

ROTEIRO DE RESOLUÇÃO – N26 DATA MODELING TEST

❖ Depois da análise do arquivo do teste por inteiro, a conclusão foi:

O teste consiste em criar um diagrama MER e um script SQL que resolvam os problemas citados no arquivo. A seguir a lógica usada na resolução:

Análise dos problemas:

1) O principal possível problema observado na lentidão de alteração e processamento de dados está relacionado com o tamanho e organização da tabela. Mesmo com o uso do Spark, a quantidade de informações pode ser muito grande e pesada para executar com eficiência.

2) No caso dos problemas de consistência, possivelmente estão relacionados a formatação dos dados ou dados ausentes. Como a tabela é muito grande, fica difícil enxergar ausências ou má formatação de dados em campos específicos, e isso pode acarretar inconsistências nos relatórios.

3) O modelo excede o tempo de rodar (no caso do dashboard nulo), pois são muitos dados a serem carregados no mesmo lugar ao mesmo tempo.

4) Por mais que seja possível criar uma tabela de versionamento, esta é uma estratégia pouco utilizada e que adiciona dados desnecessários ao banco. A maneira mais fácil de fazer o versionamento é utilizando sistemas de controle de versão ou plataformas online próprias para esse fim.

Solução proposta:

A melhor solução encontrada para resolver os problemas citados é a divisão da tabela em tabelas menores. O modelo utilizado foi o Star Schema, pois facilita a visualização dos dados e a utilização de funções como 'join'. Trata-se de um modelo desnormalizado, mas útil para esse tipo banco (composto principalmente por informações que se relacionam apenas com a chave primaria).

Conclusão:

- 1) Diagrama MER: Os dados foram organizados no formato estrela.
- 2) Script SQL: Foi feita a implementação do modelo utilizando o MySQL.
- 3) Diagrama DER: Foi feito adicionalmente para melhor visualização da modelagem.
- 4) GitHub: Foi feita a postagem na plataforma como uma possível forma de versionar o código.