

10801 雲端計算技術與應用 (CLOUD COMPUTING) 課程大綱

授課教師: 詹毓偉

Email: ywchan@gm.pu.edu.tw

Office: Providence Hall 577

<https://sites.google.com/site/profyuwei/>

資訊學院 上課公約

- ❑ 上課時應專心，不可嘻戲或閱覽課程無關之書籍資料。
- ❑ 背包或非上課所需之物品（包含手機）必須置於桌椅下方。
- ❑ 不得於上課時間於教室內食用食物或喝飲料。
- ❑ 上課時不可睡覺或有妨礙他人上課之行為。

課程目標

□ 此課程主要分為二大部份

Part 1: 雲端技術的介紹與實際操作

- Hadoop Ecosystems 介紹與實作
 - (1) MapReduce計算架構概念與實作
 - (2) HBase資料庫系統介紹與實作
 - (3) Sqoop系統介紹與實作

Part 2: 雲端技術的應用與開發

- 巨量資料分析
 - (1) Python語法介紹
 - (2) 機器學習演算法

上課用書與參考網站

□ 上課用書

- Hadoop+Spark大數據巨量分析與機器學習整合開發實戰,
林大貴, 博碩文化

□ 聯合大學 陳士杰教授:

- 巨量資料技術與應用課程
- 課程網址:

http://debussy.im.nuu.edu.tw/sjchen/BigData_final.html

參考用書

- ❑ 林峰正, 王耀聰, 辜文元, 施頴陽, 周天穎, Hadoop技術手冊 (Tom White, 第四版), 歐萊禮
- ❑ 林大貴, Python+Spark 2.0 + Hadoop機器學習與大數據分析實戰, 博碩文化
- ❑ 陳允傑, Python資料科學與人工智慧應用實務, 旗標科技

成績評分

項目	配分
期中考	30%
上課Lab練習+小考	40%
期末Project分組實作報告 (以2個人為一組)	30%

Attention: 期末Project分組實作報告 (以2個人為一組):

- Each group should PRESENT (10 min) and DEMO (5 min) your outstanding works.
- Please turn-in documents as following when you present your final project:
 - (a) Presentation Slides
 - (b) Final Report
 - (c) Source Code (include readme.txt)

上課實作練習

- Lab 1: Hadoop Installation
- Lab 2: HDFS Operations
- Lab 3: MapReduce Programming (WordCount)
- Lab 4: HBase Programming
- Lab 5: Sqoop Operations
- Lab 6: Python資料爬蟲實作(1)
- Lab 7: Python資料爬蟲實作(1)
- Lab 8: 資料清理與視覺化
- Lab 9: 資料分析 – 機器學習線性迴歸演算法
- Lab 10: 資料分析 – 機器學習邏輯迴歸演算法(分類)
- Lab 11: 資料分析 – 機器學習決策樹演算法
- Lab 12: 資料分析 – 機器學習分群演算法

→ UCI Machine Learning Repository: 提供資料集