



TECH CHALLENGE

Tech Challenge é o projeto da fase que engloba os conhecimentos obtidos em todas as disciplinas. Esta é uma atividade que, a princípio, deve ser desenvolvida em grupo. É importante atentar-se ao prazo de entrega, uma vez que essa atividade é obrigatória e vale 90% da nota de todas as disciplinas da fase.

O problema

Pipeline Batch Bovespa: ingestão e arquitetura de dados

Construa um pipeline de dados completo para **extrair, processar e analisar** dados de ações ou índices da B3, utilizando AWS S3, Glue, Lambda e Athena. Você pode extrair esses dados através de bibliotecas como [yfinance](#) ou realizar scrapping de sites que contenham esses dados (Exemplo: [ibovespa](#)).

Para esse desafio, sua entrega deve conter os seguintes requisitos:

- **Requisito 1:** scrap de dados de ações ou índices da B3 (granularidade diária).
- **Requisito 2:** os dados brutos devem ser ingeridos no s3 em formato parquet com partição diária.
- **Requisito 3:** o bucket deve acionar uma lambda, que por sua vez irá chamar o job de ETL no glue.
- **Requisito 4:** a lambda pode ser em qualquer linguagem. Ela apenas deverá iniciar o job Glue.
- **Requisito 5:** o job Glue pode ser feito no modo visual ou via código. Este job deve conter as seguintes transformações obrigatórias:
 - A: agrupamento numérico, sumarização, contagem ou soma.
 - B: renomear duas colunas existentes além das de agrupamento.
 - C: realizar um cálculo com base na data, como por exemplo média móvel, diferença entre períodos, valores extremos no período etc.
- **Requisito 6:** os dados refinados no job glue devem ser salvos no formato parquet em uma pasta chamada refined, particionado por data e pelo nome ou código da ação/índice.

Tech Challenge

- **Requisito 7:** o job Glue deve automaticamente catalogar o dado no Glue Catalog e criar uma tabela no banco de dados (pode ser o default).
- **Requisito 8:** os dados devem estar disponíveis e serem consultados usando SQL através do Athena.

Lembre-se que você poderá apresentar o desenvolvimento do seu projeto durante as lives com docentes ou nas mentorias. Essa é uma boa oportunidade para discutir sobre as dificuldades encontradas e pegar dicas valiosas com docentes especialistas e colegas de turma. O Tech Challenge é obrigatório! Se atente ao prazo da entrega até o final da fase.

Entregáveis

A entrega deve ser feita no formato de um **vídeo** apresentando o desenho da arquitetura, a demonstração dos serviços implementados e a realização dos requisitos. O vídeo deve ter no máximo **10 minutos**.

Boa sorte! 

The background is a dark, abstract network visualization. It features a complex web of thin, light-colored lines connecting numerous small, glowing nodes. The nodes are primarily white and light blue, with some larger, more prominent nodes in teal and orange. The overall effect is a sense of dynamic connectivity and data flow, typical of a digital or biological network map.

POSTECH