

2024/25

Trabalho Prático I

INTRODUÇÃO

Com este trabalho da Disciplina de Integração de Sistemas de Informação (ISI) pretende-se focar a aplicação e experimentação de ferramentas em processos de ETL (*Extract, Transformation and Load*), inerentes a processos de Integração de Sistemas de informação ao nível dos dados.

Pretende-se que sejam desenvolvidos processos de ETL que envolvam *scripts* próprias ou que recorram a ferramentas disponíveis como o Pentaho Kettle, Microsoft SQL Server Integration Services (MSSIS), Knime, Talend open studio, ou outras. Ferramentas complementares como *Node-RED*¹, *Home-Assistant*² ou outras do género, poderão também ser exploradas e integradas nos processos.

MOTIVAÇÃO

Uma vez que os processos de negócio não param de se reconfigurar, as empresas enfrentam desafios constantes de análise e aquisição de novas soluções informáticas. A necessidade de rentabilizarem anteriores aquisições, tanto pelo investimento financeiro envolvido como pela dependência dos processos, dos dados, etc., que delas fazem parte, as empresas procuram formas capazes de analisarem mais-valias e menor impacto com as novas aquisições. Processos como: i) auditorias a dados, processos, segurança, outros; ii) migração e reorganização de dados; iii) análise e processamento de dados (*datamining*, etc.); iv) recomendações e previsões sobre estados com processamento recorrente a *big data*; são exemplo de cenários onde processos de ETL poderão desempenhar papel preponderante.

No contexto emergente dos *smart environments* (*Cities, Health Care, Social*, etc.) a evidência deste tipo de soluções é clara e crescente. A integração de soluções ditas mais inteligentes, que reagem a eventos, em processos existentes ou mesmo legados, são desafios constantes.

Pretende-se que o aluno defina o objetivo a explorar no trabalho que poderá ser real ou imaginário. Será disponibilizado um url onde cada aluno deve registar: o tema do trabalho; a plataforma que pretende utilizar.

¹ <https://nodered.org/>

² <https://www.home-assistant.io/>

OBJETIVOS

Os objetivos do desenvolvimento deste trabalho são os seguintes:

- Consolidar conceitos associados à Integração de Sistemas de Informação usando Dados;
- Analisar e especificar cenários de aplicação de processos de ETL;
- Explorar ferramentas de suporte a processos de ETL;
- Explorar novas Tecnologias, Frameworks ou Paradigmas;
- Potenciar a experiência no desenvolvimento de software;
- Facilitar a assimilação do conteúdo da Unidade Curricular.
- Contribuir para qualidade da escrita de relatórios

MODELO DE FUNCIONAMENTO

- O trabalho poderá ser realizado quer de forma individual, quer em grupo de dois elementos. No caso de um grupo de dois elementos, cada elemento deve explorar uma plataforma diferente. A defesa do trabalho será individual.
- A plataforma de desenvolvimento pode ser a que o aluno pretender, desde que suporte os processos de ETL identificados.
- Os processos a desenvolver são do critério do aluno. Contudo, nesses processos são critérios de mais-valia:
 - Utilização de Expressões Regulares (ER) em processos de tratamento de dados: normalização, limpeza, etc;
 - Lidar com importação/exportação (serialização) de dados de/para XML e JSON;
 - Desenvolver *Jobs* ou *Process Controls* onde se consiga definir um projeto completo envolvendo acesso a serviços remotos como *ftp*, *email*, outros;
 - Joins, Agrupamentos, Lookups
 - Operações sobre valores: concatenar, separar, calcular, substituir, ordenar, “eliminar”, converter, selecionar, filtrar, validar, clonar, anonimizar, normalizar, corrigir.
 - Geração de Logs
 - Explorar o acesso a APIs (exemplo: serviços web) remotas;
 - Operações sobre Bases de Dados.
 - Processos de visualização dos resultados conseguidos utilizando *dashboards*, se necessário com a integração de outras ferramentas como NodRed, Apache Airflow® , Home Assistant, ou outras.
- O trabalho deverá ser entregue até ao dia **25 de outubro**, através de submissão na plataforma de e-learning;
- O trabalho deve fazer-se acompanhar de um **relatório** final que documente devidamente todo o trabalho desenvolvido, nomeadamente:

1. Enquadramento

Equipa, Unidade Curricular, Curso

2. Problema

Descrição do que se pretende demonstrar e/ou resolver

3. Estratégia utilizada

Que operadores e processos estarão envolvidos

4. Transformações

Diagramas

Explicação

5. Jobs

Diagramas

Explicação

6. Vídeo com demonstração (QR Code)

7. Conclusão e Trabalhos futuros

8. Referências bibliográficas utilizadas, incluindo referências web que tenham sido relevantes.

- O trabalho será apresentado presencialmente e individualmente ao docente em data a combinar com cada aluno.
- Cada aluno deverá ainda criar um repositório Git onde deverá colocar também todo o material produzido durante o trabalho (serão disponibilizadas instrução para criar o repositório).

CRITÉRIOS DE AVALIAÇÃO

São critérios de avaliação do trabalho

1. Qualidade do trabalho produzido

- Pertinência e Dificuldade do Problema a resolver
- Diversidade de Operadores de Transformação
- Diversidade de formatos de representação de dados
- Utilização de Expressões Regulares (ER), Normalização de dados, *Merging* de dados
- Orquestração de processos de transformação desenvolvidos
- Todos os restantes critérios de mais-valia indicados acima no modelo de funcionamento

2. Qualidade da Apresentação e Relatório do trabalho

Bom Trabalho

Luís Ferreira & Óscar Ribeiro