

STORAGE

存储经理人

2009 年 10 月刊

快照技术指南



常见的备份问题
其实都可以用快照来解决。但是，我们也不能单纯的将快照视为解决所有问题的灵丹妙药！

更多精彩内容：

- ◇ 云存储：谨而慎之
- ◇ 存储终极版：重振采购计划
- ◇ 剖析内部存储云
- ◇ *Storage Decisions* 专题报道

目 录

刊首寄语

经济危机探底 存储经理人需要思考什么..... 3

总编看存储

云存储：谨而慎之..... 4

技术解析

全面解析基于快照的数据备份..... 6

SPC-1 测试对存储设备采购的指导意义..... 36

业界观察

存储终极版：重振采购计划..... 17

技巧剖析

剖析内部存储云..... 26

行业动态

NFS 4.1 版的 pNFS：巨大的 NAS 性能提升... 33

存储调查

邮件归档的需求日益增长..... 35

专题报道

Storage Decisions 2009..... 49

我们的编辑团队..... 54

联系我们..... 57



经济危机探底 存储经理人需要思考什么

上个月 TechTarget 在纽约主办了“存储决策大会”（Storage Decisions Conference）。该大会每年分别在美国、加拿大、欧洲主要城市举办，场场爆满，是存储业界非常重要和权威的盛会。

2008 年开始席卷全球的经济危机给各行各业带来了严重打击，面对着成倍增长的数据和紧缩的预算，存储经理人该思考些什么？在本次会议上，来自最终用户企业的存储经理人齐聚一堂，聆听存储分析师的行业分析，共同探讨今年最实用的存储技术。除了传统的备份、容灾、容量管理等话题外，与会者还重点关注数据保留、法规遵从以及云存储等话题。FCS 咨询公司信息管理经理 Barclay Blair 认为，作为存储经理人，就有责任帮助企业认识到技术的发展现状和局限性，并且遵守法律的规定。在经济回暖之际，存储经理人也应该更多地考虑未来。

在中国未尝不是这样。中国版塞班斯法案的推出使很多企业已经认识到了数据保留、归档的重要性。尤其是一些上市公司和大中型企业已经走在了前列，开始按照《企业内部控制基本规范》的相关规定对数据进行存储备份，并对邮件实施了归档。对于更多的存储经理人来说，应该进行长远打算，对这部分数据预留存储空间，并且开始考虑选择归档软件了。

“云”这个字眼大家恐怕已经不再陌生了。“云”绝不代表云山雾罩，它只是分布式计算的一种，是本地归档存储的一种廉价的替代方案。虽然云目前有种种缺点，如可靠性、安全性、厂家保证等等，但只要对这些问题进行规范和解决，云将势不可挡的成为未来存储的一项重要选择。将来，越来越多的企业将会把数据存储这部分业务外包给内部私有云厂商、公共云服务供应商或公共云应用程序供应商。存储经理人要做的事不是关注“云”本身的概念，而是要更多地思考当前的“云”存在的优势和劣势，自己的企业是否适合部署云存储，如何与云服务商进行沟通，关注价格、数据安全责任等问题，以便更加从容地享受云带来的种种益处。

存储经理人身上的担子还很重，但经济的回暖也会给存储经理人更大的施展空间！



吴迪，TechTarget 中国存储编辑

云存储：谨而慎之

云存储即将来临，毫无疑问将带来好处，但仍然要考虑种种影响。

在被指责为厌恶技术或两耳不闻窗外事之前，我想先声明一下，我非常支持创新，热爱技术。但与此同时，我也不会盲目的随波逐流。就存储而言，虽然我相信将来在很多存储部门中，云存储将发挥一定的作用，但我也认为我们应该对此多加思考，全面考虑云存储许诺的新存储世界将带来的变化。

我从存储经理们那里了解到，他们不愿意把数据放到云提供商的站点。的确，这是让人不放心。不过我认为，这个障碍相对会很快得以清除。多年来，很多 IT 部门其实一直都在使用第三方提供的灾难恢复服务。但是，使用第三方的设备来存储第三份或第四份数据备份以备紧急之用，与使用第三方设备进行主存储或甚至进行备份大相径庭。

还有一些其他的问题已经被提出，但可能还没有得到应有的重视，例如对更快速的通信网络的需求。使用云存储可以节省费用，但为此需要搭建电信基础设施，由此产生的费用是否有可能与节省下来的费用相持平，甚至大大超出？

使用云存储可以节省费用，但为此需要搭建电信基础设施，由此产生的费用是否有可能与节省下来的费用相持平，甚至大大超出？

我认为，更加困难的是对与存储相关的服务进行商品化。基于云存储的服务，无论可能提供多少的定制服务，都不可能像内部安装一样，完全根据公司的特定需求进行修改和调整。但是，对管理层来说，这些服务的相对低成本具有吸引力。而且，这可以让公司减少对 IT 人员专业技能的需求，这样的诱惑也很难抵挡。IT 的传统角色可能发生很大的变化，逐步转变为外部服务的交换所。再者，云存储带来的人员和资本支出的节省可能极具诱惑力。但是，如果又重新需要那些逐渐消失的专业技能时，那将怎么办？很可能那些特殊的技巧和专长也将转入云中。也许有人会对此有所异议：以前也曾出现过类似的变革，但 IT 还是幸存下来，巍然不动。在许多公司中，应用开发团队已经演变为对市场流行的软件进行评估或定制，公司内部越来越少地进行实际的开发工作。但是存储需要的是不一样的

关注和实施。现有的存储解决方案种类繁多，应用到独特环境中的方式也各异，我认为任何大规模的商品化不会带来益处，或者不太现实。

这又引出另一个观点：如果公司的存储转移出来，转入云中，可能不仅仅是该公司的数据中心配置受到影响。所有存储系统的设计和制造可能发生较大的变化。云服务提供商使用的存储系统与公司的不尽相同。这似乎又回到了十年前左右，第一次推出“存储作为服务”时的情景：把大型的企业级阵列划分给一些客户，以节省成本。这样的方法是永远行不通的。今天的云提供商倾向于把他们的存储系统集中起来，形成更加经济实用、类似于公共设施的配置。与企业的存储系统相比，这些安装看上去不复杂。从云中交付存储，将行得通；但是不太可能符合对性能要求高的内部系统。

因此，随着云存储的不断发展，是否存储厂商将把他们的精力转向为云提供商建设系统？在今年七月份《纽约时报》的一篇文章中，一名哈佛法学教授 Jonathan Zittrain 表达了对云计算的一些担忧，以及云计算可能对创新造成的影响：“但是，云最困难的挑战，既难于把握又难于解决，就是它对于我们创新自由的影响。”虽然这篇文章大部分是讨论个人电脑用户，但 Zittrain 最关心的是软件开发将从很多人的手中落到一些服务供应商的手中。软件是存储系统中的主要区别，可能类似的问题也会出现在与新存储产品的开发中。

Wikibon.org 的 David Floyer 在最近的 Wikibon 新闻邮件中也表达了类似的看法，不过更多的是从存储开发者的角度来考虑。他写道：“Wikibon 预测（在接下来的十年里），外部 IT 服务的总支出将超过大部分公司日趋减少的传统 IT 预算。”他继而描述存储和其他 IT 厂商的关注度可能或者应该如何进行转移，并且表示“向 IT 组织进行销售的传统模式将成为一个逐渐缩减的市场。”

我不知道 Zittrain 和 Floyer 是否提出了云存储和提供商永远无法回避的问题，或者他们只是看得比我们超前一些（至少比我看得远）。最终，公司的存储采购模式将任重道远，决定着描绘的愿景是否能够如愿以偿。

（作者：Rich Castagna 译者：谢小丽）

快照技术指南

全面解析基于快照的数据备份



快照技术通常用于增强数据备份系统功能，极大的缩短 RTO 和 RPO。但你也需要了解快照实施有哪些不同，而这些不同取决于 IT 环境。

Marc Staimer

“快照”通常被定义为一组文件、目录或卷在某个特定时间点的副本。“快照”这个名字的含义与“照片”相似，它所捕获的是一组特定数据在某个时间点的映像。

快照技术的出现最初是为了解决备份的难点问题，其中常见的包括：

- 数据量太大，以至于无法在有限的时间段内完成备份；
- 备份失败。例如在备份时，有数据从一个未备份的目录移动到已经备份过的目录；
- 备份的数据不能使用。因为备份时该文件正在进行写操作；
- 热备份严重影响应用系统的性能，等等。

所有上述常见的备份问题其实都可以用快照来解决。但是，我们也不能单纯的将快照视为解决所有问题的灵丹妙药，因为快照技术还有待进一步完善。（见“快照的一致性问题”）

创建一个快照需要一系列步骤：

- 1、首先发起创建指令；
- 2、在发起时间点，指令通知操作系统暂停应用程序和文件系统的操作；
- 3、刷新文件系统缓存，结束所有的读写事务；
- 4、创建快照；
- 5、创建完成之后，释放文件系统 and 应用程序，系统恢复正常运行。

现在，快照技术已经超越了简单的数据保护范畴。我们可以用快照进行高效且无风险的应用软件测试。用快照数据做测试，不会对生产数据造成任何的破坏。对于数据挖掘（data mining）和电子发现（eDiscovery）应用，快照也是理想的数据源。在灾难恢复方面，快照是一种非常有效的方法 -- 甚至是首选，非常适合恶意软件攻击、人为误操作和数据损坏等逻辑错误的恢复。

现在，快照技术已经超越了简单的数据保护范畴。我们可以用快照进行高效且无风险的应用软件测试。用快照数据做测试，不会对生产数据造成任何的破坏。

哪些系统集成了快照功能？

过去我们认为只有磁盘阵列具备快照功能，但事实上磁盘阵列只是其中之一而已。广义的快照技术通常可有 7 个不同类型的实现主体：

- 1、主机文件系统（包括服务器、台式机、笔记本电脑）

- 2、逻辑卷管理器 (LVM)
- 3、网络附加存储系统 (NAS)
- 4、磁盘阵列
- 5、存储虚拟化装置
- 6、主机虚拟化管理程序
- 7、数据库

1、基于文件系统的快照

很多文件系统都支持快照功能，微软的 Windows NTFS 有 VSS 卷影复制服务 (Volume Shadow Copy Services , Vista 称作 Shadow Copy) ； Sun Solaris 的最新文件系统 ZFS (Zettabyte File System) ； Apple 公司的 Mac OS X 10.6 (雪豹) ； Novell NetWare 4.11 (或更高版本) 的 Novell Storage Services (NSS) ； Novell SUSE Linux 操作系统下的 OES-Linux 等等。

“免费”是文件系统快照的优势之一，因为它集成在文件系统内部；另一个优点是好用，最新版文件系统的快照功能通常使用起来很简单。不利的一方面是，每个文件系统都必须独立进行管理，当系统数量激增时，

“免费”是文件系统快照的优势之一，因为它集成在文件系统内部。

管理工作会变得非常繁重。想象一下，如果我们要做快照复制的话，需要给每一个文件系统都配置一套复制关系，而且还只能复制该文件系统自己的快照。此外，不同文件系统所提供的快照种类、快照频率、预留空间等参数也可能不一样，当然也包括设置、操作和管理上的差异。总之，需要管理的服务器和文件系统越多，复杂程度就越高。

2、基于 LVM (逻辑卷管理器) 的快照

带有快照功能的 LVM 也很多，比如惠普 HP-UX 操作系统的 Logical Volume Manager ； Linux 平台的 Logical Volume Manager 和 Enterprise Volume Management System ；微软 Windows 2000 及后续版本自带的

Logical Disk Manager ; SUN Solaris 10 操作系统的 ZFS ; 以及赛门铁克公司的 Veritas Volume Manager (注 : Veritas Volume Manager 是赛门铁克 Veritas Storage Foundation 产品的一部分) 。

我们可以创建跨多个文件系统的 LVM 快照。像赛门铁克的 Veritas Volume Manager 可以支持大多数常见的操作系统和文件系统。LVM 通常还包括存储多路径和存储虚拟化等功能。

使用 LVM 时，通常要付出额外的成本，包括为每台服务器购买 license (许可证) 和维护费。而且，像基于文件系统的快照一样，我们可能还要面对系统之间的协调问题和复杂的技术实施问题。

3、基于 NAS 的快照

NAS 本质上就是一个经过优化的、或是专门定制的文件系统，运行在特定的装置上，或集成在存储设备里。大多数中端和企业级 NAS 系统都提供快照功能，其中既有使用专有操作系统的设备，也包括大量基于 Microsoft Windows Storage Server 软件的各种 NAS。

通过网络连接到 NAS 的计算机系统都可以使用这种标准的通用快照，包括物理服务器、虚拟机、台式机和笔记本电脑。它也非常容易操作和管理。基于 NAS 的快照往往同 Windows Volume Shadow Copy Services (卷影复制服务 VSS)、备份服务器和备份 Agent 等软件集成在一起使用。一些 NAS 厂商还为非 Windows 平台的数据应用系统开发了 Agent 代理程序。其他一些与 NAS 快照有关的技术还包括重复数据删除 (EMC 公司，FalconStor 软件公司和 NetApp 的产品)，有些厂商甚至提供了带有自动精简配置功能的快照，目的是让快照占用的空间变得更少。

基于 NAS 的快照往往同 Windows Volume Shadow Copy Services (卷影复制服务 VSS)、备份服务器和备份 Agent 等软件集成在一起使用。

但是，使用便利的工具和附加功能也需要成本，软件 license 和维护费相当昂贵，一般是按照机器数量和磁盘卷容量来计算。大多数公司的数据量增长很快，需要使用 NAS 快照的地方也越来越多，因此，操作和管理也将更复杂。

快照的一致性问题

如果用快照来处理结构化数据，可能会存在一些问题。结构化数据涉及到数据库，以及数据库类应用（例如邮件系统、ERP 或 CRM 等等）。许多产品中的快照并不能与这些应用程序集成或被直接调用。有一种可能的情况是，在我们创建快照的瞬间，数据库恰好不在静止状态（缓存正在刷新、写操作事务尚未完成、索引和元数据正在更新等等），此刻生成的快照数据是不一致的，很有可能无法正常使用。

在微软的 Windows Server 平台上，这个问题要简单得多，利用 Windows Volume Shadow Copy Services (VSS) 和它的 API，数据库应用程序可以集成并调用快照工具。VSS 是专门为结构化数据应用设计的服务框架，可以驱动数据库等应用进入数据一致性的静止状态，在快照开始初始化之前，完成刷新缓存、结束写操作以及系统状态的更新。

遗憾的是，目前在 Linux 和 Unix 操作系统平台上还没有类似 VSS 的服务或 API。VMware 公司的 vCenter storage API 可以说是一个部分解决方案。快照的发起者可以通过 vCenter storage API 给 vCenter 发出一个指令，让虚拟机进入静止状态，然后再执行快照。但这个时候，快照由于没有通过应用程序感知，也许会存在不一致的问题。

这里还有一个好办法，可以不通过 Windows VSS，获得数据库的一致性快照。这个办法需要备份软件的配合。将快照的 API 同备份软件集成，就可以从备份服务器端驱动备份软件的数据库代理 Agent。Agent 备份代理程序可以驱动数据库进入静止状态，然后反向让备份服务器通知快照工具开始执行创建快照的操作。这也是一个比较有效的办法。

4、基于磁盘阵列的快照

大多数磁盘阵列的软件系统里都含有快照功能。基于磁盘阵列的快照与基于 NAS 的快照有非常相似的优点，即所有与磁盘阵列相连的计算机系统都可以使用这种标准的通用快照功能，包括物理服务器、虚拟机、台式机和笔记本电脑等等。快照的实施、操作和管理也都很简单。像 NAS 一样，很多磁盘阵列的快照功能也可以被 Windows VSS、备份服务器和备份 Agent 等软件直接调用。一些磁盘阵列厂商还有可供非 Windows 平台应用系统使用的 Agent 代理程序。

基于磁盘阵列的快照也有一些缺点：license 和维护费用昂贵；对非 Windows 平台的应用程序支持有限；磁盘阵列的数量越多，快照的管理也就越复杂。

5、基于存储虚拟化装置的快照

这里所说的存储虚拟化装置主要用于 SAN 光纤网络环境，不同于基于文件（NFS）应用的网络设备，像 F5 Network 公司的 Acopia ARX 产品就是排除在这个范畴之外的。主要的存储虚拟化软硬件装置（或融合了虚拟化功能的存储系统）包括：Cloverleaf Communication 公司的 Intelligent Storage Networking System (iSN)；DataCore Software 公司的 SANsymphony 和 SANmelody；EMC 的 Celerra Gateway blades；FalconStor 公司的 IPStor；HP 的 XP 系列存储；HDS 的 Universal Storage Platform V/VM；IBM 的 SAN Volume Controller；LSI 的 StoreAge Storage Virtualization Manager (SVM) 以及 NetApp 的 V-Series storage controllers 等等。

随着服务器虚拟化应用的普及，基于主机虚拟化管理软件（hypervisor）的快照技术也逐渐流行起来。

磁盘阵列和 NAS 快照所具备的优点在存储虚拟化装置上同样能够体现，而且某些方面还能做的更好。我们可以将来自不同厂商的很多存储设备聚集在少量的几个控制点或单一控制点上进行管理，提供通用的标准化快照。这样做最大程度的简化了快照的管理操作成本和学习成本。

存储虚拟化快照的缺点与上述类型相比则有些不同。使用存储虚拟化装置会导致 I/O 延迟的增加，即使是采用旁路架构的设计，最终还是会影晌应用程序的响应时间。增加存储虚拟化装置还会使故障分析变得更加困难，潜在的还可能激化厂商之间对故障责任的推诿。从另一个角度看，虽然增加额外的虚拟化存储硬件或软件要产生一定的费用，但是与每个存储系统都独立购买快照功能相比，它的软件 license 和维护费用都要低一些。

6、基于主机虚拟化软件的快照

随着服务器虚拟化应用的普及，基于主机虚拟化管理软件（hypervisor）的快照技术也逐渐流行起来。像 Citrix 公司的 XenServer、微软的 Hyper - V、

SUN 的 xVM Ops Center、以及 VMware 的 ESX 和 vSphere4 等主机虚拟化产品都支持快照功能。

在主机虚拟化软件层实现快照的优点是简单直接。由于同虚拟机管理软件绑定在一起，因此可以为所有的虚拟机（VMs）提供统一的快照，并且还可以同微软的 VSS 集成，随时调用。相对而言，基于虚拟机的快照很容易部署、使用和管理。

但是，如果非要找出不喜欢这种快照的理由？我想应该是每一套虚拟机软件的快照需要单独管理；而且当我们在非 Windows 平台下使用这种快照技术时，必须针对整个 VM。这意味着我们只能做粗粒度的数据恢复，还要消耗更多的恢复时间。这种快照是在 Windows 操作系统外部创建，所以它不能架构在应用软件感知的层面，导致快照出来的映像数据有可能是不一致状态。

7、基于数据库的快照

在数据库中，快照动作被称为“snapshot isolation（快照隔离）”。像 Oracle 和 PostgreSQL 这样的数据库需要做快照隔离以确保所有的交易命令序列化，就好像被一个个隔开一样，然后再逐个执行。其他的一些数据库也支持快照隔离，但并不要求将交易序列化。在一般情况下，数据库备份工具会利用快照隔离的功能，用快照来恢复崩溃（出现一致性问题）的数据表。

针对数据库内部数据和基于该数据库的相关应用，使用数据库自带的快照比较有效。

相反，数据库快照的重要缺欠就是覆盖的范围非常有限，其作用仅限于特定的数据库内部和数据库相关的应用，无法管理同在一台服务器上的文件系统、文件类应用或其他数据库，更不用说管理到其他的服务器了。有时候我们不得不通过其他层次的快照技术来解决数据库之外的数据保护问题，这样，操作和管理将变得有些复杂。

不同类型的快照及工作原理

通常，我们会提到 6 种类型的快照技术：

1、Copy-on-write 复制写

- 2、Redirect-on-write 重定向写
- 3、Clone or split mirror 克隆或镜像
- 4、Copy-on-write with background copy 后台拷贝的复制写
- 5、Incremental 增量快照
- 6、Continuous data protection 持续数据保护

1、Copy-on-write (COW) 复制写快照

COW 快照需要消耗一些存储空间--建立快照卷。当我们为一个数据卷创建一个快照之后，这些预留的空间用来存放被变化数据更新的旧数据。COW 快照在初始化的过程中仅仅创建用来描述源数据块位置的指针信息（元数据），而不是完整的将源数据块拷贝过来。因此初始化的过程几乎可以在瞬间完成，对系统的影响也很小。

COW 快照会跟踪数据卷的写操作和数据块变化。当某个数据块发生改变时，在将旧的数据覆盖之前，首先将该块的旧数据复制到预留的快照卷，该步骤仅在数据卷相应数据块位置发生第一次写操作请求时进行。这个处理过程确保快照出来的数据与发起快照的那个精确时间点保持完全一致。这个过程也描述了“copy on write”这个名字的含义。

如果我们需要访问某个时间点的快照数据，对没有改变过的块直接从数据卷读取；对已经改变并被复制的块则从快照空间读取。从快照被创建那一刻开始，每个快照都会跟踪记录描述块改变的元数据信息。

COW 快照的主要优势在于空间的高效利用，因为快照卷只需要保留发生过变化的数据块，与数据卷相比要小得多。但是我们也知道 COW 快照有个缺点，它会引起数据卷性能的下降，这是因为创建快照之后，对数据卷的写操作会增加一个等待的过程 --即旧数据块复制到快照卷的过程。另外一个关键问题是每个快照卷必须依赖一个完整的数据卷。

*但是我们也知道 COW 快照
有个缺点，它会引起数据卷
性能的下降.....*

2、Redirect-on-write (ROW) 重定向写快照

“ROW 重定向写”与“COW 复制写”是相对的概念，它可以避免两次写操作引起的性能损失。ROW 同 COW 一样在空间利用方面效率非常高。那是什么让 ROW 快照避免了写性能的损耗？其中的原因是 ROW 把对数据卷的写请求重定向给了快照预留的存储空间，而写操作的重定向设计则把需要两次写才能完成的操作减少为一次写。我们知道 COW 的两次写包括：1、将旧数据写入快照卷；2、在数据卷写入新数据。而 ROW 只有写入新数据一步。

使用 ROW 快照，数据卷存放的是上一个快照时间点的旧数据，新数据最终存放在预留的快照空间。这里也有一个复杂的问题，就是快照的删除。被删除的快照上的数据必须被复制到原始数据卷，并且做一致性回退。创建的快照越多，维护快照的复杂度也会以指数级别上升。这些复杂性包括对原始数据的访问、快照数据和原始数据卷的跟踪、以及快照删除后的数据调整。另一个直接引发的严重问题是，原始数据集中会产生大量的碎片。

3、Clone or split-mirror 克隆或分割镜像快照

Clone（或 split-mirror）快照所创建的是数据的完整副本。Clone（或 split-mirror）快照的对象可以是一个存储卷、一个文件系统或者是一个 LUN（logical unit number 逻辑单元号）。

	快照类型					
	Copy-on-write (COW)	Redirect-on-write (ROW)	Clone or split-mirror	COW with background copy	增量备份	连续数据保护 (CDP)
快照是否依赖于数据卷	是	是	否	仅在初始化过程中复制数据卷	取决于第一次快照的类型	否
快照是否节约空间	是	是	否	否	否	是，与多个快照累计相比，CDP 更加节省空间
系统的IO和CPU开销	高	中	低	低	低	低
数据卷的写开销	高	无	无	高	高	高
支持对逻辑错误的恢复	是	是	是	是	是	是
对数据卷的物理保护	否	否	是	是 前提是后台复制完成	取决于第一次快照的类型	是

Clone 快照的优点是它们具有高可用性；缺点是所有的数据都要完整的复制一份，复制的过程也不可能在瞬间完成。我们可以分割一对保持同步状态的镜像卷来启用 Clone 快照，分割的过程瞬间即可完成。然而，当镜像被分割成 Clone 快照之后，数据卷也就失去了他的同步镜像。

使用 Clone 快照需要面对的一个非常严重的问题是每个快照都需要和数据卷一样大的存储空间。尤其是当我们在任何时刻都需要保持一份以上 Clone 卷的情况，这个成本会非常高。另一个缺点是影响性能，因为在镜像卷之间保持写同步需要一定的系统开销。

4、Copy-on-write with background copy 后台拷贝的复制写快照

Copy-on-write with background copy 快照有两个生成步骤：首先创建一个瞬时即可生成的 COW 快照；然后利用后台进程将数据卷的数据复制到快照空间，最后生成一份数据卷的克隆或镜像。

创建这种快照的目的是发挥 COW 快照的优势，同时尽量屏蔽它的不足。因此，这种快照常常被形容为 COW 和 Clone 快照的混合体。

5、增量快照

增量快照的特点是可以跟踪数据卷和快照卷的变化。当一个新的增量快照生成之后，旧的快照数据将被刷新。第一个快照和随后创建的每一个增量快照数据上都有时间戳标记，利用时间戳我们能够将快照数据回滚到任意的一个时间点。增量快照技术能够加快后续快照的生成速度，而且仅仅在名义上多消耗了一点空间而已。由此，我们可以提高创建快照的频率，也能让快照保留得更久一点。

增量快照的不足之处是它需要依靠上面所提到的其他基础技术来创建第一个快照（COW、ROW、clone/split mirror、copy-on-write with background copy）。如果用 Clone 方式，那么第一个快照需要较长的初始化时间；如果用 COW 方式，数据卷的性能会降低。

6、持续数据保护（CDP）

CDP 的出现是为了实现零数据丢失的 RPO 指标，以及瞬时数据恢复的 RTO 指标。它本身与同步数据镜像很类似，不同之处在于 CDP 还可以对软性灾难进行恢复。包括人为误操作、恶意软件攻击、意外删除、数据损坏等情况。

持续数据保护颇像频率很高的增量快照。它会捕获并复制任何时刻发生的数据变化，并且给这些数据块打上时间戳。CDP 本质上相当于每个时刻都创建一份增量快照，提供细粒度的精确数据恢复。有些 CDP 产品同时提供基于时间和基于事件（例如应用程序升级事件）两种粒度的恢复方式。还有一个理解 CDP 概念的好方法就是将它看成一个快照的 journal 日志。

对于邮件系统、数据库和基于数据库的应用来说，CDP 是一个极好的保护方案，能将数据回滚到任意的历史时间点，恢复过程也简便、迅速。最有代表性的 CDP 产品是飞康公司的 IPStor，它是一个集成了 CDP 功能的存储系统兼存储虚拟化装置。

持续数据保护颇像频率很高的增量快照。它会捕获并复制任何时刻发生的数据变化，并且给这些数据块打上时间戳。

随着越来越多的数据需要保护，备份窗口也变得越来越紧张，因此需要快照技术来帮助我们解决备份问题。在现实的应用环境中，快照利用的是否恰当对数据保护的等级和恢复的速度有着很大的影响。尽管各类型快照之间存在的技术差异不太容易理解，但无论如何，快照技术都将在数据保护领域和日常存储管理中扮演重要的角色。

（作者：Marc Staimer 译者：郭镭）



存储终极版

重振采购计划

当前的经济危机还没有结束，但存储经理人将会看到一个更加光辉的未来。

不存储经理人可能终于看到了经济隧道终点的一线光明，不过像老笑话说的，那也可能是迎面过来的火车。不过经历过像火车碾压过一样的经济形势之后，貌似存储预算可能会恢复到貌似正常的水平。当然，“正常”是相对的，因为这些预算仍然徘徊在负数范围。

在上个春天经济衰退的全面影响开始展现的时候，存储经理人也和几乎所有其它行业一样被刺痛得步履蹒跚。在我们之前的存储采购意向调查中，存储经理们反馈的数据显示 2009 年存储预算可能会比 2008 年低 1.9%。尽管这看起来不像一个很大幅度的缩减，但是这个数据相对于他们在 2008 年秋季的数据要低将近 5 个百分点。而且是 7 年来我们第一次让存储经理人把他们的预算和上年相比时看到负增长。在我们 9 月初做的最新调查中，情况乐观了一点。和 2008 年

相比，存储预算只有微小的 0.4% 的降低，虽未回到正值，但肯定比 6 个月前要积极多了。总的来说，调查中的 30% 说预算减少了，28% 说预算持平，而接近 31% 说他们的预算有所增加。

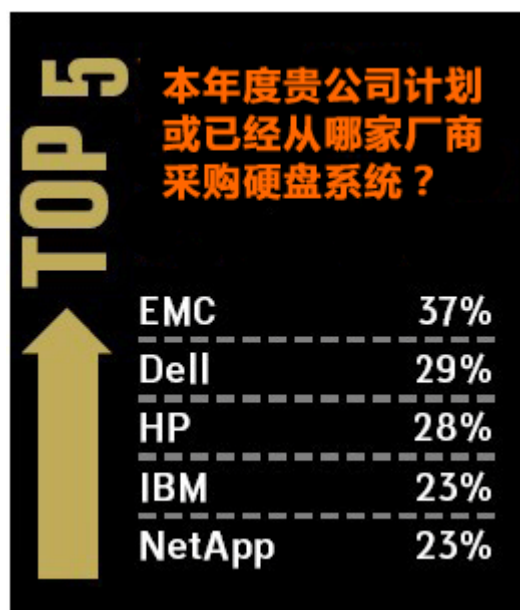
从春天到秋天的预算水平和采购计划的转变是常见现象。经理们在春天预计他们一年的开销的时候会显得比较小心；到了秋天，很多经理会发现他们的预算增加了一些。新的项目，计划外的开支和新的经费来源都是这种年度摇摆的原因。单纯从金额上看，平均报告的存储预算为 2 百 90 万美金。和去年春天的数字相同。一些巨大的预算拉高了平均值；57% 的返回数据的受访者预算少于 1 百万美金。。

有关调查

Storage 杂志/SearchStorage.com 的采购意向调查每年进行两次；今年是进行调查的第 7 个年头。Storage 杂志的读者受邀参加旨在收集存储经理们对各种存储产品类别的采购计划信息的调查。本次有 826 个有效的反馈，分布在各种行业，平均公司大小为年营业额 15 亿美金。

仍然小心并要应对容量

调查结果暗示，相对于大刀阔斧地削减存储预算，存储经理们更倾向于采取“这修修，那剪剪”的方式，并利用新技术的实施来抑制成本。



不论预算有多大，预算分配的比例都和我们历年来看到的非常相似，磁盘系统的开销占掉总预算中最大的一块（39%）。人们可能会期望随着磁盘价格在这几年的大幅下降，磁盘占有的比例应该成比例地减少。但是，这里需要乘上的系数是对新容量的需求；随着容量需求的飙升，磁盘价格降低的因素被抵消了。

对新容量的需求从未松懈，尽管调查数据显示存储经理们本年度剩下的时间可能终于有机会松口气了。平均来说，每个公司会增加 34TB 的新磁盘容量，这是个

很大的数目，但是相比去年春天的 43TB 有明显的下降（尽管较大的企业还是准备平均增加 68TB）。看来存储经理们采用了不同的战术来对待容量需求，比如使用类似自动精简配置、压缩和重复数据删除之类的工具来更好地利用现有的装机容量，并试图避免采购新的。当然，他们或许只是能花的钱减少了，导致一些项目被推迟到明年或者更晚。

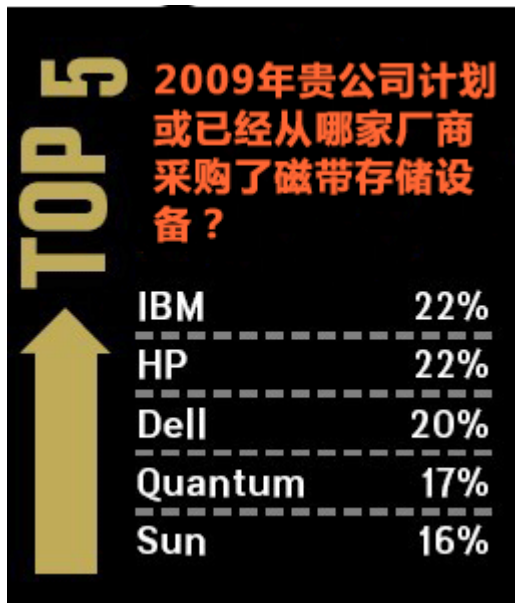
当存储经理购买磁盘的时候，他们更倾向于扩容现有的阵列，而不是购买全新的系统。他们中的 73% 说他们的主要磁盘支出会是在旧系统上增加新磁盘。我们在过去的三年内一直看到这种扩容替代新建的趋势；我们很乐意看看当已经安装的系统到寿或者租约到期的时候，这个趋势是否会重新回到新建系统上来。

对于那些打算购买新系统的人来说，中端产品占到了受访者购买计划的将近 50%。尽管光通道(FC)产品仍然最受欢迎，iSCSI 持续地缓增；12% 说他们今年会采购 iSCSI，相比去年春天有几个点的提升，而其它磁盘系统则持平或小幅减少。

iSCSI 系统正变成各种大小公司数据中心的不可或缺的一部分。在我们的受访者中，平均已安装的磁盘容量为 75TB，而其中的大部分是光通道 SAN（62%）或者网络存储（NAS）系统（63%）。而 35% 的受访者说他们正在使用 iSCSI 存储，相比去年春天和秋天的数据有几个点的增加。而且这一数字看起来将会继续增加，因为有 43% 受访者说他们今年计划或者已经部署了 iSCSI 系统。这比去年秋天有大约 3% 的增加，而且是我们曾经见过的最高比例。iSCSI 系统的使用方式也表明这一技术业已成熟；47% 说他们会在关键业务应用中使用他们的 iSCSI 系统，为历史最高数据。对小型公司来说，iSCSI 扮演了更重要的角色，有 61% 在关键应用中使用它（同为历史新高）。

在受访者 2009 年采购的系统来自的厂家中。EMC 仍居榜首，但是它领先 Dell 和 HP 的距离缩小了几个点；而前五名的另外两个是 IBM 和 NetApp。选择某个特定磁盘供应商的主要原因仍然是产品提供的功能和特性（29%），但是当前的经济形势还是影响了选择条件的。第二个经常被提到的选择标准是公司是否正在使用该厂商的其它产品（23%）；信心指数明显具有影响力，但是也有可能是由于厂家会为现有客户提供特别优惠。价格因素一直是选择标准里面影响力比较弱的，但是现在增加到了 16%，只比技术支持少一个百分点。

新的存储技术找到了立足之地



固态存储是现在的存储世界里的红人，吸引了新技术领域最多的关注。对很多人来说，尽管固态存储仍然还受糟糕的“太”之困扰——太新、测试太少和太贵。但是我们本期的调查显示了一些有趣的结果：8%受访者说他们正在使用固态硬盘（SSD），而另外有3%说他们今年打算实施。而35%在评估这一技术。

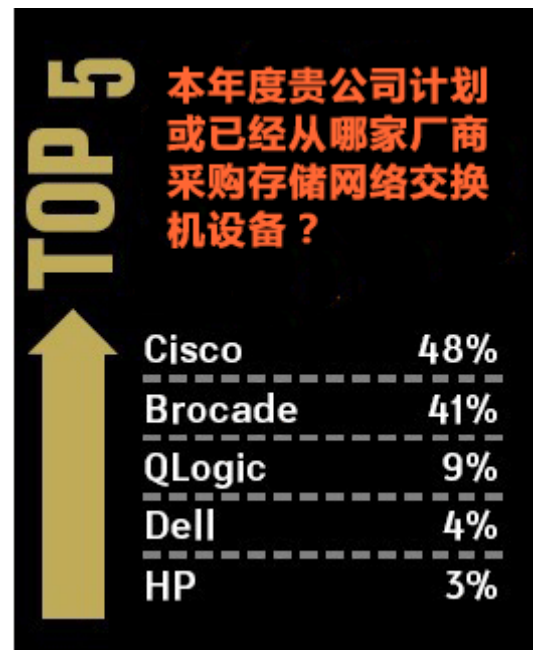
在正在使用 SSD 的人之中，48%的人将其安装在阵列里，25%把 SSD 作为服务器的直连存储，而14%说阵列和服务器中都有。显然，厂家为了对付它高每 GB 价格而采取的，更多地和磁盘进行性能比较的手段开始奏效了。而且 SSD 用户可能还会强调低功耗电量的好处，尽管节能并不是多数数据存储经理的首要考虑。39%的受访者说能效是选择阵列最重要或者主要条件，和去年秋天的数字基本一致。

高速存储网络协议也在将 SSD 和去重复推到存储舞台的聚光灯下，现在10Gbps 以太网和8Gbps 光通道已经可以买到了。尽管网络基础设施的改变通常是非常缓慢的，对这两个新协议还是有一些值得注意的动向。

目前，使用最广泛的存储网络协议是4Gbps 光通道（48%）和1Gbps 以太网（33%），但是8Gbps 光通道已经在14%的受访者处得到应用，而有13%说他们已经升级到10GbE。

虚拟化的价值

服务器虚拟化遍地开花，84%的受访者说他们虚拟化了他们全部或者部分的服务器。而服务器虚拟化对存储的影响已经变成很多这种系统的关键问题。49%受访者选择光通道作为他们的虚拟服务器存储，但是 iSCSI 的正变得越来越多。去年秋天，只有12%的受访者说他们把虚拟服务器接在 iSCSI 存储系统上；而这次，18%表示出对这一技术的肯定。在较小的公司中，对 iSCSI 的偏向更加明显，25%说他们在虚拟服务器中使用它。（直连存储



DAS 通常不被认为是虚拟服务器的一个可行存储方案，但是 11%受访者表示他们的选择是 DAS)

备份一直是虚拟服务器环境存储的最大痛处。44%的受访者说他们使用传统的备份软件和方法来备份他们的虚拟服务器，在每个虚拟机（VM）上放置代理（agents）。VMWare Consolidate Backup(VCB)被吹嘘为节省软件授权成本的备份方法，却只有 20%的受访者使用——这是个有些令人意外的数据，因为 VCB 曾受到如此广泛的赞誉。针对虚拟机的备份产品，例如 PHD 虚拟技术的 esXpress，Veeam 软件的 Veeam Backup & Replication 以及 Vizioncore 公司的 vRanger Pro 也曾吸引了很多的注意力，但是只有 6%的受访者所他们在他们的虚拟机备份中使用类似的产品。

据 27%的受访者说，用户备份他们的虚拟服务器时遇到的最大问题是备份了过多的数据，这可能和他们使用传统备份方法有关。23%说虚拟机备份太过复杂，而 17%指出对单个文件的访问（虚拟机备份的一个常见问题）是他们最头疼的问题。

尽管被服务器虚拟化从装机数量上远远超过，但是存储虚拟化仍然不可避免地在存储机房得到越来越多的重视。存储虚拟化相对较慢的发展的原因之一是，它比服务器虚拟化相对更难实施（当然也更贵），而且多数的虚拟化选择都带有不祥的“厂家锁定”。

然而，31%说他们至少部分的存储上已经有虚拟化，高于去年秋天的 26%——在不长时间内非常显著的 5%的增长。块存储仍然是虚拟化的初始选择，18%的受访者虚拟化了他们所有的块存储，而 61%指出他们虚拟化了部分的块存储。大约 8%的受访者虚拟化了他们所有的文件存储，而 44%虚拟化了部分的文件存储。考虑到虚拟化产品已经上市的时间，这些数字还是很低的，但是所有这些数字都比去年春季有所增加。

更高效的备份操作

节省空间的存储技术，包括最引人注目的重数据删除和压缩技术，是存储经理对付容量增加和预算缩减的效率法宝中的主要部分。



每年我们都看到磁带在多数公司的存储运营中的重要性越来越低，但是这个趋势也说明存储经理为了把数据压缩到他们能够管理的大小所做出的努力。去年春天，我们看到史上最低的数字，19%的受访者说他们将会增加磁带的使用而，反过来，一个最高的数字，29%说他们打算降低他们对磁带的依赖。那些数字在最新的调查中稍有缓和，但是趋势是很明显的，尽管 80%说他们所有或者一部分的备份数据最终会转移到磁带上（低于两年前的 86%）。

即便是那些在 2009 年计划或者已经买了磁带库的受访者，多数也倾向于较小的带库。他们打算购买的磁带库的平均槽位数为 109，比去年春天的 101 稍有提高，但是仍然是我们看到过的第二低的数字。但降低的槽位数也有积极的一面，因为这意味着更高的磁带库效率，特别是由于使用更新的驱动器。49%的受访者说他们的磁带库会有 LTO-4 驱动器来提供更高的容量和速度，还有加密。对加密的兴趣开始转变为实际行动，51%受访者说他们至少在加密部分的备份数据，相对去年秋天有 7 个百分点的提升。而且，加密再一次成为受访者今年打算使用或评估的新技术中的第二名。（参考 2009 年被实施和评估最多的 10 项技术）

当然，磁盘是近期备份运营中的必然之选，而存储经理会持续地在能够提高效率的磁盘到磁盘（D2D）技术上投资。D2D 的总体开销相比上个春天显示出有些反弹；只有 10%说他们会降低 D2D 开销（相对于去年春天的 13%），而 44%计划增加开销（去年春天是 34%）。这些不论对备份厂家还是对疲于奔命的备份管理员来说都是好消息。

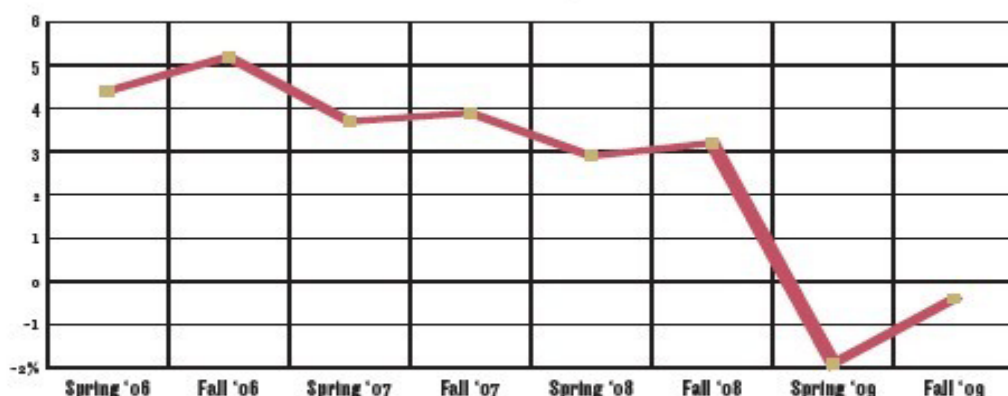
多数的公司（43%）都在他们的备份系统中集成了磁盘作为基于文件系统的目标设备或者是缓存；虚拟带库（VTLs）是排名第二的方式（26%）。在我们最近的调查中，35%的受访者说他们会在他们的备份系统中增加文件系统目标设备，而 24%倾向于 VTL。这个数字表明这一技术有了微弱的回升。

而重复数据删除仍然是大事件。21%的受访者在使用去重复——我们看到过的最高数字——而 26%在本年度的计划中增加了去重复。那个数字也是我们调查中的新高。只要大家仍关心预算，去重复就总是优先的，38%打算增加他们在去重复上的花销（比去年春天增加 8%）而只有 6%打算降低花销。

连续数据保护（Continuous data protection，CDP），是另一个节约空间和时间的方案，有 12%受访者正在使用，之比我们去年春天看到的 CDP 最高值低一个百分点。16%计划今年实施 CDP，也比上次的高点低一个百分点。

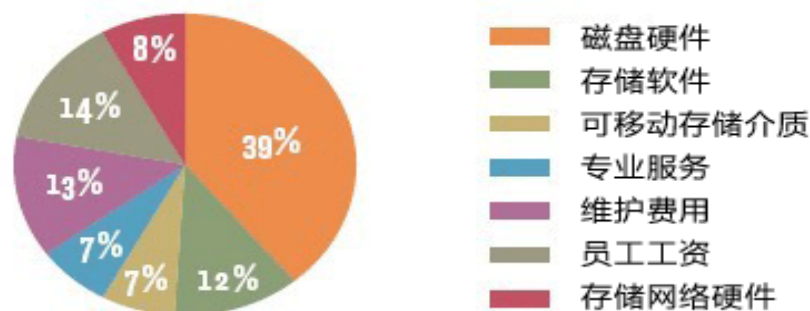
“云备份”的数量在过去的一年有明显的上升。然而，在我们的受访者中仍

同去年相比存储预算发生的变化 (%)



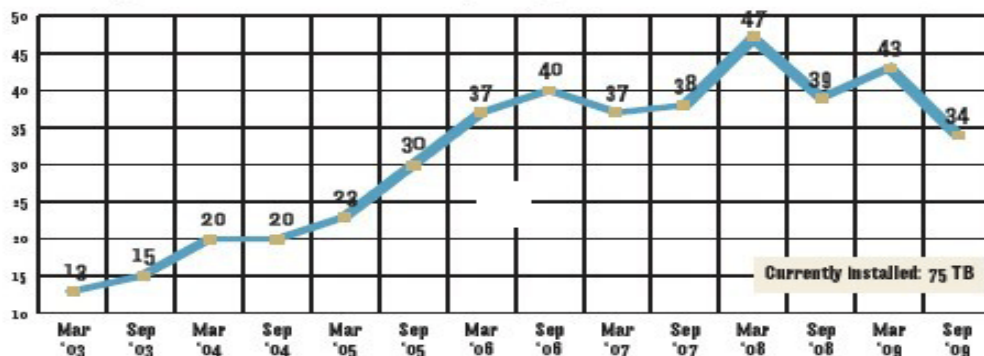
与2008年相比，今天春天的存储预算变化已经为负值（-1.9%），但在今天秋天的调研中已经发生反弹，只比2008年同期微降了0.4%。

存储预算如何分配



和去年同期调查相比，存储预算分配没有发生多少变化。由于持续增长的容量需求，磁盘硬件的预算占到了40%左右。

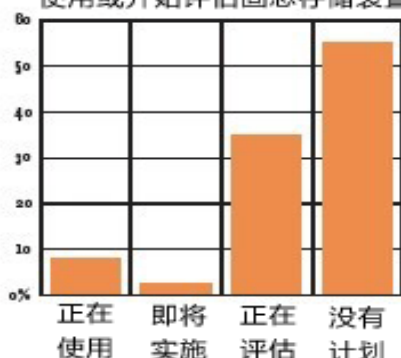
本年度平均硬盘容量增长量 (TB)



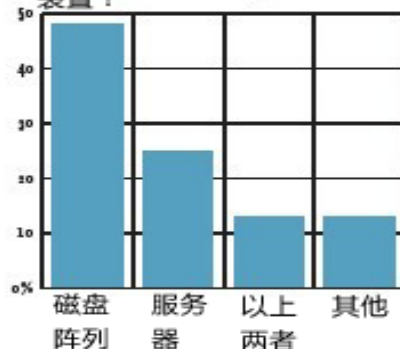
虽然大多数受访者称本年度增加的平均硬盘容量比去年同期有所下降，34TB仍然是一个不小的数字。大一些的企业平均硬盘容量增加值为68TB。

固态存储开始在数据中心崭露头角

当前您是否正在使用、计划使用或开始评估固态存储装置？

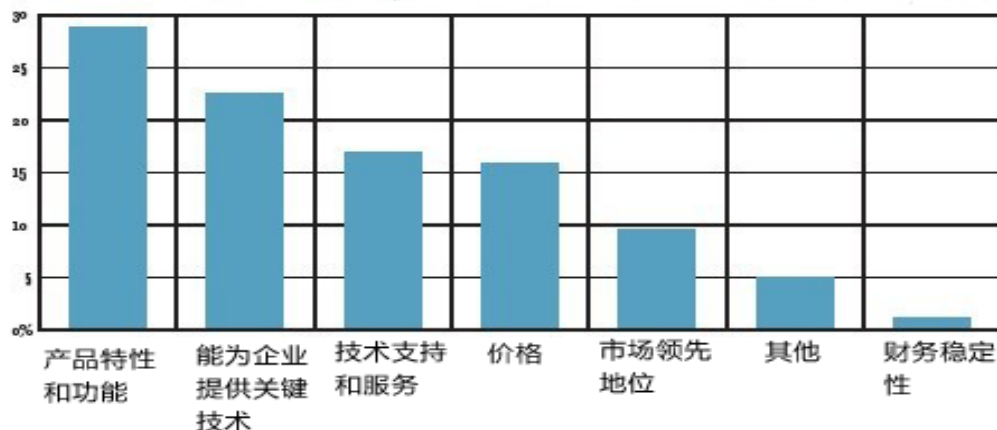


贵公司在哪里实施了固态存储装置？



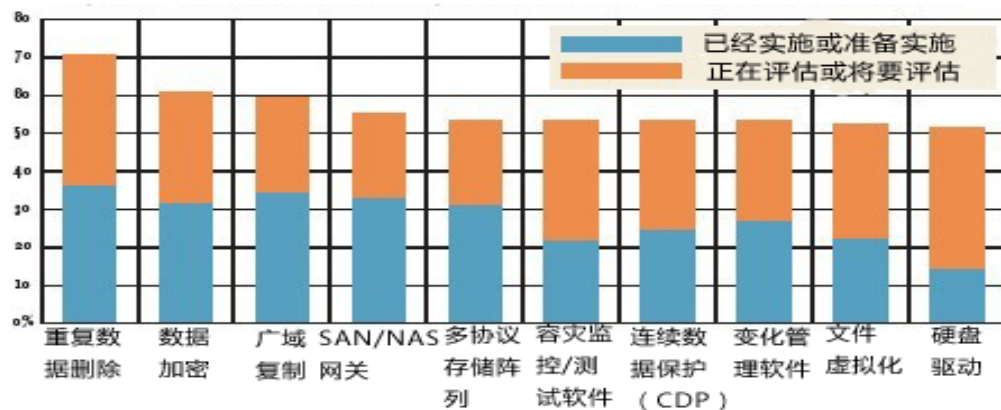
对一些企业而言，固态存储仍然是一项实验性的存储解决方案。已经有了一些实施案例，并且效果非常好。

在选择主要存储供应商时，贵公司的主要考虑因素是？



当选择硬盘系统时，产品的特性和功能是最重要的因素，但是，与提供了其他IT技术并且比较熟悉的厂商合作的变得更加重要，超过了价格和技术支持因素。

Top10：2009实施或评估的前十项技术



在新技术中，重复数据删除是最受关注的。加密，虽然没有被广泛应用，仍然被许多企业列入应用计划。

然是关注多于实施，经济形势影响了兴趣和采购决定。去年春天，我们看到打算通过外包或在线备份服务来削减备份开销的有大幅增加，21%计划使用某种备份服务。而在最新的调查中，这些数字有所回落，19%使用云备份。但是使用某种特定的备份服务——为桌面/笔记本电脑文件和数据库——的比例实际上有些许增长。

腰带仍然紧扎

对多数的存储经理来说，和从来都不太充足的存储预算斗争已经不是新鲜事了，而这种每年一度的挣扎也帮他们掌握了最近的经济风暴的风向。存储厂家也做出了他们的努力，旨在提高效率的产品和技术似乎刚好在合适的时间出现。有一件事是肯定的：面对容量需求和保护越来越多数据的需要——通过更少的资金来搞定所有事——存储经理们又一次证明了他们的神通广大。

(作者：Rich Castagna 译者：Steven Fu)

剖析 内部存储云

几乎所有的厂商都在大肆宣扬他们的某些云存储产品或服务。本文让我们冷静下来，先看看内部存储云的组成为何物。 ———Alan Radding

“如今已经没有什么所谓的私有存储云了，” Nirvanix 公司（一家公有云存储厂商）的咨询主管 Stephen Foskett 宣称。也许的确如此，但是这从未使美国总务管理局（GSA）在八月初停止发布询价单（RFQ），而此询价单最先针对的似乎就是私有存储云

GSA 所认为的私有或者说内部存储云，也许同多数企业所认为的有所不同。如询价单上所说的：“通过 GSA 的 Cloud Computing Storefront Site 站点，可以帮助最初很快的取得这些服务，而这则可以促使政府采购者通过普通的网页平台（使用信用卡或者其他可以接受的支付方式）来按需购买“架构即服务”

(IaaS) , 这就是所谓的 Cloud Computing Storefront , 它由 GSA 来管理和维护。

即使内部存储云也没有普遍接受的定义 , GSA 的 RFQ 似乎也给出了一个完全不同的解释。联邦政府不是寻求一个内部存储云或者一个公共存储云 , 而是需要给 Cloud Computing Storefront 做一个分类 , 可以使联邦机构通过一个平台或者网关 , 购买和实现让公共云存储服务的内部使用。甚至 Nirvanix 公司正在准备答复 RFQ 的 Foskett 似乎也陷入了困惑。

在这点上 , 政府看起来是做对了。无论怎么定义内部存储云 , 他们都保证可以降低存储成本和简化存储过程。根据 GSA , “云计算有能力通过利用市场上容易买到的技术来降低 IT 设备的成本 , 这些技术仅仅是基于服务器、数据库和应用程序的虚拟化技术 , 从而实现资金成本的节约。” GSA 首次将存储和计算云联系在了一起。

内部存储云的问题不在于它们不存在 , 而是它实在是可以存在太多的版本了。 “云是一个抽象的东西 , ” Stillwater 的创始人和高级分析师 Greg Schulz 说 , “几乎任何存储产品都能够配置成一个内部存储云的一部分。这取决于你的定义。所以 , 不论厂商买的是什么 , 他都可以相应地来定义存储云。”

存储云的类型

- **公共存储云** 像 Amazon 的 Simple Storage Service (S3) 和 Nirvanix 公司的 Storage Delivery Network 使得大量的文件能够实现低成本存储。多重租赁允许提供商独立保持每个客户的存储和应用程序。公共存储云的部分能够被分拆去创建对等数量的私有存储云。
- **私有存储云** 使用私有存储云 , 公司可以拥有或者控制基础设施以及决定如何配置应用程序。私有云也许在一个公司数据中心中或者在主机托管设备中进行配置。公司自己的 IT 组织或者一个服务提供商就可以构建或者管理私有云。
- **内部存储云** 这个存储云类型 , 除了是保持在组织的防火墙内这点不同外 , 类似于私有存储云。它可以在咨询师或者整合者的帮助下构建 , 但是它的管控和维护必须由 IT 部门进行。
- **混合存储云** 混合存储云综合了公共和私有或者内部云的属性。它主要用来实现按需索取 , 在临时基础上进行的外部容量精简。以公共云的容量来扩大私有或者内部云的能力 , 公司可以在面对快速的工作量变动或者计划工作量的取消时维持服务水平。然而 , 混合云带来了如何在公共和私有云之间分配应用程序的复杂性。

尽管对内部存储云尚未有广泛接受的定义，行业分析师们已经逐渐明确它的必要组成要素，同时开始可以解释它们是如何联结的。云并不神秘，“任何人都可以做，” Illuminata 公司的首席 IT 顾问 John Webster 说，内部云存储不是脑手术。

如今内部存储云并不多见，但对它们的需要已经明确了。“这是性能和成本之间的较量。内部存储云主要着眼于成本，” Ocarina Networks 公司的产品副总裁 Carter George 说。反之，一般的存储则主要看重的是性能和数据保护，其包括成熟的存储阵列、存储区域网络（SANs），高性能的磁盘驱动器及精密的备份和恢复。

根据 Bend 的独立存储咨询师 Abbott Schindler，低成本并没有必要成为主要的关注点。然而如今人们最看重的仍是成本，Schindler 说，因为“大部分开始使用云是基于归档存储或者数据保护的考虑，因此他们设计得比较便宜，速度较慢。然而，在云的概念里没有说它不能用来处理交易数据。”

内部存储云的定义

你可以说内部存储云和公共存储云一样——通过网络交付的存储服务——除了内部存储云的组件是在防火墙的后面之外。但这样的定义仍不完全正确。举例来说，公共存储云提供商能够预留一部分空间给某个客户专用，使之成为私有的存储云，尽管其对客户来说不是内部的（见“[不同类型的存储云](#)”）。

与其要给内部存储云一个具体的定义，行业分析师和咨询师更愿意去对它的属性进行解释。

基本内部存储云组件

这是为构建内部存储云的 DIY 的部分列表：

- 全局文件或者集群文件（NAS）系统（包括虚拟化和管理能力）
- 商用服务器和低成本存储（SAS DAS、JBOD、PCIe RAID）
- 节省成本的网络带宽服务
- 云个性界面取决于应用需要（NFS、SMTP/POP、HTTP、DICOM、REST、SOAP 和 XML）
- 对象或者元数据管理层，包括应用数据保护和安全工具

作者：Greg Schulz，StorageIO 集团

举例来说，焦点是明确集中在低成本和简单可扩展性上的。“对于存储云来说有个大的财务方面的问题” CommVault 的产品拓展副总裁 Anand Prahlaad 说，“客户不仅要求低成本，还希望只对自己所使用的部分付费。”简单的说，大家期望内部存储云交付的是便宜的存储。

廉价的存储，速度也慢。不过，Schindler 的咨询师并不认为作为内部存储云的一部分不存在更好的存储性能或不同的服务水平。

易管理是另一个显著的方面。“你使用内部存储云一定是想去除存储的复杂性，”希捷公司的首席技术官（CTO）David Allen 认为。最终私有存储云应该更易于管理，使管理员一个人就能处理成百的存储节点和上 PB 的存储。也许管理员就只用负责少数的简单工作了。

最后的关键是如何获得私有存储云。HTTP 是一个主要的存取协议。“你所要的就是 HTTP 或者 HTTPS 的接口和网络浏览器，” SentryBlue 的首席技术官 Ken Satkunam 认为。

“内部存储云的一个显著的区别在于它的获得是通过 API 而不是协议，” Nirvanix' s Foskett 说。“它将会如同网页一样拥有有可编译的 API，使用基于 HTTP 的 REST。”与光纤通道（FC）相反，状态传输（REST）是一个没有状态的协议，包括所有通信的状态。REST 帮助实现基于 HTTP 的网页服务；对于存储云，REST 则使得我们可以像使用服务一样来获得存储资源。

在最近的白皮书中，Sun 公司坚持存储云的可编程性。“取代部署物理服务器、存储和网络资源，支持应用程序，开发者明确了相同有效成配是如何配置和相联系的，包括虚拟机是如何映像的，应用数据是如何从存储云中存储和恢复的。他们详细说明了组件通过 API 进行配置的方式及时间。

然而行业还没有将云的 API 标准化，StorageIO 集团的 Schulz 说，每个云提供商提供自己的 API。在 7 月末，Rackspace Hosting 公开了 Creative

内部存储云的一个显著的区别在于它的获得是通过 API 而不是协议。它将会如同网页一样拥有有可编译的 API，使用基于 HTTP 的 REST。

——Greg Schulz , StorageIO 集团

Commons 3.0 版本许可下的 Cloud Servers 和 Cloud Files 的 API 技术说明。最终这也许会为希望构建内部存储云的人提供一个开放的 API，使他们可以开始行动。

最后一个特征——多重租赁——定义了公共存储云。“对存储云甚至是内部存储云，多重租赁都是一个重要的部分，” CommVault 的 Prahlad 说。通过内部或者私有的云存储，多重租赁可以让一个组织按照需要分成部门、项目和工作团队。

那么什么是内部存储云？一致的定义是：存储云是公司拥有的私有存储空间（或至少可以控制的），它可以通过基于 HTTP 连接的编程方式来实现，通过易于管理的多重租赁，它可以交付低成本、高可扩展性的存储。ParaScale 公司补充道，一个内部存储云可以很小（只有 3-5 个节点），但仍然可以提供云所能带来的易于管理和可扩展性的经济。

内部存储云的选项

如果对内部存储云有点眼熟，那就对了。“存储网格已经变体成为私有存储云，” 咨询师 Schindler 说。在存储网格之前，公共计算将计算和存储资源打包作为一个计费服务。尽管技术和构造不同，但这两个概念是相类似的。“它们都是关于存储的变革——无论何时何地，以所希望的成本存取你所想要的数据，” Schindler 补充说，而不用关心真正的存储设备是什么、它在网络中的什么地方。

内部存储云同样类似于 NAS 集群，但需要注意一些区别。“我不确定集群的 NAS 可以扩展成文真正存储云的规模，” CommVault 的 Prahlad 说。尽管一个内部存储云开始的时候可以很小，但是公司希望通过增加设备将它进行横向扩展。

**我不确定集群的 NAS
可以扩展成文真正存储云的
规模。**

——Prahlad, CommVault 产
品拓展副总

当讨论到内部存储云产品时，现有的选择可以相当少，也可以特别多，主要是取决于你如何定义内部存储云。对于真正的产品，EMC 公司的赞助商 Atmos 形容它为信息存储和分配的提供者。EMC 根据自身所希望的服务水平，用 Atmos 通过全球网络存储和复制公司的数据。它在大量的存储云中使用商务政策，政策导向的自动化和

元数据来管理公司的数据，并且保证操作上的效率，降低了管理的复杂度，节约了成本。

AT&T 作为一个私有的存储云是 EMC 为 Atmos 而准备的客户展示柜。但是事实上 AT&T 没有作为私有云来使用。它通过 Atmos 为自己的客户提供包括存储的服务——更像公共云分销商。

与一般的假设相反，在 Atmos 背后不存在巨大的 EMC 存储阵列。“那样做的费用太高了，” Nirvanix 的 Foskett 说。Atmos 的扩展能力是通过 JBOD 来交付的。用 Atmos，你可以通过 API 和 NAS 界面获得数据中心中所有的数据。或者你可以使用 Atmos 公共存储云的数据块来作为私有云。

ParaScale 公司专为创建和管理内部存储云提供软件。不像云服务提供商，他只卖工具，让商户构建自己的存储云。软件运行于标准的 Linux x86 服务器，将直连在多个服务器硬盘与 PB 级的文件存储聚集在一个单一的命名空间中。

除了 Atmos 和 ParaScale，商业化的内部存储云产品是非常少的。“除此之外，任何人所称的私有云都不是真正的云，” Foskett 说。他们可能提供在某程度上可以虚拟化的存储产品，就将其视为云。“通常他们提供的是一般的产品，但也将其挂名为云，”他接着补充道。类似的，几乎任何 NAS 集群都与内部存储云很相像。

构建内部存储云

“DIY 对内部云很重要，” 咨询师 Schindler 说。DIY 非常流行，正如 Illuminata 的 Webster 所说的，这是因为配置私有存储云并不难（见“基本的内部存储云组件”）。

设计和构建一个私有存储云有很多方法。最简单的方式也许是“以 NAS 集群开始，最好还要结合全局文件系统，配上云（网页）前端，” i365 的 Allen 说。在内部私有云背后的真正的存储各不相同。你可能不会用一个存储阵列来作为私有云的一部分。“大多会使用商用服务器，用低成本的磁盘插满插槽，” Nirvanix 的 Foskett 说。

**DIY 非常重要，尤其是对
于内部存储云来说。**

——Abbott Schindler，独立
存储顾问

一种情况是：“使用插满 Linux 服务器刀片和磁盘的机架或者刀片笼子，” SentryBlue 的 Satkunam 补充说，“可以连接本地磁盘，使得其价格没有 SAN 存储那么贵。”

构建一个可扩展的内部存储云的关键“是以很多小型机箱开始，通过逐渐增加更多的机箱以扩大规模，” CommVault 的 Prahlad 说。通过在不同的节点间复制数据构建冗余，你可以使数据得到保护。在获得服务质量方面，不同的节点有不同的服务性能属性。

将这些都联合起来的粘合剂是“代表单一命名空间的全局文件系统，” Prahlad 说。这也可能需要虚拟化和元数据层。

内部存储云的管理应当要简单。“为了你的模型，你不得不看像亚马逊和 Facebook 这样的网站。你希望整个文件存储是通过 HTTP 的，” Ocarina Networks 的 George 解释道。为了简单化，尽量限制你的文件管理选项去创建、读取、更新、删除、移动/复制。

内部存储云不能替代组织的 Tier1 存储。生产数据仍然继续在运行高性能的 FC SAN 或者初级的 iSCSI SAN 以得到备份和保护。相反，内部云可以用来存储那些耗尽磁盘空间、令备份和恢复策略复杂化的基于文件的数据，如邮件、档案、媒体和合规性数据等。那些数据仍然是活动的、被广泛使用和改变的；它们需要储存和共享，但是不需要 Tier1 生产数据存储那样的费用、性能和服务的水平。

来自基于纽约的最新的 Wave 研究（2009 年 1 月到 2009 年 5 月）TheInfoPro 大体上询问了受访者对云的兴趣。“他们的兴趣水平并不高，大概在 12% 到 15% 之间，”据 TheInfoPro 的存储研究的管理负责人 Robert Stevenson 报告。“大多对于云没有购买计划。”目前大公司很明显对于内部存储云或者云计算并不怎么感冒。

他们也许已经仿造了内部存储云，但尚未意识到他们将带有附属磁盘的虚拟化服务器带到了网络。那是向真正的内部存储云迈了一小步。

（作者：Alan Radding 译者：Tony Tang）

NFS 4.1 版的 pNFS : 巨大的 NAS 性能提升

去年年底，当网络文件系统（NFS）4.1 版在互联网工程工作小组（IETF）的状态由“最后通知”（Last Call）变成“征求意见”（Request for Comment）的时候，出现了大量的关于并行 NFS（pNFS）的出版物，pNFS 包括在 NFS 协议中，这些出版物描述了 pNFS 如何在网络附加存储（NAS）的性能方面产生了巨大的突破，它可以用于像高性能计算（HPC）那种带宽密集型的应用。NFS 4.1 版代表了途中的重要一步，使 NFS 能够更好地服务于吞吐量的需要，而这一直被持续增长的文件大小和苛刻的 HPC 环境所驱动。随着有关 pNFS 的各种声音逐渐平息，厂商正在开发融合了标准和支持 pNFS 产品的解决方案，并且会在大约 6 个月内推向市场。600 多页的征求意见（RFC）正在编撰过程中，并且很快就会得到全面的正式批准。

网络文件系统协议

NFS 协议使用户能够远程地访问和共享那些存储在中心文件服务器或者 NAS 阵列的目录和文件，就好像这些数据存储在本地磁盘一样。通过部署专用的 NAS 设备，您将获得存储资源的集中管理，更方便的文件共享和协作，更好的数据保护和灾难恢复（DR）计划，存储优化，通过配额节省空间等等好处。ESG 预测，到 2012 年，基于文件的数据将会达到总存储容量的 70% 以上，那时这些好处对于存储管理员来说将变得更加重要。

NFS 的主要挑战之一，就是性能受限于 NAS 头或处理器节点的带宽，而它们正控制着，或者说“拥有”被访问的目录和文件。NFS 4.0 把文件的归属限定到一个节点（有很多方法绕过单节点归属的限制，并不是没有折衷的办法）。当一个文件被客户端访问的时候，所有发送到客户端的数据必须通过 NAS 头进行路由。在此期间，NAS 头也同时处理像锁定、权限和文件元数据管理的 NFS 的工作。访问一个大文件会大大影响 NAS 头的性能，导致其他通过这个头等待文件的用户无法进行文件共享访问。在高性能计算环境中，这个问题会加剧，它已经经历了到并行处理的转变过程，多个处理器访问共享数据的方式可轻易战胜单个的 NAS 节点。

NFS 4.1 版代表了途中的重要一步，使 NFS 能够更好地服务于吞吐量的需要，而这一直被持续增长的文件大小和苛刻的 HPC 环境所驱动。

为了满足这一需求，很多厂商已经采用了并行文件服务技术，但是由于它们的专利性质，这些产品的采用是受限制的，并且需要与专门的客户端相结合。广泛的采用并行文件服务，如果这确实会发生，需要一个标准的途径。这时 pNFS 出现了。

pNFS

并行 NFS 把 NAS 的性能提升了一个等级。文件可以跨 NAS 头进行分割和条带化，并且利用多个数据路径和处理器，并行地向请求者传送数据，从而大幅地提升性能。pNFS 还提供了完全可以用于文件传送的 NAS 头旁路能力。

提供并行数据传送的关键之一就是增加一台带元数据服务器。元数据服务器包含一个地图，在 NFS 4.1 里叫“Layout (数据表)”，详细说明如何以及在何处存储数据。元数据服务器也处理文件的语义和权限。当客户端发出一个文件请求的时候，这个请求首先被发给元数据服务器。元数据服务器把文件“居住”在哪个相关的文件服务器的信息反馈给客户端，然后客户端会直接得到这个信息。如果文件跨多个处理器节点进行了条带化，那所有的处理器节点都可以来响应这个请求，以提升带宽和处理能力。

与在采用了支持直接块数据访问和面向对象的存储设备 (OSDs) 的 IP 网络上，进行并行数据传输的方式对比，pNFS 迈出了更深远的一步，在传送文件数据的时候，它完全从根本旁路了 NAS 头。授权客户端在块数据模式下发送文件访问请求的时候，文件以真正块的形式返回给发出请求的客户端，而不是以一个文件的形式。客户端可以直接到存储设备本身，而不通过 NAS 头，利用 SCSI 协议来获取数据。在典型的高性能计算环境中，客户端经常指的是数据中心的服务器，这意味着它们可以通过高速管道直接连接到块存储设备，像 10GbE 或 InfiniBand，同时通过多个并行路径访问文件（以块数据方式）——对比通过 NFS 和单节点 NAS 进行共享文件访问的方式，能够产生巨大的性能提升。对于对象的请求将会遵循类似的途径。

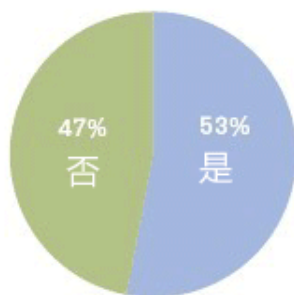
随着 pNFS 并行文件传送的强制标准化，用户将会在 NFS 4.1 中看到 NAS 性能的提升，无论他们部署的是谁的 NAS 存储。

(作者: Terri McClure 译者: 曹一玺)

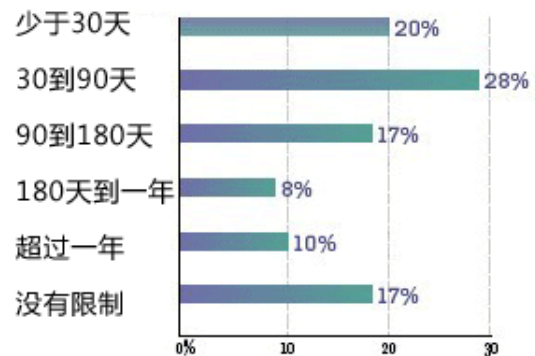
邮件归档需求日益增长

Email，曾经只是交流工具中的一种，现在成了许多企业的一个关键性应用。因此，Email 需要更多的关注——尤其是在法规遵从方面——邮件需要归档。在本月进行的调查中，53%的受访者称他们正在使用邮件归档，而去年同期这个比例只有 45%。日益增长的邮件数量使得许多企业开始进行邮件的归档，30%的受访者称使用邮件归档的主要原因是为了对邮件进行更好的管理。法律准备和法规遵从比例各占 26%和 28%。将近一半（47%）的受访者选择了第三方归档应用，但仍有 37%的受访者只依赖邮件系统本身的功能。越来越多的人开始考虑外部服务，8%的受访者考虑使用归档路径，这个比例是去年的两倍。对于打算应用邮件归档的这部分人，52%的受访者称已经列入计划。随着邮箱容量限制的扩大，65%受访者选择了 500 MB 甚至更大的邮箱，而去年这一比例为 49%——将邮件归档作为首要选择是个不错的想法。

贵公司的邮件是否进行了归档？



在进行归档前，邮件保留在使用者邮箱中的时间长度是？



您面临的三个主要的邮件归档方面的挑战是？



63

的受访者会对所有的邮件进行归档，而不是有选择的进行。

做这件事还有很多隐性花费，这些很难对非 IT 人员解释。

——调查反馈

（作者：Rich Castagna 译者：吴迪）

SPC-1 测试对存储设备采购的指导意义

一、存储性能指标

要对产品进行比较就必须有相应的标准，首先笔者在这里介绍一下三个基本的存储性能指标：

IOPS (I/O (Input/Output) operations per second)：每秒的输入输出次数。

对于存储系统来说，一个 I/O 是相对于磁盘的一次输入（写入数据到磁盘）或者输出（从磁盘读出数据）。存储端的 I/O 和主机（应用服务器）端的 I/O 是不一致的。例如，主机端发出一次的最简单的向磁盘写入数据的一个命令，就可能涉及到 N 次存储端的 I/O。

平均响应时间 (Average Response Time)：对于现代存储系统来说，除了 IOPS—每秒输入输出这个参数，另外一个重要的参数是在不同的存储 I/O 负荷下，完成每个 I/O 所需的响应时间。

数据传输率 (Data transfer rate)：即通常所说的带宽。在数据传输 I/O 负载期间，单位时间内 I/O 总线上传输的数据总量。I/O 子系统的数据传输能力是任何 I/O 负载的数据传输率上限。对于磁盘子系统 I/O 而言，数据传输率通常使用 MBytes/second 每秒兆字节来度量。

二、如何利用存储性能参数来选择适合的存储系统

在实际项目中，我们已经看到越来越多的用户采用一些特定的存储性能指标来选择存储产品。例如以下存储招标参数(均来源于网络公开发布资料)：

1. IOPS>125000（厂商必须提供公开的第三方的性能测试指标，该指标可以在公开网站查询（如：SPC-1，SPC-2），或者提供厂商网站上公布的 IOPS 值的链接地址）。

2.

磁盘阵列性能	★	磁盘阵列的 IOPS≥140,000
	★	磁盘阵列的最大数据传输率≥

		1550MB/s
		具备第三方发布的 SPC-1 或者 SPC-2 测试成绩

3.某省级政府采购招标文件：

技术参数指标及要求	
☆光纤通道（前端）	双活动热插拔 4Gb*4 FC
☆高速缓存	2GB
☆高速缓存电池	断电后至少 72 小时
☆硬盘接口（后端）	4Gb*4 FC
☆IOPS Cache	120000
☆MB/s Cache	1550
☆MB/s Disk	900
☆最大硬盘数量	112
☆驱动器支持	73, 146, 300,450 GB 15K FC 磁盘； 500, 750 ,1000 GB 7200 转 SATA 磁盘
驱动器混插	支持
☆当前容量及支持最大容量	当前容量 3TBFC 最大 FC 50TB/SATA 112TB
支持最大 LUNs	1024
SPC-1	17195.84
SPC-2	804.29

我们看到，在以上的存储性能指标参数中，均出现了“SPC-1”和“SPC-2”，另外还有 IOPS、数据传输率等指标。本文的以下部分将重点介绍 SPC 的存储性能测试，尤其是基于随机 I/O 应用的 SPC-1 存储性能指标和测试分析。希望能使读者对存储系统的性能指标和 SPC-1 测试有更深入的了解。

三、SPC 简介

SPC (Storage Performance Council , 存储性能理事会) 成立于 1998 年 , 它的建立就是为了定义、标准化以及促进存储子系统的基准 , 同时向计算机厂商和用户发布客观、权威的性能数据。

根据存储行业的需要和行业所关心的问题 , SPC 在 2001 年创建了第一个行业标准性能基准测试 SPC Benchmark-1(SPC-1) , 它是第一个基于工业标准的存储性能测试体系 , 它能够在直连或网络存储中科学、准确地测量出被测产品的性能 , 并进而判断其性价比。

- 具体来说 , SPC-1 具有以下显著的优点 :
- 为不同的系统平台提供一个跨厂商的统一的测试环境。
- 受到事实上的整个存储业的推崇 , 因而遵循严格的标准。
- 给厂商、用户、分析家和出版社一个有力的和能简单使用的结果。
- 为存储系统的整个生命周期提供价值 , 包括产品需求的发展、产品运行、性能调节、市场定位和购买评估。
- 运行方便 , 检验方便 , 官方结果报告简单。

目前 , SPC 的 SPC-1 基准测试主要是针对随机 I/O 应用环境的 , SPC-2 基准测试主要是针对顺序 I/O 应用环境的。其中 SPC-1 基准测试按下表的读写比例 , 以不同的数据块尺寸产生 I/O 数据流对存储系统进行测试 , 使 SPC-1 基准测试很好地模拟了 OLTP、数据库和 e-mail 等真实应用环境 , 使 SPC-1 基准测试结果具有很高权威性和可比性。

SPC 公布的 SPC-1 基准测试报告详细地描述了被测存储产品的配置、价格清单及测试的系统结构图和被测存储产品的以下各项指标在 SPC-1 基准测试中的测试结果:

- SPC-1 IOPS™ (Maximum Throughput:每秒输入输出次数的最大值)

- SPC-1 LRT™(Average Response Time at the low level of demand:最小响应时间)
- SPC-1 Price-Performance(\$/ SPC-1 IOPS™:每 SPC-1 IOPS™的价格)
- Storage Total Price(被测存储系统的总价)
- Storage Total Capacity (被测存储系统的总容量)
- Data Protection Level:RAID1 or RAID5 or Other Protection Level (数据保护级别：如 RAID1、RAID5 或其他 RAID 级别)

总结：

SPC-1 测试的是存储系统的 IOPS 和平均响应时间以及 SPC-LRT (最小响应时间)。

SPC-2 测试的是带宽，单位是 MBPS(每秒的数据传输量，单位 MB)

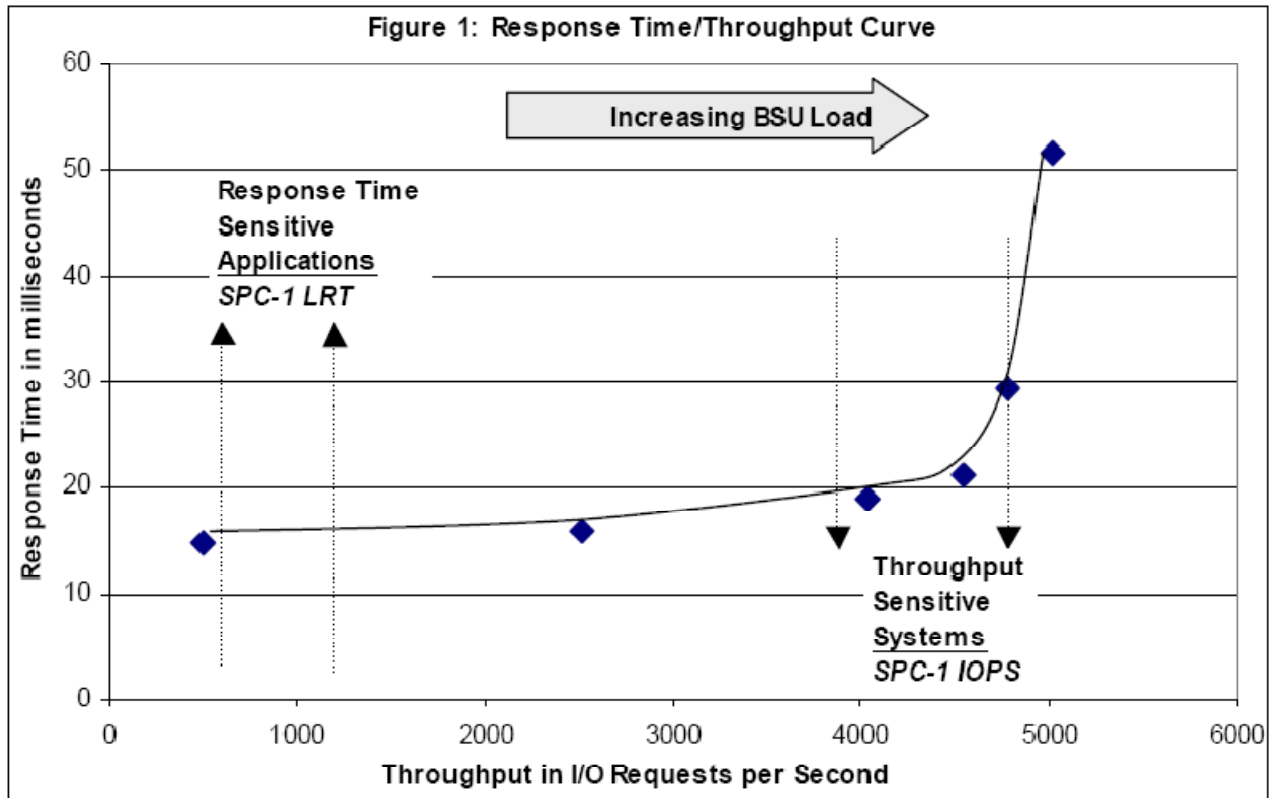
3.1 SPC-1 性能测试参测厂商

3PAR、DataCore、Dell、Fujitsu、HP、HDS、IBM、LSI、NetApp、Pillar、Sun、TMS 和 Xiotech

3.2 SPC-2 性能测试参测厂商

Fujitsu、HP、HDS、IBM、Sun 和 Xiotech

3.3 SPC-1 性能测试示例



上图是 **SPC BENCHMARK-1 (SPC-1) OFFICIAL SPECIFICATION** Revision 1.10.1 (SPC-1基准测试官方规范，版本1.10.1) 给出的IOPS和响应时间曲线图示例。这也是目前最新版的SPC-1规范文档。

我们先来看SPC-1规范中的以下定义：

响应时间 (Response Time)：一个被测的I/O请求的响应时间是它完成的时间减去开始的时间所得的时间段。实际就是一个I/O请求完成的时间。

平均响应时间 (Average Response Time)：被测的所有 I/O 请求时间之和除以被测的 I/O 请求数量，平均值。

每秒中的I/O请求数量 (I/O Request Throughput) , IOPS。

上图中纵坐标是响应时间，单位：毫秒。横坐标是每秒 I/O 请求数量，即 IOPS。

图中的 6 个蓝色菱形图标，表示的是测试该存储系统在 I/O 负荷分别为 10%、50%、80%、90%、95%、100%的情况下，所对应的平均响应时间。这也是 SPC-1 对送测存储系统的测试规定。

SPC-1 LRT：轻量级负载时的响应时间（Lightly Load Response Time），也称为称为最小响应时间。在SPC-1测试中，它是10% load level（即最大IOPS的10%）时的响应时间，即上图中左起第一个蓝色菱形方块对应的响应时间。这个参数，可以作为衡量对响应时间敏感的一些关键应用程序的性能参数，典型的例子是重建一个大型的数据库。这类应用的特征是对存储系统总体的total I/O Request throughput数量要求不高，但是事务的关键应用能否成功依赖于成千上万的并行I/O请求能否快速完成，因此要求尽可能小的响应时间。

另外，如图所示，在大约80%到95% I/O负载的这个区间的响应时间变化曲线，可以用来衡量在大量的多任务、高并行的应用情况下，存储系统的单个I/O的响应时间。典型的例子是OLTP(在线联机事务处理)系统，例如奥运门票系统。这类系统是拥有很多应用或很多并行执行的线程而导致I/O请求处理达到饱和的系统，并且要求是在大量并发事务的请求下，存储系统应该能提供可以接受的响应时间。如果在此区间，评测曲线出现大的拐点，表明该存储系统的响应时间突然增长，有可能会造成无法响应。例如奥运门票系统，当购票的用户数陡增到一个临界点时，意味着主机端(应用层)对存储系统的I/O请求数也会陡增，底层的存储系统如果无法及时响应，或者造成响应时间变得很长，将会导致整个系统的崩溃。

需要注意的是，上图给出的是在60毫秒内的测试曲线，实际SPC给出的SPC-1报告中，给出的是响应时间在30毫秒内的测试结果。

3.4 SPC-1 测试报告下载地址

SPC-1测试报告提供公开下载。SPC-1测试报告可以在SPC官方网站下载：
http://www.storageperformance.org/results/benchmark_results_spc1

每款送测的存储系统的 SPC-1 报告都分为两个版本：Executive Summary 和 Full Disclosure Report，分别是精简版和完整版的报告。完整版报告包括了详细的测试环境、定义、流程、数据表等信息。对于一般用户参考来说，查看 Executive Summary 版本的报告就足够了。

Test Sponsor	SPC-1 Submission Identifier - Tested Storage Product (TSP)	SPC-1 Version	Peer Review Status Active/Inactive/Withdrawn Status TSP Category	Documentation, Submission Date, and Revision Status	
3PAR Inc.	A00079 3PAR InServ® F400 Storage Server	V1.10.1	Accepted: 26 June 2009 Active SPC-1 Result TSP includes all storage	Executive Summary Full Disclosure Report Submitted: 27 April 2009	No Revisions

四、高端存储系统 SPC-1 Benchmark

现在我们看看目前业界的高端存储产品中，SPC-1 评测结果比较高的一款磁盘存储系统：HDS USP V。

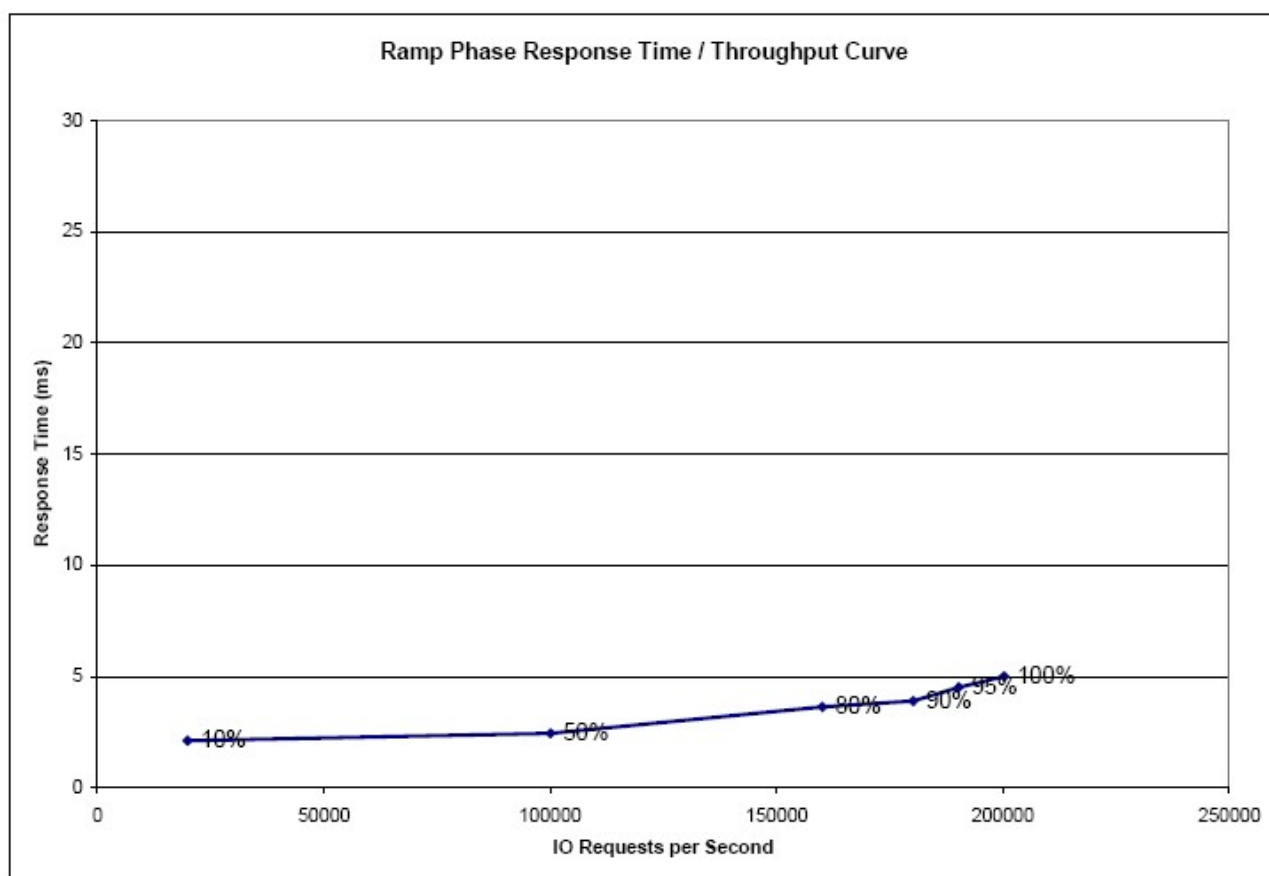
测试结果摘要

Summary of Results

SPC-1 Results	
Tested Storage Configuration (TSC) Name: Hitachi Universal Storage Platform™ V	
Metric	Reported Result
SPC-1 IOPS™	200,245.73
SPC-1 Price-Performance	\$17.61/SPC-1 IOPS™
Total ASU Capacity	26,000,000 GB
Data Protection Level	Mirroring
Total TSC Price (including three-year maintenance)	\$3,525,389

可以看到，最大 IOPS 为 200,245.73

测试曲线



测试结果数据表

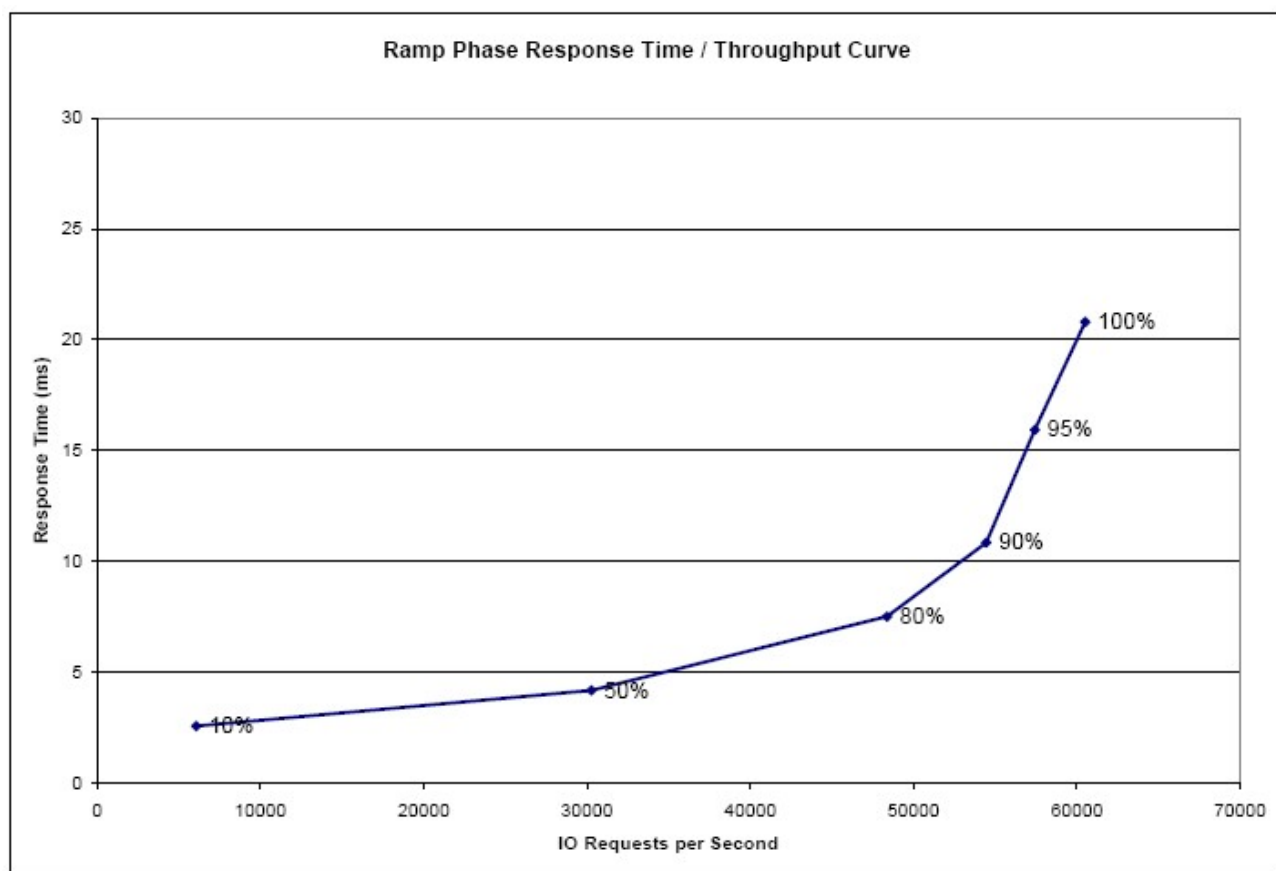
Response Time - Throughput Data

	10% Load	50% Load	80% Load	90% Load	95% Load	100% Load
I/O Request Throughput	19,994.33	100,105.69	160,170.55	180,202.31	190,196.63	200,245.73
Average Response Time (ms):						
All ASUs	2.10	2.43	3.62	3.89	4.50	4.99
ASU-1	2.92	3.20	4.68	4.92	5.54	6.10
ASU-2	2.11	2.63	3.75	4.23	4.82	5.37
ASU-3	0.34	0.70	1.31	1.58	2.15	2.47
Reads	4.82	5.10	6.99	7.46	8.12	8.89
Writes	0.33	0.69	1.43	1.57	2.14	2.45

可以看到,HDS USP V 存储系统,测得的 SPC-1 LRT 最小响应时间为 2.10 毫秒。随着 I/O 负载逐渐增大,平均响应时间增加平缓,没有出现大的性能拐点。且在 100% I/O 负载情况下,系统的平均响应时间在稳定维持在 5 毫秒以下。性能表现优异。

五、中端存储系统 SPC-1 Benchmark

NetApp FAS3170A

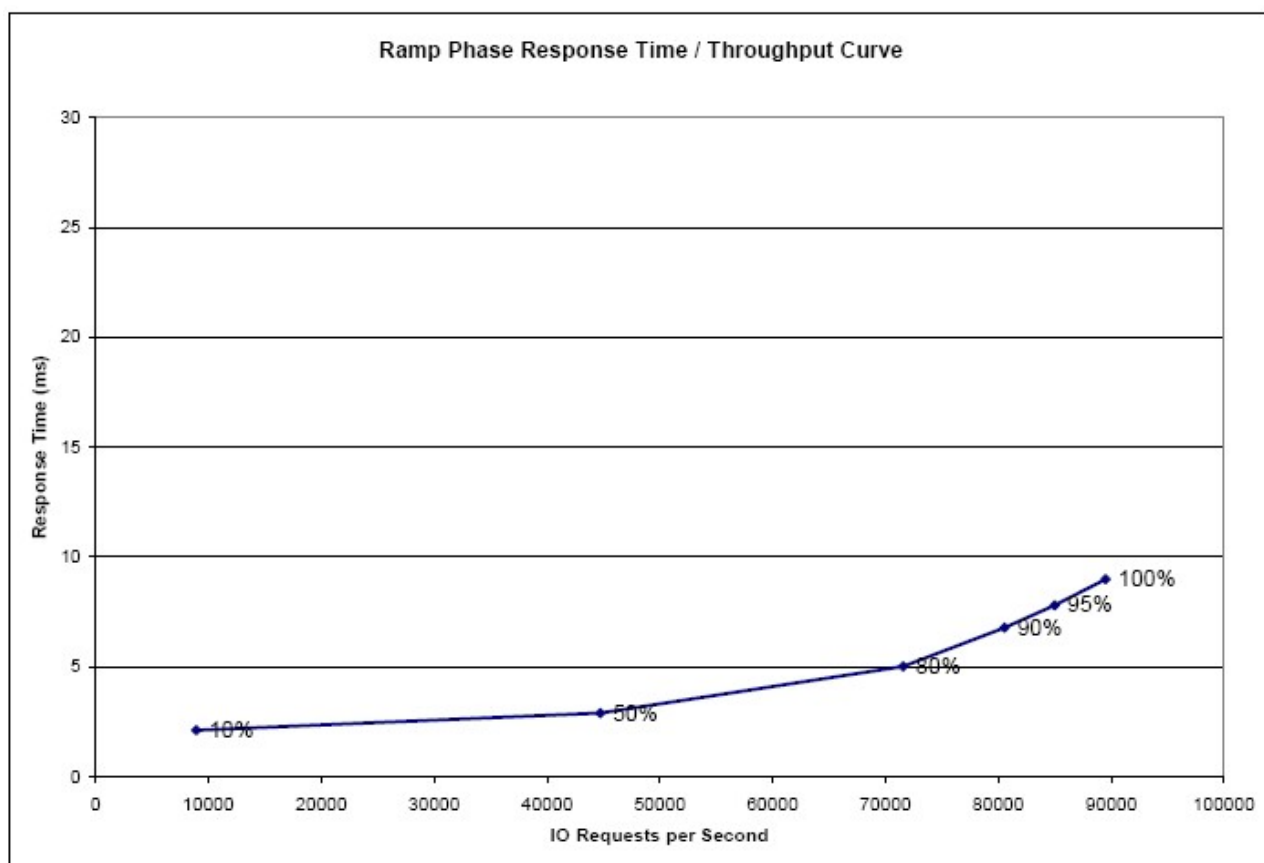


NetApp FAS3170A，测试配置为 224 块 146GB 15K RPM FC 磁盘。系统在 80%到 100%的 I/O 负载变化区间内，平均响应时间从 10 毫秒以内增加到了 20 毫秒以上，性能出现了明显的变化。

HDS AMS 2500

Summary of Results

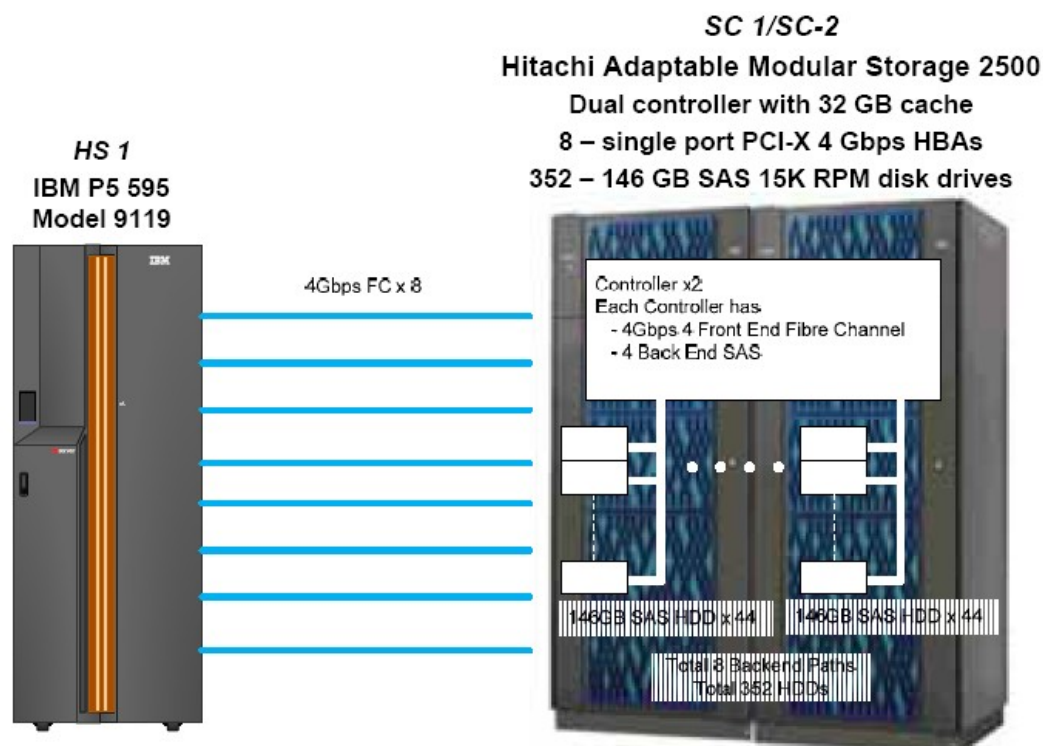
SPC-1 Results	
Tested Storage Configuration (TSC) Name: Hitachi Adaptable Modular Storage 2500	
Metric	Reported Result
SPC-1 IOPS™	89,491.81
SPC-1 Price-Performance	\$6.71/SPC-1 IOPS™
Total ASU Capacity	15,900.000 GB
Data Protection Level	Mirroring
Total TSC Price (including three-year maintenance)	\$600,581



HDS AMS2500 SPC-1 IOPS 为 89,491.81。

值得注意的是，HDS AMS2500 是 HDS AMS 2000 系列中端产品中配置最高的一款产品。AMS 2000 系列完全摒弃了目前其他主流磁盘存储厂商通常采用的 FC 磁盘和 4Gb FC 仲裁环式后端磁盘连接，采用了 12Gb/s (4xSAS Wide Link) 的 SAS 磁盘后端连接和 SAS 磁盘，并向下兼容低成本的 SATA 磁盘。截止到 2009 年 3 月，是 SPC-1 测试的中端磁盘存储产品中 IOPS 性能最高的产品。究其原因，第一是高端 SAS 磁盘的性能，已经和 FC 磁盘不相上下，第二是 SAS 后端磁盘连接，在数据传输率和工作方式、设备连接数上都优于目前的 4Gb FC 后端磁盘连接方式。

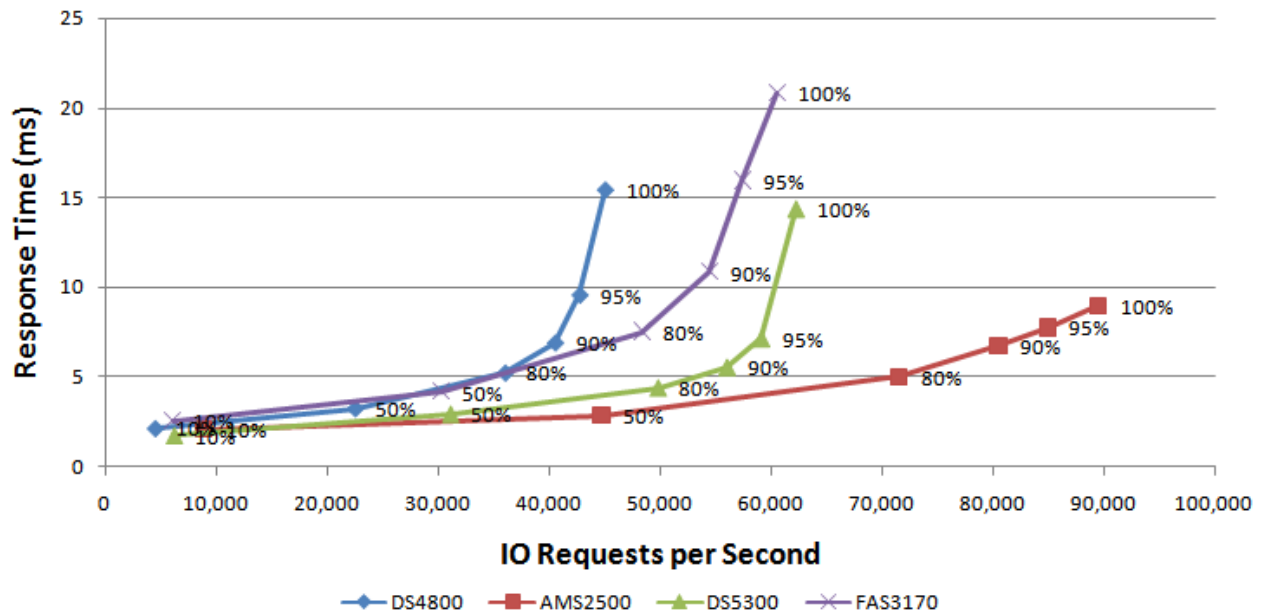
HDS AMS2500 的测试配置为 352 块 146GB 15K RPM SAS 磁盘。见下图。



存储系统的性能纪录类似于体育竞技赛场上的纪录，总是不断被刷新。HDS AMS2500 的中端存储 SPC-1 最高测试纪录，在 2009 年 4 月又被 3PAR 公司的 InServ F400 存储服务器打破。3PAR InServ F400 的表现是 93,050.06 SPC-1 IOPS，平均响应时间为 8.85 毫秒。

3PAR InServ F400 的测试配置为 384 块 146.8GB 15K RPM FC 磁盘，4 控制器（一般测试的中端存储产品均为 2 控制器）。

下图为几款目前业界主流磁盘存储产品 SPC-1 测试结果对比。



六、固态存储系统 SPC-1 Benchmark

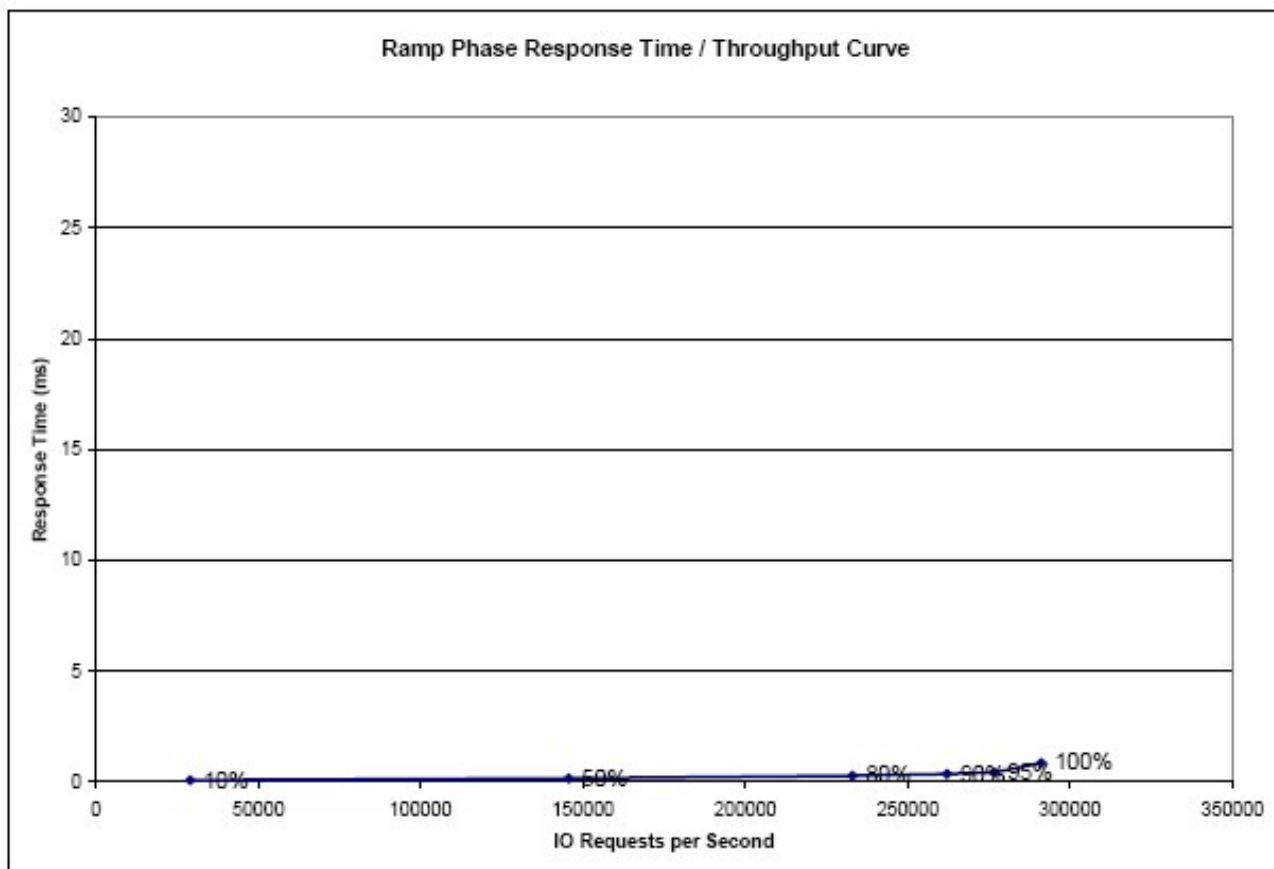
最后我们看看目前整个 SPC-1 Benchmark 的最高测试结果：Texas Memory Systems 公司的 RamSan-400 固态硬盘存储系统。SPC-1 IOPS 为 291,208.58。最小响应时间为 0.09 毫秒。平均响应时间为 0.86 毫秒。测试曲线平滑。比传统磁盘存储系统的 I/O 响应时间提升了一个数量级。

Summary of Results

SPC-1 Results	
Tested Storage Configuration (TSC) Name: Texas Memory Systems RamSan-400	
Metric	Reported Result
SPC-1 IOPS™	291,208.58
SPC-1 Price-Performance	\$0.67/SPC-1 IOPS™
Total ASU Capacity	137.439 GB
Data Protection Level	Other Data Protection
Total TSC Price (including three-year maintenance)	\$194,785

Response Time - Throughput Data

	10% Load	50% Load	80% Load	90% Load	95% Load	100% Load
I/O Request Throughput	29,108.36	145,605.79	232,991.56	262,089.32	276,655.51	291,208.58
Average Response Time (ms):						
All ASUs	0.09	0.16	0.29	0.37	0.43	0.86
ASU-1	0.08	0.15	0.28	0.36	0.41	0.90
ASU-2	0.09	0.16	0.27	0.35	0.42	0.74
ASU-3	0.11	0.20	0.33	0.41	0.49	0.83
Reads	0.08	0.15	0.28	0.35	0.40	0.87
Writes	0.10	0.18	0.31	0.39	0.45	0.86



七、结论

对于磁盘存储系统来说，在大负载 I/O 请求的压力下，存储系统的磁盘读写性能，成为系统最主要的性能瓶颈。某些高端磁盘存储，例如 HDS USP V，在最大 I/O 负载时平均 5 毫秒的响应时间，基本上已经达到了物理磁盘机械运动的上限（目前的企业级高端 FC 和 SAS 硬盘，平均寻道时间加上寻道延时再加上平均读写时间，约为 5.7 毫秒，即磁盘完成一次 I/O 的平均时间。（参见以下企业级硬盘 datasheet））。**对于磁盘存储系统来说，通过优化存储系统 Cache 算法，提高 Cache 命中率，增加磁盘数量，尽可能采用低容量、高转速的磁盘，是优化磁盘存储系统性能、增加 IOPS 和提高 I/O 响应时间的常用方法。**

规格	300GB ¹	146.8GB ¹	73.4GB ¹
型号	ST3300655LW/LC/FC/SS	ST3146855LW/LC/FC/SS	ST373455LW/LC/FC/SS
容量:			
格式化为 512 KB/扇区 (GB)	300	146.8	73.4
外部传输率 (MB/秒)			
Ultra320 SCSI	320	320	320
4 Gb/秒 FC	400	400	400
3Gb/秒串行连接 SCSI	300	300	300
性能			
转速 (RPM)	15K	15K	15K
平均延迟 (毫秒)	2.0	2.0	2.0
寻道时间			
平均读/写时间 (毫秒)	3.5/4.0	3.5/4.0	3.5/4.0
平均寻道读/写时间 (毫秒)	0.2/0.4	0.2/0.4	0.2/0.4
传输率			
内部 (MB/秒)	960 至 1607	960 至 1607	960 至 1607

而 SSD 固态硬盘存储系统，由于先天的结构优势，摒弃了传统磁盘存储系统的磁头移动加盘片旋转的机械物理运动，消除了磁盘系统中由于磁盘寻道加延时造成的大量 I/O 等待消耗时间，提供了极高的 IOPS 和 I/O 响应时间读写性能。

因此，越来越多的主流存储厂商，如 EMC、HDS、NetApp 等，都已经或准备在其中高端磁盘存储产品中，加入固态存储介质的支持，提供一种所谓的分层存储解决方案，将不同 I/O 性能要求的数据保存在相应不同 I/O 性能的存储介质中，以使存储系统的性能达到合理分布和最优化。目前已经有存储系统同时配置了适应极高性能要求的固态硬盘、中等性能要求的 FC 或者 SAS 硬盘、一般性能要求和低成本的 SATA 硬盘。例如，EMC CLARiiON CX4 型号 480、960 这两款存储阵列。

八、SPC-1 测试结果对于存储系统选择的指导意义

SPC-1 存储性能测试结果，并不只是理论结果，它的结果建立在模拟现实中的随机 I/O 处理系统，对于用户实际购买的存储系统，可以利用下面的公式，通过查询这些报告中的 SPC-1 IOPS™值和测试报告中所描述的被测存储产品的磁盘总数，依据已有的 SPC-1 测试结果，用户就可以根据下列公式快速估算所采购配置的某一存储产品实际性能：

实际性能 = [实际采购存储系统的磁盘数 / 测试报告中被测系统的磁盘数] × 测试报告的 SPC-1 IOPS™

(作者：徐健)

关注数据保留、邮件归档和法规遵从

在纽约召开的 Storage Decisions 研讨会上，一位专注信息管理的存储顾问称他们必须帮助企业认识到技术的现状和局限性，因为技术需要符合法规遵从。

“如果你不这样做，其他人也不会。”

“法律工作者们制定法规条文，非常的完美但却不实用，” FCS 咨询公司信息管理经理 Barclay Blair 这样说道，“必须有人来帮助企业了解技术发展的实际情况，必须有人来解释信息管理如何实施。”

在讨论中，Blair 给出了信息管理的定义——“人、过程和工具帮助企业更好的管理信息。”

Blair 认为邮件归档是当前企业信息管理的重中之重，由于保留了过多或者过少的信息，企业往往遭受了不少损失。Blair 称企业不可能保留所有的信息也不能把所有的信息都删除掉，他们必须根据法规的要求来决定什么该留，什么该删，留的该保存多长时间。

“但这是谁该负责的问题？是存储管理者？法务部门？诉讼律师还是业务经理？这其实是个很复杂的问题。但毫无疑问存储管理者在信息管理中决不能袖手旁观。”

Blair 说存储管理者在信息管理中的作用决定了在管理策略制定时他们必须有自己的看法。他认为存储管理者必须学习基本的法规原则，包括那些和企业相关的规定；存储管理者还需要告诉业务人员备份和归档的区别，帮助企业了解进行存储归档所需的软、硬件花费，精确到每一 Gb 需要多少钱。

解决这些问题包括创建基础架构（策略）、实施（采购工具和培训）、修复（对过去的一些东西进行重新调整）以及连续提高（审计和调整）。目的只有一个就是降低风险、节约资金。

Blair 说：“清理和调整过去的一些东西是基础，这会伴随着今后的发展。我们需要先创建一个我们想要进入的世界。”

Blair 建议存储管理者必须访问 PST/NSF 文档，实施一个法律或诉讼执行程序，应用存储和归档方案减少卷，并使用数据分类工具。他建议不要在没有数据保留和归档策略的情况下实施邮箱大小的限制，也不要让员工自己决定多长的邮件需要被保留。

他还补充道邮件是当今的罪魁祸首，因为对于法规遵从来说，邮件就是灾难。地球上没有人决定使用 Email，而忽然有一天邮件就充斥每一个角落了。但是没有人考虑到法规遵从对技术的要求和影响。“这个世界到处都是电子邮件，而我们对此失去了控制。”

然而，企业必须对社会网络的伸展及其他新型协作工具进行管理。“如果我们打算使用这些工具，我们就必须对其进行限制、规划，最终推向社会化。” Blair 说道。

(作者：Dave Raffo 译者：吴迪)

探讨云存储技术的优劣

通常云存储被用来替代或者补充传统的数据存储网络，它的定义还没有最终确定。Storage Decisions 的企业数据存储专家通过具体的案例给我们讲解了云存储的优势和劣势。

Beth Israel Deaconess 医学中心 (BIDMC) 的存储架构师 Michael Passe 表示，BIDMC 计划采用 EMC 公司的私有云存储产品 Atmos 来整合现有的 SAN 和 NAS。另外，GlassHousek 科技公司的高级顾问 Ron Scruggs 给我们全面分析了云存储和本地归档存储的不同。

虽然有时采用云存储会影响性能和功能，而且可靠性，安全性和厂家的承诺有时也得不到保证，但专家仍然认为云存储是一种更为廉价的替代方案。Passe 表示，除节约资金外，采用云存储还可以获得的其他好处。

许多人肯定会问，“这是真正的云存储吗？是不是只是一些便宜的存储设备而已？”起初，它可能只是些廉价的存储设备，但从长远来看，它的价值就会显现出来，因为它可以在内部或者外部或者同时向用户提供多租户和策略导向的存储。

Passe 表示，“并不是所有的用户都适合使用云存储，但是绝对值得去看一下。

法律规定医疗测试的结果需要保存几十年，而这些数据通常占据大量的硬盘空间，这迫使 Passe 必须选择一些相对便宜的存储设备。“我每天都要存储各个科室扫描的文件，大部分的存储任务都是临时才告诉我，而且这些数据必须要永久保存。

Passe 将提供云服务的厂家分为三类：内部私有云软件和硬件厂家，公共云服务供应商和公共云应用程序供应商。

公共云服务供应商包括 Amazzon S3 , Nirvanix 公司和 Rackspace 公司。公共云应用程序供应商包括 Carbonite , Mozy (数据备份) , MobileMe (桌面和设备同步) , Flickr 和 Snapfish (照片存储及共享) 。内部私有云存储供应商包括 Bycast 公司 , Caringo 公司 , DataDirect Networks , EMC , ParaScale 公司以及 Permabit 科技公司。

Passe 表示,内部私有云存储主要应用在企业内部,它可以更有效的控制安全性,更加方便的制定保护策略以及更加有效的降低成本。不过这种方法在其他方面可能会出现问题,例如,对厂家过份依赖,NAS 协议支持有限,厂家承诺能否兑现等。

他预计 BIMDC 最终可能会采用内部云存储和公共云存储混合的模式。内部云存储可以更加安全的存储病人的医疗记录。外部云存储则可以更加高效的跟合作者进行数据复制。

个人部署还是租赁云服务？

Scruggs 把"选择本地归档还是选择云服务？"比喻为"自己建一座房子还是租用其他人的空间？"

云服务较本地归档程序相比是一种更为廉价的选择。设备管理得到有效简化,而且用户不用操心技术的变更。但在可靠性,安全性和访问时间方面,云服务还有待提高。

Scruggs 跟 Passe 在很多方面都持有相同的观点,例如,价格,对供应商认真审核,云存储定义难以确定等等。

云存储价格

假设有 40T 的数据,每年的检索率是 33%,使用云存储时三年的费用是 227,000 美元,而使用本地归档则需要 973,000 美元。如果是 100TB 数据,云存储是 566,000 美元,本地归档是 1.475 百万美元。本地归档的费用包扩交换和离线磁带备份方面的花费。云存储的花费不包括请求服务和将数据迁入云内的费用。

BIDMC 采用的全部都是 EMC 的产品,每种数据的费用如下:在 Cisco SAN 环境中,运行在存储第一层设备 Symmertrix 上的块数据,例如 Exchange , SQL 和 Oracle ,花费是每 GB12 美元,在存储第二层 Clariion 设备上的花费是每 GB 4 美

元。非结构化的数据通过 F5 ARX 交换机存储到 Celerra NAS 设备上，采用 Data Domain 和 Centera 分别进行备份和归档数据，费用分别是\$.50/GB 和 \$8/GB。通过 SOAP 存储在 Atmos 的数据花费是\$.70/GB。

核实清楚

在确定云服务商之前，确保所有关键的问题都已经被解决。Scruggs 和 Passe 都表示，仔细阅读服务水平协议是非常必要的。

"如果数据出现了问题，谁来负责恢复？如果数据发生泄漏，是谁的问题？" 这些问题都要问清楚。

有一些云服务提供商在服务水平协议中对恢复保证和数据破坏保证并没有明确的规定，这需要用户自己去跟服务提供商协商。

我们曾经见过“云”吗？

如果用户感觉云计算好像并没有那么陌生，那就对了。专家建议不要过分关注云的定义。

Passe 表示，“它只是老技术的一种时髦说法。对用户和用户的业务能够产生什么影响完全由用户自己决定。”

Scruggs 表示，“云计算事实上是分布式计算的一种，并不是新技术”。

(作者：Dave Raffo 译者：曹同举)

我们的编辑团队

您若有何意见与建议，欢迎进入[《存储经理人》讨论区](#)，或者[与我们的编辑联系](#)。

诚挚感谢以下人员热情参与 TechTarget 中国《存储经理人》的内容编辑工作！

诚邀更多的存储专业人士加入我们的内容建设团队！

Rich Castagna



TechTarget 存储媒体集团主编，全面负责存储网站及电子杂志的内容建设，包括：《Storage》、

www.SearchStorage.com

www.SearchDataBackup.com

www.SearchSMBStorage.com

www.SearchDisasterRecory.com

www.Storage.co.UK

www.SearchStorageChannel.com



Marc Staimer

Dragon Slayer 咨询公司总裁、高级分析师，熟悉 SAN 存储、服务器以及 I/O 市场等，对企业存储战略颇有研究。



Alan Radding

Storage 杂志特邀技术撰稿人，专注于 IT 产业相关话题。



Dave Raffo

TechTarget 存储媒体集团高级新闻记者，负责存储网站的内容建设：www.SearchStorage.com。



Terri McClure

ESG (Enterprise Strategy Group) 分析师，专注于 NAS、文件系统、文件系统管理以及基于文件的存储虚拟化。



Steven Fu

TechTarget 中国特邀技术编辑。目前供职于某国际知名存储厂商的全球技术支持中心，负责 NAS 产品全球技术支持。自 2003 年起曾先后在五百强外企从事 UNIX、IP、存储相关产品技术支持。



曹一玺

TechTarget 中国特邀技术编辑。十二年 IT 从业经历，曾就职于 ChinaLife、ICSS、Teamsun 等名企，对系统软硬件架构有较全面的认识。现为某国际著名存储厂商技术顾问。



郭镭

TechTarget 中国特邀技术编辑。七年存储行业工作经验，曾服务于浪潮、GDS 等公司，主要从事存储网络、备份容灾、云存储、ITIL 管理等技术工作。目前就职某知名存储厂商，任技术顾问。



徐健

TechTarget 中国特邀技术编辑。现就职于一家专业存储备份系统集成商，企业级存储/备份解决方案售前咨询工程师，担任 TechTarget 中国“存储技术-厂商资料交流区”版主。



曹同举

TechTarget 中国特邀技术编辑。现就职于一家外企做软件技术支持工作。曾任职国内某著名的服务器存储厂商，拥有丰富的售后服务经验。

STORAGE

Tony Tang

TechTarget 中国特邀技术编辑。目前就职于存储领域知名外企。



谢小丽

TechTarget 中国地区总经理。毕业于美国 Northeastern University，获得计算机信息系统硕士学位。同时还拥有湖南大学语言学硕士学位和英语学士学位。



吴迪

TechTarget 中国存储编辑，五年网络媒体从业经验，负责“TT 存储”（www.SearchStorage.com.cn）和《存储经理人》的内容建设。拥有英语语言文学学士学位。

与我们联系

投稿

诚邀关注存储技术与发展、从事存储行业的专业人士共同参与《存储经理人》的内容建设！请联系

吴迪

wudi@techtarget.com.cn

010-62681688 转 803

广告合作

罗湘君

luoxiangjun@techtarget.com.cn

010-62681688 转 817