出國報告（出國類別：國際會議）

參加美國 2016 年全球數據科學研討會 (Global Data Science Conference) 出國報告

服務機關：行政院主計總處主計資訊處
姓名職稱：周科長采蓉
派赴國家：美國
出國期間：105 年 3 月 6 日至 105 年 3 月 11 日
報告日期：105 年 6 月

摘　要

全球數據科學研討會(Global Data Science Conference)係由數據科學中心(Data Science Central)舉辦之年度國際研討會，本會議透過各種數據管理技術探討數據科學，主要包含大數據、預測分析、機器學習、Hadoop、物聯網等議題。2016 年全球數據科學研討會 3 月 7 至 9 日於美國聖克拉拉舉辦，從初創公司到大型企業的高級主管，由超過 20 位數據科學領域的頂尖專家主講。

考量數據科學已逐漸形成統計、資訊科技、行政管理等跨領域之學科，為未來政府深入探討施政績效及民間潛在需求之重要工具。為應用數據分析、精進公共政策與服務，協助總處有效發掘歷史數據，茲派員參加本研討會，以熟悉數據應用潛力，推升資料管理利用及數據加值之廣度，有效促進政府機關間資料整合及加值運用，優化主計資訊服務品質，再造主計資料創新價值。

本報告分五章：第一、二章為目的及過程，第三章為重要內容摘要，第四章為心得及建議，第五章為結語，提供後續數據應用規劃之參考。

目　次

一、目的.....1

二、過程.....2

三、重要內容摘要.....7

四、心得及建議.....17

五、結語.....25

參考資料.....26

附錄.....27

一、目的

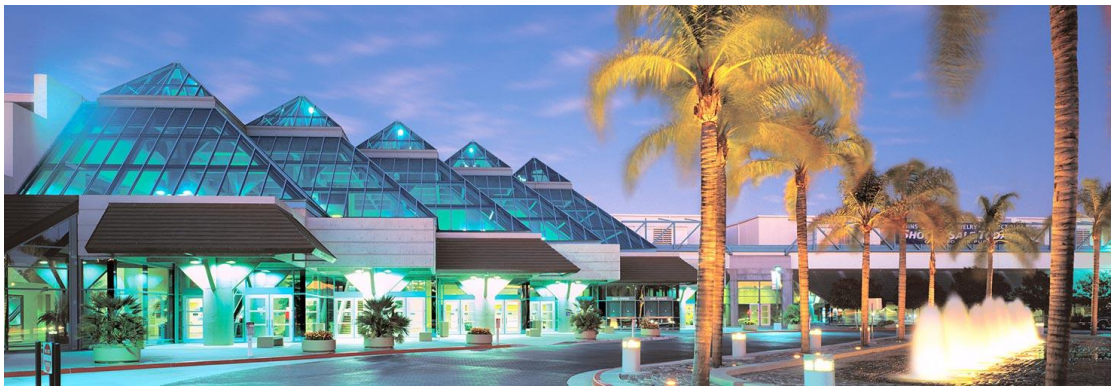
全球數據科學研討會(Global Data Science Conference)係由數據科學中心(Data Science Central)舉辦之年度國際研討會，本會議透過各種數據管理技術探討數據科學，主要包含大數據、預測分析、機器學習、Hadoop、物聯網等議題。數據科學使組織能有效地發掘歷史數據並更了解用戶的行為，各種電子商務交易、移動平臺訊息、社交媒體互動等創造了大量數據，物聯網也使組織能有效分析數據以調整所提供的服務，適當地運用數據科學可為組織帶來重要的競爭優勢。

考量數據科學已逐漸形成統計、資訊科技、行政管理等跨領域之學科，為未來政府深入探討施政績效及民間潛在需求之重要工具。為應用數據分析、精進公共政策與服務，協助本總處有效發掘歷史數據，茲派員參加 2016 年全球數據科學研討會，以熟悉數據應用潛力，推升資料管理利用及數據加值之廣度，有效促進政府機關間資料整合及加值運用，優化主計資訊服務品質，再造主計資料創新價值。

二、過程

2016 年全球數據科學研討會 3 月 7 至 9 日於美國聖克拉拉舉辦，邀請從初創公司到大型企業的高級主管，由超過 20 位數據科學領域的頂尖專家主講。以下為研討會官網及舉辦地點「聖克拉拉會議中心」:

http://globalbigdataconference.com/67/santa-clara/global-data-science-conference/event.html



議程如下：

表 1 議程

Day-1 (105 年 3 月 7 日)	
9:00AM - 12:00 PM	Introduction to Machine Learning & Predictive Analytics using R Workshop (Sudhir Wadhwa, CTO, DataTiles.io)
1:00PM - 1:40PM	An Analytics Odyssey: From Predictive to Cognitive (Piyush Malik, Director, IBM)
1:50PM - 6:00PM	Introduction to Machine Learning & Predictive

	Analytics using R Workshop (Sudhir Wadhwa, CTO, DataTiles.io)
Day-2 (105年3月8日)	
9:00 AM - 9:40AM	How Comcast uses Data Science and Machine Learning to improve the Customer Experience (Jan Neumann, Data Science Team Lead, Comcast)
9:40 AM - 10:20AM	Big-Data analytics misconceptions (Irad Ben-Gal, Chairman & Professor, Stanford University)
10:40 AM - 11:20AM	Hunting Criminals with Hybrid Analytics (David Talby, Chief Technology Officer, Atigeo)
11:20 AM - 12:00PM	Detect Sensitive Data in Hadoop Clusters (Benoy Antony, Founder, DataApps)
1:00PM - 1:40PM	Monitoring and Troubleshooting Real-time Data Pipelines (Alan Ngai, VP of Engineering, OpsClarity & Premal Shah, Co-Founder, 6sense)
1:40PM - 2:20PM	Netflix Keystone - How we built a 700B/day stream processing cloud platform in a year

	(Peter Bakas, Director, Netflix)
2:20PM - 3:00PM	Create impact at scale by data-driven applications (Leo Li, Analytics Leader, LinkedIn & Wendy Shi, Manager, LinkedIn)
3:20PM - 4:00PM	Predicting The Future: Surprising Revelations From Truly Big Data (Pushpraj Shukla, Principal Data Scientist, Microsoft)
4:00PM - 4:40PM	Detecting Anomalies in Streaming Data, Evaluating Algorithms for Real-world Use (Alexander Lavin, Research Engineer, Numenta)
4:40PM - 5:30PM	Keynote Panel: The Future of Data Science Milind Bhandarkar(CEO, Ampool) - Moderator Peter Bakas(Director, Netflix) Leo Li (Analytics Leader, LinkedIn) Pushpraj Shukla (Principal Data Scientist, Microsoft) Mike Tamir (Chief Science Officer, Galvanize) Alexander Lavin (Research Engineer, Numenta)

Day-3 (105年3月9日)	
9:00 AM - 9:40 AM	Using Analytics to Drive Change in the Workforce (Genetha Gray, Data Scientist, Intel)
9:40 AM - 10:20 AM	Random Forests: How a Chance Driven Learning Machine Does So Spectacularly Well on Marketing Datasets (Dan Steinberg, CEO, Salford Systems)
10:40AM - 11:20 AM	Structuring Data for Self-Serve Customer Insights (Jim Porzak, Data Scientist, DS4CI)
11:20AM - 12:00 PM	Case studies in connecting Devices to cloud - experience of an Indian Startup (Janakiram Dharanipragada, Senior Researcher, IIT Madras)
1:00 PM - 1:40PM.	Building a Predictive Intelligence Engine (Viral Bajaria, Chief Technology Officer, 6sense)
1:40 PM - 2:20 PM	Predictive Analytics, Inventory Management, Machine Learning, Ecommerce (Paolo Massimi, Director, Stitchfix)
2:20 PM - 3:00 PM	How to Build a Better Network of Community Care Providers Using Data

	(Mohamed Elmallah, Manager, Children's Hospital of Los Angeles)
3:10 PM - 3:50 PM	User behavioral predictive analytics through deep learning-based emotion recognition (Jr Alaoui, CEO, Eyeris)
3:50 PM - 4:30 PM.	Hybrid Artificial Intelligence (Manuel Ebert, CEO, Summer.ai)
4:30 PM - 5:30 PM.	Solving Business Problems With Data Science Krishna Sankar (Chief Data Scientist, Blackarrow.tv) Robert Bernard (Data Scientist, Boeing) Mohamed Elmallah (Manager, Children's Hospital of Los Angeles) Shankar Vedaraman (Manager, Netflix)

三、重要內容摘要

1.介紹機器學習及預測分析—使用R工作坊(Introduction to Machine Learning & Predictive Analytics using R Workshop)

R軟體是一個程式語言、統計分析及繪圖功能的整合環境，具完整且彈性的物件導向程式語言架構及強大的繪圖功能，能跨平臺及互動式處理資料之特質，可應用於各式各樣的資料分析任務。

本工作坊從R軟體基本概念、資料整理及組織切入，著重於資料理解、統計分析及資料視覺化等基本操作方法的運用，以整合式開發環境實機練習，配合圖形化使用者界面操作及應用，幫助初學者快速跨越R陡峭的學習曲線。有鑑於功能強大的R軟體係以大量物件轉換及分析資料，入門者不僅要有程式設計觀念，還須具備統計知識。內容綱要如下：

(1) R基本觀念。

(2) 各式資料匯入、儲存、匯出。

(3) 機率分佈及抽樣。

(4) 資料整合、資料清理、資料轉換。

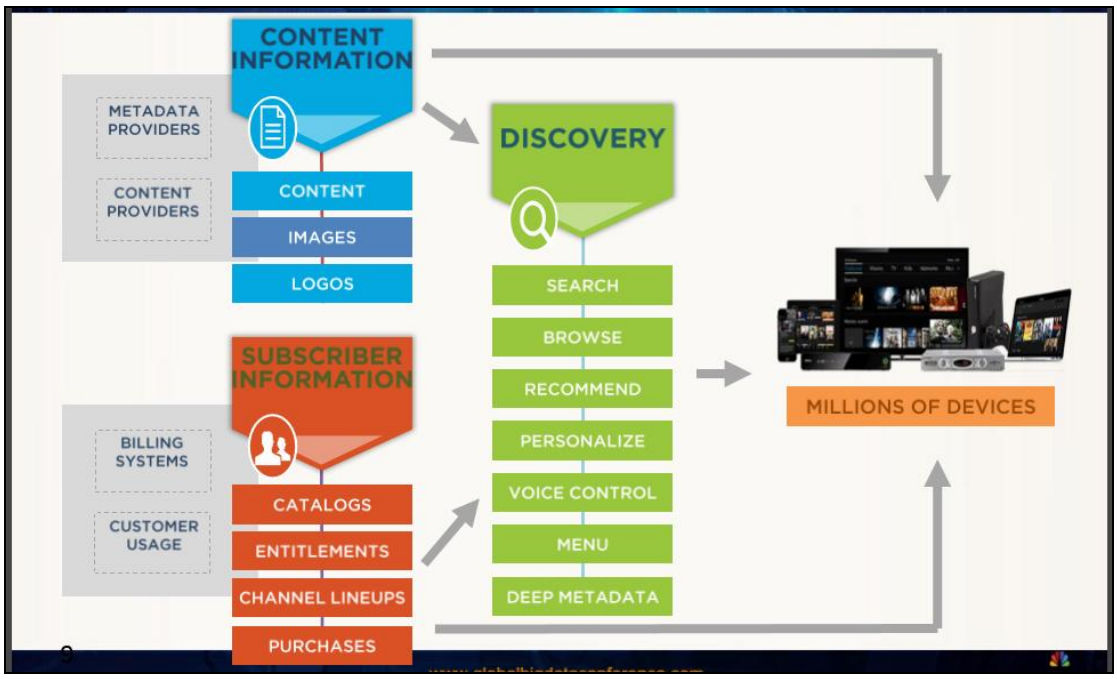
(5) 資料子集、資料變形、資料排序。

(6) 線性迴歸及模型選擇。

(7) 資料視覺化簡介。

2.如何運用數據科學及機器學習，改善客戶體驗(How Comcast uses Data Science and Machine Learning to improve the Customer Experience)

據統計，美國家庭平均每天花費 3 至 5 小時看電視，大約佔了一半的休閒時間。本節主要說明這個團隊及平臺如何打造世界一流的電視搜尋引擎，透由尖端科技的搜尋、瀏覽、語音及個性化功能的創新，幫助大約 2 千 2 百萬用戶，無論是從電視機、網路、手機及平板等設備，隨時隨地找到喜歡的節目及電影。

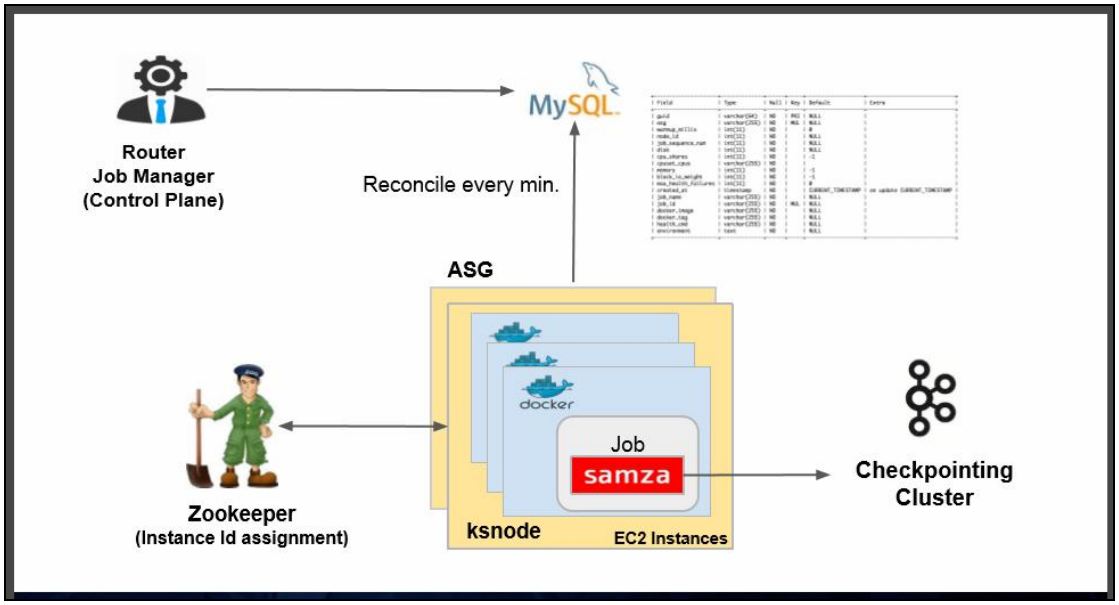


他們收集及建立每個節目及電影的 Metadata，如：簡介、片長、排名等，研究、開發並測試了先進的內容分析演算法及基礎設施，並提供每位用戶一個免費的語音介面電視遙控器，用來搜尋片名、演員、主題及價格等。藉由分析用戶搜尋及瀏覽等收視習慣，結合社群網站的數據，利用機器學習、自然語言理解、計算機視覺、深度學習及分散式計算等技術，研發轉換的電視及智能家居經驗，每 20 秒更新對下一小時、明天、熱門時段的收視建議，事先下載建議的節目內

容，減少儲存空間及網路雍塞，強化用戶未來的娛樂體驗，甚至還能建議未來節目及電影內容的發展方向，讓它們更受歡迎。

3.Netflix 如何在一年內建立每天處理 700B 的雲平臺(Netflix Keystone - How we built a 700B/day stream processing cloud platform in a year)

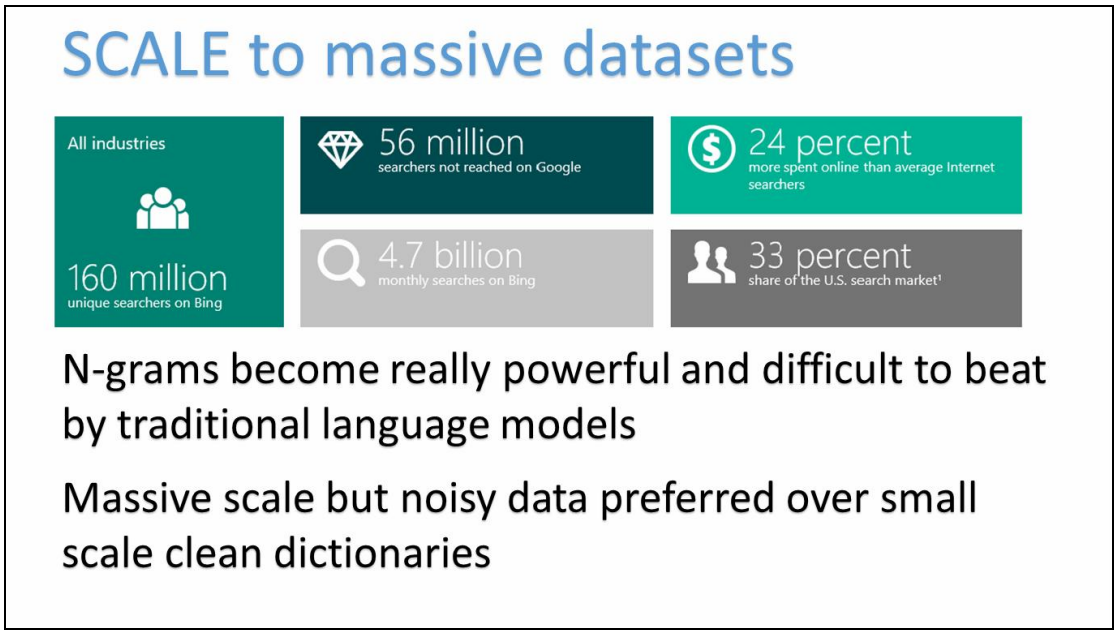
Netflix 於 2010 年將其 IT 基礎架構整合至 Amazon EC2 平臺，電影製片廠提供的數位影片主拷貝則存放在 Amazon S3 伺服器中。Netflix 在其後端也使用了數種開源軟體技術，包括 Java、MySQL、Gluster、Apache Tomcat、Hive、Chukwa、Cassandra 及 Hadoop。



本節主要探討：Netflix 如何在一年之內大規模地修改及利用 Kafka、Samza、Docker 及 Linux，以實作在亞馬遜 AWS 雲的多租戶型態；如何在雲中有效地將這些技術融合在一起，並提出性能數據、部署及監控細節、容錯及故障轉移策略。

4.預測未來：來自真正大數據的驚人啟示(Predicting The Future: Surprising Revelations From Truly Big Data)

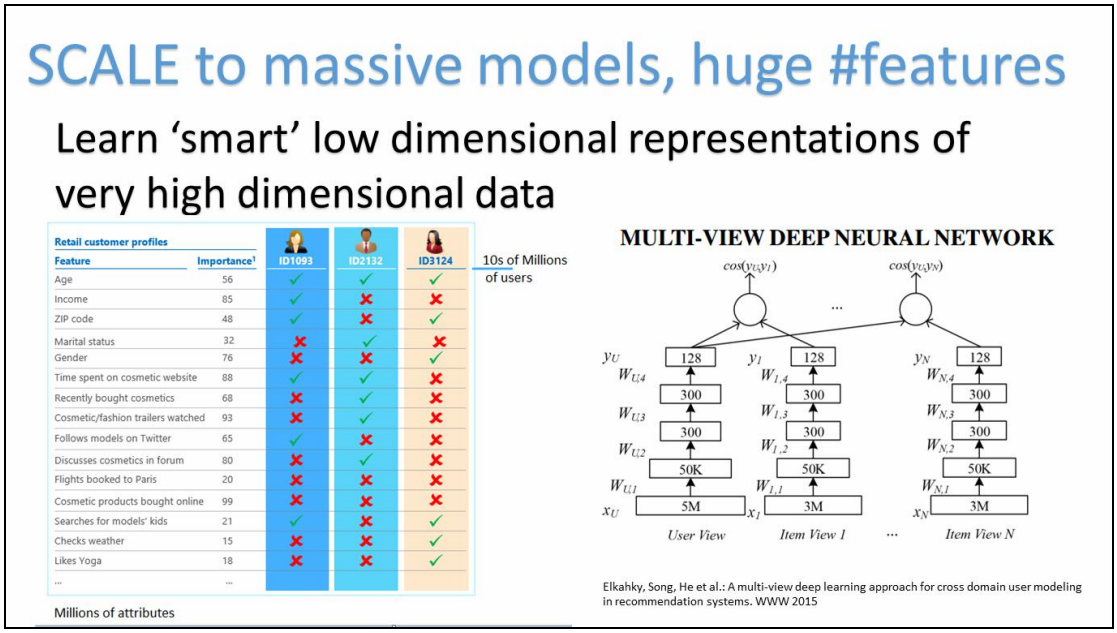
過去， Bing.com 在網頁及圖片的搜索上已經做了深入的改善。利用類似的搜索技術，整個團隊走向一個新的領域「Bing Predicts」，此功能結合演算法及訊息資產，建立預測引擎，分析網路活動、社會情緒及其他訊息，來預測事件的結果，客觀地提供用戶更多的知識及量身打造的見解，幫助他們跨越各種主題做出更有自信的決策。



本節主要說明：「Microsoft Bing」如何利用搜尋網站、社群網站及其他相關數據，對於即將舉辦的活動，如：TV 真人實境秀、體育賽事等，做出更智能化的預測。

因搜尋引擎幾乎已經成為所有用戶在網路上搜尋訊息的切入點，社群網站更

讓用戶創造、分享及交換訊息。挖掘這些網站數據及社群媒體的內容，讓我們有機會去發現用戶對某些事件的情緒、預測流行趨勢，並運用搜索網站及社群數據的演算法，透過機器學習，持續調校預測的方法。



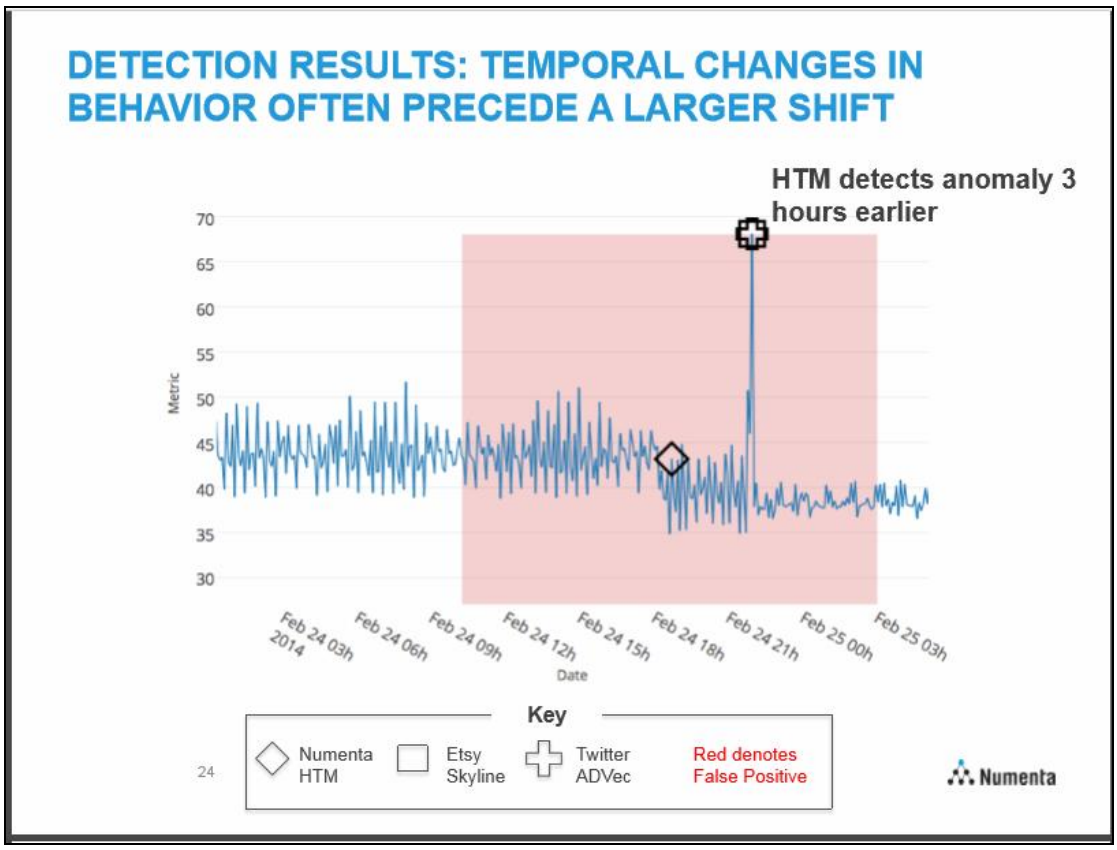
觀察網路活動，可以藉由群眾的智慧，較傳統的諮詢方式擷取更多值得信賴的見解，Microsoft 將之運用於（1）用戶投票的比賽或活動、（2）基於評審的比賽，如：奧斯卡獎等、（3）體育等現場比賽方面，都得到不錯的成果。

5.偵測異常數據流—實際應用的評估演算法(Detecting Anomalies in Streaming Data, Evaluating Algorithms for Real-world Use)

本節主要說明：隨著物聯網、偵測器等即時數據收集的快速增長，導致數據流爆炸般的成長。偵測的目標往往在於預防，而機器學習的最大應用就是偵測異常，如：監控 IT 基礎設施、監控能源消耗、監控即時健康狀況、揭開詐欺交易、

跟蹤車輛等。

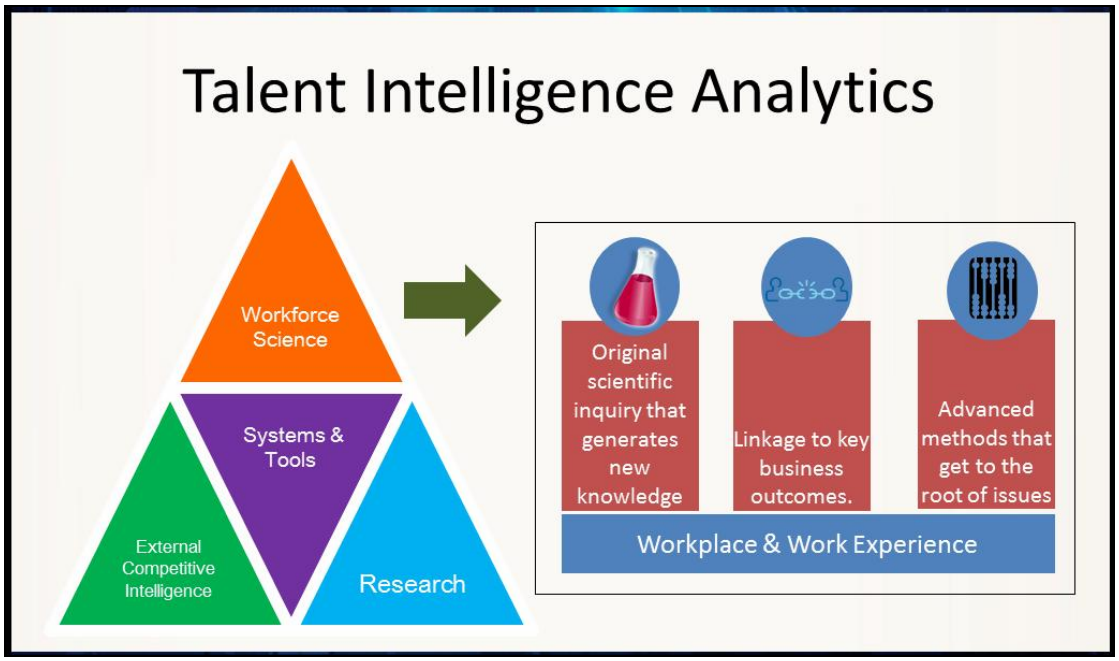
Numenta Anomaly Benchmark(NAB)是第一個設計用來評斷時間序列數據的基準，它運用跨域的真實世界數據集及標記數據檔案，發展了一套獨特的評分機制，以獎勵早期發現異常，並懲罰遲到或虛假的結果。

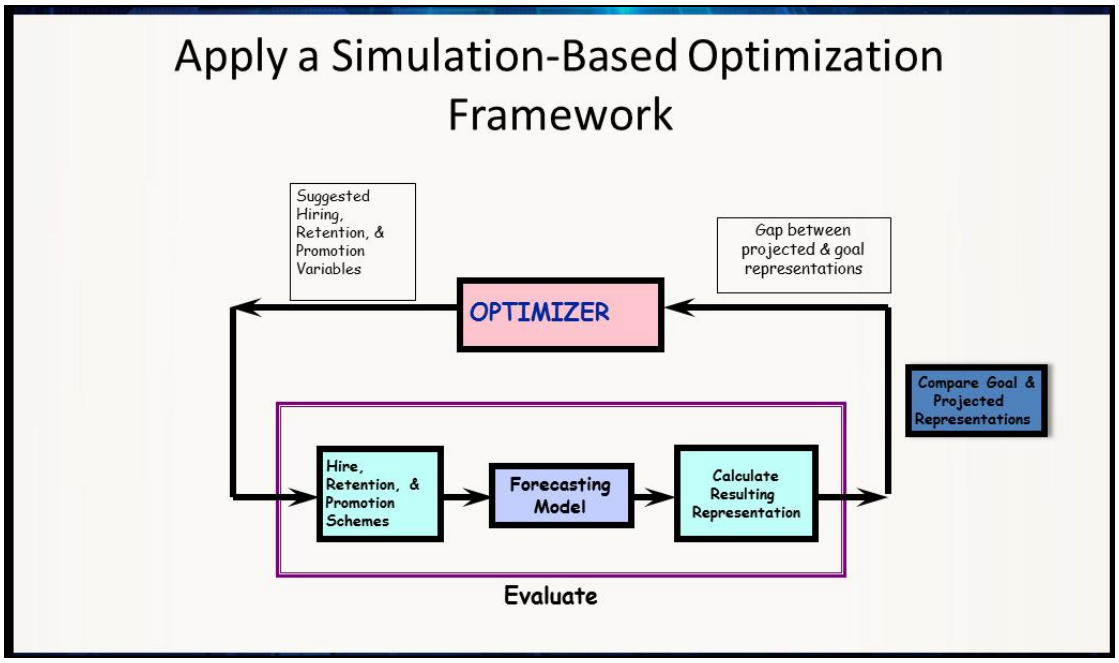


NAB 是一個模組化的開源程式碼庫，包含所有的數據檔案，演算法及文件，這些寶貴的資料是經由多年與客戶合作、解決他們的異常問題所累積的，相關的研究論文也說明了 NAB 的技術細節及評分機制的數學解釋。未來將建立一個社群，以增加數據檔案並測試更多的演算法。

6.運用分析推動勞動力的變革(Using Analytics to Drive Change in the Workforce)

部分婦女及少數民族在科學、技術、工程及數學領域有顯著的貢獻，但在組織高層擔任重要職務的卻不多。2014 年夏天，許多矽谷的大公司揭示，其在技術職位的女性員工僅佔 10-20%。為了改變這個趨勢並符合業務需求，2015 年 1 月英特爾執行長宣布，將於 2020 年前促進婦女和少數民族的就業發展。

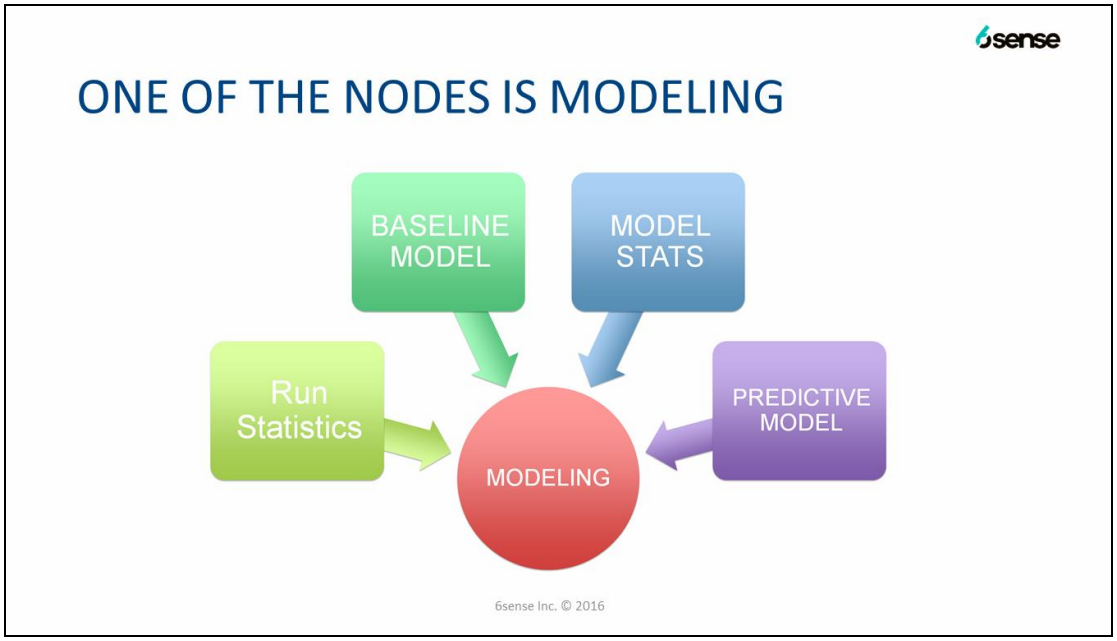




本節主要說明：Intel 人力資源部門如何進行文化上的轉變及新的分析工作，從根本上改變勞動力的供需問題，運用數據融合、數據挖掘、數據分析等新方法，及數據可視化技術來預測的不確定性，並於 2015 年底就看見初步的成效。

7.構建具預測能力的智慧引擎(Building a Predictive Intelligence Engine)

我們生活在一個到處留下數字足跡的世界。在消費領域，像 Amazon 和 Google 就利用這些洞察力預見到你想要什麼—在你知道你想要它之前！但在 B2B 領域，數據的不規則和複雜性帶來一個先所未見的難題：瞭解誰是買方？



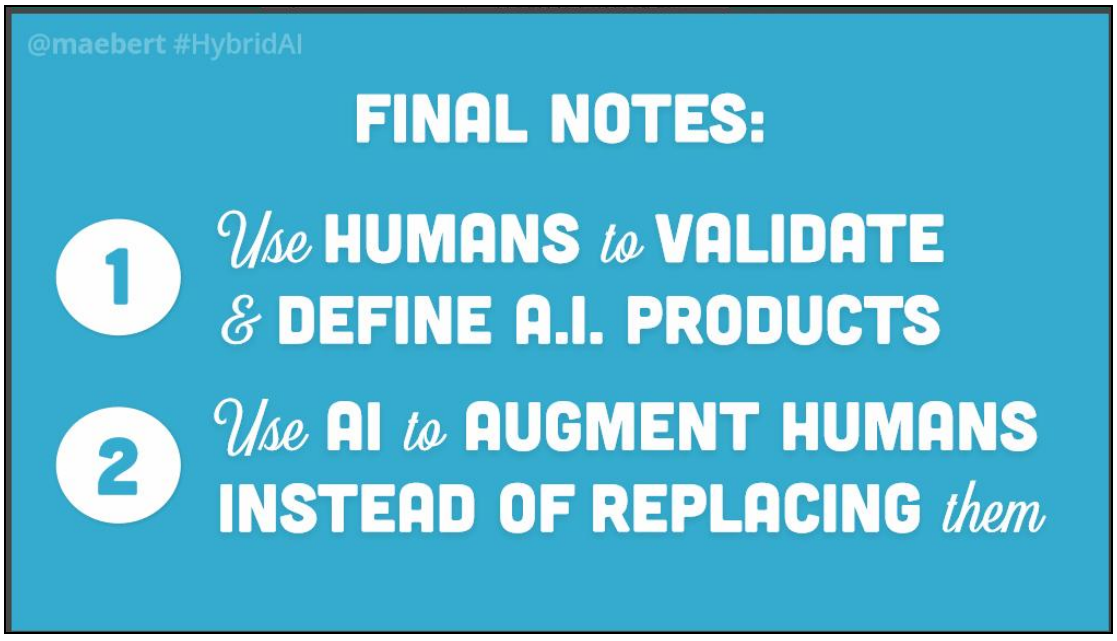
本節主要探討：為了在銷售週期早期就找到 B2B 的買家，結合數十億客戶接觸數據及預測智能技術、非結構化的時間敏感數據（例如上千個 B2B 網站上活動）及結構化的行為數據，創建一個行為目錄，可以分析來自所有網站的數據。如何重新再利用大多數的模型建立管道；如何使用 H2O、Docke、Mesos 及 Hive，每天產生數百款模型；如何在同一個數據集執行多種模型以監控效能。這些方法已經運用在財富前 1000 大公司，如：思科等，思科藉著這些預測技術發現新的買方，帶來近 3 億美元的新業務量。

8.混合型人工智慧(Hybrid Artificial Intelligence)

雖然人工智慧技術已經大大地進步，但這些服務真正的突破是在一個概念上的轉變—應用人工智慧來增加人類的智慧，而不是取代人類。過去一年出現了一波以新人工智慧為基礎的服務，如：安排會議行程、計劃旅行等新的個人助理。

這種半人類、半機器的混合，所需成本不高，卻完成了超乎人類水準的精準任務，並讓與人類的互動真正擴展為全新的商業應用模式。

本節主要探討：混合型人工智慧如何運作？為什麼它們會改變企業的運作、創造全新的市場？它們如何應用在現有的產業？以及預期混合型人工智慧未來幾年的發展。並舉 Siri 為例，它可以幫我們播電話、送簡訊、設鬧鐘、作筆記、找地點、上網、查詢股票價格及天氣資訊等。當人類詢問天氣時，它首先將感測到人類的語音作辨識，自動加上手機提供的時間點、GPS 提供的地點做為參數，呼叫相關提供天氣資訊的應用程式，最後發音將結果讀出。也就是說，混合型人工智慧並不是一個技術上的突破，而是一個商業應用上的突破。



四、心得及建議

「數位革命」之後，「資料革命」登場，這是一個從數據到知識、從知識到行動的時代！數據成為新的生產原料、重要的經濟資源，可以創造出新的經濟價值；只要心態及思維正確，就能巧妙地重複運用數據，不斷帶來創新和不同的服務；只要夠謙卑、有意願傾聽，也有工具，數據就能讓秘密躍然眼前。

當累積的數據量到達相當規模，帶來全新且特殊的改變，從量變引發了質變，掀起生活、工作及思考方式的全面革新，改變現有公民、組織及政府間的關係。隨著大數據(Big Data)時代的來臨，各國政府也積極部署各項大數據的應用，以提高政府部門的行政效率，並提供民眾更好的資訊服務。

數位政府服務需讓民眾以最方便的方式、最低的成本來跟政府打交道。因為不知道有創意的人能使用這些資料，取得哪些實用且價值非凡的見解。許多意想不到的應用便是從看似毫無關聯的資料中，交叉分析衍生出的成果。我們應樂觀期待資料應用的無限潛能！

另外，政治決策者要懂得溝通，且以淺顯易懂的方式與民眾溝通，避免擺出官僚的高姿態；並突破因政府制度設計偏重機關內部垂直分工，當需跨部會合作時，容易受法規、權責劃分等的諸多限制。試想，民眾不一定清楚哪些業務依政府制度設計分工是在哪個部會？再者，民眾所需的資料不一定剛好集中在某個部會，可能需搜尋多個部會才能找齊，故資料之跨領運用著實重要，是提升政府效能的重要關卡。畢竟，資料能夠發揮更大的價值是一件好事，收集的資料就是要

拿來用啊！

由於數據科學事涉諸多面向，以下謹提出個人淺見以供參考：

1.設置國家級資料銀行，提供橫跨各級政府資料之整合性深化服務，實現公共價值

(1)對政府的建議：

建議政府規劃設置國家級資料銀行，將資料視為國家資產，以跨政府之經濟規模，建立資料資產註冊機制，集中管理各級政府所有資料，包括能帶動經濟及具商業價值的資料、也包括能監督政府及參與政府治理的資料。於保護個人隱私及機密機制下，分享可再利用的資料；同時結合友善查詢介面，輔以視覺化、地理圖資等方式呈現，利用民間力量進行創意加值運用。至於不適合分享的資料，公務機關或學術研究機構可基於公共利益之必要下連結應用，以有效促進政府機關間資料整合。

近年政府推動資料開放政策，建置政府資料開放平臺，或可做為國家級資料銀行的基礎，惟有幾項值得注意之處：

<1>兼顧數字管理及資料完整性：誠如總處政府資料開放諮詢小組會議上委員所建議的：「考量我國資料開放永續發展，及各機關業務範圍，請評估不宜以資料開放數量作為政策績效指標(KPI)，以利資料整體運用。」

<2>強化民眾查詢介面及分析功能：政府資料開放平臺迄今已納入1萬7,368項資料集，如何在這麼多的資料中，快速找得想要的資料？提供民眾方便的查詢

介面，可作為未來優化的方向之一。另外，輔以簡易的資料分析工具，並以視覺化、地理圖資呈現查詢結果，亦應是未來精進的方向之一。

(2)對總處的建議：

總處研提「普查資訊發展暨數據應用計畫」之「建立民眾數據查詢平臺」即是一個開端。因總處負責政府歲計、會計、統計工作，擁有豐富主計業務資料，為我國政府決策施政之重要參考，亦為專家學者或社會大眾關注之主要目標。且總處掌理全國主計人事管理，主計同仁分佈各級政府機關（構）、公立學校、公營事業機構，可從集中管理全國主計機構之主計資料開始著手，加上提供民眾方便的查詢介面，輔以簡易的資料分析工具，並以視覺化、地理圖資呈現查詢結果，建構國家級的主計資料銀行！

2.面對資料連結應用需求，兼顧擴大資料應用效益及保護敏感資料

大數據分析技術能找到資料間意料外的關係、從資料中挖掘隱含的價值、為存在的問題提供重要的見解、找到創新的解決方案、幫助政府制訂政策、在資訊爆炸的年代快速地對訊息及趨勢產生反應，其中重要方法之一就是不同資料間的連結應用。

但根據直屬美國總統的科學及技術顧問委員會(PCAST)調查結果顯示：高達8成美國受訪民眾非常在意政府如何收集和使用資料。功能強大的大數據分析技術涉及多個資料庫的合併，有些匿名的身分可能因此被識別出來，更多民眾的私

密個資可能被揭露。因此，大數據分析必須加入隱私設計，以保護敏感資料不外洩。

面對資料連結應用需求，擴大資料應用效益、保護敏感資料不外洩之間如何取得平衡？

(1)對組織面的建議：

誠如行政院資料開放諮詢小組會議上委員所建議的：「資料去識別化後將失去部分價值，建議可允許外界提出原始資料分析需求，其分析結果經去識別化後提供，以發揮資料之效益。」，為逐步實踐此一建議，可參採總處「提供普(抽)查資料管制作業要點」做法，略以：「…對去除個別識別碼或將其亂碼後之資料內容仍有涉及隱私之虞或具敏感性 ，須於指定資料處理場地臨場作業，並經許可後連結其他檔案，攜出資料僅限彙總之統計結果。…」，擴大至各部會；至於「指定資料處理場地」，因須符合一定資訊安全規範，部會可建立共用機制資源分享，以撙節公帑。

(2)對法制面的建議：

至於公部門之資料連結應用需求，實有法源依據。根據「個人資料保護法」第 5 條規定：「個人資料之蒐集、處理或利用…不得逾越特定目的之必要範圍，並應與蒐集之目的具有正當合理之關聯。」又第 16 條規定：「公務機關對個人資料之利用...，應於執行法定職務必要範圍內為之，並與蒐集之特定目的相符。但有下列情形之一者，得為特定目的外之利用：…二、為…增進公共利益所必

要。…五、公務機關或學術研究機構基於公共利益為統計或學術研究而有必要，且資料經過提供者處理後或經蒐集者依其揭露方式無從識別特定之當事人。」

惟為求周妥，避免公務同仁因業務需要觸法，可安排適當機關，負責公部門提出前揭適法性之解釋。

3.重視網路輿情，促進民眾有感施政

近年來，隨著民主深化發展，「傾聽民意、回應民意」已成為民主政府必要的功課。受到民眾使用網路能力提升，及各式連網設備普及的影響，電子媒體及社群網站等網路民意資料正大量、迅速及多元地蓬勃發展，使得過去政府部門熟悉的輿情分析方法已無法有效分析這些迅速積累的非結構化巨量資料。

為體察民意，建議未來可發展為由民間公司收集電子媒體、臉書、PTT 等社群網站的爬網資料，政府機關以訂閱關鍵字的方式採購。民間公司專注於提高網路民意的可信度或掌握度，政府機關則適當調整網路民意及傳統大眾媒體的重視程度，並專注於運用網路民意導入政府決策分析，實踐良善治理、加強施政績效品質。

4.運用大數據分析技術，將統計業務推上更高層次

過去統計調查收集的資料是部分的、有關連的、資料結構是被定義好的、資料分析會有延遲；現在的資料是量大、多面向且快速變動。因此，大數據與統計

學有許多的觀念不同，如：

(1)不需抽樣：因為數據取得較過去相對容易，樣本幾乎等於母體，所以不需抽樣。

(2)可處理非結構化資料，並容忍較差的資料品質：因為過去數據取得困難，要從有限的樣本推導出有效結果，資料必須結構化且品質必須非常高。現因技術進步，可以處理非結構化資料，又因資料量大，少數較差品質的資料對結果不致造成太大的影響。

(3)統計學是提出假設，再透過數據來驗證；大數據是透過程式自動建立大量的機器假設，將所有可能的假設通通放進來，再利用雲端運算一次處理高達數億個機器假設，從大量的機器假設中找出可能超過人類概念範圍，但卻最合理的相關性。

因此，統計同仁如能運用大數據分析技術，實可克服抽樣的限制、非結構化資料、較差的資料品質及假設的限制，再將統計業務推上更高層次！

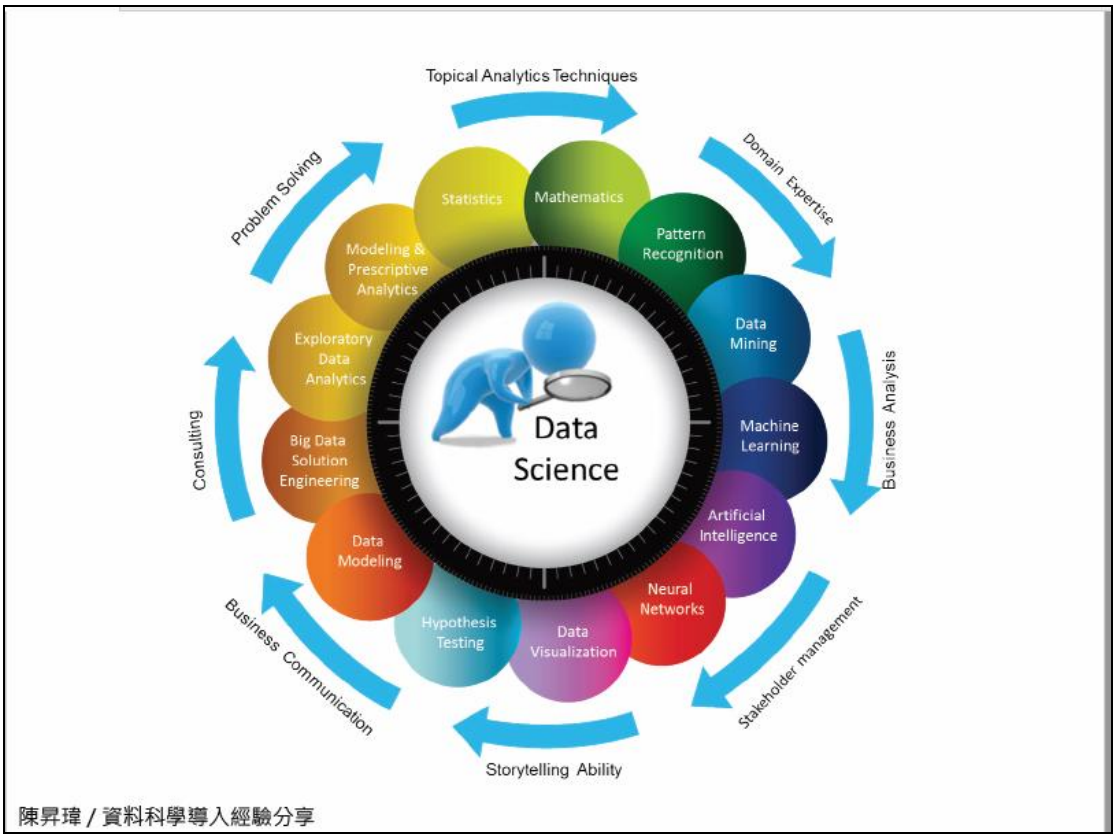
5.大數據不是魔法杖，靈活混搭(mashup)開創新局

大數據可幫助政府掌握社會脈動，提供未來遵循方向，但如何因應還是掌握在政府手中，強大的分析工具可能造成差別待遇：越來越多商業及個人生活上的應用透過大數據演算法和自動化的流程來決定，這將可能產生不公平的偏見，所以不可過度依賴預測資料。爰大數據分析方法的導入可考慮先與現行的方法並

行，比較兩者的結果，檢討發生誤差的原因，並加以修正。待應用較成熟後，倚靠大數據判釋的比例可以再增加。

6.以總處「大數據應用與發展推動分組」為基礎，逐步結合主計機構人才，型塑資料專家團隊

我們可以從中央研究院陳昇瑋博士蒞臨總處演講的簡報資料(如下圖)看出:



資料科學綜整應用了多門學科，如：數學、統計學（機率、高等運算、建模）、資訊工程（人工智慧、機器學習）、大數據等，要組成具備資料管理、資料分析、政策洞見等核心能力的優秀資料專家團隊，具備良好的數字、統計、資訊背景更是勝人一籌。

主計體系同仁大多通過會計、統計、資訊職系國家考試，又分佈各級政府機關（構）、公立學校、公營事業機構，具備優秀的專業能力，熟悉數字及技術，又接近業務，較其他部門更具備發展大數據分析的優越條件。

因為資料科學涉及領域廣泛，無法人人同時精通所有技術，因此需要透過團隊合作的模式，讓業務領域專家、資料科學工程師、統計學家在其專業領域各展所長，又在團隊的交流中涉獵其他領域的相關知識，彼此成長、互補。

建議以總處「大數據應用與發展推動分組」為基礎，先從小部分、明確、合理範圍的需求出發，一邊累積經驗，精進系統與團隊，建立數據應用之專業及信譽，一邊將成果複製至其他主計機構，並逐步結合主計機構人才，型塑資料專家團隊，再造主計資料創新價值。

五、結語

為應用大數據分析，本總處 105 年辦理「主計資料之大數據分析案」研究計畫，從專案執行中對各業務處主管訪談時可以發現：大數據分析或可先局部應用在總處的某些業管領域，輔助如：公務預算處的人事費及保費額度匡列、國勢普查處的常住人口推估、綜合統計處的物價統計等。本總處數據分析發展規劃可朝先從局部應用開始，再逐步調校修正，未來並加強向外擴大方向辦理。

未來，智慧城市及物聯網時代來臨，即時分析大量感知器資料，以提供滿足商業行為及公民需求的適性服務應用將越來越多，這都將挑戰運算資料量及即時性。再者，僅靠政府的力量是不夠的，政府希望能邀請產業、私人企業、學術單位、非營利組織及民間團體一起跟上腳步，建立政府機關、產業界及學術界數據分析成果的分享機制，促進組織的合作關係，如此將有助於強化公私部門兩方的大數據分析技術及能力，激發更多的創新應用，抓住大數據帶來的機會！

在這資料經濟的時代，統計機關不只是製造資料庫的單位，如何由資料庫中找到寶藏，做為支援決策及檢驗政策的方針，應為努力的方向。猶記長官期勉大家：努力工作產製有效資料服務他人之餘，也要思考如何運用他人產製的資料來增進自己的工作效率。請大家一起「資料再利用、重組資料、找出資料的多種用途」，釋放潛藏的資料新價值、幫助我們更瞭解世界、改變決策及行為、化資料為力量！

參考資料

1. 2016 年全球數據科學研討會官網：http://globalbigdataconference.com/67/santa-clara/global-data-science-conference/event.html
2. COMCAST：http://dclabs.comcast.com/
3. Leveraging Search Algorithms for Bing Predicts：http://blogs.bing.com/search-quality-insights/2015/03/15/leveraging-search-algorithms-for-bing-predicts?FORM=MA133A&OCID=MA133A&wt.mc_id=MA133A
4. Numenta Anomaly Benchmark (NAB)：http://numenta.com/numenta-anomaly-benchmark/
5. 陳昇瑋「資料科學的第一堂課－理論、案例及企業導入方法」研討會資料
6. 運用巨量資料實踐良善治理:網路民意導入政府決策分析之可行性研究計畫書
7. 大數據，遠見天下文化
8. E 政府，電週文化
9. 維基百科

附錄

附錄一、研討會門票

4953641776248993940 01

Event

Global Data Science Conference Santa Clara March 2016

Date+Time

Monday, March 7, 2016 at 8:00 AM - Wednesday, March 9, 2016 at 8:00 PM (PST)

Location

Santa Clara Convention Center
5001 Great America Parkway
Santa Clara, CA 95054

Name

CHOU TSAI-JUNG

Payment Status

Eventbrite Completed

Order Info

Order #495364177. Ordered by CHOU TSAI-JUNG on February 26, 2016 8:08 AM

Type

Global Data Science Conference Santa Clara- 3 Days $1,299.00

4953641776248993940 01

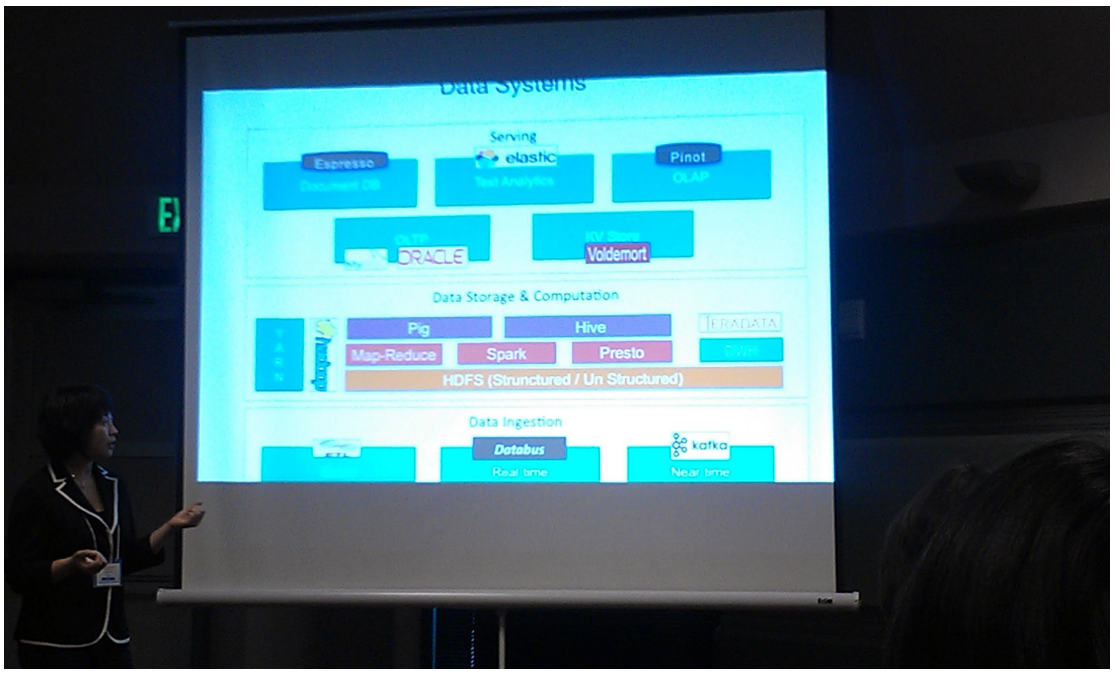
Eventbrite

Do you organize events?

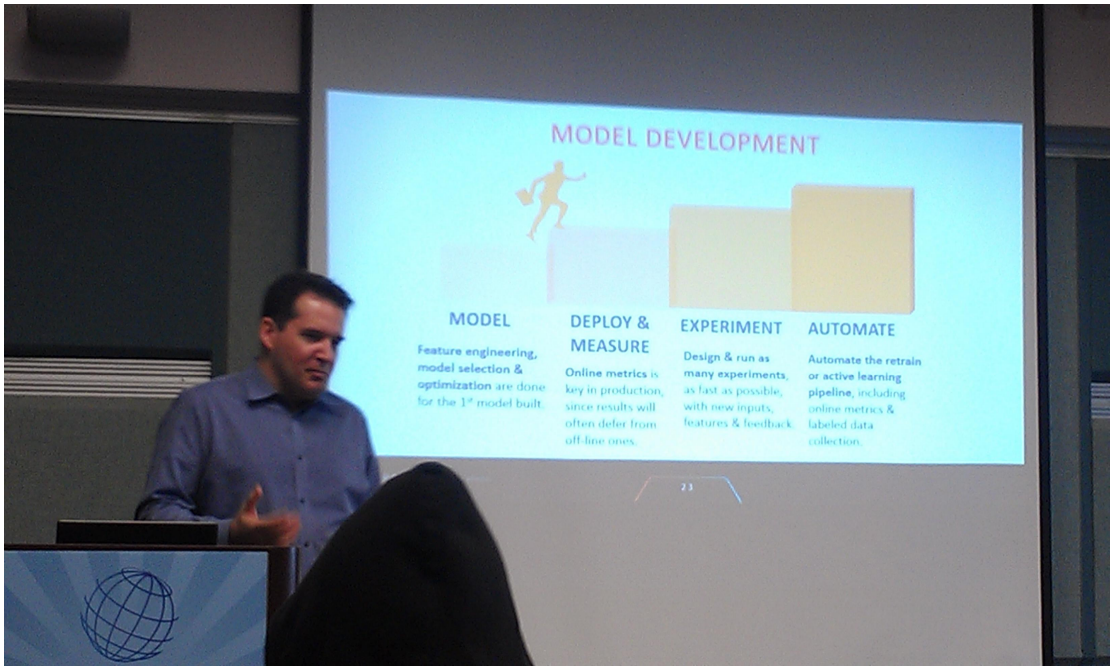
Start selling in minutes with Eventbrite!
www.eventbrite.com

附錄二、研討會相關照片

會場入口



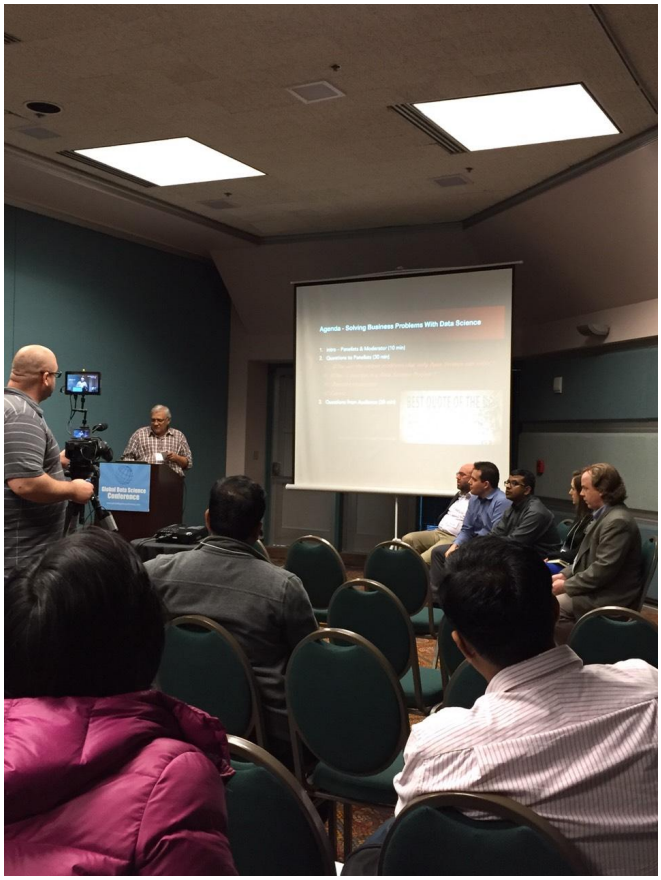
Session: Create impact at scale by data-driven applications



Session: Hunting Criminals with Hybrid Analytics

Session: Predictive Analytics, Inventory Management, Machine Learning, Ecommerce

Session: User behavioral predictive analytics through deep learning-based emotion recognition



Session: Solving Business Problems With Data Science