

PROCESSO DE PENSAMENTO, PREOCUPAÇÕES E CONSEQUÊNCIAS

PENSAMENTO

Comecei a entender o objetivo e vi que exigia que as respostas de cinco endpoints de API sejam armazenadas de forma eficiente, organizada e resiliente, em um data lake. O objetivo central é preservar os dados brutos para análises futuras, possibilitando reprocessamentos, validações e integração com ferramentas de BI.

PREOCUPAÇÕES

Mudanças no schema exemplo taxes -> taxation, mudanças silenciosas e repentinas podem quebrar pipelines e gerar inconsistência em relatórios.

Volume de dados a longo prazo, como a ingestão pode gerar um grande volume de dados, por isso o formato de texto compactado pode ser considerado, armazenamento em data lakes distribuídos com políticas de ciclo de vida pode ser adotado.

CONSEQUÊNCIAS

Decisão	Consequência positiva	Possível risco / mitigação
Armazenar JSONs brutos	Preserva dados para reprocessamento ou auditoria futura	Pode ser redundante — compressão ou limpeza periódica ajudam
Estrutura particionada por data e loja	Consulta rápida por tempo e storeId	Pode gerar muitas pastas — usar compactação e merge depois
Manter dados mesmo com schema alterado	Permite histórico completo e comparação temporal	Processamentos posteriores devem prever múltiplos schemas

Decisão	Consequência positiva	Possível risco / mitigação
Não validar schema no momento da ingestão	Evita falhas imediatas por causa de pequenas mudanças	Validação pode ser feita em etapa posterior