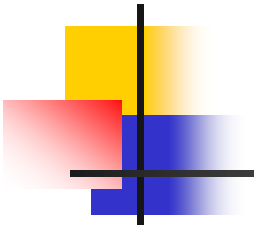
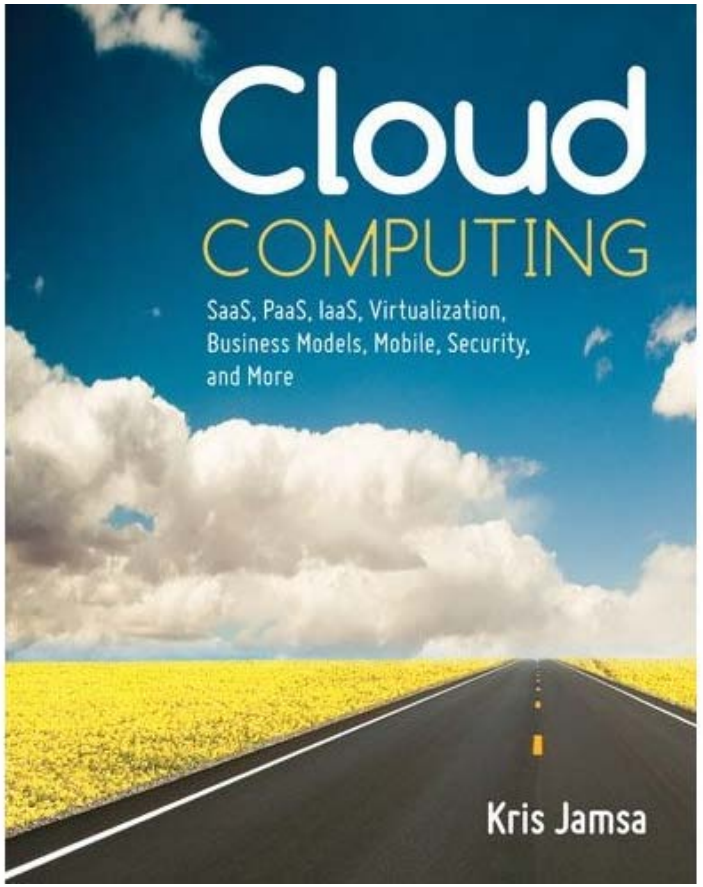


Chapter 0

Cloud Computing Overview

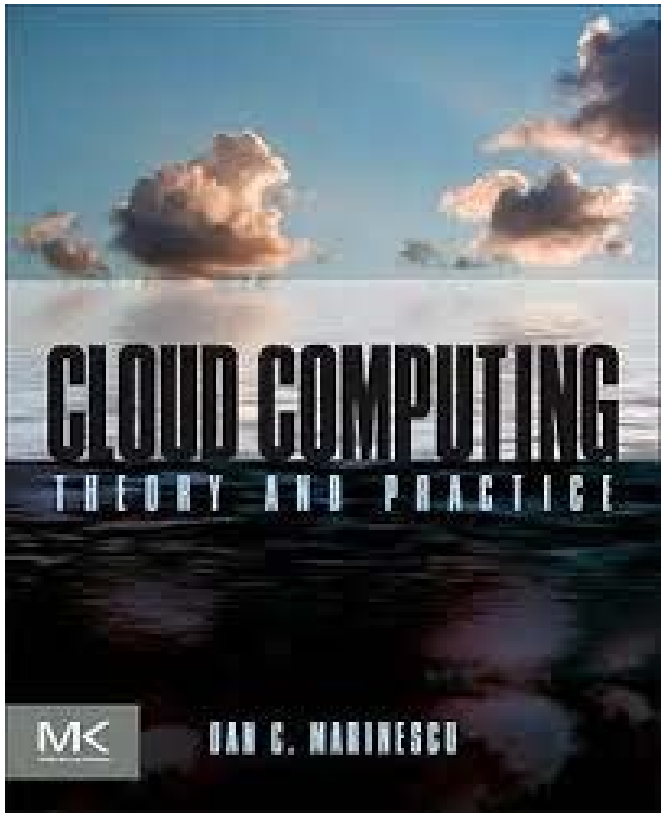
Books

- Cloud Computing, “Dr. Kris Jamsa”, 1 Edition, ISBN-10: 1449647391



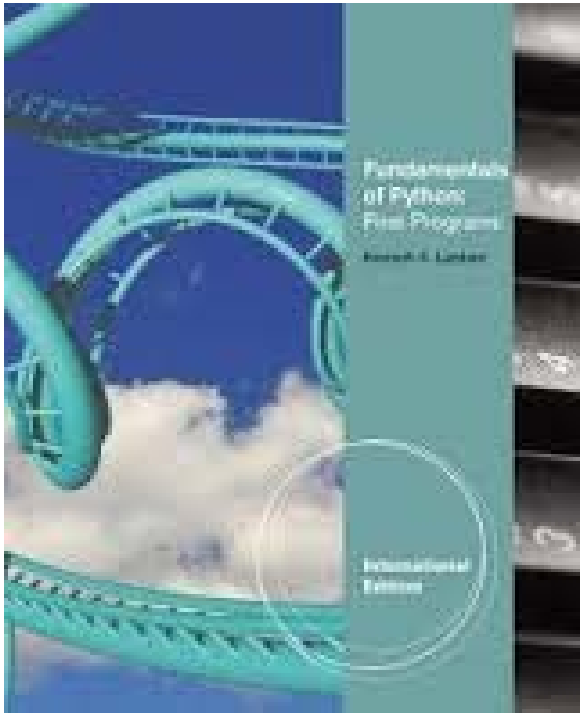
Reference Book

- Cloud Computing, “D Marinescu”, 1 Edition, ISBN-10: 0124046274



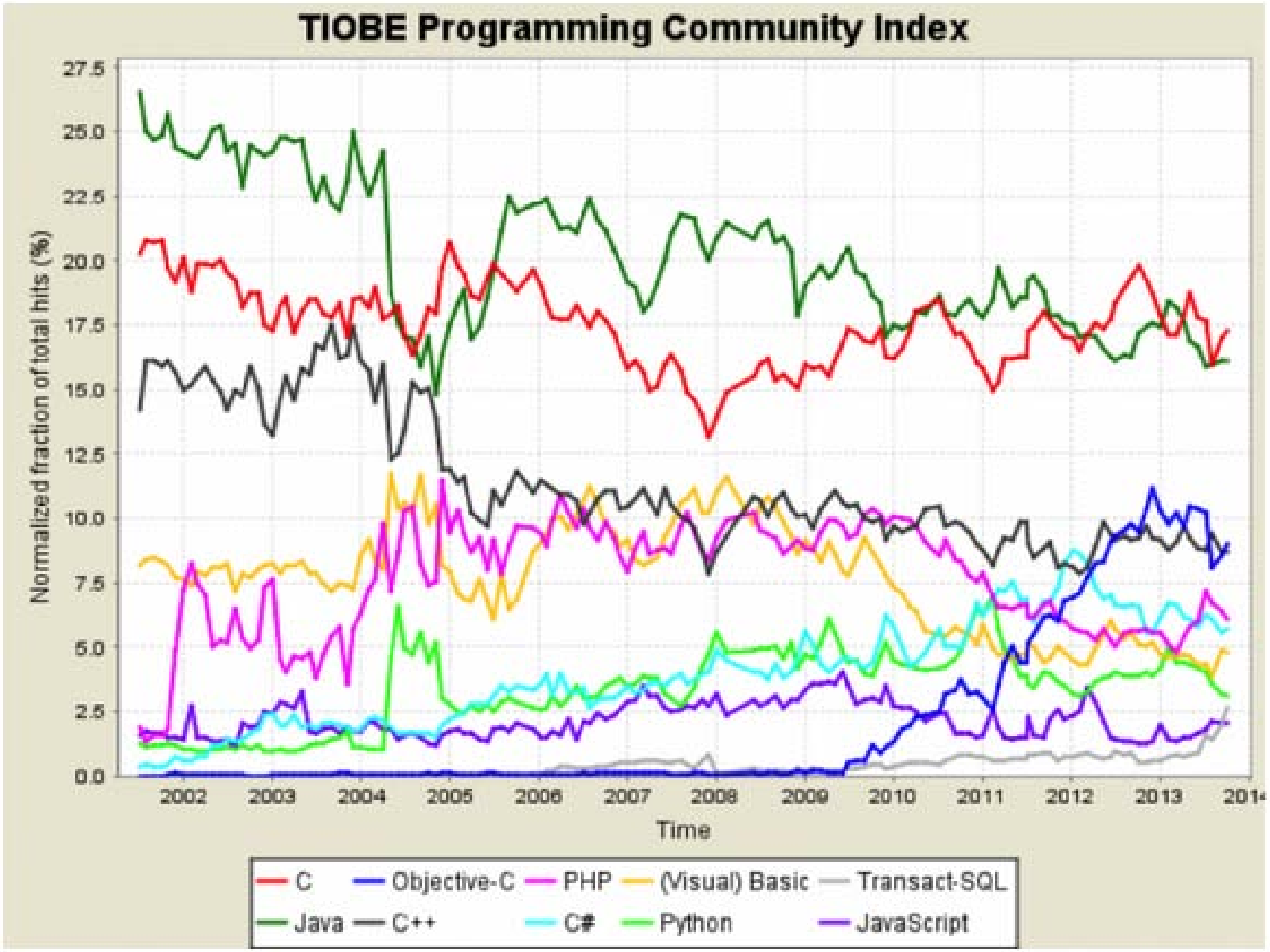
Reference Book

- Fundamentals of Python: First Programs, “Kenneth A. Lambert”, International Edition, 1111822700



Administration

- Instructor:
  - 曾學文 資工系助理教授
  - Office: Room 908
  - Email: hwtseng@nchu.edu.tw
  - Tel: 04-22840497 ext. 908
  - http://wccclab.cs.nchu.edu.tw/www/index.php/course
- Office Hours:
  - (Wednesday) 14:00~17:00; (Thursday) 14:00~17:00.
- Grade:
  - Homework (Project) 20%
  - Quiz 20%
  - Midterm Exam 30%
  - Final Exam 30%


TIOBE Programming Community Index
Normalized fraction of total hits (%)
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Time
C
Objective-C
PHP
(Visual) Basic
Transact-SQL
Java
C++
C#
Python
JavaScript

Position Nov 2013	Position Nov 2012	Delta in Position	Programming Language	Ratings Nov 2013	Delta Nov 2012	Status
1	1	=	C	18.155%	-1.07%	A
2	2	=	Java	16.521%	-0.93%	A
3	3	=	Objective-C	9.406%	-0.98%	A
4	4	=	C++	8.369%	-1.33%	A
5	6	↑	C#	6.024%	+0.43%	A
6	5	↓	PHP	5.379%	-0.35%	A
7	7	=	(Visual) Basic	4.396%	-0.64%	A
8	8	=	Python	3.110%	-0.95%	A
9	23	↑↑↑↑↑↑↑↑↑↑↑↑	Transact-SQL	2.521%	+2.05%	A
10	11	↑	JavaScript	2.050%	+0.77%	A
11	15	↑↑↑↑	Visual Basic .NET	1.969%	+1.20%	A
12	9	↓↓↓	Perl	1.521%	-0.66%	A
13	10	↓↓↓	Ruby	1.303%	-0.44%	A
14	14	=	Pascal	0.715%	-0.17%	A
15	13	↓↓	Lisp	0.706%	-0.25%	A
16	19	↑↑↑	MATLAB	0.656%	+0.04%	B
17	12	↓↓↓↓↓	Delphi/Object Pascal	0.649%	-0.35%	A-
18	17	↓	PL/SQL	0.605%	-0.03%	A-
19	24	↑↑↑↑↑	COBOL	0.585%	+0.11%	B
20	20	=	Assembly	0.532%	-0.05%	B

Course Goal

- What is cloud?
- To know what is the cloud computing!!!
- To understand how to design the data center networks of cloud computing.
  - QoS
  - Throughput
  - Routing and Failover
  - Transmission Delay
  - Scalable
  - Power and Thermal
  - ...

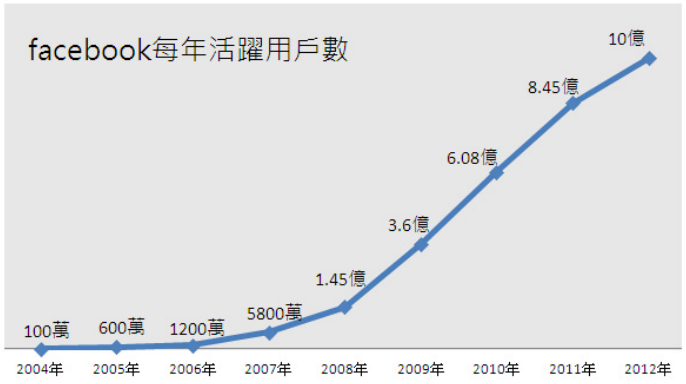
How much data?

- Wayback Machine (網站時光機) has 2 PB + 20 TB/month (2006)
- “all words ever spoken by human beings” ~ 5 EB
- NOAA (美國國家海洋暨大氣總署) has ~1 PB climate data (2007)
- CERN’s LHC (大型強子對撞機) will generate 15 PB a year (2008)
- Google processes 24 PB a day (2009)

- 1 Terabyte (TB) = 1024 GB
- 1 Petabyte (PB) = 1024 TB
- 1 Exabyte (EB) = 1024 PB
- 1 Zettabyte (ZB) = 1024 EB
- 1 Yottabyte (YB) = 1024 ZB

Hugh Data

- Huge multicast traffic in DCN
  - Google MapReduce over 400 PB in one month
  - Facebook registered users over one billion, the amount of data generated every day more than 300 TB
  - 2011 global digital data using about 1.8 ZB. According to IDC (International Data Corporation) made the prediction research report, the total to 2020 will be 44 times now, about 35.2 ZB



1 Byte = 8 Bits 1 Kilobyte (KB) = 1024 Bytes
1 Megabyte (MB) = 1024 KB
1 Gigabyte (GB) = 1024 MB
1 Terabyte (TB) = 1024 GB
1 Petabyte (PB) = 1024 TB
1 Exabyte (EB) = 1024 PB
1 Zettabyte (ZB) = 1024 EB
1 Yottabyte (YB) = 1024 ZB

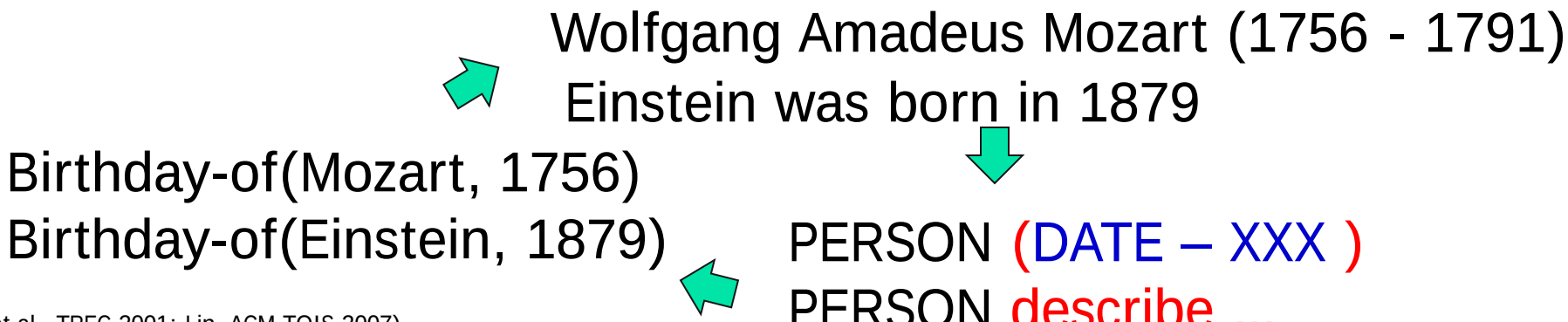
http://techorange.com/2013/04/16/why-virtual-machine-is-so-attractive/

How to create more data?

- Answering factoid questions
  - Pattern matching on the Web
  - Works amazingly well

Who shot Abraham Lincoln? → XXX shot Abraham Lincoln

- Learning relations
  - Start with seed instances
  - Search for patterns on the Web
  - Using patterns to find more instances



(Brill et al., TREC 2001; Lin, ACM TOIS 2007)
(Agichtein and Gravano, DL 2000; Ravichandran and Hovy, ACL 2002; ...)

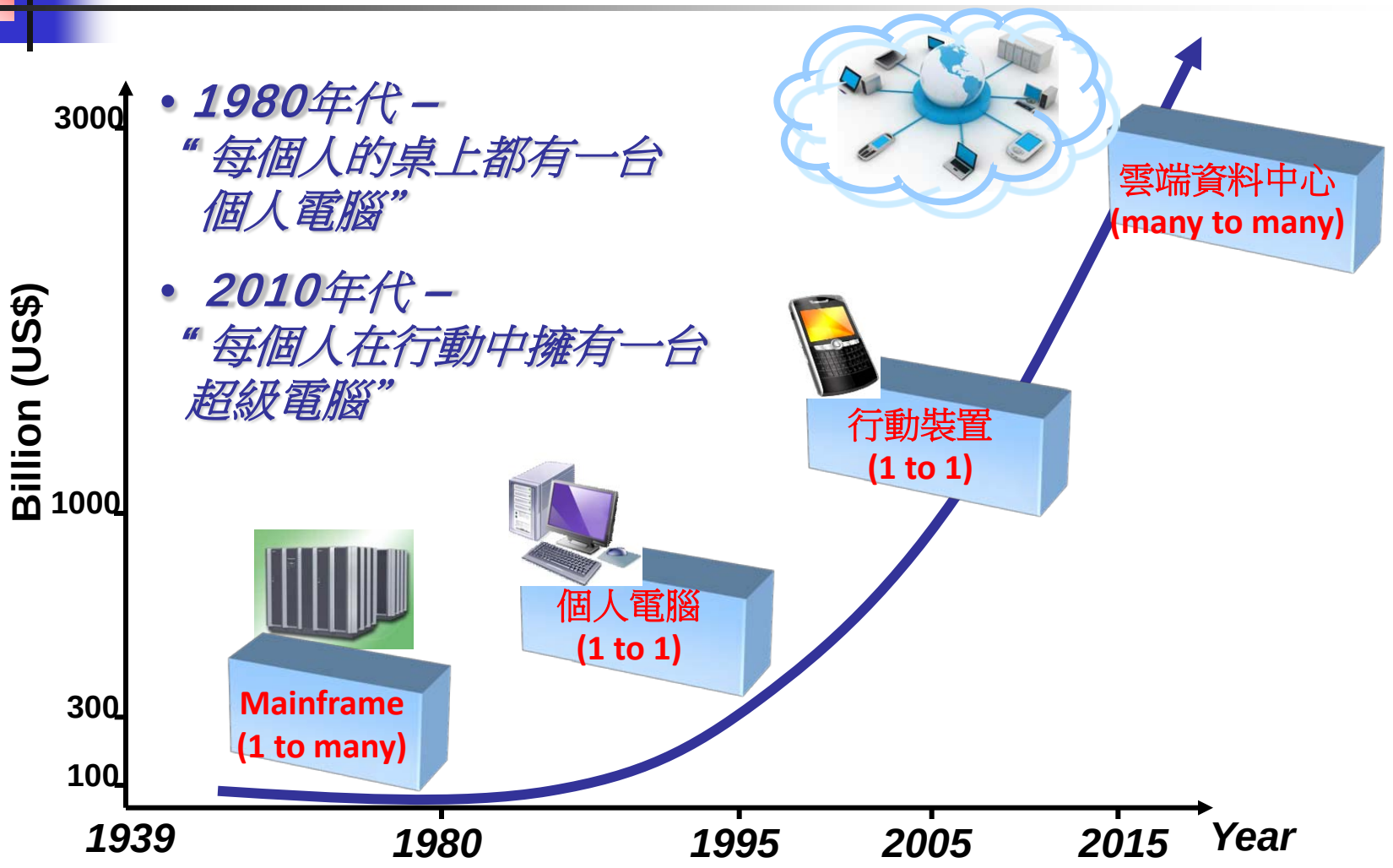
Large Data Centers

- Web-scale problems? Need more machines!!!
- Clear trend: centralization of computing resources in large data centers
  - Necessary ingredients: fiber, juice, and space
- Important Issues:
  - Redundancy --> fault tolerance, load balance.
  - Efficiency --> transmission latency
  - Utilization --> bandwidth utilization
  - Management --> virtualization, cooling system

五個主要趨勢正在進行，創造出無邊界的ICT架構

無邊界ICT 架構的關鍵技術

雲端運算新世代



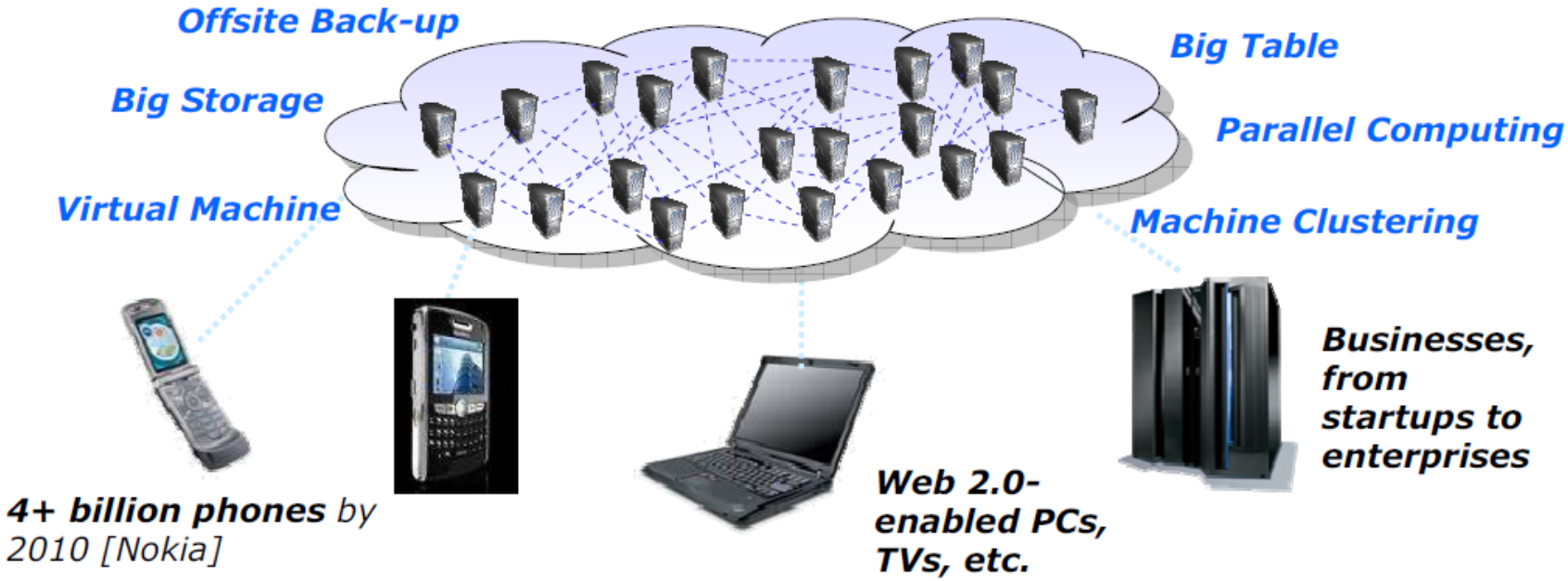
「雲端資料中心」讓電腦運算就像是水、電 一樣，只要連上網路就可以**pay-as-you-go**無限量提供服務。

雲端運算的定義

雲端運算是一種經由網際網路進行電腦運算的技術組成與使用模式-

(1)資料(data)與服務(service)放置在網際網路上之大型可延展(massively scalable)的資料中心

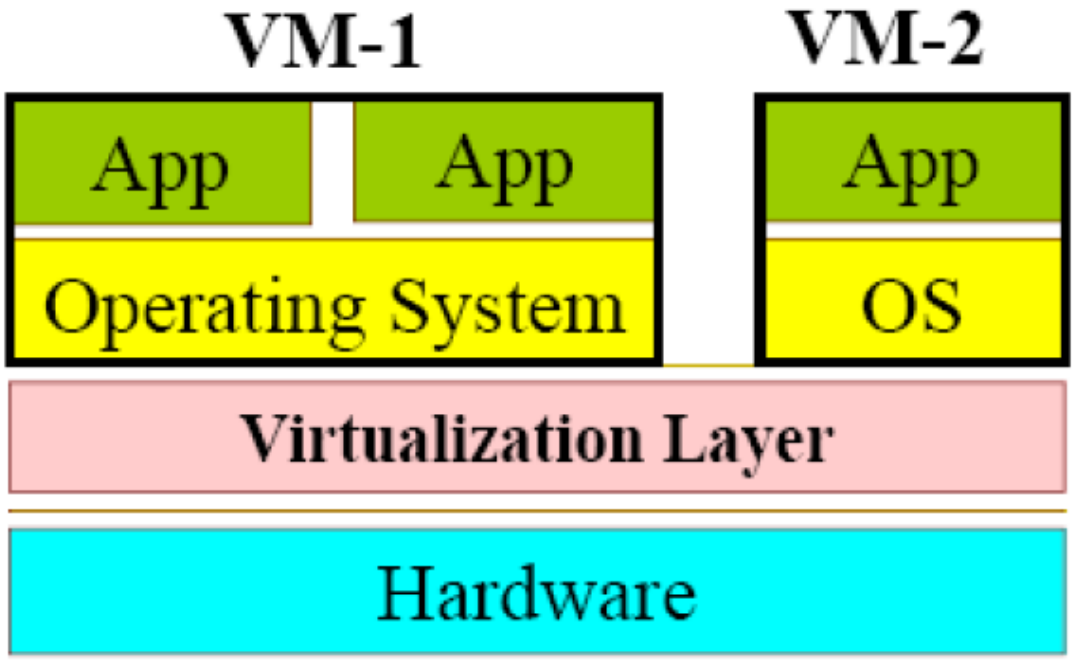
(2)使用者可以利用各種具備網際網路連線能力的電腦終端裝置(device)，無所不在(ubiquitous)的使用資料與服務



雲端運算的核心技術(虛擬運算技術)

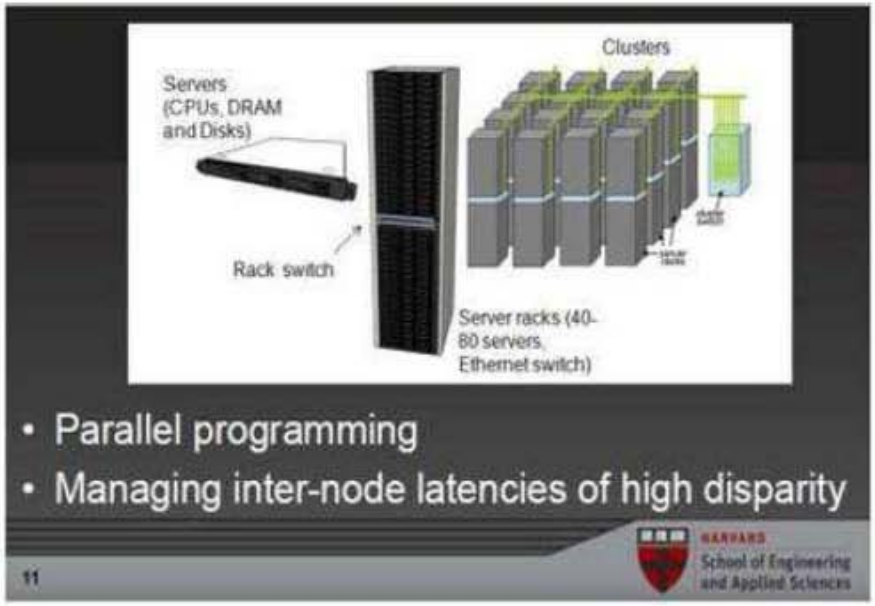
虛擬化/虛擬運算技術 (Virtualization) 是藉由一種對應方式 (virtual machine monitor, hypervisor, or virtualization layer)，將電腦硬體資源，如伺服器、儲存媒體，轉成一群可以被共用的裝置 (即虛擬裝置 virtual devices)，讓軟體與應用服務能共同使用這一群硬體

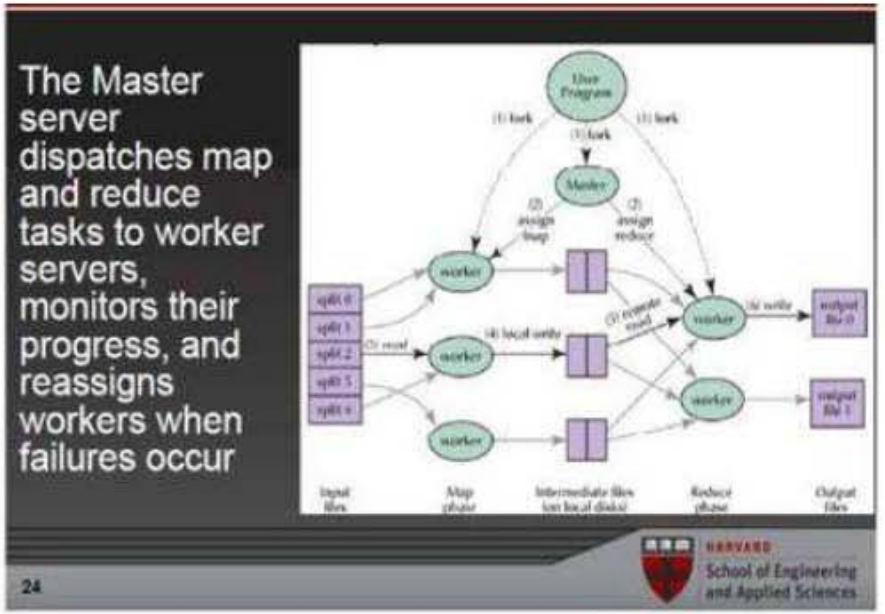
Source: Mendel Rosenblum
Stanford U., 1998



雲端運算的核心技術(叢集運算技術)

- 將許多實體電腦(通常是相同規格)，以網路連結，實現高延展與高效能的分散式運算(例如:Google Search)
- 利用 Hadoop 處理大量資料 (例如: 趨勢科技)



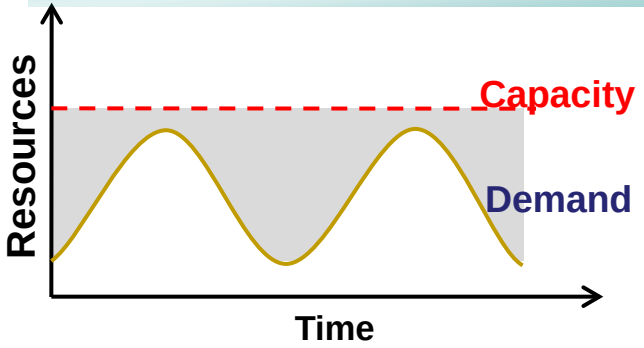


圖來源:孔祥重 院士

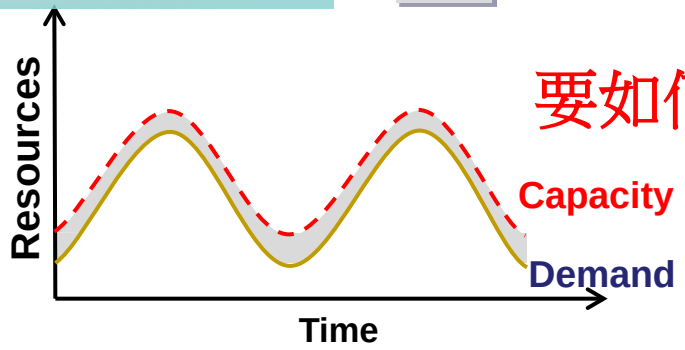
雲端運算經濟學

Pay by use instead of provisioning for peak

Unused resources



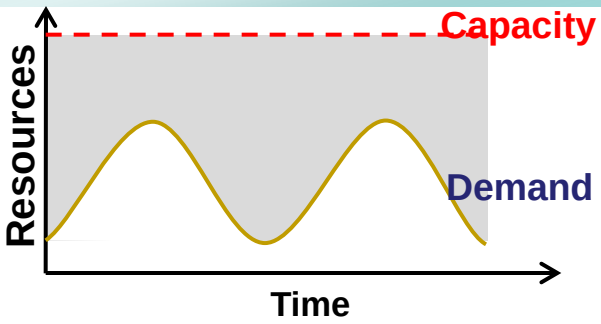
Static data center



Data center in the cloud

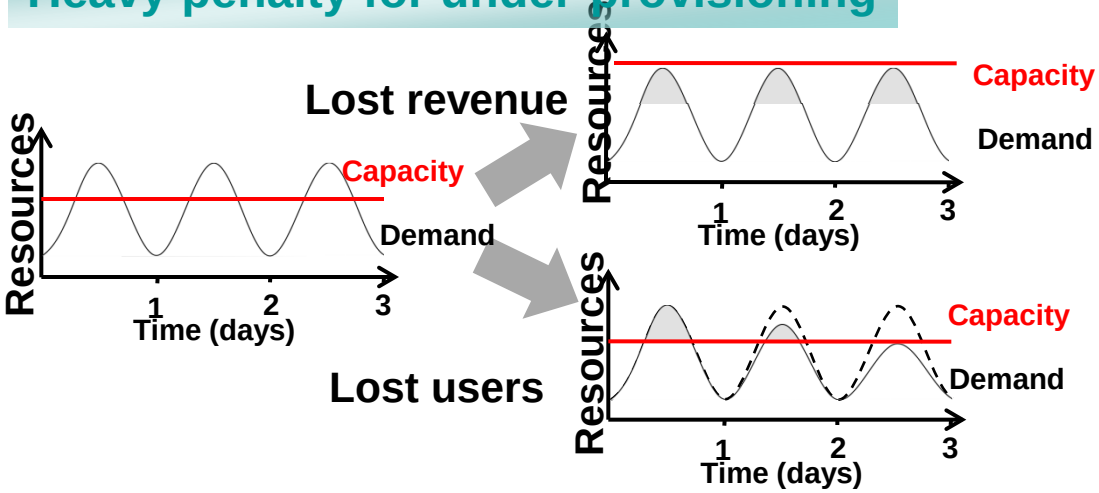
要如何做好資源管理

Risk of over-provisioning: underutilization



Static data center

Heavy penalty for under-provisioning

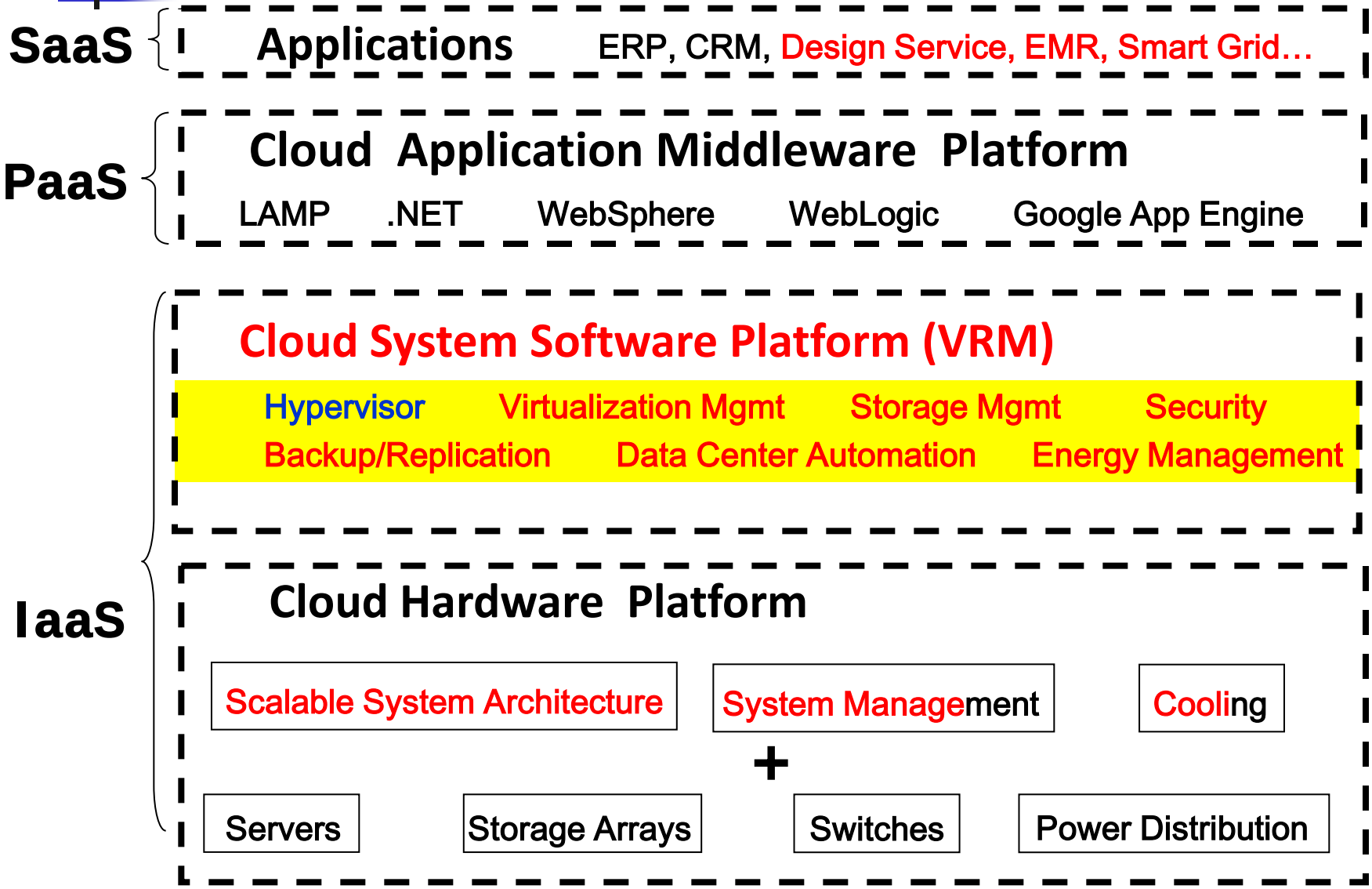


資料來源：UC Berkeley RAD Lab

雲端運算商業模式

Cloud Ecosystem

雲端運算技術藍圖


SaaS
Applications
ERP, CRM, Design Service, EMR, Smart Grid…
PaaS
Cloud Application Middleware Platform
LAMP
.NET
WebSphere
WebLogic
Google App Engine
IaaS
Cloud System Software Platform (VRM)
Hypervisor
Virtualization Mgmt
Storage Mgmt
Security
Backup/Replication
Data Center Automation
Energy Management
Cloud Hardware Platform
Scalable System Architecture
System Management
Cooling
+
Servers
Storage Arrays
Switches
Power Distribution

全球雲端運算產值現況與預測

•預估全球雲端運算服務市場規模將從2009年的102.6億美元成長至2013年的255.2億美元，CAGR為27.4%，未來雲端運算服務市場商機成長可期。

•全球2009-2013年IaaS(CAGR 42.6%)、PaaS (CAGR 54.5%)呈現高度成長，至2013年，IaaS、PaaS市場規模達107.8億美元，IaaS、PaaS市場占有率分別提升至40.6%及1.6%。

•2009年全球SaaS市占率最大(占74.9%)，至2013年，SaaS市場規模達147.4億美元，排行第一，2009-2013 CAGR達17.7%。

XaaS對台灣資訊服務業的新挑戰與機會?

全球雲端運算服務市場規模暨市占率(單位：十億美元、%)

年份 產品區隔	2009年	2010年	2011年	2012年	2013年	2009-2013年 CAGR(%)
IaaS市場規模	2.51	3.64	5.38	7.56	10.36	42.6%
PaaS市場規模	0.07	0.10	0.14	0.21	0.42	54.5%
SaaS市場規模	7.68	8.97	10.53	12.39	14.74	17.7%
全部市場規模	10.26	12.71	16.05	20.16	25.52	27.4%
IaaS市占率	24.5%	28.6%	33.5%	37.5%	40.6%	-
PaaS市占率	0.7%	0.8%	0.9%	1.0%	1.6%	-
SaaS市占率	74.9%	70.6%	65.6%	61.5%	57.8%	-

資料來源： MIC，2010年6月

CAGR:年複合平均成長率

雲端運算產業轉型新契機

半導體產業

IDM
整合元件製造廠

intel AMD

Foundry
晶圓廠

UMC tsmc

Design service
設計服務公司

FARADAY 創意電子 GLOBAL UNICHIP CORP.

Fabless
晶片設計公司

MEDIATEK REALTEK MORNINGSTAR NOVATEK

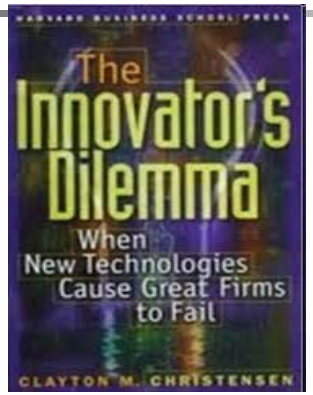
雲端運算產業

自建資料中心
自己營運並提供服務

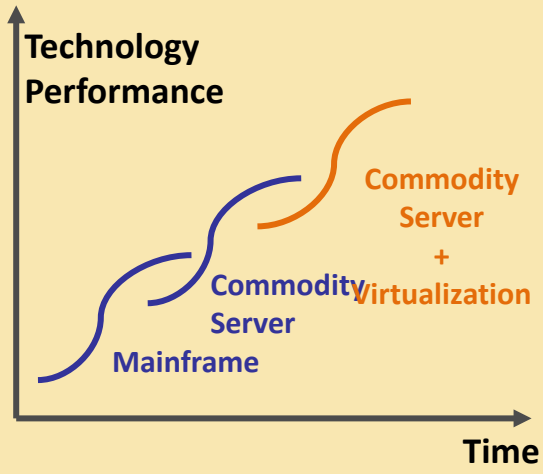
Google Microsoft

台灣的機會

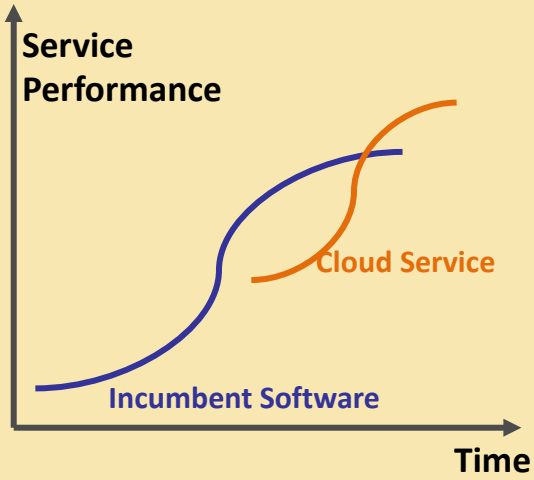
雲端運算帶來破壞式創新



Clayton M. Christensen



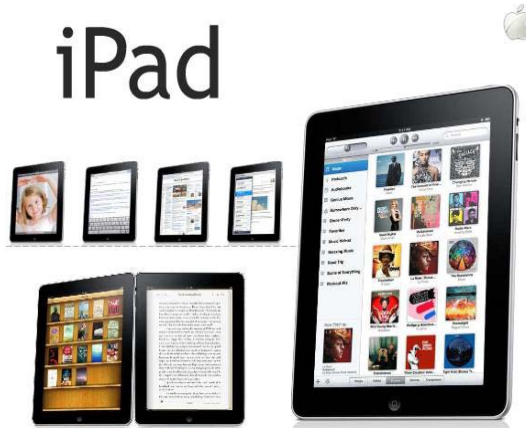
- 雲端技術是"Sustaining Innovation"
  - 雲端技術提供更好的功能與表現，且滿足相同企業用戶
  - 雲端技術無法產生破壞性創新，而是技術的延續



- 雲端服務是"Disruptive Innovation"
  - 雲端服務以使用量計費方式取代高額授權金或硬體
  - 鎖定中小客群而非主流大型企業客戶
  - 不需具備IT專業知識即可快速使用雲端服務

資料來源：工研院 IEK

Apple 雲端資料中心



1. iPad系列的優點包括運行速度相當快速，具多點觸控功能、直覺的操作設計， 能持續使用12小時等特色，大幅增加iPad的吸引力。
2. 蘋果公司有一項重大東海岸資料中心建設，以提高在線服務的能力。此次投資金額高達10億美元，目標建設和運營大型server farm

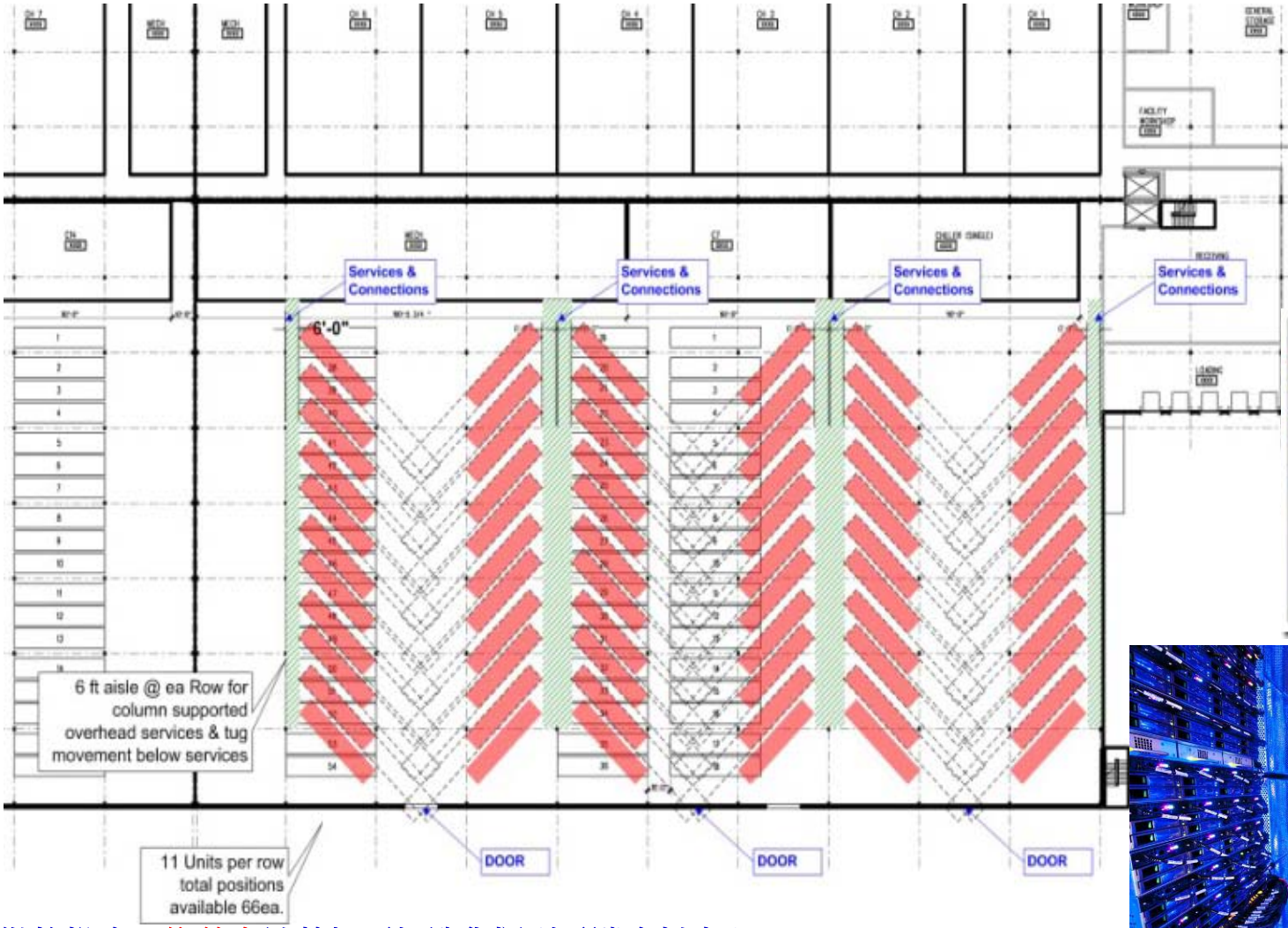
1.Apple's existing Newark, CA., Data Center is around 109,000 square feet--the new one is over 500,000. That represents either a ridiculously big scaling-up of business or a whole new thing 2.500,000 square feet is among the largest centers being built in the World on a single site. Microsoft's new one in Chicago is around 400,000, in comparison

www.datacenterknowledge.com

資料來源:經濟部科專辦公室

Microsoft 雲端資料中心

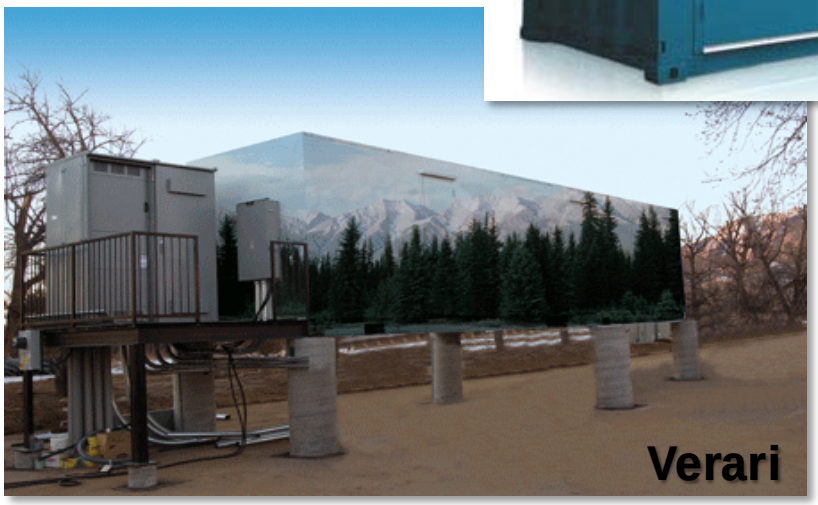
Microsoft…




Phase 1A Cblox area – load banks connected & operating

微軟投資五億美金於芝加哥打造貨櫃型雲端資料中心

雲端貨櫃型電腦



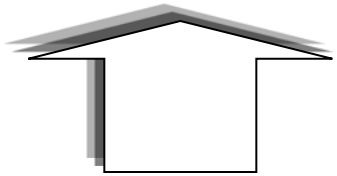
Source: Daniel Costello, Microsoft Data Center Transformation, Microsoft WinHEC 2008

Why Container Computer?

- 能源使用效率(PUE)較佳
  - 貨櫃內密閉式空間，冷卻成本大幅降低
- 擴充彈性高
  - 方便運輸、安裝與卸除
  - 能以不同貨櫃尺寸為出貨單位，支援不同的運算需求

台灣雲端運算產業推動策略

任務B.從全新雲端運算平台，
掌握利基應用機會
-催生雲端服務業



任務A.從既有資訊產業基礎，
建構成本優勢
-催生雲端設施/平台服務業

策略二.鼓勵雲端應用服務創新:
(Type1)鼓勵4C軟硬整合公有/私有雲
End-to-End Solution
(Type2)鼓勵雲端資料中心發展類似Apple A小型ISV業者有出頭機會

個人、企業、政府

SaaS
(國產)雲端應用服務

策略一.鼓勵優質平價國產雲端硬軟整合系統研發:
(1)國產Cloud OS與Green IDC系統軟體
(2)基於國產元件，發展大型資料中心或企業機房之雲端機櫃

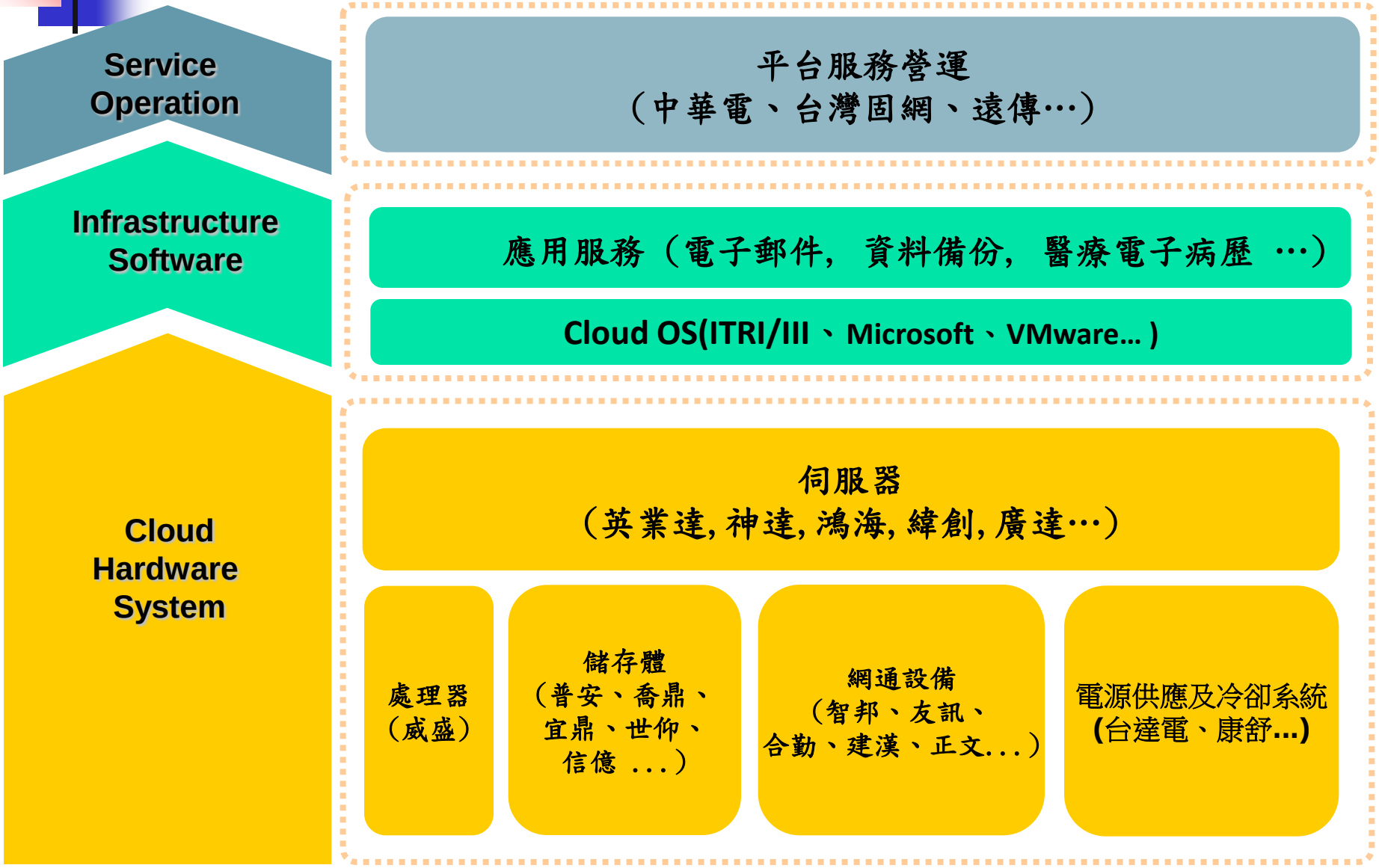
PaaS
(國產)雲端平台服務業

IaaS
(國產)雲端設施服務業

資料來源：經濟部科專辦公室

Independent software vendor (ISV)

台灣雲端上下游產業價值鏈



“Green” Cloud Computing

The Power of Evolution of VLSD

1990 — Generation 1: 10K Servers

1998 — Generation 2: 100K Servers

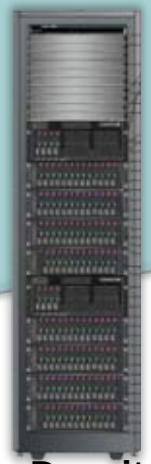
2008 — Generation 3: 300K Servers

~ 500 KWatts

~ 10 MegaWatts

~ 60 MegaWatts → ??

Server *Capacity*



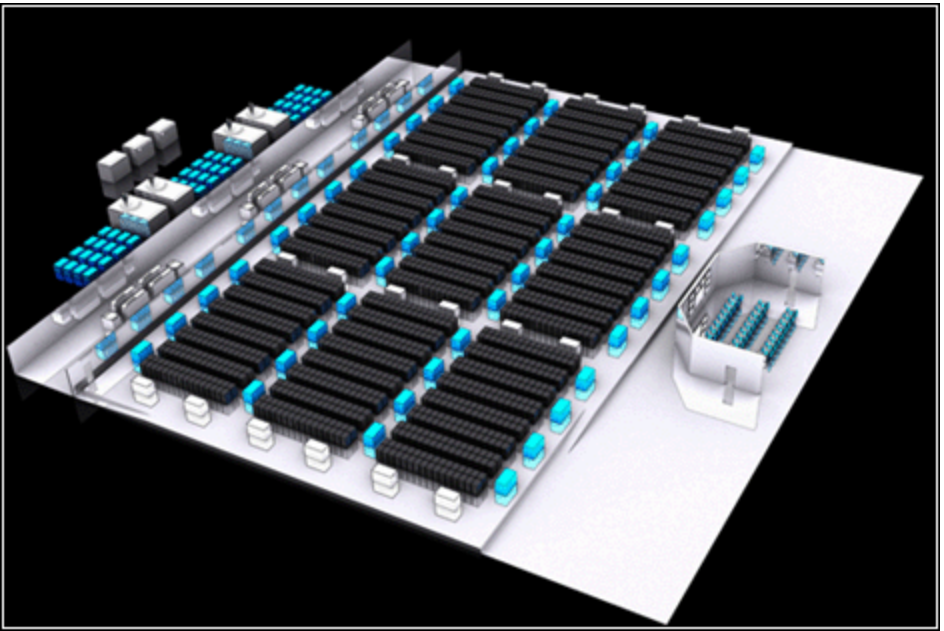
Server Density and Manageability

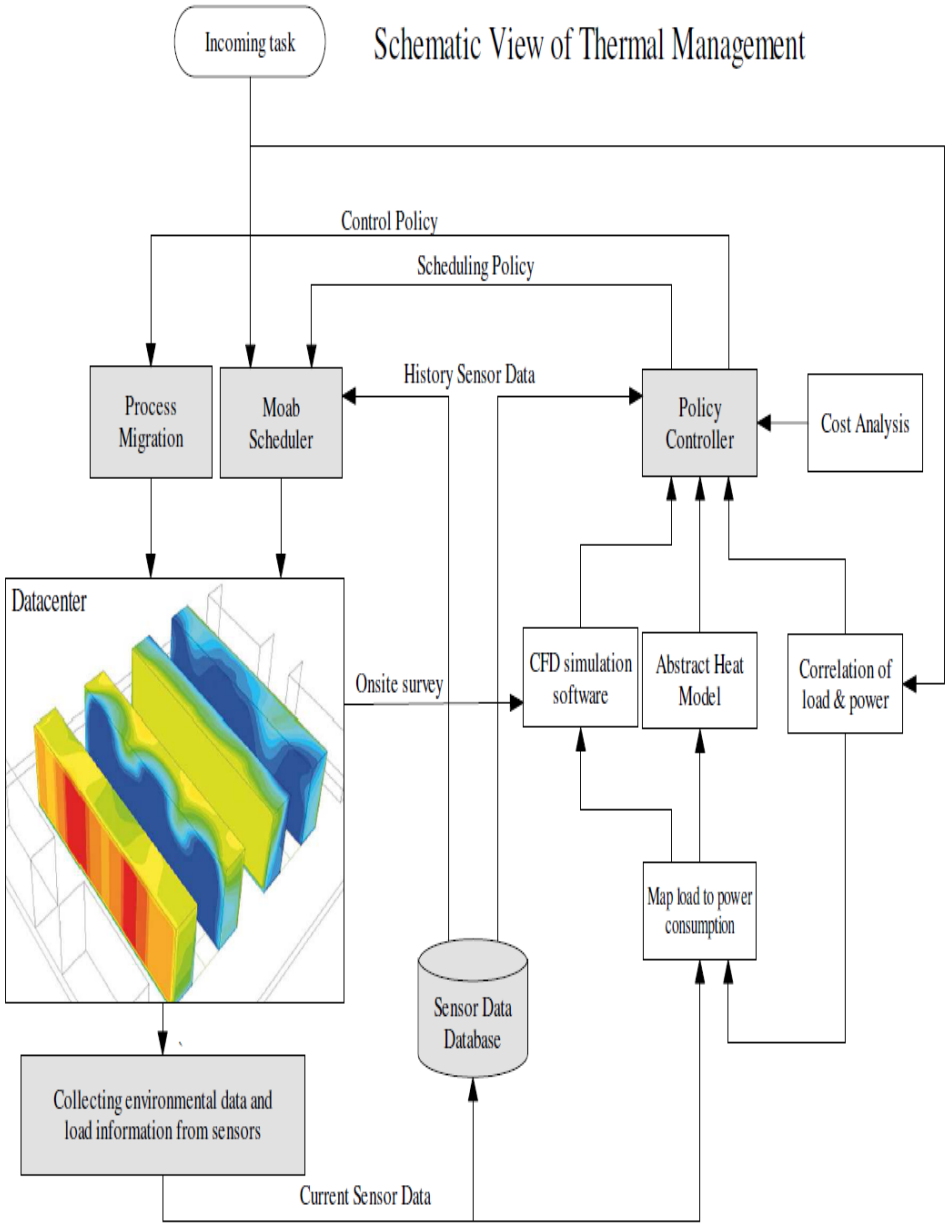
Container Scalability

Power Usage Efficiency (PUE)

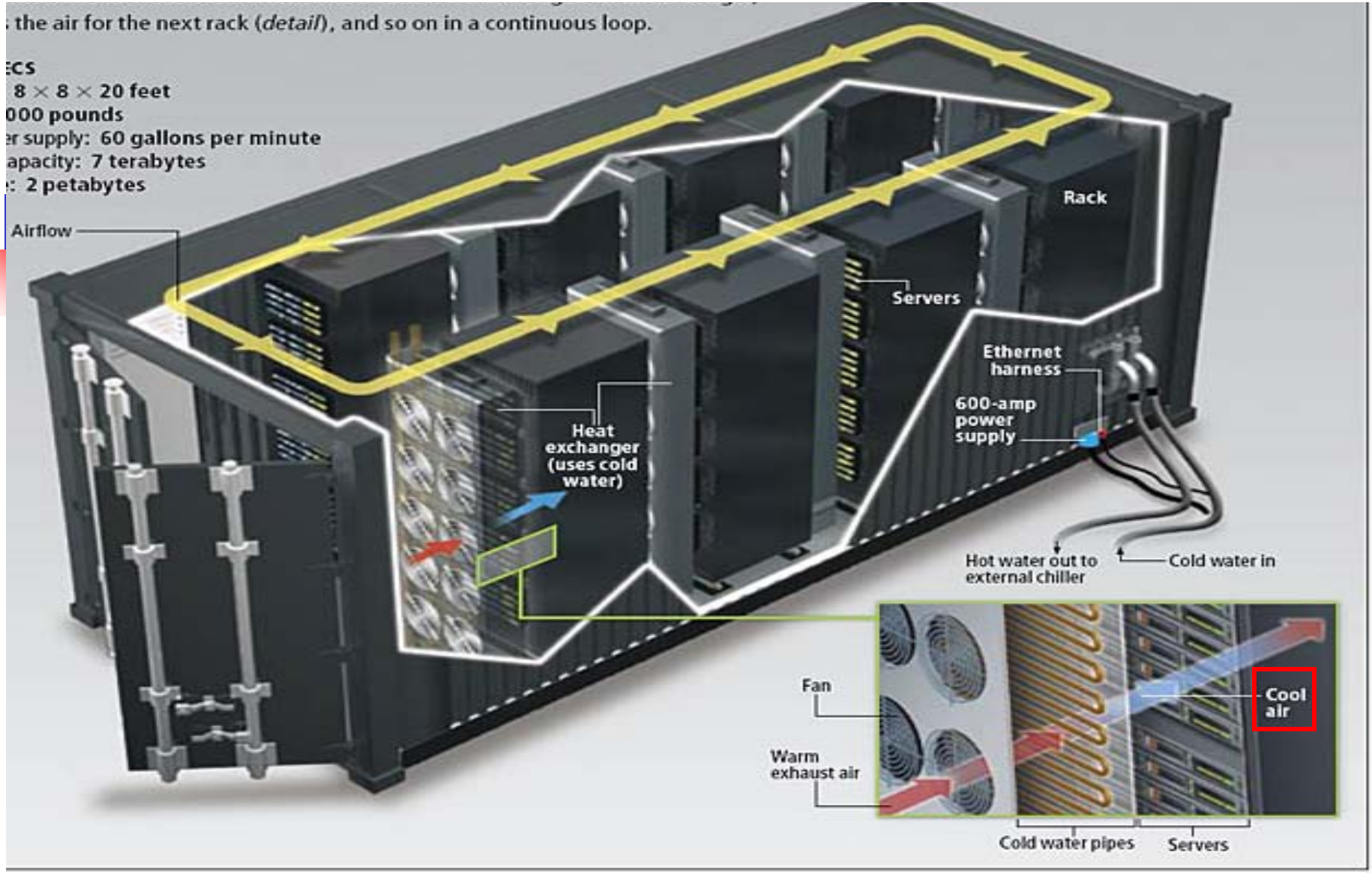
VLSD: Very Large Scale Datacenter

Thermal Management





Highly efficient container computer based VLSD

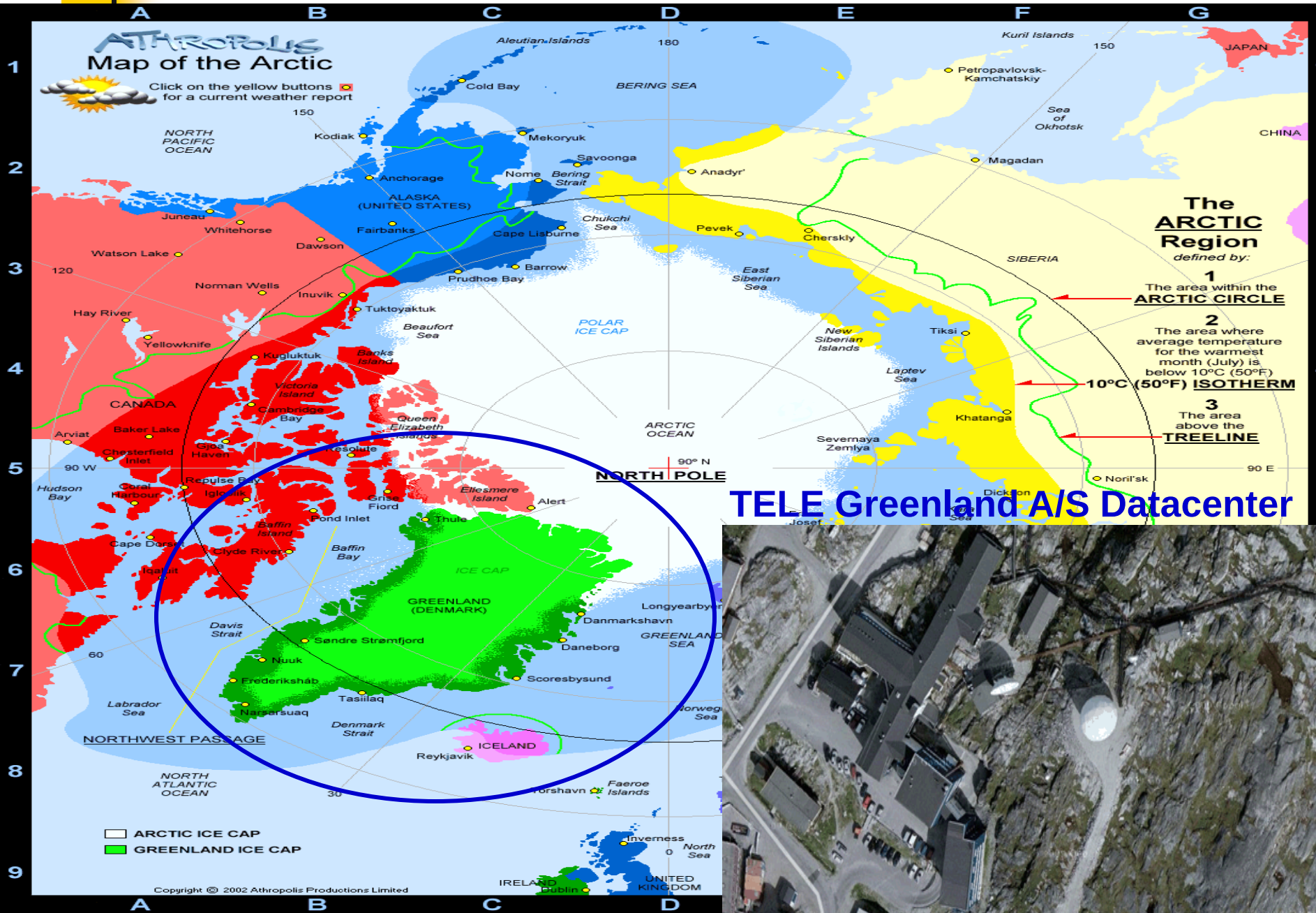


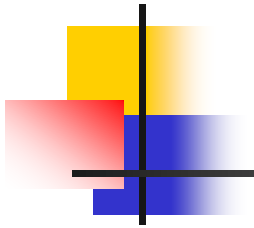
Cooling is a BIG problem in VLSD

Today's VLSD needs a lot of Power + Water

Google Datacenter at Columbia river, Oregon

Greenland (格陵蘭) Datacenter

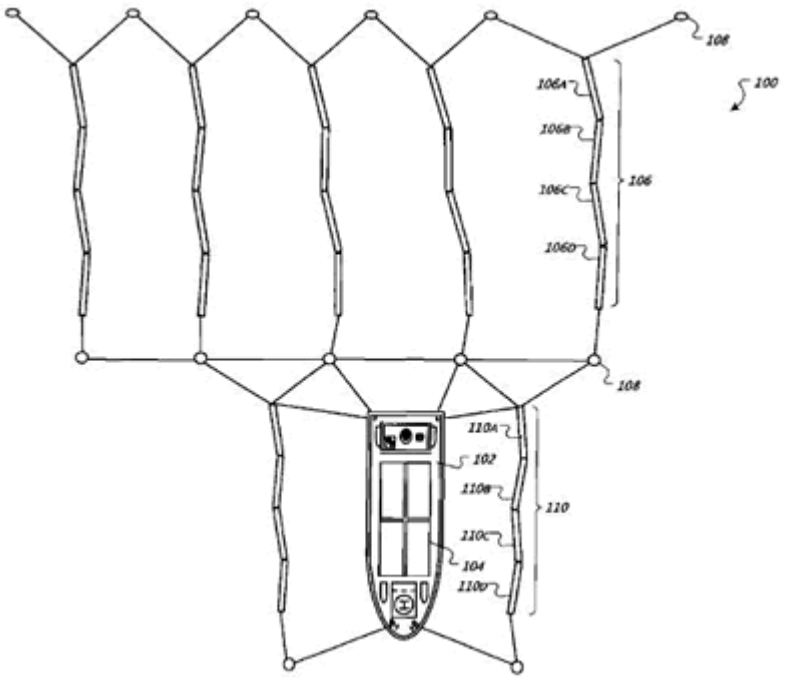




Interesting ideas for building “Green” VLSD

Interesting ideas for building VLSD

Interesting ideas for building “green” VLSD



(Google Navy floating data centers)

The sea-going computer platforms will be sustainably powered by wave energy converters.

Interesting ideas for building “green” VLSD

Google Navy floating data centers

The sea-going computer platforms will be sustainably powered by wave energy converters.

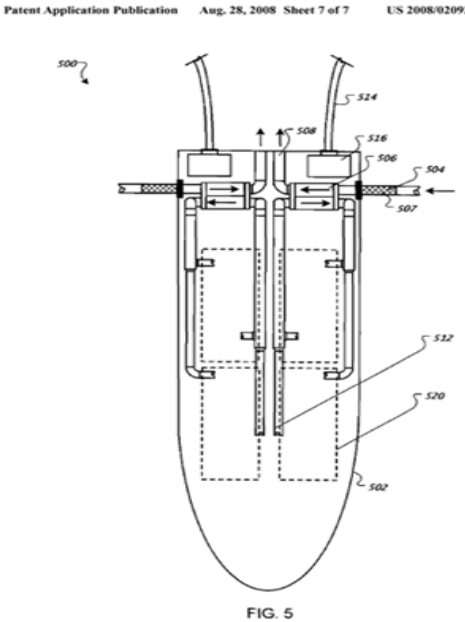


FIG. 5

USPTO PATENT FULL-TEXT AND IMAGE DATABASE

Home | Quick | Advanced | Pat Num | Help

Bottom

View Cart | Add to Cart

Images

United States Patent
Clidaras , et al.

Water-based data center

Abstract

A system includes a floating platform-mounted computer data center comprising a plurality of computing units, a sea-based electrical generator in electrical connection with the plurality of computing units, and one or more sea-water cooling units for providing cooling to the plurality of computing units.

Inventors: **Clidaras; Jimmy** (Los Altos, CA), **Stiver; David W.** (Santa Clara, CA), **Hamburgen; William** (Palo Alto, CA)
Assignee: **Google Inc.** (Mountain View, CA)
Appl. No.: **11/679,013**
Filed: **February 26, 2007**

Current U.S. Class: **290/43** ; 290/42; 290/53; 290/54
Current International Class: F03B 13/10 (20060101); H02P 9/04 (20060101)
Field of Search: 290/42,43,44,53,54,55 415/2

Green Datacenter + swimming pool

- Heat generated by VLSD is used to heat the swimming pool
- Cold water from the swimming pool is used to cool the VLSD

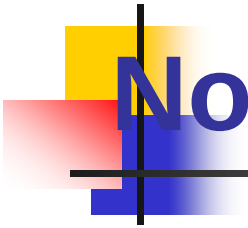
Wind “Green” Power generation for VLSD

Summary

- IT Datacenter is at crossroad
  - Cloud Computing is driving VLSD demand
- Today's cooling approaches are lacking
  - Especially in sub-tropical climates (China/Taiwan)
- Today's power generation is expensive
- Many new ideas – May the best man win

Cloud Challenge

- Non-uniform communication costs
- Power management issues
- Load balance issue
- Networking issues in datacenter



Non-uniform communication costs

- Communications costs within servers, racks, clusters, and clouds are different. (i.e., distance)
- How to justify different data placement strategies?

Power management issues

- High power consumption at data center
  - It is estimated 160 Megawatts for 140M email users with 1GB storage.
- High power consumption at multi-mode mobile handset and laptop computer.

Load balance issue

- A single slower worker can determine the response time of a huge parallel task in parallel computing.
- How to identify such a situation and effectively start redundant workers only for those slower jobs.
- Load consolidation to save power.
- Servers don’t save power proportionally with reduced load.

Networking issues in datacenter

- How to build a flat network, but scalable to the data center scale.
- Solutions ???