Avaliação de Desempenho de Virtualização Baseada em Contêiner

[1]**Naylor G. Bachiega, Paulo S. L. Souza, Sarita Bruschi**

[1]Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo
Avenida Trabalhador São-carlense, 400 - 13566-590 - São Carlos - SP

naylor@usp.br, {pssouza,sarita}@icmc.usp.br

Abstract. *Virtualization is an important technology used in large data centers and the availability of on-demand resources. It has gained adherence across industry segments and is widely studied to deliver significant cost performance. As such, containers have gained prominence as a new form of virtualization. Unlike a hypervisor-based system, containers share the same host operating system to create multiple isolated instances of virtualized programs. Containers perform is equal to or greater than the performance of a hypervisor-based system, however, they have different architecture and implementation. Thus, performance evaluation of the containers is fundamental for adoption and creation of new projects, since containers have gained adhesion from large companies.*

Resumo. *A virtualização tem ganhado adesão e suporte em vários segmentos da indústria e é amplamente estudada para oferecer o desempenho desejado a um custo significativo. Os contêineres ganharam destaque como uma nova forma de virtualização. Ao contrário de um sistema baseado em hipervisor, os contêineres compartilham o mesmo sistema operacional para criar várias instâncias isoladas de programas virtualizados. Estes possuem desempenho similar ou superior ao de uma máquina virtual, no entanto, têm arquitetura e implementação distintas. Dessarte, avaliar seu desempenho é fundamental para a adoção e criação de novos projetos, uma vez que contêineres têm ganhado a adesão de grandes empresas.*

1. Introdução

A virtualização é uma tecnologia amplamente utilizada pela indústria e está em constante expansão com o desenvolvimento de aplicações, gestores, modelos e simuladores para promover e melhorar o desempenho e gerenciamento.

No entanto, recentemente, esta área está aderindo a um novo tipo virtualização, o contêiner. Nesse tipo de virtualização, é possível a existência de múltiplas instâncias no espaço de usuário do sistema operacional. Essa nova modalidade tem ganhado adesão e apoio de grandes empresas.

De acordo com [Chung et al. 2016], existem dois modelos de virtualização: baseado em hipervisor e baseado em contêiner. Para [Turnbull 2014], as execuções baseadas em hipervisor utilizam máquinas virtuais como instâncias e no caso dos contêineres, as instâncias são executadas no espaço de usuário na parte superior do kernel do sistema operacional.

Existem algumas desvantagens na utilização de contêineres em comparação com os hipervisores, que estão relacionadas com o isolamento e segurança. Porém, uma das vantagens é o desempenho, podendo ser igual ou superior ao processamento das máquinas virtuais (VMs) em determinadas situações.

Analisar essas diferenças de desempenho é importante para a criação de novos projetos ou avaliar projetos existentes. O contêiner possui diferente arquitetura e implementação em relação a virtualização baseada em hipervisor [Chung et al. 2016], portanto a avaliação de desempenho para este novo modelo deve ser analisada, com modelos de previsão de carga de trabalho, simulação entre outras técnicas de medição.

2. Virtualização Baseada em Contêiner

O contêiner é um método de virtualização no nível do SO (sistema operacional). Tornou-se popular, pois pode executar aplicações distribuídas sem a necessidade de criar uma instância de máquina virtual.

Desta forma, é possível criar múltiplos sistemas isolados que funcionam em um único SO e acessam um único kernel. Ao contrário das VMs, os contêineres podem ser mais eficientes, pois podem executar várias instâncias em um único sistema operacional.

Como pode ser visto na Figura 1, os contêineres não têm a sobrecarga exigida pelo hipervisor, pois há apenas um SO gerenciando a interrupção de hardware.

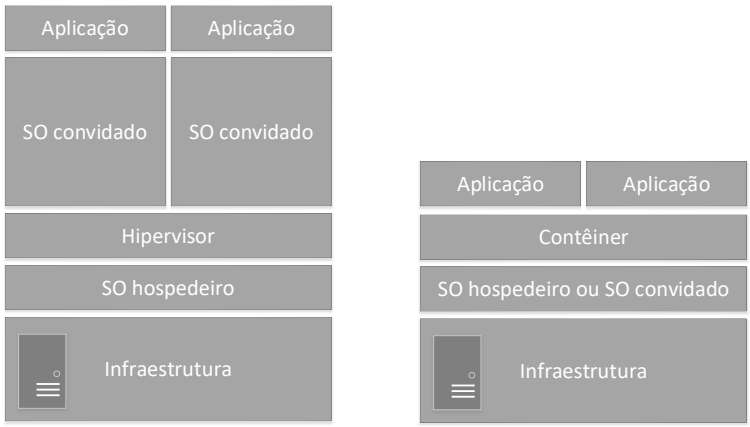

Figura 1. Virtualização baseada em hipervisor e contêiner.

Porém, eles são vistos como menos flexíveis, pois normalmente executam apenas o mesmo ou um sistema operacional semelhante.

Apesar dessa desvantagem, os contêineres podem ser criados mais rapidamente do que as máquinas virtuais. Os contêineres contêm todos os recursos necessários para executar o software, como arquivos, bibliotecas e acesso à rede. Para impedir que qualquer contêiner use todos os recursos do sistema operacional em uma disputa, o SO faz o controle de acesso aos recursos físicos.

[Turnbull 2014] cita que os contêineres foram considerados menos seguros do que o isolamento completo do hipervisor, pois eles compartilham o mesmo sistema operacional. No entanto, os contêineres têm uma área de ataque menor do que um SO completo

necessário para uma VM, assim como a camada do hipervisor também está sujeita a ataques.

3. Avaliação de Desempenho

Avaliação de desempenho é essencial para medir a eficácia de um sistema dentro de um limite aceitável. Os testes são realizados com base em modelos matemáticos ou mesmo simulações complexas. O desempenho é um fator importante no desenvolvimento de um programa, assumindo que outras características podem afetar o desempenho global, como seus respectivos subsistemas, componentes de hardware e software [Heidelberger and Lavenberg 1984].

O desempenho está presente durante toda a vida útil de um sistema. Esta é uma área complexa, que possui várias metodologias e ferramentas para analisar e determinar a eficácia de um programa. Além disso, para avaliar o desempenho é necessário selecionar as medidas corretas que são esperadas dentro de um intervalo estabelecido, o ambiente de medição correta e as técnicas corretas [Jain 1991].

[Heidelberger and Lavenberg 1984] classificam os métodos de avaliação do desempenho computacional em três técnicas: **modelagem analítica**, **simulação** e **medição**.

De acordo com [Jain 1991], a técnica de medição é usada quando o sistema ou parte dele já existe. As técnicas de modelagem analítica e simulação são usadas quando o sistema não existe ou não está disponível para avaliação.

4. Avaliação de Desempenho Baseada em Contêiner

As técnicas de avaliação de desempenho estão sendo utilizadas em contêineres para avaliar sua eficácia e eficiência em comparação com outros tipos de virtualização.

No entanto, como esse tipo de virtualização possui implementação e arquitetura diferente de um hipervisor, outras técnicas de medição, cargas de trabalho e modelos de desempenho devem ser utilizados.

Uma pesquisa mais recente, apresentada por [Chung et al. 2016], mostra a evolução do desempenho do contêiner em relação às VMs e à máquina física (PM). Na Figura 2 é possível observar esta evolução.

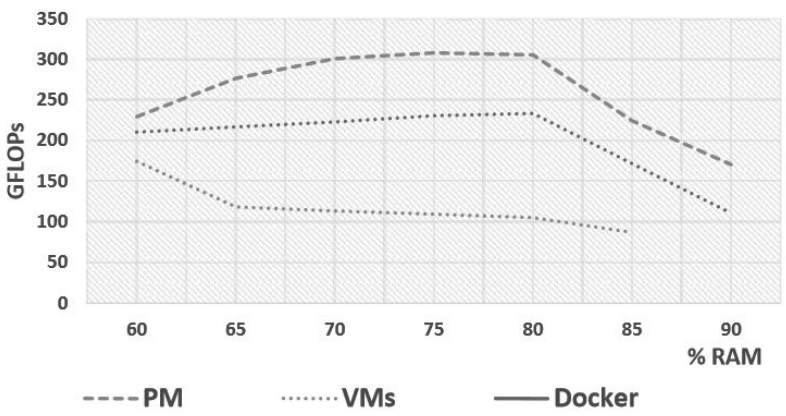

Figura 2. HPL benchmark [Chung et al. 2016].

[Chung et al. 2016] demonstrou na Figura 2 que as máquinas virtuais funcionam melhor com matrizes menores que 65% de RAM e os contêineres funcionam melhor com arrays que variam de 75% a 80% de RAM. Este resultado demonstra uma evolução nos resultados de referência para processamento de grandes quantidades de dados por contêineres.

Em outra comparação do IBM Research Report, [Felter et al. 2015] comparou o uso de máquinas virtuais (KVM) e contêineres (Docker). Os autores demonstraram que o desempenho do Docker é igual ou superior ao desempenho do KVM em todos os casos testados.

Todavia, os trabalhos encontrados utilizam apenas benchmarking (medição) para avaliar o desempenho dos contêineres em comparação com outras tecnologias de virtualização. As medições são suscetíveis a erros e bugs experimentais, assim como as demais técnicas, porém ela pode ser utilizada em conjunto com outras técnicas para fins de validação [Jain 1991].

5. Conclusão

Ao contrário da virtualização baseada em hipervisor, o contêiner possui problemas de isolamento, segurança e seu hardware não pode ser personalizado. Outro recurso ausente no contêiner é a capacidade de fornecer meios para a computação distribuída e serviços baseados em políticas.

Porém, possui capacidades importantes para sua adoção, como redução na complexidade das abstrações, orquestração, gerenciamento de cluster e desempenho.

Os contêineres têm uma arquitetura diferente de um hipervisor, com base nisso, o modelo analítico e simulação também são técnicas necessárias para a avaliação de desempenho e validação de modelos. Assim, como este processo ocorreu com a virtualização em níveis inferiores, ele também deverá ocorrer no nível do sistema operacional.

Referências

Chung, M. T., Quang-Hung, N., Nguyen, M. T., and Thoai, N. (2016). Using docker in high performance computing applications. In *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*, pages 52–57.

Felter, W., Ferreira, A., Rajamony, R., and Rubio, J. (2015). An updated performance comparison of virtual machines and linux containers. In *2015 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, pages 171–172.

Heidelberger, P. and Lavenberg, S. S. (1984). Computer performance evaluation methodology. *IEEE Transactions on Computers*, C-33(12):1195–1220.

Jain, R. (1991). *The Art of Computer Systems Performance Analysis: techniques for experimental design, measurement, simulation, and modeling*. Wiley.

Turnbull, J. (2014). *The Docker Book: Containerization is the New Virtualization*. James Turnbull.