**DSBA 6190 | Distributed Computing Lab**

**Description:** The goal of the “containerization” portion of this lab is to learn how to build a Docker image that can train a machine learning model from data housed in your data lake and to execute this model training from a Kubernetes cluster. Then, the Databricks portion of this lab will walk you through a similar machine learning exercise

**Notes:**

* Only provision the requested resources using the defined settings. Remember, the class cloud budget is everyone’s responsibility.

## Steps:

1. Build Your Docker Image. (EACH GROUP MEMBER)
2. Push to the Azure Container Registry. (EACH GROUP MEMBER)
3. Connect to the Azure Kubernetes Service
4. Run Your Container in Kubernetes and Train a Model. (EACH GROUP MEMBER)
5. Complete Distributed ML Training on Databricks with Apache Spark. (EACH GROUP MEMBER)

# Part 1: Containerization with Docker and Kubernetes

Step 1: Build Your Docker Image

Step 2: Push to the Azure Container Registry

Step 3: Connect to the Azure Kubernetes Service

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Step 4: Run Your Container in Kubernetes and Train Your Model

# Part 2: Distributed Machine Learning on Azure DataBricks

Step 5: Follow the steps in the Azure Databricks notebook.

Turn in this Word document for Part 1.

For Part 2, leave your notebook in your user folder in the Azure Databricks workspace.