

# 电力大数据中信息化工具的应用与探索

曾 愚

(国网四川省电力公司信息通信公司, 四川 成都 610041)

**摘要:**坚强的智能电网能够运用先进的通信、信息等技术来实现完全自动化的电力传输,能够监控每个电网节点与用户节点,并且在电力传输的整个网络中能够保证信息和电能的双向流动。而我国的智能电网发展还处于起步阶段,因此,选择合适的信息化工具可以极大地加快智能电网信息化的建设进程。

**关键词:**电力;大数据;信息系统;智能电网

**中图分类号:** TM73; TP39; TN8; TM93

**文献标识码:** B

**文章编号:** 1001-2184(2015)05-0100-04

## 1 概述

随着互联网、物联网和云计算技术的应用和普及,各类数据的规模均呈现出爆发性增长态势,数据的存储单位也从过去的 GB、TB 上升到 PB、EB 甚至是 ZB 级别。在动辄数千 PB 的数据面前,过去常用的信息化工具和系统已无法满足对这些数据进行分析、处理的需求。另一方面,越来越多的政府、企业开始关注、了解和接受大数据的概念,并尝试运用大数据技术来协助它们进行决策层面的分析与预测<sup>[1]</sup>。

而在未来竞争和机遇都倍增的大数据“互联网+”时代,若要紧跟市场的步伐,甚至走在市场的前列,对电力企业而言,就必须根据适合自身行业特点,选择能够最大限度地适合与满足电力生产需求的各类信息化工具和技术来解决在其发展中遇到的各类问题。

## 2 大数据时代中的信息化工具

### 2.1 Hadoop 工具

#### 2.1.1 Hadoop 工具的内容

Hadoop 是由 Apache 基金会开发的分布式系统基础架构。作为开源系统,Hadoop 在当今互联网的大数据解决方案中应用广泛。通常,它部署在成千上万台主机的集群中进行大数据的存储和分析,而用户可以在不了解底层细节的情况下,开发分布式程序并充分利用集群的运算能力进行大数据集的快速处理和存储<sup>[2]</sup>。

#### 2.1.2 Hadoop 具有的特点

在 Hadoop 架构中,其核心包括两部分,分别

是 Hadoop Distribute File System(以下简称 HDFS)和 Map Reduce,其特点如下<sup>[3]</sup>:

(1)高容错性:文件的所有数据块都会保留对应的副本,并且每个文件的数据块大小和副本系数都是可配置的。

(2)高数据一致性:所有文件都是一次性写入的,并且严格要求在任何时候只能有一个写入者。

(3)低成本性:Hadoop 是开源的并且可部署在普通的 PC 服务器上,企业可以通过购买一定数量的 PC 服务器就可以建立起一个大数据集群,用于替代价格昂贵的小型机服务器。

(4)数据的高可靠性:,Hadoop 会采取将同一个文件的块副本同时分别存储到已互联的另几台设备上上来防止某台设备故障或网络通信问题而导致的无法读取的问题,从而有效地提升了系统数据的可靠性。

(5)高效性:Hadoop 将大型数据集进行分解至数个小型数据集并发送到多个节点(Data Node)上,再将这些小型数据集以单个的形式提取出来进行加载汇总分析。其中每一步都可以通过简单的需求命令来实现,从而大幅度提升了工作效率<sup>[4]</sup>。

### 2.2 SAP HANA 工具

#### 2.2.1 SAP HANA 工具的内容

SAP HANA 是思爱普(以下简称为 SAP)公司近年来推出的一项用于解决企业大数据管理的新软硬件结合体。

收稿日期:2015-05-05

### 2.2.2 SAP HANA 工具具有的特点

SAP HANA 数据库是这套体系的核心,它实时地参与到数据查询分析与数据密集型业务流程的分析当中。与传统数据库相比,SAP HANA 的特点可以归纳为以下三方面<sup>[5]</sup>:

一是数据快速访问。HANA 支持对内存数据进行压缩,最小化了传输数据量,并且将逻辑数据的处理转移到数据库层面完成,从而节约了应用层到数据库之间传输消耗的资源,也解决了传输速率和吞吐量的瓶颈问题。

第二是数据支持多节点计算与大规模并行处理<sup>[6]</sup>。HANA 可以将大量数据分发至不同的处理器进行运算,同时单组数据也可同时被同组的服务器所共享,由此而使整个系统的运行安全得到了提升,即单一服务器宕机将不再影响到数据的运算。

第三是 SAP HANA 数据库在支持非结构化数据的同时提供了三种内存处理引擎:关系型引擎、图形引擎和文本引擎<sup>[6]</sup>。除去关系型引擎,HANA 数据库的非结构化搜索是 SAP HANA 的基础。其文本引擎提供了单个词组、段落以及具有容错性的模糊查询等查询方式,并且可对结果进行操作,而 HANA 的图形引擎则可用于海量独立资源的计划分配与应用。

## 3 大数据时代下的电力企业信息系统

### 3.1 智能电网与信息系统

未来的电网需要适应和整合所有新型可再生能源,而配套的能源管理系统也需要迅速满足这些日益增长的能源需求。在这种情况下,电网的发展就需要大步朝着智能电网的方向发展,随之而来的是急需更可靠的通信、更先进的信息化技术与设备去打造一个高度智能化的分布式能源自动传递网络<sup>[7]</sup>。

### 3.2 电力企业信息系统具有的特点

电力信息系统是一个覆盖面极其广泛的概念,涵盖发电、输电、变电、配电以及售电等环节。各个环节都涉及到与信息技术紧密相关的众多智能设备和双向通信系统,这就导致其在进行操作与监控工作的同时为电力信息系统带来相当庞大的数据流,同时,这些数据由于业务上的需求需要

被妥善保存与管理。

### 3.3 电力企业信息系统的的核心需求

处于不同环节的电力信息系统均对其数据的要求有着不同的苛刻要求,但它们具有的共同点均集中在加强对资源消耗的控制与预测的同时,保持与数据源之间交互的连续性。因此,电力企业信息系统所遇到的这类问题需要用大数据的方案来解决。随之而来的问题便是如何将智能电网信息系统中的数据融入到分布式的云计算的架构中来完成大数据的处理<sup>[8]</sup>。

### 4 电力大数据中信息化工具的比较与应用

笔者分别从实施部署成本、技术改造难度以及安全风险等方面分别比较以下两种工具。

#### 4.1 实施部署成本

在电力行业的信息化建设中,实施部署的成本是一个重要的衡量标准。高昂的部署成本如果无法带来相应的经济效益,对企业而言,其不仅是资源上的浪费,同时亦将成为企业的负担。以国网四川电力公司为例,各类业务信息系统的数据量一般为 1~50 TB 左右,尚未达到 PB 级别,在这种情况下:

(1)SAP HANA 的实施成本中内存数据库硬件占有很大比例且价格昂贵。SAP HANA 由于会对数据进行压缩后再进行处理,所以,在其实施前会对目标数据库的大小进行评估,从而估算出需要配备的内存数据库规模,通常,该比例保持在 1:7。从 SAP 公司公布的 HANA 硬件费用看,如果按照单个系统平均数据量为 5 TB 计算,那么,平均每套硬件的实施费用将高达上百万元人民币,额外再加上人员实施费用,故 SAP HANA 部署的成本十分高昂。

(2)反观 Hadoop,其可以部署在价格低廉的 PC 服务器上,并提供相当可观的计算处理资源。与相比动辄上百万元的小型机服务器相比,同等费用支出下运用 Hadoop 部署在 PC 服务器上的服务器集群的运算、数据存储、管理、处理能力均远远高出小型机。

从上述比较中不难看出:Hadoop 在实施部署成本上较 SAP HANA 具有较大的成本优势。

#### 4.2 技术改造难度

电力行业作为高速发展的行业,企业内部的业务需求会随着行业和企业的持续发展而不断发生变化。因此,与之相关的信息系统的硬件资源、系统配置、业务流程、程序代码、接口交互等方面都应随之不断优化和改造升级以满足生产要求:

(1)SAP HANA 针对业务流程变更的适应性良好,它提供了一套完善的流程变更机制去配合生产中的业务需求。而在新增功能开发上,由于SAP HANA 采用的是独有的程序语言,因此需要另行配备具有相关资质的专业技术开发人员而导致新增功能开发的周期延长和成本上升,具有一定的技术改造难度。

(2)Hadoop 在技术改造方面的优势在于系统代码透明化,企业可自行进行优化和更新改善以符合需求。其可拓展性较高,新增功能较为方便;但它所具有的缺点也是显而易见的,在执行技术改造期间,如果不支出费用、聘请第三方公司作为技术支持与协助,技术改造的难度与风险会明显升高。

#### 4.3 安全风险

电力信息系统的安全性可以分别从物理层面、网络层面、应用层面和数据安全层面四个层面进行衡量。由于SAP HANA 与Hadoop 均属于数据库层面的工具,所以,在网络安全与应用安全层面二者更易受外部因素影响,而且比较集中在物理安全与数据安全层面:

(1)物理层面安全:物理层通常是从信息系统的运行硬件、介质、基础设施和外界环境等方面去评估其安全性。SAP HANA 硬件的供应与部署可由厂商完成,其后续的维护保养也可以得到厂商支持;而Hadoop 的硬件设备则多数依靠企业自行采购并部署,其维护保养则可由相应的硬件厂商完成;不足之处在于企业需要提前对自行采购的硬件是否满足系统需求做出较为准确的评估,而SAP HANA 则可由厂商完成这一评估。

(2)数据安全方面:无论是SAP HANA 或是Hadoop,它们都提供了较完善的数据安全机制:Hadoop 的核心准则是每一份数据都会拷贝成三份并分别存放在不同的服务器中,从而避免了某一台服务器出现故障时引起的数据丢失;虽然

SAP HANA 会将内存中的数据自动地、持久化写入到硬盘之中,避免数据库故障引起的数据丢失,但是,对于硬盘故障造成的数据丢失则只有通过自带的软件或第三方软件进行数据备份来避免。从企业角度看,Hadoop 的安全机制适用性更佳。

#### 4.4 信息化工具的应用方案

##### 4.4.1 高速数据访问平台架构

从上述各方面的比较中不难看出:无论是SAP HANA 或是Hadoop,都无法独立承担起电力企业信息系统的全部需求,但二者的特点却互为补充。基于此,构建一个以Hadoop 与SAP HANA 相结合的、实时高速数据查询平台去满足当前电力企业信息系统的各方面需求不失为当前电力企业大数据一种可行的解决方案。

##### 4.4.2 高速数据访问平台运行机制

这个平台运用Hadoop 的廉价硬件部署成本为企业建立起庞大的数据源,再利用SAP HANA 的高速访问能力对这些数据进行快速批处理更新以及载入,从而在整体上得到提升。但亟待解决的问题是SAP HANA 与Hadoop 之间如何交互?由于两者的差异性,应当首先需要考虑的是如何将Hadoop 数据源中的数据合理地分类与部署,建立“数据湖”模式的数据源,以便SAP HANA 快速访问。数据湖(Data Lake)与云资源池相仿,但是其数据的储存与分类更具有可管理性,其还可以在数据不移动的情况下进行计算与分析各种来源、格式的企业级数据。因此,可以通过将各信息系统的历史数据以原始格式迁移到Hadoop 构建的数据湖中用以降低保存数据的前端成本。

在Hadoop 实现的数据湖层级之上是SAP HANA 构成的数据分析环境。对于SAP HANA 而言,Hadoop 可以以单一数据库和数据过滤器两种形式与其进行交互。当Hadoop 与SAP HANA 完成数据交互后,SAP HANA 将数据传输到应用层,便可将这些数据展现给企业,供其展示或进行分析与预测。

##### 4.4.3 高速数据访问平台具有的优势

在二者相互访问的整个过程中,数据均处于实时高速交互状态,从而高效地契合了企业快速而多变的业务流程,缓解了海量历史数据读取带

来的信息系统压力,不仅为电力企业节约了大量的资源与时间成本,还能够提供准确地实时统计与分析。不难看出,二者的结合运用在电力企业发展的决策层支持方面具有单一工具所不具备的巨大优势。

## 5 结 语

随着智能电网系统的进一步发展,大数据分析预测这一先进的数据技术为电力企业带来的收益将被越来越多的人所认同。但是,由于目前电力行业的大数据发展尚处于初期的数据治理阶段,且因信息系统历史数据种类繁多,复杂程度大,对于有效数据的甄别还存在困难,故在当前阶段更应注重对历史数据的梳理和分析,继而能够准确挖掘出企业所急需的数据。但在数据治理完成后,则应逐渐转向选择合适的大数据信息化工具,对已有的数据进行宏观层面地分析和预测,从而为智能电网的建设发展提供可靠的技术支撑。

### 参考文献:

- [1] Benjelloun, F. - Z. ; Lahcen, A. A. ; Belfkih, S. Intelligent Systems and Computer Vision (ISCV), "An overview of big data opportunities, applications and tools", 2015 Year: 2015 Pages: 1 - 6, DOI: 10.1109/ISACV.2015.7105553.
- [2] Apache Software Foundation, "Hadoop Wiki", [EB/OL], <http://wiki.apache.org/hadoop/>, 2015 - 6 - 14.

- [3] Azzedin, F., "Towards a scalable HDFS architecture", Collaboration Technologies and Systems (CTS), 2013 International Conference on Year: 2013 Pages: 155 - 161, DOI: 10.1109/CTS.2013.6567222.
- [4] Guanghui Xu; Feng Xu; Hongxu Ma, "Deploying and researching Hadoop in virtual machines", Automation and Logistics (ICAL), 2012 IEEE International Conference on Year: 2012 Pages: 395 - 399, DOI: 10.1109/ICAL.2012.630824.
- [5] J. Krüger, M. Grund, C. Tinnefeld, H. Plattner, A. Zeier, and F. Faerber., "Performance for Read Optimized Databases", In DASFAA Conference, pages 291 # 305, 2010.
- [6] Faerber, F., Cha S. K., Primsch, J., Bornhövd, C., Sigg, S., Lehner, W., "SAP HANA database: data management for modern business applications", SIGMOD Rec on Year 2011 Vol. 40, No. 4.
- [7] Bitzer, B.; Gebretsadik, E. S., "Cloud computing framework for smart grid applications", Power Engineering Conference (UPEC), 2013 48th International Universities Year: 2013 Pages: 1 - 5, DOI: 10.1109/UPEC.2013.6714855.
- [8] McHann, S. E., "Grid analytics: How much data do you really need?", Rural Electric Power Conference (REPC), 2013 IEEE Year: 2013 Pages: C3 - 1 - C3 - 4, DOI: 10.1109/REPCon., 2013. 6681858.

### 作者简介:

曾 愚(1986-),男,四川成都人,工程师,双硕士,从事信息系统运维技术与管理工。 (责任编辑:李燕辉)

(上接第 96 页)

运行方向和急刹车,以免筋带被拉动变位并产生超量变形。面板附近 1.5 ~ 2 m 范围内,尤其是靠近面板背部不允许用大、中型压路机进行碾压,应使用小型机械并配以人工方式压实。

### 6.10 变形观测

施工过程中实施的变形观测:在挡土墙施工过程中,每安装两层格栅面板后需进行变形观测一次。变形观测主要采用全站仪、电子水平尺和垂球进行,主要目的是控制整个格栅面板的平整度,确保格栅面板安装轴线位置符合设计要求。

施工结束后的变形观测:施工结束后,为了检查挡土墙施工质量和挡土墙在工作情况下的变形情况,工程竣工后应进行一次全面的变形观测并作好记录,在以后的三个月内每个月观测一次、雨季前后各观测一次。

## 7 结 语

通过对镀锌格栅面板加筋土挡土墙的设计理论和施工技术的研究,结合该挡土墙所具有的结构新颖、造型美观、施工快捷等特点,依姆铀矿项目已将其进行推广并应用于部分临建工程,包括 2 × 500 t 稳定土拌和站和大型条筛挡土墙的设计与施工,主要采用集装箱作为挡土墙面板、后背加筋土回填的方式。目前,我国正在大力开展基础设施建设,包括道路、桥梁、水电、矿山以及市政等工程建设,因此,对镀锌格栅面板加筋土挡土墙进行推广应用具有重要意义,必将会带来较大的社会效益。

### 作者简介:

刘 洪(1968-),男,四川简阳人,国际公司总工程师,高级工程师,学士,从事水利水电工程施工技术与管理工;  
阿里木江(1985-),男,新疆喀什人,助理工程师,学士,从事水利水电工程施工技术与管理工。

(责任编辑:李燕辉)