**Gcloud Setup**

Create your google cloud account and spin up a dataproc cluster with API permissions enabled.

**Transferring file from local to VM:**

gcloud compute scp car\_insurance\_cold\_calls\_dataset.csv powerpavan23@hadoop-cluster-m:/home/powerpavan23

**Schema Creation:**

create external table car\_insurance\_ext

(Id string,Age int,Job string,Marital string,Education string, Default int,Balance int,HHInsurance int,CarLoan int,Communication string,LastContactDay int,LastContactMonth string,NoOfContacts int,DaysPassed int,PrevAttempts int,Outcome string,CallStart string,CallEnd string,CarInsurance int)

row format delimited

fields terminated by ',';

**If ^m is faced:**

(optional) sed -i 's/\r$//' car\_insurance\_cold\_calls\_dataset.csv --> to remove ^m characters

**Loading the Data:**

load data local inpath 'file:///home/powerpavan23/car\_insurance\_cold\_calls\_dataset.csv' into table car\_insurance\_ext;

**Get Header column:**

set hive.cli.print.header = true;