基于云存储的网盘系统架构及关键技术研究 *

杨岳湘,邓文平,邓劲生,李　阳
(国防科技大学信息中心　长沙 410073)

摘　要　面向企业网或校园网的移动办公与存储的网盘系统有着广泛的市场需求,传统的网盘技术在性能、用户共享、安全性、可扩展性等方面存在诸多缺陷。针对这些不足,本文提出了一种基于云存储的高性能网盘系统架构:采用分布式文件系统 MooseFS 实现用户数据存储与访问的集群架构;在安全性方面,结合 SAMBA 实现用户权限管理, 用户数据存储支持 128 bit AES 加密,SSH 保证了传输链路的安全;最后,结合用户的实际需求,提供基于 Web 的访问方式以及客户端的同步盘模式。结果表明,系统在性能、安全性、可扩展性等多方面具有显著优势。

关键词　网盘;云存储;分布式文件系统;集群;用户空间
文献标识码　A doi: 10.3969/j.issn.1000-0801.2012.10.012

1　引言

网络硬盘(简称网盘)是一种基于用户空间的在线存储服务,用户可以通过网络进行数据上传、下载、共享等操作。用户对面向企业网或校园网移动办公与存储的网盘系统有着广阔的市场需求。随着存储技术的不断发展,传统的网盘技术已经显得力不从心,存在传输速度慢、容灾备份及恢复能力低、安全性差、营运成本高等诸多瓶颈。最新应用的云存储技术,为网盘行业带来了新的革命。云存储是构建在高速分布式存储网络上的数据中心,它将网络中大量不同类型的存储设备通过应用软件集合起来协同工作,形成一个安全的数据存储和访问系统,适用于各大中小型企业与个人用户的数据资料存储、备份、归档等一系列需求。近年来, 国外典型的云存储服务产品有 iCloud、Cloud Drive、Dropbox、SkyDrive 和 Amazon S3 等。自 2009年以来,云存储在我国得到了大力发展和推广,联想、新浪、华为、金山、奇虎 360、腾讯等著名企业分别推出了相应的云存储产品或服务,清华大学也为其内部用户推出了云存储服务(Corsair 和 Meepo)[1]。

尽管业内已经提供了大量云存储服务相关的产品,由于目前的云存储服务大多是商业化产品,对企业网和校园网的信息化建设来说,存在几个显著的缺陷。首先,现有的云存储服务需要依托第三方提供的数据中心设施,用户将数据托管给第三方,通过公有云、私有云或混合云形式对数据进行按需存取操作,这种非完全自主的管理模式存在泄密的风险;其次,第三方提供的存储服务平台价格高昂,对上层用户完全透明,缺乏可靠性保障。为此,业内针对云存储的私密性保护和安全存储进行了大量的研究。参考文献[2]提出了一种保护云存储平台上用户数据私密性的方法;参考文献[3]针对云存储的安全存储策略进行了深入的研究,提出了一种面向云存储的安全存储模型及存取策略;参考文献[4]设计了一种支持隐私保护的云存储框架;参考文献[5~7]研究了云存储的密文访问控制方法。本文在以上安全模型和方法的基础上,实现云存储网盘系统的

* 国家自然科学基金资助项目(No.61170286,No.61202486)

安全存储与传输。

针对现有网盘系统的不足,本文提出了基于云存储的网盘解决方案。物理存储采用快速可靠的 SAN 存储技术及磁盘 RAID;通过 VMware Sphere 实现底层硬件资源池的构建与统一管理; 采用分布式文件系统 MooseFS (Moose file system)[8]实现用户数据存储与访问的集群架构,同时实现了业务流与数据流的分离;在安全性方面,结合 SAMBA 实现用户权限管理, 存储支持 128 bit AES 加密,SSH 保证了传输链路的安全; 最后, 针对用户的网盘需求, 实现了基于 Web 的访问方式以及客户端的同步盘模式。实验结果表明,系统在性能、安全性、可扩展性等多方面具有显著优势。

2 总体设计

如图 1 所示, 云存储网盘系统的结构模型由存储层、基础管理层、应用接口层、访问层 4 层组成。

(1)存储层

存储层是云存储最基础的部分。云存储中的存储设备通过广域网、互联网或者光纤通道(FC)网络连接在一起。本系统的存储设备主要是 SAN 存储设备(高速),也可以是 NAS、FC、iSCSI 等其他存储设备。为确保底层数据的可靠性, 物理磁盘做了磁盘 RAID。存储设备之上通过 VMware Sphere 实现了对硬件资源的统一管理, 实现了设备的虚拟化集中管理、多链路冗余管理以及硬件设备的状态监控和故障维护。

(2)基础管理层

基础管理层是云存储最核心的部分,也是云存储中最难以实现的部分。基础管理层通过服务器集群和分布式文件系统等技术,实现云存储中多个存储设备之间的协同工作,使多个存储设备可以对外提供同一种服务,并提供更大、更强、更好的数据访问性能。数据加密技术保证云存储中的数据不会被未授权的用户所访问;通过各种数据备份技术保证云存储中的数据不会丢失,保证云存储自身的安全和稳定;通过动态负载均衡技术将每次用户连接指派到负载最低的服务器,实现系统的高效服务。

(3)应用接口层

应用接口层是云存储最灵活多变的部分。不同的云存储运营单位可以根据实际业务类型,开发不同的应用服务接口,提供不同的应用服务。应用接口层融合了 SAMBA、邮件系统、Web Service,提供用户认证、用户权限管理、资源的共享与输出等。

(4)访问层

任何一个授权用户都可以登录到系统享受独立的网盘服务。系统为用户提供两种访问模式: 客户端模式和 Web 模式。在客户端模式下,用户将自己的网盘挂载到本地,并在本地建立同步盘,以支持用户离线操作。通过鼠标右键的下拉菜单可灵活地对文件/文件夹进行压缩和加解密。在 Web 模式下,用户可对文件夹进行压缩,对文件进行上传下载和加解密等,还可以直接对单个文件进行邮件转发,实现与其他用户之间的文件共享。

3 关键技术

基于云存储实现网络硬盘的关键技术包括:分布式文

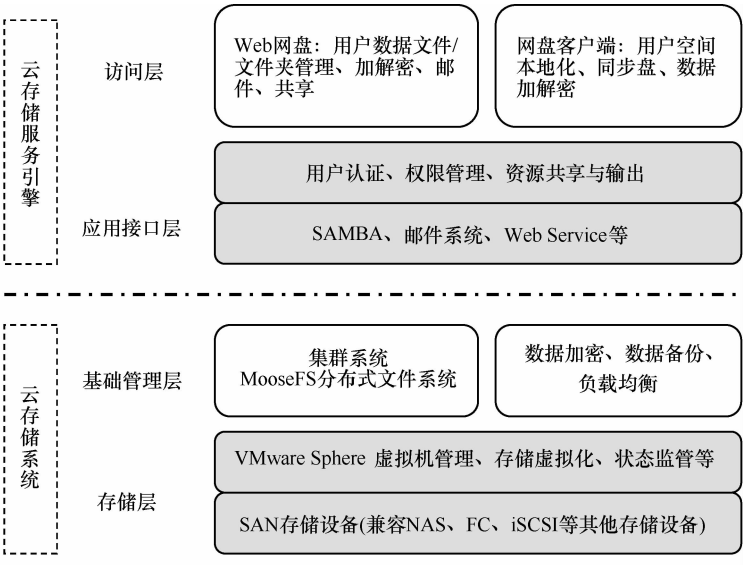

图 1 云存储网盘系统结构

件系统与集群架构、用户权限与数据安全、包含客户端及Web的双重应用模式,下面分别对这些关键技术进行介绍。

3.1 分布式文件系统与集群架构

分布式文件系统是云存储技术的核心。通过分布式文件系统,可以实现大规模、安全可靠的集群文件系统,可以支持系统的大规模扩展。目前,应用较广泛的文件系统包括Sun公司的Lustre、Hadoop架构的HDFS、MogileFS、FastDFS、MooseFS等,这些分布式文件系统各有特点,有些不具有通用性,而只适用于特殊的应用场景,比如,HDFS主要适用于超大数据集的应用程序处理。

MooseFS是一个开源免费的分布式文件系统,它是一种通用文件系统,不需要修改上层应用就可以使用。与其他分布式文件系统相比,MooseFS具有以下优点:

- 文件被分块存储于不同的数据节点,故障硬盘不存在数据泄露的风险;
- 可以在线扩容,体系架构可伸缩性极强;
- 部署简单;
- 体系架构高可用,所有组件无单点故障;
- 文件对象高可用,可设置任意的文件冗余程度;
- 提供Web GUI监控接口;
- 提高随机读或写的效率;
- 更重要的一点是,它可以提高海量小文件的读写效率,非常适用于企业网或校园网用户的随机文件存取。

结合应用背景,选取MooseFS作为分布式集群的基础架构,如图2所示。

- 主控节点:负责用户的业务控制流以及各个数据存储节点的管理、文件读写调度、文件空间回收以及恢复、多节点之间的拷贝。
- 存储节点:负责和用户之间的数据流,听从主控节点的调度,提供存储空间,并为客户提供数据传输。
- 元数据日志服务器:负责备份主控节点的变化日志文件,以便于在主控节点出故障时接替其进行工作。

用户在访问网盘空间时,首先发送请求到主控节点,主控节点再根据存储节点的负载情况动态指定一个存储节点服务器与用户进行数据交互。因此,业务流通过访问主控节点获取,而数据流通过访问存储节点获取,这种业务流与控制流分离的模式大大提高了数据的存取效率。

在可靠性方面,每个文件经过分块双份存储于两个不同的MooseFS存储节点上(在此基础上,MooseFS保证了数据自愈),保证了任意单节点出现故障时不影响数据的完整性,并通过多种安全加密措施最大程度地保证了数据存储的可靠性。

在负载均衡方面,系统根据存储节点的硬件资源利用率和各节点服务请求处理的情况,动态地在各存储服务节点上分发读写请求。

在扩展性方面,基于MooseFS集群,系统支持对存储节点、服务节点的在线扩展以及对存储域的在线扩容,实现业务不中断的前提下对服务节点、存储节点的添加。对于存储容量的扩展性,系统对业务流与数据流进行了分离,业务流通过访问主控节点获取,而数据则直接通过访问存储节点进行存储,大大提高了数据存取效率。系统支持对单个存储节点的容量扩展以及集群存储节点数的任意扩展。

3.2 用户权限与数据安全

数据安全机制方面,通过基于用户空间的权限控制、数据隔离、信息加密、传输加密等技术手段,保障用户数据的私密性与安全性。

(1)权限控制

系统面向多业务和多用户,任何对资源的访问都经过严格的权限控制。只有用户确认共享的资源才能被其他用户或业务进行访问。通过用户管理,实现用户身份的认证,为用户分配密钥,维护用户的密钥链。

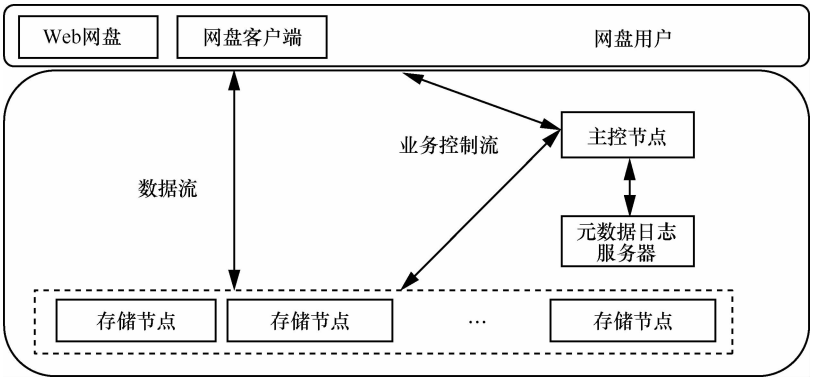

图2 MooseFS与集群架构

(2)数据隔离

系统为每个用户创建独立的存储空间,根据用户标识和对应权限对用户空间的数据进行访问控制,避免未授权用户访问到其他用户的数据以及用户信息;各种存储服务在进行数据存储和读取时, 每种应用都必须拥有自己独立的权限,系统根据不同的应用将数据隔离,避免数据被越权访问。

(3)数据加密

系统采用 128 bit AES 加密算法对用户文件进行加密,防止用户关键信息被盗取。用户在上传文件时可选择对文件数据进行加密,在获取数据时,可自动或手动解密,即使是系统管理员,也不能获得用户的加密数据。用户口令由用户自主选择,用户认证身份,同时可以辅助 UKey、口令卡等手段进行认证。由用户口令可以通过公钥加密标准(public-key cryptography standard,PKCS)(RFC3447)协议生成用户密钥, 该密钥只在每次用户成功登录后生成,会话结束即予以清除。因此,只有授权用户才能获取其对应的用户密钥。用户私钥由系统管理者分配,由基于密文策略的属性加密 (ciphertex-policy attributes based encryption, CP-ABE)机制生成,隐含式地包括了用户属性集合。文档密钥用来加密文档,每个文档对应一个密钥,该密钥又通过 CP-ABE 机制加密来实现访问控制。

3.3 双重应用模式

(1)网盘客户端

网盘客户端为用户提供网盘本地挂载。通过把网盘挂载到本地, 用户可以像操作自己的本地磁盘一样对网盘进行访问,与此同时,用户还可通过鼠标右键的下拉菜单实现对文件和文件夹的压缩/解压以及加解密。富客户端提供丰富的操作功能, 基于 SSH 和 RSYNC 实现加密传输和同步盘机制。瘦客户端提供基本的挂载功能,基于 SAMBA 实现,在 Windows 平台下,客户端的开发采用 Java 实现对以下脚本的封装(黑体部分用实际的用户名和密码替换)。

net use */del /y

net use W: \\hostIP**user** "**password**" /user:"**user**"

explorer W:

在 Linux 平台下,实现对以下脚本的封装。

mkdir /mnt/mynetdisk

mount-o username=**user** //localhost/**user** /mnt/mynetdisk

(2)Web 网盘

提供网盘的 Web 服务, 综合了邮件系统和基于 PGP (pretty good privacy)的文件加密存储:支持文件的上传与下载;支持文件目录的浏览;支持文件和文件夹的移动、重命名与删除;支持文件的在线编辑与打开;支持文件以邮件方式外发;支持文件和文件夹压缩;通过加解密插件,实现文件上传下载过程中的加解密。

4 实验分析

实验系统的 MFS 分布式文件系统采用 mfs-1.6.20 版,集群服务器为 VMware Sphere 统一分配的虚拟机,包括 1个主控节点(1.87 GHz 双核 CPU,4 GB 内存)、1 个日志服务器(1.87 GHz 单核 CPU,1 GB 内存)和 4 个存储节点服务器(1.87 GHz 单核 CPU,1 GB 内存,各自挂载 1~2 TB 的磁盘空间), 服务器操作系统为麒麟安全操作系统服务器版(Kylin Server 3.2.2),网络环境为吉比特光纤网络,用户通过无线接入网络,无线终端的平均带宽大于 1 Mbit/s,带宽上限可达到 54 Mbit/s。在此网络环境下,网盘的下载速度达到了 12 MB/s,上传速度达到了 6 MB/s。

4.1 用户压力测试

图 3 给出了在多个客户端并发连接的情况下主控节点的内存压力测试。当 1 000 个客户端并发连接时,大约 1.15 GB 的系统内存被占用; 当并发用户数达到 3 000 个时, 大约 2.1 GB 内存被占用, 每个用户占用的内存为 450~500 KB;系统支持的最大并发用户连接数为 7 000 个。

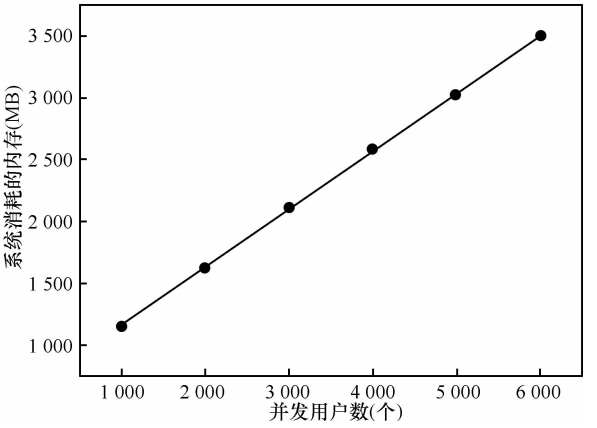

图 3 服务器内存消耗与并发用户数

4.2 MooseFS 集群的磁盘 I/O 性能

表 1 给出了在 MooseFS 集群各存储节点的磁盘运行时状态 (*chunk* 表示当前存储节点上数据块的数目,*fsync* 是块的同步时间)。各存储节点的读写速度均可达到每秒数百兆字节,不同节点之间数据块的最大同步时间是 124 ms,单个块的最大读写时间也在毫秒级。

表 1　存储节点的磁盘运行时状态

磁盘									
信息			I/O 性能统计					空间	
			迁移速度 (MB/s)		最大时间(ms)				
IP 路径	*chunk*	状态	读	写	读	写	*fsync*	已用(GB)	总大小(GB)
12.254.32.42:9422:/mnt/hd1/	8 965	OK	226	241	5.6	1.8	53	108	1.8
12.254.32.43:9422:/mnt/hd1/	11 229	OK	393	191	0.2	0.5	41	111	1.8
12.254.32.44:9422:/mnt/hd1/	10	OK	0	47	0	0.3	124	126	1.8
12.254.32.45:9422:/mnt/hd1/	8 072	OK	233	183	0.3	0.4	21	76	1.8

5　结束语

本文基于云存储技术为企业网和校园网设计了一种高性能、高可靠、安全、可扩展的网络硬盘。采用 MooseFS 分布式文件系统集群架构，实现了网盘存储系统的高性能、可靠性以及可扩展性；通过 SAMBA 用户权限控制、SSH 安全传输以及基于 PGP 的数据加密，实现了用户数据的安全存储与传输；最后，系统提供了 Web 网盘和客户端网盘双重模式，方便了用户的使用。

参考文献

1　武永卫. 清华云存储：从 Corsair 到 MeePo. Hadoop2011 云计算大会，北京，2011

2　侯清铧，武永卫，郑纬民等.一种保护云存储平台上用户数据私密性的方法. 计算机研究与发展，2011，48 (7)：1146~1154

3　林秦颖，桂小林，史德琴等.面向云存储的安全存储策略研究. 计算机研究与发展，2011，48(Z1)：240~243

4　黄汝维，桂小林，余思等.支持隐私保护的云存储框架设计.西安交通大学学报，2011，45(10)：1~6,12

5　洪澄，张敏，冯登国. AB-ACCS:一种云存储密文访问控制方法.计算机研究与发展，2010，47(Z1)：259~265

6　洪澄，张敏，冯登国.面向云存储的高效动态密文访问控制方法.通信学报，2011，32(7)：125~132

7　Yu S, Wang C, Ren K, *et al*. Achieving secure, scalable, and fine-grained data access control in cloud computing. Proceedings of IEEE INFOCOM'10, San Diego, USA, 2010

8　MooseFS. http://www.moosefs.org

System Architecture and Key Technology of Network Disk Based on Cloud Storage

Yang Yuexiang, Deng Wenping, Deng Jinsheng, Li Yang

(Information Center, National University of Defense Technology, Changsha 410073, China)

Abstract　Network disk for mobile office and storage in enterprise networks or campus networks has a broad market, however, the traditional technology has various limitations in performance, security and scalability. To address these issues, a solution based on cloud storage is proposed: by using the distributed file system MooseFS, all data servers are integrated into a cluster; in terms of security, user rights management is implemented by SAMBA, the storage supports 128 bit AES encryption, and the transmission link is protected by SSH; finally, the system provides both Web and client application modes for users. The experimental results reveal that the system has significant advantages in performance, security, scalability and other aspects.

Key words　network disk, cloud storage, distributed file system, cluster, user space　　(收稿日期：2012-06-29)