

Konfiguracja testowego uruchomienia

Cluster: MPR_cluster_2 Waiting Cluster ready after last step completed.

Summary	Application user interfaces	Monitoring	Hardware	Configurations	Events	Steps	Bootstrap actions
Add task instance group							
Instance groups							
Filter: Filter instance groups ... 2 instance groups (all loaded)							
ID	Status	Node type & name	Instance type	Instance count	Purchasing option		
ig-NV4KABWU71OI	Running	CORE Core Instance Group	m4.large 2 vCore, 8 GiB memory, EBS only storage EBS Storage: 32 GiB	2 Instances Resize	On-demand ⓘ		
ig-2PMPUA54QG38Y	Running	MASTER Master Instance Group	m4.large 2 vCore, 8 GiB memory, EBS only storage EBS Storage: 32 GiB	1 Instances	On-demand ⓘ		

Wykonane joby

```
hadoop@ip-172-31-46-7:~/books
CPU time spent (ms)=317350
Physical memory (bytes) snapshot=68386680832
Virtual memory (bytes) snapshot=484477640704
Total committed heap usage (bytes)=55462854656

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=64860323
File Output Format Counters
  Bytes Written=4495199
21/05/04 21:06:07 INFO streaming.StreamJob: Output directory: books-output

real    10m0.416s
user    0m9.900s
sys     0m0.712s
[hadoop@ip-172-31-46-7 books]$ time hadoop jar /usr/lib/hadoop/hadoop-streaming.jar -f
les mapper.py, reducer.py -mapper mapper.py -reducer reducer.py -input books-input -outp
ut books-output
```

```
hadoop@ip-172-31-46-7:~/books
1302.txt 1316-0.txt 1332-0.txt 1348-0.txt 1364.txt 1377.txt 1390.txt
1303-0.txt 1316.txt 1332.txt 1348.txt 1365-0.txt 1378-0.txt 1391-0.txt
1303.txt 1317-0.txt 1333.txt 1349-0.txt 1366-0.txt 1378.txt 1391.txt
1304.txt 1317.txt 1334-0.txt 1350-0.txt 1366.txt 1379-0.txt 1392.txt
1305-0.txt 1318.txt 1334.txt 1351.txt 1367.txt 1379.txt 1393.txt
1305.txt 1319.txt 1335.txt 1352-0.txt 1368-0.txt 1380-0.txt 1394.txt
1306-0.txt 1320-0.txt 1336.txt 1352.txt 1369-0.txt 1380.txt 1395.txt
1306.txt 1321-0.txt 1337.txt 1353-0.txt 1369.txt 1381-0.txt 1396-0.txt
1307-0.txt 1322-0.txt 1338.txt 1353.txt 1370.txt 1382-0.txt 1396.txt
1307.txt 1323.txt 1339-0.txt 1354-0.txt 1371-0.txt 1383-0.txt 1397.txt
1308-0.txt 1324.txt 1340.txt 1354.txt 1371.txt 1384.txt 1398.txt
1309.txt 1325.txt 1341.txt 1355-0.txt 1372-0.txt 1385-0.txt 1399-0.txt
1310-0.txt 1326.txt 1342-0.txt 1356.txt 1372.txt 1385.txt 1400-0.txt
1311.txt 1327-0.txt 1343-0.txt 1357-0.txt 1373.txt 1386-0.txt mapper.py
1312-0.txt 1327.txt 1343.txt 1357.txt 1374-0.txt 1386.txt reducer.py
1312.txt 1328-0.txt 1344-0.txt 1358.txt 1374.txt 1387-0.txt
1313-0.txt 1328.txt 1344.txt 1359.txt 1375-0.txt 1387.txt
1313.txt 1329-0.txt 1345-0.txt 1360.txt 1375.txt 1388.txt

[hadoop@ip-172-31-46-7 books]$ touch sequential_mr.py
[hadoop@ip-172-31-46-7 books]$ vi sequential_mr.py
[hadoop@ip-172-31-46-7 books]$ python sequential_mr.py
Time execution 6.368947267532349 [s]
[hadoop@ip-172-31-46-7 books]$
```

Pomiar czasów

- sekwencyjny: 6.4 sekund
- równoległy: 10 minut

Tak drastyczna różnica między algorytmem sekwencyjnym a równoległym mogła wynikać tak z małej ilości danych (łącznie rozmiar plików tekstowych wynosił 64 MB) jak i małej liczby procesorów (1 master, 2 slave'y)