

frilly / July 03, 2017 05:11PM

[電力+大數據項目中的架構研究與思考](#)

[b]閱讀目錄：[/b]

- 1、前言
- 2、業務架構
- 3、數據架構
- 4、技術架構
- 5、實施架構
- 6、示範架構
- 7、小結

[b]前言[/b]

智慧電網(Smart Grid)是以物理電網為基礎，將現代先進的感測測量技術、通信技術、信息技術、計算機技術和控制技術與物理電網高度集成而形成的新型電網。

電力大數據(Power Big Data)是實現智慧電網的關鍵技術之一，它通過挖掘數據之間的關係與規律，提高電網企業在生產、經營、管理等方面的質量與效率。如開展電網設備狀態監測的大數據應用，實現電網設備狀態的智慧監測，實時分析電網線損、配電負載等等。

本文旨在跟讀者分享某電網公司在配用電大數據項目中所採用的多維架構(包含數據架構、業務架構、技術架構等)，為本系列的後續文章打下鋪墊。

[b]業務架構[/b]

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017062109.png[/img]

配用電大數據項目的業務架構，是指從業務角度說明配用電大數據項目要做什麼事。此架構不會過多牽涉技術細節，它的重要性要高於其他幾類架構。一般來說，這類架構要在項目啟動前，通過多次的調研、分析、專家研討後方可決定。

上圖的業務架構主要將業務劃分為五大層次，其中最為關鍵的是數據源層和應用層：

1. 數據源層：規定配用電大數據項目能從哪些地方獲得數據資源。這是非常重要的一環，尤其是在電網領域。因為當前電力信息系統中的「網路孤島」現象比較嚴重，要梳理清楚哪些數據能采、哪些數據采上來有意義，是非常不容易的。
2. 應用層：明確配用電大數據能為電力系統實現哪些業務。規劃該層次時，行業化大數據從業人員需要和電力專業的人員進行多次深入地溝通交流。從筆者親身經歷來看，這一層切不可假大空，一定要確保落地。通俗點來說，若這層寫得太虛，可能會把後續開發人員，甚至是自己給坑了...

至於其他幾個層，則是從一個較為宏觀的角度去設計系統組件。一般來說在業務架構的側重點在系統的功能性方面，對於技術細節不過多糾結。

[i][u]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]！[/u]/i]

[b]數據架構[/b]

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017062110.png[/img]

電網企業的數據主要包括三類：

1. 電力設備數據：主要包括電網設備監測數據、設備地理位置數據、設備狀態數據等；

2. 企業管理數據：主要包括跨單位、跨部門的電網企業職工數據、財務數據等；

3. 企業運營數據：主要包括客戶信息、客戶用電數據、電費數據等。

但是上述只是一個特粗略的分類。筆者在項目實施過程中發現，數據的分類在每一個環節都需要按照不同標準重新做一次。

為何要這麼麻煩？這是因為，[數據類型]+[業務需求]將決定你選用何種[url=http://www.finereport.com/tw/]大數據分析軟體[url]組件去處理它。

這裡先以電網的拓撲結構數據為例：這類數據大都存在電力系統的RDBMS里，那麼我們顯然可以考慮使用Sqoop來做同步；而其後為高效實現電網拓撲分析業務，顯然應將其放至HIVE這類數據倉庫工具里合適。

再以電網設備檢測數據為例：這類數據由於具有事實性，用Storm或者Spark Streaming來同步就顯然更加合適了；而這類數據有部分業務環境是不需要做太多[url=http://www.finereport.com/tw/]數據分析[url]的，因此可考慮將其導入到HBase這類NoSql數據里，實現高效存取。

讀者看到這裡，應該明白了需要時刻思考數據分類的原因了吧？上述兩個例子都屬於電力設備數據，然而它們被處理的方式顯然是不同的。在實際中，我們往往根據當前架構所在層次的屬性來決定使用何種組件來處理數據。個人真心建議針對將來數據特別複雜的情況，可以考慮引入「數據畫像」這個概念，根據不同的處理方式為各類數據打上標籤，以便於管理。

[b]技術架構[/b]

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017062111.png[/img]

總的來說，針對配用電大數據的技術研究可以分為三個層面來展開：

1. 數據集成層面：研究電力系統中多源數據的分類方式、集成與融合方法，並設計出面向雲環境的多源異構數據集成模型。
2. 基礎架構層面：結合線上流處理與離線批處理的應用需求，研究可拓撲分解的流處理計算技術、分布式並行批處理計算技術，並提供應用編程介面。
3. 支援系統層面：研究電力大數據項目的建設規範，大數據集群系統的綜合管理工具、大數據可視化組件，並提供多種形式的集成介面，以便支援不同上層應用對大數據以及分析結構的調用需求。

需要特別說明的是，在這三個層面之上是真正的「電力應用層」。

[b]實施架構[/b]

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017062112.png[/img]

對於配用電大數據項目的具體實施，需要明確的主要是將計算機集群具體分成哪些區，每個區又具體採用哪些組件。

這部分內容比較繁雜，以下僅針對其中某類實時數據的處理做個大致的介紹：

1. 各業務系統和數據採集系統的秒級數據通過專線網路，經過加密壓縮傳輸到總部的負載均衡器；
2. 負載均衡器將數據分發給Kafka集群落地；
3. Storm集群從Kafka集群接收所訂閱的數據，負責對數據進行清洗、按照設定的告警條件實時監測數據並發出告警；
4. Storm清洗和標註後的數據，直接存入HDFS落地；
5. HDFS中的數據同步到數據存儲和查詢模塊（時序數據管理平台），方便在其中進行線上查詢；
6. 數據分析平台上根據預訂的作業隊列，調度數據分析程序在Hadoop集群中運行，結果存入HDFS或者按用戶程序定義寫入相應存儲位置；

7. 數據分析平台將秒級數據匯總成十分鐘級數據、根據定義的數據種類、數據格式和存儲方式將數據分發給計算存儲群組及HBase資料庫；統計報表程序通過Hive集群執行各種類SQL完成統計查詢和[[url=http://www.finereport.com/tw/](http://www.finereport.com/tw/)]報表製作[/[url](#)]生成。

( 上述介紹僅是針對其中某類實時數據的處理，而不同類型數據的處理方式是不同的 )

[b]示範架構[/b]

[img]<http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017062113.png>[/img]

在項目後期，需要將配用電大數據平台部署到部分地市局來進行試點，因而需要明確網 – 省兩地，或者網 – 省 – 市三地的綜合示範架構。

在本文給出的參考架構中，我們首先利用高速4G專網和GPRS /230M無線專網實現低壓居民用戶和專變/公變終端的採集；採集的數據通過智慧一體化終端進行簡單轉換後，上傳至區域分布式大數據中心；區域大數據中心將對電量和非電量數據，結構化與非結構化數據進行大數據集成與融合。

在區域大數據中心，可基於大數據聚類與分析技術，實現用電用戶類型的精細化劃分、分析用戶的用電行為、評估非介入式用戶的能效水平，形成一系列面向配用電網的通用知識模型與關鍵技術，為省級大數據中心提供數據與關鍵演算法支撐。

[b]小結[/b]

作為該系列博文的開篇，本文從各類架構的角度出發讓讀者對配用電大數據的項目有了全方位的整體認識。

後續的文章將涉及到真正的電力+大數據研究，這也是電力專業與計算機專業的綜合領域，讀者或許需要具備一定的電力系統知識才能消化。

簡單回顧下電力系統相關的知識，然後一起開始智慧電網之旅吧^\_^。

作者 | 穆晨

原文 | 配用電大數據項目中的架構研究與思考

---