



**POLITÉCNICO
DE LEIRIA**

SISTEMAS DE APOIO À DECISÃO

Engenharia Informática, Regimes Diurno e Pós-Laboral

3º Ano, 1º Semestre, Ano Letivo 2025/2026

Implementação do Processo ETL em Data Warehouses

Objetivos

- Analisar documentação inerente a um Data Warehouse (DW)
- Usar programação com as linguagens SQL e PL/SQL para gerir o ciclo ETL num protótipo de DW

Competências a adquirir pelos estudantes

- Saber analisar e modificar a documentação associada a um projeto de DW
- Saber definir e implementar estruturas e mecanismos adequados à extração de fontes de dados organizadas segundo plataformas diversas
- Saber definir e implementar estruturas e mecanismos adequados para tratar conjuntos de dados fonte extraídos bem como o correspondente registo de problemas de qualidade encontrados
- Saber aplicar os vários métodos de carregamento de tabelas de dimensão e de factos

Cenário

A cadeia de supermercados *Super Barato* dedica-se à venda por retalho de produtos alimentares, de vestuário e de utilidades. A administração da cadeia requereu ao seu Departamento Informático um Sistema de Apoio à Decisão (SAD) para apoiar a tomada de decisões relativas à venda de produtos nas lojas. Com este intuito, o Departamento Informático assumiu a tarefa de produzir um protótipo de um *Data Warehouse*.

Depois de uma análise profunda às fontes de dados disponíveis e aos requisitos analíticos pretendidos para o SAD, foram produzidos os seguintes documentos:

- Mapa geral da arquitetura do DW
- Mapa Lógico de Dados
- Modelo lógico de dados da base de dados dimensional
- Algoritmos de processamento das fases do ciclo ETL

No seguimento da elaboração destes documentos, foi dado início à implementação de um protótipo do DW que se encontra ainda em estado de construção: todas as estruturas e alguns processos do ciclo ETL foram já implementados.

Nota: O código fornecido constitui uma simplificação da plataforma *ETL² - Educational Tool for Learning ETL*, que implementa de forma mais completa o protótipo do DW.

GRUPO I – EXTRAÇÃO

1. Analise o documento *mapa geral da arquitetura* e identifique as tarefas a realizar neste grupo.
2. Na sua conta execute o ficheiro `sad_install_dw_vendas.sql`. Interprete as estruturas e código criados após a execução do ficheiro.
3. De acordo com o mapa lógico de dados, planeie e depois complete as estruturas e processos da *Data Staging Area* para a correta extração dos dados das fontes.
4. Com a orientação do docente, teste os processos implementados neste grupo.

GRUPO II – TRANSFORMAÇÃO

1. Analise o documento *mapa geral da arquitetura* e identifique as tarefas a realizar neste grupo.
2. De acordo com os critérios definidos na aula prático-laboratorial:
 - a. Construa o `screen screen_product_weight` que identifique e registe erros de nulidade no atributo `liq_weight` nos produtos extraídos mais recentemente;
Nota: quaisquer erros de qualidade encontrados deverão ser registados no subsistema *Transformation Error Logger* e as linhas com erros de qualidade serão rejeitadas.
 - b. Edite o `screen screen_product_brands` para identificar e registar erros de correção no atributo `brand` nos produtos extraídos mais recentemente.
Nota: quaisquer erros de qualidade encontrados deverão ser registados no subsistema *Transformation Error Logger* mas as linhas com erros de qualidade não serão rejeitadas.
 - c. Complete o método de transformação que realiza as transformações definidas para os produtos, corrigindo os erros detetados pelo `screen` na coluna `brand`.
 - d. Construa o método de transformação que realiza as transformações definidas para promoções. Inclua o método criado no fluxo de execução da fase de transformação.
 - e. Construa o método de transformação que realiza as transformações definidas para os dados provenientes do IPMA. Inclua o método criado no fluxo de execução da fase de transformação.
 - f. Identifique um erro potencialmente grave que tenha sido ignorado na construção da tabela `t_lookup_calories`. Apresente possíveis soluções.

3. Com a orientação do docente, teste os processos implementados neste grupo.

Nota: a vista `v_last_iteration_info` pode ajudar a interpretar os erros de qualidade detetados pelos `screens` durante a última iteração do processo ETL.

GRUPO III – CARREGAMENTO

1. Analise o documento *mapa geral da arquitetura* e identifique as tarefas a realizar neste grupo.
2. Planeie e depois implemente os processos em falta para o correto carregamento de dados nas dimensões `t_dim_promotion`, `t_dim_product` e `t_dim_date`.
3. Planeie e depois implemente os processos em falta para o correto carregamento de dados na tabela de factos `t_fact_lineofsale`.
4. Com a orientação do docente, teste os processos implementados neste grupo.

Tarefas extra-aulas: Pipeline Clientes

GRUPO I - EXTRAÇÃO

1. Implemente os objetos em falta na etapa de extração, referentes ao pipeline dos clientes, de forma a garantir a correta extração dos dados.

GRUPO II - TRANSFORMAÇÃO

2. Implemente os objetos em falta na etapa de transformação, referentes ao pipeline dos clientes, de forma a garantir a correta transformação dos dados. Deverá também implementar os seguintes *screens* para garantir a qualidade dos dados:
 - a. `screen_customer_gender`: para identificar e registrar erros de consistência no atributo *gender*.
 - b. `screen_customer_marital_status`: para identificar e registrar erros de clareza no atributo *marital_status*.

GRUPO III - CARREGAMENTO

3. Implemente os objetos em falta na etapa de carregamento, referentes ao pipeline dos clientes, de forma a garantir o correto carregamento dos dados.