

Research and Deployment of Power User Behavior Analysis Platform Based on Big Data

Yingqian Liao, Wenting Jiang

Power Grid Dispatching Control Center of Guangdong Power Grid Co., Ltd., Guangzhou Guangdong
Email: 715147516@qq.com

Received: Nov. 15th, 2018; accepted: Nov. 26th, 2018; published: Dec. 12th, 2018

Abstract

Analysis technology of power user behavior can coordinate the power resources and power consumption, which is very conducive to improving the utilization of resources and customer satisfaction. This paper first designs a platform based on big data for power user behavior analysis. The platform includes data presentation layer, data analysis layer, storage computing layer and data source layer, which can solve the problem of power behavior analysis. Secondly, the key technologies used in the data processing used in the platform are studied. Finally, big data platform is deployed and implemented. After more than six months of trial running, the platform presented by this paper has a better experience, and is more convenient for developers.

Keywords

Smart Grid, Power User Behavior Analysis, Big Data, Memory Calculation

基于大数据的用户用电行为分析平台的研究与部署

廖颖茜, 姜文婷

广东电网有限责任公司电力调度控制中心, 广东 广州
Email: 715147516@qq.com

收稿日期: 2018年11月15日; 录用日期: 2018年11月26日; 发布日期: 2018年12月12日

摘要

用户用电行为分析技术, 能够协调好电力资源和电力消费, 非常有利于提高资源的利用率, 从而提高客

户的满意度。本文首先设计了专门针对用户用电行为分析的基于大数据的平台,该平台包括数据表现层、数据分析层、存储计算层、数据来源层,能够较好的解决用电行为分析中的问题。其次,研究了平台中使用的数据处理的关键技术。最后,部署和实现了大数据平台。经过半年多的试运行,该平台相比于已有平台,对于技术研发人员和产品开发人员,具有更好的用户体验和更加便捷的使用效果。

关键词

智能电网, 用户用电行为分析, 大数据, 内存计算

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着大数据、云计算、物联网、人工智能等技术的快速发展和应用,智能电网近几年迎来了高速发展期,并且随着新能源和电动汽车等技术的发展,智能电网产生的数据量快速增长[1]。在传统的电力系统中,电力资源的生产和消费属于两个不同的部门管理,没有得到很好的协调。如果能够协调好电力资源和电力消费数据和管理,将非常有利于提高资源的利用率,从而提高客户的满意度。用户用电行为分析技术,非常好的解决了这个问题。用户用电行为分析是指采用大数据技术,将用户用电数据和用户的用电模式进行关联,从而发现其中的联系。基于用户用电量和用户的用电模式,可以制定出更加合理的电价和相关营销策略,从而提高电力资源利用率和客户的满意度。

在设计电力大数据应用平台时,需要考虑电力系统自身的特点。电力系统具有用户群体庞大、用户用电数据类型多样的特点,决定了电力大数据平台需要较高的存储容量。另外,电力数据受经济环境、人口迁移、窃电行为等因素影响较快,对电力大数据平台的性能提出了较高的要求。所以,电力大数据应用平台与其他大数据应用平台最大区别在于平台的容量更大、计算性能要求更高。已有研究提出了各种基于大数据的解决方案,主要成果包括生产控制系统、电力经营系统、电力营销系统、状态检测系统、风险评估系统等平台应用系统[2] [3] [4]。但是,大部分研究主要集中于发电、变电、输电、调度等智能电网的应用环节[5]。同时,已有研究大部分基于离线数据分析,或者批处理技术,这些解决方案的分析结果一般会有延迟,不能很好的适用于用户用电行为分析[6]。

为解决上述问题,本文首先设计了专门针对用电行为分析的基于大数据的平台,该平台包括数据表现层、数据分析层、存储计算层、数据来源层,能够更好的解决用电行为分析中的问题。其次,研究了平台中使用的数据处理的关键技术,并实现和部署了大数据平台。经过半年多的试运行,相比于已有系统,本文提出的大数据分析平台拥有更好的使用体验和更加便捷的使用效果。

2. 平台架构

基于对已有研究成果的分析可知,为了更好的支撑用户用电行为分析业务,本文提出的用电行为分析的基于大数据的平台,需要支撑现有的用电行为分析的主要业务。所以,需要对各个电力公司的现有相关业务进行调研和分析。例如,通过调研可知,大部分电力公司提供的用电行为分析业务,都能够结合专家预测、机器学习等模型进行集中分析,建立实时判断电力资源生产状况、电力资源质量问题发现与预警、用户用电量远程监测、用电安全形势预测评估等。

为了支撑好现有的和未来可能具有的用电行为分析业务, 确保平台的便利性和易用性, 本文提出的平台架构如下图 1 所示, 具体包括数据表现层、数据分析层、存储计算层、数据来源层。其中, 数据表现层主要用于实现数据的展现, 具体包括图表呈现、数据监控、报表推送、查询服务等功能。数据分析层主要用于用户用电数据和电力资源数据的分析, 包括元数据的存储, 数据挖掘工具的支撑。存储计算层主要用于数据的存储和计算; 数据存储方面, 采用分布式的架构, 可以实现海量数据的实时存储; 在数据的计算方面, 可以实现批量数据计算、流式计算。数据来源层用于实时获取智能电表、智能用电设备功率等数据, 包括结构化数据、非结构化数据。

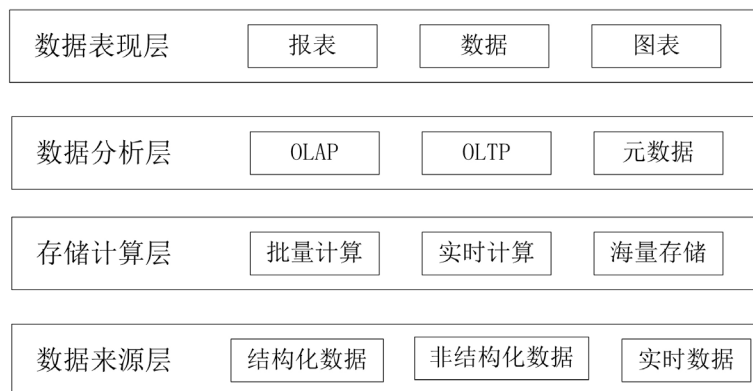


Figure 1. Platform for user behavior analysis platform based on big data
图 1. 基于大数据的用户用电行为分析平台架构

3. 数据处理的关键技术

本文提出的基于大数据的用电行为分析平台, 可用于基于大数据挖掘和分析的电力资源生产、消费及突发事件的全过程管控, 实现电力资源安全生产和消费环境的“透明化”, 在电力资源灾害的早期发现与预防领域也有较大的应用价值。数据挖掘和分析的一般过程为抽取数据、净化数据、数据引擎、算法引擎、运行挖掘算法、分析结果等。下面对本文提出的大数据用电行为分析平台中的关键技术进行研究, 以确保平台的高可靠、高扩展、高存取性能, 从而实现更好的大数据平台服务模式。

3.1. 数据的存储

数据的存储方面, 平台根据具体应用场景的需要, 既支持传统关系型数据库, 也支持内存数据库、分布式数据库、NoSQL 数据库等新型数据库。其中, 内存数据库主要用于 OLTP 事务处理类型的业务中, 包括 TimesTen、SolidDB 等数据库; 海量存储系统主要用于 OLAP 分析型的业务中, 包括 DHSS、Teradata 等数据库; NoSQL 数据系统主要用于互联网类型的业务中, 包括 HBase、BigTable 等数据库。

3.2. 数据预处理中心

数据预处理中心方面, 业务网关将原始数据收集上来后, 通过 ESB 传给数据预处理中心。数据预处理中心的主要工作是对各种不同来源的数据进行数据融合, 提高数据质量, 提高可靠性、准确性和一致性。弥补感知采集技术中的缺陷, 减轻各个传感器故障的影响。由于传感器和技术的限制, 有些信息无法直接由传感器测出, 但通过数据融合可以做到。有效减少数据量, 减轻后续处理和存储负担, 提高效率。

数据预处理中心首先对不同来源的数据做时间和空间的对准, 然后对数据进行清洗和整理。其中使用的处理算法主要包括数据整合算法、数据抽样算法、数据归一化算法、数据离散化算法、数据属性算法、数据区间化算法、缺失值处理算法、行内去重算法、数据修改算法、数据统计算法。

3.3. 挖掘与分析

挖掘与分析方面, 对数据进行面向主题的、一般性的和挖掘性的分析, 获得特征信息, 它包括电力信息整合、分析和度量, 也包括电力资源专题的分析与管理, 还包括电力资源管理与决策支持。

通用的数据挖掘算法分为基于统计学习的算法、基于机器学习的算法和基于数据库技术的算法。1) 基于统计学习的算法: 包括回归分析、判别分析、聚类分析、主成分分析、相关分析、分类算法等算法。分类算法包括决策树算法、KNN 算法、朴素贝叶斯算法; 聚类算法包括 K-means 聚类算法、CLARANS 算法、DBSCAN 算法; 2) 基于机器学习的算法: 包括决策树、SVM、关联规则、遗传算法、贝叶斯学习、KNN、神经网络等算法。关联规则包括 Apriori 算法、FP-growth 算法、WFP 算法; 3) 基于数据库技术的算法: 包括多维分析、OLAP 技术、多属性归纳等算法。

另外, 还提供了大量智能电网特有的挖掘分析算法, 如基于非参数回归的故障预测模型、矿工群体行为模型、多目标粒子群优化算法、模糊控制算法等。这些算法以服务的形式对外开放, 开发者可以根据具体问题的需要选择合适的服务, 完成具体应用的开发。

3.4. 数据应用

数据应用方面, 数据挖掘技术在用户用电行为分析的应用路线为: 选择数据挖掘算法、数据建模、模型验证、知识可视化和应用。例如, 应用于用户消费模式识别时, 可以实现电力资源的精细化管理、用户操作行为识别、操作行为分析、技术学习培训、故障诊断等。

4. 大数据平台体系架构的实现与性能分析

4.1. 架构实现

为了实现用电行为分析平台, 并确保平台的标准化、可扩展性, 大数据平台组成如下图 2 所示, 主要包括统一数据采集层、平台中使用的大数据基础服务组件、平台中使用的大数据应用框架组件、统一服务层。

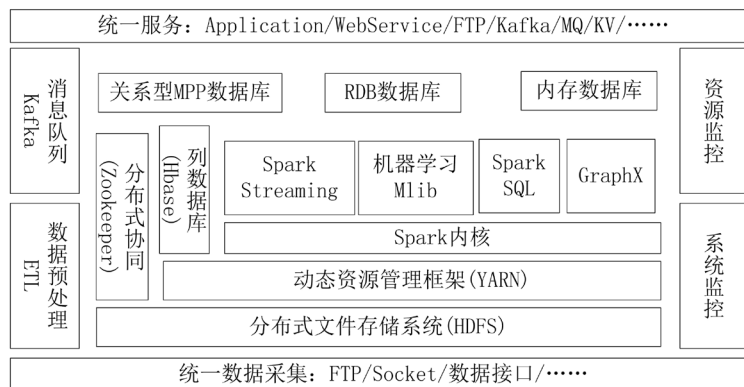


Figure 2. Implementation of big data platform architecture
图 2. 大数据平台体系架构的实现

统一数据采集层采用 FTP、socket、数据接口等, 实现智能电网中智能用电设备的数据采集。平台中使用的大数据基础服务组件包括: HDFS、YARN、HBase 等。HDFS 实现分布式文件的存储。YARN 实现资源的动态管理, 为上层应用提供可靠的资源。HBase 提供高性能的列式数据存储功能。平台中使用的大数据应用框架组件: Spark、Kafka 等。其中, Spark 用于实现流计算、机器学习、sql、图计算等能力。Kafka 用于实现消息队列的订阅、发布。统一服务层主要是为用电行为分析的相关人员提供接口, 方便资源的使

用和管理。对外提供资源的方式包括 application、web service、FTP、KAFKA、MQ 等多种形式。

4.2. 主要功能

基于大数据的用户用电行为分析平台主要包括添加租户、给租户分配资源、租户运行任务三个功能。

添加租户的界面如图 3 所示, 主要信息包括租户类型、租户名称、租户密码、租户组、队列、hdfs 数据根路径等信息, 其中租户类型分为普通用户和高级用户, hdfs 数据根路径用于定义存储该租户所有数据文件的 hdfs 路径。给租户分配资源的界面如图 4 所示, 图的左边显示了集群存储资源、集群计算资源的总容量、已使用量、剩余容量等数据信息, 图的右边显示了租户可以申请的资源类型, 包括 HDFS 资源、YARN 资源、HIVE 资源、HBASE 资源等四种类型。租户运行任务的界面如图 5 所示, 租户使用该功能执行特定程序。其中, 任务信息主要包括项目名称、系统名称、流程标识、执行周期、时间偏移量、使用的队列, 这些关键信息决定了当前任务执行的时间, 以及执行任务时可以使用的资源。

Figure 3. Add tenants

图 3. 添加租户

Figure 4. Tenant resource application

图 4. 租户资源申请

Figure 5. Tenants perform tasks

图 5. 租户执行任务

4.3. 性能分析

为了验证基于大数据的用户用电行为分析平台的性能, 本小节分析了传统平台和本文平台下用户用电行为分析的测试程序。该程序使用客户类型、负荷量、天气情况、日期类型、电价等用户用电特征, 对用户用电模式进行识别。

传统平台下, 程序的接口数据处理包括单个文件处理、全部文件关联处理两部分, 其中单个文件大小最大约 500 G, 最小约 300 G。全部数据总量约 12 T (全月数据)。每个文件处理时间 2 小时(每天数据), 全部文件、文件关联总共耗时约 69 小时。使用本文提出的基于大数据的用户用电行为分析平台处理数据时, 包括数据预处理、数据合并、规则运算三个过程。其中, 数据预处理约 4 小时、数据合并约 5 小时、规则运算约 2 小时。所以, 本文平台在 11 个小时完成用户用电模式分析程序。本文提出的基于大数据的用户用电行为分析平台的运行效率较高。

5. 总结

在智能电网大发展的时代, 通过用户用电行为分析技术, 可以基于用户用电量和用户的用电模式, 制定出更加合理的电价和相关营销策略, 从而提高资源利用率和客户的满意度。为解决这个问题, 本文首先设计了专门针对用电行为分析的基于大数据的平台, 该平台包括数据表现层、数据分析层、存储计算层、数据来源层, 能够更好的解决用电行为分析中的问题。其次, 详细介绍了平台中使用的数据处理的关键技术。最后, 介绍了大数据平台体系架构的实现。相比于已有系统, 对于技术研发人员和开发人员, 具有更好的使用体验和更加便捷的使用效果。

参考文献

- [1] Fang, X., Misra, S., Xue, G.L., et al. (2012) Smart Grid, the New and Improved Power Grid: A Survey. *IEEE Communications Surveys and Tutorials (COMST)*, 14, 944-980. <https://doi.org/10.1109/SURV.2011.101911.00087>
- [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.
- [3] 吴凯峰, 刘万涛, 李彦虎, 等. 基于云计算的电力大数据分析技术与应用[J]. 中国电力, 2015, 48(2): 111-116.
- [4] 刘世成, 张东霞, 朱朝阳, 等. 能源互联网中大数据技术思考[J]. 电力系统自动化, 2016, 40(8): 14-21.
- [5] 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.
- [6] 曹伟, 王炎初, 顾全, 等. 基于电网省地一体化维护的图模集成研究[J]. 中国电力, 2015, 48(2): 81-84.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8763，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：sg@hanspub.org