Cap. 6 Hardware Virtual

TECNOLOGIAS DE VIRTUALIZAÇÃO ENGENHARIA INFORMÁTICA WWW.IPLEIRIA.PT

Cap. 6 – Hardware virtual

- CPU virtual
- Memória virtual
- Storage virtual
- Rede virtual

CPU virtual

- vCPU: processador virtual
- pCPU: processador físico (physical CPU)

CPU virtual - Oversubscription

- Oversubscription: quando se alocam mais vCPUs que a quantidade de pCPUs disponíveis (cores, Hyper-threading)
- Isto pode tornar mais lentos os vCPUs

CPU virtual - Oversubscription

- A Oversubscription pode ser uma boa solução de curto prazo
- Deve-se tentar ter um número total de vCPUs aproximadamente igual ou inferior ao número de pCPUs.

Alguns hypervisores permitem atribuir prioridades diferentes aos vCPUs.

CPU hot-plugging

- Num computador físico, suportar CPU hotplugging significaria que um CPU poderia ser encaixado ou removido com a máquina ligada
- Com a virtualização, os SO de servidores modernos passaram a suportar esta possibilidade



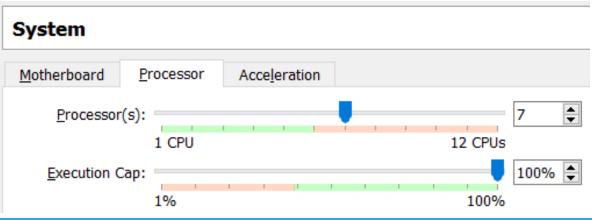
- O VirtualBox suporta CPU hot-plugging para os SO guest Linux e Windows
- Com guest Windows suporta apenas hot-addd
- Com guest Linux suporta hot-add e hot-remove

- Para utilizar mais que um vCPU na VM é necessário suporte de virtualização em hardware no host
- Uma VM Linux de 32 bits suporta no máximo 8 vCPUs

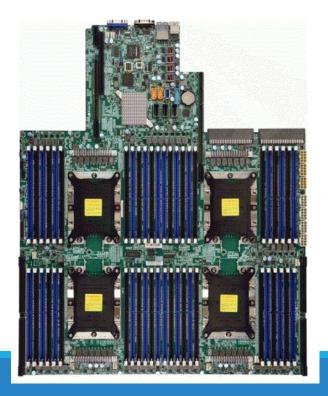
- O nº de vCPU que o VirtualBox permite atribuir à VM corresponde ao dobro de cores no CPU do host (cores reais, sem hyperthreads), até um máximo de 32
- No entanto, utilizar mais vCPU que o nº de cores do host vai degradar o desempenho da VM

 Execution cap: limita a % de tempo que um core do host passa a emular um vCPU (100% por omissão)

hexa core CPU host:



- Para utilizar o CPU hot-plugging é necessário primeiro ativar a funcionalidade
 - \$ VBoxManage modifyvm VM-name --cpuhotplug on
- Especificar o máximo de vCPU que a VM pode ter:
 - \$ VBoxManage modifyvm VM-name --cpus 8



- O vCPU 0 nunca pode ser removido
- Com a VM desligada, pode adicionar ou remover um determinado vCPU:
 - \$ VBoxManage modifyvm VM-name --plugcpu 3
 - \$ VBoxManage modifyvm VM-name --unplugcpu 3
- Com a VM ligada, pode adicionar ou remover um determinado vCPU:
 - \$ VBoxManage controlvm VM-name --plugcpu 3
 - \$ VBoxManage controlvm VM-name --unplugcpu 3

- Num guest Linux:
 - Para evitar ejetar um vCPU em utilização, o mesmo tem que ser primeiro ejetado dentro do guest
 - As guest additions do Linux captam evetos host-remove e ejetam o vCPU
 - Da mesma forma, um vCPU adicionado à VM não é utilizado automaticamente pelo Linux. As guest additions tratam depois disso.
 - Se as guest additions não estão instaladas pode ativar-se depois o vCPU manualmente:
 - \$ echo 1 > /sys/devices/system/cpu/cpu<id>/online

Cap. 6 – Hardware virtual

- CPU virtual
- Memória virtual
- Storage virtual
- Rede virtual

Memória virtual

- Memória nas VMs
- O hypervisor fornece acesso à memória a todos os guests
- Cada VM recebe uma fatia de memória física no host

Memória virtual

- Todos os hypervisors permitem definir um valor fixo de memória para a VM
- Muitos hypervisors permitem definir uma gama entre um valor mínimo e um valor máximo.
 - Quando o guest começa a ficar sem memória livre, o hypervisor aloca mais memória até ao limite máximo definido

 Para reduzir a quantidade de memória de uma VM, é normalmente necessário desligar primeiro a VM

- O hypervisor poderá também recuperar memória dos guests quando o host está a ficar sem memória.
- Com as guest additions é possível fazer memory ballooning, permitindo que a memória que tinha sido alocada a uma VM que está em execução, possa ser atribuída a outra VM
- Ballooning processo existente nos hypervisores para evitar a perda de dados
- Um pequeno programa é inserido na memória do guest
- Esse programa cresce então na memória, como um balão que se enche

- Toda a memória usada pelo balão pode ser recuperada pelo hypervisor
- O hypervisor tem a certeza que a memória usada pelo balão não contém dados importantes.

- A longo prazo deve reduzir-se a quantidade de ballooning numa máquina
- Este processo tem sempre algum overhead associado e pode tornar o hypervisor mais lento
- O melhor é adicionar memória física ao host, para reduzir a quantidade de ballooning necessária

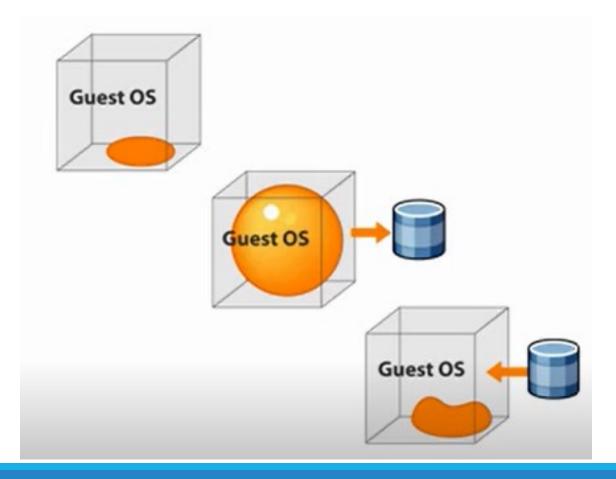
- No VirtualBox o balloning só é suportado por linha de comando (<n> MB):
 - VBoxManage controlvm "VM name" guestmemoryballoon <n>
- Utilizar modifyvm em vez de controlvm se se pretender a utilização do balão de forma permanente

Memória virtual – Ballooning - VMware

- Quando a memória no host começa a escassear
- Mecanismo que reclama memória do Guest OS
- Requer as VMware Tools
- A VM corre o processo vmmemctl (VM memory control)
- O VBoxWatchdog é o serviço que gere o ballooning configurado nas VM

Memória virtual – Ballooning - VMware

- O vmmemctl envia páginas antigas da VM para disco e enche o balão para ocupar esse espaço, não permitindo que a VM o reutilize
- Essas páginas vão para uma área de memória partilhada por todas as VMs



Libertar RAM - VirtualBox

- VM ligada com 1 GB de RAM
- > VBoxManage controlvm "Windows10" guestmemoryballoon 700
- > VBoxManage controlvm "Windows10" guestmemoryballoon 0

Gestor de tarefas do Guest:

Memory usage

1.0 GB

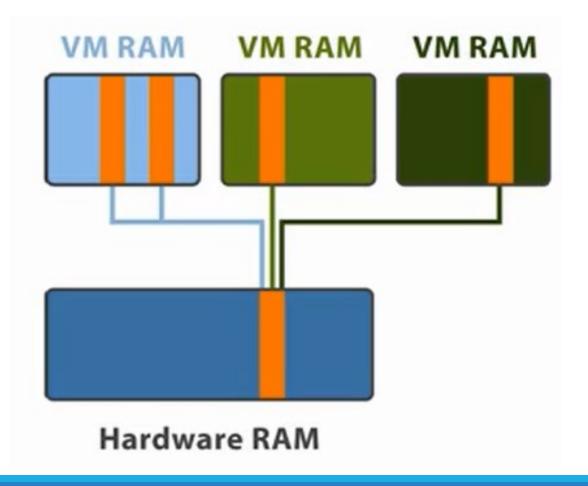
60 seconds

Libertar RAM - VirtualBox

- Duas VM Windows com 3 GB de RAM cada
- Iniciar VM1 (consome 3 GB do host)
- Fazer ballooning de 2000 MB nesta VM (disponíveis para próximas VMs)
 - Consumo no host mantém-se em 3 GB
- Iniciar depois a VM2
 - As VMs consomem apenas 4 GB de RAM no host (1 GB + 2 GB + 1 GB)
 - A VM2 utiliza 2 GB do balão da VM1
- Repor ballooning de 0 MB na VM1
 - A VM1 vai alocar mais 2 GB (consumo total de 6 GB no host)

Memória – Transparent Page Sharing

- A tecnologia TPS permite que várias VMs partilhem a mesma página de memória.
- Eliminam-se assim cópias redundantes de páginas (normalmente 4 KB)
- Isto permite usar mais memória que aquela que realmente se tem.
- O TPS funciona melhor quando existem várias VMs similares no mesmo host (com o mesmo SO)



Memória – Transparent Page Sharing

- Normalmente o host só recorre a esta técnica em situação de overcommitment da memória
- Sistemas de virtualização de memória assistidos por hardware:
 Intel EPT e AMD RVI
- No VirtualBox a partilha de páginas chama-se Page Fusion. Só suporta guests Windows e não suporta hosts Mac OS X.

Memória – Partilha de páginas

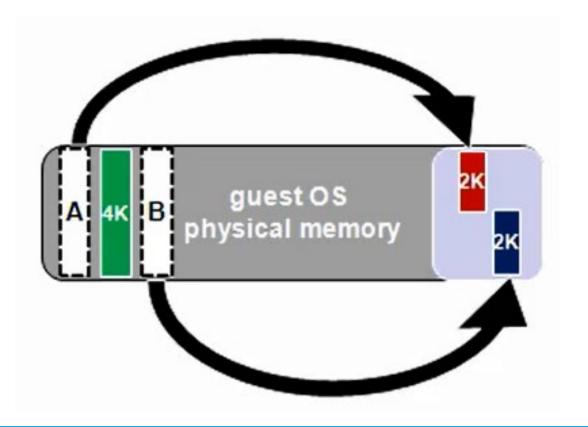
- Na técnica tradicional de partilha de páginas, o hypervisor calcula checksums (hashes) para todas as páginas da memória
- Só as páginas com o mesmo hash poderão ser iguais. Este é um processo demorado e com overhead considerável de utilização de CPU
- E quando uma página partilhada é escrita por uma VM, é preciso realocar uma nova cópia da página (mais overhead)

Memória – Partilha de páginas

- Novas técnicas utilizam as tools ou additions para identificar páginas mais prováveis de serem iguais entre VMs
- Pode assim poupar-se memória quase de forma imediata e com muito baixo overhead
- O VirtualBox utiliza esta técnica. Comando para ativar o Page Fusion numa VM: VBoxManage modifyvm "VM name" --pagefusion on
- Para acompanhar quantas páginas estão partilhadas no host ou numa VM: VBoxManage metrics query "VM name" RAM/VMM/Shared, Guest/RAM/Usage/Shared

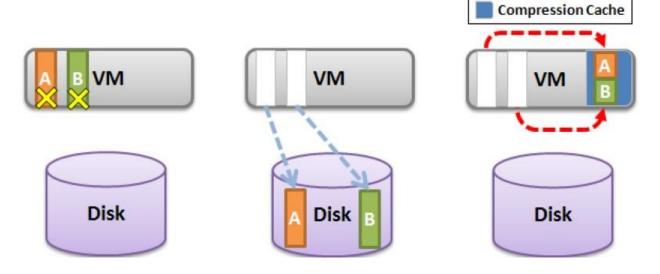
Memória virtual – Compressão - VMware

- Quando a memória no host começa a escassear
- Algumas páginas que podem ser comprimidas a 50% ou mais são guardadas numa cache em memória vez de irem para swap



Memória virtual – swap

- Depois da compressão, se ainda for precisa mais memória, recorre-se ao swap
- Mas uso excessivo afeta o desempenho



Alterar RAM e CPU

	RAM			Alterar CPUs	
	Reduzir		Aumentar	Quantidade	Alterar Prioridade
	Settings	Ballooning	Admental	Quantidade	Alteral Prioridade
Vmware Workstation (Player + Pro)	VM off		VM On!!	VM off	
Virtual box	VM off	Utilizar ballooning para ceder RAM a outras VMs Virtual Box. O SO da VM continua a ver toda a sua RAM, mas não consegue utilizar o balão, que pode então ser utilizado por outras VMs VirtualBox	VM off		Com VM On!!, Alterar Excecution Cap %
				VM On!!	
				Windows só hot-add	
Hyper-V	VM On!! Ballooning automático.		VM On!!	VM-Off	Com VM On!!, Alterar % Reserva, % Limite e peso relativo
VMware ESXi	Reduzir: VM off	Automático	VM On!! se opção	Com VM On!!, Alterar	Com VM On!!, Alterar
	VM On!!, Alterar		"Memory Hot Plug"	nº de CPUs (cores per	Reservation Hz, Limit Hz e
	Reservation MB, Limit MB e Shares		ativa nas Settings	socket: VM off).	Shares
Citrix XenServer	VM off	Automático com o Dynamic Memory Control (DMC)	VM off	VM-Off	VM-Off

Cap. 6 – Hardware virtual

- CPU virtual
- Memória virtual
- Storage virtual
- Rede virtual

- Storage virtual:
 - Tipos de ficheiros usados na virtualização
 - Sistemas de ficheiros dos discos físicos nos hosts

- O hypervisor redireciona os acessos do guest a disco para o ficheiro imagem (image file, disco virtual)
- Embora a capacidade do disco virtual seja especificada na criação do disco, esta pode ser expandida posteriormente, mesmo que já tenha dados.
- Duas opções para criar um disco virtual:
 - fixed-size: demora mais tempo a criar
 - dynamically allocated: começa muito pequeno e vai crescendo até ao limite indicado, cada vez que um setor é escrito pela primeira vez; as operações de escrita que levam ao crescimento do disco são mais lentas

- Tipos de ficheiros imagem mais comuns:
 - VDI virtual disk image, VirtualBox
 - VHD, VHDX virtual hard disk, Hyper-V
 - VHDX surgiu com o WServer2012, permite maior capacidade, melhor desempenho e proteção contra corrupção de ficheiros
 - VMDK Virtual Machine Disk Open format
 - Desenvolvido pela Vmware para os seus produtos, passou a open format e é um dos formatos de disco usado pelos OVF
 - HDD Hard Disk Drive Parallels Desktop
- Existem ferramentas que convertem um tipo de ficheiro para outro

Storage virtual – Sistemas de ficheiros

- Alguns dos sistemas de ficheiros utilizados para guardar os ficheiros das VMs:
 - VMFS virtual machine file system
 - NFS network file system
 - NTFS New technology file system
 - RDM raw device mapping

- A maioria dos hypervisors suporta acesso ao storage local
- Com várias VMs a tentar aceder simultaneamente ao storage local do host, vão existir problemas de desempenho. É necessário utilizar storage de rede (SAN ou NAS)
- Um dos pontos de desempenho mais importantes num sistema virtual é o throughput do disco e do sistema de storage

Cap. 6 – Hardware virtual

- CPU virtual
- Memória virtual
- Storage virtual
- Rede virtual

Rede virtual – switch virtual

- Switch virtual
- Tem a mesma funcionalidade que um switch físico, mas apenas existe na memória do computador
- Quando as VMs comunicam com o exterior, o tráfego é enviado pela NIC física do host.
- Na rede física o tráfego das VMs não se distingue do das máquinas físicas

Cap. 6 – Hardware virtual

- CPU virtual
- Memória virtual
- Storage virtual
- Rede virtual

Apoio à Prática – Open Virtualization Format

TECNOLOGIAS DE VIRTUALIZAÇÃO ENGENHARIA INFORMÁTICA WWW.IPLEIRIA.PT

- Open Virtualization Format (OVF) specification
- Forma normalizada de representar os metadados de uma VM
- Norma da Virtualization Management Initiative (VMAN): projeto da Distributed Management Task Force (DMTF)
- Este formato permite a criação e utilização de VMs dispensando formatos proprietários, facilitando a passagem de VMs de uma plataforma para outra

- Metadados de uma VM:
 - Nome da VM
 - Memória configurada
 - CPU
 - Rede
 - Parâmetros de Storage
 - Outras definições da VM



(Metadados de uma VM)

- O OVF permite ainda ao fornecedor da virtual appliance adicionar:
 - um EULA
 - comentários acerca da VM,
 - parâmetros de arranque,
 - requisitos mínimos,
 - •

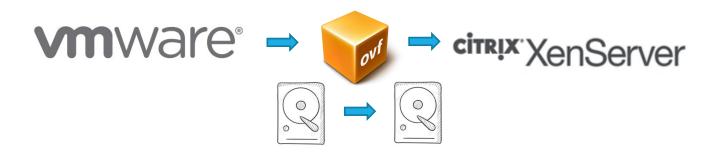
- O OVF não está limitado a uma única VM, podendo descrever várias VMs e a sua relação
- Estas VMs podem constituir assim uma suite de virtual appliances.

Há ainda a possibilidade de encriptação, compressão e assinatura digital do

conteúdo

- O OVF não é uma especificação que descreve um disco virtual
- Para importar um OVF pode ser necessário converter os discos virtuais para um formato compatível com o hypervisor recetor

- Exemplo1: importar no Citrix um OVF produzido em VMware, que inclui um disco VMDK
 - converter primeiro o disco VMDK para um formato compatível com o Citrix
- Ferramenta XenConvert trata da conversão dos discos na importação de conteúdo VMware OVF/VMDK para um ambiente Citrix



 Se a fonte do OVF é o Hyper-V da Microsoft, não é necessária conversão do disco, pois o XenServer suporta de forma nativa o formato VHD

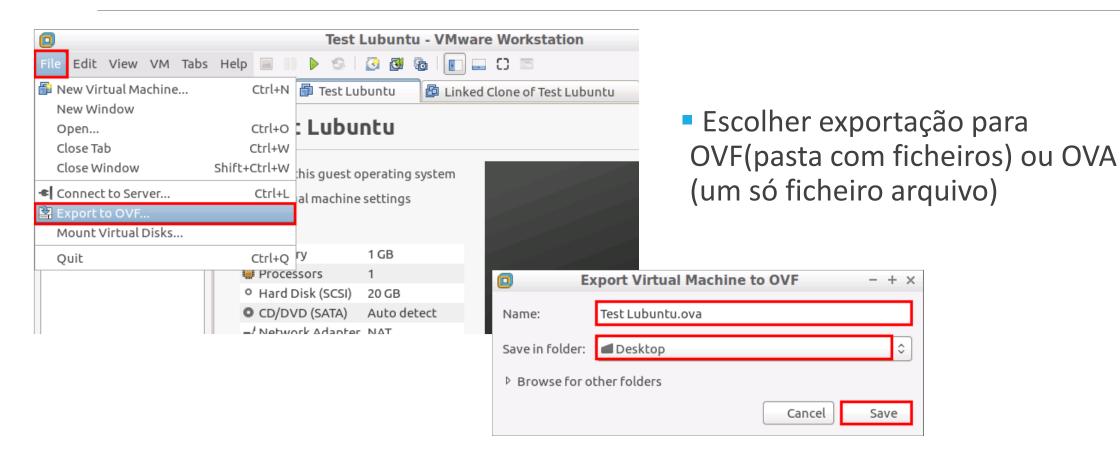


- Exemplo2: importar no VMware um OVF produzido em XenServer, que inclui um disco VHD
 - converter primeiro o disco VHD para um formato compatível com o VMware
- Ferramentas VMware ou VBoxManage (VirtualBox !)

- O OVF é um conjunto de ficheiros
- Um dos ficheiros é um documento XML com extensão .ovf que contém os metadados de descrição da VM (ou VMs)
- Acompanha o ficheiro .ovf um ou mais discos virtuais (.vhd para Citrix, .vmdk para VMware, ...)
- O ficheiro .ovf descreve a localização dos discos

- A norma OVF define 2 maneiras de agrupar os ficheiros:
 - Pacote OVF: grupo de ficheiros numa mesma pasta
 - Pacote OVA (Open Virtualization Appliance) package: ficheiro arquivo único com os vários ficheiros no formato TapeARchive (TAR)

VMware Workstation Pro - OVF



- Formato XVA: descritor de metadados proprietário do Citrix
- O formato OVF é o sucessor natural do XVA
- Utilizar o XenConvert para converter XVA numa combinação OVF + VHD

Microsoft Hyper-V

- O Hyper-V não suporta o formato OVF diretamente
- É necessário utilizar o Virtual Machine Manager (VMM) do System Center
- Pode também ser necessário utilizer o Microsoft Virtual Machine Converter (MVMC), para converter o disco (e.g. VMDK) para VHD.
- É possível exportar / importar em formato próprio entre hosts Hyper-V
- Utilizar este mecanismo também para clonagem de VMs

Apoio à Prática – VBoxManage

TECNOLOGIAS DE VIRTUALIZAÇÃO ENGENHARIA INFORMÁTICA WWW.IPLEIRIA.PT

VBoxManage

- VBoxManage: interface de linha de comando com o VirtualBox
- Permite controlo total do VirtualBox a partir da linha de comando do host
- Suporta todas as funcionalidades da interface gráfica:
 - arrancar, criar, modificar, ... VMs
 - e muitas mais (incluindo algumas experimentais avançadas)!

VBoxManage - Conversão de discos

VMDK -> VDI

C:\> VBoxManage clonehd foo.vmdk foo.vdi --format VDI

VMDK -> RAW

C:\> VBoxManage clonehd foo.vmdk foo.img --format RAW

RAW -> VMDK

C:\> VBoxManage convertdd foo.img foo.vmdk --format VMDK

VBoxManage - Conversão de discos

- Novos nomes para os comandos de conversão:
 - clonehd -> clonemedium
 - convertdd -> convertfromraw
- VMDK -> RAW

C:\> VBoxManage clonemedium foo.vmdk foo.img --format RAW

RAW -> VMDK

C:\> VBoxManage convertfromraw foo.img foo.vmdk --format VMDK

Oracle VirtualBox -Virtual Storage

TECNOLOGIAS DE VIRTUALIZAÇÃO ENGENHARIA INFORMÁTICA WWW.IPLEIRIA.PT



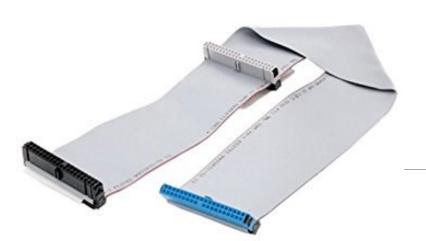
Virtual Storage - métodos

- Existem 3 métodos para apresentar à VM um disco virtual:
 - Disk image files (VDI, VMDK, VHD, HDD): usar grandes ficheiros de imagem num disco real e apresentá-los a um guest como um disco virtual
 - Servidores de storage iSCSI
 - Usar um disco do host diretamente
- Qualquer um destes discos é ligado ao controlador virtual de disco que o hypervisor apresenta à VM



Virtual Storage - controladores

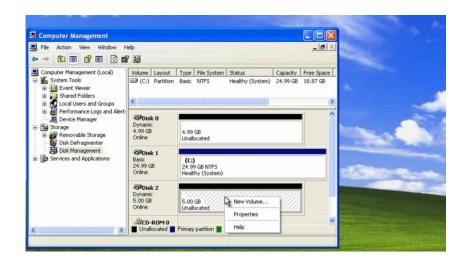
- Os discos são ligados ao controlador de discos que controla a operação dos mesmos e as transferência de dados.
- O hypervisor emula os principais tipos de controladores de discos:
 IDE, SATA (AHCI), SCSI, SAS, USB-based e NVMe



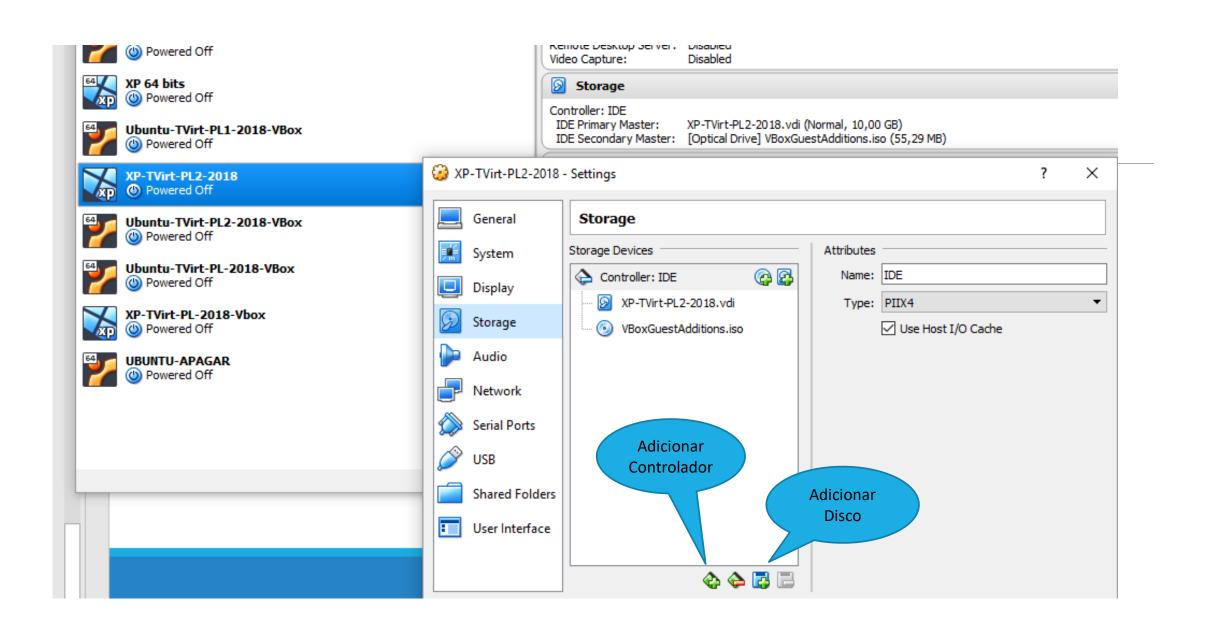
Controladores IDE (ATA)

- Num PC real, cada controlador IDE suporta 2 discos em paralelo (master e slave)
 num cabo, e existem 2 controladores (primário e secundário), permitindo 4 discos
- No virtual box cada VM tem um só controlador IDE que permite os 4 discos
- Pode ser necessário alterar o tipo de controlador IDE (PIIX3, PIIX4 or ICH6) de uma VM para eventual compatibilidade com o seu SO
- Tipicamente um dos 4 discos (o secondary master) está pré-configurado para drive virtual de CD/DVD (mas pode ser alterado)

Controladores IDE (ATA)



- IDE é a solução para quando o SO guest não suporta discos SATA ou SCSI
- No Windows XP para utilizar SATA é necessário:
 - instalar drivers adicionais e
 - adicionar o controlador SATA à VM



Controladores SATA

- Serial ATA (SATA)
- Advanced Host Controller Interface (AHCI): interface para controladores SATA
- Suporta velocidades mais elevadas
- Suporta hot swap
- Controlador mais leve: utiliza menos CPU que o IDE
- Permite até 30 discos numa VM



Controladores SCSI

- Small Computer System Interface (SCSI)
- Cada controlador SCSI suporta até 15 discos
- O VirtualBox suporta dois tipos de controladores SCSI:
 LSI Logic e BusLogic
- Serial Attached SCSI (SAS): versão série do SCSI, mais rápido e mais fiável, permitindo até 8 discos por controlador

Outros Controladores

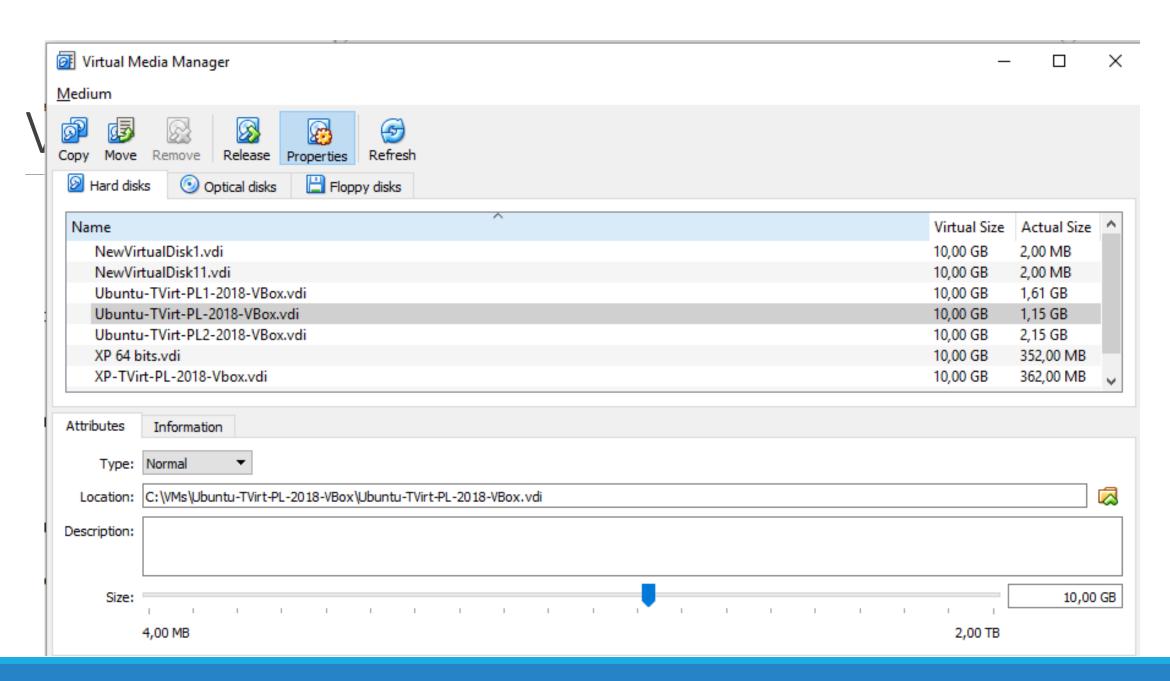
 USB mass storage device class: permite ligar até 8 discos discos externos através de USB

- Non volatile memory express (NVMe):
 - permite ligar até 255 discos SSD sobre PCI express
 - maior largura de banda que os anteriores discos SSD que utilizavam ligação SATA

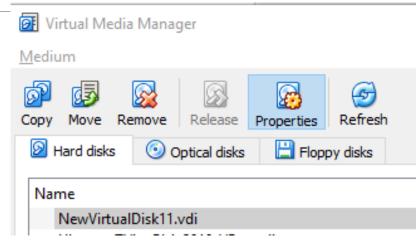


VirtualBox - Virtual Media Manager

- O VirtualBox regista todos os discos usados pelas VMs, designando-os "known media"
- O Virtual Media Manager (Menu file) permite visualizar e alterar estes discos, agrupados em
 - Hard disks (VDI ou outro formato)
 - Optical disks (DVD em formato ISO)
 - Floppy disks (formato RAW)
- Permite visualizar para cada disco: o caminho para a imagem, se está ligado a alguma VM e qual VM, o tipo, formato, tamanho atual, tamanho virtual



VirtualBox - Virtual Media Manager



- Operações disponíveis sobre o disco:
 - Copy: copiar um disco virtual para outro, podendo mudar o formato (VDI, VHD, VMDK)
 - Move: mover para outra localização
 - Remove: remover do registo com opção de apagar ou não o ficheiro imagem
 - Release: libertar uma imagem de uma VM
 - Alterar o tipo de disco: Normal, Immutable, Writethrough, Shareable, Multi-attach

VirtualBox - Virtual Media Manager



- O VirtualBox atribui um identificador único a cada disco (UUID)
- Para copiar um disco, evitar copiar apenas o ficheiro imagem, pois ao importar a cópia o VirtualBox dará erro de UUID duplicado: utilizar o botão de cópia
- Da mesma forma, para copiar uma VM, é melhor utilizar a ferramenta de clonagem de VMs



Existem 6 modos de escrita em disco:
 normal, write-through, Shareable, immutable, multiattach e read-only

- Por omissão as imagens de disco estão no modo normal:
 - podem ser lidas e escritas sem restrições
 - num snapshot o seu estado é guardado, e restaurado quando se reverte para ele
 - pode ligar-se a mais que uma VM, mas não se pode executar mais que uma dessas
 VMs simultaneamente

- Discos write-through:
 - são ignorados pelos snapshots



- também são ignorados pelos snapshots,
- mas podem ser ligados a várias VMs que podem correr de forma concorrente
- utilizados para sistemas de ficheiros em clusters com aplicações preparadas explicitamente para aceder o disco de forma concorrente a partir de várias VMs
- Só são permitidos discos de tamanho fixo (não de alocação dinâmica)





Discos immutable

- apenas relembram acessos de escrita temporariamente enquanto a VM está a correr
- todas as alterações são perdidas quando a VM é ligada novamente no hypervisor (não se perdem num reboot a partir da VM)
- tecnicamente o hypervisor nunca escreve diretamente num disco immutable. Todas as operações de escrita da VM são direcionados para uma imagem diferencial.
- logo, podem ser usados por várias VMs em simultâneo sem restrições
- não faz normalmente sentido criar um disco immutable de raíz
- normalmente cria-se uma imagem normal e depois torna-se immutable





- Discos em multiattach mode
 - iguais aos discos immutable
 - mas as alterações não são perdidas quando a VM é ligada novamente no hypervisor
 - muito útil para partilhar ficheiros que quase nunca são escritos, como galerias de fotos
- Disco read-only:
 - usado automaticamente nas imagens CD/DVD





- Considerar que uma VM com snapshot realizado, é infetada por um vírus
- Para o vírus desaparecer:
 - Com um disco normal, reverte-se para o snapshot
 - Com um disco immutable, basta desligar e voltar a ligar a VM no hypervisor
 - Com um disco write-through é preciso desinfetar a VM
- Utilização possível:
 - Disco immutable para disco do SO da VM
 - Disco write-through para o disco de dados da VM

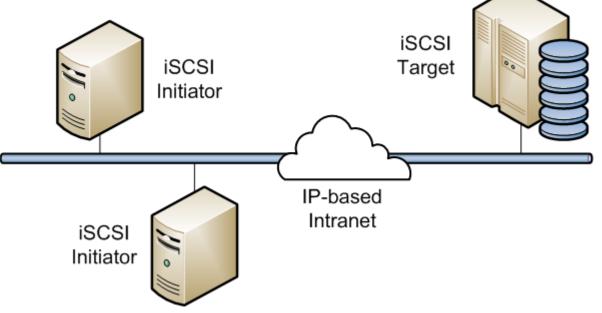




- Comparação com os modos dos discos VMware:
 - discos write-through discos Independentes Persistentes.
 - discos immutable discos independentes Não Persistentes.

Servidores iSCSI

- iSCSI Internet SCSI
- Norma que permite utilizer o protocolo SCSI sobre ligações de rede IP
- O servidor que disponibiliza os recursos de storage é o iSCSI target
- Os clientes são os iSCSI initiators



Servidores iSCSI - VirtualBox

- O VirtualBox tem um iSCSI initiator integrado, que apresenta o storage remoto às VMs como um disco virtual.
- O target é configurado por linha de comando: VBoxManage storageattach

Referências

- Virtualization Essential Training, Martin Guidry, Lynda.com
- Virtualization Essentials, Matthew Portnoy, Sybex, Wiley
- www.virtualbox.org
- support.citrix.com