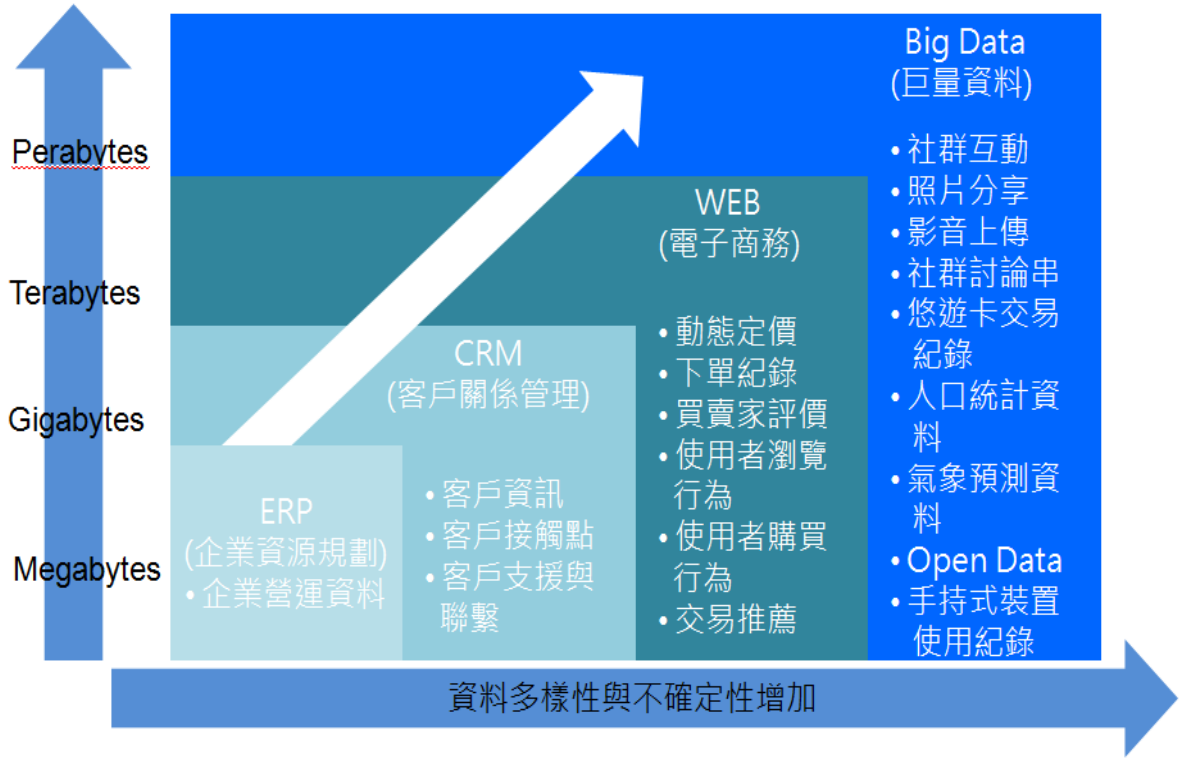
關貿第 46 期電子報

哈燒話題: 巨量資料新生活

Big Data 爲 2013 年台灣 IT 產業最熱門的名詞之一，在數位化的潮流之下，「巨量資料」分析已成爲創新競爭力的一大關鍵，也是世界各先進國家與企業投入研究的重要領域。

資料分成結構化資料、半結構化資料、非結構化資料三種型式。結構化資料是指資料經過分析後可分解成多個互相關聯的組成部分，各組成部分間有明確的層次結構，其使用和維護通過資料庫進行管理，並有一定的操作規範。我們通常接觸的資料庫庫管理的資料，包括生產、業務、交易、客戶資料的記錄都屬於結構化資料。非結構化資料，資料格式不固定，常常是各種形式的資料，例如圖像、聲音、影片等資料皆屬於非結構化資料。半結構化的資料介於結構化資料和非結構化資料之間，資料格式以文字爲主，但每個欄位填入資料的內容和長度則不固定。電子郵件、網路社群討論文章皆屬於半結構化資料。下圖爲商業智慧分析(Business Intelligent)到巨量資料分析(Big Data)的過程，從圖中可以發現，越往右上分析資料多樣性、不確定和資料大小亦逐步增加。且商業智慧分析(BI)主要分析企業內部營運和客戶資料，巨量資料分析(Big Data)分析的範圍更加廣泛，舉凡社群互動、悠遊卡交易紀錄、手持裝置使用紀錄皆在分析的範圍。



(圖形來源: TERADATA 資料內容: TERADATA & 關貿網路整理)

Big Data 應用的目的是在預測未來

目前全球最大的資料庫就是網際網路，想要了解消費者的真實需求和想法，到網路上「撈」資料，是最精準快速的方法，「了解使用者想什麼」，正是臉書（Facebook）和谷哥(Google)身價持續飆漲的關鍵。不過，Big Data 所歸屬的商用智慧（Business Intelligence，BI），已行之有年，分析大量「既有」資料的作法，並非新鮮事；如今 Big Data 累積龐大的使用者資料，將影音、照片等非結構化資料加以分析，並加入社群討論字串，此兩項新元素讓既有的 BI 產生空前變化。因此巨量資料(Big Data)應用和商用智慧(BI)最大的不同在於，Big Data 分析的目的是在預測未來。

Big Data 應用的實際案例

一、智慧行銷: 品牌聲譽分析

掌握心佔率，進而得到市佔率。隨著網路的發達，一般民眾的消費行爲也逐漸改變，網路成爲人們表達意見的主要管道，除了在網路上發表評論，也開始搜尋其他網友們的討論和建議作爲購買或評斷的決策參考依據，網路輿情因此比以往產生更大的影響力，進而影響民眾的消費行爲和生活滿意度。透過網路輿情分析，企業能了解消費者對於產品或品牌的意見和想法，政府也更能貼近民眾的心聲。由於網際網路使用人數快速增長，網路輿情的重要性也逐漸增加，透過網路輿情行銷，除了能評估產品的市場喜好、競爭對手的概況，政府政策的推行效能，也能隨時掌握企業、品牌在網路上的影響力，建立市場的滿意度、增加產品市占率和獲得民意支持。

二、數字三角洲：國際水資源管理

荷蘭爲一填海造陸著名的國家，其中 66%的人口居住在容易淹水的區域。每一個與水有關的事件都是至關重要的，並且還會影響到商業活動、農業和普通民眾的日常生活。每一年荷蘭花在水資源管理的預算高達 70 億歐元，荷蘭全國共築長 16,496 km 的海堤以防止海水倒灌，每年因爲水災造成的農業損失爲 400 萬歐元，因爲超抽地下水造成地層下陷，進而引發基礎設施維護的費用高達 50 億歐元。[註一]

因此荷蘭政府痛定思痛，於 2013 年推出一項數字三角洲計畫，旨在爲治水專家提供一個即時的智慧監控平台。此計畫將整合多個政府和研究單位的水資源監測資訊，包含 Rijkswaterstaat(荷蘭基礎設施和環境部)、代爾夫特大學、代爾夫蘭地方水資源管理部門以及 Deltares 科學研究所。這些機構將負責分析整合荷蘭 100 多個與水管理相關的不同項目數據，包括降水量預測、水位和水質監測、堤防感應器、雷達偵測數據、模型預測值，以及當前和歷史的維護數據，包括自來水閘、泵站和閘壩的數據以及通過這些數據加快水管理創新進度的可行性。另外，此計畫亦整合海堤的衛星空照資訊。透過衛星資訊可以監測到一個堤防的其中一部分（100 公尺）正在下陷，那麼堤防監測

單位則可以在這個堤防下陷的部分放置感應器，而不是在整個堤防放置感應器，如此以達到節省成本的目的。

透過即時的智慧監控平台掌握水資源管理相關訊息，可促進各大組織和機構間的資訊分享。水資源管理專家可於第一時間採取最正確的因應方案，包括蓄水、從低窪地區分流積水、避免海水入侵飲用水、水污染防治等方面的工作。另外，透過結合氣象數據、水系統模擬模型以及河川水位測量數據，可以研發出一個靈活且具備彈性的早期洪水預警方法。透過數字三角洲平台，荷蘭政府預估每年可省下 20%~30%的水資源管理經費，並且大大的保護民眾生活、農業和北歐運輸業的正常運轉。

Big Data 的真正價值操之於人

講了許多的 Big Data 專業知識和應用，讓我們回歸到 Big Data 的本質。Big Data 應用的培養關鍵在於人。因爲，真正擁有智慧的是透過分析平台做出判斷的決策者。機器可以很快很有效率地處理非結構化以及結構化資料，透過語意分析和情緒分析等工具讓我們看到消費者、使用者喜好和習性，協助我們找到過去不曾發現的問題。但是，真正擁有智慧的仍然是接著做出判斷的決策者。

施拉吉（Michael Schrage）在哈佛商業評論部落格裡寫道，「太多組織尚未明白，若要以巨量資料爲運作基礎，人員判斷比雲端機器學習更重要，」應用程式無論多麼易於使用、功能多麼強大，仍無法取代人類理解能力。

所以人類做爲智慧平台終端用戶，應當思考如何透過機器有效的處理並且分析大量的資料，透過視覺化報表的呈現消除數字與商業價值之間的落差。而今，商業世界正邁向由雲端技術、巨量資料和機械學習所建構的未來，亦唯有人能充分運用這些新技術並展現其商業價值。

資料來源:

1. IBM Digital Delta:
URL:http://www-03.ibm.com/press/us/en/pressrelease/41385.wss
Accessed 2/17/2014

新聞播報：關貿網路 CSR

關貿網路爲善盡企業社會責任(CSR Corporate Social Responsibility)於今(2014)年持續認養南港車站站前的南興公園，關貿網路員工並自發性組成志工隊並派員巡視公園狀況，並配合南港區公所安排的時間與區公所技工及清潔人員進行公園打掃與整理。

南興公園於民國 67 年設置，位於南港火車站前，南港行政中心對面，形狀呈矩形，面積約 4,801 平方公尺。公園內遊客平時來往頻繁，午間常有民眾在園椅上打盹。南興公園爲南港區較早開闢之公園。台北市政府民政局特別於民國 102 年將南興公園重新改建，讓這個具有 40 幾年歷史的公園徹底改頭換面。

新的南興公園，沒有一般傳統公園的缺失，採取了開放空間及明亮的設計，增加樹叢間距也降低蚊蟲聚集，引進了無障礙空間的設計讓整個穿越動線更爲順暢，簡潔的休憩空間擁有更好的空間視野，加上下凹式的表演廣場以及兒童遊戲區，讓南興公園煥然一新。

此外公園是「都市之肺」，提供一個充滿綠意、空氣清新、休憩調養的功能，公共場所禁菸已是全球趨勢，美國加州、明尼蘇達州陸續推動無菸公園政策，連鄰近我國的香港也於 2005 年起規定公園、海灘全面禁菸。

故南興公園也是臺北市政府所擇定試辦的無菸公園，爲推動無菸環境的起始點，建立一個未來其他區域推動時之參考典範，關貿網路也贊助兩面「無菸公園」的宣導牌，分別設置於入口處及兒童遊戲區，希望使用南興公園的民眾能配合與諒解。

此外關貿網路受南港區公所委託在 facebook 上建立南興公園 FB 粉絲團，這是全台公園的首創，歡迎一起來參觀及按讚!

南興公園 FB

https://www.facebook.com/?ref=home#!/pages/%E5%8D%97%E8%88%88%E5%85%AC%E5%9C%92/155216177879943

資安小常識：公司被駭客攻擊了嗎?!?!?!

前陣子新聞報導 F 公司遭駭客 DDoS 攻擊，3.5 小時內高達 82 億次，導致 APP 系統癱瘓，經過資安辦的調查，發現事實卻非如此，除此之外，還陸續被爆料出機密資訊外洩、繞過圖形驗證碼等攻擊事件，到底發生了什麼事情呢?

何謂 DDoS? DDoS 攻擊乃駭客透過殭屍電腦，在指短時間之內，讓目標伺服器遭受來自不同來源 IP 的大量連線，使目標伺服器與網路設備無法負荷，進而癱瘓該項服務。此事件經過資安辦調查，所謂的駭客其實就是無法連上官網的 200 萬名用戶，因程式設計不良，APP 若連不上線，會不斷嘗試自動連線，幾百萬名用戶就這樣不斷地重新連線，才導致 F 公司被自己客戶 DDoS 的局面。

機密資訊部分外洩則是被稱做 Directory Traversal (Local File Inclusion) 的駭客攻擊模式，導致可未經授權存取設定檔、使用者帳號密碼清單等。此弱點發生的原因是利用檔案上傳、下載或是讀取檔案的參數未過濾危險字元以及限定特定目錄，同時，Server 上未限定 web service 權限。 攻擊者透過輸入跳脫資料夾的字元，如: linux 的../ 、windows 的..\，跳脫原本的資料夾拿到上層或是根目錄的權限，進而取得所有檔案。

F 公司採用驗證碼 + 密文爲驗證碼方式，將驗證碼的圖片檔及其相關驗證碼資訊「存到伺服器」，再把驗證碼的「密文」交給使用者的瀏覽器網頁表單保管，每次驗證碼產生的「密文」與驗證碼「輸入值」都是一對一對應，產生方式都是不變的，因此僅需取得一組 驗證碼 + 密文，圖形驗證碼的防暴力密碼解機制等於無效 。

由此案例可發現很多資安事件發生在於開發過程的疏忽，以及權限控管不當，在開發過程中就已經埋下的漏洞，並非皆爲駭客的蓄意攻擊。程式開發者應適時學習與程式開發相關資安弱點，才能在開發過程中避免重蹈覆轍。建議在系統上線前，應建立完善的上線機制，如：執行網站弱掃、滲透測試和原碼檢測等作業，在上線前發現問題，就可以降低駭客攻擊成功的機率。