

# Distribuição de carga em execução de scripts no sistema VisionDataset

André Spanguero Kanayama  
ask@ime.usp.br

Supervisor: Roberto Hirata Jr.  
hirata@ime.usp.br



## O Problema

Os processamentos e análises de imagens consomem muitos recursos de memória e de CPU. Em particular, em sistemas colaborativos desse tipo, com essa finalidade, é desejável que os processamentos ocorram fora do servidor para não sobrecarregá-lo, mas próximos a ele para não sobrecarregar os sistemas de comunicação. Neste trabalho implementamos um sistema para a execução e edição de scripts em Python dentro do VisionDataset, sendo que tal execução deve ser feita de forma eficiente e não pode afetar o serviço para outros usuários. Com esta implementação desejamos que seja possível gerenciar e executar scripts dentro do sistema, com bastante liberdade, sem prejudicar o sistema como um todo.

## VisionDataset

O VisionDataset é um sistema colaborativo feito em Java e utiliza o sistema gerenciador de banco de dados PostgreSQL, e o framework Apache Struts. Seu objetivo é centralizar no mesmo local as imagens dos usuários facilitando assim a colaboração científica. Com ele é possível criar e gerenciar usuários e álbuns de imagens, fazer o upload de arquivos em massa e criar anotações e tags em imagens. Após este trabalho também será possível executar scripts sobre as imagens e obter os resultados.



## Máquinas Virtuais

O uso de máquinas virtuais para realizar o processamento faz com que tenhamos várias vantagens em relação à simples execução do script no próprio servidor, como:

- Faz com que não seja possível derrubar o servidor com um script;
- Facilita o controle de recursos que desejamos utilizar para realizar o processamento;
- Podemos ter vários sistemas operacionais no mesmo *hardware*. Assim conseguimos comparar qual se mostra mais eficiente.

## Balanceamento de carga

Não basta apenas distribuir o processamento das imagens de forma arbitrária, é preciso achar formas eficientes de balancear a carga de execução e utilizar todas as máquinas disponíveis da melhor maneira possível. Esté sendo realizado um estudo sobre maneiras de evitar o uso excessivo de memória, que se mostrou o grande limitante no desempenho do processamento. Além disso temos que evitar que algumas máquinas trabalhem demais, enquanto outras fiquem ociosas por muito tempo.

## Referências

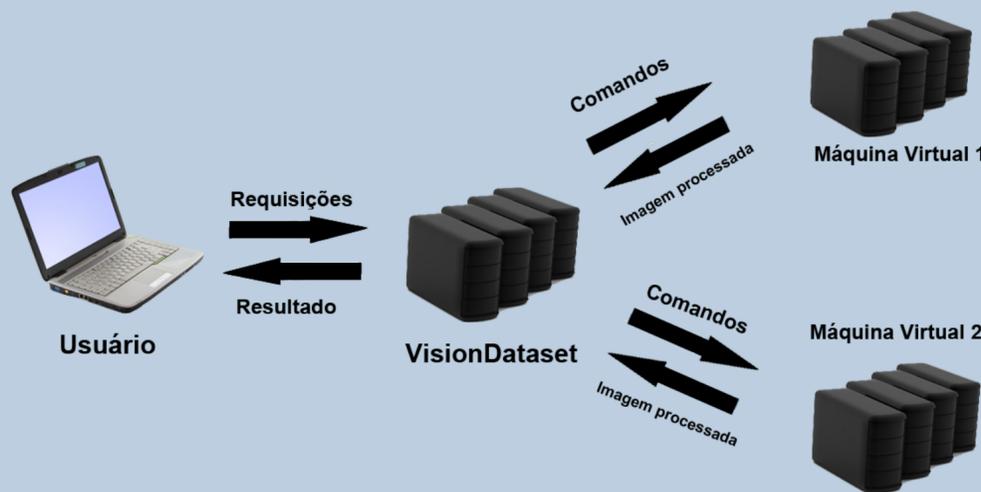
- [1] KLAVA, B. *Base de imagens do grupo de Visão Computacional do IME/USP*. 2010. 3 p. - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.
- [2] MATSUMOTO, F. T. *Ferramentas semelhantes ao VisionDataset*. 2012. 6 p. - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.
- [3] *MÁQUINA VIRTUAL*. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2013. Disponível em: <[http://pt.wikipedia.org/w/index.php?title=M%C3%A1quina\\_virtual&oldid=37155482](http://pt.wikipedia.org/w/index.php?title=M%C3%A1quina_virtual&oldid=37155482)>. Acesso em: 5 nov. 2013.

As imagens foram uma cortesia de SweetCrisis, Keerati e renjith krishnan / FreeDigitalPhotos.net

## Solução Proposta

A solução encontrada para que o usuário possa executar seus scripts com liberdade, sem o risco de derrubar o servidor, foi delegar o processamento sobre as imagens para outras máquinas, inicialmente máquinas virtuais. Sendo assim, mesmo que o usuário faça um processamento muito pesado sobre as imagens e que demore muito, o funcionamento do VisionDataset não é afetado, nem mesmo se a máquina virtual onde o processamento é feito for derrubada. O banco de dados e as imagens ficam no servidor, sendo que as máquinas virtuais conseguem carregar as imagens e os scripts através de um programa chamado *sshfs*. O processo ocorre da seguinte maneira:

- Os usuários requisitam a execução de um script sobre as imagens para o VisionDataset, escolhendo as opções desejadas;
- Um script gerenciador entra em ação e decide onde este processamento deve ser feito;
- As máquinas virtuais rodam um *daemon*, que fica escutando em um socket, assim recebem a mensagem do servidor dizendo o que deve ser feito;
- O processamento é realizado e o resultado é gravado no servidor através do *sshfs*.



## Resultados

Com os testes realizados até o momento fica claro que a execução de scripts é muito mais eficiente quando se utiliza duas ou mais máquinas virtuais em relação à execução no próprio servidor do VisionDataset. Observando o uso das máquinas e sua relação com o tempo chega-se à conclusão de que o limitante principal é a memória. Quando a execução de scripts começa a utilizar o *swap*, o processo torna-se muito lento. Em seguida podemos ver os gráficos que ilustram os resultados de um teste de processamento de 14 imagens de tamanhos variando de 1 a 10 megapixels.

