

云存储架构详解

云存储 (cloud storage) 这个概念一经提出,就得到了众多厂商的支持和关注。Amazon 在两年前就推出的 Elastic Compute Cloud(EC2 :弹性计算云)云存储产品,旨在为用户提供互联网服务形式同时提供更强的存储和计算功能。内容分发网络服务提供商 CDNetworks 和业界著名的云存储平台服务商 Nirvanix 发布了一项新的合作,并宣布结成战略伙伴关系,以提供业界目前唯一的云存储和内容传送服务集成平台。

半年以前,微软就已经推出了提供网络移动硬盘服务的 Windows Live SkyDrive Beta 测试版。近期,EMC 宣布加入 道里可信基础架构项目,致力于云计算环境下关于信任和可靠度保证的全球研究协作,IBM 也将云计算标准作为全球备份中心的 3 亿美元扩展方案的一部分。

云存储变得越来越热,大家众说纷纭“云”,而且各有各的说法,各有各的观点,那么到底什么是云存储?

云状的网格结构

云存储在云计算 (cloud computing) 概念上延伸和发展出来的一个新的概念。云计算是是分布式处理 (Distributed Computing)、并行处理 (Parallel Computing) 和网格计算 (Grid Computing) 的发展,是透过网络将庞大的计算处理程序自动分拆成无数个较小的子程序,再交由多部服务器所组成的庞大系统经计算分析之后将处理结果回传给用户。通过云计算技术,网络服务提供者可以在数秒之内,处理数以千万计甚至亿计的信息,达到和“超级计算机”同样强大的网络服务。

云存储的概念与云计算类似,它是指通过集群应用、网格技术或分布式文件系统等功能,将网络中大量各种不同类型的存储设备通过应用软件集合起来协同工作,共同对外提供数据存储和业务访问功能的一个系统。如果这样解释还是难以理解,那我们可以借用广域网和互联网的结构来解释云存储。

相信大家对局域网、广域网和互联网都已经非常了解了。在常见的局域网系统中,我们为了能更好地使用局域网,一般来讲,使用者需要非常清楚地知道网络中每一个软硬件的型号和配置,比如采用什么型号交换机,有多少个端口,采用了什么路由器和防火墙,分别是如何设置的。系统中有多少个服务器,分别安装了什么操作系统和软件。各设备之间采用什么类型的连接线缆,分配了什么 IP 地址和子网掩码。

但当我们使用广域网和互联网时,我们只需要知道是什么样的接入网和用户名、密码就可以连接到广域网和互联网,并不需要知道广域网和互联网中到底有多少台交换机、路由器、防火墙和服务器,不需要知道数据是通过什么样的路由到达我们的电脑,也不需要知道网络中的服务器分别安装了什么软件,更不需要知道网络中各设备之间采用了什么样的连接线缆和端口。

广域网和互联网对于具体的使用者是完全透明的，我们经常用一个云状的图形来表示广域网和互联网。

虽然云状的图形中包含了许多许多的交换机、路由器、防火墙和服务器，但对具体的广域网、互联网用户来讲，这些都是不需要知道的。这个云状图形代表的是广域网和互联网带给大家的互联互通的网络服务，无论我们在任何地方，都可以通过一个网络接入线缆和一个用户、密码，就可以接入广域网和互联网，享受网络带给我们的服务。

参考云状的网络结构，创建一个新型的云状结构的存储系统系统，这个存储系统由多个存储设备组成，通过集群功能、分布式文件系统或类似网格计算等功能联合起来协同工作，并通过一定的应用软件或应用接口，对用户一定类型的存储服务 and 访问服务。

当我们使用某一个独立的存储设备时，我们必须非常清楚这个存储设备是什么型号，什么接口和传输协议，必须清楚地知道存储系统中有多少块磁盘，分别是什么型号、多大容量，必须清楚存储设备和服务器之间采用什么样的连接线缆。为了保证数据安全和业务的连续性，我们还需要建立相应的数据备份系统和容灾系统。除此之外，对存储设备进行定期地状态监控、维护、软硬件更新和升级也是必须的。

如果采用云存储，那么上面所提到的一切对使用者来讲都不需要了。云状存储系统中的所有设备对使用者来讲都是完全透明的，任何地方的任何一个经过授权的使用者都可以通过一根接入线缆与云存储连接，对云存储进行数据访问。

云存储简易架构图

经常看到人们在谈论云存储，但是没看过实际的图，人们很难想象到底云存储是什么模样，下面就是一个云存储的简易架构图：

橘色的存储节点 (storage node) 负责存放文件，蓝色的控制节点 (control node) 则是作为文件索引，并负责监控存储节点间容量及负载的均衡，这两个部分合起来便组成一个云存储。存储节点与控制节点都是单纯的服务器，只是存储节点的硬盘多一些，存储节点服务器不需要具备 RAID 的功能，只要能安装 Linux 即可，控制节点为了保护数据，需要有简单的 RAID level 01 的功能。

每个存储节点与控制节点至少有 2 片网卡 (千兆、万兆卡都可以，有些也支持 infiniband)，一片网卡 internal 负责内部存储节点与控制节点的沟通、数据迁移，一片 external 负责对外应用端的数据读写，一片千兆卡，读可以达到 100MB，写可以达到 70MB，如果你觉得对外一片网卡不够，也可以多装几片。

上面灰色的方块 (NFS、HTTP、) 是应用端，左上角的灰色方块 (mgmt console) 是一台 PC，负责云存储中存储节点的管理。对应用端看来，云存储只是个文件系统，而且一般来说支持标准的协议，例如 NFS、HTTP、等等，所以很容易把旧有的系统与云存储结合，应用端不需要作什么改变。

云存储不是要取代现有的盘阵，而是为了应付高速成长的数据量与带宽而产生的新形态存储系统，因此云存储在设计时通常会考虑以下三点：

1. 容量、带宽的扩容是否简便

扩容是不能停机，会自动将新的存储节点容量纳入原来的存储池，不需要做繁复的设定

2. 带宽是否线形增长

使用云存储的客户，很多是考虑未来带宽的增长，因此云存储产品设计的好坏会产生很大的差异，有些十几个节点便达到饱和，这样对未来带宽的扩容就有不利的影响，这一点要事先弄清楚，否则等到发现不符合需求时，已经买了几百 TB，后悔就来不及了

3. 管理是否容易

不说 google 有五万台存储服务器，即使国内也有很多客户有超过 500 台存储的，若不使用云存储来统一管理，管理 500 台存储是一个巨大的工作，一不小心就可能某些应用的崩溃，因此云存储的应用是一个必然的趋势，当用户把应用迁移到云存储，他管理的就是一台存储，而不是 500 台甚至五万台存储。管理一台存储不容易出错，分别管理五万台要不出错就很难了

上面我介绍的是一个纯软件的云存储解决方案，有的产品是硬件的解决方案，他们把橘色的存储节点和蓝色的控制节点，放在一台设备上，这样做的缺点是成本比较高，客户也不能够按照自己的需求，任意选择适合自己规格的硬件，例如读写性能、网卡、硬盘容量等等，因此我个人观点觉得软件的解决方案会成为最后的赢家，因为以云存储使用者的角度来看，他们对成本的要求很高、也不希望放弃他们原有的硬件投入，这些都是硬件的解决方案无法满足的。

云存储的结构模型

与传统的存储设备相比，云存储不仅仅是一个硬件，而是一个网络设备、存储设备、服务器、应用软件、公用访问接口、接入网、和客户端程序等多个部分组成的复杂系统。各部分以存储设备为核心，通过应用软件来对外提供数据存储和业务访问服务。

云存储系统的结构模型由 4 层组成。

一、存储层

存储层是云存储最基础的部分。存储设备可以是 FC 光纤通道存储设备，可以是 NAS 和 iSCSI 等 IP 存储设备，也可以是 SCSI 或 SAS 等 DAS 存储设备。云存储中的存储设备往往数量庞大且分布多不同地域，彼此之间通过广域网、互联网或者 FC 光纤通道网络连接在一起。

存储设备之上是一个统一存储设备管理系统，可以实现存储设备的逻辑虚拟化、多链路冗余管理，以及硬件设备的状态监控和故障维护。

二、基础管理层：

基础管理层是云存储最核心的部分，也是云存储中最难以实现的部分。基础管理层通过集群、分布式文件系统和网格计算等技术，实现云存储中多个存储设备之间的协同工作，使多个的存储设备可以对外提供同一种服务，并提供更大更强更好的数据访问性能。

CDN 内容分发系统、数据加密技术保证云存储中的数据不会被未授权的用户所访问，同时，通过各种数据备份和容灾技术和措施可以保证云存储中的数据不会丢失，保证云存储自身的安全和稳定。

三、应用接口层：

应用接口层是云存储最灵活多变的部分。不同的云存储运营单位可以根据实际业务类型，开发不同的应用服务接口，提供不同的应用服务。比如视频监控应用平台、IPTV 和视频点播应用平台、网络硬盘引用平台，远程数据备份应用平台等。

四、访问层：

任何一个授权用户都可以通过标准的公用应用接口来登录云存储系统，享受云存储服务。云存储运营单位不同，云存储提供的访问类型和访问手段也不同。