

Bootcamp: Engenheiro(a) de Dados Cloud

Trabalho Prático

Módulo 3: Desenho de Arquiteturas de Dados Escaláveis

Objetivos de Ensino

Exercitar os seguintes conceitos trabalhados no Módulo:

1. Utilizar o Docker para encapsulamento de soluções de dados.
2. Utilizar o Kubernetes para deploy automatizado de soluções de dados.
3. Praticar os principais comandos de kubectl.

Enunciado

Você é engenheiro de dados em uma empresa de dados que é referência no mercado. A empresa está neste momento estudando a possibilidade de trocar de provedor de nuvem. O diretor da área de dados já demonstrou sua preocupação com o grande tempo necessário para a migração de todos os workloads do time de dados para outro provedor de nuvem e designou você para desenvolver uma POC em uma estrutura “portável” e agnóstica.

Desse modo, você optou por demonstrar o uso do Kubernetes para soluções de dados. Você vai implementar um processo de dados simples, encapsular sua execução em uma imagem docker e implementar sua execução em k8s.

Você está livre para escolher qual solução será implementada: uma API, um processo Python com algum tratamento de dados, alguma consulta a um banco de dados, como preferir. O importante é demonstrar a solução funcionando no ambiente kubernetes.

Por se tratar de uma POC e de um processo de dados muito simples, você está livre para escolher trabalhar em um cluster em nuvem ou em um cluster local utilizando kind.

Divirta-se!

Atividades

Você deverá desempenhar as seguintes atividades:

1. Desenvolver um processo de dados de sua escolha (API, processo Python, tratamento de um dataset com pandas, consulta a um banco de dados, ou algum outro de sua escolha).
2. Desenvolver o Dockerfile para o encapsulamento de sua solução em uma imagem Docker.
3. Buildar a imagem docker e fazer o push em algum repositório de imagens (público ou privado).
4. Implementar um manifesto para a execução de sua solução, seja um *job* ou um *deployment* com um *service* associado.
5. Executar comandos kubectl para implantar sua solução em seu cluster kubernetes.
6. Verificar que a solução foi implantada com sucesso.
7. Verificar os logs de execução.