

# WORDCOUNT

Compare wordcount with hadoop and with Java

## Group 4

Zulfikar Hadzalic 2106636224

Fayza Nirwasita 2106635700

Rafi' Noval Hady 2106703153

Muhammad Zaki Nur Said Hanan 2106733856

## COMPARISON

# File 1 mb

### HADOOP

Pada wordcount dengan menggunakan hadoop dan file sebesar 1 mb, diperoleh running time sebesar:

**22 Detik**

User:	rafin
Name:	word count
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Thu Jun 08 00:15:24 +0700 2023
Launched:	Thu Jun 08 00:15:25 +0700 2023
Finished:	Thu Jun 08 00:15:47 +0700 2023
Elapsed:	22sec
Tracking URL:	<a href="#">History</a>
Log Aggregation Status:	DISABLED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

```
Run: Main ×
cry 3474
any 579
the 1737
what 579
ask 579
to 2316
desert 3474
Execution time: 145122800 nanoseconds

Process finished with exit code 0
```

Version Control Run TODO Problems Terminal

### JAVA

Pada wordcount dengan menggunakan Java dan file sebesar 1 mb, diperoleh running time sebesar:

**0.1451 Detik**

## COMPARISON

# File 10 mb

### HADOOP

Pada wordcount dengan menggunakan hadoop dan file sebesar 10 mb, diperoleh running time sebesar:

**23 Detik**

User:	rafin
Name:	word count
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Thu Jun 08 00:13:55 +0700 2023
Launched:	Thu Jun 08 00:13:56 +0700 2023
Finished:	Thu Jun 08 00:14:19 +0700 2023
Elapsed:	23sec
Tracking URL:	<a href="#">History</a>
Log Aggregation Status:	DISABLED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

```
Run: Main x Main x
cry 34758
any 5793
the 17379
what 5793
ask 5793
to 23172
desert 34758
Execution time: 475085800 nanoseconds

Process finished with exit code 0

Version Control Run TODO Problems Terminal
Build completed successfully in 1 sec, 722 ms (moments ago)
```

### JAVA

Pada wordcount dengan menggunakan Java dan file sebesar 10 mb, diperoleh running time sebesar:

**0.4751 Detik**

## COMPARISON

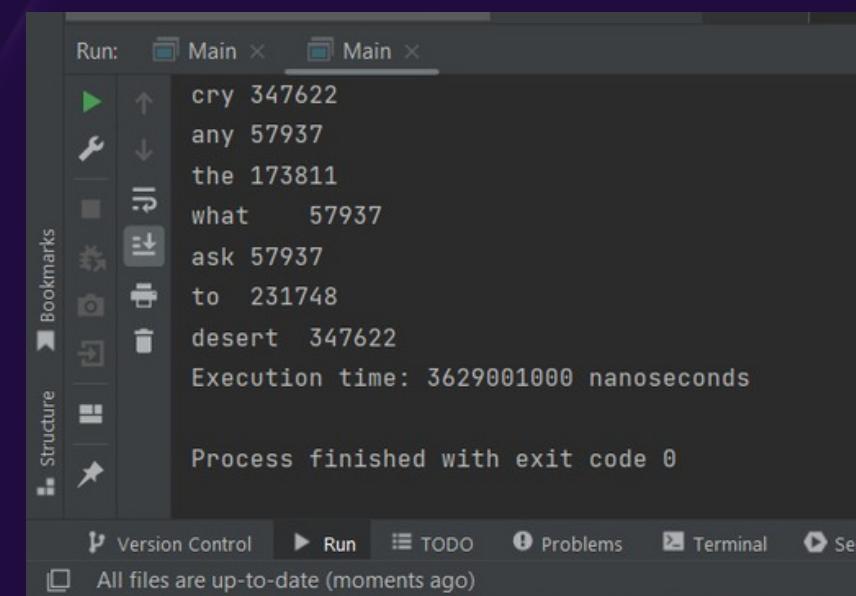
# File 100 mb

### HADOOP

Pada wordcount dengan menggunakan hadoop dan file sebesar 100 mb, diperoleh running time sebesar:

**32 Detik**

User:	rafin
Name:	word count
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Thu Jun 08 00:16:38 +0700 2023
Launched:	Thu Jun 08 00:16:38 +0700 2023
Finished:	Thu Jun 08 00:17:11 +0700 2023
Elapsed:	32sec
Tracking URL:	<a href="#">History</a>
Log Aggregation Status:	DISABLED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>



```
Run: Main x Main x
cry 347622
any 57937
the 173811
what 57937
ask 57937
to 231748
desert 347622
Execution time: 3629001000 nanoseconds

Process finished with exit code 0

Version Control Run TODO Problems Terminal Search
All files are up-to-date (moments ago)
```

### JAVA

Pada wordcount dengan menggunakan Java dan file sebesar 100 mb, diperoleh running time sebesar:

**3.6290 Detik**

## COMPARISON

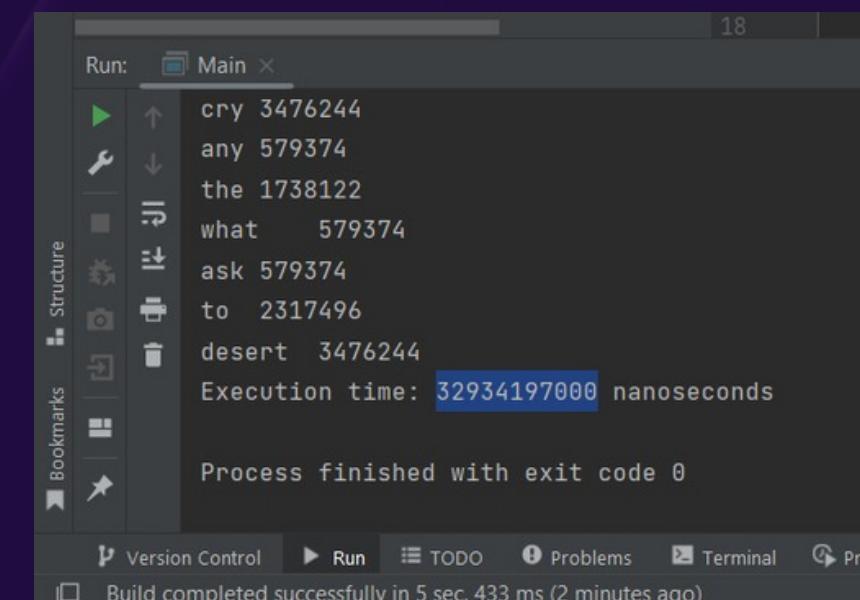
# File 1000 mb

### HADOOP

Pada wordcount dengan menggunakan hadoop dan file sebesar 1000 mb, diperoleh running time sebesar:

**107 Detik**

User:	rafin
Name:	word count
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Thu Jun 08 00:17:52 +0700 2023
Launched:	Thu Jun 08 00:17:53 +0700 2023
Finished:	Thu Jun 08 00:19:40 +0700 2023
Elapsed:	1mins, 47sec
Tracking URL:	History
Log Aggregation Status:	DISABLED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>



```
Main x
cry 3476244
any 579374
the 1738122
what 579374
ask 579374
to 2317496
desert 3476244
Execution time: 32934197000 nanoseconds

Process finished with exit code 0
```

Version Control Run TODO Problems Terminal
Build completed successfully in 5 sec, 433 ms (2 minutes ago)

### JAVA

Pada wordcount dengan menggunakan Java dan file sebesar 1000 mb, diperoleh running time sebesar:

**32.934 Detik**

## COMPARISON

# File 2000 mb

### HADOOP

Pada wordcount dengan menggunakan hadoop dan file sebesar 2000 mb, diperoleh running time sebesar:

**200 Detik**

User:	rafin
Name:	word count
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Thu Jun 08 00:20:28 +0700 2023
Launched:	Thu Jun 08 00:20:29 +0700 2023
Finished:	Thu Jun 08 00:23:49 +0700 2023
Elapsed:	3mins, 20sec
Tracking URL:	<a href="#">History</a>
Log Aggregation Status:	DISABLED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

```
Main x
cry 6952488
any 1158748
the 3476244
what 1158748
ask 1158748
to 4634992
desert 6952488
Execution time: 61716184100 nanoseconds

Process finished with exit code 0
```

### JAVA

Pada wordcount dengan menggunakan Java dan file sebesar 1000 mb, diperoleh running time sebesar:

**61.716 Detik**

## COMPARISON

# File 10000 mb

### HADOOP

Pada wordcount dengan menggunakan hadoop dan file sebesar 2000 mb, diperoleh running time sebesar:

**899 Detik**

User:	rafin
Name:	word count
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Thu Jun 08 00:26:06 +0700 2023
Launched:	Thu Jun 08 00:26:07 +0700 2023
Finished:	Thu Jun 08 00:41:06 +0700 2023
Elapsed:	14mins, 59sec
Tracking URL:	<a href="#">History</a>
Log Aggregation Status:	DISABLED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

```
Main x
cry 34762452
any 5793742
the 17381226
what 5793742
ask 5793742
to 23174968
desert 34762452
Execution time: 328500414500 nanoseconds

Process finished with exit code 0
```

The screenshot shows the Java IDE's run console. It displays the output of a wordcount program, which includes words and their counts (cry, any, the, what, ask, to, desert) and their respective counts (34762452, 5793742, 17381226, 5793742, 5793742, 23174968, 34762452). Below the output, it says "Execution time: 328500414500 nanoseconds". At the bottom, it states "Process finished with exit code 0".

### JAVA

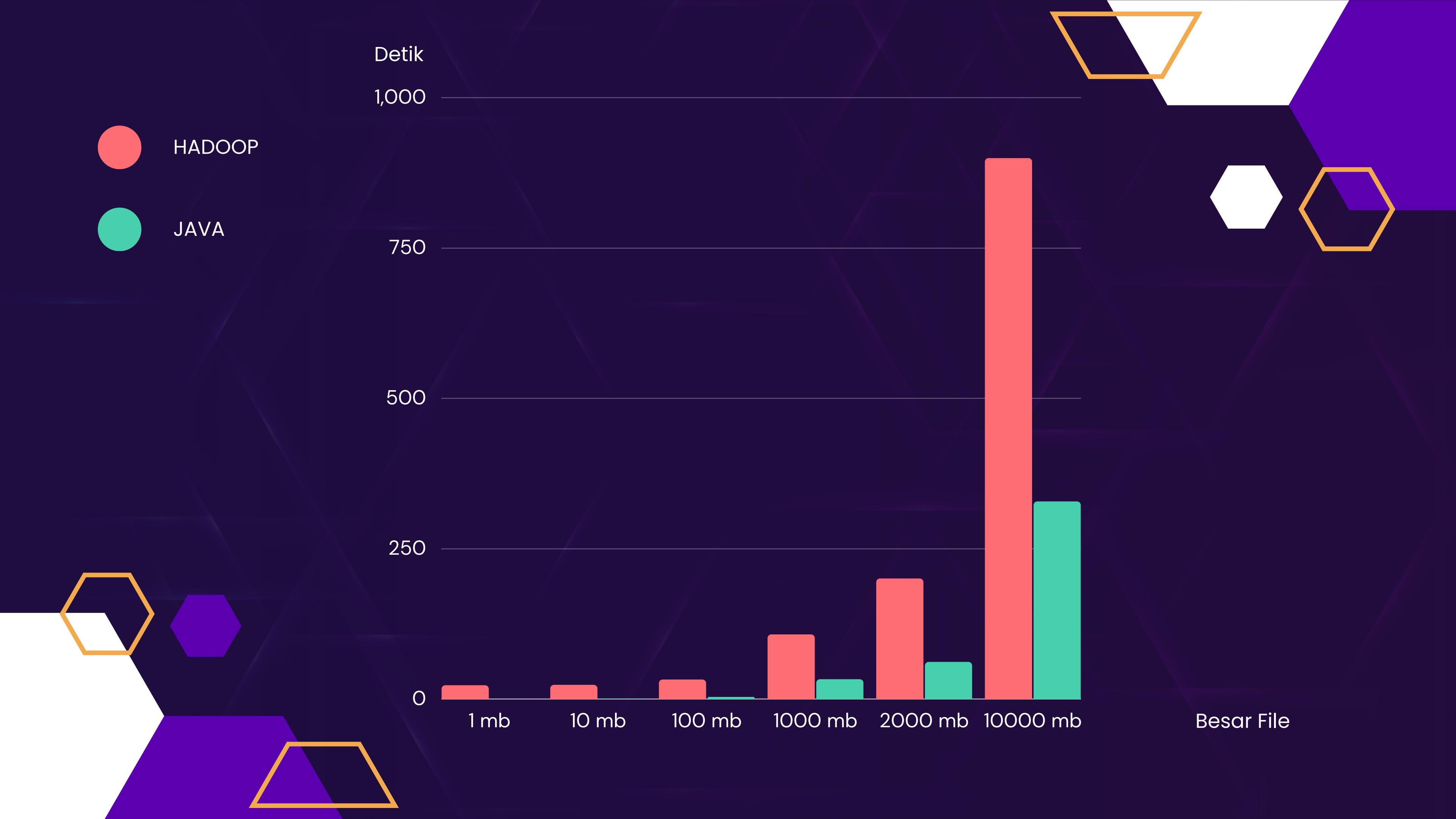
Pada wordcount dengan menggunakan Java dan file sebesar 1000 mb, diperoleh running time sebesar:

**328.50 Detik**

# GRAFIK

Comparasion





# Analisis

## Processing Speed

Java: lebih cepat dibandingkan dengan Hadoop untuk semua ukuran file.

Hadoop: Dibutuhkan lebih banyak waktu untuk memproses data, yang menunjukkan bahwa Hadoop dalam hal pengaturan dan koordinasi

## Performance

Performa Hadoop dapat ditingkatkan dengan baik dengan ukuran file yang lebih besar, tetapi mungkin ada biaya tambahan dalam menyiapkan infrastruktur Hadoop.

## File Size

Hadoop dan Java sama-sama membutuhkan lebih banyak waktu untuk memproses data. Menandakan file yang lebih besar membutuhkan lebih banyak sumber daya komputasi dan waktu untuk membaca, memproses, dan menulis data.

# Java vs Hadoop

## Java

- Waktu pemrosesan Java yang lebih cepat
- Performa lebih baik dalam simple task
- Lebih mudah didevelop
- Efisien untuk infrastruktur yang lebih sederhana
- Lebih cocok untuk file dataset dengan skala kecil-menengah

## Hadoop

- Hadoop dirancang untuk pemrosesan terdistribusi dan unggul dalam menangani set data berskala besar, overhead dan setup kompleksitas mungkin kurang untuk task wordcount
- Lebih unggul jika dalam dataset dengan infrastruktur dan konfigurasi yang kompleks
- Skalabilitas dan kapabilitas computing yang terdistribusi