

文章编号:1674-2869(2016)06-0611-05

一种数据存储系统的研究

杨烈君,吴玉芹,陈 杨

宁德师范学院计算机系,福建 宁德 352100

摘 要:提出了一种新的数据存储系统——软件定义存储系统(SDS)的改进实施方案.基于一种虚拟化的存储方式可以自动分配存储、安全、网络等资源,并将这些资源池化,从而实现数据保护、复制、压缩等一系列存储功能.与传统的DAS和SAN系统的存储模式的存储性能进行了比较,结果表明该系统对于小容量的数据存储,存储速率略高其他存储方式;对大容量数据存储性能与吞吐量优化效果明显,当数据块容量达5 Tbit时,其存储速率超过SAN存储系统约3.9倍.

关键词:数据存储;软件定义存储;存储性能

中图分类号:TM344.1 **文献标识码:**A **doi:**10.3969/j.issn.1674-2869.2016.06.018

Study on a Data Storage System

YANG Liejun, WU Yuqin, CHEN Yang

Computer Department, Ningde Normal University, Ningde 352100, China

Abstract: This paper proposes a new data storage system named an improved implementation of Software Definition Storage System (SDS), which can automatically allocate storage, security, network and other resources on basis of a virtualized storage mode. The system can also pool these resources to achieve data protection, replication, compression and other storage functions. By comparing storage performance of SDS with traditional Direct-Attached Storage and Storage Area Network (SAN) storage models, the storage rate of SDS is slightly higher than that of other storage systems for small-capacity data storage, and the SDS has obvious advantages in the storage performance and throughput optimization for large-capacity data storage. The storage rate of SDS is 3.9 times as much as that of SAN when the capacity amounts to 5 Tbits.

Keywords: data storage; software defined storage; storage performance

1 引 言

云计算目前受到业界越来越多的关注,他是一种新兴的计算模式.在云存储由概念变成了具有巨大商业价值的实际应用后其存储的安全性问题却一直没有得到有效的解决,好的数据存储方式不仅可以提高存储性能,更重要的是它也应该能保障数据存储的安全.如亚马逊是国外著名的云服务提供商,把安全性看成云存储服务中重要的一部分,他们根据安全性制定了一系列安全

方案,其中提出的最典型的云存储应用是S3(Simple Storage Service),S3特别强调存储的安全机制,用户可以根据自己的意愿选择谁访问他的数据,用户可以自由控制这一操作,选择什么时候给予访问,访问什么类型的数据等,这些都在用户的控制之内^[1].中国电信云计算公司作为国内首家具有服务级别的云计算公司也提出“宽带入户,云服务到家;终端在手,云服务随身”的口号,这样一个理念可以看出云服务占主要位置,但他们在提供云服务时都面临数据安全问题的困扰.

收稿日期:2016-10-10

基金项目:福建省自然科学基金项目(2015J01660);福建省中青年教师教育科研项目(JAT160538);宁德师范学院服务宁德项目(2013F30)

作者简介:杨烈君,硕士,讲师. E-mail:slocky@126.com

云存储受到的安全威胁可能来自访问控制,也可能来自数据传输、安全漏洞和数据所存在的物理位置等方方面面,所以如何保证各方面的安全性,是云存储现在首要完成的任务.在一项研究中表明,数据安全是所有安全问题的核心,数据安全也是用户最关注的焦点.不难理解,当今是数据的时代,拥有数据就相当于拥有一切,对数据的保护尤为重要,对数据的泄漏和恶意修改将给企业带来巨大损失.由此可见数据安全性也将是云存储技术一直而且以后长期要关注的课题之一,而在大数据时代,传统存储模式已经不能适应现在的潮流满足数据安全的需求,因此软件定义存储(Software Defined Storage,简称 SDS)应用而生,软件定义存储(SDS)以其灵活性和高扩展性被云存储服务提供商所重视并凭借其所具有的巨大优势将逐渐超越传统的存储模式^[2].

2 传统存储模式的弊端

传统的存储架构如磁盘、DAS、SAN 模式是现在大多数厂家提供的存储模式,是基数最大的存储模式^[3].可是随着时代的发展,数据的几何式增长,传统的存储模式因为存储容量的不足而饱受诟病.一方面,数据是要永久保存的,将一直占有空间;另一方面,数据是持续增加的.这种形势下,用户必须增大容量来存放自己所需的数据;磁盘和 DAS 架构的固有容量自不必说, SAN 的扩容必须配有机头,导致数据的增长,这种存储模式性能降低.而且,增大的容量又使成本增加,这种成本的增加将是持续性的,是一个无底洞.传统存储模式的数据冗余也是一大问题,很多没用的重复的数据占据了磁盘大量的容量,有用的数据寥寥可数,数据的存储效率低下^[4].

3 软件定义存储原理

软件定义存储(SDS)是一种数据存储方式,所有存储相关的控制工作都仅在相对于物理存储硬件的外部软件中.如何理解呢?简单来说, SDS 是用来存储的,是存储的一种方式.什么是软件,用户和硬件之间的接口界面就是软件.单纯从字义上理解,软件比硬件更具灵活性. SDS 就是把控制平面单独抽离出来,放在软件上实现这些存储功能的一种数据存储方式^[5].

SDS 的核心是存储虚拟化,在一定程度上 SDS 可以认为就是存储虚拟化. SDS 可以自动分配存储、安全、网络等资源,将这些资源池化,并且 API

来提供给消费者一个友好的用户界面,从而实现数据保护、复制、压缩等一系列存储功能,这些功能都是依靠软件来控制的, SDS 的解决方案,用户在不增加任何工作量的情况下进行纵向扩展或横向扩展^[6]. 2012 年 8 月盛大云因一台物理服务器磁盘发生损坏导致一些用户的数据丢失,原因是没有备份.要是当时采用的是 SDS 这种存储方式的话,也许就不会出现这种问题从而避免损失.

软件定义存储(SDS)的定义出现至今已将近三年,但目前仍然没有一致的意见,所有研究机构和各大厂商,他们对软件定义存储有着不同的定义或描述.但作为数据存储领域的权威全球网络存储工业协会(SNIA)则认为 SDS 应允许异构的或者专有的平台,并且平台能够提供管理和部署其虚拟存储空间的自助服务接口.除此, SDS 还包括以下特征:

1)自动化的标准即为管理比较简化,维护存储架构的成本也降低;

2)标准接口指对于部署、管理、维护存储服务和存储设备,并且能够提供良好的应用编程接口;

3)扩展性即提供可靠性和性能的无缝扩展,其中无需中断应用;

4)虚拟数据路径即可以提供文件、块和对象的接口,又能够支持这些接口写入数据;

存储系统和数据中心也可以为数据管理员都能用于部署和控制路径服务^[3].软件定义存储包括控制路径和数据路径,影响到所有存放在相应虚拟存储空间的数据,而软件定义存储 SDS 则通过以下路径实现存储控制:

步骤 1:确定存储需求,设计一种存储解决方案的第一步是确定您的工作负载(应用程序、服务器和客户端),然后确定存储需求.

步骤 2:检查的设计,并根据需要进行调整,完成步骤 1 和软件定义存储设计中的任务,将获取您的硬件和软件的建议的设计.

步骤 3:使用存储空间的硬件供应商以最终确定设计.

步骤 4:准备部署解决方案.

具体方案如下:在交换机启动时进行初始化将交换机的转发缓存空间置空,为交换机设置缓存阈值,用于检测交换机数据处理状况.当交换机收到数据分组时,先将数据分组放入到转发缓存中,然后判断交换机转发缓存是否达到阈值,如果没有达到阈值,则直接在缓存中等待处理,匹配流表,执行相应的 Action.当缓存中的数据达到阈值

时,则发送达到阈值的信息告诉控制器,控制器收到消息后,对经过该交换机的数据分组重新进行路由选择,有其它路径的就按照新路径进行转发;对于没有其它路径的数据分组,则将数据暂时还存存在其下一跳交换机中,通知源节点对必须通过该路径的数据分组进行延迟发送,等待节点拥塞解决后再重新进行发送. 控制器调取所有连接到它的交换机的缓存状况,将还有足够缓存空间的所有相邻交换机的信息封装到流表中,对交换机下发流表,让出现拥塞的交换机把缓存中的数据分组,平均分配发送到可以代为缓存的交换机

中. 数据分组到达相邻交换机后,为数据分组重现计算路由,有可用路径的则代为转发,反之,就放入缓存中,等待出现拥塞的交换机恢复正常. 如果相邻交换机中都有足够的缓存空间,则优先将数据分组进行二次转发到下一级相邻交换机,执行上述操作. 拥塞解决后,交换机发送拥塞已解决的信息到控制器,控制器对全网交换机发出通知信息,将暂存在相邻交换机中的数据分组取回重新进行发送同时暂停发送的数据分组继续发送过来. 如图1所示.

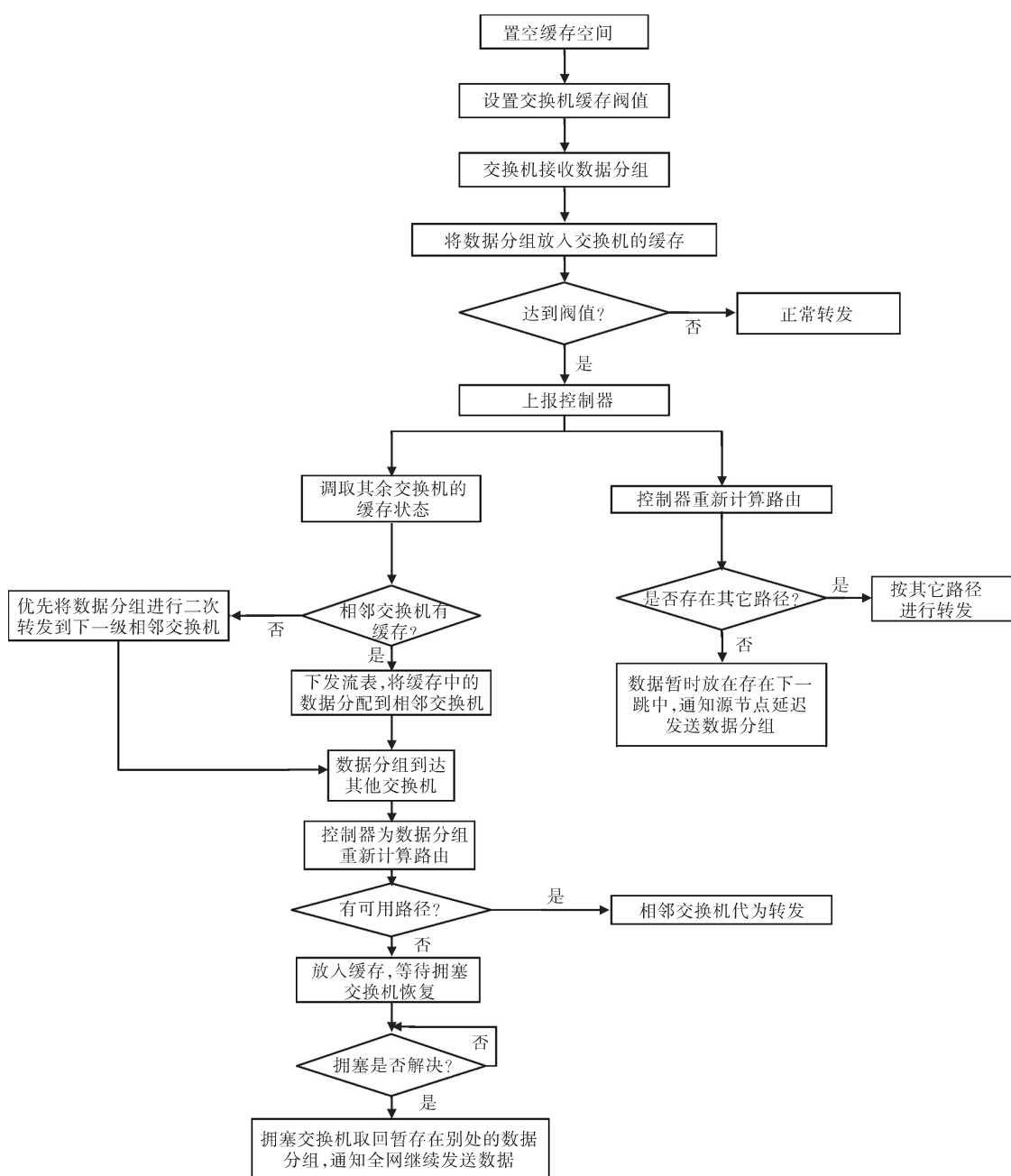


图1 实现的方案

Fig. 1 Implementation of the program

4 软件定义存储的优势

软件定义存储(SDS)之所以被云存储服务提供商所重视并逐渐超越传统的存储模式正是由于其相对于传统存储方式具有众多的优势,下面将对软件定义存储的主要优势进行初步分析。

4.1 数据处理的实时性

企业可以从数据中获得新的资料情报,从而获得新商机,所以,能否即时访问数据成为企业找到商机的重要因素。现在的存储对资源的访问提出更大的需求。显然,以硬件为中心的传统方式无法实现这一功能,缺乏灵活性。随着用户需求的提高,硬件定义存储呈现出了诸多弊端。要想实现更多功能,硬件必须更新换代,专门制造根据用户需要的产品,否则缺乏灵活性。而且生产出来的产品有了一些小问题,就得回炉重造,成本和效率得不到提升是可以预见的^[7]。SDS就没有这种担忧,它可以根据用户的需求专门定制一套服务。当用户要求增加某项服务和功能时,只需要在架构的系统中动刀就可以了,实现了高扩展性,这是传统存储架构所欠缺的^[8]。

4.2 低成本

以前的存储功能相对固化,大多数存储功能都集中在硬件上,造成了想要某一种存储设备必须从特定的厂商购买,成本自然高^[9]。值得注意的是,软件定义存储的出现,让用户不必从特定厂商购买存储介质,也就打断了这些厂商的垄断,它可以架构在任何类型的服务器硬件上。也就是说,它创建于该硬盘,但可以使用在其他硬盘上,SDS更像是一个系统。就像手机App一样,任何手机通过下载都能够得到这款应用程序,不是谁创建了它就专属谁。所以买了SDS等同于买了特定存储设备,这是一个巨大的进步^[10]。SDS的最大成本就在于它的架构及其管理,没有了特定厂商的垄断,成本自然降低。

4.3 高灵活性

SDS可以在硬件上运行,同样可以运行在云存储的架构中,它可以存在于云存储服务中作为云存储的一部分^[11]。虽然SDS是将控制平面单独分离出来的,但是它拥有硬件应该有的所有存储功能,这点是最重要的。那么,使用SDS架构的云存储,SDS将为用户处理大部分网络管理工作,因为SDS可以自动分配这些工作。提高存储的使用率,SDS的灵活性和扩展性将为云存储如虎添翼^[12]。

5 实验模拟

5.1 实验环境

网络环境的搭建方案^[13]:

- 1) 硬件平台是PC物理机一台;
- 2) 在PC物理机上安装VMware Workstation 12.1.0;
- 3) 在VMware中创建Ubuntu 14.04虚拟机^[14];
- 4) 在Ubuntu中配置安装IOMeter。

在实验中主机上采用Iometer对远端的逻辑盘进行测评,选用5个数据量不同的样本,分别为5 Mbit, 50 Mbit, 500 Mbit, 5 Gbit 和 5 Tbit 这5个容量。分析三种网络存储方式在不同数据量下存储效率的表现,如图2所示。为了更精确比较这三种方式,做出以下规定:

- 1) 对于每个数据量,三种存储方式在相同时间内测试存储速率;
- 2) 每个数据量测试两组,取平均值(四舍五入);
- 3) 除了存储方式与数据量的不同,其他无关变量尽量避免。

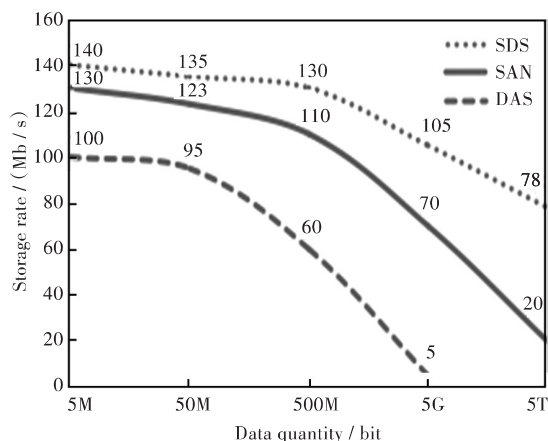


图2 不同存储结构在不同数据块下的存储速率

Fig. 2 Storage rates of different storage structures under different data blocks

5.2 实验分析

从图2可以直观地看出,软件定义存储性能优于其他存储结构:

- 1) 在小容量的数据块中,软件定义存储的存储速率跟SAN存储结构是差不多的,DAS存储结构次之。
- 2) 随着数据块容量增大,软件定义存储没有令人失望,虽然三者的处理能力都有了下降,可是

软件定义存储还是遥遥领先,三者之间的差距越来越明显.软件定义存储的存储性能明显优于SAN存储结构与DAS存储结构.

3)在数据量达到5 Tbit的时候,三种存储结构下降的幅度都很明显,DAS存储结构甚至达到了不能处理的情况,SAN存储虽然能够处理,相对于软件定义存储它的存储速度就有点脱节了.

分析比较之后可以看出,在信息量少的处理中,三种存储结构都可以驾驭.所以DAS存储更适合处理信息量少的数据,比如小型数字视频监控项目,DAS的缺点这么明显,它的优点也是显而易见的,结构简单,成本廉价是DAS存储的优势^[15];SAN存储结构的存储性能还是比较优秀的,扩容能力较强,是一种比较优秀的存储方式.可是面对大数据时代,它的优势就不能很好地展现了,它更适用于SQL数据库;软件定义存储在系统软件的修改方面实现用户提出的新功能,在处理超大数据时达到78 Mbit/s的存储速率,约为传统SAN存储技术的3.9倍,这些优势将极大的推动云存储的进一步发展.对于超大数据还是需要软件定义存储来胜任,其作为一种新的存储方式,它表现出强大的性能,可以较好的应对大数据时代发出的挑战.

6 结 语

以上介绍了软件定义存储这种新的数据存储方式,将它与现有的几种存储方式的各种参数进行对比分析,展示其优势.然后通过实验模拟将它与其他存储结构进行性能对比,从严谨的实验中证明软件定义存储的存储性能优于传统的大多数网络存储结构.这种新的存储方式具有一定的优势,在大数据时代将发挥它的作用.本文没有对软件定义存储能自定义分配存储资源这一特点进行分析,后续的研究可对软件定义存储如何自定义分配存储资源进行深入研究,进而为存储的安全问题提供有效的解决方案,进一步完善存储技术,一个安全可靠的存储服务将会使用户用得更安心更放心.

参考文献:

[1] 刘雪.云存储中的安全研究及应用[D].北京:北方工

业大学,2012.

- [2] 黄韬,刘江,魏亮,等.软件定义网络核心原理与应用实践[M].北京:人民邮电出版社,2014:9-32.
- [3] 梅胜民.云计算存储安全技术研究[D].长春:吉林大学,2014.
- [4] 高煜红.云计算数据存储安全的研究[D].杭州:浙江工业大学,2014.
- [5] 梁昊驰.SDN可扩展路由及流表资源优化研究[D].合肥:中国科学技术大学,2015.
- [6] 张咏悦.SDN网络管理系统的设计与实现[D].北京:北京邮电大学,2015.
- [7] 熊平.信息安全原理及应用[M].北京:清华大学出版社,2012:116-126.
- [8] 高东升.大数据的云存储技术[D].沈阳:辽宁大学,2010.
- [9] 张富政.云存储安全技术的研究[D].长春:长春理工大学,2014.
- [10] 许妍.数据存储安全技术研究[D].南京:南京邮电大学,2014.
- [11] 张卫峰.深度解析SDN利益、战略、技术、实践[M].北京:电子工业出版社,2014.
- [12] 左青云,陈鸣,赵广松,等.基于Open Flow的SDN技术研究[J].软件学报,2013,24(5):1078-1097.
ZUO Q Y, CHEN M, ZHAO G S, et al. Research on open flow-based SDN technologies [J]. Journal of software, 2013, 24(5): 1078-1097.
- [13] 刘惠临,林亚飞.基于控制集群的SDN流表更新机制[J].长春工业大学学报(自然科学版),2014,35(4):429-433.
LIU H L, LIN Y F. SDN flow table update mechanism based on controller cluster [J]. Journal of Changchun university of technology (natural science edition), 2014, 35(4): 429-433.
- [14] 杨云飞.基于应用层策略转换一致性的SDN流表更新方法[J].南京:南京邮电大学,2015.
- [15] 顾强,张庆顺,臧军.软件定义网络发展历程及面临的挑战[J].邮电设计技术,2015(11):64-68.
GU Q, ZHANG Q S, ZANG J. The development of software-defined network and the challenges [J]. Designing techniques of posts and telecommunications, 2015(11): 64-68.

本文编辑:陈小平