



## 雲報專欄：美國大學如何研究雲端運算

### — 郭耀煌/成功大學資訊工程學系特聘教授，技術專家 委員會委員、蘇維宗/真理大學資訊工程學系助理教授

近年來，在各類資通訊展覽中不難看出「雲端運算」儼然已經成為各家廠商爭相拉攏的對象，似乎只要產品掛上這四字就能夠熱賣。然而，多數人對於「雲端運算」四字雖然朗朗上口，但對其核心概念與關鍵技術卻是似懂非懂。就如同 Olafur Ingthorsson 分析師在”Cloud Computing Frees Companies from the Burden of IT”一文中提到”the definition of cloud is cloudy”；不同背景的人對於「雲端運算」的理解可能不盡相同。從學術研究的角度來看，雲端運算雖然是以網際網路(Internet)為基礎，但並非單純地利用網際網路來存取資料庫那麼簡單。因此，本文藉由從美國各大學所進行的雲端運算研究來深入了解其主要研究議題。

2008 年，Google 與 IBM 共同啟動的 [Academic Cloud Computing Initiative](#) (ACCI) 計畫與美國的國家科學基金會(National Science Foundation, NSF)合作，透過由 Jim French 博士所主持的 Cluster Exploratory (CLuE)計畫，以高達 5 百萬美金的經費補助了 14 所美國大學[註 1]，進行雲端運算的相關研究。2011 年，NSF 舉辦了 NSF PI Meeting: The Science of Cloud 研討會。在此研討會中，聚集了來自於幾十所獲得補助之大學與研究機構的 63 位的學者專家共同探討在雲端運算研究議題上的發現，其主題包含：

- Cloud Architectures and Systems
- Network Support for the Cloud
- Data Portability, Consistency, and Management





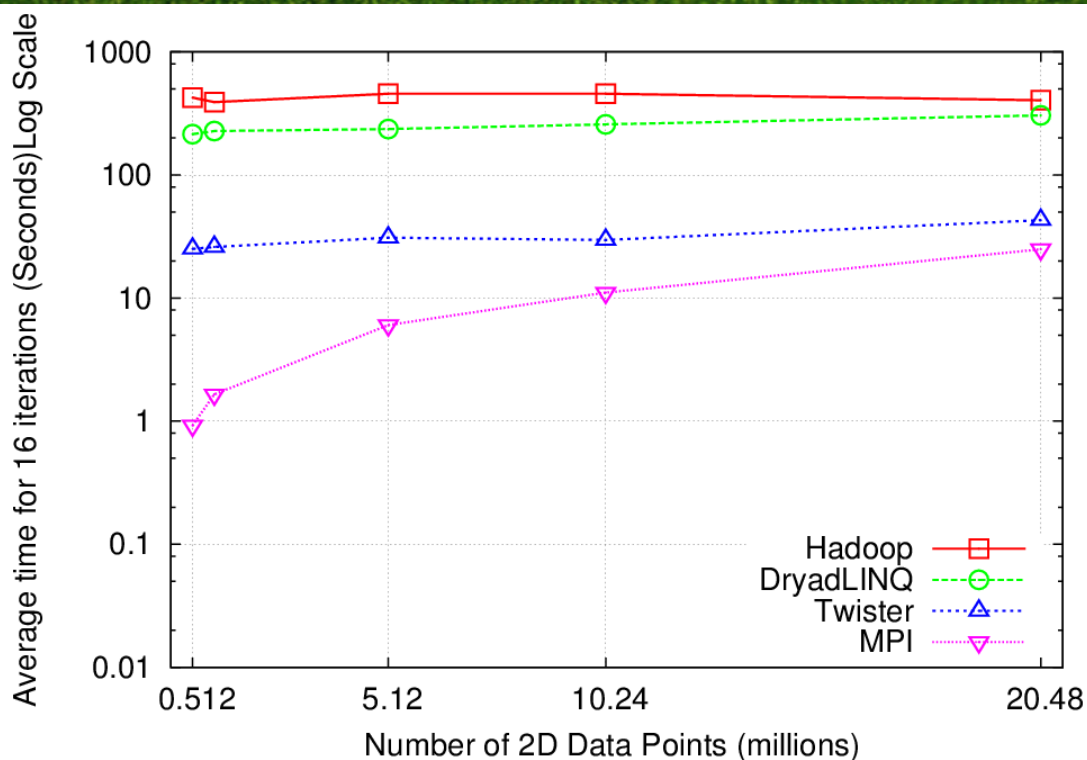
- Programming Models for the Cloud
- Fault Masking in the Cloud
- Cloud Security, Privacy, and Auditing
- Cloud Debugging, Certification, and Update
- Cloud Self-Monitoring and Autonomic Control
- Cloud Inter-operability and Standardization
- Green Clouds
- Cloud Test-Beds

除了研究如何運用雲端運算來協助需要處理大量資料的科學應用研究，例如加州大學爾灣分校進行的文件資訊處理(document cleaning)、馬里蘭大學進行的基因體研究(genomics research)與卡耐基美隆大學進行的語言翻譯(machine translation)之外，還有許多其他與雲端運算研究議題的討論。以下只針對幾個與雲端運算效能相關的研究議題，說明如下：

#### (一) MapReduce 架構的應用性(applicability)

事實上，雲端運算中的 MapReduce 架構並不適用於所有應用。例如，普渡大學與印第安納大學的研究團隊在其研究中發現了傳統的 MapReduce 運作並不適合用來執行像是 K-Means 這類的迭代(iterative)演算法。圖一為印第安納大學的 Judy Qiu 以不同的運算架構，包含知名的雲端運算開放式軟體，Hadoop，執行 K-Means 所得到的效能數據。主要原因是在每次迭代過程中，MapReduce 都需要將資料從分散式檔案系統讀出和寫入，因此在資料同步上需要花費相當多的時間。





圖一 Hadoop 處理迭代演算法的效能(資料來源：印第安納大學 Judy Qiu 博士簡報)

## (二) MapReduce 與資料庫的差異

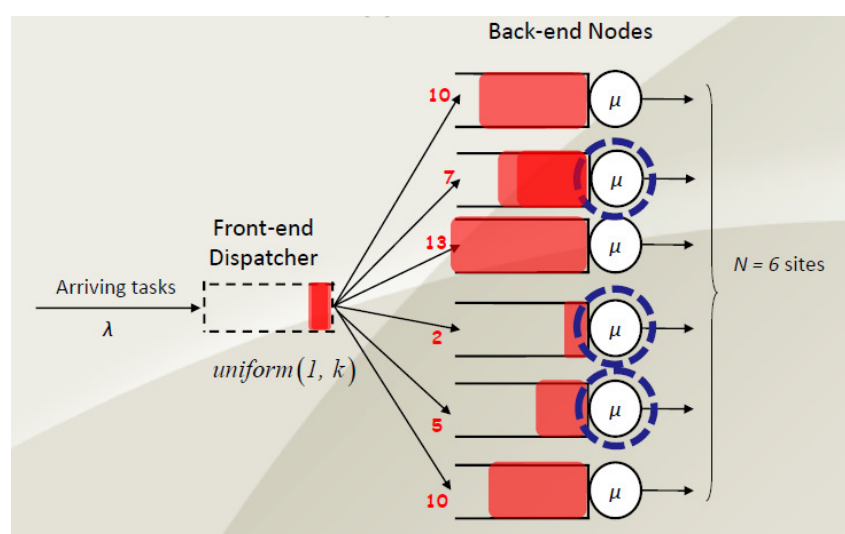
傳統的資料庫架構常常被拿來與雲端運算進行比較。由麻省理工學院、威斯康辛大學麥迪遜分校、耶魯大學、微軟與布朗大學所組成的研究團隊主要是研究 MapReduce 與資料庫在架構與效能上的差異。此研究團隊發現 MapReduce 的資料載入時間(load times)比資料庫還要快，且隨著結點數量的增加呈線性遞減。然而，MapReduce 在資料查詢(query)的速度雖然仍然隨著結點數量的增加呈線性遞減，但是比起傳統的資料庫卻慢上許多。從該研究團隊的研究成果可以看出，MapReduce 雖然有較高的擴充性(scalability)，然而在資料查詢的效能上卻不及傳統的資料庫架構。因此，發展一個能夠與資料庫結合的 MapReduce 架構是一個相當重要的研究議題。





### (三) 雲端運算資源配置問題

在異質(heterogeneous)的雲端運算環境中，有個重要的研究議題就是如何有效管理資源以達到最大的系統效益。南加州大學的研究團隊認為雲端運算的資源管理必須能夠根據需求來進行適當的配置，如此一來才能夠在成本（如：損耗電力）與運算效能間取得平衡。此外，東北大學(Northeastern University)也提出雲端運算的資源必須能夠根據節點的工作狀況來動態配置才能最佳化系統的效益，如圖二。



圖二 根據節點工作狀況配置資源(資料來源：東北大學 Ningfang Mi 博士簡報)

### (四) 資料與運算外包的安全性

雲端運算的安全性一直被認為是相當重要的議題。伊立諾理工大學提到資料外包雖然目前已經有可搜尋的加密方法(searchable encryption)相關研究，但仍然需要更多努力才能夠達到像是模糊搜尋、相似度排序、或非文字資料搜尋等需求。另外，運算外包雖然 Craig Gentry 博士在理論上已經證明了利用 Fully Homomorphic Encryption (FHE)來對密文進行運算是可行的，但目前仍未有適用於電力與運算能力受限之終端裝置(end-device)的實作方案。另外，在多租戶(multi-tenancy)雲端架構下如何確保虛擬機器的隔離亦是需要研究的安全性議題。



由此可知，雲端運算非只是單純地利用網際網路來存取資料庫那麼簡單，其 MapReduce 與虛擬化(virtualization)等核心技術在效能與安全性等議題上仍有許多尚待解決的問題。另外，台灣雖然在這一波雲端運算產業中以過去累積的硬體研發實力能夠在雲端伺服器市場中佔下一席之地，然而從上述核心研究議題不難看出，台灣必須投入更多的資源開發雲端運算系統軟體技術，才不致被國外系統大廠所左右。

註 1：獲得補助的美國大學包含：卡耐基美濃大學、佛羅里達國際大學、麻省理工學院、普渡大學、加州大學爾灣分校、加州大學聖地牙科分校、加州大學聖塔芭芭拉分校、馬里蘭大學、麻州大學、維吉尼亞大學、華盛頓大學、威斯康辛大學、猶他大學、耶魯大學等 14 所學校。

