



Integração de Sistemas de Informação

Trabalho Prático I

Cláudio Fernandes - N° 30517

Docente

Luís Ferreira

Ano letivo 2025/2026

Licenciatura em Engenharia de Sistemas Informáticos

Regime Pós-laboral

Escola Superior de Tecnologia

Instituto Politécnico do Cávado e do Ave

RESUMO

O trabalho desenvolvido no âmbito da unidade curricular de Integração de Sistemas de Informação (ISI) teve como principal objetivo aplicar técnicas e ferramentas de ETL (Extract, Transform and Load), através da utilização do KNIME como plataforma principal de integração e o Node-RED para a visualização e monitorização dos resultados.

O projeto consistiu na extração de dados a partir de uma API pública, respetiva transformação e limpeza, e posteriormente carregamento e visualização em dashboards dinâmicos.

Durante o processo foram explorados operadores de transformação, expressões regulares, serialização de dados em JSON e XML, e integração entre diferentes plataformas.

Como resultado, foi implementado um pipeline completo e funcional, que demonstra a aplicação prática dos conceitos de integração de sistemas e automação de fluxos de dados.

ÍNDICE

1. Introdução	5
1.1. Motivação	5
1.2. Enquadramento	5
1.3. Objetivos	5
1.4. Metodologia de Investigação	5
1.5. Plano de trabalho	6
1.6. Estrutura do Documento	6
2. Estado da Arte / Enquadramento Teórico	7
2.1. Fundamentos Teóricos	7
2.2. Soluções Existentes	7
3. Trabalho Desenvolvido	8
3.1. Análise e Especificação	8
3.2. Implementação	9
3.2.1. Processo KNIME – Pipeline ETL	9
3.2.2. Processo Node-RED	13
4. Análise e Discussão de Resultados	17
4.1. Análise do Processo KNIME	17
4.1.1. Fiabilidade da Extração	17
4.1.2. Qualidade das Transformações	17
4.1.3. Exportação e Interoperabilidade	18
4.2. Análise do Processo Node-RED	18
4.2.1. Funcionalidade do Dashboard	18
4.2.2. Design e Usabilidade	18
4.2.3. Robustez e Fiabilidade	19
4.3. Integração KNIME–Node-RED	19
4.4. Discussão dos Resultados	19
Conclusão	20

ÍNDICE DE FIGURAS

Figura 1 - Fluxo ETL implementado no Knime	9
Figura 2 - Extract.....	10
Figura 3 - Transform.....	11
Figura 4 - Load	12
Figura 5 - Fluxo Node-RED	13
Figura 6 - Dashboard final do Node-RED (Futebol_ETL).....	14

ÍNDICE DE TABELAS

Tabela 1 - Tarefas e Ferramentas.....	6
Tabela 2 - Soluções existentes	7

1. Introdução

1.1. Motivação

Num contexto em que os dados são um ativo estratégico, as organizações necessitam de ferramentas capazes de integrar, transformar e visualizar informação proveniente de múltiplas fontes.

Este projeto surge como resposta à necessidade de consolidar competências em processos de ETL, com a exploração das tecnologias de integração visual low-code (KNIME e Node-RED) e abordar desafios reais de tratamento e interoperabilidade de dados.

1.2. Enquadramento

Este trabalho foi desenvolvido no âmbito da unidade curricular Integração de Sistemas de Informação, do curso de Licenciatura em Engenharia de Sistemas Informáticos (LESI-PL).

1.3. Objetivos

- Aplicar o paradigma ETL para integração de diversos dados.
- Implementar processos automatizados de extração, transformação e carga.
- Aplicar expressões regulares, limpeza e normalização de dados.
- Utilizar Node-RED para dashboard.

1.4. Metodologia de Investigação

Foi seguida uma abordagem experimental e iterativa, com fases distintas de planeamento, implementação, validação e avaliação dos resultados obtidos nas ferramentas KNIME e Node-RED.

1.5. Plano de trabalho

Fase	Atividade	Ferramenta	Resultado Esperado
1	Extração de dados da API	KNIME	Dataset bruto em JSON
2	Transformações e limpeza	KNIME	Dataset estruturado
3	Exportação e armazenamento	KNIME	Ficheiros JSON/XML/CSV/SQL
4	Visualização	Node-RED	Dashboard dinâmico
5	Testes e Validação	Ambos	Pipeline funcional
6	Documentação	Word/GitHub	Relatório e README

Tabela 1 - Tarefas e Ferramentas

1.6. Estrutura do Documento

O relatório encontra-se organizado em cinco capítulos principais, seguidos das referências e anexos.

2. Estado da Arte / Enquadramento Teórico

2.1. Fundamentos Teóricos

- Conceitos de ETL, Integração de Dados e Criação de Processos.
- Arquiteturas de integração: Batch vs. Streaming.
- Ferramentas de ETL: KNIME, Talend, Pentaho, SSIS.
- Visualização e automação com Node-RED.

2.2. Soluções Existentes

Plataforma	Tipo	Vantagens	Limitações
KNIME	ETL	Interface visual	Aprendizagem
Talend	ETL	Forte integração DB	Mais pesado
Node-RED	IoT / Dashboards	Simples, extensível, ideal para fluxos leves	Menos indicado para grandes volumes

Tabela 2 - Soluções existentes

Conclusão: KNIME e Node-RED oferecem o equilíbrio ideal entre simplicidade, integração visual e modularidade, sendo apropriados para um projeto académico de ETL com dashboards.

3. Trabalho Desenvolvido

3.1. Análise e Especificação

Requisitos

- Extrair dados de uma API pública (SportMonks).
- Realizar transformações e limpeza de dados com expressões regulares.
- Exportar resultados em formatos JSON e XML.
- Apresentar resultados em dashboard (Node-RED).

Arquitetura Lógica

- API → KNIME ETL → JSON/XML/SQL/CSV.
- Node-RED Dashboard

Modelação de Dados

- Tabelas/Views: Equipas, Jogos, Estatísticas.
- Campos principais: DataHora, ResultadoInfo, EquipaCasa, EquipaFora, Jogo

3.2. Implementação

3.2.1. Processo KNIME – Pipeline ETL

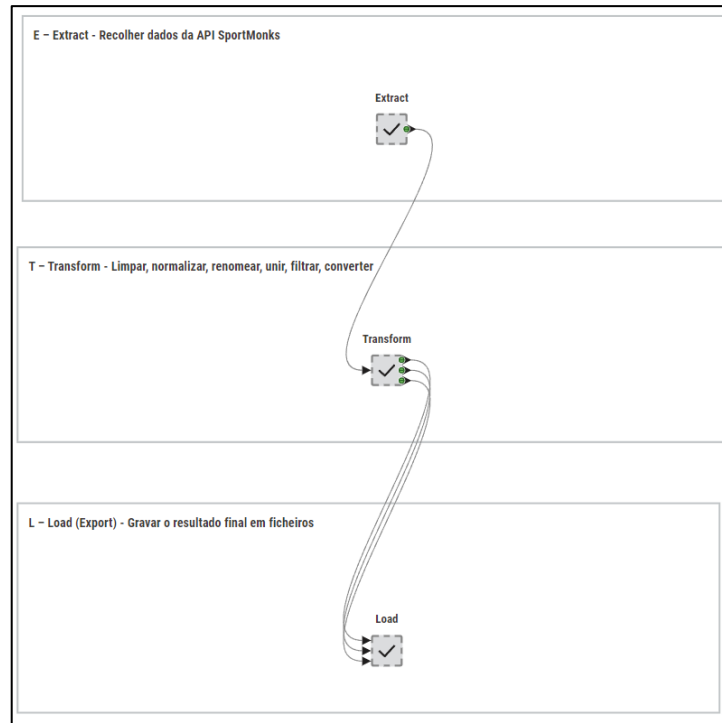


Figura 1 - Fluxo ETL implementado no Knime

O fluxo foi desenvolvido no KNIME para executar todas as fases do processo ETL (Extract, Transform and Load), desde a recolha inicial dos dados através de uma API REST até à exportação final em múltiplos formatos (CSV, JSON, XML e SQL), assegurando a interoperabilidade entre sistemas e a integração direta com o Node-RED.

Extract – Extração dos Dados

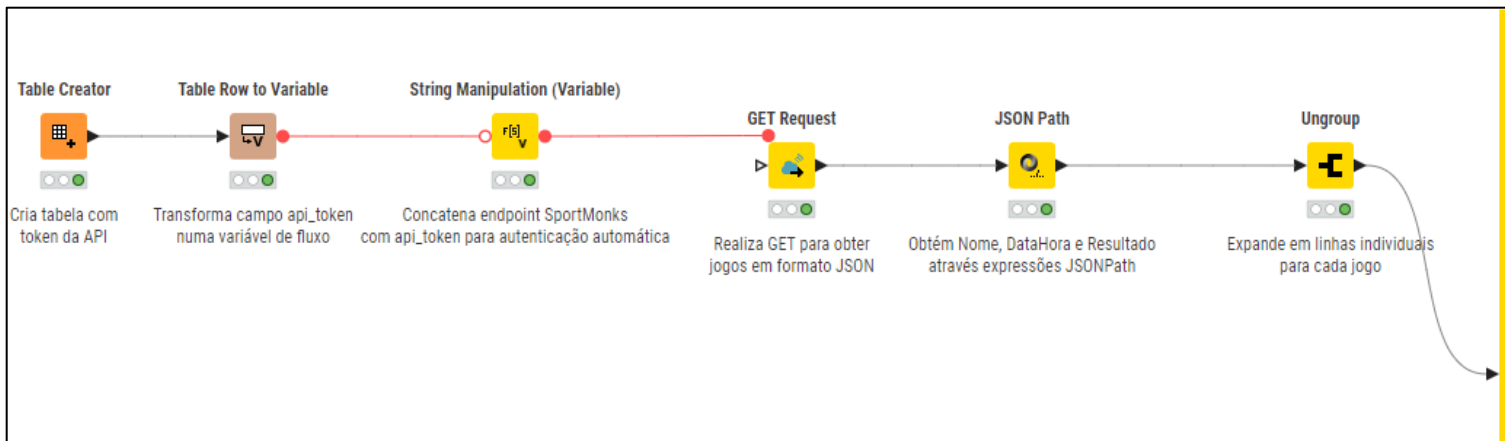


Figura 2 - Extract

A primeira etapa do pipeline é responsável pela extração de dados da API SportMonks Football, que fornece informações sobre ligas, equipas, jogos e resultados.

Para tal, foram utilizados os nós API Token, GET Request, JSON Path e Ungroup:

- API Token – Define o token de autenticação necessário para aceder à API, garantindo uma comunicação segura e autorizada.
- GET Request – Realiza a chamada HTTP ao endpoint configurado (por exemplo, `/fixtures?include=teams`), obtendo os dados em formato JSON.
- JSON Path – Extrai apenas os campos relevantes, como o nome das equipas, a data e o resultado de cada jogo.
- Ungroup – Converte listas aninhadas em linhas individuais, permitindo o tratamento granular de cada registo de jogo.

Através desta etapa, foi obtido um dataset completo com todos os jogos disponíveis, pronto para as etapas seguintes de limpeza e transformação.

Transform – Limpeza, Normalização e Estruturação

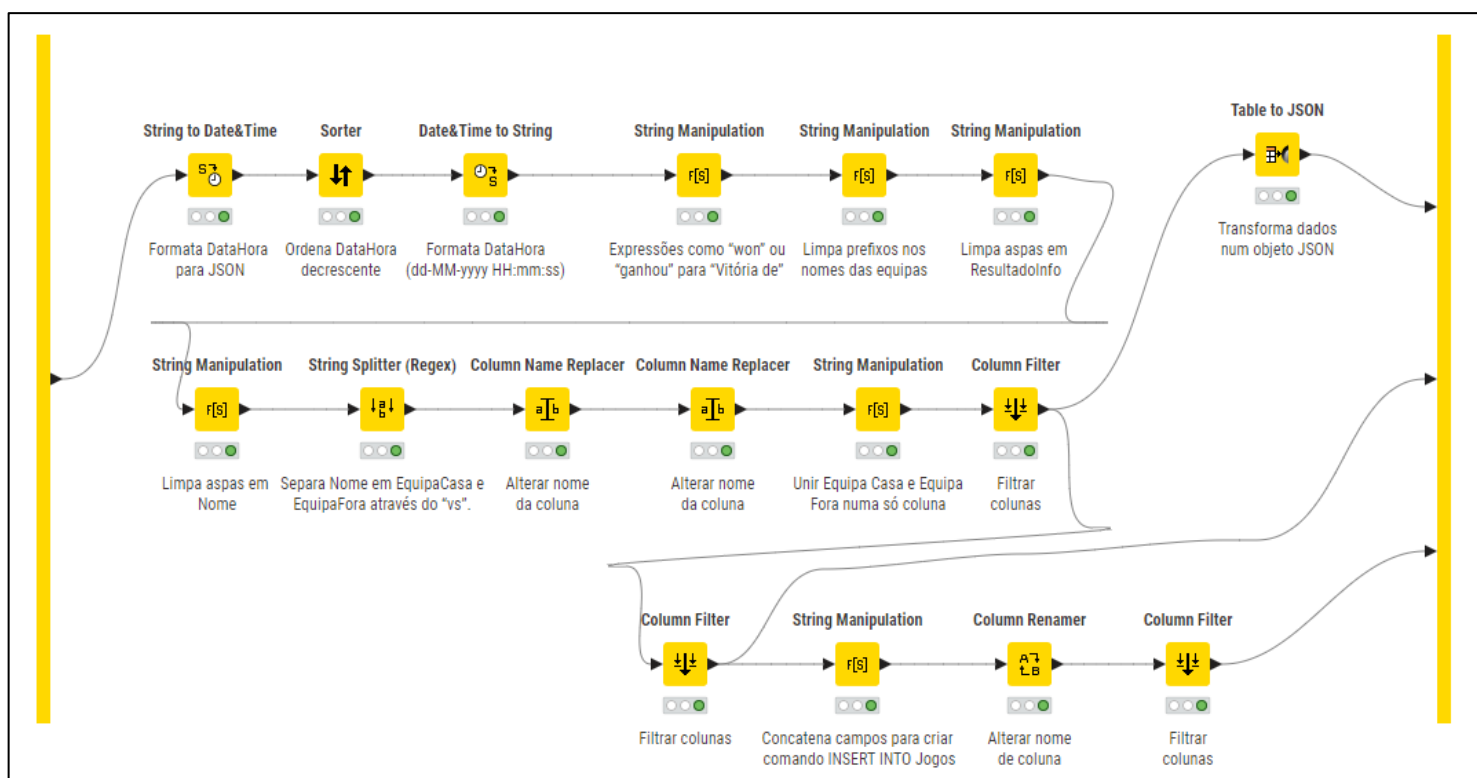


Figura 3 - Transform

Nesta fase, os dados extraídos foram convertidos, normalizados e enriquecidos. Foram aplicados vários processos utilizando nós de manipulação de texto, datas e expressões regulares:

- String to Date&Time e Date&Time to String – Padronizam os formatos de data/hora, convertendo valores ISO (ex.: 2025-10-04T14:00:00) em formatos legíveis (04/10/2025 14:00).
- String Manipulation – Normaliza textos e corrige inconsistências em nomes de equipas e campos.
- Regex Replace / String Splitter (Regex) – Utiliza expressões regulares para separar campos compostos em colunas independentes (EquipaCasa e EquipaFora) e remover caracteres especiais.
- Column Filter – Elimina colunas desnecessárias, mantendo apenas as variáveis relevantes: DataHora, ResultadoInfo, EquipaCasa, EquipaFora, Jogo.

Load – Exportação e Carregamento dos Dados

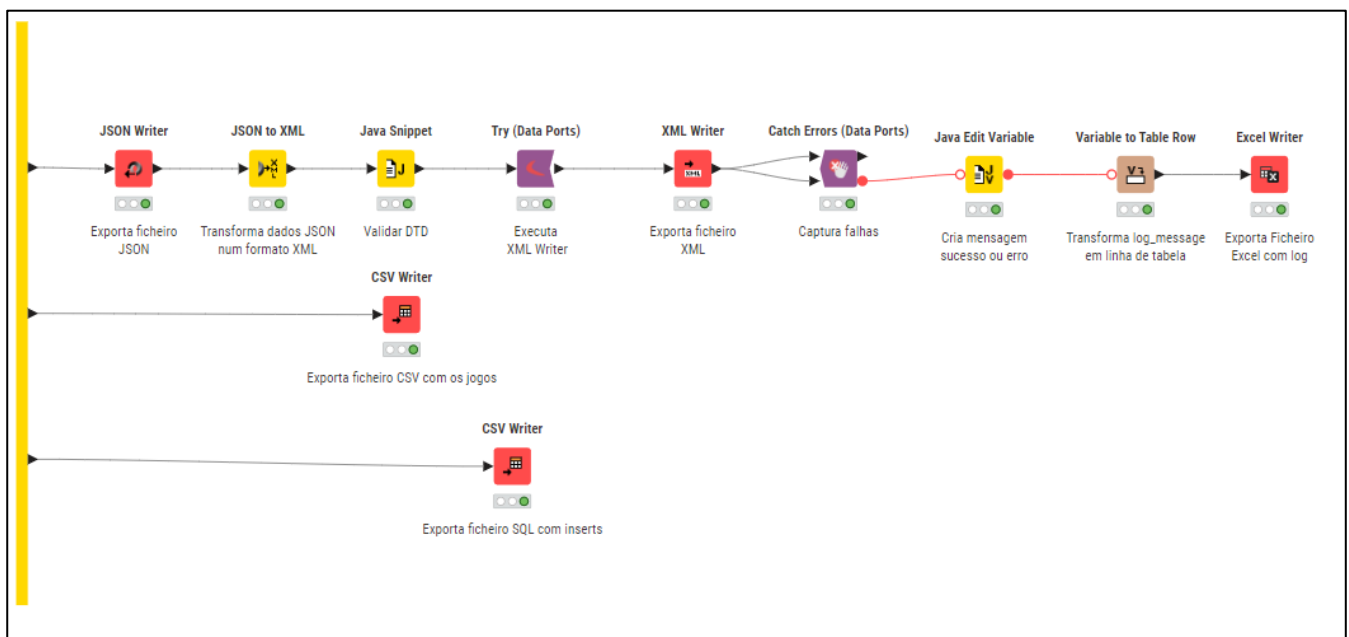


Figura 4 - Load

Após o tratamento e a verificação, os dados são exportados automaticamente em múltiplos formatos, e assim permite a sua integração com outras aplicações:

- Table to JSON / JSON Writer – Cria o ficheiro JSON_0.json, que irá servir como base para o dashboard em Node-RED.
- XML Writer – Exporta o ficheiro XML_0.xml, compatível com validação DTD.
- CSV Writer – Cria o ficheiro CSV_Jogos.csv, adequado para importação em Excel, Power BI ou bases de dados.
- CSV Writer (SQL) – Cria o ficheiro DB_export.sql, que permite o carregamento dos resultados dos jogos para uma base de dados relacional.

Para garantir a robustez e fiabilidade do processo, foi implementado um mecanismo de tolerância a falhas com os nós Try e Catch Errors.

Estes identificam automaticamente erros de execução (como problemas de escrita de ficheiros) e registam-nos num log, sem interromper o fluxo principal.

Após a exportação, o Node-RED utiliza o ficheiro JSON_0.json criado pelo KNIME para construir o dashboard dinâmico.

O vídeo de demonstração (Video_ETL.mp4) ilustra todo o processo, desde o Extract até ao Load.

3.2.2. Processo Node-RED

A componente de visualização e interação do projeto foi desenvolvida na plataforma Node-RED, com o objetivo de criar um dashboard dinâmico que fosse capaz de consultar, filtrar e analisar os resultados de jogos obtidos através do fluxo ETL criado no KNIME.

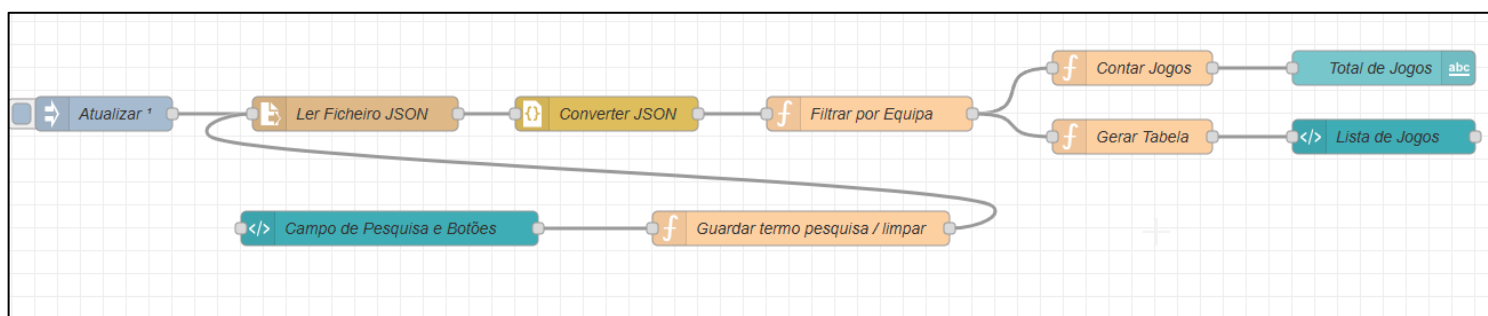


Figura 5 - Fluxo Node-RED

O fluxo implementado no Node-RED lê o ficheiro JSON exportado pelo KNIME, converte-o em objetos manipuláveis e apresenta-os no dashboard.

O utilizador pode introduzir o nome de uma equipa num campo de pesquisa e visualizar apenas os jogos correspondentes, bem como o número total de jogos encontrados.

O dashboard Futebol_ETL apresenta uma interface limpa e responsiva, composta por:

- Um indicador com o total de jogos.
- Um campo de pesquisa com botões de controlo.
- Uma tabela dinâmica que lista os jogos conforme o filtro aplicado.

Indicadores

Total de Jogos 25

Pesquisar Equipa:

Data e Hora	Equipa Casa	Equipa Fora	Resultado
14-05-2006 13:00:00	AGF	OB	Vitória de OB
30-04-2006 13:00:00	Midtjylland	Esbjerg	Empate
29-04-2006 15:00:00	SønderjyskE	AGF	Vitória de SønderjyskE
26-04-2006 16:00:00	Nordsjælland	København	Empate
08-04-2006 15:00:00	Brøndby	AaB	Vitória de Brøndby
02-04-2006 13:00:00	Midtjylland	Horsens	Empate
02-04-2006 13:00:00	Esbjerg	Silkeborg	Vitória de Esbjerg
29-03-2006 17:00:00	Silkeborg	AGF	Vitória de Silkeborg
26-03-2006 13:00:00	Brøndby	Silkeborg	Vitória de Brøndby
25-03-2006 16:00:00	Esbjerg	OB	Vitória de Esbjerg
19-03-2006 14:00:00	Horsens	Viborg	Vitória de Horsens
12-03-2006 14:00:00	OB	Nordsjælland	Vitória de OB
27-11-2005 16:30:00	AaB	København	Vitória de FC Copenhagen

Figura 6 - Dashboard final do Node-RED (Futebol_ETL)

Inicialização do Processo

A execução do fluxo tem início no nó Atualizar, do tipo inject, configurado para ser executado automaticamente após o arranque do sistema.

Este nó atua como ponto de entrada do processo, e garante que o dashboard é atualizado com a informação mais recente assim que o sistema é iniciado ou quando o utilizador o reativa manualmente.

Leitura e Conversão dos Dados

O nó Ler Ficheiro JSON (file in) é responsável por aceder ao ficheiro localizado no diretório de trabalho e ler o seu conteúdo em formato UTF-8.

De seguida, o nó converter JSON transforma o texto lido num objeto JSON manipulável, que permite o tratamento programático dos dados dentro do fluxo.

Esta conversão é essencial para possibilitar a aplicação de filtros e cálculos sobre os jogos obtidos.

Filtrar por Equipa

O filtro dos dados é realizado pelo nó Filtrar por Equipa, do tipo function.

Este nó recupera o termo de pesquisa armazenado na memória de fluxo (flow context) e aplica um filtro aos jogos, assim mantém apenas aqueles em que o nome da equipa surge como equipa da casa ou equipa visitante.

A lógica implementada permite uma pesquisa flexível e responsiva, adaptada à entrada do utilizador e executada em tempo real.

Contagem de Jogos

O nó Contar Jogos calcula o número total de jogos resultantes da filtragem, e atribuiu esse valor ao campo da mensagem.

O resultado é apresentado no dashboard através do nó Total de Jogos (ui_text), que exhibe o número de jogos encontrados de forma clara e destacada na secção de indicadores.

Esta funcionalidade oferece uma visão resumida e imediata da quantidade de registos correspondentes à pesquisa efetuada.

Criação da Tabela HTML

O nó Criar Tabela, também do tipo function, constrói dinamicamente uma tabela HTML com os dados dos jogos filtrados.

A tabela inclui colunas para Data e Hora, Equipa Casa, Equipa Fora e Resultado, que apresenta os dados de forma organizada e legível.

Caso não sejam encontrados jogos, é gerada uma mensagem informativa para o utilizador.

O conteúdo HTML é então enviado para o nó Lista de Jogos (ui_template), responsável por formatar o código no dashboard.

Interface de Pesquisa

A interface de pesquisa é implementada através do nó Campo de Pesquisa e Botões, também do tipo `ui_template`.

Este nó define um formulário interativo composto por um campo de texto e dois botões, pesquisar e Limpar, formatados através de CSS para garantir consistência visual.

O utilizador pode introduzir o nome da equipa pretendida para filtrar os resultados ou clicar em limpar para restaurar a lista completa de jogos.

Gestão do Termo de Pesquisa

O nó guardar termo pesquisa / limpar, do tipo `function`, interpreta a ação do utilizador e atualiza a variável de fluxo correspondente.

Se a ação for pesquisar o termo é guardado, se for limpar o valor é eliminado.

Para ambos os casos, o nó envia uma nova mensagem de atualização que força a execução do fluxo, assim garante que os dados apresentados refletem imediatamente a nova condição de pesquisa.

4. Análise e Discussão de Resultados

O presente capítulo apresenta a avaliação e interpretação dos resultados obtidos com a implementação do processo de integração de dados desenvolvido no KNIME e na plataforma Node-RED.

A análise incide sobre o desempenho técnico, a qualidade dos dados transformados, a usabilidade do dashboard e o cumprimento dos objetivos definidos no início do projeto.

4.1. Análise do Processo KNIME

O workflow desenvolvido em KNIME demonstrou uma integração eficiente e robusta com a API SportMonks Football, que conseguiu extrair e transformar dados com elevada consistência.

4.1.1. Fiabilidade da Extração

Durante os testes, as chamadas à API foram executadas de forma estável, sem erros de autenticação ou perdas de dados.

A utilização do nó GET Request, combinado com a configuração do token de autenticação, permitiu assegurar o acesso controlado e repetível às informações de jogos e equipas.

4.1.2. Qualidade das Transformações

As transformações aplicadas, que inclui limpeza com expressões regulares, normalização de formatos de data e eliminação de caracteres indesejados, resultaram em dados uniformes e prontos para integração.

A etapa de validação final confirmou que:

- Não existiam registos duplicados;
- As colunas estavam corretamente tipificadas;
- As datas e resultados mantinham coerência com as informações originais da API.

4.1.3. Exportação e Interoperabilidade

A exportação dos dados em múltiplos formatos (JSON, XML, CSV e SQL) foi bem-sucedida, que comprova a interoperabilidade do processo ETL.

Estes formatos serviram de ponto de ligação direto ao Node-RED, assim demonstra a integração entre ferramentas de diferentes naturezas (ETL e IoT/dashboarding).

4.2. Análise do Processo Node-RED

A solução implementada em Node-RED demonstrou grande eficácia na visualização e interação com os dados transformados.

O dashboard permitiu ao utilizador pesquisar equipas, filtrar resultados e visualizar estatísticas de forma imediata.

4.2.1. Funcionalidade do Dashboard

O painel Futebol_ETL exibiu com sucesso:

- A lista completa de jogos exportados pelo KNIME.
- O número total de jogos encontrados de acordo com o filtro aplicado.
- A atualização automática dos dados ao iniciar o fluxo ou ao clicar em “Atualizar”.

O campo de pesquisa revelou-se intuitivo e eficiente, que permite localizar com facilidade jogos específicos. O botão Limpar repõe a lista integral, assegura uma navegação fluida e sem necessidade de recarregar a página.

4.2.2. Design e Usabilidade

O layout do dashboard foi ajustado com CSS personalizado, de forma a manter todos os elementos (campo de texto e botões) alinhados na mesma linha e com cores consistentes.

Esta abordagem contribuiu para uma experiência de utilização limpa, clara e profissional, em conformidade com os princípios de design de interfaces simples e funcionais.

4.2.3. Robustez e Fiabilidade

Os mecanismos de atualização e filtros funcionaram de forma fiável, sem bloqueios ou erros de execução. A utilização de variáveis de fluxo (flow context) no Node-RED demonstrou ser uma abordagem eficaz para armazenar o termo de pesquisa e controlar a reatividade do sistema.

4.3. Integração KNIME–Node-RED

A ligação entre as duas plataformas representou um dos principais resultados positivos deste projeto. O ficheiro JSON_0.json, gerado automaticamente pelo KNIME, foi importado diretamente pelo Node-RED, que constitui um elo de integração entre o processo ETL e a camada de visualização.

O sistema demonstra assim o potencial do KNIME e do Node-RED como soluções complementares para integração, transformação e análise de dados.

4.4. Discussão dos Resultados

Os resultados obtidos demonstram que:

- O objetivo principal, implementar um processo ETL funcional com visualização em dashboard, foi completamente atingido.
- As ferramentas utilizadas permitiram demonstrar de forma prática os conceitos teóricos abordados na unidade curricular de Integração de Sistemas de Informação.

Contudo, observam-se alguns pontos de melhoria:

- A limitação do volume de dados processados depende da API e do formato de exportação.
 - O dashboard, embora funcional, pode ser expandido com gráficos estatísticos (por exemplo, número de vitórias por equipa, golos marcados, etc.).

Conclusão

O projeto desenvolvido no âmbito da unidade curricular Integração de Sistemas de Informação (ISI) teve como objetivo principal implementar um processo completo de ETL (Extract, Transform and Load), com a integração das plataformas KNIME e Node-RED. Através desta combinação, foi possível demonstrar a aplicabilidade prática dos conceitos teóricos de extração, transformação, normalização e visualização de dados, recorrendo a uma API real, a SportMonks Football API, que fornece informação sobre equipas, jogos e resultados de futebol.

Ao longo do desenvolvimento, foram cumpridas todas as etapas fundamentais de um processo ETL moderno. No KNIME, procedeu-se à extração dos dados via chamada à API autenticada, seguida da limpeza e transformação dos registos com o uso de expressões regulares e operações de manipulação de texto, garantindo a coerência e qualidade dos dados recolhidos. Posteriormente, os resultados foram exportados em múltiplos formatos (CSV, JSON, XML e SQL), de modo a assegurar a interoperabilidade com outras ferramentas.

No Node-RED, os ficheiros exportados foram lidos e processados para criar um dashboard dinâmico e interativo, que permite ao utilizador pesquisar equipas, filtrar jogos e visualizar o número total de partidas de forma automática.

O pipeline implementado evidenciou a importância da automação na integração de dados e a utilidade de ferramentas para construir soluções adaptáveis.

Durante o projeto foram consolidadas diversas competências técnicas e analíticas, nomeadamente:

- Utilização de expressões regulares e técnicas de limpeza e normalização de dados.
- Manipulação de dados em diferentes formatos de serialização (JSON, XML, CSV e SQL).
- Implementação de dashboards interativos.

Estas aprendizagens traduzem-se num ganho significativo de experiência na integração e interoperabilidade de sistemas, aproximando o projeto de cenários reais de aplicação empresarial.

Referências

KNIME Documentation – <https://www.knime.com/documentation>

Node-RED Documentation – <https://nodered.org/docs/>

SportMonks API – <https://www.sportmonks.com/api/>

Ferreira, L. & Ribeiro, Ó. (2025). Enunciado do Trabalho Prático I – ISI IPCA

Anexos

Repositório GitHub: https://github.com/claudsonandes/ISI_TP1_a30517_ETL