Tutorial Instalasi Hadoop

1. Masuk ke website oracle berikut :

   https://www.oracle.com/java/technologies/javase/javase-jdk8-downloads.html

2. Download Java Development Kit 8 (JDK 8) windows x64 kemudian install.

Linux	macOS	Solaris	**Windows**		
Product/file description			File size		Download
x86 Installer			136.77 MB		🔒 jdk-8u371-windows-i586.exe
x64 Installer			145.50 MB		🔒 jdk-8u371-windows-x64.exe

3. Set Environment Variables (Java)

   Tambahkan pada system variable dengan nama JAVA_HOME, kemudian dengan value yaitu path dari folder java yang telah terinstall.

   System variables

Variable	Value
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
HADOOP_HOME	C:\hadoop-3.2.2
JAVA_HOME	C:\Program Files\Java\jdk-1.8
NUMBER_OF_PROCESSORS	8
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\javapath;C:\Prog...

   Kemudian tambahkan /bin pada variable Path.

C:\Program Files\nodejs\
C:\Program Files\Java\jdk-1.8\bin
C:\hadoop-3.2.2\bin

4. Verifikasi Java

```
C:\Users\Asus>java -version
java version "1.8.0_371"
Java(TM) SE Runtime Environment (build 1.8.0_371-b11)
Java HotSpot(TM) 64-Bit Server VM (build 25.371-b11, mixed mode)
```
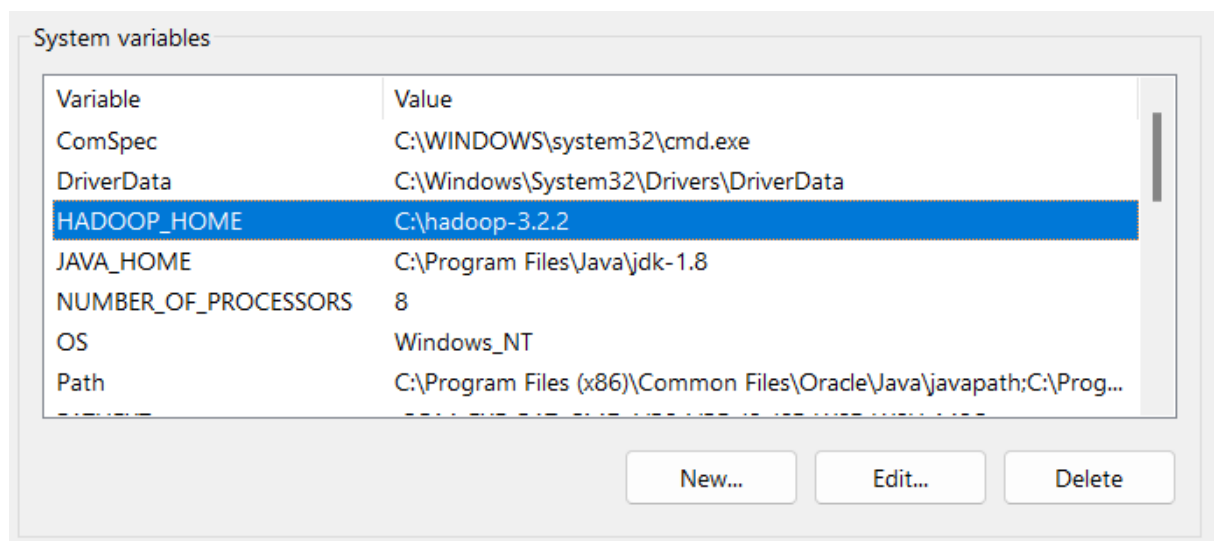
5. Download Hadoop (Recommended ver. 3.2.2)

   https://archive.apache.org/dist/hadoop/common/

hadoop-3.2.2-src.tar.gz.sha512	2021-01-13 18:48	154
hadoop-3.2.2.tar.gz	2021-01-13 18:48	377M
hadoop-3.2.2.tar.gz.asc	2021-01-13 18:48	833

   Pilih hadoop-3.2.2.tar.gz. kemudian extract file hadoop yang telah di download

6. Set Environment Variables (Hadoop)

System variables

Variable	Value
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
HADOOP_HOME	C:\hadoop-3.2.2
JAVA_HOME	C:\Program Files\Java\jdk-1.8
NUMBER_OF_PROCESSORS	8
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\javapath;C:\Prog...

New...   Edit...   Delete

   Kemudian tambahkan /bin dan /sbin pada variable Path.

C:\Program Files\Java\jdk-1.8\bin
C:\hadoop-3.2.2\bin
C:\hadoop-3.2.2\sbin

7. Tambahkan folder data pada file hadoop yang telah di extract.

kemudian di dalam folder data tersebut buat dua folder dengan nama namenode dan datanode.

8. Edit core-site.xml, hdfs-site.xml, mapred-site.xml, dan yarn-site.xml.

etc/hadoop/core-site.xml :

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

etc/hadoop/hdfs-site.xml :

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

etc/hadoop/mapred-site.xml :

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

<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>
    </property>
</configuration>
```

etc/hadoop/yarn-site.xml :

```xml
<configuration>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>
    <property>
        <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_HOME,PATH,LANG,TZ,HADOOP_MAPRED_HOME</value>
    </property>
</configuration>
```

9. Edit Hadoop Env pada hadoop env.cmd dengan path folder jdk yang telah didownload

```
@rem The java implementation to use.  Required.
set JAVA_HOME=C:"\Program Files\Java\jdk-1.8"
```

10. Download Patch file hadoop untuk Windows.

https://github.com/cdarlint/winutils

📁	hadoop-3.2.1/bin	add 321 winutils	4 years ago
📁	hadoop-3.2.2/bin	compile hadoop-3.2.2	2 years ago

11. Extract file bin tersebut kemudian pindahkan pada folder bin yang ada pada folder hadoop.

📁 bin	5/25/2023 12:32 AM	File folder
📁 data	5/25/2023 12:23 AM	File folder
📁 etc	5/25/2023 12:13 AM	File folder

12. Verifikasi Hadoop

```
C:\Users\Asus>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-3.2.2/share/hadoop/common/hadoop-common-3.2.2.jar
```

Hadoop WordCount (JAVA)

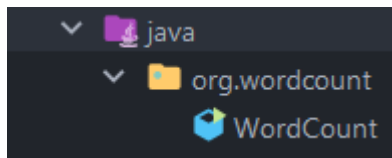1. Create Maven Project (Using IntelliJ) dengan Language Java dan Build System Maven.
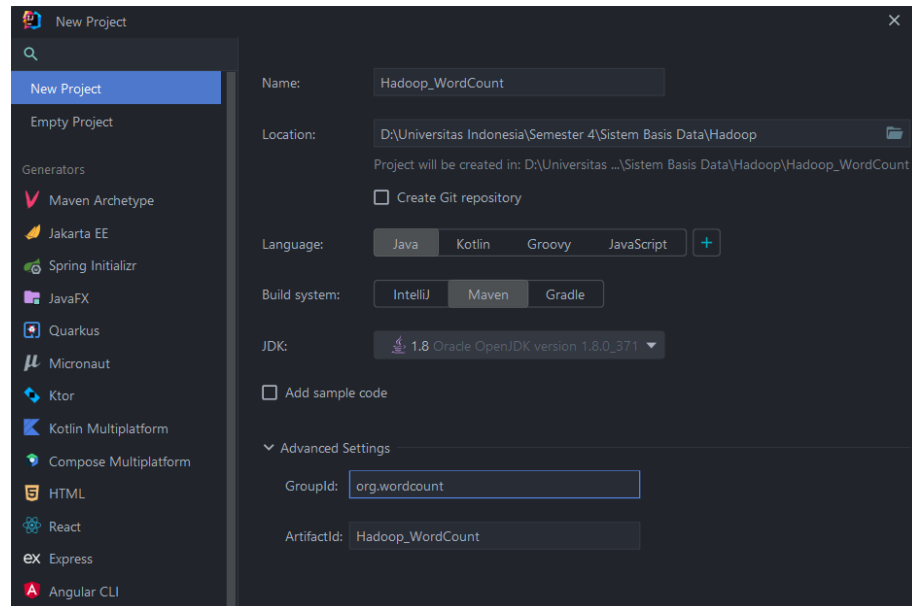
2. Tambahkan dependencies pada pom.xml

```
<dependencies>
    <dependency>
        <groupId>org.apache.hadoop</groupId>
        <artifactId>hadoop-common</artifactId>
        <version>3.2.2</version>
    </dependency>
    <dependency>
        <groupId>org.apache.hadoop</groupId>
        <artifactId>hadoop-mapreduce-client-core</artifactId>
        <version>3.2.2</version>
    </dependency>
</dependencies>
```
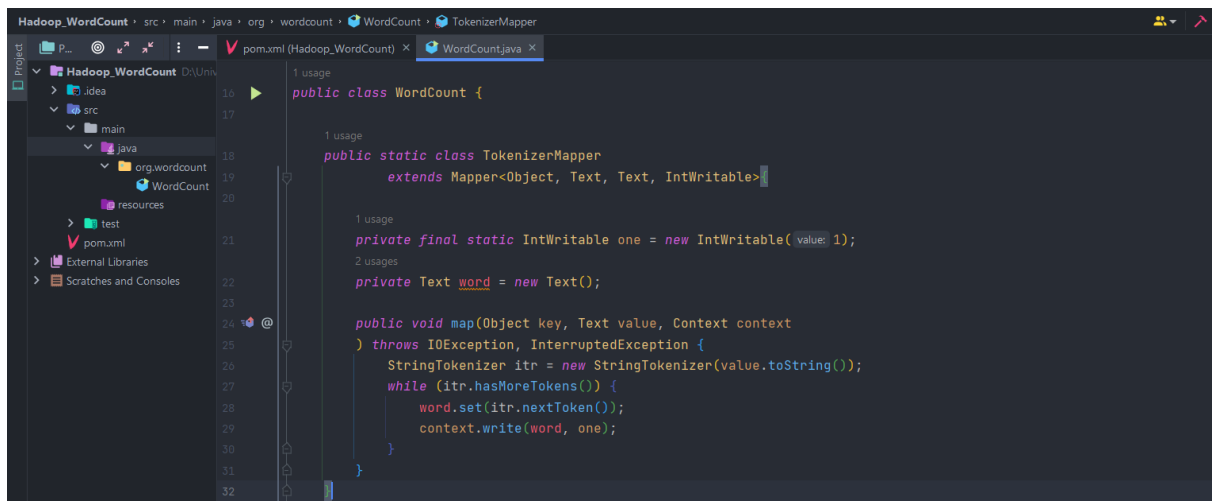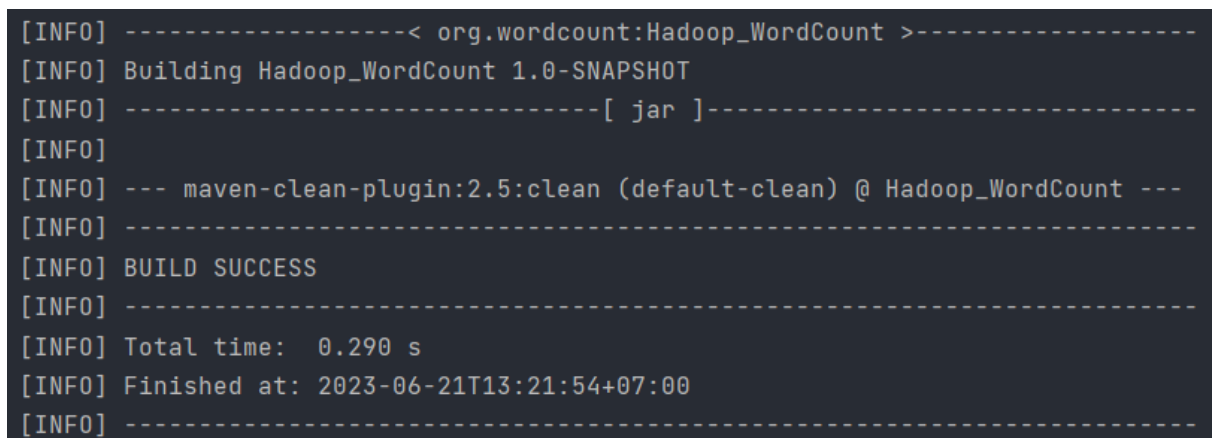
3. Create Package dan didalam package tersebut buat file java berisi kode wordcount.

kodenya dapat dilihat pada link :

https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html#Source_Code

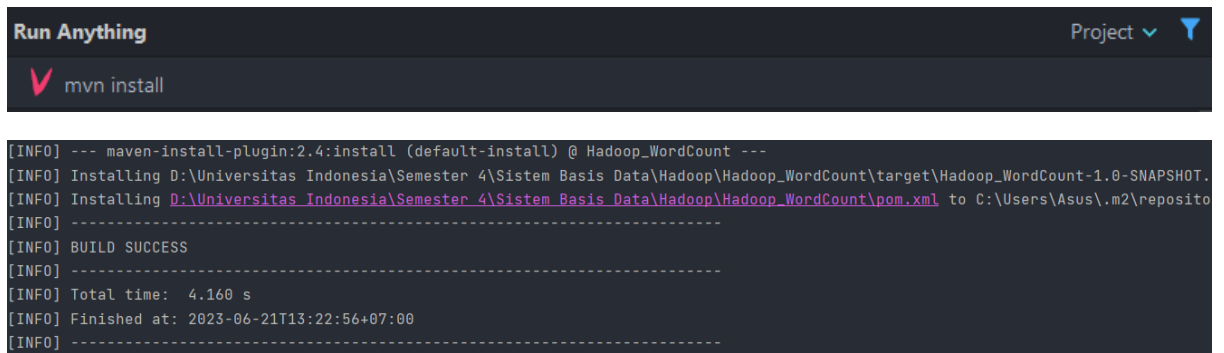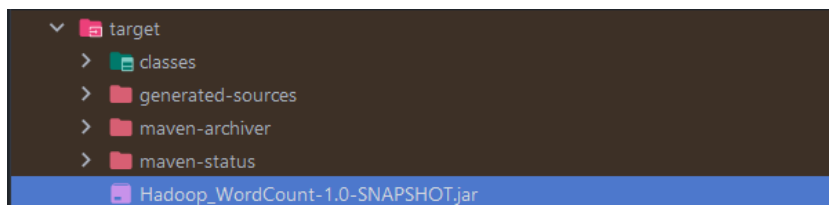4. Kemudian membuat jar file dengan maven.

mvn clean

mvn install

Maka pada target folder terdapat file jar.
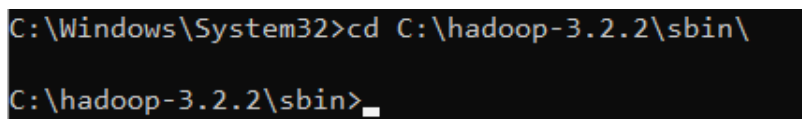

5. Buat input file (.txt)

example :
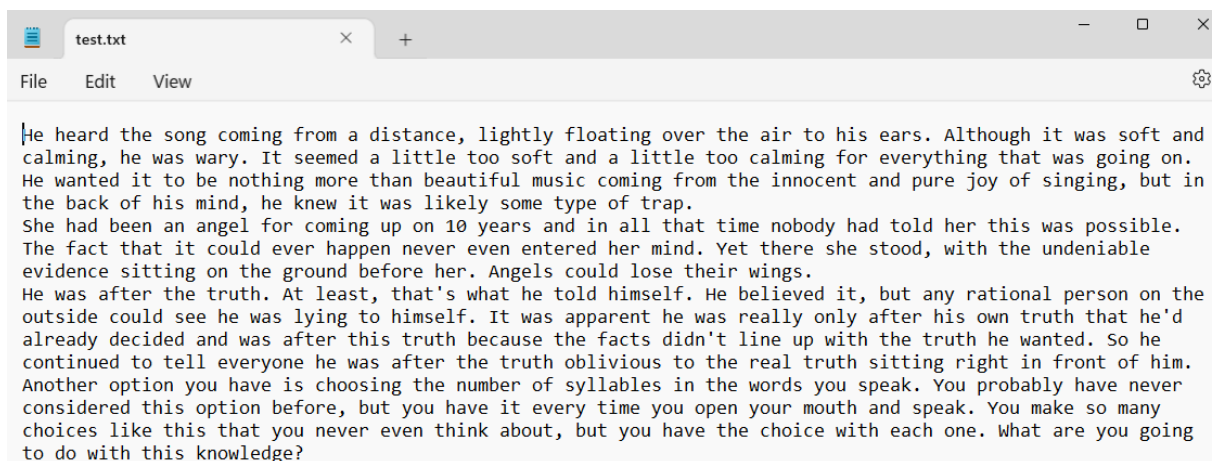
6. run cmd as administrator kemudian cd ke folder sbin

7. Menjalankan Hadoop serta jps untuk memastikan hadoop berjalan semestinya.

start-all.cmd

```
C:\hadoop-3.2.2\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\hadoop-3.2.2\sbin>jps
13008
288 RemoteMavenServer36
11236 NameNode
7332 NodeManager
9224 Jps
13612 DataNode
3180 ResourceManager

C:\hadoop-3.2.2\sbin>
```

hadoop fs -mkdir /testinp

```
C:\hadoop-3.2.2\sbin>hadoop fs -mkdir /testinp
```

hadoop fs -put "D:\Universitas Indonesia\Semester 4\Sistem Basis Data\Hadoop\1M.txt" /testinp

```
C:\hadoop-3.2.2\sbin>hadoop fs -put "D:\Universitas Indonesia\Semester 4\Sistem Basis Data\Hadoop\1M.txt" /testinp
```

hadoop jar "D:\Universitas Indonesia\Semester 4\Sistem Basis Data\Hadoop\Hadoop_WordCount\target\Hadoop_WordCount-1.0-SNAPSHOT.jar" org.wordcount.WordCount /testinp /output_test

```
C:\hadoop-3.2.2\sbin>hadoop jar D:\Universitas Indonesia\Semester 4\Sistem Basis Data\Hadoop\Hadoop_WordCount\target\Hadoop_WordCount-1.0-SNAPSHOT.jar or
g.wordcount.WordCount /testinp /output_test
JAR does not exist or is not a normal file: D:\Universitas

C:\hadoop-3.2.2\sbin>hadoop jar "D:\Universitas Indonesia\Semester 4\Sistem Basis Data\Hadoop\Hadoop_WordCount\target\Hadoop_WordCount-1.0-SNAPSHOT.jar"
org.wordcount.WordCount /testinp /output_test
2023-06-21 13:37:54,318 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-06-21 13:37:55,072 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute yo
ur application with ToolRunner to remedy this.
2023-06-21 13:37:55,086 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Asus/.staging/job_1687329215005_0
001
2023-06-21 13:37:55,332 INFO input.FileInputFormat: Total input files to process : 1
2023-06-21 13:37:55,397 INFO mapreduce.JobSubmitter: number of splits:1
2023-06-21 13:37:55,560 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1687329215005_0001
2023-06-21 13:37:55,562 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-06-21 13:37:55,752 INFO conf.Configuration: resource-types.xml not found
2023-06-21 13:37:55,753 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-06-21 13:37:56,036 INFO impl.YarnClientImpl: Submitted application application_1687329215005_0001
2023-06-21 13:37:56,092 INFO mapreduce.Job: The url to track the job: http://LAPTOP-7QHLF5HQ:8088/proxy/application_1687329215005_0001/
2023-06-21 13:37:56,092 INFO mapreduce.Job: Running job: job_1687329215005_0001
2023-06-21 13:38:10,335 INFO mapreduce.Job: Job job_1687329215005_0001 running in uber mode : false
2023-06-21 13:38:10,336 INFO mapreduce.Job:   map 0% reduce 0%
2023-06-21 13:38:17,460 INFO mapreduce.Job:   map 100% reduce 0%
2023-06-21 13:38:24,540 INFO mapreduce.Job:   map 100% reduce 100%
```

8. Output

hadoop fs -cat /output_test/part-r-00000

```
C:\hadoop-3.2.2\sbin>hadoop fs -cat /output_test/part-r-00000
"Begin   60
"I'll    60
"It      60
"You     61
"Your    61
"bridge",        60
"easy"  60
(I       60
--       120
1        61
10       61
100      121
100-foot         60
24       60
24-hour 60
25       61
A        182
Actually,        60
After   60
All      120
Although         61
```

stop-all.cmd

```
C:\hadoop-3.2.2\sbin>stop-all.cmd
This script is Deprecated. Instead use stop-dfs.cmd and stop-yarn.cmd
SUCCESS: Sent termination signal to the process with PID 9240.
SUCCESS: Sent termination signal to the process with PID 6824.
stopping yarn daemons
SUCCESS: Sent termination signal to the process with PID 11268.
SUCCESS: Sent termination signal to the process with PID 15348.

INFO: No tasks running with the specified criteria.
```