

Lab Assignment 07 - Batch 01

Aim:

Movie Ratings Analysis in Pig

Objective:

Perform analysis on a dataset containing movie ratings to identify popular movies and analyze user preferences.

Tasks:

1. Load the movie ratings data into Pig.
2. Calculate the average rating for each movie.
3. Identify the top-rated movies.
4. Determine the most frequently rated movies.
5. Analyze user preferences by identifying the most active users (users with the highest number of ratings).
6. Store the results into HDFS.

Theory:

- Explain Pig and its commands

Instructions:

1. Set up a Hadoop cluster with Apache Pig installed.
2. Download the datasets for movie ratings, customer data, and web server logs.
3. Write Pig scripts to perform the analysis tasks outlined in each lab exercise.
4. Run the Pig scripts on the Hadoop cluster to execute the analysis.
5. Store the results into HDFS for further analysis or reporting.
6. Document your findings, observations, and any insights gained from the analysis.

Lab Assignment 07 - Batch 02

Aim:

Customer Segmentation Analysis in Pig

Objective:

Segment customers based on their demographic and behavioral attributes.

Tasks:

1. Load customer data into Pig.
2. Segment customers based on age and income.
3. Segment customers based on spending score.
4. Analyze customer segments to understand demographic and behavioral differences.
5. Store the results into HDFS.

Theory:

- Explain Pig and its commands

Instructions:

1. Set up a Hadoop cluster with Apache Pig installed.
2. Download the datasets for movie ratings, customer data, and web server logs.
3. Write Pig scripts to perform the analysis tasks outlined in each lab exercise.
4. Run the Pig scripts on the Hadoop cluster to execute the analysis.
5. Store the results into HDFS for further analysis or reporting.
6. Document your findings, observations, and any insights gained from the analysis.

Lab Assignment 07 - Practice Assignment

Aim:

Web Server Log Analysis in Pig.

Objective:

Analyze web server logs to understand website traffic patterns and user behavior.

Tasks:

1. Load web server log data into Pig.
2. Extract IP addresses, timestamps, and URLs accessed.
3. Count the number of unique visitors.
4. Analyze traffic patterns over time (e.g., hourly, daily).
5. Identify the most frequently accessed URLs.
6. Store the results into HDFS.

Theory:

- Explain Pig and its commands

Instructions:

1. Set up a Hadoop cluster with Apache Pig installed.
2. Download the datasets for movie ratings, customer data, and web server logs.
3. Write Pig scripts to perform the analysis tasks outlined in each lab exercise.
4. Run the Pig scripts on the Hadoop cluster to execute the analysis.
5. Store the results into HDFS for further analysis or reporting.
6. Document your findings, observations, and any insights gained from the analysis.

Outcome:

These lab exercises provide hands-on experience in using Apache Pig for data analysis and processing. They cover a variety of scenarios commonly encountered in data analytics and can help develop skills in data manipulation, aggregation, and analysis using Pig.