



# **Relatório Integração de Sistemas de Informação**

## **Trabalho Prático I**

No.23000 – Diogo Marques

**Licenciatura em Engenharia Sistemas Informáticos**

**3ºano**

Barcelos | Outubro, 2024

# **Relatório Integração de Sistemas de Informação**

---

## **Índice de Figuras**

Figura 1 Diagrama do tratamento de dados .....	7
Figura 2 Diagrama do JOB .....	8
Figura 3 QR Code demonstração projeto.....	9

# **Relatório Integração de Sistemas de Informação**

---

## **Índice**

1. Introdução .....	4
2. Problema .....	5
3. Estratégia utilizada .....	6
4. Transformação .....	7
5. Job .....	8
6. Víde com demonstração .....	9
7. Conclusão e Trabalhos futuros .....	10

# Relatório Integração de Sistemas de Informação

---

## 1. Introdução

O presente relatório aborda o primeiro trabalho prático realizado em contexto da unidade curricular de Integração de Sistemas de Informação. O objetivo principal deste projeto é a aplicação prática de ferramentas e técnicas de ETL (*Extract, Transformation, Load*) em processos de integração de dados, através da utilização de ferramentas como o Pentaho.

Este trabalho aborda também o papel crucial do ETL em diversos contextos empresariais e em ambientes inteligentes, onde a integração de dados e sistemas é cada vez mais necessária para enfrentar os desafios modernos de análise, migração e processamento de grandes volumes de dados.

# Relatório Integração de Sistemas de Informação

---

## 2. Problema

O mercado de carros usados na Espanha gera uma vasta quantidade de dados que, quando bem estruturados e analisados, podem oferecer insights valiosos para empresas, compradores e vendedores. No entanto, esses dados muitas vezes apresentam problemas de inconsistência, duplicação e falta de padronização, tornando o processo de análise e integração mais desafiador.

Este projeto visa o tratamento e transformação de um arquivo CSV contendo dados de carros usados à venda na Espanha, usando processos de ETL para preparar a informação para análise e potencial integração com outras plataformas. Os principais objetivos incluem a normalização e limpeza dos dados, remoção de caracteres especiais, tratamento de valores ausentes, e transformação de campos de forma a garantir consistência. Com isso, busca-se possibilitar uma análise mais precisa sobre preços, quilometragem, modelos, entre outros atributos dos veículos, facilitando a obtenção de insights e auxiliando no processo de tomada de decisão.

Esse processo de ETL simulado ilustra a importância de práticas de integração e tratamento de dados no contexto de mercados dinâmicos, como o de veículos usados, e demonstra o impacto positivo dessas técnicas na qualidade e confiabilidade das informações.

# Relatório Integração de Sistemas de Informação

---

## 3. Estratégia utilizada

Para este projeto de ETL, foi utilizado o Pentaho como ferramenta principal para o tratamento e transformação dos dados extraídos do arquivo CSV, que contém informações sobre carros usados à venda na Espanha. A estratégia adotada pode ser descrita nas seguintes etapas:

- **Extração e Tratamento de Dados:** Os dados foram inicialmente extraídos do arquivo CSV, sendo submetidos a processos de normalização no Pentaho para garantir a consistência e qualidade das informações. Durante esta etapa, foram aplicadas operações como remover caracteres especiais, preencher valores ausentes e padronizar os dados, garantindo que todas as informações estivessem no formato adequado para análise e armazenamento.
- **Armazenamento e Exportação:** Após o tratamento, os dados foram salvos em XML e JSON, e em uma base de dados PostgreSQL. Estes formatos permitem uma maior flexibilidade de integração com outras aplicações e facilitam a visualização e análise de dados por diferentes ferramentas.
- **Automação e Relatórios:** Para tornar o processo mais automatizado, foi criado um job no Pentaho que executa o pipeline de tratamentos de dados e, no final, envia um email com os arquivos gerados (XML, JSON e logs do tratamento). Esse processo assegura o envio de relatórios por email para monitoramento da execução e qualidade dos dados.
- **Análise e Visualização de Dados:** Com os dados armazenados e tratados, foi desenvolvida uma dashboard no Power BI para análise e visualização das principais métricas e insights do mercado de carros usados.

Essa estratégia integrada facilita o fluxo de dados desde a extração até a análise, assegurando que o processo de ETL seja automatizado, eficiente e pronto para integração com sistemas de análise e relatórios.

# Relatório Integração de Sistemas de Informação

## 4. Transformação

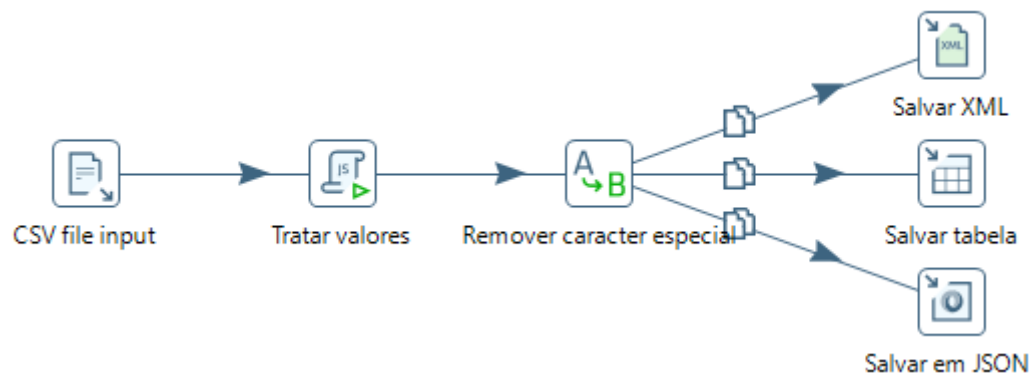


Figura 1 Diagrama do tratamento de dados

O diagrama acima representa o fluxo de transformação de dados realizado no Pentaho. Abaixo estão detalhadas as principais operações realizadas em cada etapa:

- **CSV file input:** Esta etapa lê o arquivo CSV que contém os dados dos carros usados à venda na Espanha. A partir deste arquivo, são extraídas as informações que serão processadas nas etapas seguintes.
- **Tratar valores:** Após a leitura dos dados, é realizada uma etapa inicial, que inclui a normalização de valores e preenchimento de valores ausentes. Este processo garante que os dados estejam completos e formatados corretamente para as etapas posteriores.
- **Remover caracteres especiais:** Esta etapa remove os caracteres especiais presentes nos campos de versão e modelo dos veículos com o uso de um regex. Este tratamento é importante para garantir a compatibilidade dos dados com os sistemas que irão utilizá-los posteriormente, como base de dados e sistemas de visualização.
- **Salvar XML:** Os dados são convertidos e salvos no formato XML, adequados para a integração com outros sistemas que utilizam este formato de dados.
- **Salvar Tabela (PostgreSQL):** Os dados são salvos diretamente em uma tabela de banco de dados, permitindo o armazenamento eficiente e a consulta rápida de grandes volumes de dados.
- **Salvar JSON:** Os dados são exportados em formato JSON, que é amplamente utilizado para integração com APIs e outras aplicações web.

# Relatório Integração de Sistemas de Informação

## 5. Job

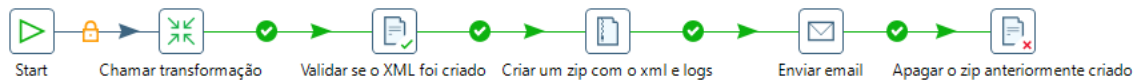


Figura 2 Diagrama do JOB

O diagrama acima representa o fluxo do job desenvolvido no Pentaho, responsável por executar o processo de transformação dos dados e, ao final, enviar um email com os arquivos gerados e os logs. As etapas deste processo são descritas a seguir:

- **Start:** Esta é a etapa inicial, onde o job e fluxo de execução são iniciados.
- **Chamar Transformação:** Nesta fase, o job chama e executa a transformação mencionada no ponto anterior, salvando as logs de execução da mesma.
- **Validar se o XML foi criado:** Após a transformação, é feita uma verificação para garantir que o arquivo XML foi criado corretamente.
- **Criar um ZIP com XML, JSON e logs:** Nesta etapa, os arquivos gerados durante a transformação são compactados em um arquivo ZIP e eliminados da pasta output. O arquivo gerado será posteriormente enviado em anexo no email.
- **Enviar email:** O job, envia o email contendo o arquivo zip gerado.
- **Apagar o ZIP anteriormente criado:** Por fim, o job remove o arquivo ZIP criado, evitando que na próxima execução o arquivo zip não seja enviado junto com os novos dados.



### 6. Víde com demonstração



*Figura 3 QR Code demonstração projeto*

# Relatório Integração de Sistemas de Informação

---

## 7. Conclusão e Trabalhos futuros

Este projeto de ETL, desenvolvido com a utilização do Pentaho, permitiu a implementação de um pipeline completo para o tratamento e análise de dados provenientes de um arquivo CSV com informações de carros usados à venda na Espanha. Através das etapas de extração, transformação e carregamento dos dados, foi possível não só normalizar e limpar as informações, como também exportá-las em formatos estruturados (XML, JSON e PostgreSQL), facilitando futuras análises e integrações.

Adicionalmente, o desenvolvimento de uma dashboard no Power BI permitiu visualizar de forma clara e interativa os dados tratados, destacando insights valiosos sobre o mercado de carros usados. O job automatizado, que executa o processo de ETL e envia por email os resultados e logs, reforça a eficiência e a escalabilidade deste sistema, eliminando a necessidade de intervenção manual e garantindo que os dados sejam tratados e distribuídos corretamente.

Com a execução deste projeto, ficou evidente a importância e a eficácia dos processos de ETL na integração de dados, demonstrando como estas práticas podem melhorar a qualidade da informação, a precisão das análises e a capacidade de integração com outras plataformas e sistemas. A automação do pipeline, aliada à flexibilidade de exportação e visualização dos dados, garante um processo robusto e confiável, pronto para ser aplicado em cenários reais de negócios.

No futuro, pode-se explorar a integração com APIs externas para enriquecer ainda mais os dados, além de otimizar o job para lidar com volumes maiores de informação, expandindo a escalabilidade e o impacto das soluções implementadas.