

# 基于虚拟化与分布式技术的云存储研究

刘亮

(中国移动设计院 安徽分院第三设计研究所,安徽 合肥 230041)

**摘要:**云存储是云计算与云服务实现的核心基石之一,目前业界已有基于虚拟化技术和分布式存储技术的多种云存储解决方案,但是其技术比较与演进趋势仍不明朗。该文从研究目前的虚拟化和分布式云存储技术着手,详细分析对比了各类技术的实现原理、适用场景与应用情况,进而提出了云存储发展的预期,为各种云存储应用与研究提供参考。

**关键词:**虚拟化;分布;云存储

**中图分类号:**TP393 **文献标识码:**A **文章编号:**1009-3044(2012)11-2641-02

云存储是实现云计算服务能力的不可或缺的关键要素。从狭义上来说,云存储是指通过虚拟化、分布式技术、集群应用、网格技术、负载均衡等技术,将网络中大量的存储设备通过软件集合起来高效协同工作,共同对外提供低成本、高扩展性的数据存储服务。从广义上来讲,云存储可以理解为由按需提供的虚拟存储资源,如同云计算的Paas、IaaS服务一样,可称为数据存储即服务(Data Storage As a Service, DaaS),即基于指定的服务水平请求,通过网络提供适当的虚拟存储和相关数据服务。

云存储技术的发展,是因为传统的盘阵存储由于高成本和系统的高负载性,已经难以适应云计算服务对存储的需求。传统盘阵虽然利用专有的、定制设计的硬件组件和互连接口设计集成在一起,保障了存储的性能和可靠性,但是其单点扩展存在容量上限和接口带宽等限制,面对PB级的海量存储需求,无论是SAN或NAS在容量和性能的扩展上均无法满足应用的需要;且盘阵造价昂贵,标准化程度较低,组件的更换常常导致系统故障或者硬件的升级,从而导致存储成本居高不下。

由此,为实现存储的低成本、高可扩展与资源池化,云存储技术应运而生,这其中最关键的是存储虚拟化技术与分布式存储技术的应用<sup>[1]</sup>。从短期来看,存储化技术可以充分提升现有存储资源的应用水平,实现资源池化,提升设备使用率,降低管理成本。从长远发展来看,分布式存储技术则是未来云存储的发展趋势,分布式存储底层可构建在标准X86服务器集群之上,通过分布式文件系统对外提供弹性存储资源,可选择SATA作为存储实现方式,根据应用需要提供丰富的服务接口,从而大大降低了存储成本。同时基于X86的分布式存储的集群架构也实现了较强的Scale-out扩展能力。

该文的第二章对现有的三种类型的存储虚拟化技术原理及应用进行了详细分析,第三章对分布式云存储的技术发展和应用进行了分析和研究,第四章给出了研究结论。

## 1 存储虚拟化

存储虚拟化技术主要指通过在物理存储系统和服务器之间增加一个虚拟层,使服务器的存储空间可以跨越多个异构的磁盘阵列,实现从物理存储到逻辑存储的转变。从而能够把不同品牌、架构甚至不在统一物理居址的存储设备统一到一个(若干个)虚拟的存储资源池,实现存储资源的统一调度、管理和使用。存储虚拟化技术是传统存储向云存储过渡的重要中间技术,在实现存储资源管理统一、高效的同时,使得存储资源的调度、增减对用户完全透明,使存储资源具备了支撑云计算服务的能力。

### 1.1 主要存储虚拟化技术

存储虚拟化技术虽然不同设备与厂商之间略有区别,但从总体来说,可概括为基于主机虚拟化、基于存储设备虚拟化和基于存储网络虚拟化三种技术<sup>[2]</sup>。

1)基于主机的虚拟化存储的实现,其核心技术是通过增加一个运行在操作系统下的逻辑卷管理软件将磁盘上的物理块号映射成逻辑卷号,并以此实现把多个物理磁盘阵列映射成一个统一的虚拟的逻辑存储空间(逻辑块)实现存储虚拟化的控制和管理。从技术实施层面看,基于主机的虚拟化存储不需要额外的硬件支持,便于部署,只通过软件即可实现对不同存储资源的存储管理。但是,虚拟化控制软件也导致了此项技术的主要缺点:首先,软件的部署和应用影响了主机性能;其次,各种与存储相关的应用通过同一个主机,存在越权访问的数据安全隐患;最后,通过软件控制不同厂家的存储设备存在额外的资源开销,进而降低系统的可操作性与灵活性。

2)存储设备虚拟化技术依赖于提供相关功能的存储设备的阵列控制器模块,常见于高端存储设备,其主要应用针对异构的SAN存储构架。此类技术的主要优点是不占主机资源,技术成熟度高,容易实施;缺点是核心存储设备必须具有此类功能,且消耗存储控制器的资源,同时由于异构厂家磁盘阵列设备的控制功能被主控设备的存储控制器接管导致其高级存储功能将不能使用。

3)基于存储网络虚拟化的技术的核心是在存储区域网中增加虚拟化引擎实现存储资源的集中管理,其具体实施一般是通过具有虚拟化支持能力的路由器或交换机实现。在此基础上,存储网络虚拟化又可以分为带内虚拟化与带外虚拟化两类,二者主要的

收稿日期:2012-03-05

作者简介:刘亮,博士,就职于中国移动设计院有限公司安徽分公司第三设计研究所,任咨询工程师。

本栏目责任编辑:梁书

区别在于:带内虚拟化使用同一数据通道传送存储数据和控制信号,而带外虚拟化使用不同的通道传送数据和命令信息。基于存储网络的存储虚拟化技术架构合理,不占用主机和设备资源;但其存储阵列中设备的兼容性需要严格验证,与基于设备的虚拟化技术一样,由于网络中存储设备的控制功能被虚拟化引擎所接管,导致存储设备自带的高级存储功能将不能使用。

1.2 存储虚拟化技术对比

被节用表格的方式对三种存储虚拟化技术的技术优点与缺点、适应场景等进行了分析对比,结果见表1。

表1 存储虚拟化技术对比

实现层面	主机	网络	设备
优点	支持异构的存储系统;不占用磁盘控制器资源	与主机无关,不占用主机资源;能够支持异构主机、异构存储设备;对不同存储设备构建统一管理平台,可扩展性好	与主机无关,不占用主机资源;数据管理功能丰富;技术成熟度高
缺点	占用主机资源,降低应用性能;存在操作系统和应用的兼容性问题;主机数量越多,管理成本越高	占用交换机资源;面临带内、带外的选择;存储设备兼容性需要严格验证;原有的磁盘阵列的高级存储功能将不能使用	受制于存储控制器接口资源,虚拟化能力较弱;异构厂家存储设备的高级存储功能将不能使用
主要用途	使服务器的存储空间可以跨越多个异构磁盘阵列,常用于在不同磁盘阵列之间做数据镜像保护	异构存储系统整合和统一数据管理(灾备)	异构存储系统整合和统一数据管理(灾备)
适用场景	主机已采用 SF 卷管理,需要新接多台存储设备;存储系统中包含异构阵列设备;业务持续能力与数据吞吐要求较高	系统包括不同品牌和型号的主机与存储设备;对数据无缝迁移及数据格式转换有较高时间性保证	系统中包括自带虚拟化功能的高端存储设备与若干需要利旧的中低端存储
不适用场景	主机数量大,采用 SF 会涉及高昂的费用;待迁入系统数据量过大,如果只能采取存储级迁移方式,数据格式转换将耗费大量的时间和人力	对业务持续能力和稳定性要求苛刻	需要新购机头时,费用较高;存在更高端的存储设备

2 分布式存储

分布式存储相比传统的集中阵列存储设备,其技术和解决方案,还处于发展初期,总体来看只具备部分场景下的存储需求实现能力。但是从发展趋势来看,通过一个可扩展的网络进行连接各离散的处理单元的分布式存储系统,其高可扩展性、低成本、无接入限制等优点是现有存储系统所无法比拟的。该文的分析研究将以目前发展的分布式块存储、分布式文件系统存储、分布式对象存储和分布式表存储这四种分布式存储技术为对象进行。

1) 分布式块存储:块存储就是服务器直接通过读写存储空间中的一个或一段地址来存取数据。由于采用直接读写磁盘空间来访问数据,相对于其他数据读取方式,块存储的读取效率最高,一些大型数据库应用只能运行在块存储设备上。分布式块存储系统目前以标准的 Intel/Linux 硬件组件作为基本存储单元,组件之间通过千兆以太网采用任意点对点拓扑技术相互连接,共同工作,构成大型网格存储,网格内采用分布式算法管理存储资源。此类技术比较典型的代表是 IBM XIV 存储系统,其核心数据组件为基于 Intel 内核的磁盘系统,卷数据分布到所有磁盘上,从而具有良好的并行处理能力;放弃 RAID 技术,采用冗余数据块方式进行数据保护,统一采用 SATA 盘,从而降低了存储成本。

2) 分布式文件系统存储:文件存储系统可提供通用的文件访问接口,如 POSIX、NFS、CIFS、FTP 等,实现文件与目录操作、文件访问、文件访问控制等功能。目前的分布式文件系统存储的实现有软硬件一体和软硬件分离两种方式。主要通过 NAS 虚拟化,或者基于 X86 硬件集群和分布式文件系统集成在一起,以实现海量非结构化数据处理能力。

软硬件一体方式的实现基于 X86 硬件,利用专有的、定制设计的硬件组件,与分布式文件系统集成在一起,以实现目标设计的性能和可靠性目标;产品代表 Isilon, IBM SONAS GPFS。软硬件分离方式的实现基于开源分布式文件系统对外提供弹性存储资源,软硬件分离方式,可采用标准 PC 服务器硬件;典型开源分布式文件系统有 GFS<sup>[3]</sup>、HDFS。

3) 分布式对象存储:对象存储是为海量数据提供 Key-Value 这种通过键值查找数据文件的存储模式;对象存储引入对象元数据来描述对象特征,对象元数据具有丰富的语义;引入容器概念作为存储对象的集合。对象存储系统底层基于分布式存储系统来实现数据的存取,其存储方式对外部应用透明。这样的存储系统架构具有高可扩展性,支持数据的并发读写,一般不支持数据的随机写操作。最典型的应用实例就是亚马逊的 S3(Amazon Simple Storage Service)。对象存储技术相对成熟,对底层硬件要求不高,存储系统可靠性和容错通过软件实现,同时其访问接口简单,适合处理海量、小数据的非结构化数据,如:邮箱、网盘、相册、音频视频存储等。

4) 分布式表存储:表结构存储是一种结构化数据存储,与传统数据库相比,它提供的表空间访问功能受限,但更强调系统的可扩展性。大多数表存储的数据可靠性依赖于表存储底层采用的文件存储,对底层硬件要求不高,一般常采用开源软件和通用 PC,具

(下转第 2661 页)

## 4 开发注意事项

### 4.1 ASP 站点的建立

首先要安装 IIS, 完成之后在 Dreamweaver 里点击管理站点分别按照提示填写本地信息、远程信息和测试服务器, 测试服务器中服务器模型选择 ASP VBScript, 访问选择本地/网络, URL 前缀填 `http://localhost/`。

再在控制面板中打开 Internet 信息服务管理器, 右击新建网站, 分别填写网站名称, 物理路径和主机名 `localhost`。

按照上述步骤就可正确建立站点, 并完成网上订餐的设计与实现。

### 4.2 与数据库的连接

很多 ASP 初学者都在建立与数据库的连接问题上头疼不已, 现介绍手动方式的连接方法。在控制面板中打开数据源 ODBC, 在系统 DSN 中添加 Microsoft Access Driver。找到之前建立的 .mdb 文件, 给数据源取名为 `myconn`。回到 Dreamweaver 界面, 在应用程序的数据库中添加数据源名称, 给连接取名为 `aspconn`。

## 5 结束语

本系统可以满足顾客和企业绝大多数的要求, 但还存在需要改进的地方。比如搜索方式较为单一, 只能通过关键词查询, 可以改进为通过多种方法检索, 如餐品种类、餐品口味、餐品价格区间、新品搜索等方式。除此之外, 并没有将系统和短信服务联系, 订单形成后要手动确认。改进的方法可以是在一笔订单确认完成后自动获取表单中顾客的联系方式, 对顾客的手机或邮箱分别发送短信或邮件。当顾客回信之后, 将订单信息交由后厨, 再进行餐品的制作。这就可以使订餐流程更加精炼。

### 参考文献:

- [1] 张海永, 章伟德. 精通 ASP+XML+CSS 网络开发混合编程[M]. 北京: 电子工业出版社, 2006.
- [2] 李素若, 严永松, 陈万华. ASP 动态网页设计[M]. 北京: 化学工业出版社, 2009.
- [3] 鲍尔斯. PHP 与 Dreamweaver 基础教程[M]. 陈宗斌, 译. 北京: 人民邮电出版社, 2011.

---

(上接第 2642 页)

有低成本的特性。提供分布式表存储的云存储系统的特征就是同时提供高并发的数据访问性能和可伸缩的存储、计算架构, 适合存储海量结构化或半结构化数据。其典型产品代表有 Greenplum 和谷歌的 Bigtable<sup>[4]</sup>。以 Greenplum 为例, 一般采用 MPP (Massive Parallel Processing) 架构实现海量数据存储和处理、以及高并发数据读写能力, 它实现了 SQL 到 MapReduce 的翻译、优化、执行和结果收集, 具有良好的扩展能力。

## 3 结论

综上所述, 从目前云存储主要技术和解决方案的发展来看, 已经具备从传统存储架构向云存储架构演进的条件。具体到技术选择和方案实施层面, 首先需要关注存储系统的应用需求, 特别是其数据需求量与数据结构特点, 从而确定选择的存储技术与方案类型; 其次需要考虑是否需要利用现有存储设备, 分布式存储技术多数应用了全新的存储架构, 一般无法和现有存储设备整合; 最后需要从性能实现和存储成本等方面对技术方案进行比较。

### 参考文献:

- [1] 李煜民, 章才能, 谢杰. 云计算环境下的数据存储[J]. 电脑知识与技术, 2010, 6(5): 1032-1034.
- [2] 白翠琴, 王建, 李旭伟. 存储虚拟化技术的研究与比较[J]. 计算机与信息技术, 2008(7).
- [3] Ghemawat S, Gob Dffh, Leung P T. The Google file system, Proceedings of the 19th ACM Symposium on Operating Systems Principles[M]. New York: ACM Press, 2003: 29-43.
- [4] Yu Y, Isard M, Fetterly D, et al. DryadLINQ: A system for general-purpose distributed data-parallel computing using a high-level language[C]. Proceedings of the 8th Symposium on Operating Systems Design and Implementation (OSDI), 2008.