

Trabalho Prático 1

Integração de Sistemas de Informação

DIOGO MACHADO Nº26042

Instituto Politécnico do Cávado e do Ave

19 de outubro de 2025

Índice

1	Introdução	3
2	Estratégia Utilizada	4
2.1	Ferramentas e Tecnologias	4
2.2	Pré-requisitos e dependências	4
3	Fluxo Principal (Job)	5
4	Descrição das Transformações	6
4.1	ImportHoteis.ktr	6
4.2	ImportClientes.ktr	7
4.3	ImportReservas.ktr	8
4.4	IntegrarDados.ktr	9
4.5	GeraGraficos.ktr	10
4.6	EnviaEmail.ktr	11
5	Fluxo node-red	11
5.1	Fluxo de dados aleatórios	12
5.2	Fluxo de gerar gráficos	13
6	Configuração de caminhos de ficheiros	14
7	Conclusão	14
8	Repositório Git	15
9	Visualização do video	15

Lista de Figuras

1	Fluxo principal (Job)	5
2	Transformação dos hotéis	6
3	Transformação dos clientes	7
4	Transformação das reservas	8
5	Transformação das reservas	9
6	Transformação de gerar gráficos	10
7	Transformação de enviar email	11
8	Fluxo node-red para gerar ficheiros	12
9	Fluxo node-red para gerar graficos	13
10	QR code	15

1 Introdução

O presente trabalho foi desenvolvido no âmbito da unidade curricular de Integração de Sistemas de Informação, tendo como tema a gestão de reservas de hotéis. O objetivo principal consiste em aplicar e consolidar conhecimentos sobre processos de *ETL* (Extract, Transform, Load), explorando a integração e transformação de dados provenientes de diferentes fontes.

Para a concretização deste projeto, foram utilizados três conjuntos de dados distintos — *reservas*, *hotéis* e *clientes* — com o intuito de demonstrar a importância da integração de informação dispersa num sistema coerente e estruturado. Ao longo do trabalho, são aplicadas diversas operações de tratamento, limpeza e normalização de dados, de modo a garantir a consistência e a fiabilidade da informação processada.

Este projeto permite, assim, compreender de forma prática como os processos de *ETL* podem ser utilizados para suportar a integração de sistemas no contexto da hotelaria, potenciando uma melhor análise e gestão das reservas e clientes. O trabalho contribui ainda para o reforço das competências técnicas e metodológicas associadas ao desenvolvimento de soluções de integração de dados.

2 Estratégia Utilizada

A estratégia adotada para o desenvolvimento deste trabalho centrou-se na implementação de um processo de integração de dados aplicado ao contexto da gestão de reservas de hotéis. O principal objetivo consistiu em construir um fluxo de *ETL* (Extract, Transform, Load) capaz de consolidar informação proveniente de diferentes fontes, garantindo a consistência, qualidade e utilidade dos dados.

2.1 Ferramentas e Tecnologias

Para a realização deste trabalho foram utilizadas as seguintes ferramentas e tecnologias:

- **Pentaho Kettle (PDI – Pentaho Data Integration)**: utilizado para a criação e execução dos processos de *ETL*, responsáveis pela extração dos dados dos três ficheiros principais — *reservas*, *hotéis* e *clientes* —, pela sua transformação (limpeza, normalização e junção) e posterior carregamento num formato integrado e estruturado.
- **Node-red**: utilizado para a visualização dos dados processados e para a geração de dados aleatórios, de forma a simular um cenário mais próximo da realidade.

2.2 Pré-requisitos e dependências

Para o desenvolvimento do projeto, foi necessária a instalação de algumas ferramentas e bibliotecas essenciais. Foi instalado o **Node-RED Dashboard**, que permitiu a visualização dos dados em gráficos interativos diretamente na interface web do Node-RED. A instalação pode ser realizada através do comando:

```
npm install node-red-dashboard
```

Além disso, para a geração de gráficos adicionais a partir dos dados processados, foi necessária a instalação da biblioteca **Matplotlib** no ambiente Python. Esta biblioteca está incluída no ficheiro `requirements.txt`, permitindo que todas as dependências Python sejam instaladas de forma automática com o comando:

```
pip install -r requirements.txt
```

Desta forma, é possível garantir que todas as bibliotecas necessárias para a execução dos scripts de geração de gráficos estejam corretamente configuradas.

3 Fluxo Principal (Job)

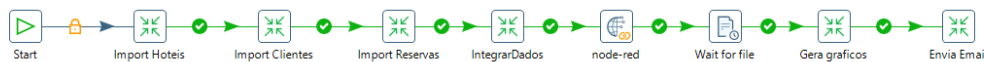


Figura 1: Fluxo principal (Job)

O fluxo principal representa um processo automatizado de integração e processamento de dados. O seu objetivo é importar, consolidar, tratar e comunicar informações através de diferentes etapas sequenciais:

- **Import Hoteis** – Importa os dados dos hotéis de um ficheiro XML.
- **Import Clientes** – Importa a informação relativa aos clientes de um ficheiro JSON.
- **Import Reservas** – Importa os dados das reservas efetuadas de um ficheiro CSV.
- **IntegrarDados** – Integra e consolida os dados importados, estabelecendo relações entre hotéis, clientes e reservas.
- **node-red** – Executa um fluxo no Node-RED para gerar gráficos com os dados transformados e gerar também ficheiros JSON que serão utilizados de seguida no script *python* para gerar gráficos localmente.
- **Wait for file** – Aguarda a chegada de um ficheiro externo do node-red para gerar novos gráficos.

- **Gera gráficos** – Gera representações gráficas a partir dos dados processados.
- **Envia Email** – Envia um email automático com os gráficos gerados em anexo.

Este job representa um pipeline completo de integração de dados, desde a importação até à comunicação dos resultados.

4 Descrição das Transformações

4.1 ImportHoteis.ktr

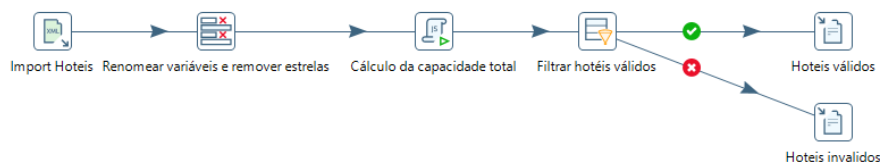


Figura 2: Transformação dos hotéis

- **Import Hoteis** – Importa os dados brutos de hotéis a partir de um ficheiro externo em formato XML.
- **Renomear variáveis e remover estrelas** – Efectua a limpeza e normalização dos dados, renomeando colunas e removendo o campo de classificação por estrelas.
- **Cálculo da capacidade total** – Calcula a capacidade total de cada hotel, no caso multiplica o número de quartos por 2 considerando que cada quarto tem capacidade para duas pessoas.
- **Filtrar hotéis válidos** – Verifica se o número de quartos do hotel é superior a 0.
- Os **hotéis válidos** são exportados para um ficheiro designado *Hotéis válidos*.
- Os **hotéis inválidos** são guardados num ficheiro separado chamado *Hotéis inválidos*.

4.2 ImportClientes.ktr

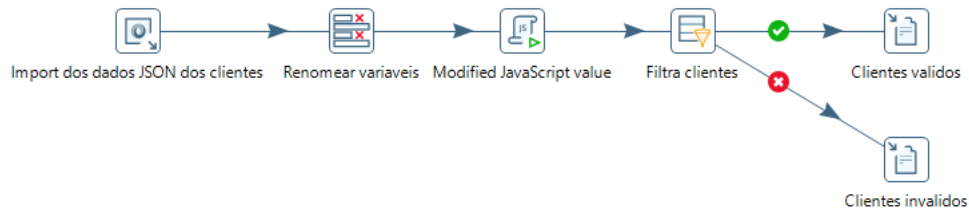


Figura 3: Transformação dos clientes

- **Import dos dados JSON dos clientes** – Importa a informação dos clientes a partir de um ficheiro no formato JSON.
- **Renomear variáveis** – Normaliza os nomes das variáveis, tornando-os consistentes e adequados ao modelo de dados interno.
- **Modified JavaScript value** – Verifica se o email do utilizador é válido ou não.
- **Filtra clientes** – Filtra os registos de clientes de acordo com critérios de validade.
- Os **clientes válidos** são exportados para um ficheiro denominado *Cientes válidos*.
- Os **clientes inválidos** são guardados num ficheiro separado chamado *Cientes inválidos*.

4.3 ImportReservas.ktr

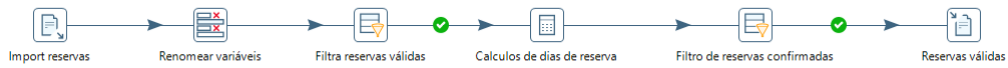


Figura 4: Transformação das reservas

- **Importar reservas** – Importa os dados das reservas a partir de um ficheiro csv.
- **Renomear variáveis** – Normaliza os nomes das colunas, tornando-os consistentes e adequados ao processamento interno.
- **Filtrar reservas válidas** – Filtra os registos de reservas com preços superiores a 0.
- **Cálculo de dias de reserva** – Calcula o número de dias de cada reserva com base nas datas de check-in e check-out.
- **Filtrar reservas confirmadas** – Mantém apenas as reservas que estão confirmadas.
- **Reservas válidas** – O resultado final é um conjunto de reservas válidas e confirmadas, pronto para utilização ou exportação.

4.4 IntegrarDados.ktr

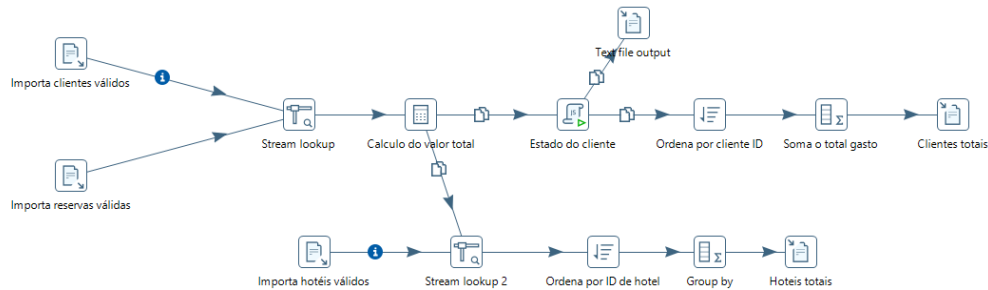


Figura 5: Transformação das reservas

Esta transformação tem como objetivo integrar e analisar dados de clientes, reservas e hotéis, de forma a calcular o valor total gasto por cliente e as estatísticas agregadas por hotel.

- **Importa clientes válidos** – Importa os registos de clientes previamente validados.
- **Importa reservas válidas** – Importa as reservas consideradas válidas após o processo de filtragem.
- **Stream lookup** – Junta os dados de clientes e reservas com base no identificador comum, permitindo combinar informação relevante de ambos.
- **Cálculo do valor total** – Calcula o valor total associado a cada reserva, considerando as variáveis necessárias (como número de dias e preço).
- **Importa hotéis válidos** – Importa os dados dos hotéis previamente validados.
- **Stream lookup 2** – Associa as reservas aos respetivos hotéis, enriquecendo os dados com informação do estabelecimento.
- **Estado do cliente** – Verifica se o cliente é válido ou inválido.
- **Ordena por cliente ID** – Organiza os dados por identificador de cliente para facilitar o agrupamento e cálculo posterior.

- **Soma o total gasto** – Calcula o total gasto por cliente, consolidando todas as suas reservas.
- **Clientes totais** – Exporta o resultado agregado por cliente, representando o total gasto.
- **Ordena por ID de hotel** – Ordena os registos por identificador de hotel para o processamento seguinte.
- **Group by** – Soma o valor ganho por cada hotel com as reservas.
- **Hotéis totais** – Gera o resultado final com as estatísticas globais de cada hotel.
- **Text file output** – Exporta os dados das reservas para um ficheiro.

4.5 GeraGraficos.ktr



Figura 6: Transformação de gerar gráficos

- **JSON input** – Importa os dados dos gráficos gerados pelo node-red.
- **Add constants** – Cria constantes com o valor do comando a ser executado no nó a seguir, neste caso o comando *python*, e o path do ficheiro *.py* a ser executado.
- **Gera graficos** – Executa o script python que cria os gráficos.

4.6 EnviaEmail.ktr



Figura 7: Transformação de enviar email

- **Get file names** – Importa o nome dos ficheiros com os gráficos em .png.
- **Add constants** – Cria constantes com o que é necessário para enviar o email.
- **Mail** – Envia email com as imagens dos gráficos em anexo.

5 Fluxo node-red

Foram desenvolvidos dois fluxos distintos na plataforma Node-RED com objetivos complementares. O primeiro fluxo tem como finalidade a **criação de ficheiros com dados aleatórios**, simulando informação de clientes, reservas e hotéis, utilizada posteriormente para análise e visualização. O segundo fluxo é responsável pela **geração de gráficos** a partir desses dados, permitindo representar visualmente as receitas por hotel e o total gasto por cliente. Este último fluxo inicia-se através de um pedido HTTP e integra etapas de leitura, conversão, tratamento e exportação dos dados em formato JSON, assegurando a automatização do processo de criação dos gráficos.

5.1 Fluxo de dados aleatórios

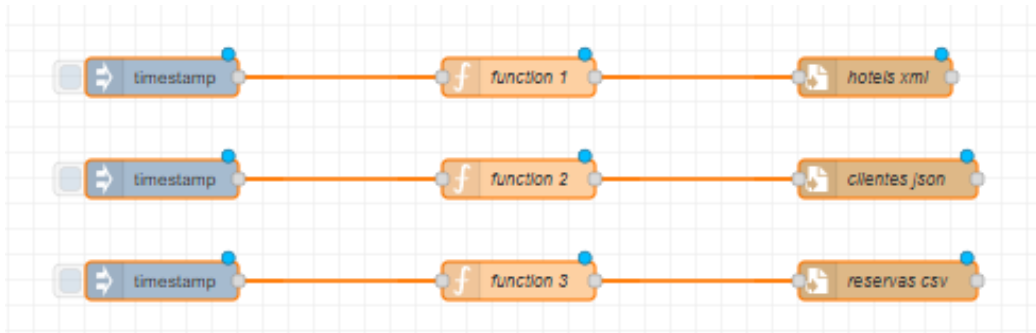


Figura 8: Fluxo node-red para gerar ficheiros

Neste caso será descrito apenas o fluxo de geração de dados dos **hotéis**, uma vez que os fluxos relativos aos **clientes** e às **reservas** apresentam a mesma estrutura e funcionamento, variando apenas no tipo de dados gerados.

- **Timestamp** – Inicia o processo de geração de dados de forma manual, permitindo executar o fluxo sempre que necessário.
- **Function 1** – Cria dados aleatórios referentes aos hotéis, simulando atributos como identificadores, nomes, estrelas, país, cidade e número de quartos.
- **Hotels XML** – Exporta os dados gerados para um ficheiro no formato XML, armazenando a informação simulada para utilização posterior nos restantes fluxos.

5.2 Fluxo de gerar gráficos

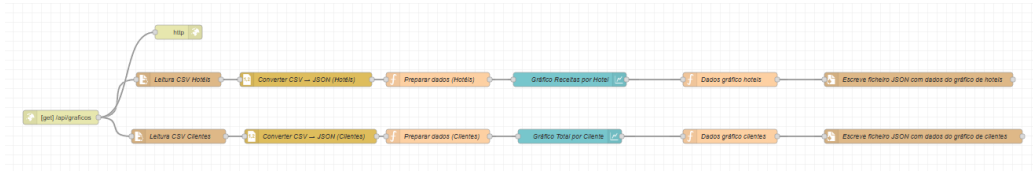


Figura 9: Fluxo node-red para gerar graficos

Neste ponto será apresentado apenas o fluxo correspondente aos **hotéis**, uma vez que o processo relativo aos **clientes** é idêntico em estrutura e funcionamento, diferindo apenas nos dados utilizados.

- **HTTP GET /apigraficos** – Inicia o fluxo quando é feita uma requisição HTTP para o endpoint definido, ativando o processo de geração dos gráficos.
- **Leitura CSV Hotéis** – Lê o ficheiro CSV contendo a informação dos hotéis, incluindo dados relevantes como identificadores e receitas.
- **Converter CSV → JSON (Hotéis)** – Converte os dados do ficheiro CSV para o formato JSON, facilitando o seu tratamento nas etapas seguintes.
- **Preparar dados (Hotéis)** – Processa e organiza os dados dos hotéis, calculando os valores necessários para gerar o gráfico.
- **Gráfico Receitas por Hotel** – Cria o gráfico que representa as receitas obtidas por cada hotel com base nos dados processados.
- **Dados gráfico hotéis** – Extrai e formata os dados do gráfico, preparando-os para serem guardados num ficheiro.
- **Escreve ficheiro JSON com dados do gráfico de hotéis** – Exporta os dados do gráfico para um ficheiro em formato JSON, permitindo o seu uso posterior.

6 Configuração de caminhos de ficheiros

Nos nós de entrada e saída de ficheiros do Node-RED foram definidos *paths* absolutos, o que significa que os ficheiros são lidos e gravados num local específico do computador onde o projeto foi desenvolvido.

Para garantir o correto funcionamento do programa noutros dispositivos, estes caminhos devem ser alterados de acordo com a estrutura de pastas existente no computador em que o sistema for executado. Por exemplo, atualmente os fluxos utilizam caminhos como:

```
C:\Users\Diogo Machado\Documents\GitHub\ISI\data\output\hoteis_totais.csv
```

Este caminho deve ser ajustado para a diretoria correspondente no novo computador, assegurando que o Node-RED consegue aceder corretamente aos ficheiros de entrada e criar os ficheiros de saída.

7 Conclusão

O desenvolvimento deste trabalho permitiu compreender de forma prática e integrada o funcionamento dos processos de ETL (*Extract, Transform, Load*) e a sua importância na consolidação de dados provenientes de diferentes fontes. Através da utilização das ferramentas Pentaho Data Integration e Node-RED, foi possível construir um sistema completo, capaz de importar, transformar, integrar e visualizar dados de forma automatizada.

A implementação dos fluxos desenvolvidos demonstrou a relevância da normalização e validação da informação, assegurando a consistência dos resultados obtidos. A criação de ficheiros com dados aleatórios e a posterior geração de gráficos permitiram simular um cenário realista de gestão de reservas hoteleiras, evidenciando a utilidade prática dos processos de integração de dados na análise e apoio à decisão.

Para além dos aspetos técnicos, este projeto contribuiu também para o reforço de competências na configuração e interoperabilidade entre diferentes ferramentas, nomeadamente na comunicação entre o Pentaho e o Node-RED. A atenção aos detalhes, como a definição de *paths* e dependências externas, revelou-se essencial para garantir a portabilidade e o correto funcionamento do sistema em diferentes ambientes.

Em suma, o trabalho cumpriu os objetivos propostos, demonstrando a aplicação eficaz dos conceitos de integração de sistemas de informação e

destacando o potencial das ferramentas de ETL e automação na gestão e visualização de dados empresariais.

8 Repositório Git

O repositório de GitHub pode se encontrar aqui.

9 Visualização do video

Para visualizar o vídeo com a demonstração faça scan do seguinte QR code:

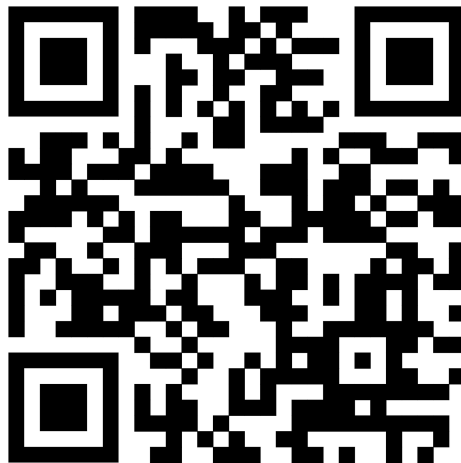


Figura 10: QR code