

块和文件之外的存储形式-----对象存储介绍

郝常杰

块是存储的基础架构，对块的管理是简单的。从全球第一个磁盘存储技术 50 年前诞生到现在，块的技术没有发生太大变化。磁盘工业生产遍及全球。最近几年，存储行业兴起存储虚拟化和数据整合的新概念，在这些概念和方案中，数据还是以块的形式通过网络进行读写操作。

块形式的存储在满足数据可扩展性和数据安全性的增长方面，日益显现出其局限性和不足。国际上主要有两类网络化存储架构，它们是通过命令集来区分的。第一类是 SAN (Storage Area Network) 结构，它采用 SCSI 块 I/O 的命令集，通过在磁盘或 FC (Fiber Channel) 级的数据访问提供高性能的随机 I/O 和数据吞吐率，它具有高带宽、低延迟的优势，在高性能计算中占有一席之地，但是由于 SAN 系统的价格较高，且可扩展性较差，已不能满足成千上万个 CPU 规模的系统。第二类是 NAS (Network Attached Storage) 结构，它采用 NFS 或 CIFS 命令集访问数据，以文件为传输协议，通过 TCP/IP 实现网络化存储，可扩展性好、价格便宜、用户易管理，如目前在集群计算中应用较多的 NFS 文件系统，但由于 NAS 的协议开销高、带宽低、延迟大，不利于在高性能集群中应用。

针对 Linux 集群对存储系统高性能和数据共享的需求，国际上已开始研究全新的存储架构和新型文件系统，希望能有效结合 SAN 和 NAS 系统的优点，支持直接访问磁盘以提高性能，通过共享的文件和元数据以简化管理，目前对象存储系统已成为 Linux 集群系统高性能存储系统的研究热点，如 Panasas 公司的 Object Base Storage Cluster System 系统和 Cluster File Systems 公司的 Lustre 等。

基于对象的存储

在 2004 年，ANSI 推出了基于对象的存储设备 (OSD) 的 1.0 版本规范。它定义了基于对象的存储设备的通讯协议。OSD 规范描述了一个 SCSI 命令集合，由他提供一个高水平的 OSD 接口。这个接口允许客户端，比如文件系统和数据库存放和索引数据。SNIA' S 技术工作组当前正在开发 OSD 规范的 2.0 版本，这个版本年内完成。

一个 OSD 设备存放对象，他控制着从对象到物理介质的映射图。设备同时也跟踪作为属性的元数据，例如建立时间标记，从而允许在客户端非常容易地共享数据。

可以说，OSD 最大的卖点在于它结合了 SAN 的可扩展性和 NAS 的数据共享。早期的 NAS 架构的扩展性能并不好，因为所有的元数据的处理都集中在 NAS 服务器上。在有限的 NAS 头下扩张更多的存储受到限制，而且这个时候，NAS 上的元数据处理变成了瓶颈。如果想扩展，就需要增加更多的 NAS 服务器，但是此时的管理成为头疼的事情，因为数据是分散的，这就

是我们常说到的“NAS 孤岛”

OSD的能力在于它将客户端和OSD设备直接联系起来，并不需要中间环节管理元数据。Panasas 公司，全球第一家提供商用OSD产品的公司，同时提供面向对象的存储和并行文件系统。Panasas 公司的 DirectFLOW 的设计，客户端从带外管理的控制刀片得到目标的分布和安全属性。所有的数据流都直接从OSD存储刀片到客户端。尽管商用的OSD产品还只是凤毛麟角，但是OSD技术还是在日新月异。西捷和IBM已经展示了OSDc产品。HP已经和开放源代码厂家 Lustre 文件系统合作，使用OSD作为他的StorageWork 可扩展文件系统的重要部分。

对于对象存储来讲，不光是解决了数据存储的问题，同时它还解决了数据安全性的问题。存储通常已经依赖于客户端和私有网络的认证来保护系统的安全性，不管在文件服务器内用的是FC SAN或SCSI阵列。对象存储体系结构在每一个级别上都提供了安全性：存储系统对存储设备的认证；存储系统对计算节点的认证；存储系统对计算节点命令的认证；所有命令都经CRC的完整性检查；数据和命令经由IP的私有性。这样的安全水平能给用户以信心，他们可以用更经济高效，可管理并容易访问的网络，如以太网，作为存储的传输工具，同时还提高了整个存储体系的安全性。

“对象存储设备并不适合所有的用户。”Panasas 公司的CTO Garth Gibson博士说。经过几年的努力，Panasas 在政府，科研，能源，媒体和金融服务有了很多非常成功的案例。

“Panasas 是使用面向对象的存储集群，来解决计算集群的并行存储的问题”。高性能计算和存储专家胡家鏊教授说。

高性能计算领域

对象存储体系结构提供了一个带有NAS系统的传统的文件共享和管理特征的单系统映像(single-system-image)文件系统，并改进了SAN的资源整合和可扩展的性能。这种性能，可扩展性，可管理性以及安全性的结合，只能通过在存储体系结构上的重大革命才能被完成。第一个支持对象体系结构的产品现在已经问世，它就是Panasas ActiveScale Storage Cluster (存储集群)。此产品已经被部署在国家实验室，地震资料处理机构，以及生物技术组织，它们都在用Linux集群去解决关键的科学问题，这些问题在过去是用巨型超级计算机也不能解决的。它们都期盼能解决需要高性能，可扩展的共享存储才能解决的新问题，这种存储在Panasas的存储体系之前，在市场上是没有的。Panasas存储集群及基于对象的存储体系结构已经展示了，它们能迎接由Beowolf/Linux集群计算体系结构提出的挑战，在那里，传统的基于SAN和DAS的产品都会败下阵来。Panasas在美国能源部(U.S. Department of Energy: DOE)、Lawrence Livermore国家实验室，Los Alamos国家实验室，Sandia国家实验室，Pacific Northwest国家实验室的高性能计算系统中已得到了初步的应用。在全球范围内，Panasas对象存储集群系统的性能在企业级HPC集群的关键应用中得以淋漓体现和发挥。OSD将来会从HPC存储过渡成主流的企业级存储吗？我们拭目以待。试想，当年SUN公司的NFS不就是从最早的技术环境走到主流商业应用的最前线的吗？

对象存储成为主流

无论 OSD 是从 HPC 走到企业级应用中，或是通过其他方式，均面临需要被企业级的用户所接受这样的事实。从传统的存储变成 OSD 存储，需要在多个层面上作改变，这包括，存储设备，存储网络，文件系统和数据库。

OSD 面临着巨大的机遇。如 Garth Gibson 博士所讲，大规模并行数据存储的问题通过数据对象，和属性可以得到解决。在这个时候，对象存储的特性才得以发挥。而这些都是基于块的存储所不能解决的。

对象存储可以以这种方式使用而与应用无关。应用系统可以和以前使用文件导向的系统一样使用，所有的工作通过文件系统实现对文件的存取。这个时候，应用直接存储数据对象而不是向文件系统写数据。

对象存储自诞生的那一天起已经表现出其巨大的活力，它可以提供数据安全，容易的数据共享，强大的可扩展功能，完全具备走向企业级数据中心的能力。我们很高兴地看到，对象存储的国际标准已经制定，很多全球一线的存储厂商和服务器厂家均加入到此阵营中。各个厂家的对象存储产品已经或即将面世。但是，同时我们也看到，对象存储全面取代传统的块存储还需时日。