

基于 Ceph 对象存储的云网盘设计与实现

杨 飞¹, 朱志祥², 梁小江³

(1. 西安邮电大学 计算机学院, 陕西 西安 710061; 2. 陕西省信息化工程研究院, 陕西 西安 710061;

3. 未来国际信息股份有限公司 研发部, 陕西 西安 710065)

摘 要 针对基于传统 NAS 存储实现云网盘存在扩展性差, 响应性能低、数据不可靠性、安全性低等问题, 提出一种基于 Ceph 分布式对象存储系统构建高性能、高可靠、可扩展的云网盘设计与实现方案, 底层基于 X86 服务器作为 Ceph 存储集群池, 通过设计多对象网关发布对象存储服务, 采用软负载均衡, 提升云网盘请求响应吞吐能力, 并集成基于 Token 认证系统实现对象数据的隔离与安全。经实验测试表明, 该设计实现方案可满足现实应用需求。

关键词 Ceph; 对象存储; Token; 云网盘

中图分类号 TP302.1 文献标识码 A 文章编号 1007-7820(2015)10-096-04

Design and Implementation of Cloud Disk Based on Ceph Object Storage Cluster

YANG Fei¹, ZHU Zhixiang², LIANG Xiaojiang³

(1. School of Computer Science, Xi'an University of Posts and Telecommunications, Xi'an 710061, China;

2. Institute of Communication Technology, Xi'an 710061, China;

3. R & D Department, Future International Information Co., Ltd., Xi'an 710065, China)

Abstract Traditional cloud disks based on NAS storage have the disadvantages of poor scalability, low response performance, unreliable data and low security. This paper presents a design and implementation scheme based on the distributed object storage system of Ceph. The scheme can build high performance, high reliability and scalable cloud disks. The pools of Ceph storage clusters are based on the X86 server, in order to the service of object storage is published through multi-object gateway with soft load balancing to enhance the capacity of request and response in cloud disk and achieve the functions of isolation and security in object data by integrating with token-based authentication systems. Experiments show that the design and implementation scheme meets the requirements of practical application.

Keywords Ceph; object storage; Token; cloud disk

随着云计算和大数据应用的快速增长,网络数据呈海量的增长方式,人们对数据存储能力、计算能力的需求呈爆炸式增长,这对存储系统的容量、可扩展性、数据可靠性以及安全性提出更高的要求,这种需求正在影响着数据存储基础设施的发展。

传统的 NAS 存储系统,存在着存储性能的局限性、可靠性低、扩容成本高、易形成单点故障等问题,而 Ceph 分布式存储系统是集高性能、高可靠性和可扩展性为一身的存储系统,能有效避免传统存储系统的缺点^[1]。

本文以 Ceph 分布式系统为研究对象,设计基于 Ceph 对象存储的健康集群,通过设计多个对象网关实现对应用请求数据的存储和管理,使用负载均衡提高

云网盘的请求、响应和吞吐能力,然后设计统一的 Token 认证系统,实现对象存储集群的数据安全性和隔离性,最终设计和实现了基于 Ceph 对象存储集群的云网盘。

1 系统设计与实现

Ceph 是一种具有高性能、高扩展性、高可靠性的分布式存储系统。Ceph 可为一套存储系统同时提供对象存储、块存储和文件系统存储 3 种功能,以便在满足不同应用需求的前提下简化部署和运维。在实践中, Ceph 可被部署于上千台服务器上^[2]。

首先设计基于 Ceph 对象集群的简单集群系统,保证整个 Ceph 对象存储集群的可用性,同时 Ceph 存储集群也是整个云网盘的存储系统。

整个 Ceph 存储集群拥有 4 个 Gateway 节点,分别为 EAST、WEST、SOUTH、NORTH。这些节点能对外提供和

收稿日期: 2015-03-13

作者简介: 杨飞(1989—), 男, 硕士研究生。研究方向: 云计算, 大数据等。E-mail: yangfeigogo@sina.com

发布对象存储服务,实现基于 Ceph 对象存储的多区域网络访问,所有的应用请求必须通过这 4 个 Gateway 节点实现。根据实际网络用户的数量和分布情况,动态的调整 Ceph 对象存储集群的架构,满足不同区域用户的动态请求服务^[3]。

基于 Token 的统一认证系统,保证整个 Ceph 云存储集群的安全性,同时对不同用户进行隔离,保证整个 Ceph 对象存储集群的高可靠性。

最后设计和实现基于 Ceph 对象存储集群的云网盘。通过大量网络应用测试和分析,本系统实现了基于 Ceph 对象存储集群的云网盘基础功能^[3]。

1.1 总体设计框架

用户通过网络进行存储应用访问,首先要通过统一的 Token 认证系统,区分不同用户的权限和工作区域,实现数据管理和数据操作的安全性和隔离性^[4]。

根据用户的分布和数量,设计合理的 Gateway 区域,实现对不同用户请求的对象存储服务,满足用户访问请求的实时性和可用性。

本系统设计出 4 个不同的区域,分别为 EAST 区、WEST 区、SOUTH 区、NORTH 区,其能提供不同区域的对象存储服务。整体设计框架如图 1 所示。

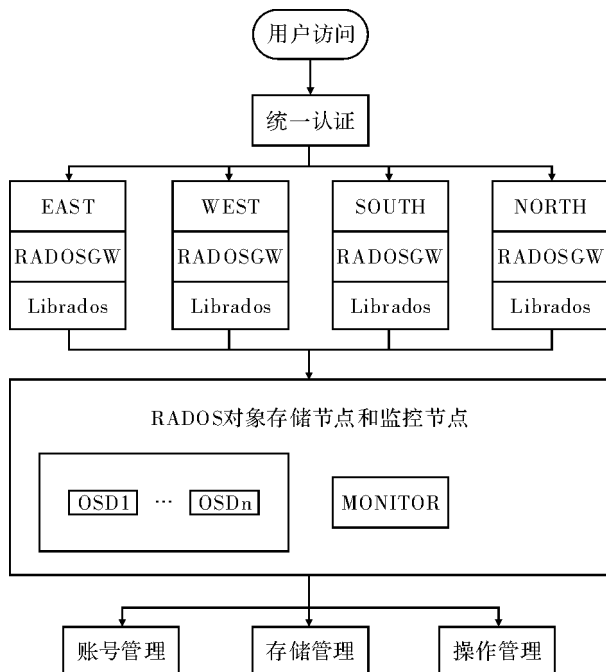


图 1 整体设计框架

1.2 多区域的统一认证

在整个 Ceph 存储集群中,大量网络用户通过用户访问节点进行统一的认证,在统一认证系统中安装和配置 Keystone 服务,产生统一的 Token 认证码,在 4 个 Gateway 节点上均使用这一 Token 认证码,这样保证了用户请求能在 4 个不同的节点上对 Ceph 存储集群进

行合法访问^[5]。

在基于 Ceph 对象存储集群中的 4 个不同区域节点,均需要安装配置 MySQL 和 Keystone,首先在 MySQL 中创建 Keystone 数据库,然后在 Keystone 的配置文件中,将数据库修改为 MySQL 并设置用户名和密码,这样就完成了 Keystone 的单点认证服务。

在 Ceph 集群的配置文件中,分别修改 4 个不同区域节点的 Gateway 配置文件,将其中 keystone_url 选项中的 IP 修改为用户访问节点的 IP,这样用户访问节点每产生一个 Token。在 Ceph 对象存储集群中的 4 个不同 Gateway 区域节点中,当有对象存储服务请求时,不同的区域节点能使用统一的 Token 验证码对用户访问节点进行数据的操作和数据的管理^[6]。如图 2 为统一认证框架。

在整个 Ceph 对象存储集群中,通过 Token 统一认证系统,能保证用户请求服务的安全性和隔离性,因为只有获取到 Token 才能实现对云网盘的访问和控制。

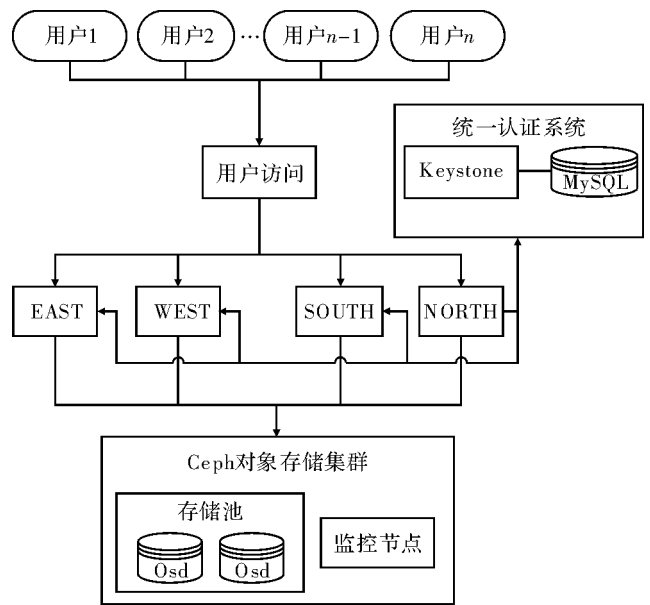


图 2 统一认证框架

1.3 多区域数据操作

在整个 Ceph 对象存储集群中有 4 个不同的区域节点对外提供服务接口,通过 Token 统一认证系统,只有合法用户才能对云网盘进行基本操作。从而实现不同区域用户对整个 Ceph 对象存储集群的数据访问和数据操作。对多区域的数据操作包括文件上传、下载、复制、删除、共享、权限修改等操作。

通过使用软均衡负载设置,实现 4 个不同的 Gateway 节点对数据的轮询访问,由此便提高了云网盘的请求、响应和吞吐能力,有效降低了用户访问节点的压力,从而提高整个 Ceph 对象存储集群的可靠性^[7]。

2 系统测试

2.1 Keystone 统一认证

在用户访问节点上安装和配置 Keystone,并初始化 Keystone 数据库,重新启动 Keystone 服务,通过设置和添加环境变量,最后执行 Keystone 的初始化脚本。这样在用户访问节点上就能够产生统一的 Token 认证码,实现云网盘的统一化的认证和管理。

在 EAST、WEST、SOUTH、NORTH 这 4 个不同的区域节点,均配置和安装 Keystone 和 MySQL 服务,首先设置 MySQL 的用户名和密码,允许客户端进行访问,在 MySQL 中创建 Keystone 数据库,然后在 Keystone 配置文件中修改数据库指定为 MySQL,并初始化数据库,然后重启。这样在 4 个不同的区域节点上就能产

生 Token 认证码^[8]。

在 4 个区域节点的 Ceph 配置文件中,分别修改 Keystone的地址为用户访问节点的地址,则当用户访问节点产生统一的 Token 认证码后,在 4 个区域节点上,可使用该 Token 认证码对云网盘进行数据操作和数据管理。

2.2 云网盘数据操作

在用户访问节点上设置软均衡负载,所有的网络访问用户通过该节点实现对整个 Ceph 对象存储集群的轮询访问。

在用户访问节点的均衡负载配置文件中添加 4 个不同区域的监控节点信息。当有用户请求服务时,通过数据监控界面可查看当前的数据操作记录^[9]。

通过网络对 Ceph 对象存储进行监控,查看不同区域的数据的监控界面,如图 3 所示。

DNS Server	Queue			Session rate			Sessions					Bytes	
	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LbTot	In	Out
EAST	6	0	-	0	10		0	0	-	7	7	2 761	2 785
WEST	5	0	-	0	8		0	0	-	8	8	1 502	2 668
SOUTH	5	0	-	0	8		0	0	-	9	9	3 261	3 358
NORTH	5	0	-	0	8		0	0	-	8	8	2 649	2 964

图 3 数据监控界面

基于 Ceph 对象存储集群的云网盘可实现基本的数据操作和管理功能,如图 5 所示。



图 4 云网盘界面

2.3 云网盘性能测试

云网盘的服务请求入口的 IP 为 http://10.10.10.5:8090/yunpan,使用该 IP 作为测试的 URL,使用 Apache Bench 软件设置并行连接数和请求数,通过实验测试云网盘的性能。

如表 1 所示,请求数为 1 000,并行连接数分别为 10、100、1 000。

请求数	连接数	时间/s	速率/Mbit · s ⁻¹
1 000	10	0.09	1.54
1 000	100	0.07	1.98
1 000	1 000	0.44	0.32

如表 2 所示,请求数为 10 000,并行连接数分别为 10、100、1 000。

请求数	连接数	时间/s	速率/Mbit · s ⁻¹
10 000	10	0.81	1.73
10 000	100	0.63	2.24
10 000	1 000	0.75	1.86

如表 3 所示,请求数为 100 000,并行连接数分别为 10、100、1 000。

请求数	连接数	时间/s	速率/Mbit · s ⁻¹
100 000	10	8.53	1.73
100 000	100	6.93	2.13
100 000	1 000	7.81	1.89

如表 4 所示,请求数为 1 000 000,并行连接数分别为 10、100、1 000。

请求数	连接数	时间/s	速率/Mbit · s ⁻¹
1 000 000	10	86.67	1.83
1 000 000	100	66.36	2.39
1 000 000	1 000	76.62	2.07

经过对上述测试结果的对比分析,得出总体的测试对比图,如图 5 所示。

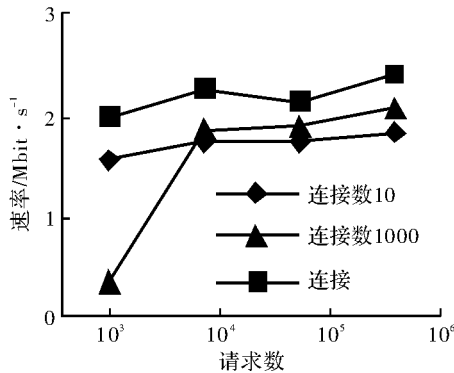


图 5 测试对比图

3 结束语

设计和部署 Ceph 存储集群,根据实际生产需求设计多对象区域架构,实现多对象存储服务。采用软负载均衡,提升云网盘请求、响应吞吐能力,并集成基于 Token 认证系统实现对象数据的隔离与安全。通过测试,本系统能实现基于 Ceph 对象存储集群的云网盘数据操作和数据管理功能,同时保证了整个云网盘的良好性能。

(上接第 95 页)



图 5 系统设置

由实际压力曲线和目标压力曲线的比较可知:无论增加水压还是减少水压、系统调整时间短、响应速度较快、超调量微小、抗干扰能力强、运行稳定。

4 结束语

传统的供水系统多采用 PLC 或单片机作为控制器,不能较好地实现逻辑复杂的控制程序,开发周期长,消耗人力和物力大。该供水系统利用 Sedona 平台采用变频调速运行方式,根据实际用水水压的变化实时自动调节水泵电机的转速来加减泵,实现恒压供水,节能降耗。根据工作时间合理选择工作泵大小,提高整个泵组的使用寿命。人机界面做在组态屏里,方便改变参数,整套系统经过多种环境测试,运行情况良好、性能优越、操作简单。

参考文献

- [1] 李翔. Ceph 分布式文件系统的研究及性能测试 [D]. 西安: 西安电子科技大学, 2014.
- [2] 段剑弓. 存储系统 NAS 和 SAN 的差异和统一 [J]. 计算机应用研究, 2004, 21(12): 94-97.
- [3] Weil S A, Brandt S A, Miller E L, et al. Ceph: A scalable, high-performance distributed file system [C]. Berlin: Proceedings of the 7th Symposium on Operating Systems Design and Implementation (OSDI), 2006: 307-320.
- [4] 肖培棕. 分布式文件系统元数据负载均衡技术研究与实现 [D]. 合肥: 中国科学技术大学, 2009.
- [5] Gritzalis S, Katsikas S K, Lekkas D, et al. Securing the electronic market: the KEYSTONE public key infrastructure architecture [J]. Computers & Security, 2000(19): 731-746.
- [6] 蔡官明. 开放式云存储服务设计及其移动云盘应用开发 [D]. 广州: 华南理工大学, 2013.
- [7] Cheswick B. The design of a secure internet gateway In Proc [C]. Salt Lake: Summer USENIX Conference, 1990: 233-237.
- [8] 覃灵军, 冯丹, 曾令仿, 等. 基于对象存储系统的动态负载均衡算法 [J]. 计算机科学, 2006, 33(5): 88-91.
- [9] Holmquist L E, Redström J, Ljungstrand P. Token-based access to digital information proceeding [M]. Berlin: Springer-Verlag, 2000.
- [10] 王芳, 陈亮. 对象存储系统中基于负载均衡的设备选择算法 [J]. 华中科技大学学报: 自然科学版, 2007, 35(10): 46-49.

参考文献

- [1] 方大寿, 李纪扣. 模糊-PI 复合控制恒压供水系统 [J]. 微小型计算机开发与应用, 1996, (4): 32-35.
- [2] 王兴, 齐向东, 郭建国. 智能供水系统设计及应用 [J]. 计算机测量与控制, 2007, 15(5): 624-627.
- [3] 孙小权等. 基于 PLC 的变频恒压供水控制系统优化设计 [J]. 自动化仪表, 2007, 28(1): 47-51.
- [4] 杨美芬. 自动恒压供水控制系统的设计 [J]. 云南民族大学学报: 自然科学版, 2005, 14(3): 260-264.
- [5] 肖进. 基于 Niagara 的建筑智能化系统集成的设计与开发 [D]. 济南: 山东大学, 2011.
- [6] 吴再群. 基于单片机的水位控制系统 [J]. 电子科技, 2013, 26(6): 80-81.
- [7] 姚福强. 基于 PLC 和变频器串行通讯的变频恒压供水系统 [J]. 微计算机信息, 2004, (20): 9-10.
- [8] 林俊赞. PLC 在恒压供水控制系统中的应用 [J]. 电机电器技术, 1999(3): 45-48.
- [9] 尚志红, 刘斌. 变频调速恒压供水的自动控制系统 [J]. 信息科技, 2007(10): 56.
- [10] 谭远博, 黄珊珊, 雷红玲. 基于专家知识的恒压供水控制算法研究 [J]. 现代电子技术, 2009(15): 178-180.