Análise e Visualização de Dados

Apresentação

Cláudio Miguel Sapateiro

Masters Software Eng.

Escola Superior de Tecnologia de Setúbal – IPS

claudio.sapateiro@estsetubal.ips.pt

Summary

- Main objectives
- Syllabus
- Evaluation modes
- Bibliography and Software

Main Objectives

- Recognize the several steps on the development of *BI/Analytics projects*
- Acknowledge challenges and specific requirements on big data volumes storage
- Data warehousing data modelling
- Acknowledge the good practices on each function of the ETL process
- Understand OLAP cubes
- Build data analytics visualizations

Syllabus

1. BI e *Datawarehousing* Projects development
    a. Operational vs Decision Support Systems (DSS)
    b. Brief review on the history of DSS
    c. BI projects workflows
    d. Data modelling on datawarehousing

Syllabus

2. Dimensional Modelling
    a. Defining dimensions
    b. Data cubes and browsing operations
    c. Dimension tables
    d. Facts tables and business processes measures
    e. Star schema
    f. Data granularity
    g. Degenerated dimensions
    h. Surrogate keys
    i. Hierarchies
    j. Slowly changing dimensions

Syllabus

4. *ETL Process*
   - Extraction, Transform and Loading process(es)

4. *Data Visualization*
   - *Formats*
   - *Interaction*
   - *Tools: powerBI*

Evaluation

- Final grade calculation:
  - 30% written test
  - 30% integrated practical tasks (individual)
  - 40% presenting seminar on related subject
    - Groups of 2
    - Produce a poster
    - Any related subject since validated with teacher

  (minimum in each: 9)

- Alternative Exams dates:
  - 50% Practice Project + 50% written exam

Bibliography

Data Warehousing:

- Ralph Kimball e Margy Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, Wiley, 2002
- William H. Inmon, Building the Data Warehouse, Wiley, 2005
- Ralph Kimball e Margy Ross, The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence, Wiley, 2010
- Joy Mundy e Warren Thornthwaite, The Microsot Data Warehouse Toolkit, Wiley, 2ª edição, 2011

Data Mining:

- Jiawei Han, Micheline Kamber, e Jian Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2005
- Michael Berry e Gordon Linoff, Data Mining Techniques for Marketing, Sales, and Customer Support, Wiley, 3ªedição,2011
- Trevor Hastie, Robert Tibshirani, e Jerome Friedman, The Elements of Statistical Learning: Data Mining Inference and Prediction, Springer, 2ª edição, 2003

Bibliography

Business Intelligence:

- Ramesh Sharda, Dursun Delen, Efraim Turban; Business Intelligence and Analytics: Systems for Decision Support (10th Edition-2014). Prentice Hall. ISBN-10: 0133050904. ISBN-13: 978-0133050905.
- Rick Sherman; Business Intelligence Guidebook: From Data Integration to Analytics (2014); Morgan Kaufmann. ISBN-10: 012411461X. ISBN-13: 978-0124114616.
- Hugh Watson, George Houdeshel, e Rex Rainer, Building Executive Information Systems and Other Decision Support Applications, Wiley, 1997

Databases:

- Raghu Ramakrishnan e Johannes Gehrke, Database Management Systems, McGraw--Hill, 3ª edição, 2003
- M. Tamer Ozsu and Patrick Valduriez. Principles of Distributed Database Systems, Springer, third edition, 2011. ISBN: 978-1-4419-8833-1.
- Database System Concepts, 6th Edition, Avi Silberschatz, Henry F. Korth and S. Sudarshan; McGraw Hill; ISBN 0-07-352332-1
- Database Systems: A Practical Approach to Design, Implementation and Management; Thomas M. Connolly, Carolyn E. Begg; Pearson Education Limited; ISBN10 1292061189.

Bibliography

Complementar:

- Foster Provost , Tom Fawcett. Data Science for Business: What you need to know about data mining and data-analytic thinking (2013). O'Reilly Media. ISBN-10: 1449361323. ISBN-13: 978-1449361327.
- Rick Sherman; Business Intelligence Guidebook: From Data Integration to Analytics (2014); Morgan Kaufmann. ISBN-10: 012411461X. ISBN-13: 978-0124114616.
- Nathan Marz, James Warren. Big Data: Principles and best practices of scalable realtime data systems (2015). Manning Publications. ISBN-10: 1617290343. ISBN-13: 978-1617290343
- Avinash Kaushik ; Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity (2009). Sybex . ISBN-10: 0470529393. ISBN-13: 978-0470529393.