

CSDN

博客学院下载图文课TinyMind论坛问答商城VIP会员活动招聘ITeyeGitChat

搜博主文章

写博客发Chat

转

块存储、文件存储、对象存储这三者和分布式文件存储系统的本质区别

2017年11月10日 16:13:33 hNicholas 阅读数：1215

01

块存储和文件存储是我们比较熟悉的两种主流的存储设备（Object-based Storage Device）简称OSD。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

首先，我们介绍这两种传统的存储类型。通常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。

首先我们先来了解一下，什么叫做块级？什么叫文件级？

1.块级概念：

块级是指以扇区为基础，一个或我连续的扇区组成一个块，也叫物理块。它是在文件系统与块设备（例如：磁盘驱动器）之间。

2.文件级概念：

文件级是指文件系统，单个文件可能由于一个或多个逻辑块组成，且逻辑块之间是不连续分布。逻辑块大于或等于物理块整数倍，

3.物理块与文件系统之间的关系图：

映射关系：扇区→物理块→逻辑块→文件系统

文件级备份：

文件级备份是指在指定某些文件进行备份时，首先会查找每个文件逻辑块，其次物理块，由于逻辑块是分散在物理块上，而物理块也是分散在不同扇区上。需要一层一层往才完成整个文件复制。文件级备份时比较费时间，效率不高，实时性不强，备份时间长，且增量备份时，单文件某一小部份修改，不会只备份修改部份，而整个文件都备份

块级备份：

块级备份是指物理块复制，效率高，实时性强，备份时间短，且增量备份时，只备份修改过的物理块。

目前文件级备份工具：

Symantec NBU/BE 备份软件、Commvault、CA、Networker

目前块级备份工具：

飞康CDP、Recoverpoint、杭州信核CDP、Novell CDP

备份时间点保留周期：

传统备份软件（文件级备份），可以保留备份时间点多，恢复颗粒度大

CDP备份（块级备份），可以保留备份时间点少，恢复颗粒度小

【块存储】

典型设备：磁盘阵列，硬盘，虚拟硬盘

【文件存储】

典型设备：FTP、NFS服务器，SamBa

【对象存储】

典型设备：内置大容量硬盘的分布式服务器

分布式存储的应用场景相对于其存储接口，现在流行分为三种：

对象存储：也就是通常意义的键值存储，其接口就是简单的GET、PUT、DEL和其他扩展，如七牛、又拍、Swift、S3

块存储：

这种接口通常以QEMU Driver或者Kernel Module的方式存在，这种接口需要实现Linux的Block Device的接口或者QEMU提供的Block Driver接口，如Sheepdog，AWS的EBS云硬盘和阿里云的盘古系统，还有Ceph的RBD（RBD是Ceph面向块存储的接口）

文件存储:

通常意义是支持POSIX接口，它跟传统的文件系统如Ext4是一个类型的，但区别在于分布式存储提供了并行化的能力，如Ceph的CephFS(CephFS是Ceph面向文件存储的接口)又会把GFS，HDFS这种非POSIX接口的类文件存储接口归入此类。

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 Storage）：是直接连接于主机服务器的一种储存方式，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（光纤）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，它使用高速式，如SCSI, ESCON 及 Fibre- Channels。一般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，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

存储系统架构	DAS	NAS	SAN
安装难易度	不一定	简单	困难
数据传输协议	SCSI FC/ATA	TCP/IP	FC
传输对象	数据块	文件	数据块
使用标准文件共享协议	否	是（NFS/CIFS...）	否
异种操作系统文件共享	否	是	需要转换设备
集中式管理	不一定	是	需要管理工具
管理难易度	不一定	以网络为基础，容易	不一定，但通常很难
提高服务器效率	否	是	是
灾难忍受度	低	高	高，专有方案
适合对象	中小企业服务器 捆绑磁盘（JBOD）	中小企业 SOHU 族 企业部门	大型企业 数据中心

应用环境	局域网 文档共享程度低 独立操作平台 服务器数量少	局域网 文档共享程度高 异质格式存储需求高	光纤通道存储域网 网络环境复杂 文档共享程度高 异质操作系统平台 服务器数量多
业务模式	一般服务器	WEB 服务器 多媒体资料存储 文件资料共享	大型资料库 数据库等
档案格式复杂度	低	中	高
容量扩充能力	低	中	高

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Obj 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，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R分布和服务质量等，而传统的存储系统中用文件基本的存储单位，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。（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
- （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努力，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CI 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等接口存储。

<http://www.openstack.org/blog/2011/07/announcing-the-gluster-connector-for-openstack/>

## ——详细介绍

【块存储】

典型设备：磁盘阵列，硬盘

块存储主要是将裸磁盘空间整个映射给主机使用的，就是说例如磁盘阵列里面有5块硬盘（为方便说明，假设每个硬盘1G），然后通过划逻辑盘、做Raid、或者LVM种种方式逻辑划分出N个逻辑的硬盘。（假设划分完的逻辑盘也是5个，每个也是1G，但是这5个1G的逻辑盘已经于原来的5个物理硬盘意义完全不同了。例如第一个逻辑盘能第一个200M是来自物理硬盘1，第二个200M是来自物理硬盘2，所以逻辑硬盘A是由多个物理硬盘逻辑虚构出来的硬盘。）  
接着块存储会采用映射的方式将这几个逻辑盘映射给主机，主机上面的操作系统会识别到有5块硬盘，但是操作系统是区分不出到底是逻辑还是物理的，它一概就认为只是硬盘而已，跟直接拿一块物理硬盘挂载到操作系统没有区别的，至少操作系统感知上没有区别。  
此种方式下，操作系统还需要对挂载的裸硬盘进行分区、格式化后，才能使用，与平常主机内置硬盘的方式完全无异。

优点：

1、这种方式的好处当然是因为通过了Raid与LVM等手段，对数据提供了保护。  
2、另外也可以将多块廉价的硬盘组合起来，成为一个大容量的逻辑盘对外提供服务，提高了容量。  
3、写入数据的时候，由于是多块磁盘组合出来的逻辑盘，所以几块磁盘可以并行写入的，提升了读写效率。  
4、很多时候块存储采用SAN架构组网，传输速率以及封装协议的原因，使得传输速度与读写速率得到提升。

缺点：

1、采用SAN架构组网时，需要额外为主机购买光纤通道卡，还要买光纤交换机，造价成本高。

2、主机之间的数据无法共享，在服务器不做集群的情况下，块存储裸盘映射给主机，再格式化使用后，对于主机来说相当于本地盘，那么主机A的本地盘根本不能给主机B共享数据。

3、不利于不同操作系统主机间的数据共享：另外一个原因是因为操作系统使用不同的文件系统，格式化完之后，不同文件系统间的数据是共享不了的。例如一台装了WIN系统是FAT32/NTFS，而Linux是EXT4，EXT4是无法识别NTFS的文件系统的。就像一只NTFS格式的U盘，插进Linux的笔记本，根本无法识别出来。所以不利于文件共享。

**【文件存储】**

**典型设备：FTP、NFS服务器**

为了克服上述文件无法共享的问题，所以有了文件存储。

文件存储也有软硬一体化的设备，但是其实普通拿一台服务器/笔记本，只要装上合适的操作系统与软件，就可以架设FTP与NFS服务了，架上该类服务之后的服务器，就是一种了。

主机A可以直接对文件存储进行文件的上传下载，与块存储不同，主机A是不需要再对文件存储进行格式化的，因为文件管理功能已经由文件存储自己搞定了。

**优点：**

1、造价交低：随便一台机器就可以了，另外普通以太网就可以，根本不需要专用的SAN网络，所以造价低。

2、方便文件共享：例如主机A（WIN7，NTFS文件系统），主机B（Linux，EXT4文件系统），想互拷一部电影，本来不行。加了个主机C（NFS服务器），然后可以先A到B就OK了。（例子比较肤浅，请见谅.....）

**缺点：**

读写速率低，传输速率慢：以太网，上传下载速度较慢，读写都要1台服务器里面的硬盘来承担，相比起磁盘阵列动不动就几十上百块硬盘同时读写，速率慢了许

**【对象存储】**

**典型设备：内置大容量硬盘的分布式服务器**

对象存储最常用的方案，就是多台服务器内置大容量硬盘，再装上对象存储软件，然后再额外搞几台服务作为管理节点，安装上对象存储管理软件。管理节点可以管理其他供读写访问功能。

之所以出现了对象存储这种东西，是为了克服块存储与文件存储各自的缺点，发扬它俩各自的优点。**简单来说块存储读写快，不利于共享，文件存储读写慢，利于共享。自快，利于共享的出来呢。于是就有了对象存储。**

首先，一个文件包含了了属性（术语叫metadata，元数据，例如该文件的大小、修改时间、存储路径等）以及内容（以下简称数据）。

以往像FAT32这种文件系统，是直接将一份文件的数据与metadata一起存储的，存储过程先将文件按照文件系统的最小块大小来打散（如4M的文件，假设文件系统要求么就将文件打散成为10000小块），再写进硬盘里面，过程中没有区分数据/metadata的。而每个块最后会告知你下一个要读取的块的地址，然后一直这样顺序地按图索整份文件的所有块的读取。

这种情况下读写速率很慢，因为就算你有100个机械手臂在读写，但是由于你只有读取到第一个块，才能知道下一个块在哪里，其实相当于只能有1个机械手臂在实际工作。而对象存储则将元数据独立了出来，控制节点叫元数据服务器（服务器+对象存储管理软件），里面主要负责存储对象的属性（主要是对象的数据被打散存放到了那几台分的信息），而其他负责存储数据的分布式服务器叫做OSD，主要负责存储文件的数据部分。当用户访问对象，会先访问元数据服务器，元数据服务器只负责反馈对象存储。假设反馈文件A存储在B、C、D三台OSD，那么用户就会再次直接访问3台OSD服务器去读取数据。

这时候由于是3台OSD同时对外传输数据，所以传输的速度就加快了。当OSD服务器数量越多，这种读写速度的提升就越大，通过此种方式，实现了读写快的目的。

另一方面，对象存储软件是有专门的文件系统的，所以OSD对外又相当于文件服务器，那么就不存在文件共享方面的困难了，也解决了文件共享方面的问题。

所以对象存储的出现，很好地结合了块存储与文件存储的优点。

**最后为什么对象存储兼具块存储与文件存储的好处，还要使用块存储或文件存储呢？**

1、有一类应用是需要存储直接裸盘映射的，例如数据库。因为数据库需要存储裸盘映射给自己后，再根据自己的数据库文件系统来对裸盘进行格式化的，所以是不能够采格式化为某种文件系统的存储的。此类应用更适合使用块存储。

2、对象存储的成本比起普通的文件存储还是较高，需要购买专门的对象存储软件以及大容量硬盘。如果对数据量要求不是海量，只是为了做文件共享的时候，直接用文件了，性价比高。



龙脊梯田旅游攻略

龙脊梯田

想对作者说点什么？

我来说一句

逃离地球的小小呆： 汇集的很好，博主辛苦了。 （1个月前 #1楼）

**文件存储，块存储，对象存储的区别**

2566

文件存储 设备：FTP、NFS服务器 特点：一个大文件夹，大家都可以获取文件 优点：可以共享 缺点：传输速率低 块存储 设备：cinder，硬盘 特点：分区...

**对象存储Object，分布式文件存储NAS，分布式块存储（ServerSAN）**

4476

差异点\产品 ServerSAN 分布式NAS 分布式对象存储 接口协议 块（SCSI协议） 文件（NFS、CIFS协议） ...



多用户商城系统解决方案

百度广告

**块储存、对象存储、文件存储的区别和联系**

1.6万

块储存、对象存储、文件存储的区别和联系 通常来讲，队友磁盘阵列都是基于Block块的存储，而所有的NAS产品都是文件级存储。 1. 块存储：DAS SA...

**块存储、文件存储、对象存储这三者的本质差别是什么？**

1496

本质是一样的，底层都是块存储，只是在对外接口上表现不一致，分别应用于不同的业务场景。分布式存储的应用场景相对于其存储接口，现在流行分为...