

DOI: 10.7667/PSPC170281

基于大数据平台的电网线损与窃电预警分析关键技术

李端超¹, 王松¹, 黄太贵¹, 程栩¹, 许小龙², 窦万春²

(1. 国网安徽省电力有限公司电力调度控制中心, 安徽 合肥 230022;

2. 南京大学计算机软件新技术国家重点实验室, 江苏 南京 210023)

摘要: 提出了一种基于营配调贯通的海量数据分析技术, 采用电力大数据平台关键技术构建电网线损与窃电预警分析系统, 实现线损的一体化计算、分析与展示。在线损计算结果的基础上, 综合利用电网企业现有海量数据, 通过采用 Hadoop 离线分布式计算、Spark 内存计算等大数据技术对线损率异常线路或台区进行深度挖掘, 识别出可能存在的窃电行为, 为供电企业反窃电稽查提供窃电预警和数据支持服务, 进一步提升供电企业的经营效益。本系统的构建为大数据技术在电力行业的应用进行了验证和实践。

关键词: 海量数据; 大数据平台; 电网线损; 窃电预警; 一体化计算

Key technologies of line loss and stealing electricity prediction analysis based on big data platform

LI Duanchao¹, WANG Song¹, HUANG Taigui¹, CHENG Xu¹, XU Xiaolong², DOU Wanchun²

(1. Dispatch & Control Center, State Grid Anhui Electric Power Limited Company, Hefei 230022, China;

2. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China)

Abstract: Based on the adjustment of marketing and distribution massive data, this paper proposes a system for power grid line loss and stealing electricity prediction with the key technologies of electric power big data platform. The system provides the integration of calculation, analysis and visualization for line loss. According to the result of line loss calculation and the existing massive data of power grid enterprise, the system adopts the big data technology of Hadoop offline distributed computing and Spark memory computing algorithm to identify the potential acts of stealing electricity by big data mining for the abnormal line loss rate of line and supplying district. Therefore, it provides the stealing electricity prediction and data support services for power supply enterprise with anti-stealing inspection and improves the economic efficiency of enterprise. Through the construction of the system, the application of big data technologies in the electric power industry is verified and implemented.

This work is supported by National Natural Science Foundation of China (No. 61672276).

Key words: massive data; big data platform; power grid line loss; stealing electricity prediction; integrated calculation

0 引言

随着云计算、物联网、移动互联网和社交网络等信息化技术的快速发展, 以分布式存储、分布式计算和海量数据挖掘等为代表的大数据技术得到了空前的发展, 并日趋成熟^[1]。大数据术语最早期来源于 Apache 软件基金会的开源项目 Nutch, 该项目中大数据用以描述更新网络搜索索引需要进行批量处理的海量数据集。2012 年美国奥巴马政府宣布投资 2 亿美元启动“大数据研究和发展计划”, 从而将

对大数据的研究上升为国家意志, 掀起了世界各国对大数据的研究热潮^[2-3]。维克多教授在《大数据时代》著作中指出大数据强调的是相关关系而不是因果关系。目前国内应用大数据技术的企业, 多以 Hadoop 开源系统作为自己大数据解决方案的核心, 如中国移动、中国联通等, 或者是借鉴 Hadoop 思想自主研发类似产品, 如阿里巴巴公司。

大数据的显著特征可概况为“4V”, 分别为 Volume、Velocity、Variety、Value, 即数据量巨大、产生速度快、数据类型多、价值密度低^[4]。进入 21 世纪, 电网企业信息化建设与应用不断深入, 在多个专业领域建成并应用了多套业务系统, 以营配调

贯通相关系统为例，包括营销 SG186 系统、生产管理 PMS 系统、SCADA 系统、电能量采集系统、用电信息采集系统和电网 GIS 系统等，产生了海量的电网运行数据，这些数据具有典型的大数据 4V 特征。目前电网企业采用的传统存储和计算技术越来越难以支撑日益增长的海量数据对存储、计算处理的需求，对数据的利用仍然停留在以统计报表为主的浅层应用层面，难以挖掘出新的业务模式和数据蕴含的深层次业务价值，迫切需要采用新的技术进行处理。将大数据技术广泛应用到智能电网领域，并在坚强智能电网基础上构建全球能源互联网^[5]，是国家“一带一路”战略的重要组成部分，也是“互联网+”电力业务的具体体现。电网线损与窃电预警分析需要依赖营配贯通中各类电网运行管理系统与自动采集系统中的海量数据，具有大数据应用的典型特征，目前国内该项研究工作主要侧重于电网线损的一体化计算与展示，对线损率异常台区的深度挖掘和窃电行为识别的研究还处于起步阶段。

1 电网线损与窃电预警系统总体架构设计与分析

电力大数据平台架构设计时需充分考虑电力行业业务数据特点，具体包括：(1) 支持多种类型的数据存储和处理，涵盖结构化数据、半结构化数据和非结构化数据等；(2) 能够以较低的成本存储海量数据，大量价值密度低的数据不应采用传统的高成本存储技术；(3) 平台能根据用户需要灵活扩展以支撑不断增长的数据量，同时保证系统性能的稳定性；(4) 平台应具有较好的通用性，同时保持一定的技术先进性。综合以上分析，在电力大数据平台总体架构设计时借鉴开源系统 Hadoop 平台及 Spark 平台的设计思想，以分布式文件处理技术为基础，采用 HDFS (Hadoop distributed file system) 的分布式文件处理系统作为大数据的数据存储框架，采用分布式计算技术作为大数据的数据处理框架。基于大数据平台的电网线损与窃电预警分析系统总体技术架构如图 1 所示，系统技术架构分为数据源层、数据采集层、数据存储层、数据处理层、业务模型层、应用展示层以及典型应用场景构建。

数据源层为系统数据的来源，包括生产管理 PMS 系统(电网基础台账等相关数据)、电网 GIS 系统(电网拓扑图形数据)、SG186 营销系统(用户档案与抄录电量相关数据)、电能量采集系统(开关台账与自动采集电量指数数据)、用电信息采集系统(用户及配变自动采集数据)以及 SCADA 系统(开关负荷



图 1 基于大数据平台的电网线损与窃电预警分析系统总体技术架构

Fig. 1 Overall technology architecture of the line loss and stealing electricity prediction system based on big data platform

及开关状态数据)。

数据采集层通过数据抽取、数据转换和数据清洗等步骤，将每天需要的海量数据同步到分布式文件系统中。电网线损及窃电预警分析系统需要每天从调度 SCADA 系统、电能量采集系统、用电信息采集系统和营销系统等多个系统同步电网数据至大数据平台。

数据存储层主要为计算层和应用层提供数据的存储服务，本系统采用 HDFS 对数据进行存储，主要面向全类型数据(结构化、半结构化、实时、非结构化)的存储查询，以海量规模存储、快速查询读取为特征，在传统底层硬件、文件系统的基础上，采用包括分布式文件系统、分布式关系型数据库和列式数据库等技术，支撑数据处理高级应用。

数据处理层根据线损分析及窃电分析构建需求，设计数据挖掘业务模型，采用常用的聚类、回归、分类和关联分析等数据挖掘算法，挖掘数据潜在价值，分析预测线损率及窃电规律。根据样本数据训练优化数据挖掘模型，并在大数据环境下并行运行，得出线损分析和窃电分析所需隐含电网运行模式及结果。数据处理层提供基于批量计算、内存

计算的分布式计算框架, 以及数据挖掘、多维度分析等数据分析算法, 为构建上层的高级数据应用提供算法库。

业务模型层基于数据处理层, 为上层应用的构建提供业务模型。根据不同业务应用的需求, 业务模型结合数据存储层的相关数据与数据处理层提供的相关算法, 为特定业务提供具体实现。

应用展示层基于数据处理层提供的数据分析挖掘算法及数据存储层提供的数据服务, 构建丰富的大数据分析应用。该系统提供的应用主要包括线损与窃电分析以及基础数据管理分析平台两大类。

大数据的安全和治理框架用于大数据的信息安全保护和基础数据管理, 其中数据安全框架包括权限、计算、存储和接入 4 个模块, 数据治理框架包括系统监控、数据治理、配置管理和部署升级。

2 基于大数据分析的电网线损与窃电预警系统关键技术

2.1 大数据采集技术

电力大数据平台中的数据主要来源于电网企业现有的关系数据库和数据仓库等。数据在进行传输的过程中, 既需要保障数据能在关系型数据库和分布式存储之间无缝传输, 又需要保障大规模数据的传输与流转效率, 同时不能影响现有业务系统的正常运行。传统的 ETL 抽取工具一般只具有关系数据库、文件、服务等数据源之间同步的功能, 缺少关系数据库与分布式存储之间数据同步能力。本系统利用关系数据库与分布式存储同步技术实现不同存储机制下的数据双向同步, 即一方面可以将关系数据库的数据抽取到分布式存储中, 另一方面又可以将分布式存储中的数据回写到关系数据库中。数据采集关键技术主要包括数据抽取、数据清洗和数据转换三个部分。

数据抽取技术通过 Sqoop 进行系统之间的数据抽取。Apache Sqoop 能够实现关系数据库(RDBMS)与 Hadoop 之间进行高效的大数据交流。通过 Sqoop 可以将关系型数据库的数据导入到 Hadoop 中的数据存储服务(如 HBase 和 Hive) 中, 同时也可以把数据从 Hadoop 系统里抽取并导出到关系型数据库里。数据清洗技术通过利用数理统计、数据挖掘或预定义的清理规则将脏数据转化为满足线损分析及窃电分析模型质量要求的数据。数据转换技术将导入的业务信息按照线损分析及窃电分析数据挖掘模型进行转换, 典型的数据转换包括: 实体合并及拆分, 如将多个变电设备信息合并到一个变电设

备信息表中; 字段合并及拆分, 如“20170115”拆分成“2017”, “01”; 数据聚合, 根据维度进行数据聚合, 如日->月->年; 数据离散化、数据挖掘关联分析时, 对数值数据需转换成非连续的字典类别, 利用离散化数据以进行关联。

2.2 大数据存储技术

数据存储技术通过以较低的成本保存海量的多类型数据, 并提供对数据的多种访问方式, 涉及的关键技术主要包括: (1) HDFS 分布式文件系统。文件数据存储在分散的低成本存储介质上, 对外提供一致的文件访问接口, 具有良好的容错性和并发吞吐。(2) HBase 列式存储数据库。基于 HDFS 分布式文件系统上的以列相关存储架构进行数据存储的数据库, 主要适用于批量数据处理和即时查询。(3) Hive 为基于 Hadoop 的一个数据仓库工具, 可以将结构化的数据文件映射为一张数据库表, 并提供类 SQL 的查询功能, 同时可以将查询语句转换为 MapReduce 任务进行运行, 执行简单的数据计算任务。

本系统中电网运行数据, 如开关、公变和配变运行数据, 用户用电信息等历史/海量实时数据, 数据量巨大, 需要存放在 HDFS 中进行分布式存储, 供数据处理程序进行分布式离线计算调用。设备台帐、客户档案等结构化档案数据存放在 HBASE 中, 供系统随时调用。电网设备拓扑等非结构化数据存放在 HDFS 中, 供系统进行分布式电网设备拓扑关系分析时使用。

2.3 大数据分析技术

数据处理与分析技术包括分布式计算技术、内存计算技术、流处理技术、数据挖掘和多维度分析等。本系统中对海量数据不需要多次循环迭代的算法, 如数据聚合处理、数据离散化、演化分析和异类分析等, 采用 Hadoop 离线分布式计算技术完成。对于需要多次对数据进行全量扫描的算法, 如关联分析和聚类算法采用 Spark 内存计算技术实现^[6]。

线损分析与窃电分析通过采用数据挖掘技术对电网运行数据进行业务模式挖掘, 数据挖掘基于人工智能、机器学习和统计学技术, 通过对原始数据自动化的分析处理, 做出归纳性的推理, 从而得到数据对象间的关系模式及内在特性。本系统采用的数据分析挖掘算法主要有聚类分析、关联分析、演化分析和异类分析, 通过数据挖掘技术识别出线损率异常线路和台区, 然后再通过更深层次的应用分析, 识别出潜在的窃电用户。

系统通过建立电网能量节点数据模型, 实现电网能量节点主数据和能量数据的管理, 利用 GIS 中

电网拓扑数据建立电网拓扑分析模型，并在此基础上根据现有业务需求分析结果建立主网电量与损耗分析模型、配网分线损耗分析模型、窃电预警分析模型等与电网线损计算与窃电预警分析业务相关的分析模型。

系统通过利用多维度分析展示技术，使线损及窃电预警结果分析充分体现企业线损管理主题数据的集成应用，实现电网主网运行方式公式化，配网拓扑结构图形化。在此基础上，依托数据仓库多维度建模设计思路，实现主、配网动态拓扑管理。利用下钻等分析操作，实现拓扑历史结构追溯及电力关系追溯，实现拓扑结构精细化操作。结合准实时的线损计算能量节点主体数据库，实现任意时间段范围内的、准实时主配网四分线损计算、统计、分析。最后通过对各类关联信息集成分析，多维度、多形式展示计算结果，为提高电网企业线损管理水平和反窃电预警能力提供有效技术保障。

2.4 大数据安全技术

大数据的多源异构、分布广泛、动态增长等特点，使得其安全风险大幅增加。面对日渐突显的大数据安全风险和日新月异的网络入侵及攻击手段，迫切需要采用大数据安全技术对电力大数据的数据安全进行有效保护^[7]。本系统各业务应用系统、终端等数据源的接入采用传统的安全接入方案，通过网络安全、主机安全、访问控制和数据加密等技术手段予以解决。数据存储安全主要采用数据加密方式保护关键业务数据或用户隐私数据。通过采用统一的权限平台对用户进行统一的认证、授权管理，实现对各类大数据用户进行权限区别管理。

3 电网线损与窃电预警分析系统典型业务模型分析

3.1 系统业务架构设计

电网线损与窃电预警分析业务架构包括营配调数据贯通业务与电网线损与窃电预警分析业务两个部分，如图 2 所示。营配调数据贯通业务模块包括系统管理、服务接口管理、台账对照查询、台账对照管理和基础台账查询等业务；电网线损与窃电预警分析业务包括报表统计、电量管理、线损计算、线损分析和窃电预警等业务。

3.1.1 营配调数据贯通业务

营配调数据贯通业务通过构建电网能量节点基础数据管理模型，实现各类基础数据查询分析、关联对照与异常诊断，建立基础数据维护工作的指导

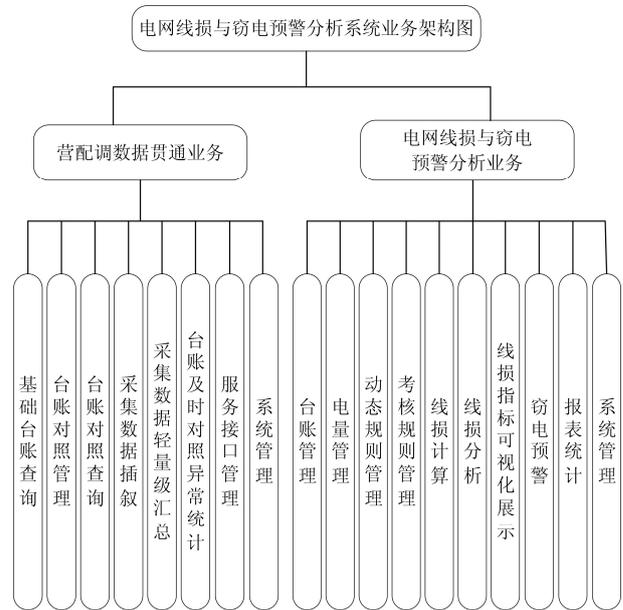


图 2 电网线损与窃电预警分析业务架构图

Fig. 2 Service architecture of the line loss and stealing electricity prediction

与考核机制，提升基础数据维护整体质量，为营配调数据贯通提供校核规范和标准，通过营销 SG186 系统、生产管理 PMS 系统、调度 SCADA 系统、电能量采集系统、用电信息采集系统和电网 GIS 系统之间的互通互联，及时全面地获取各类基础台账数据和自动采集数据，提升数据质量，构建与生产运行相关的基础数据管理平台，为电网线损与窃电预警分析提供基础数据支撑，实现的主要业务有：

- (1) 主网开关、公变、专变、低压用户和分布式电源等电网能量节点基础台账、计量资料和电网 GIS 拓扑等相关信息管理；
- (2) 变电站、线路、主变、母线、开关、公变、专变用户、电容、电抗、环网柜和分支箱等设备的基础台账、计量点信息和电网 GIS 拓扑数据关键属性的异常分析；
- (3) 异构系统间基础设备台账的关联对照以及对照结果分析；
- (4) 开关、公变、专变、低压用户以及分布式电源等能量节点自动采集电量指数和负荷相关原始数据查询与统计分析；
- (5) 开关、公变、专变、低压用户以及分布式电源等能量节点电量指数和负荷数据的自动采集率分析；
- (6) 对本系统相关源头系统数据接入的服务接口的配置与管理；
- (7) 建立基础数据维护工作的指导与考核机制，

促进营配调贯通数据质量提升。

3.1.2 电网线损与窃电预警分析业务

电网线损与窃电预警分析业务基于营配调贯通的海量数据, 利用大数据相关技术, 构建电网线损分析相关模型和全网联络图智能拓扑分析模型, 实现省市县级全面、透明和动态的分区、分压、分元件和分台区的四分理论线损与统计线损的计算和多维度分析, 以及在线监测的可视化展示, 服务于电网企业发展、运检、营销和调度等专业的线损精细化管理, 实现的主要业务有:

(1) 主网开关、公变、专变、低压用户和分布式电源等电网能量节点自动采集周期性电量计算;

(2) 电网能量节点电量数据异常分析和可用率分析;

(3) 主网公式和方案配置以及主网运行方式变化管理;

(4) 配网智能拓扑分析以及配网运行方式变化管理;

(5) 电网分区、分压、分元件和分台区的各类线损指标计算、分析, 在线监测的可视化展示、考核排名等;

(6) 利用线损计算结果和用户用电数据进行数据挖掘分析, 快速查找用户用电异常行为, 及时发现窃电嫌疑, 并结合预定的预警阈值, 及时发出窃电预警。

3.2 典型业务模型分析

3.2.1 配网分线线损模型分析

配网设备管辖范围广, 拓扑结构复杂多变, 运行方式变换频繁, 这些情况对配网分线线损计算结果有较大影响^[8-9]。配网分线联络图如图 3 所示。

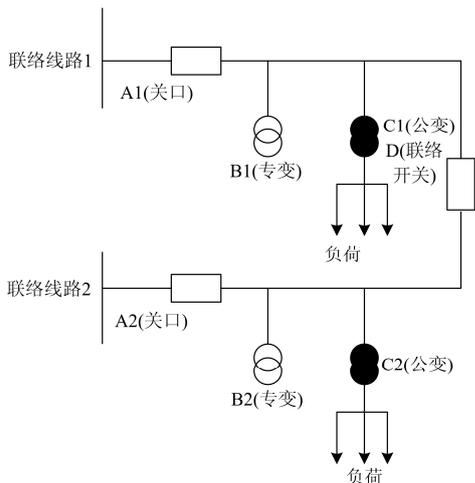


图 3 配网分线联络图

Fig. 3 Sub-line connection of distribution network

针对配网当前管理现状, 中压侧配网通过拓图形方式实现静态拓扑维护, 为降低运行方式变化对配网线损在线计算的影响, 建立与管理相配套的电网拓扑变更和运行方式变化流程管理, 及时维护电网拓扑变更和运行方式变化, 并将记录的供电关系变化涉及的电网设备和影响时间应用到线损计算中, 避免在复杂电网结构下运行方式变化给线损计算带来的偏差。

以图 3 所示配网分线模型为例, 当联络开关 D 断开时, 关联线路 1 的损耗电量为

关联线路 1 损耗电量=A1 电量-(B1 电量+C1 电量)。关联线路 1 的配网分线损耗模型如图 4 所示。

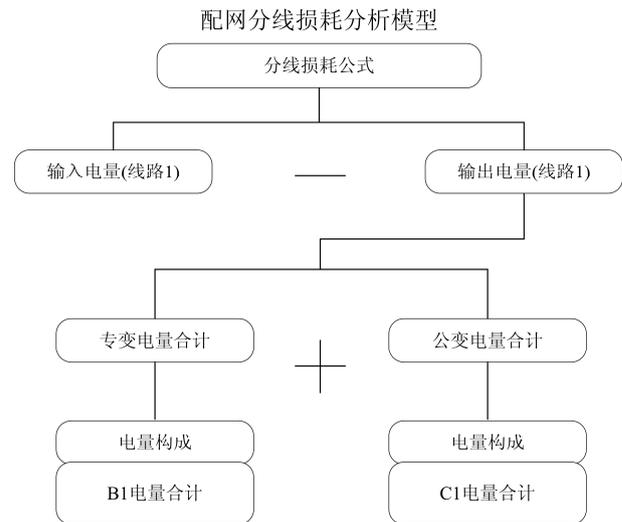


图 4 联络开关断开时配网分线损耗分析模型

Fig. 4 Analytical model of sub-line loss for distribution network while turning off the switches

当关联线路 2 电源点故障, 联络开关 D 闭合时, 电网运行方式发生变化, B2、C2 负荷由关联线路 1 电源代供, 此时关联线路 1 的配网分线损耗模型如图 5 所示, 与联络开关 D 断开时相比, 关联线路 1 的输入电量应去除关联线路 2 上专变 B2 和公变 C2 的代供电量, 即: 关联线路 1 的输入电量=A1 电量-(B2 转出电量+C2 转出电量), 其中:

B2 转出电量=(转点结束时间-开始时间)/(总时长)×B2 总电量;

C2 转出电量=(转点结束时间-开始时间)/(总时长)×C2 总电量。

上式中, 转点结束时间为联络开关 D 由闭合状态转为断开状态的时间点, 开始时间为联络开关 D 闭合时的时间点, 总时长一般根据线损计算周期确定。

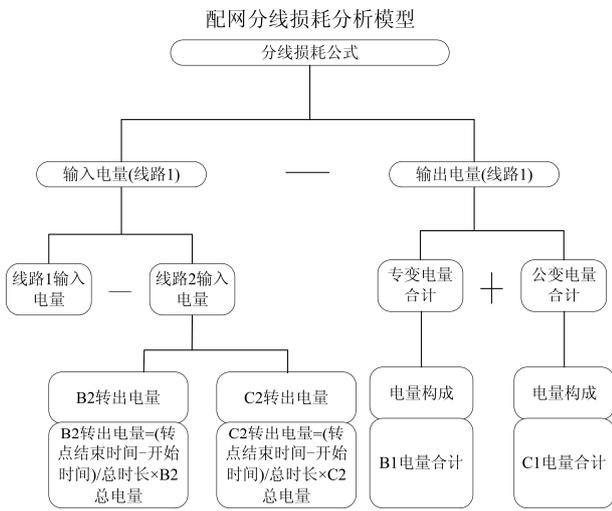


图 5 联络开关闭合时配网分线损耗分析模型

Fig. 5 Analytical model of sub-line loss for distribution network while turning on the switches

3.2.2 窃电预警模型分析

线损率是供电企业的一项重要经济技术指标，综合反映了电网经济运行管理水平和供电企业经济效益。线损根据其特点可分为统计线损、理论线损和管理线损^[10]。统计线损为电网的实际线损，其值为电表计量统计出来的供电量与售电量的差值，反映了电网的实际损耗情况^[11-12]。理论线损又称技术线损，根据供电设备的参数和电网运行的实时负荷数据，通过理论计算而得出。理论线损反映了在特定电网结构和运行方式下，电网理论上应该损耗的电量。管理线损为电网实际损耗除去理论线损外的其他损耗，其值为统计线损与理论线损的差值。窃电预警模型分析通过理论线损与统计线损进行对比分析，快速发现线损异常的线路和台区，然后进一步分析在线损率异常的线路或台区中潜在的用户用电异常行为，整体窃电预警分析模型如图 6 所示。

在图 6 所示的模型中，通过设定窃电预警阈值，当统计线损与理论线损值超过窃电预警阈值时启动窃电预警分析模型，进一步分析可能存在的窃电用户，其中公变用户可分解到低压居民用户和非居民用户，专变用户分解到三相不平衡用户、波动电量用户、采集抄录对比异常用户等，从而细化和缩小可疑用户范围。目前窃电方式主要包括改变电流、电压方式，改变电能表结构和接线方式等，无论哪种方式都会引起线损及各类数据的异常，通过线损异常偏差率并结合各类数据进行异常分析，确定窃电嫌疑的范围，为营销工作人员反窃电稽查提供技术支持。

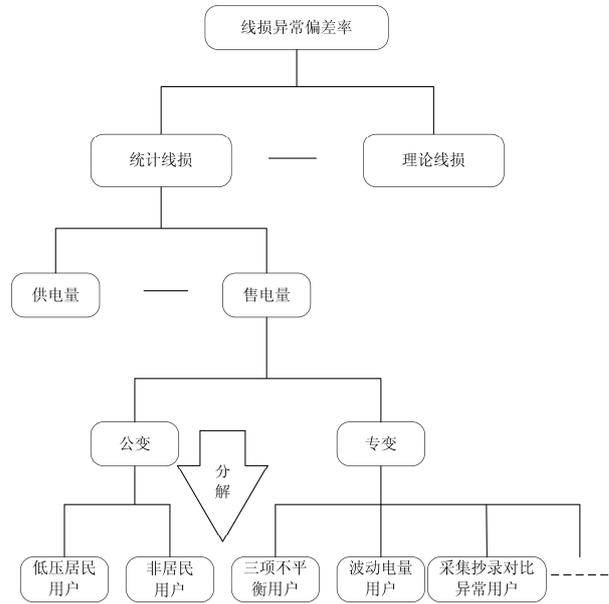


图 6 窃电预警分析模型

Fig. 6 Analytical model of stealing electricity prediction

4 电网线损与窃电预警分析系统典型应用与成效分析

4.1 系统性能指标验证

基于大数据平台的电网线损与窃电预警分析系统测试环境采用集中式架构部署^[13]，在某省级电网公司端部署基础数据平台、线损与窃电预警系统应用服务器以及相关数据服务器集群，地市端仅部署基础数据平台前置服务器。系统接入数据为某省级电网营配调相关系统数据，按照每条记录平均 400 个字节进行存储容量计算，接入全省范围内所有能量节点 2 年时间历史数据，存储容量约为 12 TB，以后每日按照 16 GB 数据量增长。

线损与窃电预警分析数据库和基础数据平台数据库采用 hadoop/Spark 服务器集群，实现负载均衡，该集群具有一台 Master 节点，按照默认的 1:3 冗余比配置 5 台 Slave 节点，分布式计算集群节点配置清单和物理结构图分别如表 1 和图 7 所示。

系统在图 7 所示物理配置环境下的性能测试情况分别如图 8 和图 9 所示。其中读性能比较将常规 Oracle 数据库和基于 Hadoop 的分布式存储性能进行比较，当读取的数据库规模从 1 亿条增长到 8 亿条时，Oracle 数据库每秒读取条数均在 50 000 条以下，Hadoop 平台每秒平均读取条数为 100 000 条，读取数据性能平均提升两倍以上。写入数据性能方面，Hadoop 平台较 Oracle 数据库平均每秒插入条数提升四倍以上，其中插入数据性能提升尤其明显。

表 1 节点配置清单

设备名称	设备型号	设备配置	节点类型
Hadoop1	曙光 A650-G	CPU: AMDOpterin, 内存 12 GB, 硬盘 1.2 TB	Master 节点
Hadoop2	惠普 DL580	CPU: Xeon E7310, 内存 4 GB, 硬盘: 292 GB	Slave 节点
Hadoop3	惠普 DL580	CPU: Xeon E7310, 内存 4 GB, 硬盘: 876 GB	Slave 节点
Hadoop4	曙光 A650-G	CPU: AMDOpterin, 内存 4 GB, 硬盘 558 GB	Slave 节点
Hadoop5	联想	CPU: Intel 奔腾双核 E5300, 内存: 4 GB, 硬盘: 640 GB	Slave 节点
Hadoop6	联想	CPU: Intel 奔腾双核 E5300, 内存 2 GB, 硬盘: 320 GB	Slave 节点

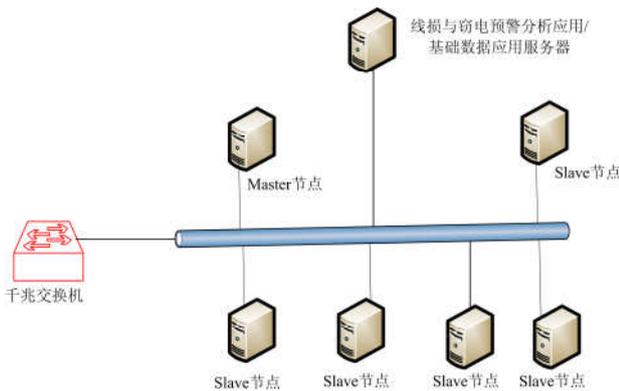


图 7 分布式计算集群物理架构

Fig. 7 Physical architecture of distributed computing clusters

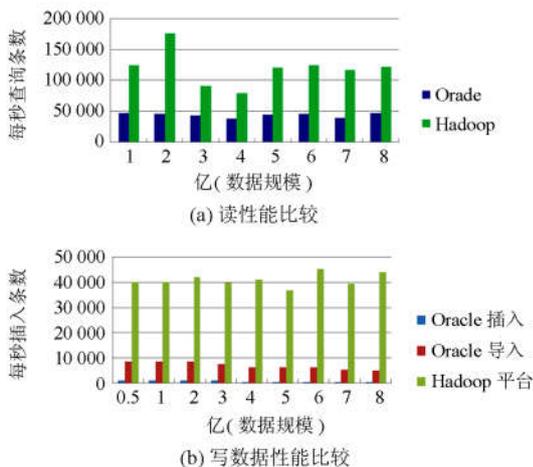


图 8 系统读写性能比较

Fig. 8 Comparison of read/write performance

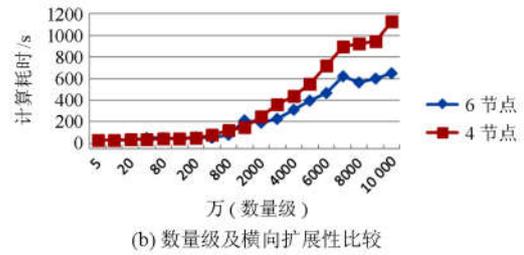
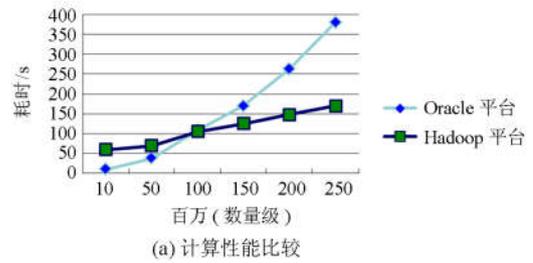


图 9 系统计算性能和扩展能力比较

Fig. 9 Comparison of computation performance and extension power

计算性能方面，当参加计算的数据规模从 1000 万条增长到 2.5 亿条时，Oracle 计算耗时上升明显，而 Hadoop 平台计算耗时增长较为缓慢，其计算性能明显优于 Oracle。性能扩展方面，当 Hadoop 平台节点从四个增加到六个节点时，计算时间显著下降，尤其是当系统数据量超过 2 亿条时，六节点较四节点计算耗时下降明显。

4.2 系统应用成效展示

4.2.1 线损计算及线损成因分析展示

系统基于营配调贯通的海量数据，利用大数据相关技术实现线损一体化计算与分析，通过分析线损的构成与原因，识别出线损异常的区域、元件和时间段等信息，同时结合电网网架数据提出合理的降损建议和电网改造建议。通过线损四分计算并结合 GIS 地理信息实现全面、透明、动态的主配网分压、分线、分台和分区域的线损计算与基于 GIS 地图的动态展示与分析，使各级电网企业管理者能够及时掌握供电量及线损情况。系统采用聚类、回归、分类和关联分析等数据挖掘算法以及多维度分析技术，分析电网线损的构成及原因，为有效降低电网线损提供数据支撑^[14-15]。

电网线损计算及线损成因分析结果展示如表 2 和图 10 所示。从表 2 可以清晰地得到全网线损的在线统计结果分析与展示；图 10 为台区线损率异常成因分析结果展示，为电网企业线损管理人员提供了直观的可视化的展示手段，有效提高了线损管理的工作效率。

表 2 电网线损计算结果展示

Table 2 Display of calculation results for line loss

电压等级	供电量/万 kWh	售电量/万 kWh	线损率/%
220 kV	55 271.27	54 701.98	1.03
110 kV	47 768.7	47 228.91	1.13
35 kV	12 738.32	12 407.12	2.6
400 V~10 kV	13 264.81	12 531.27	5.53

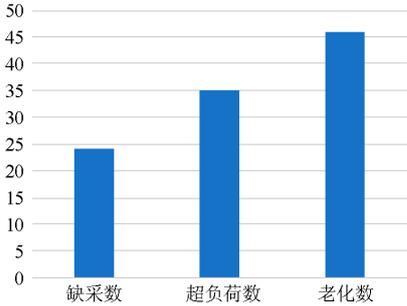


图 10 台区线损率异常成因分析展示

Fig. 10 Display of anomaly line loss rate for supplying district

4.2.2 窃电预警可视化展示

系统基于线损数据、用户数据、电网网架数据和 GIS 数据等海量数据，通过大数据技术分析出线损率异常的线路或台区，同时结合用户日常用电行为，快速发现线损率异常的线路或台区中可疑的用电客户。通过结合 GIS 地图信息并基于反窃电相关模型及预定的预警阈值，进一步缩小可疑用电客户范围，从而快速确定潜在的窃电用户并及时发出窃电预警，为电网企业用电稽查人员开展反窃电工作提供技术支持，有效减少窃电损失。图 11 所示为窃电预警可视化展示结果，GIS 图上红色告警客户为可疑用户，用电稽查管理人员点击红色区域后可以进一步查询到窃电预警客户的详细信息，便于下一步开展反窃电稽查工作。



图 11 窃电预警区域可视化展示

Fig. 11 Display of stealing electricity prediction on the map

5 结论与展望

本文探讨了以分布式存储、分布式计算和海量数据挖掘为代表的大数据关键技术，基于电网线损管理与反窃电的业务需求，提出基于大数据平台的电网线损与窃电预警系统总体架构，对电力企业大数据平台的搭建和应用具有重要参考价值。本文阐述了数据采集、数据存储、数据处理与分析和大数据安全等电力大数据关键技术，通过业务架构设计和配网分线损耗模型分析、窃电预警模型分析两个典型业务场景的建模和分析，将大数据平台和电力企业具体业务应用需求紧密结合，实现了大数据技术在电力企业的实际应用。文章最后通过搭建大数据测试平台对系统的关键指标和性能进行了比较和验证，通过可视化技术对电网线损计算结果、线损率异常成因分析和窃电预警分析功能进行了可视化成果展示。

大数据技术在电力行业的应用目前仍处于起步阶段，如何更好地发挥大数据的技术优势，充分挖掘电力企业大数据的潜在价值将是今后一段时间电力大数据的研究方向。下一步将在现有研究成果的基础上，进一步拓宽大数据技术在电力企业的应用范围，开展大数据技术在电网实时运行分析、负荷预测、节能发电调度、新能源并网和用户双向用电互动等方面的实践应用。通过大数据技术在电力行业的深化应用为智能电网建设和全球能源互联网的发展带来新的发展契机。

参考文献

- [1] MENG S, ZHOU Z, HUANG T, et al. A temporal-aware hybrid collaborative recommendation method for cloud service[C] // IEEE International Conference on Web Services: IEEE, 2016: 252-253.
- [2] 彭小圣, 邓迪元, 程时杰, 等. 面向智能电网应用的电力大数据关键技术[J]. 中国电机工程学报, 2015, 35(3): 503-511.
PENG Xiaosheng, DENG Diyuan, CHENG Shijie, et al. Key technologies of electric power big data and its application prospects in smart grid[J]. Proceedings of the CSEE, 2015, 35(3): 503-511.
- [3] 张东霞, 苗新, 刘丽平, 等. 智能电网大数据技术发展研究[J]. 中国电机工程学报, 2015, 35(1): 2-12.
ZHANG Dongxia, MIAO Xin, LIU Liping, et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE, 2015, 35(1): 2-12.
- [4] 薛禹胜, 赖业宁. 大能源思维与大数据思维的融合(一)大数据与电力大数据[J]. 电力系统自动化, 2016, 40(1): 1-8.

- XUE Yusheng, LAI Yening. Integration of macro energy thinking and big data thinking part one big data and power big data[J]. Automation of Electric Power Systems, 2016, 40(1): 1-8.
- [5] 刘振亚. 全球能源互联网[M]. 北京: 中国电力出版社, 2015.
- [6] 孟建良, 刘德超. 一种基于 Spark 和聚类分析的辨识电力系统不良数据新方法[J]. 电力系统保护与控制, 2016, 44(3): 85-91.
- MENG Jianliang, LIU Dechao. A new method for identifying bad data of power system based on Spark and clustering analysis[J]. Power System Protection and Control, 2016, 44(3): 85-91.
- [7] 冯登国, 张敏, 李昊. 大数据安全与隐私保护[J]. 计算机学报, 2014, 37(1): 246-258.
- FENG Dengguo, ZHANG Min, LI Hao. Big data security and privacy protection[J]. Chinese Journal of Computers, 2014, 37(1): 246-258.
- [8] 雷正新, 韩蓓, 聂萌, 等. 配电网大数据环境下的多点负荷预测算法与应用研究[J]. 电力系统保护与控制, 2016, 44(23): 68-78.
- LEI Zhengxin, HAN Bei, NIE Meng, et al. Research and application of multi-node load forecasting algorithm under the environment of distribution network's big data[J]. Power System Protection and Control, 2016, 44(23): 68-78.
- [9] 何艺, 陈俊. 基于 GMDH 算法的配电网线损数据预处理研究[J]. 电力系统保护与控制, 2015, 43(9): 42-46.
- HE Yi, CHEN Jun. Research on line loss data pretreatment in distribution network based on GMDH algorithm[J]. Power System Protection and Control, 2015, 43(9): 42-46.
- [10] 路书军, 董国伦, 王正风, 等. 电力系统实时运行分析技术及应用[M]. 北京: 中国电力出版社, 2015.
- [11] 卢志刚, 李学平. 基于蚁群的在线理论线损分析用输电电网单线图自动布局[J]. 电力系统自动化, 2011, 35(21): 74-77.
- LU Zhigang, LI Xueping. Automatic plating of single-line diagrams for power transmission network online theoretical line loss analysis based on ant colony algorithm[J]. Automation of Electric Power Systems, 2011, 35(21): 74-77.
- [12] 陈哲, 孙毓婕, 韩学山, 等. 电网分压线损率计算方法[J]. 中国电力, 2015, 48(11): 155-159.
- CHEN Zhe, SUN Yujie, HAN Xueshan, et al. Calculation method of divided voltage line loss rate in power grid[J]. Electric Power, 2015, 48(11): 155-159.
- [13] 章元德, 史亮, 陆巍, 等. 线损信息化统计中数据质量管理控制机制及实现[J]. 电力系统自动化, 2016, 40(7): 128-133.
- ZHANG Yuande, SHI Liang, LU Wei, et al. Realization of data management and control mechanism of line loss statistical information[J]. Automation of Electric Power Systems, 2016, 40(7): 128-133.
- [14] 姜俊锋, 丁香乾, 侯瑞春, 等. 基于 Citespace III 的大数据研究的可视化分析[J]. 计算机与数字工程, 2016(2): 291-295.
- JIANG Junfeng, DING Xiangqian, HOU Ruichun, et al. Visualization analysis about big data research based on CitespaceIII[J]. Computer and Digital Engineering, 2016(2): 291-295.
- [15] 杨东华, 李宁宁, 王宏志, 等. 基于任务合并的并行大数据清洗过程优化[J]. 计算机学报, 2016(1): 97-108.
- YANG Donghua, LI Ningning, WANG Hongzhi, et al. The optimization of the big data cleaning based on task merging[J]. Chinese Journal of Computers, 2016(1): 97-108.

收稿日期: 2017-03-01; 修回日期: 2017-06-05

作者简介:

李端超(1971—), 男, 硕士, 高级工程师, 主要研究方向为电网调度自动化; E-mail: lidc99@163.com

王松(1983—), 男, 通信作者, 硕士, 高级工程师, 主要研究方向为电网调度自动化; E-mail: wsong09@163.com

黄太贵(1963—), 男, 硕士, 教授级高级工程师, 主要研究方向为电网调度自动化。E-mail: huangtg2053@ah.sgcc.com.cn

(编辑 魏小丽)