

Lab – 4

Aim: Hello World of MapReduce Framework - Java (WordCount)

Rahil Acharya

CE004

20CEUOD004

```
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ script Lab4BDA.txt
started, output log file is 'Lab4BDA.txt'.
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-dfs.sh
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
g namenodes on [localhost]
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
g datanodes
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
g secondary namenodes [localhost]
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
-28 08:50:05,567 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where appl
```

```
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-yarn.sh
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
g resourcemanager
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
g nodemanagers
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-historymanager.sh
istorymanager.sh: command not found
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-historymanager.sh
istorymanager.sh: command not found
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-hcbase.sh
case.sh: command not found
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-hbase.sh
base.sh: command not found
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-yarn.sh
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
g resourcemanager
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
emanager is running as process 7353. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
g nodemanagers
: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
st: nodemanager is running as process 7693. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-historymanager.sh
```

```
SUBCOMMAND may print help when invoked w/o parameters or with -h.
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ echo $HADOOP_HOME
/opt/hadoop
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ whoami
hadoop
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ stop-historyserver.sh
stop-historyserver.sh: command not found
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ pwd
/home/hadoop/Desktop
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ ls
input Lab4BDA.txt
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ jps
6736 DataNode
6960 SecondaryNameNode
8755 Jps
6520 NameNode
7353 ResourceManager
7693 NodeManager
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-dfs.sh
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting namenodes on [localhost]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
localhost: namenode is running as process 6520. Stop it first and ensure /tmp/hadoop-hadoop-namenode.pid file is empty before retry.
Starting datanodes
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
localhost: datanode is running as process 6736. Stop it first and ensure /tmp/hadoop-hadoop-datanode.pid file is empty before retry.
Starting secondary namenodes [localhost]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
localhost: secondarynamenode is running as process 6960. Stop it first and ensure /tmp/hadoop-hadoop-secondarynamenode.pid file is empty before retry.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 08:56:34,906 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ jps
6736 DataNode
6960 SecondaryNameNode
```

```

applicable
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ jps
6736 DataNode
6960 SecondaryNameNode
6520 NameNode
7353 ResourceManager
9274 Jps
7693 NodeManager
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ start-historyserver.sh
start-historyserver.sh: command not found
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ $HADOOP_HOME/sbin/mr-jobhistory-daemon.sh start historyserver
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Use of this script to start the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon start" instead.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ jps
6736 DataNode
6960 SecondaryNameNode
9799 JobHistoryServer
6520 NameNode
7353 ResourceManager
7693 NodeManager
9901 Jps
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -rm -r /projects/wordcount/output
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:02:14,841 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted /projects/wordcount/output
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -rm -r /projects/wordcount/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:02:39,592 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted /projects/wordcount/input
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -ls /projects/wordcount/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:03:00,481 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

```

```

hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -rm -r /projects/wordcount/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:02:39,592 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted /projects/wordcount/input
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -ls /projects/wordcount/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:03:00,481 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ls: '/projects/wordcount/input': No such file or directory
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -mkdir -p /projects/wordcount/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:03:13,552 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -ls /projects/wordcount/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:05:38,052 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -put input/*.txt /projects/wordcount/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:07:08,003 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -ls /projects/wordcount/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:08:47,811 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--  3 hadoop supergroup      20 2022-07-28 09:07 /projects/wordcount/input/lab4.txt
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ ls
input  Lab4BDA.txt  WordCount.java
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ cat import
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.io.Text;

```



```

bash: syntax error near unexpected token `}'
bash: syntax error near unexpected token `}'
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ cat WordCount.java
import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount
{
    public static class TokenizerMapper extends Mapper < LongWritable,
        Text, Text, IntWritable >
    {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();
        public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException
        {
            StringTokenizer tokenizer = new StringTokenizer(value.toString());
            while (tokenizer.hasMoreTokens()) {
                word.set(tokenizer.nextToken());
                context.write(word, one);
            }
        }
    }
    public static class IntSumReducer extends Reducer < Text, IntWritable,
        Text, IntWritable >
    {
        private IntWritable result = new IntWritable();
        public void reduce(Text key, Iterable < IntWritable > values, Context context) throws

```

```

}
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ javac -cp 'hadoop classpath' WordCount.java
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ jar cvf wc.jar *.class
added manifest
adding: WordCount$IntSumReducer.class(in = 1755) (out= 750)(deflated 57%)
adding: WordCount$TokenizerMapper.class(in = 1862) (out= 780)(deflated 58%)
adding: WordCount.class(in = 1511) (out= 824)(deflated 45%)
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hadoop jar wc.jar WordCount /projects/wordcount/input /projects/wordcount/output
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:15:01,367 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-07-28 09:15:01,640 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at localhost/127.0.0.1:8032
2022-07-28 09:15:01,780 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2022-07-28 09:15:01,790 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1658978455578_0001
2022-07-28 09:15:01,897 INFO input.FileInputFormat: Total input files to process : 1
2022-07-28 09:15:01,928 INFO mapreduce.JobSubmitter: number of splits:1
2022-07-28 09:15:02,030 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1658978455578_0001
2022-07-28 09:15:02,030 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-07-28 09:15:02,102 INFO conf.Configuration: resource-types.xml not found
2022-07-28 09:15:02,102 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-07-28 09:15:02,210 INFO impl.YarnClientImpl: Submitted application application_1658978455578_0001
2022-07-28 09:15:02,230 INFO mapreduce.Job: The url to track the job: http://celab2-ThinkCentre-neo-50s-Gen-3:8088/proxy/application_1658978455578_0001/
2022-07-28 09:15:02,230 INFO mapreduce.Job: Running job: job_1658978455578_0001
2022-07-28 09:15:06,276 INFO mapreduce.Job: Job job_1658978455578_0001 running in uber mode : false
2022-07-28 09:15:06,277 INFO mapreduce.Job: map 0% reduce 0%
2022-07-28 09:15:09,326 INFO mapreduce.Job: map 100% reduce 0%
2022-07-28 09:15:13,346 INFO mapreduce.Job: map 100% reduce 100%
2022-07-28 09:15:13,358 INFO mapreduce.Job: Job job_1658978455578_0001 completed successfully
2022-07-28 09:15:13,412 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=56
  FILE: Number of bytes written=551581

```

```

Bytes Written=30
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ ls
input          wc.jar         'WordCount$TokenizerMapper.class'  WordCount.java
Lab4BDA.txt    'WordCount$IntSumReducer.class'    WordCount.class
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -get /projects/wordcount/output .
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:15:40,839 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ cat input/*
350
250
150
450
120
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ cat output/*
120 1
150 1
250 1
350 1
450 1
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ $HADOOP_HOME/sbin/mr-jobhistory-daemon.sh stop historyserver
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Use of this script to stop the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon stop" instead.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ end
Command 'end' not found, did you mean:
  command 'ent' from deb ent (1.2debian-3)
  command 'and' from deb and (1.2.2-4.2)
  command 'nd' from deb nd (0.8.2-8build1)
  command 'ed' from deb ed (1.18-1)
  command 'env' from deb coreutils (8.32-4.1ubuntu1)
  command 'bnd' from deb bnd (5.0.1-3)
  command 'snd' from deb snd-gui-jack (22.1-1)
  command 'snd' from deb snd-gui-pulse (22.1-1)
  command 'snd' from deb snd-nox (22.1-1)

```

```

2022-07-28 09:40:19,143 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
get: 'output/_SUCCESS': File exists
get: 'output/part-r-00000': File exists
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ cat input/*
Books which have more than 300 pages should be in the big book category. Books which have less than 300 pages should be in the small book category.
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ cat output/*
120 1
150 1
250 1
350 1
450 1
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ hdfs dfs -get /projects/wordcount/output .
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 09:42:45,570 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ cat input/*
Books which have more than 300 pages should be in the big book category. Books which have less than 300 pages should be in the small book category.
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop$ cat output/*
300 2
Books 2
be 2
big 1
book 2
category. 2
have 2
in 2
less 1
more 1
pages 2
should 2
small 1
than 2
the 2

```

HadoopOverviewDatanodes

Browse Directory

/user/hadoop/output

Show25entries

☐

Permission

Owner

☐

-rw-r--r--

hadoop

☐

-rw-r--r--

hadoop

Showing 1 to 2 of 2 entries

Hadoop, 2022.

File information - part-r-00000

DownloadHead the file (first 32K)Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741851
Block Pool ID: BP-256671721-127.0.1.1-1655192105741
Generation Stamp: 1027
Size: 55
Availability:

- celab2-ThinkCentre-neo-50s-Gen-3

File contents

dfhfd1

dfhgdfhlkj1

djfkghndkl1

hello3

hi3

world3

Block Size

Name

MB_SUCCESS

part-r-00000

Previous1Next

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

Browse Directory

/user/hadoop/input

Show25entries

☐

Permission

Owner

☐

-rw-r--r--

hadoop

Showing 1 to 1 of 1 entries

Hadoop, 2022.

File information - temp_file

DownloadHead the file (first 32K)Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741827
Block Pool ID: BP-256671721-127.0.1.1-1655192105741
Generation Stamp: 1003
Size: 73
Availability:

- celab2-ThinkCentre-neo-50s-Gen-3

File contents

hi hello world

dfhfd hi hello world

dfhgdfhlkj hi hello world djfkghndkl

Block Size

Name

28 MBtemp_file

Previous1Next

Browse Directory

/user/hadoopGo!

Show25entriesSearch:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	2.03 KB	Jul 28 09:03	3	128 MB	WordCount.java	
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	Jul 28 09:19	0	0 B	input	
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	Jul 28 09:35	0	0 B	output	

Showing 1 to 3 of 3 entriesPrevious1Next

Hadoop

Overview

Datanodes

Browse Directory

/user/hadoop/output

Show 25 entries

Permission	Owner
-rw-r--r--	hadoop
-rw-r--r--	hadoop

Showing 1 to 2 of 2 entries

Hadoop, 2022.

Block information

Block 0

Block ID: 1073741862

Block Pool ID: BP-256671721-127.0.1.1-1655192105741

Generation Stamp: 1038

Size: 26

Availability:

- celab2-ThinkCentre-neo-50s-Gen-3

File contents

Big Books	2
Small Books	3

Close

Search:

Block Size	Name
MB	_SUCCESS
MB	part-r-00000

Previous 1 Next

```
home > hadoop > Desktop > WordCount.java
6 import org.apache.hadoop.io.Text;
7 import org.apache.hadoop.mapreduce.Job;
8 import org.apache.hadoop.mapreduce.Mapper;
9 import org.apache.hadoop.mapreduce.Reducer;
10 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
11 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
12 public class WordCount
13 {
14     public static class TokenizerMapper extends Mapper < LongWritable,
15         Text, Text, IntWritable >
16     {
17         private final static IntWritable one = new IntWritable(1);
18         private Text word = new Text();
19         public void map (LongWritable key, Text value, Context context) throws IOException, InterruptedException
20         {
21             StringTokenizer tokenizer = new StringTokenizer(value.toString());
22             while (tokenizer.hasMoreTokens()) {
23                 word.set(tokenizer.nextToken());
24                 if (Integer.parseInt(word.toString()) > 300) {
25                     context.write(new Text("Big Books"), one);
26                 } else {
27                     context.write(new Text("Small Books"), one);
28                 }
29             }
30         }
31     }
32
33     public static class IntSumReducer extends Reducer < Text, IntWritable,
34         Text, IntWritable >
35     {
36         private IntWritable result = new IntWritable();
37         public void reduce(Text key, Iterable < IntWritable > values, Context context) throws
```