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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

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**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

Practical Number	1
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Installation of Apache Hadoop on Linux System.
Theory (100 words)	Apache Hadoop is an open-source framework that enables distributed storage and processing of large datasets across clusters of computers. It follows the master-slave architecture with HDFS for storage and MapReduce for data processing. Installing Hadoop on a Linux system involves setting up Java, configuring environment variables, and creating a pseudo-distributed or fully distributed setup. The configuration includes editing XML files like core-site.xml, hdfs-site.xml, and mapred-site.xml to define directories, replication factors, and file system paths. Once installed, Hadoop allows efficient handling of big data through parallel computation and fault-tolerant data management.
Procedure and Execution (100 Words)	Steps of implementation: • Install Java Development Kit (JDK). • Download and extract Hadoop. • Configure environment variables in .bashrc. • Edit core, HDFS, and MapReduce XML files. • Format the NameNode. • Start NameNode and DataNode daemons. • Verify installation using Hadoop web UI or shell commands.

	Code:
	`zubair@zubair-virtual-machine:~$ java --version`
	`openjdk 11.0.14.1 2022-02-08`
	`OpenJDK Runtime Environment (build 11.0.14.1+1-Ubuntu-0ubuntu1)`
	`OpenJDK 64-Bit Server VM (build 11.0.14.1+1-Ubuntu-0ubuntu1, mixed mode, sharing)`
	`zubair@zubair-virtual-machine:~$`

<table>
<tr><td></td><td>



```
zubair@zubair-virtual-machine:~$ sudo wget https://downloads.apache.org/hadoop/c
ommon/hadoop-3.3.2/hadoop-3.3.2.tar.gz.sha512
[sudo] password for zubair:
--2022-04-20 07:58:10--  https://downloads.apache.org/hadoop/common/hadoop-3.3.2
/hadoop-3.3.2.tar.gz.sha512
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 135.181.2
14.104, 2a01:4f8:10a:201a::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|88.99.95.219|:443... c
onnected.
HTTP request sent, awaiting response... 200 OK
Length: 160 [text/plain]
Saving to: 'hadoop-3.3.2.tar.gz.sha512'

hadoop-3.3.2.tar.gz 100%[===================>]     160  --.-KB/s    in 0s

2022-04-20 07:58:10 (21.4 MB/s) - 'hadoop-3.3.2.tar.gz.sha512' saved [160/160]
```

```
zubair@zubair-virtual-machine:~$ cd Downloads/
zubair@zubair-virtual-machine:~/Downloads$ sudo mv hadoop-3.3.2 /usr/local/hadoo
p
```

```
zubair@zubair-virtual-machine:~$ readlink -f /usr/bin/java | sed "s:bin/java::"
/usr/lib/jvm/java-11-openjdk-amd64/
zubair@zubair-virtual-machine:~$ sudo nano /usr/local/hadoop/etc/hadoop/hadoop-e
nv.sh
```

```
zubair@zubair-virtual-machine:~$ sudo /usr/local/hadoop/bin/hadoop
Usage: hadoop [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]
 or    hadoop [OPTIONS] CLASSNAME [CLASSNAME OPTIONS]
  where CLASSNAME is a user-provided Java class
```

</td></tr>
<tr><td>Output Analysis</td><td>After installation, Hadoop services such as NameNode, DataNode, ResourceManager, and NodeManager start successfully. The web interfaces display cluster information and running nodes. The hadoop fs -ls / command confirms access to HDFS, verifying that Hadoop is properly configured and operational.</td></tr>
<tr><td>Github Link</td><td>https://github.com/yashsgiri/BDH-Lab/upload</td></tr>
<tr><td>Conclusion</td><td>Installing Apache Hadoop on Linux provides a robust environment for distributed data storage and parallel processing. Once set up, the framework can efficiently handle large-scale datasets, forming the foundation for big data analytics and scalable computation.</td></tr>
</table>

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 Computer Technology

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	30 October 2025

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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

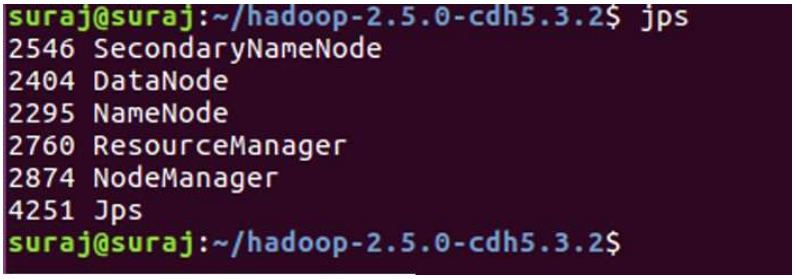
Vision of the Department

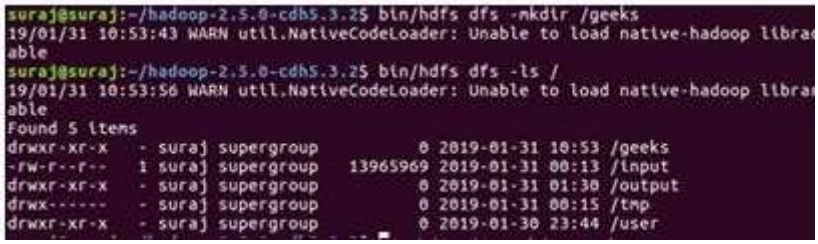
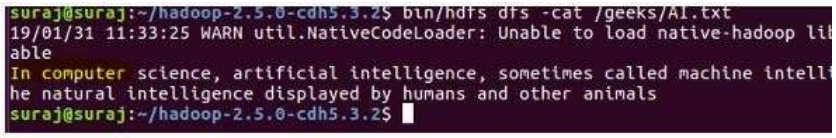
To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

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**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

Practical Number	2
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Implementation of File management operations in Hadoop.
Theory (100 words)	Hadoop Distributed File System (HDFS) provides scalable and reliable storage for big data applications. It allows users to manage files across a distributed cluster using commands similar to traditional UNIX file systems. File management operations in Hadoop include creating directories, uploading and downloading files, viewing file contents, and deleting files from HDFS. These operations are performed through the Hadoop shell or API commands, ensuring data replication, fault tolerance, and efficient access. Understanding these operations is crucial for effectively handling data within the Hadoop ecosystem.
Procedure and Execution (100 Words)	Steps of implementation: • Start Hadoop services (NameNode and DataNode). • Create a directory in HDFS using hadoop fs -mkdir. • Upload files with hadoop fs -put. • View files using hadoop fs -ls or -cat. • Download files using hadoop fs -get. • Delete files or directories using hadoop fs -rm or -rmdir.
	Code: 

	    
Output Analysis	Each command execution confirms successful file operations in HDFS. For instance, files uploaded with -put appear when listing directories, and -cat displays their contents. The successful execution of create, view, and delete commands verifies that the Hadoop file management system is functioning properly.
Github Link	https://github.com/yashsgiri/BDH-Lab/upload
Conclusion	Implementing file management operations in Hadoop demonstrates efficient handling of files within a distributed system. These commands help manage large datasets seamlessly, ensuring scalability, reliability, and accessibility across the Hadoop cluster.

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 Computer Technology

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	30 October 2025

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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

Vision of the Department

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

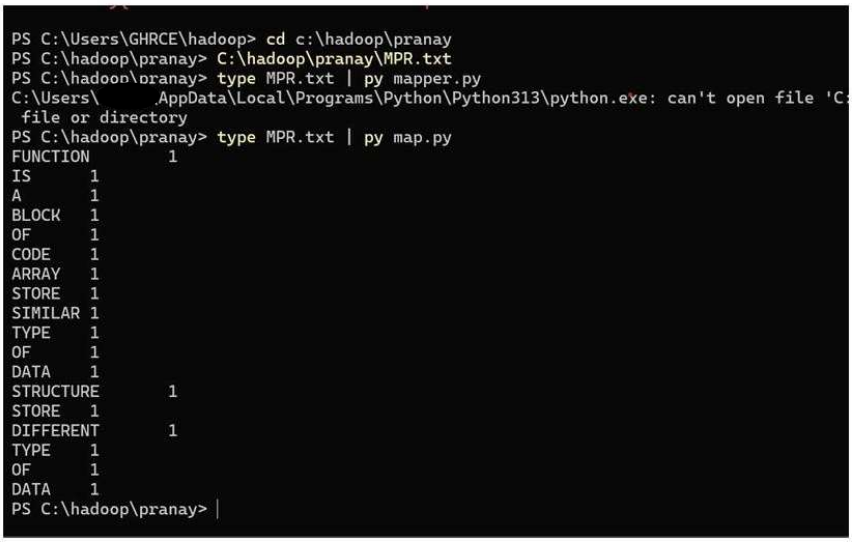
Mission of the Department

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**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

Practical Number	3
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Write a MapReduce Program to Calculate Frequency of Words from Datasets.
Theory (100 words)	MapReduce is a programming model in Hadoop used to process large datasets in parallel. It works in two main phases: the Mapper and the Reducer. In a word frequency program, the Mapper reads text input and emits each word as a key with a count of one. The Reducer then aggregates these counts for each unique word, producing the total frequency of each word in the dataset. This approach efficiently distributes processing tasks across multiple nodes, enabling fast and scalable text analysis on big data stored in HDFS.
Procedure and Execution (100 Words)	Steps of implementation: • Prepare input text files and store them in HDFS. • Create Mapper class – split lines into words and emit (word, 1). • Create Reducer class – sum counts for each word. • Set up the Driver program to configure and run the job. • Compile and run the program on Hadoop. • View the output file in HDFS to see word frequencies.
	Code: 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.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;

```
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());
            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: 
Output Analysis	The output displays each unique word from the dataset followed by its frequency count. For example: Hadoop 3, Big 2, Data 5. This confirms that the MapReduce program successfully reads, processes, and counts words in parallel across the cluster.
Github Link	https://github.com/yashsgiri/BDH-Lab/upload
Conclusion	The MapReduce word frequency program demonstrates Hadoop's capability to perform distributed data processing. By splitting and aggregating data across nodes, it efficiently computes word occurrences in large datasets, showcasing Hadoop's strength in handling scalable and parallel data analysis tasks.

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 Computer Technology

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	30 October 2025

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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

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**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

Practical Number	4
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Installation of Apache Hive on Linux with Hadoop Integration.
Theory (100 words)	Apache Hive is a data warehouse system built on top of Hadoop that provides a SQL-like interface (HiveQL) for querying and managing large datasets stored in HDFS. It converts SQL queries into MapReduce or Tez/Spark jobs, enabling users to perform data analysis without complex programming. Installing Hive on a Linux system integrated with Hadoop involves configuring environment variables, setting up Hive metastore, and connecting it to Hadoop's HDFS. This setup allows users to manage structured data efficiently using queries while leveraging Hadoop's distributed processing power.
Procedure and Execution (100 Words)	Steps of implementation: • Ensure Hadoop is installed and running. • Download and extract Apache Hive. • Configure environment variables in .bashrc. • Edit hive-site.xml to set metastore and warehouse paths. • Initialize the Hive metastore database. • Start Hive using the hive command. • Create and query tables to verify integration with Hadoop.
	Code: 

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

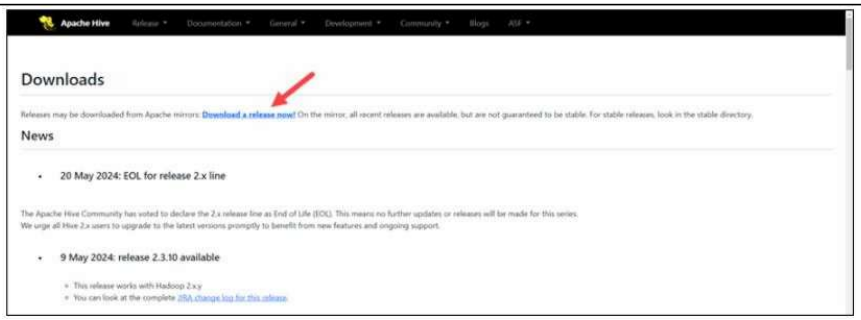
Department of Computer Technology

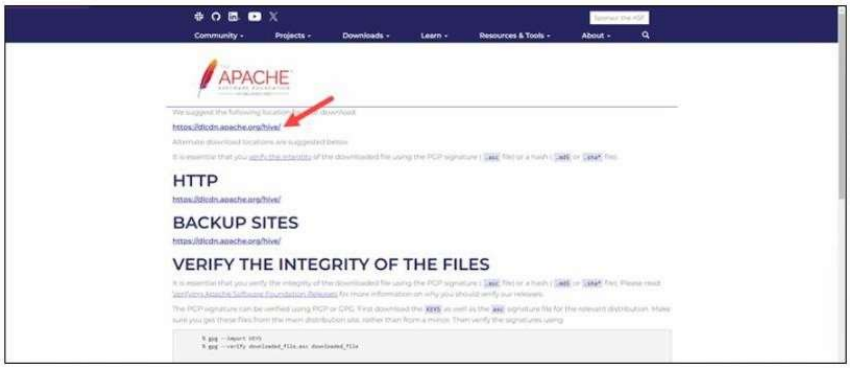
Vision of the Department

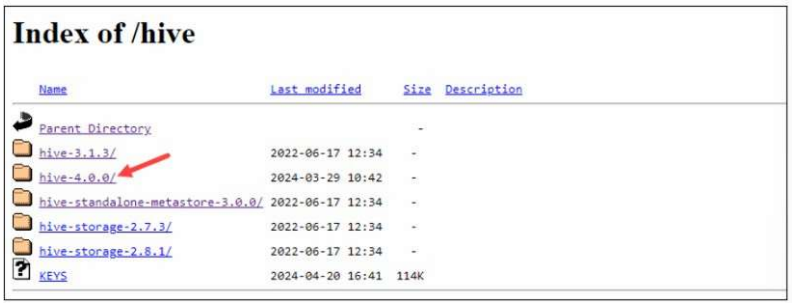
To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

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: 
Output Analysis	After successful installation, Hive starts with the prompt hive>. Commands like CREATE TABLE, LOAD DATA, and SELECT execute successfully using HDFS as storage. Query outputs confirm proper communication between Hive and Hadoop, ensuring seamless data processing and retrieval.
Github Link	https://github.com/yashsgiri/BDH-Lab/upload
Conclusion	Installing Apache Hive on Linux with Hadoop integration enables efficient management and querying of large datasets through a familiar SQL-like interface. This setup simplifies big data analytics by combining Hive's ease of use with Hadoop's scalability and distributed processing power.
Plag Report (Similarity index < 12%)	
Date	30 October 2025

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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

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**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

Practical Number	5
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Perform Hive Operations: Create, Alter and Drop Databases, Tables, Views, and Indexes.
Theory (100 words)	Apache Hive provides a SQL-like interface for managing structured data in Hadoop. It allows users to perform Data Definition Language (DDL) operations such as creating, altering, and dropping databases, tables, views, and indexes. These operations help organize and optimize large datasets stored in HDFS. Databases group related tables, tables store structured data, views act as virtual tables for simplified queries, and indexes improve query performance. Hive translates these SQL operations into MapReduce or Tez/Spark jobs, enabling scalable and efficient data management within the Hadoop ecosystem.
Procedure and Execution (100 Words)	Steps of implementation: • Start Hive shell using hive command. • Create a database: CREATE DATABASE dbname; • Use the database: USE dbname; • Create a table: CREATE TABLE tablename (...); • Alter table: ALTER TABLE tablename ADD COLUMNS (...); • Drop table: DROP TABLE tablename; • Create a view: CREATE VIEW viewname AS SELECT ...; • Drop view: DROP VIEW viewname; • Create index: CREATE INDEX idx_name ON TABLE tablename (column); • Drop index: DROP INDEX idx_name ON tablename;

	Code:
	`hive> create database test;`
	`OK`
	`Time taken: 0.2 seconds`
	`hive> use test;`
	`OK`
	`Time taken: 0.141 seconds`

```
hive> show tables;
OK
values__tmp__table__3
Time taken: 0.186 seconds, Fetched: 1 row(s)
hive> create table test.emp
    > (
    > sno int,
    > usr_name string,
    > city string)
    > ROW FORMAT delimited fields terminated by ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE;
OK
Time taken: 0.77 seconds
hive> show tables;
OK
emp
values__tmp__table__3
Time taken: 0.111 seconds, Fetched: 2 row(s)
hive>
```

```
hive> DROP DATABASE IF EXISTS Test;
OK
Time taken: 0.055 seconds
hive> DROP DATABASE Test1;
OK
Time taken: 0.033 seconds
hive> show databases;
OK
default
Time taken: 0.012 seconds, Fetched: 1 row(s)
hive>
```

```
hive> create view region_wise_profit as
    > select region, count(country), cast(avg(total_profit) as int) from sales_tracker group by region
    > ;
OK
Time taken: 0.139 seconds
hive> select * from region_wise_profit;
Query ID = root_20200513001212_cb01727d-7bcf-4370-8bab-2d2115c15843
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1589327438560_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/applicatio
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1589327438560_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-05-13 00:20:12,484 Stage-1 map = 0%,  reduce = 0%
2020-05-13 00:20:17,664 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.06 sec
2020-05-13 00:20:24,996 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.86 sec
MapReduce Total cumulative CPU time: 2 seconds 860 msec
Ended Job = job_1589327438560_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.86 sec   HDFS Read: 34575 HDFS Write: 199 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 860 msec
OK
Asia    25      349386
Australia and Oceania   12      280747
Central America and the Caribbean       14      371765
Europe  60      465092
Middle East and North Africa    24      422257
North America   8       413719
Sub-Saharan Africa      56      349561
Time taken: 20.359 seconds, Fetched: 7 row(s)
hive>
```

	Output:
Output Analysis	Each command executes successfully, displaying confirmation messages like "Table created," "Database dropped," or "Index created." Listing commands (SHOW DATABASES;, SHOW TABLES;) verify the changes. The output demonstrates successful management of Hive databases, tables, views, and indexes within Hadoop.
Github Link	https://github.com/yashsgiri/BDH-Lab/upload
Conclusion	Performing Hive operations such as creating, altering, and dropping databases, tables, views, and indexes enables structured organization and optimized querying of big data. These operations simplify data management while leveraging Hadoop's scalability and Hive's SQL-like interface for efficient analytics.
Plag Report (Similarity index < 12%)	
Date	30 October 2025

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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

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**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

Practical Number	6
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Perform Case Study: Analyzing Olympic Data Set Using Hive.
Theory (100 words)	Apache Hive enables large-scale data analysis using SQL-like queries on data stored in Hadoop. In this case study, the Olympic dataset containing details like athletes, countries, sports, medals, and years is analyzed using Hive. By loading the dataset into Hive tables, users can execute queries to extract insights such as total medals by country, top athletes, or performance trends over the years. Hive translates these queries into MapReduce or Tez/Spark jobs, enabling efficient distributed processing of massive datasets and simplifying analytics through its structured query interface.

Procedure and Execution (100 Words)	Steps of implementation:
	• Start Hadoop and Hive services.
	• Create a database, e.g., CREATE DATABASE olympics;
	• Create an external table with appropriate schema.
	• Load the dataset using LOAD DATA INPATH '/path/olympic.csv' INTO TABLE olympic;
	• Run queries like:
	  ○ SELECT country, COUNT(medal) FROM olympic GROUP BY country;
	  ○ SELECT athlete, COUNT(*) FROM olympic GROUP BY athlete ORDER BY COUNT(*) DESC;
	• Save or export query results if needed.

Department of Computer Technology

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.

Code:

```
hive> create table num_of_sports_by_olympic as select city, year, count(distinct sport) as no_of_sports from summer_olympics group by city, year;
Query ID = hduser_20190427163204_8354edce-e9ed-49f2-b47f-f462b8a64384
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1556275802324_0009, Tracking URL = http://shikhar-VirtualBox:8088/proxy/application_1556275802324_0009/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1556275802324_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-04-27 16:32:34,693 Stage-1 map = 0%,  reduce = 0%
2019-04-27 16:32:50,593 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.88 sec
2019-04-27 16:33:06,966 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.93 sec
MapReduce Total cumulative CPU time: 8 seconds 930 msec
Ended Job = job_1556275802324_0009
Moving data to directory hdfs://localhost:54310/user/hive/warehouse/pda_project.db/num_of_sports_by_olympic
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.93 sec   HDFS Read: 11578330 HDFS Write: 593 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 930 msec
OK
Time taken: 66.2 seconds
hive> select * from num_of_sports_by_olympic limit 5;
OK
Athina  1896    2
Paris   1900    6
St. Louis       1904    8
Athina  1906    8
London  1908    11
Time taken: 0.706 seconds, Fetched: 5 row(s)
```

```
hive> create table num_of_players_by_olympic as select city, year, count(distinct name) as no_of_players from summer_olympics group by city, year;
Query ID = hduser_20190427163626_167a525e-b952-4317-8870-e5894568dc26
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1556275802324_0010, Tracking URL = http://shikhar-VirtualBox:8088/proxy/application_1556275802324_0010/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1556275802324_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-04-27 16:36:45,608 Stage-1 map = 0%,  reduce = 0%
2019-04-27 16:37:04,973 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 8.55 sec
2019-04-27 16:37:21,900 Stage-1 map = 100%,  reduce = 89%, Cumulative CPU 13.8 sec
2019-04-27 16:37:22,976 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 14.8 sec
MapReduce Total cumulative CPU time: 14 seconds 800 msec
Ended Job = job_1556275802324_0010
Moving data to directory hdfs://localhost:54310/user/hive/warehouse/pda_project.db/num_of_players_by_olympic
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 14.8 sec   HDFS Read: 11578421 HDFS Write: 640 SUCCESS
Total MapReduce CPU Time Spent: 14 seconds 800 msec
OK
Time taken: 58.213 seconds
hive> select * from num_of_players_by_olympic limit 5;
OK
Athina  1896    11
Paris   1900    30
St. Louis       1904    40
Athina  1906    51
London  1908    132
Time taken: 0.666 seconds, Fetched: 5 row(s)
```

```
hive> create view sports_and_players_by_olympic  as select a.city, a.year, a.no_of_sports, b.no_of_players from num_of_sports_by_olympic a inner join num_of_players_by_olympic b on a.city=b.city and a.year=b.yea
r;
OK
Time taken: 1.593 seconds
hive> show views;
OK
sports_and_players_by_olympic
Time taken: 0.121 seconds, Fetched: 1 row(s)
hive> select * from sports_and_players_by_olympic limit 5;
Query ID = hduser_20190427164616_0d2a0cd5-accb-4977-8718-8df5a03254fb
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1556275802324_0011, Tracking URL = http://shikhar-VirtualBox:8088/proxy/application_1556275802324_0011/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1556275802324_0011
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2019-04-27 16:47:06,024 Stage-3 map = 0%,  reduce = 0%
2019-04-27 16:47:27,618 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 4.32 sec
MapReduce Total cumulative CPU time: 4 seconds 320 msec
Ended Job = job_1556275802324_0011
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 4.32 sec   HDFS Read: 10999 HDFS Write: 236 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 320 msec
OK
Athina  1896    2       11
Paris   1900    6       30
St. Louis       1904    8       40
Athina  1906    8       51
London  1908    11      132
Time taken: 73.784 seconds, Fetched: 5 row(s)
```

```
hive> create view Last_4_Olympics as select * from summer_olympics where year in ('2016','2012','2008','2004');
OK
Time taken: 0.569 seconds
hive> select * from Last_4_Olympics limit 10;
OK
13877   Abdellah Fall   M       32.0    163.0   63.0    Morocco MAR     2008 Summer     2008    Summer  Beijing Athletics       Athletics Men's 10      000 metres
13873   Katelyn Falgowski       F       19.0    168.0   63.0    United States   USA     2008 Summer     2008    Summer  Beijing Hockey  Hockey Women's Hockey   No Medal
13873   Dana Faletic    F       27.0    178.0   74.0    Australia       AUS     2004 Summer     2004    Summer  Athina  Rowing  Rowing Women's Quadruple Sculls Bronze
13870   Ghor Faldun     M       28.0    172.0   65.0    Hungary HUN     2016 Summer     2016    Summer  Rio de Janeiro  Triathlon       Triathlon Men's Olympic Distance        No Medal
13865   Nara Lorena Falco Artega        F       22.0    168.0   55.0    Mexico  MEX     2004 Summer     2004    Summer  Athina  Synchronized Swimming   Synchronized Swimming Women's Duet      No Medal
13850   Yanaguchi Falco Florentina      M       24.0    181.0   77.0    Brazil  BRA     2012 Summer     2012    Summer  London  Boxing  Boxing Men's Light-Heavyweight  Bronze
13844   Marinella Falco F       18.0    167.0   58.0    Italy   ITA     2004 Summer     2004    Summer  Athina  Rhythmic Gymnastics     Rhythmic Gymnastics Women's Group       Silver
13844   Natlia Falavigna Silva  F      20.0    178.0   69.0    Brazil  BRA     2004 Summer     2004    Summer  Athina  Taekwondo       Taekwondo Women's Heavyweight   No Medal
13840   Lucia Falasca   F       23.0    171.0   64.0    Argentina       ARG     2016 Summer     2016    Summer  Rio de Janeiro  Sailing Sailing Women's One Person Dinghy        No Medal
13835   Omnia Abdel Khalek Fakhry       F       26.0    165.0   53.0    Egypt   EGY     2008 Summer     2008    Summer  Beijing Modern Pentathlon       Modern Pentathlon Women's Individual    No Medal
Time taken: 0.621 seconds, Fetched: 10 row(s)
```

Output Analysis	Hive queries produce analytical summaries such as the number of medals won by each country, the most successful athletes, and trends across different Olympic years. The results confirm Hive's ability to efficiently process and analyze large structured datasets stored in HDFS.

Github Link	
Conclusion	Analyzing the Olympic dataset using Hive demonstrates how big data analytics can be simplified through HiveQL. The case study highlights Hive's capability to manage and query large datasets efficiently, providing meaningful insights from complex data with minimal coding effort.
Plag Report (Similarity index < 12%)	
Date	30 October 2025

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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

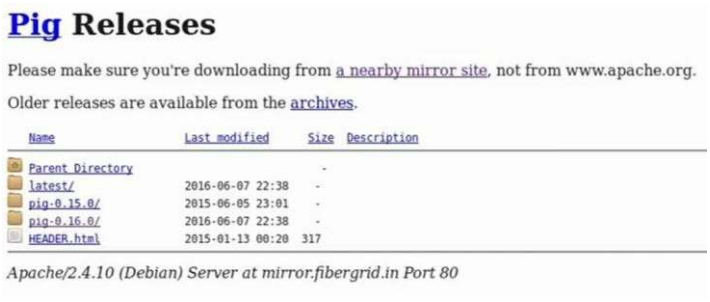
Vision of the Department

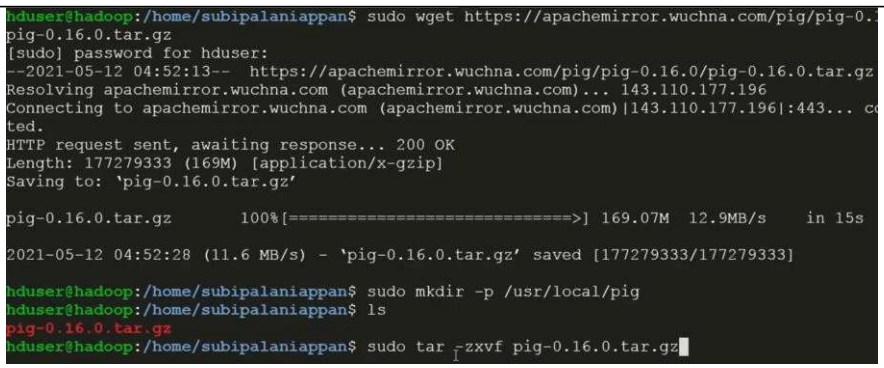
To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

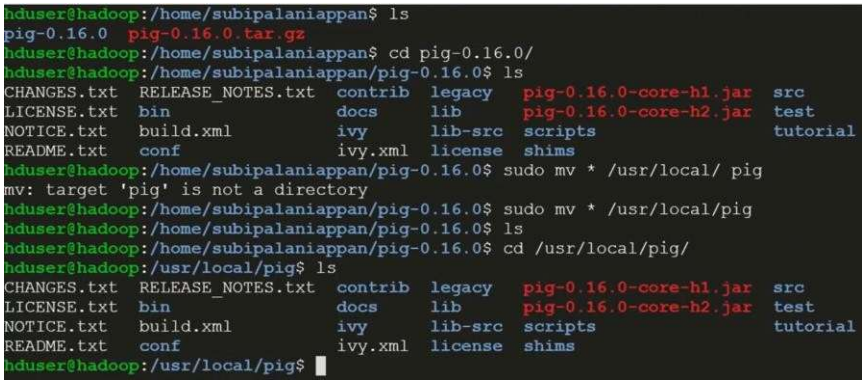
Mission of the Department

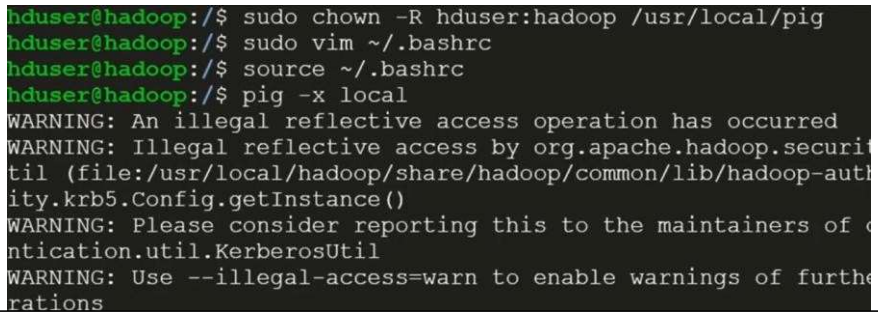
To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

Session	2025-26 (ODD)	**Course Name**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

Practical Number	7
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Installation of Apache Pig on Linux with Hadoop Integration
Theory (100 words)	Apache Pig is a high-level platform built on top of Hadoop for processing and analyzing large datasets using a scripting language called Pig Latin. It simplifies the development of MapReduce programs by providing easy-to-understand commands for data transformation, filtering, and aggregation. Installing Pig on a Linux system integrated with Hadoop allows users to execute Pig scripts that run as MapReduce jobs on the cluster. This integration enables efficient handling of structured and semi-structured data without writing complex Java code, making Pig a powerful tool for big data analytics.
Procedure and Execution (100 Words)	Steps of implementation: • Ensure Hadoop is installed and running. • Download and extract Apache Pig. • Configure environment variables in .bashrc (set PIG_HOME and update PATH). • Verify Java and Hadoop paths in Pig configuration. • Start Pig in local or MapReduce mode using pig command. • Run sample Pig Latin scripts to test Hadoop integration.
	Code:  **Pig Releases** Please make sure you're downloading from a nearby mirror site, not from www.apache.org. Older releases are available from the archives. Name Last modified Size Description Parent Directory - latest/ 2016-06-07 22:38 - pig-0.15.0/ 2015-06-05 23:01 - pig-0.16.0/ 2016-06-07 22:38 - HEADER.html 2015-01-13 00:20 317 *Apache/2.4.10 (Debian) Server at mirror.fibergrid.in Port 80*

<table>
<tr><td></td><td>






</td></tr>
<tr><td>Output Analysis</td><td>After installation, the pig shell starts successfully, displaying the grunt> prompt. Executing Pig Latin scripts (like word count) produces output stored in HDFS. Successful job completion messages confirm that Pig is properly integrated with Hadoop and capable of running MapReduce tasks.</td></tr>
<tr><td>Github Link</td><td>https://github.com/yashsgiri/BDH-Lab/upload</td></tr>
<tr><td>Conclusion</td><td>Installing Apache Pig on Linux with Hadoop integration provides a simplified way to process large datasets using Pig Latin scripts. This setup enhances productivity by abstracting complex MapReduce logic and enabling efficient big data analysis within the Hadoop ecosystem.</td></tr>
</table>

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 Computer Technology

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	30 October 2025

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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

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**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

Practical Number	8
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Perform Pig Operations: Load & Store Data, Aggregation Operations, Filtering Data and Joining Datasets.
Theory (100 words)	Apache Pig is a data flow scripting platform built on Hadoop that uses Pig Latin for analyzing large datasets. It simplifies complex MapReduce operations with simple commands for loading, transforming, and storing data. Pig allows users to load data from HDFS or local systems, apply filters, perform aggregation functions (like COUNT, SUM, AVG), and join multiple datasets. These operations are internally converted into MapReduce jobs, making Pig an efficient tool for scalable data analysis while maintaining simplicity and flexibility in handling structured and semi-structured data.

Procedure and Execution (100 Words)	Steps of implementation:
	• Start Hadoop and Pig.
	• Load data: A = LOAD '/input/data.csv' USING PigStorage(',') AS (id:int, name:chararray, marks:int);
	• Filter data: B = FILTER A BY marks > 50;
	• Perform aggregation: C = GROUP B BY name; D = FOREACH C GENERATE group, COUNT(B);
	• Join datasets: E = JOIN A BY id, Other BY id;
	• Store output: STORE D INTO '/output/result' USING PigStorage(',');
	
	Code:
	`grunt> a = load 'retail_stage1.order_items' using org.apache.hive.hcatalog.pig.HCatLoader();`
	`grunt> b = group a by order_item_order_id`
	`grunt> c = foreach b generate group, AVG(a.order_item_product_price)`

```
grunt> dump c;
```

```
hive> select order_item_order_id, AVG(order_item_product_price)
    > from order_items
    > group by order_item_order_id;
Query ID = cloudera_20160913214646_7a4abeba-4526-4028-8da3-97c5f5e10575
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
```

```
grunt> a = '/user/cloudera/department' using PigStorage(',');
2016-08-29 21:36:27,468 [main] ERROR org.apache.pig.tools.grunt.Grunt
rror, unexpected symbol at or near 'a'
Details at logfile: /home/cloudera/pig_1472530648111.log
```

```
grunt> a = load '/user/cloudera/department' using PigStorage(',');
2016-08-29 21:36:39,833 [main] WARN  org.apache.pig.PigServer - Enc
s).
```

```
grunt> c = filter a by $0 > 4;
```

Output:

```
(68852,50.0)
(68855,199.99000549316406)
(68856,129.99000549316406)
(68858,481.99199905395506)
(68859,699.9899965922037)
(68860,64.98499965667725)
(68861,103.32666905721028)
(68862,331.98400497436523)
(68863,101.9900032043457)
(68865,399.9800109863281)
(68866,119.99000358581543)
(68868,114.9900016784668)
(68869,409.98999786376953)
(68870,59.9900016784668)
(68871,174.99000549316406)
(68873,131.9880027770996)
(68875,1049.9899940490723)
(68878,154.98750591278076)
(68879,419.99000040690106)
(68880,79.9860008239746)
(68881,129.99000549316406)
(68882,54.9950008392334)
(68883,1024.9949951171875)
```

	
Output Analysis	The Pig script runs successfully and outputs processed data stored in HDFS. Results show filtered records, aggregated values, and joined datasets. Each operation generates a corresponding MapReduce job, confirming correct data loading, filtering, and transformation within the Hadoop environment.
Github Link	https://github.com/yashsgiri/BDH-Lab/upload
Conclusion	Performing Pig operations like load, store, aggregation, filtering, and joining demonstrates how Pig simplifies big data processing. These operations efficiently transform and analyze datasets through concise Pig Latin scripts, leveraging Hadoop's distributed framework for large-scale data computation.
Plag Report (Similarity index < 12%)	
Date	30 October 2025

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 Computer Technology

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: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	**Mission:** To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.

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.

Yash Giri 30 October 2025
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 Computer Technology

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

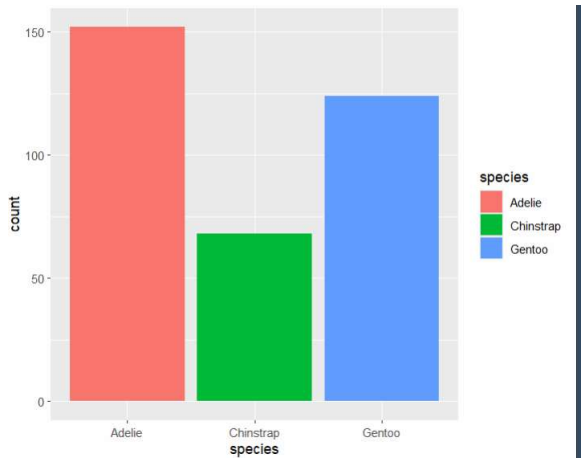
To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

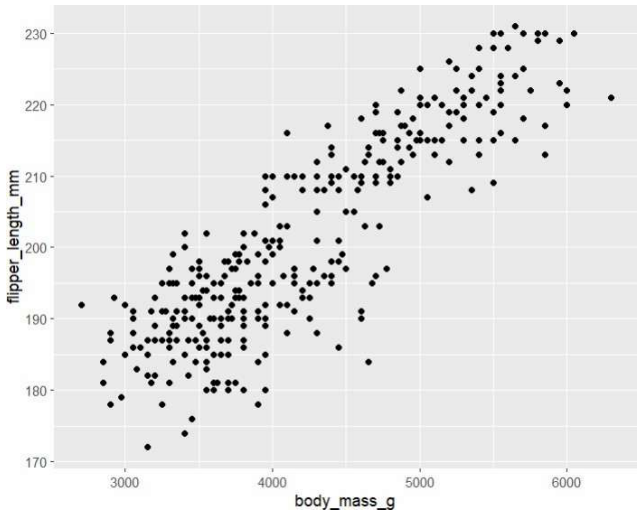
Session	2025-26 (ODD)	**Course Name**	BDH Lab
Semester	7	**Course Code**	22ADS704
Roll No	71	**Name of Student**	Yash Giri

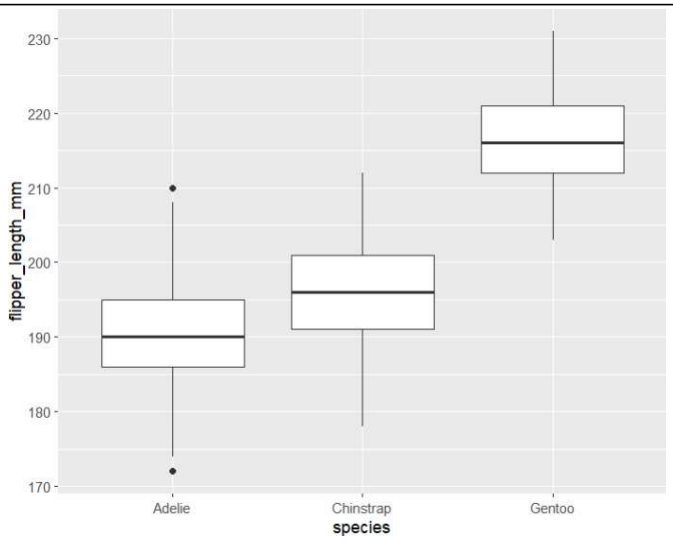
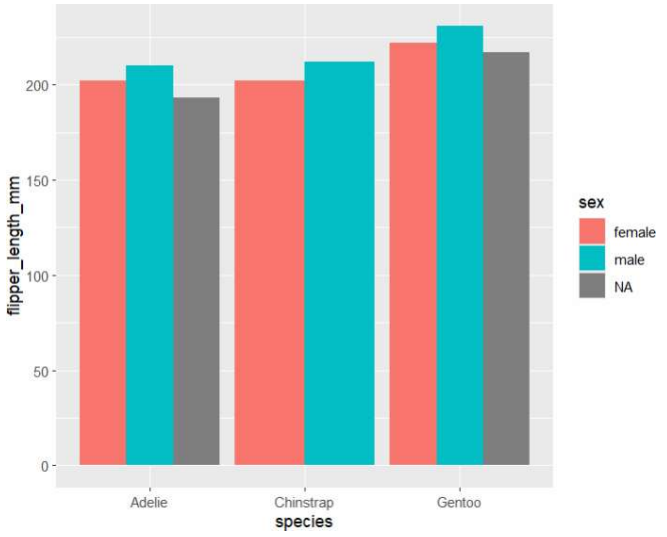
Practical Number	9
Course Outcome	**CO1:** Understand big data analytics and its business applications. **CO2:** Analyze the HADOOP and Map Reduce technologies associated with big data analytics. **CO3:** Apply Big Data Analytics Using Pig and Hive.
Aim	Analyzing Various Data Visualization Methods Using R
Theory (100 words)	R is a powerful programming language widely used for statistical computing and data visualization. It provides rich libraries like ggplot2, lattice, and plotly for creating various types of visualizations, including bar charts, histograms, scatter plots, box plots, and line graphs. Data visualization in R helps interpret complex datasets through graphical representations, revealing trends, patterns, and relationships. By analyzing different visualization methods, users can choose the most appropriate plots to effectively communicate insights and make data-driven decisions in research, analytics, and business intelligence.
Procedure and Execution (100 Words)	Steps of implementation: • Install and open R/RStudio. • Import dataset using read.csv() or similar functions. • Use base R plots like plot(), hist(), barplot(). • Install and load ggplot2 package. • Create advanced plots: ggplot(data) + geom_bar(), geom_point(), geom_boxplot(). • Customize plots with titles, labels, and colors. • Optionally use interactive libraries like plotly for dynamic visualizations.
	Code: library(tidyverse) library(palmerpenguins) # Bar Plot penguins |> ggplot(aes(x = species)) + geom_bar(aes(fill = species)) # Scatter Plot penguins |> ggplot(aes(x = body_mass_g, y = flipper_length_mm)) + geom_point()

```
# Box Plot
penguins |> ggplot(aes(x = species, y = flipper_length_mm)) +
geom_boxplot()
# Line Plot
penguins |> ggplot(aes(x = body_mass_g, y = flipper_length_mm)) +
geom_line()
# Stacked Bar Plot
ggplot(penguins, aes(fill=sex, y=flipper_length_mm, x=species)) +
geom_bar(position="dodge", stat="identity")
# Histogram Plot
penguins |> ggplot(aes(x = body_mass_g)) + geom_histogram()
```

Output:





	 
Output Analysis	Various graphs such as bar charts, scatter plots, and histograms are successfully generated, visually representing data trends and relationships. Each plot highlights different aspects of the dataset, making analysis clearer and more insightful. The output confirms R's effectiveness in producing high-quality, customizable visualizations.
Github Link	https://github.com/yashsgiri/BDH-Lab/upload
Conclusion	Analyzing data visualization methods in R demonstrates how visual tools enhance understanding of complex datasets. Using libraries like ggplot2 and plotly, R provides flexibility and precision in presenting data, enabling clearer insights and better decision-making through visual analytics.

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 Computer Technology

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	30 October 2025