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 | **8** |
|---|---|
| Course Outcome | **CO1:-** 1. Understand big data analytics and its business applications.<br>**CO2:-** Analyze the HADOOP and Map Reduce technologies associated with big data analytics.<br>**CO3:-** Apply Big Data analytics Using Pig and Hive. |
| Aim | Perform Pig Operations: Load & Store Data, Aggregation Operations, Filtering Data and Joining Datasets. |
| Problem Definition | |
| Theory<br>(100 words) | Apache Pig provides a high-level platform to process large datasets on Hadoop using Pig Latin scripts. Key operations include Load and Store for reading from and writing to HDFS, Aggregation functions like GROUP, COUNT, SUM, and AVG for summarizing data, Filtering with FILTER to select specific records, and Joining datasets to combine related tables. Pig simplifies writing complex MapReduce jobs by allowing declarative scripting. It processes data in a pipeline, generating intermediate results that Hadoop executes in parallel. Pig is ideal for ETL tasks, large-scale data transformation, and analytics on distributed datasets. |
| Procedure and Execution<br><br>(100 Words) | Steps of Implementation:-<br>1. Start Pig Grunt Shell: Launch Pig in interactive or script mode.<br>2. Load Data: Use `LOAD` to read datasets from HDFS into Pig relations.<br>3.Perform Aggregation: Apply `GROUP`, `COUNT`, `SUM`, `AVG` functions to summarize data.<br>4. Filter Data: Use `FILTER` to select records meeting specific conditions.<br>5. Join Datasets: Use `JOIN` to combine multiple datasets based on keys.<br>6. Store Results: Use `STORE` to write processed data back to HDFS.<br>7. Verify Outputs: Use `DUMP` or check HDFS to ensure operations executed correctly. |

**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:

```
grunt> a = load 'retail_stage1.order_items' using org.apache.hive.hcatalog.pig.HCatLoader();
```

```
grunt> b = group a by order_item_order_id
```

```
grunt> c = foreach b generate group, AVG(a.order_item_product_price)
```

```
grunt> dump c;
```

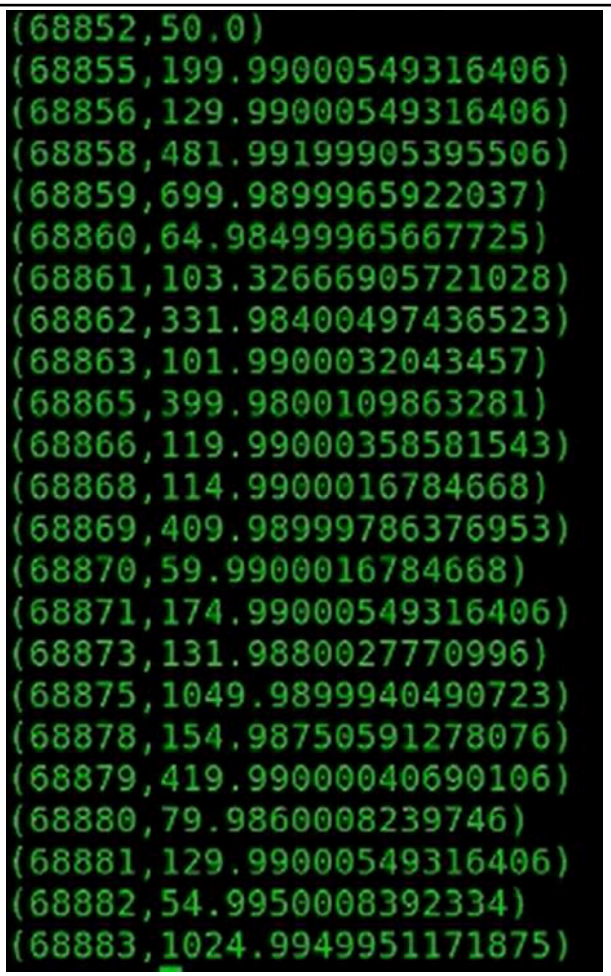
```
hive> select order_item_order_id, AVG(order_item_product_price)
    > from order_items
    > group by order_item_order_id;
Query ID = cloudera_20160913214646_7a4abeba-4526-4028-8da3-97c5f5e10575
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
```

```
grunt> a = '/user/cloudera/department' using PigStorage(',');
2016-08-29 21:36:27,468 [main] ERROR org.apache.pig.tools.grunt.Grunt
rror, unexpected symbol at or near 'a'
Details at logfile: /home/cloudera/pig_1472530648111.log
```

```
grunt> a = load '/user/cloudera/department' using PigStorage(',');
2016-08-29 21:36:39,833 [main] WARN  org.apache.pig.PigServer - Enc
s).
```

```
grunt> c = filter a by $0 > 4;
```

Output:

| | |
|---|---|
| |  <br>  |
| Output Analysis | After executing Pig scripts, outputs are stored in HDFS or displayed in the Grunt shell. **Load/Store** operations confirm successful data reading and writing. Aggregation results show accurate summaries, filters return only selected records, and joins correctly combine datasets based on keys. Output verification using `DUMP` and HDFS commands ensures scripts execute as expected. These results demonstrate Pig's ability to perform |

| | |
|---|---|
| | complex data transformations efficiently, leveraging Hadoop's parallel processing for large datasets. |
| Link of student Github profile where lab assignment has been uploaded | https://github.com/sanskruti-1234/BDH.git |
| Conclusion | Pig operations for loading, storing, aggregating, filtering, and joining datasets were successfully performed. The experiment highlights Pig's ease of use, efficiency in data manipulation, and seamless Hadoop integration, proving it to be a powerful tool for large-scale data processing and analytics. |
| Plag Report (Similarity index < 12%) |  |
| Date | 30/10/2025 |