

UJIAN TENGAH SEMESTER

MATA KULIAH BIG DATA



Disusun Oleh:

Angelia Anna P.R	2241760016
Dimas Rosyidin	2241760054
Muhammad Khasbul Hadi Fauzan	2241760039
Muhammad Ivan Yoda Bellamy	2241760094
Rega Dane Wijayanta	2241760113

PROGRAM STUDI D4 SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

2025

1. Apakah semakin banyak workers eksekusi semakin cepat

Jawab :

Tabel Eksekusi Pyspark

Worker	Waktu (Detik)	Matriks
5	102,90	1000 x 1000
10	182,41	1000 x 1000
20	343,27	1000 x 1000
5	error	10.000 x 10.000
10	error	10.000 x 10.000
20	error	10.000 x 10.000
5	error	20.000 x 20.000
10	error	20.000 x 20.000
20	error	20.000 x 20.000
5	error	40.000 x 40.000
10	error	40.000 x 40.000
20	error	40.000 x 40.000

Berdasarkan tabel di atas, semakin banyak workers tidak membuat eksekusi semakin cepat. Hal ini disebabkan karena spesifikasi VM yang kami gunakan hanya memiliki 6 core CPU. Sehingga penggunaan resource akan dialokasikan secara bergantian. Selain itu, hasil eksekusi terhadap matrix 10.000 x 10.000 menghasilkan error, karena RAM dari VM kami tidak sanggup mengatasinya. Berikut spesifikasi VM kami.

```
(sparkenv) root@hadoop-namenode:/home/hadoopuser# lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Address sizes:          48 bits physical, 48 bits virtual
Byte Order:             Little Endian
CPU(s):                 6
On-line CPU(s) list:   0-5
Vendor ID:              AuthenticAMD
Model name:             AMD Ryzen 5 6600H with Radeon Graphics
BIOS Model name:        CPU @ 0.0GHz
BIOS CPU family:        0
CPU family:             25
Model:                  68
Thread(s) per core:     1
Core(s) per socket:     6
Socket(s):              1
Stepping:               1
BogoMIPS:               6587.62
```

```
(sparkenv) root@hadoop-namenode:/home/hadoopuser# free -h
               total        used        free      shared  buff/cache   available
Mem:           7.8Gi        511Mi        7.0Gi         1.1Mi        488Mi        7.3Gi
Swap:           0B           0B           0B
```

(sparkenv) root@hadoop-namenode:/home/hadoopuser# |

2. Apakah linear peningkatan kecepatan itu, atau seperti apa polanya?

Jawab :

Linear atau tidaknya peningkatan kecepatan eksekusi bergantung pada spesifikasi VM yang digunakan. Semakin banyak core CPU dan semakin besar ukuran RAM, maka semakin linear juga peningkatan kecepatan berdasarkan jumlah worker yang ditambah.

3. Bandingkan performa Scala dan Python. Sama atau beda

Jawab :

Performa scala dan python cukup berbeda tergantung konteks penggunaannya:

Kecepatan	Scala	Python
Kecepatan	cepat	lebih lambat
Sintaks	kompleks	mudah
Ekosistem	big data, JVM	data science, AI, web
Concurrency	kuat	terbatas
Learning Curve	menengah-tinggi	rendah

Python

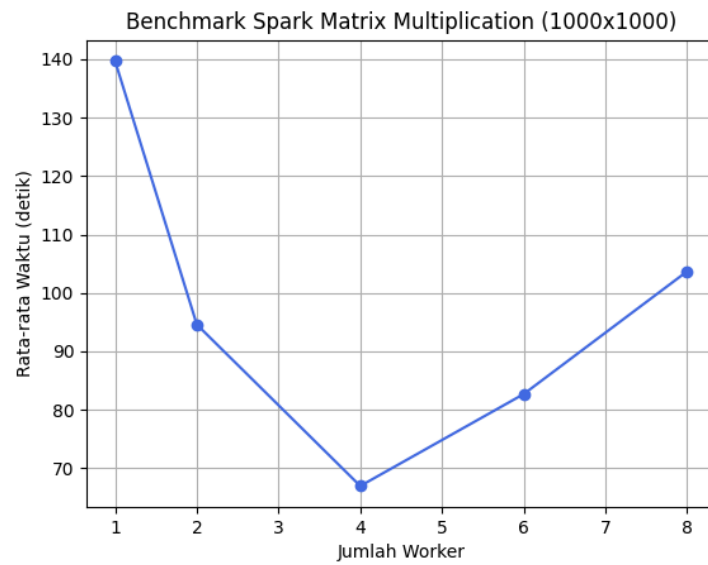
```
Hasil perkalian matriks (5x5 pertama):
[[252.46630166 253.08596109 252.15522702 243.38453959 248.77492917]
 [253.36291787 261.75296647 257.18972024 256.09532236 253.28106759]
 [263.09037407 265.93525467 264.17466956 260.32532039 254.34684228]
 [253.29876202 265.86406775 255.90487838 255.98188485 254.86321245]
 [254.03601781 259.61299093 255.94831653 257.37153895 251.98517111]]
Waktu eksekusi: 84.54 detik
```

Scala

```
Hasil perkalian (5x5 pertama):
244.0491597982624    241.41166045483618    233.14434927312803    238.79089316514487    242.511216833132
259.37228302276463    254.0460217294455    246.84710542386642    251.7760331946147    258.9101665268854
257.3476190486054    249.1721196929347    246.48205650999847    249.4936896834159    252.70378003137515
245.15208198212633    249.2155184415875    240.97714638999392    245.4948749082434    246.52438667496762
246.99793923185712    246.4416766247512    243.00993040594778    238.23447972649427    251.09502717295123
Total waktu eksekusi: 30.21 detik
```

Dari sini terbukti bahwa Scala memiliki performa lebih cepat daripada Python.

4. Tunjukkan dalam grafis dan Laporan yang bagus
Jawab :



Worker	Rata-rata Waktu (detik)
1	139.6490399837494
2	94.59789355595906
4	66.98939561843872
6	82.710307041804
8	103.70208692550659