



Data Management and Databases

Instructor: Peter Baumann

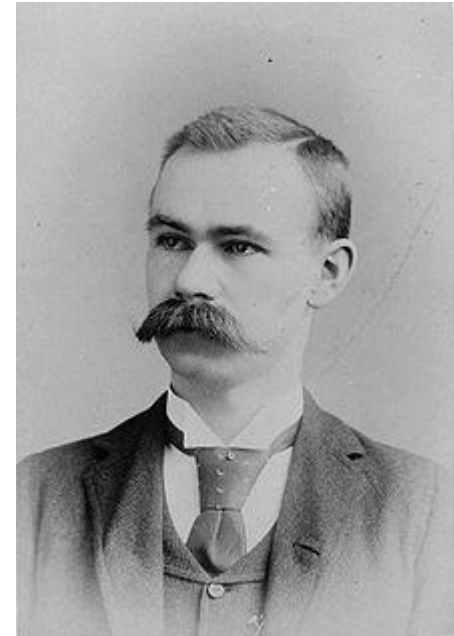
email: pbaumann@constructor.university

office: room 88, Research 1

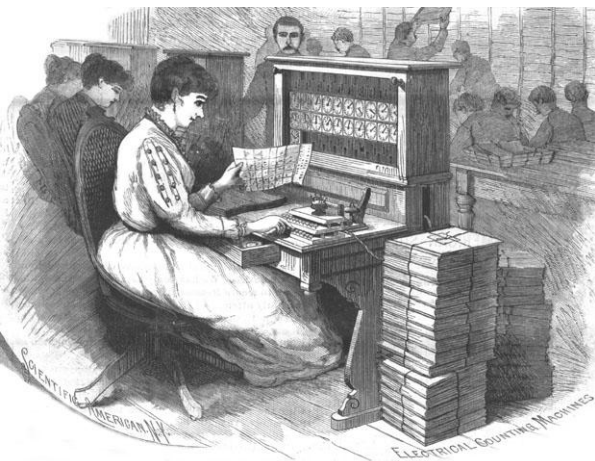
Where It All Started

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]

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]

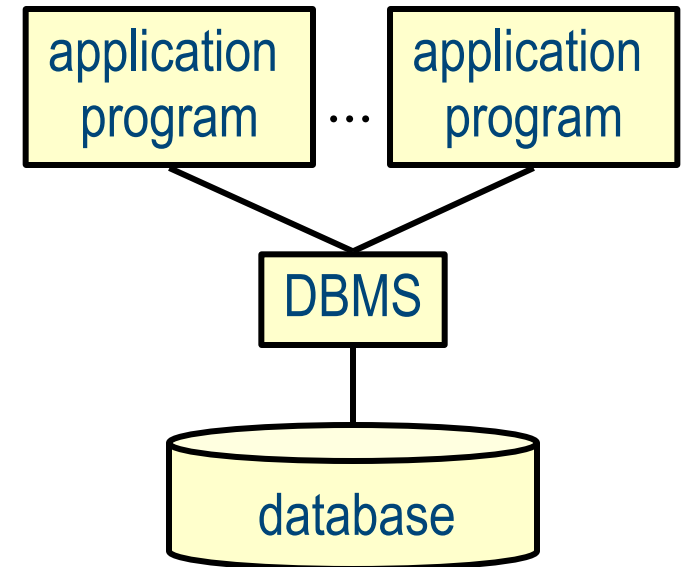


Data Management: The Task

- 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, ...)



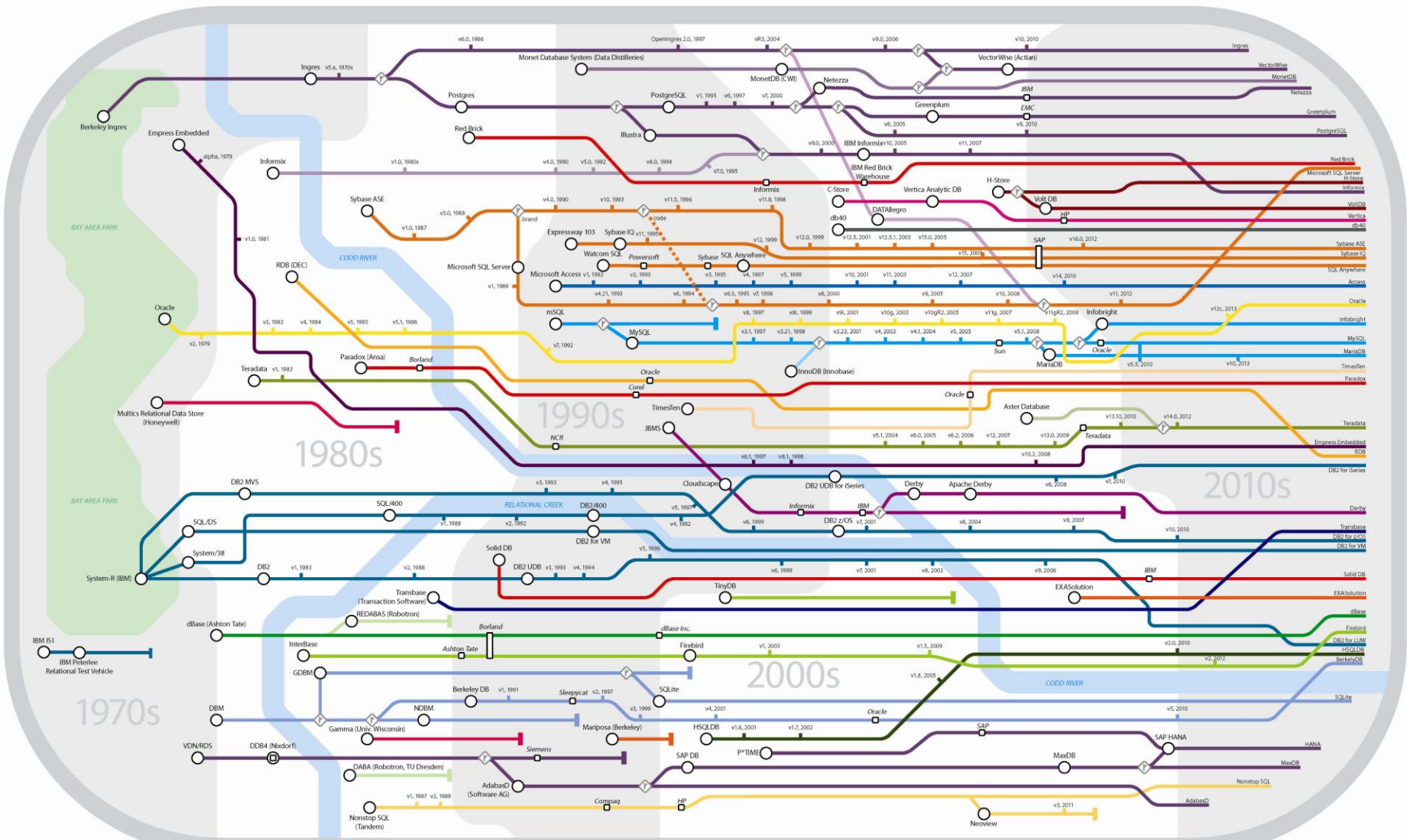
DBMS History

- 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
- v5, 2006CC Versions
- Discontinued
- Branch (intellectual and/or code)
- Crossing lines have no special semantics

Database Landscape Map – December 2012

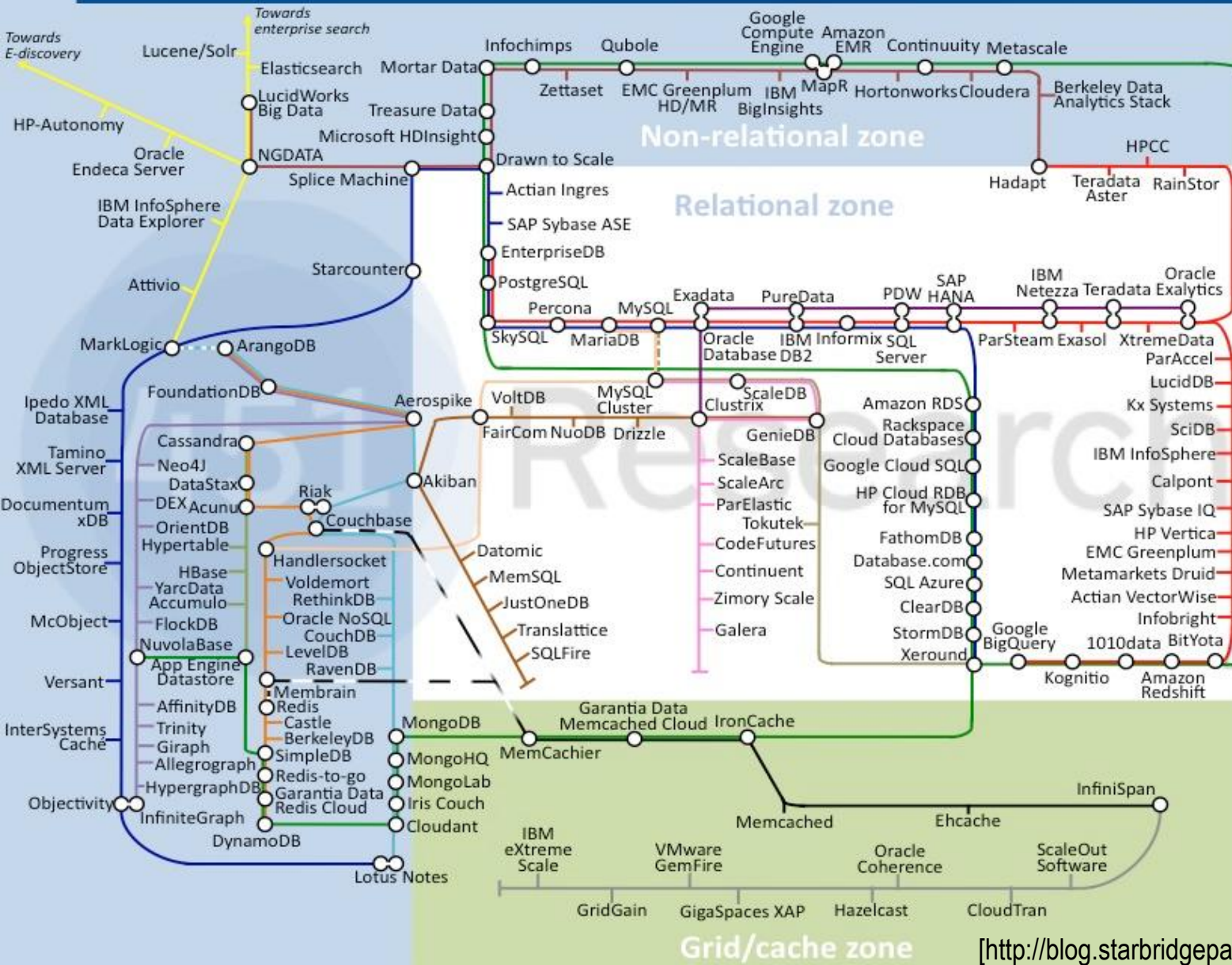
451 Research

Key:

- Operational
- Analytic
- as-a-Service
- - - NoSQL extension
- BigTables
- Graph
- Document
- Key value stores
- Key value direct access
- Hadoop
- - - NewSQL extension
- Storage engines
- Advanced clustering/sharding
- New SQL databases
- - - Data caching extension
- Data caching
- Data grid
- Index-based data management
- Appliances

www.451research.com

@maslett



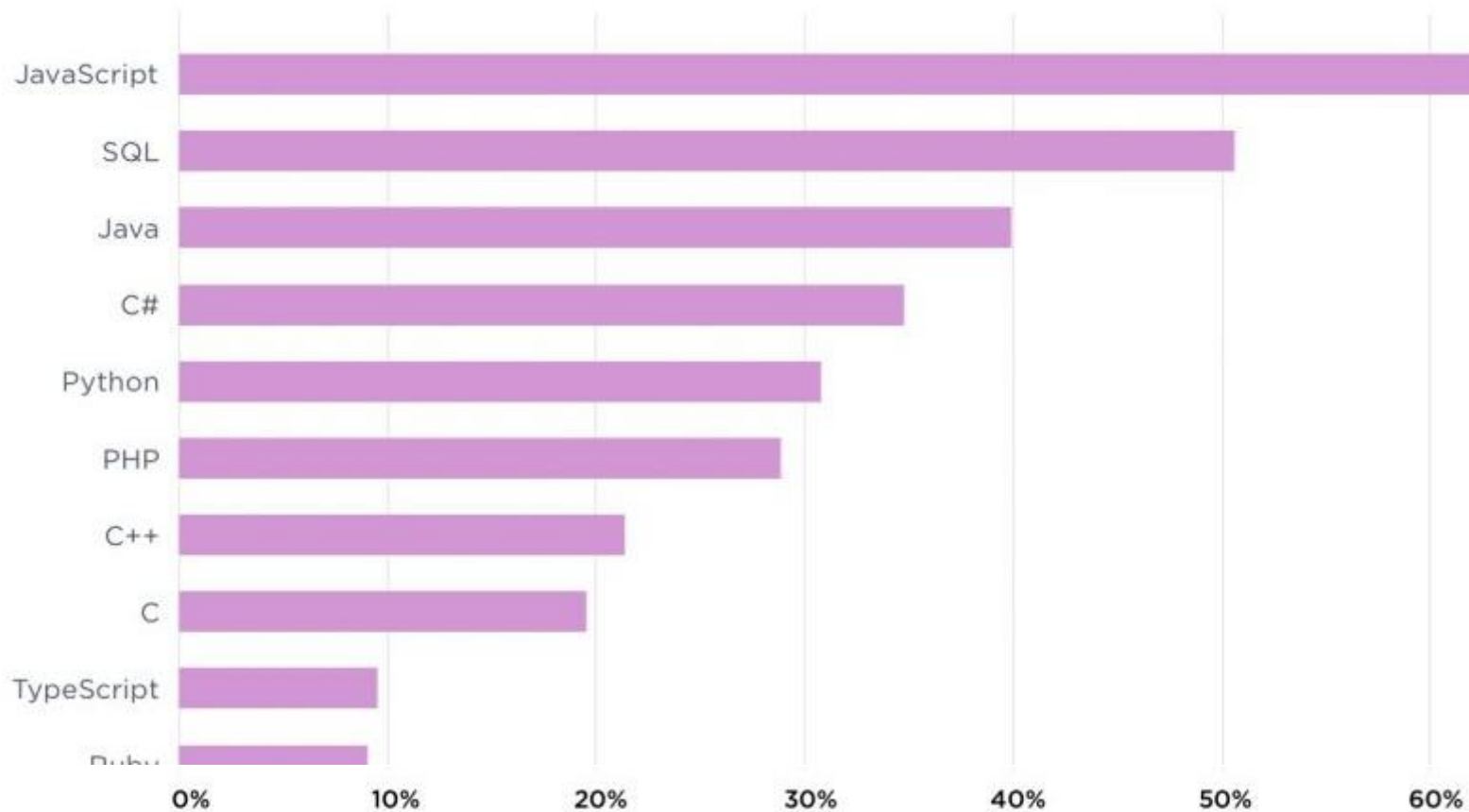
[http://blog.starbridgepartners.com, 2013-aug19]

...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:



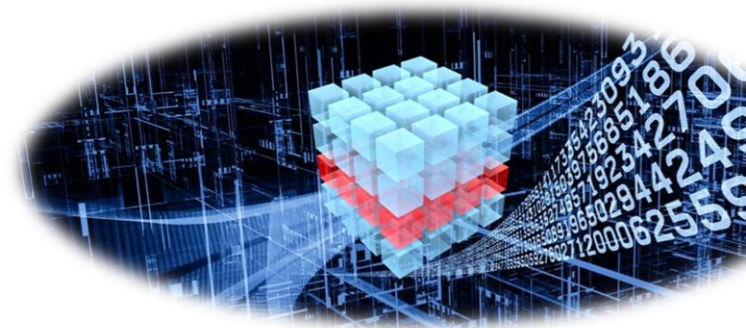
OUR RESEARCH

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



LOGISTICS

About the Course

4 th Semester	Master Thesis / Seminar m, 30 CP							
3 rd Semester	Data Warehousing: Concepts and Technologies m, 5 CP	Image Processing for Data Engineers m, 5 CP	Data Security and Privacy m, 2.5 CP	Elective me, 5 CP	Elective me, 5 CP	Ethics and the Information Revolution m, 2.5 CP	Advanced Project II* me, 5 CP	Internship* me, 10 CP
2 nd Semester	Cloud Computing m, 5 CP	Data Pipeline Engineering m, 5 CP	IT Law m, 2.5 CP	Elective me, 5 CP	Elective me, 5 CP	Acad. Writing Skills Intercultural Training m, 2.5 CP	Advanced Project I* me, 5 CP	
1 st Semester	Data Management and Databases m, 5 CP	Data Analytics m, 5 CP	Python Programming for Data Engineers m, 5 CP	Elective, Remedial me, 5 CP	Communication & Presentation Skills m, 2.5 CP	Language m, 2.5 CP	Current Topics in Data Engineering m, 5 CP	
CORE				Elective	Career	Discovery		

Components	Examination Type	Duration/ Length	Weight (%)	Minimum	ILOs
Data Management and Databases	Written Examination	120 min	50	45%	All
Data Management and Databases - Tutorial	Program Code		50	45%	All

About the Course

- introduction to data management & database systems
- principles of database design, [not only] SQL querying, data management
- Tailored to Data Engineering (not Computer Science)

No	Competence	ILO
1	Design	Design and optimize complex database systems tailored to specific requirements.
2	Apply	Apply advanced querying techniques and utilize sophisticated database features.
3	Evaluate	Evaluate and implement various database technologies, including NoSQL solutions.
4	Address	Address challenges related to performance, scalability, and data management.
5	Understand	Understand the importance of data governance, security, and compliance in database systems.

Grading

Components	Examination Type	Duration/ Length	Weight (%)	Minimum	ILOs
Data Management and Databases	Written Examination	120 min	50	45%	All
Data Management and Databases - Tutorial	Program Code		50	45%	All

- Course: Mon 08:15
 - lecture
- Tutorial: Fri 08:15
 - Practicing material
 - Periodic code review

Prerequisites

- Interest, Curiosity, Engagement
- Some programming
 - data structures (trees!), object-oriented concepts
 - general programming experience
 - Linux, shell
- On any difficulties, **contact** TAs/me

Resources

- Course material: <https://peter-baumann.org> → teaching → DB-DE
- Instructor: pbaumann@constructor.university
- DB forum of Teams course & project
- Textbooks: next slide

Literature

- Silberschatz, Korth, and Sudarshan: **Database System Concepts**, McGraw-Hill.
- Elmasri and Navathe: **Fundamentals of Database Systems**, Pearson
- ~~▪ Kristina Chodorow and Michael Dirolf: **MongoDB: The Definitive Guide**, O'Reilly Media.~~
- Ben Forta: **SQL in 10 Minutes, Sams Teach Yourself**, Sams Publishing

Also recommendable:

- Garcia-Molina, Ullman, Widom: **Database Systems: The Complete Book**. 2nd ed, Pearson, 2008
- P. Trivett: **Python Programming and SQL**. ISBN-13: 979-8868124884
- M. Grinberg: **Flask Web Development: Developing Web Applications with Python**. O'Reilly, 2018

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

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!**

