

數據浪潮即將襲捲全球 「巨量資料」帶來的結構變化

BIG DATA 的跨域技術與人才

作者／臺灣產業科技前瞻研究計畫團隊 高雅玲 (MIC)

備受關注的技術新星

根據 Gartner 每年所發表的技術成熟度曲線 (Hype Cycle)，將技術的發展分為五個階段，分別為萌芽期、過熱期、衰退期、復甦期和成熟期，讓企業可依此評估新興技術未來的投資方向與時間點。巨量資料技術的快速發展，近年也深被 Gartner 所掌握，並納入每年定期觀測的資料項目中。

Gartner 於 2011 年時首度將巨量資料技術納入於技術成熟度曲線中，2012 年時巨量資料技術更從萌芽期迅速移動，預估在 2 到 5 年內將進入過熱期，與 3D Printing、Internet TV、NFC、Activity

Streams、Cloud Computing、Media Tablets、Idea Management 等，並列為新興科技的重點項目。

更進一步檢視 Gartner 從 2011 年到 2013 年的十大策略性技術之變化，可發現巨量資料技術在短短三年之內，從十名之外迅速竄升到第六名。由此可知巨量資料技術的來勢洶洶，未來也將持續影響各產業的發展，甚至帶動其他技術議題邁向創新領域。此外，其他像是個人雲、物聯網、In-Memory 運算、App Store……的發展也都隨著巨量資料備受重視，一起熱門了起來，相互拉抬技術發展的聲勢。

圖 1. 2011 ~ 2013 年之十大策略性技術排名

2011		2012		2013	
1	Cloud Computing	1	Media Tablets and Beyond	1	Mobile Devices Battles
2	Mobile Applications & Media Tablets	2	Mobile-Centric Applications and Interfaces	2	Mobile Applications & HTML5
3	Next-Generation Analytics	3	Contextual and Social User Experience	3	Personal Cloud
4	Social Analytics	4	Internet of Things	4	Internet of Things
5	Social Communications & Collaboration	5	App Stores and Marketplaces	5	Hybrid IT & Cloud Computing
6	Video	6	Next-Generation Analytics	6	Strategic Big Data
7	Context-Aware Computing	7	Big Data	7	Actionable Analytics
8	Ubiquitous Computing	8	In-Memory Computing	8	Mainstream In-Memory Computing
9	Storage Class Memory	9	Extreme Low-Energy Servers	9	Integrated Ecosystems
10	Fabric-Based Infrastructure and Computers	10	Cloud Computing	10	Enterprise App Stores

資料來源：Gartner、MIC 整理，2012 年 12 月

關鍵技術的混搭使用

根據牛津大學與 IBM 在 2013 年初所發表的研究報告顯示，28% 的全球企業已經開始使用巨量資料。巨量資料技術發展的挑戰在於如何妥善處理資料的三個特徵：數量、速度與多樣性。隨著資料量的增幅加大，以及資料格式的更多樣性，企業渴望得到即時分析的需求也愈明確，舉凡像是醫療、製藥、製造、通訊、零售、能源、運輸、汽車、金融等領域皆可能大量運用巨量資料技術。

企業若有效運用巨量資料，須從資料的擷取、分類、整合、處理、分析到預測搭配適當的技術，才能讓這麼龐大的數據資料為人們所用。因此，常見的巨量資料處理技術，主要為機器學習 (Machine Learning)、資料探勘 (Data Mining)、資料分群 (Data Clustering)、神經網路 (Neural Network)、迴歸分析和關聯分析等，依據資料的型態、分析目的等的不同，相互搭配使用，以做出最佳的決策建議。

機器學習是一種能透過經驗而自動改進的演算法，可持續優化演算結果的正確性，是屬於人工智慧的一門技術，可運用於語音辨識、垃圾郵件過濾、推薦引擎、氣象預報、路線規劃、機器故障等領域。而資料探勘則是藉由針對大量資料進行分析，它可用自動或半自動的方式，找出不為人知且蘊藏其中數字意涵，通常會混用資料分群，把相類似的個體聚集在一起，並給予分類，如把背景、興趣、嗜好相近的人歸類在一起，提高行銷活動的成功率。或是經由關聯分析，把買 A 的人也買 B 的消費傾向，從繁雜的消費資訊中梳理出來，或是找出潛在客戶提升客戶忠誠度等。

而巨量資料的應用，常見的有進行推薦商品、行為定位廣告、運用地點資訊的行銷、消費者行為分析和故障與異常預測等。

所謂的推薦商品，是利用消費者的屬性、行為、消費資料等資訊為基礎，進行相似度的分析，以提供消費者購買商品的建議清單，如 Amazon 的「購買

這個商品的人，也同時買了這些商品」功能、Facebook 的「你可能認識的人」功能，許多人因此增加了消費，也有許多人找到自己失散多年的同窗好友，重新拉近人與人之間的距離。

而行為定位廣告，則是透過在電子商務平臺上的各種消費資訊，以及消費者的興趣、嗜好等資料將其分群，並針對不同分群提供不同內容的廣告服務，此類型的技術已被廣泛運用於搜尋引擎的入口網站，如：Google 或 Yahoo。

另外一項為人們所廣泛利用且饒富趣味的應用技術，則需透過現在人手一機的智慧型手機才能達成，透過智慧型手機當中的 GSP 定位功能，蒐集其所在位置資訊，結合當地的地理資訊、店家、生活空間，以及其他使用者的資訊，進行資訊的串聯，更進一步將周邊的人們進行分群，提供最合適的廣告簡訊或電子折價券，刺激消費，建立社群關係等。巨量資料分析結果良窳，除了受限於軟體、人力等資源條件外，亦受限於資料本身的正確性與完整性，正所謂「Garbage in, Garbage out」，垃圾進去，出來就是垃圾，如何有效去除雜訊，提高分析運算的效能，亦為巨量資料使用者關心之所在。

有了優質且大規模的資料來源，加上適度的組織資源，完善的儲存設備，都不能保證企業就能有效掌握巨量資料所帶來的甜美果實，根本在於企業有沒有能力運用這些工具的專業人才？

技術論壇 INNOVATION TO-BE

人才才是勝出的根本

有了優質且大規模的資料來源，加上適度的組織資源，完善的儲存設備，都不能保證企業就能有效掌握巨量資料所帶來的甜美果實，根本在於企業有沒有能力運用這些工具的專業人才？

這類專業人才被稱之為「資料科學家」，他們是一群能掌握分析工具，從複雜且龐大的資料礦山中，挖掘出金礦的人，並將所分析出的成果及可能的價值，藉由簡單的圖表表達出來，供決策高層做參考。資料科學家在企業的工作中，因為需要與許多內、外部的成員進行積極的交流，並且把難懂的資料分析結果，透過簡單、清晰的表達方式，傳達給不具專業知識的相關同仁或主管，所以在選才的過程中，應重視其溝通能力。同時，資料科學家也應充滿好奇心，才能發揮創意，把不同的生活觀察與領域做結合，找出意想不到、深具價值的洞見。

除了選才難，這類型人才的培養過程也十分困難，不僅需要具備傳統統計學的知識，還需要接受數學、多變量分析、管理科學、行銷、人類學、心理

學、資訊工程、經濟學等學科領域的洗禮，才能把分析結果轉換出商業價值。

有了電腦科學相關的認識，資料科學家才能正確的與軟體工程師進行溝通，提升資料處理技術或機器學習能力之精進；同時，嫻熟數學、統計之邏輯，才能明確的知道各式分析工具的適用性與侷限性，以正確的方式進行資料的處理與分析；有行銷學的瞭解，才能將分析的結果，轉化為創新的行銷內容與模式等。因此，多元能力的養成是培育此類人才最為困難的地方。

有了高效的 IT 設備與資料處理、分析工具，搭配睿智又敏感的資料科學家，最後一項，可能影響企業使用巨量資料成效的，就是企業本身的組織與文化，無論分析出來的結果有多麼優異，企業欠缺積極、創新的文化，過度秉持傳統，缺乏高層的承諾與支持，也很難將分析結果具體規劃出策略，進而落實。因此，巨量資料時代的來臨，不僅是企業排名重新洗牌的機會，也是讓企業組織與文化調整的好時機。

