

Escalando nossas instâncias do EC2

Transcrição

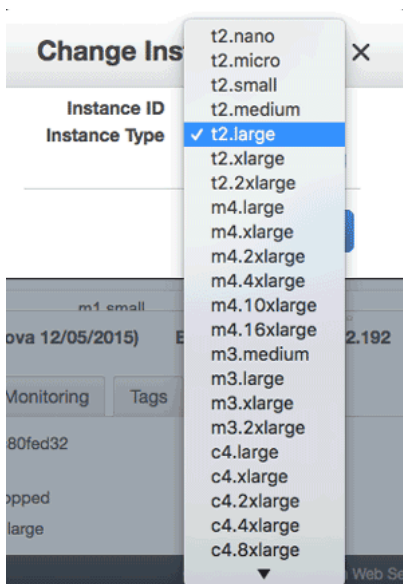
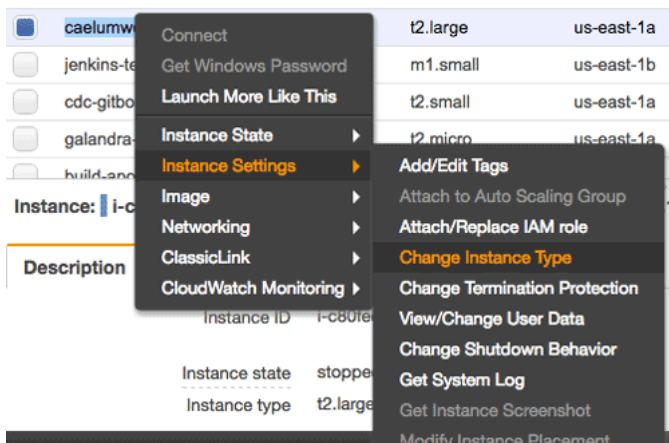
No capítulo anterior terminamos as configurações no EC2 para usar o AWS RDS. O nosso banco de dados está funcionando agora em uma máquina separada só para ele, num ambiente escalável e bem configurado pela Amazon. Como vimos, isso é um exemplo do SaaS (Software as a Service), pois não estamos mais responsáveis pela instalação do banco, apenas configuramos pela interface web.

Mais ainda tem uma questão referente a escalabilidade: Será que nosso `catalogo-teste` vai sempre aguentar qualquer número de requisições? Quando a demanda da nossa aplicação cresce será que essa máquina consegue ainda atender bem os nossos clientes?

Falando sobre escalabilidade, sempre há duas formas de escalar: vertical e horizontal

Escalabilidade vertical ou horizontal?

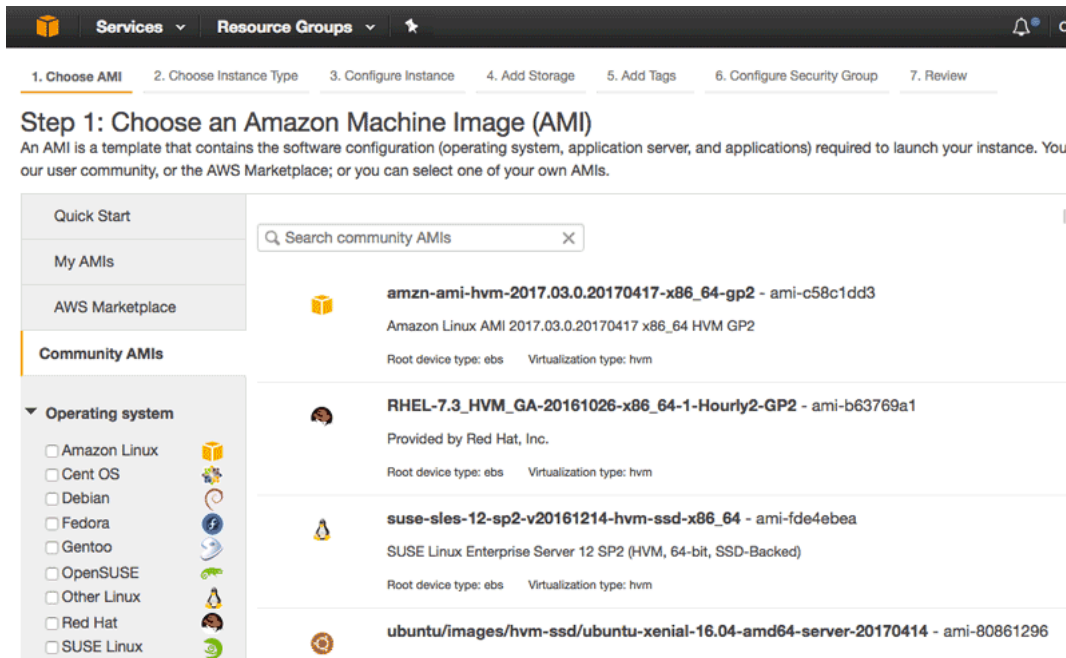
Vertical significa melhorar a máquina em si, por exemplo a CPU, Memória RAM ou HD. Isso é **limitado** e também pode ficar muito **caro**. Escalabilidade vertical é sempre limitado mas também é o mais simples, pois basta melhorar o Hardware da máquina. Isso é feito no Amazon através do *instance type*:



Para trocar o tipo da instância a máquina deve estar desligada e, novamente, quando gravamos o curso apenas o tipo `t2.micro` é gratuito.

Imagem com AMI

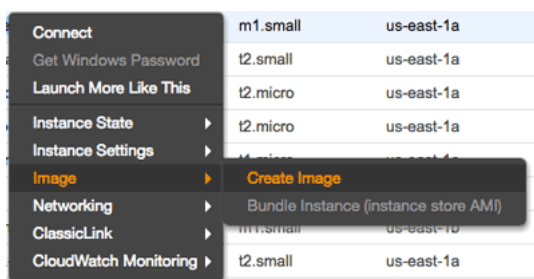
No nosso caso queremos continuar com a faixa gratuita da Amazon EC2 e por isso escalaremos a nossa aplicação horizontalmente. Isso significa que poderíamos criar uma segunda instância como já aprendemos. No entanto, não basta só criar a instância, também é preciso instalar todo o software nela, no nosso caso Apache Tomcat junto com a aplicação. Ideal seria criar uma nova instância a partir dessa instalação já feita. Isso é possível no Amazon AWS através das imagens que inclusive já usamos no curso. Essas imagens tem o nome de **Amazon Machine Instance** (AMI) e já existem vários disponíveis vimos elas no primeiro passo da criação da máquina:



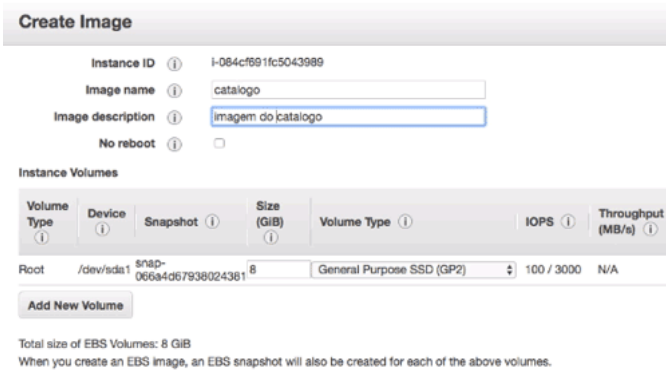
A diferença é que criaremos a nossa própria imagem (ou template) baseado na instância que já configuramos!

Criando imagem

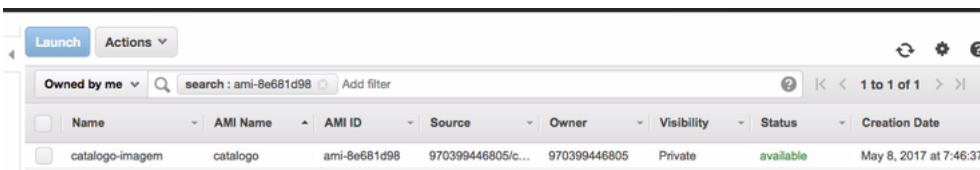
Para criar uma nova imagem basta selecionar a uma instância e clicar no `Image` - `>` `Create Image`



No formulário preenchemos o nome e descrição e basta confirmar para criar a imagem.



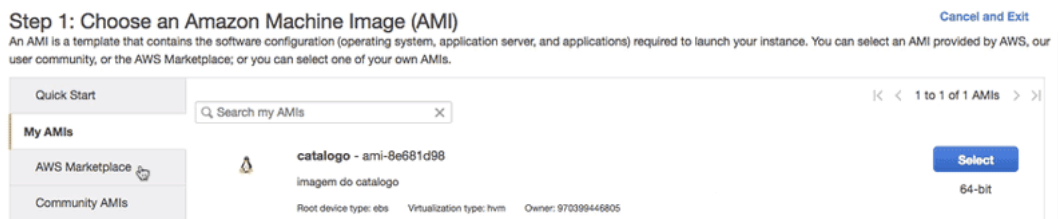
A criação pode demorar alguns minutos e depois disso a nova imagem deve estar disponível no menu *IMAGES* -> *AMIs*. Com a nossa imagem disponível podemos iniciar uma nova instância a partir dela. Só vamos renomear para "catalogo-imagem":



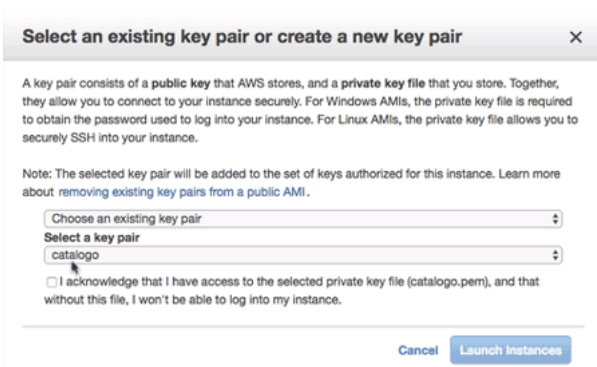
Name	AMI Name	AMI ID	Source	Owner	Visibility	Status	Creation Date
catalogo-imagem	catalogo	ami-8e681d98	970399446805/c...	970399446805	Private	available	May 8, 2017 at 7:46:37

Nova instância

Agora vamos realmente criar uma nova instância a partir da nossa. No menu *INSTANCES* -> *Instances* vamos clicar no botão *Launch instance* e escolher a nossa imagem (menu *My AMIs*):



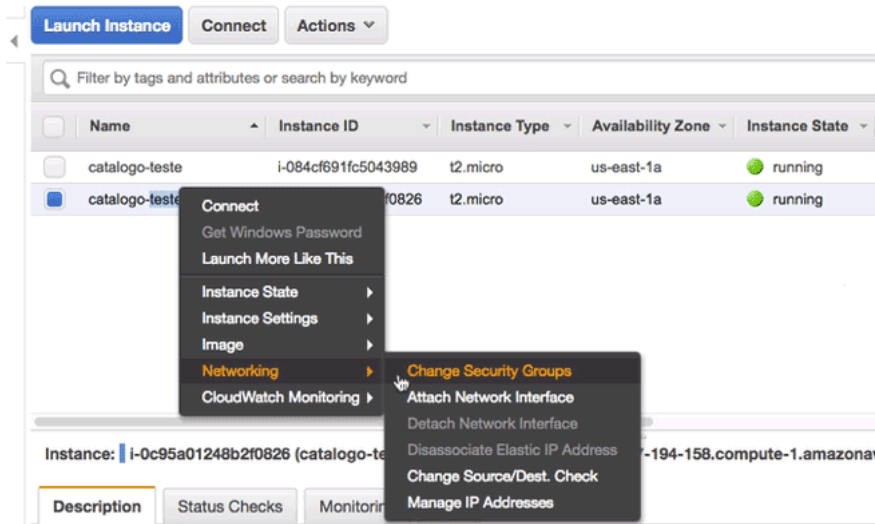
Como já vimos anteriormente, o tipo da instância deve ser novamente aquele que faz parte da faixa gratuita da Amazon (*Free Tier*), o `t2.micro`. Todos os outros passos que já aprendemos. Importante é que você escolhe o mesmo `keypair` que já definimos antes (claro, pode mudar tbm). Dessa forma podemos usar o mesmo arquivo `catalogo.pem` para ambas as máquinas:



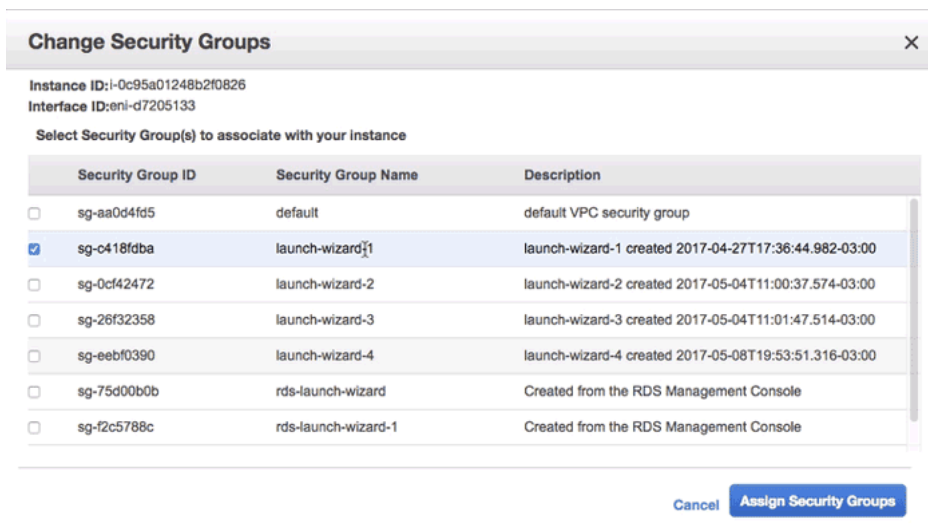
Ajustando Security Group

A instância já está configurado e rodando mas ainda não temos acesso na parte `8080` (e o servidor Tomcat usa por padrão essa porta). A configuração das portas não é da imagem e sim um serviço externo da Amazon e por isso devemos liberar

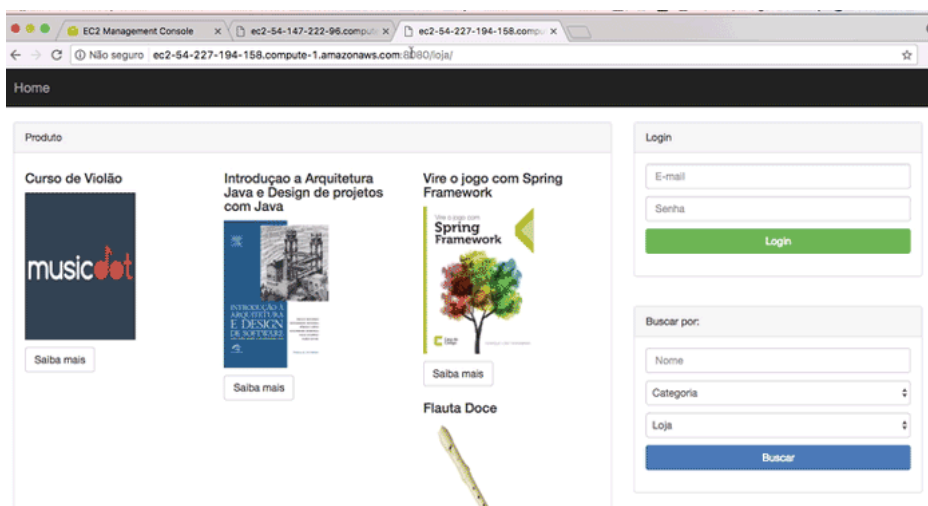
ainda a porta. Para simplificar e unificar a configuração vamos usar a mesma *Security Group* já configurada da outra instância. Para selecionar a nova instância e clique no botão direito *Instance Settings* -> *Change Security Group*:



Na nova janela desabilita o grupo atual e escolhe o grupo da primeira máquina configurada (no nosso caso `launch-wizard1`):



Uma vez configurado o mesmo primeiro *security group* para ambos as máquinas podemos testar o acesso direto à segunda instância.



Agora já podemos testar e qualquer alteração dos dados deve ser visível em ambas as aplicações. Isto é pois as duas instâncias estão usando o mesmo banco de dados.