

Integração de Sistemas de informação

Rafael Faria – 17004

Licenciatura em Engenharia de Sistemas Informáticos

Escola Superior de Tecnologia

Instituto Politécnico do Cavado e do Ave

Indicie

Introdução	3
Descrição do problema	4
Desenvolvimento	5
Conclusão	7

Introdução

O presente relatório é referente a unidade curricular de Integração de Sistemas de informação, inserida no curso de Engenharia de Sistemas Informáticos. Este primeiro trabalho prático tem como objetivo focar na aplicação e experimentação de ferramentas em processos de ETL. Tendo em vista que sejam desenvolvidos processos ETL recorrendo a ferramenta Pentaho kettle.

ETL é um tipo de data integration em três etapas (extração, transformação, carregamento) usado para combinar dados de diversas fontes.

Descrição do problema

Como objetivo próprio, decidi utilizar um ficheiro .csv do website <https://registos.turismodeportugal.pt/>. A minha escolha foi um ficheiro onde se encontravam todos empreendimentos turísticos, tais como turismo rural (agroturismo, casas de campo, hotel rural), estabelecimentos hoteleiros e parque de campismo e/ou caravanismo. Para cada empreendimento há várias informações referentes a cada um, e algumas características de cada.

Desenvolvimento

Primeira transformação (Tranf1)

Inicialmente foi importando um ficheiro .csv com vários tipos de empreendimentos turísticos, para facilitar as transformações seguintes foi ordenado o ficheiro.

De seguida foram mudados nomes de algumas colunas para que um ficheiro .xml possa ser compreendido corretamente, seguindo com uma reposição de valores na coluna do “Localizacao_Endereco” para poder retirar algumas das abreviaturas existentes nem ficheiro.

Por fim, o ficheiro é filtrado e dividido pelas várias tipologias (Turismo rural, hoteleiro, campos de campismo e outros), sendo que os dados são exportados como um ficheiros .xml.

Segunda transformação (Transf2)

Inicialmente foi importado um ficheiro .csv com vários tipos de empreendimentos turísticos, tal como na transformação anterior e para facilitar as transformações seguintes foi ordenado o ficheiro.

Se seguida todas colunas foram apagadas exceto as que continham qualquer tipo contactos dos empreendimentos turísticos, sendo esta informação exportada como um ficheiro .sql.

Job

Neste processo começou por verificar se os ficheiros a utilizar na transformação. Após a verificação, foi realizada uma transformação com o objetivo de obter os ficheiros .xml obtidos na primeira transformação. Após obter o ficheiro .xml, foi realizada uma verificação com o .dtd.

De seguida é feita a segunda transformação onde os dados são exportados para o ficheiro .sql. E por fim é enviado um e-mail de confirmação.

Conclusão

Com este trabalho prático, posso concluir que foi fundamental a utilização de processos ETL para a sua realização.

No decorrer da realização do trabalho, vários problemas foram encontrados e também resolvidos, porém não foi possível a utilização de uma base de dados para o melhor funcionamento da aplicação.

Contudo, todos os objetivos foram alcançados com sucesso e pode-se dizer que se obteve mais um novo conhecimento de uma boa ferramenta para uso futuro.