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**	BIG DATA AND HADOOP-LAB
Semester	7 AIDS	**Course Code**	22ADS704
Roll No	03	**Name of Student**	**Debasrita Chattopadhyay**

Practical Number	08
Course Outcome	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.
Aim	Perform Pig Operations: Load & Store Data, Aggregation Operations, Filtering Data and Joining Datasets.
Problem Definition	Perform Pig Operations: Load & Store Data, Aggregation Operations, Filtering Data and Joining Datasets.
Theory (100 words)	Apache Pig is an advanced platform for manipulating big datasets in Hadoop. It offers a high-level scripting language called Pig Latin, which allows users to define transformations (for example, loading, filtering, joining, grouping, and storing data) on the datasets. Pig abstracts away the often complex process of writing MapReduce programs to process big data faster and more easily. Some of the important operations include: Load & Store Data – Load data from HDFS or local file systems to Pig relations, and store the resultant data after processing. Aggregation Operations – Perform operations such as COUNT, SUM, AVG, MAX, and MIN to summarize data. Filtering Data - Selecting specific records that meet a condition by using the FILTER operator. Joining Datasets – Joining two or more datasets using the JOIN operator, where datasets are joined based on common keys.

Procedure and Execution
 	
(100 Words)
	Steps of Implementation: -
	1. **Start Pig in local or Hadoop mode:**
	pig -x local # local mode
	pig -x mapreduce # Hadoop mode
	2. **Load Data:**
	-- Load a CSV file with schema
	students = LOAD 'students.csv' USING PigStorage(',') AS (id:int, name:chararray, age:int, marks:int);
	courses = LOAD 'courses.csv' USING PigStorage(',') AS (id:int, course:chararray, student_id:int);
	3. **Display Data:**
	DUMP students;
	4. **Aggregation Operations:**
	-- Group data
	grouped_data = GROUP students ALL;
	
	-- Count total students
	total_students = FOREACH grouped_data GENERATE COUNT(students);
	
	-- Average marks
	avg_marks = FOREACH grouped_data GENERATE AVG(students.marks);
	5. **Filtering Data:**
	-- Filter students with marks > 70
	high_scorers = FILTER students BY marks > 70;
	6. **Joining Datasets:**
	-- Join students with courses
	student_courses = JOIN students BY id, courses BY student_id;
	7. **Store Data:**
	STORE high_scorers INTO 'high_scorers_output' USING PigStorage(',');
	STORE student_courses INTO 'student_courses_output' USING PigStorage(',');
	
	Code:
	-- Load students' data
	students = LOAD 'students.csv' USING PigStorage(',') AS (id:int, name:chararray, age:int, marks:int);
	
	-- Load courses data

```
courses = LOAD 'courses.csv' USING PigStorage(',') AS
(course_id:int, course_name:chararray, student_id:int);

-- Display loaded data
DUMP students;
DUMP courses;

-- Aggregation Operations
grouped_students = GROUP students ALL;

-- Count total students
total_students = FOREACH grouped_students GENERATE
COUNT(students);
DUMP total_students;

-- Average marks
avg_marks = FOREACH grouped_students GENERATE
AVG(students.marks);
DUMP avg_marks;

-- Filtering: Students with marks > 70
high_scorers = FILTER students BY marks > 70;
DUMP high_scorers;

-- Joining students with courses
student_courses = JOIN students BY id, courses BY student_id;
DUMP student_courses;

-- Store the results
STORE high_scorers INTO 'high_scorers_output' USING
PigStorage(',');
STORE student_courses INTO 'student_courses_output' USING
PigStorage(',');
```

Output:

```
grunt> courses = LOAD 'courses.csv' USING PigStorage(',') AS (course_id:int, course_name:chararray, student_id:int);
```

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.

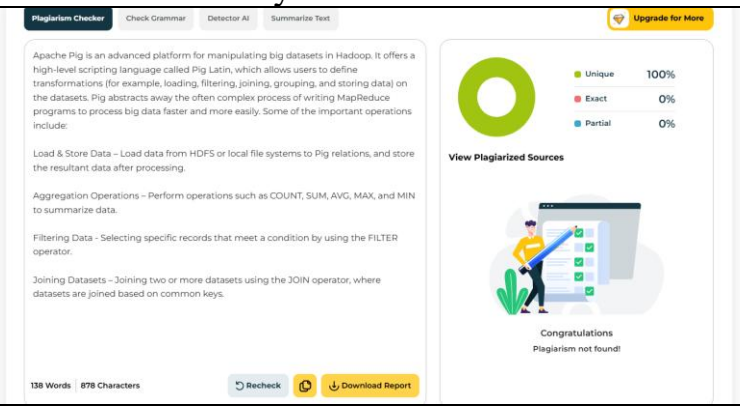
Vision of the Department

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.

```
(101,Math,1)
(102,Physics,2)
(103,Chemistry,3)
(104,English,4)
(105,Biology,5)

grunt> grouped_students = GROUP students ALL;
grunt> total_students = FOREACH grouped_students GENERATE COUNT(students);
grunt> DUMP total_students;

(5)

grunt> avg_marks = FOREACH grouped_students GENERATE AVG(students.marks);
grunt> DUMP avg_marks;
(74.6)

grunt> high_scorers = FILTER students BY marks > 70;
grunt> DUMP high_scorers;

(1,John,20,85)
(2,Alice,22,78)
(4,Emma,21,90)

grunt> student_courses = JOIN students BY id, courses BY student_id;
grunt> DUMP student_courses;
(1,John,20,85,101,Math,1)
(2,Alice,22,78,102,Physics,2)
(3,Bob,19,65,103,Chemistry,3)
(4,Emma,21,90,104,English,4)
(5,Mike,23,55,105,Biology,5)

uccess!
grunt> STORE high_scorers INTO 'high_scorers_output' USING PigStorage(',');
grunt> STORE student_courses INTO 'student_courses_output' USING PigStorage(',');
```

Output Analysis	1. **Load Data:** Successfully loaded students.csv and courses.csv into Pig relations. 2. **Display Data:** Printed tuples showing student and course records. 3. **Count:** Total number of students = **5**. 4. **Average:** Average marks of students = **74.6**. 5. **Filter:** Students with marks > 70 → **John, Alice, Emma**. 6. **Join:** Combined student and course data based on student ID. 7. **Store:** Results saved in high_scorers_output and student_courses_output folders.
Link of student Github profile where lab assignment has been uploaded	
Conclusion	Pig Operations: Load & Store Data, Aggregation Operations, Filtering Data and Joining Datasets implemented successfully.
Plag Report (Similarity index < 12%)	
Date	**18 /9 / 25**