



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-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)



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	Sanskruि. 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: <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_20190427163204_8354edce-ebed-40f2-b477-f462ba6a6100 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 16:36:16,044 Stage-1 map = 0%, reduce = 0% 2019-04-27 16:36:16,044 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 16:37:21,908 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 1096 11 Paris 1900 30 St. Louis 1904 40 Athina 1906 51 London 1908 132 Time taken: 0.666 seconds, Fetched: 5 row(s)</pre>	
<pre>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_201904271636026_317a525e-b952-4317-8870-e5894508dc26 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 16:36:17,044 Stage-1 map = 0%, reduce = 0% 2019-04-27 16:36:17,044 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 16:37:21,908 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 1096 11 Paris 1900 30 St. Louis 1904 40 Athina 1906 51 London 1908 132 Time taken: 0.666 seconds, Fetched: 5 row(s)</pre>	
<pre>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: 1.959 seconds hive> show create view; 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_20190427164645_6262cccd-0777-8718-8ef5a0324e49 Total jobs = 1 MapReduce task succeeded MapReduce job succeeded 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: 0; number of reducers: 0 2019-04-27 16:47:27,618 Stage-1 map = 0%, 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: 0 Reduce: 0 Cumulative CPU: 4.32 sec HDFS Read: 109999 HDFS Write: 236 SUCCESS Total MapReduce CPU Time Spent: 4 seconds 320 msec OK Athina 1096 2 11 Paris 1900 30 39 St. Louis 1904 8 40 Athina 1906 51 112 London 1908 132 112 Time taken: 73.784 seconds, Fetched: 5 row(s)</pre>	
<pre>hive> create view Last_4_Olympics as select * from summer_olympics where year in ('2010', '2012', '2008', '2004'); OK Time taken: 0.569 seconds hive> select * from Last_4_Olympics limit 10; OK +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 33877 Abdellah M' 32.0 163.0 63.0 Morocco M'R 2008 Summer 2008 Summer Beijing Athletics Athletics Men's 100 metres 33878 Dany Gheorghe P 27.0 159.0 63.0 United States P 2008 Summer 2008 Summer Beijing Athletics Athletics Men's 100 metres 33879 Dany Tafetic F 27.0 178.0 74.0 Australia AUS 2008 Summer 2008 Summer Beijing Rowing Rowing Women's Quadruple Sculls Bronze 33880 Daniel Gushyan D 28.0 172.0 65.0 Hungary HUN 2016 Summer 2016 Summer Rio de Janeiro Rowing Rowing Men's Olympic Double scull No Medal 33881 Daniel Lescure F 27.0 178.0 63.0 France FRA 2008 Summer 2008 Summer Rio de Janeiro Swimming Swimming Women's 4x100m Medley Relay No Medal 33882 Danilo Salazar F 24.0 181.0 77.0 Brazil BRA 2012 Summer 2012 Summer London Boxing Boxing Men's Light-Heavyweight Bronze 33883 David Tlusty D 20.0 180.0 67.0 United States U.S. 2008 Summer 2008 Summer London Gymnastics Artistic Gymnastics Men's Floor Exercise Silver 33884 Nelli Palavina Silva F 20.0 176.0 69.0 Brazil BRA 2008 Summer 2008 Summer Rio de Janeiro Taekwondo Taekwondo Women's Heavyweight No Medal 33885 Delia Palacio F 23.0 171.0 64.0 Argentina ARG 2016 Summer 2016 Summer Rio de Janeiro Taekwondo Taekwondo Women's Heavyweight One Person Duel Gold 33886 Denisa Khatek Falak D 27.0 176.0 66.0 Egypt EGY 2008 Summer 2008 Summer Beijing Modern Pentathlon Modern Pentathlon Women's Individual No Medal +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ Time taken: 0.621 seconds, Fetched: 10 row(s)</pre>	
Output Analysis	<p>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</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.

	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/sanskruti-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%)	 <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	30/10/2025