

# 大數據應用簡介

樹德科技大學資管系  
蕭銘雄/董信煌

# 本週課程

- 大數據簡介
- 大數據分析
- 大數據應用
- 結論
- 參考資料

# 大數據簡介

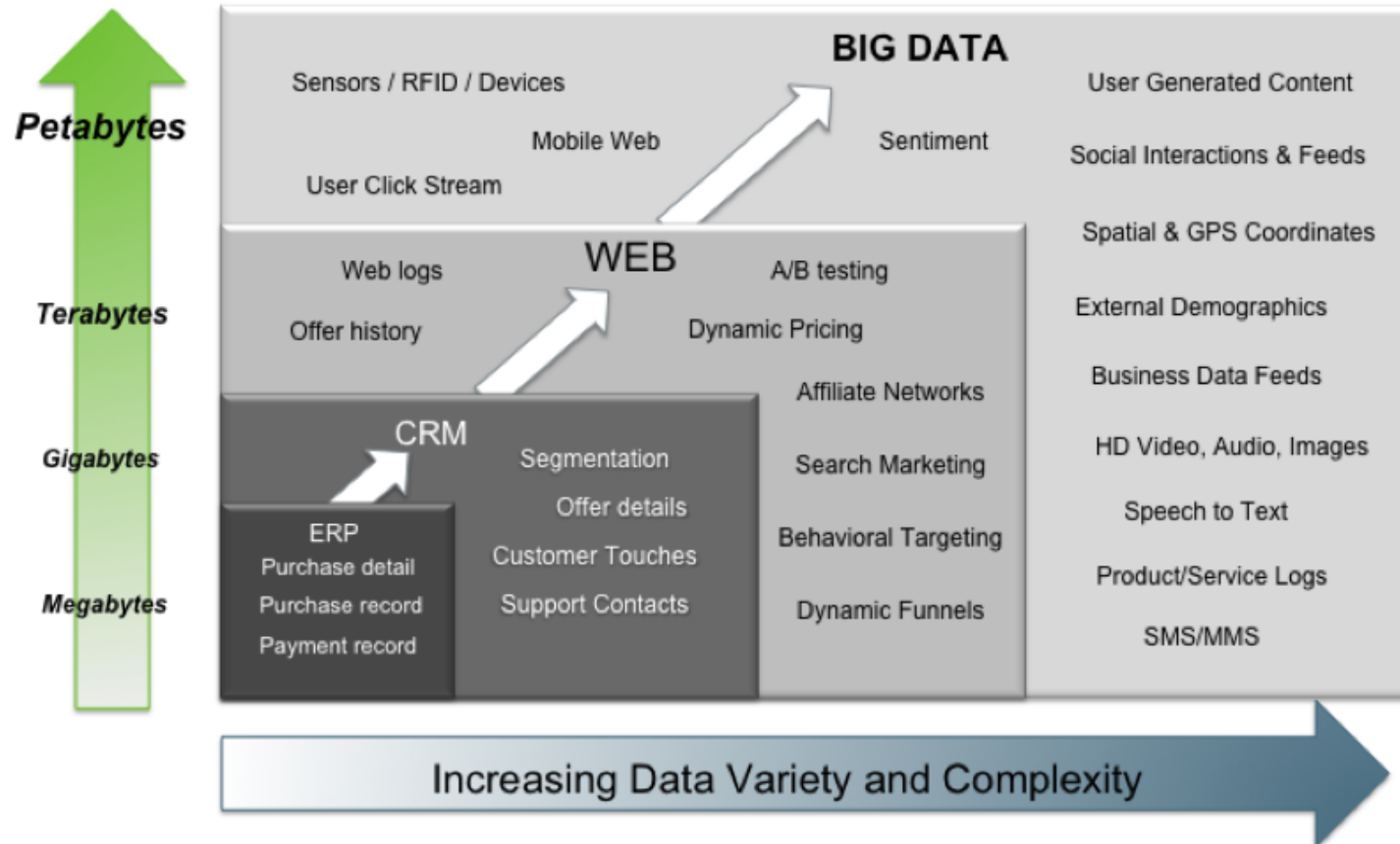
- 同義詞: big data, 巨量資料, 海量資料
- 重要里程碑
  - 2000初電子商務
  - 2000中社群網站、行動通訊
  - 2010初雲端運算、感測器網路、物聯網
  - 2012 美國政府開始投入研發經費
  - 2013政府開放資料啟動

# 大數據定義

- 維基百科: : big data指的是所涉及的資料量規模巨大到無法透過人工，在合理的時間內達到擷取、管理、處理、並整理成為人類所能解讀的形式的資訊
- 什麼是大數據? 大數據又會面臨什麼挑戰?

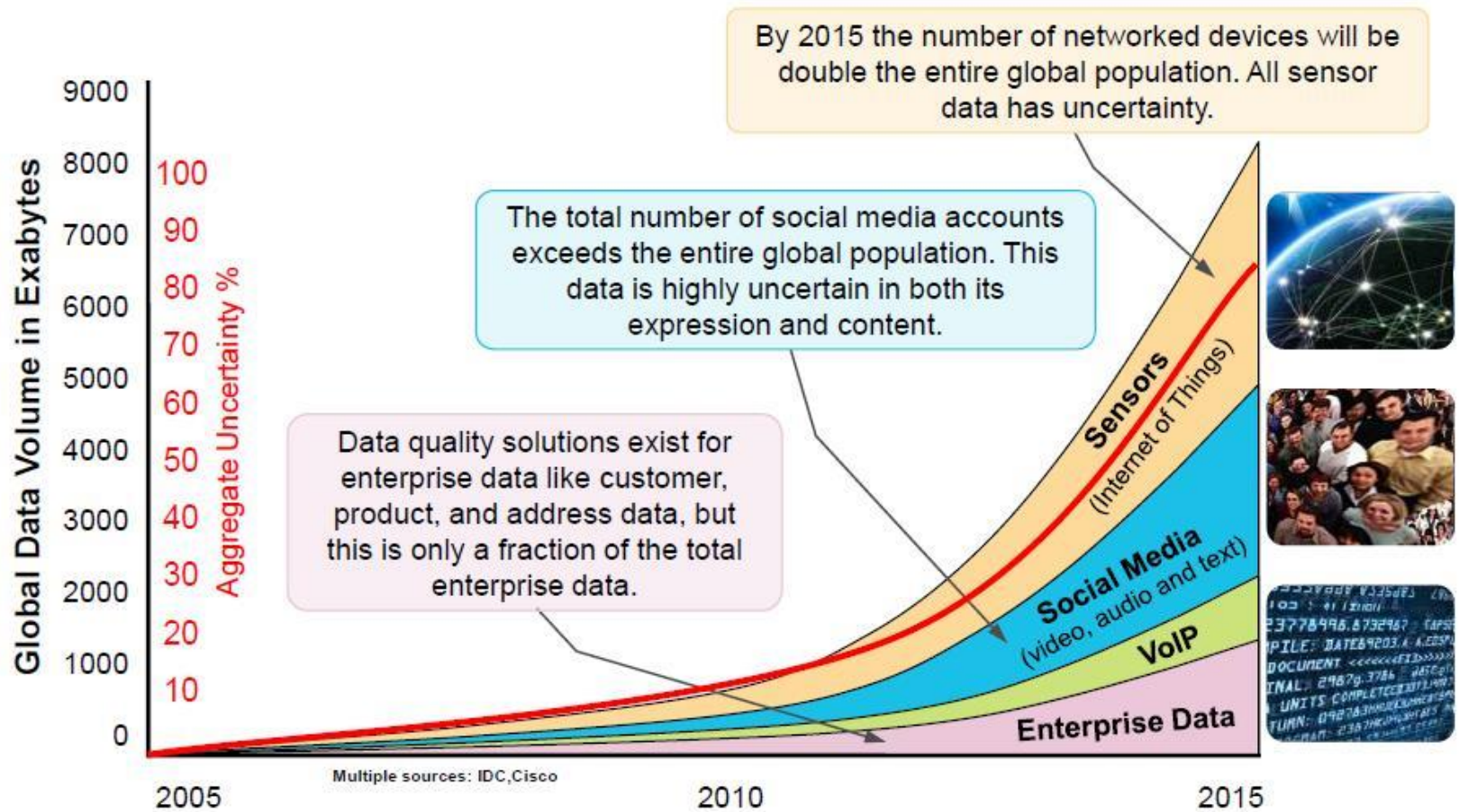
# 大數據來源

Big Data = Transactions + Interactions + Observations



Source: Contents of above graphic created in partnership with Teradata, Inc.

# 急速成長的數據



Source: IBM

# 大數據特徵(3V 或 5V)

- 數量大 (volume)
  - 每日推文數、智慧電表資料
- 速度快 (velocity)
  - 網路傳遞
- 樣式雜 (variety)
  - 圖片、社群關係
- 真實性 (veracity)
- 價值 (value)

# 大數據分析

- 從資訊科學到資料科學
  - 科技的進步使得處理大量資料能力增加
- 相關領域
  - 電腦科學
  - 統計學
  - 資料探勘
  - 機器學習
  - ...

# 資料分析技術

- 結構化資料: 橫列、直行，規格清楚，如 Access, SQL Server。資料探勘工具有
  - 關聯規則(買牛奶 => 買啤酒)
  - 分類預測(30歲以下，男性，單身 => 喜歡 Pokemon GO)
  - 分群分析(軍公教、製造業、交通業)
- 非結構化資料: 樣式雜是大數據特徵之一
  - 影片 => 影像處理技術
  - 文件 => 文件探勘
  - 網路 => 社會網路分析

# 非結構化資料分析

- 許多非結構化資料的分析還是使用傳統資料探勘工具(關聯規則、分類預測、分群分析)，但是在使用這些工具前會先從非結構化資料中萃取結構化屬性，滿足工具需求
- 文件分析
  - 本文萃取關鍵字
  - 以關鍵字建立向量(結構化屬性)
  - 使用分類演算法於結構化屬性，建立分類預測模式
- 圖片分析(分類、分群)
  - 萃取結構化屬性，使用傳統資料探勘工具
  - 近期: 由演算法來協助擷取屬性，如深度學習 (deep learning)

# 大數據分析基礎架構

- 雲端運算架構
- Hadoop 生態系
  - HDFS: 儲存大量資料
  - MapReduce: 分析大量資料
    - Map
    - Reduce
    - 搬計算資源比搬資料便宜

# 大數據應用

- 電子商務
- 國土安全
- 醫療照護
- 工程(自動車駕駛)

## 美國醫療照護

- ✓ 每年3,000億美元成效
- ✓ 每年0.7%生產力之提升 (約當21億美元)



## 歐盟公部門施政

- ✓ 每年2,500億歐元成效
- ✓ 每年0.5%生產力之提升 (約當12.5億歐元)



## 全球個人定位資料

- ✓ 服務供應商每年至少1,000億以上美元營收
- ✓ 消費者創造高達7,000億美元產值



## 美國零售業

- ✓ 提升60%以上的毛利
- ✓ 每年有0.5%~1%生產力之提升



## 製造業

- ✓ 產品開發與組裝成本降低50%
- ✓ 7%營運成本之降低



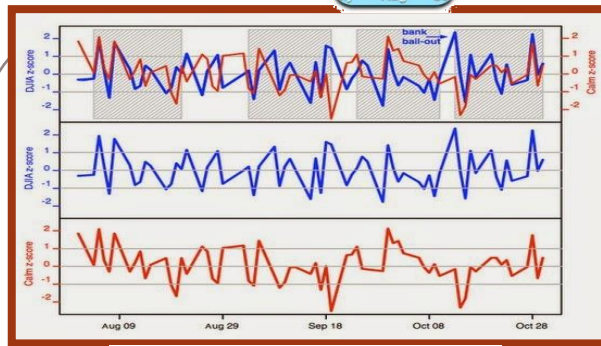
資料來源: McKinsey, 2013

# 廣告行銷運用

- 找金城武代言4G 電信商用大數據算出來？

# 社群網路大數據

## 股價



twitter

上億條社群推特/討論  
語意與情感分析

準確率達**87.6%**  
**15%**投資報酬率

## 健康



google.org Flu Trends

和美國疾病控制及預防  
中心合作，以關鍵字搜尋  
次數掌握流感

提前**2週**  
掌握流感爆發關鍵

# 結論

- 大數據對於未來生活各層面有重大影響
- 大數據是新石油，是一種策略性資源
- 大數據分析延伸資料探勘技術，處理更大量 (volume)、快速(velocity)與多元(variety)的資料
- 資料科學家(大數據分析師)是新興工作來源
- 常見的大數據來源有
  - 政府開放資料
  - 社群網路資料

# 大數據改變生活、創照新生意

- 遠見雜誌，看見未來5分鐘

# 電子商務日益繁榮，數據產業前景

- 民視新聞，財經火線

# 參考資料

- 美國國家標準局(NIST)大數據應用架構文件
- 政府資料開放平台
- 推特(Twitter)資料擷取API
- Kaggle 大數據平台(資料、競賽)