

Master of Computer and Information Sciences

COMP810 Data Warehousing and Big Data

Big Data Tutorials v6.0



Dr Weihua Li

July 2024

Contents

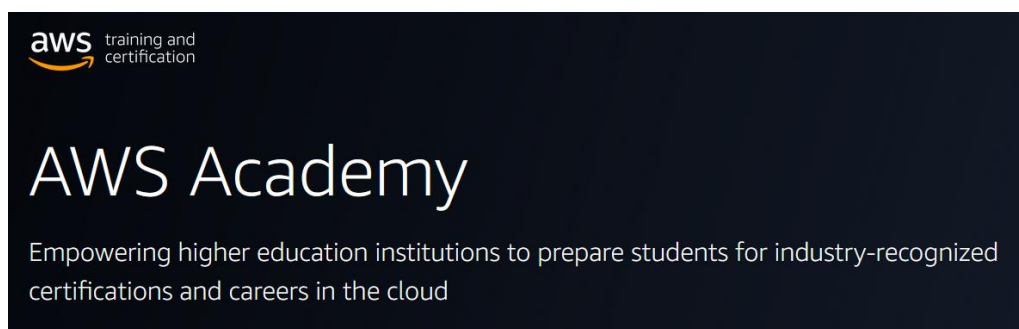
Lab 1 – Preparation and Assignment Focused	3
Task 1: AWS Academy Account	3
Task 2: Visit AWS Academy Canvas	4
Task 3: Identify a potential topic for your report.....	6
Task 4: Install JDK 1.8 and NetBeans 14 (optional)	6
Lab 2 Hadoop - AWS EMR.....	7

Task 1 Launch an Amazon EMR cluster	7
Task 1.1 Login AWS Academy account	7
Task 1.2 Launch Your First Amazon EMR Cluster	9
Task 2 Elastic MapReduce Hadoop Job Using Custom Jar	14
Task 2.1 Prepare Java map-reduce program (Optional).....	14
Task 2.2 Create Amazon S3 Bucket	14
Task 2.3 Execute the Java program using Map-reduce	16
Task 2.4 Terminate your cluster	19
Lab 3 Hive QL.....	20
Task 1 Preparation – Update Security Policy.....	20
Task 2 Hue and HQL.....	21
Task 3 Employee Data Analysis with Hive	23
Task 4 Analysing Oil Import Prices with Hive	28
Task 5 Joining Tables in Hive (Optional)	31
Task 6 Hive Thrift – JDBC (Optional).....	33
Reference and Resources	34
Lab 4 Elasticsearch and Kibana (I)	35
Task 1 Getting started with Elasticsearch and Kibana.....	35
Task 2 Domain Specific Language (DSL)	38
Task 3 Using Analyzer API.....	40
Task 4 Dynamic and Explicit Mapping	41
Task 5 Understand Coercion	43
Reference and Resources	44
Lab 5 Elasticsearch and Kibana (II)	45
Task 1 Customize Analyzer	45
Task 2 Use Match and Term Query	48
Task 3 Use Aggregations.....	48
Task 4 Visualization using Kibana	53
Task 5 Elasticsearch – Java API (Optional).....	57
Reference and Resources	58

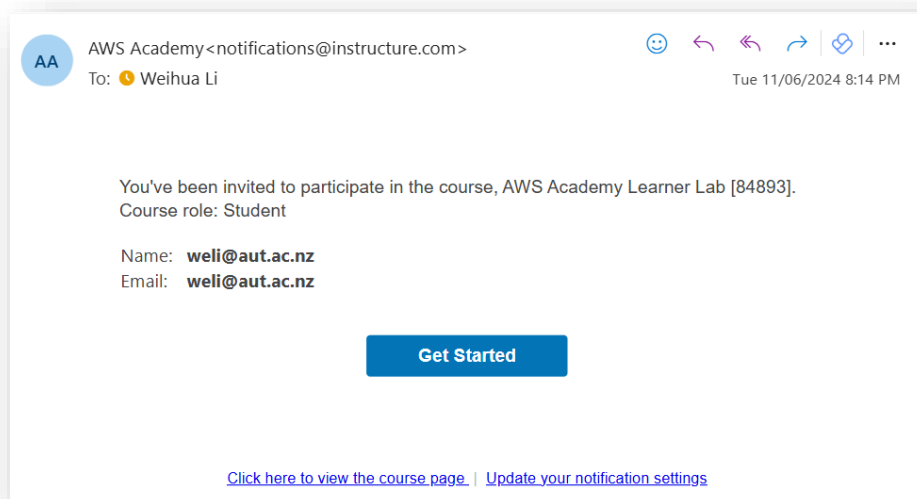
Lab 1 – Preparation and Assignment Focused

*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”.




Welcome Aboard!

You've been invited to join **AWS Academy Learner Lab - Foundation Services [15876]**. To accept this request you need a Canvas account. Click the link below to create a Canvas account.

[I Have a Canvas Account](#) [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.



Welcome Aboard!

In order to finish signing you up for the course **AWS Academy Learner Lab - Foundation Services [15876]**, we'll need a little more information.

Login:

Password:

Time Zone:

☐ Yes, I'd like Canvas to provide my contact information to [Amazon Web Services\(AWS\)](#) so AWS can share the latest news about AWS services and related offerings with me by email, post or telephone.

You may unsubscribe from receiving AWS news and offers from at any time by following the instructions in the communications received. AWS handles your information as described in the [AWS Privacy Notice](#). Providing Canvas with your information may involve transferring it to another country. For questions about how Canvas will handle your information, please contact Canvas directly or refer to its privacy policy.

☒ I agree to the Canvas [Instructure Acceptable Use Policy](#) and to the [AWS Learner Terms and Conditions](#) The information you provide will be handled by AWS as described in the [AWS Privacy Notice](#).

[Register](#)

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.



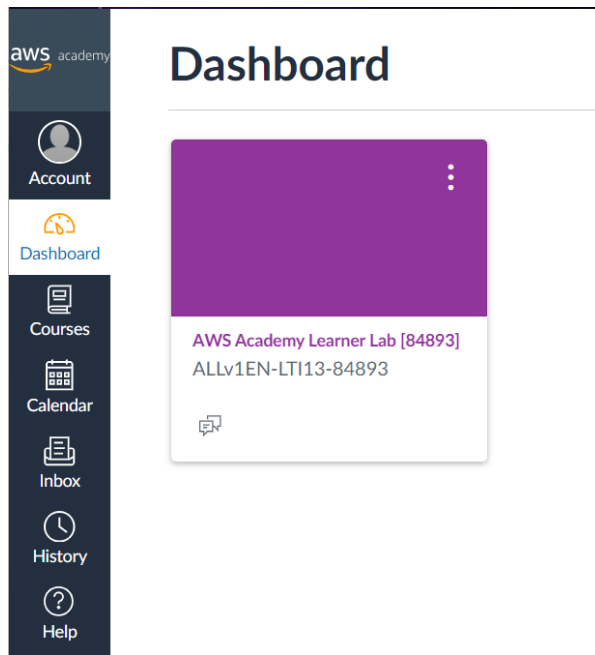
The image shows the AWS Academy login page. At the top is the AWS Academy logo. Below it, the 'Student Login' button is highlighted with a red rectangle. Underneath this button, there is text in English: '(For students enrolled in a class)' and in Japanese: '学生の方はここからログインしてください。'. Below that is the Chinese text: '已注册课程的学生请在这里登录'. A horizontal line separates this section from the 'Educator Login' section below. The 'Educator Login' button is also visible. Below it, there is text in English: '(For educators who have access to the AWS Academy Portal)' and in Japanese: '講師の方（AWS Academyメンバーポータルアカウントをお持ちの方）はここからログインしてください。'. At the bottom of the educator section is the Chinese text: '教师请在这里登录（您需使用AWS Academy Portal账户登录）'.

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



The image shows the AWS Academy login form. It has a dark blue background with the AWS Academy logo at the top. Below the logo, there are two input fields: 'Email' and 'Password'. The 'Email' field contains the text 'weli@aut.ac.nz'. The 'Password' field is filled with dots. Below the password field, there is a checkbox labeled 'Stay signed in' which is checked. To the right of the checkbox is a 'Log In' button. Below the checkbox is a link that says 'Forgot Password?'. At the bottom of the page, there are links for 'Help', 'Privacy Policy', 'Acceptable Use Policy', 'Facebook', and 'Twitter'. At the very bottom is the 'INSTRUCTURE' logo.

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.