

DOI:10.20033/j.1003-7241.(2026)01-0001-05

大数据存储技术的研究综述

王伟晨¹, 王瑾婷²

(1. 陕西工业职业技术学院, 陕西 咸阳 712000; 2. 咸阳市气象局, 陕西 咸阳 712000)

摘要: 大数据存储是大数据处理和分析的基础, 调研分析国内外主流的数据存储技术, 比较不同存储方式的特点和应用场景, 并针对不同需求提出参考建议, 有利于从不同角度把握数据存储技术的发展。对分布式文件系统的文件分割功能、适用场景及优缺点的总结, 并重点分析总结 NoSql 数据库的 4 种数据存储模型的原理及适用场景, 以及数据库一体机的发展及特征进行详细的调研与归纳, 同时概述大规模并行处理系统 (massively parallel processing, MPP) 架构新型数据存储技术和多模型数据库在大数据存储技术中的应用。最后对未来存储技术的可能研究趋势进行展望。研究表明, 选择合适的存储技术需要综合考虑数据规模、性能需求、成本等因素。

关键词: 大数据存储; NoSql; 分布式文件系统; 数据库一体机

中图分类号: TP311

文献标志码: A

文章编号: 1003-7241(2026)01-0001-05

Research survey on big data storage technology

WANG Weichen¹, WANG Jinting²

(1. Shaanxi Polytechnic Institute, Xianyang 712000, Shaanxi, China;

2. Xianyang Meteorological Service, Xianyang 712000, Shaanxi, China)

Abstract: Big data storage is the foundation of big data processing and analysis, this paper respectively investigates and analyzes the domestic and international mainstream processing technology of data storage, in view of the current popular big data storage technology, compares the characteristics and application scenarios of different storage methods, and puts forward reference suggestions for different application needs, which is conducive to grasping the development of data storage technology from different angles. This paper summarizes file segmentation, appropriate scenarios and merits and faults of distributed file system, and mainly analyzes and summarizes the principle and scenarios of four data storage models of NoSql database. Furthermore, it investigates and concludes the developments and features of database appliance minutely. At the same time, it outlines the applications of the massively parallel processing (MPP) architecture, emerging data storage technology, and the application of Multi-Model Database in big data storage technology. At last, the possible research trends of future storage technologies are prospected, research indicates, choosing the right storage technology requires comprehensive consideration of factors such as data size, performance requirements, and cost.

Keywords: big data storage; NoSql; distributed file system; database appliance

自 20 世纪 90 年代以来, 随着应用程序规模的扩张, 网络服务从单一的形式演变成多媒体的形式, 导致数据的形式和结构产生了多样化, 全球数据量正以每年 40% 的速度增长, 国际数据公司 IDC 给出数据规模的预测今后将以每两年翻一倍的速度增长^[1]。大数据研究的先驱——美国咨询公司麦肯锡给出的大数据定义是大数据指的是大小超出常规的数据库工具获取、存储、管理和分析能力的数据集^[2]。由于大数据具有规模性、多样性、高速性、价值性等特征^[2], 对传统的数据存储技术带来了挑战, 传统的数据存储技术难以满足大数据应用对海量数据存储、高并发访问、高可靠性和可扩展性的需求, 无法及时

地完成数据处理任务, 分布式存储技术应运而生。

然而, 目前针对不同应用场景下如何选择合适的存储技术的研究还较为缺乏, 因此, 本文对国内外主流的数据存储技术进行调研分析, 旨在为大数据存储技术的选择和应用提供参考。本文针对大数据存储面临的程序规模迅速扩张、数据量飞速增长、数据类型多样化等问题, 对新型的数据存储技术进行了研究总结。首先对分布式文件系统国内外研究现状进行了调研分析, 针对当前主流的分布式文件系统如 GFS、HDFS、GlusterFS、GridFS、TFS、Lustre、FastDFS 的各类特征进行了分析比较; 另外通过研究 NoSql 数据库的四大数据模型, 分别对键值模型、列式模

收稿日期: 2024-08-06; 录用日期: 2024-10-14

基金项目: 国家自然科学基金项目 (61462070); 陕西省自然科学基金研究计划项目 (2023-JC-QN-0311); 陕西工业职业技术学院科研基金资助项目 (2024YKYB-026); 咸阳市 2024 年度重点研发计划项目 (L2024-ZDYF-ZDYF-NY-0027)

作者简介: 王伟晨 (1992—), 男, 硕士, 讲师, 研究方向: 大数据存储技术和移动计算。

通信作者: 王瑾婷 (1992—), 女, 硕士, 工程师, 研究方向: 强对流天气预报和数值模拟。

引用本文: 王伟晨, 王瑾婷. 大数据存储技术的研究综述 [J]. 自动化技术与应用, 2026, 45(1): 1-5. (WANG Weichen, WANG Jinting. Research survey on big data storage technology [J]. Techniques of Automation and Applications, 2026, 45(1): 1-5.)

型、文档模型、图形模型的存储技术进行了对比研究;同时针对当前热门的大数据库一体机的适用性和架构特征进行了研究;接着对 MPP 架构的新型数据库集群的定义和适用场景进行了简要的介绍;最后对多模型数据库系统在大数据存储中的应用进行了探讨,并对未来存储技术的可能研究趋势进行了展望。

1 大数据存储技术

本文从 4 个方面对大数据存储技术进行了分析研究,分别是分布式文件系统、NoSql 数据库、MPP 架构新型数据存储技术和大数据库一体机。并给出了适用环境的参考建议,有利于从不同角度把握数据存储技术的发展。

1.1 分布式文件系统

文件系统是上层应用程序的基础,然而随着网络应用的不断发展,数据也随之大量增长,因此大数据存储技术已经成为企业和研究机构主要任务。由于文件系统容量

的拓展受到限制,已经很难使用传统的存储技术来解决大数据的存储问题。因此使用分布式文件系统将系统的负载转移到多个节点上已经成为目前大数据存储的共识。分布式文件系统能够提供聚合的存储容量和 I/O 带宽,对于文件系统规模进行扩展带来了极大的便利^[3]。

判断一个分布式文件系统是否成功,主要取决以下 3 个因素:数据的存储方式;数据的读取速率;数据的安全机制。然而分布式文件系统也存在一些有待完善的地方,例如为大文件设计的 GFS 和 HDFS 就无法存储大量的小文件,因为小文件访问频率相对较高,导致对磁盘的访问次数也增多,进而降低 I/O 的性能;并且大规模小文件也会造成元数据的负担,它会影响元数据服务器的管理和恢复性能,拖累集群的整体性能;而且小文件的数量较多容易产生文件碎片,进而造成磁盘空间的浪费^[3]。

表 1 对当前几种主流的分布式文件系统进行了概括的总结和对比,例如 GFS, HDFS, TFS, Lustre 等等。

表 1 常用的分布式文件系统对比^[4-8]

Tab. 1 Comparison of commonly used distributed file systems^[4-8]

名称	文件分割	系统备份	优点	缺点	应用场景
GFS	文件被分割成固定大小的块	每一个块被复制到多个块服务器上,默认会保存 3 个副本	硬件失败不被视为异常。更新通过添加新数据而不是对现有的数据进行改变。	不适用于对小文件的存储,过多的小文件会降低性能	大型分布式海量数据集,数据量一般 4~40 GB 之间
HDFS	文件被分割成固定大小的 Block,每个块有副本备份	支持数据复制,将文件的多个副本存放在不同的节点	扩展能力强,成本低,具有可靠性和高容错性	不能用于对数据访问要求低延迟的场景,不能存储大规模小文件	大小在 GB 级别到 TB 级大规模的超大数据集
GlusterFS	不支持文件分割	支持数据复制,并提供全局的命名空间	支持 CIFS, NFS 和本机使用 GlusterFS 客户端;多文件系统可以被部署在虚拟的分布式文件系统上;所有节点可用来获取数据	只能在一个服务器上管理系统,没有冗余;不能在中途添加新节点;没有安全策略	支持 PB 级别的海量大文件存储
TFS	不支持文件分割,大量的小文件会合并成一个大文件	以块文件的形式存放数据文件,并且存放多份保证数据安全	操作简单,完全平滑扩容与负载均衡;支持线性扩展。可以轻松地扩展到 PB 级	在并发过高,文件大小超过 5 MB, TFS 出现 bug。很少支持大文件的存储	针对海量非结构化小数据,淘宝网站海量图片的存储与处理
Grid FS	支持将一个文件分割成多个较小的文档文件	数据和元数据保存在 MongoDB 里,通过复制来解决故障切换,还可以用来做读扩展,热备份或者作为离线批处理的数据源	基于对象存储结构模式,减少运行时文件访问的开销;调节访问机制可以提供应用集群需求最快的 I/O 性能;确保负载均衡	读取文件相对较慢;修改大文档时无法对所有的文件块加锁,并且只有先删除旧文档才能重新保存文档。	较适合存放不常修改的大规模文件
Lustre	将数据分割到固定数量的对象中去,每个对象包含若干个数据块	提供两个备份工具,一个用来扫描文件系统,一个用来打包备份并加压恢复	提供数据共享和并行的能力;具有极强的扩展性,可以对元数据和目标数据提供故障切换技术,实现高可靠性的访问;分布式的管理机制可以实现并发控制;提供多网络协议访问	节点间故障切换要依赖于第三方的心跳技术;只有两个元数据管理节点,系统规模如果达到一定程度,管理节点会过载;只能在 Linux 上部署 lustre 的内核,具有一定的局限性	支持 PB 级别的海量大文件存储,适用于大型计算机集群或者超级电脑
Ceph	采用 RAID0 模式横跨多个硬盘,把连续的数据分散到多个磁盘上存取,以适应负载均衡	支持数据备份,存在多个元数据服务器	将数据和元数据分开存储,动态分布元数据管理,具有可靠的自动分布式对象存储。	技术尚不成熟,还不能适用于生产环境。	支持 PB 级别的海量大文件存储
Fast DFS	不支持文件分割	采用分组存储方式,同组内多台服务器之间是互相备份的关系	服务端只有跟踪器和存储节点两个角色,具有轻量级特点;采用分组存储方式,灵活,可控性强;的每个节点都是主节点,具有对等结构,可根据服务器的压力随时改变跟踪器的数量	存储文件时没有进行分块存储,不适合分布式计算场景;存储能力受到单个存储服务器的限制	适合以文件为载体高访问量服务,如相册网站、视频网站等,海量小规模文件,建议范围 4 kB~500 MB

1.2 NoSql 数据库

随着企业用户数据量的迅速增长,以及用户对服务水平的需求不断提高,传统的关系数据库具有很大的局限性;传统的数据库将以结构化记录为基础的平面文件的形式存储所有的应用数据,其导致阻抗应用程序与数据库之间的不匹配;这种情况是应用程序被编码在非声明性的语言中,其结构与这些数据库是完全不同的^[9]。以往为了提高系统的性能,会纵向扩充组件与资源;然而,自从存储与应用程序不再分开之后,每一次对资源的扩充都是服务中断和应用程序重置。

当前所产生的大部分数据都是异构数据,大量非结构化、结构化数据的存在,对事先确定完善、统一的数据关系模型带来了挑战,同时关系数据库的横向扩展能力相对较低。多数关系数据库也不支持大规模分布式存储;难以满足高并发,大数据规模的实时性要求。因此底层存储技术不仅应该灵活地允许数据以它的自然形式进行存储,还要满足前沿的需求。

相较关系型数据库, NoSql 数据库存储系统支持海量数据的存储和动态管理,它避免了不必要的复杂性,具有较高的吞吐量,可以很好地进行横向扩展,而且高度容错可以存储结构化,半结构化和非结构化数据避免了对对象—关系映射。NoSql 数据库的设计思想是提取出关系型数据库的索引机制,并融合分布式存储,把 SQL 体系中那些在某些问题上并不需要的东西全部删除,因此实现了较为优秀的效率、扩展性与灵活性^[10]。

NoSql 数据库主要分为:键值存储,列式存储,文档存储,图形存储。

键值数据库的目的是支持简单的查询操作,将复杂的操作留给应用层。数据集将键映射到一个或者一组值,即 key 键是查找每条数据地址唯一的關鍵字,这也意味着它是不可或缺的^[11];value 是该数据实际存储的内容。键值

存储提供了一个装有键值对的哈希表在一个分布式集群上的多台远程服务器上,用来实现 key 键到值的映射。基于 key 键的 hash 值直接定位到数据所在的地址,实现快速高并发查询,也支持大数据量的操作。键值存储也分为 key-value 型、key-document 型和 key-column 型^[12]。key-column 型是 key-value 型键值对的典型扩充,因其简单以及具有灵活的可扩展性因此也是目前主流的数据模型。

广义的列式存储是以列族取代行,关系存储数据库的思想是将在磁盘上用一行存储所有表项内容,即关联同一个特定的行 id 的所有列条目将被存储在一起^[11]。由于银行或金融机构需要维护大量的相关记录,并不能保证所有值总是以连续的方式存储。在列数据的数据库中,表的一整列存储在一起,映射到一个键。由于所有列条目都有索引,因此可以仅在表的一部分进行搜索。一列也可以有层次结构的嵌套列,其中有一个超级列。这提供简单的查询和快速访问,同时避免寻找一条记录的单独键这种不必要的开销。

文档数据库是另一种关系数据库,用于存储半结构化数据,存在于 XML(可扩展标记语言),JSON(JavaScript 对象表示法),或其他类似的格式^[11]。一个文档可以比作关系数据库的一行,其中包含所有文档的相关信息。一组包括多个文档,每个文档可以有不同的模式和不同数据存储数量及类型。存储文本信息特别优化。由于相关的数据集中存储,节省 SQL JOIN 操作的开销。虽然数据库是无模式设计,但存储记录半结构化并存在层次结构。

图数据库最适合遍历和搜索应用程序,例如在 LinkedIn 上寻找相关链接,在 Facebook 上查找朋友等等。它更重视数据项之间的关系而非数据本身,他们高度优化快速遍历并高效利用图的算法,例如最短路径优先为了找到信息之间的关联等等。

表 2 将 NoSql 数据库主要存储类型进行了分析与总结。

表 2 NoSql 数据库的 4 种存储模型^[9-12]

Tab. 2 Four storage models of NoSql databases^[9-12]

数据库类型	优点	缺点	数据模型	应用场景	实例
键值存储	具有极高并发读写性能;数据按 key 值进行索引、切分;查找迅速,数据模型简单	数据无结构化,不支持复杂的逻辑数据操作	key 与 value 间建立的键值映射	内容缓存,主要用于日志系统	Dynamo、Redis、Voldemort
列式存储	查找迅速、扩展性好、节省大量的 I/O 操作,容易进行分布式扩展	功能相对局限	列式存储,同一列数据存储在上一页	分布式文件系统	Bigtable、HBase、Cassandra
文档存储	不需要提前定义数据结构,采用特定格式的文档替代元组作为数据存储的单位	查询效率不高,缺乏统一的查询语法	value 指向结构化数据	Web 应用	MongoDB、XML Database
图形存储	利用图论及相关算法提高存储性能,管理和操作数据	功能相对局限	图结构	社交网络、关系图谱	Neo4j、GraphDB InfoGrdi

1.3 数据库一体机

近年来,数据库一体机作为一种面向海量数据处理和存储的一体化解决方案,逐渐成为研究热点。表 3 将主流数据库一体机进行了分析与总结。

当前主流数据库一体机均使用了 Shared-Nothing 与 MPP 架构,实现数据并行加载、存储、计算;同时,主流厂商

采用具有各自特色的技术手段避免了数据库一体机的 I/O 瓶颈,对数据处理分析技术进行了优化。数据处理速度比传统解决方案有大幅提升,性能能够随着存储节点服务器的增加快速增长,具有较好的横向扩展能力,可以保证低延迟、高吞吐量的性能和业务的连续性^[14]。能够根据用户自身的具体需求来选择不同配置的产品,实现个性化按需定制。

表3 主流数据库一体机分析比较^[13-14]

Tab. 3 Analysis and comparison of mainstream database appliance systems

数据库类型	数据处理架构	响应能力	数据分布方式	适用场景
EMC greenplum	采用 Share-Nothing 的并行处理架构	数据加载速率 10 T/h	Hash 方式或者 Random 方式	主要集中在需要高性能数据分析、决策支持以及数据挖掘的企业级环境中
IBM Netezza 1000-12	采用 Share-Nothing 的并行处理架构	可与 CDC 整合实现实时数据加载,其他加载近实时	哈希分区或随机分区	需要处理大量数据、进行深度分析和支持 AI/ML 工作负载的场景,支持从传统数仓向现代化数仓迁移
Oracle Exadata	采用 Share-Disk+ Share-Nothing 的混合结构	支持在线 OLAP	采用 ASM 存储管理,实现存储镜像和条带化	用于处理大规模的高性能数据处理、复杂的数据仓库操作等场景
浪潮 inData	采用 Share-Nothing 的并行处理架构	采用 RDMA 协议,低延迟网络传输	全闪存配置,支持混合存储模式	替换传统小型机与高端存储,广泛运用于政府、能源、通信、金融等各行各业

1.4 MPP 架构的新型数据库集群

大规模并行处理系统是一种进行系统资源扩展的方法,主要是并行处理,意味着单个计算机有多个网络处理器^[15]。横行扩展是 MPP 架构数据库的主要设计目标;它由多个 SMP 服务器通过固定节点的互联网进行链接,互相协作完成共同的任务,从用户层面来看是一个服务器系统,支持严格的数据关系模型。最大的特点就是每个节点只能访问自身本地资源,资源不共享。

使用 MPP 架构的数据库集群能够存储 PB 级别的海量结构化数据,基于 Shared Nothing 架构,通过列存储、粗粒度索引等大数据处理技术,并结合其高效能的分布式计算模式,完成对分析类存储应用的技术支撑,运行环境多为低成本 PC,具有高性能和高扩展性的优点。MPP 架构的数据库集群在处理大数据分析和查询场景时表现出色,可以提升数据处理的性能、提升数据处理量、提升海量数据处理的效率并降低处理每一个 TB 的整体成本,能够提供更高的性能和可扩展性,因此在企业新一代的数据仓库和结构化数据分析类的应用领域中获得非常广泛的应用;而传统数据库集群则更适合一般的业务处理和数据存储需求。

虽然 MPP 架构具有并行处理、分布式存储等优势,但在实际应用中存在数据孤岛、对节点配置同质性要求高以及处理异构机器时性能下降等局限性,这些局限性在一定程度上限制了 MPP 架构的广泛应用和优化。

1.5 多模型数据库

多模型数据库系统是一种高度灵活、可拓展、能够同时支持多种数据模型的数据库管理系统。这些数据模型中典型的代表有关系型数据模型、文档数据模型、图形数据模型、键值数据库模型、列族数据库模型等。其关键特点在于其能够有效地存储、管理和检索多种数据类型,以满足不同应用场景下的复杂数据需求,而无须引入多个独立的数据库系统,具有较强的通用性和扩展性^[16]。

正是由于多模型数据库系统支持多种数据模型这一核心特点,使得它可以轻松处理大数据中的结构化数据、半结构化数据、非结构化数据,为种类繁多的大数据提供了一个统一的存储和管理平台,有助于降低数据处理的复杂性。

大数据处理过程中需要将多源数据整合在一起,以便于分析和挖掘,而多模型数据库系统可以将不同数据模型的数据整合在同一个数据库中,然后使用查询语言进行跨模型的复杂分析,从而为数据分析提供便利。另外大数据处理过程中既包含了实时处理又包含批处理,多模型数据库系统可以在同一个平台上实现这两种处理方式,简化了数据处理的架构和管理。大数据存储中数据安全和隐私保护的重要性不言而喻,而多模型数据库系统通常提供强大的安全性控制和权限管理功能,以保护数据免受未知的访问和攻击,这对于处理包含敏感信息的大数据至关重要^[17]。

综上,多模型数据库系统与大数据存储之间有着密切的关联,通过将不同数据模型的数据整合在一个统一的数据库中,多模型数据库系统可以简化大数据处理的复杂性,提高数据处理的效率和拓展性,为数据驱动的决策和创新提供有力支持。因此多模型数据库系统的应用场景非常广泛,涵盖医疗、金融、社交媒体、电子商务平台、社交媒体分析以及物联网等领域。这些应用场景展示了多模型数据库系统的灵活性和效率,能够有效地处理和整合不同类型的数据,满足各种领域的复杂数据处理需求。

2 结论

本文调研分析了国内外主流的数据存储技术,比较了不同存储方式的特点和应用场景,并针对不同应用需求提出了参考建议,为大数据存储技术的选择和应用提供了参考价值。大数据存储技术的发展趋势受到多种因素的推动,如数据量的快速增长、前沿技术的不断创新以及企业对数据驱动决策的需求等。未来大数据存储技术的研究方向包括:研究更高效的存储算法、设计更轻量级的存储模型、探索更广泛的应用场景等。

类似于 NoSql 数据库以及分布式文件系统的新型存储技术,无论是从容错性、可扩展性、数据的移动性上明显优于传统的存储方式,适用于数据持久存储和模式自由的海量数据存储管理。但是从处理数据的实时性能上,与关系型数据库还有一定差距;因此未来将关系型数据库和分布式并行处理系统进行结合的混合存储解决方案,可使数据在存储效率以及处理分析速度上都得到很大的提高。

分布式存储系统由于其具有高容错性和拓展性,将会越来越受到企业的青睐,未来分布式存储系统将进一步优化,支持更高的性能和更低的延迟。在后续的发展中,新型数据库将逐步与 Hadoop 生态系统或者 Spark 生态系统结合使用^[18],为应用提供丰富的 SQL 和事务支持。所以未来存储方式也应该会朝着新一代的 MPP 并行数据库集群+Hadoop/Spark 集群方向发展。另外,数据库一体机和多模型数据库也势必成为研发的热门技术,借助数据库一体机,可以满足 OLTP、OLAP 与混合型工作应用,用户可以进行大量数据库的整合,提升系统运行效率,加速创新发展与快速响应业务对 IT 的要求;而多模型数据库系统的灵活性、模型多样性、分布式架构和数据整合等特性正使其成为处理大数据的理想选择,在大数据时代的背景下有望在各个领域发挥重要作用。除了这些主流存储技术之外,云存储具有的弹性、可扩展性和按需付费的特性将使企业能够更高效地管理海量数据;未来,混合云和多云架构应用也许会更加广泛。对象存储由于其可扩展性和成本效益,能够高效地存储非结构化数据,未来会朝着成为大数据存储主要方式的方向发展。边缘计算的兴起带动了边缘存储的发展,随着物联网设备的大量普及,越来越多的数据在网络边缘生成,而边缘存储能够在数据源附近存储和处理数据,减少延迟并提高实时分析的能力,因此也会成为未来大数据存储技术的研究热点之一。数据量的爆炸式增长除了对存储技术的迫切需求之外,高效的数据压缩和去重技术也越来越重要,未来优化存储效率和降低存储成本肯定会成为大数据存储算法的主要研究方向。

这些趋势表明,大数据存储技术将朝着更加智能、高效和灵活的方向发展,以满足未来企业和社会对数据存储的更高要求。

参考文献

[1]肖英,赵林洁,张宇,等.支持快速索引的高效大数据存储结构

- [J]. 计算机应用与软件, 2024, 41(3):28-33.
- [2]涂新莉,刘波,林伟伟. 大数据研究综述[J]. 计算机应用研究, 2014, 31(6):1612-1616.
- [3]梁懿,刘迪,陈又咏,等. 国产化环境下的海量小文件数据分布式存储技术[J]. 计算技术与自动化, 2023, 42(3):141-146.
- [4]RAM K, SINGLA M. Hadoop distributed file system hybrid encryption algorithm for big data security[J]. Semantic Scholar, 2023(36):16-27.
- [5]程稳,李焱,曾令仿,等. 面向 Lustre 集群存储的应用日志分析及系统自动优化框架[J]. 计算机工程与科学, 2022, 44(4):594-604.
- [6]周添锦. 基于 FastDFS 安全文件系统设计与实现[D]. 西安:西安电子科技大学, 2019:65-58.
- [7]郑通,郭卫斌,范贵生. HDFS 中海量小文件合并与预取优化方法的研究[J]. 计算机科学, 2017, 44(增刊2):516-519, 541.
- [8]李文武,张建锋,王景林. 基于 EHDfs 的海量小文件存储与检索方法[J]. 计算机工程与设计, 2022, 43(2):376-383.
- [9]刘志强. 基于非关系型数据库的大数据分布式存储技术[J]. 科技创新与应用, 2023, 13(22):27-30.
- [10]李立猛. 关系型数据库与 NoSql 数据库的应用场景[J]. 电子技术与软件工程, 2022(16):184-187.
- [11]LANDUYT D V, WIJSHOFF V, JOOSEN W. A study of NoSql query injection in Neo4j[J]. Computers & Security, 2024, 137(2):48-56.
- [12]PEDONE F. MoSQL: a relational database using NoSql [J]. Computers & Security, 2022(125):26-47.
- [13]张东,亓开元,吴楠,等. 云海大数据一体机体系结构和关键技术[J]. 计算机研究与发展, 2016, 53(2):374-389.
- [14]陈国良,汤晓宇,尤帅,等. 基于国产处理器的智能大数据一体机架构及应用研究[J]. 南京邮电大学学报(自然科学版), 2024, 44(4):1-16.
- [15]NIKOLA Y, GOLO V, LARS R. Big data normalization for massively parallel processing databases [J]. Computer Standards & Interfaces, 2017 (54):86-93.
- [16]杨颀,韩路,许勤,等. 大型商业银行基于 Hadoop 分布式数据仓库建设初探[J]. 计算机应用与软件, 2017, 34(8):72-75.
- [17]JIA C, SUN S, GENG Y. Financial risk early warning model based on intelligent data monitoring [C]. 2022 International Conference on Big Data, Information and Computer Network (BDICN) Sanya, Jan, 2022:161-164.
- [18]周润松. 2022 中国优秀大数据产品与解决方案测评发布[J]. 软件和集成电路, 2022, 26(8):30-36.