CS 519 Cloud Computing Overview

**VL10: Scale & Load Balance your Architecture**

School of Technology and Computing

## **Instructions**

## For this activity you will be using AWS Academy Labs - **Scale & Load Balance your Architecture.**

Please complete the provided lab.

**Write a 150-word summary to explain your understandings and findings from this lab assignment.**

Auto Scaling helps to achieve maximum benefits of using the cloud. This service automates the process of scaling out or in. For example, EC2 instances can be added automatically when the load average spikes above a threshold.  The advantages of using Auto Scaling are:

* The possibility of maintaining the amount of compute resources on application requirements level, so the system is highly available and much cheaper.
* Controlling the health of instances and replacing unhealthy instances makes the system fault tolerant.

Paired together, auto scaling and load balancing provide useful benefits. Say you want to smoothly handle traffic surges to your website. When load increases, you want the website infrastructure to have enough capacity to serve the traffic. During bouts of low activity, naturally you want to reduce capacity.

With load balancing alone, you’ll have to know ahead of time how much capacity you need so you can keep additional instances running and registered with the load balancer to serve higher loads. Or you could manually stop worrying about it and auto scale based on say CPU usage so that instances increase or decrease dynamically based on the load. Now this should give you a good idea of why it makes sense to have both.