

中國文化大學

Big Data 系統建置與資料分析講師研習

實施計畫

壹、目標

"何謂 Big Data ?"

我們每天建立幾百 G 的資料 — 數量之大，光是過去兩年所建立的資料就佔當今世界總量的 90%。這些資料來源廣泛，像是用於搜集氣候資訊的感應器(物聯網)、社交媒體網站的貼文、數位圖片與影像、採購交易記錄以及行動電話 GPS 訊號等，不一而足，這類的資料就是 Big Data。

Big Data 含括四種層面：巨量、即時性、多樣性及真實性。

1. 巨量 (Volume)

Big Data 的特色就在於：龐大。企業資料包羅萬端，很容易便達到數兆位元組，甚至千兆位元組之譜。

2. 即時性 (Velocity)

Big Data 通常具有時效性，一旦串流至企業便須立即使用，方能發揮其最大價值。

3. 多樣性 (Variety)

Big Data 的範疇不僅止於結構化資料，還包含各類非結構化的資料：諸如文字、音訊、視訊、點擊串流 (click stream)、日誌檔等等。

4. 真實性 (Veracity)

指的是當資料的來源變得更多元時，這些資料本身的可靠度、品質是否足夠，若資料本身就是有問題的，那分析後的結果也不會是正確的

這 4 個資料特性，已經是現在式，而不是未來式。然而該如何解決日漸緊迫的巨量資料處理問題呢？像 Facebook、Twitter 這樣面臨資料量大爆炸的網路公司，開始用 Hadoop、NoSQL 等新興技術來解決問題。

"挑戰還是機會？"

Big Data 不只是一項挑戰，更是絕佳的機會，讓您能夠洞悉新興的資料類型、使企業運作更加靈敏並為過往所無法企及的問題提供解答。但在此之前，這種機會並無實際方法可以掌握。今天，Big Data 平台採用 Hadoop 等技術，能為充滿各種可能性的世界開啟一扇大門。

"Hadoop 技能有高度需求"

由於成長型企業在 Hadoop 和相關技術方面的興趣，同時也推動了企業在 Big Data 方面的專業人力需求，而資料分析師、資料科學家和資料管理師將獲得大鳴大放的機會。

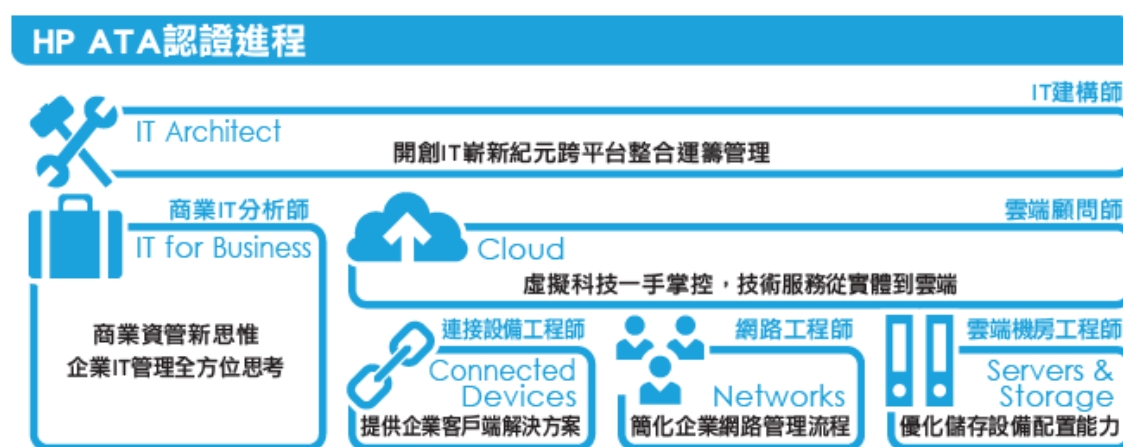
在紐約舉辦的 Hadoop World 會議上，分析師和 IT 經理人一再指出，採用 Hadoop 技術的公司面臨的主要挑戰之一就是技術可用性(skills availability)，並表示具有 Hadoop 技能的人同時也掌握了更多收入。

貳、研習活動課程內容

面對 Big Data 這波巨大資訊 (料) 的洪流，需要依靠新一代電腦機房(低排碳量)及 Apache 開源碼 Hadoop 專案，來當做驅動 Big Data 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。機房建置與佈署 Hadoop 分散運算系統，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。

1. 低排碳量 Big Data 電腦機房
2. 自造者 - Hadoop 樹莓小刀鋒
3. 貨櫃虛擬化技術 - Linux Container (LXC)
4. 多點 Hadoop 分散運算系統 (HDFS, YARN)
5. Pig 資料分析工具
6. Hive 資料倉儲工具

除了掌握 Big Data 關鍵技術, 研習中還會輔導學員取得 HP 伺服器與儲存管理(Servers and Storage)國際證照.



- HP ATA 認證說明網址：

<http://www.certipoint.com/PORTAL/desktopdefault.aspx?tabid=693&roleid=101>

- 報考雲端管理(Cloud)認證，需先取得 端點連接設備管理(Connected Devices)、網路管理(Networks)及伺服器與儲存管理(Servers and Storage)認證。
- 即測即評：線上測驗 HP ATA 國際認證。

參、實施對象：

- 一、高中職、大專院校教職員，或對於 Big Data 有興趣之業界人士。
- 二、名額 30 名，依報名時間順序額滿為止逾期恕不接受報名。

肆、研習活動與報名資訊:

- 一、研習費用：NT\$ 1000
- 二、報名繳費網址：<http://goo.gl/WiInf5>
- 三、研習地點：中國文化大學推廣教育部 302 電腦教室
(地址：台北市大安區建國南路二段 231 號)
- 四、研習日期：2015/01/16 (星期五) 13：00~18：30
- 五、研習時數：全程參加研習教師核予研習時數 6 小時。

時間	上課內容
13:00~16:00	<ol style="list-style-type: none"> 1. Big Data 機房建置與管理 2. 說明冷卻管理技術和理念及物理空間佈局要求 3. 機房消防 (獨立防火區規劃, 氣體滅火系統,..) 4. 機架式伺服器 (機器散熱, PDU, ...)
16:00~17:40	<ol style="list-style-type: none"> 1. 自造者 - Hadoop 樹莓小刀鋒 2. 實作多點 HDFS 分散檔案系統 3. 實作多點 YARN 分散運算系統 4. Pig 資料分析工具 -人力資源調查失業率 (XML) 5. Hive 資料倉儲工具 -分析大專校院校別學生數 (TXT)
17:40~18:30	HP Server and Storage 國際認證線上測試 (50 分鐘)

伍、其他

- 1 通過線上即測即評認證考試者，可取得 HP ATA 該科原廠國際認證。
- 2 HP ATA 認證參考資料：

(英文) http://www.hp.com/certification/hp_institute_portfolio.html

(英文)

http://www.certiport.com/Portal/desktopdefault.aspx?page=common/pagelibrary/HP_Student.html