

**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	<b>Preparation</b>	<b>P: Preparation</b>	<b>Pep-CL abbreviation pronounce as Pep-si-IL easy to recall</b>
PEO2	<b>Core Competence</b>	<b>E: Environment (Learning Environment)</b>	
PEO3	<b>Breadth</b>	<b>P: Professionalism</b>	
PEO4	<b>Professionalism</b>	<b>C: Core Competence</b>	
PEO5	<b>Learning Environment</b>	<b>L: Breadth (Learning in diverse areas)</b>	

**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)



## Department of Artificial Intelligence &amp; 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	06
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 Case Study: Analyzing Olympic Data Set Using Hive.
Problem Definition	Perform Case Study: Analyzing Olympic Data Set Using Hive.
Theory (100 words)	<p>Case Study: Analyzing Olympic Dataset Using Hive (Brief Theory)</p> <p>The Olympic dataset is analyzed using the Hive tool which is a data warehousing solution based on Hadoop, and which uses HiveQL for querying large-scale datasets, as opposed to any relation database model. The dataset (athletes, countries, sports, medals, etc.) is loaded into HDFS where it is stored in tables accessible by Hive. Queries to answer analytical questions such as the total number of medals by country, the most medal-winning athletes, and trends in participation and participation by country are written using SQL-like commands. Hive allows for static distributed data processing and analytics, but is not a traditional "no-code" visualization tool; rather, one can conduct large-scale analysis and develop visualizations using BI tools, thus gaining better insights into Olympic performance and Olympic trends over the years.</p> <p>Hive performs some common yet tedious work of partitioning and parallel execution for the user, which is one of the primary benefits of using Hive for big data analytics. Hive also has distinct analytical capabilities, such as the ability to GROUP BY, ORDER BY, JOIN, and so on. The end user can use Hive's output. to either view queries in a BI</p>

**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.*

	<p>tool or export them to work in their own analytical environment.</p> <p>Overall, this case study demonstrates how powerful the value of distributed data processing can be, and how data warehousing concepts can provide some simplification, and yet still be able to support a complex analytical workload, on large-scale datasets for a study of Olympic performance and global participation over the years.</p>
<p>Procedure and Execution</p> <p>(100 Words)</p>	<p>Steps of Implementation: -</p> <p><b>Start Hadoop and Hive Services</b></p> <ul style="list-style-type: none"><li>• start-dfs.sh</li><li>• start-yarn.sh</li><li>• hive</li></ul> <p><b>Create a Hive Database</b></p> <ul style="list-style-type: none"><li>• CREATE DATABASE olympicdb;</li><li>• USE olympicdb;</li></ul> <p><b>Create a Hive Table</b></p> <ul style="list-style-type: none"><li>• Define columns (athlete, country, year, sport, gold, silver, bronze, total)</li><li>• Use ROW FORMAT DELIMITED and FIELDS TERMINATED BY ','</li></ul> <p><b>Upload Dataset to HDFS</b></p> <ul style="list-style-type: none"><li>• hdfs dfs -mkdir /user/hive/olympics</li><li>• hdfs dfs -put olympics.csv /user/hive/olympics/</li></ul> <p><b>Load Data into Hive Table</b></p> <ul style="list-style-type: none"><li>• LOAD DATA INPATH '/user/hive/olympics/olympics.csv' INTO TABLE olympic_data;</li></ul> <p><b>Perform Analysis Queries</b></p> <ul style="list-style-type: none"><li>• Total medals by country</li><li>• Top athletes</li><li>• Medals by year</li><li>• Country with most gold medals</li></ul> <p><b>View Results</b></p> <ul style="list-style-type: none"><li>• SELECT * FROM olympic_data LIMIT 10;</li></ul> <p><b>Exit Hive and Stop Hadoop Services</b></p> <ul style="list-style-type: none"><li>• EXIT;</li><li>• stop-yarn.sh</li><li>• stop-dfs.sh</li></ul>



**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.*

Code:

```
1 sudo apt update
2 sudo apt install openjdk-11-jdk -y
3 CREATE DATABASE olympicdb;
4 USE olympicdb;
5 java -version
6 sudo apt update
7 sudo apt install openjdk-8-jdk -y
8 sudo update-alternatives --config java
9 java -version
10 cd /usr/local
11 sudo wget
https://downloads.apache.org/hadoop/common/hadoop-
3.3.6/hadoop-3.3.6.tar.gz
12 sudo tar -xzvf hadoop-3.3.6.tar.gz
13 sudo mv hadoop-3.3.6 hadoop
14 sudo nano ~/.bashrc
15 source ~/.bashrc
16 hadoop version
17 cd /usr/local
18 sudo wget https://downloads.apache.org/hive/hive-
3.1.3/apache-hive-3.1.3-bin.tar.gz
19 sudo tar -xzvf apache-hive-3.1.3-bin.tar.gz
20 sudo mv apache-hive-3.1.3-bin hive
21 cd /usr/local
22 sudo wget https://archive.apache.org/dist/hive/hive-
3.1.3/apache-hive-3.1.3-bin.tar.gz
23 sudo tar -xzvf apache-hive-3.1.3-bin.tar.gz
24 sudo mv apache-hive-3.1.3-bin hive
25 nano ~/.bashrc
26 source ~/.bashrc
27 schematool -initSchema -dbType derby
28 mkdir -p ~/metastore_db
29 export HIVE_METASTORE_DB_DIR=~/metastore_db
30 schematool -initSchema -dbType derby
31 hive
32 hiv
33 hive
34 export HIVE_HOME=/usr/local/hive
35 mkdir -p ~/metastore_db
36 export DERBY_HOME=~/metastore_db
37 hive
38 history
```



## Department of Artificial Intelligence &amp; 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.

	<p>Output:</p> <pre>java HotSpot(TM) 64-Bit Server VM (build 21.0.8+11-LTS-250, mixed mode, : theia@theiadocker-u22070346:/home/project\$ sudo apt update sudo apt install openjdk-8-jdk -y  theia@theiadocker-u22070346:/home/project\$ sudo update-alternatives --config java  theia@theiadocker-u22070346:/home/project\$ java -version openjdk version "1.8.0_462" OpenJDK Runtime Environment (build 1.8.0_462-8u462-ga~us1-0ubuntu2~22.04.2-b08) OpenJDK 64-Bit Server VM (build 25.462-b08, mixed mode) theia@theiadocker-u22070346:/home/project\$  theia@theiadocker-u22070346:/home/project\$ cd /usr/local sudo wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz sudo tar -xzf hadoop-3.3.6.tar.gz sudo mv hadoop-3.3.6 hadoop  theia@theiadocker-u22070346:/usr/local\$ source ~/.bashrc  theia@theiadocker-u22070346:/usr/local\$ hadoop version Hadoop 3.3.6 Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c Compiled by ubuntu on 2023-06-18T08:22Z Compiled on platform linux-x86_64 Compiled with protoc 3.7.1 From source with checksum 5652179ad55f76cb287d9c633bb53bbd This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-3.3.6.jar theia@theiadocker-u22070346:/usr/local\$  theia@theiadocker-u22070346:/usr/local\$ cd /usr/local sudo wget https://downloads.apache.org/hive/hive-3.1.3/apache-hive-3.1.3-bin.tar.gz sudo tar -xzf apache-hive-3.1.3-bin.tar.gz sudo mv apache-hive-3.1.3-bin hive  theia@theiadocker-u22070346:/usr/local\$ nano ~/.bashrc  theia@theiadocker-u22070346:/usr/local\$ source ~/.bashrc  theia@theiadocker-u22070346:/usr/local\$ schematool -initSchema -dbType derby  hive&gt; SHOW DATABASES;</pre>
Output Analysis	<p><b>Total Medals by Country:</b> Shows leading countries like USA, Russia, China.</p> <p><b>Top Athletes:</b> Identifies athletes with highest total medals.</p> <p><b>Medals by Year:</b> Highlights trends and exceptional Olympic years.</p> <p><b>Country with Most Gold Medals:</b> Shows top-performing nation in gold medals.</p>
Link of student Github profile where	



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](http://www.ycce.edu)



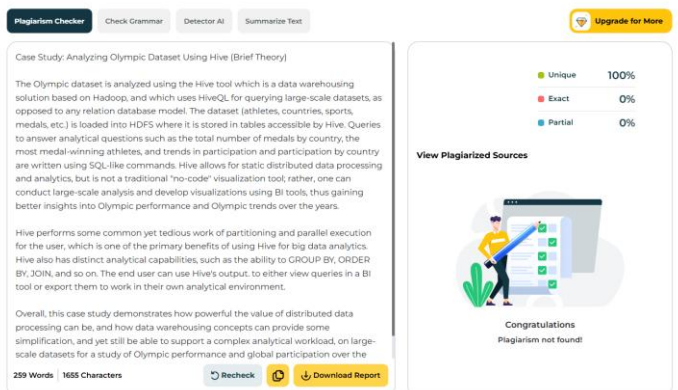
### Department of Artificial Intelligence & Data Science

#### 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.

lab assignment has been uploaded	
Conclusion	Case Study: Analyzing Olympic Data Set Using Hive implemented successfully.
Plag Report (Similarity index < 12%)	
Date	28 / 8 / 25