

基于大数据的智能变电站二次状态监测系统研究

张 艺¹ 李大中¹ 郑永康²

(1. 华北电力大学 河北 保定 071000; 2. 国网四川省电力公司电力科学研究院 四川 成都 610041)

摘 要: 在传统变电站往智能变电站这一过渡过程中,随着变电站网络结构改变所带来的监测难点体现在难以在智能变电站二次系统状态监测的庞大数据库中,进行有效信息的高效提取以及对应故障定位与溯源。在此背景下,首先探讨学习变电站网络架构的改变;其次在智能变电站二次状态监测系统的基础上分析了其监测概念以及监测信号;最后为了解决当前智能变电站海量数据的处理问题,分析总结基于 Hadoop 数据处理平台的优势以及大数据在智能变电站中的应用发展。

关键词: 智能变电站; 大数据; 二次状态

中图分类号: TM76 文献标志码: A 文章编号: 1003-6954(2019)04-0043-06

DOI:10.16527/j.cnki.cn51-1315/tm.2019.04.010

Research on Secondary State Monitoring System in Smart Substation Based on Big Data

Zhang Yi¹, Li Dazhong¹, Zheng Yongkang²

(1. North China Electric Power University, Baoding 071000, Hebei, China;

2. State Grid Sichuan Electric Power Research Institute, Chengdu 610041, Sichuan, China)

Abstract: In the process of transition from traditional substation to smart substation, the monitoring difficulties brought by the change of substation network structure are reflected in the difficulty of efficient extraction of effective information and corresponding fault location and traceability in the huge database of secondary system condition monitoring in smart substation. Under this background, firstly, the change of substation network structure is discussed as well as the intelligence. Secondly, based on substation secondary condition monitoring system, the concept of monitoring and monitoring signals are analyzed. Finally, in order to solve the problem of processing massive data in smart substation, the advantages of Hadoop data processing platform, and the application and development of big data in smart substation are analyzed and summarized.

Key words: smart substation; big data; secondary state

0 引 言

中国变电站自动化系统(substation automation system)从20世纪90年代开始至今已经到了全面推广的阶段^[1]。传统智能变电站中各保护装置之间存在较多硬开入连线,导致二次回路接线比较复杂,可靠性不高。相比之下,传统依靠电缆传输的二次回路被智能变电站的网络通信取代,物理信号被数字信号所取代,实现了二次设备的网络化^[2-5]。

目前国内外市场上对于智能变电站二次回路的研究,主要还是根据网络分析仪的报文进行异常原因分析,缺乏直观有效的手段对网络信息中的故障

特征进行定性分析,无法形成完整的二次回路预警和分析策略^[4]。智能变电站电力状态监测数据呈现以下特点:数据量爆炸性增长,规模急速扩大;复杂的数据结构;多样化的历史与实时监测信息(设备信息、试验数据、噪声数据等)^[5];数据种类分布跨度大,不仅仅包含结构化与非结构化数据,而且各型数据在平台的查询与处理方式及要求也不大一致;各平台的交互通信也存在较大的难度^[5-6]。

下面对智能变电站的网络架构与二次状态监测的概念、Hadoop 大数据处理技术的优势进行了总结;分析验证了大数据技术在智能变电站海量状态监测数据下,基于 Hadoop 平台的储存与读写优势。

1 智能变电站网络结构

早期的智能化变电站利用 IEC 61850 协议将变电站设备模型化。这个时期的智能变电站的网络化实现仅仅存在于间隔层与站控层,过程层设备的信息交互仍然依靠传统的模拟信号^[7]。网络结构如图 1。

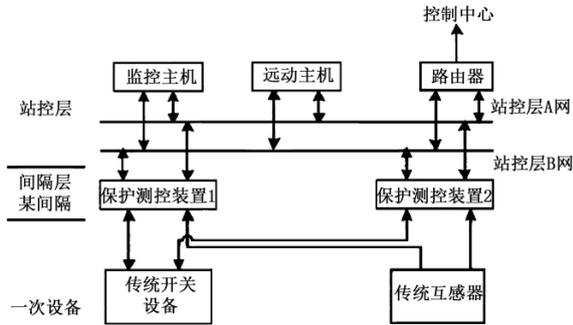


图 1 早期智能变电站网络结构

基于早期的智能变电站的不足,理想的智能变电站为了将信息采集到输出的整个过程全部数字化,采用智能开关、EVCT 等智能设备来满足设备智能化、网络通信协议一体化、运维管理全程自动化等基本性能需求^[7]。但是考虑在实际应用中的设备条件,目前只能通过智能终端与传统的二次开关进行结合链接,来填补实际的技术缺陷以达到较高的网络数字化程度。工程实践中,采用“直跳直采”方案来实现过程层信息共享并加强了测控保护装置的可靠性。具体网络结构如图 2 所示。

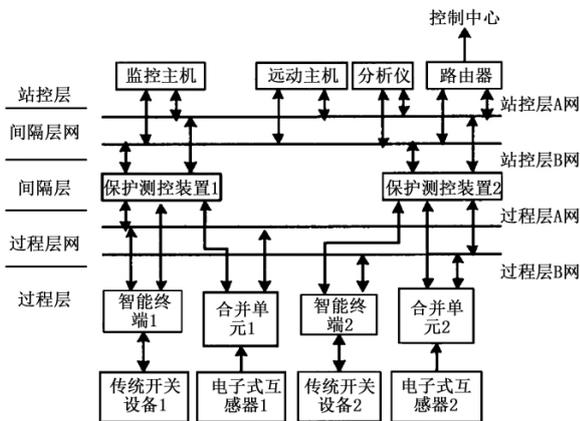


图 2 智能变电站实用网络结构

2 智能变电站二次系统的在线监测

智能变电站中,信息采集输入时需要光纤以太网传输的采样值报文,保护动作出口信息由光纤以

太网传输的 GOOSE 开关量取代了跳闸、重合闸接点动作;二次设备一般都具有在线自检功能及通讯功能,在线监测的实现就是利用装置本身自检及装置之间的互相监测^[8-11]。

文献[12]认为继电保护装置的电流、电压、光纤链路状态,装置的遥信遥控等 GOOSE 状态,直流逆变电源状态等都应该纳入装置的主要监测对象,甚至本身自检的 FLASH 擦写次数、RAM 是否出错等信息也纳入监控范围。文献[13]在考虑将含有端口连接状态、端口双工模式、速率和吞吐量等表征通信网络运行状态参数的网络设备(交换机等)也纳入监控范围,以更好地完成状态监测、故障诊断及故障定位等工作。总而言之,智能变电站二次设备状态监测系统正在往全面智能化方向发展^[14]。智能变电站的二次监测涵盖内容十分广泛,不仅仅局限于某些具体的信号丢失或者设备故障,具体的监测信息概念如图 3 所示。

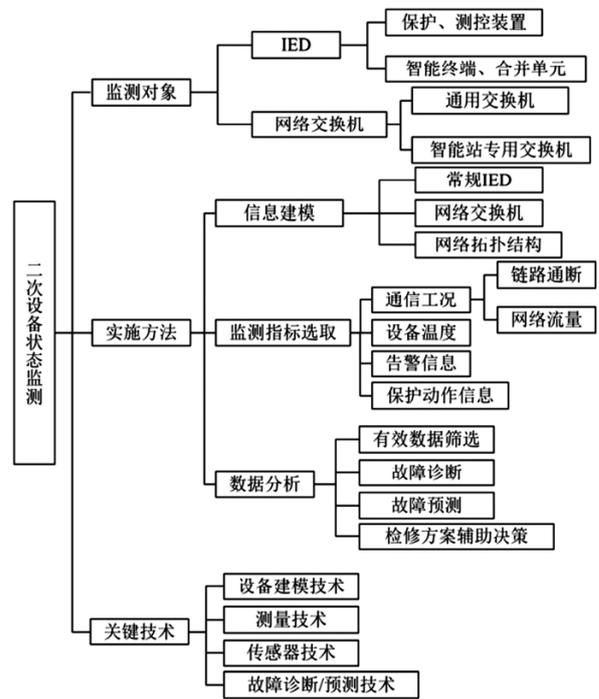


图 3 二次设备状态监测概念

3 大数据技术在智能变电站海量信息中的应用

智能变电站二次系统监测包含海量监测数据信息。其中,单个变电站公用信号分类图中包含 PMU (制造主柜相量测量单元)、TMU (时钟检测单元) 装置电源失电、PDM (局部放电在线监测) 等 60 多种

告警信号,该告警信息具体可分为14类,遥信、遥控、SOE(时间顺序记录)、开关刀闸动作等^[15-16]。因此,将大数据技术应用于变电设备状态监测数据的存储与管理,以此来适应电网安全经济的运行和用户供电可靠性的要求。

3.1 Hadoop 的提出

大数据云计算原理简单,操作性强,可靠性高。文献[15]提出一种基于大数据平台的智能变电站二次装置故障追踪方法,将故障诊断数据源延展至变电层,利用Spark作为大数据处理工具对各类故障信息进行处理,实现故障源的准确定位。文献[16]提出一种基于Hadoop框架的智能变电站数据管理系统。为了对数据更好地进行分布式管理,HDFS和HBase数据库发挥着至关重要的作用,采用并行计算框架MapReduce作为数据查询分析的计算模式,还致力于将真实运维数据进行系统测试,在线对比该数据管理系统的存储、查询、读写延迟对比、分布式索引对比等性能参数。

3.2 Hadoop 处理技术

Hadoop是一种开源分布式的计算框架,其扩展性、利用率、可靠性等优点使其在数据处理领域广泛应用。Hadoop分为分布式文件系统HDFS(Hadoop Distributed File System)和MapReduce,其中:HDFS是其独有的分布式文件系统,容错性与可靠性较高,在文件存储及校验中有一定的优势;MapReduce作为Hadoop的并行计算框架,对于1TB以上的数据集有着不容忽视的并行计算优势。基于变电站大数据背景下,Hadoop技术带来了新的解决思路,可提供高性能的计算环境。

3.3 基于Hadoop的智能变电站监测数据处理平台

目前大多设备状态监测系统主要分为数据获取层、数据存储与管理层、数据访问层^[16]。其系统结构如图4所示。

原始数据信息经过第一功能层获取,进行ETL(抽取、转化、清洗、装载)至数据库加载。图5展示了大数据技术下变电设备状态监测平台架构。

1) 数据采集层中状态接入网关机(CAG)接收来自获取数据的传感器和状态接入控制器(CAC)以Web服务方式传送的信号。所采集的信息需要Sqoop这一开源工具来对电力设备信息、异常告警数据等结构数据以及图像、视频等非结构数据进行

ETL,进而进行存储。

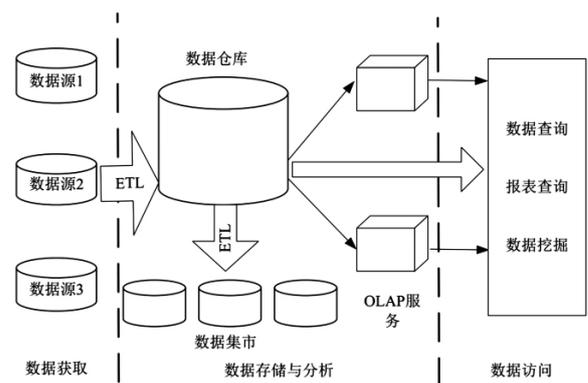


图4 传统变电设备状态监测平台架构

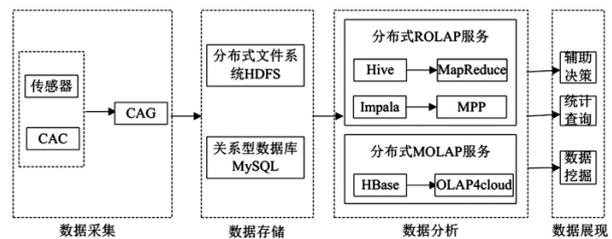


图5 大数据技术下变电设备状态监测平台架构

2) 数据存储层根据文件系统(HDFS)和数据库(MySQL)的优势进行整合。

3) 数据分析层可根据具体的方案需求进行不同的分布式ROLAP服务和分布式MOLAP服务设计。Hive和Impala都是ROLAP下的分布式服务,可共用数据库和存储资源池等。Hive适用于长时间的批处理数据分析,但是需要通过MapReduce分解任务才能完成操作。Impala适合于实时数据分析,通过使用Select和统计函数直接从HDFS抽取数据进行分析处理^[17]。不同于前者,HBase OLAP隶属于分布式MOLAP服务,引擎为OLAP4cloud。它将维信息完全压缩至事实表中,使用计算数据立方体的方法同时加快寻址的速度。

4) 数据展现层经过数据的统计查询,加以运维经验的辅助决策以及最后的数据挖掘等功能模块,为实现智能变电站二次状态监测的全面管理提供了有效保障。

4 智能变电站云实验平台及性能测试

实验平台由10台PC机(INTEL core i5 3.5GHz CPU 4GB内存,500G硬盘)搭建Hadoop分布式集群,分别安装Centos虚拟机、Apache Hadoop云平

台。Datanode 各个节点的空间、大小、使用率等将会在 HDFS 管理界面中显示。利用 TearSort、Sort、TestDFSIO、YCSB 等辅助工具对智能变电站二次状态监测数据性能进行测试。

4.1 监测数据分类排序能力测试

Sort 是 Hadoop MapReduce 中一种衡量分布式数据处理框架数据处理能力的工具。将所有需要处理的数据划分成 N 个数据模块 (Hash 处理),接着每个 Map task 对单独的数据模块进行局部排序之后,Reduce task 将对所有数据进行全部排序处理。为了改善 sort 分类在 Reduce 阶段无法并行的弊端, TearSort 在 Map 这一阶段经过 Map task 划分需处理数据的 M (reduce task 数量为 M) 个模块,设定第 i ($i > 0$) 个模块中的所有数据要均要大于第 $i + 1$ 个;而 Reduce 处理阶段中,第 i 个经过 Reduce task 进行排序处理后的所有 Map task 的第 i 个模块所产生的结果均会大于 $i + 1$ 个,最后将 $1 \sim M$ 个 Reduce task 所有排序结果按照顺序输出,得到最终的数据结果,完成排序操作。图 6 为 tearsort 算法结构图。

为了验证该实验平台数据处理的优越性,首先利用 Random Writer 自动生成随机数据,将 Map 方式与 MapReduce 进行 10:1 的运行作业,每一个 Map 会生成大概 10 GB(二进制)的不同幅值和键长的数据。在此 Hadoop 软件平台上进行 tearsort 排序,取 10 次实验结果数据,如表 1 所示。

表 1 tearsort 排序测试

实验次数	时间/s
1	324.12
2	341.25
3	299.88
4	387.18
5	352.56
6	302.25
7	282.14
8	345.29
平均值	329.33

上述结果表明,此在线监测平台在数据处理上具有明显优势,拥有良好的数据计算能力。

4.2 读写能力测试

智能变电站二次设备监测系统中信息量庞大,海量的告警信息以及各节点的采样值在日常监测系统中给监测计算机载体造成了严重的负担。仅告警信息的统计就分为事故级、一般级、预告级三大等级,如表 2 所示。

表 2 告警信息分类表

告警等级	告警信息
预告级	U/I 采样异常、无效; 局部在线监测、复归等信息
一般级	馈线两端重合闸误动; 智能终端通信链路断链; 合并单元网络通信链路出错等;
事故级	保护动作/启动,主变压器低智能终端启动

智能变电站二次状态监测系统的监测难点在于监测设备数量多,信息数据结构复杂。为了保证智

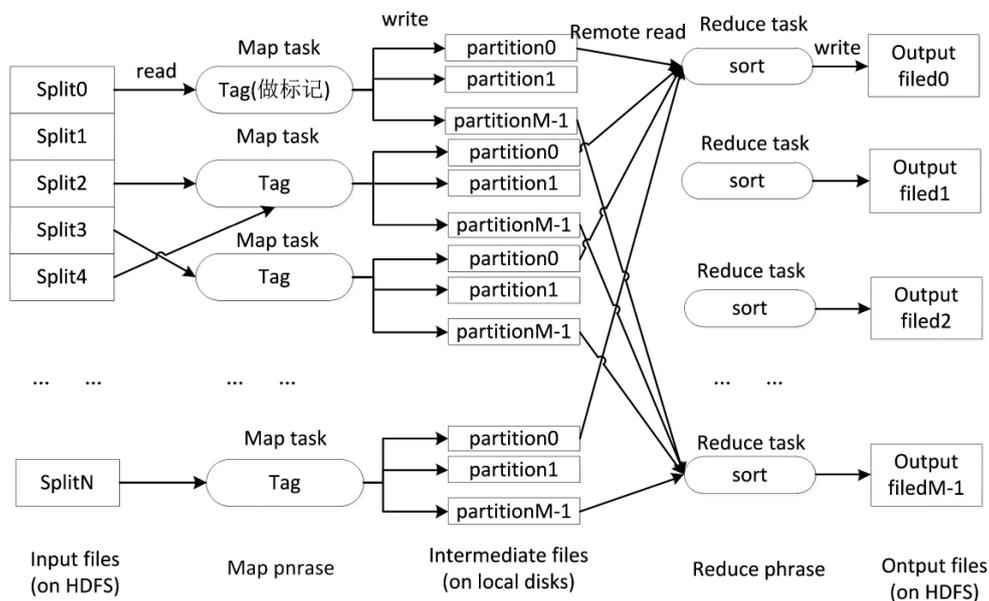


图 6 tearsort 算法结构

能电网的稳定运行,所以对试验软件平台的基准性能测试非常必要的。一般常用的测试工具是TestDFSIO,通过生成数据提交、统计整个平台运作时间进而完成对整个I/O性能测试。基于控制变量法原则,分别选取文件数量与文件大小这两个变量进行试验,具体仿真结果见表3。

表3 I/O性能测试

性能测试	文件数量/个	文件大小/MB	运行时间/s
write	10	1500	351.45
write	10	3000	546.28
write	10	4500	685.25
write	30	1500	396.85
write	30	3000	438.36
write	30	4500	620.69
read	10	1500	205.35
read	10	3000	241.67
read	10	4500	283.35
read	30	1500	224.15
read	30	3000	276.22
read	30	4500	343.69

根据试验结果可以得到以下信息:

相比于传统数据管理平台(人工长期的运维经验)基于Hadoop的在线监测数据处理系统在读写吞吐量测试上具有了明显的优势。

当数据大规模地增加时,实验平台读写测试所需要的时间增加,访问效率变低。

传统数据平台由于测试信息的大规模增加、负荷急剧增大导致平台系统性能降低,无法满足目前智能变电站的监测需求。总的来说,基于Hadoop数据处理软件平台中集群在大数据驱动下展现了强大的处理能力。

4.3 在线监测数据查询测试

经过Hadoop平台的吞吐读写测试,在线监测的数据信息存储于HDFS中,为了提高该试验平台的处理效率,满足目前智能变电站信息的查询需求,需要在存储平台上建立对应的搜索引擎,并且设置相

应的关键性索引。查询过程的结构如图7所示,可以在存储文件更新状态下进行监测,有效缓解负载平台的运作压力。

基于HBase变电站设备在线监测数据概念,索引主要是由行健、时间戳和列族组成的^[18]。表4是以二次设备中合并单元上传到站控层的SV通信状态中报文处理结果的监测举例:行健设置为可监测类型的编码,例如021001代表SV通信链路状态,021002代表是其根据SV报文计算出来的电流有效值;被监测设备码根据国家统一规定,由3段共17位字符组成,其中前两位为省公司标识,M表示固定字符,后14位表示具体流水号。基于HBase低冗余性与一致性不能满足目前智能变电站二次状态监测的需求,当数据规模超过一定数量级时,这种一对一的查询方式给大数据平台的数据处理上带来了不便。为了提高数据信息的查询效率,将以组合索引的方式进行拼接,如表5所示。

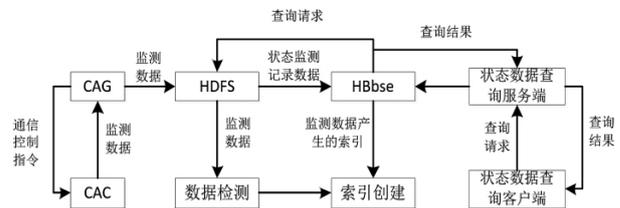


图7 查询过程结构

将监测时间aaaa-MM-dd与具体的监测量组合在一起,如表5中021002_aaaa-MM-dd就表示检测时间与采样节点的电流有效值结合,这种组合方式最大的便利在于可以根据具体需求进行改变,减少索引时间提高了查询效率,用户也能根据自身查询条件建立组合索引,最终获得数据结果集。基于智能变电站二次状态监测背景,在Hadoop实验平台上建立索引并进行测试,对通信链路中采样节点的电流有效值查询进行性能测试对比,如图8所示。

表4 基于HBase变电站设备在线监测数据概念

行健	时间戳	列族		
		被监测设备码	日期	I(有效值)/A
021001	T1	23M00000005482672	2016-02-18 09:30:00	
021002	T2		2016-02-18 09:30:00	95
021002	T3	23M00000005482672	2016-02-18 09:30:00	99

表5 组合索引优化

行健	列族		
	被监测设备码	检测装置标识	相别
021002_aaaa-MM-dd	17位字符	17位字符	XX

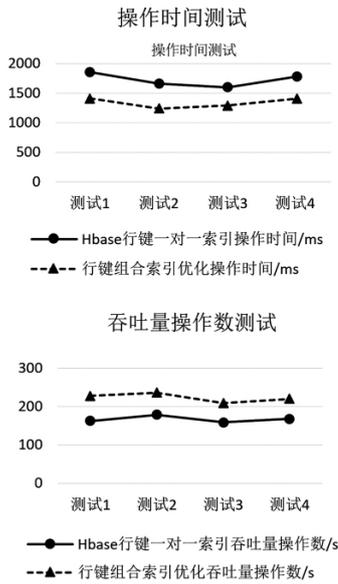


图8 索引优化性能测试对比

根据表5中改进的行键组合优化后,不仅在吞吐量每秒操作数上有了大幅度的提高,而且平台的操作时间降低了不少,提高了工作效率,为智能变电站二次状态监测中的海量数据处理提供了新的思路。

5 结语

在智能变电站监测数据的大背景下,基于Hadoop大数据处理平台对信息的预处理、数据的存储吞吐、查询延迟的测试以及数据可靠性与延展性的保证都进行了优化,这为后续智能变电站二次系统的状态监测和故障定位研究工作提供了有力的数据支持,提高了智能变电站二次监测系统的稳定性。

参考文献

[1] 徐长宝,庄晨,蒋宏图.智能变电站二次设备状态监测技术研究[J].电力系统保护与控制,2015,43(7):127-131.

[2] 高翔,张沛超.数字化变电站的主要特征和关键技术[J].电网技术,2006,30(23):67-71.

[3] 王治民,陈炯聪,任雁铭,等.网络通信记录分析系统在数字化变电站中的应用[J].电力系统自动化,2010,34(14):92-95.

[4] 叶远波,孙月琴,黄太贵,等.智能变电站继电保护二次回路在线监测与故障诊断技术[J].电力系统保护与控制,2016,44(20):148-153.

[5] 王德文,肖磊,肖凯.智能变电站海量在线监测数据处理方法[J].电力自动化设备,2013,33(8):142-146.

[6] 张怀宇,朱松林,张扬,楼其民,张亮.输变电设备状态检修技术体系研究与实施[J].电网技术,2009,33(13):70-73.

[7] 阴玉婷,杨明玉,郑永康.智能变电站网络化二次系统及其在线监测研究综述[J].电气自动化,2014,36(1):1-4.

[8] 王红星,黄曙,马凯,等.智能变电站间隔层设备在线式自我状态监测系统设计[J].广东电力,2013,26(10):69-74.

[9] Yanh, Gao G, Huang, et al. Study on the Condition Monitoring of Equipment Power System Based on Improved Control Chart [C]//Quality, Reliability, Risk, Maintenance and Safety Engineering (ICQR2MSE) 2012 International Conference on IEEE 2012:24-27.

[10] 蔡骥然,郑永康,周振宇,等.智能变电站二次设备状态监测研究综述[J].电力系统保护与控制,2016,44(6):148-154.

[11] 彭立波.集中式保护装置在智能变电站中的应用[J].科技与企业,2012(16):120-121.

[12] 张晓华,刘跃新,刘永欣,等.智能变电站二次设备的状态监测技术研究[J].电工文摘,2011(4):68-72.

[13] 朱林,王鹏远,石东源.智能变电站通信网络状态监测信息模型及配置描述[J].电力系统自动化,2013,37(11):87-92.

[14] 张巧霞,贾华伟,叶海明,等.智能变电站虚拟二次回路监视方案设计及应用[J].电力系统保护与控制,2015,43(10):123-128.

[15] 王磊,陈青,高洪雨,等.基于大数据挖掘技术的智能变电站故障追踪架构[J].电力系统自动化,2018,42(3):84-9.

[16] 李永恒.基于Hadoop的智能变电站数据管理系统设计[J].长春:吉林大学,2017.

[17] 李岩,阎惊奇.基于Hadoop智能变电站在线监测数据平台及性能测试[J].现代工业经济和信息化,2017,7(16):54-55.

作者简介:

张 艺(1994), 硕士, 研究方向为大数据驱动下智能变电站的二次状态监测系统研究;

李大中(1961), 教授, 主要从事新能源发电研究工作;

郑永康(1977), 博士, 专责, 高级工程师, 主要从事智能变电站状态监测与故障定位研究工作。

(收稿日期: 2019-05-24)