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**	BDH Lab
Semester	7 AIDS	**Course Code**	22ADS703
Roll No	21	**Name of Student**	**Sanskruti.Paunikar**

Practical Number	**1**
Course Outcome	**CO1:-** 1. 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.
Problem Definition	
Theory (100 words)	Apache Hadoop installation on a Linux system involves setting up Java, configuring Hadoop binaries, and preparing the environment for distributed data processing. First, install Java (Hadoop's dependency). Then, download and extract the Hadoop package to a desired directory. Configure core-site.xml, hdfs-site.xml, mapred-site.xml, and yarn-site.xml files to define storage paths and resource management. Set environment variables for Hadoop and Java in `.bashrc`. Format the Hadoop Name Node and start HDFS and YARN daemons using start-dfs.sh and start-yarn.sh scripts. Finally, verify installation via the web interface or Hadoop command-line tools for successful cluster operation.
Procedure and Execution (100 Words)	Steps of Implementation:- 1. Update System: sudo apt update && sudo apt upgrade -y 2. Install Java: sudo apt install openjdk-11-jdk -y 3. Create Hadoop User: sudo adduser hadoop && su - hadoop 4. Enable SSH: ssh-keygen -t rsa -P "" && cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys 5. Download Hadoop: wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz tar -xzf hadoop-3.3.6.tar.gz && sudo mv hadoop-3.3.6 /usr/local/hadoop

6. Set Environment Variables: Add Hadoop and Java paths in ~/.bashrc.
7. Configure Files: Edit core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml.
8. Format NameNode:
   hdfs namenode -format
9. Start Services:
   start-dfs.sh && start-yarn.sh
10. Verify:
    Run jps and check UIs →

- NameNode: http://localhost:9870
- ResourceManager: http://localhost:8088

Code:

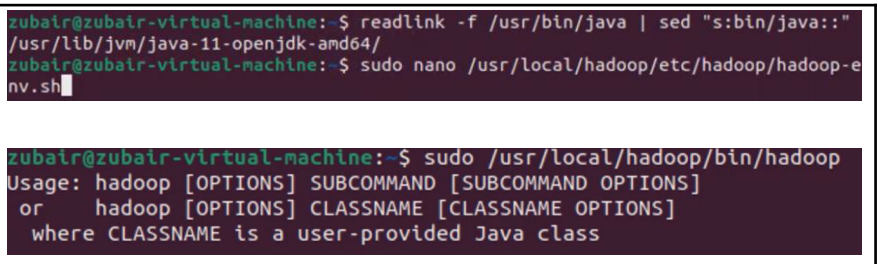
```
zubair@zubair-virtual-machine:~$ java --version
openjdk 11.0.14.1 2022-02-08
OpenJDK Runtime Environment (build 11.0.14.1+1-Ubuntu-0ubuntu1)
OpenJDK 64-Bit Server VM (build 11.0.14.1+1-Ubuntu-0ubuntu1, mixed mode, sharing
)
zubair@zubair-virtual-machine:~$
```


```
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
```

	
Output Analysis	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.
Link of student Github profile where lab assignment has been uploaded	https://github.com/sanskruti-1234/BDH.git
Conclusion	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.
Plag Report (Similarity index < 12%)	
Date	30/10/2025