

基于 CDN 和 P2P 的分布式网络存储系统

马 军 郑 炅 殷保群

(中国科学技术大学网络传播系统与控制联合实验室网络传播系统与控制安徽省重点实验室 安徽 合肥 230027)

摘 要 把用户的文件分片后均衡存储在不同的分布式存储节点上,并利用虚拟目录服务器和基于 P2P-DHT 的目录服务器把文件元数据与文件数据片高效地对应起来,以提供高效目录服务,分布式存储节点以 P2P 方式工作以快速完成用户对文件数据的请求任务。分布式网络存储系统 DNSS 充分利用了 CDN 和 P2P 的技术优势,有较高的可用性、可靠性和可扩展性。DNSS 已经在中国科学技术大学应用。

关键词 分布式存储 P2P 虚拟目录 DHT

CDN AND P2P-BASED DISTRIBUTED NETWORK STORAGE SYSTEM

Ma Jun Zheng Quan Yin Baoqun

(Joint Lab of NCSC, Key Lab of Anhui NCSC, University of Science and Technology of China, Hefei 230027, Anhui, China)

Abstract The user's file is balanced stored at different distributed storage nodes after being split. File metadata and file data slice are matched high-efficiently by virtual directory server and P2P-DHT-based directory server in order to provide high-efficient directory service. Distributed storage nodes complete quickly by means of P2P system the tasks of file data requested by users. Distributed network storage system (DNSS) takes full advantage of the technology superiority of CDN and P2P and is of high availability, reliability and scalability. DNSS has been used in University of Science and Technology of China.

Keywords Distributed storage P2P Virtual directory DHT

0 引 言

随着网络技术的不断发展、网络带宽的不断提高以及单机存储容量的不断增大,近年来,分布式网络存储系统的研究和应用已经成为一个热点^[1],如已经出现的分布式网络存储系统 OceanStore^[2]、CFS^[3]、PAST^[4,5]、Tangler^[6]、Pangaea^[7]、Granary^[8]等。CDN 作为一种提高网络内容、节省骨干网络带宽的技术,存在应用和扩展成本高的缺点^[9]。P2P 技术突破了传统的 C/S 架构的模式,具有非常好的扩展性,但存在安全性、可控性问题^[10]。本文充分利用 CDN 的资源管理优势和 P2P 的高扩展性,构建了一个在全互联网范围内使用的可靠高效的海量分布式存储系统 DNSS。

1 系统架构设计

1.1 系统架构

基于 CDN 的资源管理优势和 P2P 的高扩展性为实现 DNSS 提供了良好的支持,本文按照分层原则设计了整个系统的架构。DNSS 分成三层:用户接口层、CDN 中心管理层、以 P2P 方式工作的服务节点层。图 1 和图 2 分别表示 DNSS 的分层体系架构和系统架构。

• 用户接口层主要向用户提供操作接口,包括三个功能模块:Web 服务模块 (WS),用户管理模块 (UM),虚拟目录服务模块 (VDS)。Web 服务模块负责为用户提供文件操作界面;用户

管理模块主要负责管理和维护用户的信息;虚拟目录服务模块则负责组织和维护文件系统的树形目录结构信息。

- CDN 中心管理层主要负责维护全局文件数据分布信息、管理服务节点和实现全局负载均衡,包括三大功能模块:节点管理模块 (MM),基于 P2P-DHT 的目录服务模块 (DS),负载均衡模块 (LB)。节点管理模块主要负责管理服务节点的增加、退出、删除等;目录服务模块则管理和维护整个系统的每个文件在各个服务节点的存储信息;负载均衡模块则主要负责整个系统的负载均衡;
- 服务节点层主要负责整个系统的文件数据的存储以及对用户提供服务。该层由很多服务节点 (PN) 组成,所有服务节点间以 P2P 方式相互协作,完成 CDN 中心管理层下发的用户对文件的操作任务。

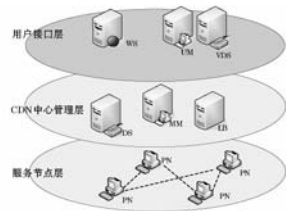


图 1 DNSS 分层体系架构

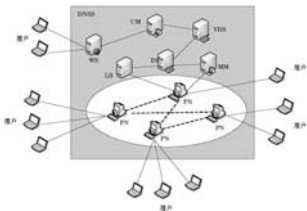


图 2 DNSS 系统架构图


```
//服务节点
Long pnLoad = MAXLOAD;
Long pnID = -1;
For (int i=0; i<pns.length; i++)
{
    Pns[i].load < pnLoad;
    pnLoad = Pns[i].load;
    pnID = Pns[i].pnID;
}
//把任务分配到这些节点中当前负载最轻的服务节点上
setServer(pnID);
```

2.4 数据备份

随着广域网技术的不断发展,数据信息访问量越来越趋向海量。在频繁地对数据进行访问的过程中,数据丢失和系统不稳定成为越来越突出的问题。从而,为了保障服务的可靠性,需要对数据备份存储。目前,高可用的数据备份存储有 RAID 和数据复制等。DNSS 采用双机备份,每个中心服务器与服务节点都有一个与其相邻的服务节点且含有与其完全一样的数据存储结构,可以认为是这两个服务节点互相成为对方的影像。服务器及其影像同时提供服务,这样虽然不能扩大系统的存储容量,但是提高了系统的可靠性和服务效率,并减少了用户的等待时延。

3 总 结

本文设计与实现的基于 CDN 和 P2P 相结合的 DNSS,能有效利用 CDN 架构的优点和 P2P 的技术优势,使系统有较高的可用性和扩展性。本文采取的文件结构能够有效地建立起目录树,并可以方便地实现文件或者目录的共享,元数据与文件数据的分离,以及文件数据的分布式存储;所采取的存储策略和调度算法能够有效地解决各个文件服务器的存储和服务负载均衡问题。

参 考 文 献

[1] 魏青松. 大规模分布式存储技术研究[D]. 电子科技大学,2004.

[2] Kubiawicz J, Bindel D, Chen Y, et al. OceanStore: An Architecture for Global-scale Persistent Storage[C]//9th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS IX). November 2000.

[3] Bek F, Kaashoek M F, Karger D, et al. Wide-area Cooperative Storage with CFS[C]//18th ACM Symposium on Operating Systems Principles (SOSP01). October 2001.

[4] Peter Druschel, Antony Rowstron. PAST: A Large-scale, Persistent Peer-to-Peer Storage Utility[C]//8th Workshop on Hot Topics in Operating Systems (HotOS VIII). May 2001.

[5] Rowstron A, Druschel P. Storage Management and Caching in PAST, a Large-scale Persistent Peer-to-Peer Storage Utility [C]//18th ACM Symposium on Operating Systems Principles (SOSP01). October 2001.

[6] Waldman M, Mazieres D. Tangler: A Censorship-Resistant Publishing System Based On Document Entanglements[C]//8th ACM Conference on Computer and Communication Security (CCS-8). November 2001.

[7] Saito Y, Karamanolis C, Karlsson M, et al. Taming Aggressive Replication in the Pangaea Wide-Area File System[C]//5th Symposium Operating Systems Design and Implementation (OSDI02). December 2002.

[8] Zheng W, Hu J, Li M. Granary: Architecture of Object Oriented Internet Storage Service [C]//IEEE International Conference on E-Commerce Technology for Dynamic E-Business (CEC-EAST 2004). September 2004.

[9] Shoaib Khan, Rüdiger Schollmeier, Eckehard Steinbach. A Perform-

ance Comparison of Multiple Description Video Streaming in Peer-to-Peer and Content Delivery Networks [C]//2004 IEEE International Conference on Multimedia and Expo (ICME).

[10] Adarand E, Huberman B. Free riding on Gnutella. Technical Report, CSL-00-3, Palo Alto; Xerox PARC, 2000.

(上接第 35 页)

SPACE(4.0)下的运行时间情况,可以知道 SPACE(δ)的 δ 越大,预处理时间越短。

表 1 3 种细化算法在不同粗糙空间下运行时间对比

	图像在 SPACE(1.0)下			图像在 SPACE(2.0)下			图像在 SPACE(4.0)下		
	FPA	文献[3]	文献[4]	FPA	文献[3]	文献[4]	FPA	文献[3]	文献[4]
图像 1	5.4370	4.9060	4.1720	3.0470	1.5310	1.2660	0.4220	0.4220	0.4220
图像 2	6.9370	6.4060	3.3440	1.5000	1.4690	1.2350	0.3750	0.3600	0.3750
图像 3	5.5780	5.3120	2.4370	1.0940	1.0940	1.0310	0.3750	0.3600	0.2970
图像 4	3.8910	3.2970	2.5000	1.8910	1.1560	1.0620	0.7030	0.4220	0.4220
图像 5	8.7650	4.8900	3.3120	1.4840	1.4680	1.0310	0.5620	0.3750	0.3750

注:时间单位为 ms

在不能破坏汉字的拓扑结构和本质特征前提下,增大 δ ,粗糙化不同粗糙空间下汉字图像,再细化粗糙化图像。500 个样本实验结果的统计数据如表 2 所示, δ 越大,去除的伪特征点越多,但当 $\delta=8.0$ 时,部分粗糙化图像拓扑结构被改变。而在 SPACE(4.0)下,拓扑结构和本质特征保持良好,同时可去除大量伪特征点,因此 δ 取 4.0 较好。

表 2 不同粗糙空间下汉字图像的总体情况

	图像在 SPACE (2.0)下	图像在 SPACE (4.0)下	图像在 SPACE (8.0)下
拓扑结构和本质特征 保持情况(%)	99.9%	99.6%	86.7%
去除伪特征点数(个)	194	358	372

4 总 结

经过粗糙化后的手写汉字图像,图像的尺寸可以减少,对粗糙化后的汉字图像细化,可以防止一些断点、伪交叉点、伪线段的产生,提高了手写汉字的识别率和识别速度。

参 考 文 献

[1] 马垣. 粗糙几何学[C]//中国第 6 届 Rough 集及软计算学术会议大会报告,2006.

[2] Chen M S, Park J S, Yu P S. Efficient Data Mining for Path Traversal Patterns[J]. IEEE Trans. Knowledge Data Engineer, 1998, 10(2): 209-211.

[3] 贺继刚, 杨晓伟, 吴广潮, 等. 基于模板保留的快速并行细化算法[J]. 计算机应用与软件, 2007, 24(12): 26-28.

[4] 廖开阳, 张学东, 章明珠. 一种新的指纹图像快速细化算法[J]. 计算机工程与应用, 2008, 44(5): 93-95.

[5] 黄铁英, 姜显明. 一种快速手写汉字细化算法[J]. 计算机工程. 2004, 30(19): 121-128.

[6] 吴选忠. Zhang 快速并行细化算法的扩展[J]. 福建工程学院学报, 2006, 4(1): 89-92.

[7] 史绍强. 一种改进型的汉字字符图像细化算法[J]. 计算机技术与发展, 2007, 17(9): 88-91.