

## 22 - INTRODUÇÃO AO SERVIÇO DE AUTOSCALING PARA EC2

ACESSAR DOCUMENTAÇÃO E COMEERRRRRRR

- Forma inteligente criada pela amazon para que o cliente consiga economizar, e a sua performace nas aplicações que utilizam o autoscaling, no caso por exemplo das maquinas EC2 ou das instancias de Banco de dados, para qualquer um que voce habilitar o autoscaling elas vão, alem de economizar dinheiro, terem uma perfomace muito boa.
- O ponto principal do autoscaling é, que se voce tiver uma aplicação que requer por exemplo a utilizaçãode dois servidores no minimo e essa aplicação sempre roda no seu pico (10 AM as 18PM). Então voce terá dois servidores rodando com essa aplicação nesse horario, so que a partir das 19h voce utiliza recursos que somente o servidor 1 poderia suprir. Mas voce tem dois servidores e esta pagando pelos dois, e se tiver um banco de dados ainda possui esse adicional.
- A partir das 19h ate as 9h da manha voce tem um periodo de baixa.
- Alternativa seria pegar uma das estruturas e não utiliza-la dentro do periodo de baixa.
- Ou definir regras para subir o segundo servidor.
- Essa função de autoscaling serve para aplicações que não rodam em seu pico de performace 24/7
- Imagine outro cenario, voce precisa de 3 servidores
- Esses servidores web estão ligados a uma instancia EFS e utilizam um Load Balancer (faz o balanceamento de carga para os 3 servers)
- Voce pode configurar o servidor e o que chamamos de Group Balance.
  - Podemos falar por exemplo que normalmente queremos 3 servers rodando.
  - Em momentos de pico, se algum deles chegar a mais de 80 ou 70% de processamento eu quero que adicione mais um servidor.
  - Se estiver 3 servidores e os 3 estiverem com menos de 50% de processamento, queremos que retire um servidor.
- Toda essa configuração e alertas voce consegue fazer atraves do group balance.
- É uma ferramenta de escalabilidade gigante.