第 50 卷 第 11 期：5151-5163　　高电压技术　　Vol.50, No.11: 5151-5163
2024 年 11 月 30 日　　High Voltage Engineering　　November 30, 2024
DOI: 10.13336/j.1003-6520.hve.20230115

基于云-边-端协同的区域级风电场大数据中心数据管理框架及优化运行方法

张扬帆[1]，王　玥[1]，李　阳[2]，沈小军[2]，梁　恺[1]
（1. 国网冀北电力有限公司电力科学研究院，北京 100045；2. 同济大学电气工程系，上海 200092）

摘　要：区域级风电场大数据中心面临运行能效低、交互性差、资源重复建设等问题，难以满足数字化特征凸显的新型电力系统对数据实时计算、绿色计算的需求。该文系统性梳理了风电场大数据中心典型数字化业务的类别、算力敏感性特征和可调节潜力，构建了基于云-边-端协同技术的数据管理架构，从数据交互、业务执行、资源调度和质量治理 4 个维度提出了大数据中心优化运行方法。算例仿真分析了所提架构在数据传输、存储和能耗等方面的性能，结果表明：相较于集中式架构，云-边-端协同可为大数据中心云服务器节约 350%的存储容量，降低 21.04%的能耗，验证了所提架构在实际工程中应用的有效性和合理性。

关键词：区域级风电场；大数据中心；数字化业务；云-边-端协同；数据管理架构；优化运行

Data Management Framework and Operation Optimization Method for Regional Wind Farms Big Data Center Based on Cloud-edge-end Collaboration

ZHANG Yangfan[1], WANG Yu[1], LI Yang[2], SHEN Xiaojun[2], LIANG Kai[1]
(1. State Grid Jibei Electric Power Research Institute, Beijing 100045, China;
2. Department of Electrical Engineering, Tongji University, Shanghai 200092, China)

Abstract: The big data center (BDC) of regional wind farms is faced with challenges such as proliferation of low operational energy efficiency, poor interactivity, and duplication of resources, which makes it difficult to adapt to the demands of data processing for real-time, green computing in new power systems. This paper systematically sorted out the typical digital services and their sensitivity to computing power in wind farms. The spatial and temporal adjustable potentials of BDC loads are also analyzed. Then, a collaborative framework is constructed for data management, which decomposes wind farm big data management into three levels, including cloud, edge and end. The corresponding key technologies on data interaction, business execution, resource scheduling and quality governance are proposed. The experiments are performed to analyze the performance of the proposed architecture in terms of data transfer, storage and energy consumption. The results show that the proposed cloud-edge-end collaborative architecture can save 350% of storage capacity and reduce 21.04% of server energy consumption for cloud servers in big data centers compared to traditional centralized architectures, which further verify the effectiveness and rationality of the proposed architecture.

Key words: regional wind farms; big data center; digitalization business; cloud-edge-end collaboration; data management framework; operation optimization

0　引言

数字化转型是推动新型电力系统建设、服务碳达峰碳中和目标的关键举措，发展以数据中心为代表的基础设施“新基建”已成为我国能源数字化转型的重要战略[1-2]。其中，风力发电经过多年的发展已形成一定的装机规模，成为我国新增电量的主体，截至 2021 年底，我国风电累计装机容量达 328.5 GW，占国内发电装机总量的 13.8%，风电发电量为 6 526 亿 kWh，同比增长 40.5%[3]，风电行业的运营模式由传统的粗放式扩容发展向运行能效提升、全寿命周期管理等精益化管理的转变，加速构建区域级风电场大数据中心成为我国风电产业数字化转型的核心驱动力[4]。

基金资助项目：国家电网有限公司总部科技项目资助(风光储基地数字孪生建模与运行能效提升技术研究及示范) (000202114069A0000)。
Project supported by Science and Technology Project of SGCC (Research and Demonstration of Digital Twin Modeling and Operational Energy Efficiency Improvement Technology for Wind and Solar Energy Storage Bases) (4000202114069A0000).

随着大量物联网智能感知设备的部署和应用，风电场正产生大量的实时数据，且类别多样、复杂度高、质量参差，大数据中心的网络资源、计算资源和存储资源都面临巨大的压力，通信网络和服务器资源的建设进程已难以满足当前数字化业务的算力需求，研究适应新业务模式下数据、资源、网络优化运行方法和管理体系对于提升在运大数据中心的算力水平、服务风电生产运营业务等意义重大[5]。云-边-端作为一种新型的数据计算范式被提出，成为适应差异化业务需求的数据处理架构[6]。其基本思想是将云中心的部分计算功能下放至网络边缘侧，缩短数据的通信链路，减小网络的通信带宽，提升系统对数据的计算和存储性能[7]。云-边-端协同计算在越来越多的领域得到发展和应用，包括：智能电网[8]、车联网[9]、智能机载[10]、轨道交通[11]等。文献[12]以云边协同的数据处理架构为基础，建立了一种新的电力信息物理系统模型，提高了电网对数据感知、传输、处理和存储的能力。文献[13]利用边缘计算技术来解决电力系统计量中心的异常数据检测问题，可提升网络的全局监控能力和虚假数据的识别能力。文献[14]分析了边缘计算在电力系统供需领域的应用，为边缘计算在电力供需领域中的发展提供参考。文献[15]提出了基于边缘计算的能源管理系统，包括架构、关键技术和实现方式。

从数据层面，现有的风电场辅助运维系统多样，服务于电网公司、设备制造单位、运维单位等多方主体，不同系统间相互独立，信息孤岛问题严重，大量数据重复采集和存储，数据质量参差，限制了大数据中心的数据应用效率；从业务层面，云-边-端协同计算在电力系统应用中尚缺乏科学的技术架构和完善的协同机制，差异化的业务需求与云-边-端之间的协同关系还需要进一步的研究；从网络资源层面，现有的研究多侧重于单一机组或单一场站局部功能的更新与升级，缺乏区域内机组、场站间网络资源的协同管控和跨域融合，没有形成“区域-场站-机组”的科学管控体系，使得大量资源重复建设，限制了大数据中心应用生态的绿色发展。

基于此，本文面向区域级风电场大数据中心运行能效提升展开研究，从数据协同交互、业务协同计算和资源协同调度 3 个维度给出解决方案，提升数据处理的实时性，提高业务的执行效率，降低网络能耗。文中系统性梳理了风电场典型的业务及其对算力的需求特征，提出了面向风电场多数据类别、多业务主体的云-边-端协同架构及关键技术，从数据传输性能、存储性能、服务器能耗等维度分析和证明了云-边-端协同架构在风电场应用的可行性和有效性。该文是在现有研究的基础上对区域级风电场大数据中心建设生态的梳理和思考，对当前存在的主要技术瓶颈和未来的研究方向进行展望。

1 区域级风电场大数据中心运行现状分析

1.1 区域级风电场大数据中心面临的挑战

1.1.1 数据处理的实时性有待进一步提升

风电场采集的运行数据数量庞大、种类多，不同场站之间核心设备差异小，运维工作重复性强，同质化程度高。但由于初期缺乏体系化的顶层设计统筹指导，风电场大数据中心目前采用集中式的数据管理模式，各场站采集的运行数据通过数据通信网络上送至区域级大数据中心进行统一处理，计算结果和控制指令再经通信网络回传至设备端。

集中式的数据管理模式虽能满足当前风电场的数据处理需求，但随着业务对算力的需求不断增加，以下问题逐渐凸显：1）复杂、冗长的数据传输链路降低了数据的交互效率，难以匹配延迟敏感型业务的实时性需求；2）数据的集中管理加剧了大数据中心的存储压力、网络带宽和算力负担，大量数据的非必要重复搬运增加了网络的负荷，加大了数据的传输时延[16]；3）服务器的计算性能与数量并不是线性叠加关系，即“1+1<2”，大数据中心可扩展性较差，难以适应数字化应用的更新迭代和可持续发展。

1.1.2 信息孤岛问题突出

现有的风电场运维辅助系统主要包括数据采集与监控系统、设备健康管理系统、故障预警系统、功率预测系统等，各系统软硬件设施采用垂直一体化的封闭管理架构且相互独立并行运行，产生的运行数据分别存储于多个异构系统中，使得信息孤岛问题突出，系统间缺乏有效的信息交互和协同管控[17]。不同系统的软硬件资源重复建设，造成大量硬件、软件资源浪费。不同系统之间采用私有化通信规约的方式进行数据传输，通信通道复杂，信息传输成本高昂，不利于数据中心的低碳高效运行。

有关风电大数据的业务应用中，涉及到气象、运行、调度、营销、负荷等领域的数据，由于存储于不同类别的数据库中，数据的存储类型、格式存在一定差异，使得数据的获取、查询、处理存在诸多不便，加大了原始数据的集成和融合难度。

1.1.3　数据的可用性较差

受网络信号、环境扰动和传感器缺陷等因素的影响，难免会出现缺测、误测等情况，原始数据中存在一定的异常项，由于数据的数量巨大、类别多样、业务需求不明确，使得数据的可用性较低，异常值会对数据的特征提取、规律挖掘等应用产生干扰，有必要结合业务的实际需求，包括数据质量和算力规模的需求，制定一套系统的、面向具体应用的数据质量治理体系，提升业务的应用效率。

1.1.4　大数据中心能耗水平激增

数字化转型推动下，风电场大数据中心网络流量呈爆发式增长，风电运营业务对算力的需求持续上涨，导致大数据中心的算网规模和能耗水平激增。截至 2021 年底，我国在用机架量达到 520 万，近 5 年年均复合增速超过 30%。2021 年我国数据中心耗电量为 1 100 亿 kWh，同比增长 14.6%，占全国总用电量的 1.3%，二氧化碳排放量为 0.78 亿吨(当量)，同比增长 16.4%，约占全国碳排放的 0.7%[18]。因此，大数据中心已成为一种体量庞大、增长迅猛的新兴负荷，优化大数据中心的数据管理架构、资源分配机制对于提升“电力算网”的绿色化水平意义重大。

1.2　风电场大数据中心典型业务类别及需求分析

随着国家电网公司对风电场数字新基建项目的逐步部署和应用[19]，风电场配备了全域的信息感知设备、信息交互网络及智能终端，形成了机组+传感器+光纤网络的“机联网”信息共享架构，应用物联网、大数据、云计算等先进技术，实现对场站运行设备的全域状态感知、多源信息共享和智能应用等数字化管理体系，以加速风电场的数字化转型。为制定合理的风电大数据中心运行能效提升方案，该文首先从时间维度和应用层级梳理了风电业务的类别及其对算力的敏感性需求，如表 1 所示。

1.2.1　大数据中心业务在时间尺度的特征分析

从时间尺度，风电场大数据中心典型的业务可分为实时运行监控业务和非实时优化运营业务 2 类，实时业务为时延敏感型业务，非实时业务属于时延容忍型业务。时延敏感型业务对数据处理的实时性要求高，需要分配高优先级的算力资源。由于业务对象主要是实时数据，单次任务需要处理的数据量级较小，算力规模需求较低，因此此类算力负荷在时间维度的可调节潜力较低。相反，时延容忍型业务对于数据处理的实时性需求不高，但对算力规模的需求较大，负荷水平较高，具备明显的调节潜力，属于大数据中心的柔性负荷，因此可通过调整算力在时间上的执行策略来优化数据中心资源分配，以最大化发挥有限资源的算力水平。

1.2.2　大数据中心业务在应用层级的特征分析

从应用层级来看，风电场大数据中心典型业务总体上可分为设备级、场站级、电网级和政府级业务，如表 1 所示。设备级业务涵盖了风电机组、升压站、箱变、测风塔等核心设备的智能运维，包括设备的运行控制、状态监控等，属于时延敏感型；场站级业务负责风电场重要节点的运行控制调度，时延敏感性和规模敏感性特征突出；电网级业务主要涉及风电场与电网电气连接设备的拓扑结构参数和运行控制数据的计算，指导风电场运行调度、调压、调功率等，同样具有时延敏感性和规模敏感性特征；政府级业务指风电场与政府的交互业务，包括风电场新建、扩建等，属于规模敏感型。

2　风电场云-边-端协同架构及其关键技术

2.1　云-边-端协同架构的构建

2.1.1　风电场云-边-端协同架构的提出

大数据中心负责区域内多个风电场的运维辅助决策工作，算力网络由多个跨区域分布的互联网

表 1　风电场大数据中心业务类别及其算力敏感型特征

Table 1　Typical businesses and their sensitivity to network resources in wind farms big data center

业务类别		数据类别	时延敏感度	规模敏感度
时间维度	实时运行监控	生产运行：生产信息、调度报表、停机列表、异常出力、功率预测、状态视图等	高	中、小
		在线监控：运行监视、故障分析、指标监控、可靠性分析等		
	非实时优化运营	故障统计：场站故障、机组故障、历史故障查询等	低	大
		运行统计：历史数据统计、报警信息统计、历史状态统计、发电量报表等		
		增值数据：风电能效分析、机组故障关联、机组出力关联等		
应用层级	设备级	风电机组运行数据、升压站运行数据、箱变运行数据、测风塔运行数据等	高、中	中、小
	场站级	场站运行监控数据、功率预测数据、故障预警数据、统计业务数据等	高、中	大、中
	电网级	潮流数据、自动电压无功控制数据、自动发电控制数据等	高	大
	政府级	机组年平均利用小时数、发电量、弃风量、装机容量等	低	大

数据中心共同组成，为风电场数字化业务提供持续的算力资源。根据前文对风电场典型业务在时间尺度和应用层级的特征分析，大数据中心的算力负荷在时间尺度具备可调节潜力，在空间维度具备可转移特性。在未来数字化业务深度应用场景下，可通过对“资源+业务+数据”的统筹布局和协同优化，如图 1 所示，完成算力资源在时间和空间维度的优化分配，以匹配业务对数据处理的差异化需求，助推我国新型电力系统和绿色算力网的建设，优化大数据中心能源利用效率(power usage effectiveness, PUE)，提升算力网络的绿色低碳等级，从而实现电网与算网的融合发展。

云-边-端协同技术作为一种新兴的计算架构被提出，为物联网系统中设备运行状态全面感知、采集数据高效计算和广域业务智能应用等提供了一种解决方案。本文将云-边-端协同技术引入到区域级风电场大数据中心运行数据管理应用中，解决多业务主体的资源调度、业务计算和数据交互的需求。有别于传统的“云-端”模式，所构建的“云-边-端”架构将部分计算云中心资源下发至靠近设备端的边缘侧，可大大降低云中心与风电场间的通信负担和算力负载，同时提升业务响应的实时性。其中，云计算可用于处理规模敏感型的全局业务分析和决策，边缘计算则存在数据处理速度快的特点，用于执行时延敏感型的局部业务计算，云计算和边缘计算的特点如表 2 所示。

边缘计算用作云中心计算资源向设备端的延伸，执行本地的实时性计算任务；云计算则执行大规模离线任务的集成计算。通过云计算与边缘计算的协同，解决风电场业务对数据处理的时延敏感型、时延容忍型、规模敏感型和规模容忍型业务的计算任务。鉴于云-边-端协同的计算架构与风电场业务的时空属性和需求特征具有相似性，将云-边-端协同计算迁移到风电场中具有理论可行性。

2.1.2 风电场云-边-端协同架构的构建

为实现风电场“资源-业务-数据”的协同管理，本文搭建了面向区域级风电场大数据中心云-边-端协同的数据管理架构，如图 2 所示。该架构包含云中心服务器、边缘节点和终端设备 3 层，其中，云服务器和边缘节点通过广域的数据传输网互连，终端设备通过本地局域网络与附近的边缘节点通信，此外，某些重要的终端设备也部署专有的广域通信网与云中心交互。通过动态优化资源调度策略、合理布局数据存储方案和有序调整业务的执行顺序，最大化发挥风电场有限网络资源的数据处理能力，减小不必要的数据重复搬运，降低系统能耗，提升场站的数据管理水平和应用效率。

终端设备由部署在风电场不同类别设备的传感器、路由器、控制器、执行器等组成，涵盖了风电场风电机组、升压站、箱式变压器、储能、测风塔等设备，负责风电场内电气、气象、环境、机械等运行数据的采集，并通过专用的光纤通信网络上传至附近的边缘服务器和云中心[20]。此外，端侧设备具备控制指令接收和执行功能，支持控制功能/指令的模块化扩展，为外部资源的接入/输出提供 I/O 接口，同时设置软件更新服务以优化系统配置。

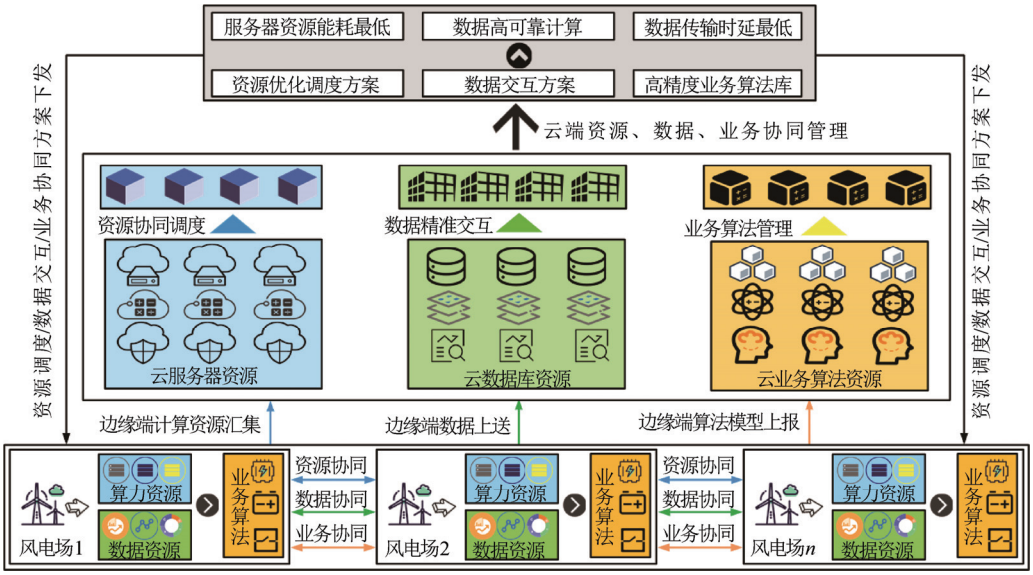


图 1 区域级风电场大数据中心协同运行示意图

Fig.1 Collaborative scheme of regional wind farms big data center

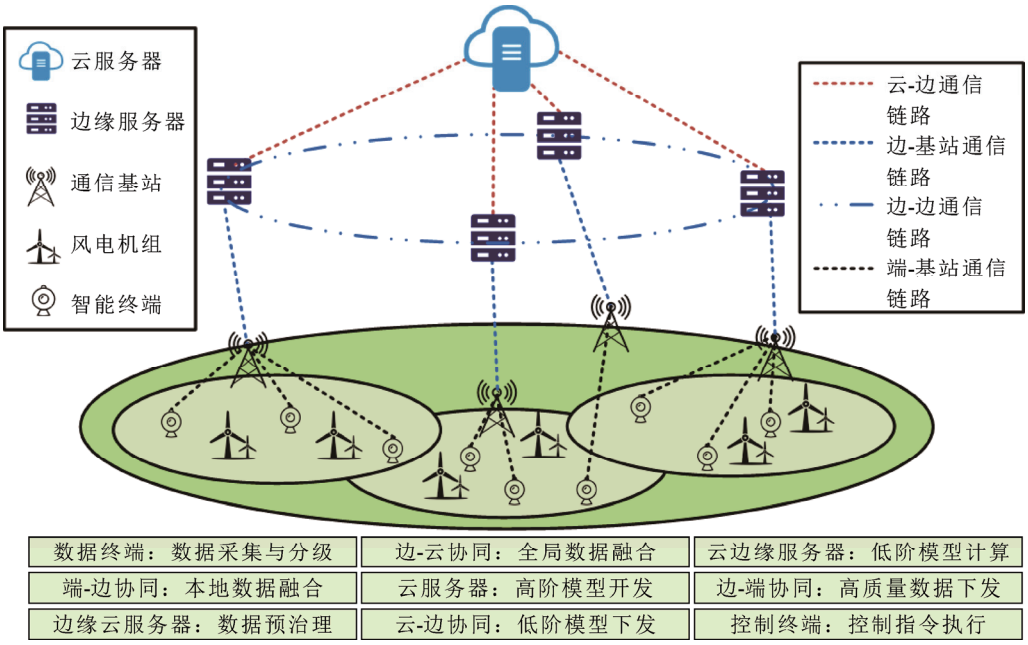


图 2　云-边-端协同的物理架构

Fig.2　Physical layer of cloud-edge-end collaboration

边缘节点包括通信基站、边缘服务器集群、通信网络等，负责终端设备实时数据的获取、汇集和初步处理，包括出力数据、状态数据、检修数据、红外图像数据等，具备数据存储、实时计算和快速决策能力。在实际应用中，边缘节点采用“空间位置+业务”的方式进行划分：1）将空间位置相邻的边缘服务器集群划为一个边缘节点；2）根据业务关系对边缘服务器集群进行二次划分，形成面向特定业务的边缘服务，例如：面向数据质量治理的边缘集群、面向运行数据预测的边缘集群等。在业务部署中，将数据采集、清洗、重构等应用功能封装到边缘算法上，完成软件程序与硬件环境的解耦，提升工程可应用性。

云中心节点集成了云-边-端协同平台、物联网平台和大数据服务，云中心负责对边缘集群的算力资源、数据和业务进行汇集和统一调度，以虚拟化、分布式及服务化等云技术为基础，建立一套服务器资源虚拟化、运行数据规范化、业务应用服务化的数据计算、存储和传输技术支撑体系。此外，云中心融合了风电场多场景的数据分析需求，学习和训练高阶智能化算法，为风电场提供智能的大数据分析模型和应用，发挥云中心大规模数据计算和存储的优势。通过对云-边-端的协同管理，实现边缘节点实时数据与云中心历史数据的交互融合，保证端侧设备控制指令的实时计算和安全下发。

表 2　云计算与边缘计算特点

Table 2　Features of cloud and edge computation

特征参数	云服务器	边缘服务器
服务器位置	大数据中心	设备端
通信网络	全域网	局域网
服务对象	全域风电场	区域风电设备
资源分布情况	集中式部署	分布式部署
访问时延	秒级	毫秒级
网络带宽	兆字节每秒级	吉字节每秒级

2.2　云–边–端协同的关键技术

为解决风电场大数据中心面临的数据处理实时性差、信息交互性弱、资源运行能效低和数据质量参差等问题，本文分别从云-边-端协同架构下的数据协同、业务编排协同、资源协同和数据质量协同治理 4 个方面给出相应的解决方案。其中，数据协同主要是解决风电场运行数据采集、存储和交互问题；业务协同指业务算法模型的拆解、重组、优化、版本管控、运行管理等；资源协同则是对大数据中心网络资源的优化调度和任务分配；数据质量协同治理是提升原始数据的可用性问题。

2.2.1　基于时延敏感性的数据协同方案

为避免由于数据重复采集和搬运而引起的资源浪费问题，本文根据业务数据对时延敏感性特征，提出一种云-边-端协同的数据存储和交互方案。

时延敏感性数据即实时运行监控业务数据，包

括设备、场站、场站与电网连接点等部件的实时运行控制和状态监测数据。由于此类数据的业务价值与其传输、计算过程产生的时延成反比，因此为提升数据的交互效率，将采集的实时数据分发至就近的边缘服务器进行存储，业务模型采用 Docker 技术封装至边缘服务器，数据分析以边缘服务器为主、云服务器辅助的方式执行。即算法的执行在边缘服务器，模型的开发和训练在云服务器，从而发挥边缘服务器集群实时计算和云服务器大规模集成计算的能力。

时延容忍型数据即非实时优化运营业务数据，通常是一些离线状态下的历史数据。业务模型往往涉及大规模、高复杂度的人工智能算法训练和优化，数据的量级较大，类别较多，模型高阶，实时性需求低，算力负荷的时间可调节性强。将时延容忍型数据分发至云服务器进行集中存储，并根据业务算法的需求及优先级，有序分配服务器资源。时延容忍型业务的计算更为复杂，需统筹全域服务器资源的占用率和业务的优先级进行统一调度。

2.2.2 基于空间可转移性的业务协同方案

区域级风电场大数据中心由多个跨区域分布的风电场数据中心边缘子站组成，子站又由面向不同类别设备级业务的服务器组成，即“一云连多边”的网络拓扑结构。不同边缘子站之间通过部署的信息传输通道进行算力负荷的传输，数据中心能耗与算力需求呈正相关的关系，可以实现大数据中心算力负荷的空间转移和算力资源的空间协同共享。例如：同一个边缘子站不同服务器的算力协同、不同边缘子站之间算力的协同、边缘子站与云中心间的算力协同等，实现设备级-设备级、设备级-场站级、场站级-场站级等多级业务间的算力资源和负荷的转移，以优化算力负荷在空间上的分布，提升大数据中心的计算、存储和传输性能，云-边-端协同的业务分配方案如图 3 所示。

3 风电场大数据中心优化运行方法

3.1 云-边-端协同的资源优化调度方法

资源优化调度的目标是以尽可能少的运行开销执行计算任务，优化算力资源在时延容忍性业务和时延敏感性业务间的调度。为此，本文提出了一种云-边-端协同的运行机制，如图 4 所示。当大数据中心接收到业务请求时，首先确定发送业务请求的服务器类别和业务的时延敏感性特征，随即评估本地资源是否能够匹配业务的算力需求。若能够独

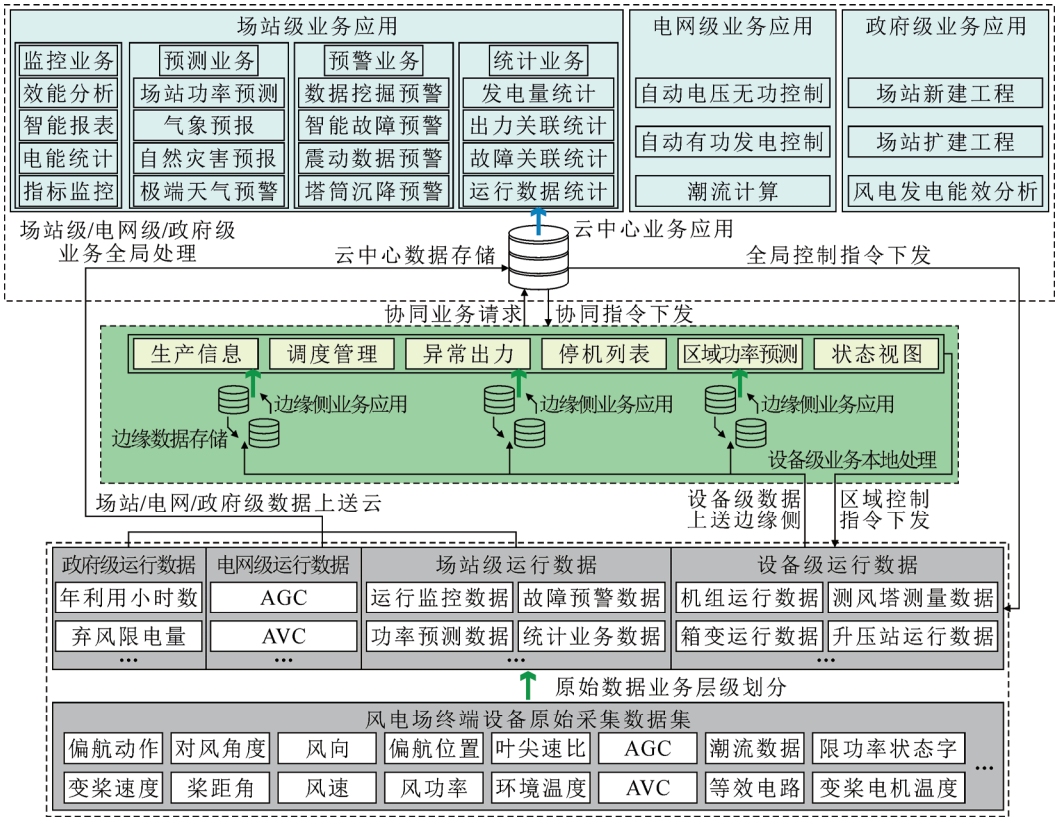


图 3 云-边-端协同的业务分配方案

Fig.3 Businesses allocation scheme under cloud-edge-end collaboration framework

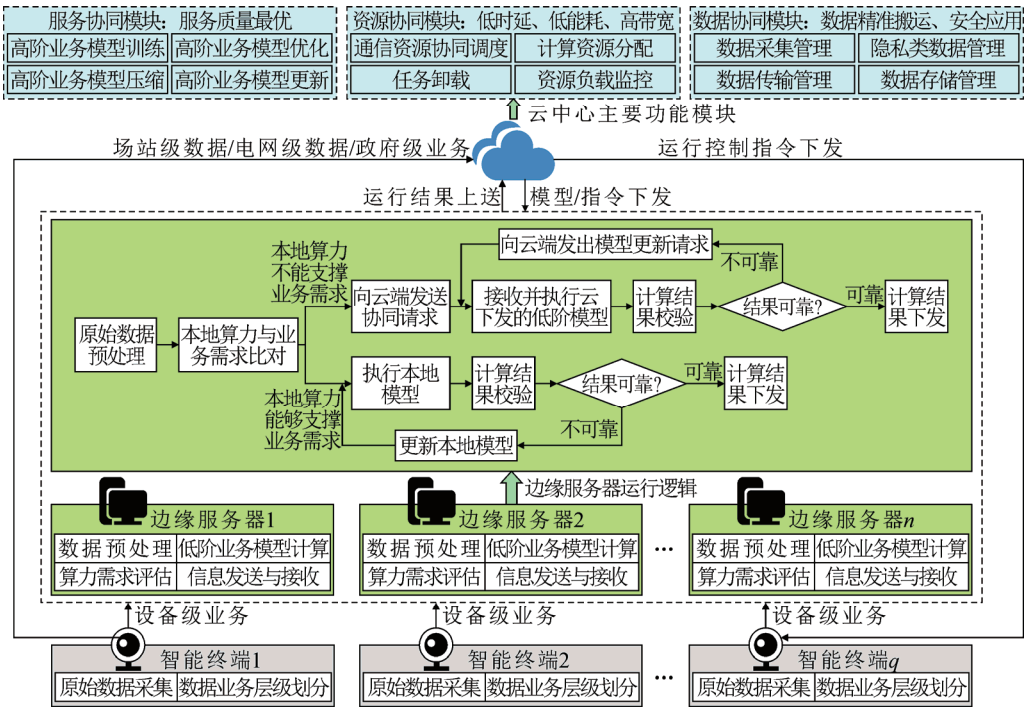


图 4　云边端协同的机制

Fig.4　Mechanisms of cloud-edge-end collaboration

立执行任务计算，则业务直接卸载至本地，由本地服务器管理业务的执行；若本地资源无法匹配业务的算力需求，则向大数据中心发送协同请求，任务则卸载至云中心，由云中心服务器制定全局服务器协同运行策略，执行相应的计算任务。

在所构建的云-边-端协同框架下，云服务器的主要任务是为业务分发合适的边缘服务器资源，基本思想是根据业务在云服务器和边缘服务器的编排情况以及时延特征，制定资源调度策略，执行步骤如图 5 所示。业务通过 HTTP(hyper text transfer protocol)协议发送资源处理请求，大数据中心利用 DNS(domain name server)服务器解析 URL(uniform resource location)域名以确定业务的请求路径和编排位置：1）若业务编排至边缘侧，且属于高复杂度时延敏感型，则采用随机分发的形式执行计算，若为低复杂度时延敏感型或时延容忍型业务，则采用局部优化调度的方式执行计算；2）若业务编排至云中心，则通过全局优化调度的方式执行计算。

综合考虑业务数据的处理时间、数据的网络传输带宽等指标，本文从运行的角度对风电场服务器资源的分配问题展开研究，提出一种基于多目标优化的算力资源调度模型：首先将业务拆解为若干子任务，以数据质量治理业务为例，依据执行关系可拆分为数据预处理、数据清洗、数据重构等子任务，任务集由 $X=\{x_i\}$表示。假设子任务 x_i 在边缘服务器处理时间为 t_{ei}，则总任务在边缘服务器的处理时间为 $T_e=\{t_{ei}\}$，在云服务器的处理时间为 $T_c=\{t_{ci}\}$，子任务 x_i 处理前的数据量 $D_0=\{D_{0i}\}$，处理后数据量 $D_1=\{D_{1i}\}$；子任务 x_i 在边缘侧和云中心执行产生的能耗为 $E_e=\{E_{ei}\}$和 E_{ci}。

根据所提出的业务执行步骤确定参与调度的服务器资源，以云-边-端协同架构下任务处理的时间最优、占用通信资源最低、能耗最小为目标函数。

1）假设业务在边缘服务器和云中心的初始化分配策略由 α 和 β 表示：$\alpha=[\alpha_1, \alpha_2, \cdots, \alpha_n]$, $\alpha_j=(0,1)$，表示任务在边缘侧执行策略；$\beta=[\beta_1, \beta_2, \cdots, \beta_n]$, $\beta_j=(0,1)$，表示任务在云中心执行策略；其中 $\alpha \oplus \beta=1$，表示任务 x_i 必须在边缘侧或云中心执行，且只能执行一次。

2）设置资源分配的多目标函数如下：

边缘处理时间最短，$f_1=\min(\alpha T_e')$；

云中心处理时间最短，$f_2=\min(\beta T_c')$；

传输数据量最小，$f_3=\min(\alpha D_1'+\beta D_0')$；

服务器能耗最小，$f_4=\min(\alpha E_e'+\beta E_c')$；

3）多目标优化方案：

$$\min \varphi_1 f_1+\varphi_2 f_2+\varphi_3 f_3+\varphi_4 f_4 \tag{1}$$

$$\varphi_1+\varphi_2+\varphi_3+\varphi_4=1 \tag{2}$$

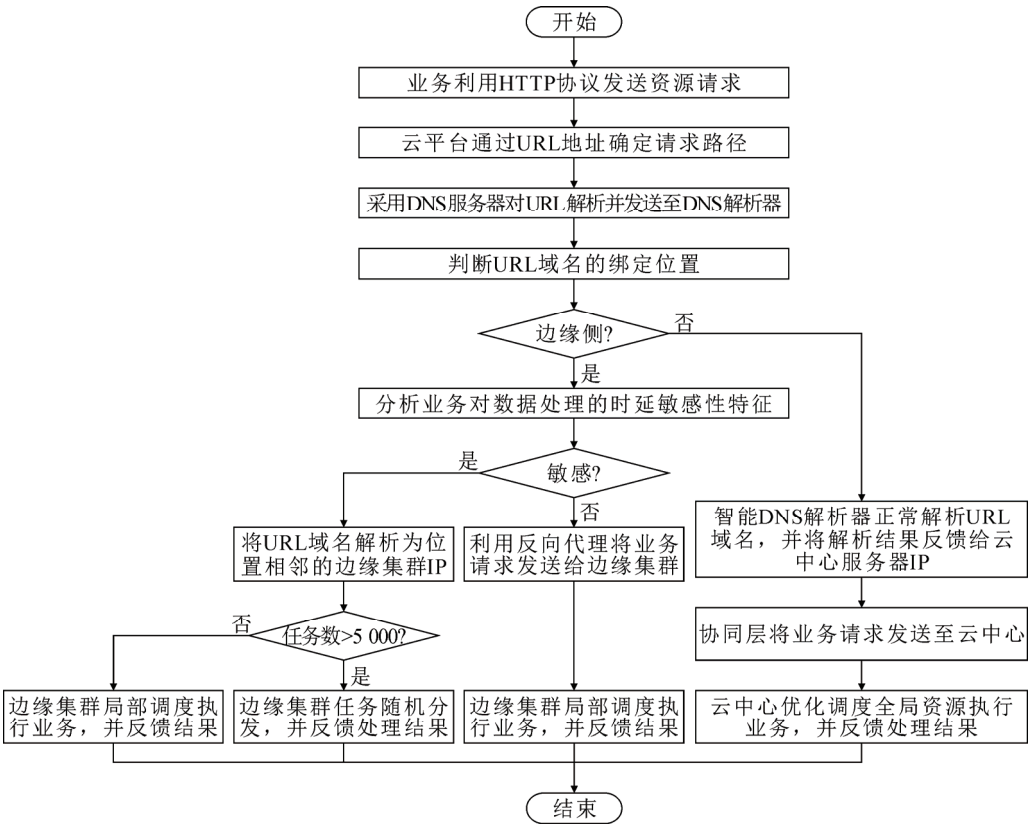


图 5 云-边-端协同的业务执行步骤

Fig.5 Businesses processing steps under cloud-edge-end collaboration framework

式中：φ_1，φ_2，φ_3，φ_4 表示不同子目标的权重分配，本文采用平均分配。

4）目标求解：上述问题为典型的 0-1 分配问题，可采用模拟退火算法进行求解，通过迭代优化的方式，得到最优分配方案。

3.2 云-边-端协同的数据质量治理方法

在梳理风电场数据应用后发现，无论是用于设备服务的行为画像、精准运检等设备运维业务，还是维护风电场安全经济运行的调度规划、资源配置、状态预警等场站经营业务，或是风电并网控制、电能质量等电网保障型业务，亦或是实现碳达峰/碳中和等多重目标的政府级综合型业务，都需要高质量的运行数据作为技术支撑[21-22]。从功能层面，数据质量治理任务可拆解为预处理、数据清洗、异常值重构、质量评估等，任务之间的关系如图 6 所示。

在云-边-端协同架构下，云中心负责执行高复杂度、高聚集度、非实时性的计算任务，诸如：数据清洗/重构规则的开发、训练、优化迭代等，此外，云中心还管理和调度边缘服务器的计算任务；边缘服务器的运行流程主要包含：1）本地模型/规则的执行；2）接收并执行云中心下发的指令和任务。云-边-端协同的数据质量治理任务重整方案如图 7 所示，主要分为：数据采集终端业务、边缘服务器业务和云服务器业务分发。

智能终端作为数据源头，主要负责各类数据的感知和汇集，包括：电气参数、气象参数、环境参数等，此外，终端设备还负责运行控制指令的执行，例如：偏航、变桨、启停等。智能终端设备通过通信网络和工业总线与边缘服务器、云中心连接，实现数据流和控制流在端-边和端-云之间的连通[23]。

边缘服务器在云-边-端协同架构中起承上启下的作用，负责云中心下发任务/指令的接收以及算法/规则的执行，完成对区域内机组采集数据的清洗和重构，并将治理结果下发给机组，实现运行控制和优化运营。边缘服务器在接收到数据采集端上送的运行数据后，首先将原始数据进行本地存储，随后分别执行本地模型和云中心下发的数据处理模型，完成数据的就地清洗，以支撑局部时延敏感性业务，如制定区域发电计划、机组状态监控等。

云服务器负责对边缘服务器上传的业务数据进行质量管理，同时兼具全局网络资源的调度、全局治理模型/规则的开发和更新等任务。

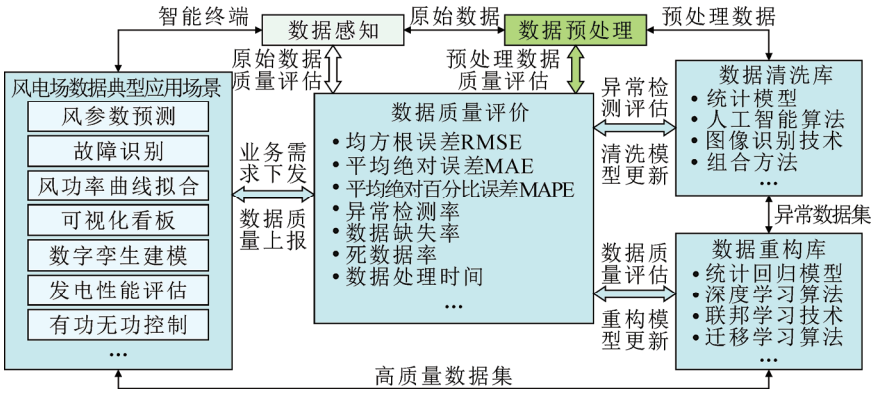


图 6　数据质量治理任务的主要功能模块

Fig.6　Key functions of data quality improvement

1）数据质量管理模块主要负责对数据的汇聚和数据质量的评价。该模块首先对边缘服务器集群上送的数据进行标准化集成，并根据数据的类别、应用场景等信息对数据进行分层管理，由下往上分为：实际运行数据、明细数据、汇总数据和应用数据。此外，该模块设置数据质量综合评估功能，对集成后的数据质量进行分析评估，生成数据质量标签，并根据标签对数据进行分类存储，当接收到某项业务的数据调用请求时，该模块自动完成业务与数据源的匹配。

2）全局网络资源任务调度模块负责大数据中心服务器资源的统一调度管理。结合数据质量治理业务，总体任务包括：数据采集、边缘侧数据预处理、数据深度清洗和重构、数据质量校验等任务。该模块首先统计全网服务器资源的占用情况，根据服务器集群的性能，基于前文所构建的资源调度模型，将计算任务分发至边缘服务器，由边缘服务器负责具体任务的执行。同时，任务调度管理模块设置了边缘服务器运行情况实时监控功能，基于各边缘服务器的业务数据管理范围、计算资源使用情况以及任务的实时执行情况，动态调整任务的分配方式，制定全域网络资源优化调度策略，在确保总体任务能够在各边缘侧被有效执行的同时，提升全局服务器的运行效率。

3）全局模型/规则管理模块负责对数据治理算法的开发、训练、评估和优化，是数据质量治理的关键。预处理模型采用离线的方式嵌入到边缘服务器集群中，包括：缺失值检测、重复值检测、非法值检测、平滑突变点等。深度治理算法的训练和更新在云服务器完成，例如：深度学习算法、机器学习算法等，算法采用离线训练、定期更新的方式进行管理。此外，云服务器还负责对算法的执行效率和性能进行监控和评估，根据数据质量模块反馈的结果动态调整模型的参数，如此反复迭代，以提升风电场全域内数据的质量。云-边-端协同的数据质量治理方案如图 8 所示。

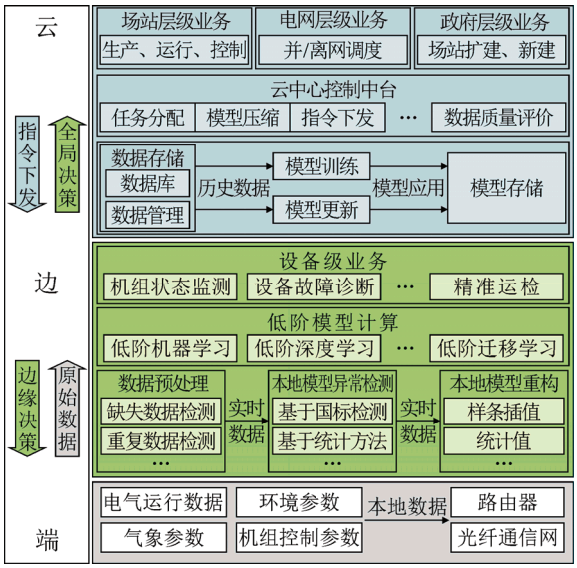


图 7　云-边-端协同的数据质量治理任务重整方案

Fig.7　Strategies of data quality improvement based on cloud-edge-end collaboration

4　算例分析

以我国某区域风电场大数据中心为研究对象，包括 1 台云服务器、7 台边缘服务器和 28 套终端设备(28 台风电机组，每台机组部署一套数据采集装置)，基于实际运行系统的性能统计和仿真测试对比集中架构与云-边-端协同架构在数据传输、存储、资源调度方面的特性，验证所提方案的有效性，大数据中心和仿真环境的物理参数分别见表 3 和表 4。

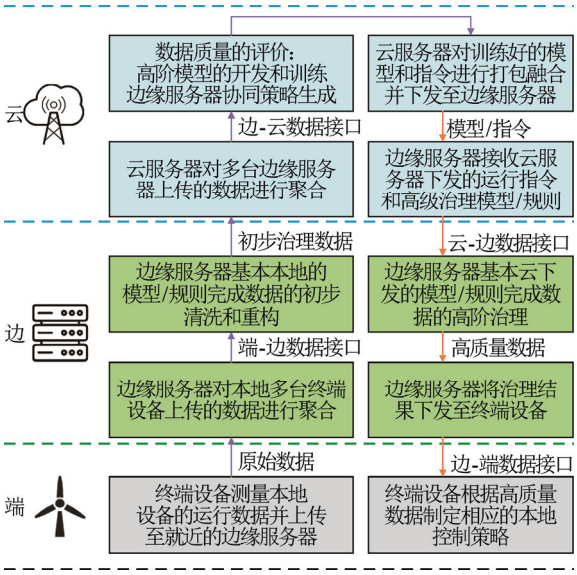


图 8 数据质量治理的云-边-端协同方案

Fig.8 Cloud-edge-end collaborative scheme for data quality improvement

表 3 大数据中心物理主机参数

Table 3 Physical server parameters in big data center

服务器类型	CPU	RAM	存储容量
云服务器	108 核	700 GB	30 TB
边缘服务器	7×2 核	7×2 GB	7×100 GB

表 4 仿真环境物理主机参数

Table 4 Physical server parameters of simulation PC

配置	CPU	RAM	频率	数据库
参数	酷睿 i7-10510U	2.30 GHz	16 GB	MySQL

4.1 数据传输性能分析

与云服务器相比，边缘服务器主要的优势是低网络延迟和高传输带宽。边缘服务器集群以风电场为单位进行划分，所研究风电场设置了 7 台边缘服务器，负责该场站设备的边缘计算业务。边缘服务器部署在风电场，通过场内的局域网络与智能终端设备通信，完成场内设备运行数据的实时采集、传输、存储和计算。边缘计算结果通过广域光纤通信网络上送至区域大数据云中心，完成算法、数据和业务的融合，形成云-边-端协同的数据管理架构。

传统集中架构实际上将边缘侧服务器集成到云中心，形成一个区域级风电场云计算中心，场站设备通过广域网与云中心进行信息交互。云服务器在空间部署上往往远离场站，场站信息需要经过一段较长的传输链路与云中心服务器进行交互：一方面，数据的传输时延会增加；另一方面，由于数据汇集的量级更高，因此网络宽带占用率更大。表 5 为一般情况下的边缘服务器和云中心服务器数据处理的延时特性和带宽特性，即云-边-端协同计算架构下数据处理的实时性和效率更高。假设 2 种数据处理架构同时传输 1 GB 的实时数据，集中架构的传输时延为秒级，协同架构的传输时延可控制在毫秒级，可以看到，云-边-端协同架构数据的传输效率更高，实时性更好。

4.2 数据存储性能分析

云-边-端协同架构的数据管理逻辑是首先安排原始数据在边缘服务器进行初步计算，若边缘服务器无法匹配业务的算力需求，则再将边缘侧的计算结果上传至云服务器进行深度计算，形成全域计算结果或调度指令，然后下发至边缘端，指导设备的运行控制。由于云-边之间只是部分特征参数或控制指令的交互，可大规模压缩数据的传输量级，减少数据的重复搬运，减轻云中心数据的存储压力。在云-边-端协同的数据管理架构下，风电场设备级的原始数据存储在边缘服务器，云服务器可通过通信网络对边缘数据进行远程访问和调用服务，在不影响数据应用的前提下，大大减少云中心的数据存储压力。

以风电机组虚拟风速传感器业务为应用来比较不同架构下的数据存储性能，该业务的输入数据为相邻上风向机组的测量风速数据，输出为目标机组的虚拟风速数据，业务计算流程主要分为如下 3 步。

1）数据清洗：根据对该风电场原始测量数据的质量分析，“脏”数据的占比达到了 10%。在进行业务应用前，需要将“脏”数据剔除，云-边-端架构下数据的清洗任务放在边缘服务器，可避免对这 10%数据的无效搬运。根据统计，1 台风机年产生的数据量大约为 60 GB，则该风电机组每年可为云中心节约的数据存储容量达 6 GB。

2）模型的训练和测试：考虑到算法的复杂度，模型的训练需要在云服务器完成，训练数据集由边缘侧上送至云服务器，云服务器再将训练好的模型参数下发至边缘服务器，模型的测试可在边缘服务器完成，可减少测试集数据上送云服务器。

3）虚拟风速数据映射：实时业务数据的计算放在边缘侧，实时采集的风速数据只需传输至边缘服务器。对于一个含 200 台风机的区域风电场，云-边-端协同的数据管理架构在虚拟风速传感器这一项业务上每年可为云服务器节约的存储空间达太字节级。随着风电机组数量的增多和业务范围的拓宽，

云-边-端协同的管理架构将为数据中心节约更多存储空间。集中架构和云-边-端协同架构下大数据中心每年的数据存储容量对比如图 9 所示，云-边-端协同架构可为云中心节约近 350%的存储空间。

4.3　资源优化调度性能分析

根据对我国某风电场数据中心运行能耗的统计分析，服务器、交换机、路由器、空调、制冷系统的能耗接近总能耗的 90%。其中，服务器设备和制冷系统能耗占比最高，分别为 45%和 35%。据排查，在正常运行状态时，风电场大数据中心的服务器资源利用率不到 20%；而在峰值状态时(时长占比低于 5%)，数据中心服务器的占用率接近 100%，使得难以通过新增计算任务的方式来提升中心服务器的利用率。大多数情况下、大多数服务器设备都处于空闲状态，数据中心能耗一直处于较高水平。

图 10 展示了该风电场数据中心某一部分服务器集群在典型日的服务需求以及在不同计算架构下的能耗情况，服务器的服务需求与其负载率相关，不同负载率下服务器的能耗如表 6 所示。在集中架构下，服务器的任务随机分发，协同架构下，采用文中所构建的多目标协同优化调度方案。可以看到，采取了资源协同优化调度策略后，协同架构的服务器能耗水平明显低于集中式架构，也证明了所提方法对于降低大数据中心能耗水平是有效的。

为了进一步探究集中架构与协同架构下资源的运行差异，选取 20:00 的运行数据进行对比分析，服务需求为 2.4，图 11 展示了集中架构和云边端协同架构下的服务器运行方式及能耗分布情况。集中架构下，任务由 7 台服务器承担，其中，3 台服务器的负载为 0.6，2 台服务器的负载为 0.3，2 台服务器处于待机状态，负载为 0，服务器总功耗为 1 754 W；云-边-端协同架构下，服务器任务分配采用多目标协同优化调度方案，服务器需求不变，业务由 4 台服务器承担，负载为 0.6，1 台处于备用，负载为 0，其余 2 台处于服务器关闭，服务器总功耗为 1 385 W。由此可计算云-边-端协同架构可节约功耗 369 W，为集中架构功耗的 21.04%。

表 7 展示了基于仿真测试生成资源调度指令产生的时延，测试环境的主机物理参数如表 4 所示。可以看到，对于边缘计算业务，往往为轻量级任务，任务数远远小于 1 000 个，由于资源调度所产生的时延为毫秒级，因此基本可以满足时延敏感型业务的任务处理需求。

表 5　云计算与边缘计算时延和带宽对比

Table 5　Comparisons of delay and bandwidth between cloud and edge servers

服务器类型	网络带宽/(Mb·s^{-1})	时延/ms
云中心服务器	<155	>100
边缘服务器	>1 000	<10

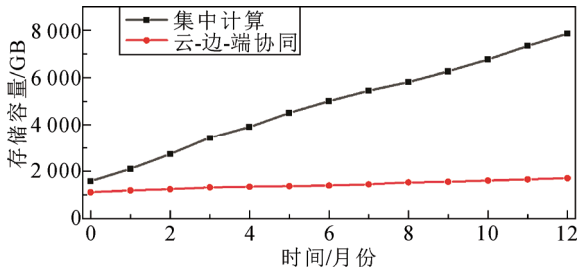


图 9　不同计算架构下云中心每年的数据存储容量对比

Fig.9　Comparisons of annual storage capacity between centralized computing and collaborative computing

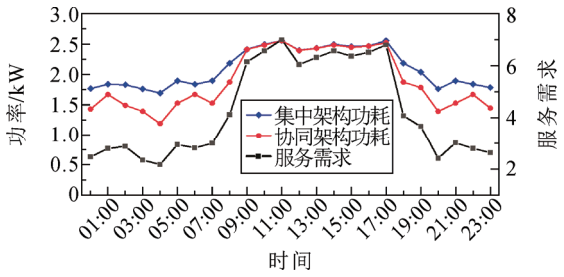


图 10　数据中心负载情况及服务器能耗

Fig.10　Loads and servers energy consumption in wind farms big data center

表 6　不同负载率下服务器能耗

Table 6　Server energy consumption with different loads

负载率	0	0.2	0.4	0.6	0.8	1.0
负荷/W	169	223	269	304	328	364

云-边-端协同架构实际上是通过优化数据中心服务器运行方式来提升资源的利用率，减少了备用服务器的数量，从而可降低系统的服务器能耗。这主要是因为数据中心的计算性能与服务器集中式叠加的数量并不是简单的线性关系，而是存在一个性能阈值拐点，当超过该点后，就难以有效地通过增加服务器的数量来提升系统的数据处理性能，从而会造成资源的浪费。云-边-端协同的架构避免了服务器的集中式叠加，以一种深度交互、资源共享的方式来最大化有限网络资源的数据处理能力，在保证系统运行可靠性的基础上，发挥每台服务器资源的运行性能，减少了冗余服务器的数量，降低了数

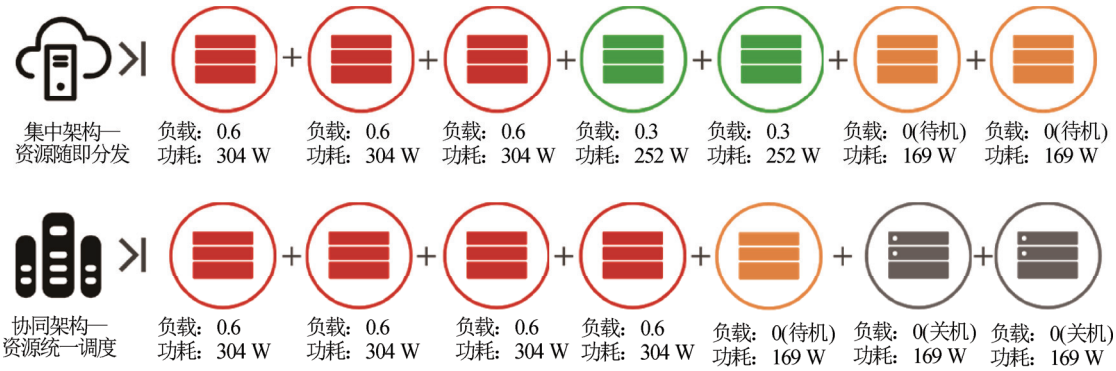


图 11 集中和协同架构下服务器的运行方式及能耗情况

Fig.11 Servers operation mode and energy consumption in centralized and collaborative framework

表 7 资源优化调度产生的时延

Table 7 Delay generated by resources allocation

任务数/个	1 000	2 000	3 000	4 000	5 000
时延/s	0.80	2.39	2.93	4.39	5.02

据中心的运行能耗。有别于传统的“云-端”架构，本文所提云-边-端协同架构从数据协同、业务协同和资源协同 3 个维度制定优化运行策略，根据数据和业务的时延、规模敏感性特征分配合理的网络资源，在保证业务执行可靠性的前提下，优化大数据中心的运行方式，从而达到降低时延、减少能耗的目的。

5 结论

1）区域级风电场大数据中心的算力负荷在时间和空间维度均具有良好的可调节特性，基于“数据-业务-资源”协同的优化调度方案对于提升大数据中心的运行能效是可行的。

2）云-边-端协同的数据管理架构能够有效应用到区域级风电场大数据中心，解决传统集中计算在网络带宽占用过大、数据计算实时性差等方面的不足，协同优化策略可为云服务器节约 350%的存储容量，降低 21.04%的能耗。

3）基于云-边-端协同的数据管理架构通过对风电场运行数据全状态感知及业务算法的全过程管控和网络资源的全周期智能调度，支撑新型电力系统中多业务主体、多时间尺度数据的管理，满足电力绿色算网建设的要求，为大规模风电并网消纳提供算力保障。

未来工作将围绕以下 3 方面的内容展开：

1）研究考虑时延敏感性和规模敏感性的风电场业务多任务拆解和云-边-端重整技术，研究风电场大数据中心算力负荷预测方法，提出考虑云-边-端业务交互同步、分发共享、版本管控、协同训练的智能协同方案，形成业务-模型-数据-算力的自适应优化匹配机制。

2）研究区域风电场大数据中心运行性能画像建模技术，分析大数据中心的性能短板，研究兼顾经济性和可靠性的大数据中心升级部署方案。

3）研究融合区块链技术、数据加密技术和安全认证技术的云-边-端网络安全防护体系，提升接入数据、业务对象和网络终端的安全水平。

参考文献 References

[1] 韩肖清，李廷钧，张东霞，等. 双碳目标下的新型电力系统规划新问题及关键技术[J]. 高电压技术，2021，47(9)：3036-3046.
HAN Xiaoqing, LI Tingjun, ZHANG Dongxia, et al. New issues and key technologies of new power system planning under double carbon goals[J]. High Voltage Engineering, 2021, 47(9): 3036-3046.

[2] 马玉草，张铁峰，许正阳，等. 信息基础设施能耗分析及现状与趋势[J]. 电力信息与通信技术，2022，20(4)：79-87.
MA Yucao, ZHANG Tiefeng, XU Zhengyang, et al. Energy consumption analysis and current situation and trend ofinformation infrastructure[J]. Electric Power Information and Communication Technology, 2022, 20(4): 79-87.

[3] 国家能源局. 2021 年全国电力工业统计数据[EB/OL]. [2023-01-26]. http://www.nea.gov. cn/2022-01/26/c_1310441589.htm.
National Energy Administration. Statistics of China’s electric power industry in 2021[EB/OL]. [2023-01-26]. http://www.nea.gov.cn/2022-01/26/c_1310441589.htm.

[4] 卢子敬，邓艺欣，蒋霖等. “东数西算”背景下大型数据中心建设对电网规划的影响研究[J]. 全球能源互联网，2022，5(6)：552-562.
LU Zijing, DENG Yixin, JIANG Lin, et al. Research on the influence of large-scale data center construction on power grid planning under the background of east digital west computing[J]. Journal of Global Energy Interconnection, 2022, 5(6): 552-562.

[5] WU X, YOU L, WU R F, et al. Management and control of load clusters for ancillary services using internet of electric loads based on cloud-edge-end distributed computing[J]. IEEE Internet of Things

Journal, 2022, 9(19): 18267-18279.

[6] KAI C H, ZHOU H, YI Y B, et al. Collaborative cloud-edge-end task offloading in mobile-edge computing networks with limited communication capability[J]. IEEE Transactions on Cognitive Communications and Networking, 2021, 7(2): 624-634.

[7] 胡金磊，朱泽锋，林孝斌，等. 变电站无人机机巡边缘计算框架设计及资源调度方法[J]. 高电压技术，2021，47(2)：425-433.
HU Jinlei, ZHU Zefeng, LIN Xiaobin, et al. Framework design and resource scheduling method for edge computing in substation UAV inspection[J]. High Voltage Engineering, 2021, 47(2): 425-433.

[8] 周振宇，陈亚鹏，潘　超，等. 面向智能电力巡检的高可靠低时延移动边缘计算技术[J]. 高电压技术，2020，46(6)：1895-1902.
ZHOU Zhenyu, CHEN Yapeng, PAN Chao, et al. Ultra-reliable and low-latency mobile edge computing technology for intelligent power inspection[J]. High Voltage Engineering, 2020, 46(6): 1895-1902.

[9] 许梦瑶，艾小猛，方家琨，等. 考虑用户积极性的电动汽车与机组联合调频的两阶段随机优化调度模型 [J]. 电网技术，2022，46(6)：2033-2041.
XU Mengyao, AI Xiaomeng, FANG Jiakun, et al. Two-stage stochastic optimal scheduling model for joint regulation of EV and thermal units considering users enthusiasm[J]. Power System Technology, 2022, 46(6): 2033-2041.

[10] LIAO Haijun, JIA Zehan, ZHOU Zhenyu, et al. Cloud-edge-end collaboration in air–ground integrated power IoT: a semidistributed learning approach[J]. IEEE Transactions on Industrial Informatics, 2022, 18(11): 8047-8057.

[11] 周　超，林　湛，李　樊，等. 城市轨道交通视频监控系统云边协同技术应用研究[J]. 铁道运输与经济，2020，42(12)：106-110，125.
ZHOU Chao, LIN Zhan, LI Fan, et al. Cloud-edge collaboration technology of CCTV system for urban rail transit[J]. Railway Transport and Economy, 2020, 42(12): 106-110, 125.

[12] 司羽飞，谭阳红，汪　泓，等. 面向电力物联网的云边协同结构模型[J]. 中国电机工程学报，2020，40(24)：7973-7979.
SI Yufei, TAN Yanghong, WANG Feng, et al. Cloud-edge collaborative structure model for power internet of things[J]. Proceedings of the CSEE, 2020, 40(24): 7973-7979.

[13] 钱　斌，蔡梓文，肖　勇，等. 基于边缘计算的电表计量系统数据协同检测方案[J]. 中国电力，2019，52(11)：145-152.
QIAN Bin, CAI Ziwen, XIAO Yong, et al. Data collaborative detection scheme of electric metering system based on edge computing[J]. Electric Power, 2019, 52(11): 145-152.

[14] 李　彬，贾滨诚，陈宋宋，等. 边缘计算在电力供需领域的应用展望[J]. 中国电力，2018，51(11)：154-162.
LI Bin, JIA Bincheng, CHEN Songsong, et al. Prospect of application of edge computing in the field of supply and demand[J]. Electric Power, 2018, 51(11): 154-162.

[15] 祁　兵，夏　琰，李　彬，等. 基于边缘计算的家庭能源管理系统：架构、关键技术及实现方式[J]. 电力建设，2018，39(3)：33-41.
QI Bing, XIA Yan, LI Bin, et al. Family energy management system based on edge computing: architecture, key technology and implementation[J]. Electric Power Construction, 2018, 39(3): 33-41.

[16] 马玉草，张铁峰，许正阳，等. 信息基础设施能耗分析及现状与趋势[J]. 电力信息与通信技术，2022，20(4)：79-87.
MA Yucao, ZHANG Yifeng, XU Zhengyang, et al. Energy consumption analysis and current situation and trend of information infrastructure[J]. Electric Power Information and Communication Technology, 2022, 20(4): 79-87.

[17] 王丽杰，张喜平，冯　强，等. 基于云边协同的新能源监控与大数据平台构建[J]. 分布式能源，2021，6(1)：44-50，55.
WANG Lijie, ZHANG Xiping, FENG Qiang, et al. Construction of new energy monitoring and big data platform based on cloud-side collaboration[J]. Distributed Energy, 2021, 6(1): 44-50, 55.

[18] 陈　敏，高赐威，郭庆来，等. 互联网数据中心负荷时空可转移特性建模与协同优化：驱动力与研究架构[J]. 中国电机工程学报，2022，42(19)：6945-6957.
CHEN Min, GAO Ciwei, GUO Qinglai, et al. Modeling and coordinated optimization for spatiotemporal load regulation potentials of internet data centers: motivation and architecture[J]. Proceedings of the CSEE, 2022, 42(19): 6945-6957.

[19] 周远翔，陈健宁，张　灵，等. “双碳”与“新基建”背景下特高压输电技术的发展机遇[J]. 高电压技术，2021，47(7)：2396-2408.
ZHOU Yuanxiang, CHEN Jianning, ZHANG Ling, et al. Opportunity for developing ultra high voltage transmission technology under the emission peak, carbon neutrality and new infrastructure[J]. High Voltage Engineering, 2021, 47(7): 2396-2408.

[20] 李庆民，于万水，赵继尧. 支撑“双碳”目标的风光发电装备安全运行关键技术[J]. 高电压技术，2021，47(9)：3047-3060.
LI Qingmin, YU Wanshui, ZHAO Jiyao. Key technologies for the safe operation of wind and solar power generation equipment in support of the “peak CO_2 emissions and carbon neutrality” policy[J]. High Voltage Engineering, 2021, 47(9): 3047-3060.

[21] LI Yang, SHEN Xiaojun. Anomaly detection and classification method for wind speed data of wind turbines using spatiotemporal dependency structure[J]. IEEE Transactions on Sustainable Energy, 2023, 14(4): 2417-2431.

[22] 沈小军，付雪姣. 基于改进光滑样条的风电机组功率曲线建模方法[J]. 高电压技术，2020，46(7)：2418-2424.
SHEN Xiaojun, FU Xuejiao. Modeling of wind turbine power curve based on improved smoothing spline[J]. High Voltage Engineering, 2020, 46(7): 2418-2424.

[23] 佘文魁，刘田豹，李　孜，等. 基于 i.MX6Q 的低压配电台区智能终端关键模块的设计与实现[J]. 电力信息与通信技术，2022，20(4)：41-47.
SHE Wenkui, LIU Tianbao, LI Zi, et al. Design and realization of key modules of smart terminal in low-voltage power distribution area based on i.MX6Q[J]. Electric Power Information and Communication Technology, 2022, 20(4): 41-47.



张扬帆
1987—，男，硕士，工程师
主要从事新能源发电并网技术的研究工作
E-mail: zhangyangfanhit@163.com

ZHANG Yangfan



李　阳(通信作者)
1992—，男，博士生
主要从事新能源场站运行设备状态感知与智能诊断、参数预测、数字孪生建模的研究
E-mail: ly18227602440@163.com

LI Yang
Ph.D. candidate
Corresponding author

收稿日期　2023-01-30　修回日期　2023-11-07　　编辑　何秋萍