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	**7**
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 Pig on Linux with Hadoop Integration
Problem Definition	
Theory (100 words)	Apache Pig is a high-level platform for processing large datasets in Hadoop using Pig Latin, a scripting language that simplifies writing MapReduce programs. Pig translates scripts into MapReduce jobs, enabling parallel data processing on HDFS. Installing Pig on Linux involves downloading the Pig binary, configuring environment variables, and integrating it with an existing Hadoop installation. This integration allows Pig to access HDFS data and leverage Hadoop's distributed computing framework. Pig is ideal for ETL operations, data transformations, and iterative data processing, offering simplicity over traditional Java MapReduce while maintaining Hadoop's scalability, fault tolerance, and parallel processing capabilities.
Procedure and Execution (100 Words)	Steps of Implementation:- 1. Ensure Hadoop is installed and running. 2. Download and extract Pig. 3. Move Pig folder to `/usr/local/pig`. 4. Set environment variables (`PIG_HOME`, `HADOOP_HOME`, update `PATH`). 5. Source `~/.bashrc` to apply changes. 6. Start Pig Grunt shell using `pig`. 7. Verify Hadoop integration by running Pig commands to load/store data in HDFS.

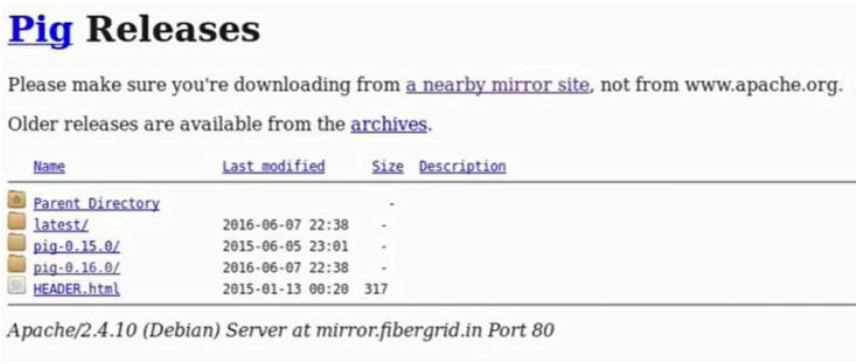
Vision of the Department

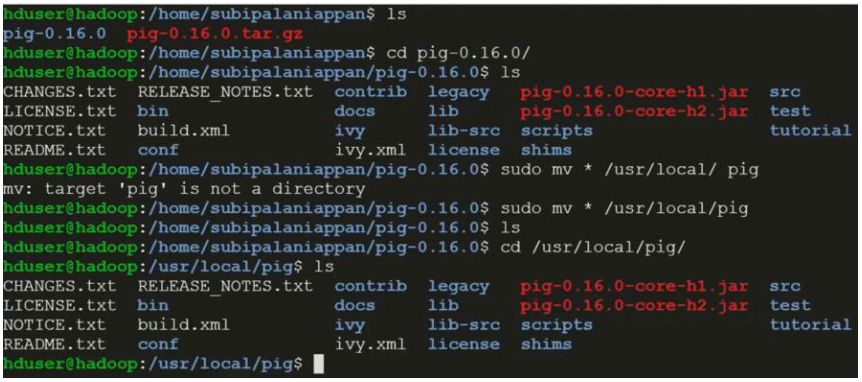
To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

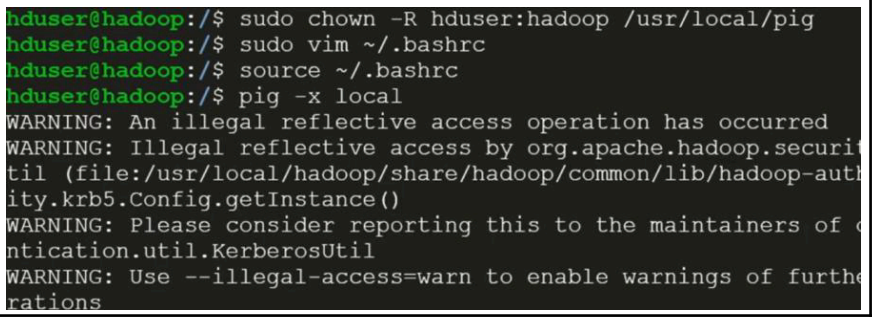
Mission of the Department

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

Code:







Output Analysis	After installation, running Pig in **Grunt shell** allows execution of Pig Latin scripts. Commands like LOAD, FILTER, GROUP, and FOREACH process data and produce intermediate and final results. The output confirms proper Hadoop integration when Pig successfully reads from and writes to HDFS, executes jobs via MapReduce, and displays aggregated or transformed results. Verification using HDFS commands (hdfs dfs -ls /path) ensures that Pig scripts correctly manipulate data in the Hadoop ecosystem. This demonstrates Pig's ability to simplify large-scale data processing while maintaining efficiency and scalability..
Link of student Github profile where lab assignment has been uploaded	https://github.com/sanskruti-1234/BDH.git
Conclusion	Apache Pig was successfully installed and integrated with Hadoop on Linux. Pig provides a high-level, script-based approach for processing large datasets, reducing the complexity of writing MapReduce programs. The experiment confirms Pig's capability to read, transform, and store data on HDFS efficiently, making it a practical tool for ETL, data analysis, and big data processing tasks in Hadoop environments.
Plag Report (Similarity index < 12%)	
Date	30/10/2025