**[Xen 4.0新特性介绍](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html)**

**虚拟化概述**

在讨论Xen之前首先谈谈虚拟，虚拟（Virtualization）指对计算机资源的抽象，一种常用的定义是“虚拟就是这样的一种技术，它隐藏掉了系统，应用和终端用户赖以交互的计算机资源的物理性的一面，最常做的方法就是把单一的物理资源转化为多个逻辑资源，当然也可以把多个物理资源转化为一个逻辑资源（这在存储设备和服务器上很常见）” 实际上，虚拟技术早在20世纪60年代就已出现，最早由IBM提出，并且应用于计算技术的许多领域，模拟的对象也多种多样，从整台主机到一个组件，其实打印机就可以看成是一直在使用虚拟化技术的，总是有一个打印机守护进程运行在系统中，在操作系统看来，它就是一个虚拟的打印机，任何打印任务都是与它交互，而只有这个进程才知道如何与真正的物理打印机正确通信，并进行正确的打印管理，保证每个任务（job）按序完成。

长久以来，用户常见的都是进程虚拟机，也就是作为已有操作系统的一个进程，完全通过软件的手段去模拟硬件，软件再翻译内存地址的方法实现物理机器的模拟，比如较老版本的VMWare, VirtualPC 软件都属于这种。

在2005年和2006年，Intel和AMD都开发出了支持硬件虚拟技术的CPU，也就是在这时，x86平台才真正有可能实现完全虚拟化。在2007年初的时候，Intel还进一步的发布了VT-d技术规范，从而在硬件上支持I/O操作的虚拟化，而在2009年随着Intel-i系列芯片的出现，下一代硬件虚拟化技术也随之诞生。随着整机处理能力的不断提高和虚拟化技术的发展，虚拟机已经成为云计算的基础架构核心，并开始深入人们的生活。

**Xen概述**

在2003年国际操作系统会议*Symposium on Operating Systems Principles* (*SOSP*)上，一篇名为*Xen and the Art of Virtualization*的论文正式宣告了Xen虚拟机的诞生。随后在2007年，Citrix公司接管了该项目，并一直致力于Xen的维护和更新工作。

Xen作为最具历史和最著名的Linux开源虚拟机已经走过了7年。在这7年中，Xen已经先后支持x86, x86\_64, 安腾（Itanium）等硬件平台。不同于传统进程虚拟机需要依赖于外部先启动的操作系统（就像Windows上的VMWare），Xen的设计目的在于开机后先启动虚拟机监控器（Hypervisor），再启动宿主OS（Host OS或Native OS）和各个虚拟机，形成一个与前者相反的过程。需要注意的是，Hypervisor并不是一个图形界面，如果以Windows操作系统比喻的话，虚拟机管理辅助工具就像是Windows界面，而Hypervisor则是Windows内核。这样做的好处在于，通过尽量减小Hypervisor和Native OS之间的联系而降低Hypervisor自身和各个虚拟机被破坏和信息泄露的风险。

为了达到这样的设计目的，Xen设计架构图如图1所示，其中Dom0指虚拟机辅助主机，在Xen架构中是不可缺少的，换句话说，任何原因造成的Dom0崩溃会导致整个虚拟机系统崩溃。而DomU则指Dom0启动后创建的虚拟机，可以动态添加和删除。

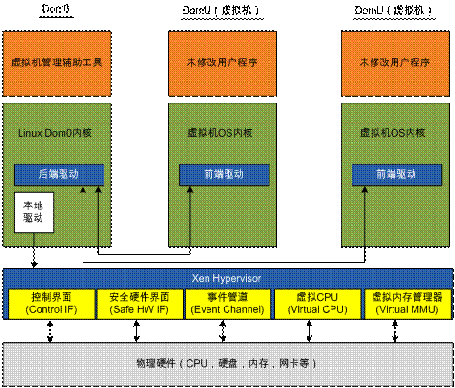
[[](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image002_2.gif)](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image002_2.gif" \t "_blank)

图1. Xen架构图

**Xen4.0新特性**

Citrix于2010年4月7号发布Xen 4.0，其新特性包括了对可扩展性、性能、可靠性和可用性的提升，下面我们将逐一介绍每个方面。

**可扩展性：**

在Xen3.x时代，对宿主机最大支持CPU数量为64，如图2所示，该图显示了两个可在64颗CPU中前16颗CPU上调度的虚拟机，图中的CPU Affinity（亲核性，指系统运行在哪个CPU上）指明这点.

而在Xen4.0中提供了更强大的可扩展性，Xen4.0支持安装1TB内存以及128颗物理CPU的宿主机，同时对虚拟机也可支持到128个虚拟CPU（VCPU）。如果将这些CPU全部用来虚拟Win7会如图3所示：

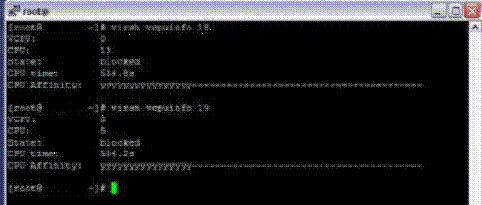
[[](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image004_2.gif)](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image004_2.gif" \t "_blank)

图2. 两个可在前16CPU/64CPU上调度的虚拟机[[1]](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html" \l "_ftn1_4540" \t "_blank)

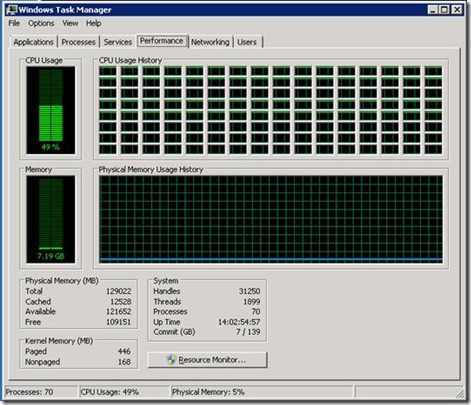
[[](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image006_2.jpg)](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image006_2.jpg" \t "_blank)

图3 一个拥有128核的Win7[[2]](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html" \l "_ftn2_4540" \t "_blank)

**性能：**

对于虚拟系统而言，所谓性能既包括虚拟机的运行性能和使用率，又包括虚拟化软件对宿主机造成的性能影响。Xen4.0性能上的提升分为两大类，第一类是对高带宽消耗的硬件进行了性能上的优化,比如：

1. 普通PCI设备：通过优化对Intel VT-d和AMD IOMMU（I/O内存管理单元，也称为I/O虚拟化）技术的使用来提升虚拟机对PCI设备的访问效率。

2. 显卡：Xen4.0支持虚拟机对宿主机显卡的独占式访问，从而大大提升了在虚拟机中图形图像的处理速度。有过一段时间玩虚拟机经验的朋友知道，过去无论所装显卡能力有多强，在虚拟机中玩3D游戏总是图像跳帧甚至根本不能玩，这主要是由于虚拟机需要引入额外的内存读写或建立共享内存而大大降低了显卡在虚拟机中的性能所致。Xen4.0出现这项技术之后，用户可以将机器上额外的显卡设备指定给某个虚拟机独占使用，从而避免这些开销而获得与宿主机一样的图像体验。

3. 网卡：Xen4.0对网卡性能的提升主要在于引入了多队列机制和对SR-IOV（单点I/O虚拟化）网卡的支持。后者是最近刚刚兴起的一种I/O设备虚拟化技术，这种技术要求I/O设备有专门的硬件实现。Intel VT-d技术为了保证高性能使用设备，所以将某个PCI设备完全指定给某个虚拟机，这就造成一旦某个虚拟机拥有设备，则在这个虚拟机被完全关闭之前，其它虚拟机甚至宿主机也不能访问该设备，这显然违背了虚拟化技术中“共享”的原则。而SR-IOV技术则在保证高性能使用设备基础上添加了“共享”的功能。以SR-IOV网卡而言，可以将这种网卡视作若干个虚拟网卡（当前一般为8个，最多64个）和一个交换机的组合各个虚拟网卡之间完全隔离，这就使得每个虚拟机可以拥有一块独占的虚拟网卡，而宿主机可以对所有虚拟网卡进行管理。

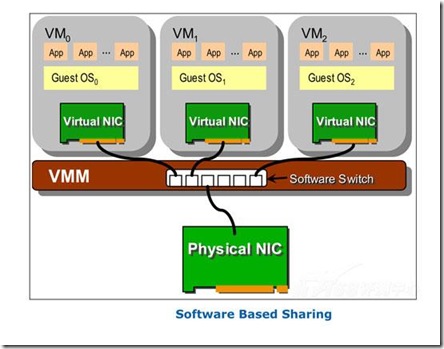
[[](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image008_2.jpg)](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image008_2.jpg" \t "_blank)

图4. SR-IOV设备分配示意图[[3]](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html" \l "_ftn3_4540" \t "_blank)

4. USB设备：引入对USB设备的并行虚拟化，从而提高了虚拟机访问USB设备的吞吐率。之前的方案是在通过模拟为USB1.1设备来实现的，相比之下，后者访问速度较慢且仅支持部分USB设备。

第二类则是结构优化带来的性能提升，比如：

1. TMEM技术（Transcendent Memory，超内存技术）：提高了半虚拟化（Paravirtualization）中虚拟机内存的使用率。这样做可以避免某个空闲（Idle）状态虚拟机占据大量不需要的物理内存，从而避免了可能造成的其它虚拟机上内存短缺现象。

2. 支持虚拟机间的页面共享，这样可以加快虚拟机间的通信。当前采用的实现是写时复制（Copy-On-Write）机制，即共享页面在某个共享者尝试写入时，会首先复制一份该页面内容作为这个共享者的私有页面并重新绑定，这样可以最大限度的减少不必要的内存复制操作。

**可靠性：**

众所周知，每台机器即使在使用寿命中每天内仍有一定的概率发生故障，这种故障称为单点故障。而对于服务器而言，由于发生故障而造成的服务中断会造成大量的金钱损失，有时这种损失甚至会远远超过机器本身的价值。一些知名公司中断服务造成的每小时损失列表如表1所示。为了将这种损失发生频率降到最低，大公司必须未雨绸缪实施多种方案，当今在工业界使用最多的是主从复制（Primary-Backup）方案，通过构建备份的方法来大大降低全系统故障概率。

表1知名公司中断服务每小时造成的损失

|  |  |
| --- | --- |
| 公司/行业 | 每小时损失(美元/每小时) |
| 金融 | 6.45M |
| 信用卡认证 | 2.6M |
| Ebay.com | 225K |
| Amazon.com | 180K |
| 飞机订票业务 | 89K |
| ATM服务 | 14K |

所以Xen4.0突出更好发挥服务器稳定性，主要有以下两点提升：

1. 借鉴了*Remus*，实现了对虚拟机状态的热备份，通过实现Primary-Backup方案来抵御因为硬件损坏造成的虚拟机失效，图5为*Remus*架构图。

2. RAS特性：支持运行时热插拔CPU和内存条。这一特性对于服务器来说很有用，由于大型服务器上拥有数量众多的内存和CPU，因而频繁发生元件损坏。热插拔技术的诞生使得服务器可以不用断电即可完成对故障元件的更换，从而避免了由于该因素造成的大量不可服务时间。Xen4.0对该特性的支持使得Xen可以充分发挥服务器的可靠性，并使得部署在更大规模服务器上成为现实。

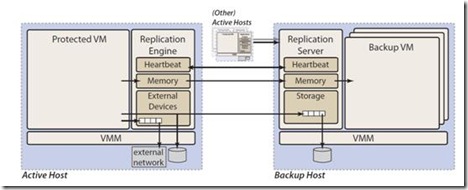
[[](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image010_2.jpg)](http://images.cnblogs.com/cnblogs_com/superymk/WindowsLiveWriter/Xen4.0_A5A7/clip_image010_2.jpg" \t "_blank)

图5. *Remus*架构图[[4]](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html" \l "_ftn4_4540" \t "_blank)

**可用性：**

Xen4.0除了上述提升外，还引入了众多新功能以提升可用性，包括：

1. 更新blktap，引入对VHD格式的支持，并且提升快照和备份的效率。在Xen3.x时代，主要支持RAW格式虚拟磁盘文件。在这里，VHD格式和RAW格式均指虚拟磁盘文件格式，两者都用于在宿主机上以文件格式存储虚拟机整个文件系统，对虚拟机提供硬盘视图。两者的诞生是由于虚拟机软件曾经出现过一个百花齐放的时代，发展到最后这两种格式分别被不同虚拟机软件阵营支持，比如VHD格式虚拟机磁盘文件就常见于Virtual PC和HyperV等微软公司出品的虚拟机软件，而RAW格式就被Xen和QEMU等开源虚拟机一直支持，除这两者外还有VMDK格式虚拟磁盘文件，常用于VMWare虚拟机软件系列。 VHD和RAW格式之间存在差异，这导致两者之间不太容易相互转换，并且在性能上RAW格式虚拟磁盘稍快，原因主要是VHD格式虚拟磁盘文件可视做在RAW格式基础上增加额外管理功能，引入一定开销。因此在Xen4.0中引入对VHD格式虚拟磁盘文件的支持可以说意在蚕食微软阵营的份额。

2. 引入运行时更改虚拟机磁盘大小的功能。这样做可以有效避免宿主机硬盘空间的浪费。

3. 对于并行虚拟化，Xen4.0对Dom0可以支持到pvops-kernel 2.6.31.x,最高到2.6.32.x，同时对已经使用很长时间（大约三年）的linux-2.6.18 Dom0内核仍然提供支持。不过根据笔者的实验，对于自己手动编译安装Xen来说，linux-2.6.18 Dom0内核仍然是最好的选择。

4. 新提供libxl库，用于开发者开发一系列上层控制Xen的工具，从而实现对各虚拟机状态的监控和管理，进而便于开发无人值守虚拟机服务器管理程序组。

5. 对Citrix WHQL认证的Windows并行虚拟化驱动的支持。

这些特性使得Xen不仅可以更好的用于个人电脑，还可以充分发挥服务器的稳定性，可以更好的应用于实际场合中。

**总结**

通过上文介绍我们可以看出，Xen4.0作为最新的虚拟机系统，充分使用了最新虚拟化技术，从可扩展性、性能、可靠性和可用性上均较前代产品有很大提升。在实际应用中，Amazon的EC2云计算平台基础架构就使用Xen构建，由此可以看出Xen并不只是一个各种新技术堆积起来的玩具，而是一个可应用到实际环境中的系统。

当然Xen也存在很多缺点。比如从可用性上来说，VMWare就做得比Xen要好很多，比如鼠标指针漂移问题就没有出现在VMWare的虚拟机中。另一个例子是VMWare提供了强大的VMWare Unity功能，使得虚拟机中的窗口可以整合在宿主机桌面上，并且可以从宿主机直接点击，编缉，拖曳这些窗口。类似的功能并未出现在当前版本的Xen上，当然这个问题受制于Xen的表现层架构设计以及其它非Xen维护的相关项目，不过相信这些问题会在未来某一天解决，从而在Linux越来越盛行的今天给用户带来更好的体验。

**参考文献**

[1]《NewBluePill：深入理解硬件虚拟机》 本作者著，即将出版

[2] *Remus: High Availability via Asynchronous Virtual Machine Replication,* NSDI 2008,

Brendan Cully, Geoffrey Lefebvre, Dutch Meyer, Mike Feeley, Norm Hutchinson, Andrew Warfield

[3] *Xen and the Art of Virtualization*, SOSP2003, Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, Andrew Warfield

[4] *Xen Architecture Overview*,

<http://wiki.xensource.com/xenwiki/XenArchitecture?action=AttachFile&do=get&target=Xen+Architecture_Q1+2008.pdf>

[5] *Virtual Hard Disk Performance*, Microsoft Corp.

http://download.microsoft.com/download/0/7/7/0778C0BB-5281-4390-92CD-EC138A18F2F9/WS08\_R2\_VHD\_Performance\_WhitePaper.docx

[[1]](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html" \l "_ftnref1_4540" \t "_blank) 摘自Installing and configuring Xen on Red Hat Enterprise Linux 5.2 on System x

[[2]](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html" \l "_ftnref2_4540" \t "_blank) 摘自 The Moth’s Blog

[[3]](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html" \l "_ftnref3_4540" \t "_blank) 摘自http://doc.chinaunix.net/server/201002/389306.shtml

[[4]](http://www.cnblogs.com/superymk/archive/2010/07/04/1770799.html" \l "_ftnref4_4540" \t "_blank) 摘自*Remus: High Availability via Asynchronous Virtual Machine Replication*