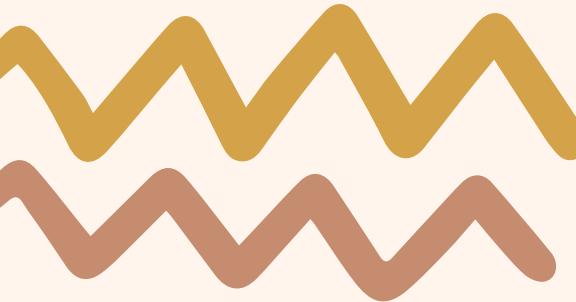


# Hadoop

K1 Group 5

START



# Our Team

Armond Harer

(2106634710)

Enricco Verindra Putra

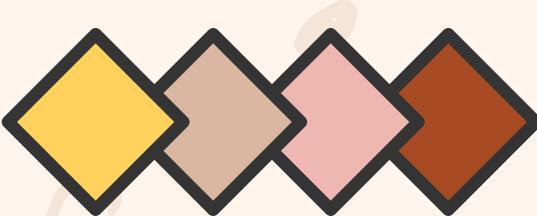
(2106651793)

Fatima Khairunnisa

(2106651515)

Zefanya Christira Deardo

(2106637214)

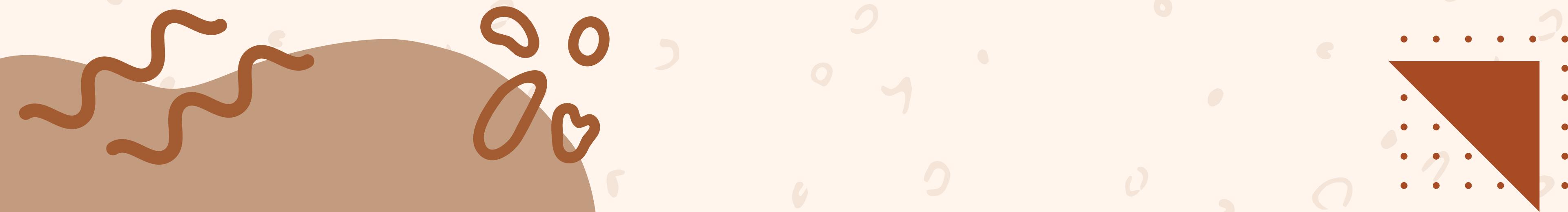


# Topic



“ Topic 1: Instalasi

“ Topic 2: Word Count



# Instalasi Hadoop

Pastikan sudah menginstall Java

```
C:\Users\Zefanya>java --version
java 11.0.16.1 2022-08-18 LTS
Java(TM) SE Runtime Environment 18.9 (build 11.0.16.1+1-LTS-1)
Java HotSpot(TM) 64-Bit Server VM 18.9 (build 11.0.16.1+1-LTS-1, mixed mode)
```

Download Hadoop (tar.gz) dan aditional binaries

(<https://archive.apache.org/dist/hadoop/common/>)

(<https://github.com/cdarlint/winutils>)

# Configurasi System Variable

System Properties --> Environment Variable --> System Variable  
--> New --> Masukkan Variable JAVA\_HOME dengan Value  
berupa path dari instalasi JAVA

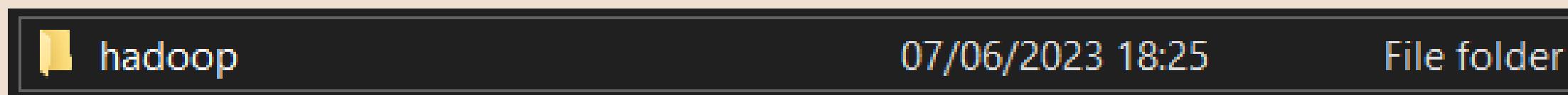
System variables	
Variable	Value
JAVA_HOME	C:\Program Files\Java\jdk-11.0.16.1

System Variable --> Path--> Edit --> Masukkan Path baru dengan  
Value menuju bin dari Java

System variables	
Variable	Value
Path	C:\Python311\Scripts\;C:\Python311\;C:\Program Files\Commo...
C:\Program Files\Java\jdk-11.0.16.1\bin	

# Extract Hadoop

- Pastikan directory dari hadoop tidak mengandung spasi
- (C:\Program Files\hadoop) akan menimbulkan masalah, karena terdapat spasi pada Program Files.



## Edit core-site.xml (/etc/hadoop)

A screenshot of a code editor displaying the contents of the 'core-site.xml' file. The file is an XML document with the following structure:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

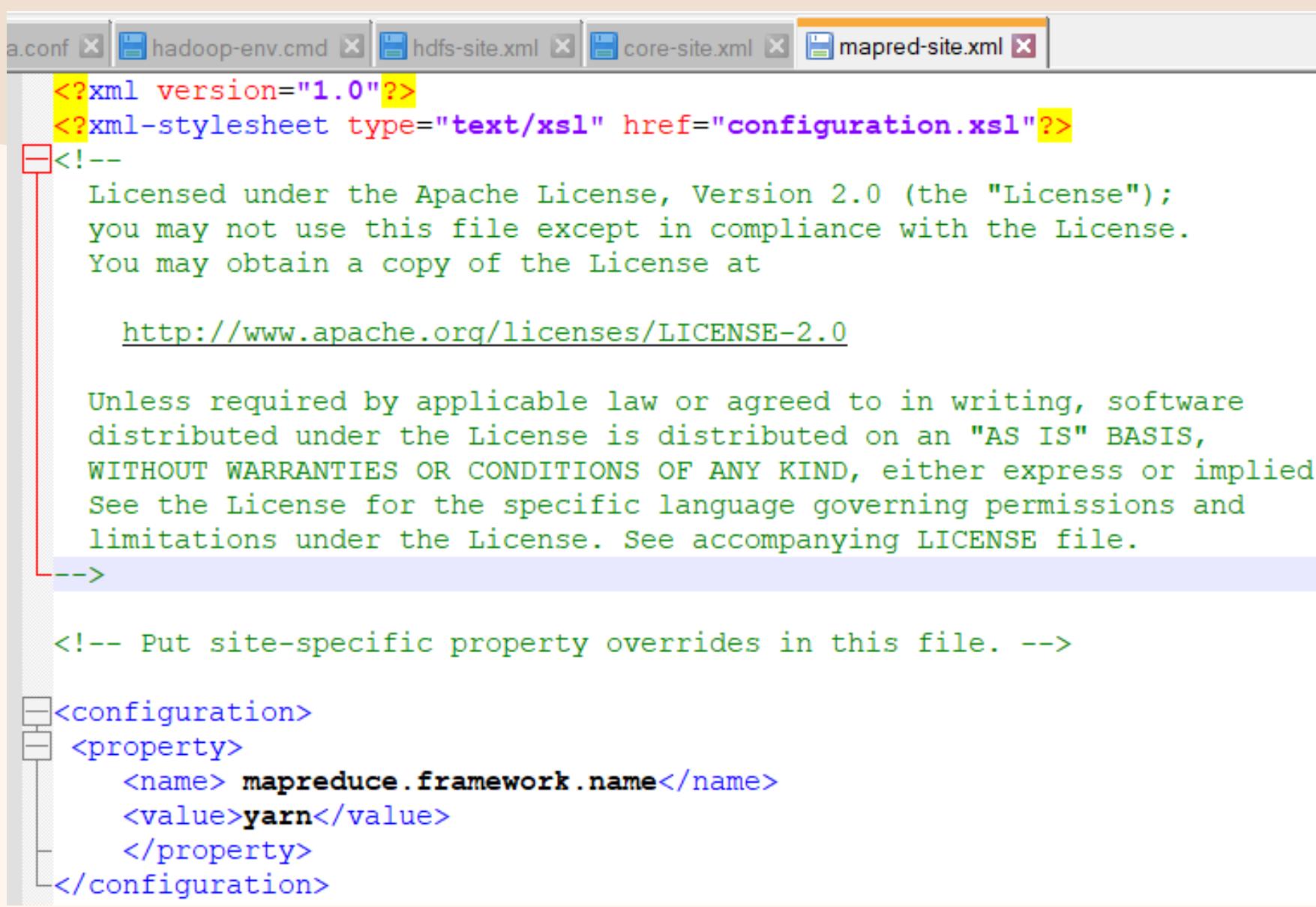
http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name> fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

# Edit mapred-site.xml dan yarn-site.xml (/etc/hadoop)



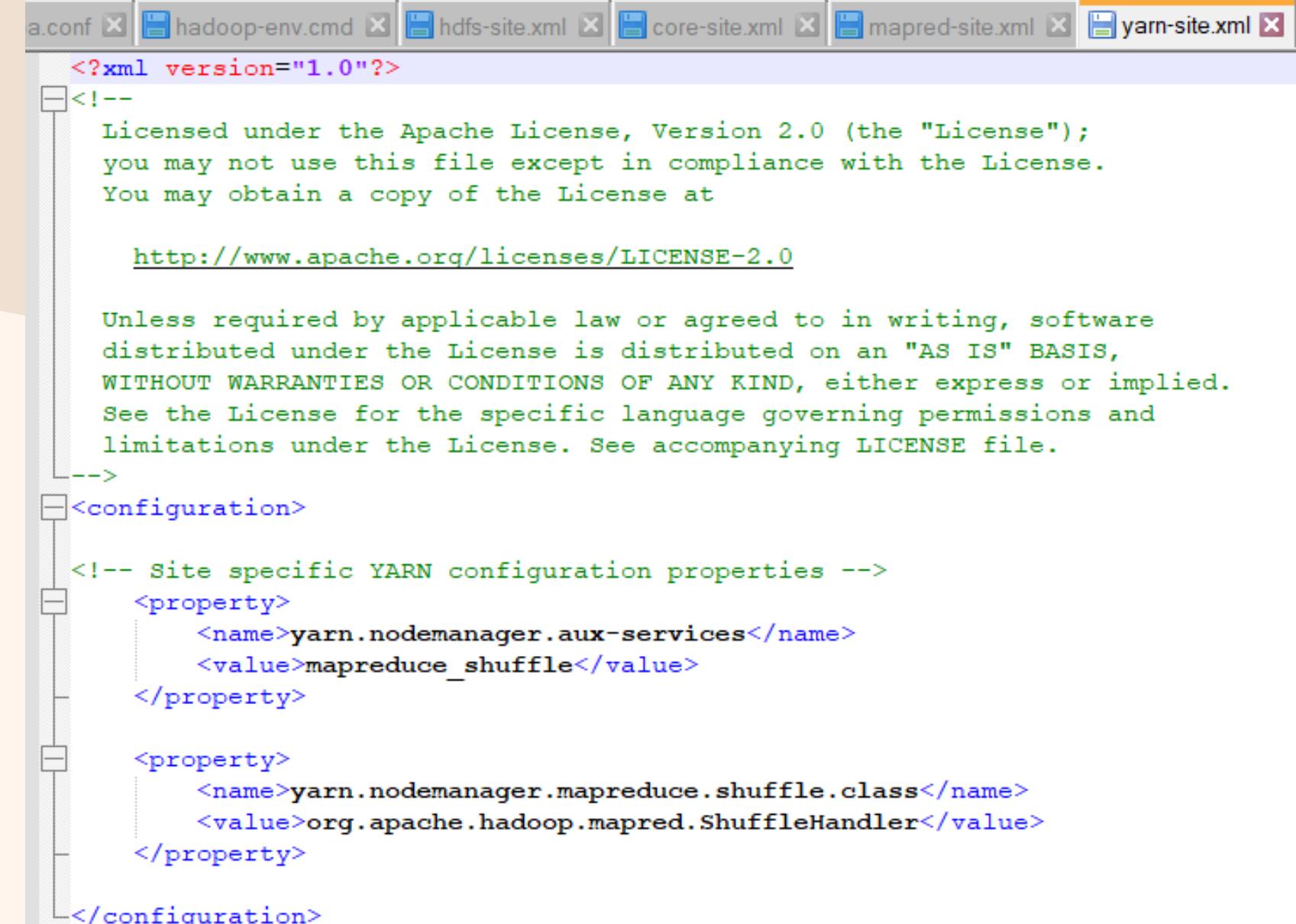
```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name> mapreduce.framework.name</name>
<value>yarn</value>
</property>
</configuration>
```

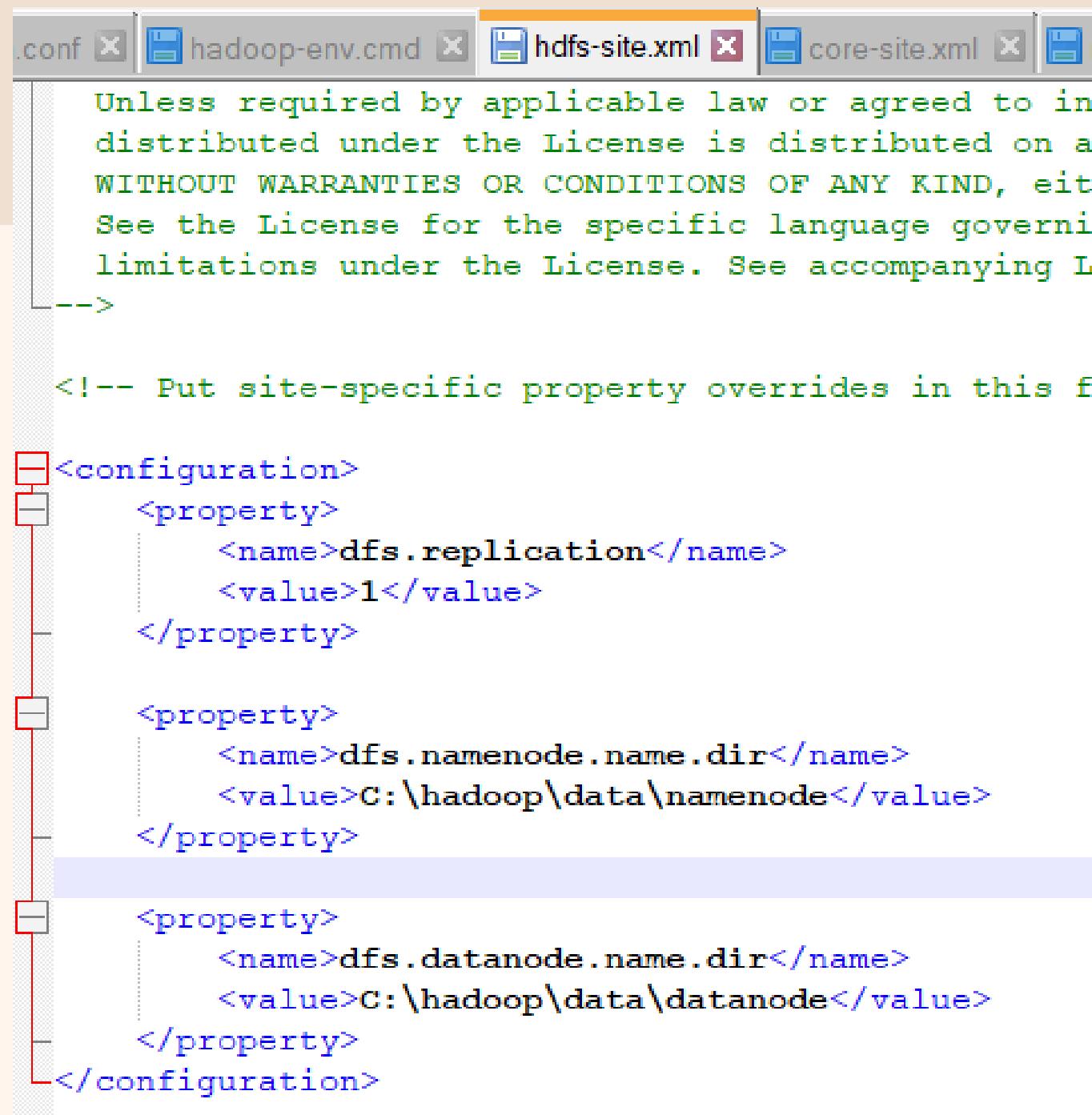
  


```
<?xml version="1.0"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.mapreduce.shuffle.class</name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
</configuration>
```

# Edit hdfs-site.xml dan hadoop-env.cmd (/etc/hadoop)



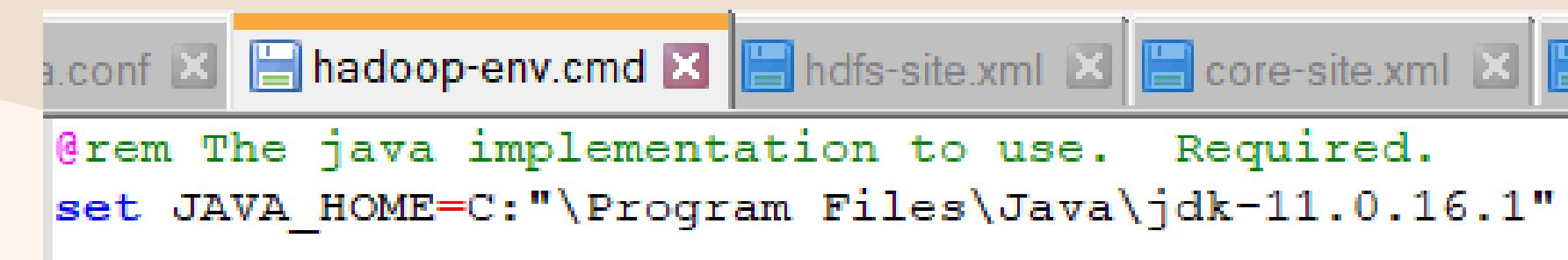
```
Unless required by applicable law or agreed to in
distributed under the License is distributed on an
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either
See the License for the specific language governing
limitations under the License. See accompanying LI
-->

<!-- Put site-specific property overrides in this fi

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>

  <property>
    <name>dfs.namenode.name.dir</name>
    <value>C:\hadoop\data\namenode</value>
  </property>

  <property>
    <name>dfs.datanode.name.dir</name>
    <value>C:\hadoop\data\datanode</value>
  </property>
</configuration>
```

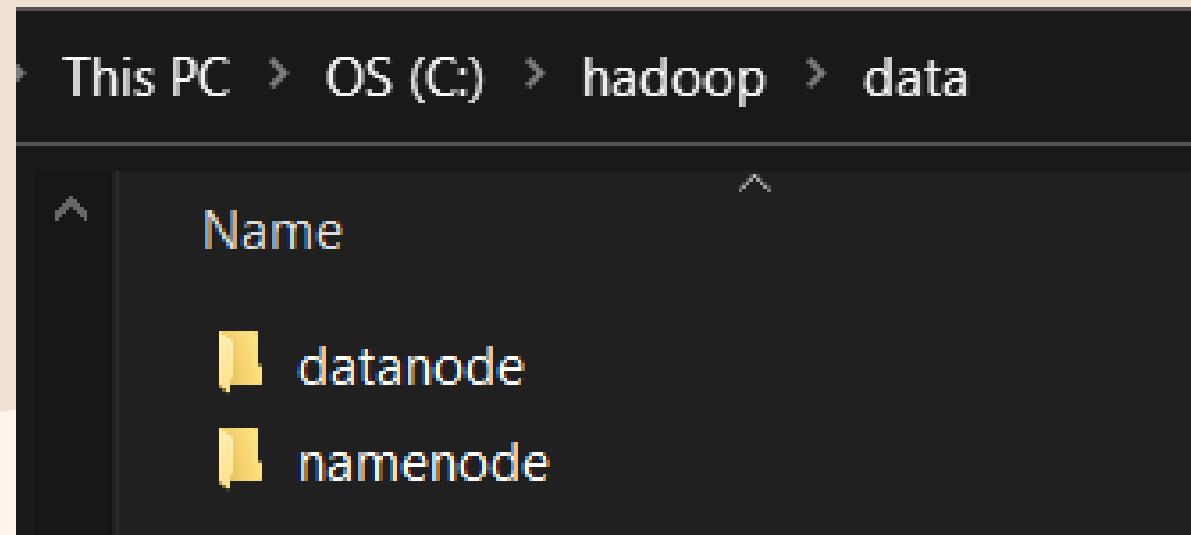


```
@rem The java implementation to use. Required.
set JAVA_HOME=C:"\Program Files\Java\jdk-11.0.16.1"
```

JAVA\_HOME di set sesuai dengan file pathnya

# Data Directory

Membuat folder baru yang bernama data dan berisi folder datanode dan namenode



# Configurasi System Variable

Masukkan Variable HADOOP\_HOME dengan Value berupa path ke HADOOP & directory bin dan sbin pada PATH

System variables	
Variable	Value
HADOOP_HOME	C:\hadoop

C:\hadoop\bin
C:\hadoop\sbin

# Additional bin

Masukkan additional in ke folder bin dan replace yang sama

Name	Date modified	Type	Size
container-executor	03/01/2021 16:54	File	433 KB
hadoop	29/09/2021 1:12	File	9 KB
hadoop	29/09/2021 1:12	Windows Comma...	11 KB
hadoop.dll	29/09/2021 1:12	Application extens...	94 KB
hadoop.exp	29/09/2021 1:12	EXP File	25 KB
hadoop.lib	29/09/2021 1:12	LIB File	41 KB
hadoop.pdb	29/09/2021 1:12	PDB File	820 KB
hdfs	29/09/2021 1:12	File	12 KB
hdfs	29/09/2021 1:12	Windows Comma...	8 KB
libwinutils.lib	29/09/2021 1:12	LIB File	1.561 KB
mapred	29/09/2021 1:12	File	7 KB
mapred	29/09/2021 1:12	Windows Comma...	6 KB
oom-listener	03/01/2021 16:54	File	29 KB
test-container-executor	03/01/2021 16:54	File	474 KB

# Cek Instalasi Hadoop

```
C:\Users\Zefanya>hadoop version
Hadoop 3.2.2
Source code repository Unknown -r 7a3bc90b05f257c8ace2f76d74264906f0f7a932
Compiled by hexiaoqiao on 2021-01-03T09:26Z
Compiled with protoc 2.5.0
From source with checksum 5a8f564f46624254b27f6a33126ff4
This command was run using /C:/hadoop/share/hadoop/common/hadoop-common-3.2.2.jar
```

## Format namenode

```
C:\Users\Zefanya>hdfs namenode format
```

# Menjalankan Hadoop



Menggunakan start-all.cmd (atau start-dfs.cmd lalu start-yarn.cmd) untuk memulai dan stop-all.cmd untuk memberhentikan (/sbin)

start-all	03/01/2021 16:28	Shell Script	3 KB
start-balancer	03/01/2021 16:32	Shell Script	2 KB
start-dfs	03/01/2021 16:32	Windows Comma...	2 KB
start-dfs	03/01/2021 16:32	Shell Script	6 KB
start-secure-dns	03/01/2021 16:32	Shell Script	2 KB
start-yarn	03/01/2021 16:54	Windows Comma...	2 KB
start-yarn	03/01/2021 16:54	Shell Script	4 KB
stop-all	03/01/2021 16:28	Windows Comma...	2 KB
stop-all	03/01/2021 16:28	Shell Script	3 KB
stop-balancer	03/01/2021 16:32	Shell Script	2 KB
stop-dfs	03/01/2021 16:32	Windows Comma...	2 KB
stop-dfs	03/01/2021 16:32	Shell Script	4 KB
stop-secure-dns	03/01/2021 16:32	Shell Script	2 KB
stop-yarn	03/01/2021 16:54	Windows Comma...	2 KB



# Menjalankan Hadoop

Cek GUI dan Resource Manager pada localhost:9870 dan localhost:8088

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

### Overview 'localhost:9000' (active)

Started:	Wed Jun 21 18:18:47 +0700 2023
Version:	3.2.2, r7a3bc90b05f257c8ace2f76d74264906f0f7a932
Compiled:	Sun Jan 03 16:26:00 +0700 2021 by hexiaoqiao from branch-3.2.2
Cluster ID:	CID-8b22e1ac-91de-471e-8e4a-30345b533ead
Block Pool ID:	BP-791238868-192.168.241.1-1686143026620

### Summary

Security is off.  
Safemode is off.  
3 files and directories, 1 blocks (1 replicated blocks, 0 erasure coded block groups) = 4 total filesystem object(s).  
Heap Memory used 110.19 MB of 254 MB Heap Memory. Max Heap Memory is 1000 MB.  
Non Heap Memory used 59.1 MB of 62.1 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	237.86 GB
Configured Remote Capacity:	0 B

 All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Re...
0	0	0	0	0	<memory:0, vCores:0>	<memory:8192,...

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>
<memory:1024, vCores:1>		

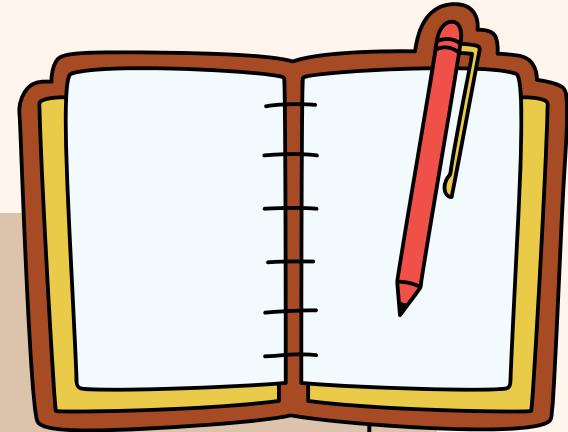
Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB
No data available in table													

Showing 0 to 0 of 0 entries



# Spesifikasi Hardware



Processor	11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz
Installed RAM	16,0 GB (15,8 GB usable)

## Virtual Machine:

The screenshot shows a virtual machine configuration window with two main sections: "General" and "System".

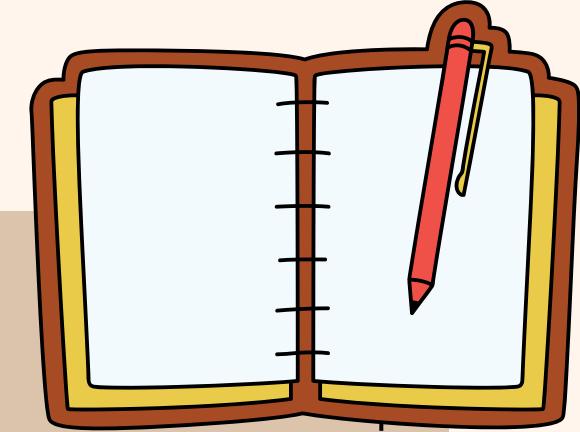
**General**

- Name: Ubuntu 20.4 kelas OS
- Operating System: Ubuntu 20.04 LTS (Focal Fossa) (64-bit)

**System**

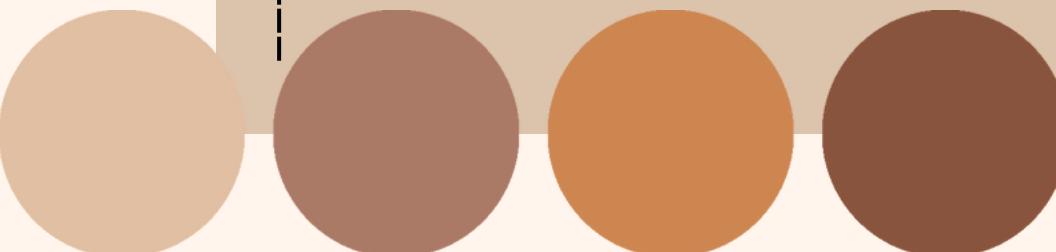
- Base Memory: 5120 MB
- Processors: 4
- Boot Order: Floppy, Optical, Hard Disk
- Acceleration: Nested Paging, KVM Paravirtualization

# Word Count



Word count akan menghitung jumlah semua kata yang ada dalam sebuah teks.

Waktu yang dibutuhkan untuk melakukan Word Count Pada Hadoop dan pada Java dengan Ukuran file 1 , 10, 100, dan 1000 mb akan dibandingkan



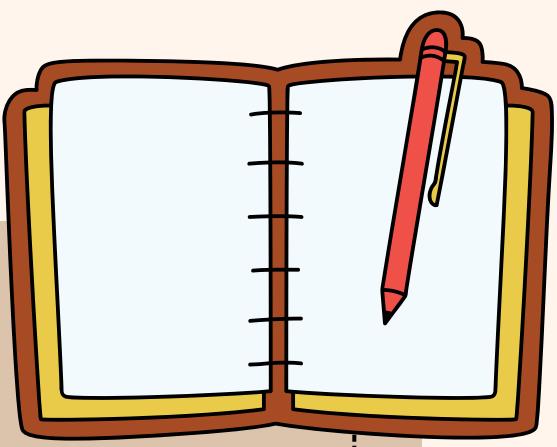
# Code:

```
import java.io.File;
import java.io.FileNotFoundException;
import java.util.HashMap;
import java.util.Map;
import java.util.Scanner;

public class CountWord {
    public static void countWords(String filename, Map<String, Integer> words) throws FileNotFoundException {
        Scanner file = new Scanner(new File(filename));
        while (file.hasNext()) {
            String word = file.next();
            Integer count = words.get(word);
            if (count != null)
                count++;
            else
                count = 1;
            words.put(word, count);
        }
        file.close();
    }

    Run | Debug
    public static void main(String[] args) {
        Map<String, Integer> words = new HashMap<>();
        try {
            countWords(filename:"data.txt", words);
            System.out.println(words);
        } catch (FileNotFoundException e) {
            System.out.println("File not found: " + e.getMessage());
        }
    }
}
```

## Referensi Code:



# Output Java:

1mb

```
real    0m0,325s
user    0m0,493s
sys     0m0,205s
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=1158, play=1158, been=2316, tell=4632, what's=1158,
do=579, (Give=579, down=3474, We've=1158, We're=579, Neve
=3474, up=3474, if=579, you=20265, heart's=1158, give=405
tand=1158, feeling=1737, it=2316, strangers=579, wanna=11
158, never=579, aching=1158, And=579, commitment's=579, w
, full=579, love=579, no=579, A=579, other=1737, game=115
158, goodbye=3474, for=1158, I=1737, Inside=1158, guy=579
ng=1158, we=1158, how=1737, see=579, known=1158, too=1737
u're=1737, so=1737, just=1158, make=4632, on=1158, wouldn
a=22581, give,=579, say=4632, cry=3474, any=579, We=1158,
sk=579, Your=1158, to=2316, desert=3474, You=1158}

real    0m0,338s
user    0m0,510s
sys     0m0,159s
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=1158, play=1158, been=2316, tell=4632, what's=1158,
do=579, (Give=579, down=3474, We've=1158, We're=579, Neve
=3474, up=3474, if=579, you=20265, heart's=1158, give=405
tand=1158, feeling=1737, it=2316, strangers=579, wanna=11
158, never=579, aching=1158, And=579, commitment's=579, w
, full=579, love=579, no=579, A=579, other=1737, game=115
158, goodbye=3474, for=1158, I=1737, Inside=1158, guy=579
ng=1158, we=1158, how=1737, see=579, known=1158, too=1737
u're=1737, so=1737, just=1158, make=4632, on=1158, wouldn
a=22581, give,=579, say=4632, cry=3474, any=579, We=1158,
sk=579, Your=1158, to=2316, desert=3474, You=1158}

real    0m0,333s
user    0m0,427s
sys     0m0,271s
```

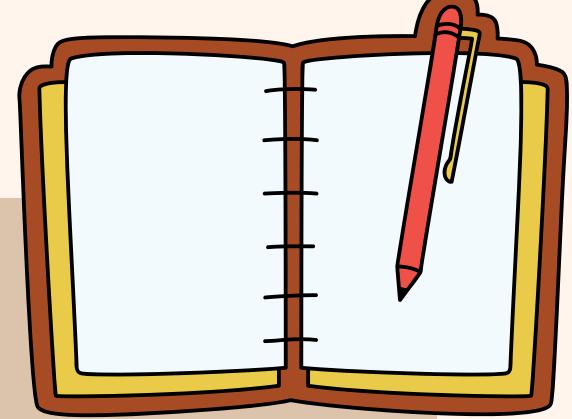
10mb

```
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=11586, play=11586, been=23172, tell=46344, what's=
6, We're=5793, Never=208548, me=11586, from=5793, let=
93, understand=11586, feeling=17379, it=23172, strange=
d=5793, commitment's=5793, we're=11586, know=28965, hu=
Gotta=11586, shy=11586, goodbye=34758, for=11586, I=1
17379, see=5793, known=11586, too=17379, and=81102, of=
793, a=28965, Don't=5793, gonna=225927, give,=5793, sa=
r=11586, to=23172, desert=34758, You=11586}

real    0m1,160s
user    0m1,315s
sys     0m0,234s
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=11586, play=11586, been=23172, tell=46344, what's=
6, We're=5793, Never=208548, me=11586, from=5793, let=
93, understand=11586, feeling=17379, it=23172, strange=
d=5793, commitment's=5793, we're=11586, know=28965, hu=
Gotta=11586, shy=11586, goodbye=34758, for=11586, I=1
17379, see=5793, known=11586, too=17379, and=81102, of=
793, a=28965, Don't=5793, gonna=225927, give,=5793, sa=
r=11586, to=23172, desert=34758, You=11586}

real    0m1,106s
user    0m1,224s
sys     0m0,286s
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=11586, play=11586, been=23172, tell=46344, what's=
6, We're=5793, Never=208548, me=11586, from=5793, let=
93, understand=11586, feeling=17379, it=23172, strange=
d=5793, commitment's=5793, we're=11586, know=28965, hu=
Gotta=11586, shy=11586, goodbye=34758, for=11586, I=1
17379, see=5793, known=11586, too=17379, and=81102, of=
793, a=28965, Don't=5793, gonna=225927, give,=5793, sa=
r=11586, to=23172, desert=34758, You=11586}

real    0m1,105s
user    0m0,946s
sys     0m0,584s
```



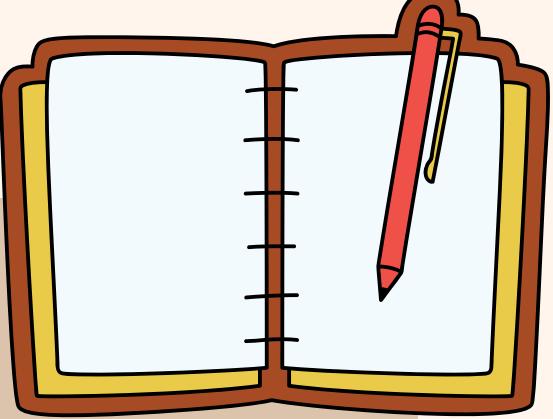
# Output Java:

100mb

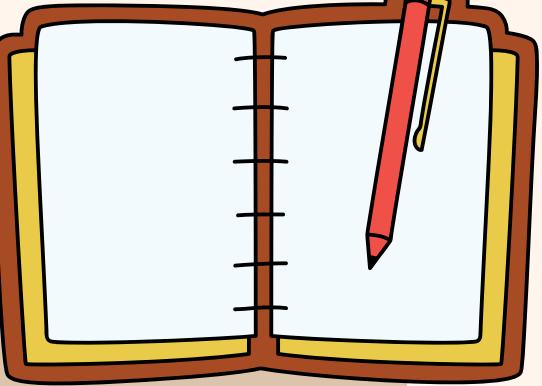
```
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=115874, play=115874, been=231748, tell=463496, wha
 We've=115874, We're=57937, Never=2085732, me=115874, f
 , going=115874, this=57937, understand=115874, feeling=
74, never=57937, aching=115874, And=57937, commitment's
57937, other=173811, game=115874, blind=57937, Gotta=11
inking=57937, around=347622, long=115874, we=115874, ho
re=173811, so=173811, just=115874, make=463496, on=1158
ry=347622, any=57937, We=115874, up)=57937, the=173811,
real    0m6,662s
user    0m6,067s
sys     0m1,013s
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=115874, play=115874, been=231748, tell=463496, wha
 We've=115874, We're=57937, Never=2085732, me=115874, f
 , going=115874, this=57937, understand=115874, feeling=
74, never=57937, aching=115874, And=57937, commitment's
57937, other=173811, game=115874, blind=57937, Gotta=11
inking=57937, around=347622, long=115874, we=115874, ho
re=173811, so=173811, just=115874, make=463496, on=1158
ry=347622, any=57937, We=115874, up)=57937, the=173811,
real    0m6,636s
user    0m6,662s
sys     0m0,420s
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=115874, play=115874, been=231748, tell=463496, wha
 We've=115874, We're=57937, Never=2085732, me=115874, f
 , going=115874, this=57937, understand=115874, feeling=
74, never=57937, aching=115874, And=57937, commitment's
57937, other=173811, game=115874, blind=57937, Gotta=11
inking=57937, around=347622, long=115874, we=115874, ho
re=173811, so=173811, just=115874, make=463496, on=1158
ry=347622, any=57937, We=115874, up)=57937, the=173811,
real    0m6,515s
user    0m6,574s
sys     0m0,339s
```

1000mb

```
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=1158748, play=1158748, been=2317496, tell=4634992
wn=3476244, We've=1158748, We're=579374, Never=2085746
=1158748, give=4055618, going=1158748, this=579374, un
317496, each=1158748, both=1158748, never=579374, achi
0, full=579374, love=579374, no=579374, A=579374, othe
or=1158748, I=1738122, Inside=1158748, guy=579374, thi
n=1158748, too=1738122, and=8111236, of=579374, get=57
9374, a=2896870, Don't=579374, gonna=22595586, give,=5
t=579374, ask=579374, Your=1158748, to=2317496, desert
real    1m1,526s
user    1m0,545s
sys     0m1,835s
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=1158748, play=1158748, been=2317496, tell=4634992
wn=3476244, We've=1158748, We're=579374, Never=2085746
=1158748, give=4055618, going=1158748, this=579374, un
317496, each=1158748, both=1158748, never=579374, achi
0, full=579374, love=579374, no=579374, A=579374, othe
or=1158748, I=1738122, Inside=1158748, guy=579374, thi
n=1158748, too=1738122, and=8111236, of=579374, get=57
9374, a=2896870, Don't=579374, gonna=22595586, give,=5
t=579374, ask=579374, Your=1158748, to=2317496, desert
real    1m2,182s
user    1m1,726s
sys     0m1,172s
hadoopusr@VirtualBox:~/Downloads$ time java Wordcount
{but=1158748, play=1158748, been=2317496, tell=4634992
wn=3476244, We've=1158748, We're=579374, Never=2085746
=1158748, give=4055618, going=1158748, this=579374, un
317496, each=1158748, both=1158748, never=579374, achi
0, full=579374, love=579374, no=579374, A=579374, othe
or=1158748, I=1738122, Inside=1158748, guy=579374, thi
n=1158748, too=1738122, and=8111236, of=579374, get=57
9374, a=2896870, Don't=579374, gonna=22595586, give,=5
t=579374, ask=579374, Your=1158748, to=2317496, desert
real    1m2,745s
user    1m1,997s
sys     0m1,532s
```



# Output Hadoop



**1mb:**

Total time spent by all map tasks (ms)=2591

Total time spent by all reduce tasks (ms)=2220

**10mb:**

Total time spent by all map tasks (ms)=4058

Total time spent by all reduce tasks (ms)=1989

**100mb:**

Total time spent by all map tasks (ms)=13979

Total time spent by all reduce tasks (ms)=1856

**1000mb:**

Total time spent by all map tasks (ms)=369406

Total time spent by all reduce tasks (ms)=28930

# Hasil

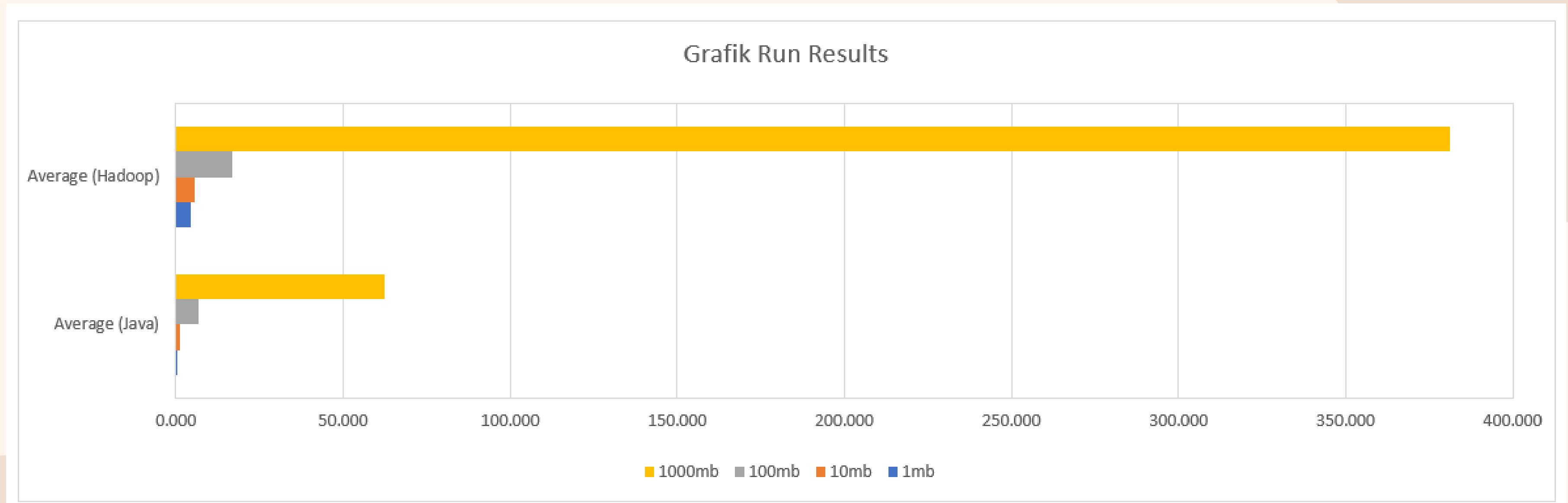
## Java

File Size	Java (Real Time)				
	Run 1 (Java)	Run 2 (Java)	Run 3 (Java)	Run 4 (Java)	Average (Java)
1mb	0,372	0,325	0,338	0,333	0,342
10mb	1,160	1,106	1,105	0,870	1,060
100mb	6,662	6,636	6,515	6,367	6,545
1000mb	61,526	62,182	62,745	62,232	62,171

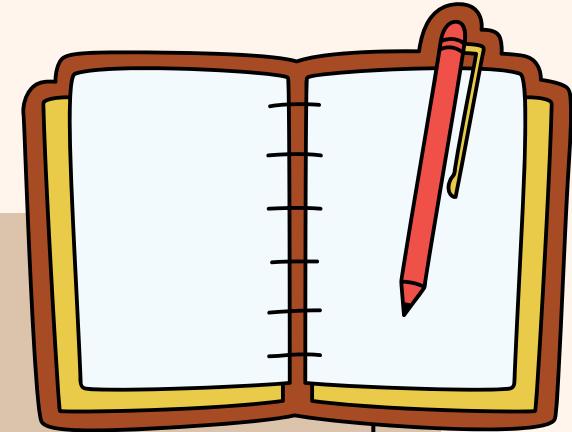
## Hadoop

File Size	Hadoop (Total mapreduce time)				
	Run 1 (Hadoop)	Run 2 (Hadoop)	Run 3 (Hadoop)	Run 4 (Hadoop)	Average (Hadoop)
1mb	4,811	4,587	3,897	3,891	4,297
10mb	6,047	5,073	5,770	5,429	5,580
100mb	15,835	17,951	17,310	16,930	17,007
1000mb	398,336	387,278	386,159	353,332	381,276

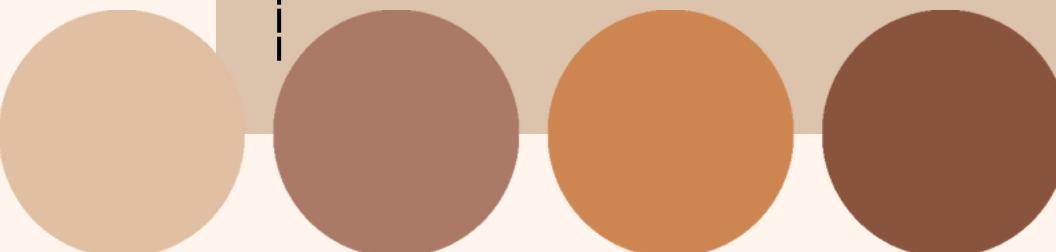
# Grafik



# Kesimpulan



Eksperimen tersebut mengindikasikan bahwa algoritma wordcount yang diterapkan menggunakan *Java Programming Language* mempunyai kecepatan eksekusi yang lebih tinggi dibandingkan dengan Hadoop saat eksperimen tersebut diuji coba menggunakan file berukuran 1MB, 10MB, 100MB, dan 1000MB. Meskipun demikian, perbedaan yang mencolok antara kedua cara tersebut berada pada output yang berupa file .txt. Dimana, dengan pemrograman Java, menghasilkan file .txt tanpa kata-kata yang diurut sesuai alphabet, sedangkan Hadoop menghasilkan output yang terurut rapi dan teratur. Hal ini juga berdampak pada kecepatan eksekusi algoritma wordcount dalam Java jika dibandingkan dengan algoritma word count dalam Hadoop.



# Github

## ArmondHarer/ HadoopSBD



1 Contributor    0 Issues    0 Stars    0 Forks



**ArmondHarer/HadoopSBD**  
Contribute to ArmondHarer/HadoopSBD development by creating an account on GitHub.  


THANK YOU  
SO MUCH!

