**AIT614 LAB 5**

This lab is specially designed for you to learn **Spark MLlib** with Python on the Databricks platform, including tools for data preprocessing, Machine Learning (ML) pipelines, several different machine learning algorithms, and corresponding ML applications. 

This lab includes **TWO tasks**/**mini-projects**:

* 1) **market basket analysis** and **recommendation**
* 2) **binary classification** using **logistic regression**

More specifically, in this lab, you will learn

* how to **preprocess**data with Spark MLlib
* how to use machine learning pipelines for **optimization**with Spark MLlib
* how to conduct *advanced*data analytics with Spark MLlib
  + **Association rules**
  + **Classification**
  + **Hyperparameter tuning** and **cross validation**
  + **Market basket analysis**
  + **Product recommendation**
* how to embed SQL query in Spark on Databricks

All of these will be very helpful for your final project.

      First, please follow **step-by-step** **instructions**in**Lab 5**to log into the [**Databricks Community Edition**](https://community.cloud.databricks.com/), then continue following the **lab instructions** and **code examples/tutorials** as shown in the lab to complete all required tasks on the Databricks platform.