網路技術與應用期末報告

Amazon.com 雲端運算服務發展

指導教授: 林永松 老師

B97705007 資管三 孫旭辰
B97705008 資管三 王祥鈺
B97705023 資管三 彭俊穎
B97705026 資管三 吳東曄
B97705031 資管三 謝維哲

目錄

目錄 1
雲端運算技術簡介 2
雲端運算的定義 2
命名的由來 2
嚴謹的定義 2
雲端運算服務的特色 2
分散虛擬架構 2
網路使用介面 3
彈性付費模式 3
隨選應用服務 3
雲端運算運算演進歷程 4
網格運算 4
公用運算 4
軟體即服務 4
雲端運算 5
雲端運算核心技術剖析 5
Hadoop 源起 5
Hadoop 架構 5
商業應用模式 6
Amazon Web Services 7
Amazon 簡介 7
AWS 簡介 7
AWS 服務 7
AWS 主要的五個核心業務 8
應用實例 9

雲端運算技術簡介

雲端運算的定義

命名的由來

雲端運算(Cloud Computing)一詞最早由 Google 提出，起源於網頁資料的搜尋過 ，透過大量的遠端伺服器進行平行運算，再將結果彙整和回傳給使用者。由於伺服器群皆在遠端網路，一般 IT 業界也習慣將網際網路描繪成 Cloud，所以 Cloud Computing 名詞由此而來。

嚴謹的定義

雲端運算嚴謹的定義是「透過網際網路的分散式運算(Distributing Computing)架構，所提供的一種服務(Service)模式，並且具備彈性(Flexibility)與可擴充(Scalability)的能力」。有別於傳統 IT 代管服務(Hosting Services) ，雲端運算特別強調「平行運算」的資源彈性和可用性(Availability) 。不過，現在資訊服務產業普遍對於雲端運算的定義採取較寬鬆的認定態度。廣義來說，任何網際網路上提供的運算資源和隨選服務都是雲端運算服務的涵蓋範圍，只要滿足彈性使用和可擴充的特性，並不一定需要完全符合分散式電腦運算架構。

雲端運算服務的特色

分散虛擬架構

雲端運算屬於分散式運算的一種。雲端運算可以採用大量規格相同的 X86 平價伺服器進行平行運算，不需要處理異質性的問題，可以簡化平行運算的系統架構，提高系統容錯能力，提供大量且穩定的網路服務給衆多的線上使用者。 又雲端運算可以隱藏底層的複雜基礎環境，利用 IT 管理工具可以將使用者需求負載(Workload)分配到多個動態的虛擬機器(Virtual Machine，VM)上面執行。對使用者來說，操作介面非常簡易，如同在桌面端執行一般；對資料中心業者而言，可以促使現有伺服器資源的分配更為彈性，能夠最佳化資料中心的所有運算資源，達到完全 IT 虛擬化的目標。

網路使用介面

雲端運算主要利用遠端伺服器的運算，再將結果回傳給本機端使用者，因此相當依賴網際網路的基礎建設。尤其為了實現企業級商用軟體雲端服務化，高速、不中綴的網路連線品質為基本要求。全球雲端運算服務領導廠商更是持續建置多個世界級的資料中心(Data Center) ，其中不少更是緊鄰在全球主要光纖骨幹(Backbone)上，以期服務品質能滿足 SLA(Service Level Agreement)保證協議。 同時網頁瀏覽器(Browser)在未來將扮演吃重的角色，任何雲端任務將透過 browser 的簡易參數設定、並傳送請求給雲端運算服務提供者去執行。本機端將不需安裝複雜的應用程式、甚至作業系統，因此，精簡型電腦(Thin-Client)可能是未來電腦市場的主流產品。

彈性付費模式

雲端運算主張用多少付多少(Pay-as-you-go)的消費方式，完全視企業的使用需求量提供服務。例如企業目前僅需要 10TB 的儲存空間，但未來因

應業務擴充需求可能要增加到 50TB 的容量，雲端運算服務提供者可以立即升級或降級服務內容。就好比申請 ISP 服務，從月租 2M／256K 升級至月租 8M／640K 一般，使用者可以自由選定適合的服務內容和相關費率。 由於中小企業占企業市場絕大多數，高額版權的商用軟體對他們是非常遥不可及的商品，因而長久以來一直是最被忽視的客群。如今標榜平價精神的 SaaS 問世，軟體大廠相繼推出商用軟體線上版本，提供多樣的產品組合方案，以拉攏中小企業客群，冀望發揮長尾效應(Long-Tail Effect) 、增加新的收益機會。

隨選應用服務

雲端運算的另一特色就是馬上申請、馬上使用。以往企業需要新的商務資訊系統，必須投入 IT 人力和漫長的本機端(On-Premise)部署過程，但等到導入完成後，卻也可能錯失最佳的市場切入時機(Time-to-market) 。雲端運算服務則免去先期部署流程，可以立即申請使用，提供極大的便利性。 另外，雲端運算服務獨立於本機端的使用者介面，不論是 PC、NB、手機或 Thin-Client 等，辦公室內或出差途中，皆可隨時隨地使用。雲端運算服務提供者也透過資料中心的建置與維護，提供企業的高可用性。

雲端運算運算演進歷程

網格運算

網格運算主要起源於網際網路(Internet)尚未盛行的 80 年代末期和 90 年代初期，主要藉助標準化溝通協定來整合異質(Heterology)伺服器、並建構強大的叢集運算系統，使組織內部的 IT 資源有良好的使用率。網格運算通常用來處理複雜的科學運算，如解析衛星照片、分析生物 DNA 結構等，一般多應用於國防與特殊領域研究中，需要單次超級運算(Super Computing)的任務。由於網格運算的實務操作情形可以抽象模擬成為一台超級電腦，所以其具體代表意象就是 Virtual Server。

公用運算

公用運算的概念最早出現在 1961 年，由 MIT 的 John McCarthy 教授在百年校慶演講時首先提出。他預言未來電腦技術將朝向多工平衡負載(Multi-Tasking)的方式演進，以後的軟體或資訊服務將類似日常民生水電的計量(Metered)付費方式提供，就像水龍頭隨開即用般的便利。不過這樣的概念一直延續到 70 年代中期，最終在硬體、軟體以及通訊技術能力仍無法滿足的情況下慢慢式微。直至 90 年代末期，網通技術有著長足的突破，網路基礎建設開始普及，ISP 業者將資料中心資源提供給外部使用者成為實際可行的商業模式，並可依使用量收費。由於公用運算可模擬成提供線上服務的平台，所以其具體代表意象是 Virtual Platform。

軟體即服務

「Software as a Service」名詞最早出現在 2001 年 2 月由美國軟體與資訊產業協會(SIIA)所發布的策略性白皮書中。書中指出網際網路將改變軟體產業的生態，未來應用軟體或資訊服務都將部署在集中化的網路資料中心，並由主要廠商提供使用者定期支付租賃費用的存取和使用方式。雖然概念上和公用運算非常相似，但 SaaS 標榜的是軟體服務帶給企業的應用價值，所以其定價模式和依使用量多寡計費的公用運算有著明顯的區隔。SaaS 提供的線上軟體應用，強調的是延續既有桌面端軟體的使用經驗，其具體代表意象是 Virtual Application。

雲端運算

雲端運算是建構在公用運算和 SaaS 等商業化服務基礎，往後發展的下一代網路運算和資料中心概念。

雲端運算核心技術剖析

Hadoop 源起

Hadoop 運算架構最早起源於 Apache.org 在 Lucene 下的一個子專案 – Nutch，由 Dong Cutting 所開發。其內容是利用 Java 設計的高效率 Full-Text Search 引擎，讓搜尋的速度比傳統逐字比對方式加快許多。

Hadoop 架構

Hadoop 運算架構主要包含三個元件：HDFS 檔案系統、Hadoop MapReduce 模式、以及資料庫系統 (HBase) 。以下將深入解析 Hadoop 運算架構的組合元件：

1、HDFS 檔案系統

當一個資料集(Dataset)的資料量過度龐大、超過單一台實體機器的儲存空間時，就有需要先將這些資料分割再存放於多台分散的機器之中。這種透過網路來管理儲存資料的檔案系統，就稱為分散式檔案系統。而 Hadoop 分散式檔案系統(Hadoop Distributed File System, HDFS)或一般稱為 DFS 就是泛指此類檔案系統，可用來實現 Google MapReduce 的雲端運算。

2、HBase

HBase 是仿傚 Google BigTable 格式、以 Java 撰寫的分散式結構化的資料儲存系統。HBase 與 BigTable 一樣，有別於一般資料庫系統使用的 Row-Oriented 儲存方式，這兩個系統都是採用 Column-Oriented 的儲存方式。Column-Oriented 的優點是每一筆 Record 可以存放不固定欄位的資料，減少像 Row-Oriented 增加新欄位的複雜度。

3、Hadoop MapReduce

Hadoop MapReduce 主要參考 Google MapReduce 的基礎應用程序, 是能夠在叢集電腦架構下處理大型資料量的平行運算方法。MapReduce 將要執行的任務, 拆解成 Map 和 Reduce 兩個工作流程來執行, 以達到分散運算的效果。分散式的計算過程是先用 Map 程式, 來計算出所有網頁中, 每一個字的位置; 接著再使用 Reduce 程式, 在每一個字的清單中, 篩選出單詞, 以及所對應的網頁網址。

商業應用模式

現在 cloud services 主要有三大商業應用模式, 分別是 Software-as-a-Service(Saas)、Platform-as-a-Service(Paas)、Infrastructure-as-a-Service(Iaas)

1、Software-as-a-Service(SaaS)

將商用軟體 Host 在資料中心, 以網路存取的方式提供服務。依訂閱(Subscription)或計次付費(Pay-per-use)的型式收費。Saas 的廠商包括 Oracle、Salesforce.com 的 SFDC CRM、BroadVision、Microsoft 的 Web application 等。

2、Platform-as-a-Service(Paas)

有兩種型態, 第一種是藉與 ISV 合作, 經營 AP 市集平台, 依流量、GB 使用量向 ISV 收費(ISV 賺取用户租賃費用)。第二種是提供 APIs 給企業用户開發 Private Cloud 資訊系統。Paas 的廠商包括了 Microsoft 的 Azure、Google App Engine、Salesforce 的 Force.com 等。

Google 日前公開資料中心的秘密: 每台伺服器配一顆 12 伏特電, 確保萬一主電源斷電時還可持續供電; 從 2005 年起, 其資料中心加入了標準貨櫃模組, 每個都有 1,160 台伺服器, 耗電量可達 250 千瓦特, 目前已是第七代的設計。

3、Infrastructure-as-a-Service(Iaas)

企業用户可以設定所需的虛擬作業運算環境: CPU、OS、Storage, 並依 CPU／Hr、GB 或 Gbps 使用量計費。Iaas 的廠商包括了 Amazon Web Service、IBM 的 New Enterprise Data Center(NEDC)等。

Amazon Web Services

Amazon 簡介

Amazon 是美國最大的一家網路電子商務公司, 創立於 1994 年 7 月, 是網路上最早開始經營電子商務的公司之一, 長久以來在線上書店及商品零售的電子商務市場居於領先地位。

2006 年起, Amazon 跨足雲端運算服務(cloud services), 提供企業用户使用 Amazon.com 的 AWS 服務, 讓他們可以在平台上自行建置網路系統服務, 意外成為目前雲端運算服務的代表廠商。

AWS 簡介

AWS 全名 Amazon Web Services, 是一些 remote computing services (也稱為 web services)所組成的集合。透過網際網路(Internet), Amazon 為開發者提供了本身擁有的後端技術平台及 AWS 軟體開發套件(SDK), 利用此技術平臺, 開發者可以在網路上自行開發所需的解決方案。

後來 Amazon.com 更進一步的將 AWS 平台開放給一般企業民衆, 讓用户可以訂閱需要的網路運算資源, 如 COU、儲存空間、伺服器 kernel 等, 演變成現在 IaaS 的商業模式。

AWS 服務

在 AWS 上建構功能, 可以簡單實現複雜的企業應用程序。這些 Web 服務本身位在 Amazon 的雲中, 任何能夠連上 Internet 的人都可以使用它。使用時只需根據已使用的資源付費, 不需任何預支付費。且因為硬體是由 Amazon 維護, 所以也不需要承擔任何維護費用。

由於利用 AWS 可以在幾分鐘內快速的獲得一個虛擬的基礎建設, 小公司不用額外花費好幾周的時間去架構基礎 IT 建設, , 此舉大大地降低了 Web 環境中的「貧富差距」。重要的是這個基礎建設是彈性的, 可以根據需求擴展和收縮, 世界各地的公司都可以使用這個彈性的雲端運算服務。

AWS 主要的五個核心業務

在這些服務之中，我們特別關注滿足幾個大多數系統的核心需求的服務：儲存、計算、訊息傳遞和資料庫。

Amazon Simple Storage Service（S3）

以網路介面提供使用者網路儲存空間的服務，儲存的對象大小可由 1KB 到 5GB，可藉由 HTTP 或 BitTorrent 進行傳送。

Amazon Elastic Compute Cloud （EC2）

提供一個虛擬的電腦環境，使用者可以自訂 Amazon Machine Image(AMI)，例如虛擬作業系統，記憶體需求等設定，標榜使用多少就支付多少（Pay-as-You-Go）的彈性訂閱服務。

Amazon Simple Queue Service（SQS）

用户可以創造一個無數目限制的 SQS，每個 SQS 可以接收/傳遞訊息，每則訊息可保留四天。最大的好處是不需安裝任何軟體，或是限制網域、設定防火牆等繁雜步驟，即可直接透過 Internet 在電腦間做訊息傳遞。

Amazon SimpleDB

提供簡易的結構化資料儲存服務，可以建立類似表單的 Domain，裡面可隨意新增及儲存 Item。完全不需要定義 Schema，系統會自動資料索引的 API 去進行資料存取的動作。

Amazon CloudFront

提供簡易的內容物件傳送(Content Delivery Network, CDN)服務，不論是聲音或是影音檔案，皆可自動導引至最近的伺服器位置，並以高效率低延遲的方式傳輸。

上述的 Infrastructure Services 彼此可以互相搭配使用，且這些服務都在 Amazon 自家的環境中運行，它們之間的通信相當快速，因此使得 AWS 成為非常具有擴充性、應用廣泛的雲端運算服務。

應用實例

1、NASA 噴氣推進實驗室的沙漠調查與訓練研究

美國國家航空暨太空總署噴氣推進實驗室開發了一種名為全地形六腿地外探測車的機器人，而操縱全地形六腿地外探測車需要依靠高解析度的衛星影像來做嚮導、定位與狀況認知。為了提高處理衛星影像的效率，噴氣推進實驗室開發了具有平行性的工作流的應用程式，這應用程式必須依靠 Amazon Web Service。

此應用程式以 Amazon Elastic Compute Cloud 上用於雲端技術的工作流編配平台模組 – Polyphony 來開發，Polyphony 為了達成目標能讓資源共同工作。利用 Amazon Simple Queue Service，噴氣推進實驗室的開發者可以用一個類別來部署 Amazon Elastic Compute Cloud 上的大量計算。噴氣推進實驗室在 2010 的現場測試中採用叢集計算環境來處理達到 3.2G 像素的圖像。

2、維基解密

以公開匿名來源的機密文檔而著名的維基解密網站在開始遭受到分散式阻斷攻擊後，曾一度使用 Amazon Web Service 的 Amazon Elastic Compute Cloud 服務架設一個雲端的伺服器並交由 Amazon 代管來維持網站的持續運作，雖然這個措施讓維基解密逃離阻斷攻擊一段時間，但在 2010/12/01, Amazon 以維基解密違反使用規約為由終止對維基解密的服務。許多人相信這與美國政府的施壓有關。