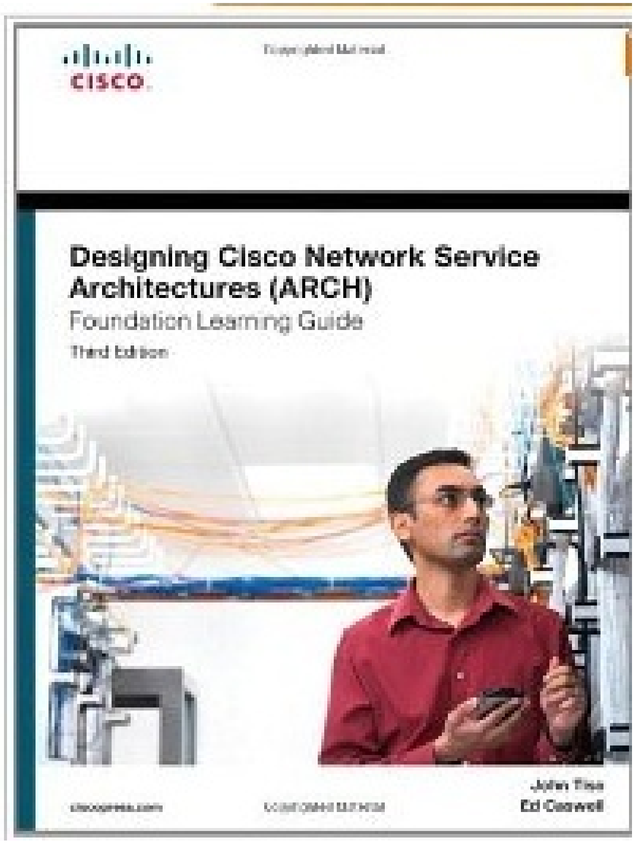
Chapter 0

Cloud Computing and Networking

Books

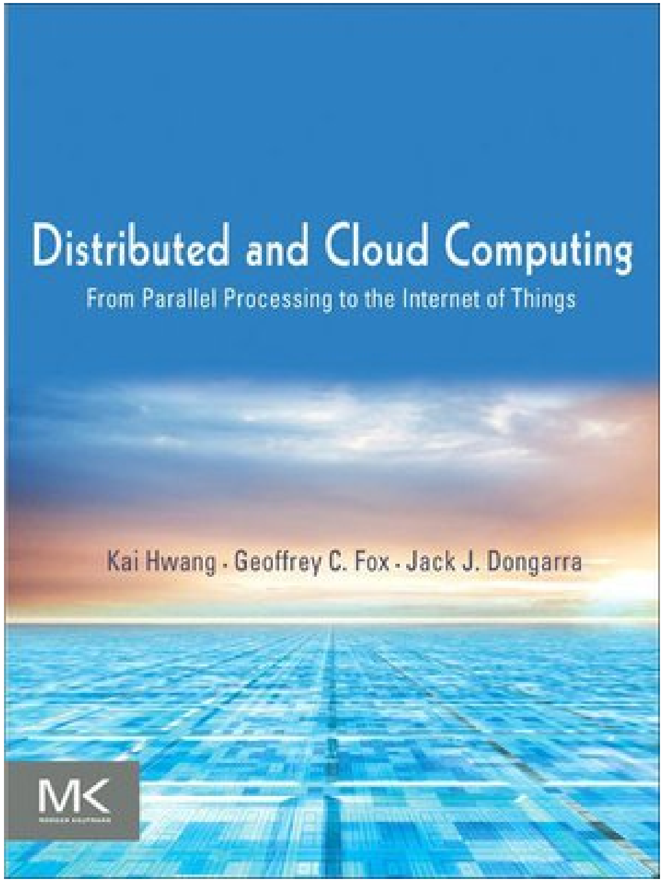
- Designing Cisco Network Service Architectures Foundation Learning Guide (3rd Edition) (Foundation Learning Guides)



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)14:00~16:00
- Grade:
  - Homework 30%
  - Paper Presentation 30%
  - Midterm Exam 20%
  - Final Exam 20%

Course Goal

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

Cloud applications create huge data to use cloud computing

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)

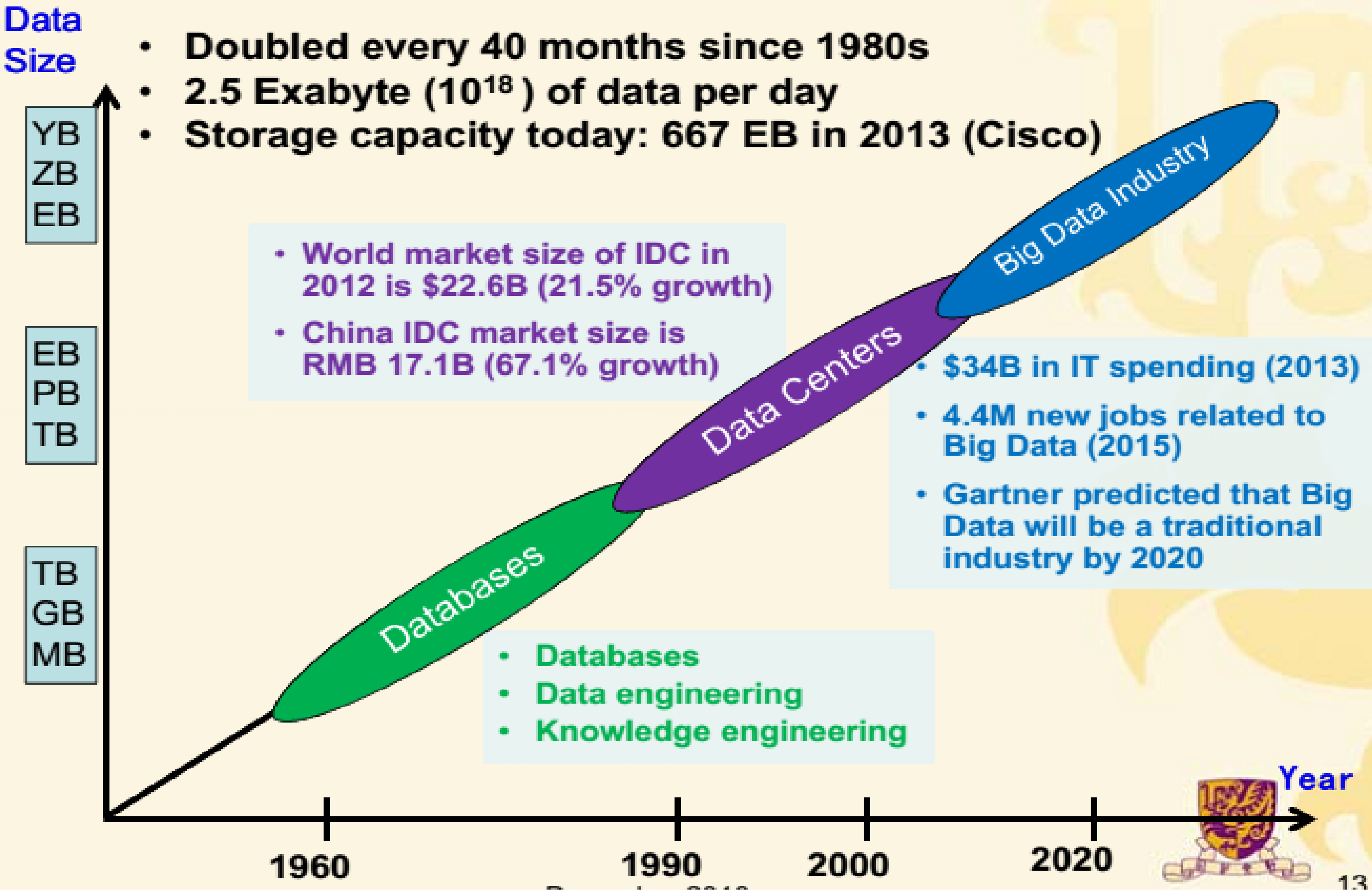
•1 Terabyte (TB) = 1024 GB
•1 Petabyte (PB) = 1024 TB
•1 Exabyte (EB) = 1024 PB
•1 Zettabyte (ZB) = 1024 EB
•1 Yottabyte (YB) = 1024 ZB

Hugh Data

- Huge multicast traffic in DCN

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

Evolution of the Big Data Industry



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

Wolfgang Amadeus Mozart (1756 - 1791)

Einstein was born in 1879

Birthday-of (Mozart, 1756)

Birthday-of (Einstein, 1879)

PERSON (DATE – XXX)

PERSON describe ...

(Brill et al., TREC 2001; Lin, ACM TOIS 2007)

(Agichtein and Gravano, DL 2000; Ravichandran and Hovy, ACL 2002; ...)

Different Use Habits

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, resource utilization
  - Management --> virtualization, cooling system

Big Data and Cloud Computing

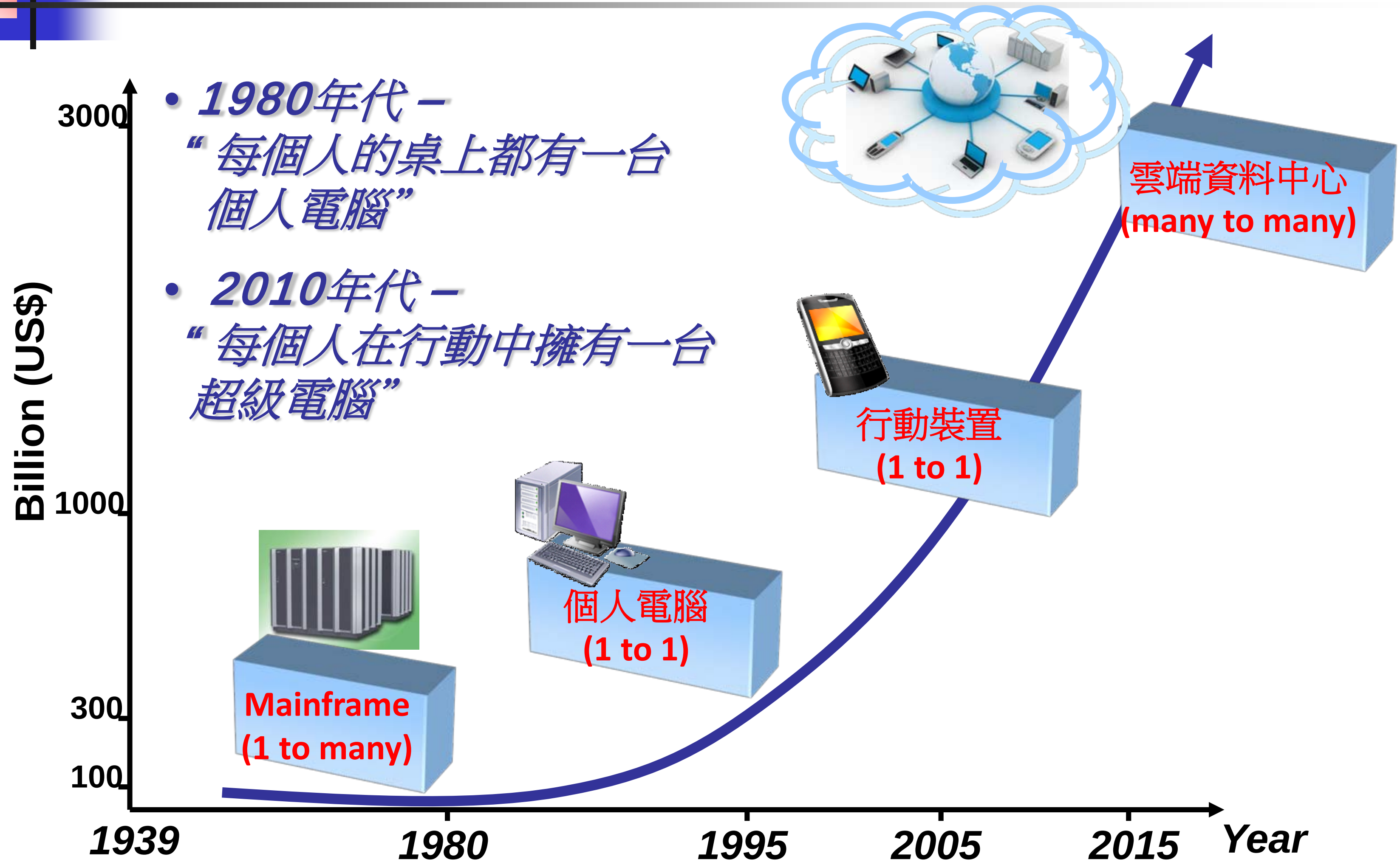
- It starts with the premise that the data services and architecture should be on servers.
  - We call it cloud computing – they should be in a "cloud" somewhere.

- 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 BlackBerry or a mobile phone 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

- ***1980年代 –***
  “每個人的桌上都有一台個人電腦”
- ***2010年代 –***
  “每個人在行動中擁有一台超級電腦”



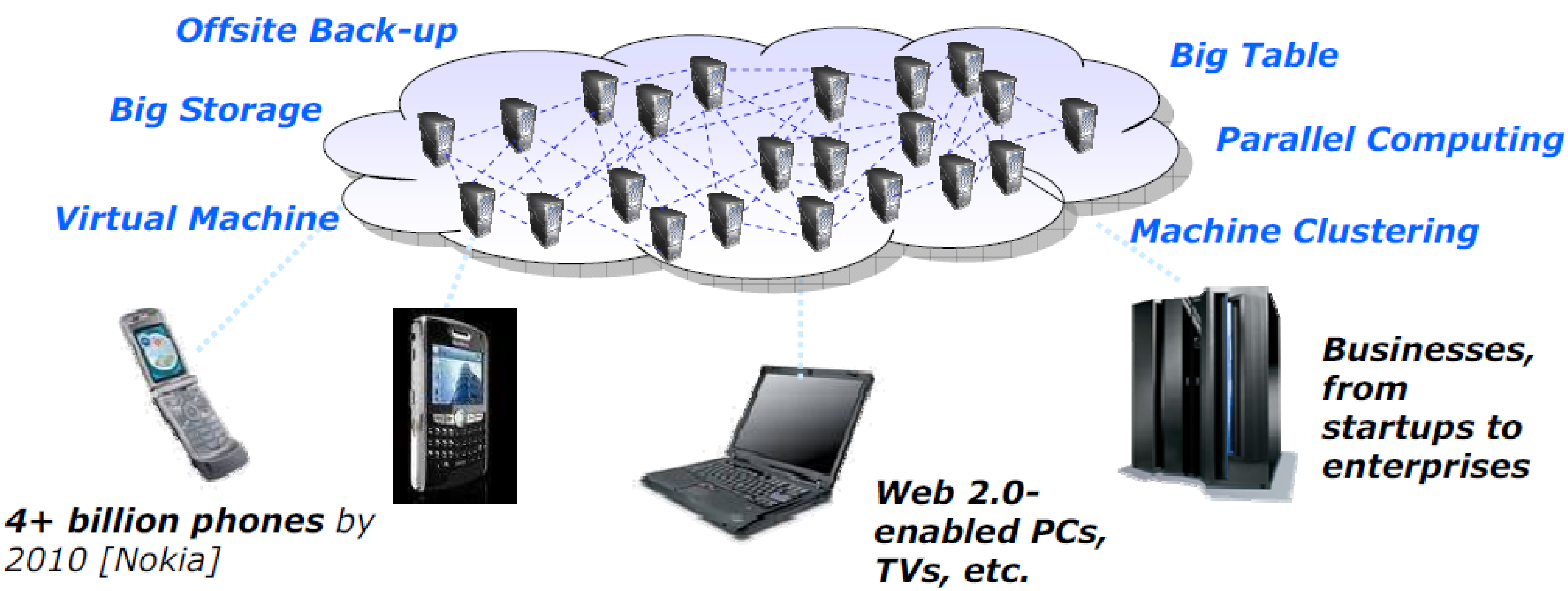
「雲端資料中心」讓電腦運算就像是水、電 一樣，只要連上網路就可以**pay-as-you-go**無限量提供服務。

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*

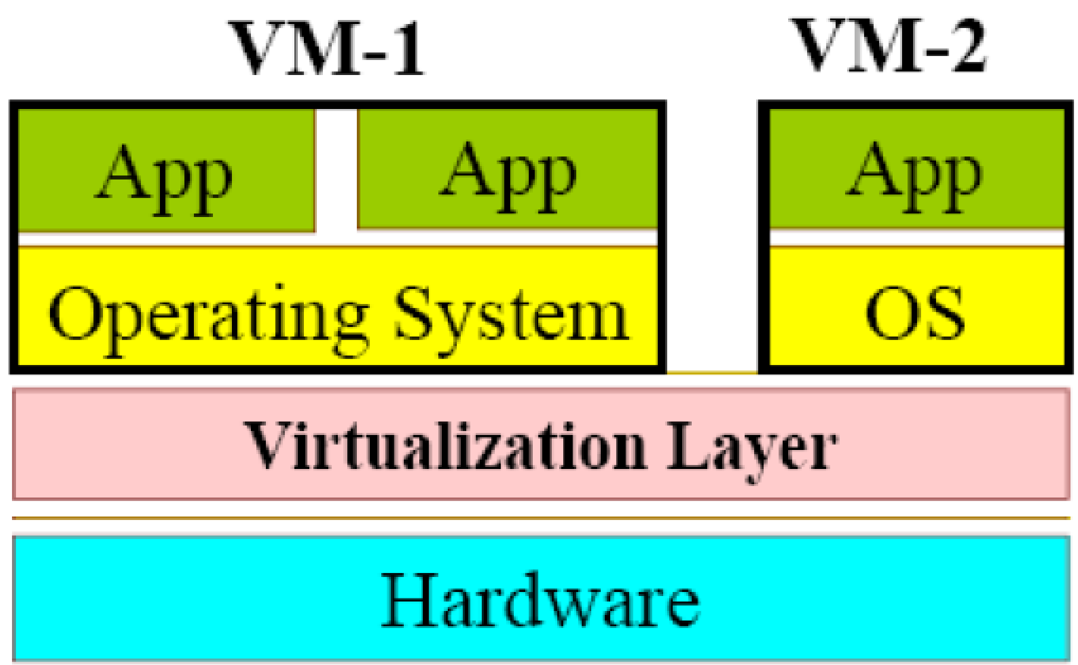
(2) *使用者可以利用各種具備網際網路連線能力的電腦終端裝置* **(device)**，*無所不在***(ubiquitous)** *的使用資料與服務*



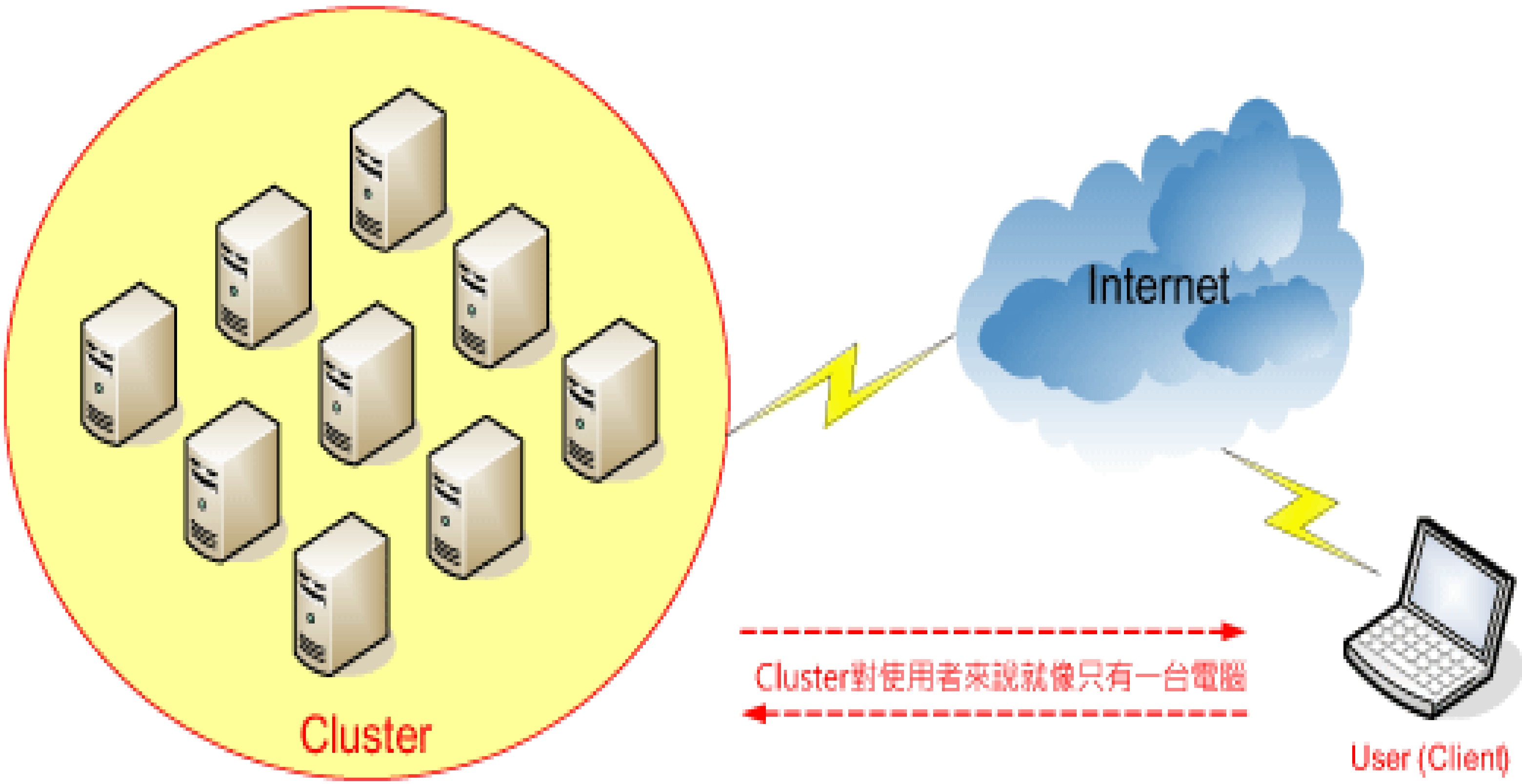
虛擬運算技術

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

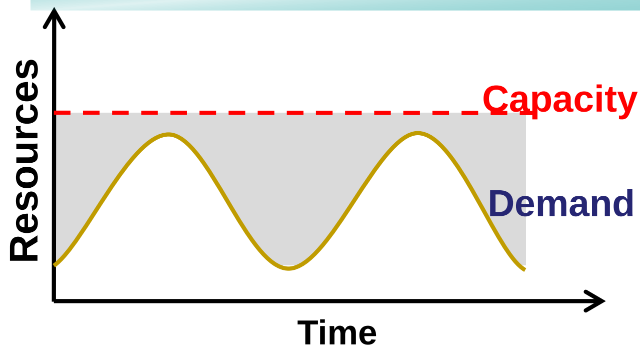


雲端運算經濟學

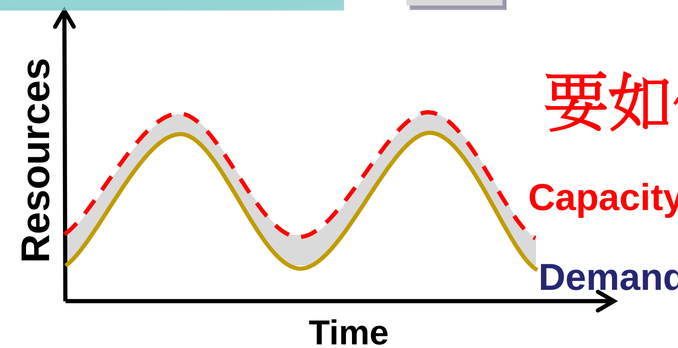
Pay by use instead of provisioning for peak

Unused resources

要如何做好資源管理

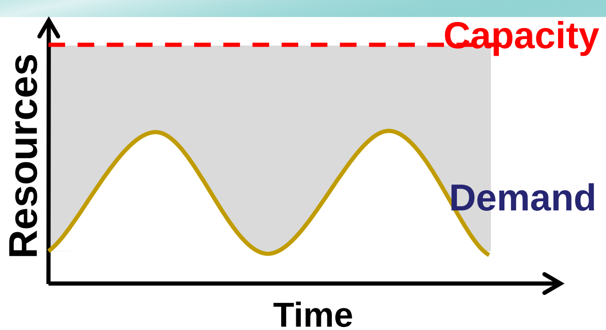


Static data center



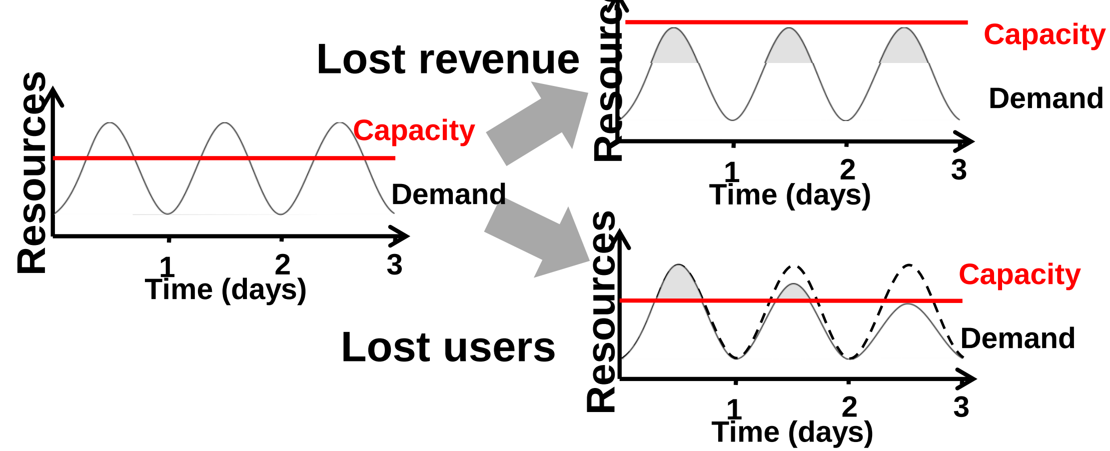
Data center in the cloud

Risk of over-provisioning: underutilization



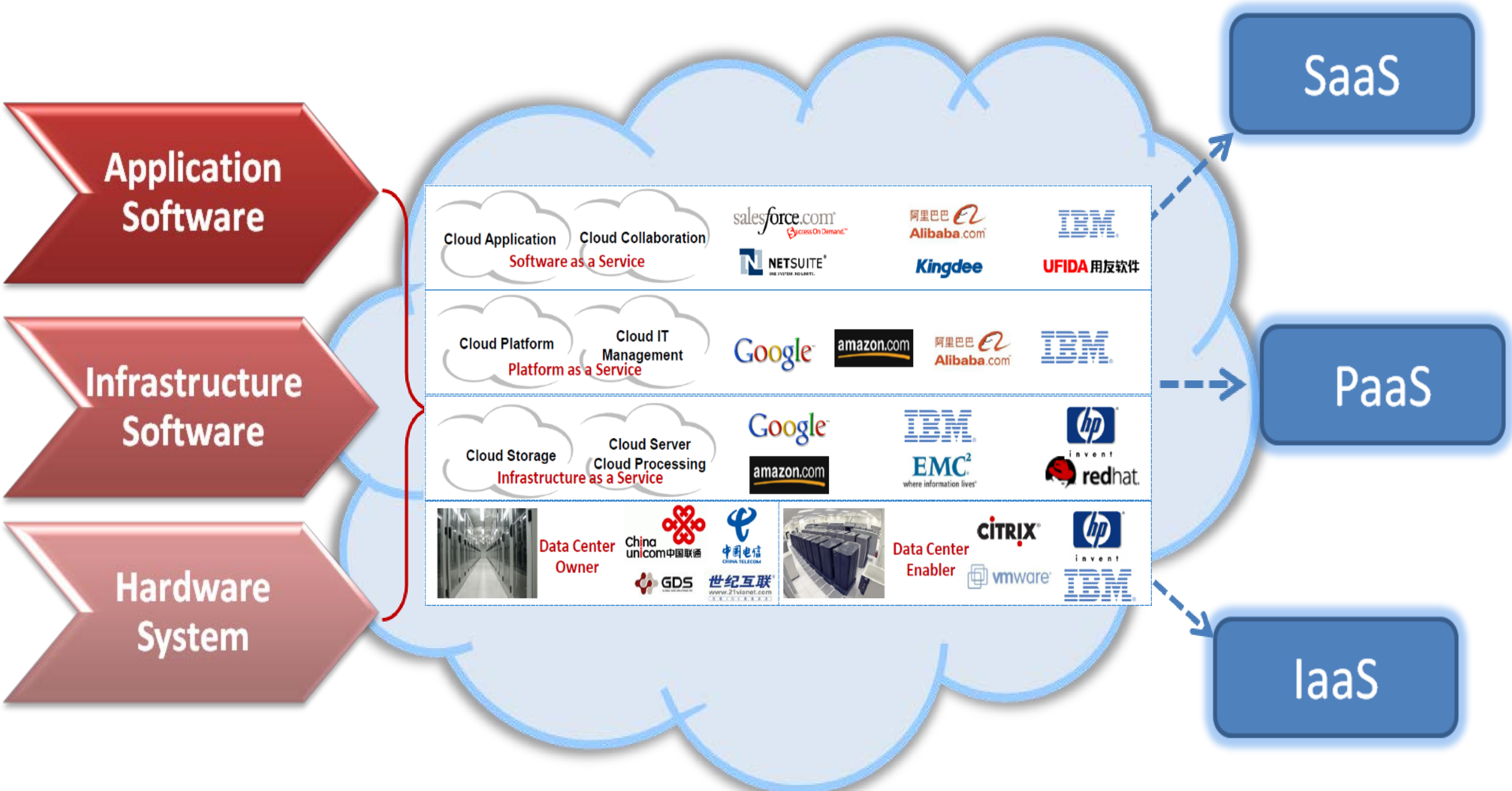
Static data center

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

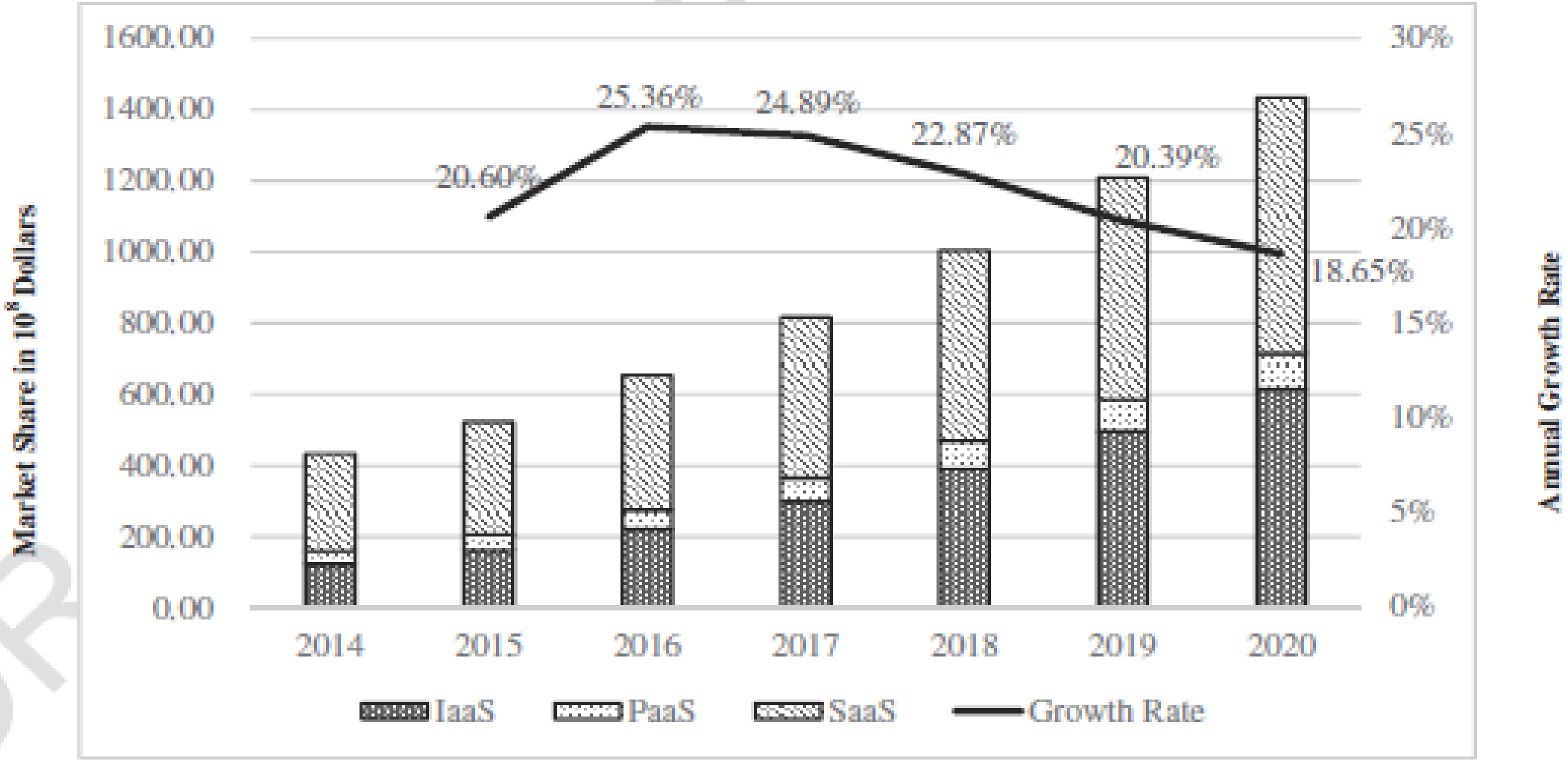


Figure 1.13
Worldwide distribution of cloud service models and the growth rate based on projections by Gartner Research from 2014–2020.

台灣雲端上下游產業價值鏈

Service Operation

平台服務營運
（中華電、台灣固網、遠傳…）

Infrastructure Software

應用服務（電子郵件，資料備份，醫療電子病歷 …）

Cloud OS(ITRI/III、Microsoft、VMware…)

Cloud Hardware System

伺服器
（英業達, 神達, 鴻海, 緯創, 廣達…）

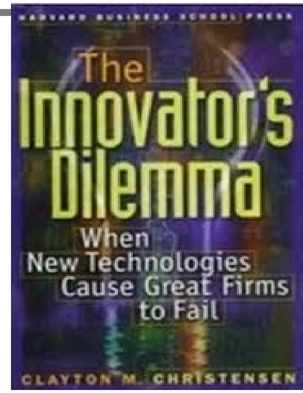
處理器
（威盛）

儲存體
（普安、喬鼎、宜鼎、世仰、信億 ...）

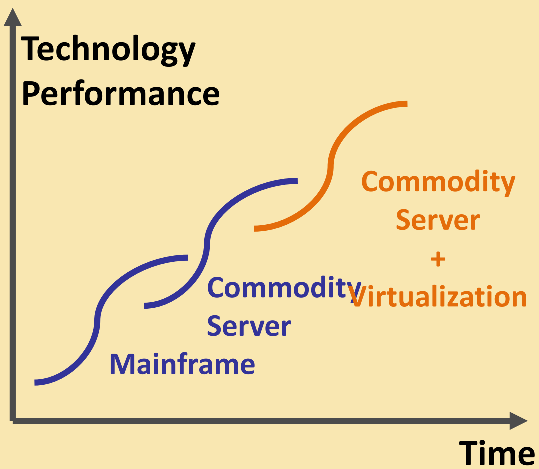
網通設備
（智邦、友訊、合勤、建漢、正文...）

電源供應及冷卻系統
(台達電、康舒...)

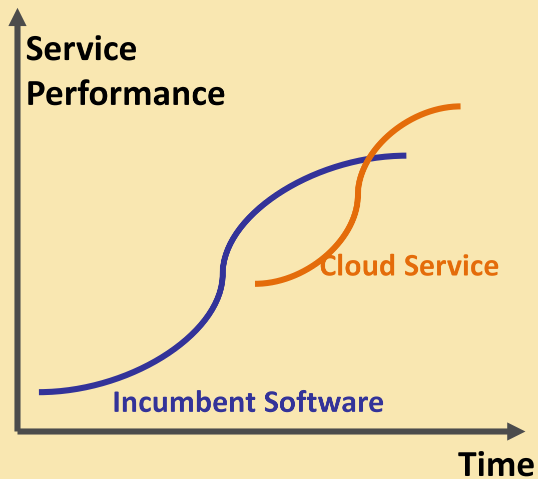
雲端運算帶來破壞式創新



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

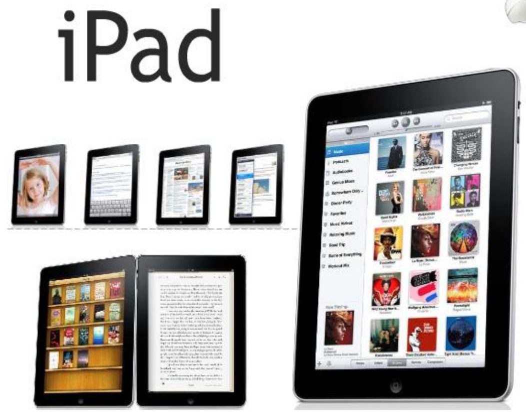
Source: Daniel Costello, Microsoft Data Center Transformation, Microsoft WinHEC 2008

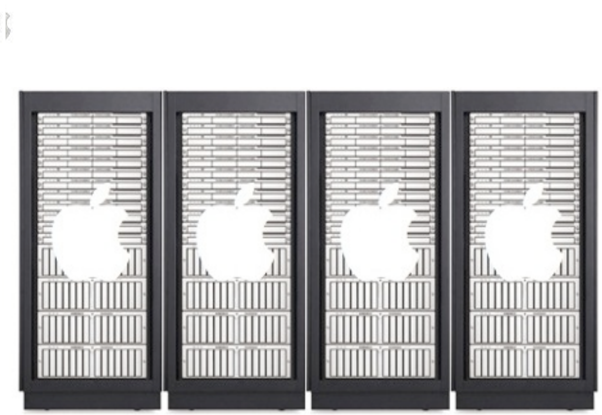
Why Container Computer?

- 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

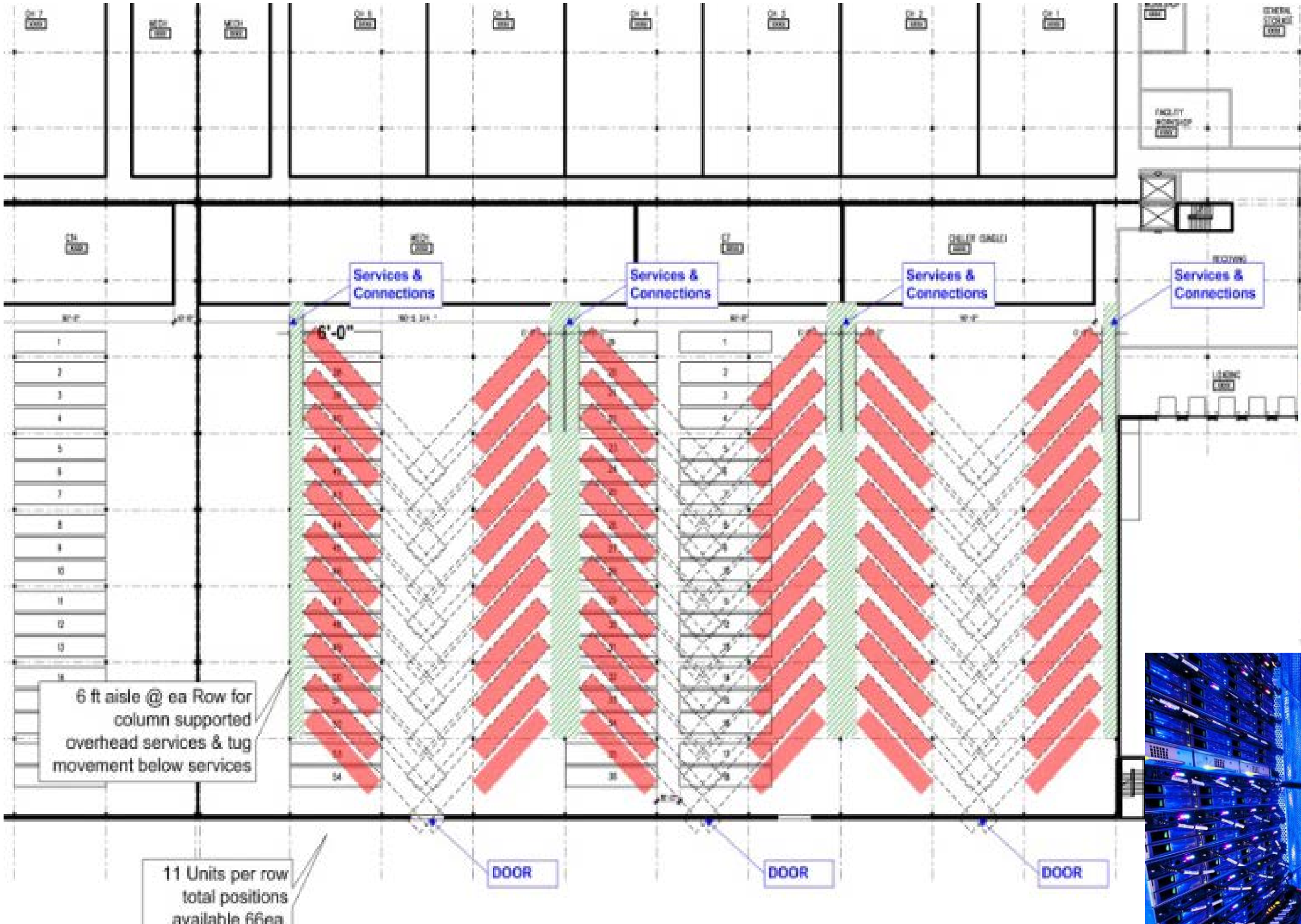
1.Apple's existing Newark, CA., Data Center is around 109,000 square feet--the new one is over 500,000. That represents either a ridiculously big scaling-up of business or a whole new thing 2.500,000 square feet is among the largest centers being built in the World on a single site. Microsoft's new one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

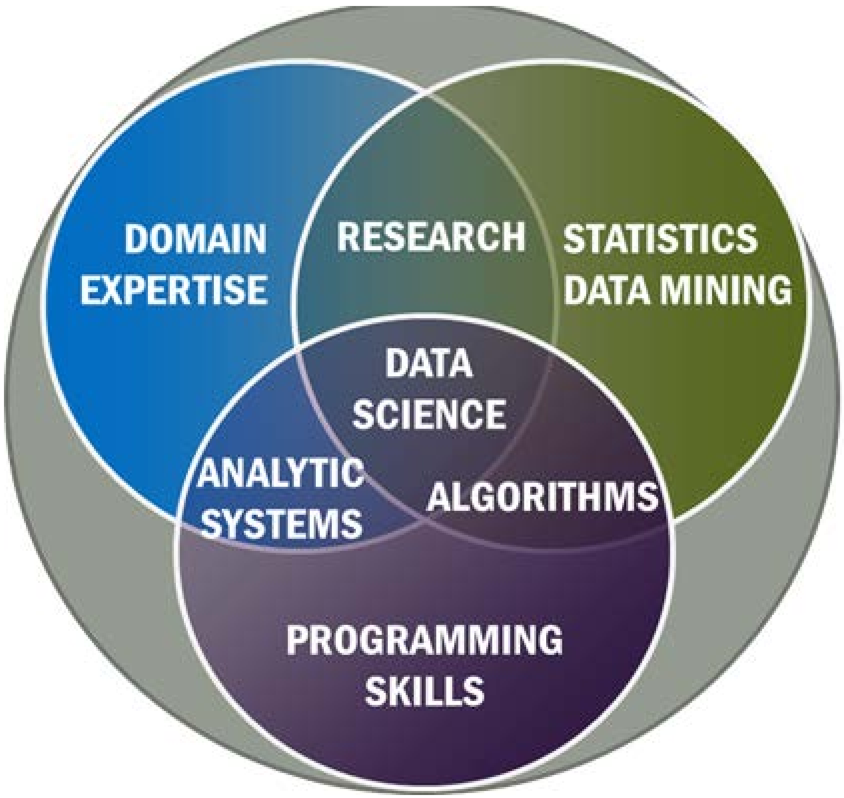
Big Data

What is Data Science?

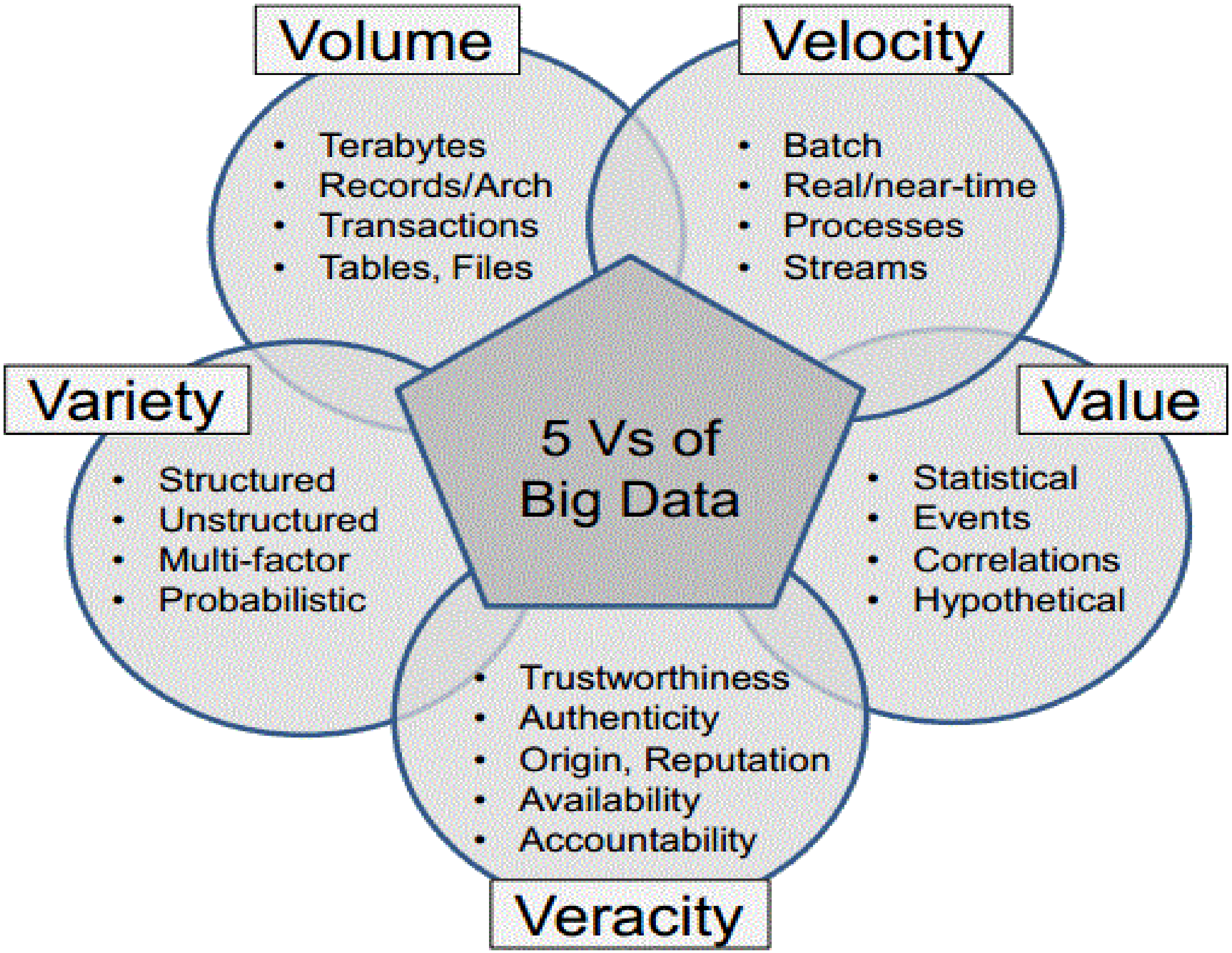
- **Data science** is the extraction of useful knowledge directly from data through a process of discovery, hypothesis, and analytical hypothesis analysis

- 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

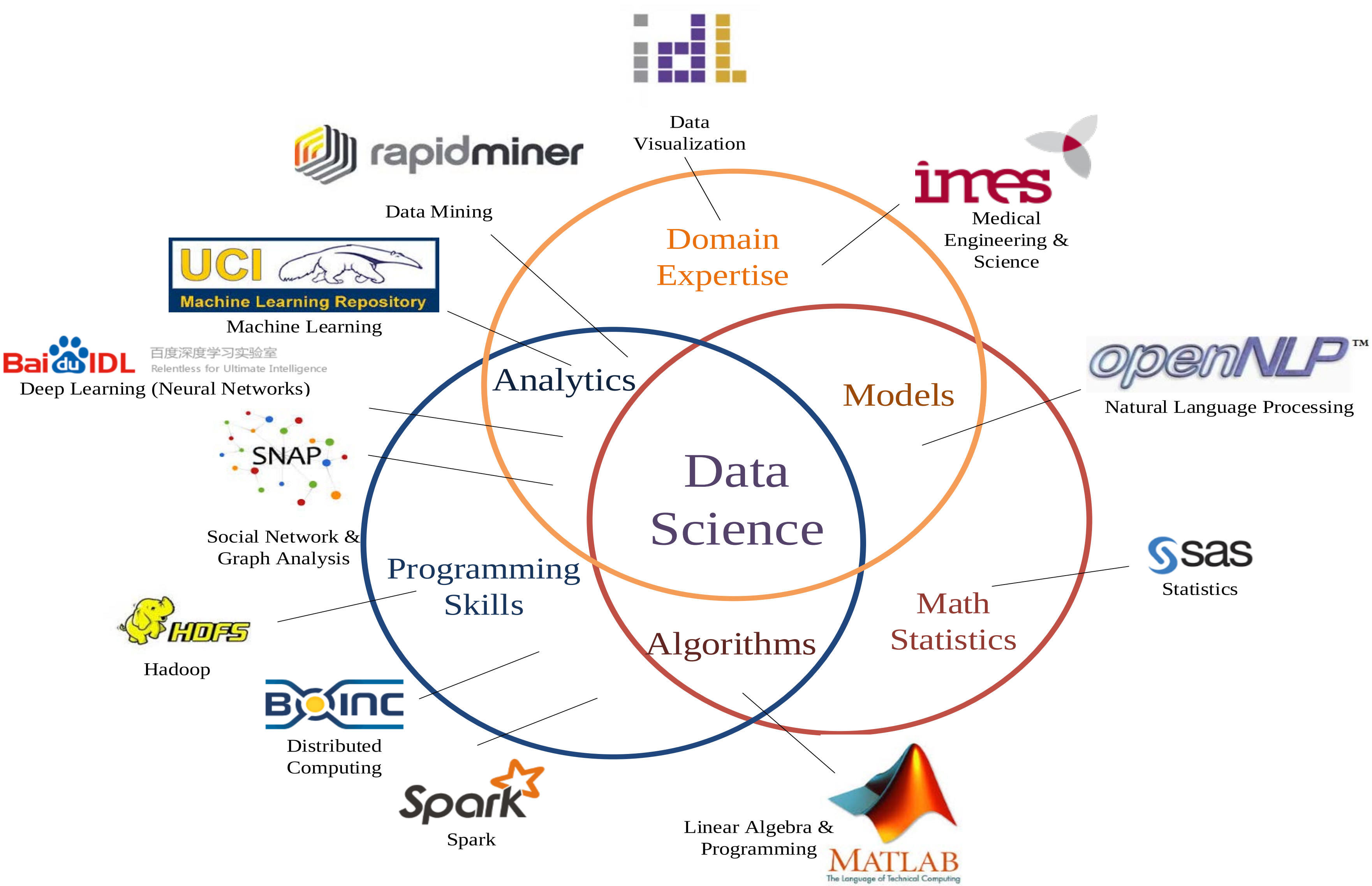
Big data refers to digital data volume, velocity and/or variety whose management requires scalability across coupled horizontal resources



The Five v's of Big Data



Today's Dig Data Software Libraries



Less than 1 MB (12)	3.7%
1.1 to 10 MB (8)	2.5%
11 to 100 MB (14)	4.3%
101 MB to 1 GB (50)	15.5%
1.1 to 10 GB (59)	18%
11 to 100 GB (52)	16%
101 GB to 1 TB (59)	18%
1.1 to 10 TB (39)	12%
11 to 100 TB (15)	4.7%
101 TB to 1 PB (6)	1.9%
1.1 to 10 PB (2)	0.6%
11 to 100 PB (0)	0%
Over 100 PB (6)	1.9%

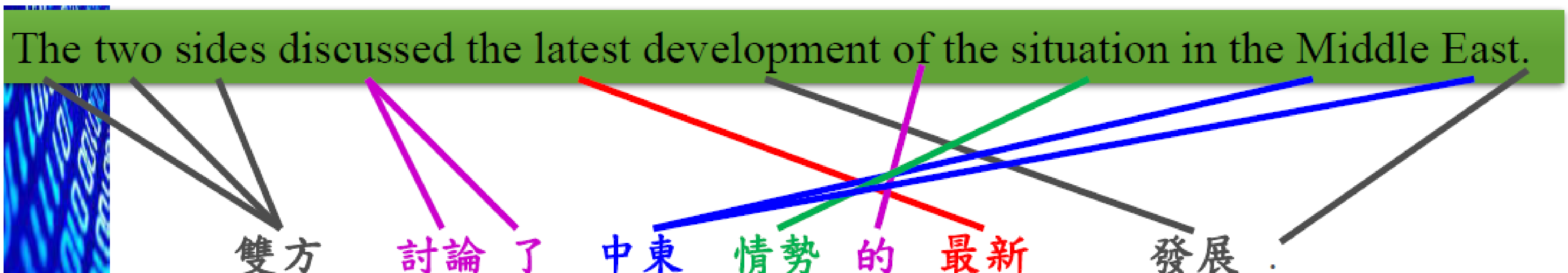
Figure 1.1 Results from a KDnuggets Poll about Largest Data Sets Analyzed

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.

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

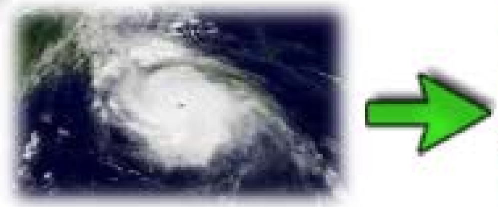


latest	
最新	25%
最遲	**40%**
最晚	30%

latest + development	
最新	**97%**
最遲	2%
最晚	1%

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%。

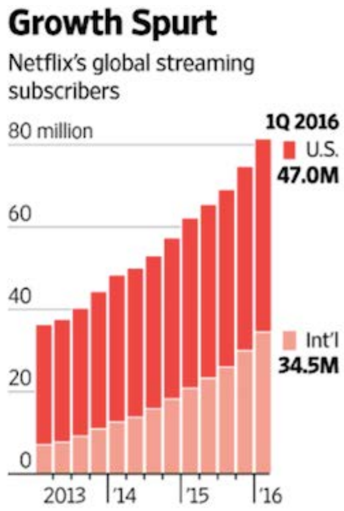
雲端雲算與巨量資料驅動企業創新

- 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》

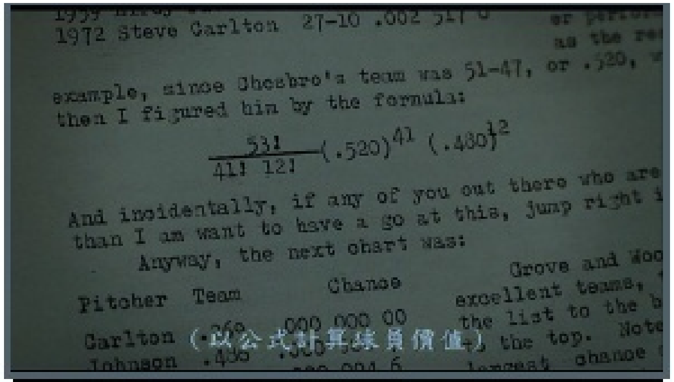
Netflix

- 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 ！



企業巨資分析商業應用的挑戰

- 電信業者利用手機可攜碼資料庫，開發名為M+Messenger的通訊軟體，消費者下載後，手機通訊錄
手機門號是哪家電信公司，原意是讓民眾知
馬，會以網外還是網內計費，可是有民眾主
資料保護法，告上法院，台北地方法院認定
五百元。

Data Sources and Collection

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

The Internet of Things (IoT)

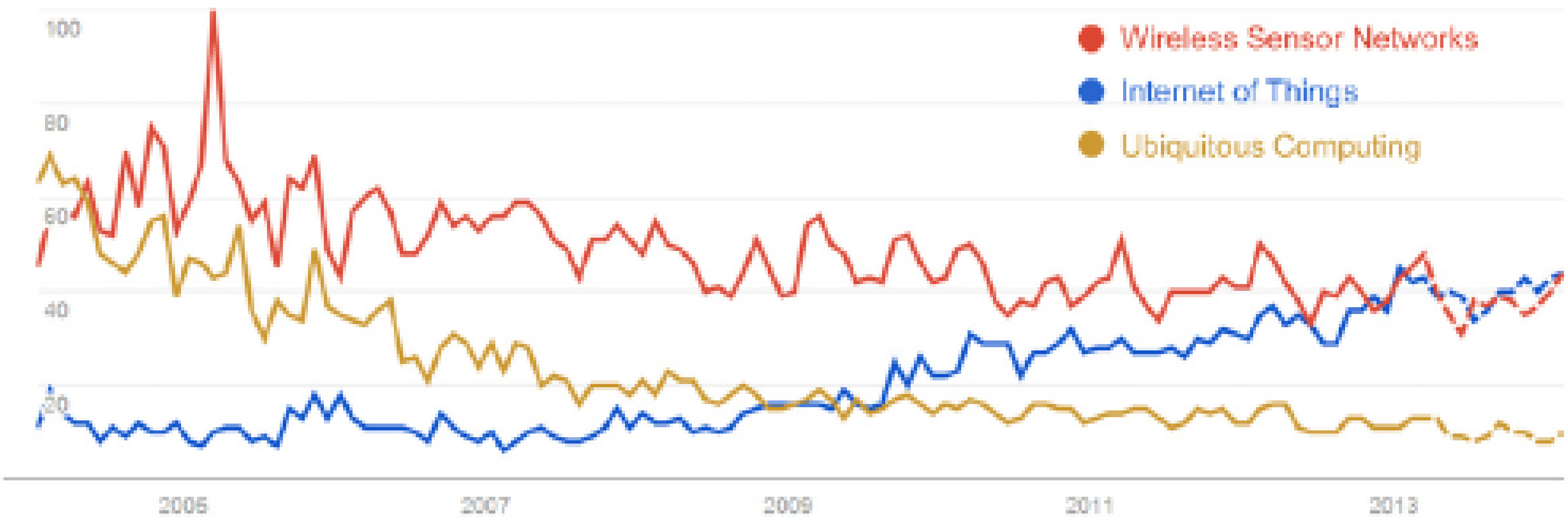
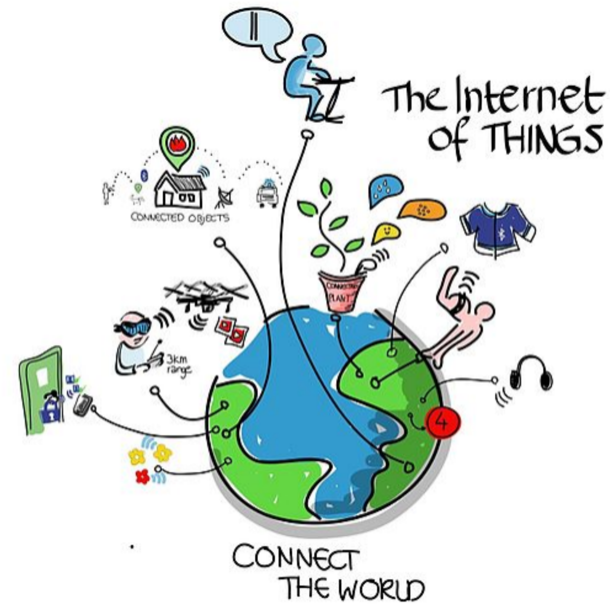


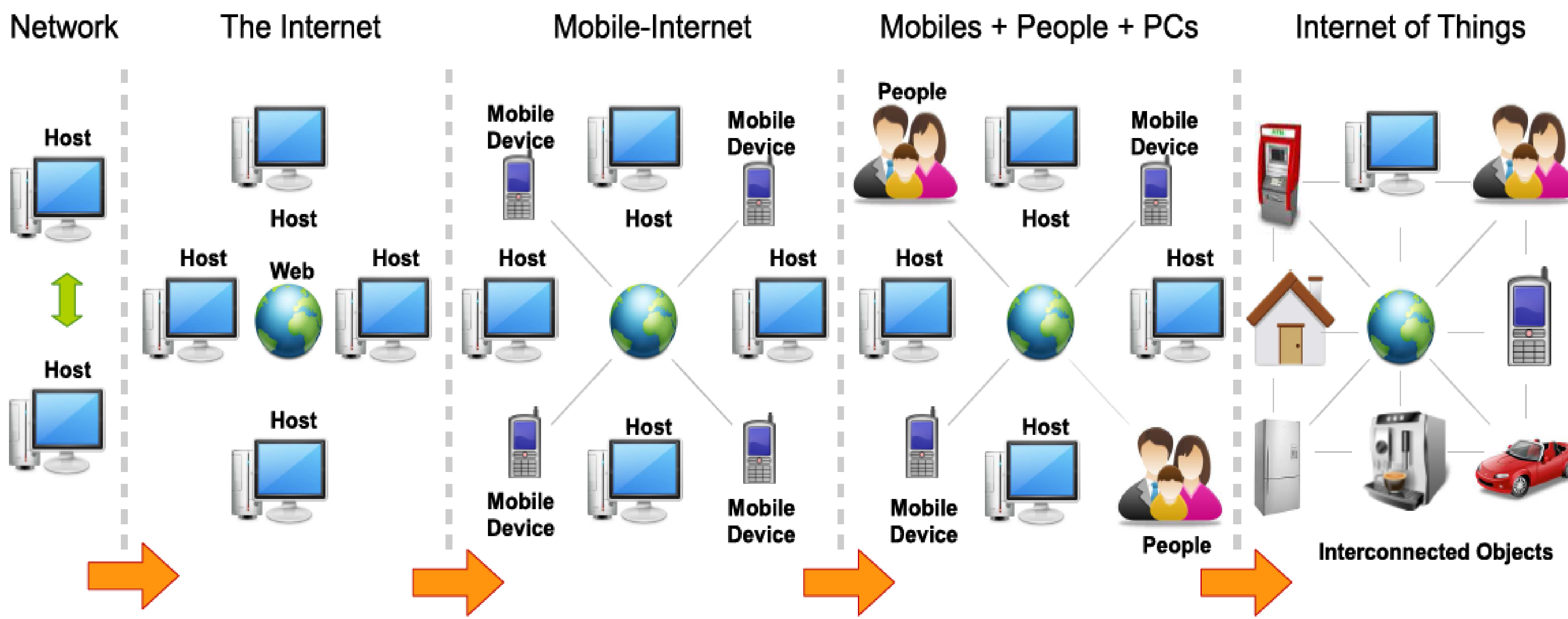
Fig. 3. Google search trends since 2004 for terms Internet of Things, Wireless Sensor Networks, Ubiquitous Computing.

Introduction

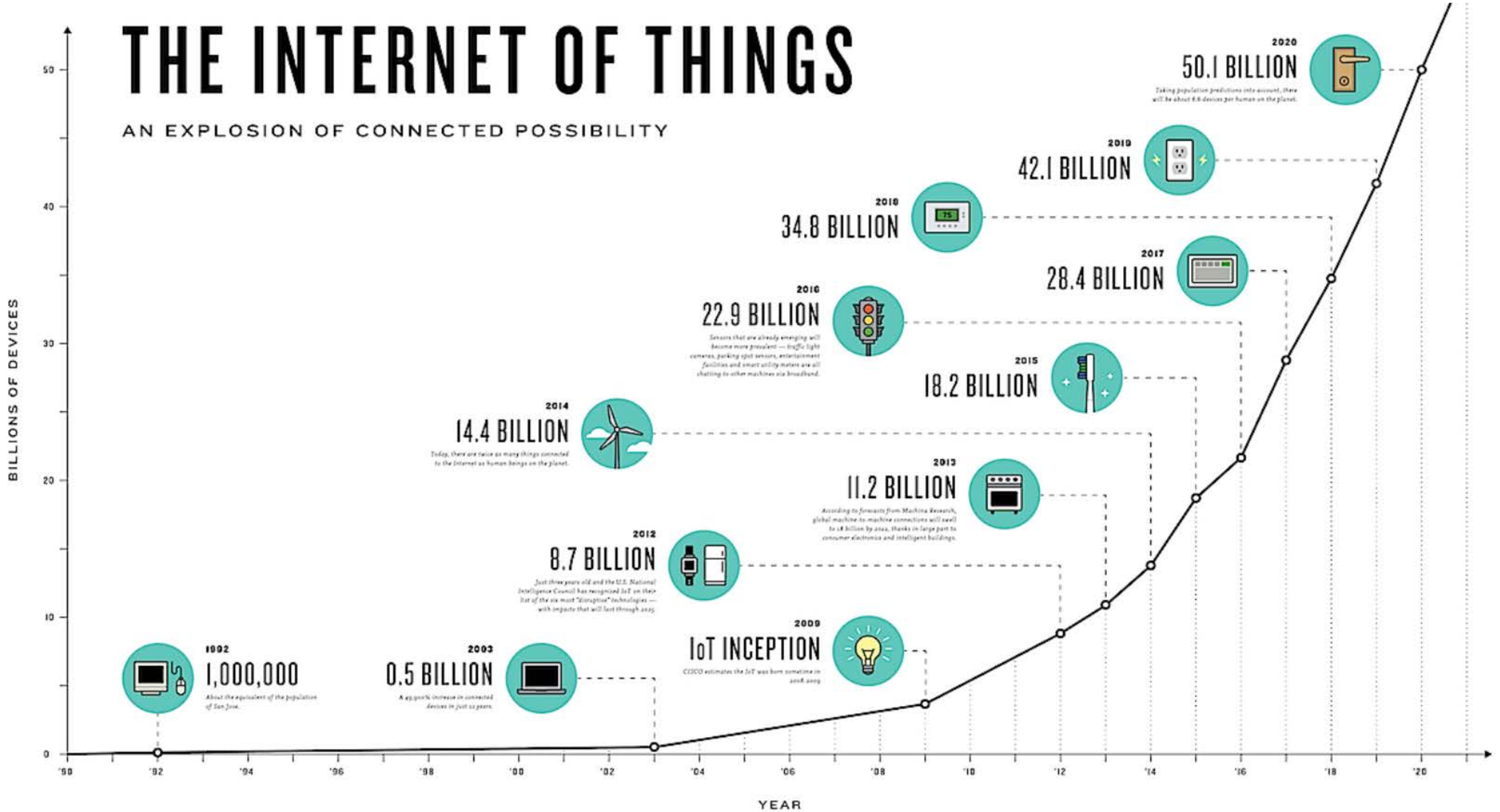


- Internet of Things(IoT)
  - Sensed and controlled remotely
    - Across existing network infrastructure
    - Physical objects in Network
      - devices, vehicles, buildings
    - Collect and exchange data
      - Electronics, software, sensors, and network connectivity
  - Create more direct integration of the physical world
    - Efficiency, accuracy and economic benefit
  - British entrepreneur **Kevin Ashton**
    - Coined the term in 1999 while working at Auto-ID Labs

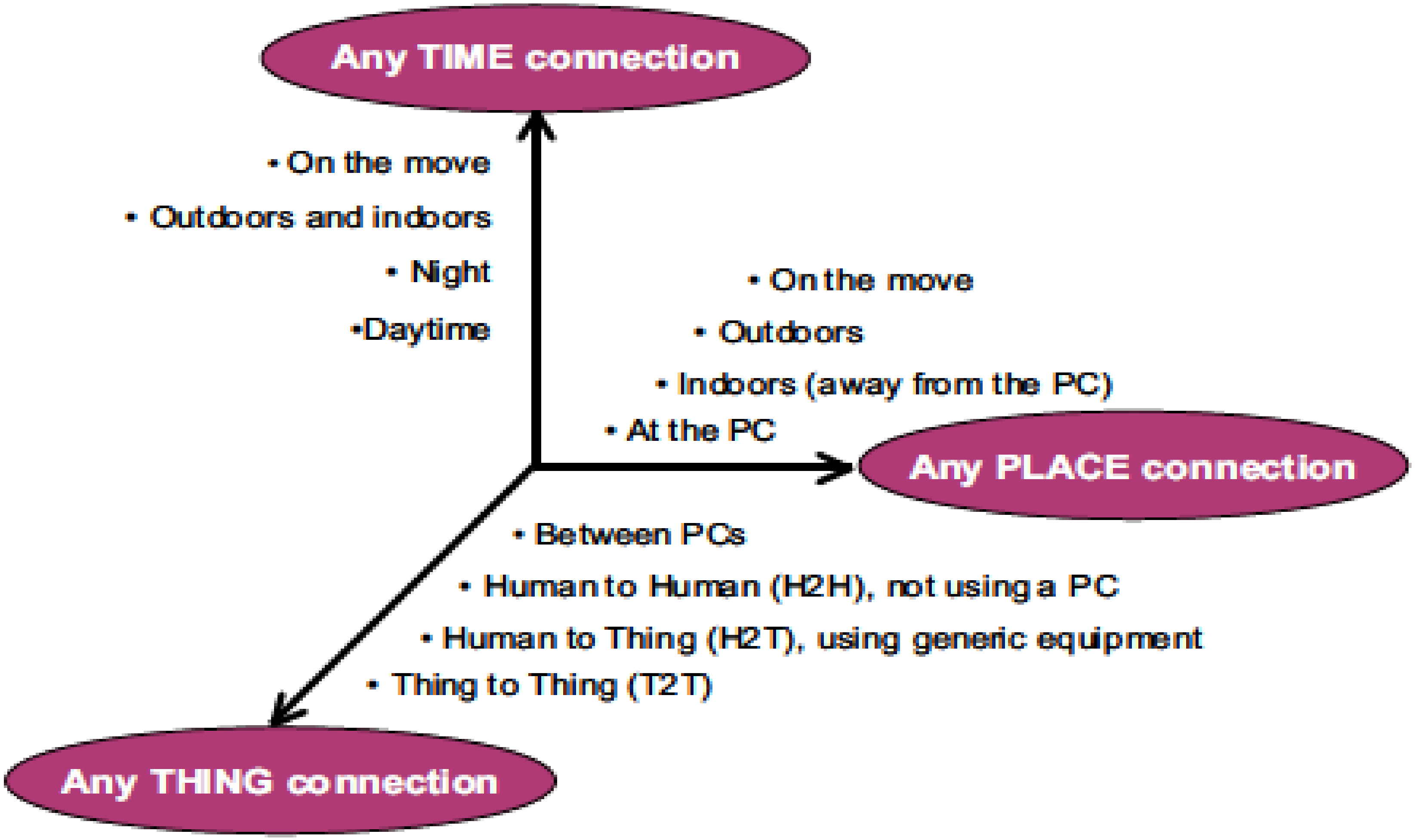
Evolution of the Internet of Things



Introduction



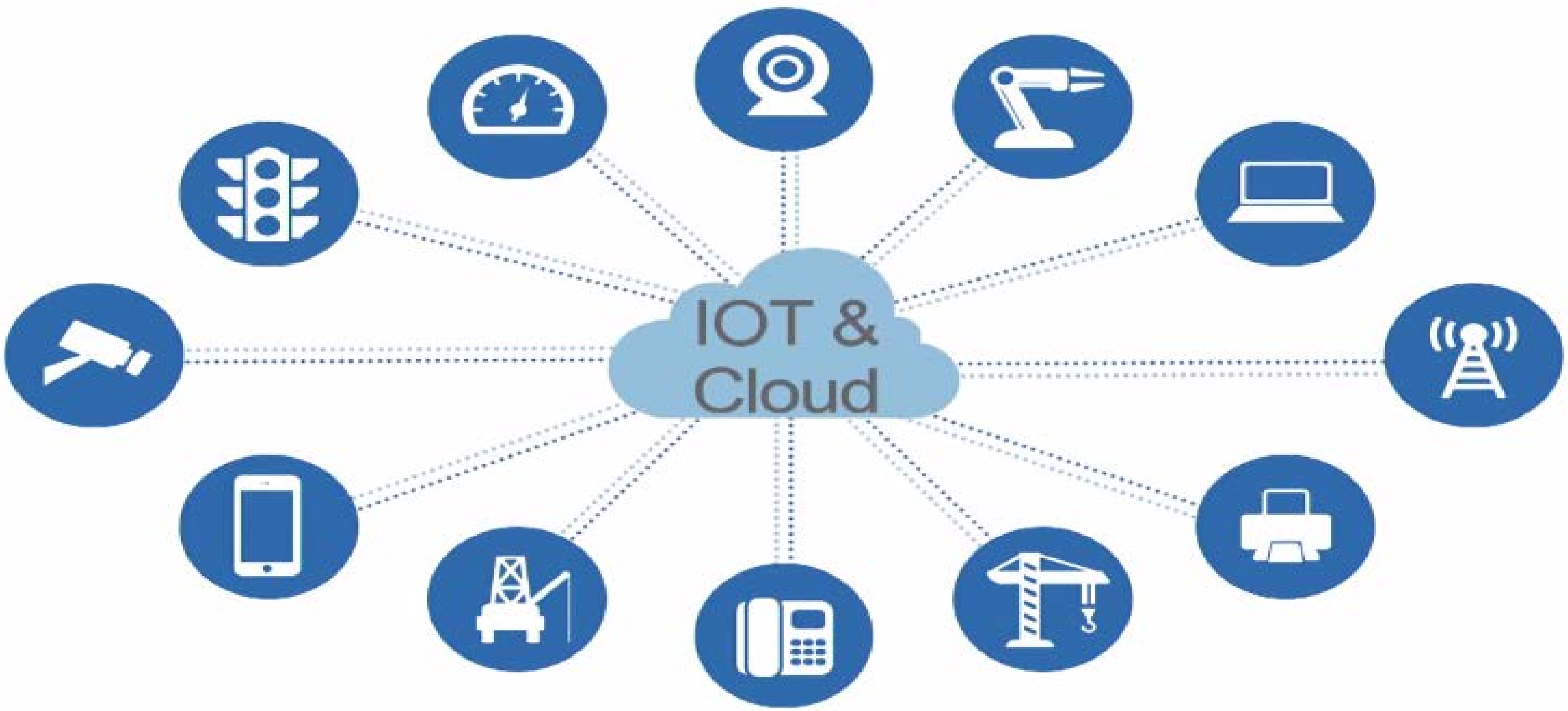
Opportunities of IoT in 3 dimensions



(courtesy of Wikipedia, 2010)

IoT & Cloud Computing

Recent years ago, the IoT started its communication with the centralized server in the cloud.



Basically, Centralized servers hosted in the core internet rather than using a local sever or personal system for huge processing/computation or storage of data.

Introduction

- The challenge of IoT
  - Requires a new computing model
    - Rapid growing 4V of IoT data
      - Volume, variety, veracity and velocity
  - Requirements
    - Minimize latency
    - Conserve network bandwidth
    - Address security concerns
    - Operate reliably
    - Collect and secure data across a wide area with different environmental conditions
    - Move data to the best place for processing

Energy Utility Co.	.5TB/day
Offshore Oil Field	.75TB/week
Large Refinery	1TB/day
Airplane	10TB/30 min of flight

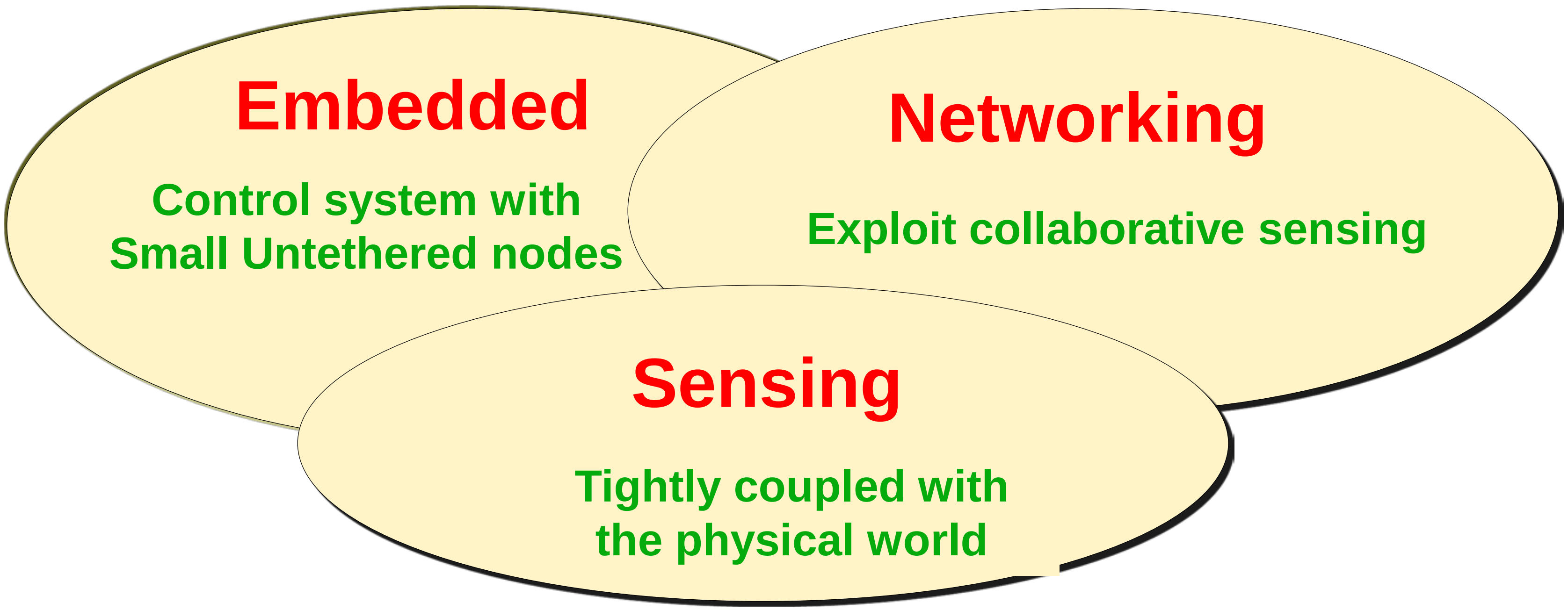
Connecting Many Things to the Cloud is impractical

Traditional cloud computing architectures do not meet all of these requirements

IoT Enabling Technologies

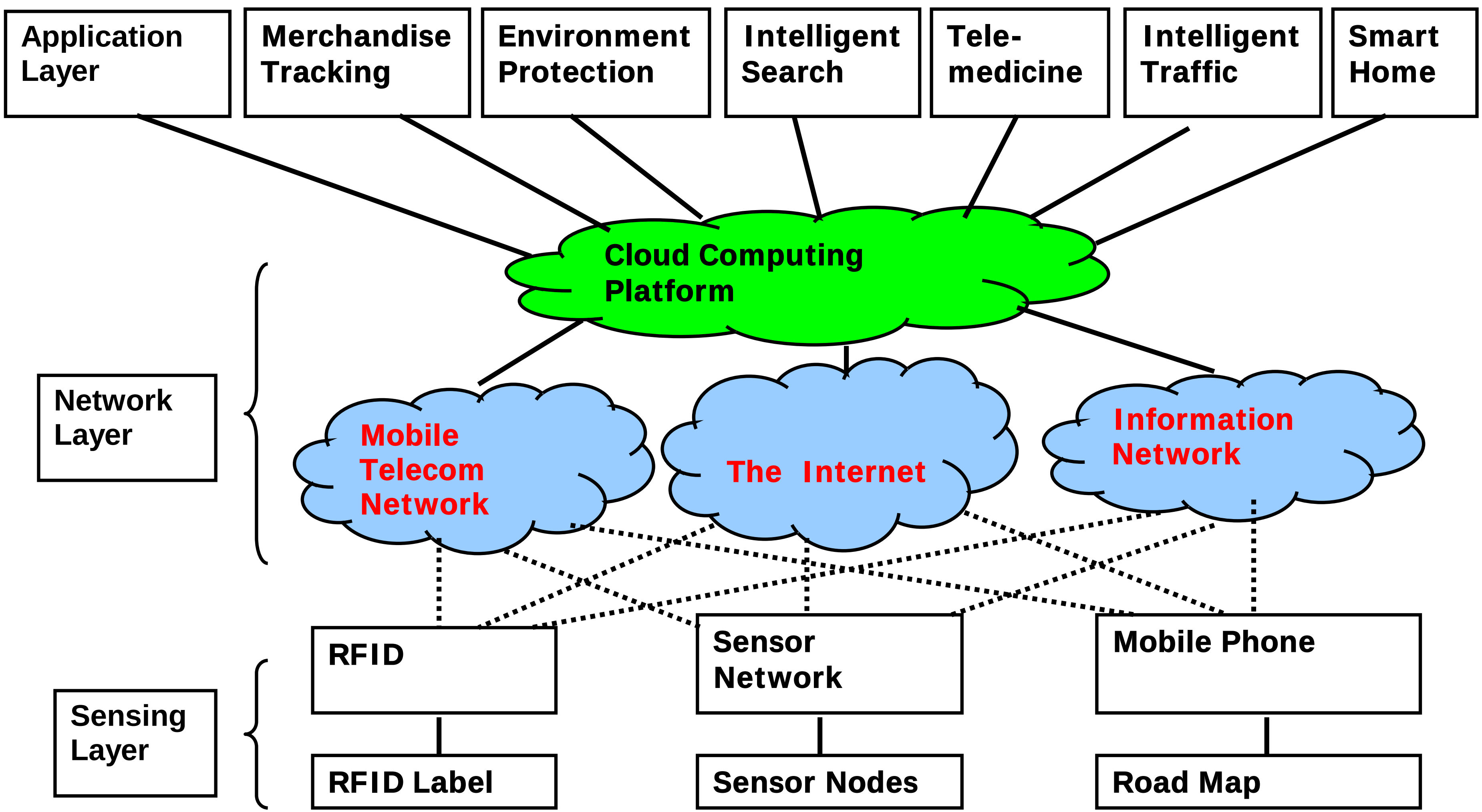
Embed numerous distributed devices to monitor and interact with physical world

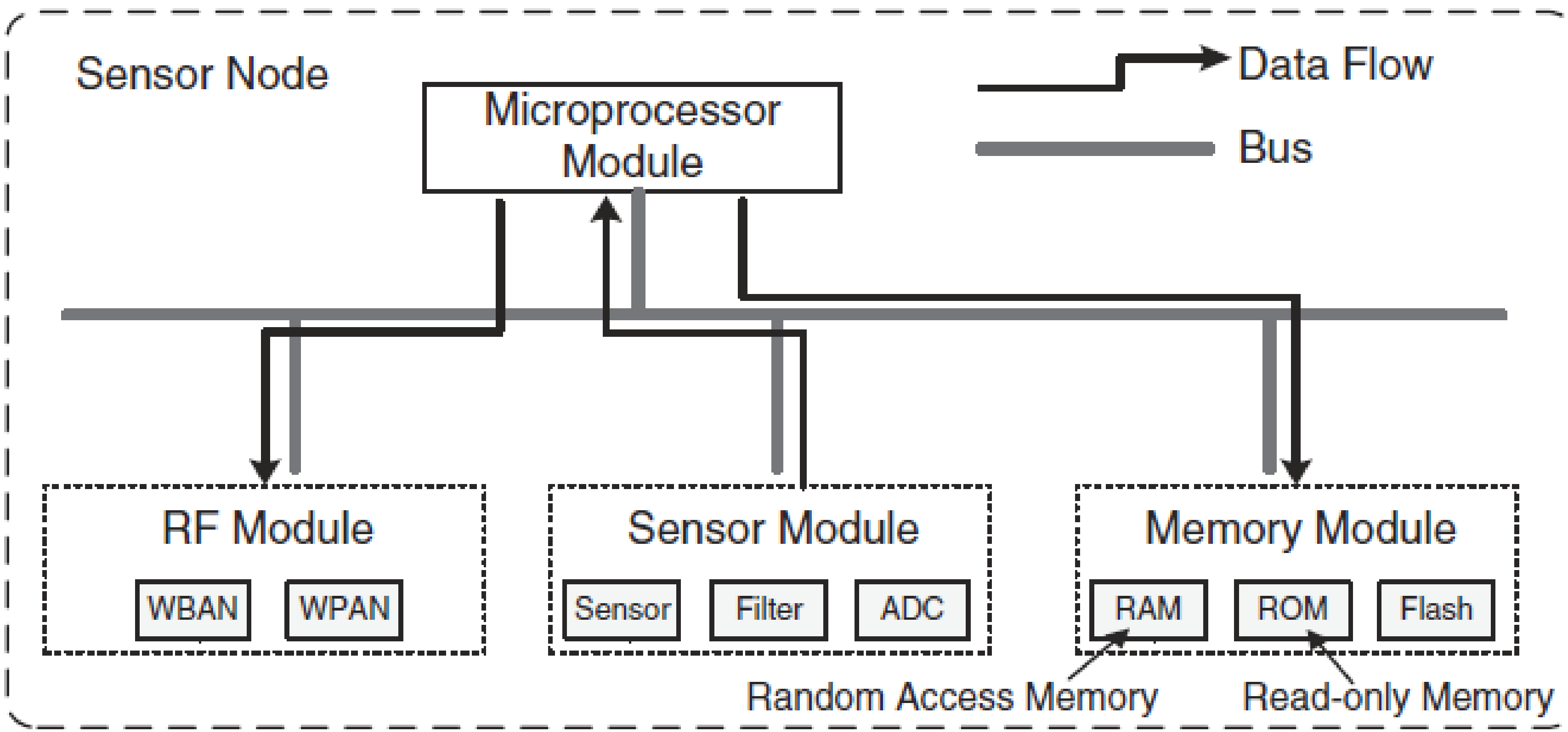
Network devices to coordinate and perform higher-level tasks



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

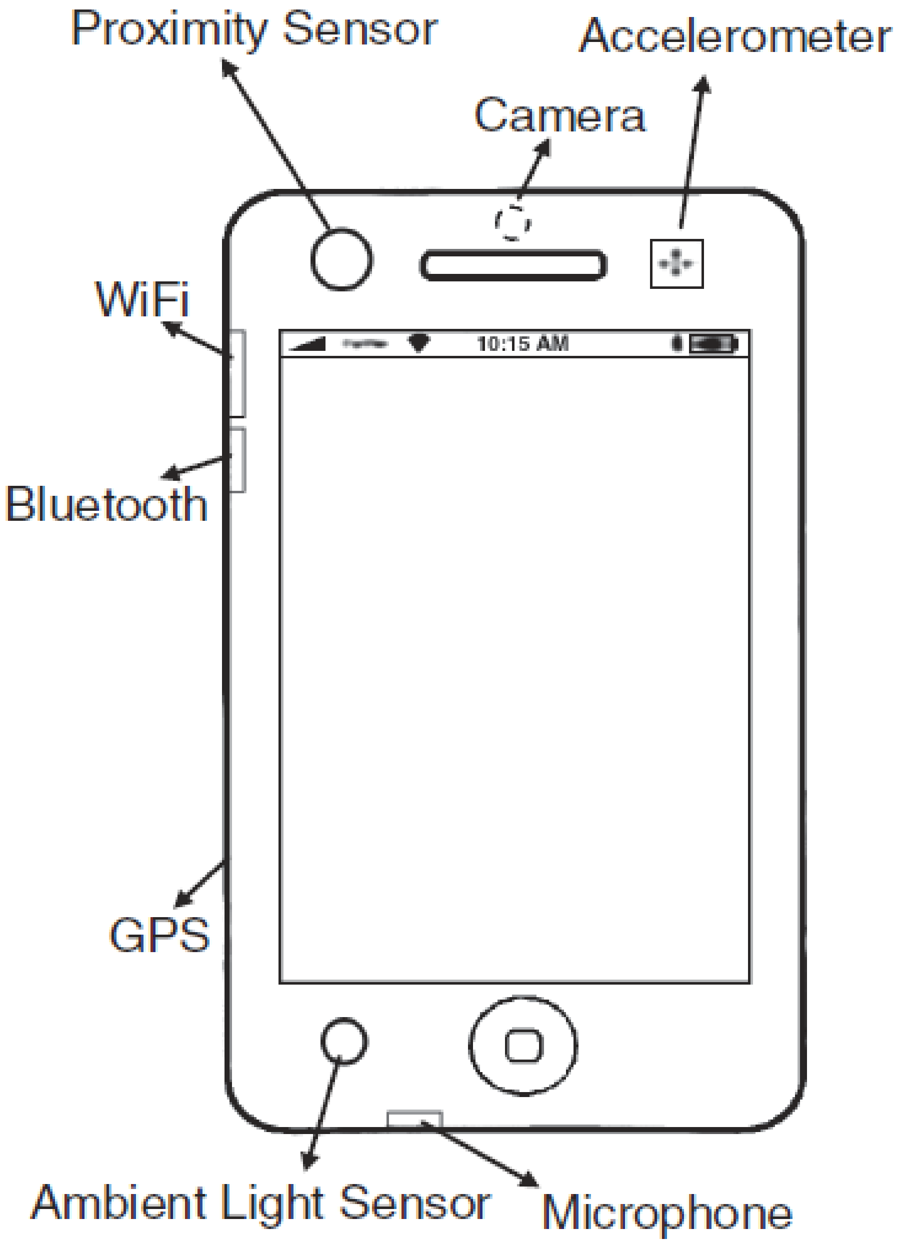
Architecture of the Internet of Things



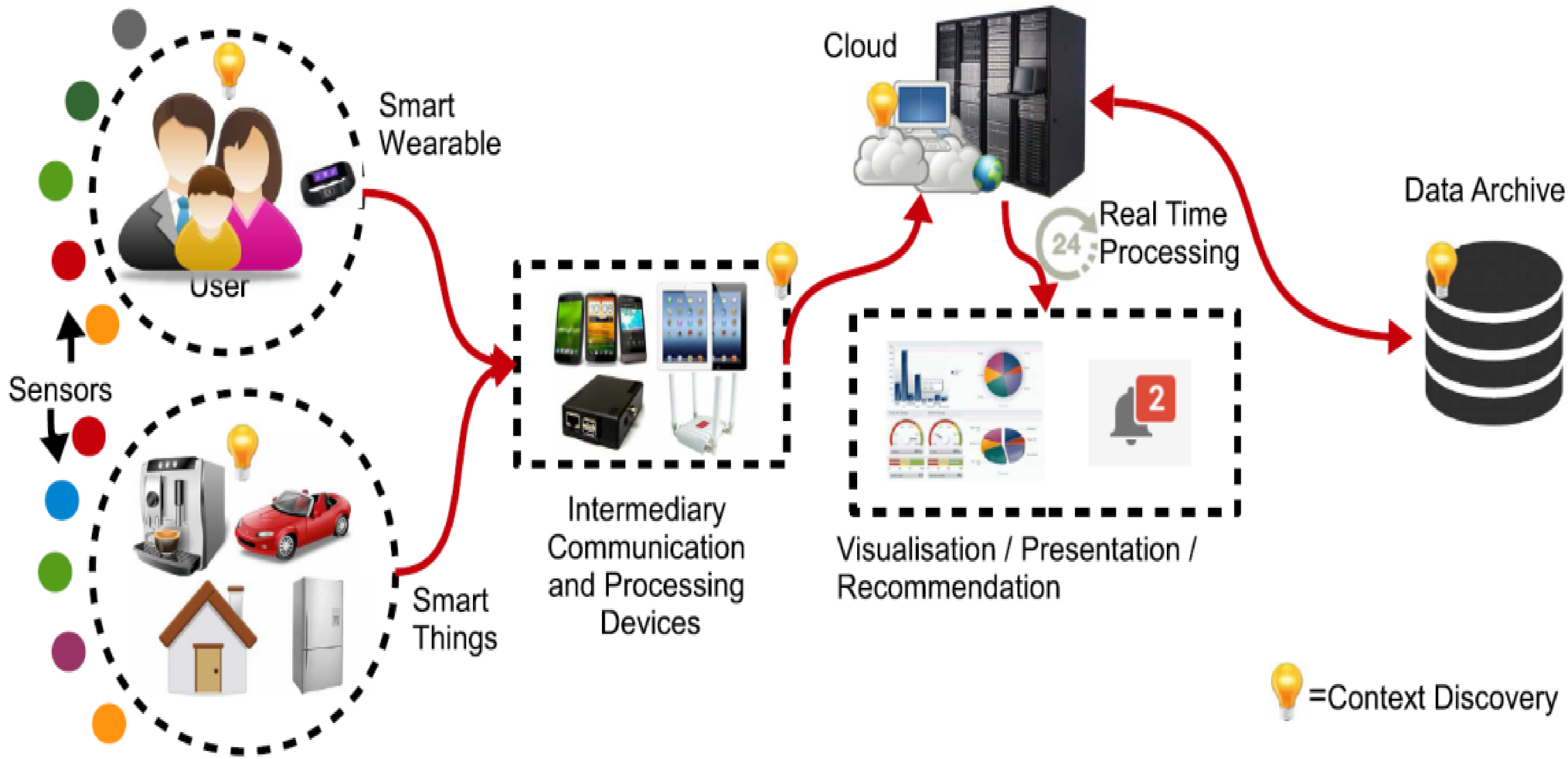


(a) **Functional modules in a smart sensor node.**

Sensor devices built inside a typical smartphone



Location-sensitive Applications of Typical IoT Processing Stages



IoT Apps Domains and Users

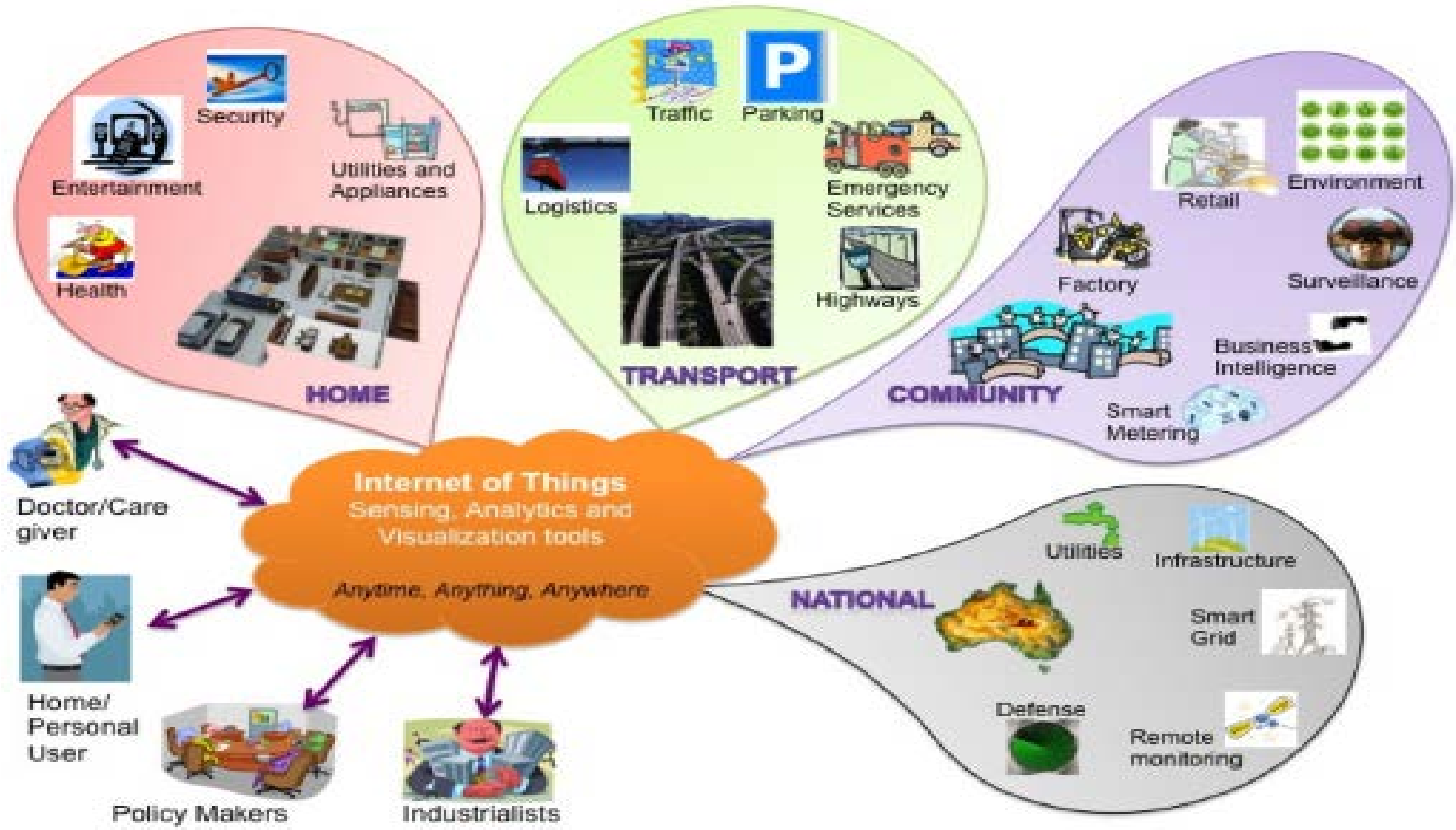


Fig. 1. Internet of Things schematic showing the end users and application areas based on data.

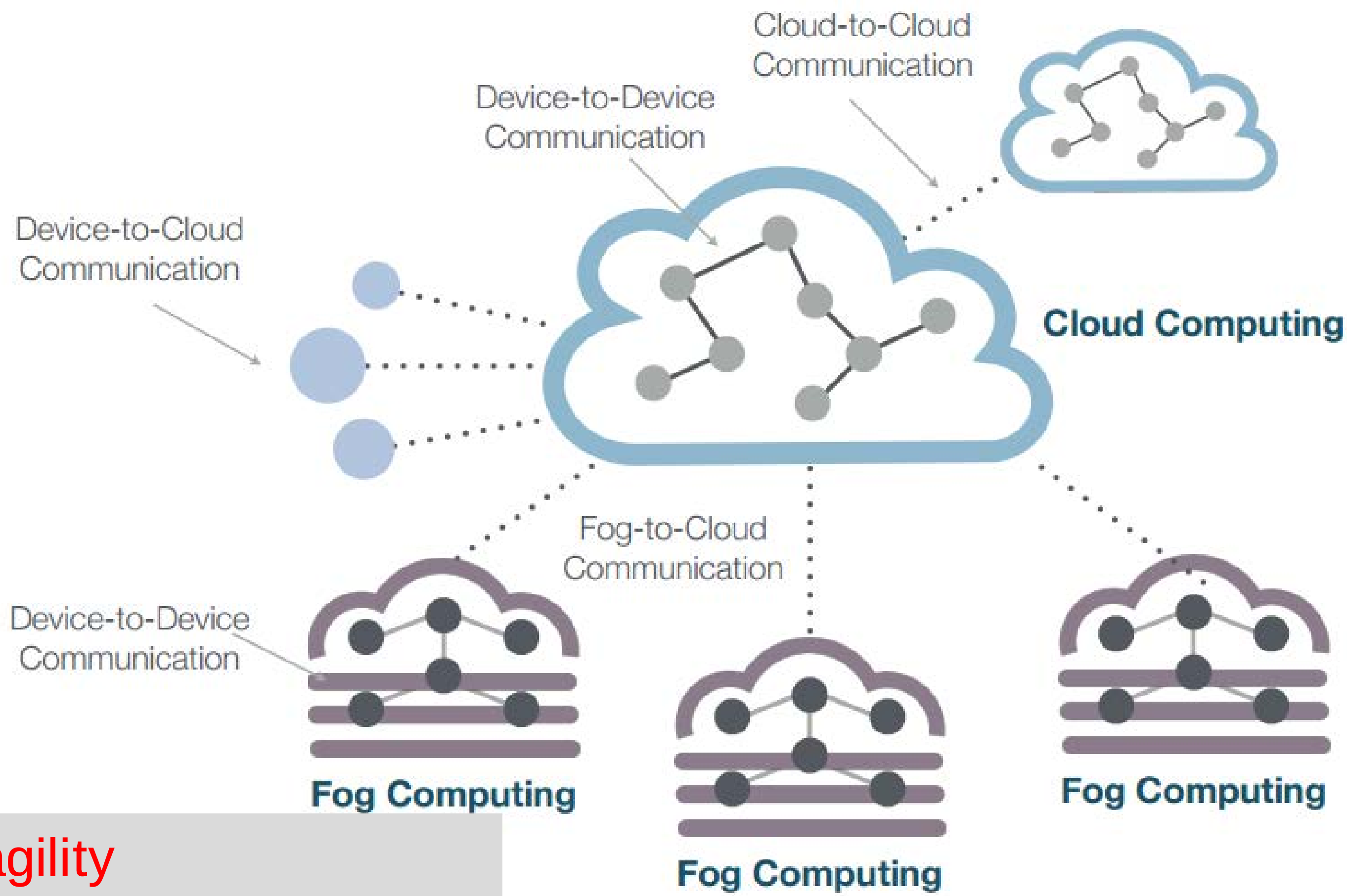
Need for fog computing

- Why can’t do all in cloud?
  - Cloud computing frees for the enterprise and the end user from many details.
  - This bliss becomes a problem for **latency-sensitive** applications.

- Why can’t do all in end systems?
  - Physical constraints: Energy, space, capacity, cost, etc.,

Introduction

- What is the fog computing?

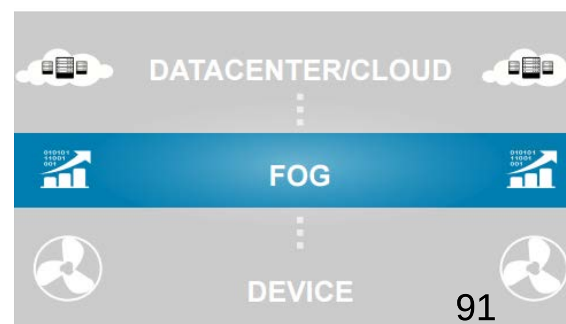


- Greater business agility
- Better security
- Deeper insights, with privacy control
- Lower operating expense

Introduction

Cisco propose in 2014

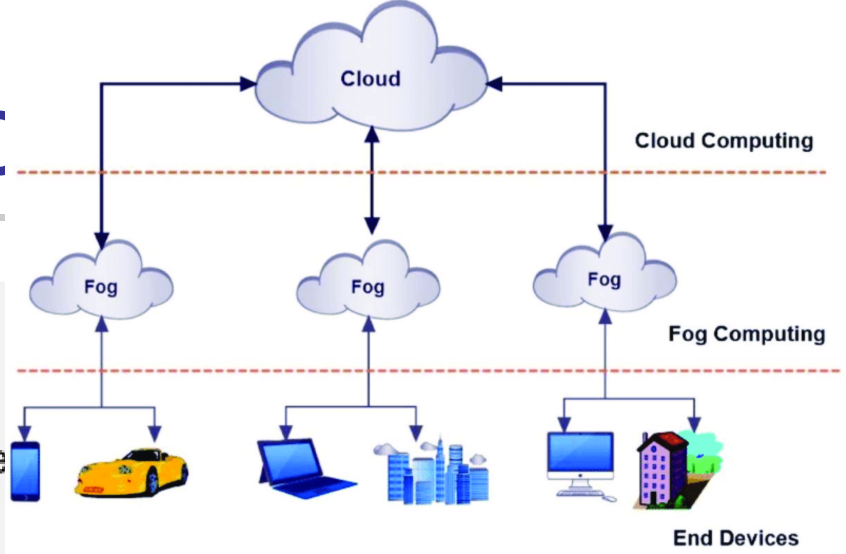
- What is the fog computing?
  - The fog extends the cloud to be closer to the things that process IoT data
    - Fog node
      - Computing, storage, and network connectivity
    - Estimation by IDC[1]
      - The data analyzed on devices that are physically close to IoTs is approaching **40** percent
    - Analyze IoT data close to where it is collected
      - Minimize latency
      - Offload gigabytes of network traffic to the core network
      - Keeps sensitive data inside the network



IDC：International Data Corporation

[1]IDC FutureScape: Worldwide Internet of Things 2015 Predictions.

Introduc



When to Consider Fog Computing

- Data is collected at the extreme edge: vehicle railways, etc.
- Thousands or millions of things across a large geographic area are generating data.
- It is necessary to analyze and act on the data in less than a second.

	Fog Nodes Closest to IoT Devices	Fog Aggregation Nodes	Cloud
Response time	Milliseconds to subsecond	Seconds to minutes	Minutes, days, weeks
Application examples	M2M communication Haptics[2], including telemedicine and training	Visualization Simple analytics	Big data analytics Graphical dashboards
How long IoT data is stored	Transient	Short duration: perhaps hours, days, or weeks	Months or years
Geographic coverage	Very local: for example, one city block	Wider	Global

Introduction

What Happens in the Fog and the Cloud

Fog nodes:

- Receive feeds from IoT devices using any protocol, in real time
- Run IoT-enabled applications for real-time control and analytics, with millisecond response time
- Provide transient storage, often 1–2 hours
- Send periodic data summaries to the cloud

The cloud platform:

- Receives and aggregates data summaries from many fog nodes
- Performs analysis on the IoT data and data from other sources to gain business insight
- Can send new application rules to the fog nodes based on these insights

Vehicular Fog Computing: A Viewpoint of Vehicles as the Infrastructures

X. Hou; Y. Li; M. Chen; D. Wu; D. Jin; S. Chen,
***IEEE Transactions on Vehicular Technology*, 2016**

Introduction

- Urban vehicular networks
  - Invoke by cellular network and cloud computing
  - Intelligent transportation systems’ important component
    - Sharing applications
      - Advertisements and entertainments
    - Information spreading services
      - Emergency operations for natural disaster and terrorist attack
  - Benefits
    - Driving safety
    - Traffic efficiency
    - Great convenience by exchanging valuable information
  - Challenge
    - Bigger data
      - Augmented Reality (AR)
    - Rapid response
      - Self-driving

Complex data processing and storing

computation and storage

For example, a company parking lot can be utilized to handle the computational tasks.

Introduction

- Vehicular fog computing(VFC)
  - Collaboration of near-located vehicles instead of sending the information to remote servers

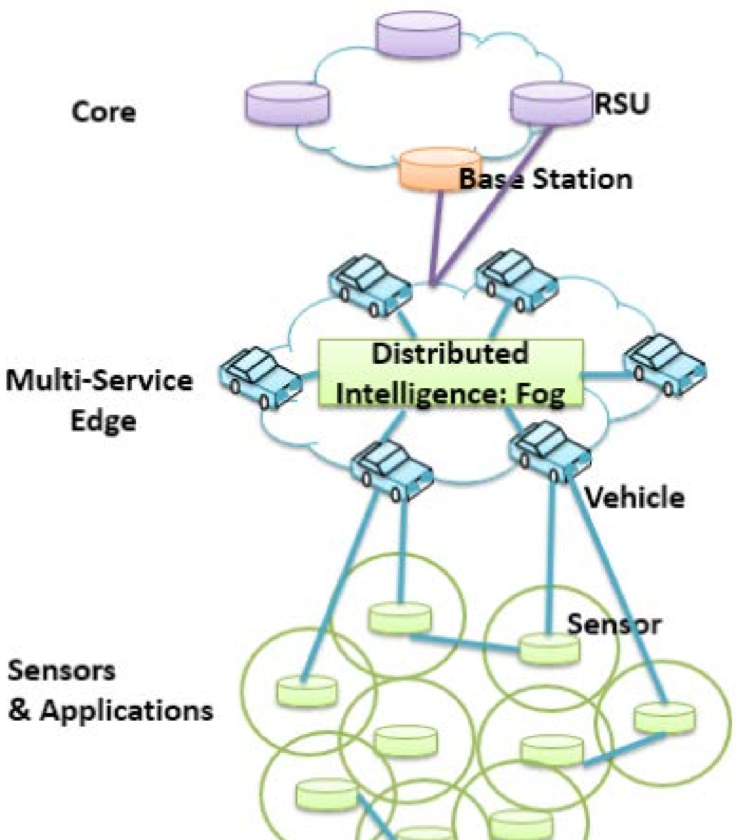


TABLE I

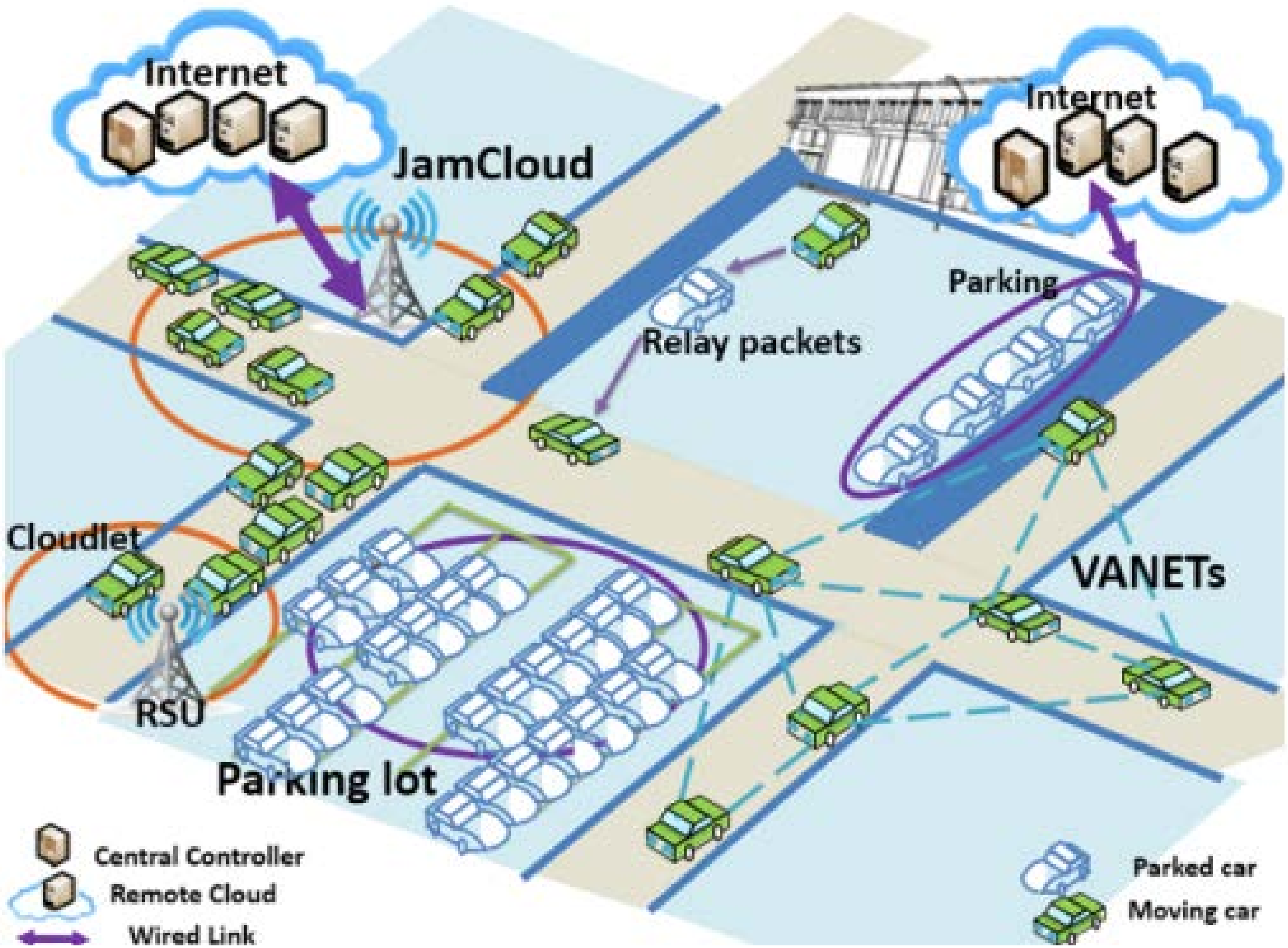
THE DIFFERENCES BETWEEN VCC AND VFC

Features	VCC	VFC
Decision Making	Remote	Local
Geo-distribution	No	Yes
Communication	Bandwidth Constrained	Real-Time Load-Balancing
Computation Capacity	Medium	Large
Deployment Cost	High	Low

RSU : Rode Side Unit

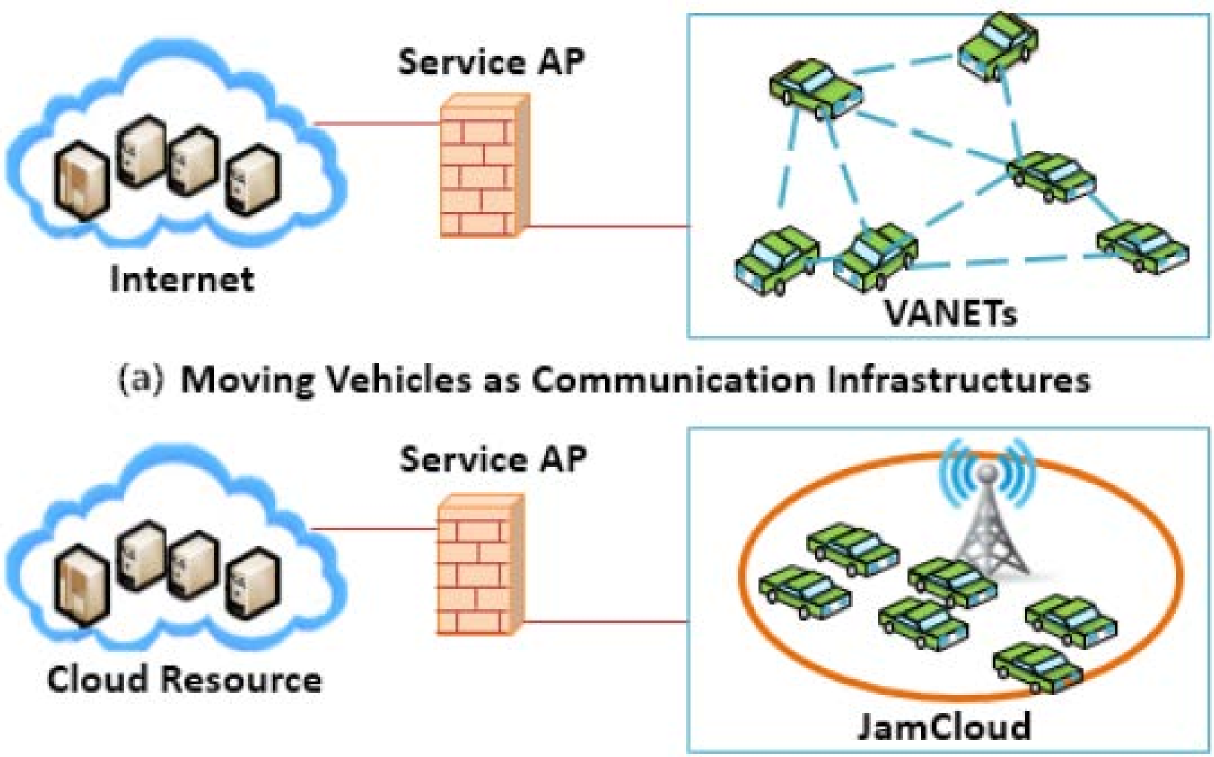
VCC : Vehicular Cloud Computing

System Overview



System Overview

- Scenarios
  - Moving Vehicles as Infrastructures
    - Carry the information from place to place
    - Computation

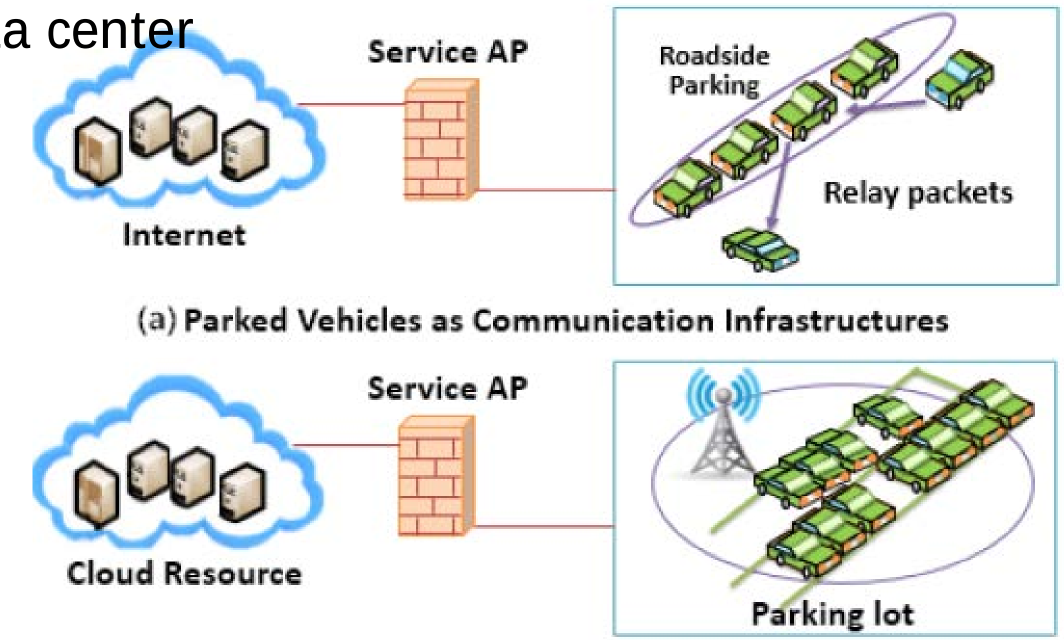


(a) Moving Vehicles as Communication Infrastructures

(b) Moving Vehicles as Computational Infrastructures

System Overview

- Scenarios
  - Parked Vehicles as Infrastructures
    - Can not carry the information from place to place
      - But improve connectivity.
        - Static backbones and service infrastructures
    - Computation
      - Form a small data center



(a) Parked Vehicles as Communication Infrastructures

(b) Parked Vehicles as Computational Infrastructures