Distributed Systems

Chun-Feng Liao

廖峻鋒

Department of Computer Science

National Chengchi University

Distributed Systems

Cloud Computing

Chun-Feng Liao

廖峻鋒

Dept. of Computer Science

National Chengchi University

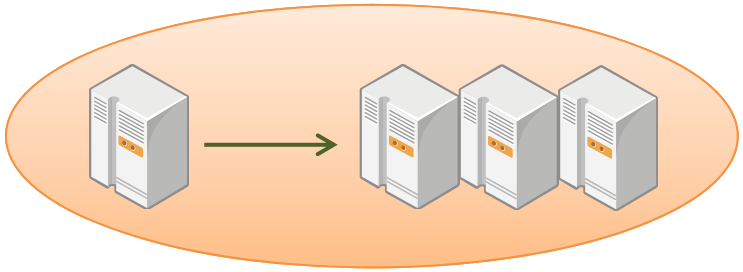
把應用服務都放在Server，就是雲端服務?

google 的東西幾乎是雲端服務(google 文件、試算表...)

很多人用時會配置很多硬體

隨需伸縮

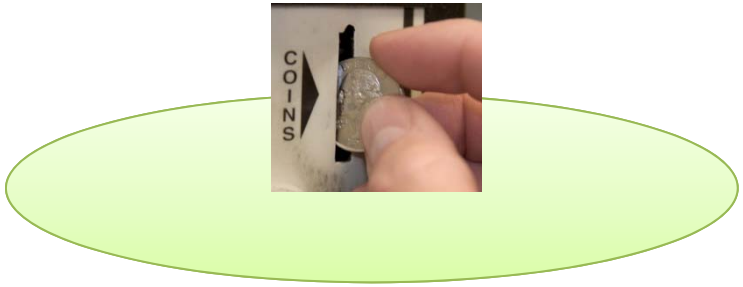
Elastic



如何量測？

以次(量)計價

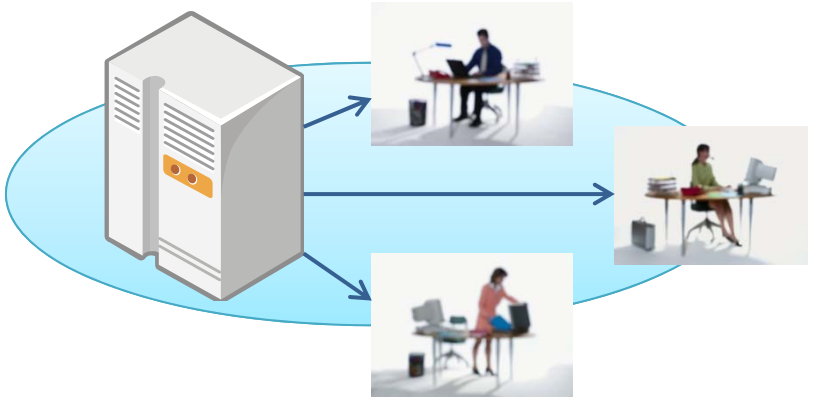
Pay-per-use



每個 user 都以為是獨占服務

多租戶

Multi-Tenant



P. Mell and T. Grance, *The NIST Definition of Cloud Computing*, National Institute of Standards and Technology, 2011

A Brief History

- The idea of computing in a “cloud” traces back to the origins of utility computing:
  - “If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility. … The computer utility could become the basis of a new and important industry.” (J. MacCarthy, 1961)

John McCarthy, 1971 Turning Award, Father of AI

A Brief History

- “As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of ‘computer utilities’ …”.
  - 1969, L. Kleinrock, ARPANET

(Kleinrock的博士論文提供了當代packet network的數學理論基礎)

A Brief History

- Mid 1990
  - Web-based applications become “utilities”
    - Search engines (Yahoo!, Google)
    - On-line E-mail services (Hotmail, Gmail)
    - Open publishing platforms (Blog, YouTube)
    - Social network (Facebook, Twitter, LinkedIn)

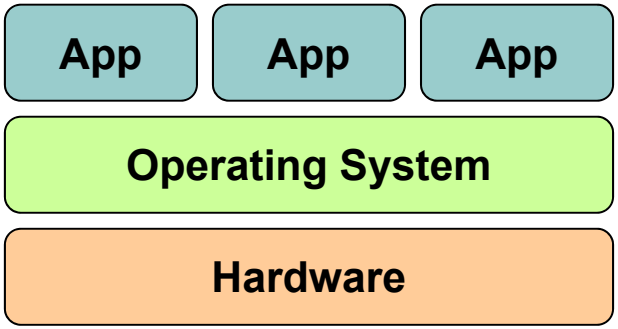
A Brief History

- 1999
  - Salesforce.com pioneered the notion of bringing “utility”-based services into the enterprise
  - Salesforce.com於1999年由當時37歲的甲骨文（Oracle）高級副總裁，馬克·貝尼奧夫創辦
  - 馬克·貝尼奧夫被譽為「軟體終結者」。提出軟體即服務(SaaS) 的理念
  - 獲富比士選為發展最迅速的科技公司，排名僅次Google

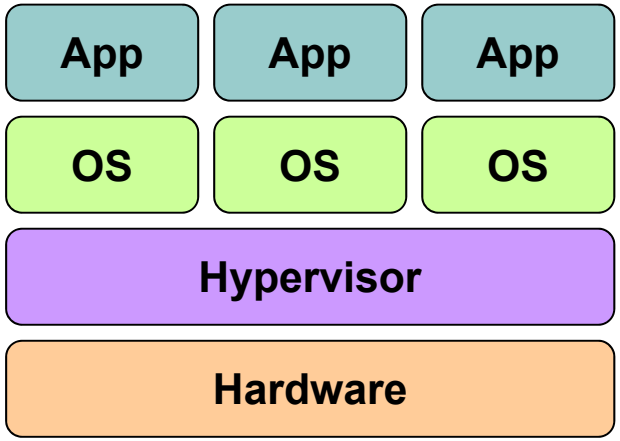
A Brief History

- 2002
  - Xen: The first open source virtual machine monitor
    - Developed in U. Cambridge
    - Largely supported by Intel, IBM, HP, Novell, Red Hat, Sun, and Oracle
    - Virtualization is one of the most important enabling technologies of Cloud Computing

Virtualization



Traditional Stack



Virtualized Stack

A Brief History

- 2006
  - Amazon Web Services (AWS) Platform
    - EC2 (Elastic Compute Cloud, based on Xen)
    - EMR (Elastic MapReduce)
    - S3 (Simple Storage Service)
    - EBS (Elastic Block Store)
    - And more…
  - Offer services using HTTP (REST or SOAP)
  - Billed based on usage

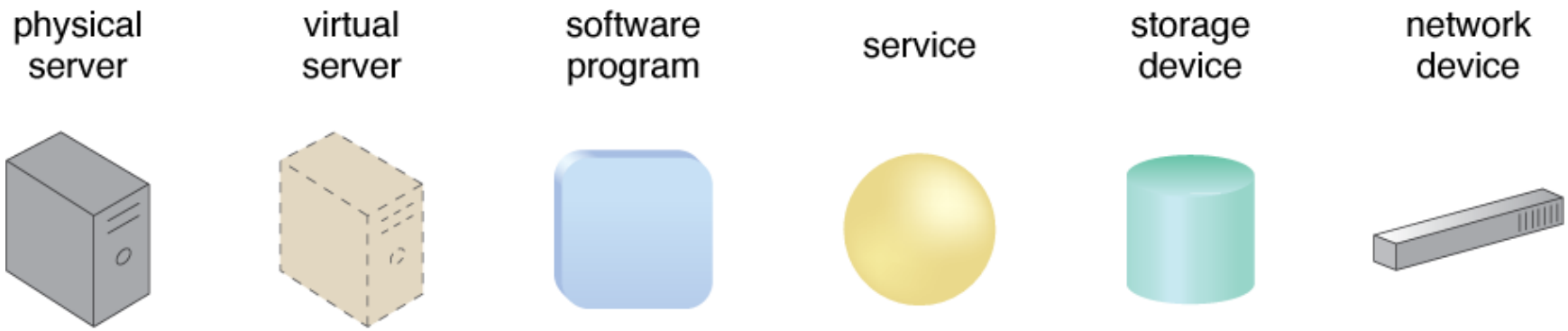
A Brief History

- 2008 Gartner proposes Cloud Computing
  - "…a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies."

Terms

隨需伸縮

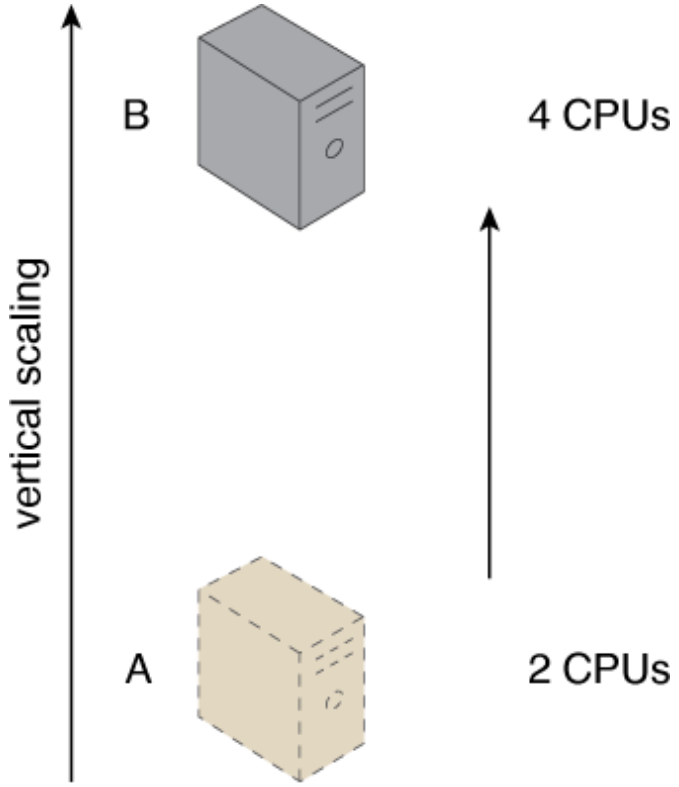
- Cloud
  - An IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources.
- Resource



Terms

升級 CPU

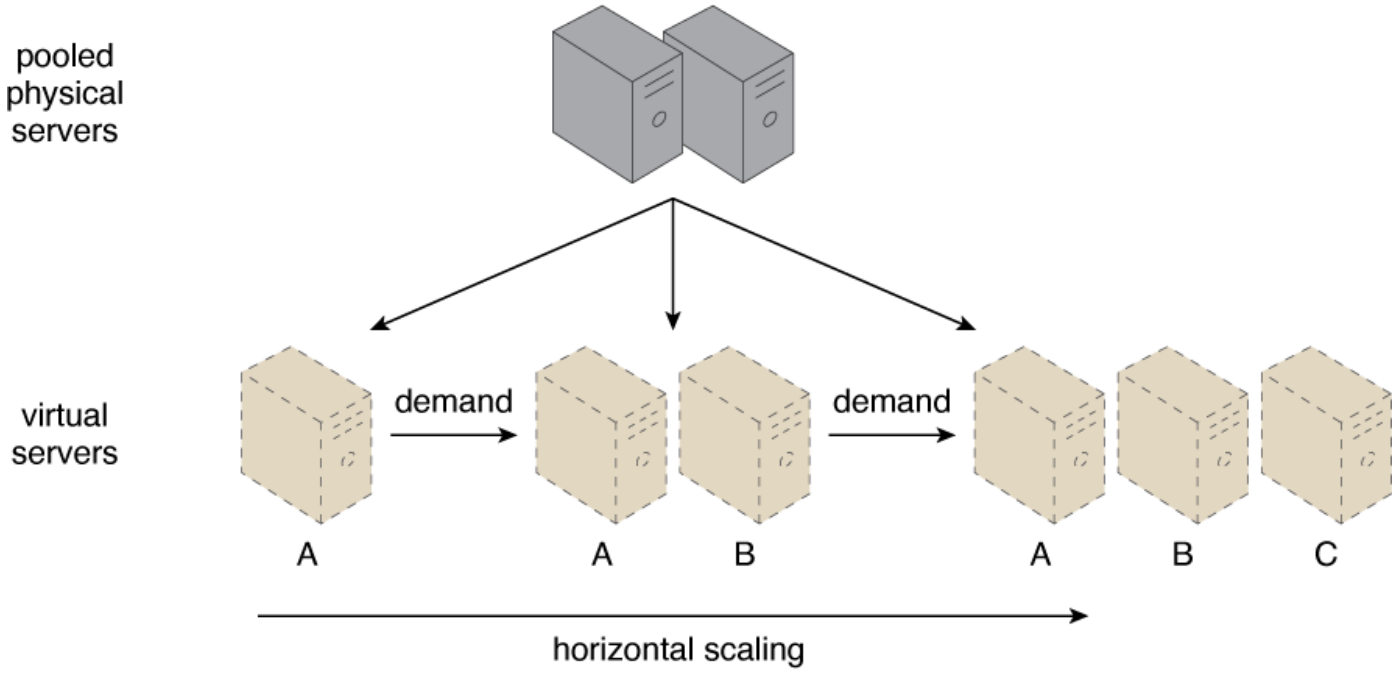
- Vertical Scaling



Terms

多台機器綁在一起
類似 cluster

- Horizontal Scaling



Definition

- 2009 NIST Defines Cloud Computing (revised in 2011)
  - A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort
  - Five essential characteristics
    - On-demand self-service, broad network access, resource pooling, rapid elasticity, measured service
  - Three service models

租用軟體　租 Server　租電腦

    - SaaS, PaaS, IaaS
  - Four deployment models
    - Private Cloud, Community Cloud, Public Cloud, Hybrid Cloud

Cloud: Essential Characteristics

- Rapid elasticity
  - 隨需伸縮，而且要夠快!
- Resource pooling (Multi-tenant)
  - 多租戶 (details are in the next few slides)
- On-demand self-service
  - 可以在要的時候再即時取得資源，而且過程高度自動化，可以由使用者即時線上申請
- Broad network access
  - 透過既有標準協定(如HTTP)，隨時隨地可存取此服務
- Measured service
  - Pay per use

Business Drivers

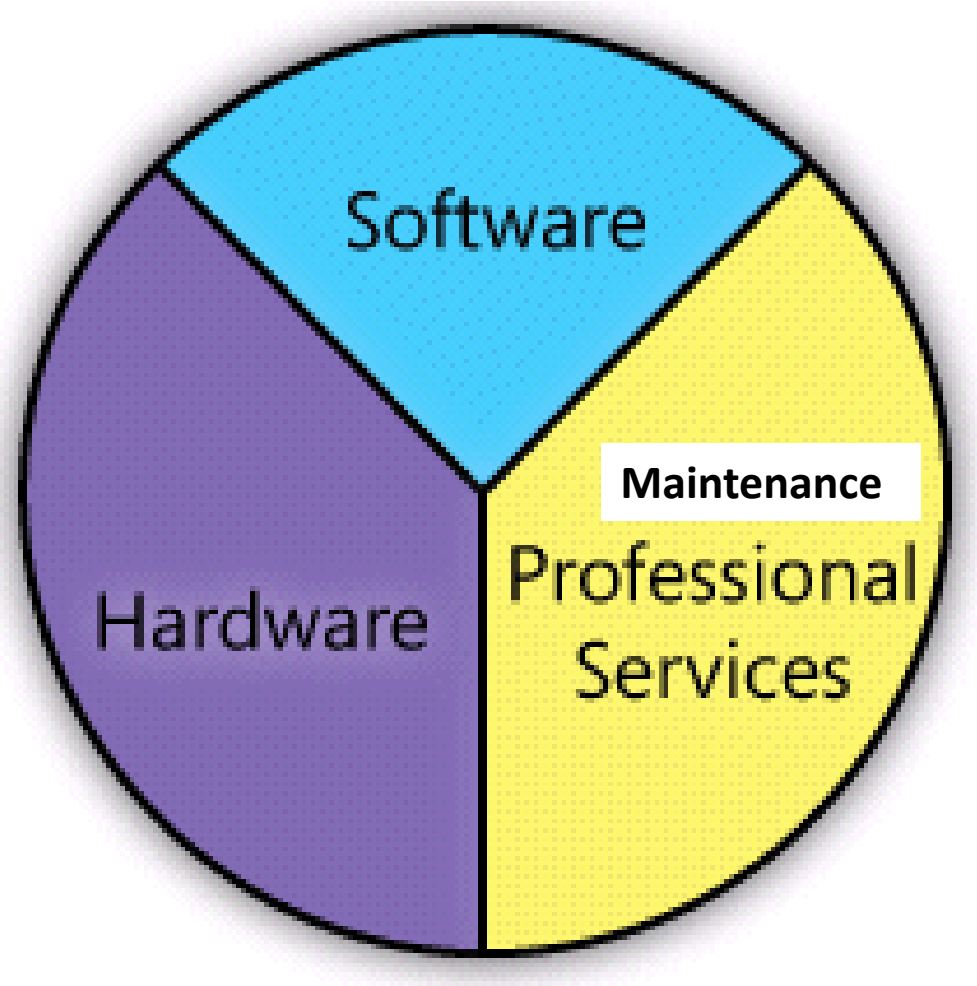
- Capacity planning
- Organizational agility
- Cost reduction
  - Cost of acquiring new infrastructure
  - Cost of maintenance
  - Includes
    - Personnel
    - Upgrades and patches
    - Utility: 電力、機房租金…
    - Security: 保安機制
    - Administrative : 相關行政費用

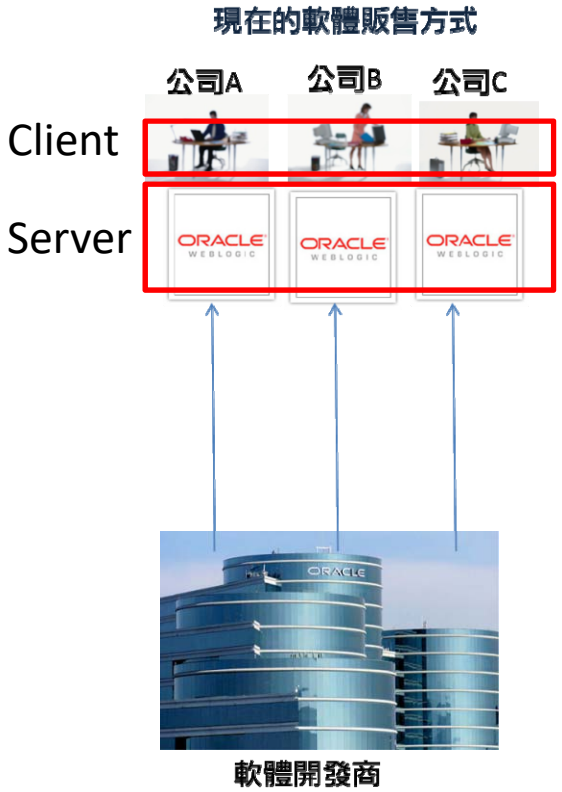
Business Model

- Traditional 系統供應商
  - Users own the software and hardware
  - Users maintains the software and hardware
- Multitenant 服務供應商
  - ASPs own the software and hardware
  - ASPs maintains the software and hardware

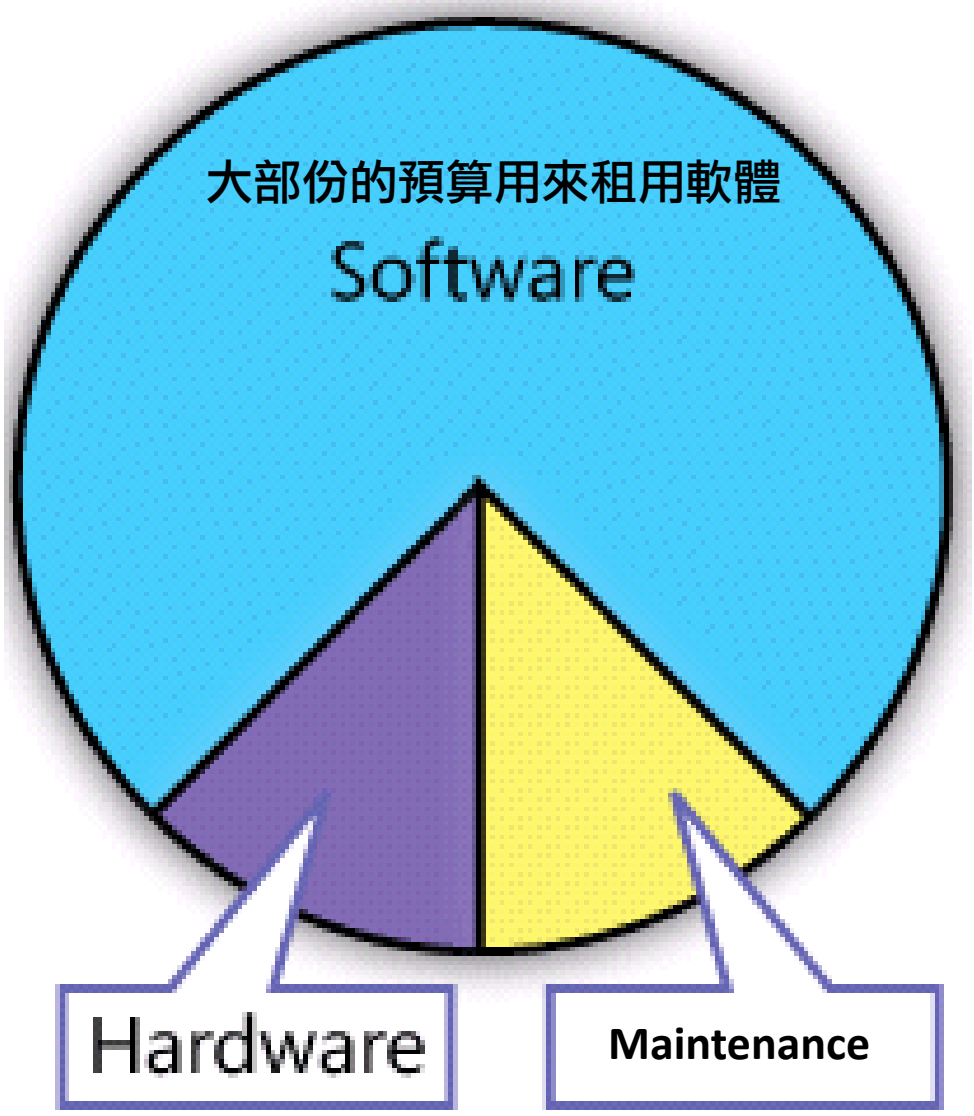
ASP=Application Service Provider

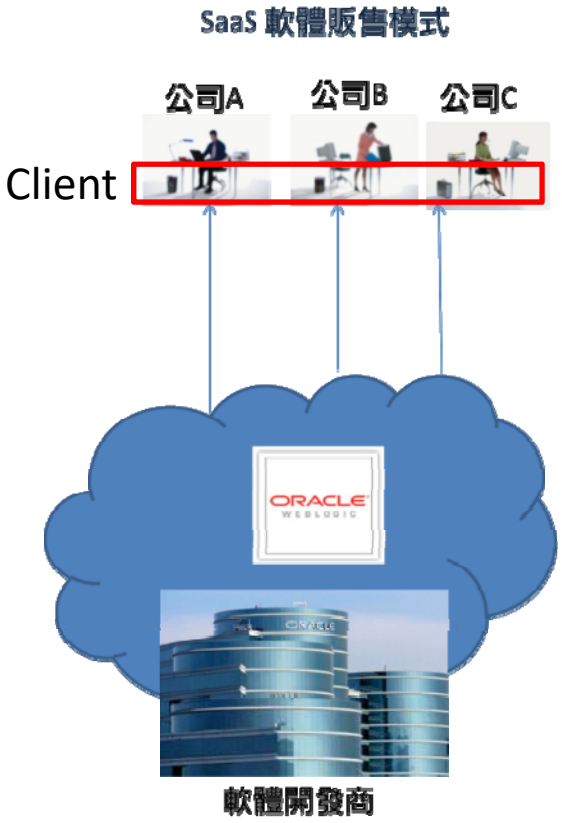
傳統軟體系統預算分配


Software
Hardware
Professional Services
Maintenance


現在的軟體販售方式
公司A
公司B
公司C
Client
Server
ORACLE
WEBLOGIC
軟體開發商

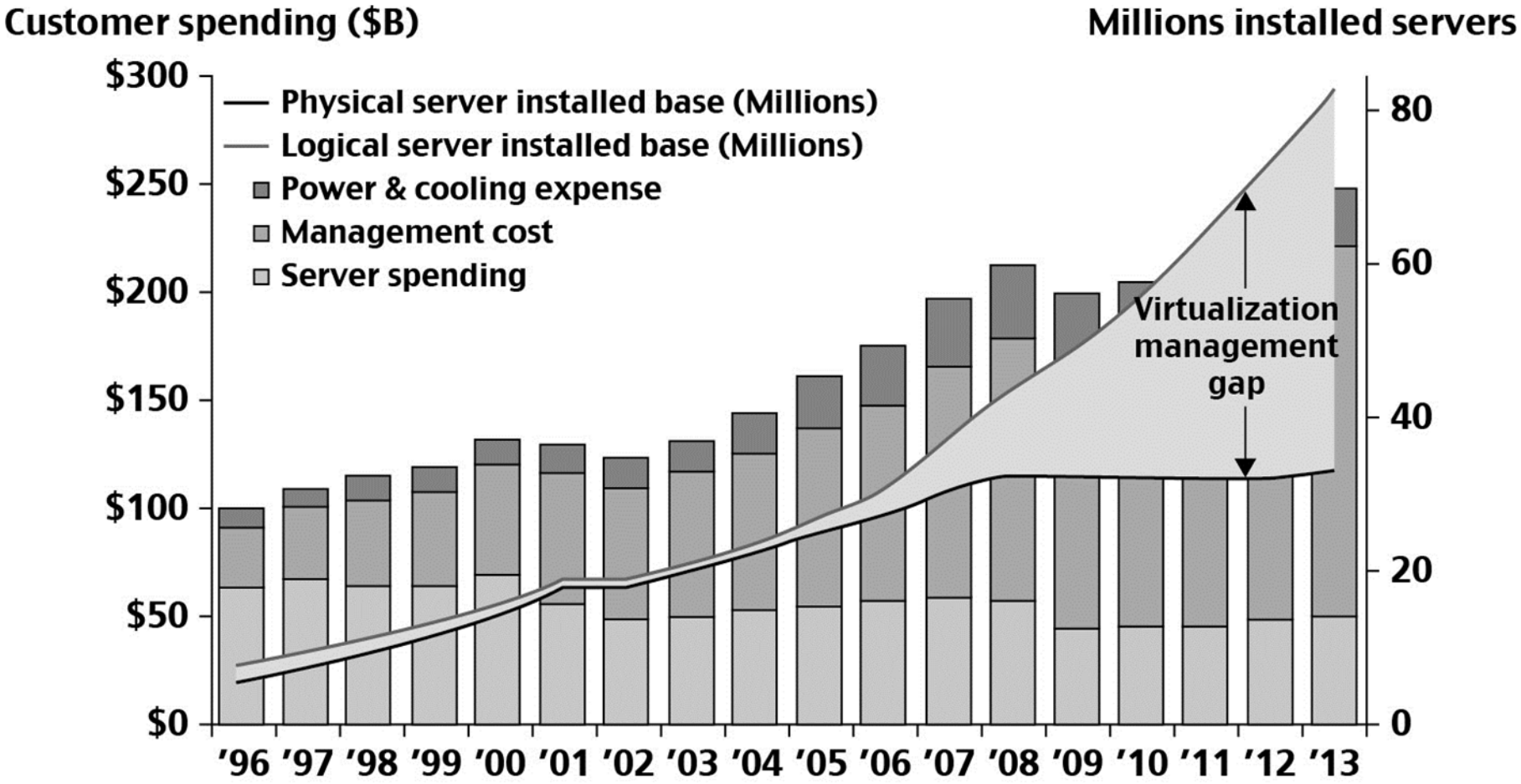
改採SaaS後，公司的軟體系統預算分配





伺服端的軟硬體都放在雲端，所以只需要建置與維護Client端的軟硬體

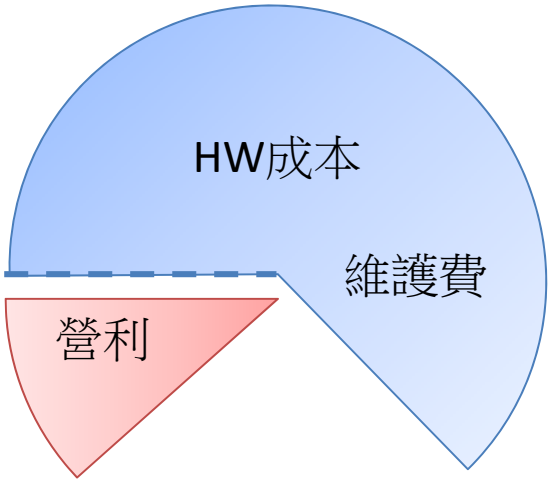
Datacenter and Server Cost Distribution



雲端公司如何營利

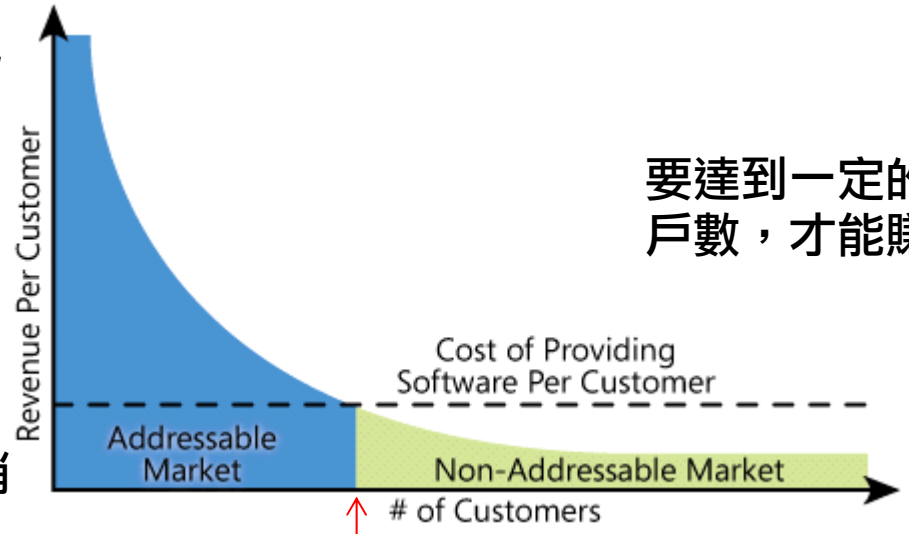
企業目標: 以最少的軟硬體，服務最多的客戶!

所買的軟硬體愈少才能賺錢



一次賺飽

薄利多銷



要達到一定的客戶數，才能賺錢

事實上，交給雲端公司的租金中，大部份是用來支付硬體及維護成本

薄利多銷的生意要達到一定的客戶數，才能賺錢

“多租戶”

雲端公司

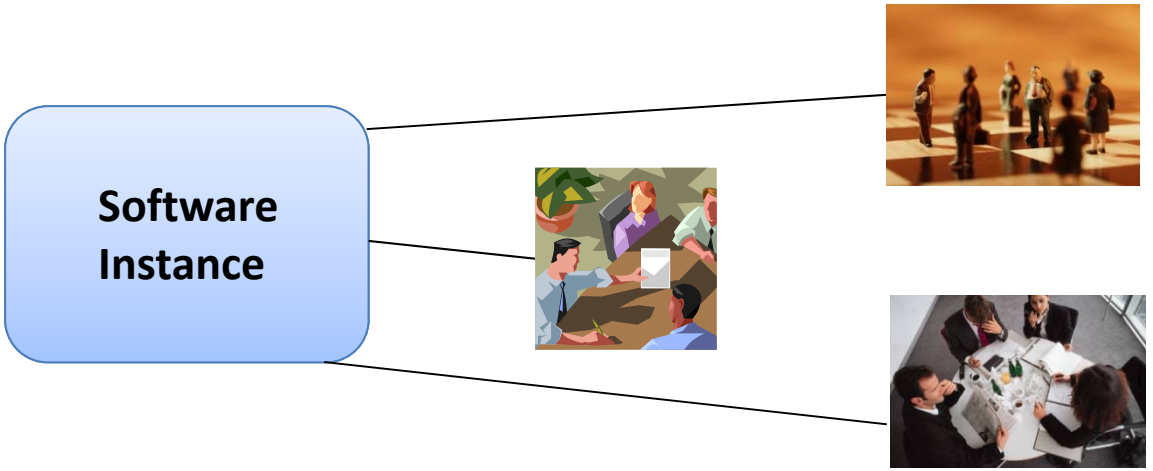
isolate

vs

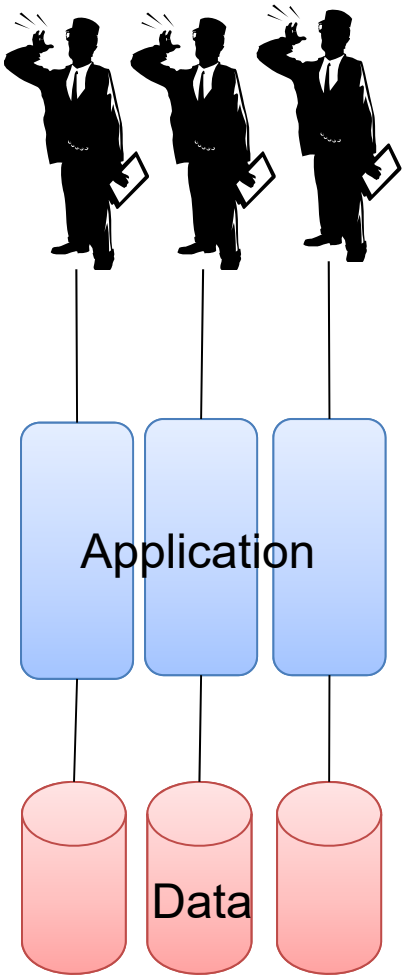
share

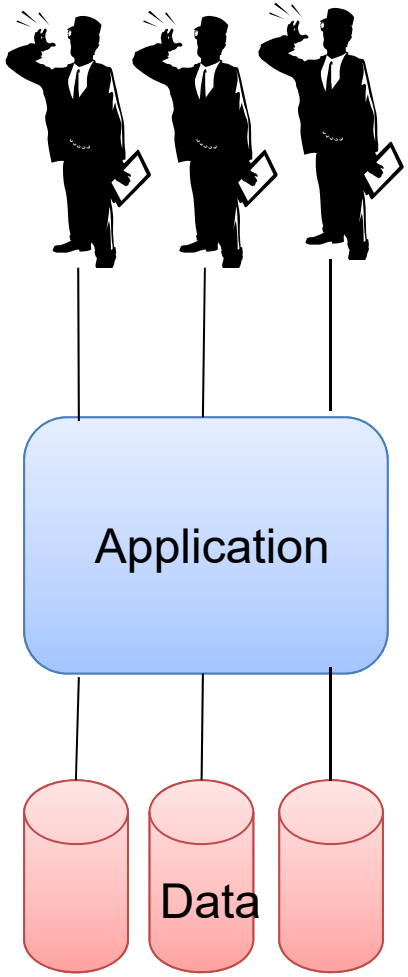
Tenant and Multi-Tenancy

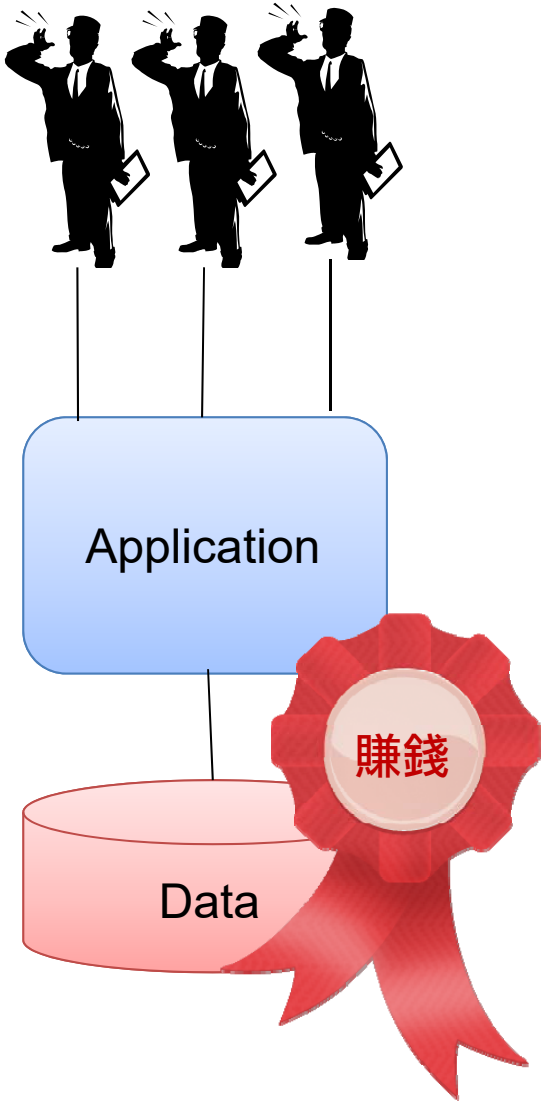
- Tenant (租戶)
  - A small group of users (e.g. a company)
- Multi-Tenancy (軟體支援多租戶的能力)　可以客製化成不同的樣子
  - The ability to deliver software to tenants from a single, shared instance of the software
  - Must be elastic and customizable



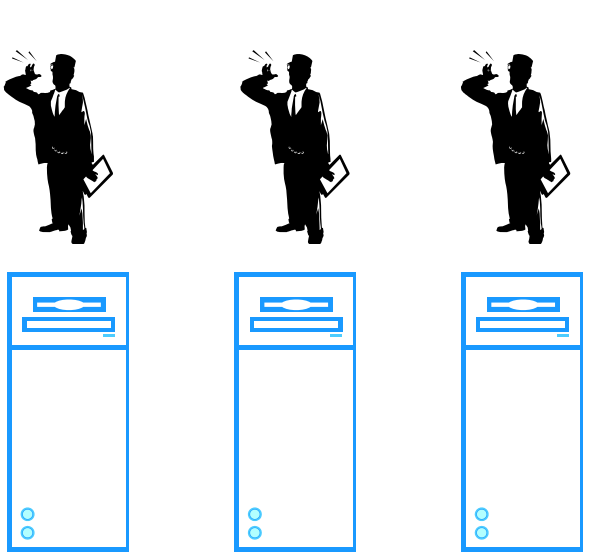
“多租戶化”的程度

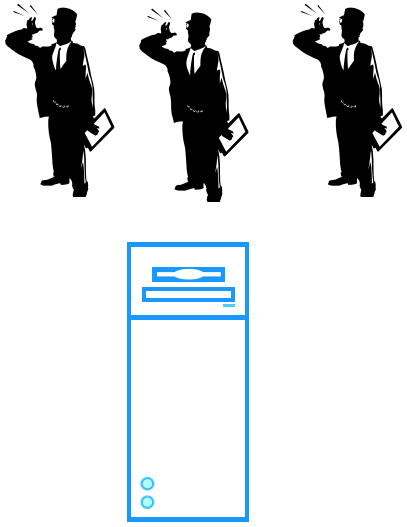






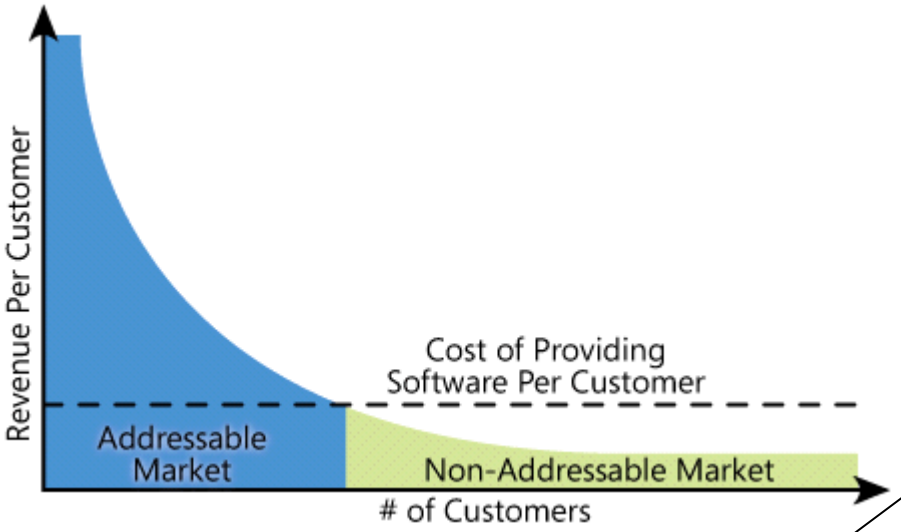
以最少的軟硬體，服務最多的客戶





雲端運算是針對一些中低階客戶

以最少的軟硬體，服務最多的客戶

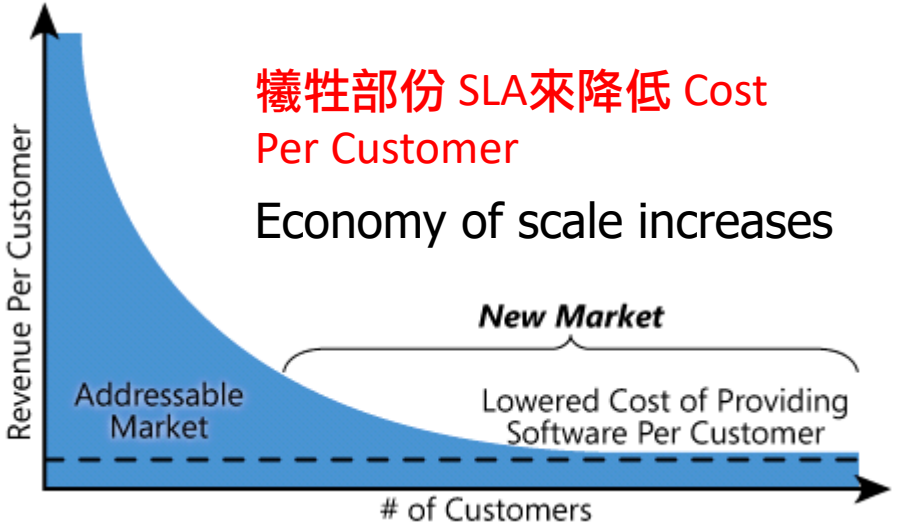


犧牲效能但價格可以壓低

Multi-Tenant Technologies

犧牲部份 SLA來降低 Cost Per Customer

Economy of scale increases



多租戶是雲端運算的宗旨

“Multi-Tenancy is the core tenet of cloud computing”

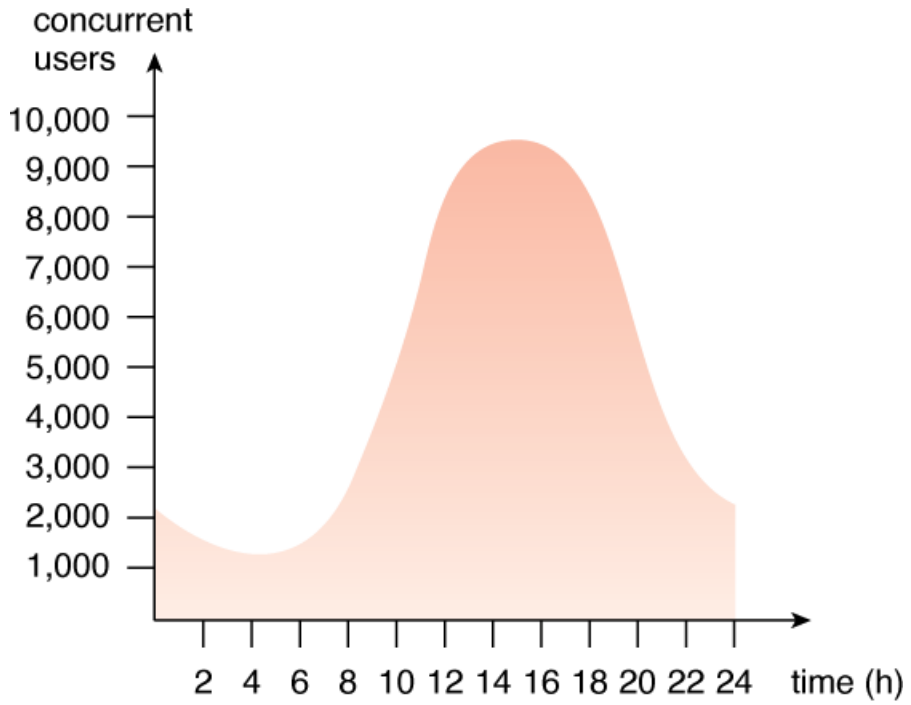
- S. Kajeeeta, CTO, Computer Science Corp.

Cloud: 規模經濟

- 早期工廠皆自行發電，後來發現由電廠統一發電反而降低成本
- 目前資訊系統皆由硬體至軟體從頭建置，如果由雲端服務商統一供應反而可降低成本

Benefits

- 降低投資成本
  - 對於短期或小型系統來說，租用計算資源比較划算
- 彈性的計算資源　可隨需伸縮
  - 如果短期間需求增加，不用特意買進新電腦，可臨時多租用一點計算資源，需求降低後可再釋放
  - 可以隨意隨時變更硬體規格(CPU, RAM, Disk size)，而不用重灌
- 可靠性的提昇
  - 系統可靠性與網路流量壓力規劃由雲端廠商負責



Service Model

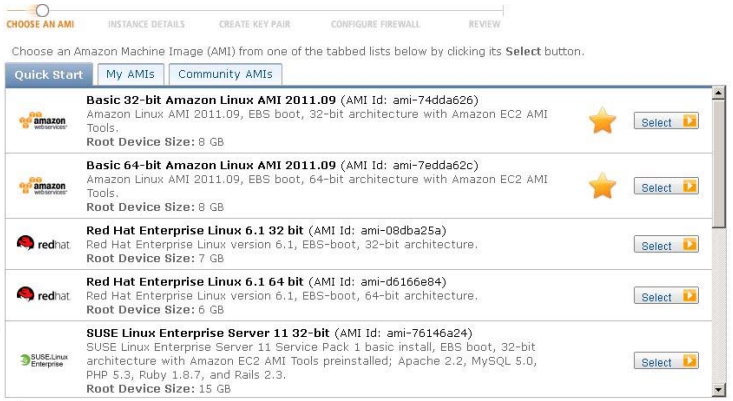
IaaS (Infrastructure as a Service)

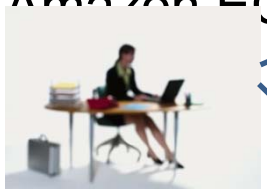
PaaS (Platform as a Service)

SaaS (Software as a Service)

設施即服務模式 (IaaS)

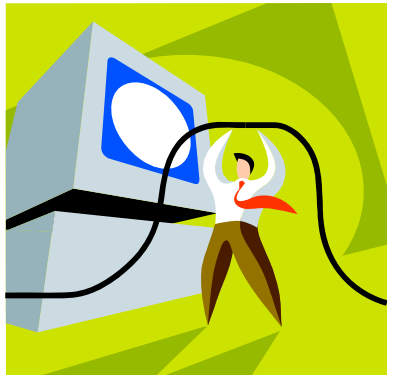
- 廠商
  - 提供虛擬的計算設施 (CPU, 儲存空間)
- 使用者
  - 根據需求訂製自已所需設施
  - 以為自己獨佔這些設施
- 範例
  - Amazon EC2





2 CPU, 1G RAM, Ubuntu…

AMI
(Amazon Machine Image)



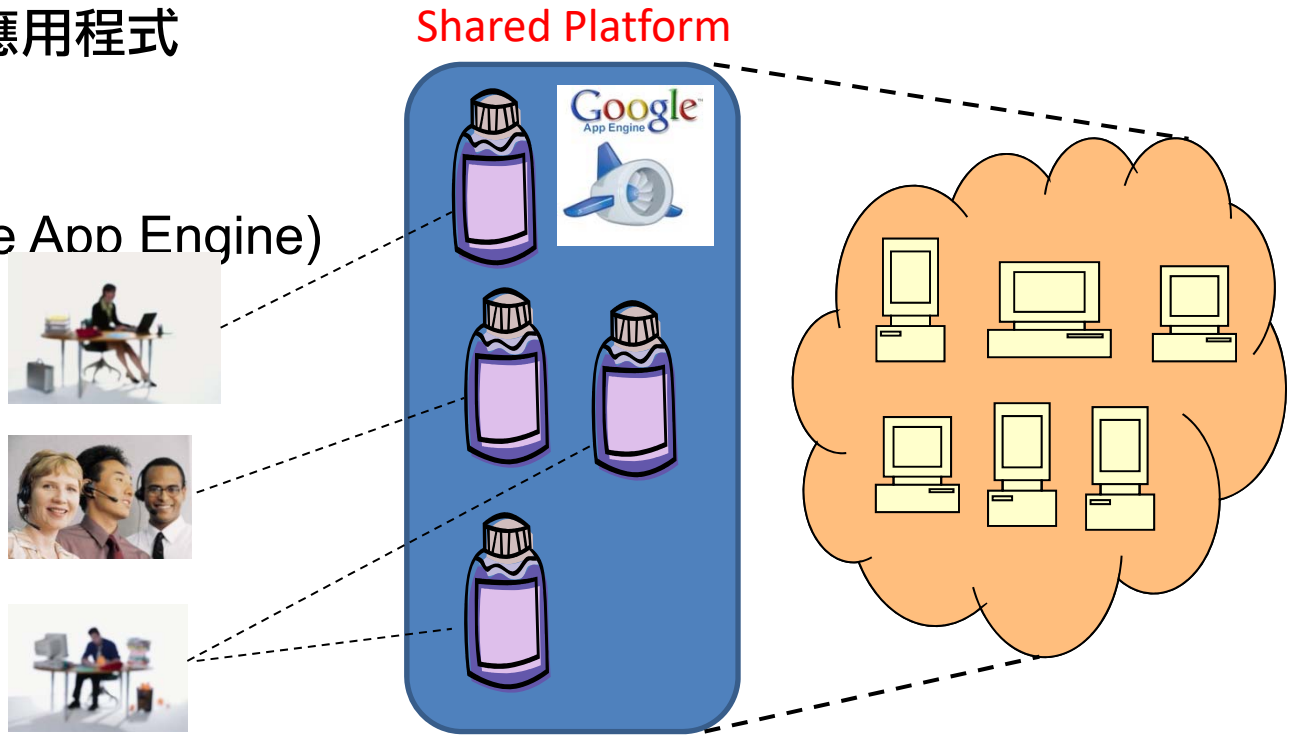
Virtualized Computer (s)

ssh

Shared Infrastructure

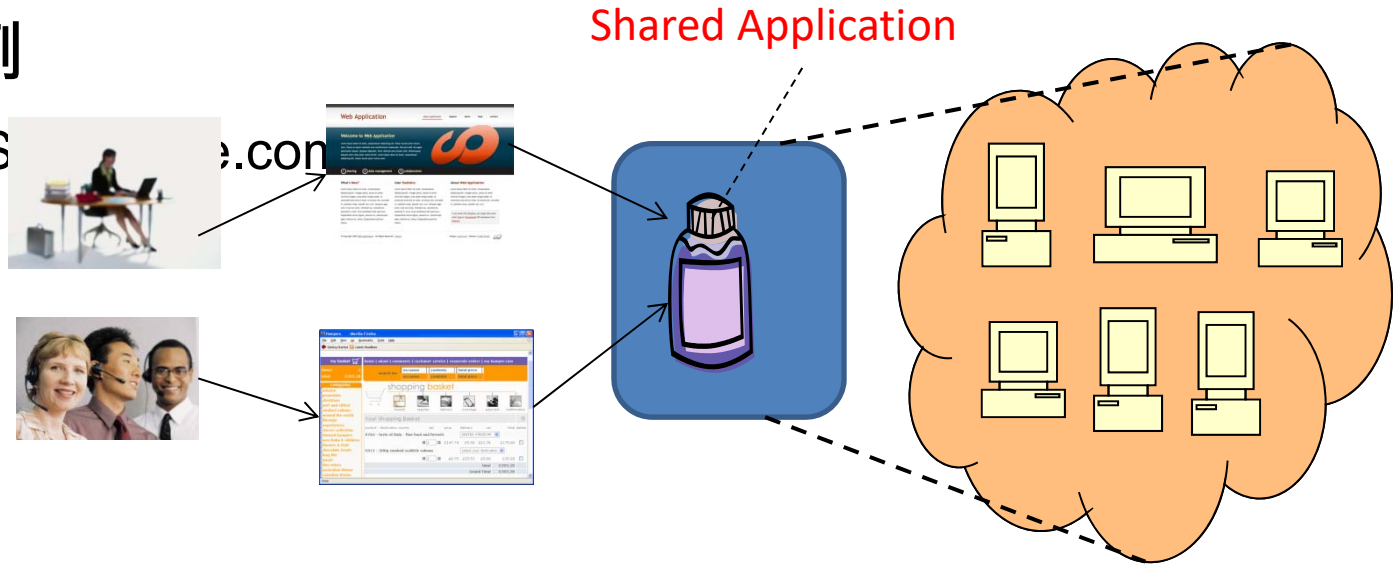
平台即服務模式 (PaaS)

- 廠商
  - 提供應用程式規格或開發工具
  - 符合規格者可佈署至其提供的容器(Container)
  - 可依應用程式負荷，自動調配計算與儲存資源
- 使用者
  - 開發並佈署應用程式
- 範例
  - GAE (Google App Engine)



軟體即服務模式 (SaaS)

- 廠商
  - 提供應用程式本身
  - 提供圖型化介面的應用程式客製化工具
  - 可依應用程式負荷，自動調配計算與儲存資源
- 使用者
  - 使用圖型化介面，客製化自己的應用程式
- 範例
  - S[illegible].com

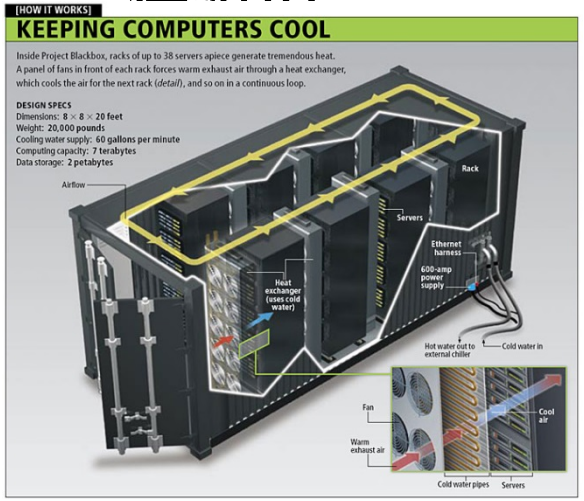


Case: Google Data Center

- 定義
  - A facility to house a large collection of computers, networking, and communications equipment
- Four (computer system) design issues
  - Structure
  - Scalability
  - Flexibility
  - Power efficiency

Data Center

- 伺服器置放於 19-inch rack cabinets (標準規格)
  - 1櫃可放42 1U伺服器
  - 成排單行放置，中間走道相隔
  - 在超大型Data Center中，可以貨櫃模組化



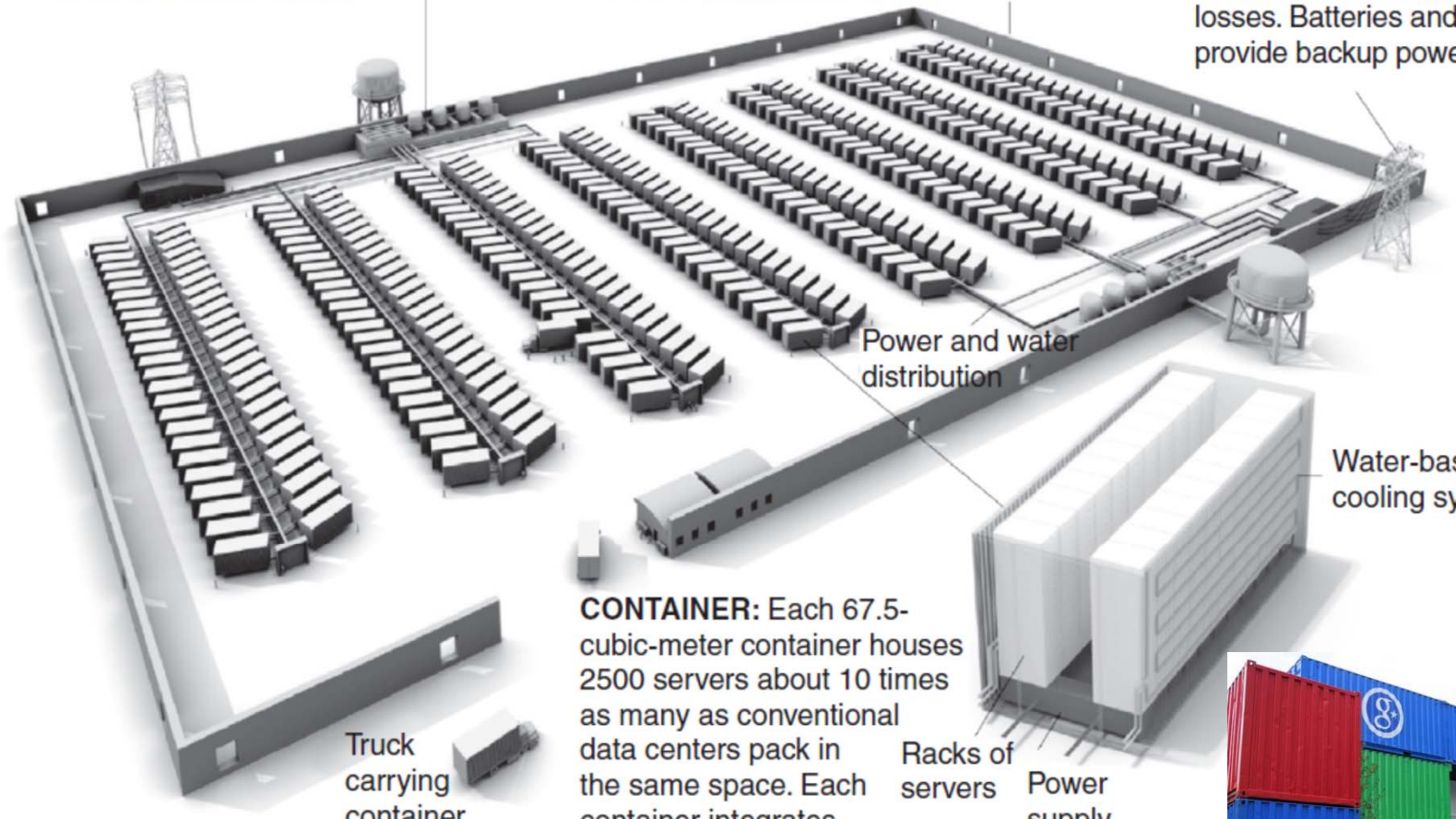
Data Center Requirements

- Unwavering power
  - Should be prepared to handle brownouts (電力短少) and even power outages (斷電)
  - Backup batteries and diesel (柴油) generators must be available to keep power flowing
- Broadband network connectivity
- Security
  - Firewalls, VPN gateways, intrusion-detection software

COOLING: High-efficiency water-based cooling systems–less energy-intensive than traditional chillers–circulate cold water through the containers to remove heat, eliminating the need for air-conditioned rooms.

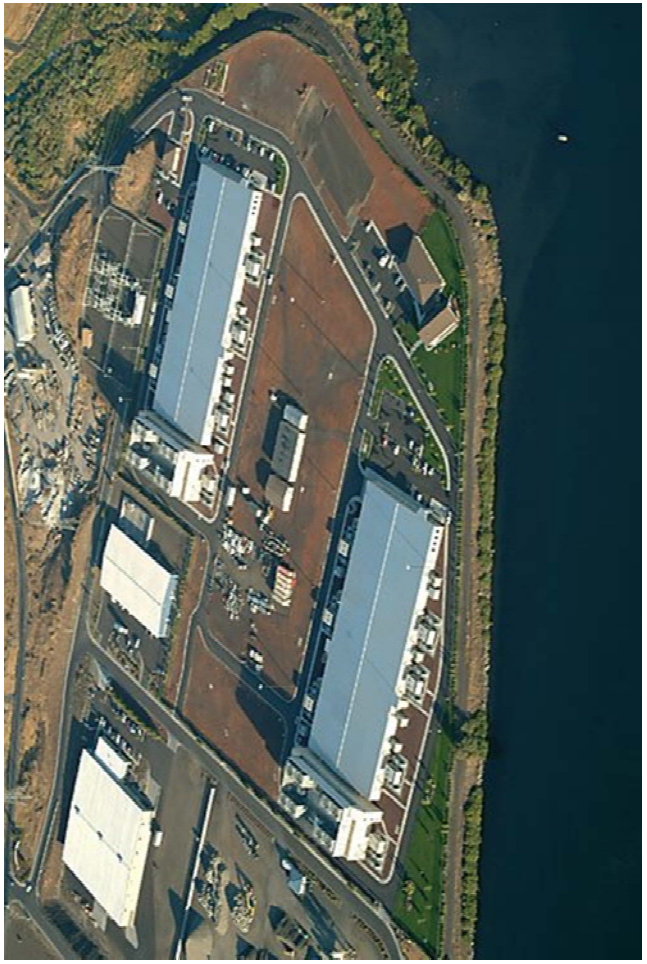
STRUCTURE: A 24 000-square-meter facility houses 400 containers. Delivered by trucks, the containers attach to a spine infrastructure that feeds network connectivity, power, and water. The data center has no conventional raised floors.

POWER: Two power substations feed a total of 300 megawatts to the data center, with 200 MW used for computing equipment and 100 MW for cooling and electrical losses. Batteries and generators provide backup power.



CONTAINER: Each 67.5-cubic-meter container houses 2500 servers about 10 times as many as conventional data centers pack in the same space. Each container integrates computing, networking, power, and cooling systems.

Google Dallas Data Center



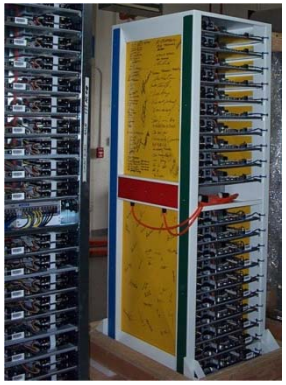
Google Data Centers

- 36 data centers / 500 Ips (2013-彰化; 2019-台南; 2021-雲林)
- Continuous evolution: 7 significant revisions in last 10 years
- An ordinary search query involves 700 to 1,000 servers

Inside a Google Data Center (2003)

- Google servers are custom made
- A rack (機櫃) consists of 40 to 80 x86-based servers
  - 1 rack = 20 2u or = 40 1u
- CPU
  - Intel-Celeron 533 MHz
  - Intel PIII Dual 1.4G
- HDD: IDE 80G
  - Index server has less disk space (cpu-intensive)
  - Doc server has more disk space
- Network
  - Core: Gigabyte Ethernet
  - Rack: 100 Mbps Ethernet

Inside a Google Data Center (2009)



- A small data center consists of a minimum of 2,400 servers
  - racks of 80 servers tied together with 10Gb Ethernet or other high-speed network fabrics
  - 30 or more of these racks are deployed into a single cluster
- Each of these servers has 16GBs of RAM with fast 2TB (Terabyte) hard drives
  - A patent on a power supply that integrates a battery, allowing it to function as an uninterruptible power supply (UPS)
  - Google-optimized Ubuntu Linux

YouTube - Google container data center tour (2009.4.7)

Google Search Engine

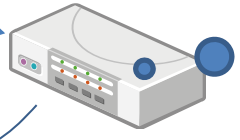
- Google查詢需耗費大量計算與I/O資源
  - 讀取 > 數百MB資料
  - 計算需耗 > 數十億CPU cycles
- Google 搜尋引擎如何建置
  - 硬體 > 15,000 一般PC
  - 軟體：需要特殊設計
    - Energy efficiency
    - Price-performance ratio

Google Search Engine

- 針對巨量資料特別設計的軟體架構
  - 在不穩定的硬體上，透過軟體架構設計建立穩定的系統
  - 二個關鍵理念
    - Provide reliability in software rather than server-class hardware
      - 以軟體取代硬體的部份
        - » Replication
        - » Failure detection
    - Tailor design for best throughput

http://www.google.com/search?q=nccu

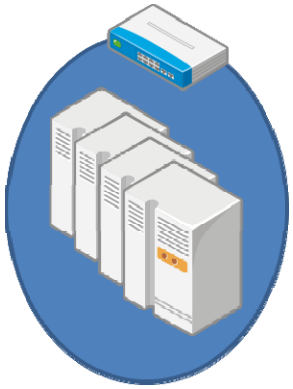
1) nslookup www.google.com



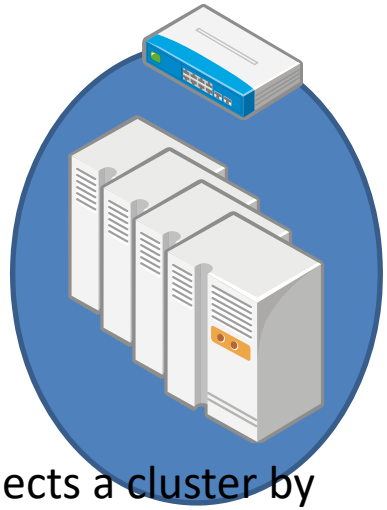
2) 173.194.127.50

3) http://173.194.127.50/search?q=nccu

173.194.127.50



173.194.127.51

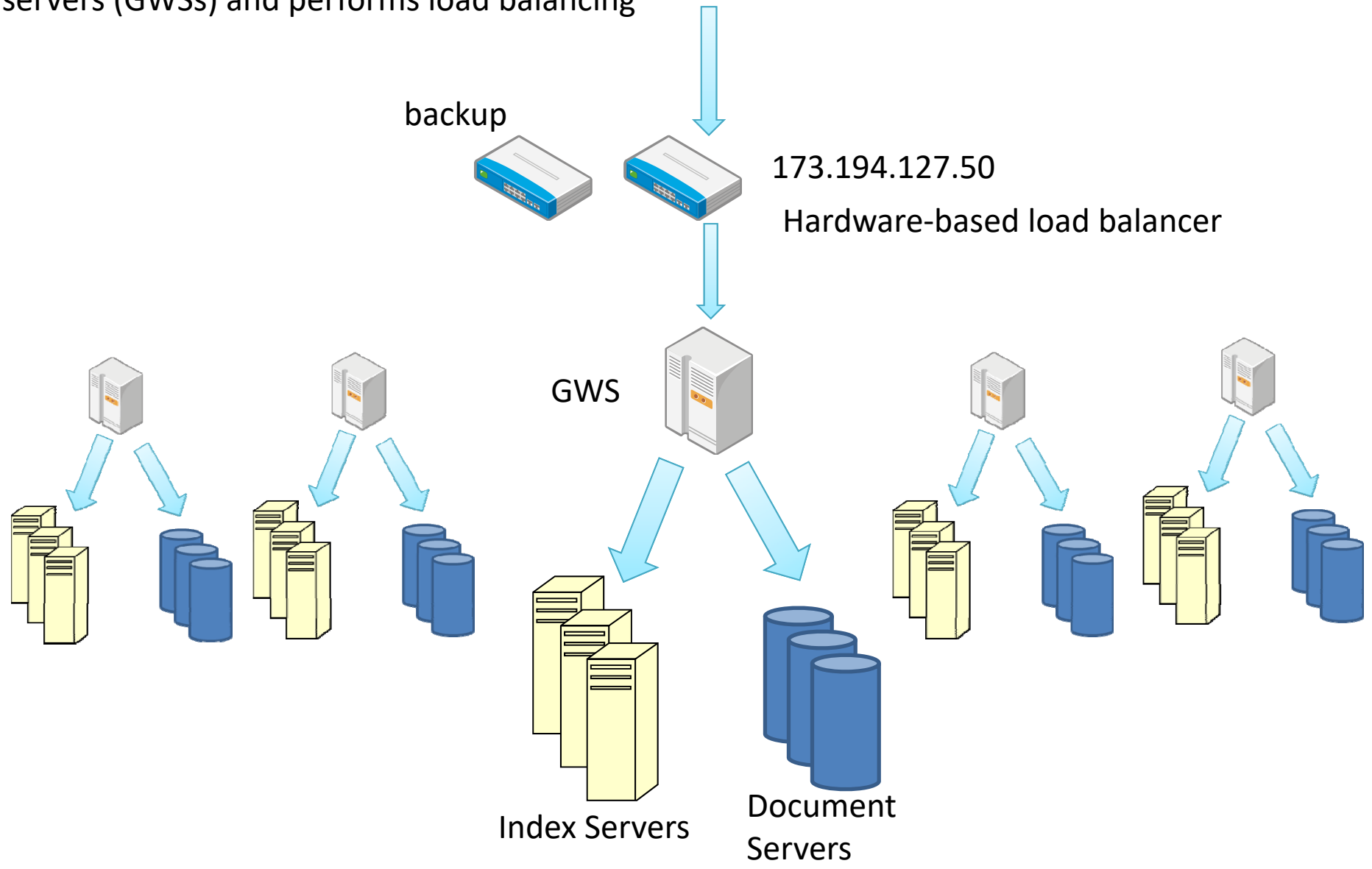


173.194.127.52

173.194.127.48

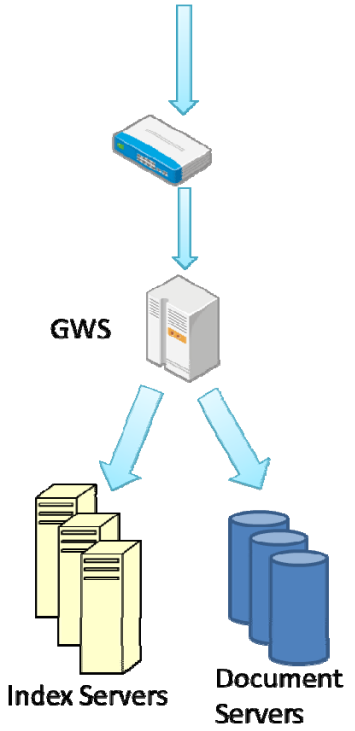
A DNS-based load-balancing system selects a cluster by accounting for the user's geographic proximity to each physical cluster

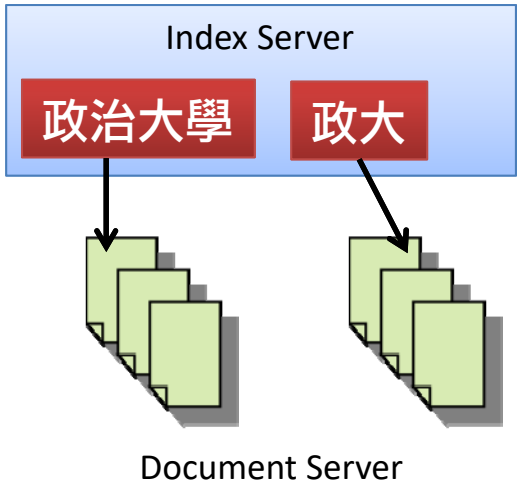
A hardware-based load balancer in each cluster monitors the available set of Google Web servers (GWSs) and performs load balancing



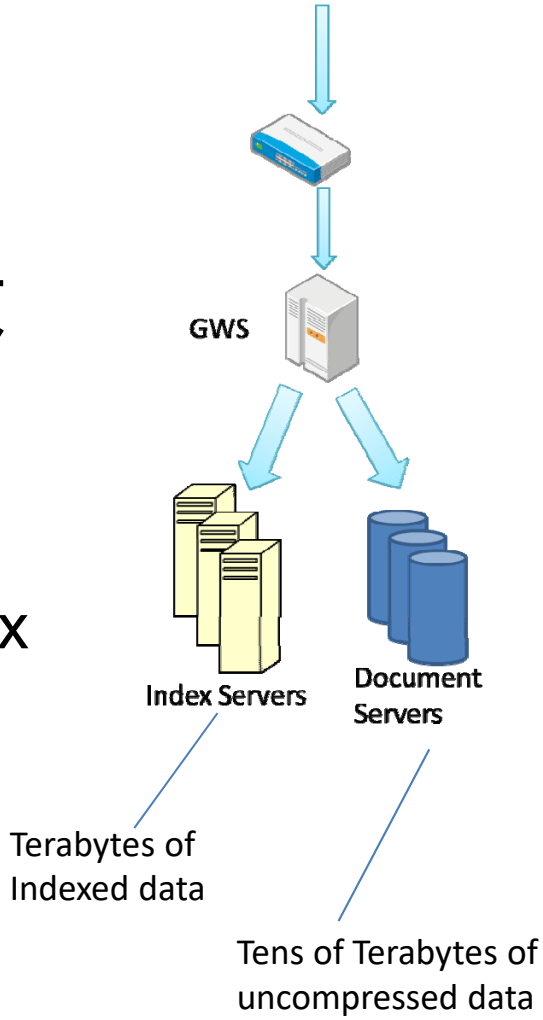
Searching and Ranking

- Index servers consult an inverted index that maps each query word to a matching list of documents (the hit list)
- The index servers then determine a set of relevant documents
  - Intersect the hit lists of the individual query words
  - Compute a relevance score for each document (PageRank)
    - The score determines the order of results on the output page

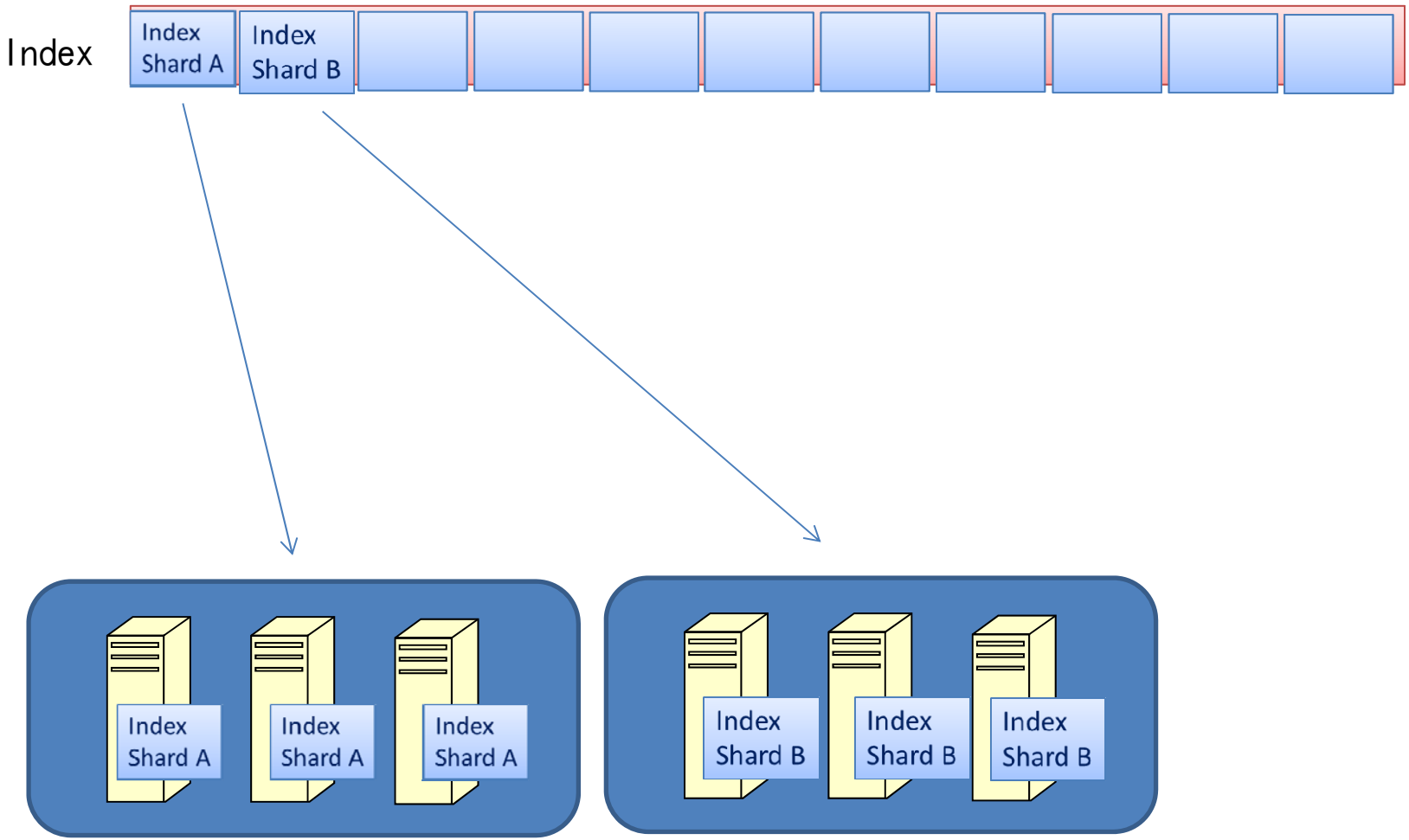


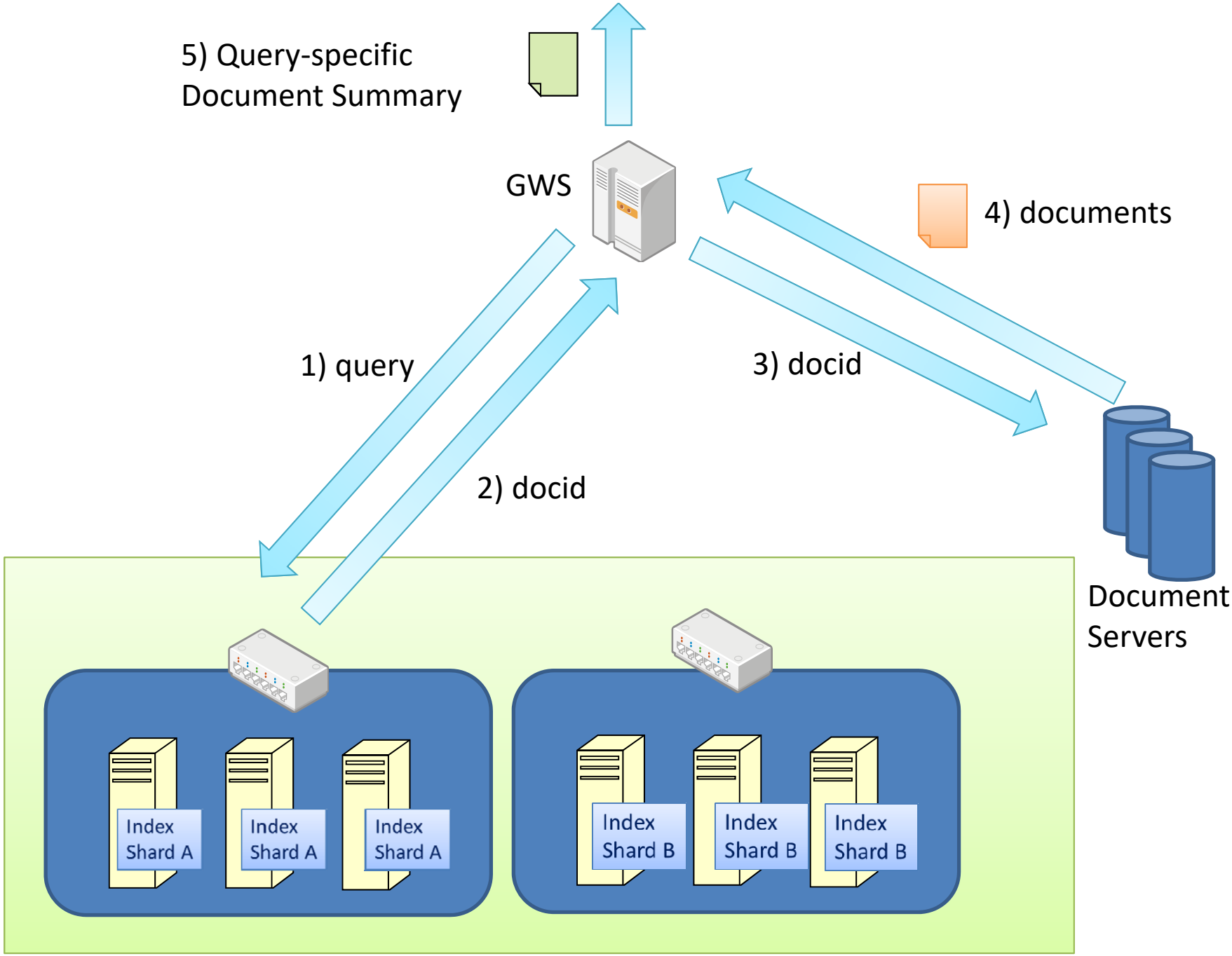


- Index本身也是Tera等級，如何有效率搜尋?
- Index shards
  - a randomly chosen subset of full index
  - A pool of machines serves requests for each shard

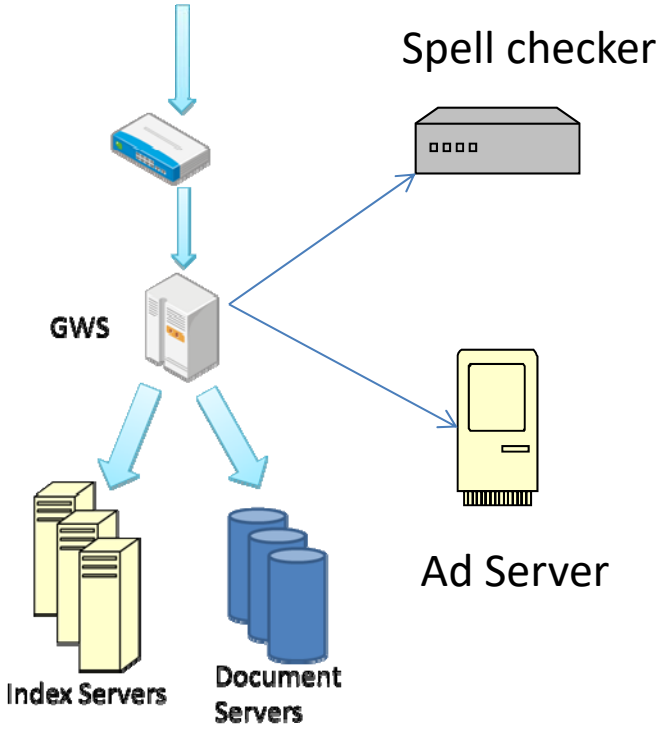


Index Shards




5) Query-specific
Document Summary
GWS
4) documents
1) query
3) docid
2) docid
Document
Servers
Index
Shard A
Index
Shard A
Index
Shard A
Index
Shard B
Index
Shard B
Index
Shard B

- Doc server
  - Online, low-latency copy of the entire Web
  - Google stores dozens of copies of the Web across its clusters
- Ancillary service
  - GWS also initiates several other ancillary tasks
    - Spell-checking
    - Ad-serving system
- Result generation
  - GWS generates the output page and returns it to the user's browser



Google's Key Patent on Cloud Computing

US2008/0262828 "Encoding and Adaptive Scalable Accessing of Distributed Models"

> "Systems, methods, and apparatus for accessing distributed models in automated machine processing, including using large language models in machine translation, speech recognition and other applications."

- Filed in February 2006
- 91 claims

System Example

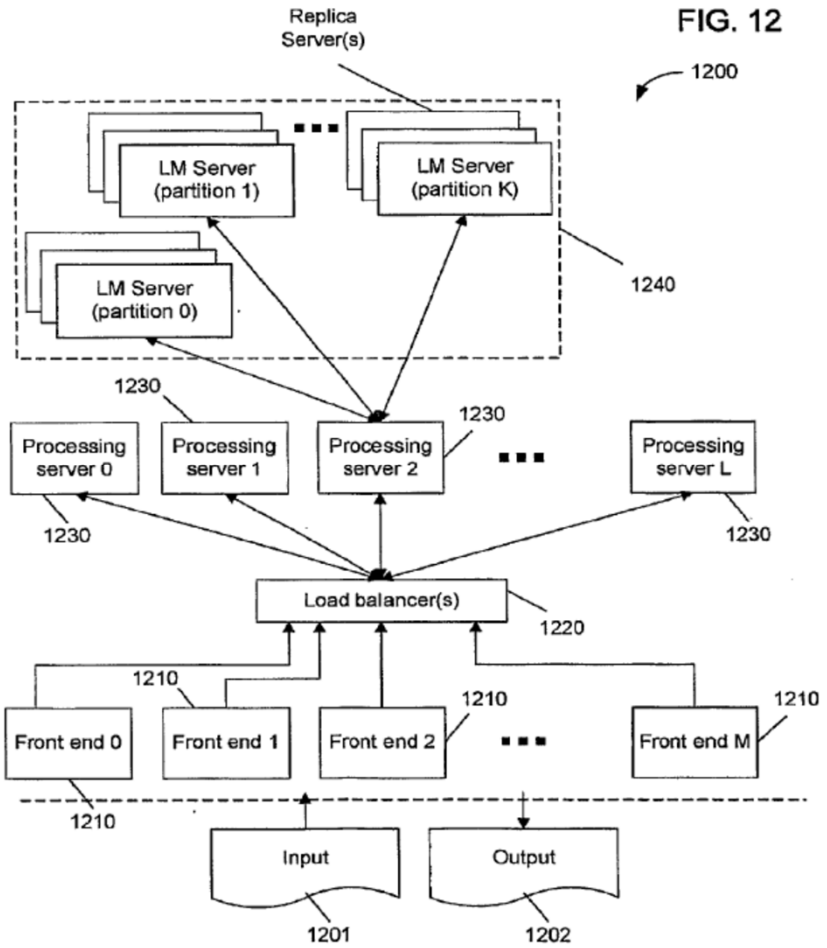


FIG. 12 An example of a distributed processing system that can be configured to provide a language processing function based on a large language model

Machine processing using machines such as computers to perform processing tasks such as machine translation

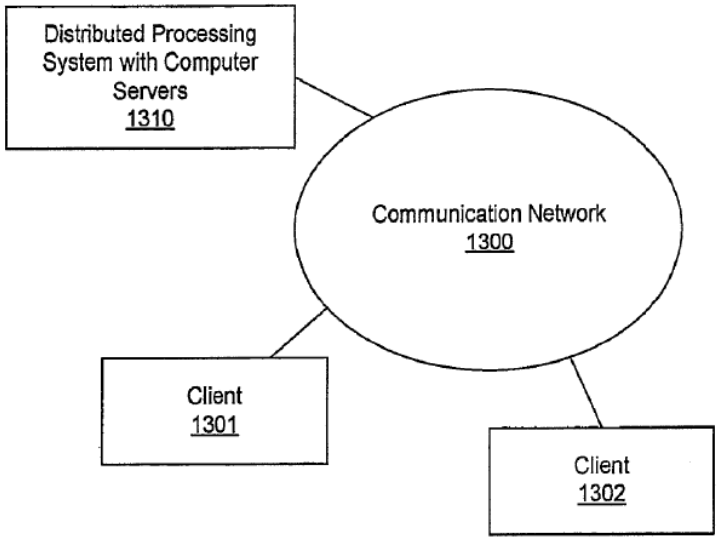
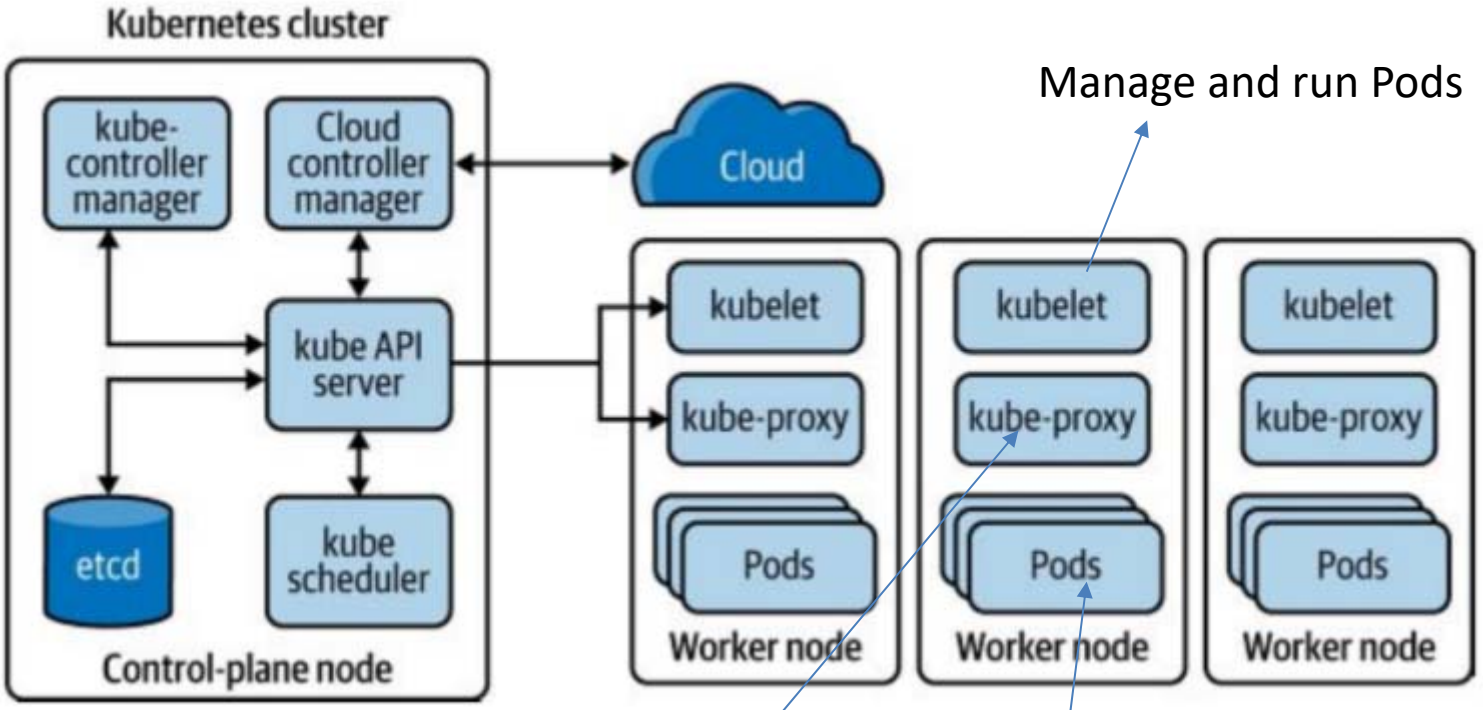


FIG. 13 An example computer system in a communication network that provides distributed processing

How to Manage Services in Data Center?

- Virtualization
  - “Hardware as Software”
- Containerization (ex: Docker)
  - Packaging system as OCI-compatible images
- Container Orchestration (ex: K8S)
  - Platform that runs OCI-compatible images
- Massive configuration management
  - Infrastructure as Code (IaC)

Kubernetes: a high-level view



Q & A