



ST LECTUREs
小 孫 學 堂

Big Data

大數據理論及實務應用

Created by 孫善堂 【小孫學堂】

講師簡歷 - 孫善堂 講師

現職

台灣跨領域人才培訓發展協會 理事長

意得彙智有限公司 CEO

勞動部勞動力發展署 關鍵就業力 KC講師

學歷

國立台灣大學 森林系 學士

元智大學 工業工程與管理 碩士班

輔仁大學 經濟學系 研究所

經歷

- 台灣人工智能產業協會 AI課程講師
- 社團法人中華勞動力職能發展協會 大數據/物聯網/AI課程講師
- 社團法人台灣創意啟發協會 大數據/物聯網/AI課程講師
- 企業內訓講師及專案輔導顧問：

貳樓餐廳、創創數位科技股份有限公司、馮佑國際貿易有限公司、ADOLE時尚皮件、巴夏喀有限公司、VESSEL中呈實業股份有限公司、WSS鐳射槍戰-力新創意有限公司、粹智國際實業有限公司、運通旅行社、品王旅行社有限公司、Webticket旅遊便利網、勤貿實業股份有限公司、麒麟股份有限公司、巴蜀香四川麵食鍋物、天下布武日月潭紅茶、元昶玻璃股份有限公司、格郁文教事業股份有限公司等



個人教育品牌：小孫學堂

歡迎關注我嘅！



臉書粉絲專頁

<https://www.facebook.com/STLecture/>

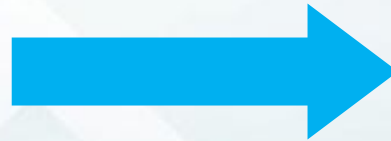
Youtube頻道

<http://www.youtube.com/c/小孫學堂>





Data Visualization



What is the Big Data ?

Big data is data sets that are so voluminous and complex that traditional data-processing application software are inadequate to deal with them.

What is the Big Data ?



What is the Big Data ?

Big data (n.名詞)

無法用”傳統方法”處理的”資料們”

(DB2 、 Oracle 、 MSSQL 、 MySQL)

What is the Big Data ?

Volume

大量

Velocity

快速

Variety

資料類型不定

Veracity

正確性不定

The “three V’s”, i.e the Volume, Variety and Velocity of the data coming in is what creates the challenge.

VOLUME



Amount of Big Data stored across the world (in petabytes)

VARIETY



PEOPLE TO PEOPLE

NETIZENS, VIRTUAL COMMUNITIES, SOCIAL NETWORKS, WEB LOGS...



PEOPLE TO MACHINE

ARCHIVES, MEDICAL DEVICES, DIGITAL TV, E-COMMERCE, SMART CARDS, BANK CARDS, COMPUTERS, MOBILES...



MACHINE TO MACHINE

SENSORS, GPS DEVICES, BAR CODE SCANNERS, SURVEILLANCE CAMERAS, SCIENTIFIC RESEARCH...



2.9
MILLION

EMAILS SENT EVERY SECOND



20
HOURS

OF VIDEO UPLOADED EVERY MIN



50
MILLION

TWEETS PER DAY

VELOCITY

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session



By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth



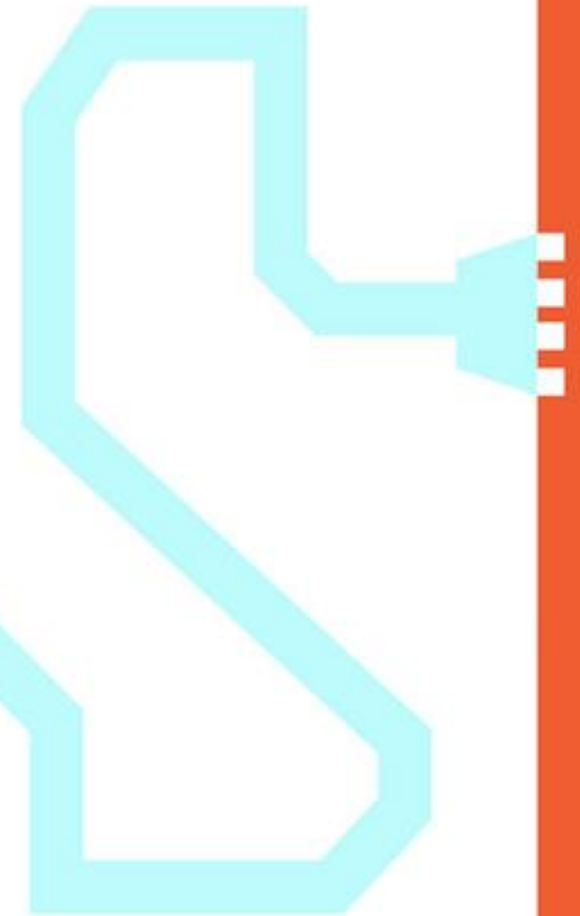
Modern cars have close to

100 SENSORS

that monitor items such as fuel level and tire pressure

Velocity

ANALYSIS OF STREAMING DATA



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]



Variety

DIFFERENT FORMS OF DATA

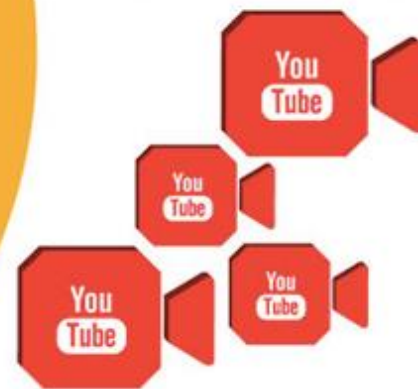
By 2014, it's anticipated there will be

420 MILLION WEARABLE, WIRELESS HEALTH MONITORS



4 BILLION+ HOURS OF VIDEO

are watched on YouTube each month



30 BILLION PIECES OF CONTENT

are shared on Facebook every month



400 MILLION TWEETS

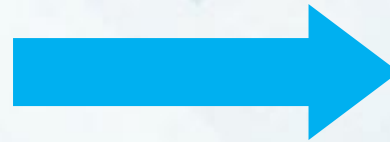
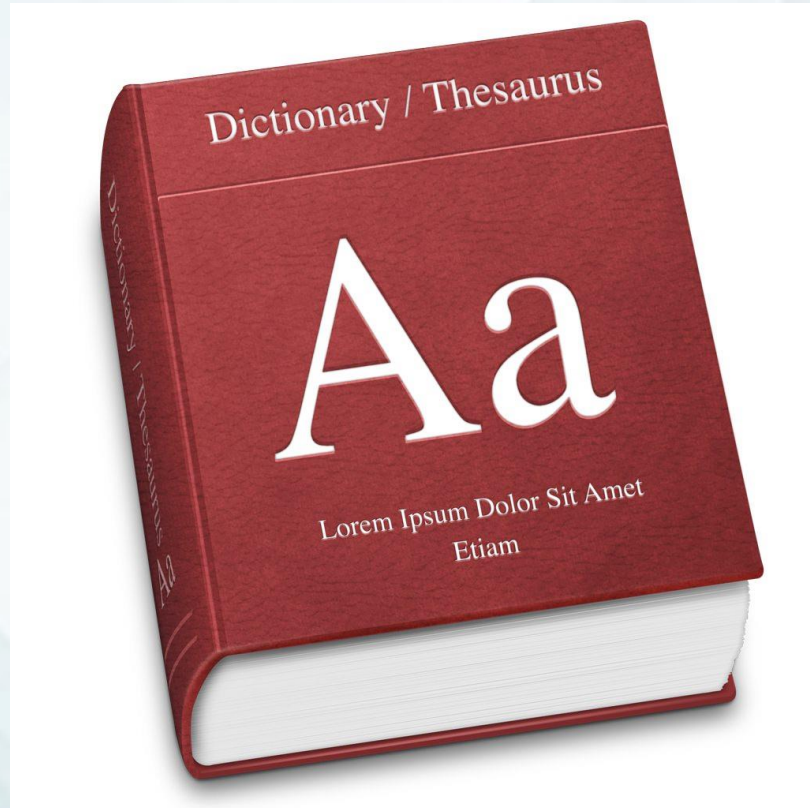
are sent per day by about 200 million monthly active users



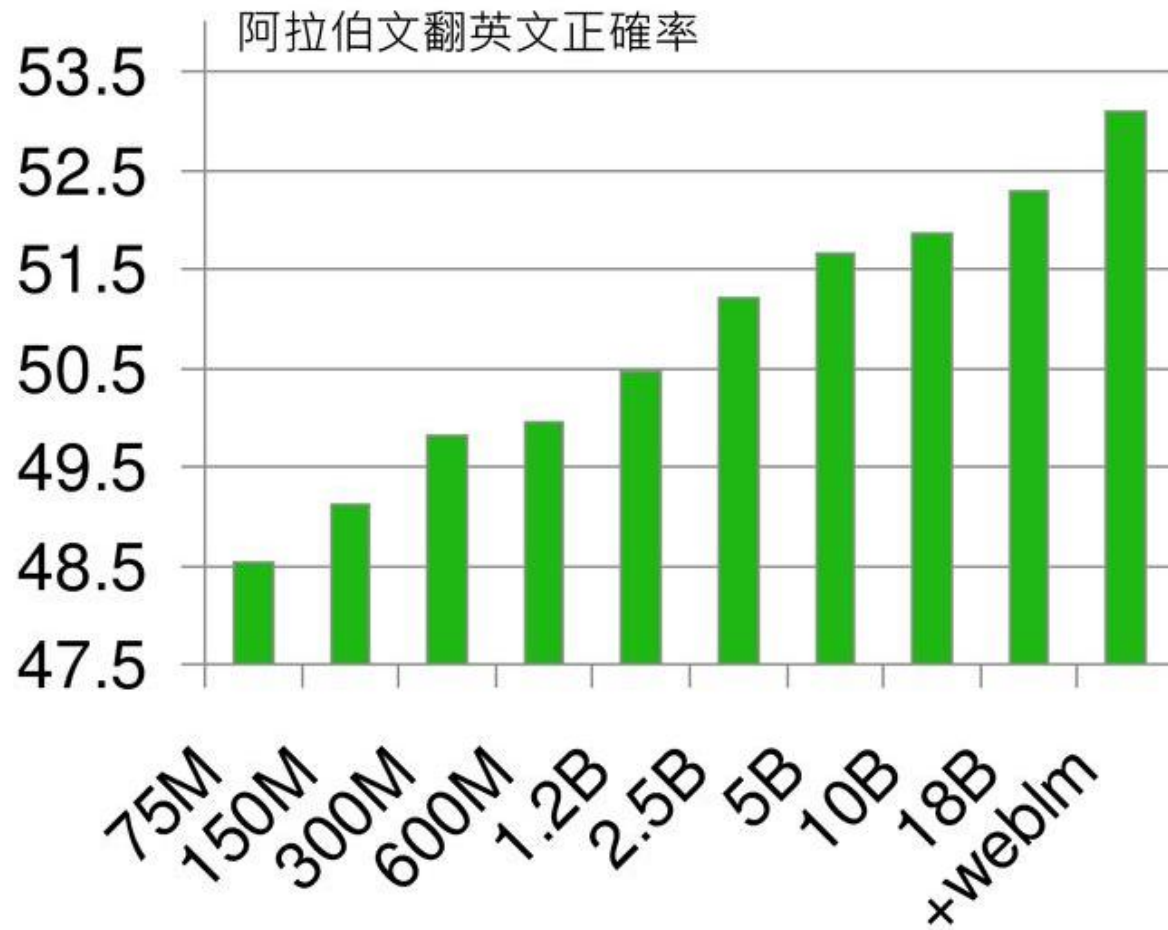
Garbage IN Garbage OUT?



Example 1 : translation



Example 1 : translation



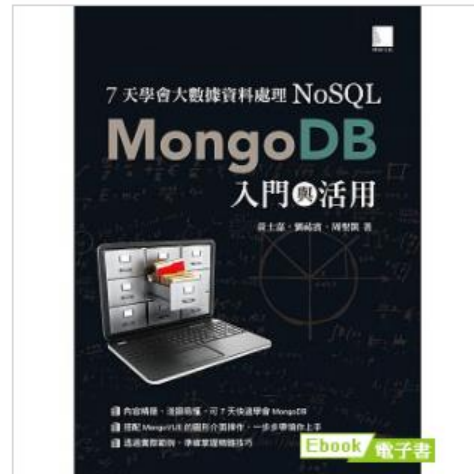
+web1m =
219B words of
web data!

Example 2 : Walmart



Example 3 : Amazon first

博客來 > 電子書 > 中文電子書 > 電腦資訊 > 資料庫 > MSSQL-Server > 商品介紹



7天學會大數據資料處理 NoSQL : MongoDB入門與活用 (電子書)

作者：黃士嘉, 劉祐宜, 周聖凱

出版社：博碩

出版日期：2016/03/01

語言：繁體中文

定價：320元

優惠價：230元

抵用購物金最多再省\$69 詳情

放入購物車

直接結帳

加入下次再買清單

電子書閱讀軟體

支援瀏覽器 ?

APP下載：



我要寫評鑑

分享

讚 0

G+

試閱
收藏
試閱本
45人收
適合
平板

內容簡介

在大資料 (Big Data) 時代，NoSQL (Not only SQL) 已經成為資料儲存的主流，NoSQL 的出現並非意味著關聯式資料庫系統 (Relational Database Management System, RDBMS) 的消失，而是在網路上資料特性更加複雜與大量，NoSQL 代表著一新型態的資料庫系統的架構類別，因應不同對於在資料儲存及處理需求，延伸出多種儲存方式，例如：文件型 (Document)、鍵值型 (Key-Value)、記憶體型 (In-memory)、圖學型 (Graph) 等，其相符於 BigData 對於資料的三大特性—Volume、Velocity、Variety—簡稱 3V 定義，將可成為新形態的資料儲存之有效解決的方案。

在資料庫網站排行榜「DB-Engines Ranking」(db-engines.com/en/ranking) 中，對於各類資料庫系統在網路上被提及的數量、Google Trend 的關鍵字搜尋頻率、Stack Overflow 相關的技術討論、業界所開出的工作需求、個人於 LinkedIn 履歷所列出的專業技能以及 Twitter 社群討論活躍度之綜合評估上，以文件類型 (Document Store) 的 MongoDB 最為活躍，2015 年位在所有資料庫系統的排名第 4 位，其中在 NoSQL 中又為最具影響力的資料庫之第 1 名。因此，本書將以 MongoDB 為主進行介紹與教學，其 MongoDB 具有的特性如下：



瀏覽此商品的人，也瀏覽 ...

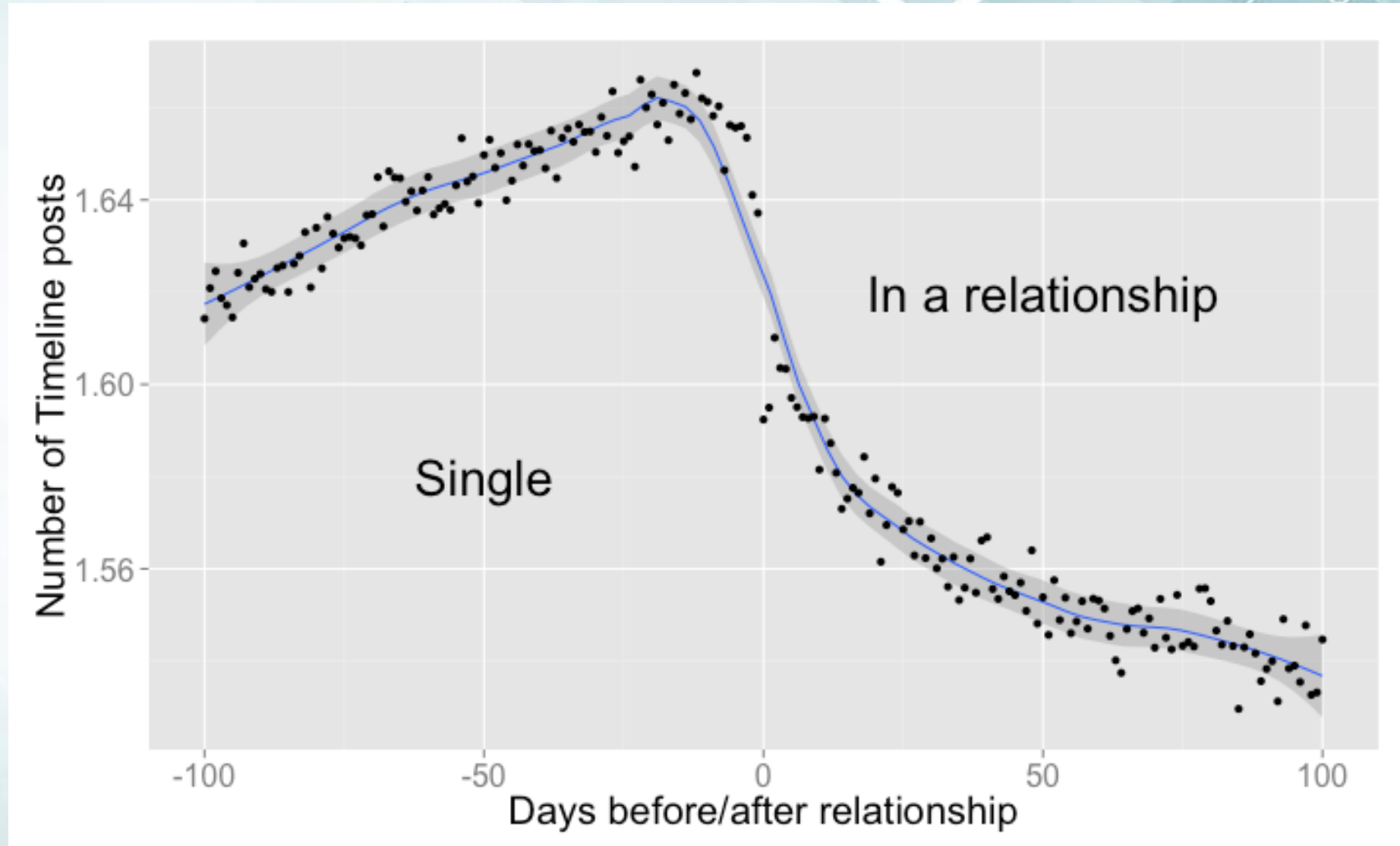


What is the Big Data ?

Big data (n.名詞.延伸)

了解如何挖掘、設計分析，
從資料中提取出價值的技術。

Example 4 : relationship



Level of the Big Data

What you can **save** is storage.

What you can **see** is Data.

What you can **understand** is information.

What you can **use** is intelligence.

Who can do the Big Data ?

擁有Big Data的人

會寫程式分析Big Data的人

知道Big Data價值何在的人

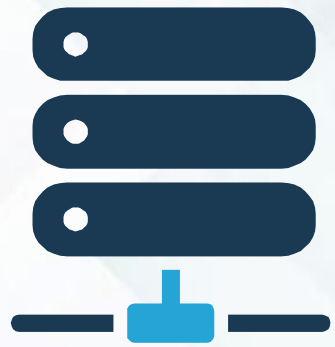
And, Who earn the most ?

資料分析優化導入

外部數據



內部數據



資料庫



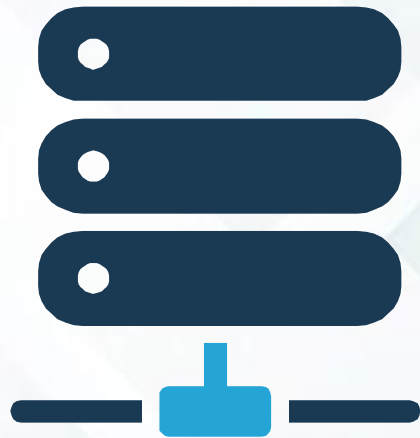
視覺化報告



資料分析優化導入

問題解決導向

數據蒐集



現有數據

針對問題



優化
決策

數據應用導向

現狀優化

資料分析優化導入流程：現有數據

產量優化

績效管理

離職預測

薪資政策

銷售策略

