ANA RAQUEL

PÓS TECH

MACHINE LEARNING ENGINEERING

TECH CHALLENGE

FASE 02

TECH CHALLENGE

Tech Challenge é o projeto da fase que engloba os conhecimentos obtidos em todas as disciplinas dela. 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 do pregão D-1 da B3, utilizando AWS S3, Glue, Lambda e Athena. Para esse desafio, sua entrega deve conter os seguintes requisitos:

Pipeline Batch Bovespa (entrega obrigatória):

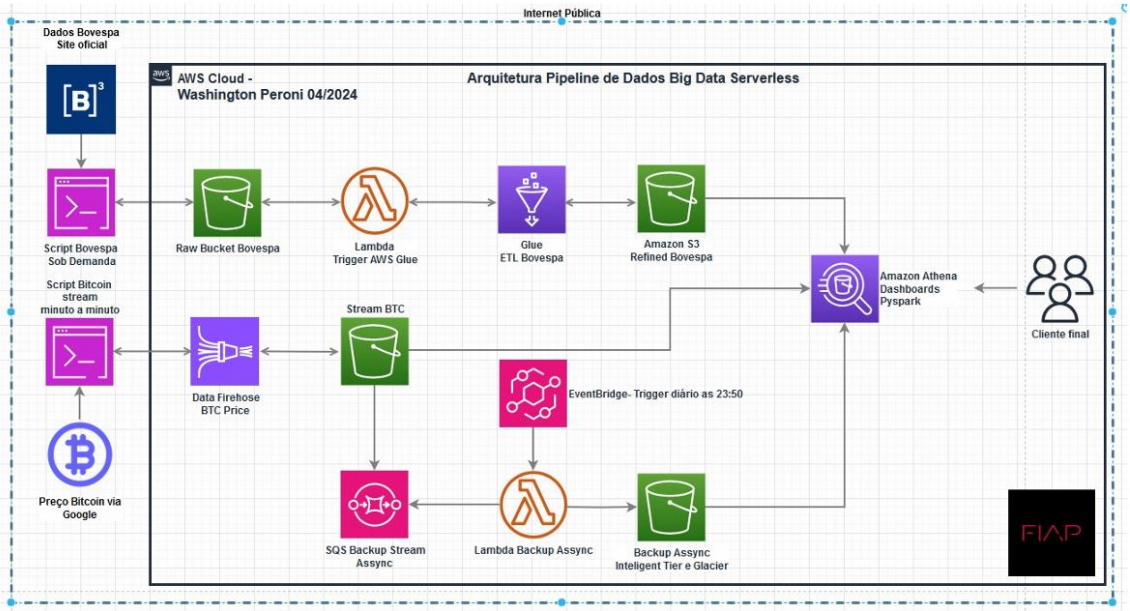
- **Requisito 1**: scrap de dados do site da B3 com dados do pregão D-1.
- **Requisito 2**: os dados brutos devem ser ingeridos no s3 em formato parquet com partição diária.
- **Requisito 3**: o bucket deve acionar uma lambda, que por sua vez irá chamar o job de ETL no glue.
- **Requisito 4**: a lambda pode ser em qualquer linguagem. Ela apenas deverá iniciar o job Glue.
- **Requisito 5**: o job Glue deve ser feito no modo visual. 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 campos de data, exemplo, poder ser duração, comparação, diferença entre datas.
- **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 abreviação da ação do pregão.

- **Requisito 7**: o job Glue deve automaticamente catalogar o dado no Glue Catalog e criar uma tabela no banco de dados default do Glue Catalog.
- **Requisito 8**: os dados devem estar disponíveis e legíveis no Athena.
- **Requisito 9**: é opcional construir um notebook no Athena para montar uma visualização gráfica dos dados ingeridos.

Pipeline Stream Bitcoin (entrega opcional):

Arquitetura de referência totalmente opcional. Só trabalhe nela se sobrar tempo após a Arquitetura **obrigatória** da Bovespa.

A seguir, temos um modelo da arquitetura a ser construída:



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

Boa sorte!


POS TECH