



Geek University

Evolua seu lado geek!

www.geekuniversity.com.br

Introdução ao Auto Scaling da AWS



Introdução ao Auto Scaling da AWS



Já conhecemos o ELB que distribui a carga para nossos servidores.

Mas o que é 'escalabilidade'?

Quando falamos em 'escalar' uma aplicação (ou serviço) estamos falando em aumentar a capacidade de suporte à carga.

Em outras palavras, imaginando uma instância EC2 que está executando um servidor web, escalar este serviço seria aumentar o número de instâncias para que a nossa aplicação suporte a carga de milhares ou milhões de acesso.

Vimos na aula passada como implementar um balanceador de cargas, mas não criamos uma segunda instância da nossa EC2 para receber parte da carga.

Poderíamos fazer isso manualmente de forma fácil, mas o Auto Scaling da AWS faz isso de forma melhor pra gente.

Introdução ao Auto Scaling da AWS



Tipos de Auto Scaling

Scaling up/down: Adiciona/remove recursos adicionais como memória, espaço em disco, de acordo com a necessidade;

Scaling in/out: Adiciona/remove instâncias adicionais, de acordo com a necessidade;

O **Auto Scaling** da **AWS** usa o **in/out**.

Introdução ao Auto Scaling da AWS



A grande magia de uso do **ELB** + **Auto Scaling** é que nossa aplicação pode ser escalada para cima ou para baixo (elástica?) de forma automática.

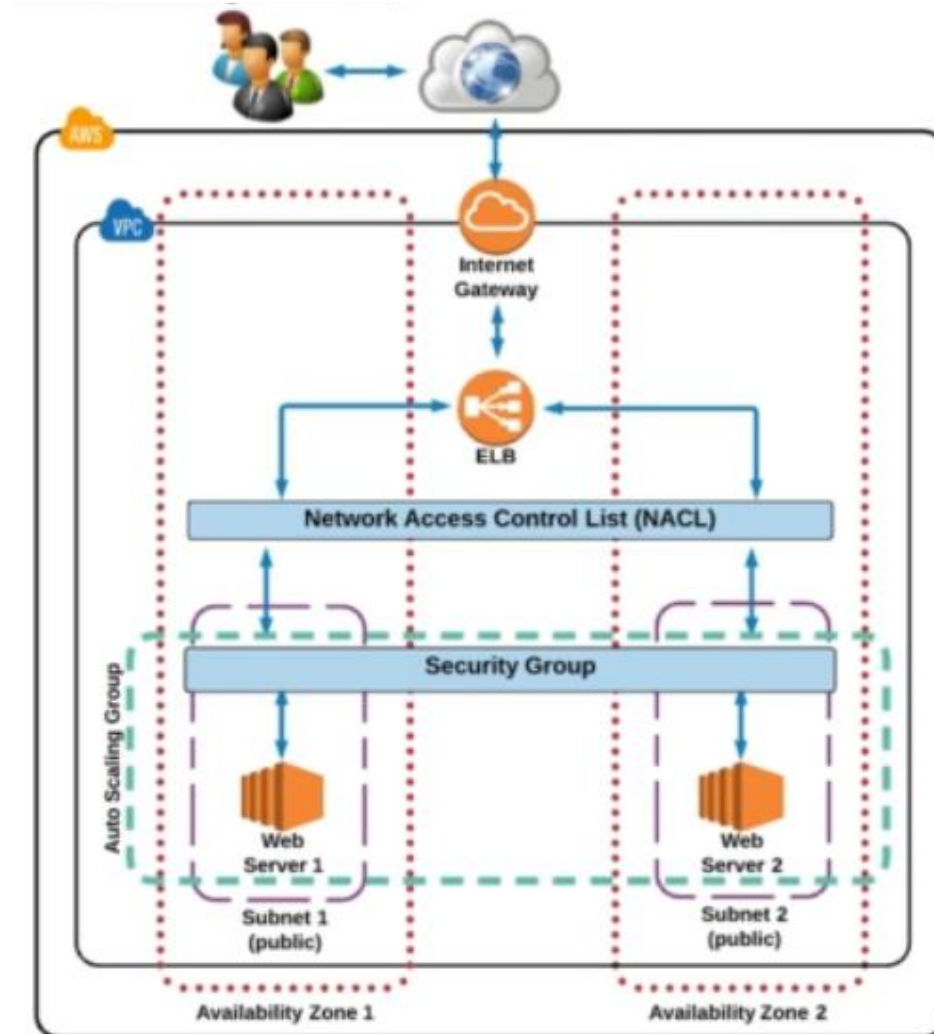
Lembra do caso do e-commerce que conversamos no início deste curso?

Imaginando que um e-commerce tenha vendas durante todo o ano mas em determinadas datas o número de acessos cresce muito, o uso do **ELB** + **Auto Scaling** iria fazer com que durante os dias normais a capacidade do servidor fosse a mínima suficiente e a medida que fosse demandada mais acesso o auto scaling iria 'escalar' a aplicação fazendo voltar ao normal quando a agitação acabasse.

Introdução ao Auto Scaling da AWS



Com o Auto Scaling criamos um Auto Scaling Group, que irá agrupar as instâncias na qual o serviço de auto scaling será ativado.

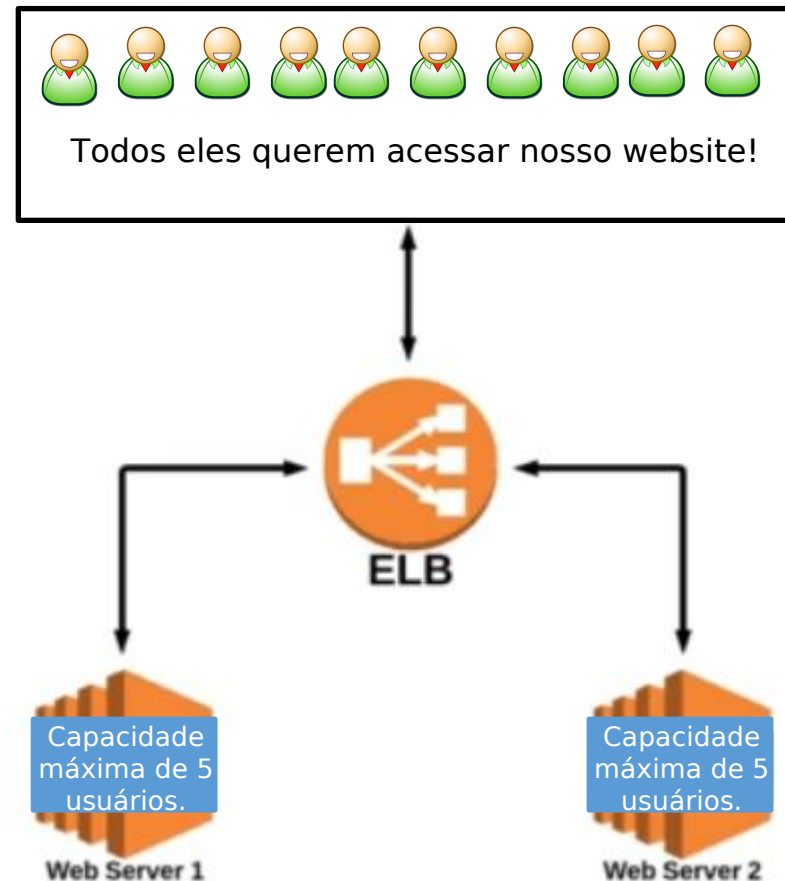


Introdução ao Auto Scaling da AWS



Como funciona o Auto Scaling?

Nosso balanceador de carga irá receber as requisições de acesso ao nosso website e distribuir os acessos igualmente entre nossos servidores web.

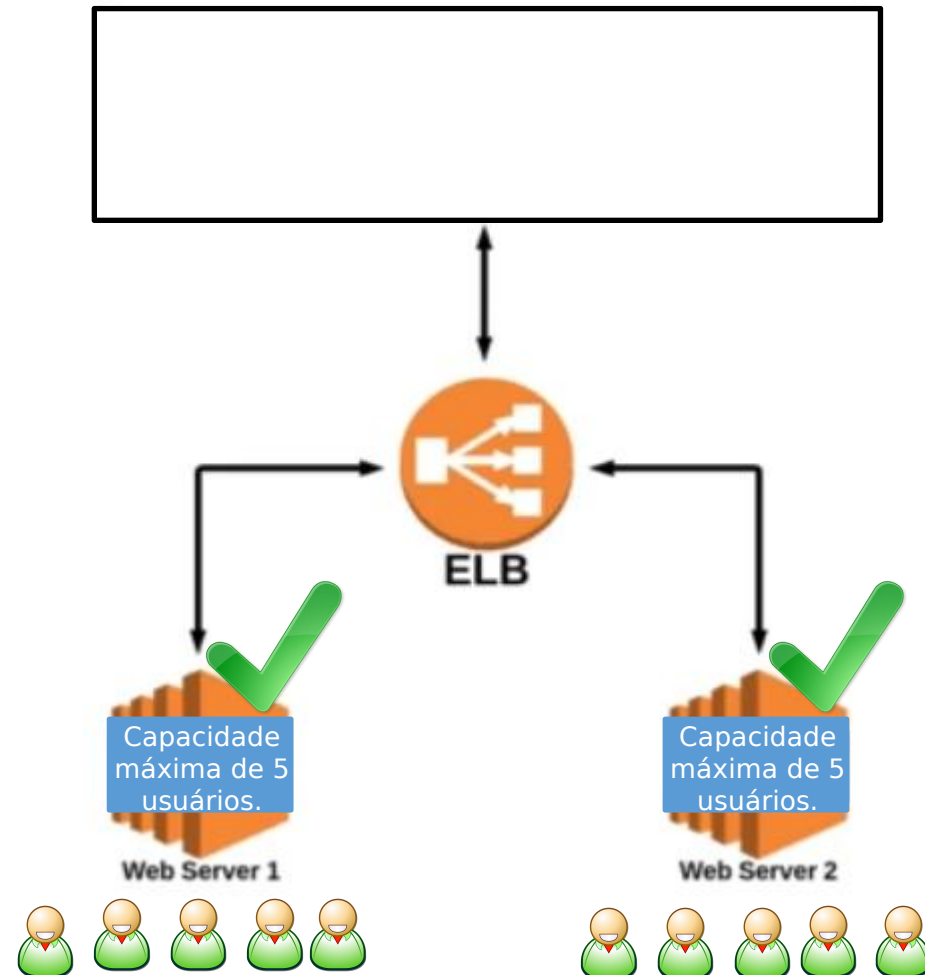


Introdução ao Auto Scaling da AWS



Como funciona o Auto Scaling?

O balanceamento de carga previne que um servidor seja sobrecarregado em detrimento ao outro.

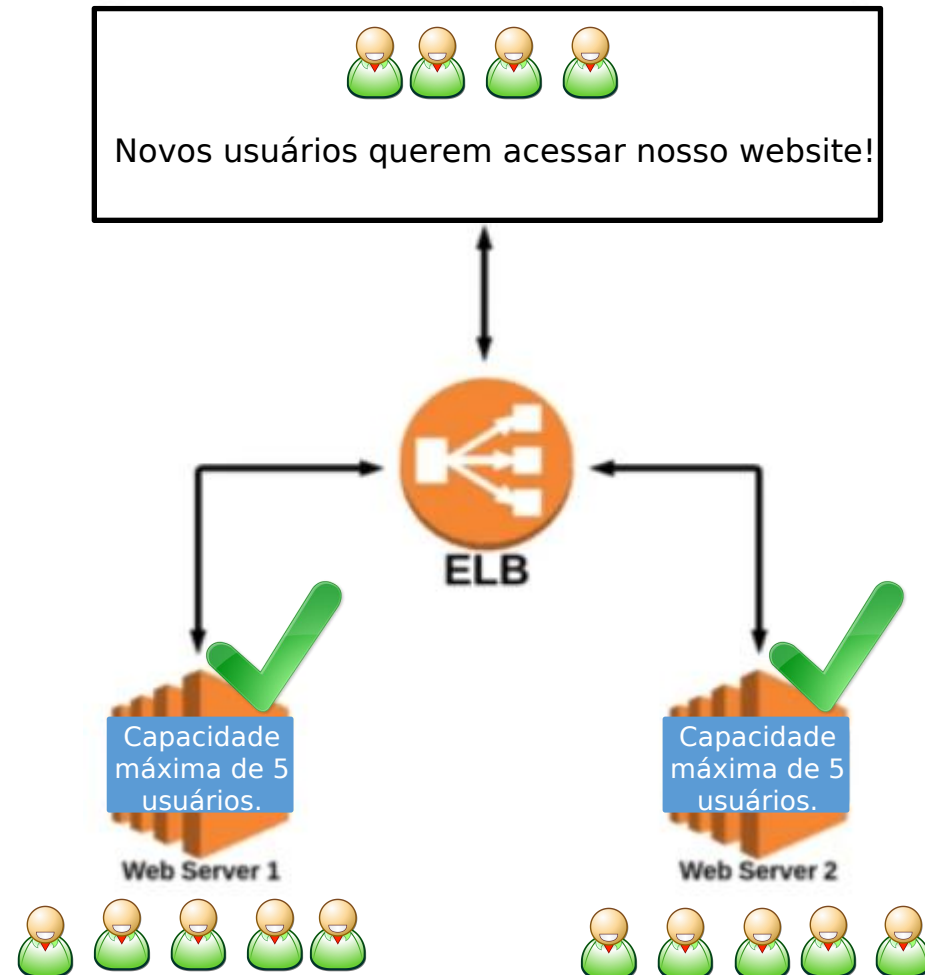


Introdução ao Auto Scaling da AWS



Como funciona o Auto Scaling?

Mas dado que nossos servidores atingiram suas capacidades máximas em números de usuários, o que ocorre se novos usuários fizerem acesso?

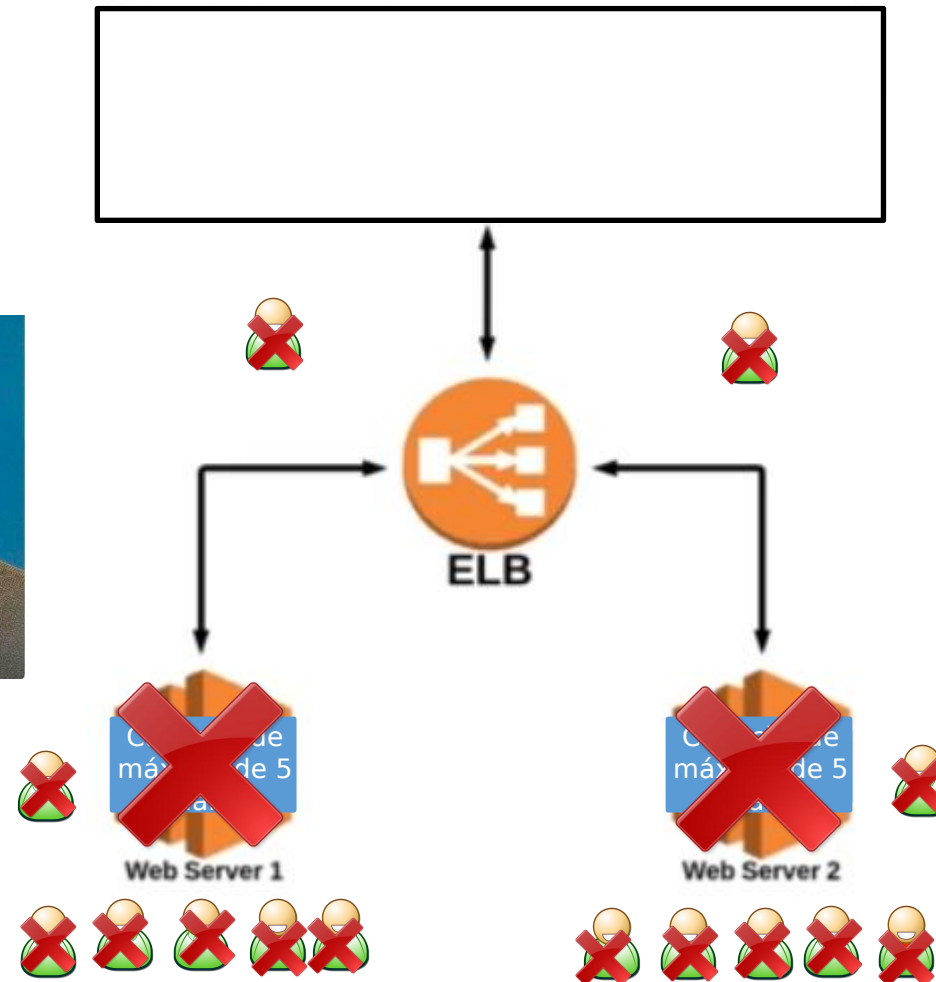
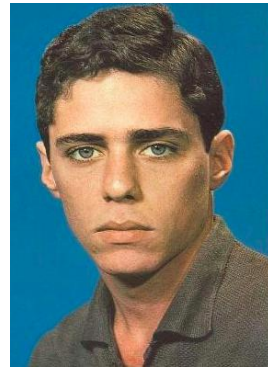


Introdução ao Auto Scaling da AWS



Como funciona o Auto Scaling?

Sem o uso do Auto Scaling, nossos servidores iriam sobrecarregar, executar o serviço extremamente lentos e possivelmente 'cair'.

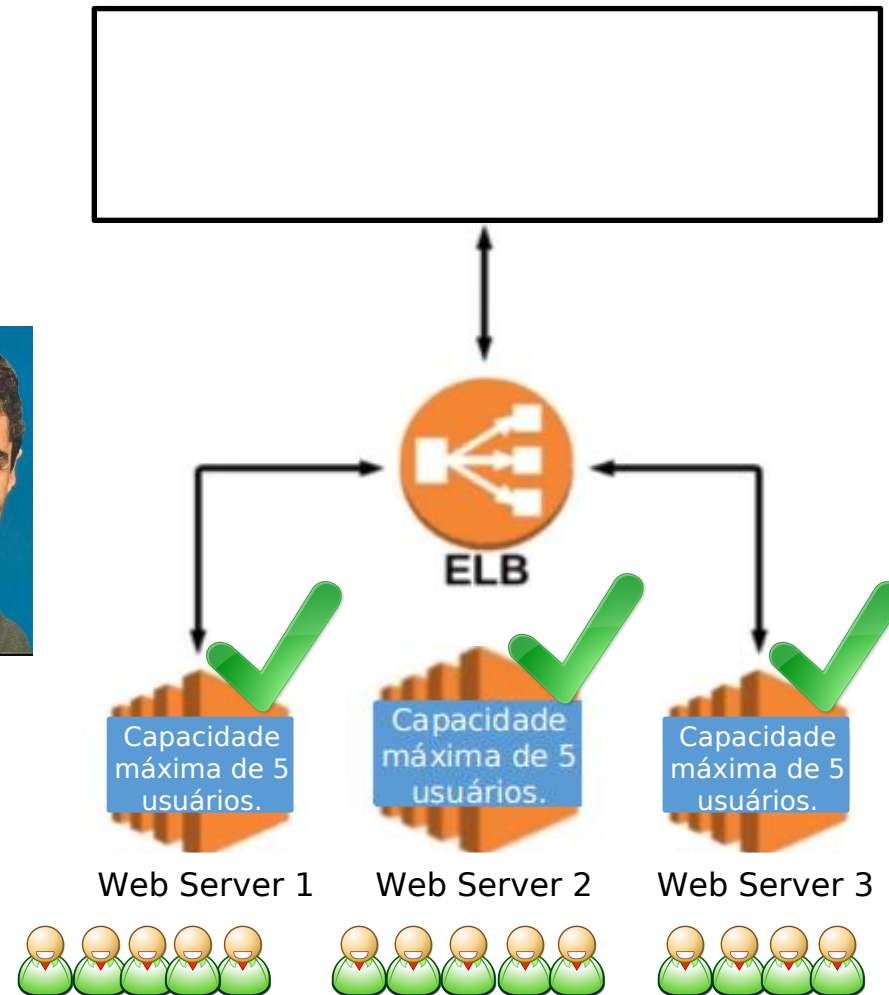
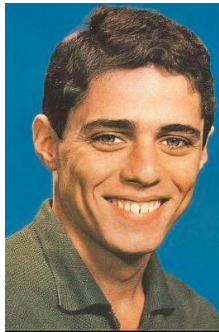


Introdução ao Auto Scaling da AWS

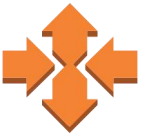


Como funciona o Auto Scaling?

Com o uso do Auto Scaling, assim que necessário será adicionado uma nova instância idêntica aos nossos servidores para que a carga seja suportada.



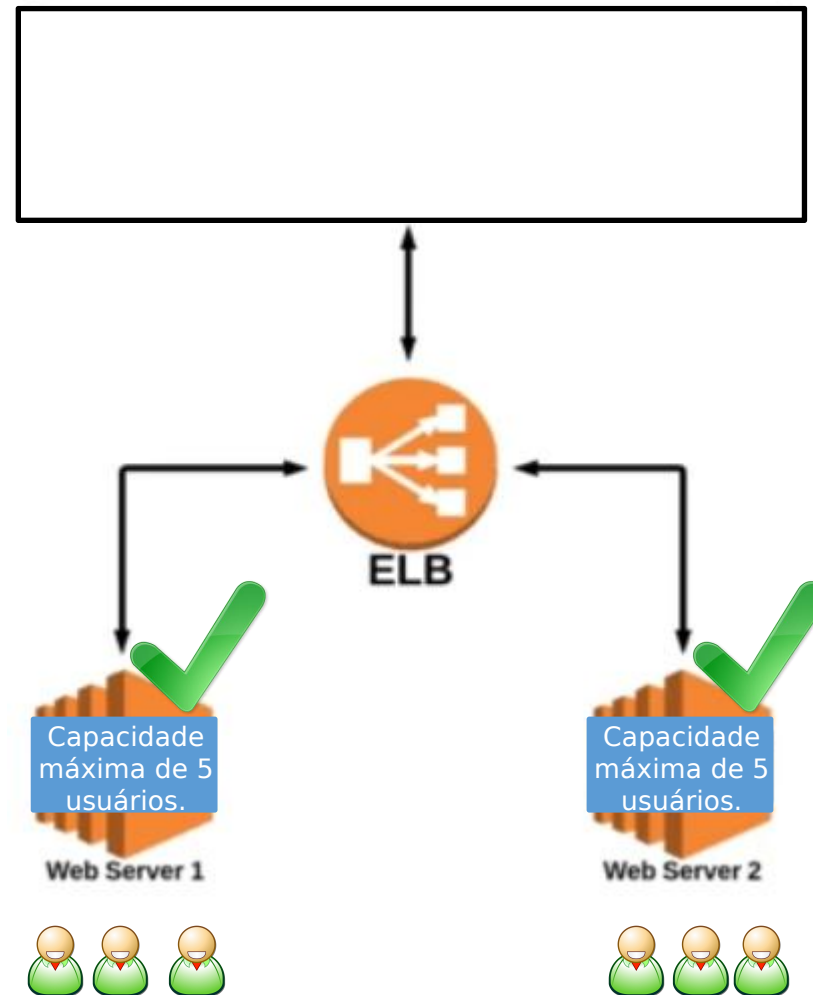
Introdução ao Auto Scaling da AWS



Como funciona o Auto Scaling?

Quando não for mais necessário o Auto Scaling irá remover a instância para que a aplicação volte a trabalhar na sua capacidade normal.

Isso é o que chamamos de 'elasticidade', pois auto scaling juntamente com o ELB funciona como um elástico aumentando e diminuindo a capacidade de carga dos servidores de acordo com a necessidade.



Introdução ao Auto Scaling da AWS



Componentes do Auto Scaling?

Modelo de lançamento (Launch Model): O modelo da instância **EC2** que será usada quando o **Auto Scaling** precisar adicionar uma instância adicional para o **Grupo Auto Scaling**.

Grupo Auto Scaling (Auto Scaling Group): Todas as regras e configurações usadas quando uma instância **EC2** é automaticamente adicionada ou removida.

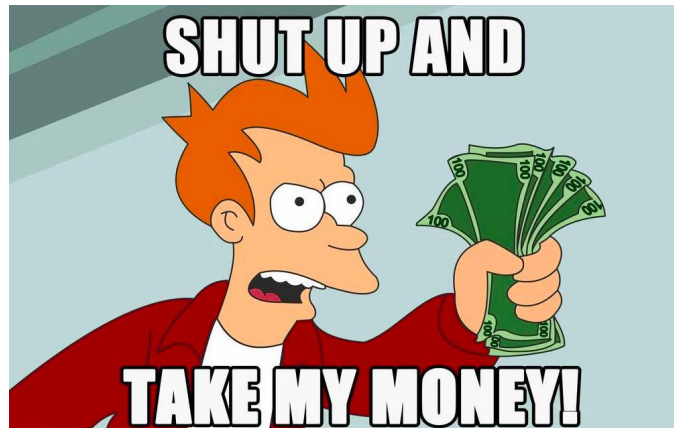
Introdução ao Auto Scaling da AWS



Como é a precificação do Auto Scaling?

O **Auto Scaling** é grátis para uso, seja na conta gratuita seja na paga, mas...

...você será cobrado pelos recursos adicionais usados durante o **Auto Scaling**, por exemplo qualquer instância **EC2** provisionada que ultrapasse o definido na conta gratuita.





Geek University

Evolua seu lado geek!

www.geekuniversity.com.br