Big Data ?

數據分析與展現

美華集團 美商傑明工程顧問(股)台灣分公司

Stantec

Grand Valley State University - Mary Idema Pew Library

有沒有聽過「Big Data」?

大約4、5年前(甚至更早)開始出現這個名詞。

有沒有發現近2年在工作上，你的長官開始會請你想想怎麼幫他們做Big Data的統計分析?

然後，你的心裡出現一堆問號???
看著長官殷殷期盼的臉，你的表情開始微笑.....

Stantec

到底什麼是「Big Data」?

從Wiki搜尋一下：

「巨量資料(Big data)，指的是傳統資料處理應用軟體不足以處理它們的大或複雜的資料集的術語。在總資料量相同的情況下，與個別分析獨立的小型資料集(Data set)相比，將各個小型資料集合併後進行分析可得出許多額外的資訊和資料關聯性，可用來察覺商業趨勢、判定研究品質、避免疾病擴散、打擊犯罪或測定即時交通路況等；這樣的用途正是大型資料集盛行的原因。截至2012年，技術上可在合理時間內分析處理的資料集大小單位為艾位元組(exabytes)。在許多領域，由於資料集過度龐大，科學家經常在分析處理上遭遇限制和阻礙；這些領域包括氣象學、基因組學、神經網路體學、複雜的物理模擬，以及生物和環境研究。這樣的限制也對網路搜尋、金融與經濟資訊學造成影響。資料集大小增長的部分原因來自於資訊持續從各種來源被廣泛收集，這些來源包括搭載感測裝置的行動裝置、高空感測科技(遙感)、軟體記錄、相機、麥克風、無線射頻辨識(RFID)和無線感測網路。自1980年代起，現代科技可儲存資料的容量每40個月即增加一倍；截至2012年，全世界每天產生2.5艾位元組(2.5×1018位元組)的資料.....ooxx.....」

講到這裡，你的長官已經開始睡著了.....

Stantec

簡單一點的描述呢?

關於Big Data聽過很貼切的一句話：

大家都在講大數據

沒人知道到底怎麼做

每個人都以為別人正在做

所以每個人都說自己也在做

Stantec

那Big Data是個假議題嗎?

其實不是假議題，舉工作實例來說，你會發現當我們要做計劃中所遇到的數據分析時，常常不是原始資料不足(例如：監測數據/水理條件/管線屬性.....等)、資料品質不佳、不然就是有資料但是資料跟資料間的關聯性不足，以至於無法繼續分析。

再加上數據分析需要一些工具輔助，而這些工具一般來說主要是資訊軟體、模式，因為其專業性與專業領域的不同，常常造成使用者入門的門檻過高。

這樣情形是否似曾相識? 恭喜你，你已經在做Data分析了，至於Big不Big，那是下一步的尺度問題了。

Stantec
Stantec

是否有好用的工具可以幫助分析過程簡化呢?

市場上其實過去一直有所謂的「自助化分析工具」，早期比較常用在財務或銷售等數字型的分析上，這類自助化分析工具在市場上稱為「商業智慧分析」簡稱BI (Business Intelligence)

tantec
Stantec

BI 是...

為了將資料轉化為知識，需要利用資料庫、線上分析處理（OLAP）工具和資料挖掘等技術。
因此，從技術層面上講，商業智慧不是什麼新技術，它只是ETL、資料庫、OLAP、資料挖掘、資料展現等技術的綜合運用。

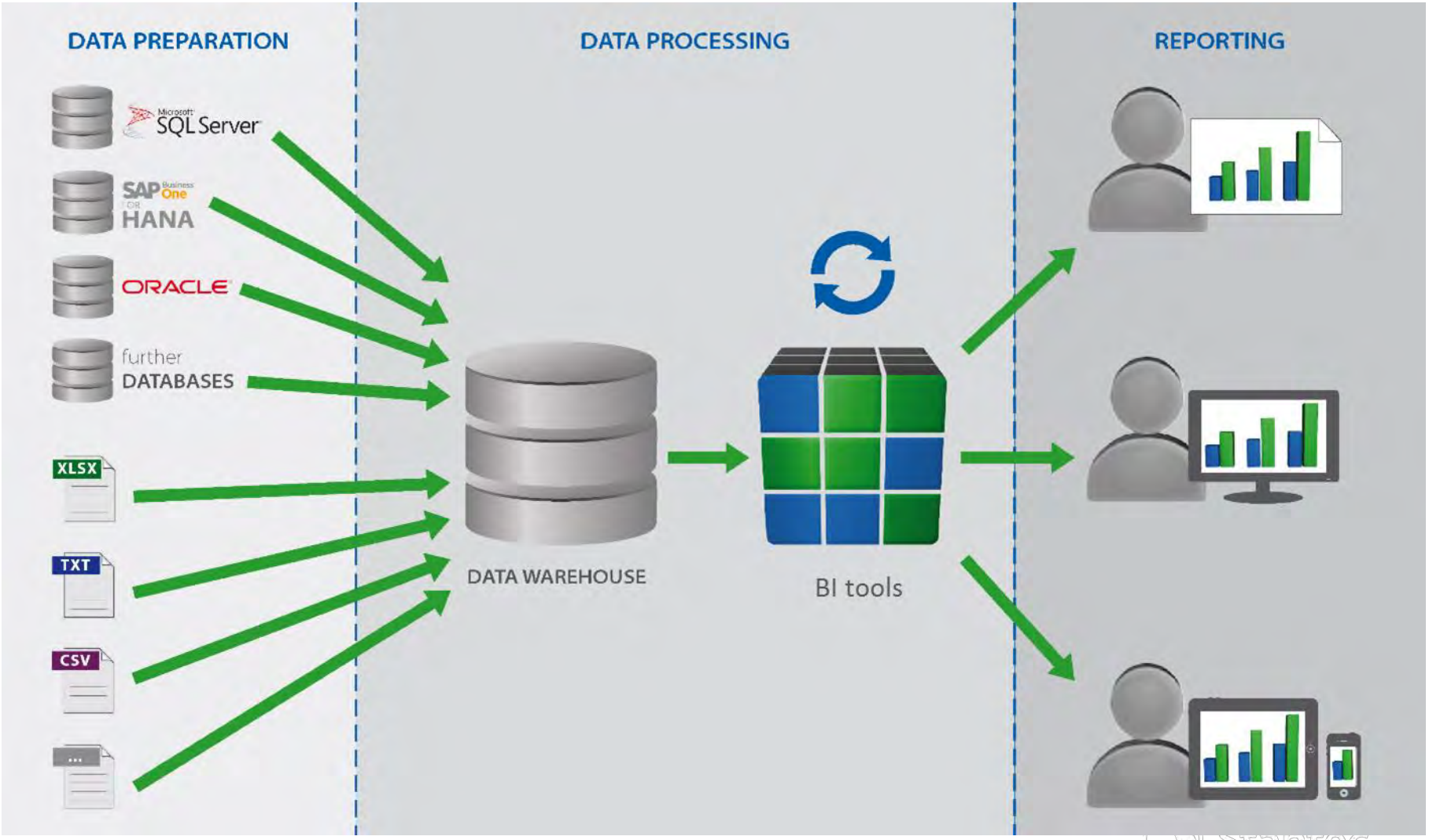
故將商業智慧看成是一種解決方案較為恰當。

商業智慧的關鍵是從許多來自不同運作系統的資料中提取出有用的資料並進行清理，以保證資料的正確性，然後經過抽取（Extraction）、轉換（Transformation）和裝載（Load），

合併到一個資料庫裡，從而得到企業資料的一個全局視圖。
在此基礎上利用合適的查詢和分析工具、資料挖掘工具、OLAP工具等對其進行分析和處理（這時資訊變為輔助決策的知識），最後將知識呈現給管理者，為管理者的決策過程提供支援。

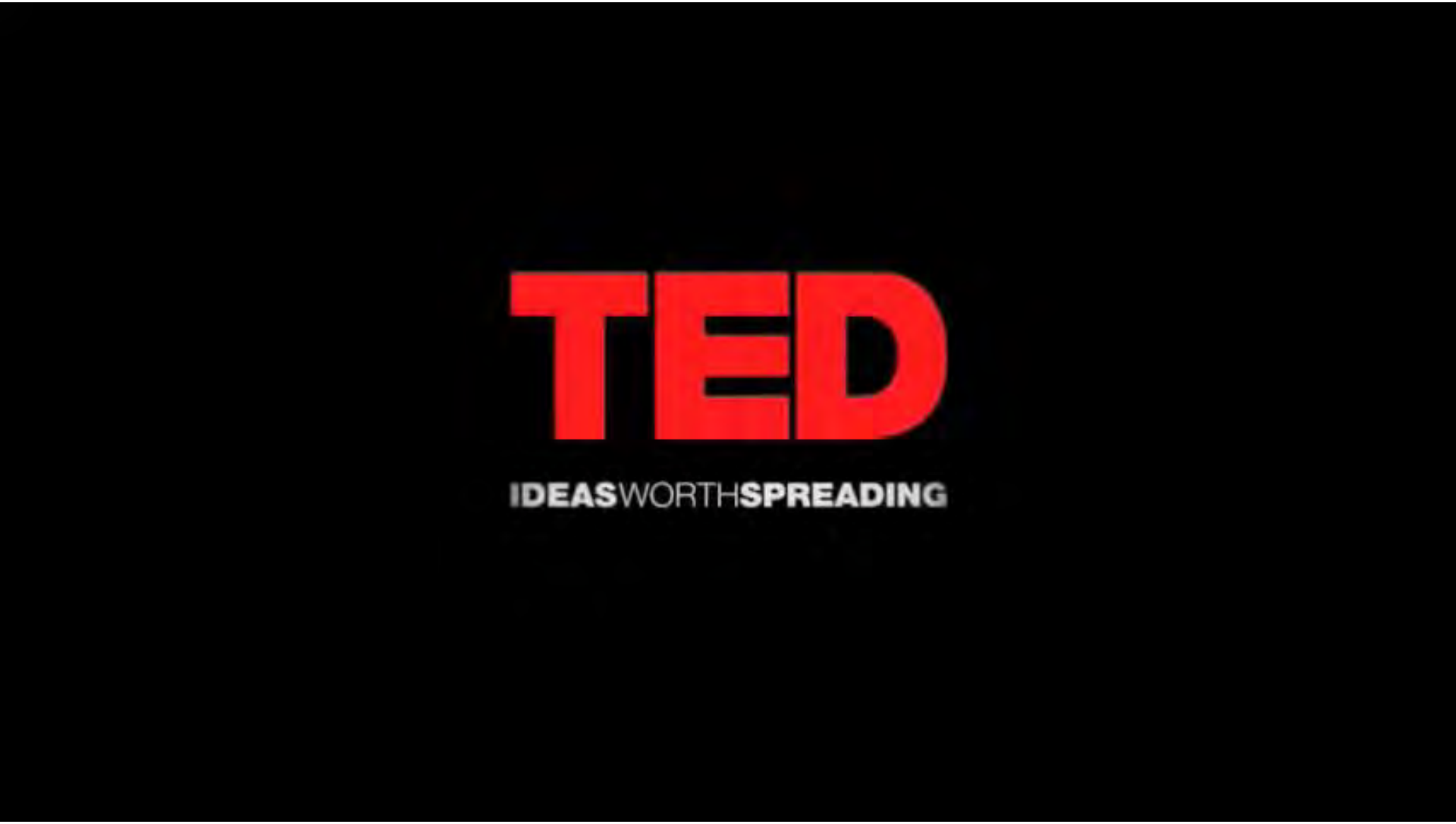
Stantec
Stantec

簡單的說...



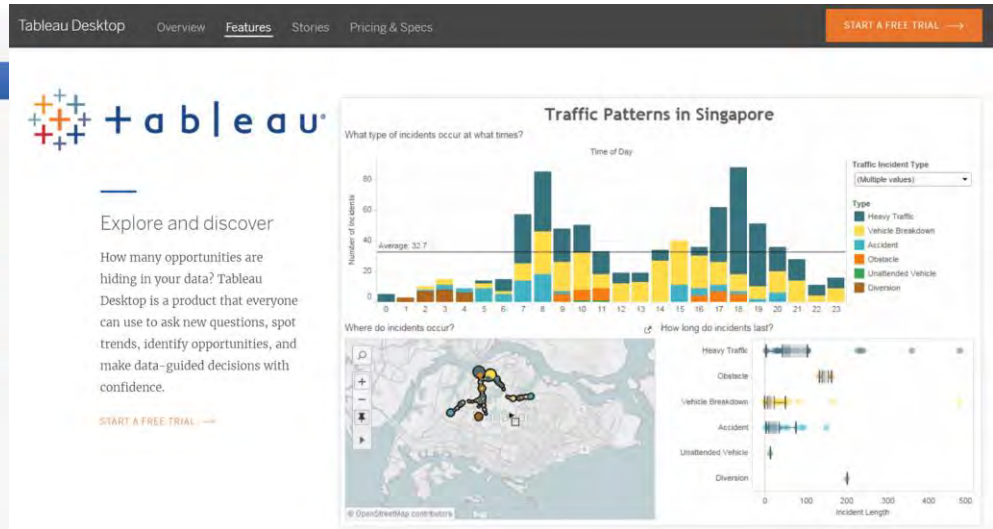
Stantec

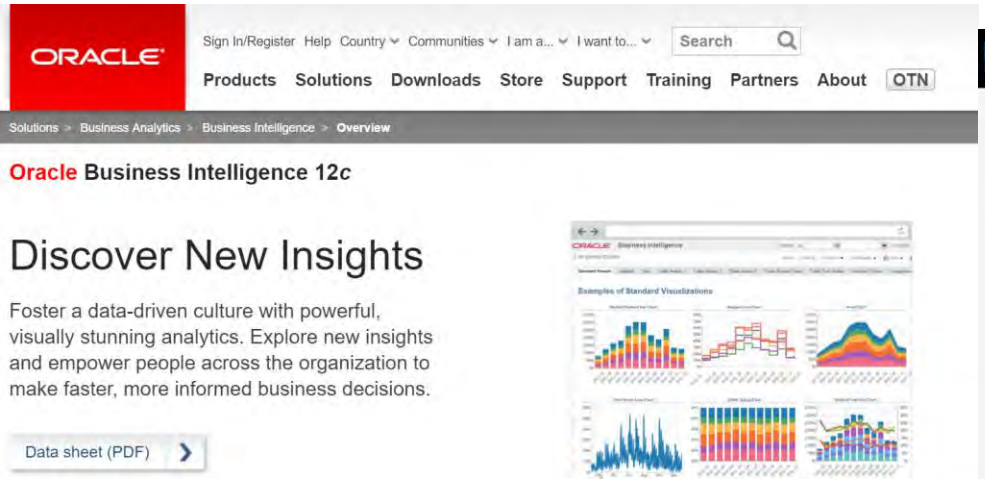
Dr. Hans Rosling – Gapminder 基金會



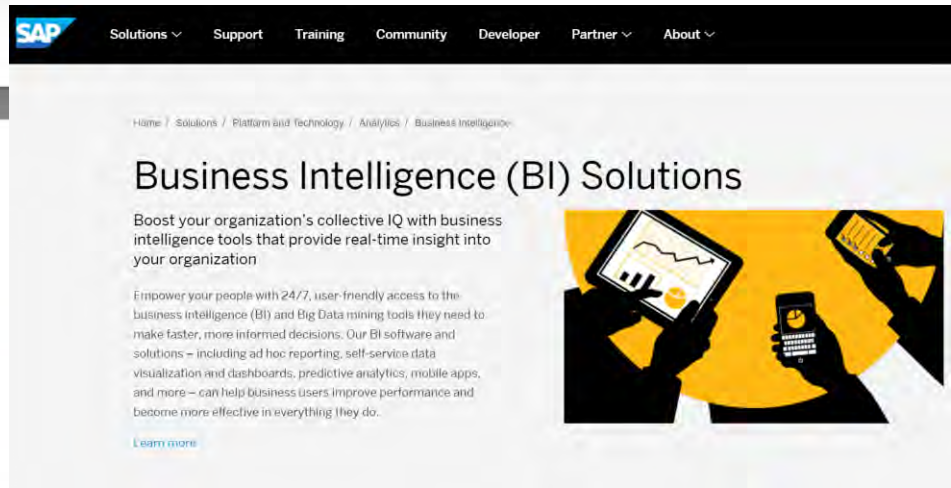
Stantec

常見 BI 工具





Rich library of visual components makes the presentation of data intuitive, relevant, and easy to understand.



Stantec

喔?! 那我也能使用嗎?

有使用過excel嗎? 你是否曾經在excel中撰寫程式來計算加總數字? 我過去也是這樣，一直到開始做數據分析後，才知道，excel不是這樣用的!

看過真正高手的excel使用法後，我才知道原來使用excel只要「拖、拉、點、選」就可以完成數據分析! 也就是，excel工作表示拿來放row data的，要做數據分析乃是使用「樞紐分析」的功能。

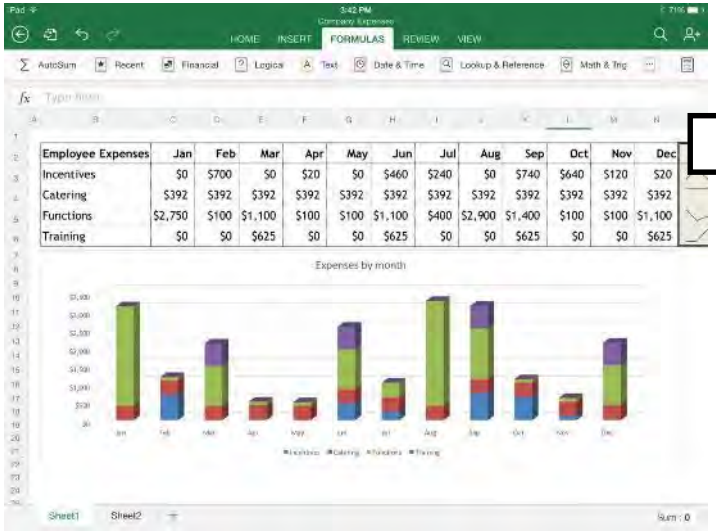
如果你有用過excel和樞紐分析，那我們可以開始來用BI工具了!

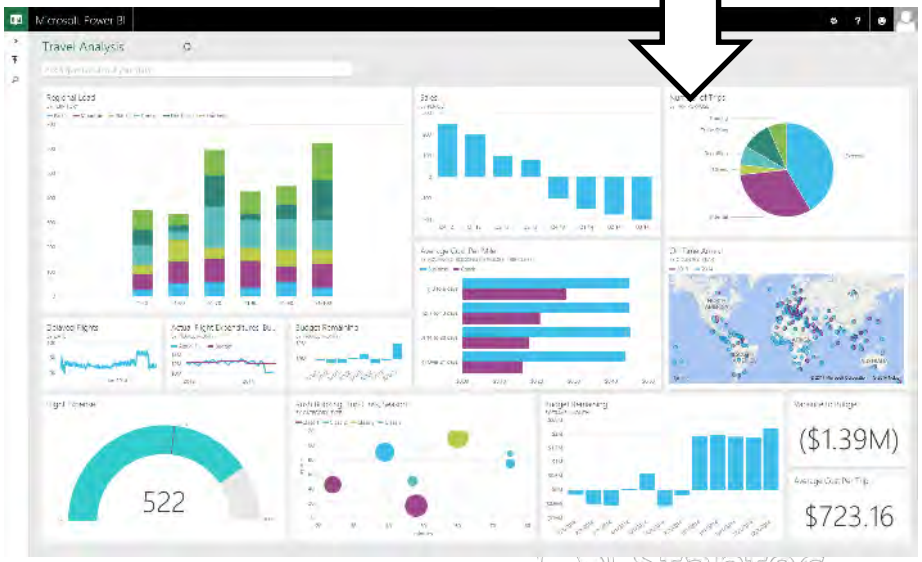
嗯? 沒用過樞紐分析.....沒關係，我們還是繼續.....

Stantec
Stantec

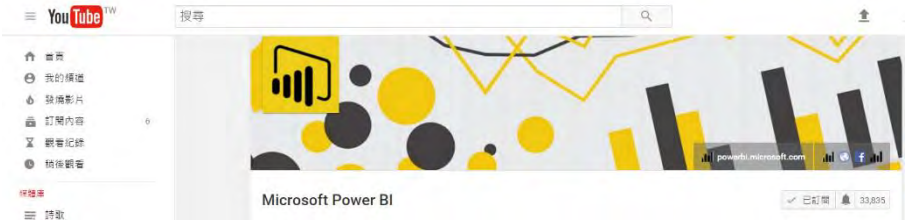
Microsoft Power BI

從 Excel > Pivot > Power BI





https://www.youtube.com/user/mspowerbi



Stantec

我們來試試看!
Power BI 實作

資料清理

匯入檢查

產製報表

60 %

20 %

20 %

Stantec

Stantec