

**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-IL 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)



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 | Sanskruir.Paunikar |

| | |
|--|---|
| Practical Number | 6 |
| 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 | Perform Case Study: Analyzing Olympic Data Set Using Hive. |
| Problem Definition | |
| Theory (100 words) | Apache Hive enables querying and analyzing large datasets stored in Hadoop's HDFS using HiveQL, a SQL-like language. In a case study with an Olympic dataset, Hive can be used to extract insights such as medal tallies, country-wise performance, athlete participation, and trends over time. By creating structured tables from raw CSV or text files, Hive allows filtering, aggregation, sorting, and joining data efficiently using MapReduce or Tez. This approach leverages distributed processing to handle large-scale datasets, making it suitable for historical Olympic records analysis, performance tracking, and generating meaningful statistics for sports analytics.. |
| Procedure and Execution (100 Words) | Steps of Implementation:- 1. Start Hive shell: `hive` 2. Create and use a database: `CREATE DATABASE olympics_db; USE olympics_db;` 3. Create a table for Olympic data with appropriate schema. 4. Load dataset into the table: `LOAD DATA LOCAL INPATH 'olympic_data.csv' INTO TABLE medals;` 5. Run Hive queries for analysis (e.g., medal counts, country-wise or year-wise trends). 6. Verify tables and schema: `SHOW TABLES;`, `DESCRIBE medals;` 7. View and interpret query results. |



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:

```
hive> create table num_of_sports_by_olympic as select city, year, count(distinct sport) as no_of_sports from summer_olympics group by city, year;
Query ID = hduser_20190427163204_8354edce-e9ed-49f2-b47f-f462b8a64384
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>
Starting Job = job_1556275802324_0009, Tracking URL = http://shikhar-VirtualBox:8088/proxy/application_1556275802324_0009/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1556275802324_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-04-27 16:32:34,693 Stage-1 map = 0%, reduce = 0%
2019-04-27 16:32:50,593 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.88 sec
2019-04-27 16:33:06,960 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.93 sec
MapReduce Total cumulative CPU time: 8 seconds 930 msec
Ended Job = job_1556275802324_0009
Moving data to directory hdfs://localhost:54310/user/hive/warehouse/pda_project.db/num_of_sports_by_olympic
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.93 sec HDFS Read: 11578330 HDFS Write: 593 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 930 msec
OK
Time taken: 66.2 seconds
hive> select * from num_of_sports_by_olympic limit 5;
OK
Athina 1896 2
Paris 1900 6
St. Louis 1904 8
Athina 1906 8
London 1908 11
Time taken: 0.786 seconds, Fetched: 5 row(s)
```

```
hive> create table num_of_players_by_olympic as select city, year, count(distinct name) as no_of_players from summer_olympics group by city, year;
Query ID = hduser_20190427163626_167a525e-b952-4317-8870-e894568dc2c0
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>
Starting Job = job_1556275802324_0010, Tracking URL = http://shikhar-VirtualBox:8088/proxy/application_1556275802324_0010/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1556275802324_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-04-27 16:36:45,608 Stage-1 map = 0%, reduce = 0%
2019-04-27 16:37:04,973 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 0.55 sec
2019-04-27 16:37:21,900 Stage-1 map = 100%, reduce = 89%, Cumulative CPU 13.8 sec
2019-04-27 16:37:22,976 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 13.8 sec
MapReduce Total cumulative CPU time: 14 seconds 800 msec
Ended Job = job_1556275802324_0010
Moving data to directory hdfs://localhost:54310/user/hive/warehouse/pda_project.db/num_of_players_by_olympic
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 14.8 sec HDFS Read: 11578421 HDFS Write: 640 SUCCESS
Total MapReduce CPU Time Spent: 14 seconds 800 msec
OK
Time taken: 58.213 seconds
hive> select * from num_of_players_by_olympic limit 5;
OK
Athina 1896 11
Paris 1900 38
St. Louis 1904 40
Athina 1906 51
London 1908 132
Time taken: 0.666 seconds, Fetched: 5 row(s)
```

```
hive> create view sports_and_players_by_olympic as select a.city, a.year, a.no_of_sports, b.no_of_players from num_of_sports_by_olympic a inner join num_of_players_by_olympic b on a.city=b.city and a.year=b.year;
Time taken: 0.093 seconds
hive> show views;
OK
sports_and_players_by_olympic
Time taken: 0.121 seconds, Fetched: 1 row(s)
hive> select * from sports_and_players_by_olympic limit 5;
Query ID = hduser_20190427164035_6d2abcc5-acc5-4977-8718-bef5a03254f0
Total Jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1556275802324_0011, Tracking URL = http://shikhar-VirtualBox:8088/proxy/application_1556275802324_0011/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1556275802324_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2019-04-27 16:40:36,204 Stage-1 map = 0%, reduce = 0%
MapReduce Total cumulative CPU time: 4 seconds 328 msec
Ended Job = job_1556275802324_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 4.32 sec HDFS Read: 10999 HDFS Write: 236 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 328 msec
OK
Time taken: 73.784 seconds, Fetched: 0 row(s)
```

```
hive> create view last_olympics as select * from summer_olympics where year in ('2016', '2012', '2008', '2004');
Time taken: 0.569 seconds
hive> select * from last_olympics limit 10;
OK
13877 Abdelilah Falli M 22.0 163.0 63.0 Morocco MOR 2008 Summer 2008 Summer Beijing Athletics Athletics Men's 10 800 metres
13875 Katelyn Falgout M 19.0 168.0 63.0 United States USA 2008 Summer 2008 Summer Beijing Hockey Hockey Women's Hockey No Medal
13873 Dana Falatic F 27.0 178.0 74.0 Australia AUS 2004 Summer 2004 Summer Athina Rowing Rowing Women's Quadruple Sculls Bronze
13870 Omar Falson M 29.0 172.0 65.0 Hungary HUN 2004 Summer 2004 Summer Rio de Janeiro Triathlon Triathlon Men's Olympic Distance No Medal
13865 Nara Lerena Falco Artega F 22.0 168.0 55.0 Mexico MEX 2004 Summer 2004 Summer Athina Synchronized Swimming Synchronized Swimming Women's Duet No Medal
13858 Yorgosch Falco Florentino M 24.0 181.0 77.0 Brazil BRA 2002 Summer 2002 Summer London Boxing Boxing Men's Lightweight Flyweight Bronze
13846 Mariella Falca F 18.0 167.0 58.0 Italy ITA 2004 Summer 2004 Summer Athina Rhythmic Gymnastics Rhythmic Gymnastics Women's Group Silver
13844 Natalia Falencia Gillo F 26.0 174.0 69.0 Russia RUS 2004 Summer 2004 Summer Athina Judo Judo Women's Heavyweight No Medal
13840 Lucia Falasca F 23.0 171.0 64.0 Argentina ARG 2016 Summer 2016 Summer Rio de Janeiro Sailing Sailing Women's One Person Dinghy No Medal
13819 Omar Fakhri M 28.0 165.0 53.0 Egypt EGY 2008 Summer 2008 Summer Beijing Modern Pentathlon Modern Pentathlon Women's Individual No Medal
Time taken: 0.631 seconds, Fetched: 10 row(s)
```

Output Analysis

After loading the Olympic dataset into Hive tables, queries such as SELECT country, COUNT(*) FROM medals GROUP BY country; or SELECT year, SUM(gold) FROM medals GROUP BY year; produce aggregated results. The output displays medal counts per country, athlete participation trends, and yearly performance metrics. Verification with SHOW TABLES; and DESCRIBE table name; confirms correct schema and data



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.

| | |
|---|---|
| | loading. The results demonstrate Hive's capability to handle large datasets, perform complex queries, and provide actionable insights efficiently. |
| Link of student Github profile where lab assignment has been uploaded | https://github.com/sanskriti-1234/BDH.git |
| Conclusion | Using Hive, the Olympic dataset was successfully analyzed to extract meaningful trends and statistics. Hive simplified large-scale data querying with SQL-like commands while leveraging Hadoop's distributed processing for scalability and efficiency. This case study highlights Hive's effectiveness for big data analytics in sports and other domains requiring structured data analysis. |
| Plag Report (Similarity index < 12%) | |
| Date | 30/10/2025 |