

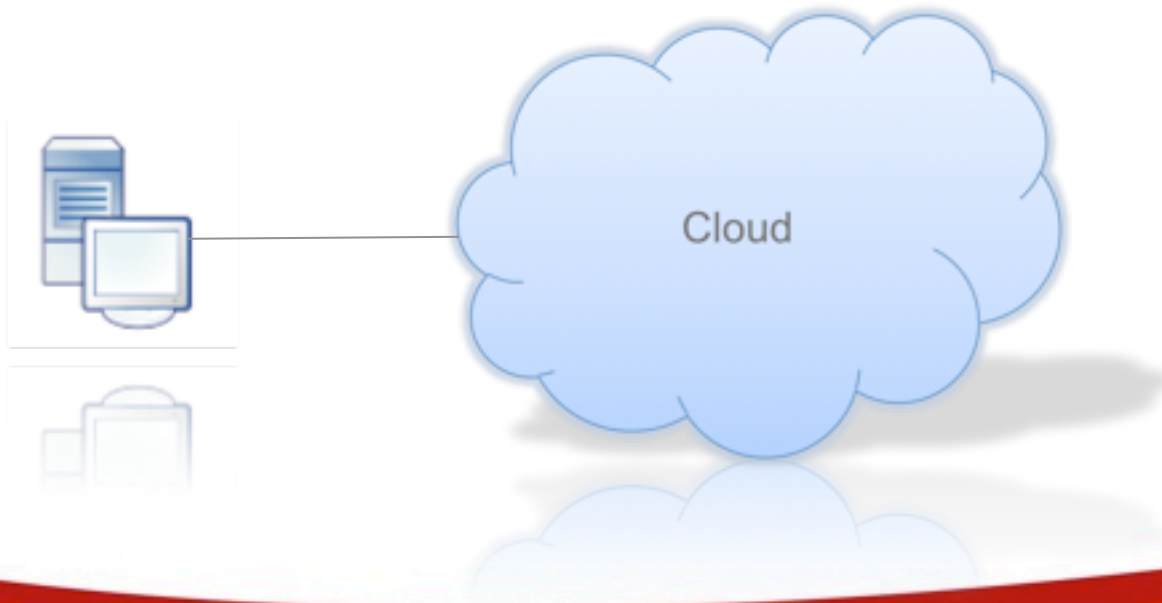
雲端運算簡介

趨勢科技騰雲駕霧程式競賽

趨勢科技研發實驗室

雲端(Cloud)名稱由來

- 繪製網路配置示意圖時(Computer Network Diagrams)，習慣將網際網路(Internet)呈現為一朵雲(Cloud)。
- 劃一台電腦並以線與雲相連，即代表此電腦連接至 Internet，透過 Internet 存取服務(Services)或內容(Content)，比如電子郵件服務(Web Mail)、即時通訊服務(IM)、網頁內容(Web Pages)、線上掃毒。





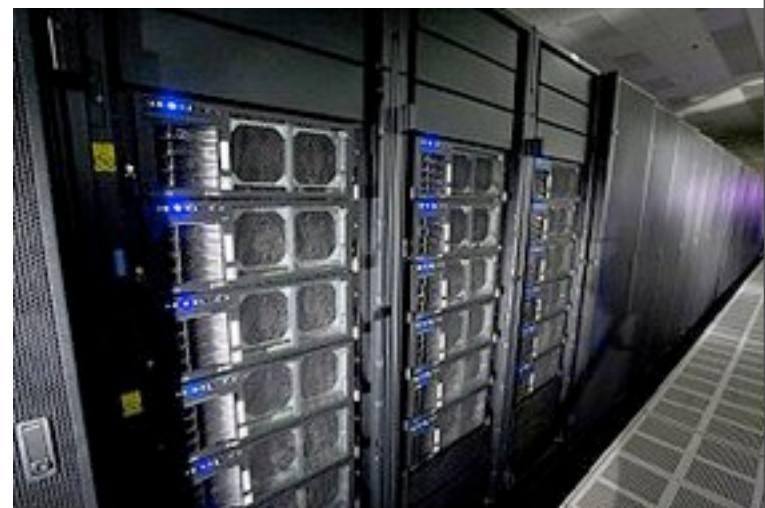
Super Computing Cluster Computing Grid Computing

Grid Computing
Cluster Computing
Super Computing

- 「雲端運算」的基礎概念可回溯至 1960 年，要歸功於電腦科學家 John McCarthy。
- 當時 John McCarthy 提出了「如同公用事業般彙整與提供運算作業」的概念。
- 這個概念的性質類似於 1960 年代的「第三方服務供應商」（即收費提供商業服務的公司）。

超級運算

- 超級電腦是在 1960 年代引進的，主要設計者是「超級運算之父」Seymour Cray。
- 超級電腦是一種具有數十萬個 CPU，並且共享記憶體與 I/O 的超大型電腦。
- IBM Roadrunner 是目前世界上最快速的超級電腦，截至 2008 年 5 月 25 日為止，其運算效能達到 1.026 Petaflop。



叢集運算

- 另一種方法建置超級電腦, ex: Google。
- 讓數百、千部 COTS 低成本電腦透過 LAN 互連以形成叢集（群組），然後將它們配置成看起來像是單一機器。
- 達到高可用性與可擴充性。
- 一般而言，叢集運算的成本效益高出單一超級電腦甚多。

叢集運算 – 程式設計工具

- 訊息傳遞介面 (Message Passing Interface)
 - 用以傳送／接收訊息的 API
 - 專為在大量平行機器與工作站叢集上達成高效能運算而設計。
 - 群組通訊、可擴充檔案 I/O、動態程序管理、同步化。
- 平行虛擬機器 (Parallel Virtual Machine)
 - 在互連電腦上，模擬一般用途異質運算架構。
 - 呈現虛擬處理元素檢視。
 - 建立與管理處理作業、基本訊息傳遞。

叢集運算 – 以 Google 為例

- 2006 年時，估計使用了 45 萬部低成本的普通伺服器。
- 在 Google 的每部機架上，每一側都設置 20 部 2U 或 40 部 1U 伺服器。
- 機架上每一側的伺服器，都透過 100-Mbps 的乙太網路交換器相互連結。
- 機架上的每一台乙太網路交換器都具有一或兩個 Gigabit Uplink，以連結負責連接所有機架的核心 Gigabit 交換器。

叢集運算 – 以 Google 為例



網格運算

- 1990 年代初期，Ian Foster 與 Carl Kesselman 發展出一種全新概念，稱為「網格：全新的運算基礎架構藍圖」。
- 擴充叢集運算技術（其中多重的獨立電腦叢集像網格般運作，因為實質上它們並非位於單一的管理網域中），提供分散式且大規模的叢集運算，以及一種網路分散式平行處理。
- 網格運算與一般叢集運算系統最大的差異在於，網格比較傾向於loosely coupled、heterogonous與geographically diversified。

網格運算

- 做為志願運算 (Volunteer computing)，應用在需要進行大量運算的科學、數學與學術問題上。
- 做為公用運算 (Utility computing)，供商業組織的各種應用程式與商業交易使用，以支援電子商務與 Web 服務。

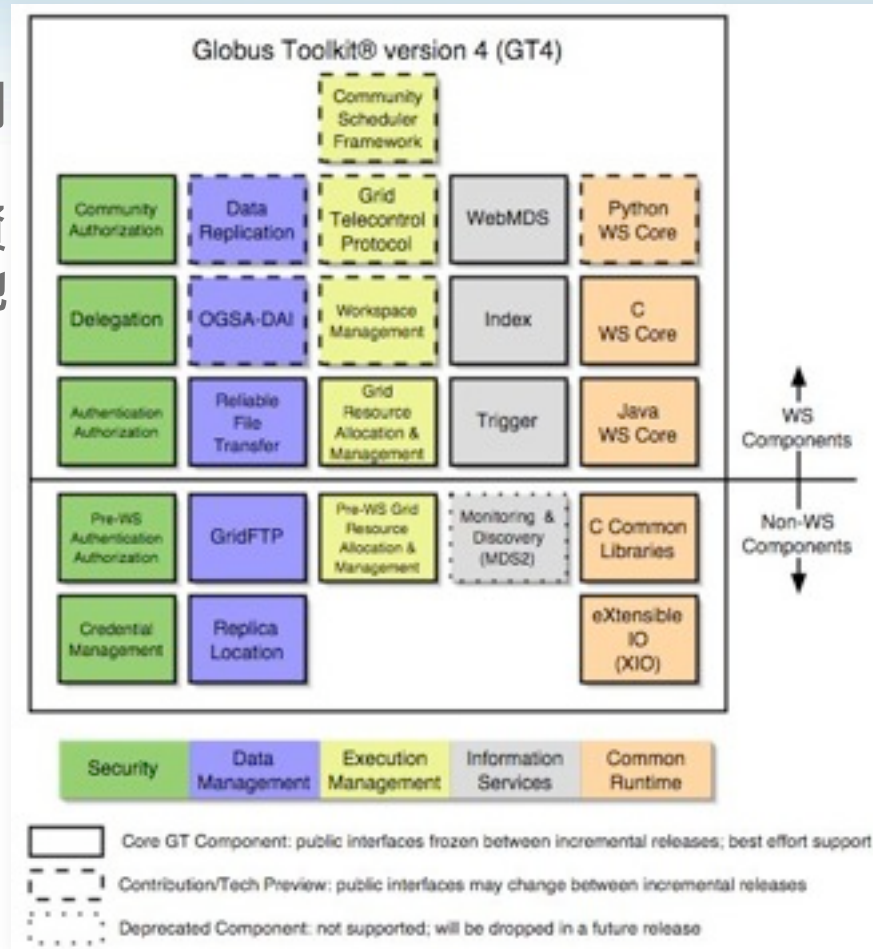


網格運算

- SETI@home
 - 在柏克萊大學的 Search for Extraterrestrial Intelligence 研究計劃中，一項使用網際網路互連以利科學實驗的網格運算專案。
- BOINC
 - 供志願與網格運算使用的非商業中介軟體系統。
 - 讓研究人員得以善用全球個人電腦的龐大處理能力。
 - 截至 2009 年 3 月為止，全球 55 萬部作用中主機的平均處理能力達到 1.5 Petaflop。

網格運算

- Globus Toolkit
 - 基礎的「網格」啟用技術，讓人們得以跨越企業、機構與地理的疆界，安全地線上共享運算能力、資料庫與其他工具，而無需犧牲本地自主權。
 - 開放式網格服務架構
 - GRIP（網格資源資訊通訊協定）
 - GRAM（網格資源存取與管理）
 - GridFTP



Distributed Computing

Distributed Computing

分散式運算

- 讓多部電腦能透過網路一起工作，以解決分散式運算問題。
- 將程序分割為多個部分，然後在多部電腦上透過網路通訊同時執行。
- 一種平行程式設計。平行程式設計通常是指向量資料處理，或多重作業程式設計。

分散式運算

- 分散式問題涵蓋：
 - 大量資料
 - 許多運算作業
 - 可平行執行的應用程式
 - 為數眾多的電腦
- 為全球資訊網編製索引 (Google)
- 資訊安全分析 (Trend Micro)

分散式運算

- 分散式系統

- 一種執行多種通訊協定，以針對網路上多種程序進行動作協調的應用，諸如所有元件共同合作，以執行一組或一個小組的相關作業。
- 容錯
- 高可用性
- 可回復性
- 一致性
- 可擴充性
- 效能預測
- 安全性

分散式運算

- 預期及處理硬體與軟體上的失敗。
- 8 種謬誤：
 - 1. 網路是可靠的。
 - 2. 零延遲。
 - 3. 頻寬無限。
 - 4. 網路是安全的。
 - 5. 拓撲不會變更。
 - 6. 只有一位管理員。
 - 7. 零傳輸成本。
 - 8. 網路是同質的。

分散式運算

- 分散式運算的基本架構之一: Client-Server
- 用戶端透過電腦網路，要求伺服器提供服務。
- 由一群特定類型伺服器提供服務。
- DNS 服務、Web 服務。



分散式運算

- 伺服器在處理用戶端要求時，可能會根據地區或負載平衡提供實際服務。
- 地區
 - 本端機器 -> 本端伺服器 -> 遠端伺服器
- 負載平衡
 - 在服務伺服器之間分配要求工作。
 - 資源使用率最佳化、效能最大化、回應時間最小化。

分散式運算

- 資料複製 (Data Replication)
 - 在多部伺服器上放置多份服務資料
 - 允許存取多重位置
 - 提高可用性
 - 避免 SPOF (單一失敗點)
- 快取 (Cache)
 - 可供快速存取的本端資料複本
 - 在使用之前驗證快取資料，通常是 TTL。
 - 瀏覽器與 Web Proxy 快取經常存取的網頁。

分散式運算

- 透過 TCP/IP 通訊協定進行可靠通訊
- 網際網路通訊協定 (IP)
 - 用來跨封包交換式網際網路進行資料通訊的通訊協定。
 - 定址方法：IPv4 和 IPv6。
- 傳輸控制通訊協定 (TCP)
 - 一種可靠的串流遞送服務，可保證在傳送者與接收者之間依序遞送資料串流。
 - Socket 程式設計
socket()、bind()、connect()、accept()、send()、recv() 等等

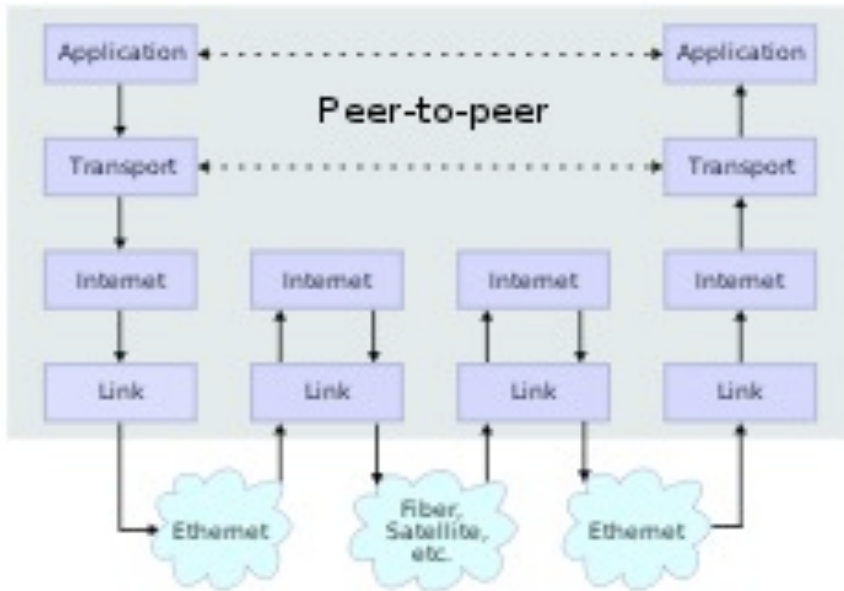
分散式運算

- 應用程式層
 - HTTP、FTP、RPC、SSH

- 傳輸層
 - TCP、UDP

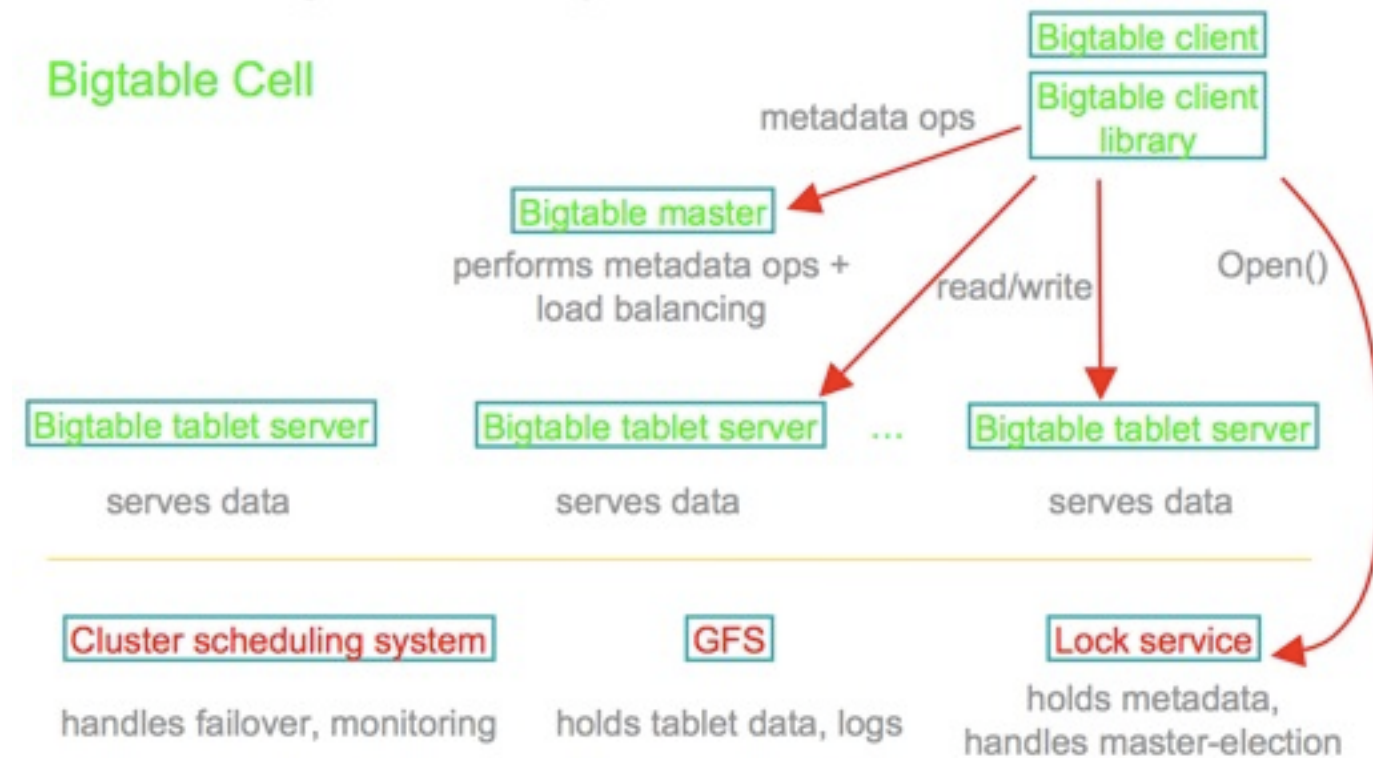
- 網際網路層
 - IP、ICMP

- 鏈結層
 - 實體資料傳輸
 - Ethernet、MAC



分散式運算

- Google Big Table
 - 大規模分散式結構儲存系統



分散式運算

- Hadoop 分散式檔案系統 (HDFS)
 - 偵測失敗與自動回復
 - 大量的串流資料存取
 - 批次處理，而非互動式處理
 - 大型資料集，大型檔案優先
 - Write-once-read-many-times consistency model
 - 移動計算作業比移動資料成本更低
 - 可移植性 (Java)



MapReduce

MapReduce

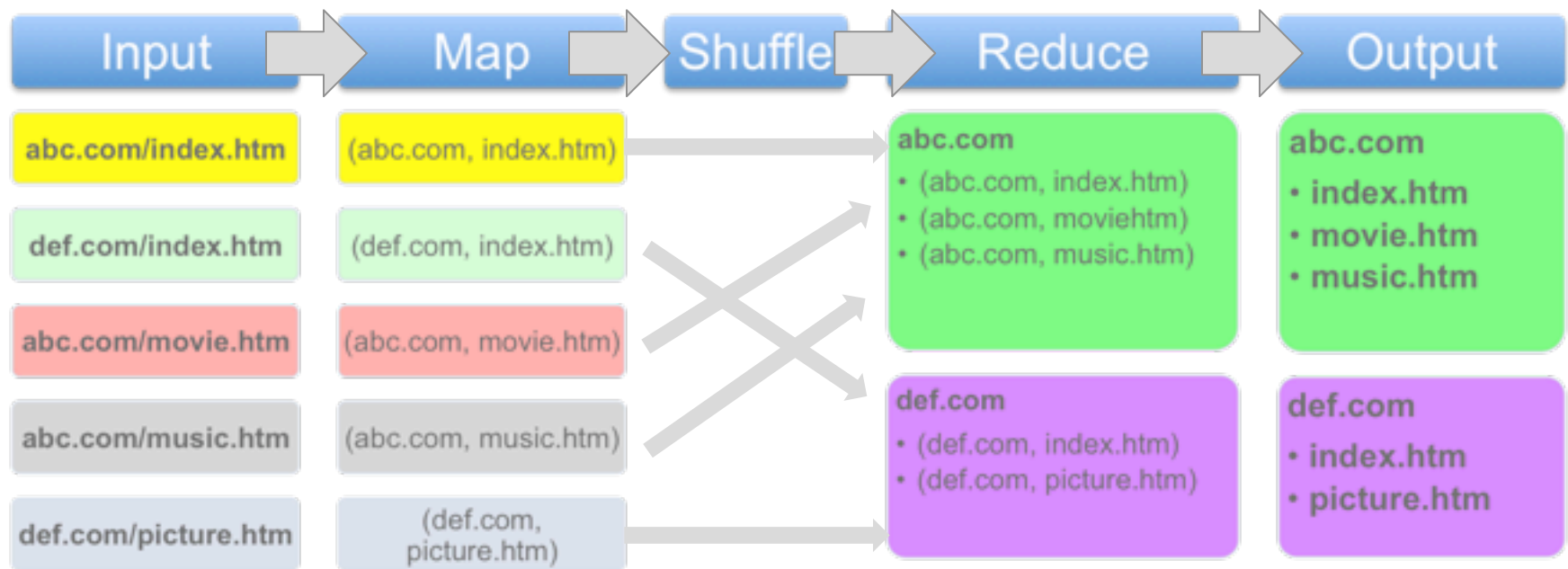
MapReduce

- 一種簡易的程式設計模型，可應用於許多大規模運算問題，並處理大型資料集。
- 在大型普通機器叢集上，自動平行化與分配大規模運算作業。
- 在平行與分散式系統方面毫無經驗的程式設計師，亦可輕鬆運用大型分散式系統的資源。
- 與分散式檔案系統一起執行，即 GFS (Google File System) 與 HDFS。

MapReduce

- 自動平行化與分配
- 透過重新執行達成容錯
- I/O 排程
- 狀態與監控

MapReduce



Concept & Service Models

concept & service models

Web 2.0

- Web 2.0 是一個架構在網際網路上的平台，人與人之間互動和分享而產生出的內容，經由在服務導向的架構中的程式，在這個平台被發佈，管理和使用。其中服務是透過一個應用程式介面(Application Programming Interface, API)來叫用的。
- Web 2.0 是網路使用的一種新的改變，從一系列網站發展到為使用者提供網路應用服務的平台。

Web 2.0 網站



軟體即服務

- SaaS 是一種軟體應用和提供模式。在此模式中，應用軟體是由服務提供商(Service Providers)所控管，將軟體及應用程式以網路服務形式進行，提供使用者和客戶按需(on-demand)軟體應用服務。
- 在使用者及客戶端，無須安裝、維護、更新應用軟體和硬體設備。
- SaaS 的好處包括：更方便管理(Administration)、自動更新及維護(Update/Maintenance)、更好相容性(Compatibility)、節省IT成本(Total Cost of Ownership)、隨時隨地存取(Global accessibility)等等。

軟體即服務供應商

- 趨勢科技 (Trend Micro) 資安代管
 - <http://us.trendmicro.com/us/solutions/software-as-a-service/index.html>



- Google 應用服務
 - <http://www.google.com/a/help/intl/zh-TW/index.html>



- Salesforce.com 客戶關係管理(CRM)
 - <http://www.salesforce.com/tw/>



平台即服務

- 這個觀念由 Salesforce.com 提出。主要是在於提供一個開發平台，給開發人員更大的彈性依不同需求客置化 Web Applications。
- PaaS 是因應 SaaS 而生的應用系統佈建模式。PaaS 模式提供建構與發佈 Web Applications 與服務所需的設備與建置軟體所需的 Life cycle，藉由 Internet 即可使用，無須軟體的下載與安裝，包括開發人員，IT 管理者，使用者等皆可以享用此平台提供的好處。也就是一般所說的 Cloudware。

平台即服務供應商

- Force.com (源自 Salesforce.com)
 - <http://www.salesforce.com/platform/>
- Google App Engine
 - <http://code.google.com/appengine/>
- Microsoft Azure 服務平台
 - <http://www.microsoft.com/azure/default.msp>
- Amazon Web 服務
 - <http://aws.amazon.com/>

基礎架構即服務

- (Originally Hardware as a Service, HaaS) 提供電腦運算基礎設施，通常是一個虛擬環境的平台(Platform virtualization)，作為一種服務。
- 資源包括：伺服器(Servers)、網路設施(Network equipment)、記憶體(RAM)及儲存硬體(Disk)、CPU、資料中心設施、等等。
- 動態資源配置、增加或減少依據應用程式運算資源需求。
- 使用者付費

基礎架構即服務供應商

- Amazon EC2 (Elastic Compute Cloud)
 - 「可在網路雲端提供可靠運算產能的 Web 服務 ...」
 - <http://aws.amazon.com/ec2/>



United States		Europe	
Standard On-Demand Instances		Linux/UNIX Usage	Windows Usage
Small (Default)		\$0.10 per hour	\$0.125 per hour
Large		\$0.40 per hour	\$0.50 per hour
Extra Large		\$0.80 per hour	\$1.00 per hour
High CPU On-Demand Instances		Linux/UNIX Usage	Windows Usage
Medium		\$0.20 per hour	\$0.30 per hour
Extra Large		\$0.80 per hour	\$1.20 per hour

- Small instance (Default) 1.7 GB of memory, 1 EC2 Compute Unit (1 virtual core with 1 EC2 Compute Unit), 160 GB of instance storage, 32-bit platform

Cloud Computing

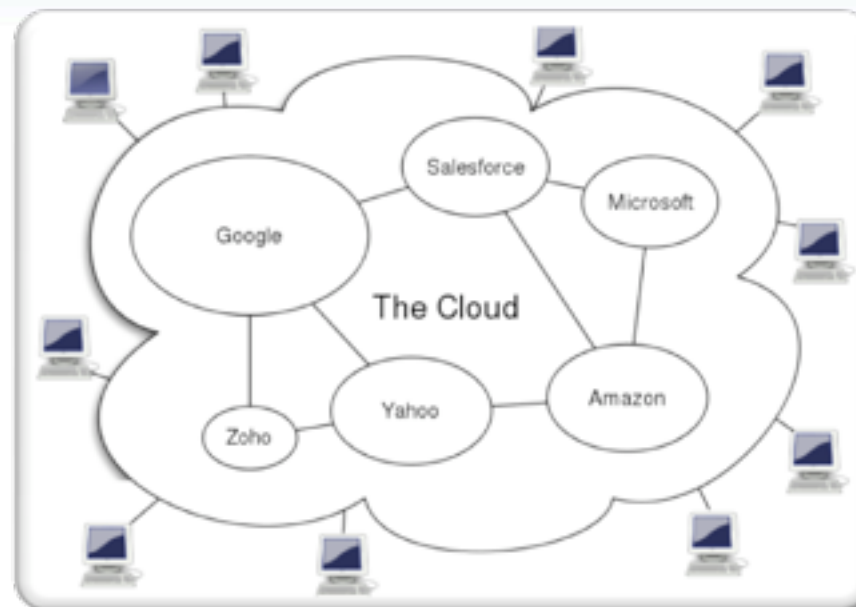
Cloud Computing

Public 2009/5/13



雲端運算

- 雲端運算的概念結合了 IaaS、PaaS、SaaS、Web 2.0、和其它相關技術(像是 MapReduce、Ajax、Virtualization 等等)，共同在 Internet 架構上，來滿足使用者在運算資源的需求。



雲端運算

- 一種分散式運算(Distributed Computing)的模式和概念，透過在 Internet Cloud 中，由動態可擴展性和虛擬化運算資源來提供 Web Services，將龐大而複雜的運算處理程序自動分拆成無數個較小的子程序，交由多部伺服器所組成的龐大電腦集群系統(Cluster)進行分散和平行運算分析後，將處理結果回傳給使用者(Cloud Client)。
- 對於這些雲端技術和基礎設施，使用者無須擁有專業知識和任何控治權。透過雲端運算，Web Service Providers 可以在數秒之內，達成處理數以千萬計甚至億計的資訊，達到和超級電腦同樣強大效能的各式各樣網路服務。

Cloud Architecture

Cloud Architecture

Public 2009/5/13



雲端架構

- 提供雲端運算時，會涉及的軟體系統架構，[...]。 通常涵蓋多重的雲端元件，這些元件會透過應用程式設計介面（通常是 Web 服務）相互通訊。
- 雲端架構會延伸至用戶端，讓用戶端的 Web 瀏覽器和／或軟體應用程式得以存取雲端應用程式。



雲端架構

- 軟體應用程式的設計，可透過網際網路隨需使用服務。以雲端架構做為建置基礎的應用程式，是一種基本的運算基礎架構，有需要時才會使用（例如處理使用者要求）；可以隨需獲取必要資源（例如運算伺服器或儲存設備）、執行特定工作，然後放棄不需要的資源，通常會在完成工作後自我處置。

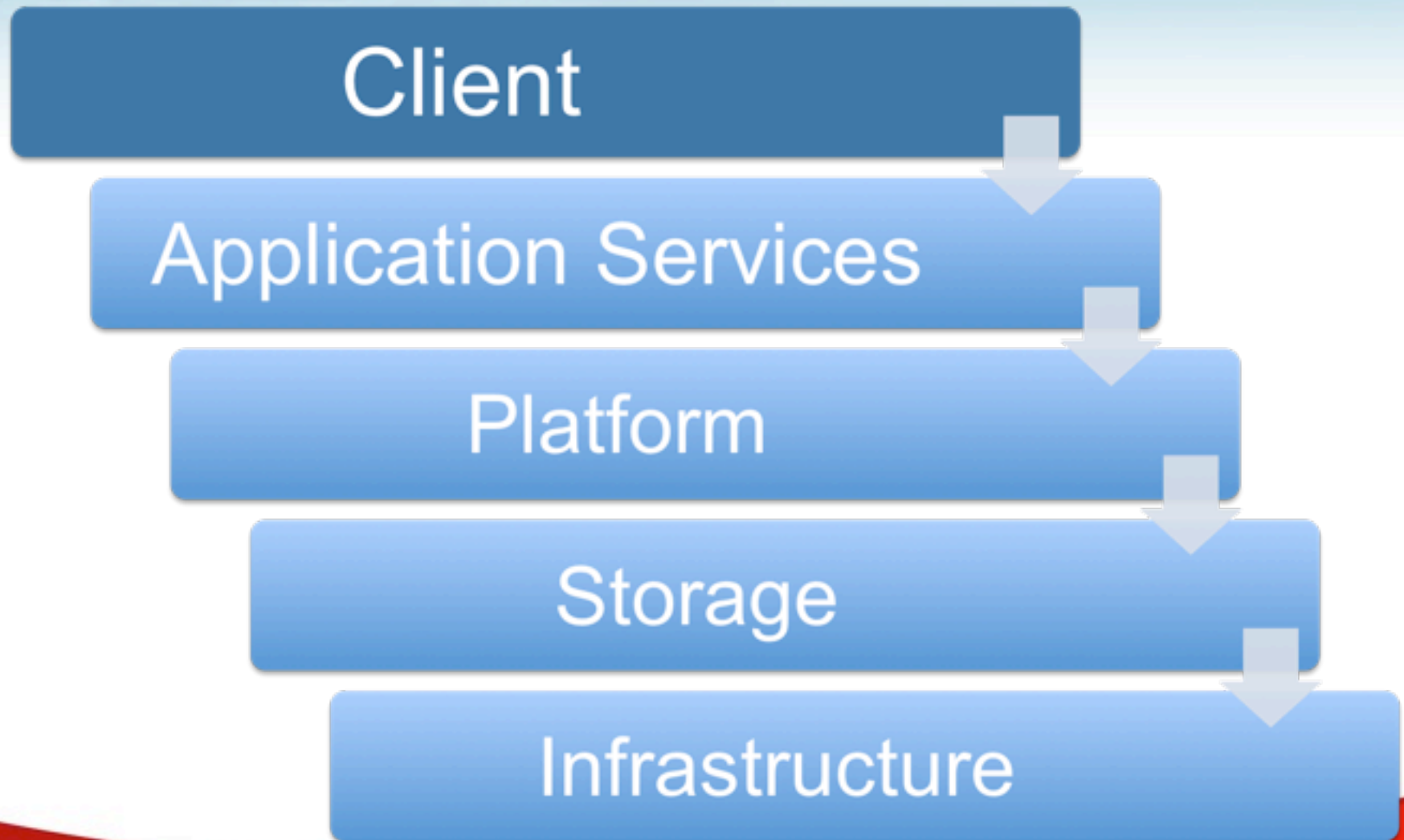


雲端架構

- 解決與大規模資料處理有關的重大難題。
 - 難以取得應用程式所需的大量機器。
 - 很難在有需要的時候取得機器。
 - 難以在不同機器上分配與協調大規模工作、在不同機器上執行程序，並在某一部機器故障時，提供另一部機器以供回復。
 - 難以根據動態工作量，自動向上與向下調整。
 - 難以在完成工作時擺脫這些機器。



雲端架構



雲端架構 - 用戶端

- 仰賴雲端運算架構上應用程式服務的雲端用戶端。
- 瀏覽器
 - Microsoft Internet Explorer
 - Mozilla Firefox
 - Google Chrome
- 行動式裝置
 - Android
 - iPhone
 - Windows Mobile

雲端架構 – 平台

- 用來開發與執行應用程式服務的雲端平台。
- Google AppEngine
 - 可讓您在 Google 基礎架構上，執行 Web 應用程式
 - Java 與 Python 執行時期環境
 - 完全支援一般的 Web 技術與服務
 - 分散式資料儲存庫 (BigTable)
- Hadoop MapReduce
 - 用來執行 MapReduce 應用程式的開放原始碼分散式運算平台
 - 批次處理

雲端運算 - 儲存設備

- 雲端上的分散式持續資料儲存。
- 未結構化的資料儲存
 - 傳統式檔案系統，例如：
 - HDFS
 - Key-Object pair
 - Amazon S3（簡式儲存服務）
- 結構化的資料儲存
 - Amazon SimpleDB
 - Google AppEngine DataStore

雲端架構 - 基礎架構

- 基礎架構即服務
 - 用多少、付多少
 - 隨需使用
- 虛擬化運算基礎架構
 - 電腦叢集
 - 硬體抽象化
 - 虛擬化
 - Xen、VMware
- 可由使用者配置的運算環境
 - 作業系統、網路
 - 記憶體、磁碟、CPU



Smart Protection Network

Smart Protection Network

趨勢科技智慧防護網絡(Smart Protection Network)

- <http://itw.trendmicro.com/smart-protection-network/index.php>
- 近年，在個人電腦及網絡上耗費記憶體的 Virus Patterns 數目正不斷上升。在2005 年，一天約有50 個新 Virus Patterns 增加至 Pattern database。到2008 年，數目已經增加到5,000 個。趨勢科技把大多數的 Virus Patterns 儲存到雲端資料庫，目的是減低儲存在端點的數目，並減少佔用在網絡或端點的頻寬，而且更提供即時的網絡保障。
- 超越傳統的阻擋威脅方法，將安全保護範圍延伸至 Internet，並採用趨勢最新的雲端運算技術將 Internet 上的威脅阻擋在網絡或電腦之外。

趨勢科技智慧防護網絡(Smart Protection Network)

- Smart Protection Network 利用 Reputation Technologies，連結來自網絡、檔案、及電子郵件的威脅，並持續更新雲端的威脅資料庫，利用 Cloud-oriented Architecture 以偵測與分析最新的網絡威脅。
- Smart Protection Network 結合雲端技術，確保用戶無論在何時何地，或是以任何方式上網，均能隨時獲得最新最強的安全保護。

趨勢科技智慧防護網絡(Smart Protection Network)

- 網站信譽服務(Web Reputation Service)
 - 趨勢科技的網站信譽資料庫屬全世界最大之一，其網站信譽技術能夠追蹤網域的可靠程度，方法是以網站的歷史、網址的轉變、以及惡意軟件行為分析所發現的可疑活動等因素，給予一個信譽評分，並阻止使用者進入風險較高的網站。



趨勢科技智慧防護網絡(Smart Protection Network)

- 電子郵件信譽服務 (Email Reputation Service)
 - 電子郵件信譽技術藉著查核已知垃圾郵件來源的信譽資料庫，以及使用時評估發信者信譽的動態服務來驗證 IP 地址。信譽評分是根據 IP 地址的行為、活動規模、與先前紀錄而制定。發信者的 IP 地址一經查明，惡意電子郵件在抵達網絡之前就可被阻擋，並有效防止如：zombies 或 botnet 的威脅進入網絡或是用戶的個人電腦。



趨勢科技智慧防護網絡(Smart Protection Network)

- 檔案信譽服務 (File Reputation Service)
 - 結合網站與電子郵件信譽技術，雲端檔案信譽技術能夠檢查每份流經端點、伺服器、或閘道的檔案，並與防毒模式系統作比較。高效能內容傳輸網絡與區域快取伺服器能確保在檢查過程中的存取時間減到最低。
 - 由於有關惡意軟件的資料已儲存在雲端網絡，所以用戶可以隨時獲得相關資訊。與下載防毒模式檔案的傳統方法比較，檔案信譽技術不但可以節省用戶時間，亦可以大幅減少下載檔案的端點儲存空間。

趨勢科技智慧防護網絡(Smart Protection Network)

- 行為分析的相關性技術 (Correlation technology with behavior analysis)
 - 以行為分析來檢查一系列網絡行為的相互關係，以判斷其惡意程度。獨立的網絡行為可能看似無害，但是將若干行為串連在一起時，則可能發現當中的惡意企圖。透過不斷更新威脅資料庫和比對不同的惡意行為，趨勢科技可以即時提供自動防禦功能。



趨勢科技智慧防護網絡(Smart Protection Network)

- 即時回應及更新功能
 - 可以保持趨勢科技的產品技術和公司24x7 的網絡威脅研究中心之間的緊密聯繫，提供雙向的溝通橋樑。例如，每當在用戶的電腦中偵測到新威脅，便會自動更新公司遍佈全世界的威脅資料庫，避免有更多用戶遇到同一威脅問題。趨勢科技不斷收集及處理由honeypot、客戶、及業務夥伴所組成的環球網絡威脅資料，對最新的網絡襲擊提供自動防禦功能，並實踐「團結就是力量」的保安理念，與其他用戶共同發揮網絡保安的強大功能。



參考資料

- [1] http://en.wikipedia.org/wiki/Cloud_computing
- [2] <http://aws.amazon.com/>
- [3] <http://www.davidchappell.com/CloudPlatforms--Chappell.pdf>
- [4] <http://code.google.com/edu/parallel/dsd-tutorial.html>
- [5] <http://code.google.com/edu/submissions/rutgers/index.html>
- [6] http://en.wikipedia.org/wiki/Internet_protocol_suite
- [7] <http://hadoop.apache.org/core/>
- [8] <http://labs.google.com/papers/mapreduce.html>

<http://www.trend.org/fd/tabid/66/Default.aspx>



<http://www.wretch.cc/blog/trendnop09>