# Practice Project: Data Processing Using Spark

## Scenario

You have been hired as a Junior Data Engineer by BDPS Corporation and you have been provided with links to two raw datasets that you need to acquire and perform ETL on using PySpark and Hive warehouse.

## Project Overview

This practice project focuses on data transformation and integration using PySpark. You will work with two datasets, and perform various transformations such as adding columns, renaming columns, dropping unnecessary columns, joining dataframes, and finally, writing the results into a Hive warehouse and an HDFS file system.

Skills Network Labs (SN Labs) is a virtual lab environment used in this course. Upon clicking the "Launch App" button below, your Username and Email will be passed to SN Labs and will be used in strict accordance with IBM Skills Network Privacy policy, such as for communicating important information to enhance your learning experience.