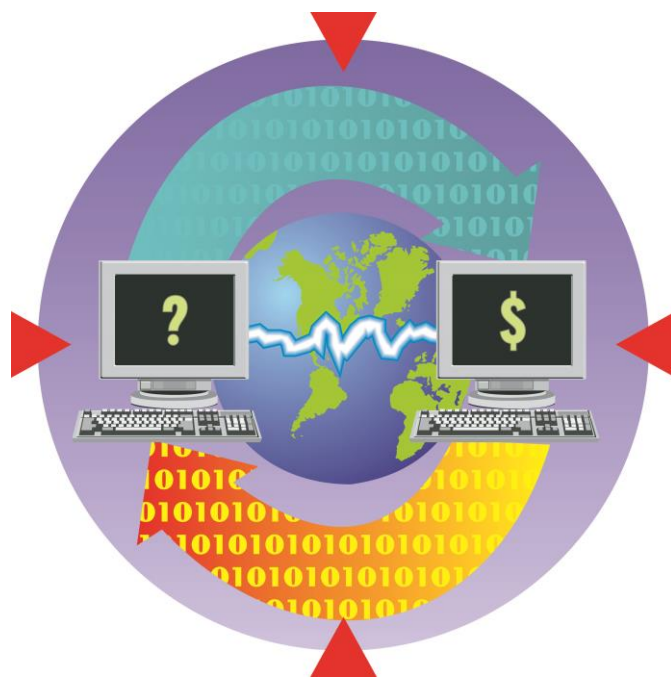


Spark

大數據分析實務班

2018/7/28週六開課，實作演練!

7/20 前報名，享早鳥優惠!



主辦單位：財團法人資訊工業策進會

相關訊息：<http://www.iiedu.org.tw/ites/SPARK.htm>

一、課程緣起：

隨著物聯網、社交引擎普遍使下，巨量資料的浪潮席捲而來，而巨量資料的管理、分析、應用等儼然成為巨大的挑戰，資料的時效性也是另一個重要的課題，往往取得資料時，必須在最短時間內反映，否則將會錯失商機。

近年來，有大量的業界、學者投入巨量資料相關技術，根據 Hadoop Summit 2014、2014 大數據技術大會等國外指標性論壇所討論的主軸，In-Memory 技術已成為近期最熱門的巨量資料技術，而最為人熟知的是由加州柏克萊大學所發展的 BDAS (Berkeley Data Analytics Stack)，其中從儲存、運算、資料檢索都利用了 In-Memory 技術，提供優於 MapReduce 運算框架 10~100 倍的效能。而 Apache Spark 便是其中扮演重要的運算框架角色，Spark 憑藉其特有的資料封裝操作框架 RDD 以及 In-memory 資料儲存技術，解決傳統 MapReduce 框架在迭代式演算法中所遇到的效能瓶頸。基於 Spark 框架上，有許多擴充模組，包含以 SQL 指令操作資料的 SparkSQL、豐富的機器學習函式庫 Mllib 以及即時串流處理技術 Spark Stream 等足以覆蓋巨量資料 90% 的處理需求。

為了協助國內產業界引進最新巨量資料技術，資策會特規畫「Spark 大數據分析實務班」課程，課程注重於巨量資料實務開發，內容從開源巨量資料平台建設到 Spark 大數據分析應用皆有含括。在課程中，亦會導入真實資料進行實際應用開發，使學員能充分利用 Spark 工具來解決問題，加強學習效果，讓所學能真正導入實際工作中。

二、課程目標：

本課程將學習如何利用開源軟體，建置大型巨量資料管理平台，並能透過該平台進行 Spark 開發應用，讓學員能夠了解 Spark 核心技術、SparkSQL 操作方式、機器學習函式庫應用，更進一步讓學員了解 Spark 程式設計模式、內部核心運作方式與效能調教。課程最後將透過實際分組演練真實數據，透過資料探索的過程，彼此分享學習，使學員更熟悉 Spark 工具的使用。

三、課程特色：

本課程旨在建立 In-Memory 巨量資料處理、分析、應用的根基，讓參訓學員瞭解正確的觀念與方法，課程重點在於透過採體驗式教學方式的實作，學員能夠了解如何快速搭建一巨量資料分析叢集，並能在自建平台上實際演練巨量資料操作與分析，以從體驗中驗證課程所學。課程最後將透過實際分組演練真實數據，學習如何將巨量資料工具用在實際工作上。

四、適合對象：

- IT 經理或系統網路部門主管
- 專案經理、系統架構師或系統網路管理人員
- 企業或技術決策人員
- 對於大量資料處理、分析、應用之人員、分析師、主管

五、上課地點：

資訊工業策進會 數位教育研究所，台北市信義路三段 153 號 10 樓。

上課地點位於捷運大安站 1 號出口左轉約 20 公尺。

* 上課地點與教室之確認，以上課通知函為主。

六、課程日期：

2018/7/28~7/29 (週六日白天二天 9:00 ~17:00) · 共 2 天、計 14 小時。

七、報名方式：

本課程採線上報名，請至課程及報名網頁：<http://www.iiiedu.org.tw/ites/SPARK.htm>。

課程諮詢：(02)6631-6533 課程經理 吳小姐，或 E-mail：fionawu@iii.org.tw

報名確認：(02)6631-6535 郭小姐，聯絡 E-mail：florakuo@iii.org.tw

八、課程大綱：

課程單元	課程內容	時數
巨資平台簡介與安裝	<ul style="list-style-type: none">• 巨量資料生態系統簡介• OpenSource 巨量資料管理監控平台安裝• Spark 互動式開發環境建置• 資源調度管理框架介紹	2
Spark 技術實作與核心技術	<ul style="list-style-type: none">• Spark 的運行模式介紹• Spark 開發流程介紹• RDD Transformation 與 Action 方法解析與應用• Spark 進階用法與優化技巧• Spark 開源程式碼之範例設計模式解析	4
Spark 資料視覺化	<ul style="list-style-type: none">• Spark 資料運算與視覺化應用	1
Spark SQL 結構化資料處理應用	<ul style="list-style-type: none">• 將 RDD 與外部來源導入 Spark SQL• SparkSQL 解析與操作• DataFrame 解析與操作• 實際案例練習	3
Spark MLlib 機器學習函式庫	<ul style="list-style-type: none">• MLlib 分類模組使用• MLlib 分群模組操作• 文件探勘前處理模組使用	2
實際資料練習	<ul style="list-style-type: none">• 分散式 TFIDF 內部實作介紹• 媒體新聞摘要系統實作	2

*課程執行單位保留調整課程內容、日程與講師之權利

九、課程費用與繳費：

1. 本課程費用 NT\$14,000 元(含稅) · 費用含課程、講義、餐點。
2. 凡於 **2018/7/20** 前完成報名並填寫繳費資料者，享早鳥優惠 **NT\$9,800 元**(含稅)。
3. 團報優惠：二人團報可打 95 折、四人團報可打 9 折優惠。
 - 團報優惠與早鳥優惠可一併使用。

- 課程報名系統每個帳號僅提供一位學員報名，若為團報學員，每位皆必須完成線上報名，並於其他金額說明欄位註記「與 XXX 一起團報」，以利事後核對名單。

4. 請以信用卡或即期支票支付—

- ◆ 以信用卡支付者：請務必將正確信用卡資料填寫於繳費方式資料表。
- ◆ 以即期支票支付者：抬頭為「財團法人資訊工業策進會」，劃線並禁止背書轉讓，以掛號寄至台北市 106 信義路三段 153 號 11 樓，邱小姐 收。

十、報名確認與取消：

1. 已完成報名與繳費之學員，課程主辦單位將於開課三天前以 E-mail 方式寄發上課通知函；若課程因故取消或延期，亦將以 E-mail 方式通知，如未收到任何通知，敬請來電確認。
2. 已完成繳費之學員，若於開課前 8 天(含)內欲取消報名，請於開課日前以書面通知業務承辦人，主辦單位將退還 90% 課程費用。
3. 學員於培訓期間如因個人因素無法繼續參與課程，將依課程退費規定辦理之：上課未逾總時數三分之一，欲辦理退費，退還所有上課費用之二分之一，上課逾總時數三分之一，則不退費。
4. 本單位保留是否接受報名之權利。
5. 如遇不可抗拒之因素，課程主辦單位保留修訂課程日期及取消課程的權利。

十一、結業證書：

依本會 ISO 品質系統之「教育訓練服務程序」規定，參訓學員缺課未超過總時數五分之一者，結業時由本會核發結業證書。

十二、學習護照：

資策會為人事行政局首批認證審定之民間學習機構，參加本課程之中小企業員工，亦可獲得「中小企業終身學習護照」認可之時數；此外，公務人員參加資策會課程，學習時數可登錄「公務人員終身學習護照」。

十三、餐點服務：

本課程上課期間，每天都會提供午餐，學員不需自行準備或額外付費。

十四、開課人數：

最低招生人數至少為 10 人，預計開班人數為 25 人。