Modelagem e Avaliação de Comportamento de Aplicações do *Tipo Bag of Tasks* em uma Nuvem Gerida pelo *OpenStack*

Fernando Angelin, Maicon Ança dos Santos,
Vilnei Marins de Freitas das Neves, Gerson Geraldo H. Cavalheiro

[1]Bacharelado em Ciência da Computação
Centro de Desenvolvimento Tecnológico – Universidade Federal de Pelotas (UFPel)
Campus Porto – Rua Gomes Carneiro, 1 – 96010-610 – Pelotas – RS – Brasil

{fangelin, madsantos, vilnei, gerson.cavalheiro}@inf.ufpel.edu.br

Resumo. *O cenário neste trabalho consiste em uma nuvem que oferece serviço de processamento para computação intensiva. As aplicações do tipo Bag of Tasks (BoT) podem tirar vantagem desse cenário. Propôs-se uma modelagem que será validada por meio de um novo componente do OpenStack, que é o Neat, que realiza a consolidação dinâmica de máquinas virtuais utilizando a migração em tempo real, alcançando economia energética para o provedor deste serviço.*

1. Introdução

A computação na nuvem revolucionou a indústria de Tecnologia da Informação e Comunicação possibilitando o provisionamento sob demanda de recursos computacionais elásticos com uma política pague pelo que utilizar [Beloglazov 2013]. Desta forma uma organização pode terceirizar suas necessidades computacionais para a nuvem e assim, evitar investimentos em infraestrutura computacional privada. A computação em nuvem vem crescendo nos últimos anos, oferecendo acesso a computadores de uma forma eficiente e remota. Neste tipo de modelo computacional, o uso dos recursos pode ser otimizado dinamicamente conforme as demandas de seus usuários. As mudanças na alocação dos recursos podem ser feitas em tempo real permitindo assim sistemas mais eficientes, evitando o super ou subdimensionamento dos recursos [Armbrust et al. 2010].

Muitas aplicações do nosso cotidiano necessitam de altas cargas de processamento. Porém existem certos períodos de tempo em que todo este processamento não está sendo realizado. Caso um usuário conte com recursos de processamento privado, é provável que se observe uma subutilização de seus recursos. Preocupados em incluir nos seus ambientes soluções que visam a diminuição dessa ociosidade de processamento, e o consequente gasto financeiro, os administradores de redes tem utilizado a técnica de virtualização. O uso da virtualização representa a ilusão de várias máquinas virtuais independentes, cada uma rodando uma instância de um sistema operacional virtualizado [Smith and Nair 2005].

O cenário proposto para este trabalho é composto por uma nuvem oferecendo o serviço de processamento para computação intensiva, atendendo aplicações que necessitem altas demandas de processamento durante a execução. Muitas aplicações podem usufruir deste cenário, dentre elas estão as aplicações do tipo *Bag of Tasks (BoT)*. Estas aplicações são compostas por uma coleção de tarefas independentes entre si, podendo ser executadas em qualquer ordem. Para isso, neste trabalho é proposto um modelo para este tipo de aplicação para que seja possível avaliar sua execução em uma nuvem computacional.

Daremos preferência a uma nuvem gerida pelo *OpenStack*, entendido como um sistema operacional que visa gerenciar uma nuvem computacional de forma simplificada e eficiente [OpenStack 2014].

Para a validação desta modelagem foi utilizado um simulador de cargas (aplicação que efetua uma série de cálculos gerando cargas sintéticas de processamento no CPU) que foi desenvolvido para a validação do *Neat*. Este componente também é utilizado, permitindo que o *OpenStack*, realize a consolidação dinâmica de máquinas virtuais utilizando a migração destas máquinas virtuais em tempo real.

Diferentes estratégias de escalonamento podem ser usadas para diferentes tipos de perfis de usuários e, neste caso, cabe avaliar o impacto de cada estratégia de escalonamento escolhida e aplicada às máquinas virtuais para essa classe de aplicações.

2. Desenvolvimento

O *OpenStack* é entendido como um sistema operacional para gerenciamento de forma simplificada de uma nuvem. Com ele é possível controlar grandes centros de computação, armazenamento e recursos de rede em um *datacenter* e, esta ferramenta conta com um painel administrativo, de onde são controlados os recursos que serão provisionados aos seus usuários, por meio de uma interface *web* [OpenStack 2014]. O projeto *OpenStack* é *open source*, ou seja, várias pessoas/desenvolvedores ao redor do mundo contribuem na sua construção. Seu principal objetivo é prover escalabilidade da disponibilidade dos recursos de hardware às aplicações submetidas. Assim, o *OpenStack* visa fornecer computação em nuvem de forma elástica tanto para nuvens públicas quanto para nuvens privadas, independentemente também de seu tamanho, com uma implementação simples e possuir um rico conjunto de recursos, chamados de componentes. Uma nuvem gerida pelo *OpenStack* provê uma solução de *IAAS - Infrastructure as a Service* por meio de uma variedade de serviços complementares. Cada serviço oferece uma *API* - Application Programming Interface que facilita esta integração.

Os principais componentes são: *Nova*, *Swift*, *Cinder*, *Glance*, *Keystone*, *Horizon*, *Neutron*, *Ceilomenter*, *Heat*, *Trove*, *Sahara*, etc [OpenStack 2014]. Ainda pode-se instalar outros componentes, conforme a demanda do usuário, como por exemplo o *Neat*, que provê migração em tempo real de máquinas virtuais dentro de clusters, para fins de reduzir o consumo energético dos mesmos [Neat 2014].

Neat é uma extensão para o *OpenStack* que implementa a consolidação dinâmica de maquinas virtuais utilizando a migração em tempo real. O principal objetivo da consolidação dinâmica de máquinas virtuais é melhorar a utilização dos recursos físicos e reduzir o consumo de energia por re-alocação de máquinas virtuais usando a migração em tempo real de acordo com sua demanda de recuros atual e mudar o estado dos servidores *hosts* de ociosos para o estado de suspensão [Neat 2014].

Aplicações do tipo *Bag of Tasks* são descritas por um conjunto de tarefas independentes. Não raro, estas tarefas possuem, isoladamente, alto custo computacional. Para evitar grandes investimentos em recursos computacionais, usuários podem submeter à nuvem, de forma oportunista, suas demandas de processamento para este tipo de aplicação.

Bag of Tasks é uma estratégia de processamento paralelo em que as tarefas são colocadas em um saco de tarefas que é compartilhado pelos processadores clientes, chama-

dos de *workers*. Cada *worker*, repetitivamente, pega uma tarefa e executa-a. Possivelmente, essa execução gere novas tarefas que são colocadas dentro do mesmo saco de tarefas. A computação encerra quando o saco de tarefas estiver vazio e os *workers* estiverem ociosos[Andrews 1999].

A modelagem de um *Bag of Tasks* proposta é como segue: o *Bag of Tasks* é descrito por n triplas, onde cada valor, na sequência, representa: número de tarefas, tempo, em segundos, que cada tarefa ocupa no CPU e, a porcentagem de CPU que cada tarefa ocupa. Um exemplo de *Bag of Tasks*:$\{[10, 30, 30], [20, 40, 90], \ldots, [55, 7, 10], [12, 140, 95]\}$.

Neste trabalho, foi utilizado o algoritmo de detecção de sobrecarga LRR – *Local Regression Robust* [Beloglazov 2013] para a verificação de sobrecarga dos *hosts* feita pelo *Neat*. Este algoritmo calcula as estimativas dos parâmetros de *Loess* e, usa-as para prever o futuro de utilização de CPU no próximo passo de tempo, tendo em conta o tempo de migração da máquina virtual. Além disso, este algoritmo aceita um parâmetro de segurança, que é usado para dimensionar a utilização da CPU prevista para aumentar ou diminuir a sensibilidade do algoritmo para potenciais sobrecargas.

Para causar estresse nas máquinas virtuais, foi utilizado um *script* que gera *Lookbusy*, ou seja, ele aplica cargas sintéticas no CPU, como se fosse um trabalho real. O servidor controle lê o *Bag of Tasks* descrito e passa a distribuir para as máquinas virtuais, hospedadas nos nós físicos.

3. Metodologia

Os experimentos foram realizados em uma plataforma contendo cinco nós com a seguinte configuração: Processador Intel(R) Xeon(R) CPU E5-2420 1.90GHz (6 cores, 12 threads) e 8GB RAM. Destes nós, um é dedicado a realizar as operações de controle do *OpenStack* (controler) e os demais a hospedar as máquinas virtuais. Os experimentos realizaram a monitoração do consumo energético e a migração de máquinas virtuais na execução de uma aplicação *BoT* descrita anteriormente, gerada visando o alto consumo de processamento. Os valores descritos em cada tripla (de tarefas homogêneas), se refere: a quantidade de tarefas, o tempo (em segundos) que cada tarefa demanda de CPU e, a porcentagem de utilização de CPU. Esta descrição é compactada de forma a agrupar as tarefas que maximizem o uso das CPUs disponíveis. No estudo de caso realizado a aplicação é composta de 2700 tarefas que consumiriam no total cerca de 60 horas em uma execução sequencial monoprocessada.

Foram realizados experimentos considerando a existência de 10, 20 e 40 máquinas virtuais e duas estratégias de detecção de sobrecarga, uma sem controle, não realizando migração de máquinas virtuais, outra migrando máquinas virtuais quando algum nó real é identificado como sobrecarregado por meio do LRR. Foram realizadas três rodadas com cada configuração. Os resultados estão sumarizados na Tabela 1. Não utilizando a detecção de sobrecarga, podemos perceber que não ocorreram migrações, diferentemente de quando utilizado o LRR. A economia de energia se refere a quantidade de energia que, na totalidade, os nós deixaram de gastar por se utilizarem do *Neat* com relação a energia que seria gasta na situação em que os nós não usem o referido componente.

Tabela 1. Relação dos resultados obtidos

Sem detecção de sobrecarga						
Experimento	10 MVs		20 MVs		40 MVs	
	Migrações	Economia (%)	Migrações	Economia (%)	Migrações	Economia (%)
1	0	56,50	0	43,91	0	51,71
2	0	56,87	0	44,20	0	54,66
3	0	56,52	0	43,79	0	51,49
Detecção de sobrecarga - LRR						
Experimento	10 MVs		20 MVs		40 MVs	
	Migrações	Economia (%)	Migrações	Economia (%)	Migrações	Economia (%)
1	6	55,58	39	47,33	74	50,05
2	18	54,64	39	51,75	76	49,90
3	14	53,28	39	50,61	74	50,21

4. Conclusão

No estudo realizado foi possível perceber que diferentes estratégias de detecção de sobrecarga implicam diferentes padrões de migração de máquinas virtuais. Pode-se perceber economia de energia quando não se usa detecção de sobrecarga. Isso acontece pelo fato de o próprio *OpenStack* realizar um controle sobre os nós computacionais ociosos. Também, não utilizando deteção de sobrecarga e com mais de 20 máquinas virtuais, os resultados apresentam economia de energia, porém, as aplicações presentes no *Bag of Tasks* não são executadas na sua totalidade.

Na sequência, o estudo de outras estratégias de detecção de sobrecarga serão realizados, bem como avaliados o padrão de execução considerando diferentes configurações de aplicações *Bag of Tasks*.

Como trabalho futuro, a expectativa é desenvolver uma estratégia de detecção de sobrecarga eficiente para aplicações *Bag of Tasks* sobre uma nuvem gerida pelo *OpenStack*.

Referências

Andrews, G. R. (1999). *Foundations of Multithreaded, Parallel, and Distributed Programming*. PEARSON, Tucson, AZ, USA.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. (2010). A view of cloud computing. *Commun. ACM*, 53(4):50–58.

Beloglazov, A. (2013). *Energy-Efficient Management of Virtual Machines in Data Centers for Cloud Computing*. PhD thesis, THE UNIVERSITY OF MELBOURNE, Melbourne-Australia.

Neat, O. (2014). Openstack neat. `http://openstack-neat.org/`. Acessado em 21 de Junho de 2014, traduzido.

OpenStack (2014). Openstack. `https://www.openstack.org/`. Acessado em 20 de Junho de 2014, traduzido.

Smith, J. E. and Nair, R. (2005). The architecture of virtual machines. *Computer*, 38(5):32–38.