**三种存储类型比较-文件、块、对象存储**

块存储和文件存储是我们比较熟悉的两种主流的存储类型，而对象存储（Object-based Storage）是一种新的网络存储架构，基于对象存储技术的设备就是对象存储设备（Object-based Storage Device）简称OSD。

        首先，我们介绍这两种传统的存储类型。**通常来讲，所有磁盘阵列都是基于Block块的模式，而所有的NAS产品都是文件级存储。**

**1、块存储**

        以下列出的两种存储方式都是块存储类型：

        1） DAS（Direct Attach [ST](http://www.dzsc.com/data/iccompany/detail3.html)orage）：是直接连接于主机[服务器](http://www.dzsc.com/product/searchfile/5368.html)的一种储存方式，每一台主机服务器有独立的储存设备，每台主机服务器的储存设备无法互通，需要跨主机存取资料时，必须经过相对复杂的设定，若主机服务器分属不同的操作系统，要存取彼此的资料，更是复杂，有些系统甚至不能存取。通常用在单一网络环境下且数据交换量不大，性能要求不高的环境下，可以说是一种应用较为早的技术实现。

        2）SAN（Storage Area Network）：是一种用高速（[光纤](http://www.dzsc.com/product/searchfile/2870.html)）网络联接专业主机服务器的一种储存方式，此系统会位于主机群的后端，它使用高速I/O 联结方式， 如 SCSI, ESC[ON](http://www.dzsc.com/data/iccompany/detail12.html) 及 Fibre- Channels。一般而言，SAN应用在对网络速度要求高、对数据的可靠性和安全性要求高、对数据共享的性能要求高的应用环境中，特点是代价高，性能好。例如电信、银行的大数据量关键应用。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。

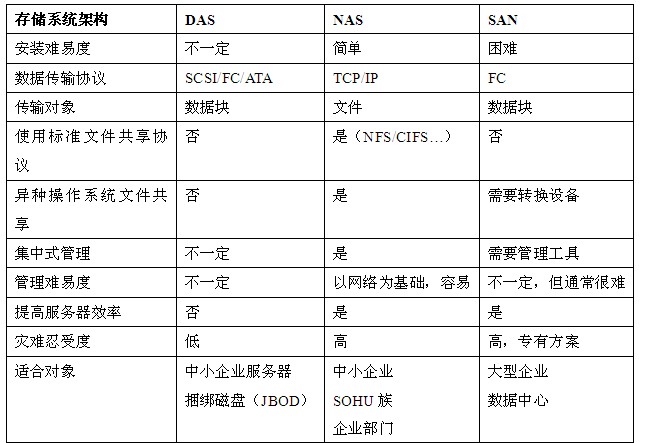
**2、文件存储**

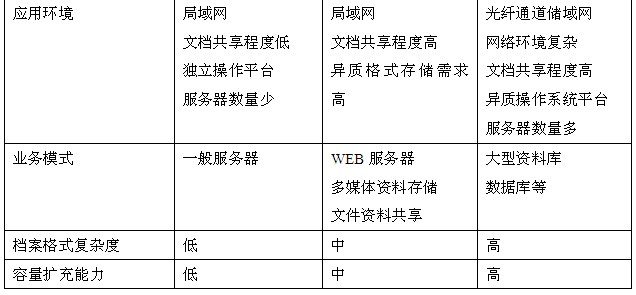
        通常，NAS产品都是文件级存储。  NAS（Network Attached Storage）：是一套网络储存设备，通常是直接连在网络上并提供资料存取服务，一套 NAS 储存设备就如同一个提供数据文件服务的系统，特点是性价比高。例如教育、政府、企业等数据存储应用。

        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。

     下面，我们对DAS、NAS、SAN三种技术进行比较和分析：

表格 1 三种技术的比较





        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。下面将详细介绍对象存储系统。

**3、对象存储**

**总体上来讲，对象存储同兼具SAN高速直接访问磁盘特点及NAS的分布式共享特点。**

        核心是将数据通路（数据读或写）和控制通路（元数据）分离，并且基于对象存储设备（Object-based Storage Device，OSD）构建存储系统，每个对象存储设备具有一定的智能，能够自动管理其上的数据分布。  
        对象存储结构组成部分（对象、对象存储设备、元数据服务器、对象存储系统的客户端）：  
      **3.1、对象**  
        对象是系统中数据存储的基本单位，一个对象实际上就是文件的数据和一组属性信息（Meta Data）的组合，这些属性信息可以定义基于文件的RAID参数、数据分布和服务质量等，而传统的存储系统中用文件或块作为基本的存储单位，在块存储系统中还需要始终追踪系统中每个块的属性，对象通过与存储系统通信维护自己的属性。在存储设备中，所有对象都有一个对象标识，通过对象标识OSD命令访问该对象。通常有多种类型的对象，存储设备上的根对象标识存储设备和该设备的各种属性，组对象是存储设备上共享资源管理策略的对象集合等。   
       **3.2、对象存储设备**  
        对象存储设备具有一定的智能，它有自己的CPU、内存、网络和磁盘系统，OSD同块设备的不同不在于存储介质，而在于两者提供的访问接口。OSD的主要功能包括数据存储和安全访问。目前国际上通常采用刀片式结构实现对象存储设备。OSD提供三个主要功能：  
      （1） 数据存储。OSD管理对象数据，并将它们放置在标准的磁盘系统上，OSD不提供块接口访问方式，Client请求数据时用对象ID、偏移进行数据读写。  
      （2） 智能分布。OSD用其自身的CPU和内存优化数据分布，并支持数据的预取。由于OSD可以智能地支持对象的预取，从而可以优化磁盘的性能。  
      （3） 每个对象元数据的管理。OSD管理存储在其上对象的元数据，该元数据与传统的inode元数据相似，通常包括对象的数据块和对象的长度。而在传统的NAS系统中，这些元数据是由文件服务器维护的，对象存储架构将系统中主要的元数据管理工作由OSD来完成，降低了Client的开销。  
        **3.3、元数据服务器（Metadata Server，MDS）**          MDS控制Client与OSD对象的交互，主要提供以下几个功能：  
      （1） 对象存储访问。  
        MDS构造、管理描述每个文件分布的视图，允许Client直接访问对象。MDS为Client提供访问该文件所含对象的能力，OSD在接收到每个请求时将先验证该能力，然后才可以访问。  
      （2） 文件和目录访问管理。  
        MDS在存储系统上构建一个文件结构，包括限额控制、目录和文件的创建和删除、访问控制等。  
     （3） Client Cache一致性。  
        为了提高Client性能，在对象存储系统设计时通常支持Client方的Cache。由于引入Client方的Cache，带来了Cache一致性问题，MDS支持基于Client的文件Cache，当Cache的文件发生改变时，将通知Client刷新Cache，从而防止Cache不一致引发的问题。  
       **3.4、对象存储系统的客户端Client**        为了有效支持Client支持访问OSD上的对象，需要在计算节点实现对象存储系统的Client，通常提供POSIX文件系统接口，允许应用程序像执行标准的文件系统操作一样。  
 **4、GlusterFS 和对象存储**

       GlusterFS是目前做得最好的分布式存储系统系统之一，而且已经开始商业化运行。但是，目前GlusterFS3.2.5版本还不支持对象存储。如果要实现海量存储，那么GlusterFS需要用对象存储。值得高兴的是，GlusterFS最近宣布要支持对象存储。它使用openstack的对象存储系统swift的上层PUT、GET等接口，支持对象存储。