Virtualização de Servidores

Paulo Pires
Maj Cav Engº Informático

Por opção do autor, este artigo está redigido segundo os instrumentos ortográficos anteriores ao Acordo Ortográfico da Língua Portuguesa de 1990.

Resumo

Virtualização, pedra base de toda a tecnologia cloud, apresenta-se como uma tecnologia de mercado cada vez mais em uso e numa forma eficaz de fazer mais com menos investimento. No boletim anterior do IGeoE, foi feita uma abordagem à virtualização na sua componente histórica desde a década de 60 com a virtualização de simples desktops até aos dias de hoje com sistemas completos como o ESX Server da VMware ou o XenServer da Citrix Inc. Este artigo aborda o dinamismo constante do triângulo existente entre fornecedores de software, hardware e seus clientes, inserido em tecnologias como virtualização de servidores, armazenamento e rede.

O dinamismo do triângulo existente entre fornecedores de *software, hardware* e seus clientes é uma constante. Fornecedores como a VMware e a Microsoft publicitam as suas soluções de virtualização destacando as mais-valias como a utilização eficiente de recursos partilhados, gestão centralizada, e migração dinâmica; os fornecedores de *hardware* publicitam menores consumos de energia, e maior interoperabilidade e suporte; as organizações olham para este mundo com o objectivo último da eficiência, produtividade e custos. A virtualização aparece como uma forma eficaz de fazer mais com menos investimento.

A primeira razão pela qual se torna muito interessante executar múltiplas VMs num único servidor é a mesma pela qual executamos múltiplas aplicações num único servidor: os recursos estariam, se não o fizéssemos, muito subaproveitados. Sendo assim, porque não executar simplesmente múltiplas aplicações no servidor usando um só SO? Podemos avançar desde já com uma razão simples: os fornecedores certificam os seus produtos, aplicações, e serviços apenas para determinados ambientes e pilhas de software; vejamos alguns exemplos: 1) a Microsoft não certifica configurações MSCS (Microsoft Cluster Server) que incluam o serviço Active Domain num dos nós dos cluster; 2) a VMware disponibiliza o seu produto de gestão vCenter Server sobre Unix (SUSE Linux) na forma de uma appliance fornecida pela própria VMware. Assim, se dispusermos de um servidor que tem os recursos apropriados para suportar um nó MSCS e uma instância vCenter teremos de os instalar sob a forma duas VMs a correrem sob um hipervisor instalado nesse servidor. Uma outra situação que

"obriga" (do ponto de vista financeiro) à utilização de VMs é a de aplicações *legacy*: uma aplicação antiga (e é demasiado oneroso actualizá-la) está certificada apenas para uma dada versão de SO, também antigo, que não possui *drivers* adequados para os servidores mais actuais, sendo que o servidor antigo já não tem capacidade para executar a aplicação (e é muito oneroso mantê-lo). Uma solução simples é instalar um hipervisor num servidor recente, criar uma VM, e nela instalar a versão antiga do SO e a aplicação.

A base de uma arquitectura de virtualização é o VMM ou hypervisor, o Virtual Machine Manager responsável pela criação, gestão, isolamento e preservação do estado da VM, assim como toda a orquestração do acesso aos recursos do sistema host.

O VMM permite que seja possível a execução de vários SO's em VMs, contudo está limitado a SO's que possam ser executados nativamente no processador físico do sistema. Essa valência torna-se hoje uma das maiores procuras de todos os sistemas de virtualização.

Em termos de implementações arquitecturais temos VMM's: Tipo-2, modelo "hybrido" e Tipo-1. O VMM ou hypervisor do Tipo-2, é implementado dentro do próprio SO real (Linux ou Windows - hostsystem) e executando paralelamente, é mais um processo. Neste caso, trata-se de uma camada hypervisor própria como um segundo e distinto nível de software. Os sistemas operativos convidados correm num terceiro nível acima do hardware. Nesta arquitectura, o SO guest acede directamente o SO host (nativo) por intermédio de uma API cedida pelo hipervisor ao SO guest. O SO guest acede assim ao hardware através de um device driver específico pelo hypervisor e pelo SO host, sendo por isso uma implementação que apresenta menor desempenho e maior sobrecarga ao sistema. O Java VM, VMware Workstation, VMware Player, Sun Microsystems VirtualBox e KVM são alguns exemplos que usam esta arquitectura.

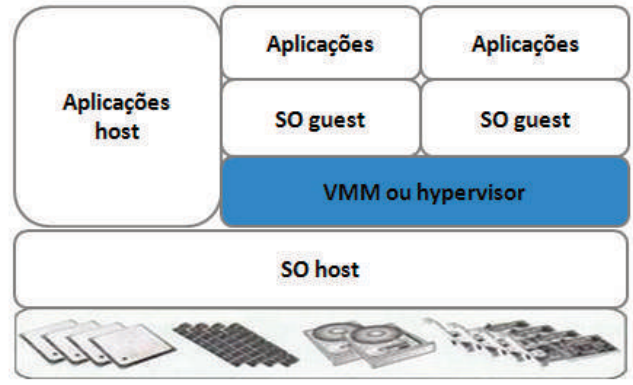

Figura 1 - VMM ou hypervisor do Tipo-2

No modelo "híbrido", o VMM não é executado uma camada acima do SO host (ou abaixo) mas sim concorrentemente (peer) com o SO host e não depende de instruções específicas no processador. Microsoft Virtual Server 2005 R2 e Virtual PC são alguns exemplos que usam esta arquitectura.

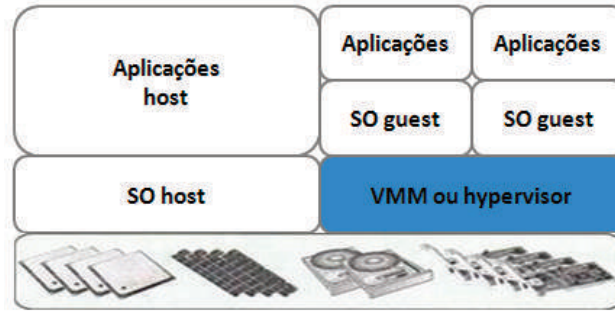

Figura 2 - VMM ou hypervisor do "híbrido"

No VMM ou hypervisor do Tipo-1, é executado directamente no hadware, a camada mais baixa de todas as partições da VM, sem necessidade da existência de qualquer SO host. O SO guest acede directamente ao hardware através do hipervisor, tal é possível pelas modificações feitas no SO guest e no hipervisor. Esta é a implementação que atinge níveis mais elevados de eficiência por isso permite uma maior densidade de VMs, é a implementação clássica de arquitecturas VM. VMware ESX/ESXi, Citrix XEN Server e Microsoft Hyper-V são alguns exemplos que usam esta arquitectura.

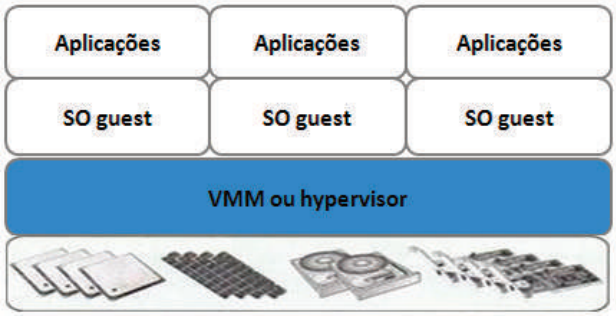

Figura 3 - VMM ou hypervisor do Tipo-1

Conceitos como a arquitectura da virtualização descrita e os tipos de Implementação são abordagens essenciais que ajudam a perceber o mundo desta tecnologia.

Na Emulação de Hardware, pretende-se a virtualização completa do hardware, o seu comportamento/estados de execução (ciclos de clock, conjunto de instruções, pipeline, memória cache), por isso é a forma de virtualização mais complexa. Todo o hardware da VM é criado via software no sistema hospedeiro para emulação do hardware proposto criando assim grandes overheads[1], com fracos desempenhos comparativamente ao hadware real. Este tipo de virtualização apresenta mais-valias como o facto de poder ser usado um SO guest sem qualquer modificação ou adaptação, os programadores podem fazer testes de firmware e hardware em hardware que não o real nem necessitam da existência deste e ainda a valência de ser possível emular hardware que, na maior parte das vezes é bastante diferente do hardware real.

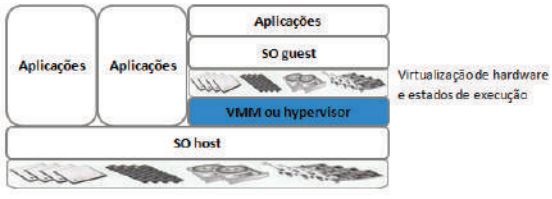

Figura 4 - Emulação de Hardware

1 Overhead: Diferença de performance (latência/lentidão) no SO guest

A virtualização completa, também conhecida como "nativa", é em todo igual a emulação de hardware excepto à não representatividade dos estados de execução do hardware emulado, por esse facto, esta técnica de virtualização apresenta resultados mais satisfatórios no seu desempenho mas ainda aquém dos SO's em execução nativa. Esta técnica permite assim a execução de todos os SO's originais, sem alteração, fornecendo uma réplica de todo o hardware dessa máquina, levando-os a ter a ilusão de estar a executar directamente no hardware.

Os grandes obstáculos desta técnica foram desde logo as diferenças arquitecturais e os comportamentos particulares das instruções, muitas impossíveis de serem emuladas ou capturadas pois muitas dependem do nível de privilégio. Assim, para lidar com a heterogeneidade de processadores e comportamento de instruções, usa-se uma abordagem chamada de Tradução Binária. Nesta abordagem, a VMM analisa todas as instruções da VM, quer instruções não privilegiadas que depois acedem ao hardware com drivers genéricos, quer instruções privilegiadas e quando na presença destas, faz a emulação e reescreve dinamicamente o código. Este teste a todas as instruções acarreta uma latência significativa no desempenho.

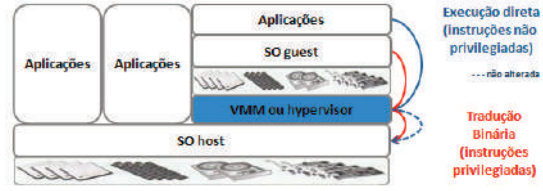

Figura 5 - Virtualização Completa

A Para-Virtualização ou virtualização assistida por SO, é uma alternativa á virtualização completa. O SO guest é modificado (perdendo a portabilidade) para melhorar a eficiência usando todos os "limites" do sistema. A ideia é acelerar a execução das instruções e para isso a VMM testa apenas instrução que podem alterar o estado do

sistema (instruções sensíveis) aumentando significativamente o desempenho. Esta substituição de uma instrução sensível pelo tratador de interrupção de software é chamada de hypercall, são instruções executadas através de tradução binária. O SO guest é assim modificado para chamar a VMM sempre que estamos na presença dessas instruções sensíveis.

Esta alternativa de virtualização com a modificação do kernel, também permite que o SO guest aceda directamente aos recursos de hardware com os drivers da própria máquina virtual e deste modo, não usando os drivers genéricos (virtualização completa), usam a capacidade total dos dispositivos. Assim, a para-virtualização apresenta um ganho significativo em relação à virtualização completa, compensando as modificações implementadas no SO guest.

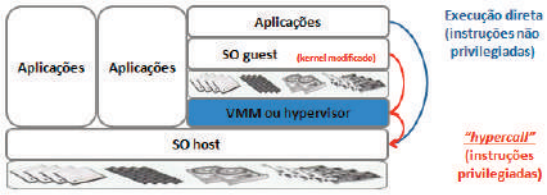

Figura 6 - Para-Virtualização

As técnicas descritas apresentam alguns inconvenientes que enfraquecem o desempenho como o teste a todas as instruções e tradução das instruções privilegiadas na Virtualização Total e a alteração do SO guest na Para-Virtualização, agarrando o SO á arquitectura, limitando a compatibilidade e suporte.

A Virtualização assistida por Hardware, é aplicada directamente nos processadores e restante hardware, são usadas extensões de virtualização do processador e hardware para virtualização dos guest's, fornecendo assim um recurso que permita ao hypervisor executar ainda mais próximo do hardware. As instruções sensíveis são agora entregues sem a necessidade da tradução binária aumentando significativamente o desempenho mas ainda um pouco mais lento que a PV. A opção do uso de drivers como a PV torna-se um avanço notório neste tipo de implementação. Não emular I/O e armazenamento e ainda oferecer uma plataforma de VMs com SO's não modificados é uma junção das mais-valias dos dois mundos (PVHVM ou PV-on-HVM drivers). Torna-se assim possível, por exemplo, obter um óptimo desempenho no conjunto SO guest Windows + XenServer.

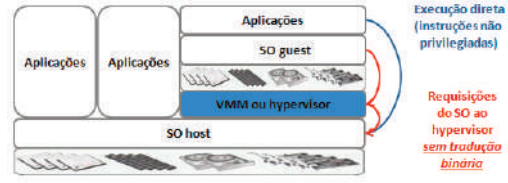

Figura 7 - Virtualização assistida por Hardware

A Recompilação Dinâmica (dynamic recompilation - DynaRecs), também denominada como tradução dinâmica (dynamic translation) é uma técnica de virtualização bastante utilizada. Traduz as instruções de um determinado formato para outro formato, durante a própria execução do programa, permitindo assim a criação do ambiente nativo do programa. Essa tradução é relativa a instruções do SO guest e das suas aplicações, mais próximas do SO host. Assim, o VMM ou hipervisor analisa, reorganiza e traduz as sequências de instruções emitidas pelo SO guest em novas sequências de instruções (código de mais alto nível) compiladas na linguagem nativa do sistema host para que esse código gerado seja mais eficiente.

Essa eficiência de código prende-se com adaptação de instruções à interface ISA do sistema real, detectar e tratar instruções sensíveis e ainda analisar, reorganizar e optimizar sequências de instruções do SO guest com o fim último da eficiência e desempenho da sua execução. Nesta ultima análise, é usual guardar em cache a tradução dos blocos de instruções frequentes, melhorando assim ainda mais o desempenho. É notória a vantagem desta técnica como o acesso a

código em tempo de execução não alcançável a um compilador estático mas em contrapartida, também exige muito mais processamento.

Actualmente o número de tecnologias disponíveis para implementação e gestão de máquinas virtuais tem vindo a crescer, destacando-se algumas pela disponibilização de um conjunto de ferramentas que as tornam vencedoras nos mercados. Seguem-se quatro exemplos das empresas (e suas tecnologias) que mais se destacam hoje em dia:

- *VmWare, Inc.:* VmWare Player, Workstation, Fusion, ESXi
- *Citrix Systems, Inc:* XenServer
- *Hewlett-Packard:* HP Integrity Virtual Machines
- *Microsoft:* Virtual PC, Hyper-V

1.1.1. VMware (VMware, Inc.)

O VMware apresenta-se como uma plataforma líder (80%) no mercado da virtualização, com uma tecnologia de virtualização do tipo 1 (virtualização completa, também designada nativa ou *bare metal*). Oferece um vasto leque de produtos, desde os hipervisores para *desktops* e *laptops* (VMware Player/Workstation/Fusion) aos para servidores de grande porte (VMware vSphere ESXi) e ainda um vasto conjunto de *software*: de gestão de VMs, orquestração de operações, *Cloud* IaaS, etc..

De todos os produtos da VMware, interessa-nos destacar aqui apenas a tecnologia de virtualização ESXi, oferecida para servidores, que descrevemos em seguida muito sucintamente:

- *VMware vSphere ESXi*: É uma VMM ou hipervisor tipo 1 garantindo um elevado nível de isolamento entre os recursos oferecidos às VMs, sejam estes o processador, memória, discos ou adaptadores de rede. É instalado directamente sobre o *hardware* do servidor, eliminando assim a sobrecarga de ter um SO standard sobre o qual corre um hipervisor; os hipervisores de tipo 1 exibem, portanto, melhor desempenho e aumentam a segurança. O facto de permitir que tudo seja virtualizado torna a VM ainda mais completa. O pacote inclui apenas o hipervisor ESXi e ferramentas básicas de gestão.

A forma como as instruções do sistema hospedado (*guest*) são executadas num ambiente ESXi pode ser descrita com ajuda da figura seguinte: 1) as instruções oriundas de aplicações que se executam sobre o SO hospedado são executadas directamente no processador real; 2) as instruções oriundas do próprio SO hospedado são verificadas antes de serem executadas, sendo que as privilegiadas são, traduzidas ou modificadas para comandos ou instruções do próprio hipervisor, uma técnica que é denominada como "Tradução Binária".

Esta técnica, complementada com outras técnicas de "aceleração" de operações sobre I/O, memória e de gestão de recursos entre VMs fazem com que a VMware consiga atingir desempenhos muito próximos de um ambiente nativo, i.e., não virtualizado.

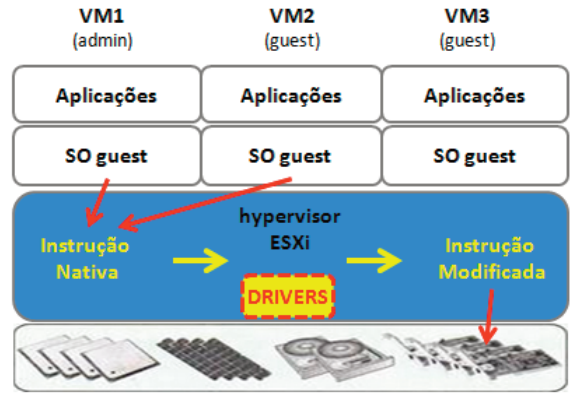

Figura 8 - Arquitectura VMware ESXi (Tipo-1)

Na versão vSphere para servidores de grande porte estão disponíveis muitas funcionalidades, como a migração automática de VMs (balanceamento de carga), a alta disponibilidade para vCenter (suportada pelo vCenter Server

Heartbeat), a recuperação de destastres (Site Recovery Manager).

1.1.2. Xen (Citrix Systems, Inc.)

O XenServer é a plataforma de virtualização da Citrix, baseada no hipervisor Xen; a versão base, designada Standard, é gratuita e oferece, para além do hipervisor, ferramentas como o XenCenter, uma consola de gestão para múltiplos servidores, e o XenMotion para migração "in vivo" de VMs.

O hipervisor Xen, na sua mais recente versão, é uma tecnologia de virtualização do tipo 1, instalada num núcleo Linux (x86, x86-64, ou IA-64 SUSE Linux), que permite a criação de várias VMs com recursos e aplicações partilhadas. Contrariamente ao VMware, não usa a tradução binária, mas uma combinação de paravirtualização (eliminando a latência no desempenho introduzida pela tradução binária) e virtualização assistida por *hardware*. Na paravirtualização a máquina abstracta é quase, mas não completamente igual ao *hardware* hospedeiro; por isso, o SO hospedado (*guest*) tem de ser modificado. Como tal só é possível para SOs disponíveis em código aberto, o Xen executa os outros (e.g., Windows) recorrendo a uma combinação da paravirtualização com a virtualização assistida por *hardware* oferecida nos processadores Intel (Intel-VT) e AMD (AMD-V).

Na arquitectura XenServer, existe uma VM privilegiada para controlo do hipervisor, denominada "Domo"; esta VM é parte integrante do ambiente XenServer, executa uma versão paravirtualizada do Linux, e contém os *drivers* dos dispositivos disponíveis no sistema. As outras VMs são os "DomUs" (domínios do utilizador). Nesta arquitectura, os DomUs comunicam directamente com o hipervisor que controla a memória e o processador do hospedeiro, e, se usam paravirtualização, comunicam com o Domo para operações de I/O.

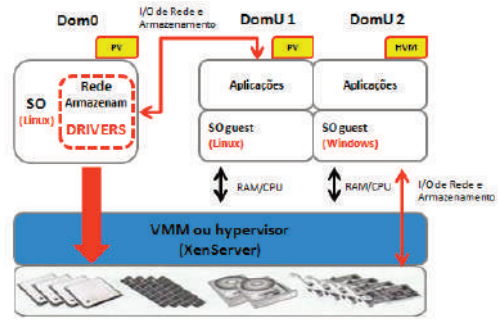

Figura 9 - Arquitectura XenServer

1.1.3. Integrity VM (HP)

HP Integrity Virtual Machines (Integrity VM ou HPVM) é uma tecnologia de virtualização do tipo 1, instalada num servidor da linha Integrity IA-64 com o sistema de operação HP-UX 11i v3.

A gestão do hipervisor é efectuada com a ferramenta HP Integrity Virtual Machines Manager (HPVMMGR); mas, também é possível gerir os próprios sistemas hospedados desde que nestes sejam instalados os agentes HPVirtualMachine e HPVSwitch. Esta ferramenta apresenta benefícios como flexibilidade e maximização dos recursos do servidor; consolidação de servidores organizacionais; rapidez na implementação e disponibilização de novos ambientes; permite a migração online e offline; isolamento de VMs; oferece alta disponibilidade; monitoriza automaticamente todas as aplicações; e permite a visualização e configuração simplificada com ferramentas como HP Integrity Virtual Machines Manager (HPVMMGR), System Management Home (HPSMH) e HP Virtualization Manager (HPVMAN).

1.1.4. Hyper-V (Microsoft)

O Hyper-V é uma tecnologia de virtualização do tipo 1, da Microsoft, para servidores x86-64 com suporte à virtualização assistida por *hardware*. O produto é disponibilizado de forma isolada (*standalone*), ou integrado em versões

Windows Server (2008 R2 ou 2012).

A gestão do ambiente virtual é efectuada com a ferramenta System Center Virtual Machine Manager (SCVMM), que fornece 3 formas distintas de interactuar com o hipervisor: consola de administração, portal self-service (para utilizadores finais) e Windows PowerShell (usando commandlets).

Na arquitectura Hyper-V, existem dois tipos de partição: 1) a "partição pai", uma VM de administração de todo o sistema, com um Windows Server 2008, 2008 R2 ou 2012, *drivers* de acesso e controlo do *hardware* e ferramentas de gestão; 2) as partições filhas, nas quais são criadas as restantes VMs, completamente isoladas e sem acesso directo ao hardware. No acesso ao *hardware*, há 3 componentes importantes a mencionar: a) o VSP (Virtual Service Provider) responsável por receber as chamadas das partições filhas e aceder aos *drivers* dos periféricos; b) VSC (Virtual Service Client), um módulo instalado no SO hospedado que solicita, via VMBus, o acesso aos periféricos da partição pai; c) o VMBus, canal de comunicação entre VSC e VSP onde se desenrolam as comunicações entre as partições.

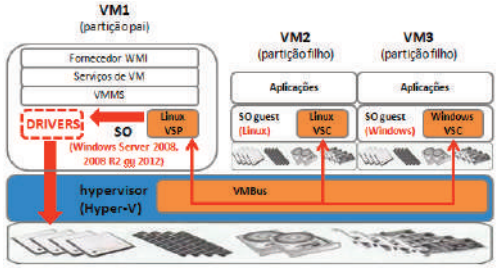

Figura 10 - Arquitectura Hyper-V

A Microsoft apresenta-nos, na recente versão Hyper-V 2012, uma panóplia de funcionalidades que a colocam na primeira linha das soluções de virtualização, a par da VMware: migração online e replicação de VMs, migração de armazenamento, e *Clustering*.