

基于大数据的智能电网数据分析和可视化处理

王 睿

承德石油高等专科学校 河北承德 067000

摘 要: 随着智能电网的普及,可以对海量的数据进行汇总、分析,但传统的智能电网数据管理系统不能满足数据的扩充和处理。针对这一问题,我们设计了一个以Lambda体系结构为基础的智能电网大数据分析系统。本系统具有存储各类智能电网数据的能力,可以实现并行批量、实时运行。通过对360个居民小区的电网数据进行数据的分析与可视化处理,以检验该系统的运行效果。通过对智能电网的实际应用,验证了基于Lambda结构的大容量网络模型存在的延时问题,并对负载进行了分类,并对结果进行了实时可视化。

关键词: 大数据;智能电网;Lambda构架;数据可视化;需求响应

Smart grid data analysis and visualization processing based on big data

Rui Wang

Chengde Petroleum College, Hebei Chengde 067000

Abstract: With the popularity of smart grid, massive data can be summarized and analyzed, but the traditional smart grid data management system can not meet the data expansion and processing. To address this problem, we design a smart grid big data analysis system based on the Lambda architecture. The system has the ability to store various smart grid data, and can realize parallel batch and real-time operation. The operation effect of the system is tested by processing the data analysis and visualization of 360 residential areas. Through the practical application of smart grid, we verify the delay problem of large capacity network model based on Lambda structure, classify the load, and visualized the results in real time.

Keywords: big data; smart grid; Lambda architecture; data visualization; demand response

引言:

随着社会的飞速发展,与互联网相关的各方面无时无刻不在制造着海量的大数据。本文利用Lambda架构,构建了一个智能化电网大数据分析平台。该系统可以对智能仪表数据、传输监控数据、图像数据、结构数据、非结构数据等智能电网数据进行采集和分析,对海量数据进行批量实时处理^[1]。最后通过对实际电网数据的分析和处理,证明了本系统的实用性。

一、Lambda构架结构特征

(一) Lambda体系结构

Lambda架构是一种将批量运算与实时运算相结合的大数据处理架构,该架构集成了不可变、读写分离、复杂隔离等一系列架构特点。

1. 批处理层

在处理中,我们把所有的数据都看成是不断增加的

一组不变的资料。在成批过程中,对全部的资料进行批次操作,产生批次视图,存储到ServiceLife中,而服务层则可以针对不同的查询情况,重新组织批视图。批处理层定期的更新重复的批次视图,以保证高的数据容错率。然而,批量数据的运算时间一般较长,延迟时间也较大,因此很适于进行离线的数据分析与预测。

2. 实时处理层

实时性的处理级是对动态变化的数据进行即时运算。但由于需要大量的运算,因此无法对实际的增量式资料进行深度的运算。在实时处理层面,可视化系统通过对当前的数据进行快速的递增操作,确保了视图的实时性,并能很好地克服了许多视图的高时延问题。在整个数据流计算过程中,由于存在着许多的视图重叠的情况,因此在实时性的过程中动态选择是非常重要的,信息处理可以在任何时候进行新的信息的读出。

3. 服务层

服务层的工作就是根据数据分析的情况，为用户的查询提供支持。服务层随机存取视图，通过对批量视图和实时视图的分析，向用户提供最终的反馈。服务层一般是通过NoSQL数据库来实现的，降低了复杂性，但是不能随意地写视图，仅能装载和随机读取批量视图。

(二) Lambda 构架特性

一是容错性。在智能电网中，计算失误和节点崩溃是非常普遍的，Lambda体系结构对于机器失效和数据破坏具有很高的容忍能力，并且可以通过批量和实时的方式对主数据进行重算，每一个节点数据都有备份，可以迁移到其它正常节点上完成。

二是低延时性。Lambda结构采用了并行运算技术，能够在不影响系统健壮性的情况下，实时地进行数据的读取与更新。智能电网的运行人员可以对电力系统的运行状况进行实时监测，并对其进行需求侧的管理。

三是可扩展性。Lambda体系结构可以单独地进行扩展。这使得系统可以在不影响集群结点功能的前提下，自动地进行数据的再分配和运算。另外，当新的智能仪表安装完成之后，可以在不影响现有的基础结构的情况下，向现存的Lambda体系结构集群增加计算节点和存储装置。

四是灵活性。Lambda体系结构可以储存、计算来自不同资源的各类数据，适合在智能电网中使用^[2]。

二、智能电网大数据分析系统

基于大数据的智能化网络分析体系，可以对其决策的电力系统的可靠度造成的影响进行有效的监控。基于Lambda架构的智能电网数据处理可以划分为智能电网数据生成、智能电网数据采集和数据存储；本文主要讨论资料的处理、资料的检索及资料的分析。

(一) 数据生成

在整个电力系统内，都安装了智能化的仪器和传感器，能够产生大量的数据，包括了发电侧的数据（风力发电站、光伏电站）、消费侧（住宅）、工厂（电动车）、用户（住宅）、光电板（电动车）。在电力中断监测中，不仅有图像和视频数据，而且还可以把天气数据和自然灾害信息整合到智能电网中。智能电网数据量大、速度快、类型多样，因此，将智能电网大数据与已有的多源智能电网进行数据整合后，其应用价值将会越来越大，并有可能提高智能电网的性能。

(二) 数据采集

通过Flume实现数据的整合与传送，确保了数据存储在一个终端上，实现了数据的实时传输，即便出现了断开或中断。Flume的目的是把真实的资料储存在一个虚

拟的记忆通道中，直至所有的资料都被储存在资料库中。利用资料湖的优点是可以结构化、半结构化及非结构化的资料储存智慧电网资料，方便资料的阅读及分析。

(三) 数据存储和处理

提出的分析系统是在Lambda架构基础上，实现对电网的智能数据的存储与处理。数据采集完成后，将数据传输到Hadoop主节点，Hadoop主机负责数据的存储，而Hadoop MapReduce则负责数据的处理。在Lambda体系结构中，HDFS组件是一个批判层，它负责对多个节点进行存储。另外，在批量处理层中，使用MapReduce处理部件，并利用预测计算对分布数据进行批量分析。当进行批量操作时，Lambda架构中的即时处理器储存，更新和计算Flume的资料。通过Apache Spark技术，可以对Lambda架构进行即时的处理，Spark的最大特点是其内存容量可以支持集群操作，因此可以加速程序的运行。最后，将Lambda系统中的批量处理和即时处理层次相结合，实现了对未来的数据检索过程的实时分析^[3]。

三、云计算平台的系统实现

(一) 系统组件层次

本文采用谷歌云计算技术建立了一套全面的模型，并通过六个计算机作为节点对其进行了系统的研究。此六结点由1个主结点、5个工作结点、8个虚拟化CPU、30 G RAM、4个虚拟化GPU、15 G RAM构成5个工作结点，所有结点都是64位Linux。要在云计算平台上使用Lambda体系结构，以下的部件必须要在它上：（1）Hadoop，包括HDFS和MapReduce。（2）包含分配，记忆管理，故障恢复和与存储系统交互的Spark核心。（3）数据的查询采用Spark SQL。（4）Hive帮助对数据湖中储存的资料进行读取、编写和管理。（5）Impala用来读取、写入和管理结点内存中的平行储存的资料。

(二) 系统与集群节点连接建立

利用Secure-Shell (SSH) 技术，在保证网络的安全性基础上，自动向主机IP地址传送各类智能电网的数据。利用云计算平台进行资料采集的优势在于，可以从任意地点经由网路连接传送资料。在主节点上设置Flume，以积极地对数据进行轮询，并将其输入到DataCast存储。以致于必须建立多个Flume代理来管理数据库。例如，把家用智能化仪表与发电站所生产的资料分开存储，这样就可以避开所谓的“数据沼泽”——即资料库冗余。通过使用数据分析的方法和多个集群进行远程的连接，可以使用SparkSQL、Hive和Impala进行查询，通过ODBC和TCP协议进行了网络的互联。当这套庞大的数据解析体系构建完毕，并且与该集群建立了一

个安全的远距离链接之后，该云服务平台将能够对该体系进行解析和运用^[4]。

四、应用分析

(一) 数据的组织和存储

每15分钟Flume把智能仪表数据输入到批量处理层(在具体位置的数据湖泊)和实时处理层中。这数据包含了智能计时器、ID及负载情况。为了收集家庭的日常负载，MATLAB可以通过Spark SQL存取这些数据，并将它们组织为每日负载数据。为了提取、处理和存储海量数据，本文提出了EPS流程。EPS过程利用大数据湖泊中的智慧电网数据，通过抽取合适的信息，对其进行结构化、查询，并将其按特定的格式进行存储。通过Hive查询，将每天的住户家庭负载数据存入HDFS仓库。这一步可以获得智能电网的实时智能仪表读数，显示家庭的负载程度，也可以用来计算每天的家庭负载程度。

(二) 智能电网负荷的可视化

本系统的目标是利用智能仪表的实时记录，将住户的负载显示出来，并在这里利用Spark技术进行数据的分析和可视化。如果直接使用Impala，则会将最新的智能仪表资料储存在HDFS的仓库中，而HDFS则是一种批处理，这会造成资料显示上的延迟，特别是在系统中有大量的智能电表时，该缺陷使得它不能实时显示智能电网负载。但是在Lambda体系结构中，可以采用实时处理层来存取最新的数据，从而减少了系统的时延。通过与Tableau相连的Spark，可以实时监测电力系统的复杂情况，如图1所示。通过这种方式，可以获得一个综合住户的每日负载。另外，该系统还可以将数据显示出来，并将其与使用者进行实时的联网监测，从而提高了电力系统的可靠性。

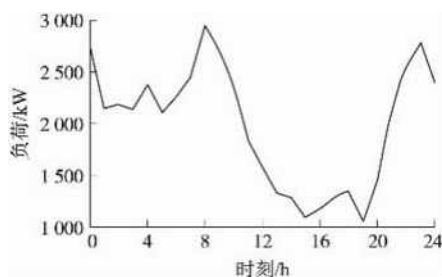


图1 居民住宅日负荷曲线

(三) 居民住宅日常负荷数据的处理

在本应用中，系统期望可以利用由智能电网大数据分析系统所获得的负载数据将其归类。为了实现这个海量数据的挖掘，我们可以通过Spark来连接Radoop和一个簇的结点。图2显示了Radoop对负荷资料进行预置并对本地使用者进行分组的流程。第1步是利用服务层从

分布式HDFS数据库中抽取出使用者的负荷。第2步是在服务层中，通过对大量地处理和实时的计算，可以得到大量的数据。当需要的资料视图成型后，需要对所撷取的资料进行预处理，如删除某些资料不全的每日负载资料，选取要群集的负载属性与使用者。第3步是经过预处理，选择适当的聚类算法进行数据分析。

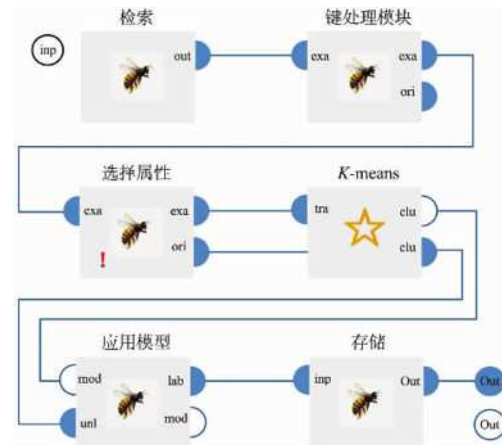


图2 Radoop 负荷数据预处理及分组流程示意图

应该指出，所选的群集算法对于分布数据应该是平行的。目前，基于分布式数据并行处理的数据挖掘算法还比较局限，还处在探索阶段。在该系统中，利用K-means聚类算法，将其按照用户的负载情况进行分组。通过K-means聚类方法，将所得的数据保存到HDFS数据库中，以便对其进行进一步的分析。此外，这些集群还可以详细介绍负荷特征，帮助开发需求反应。

五、结论

本文采用Lambda体系结构，解决了以往智能电网大数据结构存在的时延问题，并实现了实时可视化。因为缺少能够实施数据分析的成熟的工具，所以这个系统采用了一系列的分析工具，这些工具足够进行可视化和数据分析。但是，该智能电网的大数据分析系统在实际应用中的作用远远超过了本文的研究领域。另外，需要指出的是，其它的分析工具可以用于该系统的分析，而且当数据挖掘算法可以并行处理分布式数据时；他们可以在Lambda体系结构中使用。

参考文献：

- [1]刘振亚.智能电网技术[M].北京：中国电力出版社，2010：56-98.
- [2]李俊楠，李伟，李会君，等.基于大数据云平台的电力能源大数据采集与应用研究[J].电测与仪表，2019，56（12）：104-109；56-98.
- [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.