

Big Data 移動實驗室 BD-101



大數據時代是近年最為流行的詞語之一，越來越多的行業開始採用大數據分析的結果來制定未來的策略走向，因此大數據相關人才需求也日漸上升。BD-101 大數據移動實驗室是針對大數據相關人才培訓所設計的，豐富及全面化的培訓課程使學生們可以有效且快速的學習大數據相關技能。

● 什麼是大數據？

大數據是由各式各樣龐大的數據集合所組成的。在過去，大數據通常用於企業內部的資料分析或是統計應用，僅止於分析過去的歷史資料。時至今日雲端和物聯網環境發展成熟、儲存設備成本下降、資料量急速增長以及軟體技術的進步，這些條件的滿足使得我們可以從過去的歷史資料分析進階到預測未來並且提高準確性。大數據已經不在只是資料處理，而是前所未有的思維和商業模式。

● 大數據的特性

Volume 資料量

無論是金融服務、能源管理、生物醫療、多媒體社群，每分每秒都在產生龐大的數據。

Velocity 即時性

使用者每秒都在產生大量的數據，當資料送到伺服器時便立即進行分析，即時得到結果後立即做出修正，發揮資料的最大價值。

Variety 多樣性

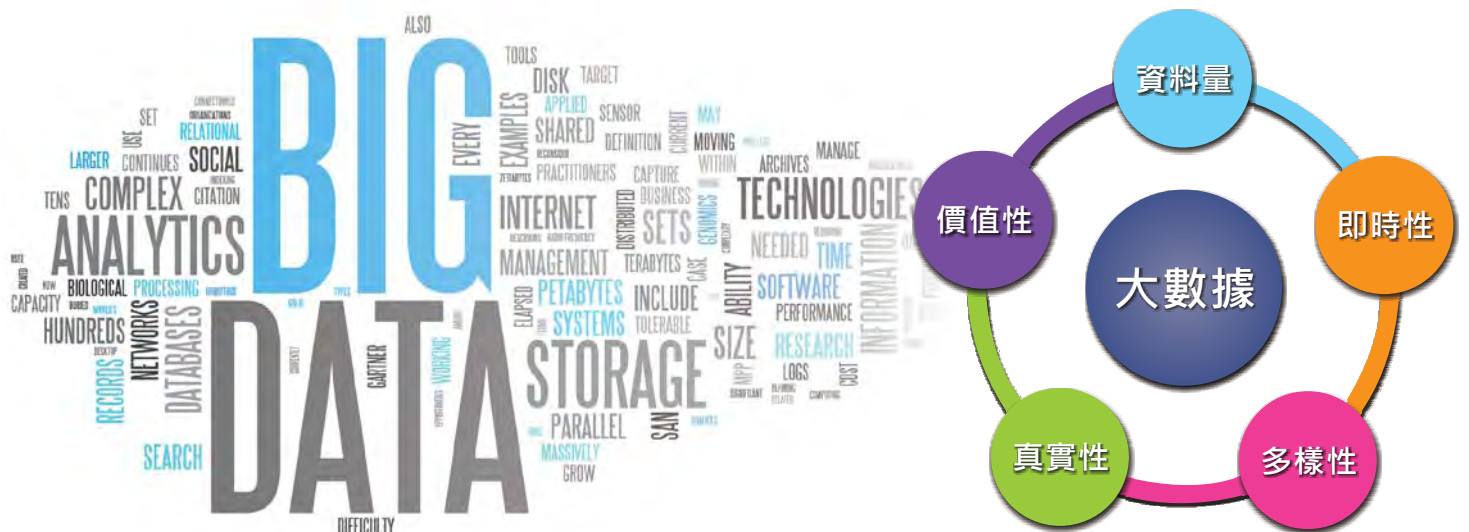
多樣化的資料包括文字、位置、語音、影像、圖片...等結構化與非結構化的資料，彼此間能進行互 分析，尋找數據間的關聯性。

Veracity 真實性

資料來源是否正確，即使是真實資料，是否有準確的紀錄，資料中有沒有異常...等。錯誤的資料來源會造成分析結果的誤差，甚至影響到預測的準確性，因此確保資料來源的真實性也是大數據的重點之一。

Value 價值性

大數據最大的價值在於從龐大的數據中，挖掘出對未來趨勢有價值的數據，並通過人工智慧、機器學習進行深度分析，提高準確性。



大數據的基礎 - Hadoop

Hadoop 是開源的軟體框架，它成功解決了檔案存放、備份及資料處理...等問題，因而被廣泛應用，成為大數據的主流技術。

大數據的資料處理 - Spark

速度在大數據的資料處理中非常重要，Spark 的一個重要特性是其可以在記憶體中運算，這使得 Spark 在資料分析及運算的速度比 MapReduce 更有效率。

大數據的資料擷取 - Python

Python 是通用於各種領域的程式語言，它能透過低成本、自動化的方式由網路中爬取大量且有效的資料，強大的數據處理能力是 Python 成為分析大數據時重要程式語言的主要原因。



● BD-101系統特色

1. 獨立性：可直接使用，不需要網路連接或額外的軟硬體安裝。機櫃式設計，易於在空間內移動。
2. 方便性：快速故障排除功能提供系統還原功能，使用者不須花費大量時間進行故障排除。透過 6 種不同模型的隨機資料生成，可以一鍵生成出符合演算法資料集的資料。
3. 擴充性：可運用於不同的專題研究與大數據實驗，也可結合 IoT 設備，儲存、分析各種感測器資料，實現跨領域的真實應用。
4. 豐富性：提供全面的大數據生態系統培訓課程
 - (1) 9 種不同的演算法，20 多個經典大數據範例
 - (2) 大數據生態圈包括 Hadoop、Yarn、Spark、Hive、HBase...等各種工具的運用訓練。

● BD-101學習主題

1. 資料清洗、正規化與標準化
2. 大數據生態圈架構與組態設定
3. 不同資料庫對比
4. 運用各種演算法提取、儲存、檢索和分析大數據
5. 大數據如何與 AI 結合

