

Assignment 2

Alisa Gazizullina

Files here: <https://github.com/Alicegaz/MapReduceHW.git>

1. Run jobs by

```
hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar \  
-files mapper.py, reducer.py -mapper mapper.py -reducer reducer.py \  
-input /input -output /output
```

2. There are three clusters created: one master and two workers

Master: 2 virtual CPU, 7,75 GB

Disk size 100+400GB

Worker: 2 virtual CPU, 7,75 GB (I later changed the machine type)

Disk size 500GB

Название	cluster-451d
Регион	global
Зона	europe-west1-c
Главный узел	Стандартный (1 главный узел; рабочих узлов: N)
Тип машины	n1-standard-2 (количество виртуальных ядер: 2, объем памяти: 7.50 ГБ)
Размер основного диска	500 ГБ
Рабочие узлы	2
Тип машины	n1-standard-1 (количество виртуальных ядер: 1, объем памяти: 3.75 ГБ)
Размер основного диска	100 ГБ
Локальные SSD-диски	0
Вытесняемые рабочие узлы	0
Сегмент для временного хранения файлов	dataproc-7aed22a5-6811-4f1c-b079-5e6c4945ca30-eu
Подсеть	default
Только внутренние IP-адреса	Нет
Версия образа	1.2.12
Дата создания	15 нояб. 2017 г., 12:57:22
Ярлыки	
goog-dataproc-cluster-name	cluster-451d
goog-dataproc-cluster-uuid	4a79d9c4-1c97-40f7-8bc2-ebc61dba8c50

Эквивалентный [запрос/ответ REST](#)

3. Map reduce job was started at 17:48:18, continued till 19:43: 36

2671856Darth_Vader 20171001:85678 20171002:85945 20171003:86218 20171004:85218
20171005:84326 20171006:85854 20171007:87376 20171008:87280 20171009:86521 20171010:88451
20171011:80412 20171012:86136 20171013:87686 20171014:86996 20171015:87277 20171016:86090
20171017:86559 20171018:85415 20171019:85523 20171020:86074 20171021:86188 20171022:86497
20171023:86738 20171024:85352 20171025:86664 20171026:86668 20171027:86742 20171028:87095
20171029:87173 20171030:85427 20171031:86277
2445427Harvey_Weinstein 20171001:279 20171002:298 20171003:319 20171004:753 20171005:40145
20171006:89668 20171007:60317 20171008:40232 20171009:170817 20171010:267581 20171011:440328
20171012:301996 20171013:187869 20171014:102363 20171015:139608 20171016:119625 20171017:81199
20171018:66547 20171019:44903 20171020:54226 20171021:25703 20171022:21348 20171023:25577
20171024:25306 20171025:22456 20171026:32110 20171027:18428 20171028:14300 20171029:13859
20171030:18207 20171031:19060

Assignment 2

Alisa Gazizullina

2051566Tom_Petty 1962149039 20171001:802 20171002:657266 20171003:837800 20171004:168347
20171005:67491 20171006:45504 20171007:32052 20171008:41849 20171009:33556 20171010:24164
20171011:14120 20171012:11480 20171013:9813 20171014:7903 20171015:7001 20171016:7867
20171017:9096 20171018:7667 20171019:7482 20171020:9173 20171021:5623 20171022:5155
20171023:5903 20171024:5304 20171025:5289 20171026:4658 20171027:4451 20171028:3497
20171029:3537 20171030:4005 20171031:3711

1993599Deaths_in_2017 20171001:56819 20171002:83655 20171003:86579 20171004:66631
20171005:62825 20171006:63862 20171007:57341 20171008:52901 20171009:66995 20171010:66241
20171011:65481 20171012:63463 20171013:60564 20171014:48607 20171015:52623 20171016:75302
20171017:71693 20171018:72642 20171019:70255 20171020:64206 20171021:52055 20171022:54668
20171023:70736 20171024:69722 20171025:73147 20171026:65822 20171027:64079 20171028:50939
20171029:50397 20171030:65791 20171031:67558

1516233Blade_Runner_2049 242076834 20171001:27700 20171002:36550 20171003:43109
20171004:56762 20171005:78119 20171006:116450 20171007:99701 20171008:111531 20171009:124818
20171010:84333 20171011:68031 20171012:55460 20171013:50006 20171014:47435 20171015:57198
20171016:58535 20171017:38576 20171018:35206 20171019:31659 20171020:29218 20171021:28207
20171022:35520 20171023:38779 20171024:24671 20171025:23084 20171026:19719 20171027:18354
20171028:17436 20171029:20258 20171030:22611 20171031:17197

1338198Curse_Inc. 388398734 20171001:67 20171002:174956 20171003:1069249 20171004:140
20171005:106 20171006:34641 20171007:56069 20171008:101 20171009:117 20171010:147 20171011:99
20171012:134 20171013:126 20171014:100 20171015:99 20171016:115 20171017:138 20171018:180
20171019:139 20171020:130 20171021:106 20171022:109 20171023:143 20171024:174 20171025:130
20171026:161 20171027:105 20171028:92 20171029:104 20171030:130 20171031:91

1322438Catalonia 266131698 20171001:115696 20171002:208674 20171003:86612 20171004:109250
20171005:54871 20171006:32561 20171007:20909 20171008:24087 20171009:29730 20171010:51358
20171011:48036 20171012:19960 20171013:9709 20171014:5743 20171015:5950 20171016:10566
20171017:7730 20171018:6896 20171019:19795 20171020:9410 20171021:15344 20171022:16479
20171023:13115 20171024:7596 20171025:6416 20171026:8273 20171027:180840 20171028:93779
20171029:43499 20171030:37305 20171031:22249

1213098Stranger_Things 1364391934 0171001:8658 20171002:10469 20171003:9135 20171004:11093
20171005:10908 20171006:9865 20171007:7934 20171008:8892 20171009:10953 20171010:10146
20171011:9252 20171012:10128 20171013:19671 20171014:14803 20171015:13345 20171016:15671
20171017:14333 20171018:12583 20171019:13343 20171020:14191 20171021:13786 20171022:16971
20171023:27830 20171024:26231 20171025:26804 20171026:37696 20171027:131602 20171028:175816
20171029:173939 20171030:200256 20171031:146794

11837992018_FIFA_World_Cup_qualification 30915376 20171001:6038 20171002:11454 20171003:14662
20171004:17200 20171005:56441 20171006:108953 20171007:107825 20171008:116542 20171009:131333
20171010:157825 20171011:222159 20171012:42939 20171013:24357 20171014:14990 20171015:12171
20171016:19144 20171017:28949 20171018:13341 20171019:9204 20171020:7185 20171021:5537
20171022:5337 20171023:6797 20171024:6318 20171025:5896 20171026:5946 20171027:4988 20171028:4157
20171029:4449 20171030:5972 20171031:5690

1010646Blade_Runner 240494360

959442 Fridtjof_Nansen 525052251

947288 2017_Las_Vegas_Strip_shooting 65750337

908817 Selena 1287244117

888331 Halloween 633851035

812043 Mindhunter_(TV_series) 984820699

755019 Kevin_Spacey 811476949

Assignment 2

Alisa Gazizullina