Master of Computer and Information Sciences

COMP810 Data Warehousing and Big Data

Big Data Tutorials v6.0

Dr Weihua Li

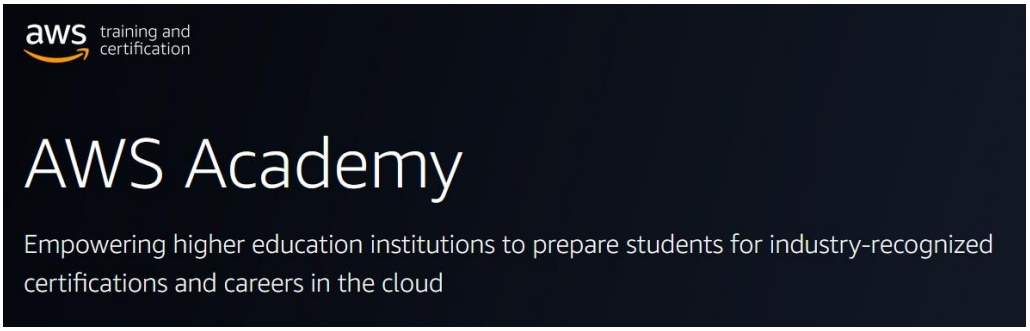
July 2024

Contents

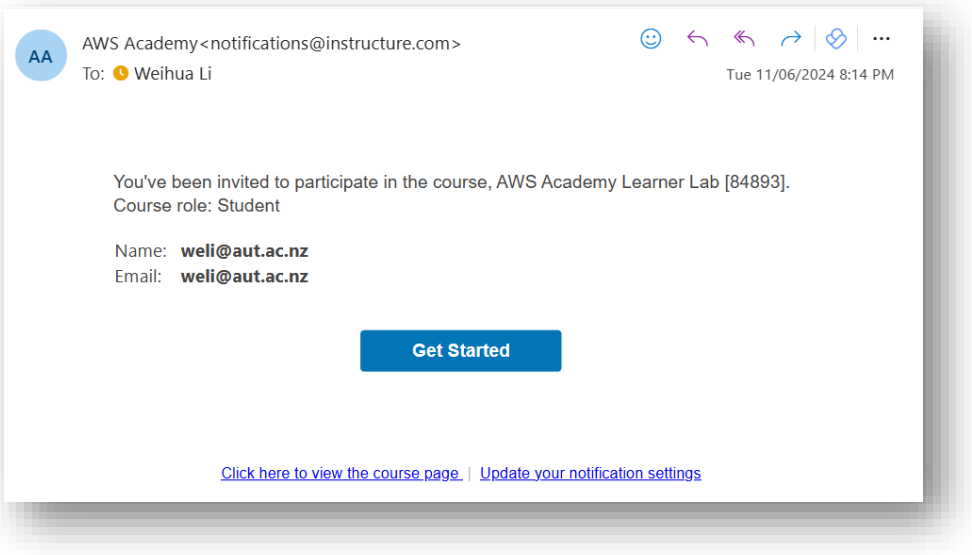
Lab 1 – Preparation and Assignment Focused

This lab is to prepare for your assignment (report) and the following week's labs. The first and most important task for you is to complete Amazon Academy Account registration. Because Amazon EMR is required in the next few labs. You should already receive an email (your AUT mailbox) from Amazon regarding AWS Academy account registration. After that, please find a teammate if you want to work on the big data report in a group of two students. Identify a potential research topic together by searching Google Scholar.

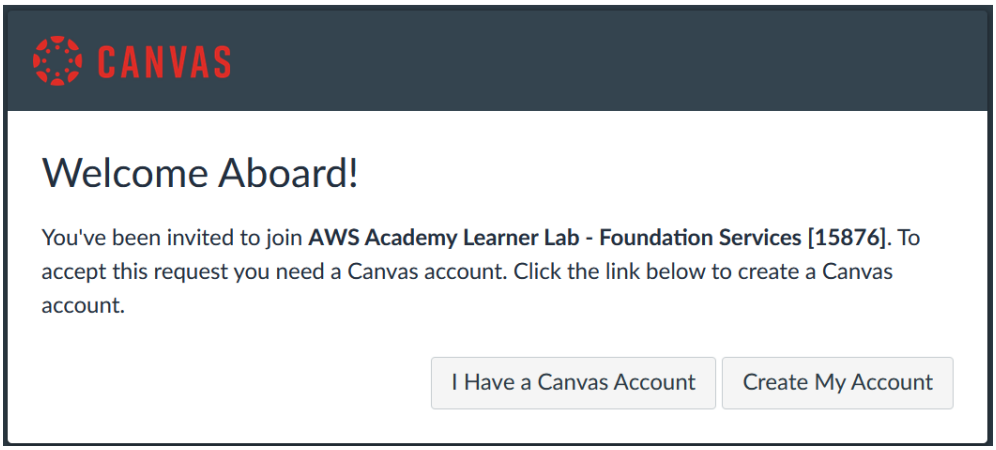
Task 1: AWS Academy Account



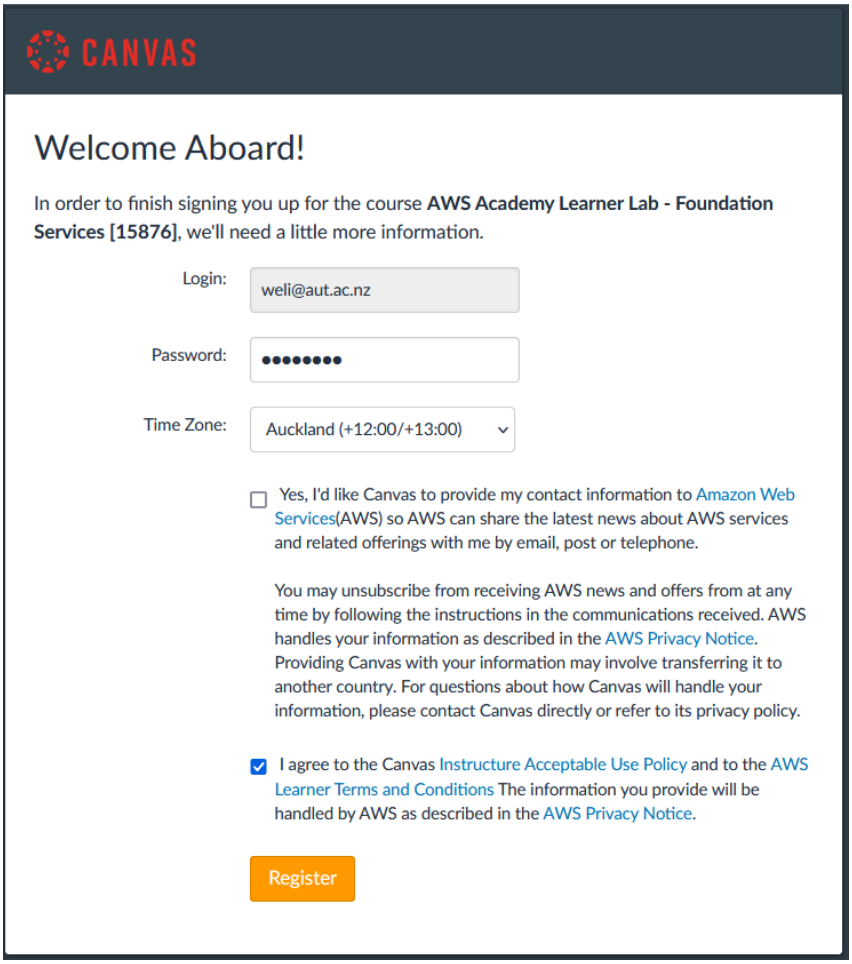
You should already receive an email (your AUT mailbox) from Amazon regarding AWS Academy account registration. An example is as below.



Click the “Get Started” in your email and then click “Create My Account”.



Please provide a password for your account and select “Auckland” as the time zone. Then tick the “I agree…” and click the register button.

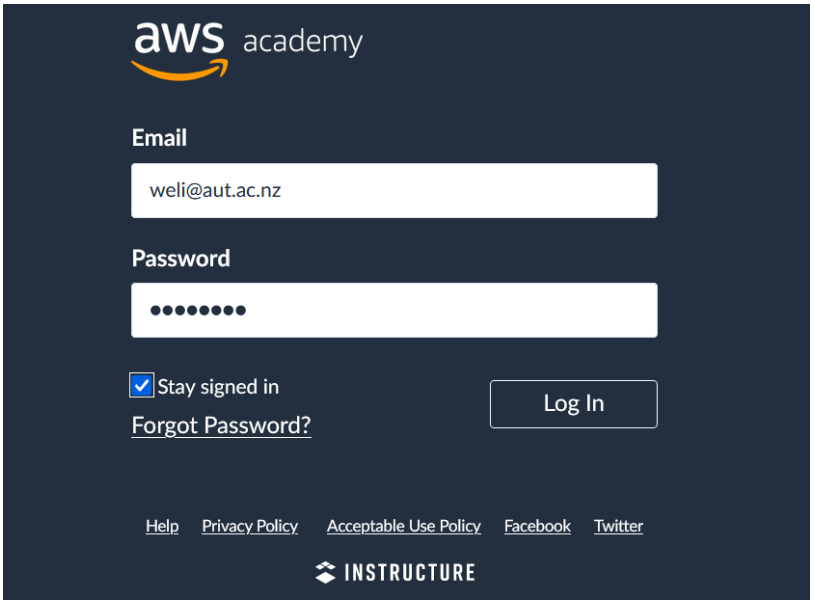


Task 2: Visit AWS Academy Canvas

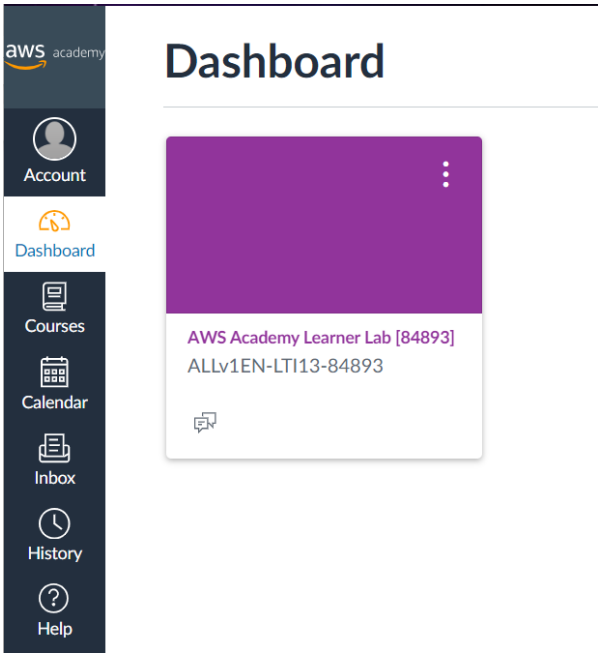
Once you finish the AWS Academy Account registration, please log in your account via: https://awsacademy.instructure.com/ ***(Please bookmark this URL)***

Please select Student Login.

Key in your AWS Academy credential. The email refers to your AUT email address, and the password is set during the registration process.



After login, under the Dashboard of AWS Academy Canvas, you should be able to see the course “AWS Academy Learner Lab – Foundation Services”. Please click it and get in the course.



Task 3: Identify a potential topic for your report

Walkthrough the slides of Big Data Use Cases. Find one or come up with your own. Identify challenging issues. You need to search articles using Google Scholar (access via AUT proxy):

https://scholar-google-co-nz.ezproxy.aut.ac.nz/

Task 4: Install JDK 1.8 and NetBeans 14 (optional)

Perform this if you use your personal laptop.

NB: The AUT Lab comes with JDK 11 with NetBeans 14. However, the Hadoop supports JDK 8 well. Thus, the examples I will give to you work well with JDK 8. The map-reduce simulation works for JDK 8 and JDK 11. However, the programs uploaded to Amazon EMR clusters for processing are only compatible with JDK 8.