

Comentários sobre o Processo ETL e Modelagem Dimensional

Breno Valente Manhães – 122038517
João Pedro Moretti Fontes Ferreira – 122081366
Murilo Jorge de Figueiredo – 122079597

22 de junho de 2025

Link do Repositório

Acesse o repositório completo no GitHub clicando aqui.

Introdução

Este documento tem como objetivo apresentar um resumo crítico do processo de desenvolvimento da solução de Data Warehouse (DW), com foco no projeto de ETL (Extract, Transform, Load) e na modelagem dimensional. A construção deste ambiente teve como base a necessidade de integrar os dados transacionais de seis empresas de aluguel de veículos, possibilitando análises históricas, relatórios gerenciais e suporte à decisão de forma unificada.

Escolhas Técnicas

Durante o projeto, algumas decisões importantes foram tomadas para garantir a consistência dos dados e a efetividade das análises:

- **Inclusão do pátio de retirada e devolução no fato de locação:** Essa decisão foi fundamental para possibilitar tanto o rastreamento operacional dos veículos quanto a posterior geração da matriz de movimentação entre pátios.
- **Padronização dos status de reserva e locação:** Como os status variavam entre sistemas (ex: “ativa”, “em andamento”, “confirmada” etc.), optou-se por aplicar um mapeamento uniforme em uma etapa de transformação na *staging area*.
- **Permissão de datas nulas:** Para locações e reservas ainda ativas ou futuras, foi necessário permitir campos de data (como data de fim) com valor nulo. Isso evita a perda de dados durante os joins com a dimensão tempo.
- **Separação clara entre as etapas ETL:** Adotamos a estratégia de transformação direta na área de staging, mantendo scripts independentes para extração, transformação e carga, seguindo boas práticas de modularidade.

Desafios Enfrentados

- **Heterogeneidade dos dados entre os grupos:** Cada empresa tinha modelagens distintas, com nomes de campos e granularidades diferentes. Esse fator exigiu atenção especial na fase de transformação para conformar os dados em um formato único.
- **Ambiguidade em datas no grupo Rickauer:** A data de devolução do veículo não estava explicitamente disponível, levando à necessidade de decisão por utilizar a data da fatura associada ao contrato como proxy para a data de fim da locação.
- **Dificuldades técnicas e de modelagem:** Houve complexidade em evitar joins cartesianos causados por chaves naturais repetidas entre empresas. A inclusão do campo *fonte_dados* foi essencial para compor chaves compostas durante os joins.

Resultados Obtidos

O ambiente DW resultante permite gerar todos os relatórios gerenciais propostos, incluindo:

- Controle de pátio por grupo de veículos e origem da frota;
- Locações ativas e estatísticas de duração;
- Reservas futuras com agrupamento por pátio e cidade de origem do cliente;
- Grupos de veículos mais alugados por perfil de cliente;
- Matriz estocástica de movimentação entre pátios (Cadeia de Markov).

Além disso, o modelo estrela e as transformações aplicadas tornam o DW robusto para expansões futuras, caso novos relatórios ou análises sejam necessários.

Organização dos Arquivos do Projeto

A estrutura de arquivos da Parte 2 do projeto foi organizada de forma lógica e modular, com diretórios separados para cada grupo (por exemplo: `brenopprufrj`, `rickauer`, `wesleyConceicao` etc.), contendo os scripts de extração e seus respectivos modelos lógicos.

Os arquivos de carga, transformação, staging, modelo dimensional (esquema estrela), bem como os scripts de relatórios SQL e geração da matriz de movimentação entre pátios, estão centralizados na raiz da pasta `Parte_2`. Há também um subdiretório exclusivo para os relatórios e a geração da matriz de percentuais (`relatorios`) e arquivos de documentação, como o diagrama do modelo estrela e sua descrição em PDF.

Essa organização visa facilitar a reprodutibilidade, depuração e colaboração entre os membros do grupo.

Conclusão

O desenvolvimento desta solução demonstrou o papel crítico da padronização e da governança de dados em ambientes colaborativos entre múltiplas fontes. A construção do

DW atendeu aos objetivos de integrar os dados centrais das operações de aluguel de veículos, tornando viável a análise global do negócio e a extração de insights estratégicos. As dificuldades encontradas foram superadas com decisões técnicas fundamentadas, e os resultados obtidos mostram a aderência da solução proposta aos requisitos do problema.