

# RELATÓRIO

**Data Science & Business Analytics – 13.ª Edição**  
Data Warehousing



André Garcia (31811) | Dário Sequeira (31772) | Mariana António (31635) | Rodrigo Relvas (31783)

O presente relatório sintetiza os passos essenciais do desenvolvimento do projeto “Data Integration”, destacando a utilização do *Azure Data Factory* (ADF). Este projeto teve como objetivo principal a criação de um processo ETL (*Extract, Transform, Load*) para a ingestão, transformação e carregamento de dados, permitindo a sua posterior exploração através de um relatório em *Power BI*.

Os principais passos do projeto foram os seguintes:

1. Começámos por criar um *pipeline* no ADF (*'g15\_PL\_Ingest'*) para ingerir os dados armazenados em formato tabular (csv), os quais foram carregados manualmente num *container* em *Azure Data Lake Storage*;
2. Em seguida, criámos tabelas SQL no *Data Warehouse* para posterior carregamento dos dados transformados, procurando replicar o modelo dimensional normalizado fornecido. As tabelas criadas incluem uma tabela de *staging* e tabelas de dimensão, de ponte e de factos;
3. Em seguida, aplicámos processos ETL, implementando atividades de *pipeline* para importar os dados para as tabelas no *Data Warehouse*, criando atividades de *data flow* para processar e carregando os dados da *staging* para as tabelas finais no *Data Warehouse*;
4. Por fim, para completar o processo ETL, desenvolvemos um *pipeline* final (*'g15\_PL\_Master'*) para orquestrar a chamada de todos os *pipelines* e respetivos *data flows* previamente criados.

Todos os artefactos criados no âmbito deste projeto (*Pipelines, DataSets e DataFlows*) foram submetidos conforme as especificações do enunciado, tendo sido guardados na pasta “g15”.