

# Relatório Explicativo do Projeto ETL

## Introdução

Este relatório detalha as etapas do pipeline ETL desenvolvido para o projeto da Neotass Marketing. O objetivo foi consolidar, transformar e estruturar dados de oportunidades de negócios e vendas de parceiros em um banco de dados dimensional, facilitando análises avançadas e geração de insights estratégicos.

## 1. Etapas do ETL

A extração foi realizada a partir de dois arquivos:

- registros\_oportunidades.json: contém informações de oportunidades de negócios.
- sellout.parquet: inclui dados detalhados das vendas realizadas (sellout).

O pandas foi utilizado para carregar e validar os dados dessas fontes.

Na etapa de transformação, os seguintes passos foram realizados:

- Normalização de nomes de colunas para facilitar o acesso.
- Conversão de datas para o formato datetime, assegurando consistência temporal.
- Cálculo de valores totais (quantidade \* preço unitário) para ambos os datasets.
- Adição de identificadores únicos (id) para cada linha dos dados transformados.

Os dados transformados foram carregados em um banco de dados MySQL, utilizando as tabelas fato\_registro\_oportunidade e fato\_sellout. Inserções foram feitas linha a linha utilizando conexões estabelecidas com a biblioteca mysql.connector.

## 2. Desafios Enfrentados

## **Relatório Explicativo do Projeto ETL**

Os principais desafios incluíram:

- Tratamento de dados inconsistentes no dataset de oportunidades.
- Ajustes na performance ao lidar com arquivos grandes.
- Garantia da integridade das conexões ao banco de dados durante a etapa de carga.

### **3. Decisões Tomadas**

As principais decisões incluem:

- Uso do Airflow para orquestração, permitindo a automação do pipeline.
- Implementação do modelo dimensional para facilitar consultas analíticas.
- Adoção do MySQL como base de dados devido à familiaridade e robustez para o contexto.

### **Conclusão**

O pipeline ETL desenvolvido cumpre o objetivo de integrar e transformar os dados em um formato adequado para análises estratégicas. As etapas foram projetadas para assegurar flexibilidade, automação e escalabilidade.