DDUC：数据更新与编码解耦的纠删码系统

屠要峰[1,2]，肖蓉[2]，韩银俊[1,2]，陈正华[2]，金浩[2]，齐学成[2]，孙辛远[2]
[1]移动网络和移动多媒体技术国家重点实验室，中国深圳市，518000
[2]中兴通讯股份有限公司，中国南京市，210000

摘要：在分布式存储系统中，常用的数据冗余方法包括副本和纠删码(erasure code，EC)。相较于副本，EC 具有更好的存储效率，但是在更新方面的开销更大。此外，并发更新带来的一致性和可靠性问题给 EC 应用带来了新的挑战。许多研究工作都致力于优化 EC 技术，包括算法优化、数据更新方法创新等，但并发更新的一致性和可靠性问题尚未得到很好解决。本文介绍了一种将数据更新与 EC 编码解耦的存储系统，命名为 DDUC，并提出了一种副本与校验块结合的放置策略。对于(N, M)的 EC 系统，按照 N 组 M+1 的副本进行数据布局，并将同一条带的冗余数据块都放置在校验节点上，使得校验节点可以自主地执行本地 EC 编码。基于上述策略，实现了一种两阶段数据更新方法，在第一阶段按照副本模式进行数据更新，在第二阶段由校验节点独立完成 EC 编码。这样在保证高并发性能的同时，解决了并发更新导致的数据可靠性降低的问题。同时利用 PMem 硬件的字节寻址和 8 字节原子写特性实现了一种轻量级的日志机制，在提升性能的同时保证了数据的一致性。实验结果表明，和当前主流的存储系统 Ceph 相比，本文所提出的存储系统并发访问性能提升至 1.70–3.73 倍，时延仅为 Ceph 的 3.4%–5.9%。

关键词：并发更新；高可靠性；纠删码；一致性；分布式存储系统
https://doi.org/10.1631/FITEE.2200466