|  |  |  |
| --- | --- | --- |
| Yeshwantrao Chavan College of Engineering | | |
| Department of Artificial Intelligence & Data Science | | |
| Semester 7 | Section: I | Session: Odd 2024-25 |
| Course: Lab: Big Data and Hadoop | | |
|  | | |
| Practical No: 07 | | |
| **Aim**: Installation of Apache Pig on Linux with Hadoop Integration. | | |
| **Theory:**  Apache Pig is a high-level platform for processing large data sets using Hadoop. Pig allows for the creation of programs (called "Pig scripts") that run on Hadoop's MapReduce framework, simplifying the process of writing complex transformations. The language used in Pig is called Pig Latin, and it offers constructs for data manipulation, such as filtering, joining, and grouping, similar to SQL. One key advantage of Pig is its ability to handle both structured and semi-structured data, making it versatile for various data formats like XML, JSON, or text. Pig can be run in two modes: local mode (where the script runs on a single machine) and MapReduce mode (where it runs on a Hadoop cluster).\  **Steps:**   1. **Download Apache PIG Distribution**      1. **Extract Apache Pig** | | |

|  |  |  |
| --- | --- | --- |
| **3) Using ls Command to show files extracted** | | |
| 1. **mv command to move files to /usr/local/pignew**      1. **Setting env variables and path in Enviorment**      1. **Running pig using pig Command** | | |
| Submitted to: Mrs. Kiran Khandare Ma’am | | |
| Name: Himanshu Dhomane | Roll No.: 43 | Reg. No.: 21070297 |