

Desafio Técnico

Objetivo

Esse teste consiste em implementar um pipeline de dados que faça:

- Importação de um arquivo CSV para processamento
- Limpeza dos dados importados
- Criação de uma tabela-resultado que:
 - Liste em ordem decrescente as "location_region" por média de "risk score"
- Criação de uma tabela-resultado que:
 - Considerando somente a transação mais recente ("timestamp") com "transaction_type" igual a "sale" de cada "receiving address", liste os 3 "receiving address" com maior "amount" dentre estas transações (apresente o "receiving address", o "amount" e o "timestamp").

Fontes

Arquivo CSV disponível em:

https://drive.google.com/file/d/1nOdebavhyRPsqOTS5YMMDxpA68RMN4I/view?usp=drive_link

Requerimentos

Orquestração	Gostaríamos de avaliar sua sugestão de solução para orquestração automatizada dessa pipeline de dados.
Data Quality Automatizado	<p>Esta solução deverá monitorar através de indicadores e métricas de qualidade dos dados, como a quantidade de registros, a quantidade de erros, o percentual de conformidade (qtd. erros/ qtd. registros), entre outros, além de reportar os problemas e as anomalias de qualidade dos dados, como os valores faltantes, os valores inconsistentes, os valores incorretos, entre outros.</p> <p>Simplificar:</p> <ul style="list-style-type: none">Quantidade total de registrosQuantidade de registros inseridosQuantidade de registros com erro

Entrega

O produto final deste desafio técnico pode ser entregue em um repositório compartilhado conosco (GitHub por exemplo) ou mesmo, de forma mais simples, uma pasta compactada contendo toda a solução enviada por e-mail ou compartilhada via link.