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.

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-lL 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)

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.

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.

	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);
	}
	}

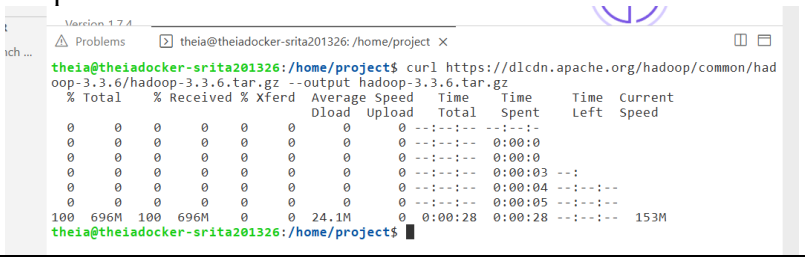
```
        }
    }

    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:

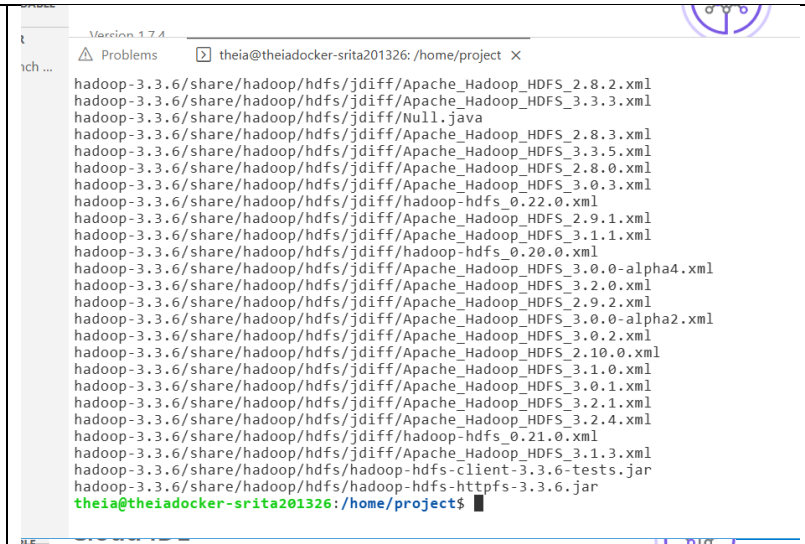


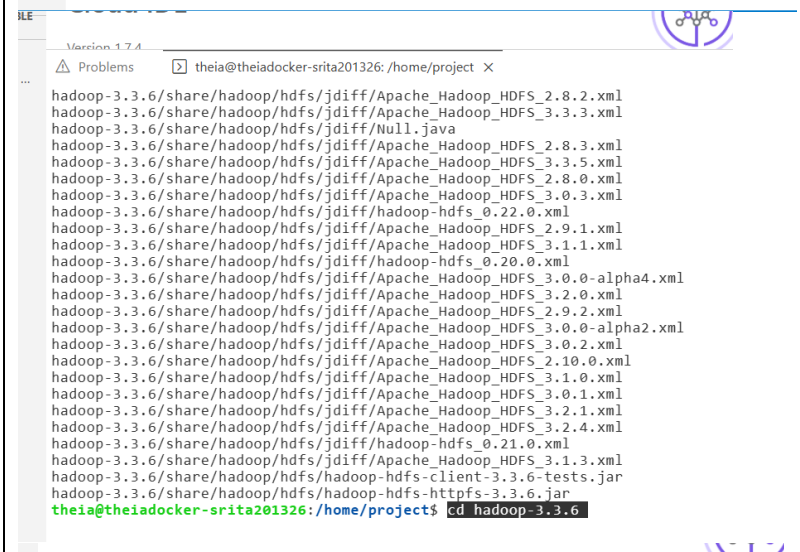
Vision of the Department

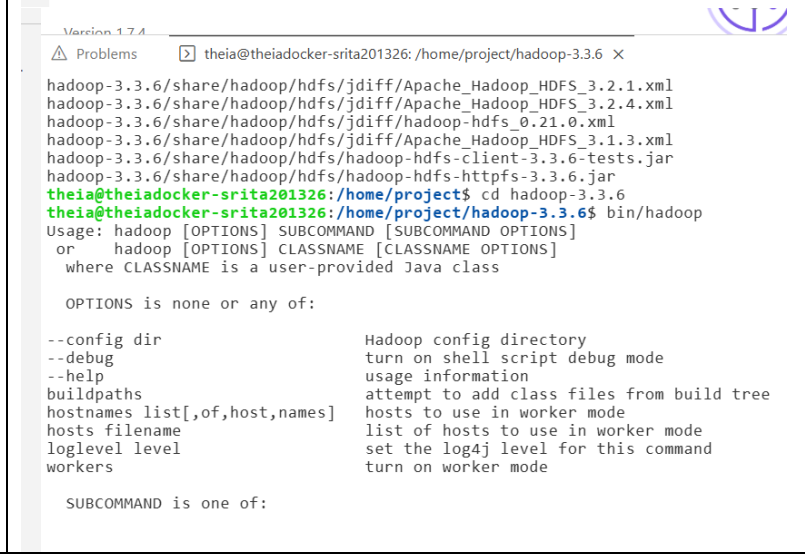
To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

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.





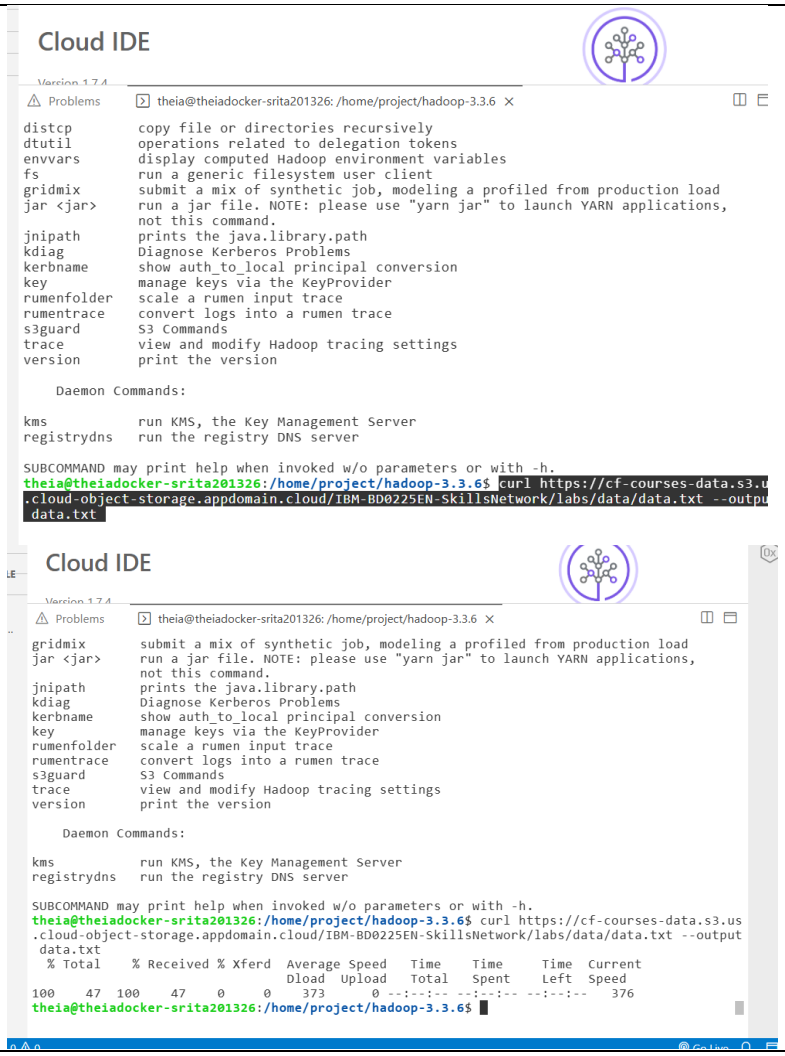


Vision of the Department

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.



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.

	
Output Analysis	The output will be stored in the HDFS output directory in **part-r-00000** file. The output confirms the total occurrences of each word across all input files. Can handle large datasets distributed over multiple nodes efficiently.
Link of student Github profile where lab assignment has been uploaded	
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

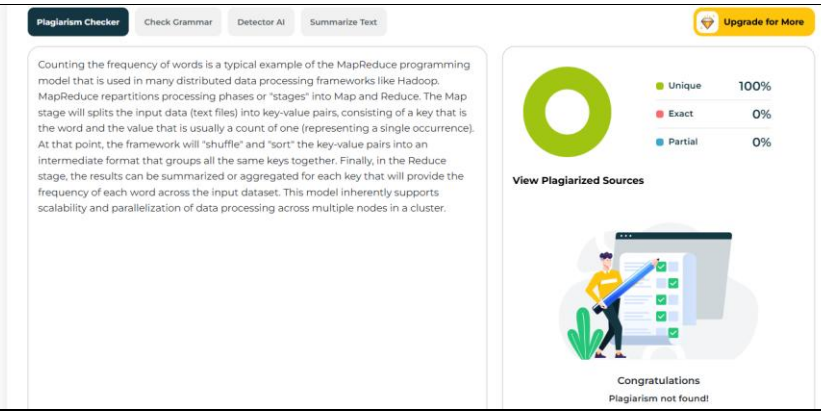
Department of Artificial Intelligence & Data Science

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%)	
Date	**07 /08 / 25**