Trainer : Naveen Pn /
www.linkedin.com/in/naveen-pn

Structured Data Analysis using Hive

Basic Hands-On Assignment

Total Points : 2 (Exercise) X 20 = 40 pts

Problem Statement 01

Prerequisite :

Create a table with the schema as specified below and load the data.

Download Dataset

id	tenure	designation	salary
INT	INT	STRING	BIGINT

Write a query to derive a new column extra_vacation based on the tenure served, the logic is as given below.

1. If tenure < 2, Then 20
2. If tenure is 2-10 then 30 days
3. If tenure > 10 then 40 days

Problem Statement 02

Prerequisite :

Create a table “temperature” to store the dataset as mentioned in the schema and load the data.

Download Dataset

Date	State	Temperature

STRING	STRING	Array<DOUBLE>

Write a query to calculate the maximum temperature of each state.

Problem Statement 03

Prerequisite :

Create a table 'student_marks' with schema as shown above and load the data into the 'student_marks' table.

 Download Dataset

Name	Marks
STRING	Map<STRING, INT>

Write a query to perform below mentioned tasks:

1. Display NAME who have scored more than 90 in subject Maths subject
2. Display NAME and <Marks Scored in Physics subject>
3. Display NAME, and <Maximum-Subject-Marks>
4. Display NAME and <Average Marks Scored>
5. Display NAME and <Percentage of marks>

Problem Statement 04

Prerequisite :

Create a table “student_info” with schema as show below and load the data

 Download Dataset

Name	Marks	Address

STRING	Map<STRING, INT>	Struct<doorNo INT,Location String,Pincode INT>

Write a query for the below mentioned tasks

1. Display all “NAME” who is located in Banashankari
2. Calculate the total count who is staying in pin code 560001