



# 計算思維與人工智慧應用導論 電腦運算結構 & 大數據應用

Dr. Chih-Hsun Wu

吳致勳 助理教授

1

本投影片僅供教學用途，  
所用圖檔都盡量附上原始來源，  
如有侵權煩請告知，將立即修正

國立政治大學人工智慧跨域研究中心

Date: 2023/10/24

# 課程進度

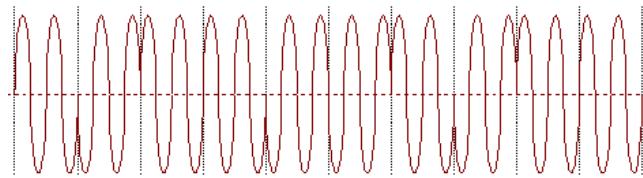
| 單元 | 時間         | 標題           |
|----|------------|--------------|
| 1  | 09/26 第三週  | 程式桌遊(海霸)     |
| 2  | 10/3 第四週   | micro:bit 1  |
| 3  | 10/17 第六週  | micro:bit 2  |
| 4  | 10/31 第八週  | RapidMiner 1 |
| 5  | 11/14 第十週  | RapidMiner 2 |
| 6  | 11/28 第十二週 | RapidMiner 3 |
| 7  | 12/12 第十四週 | 期末報告討論       |

| 週次<br>課程主題         | 課程內容與指定閱讀   |              |
|--------------------|---|--------------|
| 1 計算思維簡介           | 社會情境脈絡與未來發展<br>書目：1, 2, 3                               |              |
| 2 計算思維             | 基本內涵與核心概念<br>書目：1, 2, 3                                 |              |
| 3 功能模組             | 問題拆解與型態辨認<br>書目：4, 5, 6                                 |              |
| 4 功能模組             | 抽象思考與演算邏輯<br>書目：4, 5, 6                                 |              |
| 5 國慶日              | 國定假日  |              |
| 6 類比至數位轉換 & 電腦運算架構 | 類比與數位訊號的基礎概念及類比轉換至數位訊號的原理 &<br>書目：7 chapter 1 & 4 & 5   | 電腦組成元件與其運算架構 |
| 7 大數據應用            | 大數據中資料科學的基礎分析概念與商業相關應用                                  |              |
| 8 學習成果測試           | 期中評量/作業活動   |              |
| 9 運算思維測驗           | 國際運算思維挑戰賽   |              |
| 10 人工智慧發展          | 人工智慧發展歷程與未來趨勢<br>人工智慧各式技術與應用案例<br>書目：8                  |              |
| 11 人工智慧技術與應用       |   |              |
| 12 人工智慧應用場景        | 人工智慧跨域應用  |              |
| 13 人工智慧學習模型實作      | Nocode AI 練習 – Rapidminer<br>書目：9                       |              |
| 14 人工智慧倫理          | 生成AI (如：ChatGPT、Deepfake、Midjourney)、假新聞及未來人工智慧應用上的倫理問題 |              |
| 15 人工智慧專題          | 海報展示  |              |
| 16 計算思維與人工智慧       | 期末報告  |              |
| 17 彈性補充教學          | 人工智慧相關競賽經驗交流  |              |
| 18 彈性補充教學          | 校園人工智慧應用發想  |              |

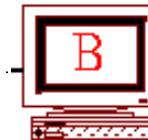
# 訊號傳輸技術

AM & FM

# 數位與類比



0010101010111100



## ●類比訊號 Analog Signal

- 為連續值，大自然的所有訊號都屬於類比訊號。可透過多種傳輸媒介進行傳送
- 例如：聲音、光、溫度、位移、壓力。
- 舉例：耳機、麥克風、水銀溫度計、傳統電話、收音機、無線電視、無線對講機等等。
- 缺點：容易被雜訊影響，會使類比訊號失真。

## ●數位訊號 Digital Signal

- 不連續的電壓脈波，主要是以不同的電位狀態來表示
- 不連續的值，以一連串的二進位表示，0 或 1。0 為低電位，1 為高電位。
- 例如：電腦處理資料、網路傳輸。
- 優點：容易處理資料
  1. 儲存、傳輸
  2. 壓縮、解壓縮
  3. 加密、解密
  4. 偵錯、除錯

# 訊號傳輸技術

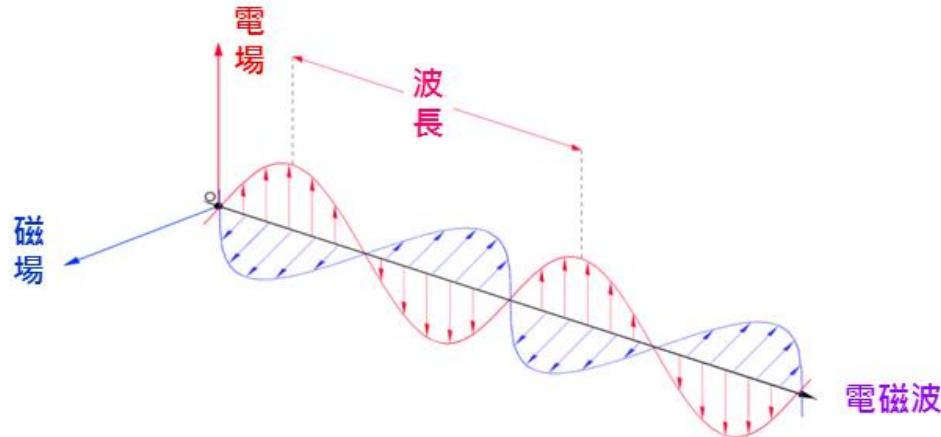
- 訊號本身
  - 類比訊號: a continuous representation of data, analogous to the actual information it represents
  - 數位訊號: a discrete representation of data, breaking the information up into separate elements

# 無線通訊運作原理1

- 使用高頻率交流電產生電磁場，進而藉由電磁場產生無線電波（即電磁波），利用無線電波不需介質就能長距離傳輸的特性，讓發射器可「憑空」將無線電波傳送到接收器，並讓接收器產生對應共振，如此一來便能解讀傳送過來的資訊
- 例如日常生活的收音機、手機、Wi-Fi 無線網路，都是這種技術的實際應用。

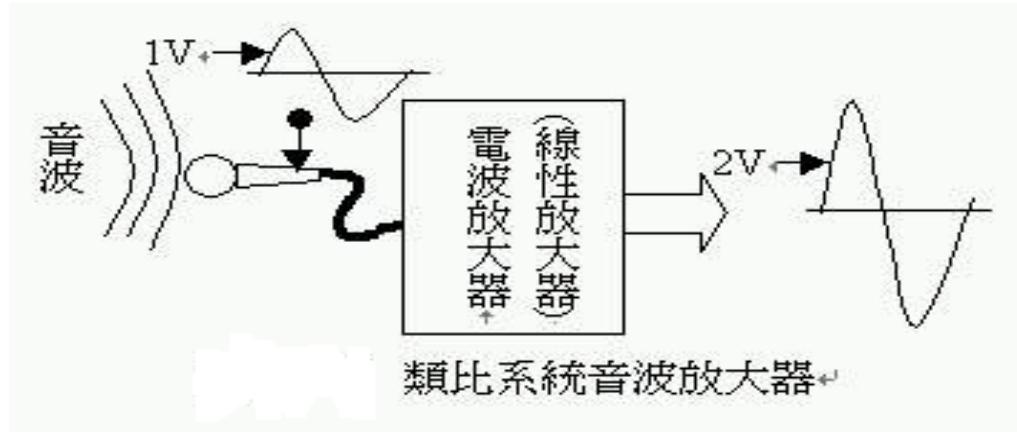
# 電磁波

- 是由互相垂直的電場與磁場交互作用而產生的一種能量，在前進時就像水波一樣會依照一定的頻率不停振動

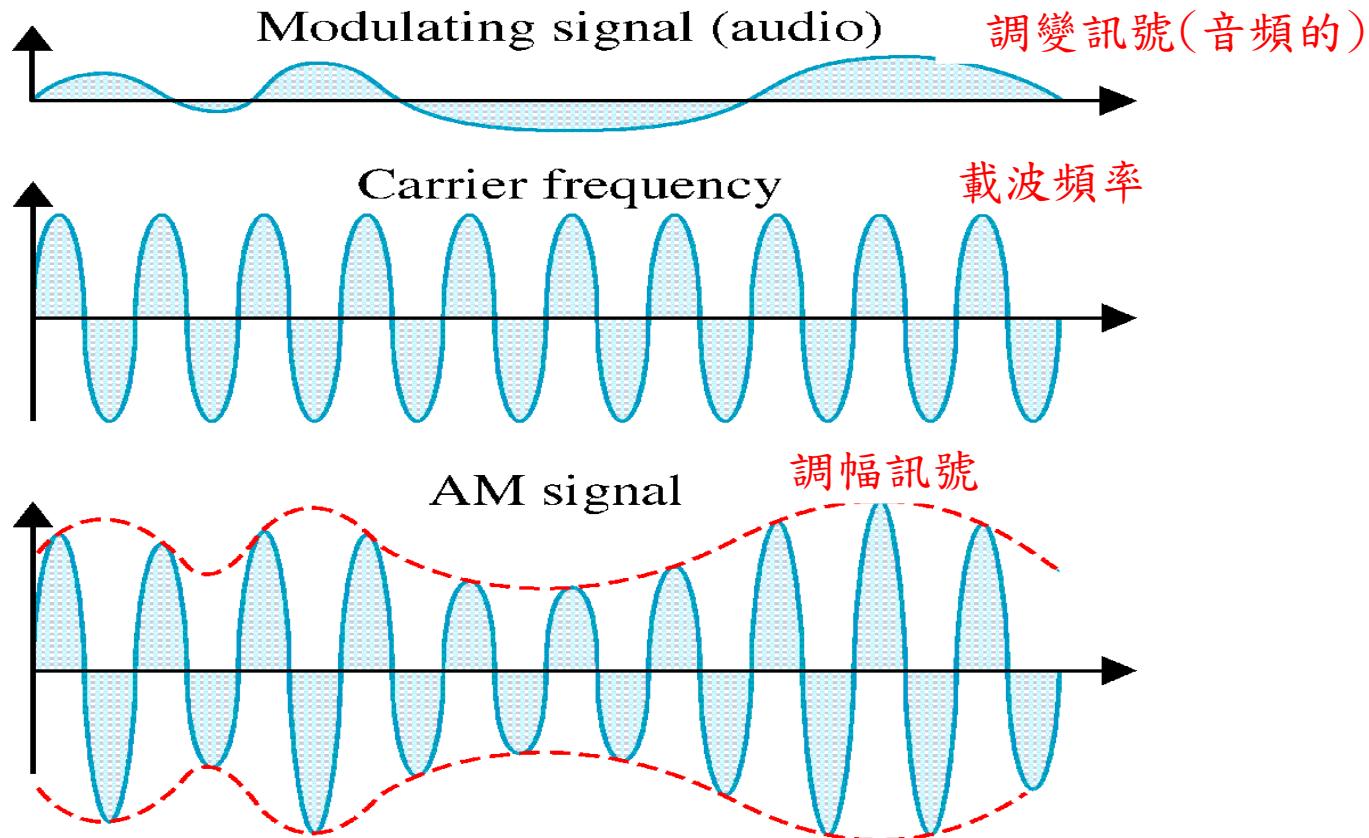


# 數位訊號類比傳輸

- 調幅 (Amplitude Modulation, AM)
- 調頻 (Frequency Modulation, FM)



# 調幅 (Amplitude Modulation, AM)

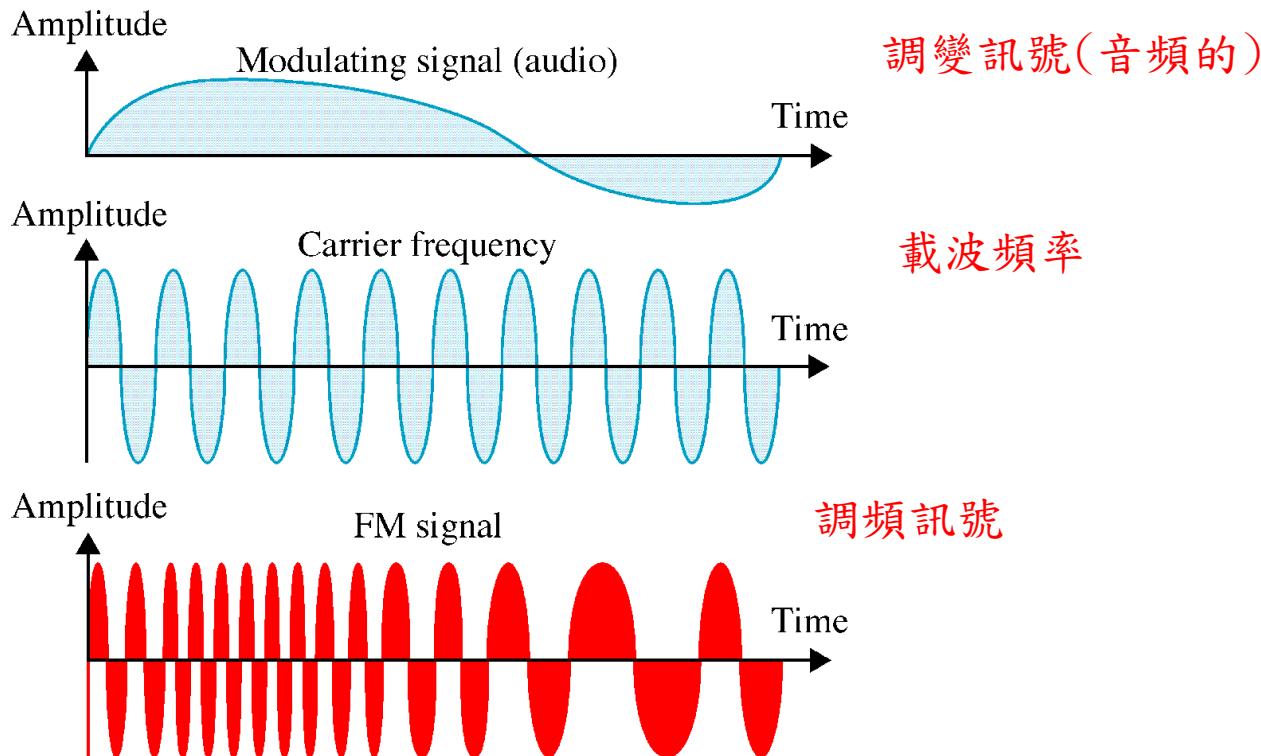


# AM 缺點

調幅傳送訊息的問題是容易受干擾和雜訊影響，因為干擾和雜訊會造成振幅隨機的改變。

- 使用振幅調變傳輸的方法（如AM廣播），品質都比較差。

# 調頻 (Frequency Modulation, FM)



# FM 優點

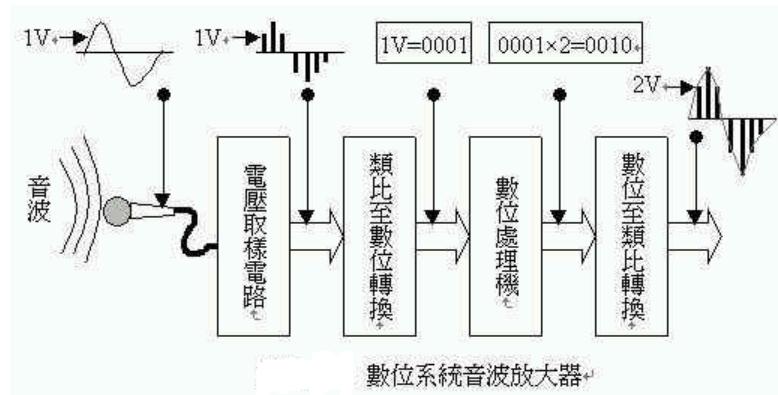
- 與AM調幅方式相較，FM調頻傳輸方式比較不受雜訊和干擾影響，因為無論振幅如何改變，接收器只對頻率有興趣，而不是振幅。
- FM 訊號可以使用較大的頻寬，將更多的訊息放入訊號，如此就可以傳送較高品質的訊息，例如可以用來傳送立體聲的音樂，而AM因為干擾的問題，只能播送單音。

# 數位訊號數位傳輸

- 直接以低電壓(0V)代表0，高電壓(1.5V)代表1
- 目前較常應用在電子產品內各電路板之間的訊號傳輸、乙太網路(ethernet)等短距離訊號傳輸

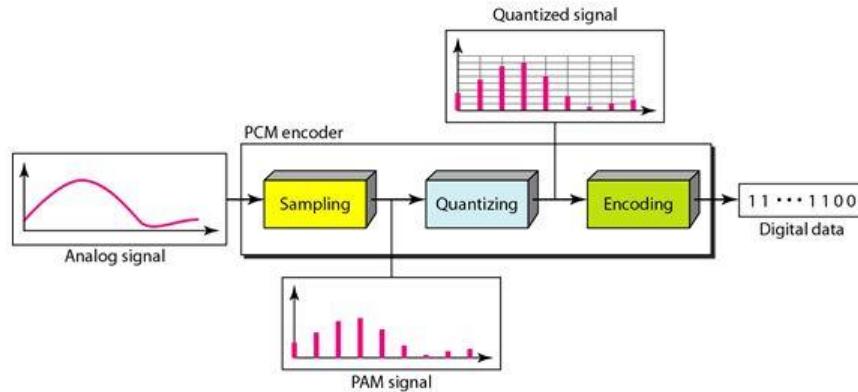
# 類比訊號數位傳輸

- 先將類比訊號轉換為數位訊號再傳輸，基本上就是「訊號數位化」的方法



# 類比到數位轉換 (Analog-to-digital conversion)

- Sampling (取樣)
- Quantization (量化)
- Encoding (編碼)



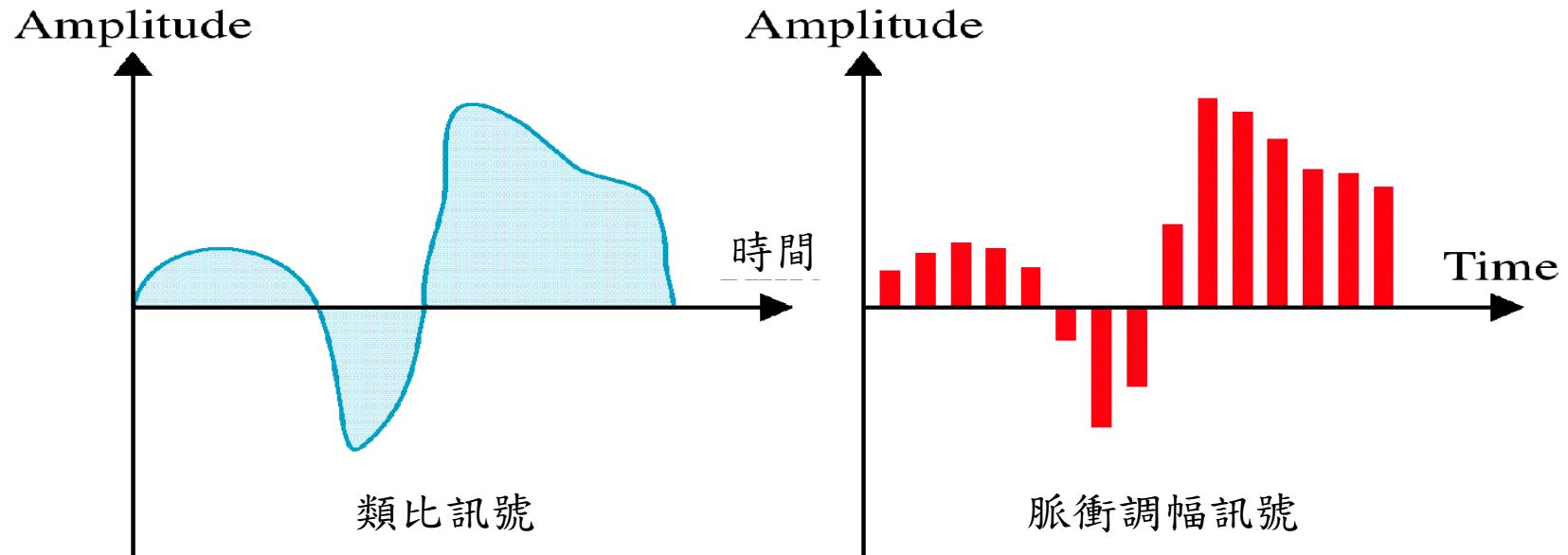
# Sampling and Quantization

- 脈衝振幅調變 (Pulse Amplitude Modulation, PAM) , 以 8KHz 取樣
- 取樣的時間間隔稱為「取樣週期」 ( Sampling Cycle ) , 單位為「秒」 , 將取樣週期取倒數可得每秒的取樣次數 , 稱為「取樣頻率」 ( Sampling Frequency ) , 單位為 Hz ( 次/秒 )

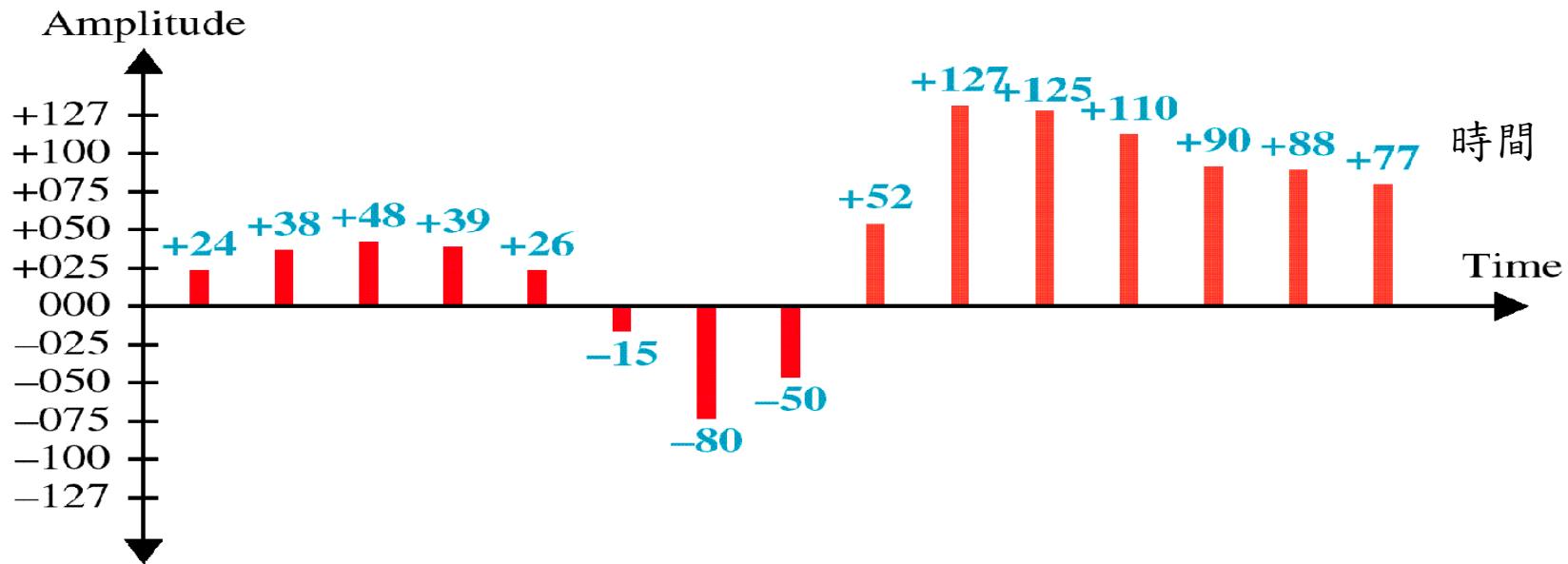
# Encoding

- 脈衝編碼調變 (Pulse Code Modulation, PCM) ,  
以 8/16 位元表示電位大小

# 脈衝振幅調變 (Pulse Amplitude Modulation, PAM)



# 量化PAM訊號

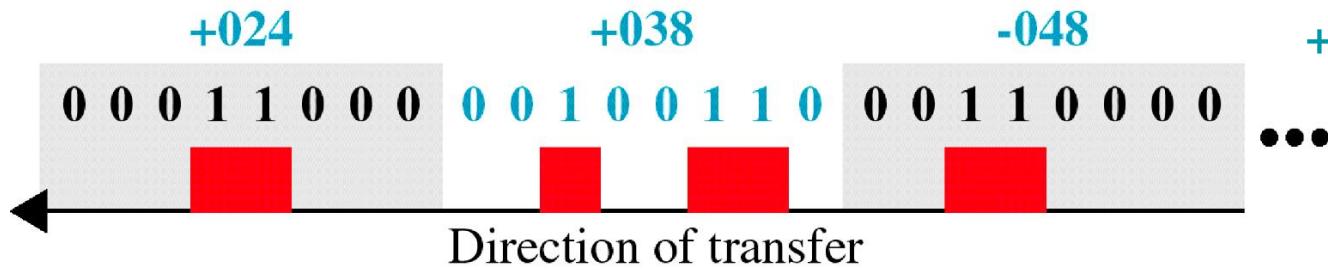


# 使用正負符號及數值大小以量化

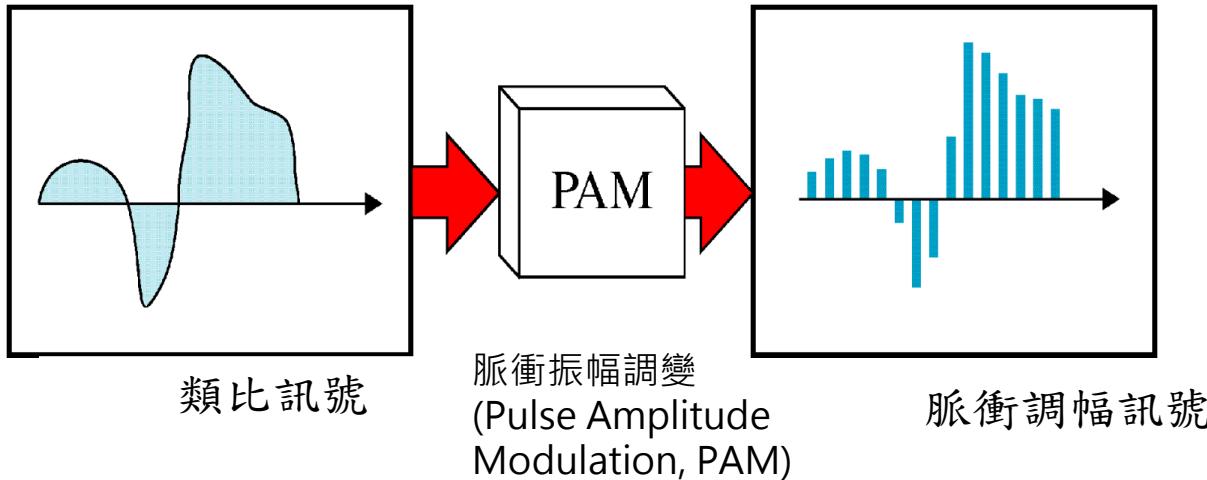
|      |          |      |          |      |          |
|------|----------|------|----------|------|----------|
| +024 | 00011000 | -015 | 10001111 | +125 | 01111101 |
| +038 | 00100110 | -080 | 11010000 | +110 | 01101110 |
| +048 | 00110000 | -050 | 10110010 | +090 | 01011010 |
| +039 | 00100111 | +052 | 00110110 | +088 | 01011000 |
| +026 | 00011010 | +127 | 01111111 | +077 | 01001101 |

正負號位元  
+是0，-是1

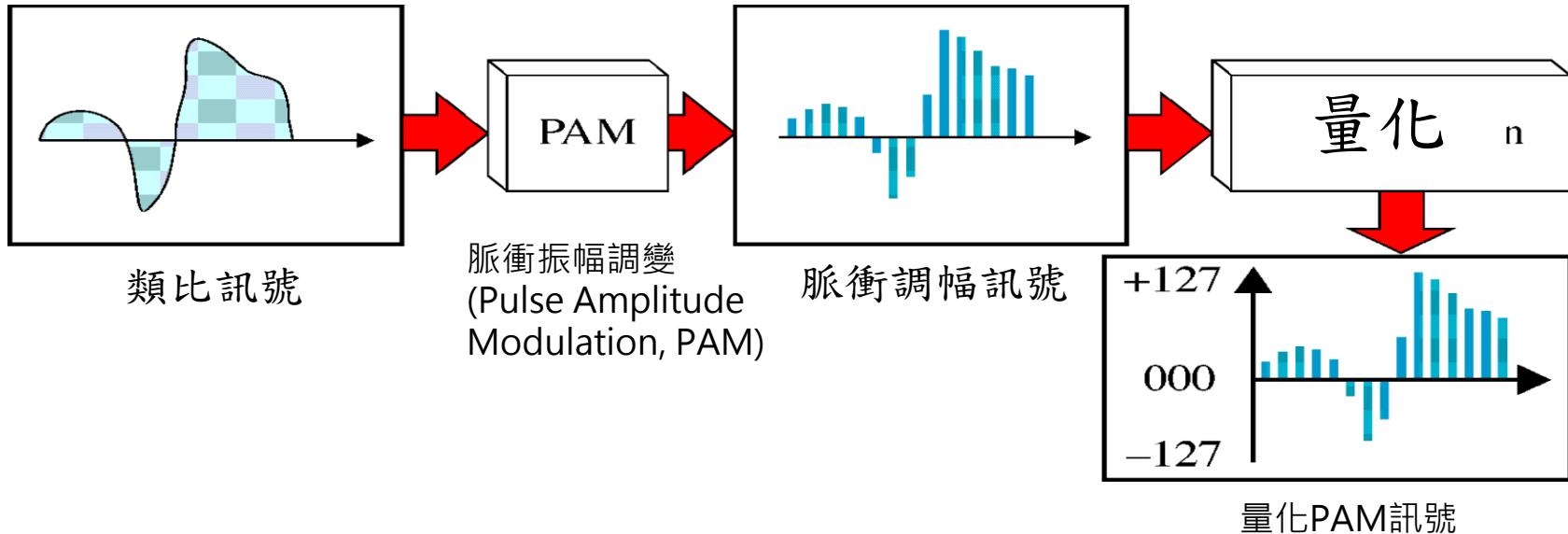
# PCM編碼



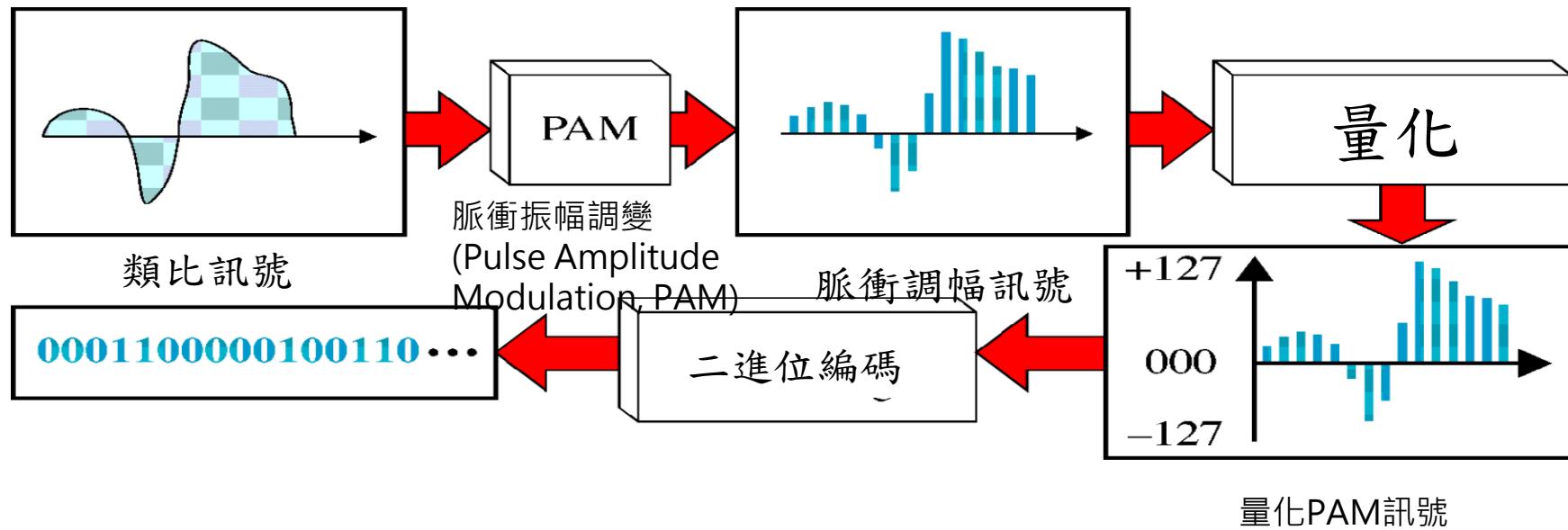
# 從類比訊號到PCM數位碼



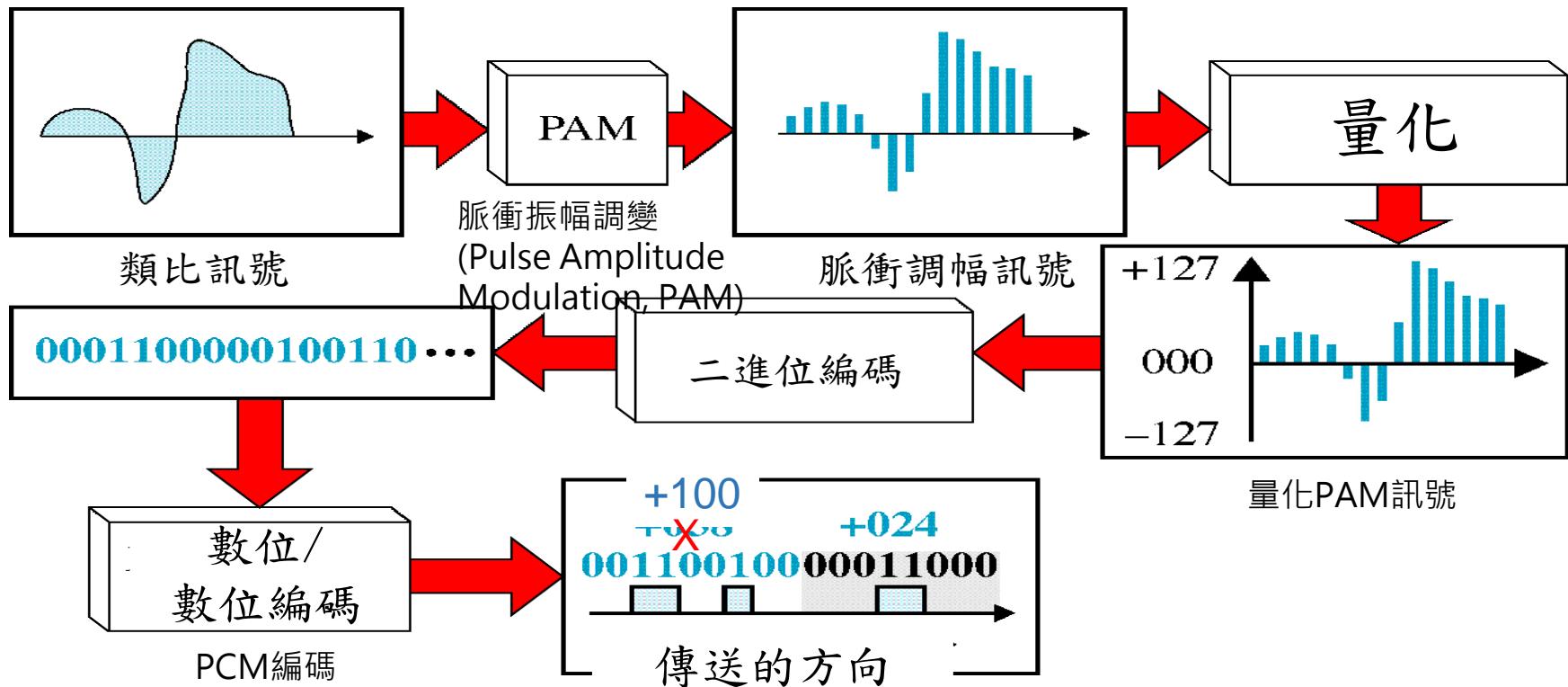
# 從類比訊號到PCM數位碼



# 從類比訊號到PCM數位碼



# 從類比訊號到PCM數位碼

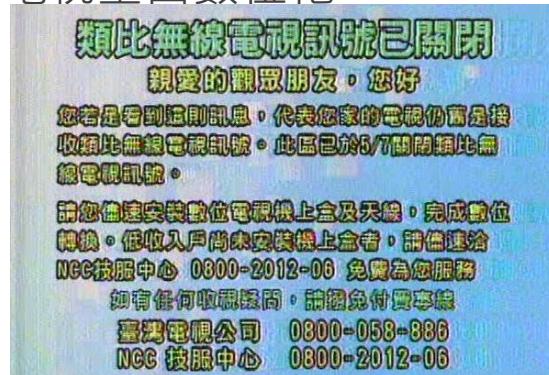
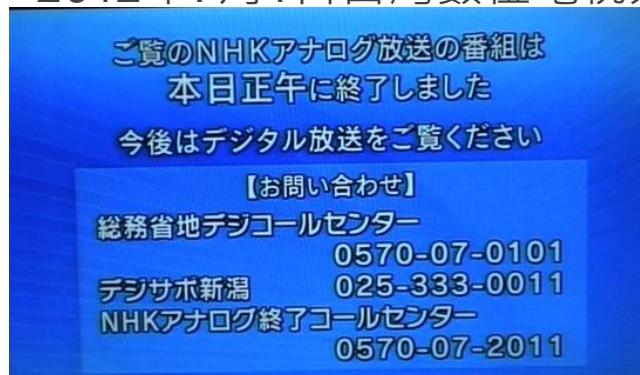


# Digital TV(數位電視)

- 視訊壓縮技術之進步，因此可將類比的畫面信號經數位化處理後，變成一串「二進位」型式數據資料
- 透過數位訊號處理，可以消除雜訊和干擾，畫面將會更清楚、細緻
- 再經數位調變傳送到家，因此可獲得比原來的類比電視更好的影像及聲音品質

# Digital TV (數位電視)

- 電視及節目訊號數位化後，數位訊號經「壓縮」，使得一個數位頻道可播送三至四個節目
- 2011年7月24日中午整點，日本電視正式全面數位化
- 2012年7月1日台灣數位電視元年：完成無線電視全面數位化



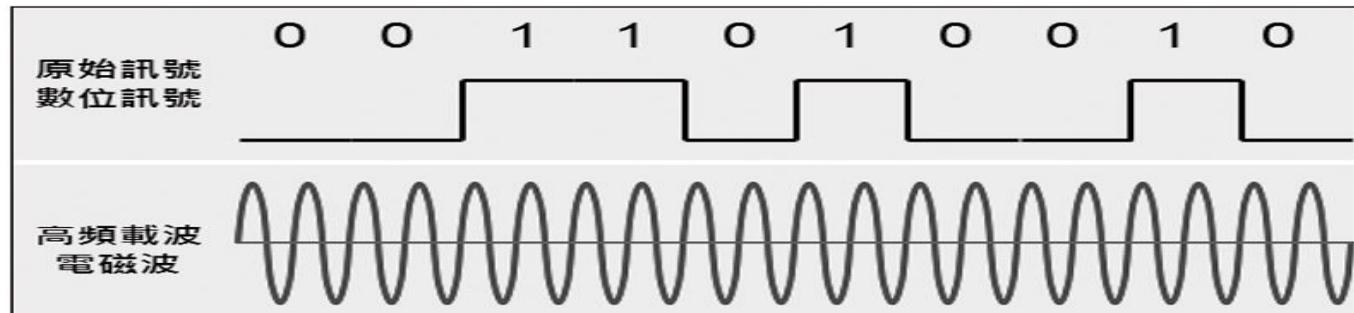
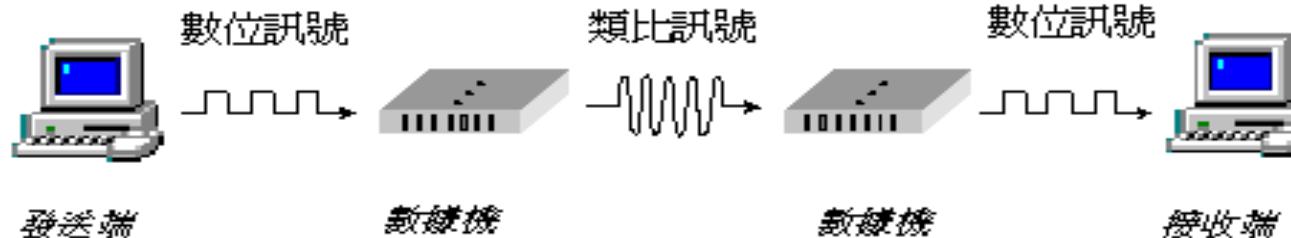
# 通訊技術

# 數位訊號類比傳輸

- 2G(GSM、CDMA)、3G(WCDMA、CDMA2000)、無線式行動電話(PHS)、無線區域網路(IEEE802.11)
  - Modulation (調變): 傳送端將數位訊號 (0 與 1) 轉變成不同的電磁波波

# 數據機 Modem

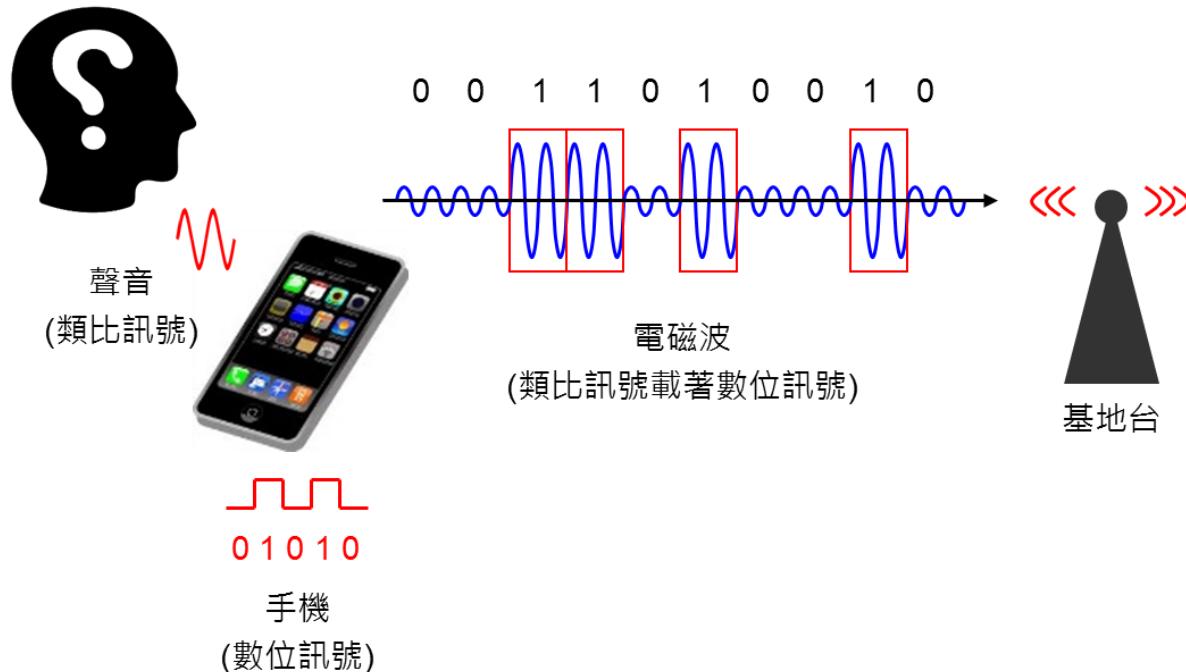
• D



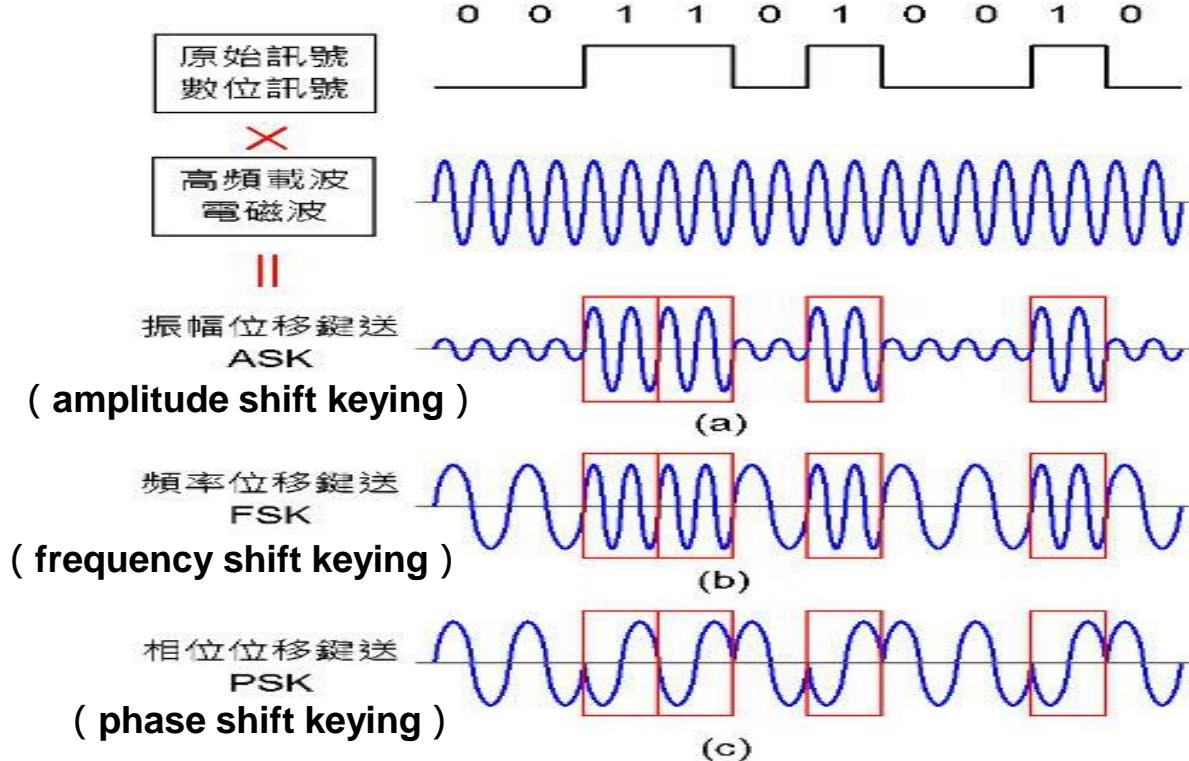
[http://scimonth.blogspot.tw/2014/09/blog-post\\_3.html](http://scimonth.blogspot.tw/2014/09/blog-post_3.html)

Credit by <http://technews.tw/2015/10/12/3g%E3%80%814g%E3%80%815g-meaning-part-two/>

# 數位通訊



# 數位通訊



振幅位移鍵送 ( ASK )：利用電磁波的「振幅大小」載著數位訊號 ( 0 與 1 ) 傳送出去，振幅小代表 0，振幅大代表 1，圖二 ( a ) 。

頻率位移鍵送 ( FSK )：利用電磁波的「頻率高低」載著數位訊號 ( 0 與 1 ) 傳送出去，頻率低代表 0，頻率高代表 1，圖二 ( b ) 所示。

相位位移鍵送 ( PSK )：利用電磁波的「相位不同 ( 波形不同 ) 」載著數位訊號 ( 0 與 1 ) 傳送出去，相位  $0^\circ$  代表 0，相位  $180^\circ$  代表 1

# 多工技術 ( Multiplex )

多人共同使用一條資訊通道的方法稱為「多工技術」 ( Multiplex )

- 分時多工接取 ( TDMA )

- 使用者依照「時間先後」輪流使用一條資訊通道
- 與B先講一句，再換C與D講一句
- 大家輪流（分時）講話彼此就不會互相干擾。



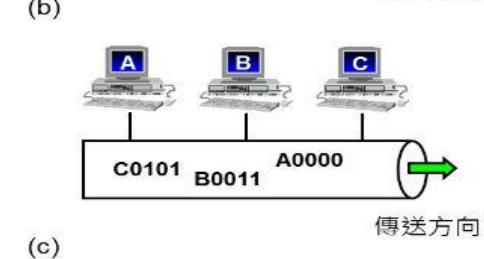
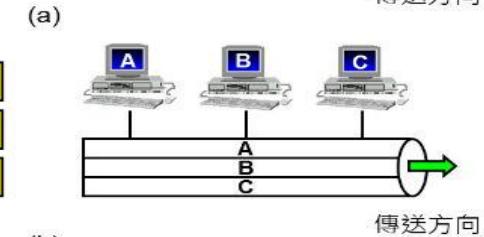
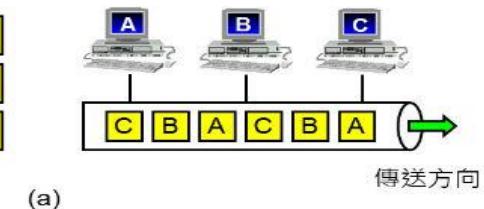
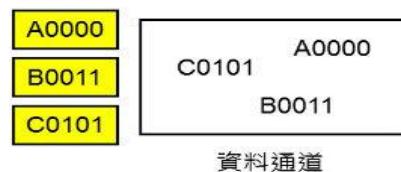
- 分頻多工接取 ( FDMA )

- 使用者依照「頻率不同」同時使用一條資訊通道
- A與B在客廳講話，C與D在書房講話
- 大家在不同的房間（分頻）講話彼此就不會互相干擾。



- 分碼多工接取 ( CDMA )

- 將不同使用者的資料分別與特定的「密碼 ( Code ) 」運算以後，再傳送到資料通道，接收端以不同的密碼來分辨要接收的訊號
- A與B用中文講話，丙與丁用英文講話
- 這樣雖然大家在同一個房子裡講話，各自仍然可以分辨出各自不同的語言

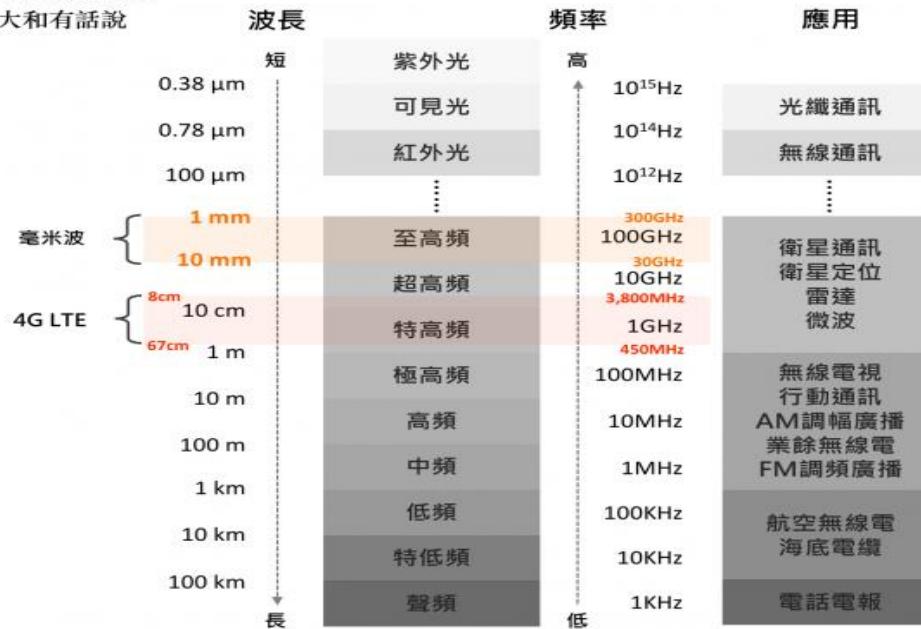


# 通訊世代比較

# 無線電波的頻譜

## ■ 通訊電磁波頻譜

Created by 大和有話說



- Credit by 大和有話說 - [5G科普, 只要9張圖, 看懂什麼是5G](#)

# 行動通訊技術演進史

- 1G 僅擁有 2.4kbps 傳輸速度，已跨出一大步，突破過去「定點」打電話限制
- 2G 速率介於 56k-114kbps 之間，實現語音通訊數字化，透過手機上的小螢幕傳遞文字內容，開啟簡訊時代
- 3G 標準自 2000 年正式公告，擁有 384kbps 速率，讓使用者瀏覽網站與串流音樂。行動網路的語音傳輸會把語音和多媒體訊息協定分開，這種網路的使用方法非常沒有效率
- 4G 利用封包交換系統 ( packet - switching system ) 的方式收發，讓語音、簡訊和多媒體傳訊完全以 IP 為基礎。4G 的穩定速度，被認為是應用程式 ( App ) 經濟推手，更帶動社群網站發展

| 行動通訊技術     | 功能                              | 峰值速率       | 頻率           |
|------------|---------------------------------|------------|--------------|
| 1G (1980s) | 通話                              | 2 Kbps     | 800-900 MHz  |
| 2G (1990s) | 通話、簡訊、Mail (純文字)                | 10 Kbps    | 850-1900 MHz |
| 3G (2000s) | 通話、簡訊、網路、音樂串流                   | 3.8 Mbps   | 1.6-2.5 GHz  |
| 4G (2010s) | 通話、簡訊、網路、1080p 影片串流             | 0.1-1 Gbps | 2-8 GHz      |
| 5G (2020s) | 通話、簡訊、網路、4K 影片串流、VR 直播、自駕車、遠距手術 | 1-10 Gbps  | 3-300 GHz    |

# 通訊技術世代

## 數位通訊系統的頻譜效率比較表

| 世代    | 系統名稱  | 多工方式 | 調變方式  | 通道頻寬   | 資料傳輸率<br>上傳/下載<br>(bps) | 頻譜效率<br>上傳/下載<br>(bps/Hz) |           |
|-------|-------|------|-------|--------|-------------------------|---------------------------|-----------|
| 2G    | GSM   | FDMA | GMSK  | 200KHz | 9.6K/14.4K              | 0.05/0.07                 |           |
| 2.5G  | GPRS  |      | GMSK  | 200KHz | 9.6K/115K               | 0.05/0.58                 |           |
| 2.75G | EDGE  |      | 8PSK  | 200KHz | 384K/384K               | 1.92/1.92                 |           |
| 3G    | WCDMA | FDMA | QPSK  | 5MHz   | 64K/2M                  | 0.01/0.40                 |           |
| 3.5G  | HSDPA |      | CDMA  | 16QAM  | 5MHz                    | 384K/14.4M                | 0.08/2.88 |
| 3.75G | HSUPA |      |       | QPSK   | 5MHz                    | 5.76M/14.4M               | 1.15/2.88 |
| 4G    | LTE   | FDMA | 64QAM | 20MHz  | 50M/100M                | 2.5/5                     |           |
| 4G    | LTE-A |      | 64QAM | 100MHz | 500M/1G                 | 5/10                      |           |

# 為何需要 5G?

- 「行動流量持續上升」：國家通訊傳播委員會（NCC）統計，2018年初台灣 4G 用戶平均數據傳輸量為 14.74 GB；2019 年 5 月已成長為平均 16.85 GB，位居世界前段班。愛立信（Ericsson）也強調，全球行動寬頻流量至 2024 年會是 2019 年的 5 倍，特別是在市區密集區域，若僅依賴 4G 將無力負擔。
- 「低頻率負荷有限」：頻率低則裝載量小，速度有限；頻率越高，能傳輸越多數據、裝更多訊息。換言之，4G 採用低頻的先天限制，是 5G 不得不發展的主因，速度反而是其次。
- 4G 受到「行動流量持續上升」與「低頻率負荷有限」考驗，試著回想看看，當你在跨年、參加演唱會時，是否連打開社群網站都很困難，更不用說開直播。

# 5G v.s. 4G

- 美國無線產業協會 (CTIA) 表示下載一部兩小時左右的電影
  - 現在的 4G LTE 網路提供平均 10-20 Mb/s 的連線速度，4G 在理想狀態下得花 6 分鐘時間
  - 5G 僅需 3.6 秒就能完成任務。
- 5G 理論速度能比 4G 快上一百倍，但相較之下，電信業者看法仍略顯保守。如中國移動表示，5G 網路速度約是 4G 的 11.2 倍，「同樣時間內，用 4G 只能下載半集戲劇，5G 可以搞定 10 集。」

# 5G v.s. 4G

5G:高速speed

+低延遲latency(遠距醫療、無人自駕車車速來到100公里時緊急煞停  
→產業發展)

+廣連結connections(同時連接大量裝置→產業發展)

|                  | 4G                    | 5G                       |
|------------------|-----------------------|--------------------------|
| 峰值速率 (理想上傳、下載速度) | 0.1-1 Gbps            | 1-10 Gbps                |
| 延遲 (回應時間)        | 15-25 毫秒 (0.02-0.03秒) | 1 毫秒 (0.001秒)            |
| 頻率               | 10 GHz 以下             | 30到300 GHz               |
| 優點               | 低頻覆蓋廣，不用大量基地台         | 高頻提升傳輸速率                 |
| 缺點               | 頻寬小易壅塞                | 難穿透固體，訊號隨距離快速下降，需建置更多基地台 |

# 5G 高頻毫米波容易被干擾

- 就物理特性而言，頻率愈高，波長愈短，「穿透能力」也就愈強。
- 然而，高頻信號的指向性也較強，它們遇到障礙物會想直接穿過去，而不是繞過去（也就是「繞射能力」差），因此其穿透障礙物所帶來的能量消耗，也會使傳輸距離變短。

## ■ 無線通訊之高、低頻訊號比較

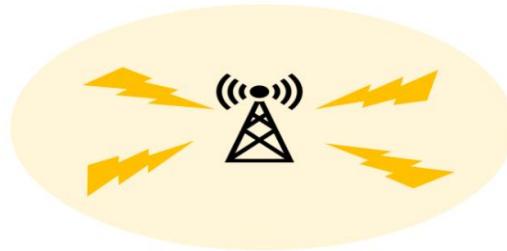
資料來源：MIC ( 2009 )  
Created by 大和有話說

|         | 高頻訊號 | 低頻訊號 |
|---------|------|------|
| 波長      | 短    | 長    |
| 方向性     | 定向   | 無方向性 |
| 繞射能力    | 弱    | 強    |
| 穿透能力    | 強    | 弱    |
| 電波衰減    | 大    | 小    |
| 覆蓋能力    | 弱    | 強    |
| 基地台承載量  | 多    | 少    |
| 功率傳送增益比 | 小    | 大    |
| 終端天線    | 短    | 長    |

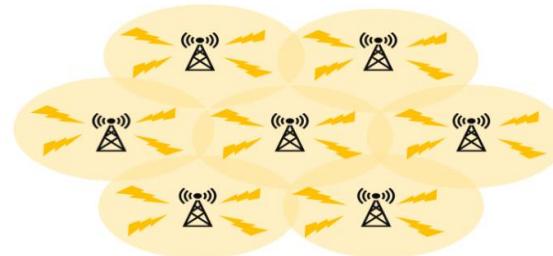
# 5G 微型基地台 ( small cell )

- 由於 5G 高頻毫米波容易被干擾，因此需布置數量更多的基地台，才能達到一定的覆蓋率。
- 由於 5G 技術特性，基地台數量將是 4G 的二到四倍，目前是 5G 發展初期，基地台的價格也是 4G 基地台的二到四倍
- 如日本政府正規劃將紅綠燈、路燈都裝上 5G 基地台，就是希望提高基地台密度。這也代表微型基地台 ( small cell ) 重要性更勝以往。

■ 相同範圍下，5G 的基站需求將遠大於 4G  
Created by 大和有話說



4G 概念圖



5G 概念圖



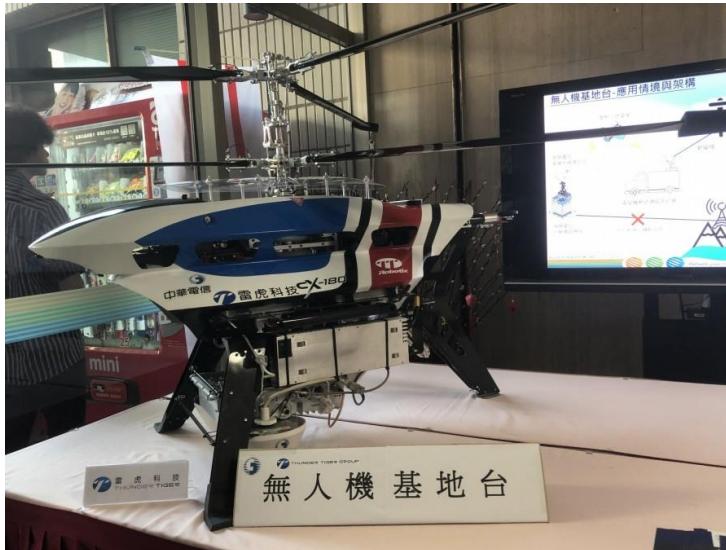
由於高頻段下的傳輸距離較短，使得覆蓋能力也相對較弱。因此，不同於頻段較低的 4G 採用一個大型基地台 (Macro Cell) 即可覆蓋 2~40 公里，5G 則必須採用許多個微型基地台 (Small Cell) 才能覆蓋到同樣的區域。

Credit by 大和有話說 - [5G 科普, 只要 9 張圖, 看懂什麼是 5G](#)

Credit by 2019-08-12 數位時代吳元熙 <https://www.bnnext.com.tw/article/54075/5g-4g-difference>

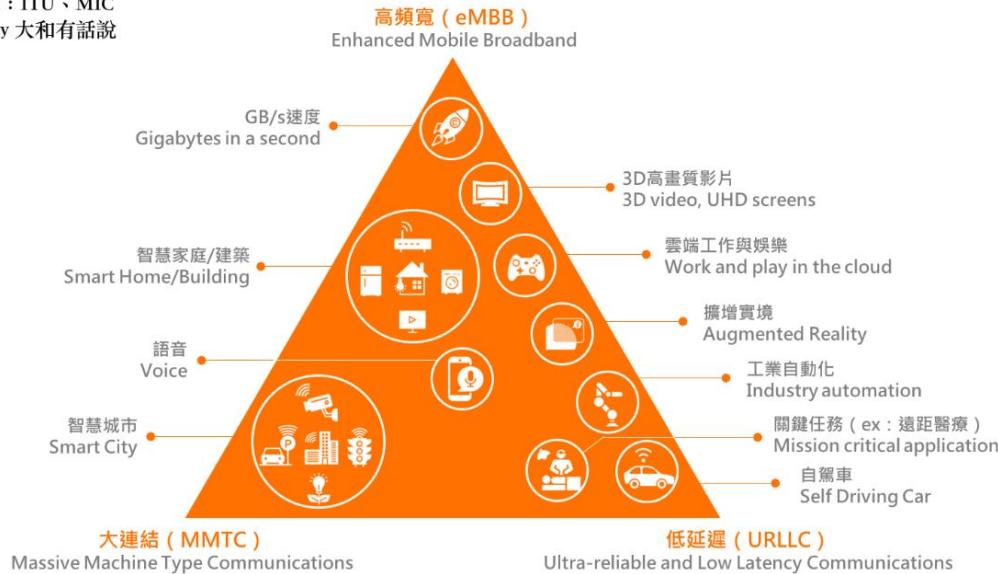
# 無人機搜救災難現場搭載 5G 減少因延遲導致的遠距操控失誤機率

- 中華電信和無人機業者「雷虎科技」合作，將5G基地台裝在機身上，能運用在高山救難上。
- 遠距醫療、VR 直播、4K 畫質直播，都因此有發展契機。



## ■ 5G 關鍵技術及應用場景

資料來源：ITU、MIC  
Created by 大和有話說



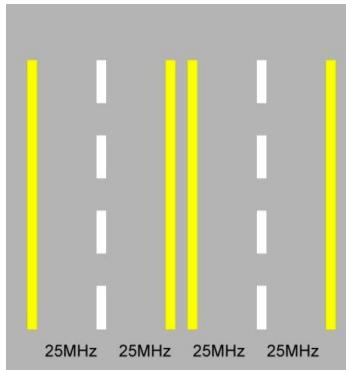
# 5G 頻段競價

# 先買車道，再選位置

- 5G 頻段競價作業時，要分配的資源共
  - 頻寬：電信業者可使用的頻段總量
  - 位置：這些頻段的頻率分布位置
- 如果把頻段比喻為馬路，無線電訊號就是在路上來往的車
  - 頻寬 = 車道寬度
    - 頻寬越大代表車道數量越多，就像 3 線道馬路運載量比 2 線道馬路多，所以頻寬越大能傳輸的資料與容納的用戶數量就越多，等於可為更多用戶提供更快的 5G 通訊服務。
  - 位置 = 車道位置
    - 頻段位置也會影響通訊品質，走在快車道會比較順暢，如果走在慢車道，可能時常被公車、路人干擾，較差的頻段也容易被其他無線電波干擾，傳輸速度受到影響
- 連續頻段有利於提供較佳通訊品質，以 5G 通訊而言，約需 80MHz 連續頻寬才能發揮高速連網特性，就像如果我們可以支配 2 個相鄰車道，會比支配隔開的 2 個車道更有效率。

# 先買車道，再選位置

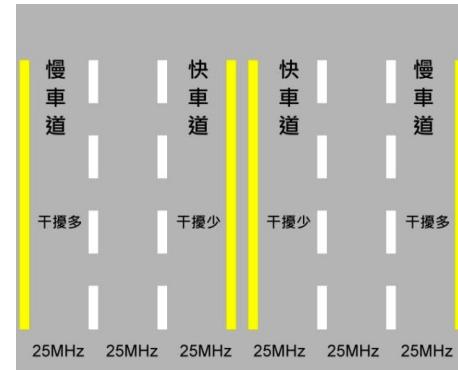
- 每 25MHz 的頻寬代表 1 個車道



100MHz 頻寬則代表雙向 4 線道  
馬路，流量較有限也容易塞車



150MHz 頻寬代表雙向 6 線道大馬路，  
交通更流暢



取得較好頻段位置與連續頻段，也有助  
提升 5G 通訊品質

[完全看懂 5G 頻段競價作業，高達 1,422 億元的標金是在搶什麼？](#)

# 競價結果 @ 2020/02/21



## 競價結果



| 得標者名單 | 得標金<br>(億元) | 3.5GHz  |         |         |        | 28GHz       |         |        |             |       |
|-------|-------------|---------|---------|---------|--------|-------------|---------|--------|-------------|-------|
|       |             | 數量競價    |         | 位置競價    |        | 總標金<br>(億元) | 數量競價    |        | 總標金<br>(億元) |       |
|       |             | 頻寬(MHz) | 標金(億元)  | 位置      | 標金(億元) |             | 頻寬(MHz) | 標金(億元) |             |       |
| 中華電信  | 483.73      | 90      | 456.75  | F13~F21 | 20.8   | 477.55      | 600     | 6.18   | G10~G15     | 6.18  |
| 台灣大哥大 | 306.56      | 60      | 304.5   | F22~F27 | 0      | 304.5       | 200     | 2.06   | G24~G25     | 2.06  |
| 台灣之星  | 197.08      | 40      | 197.08  | F1~F4   | 0      | 197.08      | 0       | 0      |             | 0     |
| 亞太電信  | 4.12        | 0       | 0       |         |        | 0           | 400     | 4.12   | G20~G23     | 4.12  |
| 遠傳電信  | 430.42      | 80      | 406     | F5~F12  | 20.3   | 426.3       | 400     | 4.12   | G16~G19     | 4.12  |
| 總計    | 1421.91     | 270     | 1364.33 |         | 41.1   | 1405.43     | 1600    | 16.48  |             | 16.48 |

註：1. 得標者名單依競價者名稱筆劃順序排列。

2. 得標金=數量競價得標總價+位置競價得標金。

### 3.5GHz頻段

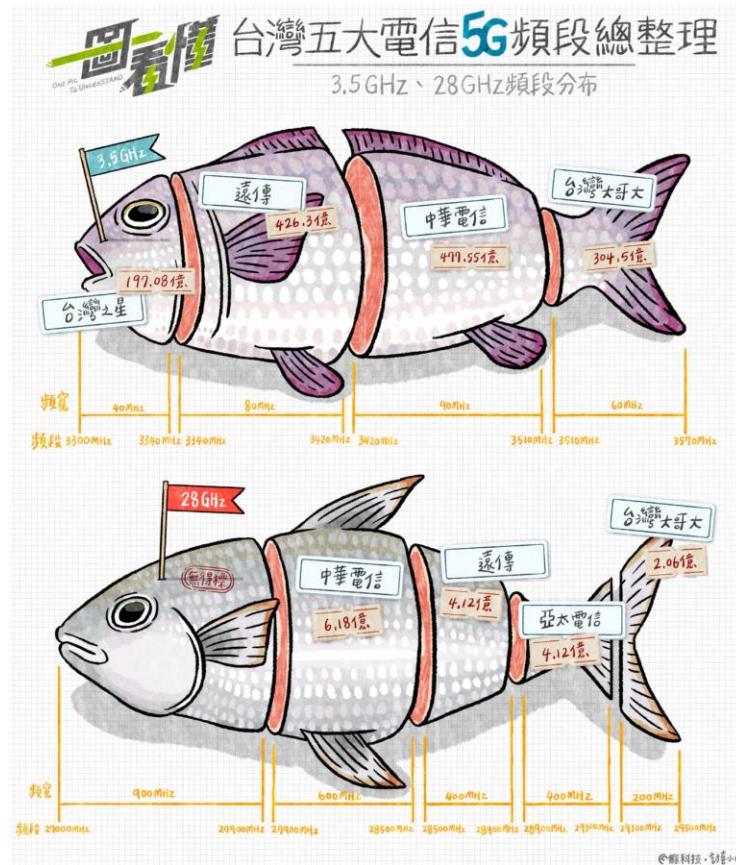
| 3300          | 3340          | 3420          | 3510           | 3570MHz |
|---------------|---------------|---------------|----------------|---------|
| 台灣之星<br>40MHz | 遠傳電信<br>80MHz | 中華電信<br>90MHz | 台灣大哥大<br>60MHz |         |

### 28GHz頻段

| 27000 | 27900 | 28500          | 28900          | 29300          | 29500MHz      |
|-------|-------|----------------|----------------|----------------|---------------|
|       | 無得標者  | 中華電信<br>600MHz | 遠傳電信<br>400MHz | 亞太電信<br>400MHz | 台灣大<br>200MHz |

# 魚頭、魚肚、魚尾

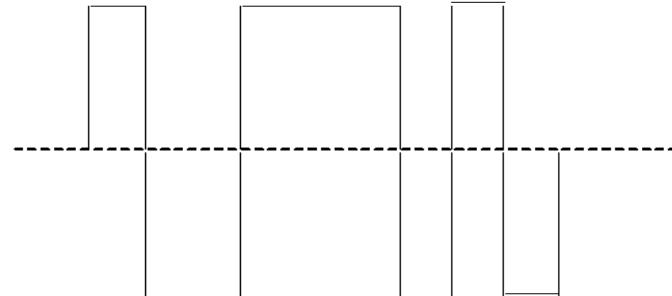
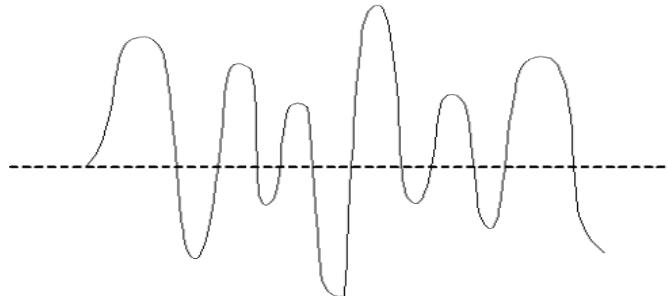
- Credit by 癱科技 討喜小姐, Zero圈  
<https://www.cool3c.com/article/151962>
- 延伸閱讀: 林之晨帶你科普 5G : 為什麼 3.5 GHz 頻段會有魚頭、魚肚、魚尾之分?
  - <https://buzzorange.com/techorange/2020/03/04/5g-3-5ghz-issue/>



# 總結

# 類比與數位訊號比較

- Analog: 連續變化的電磁波，可透過多種傳輸媒介進行傳送
- Digital: 不連續的電壓脈波，主要是以不同的電位狀態來表示



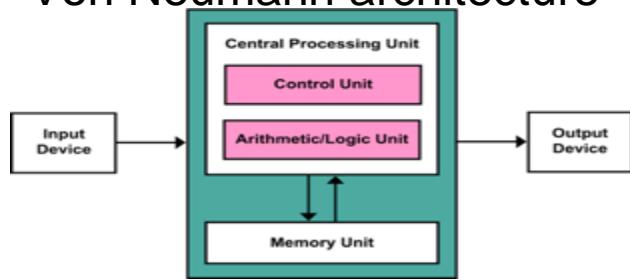
# 數位化的好處

- Information storage is easy.
- Operation can be programmed.
- Digital systems are generally easier to design.
- Accuracy and precision are greater.
- Digital circuits are less affected by noise.

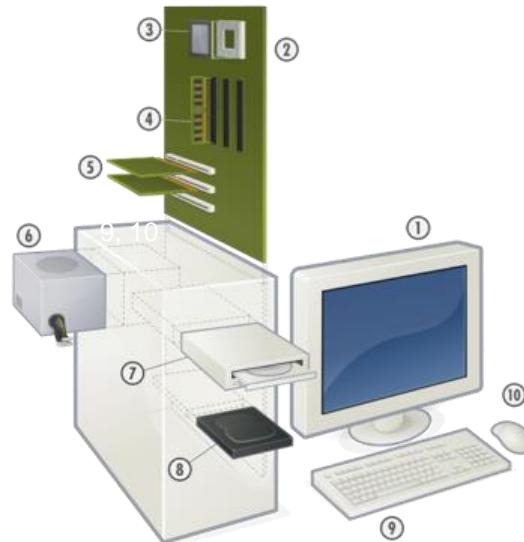
# 電腦硬體

# 電腦硬體

## Von Neumann architecture



| No. | Name           |
|-----|----------------|
| 1   | Monitor        |
| 2   | Motherboard    |
| 3   | CPU            |
| 4   | RAM            |
| 5   | Expansion card |
| 6   | Power supply   |
| 7   | DVD-ROM/BluRay |
| 8   | Hard Disk      |
| 9   | Keyboard       |
| 10  | Mouse          |



# 個人電腦外觀

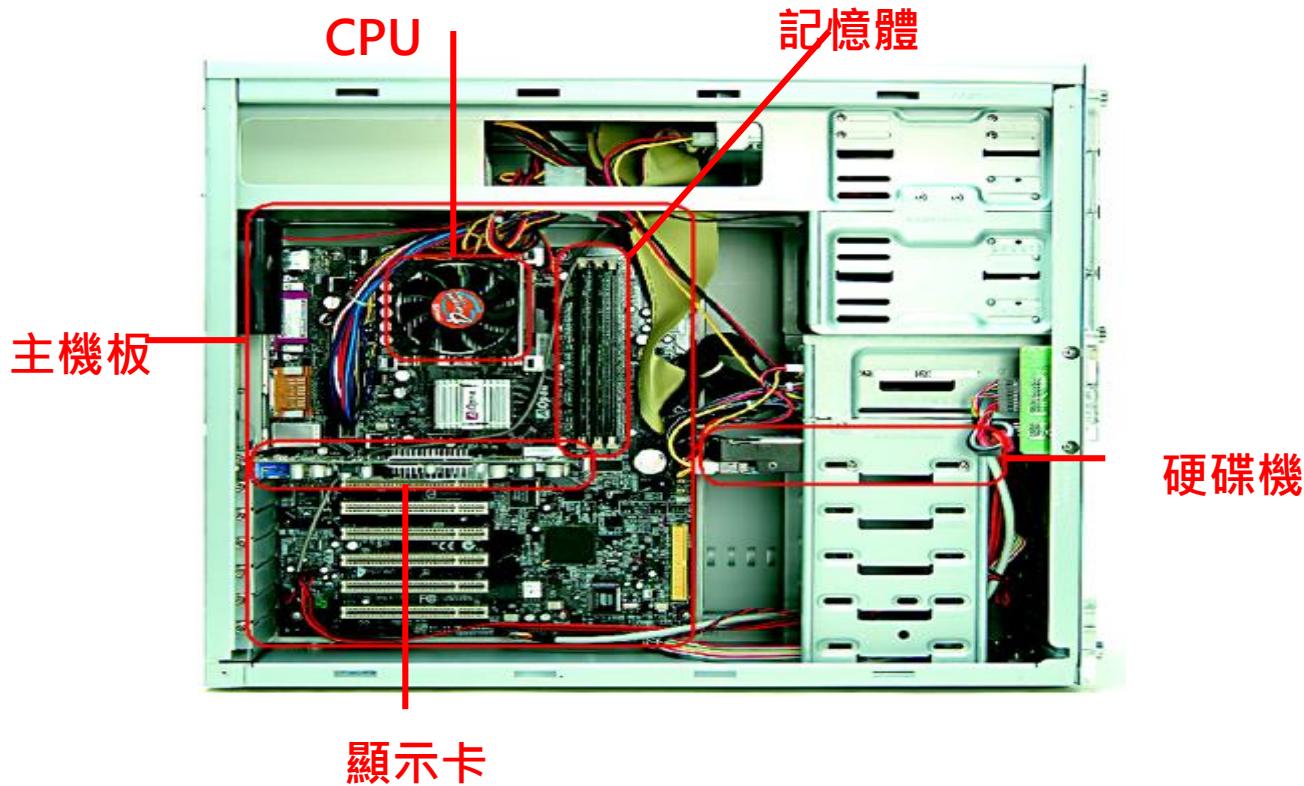


<https://www.asus.com/tw/Tower-PCs/K31AD/overview/>

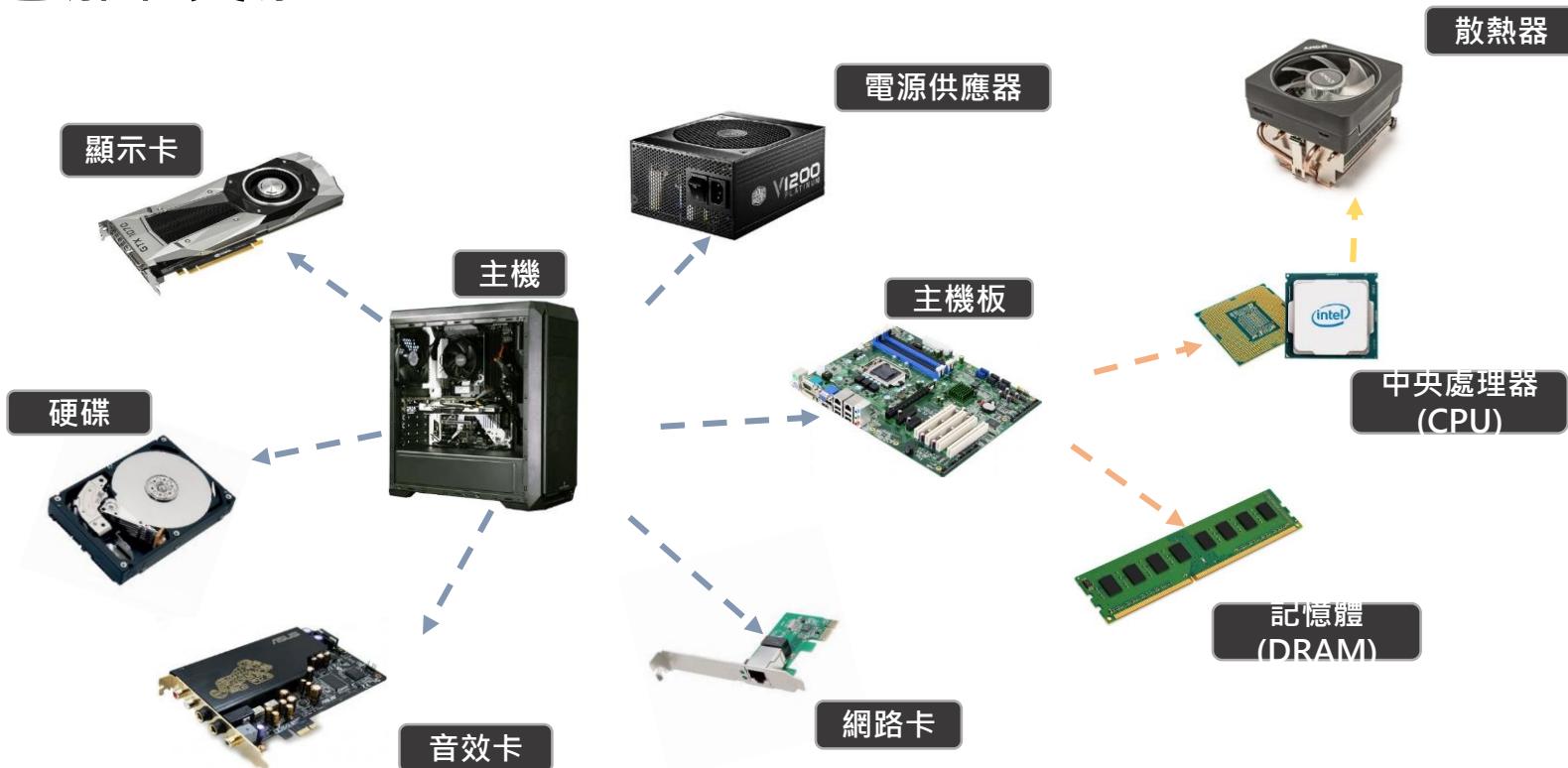


<http://blog.ilc.edu.tw/blog/blog/2206/resource/12526/159715>

# 主機內部



# 電腦硬體



# 主機板

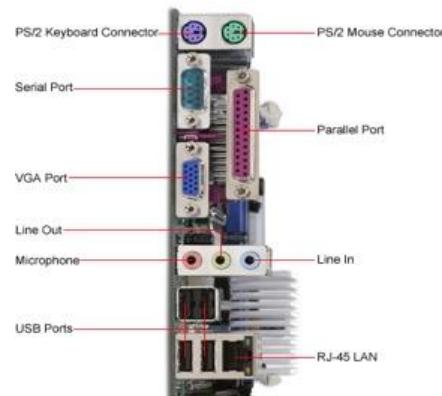
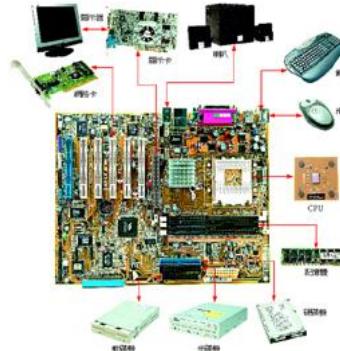
MainBoard/MotherBoard

# 主機板 MainBoard/MotherBoard

電路板，英文簡稱**MB**，是電腦最主要的部份。

電路板上焊/插滿大大小小的電子零件，包括中央處理器(CPU)、晶片組、記憶體及各種介面卡的擴充槽，並以外殼保護，透過電源供應器，將外部電力供應運轉。

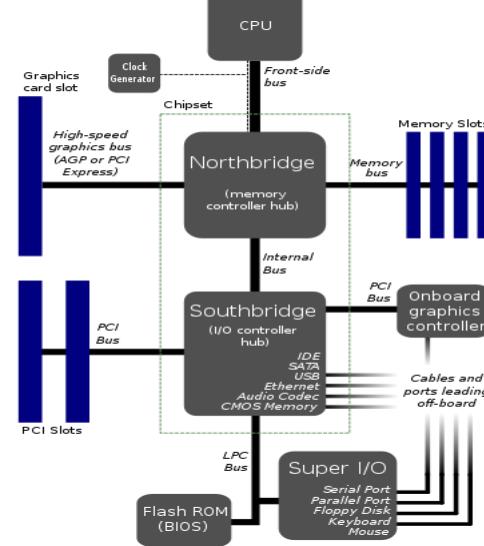
穩定的主機板可以使軟體執行得很平穩，而好的晶片組可以與CPU配合良好，以充分發揮電腦系統的最大功能



# Motherboard Diagram

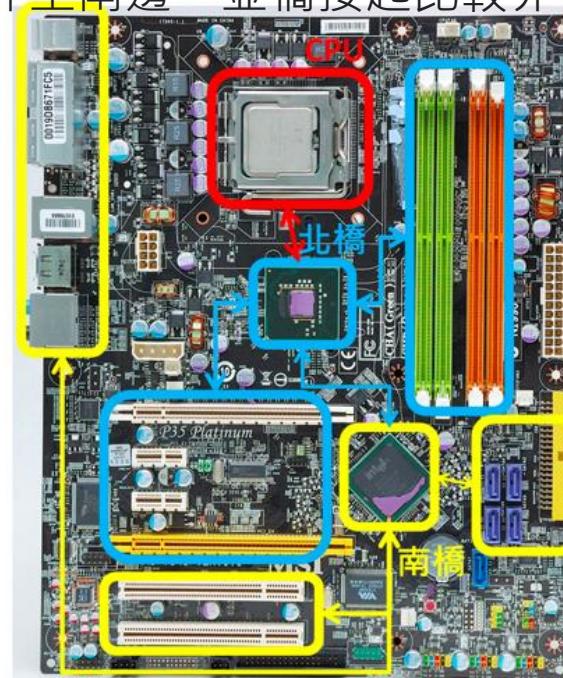
北橋晶片將PCI匯流排主幹延伸至北邊，以支援CPU、記憶體或快取(Cache)、以及其他攸關效能的功能。

反之，南橋晶片將PCI匯流排主幹延伸至南邊，並橋接起比較非攸關效能的功能，例如磁碟介面等、音效等。



<http://www.mobile01.com/topicdetail.php?f=488&t=393140>

<https://zh.wikipedia.org/zh-tw/%E5%8D%97%E6%A1%A5>



[https://commons.wikimedia.org/wiki/File%3AMotherboard\\_diagram.svg](https://commons.wikimedia.org/wiki/File%3AMotherboard_diagram.svg)

# 南北橋晶片整合

- Intel 在 2011 年發表的 Nehalem 系列 CPU，就是把原本負責對記憶體溝通的北橋晶片直接整合進處理器、並把北橋其他剩餘的功能和南橋晶片整合在一起，從此讓北橋晶片從此走入歷史。AMD 在同年發表的 Fusion 處理器同樣取消了北橋。

A newer Motherboard (Core i9 3rd Gen) [NO Northbridge]



儲存裝置 記憶體

Memory Unit

# 儲存架構

## 二級儲存體

速度慢、關機資料不消失

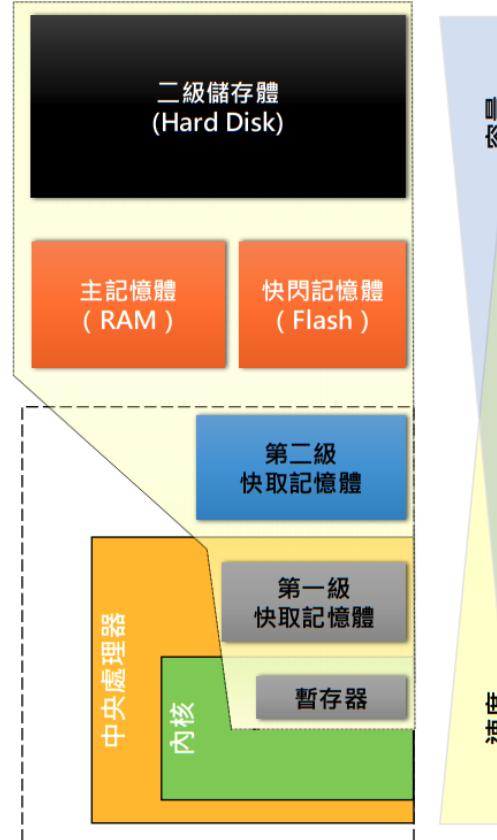
|        |
|--------|
| USB 碟  |
| SD 記憶卡 |
| 固態硬碟   |
| 磁碟     |
| 光碟     |
| 磁帶     |

## 一級儲存體

速度快、關機資料會消失

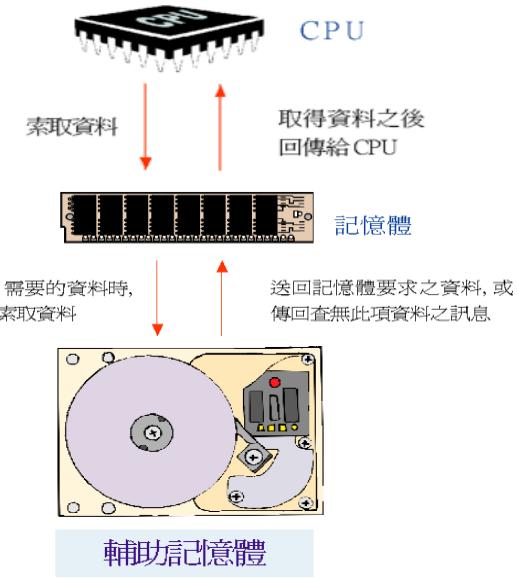
|       |
|-------|
| 暫存器   |
| 快取記憶體 |
| 主記憶體  |

## 資料存取途徑



# 主記憶體 (RAM, Random Access Memory)

- 電腦的主記憶體-隨機存取記憶體 ( RAM ) , 其功能是 CPU 在運作的過程中, 會不斷地索取或產生新的資料, 但是 CPU 本身並沒有足夠的儲存空間可放置這些資料, 因此必須由記憶體來提供。
- 電腦把一個不活動的程式的某些部分移出, 擾出空間給別的程式, 這需要花些時間。若想要讓電腦運轉得更快, 最佳策略可能是購買更多的 RAM
- 優點 : 因為記憶體是以晶片製造而成, 屬於電子式的儲存裝置, 所以存取速度快、體積小
- 缺點 : 當沒有電源時就無法保存資料



# RAM v.s. ROM

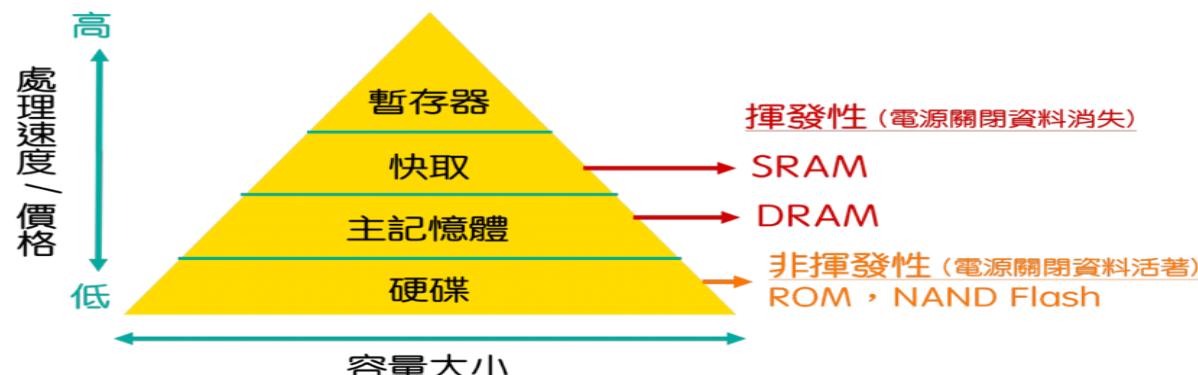
揮發性 (VOLATILE)：停止供應電源，記憶資料便會消失。

- 動態隨機存取記憶體(Dynamic Random Access Memory, DRAM)：主記憶體將每一位元(bit)資料儲存在積體電路(integrated circuit)內的單獨電容器(capacitor)中。
- 靜態隨機存取記憶體(Static Random Access Memory, SRAM)：快取記憶體(Cache)存取速度比主記憶體更快的儲存裝置，減少CPU因等待慢速設備(如主記憶體)所導致的延遲，進而改善系統的效能

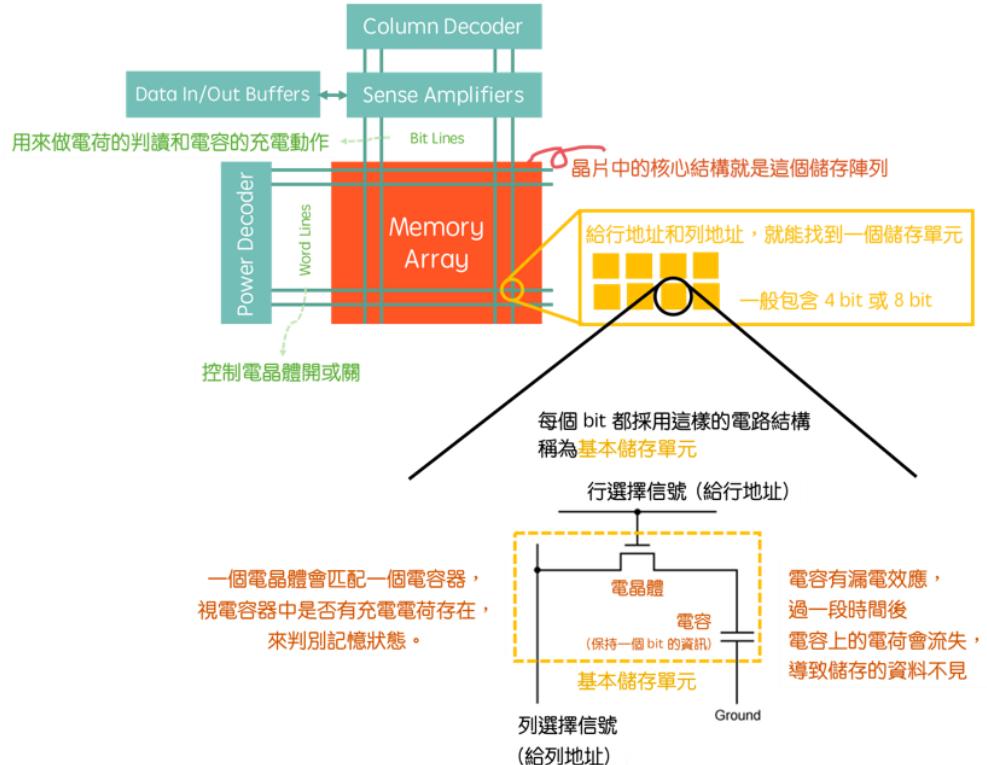
●非揮發性(NON-VOLATILE)：即使沒有供應電源，也能保存已經寫入的資料。

- 唯讀記憶體(Read Only Memory, ROM)
- 快閃記憶體(Flash)

電腦中的儲存單元比較

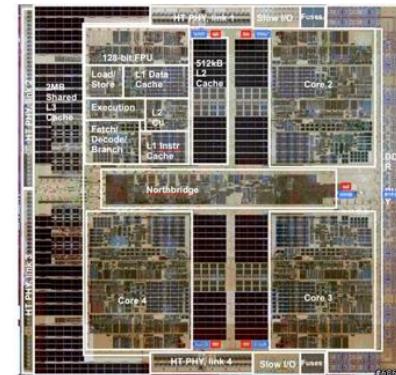
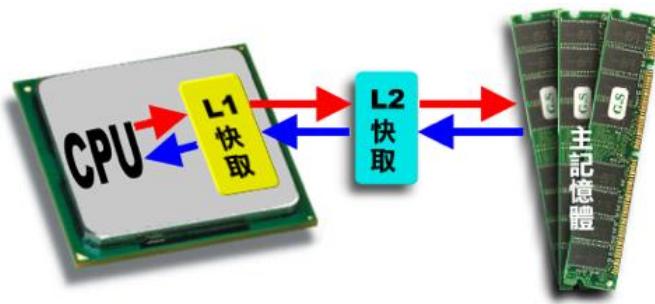


## DRAM 晶片中的結構



# 快取記憶體 (Cache)

- 目前電腦內部以距離 CPU 核心的層數來分, 有 L1、L2 與 L3三種 Cache



# 記憶體 Memory

- 時常會稱記憶體為DDR (Double Data Rate)，雙倍資料率動態隨機存取記憶體 ( Double Data Rate Dynamic Random Access Memory ) 具有雙倍資料傳輸率
- DDR (Double Data Rate), DDR-2, DDR-3



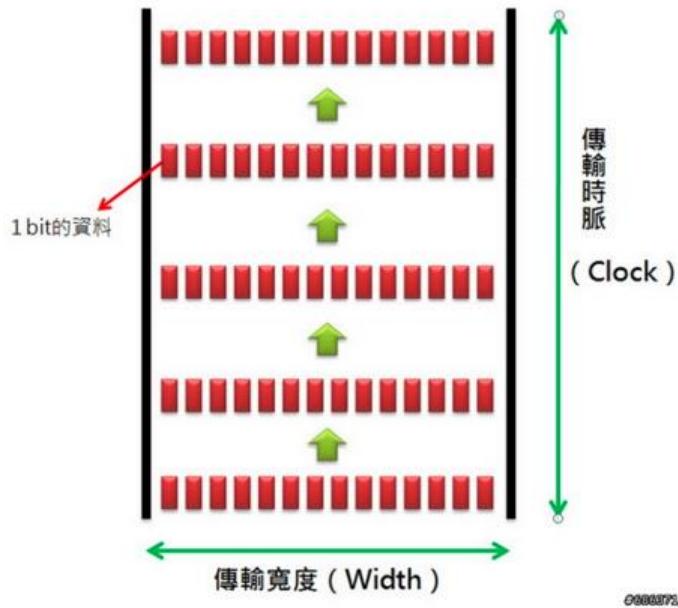
# 記憶體容量單位

- 電腦最基本的儲存單位為位元 (bit)
- Byte = 儲存的單位
  - $1\text{Byte} = 2^3 = 8\text{ bits}$
  - $1\text{KB} = 2^{10} = 1024\text{ Bytes}$
  - $1\text{MB} = 2^{20} = 1024\text{KB} = 1,048,576\text{ Bytes}$  (約1000KB)
  - $1\text{GB} = 2^{30} = 1024\text{MB}$  (約1000MB)
  - $1\text{TB} = 2^{40} = 1024\text{GB}$  (約1000GB)
- 主記憶體 (RAM) 是以 GB來計算
- 硬碟則是以 TB 來計算
  - 例如：硬碟容量2TB約等於 2,000 GB 或 2,000,000MBytes

# 揭開 64 位元 Windows 之謎

- 電腦的位元數表示電腦可處理的資料量、處理資料的速度和記憶體的最大容量
- 64 位元版本 Windows 的電腦通常較 32 位元版本 Windows 的電腦能使用更多記憶體
  - 64 位元版本最少為 4 GB
  - 32 位元版本最大為 3.5 GB。即使電腦已內附安裝 4 GB 或以上的記憶體，32 位元版本的 Windows 僅能使用約 3.5 GB 記憶體。
- 記憶體愈多，愈能同時開啟更多檔案與程式，而不減緩電腦速度。除非同時開啟大量項目，否則即使擁有超過 3.5 GB 的可用空間，通常也並沒有任何差別。

# 記憶體的時脈與頻寬



資料進出記憶體時，傳輸通道有一定的寬度，就是「每一次可傳的資料量」，就系統主記憶體來說是64位元的寬度。時脈(clock)則是「每秒可以傳幾次」，800MHz的記憶體就是每秒800百萬次的傳輸，每次傳輸64位元的資料。

「記憶體頻寬」(Memory Bandwidth)是寬度乘上時脈，變成每秒的資料傳輸量，以800MHz為例就是每秒6.4GB的頻寬。

$$64\text{位元}(\text{傳輸寬度}) * 800\text{MHz}(\text{時脈}) / 8 = \text{每秒 } 6400 \text{ MB} = 6.4\text{GB}$$

註：1 Byte等於8 bit，除8是讓bit位元變成Byte位元組

# 主記憶體中「雙通道」技術

- 指多增加一條記憶體匯流排，讓溝通北橋晶片與記憶體模組之間的頻寬變成2倍。
- 例如原為一次傳輸64Bit的256MB RAM

| 規格   | 一條記憶體 | 兩條記憶體 |        |
|------|-------|-------|--------|
|      | 採單通道  | 採雙通道  |        |
| 傳輸大小 | 64bit | 64bit | 128bit |
| 空間大小 | 256MB | 512MB | 512MB  |

# RAM v.s. ROM

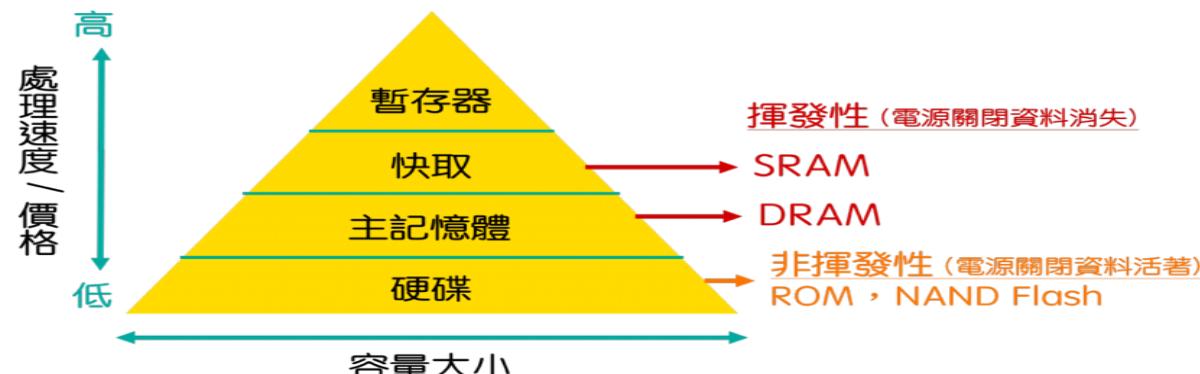
揮發性 (VOLATILE)：停止供應電源，記憶資料便會消失。

- 動態隨機存取記憶體(Dynamic Random Access Memory, DRAM)：主記憶體將每一位元(bit)資料儲存在積體電路(integrated circuit)內的單獨電容器(capacitor)中。
- 靜態隨機存取記憶體(Static Random Access Memory, SRAM)：快取記憶體(Cache)存取速度比主記憶體更快的儲存裝置，減少CPU因等待慢速設備(如主記憶體)所導致的延遲，進而改善系統的效能

● 非揮發性 (NON-VOLATILE)：即使沒有供應電源，也能保存已經寫入的資料。

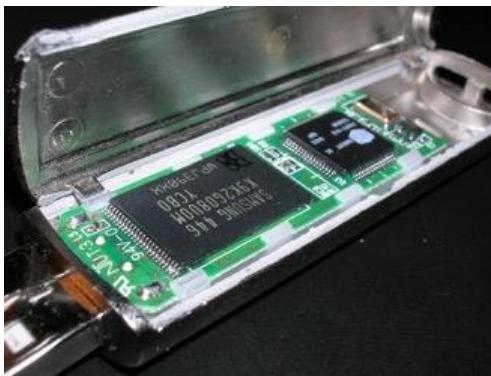
- 唯讀記憶體(Read Only Memory, ROM)
- 快閃記憶體(Flash)

電腦中的儲存單元比較



# 快閃記憶體 ( Flash memory )

- 快閃記憶體 ( Flash memory ) 允許對資料進行多次的刪除、加入或覆寫。
  - 廣泛用於記憶卡、隨身碟之中，其可迅速改寫的特性非常適合手機、筆記型電腦、遊戲主機、掌機之間的檔案轉移。



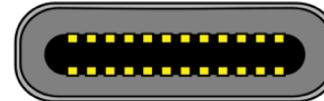
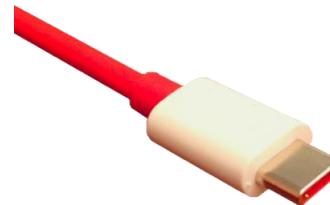
2005年拆解的USB。左邊的晶片是快閃記憶體。控制器位於右側。

# USB (Universal Serial Bus)

- 通用序列匯流排，USB 最初是由 Intel與Microsoft公司倡導發起，是目前個人電腦非常普遍的介面，眾多周邊設備均支援USB介面，使用上非常方便。
- Type-C : 24 針 USB 連接器系統，其特點是其兩折旋轉對稱連接器(two-fold rotationally-symmetrical connector)

現USB標準中，統一為USB 3.0，向下相容，分為：

| USB版本   | 速率稱號              | 頻寬      | 速度                       |
|---------|-------------------|---------|--------------------------|
| USB 3.0 | 超高速<br>SuperSpeed | 5Gbps   | 約500MB/s<br>(5000Mbit/s) |
| USB 2.0 | 高速<br>Hi-Speed    | 480Mbps | 約60MB/s<br>(60,000KB/s)  |
| USB 1.1 | 全速<br>Full Speed  | 12Mbps  | 約1.5MB/s<br>(1,500KB/s)  |
| USB 1.0 | 低速<br>Low Speed   | 1.5Mbps | 187.5KB/s<br>(192000B/s) |



Type-C

# USB

- 支援熱拔插(Hot Swapping)

- 在電腦開機的過程中，可以直接新增及移除硬體。

- 隨插即用(PnP)

- 當插入硬體時不用另外灌入驅動程式，主機會立即偵測其驅動程式，便能有效的使用該硬體。

- 優越的擴充能力，可串接127個周邊設備。

80台目辺りからエラー頻発

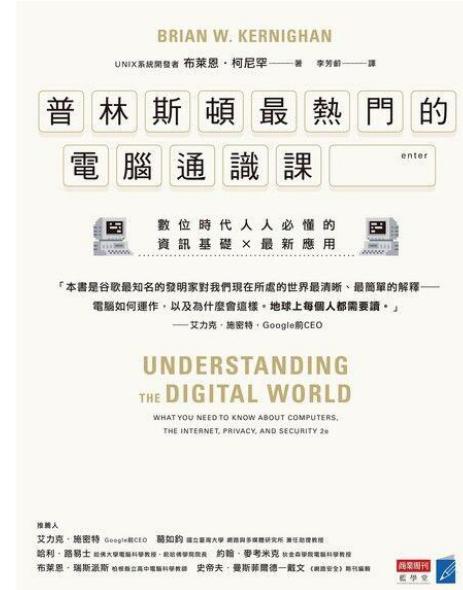


# 中央處理器

Central Processing Units (CPU)

# 中央處理器 ( Central Processing Units, CPU )

- 處理器是電腦的大腦。
- 處理器執行算術，移動資料，控制其他元件的運作
- 處理器能執行的基本運算項目有限，但它執行得飛快，每秒數十億筆。

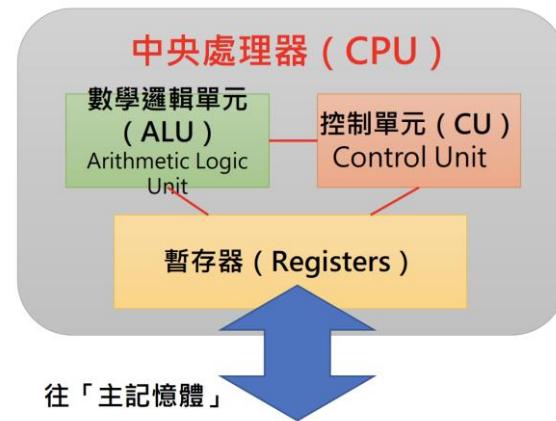
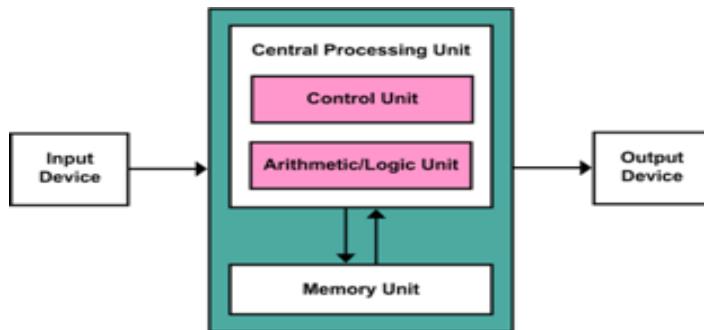


<https://pansci.asia/archives/345255>

# 中央處理器 CPU

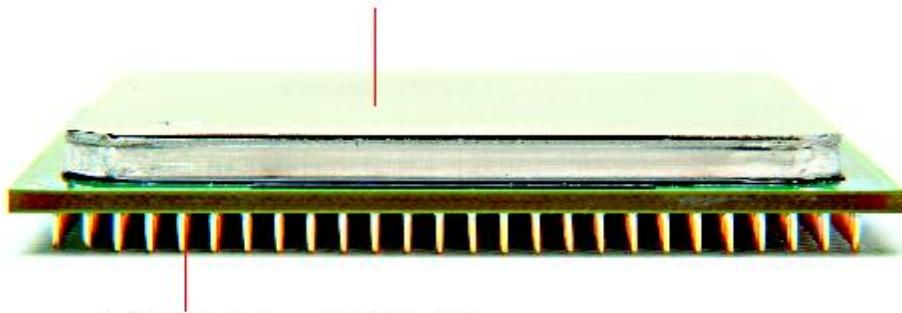
- 計算、判斷
- 指揮全局

Von Neumann architecture



# 中央處理器 CPU

護殼：具有保護晶片與散熱的功能,CPU的晶片核心(Die)就是位在護殼之中



針腳(Pin)：與主機板相接的金屬針腳



<http://news.softpedia.com/news/Intel-s-HEDT-CPU-Roadmap-Exposed-Broadwell-E-in-2015-and-Skylake-E-in-2016-446242.shtml>

# CPU速度(時脈 內頻)

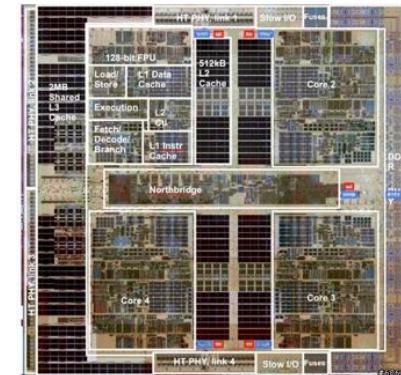
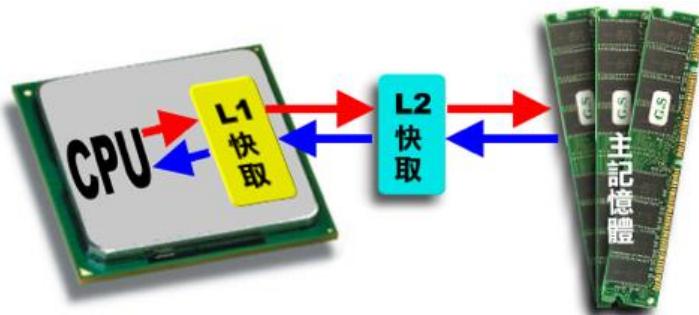
- CPU 速度指的是在單位時間內CPU 內部處理指令的頻率 (以每秒產生多少個脈衝波定義, 脈衝的頻率), 因此也稱為時脈或內頻
- 單位是 MHz (每秒百萬次) 或 GHz (每秒十億次)
  - $1\text{MHz} = 1,000,000\text{Hz}$
  - $1,000\text{MHz} = 1\text{GHz}$
  - ie, CPU:1000MHz = CPU以每秒10億次的頻率在工作
- 內頻愈高, 表示 CPU 的處理速度快
- Performance charts

# CPU 外頻

- 外頻就是指 CPU 外部與主機板之間的運作頻率, 也稱為 Host Clock
- 常見的速度是 333/400 MHz, Pentium 4 Prescott 外頻高達 800 MHz
- 如果CPU與外部「交談」非常頻繁，外頻將可能成為運算的瓶頸

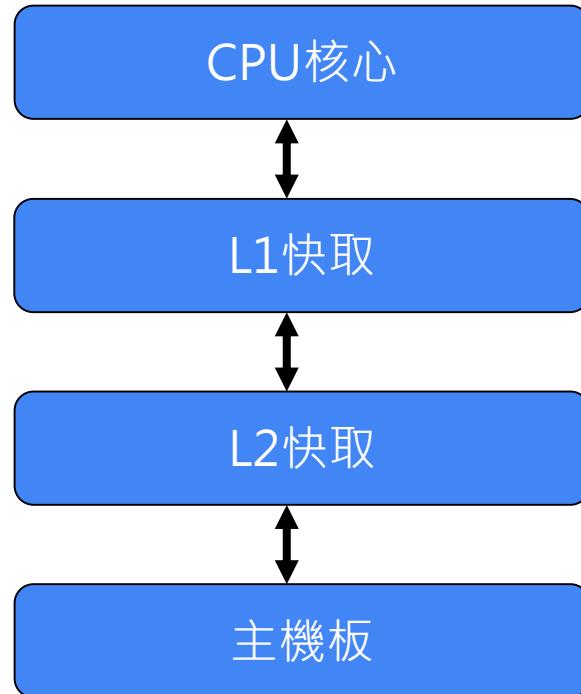
# 快取記憶體 (Cache)

- 快取記憶體存取速度比主記憶體更快的儲存裝置，減少 CPU 因等待慢速設備 (如主記憶體) 所導致的延遲，進而改善系統的效能
- 目前電腦內部以距離 CPU 核心的層數來分，有 L1、L2 與 L3 三種 Cache



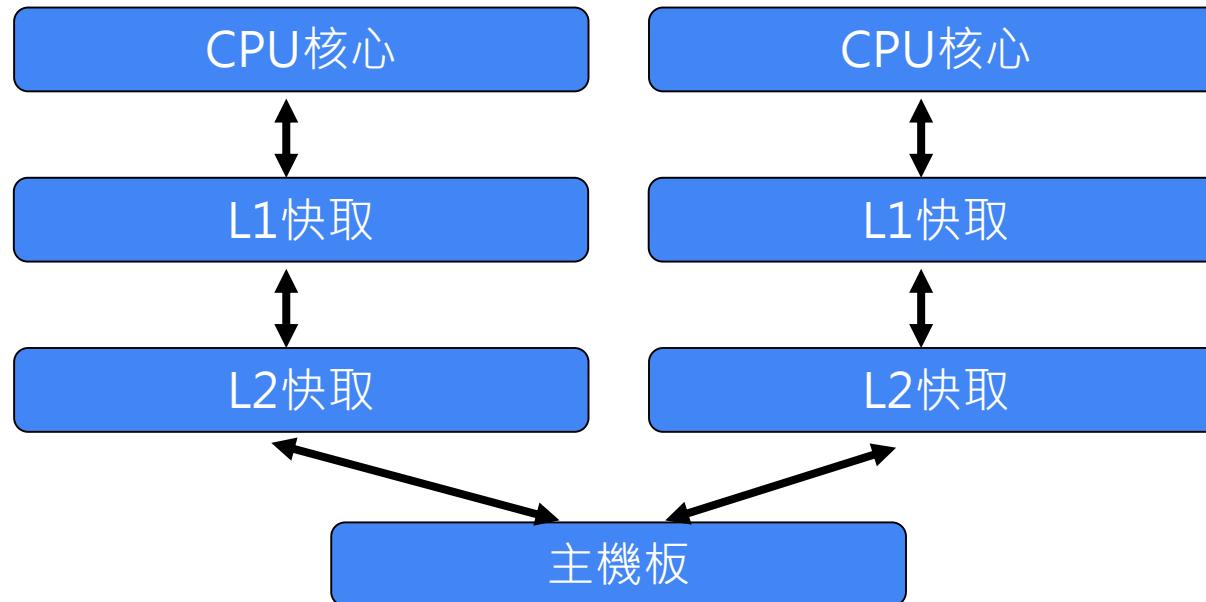
# 單核心CPU

單核心指的是一顆CPU裡只有一個實體核心晶片



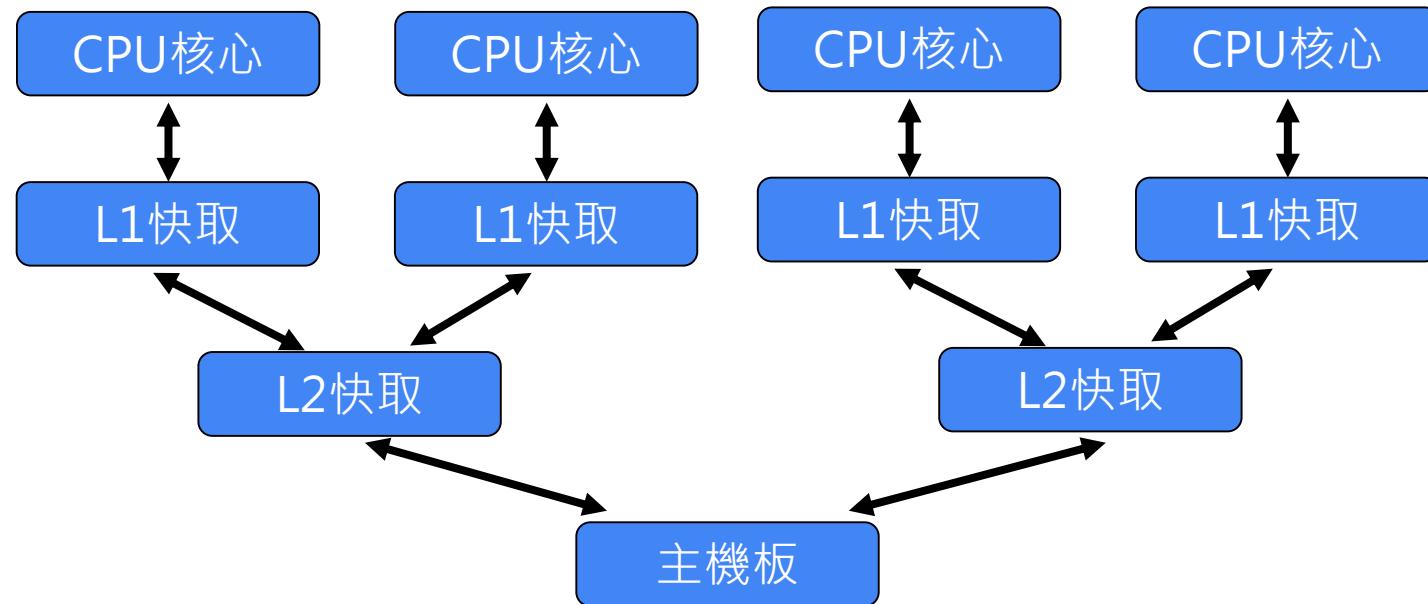
# 雙核心CPU

- 雙核心是指一顆CPU裡面有『兩個』獨立的運算單元，有兩個實體核心晶片。

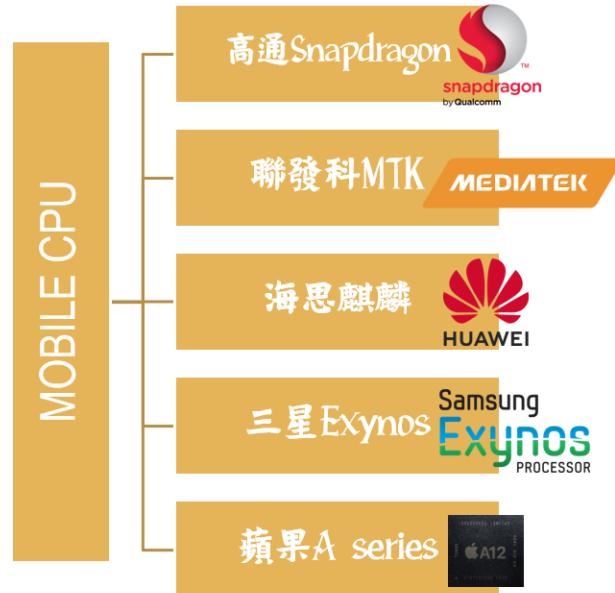


# 四核心CPU (Core 2 Quad)

- 兩顆雙核心晶片封裝在一顆CPU內則是四核心



# 手機處理器 Mobile CPU



Credit by 109308025

<https://www.landtop.com.tw/reviews/6>

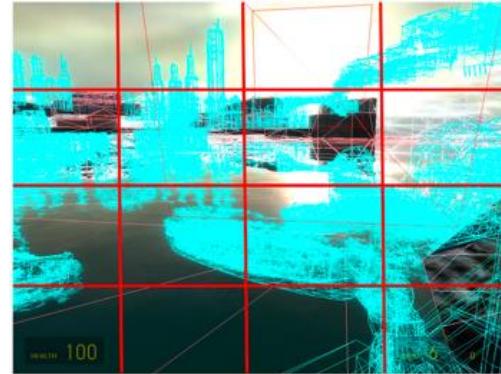
# CPU 歷史

- Intel vs AMD
- AMD CPU 製程
- Intel vs ARM
- 行動處理器爭霸戰：Intel最大危機不是AMD？看ARM如何布局行動處理器市場
- [https://kopu.chat/2019/01/30/armcpu/?fbclid=IwAR22wejyCCCN054a\\_lqEW\\_zGCUhYyiburpDMQ8-fa50Vr81LAGAkytNSIQ](https://kopu.chat/2019/01/30/armcpu/?fbclid=IwAR22wejyCCCN054a_lqEW_zGCUhYyiburpDMQ8-fa50Vr81LAGAkytNSIQ)

# 圖形處理器/顯示卡

Graphic Processing Unit (GPU)

# 顯示卡在做什麼？



<http://www.mobile01.com/topicdetail.php?f=298&t=414913&last=3759338>

# 顯示卡

- 顯示卡 (display card) 簡稱顯卡，是個人電腦上以圖形處理器 (GPU) 為核心的擴充卡
- 幫助計算圖像資訊，並將電腦系統所需要的顯示資訊進行轉換並提供逐行或隔行掃描訊號給顯示裝置，是連接顯示器和個人電腦主機板的重要元件，是「人機對話」的重要裝置之一。
- 生成並輸出圖像到顯示器，提供附加功能，例如
  - 3D場景加速渲染 (accelerated rendering of 3D scenes)
  - 影片截取
  - MPEG-2 和 MPEG-4 解碼 (decoding)
  - 非圖像處理：比如炒賣加密貨幣（「挖礦」）和AI強化學習（例如ChatGPT）。



以太幣礦機



Nvidia GeForce 9600GT, PC-e interface

# 顯示卡介面

- VGA (D-SUB)
- DVI ( Digital Visual Interface, 數位視訊介面 )
  - 透過數位化的傳送來強化個人電腦顯示器的畫面品質。比VGA介面要好，全部內容採用數位格式傳輸，保證主機到顯示器的傳輸過程中資料的完整性，無干擾信號引入，可得到更清晰的影像
- HDMI ( High Definition Multimedia Interface, 高清晰度多媒體介面 )
  - 全數位化影像和聲音傳送介面，可以傳送無壓縮的音頻信號及視頻信號
    - 被設計來取代較舊的類比影音傳送介面如 SCART或RCA等端子，可用於機上盒、DVD播放機、個人電腦、電視遊樂器、綜合擴大機、數位音響與電視機等設備
    - HDMI的纜線長度限制大約在5公尺內

# CPU 和 GPU 平行運算的效能差異

- GPU 本來的設計只是為了用在遊戲圖形渲染
- GPU 在神經網絡訓練中的計算速度要比 CPU 高很多。現在 GPU 就和 CPU 聯手負責計算機中的運算
- Google 造了自己的專用晶片 : TPU
  - Credit by 劉庭瑋, 科技橘報
  - <https://buzzorange.com/techorange/2017/09/27/what-intel-google-nvidia-microsoft-do-for-ai-chips/>
- Mythbusters Demo GPU versus CPU
  - <https://youtu.be/-P28LKWtZrl>

# Moore's Law graphed vs real CPUs & GPUs

## 1965 - 2019

- [https://youtu.be/7uvUiq\\_jTLM](https://youtu.be/7uvUiq_jTLM)
- WSJ：摩爾定律已失效，「黃仁勳定律」將正式取代！
  - By 科技橘報
  - [https://buzzorange.com/techorange/2020/09/21/huangslaw/?fbclid=IwAR3Kf7fpGleuTPhpwlhl1TTwqh-W-FuvZ0KKpnEr6EjF8U\\_I4LPAAnMpOYK4](https://buzzorange.com/techorange/2020/09/21/huangslaw/?fbclid=IwAR3Kf7fpGleuTPhpwlhl1TTwqh-W-FuvZ0KKpnEr6EjF8U_I4LPAAnMpOYK4)

# 儲存裝置 – 硬碟

Hard Disk Drive

# 儲存架構

## 二級儲存體

速度慢、關機資料不消失

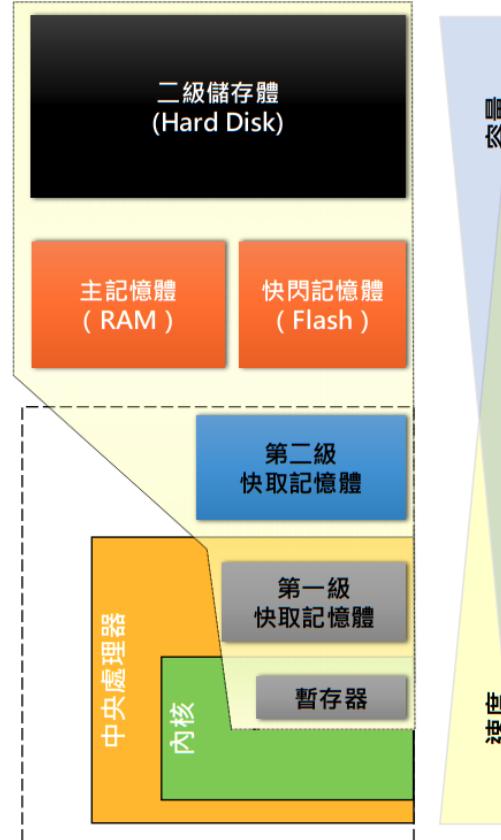
|        |
|--------|
| USB 碟  |
| SD 記憶卡 |
| 固態硬碟   |
| 磁碟     |
| 光碟     |
| 磁帶     |

## 一級儲存體

速度快、關機資料會消失

|       |
|-------|
| 暫存器   |
| 快取記憶體 |
| 主記憶體  |

## 資料存取途徑



# 硬碟 Hard Disk Drive (HDD)

- 電腦程式與資料長久儲存的主要設備
- 一種非揮發性(non-volatile)儲存設備
- 可將數位編碼資料(digitally encoded data )儲存在具有磁性表面的快速旋轉盤片上

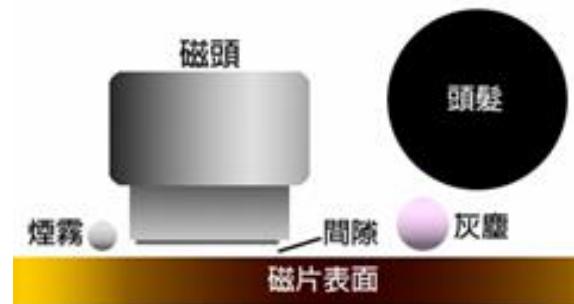
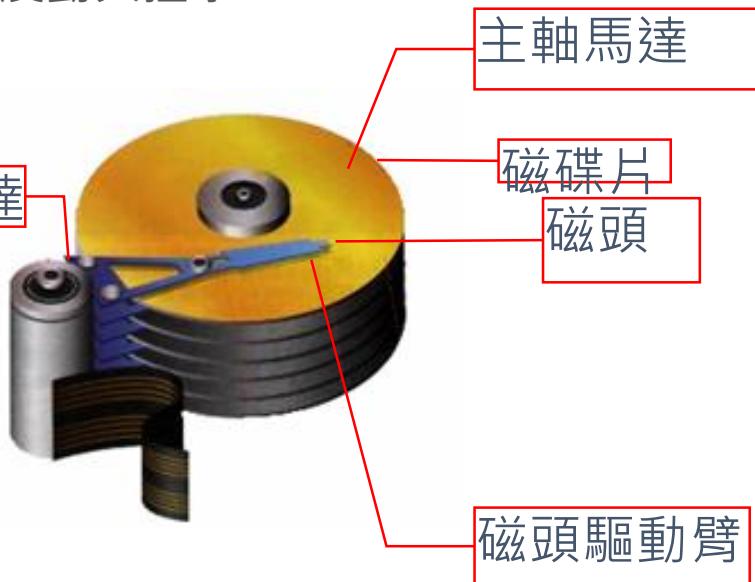


[小惡魔的電腦教室] 5-1. 認識硬碟, 硬碟的資料讀取/寫入裝置在進行資料存取前後的運作狀態 (原圖取自WD官方網頁)。

# 硬碟的結構

- 切忌震動與撞擊

磁頭定位馬達



硬碟機磁頭與磁片的間隙只有  
2/1000000吋，比頭髮更微小！

# 硬碟的轉速

- 硬碟馬達旋轉的速度單位為 RPM (Rotations Per Minute, 每分鐘多少轉)



外包装盒上所標示的  
硬碟轉速，在此為  
7,200 RPM

# 磁碟存取時間

- 存取時間 = 搜尋時間 + 旋轉時間 + 傳輸時間

- 搜尋時間 (Seek time)

- 讀寫頭 (read/write head) 移動到資料所在磁軌的正確位置，所需的時間 2-15ms。

# 硬碟的容量

- 1956年，IBM 推出了 305 RAMAC，第一台配備 HDD 的「超級」計算機重達一噸多，儲存了 5 MB 的資料。



**TOSHIBA 3TB/SATAIII/3.5吋 硬碟 7200轉 (DT01ACA300)**



- ◆ 高效能
- ◆ 高穩定
- ◆ 低耗電
- ◆ **7200轉**
- ◆ SATAIII介面
- ◆ 64MB緩衝記憶體
- ◆ 環保無鹵素設計
- ◆ 先進格式化硬碟技術
- ◆ **三年保固**

建議售價 \$69  
網路價 \$3

<http://thenextweb.com/shareables/2011/12/26/this-is-what-a-5mb-hard-drive-looked-like-in-1956-required-a-forklift/>

# 硬碟特色

- 外形尺寸
- 3.5 吋：用於大多數桌上型電腦
- 2.5 吋：便攜式高清HD
- 1.8 吋：iPod或基於HDD的便攜式MP3播放器



TOSHIBA 3TB/SATAIII/3.5吋



3TB  
2.5"/ USB3.0

# 硬碟介面(Interface)

- IDE (Integrated Drive Electronics，整合驅動電子裝置硬碟) 本意為「把控制器與盤體整合在一起的硬碟」，只需用一根電纜將它們與主機板或介面卡連起來就可以了。一般亦稱為 ATA ( Advanced Technology Attachment，先進技術附件)，是用於連接儲存裝置的介面標準硬碟，後改名為PATA ( 並列高技術組態，Parallel ATA )
- SATA (Serial ATA) : 新型硬碟介面，因採用串列方式傳輸數據而得名；速度比以往更加快捷；並支持熱插拔，使電腦運作時可以插上或拔除硬體

# 儲存架構

## 二級儲存體

速度慢、關機資料不消失

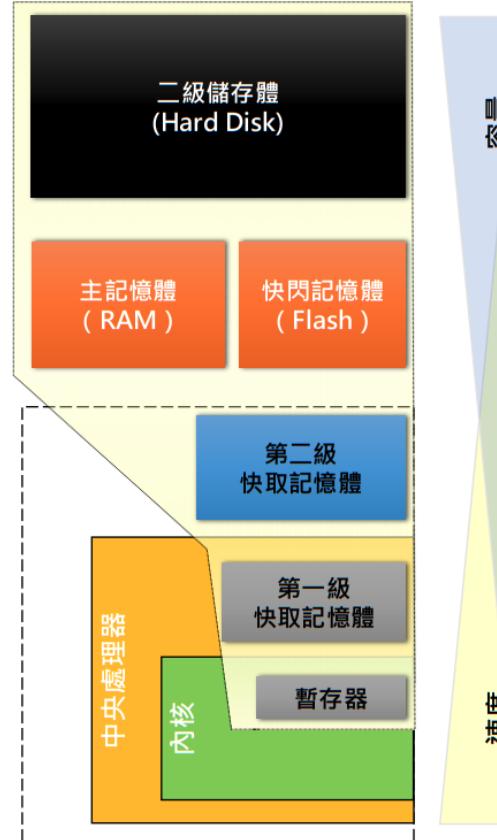
|        |
|--------|
| USB 碟  |
| SD 記憶卡 |
| 固態硬碟   |
| 磁碟     |
| 光碟     |
| 磁帶     |

## 一級儲存體

速度快、關機資料會消失

|       |
|-------|
| 暫存器   |
| 快取記憶體 |
| 主記憶體  |

## 資料存取途徑



# 固態硬碟 Solid-State Drive (SSD)

- 使用固態記憶體(solid-state memory)來儲存持久性資料(persistent data)。
- Magnetic Drive vs. SSD

- 優點

- 低功耗、無噪音、抗震動、低熱量

- 缺點

- 成本較高

- 讀寫次數 :壽命方面，由於快閃記憶體上每一個電閘都有一定的寫入次數限制，壽命結束後會無法寫入變成唯讀狀態 (wikipedia)



# 輸入/輸出裝置

# 輸入設備 - 鍵盤

## ●按鍵數

- 桌上型電腦=104

- 筆記型電腦=87

## ●QWERTY 鍵盤



<https://zh.wikipedia.org/wiki/File:Qwerty.svg>

Christopher Latham Sholes  
(02/14, 1819 – 02/17, 1890)



[https://en.wikipedia.org/wiki/Christopher\\_Latham\\_Sholes](https://en.wikipedia.org/wiki/Christopher_Latham_Sholes)

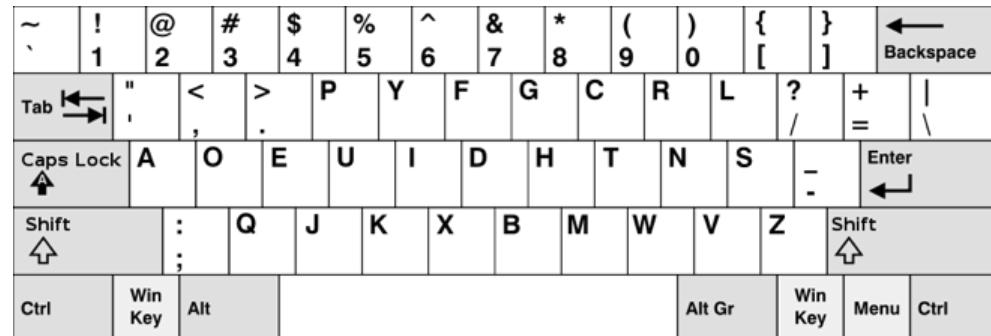
# QWERTY vs Dvorak

- QWERTY 鍵盤缺點

- 左手卻負擔了57%的工作
- 兩小指及左無名指是最沒力氣的指頭，卻頻頻要使用它們
- 排在中列的字母，其使用率僅佔整個打字工作的30%左右

- Dvorak 鍵盤 by August Dvorak

- 將常用字母和二合字母應放在鍵盤中間近食指的位置
- 右手應該有略高的輸入頻率(因為多數人都是右撇子)
- 輸入應由邊緣鍵位打到中間
- 經常相連的字母應該分開兩手輸入



# 輸出設備 - 顯示器

- 顯示器的規格與品質
  - 尺寸、像素的距離
  - 解析度
  - 低幅射
  - 符合環保與省電標準等

## › 螢幕主機規格

|               |                            |
|---------------|----------------------------|
| 螢幕尺寸          | 23吋寬 (16:9寬螢幕)             |
| 最高解析度         | 1920x1080 @60Hz            |
| 可視角度          | H:170°, V: 160°            |
| 亮度            | 250 cd/m <sup>2</sup>      |
| 對比            | 100,000,000:1(ACM)         |
| 總體訊號反應時間 (毫秒) | 2 ms 高速傳輸                  |
| 智慧型畫面自動調校OSD  | Yes                        |
| 點距(mm)        | 0.265mm                    |
| 訊號輸入          | D-Sub/HDMI™2(w/ HDCP)      |
| VESA          | 100x100 mm                 |
| 可視範圍 (公厘)     | 509.76x286.74              |
| 多媒體喇叭         | 多媒體喇叭1.5Wx2                |
| 重量            | 4.1Kgs                     |
| 尺寸            | 553x410x188 (mm)           |
| 標準配件          | VGA線 / 音源線 / 電源線 / 說明書     |
| 保固            | 三年完全保固：零件、檢修工資、液晶面板、燈管之保固。 |

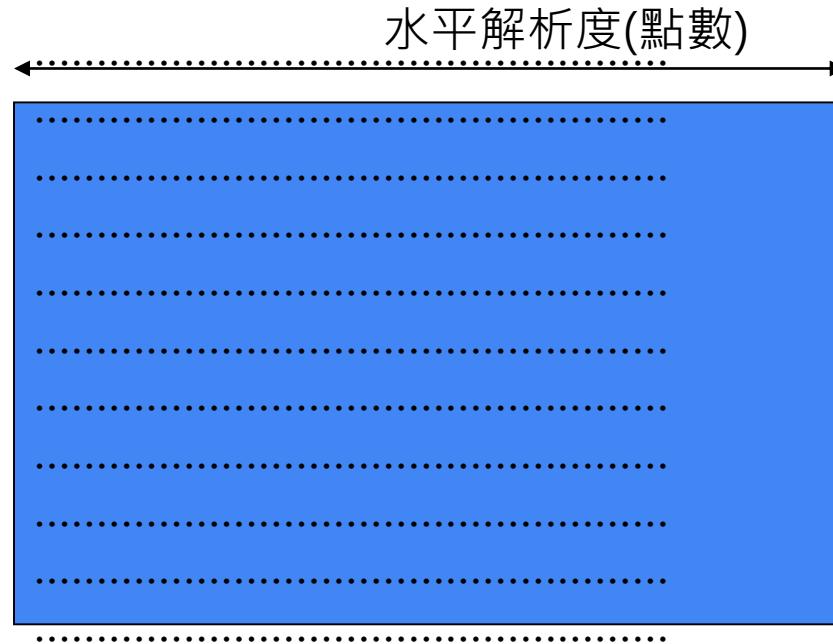


# 解析度

- $1920 \times 1080 = 1920$  是水平解析度；1080是垂直解析度。

螢幕的寬度上有 1,920 畫素，  
高度上有 1,080 畫素，因此總  
共有 2,073,600 個畫素

垂直解析度  
(掃瞄線數)



# 週邊設備 - 光碟機

- DVD-ROM或DVD光碟機
  - 光碟片容量在4.7GB以上
- 藍光光碟 ( Blu-ray Disc , BD )
  - 由Sony及松下電器等企業組成的「藍光光碟聯盟」策劃
  - 儲存大小
    - 單面25GB



# 風扇

- 幫助電腦散熱，避免因過熱所產生之硬體問題



<http://24h.pchome.com.tw/prod/DRAE2W-A9006EBMV?q=/S/DRAE3I>

# 大數據

新興資料科學的發展與應用

天下雜誌

天下雜誌出版

雲端時代的殺手級應用

**BIG DATA**

海量資料分析

全世界同一時間

有超過**5億**支智慧型手機

**10億**台電腦和數兆個感測器同時運作

每天產生的資料量高達**25億** GB

等於要用**4000萬**台64GB的iPad才能裝載

分析顧客行為

提升企業獲利

疏解交通狀況

預測失業率

預防恐怖攻擊

打造即時服務

預測股票指數

減低犯罪率

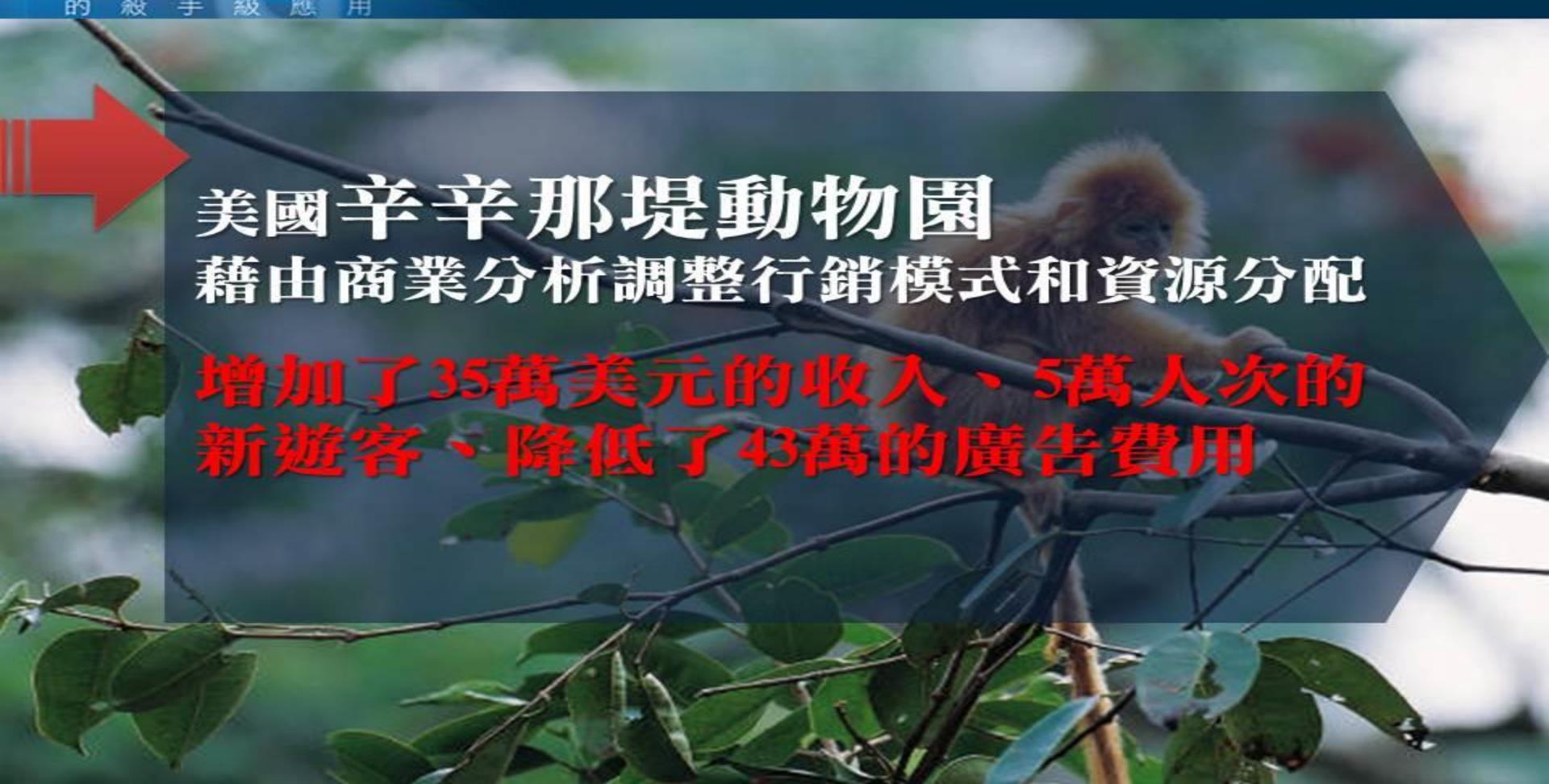
改善服務流程

控制疫情蔓延 降低營運成本

提高醫學研發成效

增加作物產量

預知通膨變化



美國辛辛那堤動物園  
藉由商業分析調整行銷模式和資源分配

增加了35萬美元的收入、5萬人次的新遊客、降低了43萬的廣告費用

美國疾病控制及預測中心  
和Google合作，以關鍵字搜尋次數發展  
了Google Flu Trends，追蹤流感傳播趨勢

提早兩星期掌握流感爆發關鍵時刻

# Google Flu Trends Data (GFD)

- 谷歌流感趨勢未卜先知的故事，常被看做大數據分析優勢的明證。
- 2008年11月谷歌公司啟動的gft項目，目標是預測美國疾控中心(cdc)報告的流感發病率。甫一登場，gft就亮出十分驚艷的成績單。2009年，gft團隊在《自然》發文報告，只需分析數十億搜尋中45個與流感相關的關鍵詞，gft就能比cdc提前兩周預報2007-2008季流感的發病率。

## LETTERS

---

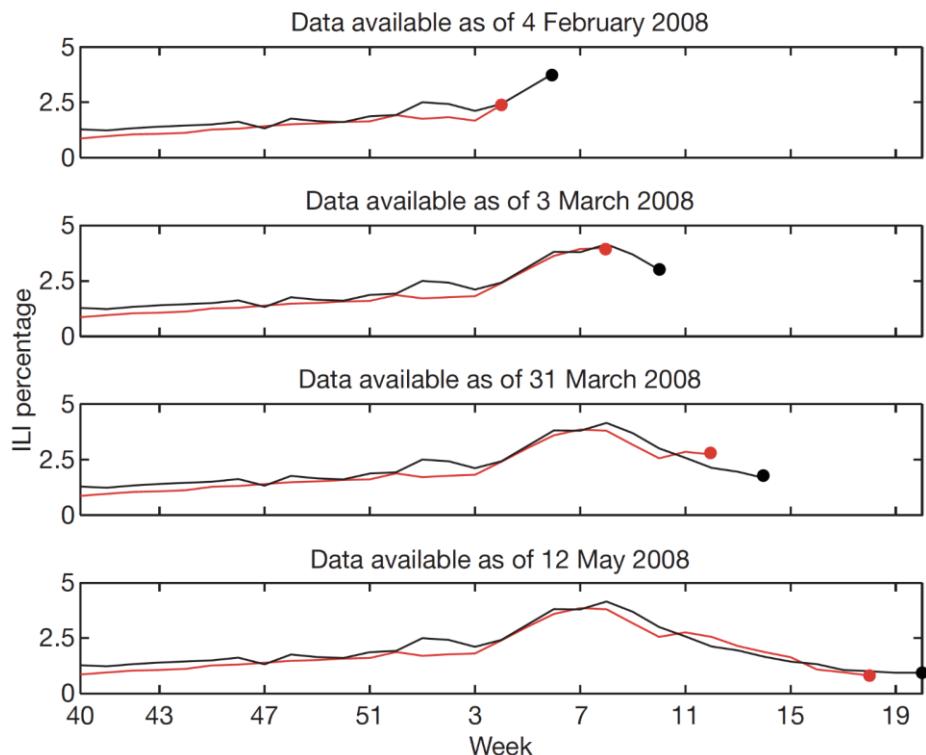
## Detecting influenza epidemics using search engine query data

Jeremy Ginsberg<sup>1</sup>, Matthew H. Mohebbi<sup>1</sup>, Rajan S. Patel<sup>1</sup>, Lynnette Brammer<sup>2</sup>, Mark S. Smolinski<sup>1</sup> & Larry Brilliant<sup>1</sup>

### ● influenza-like illness (ILI)

- $I(t)$ : the percentage of ILI physician visits at time  $t$
- $Q(t)$ : the ILI-related query fraction at time  $t$
- $\text{logit}(I(t)) = \alpha \text{logit}(Q(t)) + \epsilon$ 
  - the 50 million candidate queries
- <https://www.google.org/flutrends/about/>

# ILI percentages estimated by our model (black) and provided by the CDC (red) in the mid-Atlantic region

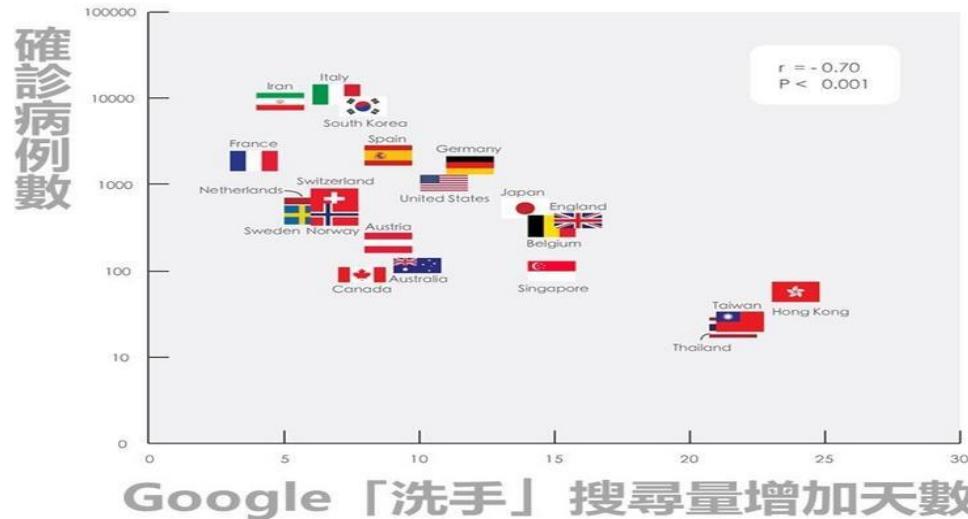


# 從谷歌流感趨勢談大數據分析的光榮與陷阱

- Credit by <https://news.cnnes.com/news/id/336695>
- 2014年, lazer等學者在《科學》發文報告了gft近年的表現
  - 2009年 : gft沒有能預測到非季節性流感a-h1n1
  - 2011年8月到2013年8月的108周里 : gft有100周高估了cdc報告的流感發病率。高估有多高呢?
    - 2011-2012季 : gft預測的發病率是cdc報告值的1.5倍多
    - 2012-2013季 : gft流感發病率已經是cdc報告值的雙倍多了
- Lazer, D., Kennedy, R., King, G. & Vespignani, A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 343, 1203–5 (2014).  
<http://science.sciencemag.org/content/343/6176/1203>

# 防疫措施有沒有做？Google最知道：「洗手」關鍵字愈熱搜、該國病例數愈少

- Credit by 報導者 2020/4/24 文字 張靜慧
- <https://www.twreporter.org/a/covid-19-taiwan-research-google-keywords-wash-hands>



加拿大安大略理工大學  
建立早產兒健康監護系統

協助醫護人員提前24小時預防早  
產兒因敗血症引發的感染

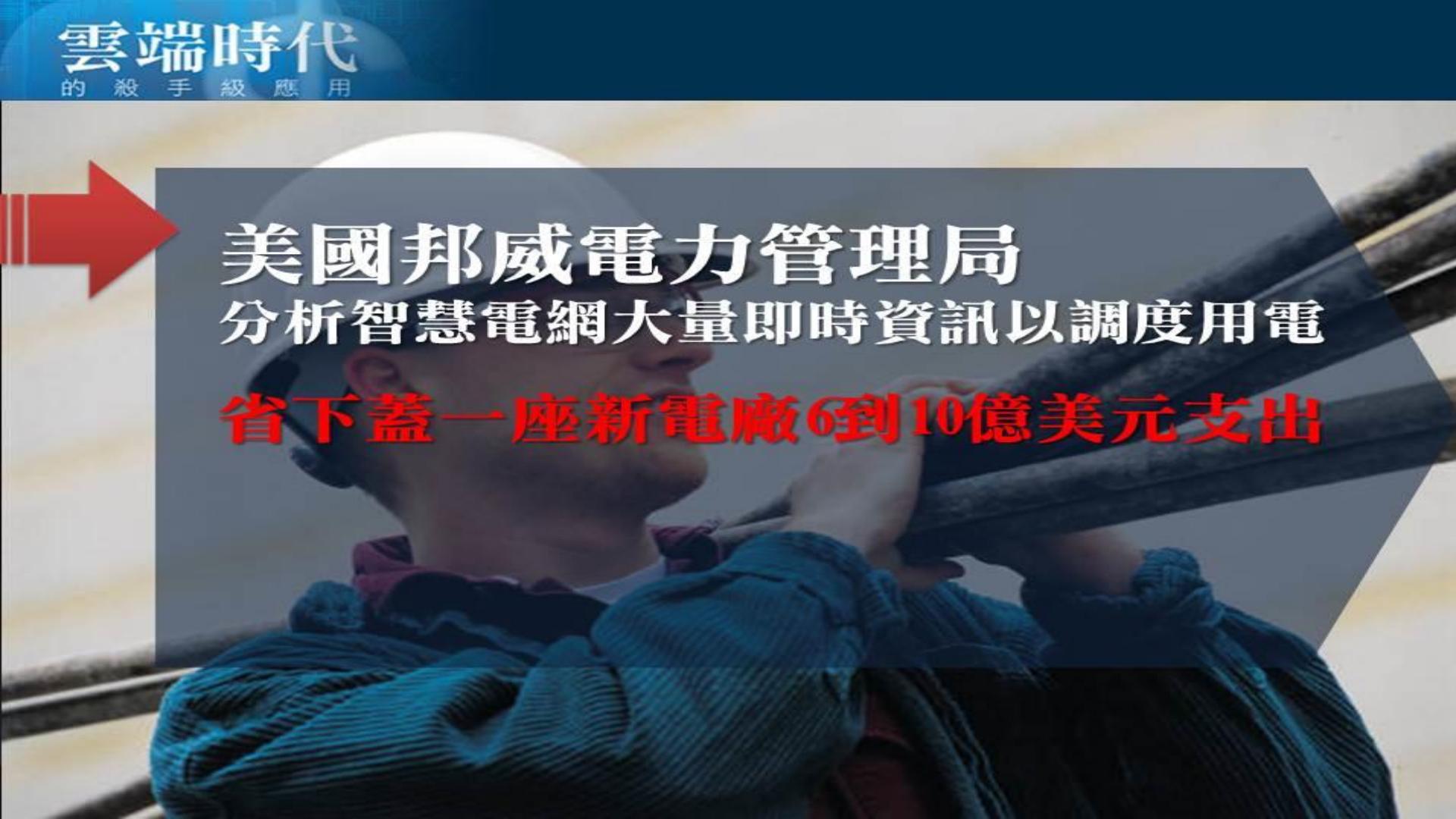


海量資料分析

A large, semi-transparent dark grey rectangular box covers the left side of the image, containing white and red text. A red arrow points from the top-left towards the text.

溫布頓網球公開賽  
運用雲端科技，推出即時賽事分析

讓球迷身歷其境體驗賽事  
打造最智慧的專業網球錦標賽

A photograph of a man from the chest up, wearing a blue corduroy jacket over a pink shirt. He is looking down at a tablet device he is holding in his hands. The background is blurred, showing what appears to be an industrial or construction setting with pipes and structures.

美國邦威電力管理局  
分析智慧電網大量即時資訊以調度用電  
省下蓋一座新電廠6到10億美元支出

羅森(Lawson)便利商店  
分析社群網站上的大量資訊，  
即時調整商品策略

稅前營收五年增加60%



海量資料分析

古都倫敦

因應奧運，整合大量資料與科技設備

讓遊客隨時查詢地鐵動態

因應人潮調度車輛

智慧型垃圾桶能自動分類  
還能預防犯罪與恐怖攻擊



海量資料分析

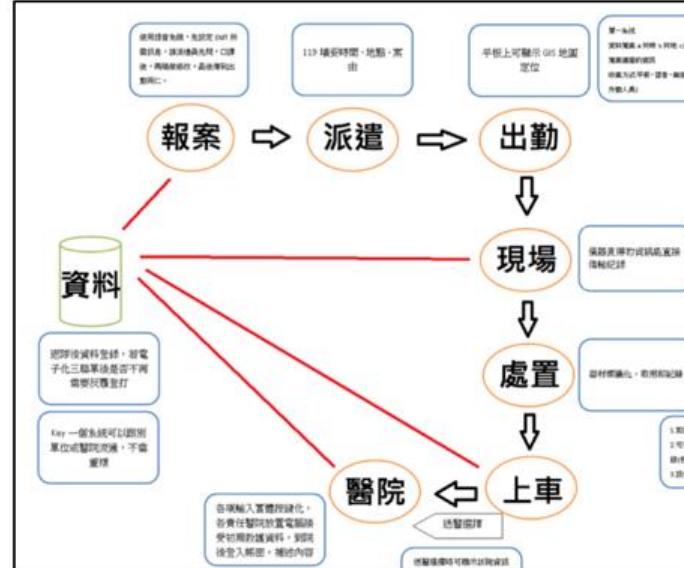
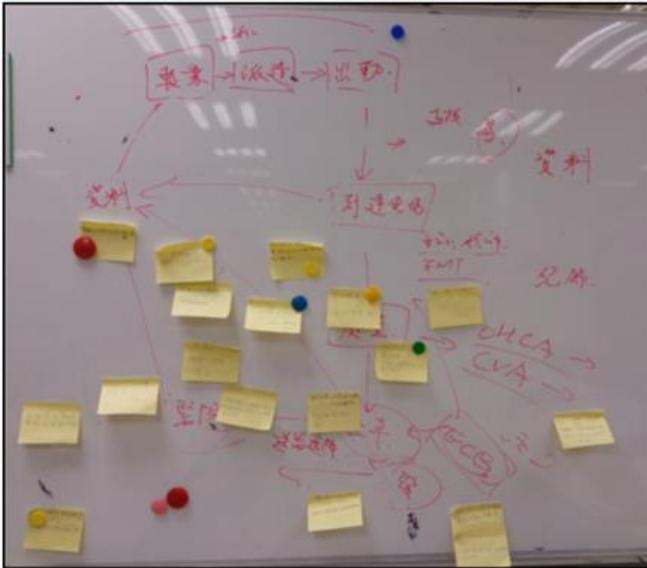


發生氣爆或地震時  
如何合理分配大量傷患？

高雄氣爆 (2014.07.31)

Credit by 謝宗震 智庫驅動

# 消防弟兄的規劃與許願清單



局部放大

處置

同步顯示醫院  
床數、ICU、  
心導管室狀態

能夠顯示患者  
過往119求救  
記錄、病史

語音辨識輸入  
譬如：口述血壓  
120/80，即自動  
輸入資料

# 每人每天產生多少資料？

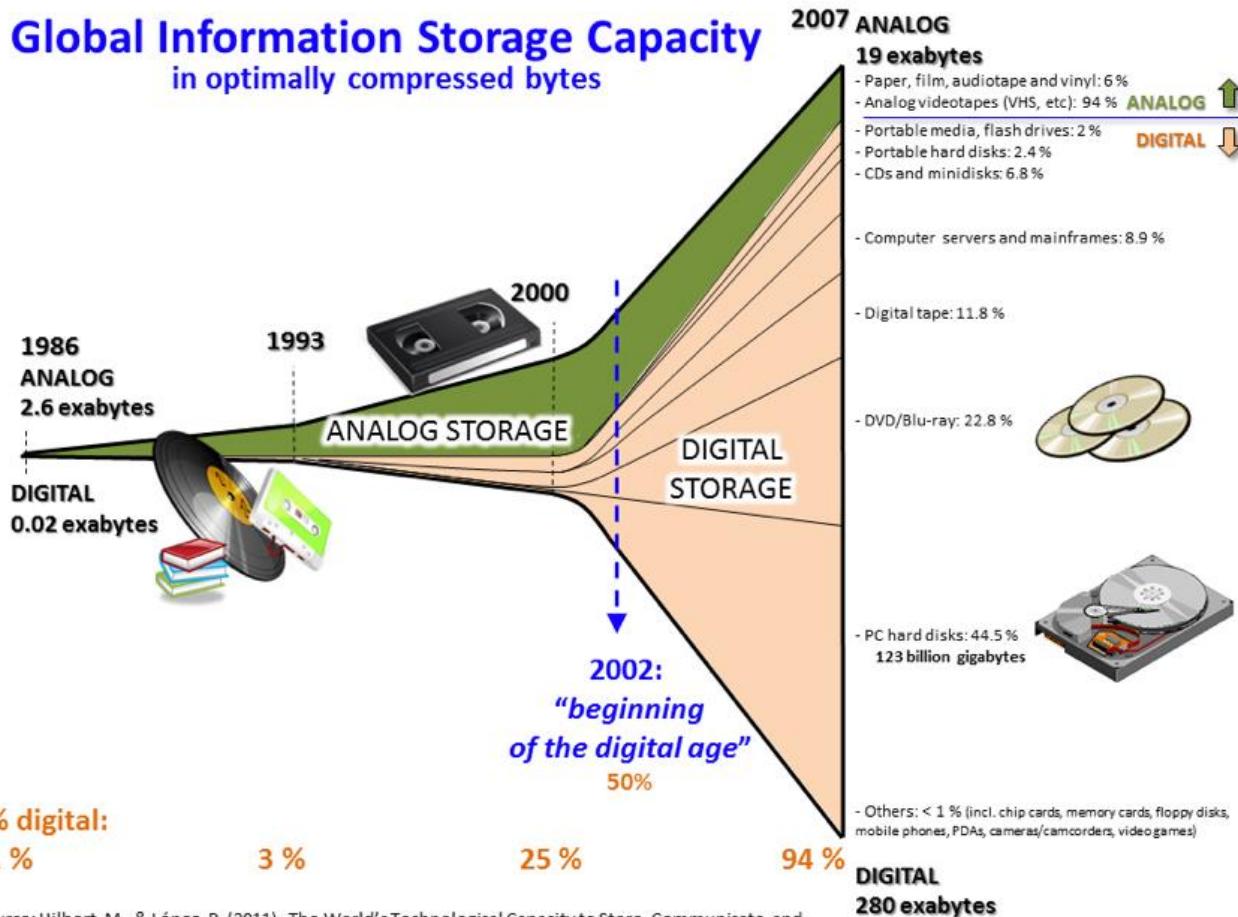
- 搭公車、搭捷運、刷悠遊卡
- 到便利商店買飲料
- 用手機上臉書、按個讚、打個卡
- 送個LINE、拍張照片上傳
- 上網瀏覽、刷卡購物、給個評價
- 過個馬路被超過十個鏡頭拍下

# 雲端大資料時代的來臨

- 隨著雲端時代的來臨，越來越多的服務需求集中於雲端伺服器當中（提供雲端服務的電腦系統稱之為雲端伺服器），因此，這些雲端伺服器將收集到來自所有客戶的資料，這是一筆比以往都要龐大的資料
- 故而雲端時代又代表著大資料時代的來臨。

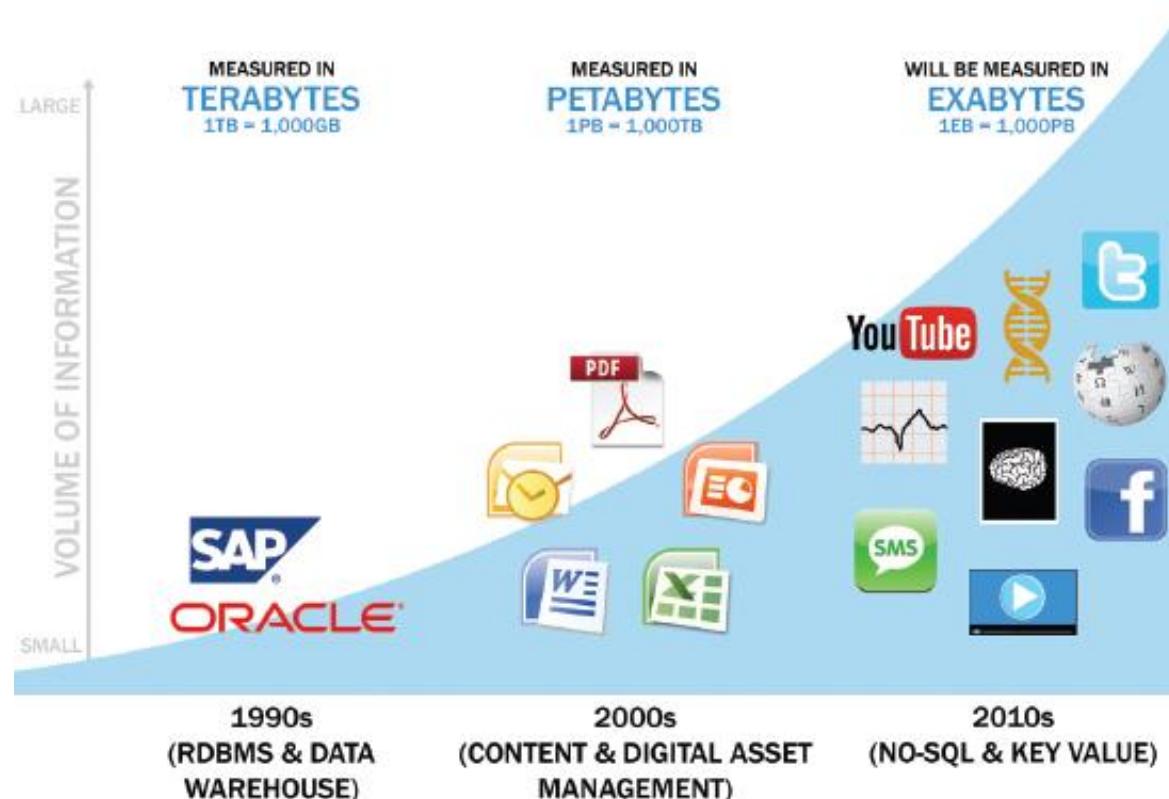
# Global Information Storage Capacity

in optimally compressed bytes



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. <http://www.martinhilbert.net/WorldInfoCapacity.html>

# Data Evolution and the Rise of Big Data Sources



# What is Big Data ?

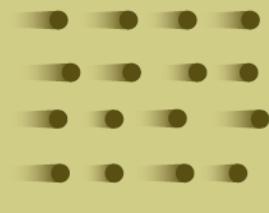
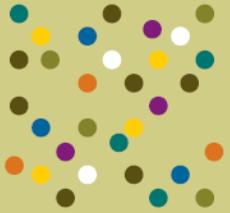
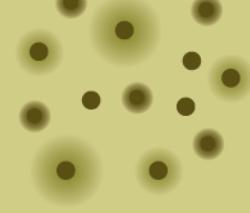
- All kinds of data
  - Large volumes
  - Valuable insight, but difficult to extract
  - May be extremely time sensitive
- Big Data is a Hot Topic Because Technology Makes it Possible to Analyze ALL Available Data



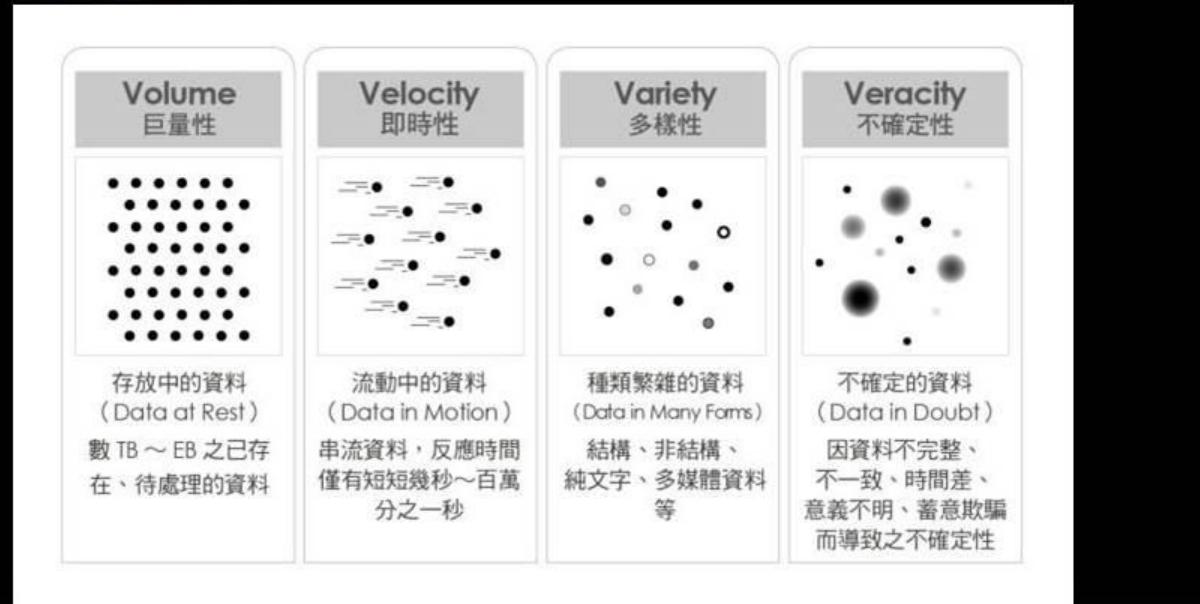
***"Big data*** technologies describe a new generation of technologies and architectures, designed to **economically extract value** from very large **volumes** of a wide **variety** of data, by enabling high **velocity** capture, discovery and/or analysis."

Source: Matt Eastwood, IDC

# Big Data 四大特性

| Volume  | Velocity  | Variety  | Veracity  |
|---|---|--|---|
|  |  |  |    |
| <b>Data at rest</b><br>Terabytes to exabytes<br>of existing data<br>to process    | <b>Data in motion</b><br>Streaming data,<br>milliseconds to<br>seconds to respond | <b>Data in many forms</b><br>Structured, unstructured,<br>text and multimedia      | <b>Data in doubt</b><br>Uncertainty due to data<br>inconsistency and<br>incompleteness,<br>ambiguities, latency,<br>deception and model<br>approximations |

Source : IBM Big Data Hub  
<http://www.ibmbigdatahub.com/>

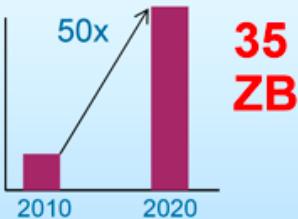


快速蔓延的海量資料，  
是已經開始，且永遠不會消失的挑戰

## The Characteristics of Big Data

- $V^4 = \text{Volume Velocity Variety Veracity}$

Cost efficiently processing the growing **Volume**



Responding to the increasing **Velocity**



**30 Billion**

RFID sensors and counting

Collectively analyzing the broadening **Variety**



**80%** of the world's data is unstructured



Establishing the **Veracity** of big data sources

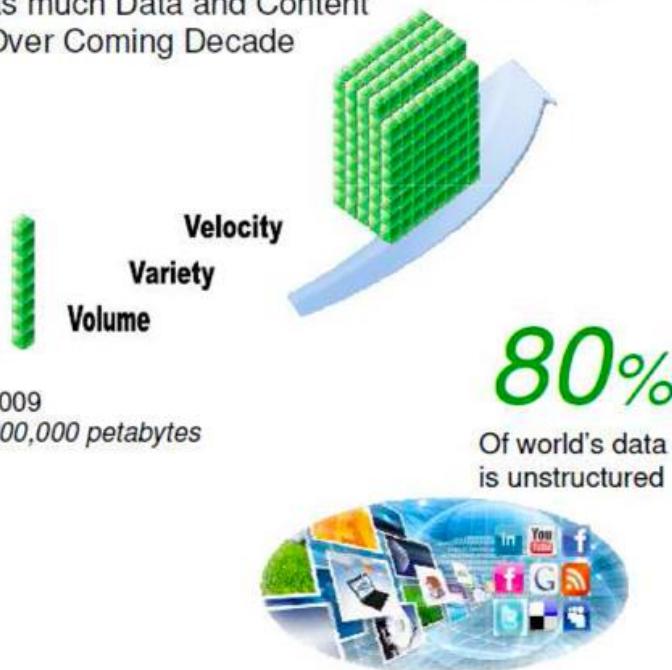
**1 in 3** business leaders don't trust the information they use to make decisions

## Information is at the Center of a New Wave of Opportunity...

**44x**

as much Data and Content  
Over Coming Decade

2020  
35 zettabytes



## ... And Organizations Need Deeper Insights

**1 in 3**

Business leaders frequently make decisions based on information they don't trust, or don't have

**1 in 2**

Business leaders say they don't have access to the information they need to do their jobs

**83%**

of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

**60%**

of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions

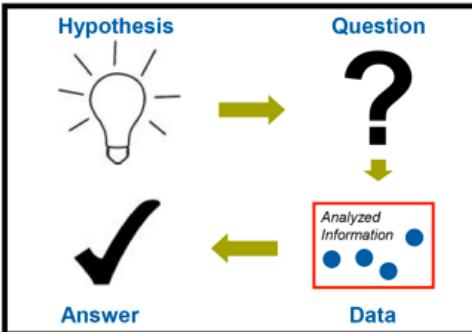
IBM CEO Survey

# Big data analysis

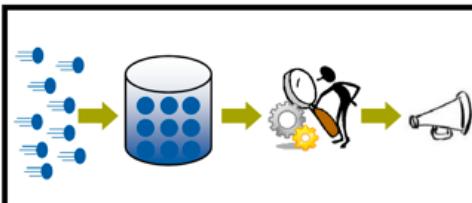
## The Big Data Approach to Analytics is Different

### Traditional Analytics

Structured & Repeatable  
Structure built to store data



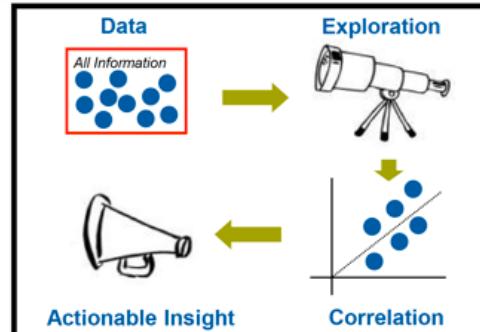
Start with hypothesis  
Test against selected data



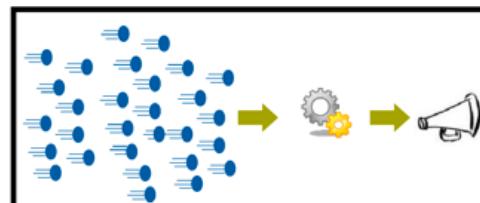
Analyze after landing...

### Big Data Analytics

Iterative & Exploratory  
Data is the structure



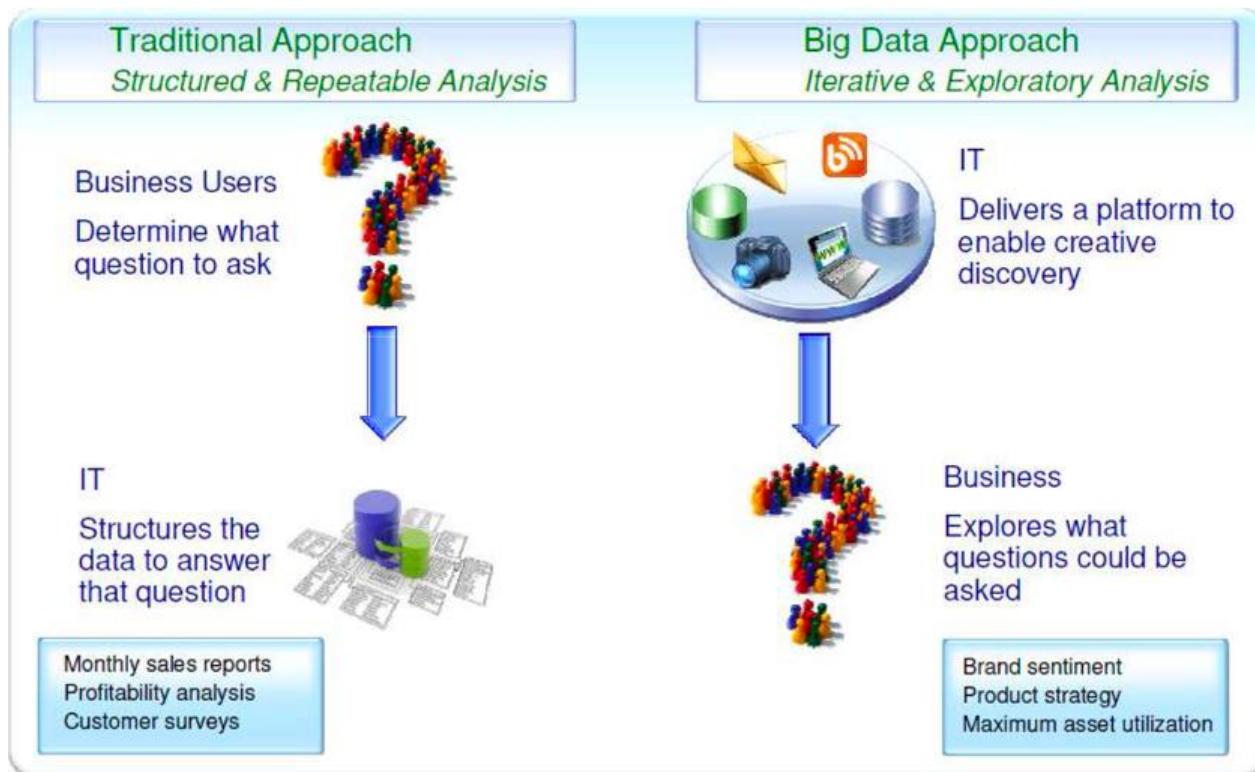
Data leads the way  
Explore all data, identify correlations



Analyze in motion...

© 2015 IBM Corporation

## Merging the Traditional and Big Data Approaches



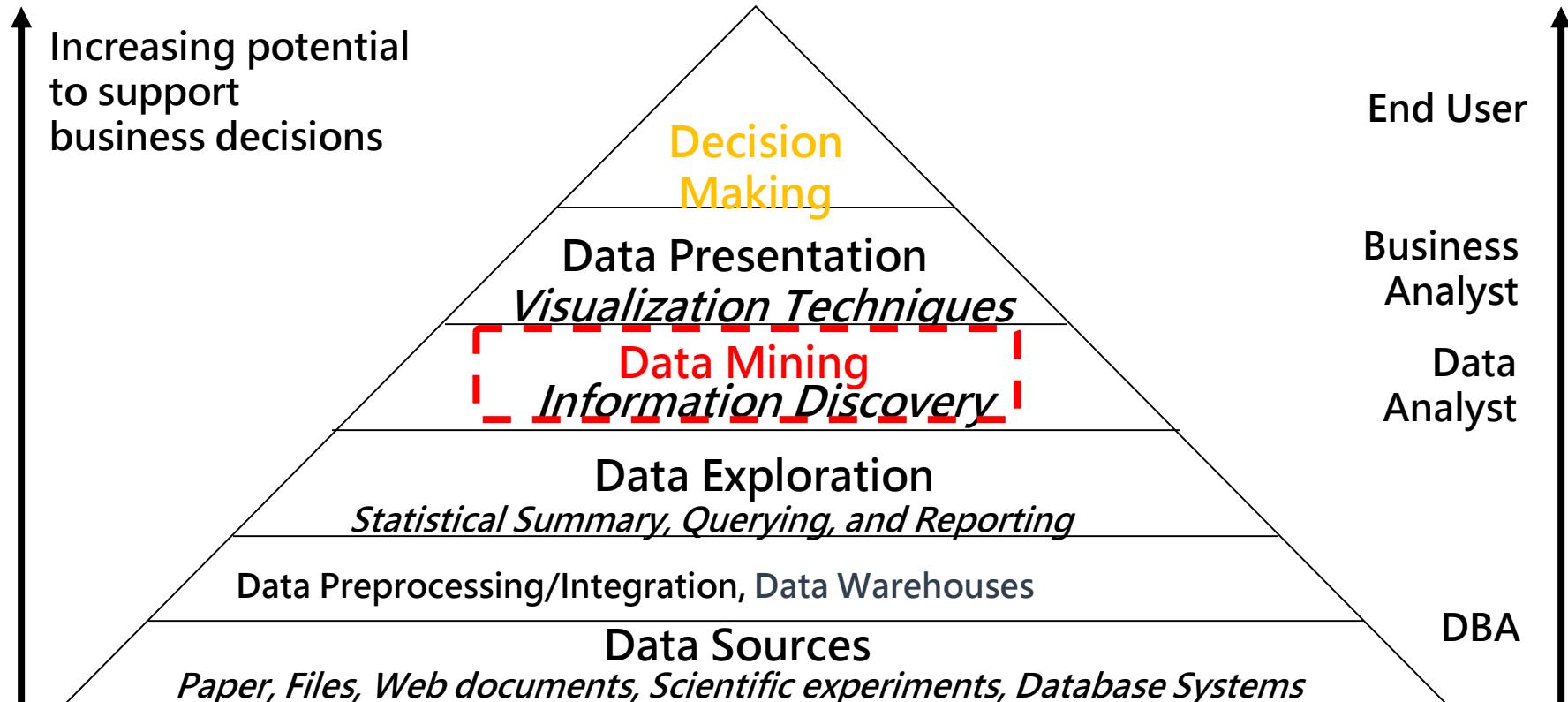
# 迎向雲端大資料的知識時代

- 電腦將資料(data) 處理後將獲得有用的資訊(Information)而這些資訊可以再被萃取出更有用的知識(Knowledge)

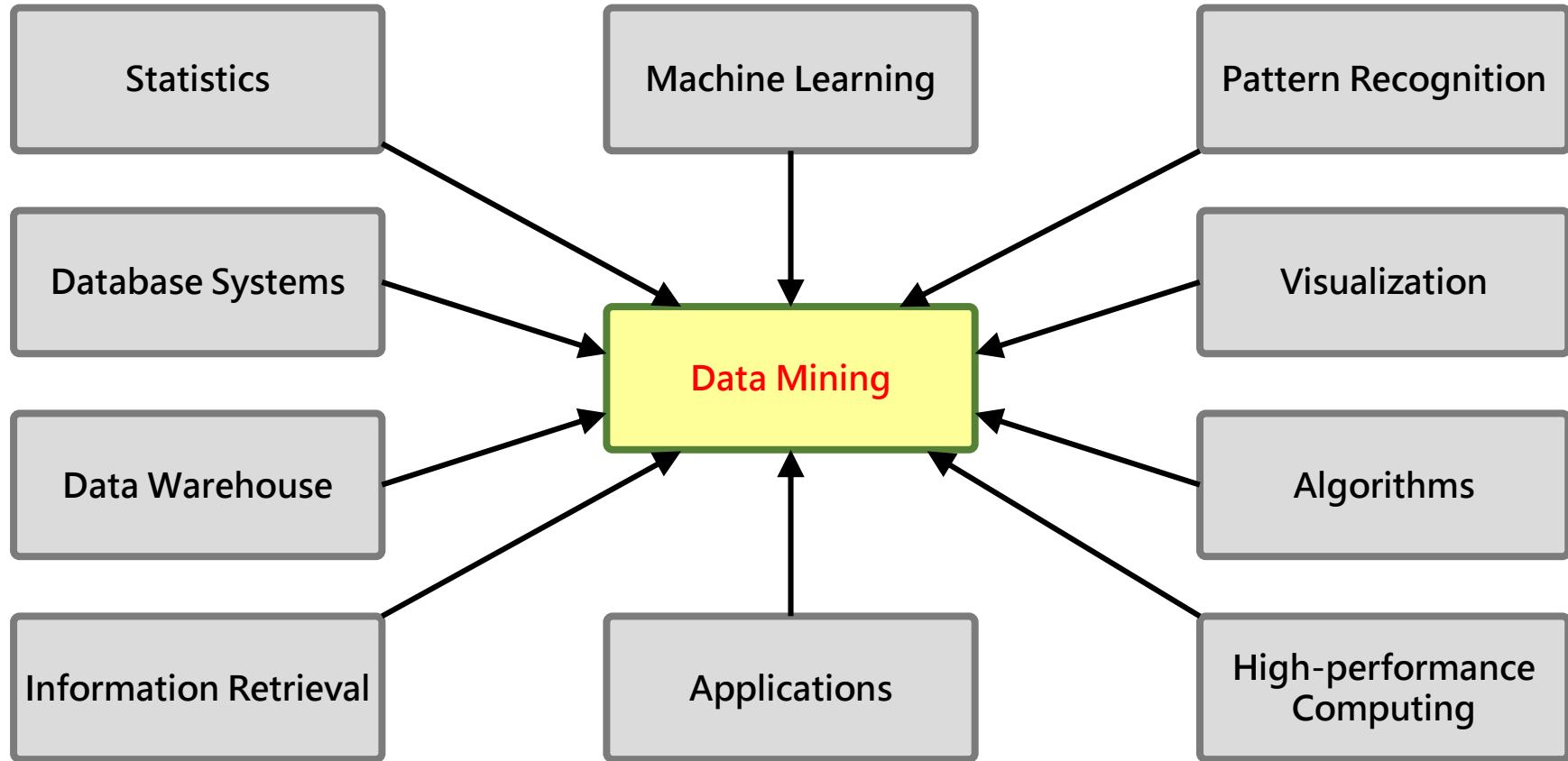
# 出版社在決定印刷一本書籍時

- 例子：出版社在決定印刷一本書籍時
  - 如果印量過大，會有非常多的庫存損失
  - 如果印量不夠需求而採分次印刷，則會提高成本，使利潤降低
- 若出版社能夠取得amazon蒐集的大量資料，那麼就有機會增加利潤
- 出版社可以將資訊以時間區分出去年的查詢次數與今年的查詢次數
- 再比較每本書在去年的銷售量與去年的查詢次數，推導出一個公式
- 當年查詢次數 $\times 0.6 =$ 當年銷售數值，今年的查詢次數  $\Rightarrow$  預測今年的銷售量
  - 因為平均每5次在amazon查詢該書（假設該書是五顆星）的使用者，會有1人向amazon購買該書
  - amazon在該出版社所佔業績為 $1/3$ ，也就是其他人在別的通路購買該書的可能性為amazon的2倍
  - 當年銷售數值 $= (\text{當年查詢次數}/5 \times 3) = \text{當年查詢次數} \times 0.6$

# Data Warehouse Data Mining and Business Intelligence



# Data Mining Technologies



# 企業資訊應用 - 資料探勘

- 人工智慧的一個分支

- 清除資料(Data Cleaning)：清除資料雜訊與矛盾的資料。
- 資料整合(Data Integration)：結合不同來源的資料。
- 資料選擇(Data Selection)：挑選有意義的屬性或資料。
- 資料轉換(Data Transformation)：將資料轉換成容易處理的格式。
- 資料探勘(Data Mining)：找出可能為知識的知識雛形(Pattern)。
- 知識雛形評估(Pattern Evaluation)：評估知識雛形(Pattern)的可用性。
- 知識表達(Knowledge presentation)：將可用的知識雛形轉化為知識表示法，並將之表達出來。

- 資料探勘所包含的資訊技術很多

- 分類(classification)、分群(clustering)、關聯規則(association Rule)等等。

# 關聯規則

# 關聯規則(Association Rule)

- 在眾多資料探勘的技術中，最容易理解的是關聯規則
- 舉例來說，一個大賣場的收銀櫃檯會把所有的銷售資料傳送到資料庫，經由整理成為資料倉儲
- 由於它會記錄每位顧客該次所購買的物品種類、數量、價格、特殊折扣、是否使用現金、使用哪一家的信用卡等等交易細節，這其中，光就物品種類一項，其實就隱含了一些知識。

# 關聯規則(Association Rule)

- 經由資料探勘的關聯規則法進行大量資料記錄的探勘後，可以發現
  - 購買吐司且同時購買牛奶者，佔了購買吐司的80%
  - 這個就是一種知識
    - (例如規則庫的 IF 買吐司 THEN 買牛奶的機率為80%)
  - 提醒業者，應該將吐司與牛奶的擺放位置盡量靠近或將吐司多放一份於牛奶附近，以節省消費者的採購時間及提高消費者的採購意願
- 而上述現象也會出現在火鍋料（如燕餃）與沙茶醬之間

**Table 8.1 A database of library transactions**

| Transaction ID | Books checked out   |
|----------------|---|
| 1              | <i>The Hobbit, The Princess Bride</i>   |
| 2              | <i>The Princess Bride, The Last Unicorn</i>   |
| 3              | <i>The Hobbit</i>   |
| 4              | <i>The Neverending Story</i>  |
| 5              | <i>The Last Unicorn</i>   |
| 6              | <i>The Hobbit, The Princess Bride, The Fellowship of the Ring</i>                     |
| 7              | <i>The Hobbit, The Fellowship of the Ring, The Two Towers, The Return of the King</i> |
| 8              | <i>The Fellowship of the Ring, The Two Towers, The Return of the King</i>             |
| 9              | <i>The Hobbit, The Princess Bride, The Last Unicorn</i>                               |
| 10             | <i>The Last Unicorn, The Neverending Story</i>  |

# Support

- T : 您的交易資料庫
- support(X)
  - 包含 X 的交易數量 / T 中的交易總數
  - # of transactions that contain X / total # of transactions in T

# confidence

- if  $X$ , then  $Y$ 
  - 每次您在交易中看到項目集  $X$  時，您也希望看到  $Y$
- $\text{conf}(X \Rightarrow Y)$ 
  - $\text{support}(\text{union}(X, Y)) / \text{support}(X)$

Table 8.1 A database of library transactions

| Transaction ID | Books checked out   |
|----------------|---|
| 1              | <i>The Hobbit, The Princess Bride</i>   |
| 2              | <i>The Princess Bride, The Last Unicorn</i>   |
| 3              | <i>The Hobbit</i>   |
| 4              | <i>The Neverending Story</i>  |
| 5              | <i>The Last Unicorn</i>   |
| 6              | <i>The Hobbit, The Princess Bride, The Fellowship of the Ring</i>                     |
| 7              | <i>The Hobbit, The Fellowship of the Ring, The Two Towers, The Return of the King</i> |
| 8              | <i>The Fellowship of the Ring, The Two Towers, The Return of the King</i>             |
| 9              | <i>The Hobbit, The Princess Bride, The Last Unicorn</i>                               |
| 10             | <i>The Last Unicorn, The Neverending Story</i>  |

support of the itemset { The Hobbit, The Princess Bride} =

3/10

confidence : "People who check out The Hobbit also check out The Princess Bride" =

3/5

confidence : "People who check out The Princess Bride also check out The Hobbit" =

3/4

# 關聯規則挖掘(association rule mining)的目標

- 至少給定的最小支持(minimum support) :  $\text{support}(X) > 10\%$
- 最小給定置信度(confidence) :  $\text{conf}(X \Rightarrow Y) > 60\%$

# 買吐司跟牛奶間的關聯規則？

|     |     | 買牛奶    |       | 總計    |
|-----|-----|--------|-------|-------|
|     |     | No     | Yes   |       |
| 買吐司 | No  | 500    | 3,500 | 4,000 |
|     | Yes | 1,000  | 5,000 | 6,000 |
|     |     | 10,000 |       |       |

已經知道買吐司的前提下，會買牛奶的機率 =  $5,000/6,000 = 83\%$

買牛奶的機率 =  $8,500/10,000 = 85\%$

Source: SAS Enterprise Miner Course Notes, 2014, SAS

# 關聯分析衡量的機率統計值— Lift增益值

- 「已知買吐司則買牛奶」這個規則的發生機率比單獨計算買牛奶的發生機率還低。

- **增益值(Lift)**

一條規則在預測結果時能比隨機發生的機會好多少。

- $\text{support}(\text{union}(X, Y)) / (\text{support}(X) * \text{support}(Y))$

- Lift (買吐司  $\cap$  買牛奶)  
 $= 0.83/0.85 < 1$

- $\text{Lift } (A \rightarrow B) = 2$

- $A \rightarrow B$  這條規則的增益值為 2

- 代表已知在買A的前提下又買B的機率，比直接買B 的機率提升 (增益)了2倍。

Source: SAS Enterprise Miner Course Notes, 2014, SAS

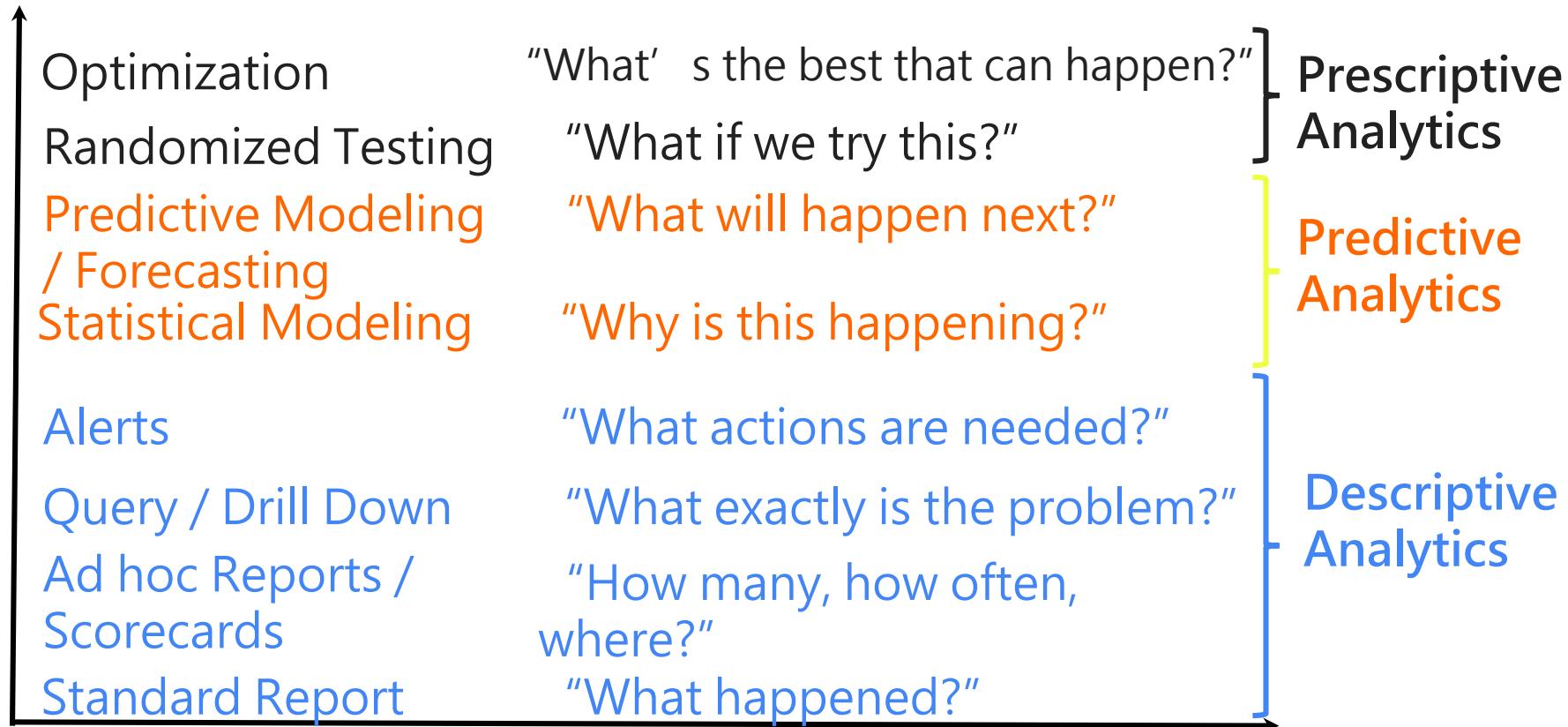
# 「買芭比娃娃就會買糖果」 你的行銷策略如何？

- 把兩項商品擺在一起
- 特意把兩項商品擺在相距較遠的地方
- 將糖果和芭比娃娃組合起來一起賣
- 糖果 + 芭比娃娃 + 銷售較差的商品一起組合銷售
- 定價策略：提供一個單價，降低另一個商品價格
- 廣告策略：芭比娃娃和糖果不需要同時廣告活動
- 產品設計：設計芭比娃娃形狀的糖果
- 提供芭比娃娃的配件，提升銷售

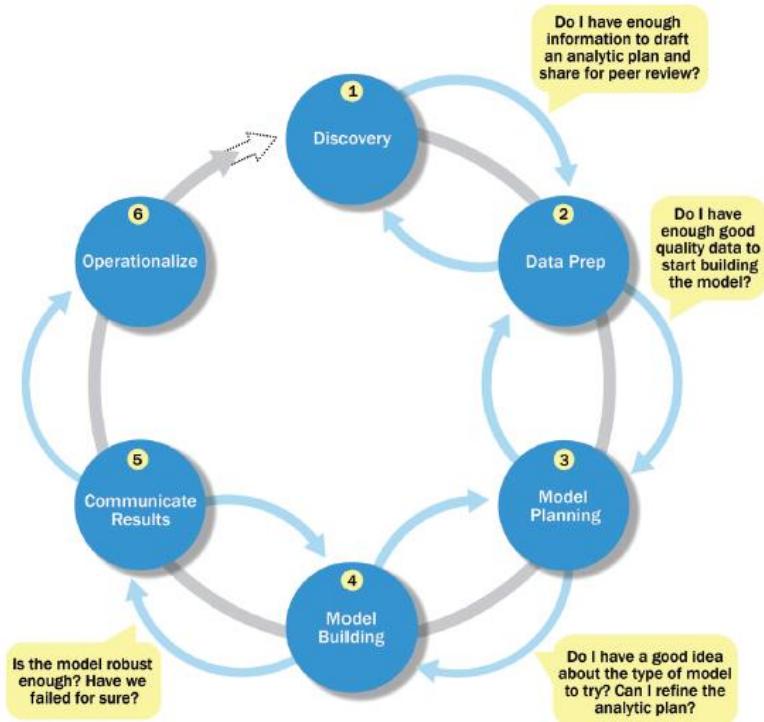
Source: SAS Enterprise Miner Course Notes, 2014, SAS

# Business Analytics

# Three Types of Business Analytics



# Overview of Data Analytics Lifecycle



# 場景：會員排序(Scenario : Member Sorting)

- 你任職於某大電信公司行銷部門
- 為了擬定促銷策略
  - 想了解目前800萬會員內，各付費方案人數比例
  - 想調出每月帳單金額超過5000元的名單
  - ...且濾出30歲以上之男生
  - 想促銷iPhone 14

## 我們以ING Direct為例，虛實整合，無縫接軌，在傳統直效行銷模式遭遇成長瓶頸時，透過多通路訊息之整合，有效地提升獲利、減少成本、縮短時程與深化顧客關係

### 公司背景

- 全歐洲最大的金融集團之一
- ING個零售銀行主要以荷蘭地區為主，貢獻集團獲利佔28%
- 每年發出的直效行銷郵件達6000萬份以上



### 面臨的挑戰

- Outbound 方式的行銷活動正面臨著成效逐漸低落的變化，再也無法從這樣的互動中尋求效益。
- 無法透過單一、跨通路的整合讓顧客在與與ING互動中感受到一致性的對待。
- 耗時、費工，有些活動設計需耗費16~22週，甚至動員10多位行銷人員配合方能完成。
- 與顧客的對話安排與設計無法集中，散落在不同的單位上。
- 行銷成本越來越高，效益卻一直在變小。

### 改善後的效果

- 不斷地創造讓顧客感受到即時、與自身相關的活動體驗，有效地提升了行銷的回應率
  - ✓ 每日透過5個通路(信件、email、網站、客服中心與分行)部署了8千5百萬份名單
  - ✓ 預期將增加2000萬歐元的獲利成長
- 直效行銷成本減少了 35%
- 有效縮減了相關流程(從22週減少為4週即可完成活動部署)

ING 每次顧客與通路進行互動的關鍵時刻，企業將能透過“行銷活動平台”之結果提供最適合顧客的 **產品服務方案 (Next Best Action)**，讓分行行銷戰力提升

**預先**將各項產品完成分群開發，每一位顧客在每個產品群中，至少會有一項**推薦商品**



Select

當顧客與**數位通路產生互動**時，系統僅會選取適合他的產品清單



Prioritise

每項產品的最終評分將依據**即時的演算法**與預定義之**行銷權重**整合分析後產生，並依據各產品最終評分結果進行順位排序



Filter

各通路僅會呈現提供給顧客最適合之方案清單



## DBS透過社群媒體,了解市場需求與服務滿意度

**金融業**



經常性的蒐集並使用大數據分析技術來自多個社群網站的討論內容，瞭解客戶對該行商品與服務的評價。當有新產品進行促銷活動之際，擴大觀察社群網站上的客戶反應，進行客戶體驗的追蹤，再根據客戶意見調整因應對策。能掌握每次促銷活動的客戶意見，可快速回應客戶聲音，讓DBS在促銷活動的成效上領先競爭者。



# “巨量資料”商業應用場景

## 網購業



**產品自動推薦系統**: 依照客戶在購物網頁的瀏覽軌跡歷史，來比對『產品』和『產品』之間的關連性，成功推薦出客戶所想要的產品。每三筆訂單中就有一筆是來自產品自動推薦系統，其跨售成績遙遙領先同業。

## 食品零售業



2009年時市占率突然下滑，委託IBM根據關鍵字調查超過150萬筆的論壇、部落格與網路新聞，發現大量關於鹹酵母醬的內容，及把酵母醬搭配酪梨、烤肉、番茄一起吃討論。於是進行行銷『你怎麼使用酵母醬？』**Vegemite 酵母醬成為熱門話題**，再度的激起人們購買意願，銷量量比高峰期再高出5%。

## 運輸業



透過每台貨車的無線電通訊和GPS定位，精確傳送車輛位置，並累積行車路徑，**找出最佳行車路線**。

UPS發現十字路口左轉最易發生意外，紅綠燈最浪費時間，只要減少通過十字路口次數，就能省油並提高安全性。因行車順暢，時間少了近一成，這等於一年省下三百萬加侖的油料及減少三萬噸二氧化碳，安全性和效率也提高了。

## 金融業



發現多數的AE信用卡持卡人都有Facebook帳號，因而通設計了一系列行銷活動，讓持卡人願意將其信用卡綁定其Facebook帳號，以便美國運通獲得用戶授權，可以**追蹤客戶在臉書上的喜好**。當發現持卡人在臉書上感興趣的商家還不是美國運通的合作夥伴時，美國運通馬上派出專員去與之建立行銷合作，然後將促銷折扣資訊推送到持卡人的臉書帳號上。

## 金融業



餘額寶是一款貨幣基金T+0產品，裡頭的資金等同活期存款，用戶隨時可取用，遇過年或節日用錢高峰，若沒足夠現金，很容易遭到擠兌，而出現流動性風險。但若預留過多的現金，則收益會被拉低。運用大數據可**對每個時點的支付寶動用趨勢做出精準的預測**，讓資金管理的難度降低。即使遇到雙十一光棍節與春節，用戶資金調動最頻繁的時段，餘額寶也能讓流動性需求預測偏離率不超過5%。



## 「使用者付費」的保險新領域

Allstate(好事達保險公司) · 美國第二大個人保險公司 · 保險範圍含蓋：汽車、房  
屋、人壽、健康....等

多樣化線上  
投保種類

線上24hr  
專屬客服

Your Allstate Agent  
Mohammad Y  
Bajwa  
1720 Columbia Rd NW  
Washington, DC 20009  
P: (202) 207-7932  
(change agent)



使用者付費車險計畫，  
人可以透過 GM 集  
OnStar 行車資訊技術服  
Ford 車廠的 Sync 車用影  
統向保險公司傳輸其駕  
駕關數據。

2013年 Allstate 表示透  
進科技幫助汽車保險定  
機制將在10年內佔據70%

## 穿戴裝置收集健康數據、使用者付費訂價模式

Medibank 澳洲第一大醫療保險公司(聯邦政府所有)

The screenshot shows the Medibank website homepage. At the top, there's a navigation bar with links for About Medibank, Careers, Privacy, HEALTH INSURANCE, OTHER INSURANCE, BETTER HEALTH, and MEMBERS. On the right side of the header, there's a phone icon followed by the number 134 190. Below the header, a large red banner has the text "JOIN AND RECEIVE A FREE FITBIT FLEX™". To the right of this banner is a call-to-action button with three options: "GET A QUOTE", "FIND A STORE", and "CONTACT US". In the bottom left corner of the red banner, there's a small red arrow pointing right with the text "Rewards for getting active". In the bottom right corner of the red banner, there's another smaller text block: "SPECIAL OFFER - Offer extended. Take out hospital and extras cover by Oct 20 and receive a Free Fitbit Flex™ Wireless Activity Tracker. Find out more >".

保險公司與可穿戴廠商合作，以補貼價為保險使用者提供裝置，使用者佩戴裝置收集的健康和運動資料傳輸回保險公司，保險公司據此組態保費上的獎勵或懲罰標準



# Data Scientist: *The Sexiest Job of the 21st Century*

**Meet the people who  
can coax treasure out of  
messy, unstructured data.**

by Thomas H. Davenport  
and D.J. Patil

# W

hen Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

# 提升資料素養

# 提升資料素養的六種方法

- 實體課程 (長期/短期) • 參加黑客松
- 線上課程 / 書籍 • 線上資料競賽
- 參與資料社群 • 資料英雄計畫 (長期專案)

在學校就好好的學理論與技術

# 實體課程 (長期)

- 統計學
- 機器學習
- 資料結構
- 資料探勘
- 平行運算程式設計
- 迴歸分析
- 時間序列分析
- 資料視覺化
- R語言、Python、SQL

短期課程就衝實戰經驗

# 實體課程 (短期)

DSP 資料爬析 Python 實戰班

DSP Data ETL and Analysis with Python

時間：2015/8/1~8/2~8/8~8/9



2016 DSP 資料科學夏令營@NCCU

DSP Data Camp @NCCU Summer 2016

Student Edition



Credit by 謝宗震 智庫驅動

自學良伴

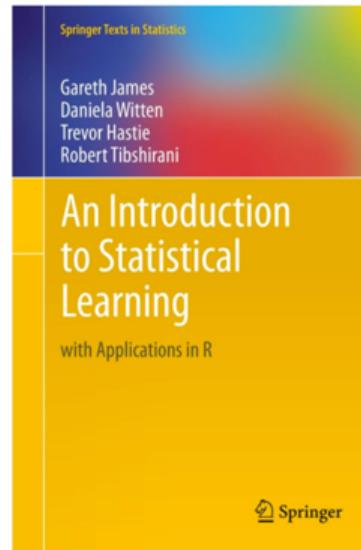
## 推薦書籍 (科普書)

- 聰明學統計的13又½堂課  
(<http://www.books.com.tw/products/0010617019>)
- 統計學，最強的商業武器  
([http://www.books.com.tw/products/0010710939?loc=P\\_asb\\_004](http://www.books.com.tw/products/0010710939?loc=P_asb_004))
- 統計學，最強的商業武器：實踐篇  
([http://www.books.com.tw/products/0010687439?loc=P\\_asb\\_001](http://www.books.com.tw/products/0010687439?loc=P_asb_001))

自學良伴

# 推薦書籍

- 60本免費的資料科學書籍  
(<http://dataology.blogspot.tw/2015/09/60.html>)
- An Introduction to Statistical Learning with Applications in R (2013)



BONUS MATERIAL: Includes multimode CD with over 50 figures and illustrations.

# HOW TO MEASURE ANYTHING

Finding the Value of "Intangibles" in Business



2<sup>nd</sup> Edition  
REVISED, EXPANDED,  
& SIMPLIFIED



DOUGLAS W. HUBBARD

READ BY DAVID DRUMMOND

**How to Measure Anything:  
Finding the Value of Intangibles  
in Business** 3rd Edition  
by Douglas W. Hubbard

# 雲端時代 的殺手級應用



天下雜誌出版  
作者/胡世忠 定價/360元

# 雲端時代的殺手級應用 BIG DATA海量資料分析

美、英、日、印、澳等標竿企業與政府應用案例  
首度揭露兩岸三地海量資料分析領先者的策略

海量資料分析

自學良伴

# 線上課程

- 慕課 (<http://course.cool3c.com>)
- R語言翻轉教室 (<http://datascienceandr.org>)
- kaggle Tutorial (<https://www.kaggle.com/wiki/Tutorials>)
- **Launch Your Career in Data Science**  
(<https://www.coursera.org/specializations/jhu-data-science>)

# 課程進度

| 單元 | 時間         | 標題           |
|----|------------|--------------|
| 1  | 09/26 第三週  | 程式桌遊(海霸)     |
| 2  | 10/3 第四週   | micro:bit 1  |
| 3  | 10/17 第六週  | micro:bit 2  |
| 4  | 10/31 第八週  | RapidMiner 1 |
| 5  | 11/14 第十週  | RapidMiner 2 |
| 6  | 11/28 第十二週 | RapidMiner 3 |
| 7  | 12/12 第十四週 | 期末報告討論       |

| 週次<br>課程主題         | 課程內容與指定閱讀  |                       |
|--------------------|--|-----------------------|
| 1 計算思維簡介           | 社會情境脈絡與未來發展<br>書目：1, 2, 3  |                       |
| 2 計算思維             | 基本內涵與核心概念<br>書目：1, 2, 3  |                       |
| 3 功能模組             | 問題拆解與型態辨認<br>書目：4, 5, 6  |                       |
| 4 功能模組             | 抽象思考與演算邏輯<br>書目：4, 5, 6  |                       |
| 5 國慶日              | 國定假日   |                       |
| 6 類比至數位轉換 & 電腦運算架構 | 類比與數位訊號的基礎概念及類比轉換至數位訊號的原理 &<br>書目：7 chapter 1 & 4 & 5  | 電腦組成元件與其運算架構          |
| 7 大數據應用            | 大數據中資料科學的基礎分析概念與商業相關應用   |                       |
| 8 學習成果測試           | 期中評量/作業活動  |                       |
| 9 運算思維測驗           | 國際運算思維挑戰賽  |                       |
| 10 人工智慧發展          | 人工智慧發展歷程與未來趨勢<br> | 人工智慧各式技術與應用案例<br>書目：8 |
| 11 人工智慧技術與應用       |  |                       |
| 12 人工智慧應用場景        | 人工智慧跨域應用   |                       |
| 13 人工智慧學習模型實作      | Nocode AI 練習 – Rapidminer<br>書目：9  |                       |
| 14 人工智慧倫理          | 生成AI (如：ChatGPT、Deepfake、Midjourney)、假新聞及未來人工智慧應用上的倫理問題  |                       |
| 15 人工智慧專題          | 海報展示   |                       |
| 16 計算思維與人工智慧       | 期末報告   |                       |
| 17 彈性補充教學          | 人工智慧相關競賽經驗交流   |                       |
| 18 彈性補充教學          | 校園人工智慧應用發想   |                       |



### 💡 最新消息

NO.1

#### 112學年度第一學期挑戰賽 重要日程

⌚ 2023/07/31

報名期間：10月2日至10月27日止（同時開放新加入教師註冊；教師登入後，點選「管理班級」即可為班級學生報名）

練習時間：11月6日至11月10日止（已報名學生可登入練習）

挑戰期間：11月13日至11月24日止（教師可於報名時，選擇挑戰期間內的一節課實施）

成績與參與證明公佈時間：12月22日

關於帳號、報名等問題，請詳見 [FAQ](#) ( 常見問題 )

\* 112學年度第一、二學期使用之題目相同。

\* 系統分流作業將於10月30日至11月3日進行，屆時網站將暫停服務；如造成不便敬請見諒。

年級：

五、六年級 七、八年級 九、十年級 十一、十二年級 大專院校

選擇挑戰賽：

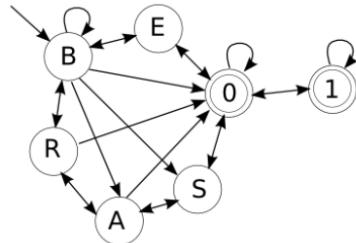
112-1國際運算思維挑戰賽大專院校組

<https://bebras.csie.ntnu.edu.tw/>

## 木筏牌照

海狸的交通工具是木筏，每艘木筏都有一個獨一無二的牌照號碼。

合法的牌照號碼必須依據下圖規則，按照邊的方向(有單向 → 跟雙向 ←→)走訪節點；依照走訪的順序，把節點上的英文字母和數字組成牌照號碼。  
而且每個牌照號碼都由字母B開始，並以數字0或1結束。

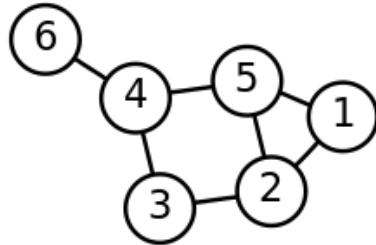


請問下列哪一個不是合法的木筏牌照號碼？

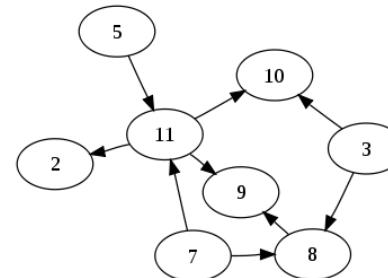
- BB0001 和 BBB011
- BBB100 和 BR00A0
- BB0100 和 BSA001
- BE0S01 和 BE01

# 什麼是圖(Graph)？

Ans: 圖是點和邊所構成的集合，分為無向圖和有向圖



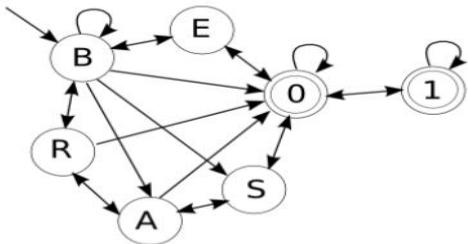
無向圖



有向圖



## 木筏牌照



請問下列哪一個不是合法的木筏牌照號碼？

- BBB001 和 BBB011
- BBB100 和 BR00AO
- BB0100 和 BSA001
- BE0S01 和 BE01

說明:

BBB100 · B無法直接到1  
BR00AO · O無法直接到A

關鍵字: 有向圖、有限狀態機

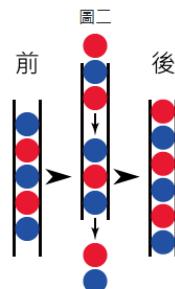
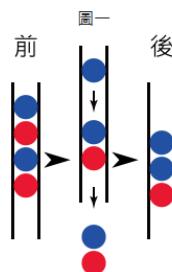
## 紅藍彈珠遊戲

海狸弟弟在海狸一日營學到了一個新遊戲。

首先他抓起一把紅色及藍色的彈珠——放進一個透明的塑膠管中，每顆彈珠都從塑膠管的上頭放入。

接下來每次從塑膠管下頭取出兩顆彈珠，若第一顆彈珠是紅色，則再從上頭放入一顆藍色的彈珠（如圖一所示）；若第一顆取出的彈珠是藍色，則從上頭依序放入紅色、藍色、紅色三顆彈珠（如圖二所示）。

此「取出 + 加入」彈珠程序持續到塑膠管內只剩下兩顆或一顆彈珠為止。

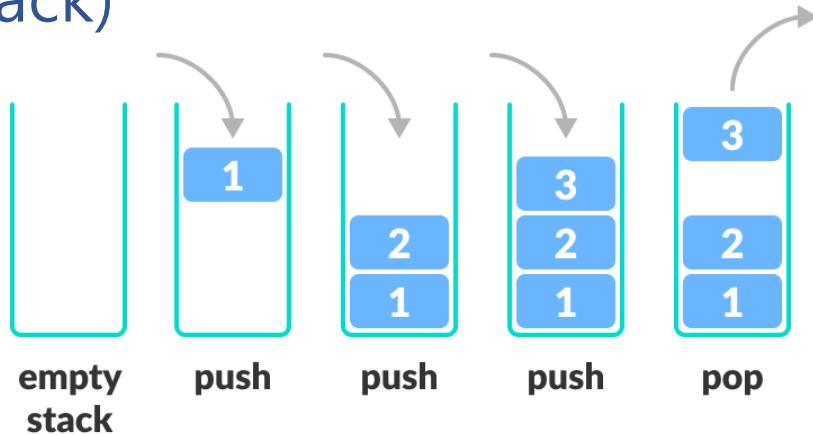


若海狸弟弟一開始依序放入了紅、藍、藍、紅、紅等五顆彈珠如下圖所示，總共需要幾次的「取出 + 加入」程序，才會使得塑膠管內的彈珠剩下一或二顆彈珠？

僅填入數字，只能輸入半形數字才會正確



## 堆疊(Stack)



先進後出(First In Last Out)

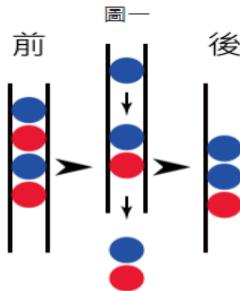
<https://www.programiz.com/dsa/stack>

# 佇列(Queue)



先進先出(First In First Out)

<https://www.programiz.com/dsa/queue>

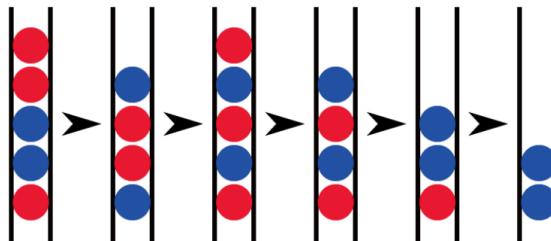
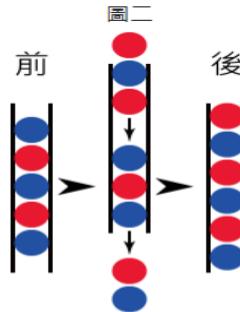


若海狸弟一開始依序放入了紅、藍、藍、紅、紅等五顆彈珠如下圖所示，總共需要幾次的「取出 + 加入」程序，才會使得塑膠管內的彈珠剩下一或二顆彈珠？

僅填入數字，只能輸入半形數字才會正確



If 取出 “藍紅”，放入 “藍”，  
else if 取出 “紅藍”，放入 “紅藍紅”



關鍵字: stack and queue, 運算模型

堆疊 - 維基百科，自由的百科全書 ([wikipedia.org](https://zh.wikipedia.org))  
併列 - 維基百科，自由的百科全書 ([wikipedia.org](https://zh.wikipedia.org))

## 聊天順序

娜娜邀請了五位好朋友參加她的生日派對，五位朋友分別是：小愛、奇哥、小蘿、戴哥、與小萱。生日派對上娜娜出了一個謎題如下：

- 我任何時候都可跟小萱聊天。
- 我必須先與小愛聊天後才能與戴哥聊天。
- 我必須先與小萱聊天後才能與奇哥聊天。
- 我必須先與戴哥及奇哥聊天後才能與小蘿聊天。
- 我必須先與奇哥及小萱聊天後才能與小愛聊天。

娜娜與五位朋友的聊天順序應為何？

( 請依聊天順序連續填入名字，不要留空白不要加入標點符號 )

請在此作答...

## 聊天順序

- 這個問題就像是修課順序一樣，課與課可能有擋修的問題，有些課必須先修，才能修另一堂。
- 這就是圖論中的 “拓撲排序” (topological sort)

關鍵字: 有向圖、拓撲排序、有向無環圖(DAG)

[拓撲排序 - 維基百科，自由的百科全書 \(wikipedia.org\)](#)

## 什麼是拓撲排序？

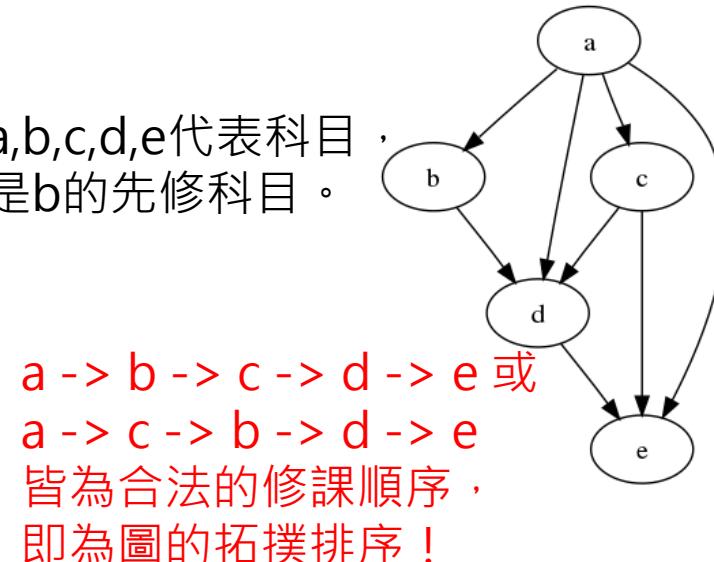
- 在某校的選課系統中，存在這樣的規則：每門課可能有若干門先修課，如果要修讀某一門課，則必須要先修讀此課程所要求的先修課後才能修讀。假設一個學生同時只能修讀一門課程，那麼，被選課系統允許的他修完他需要所有課程的順序是一個拓撲排序

[拓撲排序 - 維基百科，自由的百科全書 \(wikipedia.org\)](#)

## 什麼是拓撲排序？

舉例來說，右圖是修課的流程圖，a,b,c,d,e代表科目，每一種都必須修習，a指向b代表a是b的先修科目。

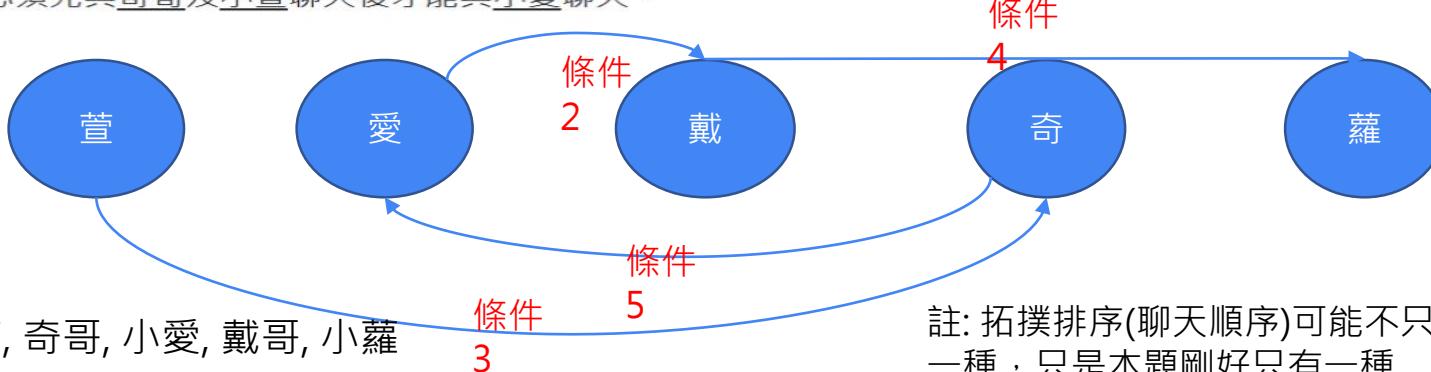
| 科目 | 先修科目  |
|----|-------|
| a  | 無     |
| b  | a     |
| c  | a     |
| d  | a b c |
| e  | a c d |



拓撲排序 - 維基百科，自由的百科全書 (wikipedia.org)

## 聊天順序

- (條件1) • 我任何時候都可跟小萱聊天。
- (條件2) • 我必須先與小愛聊天後才能與戴哥聊天。
- (條件3) • 我必須先與小萱聊天後才能與奇哥聊天。
- (條件4) • 我必須先與戴哥及奇哥聊天後才能與小蘿聊天。
- (條件5) • 我必須先與奇哥及小萱聊天後才能與小愛聊天。



## 算式化簡

小海狸有一台很特別的電腦，它提供兩種指令來計算海狸世界的數學式子，這兩種指令的使用說明如下：

- R 指令 當  $f$  是一個數學運算子，可以是  $+$ 、 $-$ 、 $\times$  或  $\div$ ；而  $X_1, X_2, \dots, X_n$  代表一串數字，那麼：  
 $(R f (X_1, X_2, \dots, X_n))$  指令將會計算  $X_1 f X_2 f \dots f X_n$

舉例來說： $(R + (1, 2, 3, 4))$  將會計算  $1 + 2 + 3 + 4$ ，而結果為 10。

- M 指令 當  $f$  是一個函數，而  $X_1, X_2, \dots, X_n$  代表一串數字，那麼：  
 $(M f (X_1, X_2, \dots, X_n))$  指令將會計算  $f(X_1), f(X_2), \dots$  及  $f(X_n)$ ，並得到計算結果所組成的數列。

舉例來說：當函數  $q(x) = -x$ ，那麼  $(M q (1, 2, 3, 4))$  將會把數列  $(1, 2, 3, 4)$  帶入  $q(x)$ ，得到結果為  $(-1, -2, -3, -4)$ 。

現在假設  $t(X) = 3X + 2$  且  $q(X) = -X$ 。請問下列式子會得到什麼結果？

$$(R + ((R + (M t (0, 2, 4))), (R + (M q (M t (3, 5))))))$$

- 7
- 0
- 7
- 4



|        |   |
|--------|---|
| ● R 指令 | 當 $f$ 是一個數學運算子，可以是 $+$ 、 $-$ 、 $\times$ 或 $\div$ ；而 $X_1, X_2, \dots, X_n$ 代表一串數字，那麼：<br>$(R f (X_1, X_2, \dots, X_n))$ 指令將會計算 $X_1 f X_2 f \dots f X_n$<br>舉例來說： $(R + (1, 2, 3, 4))$ 將會計算 $1 + 2 + 3 + 4$ ，而結果為 10。                                   |
| ● M 指令 | 當 $f$ 是一個函數，而 $X_1, X_2, \dots, X_n$ 代表一串數字，那麼：<br>$(M f (X_1, X_2, \dots, X_n))$ 指令將會計算 $f(X_1), f(X_2), \dots$ 及 $f(X_n)$ ，並得到計算結果所組成的數列。<br>舉例來說：當函數 $q(x) = -x$ ，那麼 $(M q (1, 2, 3, 4))$ 將會把數列 $(1, 2, 3, 4)$ 帶入 $q(x)$ ，得到結果為 $(-1, -2, -3, -4)$ 。 |

- 所以，R指令就是Reduce運算, M指令就是Map運算,
- Reduce運算是個純量函數(輸出是個純量)
- Map運算是向量函數(輸出是個向量)
- 舉例來說:
- (1)  $R + (a, b, c, d) = (((a + b) + c) + d)$
- (2)  $M (2x+3) (a, b, c) = (2a+3, 2b+3, 2c+3)$

關鍵字: Map&Reduce, 算式化簡,  
MapReduce - Wikipedia



現在假設  $t(X) = 3X + 2$  且  $q(X) = -X$ . 請問下列式子會得到什麼結果？

$$(R + ((R + (\underline{M t (0, 2, 4)})), (R + (\underline{M q (M t (3, 5))}))))$$

- 7
- 0
- 7
- 4

根據以上的說明，我們可以來化簡式子：

$$\underline{M t (0, 2, 4)} = M (3x+2) (0, 2, 4) = (2, 8, 14)$$

$$\underline{R + (2, 8, 14)} = 24$$

$$\underline{M t (3, 5)} = M (3x+2) (11, 17)$$

$$\underline{M q (11, 17)} = M (-x) (11, 17) = (-11, -17)$$

$$\underline{R + (-11, -17)} = -28$$

$$R + (24, -28) = -4$$

# 政治大學新創之星競賽



競賽資訊頁面：

[新創之星競賽頁面](#)

# 第八屆台積電青年築夢計畫



競賽資訊頁面：  
[2023第八屆台積電青年築夢計畫](#)

# 2023 PRIDE數據資料分析競賽



競賽資訊頁面：

[PRIDE資料庫全國大專院校數據資料分析競賽](#)

# 2023永續智慧創新黑客松



# 2023永續智慧創新 黑客松競賽

競賽資訊頁面：

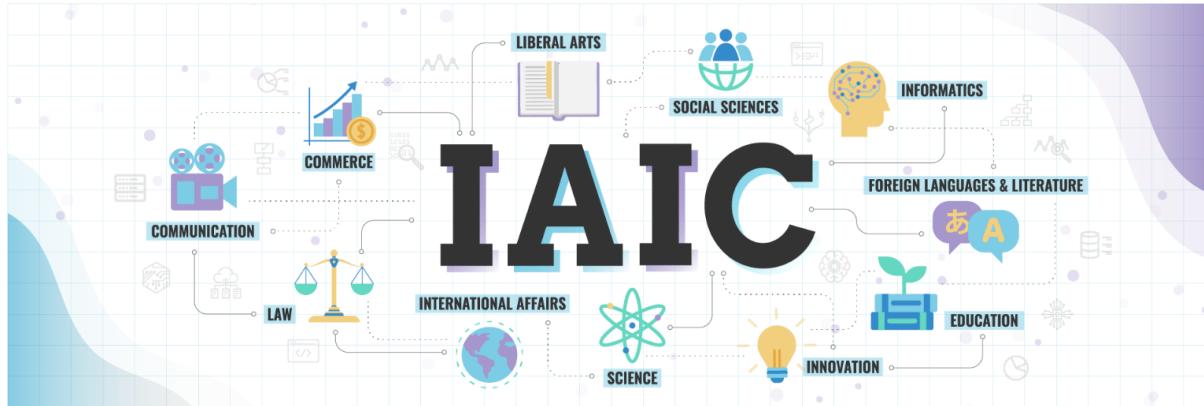
[2023永續智慧創新黑客松官網](#)

[報名連結](#)

# 課程進度

| 單元 | 時間         | 標題           |
|----|------------|--------------|
| 1  | 09/26 第三週  | 程式桌遊(海霸)     |
| 2  | 10/3 第四週   | micro:bit 1  |
| 3  | 10/17 第六週  | micro:bit 2  |
| 4  | 10/31 第八週  | RapidMiner 1 |
| 5  | 11/14 第十週  | RapidMiner 2 |
| 6  | 11/28 第十二週 | RapidMiner 3 |
| 7  | 12/12 第十四週 | 期末報告討論       |

| 週次<br>課程主題         | 課程內容與指定閱讀   |                        |
|--------------------|---|------------------------|
| 1 計算思維簡介           | 社會情境脈絡與未來發展   |                        |
|                    | 書目：1, 2, 3  |                        |
| 2 計算思維             | 基本內涵與核心概念   |                        |
|                    | 書目：1, 2, 3  |                        |
| 3 功能模組             | 問題拆解與型態辨認   |                        |
|                    | 書目：4, 5, 6  |                        |
| 4 功能模組             | 抽象思考與演算邏輯   |                        |
|                    | 書目：4, 5, 6  |                        |
| 5 國慶日              | 國定假日  |                        |
| 6 類比至數位轉換 & 電腦運算架構 | 類比與數位訊號的基礎概念及類比轉換至數位訊號的原理 & 電腦組成元件與其運算架構                | 書目：7 chapter 1 & 4 & 5 |
| 7 大數據應用            | 大數據中資料科學的基礎分析概念與商業相關應用                                  |                        |
| 8 學習成果測試           | 期中評量/作業活動   |                        |
| 9 運算思維測驗           | 國際運算思維挑戰賽   |                        |
| 10 人工智慧發展          | 人工智慧發展歷程與未來趨勢   |                        |
|                    | 人工智慧各式技術與應用案例   |                        |
|                    | 書目：8  |                        |
| 11 人工智慧技術與應用       |   |                        |
| 12 人工智慧應用場景        | 人工智慧跨域應用  |                        |
| 13 人工智慧學習模型實作      | Nocode AI 練習 – Rapidminer                               | 書目：9                   |
| 14 人工智慧倫理          | 生成AI (如：ChatGPT、Deepfake、Midjourney)、假新聞及未來人工智慧應用上的倫理問題 |                        |
| 15 人工智慧專題          | 海報展示  |                        |
| 16 計算思維與人工智慧       | 期末報告  |                        |
| 17 彈性補充教學          | 人工智慧相關競賽經驗交流  |                        |
| 18 彈性補充教學          | 校園人工智慧應用發想  |                        |



研究合作 跨域教學 多元服務



Dr. Chih-Hsun Wu

吳致勳 助理教授

20031214@nccu.edu.tw

j20031214@gmail.com