



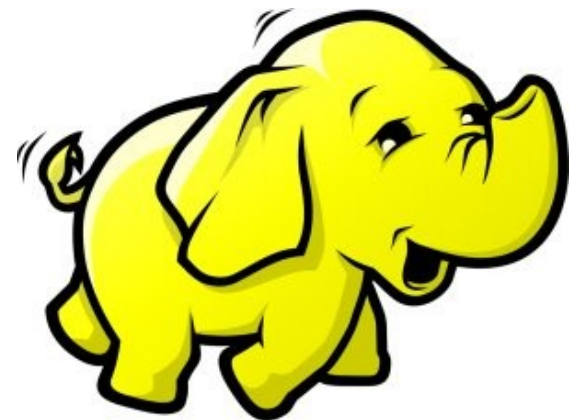
巨量資料的奇幻漂流

Life of Big Data

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw



WHO AM I ? 這傢伙是誰啊? JAZZ ?

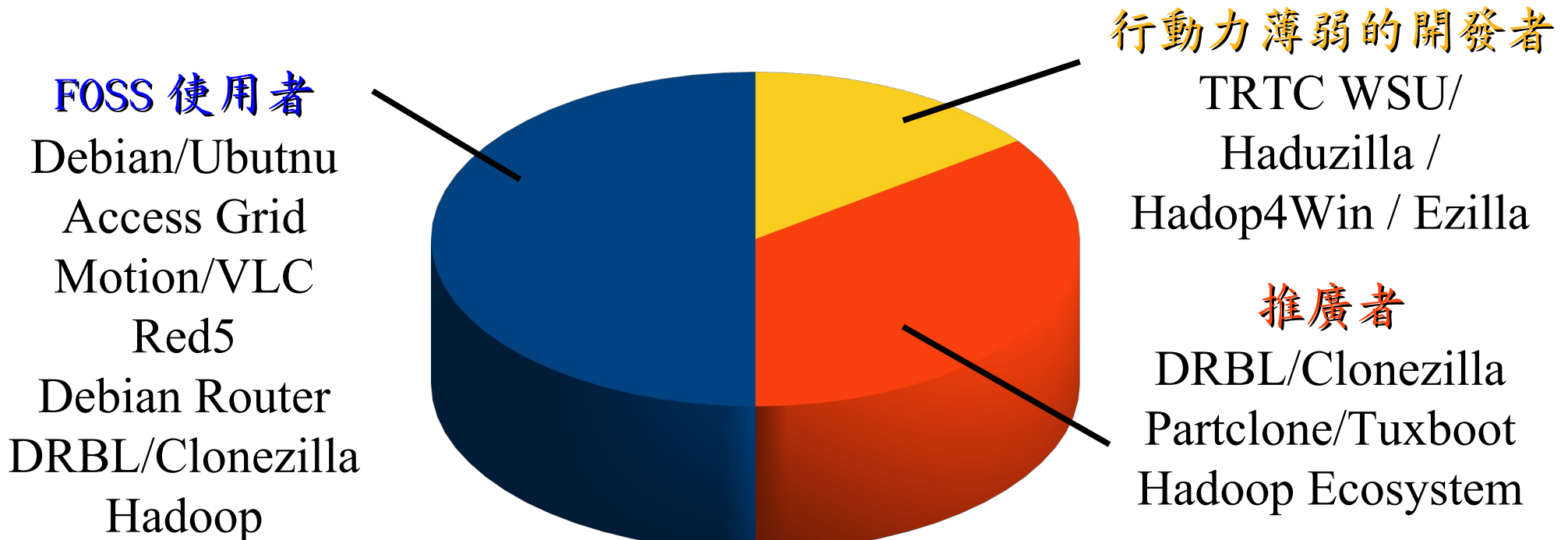
- 講者介紹：

- 國網中心 王耀聰 副研究員 / 交大電控八九級碩士
- jazz@nchc.org.tw



- 所有投影片、參考資料與操作步驟均在網路上

- <http://trac.nchc.org.tw/cloud>
- 由於雲端資訊變動太快，愛護地球，請減少不必要之列印。



【資料？ Data ？】

各位數位原住民

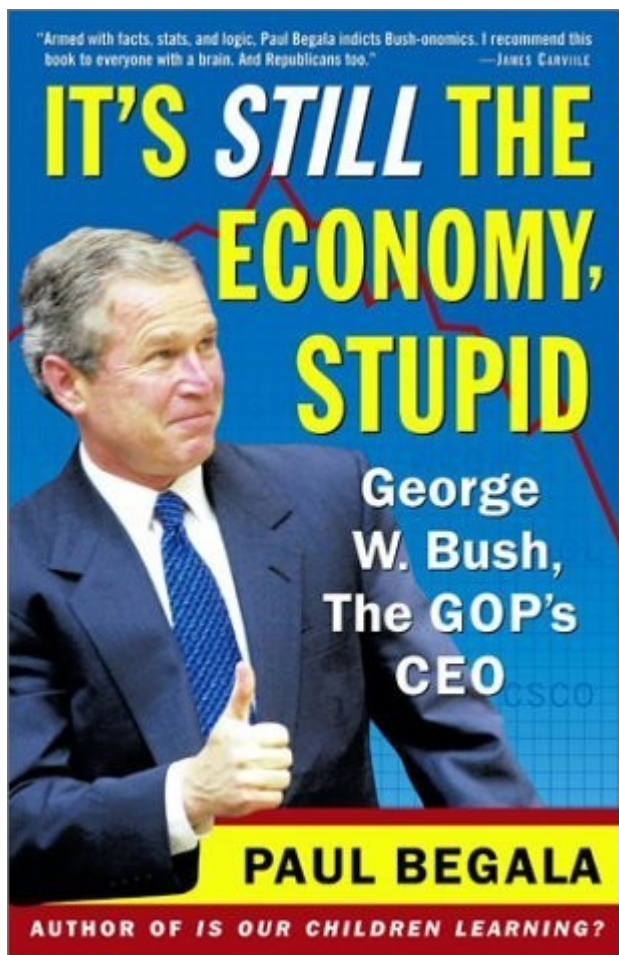
請不要忽視資料的重要性

資料會是下個世紀的黑金

淘金前，先知道資料在哪

接著學習用哪種工具採礦

誰掌握了你的資料，就有機會掌握你的荷包



「笨蛋！重點在經濟」

(**"It's the economy, stupid"**)

卡維爾 (**James Carville**) 自創這句標語，
促使柯林頓當上美國第 **42** 屆總統。

- **1992** 年

「笨蛋！重點還是在經濟」

(**"It's STILL the economy, stupid"**)

卻讓小布希嘲笑是幼稚的總統。

- **2002** 年

雲端時代，○○會說：「笨蛋！重點在資料」

(**"It's the data, stupid"**)

誰掌握了你的資料，就有機會掌握你的荷包
想想看，電腦、手機掉了，您心疼的是甚麼呢？

- **2007** 年

【問題】

如果○○○把您的信件、照片通通搞丟了，該賠償您多少錢呢？用 MB 計算嘛？

Agenda 演講大綱

From **WSN** to **IoT**?

物聯網的崛起

What is **Smart City** ?

淺談智慧城市

What is **Open Data** ?

何謂開放資料

What is **Big Data** ?

何謂海量資料

Join Communities !

加入國內社群

資料從哪裡來？往哪裡去？

如何解開資料流的限制？如何處理龐大的資料呢？

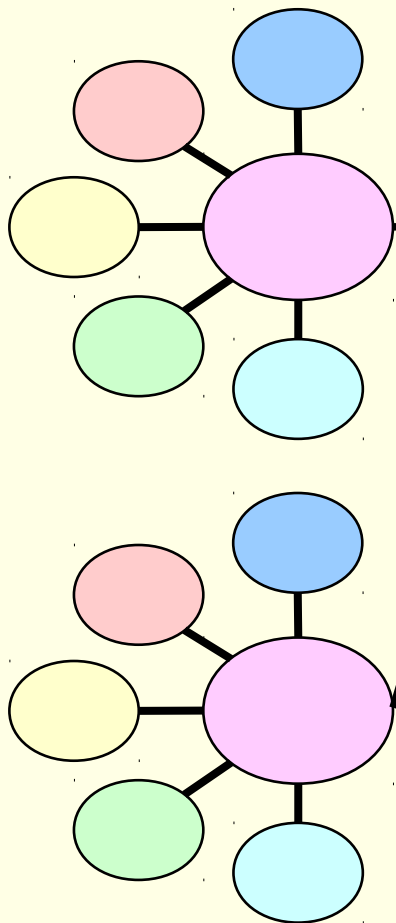
This is a story
about Data Flow

今天要講一個關於
資料的漫長旅程

巨量資料的奇幻漂流 Life of Big Data

Internet of Things
物聯網

Sensor Network



Smart Grid

雲 資料中心
提供服務

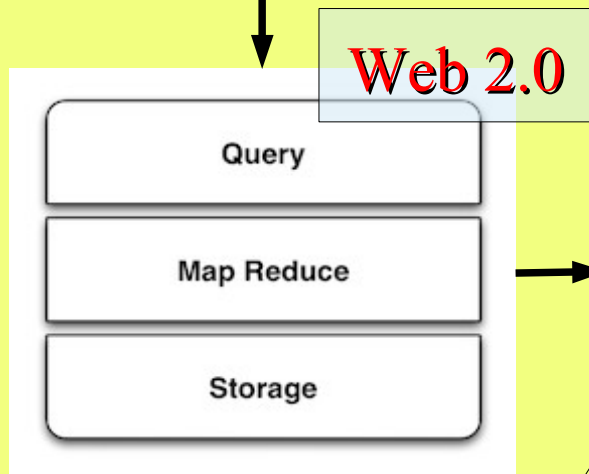


Public Data Hub
Data as a Service

開放資料
Open Data



Web 2.0



Big Data



Mobile Computing

端
各類裝置
存取服務

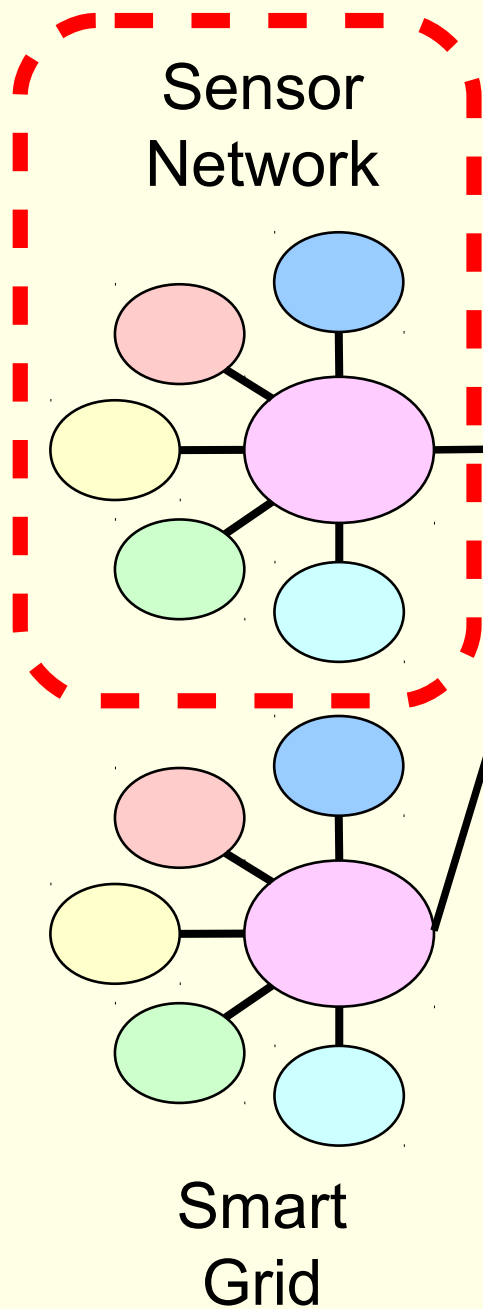
Cloud Computing

雲端運算

巨量資料的奇幻漂流 Life of Big Data

Internet of Things

物聯網



雲 資料中心
提供服務



Public Data Hub
Data as a Service

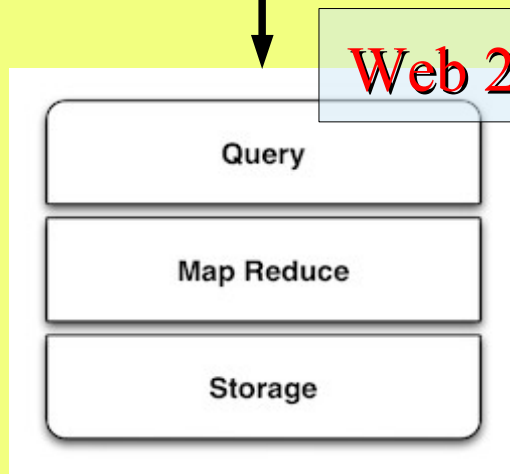
開放資料
Open Data



Cloud Computing

雲端運算

Web 2.0



Big Data



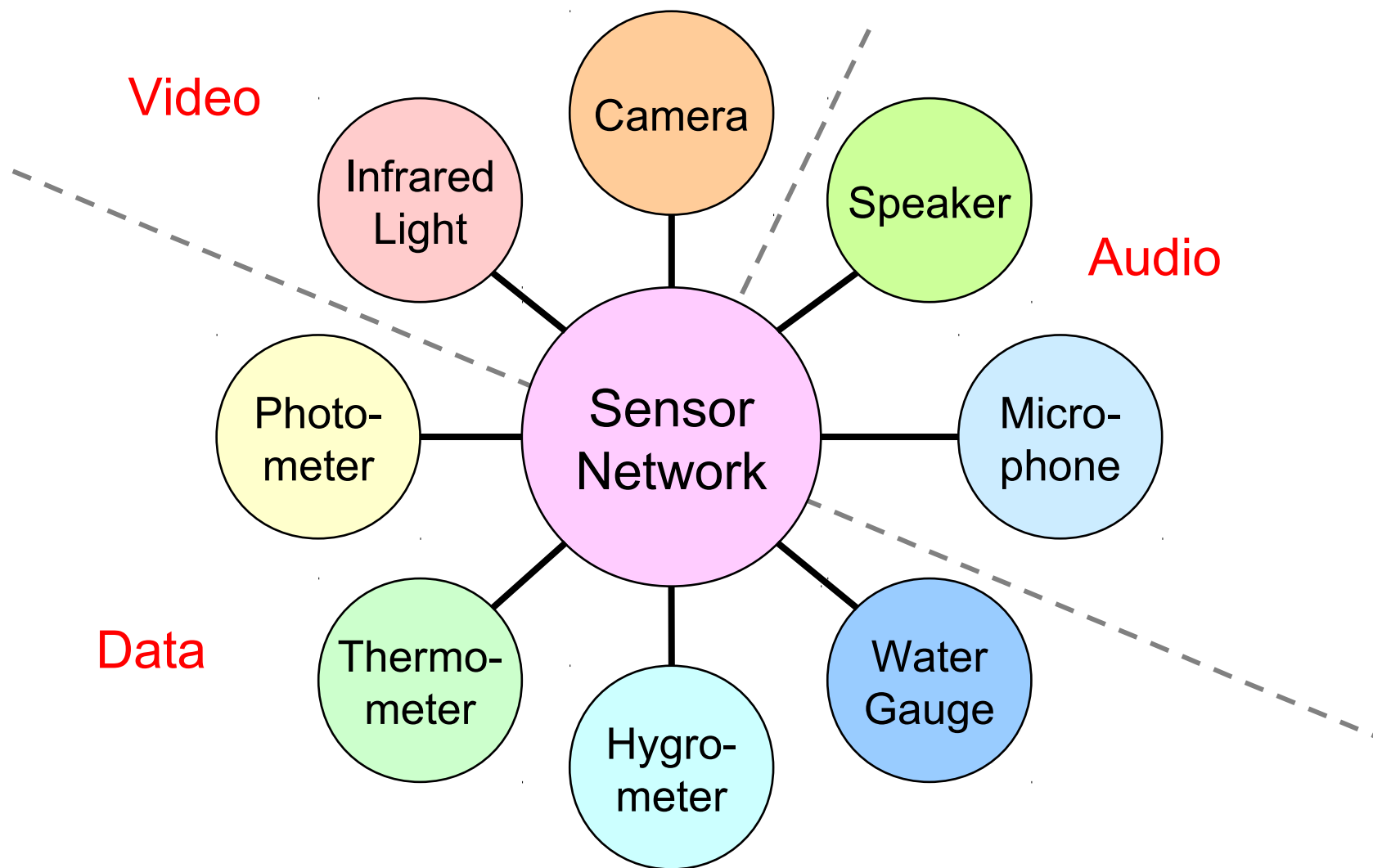
端

各類裝置
存取服務

Mobile Computing

無線感測器網路 (Wireless Sensor Network)

感測器網路泛指由各式各樣不同的感測器所組成，透過網路鏈結成一大型的資訊蒐集網。



生態觀測格網 (Ecology Grid)



[Ecogrid Insight](#) | [Events](#) | [Ecosites](#) | [Ecogrid on Media](#) | [Data Query](#) |

Ecogrid project is a joint effort among National Center for High-Performance Computing (NCHC), Taiwan Ecological Research Network (TERN) and Taiwan Forestry Research Institute (TFRI).

Using state-of-the-art grid technology, scientists at NCHC established a cyberinfrastructure for ecological research, the Ecogrid, which integrates geographically distributed sensors, computing power, and storage resources into a uniform and secure platform. Ecologists can conduct researches on this platform, including data acquisition, data analysis, and data sharing. The real time intervention capability of the Ecogrid system allows researchers to carry out interactive/control experiments that have been impossible until now.

Global Lake Observatories Network

With the support from PRAGMA, Ecogrid is now extended to international. By the cooperation with scientists from Center for Limnology of University of Wisconsin, San Diego Supercomputing Center (SDSC), and Center for Biodiversity of Academia Sinica, we deployed a lake observatories network connected LTER Trout Lake station in Wisconsin and TERN Yuan Yang Lake Station in Taiwan in April 2004. Through the ubiquitous web interface, scientists can easily access and compare data from both research stations in near real-time. The pioneer network was used as an example to call for broader partnership when scientists gathered at Scripps Institute of Oceanography in March, 2005 to discuss next move upon the global network.

Underwater Reef Monitoring

Coral reefs are precious because they are the most biologically diverse and productive marine ecosystems on earth. In addition, revenue from tourists attracted to the beauty of coral reefs can be a significant source of income for human communities in these areas. Because of the important ecological and economic roles coral reef communities fulfill, resources are pulling in to study the stresses and dangers to the reefs.

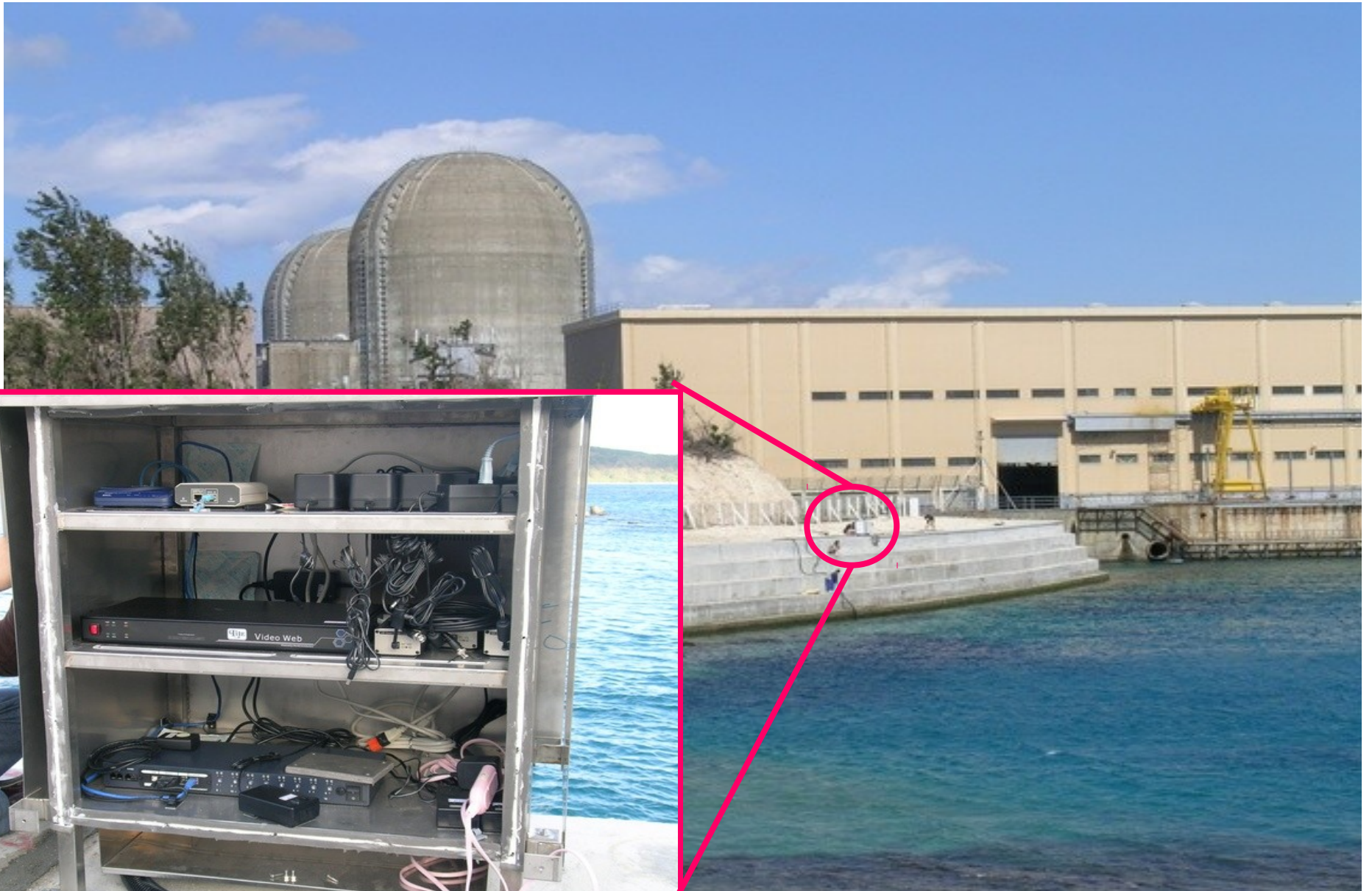
With the funding from Taipower, NCHC Ecogrid team worked together with Research for Biodiversity Center, Academic Sinica implemented an underwater reef monitoring system. The system composed with 9 cameras, located at 3 sites, lively streaming bioactivities undergoing in reef communities to marine scientists' desktop.

<http://ecogrid.nchc.org.tw>

Underwater Camera



Site A



Site B & C



農業格網 (Agriculture Grid)

農業格網 田間監測系統

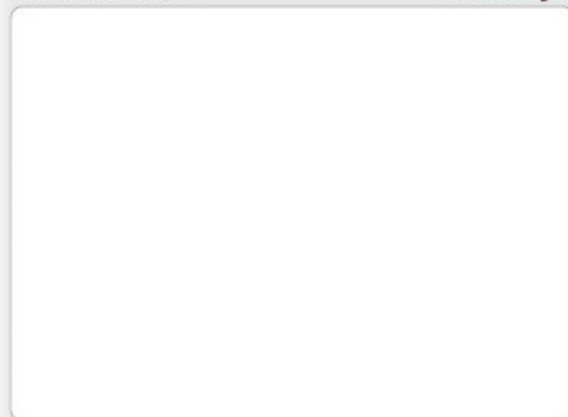
[關於本計畫](#) [計畫成員](#) [合作單位](#) [相關連結](#)

斗南地區根莖類作物種植實況



即時影像 ▶

歷史影像 ▶



如無法觀看，請先下載 [ActiveX Plug-In](#)

即時數據 ▶

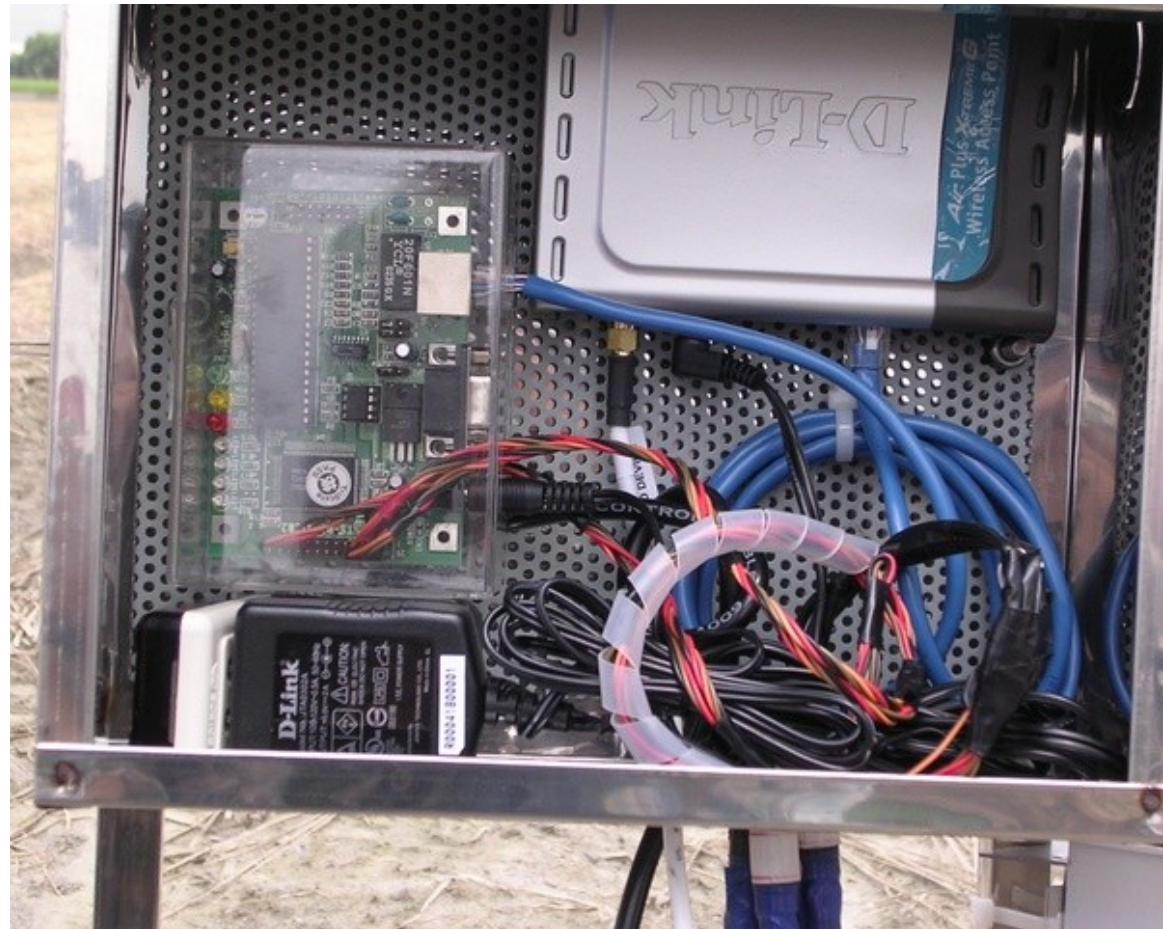
歷史數據 ▶

空氣溫度	土壤溫度	土壤溼度	雨量計
28°C	27°C	36%	0

Electric Wire Post



Passive Field Server



文湖線無線管理單元 (TRTC Neihu WSU)



<http://www.flickr.com/photos/rail02000/5292653579/lightbox/>

Wireless Fast Roaming

松山
機場

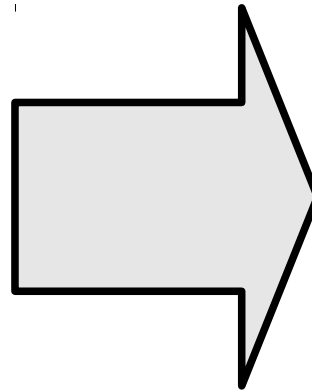
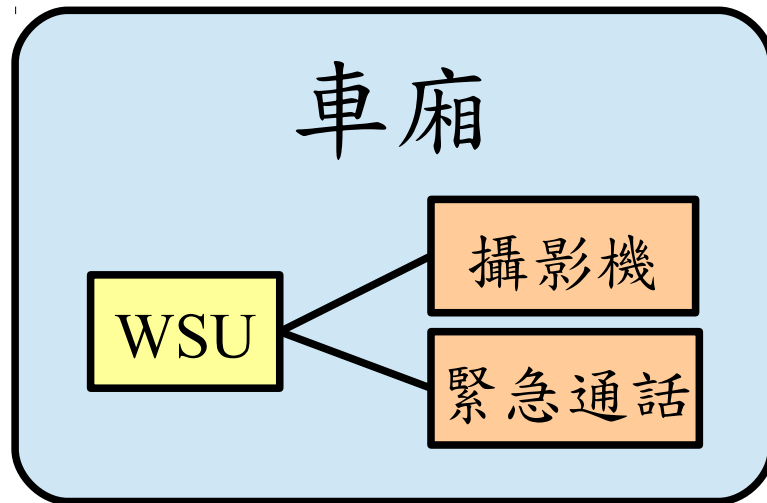
AP #1

AP #2

中山
國中

光纖

GPRS / 3G 太貴，得自建 WiFi
間隔 400 公尺，時速 60 公里，
如何快速地在無線基地台切換呢？



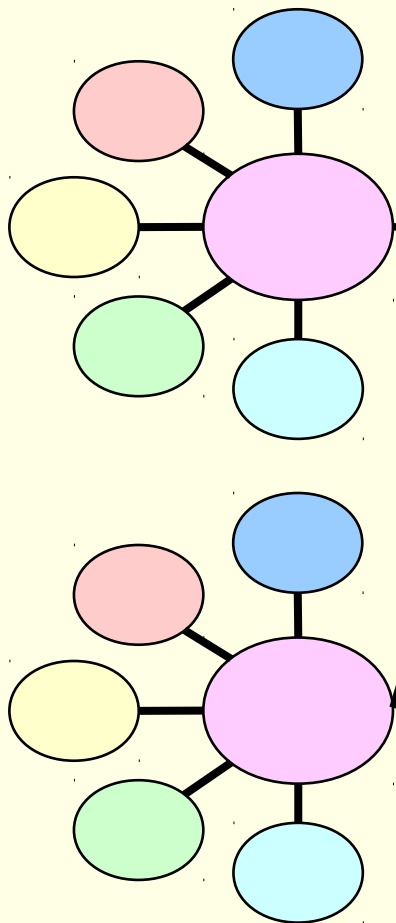
300ms 完成漫遊

切換時聲音跟影像資料不能中斷！

巨量資料的奇幻漂流 Life of Big Data

Internet of Things 物聯網

Sensor Network



Smart Grid

雲 資料中心 提供服務

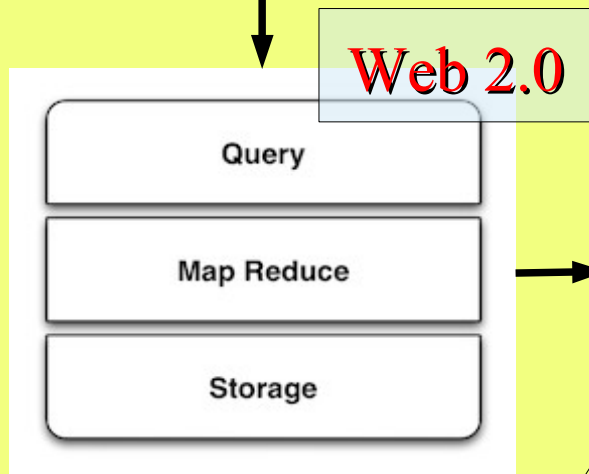


Public Data Hub
Data as a Service

開放資料
Open Data



Web 2.0



Big Data



Mobile Computing

端 各類裝置 存取服務

Cloud Computing

雲端運算

物聯網的時代來臨

Internet of Things

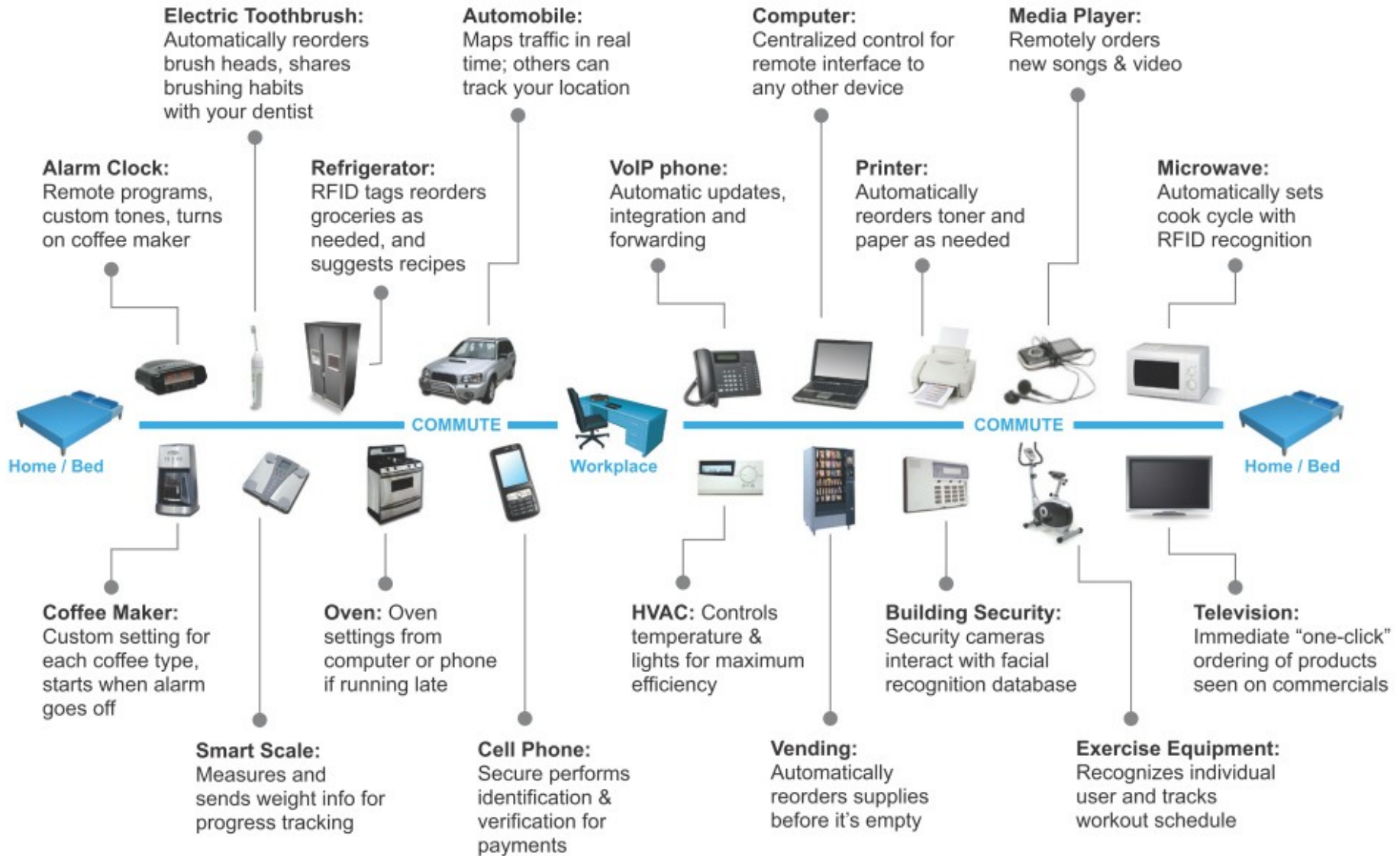


Figure 3. The Internet of Things

Wifi體重計：透過網路管理你的體重



無敵小恩恩 發表於 2009年10月05日 13:54 | [收藏此文](#)



網路入侵到家電用品只是時間的問題，現在體重計也有內建無線網路，透過網路能將你每天量的體重都紀錄下來，不用筆記也能做好更準確的體重管理。有了這麼方便的上網體重計，這下沒有藉口懶得管理你的體重了。 懶人減肥專用機

不只是記錄體重，這台體重計還能算出你的BMI，不但能將資料傳到網站上去，也能透過Wifi傳到iPhone上，經過每天的測量之後，能在網路上看到曲線圖，進而管理你的體重，如果上升的太快就表示最近吃太多動太少，對於懶得記錄自己體重的人來說，這是一台不錯的監督機器，而一台要價159美金，約5000元新台幣，並不便宜，昂貴的價格要普及化還是有一段很長的時間，而且台灣也沒販賣，要買到也不容易。

<http://www.techbang.com/posts/1258-wifi-scales-through-the-internet-to-manage-your-weight>

【另類解釋】

物聯網 =
連結多個感測網路

Ex. 智慧家庭

氣象局 + 鬧鐘 + 窗簾
+ 咖啡機 + 智慧車輛

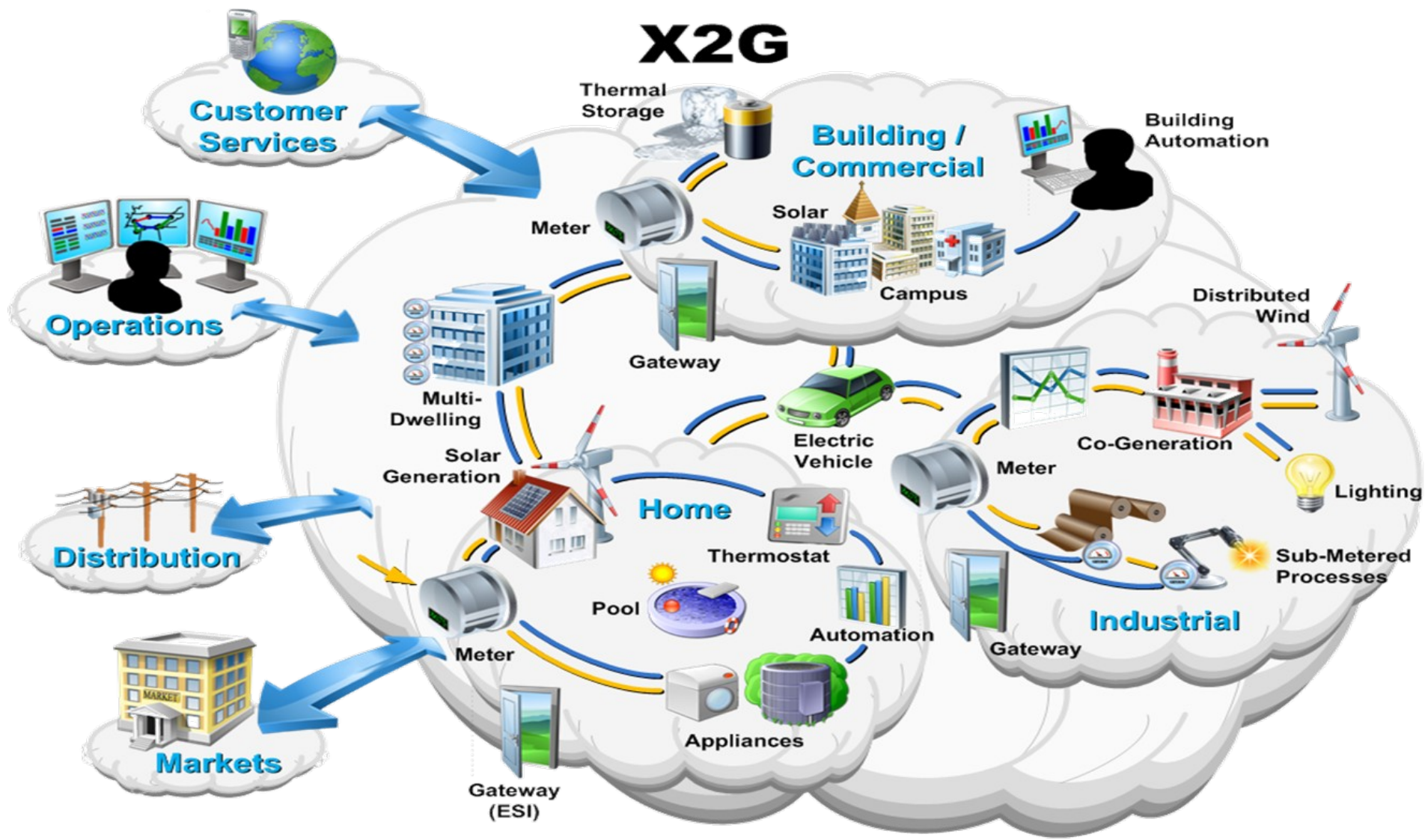
【關鍵技術】

M2M Protocol

(Machine To Machine)

哪一個才是跨感測裝置
的通訊格式標準？

智慧電網的新政策 Smart Grid



【規模擴大到城市】

智慧城市
Smart City

Agenda 演講大綱

From **WSN** to **IoT**?

物聯網的崛起

What is **Smart City** ?

淺談智慧城市

What is **Open Data** ?

何謂開放資料

What is **Big Data** ?

何謂海量資料

Join **Communities** !

加入國內社群

十九世紀是帝國的世紀
二十世紀是國家的時代

ANATOMY OF A SMART CITY

"The 19th century was a century of empires, 20th century was a century of nation states and the 21st century will be a century of cities."

- Former Denver Mayor W. Webb

二十一世紀將是城市的時代

<http://www.postscapes.com/anatomy-of-a-smart-city-full>

2008 年以後已有 50% 以上的人口居住在城市裡

THE WORLD IS NOW URBANIZED

% of population living in cities

1800



3%

1950



29%

2008



50%

2040



65%



1.3 million

people are currently moving
into cities each week

每個禮拜有一百三十萬人
搬入城市

全球有二十一個《超城市》

東京是目前全球最大的城市

There are

21

MEGACITIES

With over 10 Million people

Up until 1975 there were just **3**
New York, Tokyo and Mexico City



Largest City:

Tokyo 36 Million +

If it were a country, it would rank 35th
in population size

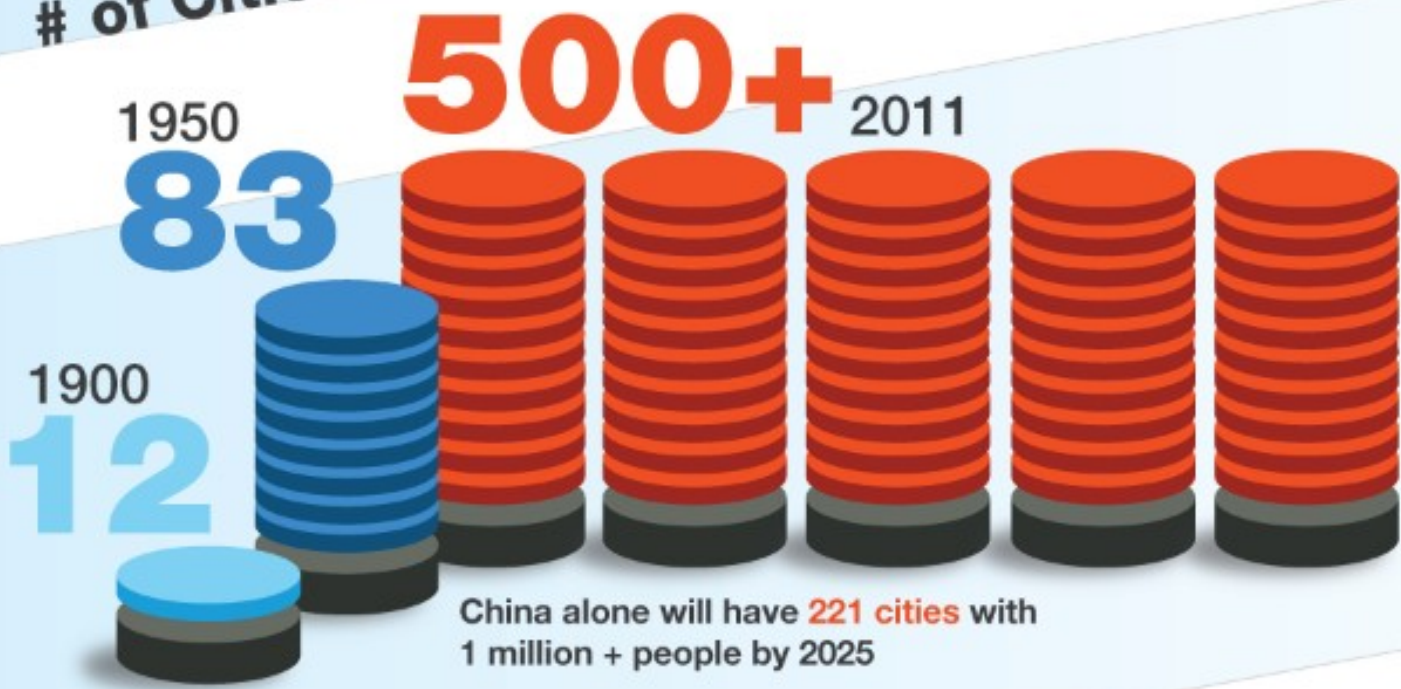


29

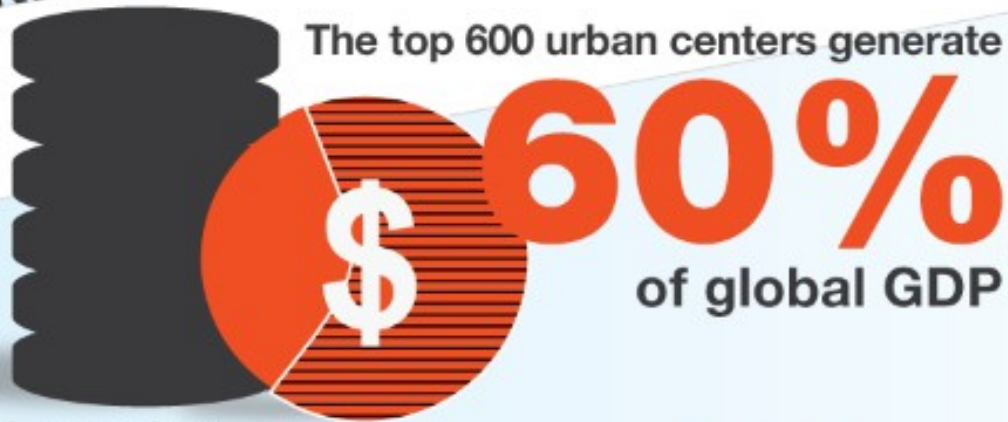
By 2025, the number of megacities is expected to reach
with an additional five in Asia (Shenzhen, Chongqing, Guangzhou, Jakarta and Lahore),
two in Latin America (Bogotá and Lima), and one in Africa (Kinshasa)

2011 年全球已有五百多個城市超過一百萬人口
前六百大城市貢獻了百分之六十的生產毛額

of Cities with more than 1 million People



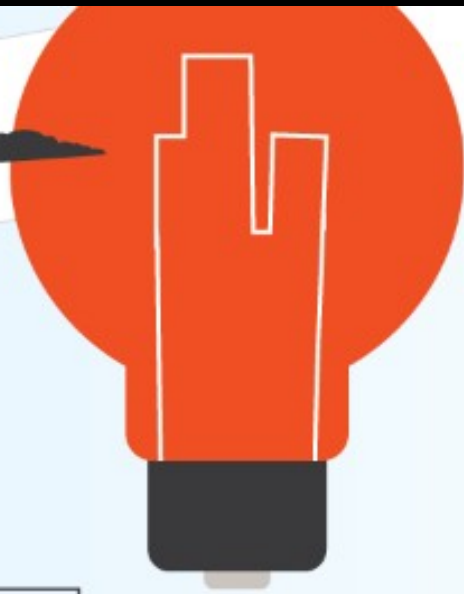
ECONOMIC INFLUENCE



In the developing world, as much as 80% of future economic growth will occur in cities

ENVIRONMENTAL IMPACT

Cities use 
60%-80%
of the world's annual energy needs



Lighting alone
represents **19%**
of the world's
total electricity
consumption

ALL OF THE ABOVE

The need for
SMARTER
CITIES

城市使用了百分之六十至八十的能源

The need for SMARTER CITIES

DRIVEN BY:

Sensors



Networks



Engagement



RESULTING IN:

REAL-TIME URBAN INFORMATICS

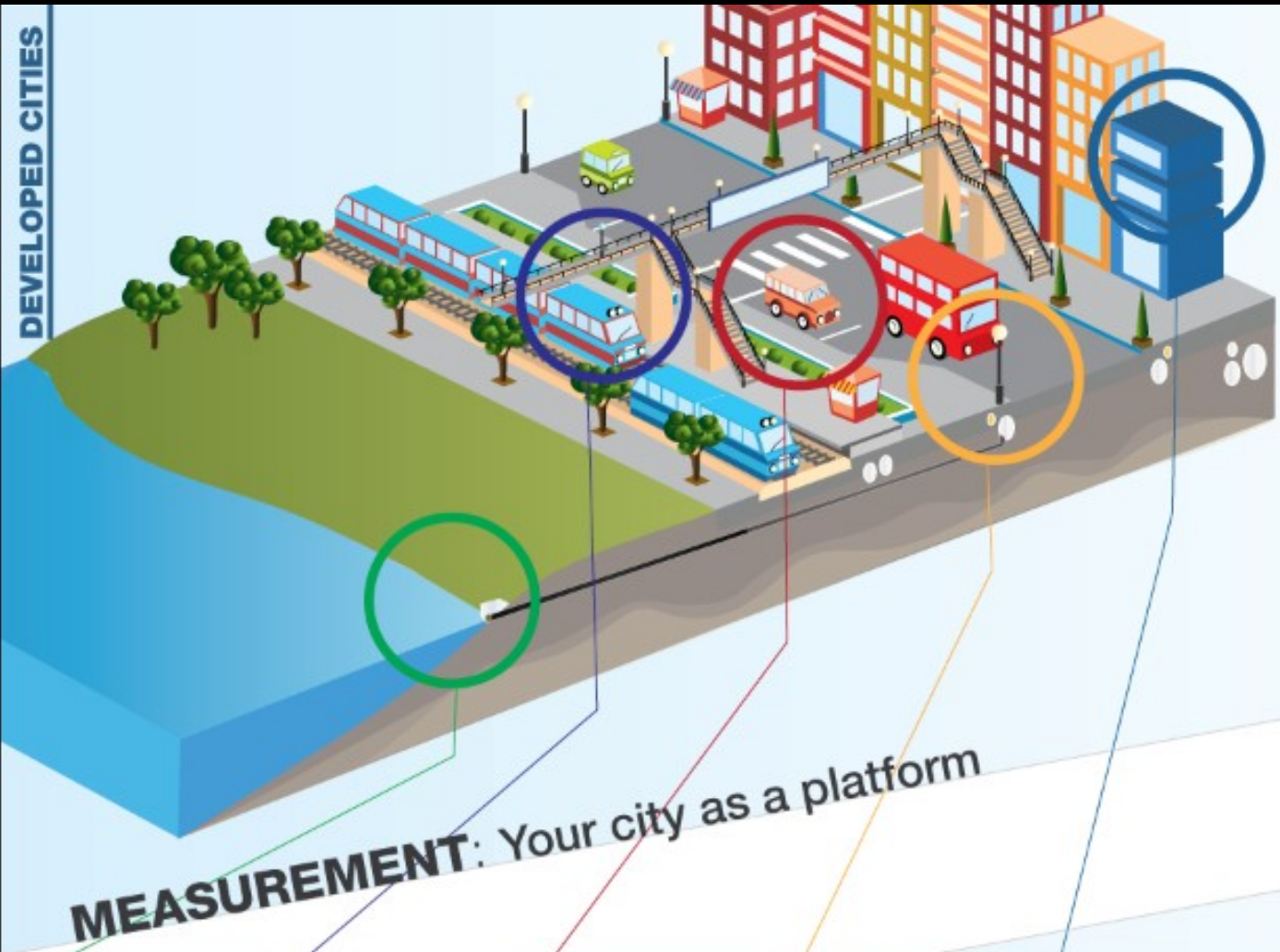
With the combination of low power sensors, wireless networks, and web and mobile-based applications, Smart Cities have arrived.

DEVELOPED CITIES













智慧城市主要基於感測器、網路與行動裝置造成了即時城市資訊系統

量測對象包括環境、安全、交通、電力、建築物



MEASUREMENT: Your city as a platform

ENVIRONMENT	SAFETY	TRANSPORATION	UTILITES	BUILDINGS
<p>By way of new sensor networks the accurate monitoring of environmental conditions like pollution levels, wildlife counts, and water runoff all become possible.</p>	<p>Structural Health Monitoring of buildings, bridges and dams as well as advanced warnings systems in emergency situations can now be put in place.</p>	<p>Through sensors embedded in roadways and street lights, real-time transit and traffic can be managed for the purpose of reducing travel time and fuel inefficiencies.</p>	<p>A smart utility grid will empower end users to be more aware of their energy uses, and allow utility companies to deliver only as much energy or water as is required.</p>	<p>Smart Buildings utilize monitoring devices that track usage and empower users and service providers to better control and reduce electricity demands.</p>
 <p> dontflush.me</p>	 <p> ガイガーマップ Jp Geiger Maps Jp</p>	 <p> STREETLINE CONNECTING THE REAL WORLD</p>	 <p></p>	 <p> nest</p>

OUTLOOK

FINANCIAL

The Smarter cities IT market opportunity is expected to be:

\$34 Billion
annually

CIVIC & ENVIRONMENTAL

Cisco estimates that cities that run on information can improve their energy efficiency by **30%** within 20 years.

YOU
+
SERVICES
+
NETWORKS
+
OPEN DATA
+
SENSORS

=

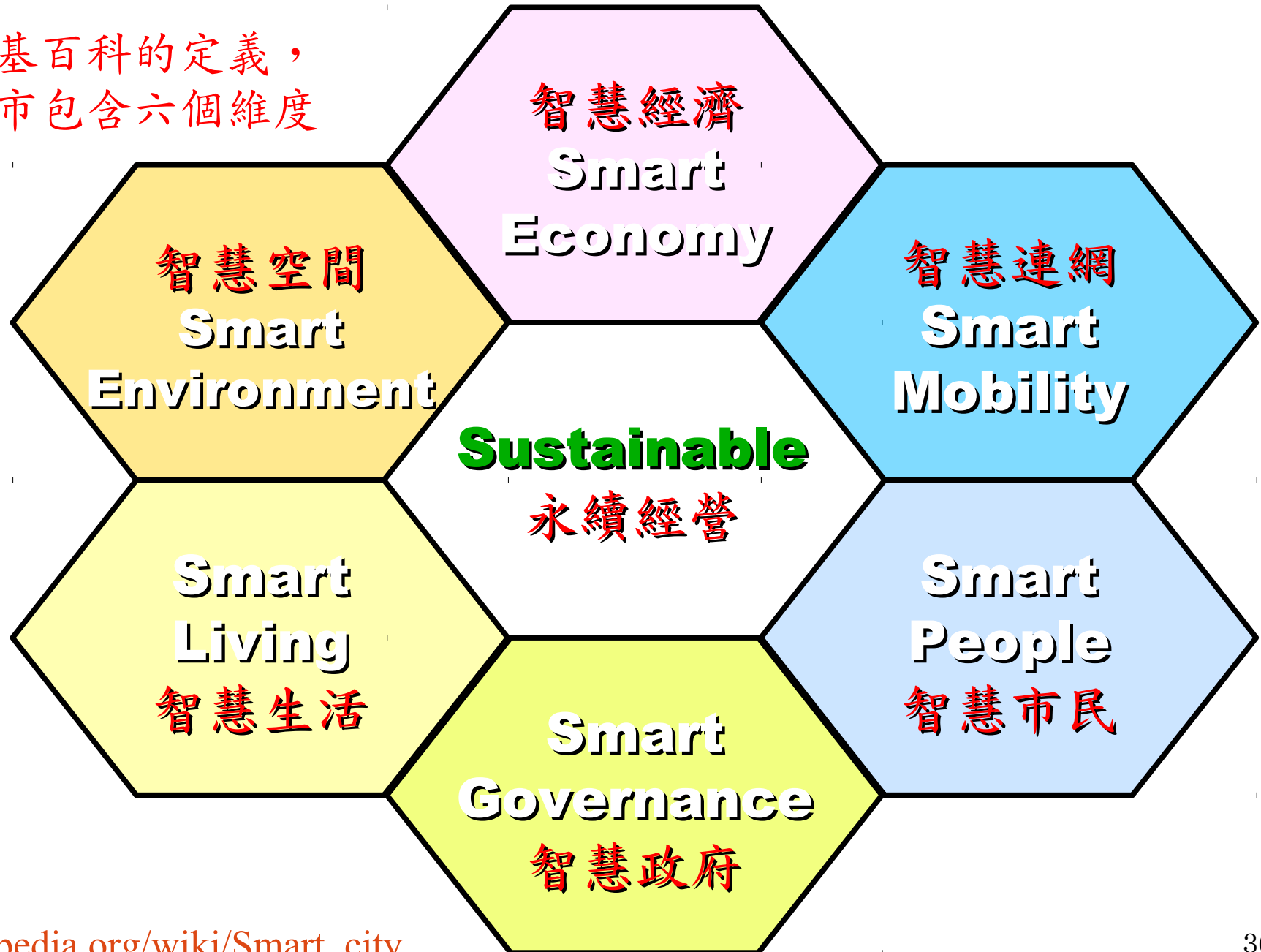
**ADAPTIVE
SENSEABLE
EFFICIENT**
Cities



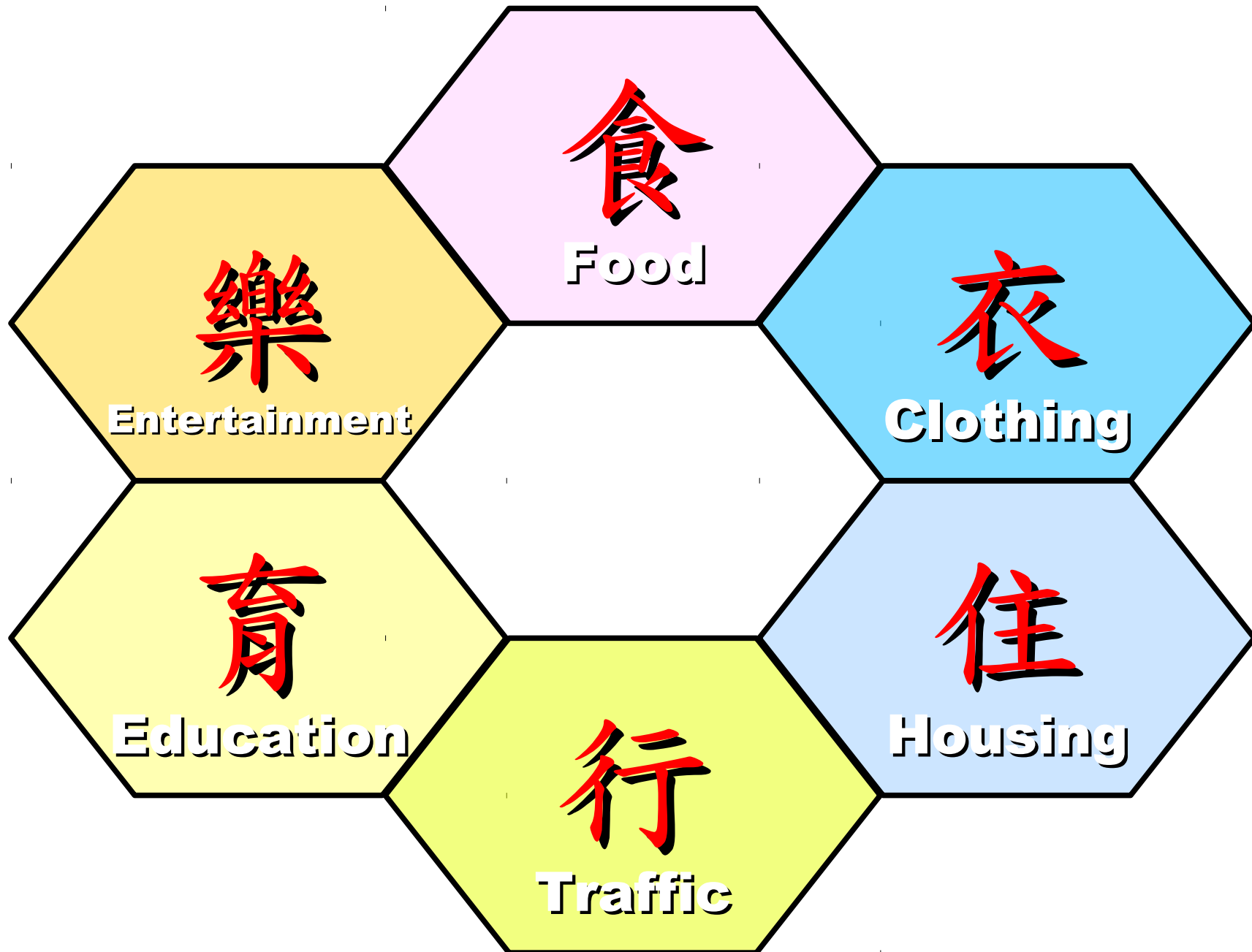
Postscapes™

智慧城市的六個維度 6D of Smart Cities

根據維基百科的定義，
智慧城市包含六個維度



讓人民有感的智慧城市 Ubiquitous Computing



幫忙您找食物的地理資訊系統服務

愛評首頁 | 娛樂藝文 | 口碑券 | 好康iCoupon | 愛評購物 | 精選情報 | 更多

早安, 訪客您好 | 登入 | 登入 | 登入 | 免費註冊會員 | NEW!! 最新分類

愛評地圖快搜



捷運美食



校園美食



高鐵美食



愛評生活通

馬上下載

主題地圖 標的列表 路徑規畫 展開搜尋列

在台中市(地址)附近, 符合"美食(分類)"共約 144 項結果, 以下為第 1-10 項結果。

排序方式: 綜合評價 | 評論數

女兒紅婚宴會館 - 顯示地圖位置>

- 綜合評鑑: ★★★★★
- 電話: 04-36006000
- 地址: 台中市南屯區文心南路99號

3篇分享

發表文章 | 想吃收藏它 | 詳細商家資訊>

COSTCO好市多 - 顯示地圖位置>

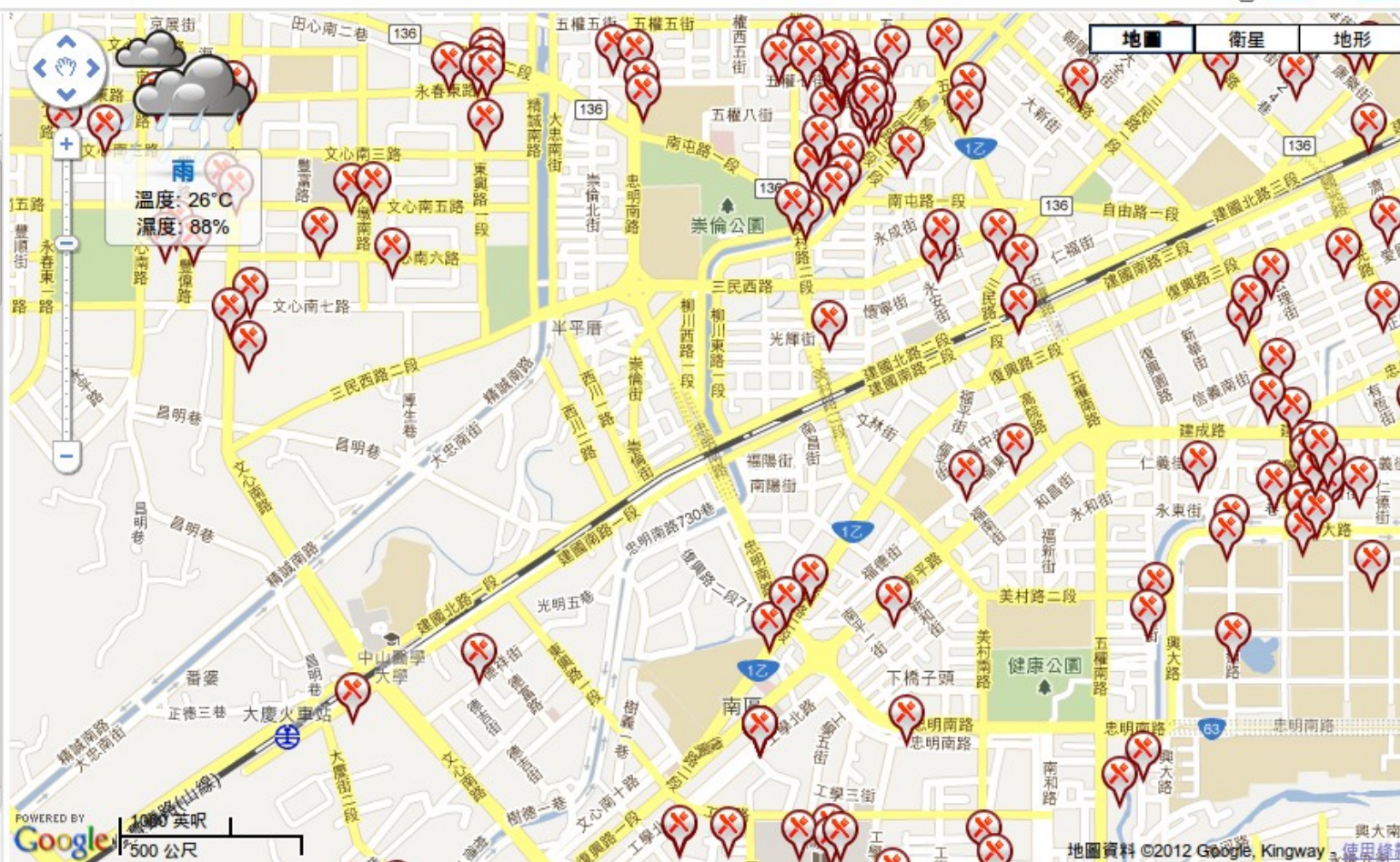
- 綜合評鑑: ★★★★★
- 電話: 04-37042855
- 地址: 台中市南屯區文心南三路289號

94篇分享

發表文章 | 想吃收藏它 | 詳細商家資訊>

泰鼎私房料理 - 顯示地圖位置>

- 綜合評鑑: ★★★★★
- 電話: 04-23780972
- 地址: 台中市西區五權五街97號



<http://www.ipeen.com.tw/map/#loc=台中市>

幫忙您決定該怎麼穿衣服的氣象服務

中區 預測明日 (8/19) 0:00

天氣狀況	風向風級	相對濕度
逐時	明天	後天
一週	① 3 6 9 12 15 18 21	

預報新鮮事 辦活動看天氣

回到全臺



- 基隆市
- 新北市
- 臺北市
- 桃園縣
- 新竹市
- 新竹縣
- 苗栗縣
- 臺中市
- 彰化縣
- 南投縣
- 雲林縣
- 嘉義市
- 嘉義縣
- 臺南市
- 高雄市
- 屏東縣
- 宜蘭縣
- 花蓮縣
- 臺東縣
- 澎湖縣
- 金門縣
- 連江縣

比例尺: 1 2 3 4 GO

<http://www.cwb.gov.tw/township/>

幫忙您找房子的地理資訊服務

591房屋交易
.com.tw
家，輕鬆自由找到

您好，歡迎來到591！會員登入

新使用者？免費加入會員

收費標準

刊登流程

| 配對電子報 | 591地圖 | 加入最愛

在線人數：8467人，租屋售屋：1534557筆(已成交440438筆)

首頁 租屋 中古屋 新房屋 頂讓 找經紀人 裝潢設計 搬家 居家/家具 討論區 會員中心

幫助

關鍵字搜尋：

請輸入街道或地標名

搜尋

縣市

捷運

台中市

西屯區 北區 南區

西區 北屯區 南屯區

東區 中區 大里區

太平區 沙鹿區 更多>>

篩選條件

選擇用途 >>

選擇格局 >>

選擇租金 >>

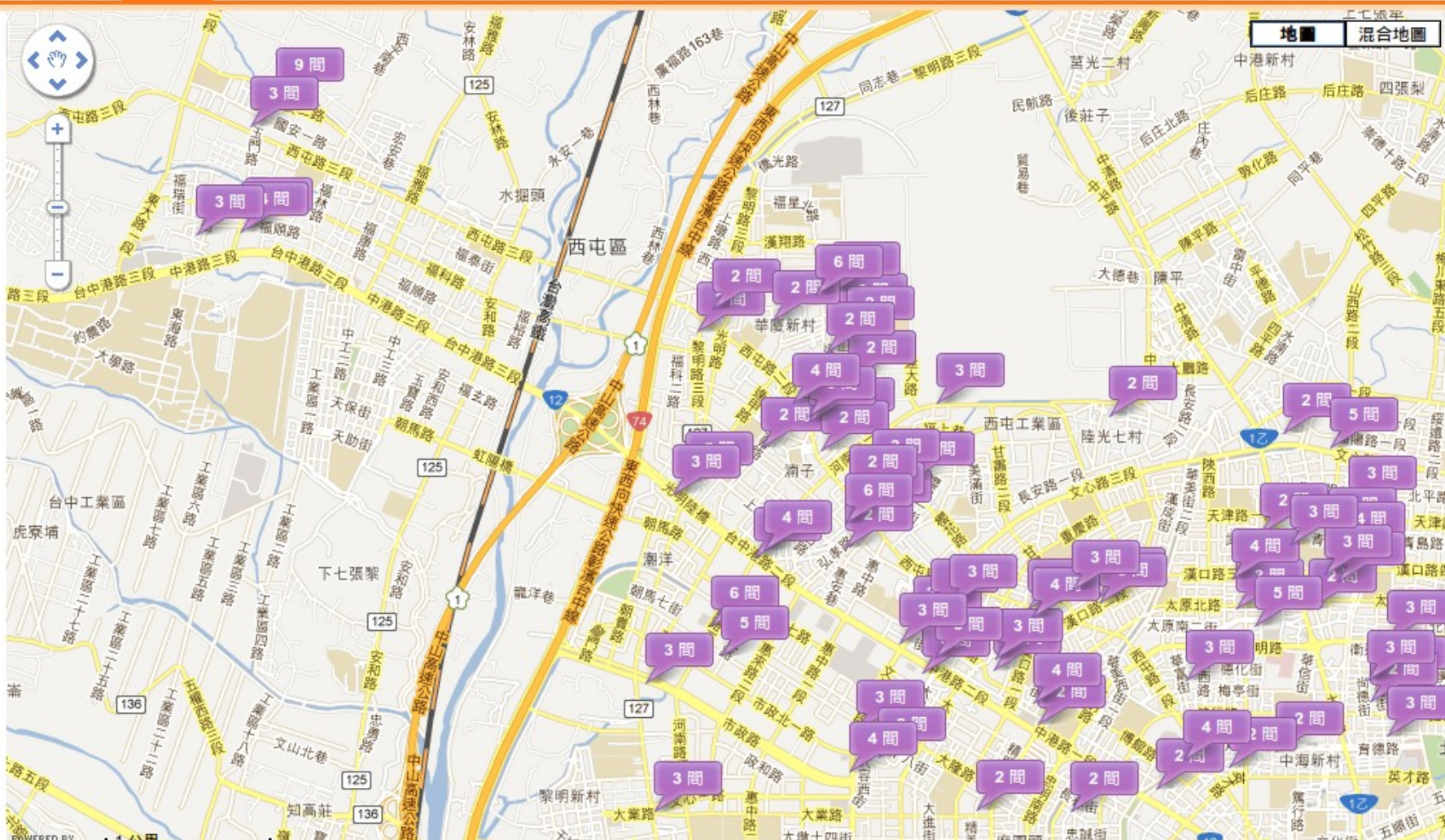
選擇坪數 >>

選擇型態 >>

房屋編號

搜尋

切換到縣市搜尋



幫忙您查即時交通路況的地理資訊服務



<http://1968.freeway.gov.tw/>

用地理資訊服務學習一個城市的歷史

The screenshot displays the hypercities website interface. At the top, a timeline slider ranges from 1700 to 2100. Below it, a satellite map of Tokyo is shown, with a historical map overlaid. The historical map is titled "RESTRICTED" and shows the city's layout in 1880. The interface includes a search bar, a "WORLD" button, and a "LOGIN" button. A sidebar on the right lists various historical maps and collections, including "1880 Rittau", "1900 Map of Tokyo (1900)", "1922 Map of Tokyo (1922)", "1930 Map of Tokyo (1930)", "1944 Japan Railroads", and "1944 Japan Roads And Water". The bottom of the interface features the hypercities logo and the text "Imagery ©2012 TerraMetrics - 使用條款".

WORLD SEARCH LOGIN HELP

POWERED BY Google 使用條款

automatically refresh maps and collections

MAP (11) COLLECTION

1880 Rittau
Rittau

1900 Map of Tokyo (1900)
Map of Tokyo (1900)

1922 Map of Tokyo (1922)
Map of Tokyo (1922)

1930 Map of Tokyo (1930)
Map of Tokyo (1930)

1944 Japan Railroads
Japan Railroads

1944 Japan Roads And Water
Japan Roads And Water Routes

TOKYO 1880 - 2012

hypercities
Imagery ©2012 TerraMetrics - 使用條款

<http://hypercities.ats.ucla.edu/>

Agenda 演講大綱

From **WSN** to **IoT**?

物聯網的崛起

What is **Smart City** ?

淺談智慧城市

What is **Open Data** ?

何謂開放資料

What is **Big Data** ?

何謂海量資料

Join Communities !

加入國內社群

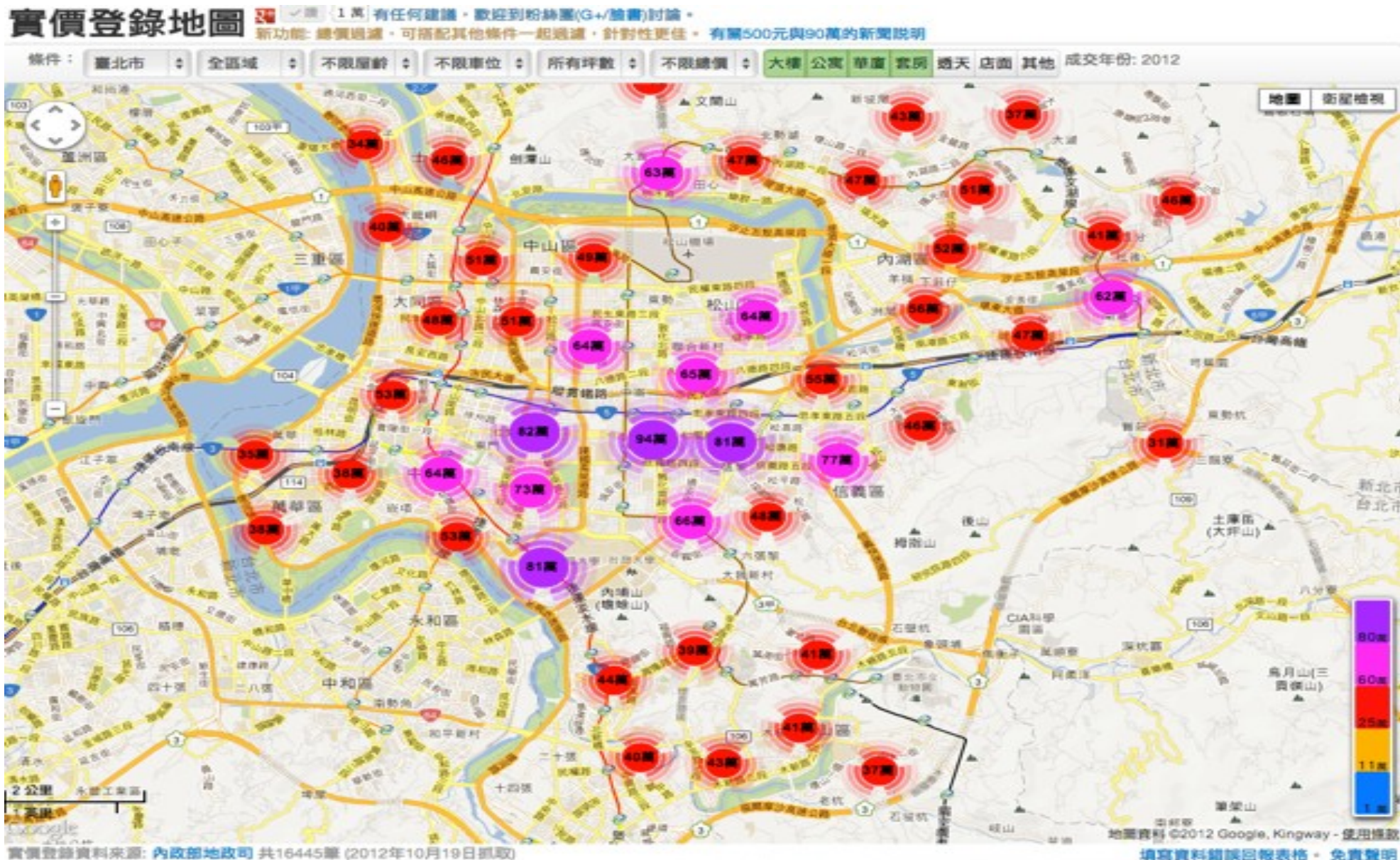
開放資料的新趨勢 Open Data

The screenshot shows the homepage of data.gov.uk. At the top left is the HM Government logo. The main header features the data.gov.uk logo with the tagline 'Opening up Government' and a 'Beta' label. A navigation menu includes links for Home, Data, Participate, Apps, Location, Linked Data, Library, Lab, and About. The main content area is a carousel with several featured items: a weather data download announcement, a blog post by Professor Brian Collins, a blog post by Tim Davies, an open data case studies section, and an open data white paper announcement. The white paper section is highlighted with a green background and the text 'OPEN DATA WHITE PAPER Unleashing the potential' and 'Read the UK Government's vision for the future of open data'. A search bar is visible at the bottom left of the page.

<http://data.gov.uk/>

開放資料的實例 Open Data

<http://real-price.appspot.com/>



開放資料的民間推動 Code for Tomorrow

Code for Tomorrow

Code for Tomorrow 將會是一個推動各種「寫程式改造社會」計劃的組織，也希望能夠成為一個媒合（供需雙方）和孵育（計畫、技術和知識）的平台。



關於我們 · 計畫 · 活動

ARCHIVE · RSS

Open Data Day 2013 台北參加者

背景 England Japan



感謝偉大的城市/程式設計師



25.media.tumblr.com/f49e749ce5e353895353e0e16850fd33/tumblr_miz0dgd5i01ribftko1_1280.png

<http://codefortomorrow.org/>

我個人的開放資料慘痛經驗 Ocean Database

台灣海洋資訊資料庫聯盟網 Taiwan Ocean Information Database - Mozilla Firefox

http://oceandb.info/oid/

HOME | MENU | NEWS | ABOUT

台灣海洋資訊資料庫聯盟網 Taiwan Ocean Information Database

地圖 衛星 混合地圖 地球

- ▶ 波浪 (定點)
 - 中央氣象局
 - 中央氣象局委託
 - 交通部觀光局
 - 港灣技術研究中心
 - 經濟部水利署
- ▶ 潮汐 (定點)
- ▶ 氣象 (定點)
- ▶ 海流 (定點)
- ▶ 溫鹽 (船測)
- ▶ 震測 (船測)
- ▶ 海流 (船測)
- ▶ 水深 (船測)
- ▶ Chirp Sonar (船測)

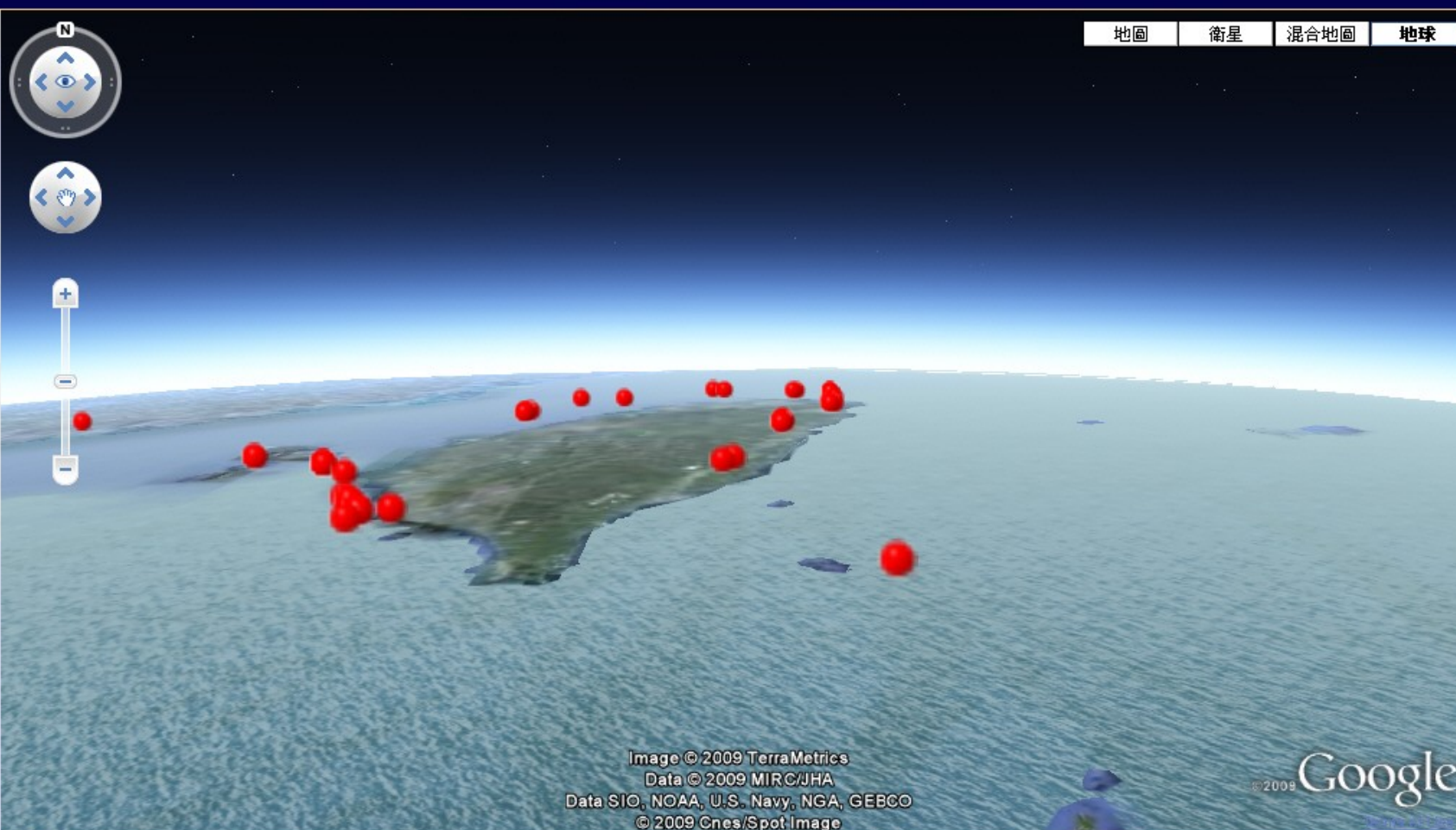


Image © 2009 TerraMetrics
Data © 2009 MIRC/JHA
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
© 2009 Cnes/Spot Image

©2009 Google

Copyright 2008 TORI & NCHC, NARL, Taiwan

台灣海洋科技研究中心 Taiwan Ocean Research Institute TORI

國家海洋科技研究中心 National Center for High Performance Computing

完成

<http://oceandb.info>

Agenda 演講大綱

From **WSN** to **IoT**?

物聯網的崛起

What is **Smart City** ?

淺談智慧城市

What is **Open Data** ?

何謂開放資料

What is **Big Data** ?

何謂海量資料

Join Communities !

加入國內社群

Why are they named by “ SMART ” ?!

智慧打哪兒來？！

Smart Phone

智慧手機

Smart Car

智慧車輛

Smart Grid

智慧電網

SMART ?

哪裡長
智慧了？

Smart City

智慧城市

Smart Home

智慧家庭

Smart Meter

智慧電錶

資料

Data

知識

Knowledge

智慧

Wisdom

Can Machine understand You? 讓機器更懂你？

iPhone

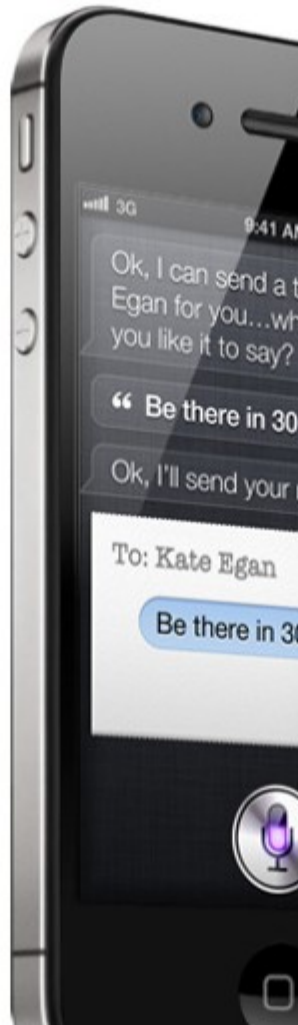
Features Built-in Apps



Siri. Beta

Your wish is
its command.

Siri on iPhone 4S lets you use your voice to send messages, schedule meetings, place phone calls, and more. Ask Siri to do things just by talking the way you talk. Siri understands what you say, knows what you mean, and even talks back. Siri is so easy to use and does so much, you'll keep finding more and more ways to use it.



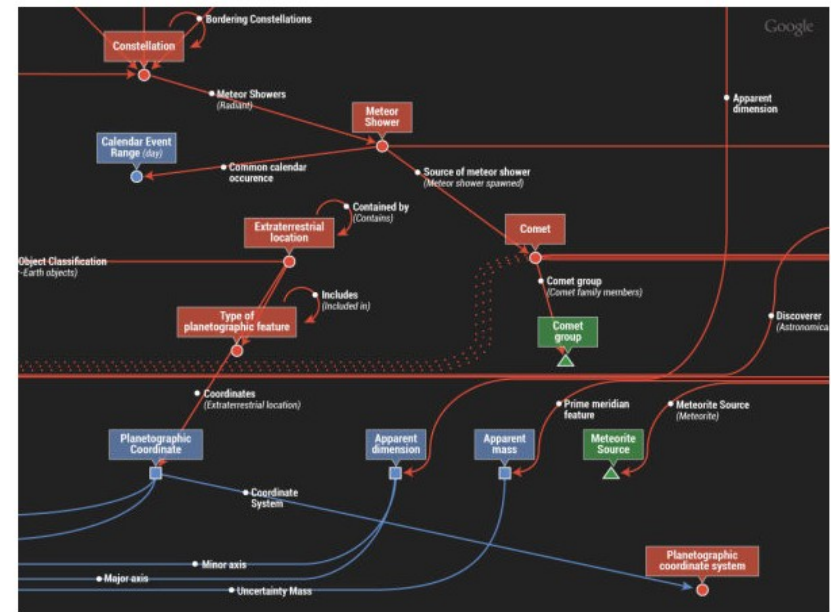
Google將發展「人工智慧」 永久改變搜尋引擎

2012年02月15日 00:11

點評：超級阿斯拉，衝啊！（阿斯拉：好的，準人！）

記者黃郁棋 / 綜合報導

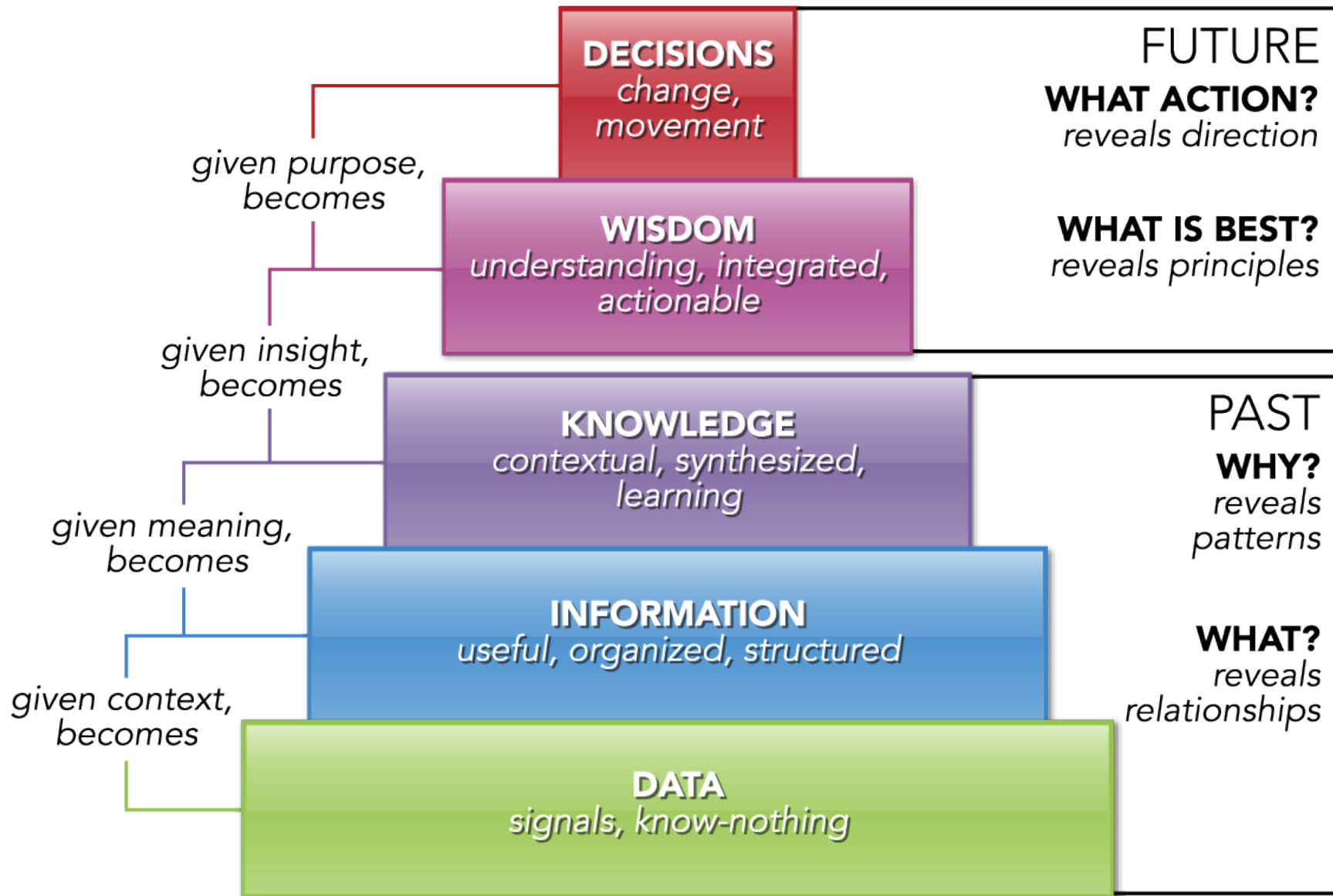
每個人都再猜，下一波網路革命是什麼？每個人都再猜，未來的世界會如何運作？Google的資深副總Amit Singhai透露了一點訊息。「Google正努力從『單字』層面進展到『意義』層面，未來搜尋引擎提供的不只是關鍵字搜尋，搜尋引擎甚至會『明白』你到底要什麼。」



▲ Google未來將會朝「人工智慧」前進。(圖 / 取自mashable.com)

Knowledge is from the Past, Wisdom aim for Future

知識源自於過去，智慧在能預測未來



What is Big Data?! 何謂『海量資料』？

海量資料泛指資料大小已無法用一般軟體擷取、管理與處理；
單一資料集大小介於數十 TB 至數 PB 的資料。

'Big Data' = few dozen TeraBytes to PetaBytes in single data set.

Definition

[edit]

Big data is a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single data set.

In a 2001 research report^[14] and related conference presentations, then META Group (now Gartner) analyst, Doug Laney, defined data growth challenges (and opportunities) as being three-dimensional, i.e. increasing volume (amount of data), velocity (speed of data in/out), and variety (range of data types, sources). Gartner continues to use this model for describing big data.^[15]

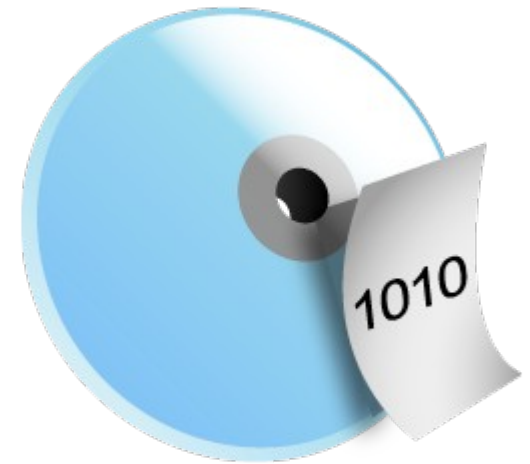
出處：http://en.wikipedia.org/wiki/Big_data



多個檔案，容量 100TB



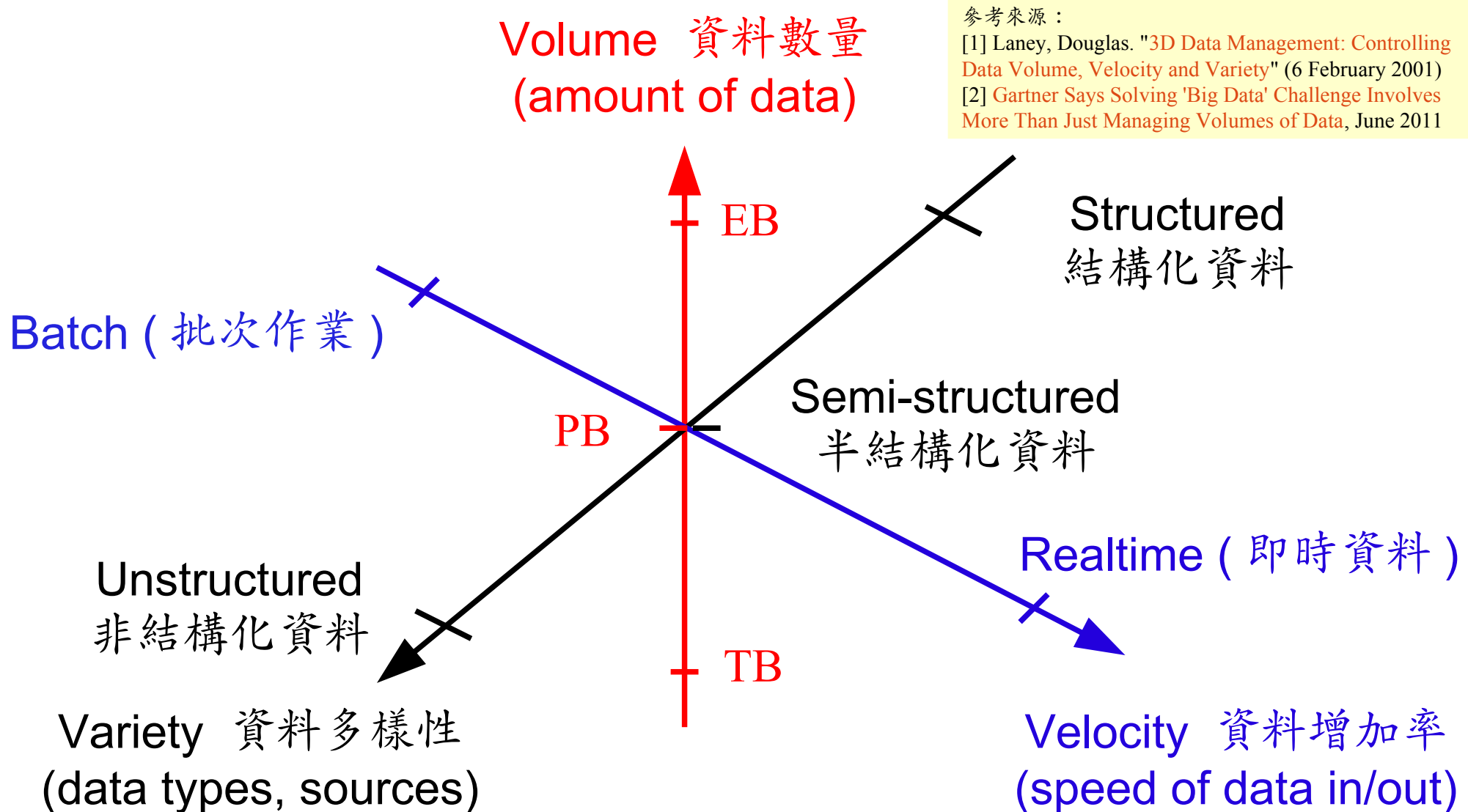
一個資料庫，容量 100TB



一個檔案，容量 100TB

Gartner Big Data Model? 海量資料的模型?

海量資料的挑戰在於如何管理「數量」、「增加率」與「多樣性」



參考來源:

[1] Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety" (6 February 2001)

[2] Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data, June 2011

12D of Information Management? 12 個維度?



Big Data
只是終極
資訊管理
的開端！

Source: Gartner (March 2011), 'Big Data' Is Only the Beginning of Extreme Information Management, 7 April 2011, <http://www.gartner.com/id=1622715>

New IT Architecture toward Cloud Computing !!

用雲掌握資料，加以分析，形成智能給端用



雲

資料中心
提供服務

雲端設計新思維：端的智能來自於雲的服務

Devices share the wisdom of Cloud

端



各類裝置
存取服務

4. 下載最佳路徑

1. 上傳行進軌跡

6. 上傳最佳路徑執行狀態
(判斷是否需要修正)

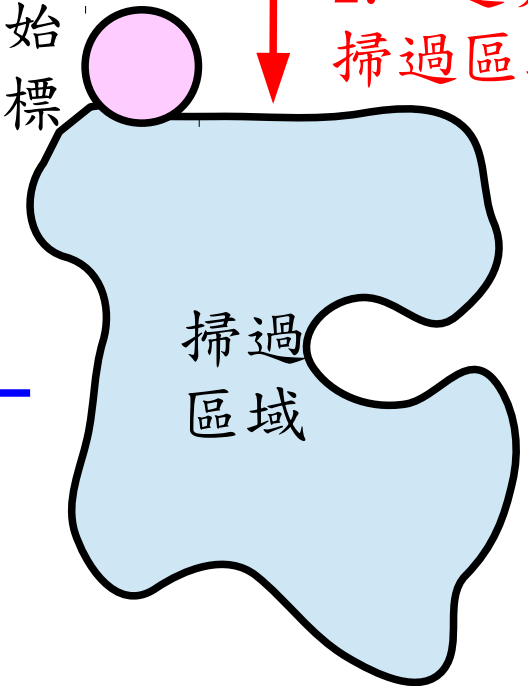
3. 運算
最佳路徑

5. 執行
最佳路徑



起始
座標

2. 運算
掃過區域



曾經有人問道：我是學電機的，那電機與控制跟雲端有甚麼關聯性？

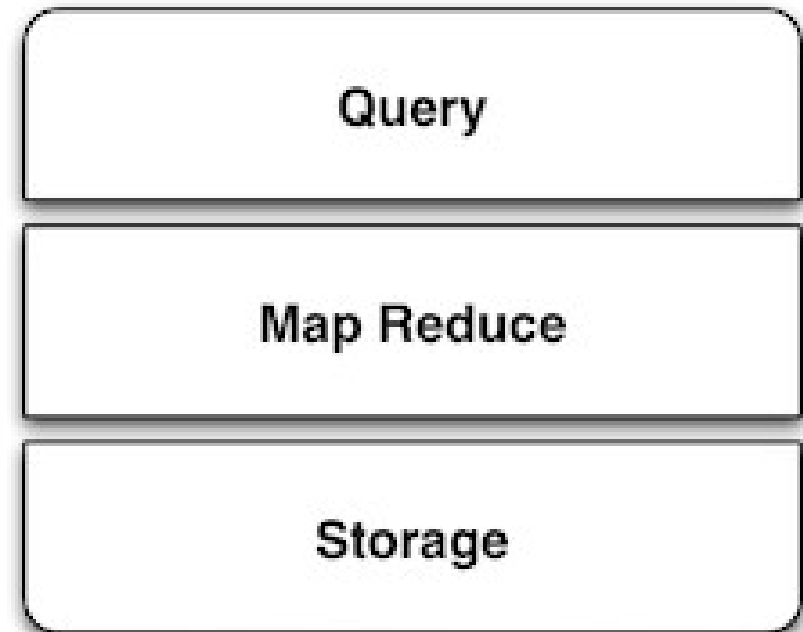
The SMAQ stack for big data

海量資料處理的資訊架構

做網頁相關的人可能聽過 LAMP



未來處理海量資料的人必需知道
SMAQ (Storage, MapReduce and Query)



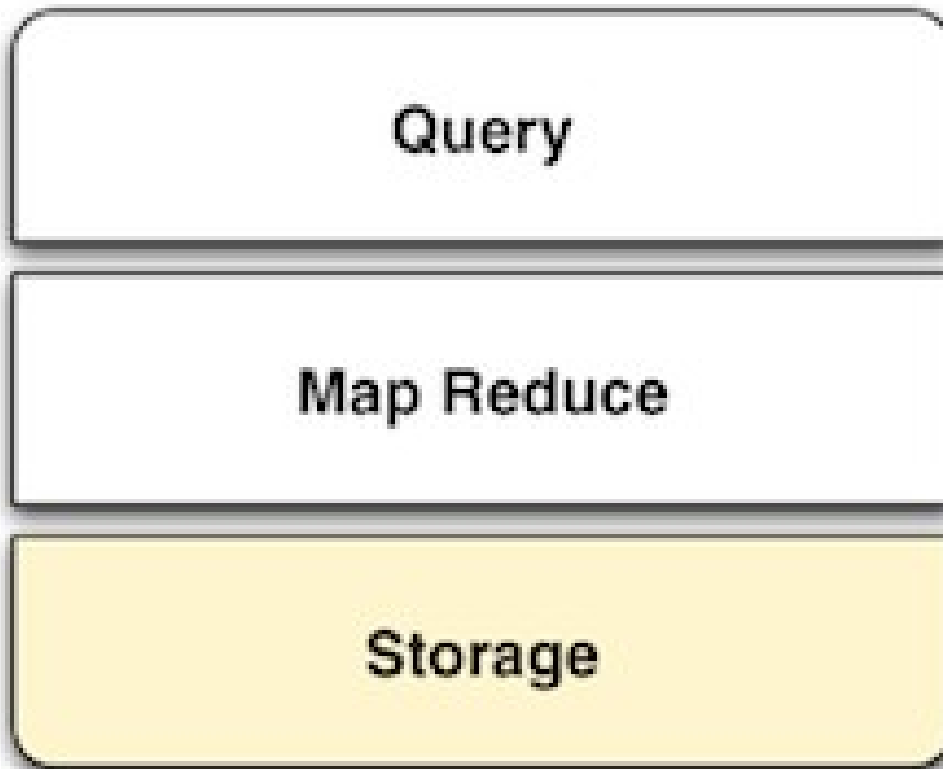
參考來源：The SMAQ stack for big data，Edd Dumbill，22 September 2010，

<http://radar.oreilly.com/2010/09/the-smaq-stack-for-big-data.html>

圖片來源：<http://smashingweb.ge6.org/wp-content/uploads/2011/10/apache-php-mysql-ubuntu.png> 57

The SMAQ stack for big data

海量資料處理的資訊架構



用來儲存分散、沒有關聯
的非結構化資料

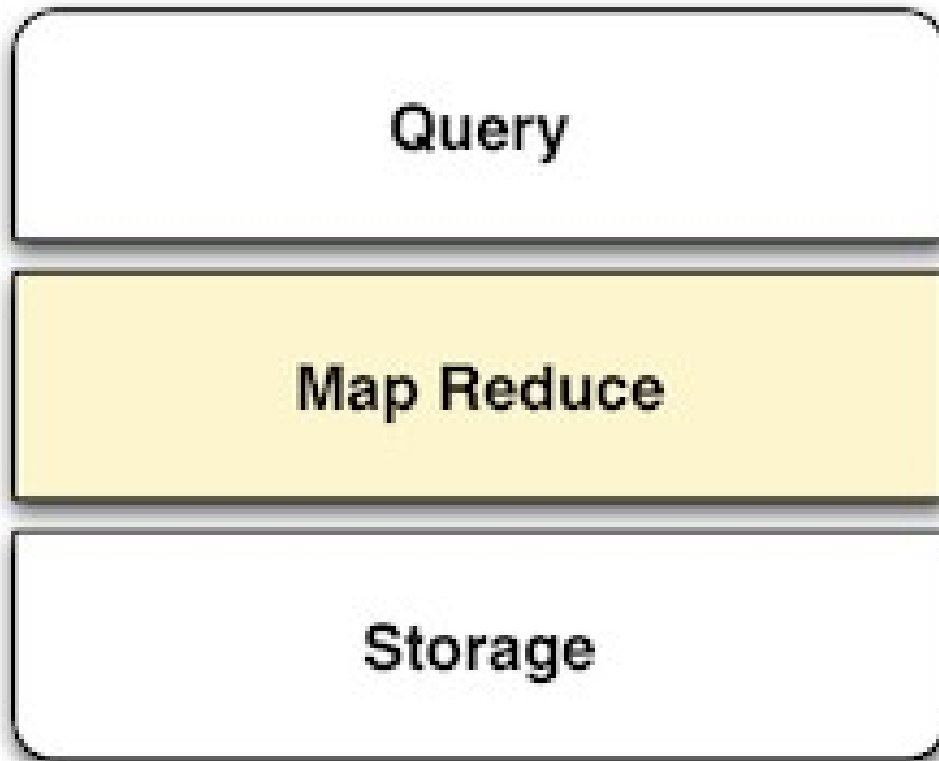
Key features

- Distributed
- Non-relational or unstructured

The SMAQ stack for big data

海量資料處理的資訊架構

運用批次處理的方式，將
運算工作平均分散到許多
的伺服器做運算。

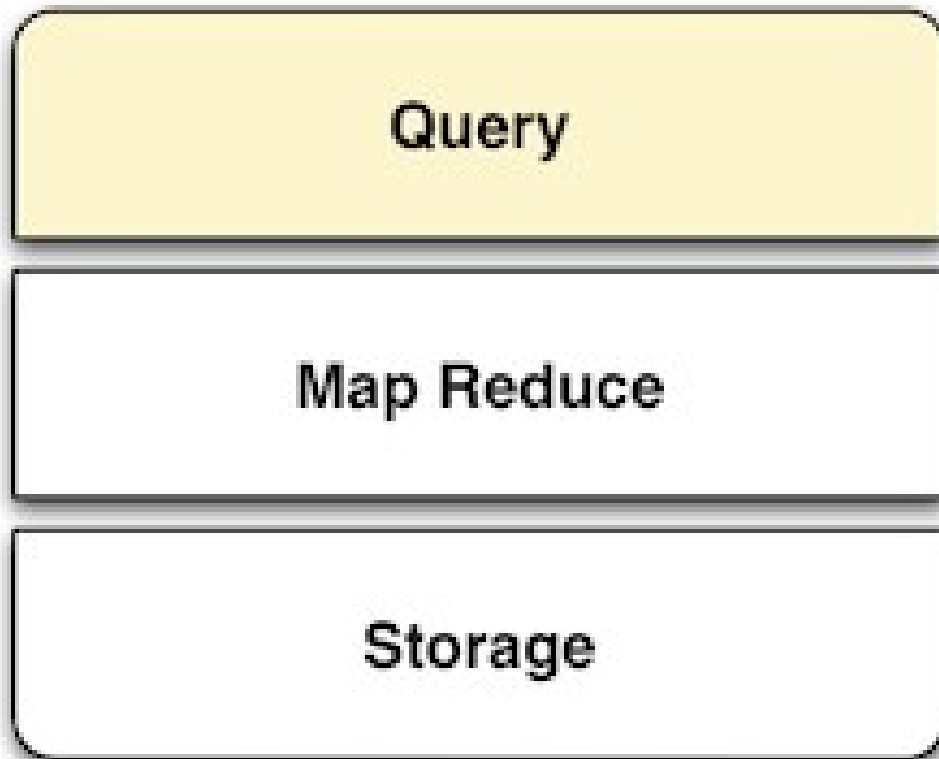


Key features

- Distributes computation over many servers
- Batch processing model

The SMAQ stack for big data

海量資料處理的資訊架構



Key features

- Efficient way of defining computation
- Platform for user friendly analytical systems

將算完的結構化資料儲存到可供查詢的資料庫系統

Three Core Technologies of Google

Google 的三大關鍵技術

- Google 在一些會議分享他們的三大關鍵技術
- Google shared their design of web-search engine
 - SOSP 2003 :
 - “The Google File System”
 - <http://labs.google.com/papers/gfs.html>
 - OSDI 2004 :
 - “MapReduce : Simplified Data Processing on Large Cluster”
 - <http://labs.google.com/papers/mapreduce.html>
 - OSDI 2006 :
 - “Bigtable: A Distributed Storage System for Structured Data”
 - <http://labs.google.com/papers/bigtable-osdi06.pdf>



Open Source Mapping of Google Core Technologies

Google 三大關鍵技術對應的自由軟體

Google 三大關鍵技術

自由軟體對應解決方案

Q = Query
BigTable

A huge key-value datastore

HBase, Hypertable
Cassandra,

MapReduce

To parallel process data

Hadoop MapReduce API
Sphere MapReduce API, ...

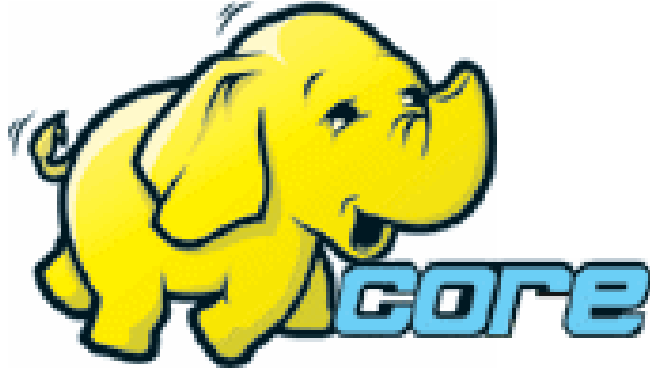
S = Storage

Google File System

To store petabytes of data

Hadoop Distributed File System (HDFS)
Sector Distributed File System

Hadoop

- <http://hadoop.apache.org>
 - Hadoop 是 Apache Top Level 開發專案
 - **Hadoop is Apache Top Level Project**
 - 目前主要由 Yahoo! 資助、開發與運用
 - **Major sponsor is Yahoo!**
 - 創始者是 Doug Cutting，參考 Google Filesystem
 - **Developed by Doug Cutting, Reference from Google Filesystem**
 - 以 Java 開發，提供 HDFS 與 MapReduce API。
 - **Written by Java, it provides HDFS and MapReduce API**
 - 2006 年使用在 Yahoo 內部服務中
 - **Used in Yahoo since year 2006**
 - 已佈署於上千個節點。
 - **It had been deploy to 4000+ nodes in Yahoo**
 - 處理 Petabyte 等級資料量。
 - **Design to process dataset in Petabyte**
- 
- Facebook、Last.fm
、Joost are also
powered by Hadoop**

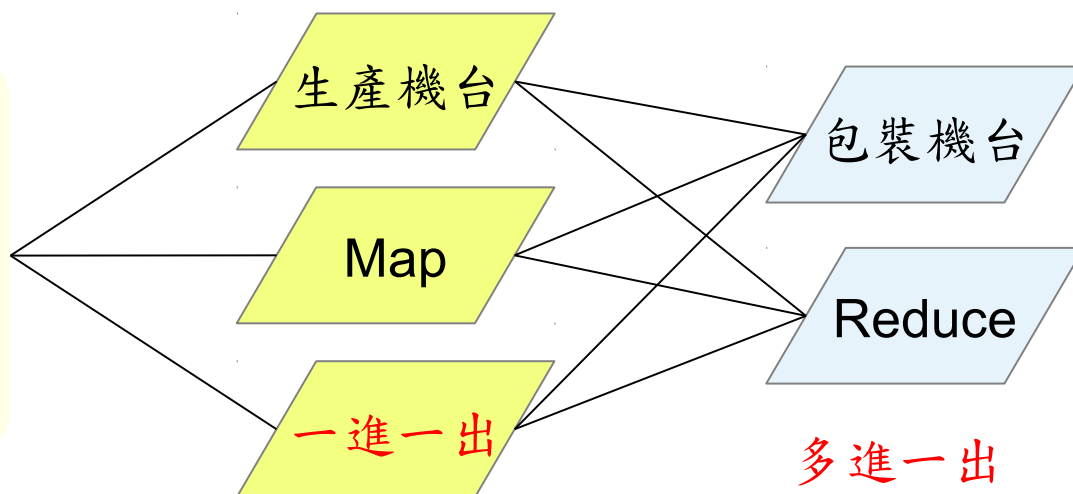
Hadoop 簡介

Hadoop 是一個讓使用者簡易撰寫並執行處理海量資料應用程式的軟體平台。

亦可以想像成一個處理海量資料的生產線，只須學會定義 **map** 跟 **reduce** 工作站該做哪些事情。

就像工廠的倉庫
存放生產原料跟待售貨物

HDFS 存放
待處理的**非結構化資料**
與處理後的**結構化資料**



Sector / Sphere

- <http://sector.sourceforge.net/>
- 由美國資料探勘中心研發的自由軟體專案。
- **Developed by National Center for Data Mining, USA**
- 採用 C/C++ 語言撰寫，因此效能較 Hadoop 更好。
- **Written by C/C++, so performance is better than Hadoop**
- 提供「類似」Google File System 與 MapReduce 的機制
- **Provide file system similar to Google File System and MapReduce API**
- 基於UDT高效率網路協定來加速資料傳輸效率
- **Based on UDT which enhance the network performance**
- Open Cloud Testbed有提供測試環境，並開發Ma1Stone效能評比軟體
- **Open Cloud Consortium provide Open Cloud Testbed and develop Ma1Stone toolkit for benchmark**

Sector-Sphere

National Center for Data Mining
University of Illinois at Chicago

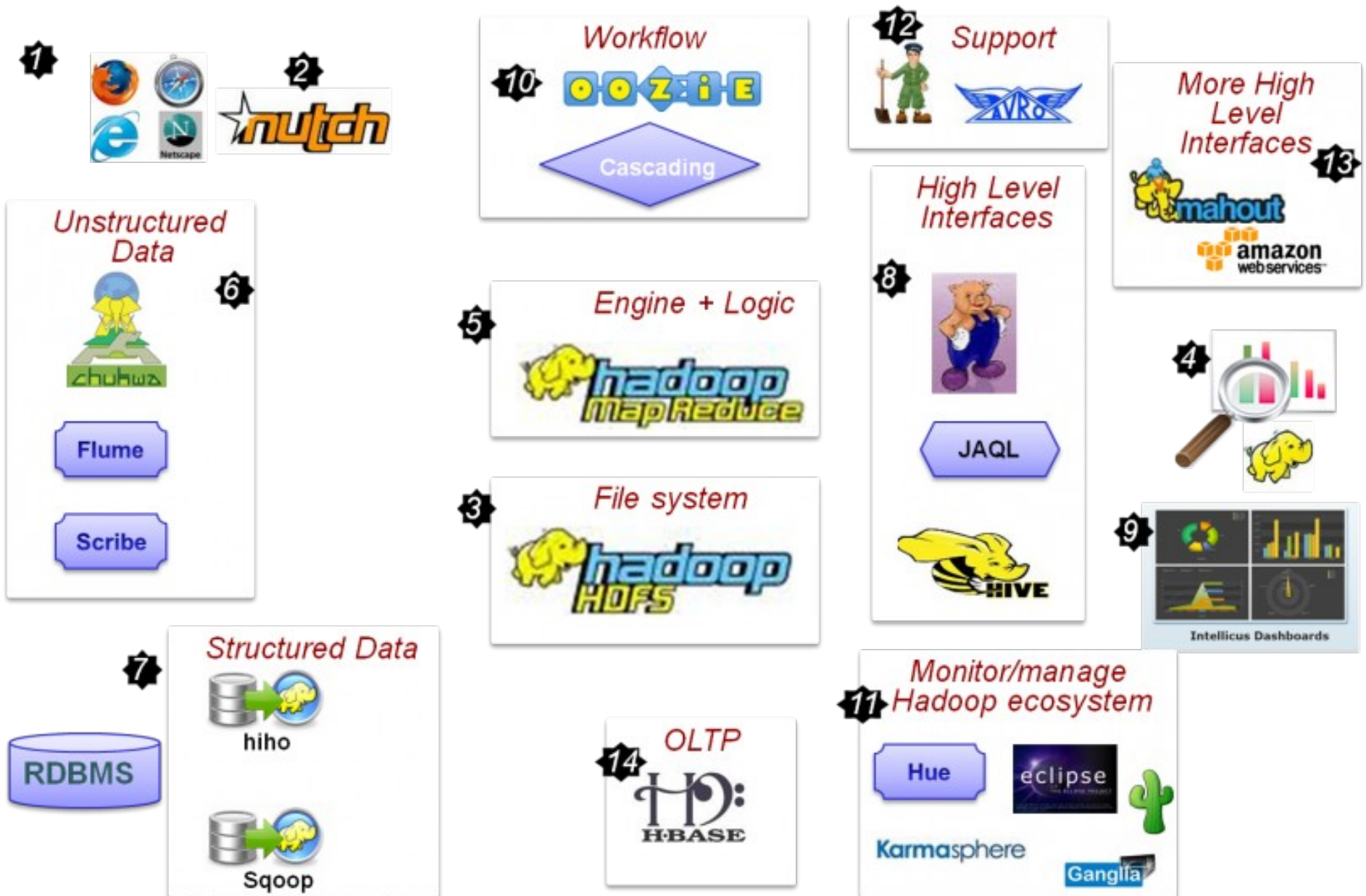


Open Data Group

<http://www.opendatagroup.com/>

Why we choice Hadoop? Good Ecosystem!

豐富的生態系建構出處理海量資料的工具庫



巨量資料的未來？ Future of Big Data ?

http://www.huffingtonpost.com/deb-roy/the-birth-of-a-word_b_2639625.html

TED WEEKENDS: Big Data Can Get Very Personal



Deb Roy ♥ Become a fan



Follow

Like

79

The Birth of a Word

Like

8.5k

share

3521

tweet

212

+1

296

email

1066

comment

222



DEB ROY
THE BIRTH OF A WORD

TED

3 years recorded
90,000 hrs video
140,000 hrs audio
200 terabytes

個人化的巨量資料

Big Data Can Get Very Personal

http://www.ted.com/talks/deb_roy_the_birth_of_a_word.html

Agenda 演講大綱

From **WSN** to **IoT**?

物聯網的崛起

What is **Smart City** ?

淺談智慧城市

What is **Open Data** ?

何謂開放資料

What is **Big Data** ?

何謂海量資料

Join **Communities** !

加入國內社群

【建議】

現在做研究也要打**團體戰**

加入技術討論社群

增加學習知識的觸角！

雲端產業供應鏈

Supply Chain of Cloud Computing

應用軟體
供應商



端

各類裝置
存取服務

行動裝置
共通平台



軟體服務
供應商



雲

資料中心
提供服務

資料中心
機房維運



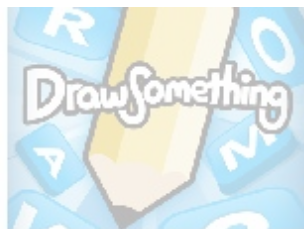
基本硬體
建設組件



雲端產業供應鏈

Supply Chain of Cloud Computing

應用軟體
供應商



端

行動裝置
共通平台

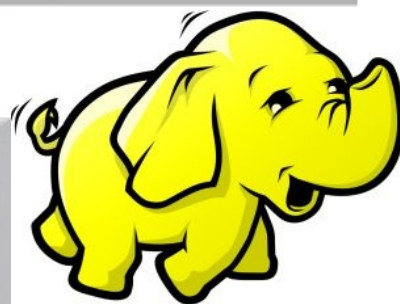


各類裝置
存取服務

軟體服務
供應商



大象擺這裡



資料中心
機房維運



資料中心
提供服務

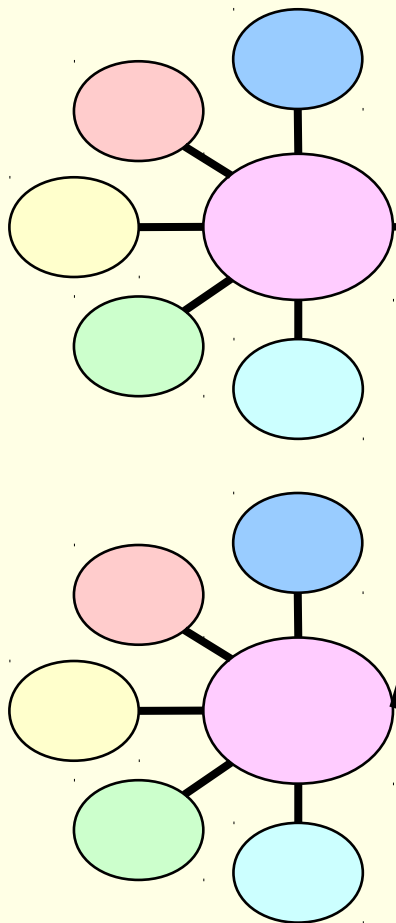
基本硬體
建設組件



巨量資料的奇幻漂流 Life of Big Data

Internet of Things
物聯網

Sensor Network



Smart Grid

雲 資料中心
提供服務

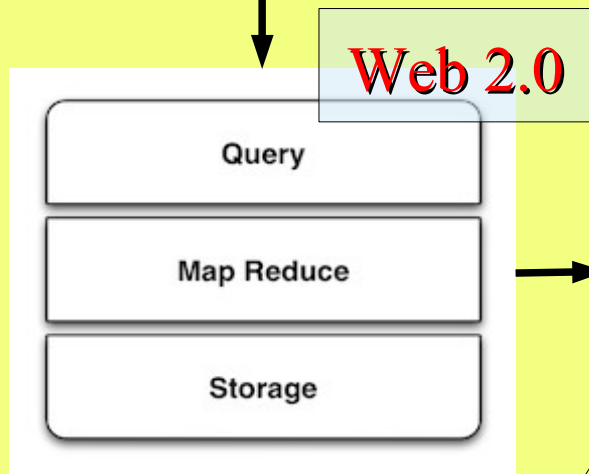


Public Data Hub
Data as a Service

開放資料
Open Data



Web 2.0



Big Data



Mobile Computing

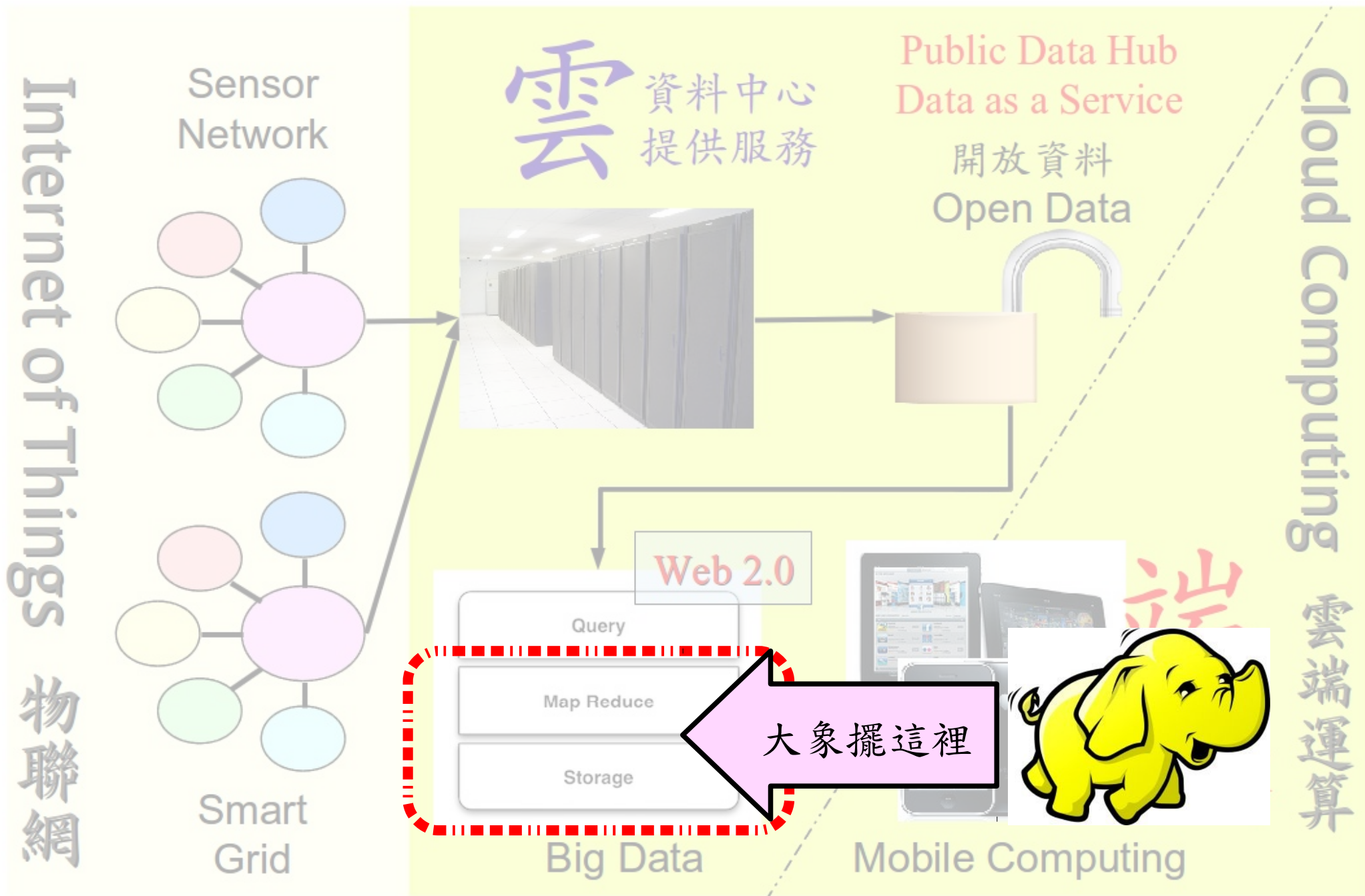
端
各類裝置
存取服務

Cloud Computing

雲端運算

巨量資料的社群關聯

Relations between Communities



Hadoop related Facebook Group in Taiwan

Hadoop.TW

<https://www.facebook.com/groups/hadoop.tw>

The screenshot shows the Facebook interface for the "Taiwan Hadoop User Group". The page header includes the group name, navigation icons, and a home button. Below the header is a banner image featuring several photos of group members. The main content area includes a "Write Post" section with a text input field, and a "RECENT POSTS" section. The first post is by Jazz Yao-Tsung Wang, dated 2013/03/22, with the text: "這幾天花了一些時間研究 Facebook Graph API, 想要了解國內 Hadoop 相關社群的關聯性。目前的狀態如附圖。" Below the text is a network diagram showing connections between various groups, with "Hadoop.TW" at the center. The right sidebar shows "533 members (9 new) · Invite by Email" and a "People You May Know" section featuring Steven Chin with 4 mutual friends. The left sidebar lists navigation options like "News Feed", "Messages", "Events", "Photos", and "Browse", as well as "ADS" and "PAGES".

Taiwan Hadoop User Group About Events Photos Files Notifications

Write Post Add Photo / Video Ask Question Add File

Write something...

533 members (9 new) · Invite by Email

+ Add People to Group

RECENT POSTS

Jazz Yao-Tsung Wang
這幾天花了一些時間研究 Facebook Graph API, 想要了解國內 Hadoop 相關社群的關聯性。目前的狀態如附圖。

People You May Know See All

Steven Chin
4 mutual friends
Add Friend

PAGES

- Hadoop in Taiwan 2012
- Pages Feed 20+
- Like Pages 1

GROUPS

- JavaScript.tw 1
- Taiwan Hadoop User...**
- S4A 3

2013/03/22 Hadoop.TW 533 members

Hadoop related Facebook Group in Taiwan

Hadoop in Taiwan

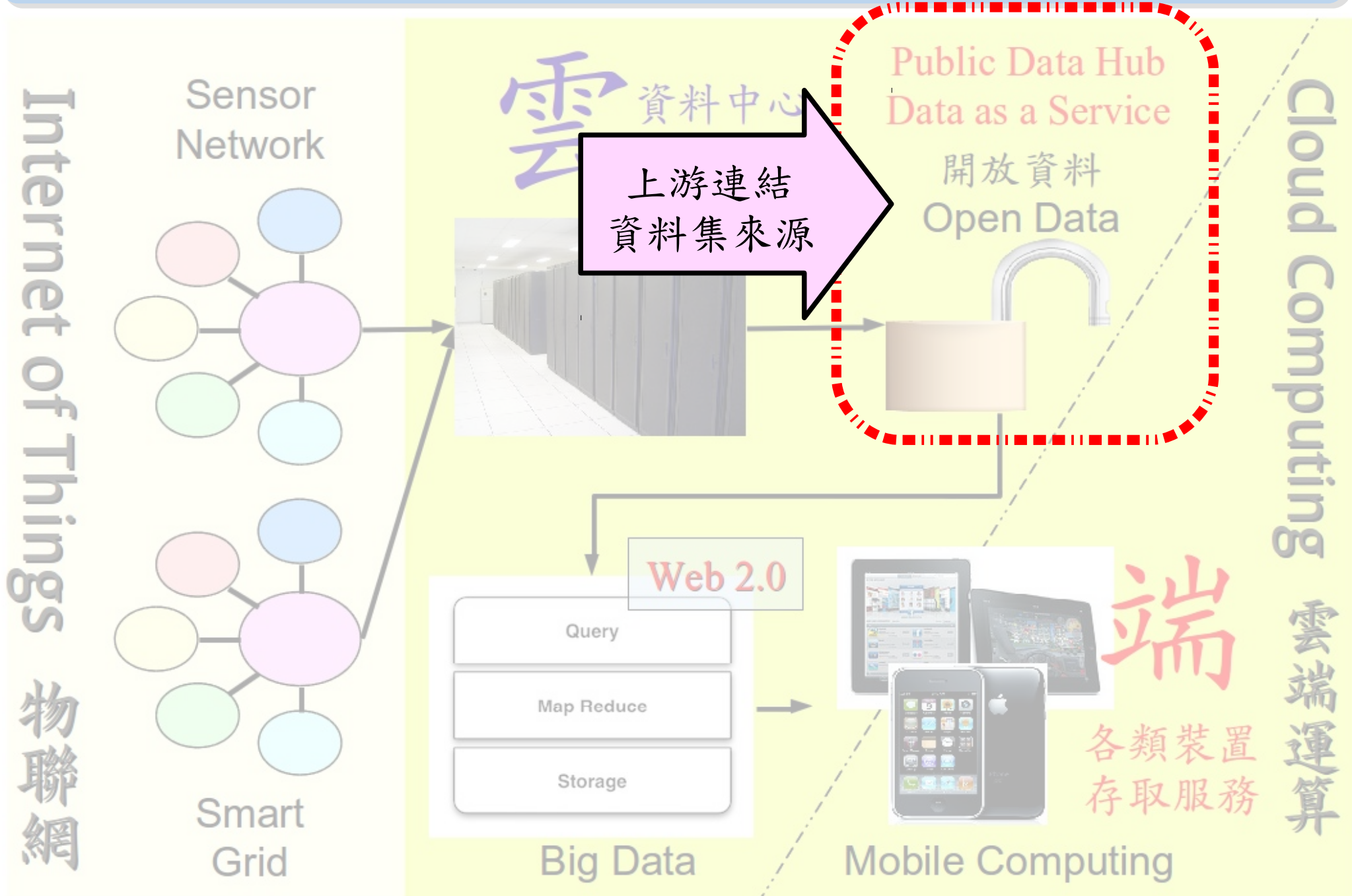
<https://www.facebook.com/groups/hadoopintaiwan/>

The screenshot shows the Facebook group page for 'Hadoop in Taiwan'. The page header includes the group name, navigation tabs (About, Events, Photos, Files), and a notification bell. Below the header is a row of profile pictures of group members. The main content area features a 'Write Post' section with a text input field and buttons for 'Add Photo / Video', 'Ask Question', and 'Add File'. To the right, it shows '59 members (2 new) · Invite by Email' and a '+ Add People to Group' button. The 'RECENT POSTS' section displays a post by Jazz Yao-Tsung Wang, dated 2013/03/22, with the text: '有鑑於今年 Big Data 熱潮持續發酵，Hadoop 社群逐漸茁壯，首先感謝去年趨勢科技的 Rice, 昆錡, Chris Huang 等建議與支持，今年開始舉辦每季一次的 Hadoop 生態系工作坊。其次要特別感謝 Etu 團隊 Fred Chiang 大的鼓勵跟建議，社群必須要開始讓更多人可以擁有強烈的參與感跟使命感。所以在此開始徵求以下工作小組：' followed by '(1) Hadoop 生態系工作坊的報到引導小組 x 2 名 (我們先邀集 3/10 可以幫忙的)'. The left sidebar contains navigation options like 'News Feed', 'Messages', 'Events', 'Photos', 'Browse', 'Ads Manager', 'Pages', and 'Groups'.

2013/03/22 Hadoop in Taiwan 59 members

巨量資料的社群關聯

Relations between Communities



Open Data Taiwan Facebook Group

OpenData / Taiwan (ODTWN)

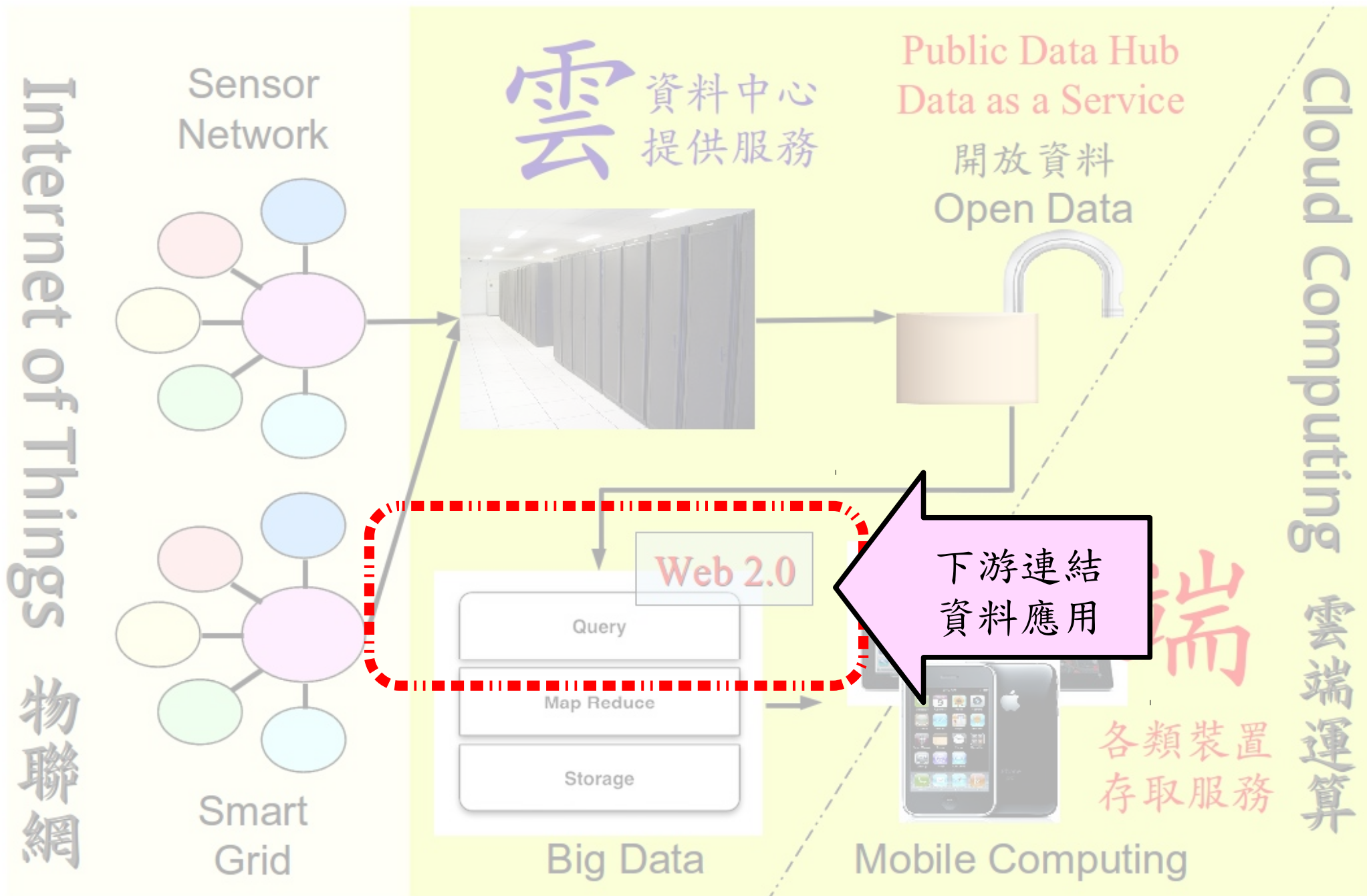
<https://www.facebook.com/groups/odtwn/>

The screenshot displays the Facebook interface for the 'OpenData / Taiwan' group. On the left sidebar, the profile of Jazz Yao-Tsung Wang is visible, along with navigation options like News Feed, Messages, and Events. The main content area features a cover photo with a network diagram of icons and text in French: 'PUBLIER SUR UNE PLATEFORME OUVERTE' and 'STATISTIQUES CARTOGRAPHIQUES HORAIRES DONNEES EC'. Below the cover, the group name and navigation tabs (About, Events, Photos, Files) are shown. A post creation bar includes options for writing, adding photos, asking questions, and adding files. The group statistics show 170 members, with 7 new members. A pinned post by user '樂樂' is partially visible at the bottom.

2013/03/22 Open Data Taiwan 170 members

巨量資料的社群關聯

Relations between Communities



NoSQL Taiwan Facebook Group

NoSQL Taiwan

<https://www.facebook.com/groups/306552142710977/>

NoSQL Taiwan Home

Jazz Yao-Tsung Wang
Edit Profile

FAVORITES

- News Feed
- Messages
- Events
- Photos
- Browse

ADS

- Ads Manager

PAGES

- Hadoop in Taiwan 2012
- Pages Feed **20+**
- Like Pages **1**

GROUPS

- JavaScript.tw **1**
- Taiwan Hadoop User Gr... **1**
- S4A **3**

NoSQL Taiwan About Events Photos Files Notifications

Write Post Add Photo / Video Ask Question Add File

683 members (**7 new**) · Invite by Email

+ Add People to Group

RECENT POSTS

Alex Lin
請問您把 Hadoop 用在那裏?

- Big data store (HBase)
- ETL (custom code, Pig) **+1**
- Data warehouse (Hive)

3 More...

Like · Comment · Follow Post · 11 hours ago

2013/03/22 NoSQL Taiwan 683 members

NoSQL & Big Data Architecture

NoSQL & BigData Architecture

<https://www.facebook.com/groups/423848814337101/>

The screenshot shows the Facebook interface for a group titled "NoSQL 中的分散式架構與 Big Data 議題小聚". The page header includes the group name and navigation options like "Home", "About", "Events", "Photos", "Files", and "Notifications". The group has 92 members, with 1 new member. The main content area features a post by Kuo-Chun Su, who is promoting Pivotal HD. The post includes a link to <http://www.greenplum.com/products/pivotal-hd> and a promotional image for Pivotal HD. The image shows the Pivotal HD logo and text: "Pivotal HD: The World's Most Powerful Distribution of Apache Hadoop | Greenplum", "www.greenplum.com", and "Dynamic Pipelining™ technology delivers 100X performance improvement with mature SQL query optimization and powerful analytics." The left sidebar contains navigation links for "News Feed", "Messages", "Events", "Photos", "Browse", "Ads Manager", "Pages", and "Groups".

2013/03/22 NoSQL-BigData 92 members

HBase Taiwan Facebook Group

HBase 小聚

<https://www.facebook.com/groups/289369481132604/>

Hbase小聚 Home

Jazz Yao-Tsung Wang
Edit Profile

FAVORITES

- News Feed
- Messages
- Events
- Photos
- Browse

ADS

- Ads Manager

PAGES

- Hadoop in Taiwan 2012
- Pages Feed 20+
- Like Pages 1

GROUPS

- JavaScript.tw 1
- Taiwan Hadoop User Gr...
- S4A 3

Hbase小聚 About Events Photos Files Notifications

Write Post Add Photo / Video Ask Question Add File

89 members (3 new) · Invite by Email

+ Add People to Group

Write something...

RECENT POSTS

余家豪

請問各位前輩們,我現在自己寫了一個bulkload的程式碼,在設定keyvalue裡頭的timestamp值用System.currentTimeMillis()發現精細度似乎不夠導致重複,若兩筆資料之row key,cf:q跟ts都一樣時,則前面那筆value會被覆蓋導致某些數據get不到,想問問各位前輩們的timestamp值都是如何設計的呢?

Like · Comment · Follow Post · March 12 at 7:54pm

Tony MoMo likes this. Seen by 31

View 4 more comments

2013/03/22 HBase.TW 89 members

MongoDB Taiwan Facebook Group

MongoDB 小聚

<https://www.facebook.com/groups/142553245867411/>

MongoDB小聚 About Events Photos Files Notifications

Write Post Add Photo / Video Ask Question Add File

114 members · Invite by Email

+ Add People to Group

RECENT POSTS

 **黃健瑋**
http://www.mongodb.org/downloads?mkt_tok=3RkMMJWWfF9wsRouvKvMZKXonjHpfsX86%2B0pW6SwlMI%2F0ER3fOvrPUfGjI4GSsFkI%2FqLAzICFpZo2FEJSueQcg%3D%3D MongoDB 2.4出囉~~

Downloads - MongoDB
www.mongodb.org
This table lists MongoDB distributions by platform and version. We recommend using these binary

FAVORITES
News Feed
Messages
Events
Photos
Browse

ADS
Ads Manager

PAGES
Hadoop in Taiwan 2012
Pages Feed 20+
Like Pages 1

GROUPS
JavaScript.tw 1
Taiwan Hadoop User Gr...
S4A 3

Jazz Yao-Tsung Wang
Edit Profile

Home

2013/03/22 MongoDB.TW 114 members

JavaScript Taiwan Facebook Group

JavaScript.TW

<https://www.facebook.com/groups/javascript.tw/>

The screenshot shows the Facebook group page for JavaScript.TW. At the top, the group name 'JavaScript.tw' is displayed with a profile picture of Jazz Yao-Tsung Wang. Below the name, there are navigation tabs for 'About', 'Events', 'Photos', and 'Files'. A large photo of a conference room with people sitting at desks with laptops is featured. Below the photo, there are options to 'Write Post', 'Add Photo / Video', 'Ask Question', and 'Add File'. The page also shows the number of members (2,177) and a pinned post by Tonyq Wang.

JavaScript.tw About Events Photos Files Notifications

Jazz Yao-Tsung Wang
Edit Profile

FAVORITES

- News Feed
- Messages
- Events
- Photos
- Browse

ADS

- Ads Manager

PAGES

- Hadoop in Taiwan 2012
- Pages Feed 20+
- Like Pages 1

GROUPS

- JavaScript.tw**
- Taiwan Hadoop User Gr...
- S4A 3

2012 JS 嘉年華會

Write Post **Add Photo / Video** **Ask Question** **Add File**

Write something...

2,177 members (76 new) · Invite by Email

+ Add People to Group

PINNED POSTS

Tonyq Wang edited a doc.
JS 參考讀物收集

2013/03/22 JavaScript.TW 2177 members

Supply Chain of Big Data Industry

Mobile

Web 2.0

Query

MapReduce

Storage

Open Data

IoT

JavaScript, TW **2177** members

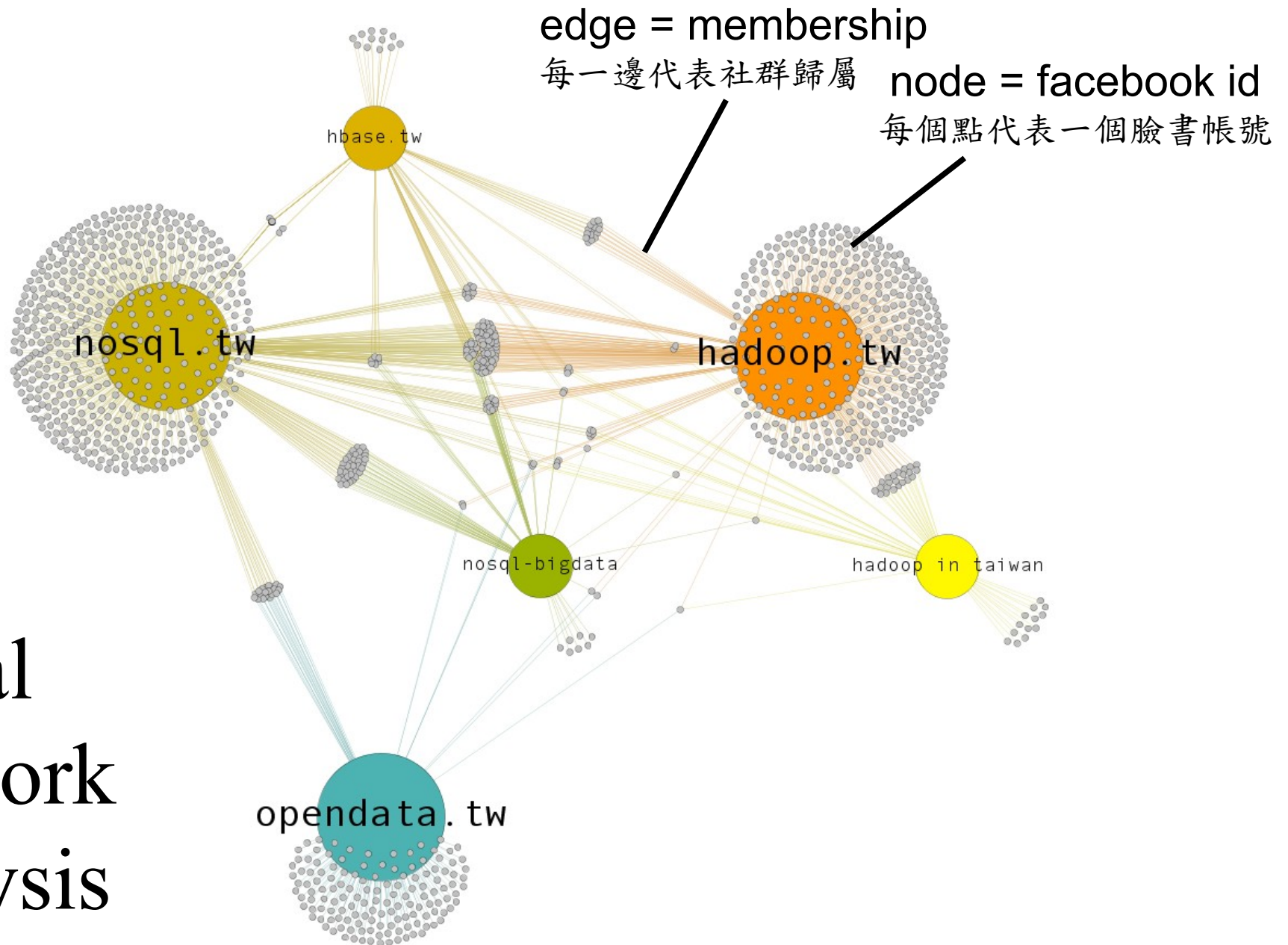
NoSQL Taiwan **683** members

Hadoop Taiwan **533** members

Open Data Taiwan **170** members

相關社群的群聚分析

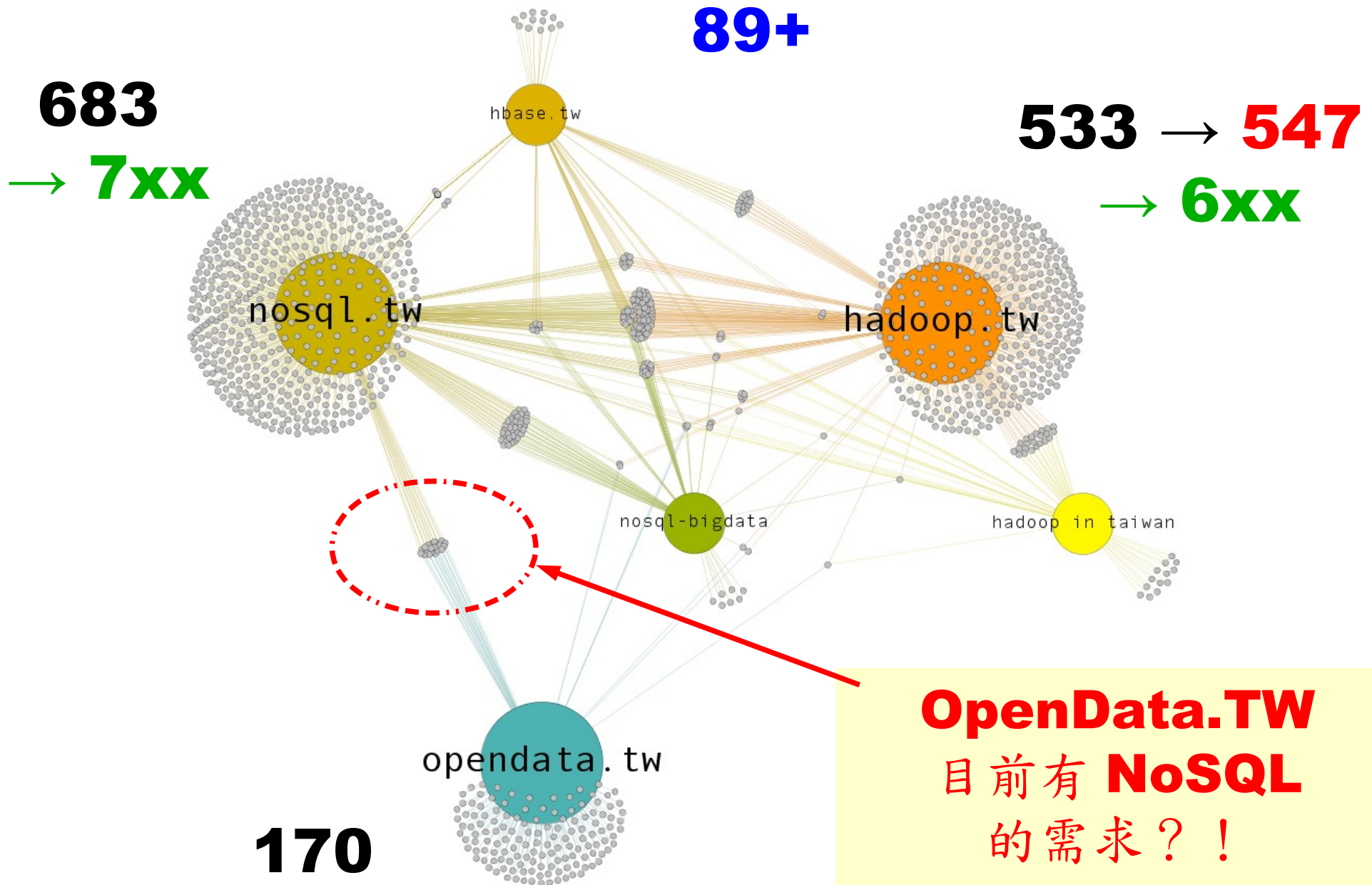
SNA of Communities



Social
Network
Analysis

從社群談未來的展望

Actions driven from SNA



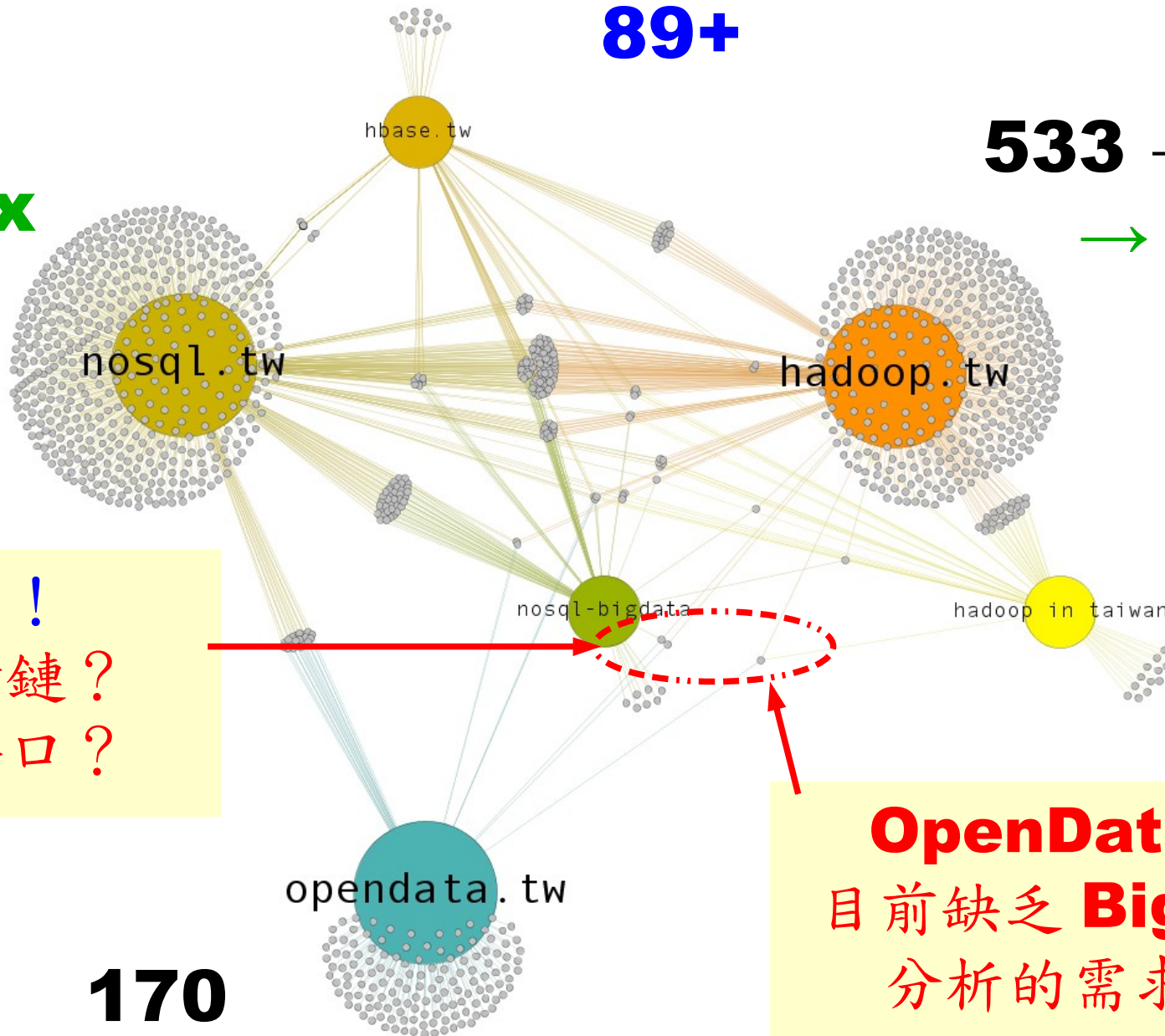
從社群談未來的展望

Actions driven from SNA

683
→ **7xx**

89+

533 → **547**
→ **6xx**



隱憂！
供應斷鏈？
人力缺口？

170

OpenData.TW
目前缺乏 **Big Data**
分析的需求？！

結語

Conclusion

Mobile

Web 2.0

Query

MapReduce

Storage

Open Data

IoT

紅海

需求明確
市場飽和

JavaScript.TW 2177 members

缺乏成功案例

NoSQL Taiwan 683 members

還算健康
持續互動

Hadoop Taiwan 533 members

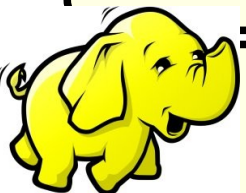
藍海

設法找出需求
強化供應鏈

Open Data Taiwan 170 members

有待法令開放

仍有待打通產業供應鍊
海量資料產業尚未成型





Questions?

Slides - <http://trac.nchc.org.tw/cloud>

Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw

