

Avaliação 02: Modelagem de Data Warehouse - Parte 2

Brian José Costa de Medeiros DRE: 121087678

Este documento detalha as escolhas e o raciocínio por trás do desenvolvimento do processo ETL (Extração, Transformação, Carga) e do modelo dimensional do Data Warehouse (DW) para o sistema de locação de veículos, conforme solicitado na Parte 2 do trabalho.

1. Considerações sobre as Fontes de Dados (Extração ETL)

A fonte primária de dados para este DW é o esquema OLTP projetado na Parte 1, detalhado no `schema.sql` e nos PDFs anexados (`dicionario_de_dados.pdf`, `Descricao_do_Projeto.pdf`). Para simular a integração com "outros grupos da turma", conforme o item 1 da Parte 2 do enunciado, a estratégia adotada na fase de extração para a área de *staging* inclui:

- **Identificação da Origem:** Cada registro extraído para a área de *staging* receberá uma `empresa_id`. Esta `empresa_id` permitirá rastrear a origem do dado, sendo "Empresa A" para o esquema fornecido (o seu grupo), e "Empresa B", "Empresa C", etc., para as fontes simuladas dos outros grupos.
- **Uniformidade na Staging:** As tabelas na área de *staging* (`staging.clientes`, `staging.locacoes`, etc.) foram projetadas para serem uma cópia direta das tabelas OLTP, com a adição da coluna `empresa_id`. Isso pressupõe que os esquemas dos outros grupos, embora independentes, possuem estruturas funcionais *equivalentes* para as entidades de negócio (clientes, veículos, locações, pátios), facilitando a conformidade posterior. Em um cenário real, a extração de diferentes esquemas exigiria mapeamentos complexos e normalização inicial para a área de *staging*.
- **Frequência de Extração:** O enunciado pede a especificação dos tempos de acionamento. Para um DW de gestão e análise, uma frequência diária (por exemplo, extrações noturnas) é recomendada para a maioria das tabelas de fatos e dimensões que refletem transações diárias. Para dimensões que mudam raramente (como `Dim_Tempo`), a carga é única ou atualizada anualmente.

2. Projeto do Modelo Dimensional (Esquema Estrela)

O modelo dimensional foi projetado seguindo o paradigma de esquema estrela, que otimiza as consultas para relatórios e análises. As escolhas de tabelas de fatos e dimensões foram baseadas diretamente nos requisitos de relatórios gerenciais globais e na análise de previsão de ocupação de pátio por cadeia de Markov,

conforme detalhado nos itens 12 e 13 do enunciado.

- **Fatos Identificados:**

- Fato_Locacao: Representa o evento de aluguel de veículos. Contém medidas (valores, durações) e chaves estrangeiras para as dimensões que descrevem o contexto da locação.
- Fato_Reserva: Representa o evento de reserva. Essencial para analisar a demanda futura e o comportamento das reservas.
- Fato_Movimentacao_Patio: Crucial para a análise de cadeia de Markov. Este fato foca na transição de veículos entre pátios (retirada e devolução), permitindo o cálculo dos percentuais de movimentação.

- **Dimensões Identificadas:**

- Dim_Tempo: Uma dimensão comum e fundamental para qualquer DW, permitindo análises por ano, mês, dia, etc. População prévia é essencial.
- Dim_Cliente: Contém os atributos dos clientes.
- Dim_Motorista: Informações dos condutores. Mantida separada de Dim_Cliente devido à relação N:M (um cliente PJ pode ter vários motoristas).
- Dim_Veiculo: Detalhes sobre os veículos, incluindo características e agrupamentos (nome_grupo_veiculo, tarifa_diaria_base_grupo). É aqui que a origem_frota (Própria/Externa) será integrada, a partir da empresa_id da área de staging, para atender ao relatório de "Controle de pátio".
- Dim_Patio: Informações sobre os pátios de retirada e devolução.

3. Processo de Transformação (ETL Transform)

A fase de transformação é onde os dados brutos da área de *staging* são limpos, padronizados, enriquecidos e preparados para o modelo dimensional.

- **Limpeza e Padronização:** Tratamento de valores NULL, formatação de datas e textos.
- **Derivação de Atributos:** Criação de novas medidas (ex: duracao_em_dias, dias_restantes_para_devolucao) e atributos dimensionais (ex: origem_frota no Dim_Veiculo com base na empresa_id de *staging*).
- **Conformidade de Dimensões:** Embora com fontes simuladas, o objetivo é garantir que as dimensões (Tempo, Cliente, Veículo, Pátio, Motorista) sejam consistentes e reutilizáveis por diferentes tabelas de fatos e por dados de diferentes origens (simuladas). As chaves substitutas (sk_) são geradas neste estágio para isolar o DW das chaves naturais dos sistemas OLTP.

4. Processo de Carga (ETL Load)

A fase de carga insere os dados transformados nas tabelas de fatos e dimensões do

DW.

- **Carga de Dimensões:** As dimensões são carregadas primeiro. Para as dimensões que não mudam frequentemente (como Dim_Tempo), a carga pode ser única. Para as que mudam (Cliente, Veículo, Motorista, Pátio), uma estratégia de Tipo 1 (SCD Type 1 - sobrescrever) ou Tipo 2 (SCD Type 2 - manter histórico com novas linhas e datas de validade) pode ser aplicada. Para este projeto, priorizamos a simplicidade e a carga inicial sem SCD Type 2 explícito, focando em inserir novos registros ou atualizar existentes para chaves naturais (_oltp_id).
- **Carga de Fatos:** Os fatos são carregados após as dimensões, garantindo que todas as chaves estrangeiras (sk_) para as dimensões já existam.

5. Geração de Relatórios e Matriz de Movimentação

Os relatórios gerenciais são construídos diretamente sobre o esquema estrela, aproveitando sua otimização para consultas analíticas. A matriz de percentuais de movimentação entre pátios será derivada da Fato_Movimentacao_Patio, que foi projetada especificamente para este fim.

Conclusão

A construção deste Data Warehouse relacional oferece uma plataforma robusta e integrada para as seis empresas associadas de locação de veículos. Ao consolidar dados transacionais de múltiplas fontes (simuladas neste caso) em um modelo dimensional otimizado, o DW permite:

- **Relatórios Gerenciais Globais:** A capacidade de gerar relatórios consolidados sobre ocupação de pátios, desempenho de locações e reservas, e preferências de grupos de veículos, agora unificados em uma visão corporativa, é um ganho significativo. Isso supera as limitações dos sistemas OLTP independentes, que se focam nas operações diárias de cada empresa.
- **Análises Unificadas e Tomada de Decisão:** A estrutura do DW facilita análises mais profundas e preditivas, como a previsão de ocupação de pátio via cadeias de Markov. Essa visão integrada permite à gestão otimizar a realocação de frotas, ajustar estratégias de precificação e planejar aquisições de veículos com base em dados históricos e tendências globais.

Embora a simulação de múltiplas fontes na extração tenha sido uma simplificação necessária, o design do modelo dimensional e a lógica ETL estabelecida fornecem uma base sólida para uma implementação real, onde os desafios de integração de diferentes sistemas OLTP seriam abordados com maior complexidade na fase de transformação. O DW proposto atende plenamente aos objetivos de fornecer

inteligência de negócio integrada e apoiar decisões estratégicas em um ambiente de negócios colaborativo.