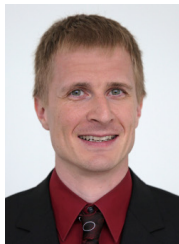


Data Warehousing

Jens Teubner, TU Dortmund
`jens.teubner@cs.tu-dortmund.de`

Winter 2015/16

A Few Words About Me



Jens Teubner

DBIS Group

Otto-Hahn-Strasse 14

jens.teubner@cs.tu-dortmund.de

- 1996–2001 Diploma in Physics, U Konstanz
- 2001–2007 Research assistant, U Konstanz, TU München
- Oct 2006 PhD in Computer Science (XML query processing)
- 2007–2008 Postdoc, IBM T. J. Watson Research Center, NY, USA
- 2008–2013 Senior Researcher, Systems Group, ETH Zurich
- since 4/2013 Full Professor, DBIS Group, TU Dortmund University

Topic: Database systems on modern computing hardware

Motivation

Frequently Bought Together



Price for all three: **\$45.76**

[Add all three to Wish List](#)

Amazon Movers & Shakers

Our biggest gainers in sales rank over the past 24 hours

Topics

Shoe

Garment

Laptop

Product category

[+ Add term](#)

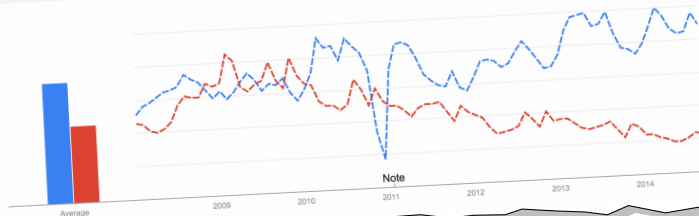
[Subscribe](#)



Beta: Measuring search interest in topics is a beta feature which quickly provides accurate measurements of overall search interest. To measure search interest for a specific query, select the "search term" option. [?](#)

Interest over time [?](#)

☐ News headlines [?](#) ☐ Forecast [?](#)



How can we...

- ...**model** business data to prepare for analyses?
- ...define an appropriate **physical database design**?
- ...**implement** analysis tasks efficiently?
- ...successfully **manage** a BI project?

We'll see how these tasks can be realized through **data warehouses**.

Course Organization

Lecture:

- Mondays, 16–18h, OH 14 / E 23

- Course website:

<http://dbis.cs.tu-dortmund.de/cms/en/teaching/ws1516/dw>

Please visit this website **regularly**. We will frequently post new information during the semester.

Exercises:

- Tuesdays 14–16h, 16–18h, OH 14 / E02 (biweekly)
- Organizer: **Cornelia Tadros** (cornelia.tadros@cs.tu-dortmund.de)
- Register via **LimeSurvey** to one of the exercise groups (see course website).
- Exercises start **next week**.

This course is meant to be taken

- as a “Wahlmodul” for CS Bachelors
- as an alternative to “Betriebliche Informationssysteme DLI (BIS)” by “Dienstleistungsinformatik” Bachelors (BIS is not offered this year), and
- by “Datenanalyse und Datenmanagement” students as a “Wahlpflichtveranstaltung zu Datenmanagement” (BD XV)

(everyone else is, of course, welcome, too).

Exercises will be **practical**.

- Based on **Microsoft Data Warehousing products**.

- Install at home (↗ MS DreamSpark)
- Or use in IRB pools
- More information on info sheet and website

Organization:

- **Register by October 21, 2015.**
- Assignment to specific group **by October 23, 2015.**
- First exercise group **on Oct 26/Nov 2.**

If you want this course to be counted as INF-BSc-AF-DLI-006:

- Exercises are **mandatory** (“Studienleistung”).
- Must do **one extra exercise within the semester**
 - Presentation (~ 10 min; present problem and solution)
 - written report (problem formulation, background, solution, tools/code; ~ more information in exercises)
 - problems posted in **Moodle**, grab one

Important:

- **Hard deadline** for written report, posted together with problem
- Please upload sources and materials to Moodle

Surviving the Exam

There will be a **written exam** (60 min) at the end of the semester.

- Date: **February 22, 2016, 15h**
- More information during the semester

Best preparation for the exam? Do the exercises!

- Do exercises **before** they are discussed in the group.

“I don’t understand this one thing. I need help!”

- Don’t hesitate to ask me or your TA.
- Speak up during the lecture!

I will post all **lecture slides** on the course web site.¹

Good **text books**:

- Ralph Kimball *et al.* *The Data Warehouse Lifecycle Toolkit*. Wiley Publishing, Inc., 2008.
- Christian Jensen *et al.* *Multidimensional Databases and Data Warehousing*. Synthesis Lectures on Data Management. Morgan & Claypool, 2010.
- Alejandro Vaisman *et al.* *Data Warehouse Systems*. Springer Verlag, 2014.

Recommended:

- “video2brain” — available **until end of 2015**.

... plus any other book on Data Warehousing that you'll find in the library.

¹Except parts that I mark with  on the slide.

Experiment with a Database!

I **strongly** recommend you exercise the material of this course on a **real database system**.

Examples:

- **Microsoft SQL Server** (<http://www.microsoft.com/en-us/server-cloud/products/sql-server/>)
 - Used in the exercises (together with other MS tools)
 - Available via DreamSpark
- **Oracle** (<http://www.oracle.com/us/products/database/>)
 - You might have worked with it in “Information Systems”
- **IBM DB2** (<http://www.db2express.com/>)
 - Full-featured, industry-strength database
 - Available **for free** (Win/Linux/Mac)
- **PostgreSQL** (<http://www.postgresql.org/>)
 - Very powerful and feature-rich **open source** database

Our topics for this semester

- Overview of data warehousing
- Planning a data warehouse
- Modelling your data for BI
- Tuning and physical optimization
- ETL–Getting your data into a data warehouse
- Web-scale analytics
- Novel technology (e.g., for real-time BI)