

基于嵌入式工业时序数据的存储策略及应用研究

周俊鹏^{1,2}, 王卓^{1,2}, 康永乐^{1,2}, 任德旺^{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: 嵌入式工业计算领域具备独特的数据优势, 海量数据呈现时序递增、多源异构、复杂多变等特点。针对现有嵌入式工业时序数据发展特性及高效存储管理代价问题, 融合时序数据库技术, 提出一种基于嵌入式工业时序数据的存储策略, 有效缓解嵌入式工业计算设备的海量时序数据高效存储及组织管理瓶颈, 并实施了时序数据库系统的有效性验证, 进而加速推进时序数据库技术在嵌入式工业计算领域的应用实施。

关键词: 海量数据; 时序数据库; 高效存储; 组织管理

中图分类号: TP311.5

文献标志码: A

DOI: 10.19358/j.issn.2097-1788.2026.02.006

中文引用格式: 周俊鹏, 王卓, 康永乐, 等. 基于嵌入式工业时序数据的存储策略及应用研究 [J]. 网络安全与数据治理, 2026, 45(2): 44-49.

英文引用格式: Zhou Junpeng, Wang Zhuo, Kang Yongle, et al. Research on storage strategy and applications based on embedded industrial time series data [J]. Cyber Security and Data Governance, 2026, 45(2): 44-49.

Research on storage strategy and applications based on embedded industrial time series data

Zhou Junpeng^{1,2}, Wang Zhuo^{1,2}, Kang Yongle^{1,2}, Ren Dewang^{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: The field of embedded industrial computing has a unique data advantage, with massive data characterized by incremental sequencing, multi-source heterogeneity, and complex variability. Aiming at the development characteristics of embedded industrial time series data and the problem of high costs associated with efficient storage management, the paper proposes a storage strategy based on embedded industrial time series data, incorporating time series database technology. The strategy effectively alleviates the bottleneck challenges of efficiently storing and organizing management of massive time series data in embedded industrial computing devices. The paper implements the effectiveness validation of the time series database system, and accelerates the applications implementation of time series database technology in the field of embedded industrial computing.

Key words: massive data; time series database; efficient storage; organizing management

0 引言

随着工业大数据、物联网、智能制造等技术的迅猛发展, 越来越多的源端设备正在无时无刻地产生数据。嵌入式工业计算领域被赋予先天的数据基因^[1], 各个计算设备在全生命周期内会形成一系列的设备数据集, 其中, 80% 以上的数据集都是实时性数据^[2], 均按照时间序列顺序产生, 包括浮点数据、音视频数据、文本数据等多样化数据集。通过对这些数据集进行高效采集、存储及组织管理, 深度挖掘数据发展的

潜在性特征, 能够提升嵌入式工业计算设备的安全可靠及高度智能。

当前, 由于嵌入式工业计算设备的硬件成本要求、存储资源限制及系统环境特殊等制约因素, 致使大量实时性数据流失且难以发挥应用效能。源端设备数据呈现分散、孤立现象^[3], 存在数据获取受限、单一化服务及存储管理效能滞后等问题。而传统数据库技术主要从数据本身实体关系之间设计模型结构, 难以从业务角度直观性反映某个类型设备的数据变化情况。

数据建模方面更加注重表之间的关联关系和结构特性，数据的存储成本较高，读写性能相对较低，可维护成本略高，难以满足嵌入式工业时序数据的存储、管理及分析等应用场景^[4]。同时，现有的存储策略并未充分利用开源时序数据库技术的优势，导致存储系统的扩展性、灵活性较差，难以应对工业生产中不断变化的数据需求。面对有限的嵌入式工业计算设备资源，提升海量实时性数据的高速读写、高效存储管理、低延迟查询及高效聚合分析等能力，成为解决嵌入式工业时序数据泛滥的本质策略和首要任务。

本文针对存在问题，深入研究嵌入式工业计算设备的时序数据结构发展特性，融合现有开源时序数据库技术^[5]，构建了一种基于嵌入式工业时序数据结构模型和存储策略，提出了时序数据库存储体系结构及设计方法。新模型和策略能够更高效地处理海量时序数据，提升数据存储和访问效率。本文通过嵌入式工业计算设备环境及时序数据库系统的有效性应用验证，表明研究策略的可行性效果。

1 嵌入式工业时序数据结构模型

1.1 数据结构特性分析

基于嵌入式工业计算应用的特殊场景，工业设备数据呈现类型多样化和爆炸式扩增。工业数据通常由成千上万台传感设备全生命工作周期产生，数据采样周期频率堆叠，数据之间存在较复杂的依赖性关系，具有周期性采样密集和强关联特性。数据产生规律的统一指标，本质是按时间顺序记录的一系列有序数据集，不存在复杂嵌套的数据结构，且每个时间点上都存在对应的指标性数据，数据结构主要由数据主体、时间点及采样值三部分共同组成^[6]，如图1所示。数据主体表示采样对象，一个数据主体包含一个或多个采样值，每个采样值存在相对应的一个具体时间点和采样指标。



图1 时间序列数据结构

嵌入式工业计算设备产生的数据主要包括飞参数据、总线数据、雷达数据、传感器数据、制造数据、状态数据、设备参数、信息系统数据及维护数据等^[7]，这种类型的数据存在时间序列化、产生频率化、采样时段集中化及采样指标唯一化等一系列特性规律，潜在反映各个计算设备在时间变化过程中的工作状态及异常性问题。围绕嵌入式工业时序数据演变过程，以数据高效存储及组织管理为核心目标，精准分析数据结构特性，如图2所示，主要包括海量性、时效性、时序性、重要性、读写频率、实时性及重要性。

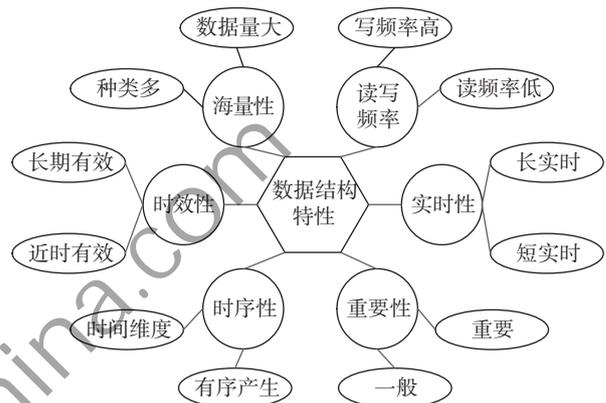


图2 嵌入式工业时序数据结构特性

(1) 海量性：嵌入式工业时序数据的海量性特征主要体现在数据产生源种类繁多，且数据量呈现TB或PB甚至更高规模，难以通过传统的数据库技术进行存储、处理及管理。

(2) 时效性：嵌入式工业时序数据具备一定的时效性，长期有效存储的历史数据需要反复使用或者周期性训练，批量分析历史变化的趋势或决策研判，数据集通常需要存储在廉价式大容量的电子磁盘，使用之后需要进行定期批量删除。近时有效的数据更多关注的是短期性数据变化情况，数据从产生到消耗所经历的时间滑动窗口非常小，注重数据的瞬息万变，处理速度要求相对较高^[8]，大多都存放在内存上，快速处理后，只需将中间计算结果保留至内存，数据的生命周期结束。

(3) 时序性：嵌入式工业时序设备产生的数据严格遵循全生命周期的时间维度，几乎每条数据均按照时间基准有序递增，且数据之间的结构关系通过所产生数据的时间戳相互维系，注重数据的时间演变规律。

(4) 读写频率：嵌入式工业时序数据产生频率快，更加注重数据的高性能写入，数据写入频率远高

于读取频率,只有当用户发起读请求或定期读取的程序指令时,服务端才会响应请求服务,而大部分时间都需要响应写操作请求,保证数据能够快速存储至大容量磁盘。

(5) 实时性:嵌入式工业时序设备的种类不一,工作机制不同,产生数据的方式存在差异^[9]。一部分设备需要全时工作,所产生的数据属于长实时数据,特点是数据的变化量在时间顺序下随之持续递增,直至设备停止工作。另一部分设备定期工作,只在短时间范围内配合其他设备工作,其源端产生的数据属于短实时数据,只有在某一时间范围内数据变化量处于间断递增。

(6) 重要性:嵌入式工业时序数据的重要性指代数据自身的固有属性,反映数据在一定时间范围内参与处理计算且影响趋势变化及决策研判因素的重要程度,包括重要数据和一般数据。

1.2 数据结构模型构建

基于嵌入式工业时序数据的结构特性分析,按照数据主体、时间点和采样值三部分构建数据结构模型。

其中,数据主体以采样对象形式映射,包括设备标识、设备编号、设备类型及设备端口等唯一确定设备标识信息;时间点表示设备数据产生的时间戳信息;采样值取决于设备产生数据的多元化指标参数类型^[10]。数据模型需要考量数据的存储形式,以便提供规则化、便捷化的数据高效分析能力。如图3所示,定义数据结构模式需要考虑数据结构表存储冗余、查询性能、小文件数量繁多、磁盘空间利用率等因素,结合嵌入式工业时序数据的演化特性,构建高可靠稳定的数据结构模型。首先,将设备标识、设备类型、设备编号及设备端口进行统一格式化,形成唯一可标识化的设备信息索引标识码,构建完备的索引列表。其次,按照设备索引标识码逐步构建对应的时序数据结构表,每张结构表中包括设备采样时间信息和各个测点的采样值信息。最后,按照时间信息对每张结构表进行自主化逻辑分区,确保数据能够按照设备信息索引标识码定位到对应设备的数据结构表,并通过数据结构表按照时间的逻辑分区定位到具体时间段,按照时间段定位到具体的测点采样数据。



图3 嵌入式工业时序数据结构模型

构建嵌入式工业时序数据结构模型能够有效管理嵌入式工业时序数据集,避免各个数据结构表出现大量数据冗余,减少小文件存储数量,提高磁盘利用率。从数据结构模型的构建环节优化数据高效存储及管理问题,提升后期数据的效能价值。

2 时序数据库存储策略

2.1 存储体系架构

以时间戳为核心的数据存取,是整个时序数据库存储体系架构设计的关键点。为了实现可扩展、高可靠及高可用的时序数据存储系统,本文基于嵌入式工

业时序数据的发展特性,围绕时序数据结构模型的构建过程,融合时序数据库技术原理,设计出一种基于C/S模式的时序数据存储体系架构,如图4所示。客户端主要负责与用户交互,接收用户的输入信息或命令请求,并发送请求到服务端,而服务端接收并解析客户端的请求命令,根据对应请求处理并返回执行结果。

由于时序数据通常具有高频率、快速写入和长期存储等特点,设计有效的存储策略对于性能优化至关重要。为了保证写入的高效,时序数据存储策略是必

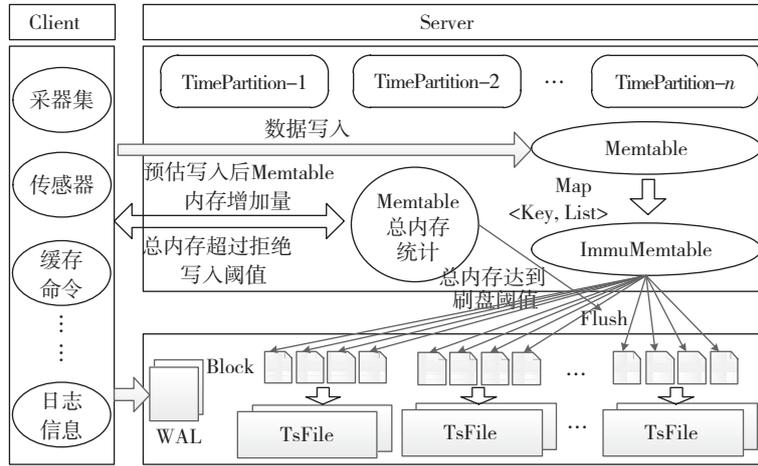


图4 时序数据存储体系架构

要基础支撑，该策略能够封装并构建对时序数据存储的数据结构模型中的抽象。基于时序数据的存储体系架构，需要通过时序数据特性构建数据结构模型，然后进行压缩，并发送到存储层。通常时序数据需先写至内存中的数据结构，当内存容量达到一定阈值后 Flush 到文件，为了避免数据丢失也会写入 WAL 文件。WAL 文件采用双副本异地备份，通过定时和定量触发机制，保证内存数据定时生成快照，或者单个 WAL 文件大小达到预设阈值，自动触发备份，确保内存数据 Flush 失败时可通过 WAL 恢复，实现数据完整性。同时，时序数据写入内存后需按照数据主体特性结构的形式组织，数据主体特性结构主要描述工业设备源端的数据结构特性，在内存中主要以唯一序列化 Key 存储在 Map 结构，Map 中每个 Key 对应各自 List 列表，而 List 中主要存储时间点数据。数据写入过程会根据工业时序数据的特性结构模型组织形成序列化 Key，同时，将对应的时间点 Timestamp 和采样值 Value 组合值写入到时间数据列表 List 中。内存中的数据 Flush 到文件后，将同一个 Key 中的时间点数据写入同一个 Block 块内，即一个 Block 块内的数据都属于同一个数据主体。时间点 Timestamp 和采样值 Value 可以在同一个 Block 内分开独立存储，独立存储可以对时间列及数值列分别高效压缩，对于时间点和采样值，可根据其各自数据类型的特点，采用特定的压缩算法提高压缩效率，进而提升数据写入的高效性。此外，时序数据的文件存储形式主要按照时间窗口组织，按照不同时间段将数据快速写入同一个文件，每个文件分为若干 Block 块，保证时序数据高维度组织及高效存储，以此提高数据在文件中的存储及查询效率。

2.2 存储设计方法

对于时序数据存储，可以分为单节点存储及多节

点分布式存储两种类型，由于时序数据库需要融合海量数据读写的高性能要求，整个数据存储过程都需要快速存储，否则容易造成上游数据采集端发生拥塞现象。因此，基于现有时序数据的特性要求，单节点存储方式难以解决这个问题，对于多节点分布式存储而言，需要考虑整个时序数据库的数据分片存储及设计问题。如图5所示，时序数据分片存储策略的设计核心主要包括两方面：一是依据不同嵌入式设备时序数据的特性，确定数据分片存储的关键字段；二是时序数据分片存储的动态时间分区设计，确保数据快速分片到不同的时区。

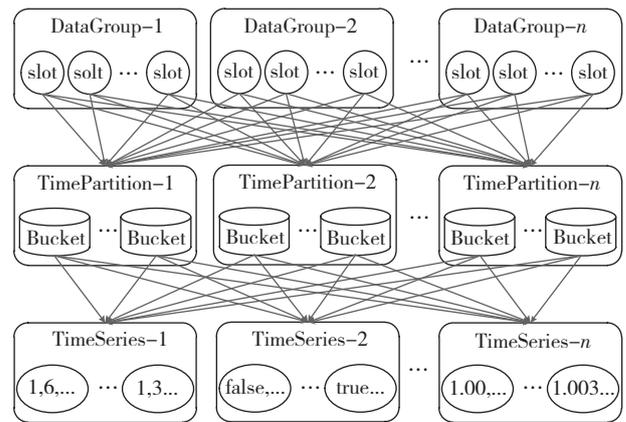


图5 时序数据分片存储策略

(1) 数据分片存储

嵌入式工业时序数据的发展特性及结构模型直接决定数据的分组策略，以设备属性和时间窗口为基准维度，通过将不同嵌入式设备及标签信息等数据主体属性进行逻辑划分，可形成数据组映射关系表，即由每对 { [MainDeviceID, DataGroup], ... } 信息组成映

射关系表。通过数据组映射关系表能够实时定位设备的逻辑存储位置, 保证设备数据的可控性及高效读写能力。由于时序数据库本身的特征, 按照时间维度进行分片存在天然优势。通过时间序列周期性创建一个数据桶 Bucket, 每个数据桶 Bucket 都有起始和结束时间用于存储对应时间段内的数据。当系统容量不足且添加新的存储节点后, 在创建新的 Bucket 时也将优先使用空闲容量最多的新节点, 已有数据不需要移动, 解决了数据存储容量的问题。其中, Bucket 是一个虚拟逻辑单元, Bucket 能够根据用户配置创建多个分片, 通过哈希函数对数据 Key 进行计算分片, 根据数据量的增长进行动态扩展, 保证数据分布的均匀性。每个 Bucket 包含多个 Block 块, 将物理块按照逻辑形式抽象成不同的 Bucket, 提高数据检索效能。

(2) 动态时间分区

分片设计的核心关键在于解决数据分布均匀性进而优化查询效率的均衡问题。以嵌入式工业时序数据库结构特性为基础, 将时间关键字进行 hashslot 分片, 保证数据分布存储。而动态时间分区的核心策略在于应用设计阶段, 由于时序数据库的关键字为时间戳, 而嵌入式工业时序数据是以时间维度产生设备数据, 按照时间区间进行动态分区, 以年/月/日形式结构对时序数据进行动态时间分区存储, 保证数据的高可用性。同一时间范围分区按照其时间形式结构能够对应到各个 Bucket 单元, 即数据写入过程按照其特性结构进行组织, 形成特有的序列化 Key, Key 能够决定其物理块 Block, 而列表 List 中时间序列能够确保对应的时间分区和 Bucket, 进而提高数据检索查询性能。

3 应用验证

3.1 应用环境

基于时序数据库技术的设计思路, 依托嵌入式系统环境对现有自研时序数据库开展了应用设计的功能模块开发工作, 并对各项常规系统指标进行试验验证。如表 1 所示, 本文通过现有公开测试的车辆设备数据集, 融合现有应用环境, 针对所提的嵌入式工业时序数据库存储策略的应用系统进行了有效性验证。

表 1 应用环境指标

| 环境指标 | 说明 |
|------|--|
| 硬件环境 | FT-D2000 开发板; CPU 4 核, 内存 16 GB, 硬盘 1 TB |
| 软件环境 | Linux Kernel: 4.19.111 5-r76-2.01.00.00-linux |
| 数据集 | 1 000 辆卡车设备数据; 视觉设备/readings, 诊断设备/diagnostics |

3.2 应用验证

分析现有嵌入式工业时序数据结构特性及高效存储管理代价问题, 根据工业时序数据的数据主体、时间点及采样值三部分构建数据结构模型, 初步开发实现了一套基于 C/S 模式的嵌入式工业时序数据库存储管理的应用系统。系统采用 QT5.13 设计开发, 运行稳定, 用户应用体验效果较好。应用系统通过对车辆数据进行大规模批量写入, 并按照不同的规约条件实时查询对应的车辆设备数据信息, 如图 6 所示。



图 6 嵌入式时序数据信息录入

通过对不同车辆数据信息的大批量写入, 按照不同的聚合函数及约束条件, 对特定时间范围内不同车主的视觉设备数据进行聚合分析, 得出基于时间维度的各个车主的实时位置信息, 如图 7 所示。由此体现出嵌入式工业时序数据应用的近实时存储及高效检索分析能力, 进而提高工业设备数据的高效管理及数据价值。

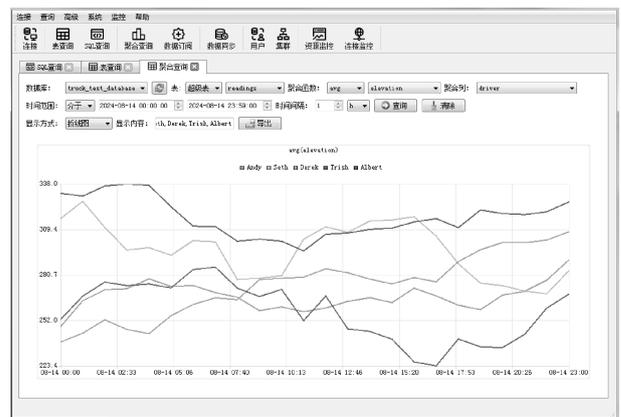


图 7 嵌入式时序数据查询分析

针对不同客户端用户的登录信息进行实时监控, 实现基于 C/S 架构的嵌入式工业时序数据库存储系统的应用特性, 如图 8 所示, 系统实现了实时管理客户端用户, 确保应用系统的稳定可靠性。

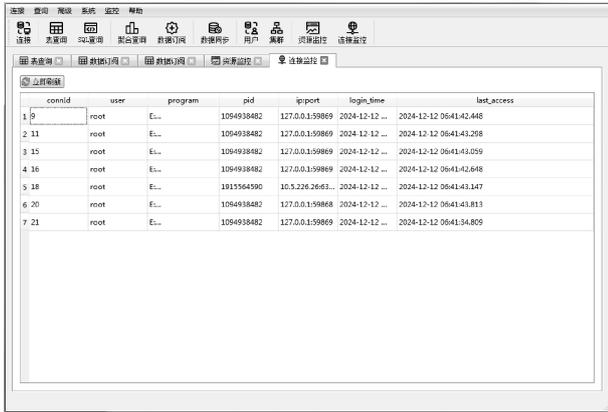


图8 嵌入式时序数据连接监控

基于嵌入式工业时序数据存储的系统应用有效性验证,通过公开的卡车车辆数据集进行不同数据量及批次大小的写入性能测试,如图9所示,横坐标表示数据写入批量大小,纵坐标表示数据写入耗时。实验结果表明,数据量与写入性能的耗时呈现正比趋势,同时从两条曲线均能够看出,在数据写入的批量大小由1 000逐步递增到10 000的过程中,性能耗时先降后增,表明批量大小 BatchSize 达到8 000时,写入性能效果最佳。

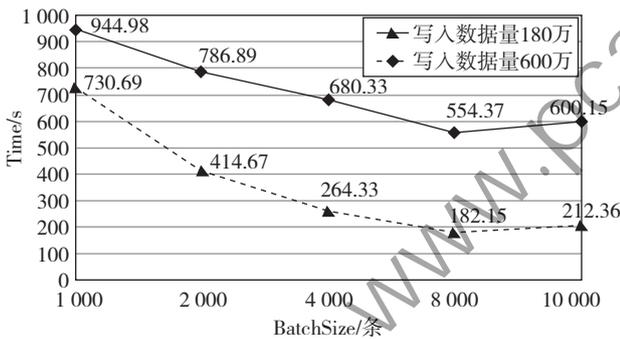


图9 嵌入式时序数据写入性能

4 结论

本文重点围绕嵌入式工业时序数据存储应用的发展现状及领域问题,归纳出工业传感设备的数据共有特性,构建了嵌入式工业时序数据的结构模型,融合时序数据库技术,设计出一种基于C/S模式的时序数据存储体系架构,进行了系统设计的有效性验证和应

用开发工作。实验验证表明,融合时序数据库技术所提出的存储体系架构及应用设计可以在嵌入式工业时序数据管理方面凸显效果,能够带来一定的应用研究价值。但综合当前嵌入式工业时序领域的硬性资源指标要求及应用发展,海量高频时序数据的高效压缩存储及实时响应是具有持续价值的研究方向,未来仍需进一步深入探究。

参考文献

- [1] 孔祥芬,蔡峻青,张利寒,等. 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): 8-23.
- [2] 郑博,王煜彤,燕钰,等. 时间序列数据库管理:技术、系统与展望 [J]. 工业技术创新, 2022, 9 (4): 12-21.
- [3] 赵清洲,刘华磊,乔文. 数据技术在军用航空装备及体系中的应用研究 [J]. 航空维修与工程, 2022, (6): 41-44.
- [4] 王妙琼,魏凯,姜春宇. 工业互联网中时序数据处理面临的新挑战 [J]. 信息通信技术与政策, 2019 (5): 4-9.
- [5] 刘帅,乔颖,罗雄飞,等. 时序数据库关键技术综述 [J]. 计算机研究与发展, 2024, 61 (3): 614-638.
- [6] 曲昌琦,毛利民,周岩,等. 面向复杂装备PHM数据体系架构的研究 [J]. 计算机测量与控制, 2021, 29 (4): 1-4, 9.
- [7] 隋佳欢,吴迪,朱盈,等. 基于机载时序数据库的数据管理方案 [J]. 网络安全与数据治理, 2023, 42 (S1): 216-220.
- [8] 李杰. 直升机试飞数据管理现状及远期规划 [J]. 电子世界, 2020 (7): 31-32.
- [9] 曹雪君,姚娜,王天秀,等. 时序数据库在航天型号试验数据管理中的应用 [J]. 航天控制, 2024, 42 (1): 51-57.
- [10] 王红涛,王志超,陈峰,等. 基于时序数据库的工业大数据应用研究 [J]. 重型机械, 2020 (4): 17-21.

(收稿日期: 2025-09-25)

作者简介:

周俊鹏 (1992-), 通信作者, 男, 硕士, 工程师, 主要研究方向: 大数据分析、数据管理技术. E-mail: 583131026@qq.com.

王卓 (1984-), 男, 博士, 高级工程师, 主要研究方向: 大数据应用、数据库技术.

康永乐 (1994-), 男, 硕士, 工程师, 主要研究方向: 数据可视化、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