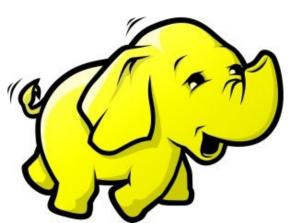


巨量資料的奇幻漂流 Life of Big Data

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





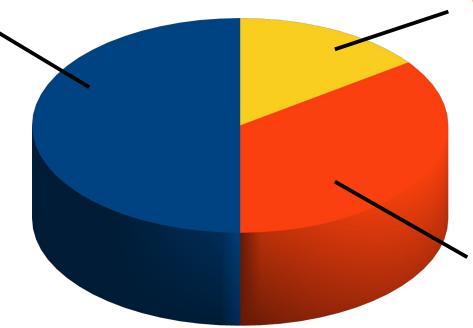
WHO AMIY 這像伙是誰啊? JAZZ ?

- 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。



FOSS 使用者

Debian/Ubutnu
Access Grid
Motion/VLC
Red5
Debian Router
DRBL/Clonezilla
Hadoop



行動力薄弱的開發者

TRTC WSU/ Haduzilla / Hadop4Win / Ezilla

推廣者

DRBL/Clonezilla Partclone/Tuxboot Hadoop Ecosystem

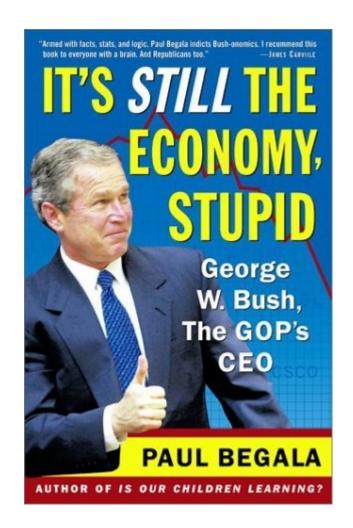
【問題】

目前還會把電子郵件用 Outlook 收信軟體 收到電腦裡的 請舉手?

【巨量資料?Big Data?】

各位數位移民 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】

如果〇包把您的信件 、照片通過語子, 該賠償您多少錢呢? 用MB計算嘛?

Agenden 演講次綱

From VYSN to lot?

物聯網的崛起

What is Open Data ? 何謂關放資料

What is Big Data?

何謂海童資料

What is **Smart City**? 淺談舒慧城市

Join Communities! 上鄉兒核人才

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?

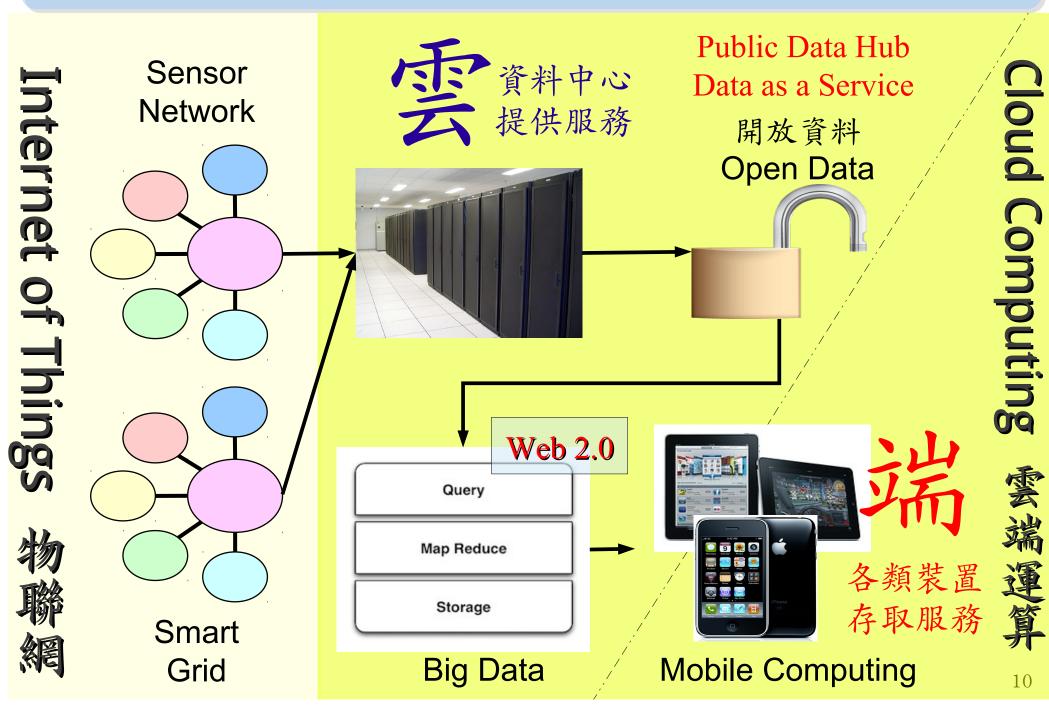
This is a story about Data Flow

今天要講一個關於資料的漫長旅程

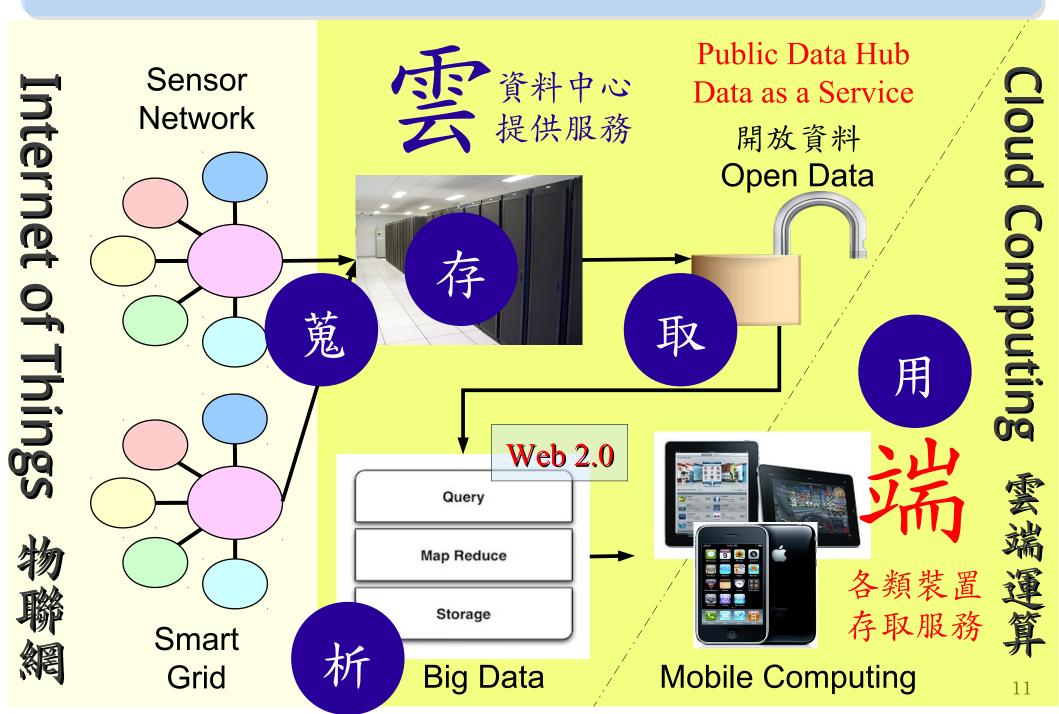
2013 三大熱門關鍵字

物聯網 Internet of Things 雲端運算 Cloud Computing 巨量資料 Big Data

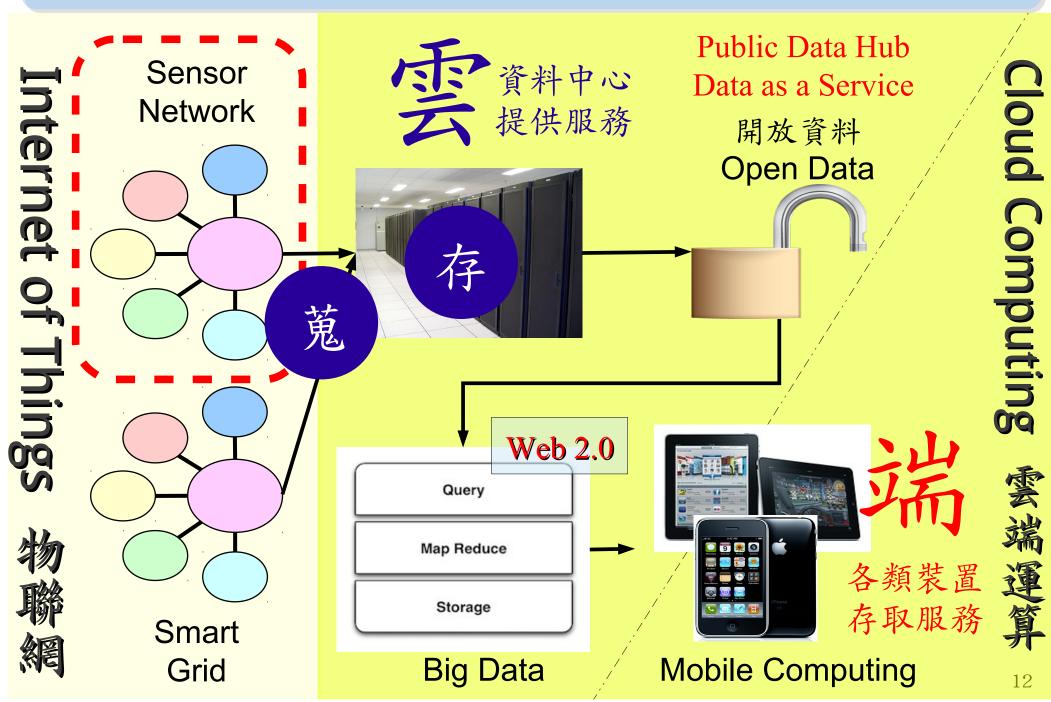
巨量資料的奇幻漂流 Life of Big Data



用五個字説明巨量資料的生命週期

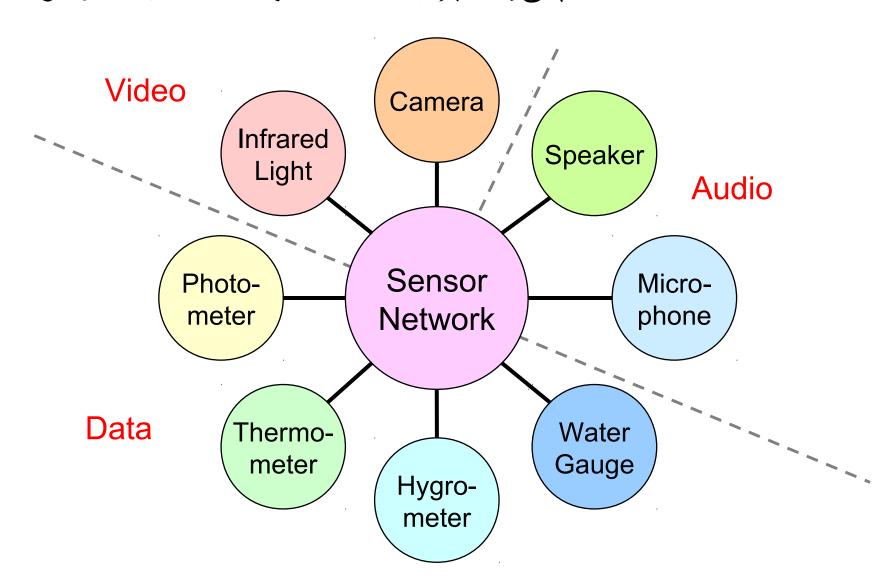


巨量資料的奇幻漂流 Life of Big Data



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 Scripp 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.

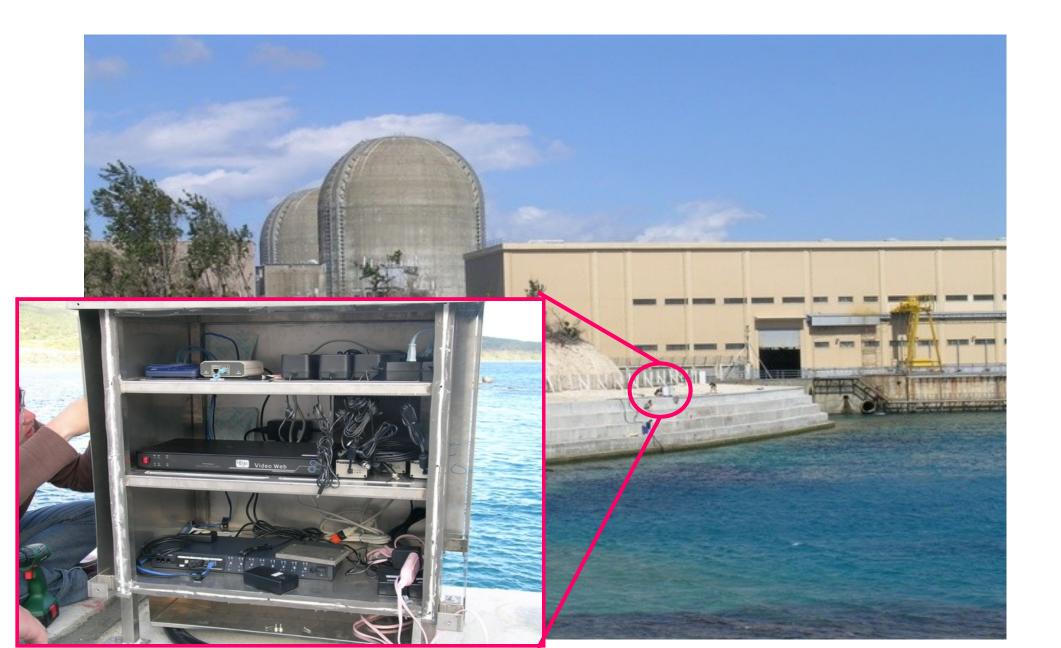
With the funding from Taipower, NCHC Ecogrid team worked together with Research for Biodiversity Center, Academic Sinica implemented an underwater reef monitoring system. The system composed with 9 cameras, located at 3 sites, lively streaming bioactivities undergoing in reef communities to marine scientists'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

http://ecogrid.nchc.org.tw

Underwater Camera



Site A



Site B & C





農業格網 (Agriculture Grid)



http://sensor.nchc.org.tw/tounan

Electric Wire Post





Passive Field Server

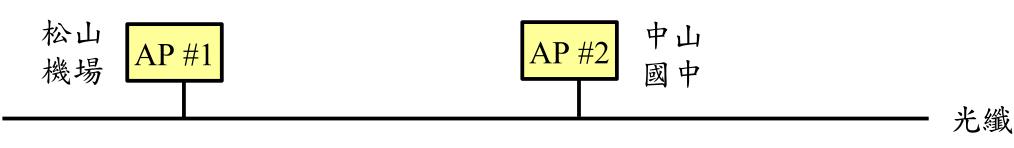


文湖線無線管理單元 (TRTC Neihu WSU)

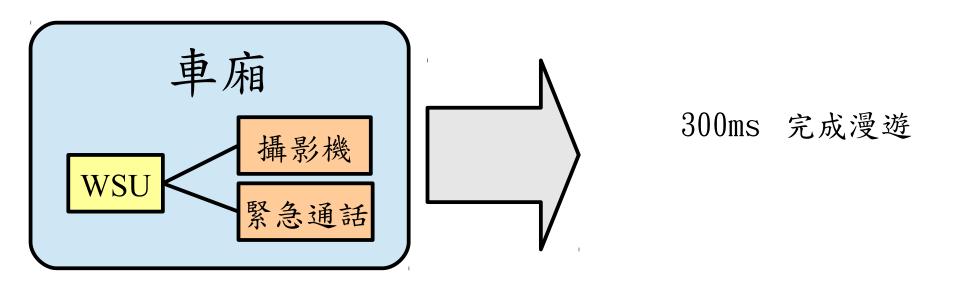


http://www.flickr.com/photos/rail02000/5292653579/lightbox/

Wireless Fast Roaming



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?



切換時聲音跟影像資料不能中斷!

科學園區管理局的 U-Taiwan 計劃



貓頭鷹復育試驗計畫

Park Owl Recovery Plan

◎勞資組 鄧慧卿/文

摘要

爲了解在高科技產業密集的科學園區之生態環境狀況,營造新竹園區有較多可供俗稱貓頭鷹的 領角鴞繁殖的條件,藉由蒐集他們捕食食物種類的概況,可作爲都會區生態多樣性的指標,進行本 復育試驗計畫,借監測其生長情形,得到比較準確且客觀的環境復育資料

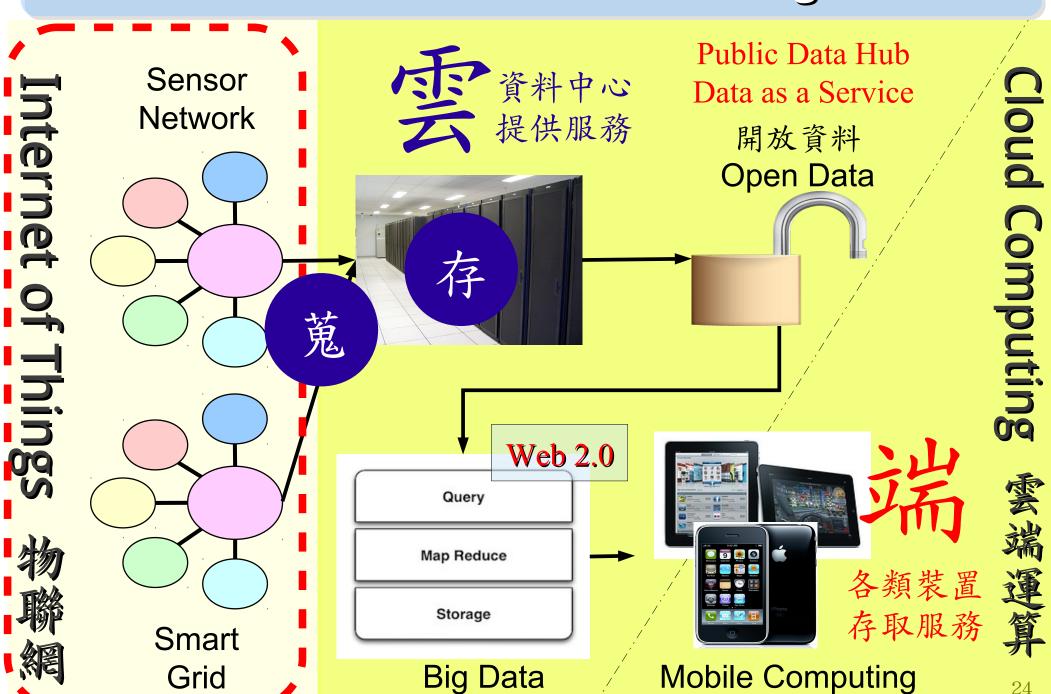
貳、計畫内容

以園區一路、介壽路北邊,靜心湖這一區 塊作爲貓頭鷹巢箱設置的試驗區,首先在試驗 區內進行巢樹的選擇,確定預裝置巢箱之樹木 位置(如楓香、相思樹、大葉桃花心木、鳳凰 木等)後,再確定搭設的高度及方向,最後將 巢箱掛設在適合的位置,總共設置20個巢箱,



貓頭鷹巢箱設置試驗

巨量資料的奇幻漂流 Life of Big Data



物聯網的時代來臨 Internet of Things

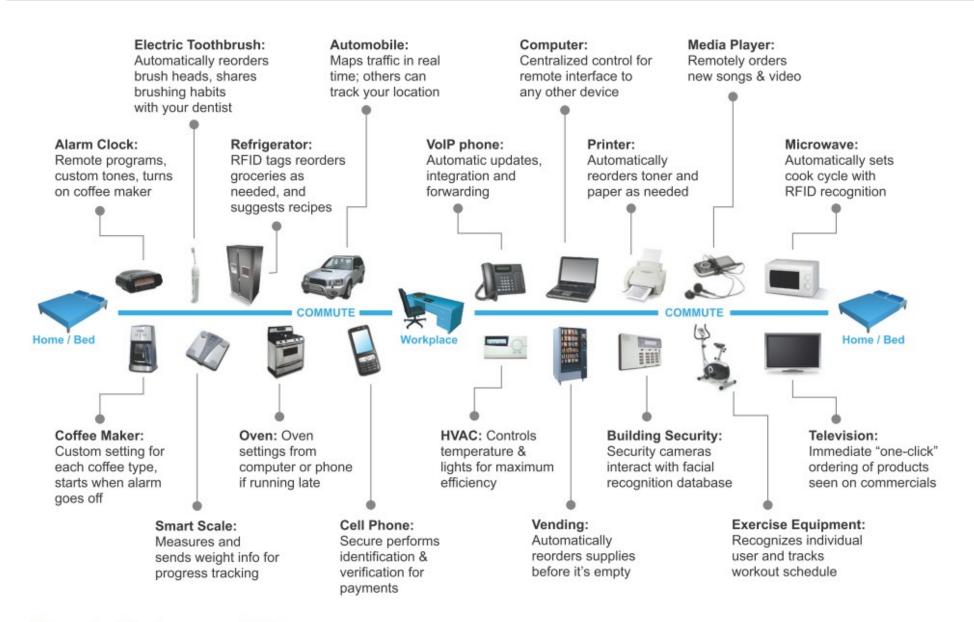


Figure 3. The Internet of Things

圖片來源: Attacks on Mobile and Embedded Systems: Current Trends by Mocana

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。

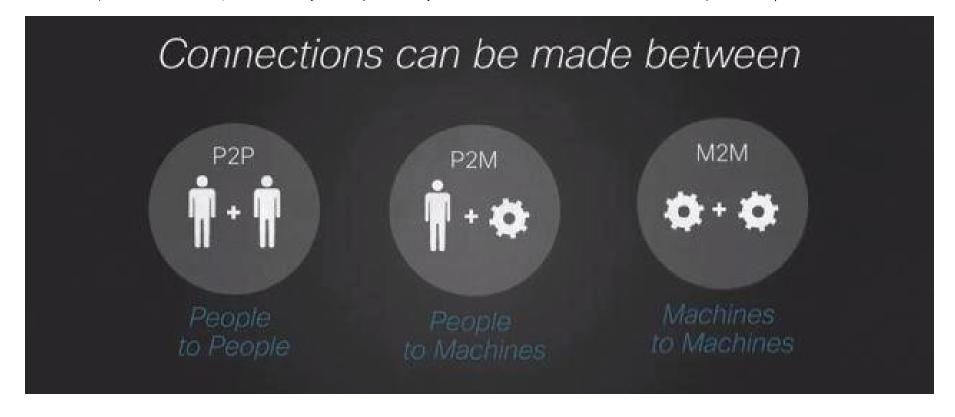
http://www.techbang.com/posts/1258-wifi-scales-through-the-internet-to-manage-your-weight

【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

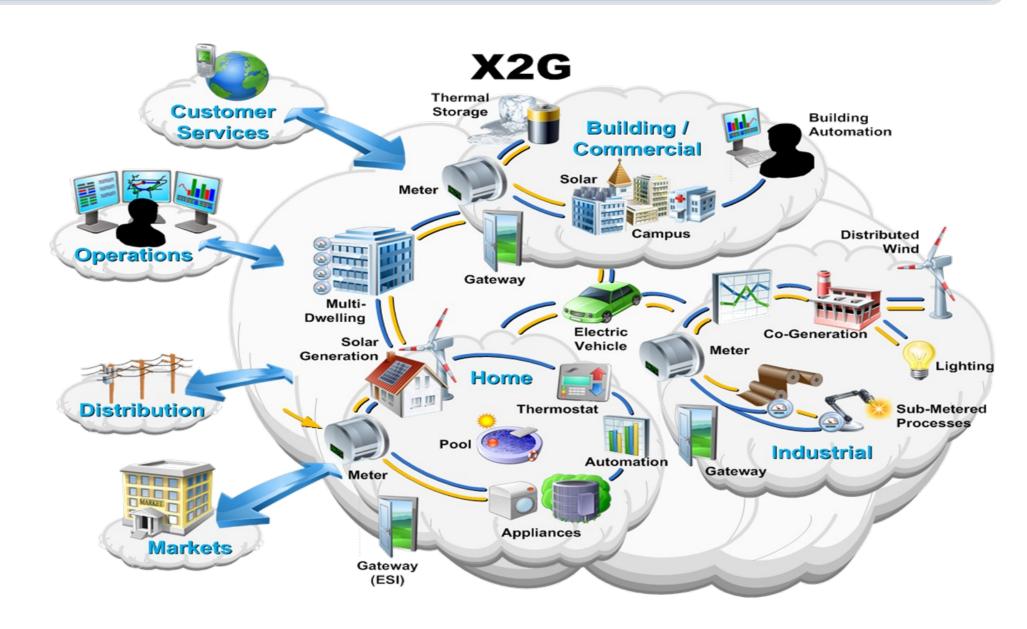
【關鍵技術】

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



圖片來源: NIST, 美國國家標準局

Agenda 演講大綱

From VYSN to lot?

物聯網的崛起

What is Open Data ? 何謂關放資料

What is Big Data?

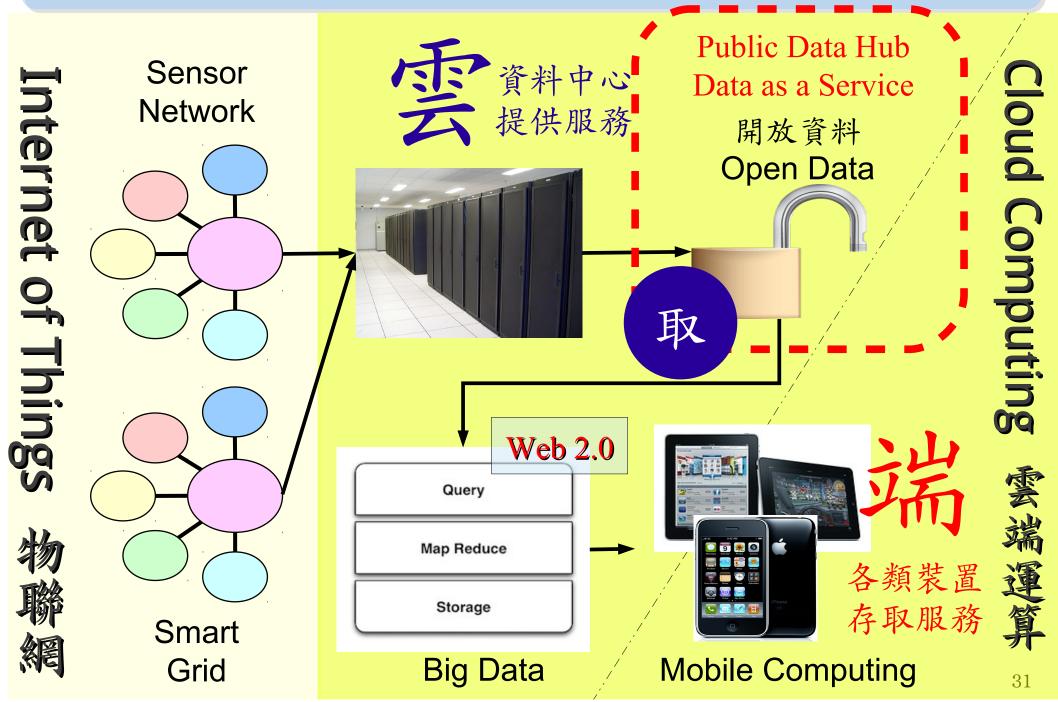
何謂海量資料

What is **Smart City** ? 淺談智慧城市

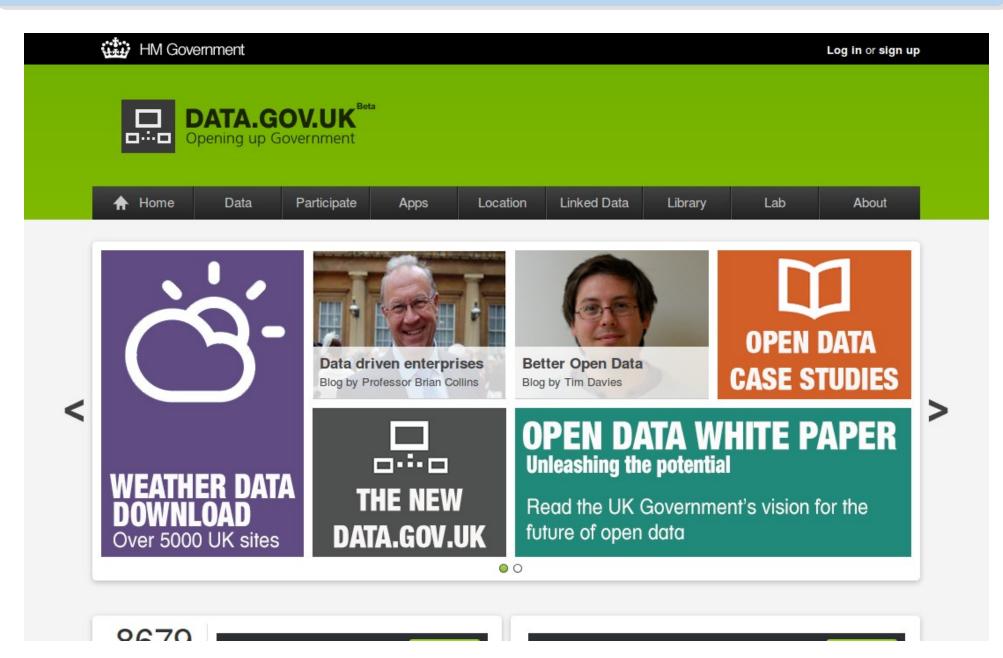
Join Communities! 上鄉兒核人才

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?

巨量資料的奇幻漂流 Life of Big Data



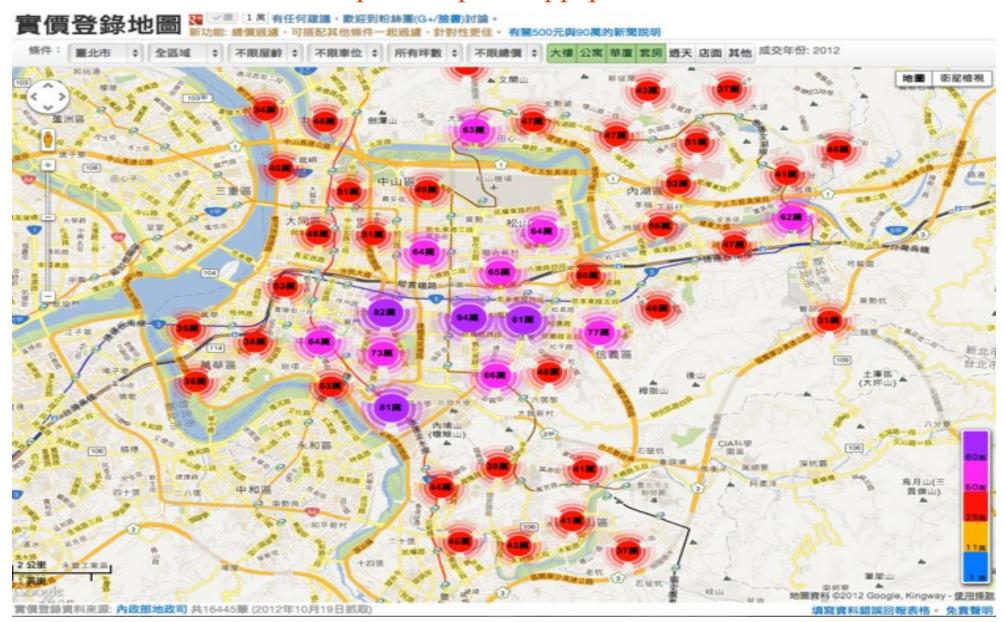
開放資料的新趨勢 Open Data



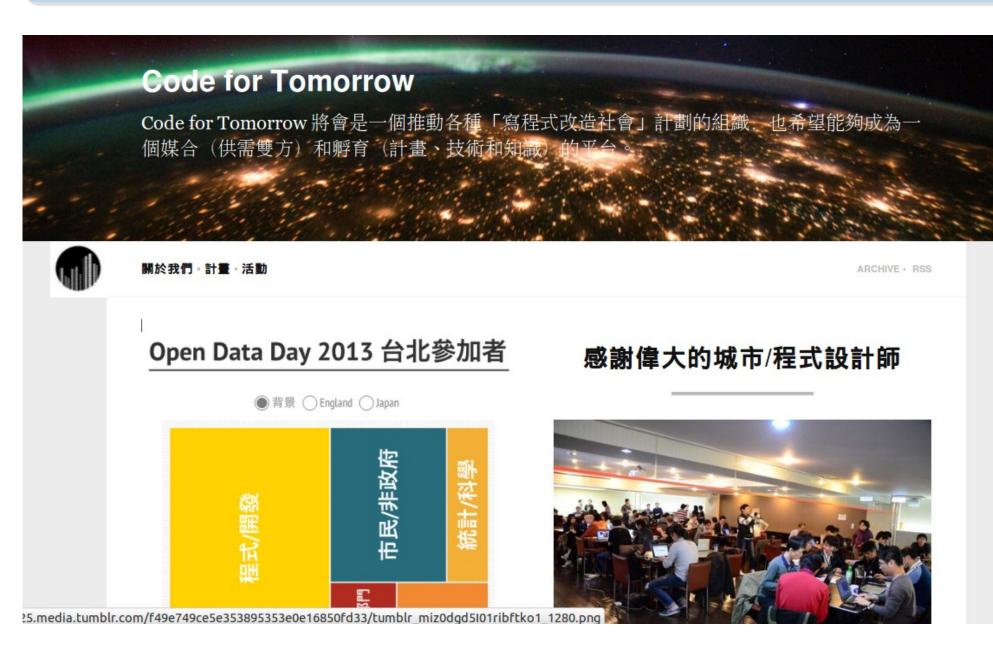
http://data.gov.uk/

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



開放資料帶來的好處: 萌典 - 因為覺得教育部的版本不好用

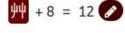
夢 ■ 國語辭典 - □ R

☑
♠
≰

請輸入欲查詢的字詞

萌▶

14



méng ☆

名

1. 草木初生的芽。

說文解字:「萌,艸芽也。」

唐・韓愈、劉師服、侯喜、軒轅彌明・石鼎聯句:「秋瓜未落蒂,凍芋強抽萌。」

2. 事物發生的開端或徵兆。

韓非子・說林上:「聖人見微以知萌,見端以知末。」

漢・蔡邕・對詔問灾異八事:「以杜漸防萌,則其救也。」

3. 人民。

如:「萌黎」、「萌隸」。

通「氓」。

4. 姓。如五代時蜀有萌慮。

動

1. 發芽。

如:「萌芽」。

楚辭・王逸・九思・傷時:「明風習習兮龢暖,百草萌兮華榮。」

https://www.moedict.tw

關放資料帶來的好處: 求職小幫手 - 图為221《很可憐

您好·歡迎來到518![登入][註冊]

外包接案 - 網站地圖

□ 用戶留言 | - addcn 析

由 推薦職缺

職務名稱

• 正職人員

• 夜班門市人員

■ 由属+外場人員

• 現台

[土城裕生店]早班計時。

肥權人員(早班長期)

• 內場計時服務人員(早班...

85度C羅東站前店早、段



門市人員-正職(竹南站前)

品股份有限公司(竹南站前) 公司簡介» 該公司其他鞣缺(0)

行業類別:餐飲業

資本額:暫不提供/員工人數:3人

職務説明

工作內容: 1. 調酒師負責調製各式含酒精或不含酒精的飲料。吧檯人員負責 製作各式餐點、飲料、或服務顧客點餐供餐等工作。



職務類別: 調酒師/吧檯人員 / 餐飲相關

需求人数: 不拘

工作待遇: 依公司規定

工作性質: 全職、兼職、工讀 2 兼差接案

身份類別: 一般求職者、應屆畢業生、日間就讀中、夜間就讀中、中高齢者、二度就業、原住民、研發替代役、

身障人士

是否出差: 不需出差/外派

2012/10/24. 勞動基準法第22條第2項工資未全額直接給付勞工(高市府勞條字第10137737900號)[<u>高雄市違反勞基法事業名單</u>]

2012/12/14. 勞基法第80條拒絕、規變或阻撓勞動檢查(北府勞條字第1013103183號函)[新北市違反勞基法事業名單] 2012/12/06. 勞基法第第27條(府授勞動字第1010219168號)[臺中市違反勞基法事業名單]

北市中正

http://jobhelper.g0v.ronny.tw/

關放資料帶來的好處:新關小幫手 - 因為網路鑑言特別多



新聞小幫手測試專用分享了 1 條連結。 8月12日在 Taipei 附近 🕝

http://www.nownews.com/2013/08/07/11635-2971267.htm



揭密/韓媒:月球發現韓國古文明 NASA刻意隱瞞

www.nownews.com

(大陸新聞中心/綜合報導)據俄羅斯《真理報》報道,黄亨錫認為,月

注意! 您可能是問題新聞的受害者

2007年大陸網路謠言 ▮

讚・留言・分享

關放資料帶來的好處:關放食庫 - 用群眾外包來管食品安全

開放食庫 Food Open Data

飲食革命, 從認識你吃的食物開始。

認識開放食庫

工作團隊 手機App

部落格

所有文章









《開放食庫》是一個民間自發性建立的包裝食品資料庫。蒐集的内容包含,各類市售包裝食品 的産品標籤資訊,例如:營養標示、成分和製造商資訊。有了開放食庫,將能讓消費者吃得更 聰明、更安心;並協助食品製造商了解消費者需求,透過使用回饋,研發更健康、更符合市場 需求的産品。

計畫緣起

身為消費者,我們幾乎每天都會接觸到包裝食品,但是卻並不清楚這些食品 的成份。即使是食品製造商,也未必可以全盤掌控自家產品的成份。

消費者

計畫介紹



Agenda 演講大綱

From VYSN to lot?

物聯網的崛起

What is Open Data ? 何謂關放資料

What is Big Data?

何謂海童資料

What is **Smart City**? 淺談舒慧城市

Join Communities! 上鄉兒核人才

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?

巨量資料的奇幻漂流 Life of Big Data



Cloud Computing

Why are they named by "SMART" 乳 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 Daita

海識 Knowledge

智慧 Wisdom

Can Machine understand You? 讓機器更懂你?

iPhone Features Built-in Apps



Siri. Siri. Siri. 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.



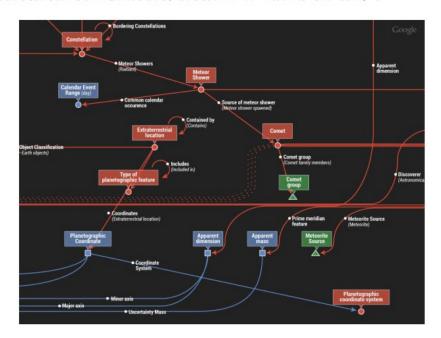
Google將發展「人工智慧」 永久改 變搜尋引擎

2012年02月15日 00:11

點評: 超級阿斯拉, 衝啊! (阿斯拉: 好的, 隼人!)

記者黃郁棋/綜合報導

每個人都在猜,下一波網路革命是什麼?每個人都在猜,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)

http://www.ettoday.net/news/20120215/25085.htm

大家都說「資料是金礦」,那就讓我們拿採礦當類比吧!

國際金價

提供給客戶的價值

產品通路

開採成本

總擁有成本

軟硬體投資

提煉廠

分析平台與工具軟體

SMAQ

含金度

資料鑑價?

商業模式

開採權

分析資料的合法性

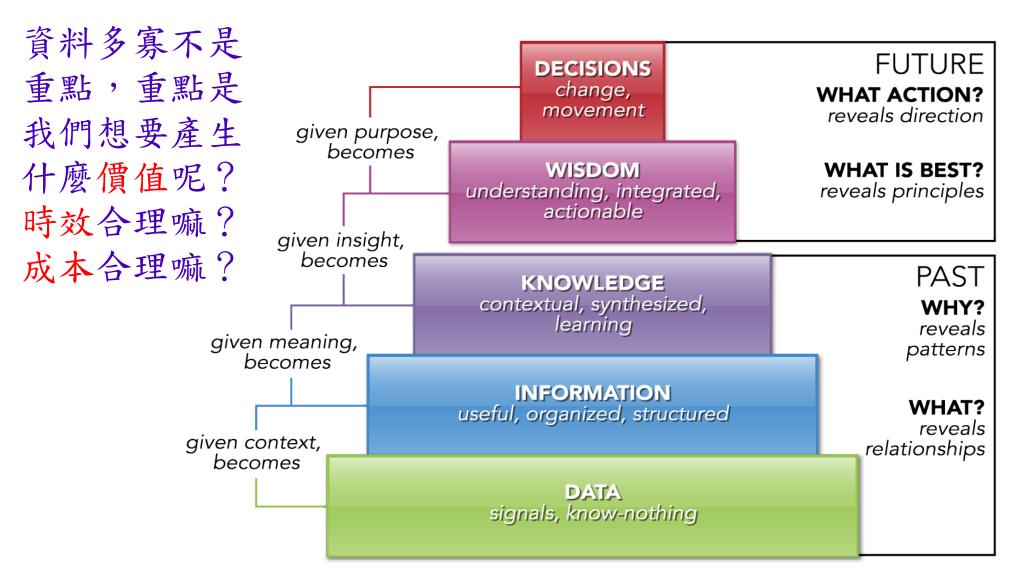
個資法

金礦

資料集

Open Data

知識源自彙整過去,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]



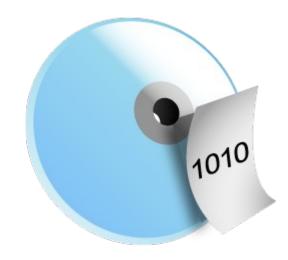
Big data is a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a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







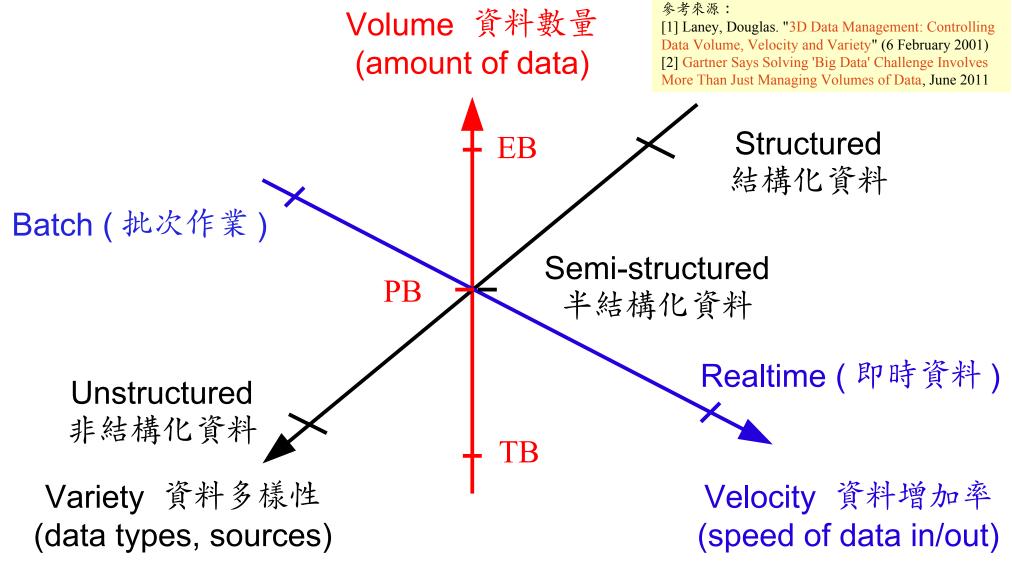
多個檔案,容量 100TB

一個資料庫,容量 100TB

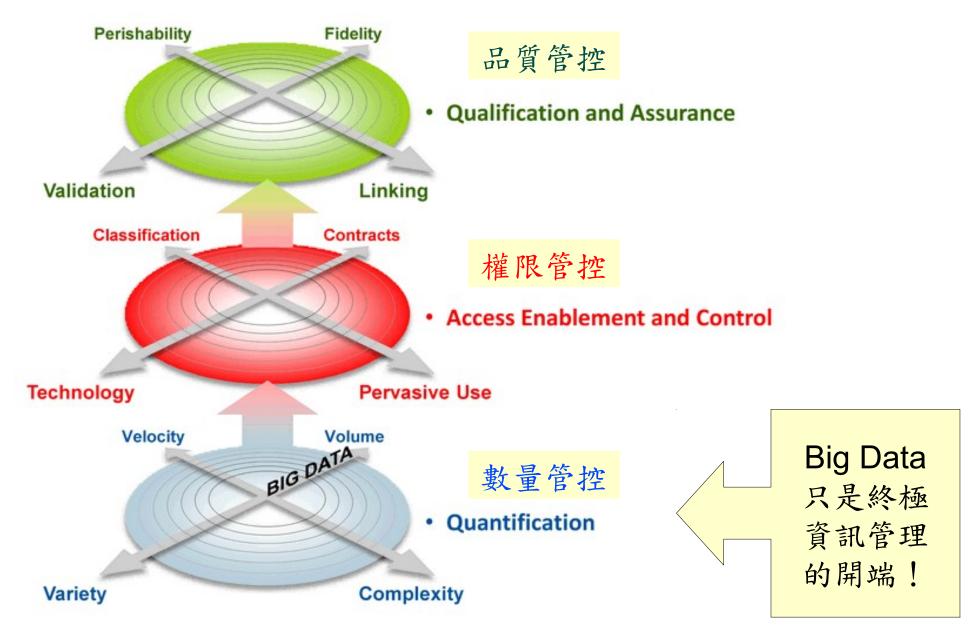
一個檔案,容量 100TB

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」



12D of Information Management? 12個維度?



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 H用雲掌握資料,加以分析,形成智能給端用



雲端設計新思維:端的智能來自於雲的服務

Devices share the wisdom of Cloud











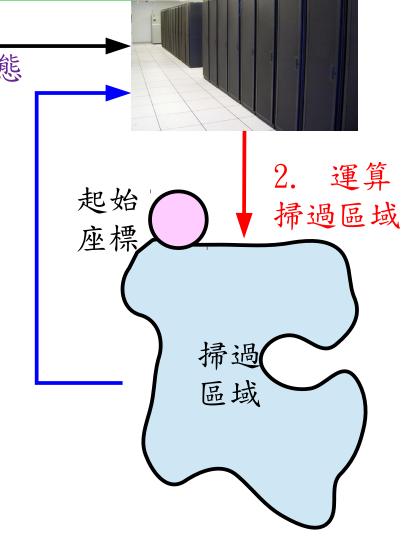


各類裝置存取服務

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



曾經有人問道:我是學電機的,那電機與控制跟雲端有甚麽關聯性?

Agenda 演講大綱

From VYSN to lot?

物聯網的崛起

What is Open Data ? 何謂關放資料

What is Big Data?

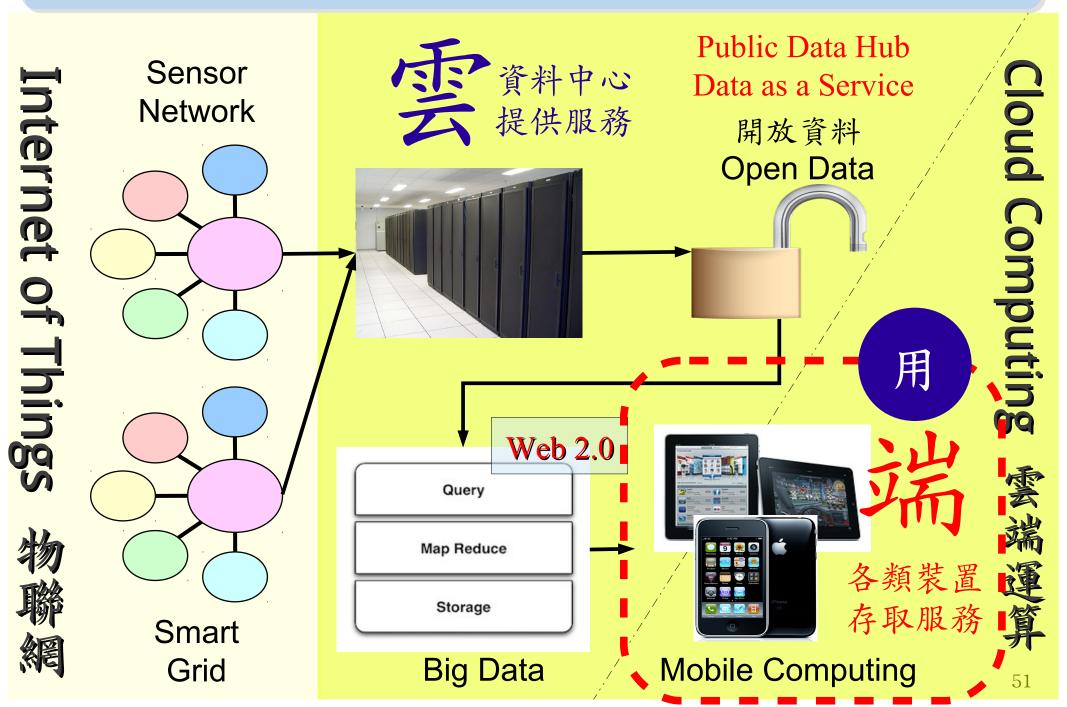
何謂海量資料

What is **Smart City** ? 淺談智慧城市

Join Communities! 上鄉兒核人才

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?

巨量資料的奇幻漂流 Life of Big Data



【與生活相關的巨量資料應用】

智慧城市 Smart City

十九世紀是帝國的世紀二十世紀是國家的時代

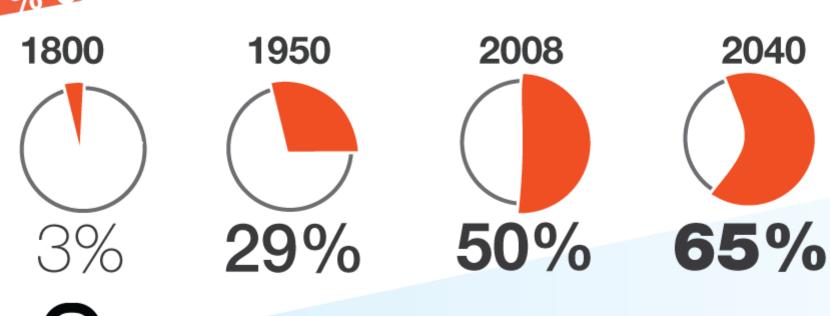
ANATOMY OF A SMARR CITY

"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."

二十一世紀將是城市的時代

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

% of population living in cities



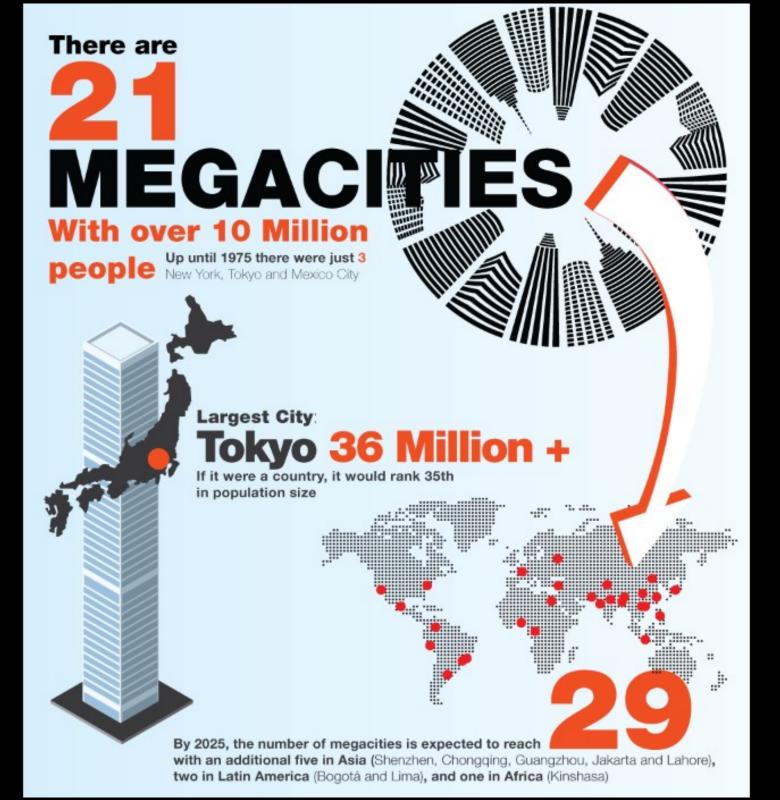


1.3 million

每個禮拜有一百三十萬人

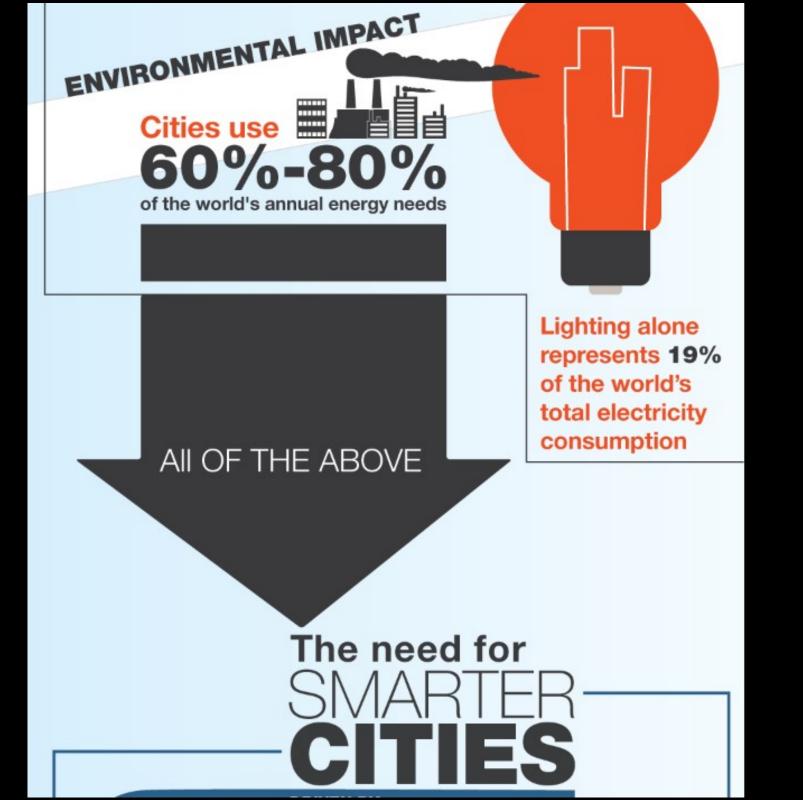
people are currently movin<mark>g</mark> into cities each week 搬入城市

的



In the developing world, as much as

80% of future economic growth will occur in cities



城市使用了百分之六十至八

十的

能

源



智慧城市 主要基於感測器 時城市資 、網路與行動裝置 則訊系統



By way of new sensor networks the accurate monitoring of environmental conditions like pollution levels, wildlife counts, and water runoff all become possible.

量測對象包括環境、安電力、建築物

安全

交通



dontflush.me

Structural Health Monitoring of buildings, bridges and dams as well as advanced warnings systems in emergency situations can now be put in place.





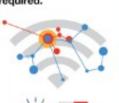
TRANSPORATION

Through sensors embedded in roadways and street lights, realtime transit and traffic can be managed for the purpose of reducing travel time and fuel inefficiencies.





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.





Smart Buildings utilize monitoring devices that track usage and empower users and service providers to better control and reduce electricity demands.





OUTLOOK

FINANCIAL

\$34

Billion

The Smarter cities IT

market opportunity

is expected to be:

CIVIC & ENVIRONMENTAL

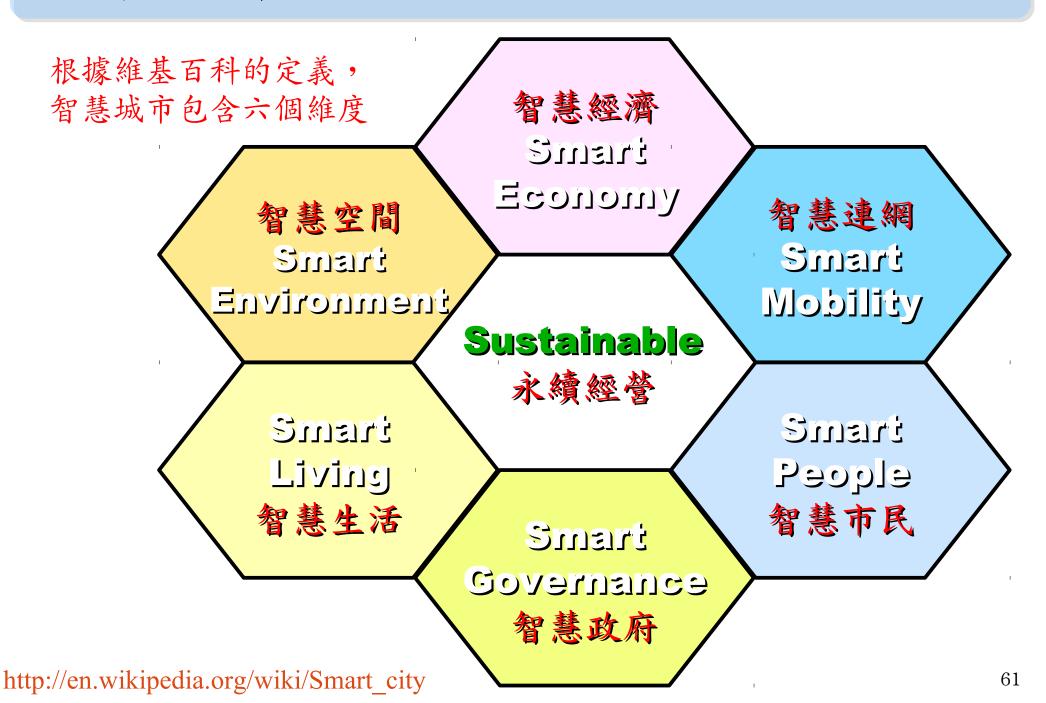
Cisco estimates that cities that run on information can improve their energy efficiency by 30% within 20 years.



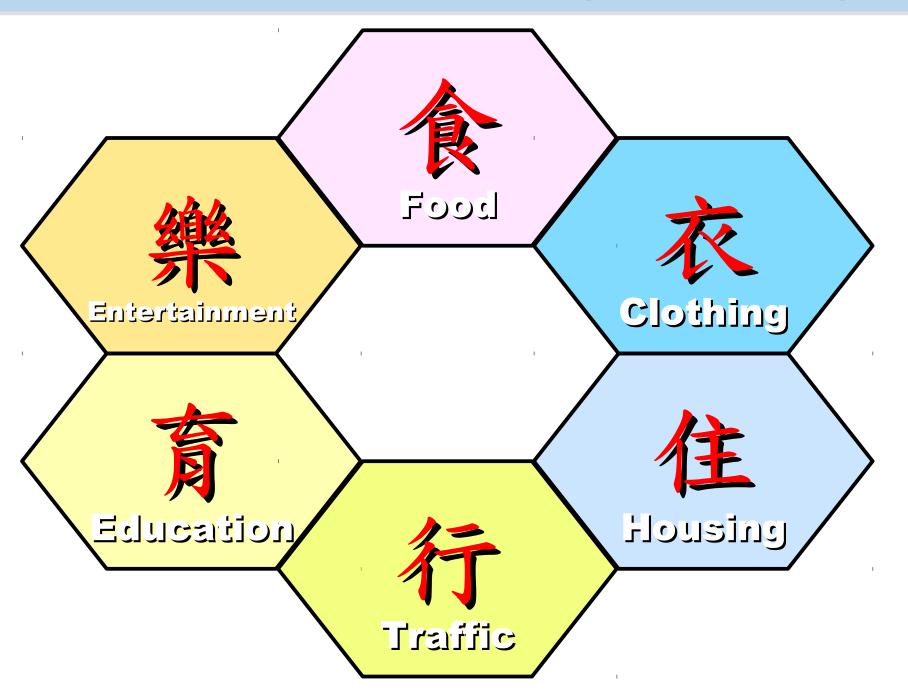


Postscapes**

智慧城市的六個維度 6D of Smart Cities



讓人民有感的智慧城市 Ubiquitous Computing

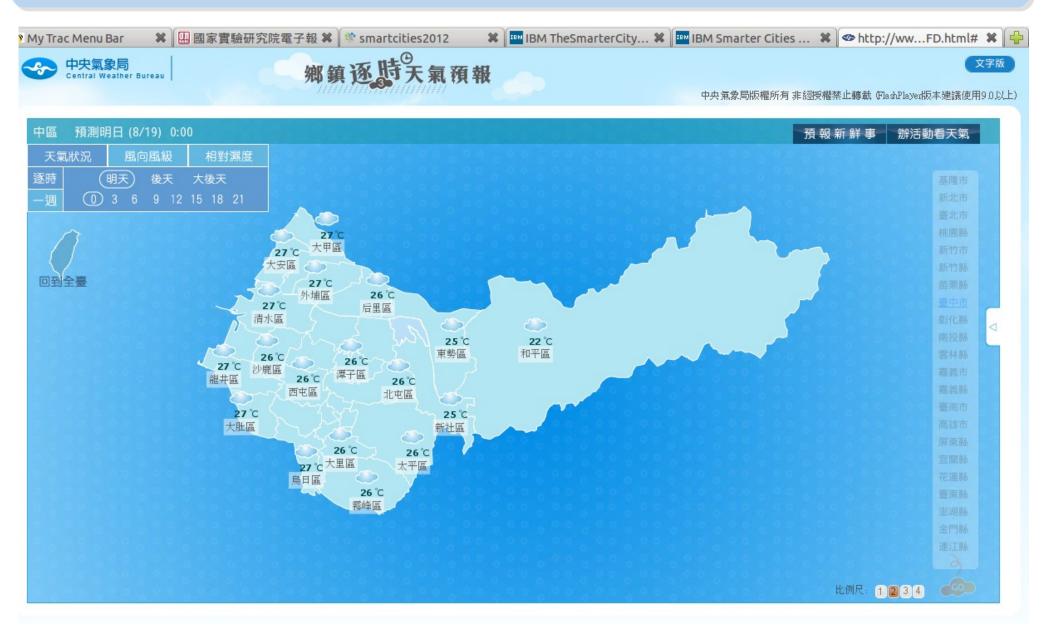


幫忙您找食物的地理資訊系統服務



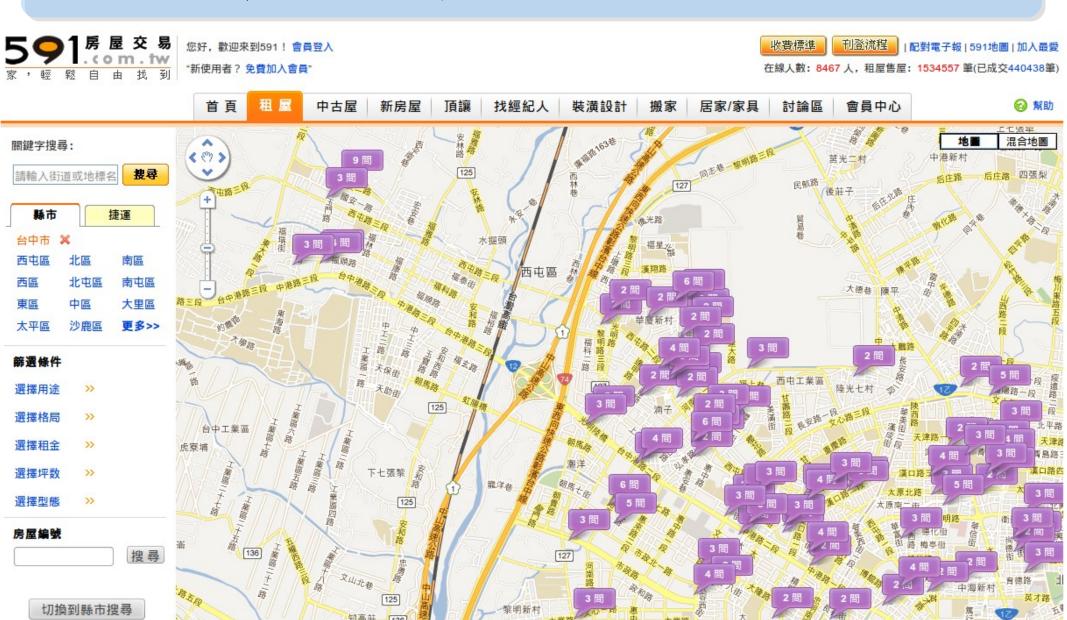
http://www.ipeen.com.tw/map/#loc=台中市

幫忙您決定該怎麼穿衣服的氣象服務

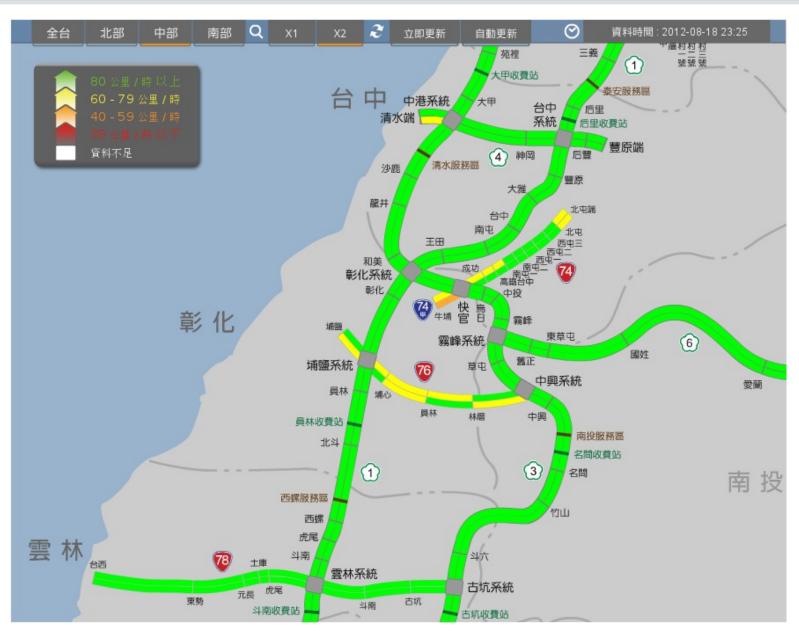


http://www.cwb.gov.tw/township/

幫忙您找房子的地理資訊服務

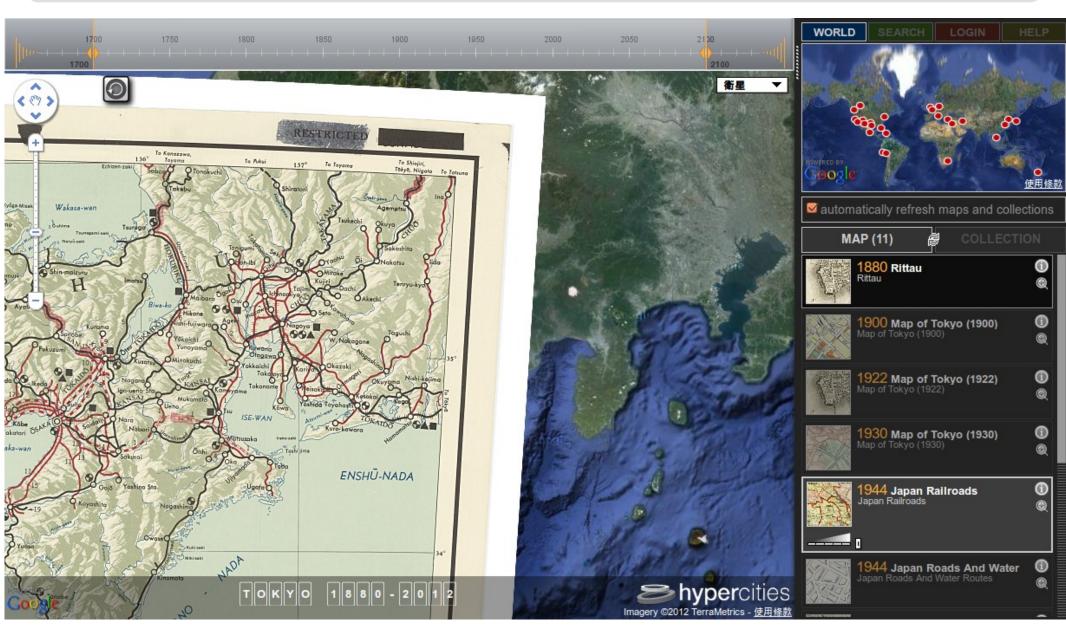


幫忙您查即時交通路況的地理資訊服務



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

用地理資訊服務學習一個城市的歷史



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

Agenda 演講次綱

From VYSN to lot?

物聯網的崛起

What is Open Data ? 何謂關放資料

What is Big Data?

何謂海量資料

What is **Smart City** ? 淺談智慧城市

Join Communities! 上鄉兒核人才

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?

【建議】

面對大陸與韓國的挑戰, 台灣產業必須形成「供應鏈聯盟」

培育組織人才,也得打團體戰 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











行動裝置 共通平台 htc quietly brilliant





各類裝置 存取服務

軟體服務供應商





資料中心機房維運









資料中心 提供服務

基本硬體建設組件



FOXCONN® 鴻海科技集團 英業達集團 Invented



雲端產業供應鏈

Supply Chain of Cloud Computing

應 用 軟體 供應 商











行動裝置 共通平台 ntc quietly brilliant





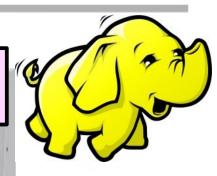


各類裝置 存取服務

軟體服務 供應商 Google



大象擺這裡



資料中心機房維運







資料中心提供服務

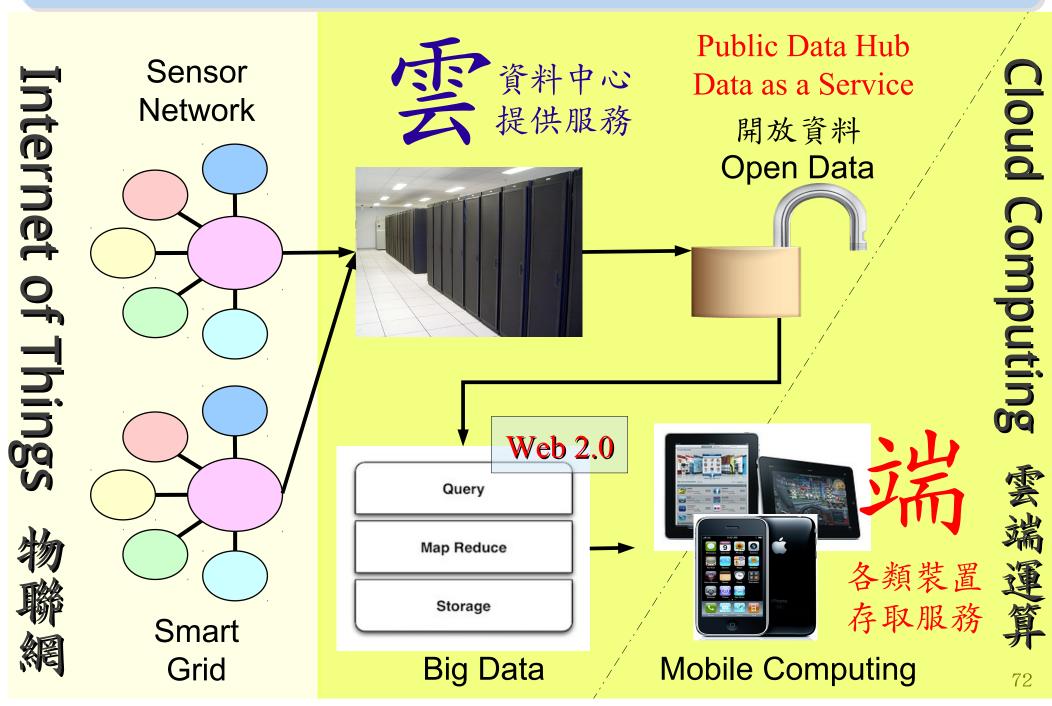
基本硬體建設組件



FOXCONN[®] 鴻海科技集團 **英業達集團**

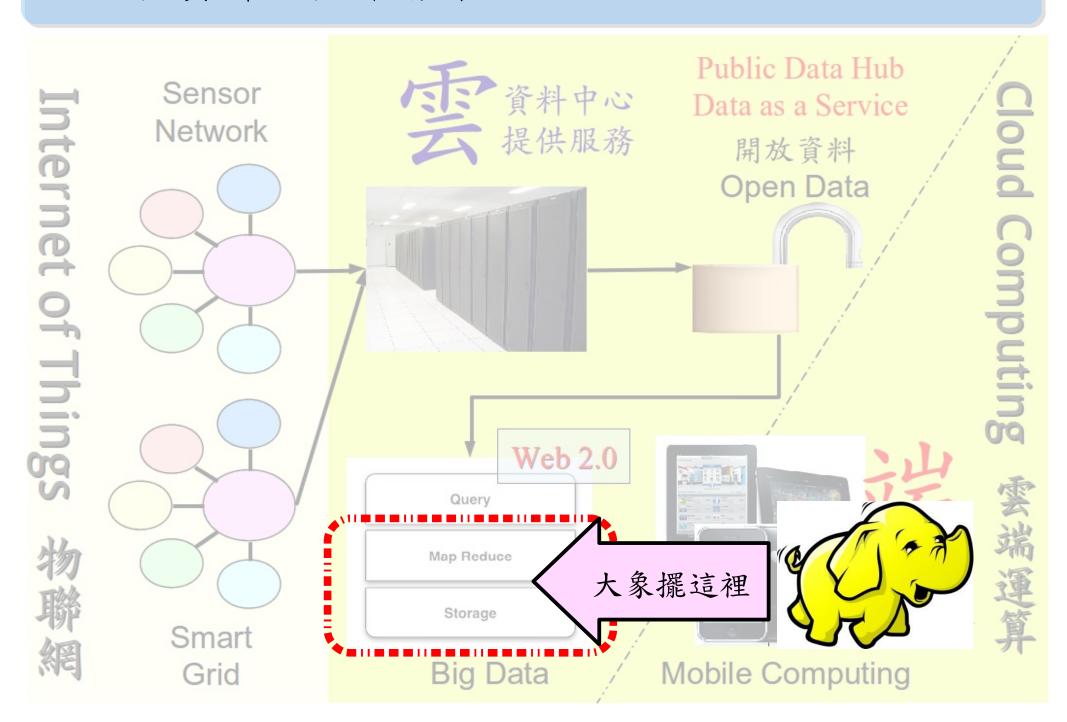


巨量資料的奇幻漂流 Life of Big Data



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



2013/03/22 Hadoop.TW 533 members

Hadoop related Facebook Group in Taiwan

Hadoop in Taiwan

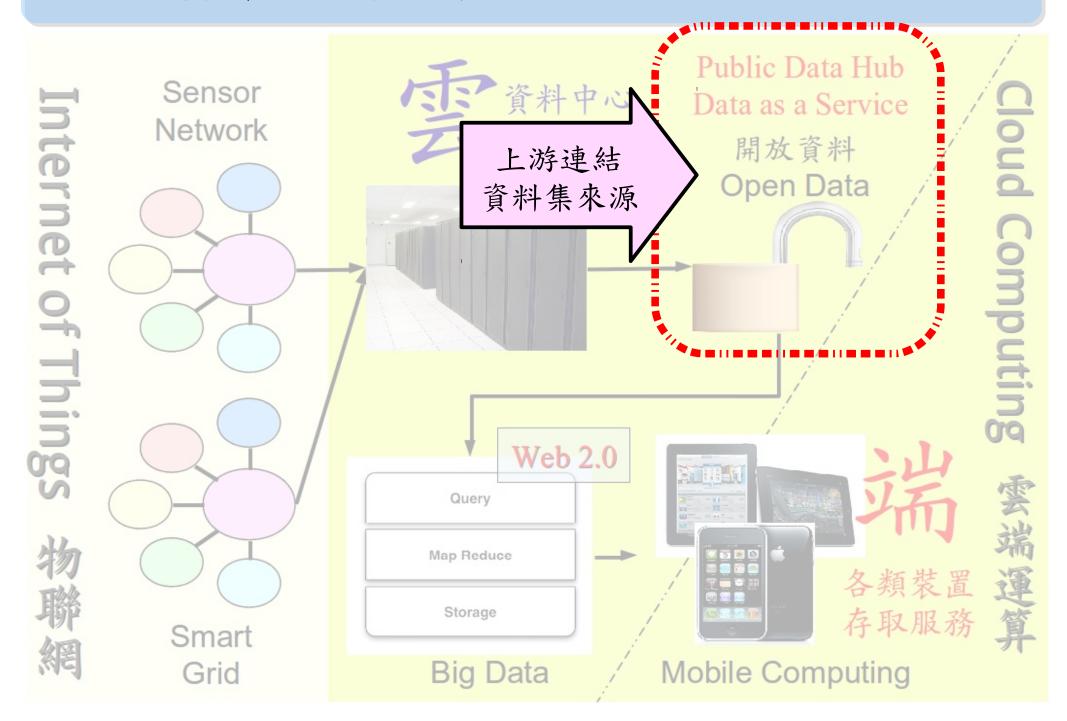
https://www.facebook.com/groups/hadoopintaiwan/



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)

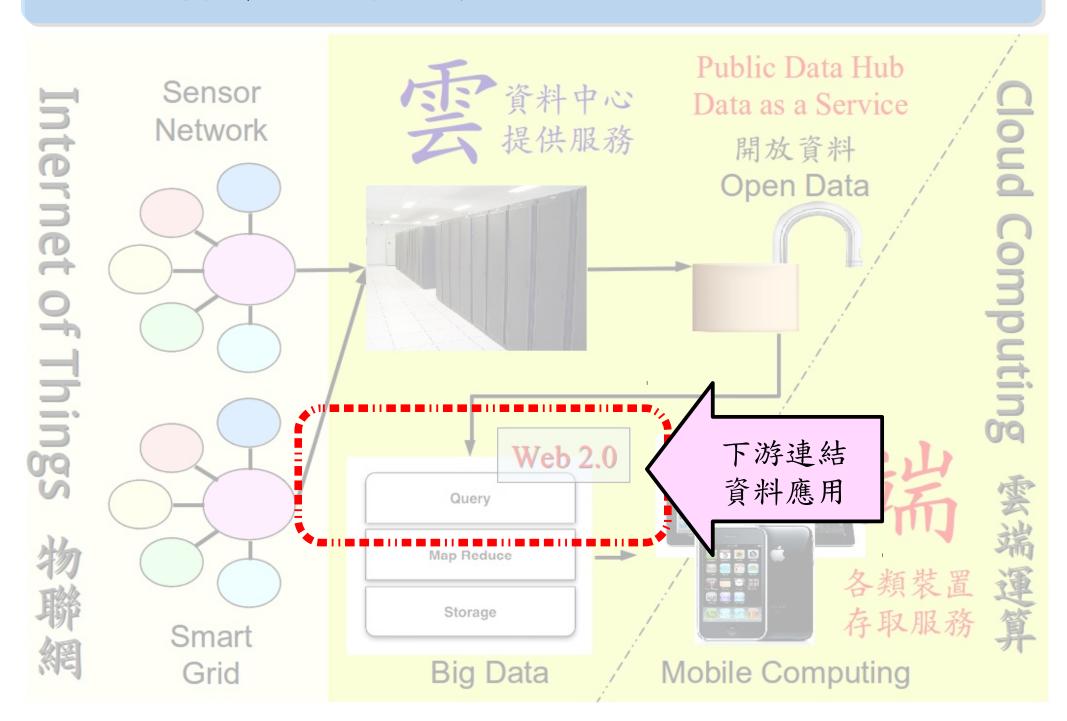
https://www.facebook.com/groups/odtwn/



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 & Big Data Architecture

NoSQL & BigData Architecture

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



2013/03/22 NoSQL-BigData 92 members

HBase Taiwan Facebook Group

HBase 小聚

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



2013/03/22 HBase.TW 89 members

MongoDB Taiwan Facebook Group

MongoDB 小聚

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



2013/03/22 MongoDB.TW 114 members

JavaScript Taiwan Facebook Group

JavaScript.TW

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



2013/03/22 JavaScript.TW 2177 members

Supply Chain of Big Data Industry



Web 2.0

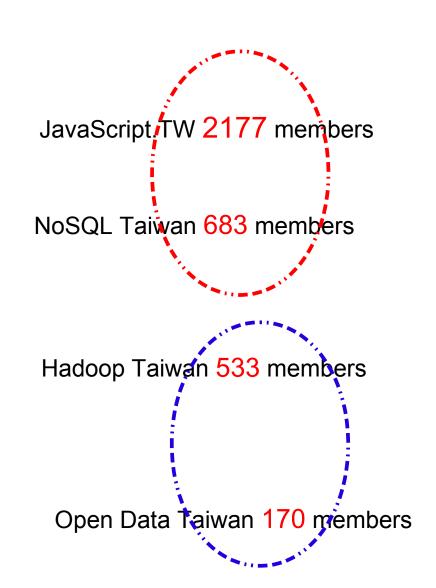
Query

MapReduce

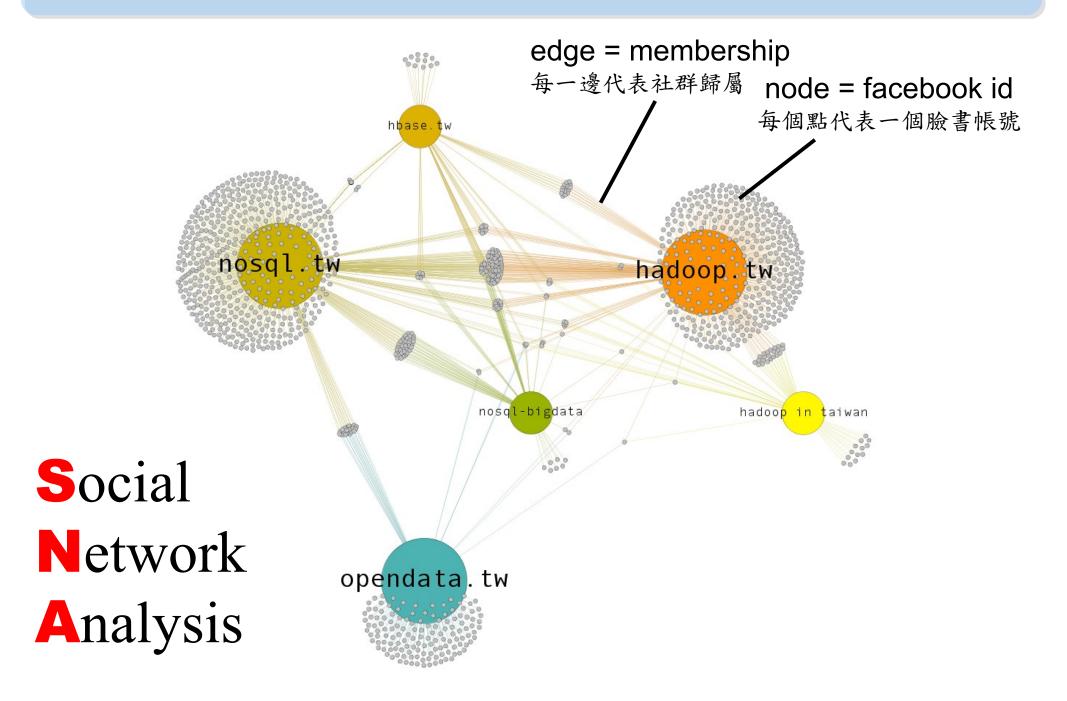
Storage

Open Data

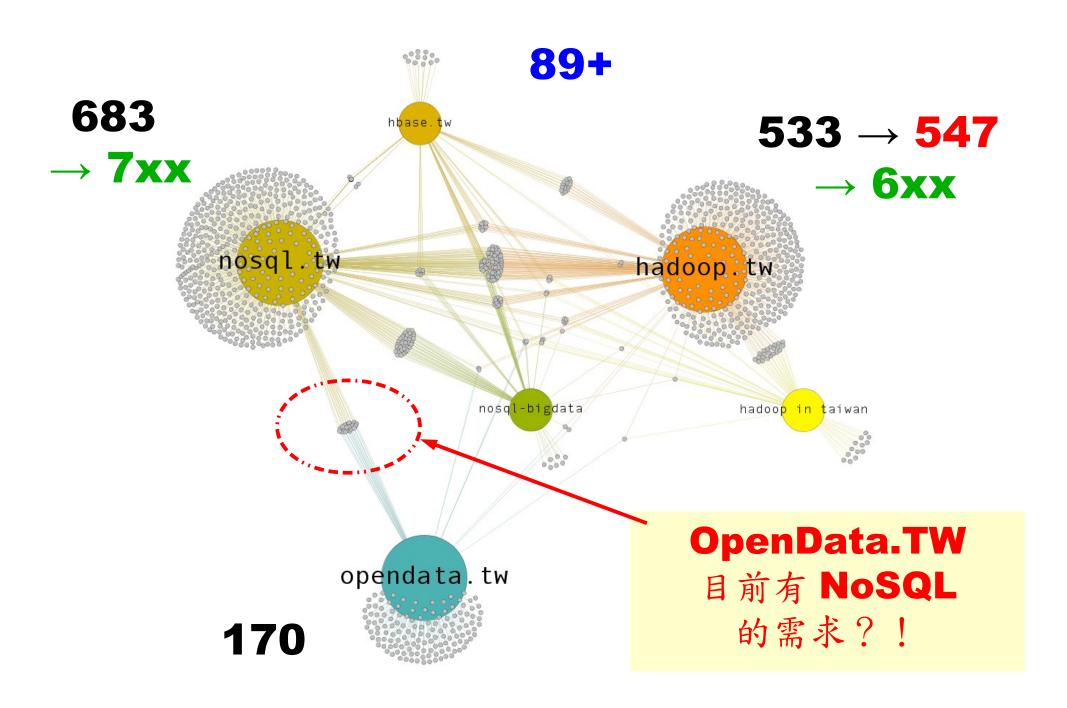
IoT



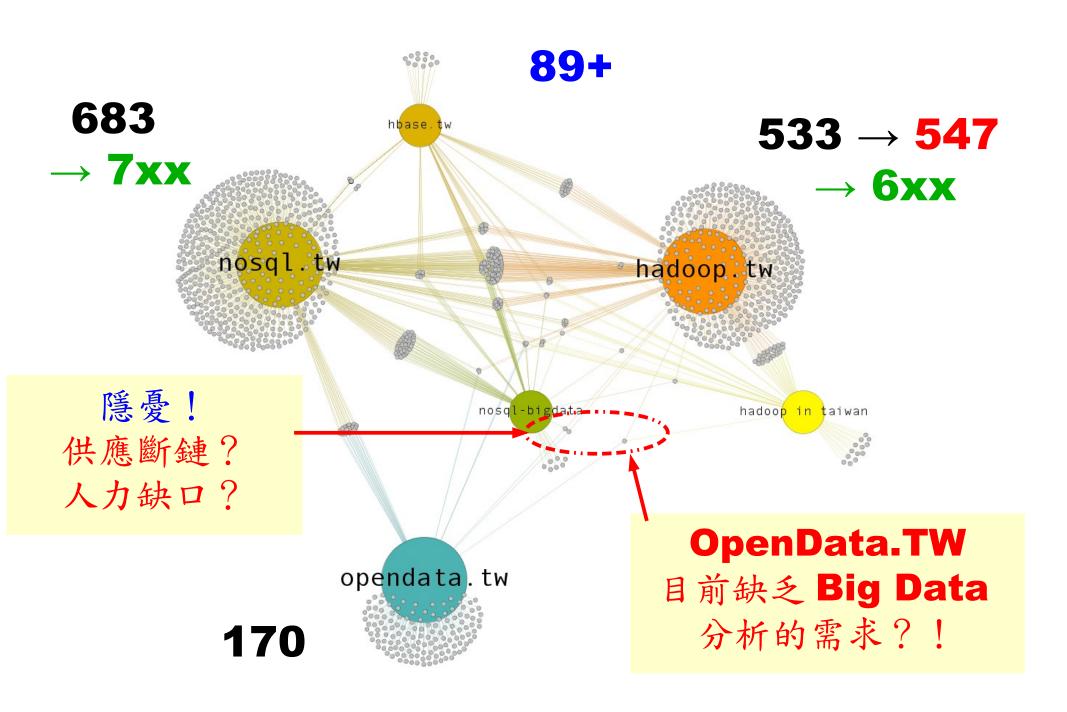
相關社群的群聚分析 SNA of Communities



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



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

仍海 有資 料 通產 產業 業 尚 供未 應成 鏈型

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



個人化的巨量資料

Big Data Can Get Very Personal







Senoiteeup

Slides - hithethireencheorghiv/cloud



