

LAB 6 WORD COUNT:-

CODE:-

```
import org.apache.hadoop.conf.Configuration; import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable; 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; import
org.apache.hadoop.util.GenericOptionsParser; import utils.MiscUtils;
```

```
import java.io.IOException; import java.util.*;
```

```
public class TopN {
```

```
public static void main(String[] args) throws Exception {
```

```
Configuration conf = new Configuration();
```

```
String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs(); if
(otherArgs.length != 2) {
```

```
System.err.println("Usage: TopN <in> <out>");
```

```
System.exit(2);
```

```
}
```

```
Job job = Job.getInstance(conf); job.setJobName("Top N");
```

```
job.setJarByClass(TopN.class); job.setMapperClass(TopNMapper.class);
```

```
//job.setCombinerClass(TopNReducer.class); job.setReducerClass(TopNReducer.class);
```

```
job.setOutputKeyClass(Text.class); job.setOutputValueClass(IntWritable.class);
```

```
FileInputFormat.addInputPath(job, new Path(otherArgs[0]));
```

```
FileOutputFormat.setOutputPath(job, new Path(otherArgs[1]));
```

```
System.exit(job.waitForCompletion(true) ? 0 : 1);
```

```
}
```

MapperN:-

$\ast/$

ReducerN:-

*/

```
public static class TopNReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
```

```

private Map<Text, IntWritable> countMap = new HashMap<>();

@Override

public void reduce(Text key, Iterable<IntWritable> values, Context context) throws
IOException, InterruptedException {

    // computes the number of occurrences of a single word
    int sum = 0;
    for (IntWritable val
: values) { sum += val.get();

}

    // puts the number of occurrences of this word into the map.

    // We need to create another Text object because the Text instance

    // we receive is the same for all the words
    countMap.put(new Text(key), new
IntWritable(sum));

}

@Override

protected void cleanup(Context context) throws IOException, InterruptedException {

    Map<Text, IntWritable> sortedMap = MiscUtils.sortByValues(countMap);

    int counter = 0;
    for (Text key : sortedMap.keySet()) {
        if (counter++ == 3) { break;

        }

        context.write(key, sortedMap.get(key));

    }

}

}

```

CombinerN:-

The combiner retrieves every word and puts it into a Map: if the word already exists in the * map, increments its value, otherwise sets it to 1.

*/

```
public static class TopNCombiner extends Reducer<Text, IntWritable, Text, IntWritable>
{
```

```
@Override
```

```
public void reduce(Text key, Iterable<IntWritable> values, Context context) throws
IOException, InterruptedException {
```

```
// computes the number of occurrences of a single word
int sum = 0;
for (IntWritable val : values) { sum += val.get();
```

```
}
```

```
context.write(key, new IntWritable(sum));
```

```
}
```

```
}
```

```
}
```

UtilsN:-

```
// MiscUtils.java package utils;
```

```
import java.util.*;
```

```
public class MiscUtils {
```

```
/**
```

sorts the map by values. Taken from:

<http://javarevisited.blogspot.it/2012/12/how-to-sort-hashmap-java-by-key-and-value.html>

```
*/
```

```
public static <K extends Comparable, V extends Comparable> Map<K, V>
sortByValues(Map<K, V> map) {
```

```
List<Map.Entry<K, V>> entries = new LinkedList<Map.Entry<K, V>>(map.entrySet());
```

```
Collections.sort(entries, new Comparator<Map.Entry<K, V>>() {
```

```
@Override public int compare(Map.Entry<K, V> o1, Map.Entry<K, V> o2) { return  
o2.getValue().compareTo(o1.getValue());
```

```
}
```

```
});
```

```
//LinkedHashMap will keep the keys in the order they are inserted
```

```
//which is currently sorted on natural ordering
```

```
Map<K, V> sortedMap = new LinkedHashMap<K, V>();
```

```
for (Map.Entry<K, V> entry : entries) {
```

```
sortedMap.put(entry.getKey(), entry.getValue());
```

```
}
```

```
return sortedMap;
```

```
}
```

```
}
```

OUTPUT:-

```
hduser@bmsce-Precision-T1700:~$ su hduser
```

```
Password:
```

```
hduser@bmsce-Precision-T1700:~$ start-all.sh
```

```
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
```

```
Starting namenodes on [localhost]
```

```
hduser@localhost's password:
```

```
localhost: namenode running as process 7871. Stop it first.
```

```
hduser@localhost's password:
```

```
localhost: datanode running as process 8043. Stop it first.
```

```
Starting secondary namenodes [0.0.0.0]
```

```
hduser@0.0.0.0's password:
```

```
0.0.0.0: secondarynamenode running as process 8265. Stop it first.
```

```
starting yarn daemons
```

```
resourcemanager running as process 8426. Stop it first.
```

```
hduser@localhost's password:
```

```
localhost: nodemanager running as process 8759. Stop it first.
```

```
hduser@bmsce-Precision-T1700:~$ jps
```

```
10049 Jps
```

```
8759 NodeManager
```

```
8265 SecondaryNameNode
```

```
8426 ResourceManager
```

```
6171 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
```

```
8043 DataNode
```

```
7871 NameNode
```

```
hduser@bmsce-Precision-T1700:~$ hdfs dfs -mkdir /nith
```

```
hduser@bmsce-Precision-T1700:~$ hdfs dfs -ls /
```

```
Found 21 items
```

```
drwxr-xr-x - hduser supergroup 0 2022-06-27 12:03 /Desktop
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:26 /FFF
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:28 /LLL
drwxr-xr-x - hduser supergroup 0 2022-06-27 12:19 /abc
drwxr-xr-x - hduser supergroup 0 2022-06-22 15:43 /cs228
drwxr-xr-x - hduser supergroup 0 2022-06-06 12:23 /hadoop_lab
drwxr-xr-x - hduser supergroup 0 2022-06-06 14:57 /harshita
drwxr-xr-x - hduser supergroup 0 2022-05-31 09:42 /hello
drwxr-xr-x - hduser supergroup 0 2022-06-27 15:16 /myth
drwxr-xr-x - hduser supergroup 0 2022-06-27 15:34 /nith
drwxr-xr-x - hduser supergroup 0 2019-10-24 12:14 /nnn
drwxr-xr-x - hduser supergroup 0 2019-10-24 11:02 /output
drwxr-xr-x - hduser supergroup 0 2019-10-24 12:16 /output1
drwxr-xr-x - hduser supergroup 0 2022-06-20 15:44 /rgs
drwxr-xr-x - hduser supergroup 0 2022-06-25 10:14 /ruben
drwxr-xr-x - hduser supergroup 0 2022-06-03 11:56 /sam
drwxr-xr-x - hduser supergroup 0 2022-06-03 12:39 /sampreeth
drwxr-xr-x - hduser supergroup 0 2019-10-24 11:01 /sid
drwxr-xr-x - hduser supergroup 0 2019-08-01 16:19 /tnp
drwxr-xr-x - hduser supergroup 0 2019-08-01 16:03 /user
drwxr-xr-x - hduser supergroup 0 2022-06-01 09:52 /yogesht_copted
```

```
hduser@bmsce-Precision-T1700:~$ hdfs dfs -copyFromLocal /home/hduser/Desktop/file.txt
```

```
copyFromLocal: '.': No such file or directory
```

```
hduser@bmsce-Precision-T1700:~$ hdfs dfs -copyFromLocal /home/hduser/Desktop/file.txt /nith
```

```
hduser@bmsce-Precision-T1700:~$ hdfs dfs -ls /
```

```
Found 21 items
drwxr-xr-x - hduser supergroup 0 2022-06-27 12:03 /Desktop
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:26 /FFF
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:28 /LLL
```

```
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:26 /FFF
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:28 /LLL
drwxr-xr-x - hduser supergroup 0 2022-06-27 12:19 /abc
drwxr-xr-x - hduser supergroup 0 2022-06-22 15:43 /cs228
drwxr-xr-x - hduser supergroup 0 2022-06-06 12:23 /hadoop_lab
drwxr-xr-x - hduser supergroup 0 2022-06-06 14:57 /harshita
drwxr-xr-x - hduser supergroup 0 2022-05-31 09:42 /hello
drwxr-xr-x - hduser supergroup 0 2022-06-27 15:16 /myth
drwxr-xr-x - hduser supergroup 0 2022-06-27 15:39 /nith
drwxr-xr-x - hduser supergroup 0 2019-10-24 12:14 /nnn
drwxr-xr-x - hduser supergroup 0 2019-10-24 11:02 /output
drwxr-xr-x - hduser supergroup 0 2019-10-24 12:16 /output1
drwxr-xr-x - hduser supergroup 0 2022-06-20 15:44 /rgs
drwxr-xr-x - hduser supergroup 0 2022-06-25 10:14 /ruben
drwxr-xr-x - hduser supergroup 0 2022-06-03 11:56 /sam
drwxr-xr-x - hduser supergroup 0 2022-06-03 12:39 /sampreeth
drwxr-xr-x - hduser supergroup 0 2019-10-24 11:01 /sid
drwxr-xr-x - hduser supergroup 0 2019-08-01 16:19 /tmp
drwxr-xr-x - hduser supergroup 0 2019-08-01 16:03 /user
drwxr-xr-x - hduser supergroup 0 2022-06-01 09:52 /yogesht_copied
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cat /nith/file.txt
ht all
this is demo file created on 6th of june.
this is the fourth lab program of bda lab.
hduser@bmsce-Precision-T1700:~$ hadoop jar /home/hduser/desktop/nithil.jar nithil.TopN /nith/file.txt /output_nyth
Not a valid JAR: /home/hduser/desktop/nithil.jar
hduser@bmsce-Precision-T1700:~$ hadoop jar /home/hduser/desktop/nithil.jar nithil.TopN /nith/file.txt /output_nyth
Not a valid JAR: /home/hduser/desktop/nithil.jar
hduser@bmsce-Precision-T1700:~$ hadoop jar /home/hduser/Desktop/nithil.jar nithil.TopN /nith/file.txt /output_nyth
22/06/27 15:46:39 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
22/06/27 15:46:39 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
22/06/27 15:46:39 INFO Input.FileInputFormat: Total input paths to process : 1
22/06/27 15:46:39 INFO mapreduce.JobSubmitter: number of splits:1
22/06/27 15:46:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1481381713_0001
22/06/27 15:46:39 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
22/06/27 15:46:39 INFO mapreduce.Job: Running job: job_local1481381713_0001
22/06/27 15:46:39 INFO mapred.LocalJobRunner: OutputCommitter set in config null
22/06/27 15:46:39 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
22/06/27 15:46:40 INFO mapred.LocalJobRunner: Waiting for map tasks
22/06/27 15:46:40 INFO mapred.LocalJobRunner: Starting task: attempt_local1481381713_0001_m_000000_0
22/06/27 15:46:40 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
22/06/27 15:46:40 INFO mapred.MapTask: Processing split: hdfs://localhost:54310/nith/file.txt:0+119
22/06/27 15:46:40 INFO mapred.MapTask: (EQUATOR) 0 kvl 26214396(104857584)
22/06/27 15:46:40 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
22/06/27 15:46:40 INFO mapred.MapTask: soft limit at 83886080
22/06/27 15:46:40 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
22/06/27 15:46:40 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
22/06/27 15:46:40 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
22/06/27 15:46:40 INFO mapred.LocalJobRunner:
22/06/27 15:46:40 INFO mapred.MapTask: Starting flush of map output
22/06/27 15:46:40 INFO mapred.MapTask: Spilling map output
22/06/27 15:46:40 INFO mapred.MapTask: bufstart = 0; bufend = 219; bufvoid = 104857600
22/06/27 15:46:40 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214296(104857184); length = 101/6553600
22/06/27 15:46:40 INFO mapred.MapTask: Finished spill 0
22/06/27 15:46:40 INFO mapred.Task: Task:attempt_local1481381713_0001_m_000000_0 is done. And is in the process of committing
22/06/27 15:46:40 INFO mapred.LocalJobRunner: map
```



```
Activities Terminal
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=119
File Output Format Counters
  Bytes Written=133
hduser@bmsce-Precision-T1700:~$ hdfs dfs -ls /
Found 22 items
drwxr-xr-x - hduser supergroup 0 2022-06-27 12:03 /Desktop
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:26 /FFF
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:28 /LLL
drwxr-xr-x - hduser supergroup 0 2022-06-27 12:19 /abc
drwxr-xr-x - hduser supergroup 0 2022-06-22 15:43 /cs228
drwxr-xr-x - hduser supergroup 0 2022-06-06 12:23 /hadoop_lab
drwxr-xr-x - hduser supergroup 0 2022-06-06 14:57 /harshita
drwxr-xr-x - hduser supergroup 0 2022-05-31 09:42 /hello
drwxr-xr-x - hduser supergroup 0 2022-06-27 15:16 /myth
drwxr-xr-x - hduser supergroup 0 2022-06-27 15:39 /nith
drwxr-xr-x - hduser supergroup 0 2019-10-24 12:14 /nnn
drwxr-xr-x - hduser supergroup 0 2019-10-24 11:02 /output
drwxr-xr-x - hduser supergroup 0 2019-10-24 12:16 /output1
drwxr-xr-x - hduser supergroup 0 2022-06-27 15:46 /output_nyth
drwxr-xr-x - hduser supergroup 0 2022-06-20 15:44 /rgs
drwxr-xr-x - hduser supergroup 0 2022-06-25 10:14 /ruben
drwxr-xr-x - hduser supergroup 0 2022-06-03 11:56 /sam
drwxr-xr-x - hduser supergroup 0 2022-06-03 12:39 /sampreeth
drwxr-xr-x - hduser supergroup 0 2019-10-24 11:01 /sid
drwxrwxr-x - hduser supergroup 0 2019-08-01 16:19 /tmp
drwxr-xr-x - hduser supergroup 0 2019-08-01 16:03 /user
drwxr-xr-x - hduser supergroup 0 2022-06-01 09:52 /yogesht_copied
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cat /output_nyth
cat: '/output_nyth': Is a directory
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cat /output_nyth/part-r-00000
ls 3
lab 2
of 2
this 2
on 1
all 1
hi 1
file 1
june 1
6th 1
created 1
it 1
ohh 1
my 1
the 1
program 1
demo 1
god 1
working 1
fourth 1
```

NAME:MITHIL RAJ

USN:-1BM19CS086