



Cloud Computing Data Science

Escalabilidade e Elasticidade em Cloud Computing

Escalabilidade e elasticidade são conceitos relacionados à capacidade de um sistema se adaptar a mudanças na demanda, mas com nuances distintas. Vamos compreender os conceitos.

Escalabilidade é a capacidade de um sistema crescer e gerenciar um aumento na demanda. Trata-se de uma característica planejada e estruturada que permite que um sistema expanda sua capacidade de processamento, armazenamento ou outras funções quando necessário. Existem 2 tipos de escalabilidade:

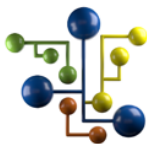
- **Escalabilidade Vertical (Scale-up):** É o aumento da capacidade de um único componente, como um servidor, adicionando mais recursos a ele (por exemplo, mais memória, CPU mais potente, etc...).
- **Escalabilidade Horizontal (Scale-out):** Refere-se à adição de mais instâncias ou nodes ao sistema (por exemplo, adicionar mais servidores), distribuindo a carga entre múltiplas entidades.

Elasticidade é a capacidade de um sistema se ajustar automaticamente e em tempo real à demanda, aumentando ou diminuindo recursos conforme necessário. É uma característica dinâmica e ágil.

Sistemas elásticos monitoram a utilização de recursos e ajustam automaticamente a capacidade, adicionando ou removendo recursos para atender às necessidades atuais, muitas vezes sem intervenção humana.

A elasticidade é ideal para lidar com padrões de demanda imprevisíveis ou picos repentinos, como um site que recebe um aumento inesperado de tráfego.

A escalabilidade é mais sobre planejamento e capacidade de expansão, enquanto a elasticidade lida com a reação automática e em tempo real às mudanças na demanda.



Equipe DSA

Muito Obrigado!
Continue Trilhando Uma Excelente Jornada de Aprendizagem.