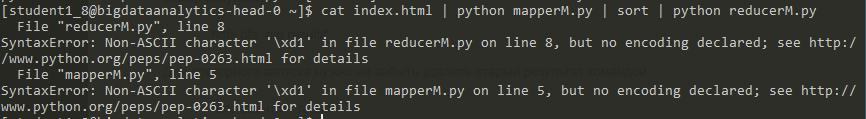




*Я создала файлы mapper.py и reducer.py в нескольких вариантах, т.к. текст пол уроком на GB и в методичке отличается (M в названии означает, что файл сделан на основе методички), и закинула их на сервер*



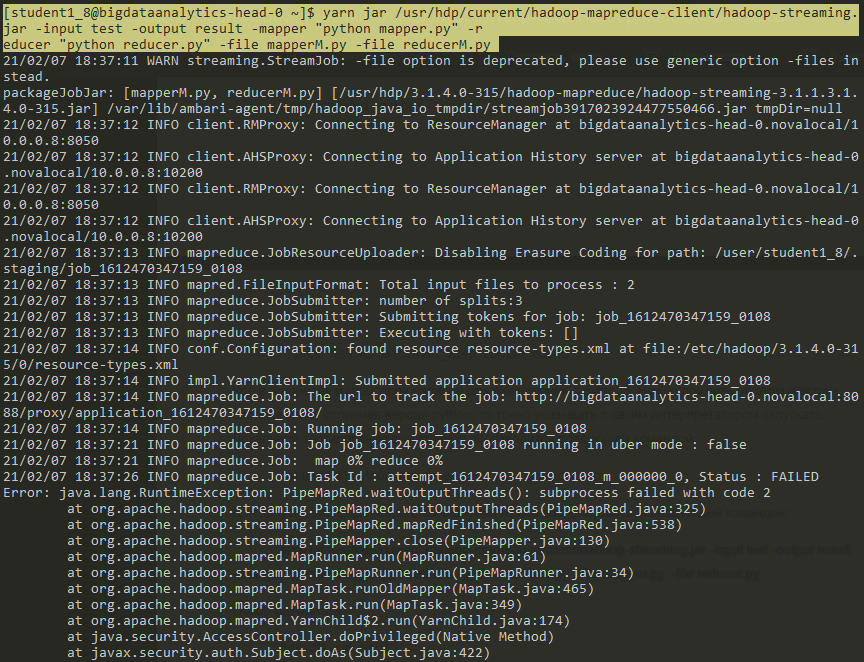
*Запускаю проверку на NameNode и получаю ошибку*

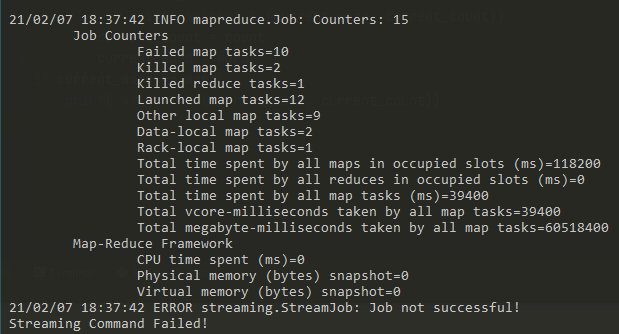


***Делаю вывод, что процесс «подавился» комментариями на русском.***

*Убираю русский текст, перезаливаю, еще раз запускаю проверку… и все работает!*

*Перемещаю файл index.html в папку test и запускаю job на hdfs*



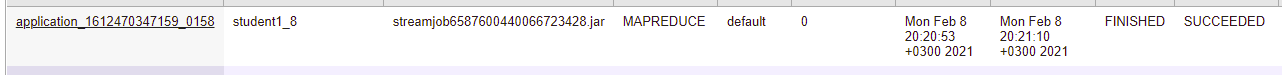


*Повылезали ошибки и в итоге job зафэйлился*

*….*

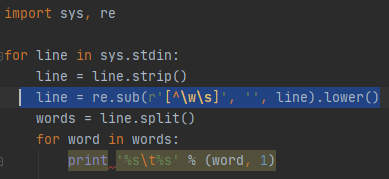
*После разбора ошибок на уроке и замечания по поводу версии питона, я переписала файлы mapper и reduce с синтаксисом как для 2го питона, перезалила их и job выполнился успешно!!!*





***…доработать скрипт из примера, чтобы удалялись знаки препинания и слова считались в нижнем регистре***

*Внесла обработку строк в mapper*



*Работает отлично!*

***3. \*реализовать алгоритм join на MapReduceсдаюсь***