P.O.Box 8718, Beijing 100080, China	Journal of Software, May 2005,16(5):1028-1038
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2005 by The Editorial Department of Journal of Software

动态盘阵D/H分布与基于控制理论的在线重构

刘 军, 杨学军, 王俊伟, 唐玉华

刘 军, 杨学军, 王俊伟, 唐玉华
(国防科学技术大学 计算机学院,湖南 长沙 410073)
作者简介: 刘军(1971－),男,湖南安化人,博士,主要研究领域为并行IO;杨学军(1963－),男,教授,博士生导师,主要研究领域为高性能计算;王俊伟(1975－),男,博士生,主要研究领域为并行计算;唐玉华(1962－),女,研究员,主要研究领域为计算机网络与通信.
联系人: 刘 军 Phn: +86-731-4573679, E-mail: muplj@tom.com, http://www.nudt.edu.cn

Received 2004-01-29; Accepted 2004-05-08

Abstract
Disk Array is adopted widely because of its high performance I/O. To adapt the need of applications' changeable I/O performance, I/O storage subsystem should be highly scalable. So DDA (dynamic disk array), which can scale adaptively, is an ideal system. The key technology of DDA is its data placement algorithm and online data reorganization algorithm. The main contribution of the paper is: first, a detailed study on DDA data placement is conducted and a new placement method, D/H, is presented. In D/H placement, the space in DDA is balanced after scale, and the reorganization cost is minimized; then, an online data reorganization algorithm based on control feedback theory is provided. With this strategy, the reorganization in DDA does little impression to the system QoS, and under this condition, data reorganization can be accomplished as quickly as possible; finally, simulation results show that Online Data Reorganization based on Control Theory is useful.

Liu J, Yang XJ, Wang JW, Tang YH. D/H placement and on-line data reorganization based on control theory in dynamic disk array. *Journal of Software*, 2005,16(5):1028-1038.
DOI: 10.1360/jos161028
http://www.jos.org.cn/1000-9825/16/1028.htm

摘要
由于能够提供高性能I/O,盘阵被广泛采用.但以往的盘阵扩展性不足.而用户或应用程序对外存容量和I/O性能需求是变化的,盘阵系统本身必须有很强的扩展性,以适应系统的I/O需求.因此,由于既具有盘阵的高性能I/O,又能通过增加或减去设备后进行数据重构实现性能的扩展,动态盘阵具有广泛的前景.动态盘阵的技术热点是数据分布算法和在线自适应数据重构技术,使得盘阵的性能和容量能够随着系统的扩展而伸缩,同时使得盘阵动态扩展时的数据重构对系统的影响非常小.主要工作是:第1,对动态盘阵的数据分布展开研究,并提出一种新的数据分布算法(D/H分布).在D/H分布中,盘阵扩展时始终保持各设备上空间和负载的平衡性,同时扩展时重构的数据最少;第2,针对D/H分布,提出基于控制理论的数据重构技术,使得盘阵动态扩展时的在线数据重构对请求QoS的影响非常小,同时使得数据重构能够尽快完成;第3,研究中针对Sprite trace和合成负载进行了大量模拟实验,结果表明,提出的基于控制理论的数据重构技术行之有效.

基金项目：Supported by the National Science Fund for Distinguished Young Scholars under Grant No.69825104 (国家杰出青年科学基金); the National High-Tech Research and Development Plan of China under Grant No.2002AA1Z2101 (国家高技术研究发展计划(863))

References:

[1] Shen XH, Choudhary AN, Matarazzo C, Sinha P. A distributed multi-storage resource architecture and I/O performance prediction for scientific computing. Cluster Computing, 2003,6(3):189-200.

[2] Harting H. An I/O architecture for microkernel-based operationg systems. Technical Report TUD-FI03-08-Juli-2003, Dresden, TU Dresden, 2003.

[3] Ho TK, Lee JYB. A row-permutated data reorganization algorithm for growing server-less video-on-demand systems. In: Lee S, Sekiguchi S, eds. Proc. of the 3rd IEEE/ACM Int'l Symp. on Cluster Computing and the Grid (CCGRID 2003). Tokyo: IEEE/ACM Press, 2003. 44-51.

[4] Liu J, Tang YH, Wang L, Jiang YH, Yang XJ. The dynamic mass storage system. Journal of Computer Research and Development, 2002,39(Suppl.):162-167 (in Chinese with English abstract).

[5] Ghandeharizadeh S, Kim D. On-Line reorganization of data in scalable continuous media servers. In: Feitelson DG, Rudolph L, eds. Proc. of the 7th Int'l Conf. on Database and Expert Systems Applications. Lecture Notes in Computer Science, Zurich, 1996. 751-768.

[6] Goel A, Shahabi C, Yao S-YD, Zimmermann R. SCADDAR: An efficient randomized technique to reorganize continuous media blocks. In: Chaudhuri S, Carey M, Garcia-Molina H, eds. Proc. of the 18th Int'l Conf. on Data Engineering (ICDE 2002). San Jose: IEEE CS Press, 2002. 473-482.

[7] Lu CY, Alvarez GA, Wilkes J. Aqueduct: Online data migration with performance guarantees. In: Chase J, Cole J, eds. Proc. of the USENIX Conf. on File and Storage Technologies (FAST). Monterey, 2002. 219-230.

[8] Abdelzaher TF. An automated profiling subsystem for QoS-aware services. In: Sztipanovits J, Abdelzaher TF, Atkins EM, eds. Proc. of the 6th IEEE Real Time Technology and Applications Symp. Washington DC: IEEE Press, 2000. 208-217.

[9] Lumb CR, Merchant A, Alvarez GA. Facade: Virtual storage devices with performance guarantees. In: Chase J, Cole J, eds. Proc. of the 2nd USENIX Conf. on File and Storage Technologies (FAST). San Francisco, 2003. 131-144.

[10] Madell T, Madell T. Disk and File Management Tasks in HP-UX. Printice-Hall: Pearson Education POD, 1997.

[11] Franklin GF, Powell DJ, Workman ML. Digital Control of Dynamic Systems. 3rd ed., New York: Addison-Wesley, 1998.

[12] Li BC, Nahrstedt K. A control-based middleware framework for quality of service adaptations. IEEE Journal of Selected Areas in Communication, Special Issue on Service Enabling Platforms, 1999,17(9):1632-1650.

[13] Lu CY, Stankovic JA, Tao G, Son SH. Feedback control real-time scheduling: Framework, modeling, and algorithms. Journal of Real-Time Systems, 2002,22(1/2):85-126.

[14] Holland M, Gibson GA. Parity declustering for continuous operation in redundant disk arrays. In: Mukherjee S, McKinley KS, eds. Proc. of the 5th Int'l Conf. on Architectural Support for Programming Languages and Operationg Systems (ASPLOS-V). Boston MA: ACM Press, 1992. 23-35.

[15] Gribble S, Manku G, Roselli E, Brewer E. Self-Similarity in file systems. In: Gibson TJ, Gribble SD, eds. Proc. of the 1998 ACM SIGMETRICS Joint Int'l Conf. on Measurement and Modeling of Computer Systems. Madison: ACM Press, 1998. 141-150.

[16] Jain R. The Art of Computer Systems Performance Analysis. New York: John Wiley & Sons, Inc. 1991.

[17] Smith W, Foster I, Taylor V. Predicting application run time using historical information. In: Feitelson DG, Rudolph L, eds. Proc. of the Workshop on Job Scheduling Strategies for Parallel Processing. Lecture Notes in Computer Science, San Juan, 1999. 202-219.

[18] Schindler J, Griffin JL, Lumb CR, Ganger GR. Track-Aligned extents: Matching access patterns to disk drive characteristics. In: Chase J, Cole J, eds. Proc. of the Conf. on File and Storage Technologies (FAST). Monterey, 2002. 259-274.

[19] Lee EK, Katz RH. An analytic performance model of disk arrays. In: Lazowska ED, Felten EW, eds. Proc. of the 1993 ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems. Santa Clara: ACM Press, 1993. 98-109.

[20] Baker MG, Hartman JH, Kupfer MD, Shirriff KW, Ousterhout JK. Measurements of a distributed file system. In: Proc. of the 13th ACM Symp. on Operating System Principles. Pacific Grove: ACM Press, 1991. 198-213.

附中文参考文献:

[4] 刘军,唐玉华,王磊,蒋艳凰,杨学军.动态海量存储系统.计算机研究与发展,2002,39(Suppl.):162-167.