 DOI: https://doi.org/10.12349/iser.v7i3.9477

Performance Evaluation of Centralized Storage Network: A Comparative Study of RDMA over Converged Ethernet (RoCE) and Fibre Channel (FC)

Yongkang Qu[1] Yizhe Sun[1] Jian Zhang[1] Meng Li[2]

1. China Railway Information Technology Group Co., Ltd., Beijing, 100844, China
2. China Railway Information Engineering Group Co., Ltd., Beijing, 100044, China

Abstract

This study addresses supply chain security and technological monopoly issues arising from long-term reliance on foreign technologies in centralized storage for critical railway information infrastructure. To mitigate potential supply chain disruptions faced by core railway systems (e.g., the 12306 ticketing system and freight management platform) under traditional FC-SAN architectures, the research utilizes the China Railway Cloud environment to conduct an in-depth comparison between FC-based centralized storage networks and RDMA-powered indigenous high-performance lossless networks. The paper systematically elucidates the technical principles of both solutions and evaluates their key performance metrics through empirical studies, aiming to provide theoretical foundations and practical guidance for the large-scale deployment of domestically developed centralized storage networks in the railway sector.

Keywords

Storage Area Network; FC; RDMA; RoCE

集中式存储网络性能评估：RDMA over Converged Ethernet（RoCE）与光纤通道（FC）的对比研究

屈永康[1] 孙轶哲[1] 张健[1] 李孟[2]

1. 中国铁路信息科技集团有限公司，中国 · 北京 100844
2. 中铁信息工程集团有限公司，中国 · 北京 100044

摘 要

本文旨在解决铁路关键信息基础设施在集中式存储领域长期依赖国外技术所带来的供应链安全与技术垄断问题。针对当前铁路行业核心业务系统（如12306票务系统、货管平台）在传统FC-SAN架构下面临的潜在断供风险，本研究结合国铁云环境，深入对比了基于FC协议的集中式存储网络与基于RDMA技术的信创高性能无损网络。论文将系统阐述两种方案的技术原理，并通过实证研究综合对比其关键性能指标，旨在为国产化集中式存储网络在铁路行业的规模化推广应用提供理论依据与实践指南。

关键词

存储区域网络；FC；RDMA；RoCE

1 引言

集中式存储网络作为关键信息基础设施的核心组件，其性能与可靠性直接影响关键领域的业务连续性。长期以来，光纤通道（Fibre Channel, FC）技术凭借其低延迟、高可靠特性，在高端存储市场占据主导地位。

当前，我国铁路行业在存储区域网络（SAN）的建设中，光纤通道存储区域网络（FC-SAN）架构仍占据高比例。这种架构的核心设备与关键技术长期被国际厂商垄断，导致国内企业面临严重的供应链安全隐患和技术依赖风险。近年来，随着 100G 乃至 400G 高速以太网的商用化，基于以太网的 RoCE 存储网络因其可降低对国外交换机芯片的依赖及良好扩展性等优势，已成为现代化存储系统的重要发展方向。

【基金项目】中国铁路信息科技集团有限公司系统性重大项目“基于信创环境的中间件、数据库双活、集中式存储网络架构关键技术研究”子课题“三集中式存储网络架构关键技术研究”（项目编号：WJZG-CKY-2024042（2024P03））。

【作者简介】屈永康（1989—），男，中国山西阳泉人，硕士，高级工程师，从事存储技术与应用研究。

本文聚焦于传统环境下的光纤通道（FC）集中式存储区域网络技术与 RDMA over Converged Ethernet（RoCE）的集中式存储网络关键技术对比研究并通过系统性实验验证，信创化 RoCE 存储网络是否具备替代 FC 存储网络的技术可行性与保障性，从而为推进存储网络基础设施的自主化演进提供实证依据。本研究不仅有助于降低对国外核心技术依赖、提升我国关键信息基础设施的自主可控能力，也为铁路行业在高性能无损存储网络的技术选型、测试验证与实践应用方面提供了重要参考，填补了该领域在系统性实验研究与性能评估方面的空白。

2 关键技术分析

区别于业务网络，存储网络通常限定于局域网范畴，其架构设计不涉及与外部广域网或公共互联网的数据交换。这一封闭性特征使其性能优化焦点高度集中于网络内在属性的协同提升：严格保障低延迟、零丢包率的确定性传输质量，同时需兼顾高带宽的可持续吞吐能力。鉴于存储操作对数据完整性和 I/O 确定性的刚性约束，任何网络层抖动或丢包均可能导致系统级故障，因此存储网络的性能指标构成其核心设计边界。

为全面评估 RoCE v2 与 FC 技术，本研究构建了四维理论评估模型：

协议架构：分析协议栈设计与分层结构

延迟机制：剖析低延迟实现原理与确定性保障

无损传输：评估无损网络保障机制的有效性

流量控制：拥塞控制算法分析。

2.1 协议架构

FC 与 RoCE 协议的核心差异之一体现在协议架构上。FC 作为专用存储网络，其协议栈高度精简（仅 5 层），通过彻底的硬件卸载（HBA）及基于信用的 BB_Credit 无损流量控制机制，实现了超低延迟与“零丢包”，但其技术生态相对封闭。而 RoCE v2 构建于通用以太网之上，依托 TCP/IP 协议栈，通过 RDMA 层实现内核旁路与零拷贝传输，在保持高性能的同时继承了以太网的开放性与扩展性。其依赖于 PFC（优先级流量控制）和 ECN（显式拥塞通知）来实现无损网络，但管理复杂度较高。同时，协议处理开销与协议栈深度呈线性关系。FC 协议栈需 5 层即可完成存储数据传输，RoCE 需 6 层，而传统 iSCSI 则需 7 层以上。

通过数学模型可以将理论协议处理延迟简化表示为：

$$T_{\text{proto}} = \mathrm{k} \times D + C \qquad (1)$$

其中：k 表示每层处理延迟系数，D 表示有效协议栈深度 C 表示固定开销。因此，通过理论推导表明，FC 和 RoCE 的延迟显著低于 iSCSI，且两者差距相对较小。

2.2 延迟机制

FC 与 RoCE 在实现低延迟的路径上体现了“专用硬件确定性”与“通用协议优化性”两种不同方式。FC 作为专用存储网络，其低延迟根植于全链路硬件保障与原生无损设计。通过直通交换技术将每跳交换延迟降至微秒级，并依靠 BB_Credit 信用机制在链路层实现预防性流量控制，从根本上杜绝了拥塞与排队，从而获得高度确定性的亚微秒级延迟抖动。整个协议栈由 HBA 硬件卸载，主机 CPU 零参与。

RoCE v2 则是在通用以太网上通过协议创新逼近专用性能。其核心是通过 RDMA 实现内核旁路与零拷贝，消除了操作系统开销，节省了数十微秒。然而，以太网本身并非无损，因此 RoCE 必须依赖 PFC（链路层暂停）、ECN（拥塞标记）及 DCQCN 等复杂算法在应用层构建一个“软”无损网络。

2.3 无损传输

FC 与 RoCE 在实现无损传输的机制上存在根本性差异，体现了预防性控制与反应性控制两种设计方式。FC 的 BB_Credit 机制是一种预防性、点对点的硬性保障。它在发送前通过信用授权确保接收端缓冲区可用，从源头上杜绝了因缓冲区溢出导致的丢包。该机制独立作用于每条物理链路，逻辑简单确定，几乎无死锁风险，实现了理论上的零丢包与高度可预测的延迟。

相比之下，RoCE v2 依赖的 PFC 与 ECN 机制是一种反应性、依赖网络协作的软性保障。PFC 在拥塞发生后通过发送 PAUSE 帧紧急刹停上游流量；ECN 则尝试在队列堆积初期标记报文、通知端到端减速。这套机制在通用以太网上构建了一个“无损”覆盖层，但其反应延迟、参数敏感性以及 PFC 可能引发的死锁与流量振荡问题，使其在复杂组网中较难达到 FC 的确定性。不过，目前大量学者对 RoCE v2 技术的缺陷，提出了多种拥塞控制算法并结合基础机制使用，推动存储网络由反应式转向预测式。

2.4 流量控制

FC 没有传统意义上的“拥塞控制算法”，它通过底层基于信用的 BB_Credit 硬件流控机制，从链路层预防了缓冲区溢出，从而在根本上避免了由此引发的拥塞。而 RoCE v2 则必须依赖 PFC/ECN 等无损传输策略，并配合上层拥塞控制算法，才能在通用以太网上实现接近无损的传输。本文以 DCQCN 算法为依据，重点分析其控制原理，在后续章节中对 RoCE 存储网络的时延特性展开实验研究与对比讨论。

DCQCN 速率控制公式如下；

（1）速率降低：在收到 CNP 后，发送端将会降速操作。

$$\begin{aligned} R_{\mathrm{t}} &= R_c\left(1-\frac{\alpha}{2}\right) \\ \alpha &= (1-g)\alpha + g \end{aligned} \qquad (2)$$

其中：α 为折减系数；g 为预配置常数（一般为 $\frac{1}{16}$）；R_c 为当前速率；R_t 为目标速率。

速率回复：在一定时间内未收到 CNP 后，发送端将会进行提速操作。

$$R_t = R_t + R_{AI} \quad (3)$$

其中：R_{AI} 为固定增长速率。

在实际应用场景中，FC SAN 的吞吐性能主要受物理带宽及信用值与网络带宽延迟积匹配度的制约，而 RoCE v2 的传输效率与稳定性，则高度依赖于其由 PFC/ECN 基础与拥塞控制算法（如 DCQCN）共同构成的流控体系的效能。

2.5 小结

本章从协议架构、延迟机制、无损传输及流量控制四个维度，系统对比了 FC 与 RoCE SAN 的技术原理。分析表明，FC 通过精简协议栈、专用硬件卸载及预防性的 BB_Credit 机制，构建了一个确定性的高性能专用存储网络；而 RoCE 则在通用以太网基础上，依托 RDMA、PFC/ECN 及 DCQCN 等机制，以复杂性与管理开销为代价，实现了面向开放的近无损高性能传输。理论层面的剖析，揭示了两者在设计思想与实现路径上的根本差异。下一章将基于此框架，设计并执行系统性实验，以实测数据量化评估 RoCE SAN 在数据库下的性能表现，特别是其延迟确定性、吞吐效率及对拥塞的响应特性，从而实证检验其在关键场景中替代 FC 的技术可行性与边界。

3 实验方法

3.1 实验环境介绍

测试环境由 3 台数据库服务器，各节点之间以 RoCE 和 FC 交换机互联，测试环境如表 1、表 2 所示，其拓扑图如图 1 所示。

表 1 测试环境硬件配置

设备类型	配置信息	数量
存储设备	OceanStora Dorado	1
服务器设备	宝德 PR210K 新华三 UniServer R4930 G3	3
网络交换机	CE6855-64 CQ EI	2
RoCE 交换机	CE 6860-SAN	2
FC 交换机	SNS3664	2
其他	1 * 2 端口 25GE RoCE 网卡 1 * 2 端口 32Gb FC HBA 卡 1 * 2 端口 10GE ETH 网卡	若干

表 2 测试软件配置

设备类型	配置信息	版本
数据库	GaussDB	505.2.RC1 及以上
操作系统	银河麒麟 V10 SP1	4.19.90-23.23.v2101.ky10.aarch64
测试工具	Vdbench 50406 BenchmarkSQL 5.0	/
数据库	GaussDB	505.2.RC1

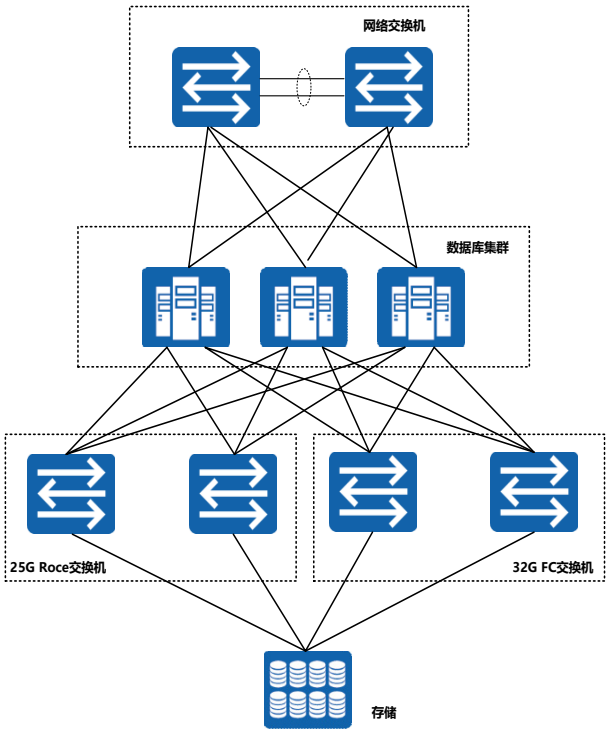

图 1 测试环境拓扑图

3.2 实验内容与方法

本次试验方法主要以数据库系统性能评估为参考，基准测试作为一种标准化的实验方法，被广泛用于量化系统在特定工作负载下的吞吐量、响应时间及可扩展性等关键指标。其核心目的在于提供可重复、可验证的性能数据，以支持系统优化、架构比较及容量规划决策。本文将使用两种主流基准测试方法：TPC-C 与 sysbench，重点分析其设计原理、测试框架及性能度量体系。

方法一：通过脚本模拟包含点查询、范围查询、更新、插入及删除等操作的典型 OLTP 读写混合业务负载。测试基于预先构建的规模为 20 张数据表、每表 2000 万行数据的测试集，配置客户端以 384 个并发线程的 PostgreSQL 数据库实例持续施加压力，测试时长为 300 秒。在测试执行过程中，设定每秒记录一次包含事务吞吐量与延迟在内的实时性能指标，并以 95% 响应时间百分位作为核心延迟统计依据，所有测试输出结果均被定向保存至系统日志文件以供分析，整个测试进程以后台服务模式运行以确保不受会话中断影响。

方法二：首登录至数据库节点并切换至数据库用户，加载相应环境变量后，通过 gsql 命令行工具连接集群主，执行创建测试数据库 db_ptcc 以及测试用户 tpcc_user 并授予全部权限的操作。随后在测试机上利用 BenchmarkSQL 测试工具向数据库集群预埋 1000 仓 TPCC 模型数据，通过系统化调整并发数进行 TPCC 模型性能测试，测试持续时间为

10 分钟，以获取最优的 tpmc 值作为性能评估依据。

3.3 实验结果

Sysbench 实验结果

组网	TPS	QPS	平均时延 (ms)	95% 时延 (ms)
RoCE	13048.79	260976.99	29.41	45.79
FC	11043.55	220872.79	34.75	50.11
FC/RoCE	84.6%	84.6%	118%	109%

基于 Sysbench 性能测试数据的分析表明，RoCE 存储网络方案在数据库负载下综合性能优于传统 FC 方案。RoCE 的 TPS 与 QPS 分别达到 13048.79 与 260976.99，较 FC 方案提升约 15.4%。其平均延迟与 95% 延迟分别为 29.41 毫秒与 45.79 毫秒，较 FC 方案降低约 15.4% 与 8.5%。方法一实验结果表明，FC 方案在吞吐性能上仅为 RoCE 的 84.6%，且平均延迟高出约 18%，体现了 RoCE 在高效数据处理方面的明显优势。

TPCC 实验结果

组网类型	TpmC
RoCE	436294.78
FC	391488.54
FC/RoCE	89.7%

基于 TPCC 基准测试结果，对比 RoCE 与 FC 两种存储网络在事务处理性能上的表现。实验数据显示，RoCE 组网下的 tpmC 值为 436,294.78，而 FC 组网下为 391,488.54，FC 性能相当于 RoCE 的 89.7%。结果表明，在 TPCC 模型所模拟的高并发事务负载环境下，RoCE 存储网络相比 FC 具有更优的事务处理吞吐能力，性能提升约 10.3%。

4 结语

本文对 RoCE 和 FC 存储网络硬件作为互联方案进行了系统的性能评估。实验结果表明，在标准数据库负载测试（Sysbench）中，RoCE 相较于 FC 展现出更优的吞吐与响应性能，其事务处理与查询能力提升约 15.4%，平均延迟降低约 15.4%。在高并发事务负载测试（TPCC）中，RoCE 方案的事务处理性能亦优于 FC，提升约 10.3%。综合来看，在数据库及高并发事务处理场景下，RoCE 存储网络的整体性能优于传统 FC 方案。这些结论为构建高性能数据存储与计算集群时的互联方案选择提供了实证依据与参考。

参考文献

[1] IDC.(2023).Worldwide Enterprise Storage Systems Tracker, Q4 2023,Doc #US49872323.

[2] 余胜生,初莹莹,周敬利,等.基于RDMA协议的零拷贝技术研究.计算机工程与应用,2004,(03):126-128.

[3] Alizadeh M,Greenberg A, Maltz D A, et al. Data center tep (dctcp) CJ/ Proceedings of the ACM SIGCOMM 2010 conference. 2010: 63-74.

[4] 汪洋,骆兰军,虞玲玲,祝春荣,张帆,高洋洋,等.NoF+存储网络解决方案,2023-10-10.

[5] Guo, X., Zhang, R., & Chen, W.»Enhancing Data Center Networks with RoCE: Performance Analysis and Optimization Strategies,»Proceedings of the International Conference on High Performance Computing, pp. 215-228, 2024.

[6] Chen, S., Liu, Y., & Wang, T.»A Comparative Study of InfiniBand and RoCE in Modern Data Centers: Protocol Stack, Congestion Control, and QoS,»Computer Networks, vol. 228, 109732, 2024.

[7] Zhang, Y., Chen, X., & Wang, L. «RoCEv2 Network Congestion Control Mechanisms: A Comprehensive Analysis,» Journal of Network and Computer Applications, vol. 185, pp. 103-120, 2023. doi: 10.1016/j.jnca.2023.103120.

[8] Fibre Channel Industry Association. Fibre Channel Storage Area Networks. 2001.