# Assignment - Data Load Activity (20%)

Due: Module 3

## Purpose:

The purpose of this assignment is to load data into a current data analysis tool.

Assignment Objective:  
For the business problem you selected, perform the required steps to: load a CSV data file into HDInsight using the Azure Data Storage Tool, create a Data Schema using Hive and verify that the data have been loaded into Hive correctly.

## Instructions:

This activity relates to the human resources analysis data.

**Click on the link below to download your data:**

* [**Human Resources analysis - Data**](https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset/data)

In this assignment you will load your business problem data into HIVE.

1. Load your csv file in HDInsight using the Azure Storage Tool. Put the data in a container called Employee
2. A) Create a schema for your data from the raw data.

B) Using your schema, load the data from Azure Storage into Hive. Call the table, “employee”.

1. Validate that the schema is correct using the “table view” tab.
2. Inspect the data were loaded properly using the SQL query interface.

## This assignment relates to the following Course Learning Requirements:

CLR 1: Appraise the strengths and limitations of current tools in data analysis;

CLR 4: Extract and load data to support problem solving and decision-making.

## Grading for This Assignment

Please take a screen shot of each step, showing the result. Also provide the query you used.

|  |  |
| --- | --- |
| **Steps required** | **Grade Earned** |
| 1. A) Create a schema for your data from the raw data.   B) Using your schema, load the data from Azure Storage into Hive. Assume that the data is in a container called “employee”. Call the table, “employee”. | 50% |
| 1. Validate that the schema is correct using the “table view” tab. | 80% |
| 1. Inspect the data were loaded properly using the SQL query interface. | 100% |