

01

## Como lidar com o aumento de acessos?

### Transcrição

É muito comum que o nosso número de acessos aumente com o sucesso das lojas relacionadas e isso não é exatamente um problema, na verdade, quanto mais gente acessar nosso catálogo melhor, certo? Mas o problema na verdade é quando a nossa máquina acaba não sendo suficiente para suportar toda essa carga de acessos simultâneos onde ela acaba caindo e a aplicação, por consequência, indisponível.

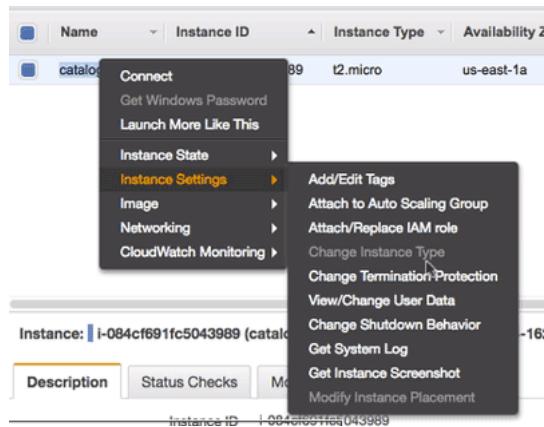


Essa situação ocorre também quando há alguma atração para os usuários o que pode ser, por exemplo, promoção e ofertas em algum determinado horário. Nesse caso, essa quantidade maior (pico) de acessos ocorrerá apenas em uma fração do tempo em que a aplicação fica indisponível.

A questão é: **como fazer a aplicação continuar disponível mesmo com o aumento dos acessos?**

### Escalabilidade vertical

A primeira ideia é melhorar a instância em si. Isso significa melhorar o CPU, HD ou a memória RAM para poder atender mais clientes. Na AWS EC2 isso é muito fácil de fazer basta parar a instância e mudar o tipo da instância (Menu *Instance Settings -> Change Instance Type*).



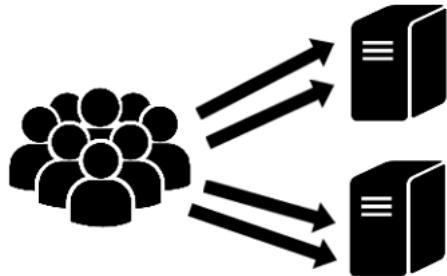
Lembra-se, para alterar o tipo da instância é preciso parar ela antes.

A escalabilidade vertical é muito simples de aplicar mas tem dois problemas: 1) saímos da faixa gratuita e é preciso pagar para instâncias mais poderosas (pode ficar caro). E 2) ela é bastante limitada pois mesmo a melhor instância na EC2 talvez não seja boa suficiente para atender todos os clientes ao mesmo tempo.

Além disso, continuamos com o problema da disponibilidade. Se essa instância cai, toda a aplicação fica offline.

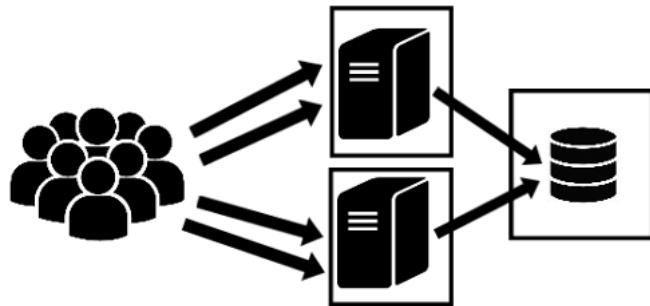
## Escalabilidade horizontal

Uma outra abordagem então seria continuar com a faixa gratuita da AWS e usar uma segunda máquina para rodar a aplicação. Essa nova instância pode atender novos clientes e aumenta a disponibilidade:



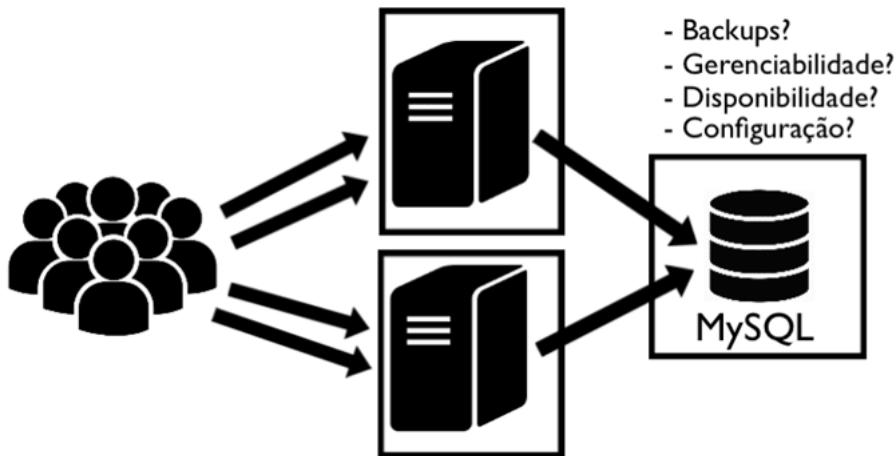
Essa forma de crescer a aplicação também é chamado de *escalabilidade horizontal*. O problema ainda é que temos a servidor web e o banco de dados instalado na mesma máquina. Os dados da aplicação devem ficar centralizados e ambos os servidores devem utilizar o mesmo banco de dados para não causar inconsistências nos nossos dados.

Para resolver isso vamos criar uma instância dedicada para o banco de dados, separados dos servidores web. Os servidores web vão acessar o mesmo banco de dados e sempre ler e gravar os dados do mesmo lugar:



## Escalar e configurar o banco

Repare que nossa arquitetura está crescendo e com isso vem outros problemas. Como gerenciar tudo isso e como garantir que as instalações e permissões estão corretas? Isso principalmente importante para o banco de dados pois agora estamos acessando ele remotamente. Como podemos garantir que o banco fica disponível para as instâncias com um bom desempenho. Também queremos uma configuração simples, de alto nível com backups automáticos:



Todas essas questões (e outras) não são tão fácil de resolver e por isso a AWS já oferece um serviço chamado de Amazon RDS (Relational Data Service) que veremos no próximo vídeo.