

2022 數據分析研習營暨成果競賽活動

一、 特訓課程緣起

大數據分析是指從大量資料中擷取寶貴資訊以制定有效的決策，大數據分析是國家發展的重要政策，也是各大學普及的研究議題，產業界亦有相當多的應用範例；在人力需求方面，根據「1111 人力銀行」2022 年 8 月刊登「數據分析師」相關的職缺近 8,000 人。

本次營隊規劃兩天共 12 小時的基礎訓練時程，讓學員能瞭解大數據分析的理論與知識，並且實機操作最具價值的 Apache Hadoop 軟體架構與分析工具。課程安排深入淺出，學員不限領域與系科，只要動手操作就能學會基礎架構，歡迎踴躍參加。

二、 課程介紹

面對 Big Data 這波巨大資料的洪流，需要依靠來自 Apache 開源碼專案 Hadoop，來當做驅動海量資料的核心動力，而 Hadoop 這宛如渦輪引擎般的動力，即可輕易駕馭這龐大到難以數計的資料流。然而，要架構此核心動力，並不需要採購高額資金的設備來建置，且不管小從個人，或大到跨國企業，只需要掌握以下的關鍵技術，就可以利用相當低廉且有限的預算，來建置可驅動 Big Data 的動力引擎，進而探究 Big Data 所帶來的下個世代的智慧分析與應用。

三、 課程目標

培訓學員具有 Big Data 初階實務分析能力

四、 課程內容

本課程旨在建立 Big Data 處理、分析、應用的根基，讓參訓學員瞭解正確的觀念與方法，重點並不在講解程式設計的細節，而在於透過體驗式教學方式的實作，透過指令來體驗實際的操作方式，以從體驗中驗證課程所學。

五、 訓練方式：採實體教室操作演練

六、 訓練期間：111 年 11 月 12 日(週六)、110 年 11 月 13 日(週日)

七、 學員資格：高中三年級學生，不限科系與專長領域、本校資訊學院在學生，全程免費，額滿為止。

八、 完訓學員：完訓研習證明

九、 競賽得獎：競賽得獎證明

十、 課程大綱

第一章 大數據簡介

- 工作環境介紹與操作
- Linux ETL(用 Linux 命令來萃取資料)

第二章 分散式檔案系統(HDFS)介紹

- 從開放平台下載資料
- 將資料傳至 HDFS
- 操作 HDFS

第三章 資料分析語言介紹與基本操作

- 認識 Hadoop 資料倉儲工具
- 資料分析的基本語法
- 資料的篩選與排序

第四章 數據分析競賽模擬演練

十一、 課表

日期	時段	課程
11/12(六)	09:10-10:00	大數據簡介
	10:10-11:00	工作環境介紹與操作
	11:10-12:00	Linux ETL(用 Linux 命令來萃取資料)
	13:10-14:00	分散式檔案系統介紹
	14:10-15:00	從開放平台下載資料
	15:10-16:00	將資料上傳至 HDFS

11/13(日)	09:10-10:00	操作 HDFS
	10:10-11:00	資料分析語言介紹與基本操作
	11:10-12:00	認識 Hadoop 資料倉儲工具
	13:10-14:00	資料分析的基本語法
	14:10-15:00	數據分析競賽模擬演練
	15:10-16:00	成果競賽