

- Guest: 客户机系统,包括CPU(vCPU)、内存、驱动(Console、网卡、I/O 设备驱动等),被 KVM 置于一种受限制的 CPU模式下运行。
- KVM:运行在内核空间,提供CPU 和内存的虚级化,以及客户机的 I/O 拦截。Guest 的 I/O 被 KVM 拦截后,交给 QEMU 处理。
- QEMU:修改过的为 KVM 虚机使用的 QEMU 代码,运行在用户空间,提供硬件 I/O 虚拟化,通过 IOCTL /dev/kvm 设备和 KVM 交互。

KVM 是实现拦截虚机的 I/O 请求的原理:

现代 CPU 本身了对特殊指令的截获和重定向的硬件支持,甚至新的硬件会提供额外的资源来帮助软件实现对关键硬件资源的虚拟化从而提高性能。以 X86 平台为例,支持虚拟化技术的 CPU 带有特别优化过的指令集来控制虚拟化过程。通过这些指令集,VMM 很容易将客户机置于一种受限制的模式下运行,一旦客户机视图访问物理资源,硬件会暂停客户机的运行,将控制权交回给 VMM 处理。VMM 还可以利用硬件的虚级化增强机制,将客户机在受限模式下对一些特定资源的访问,完全由硬件重定向到 VMM 指定的虚拟资源,整个过程不需要暂停客户机的运行和 VMM 的参与。由于虚拟化硬件提供全新的架构,支持操作系统直接在上面运行,无需进行二进制转换,减少了相关的性能开销,极大简化了VMM的设计,使得VMM性能更加强大。从 2005 年开始,Intel 在其处理器产品线中推广 Intel Virtualization Technology 即 IntelVT 技术。

QEMU-KVM:

其实 QEMU 原本不是 KVM 的一部分,它自己就是一个纯软件实现的虚拟化系统,所以其性能低下。但是,QEMU 代码中包含整套的虚拟机实现,包括处理器虚拟化,内存虚拟化,以及 KVM需要使用到的虚拟设备模拟(网卡、显卡、存储控制器和硬盘等)。 为了简化代码,KVM 在 QEMU 的基础上做了修改。VM 运行期间,QEMU 会通过 KVM 模块提供的系统调用进入内核,由 KVM 负责将虚拟机置于处理的特殊模式运行。遇到虚机进行 I/O 操作,KVM 会从上次的系统调用出口处返回 QEMU,由 QEMU 来负责解析和模拟这些设备。 从 QEMU 的角度看,也可以说是 QEMU 使用了 KVM 模块的虚拟化功能,为自己的虚机提供了硬件虚拟化加速。除此以外,虚机的配置和创建、虚机运行说依赖的虚拟设备、虚机运行时的用户环境和交互,以及一些虚机的特定技术比如动态迁移,都是 QEMU 自己实现的。

KVM

KVM 内核模块在运行时按需加载进入内核空间运行。KVM 本身不执行任何设备模拟,需要 QEMU 通过 /dev/kvm 接口设置一个 GUEST OS 的地址空间,向它提供模拟的 I/O 设备,并将它的视频显示映射回宿主机的显示屏。它是KVM 虚机的核心部分,其主要功能是初始化 CPU 硬件,打开虚拟化模式,然后将虚拟客户机运行在虚拟机模式下,并对虚机的运行提供一定的支持。以在 Intel 上运行为例,KVM 模块被加载的时候,它:

- 1. 首先初始化内部的数据结构;
- 2. 做好准备后,KVM 模块检测当前的 CPU,然后打开 CPU 控制及存取 CR4 的虚拟化模式开关,并通过执行 VMXON 指令将宿主操作系统置于虚拟化模式的根模式:
- 3. 最后, KVM 模块创建特殊设备文件/dev/kvm 并等待来自用户空间的指令。

接下来的虚机的创建和运行将是 QEMU 和 KVM 相互配合的过程。两者的通信接口主要是一系列针对特殊设备文件 dev/kvm 的 IOCTL 调用。其中最重要的是创建虚机。它可以理解成KVM 为了某个特定的虚机创建对应的内核数据结构,同时,KVM 返回一个文件句柄来代表所创建的虚机。

针对该句柄的调用可以对虚机做相应地管理,比如创建用户空间虚拟地址和客户机物理地址、真实物理地址之间的映射关系,再比如创建多个 vCPU。KVM 为每一个 vCPU 生成对应的文件句柄,对其相应地 IOCTL 调用,就可以对vCPU进行管理。其中最重要的就是"执行虚拟处理器"。通过它,虚机在 KVM 的支持下,被置于虚拟化模式的非根模式下,开始执行二进制指令。在非根模式下,所有敏感的二进制指令都被CPU捕捉到,CPU 在保存现场之后自动切换到根模式,由 KVM 决定如何处理。

除了 CPU 的虚拟化,内存虚拟化也由 KVM 实现。实际上,内存虚拟化往往是一个虚机实现中最复杂的部分。CPU 中的 内存管理单元 MMU 是通过页表的形式将程序运行的虚拟地址转换成实际物理地址。在虚拟机模式下,MMU 的页表则必须在一次查询的时候完成两次地址转换。因为除了将客户机程序的虚拟地址转换了客户机的物理地址外,还要将客户机物理地址转化成真实物理地址。

2. KVM 的功能列表

KVM 所支持的功能包括:

• 支持CPU 和 memory 超分 (Overcommit)

1. Re:Neutron 理解 (1): Neutron 所 实现的虚拟化网络 [How Netruon Virtualizes Network]

eth1 - 公共网络 (untagged),管 理网络(tag=102),存储网络 (tag=103)不好意思,大家共用同 ·个eth1端口的时候,请问这里交 换机端口是配置为tagged还是 untagged.....

--xianke9

2. Re:理解Docker (5): Docker 网络

1.12版本上网络的表现如何?

--幽灵狼

3. Re:理解Docker (5): Docker 网络

我想请问一下运行docker quickstart terminal时一直卡在"waiting for an IP"应该如何解决呢?希望楼主能解 答一下。

--silentbell

4. Re:理解Docker (6): 若干企业 生产环境中的容器网络方案

写得好!加油。

--itbi00

5. Re:理解Docker(5): Docker 网络

非常好,写得很详细。加油!

--itbi00

阅读排行榜



- 1. Neutron 理解 (1): Neutron 所实 现的虚拟化网络 [How Netruon Virtualizes Network1(22087)
- 2. 理解 OpenStack 高可用 (HA) (1): OpenStack 高可用和灾备方
- 案 [OpenStack HA and DR](13707)
- 3. Neutron 理解 (3): Open vSwitch
- + GRE/VxLAN 组网 [Netruon Open
- vSwitch + GRE/VxLAN Virutal Network](13434)
- 4. 探索 OpenStack 之(9): 深入 块存储服务Cinder (功能篇)
- 5. 理解 OpenStack + Ceph (1): Ceph + OpenStack 集群部署和配置 (12444)

评论排行榜



- 1. Neutron 理解 (1): Neutron 所实 现的虚拟化网络 [How Netruon Virtualizes Networkl(63)
- 2. Neutron 理解 (14): Neutron ML2 + Linux bridge + VxLAN 组网 (54)
- 3. Neutron 理解 (8): Neutron 是如 何实现虚机防火墙的 [How Neutron Implements Security Group](34)
- 4. Neutron 理解 (3): Open vSwitch
- + GRE/VxLAN 组网 [Netruon Open vSwitch + GRE/VxLAN Virutal Network](25)
- 5. Neutron 理解 (5): Neutron 是 如何向 Nova 虚机分配固定IP地址
- 的 (How Neutron Allocates Fixed IPs to Nova Instance) (21)

推荐排行榜



- 1. Neutron 理解 (1): Neutron 所实 现的虚拟化网络 [How Netruon Virtualizes Network](9)
- 2. 我所了解的 京东、携程、eBay、 小米的 OpenStack 云(6)
- 3. 理解 OpenStack 高可用 (HA)
- (1): OpenStack 高可用和灾备方 案 [OpenStack HA and DR](6)
- 4. Neutron 理解 (2): 使用 Open vSwitch + VLAN 组网 [Netruon Open vSwitch + VLAN Virutal
- Network](6)

- 支持半虚拟化I/O (virtio)
- 支持热插拔 (cpu , 块设备、网络设备等)
- 支持对称多处理 (Symmetric Multi-Processing, 缩写为 SMP)
- 支持实时迁移 (Live Migration)
- 支持 PCI 设备直接分配和 单根I/O 虚拟化 (SR-IOV)
- 支持 内核同页合并 (KSM)
- 支持 NUMA (Non-Uniform Memory Access,非一致存储访问结构)

3. KVM 工具集合

- libvirt:操作和管理KVM虚机的虚拟化 API,使用 C 语言编写,可以由 Python,Ruby, Perl, PHP, Java 等语言调用。可以操作包 括 KVM, vmware, XEN, Hyper-v, LXC等 Hypervisor。
- Virsh: 基于 libvirt 的 命令行工具 (CLI)
- Virt-Manager: 基于 libvirt 的 GUI 工具
- virt-v2v:虚机格式迁移工具
- virt-* 工具:包括 Virt-install (创建KVM虚机的命令行工具), Virt-viewer (连接到虚机屏幕的工具),Virt-clone(虚机克隆工 具) . virt-top 等
- sVirt:安全工具

4. RedHat Linux KVM 安装

RedHat 有两款产品提供 KVM 虚拟化:

- 1. Red Hat Enterprise Linux:适用于小的环境,提供数目较少的KVM虚机。最新的版本包括 6.5 和 7.0.
- 2. Red Hat Enterprise Virtualization (RHEV):提供企业规模的KVM虚拟化环境,包括更简单的管理、HA,性能优化和其它 高级功能。最新的版本是 3.0.

RedHat Linux KVM:

- KVM 由 libvirt API 和基于该 API的一组工具进行管理和控制。
- KVM 支持系统资源超分,包括内存和CPU的超分。RedHat Linux 最多支持物理 CPU 内核总数的10倍数目的虚拟CPU,但是不 支持在一个虚机上分配超过物理CPU内核总数的虚拟CPU。
- 支持 KSM (Kenerl Same-page Merging 内核同页合并)

RedHat Linux KVM 有如下两种安装方式:

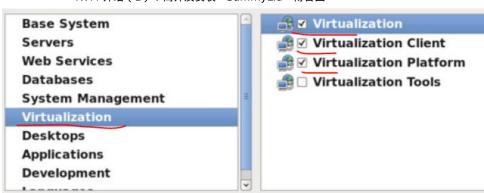
4.1 在安装 RedHat Linux 时安装 KVM

选择安装类型为 Virtualization Host :

	se Server							
○ Web Se								
	Name and Address of the Owner, where the Owner, which is the Owner, where the Owner, which is the Owner, where the Owner, which is th	nent Server						
Virtual								
O Deskto								
		ment Works	station					
O Minima	ii.							
Please selo	ct any add	fitional repos	sitories that yo	u want	to use f	or softw	are inst	allation.
	ect any ado	litional repos	sitories that yo	u want	to use f	or softw	are inst	allation.
☐ High A		litional repos	sitories that yo	u want	to use f	or softw	are inst	allation.
☐ High A	vailability alancer		sitories that yo	u want	to use f	or softw	are inst	allation.
☐ High Av	vailability alancer		sitories that yo	u want	to use f	or softw	are inst	allation.
☐ High A	vailability alancer t Enterpris		V		to use f			allation.

可以选择具体的 KVM 客户端、平台和工具:

5. 理解 OpenStack 高可用 (HA) (2):Neutron L3 Agent HA 之 虚 拟路由冗余协议 (VRRP) (5)



4.2 在已有的 RedHat Linux 中安装 KVM

这种安装方式要求该系统已经被注册,否则会报错:

[root@rh65 ~]# yum install qemu-kvm qemu-img
Loaded plugins: product-id, refresh-packagekit, security, subscription-manager
This system is not registered to Red Hat Subscription Management. You can use subscription-manager to register.
Setting up Install Process
Nothing to do

你至少需要安装 qemu-kvm qemu-img 这两个包。

yum install gemu-kvm gemu-img

你还可以安装其它工具包:

yum install virt-manager libvirt libvirt-python python-virtinst libvirt-client

4.3 QEMU/KVM 代码下载编译安装

4.3.1 QEMU/KVM 的代码结构

QEMU/KVM 的代码包括几个部分:

(1) KVM 内核模块是 Linux 内核的一部分。通常 Linux 比较新的发行版(2.6.20+)都包含了 KVM 内核,也可以从这里得到。比如在我的RedHat 6.5 上:

[root@rh65 isoimages]# uname -r 2.6.32-431.el6.x86_64 [root@rh65 isoimages]# modprobe -l | grep kvm kernel/arch/x86/kvm/kvm.ko kernel/arch/x86/kvm/kvm-intel.ko kernel/arch/x86/kvm/kvm-amd.ko

- (2)用户空间的工具即 qemu-kvm。qemu-kvm 是 KVM 项目从 QEMU 新拉出的一个分支(看这篇文章)。在 QEMU 1.3版本之前,QEMU 和 QEMU-KVM 是有区别的,但是从 2012 年底 GA 的 QEMU 1.3版本开始,两者就完全一样了。
- (3) Linux Guest OS virtio 驱动,也是较新的Linux 内核的一部分了。
- (4) Windows Guest OS virtio 驱动,可以从这里下载。

4.3.2 安装 QEMU

RedHat 6.5 上自带的 QEMU 太老, 0.12.0 版本, 最新版本都到了 2.* 了。

(1). 参考 这篇文章,将 RedHat 6.5 的 ISO 文件当作本地源

mount -o loop soft/rhel-server-6.4-x86_64-dvd.iso /mnt/rhel6/

vim /etc/fstab

=> /root/isoimages/soft/RHEL6.5-20131111.0-Server-x86_64-DVD1.iso /mnt/rhel6 iso9660 ro,loop

[root@rh65 qemu-2.3.0]# cat /etc/yum.repos.d/local.repo

[local]

name=local

baseurl=file:///mnt/rhel6/

enabled=1

gpgcjeck=0

1 yum clean all
yum update



(2). 安装依赖包包

yum install gcc yum install autoconf yum install -y glib*
yum install zlib*

(3).从 http://wiki.qemu.org/Download 下载代码,上传到我的编译环境 RedHat 6.5.

tar -jzvf qemu-2.3.0.tar.bz2
cd qemu-2.3.0
./configure
make -j 4

make install (4). 安装完成

[root@rh65 qemu-2.3.0]# /usr/local/bin/qemu-x86_64 -version qemu-x86_64 version 2.3.0, Copyright (c) 2003-2008 Fabrice Bellard

(5). 为方便起见,创建一个link

ln -s /usr/bin/qemu-system-x86_64 /usr/bin/qemu-kvm

4.3.3 安装 libvirt

可以从 libvirt 官网下载安装包。最新的版本是 0.10.2.

5. 创建 KVM 虚机的几种方式

yum install autoconf automake libtool

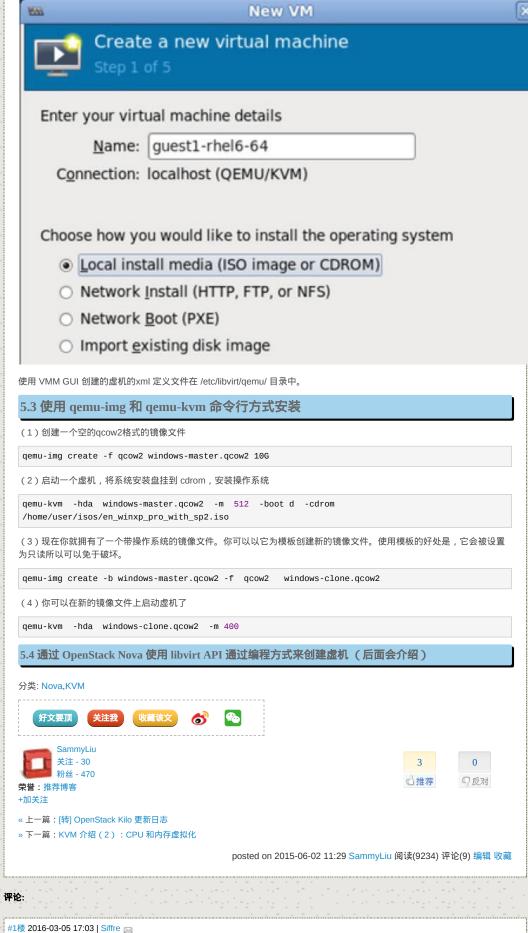
5.1 使用 virt-install 命令



- virt-install \
 --name=guest1-rhel5-64 \
- --file=/var/lib/libvirt/images/guest1-rhel5-64.dsk \
- --file-size=8 \
- --nonsparse --graphics spice \
- --vcpus=2 --ram=2048 \
- --location=http://example1.com/installation_tree/RHEL5.6-Serverx86_64/os \
- --network bridge=br0 \
- --os-type=linux \
- --os-variant=rhel5.4

5.2 使用 virt-manager 工具

Virtual Machine Manager File Edit View Help Open D D V



#1楼 2016-03-05 17:03 | Siffre ☑ 大神你好,小弟一直在跟着你的步伐前进,请多多指点,这是小弟的博客,有时间来多多指点,小弟感激不尽! 博客地址:lanzhishi.org 支持(0) 反对(0)

#2楼[楼主] 2016-03-05 17:51 | SammyLiu 🖂

@ Siffre

实在不敢当大神,就是喜欢随手记点东西,希望在帮助自己的同时对别人也有些帮助或者启发。我看了你的博客,版面和内容都挺不错,加

KVM 介绍 (1): 简介及安装 - SammyLiu - 博客园 油,贵在坚持! 支持(0) 反对(0) #3楼 2016-03-05 17:52 | Siffre 🖂 @ SammyLiu 实时在线a,加个微信吧,ctym361 支持(0) 反对(0) #4楼[楼主] 2016-03-05 17:53 | SammyLiu 🖂 @ Siffre 已加 支持(0) 反对(0) #5楼 2016-08-18 15:47 | cloudcloud 区 大大讲的太好啦 支持(0) 反对(0) #6楼 2017-02-20 13:42 | edward_lii 🖂 你的文章写得很好,但我有些疑问,能告知一下你的联系方式吗? 支持(0) 反对(0) #7楼[楼主] 2017-02-27 09:05 | SammyLiu 🖂 @ edward_lii 见站内信 支持(0) 反对(0) #8楼 2017-03-27 09:42 | 不见天高 🖂 楼主你好,可以加个微信请教你不,z157984031 支持(0) 反对(0) #9楼 2017-04-05 10:35 | 大相林 🖂 挺好,知识面非常全面,很深入。 支持(0) 反对(0) 刷新评论 刷新页面 返回顶部 🔜 注册用户登录后才能发表评论,请 登录 或 注册, 访问网站首页。

【推荐】50万行VC++源码: 大型组态工控、电力仿真CAD与GIS源码库

【报表】Excel 报表开发18 招式,人人都能做报表

【活动】阿里云海外云服务全面降价助力企业全球布局

【实用】40+篇云服务器操作及运维基础知识!



最新IT新闻:

- 知乎上线视频功能,以后看教程更方便了
- 一年只赚2万元:乐视游戏或被出售
- Unity获得4亿美元投资,现估值为26亿美元
- 直播对陌陌的意义,就像王者荣耀之于腾讯游戏
- · 死磕支付宝? 苏宁金融发布"星辰计划": 扫码支付返888元
- » 更多新闻...



Powered by: 博客园 模板提供:沪江博客 Copyright ©2017 SammyLiu