



## 雲報專欄：巨量資料的機會與挑戰—余孝先 工研院巨量資訊科技中心主任/雲端技術專家委員會 副主任委員

### 巨量資料分析的核心價值：用資料分析的結果做決策

巨量資料(Big Data)分析的目的及其核心價值何在？簡單講就是用資料分析的結果來做決策。千百年來，大至國家、企業，小至家庭、個人，其實每天都在做決策。決策的依據可以是學理，可以是經驗，可以是直覺，當然也可以是資料分析的結果。以零售業為例，傳統的柑仔店，資深的老闆可以由過往的經驗，如季節的變換、節慶的到來等多種因素，來判斷哪些貨品應逐步上架或下架。現代化的連鎖便利商店則可以用資料分析的方式來做決定，所考慮的因素可以更多，決策的頻率可以更高，決定的事項可以更詳細，通常效果也更好。

根據資料分析的結果做決策並不是一個新的觀念，例如商業智慧(Business Intelligence)，資料探勘(Data Mining)均屬之。所以巨量資料到底新意何在？或說 Big Data 跟 Data Mining 差異何在？

### 巨量資料的新意所在

巨量資料近兩年吸引了許多人的注意，也有許多新創公司爭相投入，其根本原因在於由於科技的進步，尤其是網際網路應用服務的興起，如 Twitter, Facebook, e-commerce 等；以及各式各樣感測儀器(sensor)的普及，如 GPS, RFID, 監視攝影機等。這些新興事物大幅度地豐富了資料的內容，也大幅度地增加了資料分析的困難度，使得傳統的資料分析工具與方法不再完全適用。這些改變造成新的挑戰，同時也帶來新的機會。

資料分析對象這幾年的重大改變，可用四個 V 的概念說明。首先是 Volume，也就是處理的資料量。很多研究都指出，人類活動所產生的資料量，每年將以增加 50% 以上的指數性成長，導致過去所使用的方法已無法消化這麼多的資料，因而必須產生新的資料處理系統。

第二是 Velocity，就是資料產生與處理的速度。以往我們分析如「大賣場顧客的採購行為」等資料時，可以用幾個星期甚至更長的時間來分析過去所累積的





資料，然後據以做進貨及促銷等決定。現在因為科技的發展，例如網際網路，有太多的資料是不斷的產生，必須被即時的處理與即時的回應。例如要偵測網路上(如 Twitter)的熱門議題時，由於 Twitter 上的訊息是隨時不斷地產生，也就需要即時的處理。

第三是 Variety，就是資料的多樣性。以前的分析資料多為有結構的資料，像是存在資料庫系統裡的資料，或是 Excel 檔案裡的資料等都是。但其實世界上大部分的資料都沒有結構性；例如是一本書的內容、一個 YouTube 多媒體資料、Facebook 上的一則消息。這些資料的分析，跟以往傳統資料分析的差別相當大。有研究指出，我們要面對的資料，有 95% 都是屬於這些非結構性的資料。傳統的資料分析方式對處理結構性資料比較得心應手，對非結構性資料就會左支右絀；因此眼前的挑戰，即是需要有新的分析方法，以把其餘 95% 的資料都納入討論。

最後一個 V 是 Veracity，指的是資料的可信程度。以往資料因來源一致，正確性高，但現在的資料因來源複雜，正確性和可信度都降低，例如：在網路上搜尋某產品的評價時，無法得到單一意見，而是正反兼具，彼此衝突，在這些意見中甚至可能含有許多非真實、惡意的評論，如何分辨何者為真正有價值的資訊，也是一個很大的挑戰。

綜合以上概念，我們知道，巨量資料分析有別於過去資料分析之處，就在於資料量、分析速度、資料多樣性與資料可信度四者，在新的挑戰之下，新的技術因應而生，誰能克服挑戰，就能抓住機會。

### 巨量資料分析應用 台灣產業競爭加值

巨量資料分析研究，台灣才正開始要推動，而美國與其他國家的政府、企業已經有了許多值得借鏡的經驗。當初為了找出台灣推動巨量資料處理的切入點，工研院研究了許多資料後，決定的第一個策略是由應用切入。台灣多年來都被詬病是 ICT 製造大國，卻是應用小國，所生產的 ICT 設備主要是出口幫別的國家提升競爭力與附加價值。硬體設備業是賺了錢，但其他產業的國際競爭力很多卻是相對減弱。有鑑於此，工研院推動巨量資料的目標就不是在設備上著墨，而是





要把台灣推成巨資應用大國，讓各行各業與政府部門都能提升效能與附加價值。

經過充分的討論與請教專家的意見，我們分別就台灣的製造業、服務業與政府開放資料三者規劃了切入重點。製造業的部分，我們選擇台灣特色製造業為優先，例如半導體業與機械業。資訊硬體業因為製造已全部外移，是否是個好標的還須討論。

服務業的部分則以零售與銷售業為主，因為台灣的便利商店密度全球第一，就業人口一百多萬，佔台灣就業人口 10%，其經營效率之良窳對台灣影響很大。台灣網際網路上的零售業，如博客來等，在與超商合作前後，其發展相去不可以道里計，顯示出經營內容雖不變，但引進新的元素就可能產生巨大的效益。我們認為巨量資料的技術，就是下一個影響深遠的新元素。資料分析做的越精確、考慮到更多的面向，就能更精準掌握顧客的需求，除了提升顧客的滿意度，也可指引產品的研發方向，並降低庫存的浪費。尤其對跨國界的零售、銷售業而言，巨量資料技術將會成為其增進國際競爭力之利器。

除了產業界，政府效能的提升對於全民的幫助更是全面。近年行政院推動政府雲以及開放資料計畫，把政府運作過程中所產生的許多資料，如交通資訊、工商登記等開放出來，讓民眾可以使用與創造價值。工研院也計畫透過巨量資料技術分析這些政府開放資料，挖掘出其中潛藏的價值。

### **推動軟體與服務，不能坐而言，必須起而行**

目前工研院巨資中心雖然決定以應用服務做為推動方向，但也有許多人其實並不看好，認為還是該回到做設備才是台灣的強項。我想這是在台灣的產業結構下，多數人熟悉的還是硬體，應用服務的接觸較少，總是心有疑慮。平常談到產業發展，人人都說不能偏於硬體，應多注意軟體與服務；但一旦要真的投資推動時，又很自然地回到熟悉的硬體設備項目，對於摸不著、聽不懂的軟體與服務還是猶豫不決。長期以來的現象是：投資十億開一條新的生產線，比投資一億做軟體服務還要容易做決定。我想這是大家都有責任改變的現象。台灣的 ICT 產業在多年來一波又一波的服務、應用新浪潮，幾乎都是缺席的狀態，巨資中心會盡一己之力，從人才培育、技術研發、應用推展都努力推動。希望台灣未來除了硬體







設備之外，軟體與服務的部分也能有好的發展。

