Data Warehouses

Departamento de Engenharia Informática (DEI/ISEP)
Paulo Oliveira
(PJO@isep.ipp.pt)

Summary

- Teacher presentation
- Course Purposes
- Course Programme
- Most Important Studying Material
- Class Types
- Required Software
- Evaluation Procedures

Teacher Presentation

- Activities
  - Associate Professor in DEI-ISEP
  - Researcher in GECAD
  - Subdirector of MEI

- Academic Qualifications
  - Degree in Management Informatics
    - Universidade Portucalense, 1994
  - MSc in Information Management
    - Faculdade de Engenharia da Universidade do Porto, 2002
  - PhD in Informatics
    - Universidade do Minho, 2009

Course Purposes

- The course aims to give the students the skills to plan, implement, manage and explore a project of a data warehouse and analytical processing system within an organization
- This course contributes with additional knowledge and competences of:
  - Exploring and maintaining data in environments characterized by high volumes, heterogeneous and from several sources.
  - Modelling and efficient representation of the data to be used in queries dealing with high data volumes.
  - Integration of data coming from heterogeneous sources.

Course Programme

1. Introduction to data warehouses
   - Concept
   - Main characteristics
   - Concept of Data Mart
   - General view of dimensional model
   - Kinds of schema

2. Dimensional Data Modelling
   - Star schema model
   - Dimension tables
   - Fact tables
   - Kind of facts
   - Slowly Changing Dimensions (SCD)
   - Modelling techniques

Course Programme

3. Data Warehouse Architectures
   - Corporate information factory
   - BUS architecture
   - Comparison between architectures

4. Extraction, Transformation, Cleaning and Loading Process
   - Data extraction
   - Transformation, cleaning and integration
   - Data loading

5. Data warehouse optimization
   - Creating indexes
   - Creating partitions
   - Creating Aggregates

Course Programme

6. On-Line Analytical Processing
   - Advantages
   - Kinds of analytical databases
   - Basic operations
   - Pivot tables

7. Advanced/Research topics in data warehouses
   - Data webhouses
   - Real-time data warehouses
   - Cloud data warehouses
   - Agile development of data warehouses

Most Important Studying Material

- Moodle: https://moodle.isep.ipp.pt/
- **Relentlessly Practical Tools for Data Warehousing and Business Intelligence**

  Ralph Kimball, Margy Ross

  2nd Edition

  Wiley, 2015

- **Database Systems: Introduction to Databases and Data Warehouses**

  Nenad Jukic, Susan Vrbsky, Svetlozar Nestorov

  1st Edition

  Prospect Press, 2016

Class Types

- Lecture Classes (Paulo Oliveira)
  - Introduce the concepts and techniques that are part of the course programme
- Practical Classes (Paulo Oliveira)
  - Theoretical contents will be discussed, complemented, and applied
- Lab Classes (Fábio Santos and Paulo Oliveira)
  - Resolving tutorials

Required Software

- **Microsoft SQL Server 2022 (Developer Edition)**
  - Minimal Components Required:
    - Database Engine Services
    - Integration Services
    - Analysis Services
- **Microsoft SQL Server Management Studio**
- **Microsoft Visual Studio 2022**
  - Components required:
    - Data storage and processing toolset with the *SQL Server Data Tools* option
    - Extensions:
      - *SQL Server Integration Services Projects*
      - *Microsoft Analysis Services Projects*
  - Visual Studio 2019 can also be used
- **Adventure Works DW 2022 Sample Data Base**

Evaluation Procedures

- During the semester:
  - The purpose of the project is the analysis, modelling, development and implementation of a data warehouse, which is then used to perform data analysis.
  - Group of 2 students
  - Minimum grade: 8.0
- Exam:
  - Minimum grade: 8.0
- Formula
  - Final Grade = Exam Grade * 35% + Project Grade * 65%
  - Always the same in any exams epoch (normal, appeal and special)

Project

- Carried out outside classes in group of 2 students
- Mandatory
- Launch: 6th week
- Divided into 2 parts (P1 e P2)
- Delivery P1-1st: 26/11/2023 (10th week)
  - Feedback and a grade is provided at the begining of P2
- Final delivery (P1-2nd + P2): 07/01/2023 (15th week)
- Presentation/discussion in the last week with both students present
- Formula
  - Project Grade = 37,5% * P1-1st + 12,5% * P1-2nd + 50% * P2
- Weight of 65% in the final grade
- Minimum grade: 8.0

Exam

- Written
- Duration: 1h30
- Without consulting
- Weight of 35% in the final grade
- Minimum grade: 8.0

Grade Improvement

- Partial Improvement
  - **New Exam**
  - Formula
    - Final Grade = **Exam grade** * 35% + Project grade * 65%
- Global Improvement
  - Written Exam
  - Formula
    - Final Grade = **Written Exam Grade**

Data Warehouses

Departamento de Engenharia Informática (DEI/ISEP)
Paulo Oliveira
(PJO@isep.ipp.pt)