

frilly / March 07, 2020 05:45PM

[七個步驟帶你全面了解數據分析完整流程，建議收藏！](#)

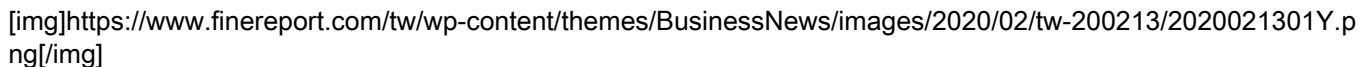
1. [1.1. 資料獲取](#)
2. [2.2. 資料存儲](#)
3. [3.3. 資料提取](#)
4. [4.4. 資料採擷](#)
5. [5.5. 資料分析](#)
6. [6.6. 數據展現](#)
7. [7.7. 資料應用](#)

作為資料分析師，無論最初的職業定位方向是技術還是業務，最終發到一定階段後都會承擔資料管理的角色。因此，一個具有較高層次的[\[url=https://www.finereport.com/tw/\]](https://www.finereport.com/tw/)數據分析師[\[url\]](https://www.finereport.com/tw/) 需要具備完整的知識結構。

那作為一個十年以上的數據從業人，想和你們分享數據分析的流程，主要分七個步驟，遵循這七種方法，一個完整的[\[url=https://www.finereport.com/tw/\]](https://www.finereport.com/tw/)數據分析[\[url\]](https://www.finereport.com/tw/)專案就出來了。

1. 1. 資料獲取

瞭解資料獲取的意義在於真正瞭解資料的原始面貌，包括資料產生的時間、條件、格式、內容、長度、限制條件等。這會說明資料分析師更有針對性的控制資料生產和採集過程，避免由於違反資料獲取規則導致的資料問題；同時，對資料獲取邏輯的認識增加了資料分析師對資料的理解程度，尤其是資料中的異常變化。



比如：

Omniure中的Prop變數長度只有100個字元，在資料獲取部署過程中就不能把含有大量中文描述的文字賦值給Prop變數（超過的字元會被截斷）。

在Webtrekk323之前的Pixel版本，單條資訊預設最多只能發送不超過2K的資料。當頁面含有過多變數或變數長度有超出限定的情況下，在保持資料收集的需求下，通常的解決方案是採用多個sendinfo方法分條發送；而在325之後的Pixel版本，單條資訊預設最多可以發送7K資料量，非常方便的解決了代碼部署中單條資訊超載的問題。（Webtrekk基於請求量付費，請求量越少，費用越低）。

當使用者在離線狀態下使用APP時，資料由於無法聯網而發出，導致正常時間內的資料統計分析延遲。直到該設備下次聯網時，資料才能被發出並歸入當時的時間。這就產生了不同時間看相同歷史時間的資料時會發生資料有出入。

在資料獲取階段，資料分析師需要更多的瞭解資料生產和採集過程中的異常情況，如此才能更好的追本溯源。另外，這也能很大程度上避免“垃圾資料進導致垃圾資料出”的問題。

2. 2. 資料存儲

無論資料存儲於雲端還是本地，資料的存儲不只是我們看到的資料庫那麼簡單。

比如：

資料存儲系統是MySql、Oracle、SQL Server還是其他系統。

資料倉庫結構及各庫表如何關聯，星型、雪花型還是其他。

生產資料庫接收資料時是否有一定規則，比如只接收特定類型欄位。

生產資料庫面對異常值如何處理，強制轉換、留空還是返回錯誤。

生產資料庫及資料倉庫系統如何存儲資料，名稱、含義、類型、長度、精度、是否可為空、是否唯一、字元編碼、約束條件規則是什麼。

接觸到的資料是原始資料還是ETL後的資料，ETL規則是什麼。

資料倉庫資料的更新更新機制是什麼，全量更新還是增量更新。

不同資料庫和庫表之間的同步規則是什麼，哪些因素會造成資料差異，如何處理差異的。

在資料存儲階段，資料分析師需要瞭解資料存儲內部的工作機制和流程，最核心的因素是在原始資料基礎上經過哪些加工處理，最後得到了怎樣的資料。由於資料在存儲階段是不斷動態變化和反覆運算更新的，其及時性、完整性、有效性、一致性、準確性很多時候由於軟硬體、內外部環境問題無法保證，這些都會導致後期資料應用問題。

3. 3.資料提取

資料提取是將資料取出的過程，資料提取的核心環節是從哪取、何時取、如何取。

從哪取，資料來源——不同的資料來源得到的資料結果未必一致。

何時取，提取時間——不同時間取出來的資料結果未必一致。

如何取，提取規則——不同提取規則下的資料結果很難一致。

[img]<https://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2020/02/tw-200213/2020021302Y.jpg>
g[/img]

在資料提取階段，資料分析師首先需要具備資料提取能力。常用的Select From語句是SQL查詢和提取的必備技能，但即使是簡單的取數工作也有不同層次。第一層是從單張資料庫中按條件提取資料的能力，where是基本的條件陳述式；第二層是掌握跨庫表提取資料的能力，不同的join有不同的用法；第三層是優化SQL語句，通過優化嵌套、篩選的邏輯層次和遍歷次數等，減少個人時間浪費和系統資源消耗。

其次是理解業務需求的能力，比如業務需要“銷售額”這個欄位，相關欄位至少有產品銷售額和產品訂單金額，其中的差別在於是否含優惠券、運費等折扣和費用。包含該因素即是訂單金額，否則就是產品單價×數量的產品銷售額。

4. 4.資料採擷

資料採擷是面對海量資料時進行資料價值提煉的關鍵，以下是演算法選擇的基本原則：

沒有最好的演算法，只有最適合的演算法，演算法選擇的原則是兼具準確性、可操作性、可理解性、可應用性。

沒有一種演算法能解決所有問題，但精通一門演算法可以解決很多問題。

挖掘演算法最難的是演算法調優，同一種演算法在不同場景下的參數設定相同，實踐是獲得調優經驗的重要途徑。

在資料採擷階段，資料分析師要掌握資料採擷相關能力。一是資料採擷、統計學、數學基本原理和常識；二是熟練使用一門資料採擷工具，Clementine、SAS或R都是可選項，如果是程式出身也可以選擇程式設計實現；三是需要瞭解常用的資料採擷演算法以及每種演算法的應用場景和優劣差異點。

5. 5.資料分析

資料分析相對於資料採擷更多的是偏向業務應用和解釋，當資料採擷演算法得出結論後，如何解釋演算法在結果、可信度、顯著程度等方面對於業務的實際意義，如何將挖掘結果回饋到業務操作過程中便於業務理解和實施是關鍵。

[img]https://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2020/02/tw-200213/2020021303Y.png[/img]

6. 6. 數據展現

資料展現即資料視覺化的部分，資料分析師如何把資料觀點展示給業務的過程。資料展現除遵循各公司統一規範原則外，具體形式還要根據實際需求和場景而定。

基本素質要求如下：

工具:FineReport是不錯的[url=https://www.finereport.com/tw/products/datavisualization]資料視覺化[/url]工具。

形式:圖文並茂的基本原則更易於理解，生動、有趣、互動、講故事都是加分項。

原則:領導層喜歡讀圖、看趨勢、要結論，執行層歡看數、讀文字、看過程。

場景:大型會議Power point最合適，彙報說明Word最實用，資料較多時Excel更方便。

最重要一點，資料展現永遠輔助於資料內容，有價值的資料包告才是關鍵。

[img]https://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2020/02/tw-200213/2020021304Y.gif[/img]

7. 7. 資料應用

資料應用是資料具有落地價值的直接體現，這個過程需要資料分析師具備資料溝通能力、業務推動能力和專案工作能力。

資料溝通能力。深入淺出的資料報告、言簡意賅的資料結論更利於業務理解和接受，打比方、舉例子都是非常實用的技巧。

業務推動能力。在業務理解資料的基礎上，推動業務落地實現資料建議。從業務最重要、最緊急、最能產生效果的環節開始是個好方法，同時要考慮到業務落地的客觀環境，即好的資料結論需要具備客觀落地條件。

專案工作能力。資料項目工作是循序漸進的過程，無論是一個資料分析專案還是資料產品專案，都需要資料分析師具備計畫、領導、組織、控制的專案工作能力。

感謝閱讀！FineReport提供最全免費功能版本，不用等待，直接點擊以下按鈕激活&下載！

[url=https://www.finereport.com/tw/products/trial]免費試用FineReport10.0>[/url]

獲得帆軟最新動態：數據分析，報表實例，專業的人都在這裡！加入[url=https://www.facebook.com/twfinereport/]FineReport臉書粉絲團[/url]

相關文章：

[url=https://www.finereport.com/tw/data-analysis/analytical-thinking.html]【值得收藏】數據分析師必備的分析思維方法，你知道多少？[/url]

[url=https://www.finereport.com/tw/data-analysis/2020-data-analysis-tools.html]這大概是2020年最值得推薦的【大數據分析】工具！[/url]

[url=https://www.finereport.com/tw/success/solution/bi-and-data-visualization-screen.html]BI+資料視覺化大屏,製造業可以這樣進行數據分析[/url]