

Data Challenge

uma iniciativa Stone.

Case Infrastructure

Data Platform - Processing Engine

Responsável: [Vitor Augusto Toledo Azevedo Pinto](#)

Data Platform Engineering Team Lead

Descrição do case

Com a necessidade de um maior poder de processamento e escalabilidade de recursos nos nossos processos, optamos por Utilizar o Spark como tecnologia de processamento de grandes volumes de dados e utilizando como infraestrutura a solução do Kubernetes para este cenário, seu desafio será criar um Cluster de Spark no Kubernetes e executar um job de spark qualquer com 3 executors pelo menos, garanta que os nós dos cluster de K8's (Kubernetes) conversem entre si e seguindo boas práticas de segurança como limitação de acesso a portas necessárias para a integração e também faça criação de Roles de usuários no Cluster de, admin, edit e view. O projeto deverá ser feito em um ambiente local e disponibilizado no seu github de forma pública para que o nosso time de jurados possa analisar o case após a entrega!

Requisitos

- Uma conta no Github
- Conhecimento em Docker/Kubernetes

Entrega 1

Desenho da solução + Imagens de Docker e todos os artefatos utilizados no projeto postado no Github (enviar link do repositório privado no formulário), esses artefatos serão utilizados para testar o funcionamento da solução como um todo.

Entrega 2

Apresentação do projeto em 20 min detalhando o problema e a solução implementada.

Avaliação do case

Serão avaliados da seguinte forma:

- A. Deployability (roteiro de implantação)
- B. Performance
- C. Estrutura de códigos/artefatos (manutenibilidade)
- D. Recuperação em caso de falhas
- E. Documentação
- F. Segurança

Pesos:

A=1.5	B=1.5	C=1.5	D=2.0	E=2.0
F=1.5				

A nota final será soma das notas da seguinte maneira:

$$\text{nota final} = (A+B+C+D+E+F)$$

3. Avaliação da apresentação

- Conhecimento da solução implementada
- Storytelling
- Maiores dificuldades
- Benefícios da solução

Banca avaliadora

Composta por engenheiros de dados da Stone.