Practical -13

AIM- To work with Hive in Hadoop.

Exercise: You work as a data analyst for a bank, which now needs you to analyse a few things mentioned below in order to launch a new scheme.

Tasks:

1. Create Table for the Bank Dataset using following columns:

   1 - age (numeric)
   2 - job: type of job
   ("admin.","unknown","unemployed","management","housemaid","entrepreneur","student",
   "blue-collar","self-employed","retired","technician","services")
   3 - marital : marital status ("married","divorced","single"; note: "divorced" means divorced or widowed)
   4 - education ("unknown","secondary","primary","tertiary")
   5 - default: has credit in default? ("yes","no")
   6 - balance: average yearly balance, in euros (numeric)
   7 - housing: has housing loan? ("yes","no")
   8 - loan: has personal loan? ("yes","no")
   9 - contact: contact communication type ("unknown","telephone","cellular")
   10 - day: last contact day of the month (numeric)
   11 - month: last contact month of year "jan", "feb", "mar", ..., "nov", "dec")
   12 - duration: last contact duration, in seconds (numeric)
   13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
   14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)
   15 - previous: number of contacts performed before this campaign and for this client (numeric)
   16 - poutcome: outcome of the previous marketing campaign (
   "unknown","other","failure","success")
   17 - y - has the client subscribed a term deposit? ("yes","no")

2. understand the schema, LOAD file in to table
3. with the use of AND Logical operator get the records who has age greater than 25 and married.
4. Who has not subscribed to a term deposit (column: y)
5. List the age group of the people who have subscribed to a term deposit
6. Find the job status for the people who has been contacted more than once.
7. Find the number of single people contacted and sort by age
8. Calculate the average balance in the month of June and July