

CO-560-A Databases

Instructor: Peter Baumann

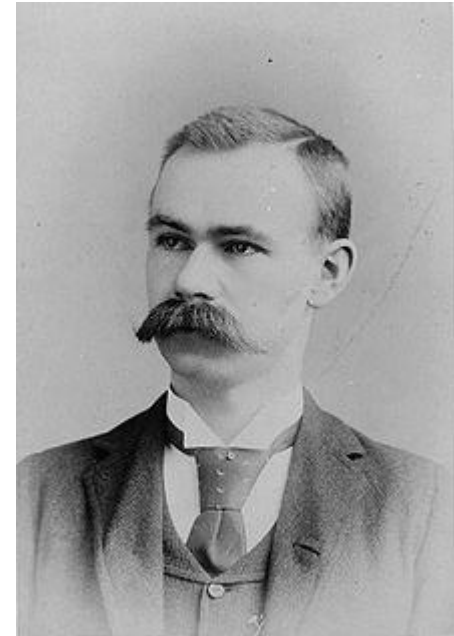
email: pbaumann@constructor.university

office: room 88, Research 1

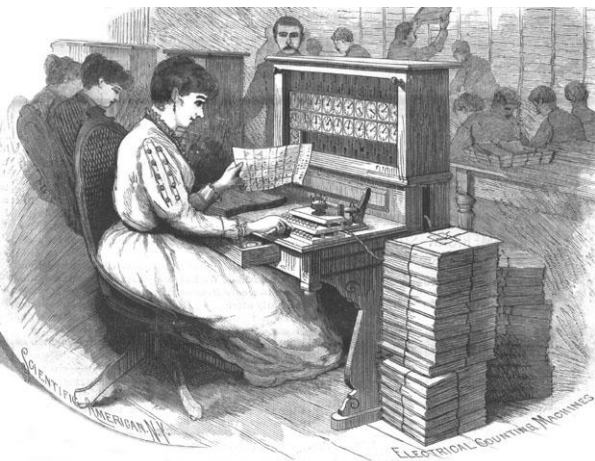
Where It All Started

Source: Wikipedia

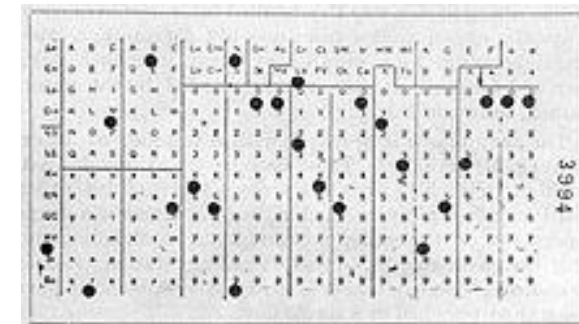
- 1890 census on 62,947,714 US population ← “Big Data”
 - was announced after only six weeks of processing
- Hollerith „tabulating machine and sorter“
- Tabulating Machine Company
 - International Business Machines Corporation



Herman Hollerith in 1888



Hollerith card puncher, used by the United States Census Bureau



Hollerith punched card

„Data Never Sleep“



[Domo 2025]

„Data Deluge“

- „It is estimated that a week’s work at the New York Times contains more information than a person in the 18th Century would encounter in their entire lifetime and the thought is that within 10 years the rate of information doubling will occur every 72 hours.“ -- P. „Bud“ Peterson, U Colorado
- “monthly global traffic 147 EB per month” – The Global Statistics, 2025
- typical car: ~100 million LoC
[\[source\]](#)
- „drowning in data,
starving from insight“

[Audi/Wired]



2012



[Mobile Marketing Association of Asia]

„Big Data“

- Internet: the unprecedented information collector
 - 2012: 200m Web servers [Yahoo]
 - estd 50+b static pages [Yahoo]
 - 40 b photos [Facebook]
 - 2012: 31b searches/m [Google]
- 2025: 463 Exabytes / day
- Typical Big Data:
 - Business Intelligence
 - Social networks - Facebook, Twitter, GPS, ...
 - Life Science: patient data, imagery
 - Geo: Satellite imagery, weather data, crowdsourcing, ...

Data = the „new gold“, „new oil“
Petrol industry: „more bytes than barrels“

Where to Store Big Data?

[technostation.com]

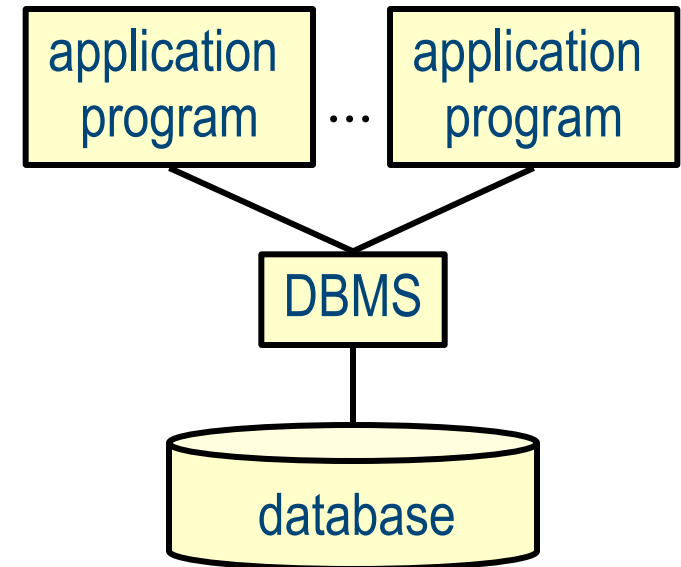


Data Management: The Task

- Manifold information,
accessed by users in manifold (often unanticipated) ways
 - Standard task
 - Many variations
- Solution: **individually configurable standard tool**
- *...is this marketing speak???*

What Is a Database [System]?

- **Database = DB** = an integrated collection of data
 - With a well-described structure = schema
- **Database [Management] System = DBMS**
= software to store and manage databases
 - ...and no one else!
- describes **excerpt** of real-world enterprise
 - "Universe of Discourse" (UoD), "mini world"
- **Example:**
 - Entities (students, courses, ...)
 - Relationships (Madonna is taking 320301, ...)



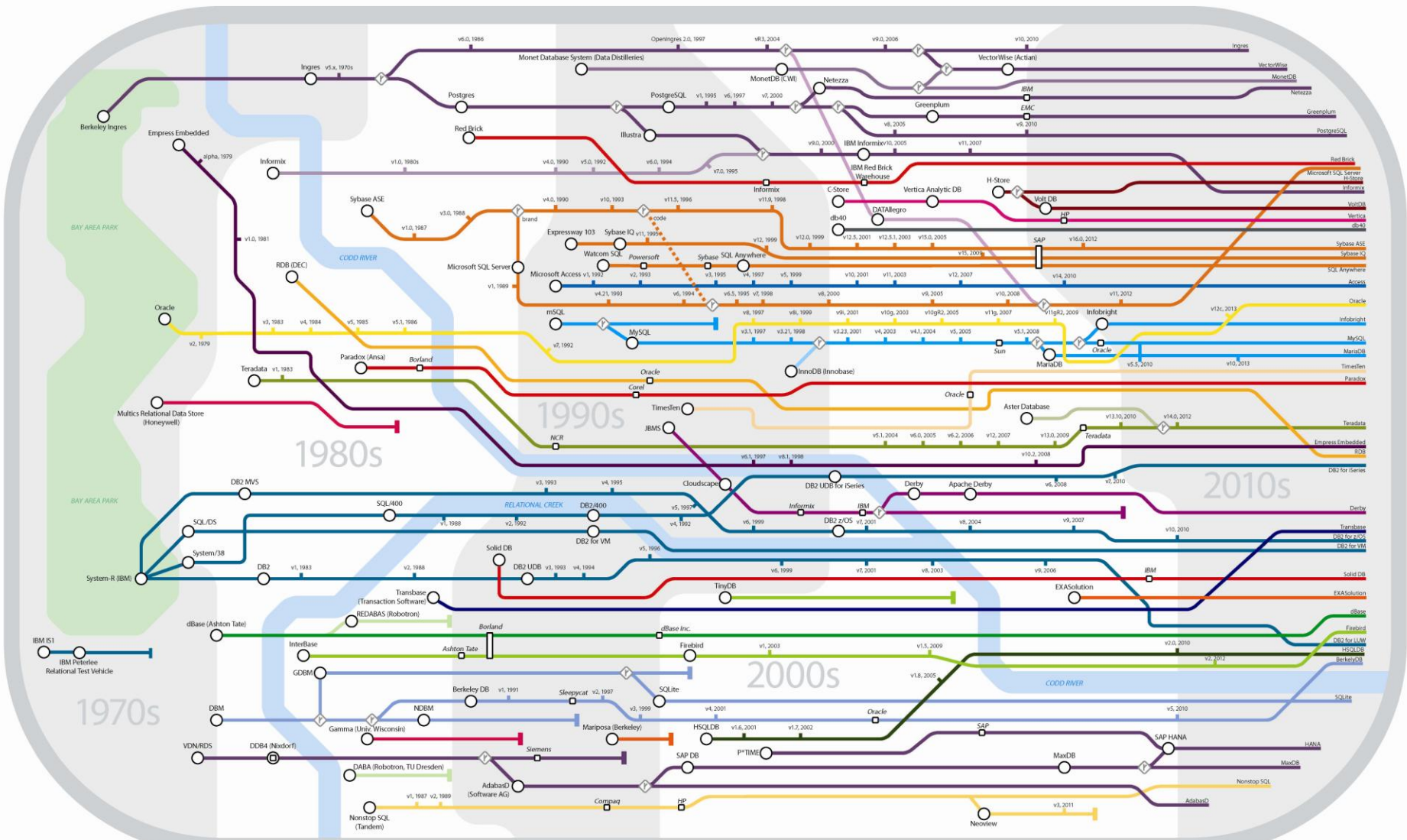
DBMS History

- 60s... IMS (hierachical model, for tapes),
CODASYL (network model, still tapes)
- 1974 SEQUEL defined (Chamberlain et al.)
- 1977 IBM prototype System R;
Oracle starts implementation
- 1979 first Oracle SQL DBMS shipped
- 1981 IBM ships SQL/DS
- 1983 IBM introduces DB2
- 1985 Ingres, Informix switch to SQL
- 1987 ISO 9075 Database Language SQL
- 1988 dBASE IV with SQL
- 1989 ISO SQL-89
- 1992 ISO SQL-92
- 1999 SQL:1999 (SQL3): extensibility
- 2003 SQL:2003

Key to success: query language

- **Intuitive** (hm...)
- Yet precise, formalised **semantics**
- **Declarative** = abstracts from internals
- ...hence **optimizable**

Genealogy of Relational Database Management Systems

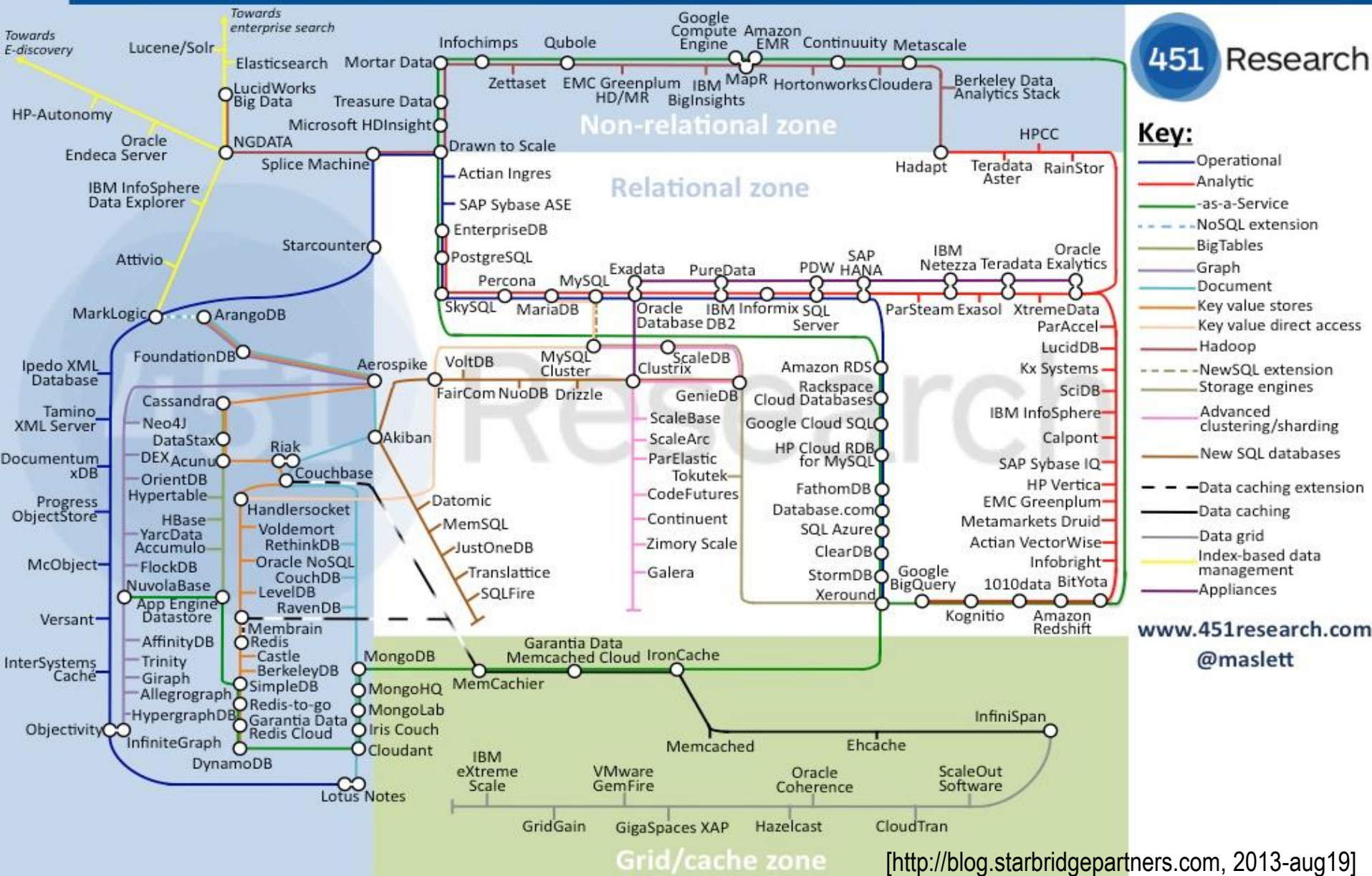


Key to lines and symbols

- Publishing Date
- ◻ Acquisition
- ↗ Versions
- ⊥ Discontinued
- ◇ Branch (intellectual and/or code)
- Crossing lines have no special semantics

Database Landscape Map – December 2012

451 Research

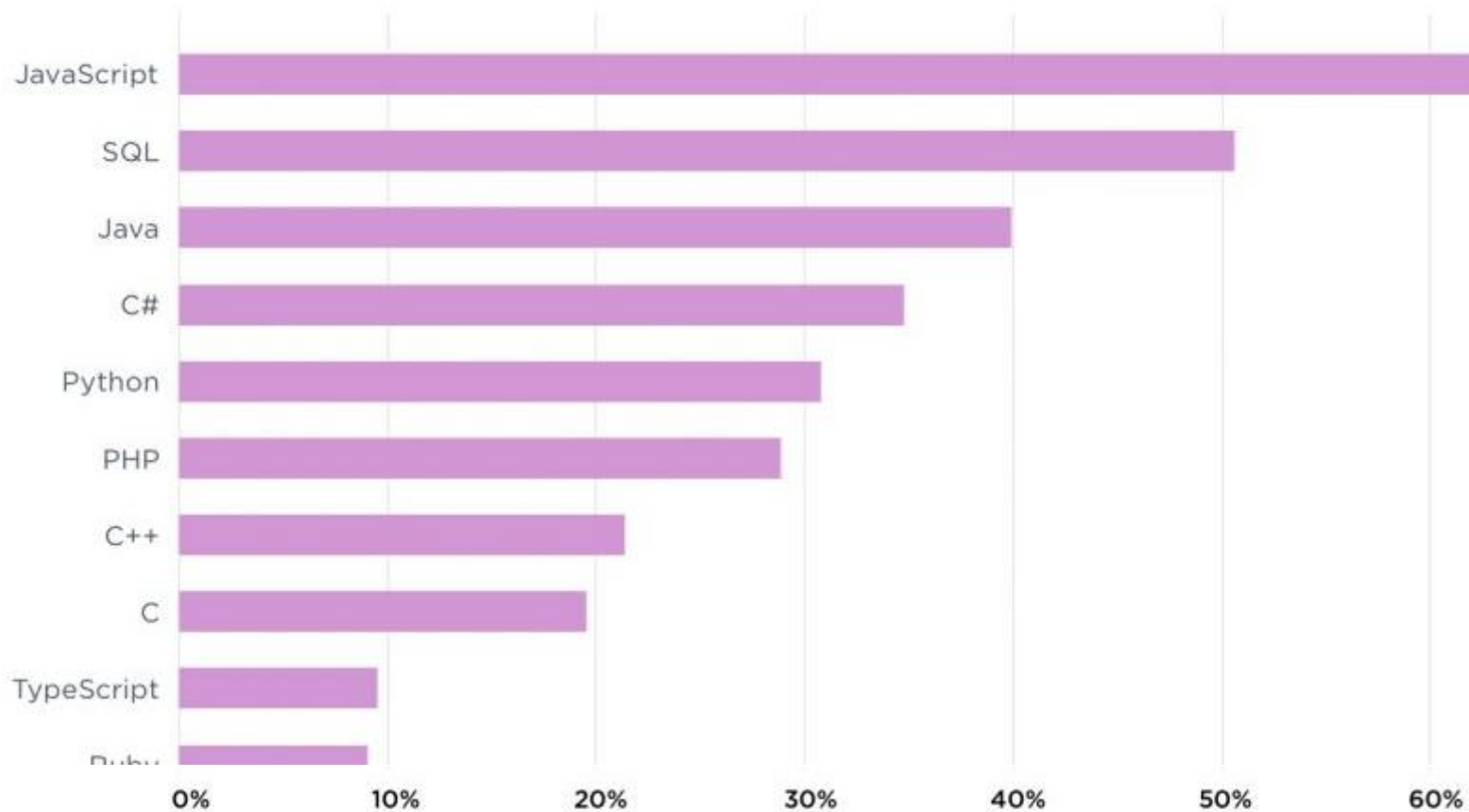


...and Then Came NoSQL

www.nosql-database.org

- original intention: modern web-scale databases
 - began early 2009, has grown rapidly
 - Broadened into “Next-Generation Databases”
- **Fast:** On >50 GB data:
 - MySQL: Writes **300 ms** avg
 - Cassandra: Writes **0.12 ms** avg
- The Empire strikes back: **NewSQL**

...but still:



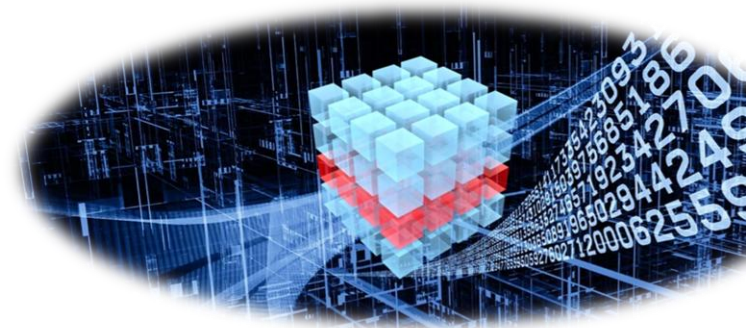
OUR RESEARCH

BIG EARTH DATA

The Digitized Planet

Our Research: Array Databases

- Large-Scale Scientific Information Services (L-SIS) Research Group
 - flexible, scalable services on massive n-D arrays
- Main visible results:
 - rasdaman Array DBMS
 - EarthServer global Earth datacube federation
 - Datacube standards in OGC, ISO, INSPIRE
...most of all: SQL/MDA
- *Got rock-solid coding skills?
C++, Java, JavaScript?
Join us!*



LOGISTICS

Grading

Requirements			
Course/Final module requirements	Requirements	Compulsory pass	Weighting
Final module requirements	Project	Yes	33%
	Written Exam	Yes	67%

Final module exams			
Exam	Date	Instructors	Compulsory pass
1 Project	No Date	Prof. Dr. Peter Baumann	Yes
1 Written Exam	Time tbd	Prof. Dr. Peter Baumann	Yes

- Overall pass: each module component at least 45%

Prerequisites

- Interest, Curiosity, Engagement
- General CS I+II, programming, basic algebra
 - data structures (trees!), object-oriented concepts
 - general programming experience
 - Linux, shell (project!)
- Non-CS majors: **contact** me!
 - more difficult due to missing prerequisites
 - This is an *advanced* CS course!
- *"reading without writing is daydreaming"*
- On any difficulties, **contact** TAs/me

Resources

- Course material: <https://peter-baumann.org> → teaching → DBWS
- Instructor: pbaumann@constructor.university
- Teaching Assistants (questions, tutorials, ...):
 - Tsonev, Nikolay Lachezarov <ntsonev@constructor.university>
 - Pico, Mateo <mpico@constructor.university>
 - Tsignadze, Nana <ntsignadze@constructor.university>
 - Tadic, Milica <mitadic@constructor.university>
 - Moryakov, Emil <emoryakov@constructor.university>
- DB forum of Teams course & project
- CLAMV: [clabsql](#) – see project session
- Textbooks: next slide

Literature

- H. Garcia-Molina, J.D. Ullman, J.D. Widom: **Database Systems: The Complete Book**. 2nd ed, Pearson, 2008
 - Fundament providing all the basics
- M. Stonebraker, J.M. Hellerstein, J. Hamilton: **Readings in Database Systems**, 5th ed, Morgan Kaufmann Publishers, 2018
 - Interesting facets and views
- Elvis C. Foster, Shripad V. Godbole: **Database Systems**. O'Reilly, 2014
 - Inspection of concrete DBMSs
- P. Trivett: **Python Programming and SQL**. ISBN-13: 979-8868124884
- M. Grinberg: **Flask Web Development: Developing Web Applications with Python**. O'Reilly, 2018
 - Two alternatives providing important how-to for the project
- the Web – find your favourite, such as [this video play list](#)

Course Plot – or: why should 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

Your entry point to the DB [dev/admin] world

Course Plot, Refined

- Database design
 - Entity-Relationship Model; UML
- Relational model
 - Relations; SQL intro; ER mapping; views
 - SQL: queries, constraints, triggers
- Database application development
- Web service architectures
 - AJAX
 - HTML, CSS, JSON, ...
- Databases revisited
 - Logical/Physical Design, Transaction Management, Security, Authorization
- NoSQL
- Big Data

CAREER RELEVANCE

Job Opportunities with DB Knowledge

- DBMS implementor (with DBMS vendor)
- DB administrator (DBA)
- Database consultants
- Software developer
 - ...without basic DB knowledge? No way!

Salaries in Germany 2024?

- Database administrator



- Database developer

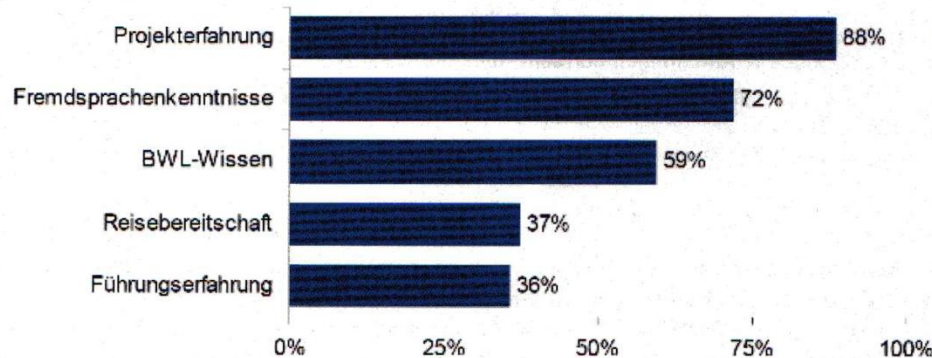


- Backend programmer



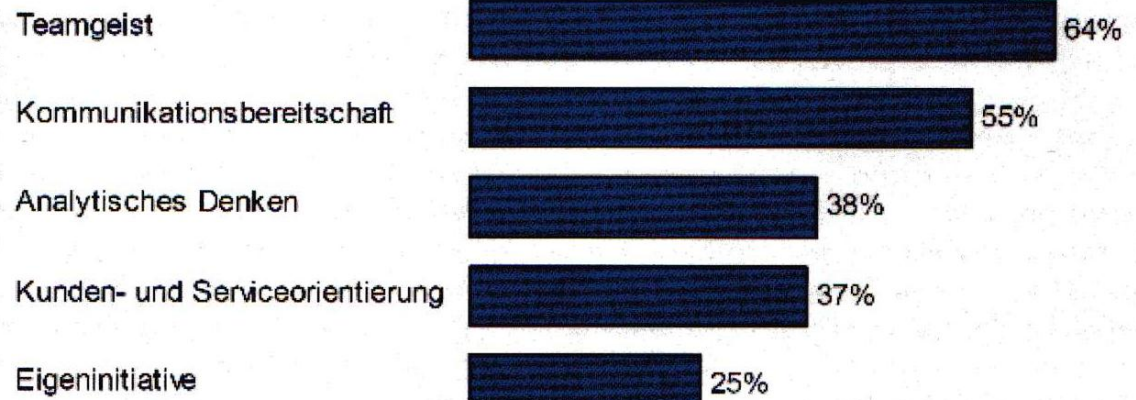
Skills Expected

IT-Karriere: Was der perfekte Bewerber mitbringen muss



Gefragte Zusatzqualifikationen: Für 72 Prozent der IT-Jobs sind Fremdsprachenkenntnisse Pflicht.

Top 5 der „weichen“ Faktoren in Jobanzeigen für ITler



Summary: Why Learn Databases?

- Fun & challenge
 - DBMS unique mix of most of CS:
OS, programming languages, complexity theory, AI, logic, statistics, hardware, ...
- Money
 - Computer experts *with database knowledge* hold responsible jobs...and are **well-paid!**

