Steps

Create cluster

Goto workspace - create - notebook

Python

Scala

Sql

R

Add data

Create table with ui

Preview table

Workspace  do coding

Select \* from name of file :

Machine Learning

Run the data

library(SparkR)

sparkR.session()

df1 <- read.df("/FileStore/tables/Model.csv", "csv", header = "true", inferSchema = "true")

head(df1)

Split The Data

split\_df1<- randomSplit(df1, c(0.7,0.3), seed=0)

split\_df1

#Training data

df1\_train <- split\_df1[[1]]

count(df1\_train)

head(df1\_train)

#test data

df1\_test <- split\_df1[[2]]

count(df1\_test)

head(df1\_test)

#Make model

model <- spark.glm(df1\_test, bmi ~ age+wt+ht, family="gaussian")

summary(model)

#Prediction

predicted <- predict(model, df1\_test)

head(predicted)