

未來一年四大資料分析趨勢

文 / Thor Olavsrud 譯 / 雲翻譯

隨著企業逐漸轉型為資料導向，資料技術和策略需要開始提供價值。以下是未來幾個月值得觀察的四項資料分析趨勢。

分析和社交、移動通訊和雲端等相關資料技術，已成為數位時代顛覆企業模式的核心角色。隨著各公司在 2017 年開始從資料生成轉向資料導向的組織，資料和分析成為許多企業的業務重心。在 2018 年，這些技術需要開始提供價值。以下是將在未來一年推動資料分析策略的方式、角色和要點。

資料湖需要展示其商業價值，否則將會消亡

多年來，資料在企業中以驚人的速度積累著。在資料源從網路移動到行動通訊、再到機器的時代，物聯網（IoT）只會加速資料的產生。

即時串流資料平台供應商 DataTorrent 執行長 Guy Churchward 說：「這使得以划算的方式擴展資料管線成為迫切的需求。」

對於許多仰賴如 Apache Hadoop 這類技術的企業而言，答案就是創建資料湖——涵蓋整個企業的數據管理平台，以原生格式儲

存企業的所有資料。資料湖的用途是，把整個組織的所有應用，從業務分析到資料探勘，都集中在單獨一個資料儲存庫，來解決訊息孤島的問題。資料湖既原始又未經治理，被視為 Big Data 的萬應良藥。

但是，雖然資料湖可以儲存大量數據，要從中獲取可以用來做為決策參考的情報卻非常困難。

「資料湖在『靜止資料』與『批次處理』的時代為各家公司提供了絕佳的貢獻，」Churchward 表示。「在 2015 年，這種架構開始有受到濫用的跡象，但現在它已成為即時資料分析的致命弱點。將流動中的資料停下，然後再行分析，讓各公司處於極大的劣勢。當要在運算能力的極限之內盡快獲得情報並採取行動時，依賴過時事件資料的公司，其可見性，行動和任何可能的即時補救措施都無法顧全大局。在這個領域若是抱持『夠好就可以了』的心態，會造成致命危機。」

Splice Machine 的執行長 Monte Zweben 表示贊同。

Zweben 預測，在 2018 年，「Hadoop 時代的幻滅將全面展開，許多公司會淹沒在資料湖中，對以 Hadoop 為主的運算引擎進行修改產生的複雜度使他們無法獲得投資回報。」

資料目錄專家 Alation 的策略及聯盟副總裁 Ken Hoang 說，為了在 2018 年生存下來，資料湖需要開始證明其商業價值。

Hoang 表示：「在過去幾年中，堆積資料的新地方——資料湖——不少公司已經完成實驗性質部署，現在除非它們證明能夠提供價值，否則將被逐一撤除，一個成功的資料湖有以下特徵：擁有一個企業目錄，將訊息發掘出來，人工智慧和訊息看管匯集在一起，為企業提供新的見解。」

不過，Hoang 並不認為資料湖完全沒救了。他預測，資料湖和其他大型資料樞紐可以通過所謂「超級樞紐」找到新的契機，通過機器學習提供「情境即服務」。

「過去 25 年來，大型資料樞紐的部署（例如資料倉庫、主

資料管理、資料湖、Salesforce 和 ERP) 導致了更多的資料孤島，它們難以理解，缺乏相互關聯或共享機制，」Hoang 說。「樞紐的樞紐將有把跨樞紐的資產相互關聯起來的能力，實現情境即服務。這一點將會進一步推動更多重要且強大的預測性分析，使公司得到更快、更出色的業務成果。」

MapR 的應用程式架構長 Ted Dunning 預測的轉變也很類似：隨著 Big Data 系統成為儲存、存取和營運方面的重心，企業將著眼於建構全球資料網路，可以全面存取來自多個來源的資料，並且在真正的多租戶系統進行運算。

Dunning 認為：「我們將看到越來越多的企業以資料流的角度看待運算，而不僅是處理完後存入資料庫的資料，這些資料流捕捉了關鍵的業務事件並反映了業務結構。統一的資料網路將成為建構這些大規模、以資料流為主的系統基礎。」

Dunning 表示，這些資料網路將支援適用不同情境的各種運算。「現在的新興趨勢是建立一個資料網路，提供流動資料和靜止資料，用在如 Kubernetes 等公司提供的多雲端運算。」

自助服務資料分析專家 Alteryx 的策略長 Langley Eide 表示，資料湖要體現價值，不僅是 IT 領域的問題：業務線 (LOB) 分析師和數

位長 (CDO) 在 2018 年也需承擔責任。

「大多數分析師沒有利用大量的非結構化資源，例如點擊流，物聯網，日誌等充斥在資料湖的資料——主要是因為其非常困難，」Eide 說。「但事實上，如果分析師不使用這些資料，他們其實沒有做好份內的工作。一般認為，作為一種資產，有許多資料湖的表現不佳——人們不知道裡面有什麼，如何存取它，或者如何從中獲取有用的情報。這一點將在 2018 年發生變化，因為有越來越多數位長和企業希望其資料湖具有更高的投資回報率。」

Eide 預測，2018 年，分析師會用如資料編目之類等更多偏向編程的技術與科技來替代如 Excel 和 SQL 等「土法煉鋼」的工具，以從資料中發掘並獲取更多價值。

數位長將進入成熟階段

身為從資料中獲得更精關分析的推手，Eide 也預測數位長的角色將在 2018 年大放異彩。

「資料基本上是一種新的石油，數位長也開始被認為是解決當今企業最重要問題的靈魂人物：從資料中獲取價值，」Eide 說。「數位長掌控的預算通常不到 1000 萬美元，但面臨的最大挑戰和機會就是讓公司的資料資產更貼近業務線用戶，從而使得備受推崇的自助服務得以成真。在 2018 年，能夠在身為領導人以及業務線領域能力之間取得平衡的數位長，最終將獲得更大筆的預算。」

Eide 相信，能使資源、技能和功能在卓越表現和業務線之間迅速轉換的數位長成果將最為豐碩。對此，Eide 說，敏捷平台和方法將是關鍵。



資料編理員(data curator)的崛起?

資料分析創業公司 Dremio 是開源 Apache Arrow 計畫背後的推手，該公司的執行長兼共同創始人 Tomer Shiran 預測，企業將開始對一種新角色有所需求：資料編理員。

Shiran 表示，資料編理員介於資料消費者（使用如 Tableau 和 Python 等工具，回答重要問題的分析師和資料科學家）和資料工程師（使用腳本語言以及 Spark, Hive 和 MapReduce 等在系統間移動和轉換資料的人）之間。要成為成功的資料編理員，必須了解資料的含義以及應用於資料的技術。

「資料編理員必須理解需要由整個組織中各團體一起進行的分析工作，哪些資料集適合這項工作，以及將資料從原始狀態轉換為適合資料消費者工作所需的形態，」Shiran 說。「資料編理員使用自助服務資料平台等系統來加速資料消費者存取重要資料集的整體流程，而不需製作無數的資料副本。」

資料治理策略將成為所有高階主管的關鍵主題

歐盟的一般資料保護條例 (GDPR) 將於 2018 年 5 月 25 日生效，在分析領域上空像幽靈一樣徘徊著，但並非所有企業都已做好準備。

Morrison & Foerster 全球隱私+資料安全組的律師團隊：Miriam

Wugmeister，全球隱私共同主席；歐洲隱私專家 Lokke Moerel 以及全球風險與危機管理主席 John Carlin（為美國司法部國安分部前助理檢查總長）如此解釋：GDPR 將直接適用於所有歐盟成員國，它徹底改變了企業徵求歐盟公民的同意以收集和處理個人資料的方式，

他們解釋：「單獨依賴使用者同意以進行其所有業務的公司將無法繼續這樣做，需要其他合法依據（即合約必要性和正當利益）。各公司將需要建立一個全新的生態系統通知使用者，以及確認使用者同意。」

雖然 GDPR 罰款高昂——行政罰款可高達 2000 萬歐元或全球年營業額的 4%（以較高者為準），許多企業，特別是美國，尚未做好準備。

「在 Y2K 熱潮來臨時，每個人都在準備迎戰他們可能會或不會面臨到的問題，」Hortonworks 的技術長 Scott Gnuau 說。「如今，2018 年 5 月開始實施的 GDPR 卻似乎沒有人做好萬全準備。這是為什麼？我們目前正處於一個階段，每個組織不僅在『下一步該怎麼作』耗費心力，它們在維護以及處理眼前需要解決的問題，就已經十分吃力。許多組織可能依靠資安長定義規則，系統，參數等，以幫助他們的全球系統整合商找出最佳的行動方案。這樣的重擔讓一個人扛下是不切實際的期望。」

Gnuau 說，為了確實遵守

GDPR，高階主管必須有所了解，做好準備並與組織的各層面做好溝通。企業組織需要加強其資料資產的整體治理。但是大規模的違規行為，例如 2017 年的 Equifax 違規事件，意味著各公司在替員工提供自助服務存取資料與保護資料免受潛在威脅侵害兩者之間，仍然難以取得平衡。

因此，Gnuau 預測，資料治理將成為 2018 年所有企業組織的焦點。

「一個關鍵的目標應該是發展一個在民主化資料、存取、自助服務分析和規範間取得平衡的系統，」Gnuau 說。「以安全的方式構建資料在未來對每個人都有影響——美國本土和海外的客戶、媒體、你的合作夥伴等。」

多雲端資料管理專家 Veritas Technologies 的解決方案行銷總監 Zachary Bosin 預測，第一家因 GDPR 被裁罰的公司將會是美國公司。

Bosin 表示：「儘管截止日期即將到來，Veritas 在全球調查的公司中只有 31% 認為它們符合 GDPR 標準，GDPR 對違規的處罰很嚴厲，而且這項規範將影響任何與歐盟公民打交道的公司。」 **CIO**