



FIGURE

圖 8-5 能力性知識地圖

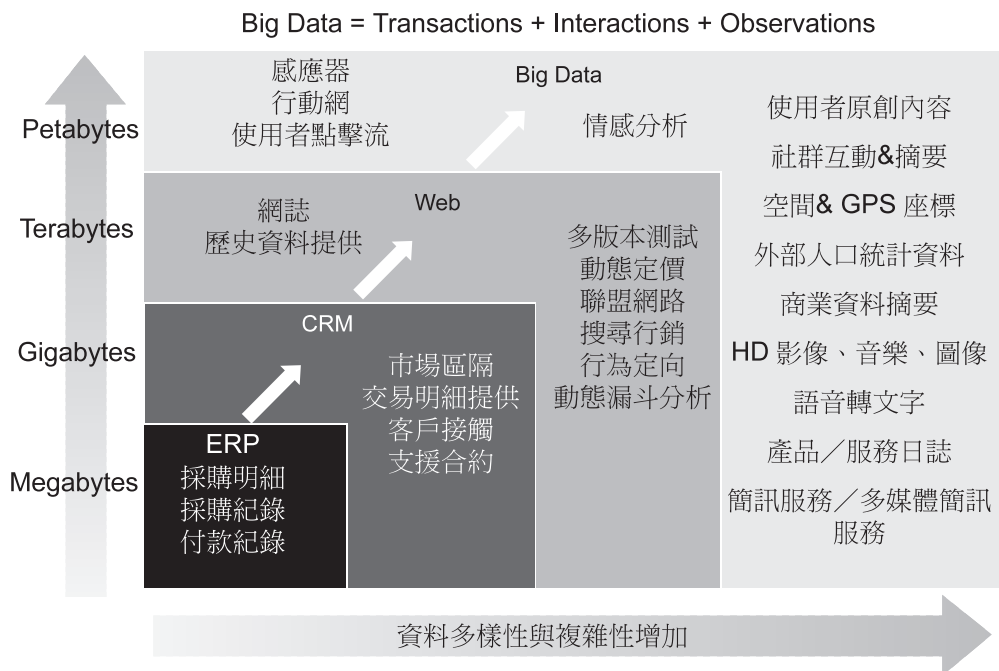
非將知識「推」(Push)向使用者。

8.3 大數據

從個人電腦的發明、網際網路、社群媒體的普及到雲端的應用，電腦科技和人們的關係愈來愈緊密。根據 eMarketer 2018 年 11 月報告預測 2019 年全球約有 37.5 億網際網路使用者，其中約有 86.2% 會使用手機上網(eMarketer, 2018)。另外，12 月報告亦指出，2018 年 12 月全球網路使用者達到 41 億，亞洲地區為網際網路使用人口排序第一，約占全球 49% 的比例，歐洲地區排序第二，約占 16.8% 的比例(eMarketer, 2018)。麥肯錫(McKinsey)在 2011 年發表《大數據：創新、競爭和生產力的下一個新領域》報告中，也明白揭露大數據(Big Data)將成為競爭的關鍵性基礎，並成為下一波生產率提高、創新和消費者創造價值的支柱。隨著網際網路的普及與雲端應用的加持，2012 年開始，Big Data 成為備受矚目的新概念、新機會，且在瑞士達沃斯論壇上（世界經濟論壇），一份題為《大數據，大影響》(Big Data Big Impact)的報告宣稱，數據已經成為一種新的經濟資產類別，是未來重要的生產要素。數據在過去被視為無用的積累，現在卻被視為資產，這是大數據時代的第一個變革，且令人驚喜的發現是數據不像機器、廠房一樣，會隨著使用次數的增多而折舊；相反的，重複使用、不斷雕琢，它可能增值，大數據的重要性不在於數據資料有多少，而是

人們如何應用軟硬體與技術，從各種數據中找出線索、趨勢以及商機。

現今大數據所涉及的資料已經和過去的資料不同了，過去的資料大部分是人工手寫下來的交易紀錄，現在則是以機器來記錄交易資料包含人們跟企業間的互動數據、以及網頁跟連結的點擊紀錄等，皆由機器自動產生與累積出後續的觀察數據，例如智慧型家居產品能記錄室溫變化等(Inside, 2015)。Connolly(2012)認為「大數據」是由**交易數據(Transactions)**、**互動數據(Interactions)**與**觀察數據(Observations)** 3 種資料型態所組成，並提出數據演進示意圖（如圖 8-6 所示）。圖中，縱軸表示資料量，橫軸表示資料之多樣性與複雜性，並闡述從 ERP、CRM、Web、以及 Big Data，其資料量和資料之多樣性與複雜性持續增加。而 ERP、SCM、CRM 與 Web 應用是典型交易處理的系統，這些系統中高度結構化的數據通常被儲存在 SQL 資料庫中。互動數據則是有關人與物彼此的互動或是業務交流，例如 Web 網誌、使用者點擊流、社群互動與回饋，以及使用者產生的內容等，都是屬於互動數據。觀察數據一般來自於物聯網，諸如來自於熱、動作、壓力與 RFID 感測器，以及行動裝置中的 GPS 晶片，甚至飛機引擎等物件也能產出觀察數據。



FIGURE

圖 8-6 數據演進示意圖

資料來源：Content of graphic created in partnership with Teradata, Inc.

網際網路數據中心(Internet Data Center, IDC)認為大數據技術描述了一種新一代技術和架構，以更經濟的方式與分析技術，快速地捕獲、發現數據，其目的在於將數據轉化為知識，探索數據的產生機制，進行預測和政策制定等。麥肯錫(McKinsey)認為大數據係指其大小超出一般典型數據庫軟體的蒐集、儲存、管理與分析等能力的數據集，且需要具備數據儲存能力、處理能力與查詢能力。當具備此三種能力後，其落實的價值大致可從兩方面來說明：(1)對內幫助企業發展、進行決策支撐，讓企業能更高效率的制定決策，支持行銷管理工作、對目標客群進行精準行銷等；(2)對外開拓數據的長尾效應，對數據進行整合與深度挖掘，並拓展新業務的機會。

IBM 將大數據的特徵歸納為 4V(Volume, Variety, Velocity, Veracity)。其中，**數據量(Volume)**說明了用戶在滑鼠的點擊、瀏覽觀看、接收訊息、網路搜尋、甚至線上交易的同時，也正在產生與累積龐大的數據，因此數據量很容易就能達到數兆位元組(Tera Bytes, TB)、千兆位元組(Peta Bytes, PB)，或是百萬兆位元組(Exabytes, EB)的等級。**數據類型多樣性(Variety)**說明數據來源包羅萬象，若是最簡單的方式來分類的話，可以分為結構化數據、半結構化數據、非結構化數據。**數據的傳輸速度(Velocity)**說明數據的流動是連續且快速的，隨著網路使用者增加，社群網站、搜尋結果每秒都在成長，因此輸出與反應數據的速度必須更加即時與迅速。然而為了發揮數據的價值，必須確認資料本身的正確性與**真實性(Veracity)**，避免資料的缺失與偽造影響分析結果，進而影響決策。

世界正在快速的改變，阿里巴巴集團董事長馬雲曾說：「我們正從以控制為出發點的 IT(Information Technology)時代，走向激活生產為目的的 DT(Data Technology)數據時代，這不僅是技術的升級，更是思想意識的巨大變革。」其背後所闡明的是 IT 時代的實現工具以硬體與軟體為主，DT 時代的工具則以雲端運算與大數據為主，且與 2015 年逐漸嶄露頭角的「互聯網+」一詞有著密切的關聯性。「互聯網+」不是單純的科技連接，而是透過整合，促進互動、分享與使用等，產生的創新與融合。「互聯網+」的本質是產業透過以網際網路為基礎設施的整套技術（例如包括行動、物聯網、雲端運算、大數據等），在經濟、社會、生活等應用所進行的線上化與數據化的過程，而其背後仰賴的基礎設施概括「雲+網+端」三種概念。其中，「雲」係指雲端運算與大數據；「網」係指網際網路與物聯網；「端」則係指各式各樣的終端設備。

阿里集團面對這樣的問題，曾經推出「千人千面」的應用，其係指當用戶進入電商平台首頁時，系統會根據當前用戶的屬性，例如性別、購物習慣、消費能力等，

展示與之相對的商品，或關聯的活動頁面來鎖定用戶的喜好，以期提高電商平台的媒合率與店鋪的成交率。然而，在產出此解決方案的背後，往往需要經過一套數據管理與運算過程，並產生用戶在特定市場或特定場景下的個性化標籤等，才能協助不同商品類別的營運員工，透過簡易地操作即可實現用戶標籤／屬性與商品的匹配。有鑑於此，從這個千人千面的例子中，其實蘊含著數據蒐集、管理、分類與營運等過程，必須將蒐集後的數據做更好的歸類與整理，才能更加清晰地識別出數據存在的價值。

數據的保存除了需要管理外，阿里集團對數據的研究，更進一步探討數據從「數據化營運」（同義於數據化運營）到數據應用，最後再進行「營運數據」（同義於運營數據）的過程。其中，「數據化營運」係指用數據幫助企業解決問題，例如當電商平台想要為童裝類別做行銷推廣時，通常會先找出目標客群，相關的作法可以找出曾經看過童裝但沒有購買的客戶群、將童裝類別儲存為喜愛的客戶群等，然後把對此類別感興趣的消費者用標籤圈起來，再透過寄送郵件或是簡訊的方式，吸引消費者關注，就是簡單的數據化營運過程；而數據從使用升級到培養新數據的過程，則稱為「營運數據」，不同於數據化營運只是在現有的數據中做進一步的挖掘與運用，營運數據則包含從主動蒐集數據，再產生新數據的過程。延續先前例子的說明，營運數據的作法可能是在用戶中，搜尋／推敲出有多少用戶家中有 12 歲以下孩子，但卻未發生過購買此類別商品的行為，例如從用戶過去曾經買過學齡圖書來推敲、或由過去的消費紀錄推算用戶孩子的年齡等，這個過程能將營運範圍從以前有過購買行為的幾百萬用戶，擴大到可能有 12 歲以下孩子的幾千萬用戶。從幾百萬到幾千萬的過程可被視為營運數據的一環，且從數據化營運到營運數據之間是不斷運行的循環，在這個循環中會容納許多新的、不同面向的數據，這些數據經過反覆循環的調整過程，再回饋到數據化營運中，並且改變原有的營運模式，就是阿里集團落實數據化的方法。

然而，數據令人期待的是只要數據量夠多，懂得分析與應用，不同場景就會產生不同的價值，有些數據可以和不同的數據進行結合，對用戶做出更好的商品推薦與精準行銷；有些數據可以強化供應鏈的反應與預測趨勢，例如雙 11 期間，阿里集團透過商品銷售的歷史數據與預購數據，來幫助商家提前將商品備貨到可能會有強勁需求的省份，以強化出貨的效率，提升服務水平。因此，數據不僅能提供後見之明，對於預測趨勢、瞭解人們的購物行為、演算行銷推廣模式等，都能有所幫助，使之成為微利時代企業不得不重視的競爭利器，對於大數據的運用，透過人工智慧