

frilly / September 01, 2017 02:11PM

[這是一份詳細的數據平台架構設計指南！](#)

[數據分析](#)平台的搭建從規模上分類，確定企業規模，明確合作點，非常重要。

以服裝行業為例，大型企業如波司登，本身的大數據系統架構已經完善，數據分析平台（[報表製作/商業智慧軟體](#)）在整個系統架構下的角色定位為「工具」更合適，發揮工具易用、高效開發、交互性強，穩定等優點。

中小型企業從成本上考慮，並沒有成熟的架構以及大量的投入。對於整合數據，構建數據中心報表BI系統，我們可以進行合理規劃，控制整個項目建設和運維成本，從而能夠達成更好的合作。

在時尚業行業中，對於各個分店的有效管理，如何構建合理高效的報表中心變得非常關鍵。這裡從兩方面講述，系統架構和技術實現方式，主要是鞋服行業，其他行業僅供借鑒。

首先是系統架構，從架構上分為分店管理系統（ERP和POS）及資料庫、伺服器 and 應用層客戶端。系統架構如圖：

（1）分店管理系統及資料庫。分店使用管理系統管理本店進銷存業務及相關管理工作，並定期按需將各個分店數據上傳至雲端伺服器。

（2）伺服器。構建伺服器集群。數據分散在不同的伺服器主機上可以並行存取，提高了數據的存取速度。伺服器負責存儲分店採集各種數據，並以這些數據為基礎構建數據倉庫。再部署帆軟數據分析平台，將處理結果給返回客戶端，供業務層和決策層使用。

（3）應用層客戶端。應用層客戶端分為決策和統籌管理兩部分。決策層根據所獲得的報表、圖形和走勢圖等來支援其決策。設置一個統籌管理的職能部門，統籌各分店統一促銷，暢銷商品管理和會員行為分析。企業通過雲改變了以前處理數據和接入數據的方式，數據更集中，數據一致性更強，數據質量提高，分店之間的聯繫更緊密更便捷，在這種環境下，企業的決策依據將更準確。

（4）伺服器的構建。架構如圖，ETL工具通過遠程訪問。各個分店的伺服器完成數據收集的任務，收集的數據是最原始的數據不做處理，先存儲在數據中心。數據中心為基礎資料庫，數據中心集中了所有分店的數據。數據上傳完成後繼續對數據進行ETL處理，並將處理後的數據存入到數據倉庫。數據分析應用程序根據客戶端的請求調用數據倉庫中的數據進行處理，並將結果返回給請求客戶端，同時將常用的分析按計劃定期自動分析並將結果保存到預定義分析結果模塊中。每個分店和總部的管理層都有接入雲的許可權，雲端數據共享。作為總部，可以監控各個分店的運營情況，作為分店可以及時了解其他分店的運行情況，借鑒經驗並制定銷售策略。

其次是技術實現方式，包含數據倉庫、ETL、數據分析平台。

數據倉庫（Data Warehouse，DW）是一個收集、組織、存儲和共享歷史數據的系統，其中數據ETL工具（選開源工具的話，可以用Kettle）。支援多種類型的數據源，還可以將資料庫文件下載到本地進行ETL工作。PDI分為兩個步驟，一個叫Transformation，另一個叫job，可以設定這些轉換的執行時間和頻率，這一點對於數據倉庫的自動化更新是很有幫助。

下面聊一聊數據採集與分析。

每個分店有各自的分店管理系統及資料庫，根據中央伺服器要求將需要的數據進行上傳。對於零售業來說，需要上傳的數據主要包括銷售數據、會員數據、商品數據、庫存數據、調研數據等。需要預定義所需採集的數據，包括數據的類型、數據結構。對於資料庫的數據，資料庫名稱、表名稱、表欄位都採取統一格式和名稱。對於文本型數據也要統一格式，或以xml方式存儲。伺服器收集各個分店管理資料庫的數據並對每個分店的數據標記以區分。統一標準數據可很大程度地提高數據採集的質量和後續處理效率。

對於除了分店以外的數據源，如商業共享數據平台等，需要根據實際情況設計相應介面和採集方法，帆軟數據分析平台內置採集數據功能，可以非常方便根據業務情況定製數據採集模塊。

數據的分析工作在按照數據倉庫對數據的要求並選擇合適的工具對不同類型的數據進行處理，然後保存到數據倉庫中。隨著時間的推移，數據中心的數據量會不斷增加，運用大數據分析軟體工具是非常有必要的。大數據工具的主要特點是通過伺服器集群中的主機並行處理數據，將一個龐大的任務分解為小任務處理。

應用程序部署到雲端以後，客戶端通過瀏覽器調用相應的功能，只需將結果返回給客戶端，在客戶端進行數據分析結果的展現。針對時尚業的數據分析可以包括多個方面，比如：銷量分析、客戶購買偏好分析、商品關聯分析、精準推送服務等。

文 | 帆軟數據應用研究院 賈強

---