

Teste técnico - Engenharia de Dados

A3Data

Introdução

Você é engenheiro de dados do "Canarinhos Futebol Clube". A diretoria do clube contratou um novo técnico para o time o qual propôs um novo estilo de treinamento do time totalmente ancorado em dados. Desse modo, ele precisa que dados históricos do campeonato brasileiro sejam disponibilizados para fácil acesso e consulta. Além disso, há uma grande preocupação do clube com relação à **segurança de dados** e à **gestão de acessos** dos dados disponibilizados. Eles escolheram trabalhar com a AWS como provedor de tecnologia de nuvem.

Você pode encontrar os dados do campeonato brasileiro neste link:

https://github.com/neilsoncrepalde/campeonato_brasileiro_dataset

Desafio

Seu desafio é:

- Construir um ambiente mínimo na AWS para realizar a **Ingestão**, o **processamento** e a **disponibilização** de dados históricos do campeonato brasileiro.
- Use a *stack* que desejar. Pode com Python, Airflow ou Step Functions, EMR ou Glue, Athena ou Redshift.
- É necessário garantir que os dados estejam protegidos em repouso. Utilize criptografia em repouso no seu ambiente e explique como ela funciona.
- É necessário garantir que o acesso seja muito controlado. Implemente no seu ambiente um segundo usuário (*usuario-negocio*) com permissões apenas de leitura na camada de consumo dos dados. Mostre as *policies* implementadas e explique como funciona a sua lógica.

Você deve desenvolver o ambiente mencionado e apresentá-lo ao vivo. Deve apresentar ainda a estratégia de criptografia e as *policies* de permissão de usuário criadas.

Na apresentação você deve explicar a construção do ambiente, a lógica de suas decisões de desenvolvimento e realizar 3 consultas com o usuário *usuario-negocio* que respondam às seguintes perguntas:

1. Em quantas partidas o Palmeiras recebeu cartão amarelo no segundo tempo quando jogava como visitante?
2. Qual foi a média de gols do Cruzeiro nas partidas em que venceu jogando fora de casa?
3. Qual foi o clube que teve mais cartões vermelhos quando jogava em casa?

Entregáveis

Para os fins deste desafio, você deve entregar os seguintes itens:

- Um repositório git público contendo os códigos desenvolvidos para a solução
- Um texto explicando os dois pontos requeridos da solução (o método de criptografia em repouso adotado e as policies para definição de restrições de acesso do usuário *usuario-negocio*).
- Apresentar ao vivo o ambiente construído explicando o racional das decisões de construção.
- Realizar as consultas para responder às 3 perguntas mencionadas acima.

Observações

Para os fins do teste técnico, é possível que haja algum custo AWS com a construção do ambiente. É recomendável que o(a) candidato(a) utilize, ao máximo quanto possível, a camada de serviços *free tier* e monitore o custo da conta para que não haja gastos excessivos. A construção do ambiente, bem como os custos associados são de responsabilidade do(a) candidato(a).