**VMware与华为存储技术比较**

[1. Data Protection 3](#_Toc492395079)

[1.1 VMWare 3](#_Toc492395080)

[1.1.1 一致性 3](#_Toc492395081)

[1.1.2 通用性 3](#_Toc492395082)

[1.1.3 API支持 4](#_Toc492395083)

[1.1.4 参考 5](#_Toc492395084)

[1.2 华为 5](#_Toc492395085)

[2. Storage vMotion 5](#_Toc492395086)

[2.1 VMWare 5](#_Toc492395087)

[2.1.1 功能 5](#_Toc492395088)

[2.1.2 限制 5](#_Toc492395089)

[2.1.3 性能 5](#_Toc492395090)

[2.1.4 参考 6](#_Toc492395091)

[2.2 华为 6](#_Toc492395092)

[3. vSphere Replication 6](#_Toc492395093)

[3.1 VMWare 6](#_Toc492395094)

[3.1.1 功能 6](#_Toc492395095)

[3.1.2 限制 6](#_Toc492395096)

[3.1.3 性能 6](#_Toc492395097)

[3.1.4 参考 7](#_Toc492395098)

[3.2 华为 7](#_Toc492395099)

[4. vCenter备份和还原 7](#_Toc492395100)

[4.1 VMWare 7](#_Toc492395101)

[4.1.1 功能 7](#_Toc492395102)

[4.1.2 限制 7](#_Toc492395103)

[4.1.3 性能 7](#_Toc492395104)

[4.1.4 参考 7](#_Toc492395105)

[4.2 华为 8](#_Toc492395106)

[5. Storage Policy Based Management 8](#_Toc492395107)

[5.1 参考： 8](#_Toc492395108)

[6. Storage DRS 8](#_Toc492395109)

[6.1 VMWare 8](#_Toc492395110)

[6.1.1 功能 8](#_Toc492395111)

[6.1.2 限制 9](#_Toc492395112)

[6.1.3 性能 9](#_Toc492395113)

[6.1.4 参考 9](#_Toc492395114)

[7. Storage I/O Control 9](#_Toc492395115)

[7.1 功能 10](#_Toc492395116)

[7.2 限制 10](#_Toc492395117)

[7.3 性能 10](#_Toc492395118)

[7.4 参考 10](#_Toc492395119)

[8. Storage APIs for Array Integration 10](#_Toc492395120)

[8.1 VMWare 10](#_Toc492395121)

[8.1.1 功能 10](#_Toc492395122)

[8.1.2 限制 11](#_Toc492395123)

[8.1.3 性能 11](#_Toc492395124)

[8.1.4 参考 11](#_Toc492395125)

[9. Storage APIs for Multipathing 11](#_Toc492395126)

[9.1 VMWare 11](#_Toc492395127)

[9.1.1 功能 12](#_Toc492395128)

[9.1.2 限制 12](#_Toc492395129)

[9.1.3 性能 12](#_Toc492395130)

[9.1.4 参考 12](#_Toc492395131)

# Data Protection

## VMWare

### 一致性

VMWare数据保护方案中最重要是Snapshot，大部分和磁盘相关的操作（如Replication、Backup、数据迁移等）会在开始工作之前创建一个临时快照，以此保证磁盘数据的一致性。VMWare在创建快照时会通过VMWare Tool冻结虚拟机，并将缓存数据和内存数据刷新到磁盘，当快照创建完毕后解冻虚拟机恢复磁盘的读写。

按照快照的一致性程度，可以分为以下几种类型：

**不一致快照：**

指乱序对文件进行复制，导致快照中不同文件的时间戳不一致。不一致快照不会处理缓存或者内存中的任何数据。

**崩溃一致快照：**

崩溃一致快照同样不会处理缓存或者内存中的任何数据，但是能够保证文件的时间戳一致。崩溃一致快照的状态相当主机突然断电时的状态，是最为常见的快照类型。

**文件系统一致快照：**

文件系统快照在进行快照之前将虚拟机静默，并刷新缓存或者内存中的数据到磁盘，以此保证文件系统的一致性，快照完成后解冻虚拟机。创建文件系统一致快照需要系统的内核支持，如在较新的 Windows 客户机上，Windows 提供了 VSS（Volume Shadow Copy Service） 服务，它可以通过 requester-writer 方式来实现有冻结需求的应用和文件系统，在快照之前进行冻结和快照之后进行解冻，而在VMWare的解决方案中，VMWare Tool工具提供了冻结、解冻虚拟机的能力。

**应用一致快照：**

需要这种一致性级别的最常见应用程序类型是数据库和消息传递系统， 它们需要确保事务在应用程序中提交并刷新到操作系统内的磁盘，从而确保备份过程中不 会丢失任何事务。一般情形下实现这种级别的快照，需要进行专门的开发。

VMware 生态系统中的大多数数据保护解决方案涉及到快照，而快照的一致性很大程度上依赖于 VMware tools 休眠，所以通常情形下VMWare的数据保护能够保证文件系统这一级别的数据一致性，无法提供应用程序级别的数据一致性。

### 通用性

在VMWare中虚拟机的磁盘设备本质上是一个VMDK文件，这些文件存放在VMWare的定制共享文件系统VMFS上。VMWare的数据保护策略绝大多数是基于VMDK文件，与存储设备无关，有相当好的通用性。

VMFS是一个共享文件系统，工作在存储设备和VMWare虚拟机之间的层级，VMFS的工作结构如下图。



相比OpenStack中为每个ISCSI卷提供一个独立LUN的做法，VMFS将所有的卷整合到一个公共的存储池中。两种方案各有优势，下面简要分析整合方案和隔离方案的区别：

**整合方案（VMWare、FusionCompute）：**

**优势：**

1. 屏蔽了设备上的差异性，存储仅需提供存储空间（数据面）。用户在创建VMFS环境时，只需要提供一个空间足够的LUN，就可以为众多虚机提供磁盘设备。

2. VMFS可以实现自己的快照、Replication、Backup等功能，而不需要依赖存储设备。

**劣势：**

1.整合方式的磁盘管理，会使产生资源竞争，导致I/O性能降低。

**隔离方案（OpenStack）：**

**优势：**

能完全发挥存储设备的I/O性能；

**劣势：**

由于绝大多数存储设备LUN数目存在上限，隔离的方案很容易到达这个上限。

上述的两种存储管理方案各有利弊，合理解决方法肯定是一种混合方案，即：为关键数据提供独立的LUN设备，其他磁盘使用整合存储池中的存储空间。VMWare和华为都是以整合存储池的方式管理虚拟机的磁盘，但是这些厂家同时也支持直连LUN的方式。

### API支持

vStorage APIs for Data Protection (VADP)是VMWare提供的一组专门的API接口集，可供备份产品用于在 VMware 虚拟机上执行备份。事实上，大多数第三方备份产品均围绕该工具集而构建，并对其频繁加以利用。VADP 并不执行实际的备份，而是提供解决方案可以实施的大量备份相关功能。当前，备份产品利用的最多的两个 VADP 组件是变更数据块跟踪 (CBT) 和备份传输模式。

### 参考

VMware灾难恢复和数据保护专家指南.pdf

VMware-VMFS最佳实践白皮书.pdf

VMware官方数据保护指导：

https://docs.vmware.com/en/VMware-vSphere/6.5/rn/data-protection-615-release-notes.html

# Storage vMotion

## VMWare

Storage vMotion是VMWare提供的存储设备热迁移功能，原理基于VMWare中的vMotion功能（VMWare的虚机迁移）：

1. 对待迁移的VMDK磁盘创建快照。在VMWare中磁盘实际上是一个VMDK文件，创建快照的原理和NFS文件系统类似，是一种层叠式的快照。 创建了快照之后磁盘文件分为两部分：只读的BASE部分和可读写的TOP部分。
2. vCenter先拷贝BASE部分到指定的迁移设备，由于BASE只读因此不会有数据一致性的问题。
3. vCenter拷贝TOP部分到迁移的存储设备，由于TOP可以进行读写，因此拷贝TOP时需要停止虚机的活动，或者将磁盘设定为只读。
4. 将VM磁盘重定向到迁移设备上。
5. 删除源磁盘文件。

上述过程中步骤3、步骤4是需要暂时锁定虚机的，根据网上收集的资料来看这个卡顿时间大概时1到2秒左右。

### 功能

实现磁盘设备的热迁移，以及变更存储空间配置模式（如厚分配切换为瘦分配）。

### 限制

1. 对虚机的磁盘类型有一定要求，裸设备映射 (RDM)或者处于持久状态下磁盘都能够进行迁移；

2. 要求证书；

3. 要求宿主机能够同时访问源存储和目标存储；

### 性能

迁移过程中虚机无需关机，但是会出现1至2秒的卡顿。迁移的速度和磁盘的大小、网络环境、设备读写速度有关。

### 参考

功能简述：

http://pubs.vmware.com/vsphere-51/index.jsp?topic=%2Fcom.vmware.vsphere.vcenterhost.doc%2FGUID-A16BA123-403C-4D13-A581-DC4062E11165.html

限制和要求：

http://pubs.vmware.com/vsphere-51/index.jsp?topic=%2Fcom.vmware.vsphere.vcenterhost.doc%2FGUID-A16BA123-403C-4D13-A581-DC4062E11165.html

配合vMotion共同使用时的限制：

http://pubs.vmware.com/vsphere-51/index.jsp?topic=%2Fcom.vmware.vsphere.vcenterhost.doc%2FGUID-A16BA123-403C-4D13-A581-DC4062E11165.html

参考原理：

<http://blog.sina.com.cn/s/blog_64da692d01011mo2.html>

# vSphere Replication

## VMWare

vSphere Replication 是 VMware vCenter Server 的扩展（vCenter Server是vSphere的管理模块，包含HA，虚机迁移等功）。官方产品文档中对其的描述是：“提供基于管理程序的虚拟机复制和恢复功能”。

完整的vSphere Replication设备包括：

1. 为 vSphere Replication 提供用户界面的 vSphere Web Client 插件；

2. 存储复制配置和管理信息的嵌入式数据库。

3. 提供 vSphere Replication 基础架构内核的 vSphere Replication Management Server 和 vSphere Replication Server。

在主站点、备站点都需要部署vSphere Replication设备。

### 功能

提供虚拟机在不同站点间的复制功能，以及从指定Replication恢复虚机的功能。

### 限制

1. 在每个主机上 vSphere Replication Management Server 最多可管理 500 个复制的虚拟机，并且在VMWare建议单个节点只部署一个vSphere Replication Management Server（不会禁止部署多个，但可能会导致意外结果）。

2. 产品文档中提及配置vSphere Replication后不支持对磁盘进行迁移（vSphere Storage vMotion）。

### 性能

VMware vSphere Replication的对象是虚机，并不是传统方案中的磁盘。此功能正是为了取代传统的“存储的复制”方案，有如下优势：

1. Relication是基于vSphere自身的能力，允许灵活选择主站点和辅助站点的存储器供应商，成本较低。

2. 采用增量复制的方案（除开第一次），每次复制的总体成本较低。

### 参考

vmware产品文档：

<http://pubs.vmware.com/vsphere-51/index.jsp?topic=%2Fcom.vmware.vsphere.replication_admin.doc%2FGUID-35C0A355-C57B-430B-876E-9D2E6BE4DDBA.html>

# vCenter备份和还原

## VMWare

vCenter备份/恢复功能的对象主要是虚机的磁盘文件（VMWare中每个磁盘用一个vmdk文件表示），通常情形一个磁盘包括三种类型vmdk文件：

1. xxx.vmdk元数据文件；

2. xxx-flat.vmdk数据文件；

3. xxx-ctk.vmdk CTK 文件；

vCenter对磁盘的备份是增量备份（当然第一次必须全量），创建备份时vCenter对磁盘创建快照，之后根据CTK文件比较当前磁盘数据和备份文件的差异，对差异部分进行拷贝。vCenter进行恢复时虚机需要关闭虚拟机，可以将虚机恢复到指定的备份副本。

### 功能

为虚机的磁盘文件创建备份，通过这些备份能够恢复虚拟机到指定备份的状态。备份程序都可以调用 VDAP 的 Connect/ConnectEx 接口来建立和 vmdk 的连接的，当传输完成之后断开连接（即备份程序可以通过VMWare的接口自己开发）。

### 限制

没有发现啥限制。

### 性能

由于备份过程基于快照机制，因此备份操作对虚机的性能影响不大。增量备份功能基于CBT（Changed Block Tracking）因此开启了这个功能后会损失一些IO（因为需要将读写变换记录到ctk文件中）。传输模式也会影响性能，VMWare支持的传输包括SAN（备份所在的主机通过FC、ISCSI等存储协议访问主机）、LAN（ESX/ESXi 主机从其存储中读取数据，再通过 LAN 网络发到备份程序所在的主机。通用性高、需要时可以通过SSL加密）。

### 参考

<http://www.cnblogs.com/sammyliu/p/5661085.html>

# Storage Policy Based Management

Storage Policy Based Management （SPBM） 是一种通用的存储策略框架，是 VMware的SDS框架。SPBM可以看做VM存储配置文件的演进，其目标是根据VM存储策略配置模板为虚拟机提供特定级别的容量，性能，可用性，冗余性等。

SPBM的实现包括以下三个方面：

**存储设备特性：**在SPBM中存储设备通过vStorage APIs将设备的特性通知vCenter Server，在这些特性的基础上，用户可以定义相应的Metadata标签，vSphere也会定义一些存储的公共特性。

**规则：**规则通过键值对表示，键表示某种特定的存储特性，值是对特性的具体数值或者特征要求。

**规则集：**规则集中包含多个规则，但是这些规则必须由唯一一个存储设备提供。SPBM中提及的存储策略实际上就是多个规则集的聚合。

在SPBM中我们可以为一个VM配置不同的存储策略，分别应用于不同磁盘。VMWare推荐见操作系统和应用程序分别存放在不同的磁盘，因此一个典型的SPBM场景就是：我们创建虚机时指定两个存储策略，策略一分配普通磁盘给操作系统，策略二分配高性能SSD给应用（如数据库应用），整个过程只需指定策略，其他与设备相关配置通过SPBM自动完成。

SPBM和传统storage框架不同的是，SPBM能最大化的减小存储管理的配置工作，为虚机提供更加精确的存储服务。同时SPBM不但能够在创建新虚拟机时发挥作用，也能为原先的旧虚拟机进行存储设备的调整。

## 参考：

vmware官方博客：

https://blogs.vmware.com/vsphere/2014/09/storage-policy-based-management-overview.html

<https://blogs.vmware.com/vsphere/2014/10/vsphere-storage-policy-based-management-overview-part-2.html>

# Storage DRS

## VMWare

Storage DRS（Storage Distributed Resource Scheduler）即VMWare在vSphere 5中提供的存储分布式资源调度程序。Storage DRS将现有的资源的整合在一个资源池里，然后设置相关的阕值（优先级等），当存储资源过载的时候会进行自动重新分配存储。从而实现7x24小时的服务。

### 功能

Storage DRS提供的能力包括以下几点：

1. 整合原先的datastore形成存储设备集群；
2. 创建VM动态定位存储位置（类似Cinder Scheduler）；
3. 存储集群负载均衡，包括：全自动的存储磁盘迁移，迁移推荐（此时要手动迁移），这里的迁移通过上文中提到的Stroage vMotion特性；
4. 维护模式，当某个datastore维护时可以将他设定成维护模式，此时属于该datastore磁盘会自动迁移到其他的datastore。一般情形下，生产环境中会有多个datastore，每个datastore表示一个存储设备。
5. 负载均衡的规则，除了对存储空间进行负载均衡，Storage DRS还支持对设备IO性能（计算IO延时）进行负载均衡。

### 限制

1. 单个datastore的限制9000个VMDK；
2. 单个datastore集群限制64个datastore；
3. 单个vCenter限制256个datastore集群；

### 性能

Storage DRS本质上是通过Storage vMotion进行磁盘迁移，因此性能主要取决于Storage vMotion功能。Storage DRS本身有全自动模式和手动模式，全自动模式下会自动进行负载迁移，手动模式下VMWare只是推荐那些磁盘需要迁移，是否迁移取决于管理员。由于一次磁盘迁移需要进行拷贝操作，会产生大量的额外磁盘IO，因此推荐使用手动模式。

### 参考

产品文档说明：

https://docs.vmware.com/cn/VMware-vSphere/6.5/com.vmware.vsphere.resmgmt.doc/GUID-8ACF3502-5314-469F-8CC9-4A9BD5925BC2.html

博客资料：

<http://virtual.51cto.com/art/201109/294564.htm>

<http://delxu.blog.51cto.com/975660/629721/>

# Storage I/O Control

存储I/O控制（SIOC）在资源争用期间，提供一个方法来公平的分配存储I/O资源。SIOC启用后，能够阻止其它的情况下的干扰（同一个数据存储上的一台虚拟机能对另一台虚拟机造成负面的影响）。

当对数据存储启用 Storage I/O Control 时，ESXi 会开始监控主机与该数据存储通信时出现的设备延迟时间。当设备延迟时间超出阈值时，该数据存储会被视为出现拥堵，此时将按访问该数据存储的每个虚拟机的份额比例向其分配 I/O资源。您可以设置每个虚拟机的份额，并根据需要调整每个虚拟机的份额数量。

配置 Storage I/O Control 分为两个步骤：

1. 为数据存储启用 Storage I/O Control。
2. 设置每个虚拟机所允许的存储 I/O 份额数量以及每秒 I/O 操作数 (IOPS) 的上限。

默认情况下，所有虚拟机的份额均设置为“正常 (1000)”且 IOPS 无限制。

## 功能

1. 根据虚拟机的存储I/O份额数量以及IOPS，防止I/O竞争。
2. 允许用户开发第三方I/O筛选器。这些I/O筛选器是安装在ESXI主机上的监控程序，可以监听特定的VMDK的IO信息，并反馈给ESXI。

## 限制

1. vSphere 4以前对存储设备有限制（不支持NFS设备），vSphere 5以后的版本已经全面支持了所有存储类型。

## 性能

无

## 参考

原理分析：

<http://virtualbox.blog.51cto.com/531002/1213279/>

功能描述：

https://cormachogan.com/2012/09/17/vsphere-5-1-storage-enhancements-part-8-storage-io-control/

SIOC功能白皮书（vSphere 4.1）：

<https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/vmware-vsphere41-sioc-white-paper.pdf>

vSphere产品文档：

https://docs.vmware.com/cn/VMware-vSphere/6.5/com.vmware.vsphere.resmgmt.doc/GUID-7686FEC3-1FAC-4DA7-B698-B808C44E5E96.html

# Storage APIs for Array Integration

## VMWare

Storage APIs for Array Integration，简称VAAI，是一组用于ESXi和存储设备通信的API。VAAI的目标是通过存储设备的硬件加速，提高一些I/O密集操作的性能。在VAAI的帮助下，如迁移、克隆等操作可以在存储设备内部执行，无需占用额外的资源。

主要应该场景包括：工作负载下放到硬件和设备状态监控。

### 功能

1. XCOPY (Extended Copy)

存储设备提供的从一个卷到另一个卷的智能拷贝过程（注意是Copy不是同一个卷内的克隆）。

1. ATS

ATS是一种共享文件加锁技术，是VMFS的SCSI Reservation Lock技术的加强版。这种锁定技术允许来自不同ESX主机上的多个访问可以到达同一个数据文件，而更重要的 一点是可以避免发生冲突。

1. Write Same (Zero)

优化磁盘的清零过程。在虚拟磁盘上常见的操作之一是初始化磁盘的大范围，并以零隔离虚拟机和促进安全。通过Write Same特性能够节省SCSI的清零命令，提升该过程的速度。

1. Thin Provisioning Stun

精简数据存储的使用率达到了100％的容量时，只有那些需要额外存储空间的虚拟机受到影响。 那些不需要额外空间的人继续运行。 额外的空间分配给精简配置的数据存储区，可以恢复暂停的虚拟机。

1. Thin Provisioning Space Threshold Warning

当精简配置的存储设备没有可用存储空间时，通过该VAAI接口可以向vCenter发出告警。

### 限制

启动VAAI能力需要存储厂商支持，并非所有存储阵列都支持VAAI。

### 性能

属于性能提升的辅助措施。

### 参考

白皮书下载：

<https://www.vmware.com/techpapers/2012/vmware-vsphere-storage-apis-array-integration-10337.html>

# Storage APIs for Multipathing

## VMWare

Storage APIs for Multipathing简称为VAMP，提供了一个可插拔存储架构框架包，允许合作伙伴提供针对每个阵列优化的性能增强，多路径和负载平衡插件。

VMWare的虚拟层（VMkernel layer）之上有一层用于管理存储多路径的框架，称为Pluggable Storage Architecture (PSA)。PSA可以加载第三方开发的插件，使用户指定存储设备的负载均衡、路径选择、failover等策略。

用户可以开发三种类型的PSA插件： Multi-Pathing Pluin (MPP)、Storage Array Type Plug-ins (SATPs)、Path Selection Plug-ins (PSPs)，三种类型的插件构成一个多路径扩展模块（MEMS）。MEMS模块能够帮助I/O设备进行：路径负载均衡、路径故障切换、传输性能提升。

VMware提供了一个通用的多路径插件（MPP），称为本地多路径插件（NMP），随vSphere一起提供，是默认的多路径。 NMP插件管理物理路径声明、 逻辑设备的注册、将物理路径与逻辑设备相关联、处理对逻辑设备的I / O请求、负载平衡、处理故障以及重试请求、 并支持管理任务（如中止或重置逻辑设备）。NMP是一个可扩展的模块，允许在其上构建两种插件：SATP和PSP。

### 功能

存储设备多路径插件开发框架，允许开发多路径扩展模块实现：路径负载均衡、路径故障切换、传输性能提升。

### 限制

存储设备的多路径能力需要存储厂商支持，并非所有设备、协议都支持多路径。

### 性能

属于性能提升的辅助措施。

### 参考

PSA框架包说明：

https://code.vmware.com/vmware-ready-programs/storage/psa