**Assignment: Assignment #1  
Name: Kenatu Bruce  
Student ID: C0770979  
Course ID: DAT1030**

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. - **Human Resources analysis - Data**

1. **Create Schema**

CREATE EXTERNAL TABLE employee  
(  
Age int,  
Attrition int,  
BusinessTravel string,  
DailyRate int,   
Department string,  
DistanceFromHome int,  
Education int,  
EducationField string,  
EmployeeCount int,  
EmployeeNumber int,  
EnvironmentSatisfaction int,   
Gender string,  
HourlyRate int,  
JobInvolvement int,  
JobLevel int,  
JobRole string,  
JobSatisfaction int,  
MaritalStatus string,  
MonthlyIncome int,  
MonthlyRate int,  
NumCompaniesWorked int,  
Over18 char,  
OverTime string,  
PercentSalaryHike int,  
PerformanceRating int,  
RelationshipSatisfaction int,  
StandardHours int,  
StockOptionLevel int,  
TotalWorkingYears int,  
TrainingTimesLastYear int,  
WorkLifeBalance int,  
YearsAtCompany int,  
YearsInCurrentRole int,  
YearsSinceLastPromotion int,  
YearsWithCurrManager int  
)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED AS TEXTFILE LOCATION '/employee'

TBLPROPERTIES("skip.header.line.count"="1");

1. **Verify Data Load**

SELECT \* FROM employee LIMIT 10;

1. **Upload file to Azure Storage**

hadoop fs -copyFromLocal WA\_Fn-UseC\_-HR-Employee-Attrition.csv /example/data/employee/hrdata.csv