

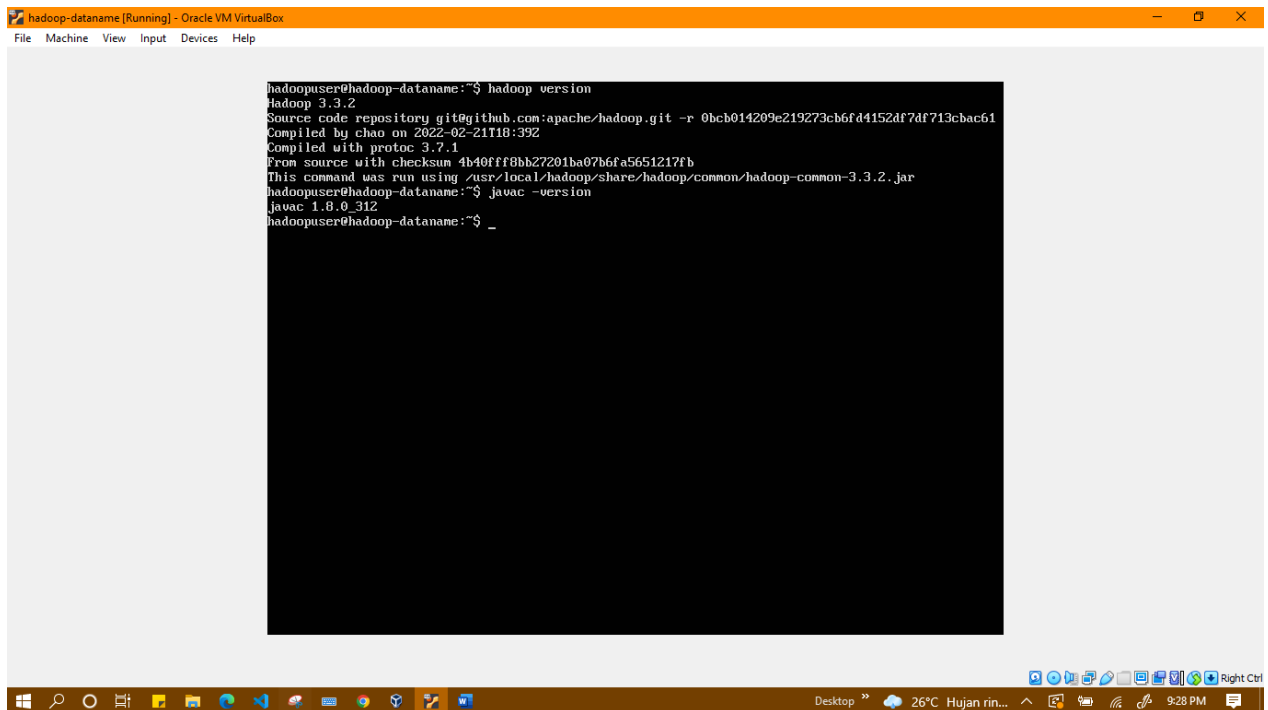
Big Data
“MapReduce II”



Oleh :
Rajendra Rakha Arya
1941720080
TI 3H

JURUSAN TEKNOLOGI INFORMASI
PROGRAM STUDI TEKNIK INFORMATIKA
POLITEKNIK NEGERI MALANG
2022

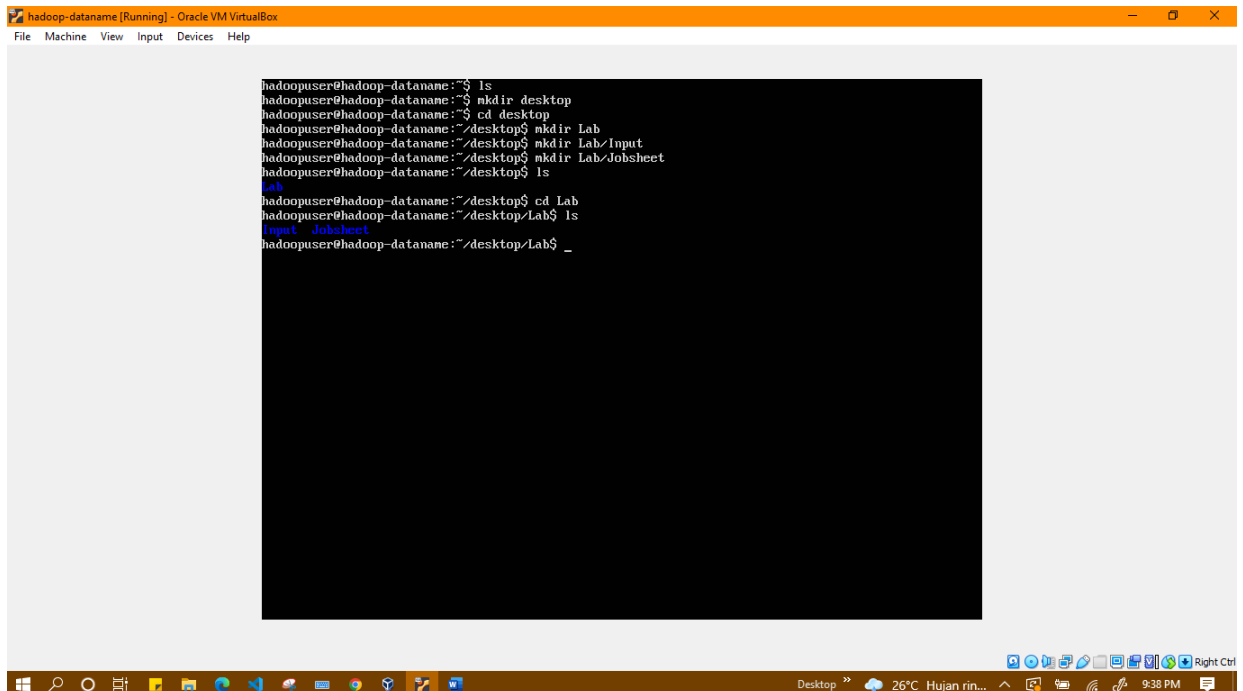
1. Pastikan Hadoop dan Java diinstal dengan benar



The screenshot shows a terminal window titled "hadoop-dataname [Running] - Oracle VM VirtualBox". The terminal output is as follows:

```
hadoopuser@hadoop-dataname:~$ hadoop version
Hadoop 3.3.2
Source code repository git@github.com:apache/hadoop.git -r 0bcb014209e219273cb6fd4152df7df713cbac61
Compiled by chao on 2022-02-21T18:39Z
Compiled with protoc 3.7.1
From source with checksum 4b40fff8bb27201ba07b6fa5651217fb
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-3.3.2.jar
hadoopuser@hadoop-dataname:~$ javac -version
javac 1.8.0_312
hadoopuser@hadoop-dataname:~$ _
```

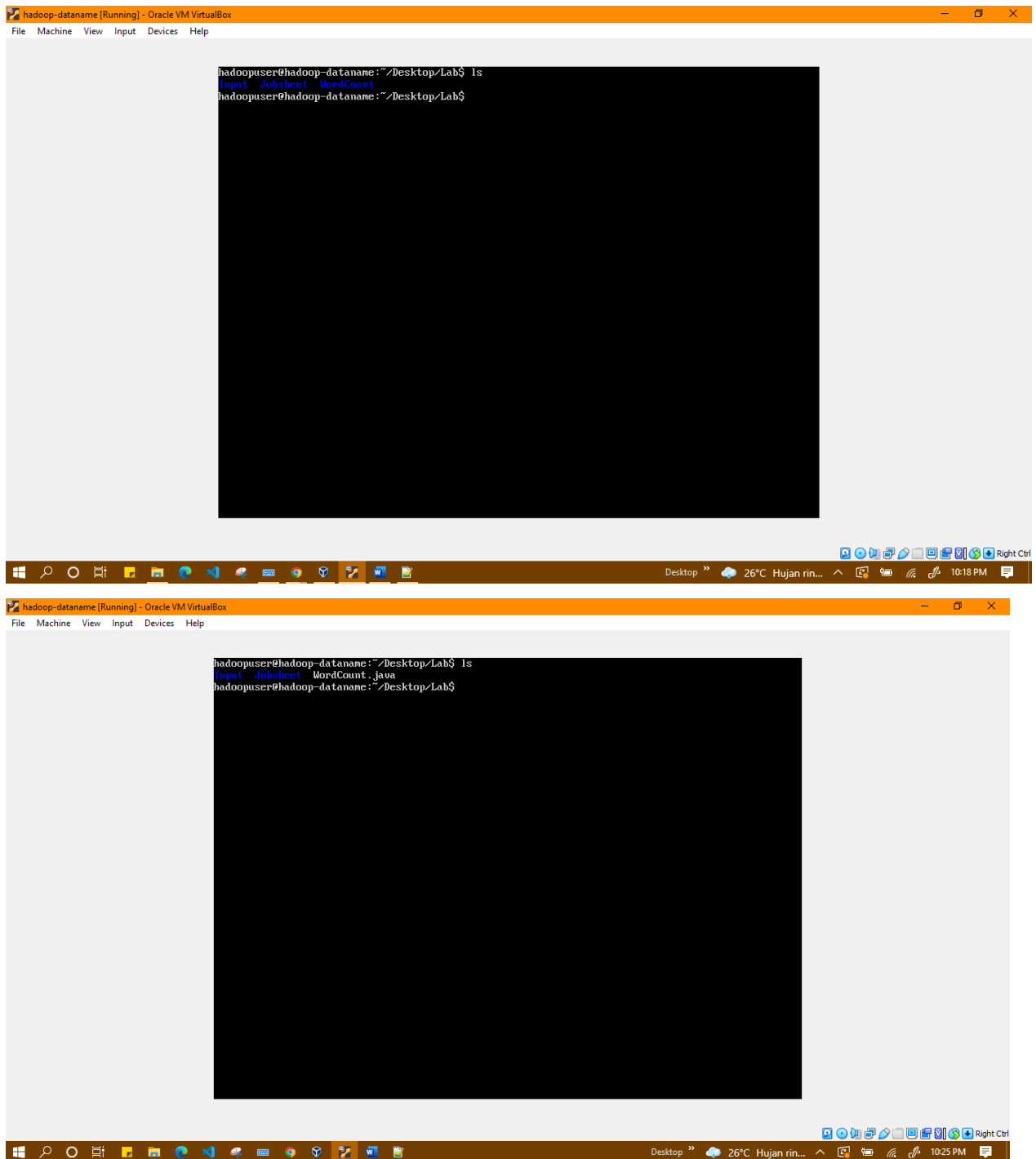
2. Buat direktori di Desktop bernama Lab dan di dalamnya buat dua folder; satu disebut "Input" dan yang lainnya disebut "Jobsheet".



The screenshot shows a terminal window titled "hadoop-dataname [Running] - Oracle VM VirtualBox". The terminal output is as follows:

```
hadoopuser@hadoop-dataname:~$ ls
hadoopuser@hadoop-dataname:~$ mkdir desktop
hadoopuser@hadoop-dataname:~$ cd desktop
hadoopuser@hadoop-dataname:~/desktop$ mkdir Lab
hadoopuser@hadoop-dataname:~/desktop$ mkdir Lab/Input
hadoopuser@hadoop-dataname:~/desktop$ mkdir Lab/Jobsheet
hadoopuser@hadoop-dataname:~/desktop$ ls
Lab
hadoopuser@hadoop-dataname:~/desktop$ cd Lab
hadoopuser@hadoop-dataname:~/desktop/Lab$ ls
Input  Jobsheet
hadoopuser@hadoop-dataname:~/desktop/Lab$ _
```

3. Buat dan Tambahkan file .java "WordCount.java" di lab direktori



hadoop-dataname [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

```
GNU nano 2.9.3 WordCount.java

import java.io.IOException;
import java.util.regex.Pattern;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;

import org.apache.log4j.Logger;

public class WordCount extends Configured implements Tool {

    private static final Logger LOG = Logger.getLogger(WordCount.class);

    public static void main(String[] args) throws Exception {
        int res = ToolRunner.run(new WordCount(), args);
        System.exit(res);
    }

    public int run(String[] args) throws Exception {
        Job job = Job.getInstance(getConf(), "wordcount");
        job.setJarByClass(this.getClass());
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        job.setMapperClass(Map.class);

        Get Help Write Out Where Is Cut Text Justify Cur Pos T-U Undo
        Exit Read File Replace Uncut Text To Spell Go To Line T-R Redo

        Read 70 lines
    }
}
```

Desktop 26°C Hujan rin... 10:28 PM

hadoop-dataname [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

```
GNU nano 2.9.3 WordCount.java

public class WordCount extends Configured implements Tool {

    private static final Logger LOG = Logger.getLogger(WordCount.class);

    public static void main(String[] args) throws Exception {
        int res = ToolRunner.run(new WordCount(), args);
        System.exit(res);
    }

    public int run(String[] args) throws Exception {
        Job job = Job.getInstance(getConf(), "wordcount");
        job.setJarByClass(this.getClass());
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        job.setMapperClass(Map.class);
        job.setReducerClass(Reduce.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        return job.waitForCompletion(true) ? 0 : 1;
    }

    public static class Map extends Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();
        private long numRecords = 0;
        private static final Pattern WORD_BOUNDARY = Pattern.compile("\\s*\\b\\s*");

        public void map(LongWritable offset, Text lineText, Context context)
            throws IOException, InterruptedException {
            String line = lineText.toString();
            Text currentWord = new Text();
        }
    }
}
```

Desktop 26°C Hujan rin... 10:28 PM

hadoop-dataname [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

hadoop-dataname [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

```
GNU nano 2.9.3 WordCount.java

job.setReducerClass(Reducer.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
return job.waitForCompletion(true) ? 0 : 1;
}

public static class Map extends Mapper<LongWritable, Text, Text, IntWritable> {
    private final static IntWritable one = new IntWritable(1);
    private Text word = new Text();
    private long numRecords = 0;
    private static final Pattern WORD_BOUNDARY = Pattern.compile("\\s*\\b\\s*");

    public void map(LongWritable offset, Text lineText, Context context)
        throws IOException, InterruptedException {
        String line = lineText.toString();
        Text currentWord = new Text();
        for (String word : WORD_BOUNDARY.split(line)) {
            if (word.isEmpty()) {
                continue;
            }
            currentWord = new Text(word);
            context.write(currentWord, one);
        }
    }
}

public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritable> {
    @Override
    public void reduce(Text word, Iterable<IntWritable> counts, Context context)
        throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable count : counts) {

```

Get Help Write Out Where Is Cut Text Justify Cur Pos Undo
Exit Read File Replace Uncut Text To Spell Go To Line Redo

Desktop 26°C Hujan rin... 10:29 PM

hadoop-dataname [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

```
GNU nano 2.9.3 WordCount.java

        for (String word : WORD_BOUNDARY.split(line)) {
            if (word.isEmpty()) {
                continue;
            }
            currentWord = new Text(word);
            context.write(currentWord, one);
        }
    }
}

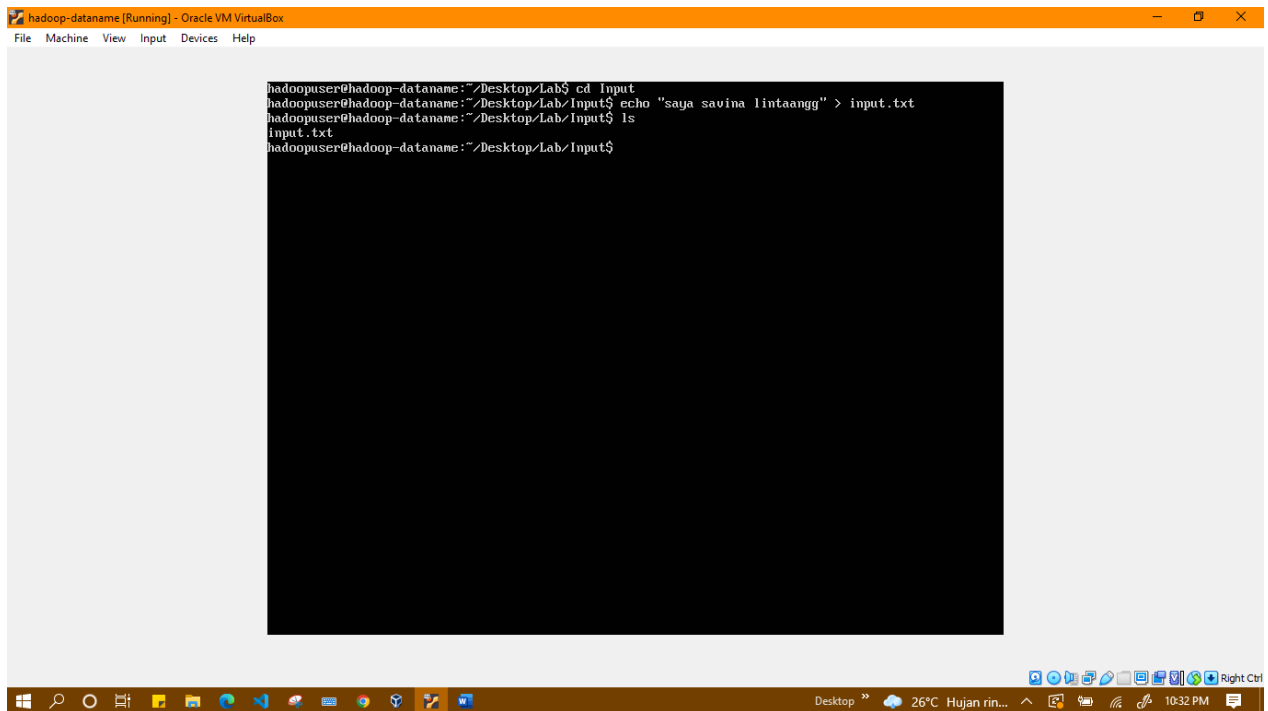
public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritable> {
    @Override
    public void reduce(Text word, Iterable<IntWritable> counts, Context context)
        throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable count : counts) {
            sum += count.get();
        }
        context.write(word, new IntWritable(sum));
    }
}

```

Get Help Write Out Where Is Cut Text Justify Cur Pos Undo
Exit Read File Replace Uncut Text To Spell Go To Line Redo

Desktop 26°C Hujan rin... 10:29 PM

4. Buat & Tambahkan file .txt dengan penamaan "input.txt" di direktori Lab/Input.

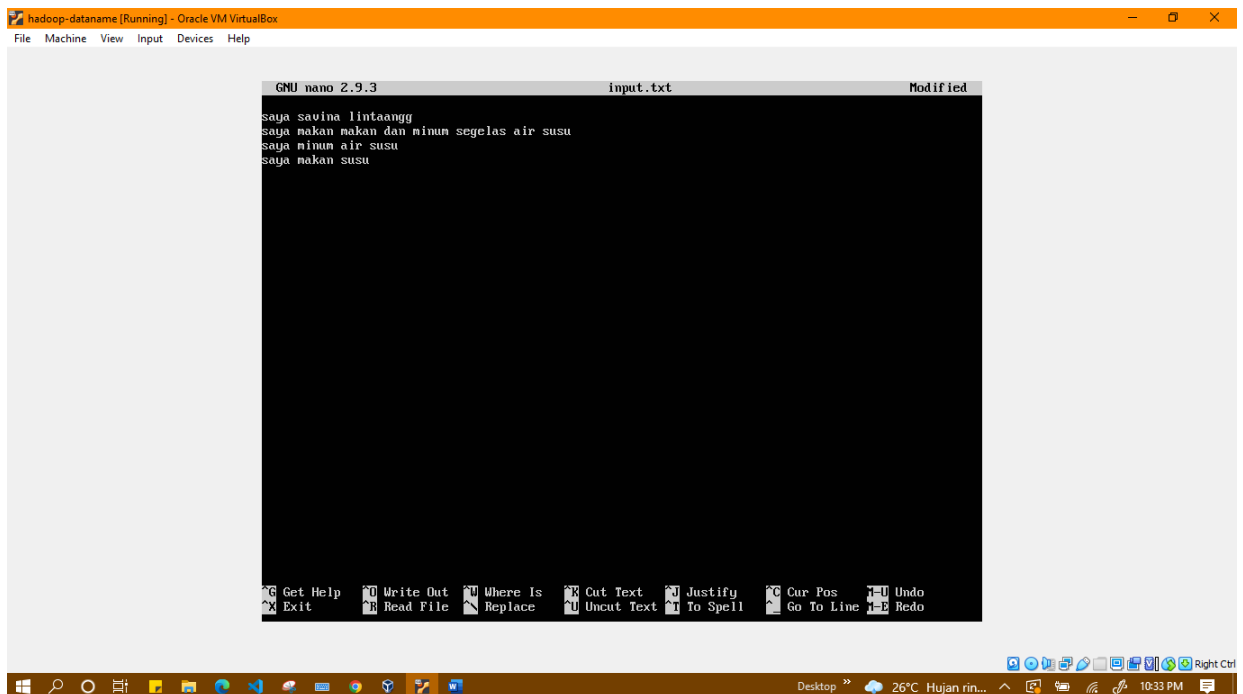


The screenshot shows a terminal window titled "hadoop-dataname [Running] - Oracle VM VirtualBox". The terminal output is as follows:

```
hadoopuser@hadoop-dataname:~/Desktop/Lab$ cd Input
hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$ echo "saya savina lintaangg" > input.txt
hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$ ls
input.txt
hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$
```

The terminal window is part of a desktop environment with a taskbar at the bottom showing various application icons and system status information like "Desktop", "26°C", and "Hujan rin...".

Tambahkan kata-kata lain pada file input



The screenshot shows the same terminal window, but now the nano text editor is open, editing the file "input.txt". The editor's title bar says "GNU nano 2.9.3 input.txt Modified". The content of the file is:

```
saya savina lintaangg
saya nakan makan dan minum segelas air susu
saya minum air susu
saya nakan susu
```

The nano editor's status bar at the bottom shows various commands like "Get Help", "Exit", "Write Out", "Read File", "Where Is", "Replace", "Cut Text", "Uncut Text", "Justify", "To Spell", "Cur Pos", "Go To Line", "Undo", and "Redo". The desktop environment and taskbar are the same as in the previous screenshot.

5. Ketik command berikut untuk mengeksport classpath hadoop ke bash.

Pastikan telah berhasil diekspor.

```
hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$ export HADOOP_CLASSPATH=$(hadoop classpath)
hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$ echo $HADOOP_CLASSPATH

hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$ echo $HADOOP_CLASSPATH
/usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/hdfs:/usr/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/mapreduce/*:/usr/local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/yarn/*
```

6. Selanjutnya adalah membuat direktori baru /WordCount/Input di HDFS dan memindahkan input.txt ke HDFS dengan mengetikkan perintah berikut.

```
hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$ hadoop fs -mkdir /WordCount
hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$ hadoop fs -mkdir /WordCount/Input
hadoopuser@hadoop-dataname:~/Desktop/Lab/Input$

hadoopuser@hadoop-dataname:~/Desktop$ hadoop fs -put Lab/Input?input.txt /WordCount/Input
hadoopuser@hadoop-dataname:~/Desktop$ hadoop fs -ls /WordCount/Input
Found 1 items
drwxr-xr-x - hadoopuser supergroup          0 2022-04-20 00:29 /WordCount/Input/Input
hadoopuser@hadoop-dataname:~/Desktop$ _
```

7. Periksa dan lihat bahwa direktori dan file telah ditempatkan di File System yang tepat.

```
hadoopuser@hadoop-dataname:~/Desktop$ hadoop fs -ls /
Found 1 items
drwxr-xr-x - hadoopuser supergroup          0 2022-04-19 22:52 /WordCount
hadoopuser@hadoop-dataname:~/Desktop$ hadoop fs -ls /WordCount/Input
Found 1 items
drwxr-xr-x - hadoopuser supergroup          0 2022-04-20 00:29 /WordCount/Input/Input
hadoopuser@hadoop-dataname:~/Desktop$ _
```

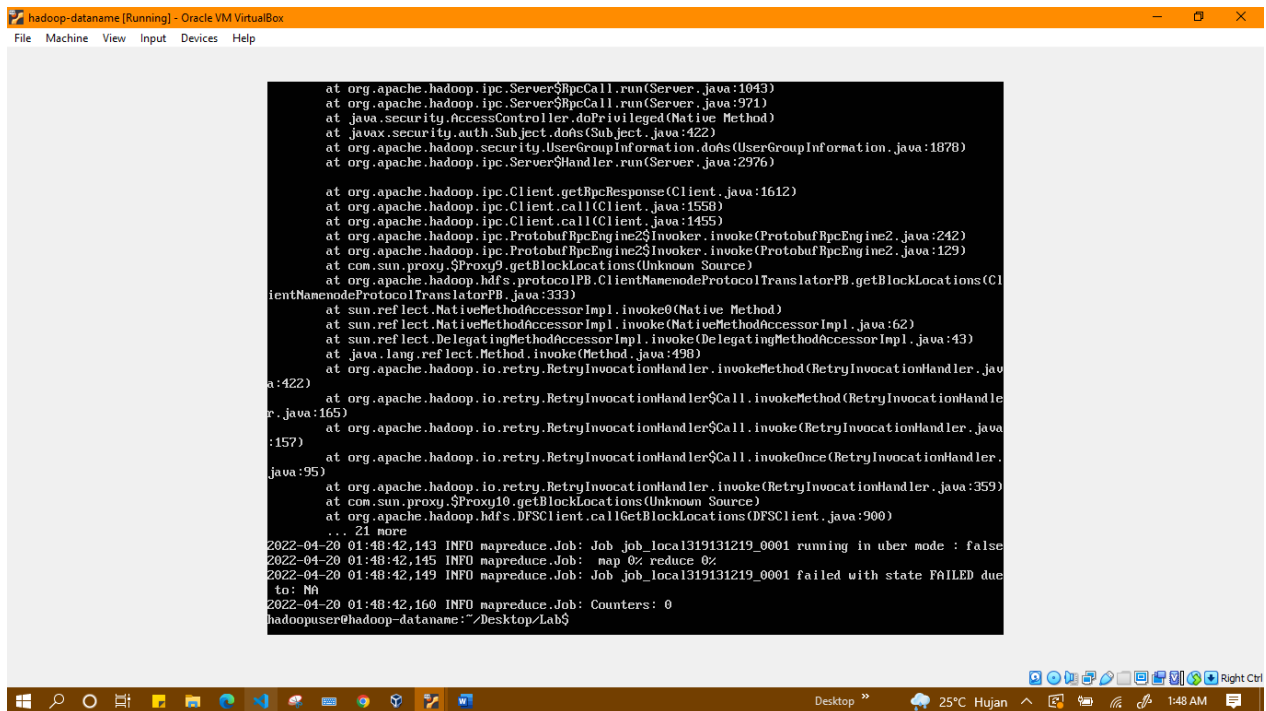
8. Kemudian, kembali ke mesin lokal di mana kita akan mengkompilasi file WordCount.java. Dengan asumsi saat ini berada di direktori Desktop.

Masukkan file output dalam satu file jar

```
hadoopuser@hadoop-dataname:~/Desktop/Lab$ javac -classpath $(hadoop classpath) -d Jobsheet WordCount.java
hadoopuser@hadoop-dataname:~/Desktop/Lab$ jar -cvf WordCount.jar -C Jobsheet/ .
added manifest
adding: WordCount$Reduce.class(in = 1627) (out= 686)(deflated 57%)
adding: WordCount.class(in = 1946) (out= 981)(deflated 49%)
adding: WordCount$Map.class(in = 2189) (out= 981)(deflated 55%)
hadoopuser@hadoop-dataname:~/Desktop/Lab$ _

hadoopuser@hadoop-dataname:~/Desktop/Lab$ jar -cvf WordCount.jar -C Jobsheet/ .
added manifest
hadoopuser@hadoop-dataname:~/Desktop/Lab$ ls
Input  Jobsheet  WordCount.jar  WordCount.java
```

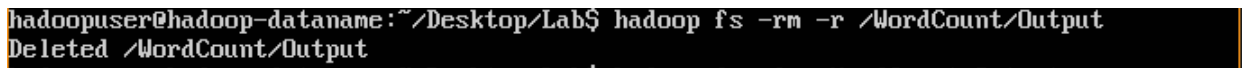
9. Sekarang, jalankan file jar di Hadoop



```
at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:1043)
at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:971)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)
at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2976)

at org.apache.hadoop.ipc.Client.getRpcResponse(Client.java:1612)
at org.apache.hadoop.ipc.Client.call(Client.java:1558)
at org.apache.hadoop.ipc.Client.call(Client.java:1455)
at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:242)
at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:129)
at com.sun.proxy.$Proxy9.getBlockLocations(Unknown Source)
at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.getBlockLocations(ClientNamenodeProtocolTranslatorPB.java:333)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:422)
at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeMethod(RetryInvocationHandler.java:165)
at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invoke(RetryInvocationHandler.java:157)
at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeOnce(RetryInvocationHandler.java:95)
at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:359)
at com.sun.proxy.$Proxy10.getBlockLocations(Unknown Source)
at org.apache.hadoop.hdfs.DFSClient.callGetBlockLocations(DFSClient.java:900)
... 21 more
2022-04-20 01:48:42,143 INFO mapreduce.Job: Job_job_local1319131219_0001 running in uber mode : false
2022-04-20 01:48:42,145 INFO mapreduce.Job: map 0% reduce 0%
2022-04-20 01:48:42,149 INFO mapreduce.Job: Job_job_local1319131219_0001 failed with state FAILED due to: NA
2022-04-20 01:48:42,160 INFO mapreduce.Job: Counters: 0
hadoopuser@hadoop-dataname:~/Desktop/Lab$
```

10. Jika Anda ingin menjalankan sampel lagi, Anda harus terlebih dahulu menghapus direktori output. Gunakan perintah berikut.



```
hadoopuser@hadoop-dataname:~/Desktop/Lab$ hadoop fs -rm -r /WordCount/Output
Deleted /WordCount/Output
```