

# ROTA: Estrutura de dados – ETL

## O PROCESSO ETL: Atualização do Data Warehouse

Neste relatório será detalhado como os dados serão extraídos, transformados e carregados no nosso Data Warehouse, garantindo que as informações estejam sempre atualizadas e prontas para análise. O processo segue uma sequência lógica para respeitar as dependências entre as tabelas.

<p><b>1.</b> <b>ÁREA DE PREPARAÇÃO (STAGING): PADRONIZAÇÃO DOS DADOS BRUTOS</b></p>	<ul style="list-style-type: none"><li>• <b>Objetivo:</b> Receber os dados brutos da fonte original e aplicar as primeiras camadas de limpeza e padronização. Isso garante que teremos um ponto de partida consistente e de alta qualidade para o carregamento no Data Warehouse.</li><li>• <b>Processo:</b><ul style="list-style-type: none"><li>◦ Extraímos os dados da tabela original da Super Store (rota-super-store.superstore.superstore).</li><li>◦ Transformamos variáveis categóricas, padronizando a capitalização (tudo para maiúsculas) e removendo espaços extras (usando TRIM(UPPER())).</li><li>◦ <b>Carregamos</b> o resultado em uma nova tabela de preparação: rota-super-store.superstore.superstore_standardized.</li></ul></li><li>• <b>Dependência:</b> Nenhuma. Este é o primeiro passo do fluxo.</li></ul>
<p><b>2.</b> <b>CARREGAMENTO DAS TABELAS DE DIMENSÃO</b></p>	<ul style="list-style-type: none"><li>• <b>Objetivo:</b> Popular as tabelas de dimensão com dados únicos e descritivos. As dimensões fornecem o "contexto" (quem, o quê, onde, quando) para as análises.</li><li>• <b>Processo:</b><ul style="list-style-type: none"><li>◦ A partir da tabela superstore_standardized, identificamos os valores únicos para cada dimensão (clientes, produtos, pedidos, modos de envio, localizações, mercados e datas).</li><li>◦ Para cada registro único, criamos uma Chave Primária (PK) numérica (ex: customer_pk, product_pk), que será o identificador principal dentro do Data Warehouse.</li><li>◦ Essas chaves e seus atributos são carregados nas respectivas tabelas de dimensão (ex: DimCustomer, DimProduct, DimLocation).</li><li>◦ A DimDate é um caso especial: ela é geralmente pré-carregada com um calendário completo ou atualizada para incluir novas datas, não dependendo diretamente das transações para sua estrutura inicial.</li></ul></li><li>• <b>Dependência:</b> As tabelas de dimensão dependem da superstore_standardized. Elas podem ser carregadas em paralelo, mas todas devem estar prontas antes da tabela de fatos.</li></ul>

<p><b>3.</b> <b>CARREGAMENTO</b> <b>DA TABELA DE</b> <b>FATOS</b> <b>(FACTSALES)</b></p>	<ul style="list-style-type: none"> <li>• <b>Objetivo:</b> Popular a tabela de fatos com as métricas quantitativas (vendas, lucros, quantidades) e as conexões (Chaves Estrangeiras) para as dimensões. Esta é a tabela central para as análises.</li> <li>• <b>Processo:</b> <ul style="list-style-type: none"> <li>◦ Para cada transação (item de linha de pedido) na superstore_standardized, as métricas (sales, quantity, profit, discount, shipping_cost) são selecionadas.</li> <li>◦ As Chaves Estrangeiras (FKs) são geradas através da "ligação" (JOIN) da superstore_standardized com cada uma das tabelas de dimensão já carregadas. Por exemplo, o customer_id da superstore_standardized é usado para buscar o customer_pk correspondente na DimCustomer.</li> <li>◦ As métricas e as FKs são então carregadas na FactSales.</li> </ul> </li> <li>• <b>Dependência:</b> A FactSales depende da superstore_standardized e de todas as tabelas de dimensão estarem carregadas.</li> </ul>
<p><b>4.</b> <b>VISUALIZAÇÃO</b> <b>DO PIPELINE DE</b> <b>ATUALIZAÇÃO</b></p>	<pre> graph TD     subgraph "Extração e Transformação Inicial (Stage)"         A[Dados Brutos (Tabela Original)] --&gt; B{Limpeza e Padronização de Categorias};         B --&gt; C(Criação/Atualização: superstore_standardized);     end      subgraph "Carregamento das Dimensões"         C --&gt; D1(Carregar DimCustomer);         C --&gt; D2(Carregar DimProduct);         C --&gt; D3(Carregar DimOrder);         C --&gt; D4(Carregar DimShipment);         C --&gt; D5(Carregar DimLocation);         C --&gt; D6(Carregar DimMarket);         D7[Carregar/Atualizar DimDate];     end      subgraph "Carregamento da Tabela de Fatos"         D1 &amp; D2 &amp; D3 &amp; D4 &amp; D5 &amp; D6 &amp; D7 --&gt; F(Carregar FactSales);         C --&gt; F;     end      F --&gt; G[Data Warehouse Pronto para Análise]; </pre>
<p><b>CONCLUSÃO</b></p>	<ul style="list-style-type: none"> <li>• <b>Ordem Lógica:</b> a ordem dos passos é crucial devido às dependências;</li> <li>• <b>Chaves Primárias e Estrangeiras:</b> as chaves são a "cola" que une tudo;</li> <li>• <b>Qualidade dos Dados:</b> o processo de ETL garante dados limpos e prontos para uso analítico;</li> <li>• <b>Preparação para Análise:</b> essa estrutura permite análises de dados mais eficientes e a geração de insights valiosos.</li> </ul>