

# T1 – Data Architecture

FIAP – MBA – Data Science & Artificial Intelligence – 10DTS – 28/03/2025

## Alunos

- RM 358829 - Rodrigo Patricio Carvalho
- RM 358024 - Lucas Nascimento dos Santos
- RM 358921 - Társis Fortes Tavares
- RM 358285 – Rafael Henrique Gallo

## Link para o vídeo

- [https://youtu.be/ysG\\_tGnJjpw](https://youtu.be/ysG_tGnJjpw)

## Pipeline de dados – Entrega 2

O processo representado pelo nosso pipeline de dados, representa uma carga em uma tabela do nosso Data Warehouse, a partir dos dados oriundos do cadastro de clientes no sistema de Vendas.

O cadastro é realizado via sistema transacional, e toda madrugada via rotina batch, é exportado um arquivo .csv que será utilizado para que possamos realizar a carga na tabela de clientes de nosso Data Warehouse. Para nosso pipeline, vamos considerar que este arquivo .csv já foi extraído e disponibilizado para a leitura.

Em nosso pipeline, iremos fazer o uso de tabelas STAGE, que são tabelas usadas para armazenar temporariamente os dados extraídos de sistemas fontes antes de serem processados e carregados no Data Warehouse. O objetivo dessas tabelas é servir como uma área intermediária onde os dados são "limpos", transformados e validados antes de serem movidos para as tabelas de fato do DW.

## Fluxo do Pipeline de Dados

### 1ª etapa

Realizamos a limpeza da área de STAGE para que esta tabela possa receber novos dados, através do comando TRUNCATE TABLE no SQL Server.

## 2ª etapa

Realizamos a leitura do arquivo .csv que foi exportado do sistema, e carregamos estes dados do csv para a tabela STAGE no SQL Server, configurando a origem (csv), o destino dos dados (SQL) e o mapeamento dos campos.

## 3ª etapa

Realizamos algumas transformações antes de carregar as informações em nosso DW. Note que o endereço estava vindo do sistema/csv com um ponto separando a rua e o número (RUA A. 123). Realizamos a correção inserindo uma vírgula ao invés do ponto (RUA A, 123).

## 4ª etapa

Após os dados serem extraídos, carregados na STAGE e transformados, vamos de fato realizar a carga destes dados em nossa tabela de clientes em nosso Data Warehouse. Concluindo assim a simulação de nosso Pipeline de Dados.

É possível visualizar no vídeo todo o processo ETL do Pipeline de Dados sendo executado.

## Resumo do Processo

1. **Extração:** Dados são extraídos do sistema fonte;
2. **Carga nas Tabelas STAGE:** Dados brutos são armazenados nas tabelas STAGE;
3. **Transformação:** Dados nas tabelas STAGE são transformados para o formato desejado;
4. **Validação:** Verificação da qualidade e consistência dos dados;
5. **Carga no Data Warehouse:** Dados transformados são movidos para o Data Warehouse;
6. **Exclusão ou Arquivamento:** Dados nas tabelas STAGE podem ser removidos ou arquivados.