

**Department of Artificial Intelligence & Data Science****Vision of the Department***To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.***Mission of the Department***To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.***Session 2025-2026**

Vision: Dream of where you want.	Mission: Means to achieve Vision
---	---

Program Educational Objectives of the program (PEO): (broad statements that describe the professional and career accomplishments)

PEO1	Preparation	P: Preparation	Pep-CL abbreviation pronounce as Pep-si-IL easy to recall
PEO2	Core Competence	E: Environment (Learning Environment)	
PEO3	Breadth	P: Professionalism	
PEO4	Professionalism	C: Core Competence	
PEO5	Learning Environment	L: Breadth (Learning in diverse areas)	

Program Outcomes (PO): (statements that describe what a student should be able to do and know by the end of a program)

Keywords of POs:

Engineering knowledge, Problem analysis, Design/development of solutions, Conduct Investigations of Complex Problems, Engineering Tool Usage, The Engineer and The World, Ethics, Individual and Collaborative Team work, Communication, Project Management and Finance, Life-Long Learning

PSO Keywords: Cutting edge technologies, Research

“I am an engineer, and I know how to apply engineering knowledge to investigate, analyse and design solutions to complex problems using tools for entire world following all ethics in a collaborative way with proper management skills throughout my life.” to contribute to the development of cutting-edge technologies and Research.

Integrity: I will adhere to the Laboratory Code of Conduct and ethics in its entirety.

Name and Signature of Student and Date

(Signature and Date in Handwritten)



Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

Session	2025-26 (ODD)	Course Name	BIG DATA AND HADOOP-LAB
Semester	7 AIDS	Course Code	22ADS704
Roll No	03	Name of Student	Debasrita Chattopadhyay

Practical Number	03
Course Outcome	1. Understand big data analytics and its business applications. 2. Analyze the HADOOP and Map Reduce technologies associated with big data analytics. 3. Apply Big Data analytics Using Pig and Hive.
Aim	Write a MapReduce Program to Calculate Frequency of Words from Datasets.
Problem Definition	Write a MapReduce Program to Calculate Frequency of Words from Datasets.
Theory (100 words)	Counting the frequency of words is a typical example of the MapReduce programming model that is used in many distributed data processing frameworks like Hadoop. MapReduce repartitions processing phases or "stages" into Map and Reduce. The Map stage will splits the input data (text files) into key-value pairs, consisting of a key that is the word and the value that is usually a count of one (representing a single occurrence). At that point, the framework will "shuffle" and "sort" the key-value pairs into an intermediate format that groups all the same keys together. Finally, in the Reduce stage, the results can be summarized or aggregated for each key that will provide the frequency of each word across the input dataset. This model inherently supports scalability and parallelization of data processing across multiple nodes in a cluster.
Procedure and Execution (100 Words)	Steps of Implementation 1. Install Hadoop and configure HDFS. 2. Prepare the input dataset (text files) and upload it to HDFS. 3. Write a Mapper class that splits lines into words and emits (word, 1) pairs. 4. Write a Reducer class that sums the values for each word key.



Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

5. Write a **Driver class** to configure and run the MapReduce job.
6. Compile the Java code and create a JAR file.
7. Run the MapReduce job using the Hadoop command.
8. Retrieve the output from HDFS and display the word frequencies.

Code:

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

```
public class WordCount {
```

```
    public static class TokenizerMapper
```

```
        extends Mapper<Object, Text, Text, IntWritable> {
```

```
        private final static IntWritable one = new IntWritable(1);
```

```
        private Text word = new Text();
```

```
        public void map(Object key, Text value, Context context
        ) throws IOException, InterruptedException {
```

```
            StringTokenizer itr = new
```

```
StringTokenizer(value.toString());
```

```
            while (itr.hasMoreTokens()) {
```

```
                word.set(itr.nextToken().replaceAll("[^a-zA-Z]",
```

```
                "").toLowerCase());
```

```
                if (!word.toString().isEmpty()) {
```

```
                    context.write(word, one);
```

```
                }
```

```
            }
```



Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

```
}  
}  
  
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 IOException, InterruptedException {  
        int sum = 0;  
        for (IntWritable val : values) {  
            sum += val.get();  
        }  
        result.set(sum);  
        context.write(key, result);  
    }  
}  
  
public static void main(String[] args) throws Exception {  
    Configuration conf = new Configuration();  
    Job job = Job.getInstance(conf, "word count");  
    job.setJarByClass(WordCount.class);  
    job.setMapperClass(TokenizerMapper.class);  
    job.setCombinerClass(IntSumReducer.class);  
    job.setReducerClass(IntSumReducer.class);  
    job.setOutputKeyClass(Text.class);  
    job.setOutputValueClass(IntWritable.class);  
    FileInputFormat.addInputPath(job, new Path(args[0]));  
    FileOutputFormat.setOutputPath(job, new Path(args[1]));  
    System.exit(job.waitForCompletion(true) ? 0 : 1);  
}
```

Output:

```
theia@theiadocker-srita201326:/home/project $ curl https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz --output hadoop-3.3.6.tar.gz  
% Total % Received % Xferd Average Speed Time Time Time Current  
Dload Upload Total Spent Left Speed  
0 0 0 0 0 0 0 0 --:--:-- --:--:--  
0 0 0 0 0 0 0 0 --:--:-- 0:00:0  
0 0 0 0 0 0 0 0 --:--:-- 0:00:0  
0 0 0 0 0 0 0 0 --:--:-- 0:00:03 --:  
0 0 0 0 0 0 0 0 --:--:-- 0:00:04 --:--:--  
0 0 0 0 0 0 0 0 --:--:-- 0:00:05 --:--:--  
100 696M 100 696M 0 0 24.1M 0 0:00:28 0:00:28 --:--:-- 153M  
theia@theiadocker-srita201326:/home/project $
```



Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

```
Version: 1.7.4
Problems theia@theiadocker-srita201326: /home/project X

hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.3.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Null.java
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.3.5.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.22.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.9.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.20.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.0-alpha4.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.9.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.0-alpha2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.10.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.4.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.21.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.3.xml
hadoop-3.3.6/share/hadoop/hdfs/hadoop-hdfs-client-3.3.6-tests.jar
hadoop-3.3.6/share/hadoop/hdfs/hadoop-hdfs-httpfs-3.3.6.jar
theia@theiadocker-srita201326: /home/project$

Version: 1.7.4
Problems theia@theiadocker-srita201326: /home/project X

hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.3.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Null.java
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.3.5.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.22.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.9.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.20.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.0-alpha4.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.9.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.0-alpha2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.10.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.4.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.21.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.3.xml
hadoop-3.3.6/share/hadoop/hdfs/hadoop-hdfs-client-3.3.6-tests.jar
hadoop-3.3.6/share/hadoop/hdfs/hadoop-hdfs-httpfs-3.3.6.jar
theia@theiadocker-srita201326: /home/project$ cd hadoop-3.3.6

Version: 1.7.4
Problems theia@theiadocker-srita201326: /home/project/hadoop-3.3.6 X

hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.4.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.21.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.3.xml
hadoop-3.3.6/share/hadoop/hdfs/hadoop-hdfs-client-3.3.6-tests.jar
hadoop-3.3.6/share/hadoop/hdfs/hadoop-hdfs-httpfs-3.3.6.jar
theia@theiadocker-srita201326: /home/project$ cd hadoop-3.3.6
theia@theiadocker-srita201326: /home/project/hadoop-3.3.6$ bin/hadoop
Usage: hadoop [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]
or hadoop [OPTIONS] CLASSNAME [CLASSNAME OPTIONS]
where CLASSNAME is a user-provided Java class

OPTIONS is none or any of:
--config dir          Hadoop config directory
--debug              turn on shell script debug mode
--help              usage information
buildpaths          attempt to add class files from build tree
hostnames list[,of,host,names] hosts to use in worker mode
hosts filename      list of hosts to use in worker mode
loglevel level       set the log4j level for this command
workers             turn on worker mode

SUBCOMMAND is one of:
```



Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

Cloud IDE

Version: 1.7.4

Problems theia@theiadocker-srita201326: /home/project/hadoop-3.3.6 X

```
distcp      copy file or directories recursively
dtutil      operations related to delegation tokens
envvars     display computed Hadoop environment variables
fs          run a generic filesystem user client
gridmix     submit a mix of synthetic job, modeling a profiled from production load
jar <jar>    run a jar file. NOTE: please use "yarn jar" to launch YARN applications,
            not this command.
jnipath     prints the java.library.path
kdiag       Diagnose Kerberos Problems
kerbname    show auth_to_local principal conversion
key         manage keys via the KeyProvider
rumenfolder scale a rumen input trace
rumentrace  convert logs into a rumen trace
s3guard     S3 Commands
trace       view and modify Hadoop tracing settings
version     print the version
```

Daemon Commands:

```
kms          run KMS, the Key Management Server
registrydns  run the registry DNS server
```

SUBCOMMAND may print help when invoked w/o parameters or with -h.

```
theia@theiadocker-srita201326: /home/project/hadoop-3.3.6$ curl https://cf-courses-data.s3.us
.cloud-object-storage.appdomain.cloud/IBM-BD0225EN-SkillsNetwork/labs/data/data.txt --output
data.txt
```

Cloud IDE

Version: 1.7.4

Problems theia@theiadocker-srita201326: /home/project/hadoop-3.3.6 X

```
gridmix     submit a mix of synthetic job, modeling a profiled from production load
jar <jar>    run a jar file. NOTE: please use "yarn jar" to launch YARN applications,
            not this command.
jnipath     prints the java.library.path
kdiag       Diagnose Kerberos Problems
kerbname    show auth_to_local principal conversion
key         manage keys via the KeyProvider
rumenfolder scale a rumen input trace
rumentrace  convert logs into a rumen trace
s3guard     S3 Commands
trace       view and modify Hadoop tracing settings
version     print the version
```

Daemon Commands:

```
kms          run KMS, the Key Management Server
registrydns  run the registry DNS server
```

SUBCOMMAND may print help when invoked w/o parameters or with -h.

```
theia@theiadocker-srita201326: /home/project/hadoop-3.3.6$ curl https://cf-courses-data.s3.us
.cloud-object-storage.appdomain.cloud/IBM-BD0225EN-SkillsNetwork/labs/data/data.txt --output
data.txt
```

% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current
			Dload Upload	Total	Spent	Left	Speed
100	47	100	47	0	0	373	0 --:--:-- --:--:-- 376

```
theia@theiadocker-srita201326: /home/project/hadoop-3.3.6$
```





Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

	 <pre>theia@theiadocker-srita201326:/home/project/hadoop-3.3.6\$ ls output SUCCESS part-r-000000 theia@theiadocker-srita201326:/home/project/hadoop-3.3.6\$ cat output/part-r-000000 BigData 2 Hadoop 1 IBM 1 MapReduce 2 theia@theiadocker-srita201326:/home/project/hadoop-3.3.6\$</pre>
Output Analysis	<p>The output will be stored in the HDFS output directory in part-r-000000 file.</p> <p>The output confirms the total occurrences of each word across all input files.</p> <p>Can handle large datasets distributed over multiple nodes efficiently.</p>
Link of student Github profile where lab assignment has been uploaded	https://github.com/srita2003/BDH_Practicals/blob/main/03_Practical%20No_03_HPC.pdf
Conclusion	MapReduce Program to Calculate Frequency of Words from Datasets implemented successfully.



Nagar Yuwak Shikshan Sanstha's

Yeshwantrao Chavan College of Engineering

(An Autonomous Institution affiliated to Rashtrasant Tukadoji Maharaj Nagpur University)

Hingna Road, Wanadongri, Nagpur - 441 110

NAAC A++

Ph.: 07104-237919, 234623, 329249, 329250 Fax: 07104-232376, Website: www.ycce.edu



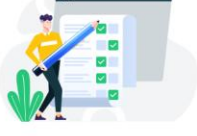
Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

Plag Report (Similarity index < 12%)	<div><div>Plagiarism Checker</div><div>Check Grammar</div><div>Detector AI</div><div>Summarize Text</div></div> <div>Counting the frequency of words is a typical example of the MapReduce programming model that is used in many distributed data processing frameworks like Hadoop. MapReduce repartitions processing phases or "stages" into Map and Reduce. The Map stage will split the input data (text files) into key-value pairs, consisting of a key that is the word and the value that is usually a count of one (representing a single occurrence). At that point, the framework will "shuffle" and "sort" the key-value pairs into an intermediate format that groups all the same keys together. Finally, in the Reduce stage, the results can be summarized or aggregated for each key that will provide the frequency of each word across the input dataset. This model inherently supports scalability and parallelization of data processing across multiple nodes in a cluster.</div> <div><div>Unique100%</div><div>Exact0%</div><div>Partial0%</div><div>View Plagiarized Sources</div><div><div>Congratulations Plagiarism not found!</div></div></div>
Date	07 /08 / 25