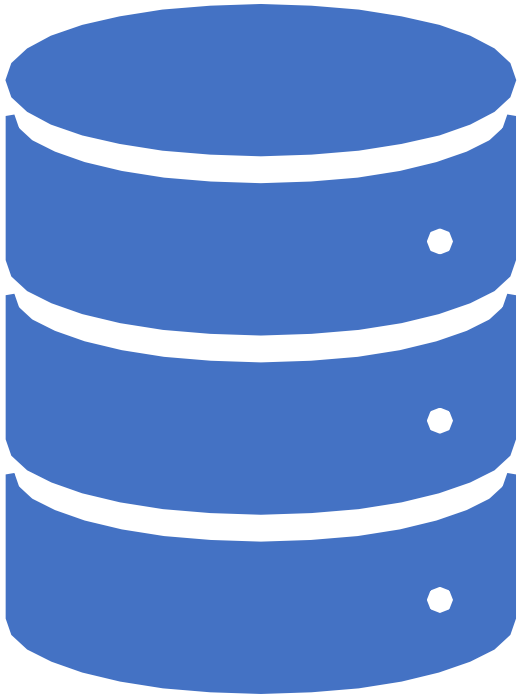

PA220: Database systems for data analytics

Course Organization

Vlastislav Dohnal

Course Overview

- Overview of data warehousing
- Planning a data warehouse
- Modelling your data for BI
- Querying your data
- Tuning and physical optimization
- ETL – getting your data into a data warehouse
- Case Study
- Novel technology (e.g., for real-time BI) – Apache Hive, Pig

Course Organization

- Lectures:
  - slides and video commentary – available for studying at anytime
- Assignments:
  - 5 home assignments with optional online consultation
    - at the time scheduled for lecture (Tuesday at 12am) – see the schedule in IS
  - at least 4 must be submitted; grading of each will be announced later
- Exam:
  - oral exam – 2-3 tasks to solve/discuss instantly
- Evaluation:
  - composite of assignment result (weight 1/3) and oral exam (weight 2/3)
  - for passing – at least 50 % of total points

Practice

- Postgresql
  - www.postgresql.org
  - may use you own installation or student's DB@FI https://www.fi.muni.cz/tech/unix/databases.html
- Microsoft Power BI
  - https://powerbi.microsoft.com/en-us/desktop/
  - install locally on your computer

Sources

- Textbooks:
  - Ralph Kimball et al.: The Data Warehouse Lifecycle Toolkit. Wiley Publishing, Inc., 2008.
  - William Inmon: Building the Data Warehouse. John Wiley and Sons, 1996.
  - Christian Jensen et al.: Multidimensional Databases and Data Warehousing. Synthesis Lectures on Data Management. Morgan & Claypool, 2010.
- Journal paper:
  - Mark Levene and George Loizou: Why is the Snowflake Schema a Good Data Warehouse Design? Information Systems, Elsevier, 2003.
- Courses:
  - Data Warehousing – Jens Teubner, TU Dortmund
  - Data Warehousing and Data Mining – Johann Gamper and Mouna Kacimi, Univ. Bolzano
  - Data Warehousing and Data Mining Techniques – Wolf-Tilo Balke, TU Braunschweig