



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

**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.*

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

<b>Practical Number</b>	01
<b>Course Outcome</b>	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.
<b>Aim</b>	Installation of Apache Hadoop on Linux System.
<b>Problem Definition</b>	Installation of Apache Hadoop on Linux System.
<b>Theory</b> (100 words)	Apache Hadoop is a free software framework for distributed storage and processing of large sets of data using the MapReduce programming model. It works on a cluster of commodity hardware with a high degree of scalability and fault-tolerance. The Hadoop ecosystem primarily consists of HDFS (Hadoop Distributed File System) for distributed storage and YARN (Yet Another Resource Negotiator) for resource management and job scheduling. To install Hadoop on a Linux system you have to install Java and enable SSH and also modify the core configuration files for pseudo-distributed mode or fully distributed mode.
<b>Procedure and Execution</b> (100 Words)	<b>Steps of Implementation</b> <ol style="list-style-type: none"> <li>1. Update the Linux system packages using <code>sudo apt update</code>.</li> <li>2. Install Java Development Kit (JDK) using <code>sudo apt install openjdk-11-jdk</code>.</li> <li>3. Verify Java installation with <code>java -version</code>.</li> <li>4. Create a Hadoop user using <code>sudo adduser hadoop</code>.</li> <li>5. Switch to the Hadoop user with <code>su - hadoop</code>.</li> <li>6. Configure passwordless SSH using <code>ssh-keygen -t rsa -P ""</code> and <code>cat ~/.ssh/id_rsa.pub &gt;&gt; ~/.ssh/authorized_keys</code>.</li> <li>7. Download Hadoop from the official Apache website using <code>wget</code>.</li> <li>8. Extract the Hadoop tar file using <code>tar -xzf hadoop-x.y.z.tar.gz</code>.</li> <li>9. Move Hadoop to <code>/usr/local/hadoop</code> using <code>sudo mv hadoop-x.y.z /usr/local/hadoop</code>.</li> <li>10. Set environment variables in <code>.bashrc</code> for <code>JAVA_HOME</code> and <code>HADOOP_HOME</code>.</li> </ol>



**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.*

	<ol style="list-style-type: none"><li>11. Edit configuration files (core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml) in the Hadoop etc/hadoop directory.</li><li>12. Format the Hadoop namenode using hdfs namenode - format.</li><li>13. Start HDFS and YARN services using start-dfs.sh and start-yarn.sh.</li><li>14. Verify installation using jps command to check running daemons.</li><li>15. Access Hadoop web interface at <a href="http://localhost:9870/">http://localhost:9870/</a>.</li></ol>
	<p>Code:</p> <pre># 1. Update system packages sudo apt update -y  # 2. Install Java sudo apt install openjdk-11-jdk -y  # 3. Verify Java installation java -version  # 4. Create a Hadoop user sudo adduser hadoop  # 5. Switch to Hadoop user su - hadoop  # 6. Configure passwordless SSH ssh-keygen -t rsa -P "" cat ~/.ssh/id_rsa.pub &gt;&gt; ~/.ssh/authorized_keys chmod 600 ~/.ssh/authorized_keys ssh localhost # test SSH  # 7. Download Hadoop (replace version if needed) wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz  # 8. Extract Hadoop tar -xzf hadoop-3.3.6.tar.gz  # 9. Move Hadoop to /usr/local directory sudo mv hadoop-3.3.6 /usr/local/hadoop</pre>



## Department of Artificial Intelligence &amp; 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># 10. Set environment variables nano ~/.bashrc</pre> <pre>Output: theia@theiadocker-srita201326:/home/project\$ hadoop version bash: hadoop: command not found theia@theiadocker-srita201326:/home/project\$ git clone https://github.com/big-data-europe/docker-hadoop.git Cloning into 'docker-hadoop'... remote: Enumerating objects: 539, done. remote: Counting objects: 100% (189/189), done. remote: Compressing objects: 100% (23/23), done. remote: Total 539 (delta 169), reused 166 (delta 166), pack-reused 350 (from 1) Receiving objects: 100% (539/539), 108.00 KiB   13.50 MiB/s, done. Resolving deltas: 100% (251/251), done. theia@theiadocker-srita201326:/home/project\$ cd docker-hadoop theia@theiadocker-srita201326:/home/project/docker-hadoop\$ docker-compose up -d WARN[0000] /home/project/docker-hadoop/docker-compose.yml: the attribute 'version' is obsolete, it will be ignored, please remove it to avoid potential confusion [+] Running 28/28   ✓ historyserver Pulled                                76.8s   ✓ datanode Pulled                                    76.8s   ✓ resourcemanager Pulled                             76.8s   ✓ namenode Pulled                                    76.8s   ✓ nodemanager1 Pulled                                76.7s [+] Running 9/9   ✓ Network docker-hadoop_default                      0.1s   ✓ Volume "docker-hadoop_hadoop_namenode"             0.0s   ✓ Volume "docker-hadoop_hadoop_datanode"             0.0s   ✓ Volume "docker-hadoop_hadoop_historyserver"        0.0s   ✓ Container resourcemanager                          13.9s   ✓ Container datanode                                 14.6s theia@theiadocker-srita201326:/home/project/docker-hadoop\$ docker ps CONTAINER ID   IMAGE                                PORTS STATUS 1ed7bc3d075c   bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8   9864/tcp Up 10 seconds (health: starting) 74e6b9be4c20   bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8 8042/tcp Up 11 seconds (health: starting) 57b2ffa5d876   bde2020/hadoop-resourcemanager:2.0.0-hadoop3.2.1-java8 8088/tcp Up 11 seconds (health: starting) fe222273db8e   bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8   0.0.0.0:9000-&gt;9000/tcp, [::]:9000-&gt;9000/tcp, [::]:9870-&gt;9870/tcp Up 11 seconds (health: starting) 15308eebda32   bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8 8188/tcp Up 11 seconds (health: starting) theia@theiadocker-srita201326:/home/project/docker-hadoop\$ docker exec -it namenode bash root@fe222273db8e:/# theia@theiadocker-srita201326:/home/project/docker-hadoop\$ docker exec -it namenode bash root@fe222273db8e:/# theia@theiadocker-srita201326:/home/project/docker-hadoop\$ docker exec -it namenode bash root@fe222273db8e:/# hadoop version Hadoop 3.2.1 Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r b3cbbb467e22ea829b3808f4b7b01d0e0bf3 Compiled by rohithsharmaks on 2019-09-10T15:56Z Compiled with protoc 2.5.0 From source with checksum 776eaf9ee9c0ff370bcb31888737 This command was run using /opt/hadoop-3.2.1/share/hadoop/common/hadoop-common-3.2.1.jar root@fe222273db8e:/# echo "Hello Hadoop Hello World" &gt; sample.txt root@fe222273db8e:/# hdfs dfs -mkdir /input root@fe222273db8e:/# hdfs dfs -put sample.txt /input 2025-09-01 11:58:14,000 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false remoteHostTrusted = false root@fe222273db8e:/# hadoop jar /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount / ut /output JAR does not exist or is not a normal file: /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar root@fe222273db8e:/# hdfs dfs -cat /output/part-r-00000 cat: /output/part-r-00000: No such file or directory root@fe222273db8e:/# 9-</pre>
Output Analysis	Successful installation displays Hadoop daemons like NameNode, DataNode, ResourceManager, and NodeManager running, and the web UI confirms cluster health and status.
Link of student Github profile where	



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](http://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.

lab assignment has been uploaded	
Conclusion	Installation of Apache Hadoop on Linux System implemented successfully.
Plag Report (Similarity index < 12%)	<div><p><b>Result</b></p><div><p>Plagiarism Checker   Check Grammar   Detector AI   Summarize Text</p><p>Upgrade for More</p><p>Apache Hadoop is a free software framework for distributed storage and processing of large sets of data using the MapReduce programming model. It works on a cluster of commodity hardware with a high degree of scalability and fault-tolerance. The Hadoop ecosystem primarily consists of HDFS (Hadoop Distributed File System) for distributed storage and YARN (Yet Another Resource Negotiator) for resource management and job scheduling. To install Hadoop on a Linux system you have to install Java and enable SSH and also modify the core configuration files for pseudo-distributed mode or fully distributed mode.</p></div><div><p>Unique 100% Exact 0% Partial 0%</p><p>View Plagiarized Sources</p><p>Congratulations</p></div></div>
Date	24 /07 / 25