

frilly / July 27, 2018 11:49AM

## 數據分析師需要掌握哪方面的電腦技能？——從基礎到進階

最近，在網路上看到一個問題：「數據分析師需要注重哪方面的電腦技能培養？」

問題的背景是這樣的：

某高校通信專業出身，畢業後在運營商工作了7年多，先後從事通信網路運維、規劃工作，近兩年負責運營商數據分析（網路部門，偏向業務分析）。

由於職業發展瓶頸，從去年11月計劃跳槽，花了半年時間學習統計學基礎、SQL、Python等。

近期跳槽到互聯網產品部門，從事互聯網產品（APP）的數據分析師，支撐產品部門的數據分析（偏向業務分析，不負責數據倉庫、ETL等偏向IT工作）。工作內容差異較大，包括分析的顆粒度、工作方式（例如自己寫shell腳本跑數）、工作內容，因此緊急提升linux（shell編程）、SQL等技能，且加快對業務的熟悉，但仍感覺亞歷山大。

個人想繼續往數據分析方向發展，也深知數據分析包括電腦科學、統計學、業務等三個部分內容，目前比較欠缺的應該是電腦科學，請問對於想往數據分析師（數據科學）方向發展，電腦科學方面的技能能否給些提升建議？

今天也是想借這個問題，系統回答下「數據分析師」的職業發展，也是最近在思考的。

根據我近10年的工作經驗，包括在甲方IT部任職**商業智慧BI**專案經理和運營部任職業務分析經理，乙方Data Analytics專案（EDW/BI/Big Data/AI Machine Learning）諮詢和專案實施經驗，按照由易到難的進階步驟，我覺得應該掌握這些技能：

### 基礎篇

#### 1、首先是Excel

貌似這個很簡單，其實未必。Excel不僅能夠做簡單二維表、複雜嵌套表，能畫折線圖/Column chart/Bar chart/Area chart/餅圖/雷達圖/Combo char/散點圖/Win Loss圖等，而且能實現更高級的功能，包括透視表（類似於商業智慧BI的多維分析模型Cube），以及Vlookup等複雜函數，處理100萬條以內的數據沒有大問題。最後，很多更高級的工具都有Excel插件，例如一些AI Machine Learning的開發工具。

#### 2、掌握SQL Server或者Oracle的SQL語句

雖然你是業務分析師，但如果取數據能少依賴於IT資訊人員和IT工具（比如BI的多維分析模型，有時候並不能獲取你想要的數據），對於做業務分析，無疑是如虎添翼，我曾經見過華為的會計能寫七層嵌套的SQL語句，很吃驚。包括join, group by, order by, distinct, sum, count, average, 各種統計函數等。

#### 3、掌握可視化工具

比如商業智慧BI，如Cognos/Tableau/FineBI等，具體看企業用什麼工具。這些工具做可視化非常方便，特別是分析報告能含這些圖，一定會吸引高層領導的眼球，一目了然了解，洞察業務的本質。另外，作為專業的分析師，用多維分析模型Cube能夠方便地自定義**報表**，效率大大提升。

總結：至此，掌握以上技能的80%，可以算是一個合格的分析師了。這個階段的數據分析師，需要既懂得如何利用工具處理數據，也要懂得業務場景，能分析解決基本的問題。這裡還是要強調一點，數據分析師最重要的是熟悉業務，最好是懂。懂業務，分析邏輯就會清晰一般，而且也能排除大部分無用的嘗試。長期以往對於了解的業務，比對一下數據就知道問題出在哪裡了。

之後，如果是要深鑽技術，甚至往數據科學家方向上發展。

## 進階篇

### 1、系統的學好統計學

純粹的機器學習講究演算法預測能力和實現，但是統計一直就強調「可解釋性」。比如說，針對今天微博股票發行就上升20%，你把你的兩個預測股票上漲還是下跌的model套在新浪的例子，然後給你的上司看。統計學就是這樣的作用。

數據探勘相關的統計方法（多元Logistic回歸分析、非線性回歸分析、判別分析等）

定量方法（時間軸分析、概率模型、優化）

決策分析（多目的決策分析、決策樹、影響圖、敏感性分析）

樹立競爭優勢的分析（通過項目和成功案例學習基本的分析理念）

資料庫入門（數據模型、資料庫設計）

預測分析（時間軸分析、主成分分析、非參數回歸、統計流程控制）

數據管理（ETL（Extract、Transform、Load）、數據治理、管理責任、元數據）

優化與啟發（整數計劃法、非線性計劃法、局部探索法、超啟發（模擬退火、遺傳演算法））

[大數據分析](#)（非結構化數據概念的學習、MapReduce技術、大數據分析方法）

數據探勘（聚類（k-means法、分割法）、關聯性規則、因子分析、存活時間分析）

其他，以下任選兩門（社交網路、文本分析、Web分析、財務分析、服務業中的分析、能源、健康醫療、供應鏈管理、綜合營銷溝通中的概率模型）

風險分析與運營分析的電腦模擬

軟體層面的分析學（組織層面的分析課題、IT與業務用戶、變革管理、數據課題、結果的展現與傳達方法）

### 2、掌握AI Machine Learning演算法，會用工具（比如Python/R）進行建模。

傳統的商業智慧BI分析能回答過去發生了什麼？現在正在發生什麼？但對於未來會發生什麼？必須靠演算法。雖然像Tableau、FineBI等自助式BI已經內置了一部分分析模型，但是分析師想要更全面更深度的探索，需要像Python/R的數據探勘工具。另外大數據之間隱藏的關係，靠傳統工具人工分析是不可能做到的，這時候交由演算法去實現，無疑會有更多的驚喜。

其中，面向統計分析的開源編程語言及其運行環境「R」備受矚目。R的強項不僅在於其包含了豐富的統計分析庫，而且具備將結果進行可視化的高品質圖表生成功能，並可以通過簡單的命令來運行。此外，它還具備稱為CRAN（The Comprehensive R Archive Network）的包擴展機制，通過導入擴展包就可以使用標準狀態下所不支持的函數和數據集。R語言雖然功能強大，但是學習曲線較為陡峭，個人建議從python入手，擁有豐富的statistical libraries，NumPy，SciPy.org，Python Data Analysis Library，matplotlib: python plotting。

最後，怎麼說呢，無論何時業務分析、[數據分析](#)

還是數據科學教，他的價值體現還是在於商業價值。數據人才到最後的發展也一定是要往企業運營VP、戰略參謀者身居。比如，數據戰略家可以使用IT知識和經驗來制定商業決策，數據科學家可以結合對專業知識的深入理解使用IT技術開發複雜的模型和演算法，分析顧問可以結合實際的業務知識與分析經驗聚焦下一個行業爆點。

所以需要你具備溝通、組織、管理能力和商業思維，這就不只局限於某個崗位了，需要你站在更高位的角度思考，為企業謀福利。同時也要思考，拿著「數據分析」這張牌，如何在公司發揮價值，用數據驅動企業運營，這是要思考的。

轉載自臉書專頁【[FineReport報表軟體](#)】

---