

基于边端协作的分布式数据分级管理技术

任德旺^{1,2}, 周俊鹏^{1,2}, 倪鑫^{1,2}, 李丽娜^{1,2}, 李亚晖^{1,2}, 李运喜^{1,2}

(1. 航空工业西安航空计算技术研究所, 陕西 西安 710068;

2. 机载弹载计算机航空科技重点实验室, 陕西 西安 710065)

摘要: 随着智能应用在嵌入式装备系统中的不断普及与发展, 对大数据的有效利用和深度融合提出了更高的要求。然而, 嵌入式装备类型多样、大数据多源异构, 由于缺乏统一的组织管理, 业务应用间存在数据壁垒, 并出现了数据孤岛现象, 导致大数据利用与开发困难。因此, 在嵌入式装备系统中, 如何有效组织管理数据是亟待解决的难点。为此, 以嵌入式装备系统的数据统一管理和共享利用为导向, 基于边侧与端侧协作技术, 从数据管理架构、数据分级协同管理方案、数据请求调度策略三个方面, 探讨了嵌入式装备系统中分布式数据协作管理技术的可行性, 以促进数据到信息、信息到知识、知识到决策的快速转化, 同时, 通过管与用的紧密结合, 推动嵌入式装备系统中数据应用的开发。

关键词: 数据管理; 分级协同; 边端协作; 分布式数据

中图分类号: TP309.2; TP311

文献标识码: A

DOI: 10.19358/j.issn.2097-1788.2024.04.006

引用格式: 任德旺, 周俊鹏, 倪鑫, 等. 基于边端协作的分布式数据分级管理技术 [J]. 网络安全与数据治理, 2024, 43(4): 35-40.

Research on hierarchical management for distributed data based on edge-terminal collaboration

Ren Dewang^{1,2}, Zhou Junpeng^{1,2}, Ni Xin^{1,2}, Li Lina^{1,2}, Li Yahui^{1,2}, Li Yunxi^{1,2}

(1. Xi'an Aeronautics Computing Technique Research Institute, AVIC, Xi'an 710068, China;

2. Aviation Key Laboratory of Science and Technology on Airborne and Missile-borne Computer, Xi'an 710065, China)

Abstract: With the continuous popularization and development of intelligent applications in the embedded equipment systems, higher requirements are put forward for the effective use and deep integration of big data. However, there are various types of embedded equipment and multi-source big data is heterogeneous. Due to the lack of unified organization and management, there are data barriers between business applications, and data islands have emerged, resulting in difficulties in the utilization and development of big data. Therefore, in the embedded equipment system, how to effectively organize and manage data is a difficult problem to be solved. To this end, this paper is guided by the unified management and sharing of data in embedded equipment systems. Based on edge-side and end-side collaboration technology, from three aspects: data management architecture, data hierarchical collaborative management scheme, and data request scheduling strategy, this paper discusses the feasibility of collaborative management technology for distributed data in the embedded equipment systems, so as to promote the rapid transformation of data to information, information to knowledge, and knowledge to decision-making. At the same time, this paper promotes the development of data applications in embedded equipment systems through the close combination of management and use.

Key words: data management; hierarchical collaboration; edge-terminal collaboration; distributed data

0 引言

人工智能、云计算、大数据等信息技术在航空嵌入式装备系统中的引入和应用, 促进了人机协同、有人无

人协同等工作模式的不断发展^[1]。协同工作模式对航空智能业务的实时性与准确性提出更高的要求, 如智能态势认知^[2]、综合目标识别与判断打击^[3]等, 需要航空智

能应用做出及时且正确的决策支持,同时需要将航空数据准确转化为算法模型可执行的信息。显然,智能化、协同化的工作模式对航空大数据的组织管理提出了新要求^[4-5]。航空大数据的有效组织管理不仅可以提升航空数据的使用效率、加快智能应用的执行过程,而且能降低数据维护成本、优化航空应用升级的训练运作模式。

在大数据相关战略的指导下,美国空军的“先进作战管理系统”(ABMS)构建统一数据库 dataONE,提供航空数据统一管理能力,为工作人员提供统一的态势感知视图^[6]。同时,美国空军的“下一代空中主宰”(NGAD)运用人工智能系统实现全域传感器感知采集的态势数据的快速解析,为飞行员提供之前无法实现的军事辅助决策能力。同时,美军推出的新型机载数据链,支持异型机之间的数据共享交互^[7]。在国内航空领域,大数据管理体系或系统已应用于飞机故障预测与健康管理^[8]、飞机发动机数据管理^[9]、航空装备质量信息管理^[10]等。然而,在航空嵌入式装备系统中,由于有限的存储、计算、网络资源的约束,大数据管理技术的发展受到限制,导致缺乏统一组织管理^[11]。数据使用困难主要表现在以下几方面:(1)数据来源众多,数据标准规范不一致,数据模型定义不明确,导致数据管理困难;(2)业务系统中存在数据壁垒,数据孤岛现象突出,导致数据共享困难;(3)任务系统相互独立、数据存储分散,导致数据缺乏有效利用和深度融合。另外,航空嵌入式装备系统中,传统的集中式数据存储管理模式为了提高数据存储可靠性,全部数据采用主从备份模式^[9],由于缺乏数据分级管理,容易造成存储浪费,同时集中管理模式^[12-13]容易出现数据访问瓶颈,不利于数据的二次开发利用。针对这些问题,本文提出基于边端协作的分布式数据分级管理技术,首先在航空嵌入式装备系统内构建边侧与端侧协作的数据管理架构,以分布式存储、关键数据集中备份存储的方式,支持数据持久化存储与检索查询;然后设计分布式数据分级管理方案,基于数据属性分析与元数据管理,实现冷热数据的分级协作管理和元数据与主数据分离存储;最后,提出基于边缘缓存的数据请求调度策略,通过数据预先缓存,支持访问请求的快速响应、定位、获取访问。

1 基于边端协作的数据管理架构

为了弥补集中式数据存储管理的不足,本文基于边缘缓存与计算的思想^[14-15],设计分布式与集中式混合的数据管理架构,将与数据相关的存储、计算、处理、分析部署于距离数据源端更近的分布式边缘节点。该架构从资源利用角度出发,有助于充分利用分散的存储与计算资源,节省数据传输带宽资源,同时促进边侧与端侧

的协同融合;从数据本身出发,支持数据持久化存储、关键数据的备份存储,支持数据的定位、访问、获取,支持对数据的计算分析和开发利用。

1.1 分布式与集中式混合的数据管理架构

基于边端协作的数据管理架构由感知层、管理与计算层、数据应用层构成,如图1所示。

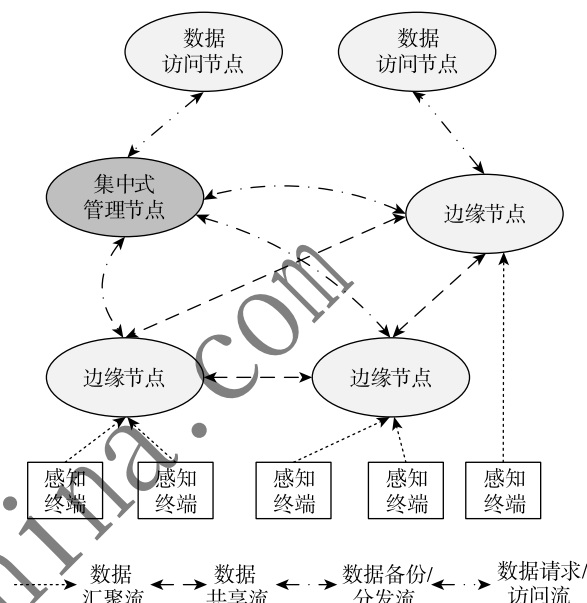


图1 边端协作的数据管理架构

(1) 感知层:感知终端主要由传感设备构成,主要负责数据的采集、按照筛选策略过滤数据、将数据回传至距离最近的边缘节点,以进行进一步的存储、计算、处理、分析。

(2) 管理与计算层:边缘节点和集中式管理节点由综合核心处理机构成,边缘节点负责将不同感知终端的数据进行接入与汇聚、根据数据的结构对数据进行预处理操作、数据本地存储管理;同时,不同边缘节点间可以数据共享与分发;集中式管理节点负责全局元数据管理、关键数据集中备份存储,为跨边缘节点的数据共享分发提供检索查询支持。

(3) 数据应用层:数据访问端主要是显控设备、任务系统等,即数据使用方,负责发起数据请求、从边缘节点获取数据,支持业务应用,根据具体的业务逻辑对数据进行深入的计算处理,以发挥、挖掘数据的价值。

通常,分层架构中各层的角色功能是根据数据生命周期而定的,边缘节点、管理节点、数据访问节点的角色是相对的,边缘节点可能也是访问节点、访问节点也可能是数据存储节点。

1.2 边-端与边-边的融合协作

基于边端协作的数据管理架构中各层节点间既有明

确的分工、又有深度的协作配合，数据从采集到应用的全链路周期中，感知层、数据管理与计算层、数据应用层的功能划分如图2所示，通常，针对某项具体数据的请求访问过程，数据拥有者属于数据管理与计算层、数据消费访问者属于数据应用层。

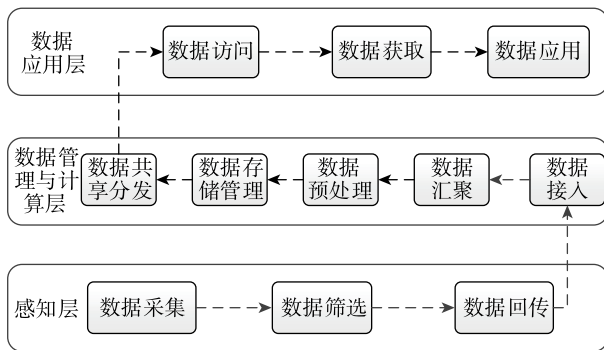


图2 数据感知、存储计算、应用功能划分

1.2.1 “端”与“边”的深度融合

感知终端的基本功能是数据采集、筛选、回传，但随着计算资源向感知终端下沉，感知终端也可以承载边缘节点的预处理功能，按照预先约定的预处理机制处理原始采集数据。相反，若终端的计算资源受限，可以将数据筛选等预处理工作卸载至资源相对比较丰富的边缘节点进行，边与端可以双向利用、相互融合。

1.2.2 “边”与“边”的协作配合

边缘节点之间以及边缘节点与集中式管理节点之间相互协作存储、协作计算，协作存储保证数据存储的持久性以及关键数据的备份管理，协作计算可以支持数据的加速计算与处理。同时，边与边的协作可以解决计算存储资源不足、单节点负载过重的问题。

1.2.3 “管”与“用”的紧密结合

数据的管理与使用相辅相成，数据的规范组织、统一管理可以更好地服务于数据的使用，反之，数据的使用需求完善可指导数据的组织管理。分布式数据管理架构为数据应用提供有效的组织管理，方便数据的快速定位访问，同时数据应用层对数据的使用需求反向促进分布式数据管理架构中数据管理策略的改进与提升。

2 分布式数据分级协作管理方案

基于分布式与集中式混合的数据管理架构，保障航空数据的持久化存储，支持全局数据的一致性共享与管理^[16]。针对航空装备系统中分布式数据管理，设计分级协作管理方案，基于数据属性分析抽取元数据集合，通过元数据定位主数据，进而支持数据“管”与“用”的相辅相成。

2.1 数据标准化与元数据构建

2.1.1 数据属性分析

航空大数据的管理需要兼顾数据的来源、类型、格式、及时性、重要性等诸多属性，综合设计数据管理方案。航空大数据来源众多，既有结构化的时间序列数据，又有非结构化的图像、视频数据；根据数据的热度或访问频次，数据又可分为热数据与冷数据，一般情况下，实时数据的热度高于历史数据，其被请求访问的概率更高；根据数据的重要性，数据可分为敏感数据与非敏感数据，敏感数据的安全可靠问题更突出。根据数据的不同属性，设计不同的数据管理策略，遵循以下基本原则：（1）热数据尽量靠近数据访问节点以便快速获取；（2）敏感数据备份存储，防止数据丢失；（3）数据索引管理轻量且灵活。

2.1.2 数据预处理

数据预处理是在保证数据准确性和完整性的前提下，将原始感知数据规范化、标准化，既方便异构数据的统一管理，又为数据的后续使用与分析提供支撑，同时，尽可能地屏蔽结构化与非结构化数据之间的差异。数据预处理流程如图3所示。针对结构化的时序数据，首先按规定格式转换原始感知数据，然后根据数据标准格式设计数据表结构模型，确定数据的主、外键，并抽取标签信息，最后根据数据所归属主题，对数据进行分类。针对非结构化数据，抽取描述非结构化数据的标签信息，标签信息包含时间戳、数据源标识、数据类型等，对标签信息建立数据表模型，确定标签信息的主、外键，根据数据主题分类数据。

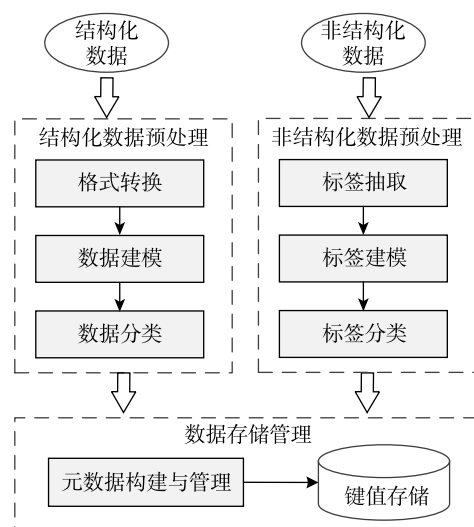


图3 结构化与非结构化数据预处理流程

2.1.3 元数据构建

元数据用于描述数据和数据相关信息，同时用于定

位、识别对应的主数据^[17]，元数据的组成结构如图4所示，主要属性要素如下：（1）标识符：唯一标识相应的元数据，通过标识符可以定位对应的元数据；（2）主题类别：确定数据实体所归属的主题类别，主题类别的设置需要结合业务的类别、属性特征；（3）数据描述：描述数据实体的结构特征，包括数据的来源、类型、格式、长度以及时间范围等信息；（4）存储路径：记录数据实体的具体存储位置；（5）访问控制策略：定义数据的关键等级，以及对应级别的访问权限控制。

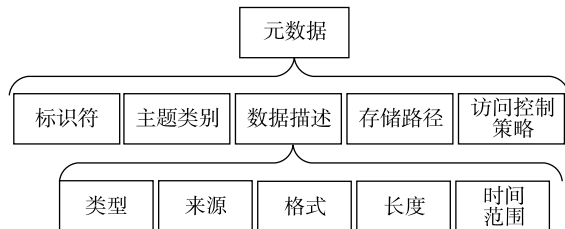


图4 元数据的组成结构

元数据管理用于精准识别、定位和访问数据实体，通过元数据管理，建立数据目录，实现以元数据定位主数据。如图3所示，结构化数据、非结构化数据采用键值存储模式，元数据为键、主数据为值，实现元数据与主数据的分离存储管理，有利于异构数据的统一管理和快速响应获取。

2.2 数据分级协作管理

在分布式与集中式混合的数据管理架构中，设计数据分级协作管理策略，如图5所示。根据数据属性分析结果，按数据热度、数据重要性对数据进行分级管理，边缘节点和集中式管理节点均管理本地的热、冷数据，根据数据访问需求，热数据会存储在不同的节点，而一般情况下冷数据只会单独存储。敏感数据不仅会存储于边缘节点，同时会备份于集中式管理节点。为了提高检索效率，边缘节点会维护本地数据的索引，同时全局数

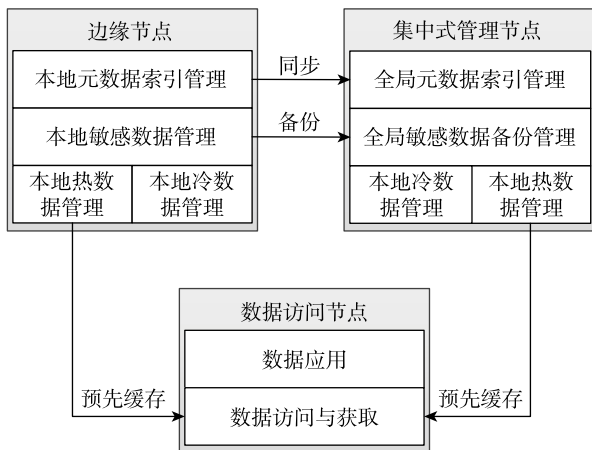


图5 数据分级协作管理架构

据的索引会同步至集中式管理节点，边缘节点的数据索引是集中式管理节点数据索引的一部分。因此，集中式管理节点管理本地数据、全局敏感数据、全局元数据索引。显然，分布式的边缘节点与集中式管理节点相互协作配合，元数据统一管理、敏感数据统一备份。

2.2.1 边缘节点与集中式管理节点间的协作

所有节点基于键值存储模式管理本地数据，元数据与主数据分离管理、热数据与冷数据分开管理。边缘节点的元数据索引信息定期同步至集中式管理节点，边缘节点上的敏感数据在集中式管理节点中备份存储，因此集中式管理节点负责维护全局元数据信息以及全局的敏感数据。集中式管理节点为全局边缘节点、数据访问节点提供统一索引和查询方式，同时，全局元数据索引信息会定期更新。

2.2.2 数据管理节点与数据访问节点间的协作

航空装备系统中，最新感知数据的热度高于历史感知数据，表征最新感知数据被请求访问的可能性高于历史感知数据。同时，根据二八定理，只有少部分的数据属于热数据，而大多数数据属于冷数据。因此，冷热数据分级管理，一般情况下，热数据的管理更有助于提高数据的使用效率。边缘节点和集中式管理节点中的数据管理策略服务于数据访问节点，数据访问请求模式、访问数据类型均会反向优化数据管理策略，数据管理节点根据数据访问模式等调整管理策略，以适应数据访问节点的访问请求。

根据移动边缘网络中数据访问、请求重定向模式，数据访问节点直接访问距离其最近的边缘节点，如果此时边缘节点缓存了请求的数据，则直接响应，否则会将请求重定向至集中式管理节点获取数据索引，由数据索引指向具体的边缘节点获取请求数据。

2.3 元数据与主数据映射

数据映射负责依据数据请求和数据的存储规则，动态地将数据请求映射成元数据，再将元数据映射成数据实体。如前文所述，元数据标识符是唯一的，用于定义一个主数据。基于哈希表，维护元数据标识符与元数据的映射关系，通过元数据识别确定对应的数据实体。元数据与主数据之间的映射逻辑如图6所示，通过解析数据请求，提取元数据标识符，并基于哈希表搜寻元数据信息，从元数据信息确定数据实体的具体存储位置，根据存储路径获取请求的数据实体。

3 基于边缘缓存的数据请求调度策略

数据“管”的目的是服务于数据“用”，进而发掘数据的价值。航空业务的实时性需求高，不仅要求精准调

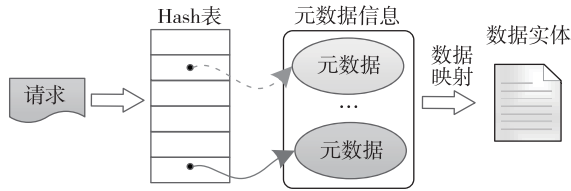


图6 请求-元数据-数据实体之间的映射关系

度数据请求，而且需要快速获取数据，为此，本文设计基于边缘缓存的数据请求调度策略，即结合边缘缓存技术与时间局部性原理，将即将或近期需要使用的数据放置在距离数据访问节点更近的位置，并根据动态变化的数据请求不断调整、更新缓存的数据。同时，请求调度遵循就近优先原则，首先检测本地是否存在请求数据，若存在则直接从本地获取；否则，根据调度策略获取数据。

3.1 请求感知的数据预先缓存策略

数据访问节点作为数据请求节点，从边缘节点或集中式管理节点获取相应的数据实体，若频繁从数据存储节点获取数据，难以保障业务的实时性需求^[18]。为此，根据数据本身属性和被请求访问频次的汇总，将数据提前缓存至数据访问节点本地或者直连的边缘节点上，以降低数据访问时延，尽可能让数据请求就近被响应，具体如图7所示。首先，结合数据的实时特性以及历史请求记录，综合计算数据的排名度；然后，根据排名度提前缓存数据，策略为：排名度靠前的数据提前缓存至数据访问节点，如图7中的A数据，中间排名度的数据放置在与数据访问节点直连的边缘节点上，如图7中的B数据，其余相关的排名度低的数据缓存至非直连边缘节点上。

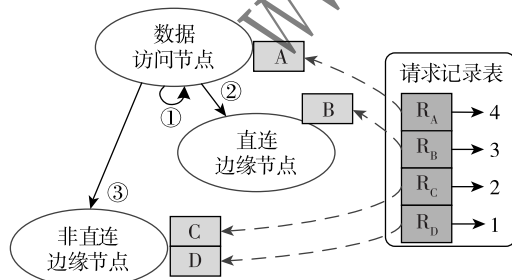


图7 基于预先缓存的数据请求调度机制

假设图7中本地访问、直连边缘节点访问、非直连边缘节点访问的代价分别为1、2、3，总请求数为10。预先缓存策略如图7所示，总访问代价为： $4 \times 1 + 3 \times 2 + 2 \times 3 + 1 \times 3 = 19$ ；不缓存策略中A数据在直连边缘节点上，总体访问代价为： $4 \times 2 + 3 \times 2 + 2 \times 3 + 1 \times 3 = 23$ 。对比基于预先缓存策略和不缓存策略的平均访问代价，如表1所示。

表1 预先缓存策略效果分析

策略	总访问代价	平均访问代价
预先缓存策略	19	1.9
不缓存策略	23	2.3

3.2 数据请求调度策略

数据访问节点获取数据的流程如图6所示，包括请求、元数据标识、元数据、数据实体，因此缓存数据包括元数据与数据实体，根据请求定位数据的存储位置包括本地、直连边缘节点、非直连边缘节点、集中式管理节点。请求调度基本流程为：解析请求、定位元数据、获取数据实体。本地解析请求，提取元数据标识，定位元数据位置，若在本地，则直接访问；若不在本地，则将请求转发至对应的边缘节点，从直连或非直连边缘节点获取。同时，请求是动态变化的，需要根据请求变化信息，动态更新缓存的数据，包括元数据和数据实体。

4 结论

针对航空嵌入式装备系统中数据组织管理的新需求，本文提出基于边端协作的分布式数据分级管理技术，旨在通过数据的统一组织管理，促进数据的有效利用与深度融合。首先，依托嵌入式装备系统内的感知设备与综合核心处理设备，构建边端融合的数据管理架构，通过边-端协作、边-边协作，完成数据从感知到应用的全链路流程；然后，基于数据标准化与数据属性分析，以数据分级、节点协作为核心，设计元数据与主数据分离的分级协作管理方案；最后，为了支持数据的快速访问，结合边缘缓存与时间局部性原理，设计请求感知的数据预先缓存策略和就近优先的请求调度策略，以提升数据的访问与使用效率。

参考文献

- [1] 石小刚, 邓武振. 基于人机混合的协同作战新变化 [N]. 中国社会科学报, 2022-01-06 (006).
- [2] 崔令飞, 郭永红, 邵鹏志, 等. 运用时空大数据增强战场态势感知能力 [J]. 国防科技, 2021, 42 (2): 127-132.
- [3] 温鸿鹏, 刘捷. 运用大数据增强战场态势感知能力 [J]. 中华医学图书情报杂志, 2018, 27 (4): 33-36.
- [4] 石起斌. 航空大数据技术的发展与应用研究 [J]. 科技经济导刊, 2019 (27): 1.
- [5] 孔祥芬, 蔡峻青, 张利寒. 大数据在航空系统的研究现状与发展趋势 [J]. 航空学报, 2018, 39 (12): 1-16.
- [6] 胡春宇, 刘卫东. 美国作战数据发展现状研究及启示 [J]. 飞航导弹, 2020 (7): 46-49.
- [7] 韩戈白, 张海越, 柳永齐, 等. 美军作战综合保障态势一张图研究综述 [J]. 信息化研究, 2021, 47 (6): 1-5.

- [8] 曲昌琦, 周锐, 杜宝, 等. 航空装备故障预测与健康管理的
数据体系研究 [J]. 航空科学技术, 2020, 31 (12): 61
- 67.
- [9] 陈庆贵, 谢静, 蔡娜, 等. 军用航空发动机大数据管理系
统设计开发研究 [J]. 兵工自动化, 2021, 40 (7): 21
- 23.
- [10] 罗璇. 浅谈基于大数据的航空装备质量信息管理 [J]. 装
备维修技术, 2018 (4): 22 - 26.
- [11] 赵学武, 吴宁, 王军, 等. 航空大数据研究综述 [J]. 计
算机科学与探索, 2021, 15 (6): 999 - 1025.
- [12] 付远祥, 张驰, 邢磊, 等. 数据元标准化在航空装备数据
治理中的应用 [J]. 中国标准化, 2024 (5): 110 - 117.
- [13] 殷华杰, 张彦, 王凯. 基于大数据的航空数据采集与处理
系统研究与设计 [J]. 航空电子技术, 2020, 51 (2): 11
- 15.
- [14] 王莹, 费子轩, 张向阳. 移动边缘网络缓存技术 [J]. 北
京邮电大学学报, 2017, 40 (6): 1 - 13.
- [15] TRAN T X, LE D V, YUE G, et al. Cooperative hierarchical cac-
hing and request scheduling in a cloud radio access network [J].
IEEE Transactions on Mobile Computing, 2018, 17 (12):
2729 - 2743.
- [16] 杨子剑, 胡俊, 蔡静. 基于分布式数据管理系统的读写性能
优化研究 [J]. 信息技术与标准化, 2022 (6): 29 - 34.
- [17] 焦利, 孙松周, 刘天须, 等. 元数据驱动的分布式数据资
源管理技术 [J]. 计算机与现代化, 2019 (3): 78 - 84.
- [18] REN D, GUI X, ZHANG K, et al. Adaptive request scheduling
and service caching for MEC-assisted IoT networks: an online
learning approach [J]. IEEE Internet of Things Journal, 2022,
9 (18): 17372 - 17386.

(收稿日期: 2024 - 03 - 19)

作者简介:

任德旺 (1989 -), 通信作者, 男, 博士, 工程师, 主要研
究方向: 大数据技术、数据管理与应用。E-mail: rende-
wang999999@163.com。

周俊鹏 (1992 -), 男, 硕士, 助理工程师, 主要研究方
向: 大数据分析、数据库管理。

倪鑫 (1993 -), 男, 博士, 工程师, 主要研究方向: 大数
据应用、数据库技术。



版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.com