

# Daftar Isi

---

- Laporan HDFS dan MapReduce
  - Unduh file teks
  - Salin source code
  - Tambah environment variable HADOOP\_CLASSPATH
  - Compile WordCount.java dan buat Jar
  - Jalankan DFS dan YARN
  - Membuat folder input dan memindahkan file teks
  - Jalankan Jar
  - Troubleshoot Error MRAppMaster
  - Tampilan dashboard Hadoop

## Laporan HDFS dan MapReduce

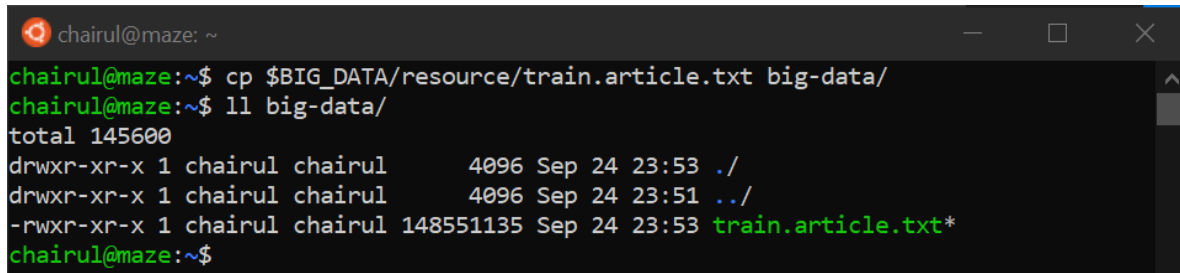
---

- Chairul Imam 1608107010040
- Nani Fiddini 1508107010060

### Unduh file teks

File teks yang akan digunakan dapat diunduh pada URL berikut [train.article.txt](#). Setelah diunduh, duplikasi file tersebut ke WSL Ubuntu dengan perintah

```
$ cp $BIG_DATA/resource/train.article.txt big-data/
```



```
chairul@maze: ~  
chairul@maze:~$ cp $BIG_DATA/resource/train.article.txt big-data/  
chairul@maze:~$ ll big-data/  
total 145600  
drwxr-xr-x 1 chairul chairul    4096 Sep 24 23:53 ./  
drwxr-xr-x 1 chairul chairul    4096 Sep 24 23:51 ../  
-rwxr-xr-x 1 chairul chairul 148551135 Sep 24 23:53 train.article.txt*  
chairul@maze:~$
```

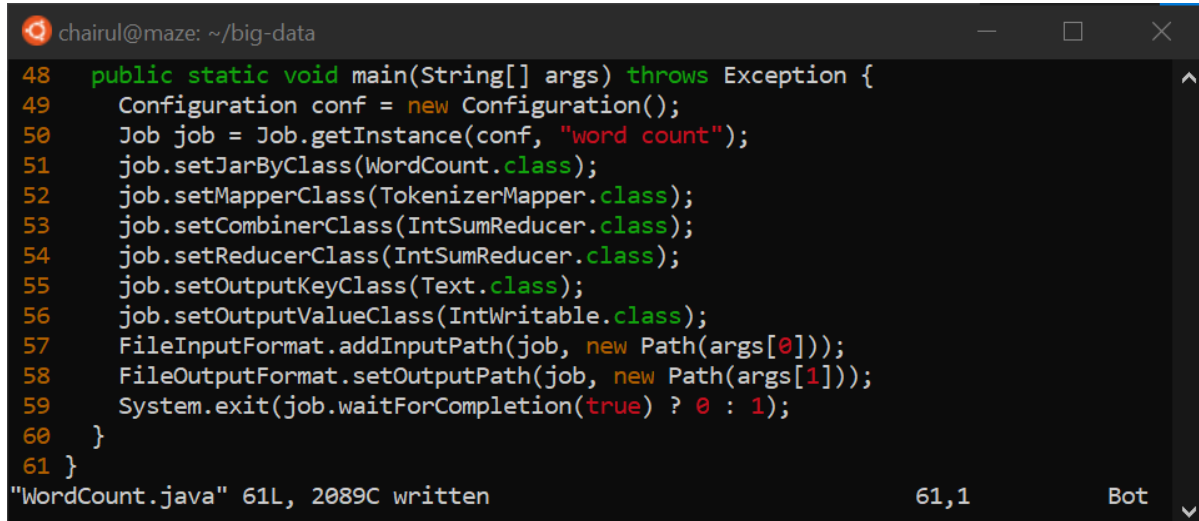
#### Info

command **cp** pada contoh diatas menerima 2 parameter, **SOURCE** dan **DEST**. Sesuaikan dengan lokasi file dan folder pada sistem anda.

## Salin source code

Salin dan simpan *source code* WordCount v1.0 ke file `WordCount.java`. *Source code* dapat diakses pada URL berikut [WordCount v1.0](#).

```
$ vi big-data/WordCount.java
```



```
chairul@maze: ~/big-data
48 public static void main(String[] args) throws Exception {
49     Configuration conf = new Configuration();
50     Job job = Job.getInstance(conf, "word count");
51     job.setJarByClass(WordCount.class);
52     job.setMapperClass(TokenizerMapper.class);
53     job.setCombinerClass(IntSumReducer.class);
54     job.setReducerClass(IntSumReducer.class);
55     job.setOutputKeyClass(Text.class);
56     job.setOutputValueClass(IntWritable.class);
57     FileInputFormat.addInputPath(job, new Path(args[0]));
58     FileOutputFormat.setOutputPath(job, new Path(args[1]));
59     System.exit(job.waitForCompletion(true) ? 0 : 1);
60 }
61 }
"WordCount.java" 61L, 2089C written 61,1 Bot
```

## Tambah environment variable HADOOP\_CLASSPATH

Pastikan bahwa variabel lain seperti **JAVA\_HOME** telah ditambahkan pada file `.bashrc`. Jalankan perintah berikut untuk menambahkan variabel **\$HADOOP\_CLASSPATH**

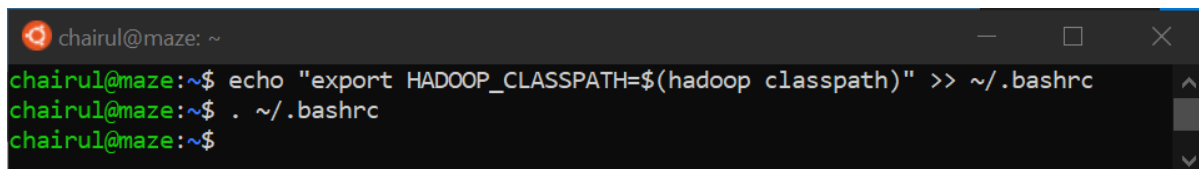
```
$ echo "export HADOOP_CLASSPATH=$(hadoop classpath)" >> ~/.bashrc
```

Kemudian *reload environment* dengan perintah

```
$ source ~/.bashrc
```

atau

```
$ . ~/.bashrc
```



```
chairul@maze: ~
chairul@maze:~$ echo "export HADOOP_CLASSPATH=$(hadoop classpath)" >> ~/.bashrc
chairul@maze:~$ . ~/.bashrc
chairul@maze:~$
```

## Compile WordCount.Java dan buat Jar

File **WordCount.java** dapat dikompilasi dengan perintah

```
$ javac -cp ${HADOOP_CLASSPATH} WordCount.java -d WordCount/
```

```
chairul@LAPTOP-JMUC1KCO: ~/big-data
chairul@LAPTOP-JMUC1KCO:~/big-data$ javac -cp ${HADOOP_CLASSPATH} WordCount.java
chairul@LAPTOP-JMUC1KCO:~/big-data$ ll
total 145092
drwxr-xr-x 1 chairul chairul 4096 Sep 26 15:22 ./
drwxr-xr-x 1 chairul chairul 4096 Sep 26 14:55 ../
-rw-r--r-- 1 chairul chairul 1739 Sep 26 15:22 'WordCount$IntSumReducer.class'
-rw-r--r-- 1 chairul chairul 1736 Sep 26 15:22 'WordCount$TokenizerMapper.class'
-rw-r--r-- 1 chairul chairul 1491 Sep 26 15:22 WordCount.class
-rw-r--r-- 1 chairul chairul 2089 Sep 25 02:15 WordCount.java
-rwxr--r-- 1 chairul chairul 48 Sep 25 13:12 del_output.sh*
-rw-r--r-- 1 chairul chairul 3667 Sep 25 13:16 error_MRAppMaster.txt
-rwxr-xr-x 1 chairul chairul 148551135 Sep 24 23:53 train.article.txt*
chairul@LAPTOP-JMUC1KCO:~/big-data$
```

Kemudian dari file-file **.class** tersebut akan dibuat sebuah file **Jar**. Lakukan dengan perintah

```
$ jar -cf wc.jar WordCount*.class
```

```
chairul@LAPTOP-JMUC1KCO: ~/big-data
chairul@LAPTOP-JMUC1KCO:~/big-data$ jar -cf wc.jar WordCount*.class
chairul@LAPTOP-JMUC1KCO:~/big-data$ ll
total 145092
drwxr-xr-x 1 chairul chairul 4096 Sep 26 15:24 ./
drwxr-xr-x 1 chairul chairul 4096 Sep 26 14:55 ../
-rw-r--r-- 1 chairul chairul 1739 Sep 26 15:22 'WordCount$IntSumReducer.class'
-rw-r--r-- 1 chairul chairul 1736 Sep 26 15:22 'WordCount$TokenizerMapper.class'
-rw-r--r-- 1 chairul chairul 1491 Sep 26 15:22 WordCount.class
-rw-r--r-- 1 chairul chairul 2089 Sep 25 02:15 WordCount.java
-rwxr--r-- 1 chairul chairul 48 Sep 25 13:12 del_output.sh*
-rwxr-xr-x 1 chairul chairul 148551135 Sep 24 23:53 train.article.txt*
-rw-r--r-- 1 chairul chairul 3069 Sep 26 15:24 wc.jar
chairul@LAPTOP-JMUC1KCO:~/big-data$
```

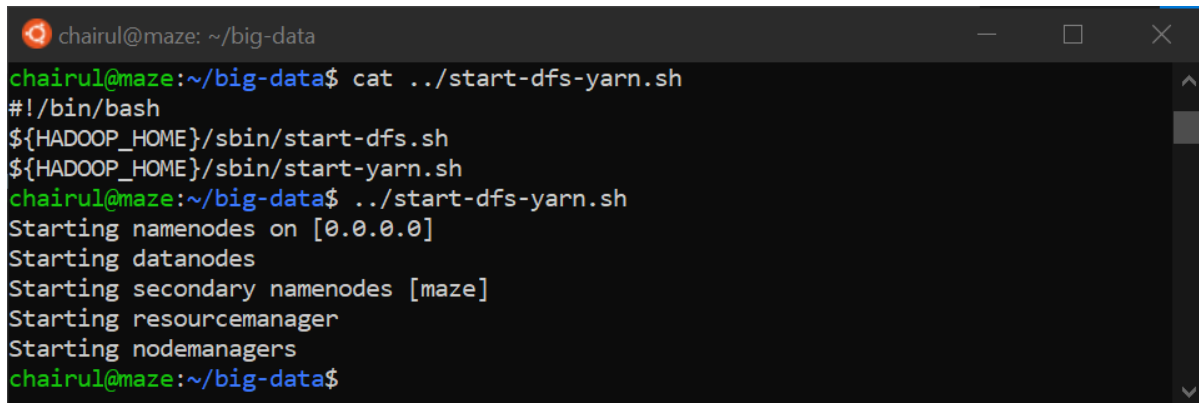
### Info

Opsi **-cf** pada perintah **jar** adalah opsi untuk membuat file jar dengan nama yang dilewatkan pada parameter/argumen.

## Jalankan DFS dan YARN

Setelah selesai membuat file jar, dilanjutkan dengan menjalankan DFS dan YARN dengan perintah

```
$HADOOP_HOME/sbin/start-dfs.sh  
$HADOOP_HOME/sbin/start-yarn.sh
```



```
chairul@maze: ~/big-data  
chairul@maze:~/big-data$ cat ../start-dfs-yarn.sh  
#!/bin/bash  
${HADOOP_HOME}/sbin/start-dfs.sh  
${HADOOP_HOME}/sbin/start-yarn.sh  
chairul@maze:~/big-data$ ../start-dfs-yarn.sh  
Starting namenodes on [0.0.0.0]  
Starting datanodes  
Starting secondary namenodes [maze]  
Starting resourcemanager  
Starting nodemanagers  
chairul@maze:~/big-data$
```

### Tip

Lelah mengetik perintah yang sama berulang kali ?

Coba pakai **Bash Script**.

Pada contoh diatas saya menggabungkan dua perintah kedalam 1 file **.sh**. Untuk dapat mengeksekusi file tersebut lakukan

```
$ chmod 744 start-dfs-yarn.sh
```

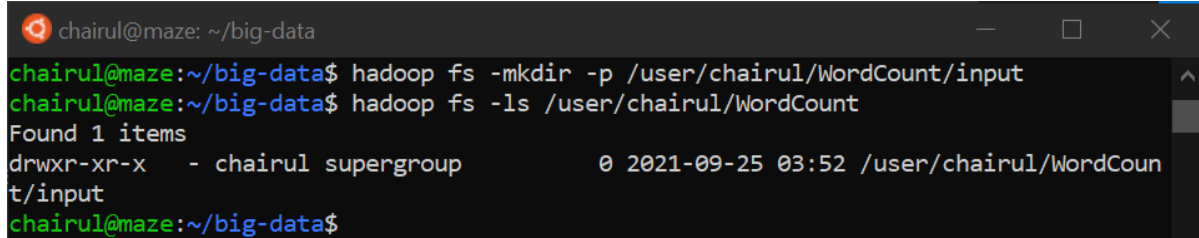
Kemudian file dapat dieksekusi dengan

```
$ ./start-dfs-yarn.sh
```

## Membuat folder input dan memindahkan file teks

Folder input ini digunakan sebagai folder penyimpanan file teks **train.article.txt**. Lakukan perintah berikut untuk membuat folder input

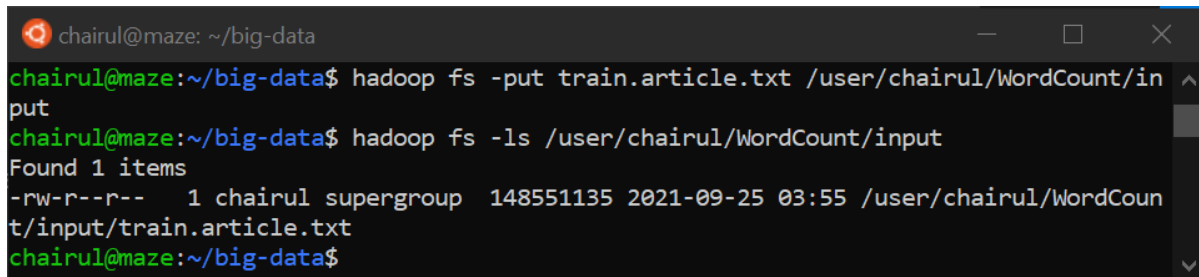
```
$ hadoop fs -mkdir -p /user/chairul/WordCount/input
```



```
chairul@maze: ~/big-data
chairul@maze:~/big-data$ hadoop fs -mkdir -p /user/chairul/WordCount/input
chairul@maze:~/big-data$ hadoop fs -ls /user/chairul/WordCount
Found 1 items
drwxr-xr-x  - chairul supergroup          0 2021-09-25 03:52 /user/chairul/WordCount/input
chairul@maze:~/big-data$
```

Kemudian lanjutkan dengan memindahkan file **train.article.txt** ke folder **/input**

```
$ hadoop fs -put train.article.txt /user/chairul/WordCount/input
```



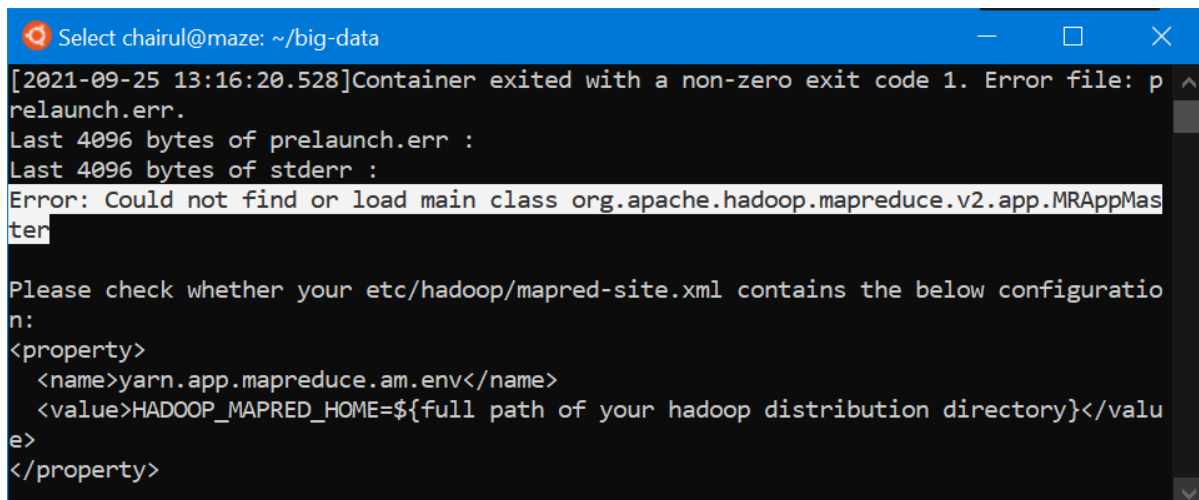
```
chairul@maze: ~/big-data
chairul@maze:~/big-data$ hadoop fs -put train.article.txt /user/chairul/WordCount/input
chairul@maze:~/big-data$ hadoop fs -ls /user/chairul/WordCount/input
Found 1 items
-rw-r--r--  1 chairul supergroup 148551135 2021-09-25 03:55 /user/chairul/WordCount/input/train.article.txt
chairul@maze:~/big-data$
```

## Jalankan Jar

Kemudian eksekusi perintah berikut untuk menjalankan program MapReduce yang menghitung frekuensi kata,

```
$ hadoop jar wc.jar WordCount /user/chairul/WordCount/input
/user/chairul/WordCount/output
```

Sesaat setelah dijalankan, terminal menampilkan **Error: Could not find or load main class org.apache.hadoop.mapreduce.v2.app.MRAppMaster**.



```
Select chairul@maze: ~/big-data
[2021-09-25 13:16:20.528]Container exited with a non-zero exit code 1. Error file: prelaunch.err.
Last 4096 bytes of prelaunch.err :
Last 4096 bytes of stderr :
Error: Could not find or load main class org.apache.hadoop.mapreduce.v2.app.MRAppMaster

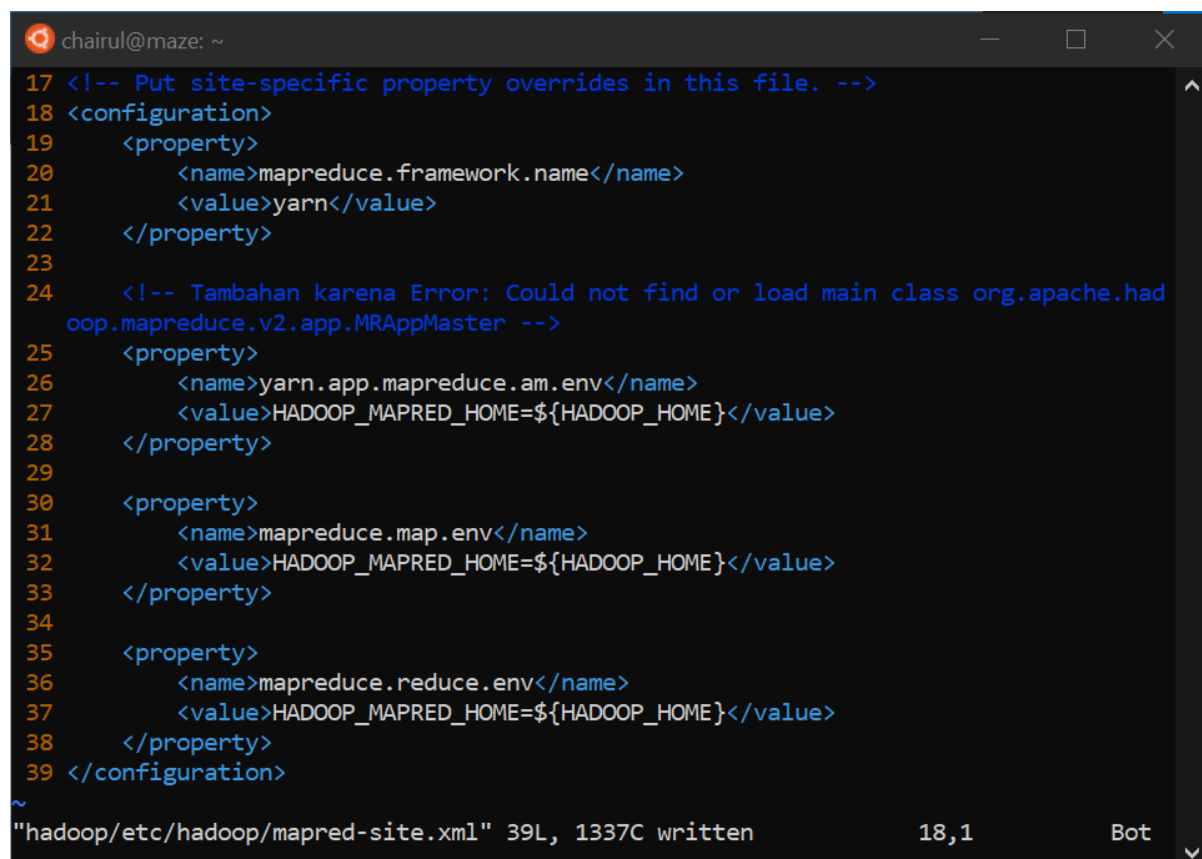
Please check whether your etc/hadoop/mapred-site.xml contains the below configuration:
<property>
  <name>yarn.app.mapreduce.am.env</name>
  <value>HADOOP_MAPRED_HOME=${full path of your hadoop distribution directory}</value>
</property>
```

# Troubleshoot Error MRAppMaster

Error ini terjadi karena *main class* **org.apache.hadoop.mapreduce.v2.app.MRAppMaster** tidak ditemukan atau dimuat. Detail error dapat diakses pada [Error MRAppMaster](#). Untuk memperbaiki error ini, tambahkan beberapa konfigurasi berikut pada file **mapred-site.xml**.

```
<!-- Tambahan karena Error: Could not find or load main class org.apache.had
<property>
  <name>yarn.app.mapreduce.am.env</name>
  <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>
</property>
<property>
  <name>mapreduce.map.env</name>
  <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>
</property>
<property>
  <name>mapreduce.reduce.env</name>
  <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>
</property>
```

Sehingga file **mapred-site.xml** terlihat seperti berikut



```
chairul@maze: ~
17 <!-- Put site-specific property overrides in this file. -->
18 <configuration>
19   <property>
20     <name>mapreduce.framework.name</name>
21     <value>yarn</value>
22   </property>
23
24   <!-- Tambahan karena Error: Could not find or load main class org.apache.had
oop.mapreduce.v2.app.MRAppMaster -->
25   <property>
26     <name>yarn.app.mapreduce.am.env</name>
27     <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>
28   </property>
29
30   <property>
31     <name>mapreduce.map.env</name>
32     <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>
33   </property>
34
35   <property>
36     <name>mapreduce.reduce.env</name>
37     <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>
38   </property>
39 </configuration>
~
"hadoop/etc/hadoop/mapred-site.xml" 39L, 1337C written      18,1      Bot
```

Kemudian coba jalankan kembali program MapReduce tersebut. Program berhasil berjalan dan hasil frekuensi kata dapat diakses pada folder **/output**.

```
$ hadoop fs -ls /user/chairul/WordCount/output
```

```
chairul@maze: ~  
chairul@maze:~$ hadoop fs -ls /user/chairul/WordCount/output  
Found 2 items  
-rw-r--r--  1 chairul supergroup          0 2021-09-25 22:36 /user/chairul/WordCount/output/_SUCCESS  
-rw-r--r--  1 chairul supergroup 2570083 2021-09-25 22:36 /user/chairul/WordCount/output/part-r-00000  
chairul@maze:~$
```

Sedangkan untuk menampilkan beberapa frekuensi kata terakhir dapat dilakukan dengan

```
$ hadoop fs -tail /user/chairul/WordCount/output/part-r-00000
```

```
chairul@maze: ~  
chairul@maze:~$ hadoop fs -tail /user/chairul/WordCount/output/part-r-00000  
) 4  
' 42  
' 95  
'10 4  
'11 4  
'15 4  
'19' 4  
'20 4  
'60 12  
'66 4  
'70 3  
'70an. 4  
'80 8  
'90 26  
'98 4  
'Beto' 4  
'Disneyland 4  
'Jalur 4  
'Kami 4
```

# Tampilan dashboard Hadoop

Folder /input dapat diakses dengan menginput **/user/chairul/Wordcount/input** pada form kemudian klik **Go** atau **Enter**.

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

## Browse Directory

Go!

Show 25 entries

Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	<a href="#">chairul</a>	<a href="#">supergroup</a>	141.67 MB	Sep 25 03:55	<a href="#">1</a>	128 MB	<a href="#">train.article.txt</a>	<div></div>

Showing 1 to 1 of 1 entries

Previous1Next

Hadoop, 2021.

Sedangkan untuk mengakses folder /output, ganti nilai form menjadi **/user/chairul/Wordcount/output**.

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

## Browse Directory

Go!

Show 25 entries

Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	<a href="#">chairul</a>	<a href="#">supergroup</a>	0 B	Sep 25 22:36	<a href="#">1</a>	128 MB	<a href="#">_SUCCESS</a>	<div></div>
<input type="checkbox"/>	-rw-r--r--	<a href="#">chairul</a>	<a href="#">supergroup</a>	2.45 MB	Sep 25 22:36	<a href="#">1</a>	128 MB	<a href="#">part-r-00000</a>	<div></div>

Showing 1 to 2 of 2 entries

Previous1Next

Hadoop, 2021.