基于目录路径的元数据管理方法

刘 仲, 周兴铭

刘 仲, 周兴铭

(国防科学技术大学 计算机学院,湖南 长沙 410073)

作者简介: 刘仲(1971－),男,湖南邵东人,博士,副教授,主要研究领域为网络存储,并行与分布处理.周兴铭(1938－),男,教授,博士生导师,中国科学院院士,CCF高级会员,主要研究领域为高性能体系结构,并行与分布处理.

联系人: 刘 仲 E-mail: zhongliu@nudt.edu.cn, http://www.nudt.edu.cn

Abstract

A metadata management method dividing directory path attribute from directory object is proposed, which extends the present object storage architecture. This method avoids efficiently the large-scale metadata migration according to the updating directory attributes, improves the cache utilization and hit rate by reducing the overlap cache of prefix directory, reduces the disks I/O demands by reducing the overhead of traversing the directory path and exploiting directory locality, and avoids overloading a single metadata server by dynamic load balancing. Experimental results demonstrate that this method has obvious advantages in improving the throughput, scalability, balancing metadata distribution, and in reducing the metadata migration.

摘要

提出目录路径属性与目录对象分离的元数据管理方法,扩展了现有的对象存储结构.该方法能够有效避免因为目录属性修改而导致的大量元数据更新与迁移;通过减少前缀目录的重迭缓存提高了元数据服务器Cache的利用率和命中率;通过减少遍历目录路径的开销和充分开发目录的存储局部性,减少了磁盘I/O次数;通过元数据服务器的动态负载均衡避免单个服务器过载.实验结果表明,该方法在提高系统性能、均衡元数据分布以及减少元数据迁移等方面具有明显的优势.

References:

[1] Nagle D, Serenyi D, Matthews A. The Panasas ActiveScale storage cluster delivering scalable high bandwidth storage. In: Benton V, ed. Proc. of the ACM/IEEE SC 2004 Conf. Washington: IEEE Computer Society, 2004. 53-62.

[2] Lustre SP. Building a file system for 1000 node clusters. In: John WL, ed. Proc. of the 2003 Ottawa Linux Symp. Ottawa: Red Hat, Inc., 2003. 401-407.

[3] Brandt SA, Xue L, Miller EL, Long DDE. Efficient metadata management in large distributed file systems. In: Miller E, Meter RV, eds. Proc. of the 20th IEEE/11th NASA Goddard Conf. on Mass Storage Systems and Technologies. San Diego: IEEE Computer Society, 2003. 290-298.

[4] Weil SA, Pollack KT, Brandt SA, Miller EL. Dynamic metadata management for Petabyte-scale file systems. In: Benton V, ed. Proc. of the ACM/IEEE SC 2004 Conf. Washington: IEEE Computer Society, 2004. 4-15.

[5] Liu Z, Zhou XM. A data object placement algorithm based on dynamic interval mapping. Journal of Software, 2005,16(11): 1886-1893 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/16/1886.htm

[6] Liu Z, Zhou XM. An adaptive data objects placement algorithm for non-uniform capacities. In: Jin H, Pan Y, Xiao N, Sun JH, eds. Proc. of the 3rd Int'l Conf. on Grid and Cooperative Computing. LNCS 3251, Berlin, Heidelberg: Springer-Verlag, 2004. 423-430.

[7] Liu Z, Xiao N, Zhou XM. RDIM: A self-adaptive and balanced distribution for replicated data in scalable storage clusters. In: Pan Y, Chen DX, Guo MY, Cao JN, Dongarra J, eds. Proc. of the 3rd Int'l Symp. on Parallel and Distributed Processing and Applications. LNCS 3758, Berlin, Heidelberg: Springer-Verlag, 2005. 21-32.

[8] Ganger GR, Kaashoek MF. Embedded inodes and explicit groupings: Exploiting disk bandwidth for small files. In: Proc. of the 1997 USENIX Annual Technical Conf. Anaheim: USENIX, 1997. 1-17.

[9] Litwin W, Neimat MA, Schneider DA. LH*—A scalable, distributed data structure. ACM Trans. on Database Systems, 1996,21(4):480-525.

[10] McKusick MK, Ganger GR. Soft updates: A technique for eliminating most synchronous writes in the fast file system. In: Proc. of the FREENIX Track: 1999 USENIX Annual Technical Conf. Monterey: USENIX, 1999. 1-18.

[11] Duchamp D. Optimistic lookup of whole NFS paths in a single operation. In: Proc. of the USENIX Summer 1994 Technical Conf. Boston: USENIX, 1994. 161-169.

附中文参考文献:
[5] 刘仲,周兴铭.基于动态区间映射的数据对象布局算法.软件学报,2005,16(11):1886-1893. http://www.jos.org.cn/1000-9825/16/1886.htm