

# 雲端運算的應用—— 資料採礦和商業智慧

■ 謝邦昌

雲端運算是一種新興的商業計算模型，  
由於具低成本、高效率、高擴展性等優質特性，  
吸引了眾多公司競相投入，即將「引爆商業革命，改寫遊戲規則」。

自2006年3月亞馬遜（Amazon.com）首開市場先例，推出雲端服務，2007年谷歌（Google）正式提出「雲端運算」（cloud computing）一詞後，愈來愈多的企業希望乘風扶搖而上，直入雲端。雲端運算即將「引爆商業革命，改寫遊戲規則」。

雲端運算是一種新興的商業計算模型，由於具低成本、高效率、高擴展性等優質特性，吸引了眾多公司競相投入。例如Google搜尋服務、Gmail、YouTube、Google Docs、Google Talk、iGoogle、Google Calendar等都使用雲端演算的技術，微軟、雅虎、亞馬遜也採用這技術提升網路服務功能，Google、微軟、蘋果（Apple）正在上演一場激烈爭搶雲端運算商機的戰爭。

不僅企業在追逐雲端運算，政府也緊跟其後。2009年9月，美國白宮資訊長坎卓（Vivek Kundra）宣布，運用雲端科技開發出來的新政府網站Apps.Gov，將取代舊有的資訊系統。

透過雲端，網路服務供應者可以在數秒內處理數以萬計的資訊，  
提供等同於「超級電腦」效能的網路服務。



超級電腦（圖片來源：財團法人國家實驗研究院國家高速網路與計算中心）



聚集運算（圖片來源：財團法人國家實驗研究院國家高速網路與計算中心）

## 雲端運算

**概念** 雲端運算是基於網際網路的運算方式，透過網際網路為個人使用者或企業使用者提供按需求取用的服務。具體來說，雲端演算法透過網際網路，把龐大的運算處理程式拆分為大量的副程式，交由大量伺服器所組成的龐大系統去處理，經過搜尋和運算分析後，再把處理結果傳回使用者端。透過雲端，網路服務供應者可以在數秒內處理數以萬計的資訊，提供等同於「超級電腦」效能的網路服務。

雲端運算的基礎是網際網路，網際網路在示意圖上常以雲狀圖案表示，「雲端」一詞即由此而生。雲端在《漢英詞典》中的解釋是「在雲中、在雲上」，這是對使用網際網路來進行運算、處理問題的一種直觀、清晰的表述。直觀來講，雲端演算法就是讓網路上不同的電腦同時幫



雲端圖

雲端運算的最終目標是，使用者只需一個連接雲端的設備（如手機）和簡單的介面（如瀏覽器），就可從雲端快速獲得自己想要的資源。

使用者做一件事情，這大幅度地提高了處理的速度和效率。雲端運算的最終目標是，使用者只需一個連接雲端的設備（如手機）和簡單的介面（如瀏覽器），就可從雲端快速獲得自己想要的資源。

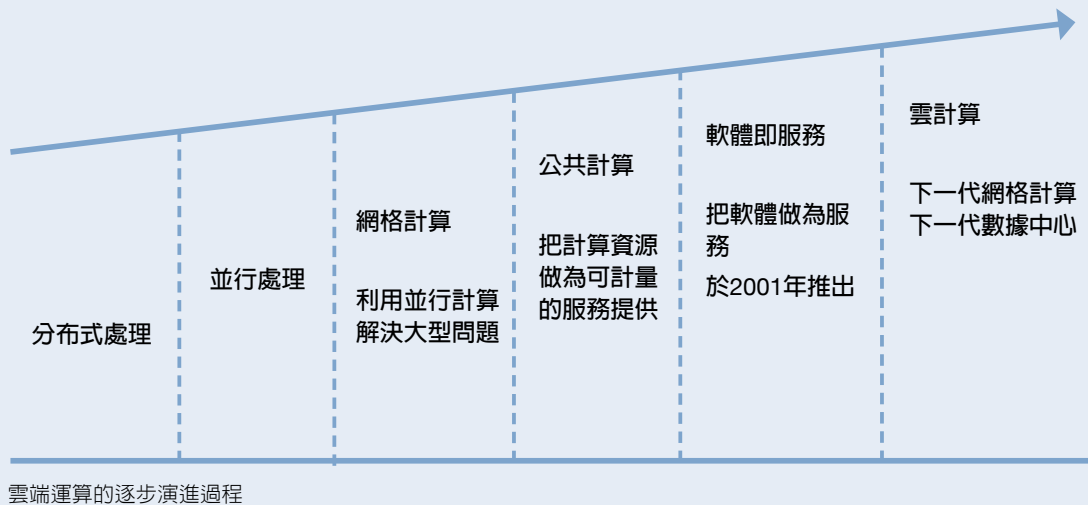
相較於其他運算，雲端運算是基於虛擬化技術快速部署資源和獲得服務的，它具有按需求提供資源、按使用量付費、成本低、通用性高、擴充性佳、使用方便、服務配置速度快等特點。

雲可以分為3種：公共雲、私有雲、混合雲。公用雲是具有公用服務（如水、電、燃氣等服務）本質的雲端技術。私有雲是為一個客戶單獨使用而構建的，是企業內部自用的雲端計算技術，可以有效地控制資料的安全和服務品質。私有雲可部署在企業資料中心的防火牆內，也可以部署在一個安全的主機託管場所。混合雲是公用雲和私有雲混合組成的雲端計算

技術。建議企業可以利用公共雲儲存敏感性不高的資料並做運算，至於敏感性高的資料，則可以在內部建立私有雲來提升資料的安全性。

**服務層次** 雲端運算包括以下幾個層次的服務：硬體即服務、軟體即服務、資料即服務、設施即服務。雲端運算服務通常提供能夠透過瀏覽器存取的線上商業應用，使用者需要雲端運算的計算平台或資訊技術（IT）基礎設施，並在這些平台與設施中運行應用。因此，計算雲給予使用者進入硬體、軟體及資料資源的服務之後，一個集合而成的計算平台便以一種透明化的方式進行服務。

**硬體即服務**—「硬體即服務」這一術語可能是在2006年產生的。由於硬體虛擬化及IT自動化的迅速發展，使用者能夠購買IT硬體或整個資料中心，做為現收現付的訂購服務。硬體即服務是以靈活的、可升級的，並且可管理的硬體設備來滿足客戶的需求。





雲端應用的流程

數據即服務—通過網際網路提供軟體或應用服務給顧客。這種模式降低了使用者的電腦在安裝軟體或運行應用的需求，因此，軟體即服務減輕了使用者在軟體維護上的負擔，並降低了購買軟體的費用。

資料即服務—網上使用者可以透過服務，從多種資源中獲取多種形式的資料。例如使用雲端控制遙遠的資料，就像在本地存取硬碟資料。

設施即服務—以前述3種服務的支援為基礎，雲端運算能夠給使用者提供設施即服務。用戶可以

按照需求訂購他們喜歡的硬體設定、軟體安裝和資料需求。

### 資料採礦和商業智慧

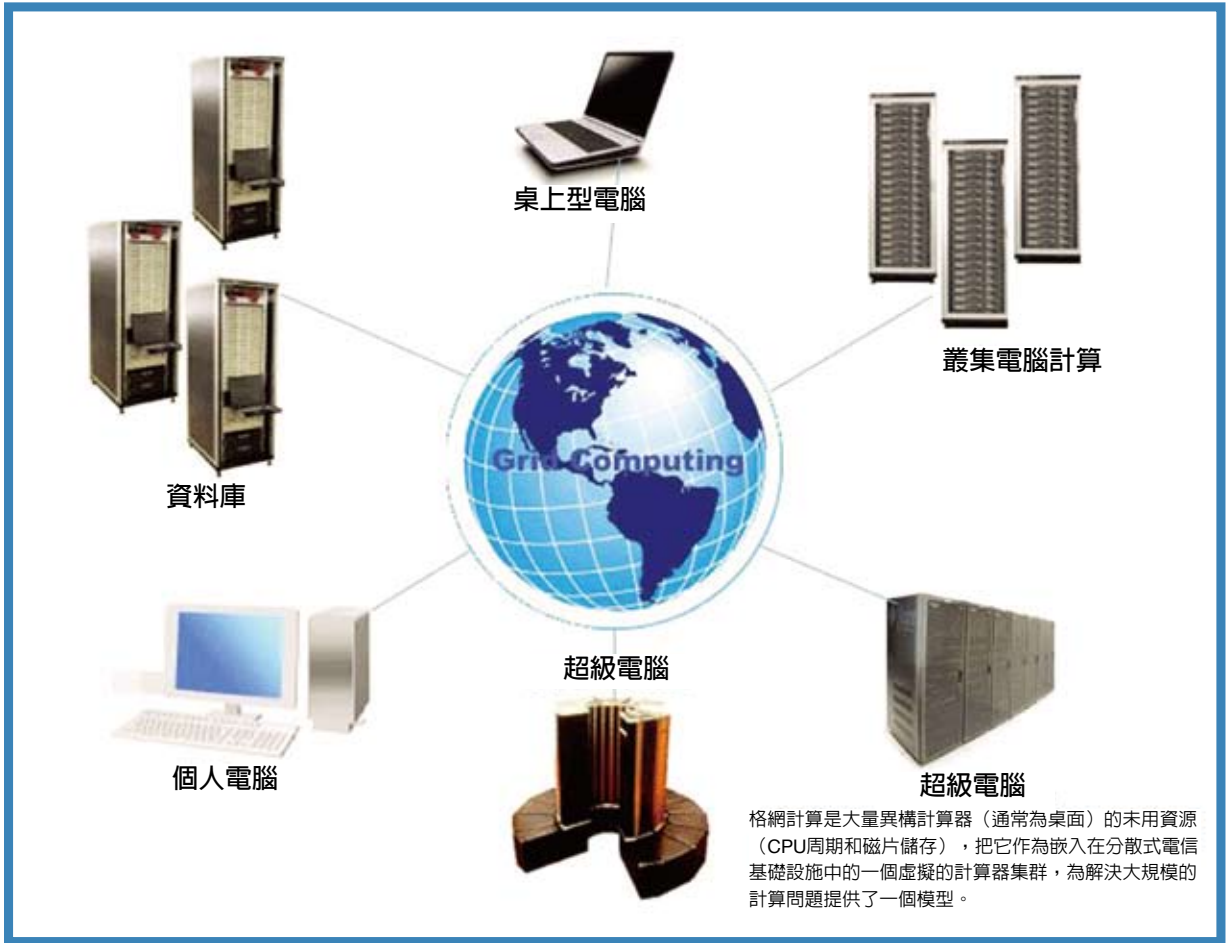
隨著電腦技術的發展，資料的存儲量成倍增長，而海量資料分析方法的發展卻難以望其項背，資料採礦（data mining）技術就是在這樣的「資料海洋、知識荒漠」的情況下應運而生。資料採礦是從海量的資料中發現潛在規律，提取有用知識的方法和技術。它不僅能分析現有的問題，也能預測未來的趨勢，且分析的結果易於理解與應用，因此一出現就得到各個領域的重視。隨著資料採礦逐漸成熟，以資料採礦為核心的商業智慧（business intelligence，簡稱BI）已經成為資訊和其他行業的必爭之地。

當前資料採礦主要集中在生物、電信、零售、銀行、醫藥等產業方面，並在許多領域有了成功的應用實例。例如當您在網路書店中點選其中一本書後，就會出現相關的推薦書目，這背後就是資料採礦在發揮它的分析功能。

**資料採礦** 資料採礦是由電腦技術、人工智慧技術、統計技術等構成的一門新學科。它採用數學、統計、人工智慧、神經網路等領域的科學方法，運用如聚類分析、關聯分析、決策樹等技術，從大量資料中挖掘出隱含的、先前未知的、對決策有潛在價值的關係、模式和趨勢，並用這些知識和規則建立用於決策支援的模型，提供預測性決策支援的方法、工具和過程。

隨著電腦技術的發展，資料的存儲量成倍增長，而海量資料分析方法的發展卻難以望其項背，資料採礦技術因而應運而生。





分散式運算（圖片來源：財團法人國家實驗研究院國家高速網路與計算中心）

資料採礦源自於統計分析，而又不同於統計分析。資料採礦是統計分析方法的擴展和延伸，資料採礦的前期資料探索、分析及後期的結果描述都離不開統計分析方法，並且部分資料採礦演算法是建立在統計學的理论基礎上（如支持向量機演算法），統計學方法對於提高資料採礦的可信性和準確性具有重要的作用。

根據「跨行業資料採礦標準過程」，資料採礦的過程可以分成以下6個步驟（但順序並非完全不變）：

商業理解—從商業的角度了解專案的要求和最終目的，並把這些目的轉化成運用資料採礦方面的計畫及目標。

資料理解—根據要求從資料庫中提取相關

的資料，並對可用資料進行評估。

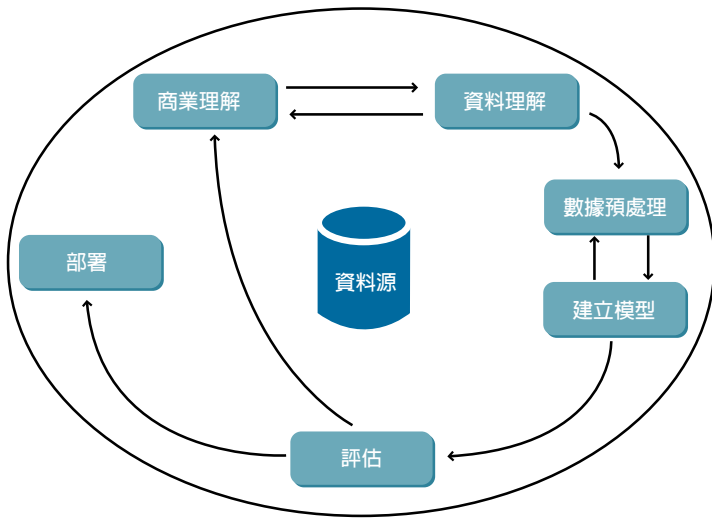
數據預處理—主要對提取的資料加工，檢查資料的完整性和一致性，填補缺失的資料，處理噪音資料等，以達到建模要求。

建立模型—運用資料採礦工具建立模型。

評估—評估建立的模型，並具體考察得到的結果是否符合商業目的。

部署—把發現的結果及認識過程組織成可讀文本形式，即撰寫資料採礦報告。

資料採礦的演算法很多，大致可以歸為以下幾類：決策樹演算法、群集演算法、時間序列演算法、時間群集演算法、關聯規則演算法、類神經網路演算法、羅吉斯迴歸、貝式機率分類、線性迴歸和資料採礦等。



跨行業資料採礦標準過程的示意圖

**商業智慧** 商業智慧就是把智慧計算技術應用於傳統商業領域，從而提高資料分析能力，改良整個業務過程，提高企業競爭力。雖然最近幾年才開始普及實施商業智慧，但它的應用範圍已遍及金融、電信、零售、醫藥、製造、政府單位等各個行業和領域，成為大中型企業經營決策的重要部分。商業智慧的核心是：如何蒐集資料、如何管理資料、如何從資料中獲得智慧，以及如何應用智慧。

實施商業智慧系統是一項複雜的系統工程，涉及到企業管理、資料倉儲、資料採礦、統計分析等眾多類別的知識，因此正確的實施方法非常重要。商業智慧專案的實施步驟可分為：

**需求分析**—需求分析是商業智慧實施的第一步，必須首先確認企業的需求和期望，確定需要分析的主題等。

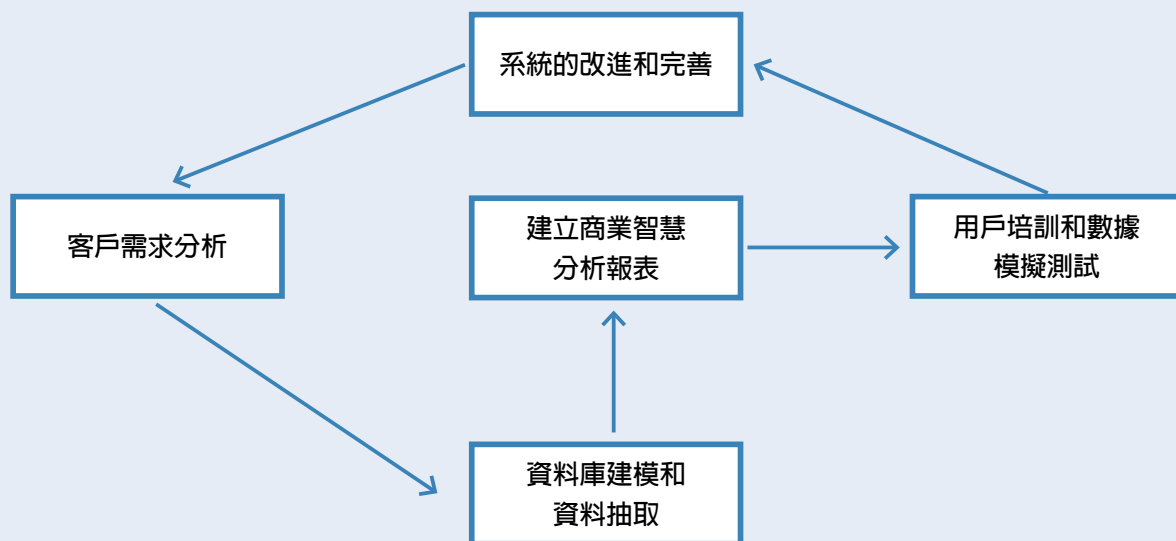
**資料庫建模與資料抽取**—通過對企業需求的分析，建立企業資料庫的邏輯模型和物理模型，並規劃系統的應用架構，把企業各類資料按照分析主題組織和歸類。

資料庫建立後，必須從業務系統中抽取相關資料到資料庫中，在抽取的過程中還必須進行資料轉換、清理，以便適用於分析的需要。

**建立商業智慧分析報表**—需要專業人員按照用戶制訂的格式開發，或由用戶自行開發。

**用戶培訓和數據模擬測試**—對於開發 / 使用分離型的商業智慧系統，最終用戶的使用是相當簡單的，只需要點選操作就可分析特定的商業問題。

**系統改進和完善**—任何系統的實施都必須不斷改善，商業智慧系統更是如此。在用戶使用一



商業智慧過程示意圖

商業智慧把先進的資訊技術應用到整個企業，不僅提高企業的資訊獲取能力，而且透過對資訊的開發，把資訊轉變為企業的競爭優勢。

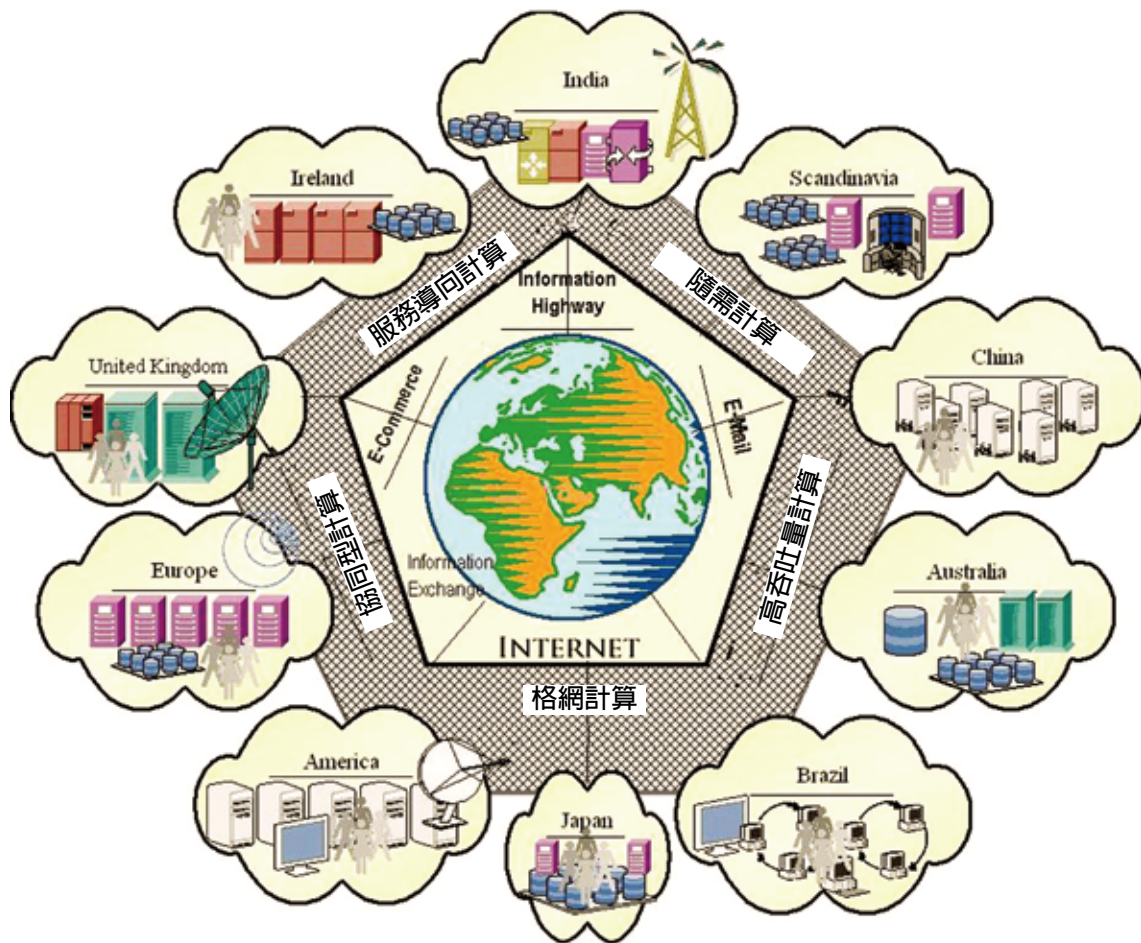
一段時間後可能會提出更多、更具體的要求，這時需要再按照上述步驟重構或改善系統。

### 應用

雲端運算可以實現使用端透過網路上傳資料或購買資料，通過雲端資料倉儲服務進行資料庫建模和資料抽取，線上支付使用資料採礦工具和商業智慧相關分析處理軟體的費用。資料採礦和商業智慧的原理相似，都是由資料

提供資訊、產生知識，再由知識累積智慧，而雲端運算可以使這個過程在網際網路上進行。也就是說，雲端運算可以提供基於「軟體即服務」的知識與智慧分析的服務。

商業智慧把先進的資訊技術應用到整個企業，不僅提高企業的資訊獲取能力，而且透過對資訊的開發，把資訊轉變為企業的競爭優勢。因此，越來越多的企業提出他們對商業智慧的需求，把商業智慧做為說明企業達到經營



網絡運算（圖片來源：財團法人國家實驗研究院國家高速網路與計算中心）

目標的一種有效方法。然而商業智慧活動，如分析、資料倉儲和資料採礦，要求的費用龐大，讓很多中小企業望之卻步。雲端運算具有按需要使用、按使用收費的特性，這會大大減少企業應用商業智慧的成本。雲端運算的出現，使得商業智慧分析不再只是中小企業的一個夢，而是能夠付諸實行。

在商業智慧上，運用雲端運算可以把資料與應用託管到「軟體即服務」供應商，一方面可以通過雲的資料倉儲實現海量資料的高效率運算，另一方面也可以實現線上使用及支付資料採礦工具和商業智慧相關的分析處理軟體的費用。

資料倉儲是資料採礦發展的基礎，也讓商業智慧得以支撐，資料採礦也是商業智慧的重要環節，由此可見資料倉儲對商業智慧而言有很重要的作用。它集成了企業的最核心資料，隨著企業對資料再次利用和深入挖掘，海量資料的高效計算問題成爲企業關注的問題之一。

雲端計算技術給資料倉儲的高效運算帶來革命性的影響。例如OnDemand公司是一個資料倉儲提供商，可以透過BusinessObjects Web智慧來分析資料；又如資料庫廠商Kognitio提供了基於雲端的資料倉儲，它標榜「資料倉儲即服務」。還有資料倉儲應用的先驅者Netezza公司，也和ppNexus一起合作發布了一個基於雲的資料倉儲服務，後者是一個硬體即服務雲端計算廠商。

線上資料採礦服務也受到一些企業的青睞，例如中國的海量資訊技術有限公司提供名爲「海納睿智」的線上資料採礦服務，可以利用線上網路資料採礦技術上的優勢，爲用戶提供個性化需求，訂製個性模組服務，來滿足網站的不同需求。

此外，在商業智慧平台的開發工具方面似乎已到達爆發的時代。例如KNIME是基於Eclipse環境的開發商業智慧工具，它是透過工作流程來控制

資料的彙整、清理、轉換、過濾，再到統計、資料採礦，最後是資料的視覺化。整個開發都在視覺化的環境下進行，運用簡單的拖曳和設置就可以完成一個流程的開發。

雲端運算在資料採礦和商業智慧上的應用非常廣泛，如構建知識脈絡、繪製知識地圖、追蹤事件、推薦知識社群等。除資料採礦和商業智慧之外，雲端演算法也可應用於其他領域，如微軟與中國電信在整合雲端運算和電腦、電視、手機等的合作專案，主要是讓消費者與企業用戶體驗到任何時間、任何地點、使用任何連網的用戶終端設備，就可快速取得所需資訊與服務，享受便利的數位生活。

當今智慧型手機、衛星導航等行動裝置，都可以透過雲端運算發展出更多的應用服務。雲端運算更可應用在生物科學，例如分析基因結構、基因圖譜定序、解析癌症細胞等，利用雲端運算架構協助，效率快又準確。當大量資訊處理不再昂貴時，許多科技就會百花齊放，應運而生。

雲端運算集低成本、高效率、高擴展性等優點於一體，但由於尚處於發展初期，它仍有一些問題等待解決。如雲運算系統故障問題（例如Sidekick服務中斷、Amazon EC2遭到阻斷服務攻擊，以及Google電子郵件服務中斷）、資料遺失或遭竊問題、遺失的資料需要使用者自行承擔責任、欠缺獨力軟體發展商支持等問題，這些都是發展雲端運算必須面臨的挑戰。漫步雲端，任重而道遠！

謝邦昌

輔仁大學商學研究所