



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

Vision: To harness the power of artificial intelligence and data science to solve real-world problems and enhance human potential.	Mission: 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	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.

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.

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

Practical Number	6
Course Outcome	CO1: 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.
Theory (100 words)	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.
Procedure and Execution (100 Words)	Steps of implementation: <ul style="list-style-type: none">• Start Hadoop and Hive services.• Create a database, e.g., CREATE DATABASE olympics;• Create an external table with appropriate schema.• Load the dataset using LOAD DATA INPATH '/path/olympic.csv' INTO TABLE olympic;• Run queries like:<ul style="list-style-type: none">◦ SELECT country, COUNT(medal) FROM olympic GROUP BY country;◦ SELECT athlete, COUNT(*) FROM olympic GROUP BY athlete ORDER BY COUNT(*) DESC;• Save or export query results if needed.



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: <pre> 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_20190427162824_8354edce-e9ed-49f2-b47f-f462bba64384 Total Jobs = 1 Launching Job 1 out of 1 Number of reducers 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 16:33:06.968 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 +-----+ city no_of_sports +-----+ Athina 1896 Paris 1900 St. Louis 1904 Athina 1906 London 1908 +-----+ 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_20190427163026_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 16:45:45.608 Stage-1 map = 0%, reduce = 0% 2019-04-27 16:37:04.973 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.55 sec 2019-04-27 16:37:21.908 Stage-1 map = 100%, reduce = 89%, Cumulative CPU 13.8 sec 2019-04-27 16:37:22.974 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 +-----+ city no_of_players +-----+ Athina 1896 Paris 1900 St. Louis 1904 Athina 1906 London 1908 +-----+ 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; OK Time taken: 1.933 seconds hive> show views; +-----+ view +-----+ sports_and_players_by_olympic +-----+ Time taken: 0 seconds hive> select * from sports_and_players_by_olympic limit 5; Query ID = hduser_20190427164030_0c2a0cd5-accc-4977-8718-8df5a05254fb Total Jobs = 1 Execution completed successfully MapReduce Jobs Launched: Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 0.0 sec Launching Job 1 out of 1 Number of reducers not specified. 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 16:47:06.924 Stage-1 map = 0%, reduce = 0% 2019-04-27 16:47:07.924 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.32 sec MapReduce Total cumulative CPU time: 4 seconds 208 msec Ended Job = job_1556275802324_0011 MapReduce Jobs Launched: Stage Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.32 sec HDFS Read: 10999 HDFS Write: 236 SUCCESS Total MapReduce CPU Time Spent: 4 seconds 208 msec +-----+ city year no_of_sports no_of_players +-----+ Athina 1896 2 11 Paris 1900 3 39 St. Louis 1904 8 40 Athina 1906 8 51 London 1908 132 132 +-----+ Time taken: 73.784 seconds, Fetched: 5 row(s) hive> create view Last_4_Olympics as select * from summer_olympics where year in ('2016', '2012', '2008', '2004'); OK Time taken: 0.609 seconds hive> select * from Last_4_Olympics limit 10; Query ID = hduser_20190427164130_0c2a0cd5-accc-4977-8718-8df5a05254fb +-----+ id athlete name gender country sport year medal value +-----+ 13877 Abdellah Falli M 32.0 163.0 63.0 Morocco MOR 2008 Summer 2008 Summer Beijing Athletics Athletics Men's 10000 metres 13875 Katelyn Falgowski F 19.0 168.0 63.0 United States USA 2008 Summer 2008 Summer Beijing Hockey Hockey Women's Hockey No Medal 13876 Ondrej Palka M 37.0 178.0 63.0 Czechia CZE 2008 Summer 2008 Summer Beijing Rowing Rowing Men's Double scull 13878 Gabor Faludi M 28.0 172.0 65.0 Hungary HUN 2008 Summer 2008 Summer Rio de Janeiro Triathlon Triathlon Men's Olympic Distance No Medal 13885 Nara Lorena Salom Artega F 22.0 168.0 55.0 Mexico MEX 2008 Summer 2008 Summer Rio de Janeiro Synchronized Swimming Synchronized Swimming Women's Duet No Medal 13886 Marcella Falice F 18.0 167.0 56.0 Italy ITA 2008 Summer 2008 Summer Athens Rhythmic Gymnastics Rhythmic Gymnastics Women's Group Silver 13884 Matilde Favino Silveira F 26.0 178.0 69.0 Brazil BRA 2008 Summer 2008 Summer Athens Team Handball Team Handball Women's Heavyweight No Medal 13883 Almudena Alarcos M 23.0 171.0 56.0 Spain ESP 2008 Summer 2008 Summer Beijing Sailing Sailing Sailboat Women's 470 13881 Onnia Abdeselam M 26.0 165.0 53.0 Egypt EGY 2008 Summer 2008 Summer Beijing Modern Pentathlon Modern Pentathlon Women's Individual No Medal +-----+ Time taken: 0.011 seconds, Fetched: 10 row(s) </pre>



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	https://github.com/Prerana-Bijekar/BDH
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>Report Generated on: Oct 31, 2024</p> <p>Plagiarism Scan Report By SmallSEOTools</p> <p>Total Words: 388 Total Characters: 453 Plagiarized Sentences: 13 Unique Sentences: 113 (91.6%)</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>
Date	30 October 2025