



### Department of Computer Technology

#### 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

<b>Vision:</b> To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	<b>Mission:</b> To acquire skills through coursework, projects, and internships, while actively engaging in research and collaboration with peers to innovate and apply AI solutions.
---	---

**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.

Prerana Bijekar      30 October 2025

**Name and Signature of Student and Date**

(Signature and Date in Handwritten)



## Department of Computer Technology

## 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.*

<b>Session</b>	2025-26 (ODD)	<b>Course Name</b>	BDH Lab
<b>Semester</b>	7	<b>Course Code</b>	22ADS704
<b>Roll No</b>	11	<b>Name of Student</b>	Prerana Bijekar

<b>Practical Number</b>	6
<b>Course Outcome</b>	<b>CO1:</b> Understand big data analytics and its business applications. <b>CO2:</b> Analyze the HADOOP and Map Reduce technologies associated with big data analytics. <b>CO3:</b> Apply Big Data Analytics Using Pig and Hive.
<b>Aim</b>	Perform Case Study: Analyzing Olympic Data Set Using Hive.
<b>Theory (100 words)</b>	Apache Hive enables large-scale data analysis using SQL-like queries on data stored in Hadoop. In this case study, the Olympic dataset containing details like athletes, countries, sports, medals, and years is analyzed using Hive. By loading the dataset into Hive tables, users can execute queries to extract insights such as total medals by country, top athletes, or performance trends over the years. Hive translates these queries into MapReduce or Tez/Spark jobs, enabling efficient distributed processing of massive datasets and simplifying analytics through its structured query interface.
<b>Procedure and Execution (100 Words)</b>	Steps of implementation: <ul style="list-style-type: none"><li>• Start Hadoop and Hive services.</li><li>• Create a database, e.g., CREATE DATABASE olympics;</li><li>• Create an external table with appropriate schema.</li><li>• Load the dataset using LOAD DATA INPATH '/path/olympic.csv' INTO TABLE olympic;</li><li>• Run queries like:<ul style="list-style-type: none"><li>○ SELECT country, COUNT(medal) FROM olympic GROUP BY country;</li><li>○ SELECT athlete, COUNT(*) FROM olympic GROUP BY athlete ORDER BY COUNT(*) DESC;</li></ul></li><li>• Save or export query results if needed.</li></ul>

## Department of Computer Technology

### 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-4972-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 16:32:04,092 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 16:33:06,966 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: 0.706 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-e5894568dc26
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 16:37:04,972 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.55 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 14.8 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 30
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=a.city and a.year=b.year;
Time taken: 1.563 seconds
hive> show views;
OK
sports_and_players_by_olympic
Time taken: 8.121 seconds, Fetched: 1 row(s)
hive> select * from sports_and_players_by_olympic limit 5;
Query ID = hduser_20190427164616_8d2a8cd5-acc0-4977-8718-8df503254f0
Total Jobs = 1
Execution completed successfully
MapReduce task succeeded
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_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 16:47:09,924 Stage-1 map = 0%, reduce = 0%
2019-04-27 16:47:07,448 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.32 sec
MapReduce Total cumulative CPU time: 4 seconds 320 msec
Ended Job = job_1556275802324_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 0 Cumulative CPU: 4.32 sec HDFS Read: 18999 HDFS Write: 236 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 320 msec
OK
Athina 1896 2 11
Paris 1900 6 30
St. Louis 1904 8 40
Athina 1906 8 51
London 1908 11 132
Time taken: 72.784 seconds, Fetched: 5 row(s)
```

```
hive> create view Last_4_Olympics as select * from summer_olympics where year in ('2010', '2012', '2008', '2004');
OK
Time taken: 0.508 seconds
hive> select * from Last_4_Olympics limit 10;
OK
13877 Abdelilah Falli M 32.0 163.0 63.0 Morocco MGR 2008 Summer 2008 Summer Beijing Athletics Athletics Men's 10 000 metres
13875 Katerlny Falgowski F 19.0 166.0 63.0 United States USA 2008 Summer 2008 Summer Beijing Hockey Hockey Women's Hockey No Medal
13873 Dena Jalevic F 27.0 170.0 74.0 Australia AUS 2008 Summer 2008 Summer Athens Rowing Rowing Women's Quadruple Scull Bronze
13870 Gbor Faldut A 28.0 172.0 62.0 Hungary HUN 2016 Summer 2016 Summer Rio de Janeiro Triathlon Triathlon Men's Olympic Distance No Medal
13865 Hera Lorenz Falcu Ortega M 24.0 181.0 77.0 Mexico MEX 2004 Summer 2004 Summer Athens Rhythmic Gymnastics Synchronized Swimming Women's Duet No Medal
13850 Vanaguchi Falco Florentino M 24.0 181.0 77.0 Brazil BRA 2012 Summer 2012 Summer London Boxing Boxing Men's Light-heavyweight Bronze
13846 Mariella Falco M 18.0 187.0 64.0 Italy ITA 2004 Summer 2004 Summer Athens Rhythmic Gymnastics Rhythmic Gymnastics Women's Group Silver
13844 Natalia Falavigna Silva F 20.0 178.0 69.0 Brazil BRA 2004 Summer 2004 Summer Athens Taekwondo Taekwondo Women's Heavyweight No Medal
13840 Lucia Falavigna F 23.0 172.0 64.0 Argentina ARG 2016 Summer 2016 Summer Rio de Janeiro Judo Judo Women's One Person Single No Medal
13835 Omnia Abdel Khalek Fahmy F 28.0 165.0 53.0 Egypt EGY 2008 Summer 2008 Summer Beijing Modern Pentathlon Modern Pentathlon Women's Individual No Medal
Time taken: 8.021 seconds, Fetched: 10 row(s)
```

### Output Analysis

Hive queries produce analytical summaries such as the number of medals won by each country, the most successful athletes, and trends across different Olympic years. The results confirm Hive's ability to efficiently process and analyze large structured datasets stored in HDFS.



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 Computer Technology

#### 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.

Github Link	<a href="https://github.com/Prerana-Bijekar/BDH">https://github.com/Prerana-Bijekar/BDH</a>
Conclusion	Analyzing the Olympic dataset using Hive demonstrates how big data analytics can be simplified through HiveQL. The case study highlights Hive's capability to manage and query large datasets efficiently, providing meaningful insights from complex data with minimal coding effort.
Plag Report (Similarity index < 12%)	<p>SmallSEOTools</p> <p>Plagiarism Scan Report By SmallSEOTools</p> <p>Report Generated on: Oct 31, 2024</p> <p>8.4% Plagiarized Content</p> <p>5.3% Exact Plagiarized</p> <p>3.1% Partial Plagiarized</p> <p>91.6% Unique Content</p> <p>Total Words: 388 Total Characters: 453 Plagiarized Sentences: 13 Unique Sentences: 113 (91.6%)</p>
Date	30 October 2025