 DOI: https://doi.org/10.12349/iser.v6i3.5180

A Brief Discussion on Centralized Storage and Distributed Storage

Peng Wang

Information Center of Northwest Survey and Design Institute of China Electric Power Construction Group Co., Ltd., Xi'an, Shaanxi, 710065, China

Abstract

With the rapid growth of data scale and the diversification of application scenarios, the selection of storage systems has become a top priority for enterprise IT infrastructure construction. This article systematically analyzes the technical principles, performance characteristics, and applicable scenarios of two mainstream architectures, centralized storage and distributed storage. The focus is on comparing the performance of the two storage systems in terms of architecture, performance, access interfaces, data security, and operation and maintenance costs. The comparative results show that centralized storage has the advantages of simple management and stable performance in low concurrency and small-scale data scenarios, while distributed storage exhibits better scalability and fault tolerance in large-scale data and high concurrency access scenarios. Based on the analysis and comparison results, this article proposes storage system selection recommendations for different application requirements.

Keywords

Centralized storage; Distributed storage; framework; High concurrency

浅谈集中式存储与分布式存储

王鹏

中国电建集团西北勘测设计研究院有限公司信息中心，中国 · 陕西 西安 710065

摘　要

随着数据规模的快速增长和应用场景的多样化，存储系统的选型已成为企业IT基础设施建设的头等大事。本文针对集中式存储与分布式存储两种主流架构，系统分析了其技术原理、性能特征及适用场景，重点比较了两种存储系统在架构、性能、访问接口、数据安全和运维成本方面的表现。对比结果表明，集中式存储在低并发、小规模数据场景下具有管理简单、性能稳定的优势，而分布式存储则在大规模数据、高并发访问场景下展现出更好的可扩展性和容错能力。基于分析对比结果，本文提出了针对不同应用需求的存储系统选型建议。

关键词

集中式存储；分布式存储；架构；高并发

1 引言

随着信息化、数字化技术的快速发展，数据存储已成为社会不可或缺的重要基础设施。从个人照片到企业数据库，从科学研究到政府档案，海量数据的存储和管理需求不断增长，对存储系统的性能、可靠性和可扩展性提出了前所未有的挑战。在此背景下，集中式存储和分布式存储作为两种主流的存储架构，各自展现出独特的优势和局限性。

本文旨在探讨集中式存储与分布式存储的技术特征、性能表现、访问接口、数据安全及运维成本，系统比较两种存储架构的优劣，为不同应用场景下的存储系统选型提供选择依据。

【作者简介】王鹏（1980–），男，中国河北衡水人，本科，正高级工程师，从事网络基础设施建设与运维、网络安全防护、私有云平台设施建设与运维研究。

2 存储技术概述

2.1 集中式存储

集中式存储是指将数据集中存储在一个或一组紧密耦合的存储设备中，这些设备由一个存储系统统一管理，数据的存储、访问和管理都通过这个核心存储系统来完成。数据存储在专门的存储阵列中，服务器通过光纤通道等高速网络连接到这个存储阵列进行数据读写，数据的管理和维护相对集中，便于统一控制。通常采用“控制器 + 磁盘柜”的方式，中高端存储支持多个控制器，以保障高可用并提高性能。多控制器为紧耦合，通过 PCIE 总线或 Infiniband 网络互连，共享磁盘阵列，共享缓存。

2.2 分布式存储

分布式存储是一种将数据分散存储在多个独立的存储

设备（节点）上的存储方式。它通过网络将这些存储节点连接起来，每个节点都可以提供数据的读写服务，节点之间通过消息传递来通信和协调，对外提供统一的存储服务。通常采用“标准的 x86 服务器 + 存储软件”的方式，实现了存储的硬件和软件解耦，支持块存储、文件存储和对象存储功能。分布式存储系统中，数据被分成多个部分，每个部分存储在不同的节点上，以实现数据的分布式存储和管理 [1]。

3 对比分析

依据目前企业对集中式存储和分布式存储的使用现状，从存储系统架构、性能、访问接口、数据安全、运维成本五个维度进行对比分析。

3.1 架构

集中式存储采用经典的主从架构（Master-Slave），其核心由中央控制器、存储介质阵列和接入网络构成金字塔式拓扑。中央控制器作为唯一决策节点，统管元数据管理、IO 调度和存储资源分配，这种设计使得数据路径呈现严格的“决策—执行”分层结构。以传统 SAN 架构为例，前端主机通过 HBA 卡接入光纤交换机，由存储控制器将逻辑块地址映射到物理磁盘的固定扇区，整个过程依赖控制器的全局视图和同步锁机制保证原子性。这种架构的确定性优势在于硬件加速路径的极致优化（如 ASIC 芯片处理 SCSI 指令），但其扩展边界受制于控制器的 PCIe 总线带宽和 CPU 指令吞吐量。

分布式存储架构则构建于去中心化的对等网络（P2P）之上，采用元数据与数据分离的扁平化设计。以 Ceph 为代表的现代分布式系统，通过 CRUSH 算法将数据对象映射到物理节点的伪随机过程取代了集中式地址映射表，每个存储节点（OSD）既是数据载体也是决策参与者。这种架构的创新性体现在三个维度：其一，客户端可直接与存储节点通信，绕过了传统架构的中心调度瓶颈；其二，动态哈希环机制使数据分布随节点增减自动平衡，形成弹性拓扑；其三，基于 Paxos/Raft 的分布式共识协议替代了中心化锁机制，在保证弱一致性的同时实现高并发访问。

3.2 性能

3.2.1 时延

集中式存储的时延主要受限于单一控制节点的处理能力与物理链路长度，其数据访问路径通常呈现“客户端—中心节点—存储设备”的线性结构。在低并发场景下，这种架构可通过硬件加速（如全闪存阵列）和短距离网络（如本地 SAN）实现微秒级延迟，表现出较高的响应效率。但随着并发请求量增加，中心节点的 CPU 处理队列、磁盘 I/O 竞争和网络带宽争用等问题会导致时延显著上升，且呈现非线性增长趋势。

分布式存储系统通过多节点并行处理机制重构了时延模型，其数据访问路径通常呈现“客户端—数据节点—数据节点群”的三层架构。数据分片存储和负载均衡策略使得单个请求的处理时延可能增加，但系统整体吞吐量的线性扩展能力有效抑制了高并发下的时延波动。

综上所述，集中式存储更适合对延迟确定性要求严格的场景（如高频交易系统）；而分布式存储更适应互联网级突发流量场景。集中式存储以单点性能最优换取时延确定性，而分布式存储则通过分布式协调容忍时延波动来换取系统弹性。

3.2.2 IO 吞吐

集中式存储的 IO 吞吐能力严格受限于单节点的硬件性能上限，其理论峰值吞吐由存储介质、网络接口及控制器处理能力共同决定。在理想条件下，高端集中式存储阵列可实现每秒数十 GB 的顺序读写吞吐，但这一性能表现呈现明显的“天花板效应”——当业务负载超过单节点处理能力时，吞吐曲线会快速进入平台期，且无法通过简单扩容突破物理限制。

分布式存储系统通过多节点并行机制重构了吞吐模型，其核心优势在于吞吐能力的线性扩展特性，每个新增节点不仅贡献存储容量，同时提升系统整体 IO 吞吐，理论上 N 节点集群可提供接近 N 倍的吞吐增益。但这一特性需要特定条件支撑——当数据访问具备强局部性特征时，跨节点数据获取导致的网络开销可能使有效吞吐下降 30%~40%，此时需依赖智能缓存策略（如客户端缓存）进行优化。

综上所述，集中式存储在顺序大块 IO 场景（如 4K 视频编辑）中可发挥硬件极致性能，而分布式存储则在混合负载（尤其是小文件随机读写）场景表现更优。

3.2.3 数据重平衡

当存储介质磁盘出现故障损坏或进行修复时，存储软件均会进行数据重构以保证数据的多副本可靠性。数据重构周期与存储的硬件性能、重构数据量、前端业务压力有关 [2]。

集中式存储的重平衡操作高度依赖中心控制器的全局调度能力，其典型场景包括 RAID 组扩容、LUN 迁移或磁盘故障替换等场景。此类系统通常采用静态分区策略，重平衡过程需要管理员手动规划数据迁移路径，并通过全量数据拷贝实现存储空间的再均衡。这种集中式重平衡的优势在于策略统一、过程可控，但在大规模数据场景下易形成“重平衡风暴”，导致业务性能断崖式下跌。

分布式存储系统通过去中心化的数据分布算法重构了重平衡范式，其核心在于将重平衡逻辑下沉至数据节点层面。这种算法级优化使得分布式存储的重平衡过程具备“热迁移”特性，业务 IO 可在迁移过程中持续服务，且迁移粒度细化为对象级别（通常 MB 级）。但拜占庭节点可能导致重平衡震荡，跨地域集群的同步延迟可能引发数据分布倾斜。

综上所述，集中式存储的重平衡机制如同精密钟表，依靠人工调校实现确定性运维；而分布式存储的重平衡则类

似生态系统，通过分布式协商达成动态平衡。集中式存储在重平衡过程中牺牲弹性换取确定性，而分布式存储则通过容忍短期不均衡获得长期扩展性。

3.3 访问接口

集中式存储通常提供统一的访问接口，如基于文件系统（如 NFS、CIFS）或块设备（如 iSCSI、FC）的标准化协议，这种单一入口的设计简化了客户端配置和管理，但可能成为性能瓶颈和单点故障的根源。分布式存储系统往往采用多节点并行访问接口，如对象存储接口（如 S3、Swift）或分布式文件系统接口（如 HDFS、CephFS），这些接口支持并发访问和负载均衡，能够有效提升系统吞吐量，但同时也增加了接口复杂性和客户端实现难度。

在一致性模型方面，集中式存储通常提供强一致性保证，而分布式存储则可能根据应用场景提供不同级别的一致性（如最终一致性、会话一致性）。此外，分布式存储接口通常需要处理节点故障、网络分区等异常情况，因此往往内置了重试机制和故障转移功能，而集中式存储的接口设计则相对简单。

综上所述，集中式存储更适合对一致性和管理简便性要求高的场景；而分布式存储则更适用于需要高并发访问和大规模扩展的场景。

3.4 数据安全

集中式存储通常依赖于中心化的安全控制措施，如访问控制列表（ACL）、身份认证和加密存储等，这种集中化的安全管理简化了策略的实施和监控，但也使得系统容易成为攻击的单一目标，一旦中心节点被攻破，整个系统的数据安全将面临严重威胁。此外，集中式存储在数据备份和灾难恢复方面通常采用定期全量备份和增量备份相结合的方式，虽然操作相对简单，但在大规模数据环境下，备份窗口和恢复时间可能较长。

分布式存储通常采用去中心化的安全机制，如基于分布式账本技术的访问控制和加密存储，这些机制能够有效防止单点故障和内部威胁，通过数据分片和冗余存储（如副本或纠删码）来增强数据的安全性和可靠性，数据被分散存储在多个节点上，即使部分节点遭受攻击或发生故障 [3]，系统仍能通过剩余节点访问完整数据。此外，分布式存储在数据备份和灾难恢复方面通常利用其并行处理能力，可以实现快速的数据重建和恢复，显著缩短了恢复时间目标（RTO）。

综上所述，集中式存储在安全管理的简便性和一致性方面具有优势，适合对安全策略统一性要求较高的场景；而分布式存储则在抗攻击能力和灾难恢复方面表现更为出色，适合对数据可靠性和恢复速度要求较高的场景。

3.5 运维成本

集中式存储的运维相对简单直接，由于其架构的单一性，管理员只需关注少数几个核心节点的状态和性能指标，如存储容量、I/O 负载和网络带宽等。日常维护工作主要包括硬件监控、固件升级和备份管理，这些操作通常可以通过统一的管理界面完成，降低了运维人员的技术门槛。系统扩展需要停机扩容，故障排查范围相对集中但影响面广，且容易形成运维瓶颈。

分布式存储系统的运维则呈现出更高的复杂性和技术要求，运维人员需要同时监控大量节点的运行状态，包括存储节点、元数据节点和网络连接等，这对监控系统的自动化程度提出了更高要求。分布式存储通常采用软件定义的方式实现资源管理，支持在线扩容和动态负载均衡，这大大提升了系统的灵活性和资源利用率。在故障处理方面，分布式存储系统具有自愈能力，能够自动检测节点故障并进行数据重建，显著降低了运维干预的频率。

综上所述，集中式存储运维的前期投入相对较低，但随着系统规模扩大，其边际运维成本会显著增加；而分布式存储虽然初始部署复杂度较高，但其线性扩展特性和自动化运维能力使得大规模部署时的边际运维成本相对较低。

4 结论

本文通过对集中式存储与分布式存储的对比，探讨了两种存储在架构、性能、访问接口、数据安全、运维成本方面的差异与优劣。集中式存储采用单一控制节点管理所有存储资源，具有结构简单、IO 吞吐高、易于管理等优势，在小规模、低并发场景中占据主导地位；而分布式存储则通过将数据分散存储在多个节点上，利用并行处理和冗余机制，有效提升了系统的整体性能和可靠性，凭借其高可扩展性、容错能力和灵活的访问接口，在云计算、大数据等领域得到广泛应用。随着数据规模的持续增长和应用场景的不断扩展，分布式存储正逐渐成为主流趋势，但集中式存储在某些特定场景下仍不可替代。

5 结语

未来的存储系统发展将更加注重架构的融合与创新，集中式存储与分布式存储的界限可能逐渐模糊，混合架构和新型存储技术（如边缘计算存储等）将不断涌现。在实际应用中，存储系统的选型应充分考虑业务需求、应用场景及数据特性，选择最适合的解决方案，满足企业日益增长的数据存储与管理需求。

参考文献

[1] 么宇光.分布式存储在超融合架构方案中的应用[J].长江信息通信,2023,36(11):86-88.

[2] 金天骄.分布式存储取代集中式存储可行性分析[J].邮电设计技术,2019,(10):84-87.

[3] 王帅.基于分布式存储的现代档案管理体系的构建[J/OL].中文科技期刊数据库（文摘版）社会科学,2024(11)[2024-11-04]. https://www.cqvip.com/doc/journal/3471184898.