

doi:10.1631/FITEE.1700626

**题目:** ONFS: 面向高性能计算的基于内存、固态硬盘和磁盘的层次式混合文件系统

**概要:** 随着超级计算机向 Eflops 规模快速发展和计算核数急剧增加, 更大规模和更复杂的应用成为可能。

大规模科学计算、新的 workflow 应用以及检查点操作均需要存储系统具有非常高的带宽和低延迟, 这使得高性能存储系统面临严峻的技术挑战。当前基于磁盘的底层存储系统难以满足新一代 Eflops 超级计算机和应用的要求。为此, 本文提出了基于计算节点内存、固态硬盘和磁盘的层次式混合存储系统 ONFS (on-line and near-line file system)。它具有三个存储层次和统一的命名空间, 支持可移植操作系统接口 (portable operating system interface, POSIX) 协议, 可提供高带宽、低延迟和超大存储容量。本文详细分析了分布式元数据管理、内存借用和归还策略、数据一致性、并行访问控制, 以及向下迁移和向上主动预迁移机制。在天河一号超级计算机上实现了 ONFS 原型系统, 测试了 I/O (input/output) 性能和可扩展性。测试结果表明, 单线程和多线程读/写性能比基于磁盘的 Lustre 分别高出 6 倍和 5 倍。与 Lustre 相比, 运行在 ONFS 上的典型数据密集型应用可获得 6.35 倍的 I/O 加速。

**关键词:** 高性能计算; 层次式混合存储系统; 分布式元数据管理; 数据迁移