

Em Direção à Comparação do Desempenho das Aplicações Paralelas nas Ferramentas OpenStack e OpenNebula

Carlos A. F. Maron¹, Dalvan Griebler², Adriano Vogel¹, Claudio Schepke³

¹Curso Superior de Tecnologia em Redes de Computadores – Faculdade Três de Maio (SETREM) – Três de Maio – RS – Brasil

²Programa de Pós-Graduação em Ciência da Computação – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) – Porto Alegre – RS – Brasil

³Laboratório de Estudos Avançados (LEA) Universidade Federal do Pampa (UNIPAMPA) – Alegrete – RS – Brasil

{francocaam, adrianovogel103}@gmail.com,
dalvan.griebler@acad.pucrs.br, claudioschepke@unipampa.edu.br

Resumo. *A infraestrutura de Computação em Nuvem vem sendo uma alternativa à execução de aplicações de alto desempenho. No entanto, o desempenho pode ser prejudicado devido a camada de virtualização e da ação das ferramentas de administração de nuvem. O objetivo deste trabalho foi comparar o desempenho de aplicações em OpenStack e OpenNebula. Os resultados apresentaram diferença significativa entre as ferramentas e positiva ao OpenNebula.*

1. Introdução

A computação de alto desempenho tradicionalmente tem sido realizada em *clusters* dedicados [Buyya et al. 2010]. Por outro lado, a computação em nuvem, permite que grandes infraestruturas de processamento sejam mantidas e gerenciadas dinamicamente. Esta baseia-se na oferta de modelos de serviços como IaaS (infraestrutura como serviço); PaaS (plataforma como serviço) e SaaS (*software* como serviço), os quais podem ser implantados em um ambiente público, híbrido, privado ou comunitário.

Infraestruturas do tipo IaaS tornam-se o cenário recomendado para execução de aplicações paralelas na nuvem. As ferramentas de administração permitem maior elasticidade em relação a *clusters* nativo, alocação sob demanda, eficiência energética, redução de custos e facilidade na manutenção e configuração do ambiente. No entanto, a camada de virtualização em conjunto com as ferramentas de administração, torna-se um ponto questionável no desempenho, podendo influenciar positivamente ou negativamente.

A literatura apresenta trabalhos que comparam o desempenho de aplicações paralelas considerando o impacto dos virtualizadores e, das plataformas provedoras de serviços de nuvem (Seção 2). Os estudos de [Thome et al. 2013, Maron et al. 2014a], mostram que cada ferramenta de administração possui características específicas. No trabalho de [Maron et al. 2014b], foram efetuados testes de performance no isolamento de recursos nas ferramentas OpenStack e OpenNebula, o qual demonstrou diferenças significativas nos resultados. Assim, o desafio deste trabalho é comparar o desempenho de aplicações paralelas nestas ferramentas e verificar o que influenciou.

O trabalho está organizado em 4 seções. A Seção 2 apresenta os trabalhos relacionados. Na Seção 3 são elencados os resultados obtidos a partir de testes com aplicações nos diferentes ambientes. Por fim, são relacionados a conclusão e os trabalhos futuros.

2. Trabalhos Relacionados

Na literatura encontra-se vários estudos relacionados, os quais avaliam e comparam o desempenho de aplicações paralelas (usando o *NAS Parallel Benchmark*) em: plataformas de nuvem pública [Navaux et al. 2012, Evangelinos and Hill 2008, Gupta and Milojevic 2011, Strazdins et al. 2012] e virtualizadores [Xavier et al. 2013, Regola and Ducon 2010]. O estudo de [Gupta and Milojevic 2011] faz uma análise da relação custo-benefício entre *cluster* tradicional e instâncias virtuais na ferramenta Eucalyptus. Neste trabalho, foi constatado que a computação em nuvem pode ser rentável e viável para determinadas aplicações, pois apresentou boa relação custo-benefício. Por outro lado, [Evangelinos and Hill 2008] não alcançou resultados satisfatórios executando aplicações paralelas em instâncias da Amazon, devido à baixa largura de banda e a alta latência na troca de mensagens.

Também, em provedores de IaaS públicos, [Strazdins et al. 2012] constatou que a virtualização tem efeitos diretos no desempenho e ficou evidente a necessidade de alta velocidade na comunicação entre os nodos. Neste cenário, [Navaux et al. 2012] concluiu que aplicações paralelas em nuvem tem ganho de desempenho e melhor custo-benefício em relação a tradicionais *clusters* (27% e 41% respectivamente). Mostrando assim, que é possível a nuvem prover uma plataforma viável apesar das desvantagens existentes no desenvolvimento, pois os provedores focam na entrega de IaaS sem se preocupar com o suporte das bibliotecas de programação paralela.

No cenários de virtualizadores por exemplo, [Regola and Ducon 2010] avaliou e comparou os virtualizadores (OpenVZ, Xen e KVM). Os resultados mostraram que as aplicações são economicamente viáveis e a sobrecarga de CPU é mínima tanto na paravirtualização, virtualização completa e na virtualização por *container*. No trabalho de [Xavier et al. 2013], foram testadas aplicações de alto desempenho em um ambiente com virtualização por *container* (OpenVZ, Vserver, LXC) e comparados com a ferramenta Xen Server (*Hypervisor*). A conclusão do trabalho é que este tipo de virtualização é uma opção atrativa para computação de alto desempenho, pois pode oferecer melhor compartilhamento de recursos e ambiente customizado.

Esta pesquisa contribui na comparação do desempenho das aplicações paralelas nas instâncias virtuais executadas em nuvem privada e gerenciadas por diferentes ferramentas de administração de nuvem, analisando se existem diferenças entre elas.

3. Resultados

A infraestrutura de testes foi baseada em dois *clusters* isolados, cada um formado por 4 computadores análogos ao ambiente nativo com processadores Intel Core i5 650 - 3.20 GHz, memória RAM de 4 GB DDR3 de 1333 MHz, disco de 500 GB *sata* II, em uma rede 10/100 *Mbits*. Nos computadores e nas instâncias virtuais foi utilizado o Ubuntu Server 12.04. Foram utilizadas ambientes baseados no virtualizador KVM com as ferramentas OpenStack e OpenNebula 4.7.80.

Para os testes utilizou-se o NPB-3.3 implementado em MPI (*Message-Passing Interface*) e OMP (*Open Multi-Processing*). Esta suíte de *benchmarks* é amplamente utilizada para testes de alto desempenho. Compilado com a classe B, os *kernels* EP, SP, IS e MG foram executados 40 vezes em cada ambiente, em ordem aleatória e contínua. Na Figura 1(a) são apresentados os resultados de *speed-up* e eficiência nos padrões MPI e OMP e na Figura 1(b) são apresentados os resultados estatísticos.

O uso do *Hyper Threading* nos processadores afeta as execuções nos padrões OMP a partir de 2 *threads* e MPI a partir de 9 processos. Consequência natural que não pode ser

resolvida com uso de virtualizadores ou de ferramentas de administração de nuvem. Além disso, os resultados do EP em MPI e OMP mostram que o desempenho deste *benchmark* é próximo ao Ideal. Isto devido ao algoritmo não exercer intensa comunicação entre os nodos e processos.

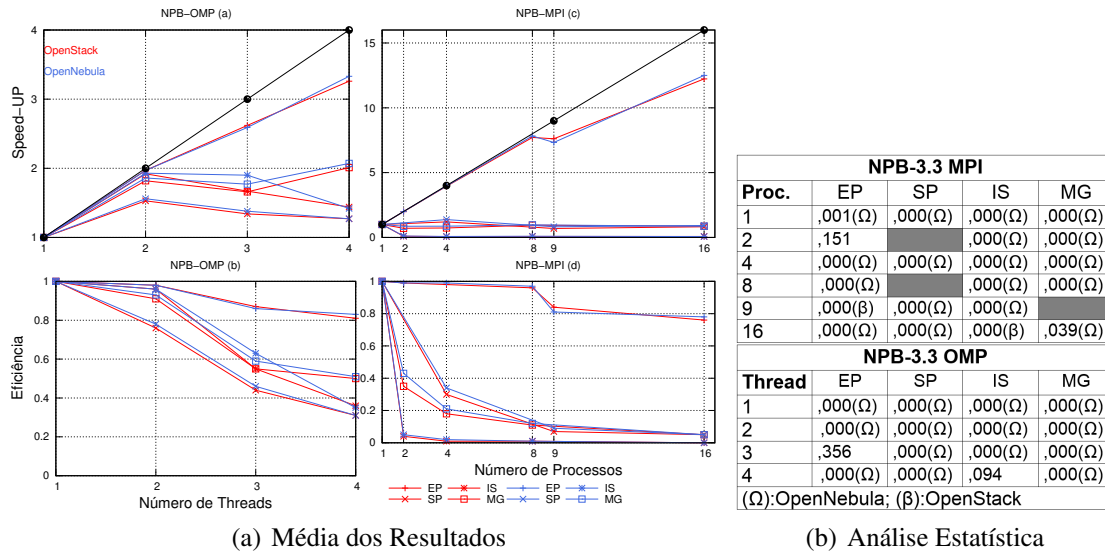


Figura 1. Resultados

OpenNebula alcançou melhores resultados com as execuções. A possível perda de desempenho do OpenStack na suíte OMP se deve ao componente *Nova-compute-KVM*, que na versão Havana, grava em tempo real as informações das instâncias (ex. métricas do uso dos recursos) em um banco de dados SQL, o qual também necessita do auxílio do *middleware* RabbitMQ para realizar esta tarefa. No MPI, acredita-se que *Nova-compute-KVM* também esteja influenciando o desempenho, mas o fator principal está no componente Neutron e o *plugin* OpenvSwitch, que cria interfaces virtuais e um tunelamento exclusivo para comunicação de alguns serviços, inserindo camadas adicionais na comunicação de rede. Enquanto o OpenNebula apenas usa um *driver* básico (*Dummy*), o qual não exerce regras específicas para o tráfego de rede.

Na Figura 1(b), identifica-se as diferenças e qual ferramenta é melhor em cada *benchmark*. Na tabela, também constam os valores da variável *Sig.* retornada pelo *software* SPSS [Field 2009], usando um nível de confiança de 95%. Os resultados mostram que a maior parte das execuções (MPI e OMP) são significativamente diferentes e favoráveis ao OpenNebula. No entanto, somente em OMP IS-4 e EP-3, e MPI EP-2 não é possível afirmar estatisticamente que existe diferença significativa entre as ferramentas.

4. Conclusão

Neste trabalho comparou-se o desempenho das aplicações paralelas nas ferramentas OpenStack e OpenNebula. Os testes estatísticos serviram para demonstrar que existem diferenças significativas entre as ferramentas. Em geral, as melhores médias obtidas foram nas instâncias do OpenNebula, mas nos testes MPI IS-16 e MPI EP-9 o OpenStack foi melhor. Além disso, a análise constatou que nas execuções OMP IS-4 e EP-3, e MPI EP-2 não foi possível definir qual é a melhor ferramenta. Efetuando um comparativo com os resultados de isolamento de recursos de [Maron et al. 2014b], nota-se que o desempenho do disco não foi determinante para as aplicações paralelas testadas neste trabalho. Isso porquê, estas aplicações demandam mais rede, processamento e memória, nos quais o OpenNebula demonstrou ser melhor em [Maron et al. 2014b].

Mesmo que OpenNebula teve melhores resultados que OpenStack, é necessário ressaltar que em termos de *speed-up* e eficiência, os resultados não foram satisfatórios. Principalmente, nas aplicações MPI que dependem de altas taxas de comunicação entre os nodos do *cluster*, conforme relatado também por [Strazdins et al. 2012]. Assim, em um ambiente com maior largura de banda, as aplicações paralelas poderão apresentar melhores resultados.

Os trabalhos futuros serão: (I) realizar os experimentos deste trabalho em uma rede *gigabit*; (II) comparar o desempenho de aplicações corporativas/reais nestas ferramentas; (III) comparar os resultados aqui obtidos com outras ferramentas de administração de nuvem;

Referências

- Buyya, R., Broberg, J., and Goscinski, A. (2010). *Cloud Computing: Principles and Paradigms*. Wiley.
- Evangelinos, C. and Hill, C. (2008). Cloud Computing for Parallel Scientific HPC Applications: Feasibility of Running Coupled Atmosphere-Ocean Climate Models on Amazon's EC2. *ratio*, 2(2.40):2–34.
- Field, A. (2009). *Discovering Statistics Using SPSS*. SAGE, Dubai, EAU.
- Gupta, A. and Milojicic, D. (2011). Evaluation of HPC Applications on Cloud. In *Open Cirrus Summit (OCS), 2011 Sixth*, pages 22–26. IEEE.
- Maron, C. A. F., Griebler, D., and Schepke, C. (2014a). Comparação das Ferramentas OpenNebula e OpenStack em Nuvem Composta de Estações de Trabalho. In *14th Escola Regional de Alto Desempenho do Estado do Rio Grande do Sul - ERAD/RS*, pages 173–176, Alegrete, RS, Brazil. Sociedade Brasileira de Computação - SBC.
- Maron, C. A. F., Griebler, D., Vogel, A., and Schepke, C. (2014b). Avaliação e Comparação do Desempenho das Ferramentas OpenStack e OpenNebula. In *12th Escola Regional de Redes de Computadores (ERRC)*, Canoas. Sociedade Brasileira de Computação.
- Navaux, P., Roloff, E., Diener, M., and Carissimi, A. (2012). High Performance Computing in the Cloud: Deployment, Performance and Cost Efficiency. In *Proceedings of the 2012 IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom)*, pages 371–378, Washington, DC, USA. IEEE Computer Society.
- Regola, N. and Ducon, J. C. (2010). Recommendations for Virtualization Technologies in High Performance Computing. pages 409 – 416, Indianapolis. IEEE.
- Strazdins, P. E., Cai, J., Atif, M., and Antony, J. (2012). Scientific Application Performance on HPC, Private and Public Cloud Resources: A Case Study Using Climate, Cardiac Model Codes and the NPB Benchmark Suite. In *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International*, pages 1416–1424. IEEE.
- Thome, B., Hentges, E., and Griebler, D. (2013). Computação em Nuvem: Análise Comparativa de Ferramentas Open Source para IaaS. In *Escola Regional de Redes de Computadores (ERRC)*, pages 1–4, Porto Alegre, RS, Brazil. Sociedade Brasileira de Computação - SBC.
- Xavier, M., Neves, M., Rossi, F., Ferreto, T., Lange, T., and Rose, C. D. (2013). Performance Evaluation of Container-based Virtualization for High Performance Computing Environments. In *21st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, pages 233–240, Belfast, UK. IEEE.