Digital Library Forum 数字图书馆论坛
http: // www.dlf.net.cn

MooseFS系统在图书馆联盟云计算架构中的应用研究*

□ 隋会民 刘万国 周秀霞 / 东北师范大学图书馆 长春 130024

摘要：文章通过对云存储基础设施以及MooseFS分布式文件存储系统的分析，提出了图书馆联盟数据云存储的科学解决方案。通过MooseFS系统将不同品牌、不同类型、不同容量的存储设备集合起来协同工作，共同对外提供数据存储和业务访问功能，服务系统不需要修改就可以使用大容量的存储空间，各种软件在整个云存储系统中协同工作，实现图书馆联盟用户的存储和访问。

关键词：图书馆联盟，云服务，云计算，云存储，分布式系统
DOI：10.3772/j.issn.1673—2286.2012.03.007

1 引言

随着各种新技术的风起云涌，信息数据总量以惊人的速度增长。据IDC2010年的研究表明，从2006年到2010年，全球信息数据总量将增长6倍以上，将从161EB增加到988EB。信息数据的光速增长，一方面对信息数据的存储、计算、提取等提出了严峻的考验，另一方面对信息数据的容灾系统、备份、归档的方案等提出了更严格的要求。在这种环境下，云存储技术应运而生。云存储是一个以数据存储和管理为核心的云计算系统，它是通过集群应用、网格技术或分布式文件系统等功能，将网络中各种不同类型的存储设备通过应用软件集合起来协同工作，共同对外提供数据存储和业务访问功能的一个系统。云存储系统的推出，成功地解决了海量数据存储和计算的挑战。作为一种应用技术，大量缩减了云服务中服务器的数量，降低了系统建设成本，减少了系统中由服务器造成的单点故障和性能瓶颈，精简了数据传输环节，提高了系统的性能和效率。

云存储的成功应用，也为图书馆联盟中海量数据的存储提供了科学的解决方案。图书馆联盟是图书馆联合资源建设、服务的最新形式，其主要目的是为了实现资源和服务的共建共享，因此，其管理平台上容纳了几十、几百、几千个图书馆的呈BT级或是呈ET级的各类数据信息，并同时提供给成千上万的用户并发使用。云存储的应用，可以彻底解决图书馆联盟资源利用中数据海量存储、总读取带宽要求较高、多个数据文件同时读取或写入、并发性能要求较高、长时间存放、利用成本低等一系列特殊性需求，为图书馆联盟海量数据存储提供了完善的解决方案。

2 云存储的基础设施

云存储系统能够改善存储利用率，增强性能和降低管理开销。它可以负载平衡并大规模提高无缝的利用率、改善了性能，同时也降低了管理开销。这一目标将通过分布式文件系统或NAS设备作为服务器节点中的一个文件系统而实现。每个节点都能够带来性能的提升，提供更多的可用容量。

* 文章系吉林省科技厅“建立国外开放获取科技期刊资源利用平台的对策研究”（项目号：20090630）、教育部基金项目《基于国家〈信息安全等级管理办法〉构建高校图书馆信息安全保障新体系》（项目编号：08JA870003）研究成果之一。

云存储中的存储设备数量庞大且分布在多个不同地域，如何实现不同厂商、不同型号甚至于不同类型（如FC存储和IP存储）的多台设备之间的逻辑卷管理、存储虚拟化管理和多链路冗余管理，是一个巨大的难题，这个问题得不到解决，存储设备就会是整个云存储系统的性能瓶颈，结构上也无法形成一个整体，而且还会带来后期容量和性能扩展难等问题。

在一个大容量的存储系统中，硬盘的成本占了很大的比重，为了保证数据读写的一致性，对硬盘的要求很高，必须是同容量、同品牌、同型号，以现在IT产业的高速发展来看，用户往往在使用2～3年后，发现硬盘坏掉，想要更新，已经没有原来采购的硬盘了，使用成本比原先估计要高得多。图书馆联盟的云存储在设计的过程中必须充分考虑这一点，任何硬盘都可以兼容，可以在同一个云存储甚至同一台存储节点使用品牌、容量、型号、介质、新旧完全不同的硬盘，而这些硬盘可以一起很好地协作，旧有的投资也不会浪费，硬盘坏掉，随便买一个插上即可。云存储对存储节点也没有什么限制，任何品牌的服务器都可以，用户可以根据成本、服务、合作关系等等，选择适合自己的服务器供货商，并且随时可以更改供货商，任何品牌的服务器都可以放在同一个云中。

3 底层存储解决方案

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。基础管理层通过集群、分布式文件系统和网格计算等技术，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。CDN内容分发系统、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，同时，通过各种数据备份、容灾技术和措施可以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。任何一个授权用户都可以通过标准的公用应用接口来登录云存储系统，享受云存储服务。云存储运营单位不同，云存储提供的访问类型和访问手段也不同。

采用分布式文件系统，可以支持大数量的节点以及PB级的数量存储。到目前为止，有数十种以上的分布式文件系统解决方案可供选择，如Lustre、Hadoop、MogileFS、FreeNAS、FastDFS、OpenAFS、MooseFS等。从应用情况来看，互联网上应用最多的是Hadoop、FastDFS、MooseFS，这里我们以MooseFS（以下简称MFS）分布式文件系统来作为图书馆联盟的云存储服务系统。

3.1 MFS系统的特点

- MFS的安装、部署、配置相对于其他几种系统来说，要简单和容易得多。
- MFS框架做好后，可以随时增加存储节点扩充容量，扩充和减少容量皆不会影响现有的服务。
- 除了MFS本身具备高可用特性外，手动恢复服务也是非常快捷的（只有元数据服务器节点存在单点故障）。
- 支持FUSE，目前的业务系统不需要修改就可以使用。

3.2 MFS系统的组成

MFS系统由以下四部分组成：

（1）元数据服务器master。在整个体系中负责管理文件系统，目前MFS只支持一个元数据服务器，需要一个运行稳定的服务器来充当。元数据服务器（master）是MooseFS部署中一个重要的元素。在硬件方面，应该被安装在一台能够保证高可靠性和能胜任整个系统存取要求的机器上。一个明智的做法是用一个配有冗余电源、ECC内存、磁盘阵列如RAID1/RAID5/RAID10的服务器。在操作系统方面，管理服务器的操作系统应该是具有POSIX兼容的系统。

（2）元数据日志服务器。负责备份master服务器变化的日志文件，以便于在master出问题的时候接替其进行工作。元数据日志守护进程是在安装master server时一同安装的，最小的要求并不比master本身大，可以被运行在任何机器上，但是最好是放置在MooseFS master的备份机上，备份master服务器的变化日志文件，文件类型为changelog_ml.*.mfs。因为主要的master server一旦失效，可能就会将这台metalogger机器取代而作为master server。

（3）数据存储服务器chunk server。真正存储用户数据的服务器。数据服务器一般是多个，数量越多，磁盘空间越大，性能越高，可靠性也越高。这些机器的磁盘上要有适当的剩余空间，而且操作系统要遵循POSIX标准（比如Linux、FreeBSD）。Chunkserver

在一个普通的文件系统上储存数据块/碎片（chunks/fragments）作为文件。

（4）客户端。通过fuse内核接口挂接远程管理服务器上所管理的数据存储服务器，使共享的文件系统和本地unix文件系统使用一样的效果。该服务器上安装的是对外提供服务的系统程序，也就是云存储业务系统的发布节点。可以有多个客户端同时存在，提高系统的可靠性。

3.3 MFS系统的部署方法

MFS部署的首选方法是从源代码安装。源代码包安装支持标准./configure && make && make install的步骤，官方网站上有详细的文档，这里不再赘述。

3.4 MFS系统的备份与恢复

一旦MooseFS master崩溃，必须及时恢复。为了从备份中恢复一个master，需要做如下工作：

（1）安装一个mfsmaster。

（2）利用同样的配置来配置这台mfsmaster（利用备份来找回mfsmaster.cfg），可见配置文件也是需要备份的。

（3）找回metadata.mfs.back文件，可以从备份中找，也可以从metalogger主机中找（如果启动了metalogger服务），然后把metadata.mfs.back放入data目录，一般为${prefix}/var/mfs。

（4）从在master宕掉之前的任何运行metalogger服务的服务器上拷贝最后metadata文件，然后放入mfsmaster的数据目录。

（5）利用mfsmetarestore命令合并元数据changelogs，可以用自动恢复模式mfsmetarestore -a，也可以利用非自动化恢复模式，语法如下：

mfsmetarestore -m metadata.mfs.back -o metadata.mfs changelog_ml.*.mfs

4 前端管理及用户服务

后台的大规模存储必须与性能优越的业务管理平台相结合，才能集中展现云存储系统的优越性能。就目前而言，一般的图书馆联盟都搭建有一个比较成熟的业务管理平台，用以管理、揭示联盟中存在的海量资源。如吉林省图书馆联盟就使用了Exlibris公司的Primo平台来进行资源的管理与集成。Primo是图书馆统一资源发现与获取门户系统，是目前正在构造的下一代数字图书馆服务系统平台中的两大核心之一，它可以帮助图书馆为用户提供统一资源的发现与获取服务，内容包括图书馆自身的物理、数字馆藏，以及图书馆订购的各类远程数据库、电子资源。所有的发现与获取服务均基于Web 2.0标准构造，并结合SFX开放链接服务系统，完全可以胜任下一代的数字图书馆用户服务门户功能。该平台的存储分为两大部分：Primo Central和本地索引库。Primo Central托管存放在Amazon，由公司负责管理和更新；本地索引库由用户负责管理，库的大小完全由数据量决定，Primo可以导入目前所有的元数据，使用MFS来管理它的数据是一个非常好的选择。

5 展望

目前，大多数用户都是通过ADSL、DDN等宽带接入设备的方式来连接云存储，因此，云存储用户需要使用宽带网络与存储系统进行连接，只有宽带网络得到充足的发展，使用者才能获得足够大的数据传输带宽，实现大容量、高速度数据传输，享受到云存储带来的便利。随着各种网络技术和云计算技术的快速发展，云存储系统必将成为多区域分布，遍布全国甚至是全球的庞大系统，接入方式也不仅限于ADSL、DDN等，而将拓展到更为广阔的各种有线、无线终端。对海量存储的可管理性、性能扩展性、可靠性、总体拥有成本低以及拥有无限的容量扩展性等将会成为云存储系统发展的技术衡量指标，自存储、自恢复、自优化、自管理将成为云存储系统的核心需求。云存储系统必将成为海量数据的主流存储模式。

参考文献

[1] MooseFS官方网站手册[OL]. [2012-01-12]. http://www.moosefs.org/reference-guide.html.
[2] 云存储_百度百科[OL]. [2012-01-12]. http://baike.baidu.com/view/2044736.htm.
[3] 雷万云.云计算：企业信息化建设策略与实践[M].北京:清华大学出版社,2010:45-105.
[4] 田逸.互联网运营智慧——高可用可扩展网站技术实战[M].北京:清华大学出版社,2011:245-282.
[5] (美)里特豪斯.云计算：实现、管理与安全[M].北京:机械工业出版社,2010:105-123.

作者简介

隋会民，男，东北师范大学图书馆副研究馆员。E-mail: suihm@nenu.edu.cn
刘万国，男，东北师范大学图书馆研究馆员。E-mail:liuwg@nenu.edu.cn
周秀霞，女，东北师范大学图书馆馆员。

Application and Research of MooseFS System in Library Alliance Cloud Computing Architecture

Sui Huimin, Liu Wanguo, Zhou Xiuxia / Northeast Normal University Library, Changchun, 130024

Abstract: Through the cloud storage infrastructure and MooseFS distributed file storage system analysis, this article puts forward a scientific solution of the library consortium data cloud storage. Through the MooseFS system, it will put different brands, types and capacity of the storage devices together to work cooperatively, and provide data storage and business visiting function. Service system will use the large capacity of storage space without modifying. All sorts of software can achieve for the library consortium users' storage and access work cooperatively in the whole cloud storage system.

Keywords: Library consortia, Cloud service, Cloud computing, Cloud storage, Distributing system

（收稿日期：2012-02-17）

业界动态

大英百科全书停印纸质版 只发行电子版

据新华社电，总部位于芝加哥的不列颠百科全书公司13日宣布，将停印已有244年历史的纸质版《不列颠百科全书》（又称《大英百科全书》），今后将只提供电子版。

公司方面说，终结纸质版的想法已有一段时间。美国媒体援引公司总裁豪尔赫·考斯的话说，做出这一决定是因为电子版《不列颠百科全书》拥有更大消费群体。此外，电子版的百科全书可以实时更新，而印刷版本在印制时就可能已经过时。

纸质版《不列颠百科全书》的销量早已今非昔比。考斯说，《不列颠百科全书》的销量最高峰是1990年，卖出12万套。但到1996年，这一数字下降至4万套。首个光盘版《不列颠百科全书》发行于1989年，网络版始于1994年，现有全球逾1亿用户。

《不列颠百科全书》网站标出的剩余库存精装版《不列颠百科全书》目前售价为每套1395美元（约合8835元人民币）。作为终结纸质版的纪念，《不列颠百科全书》网站内容自13日起供网民免费浏览一周。

（来源：http://auto.cnradio.com.cn/lishipindao/shouyetuijian2/201203/t20120316_509297382.shtml，查询日期：2012-03-17）