

一次搞懂大數據（上）

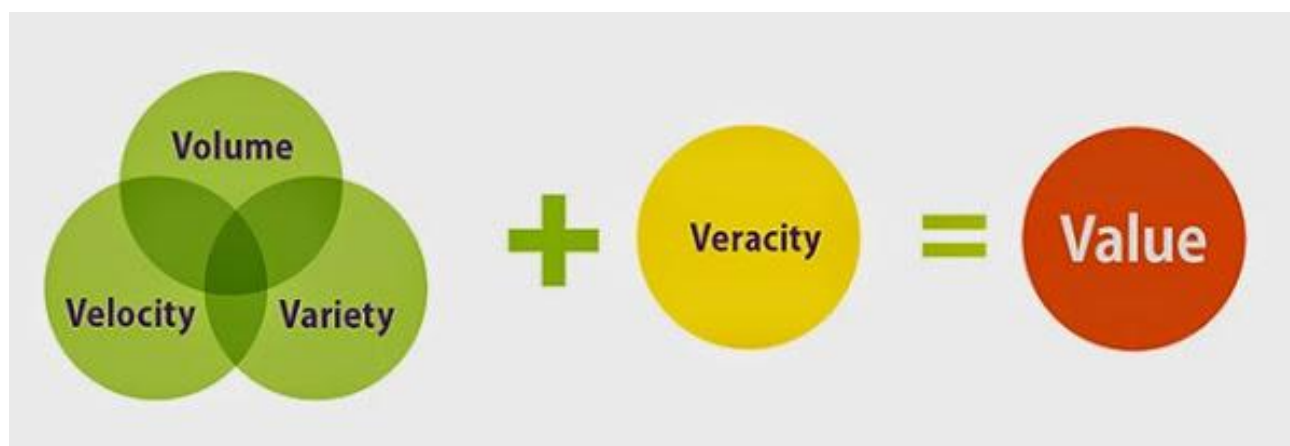
資料來源：數位時代 / 撰文者：李欣宜

當大數據充斥各種場合，從馬雲到釋昭慧都侃侃而談，你還能不懂什麼是大數據嗎？你也許已經聽過無數的大數據神話，但對於大數據仍停留在一知半解階段，《數位時代》專訪各大數據專家，整理出你最應該知道的大數據十問。

Q：大數據是什麼？

A：大數據（Big Data）又被稱為巨量資料，其概念其實就是過去 10 年廣泛用於企業內部的資料分析、商業智慧（Business Intelligence）和統計應用之大成。但大數據現在不只是資料處理工具，更是一種企業思維和商業模式，因為資料量急速成長、儲存設備成本下降、軟體技術進化和雲端環境成熟等種種客觀條件就位，方才讓資料分析從過去的洞悉歷史進化到預測未來，甚至是破舊立新，開創從所未見的商業模式。

一般而言，大數據的定義是 Volume（容量）、Velocity（速度）和 Variety（多樣性），但也有人另外加上 Veracity（真實性）和 Value（價值）兩個 V。但其實不論是幾 V，大數據的資料特質和傳統資料最大的不同是，資料來源多元、種類繁多，大多是非結構化資料，而且更新速度非常快，導致資料量大增。而要用大數據創造價值，不得不注意數據的真實性。



大數據和商業分析之異同		
	傳統商業分析	大數據分析
資料分析方法	抽樣分析	使用大數據的原始資料 (raw data) 進行統計分析、視覺化、產出報表
可預測性	制定模型以跑出期望中的數據結果	大數據最重要的，在於探索資料、發現新模型、找出資料間的關聯性，進而達到精準預測
所需耗費時間	透過資料庫正規化、視覺化，費力耗時、學習門檻高	可即時支援業務單位進行市場策略調整與配合
使用技術	資料庫、正規化、結構化基礎的分析	原始資料、大數據、非結構化和結構化分析、分散式儲存、分散式運算

資料來源：Splunk

Q：為什麼需要大數據？

A：因為當從人到機器都已經被數據解構，數據不僅僅是歐巴馬口中的石油或是黃金，它更是血液，貫穿每個人一生中每個生命階段。這並非危言聳聽，更不是科幻電影，而是正在逐步成真的現實。

例如有一款叫做 [Ovia Fertility](#) 的 App，藉由分析 30 萬名會員的數據，開發演算法，精準計算排卵期，提高懷孕的機率，這個 App 已幫助 5 萬名會員成功懷孕。又比如 [Workday](#) 推出一套軟體，預測員工的薪水漲幅和可能跳槽時間，幫助企業決定每名員工的加薪幅度、時間點和轉職時機。理財也逃不過大數據的掌控，騰訊就於年初推出第一家用大數據決定借貸與否的銀行，[微眾銀行](#) 結合辨識人臉和公安部門資料，決定借貸者的信用等級。

從懷孕生子、工作到理財，大數據將全面影響每個人與每家企業。對企業而言，大數據可望提升服務品質、增加管理效率、幫助決策和創造商業模式；對一般民眾而言，大數據是另一個自我，它可能比本人更了解本人，為你預先解決每個未知，當一切都開始數據化，你能夠不需要數據嗎？

Q：大數據一定要很大嗎？

A：雖然大數據的狹義定義是，資料量要在 100TB 到 PB 之間，但其實絕大多數的企業，都不符合這個標準，大企業如 eBay、亞馬遜或 AT&T 或許符合大數據的標準。但其實資料量只是大數據的其中一個面向，大數據揭示的是一種「資料經濟」的精神，而非只是「大」。

「大，是大數據中最無趣的部分。」天睿資訊(Teradata)首席技術長寶立明(Stephen Brobst)認為，企業真正要尋找的是非傳統的、而且未曾被挖掘過的資料，並且從這些資料中去提煉出價值，這才是對大數據應有的正確認知，而非只是執著於資料大小，只要能從看似毫無意義的數據礦坑中挖掘出金礦，有誰會在意那座礦坑原本是大得像座山還是小得像狗屋呢？和沛科技創辦人翟本喬就指出，大數據這個名字容易讓人誤導，因為真正重要的其實是大智慧。大數據不只是說資料量有多大，速度快和資料量大都可以用技術輕易解決，但種類(Variety)比較需要智慧。

Q：沒有大數據就不能用大數據嗎？

A：非也，建置大數據架構與環境的確所費不貲，一般中小企業通常無法輕易投入鉅額成本，但大數據時代的精神在於如何妥善利用既有或非傳統資料，從中挖掘出新商機，因此即使是中小企業甚或者是新創企業，都能在大數據時代用「大數據」。

就技術面來說，現在有許多業者開始提供建置成本較低的大數據處理工具和雲端系統，有些甚至跟 App 一樣，只要根據自身需求挑選需要購買的功能即可，例如科智提供的工業化數據管理工具即為一例。另一方面，很多時候中小企業其實不需要建設大數據系統。中研院資訊科學研究所研究員陳昇瑋即指出，在絕大多數情況下，大數據專案其實不需要建置 Hadoop 系統，特別是台灣的社群媒體沒那麼發達，而是直接採用國外的居多，資料都不在自己手上，與其盲目追求技術和工具，不如先用小量資料去驗證一個概念，是否能將資料轉換成商業機會，再來決定要不要建置大數據的作業環境。

大數據領域權威麥爾苟伯格(Viktor Mayer-Schönberger)在《大數據》一書中便提及，大公司有巨量資料的規模優勢，但小公司有成本及創新上的優勢，小公司因為速度夠快、靈活度高，就算維持小規模，還是能夠蓬勃發展。

Q：我要怎麼開始進行大數據專案？

A：設置專門統籌大數據專案的部門和職銜是第一步，而且層級越高越好，企業領導人必須足夠正視大數據的力量，才能帶動整個組織重視數據的文化。Etu 負責人蔣居裕便指出，大數據其實是管理問題，而非技術問題，缺少跨部門協作，大數據專案很難有個美好的開始。第二步，切勿陷入大數據迷思，與其急著想用數據變現，不如先回頭看看自己企業內部的問題為何，先定義問題，再試圖用數據找解方。阿里巴巴集團副總裁車品覺建議，與其整天想著大數據，不如先整頓自己企業內部的數據，很多時候光是企業內部的數據就問題叢生，不同部門之間的數據無法相容，「整個數據在一個中小企業裡面也是四分五裂，在這個地方沒做好的情況下，居然說你想用大數據，其實是有點難以理解。」