

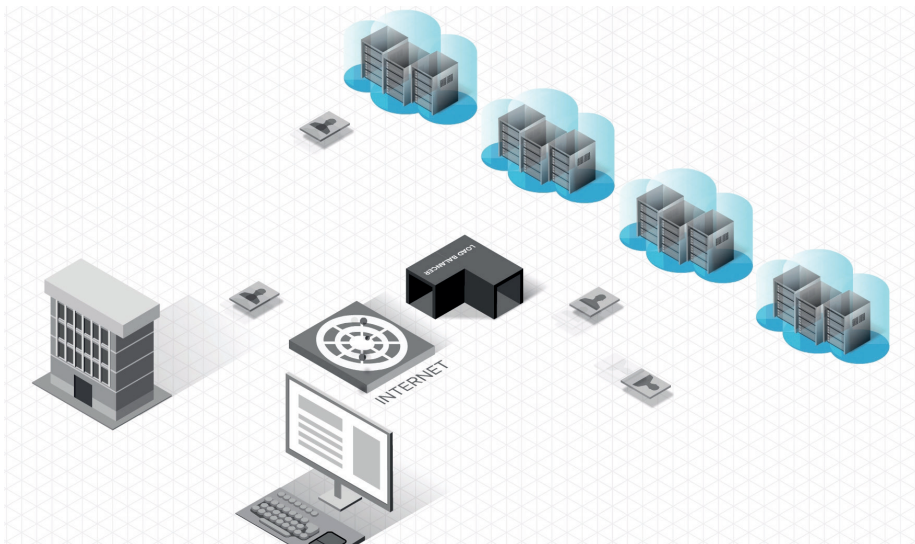


AUTO-SCALING

Após a criação de uma instância de balanceamento podemos configurar um sistema de Auto-Scaling (Escalação Automática). Consiste numa série de regras cuja verificação supõe a criação (ou eliminação) de um Servidor Virtual (Cloud) (usando recursos não atribuídos dentro do mesmo Cloud Datacenter), pode definir as regras, irão ser traduzidas num modelo que serão submetidas a um servidor específico.

Trata-se de maximizar o desempenho dos recursos e dar à sua solução a máxima flexibilidade.

O sistema de AutoScaling permite definir os automatismos de criação e eliminação de servidores de uma maneira simples e intuitiva, seja a partir das API's ou do próprio Painel de Controlo



Uso do Auto-Scaling

A modo de exemplo, podemos fazer Auto-Scaling dos seguintes recursos:

- Em função do número de sessões no balanceador. O sistema de Auto-Scaling encontra-se totalmente integrado com nosso balanceador, de maneira que é possível criar um grupo de servidores que se incorporem automaticamente no balanceador e do mesmo modo aumentar a capacidade do serviço quando for alcançado um determinado número de sessões.
- Se a carga de um dos servidores aumenta acima de uma percentagem de utilização de CPU previamente definida e durante o tempo mínimo que for estabelecido.
- Em determinados períodos de tempo pré-estabelecidos: é possível criar regras para levantar e/ou retirar servidores nesses períodos.

REQUISITOS

- Devemos contar com um modelo personalizado para determinado cloud server
- Devemos contar com uma instancia de balanceamento configurada e funcional e também com um backend configurado.
- Devemos contar com suficientes recursos disponíveis (não-assignados) na cloud para cobrir a soma das configurações de todos os cloud servers que participam no Auto-Scaling (clonados a partir do primeiro servidor com o modelo correcto) que desejamos ter em funcionamento concorrentemente.