Heurísticas para Predição de Configurações de Custo Mínimo para Execução de Aplicações em Ambientes de Nuvem de Infraestrutura

O processo de decisão pela migração de aplicações para o ambiente de nuvens computacionais envolve uma série de análises que buscam, entre outras coisas, identificar que vantagens a mudança trará de fato. Os principais benefícios buscados no ambiente de nuvem são a melhoria de desempenho, redução/otimização de custo e facilidades de gerenciamento [LI *et. al.*, 2011; RODERO-MERINO *et. al.*, 2010]. Entre esses benefícios, o custo apresenta-se como o mais difícil de prever, uma vez que depende necessariamente do tamanho da demanda exercida sobre a aplicação além do desempenho oferecido e preços cobrados pelo provedor de nuvem de infraestrutura contratado [CUNHA et. al., 2012].

Uma maneira de tentar prever o desempenho da aplicação implantada num provedor de nuvem de infraestrutura é aplicar uma carga de utilização sobre uma determinada arquitetura composta de uma ou mais configurações de máquinas. A partir daí, variações na carga, na arquitetura e nas configurações das máquinas oferecerão resultados diferentes. Cabe então ao dono da aplicação estabelecer o padrão mínimo de resposta para que o desempenho da aplicação seja considerado satisfatório.

Porém, a própria fase de ensaios pode atingir patamares elevados de custo, a depender das necessidades de variação da demanda, da arquitetura de implantação e das configurações utilizadas em cada arquitetura implantada [SILVA et. al., 2013].

-- Formalização

Seja **W** um conjunto de cargas de trabalho, onde **w1, w2, w3... w|W|** são valores que determinam o tamanho da carga de trabalho.

Seja **S** um conjunto de cenários, onde **s1, s2, s3...s|S|** são identificadores de cenários de implantação de uma aplicação.

Seja **M** um conjunto de métricas, onde **m1, m2, m3...m|M|** são identificadores de métricas usadas para avaliar o desempenho de uma aplicação. Métricas podem ser *minimizáveis* ou *maximizáveis*, a depender do objetivo da métrica quanto ao resultado desejado. Por exemplo, tempo de resposta é uma métrica minimizável, uma vez que geralmente se deseja o menor tempo possível nos resultados. Contrariamente, quadros por segundo, em um sistema de computação gráfica, é uma métrica maximizável.

Seja **E** um conjunto de execuções, onde os elementos **e1, e2, e3...e|E|** são quádruplas no formato **{s, w, m, r}** com **s S**, **w W**, **m M**, e onde **r** é o resultado de uma execução para a métrica **m**, no cenário **s** sob a carga de trabalho **w**.

Seja um valor de referência definido como parâmetro de sucesso da execução de um teste.

Seja **atende(e, alfa),** onde **e** **E**, uma função tal que:

* se a métrica **me** é minimizável
* se a métrica **me** é maximizável

Seja **P** a matriz de desempenho definida como:

Cada elemento da matriz é definido como:

onde || é a contagem de execuções para o i-ésimo cenário S, a j-ésima carga de trabalho W e a n-ésima métrica M.

Assim, temos que a matriz de desempenho P informa o percentual de execuções de sucesso considerando todas as métricas avaliadas para cada cenário sob cada carga de trabalho.

Bibliografia

CUNHA, M.; MENDONÇA, N.; SAMPAIO, A. Um ambiente programável para avaliar o desempenho de aplicações em nuvens de infraestrutura. Dissertação de Mestrado. Universidade de Fortaleza, 2012.

LI, A.; ZONG, X.; ZHANG, M.; KANDULA, S.; YANG, X. CloudProphet: Predicting Web Application Performance in the Cloud. Duke University Technical Report CS-2011-11, 2011.

RODERO-MERINO, L.; VAQUERO, L. M.; GIL, V.; GALÁN, F.; FONTÁN, J.; MONTERO, R. S.; LLORENTE, I. From Infrastructure delivery to service management in clouds. In: Future Generation Computer Systems, 2010, p. 1226 – 1240.

SILVA, M. A.; HINES, M. R.; GALLO, D.; LIU, Q.; RYU, K. D.; SILVA, D. CloudBench: Experiment Automation for Cloud Environments. IEEE International Conference on Cloud Engineering (IC2E), 2013.