



电力大数据全景实时分析关键技术

周国亮¹, 吕凇杰¹, 王桂兰²

- (1. 国网冀北电力有限公司技能培训中心, 河北 保定 071051;
2. 华北电力大学信息与网络管理中心, 河北 保定 071003)

摘要: 针对智能电网建设过程中收集的电力大数据, 基于电力系统全景实时数据分析的需求, 探讨基于大数据的电力系统安全可靠分析、实时状态监控及能源全景动态平衡调度等核心问题的解决思路。分析了利用大数据解决安全可靠、设备全寿命周期管理及能源实时平衡调度等问题的挑战及解决思路, 基于大规模实时多源细节数据和设备全景数据的计算, 有助于提高系统分析的精度和准确度, 保证电网安全运行; 探讨了内存计算、实时流式大数据处理、大规模并行计算及列存储等技术在电力大数据实时分析中的应用; 结合主流开源大数据处理技术, 设计了电力大数据分析平台的分层体系架构, 为电力系统的高效运行提供保证。

关键词: 电力大数据; 全景实时数据; 内存计算; 数据流; 大规模并行

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016103

Key technology of power big data for global real-time analysis

ZHOU Guoliang¹, LV Linjie¹, WANG Guilan²

1. Skill Training Center of State Grid Jibei Electric Power Co., Ltd., Baoding 071051, China

2. Network and Information Management Center, North China Electric Power University, Baoding 071003, China

Abstract: For power big data collected during smart grid construction process, based on the demand of power system global and real-time data analysis, ideas of solving power system security and reliability, real-time status monitoring, energy global dynamic balance scheduling and other key issues were explored. The problems of big data safety and reliability, equipment life-cycle management and energy real-time balance scheduling were analyzed and discussed, system analysis precision and accuracy based on large-scale real-time multi-source detail data and global data of equipment would be improved, then application of in-memory computing, real-time streaming data processing technology, massively parallel computing technology and column stores were explored; a layered architecture of power big data analytics platform which combined with the mainstream open source big data processing technology was proposed to provide guarantees for the efficient operation of the power system.

Key words: power big data, global real-time data, in-memory computing, data stream, massively parallel

收稿日期: 2016-01-25; 修回日期: 2016-03-15

基金项目: 河北省自然科学基金资助项目 (No. F2014502069); 中央高校基本科研业务费专项资金资助项目 (No.13MS103)

Foundation Items: Natural Science Foundation of Hebei Province (No. F2014502069), Fundamental Research Funds for Central University (No.13MS103)



1 引言

近年来,随着全球能源危机、环境问题等因素的不断加剧,世界各国对清洁能源的开发利用程度不断提高,大量分布式、间歇性能源的广泛接入对电力系统的安全稳定运行提出了更高的要求。在该前提下,智能电网应运而生。智能电网建设的最终目标是成为覆盖发、输、变、配、用及调度等过程的全景实时电力系统,而支撑系统得以准确、安全、实时及可靠运行的基础是电力系统多源异构大数据的快速采集、响应和分析^[1,2]。未来智能电网既要支持个人终端用户与电网系统的交互,也要满足控制系统对电网安全稳定性的需求,智能电网中的多数应用需要海量数据处理技术的支撑。

随着智能电网建设在广度和深度上的不断推进,在智能电网运行过程中会收集到系统内外的海量全景数据,形成电力大数据。比如,截至 2013 年底,国家电网公司累计安装智能电能表 1.82 亿只,实现用电信息采集 1.91 亿户,智能电表应用量占全球的一半,其用电信息采集系统成为世界上最大的电能计量自动化系统,将产生以 PB 级计的数据^[3]。同样在电力系统状态监测中安装的 PMU(phasor measurement unit,相量测量单元)是用于进行同步相位测量、输出和动态记录的装置。100 个 PMU 一天收集 62 亿个数据点,数据量约为 60 GB,而如果监测装置增加到 1 000 套,每天采集的数据点为 415 亿个,数据量将达到 402 GB。表 1 展示了智能电表生成数据量与采集频率的关系。

表 1 智能电表生成数据量与采集频率的关系

表数量/只	采集频率 15 min	采集频率 1 min	采集频率 1 s
10 000	32.61 GB	498.0 GB	114.6 TB
100 000	326.1 GB	4.8 TB	1.1 PB
1 000 000	3.18 TB	47.7 TB	11.2 PB

同样,在电力系统的其他各环节,随着大量传感采集装置的安装都将产生大规模的数据,而如何规划、存储、整合及综合分析这些数据,是电力系统当前需要迫切解决的关键问题之一。但电力大数据与传统互联网大数据不同,对分析的全景性、实时性和可靠性提出了更高的要求。

通过建设基于大数据分析技术的全景实时系统,并使之成为整个系统的“神经中枢”,将有助于从大数据驱动的角度解决电力系统当前面临的主要挑战,比如可靠性分析、实时调度等问题,也有利于提高数据的利用率,挖掘数据价值。电力大数据全景实时分析平台在系统中的地位如图 1 所示。

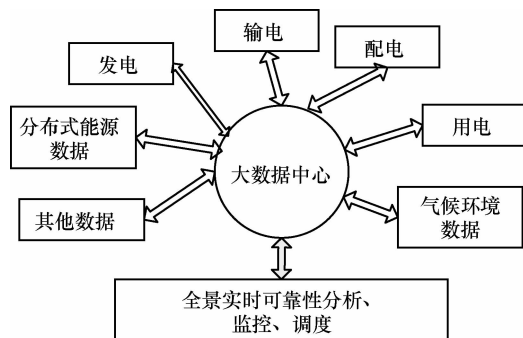


图 1 电力大数据分析平台在系统中的地位

将电力大数据中心建设成为整个电力系统的中心节点,负责全景信息收集、整合、状态监控和资源调度平衡。通过获取整个系统的实时全景数据,将电力系统变成一个“端到端”的透明系统,实现全局信息共享,消除“信息孤岛”,避免由于信息不对称造成的资源浪费;利用大数据分析技术,结合电力系统理论知识,构建系统分析模型,提高电力系统整体规划水平,促进能源动态平衡;利用全景实时大数据,实现对输变电设备的实时监控、评估、分析及状态预测,从而快速隔离事故;利用大数据,实时获取发电及负荷信息,达到资源的优化配置、调度,提高新能源接纳吸收能力,促进“削峰填谷”和高级需求响应技术的应用。

2 相关工作

云计算技术作为处理大数据的有效方式,在国内电力行业中的应用研究还处于探索起步阶段,研究内容主要集中在系统构想、实现思路和前景展望等方面。参考文献[3]针对智能电网状态监测的特点,结合开源技术 Hadoop,借助虚拟化技术、分布式冗余存储以及基于列存储的数据管理模式来存储和管理数据,以保证电网海量状态数据的可靠和高效管理,但目前还只是一个框架设想。为了解决电力系统灾备中心资源利用率低、灾备业务流程复杂等一系列问题,参考文献[4]设计了云计算资源管理平台框架和部分模块,其目标是实现电力企业 ERP 数据的备份,但尚未实现。参考文献[5]初步设计了电力系统仿真云计算中心的系统架构及其所包含的层次结构:基础设施云、数据管理云、仿真计算云等。参考文献[6]探讨了未来智能电网控制中心面临的挑战,提出物联网和云计算技术结合是新型控制中心的技术支撑。基于 Hadoop 云计算平台,设计实现了基于 Hadoop 分布式文件系统的电力设备状态监测存储系统^[7],对动态时序数据、静态数据以及视频数据进行了存

储、关键字查询与并行处理方面的研究,通过对系统的测试,验证了云计算平台高可靠性、良好的可扩展性和数据并行访问的性能优势。

在国外,云计算技术目前已用于电力系统海量数据的存储和处理,并有实际运行的系统。参考文献[8]分析了电力系统中不同用户的实时查询需求,设计了用于实时数据流管理的智能电网数据云模型,该模型较适合处理电力系统中的大规模流式数据,同时基于该模型实现了一个实时流数据的智能测量与管理系统。Cloudera公司^[9]设计并实施了基于Hadoop平台的智能电网在TVA(Tennessee Valley Authority,田纳西河流域管理局)上的项目,帮助TVA管理PB级的PMU数据,体现了Hadoop平台高可靠性以及价格低廉的优势。另外,TVA在该项目基础上开发的superPDC,通过openPDC项目将其开源,此工作将有利于推动量测数据的大规模分析处理,并可为电网其他时序数据的处理提供参考。日本Kyushu电力公司^[10]使用Hadoop平台对大规模的电力系统用户消费数据进行快速并行分析,并在该平台基础上开发了各类分布式的批处理应用软件,提高了数据处理的速度和效率。参考文献[11]对云计算平台应用于智能电网进行了详细的分析,得出的结论是:现有云计算平台可以满足智能电网监控软件运行的可靠性和可扩展性,但实时性、一致性、全局性、数据隐私性和安全性等方面的要求尚不能满足,有待进一步深入研究。

随着大数据在电力系统中的形成,电力大数据的研究也获得了各方关注。参考文献[12]探讨了大数据在电力系统生产、控制和保护中的应用,从特征量提取、系统整合和案例分析3个方面进行了论述,并指出操作计划制定、实时监控和故障检测保护是将来的研究方向。参考文献[13]认为电力系统大数据与传统商业互联网大数据不同,对实时性和安全可靠性的要求更高,需要设计新型的处理分析系统,作者设计了原型系统,并在电力公司试用,性能获得了一个数量级的提高。风机选址是风电场建设中的一项重要工作^[14],IBM利用气候大数据、潮汐相位、地理空间与传感器数据、卫星图像、森林砍伐地图等信息帮助丹麦能源公司维斯塔斯的风机选址安装,从而使风机获得最优的发电量和最低的维护成本等。

电力系统是一个复杂、庞大的系统,传统的系统可靠性评估、潮流计算等算法,具有处理数据有限、算法复杂度高、计算量大等特点,很难适应全景大数据的环境。参考文献[15]研究了实时并行潮流计算技术,通过并行计算技术

提高计算的效率。参考文献[16]研究了利用集群计算电网可靠性的技术。参考文献[17]研究了海量电网交换数据的处理技术。

综上所述,电力大数据分析处理已经成为提升电力系统安全和运行效率的重要途径,也是智能电网建设过程中必须面对的问题。目前研究思路主要集中在互联网大数据技术在电力系统中的应用,而结合电力系统特点,并与电力生产运行系统紧密结合的大数据分析技术研究较少,将是未来研究的方向。

3 全景实时分析关键技术问题

3.1 基于大数据的大电网安全可靠分析技术

随着电网规模的不断扩大、大规模可再生间歇性能源的接入,电网面临着严峻的安全可靠性问题^[16]。传统电力系统安全可靠分析技术具有采集点少、运算量大的特点,随着覆盖整个系统的信息采集系统逐步完善,智能电网可以实时获取整个系统的全景数据,而基于全景数据的分析评估计算量非常大,容易造成“计算灾难”,几千阶的微分方程求解,无法满足实时性要求;另一方面,随着大规模分布式间歇性能源的接入,由于分布式能源的不确定性和难以预测等,很容易对大电网造成冲击,从而对系统的稳定性提出了更高的要求^[18]。因此,利用大数据从全局角度来动态实时分析系统的安全可靠性是电力大数据分析面临的一项重要工作。

潮流计算是电力系统安全可靠分析的重要手段之一,而传统潮流计算的计算复杂度高、数据量小、计算精度难以保证。当面对全景实时大数据和动态实时变化的分布式能源时,传统的计算技术很难动态精确计算系统的潮流情况,而基于大数据的高性能并行快速高精度的潮流计算技术,有助于提高潮流计算的速度和精度,从而提高电力系统可靠性水平。

目前电力可靠性分析主要采用传统的 $N-1$ (即当某一设备失效后系统的安全性评价)安全准则,不考虑概率较低的多重事件,而电网事故中很多情况下会因为某一设备故障产生连锁反应,所以有必要研究多重事故发生情况下的系统可靠性问题(即 $N-m$)。而 $N-m$ 情况下的可靠性分析计算量将呈指数级增长,并且在数据规模不断增长的情况下,分析任务很难在有效时间内实时完成。

在电力系统安全稳定分析中,大多数分析归结为代价很高的高维稀疏矩阵与向量常量的乘法运算。比如在潮流



计算的过程中,如果电力网络有 n 个节点,则有节点方程:

$$I=YV \quad (1)$$

其中, $I = \begin{bmatrix} \dot{I}_1 \\ \dot{I}_2 \\ \vdots \\ \dot{I}_n \end{bmatrix}$, $V = \begin{bmatrix} \dot{V}_1 \\ \dot{V}_2 \\ \vdots \\ \dot{V}_n \end{bmatrix}$ 分别是节点注入电流列向量

及节点电压列向量。 $Y = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nn} \end{bmatrix}$ 是导纳矩

阵,对角元素是节点 $i(i=1,2,\cdots,n)$ 的自导纳,非对角元素是节点间的导纳。式(1)是典型的非线性方程,求解方法一般为迭代法,而求解的核心运算是矩阵与向量的乘积。因此,快速高维稀疏矩阵运算是电力系统可靠性分析的基础。

3.2 基于大数据的输变电设备全寿命周期管理

除了安全稳定运行非常重要外,电力系统本身是一个动态实时变化的系统,因此必须实时监测系统的运行状态,快速处理各种情况,保证系统的安全稳定运行。统计发现大多数电网故障主要是设备故障问题引发的,通过收集设备的全寿命周期数据(实验数据、运行数据及气候环境数据等),建立设备运行模型,有利于实时评估设备状态,从而避免由于设备故障造成的电网事故^[19]。目前的设备管理系统记录存储的主要是设备台账信息,而设备全寿命数据由于时间和空间跨度大,收集整理具有很大的难度。通过大数据中心,整合系统可利用的所有内外部数据,使设备的全寿命周期管理成为可能。利用设备全寿命周期大数据,构建新的设备运行评估模型,以修正由于试验环境条件有限造成的设备模型不准确、运行等级分类不足等问题。在设备全寿命周期大数据的基础上,利用聚类划分、异常点检测、深度学习等技术,实现设备的无监督分类及故障预测^[20]。

当前,电力设备状态评估主要是基于实验室建立的模拟环境产生的测试数据,利用人工智能(神经网络等)对数据进行分析、划分等级。然而这种方式存在数据样本有限、数据与真实环境不一致、算法针对大规模数据集失效等问题。借助电力大数据,利用数据分析技术,挖掘设备状态、环境等之间的关联规则,发现有价值的特征量^[21-23],实现设备状态监控,保证电网安全运行。因此,在设备全寿命周期管理中的关键问题是数据挖掘分析算法在电力系统中的

深度广泛应用。

3.3 全景实时能源动态平衡调度技术

电力系统作为一个有机的整体,系统内电能的生产、输送与使用总量在不断变化,但必须保持瞬间平衡。根据电力生产发、供、用必须同时完成的瞬时平衡规律及电能不能大规模有效存储的特点,需要对电网进行严格的科学调度,保持电网正常运行。电力系统调度当前主要应用 SCADA (supervisory control and data acquisition,数据采集与监视控制)系统,以实现数据采集、设备控制、测量、参数调节以及各类信号报警等各项功能。然而,传统调度系统由于数据处理能力和速度有限,只接受电力系统主参数,而大量细节数据由于不能利用而丢弃^[24]。但随着大规模间歇性能源的接入,电力系统的结构更加复杂多变,海量、分散、异构的实时信息大量涌入数据中心,面对大量实时信息时,传统调度系统的实时性和合理性很难满足要求,无法实现电网的精益化调度。在智能电网环境下,电力调度中心面临的形势如图2所示。

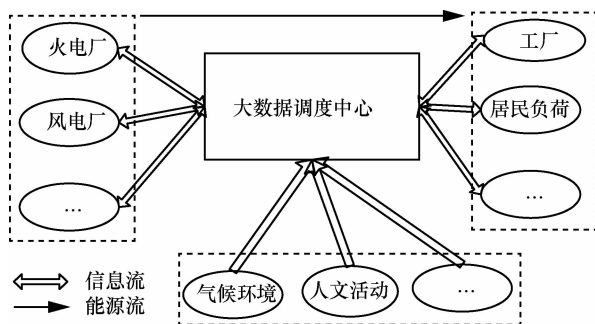


图2 电力调度中心面临的形势

在掌握全系统实时数据的情况下,电力大数据中心将能够更有效地实现全局资源的整合,实现电力供需平衡调度。在电网运行过程中,根据能量的实时需求、负荷建模信息及实时产能信息进行能量自动调配等技术^[25,26]是系统的关键。因此,基于大数据的调度技术关键问题是设计基于大规模多源细节数据的电力系统调度模型和实时流数据处理技术,实现电力系统调度从粗放型向集约型的转变。

4 电力大数据实时分析核心技术

4.1 主要技术挑战

目前主流的 Hadoop 批处理方式可以适应电力大数据的历史数据分析,但 Hadoop 采用先存储后计算的模式,且需要频繁的磁盘操作,难以满足对生产系统的快速响应需求。因此,电力大数据全景实时分析平台面临着严重的性

能 and 安全性挑战,具体而言主要包括如下两个方面。

(1) 利用大规模复杂细节多源数据的电力系统分析算法

在智能电网环境下,PMU、AMI (advanced metering infrastructure, 高级电表架构) 及 IED (intelligent electronic device, 智能电子设备) 提供了大规模可利用的复杂细节数据,这些数据为提高电力系统的操作水平、建模技术和计划调度提供了机遇。通过现代计算机高并行计算平台,设计快速、高精度的安全稳定分析算法,能够提高数据利用率,并保证系统的实时性要求。

(2) 大规模实时流数据调度、分析技术

为了保证监测、调度等关键业务的时效性,大规模实时流数据的接入和分析性能至关重要。将常用数据存储在分布式内存中,并充分利用多核或众核处理器的并行技术和大内存的缓存技术以及高效的调度算法,以保证流数据的实时处理效率。

4.2 内存计算

在电力系统中多项业务处理任务需要高性能计算技术的支撑。而随着现代计算机体系结构的发展,64 bit 计算机理论内存可达 232×4 GB,因此在内存中可以存储更多的数据,利用内存计算技术^[27,28],避免代价很高的磁盘操作,可以有效提高数据处理的性能。内存计算主要用于数据密集型计算的处理,面向数据量大且需要实时分析处理的情况。针对电力大数据价值密度低的特点,依据数据的使用频率,将电力大数据分为热数据和冷数据,热数据存储在内存中,冷数据存储在磁盘上。

目前比较成熟的内存计算平台包括 HANA 和 Spark 等。HANA 是由 SAP 提出的基于内存计算技术的高性能实时数据计算平台,采用的主要技术包括:内存数据读取和处理、行列混合存储、并行计算、数据压缩等。HANA 在辽宁电力公司应用效果明显,部分业务系统的效率提高达几十倍,但系统部署、维护成本很高。

Spark 是一款基于内存计算的开源集群计算系统,与 Hadoop 类似,但适合高效运行的数据分析算法,比如机器学习算法等。其可以在并行操作之间重用工作数据集,从而提高效率。与 Hadoop 相比,性能可获得两个数据量级的提升。电力系统中的数据分析算法可以通过 Spark 来提高效率。比如在电力系统中需要针对设备状态监测数据进行聚类划分,从而实现分类处理,提高工作效率。 k -means 是一种基本的聚类划分算法,在电力系统中应用广泛,基于

Spark 的 k -means 聚类算法实现过程如下。

(1) 读取存储在 HDFS 上的文件块(block)到内存中,每个块转化为一个 RDD,里面包含监测数据的特征量(vector)集合。

(2) 然后对 RDD 进行映射(map)操作,计算每个 vector (point) 对应的聚类编号(class),并输出键值对 (k, v) 为 $(class, (point, 1))$,生成新的 RDD。

(3) 接着在归约(reduce)操作中,对每个新的 RDD 进行混合,相同聚类的数据存放在一起,并在 RDD 内部计算每个聚类中心点。

(4) 最后判断中心点与前一个中心点之间的距离,如果满足要求,则结束,否则从步骤(2)开始,直到满足结束条件。

(5) 最后将输出结果写到 HDFS 中。

Spark 中 k -means 实现过程如图 3 所示,其中①~⑤表示数据点,黑点表示聚类的中心点。

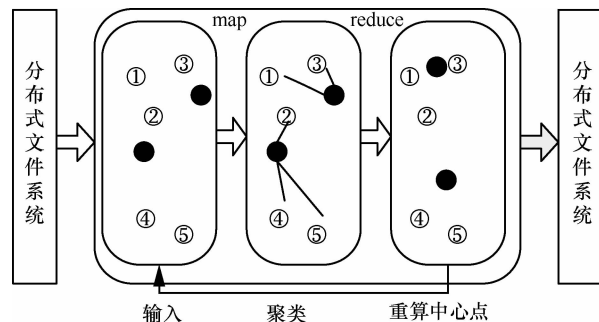


图3 Spark 中 k -means 实现过程

4.3 流式大数据分析技术

在电力系统中,各种监测、采集装置的流数据和极端天气情况下的报警数据构成了电力流式大数据。流式电力大数据具有实时性、易失性、突发性、无序性、无限性等特征^[29],并要求在有限时间内处理完成,不能遗漏数据,因此流式大数据的实时处理对系统提出了更高的要求。电力大数据分析应涵盖低时延、高吞吐量且持续可靠运行的流式大数据计算系统^[30,31]。目前在商业互联网领域应用比较多的流数据分析平台包括 Storm、S4 等,但这些系统对数据处理的时限和安全性要求很难直接应用于电力系统,还有待针对特定行业做进一步完善。

Storm 是具有分布式和高容错的实时计算系统,以简单、高效、可靠的方式处理流式数据,并保证消息不丢失,处理严格有序。其主要特征包括:编程模型简单、支持多种编程语言、作业级容错、水平扩展、快速消息计算等。但其资源分配策略并没有考虑系统拓扑结构,任务调度机制过



于简单,因此很难直接应用于电力大数据处理。在一个 Storm 系统中有两类节点:一个主节点 nimbus、多个从节点 supervisor。3 种运行环境:master、cluster 和 slave。

使用 Storm 处理流数据时,首先要设计流数据处理过程的拓扑(topology)结构,也就是数据处理的先后逻辑关系。比如对状态监测流数据的处理顺序依次为:获取数据、去噪(多种方法)、计算特征量(多种方法)、评价状态等。设计的拓扑结构如图 4 所示。

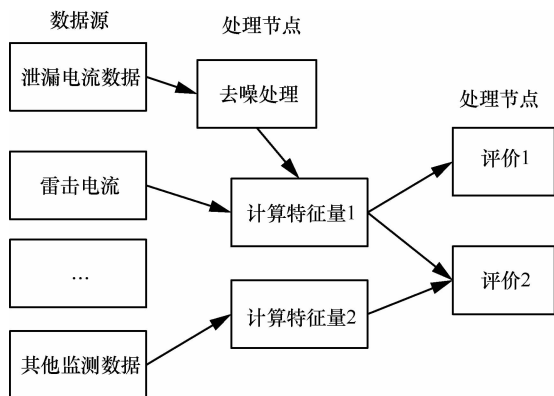


图 4 流数据处理的拓扑结构

4.4 多核/众核并行计算技术

当前电力系统中的主要算法通过串行实现,而随着需要处理数据的增加和计算任务复杂度的提高,算法处理时间不断增长,而通过并行计算技术,将部分算法通过并行来实现,可以显著提高数据分析处理的执行速度^[32-35]。同时,随着现代处理器向多核和众核方向发展,充分利用多线程并行,可以提高电力大数据平台集群中节点的利用率,从而加快数据处理的效率。目前,普通 CPU 已经配置 8 个计算核心,而图形处理器(GPU)拥有上千个计算核心,具有更强大的数据处理能力,两者混合并行计算技术,将是未来的主流计算平台。多核 CPU 与众核 GPU 的结构如图 5 所示。

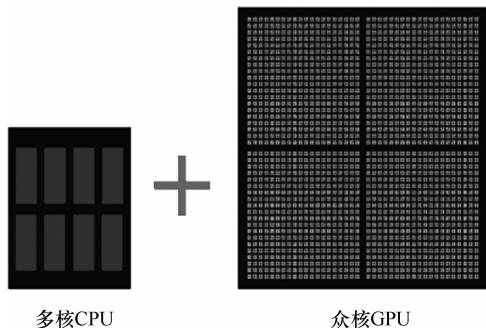


图 5 多核 CPU 与众核 GPU 的结构

近年来,GPU 从图形图像处理领域进入通用计算领域,成为理想的高性能计算平台。GPU 在众多领域的大数据分析中,有效地提升了数据处理的效率。GPU 通过大规模线程并行隐藏内存访问时延,同时支持几百万线程运行,线程间切换代价很低。GPU 的并行采用 SIMD(single instruction multiple data,单指令多数据)并行,同一指令一次操作不同的数据。GPU 适合处理电力系统中数据密集型的计算任务,利用它可以有效提高电力系统中密集数据的分析处理效率。比如处理电力系统安全性分析中的稀疏矩阵与向量乘积的计算过程如下。

首先将稀疏矩阵以偏移量+前缀和+值的形式存储,并通过前缀和记录每行中非零元素的个数及位置信息;然后将常量存放在共享内存中,其中 CUDA 的每个线程块负责稀疏矩阵中的一行和常量乘积;而块内的每个线程负责进行常量与稀疏矩阵本行内对应元素的乘法运算,并求和。大致的运行过程如图 6 所示。

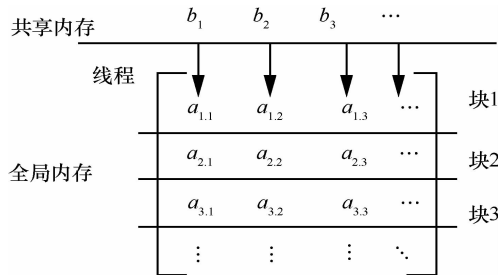


图 6 GPU 下的矩阵与向量乘法运算

4.5 列存储技术

在电力大数据全景实时分析中,应支持多种数据存储模式,比如文件存储、key-value 及结构化数据存储等^[36,37]。而在电力系统中,结构化数据占有较大的比例,比如 PMU 的数据、设备监测的数据及居民用电数据等。结构化数据采用列存储,可以有效地提高数据分析效率。

列存储要求每列单独存储,这样在执行查询时只需要读取与查询相关的列,而不相关的列不用读取,当一个表包含很多列,而查询只涉及较少的列时,可以大量减少磁盘 I/O,从而提高查询性能。另外,由于每一列内的数据单独存储,一列内的数据相似度更高,当采用压缩存储时有更高的压缩率。在列存储数据库中常用的压缩技术主要有:数据字典、位向量和 run-length 3 种方法。

以电能质量监测数据为例,讨论列存储在电力系统中的应用。电能质量监测系统需要存储大量的时序监测数据,这些数据分为不同的监测类别,刻画了电能质量的不

同方面的特性,比如谐波、间谐波及波形、参数数据等。而用户的查询分析主要是针对每个监测点或通道的查询和分析,因此将每一个通道的数据尽量存储在相邻的节点,有助于后续的数据检索分析,通过以监测点+通道号为key,以监测数据为value的键值对方式存储,从而实现监测数据的连续存储,方便后续的数据检索操作。

根据监测数据的监测类别划分为不同的列簇,其中谐波、间谐波及频谱数据划分为一个列簇,数据不是根据次数分别存储,而是采用组合压缩存储;各项电压电流波形数据划分为一个列簇;其他监测数据为另一个列簇;闪变数据也组成一个列簇。通过划分列簇,利用数据的相关性提高读取效率。其存储结构在 HBase 数据库上的表现大致如图 7 所示。

5 电力大数据分析平台

当前,电力系统虽然已经建成了规模庞大的数据中心,但主要存在如下两方面的问题:不同的业务系统被分割为不同的“信息孤岛”,彼此之间数据不能共享,不能从系统全景的角度分析解决问题;为了保证系统的实时性,大量有助于提高计算精度和准确率的信息并不接入数据中心,从而造成系统因为无法协调全部可用资源而导致整体运行效率和数据利用率低。通过建设智能电网全景实时大数据分析平台,有助于上述问题的解决。

另一方面,电力系统作为人类有史以来创造的最复杂系统之一,具有系统规模庞大、整体性强、复杂度高、安全稳定运行要求高、电能光速传输并不能大规模有效存储等特性,决定了电力大数据全景实时分析将面临着两方面的主要挑战:大数据分析挖掘技术与电力系统理论的深度融合,从数据驱动的角度探讨解决系统所面临问题的思路,促进大数据在电力系统中的应用落地;满足一定可靠性和时间约束的大数据(批处理和流数据)全景实时分析技术,

快速从系统整体层面分析解决问题、合理调度资源,满足电力系统高可靠性的要求。

参考当前电力系统面临的各项挑战和大数据实时分析主流技术,电力大数据全景实时分析平台采用分层架构,层与层之间采用松耦合模式。结合电力系统运行特点及借鉴主流开源系统,架构应包括如下几个层次。

最底层是数据存储层,为保证平台的可扩展性采用分布式文件存储系统,并设计以数据相关性和负载均衡策略为基础的数据分布方案。在分布式文件系统之上,结合高性能列存储技术,进一步优化存储代价和性能。本层可以采用 Hadoop 的分布式文件系统存储方案并结合 HBase 的列存储技术。

其次是流式大数据处理和分布式内存层,流式大数据处理模块用于响应电力系统中的实时流数据,关键是流数据的任务调度技术;分布式内存用于存储使用频率高的热数据,为上层计算提供数据缓存,减少磁盘访问代价。本层可以采用 Storm 和 Spark 技术分别实现流数据处理和内存计算。

然后是整合计算层,设计电力系统中基于大数据的并行分析算法框架,并考虑利用多核、大内存和 GPU 等现代计算机的硬件体系架构特征的核心算法库,将电力系统中常用的分析算法封装为库函数供上层高级应用调用。

最后一层是应用层,包括可靠性分析、状态监测和能源调度等各种业务应用系统,并直接和最终用户交互。

电力大数据全景实时分析平台层次体系结构如图 8 所示。

通过前置通信集群不间断采集设备及系统的事件、状态、预警等多数据源信息,形成流数据,完成高效实时数据收集功能。然后,将收集的数据传输给流计算模块,完成数据混合并实时预处理,基于 CEP(complex event processing,复杂事件处理)的快速发现和异常数据过滤,分析状态并

row-key 监测点+通道号	TimeStamp 时间戳	ColumnFamily 谐波信息			ColumnFamily 其他数据			ColumnFamily 波形数据		
		谐波	间谐波	频谱	功率	频率	基波	A相电压	B相电压	C相电压
0001.001										
			#####							
0001.002										
0001.003										

图 7 HBase 电能质量监测数据存储结构

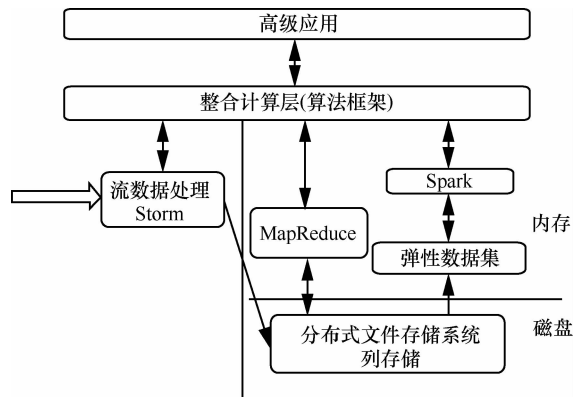


图 8 电力大数据分析平台结构

实时反馈给用户。最后将数据存储在分布式文件系统中进行持久化存储,使用内存计算技术进行数据快速分析与预警计算,保证电力系统中数据分析实时性的要求。

电力大数据分析平台的特色包括如下两个方面。

(1)大数据与电力系统深度融合的复杂电网分析、监控、调度技术和理论

随着电力大数据规模不断膨胀以及电网结构、运行模式复杂多变,从大数据角度探讨复杂电网分析、监控、调度技术和理论,促进大数据在电力系统中的深度应用,发现挖掘电力大数据的价值。

(2)面向电力行业特定领域的大数据全景实时分析技术

针对电力行业对数据分析时效性、全景性的高要求,面向特定行业领域的大数据实时分析技术,提供整个电力系统的全景视图和实时动态监控、调度技术,促进大数据在特定行业应用中落地。

6 结束语

本文主要综述了电力大数据全景实时分析面临的挑战及实现技术,并探讨平台的体系架构。然而,建设电力大数据中心需要一个较长的时间周期和各方的通力配合,其中的关键技术和挑战也需要进一步深入研究。结合电力系统应用需求,开发部署面向电力行业的大数据中心,将推动大数据在电力系统中的研究和发展。

参考文献:

- [1] 宋亚奇,周国亮,朱永利. 智能电网大数据处理技术现状与挑战[J]. 电网技术, 2013, 37(4): 927-935.
- SONG Y Q, ZHOU G L, ZHU Y L. Present status and challenges of big data processing in smart grid[J]. Power System

Technology, 2013, 37(4): 927-935.

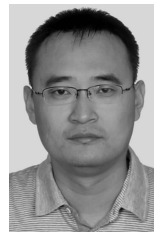
- [2] 王德文,宋亚奇,朱永利. 基于云计算的智能电网信息平台[J]. 电力系统自动化, 2010(22): 7-12.
- WANG D W, SONG Y Q, ZHU Y L. Information platform of smart grid based on cloud computing [J]. Automation of Electric Power Systems, 2010(22): 7-12.
- [3] 中国电机工程学会信息化专委会.中国电力大数据发展白皮书[R]. 2013.
- Electric Power Informatization Special Committee. China power big data development white paper[R]. 2013.
- [4] 朱征,顾中坚,吴金龙,等. 云计算在电力系统数据灾备业务中的应用研究[J]. 电网技术. 2012, 36(9): 43-50.
- ZHU Z, GU Z J, WU J L, et al. Application of cloud computing in electric power system data recovery[J]. Power System Technology, 2012, 36(9): 43-50.
- [5] 沐连顺,崔立忠,安宁. 电力系统云计算中心的研究与实践[J]. 电网技术, 2011, 35(6): 170-175.
- MU L S, CUI L Z, AN N. Research and practice of cloud computing center for power system[J]. Power System Technology, 2011, 35(6): 170-175.
- [6] 王广辉,李保卫,胡泽春,等. 未来智能电网控制中心面临的挑战和形态演变[J]. 电网技术, 2011, 35(8): 1-5.
- WANG G H, LI B W, HU Z C, et al. Challenges and future evolution of control center under smart grid environment [J]. Power System Technology, 2011, 35(8): 1-5.
- [7] 刘树仁,宋亚奇,朱永利,等.基于Hadoop的智能电网状态监测数据存储研究[J]. 计算机科学, 2013, 40(1): 81-84.
- LIU S R, SONG Y Q, ZHU Y L, et al. Research on data storage for smart grid condition monitoring using Hadoop [J]. Computer Science, 2013, 40(1): 81-84.
- [8] RUSITSCHKA S, EGER K, GERDES C. Smart grid data cloud: a model for utilizing cloud computing in the smart grid domain[C]// The 1st International Conference on Smart Grid Communications, October 4-6, 2010, Gaithersburg, USA. New Jersey: IEEE Press, 2010: 483-488.
- [9] CHRISTOPHE B. The smart grid: hadoop at the Tennessee Valley Authority(TVA) [EB/OL]. [2013-02-08].<http://www.cloudera.com/blog/2009/06/smart-grid-hadoop-tennessee-valley-authority-tva/>.
- [10] KAWASOE S, IGARASHI Y, SHIBAYAMA K, et al. Examples of distributed information platforms constructed by power utilities in Japan [C]//CIGRE Symposium, August 26-31, Paris, France. [S.l.:s.n.], 2012: 134-306.
- [11] KENNETH P B, LAKSHMI G, ROBERT V R, et. al. Running smart grid control software on cloud computing architectures[C]// Workshop on Computational Needs for the Next Generation Electric Grid, April 19-20, 2011, Ithaca, USA. New York: Department of Energy's Sci Tech Connect, 2011:1-28.

- [12] KEZUNOVIC M, LE X, GRIJALVA S. The role of big data in improving power system operation and protection [C]//Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid (IREP), August 25-30, 2013, Rethymnon, Greece. New Jersey: IEEE Press, 2013: 1-9.
- [13] JIAN Y, POORVA S, IAN G, et al. Large-scale data challenges in future power grids[C]//7th International Symposium on Service Oriented System Engineering, March 25-28, 2013, Redwood City, CA, USA. New Jersey: IEEE Press, 2013: 324-328.
- [14] 赛迪网. 维斯塔斯利用 IBM 大数据分析实现“智慧风能”[EB/OL]. [2012-03-23]. <http://news.ccidnet.com/art/1032/20120323/3709391.html>.
CCID net. The realization of “wisdom wind” using IBM big data analysis in Vestas [EB/OL]. [2012-03-23]. <http://news.ccidnet.com/art/1032/20120323/3709391.html>.
- [15] 李传栋, 房大中, 杨金刚, 等. 大规模电网并行潮流算法[J]. 电网技术, 2008, 32(7): 34-39.
LI C D, FANG D Z, YANG J G, et al. New research on parallel power-flow calculation for large-scale power system[J]. Power System Technology, 2008, 32(7): 34-39.
- [16] 刘洋, 谢开贵, 周家启, 等. 大电力系统可靠性评估高性能计算平台设计与实现[J]. 电力系统自动化, 2006, 30(18): 1-8.
LIU Y, XIE K G, ZHOU J Q, et al. Design and implementation of HPC platform for bulk power system reliability evaluation[J]. Automation of Electric Power Systems, 2006, 30(18): 1-8.
- [17] 桂勋. 基于递推算法的海量 COMTRADE 数据计算并行化[J]. 电力系统自动化, 2014, 38(1): 86-91.
GUI X. Calculation parallelization for massive COMTRADE data based on recursive algorithms [J]. Automation of Electric Power Systems, 2014, 38(1): 86-91.
- [18] 何剑, 程林, 孙元章. 电力系统运行可靠性最优控制[J]. 中国电机工程学报, 2010, 30(7): 15-21.
HE J, CHENG L, SUN Y Z. Optimal control of power system's operational reliability[J]. Proceeding of the CSEE, 2010, 30(7): 15-21.
- [19] 江修波, 吴文宣, 陈祥伟. 区间分析法在电力变压器全寿命周期成本模型中的应用[J]. 电力自动化设备, 2011, 31(9): 50-54.
JIANG X B, WU W X, CHEN X W. Transformer life cycle cost model based on interval analysis [J]. Electric Power Automation Equipment, 2011, 31(9): 50-54.
- [20] 刘吉臻, 杨光军, 谭文, 等. 基于数据驱动的电站燃烧稳定度综合评价[J]. 中国电机工程学报, 2007, 27(35): 1-6.
LIU J Z, YANG G J, TAN W, et al. Synthetic evaluation on the degree of combustion stability in power station based on data-driven[J]. Proceedings of the CSEE, 2007, 27(35): 1-6.
- [21] 张广斌, 束洪春, 于继来. 利用广义电流模量的行波实测数据半监督聚类筛选[J]. 中国电机工程学报, 2012, 32(10): 150-159.
ZHANG G B, SHU H C, YU J L. Travelling wave field data contingency screening based on semi-supervised clustering using generalized current modal components[J]. Proceedings of the CSEE, 2012, 32(10): 150-159.
- [22] 陈铁华, 陈启卷. 模糊聚类分析在水电机组故障诊断中的应用[J]. 中国电机工程学报, 2002, 22(3): 43-47.
CHEN T H, CHEN Q J. Fuzzy clustering analysis based vibration fault diagnosis of hydroelectric generating unit[J]. Proceedings of the CSEE, 2002, 22(3): 43-47.
- [23] 王一, 何奔腾, 王慧芳. 基于全寿命状态的设备可靠性研究[J]. 电网技术, 2011, 35(8): 207-211.
WANG Y, HE B T, WANG H F. Research on equipment reliability based on life cycle states[J]. Power System Technology, 2011, 35(8): 207-211.
- [24] 曹军威, 万宇鑫, 涂国煜, 等. 智能电网信息系统体系结构研究[J]. 计算机学报, 2013(1): 143-167.
CAO J W, WAN Y X, TU G Y, et al. Information system architecture for smart grids [J]. Chinese Journal of Computers. 2013 (1): 143-167.
- [25] 王蓓蓓, 刘小聪, 李扬. 面向大容量风电接入考虑用户侧互动的系统日前调度和运行模拟研究 [J]. 中国电机工程学报, 2013, 36(22): 35-44.
WANG B B, LIU X C, LI Y. Day-ahead generation scheduling and operation simulation considering demand Response in large-capacity wind power integrated systems [J]. Proceeding of the CSEE, 2013, 36(22): 35-44.
- [26] 黄海涛. 居民分时阶梯电价联合优化模型研究[J]. 电网技术, 2012, 36(10): 253-258.
HUANG H T. A Joint optimization model of residential time-of-use rlock electricity rate [J]. Power System Technology, 2012, 36(10): 253-258.
- [27] 金海, 廖小飞, 叶晨成. 内存计算: 大数据处理的机遇与挑战[J]. 中国计算机学会通讯, 2013, 9(4): 40-46.
JIN H, LIAO X F, YE C C. In-memory computing: The opportunities and challenges of big data processing [J]. Communication of China Computer Federation, 2013, 9 (4): 40-46.
- [28] PARK K H, PARK S K, SEOK H. Efficient memory management of a hierarchical and a hybrid main memory for MNMATE platform[C]//International Workshop on Programming Models and Applications for Multicores and Manycores. March 12-16, 2012, New York, USA. New York: ACM Press, 2012: 83-92.
- [29] 孙大为, 张广艳, 郑纬民. 大数据流式计算: 关键技术及系统实例[J]. 软件学报, 2014, 25(4): 839-862.

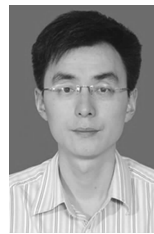


- SUN D W, ZHANG G Y, ZHENG W M. Big data stream computing: technologies and instances [J]. Journal of Software, 2014, 25(4): 839-862.
- [30] 赵加奎, 杨国凤, 沐连顺, 等. 数据流技术在电网自动化中的应用研究[J]. 电网技术, 2011, 35(8): 6-11.
- ZHAO J K, YANG G F, MU L S, et al. Research on the application of the data stream technology in grid automation[J]. Power System Technology, 2011, 35(8): 6-11.
- [31] 张之哲, 李兴源, 程时杰. 智能电网统一信息系统的框架、功能和实现[J]. 中国电机工程学报, 2010, 30(34): 1-7.
- ZHANG Z Z, LI X Y, CHENG S J. Structures, functions and implementation of united information system for smart grid [J]. Proceedings of the CSEE, 2010, 30(34): 1-7.
- [32] 郭春辉. 基于 GPU 的电力系统并行计算的研究 [D]. 济南: 山东大学, 2013: 18-34.
- GUO C H. Research on parallel computing of electric power system based on GPU [D]. Jinan: Shan Dong University, 2013: 18-34.
- [33] 李鹏, 丁承第, 王成山, 等. 基于多核心处理器的分布式发电微网系统暂态并行仿真方法 [J]. 中国电机工程学报, 2013, 33(16): 171-178.
- LI P, DING C D, WANG C S, et al. A parallel algorithm of transient simulation for distributed generation system based on multi-core CPU [J]. Proceeding of the CSEE, 2013, 33(16): 171-178.
- [34] 唐聪, 严正, 周挺辉. 基于图形处理器的广义最小残差迭代法在电力系统暂态仿真中的应用 [J]. 电网技术, 2013, 37(5): 1365-1371.
- TANG C, YAN Z, ZHOU T H. Application of graph processing unit-based generalized minimal residual iteration in power system transient simulation [J]. Power System Technology, 2013, 37(5): 1365-1371.
- [35] 马发勇, 厉启鹏, 马志斌, 等. 电力调度 SCADA 系统中历史数据压缩及存储策略[J]. 电网技术, 2014, 38(4): 1109-1114.
- MA F Y, LI Q P, MA Z B, et al. The research of historical data compression and storage strategy in power dispatch SCADA system[J]. Power System Technology, 2014, 38(4): 1109-1114.
- [36] 闫湖, 周薇, 李立新, 等. 基于分布式键值对存储技术的 EMS 数据库平台[J]. 电网技术, 2012, 36(9): 162-167.
- YAN H, ZHOU W, LI L X, et al. Database platform of EMS system based on distributed key-value pair storage technology[J]. Power System Technology, 2012, 36(9): 162-167.
- [37] ZHONG Y, SUN S, LIAO H, et al. A novel method to manage very large raster data on distributed key-value storage system[C]//The 19th International Conference on Geoinformatics, June 24-26, 2011, Shanghai, China. New Jersey: IEEE Press, 2011: 1-6.

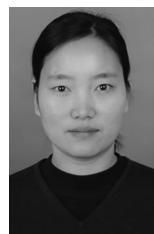
[作者简介]



周国亮(1978-),男,博士,国网冀北电力有限公司技能培训中心副教授,主要研究方向为电力大数据分析、智能电网等。



吕凛杰(1978-),男,国网冀北电力有限公司技能培训中心讲师,主要研究方向为电力通信技术。



王桂兰(1979-),女,华北电力大学信息与网络管理中心讲师,主要研究方向为风机故障定位与检测。