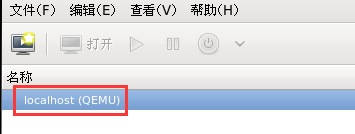
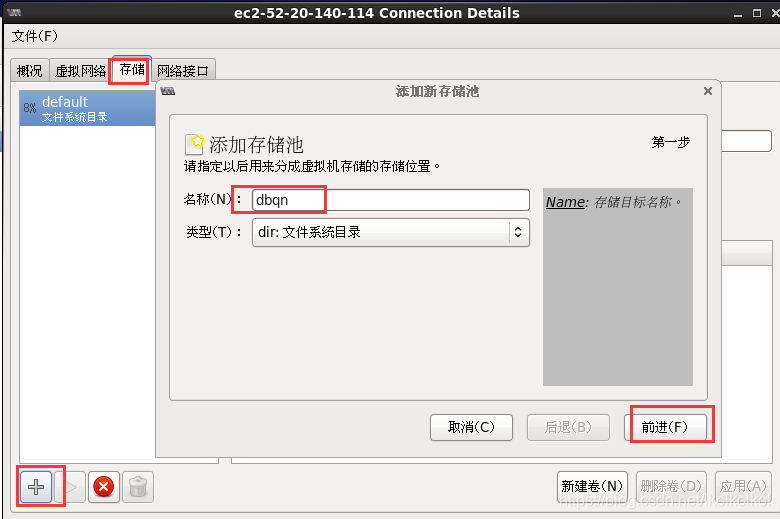
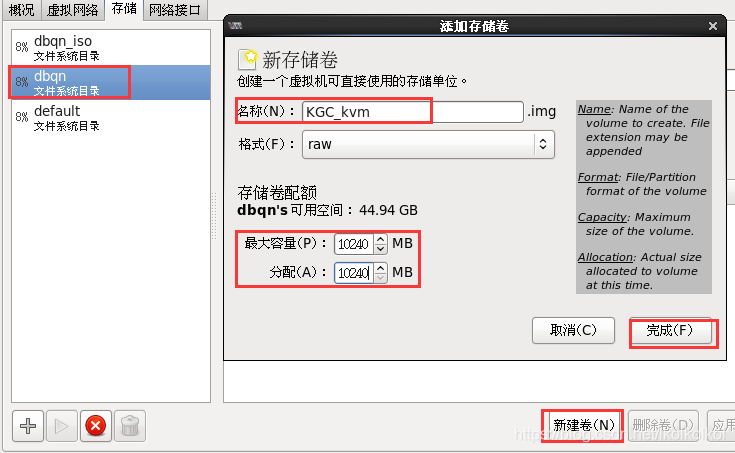
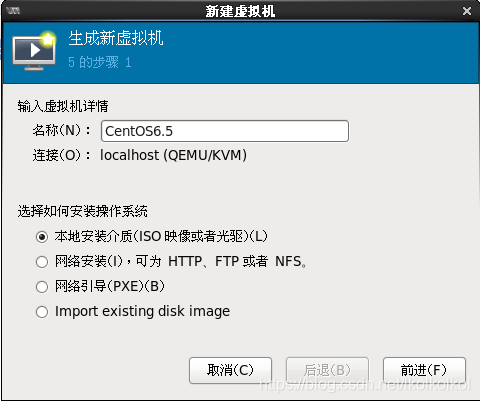
**Docker容器与虚拟化技术——部署KVM虚拟化平台**

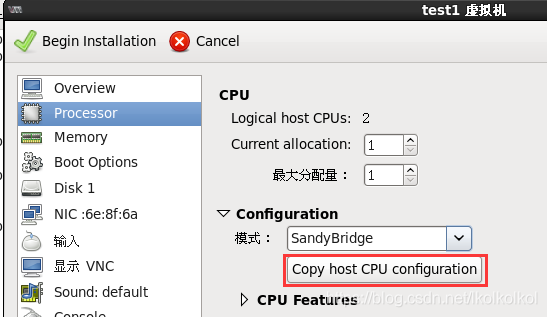
[作者：元芳很忙](https://www.huaweicloud.com/articles/70384406d948fbaa4fc38be49a5a9c68.html) 时间: 2020-04-17 09:26:06

[标签：](https://www.huaweicloud.com/articles/topic-A-1.html)[docker容器与虚拟化技术](http://www.huaweicloud.com/articles/topic_a7615ba575ee714257b5e4b41e134e52.html)[kvm](http://www.huaweicloud.com/articles/topic_41d88848a2ee197eb49ed9b016d796e6.html)

**部署KVM虚拟化平台**

一、搭建KVM虚拟化平台  
1、前置知识点  
(1)什么是虚拟化  
        **把硬件资源从物理方式转变为逻辑方式，打破原有物理结构，使用户可以灵活管理这些资源，并且允许1台物理机上同时运行多个操作系统，以实现资源利用率最大化和灵活管理的技术**  
(2)**虚拟化层**  
①X86平台指令集划分为4个特权模式：ring0-3  
②操作系统工作在ring0  
③应用程序使用ring3  
④驱动程序使用ring1-2  
(3)**虚拟化的优势**  
①减少服务器数量，**降低硬件采购成本**  
②**资源利用率**最大化  
③降低机房空间、散热、电耗成本  
④硬件资源可动态调整，**提高企业IT业务灵活性**  
⑤**高可用性**  
⑥**在不中断服务的情况下进行物理硬件调整**  
⑦降低管理成本  
⑧**具备更高效的灾备能力**  
(4)**VMware虚拟化**  
①vSphere是VMware公司2001年基于云计算推出的一套企业级虚拟化解决方案，核心组件为ESX，现已被ESXi取代。经历了5个版本改进，实现了虚拟化基础架构、高可用性、集中管理、性能监控等一体化解决方案。号称世界第一套云计算的操作系统  
②ESXi本身也是一个操作系统，采用Linux内核（VMKernel），安装方式为裸金属方式，直接安装在物理服务器上，不需要安装任何其他操作系统。  
(5)KVM虚拟化  
①KVM自Linux2.6.20版本后就直接整合到Linux内核中，它依托CPU虚拟化指令集（如Interl-VT、AMD-V）实现高性能的虚拟化支持。由于Linux内核高度整合，因此在性能、安全性、兼容性、稳定性上都有很好的表现  
  
2、案例实施  
(1)安装方式  
①**方式一：安装Linux系统时，选择桌面安装，然后选择虚拟化选项**  
②**方式二：在已有系统上安装KVM所需软件**  
1)yum -y groupinstall “Desktop”：安装桌面环境  
2)yum -y install qemu-kvm：KVM模块  
3)yum -y install qemu-kvm-tools：KVM调试工具，可不安装  
4)yum -y install python-virtinst：python组件，记录创建VM时的xml文件  
5)yum -y install qemu-img：qemu组件，创建磁盘、启动虚拟机等  
6)yum -y install bridge-utils：网络支持工具  
7)yum -y install libvirt：虚拟机管理工具  
8)yum -y install virt-manager：图形界面管理虚拟机  
③验证  
1)**查看CPU是否支持虚拟化：cat /proc/cpuinfo | grep vmx**  
2)**查看KVM模块是否安装：lsmod | grep kvm**  
(2)**设置KVM网络**  
①用户模式（NAT）：默认模式，数据包通过主机接口传送，可访问外网，但外网不能访问虚拟机网络  
②桥接模式：允许虚拟机像一台独立主机那样拥有独立网络  
1)eth0中添加BRIDGE=”br0”  
2)vim /etc/sysconfig/network-scripts/ifcfg-br0  
(3)KVM管理（DEVICE=ager）  
①创建存储池  
1)双击localhost(QEMU)  
  
2)选择“存储”选项框，单击“+”建立名称为bdqn的镜像存储池，并单击前进，之后设置存储目录点击完成即可  
  
3)以同样方式创建名为dbqn\_iso的镜像存储池，目录为/data\_kvm/iso，并将系统安装镜像上传到该目录  
4)在dbqn存储池中建立存储卷  
  
5)返回管理器，右击localhost(QUMU)，选择新建，并按步骤配置新建虚拟机  
  
  
  
6)在Overview中定位到机器设置，把时钟偏移改为localtime，在Boot Options中勾选主机引导时启动虚拟机。如果要远程管理，在显示VNC选项中，将Keymap设置为Copy Local Keymap  
  
7)完成设置后进入装机界面  
  
  
二、使用KVM命令集管理虚拟机  
1、KVM基本功能管理  
(1)查看帮助：virsh -h  
(2)查看KVM配置文件存放目录（CentOS6.5.xml是虚拟机系统实例的配置文件）  
ls /etc/libvirt/qemu/  
(3)查看虚拟机状态：virsh list --all  
(4)虚拟机关机与开机  
virsh shutdown CentOS6.5  
关机命令需要在虚拟机上安装acpid软件并运行acpid服务  
virsh start CentOS6.5  
(5)强制实例系统关闭电源：virsh destroy CentOS6.5  
(6)通过配置文件启动虚拟机实例  
virsh create /etc/libvirt/qemu/CentOS6.5.xml  
(7)挂起虚拟机  
virsh suspend CentOS6.5  
(8)恢复虚拟机  
virsh resume CentOS6.5  
(9)配置虚拟机实例伴随宿主机自动启动  
virsh autostart CentOS6.5  
(10)导出虚拟机配置  
virsh dumpxml CentOS6.5>/etc/libvirt/qemu/test02.xml  
(11)虚拟机的删除与添加  
①删除  
virsh shutdown CentOS6.5  
virsh undefine CentOS6.5  
②通过备份配置文件重新定义虚拟机  
mv test02.xml CentOS6.5.xml  
virsh define CentOS6.5.xml  
(12)修改虚拟机配置信息（内存大小、磁盘等信息）  
①通过vim /etc/libvirt/qemu/CentOS6.5.xml  
②通过virsh edit CentOS6.5  
2、KVM文件管理  
通过文件管理可以直接查看、修改、复制虚拟机的内部文件。例如在虚拟机因配置问题无法启动时，可直接修改虚拟机的文件。虚拟机磁盘文件有raw和qcow2两种，默认使用raw格式。raw性能最好、速度最快，但不支持一些新功能，如镜像、Zlib磁盘压缩、AES加密等。这里介绍本地yum安装libguestfs-tools后产生的命令行工具（针对qcow2格式，因此需要将raw格式文件转换成qcow2格式  
(1)转换raw格式磁盘文件至qcow2格式  
①查看当前磁盘格式：qemu-img info /data\_kvm/store/test01.img  
②关闭虚拟机：virsh shutdown test01  
③转换格式：qemu-img convert -f raw -O qcow2 /data\_kvm/store/test01.img  
(2)修改test01的xml配置文件  
  
(3)virt-cat命令，类似cat命令（libguestfs-tools包中）  
virt-cat -a /data\_kvm/store/test01.qcow2 /etc/resolv.conf  
(4)virt-edit命令：用法基本与vim相同  
(5)virt-df命令：查看虚拟机磁盘信息  
3、虚拟机克隆  
(1)virt-clone -o test1 -n test2 -f /data/store/test2.qcow2  
4、虚拟机快照  
KVM虚拟机要使用镜像功能，磁盘格式必须为qcow2  
(1)创建快照  
virsh snapshot-create test1  
(2)查看快照版本信息  
virsh snapshot-current test1  
(3)查看快照信息  
virsh snapshot-list test1  
(4)恢复快照  
virsh snapshot-revert test1 1584261453  
(5)删除快照  
virsh snapshot-delete test1 1584261453  
5、virt-install工具（需单独安装）  
(1)virt-install是一个命令行工具，它能够为KVM、Xen或其它支持libvrit API的hypervisor创建虚拟机并完成GuestOS安装；此外，它能够基于串行控制台、VNC或SDL支持文本或图形安装界面。安装过程可以使用本地的安装介质如CDROM，也可以通过网络方式如NFS、HTTP或FTP服务实现。对于通过网络安装的方式，virt-install可以自动加载必要的文件以启动安装过程而无须额外提供引导工具。当然，virt-install也支持PXE方式的安装过程，也能够直接使用现有的磁盘映像直接启动安装过程。  
(2)安装示例  
virt-install --name=centos6.5-2 --ram=51 --vcpus=1 --os-type=Linux --cdrom=/home/kvmiso/CentOS6.iso --file=/home/kvmimg/centos65-2.img --file-size=10 --network bridge=br0 --vnc --vncport=5904 --vnclisten=0.0.0.0  
(3)一般选项：  
①-n NAME, --name=NAME：虚拟机名称，需全局惟一；    
②-r MEMORY, --ram=MEMORY：虚拟机内在大小，单位为MB；    
③--vcpus=VCPUS[,maxvcpus=MAX][,sockets=#][,cores=#][,threads=#]：VCPU个数及相关配置；  
④--cpu=CPU：CPU模式及特性，如coreduo等；可以使用qemu-kvm -cpu ?来获取支持的CPU模式；  
(4)安装选项：  
①-c CDROM, --cdrom=CDROM：光盘安装介质；    
②-l LOCATION, --location=LOCATION：安装源URL，支持FTP、HTTP及NFS等，如ftp://172.16.0.1/pub；    
③--pxe：基于PXE完成安装；    
④--livecd: 把光盘当作LiveCD；    
⑤--os-type=DISTRO\_TYPE：操作系统类型，如linux、unix或windows等；    
⑥--os-variant=DISTRO\_VARIANT：某类型操作系统的变体，如rhel5、fedora8等；    
⑦-x EXTRA, --extra-args=EXTRA：根据--location指定的方式安装GuestOS时，用于传递给内核的额外选项，例如指定kickstart文件的位置，--extra-args "ks=http://172.16.0.1/class.cfg"    
⑧--boot=BOOTOPTS：指定安装过程完成后的配置选项，如指定引导设备次序、使用指定的而非安装的kernel/initrd来引导系统启动等 ；例如：  
1)--boot  cdrom,hd,network：指定引导次序；  
2)--boot kernel=KERNEL,initrd=INITRD,kernel\_args=”console=/dev/ttyS0”：指定启动系统的内核及initrd文件  
(5)存储配置  
①--disk=DISKOPTS：指定存储设备及其属性；格式为--disk /some/storage/path,opt1=val1，opt2=val2等；常用的选项有：    
1)device：设备类型，如cdrom、disk或floppy等，默认为disk；    
2)bus：磁盘总结类型，其值可以为ide、scsi、usb、virtio或xen；  
3)perms：访问权限，如rw、ro或sh（共享的可读写），默认为rw；    
4)size：新建磁盘映像的大小，单位为GB；    
5)cache：缓存模型，其值有none、writethrouth（缓存读）及writeback（缓存读写）；    
6)format：磁盘映像格式，如raw、qcow2、vmdk等；    
7)sparse：磁盘映像使用稀疏格式，即不立即分配指定大小的空间；  
②--nodisks：不使用本地磁盘，在LiveCD模式中常用；  
(6)网络配置  
①-w NETWORK, --network=NETWORK,opt1=val1,opt2=val2：将虚拟机连入宿主机的网络中，其中NETWORK可以为：  
1)bridge=BRIDGE：连接至名为“BRIDEG”的桥设备；    
2)network=NAME：连接至名为“NAME”的网络；

三、KVM动态迁移  
1、设置主机名、/etc/hosts，保证挽留过连接  
2、两台主机的KVM连接NFS共享存储  
3、在源主机的KVM中新建虚拟机并安装系统  
4、连接KVM，并进行迁移

四、KVM性能优化  
1、KVM为什么要调优  
KVM采用全虚拟化技术，全虚拟化要由一个软件来模拟硬件层，故由一定的损耗，特别是I/O，因此需要优化。  
KVM性能优化主要在CPU、内存、I/O这几方面。而不同场景优化方向也不同。  
2、KVM优化思路  
(1)CPU优化  
要考虑CPU数量问题，所有guestcpu总数不能超过物理机CPU总数，建议选择复制主机CPU配置  
  
(2)内存优化  
①KSM（Kernel Samepage Merging，相同页合并）  
内存分配最小单位是page，默认大小4kb。可以将host机内容相同的内存合并，以节省内存使用，特别是在虚拟机操作系统都一样的情况下。开启了KSM，则会将这些内存合并为一个，当然这个过程会有性能损耗，所以开区与否，需要考虑使用场景  
通过/sys/kernel/mm/ksm目录下的文件可查看内存页共享情况  
pages\_shared记录了KSM共享的总页面数  
pages\_sharing记录了当前共享的页面数  
run记录是否开启KSM（0为不开启，1为开启）  
临时开启：echo 1>/sys/kernel/mm/ksm/run（只能用重定向，不支持VIM）  
pages\_to\_scan记录数值决定每次扫描相同页时每次查看多少个页面，超过2000无效，需要开启ksmtuned和tuned服务  
②对内存设置限制  
防止某个虚拟机无节制的使用内存资源，导致其他虚拟机无法正常使用，就需要对内存使用进行限制  
查看当前虚拟机内存使用限制：virsh memtune vm  
1)--min-guarantee：保证最小内存  
2)--config：下次重启生效  
3)--live：在线生效  
4)--current：只在当前生效  
5)virsh memtune vm --hardlimit 1024000 --live：最大内存100M，在线生效  
③大页后端内存（Huge Page Backed Memory）  
在逻辑地址向物理地址转换时，CPU保持一个翻译后备缓冲器TLB，用来缓存转换结果，而TLB容量很小，所以如果page很小，TLB很容易就充满，这样就容易导致cache miss，相反page很大，TLB需要保存的缓存项就变少，就会减少cache miss。通过为客户端提供大页后端内存，就能减少客户机消耗的内存并提高TLB命中率，从而提升KVM性能  
Interl的x86CPU通常使用4kb内存页，通过配置能够使用大页：x86\_32是4mb，64和32PAE是2mb。  
1)echo 25000>/proc/sys/vm/nr\_hugepages：指定大页需要的内存页面数  
2)cat /proc/meminfo | grep HugePage：查看大页使用情况  
3)也可在/etc/sysctl.conf中添加vm.nr\_hugepages=2500来持久设定大页文件需要的内存页面数  
4)需要的页面数可由客户机需要的内存厨业页面大小来大体估算  
5)虚拟机大页使用  
a.virsh destroy vm：关闭虚拟机VM  
b.virsh edit vm  
a)  
c.mount -t hugetlbfshugetlbfs /dev/higepages：挂载hugetlfs文件系统  
d.service libvirtd restart  
e.virsh start vm  
f.开机自动挂载：/etc/fstab中hugetlbfs/hugepages hugetlbfs defaults 0 0  
(3)I/O优化  
实际生产环境中，为了避免过度消耗磁盘资源而对其他的虚拟机造成影响，我们希望每台虚拟机对磁盘资源的消耗是可以控制的。比如多个虚拟机往硬盘中写数据，谁可以优先写，就可以调整I/O权重weight，权重越高写入优先级越高  
①在整体中的权重，范围在100-1000  
②限制具体的I/O  
vish blkiotune vm：查看权重  
virsh blkiotune vm --weight 500：设置权重为500  
编辑虚拟机XML文件：500  
③还可以使用blkdeviotune设置虚拟机读写速度  
virsh blkdeviotune --  
(4)系统调优工具tuned/tuned-adm  
①tuned：服务端程序，用于监控和手机系统组件数据  
②tuned-adm：客户端程序，用来和tuned打交道。可使用预先配置的优化方案对系统进行调优  
③安装和启动  
yum -y install tuned  
service tuned start  
service ktune start  
④查看当前优化方案：tuned-adm active  
查看预先设定好的优化方案：tuned-adm list  
    - server-powersave  
- desktop-powersave  
- default：默认节电配置，是最基本的节点配置，只启用磁盘和CPU插件  
- virtual-guest：企业级服务器配置中使用，其中包括电池备份控制程序，缓存        保护以及磁盘管理  
- latency-performance：延迟性能调优的服务器配置  
- throughput-performance：吞吐性能调整的服务器配置，如果系统没有企业级        存储，建议使用这个  
- enterprise-storage：企业存储服务器优化方案  
- virtual-host：根据enterprise-storage配置，可减少可置换的虚拟内存，并启        用更多集合脏页写回。同时推荐在虚拟化主机中使用这个配置，包括KVM        和红帽企业版虚拟化主机  
- spindown-disk  
- laptop-ac-powersave  
- laptop-battery-powersave  
Current active profile: default：使用的某种配置  
⑤使用优化方案：tuned-adm profile virtual-gest（修改当前方案为virtual-gest）  
⑥创建自定义优化方案  
1)拷贝/etc/tune-profiles/中的预定义方案后，进行参数修改即可  
2)查看方案：tuned-adm list