



Universidade do Minho
Mestrado Integrado em Engenharia Informática
Mestrado em Engenharia Informática

Unidade Curricular de Data Warehousing
Ano Lectivo de 2018/2019

Trabalho Prático

Projeto de um Sistema de *Data Warehousing*
Outubro, 2018

1 Objetivos

Com a realização deste trabalho prático pretende-se iniciar e rotinar os alunos da disciplina de Data Warehousing no planeamento e execução de projetos de *Sistemas de Data Warehousing* (SDW), com particular ênfase na análise, planeamento, arquitetura e implementação deste tipo de sistemas, análise dimensional de dados e na modelação e caracterização de sistemas para extração, transformação e integração de dados em *Data Warehouses* (DWs).

2 O Trabalho Prático

Pretende-se que cada grupo de alunos realize um trabalho de análise, planeamento e arquitetura de um SDW para uma dada instituição que permita fornecer informação pertinente para suporte aos seus processos de tomada de decisão. A seleção e caracterização da instituição, que servirá como base de trabalho para cada um dos temas apresentados, será da exclusiva responsabilidade de cada um dos grupos de trabalho, que deverá indicar qual o tipo de instituição (fictícia, mas de inspiração real) que vai servir como modelo para a implementação do sistema requerido. Após a identificação e caracterização dos cenários aplicacionais da instituição caracterizada para cada um dos domínios de trabalho apresentados, pretende-se que, em termos gerais, cada grupo de trabalho faça a:

1. Definição, contextualização, motivação e objectivos do projeto.
2. Análise e justificação da viabilidade do projeto. Vantagens e ganhos do sistema de *data warehousing* relativamente aos sistemas operacionais implementados. Definição da equipa para a implementação do projeto. Definição e apresentação do plano e das estratégias para implementação do sistema de *data warehousing*.
3. Identificação e caracterização dos perfis dos utilizadores do sistema. Levantamento da informação relativa aos modelos de negócio implementados (requisitos para a tomada de decisão), identificação e análise das fontes de informação existentes e documentação subjacente. Validação do projeto com os utilizadores e revisão do projeto.
4. Modelação dimensional do sistema. Construção da matriz de suporte à decisão. Escolha dos *data marts* a implementar. Definição do grão, escolha de dimensões e desenvolvimento dos diagramas das tabelas de factos. Especificação detalhada das tabelas de factos projetadas. Revisão e validação dos esquemas desenvolvidos. Desenvolvimento dos esquemas lógicos para as tabelas de factos. Definição e caracterização das estruturas e processo de manipulação de dados relativas ao processo de povoamento do *data warehouse*. Validação do trabalho realizado com os utilizadores e revisão do projeto.
5. Análise das fontes de dados. Identificação e caracterização das fontes de dados candidatas. Análise e classificação dos dados contidos nas fontes. Estudo da qualidade dos dados contidos nas fontes.
6. Implementação do sistema de povoamento (ETL). Definição dos planos de extração, transformação e integração de dados – *source-to-target data map*. Modelação lógica do sistema de ETL. Implementação e teste do sistema de ETL.
7. Análise crítica do trabalho realizado. Definição de eventuais ações corretivas ou de melhoria sobre o projeto desenvolvido.

As áreas funcionais escolhidas, bem como os processos de tratamento de informação que com elas possam estar eventualmente associados, constituem importantes referenciais para o processo de análise que cada grupo de trabalho deverá realizar. Os alunos deverão investigar, inspirados tanto quanto possível em modelos reais, o que é que cada uma dessas áreas poderá envolver em termos de recursos (físicos e humanos), quais os serviços que disponibilizam, quais os elementos de informação necessários para a sua implementação e de que forma é que se poderão integrar no modelo global do sistema de informação da instituição em causa.

Na realização do trabalho prático, cada grupo de projeto deverá seguir a metodologia apresentada e discutida em *Kimball, R., Reeves, L., Ross, M., Thornthwait, W., The Data Warehouse Lifecycle Toolkit – Practical Techniques for Building Data Warehouse and Business Intelligence Systems, John Wiley & Sons, 2ª Edição, 2008.* - consultar bibliografia da disciplina - na realização do seu trabalho, em particular nas fases de levantamento e análise das fontes de dados e modelação dimensional.

3 Relatório do Trabalho

No final do trabalho, cada grupo de trabalho deverá elaborar um relatório técnico, sucinto e claro, de acordo com o modelo fornecido pelo docente responsável da disciplina, que apresente de forma clara e detalhada o trabalho efectuado, apresentando de forma completa o SDW que idealizou, assim como as principais estratégias de desenvolvimento adoptadas na sua conceptualização e implementação.

4 Entrega e Apresentação do Trabalho

O relatório final do trabalho prático deverá ser enviado por correio electrónico para o docente responsável da disciplina através do endereço obelo@di.uminho.pt, até às 24 horas do dia definido para a entrega do trabalho. No corpo da mensagem de correio electrónico enviada deverá constar o número e a constituição do grupo de trabalho, a designação do trabalho realizado e um endereço de correio electrónico para confirmação da recepção do trabalho. No assunto dessa mensagem deverá constar, obrigatoriamente, a referência “UM-MIEI-DW-2018-19 – Entrega do Trabalho Prático (Grupo XX)”. Após essa data, será feita a confirmação da recepção dos trabalhos por correio electrónico, assim como a comunicação do dia e hora da apresentação do trabalho.

5 Bibliografia

Como suporte à realização deste trabalho os grupos deverão apoiar-se na bibliografia recomendada no início do ano pelo docente responsável pela disciplina.

* * * * *