JACOBS
UNIVERSITY

Wrap-Up

Instructor:	Peter Baumann
email:	p.baumann@jacobs-university.de
tel:	-3178
office:	room 60, Research 1

JACOBS UNIVERSITY

Summary: Why Use a DBMS?

- DBMS: serving large data sets to large, heterogeneous user groups
- Quality of service
  - (optimized) query language = fast & flexible access to large data assets
  - Concurrent access
  - Data independence
- Efficiency
  - scalability; reduced application development time
- Information integration
  - Uniform data administration; concise information modelling
- Safety
  - Data integrity & security; Crash recovery

Value Gaps in Data Management, Services, & Analytics

- Arizona State University, Jan 2013: roundtable on challenges, approaches, and needs of the industry with respect to data services & analytics
  - http://wp.sigmod.org/ by K. Selçuk Candan
- Six most critical knowledge competency groups (aka research fields)
  - 1. temporal and spatial data analyses,
  - 2. summarization, cleaning, visualization, anomaly detection,
  - 3. real-time processing for streaming data,
  - 4. representations and fusion for unstructured/structured data, semantic Web,
  - 5. graph-based models, social networks, and
  - 6. performance and scalability, distributed architectures.

missing: arrays

Value Gaps in Data Management, Services, & Analytics (contd)

Core challenges & vistas spotted:

- Massive data ("Big Data")
- Data in motion ("data streams")
- Federated data storage, analysis, and modeling
- A new crop of data scientists
- No more teaching-one-DBMS paradigm
- Multi-disciplinary, business-aware program

„No One Size Fits All“

- General insight today: no singular data modeling paradigm (eg, sets) can match all requirements in semantics & performance
- Ex: SAP HANA: four main-memory storage engines
  - column-store, for OLAP-dominant & mixed workloads
  - row-store, for OLTP-dominant workloads
  - graph engine
  - text engine

http://martinfowler.com/bliki/PolyglotPersistence.html

- Refs:
  - M. Stonebraker et al. *”One size fits all”: an idea whose time has come and gone*. ICDE, 2005
  - F. Färber et al.: *The SAP HANA Database – An Architecture Overview*. IEEE Data Eng. Bull., 35(1):28–33, 2012
  - V. Sikka et al. *Efficient transaction processing in SAP HANA database: the end of a column store myth*. SIGMOD, 2012

Course Plot – or: why did I take it?

- How to design databases, and how to search them
- How to design (Internet) services

What industry expects a CS graduate to know

- Database services revisited
- Practice: set up a Web service
  - LAMP = Linux, Apache, MySQL, PHP

Your entry point to the DB [admin] world

Must-Haves for IT Job Interviews

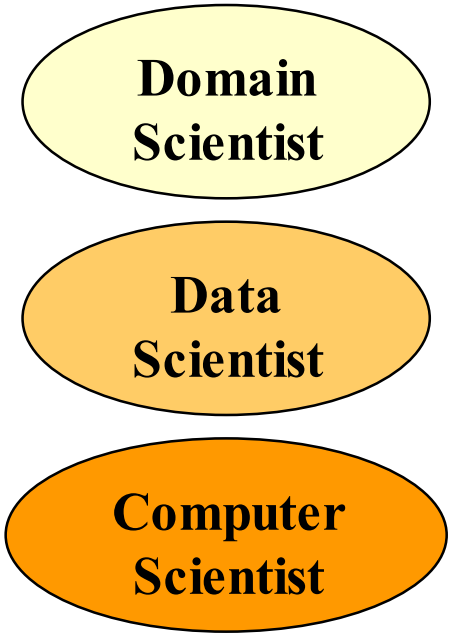
- "47% of the job ads analyzed expect economics knowledge. Also, communication skills are emphasized.
- Currently database skills are at the top of the IT companies' wish list, every 3rd IT job ad requires them. Further, Business Intelligence, Enterprise Resource Planning, and Service-Oriented Architectures are an asset. Additionally, relevant hands-on experience, e.g., in project work, plays an important role."

  -- Thomas Reher, Executive Board member, PPI AG

Data Scientists in the Market

- 2015 Data Science Salary Survey by O'Reilly:
  - https://www.oreilly.com/ideas/2015-data-science-salary-survey/page/1/executive-summary

- PS: Data Scientist ≠ Computer Scientist



But, Mind You:

- Chances are you won't use classroom knowledge as is – diversity!
  - Diversity of technology, requirements, enterprise setups, ...
- ...then why did we do it??
  - Grasp the concepts
  - Whatever gossip says – SQL is like English: y'all just need to know (at least basics)
  - Able to immerse into any DB & Web services technology rapidly