云存储架构与关键技术的研究

**曾文英**1,2**，赵跃龙**1**，宋 威**1,3

[wyzeng@126.com](mailto:wyzeng@126.com) [ylzhao1@scut.edu.cn](mailto:ylzhao1@scut.edu.cn) [song.wei@scut.edu.cn](mailto:song.wei@scut.edu.cn)

（1.华南理工大学计算机科学与工程学院，中国广州 510640；2.广东科学技术职业学院计算机工程与技术学院，中国珠海 519090；3.广东工业大学计算机学院，中国广州 510640）

**摘 要：**本文提出了云存储系统的通用架构，分析了各个模块的功能，并讨论了其中的关键技术，等。云存储是一种新型的存储服务模式，它由服务提供方通过互联网向客户提供存储空间和数据存储服务，同时客户不需要了解其中的细节、底层结构和机制。本文推荐的云存储架构是层次化的、相互协作的，且讨论的关键技术包括部署、存储虚拟化、数据组织、转移和安全性等。运行机制包括生态链、博弈论、蚁群优化、数据生命周期管理、维护和更新，同时也分析了收敛和演变机制。所以本文从总体上阐明了云存储系统，并加入了新的观点。

**学科分类：**操作系统；存储管理；分配与收集策略；辅助存储；存储层次；虚拟内存

**一般术语：**算法；管理；度量；文档；性能；设计；经济学；可靠性；实验；安全；人为因素；标准化；语言；理论；法律方面；验证

**关键词：**云存储架构；关键技术；运行机制；生态链；博弈论；蚁群优化

**Keywords**

Cloud Storage Architecture, Key Technologies, Operation Mechanism, Ecology Chain, Game Theory, Ant Colony Optimization.

1. **概述**

云计算是一种新兴的计算平台与服务模式，它基于互联网组织并安排服务。云存储就是服务之一，它使用基于云计算的远程存储服务器来提供存储资源和服务。云存储可以提供更为廉价且高可靠性和安全性的存储服务。

云存储系统是一种协作型存储服务系统，它拥有多台设备，多个应用域和多种服务形式。云存储的发展利益于网络带宽、Web 2.0，存储虚拟化，存储网络，应用存储与服务器及存储设备的集成，集群技术，网格计算，分布式文件系统，内容分发网络，对等网络，数据压缩，数据加密等技术的发展。

本文后面的内容按如下安排：第2节介绍相关工作；第3节提出云存储的通用架构；第4节讲述构建策略；第5节讨论云存储服务器的关键技术；第6节分析云存储系统的运行机制；第7节做个总结。

1. **相关工作**

现在有许多云计算和云存储提供商，如国际商业机器（IBM），谷歌（Google），太阳微系统公司（Sun Microsystems），微软（Microsoft），亚马逊（Amazon），EMC，NetApp，惠普（HP），Nirvanix，日立数据系统（HDS），赛门铁克（Symantec）等。

另外也有越来越多的云存储平台，如，HDFS，GFS，Sun Network.com，SkyDrive，Amazon S3，EMC Atoms，Hitachi Content Platform，FileStore，和KFS等。

存储网络行业协会（SNIA）提议云存储首创方（CSI）将云数据管理接口（CDMI）标准采纳为云服务标准。

Yunhong Gu等人拟议的文献[1] 允许用户像使用本地磁盘文件一样地与建立在大规模分布式节点上的大量数据集打交道。用户不必在多个节点中去定位数据，管理数据，也不必向系统中添加新的节点或删除已有节点。

文献[2] 介绍了MetaCDN，一个利用“存储云”资源来创建能为内容创建人员提供廉价且高性能的内容分发网络（CDN）的集成重叠网的系统。MetaCDN可以通过依据服务质量、覆盖范围及预算以智能地匹配和替换用户的内容到一个或多个存储供应商来去除处理多个存储供应商的复杂性。MetaCDN提供一个单独统一的命名空间使得内容创建人员及消费者驾驭多个“存储云”的性能及覆盖范围工作变得琐碎，统一命名空间也使得它更容易集成到原始网站，且对终端用户来说更为透明。

文献[3] 提出了一种在广域网之上及时存储转移机制，在存储的分布式转移方面具有参考价值。

云计算与云存储的飞速发展将产生云资源市场，并带来云服务选择的挑战，文献[4] 提出过一些这个方面相关的算法。Ying Zhan等[5] 提出的云存储管理技术也是个值得关注的紧要问题。

Albert Greenberg 等人[9] 讨论了数据中心的云服务的费用，包括服务器（45%），基础设施（25%），电源（15%）和网络（15%）。所有的资源要根据用途从资源池中动态地分配出来以适应需求和费用。为了减少花费且提高灵活性，所采取的策略有位置无关寻址，统一带宽和延迟和安全与性能隔离等。用于资源消耗塑型的市场机制被采纳来提高效率，同时地理多样化数据中心也被用来提高点对点性能和可靠性。这些想法也可做为云存储的参考。

在2009年的存储峰会[10] 中，许多人认为关于云的主题已经引起了众多信息技术的教授，开发人员，市场人员，出版社和分析专家们的注意，诸如云计算，云服务或者云存储，这里仅仅列出不同变体。

Cloud storage enables new application types [11] through SOA, Web services APIs and unified service interface via virtualization over a network at low cost, and can provide anytime and anywhere access, massive data storing, sharing and collaboration via a single namespace, and policy management of storage, etc.

There is not yet a global standard specification and general architecture to cloud computing and cloud storage. The paper will analyze the requirements of cloud storage, propose the architecture, and analyze the key technologies.

The main idea is to integrated and improvement the current architecture, distribution mode, application area, etc. to construct a low cost, fault torrent, reliable, scalable, high performance and fair cloud storage alliance system.

The creativity in cloud storage is to impose the concept of ecology chain, service market, game theory, ant colony optimization and discusses some key technologies.

1. **云存储架构**
   1. **需求分析**
   2. **云存储架构**
   3. **模块功能**
2. **构建策略**
3. **云存储服务的关键技术**
   1. **云存储部署**
      1. *需求与基础*
      2. *存储资源重定向*
      3. *优化与演变*
   2. **云存储虚拟化与可用性**
   3. **云存储数据组织**
   4. **数据转移与负载均衡**
   5. **数据去冗余**
   6. **存储安全**
4. **运行机制**
   1. **云存储生态链**
   2. **云存储中的博弈论**
   3. **蚁群优化**
   4. **数据生命周期管理**
   5. **维护与更新机制**
   6. **存储收敛与演变**
5. **总结**
6. **致谢**
7. **参考文献**
8. Yunhong Gu, Robert L. 2009. Grossman. Sector: A high performance wide area community data storage and sharing system. Future Generation Computer Systems, 20 May 2009.
9. James Broberg, Rajkumar Buyya, Zahir Tari. 2009. MetaCDN: Harnessing ‘Storage Clouds’ for high performance content delivery. Journal of Network and Computer Applications 32 (2009), 1012–1022.
10. Takahiro Hirofuchi, Hidemoto Nakada, Hirotaka Ogawa, Satoshi Itoh, Satoshi Sekiguchi. 2009. A live storage migration mechanism over wan and its performance evaluation. Proceedings of the 3rd international workshop on Virtualization technologies indistributed computing, Barcelona, Spain, 2009, 67-74.
11. Wenying Zeng, Yuelong Zhao, Junwei Zeng. 2009. Cloud service and service selection algorithm research.GEC '09: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation, Shanghai, China, June 2009, 1045-1048.
12. Ying Zhan, Yong Sun. 2009. Cloud Storage Management Technology. Second International Conference on Information and Computing Science. Manchester, England, UK, May 21-May 22, 2009, icic, vol. 1, 309-311.
13. Henry Newman. 2009. Why people don’t like to use cloud storage? http://www.cnw.com.cn/storage-Technology/htm2009/20091013\_183980\_2.shtml, 2009-10-13.
14. FalconStor Software, Inc. 2009. Demystifying Data Reduplication: Choosing the Best Solution. http://www.ipexpo.co.uk/content/download/20646/353747/file/DemystifyingDataDedupe\_WP.pdf, White Paper, 2009-10-14, 1-4.
15. Mark W. Storer Kevin Greenan Darrell D. E. Long Ethan L. Miller. 2008. Secure Data Deduplication. StorageSS’08, October 31, 2008, Fairfax, Virginia, USA. 2008, 1-10.
16. Albert Greenberg, James Hamilton, David A. Maltz, Parveen Patel. 2009. The Cost of a Cloud: Research Problems in Data Center Networks. ACM SIGCOMM Computer Communication Review, Volume 39, Number 1, January 2009:68-73.
17. SNIA CLOUD Storage Summit. 2009. http://www.snia.org/events/wintersymp2009/cloud/, Held at the WINTER SYMPO SIUM 2009.
18. Steve Lesem. 2009. Cloud Storage and The Innovator's Dilemma. http://cloudstoragestrategy.com/cloud-ecosystem/, July 19, 2009.
19. Soft Layer Technologies. 2009. CloudLayer™ Storage. http://softlayer.com/cloudlayer\_storage.html, 2009-10-15.
20. Sun Microsystems, Inc. 2009. Introduction to Cloud Computing architecture. http://www.sun.com/featured-articles/CloudComputing.pdf, White Paper, 1st Edition, June 2009:1-32.