



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

# Course Organization

- Lectures:
  - slides and video commentary – available for studying at anytime
- Assignments:
  - 4 home assignments with optional online consultation
    - consultations scheduled during lectures – see the interactive syllabus in IS
    - grading also defined there
- Exam:
  - written exam – 3-4 tasks to solve
- Evaluation:
  - composite of assignment result (weight 1/3) and exam (weight 2/3)
  - for passing – at least 50 % of total points

# Practice

- PostgreSQL
  - [www.postgresql.org](http://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 Desktop
  - <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