JIGSAW ACADEMY
THE ONLINE SCHOOL OF ANALYTICS

Non Graded Assignment

T 5.1.2Hive

In this case study, we are giving a simple real world example of how to use HIVE on Top of HADOOP for different statistical analysis. In this example, we have a predefined dataset (cricket.csv) having more than 20 columns and more than 10000 records in it. In the below example, we are explaining how to display few records based on attributes of cricket. The cricket.csv file has different attributes like name, country, Player_iD, yearID, stint, teamid, run, Igld, D, G, G_batting, AB, R, H, 2B, 3B, HR, RBI, SB, CS, BB,SO,I BB,HBP,SH,SF,GIDP,G_OLD

Description:

Name: It describe the name of the player

Player_iD: It describes the player id of the player

Year: It describes the year

R: It describes the run.

Problem statements:

1. Display year and maximum runs scored in each year.

2. Display player_id,year and maximum runs scored in each year.

3. Find out each and every year how many centuries was scored.

4. Find out how many centuries was scored in each and every year by the each and every different TeamID