引文格式：李石磊, 常晟, 邵嘉兴, 等. 基于GIS的智慧供热平台关键技术设计与实现[J]. 地理空间信息, 2023, 21(10): 49-53. doi:10. 3969/j. issn. 1672-4623.2023.10.012

基于GIS的智慧供热平台关键技术设计与实现

李石磊[1]，常　晟[1]，邵嘉兴[1]，杨红千[1]
（1. 同方节能工程技术有限公司，北京 100000）

摘　要：GIS具有空间定位特性，能很好地处理复杂的空间拓扑关系，并提供丰富的数据可视化展示方法，对海量数据有着极强的管理能力。因此，以成功应用到1.7亿m^2城市级供热中的智慧供热平台为例，详细介绍基于GIS数据融合模型在超大规模城市集中供热中的处理与应用；结合大数据技术Hadoop，介绍了基于分布式文件存储的海量数据存储方法与基于内存计算处理高并发、秒级数据处理方法；最后，介绍了边缘计算与云边协同，解决了边缘数据不准、容错率低、与云端协同效率低、调节滞后等问题。

关键词：GIS；数据融合；分布式存储；内存计算；边缘计算；云边协同
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2023）10-0049-05

Design and Implementation of Key Technologies of Smart Heating Platform Based on GIS

LI Shilei[1], CHANG Cheng[1], SHAO Jiaxing[1], YANG Hongqian[1]
(1. Tongfang Energy-saving Engineering Technology Co., Ltd., Beijing 100000, China)

Abstract: GIS has spatial positioning characteristics, can handle complex spatial topological relationships. It can provide abundant data visualization display methods and has strong management capabilities for massive data. Taking the smart heating platform successfully applied to the 170 million m^2 heating project for an example, we introduced the data integration, display and analysis methods of ultra-large-scale urban central heating based on the GIS-based data fusion model. Combining Hadoop in big data technology, we introduced the method of massive data storage based on distributed file storage and the method of high concurrency and second-level data processing based on memory computing. Finally, we introduced edge computing and cloud-side collaboration, which had solved the problems such as inaccurate data at the edge, low fault tolerance, low efficiency of collaboration with the cloud and lagging adjustments.

Key words: GIS, data fusion, distributed storage, memory computing, edge computing, cloud edge collaboration

随着国内科学技术的进步与集中供热的发展，集中供热日趋供热面积大型化、自控系统完善化、管理体系复杂化、新技术多样化，同时，对信息安全与软件安全等级的要求也越来越高。因此，解决供热过程中管理难度大、效率低、能耗高、数据乱、维修难、预警差、数据安全等问题，以及最大化提高生产效率和节约能源成本，成为急需解决的难题[1]。

因此，需要一个综合的、全面的、统一的、智慧型管理与调度的供热平台软件，以提高供热运行管理效率，满足新时期节能环保、智慧供热的新需求。鉴于以上原因，本文以成功应用在实际供热项目中的智慧供热平台为例，对在平台建设过程中的部分关键技术进行介绍。

1　技术路线

在智慧供热平台建设过程中，主要参与的主体有数据、存储与计算、处理与分析、调度与协同。因此，分别设计了数据融合模型、分布式存储与内存计算、边缘计算、云边协同，用以解决各个环节中的主要问题。具体如下：

1）建立基于GIS的多源异构数据融合模型，对结构化数据、半结构化数据和非结构化进行融合、统一存储，同时采用postGIS中间件，实现基于开源GIS引擎的城市级供热空间数据的集成、展示与分析；

2）采用Hadoop中的分布式文件系统与MapReduce中的内存计算，建立了分布式存储业务模型与实时数据业务处理模型，解决了海量历史数据的存储、实时数据的高并发传输问题；

3）采用边缘计算，建立边缘计算业务模型，将网络、计算、存储、核心应用能力等有效结合，实时对边缘数据进行处理，对负荷进行调节，使之有更快的服务响应、稳定的服务性能、高效的传输处理能力；

收稿日期：2022-01-17。
第一作者简介：李石磊（1987—），硕士，高级工程师，研究方向为地图制图和地理信息工程技术，E-mail：245183526@qq.com。

4）采用云边协同，建立云边协同业务模型，将边缘侧的服务与云端实时同步，优化资源分布、资源配置，实现了资源协同、策略协同、分布式计算管理，最终实现在海量数据、复杂应用、多场景服务、高并发数据下稳定、持久、安全、统一的云边协作方式。

2 关键技术

首先，建立数据融合模型，对多源异构数据进行处理，设计出统一存储多源异构数据的数据模型，用以展示与存储约8万条管线、约2 000个热力站、约45万个建筑物、14个热源、约1 000万热用户、约200万设备、每年约100 T的数据量，解决了异构数据的存储，打通数据壁垒，对数据进行统一维护与调度，为后期建立数据中台奠定了基础；其次，本文结合大数据中的Hadoop和MapReduce，实现了基于Hadoop的分布式存储模型与基于内存计算的实时数据处理模型，并采用机器学习算法对热源负荷进行预测与调度，解决了海量历史数据的存储与高并发、秒级数据的传输问题；再次，本文采用边缘计算与云边协同，建立了边缘计算业务模型与云边协同业务模型，解决了热力站数据不准、负荷过大、热效率低等问题。

2.1 数据融合模型设计

由于智慧供热平台包含的异构数据较多、数据量极其大、类型复杂以及要与其他业务系统（比如收费系统、客服系统、投诉系统等）进行对接，因此，高效、安全、稳定的数据模型设计显得尤其重要。本文结合实际应用，设计了LSM数据融合模型，很好地解决了生产运行中结构化数据、半结构化数据和非结构化数据。

其中，数据融合算法，采用LSM树，即Log-Structured Merge-Trees，数据格式如图1所示。

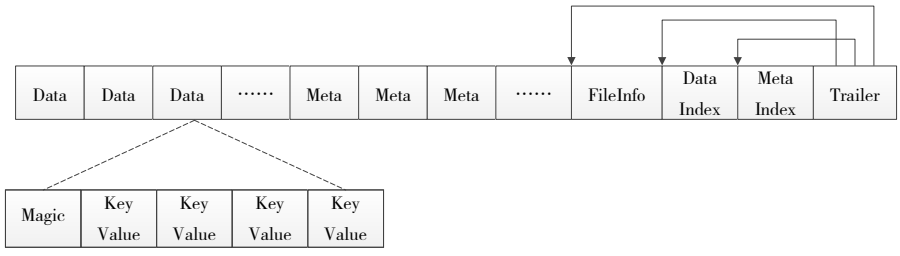


图1 数据格式

图1中，结构化数据以Data和Meta的形式进行存储，每个Data块以Key-value键值的形式进行管理，每个结构块大小约60 K，超过60 K的Data块继续以Key-value键值形式进行分割，直至数据块大小满足60 K；Meta块为元数据信息，记录数据块的大小、创建时间、索引、数据类型等关键数据块信息；FileInfo数据块存储非结构化数据，每个FileInfo数据块均由DataIndex与MetaIndex组成，DataIndex用以记录非结构化数据块信息，MetaIndex记录非结构化数据块元数据信息，包括索引、时间戳、文件类型等。通过图4数据模型，将结构化数据、非结构化数据进行融合，方便存储与管理、同时也为web端高效、安全访问提供便利。

2.2 分布式存储与内存计算

由于智慧供热平台数据量庞大、数据结构复杂、数据种类多等特征，利用数据融合对多源异构数据进行融合后，需要对海量数据和实时数据进行存储与处理。这里采用基于Hadoop的分布式存储对数据进行多节点存储，利用基于内存计算的MapReduce对实时数据进行处理，解决了海量数据存储、融合、调度的问题，同时也解决了基于海量实时数据高并发的问题[2-5]

历史数据与基础数据分别以数据块的形式存储在Hadoop上，Hadoop以分布式部署在不同服务器节点上，经过集群后进行高速运算与存储，同时采用分布式文件系统，增强了系统的容错性，以此提供高吞吐量和超大数据集的访问。

智慧供热平台历史数据庞大，主要来自于生产数据与业务系统数据。生产数据主要有传感器监测的温度、压力、流量等参数，50万点/30 s上传至服务器端；业务数据主要有用户收费数据、投诉数据、客服数据、室温数据等，定时上传至服务器。因此，庞大的数据量要求数据存储和访问采用多服务器集群、多节点访问。首先，客户端访问服务节点1，如果服务节点1不存在该数据，则继续向下索引服务二级节点；如果服务节点1存在客户端想要数据，则在该节点下索引数据节点，数据节点中包含多个数据块，数据块以Key-value即键值对的形式存放数据，最终采

用LSM树数据模型对数据单元进行索引并组织后返回给客户端。

实时数据通过集群环境，采用内存分布数据集进行迭代工作负载，实现了大型、低延迟、50万实时数据秒级的并发访问。基于MapReduce的实时数据结构设计如图2所示。

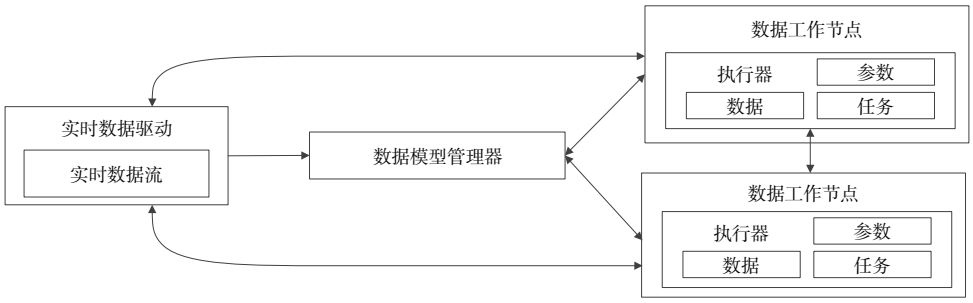


图2　实时数据结构设计

实时数据采集过程中，首先通过实时数据驱动器将实时数据按照一定的数据类型，比如温度、压力、流量、一次网、二次网等参数进行分类；其次，将分类后的实时数据通过数据模型管理器对数据进行分配，即根据不同的数据类型及数据价值特征分配到不同的数据工作节点，每个数据工作节点所承担的数据量、计算力、数据大小等参数不同，使其接受数据模型管理器的统一调度；最后，数据工作节点中的执行器，将数据、参数以及所要执行的任务进行打包执行，将结果存储在数据节点上，以实时返回给实时数据驱动器。如果下一时刻，有相同的数据类型，该数据类型不再受数据模型管理器统一调度，而是按照数据记忆参数配置直接进入相应的数据工作节点参与运算，计算完成后直接将相应结果返回给实时数据驱动器，这样节省了数据传输执行时间，最大限度提高数据的实时性与执行效率。

2.3　边缘计算与云边协同

由于城市级超大规模集中供热面积庞大，末端设备数量大、工况复杂、算法模型较多，如果将硬件、热源、热力站、热用户所有数据和模型均放到云端进行执行和调度，势必会导致云端架构复杂、维护难度大、出错率高、执行效率低等问题。因此，适当为云端“减负”，释放部分计算力给边缘，让边缘用户有自主计算力，能及时处理相应业务，使云端与边缘互相补充，这样不仅不会削弱云端计算能力，还提高了整体协调、运行能力。

在供热中热源、热力站、热用户和主要传感器设备作为整个智慧供热平台的感知层，其数据的可靠性很大程度上决定了云端对整个热网的热负荷调度，进而影响用户室温、能耗及成本；同时边缘数据的可靠性差与数据滞后性导致不能实时反映全网运行特征，成为了急需解决的问题。因此，采用边缘计算业务模型，将边缘侧、服务层、接口层与硬件传感器打包作为一个感知体，有效解决了设备容错率低、通信故障、传输延迟、计算复杂等问题。

智慧供热平台云端管理多个边缘感知体，为其提供充足的虚拟化资源，可以持久化存储和动态调整计算力，以满足不同业务场景和数据并发的需求。云端通过网络来控制边缘，并提供高效、安全连接，而在网络发生中断时，中心云端会尝试重复连接边缘，但网络中断并不影响边缘独立运行，当网络再次恢复时将任务上传到中心云端。

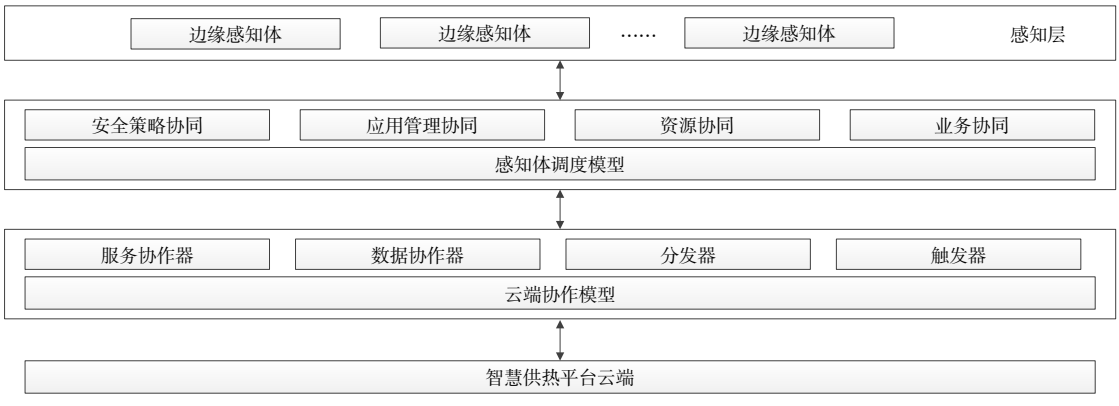


图3　云边协同业务模型

如图3所示，首先，边缘感知体通过网络进入感知体调度器，在调度器中统一对边缘各个感知体进行安全策略、资源、业务注册；其次，感知体通过调度器统一分发到中心云端协作模型中，在协作层，各个感知体以及感知体对应的请求服务会以注册ID的形式被协作器按照一定的参数进行分配，统一汇总智慧供热平台云端；最后，由云端接受各边缘感知体的服务并参与下一步的计算与结果输出，并将计算结果通过网络返回给边缘[6-11]

3 关键技术实现与应用

3.1 关键技术实现

1）数据融合模型。以智慧供热平台中的换热站为例，采用LSM树，对结构化数据和非结构化进行融合。实现过程如下：

步骤1：将换热站的空间数据和其他结构化数据转成Json格式；创建非结构化数据索引表并以base64二进制形式进行编码；

步骤2：将步骤1中的数据按照图4数据结构进行处理；

步骤3：将步骤2中数据作为一条新记录。首先，日志文件会以append形式增加一条记录，这样方便以后恢复使用；其次，将该记录的索引插入C0中，这里在内存中完成，不涉及磁盘IO操作；当C0达到一定阈值或一定时间间隔后，C0中的记录会滚动合并到磁盘C1中，当C1中的容量达到一定阈值后会向C2合并，依次类推，直至合并至Ck，至此，完成了数据的融合与存储。

通过LSM树将不同类型的数据进行融合，降低了不同类型数据存储、读写的难度，也方便对数据的维护和管理。

2）分布式存储与内存计算。智慧供热平台采用分布式文件系统HDFS和基于内存计算的MapReduce对数据进行存储和处理。其中，HDFS对大量数据进行存储，而MapReduce对大量数据进行计算。根据实际应用中数据量，智慧供热平台采用40台服务器作为数据节点，每个节点服务器主要参数配置为：内存256 G，硬盘10 T，Linux操作系统；具体设计过程参考2.2章节。实现过程如下：

步骤1：对供热中结构化数据、半结构化数据、非结构化数据进行清洗，去除掉错误、冗余数据；

步骤2：将步骤1中数据存储在大数据仓库中，并利用Hadoop平台实现数据在不同业务场景下进行转存的需要；同时利用ETL计算框架对数据进行处理，再利用Hive组件，通过HiveQL将数据仓库中的数据提交到MapReduce任务中；

步骤3：MapReduce将步骤2中的数据按照20 M（这个可以根据实际需要进行修改）划分为一个数据单元，分叉复制到集群环境内其他数据服务器上；在ETL程序分布的服务器上，用Master服务器调度Worker服务器，并给Worker服务器分配相应的Map任务或Reduce任务；

步骤4：对于步骤3中分配了Map任务的Worker，用以实现对数据的读取，当Map任务接受到数据输入时，会以键值对作为参数传递给Map函数，Map函数会返回中间键值对，并以缓存的形式将其储存在内存上；当中间键值对达到一定数量时，会被写入到本地磁盘，本地磁盘会以多个分区对其进行存储，每个分区对应一个Reduce任务；Master会把中间键值对位置转发给Reduce Worker；

步骤5：Master服务器会对Worker服务器上的任务性质进行判断，如果是Reduce任务，Master会通知Worker当前所处具体区域位置，并由Reduce Worker把相同的键值对汇集起来；Reduce Worker遍历完所有中间键值对后，向Reduce函数提交查询，Reduce函数将输出结果存储在对应的磁盘分区上；

步骤6：当全部Map任务和Reduce任务执行完毕后，Master服务器会再次执行ETL计算框架对数据进行处理，执行过程重复上述操作。

通过HDFS分布式存储与MapReduce内存计算，有效解决了数据在高并发场景下存储空间不足、读写速度慢、数据丢失等问题。

3）边缘计算与云边协同。智慧供热平台的边缘侧承担了中心云端部分计算力。其物理组成包括1台小型硬件服务器，主要参数配置：内存4 G、硬盘10 G、显示器1个、Linux操作系统；通过物理设备将终端采集器PLC连接起来，形成小型局域网，该小型服务器被称为“决策机”。在决策机上部署数据处理算法用来解决数据不准、容错率低的问题，然后将处理后的数据上报云端；中心云端将部分计算力下放到边缘的决策机上，让决策机承担底层业务工作，并定期将处理后的结果上报到云端；云端对决策机的处理结果按照一定规则进行检验与审核，最终将确认记录归档。由于边缘具有部分自主决策能力并与云端错峰交互、数据同步与共享，使得在协同过程中减少由下层业务系统同时到达云端等待分配计算力、计算资源的时

间，提高了云端与边缘的协同效率，解决了在实际供热中由于城市规模较大、供热面积较广导致调节滞后的问题。

3.2 应　用

智慧供热平台是基于GIS、物联网、大数据、人工智能于一体，集供热生产“总数据中心、总展示中心、总指挥中心、总调度中心和总管理中心”五位一体的城市级综合管控平台。目前平台完成了供热数据统一与共享交换体系，实现供热生产经营等各类数据的整条链路透视化管理；通过机器学习为供热管理人员提供科学、精准的一体化运营解决方案，也为生产人员提供调度和调控的数据指标，最终提高供热管理的精细化程度，实现供热生产的高效管理、精准调控、节能降耗和暖民服务[12-15]。

智慧供热平台由设备感知层、数据存储层、模型调度层、中间应用层与顶层应用层组成。设备感知层由硬件PLC采集器组成，组态通过modbus协议对PLC设备进行数据采集，其采集的数据为秒级实时数据，另一部分则是从其他业务系统通过TCP/IP协议集成的收费、客服与室温数据；通过OPC协议或socket即时通讯协议将组态数据存到内存库中，历史数据则存入到大数据平台的分布式文件库中；GIS数据、实时数据、历史数据通过大数据分析模型、GIS空间处理模型、数据安全模型、机器学习算法等，计算输出到基于微服务架构的中间应用层中，例如地理信息系统、生产调度系统、能耗管理系统等，最终由基于微服务架构的应用系统按需输出到智慧供热平台中。

4 结　语

本文介绍了基于GIS的智慧供热平台在超大规模城市集中供热中的部分关键技术，结合多源数据融合算法，采用LSM树建立了数据融合模型，解决了供热过程中的结构化数据、非结构化数据等多源异构数据格式不统一、来源不统一、无法统一存储、索引低效等问题；介绍了大数据技术中的Hadoop与MapReduce，并采用分布式存储与内存计算，解决了在实际应用中遇到的海量数据存储、高并发、高密度、实时性要求高的问题；通过边缘计算与云边协同，解决了在实际应用中热源、热力站、热用户边缘数据不准、容错率低、云端协同效率低、调节滞后等问题。

最后，基于以上技术的智慧供热平台在实际项目中平稳运行3 a有余，目前平台还处于继续建设阶段，接下来将重点对平台的人工智能方面进行建设，为城市级的智慧供热大脑，为智慧城市建设添砖加瓦。

参考文献

[1] 王艳东,魏广泽. 众源矢量路网融合系统的设计与实现[J]. 地理空间信息,2021,19(6):1-5

[2] 曹玲,唐家银. 基于多源数据融合Bayes修正的可靠性综合评估模型[J]. 甘肃科学学报,2021,33(3):35-41

[3] 谢翌. 基于Hadoop架构的高校科研数据存储系统设计[J]. 现代电子技术,2021,44(11):182-186

[4] 崔晓龙,张敏. Spark作业性能建模及参数优化[J].实验技术与管理,2021,38(3):146-151

[5] 金韬,庄丽婉. 基于区块链的云边协同系统研究与设计[J]. 信息安全研究,2021,7(4):310-316

[6] 袁昊,郭得科. 边缘计算中具有QoS保证的在线能耗感知任务分派[J]. 物联网学报,2021,5(2):71-76

[7] Song K, Zhang B, L W. Research on Parallel Principal Component Analysis Based on Ternary Optical Computer[J]. Optik,2021,3(10):241-248

[8] Zhang K D, Kumar M. Implementation of Parallel Computing and Adiabatic Logic in Full Adder Design for Ultra-low-power Applications[J]. SN Applied Sciences, 2020, 2(8):20-23

[9] 田兴涛. 智慧供热系统关键技术浅析[J]. 中外能源,2019, 24(11):16-20

[10] Liu Y, Li K W. A Two-sided Matching Decision Method for Supply and Demand of Technological Knowledge[J]. Journal of Knowledge Management, 2017, 21(3):592-606

[11] Tetra Laval Holdings &Finance S A. Researchers Submit Patent Application, Method Of Logging Process Data In A Plc Controlled Equipment, for Approval (USPTO 20200142380)[J]. Computers, Networks & Communications, 2020(4):52-58

[12] 齐俊维. 基于LSM-Tree的持久化存储系统设计与实现[D]. 成都:电子科技大学,2019:6-11

[13] khavan A H H, Mohsenian R H. Power Systems Big Data Analytics: an Assessment of Paradigm Shift Barriers and Prospects[J]. Energy Reports,2018,(4):91-100

[14] 许洪强. 调控云架构及应用展望[J]. 电网技术,2017,41(10):3 104-3 111

[15] 潘可佳,王鑫,杨帆,等. 面向电力大数据的多源异构数据融合技术研究[J]. 机械与电子, 2017, 35(9):7-11