

1. PROBLEMA (Contexto e Desafios Iniciais)

O desafio da *task* foi consolidar e transformar dados de vendas B2B (Sellout e Registro de Oportunidades) provenientes de duas fontes CSV (`Base_B2B_MON...` e `Base_B2B_NOTE_TAB_MOB...`) em um **Modelo Dimensional Estrela** adequado para análise de Business Intelligence (BI) no MySQL. Os requisitos incluíam:

- Focar especificamente em transações do tipo `Sellout` e `Seltru` (Registro de Oportunidade).
- Padronizar e limpar identificadores chave como CNPJs (`Tier1_PartnerCode`, `Tier2_PartnerCode`, `EndCustomer_Code`).
- Ignorar colunas com dados vazios/nulos não essenciais para a análise.
- Garantir a capacidade de comparar dados de produtos `MON` (Monitores) e `MOB` (Mobile/Notebooks).
- Estruturar os dados para facilitar consultas analíticas e reports.

Dificuldades Encontradas Durante o Desenvolvimento:

- **Formato de Data Misto:** As datas nas bases de origem apresentavam-se em múltiplos formatos (serial numérico do Excel e string `DD/MM/YYYY`), exigindo uma lógica de transformação robusta para unificar para `YYYY-MM-DD`.
- **Inconsistência de Nomenclatura de Colunas:** A coluna `Branch` em uma das bases precisou ser explicitamente mapeada para `Tier2_PartnerName` para manter a consistência do modelo dimensional.
- **Manutenção de Integridade de Chaves:** A criação das chaves primárias e estrangeiras nas dimensões e na tabela fato exigiu atenção para garantir que os DataFrames fossem unidos corretamente após a deduplicação e transformação.

2. AÇÃO (Processo de ETL Desenvolvido)

Foi desenvolvido um script Python (`etl_b2b_vendas.py`) abrangendo as fases de ETL:

- **EXTRAÇÃO:** Leitura e consolidação das duas bases CSV (`Base_B2B_MON...` e `Base_B2B_NOTE_TAB_MOB...`) em um único DataFrame Pandas.
- **TRANSFORMAÇÃO:**
 - Seleção das colunas de interesse (`COLUMNAS_CHAVE`).
 - Limpeza e padronização dos CNPJs (remoção de caracteres especiais e preenchimento com zeros à esquerda para 14 dígitos).
 - Conversão e tratamento robusto da coluna `SalesDate` para o formato `YYYY-MM-DD`.
 - Aplicação de filtros de negócio: apenas `SalesType` 'Sellout' e 'Seltru'; exclusão de `Distributor` 'NAGEM_BRANDSHOP'; e foco nos `Type` 'MON', 'LFD', 'MOBILE'.
 - Criação de 5 tabelas de dimensão (`Dim_Data`, `Dim_Tipo_Venda`, `Dim_Produto`, `Dim_Parceiro`, `Dim_Cliente_Final`) com chaves primárias (`_key`).
 - Construção da tabela Fato (`Fato_Vendas_Oportunidades`) com as chaves estrangeiras (`_key`) referenciando as dimensões e as métricas (`QTY`, `Invoice_Number`, `Projeto`).
- **CARGA (LOAD - Testado Localmente):**
 - O script foi configurado para carregar os DataFrames diretamente nas tabelas MySQL (`projeto_b2b`) utilizando `sqlalchemy` e `mysql-connector-python`.
 - Backup dos dados transformados em `dados_b2b_processados.parquet` (Fato) e `dimensoes_b2b.xlsx` (Dimensões em abas separadas) foi gerado como contingência.

3. RESULTADO (Output do ETL)

O processo de ETL gerou com sucesso:

- **Modelo Dimensional Estrela Consistente:** As 6 tabelas (5 Dimensões e 1 Fato) foram modeladas conforme os requisitos, com `Fato_Vendas_Oportunidades` contendo 108.000+ registros de eventos de venda/oportunidade.
- **Dados Limpos e Padronizados:** CNPJs uniformizados, datas corretas e filtros aplicados, resultando em um conjunto de dados pronto para análise e elaboração de *Dashboards*.
- **Atendimento aos Requisitos de Negócio:**
 - **Foco Principal:** A `Dim_Tipo_Venda` isola `Sellout` e `Seltru`.
 - **Identificadores:** `Tier1_PartnerCode`, `Tier2_PartnerCode`, `EndCustomer_Code`, PN, QTY estão presentes e tratados nas tabelas.
 - **Tipos de Produto:** A `Dim_Produto` permite a análise por `Type` (MON/LFD/MOBILE).
- **Arquivos de Saída:**
 - `etl_b2b_vendas.py`: Script Python finalizado.
 - `modelo_dimensional_b2b.sql`: Script SQL para criação da estrutura das tabelas.
 - `dados_b2b_processados.parquet`: Backup da Fato.
 - `dimensoes_b2b.xlsx`: Backup das Dimensões.