

台灣 Hadoop 社群發展現況與未來展望

Hadoop.TW : Now and Future

Jazz Yao-Tsung Wang

王耀聰 <jazz@nchc.org.tw>

Co-founder of Hadoop.TW

Free Software Lab, STD, NCHC, NARL, Taiwan

2013/3/27 @ Big Data Taiwan 2013



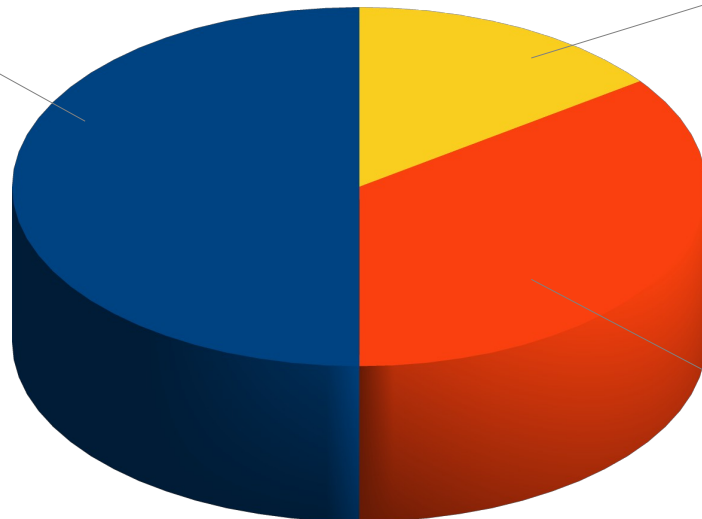
WHO AM I ? 這傢伙是誰啊 ? JAZZ ?

- 講者介紹：
 - 國網中心 王耀聰 副研究員 / 交大電控八九級碩士
 - 今天我是以台灣 Hadoop 社群傳教士的身份發言
 - 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/Ubuntu
- 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

◆ 過去 **PAST**

- **NCHC Cloud Research Group (2008.1~Now)**
- **Taiwan Hadoop User Group (2008.8~Now)**
- **Hadoop in Taiwan (2012.10~Now)**

◆ 現在 **NOW**

- 巨量資料的奇幻漂流 **Life of Big Data**
- 相關社群的群聚分析 **SNA of Communities**

◆ 未來 **FUTURE**

- 從產業供應鏈的觀點 **Supply Chain of Big Data**
- 從社群談未來的展望 **Actions driven from SNA**

Let's Go Back to Year 2007

孽緣？良緣？

這一切都從那一晚開始

2007-10-08

Source: http://googlepress.blogspot.tw/2007/10/google-and-ibm-announce-university_08.html (2007-10-08)

Source: <http://www.nytimes.com/2007/10/08/technology/08cloud.html> (2007-10-08)

Let's Go Back to Year 2007



News from Google

Search this site



[About Google](#) > [News from Google](#) > [News announcements](#) > [News announcement](#)

[News from Google](#)

[Images and B-roll](#)

[Blog directory](#)

[Google+ directory](#)

[Twitter directory](#)

[Facebook directory](#)

[YouTube directory](#)

Google and IBM Announce University Initiative to Address Internet-Scale Computing Challenges [Subscribe](#)

Mountain View, Calif., and Armonk, N.Y. – October 8, 2007 – Google and IBM today announced an initiative to promote new software development methods which will help students and researchers address the challenges of internet-scale applications in the future.

The goal of this initiative is to improve computer science students' knowledge of highly parallel computing practices to better address the emerging paradigm of large-scale distributed computing. IBM and Google are teaming up to provide hardware, software and services to augment university curricula and expand research horizons. With their combined resources, the companies hope to lower the financial and logistical barriers for the academic community to explore this emerging model of computing.

The University of Washington was the first to join the initiative. A small number of universities will also pilot the program, including Carnegie Mellon University, Massachusetts Institute of Technology, Stanford University, the University of California at Berkeley and the University of Maryland. In the future, the program will be expanded to include additional researchers, educators and scientists.]

"Google is excited to partner with IBM to provide resources which will better equip students and researchers to address today's developing computational challenges," said Eric Schmidt, CEO of Google. "In order to most effectively serve the long-term interests of our users, it is imperative that students are adequately equipped to harness the potential of modern computing systems and for researchers to be able to innovate ways to address emerging problems."

Fundamental changes in computer architecture and increases in network capacity are encouraging software developers to take new approaches to computer-science problem solving. For web software such as search, social networking and mobile commerce to run quickly, computational tasks often need to be broken into hundreds or thousands of smaller pieces to run across many servers simultaneously. Parallel programming techniques are also used for complex scientific analysis such as gene sequencing and climate modeling.

Source: http://googlepress.blogspot.tw/2007/10/google-and-ibm-announce-university_08.html (2007-10-08)

Source: <http://www.nytimes.com/2007/10/08/technology/08cloud.html> (2007-10-08)

Let's Go Back to Year 2007



News from Google

Search this site



About Google > News from Google > News announcements > News announcement

News from Google

Images and B-roll

Blog directory

Google+ directory

Twitter directory

Facebook directory

YouTube directory

Google and IBM Announce University Initiative to Address Internet-Scale Computing Challenges [Subscribe](#)

Mountain View, Calif., and Armonk, N.Y. – October 8, 2007 – Google and IBM today announced an initiative to promote new software development methods which will help students and researchers address the challenge of internet-scale applications in the future.

HOME PAGE | MY TIMES | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS | [Subscribe: Digital / Home Delivery](#) | [Log In](#) | [Register Now](#)

The New York Times

Technology

Technology All NYT

Search

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | **JOB**S | REAL ESTATE | AUTOS

CIRCUITS | CAMCORDERS | CAMERAS | CELLPHONES | COMPUTERS | HANDHELDS | HOME VIDEO | MUSIC | PERIPHERALS | WI-FI | DOWNLOADS

Google and I.B.M. Join in 'Cloud Computing' Research


By STEVE LOHR
Published: October 8, 2007

Even the nation's elite universities do not provide the technical training needed for the kind of powerful and highly complex computing [Google](#) is famous for, say computer scientists. So Google and [I.B.M.](#) are announcing today a major research initiative to address that shortcoming.

The two companies are investing to build large data centers that students can tap into over the Internet to program and research remotely, which is called "cloud computing."

 TWITTER

 LINKEDIN

 SIGN IN TO
E-MAIL OR SAVE
THIS

 PRINT

 REPRINTS

 SHARE

[More Articles in Technology »](#)

MOST POPULAR - TECHNOLOGY

E-MAILED | BLOGGED | VIEWED

1. State of the Art: New Reasons to Change Light Bulbs
2. A Strange Computer Promises Great Speed
3. Pogue's Posts: Addressing Fears About LED Light Bulbs
4. App Smart: Quickly Zeroing In on New Places to Call Home
5. 60 Seconds With Pogue: LED Light Bulbs
6. Bits: Daily Report: New Reasons to Change Light Bulbs
7. Gadgetwise: Who Has The Fastest LTE Service?

Source: http://googlepress.blogspot.tw/2007/10/google-and-ibm-announce-university_08.html (2007-10-08)

Source: <http://www.nytimes.com/2007/10/08/technology/08cloud.html> (2007-10-08)

From Data Grid to Cloud Computing

2008/01/28 trac.nchc.org.tw 開張



Wiki Timeline

Timeline

01/28/08:

18:58 WikiStart created by jazz

01/21/08:

19:04 Changeset [1] by jazz
* Initial Version

Note:

Down



Powered by Trac 0.11.7
By Edgewall Software.

2008/01/28



Login Preferences Help/Guide About Trac Forgot your password?

Wiki Timeline Roadmap Browse Source View Tickets Search

Start Page Index History Last Change

Welcome to NCHC Cloud Computing Research Group

【Project News】

倘若貴單位有關雲端運算、海量資料處理相關技術諮詢、教育訓練需求或需要搭建 Hadoop 系統，歡迎來信洽詢：[jazz @ nchc . org . tw](mailto:jazz@nchc.org.tw) ~

註：三小時左右的歸類於演講 (Speech)，三小時以上的歸類於課程 (Training Course)

[【Project News】](#)
[【Training Course】](#)
[【Speech / Workshop】](#)
[【Comittie】](#)
[【Related News】](#)
[【Hadoop News】](#)
[【HyperTable News】](#)
[【Related Conference】](#)
[【知識分享】](#)
Visitor

2013-03-10	Taiwan Hadoop Ecosystem Workshop 2013 Q1	* 活動名稱: Taiwan Hadoop Ecosystem Workshop 2013 Q1 * 活動時間: 2013/3/10(日) 10:00~16:00 * 活動地點: 精誠內湖大樓 (台北市內湖區瑞光路 318號) 1F * 報名網址: https://www.facebook.com/events/456789427727154
------------	--	---

<http://trac.nchc.org.tw/grid> , <http://trac.nchc.org.tw/cloud>



From Data Grid to Cloud Computing

**Era of
Grid Computing**

Data Grid
Cluster File System

Amazon :
Virtualization

Xen, KVM,
Eucalyptus,
OpenNebula

Google / IBM :

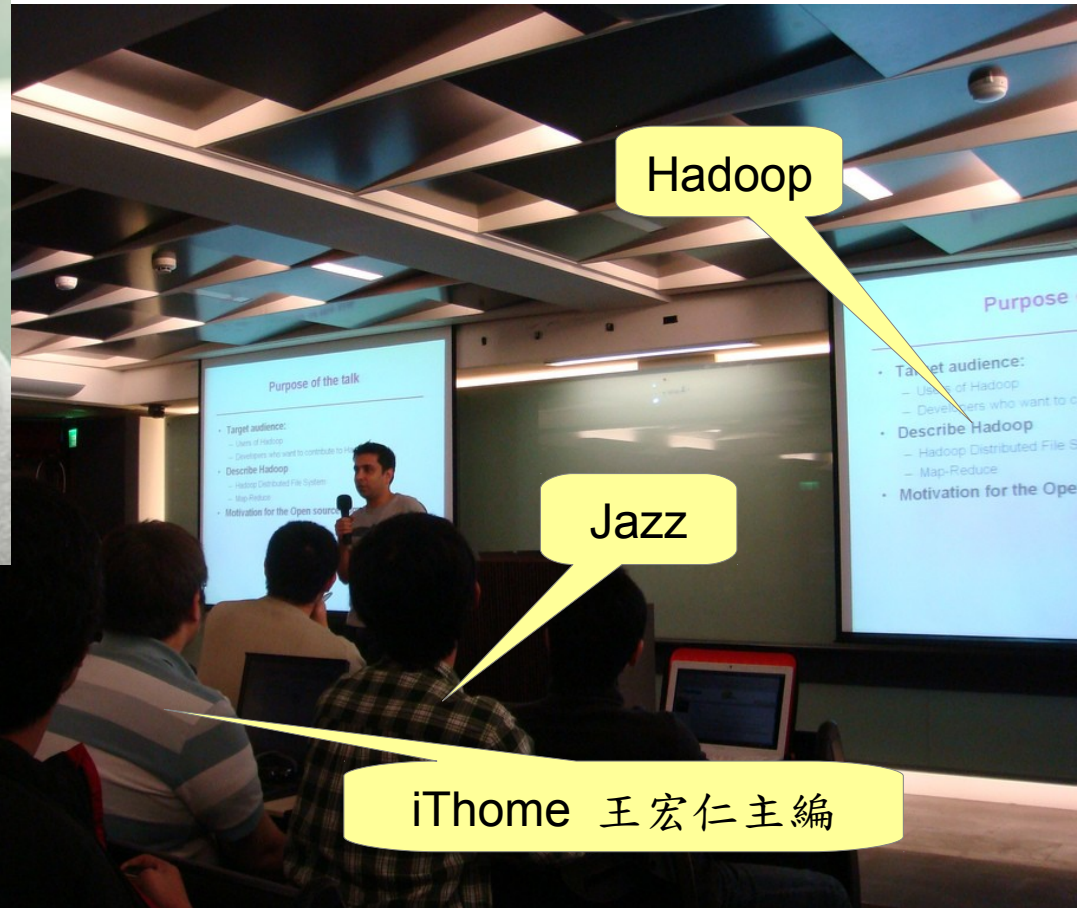
Hadoop?!
BigTable,
HBase,
HyperTable

Era of Cloud Computing ?

2008/4/12 OSDC.TW 2008 - Vivek Ratan



Source: http://farm4.staticflickr.com/3185/2406795223_9ed1366c9c_z.jpg
Source: http://farm3.staticflickr.com/2022/2407601434_cf9c65eb13_b.jpg
Source: http://blog.osdc.tw/2008/03/speaker_vivek_ratan.html
Source: http://blog.osdc.tw/2008/03/talk_hadoop.html



Vivek Ratan
Yahoo! (2008)
Amazon (Now)

2008/04/12
2008/01/28



2008/4/28 hadoop.tw 註冊 By 蔡奕楷

```
jazz@jazzbook:~$ whois hadoop.tw
Domain Name: hadoop.tw
```

Contact:

```
Record expires on 2013-04-29 (YYYY-MM-DD)
Record created on 2008-04-28 (YYYY-MM-DD)
```

Registration Service Provider:

嘗試邀請 **Vivek**
再度訪台 (X)

2008/10/16 加入
2008/12/08 移機

2008/01/28
2008/04/12
2008/04/28
2008/08/16

蔡奕楷 **Yi-Kai Tsai**

Yahoo! TW (Now)

<http://www.hadoop.tw>

2008/08/16 開張

Hadoop Taiwan User Group

August 2008 Archives

What is Hadoop ?

By [yikai](#) on August 16, 2008 5:17 PM | [Permalink](#) | [Comments \(0\)](#) | [TrackBacks \(0\)](#)

Hadoop 是 [Apache 軟體基金會](#) 旗下的一個計畫,是為了開發一套提供可靠性(reliable),延展性(scalable)以及分散式計算(distributed computing)的開源軟體,目前包括下列子計畫:

- **Hadoop 核心**, 目前的旗艦子計畫, 提供分散式檔案系統 (HDFS) 並支援 [MapReduce](#) 分散式計算模式。
- **HBase**, 依據 Hadoop 核心打造, 提供具有延展性的分散式資料庫系統。
- **ZooKeeper**, 一套高可靠度的協調管理系統, 分散式應用軟體使用 ZooKeeper 來保存, 協調重要的共享資料更新。

Hadoop 原本是為了 [Nutch](#), 一個 Apache 軟體基金會旗下的搜尋引擎系統所設計打下的基礎建設 Framework, 之後 Google 相繼發布了 [Google File System](#) 和 MapReduce 等 papers, 它們正是 Nutch 計畫所需要的, 因此 Nutch 計畫新增了 HDFS 跟 MapReduce 運算模式的實作。



2006 年, Hadoop 之父 [Doug Cutting](#) 應 Yahoo! 之邀加入公司, 專職從事 Hadoop 計畫的開發工作, 而 Hadoop 本身也從 Nutch 計畫正式獨立出來, 2008 三月 Hadoop 升格成為 Apache 軟體基金會的頂級計畫(Top Level Project), 而 Yahoo! 目前是 Hadoop 核心計畫和 ZooKeeper 計畫的主要貢獻者。

Hadoop 核心已經被證實可在 2000 台機器上執行, 目前的設計目標是朝向 10,000 台機器。已經有許多不同的公司和組織分別使用 Hadoop 從事研究或是產品開發的工作, 像是 Yahoo! 搜尋引擎的 Webmap 建立, Amazon 的 [EC2/S3](#) 服務, 甚至 Google 也使用 Hadoop 和學術單位進行合作計畫等。

Sponsored By



2008/11/04 Hadoop @ NCHC - Devaraj Das



Devaraj Das
Yahoo! (2008)
Co-founder of
Hortonworks (Now)

2008/11/04
2008/08/16
2008/04/28
2008/04/12
2008/01/28

國網中心邀請演講資訊

Hadoop 與雲端運算

雲端運算為 2008 年重大 IT 熱門議題，而 Hadoop 為 Apache Software Foundation 所開發之自由軟體，目前已廣泛應用於 Amazon 與 Yahoo! 等雲端運算服務提供者的格網架構之上。

Devaraj Das 是 Yahoo! Bangalore Grid Computing Group 的 Engineering Manager，亦為 Apache Committer，對於 Hadoop 有多年的開發經驗。此外，Yahoo! Bangalore Grid Computing Group 著重於如何打造足以處理 Peta-bytes 資料，由數千台主機組成的格網架構，將帶給中心從事格網相關研究的同仁來自於產業界的開發經驗分享。

講者簡歷

Devaraj Das (ddas@...) is the Engineering Manager of the Grid Computing group at Yahoo! Bangalore. He graduated with a Masters degree in Computer Science from Indian Institute of Science, Bangalore. Prior to Yahoo!, Devaraj was with HP. Devaraj is an Apache committer.

The Grid Computing group at Yahoo! Bangalore focuses on Grid frameworks that scale to thousands of machines and handle peta-bytes of data. The group is especially involved in the development of the Open Source Hadoop platform and its deployment within Yahoo!.

2008-11-04 上午

- 時間: 11/04 星期二 上午 11:00 - 12:30
- 地點: 國家高速網路與計算中心 北群多媒體教室 (新竹市研發六路七號) |

Source: <http://trac.nchc.org.tw/cloud/wiki/HadoopWorkshop>

國網中心邀請演講資訊

Hadoop 與雲端運算

講者簡歷

2008-11-04 上午

演講摘要

2008-11-04 下午

實作練習摘要

與會須知

2008-11-04 下午

實作練習摘要

與會須知

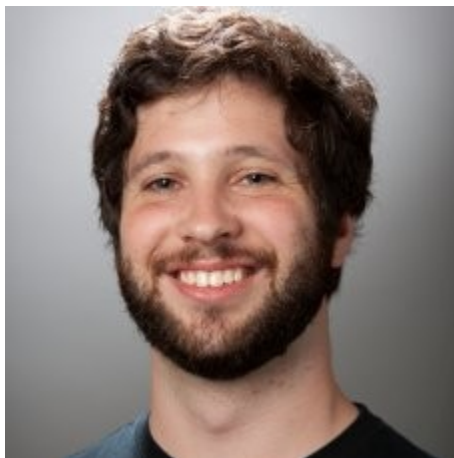
2008/11/03 開始實驗 **DRBL** 與 **Hadoop** 結合的方法
<BUG> HDFS 無法識別 **NFS** 的空間，也因為這段經驗
造就了 **2009/4/13 Hadoop** 雲端運算實驗平台的誕生



2009/4/1 Connected with Cloudera



2008/12/19 寫信給 **Tom White** 說要翻譯中文版
2009/03/30 **NCHC** 開辦第一次 **Hadoop** 課程
2009/04/01 **Tom White** 介紹 **Cloudera VP**
Christophe Bisciglia 與 **Todd Lipcon** 給我認識



Tom White
Cloudera (Now)
Hadoop 技術手冊
原作者

Todd Lipcon
Cloudera (Now)

Christophe Bisciglia
Cloudera (2009)
WibiData (Now)

2009/03/27 Hadoop.TW 釋出 0.18 DEB 套件
Source: <http://www.hadoop.tw/2009/03/-hadoop-0183-debian-ubuntu.html>

2009/04/01
2008/11/04
2008/08/16
2008/04/28
2008/04/12
2008/01/28



2009/4/13 開放 hadoop.nchc.org.tw 2009/11/10 第一屆台灣 Hadoop 使用者社群會議



```
jazz@hadoop:~$ apt-cache show hadoop-0.20
Package: hadoop-0.20
Status: install ok installed
Priority: extra
Section: misc
Installed-Size: 42320
Maintainer: Todd Lipcon <todd@cloudera.com>
Architecture: all
Version: 0.20.1+169.113-1~lenny-cdh2
```

**NCHC DRBL +
Cloudera CDH2 DEB**

- 2009/1/1/19
- 2009/1/1/10
- 2009/04/13
- 2008/11/04
- 2008/08/16
- 2008/04/28
- 2008/04/12
- 2008/01/28

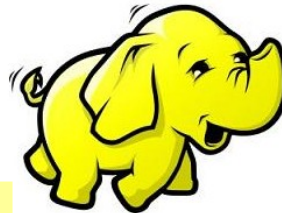
2009/04/13 1st Hadoop Public Cluster
<http://hadoop.nchc.org.tw>

2009/11/10 1st Taiwan Hadoop User Group
 2009/11/19 Taiwan Hadoop Forum
<http://forum.hadoop.tw/>

台灣 Hadoop 使用者社群會議

[「雲端運算與應用」系列演講\(一\)>>](#)

感謝各界熱烈報名，也感謝親臨現場的同好們抽空一同共襄盛舉。
 此次活動的投影片已公佈於 <http://www.hadoop.tw>，歡迎前往下載。



**Hadoop Taiwan
User Group**

<http://www.hadoop.tw>

雲端運算於海洋資訊之前瞻應用研習暨
 第一屆台灣 Hadoop 使用者社群會議

Cloud Computing for Advanced Ocean Research
 Hadoop Taiwan User Group Meeting 2009

- 會議時間：九十八年十一月十日（週二）
- 會議地點：[國家高速網路與計算中心](#) 國際會議廳

活動開始時間

2009/11/10 09:00

活動結束時間

2009/11/10 17:00

地點名稱

國家高速網路與計算中心（新竹市研發六路七號）

活動人數上限

150

**Keynote by
Christophe Bisciglia**

Like 13 people like this.
 Be the first of your friends.

<http://registrano.com/events/hadoop-tw>



2010/1/28 Connected with Andrew Purtell



Andrew Purtell
Apache HBase
Trend Micro (2010)
Intel (Now)

2010/04/26
2010/01/28
2009/11/19
2009/11/10
2009/04/13
2009/04/01
2008/11/04
2008/08/16
2008/04/28
2008/04/12
2008/01/28

2010/01/28 Connected with Andrew Purtell 2010/04/26 Apache HBase Talk

「雲端運算與應用」系列演講(五)

雲端分散式資料庫 Apache HBase

"Cloud Computing and its Application" lecture series:

Apache HBase

- 會議時間：九十九年四月二十六日（週一）下午一點三十分
- 會議地點：國家高速網路與計算中心中部事業群 3D 劇場
- （[40763 台中市中部科學工業園區科園路22號](#)）
- 視訊地點：國家高速網路與計算中心 [新竹事業群國際會議廳](#)
- 國家高速網路與計算中心 [南部事業群國際會議廳](#)
- 主辦單位：[國家高速網路與計算中心](#)
- 報名網址：<http://registrano.com/events/cloudtalk20100426>
- 轉播網址：<http://cloud.nchc.org.tw>
- 聯絡電話：王耀聰 (04)2462-0202 # 834

資料庫應用對於資訊產業已是不容或缺的一環，但隨著資料巨幅地成長，傳統關聯式資料庫將面臨無法持續擴充的效能瓶頸。知名雲端服務提供者谷歌（Google）針對大量資料的儲存、分析與查詢，分別提出 Google File System 分散式檔案系統、MapReduce 演算法與 Big Table 分散式資料庫。Apache 基金會的 Hadoop 專案實作了前兩者，而 HBase 則基於 Hadoop，提供分散式資料庫的解決方案。此次演講很榮幸邀請到現任趨勢科技資深軟體架構師，也是 HBase 專案關鍵貢獻者(Committer)的 Andrew Purtell 來為大家介紹 Apache HBase 專案的近期發展，歡迎對於雲端分散式資料庫系統感興趣的同好踴躍參加。

<http://registrano.com/events/cloudtalk20100426>

地點名稱

國家高速網路與計算中心
台中市科學園區科園路
22號

活動人數上限

150

主辦單位

[國家高速網路與計算中心](#)

Tweet <0

Like Be the first of your friends to like this.

Google Calendar

加至 Outlook 行事曆

2010/4/12 Connected with Alex Loddengaard



Alex Loddengaard
Cloudera (2010)
Co-Founder of
MemCachier (Now)

2010/04/26
 2010/01/28
 2009/11/19
 2009/11/10
 2009/04/13
 2009/04/01
 2008/11/04
 2008/08/16
 2008/04/28
 2008/04/12
 2008/01/28

2010/03/10 Cloudera Certification @ NYC 2010/04/12 Cloudera Training in Taiwan

0917527812396732001	Event	Hadoop Training for Developers - NYC - March	
	Date+Time	Wednesday, March 10, 2010 at 9:00 AM - Friday, March 12, 2010 at 5:00 PM (ET)	Name Yao-Tsung Wang
	Type	3 Days: Early Bird: Full Program + Certification \$0.00	Payment Status Free Order
	Location	Learning Tree - NYC One New York Plaza 31st Floor New York, NY 10004	
	Order Info	Ordered by Yao-Tsung Wang on Feb 07, 2010 at 8:52 AM	

Eventbrite

Create an event

Find events

Email

Share

Tweet

Like

2

Questions? Contact the organizer

Hadoop Training for Developers - Taipei - April

Cloudera, Inc

Monday, April 12, 2010 at 9:00 AM - Wednesday, April 14, 2010 at 5:00 PM (CST)

Taipei, Taiwan



Ticket Information

TICKET TYPE	SALES END	PRICE	FEE	QUANTITY
Early Bird: Full Program (3 Days) + Certification - Regular / Invoice Price: 59.000 TWD	Ended	\$888.00	\$0.00	N/A
Full Program (3 Days) + Certification - Regular / Invoice Price: 64.000 TWD	Ended	\$998.00	\$0.00	N/A

When & Where

台北市大安區金山南路2段52號1F

Taipei, T'ai Pei
Taiwan

Monday, April 12, 2010 at 9:00 AM -
Wednesday, April 14, 2010 at 5:00 PM (CST)

Add to my calendar



2010/12/02 第二屆台灣 Hadoop 使用者社群會議

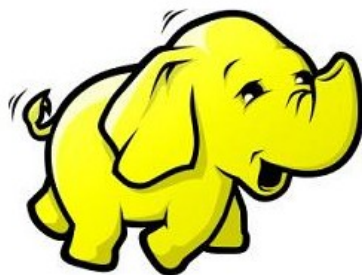


2010/10/20 hadoop.nchc.org.tw upgrade
2010/12/02 2nd Taiwan Hadoop User Group

2010 第二屆台灣 Hadoop 使用者社群會議

<<「雲端運算與應用」系列演講(五) | 第三屆台灣 Hadoop 使用者社群會議 >>

部分願意公開投影片與錄影的內容，請至 cloud.nchc.org.tw 下載。



Hadoop Taiwan
User Group

<http://www.hadoop.tw>

第二屆台灣 Hadoop 使用者社群會議

Hadoop Taiwan User Group Meeting 2010

- 會議時間：九十九年十二月二日（週四）
- 會議地點：[國家高速網路與計算中心 立體劇場](#)
（[台中市科園路 22 號](#) - [詳細交通指引](#)）
- 主辦單位：[國家高速網路與計算中心](#)

活動開始時間

2010/12/02 09:00

活動結束時間

2010/12/02 17:00

地點名稱

國家高速網路與計算中心
台中市科園路 22 號

活動人數上限

180

主辦單位

[國家高速網路與計算中心](#)

Tweet 1

Like

7 people like this.
Be the first of your friends.

2010/12/02
2010/04/26
2010/01/28
2009/11/19
2009/11/10
2009/04/13
2009/04/01
2008/11/04
2008/08/16
2008/04/28
2008/04/12
2008/01/28

<http://registrano.com/events/hadoop-tw-2010>



2011/12/05 第三屆台灣 Hadoop 使用者社群會議

全球景氣不佳，
很難邀國際講者

曾嘗試邀請
KarmaSphere
HortonWorks

最後要感謝：

中華電信
EMC Greenplum
Armorize
Trend Micro

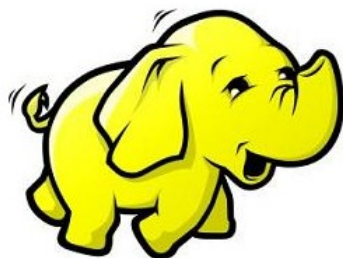
2011/12/05
2010/12/02
2010/04/26
2010/01/28
2009/11/19
2009/11/10
2009/04/13
2009/04/01
2008/11/04
2008/08/16
2008/04/28
2008/04/12
2008/01/28

2011/12/05 3rd Taiwan Hadoop User Group

2011 第三屆台灣 Hadoop 使用者社群會議

<< [第二屆台灣 Hadoop 使用者社群會議](#) | >>

部分願意公開的投影片內容，請至 cloud.nchc.org.tw 下載。



Hadoop Taiwan
User Group

<http://www.hadoop.tw>

第三屆台灣 Hadoop 使用者社群會議
Hadoop Taiwan User Group Meeting 2011

- 會議時間：民國一百年十二月五日（週一）
- 會議地點：中華電信訓練所板橋本所綜合大樓七樓會議室
（板橋市民族路 168 號 - 詳細 [交通指引](#)）
- 主辦單位：[國家高速網路與計算中心](#)
- 協辦單位：[中華電信訓練所](#)、[EMC Greenplum](#)
- 贊助單位：[CIO IT 經理人雜誌](#)、[iThome 電腦報週刊](#)

活動開始時間

2011/11/09 09:00

活動結束時間

2011/12/04 17:00

地點名稱

中華電信訓練所
板橋市民族路168號

活動人數上限

180

主辦單位

[國家高速網路與計算中心](#)



NealLee, Chao-Tung Yang and 57 others like this.



<http://registrano.com/events/hadoop-tw-2011>

2012/10/02 與趨勢合辦 Hadoop in Taiwan 2012



【 全程護航
邁向雲端 】



釋放雲端潛能 駕馭海量資料

活動首頁 精彩議程 講師陣容 報名截止 交通方式

2003及2004年, Google將分散式檔案系統、高度平行運算軟體編程平台等秘訣公諸於世

2006年, Doug Cutting 將上述秘訣寫成開源系統 Hadoop 專案

2008年, Yahoo 宣佈建立了全球最大的商業營運 Hadoop 叢集

2009年, 台灣Hadoop使用者社群會議首度開辦

2011年, 阿帕契基金會正式釋出 Hadoop 1.0 版本, 象徵 Hadoop 已經穩定到足以承載企業營運的需求

2012年, 台灣Hadoop使用者大會正式登場, 宣告Hadoop風潮即將襲捲台灣

國網中心和趨勢科技共同主辦的「Hadoop in Taiwan 2012」, 是Hadoop在台灣首次的大型會議, 精心規劃的議程內容涵蓋前瞻技術和實作應用, 從個人到企業、從資深人士到新進者, 各種角色的技術專才都能從

活動資訊

日期: 2012 / 10 / 02(二)

時間: 08:30~17:00 (08:30 開始報到)

地點: 中央研究院 人文社會科學館4F國際會議廳
(台北市南港區研究院路二段128號)

活動諮詢: 02-2782-5255#14 周小姐



立即報名

2012/10/02
2011/11/2/05
2010/12/02
2010/04/26
2010/01/28
2009/11/19
2009/11/10
2009/04/13
2009/04/01
2008/11/04
2008/08/16
2008/04/28
2008/04/12
2008/01/28

2012/10/02 Hadoop in Taiwan 2012

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



- ◆ **過去 PAST**
 - **NCHC Cloud Research Group (2008.1~Now)**
 - **Taiwan Hadoop User Group (2008.4~Now)**
 - **Hadoop in Taiwan (2012.10~Now)**
- ◆ **現在 NOW**
 - 巨量資料的奇幻漂流 **Life of Big Data**
 - 相關社群的群聚分析 **SNA of Communities**
- ◆ **未來 FUTURE**
 - 從產業供應鏈的觀點 **Supply Chain of Big Data**
 - 從社群談未來的展望 **Actions driven from SNA**

Happy Birthday to Hadoop.TW

2013/4/28 – 5 Years Old !

Happy Birthday to Hadoop.TW



Source: http://postacademic.files.wordpress.com/2011/02/800px-birthday_candles.jpg

5 Years , 10000 Hours = Domain Expert

本文出自天下雜誌最新出版《像三星一樣工作》，更多內容>>

在正式進入學習三星的業務能力之前，必須先仔細思考：「為什麼五年職場生活如此重要？」「因為是剛開始的五年，當然重要吧？」這種毫無頭緒的想法固然沒錯，不過所謂的五年，早已超越了「單純時間的累積」，而具有更特別的意義。

尤其將這「五年的意義」銘記在心，將可展望自己未來另一個五年，甚至十年，與未來的樣貌重疊，藉此將有助於強化穩定性與目標意識。

達到「一萬小時法則」的時間，五年

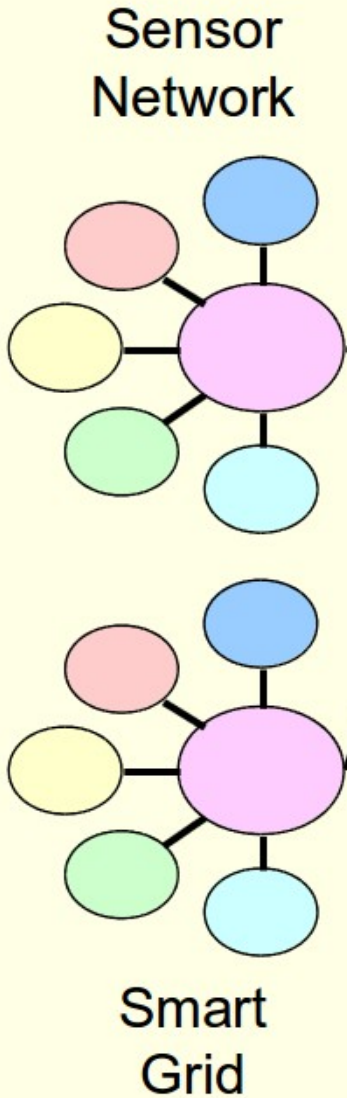
要成為真正的專家，有所謂「一萬小時的法則」之說。一天三小時，持續練習十年的話，將會在該領域達到「通達」的境界，並以其專業能力受到他人的認同。但是如果將這時間從一天的三小時增加到六小時，那麼達到「一萬小時的法則」正好是五年。

每天在職場上的時間，大約是九小時左右。不過扣除午餐時間與休息、閒聊、準備會議、接待客戶、事後業務整理等非主要業務的時間，可以完全投入在業務上的時間，約在六到七小時左右。這六到七小時的時間，便是可以成為「專家」的一萬小時。

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

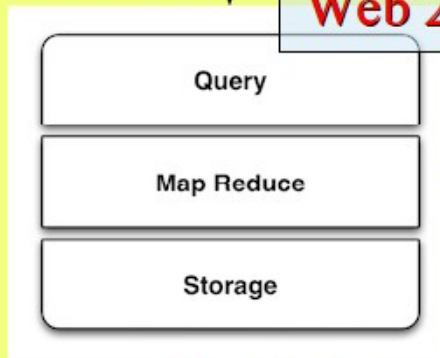


Public Data Hub
Data as a Service

開放資料
Open Data



Web 2.0



Big Data



Mobile Computing

端

各類裝置
存取服務

Cloud Computing
雲端運算

自由軟體社群最重要的是「人」！產業供應鏈也一樣！

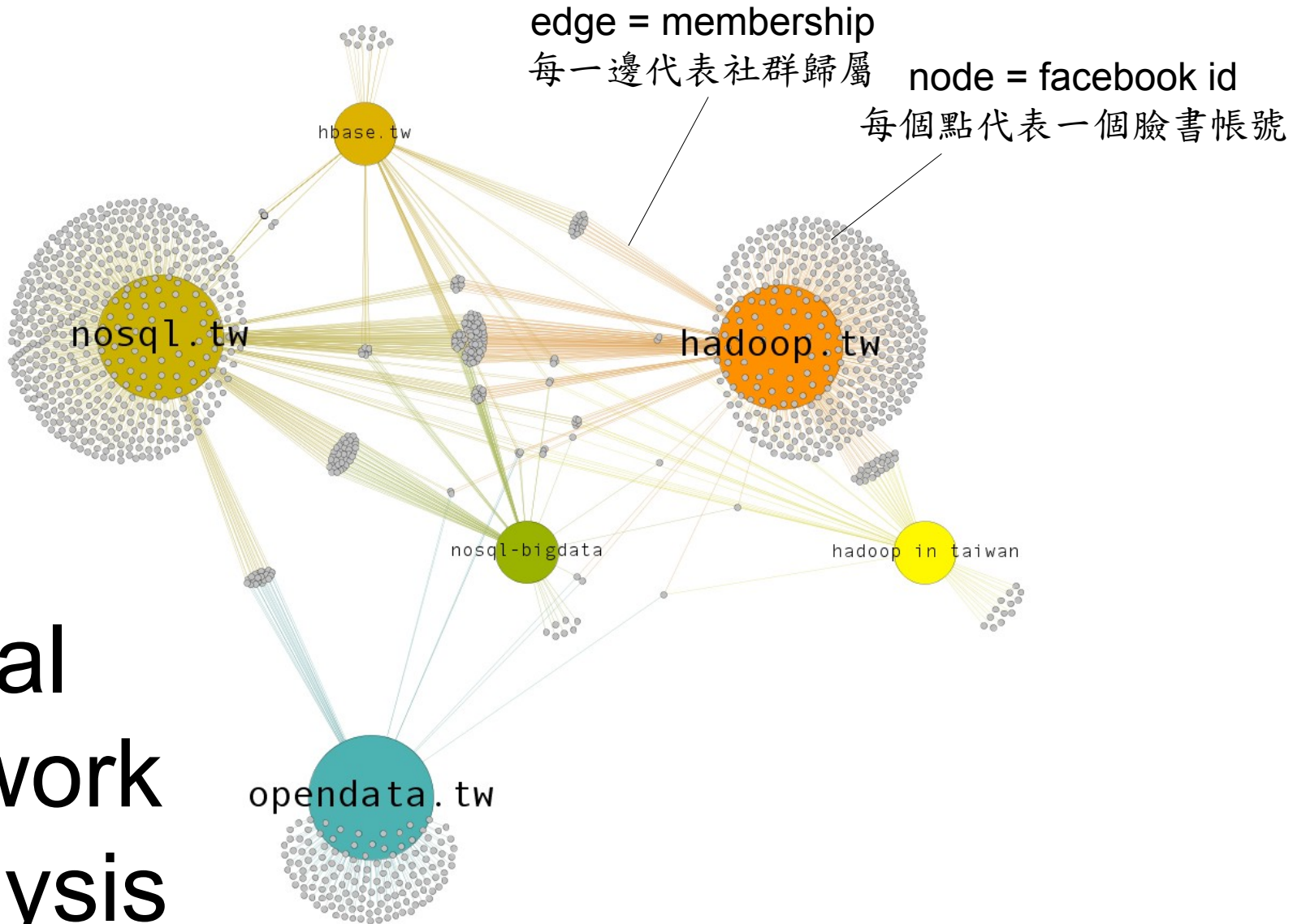
Community Evangelist

Connecting People

Connecting Values

< Mission >

**Connecting People to
build a Value Supply Chain**



Social
Network
Analysis

◆ 過去 **PAST**

- **NCHC Cloud Research Group (2008.1~Now)**
- **Taiwan Hadoop User Group (2008.4~Now)**
- **Hadoop in Taiwan (2012.10~Now)**

◆ 現在 **NOW**

- 巨量資料的奇幻漂流 **Life of Big Data**
- 相關社群的群聚分析 **SNA of Communities**

◆ 未來 **FUTURE**

- 從產業供應鏈的觀點 **Supply Chain of Big Data**
- 從社群談未來的展望 **Actions driven from SNA**

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



雲

資料中心
機房維運



資料中心
提供服務

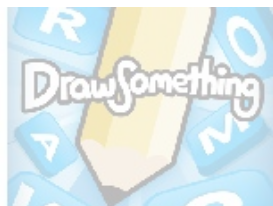
基本硬體
建設組件



廣達電腦
Quanta Computer

Supply Chain of Cloud Computing

應用軟體
供應商



端

行動裝置
共通平台

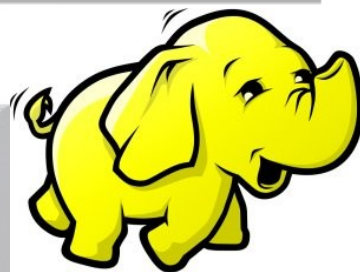


各類裝置
存取服務

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

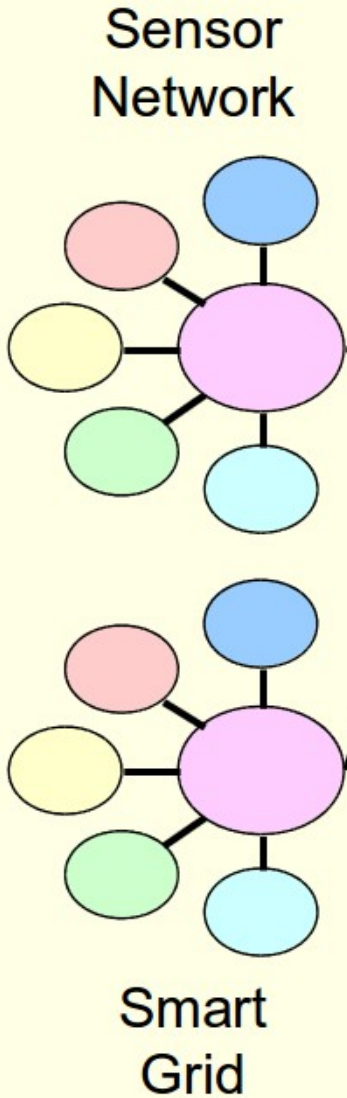


廣達電腦
Quanta Computer

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

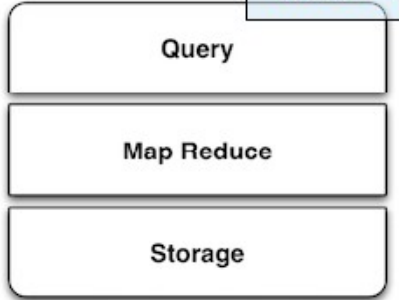


Public Data Hub
Data as a Service

開放資料
Open Data



Web 2.0



Big Data



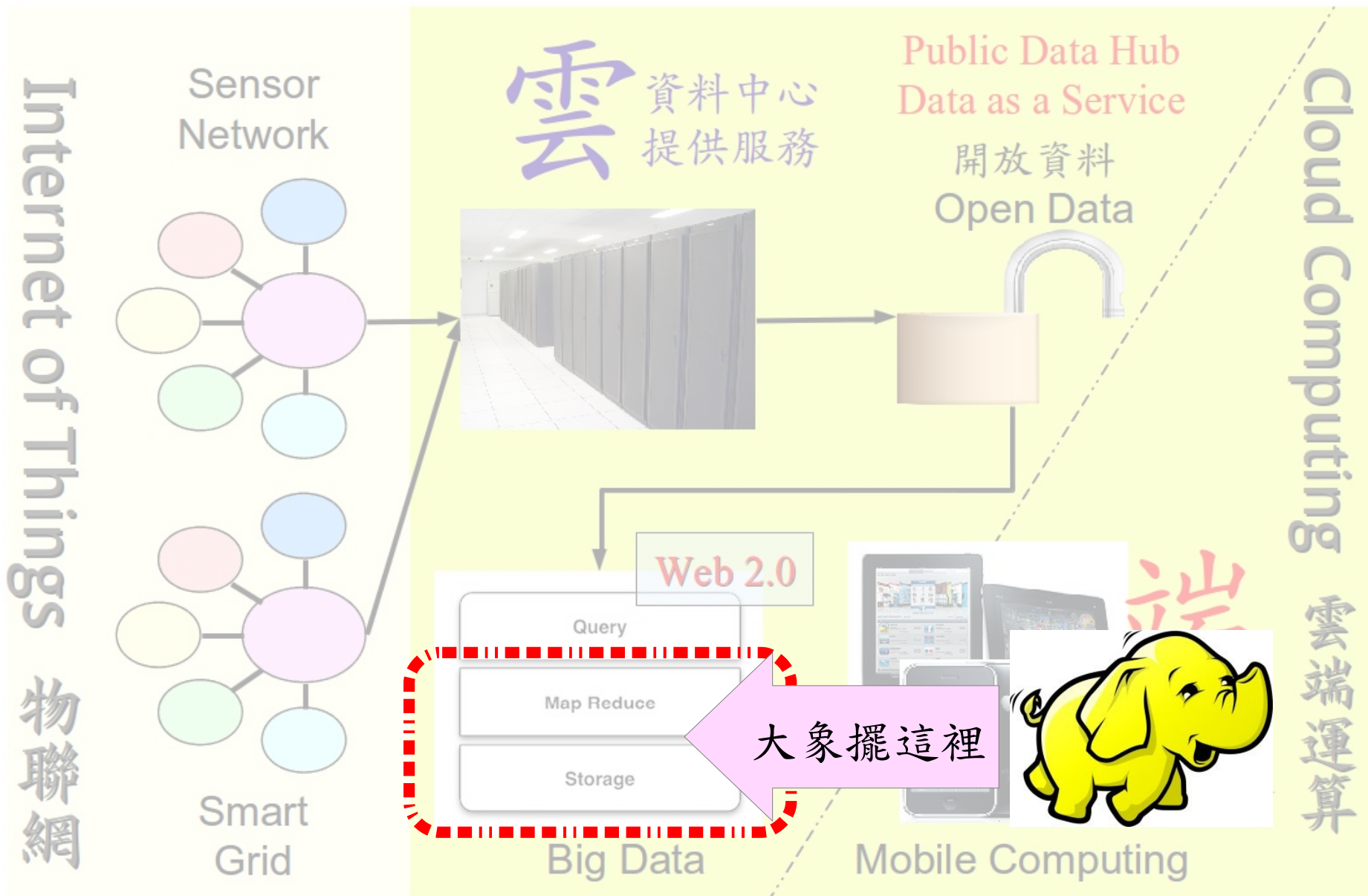
Mobile Computing

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'. At the top, the group name and navigation tabs (About, Events, Photos, Files) are visible. A banner image at the top features several photos of group members. Below the banner, there are options to 'Write Post', 'Add Photo / Video', 'Ask Question', and 'Add File'. The group has 533 members, with 9 new members. A 'Write something...' text box is present. Under 'RECENT POSTS', a post by Jazz Yao-Tsung Wang is shown, discussing his research on Facebook Graph API and Hadoop community connections. Below the post is a network diagram with 'Hadoop.TW' at the center. On the right, a 'People You May Know' section features Steven Chin. The left sidebar lists 'FAVORITES' (News Feed, Messages, Events, Photos, Browse), 'ADS' (Ads Manager), 'PAGES' (Hadoop in Taiwan 2012, Pages Feed, Like Pages), and 'GROUPS' (JavaScript.tw, Taiwan Hadoop User..., S4A).

Taiwan Hadoop User Group

News Feed
Messages
Events
Photos
Browse

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

FAVORITES
News Feed
Messages
Events
Photos
Browse

ADS
Ads Manager

PAGES
Hadoop in Taiwan 2012
Pages Feed
Like Pages

GROUPS
JavaScript.tw
Taiwan Hadoop User...
S4A

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 Notifications button. The main content area features a "Write Post" section with a text input field and options to "Add Photo / Video", "Ask Question", and "Add File". Below this is a "RECENT POSTS" section with a post by Jazz Yao-Tsung Wang. The post text discusses the growth of the Hadoop community in Taiwan and mentions a workshop. The right sidebar shows the group has 59 members (2 new) and an "Add People to Group" button. The left sidebar lists various Facebook features like News Feed, Messages, Events, Photos, and Browse, as well as Ads Manager, Pages, and Groups.

Hadoop in Taiwan About Events Photos Files Notifications

Write Post Add Photo / Video Ask Question Add File

Write something...

59 members (2 new) · Invite by Email

+ Add People to Group

RECENT POSTS

Jazz Yao-Tsung Wang

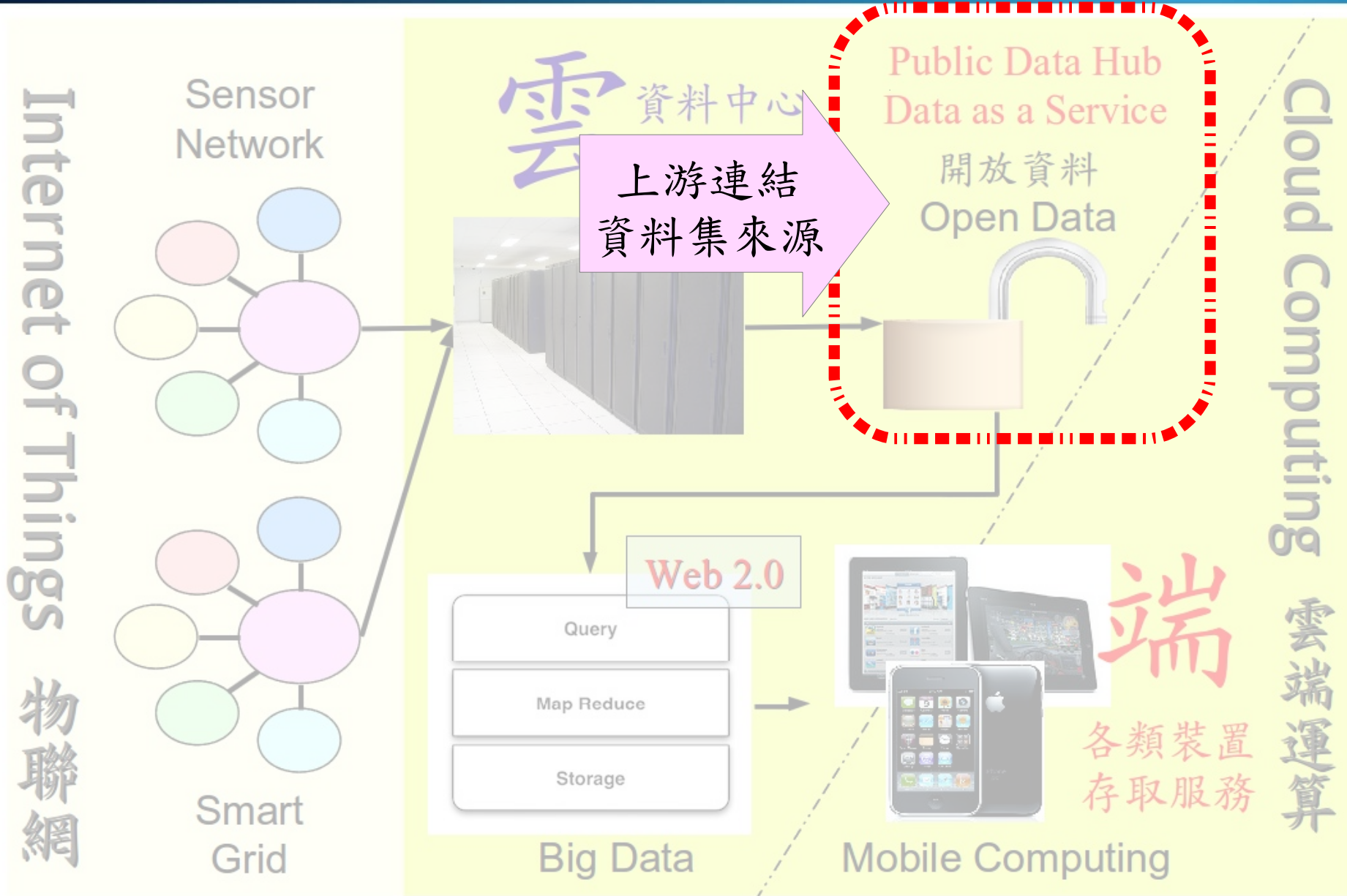
有鑑於今年 Big Data 熱潮持續發酵，Hadoop 社群逐漸茁壯，首先感謝去年趨勢科技的 Rice, 昆錡, Chris Huang 等建議與支持，今年開始舉辦每季一次的 Hadoop 生態系工作坊。其次要特別感謝 Etu 團隊 Fred Chiang 大的鼓勵跟建議，社群必須要開始讓更多人可以擁有強烈的參與感跟使命感。所以在此開始徵求以下工作小組：

(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)

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/>

OpenData / 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...
- S4A 3

OpenData / Taiwan About Events Photos Files Notifications

Write Post Add Photo / Video Ask Question Add File

170 members (7 new) · Invite by Email

+ Add People to Group

Write something...

PINNED POSTS

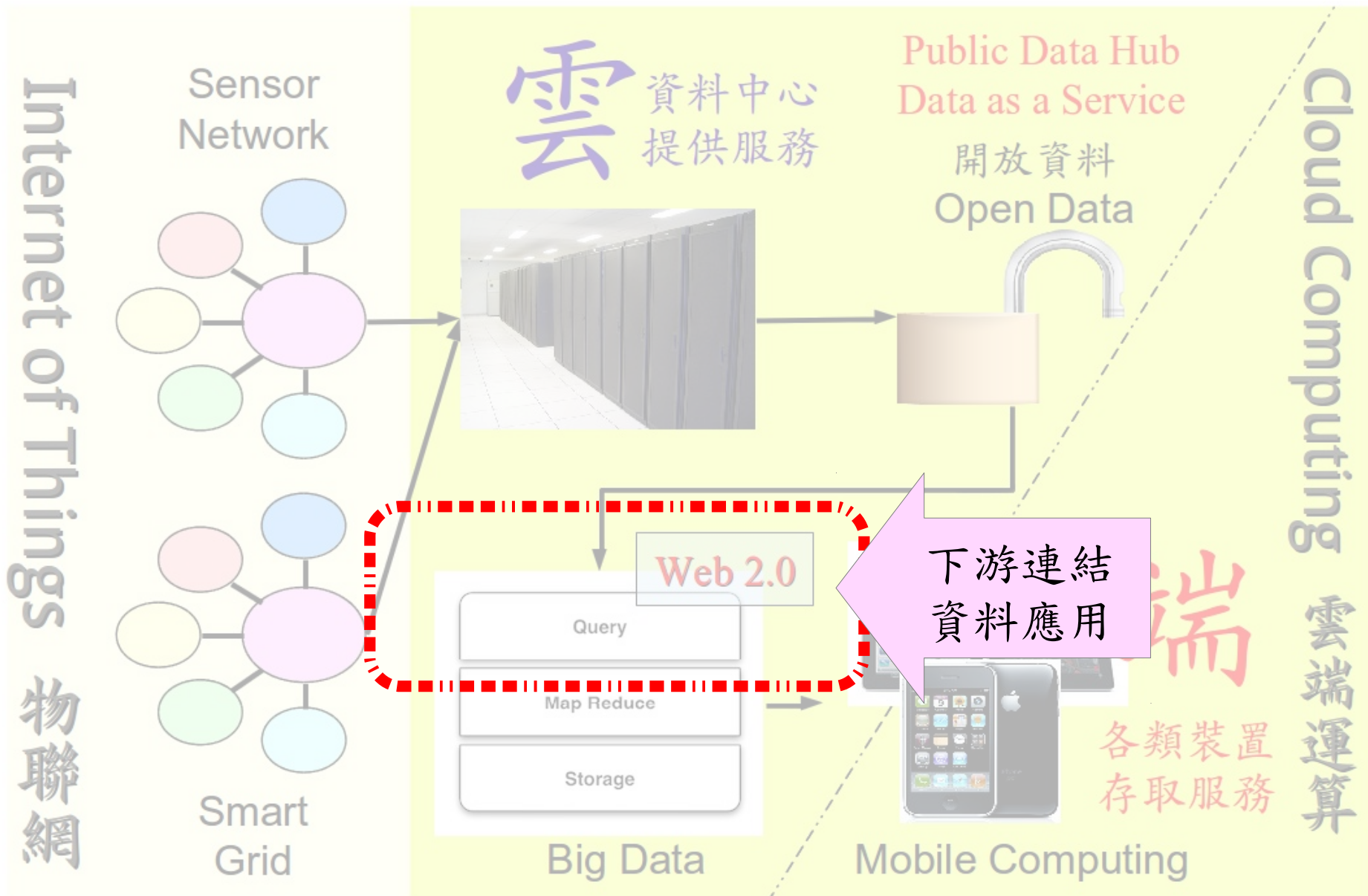
樂樂

Open Data產業新地機系列論壇(第三場)

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...
- 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 a Facebook group page for "NoSQL 中的分散式架構與 Big Data 議題小聚". The page header includes the group name, navigation icons, and a "Home" button. The main content area features a row of profile pictures, a "Write Post" button, and a "Write something..." text input. Below this is a "RECENT POSTS" section with a post by "Kuo-Chun Su" featuring a "PIVOTAL HD" logo and text about "Pivotal HD: The World's Most Powerful Distribution of Apache Hadoop | Greenplum". The left sidebar contains navigation links for "FAVORITES", "ADS", "PAGES", and "GROUPS".

NoSQL 中的分散式架構與 Big Data 議題小聚 About Events Photos Files Notifications

Write Post Add Photo / Video Ask Question Add File

Write something...

92 members (1 new) · Invite by Email

+ Add People to Group

RECENT POSTS

Kuo-Chun Su
<http://www.greenplum.com/products/pivotal-hd>

PIVOTAL HD
Pivotal HD: The World's Most Powerful Distribution of Apache Hadoop | Greenplum
www.greenplum.com
Dynamic Pipelining™ technology delivers 100X performance improvement with mature SQL query optimization and powerful analytics.

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

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... 1
- 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

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

Write something...

114 members · Invite by Email
+ Add People to Group

RECENT POSTS

 **黃健璋**
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 binaries.

2013/03/22 MongoDB.TW 114 members

JavaScript Taiwan Facebook Group

JavaScript.TW

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

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

JavaScript.tw About Events Photos Files **Notifications**

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 參考禮物收集

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

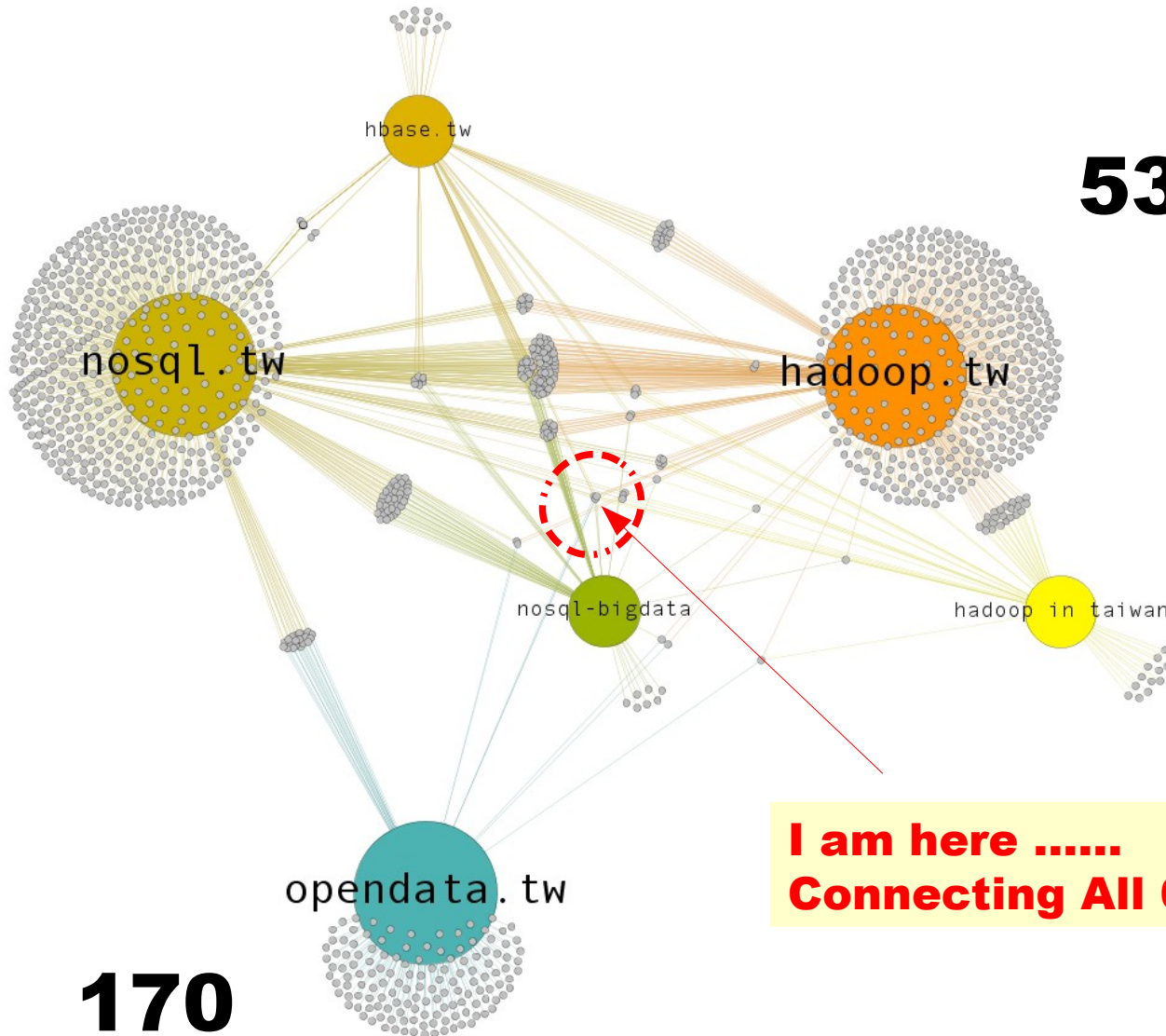
Hadoop Taiwan 533 members

Open Data Taiwan 170 members

Big Data Stack : SMAQ

Social Network Analysis of Communities

683

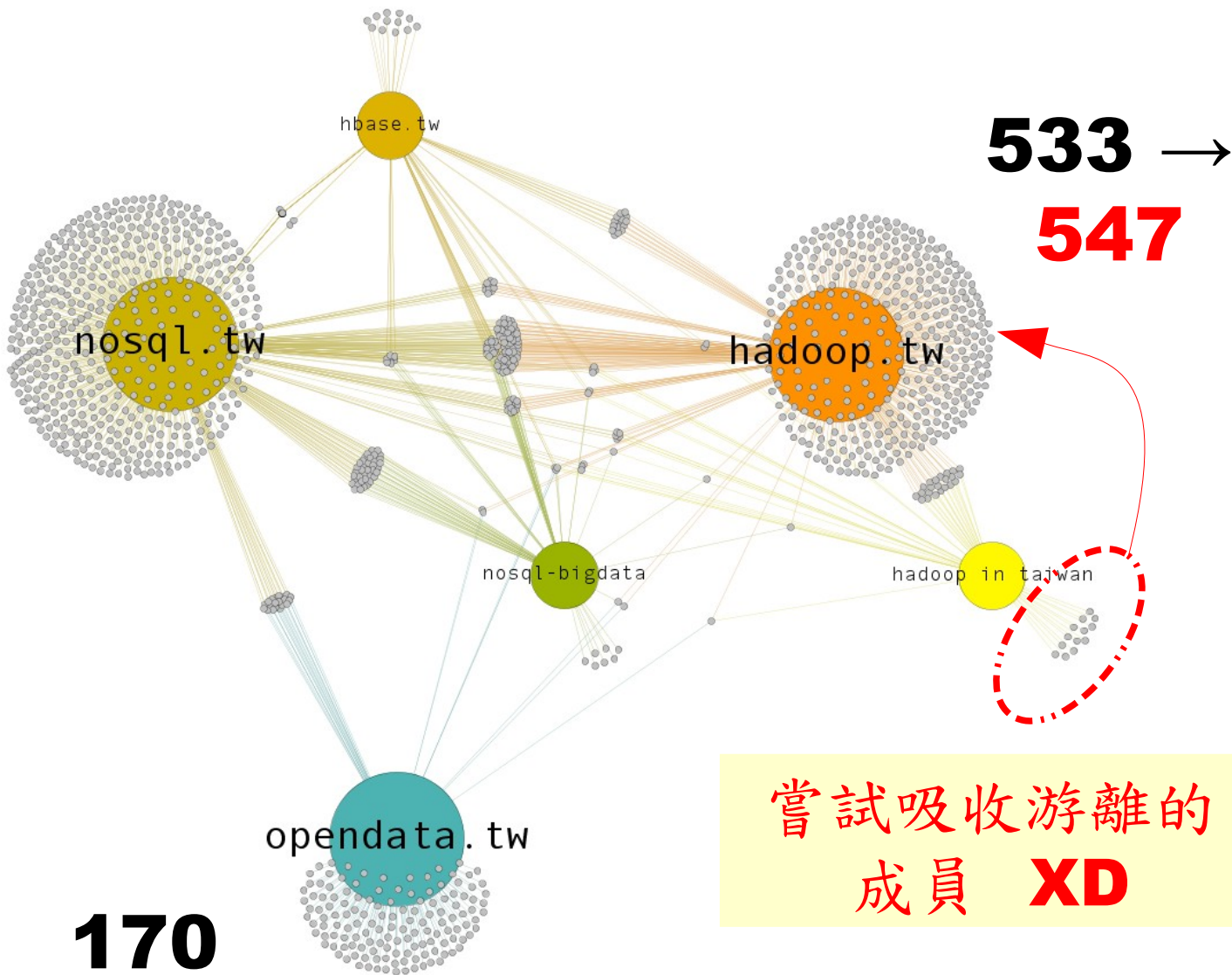


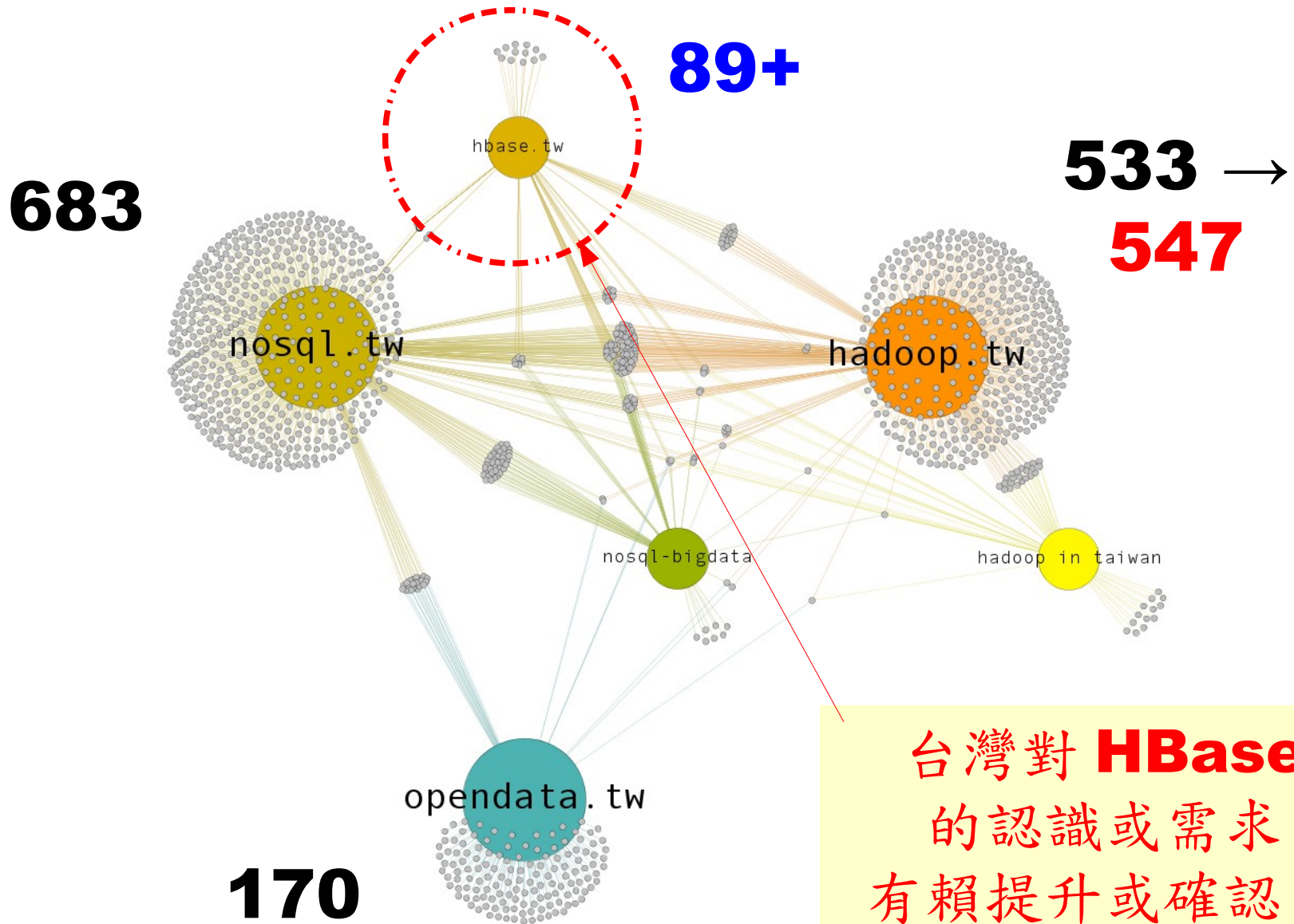
533

170

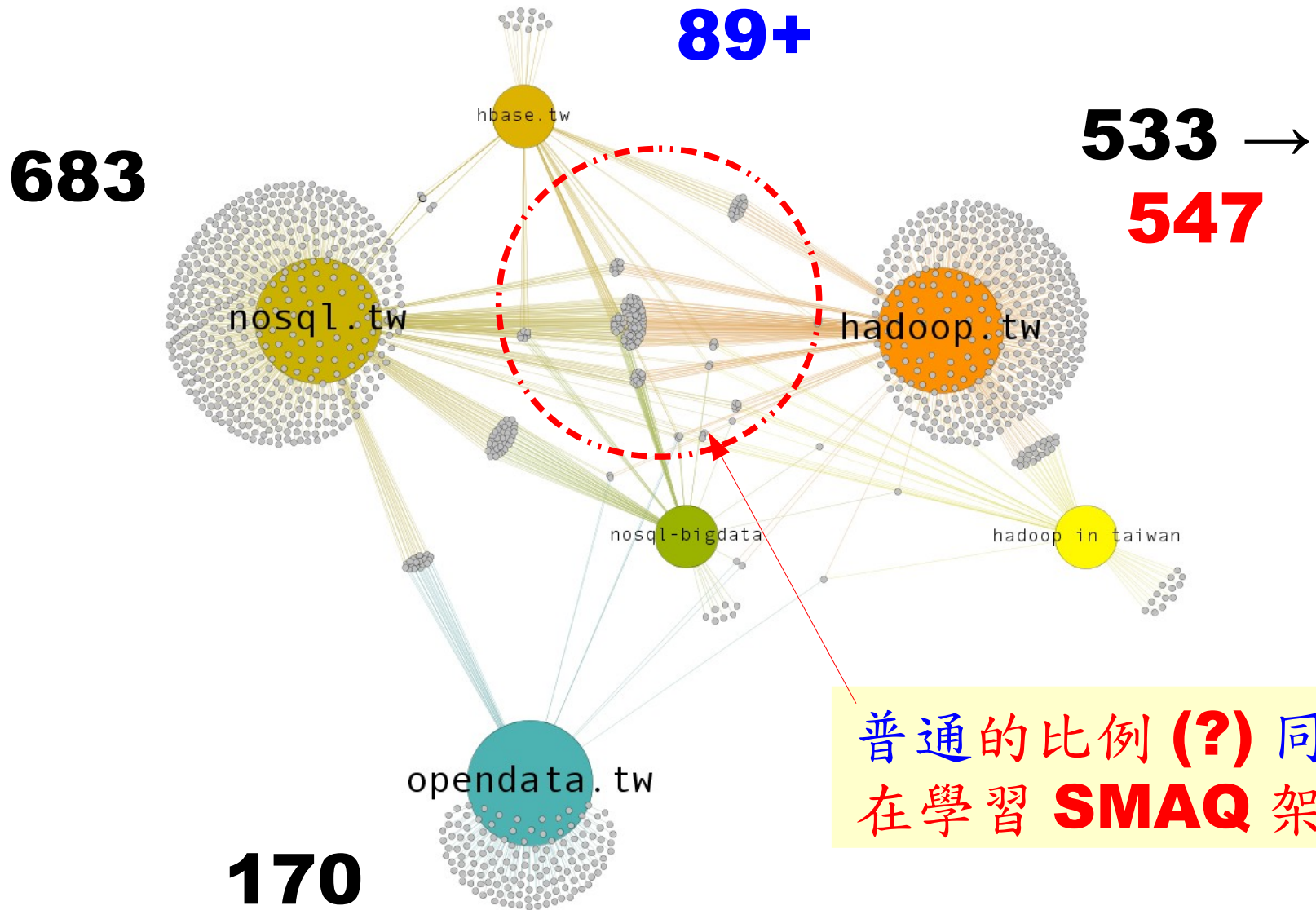
I am here
Connecting All 6 Groups

683

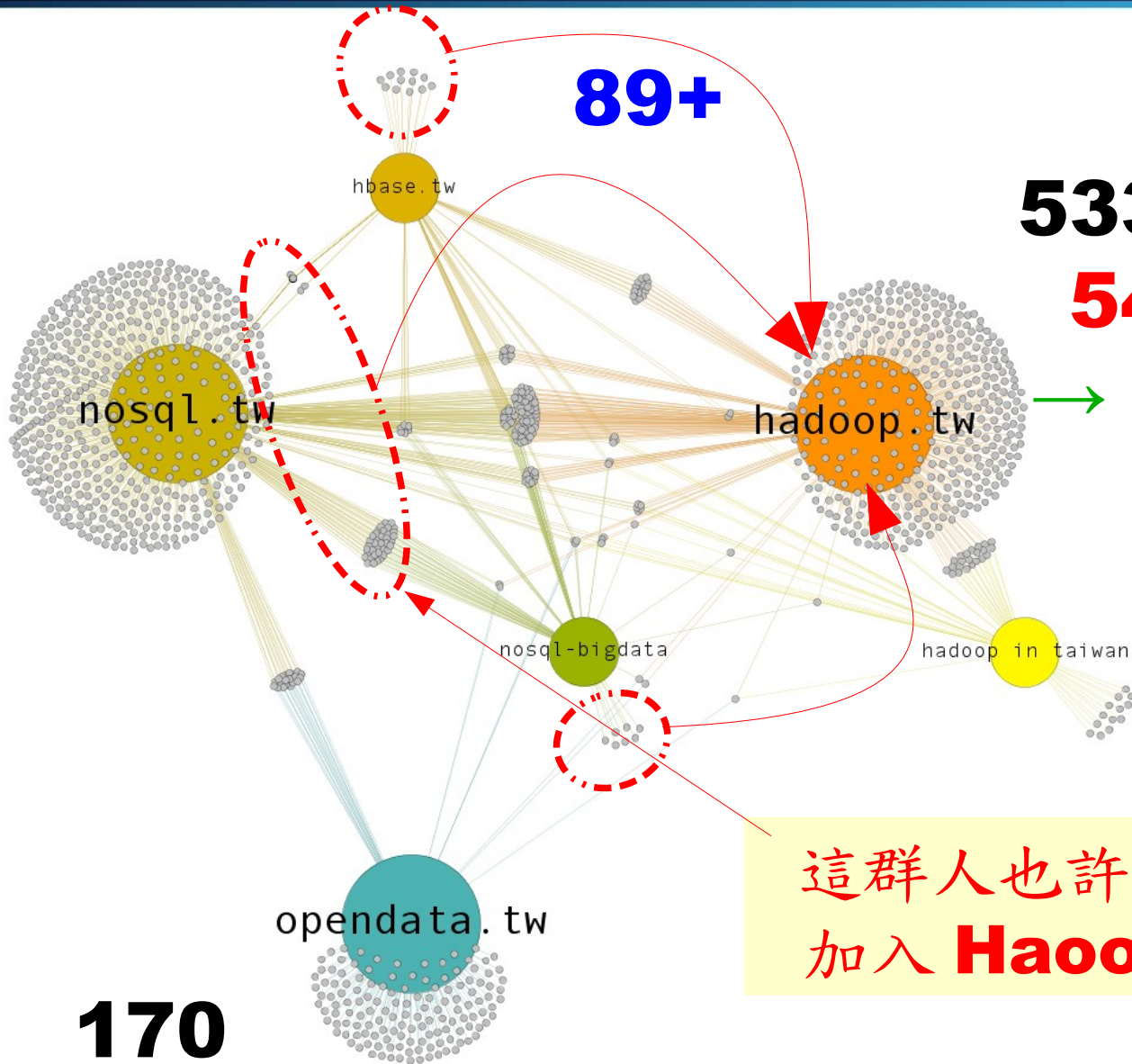




台灣對 **HBase**
的認識或需求
有賴提升或確認！



683



89+

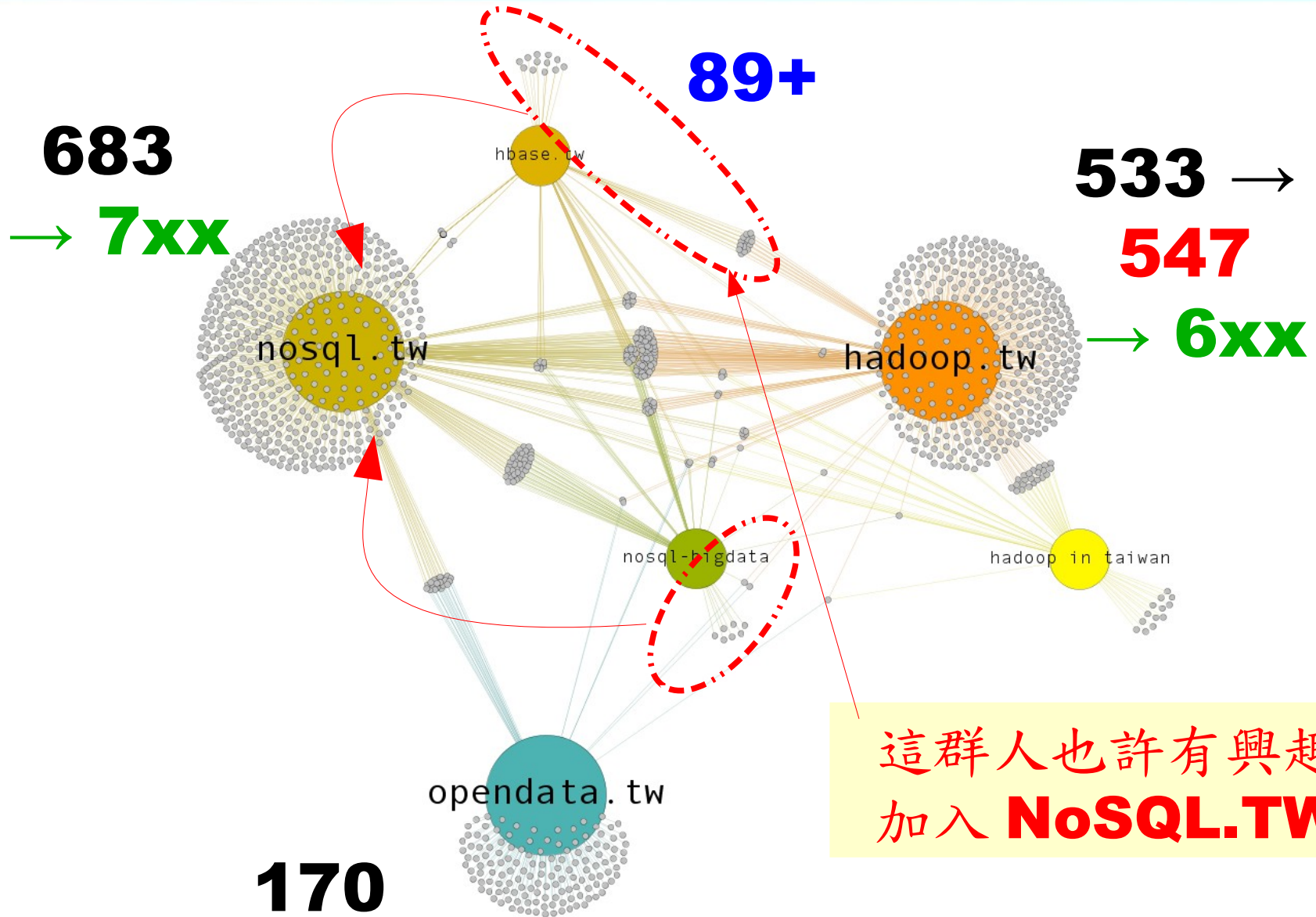
533 →

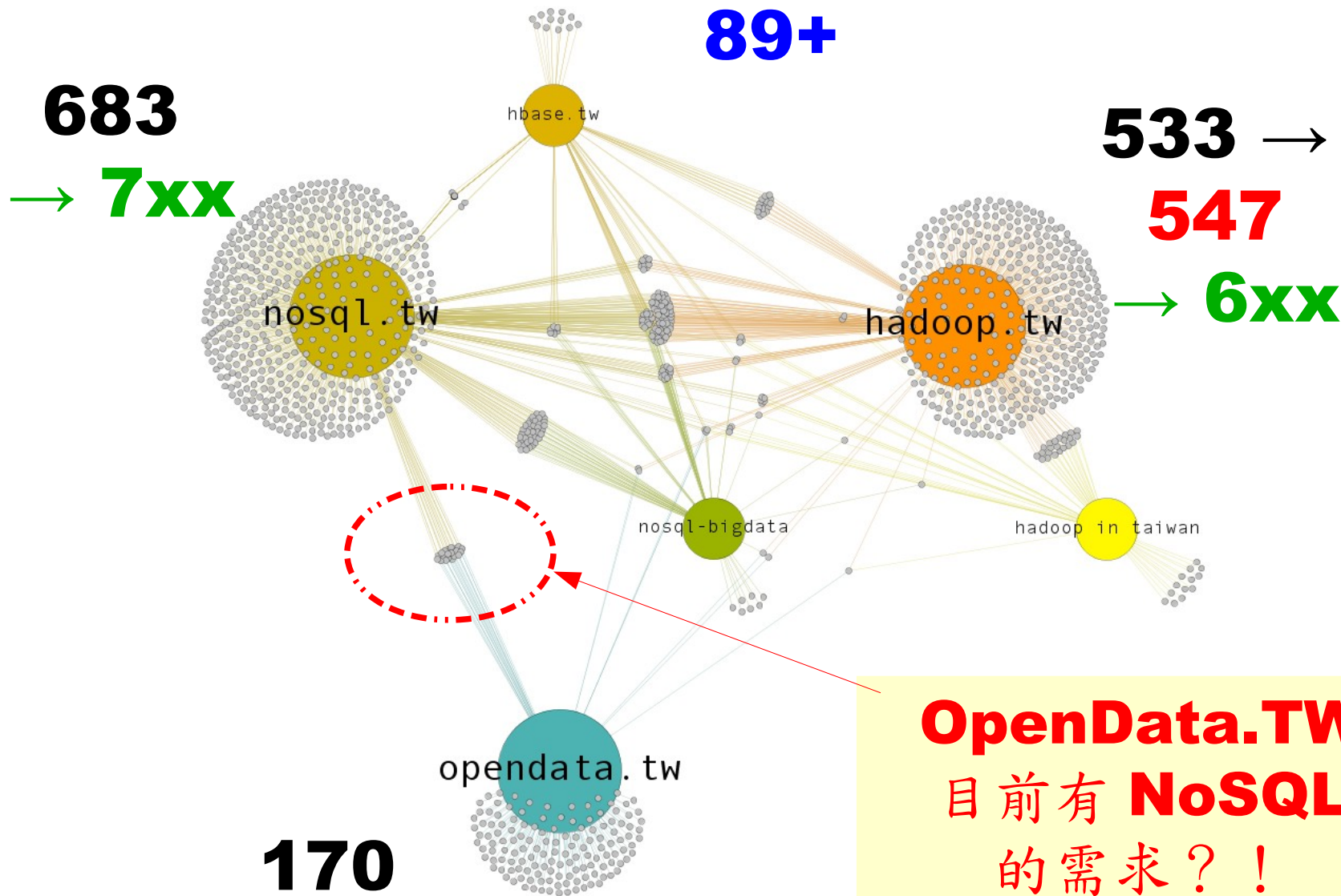
547

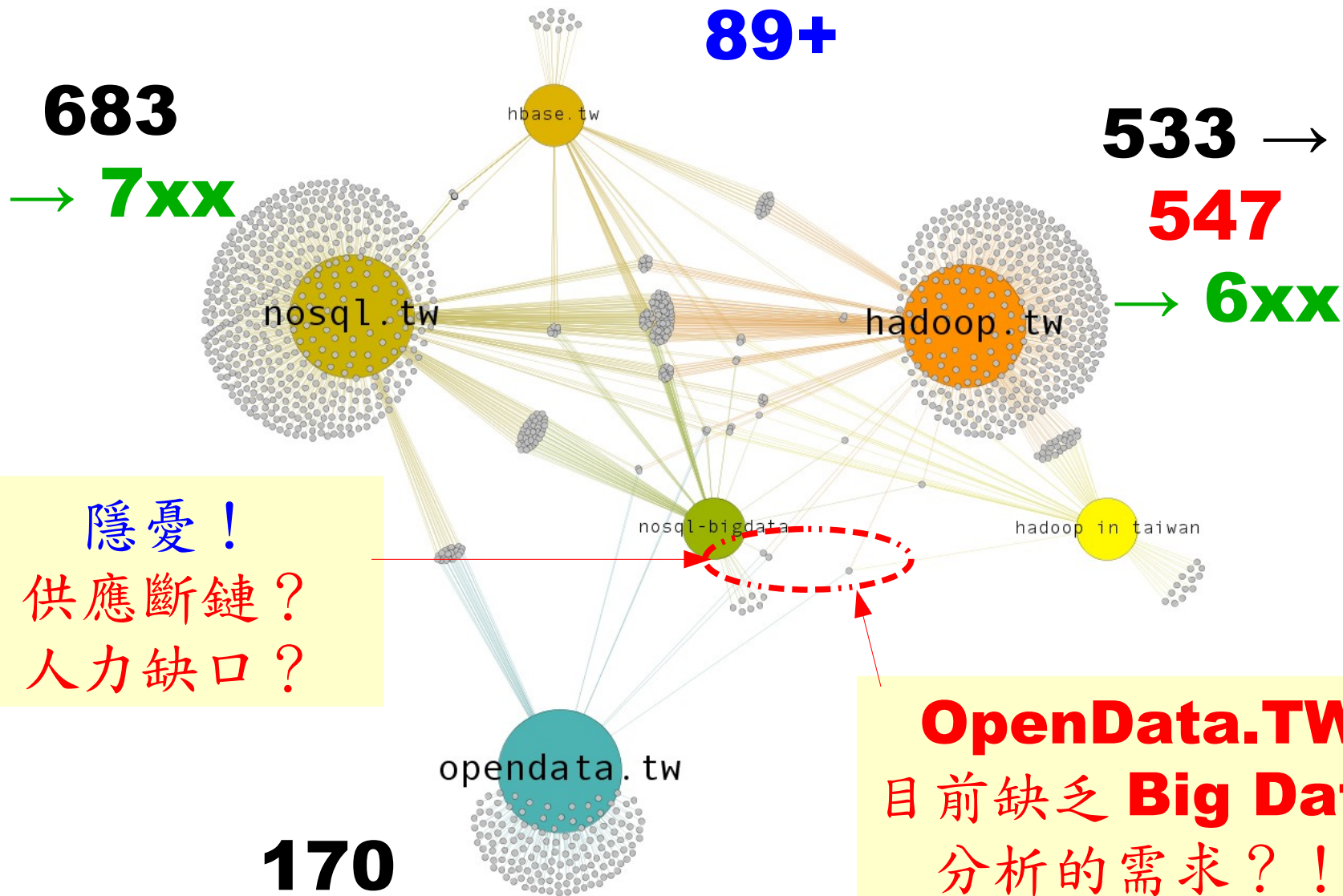
→ 6xx

170

這群人也許有興趣
加入 **Hadoop.TW**







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

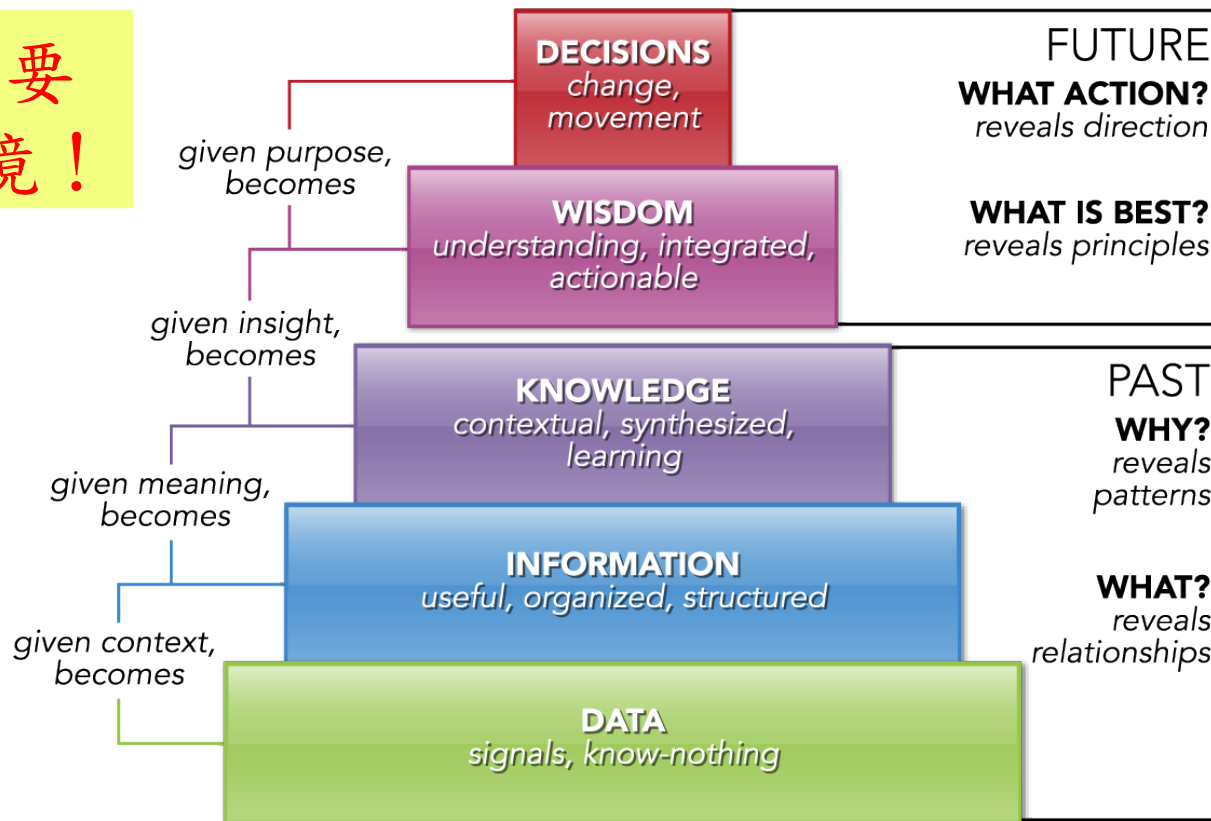
仍有待打通產業供應鍊
巨量資料產業尚未成型

Hadoop 社群如何找到需求？價值何在？

Knowledge is from the Past, Wisdom aim for Future

知識源自於過去，智慧在能預測未來

我們需要
更多情境！



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

個人化的巨量資料

Big Data Can Get Very Personal

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

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

知識有價，熱情無價！
社群也一樣需要永續經營

Questions?

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

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

