**Documento de ETL - Fluxo de Dados de Alunos da Filial de Tecnologia**

**INTRODUÇÃO**

Este documento apresenta o detalhamento do processo de Extração, Transformação e Carga (ETL) dos dados referentes ao fluxo de alunos, cursos cursados e valores médios pagos mensalmente pela empresa filial de tecnologia. Os dados são recebidos em arquivos CSV e processados usando o Power BI, sendo posteriormente armazenados em um banco de dados MySQL.

1. Extração de Dados

Os dados são coletados mensalmente em arquivos CSV enviados pela empresa filial de tecnologia. Esses arquivos contêm informações detalhadas sobre o fluxo de alunos, os cursos que estão cursando e os valores médios que estão pagando. Eles são recebidos por meio de um canal seguro e são prontamente disponibilizados para a etapa de extração.

2. Transformação de Dados

Durante a fase de transformação, os dados passam por um processo rigoroso para garantir sua consistência e integridade. Primeiramente, são realizadas operações de limpeza para remover quaisquer dados duplicados, inconsistentes ou inválidos. Em seguida, são aplicadas diversas transformações, como a padronização de formatos, a conversão de valores e a agregação de dados para gerar métricas relevantes.

As métricas incluem não apenas a quantidade de alunos novos e os valores médios arrecadados e gastos, mas também informações específicas de cada entidade:

* Alunos: Bairro, estado civil e gênero.
* Produto: Categoria, professor, valor e nome do produto.
* Venda: Faturamento e quantidade.

3. Carga de Dados

Após a conclusão da etapa de transformação, os dados são prontamente carregados em um banco de dados MySQL local. Isso é feito utilizando um script SQL pré-definido que cria as tabelas necessárias para armazenar os dados. As tabelas incluem informações sobre clientes (alunos), produtos (cursos) e vendas (transações de cursos). Este processo de carga é automatizado e garante que os dados estejam disponíveis para consultas e análises em tempo hábil.

4. Monitoramento de Dados

O monitoramento contínuo dos dados é essencial para garantir a qualidade e a consistência dos dados processados. Assim que os arquivos CSV são recebidos, eles são submetidos a uma verificação detalhada para detectar possíveis duplicatas, inconsistências nos dados e a presença de campos novos ou ausentes. Isso é feito utilizando ferramentas de monitoramento automatizadas que alertam a equipe responsável sobre quaisquer problemas identificados.

Conclusão

Este documento fornece uma visão abrangente do processo de ETL utilizado para os dados da empresa filial de tecnologia. Desde a extração dos arquivos CSV até o carregamento dos dados em um banco de dados MySQL, cada etapa é cuidadosamente projetada para garantir a integridade e a qualidade dos dados para análises futuras.