

# BDA LAB 04

- Jainil Trivedi (CE166)

- **Aim:** "Write a map-reduce program to count the frequencies of words from distributed storage source and understand the phases involved in map-reduce programming."

1. To practice MapReduce programming in Hadoop using Java Coding.  
We create a file which contains details about the number of pages in each book.

File information - input

Download

Head the file (first 32K)

Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741846

Block Pool ID: BP-96403156-127.0.1.1-1657783463122

Generation Stamp: 1022

Size: 21

Availability:

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

File contents

350

250

150

450

120

Now, apply map reduce to find number of big books ( >300 pages) and small books ( <300 pages )

```
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()) {
                String s = tokenizer.nextToken();
                if(Integer.parseInt(s)>300)
                {
                    word.set("BigBook");
                    context.write(word,one);
                }
                else
                {
                    word.set("SmallBook");
                    context.write(word,one);
                }
            }
        }
    }
}
```

Now execute the program

```
hadoop@celab2-ThinkCentre-neo-50s-Gen-3: $ 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WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME.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 09:32:33,897 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-07-28 09:32:34,167 INFO client.DefaultHadoopFollowerProxyProvider: Connecting to ResourceManager at localhost/127.0.0.1:8032
2022-07-28 09:32:34,309 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 09:32:34,319 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1658980098398_0004
2022-07-28 09:32:34,436 INFO Input.FileInputFormat: Total input files to process : 3
2022-07-28 09:32:34,467 INFO mapreduce.JobSubmitter: number of splits:3
2022-07-28 09:32:34,575 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1658980098398_0004
2022-07-28 09:32:34,575 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-07-28 09:32:34,651 INFO conf.Configuration: resource-types.xml not found
2022-07-28 09:32:34,651 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-07-28 09:32:34,676 INFO impl.YarnClientImpl: Submitted application application_1658980098398_0004
2022-07-28 09:32:34,692 INFO mapreduce.Job: The url to track the job: http://celab2-ThinkCentre-neo-50s-Gen-3:8088/proxy/application_1658980098398_0004/
2022-07-28 09:32:34,692 INFO mapreduce.Job: Running job: job_1658980098398_0004
2022-07-28 09:32:36,711 INFO mapreduce.Job: Job job_1658980098398_0004 running in uber mode : false
2022-07-28 09:32:36,712 INFO mapreduce.Job: map 0% reduce 0%
2022-07-28 09:32:36,727 INFO mapreduce.Job: Job job_1658980098398_0004 failed with state FAILED due to: Application application_1658980098398_0004 failed 2 times due to AM container for attempt1658980098398_0004_000002 exited with exitCode: 1
Falling this attempt.Diagnostics: [2022-07-28 09:32:36.504]Exception from container-launch.
Container id: container_1658980098398_0004_02_000001
Exit code: 1

[2022-07-28 09:32:36.505]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
Caused by: java.lang.ClassNotFoundException: org.apache.hadoop.mapreduce.v2.app.MRAppMaster
```

```
Error: Could not find or load main class org.apache.hadoop.mapreduce.v2.app.MRAppMaster
Caused by: java.lang.ClassNotFoundException: org.apache.hadoop.mapreduce.v2.app.MRAppMaster

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

[2022-07-28 09:32:36.506]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
Caused by: java.lang.ClassNotFoundException: org.apache.hadoop.mapreduce.v2.app.MRAppMaster

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

For more detailed output, check the application tracking page: http://celab2-ThinkCentre-neo-50s-Gen-3:8088/cluster/app/application_1658980098
```

Now, check for the output

```
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~$ hadoop dfs -cat /output/part-r-00000
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-28 10:05:39,876 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java class
  applicable
BigBook 2
SmallBook 3
```

## JOB HISTORY

```
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop/etc/hadoop$ 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.
Starting datanodes
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting secondary namenodes [localhost]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-25 17:08:56,211 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:/opt/hadoop/etc/hadoop$ mr-jobhistory-daemon.sh --config $HADOOP_CONF_DIR start historyserver
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Use of this script to start the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon start" instead.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop/etc/hadoop$ jps
26130 NameNode
21349 ResourceManager
21686 NodeManager
26312 DataNode
27114 JobHistoryServer
26652 SecondaryNameNode
28684 Jps
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop/etc/hadoop$
```



## JobHistory

▼ Application

About Jobs

Tools

### Retired Jobs

Show 20 entries

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduced
2022.07.25 17:03:09 IST	2022.07.25 17:03:11 IST	2022.07.25 17:03:19 IST	<a href="#">job_1658748139752_0002</a>	word count	hadoop	default	SUCCEEDED	1	1	1
2022.07.25 16:53:13 IST	2022.07.25 16:53:16 IST	2022.07.25 16:53:23 IST	<a href="#">job_1658748139752_0001</a>	word count	hadoop	default	SUCCEEDED	1	1	1

Submit Tim

Start Time

Finish Time

Job ID

Name

User

Queue

State

Maps T

Maps Complet

Redu

Showing 1 to 2 of 2 entries

First

**Command History:** history 81 > commands.txt

```
728 cd ~/Desktop
729 touch books.txt
730 gedit books.txt
731 vi books.txt
732 gedit books.txt
733 hdfs dfs -moveFromLocal ~/Desktop/books.txt /user/hadoop
734 hdfs dfs -moveFromLocal ~/Desktop/books.txt /user/hadoop/input
735 hdfs dfs -ls /user/hadoop
736 hdfs dfs -mkdir -p /user/hadoop
737 hdfs dfs -ls /user/hadoop
738 hdfs dfs -moveFromLocal ~/Desktop/books.txt /user/hadoop/input
739 hdfs dfs -put ~/Desktop/books.txt /user/hadoop/input
740 clear
741 gedit touch.txt
742 hdfs dfs -put ~/Desktop/books.txt /user/hadoop/input
743 gedit books.txt
744 ls
745 rm touch.txt
746 ls
747 hdfs dfs -put ~/Desktop/books.txt /user/hadoop/input
748 clear
749 touch wc.java
750 gedit wc.java
751 clear
752 hdfs dfs -put ~/Desktop/wc.java /user/hadoop/
753 cd ~/.bash
754 cd $HADOOP_HOME
755 LS
756 ls
757 cd etc
758 ls
759 cd hadoop/
760 ls
761 gedit core-site.xml
762 clear
763 gedit core-site.xml
764 clear
765 java --verison
766 java -verison
767 java
768 java -version
769 hadoop version
```

```
770 echo $HADOOP_HOME
771 WHOAMI
772 whoami
773 stop-historyserver.sh
774 cat wordcount.java
775 clear
776 cd ..
777 clear
778 cd etc
779 cd hadoop/
780 gedit core-site.xml
781 cat core-site.xml
782 cd ..
783 ls
784 export HADOOP_CLASSPATH=$(hadoop classpath)
785 hadoop classpath
786 javac -cp `hadoop classpath` WordCount.java
787 javac -cp `hadoop classpath` /user/hadoop/wc.java
788 javac -cp `hadoop classpath` /user/hadoop/wc.java
789 clear
790 export HADOOP_CLASSPATH=$(find /user/hadoop -name '*.jar')
791 export HADOOP_CLASSPATH=$(find $HADOOP_HOME -name '*.jar')
792 javac -cp $HADOOP_CLASSPATH /user/hadoop/wc.java
793 javac -cp $HADOOP_CLASSPATH /user/hadoop/wc.java /user/hadoop/input
794 ls
795 javac -cp $HADOOP_CLASSPATH /user/hadoop/wc.java /user/hadoop/input
/user/hadoop/output
796 hadoop jar wc.jar WordCount /user/hadoop/input /user/hadoop/output
797 cd ~/Desktop
798 javac -cp $HADOOP_CLASSPATH WordCount.java
799 clear
800 javac -cp $HADOOP_CLASSPATH WordCount.java
801 clear
802 javac -cp `hadoop classpath` WordCount.java
803* jar cf wc.jar WordCount*.clas
804 clear
805 hadoop jar wc.jar WordCount /user/hadoop/input /user/hadoop/output
806 hadoop jar wc.jar WordCount /user/hadoop/input /user/hadoop/uotput2
807 history
808 history 100
809 history 82 >commands.txt
```