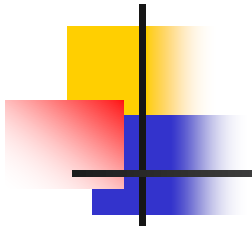
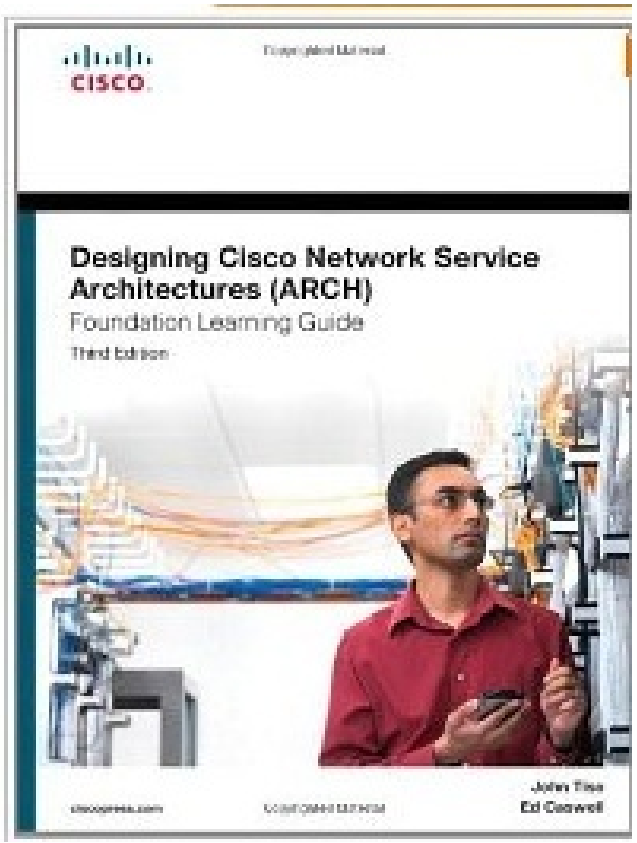


Chapter 0

Cloud Computing and Networking

Books

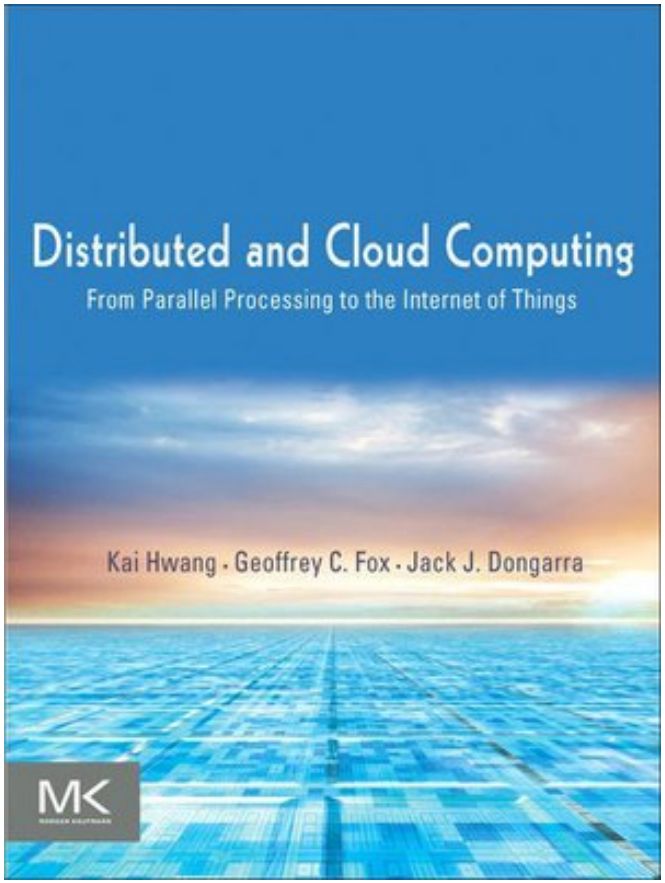
- Designing Cisco Network Service Architectures Foundation Learning Guide (3rd Edition) (Foundation Learning Guides)



Books

- Distributed and Cloud Computing: From Parallel Processing to the Internet of Things (1st Edition)

Kai Hwang



Administration

- Instructor:
  - 曾學文 資工系 副教授
  - 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:
  - (Monday)14:00~17:00; (Tuesday) 14:00~17:00.
- Grade:
  - Projects 25%
  - Paper Presentation 25%
  - Midterm Exam 25%
  - Final Exam 25%

Course Goal

- What is cloud ?
- To know what is the cloud computing
- To understand what is cloud applications or services
- To understand how to design the data center networks of cloud computing.

Cloud applications create huge data to use 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

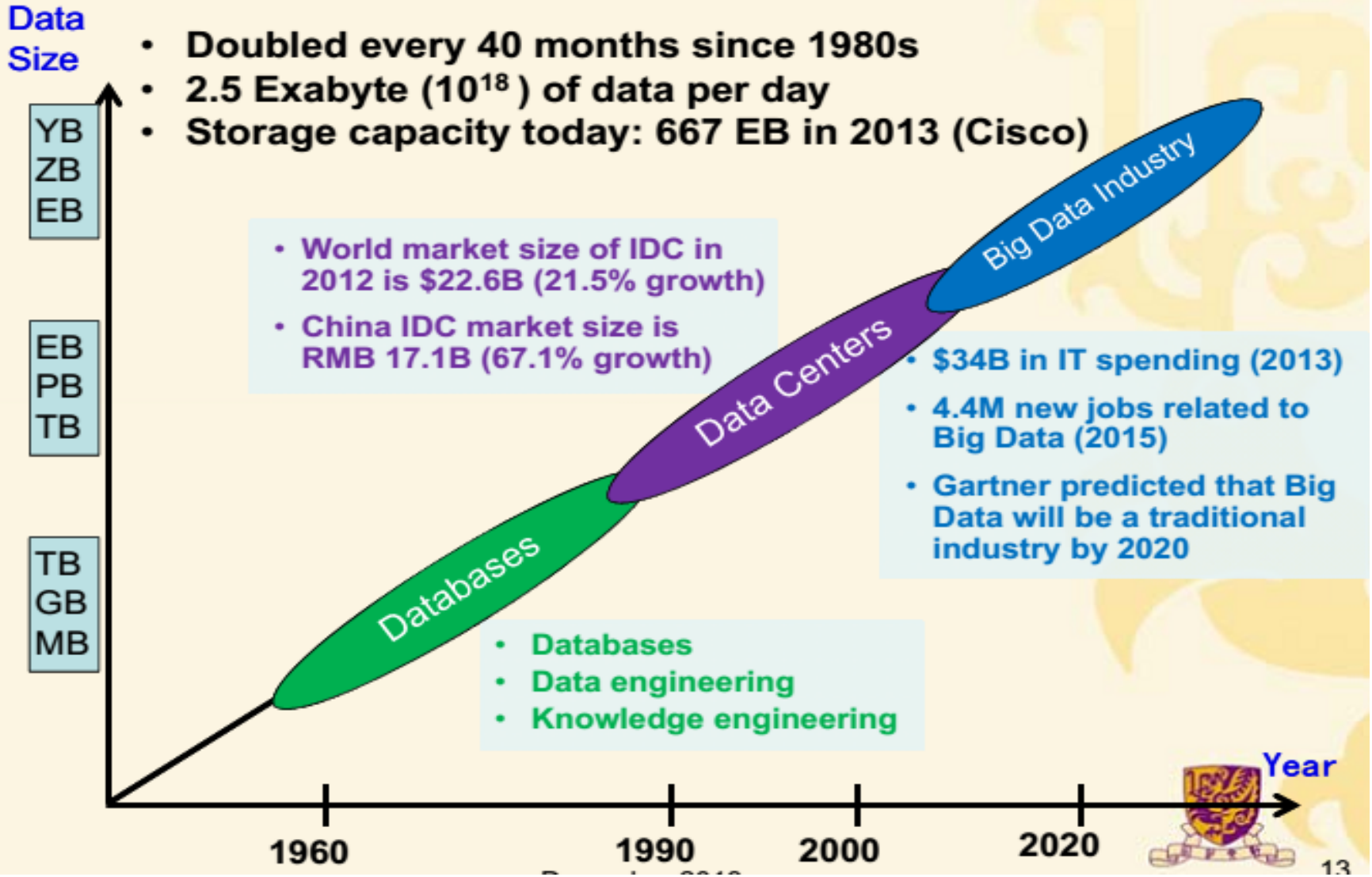
Hugh Data

- Huge multicast traffic in DCN

1 Yottabyte (YB) = 1024 ZB

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

Evolution of the Big Data Industry



- Doubled every 40 months since 1980s
- 2.5 Exabyte (10^{18}) of data per day
- Storage capacity today: 667 EB in 2013 (Cisco)

- World market size of IDC in 2012 is $22.6B (21.5% growth)
- China IDC market size is RMB 17.1B (67.1% growth)

- $34B in IT spending (2013)
- 4.4M new jobs related to Big Data (2015)
- Gartner predicted that Big Data will be a traditional industry by 2020

- Databases
- Data engineering
- Knowledge engineering

Evolution of the big data industry in three development stages.

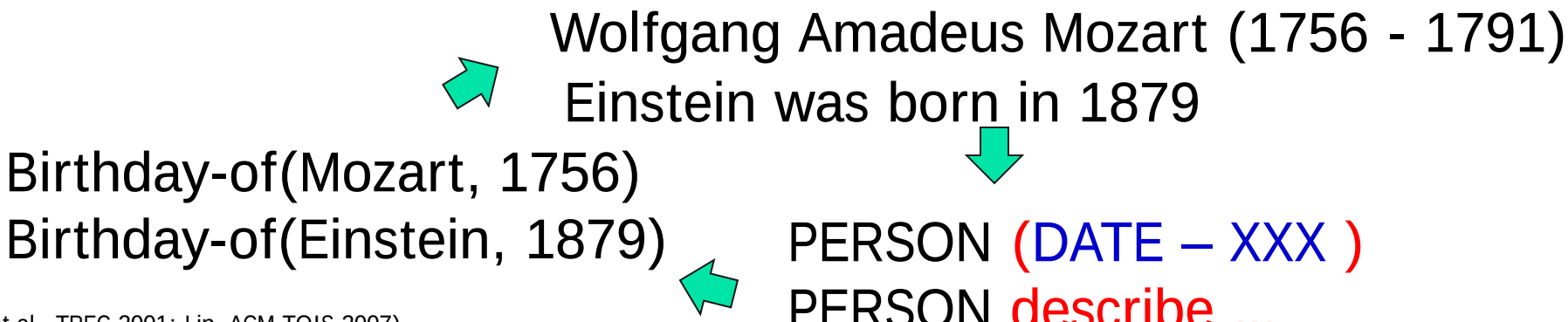
Stage	Databases	Data Centers	Big Data Industry
Time Frame	1960–1990	1980–2010	2010 and beyond
Data Sizes	MB, GB, TB	TB, PB, EB	EB, ZB, YB
Market Size and Growth Rate	Database market, data/knowledge engineering	$22.6 billion market by IDC 2012; 21.5% growth	$34 billion in IT spending (2013), predicted to exceed $100 billion by 2020; 4.4 million new big data jobs (2015)

How to create more data?

- Answering confusing questions
  - Input pattern on the Web
  - Works amazingly well

Who shot Abraham Lincoln? → XXX shot Abraham Lincoln

- Learning relations
  - Input word patterns
  - 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; ...)

Different Use Habits

2005

2013

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

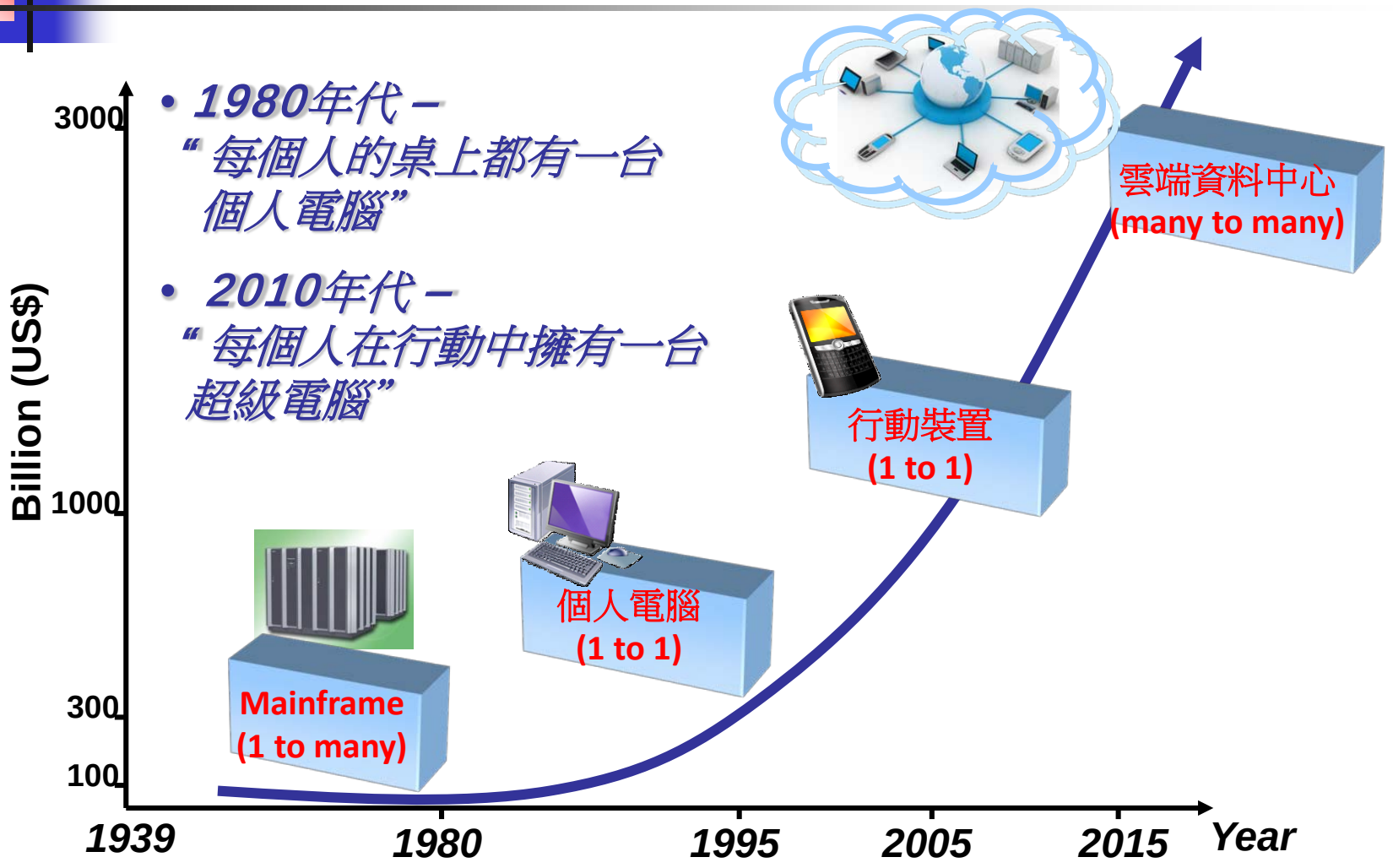
Big Data and Cloud Computing

- It starts with the premise that the data services and architecture should be on servers.
  - We call it cloud computing – they should be in a "cloud" somewhere.

- If you have the right kind of browser or the right kind of access, it doesn't matter whether you have a PC or a Mac or a mobile phone or a BlackBerry or what have you – or new devices still to be developed – you can get access to the cloud.

Danny Sullivan

雲端運算新世代



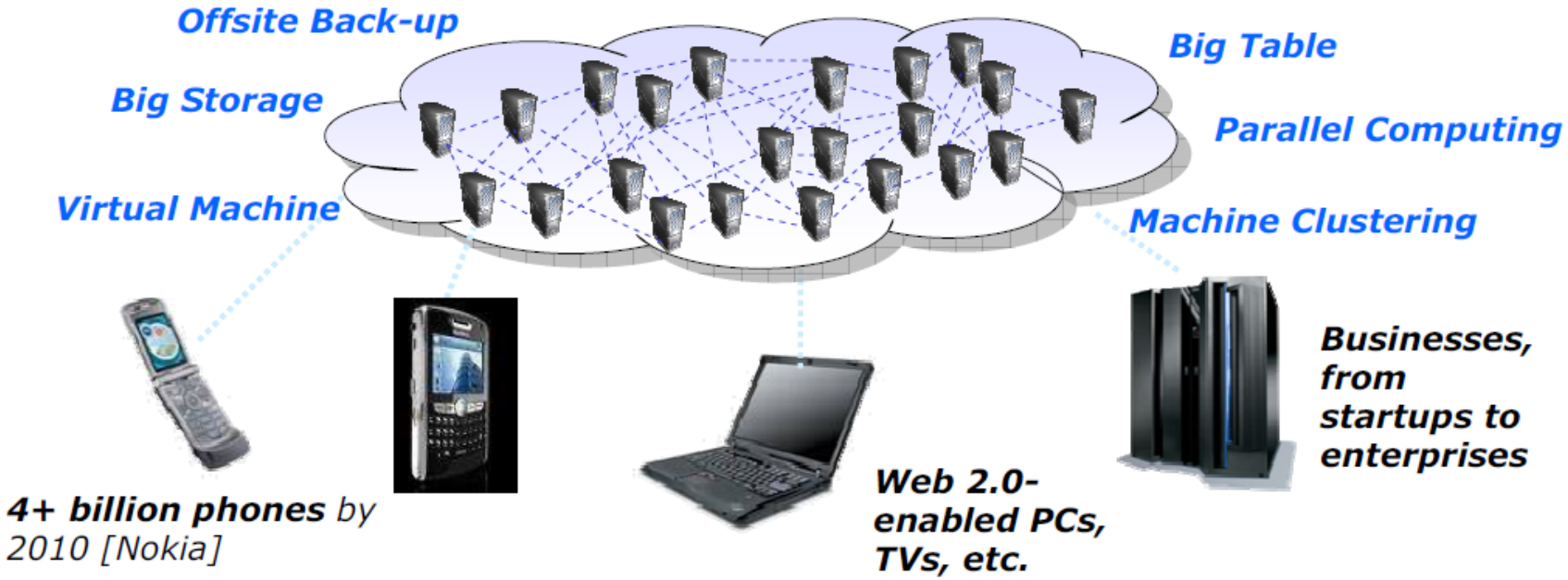
「雲端資料中心」讓電腦運算就像是水、電 一樣，只要連上網路就可以**pay-as-you-go**無限量提供服務。

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

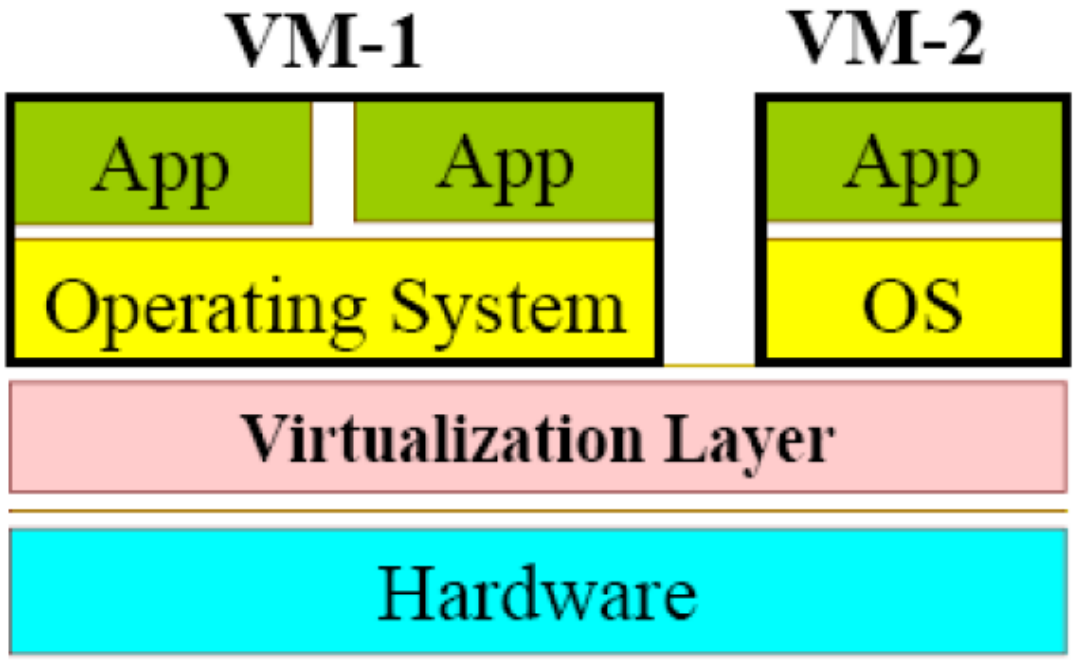
(2)使用者可以利用各種具備網際網路連線能力的電腦終端裝置(device)，無所不在(ubiquitous)的使用資料與服務



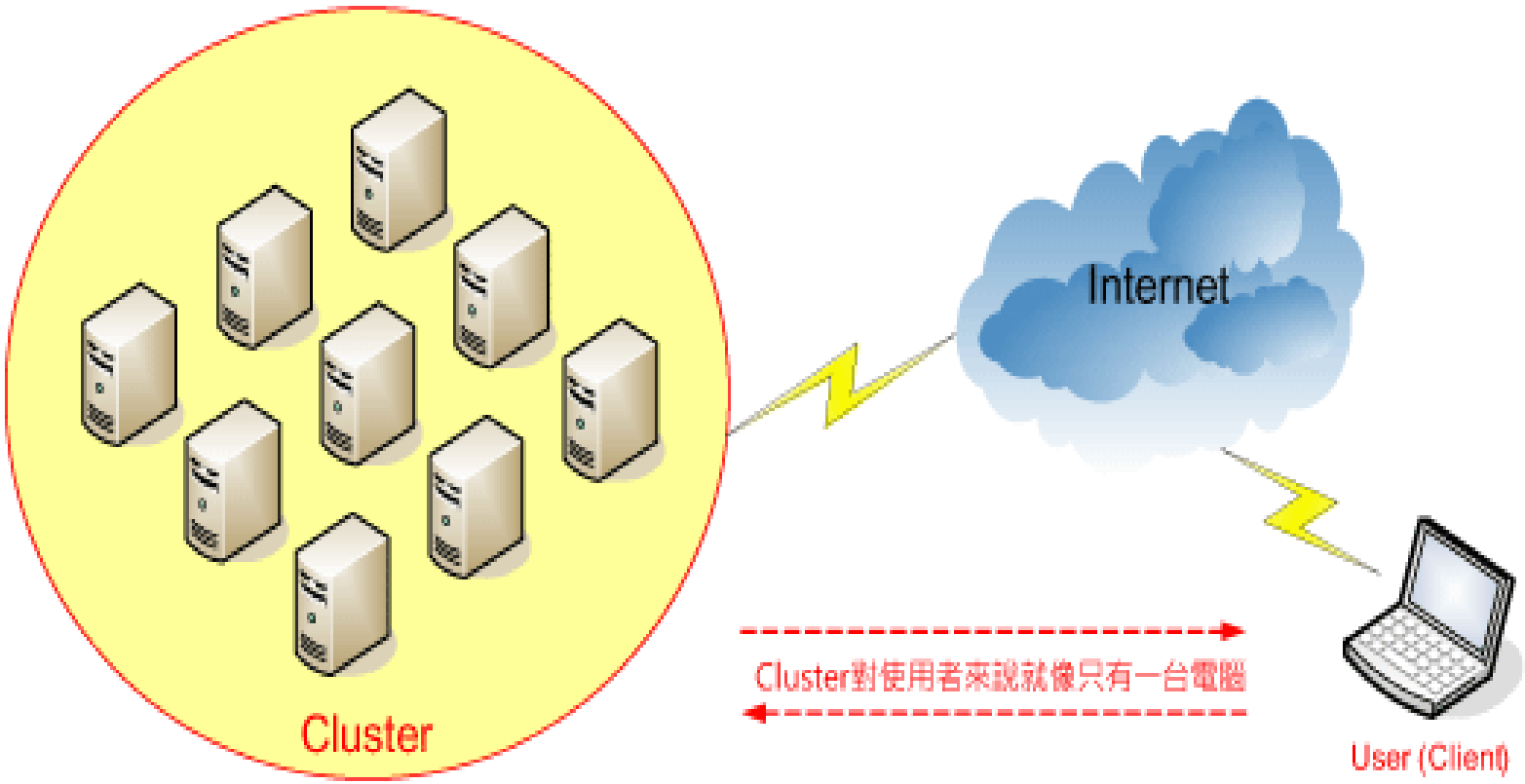
虛擬運算技術

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



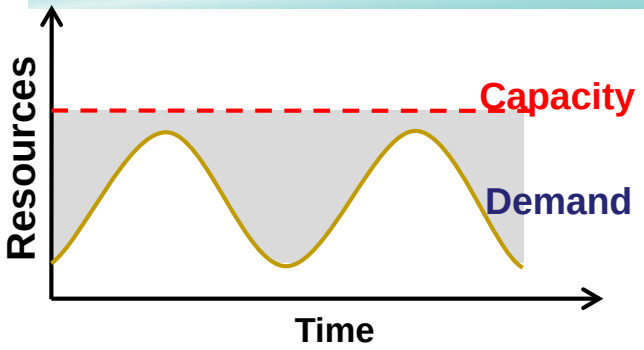
叢集運算技術



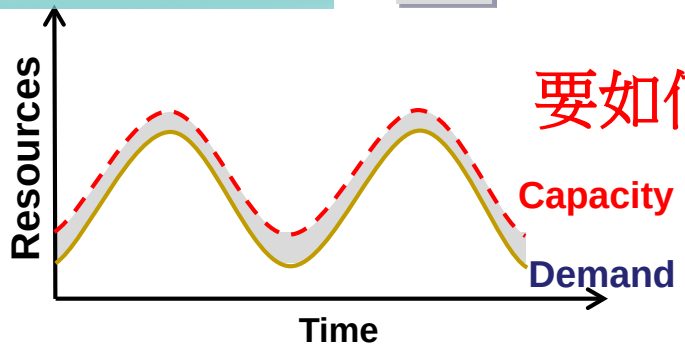
雲端運算經濟學

Pay by use instead of provisioning for peak

Unused resources



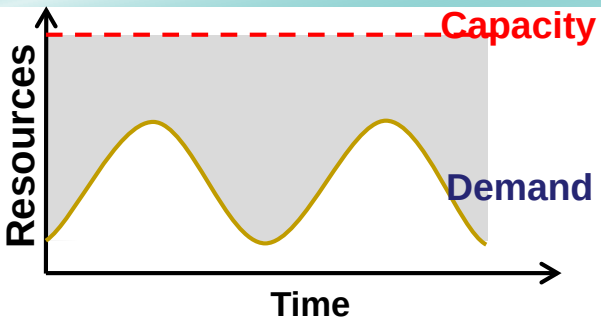
Static data center



Data center in the cloud

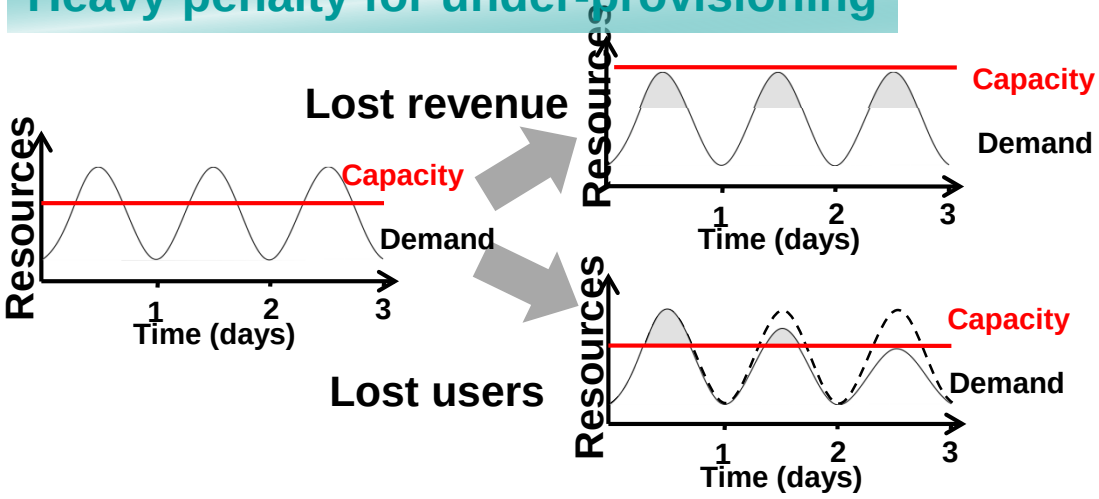
要如何做好資源管理

Risk of over-provisioning: underutilization



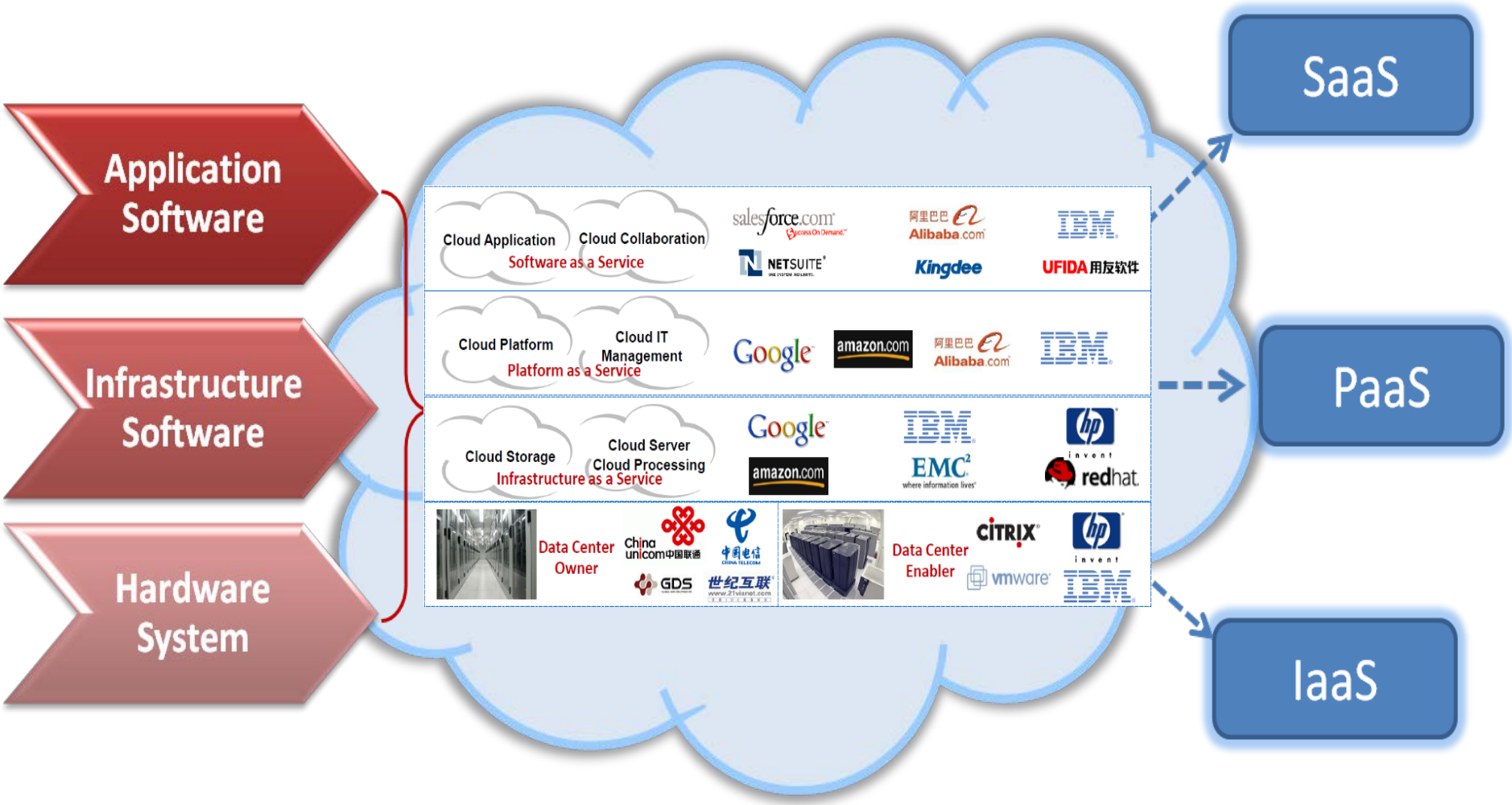
Static data center

Heavy penalty for under-provisioning



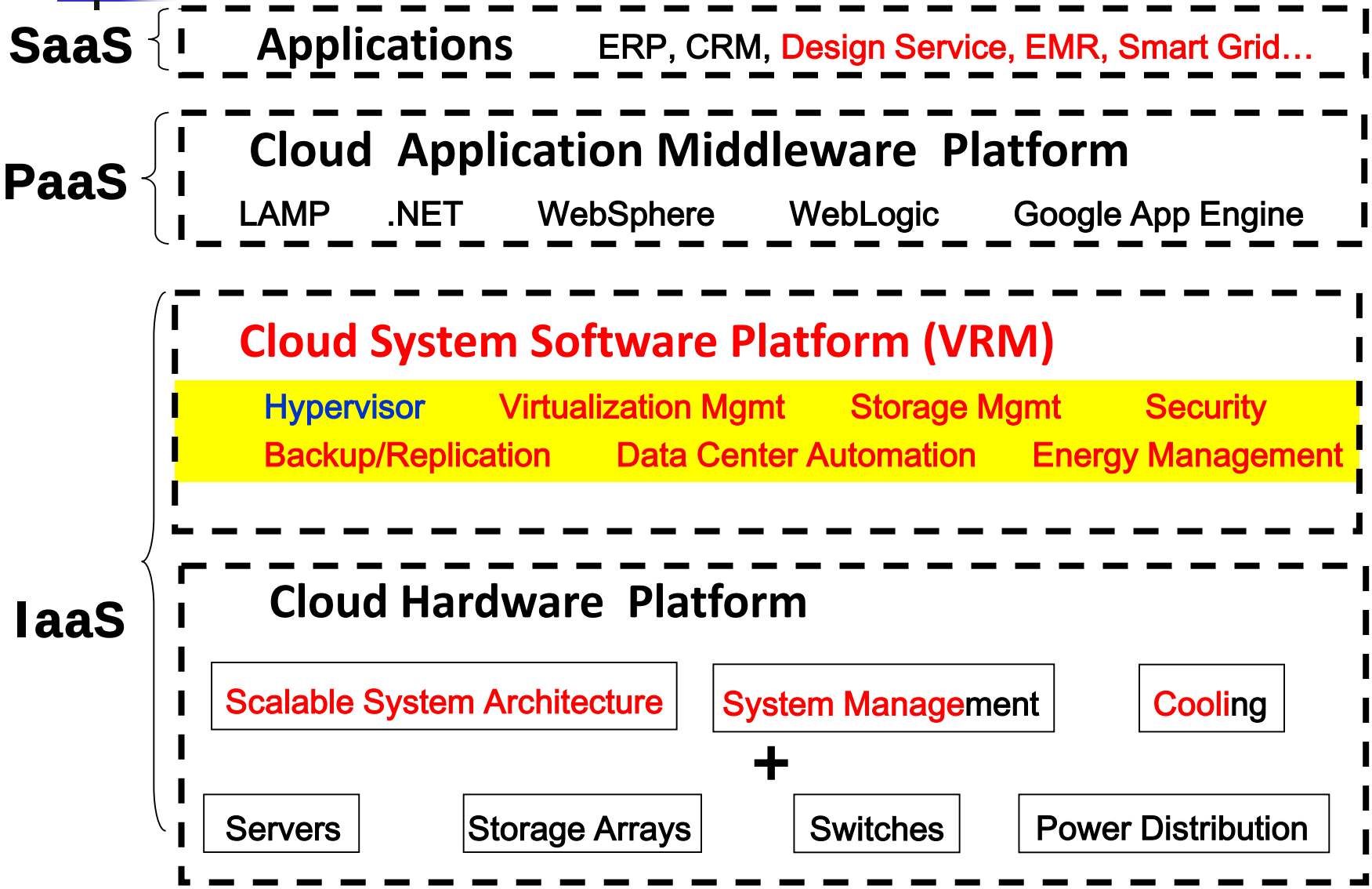
資料來源：UC Berkeley RAD Lab

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

Market share of various cloud platforms

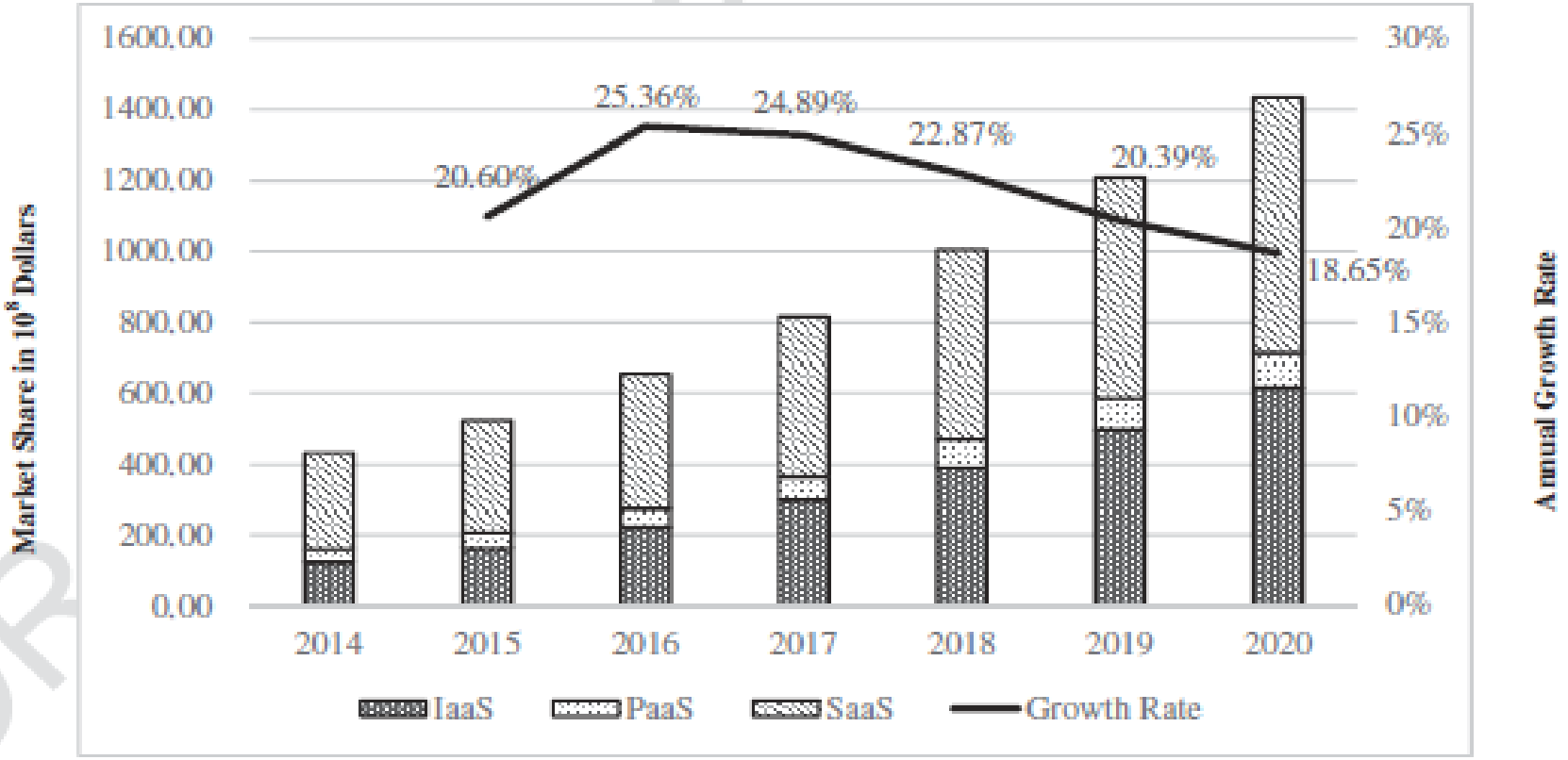
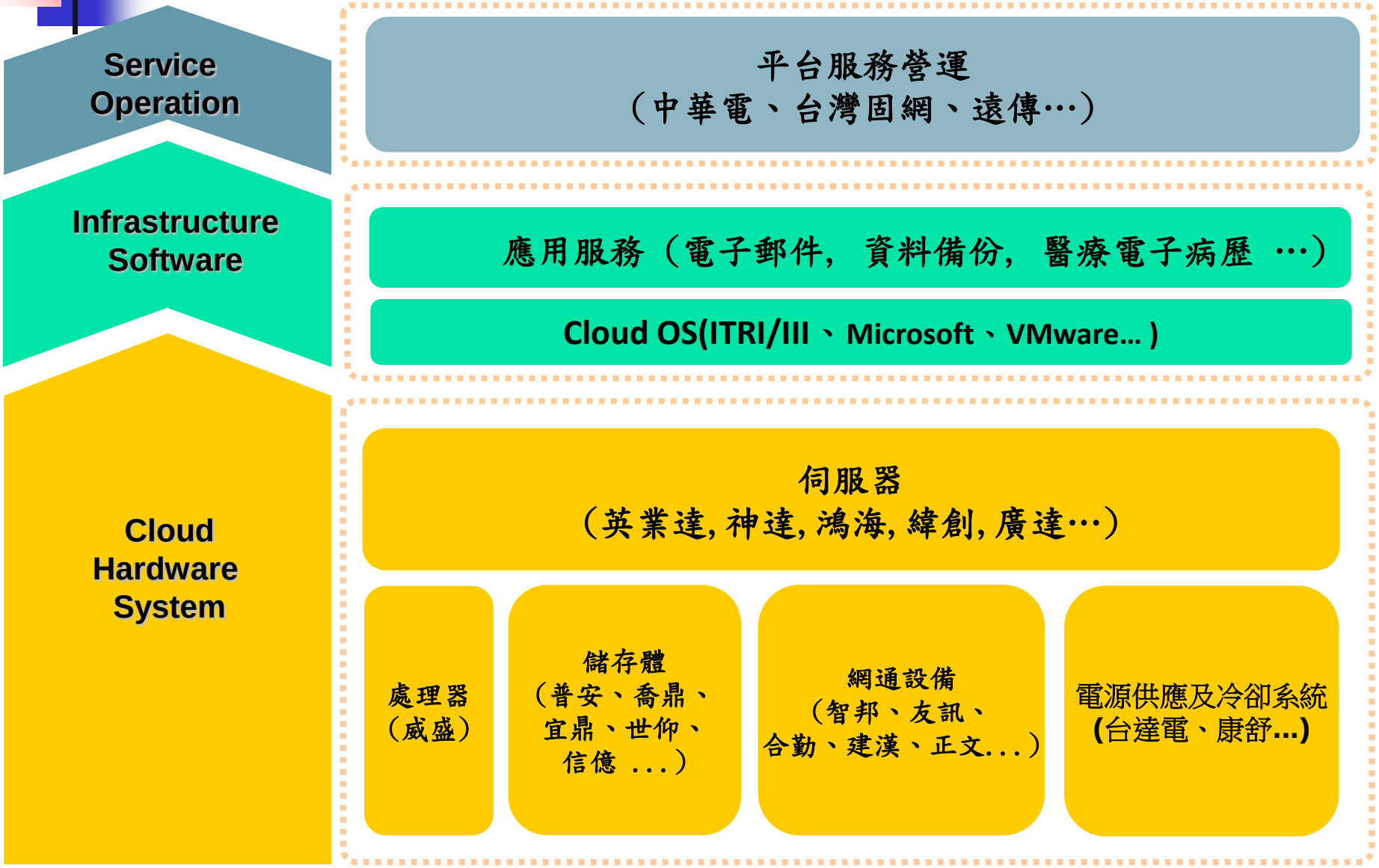
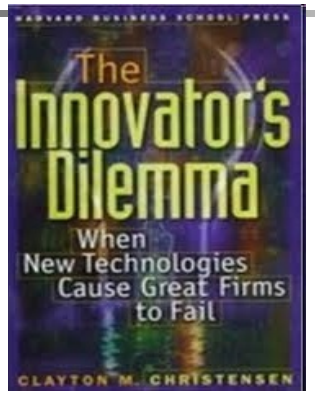


Figure 1.13
Worldwide distribution of cloud service models and the growth rate based on projections by Gartner Research from 2014–2020.

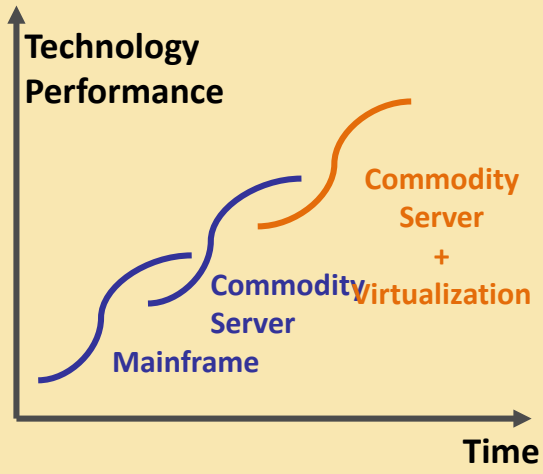
台灣雲端上下游產業價值鏈



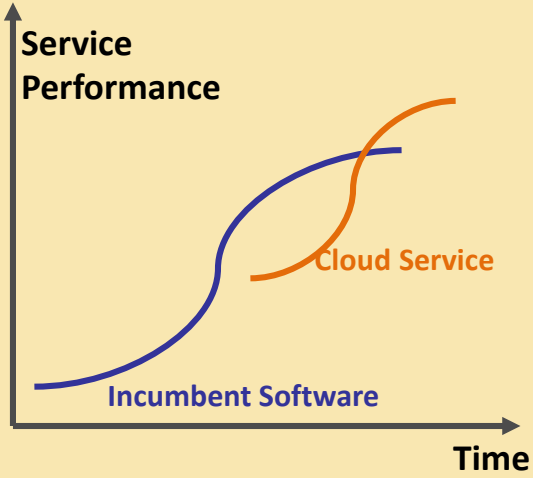
雲端運算帶來破壞式創新



Clayton M. Christensen



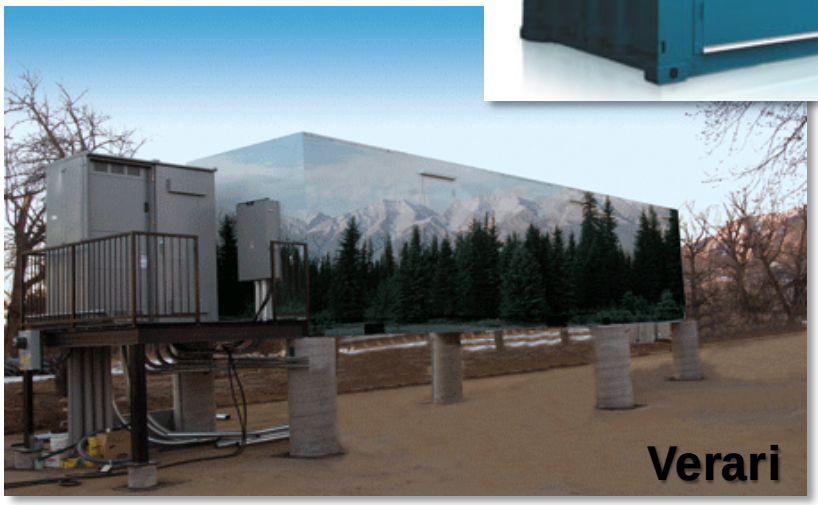
- 雲端技術是"Sustaining Innovation"
  - 雲端技術提供更好的功能與表現，且滿足相同企業用戶
  - 雲端技術無法產生破壞性創新，而是技術的延續



- 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

雲端貨櫃型電腦

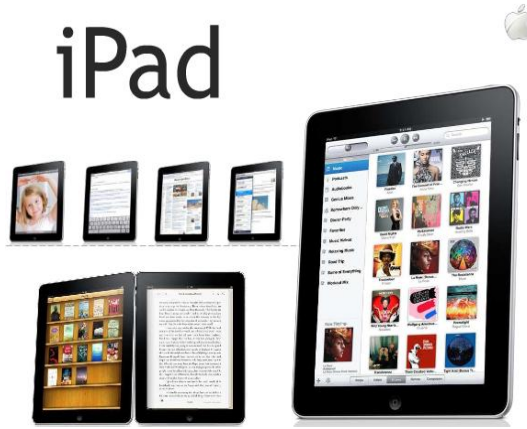


Source: Daniel Costello, Microsoft Data Center Transformation, Microsoft WinHEC 2008

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

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

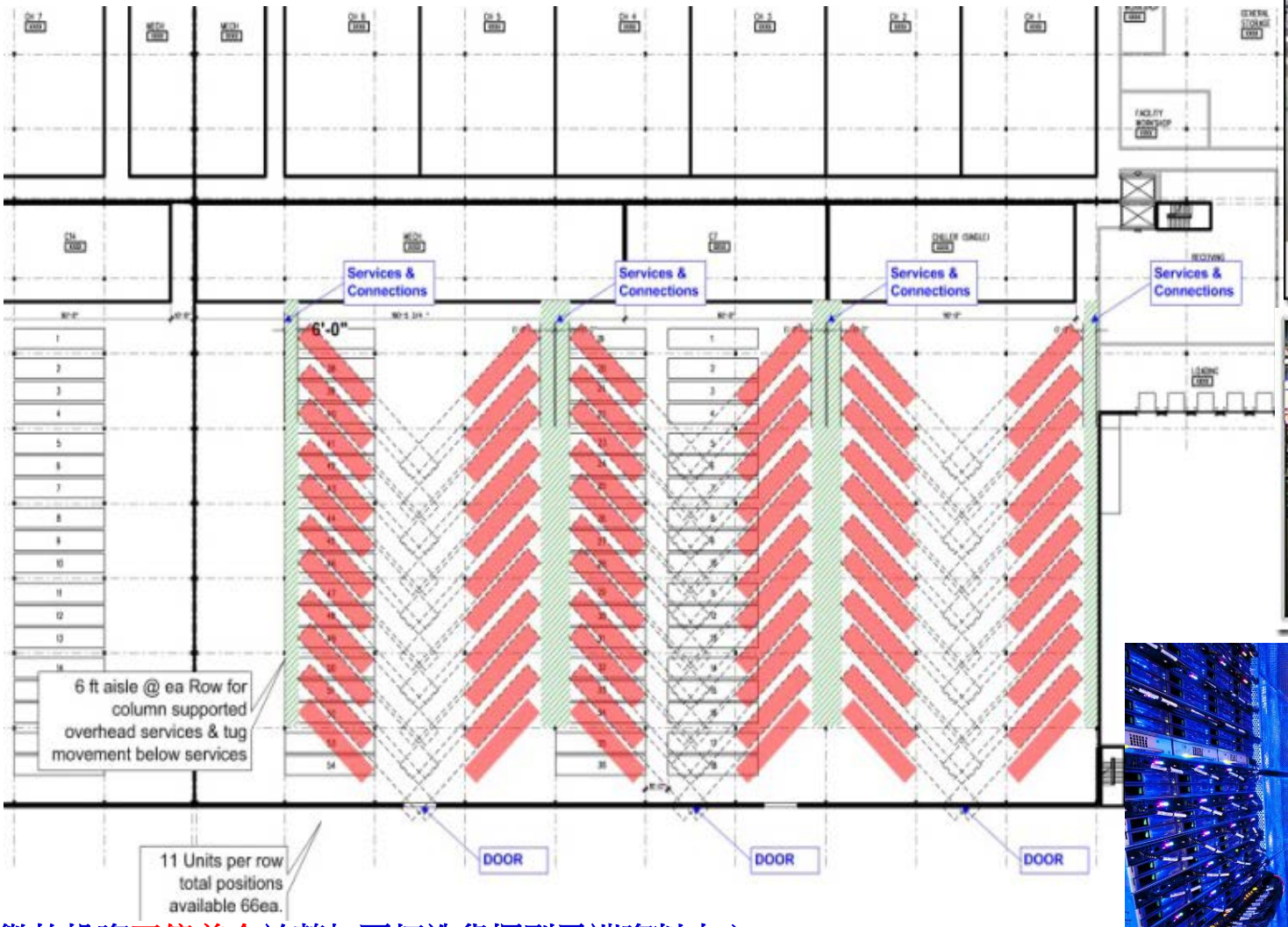
1.Apple's existing Newark, CA., Data Center is around 109,000 square feet--the new one is over 500,000. That represents either a ridiculously big scaling-up of business or a whole new thing 2.500,000 square feet is among the largest centers being built in the World on a single site. Microsoft's new one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

Big Data

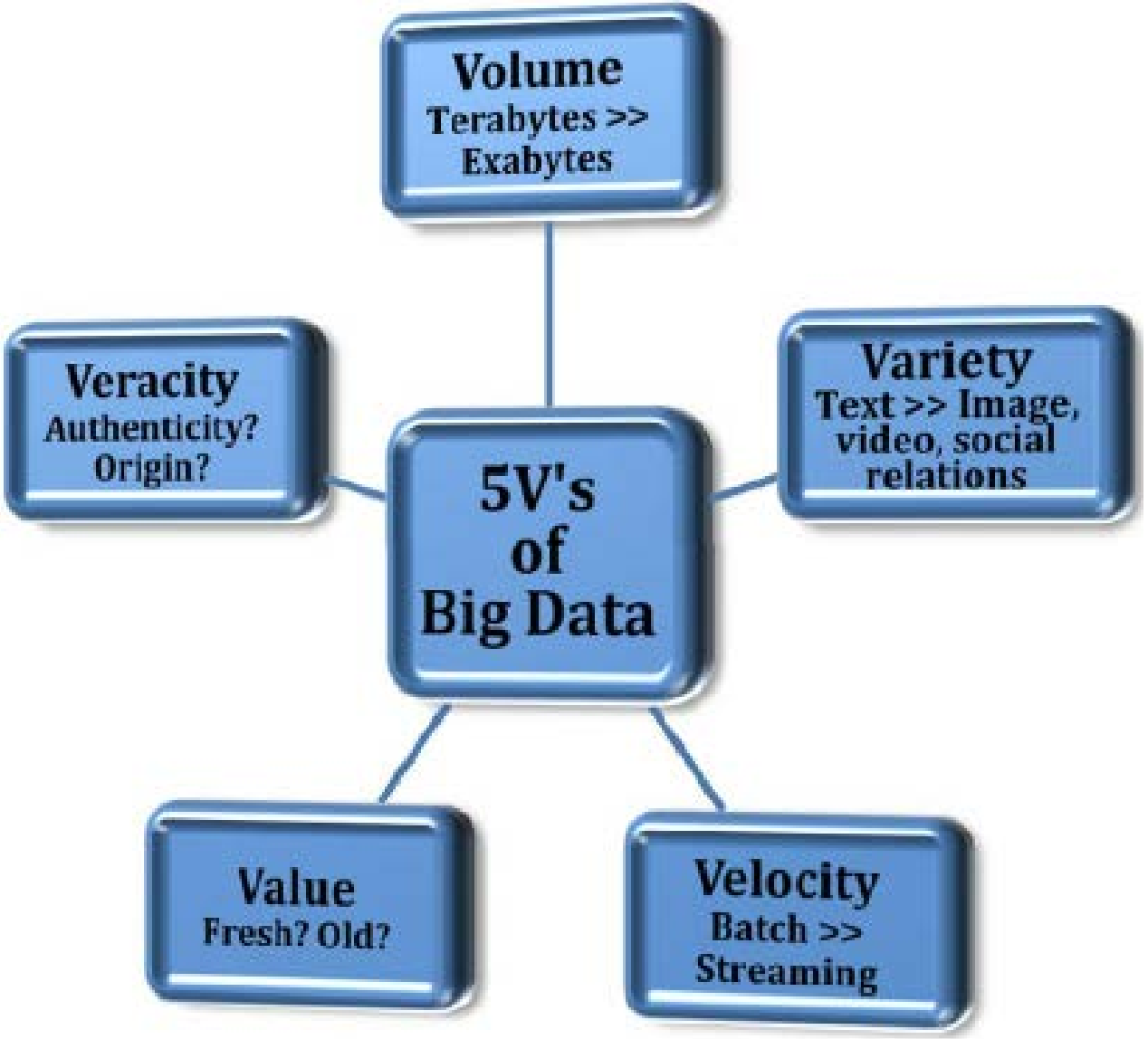
What is Data Science?

- **Data science** is the extraction of useful knowledge directly from data through a process of discovery, hypothesis, and hypothesis analysis

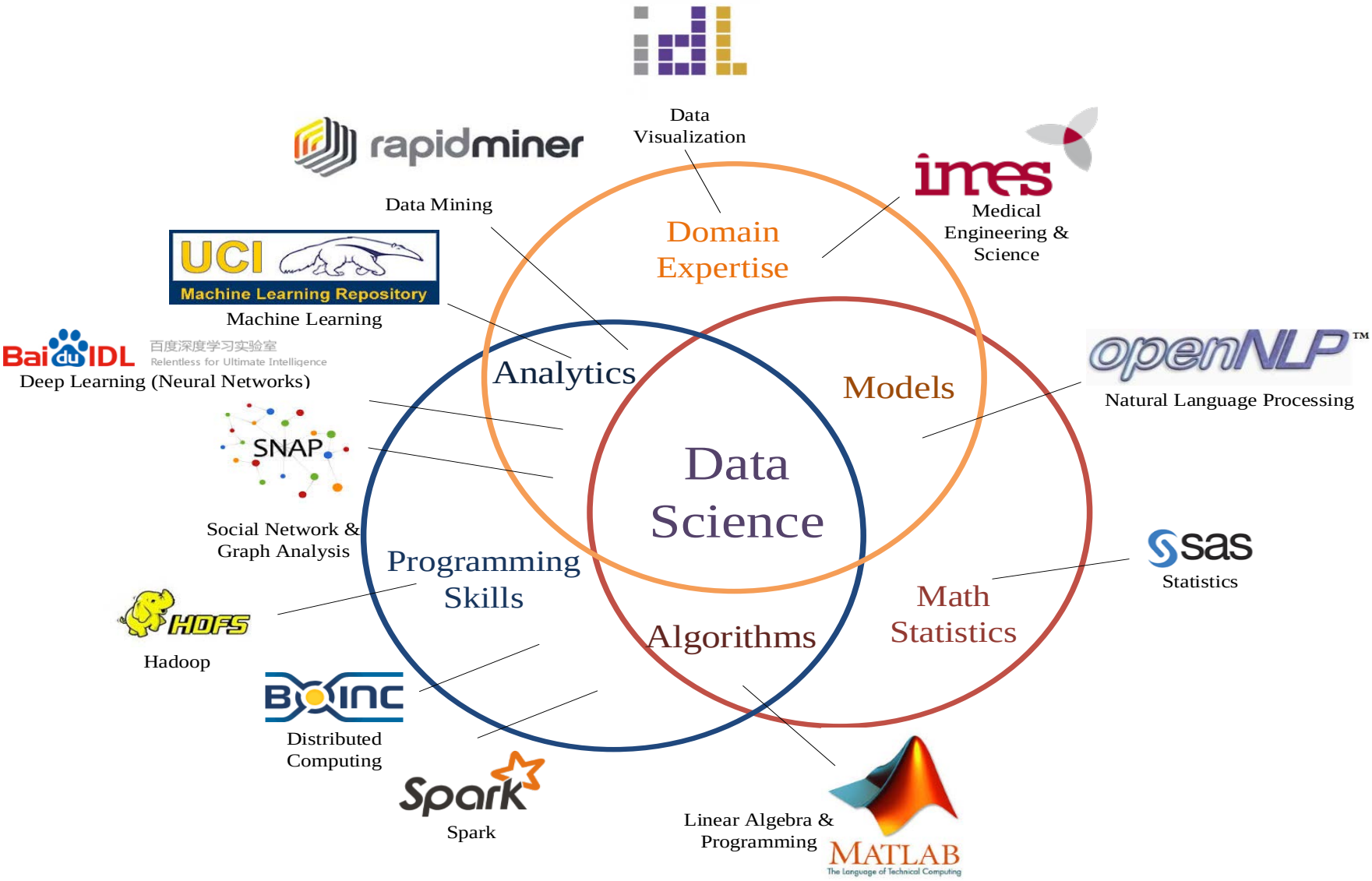
- A **data scientist** is a practitioner who has sufficient knowledge of the overlapping regimes of expertise in business needs, domain knowledge, analytical skills, and programming expertise to manage the end-to-end scientific method process through each stage in the big data lifecycle

Big data refers to digital data volume, velocity and/or variety whose management requires scalability across coupled horizontal resources

The Five v's of Big Data



Today's Dig Data Software Libraries



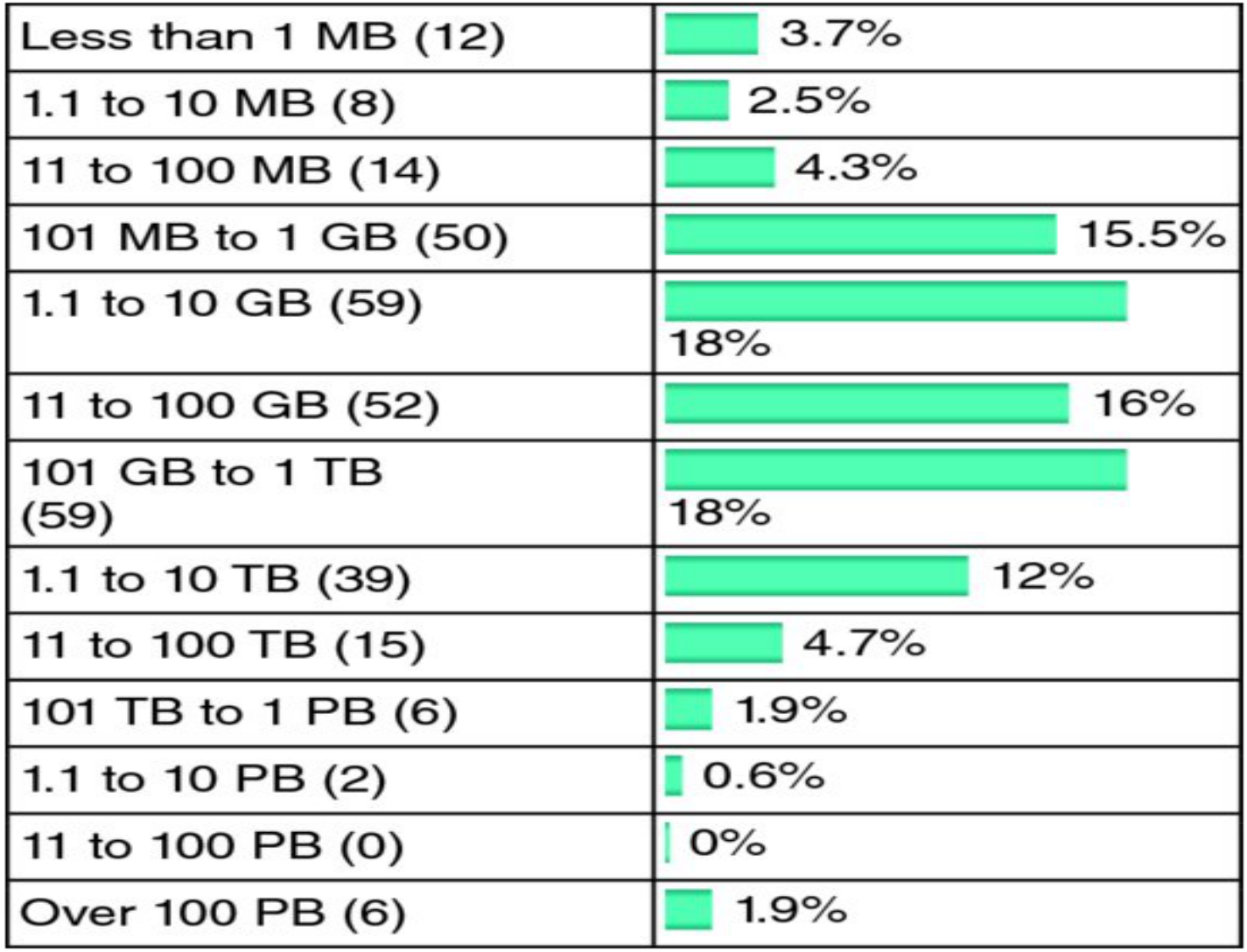


Figure 1.1 Results from a KDnuggets Poll about Largest Data Sets Analyzed

Source: www.kdnuggets.com/polls/2013/largest-dataset-analyzed-data-mined-2013.html.

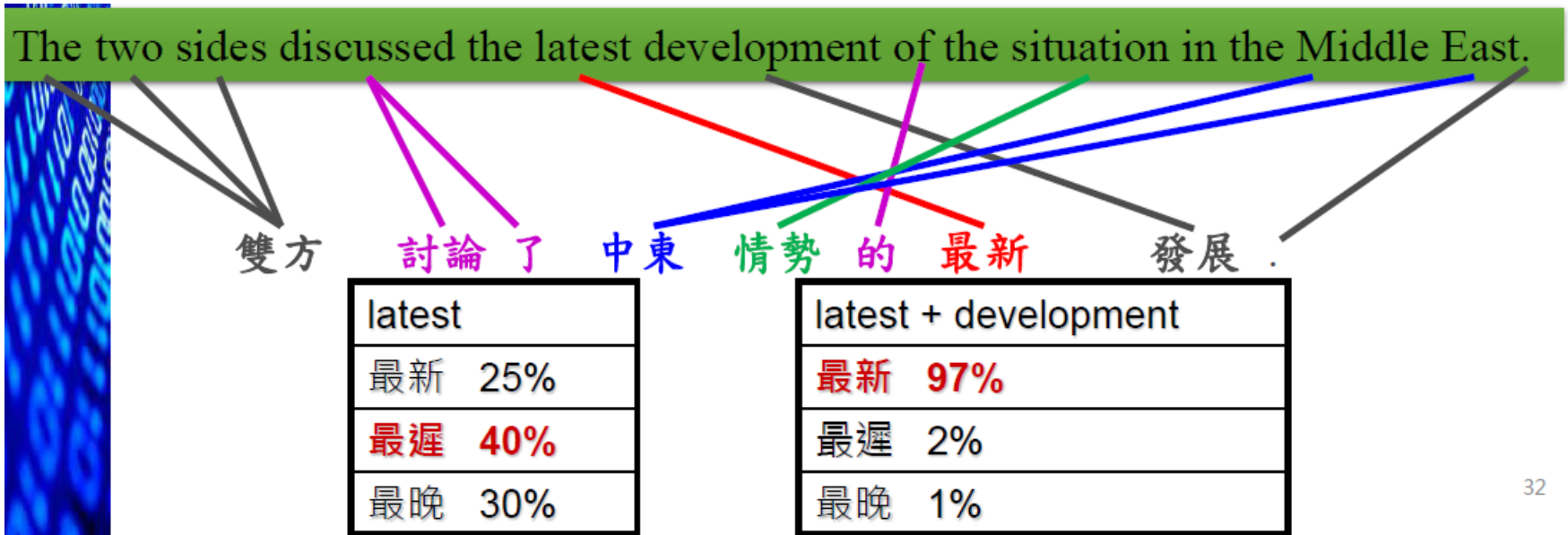
A main obstacle to fully harnessing the power of big data using analytics is the lack of skilled resources and "data scientist" talent required to exploit big data. In another poll ran by KDnuggets in July 2013, a strong need emerged for analytics/big data/data mining/data science education.[4] It is the purpose of this book to try and fill this gap by providing a concise and focused overview of analytics for the business practitioner.

Basic Dig Data Principles

- Structured vs. unstructured data and indexing
- Identification, de-identification and re-identification
- Ontologies and semantics
- Data introspection and reduction
- Design, construction, operation and description
- Data integration and software interoperability
- Immutability and immortality
- Data measurement methods
- Data range, denominators, trending, and estimation

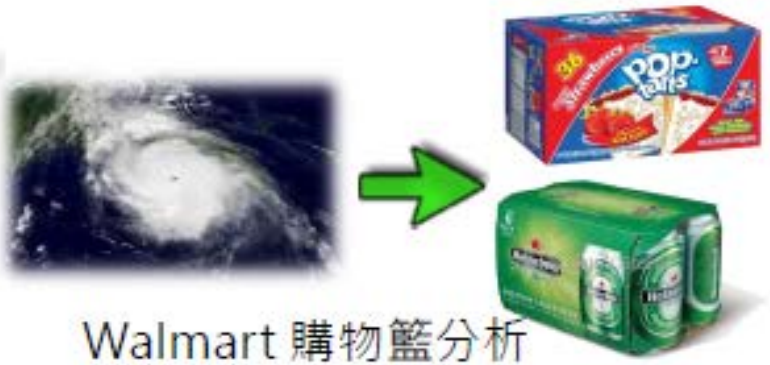
巨量資料分析

- 李開復：『一斤數據優於一兩演算法』
- 人工智慧使用在挖掘巨量資料，利用雲端中的巨量資料來解決人工智慧中難以用數學方法精確描述的、隱喻的複雜問題，且能接受精確度有瑕疵的預測
- 藉由收集龐大的雙語語料庫，訓練「自動翻譯」的統計模型，收集巨量的搜尋關鍵字，訓練「即時快搜(instant search)」的統計模型



商人的水晶球- Big Data

Walmart 購物籃分析



Walmart 購物籃分析

Target懷孕預測模型

7-ELEVEN 753感冒指數

FamilyMart 關東煮熱銷指數

Target百貨

- 美國的知名連鎖賣場Target寄發孕婦用品廣告給中學女生。
- Target特別重視懷孕的預測
- 有了小孩是家庭的一件大事，未來十餘年的物品購買習慣都會跟著改變。
- Target根據使用者的購買紀錄，包括使用的化妝品和所吃的維他命，來決定一位婦女的「懷孕預測指數」。
- 購買的商品和個人資料來預測懷孕指數值，正確率高達87%。

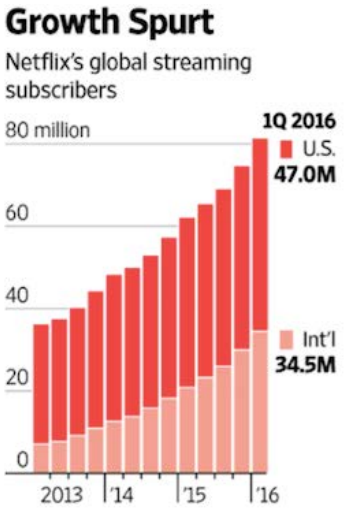
雲端雲算與巨量資料驅動企業創新

- 昔日錄影帶出租連鎖店龍頭企業
- 成立於1985年，2004年為發展發展高峰期，擁有超過9,000家店面
- 於2010年9月23日宣告破產，被Dish Network收購
- 2013年11月宣佈將關閉所有美國直營店面

- 今日雲端影片服務領導企業，運用社群媒體上的巨量資料持續創新產品與服務
- 成立於1997年，業務始於網路上提供影片DVD出租服務
- 1999年推出影片月租訂閱服務模式
- 2007年推出線上影片串流服務
- 2013年善用big data精準掌握觀眾的偏好，推出深受市場歡迎的自製影集節目《紙牌屋》

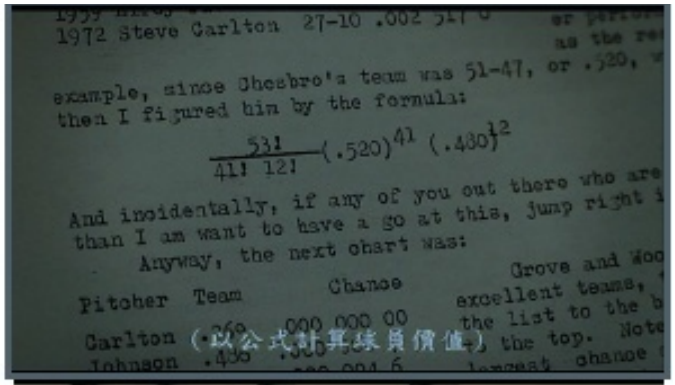
Netflix

- Netflix將觀眾尋找、評論和觀看影片的行為記錄下來，工程師再將這些數據整理為有意義的資料，後來發現有三個元素最受Netflix的美國觀眾歡迎：大衛芬奇（David Fincher）導演、BBC出品、凱文史派西（Kevin Spacey）主演。
- Netflix根據這三個元素打造自製影集《紙牌屋》，光是2013第一季就新增了305萬用戶，單季營收衝破10億2400萬美元，較2012年同期激增18% 。



魔球成真「算」出勝利

- 《魔球》（Moneyball）是一部2011年棒球題材的美國劇情片，由布萊德·彼特主演。影片根據邁克爾·路易斯於2003年發表的同名書籍 Moneyball 改編而成，講述奧克蘭運動家在球隊總經理比利·比恩(Billy Beane）的帶領下，透過數字統計分析，打破傳統，讓一支沒有明星球員、名不見經傳的小球隊，也能奪得冠軍 ！



TSMC大資料分析創造技術優勢

- 從40奈米進入到20奈米製程後，資料量大幅提升了24倍，而現有台積電一個超大型晶圓廠(GigaFab)，平均每月可以生產10萬~20萬片的晶圓，若一個12吋晶圓廠來計算，每秒可以產出約100萬筆的資料
- 台積電使用HBase作為大資料分析架構底層的資料基礎設施，並導入Hadoop平行處理系統，搭配SPSS、SAS及R語言，透過資料前處理、過濾、特徵萃取等步驟，將龐大機臺製程資料拿來進行資料採礦，找到關鍵因子，最後經由資料視覺化工具，將分析結果加以呈現。

Data Sources and Collection

- Business transactions
- Textual and multimedia contents
- Qualitative knowledge data
- Scientific discovery
- Sensing data from the IoT
- Social media sources

The Internet of Things (IoT)

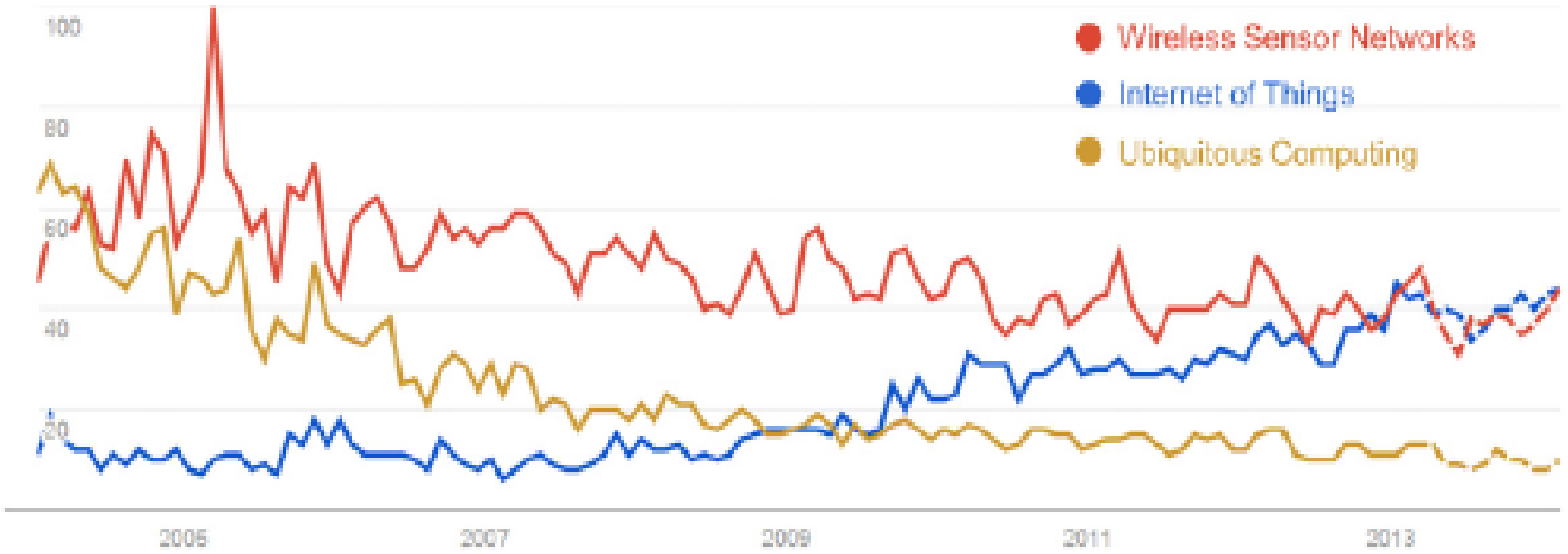
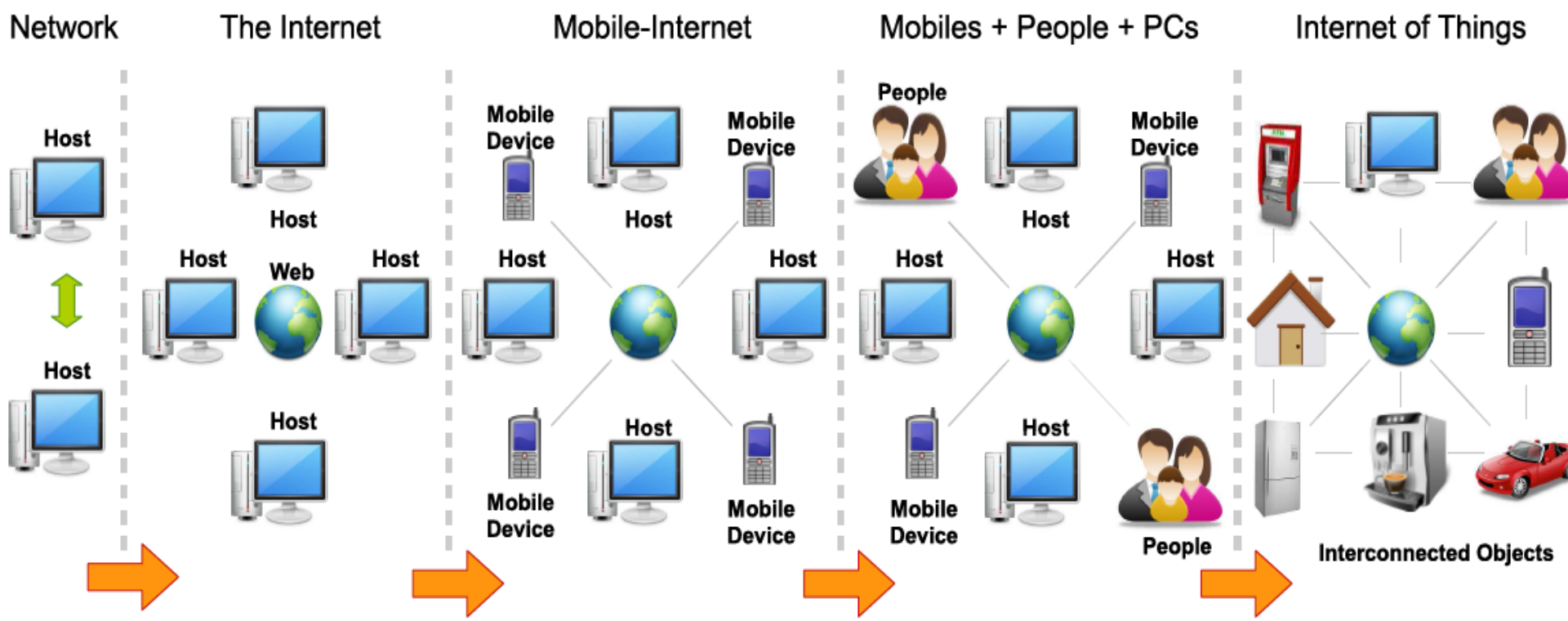
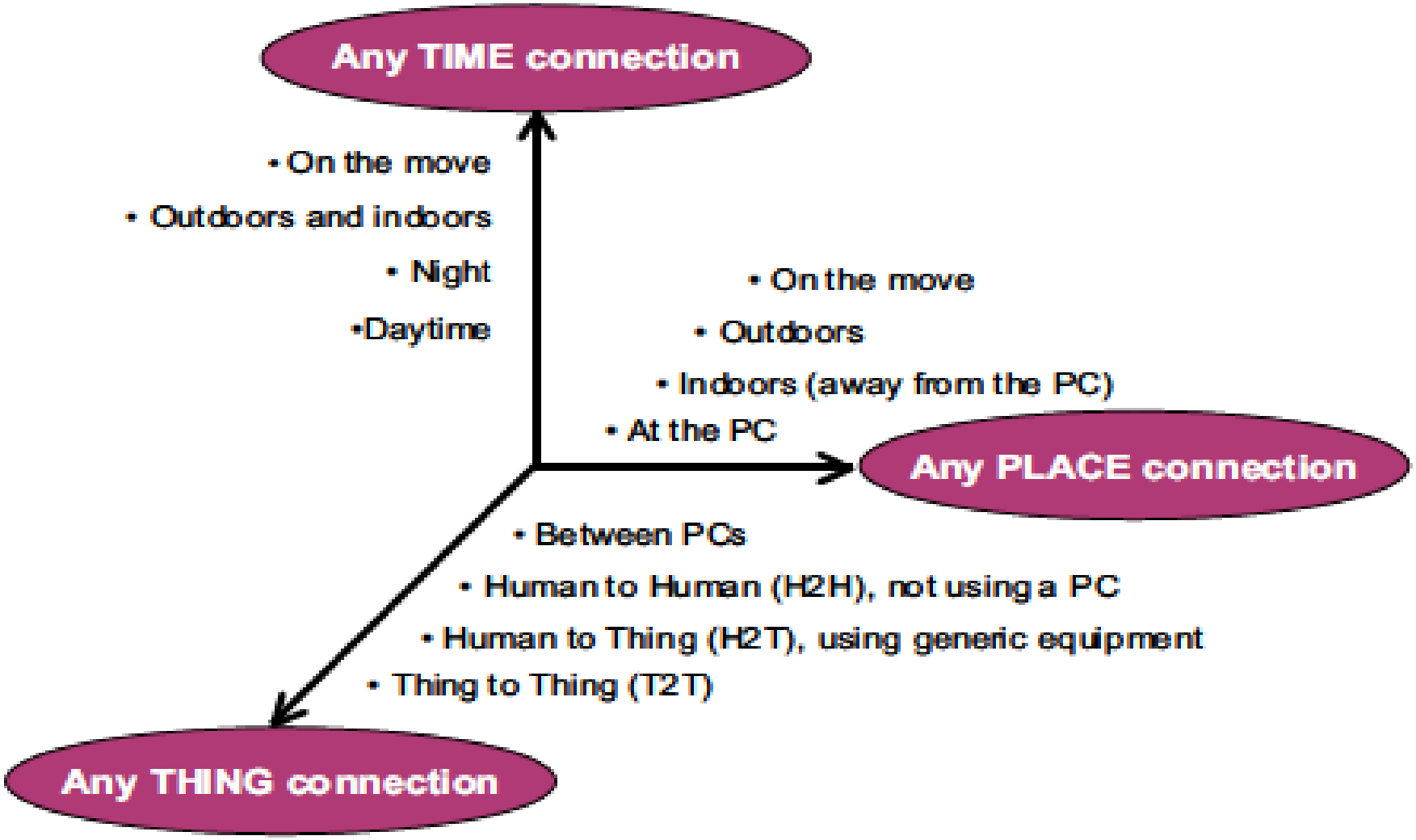


Fig. 3. Google search trends since 2004 for terms Internet of Things, Wireless Sensor Networks, Ubiquitous Computing.

Evolution of the Internet of Things



Opportunities of IoT in 3 Dimensions

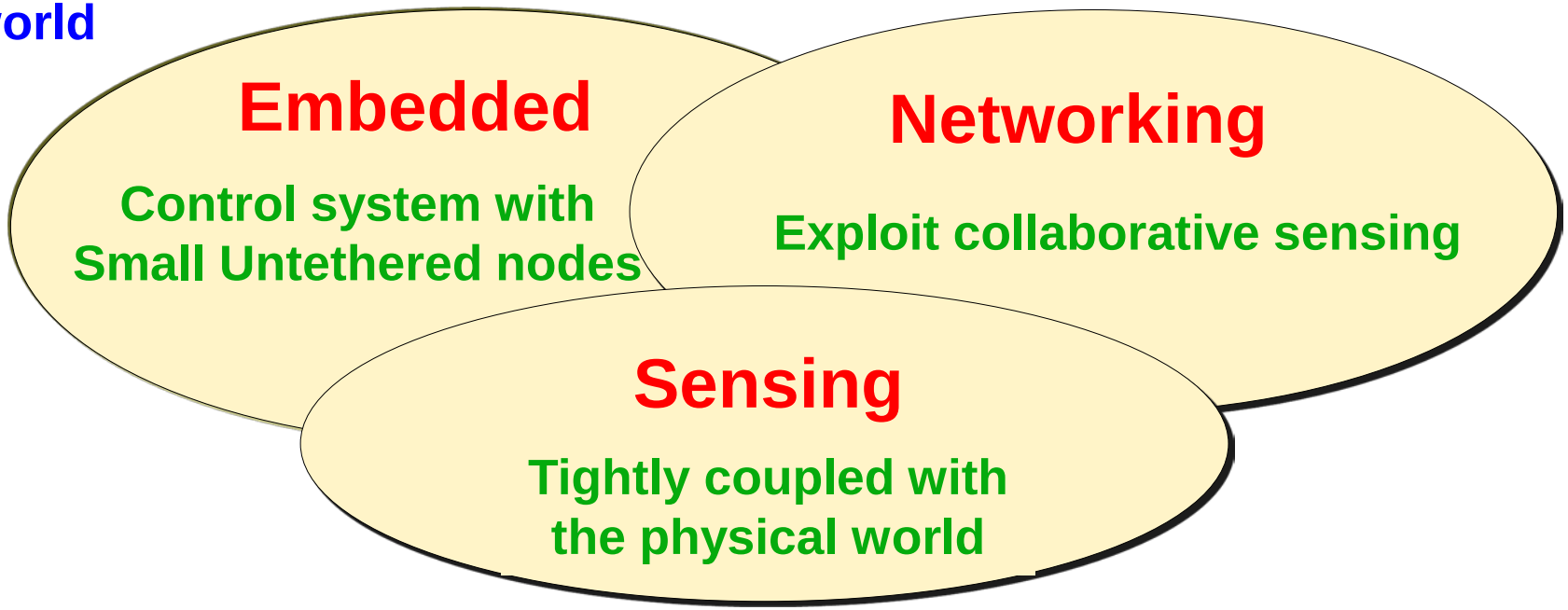


(courtesy of Wikipedia, 2010)

IoT Enabling Technologies

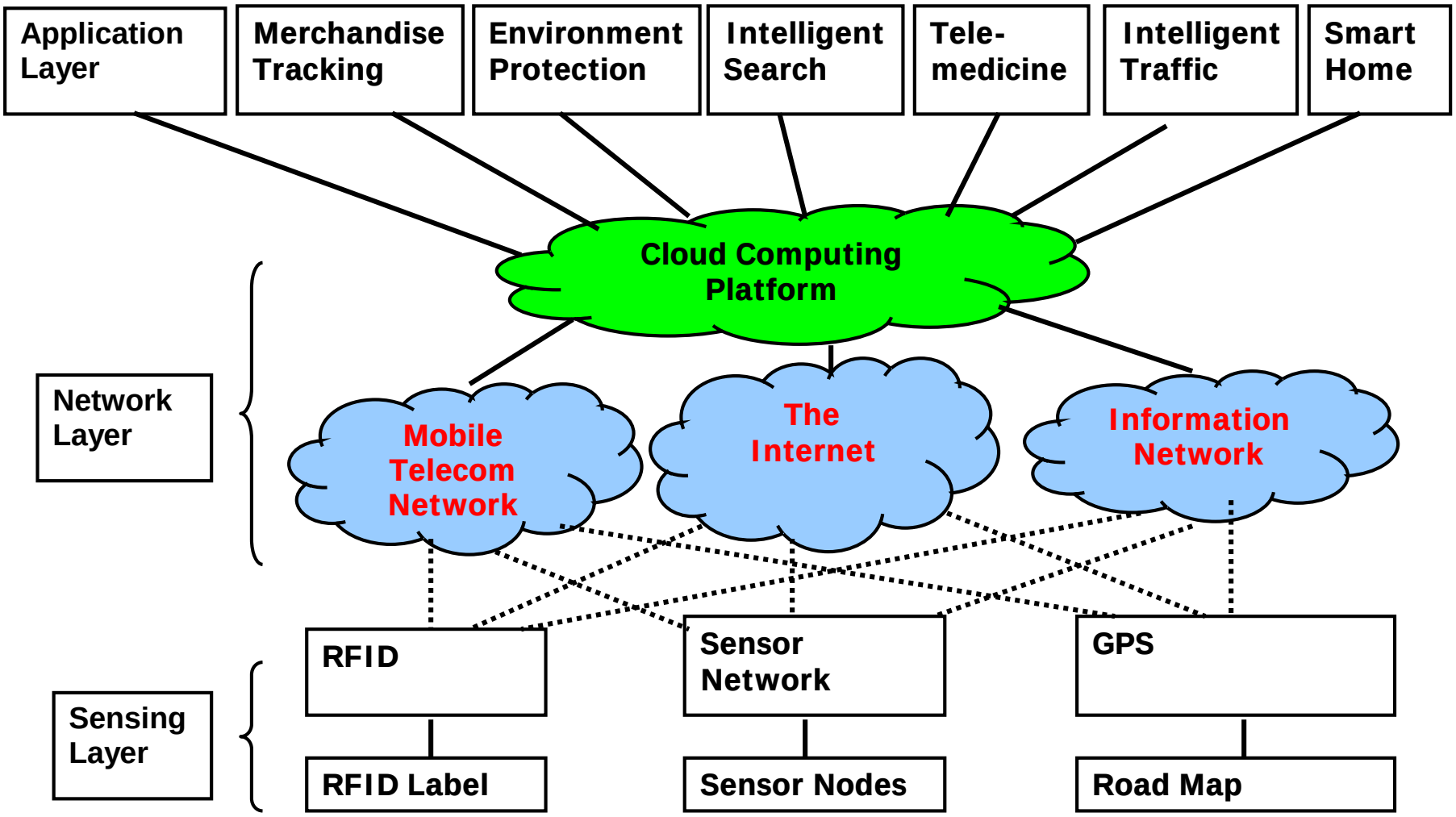
Embed numerous distributed devices to monitor and interact with physical world

Network devices to coordinate and perform higher-level tasks

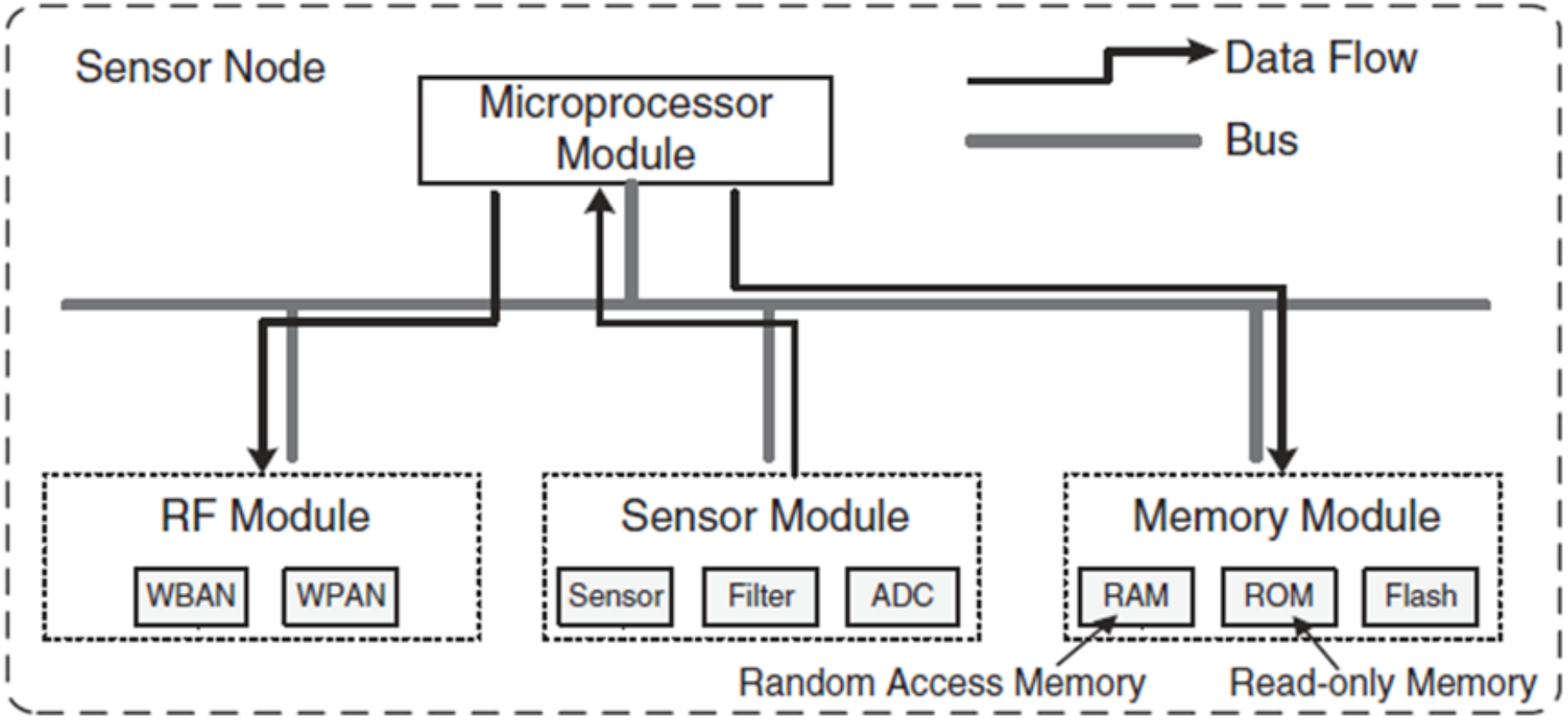


Exploit spatially/temporally dense, in situation or remote sensing or actuation

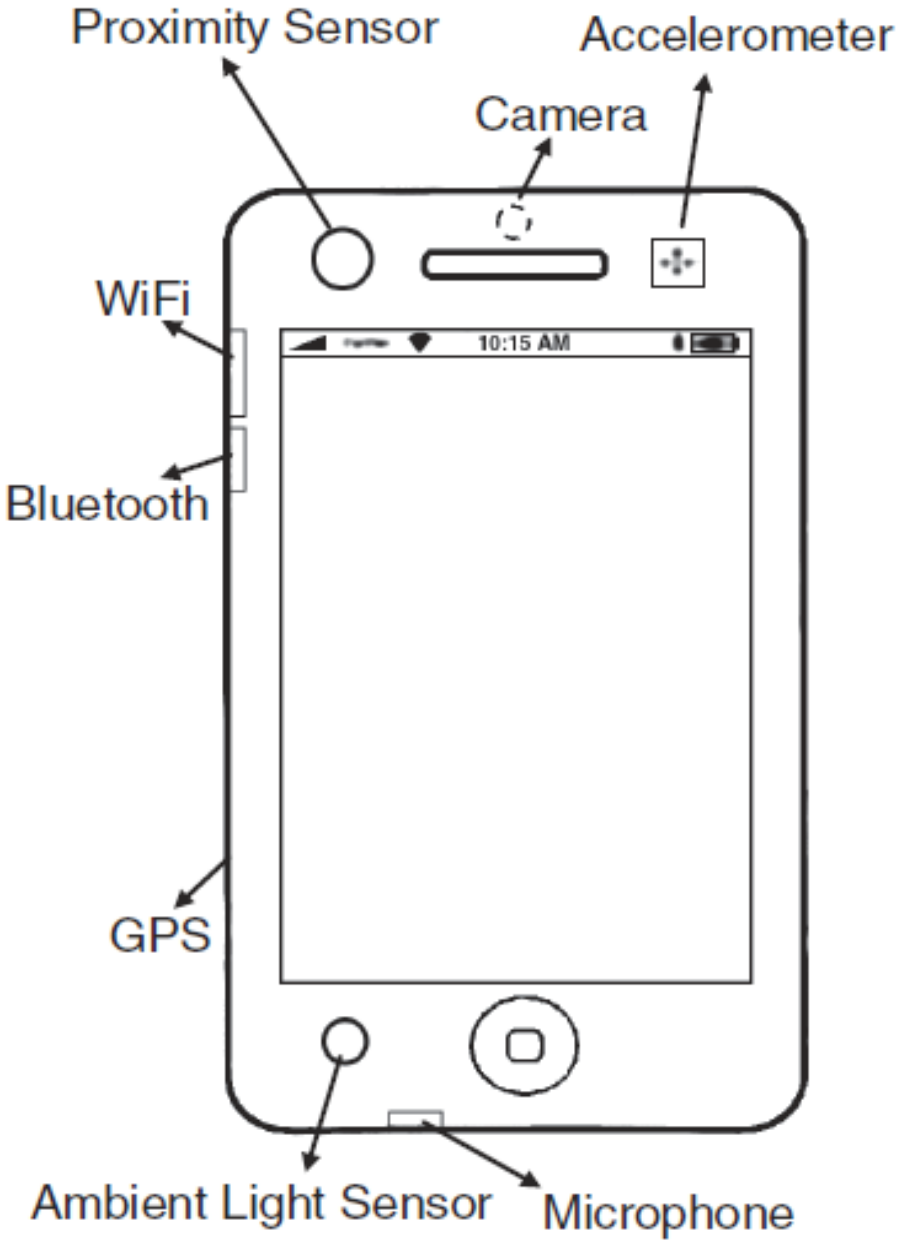
Architecture of the Internet of Things



Functional Modules of a Smart Sensor Node

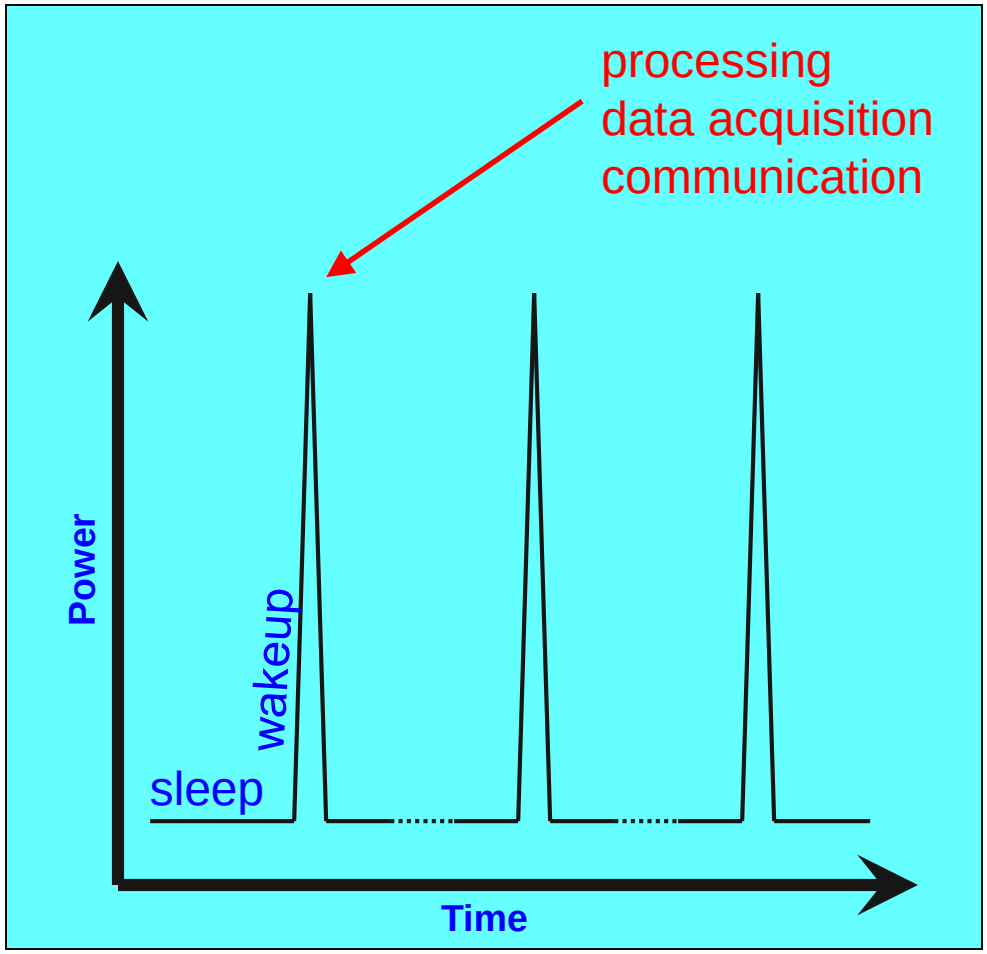


Sensor devices built inside a typical smartphone



Typical WSN Application Pattern

- Periodic
  - Data collection
  - Network maintenance
  - Majority of operation
- Triggered events
  - Detection/notification
  - Seldom occurs
    - But must be reported quickly and reliably
- Long lifetime
  - Months to years without changing batteries
  - Power management is the key to WSN success



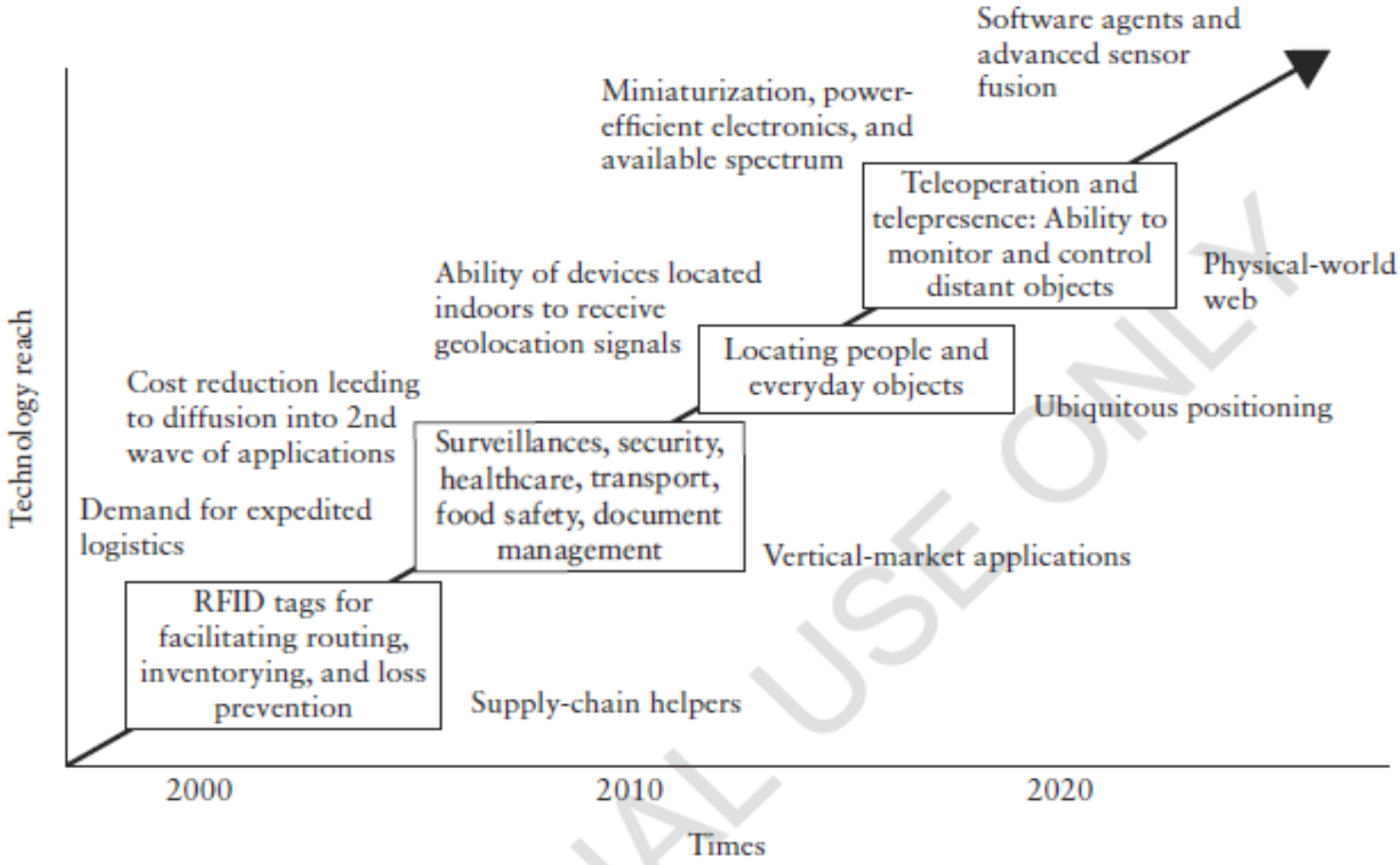


Figure 2.5
Technology road map of IoT applications. Courtesy of SRI: Consulting Business Intelligence, "Disruptive Technologies Global Trends 2025."

PERSONAL USE ONLY

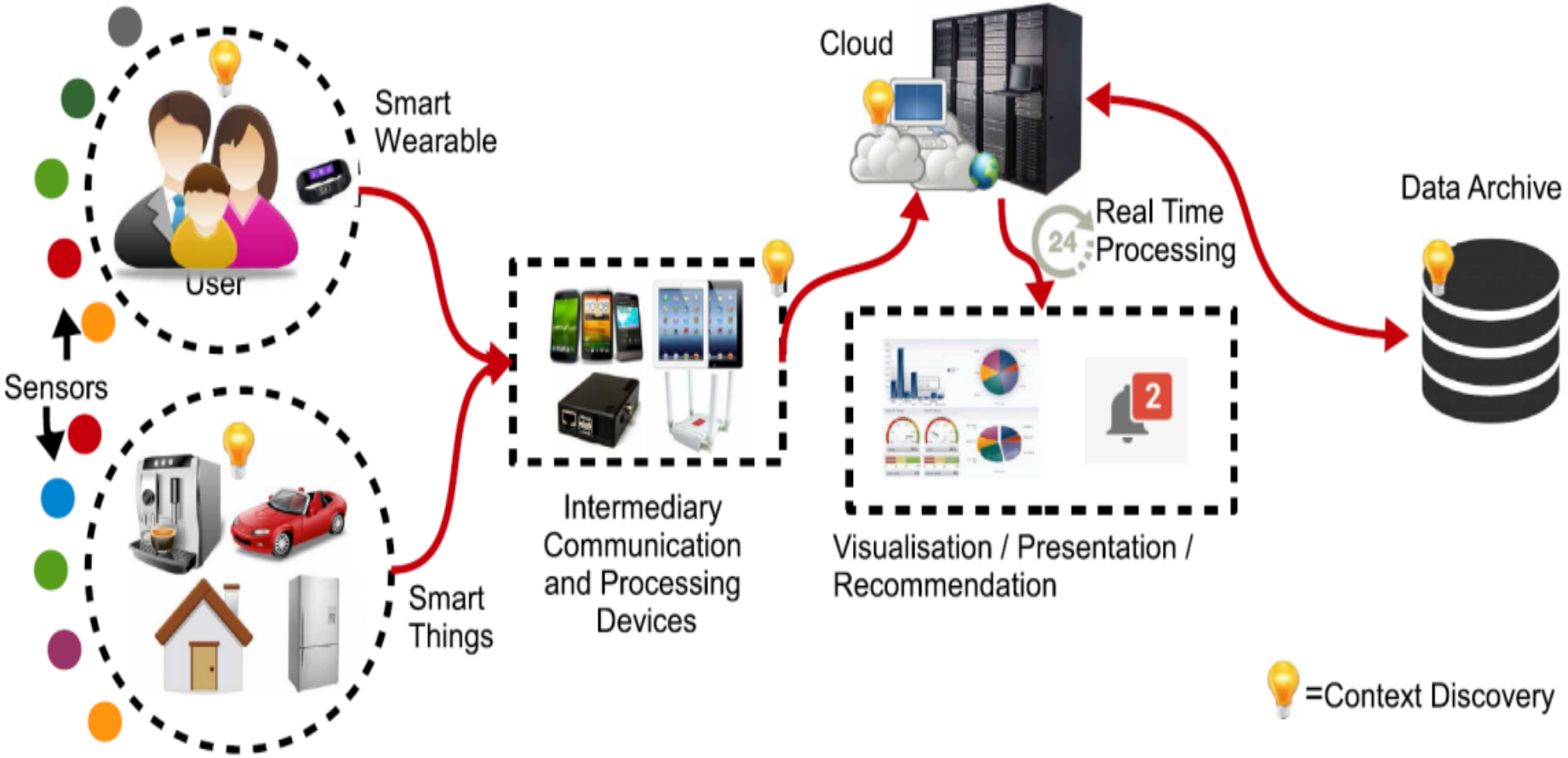
Table 4.7: IoT Wireless Communications and Networks

Network Types	Wireless WAN	WMAN	WLAN	WPAN	
Market Name Standard	GSM/GPRS CDMA/1XRTT	WiMaX 802.15.6	Wi-Fi 802.11g	ZigBee 802.15.4	Bluetooth 802.15.1
Application focus	Wide Area Voice and Data	Data, Bandwidth for WiFi AP	Web, Email, Video	Monitoring & Control	Cable Replacement
System Resources	16 MB+	8 MB+	1 MB+	4 KB – 32 KB	250 KB+
Battery Life (days)	1-7	1-7	0.5-5	100-1000+	1-7
Network Size	1	1	32	Unlimited (2^{64})	7
Bandwidth (KBs)	64-128+	75,000	54,000+	20-250	720
Range (meters)	1000KM +	100KM (LoS) 40KM (NLoS)	1 - 100	1 – 100 +	1 – 10 +
Success Metrics	Coverage, Quality	Speed, Coverage	Speed, Flexibility	Reliability, Power, Cost	Cost, Convenience

Key Applications of IoT

Domain	Brief description	Examples
Industry, commerce, and business	Activities involving financial or commercial transactions between companies and organizations	Manufacturing, logistics, service sector (服務業), banking, financial governmental authorities, retail, supply chains, dealers, intermediaries
Environment energy, and resources	Activities regarding the protection, monitoring and development of all natural resources	Agriculture and breeding, recycling, environmental management services, energy management, smart grid projects
Smart home, city, and earth development	Utilities, appliances, health care, community services, entertainments, tourist, weather services	Bill Gates smart home, smart cities projects in many countries, IBM smart Earth project, Google map services
Transport, and public services	Public transportation, traffic and crisis management, logistics, parking, emergency and disaster recovery	Intelligent highways, national parks, election, crowd sources, police services, public safety, crime prevention
Community services, national defense	Government activities/initiatives, infrastructure, military defense, battle field management	Governmental services to citizens and infrastructures, e-inclusion (seniors, disabled people)

Location-sensitive Applications of Typical IoT Processing Stages



IoT Apps Domains and Users

Fig. 1. Internet of Things schematic showing the end users and application areas based on data.

“Green” Cloud Computing

The Power of Evolution of VLSD

1990 Generation 1
10K Servers

1998 Generation 2
100K Servers

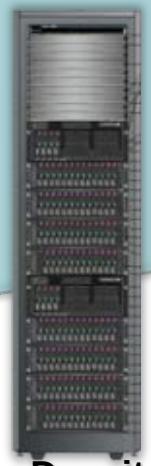
2008 Generation 3
300K Servers

~ 500 KWatts

~ 10 MegaWatts

~ 60 MegaWatts → ??

Server *Capacity*



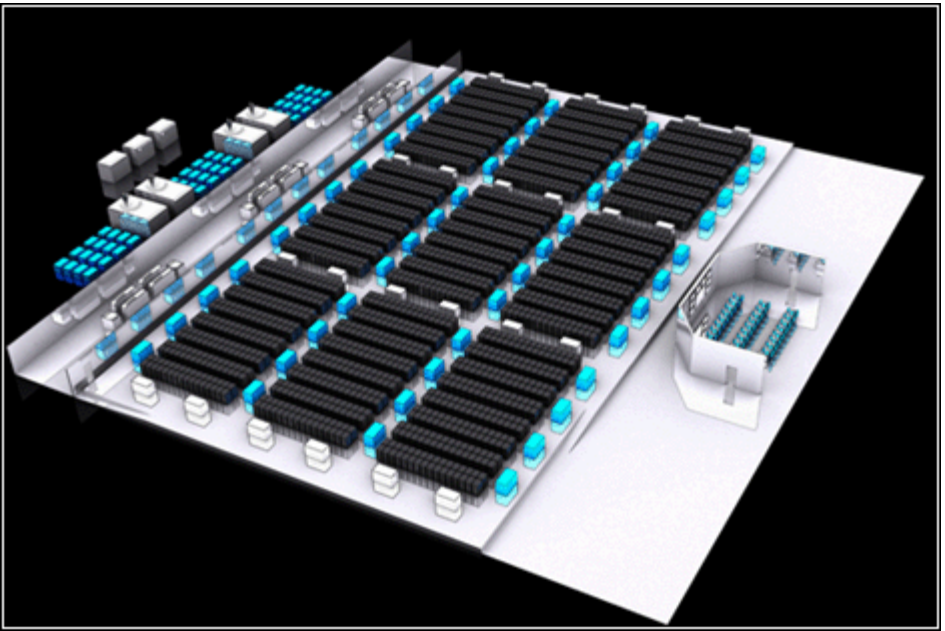
Server Density and Manageability

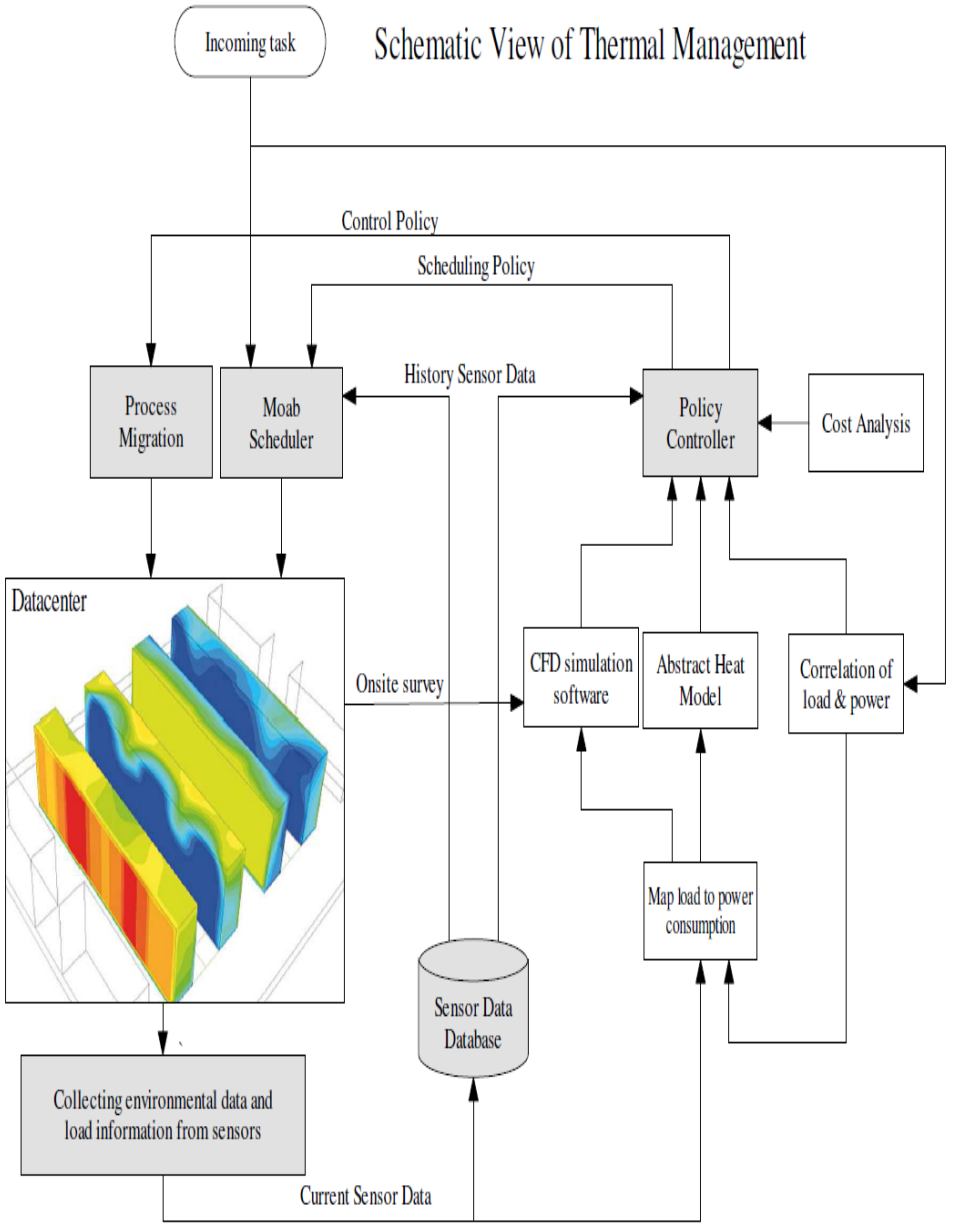
Container Scalability

Power Usage Efficiency (PUE)

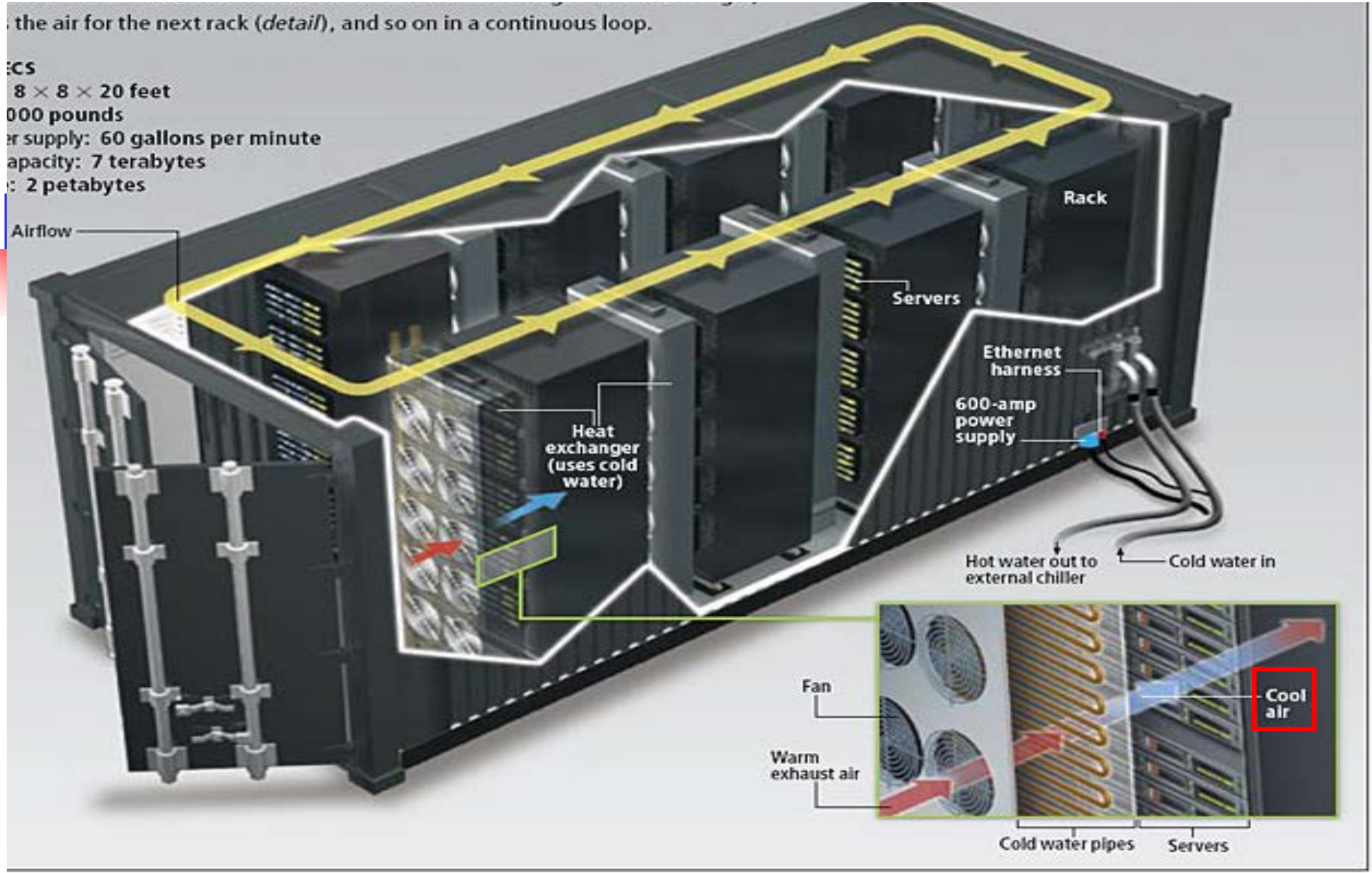
VLSD: Very Large Scale Datacenter

Thermal Management





Highly efficient container computer based VLSD

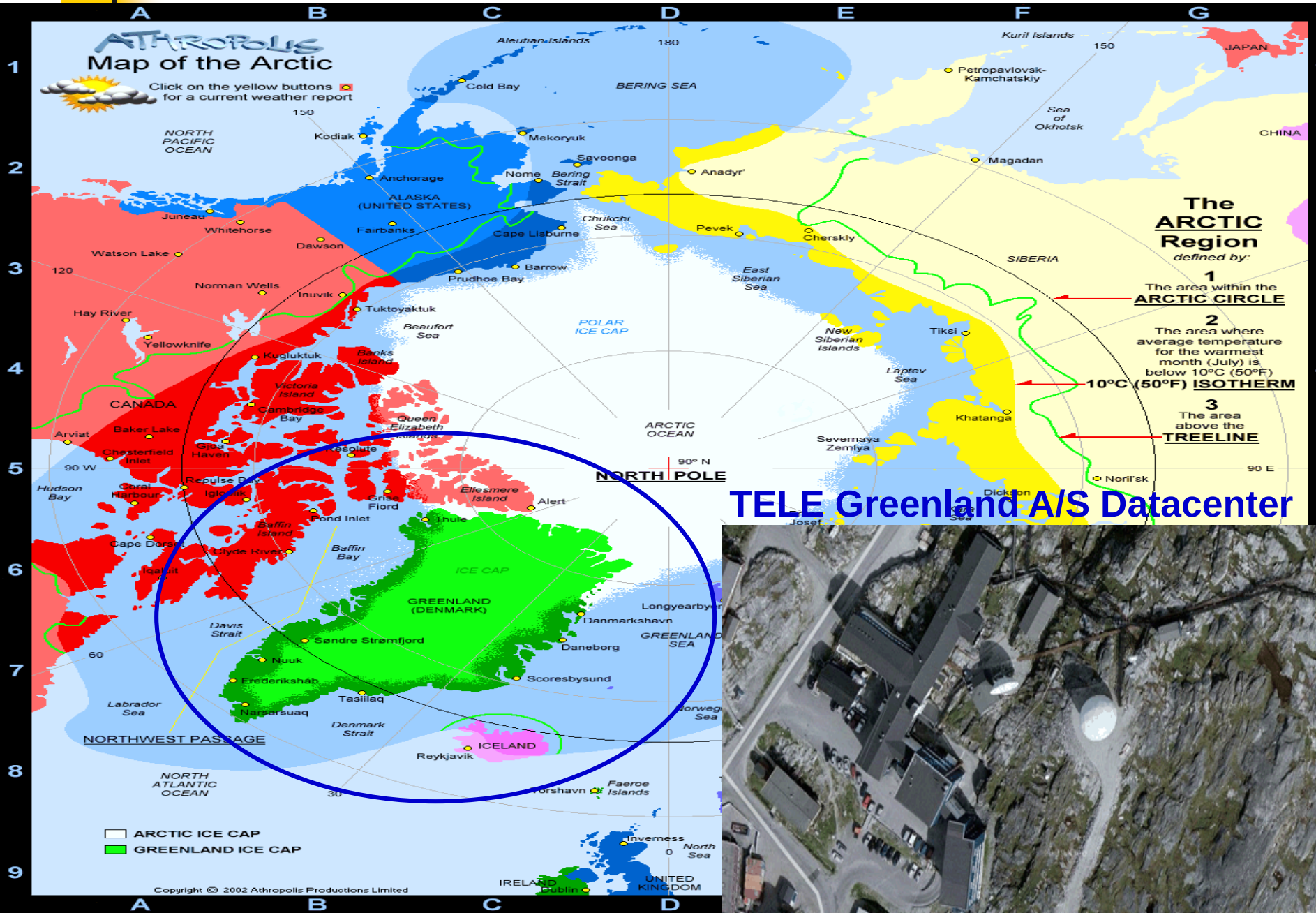


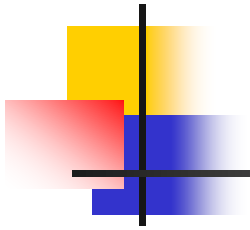
Cooling is a BIG problem in VLSD

Today's VLSD needs a lot of Power + Water

Google Datacenter at Columbia river, Oregon

Greenland (格陵蘭) Datacenter

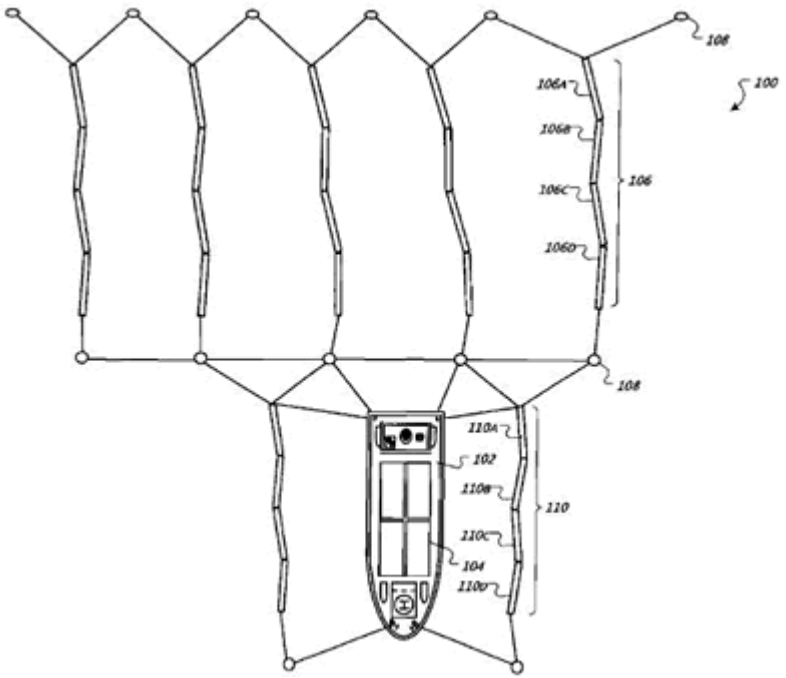




Interesting ideas for building “Green” VLSD

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.

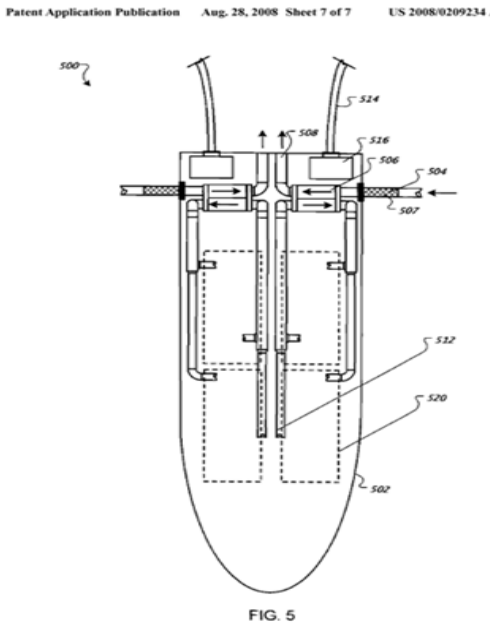


FIG. 5

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