Fundamentals of Database Systems
COMPSCI 351

Instructor: Sebastian Link

The University of Auckland

Administrative Details

Managed with CLOUDCAMPUS

Course Details

Managed with CLOUDCAMPUS

Content

- The relational model of data

Course Details

Managed with CLOUDCAMPUS

Content

- The relational model of data
- Query languages: SQL, relational algebra, and relational calculus

Course Details

Managed with CLOUDCAMPUS

Content

- The relational model of data
- Query languages: SQL, relational algebra, and relational calculus
- Database design: The Entity-relationship model and database normalization

Course Details

Managed with CLOUDCAMPUS

Content

- The relational model of data
- Query languages: SQL, relational algebra, and relational calculus
- Database design: The Entity-relationship model and database normalization
- Query processing and optimization

Learning Outcomes - Skills to put on your CV

- Create conceptual diagrams to concisely model application domains
- Transform conceptual diagrams into logical database schemata
- Normalize and de-normalize logical database schemata to process frequent database queries and database updates more efficiently
- Implement logical database schemata in the industry-standard SQL to define, manipulate and query data according to best practice
- Exploit SQL to execute semantically sound database queries and updates
- Apply relational algebra to optimize the evaluation of database queries
- Declare complex database queries in relational calculus
- Adjust database designs to evolving requirements by suitable evaluation strategies and physical data structures that help achieve good performance

Learning Outcomes - Skills to put on your CV

- Create conceptual diagrams to concisely model application domains
- Transform conceptual diagrams into logical database schemata
- Normalize and de-normalize logical database schemata to process frequent database queries and database updates more efficiently
- Implement logical database schemata in the industry-standard SQL to define, manipulate and query data according to best practice
- Exploit SQL to execute semantically sound database queries and updates
- Apply relational algebra to optimize the evaluation of database queries
- Declare complex database queries in relational calculus
- Adjust database designs to evolving requirements by suitable evaluation strategies and physical data structures that help achieve good performance

Learning Outcomes - Skills to put on your CV

- Create conceptual diagrams to concisely model application domains
- Transform conceptual diagrams into logical database schemata
- Normalize and de-normalize logical database schemata to process frequent database queries and database updates more efficiently
- Implement logical database schemata in the industry-standard SQL to define, manipulate and query data according to best practice
- Exploit SQL to execute semantically sound database queries and updates
- Apply relational algebra to optimize the evaluation of database queries
- Declare complex database queries in relational calculus
- Adjust database designs to evolving requirements by suitable evaluation strategies and physical data structures that help achieve good performance

Learning Outcomes - Skills to put on your CV

- Create conceptual diagrams to concisely model application domains
- Transform conceptual diagrams into logical database schemata
- Normalize and de-normalize logical database schemata to process frequent database queries and database updates more efficiently
- Implement logical database schemata in the industry-standard SQL to define, manipulate and query data according to best practice
- Exploit SQL to execute semantically sound database queries and updates
- Apply relational algebra to optimize the evaluation of database queries
- Declare complex database queries in relational calculus
- Adjust database designs to evolving requirements by suitable evaluation strategies and physical data structures that help achieve good performance

Learning Outcomes - Skills to put on your CV

- Create conceptual diagrams to concisely model application domains
- Transform conceptual diagrams into logical database schemata
- Normalize and de-normalize logical database schemata to process frequent database queries and database updates more efficiently
- Implement logical database schemata in the industry-standard SQL to define, manipulate and query data according to best practice
- Exploit SQL to execute semantically sound database queries and updates
- Apply relational algebra to optimize the evaluation of database queries
- Declare complex database queries in relational calculus
- Adjust database designs to evolving requirements by suitable evaluation strategies and physical data structures that help achieve good performance

Learning Outcomes - Skills to put on your CV

- Create conceptual diagrams to concisely model application domains
- Transform conceptual diagrams into logical database schemata
- Normalize and de-normalize logical database schemata to process frequent database queries and database updates more efficiently
- Implement logical database schemata in the industry-standard SQL to define, manipulate and query data according to best practice
- Exploit SQL to execute semantically sound database queries and updates
- Apply relational algebra to optimize the evaluation of database queries
- Declare complex database queries in relational calculus
- Adjust database designs to evolving requirements by suitable evaluation strategies and physical data structures that help achieve good performance

Learning Outcomes - Skills to put on your CV

- Create conceptual diagrams to concisely model application domains
- Transform conceptual diagrams into logical database schemata
- Normalize and de-normalize logical database schemata to process frequent database queries and database updates more efficiently
- Implement logical database schemata in the industry-standard SQL to define, manipulate and query data according to best practice
- Exploit SQL to execute semantically sound database queries and updates
- Apply relational algebra to optimize the evaluation of database queries
- Declare complex database queries in relational calculus
- Adjust database designs to evolving requirements by suitable evaluation strategies and physical data structures that help achieve good performance

Learning Outcomes - Skills to put on your CV

- Create conceptual diagrams to concisely model application domains
- Transform conceptual diagrams into logical database schemata
- Normalize and de-normalize logical database schemata to process frequent database queries and database updates more efficiently
- Implement logical database schemata in the industry-standard SQL to define, manipulate and query data according to best practice
- Exploit SQL to execute semantically sound database queries and updates
- Apply relational algebra to optimize the evaluation of database queries
- Declare complex database queries in relational calculus
- Adjust database designs to evolving requirements by suitable evaluation strategies and physical data structures that help achieve good performance

Recommended Textbooks

- *Database Management Systems*, Third Edition, Raghu Ramakrishnan and Johannes Gehrke, McGraw-Hill, 2003.
- *Foundations of Database Systems*, Fourth Edition, Elmasri and Navathe, Addison Wesley, 2004.

Timetable for April 5 - April 22, 2021

Lectures and in-class tutorials

- Mon 14:00-14:45; 14:55-15:40; 15:50-16:35
- Tue 8:00-8:45; 8:55-9:40; 10:00-10:45
- Tue 4:00-14:45; 14:55-15:40; 15:50-16:35
- Thu 8:00-8:45; 8:55-9:40; 10:00-10:45

Timetable for April 5 - April 22, 2021

Lectures and in-class tutorials

- Mon 14:00-14:45; 14:55-15:40; 15:50-16:35
- Tue 8:00-8:45; 8:55-9:40; 10:00-10:45
- Tue 4:00-14:45; 14:55-15:40; 15:50-16:35
- Thu 8:00-8:45; 8:55-9:40; 10:00-10:45

Rooms

- Mon afternoon 27-0206
- Tue morning 27-0304
- Tue afternoon 27-0206
- Thu morning 27-0304

Learning Opportunities and Feedback

- Actively participate in lectures, tutorials, labs and guest lectures
- Study lecture notes and the textbook
- Ask lecturers and lab supervisors specific questions
- Form study groups with regular discussion meetings
- Do strategic exercises, do textbook exercises
- Think of your own examples and do research
- Study model solutions to exercises, assignments and tests
- Study previous exams
- Install DBMS (e.g. MySQL) on your own computer to experiment
- Program

Plagiarism

- **Plagiarism defined**:

  Using the work of other scholars or students when preparing coursework or writing an examination and pretending it is your own by not acknowledging where it came from.

- **Seeking advice**:

  Course coordinators, lecturers or tutors are the appropriate people with whom you should discuss how to appropriately use and acknowledge the work of others.

- **Consequences**:

  Cheating is viewed as a serious offence by The University of Auckland. Penalties are administered by the Discipline Committee of the Senate, and may include suspension or expulsion from the University.

Guidelines to Avoiding Plagiarism

- Always do individual assignments by yourself.
- Never loan your code to another person.
- Never put your code in a public place (e.g., your web site).
- Never leave your PC without locking the screen (e.g., to get food, to have a drink, or to go to the toilet). You are responsible for the security of your account.
- Never get code from a tutor (e.g., private tutors). Several tutors have been caught giving the same code to all their students.
- Always reference the source for text you copy as part of the answer to an assignment.

Delivery

Core concepts

- Self-study of slide sets *before* lecture
- Lectures explain and illustrate core concepts
- Discuss the material in your study group

Delivery

Core concepts

- Self-study of slide sets *before* lecture
- Lectures explain and illustrate core concepts
- Discuss the material in your study group

Strategic exercises

- Attempt each set of strategic exercises on your own
- Discuss your solutions in your study group
- Solutions to strategic exercises discussed in class

Delivery

Core concepts

- Self-study of slide sets *before* lecture
- Lectures explain and illustrate core concepts
- Discuss the material in your study group

Strategic exercises

- Attempt each set of strategic exercises on your own
- Discuss your solutions in your study group
- Solutions to strategic exercises discussed in class

Optional Labs at Home

- Think about each set of lab exercises individually
- Discuss lab exercises as a team
- Work out complete solutions
- Compare your solutions with the model solutions

Instructor

- Professor Sebastian Link
  - School of Computer Science
    The University of Auckland
  - s.link@auckland.ac.nz
  - http://www.science.auckland.ac.nz/people/profile/s-link
- Associate Dean International Science
  - auckland.ac.nz/en/science.html
- Director of Data Science in the Home of *R*
  - auckland.ac.nz/en/study/study-options/find-a-study-option/data-science.html
  - r-project.org

Background

Career

- PhD from Massey on *Constraints in complex-value databases* in 2005
- Worked at Victoria University, Wellington, between 2008-2011
- Joined the University of Auckland in 2012
- Chris Wallace Award for outstanding research contributions in 2013
- DSc from Auckland on *Contributions to the theory of data dependencies* in 2015

Background

Career

- PhD from Massey on *Constraints in complex-value databases* in 2005
- Worked at Victoria University, Wellington, between 2008-2011
- Joined the University of Auckland in 2012
- Chris Wallace Award for outstanding research contributions in 2013
- DSc from Auckland on *Contributions to the theory of data dependencies* in 2015

Current research activity

- Application of logic to computer science
- Semantics in data, database design
- Data quality, data cleaning, data profiling
- Uncertainty in data, in particular possibilistic and probabilistic databases

My Top-10 Research Publications

- Ziheng Wei, Sebastian Link: Discovery and Ranking of Functional Dependencies. ICDE 2019: 1526-1537
- Tania Roblot, Miika Hannula, Sebastian Link: Probabilistic Cardinality Constraints - Validation, Reasoning, and Semantic Summaries, VLDB J. 27(6): 771-795 (2018)
- Sebastian Link, Henri Prade: Relational database schema design for uncertain data. Inf. Syst. 84: 88-110 (2019)
- Pieta Brown, Sebastian Link: Probabilistic Keys. IEEE Trans. Knowl. Data Eng. 29(3): 670-682 (2017)
- Henning Köhler, Sebastian Link: Inclusion dependencies and their interaction with functional dependencies in SQL. J. Comput. Syst. Sci. 85: 104-131 (2017)
- Henning Köhler, Uwe Leck, Sebastian Link, Xiaofang Zhou: Possible and certain keys for SQL. VLDB J. 25(4): 571-596 (2016)
- Henning Köhler, Sebastian Link: SQL Schema Design: Foundations, Normal Forms, and Normalization. SIGMOD Conference 2016: 267-279
- Sven Hartmann, Sebastian Link: The implication problem of data dependencies over SQL table definitions: Axiomatic, algorithmic and logical characterizations. ACM Trans. Database Syst. 37(2): 13:1-13:40 (2012)
- Ziheng Wei, Sebastian Link: Embedded Functional Dependencies and Data-completeness Tailored Database Design. Proc. VLDB Endow. 12(11): 1458-1470 (2019)
- Sven Hartmann, Sebastian Link: Efficient reasoning about a robust XML key fragment. ACM Trans. Database Syst. 34(2): 10:1-10:33 (2009)

Motivation: Intellectual Challenges

Bridging theory and practice

Brought forward deep theoretical results that have daily impact

Motivation: Intellectual Challenges

Bridging theory and practice

Brought forward deep theoretical results that have daily impact

Connect with seemingly disparate disciplines

Such as artificial intelligence, logic, mathematics, statistics

Motivation: Intellectual Challenges

Bridging theory and practice

Brought forward deep theoretical results that have daily impact

Connect with seemingly disparate disciplines

Such as artificial intelligence, logic, mathematics, statistics

Challenges smart people

Connects research community with users such as managers, to analysts, to developers

Motivation: SQL is the No 1 skill of Data Scientists

Data science

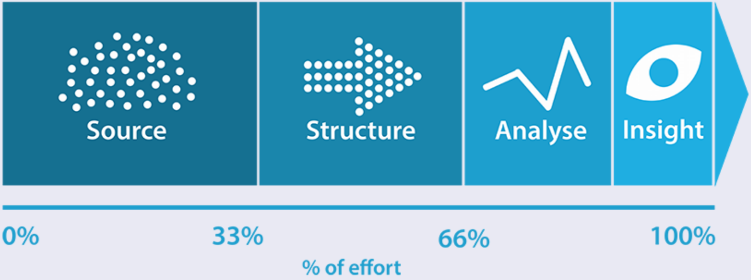
- Using Stats, Maths, CompSci to turn data into value within a domain
- Harvard Business Review: 'Sexiest job of 21st century'
- The University of Auckland is a birthplace of data science (R)

Motivation: SQL is the No 1 skill of Data Scientists

Data science

- Using Stats, Maths, CompSci to turn data into value within a domain
- Harvard Business Review: ‘Sexiest job of 21st century’
- The University of Auckland is a birthplace of data science (*R*)

Bigger picture: Databases are a cornerstone of data science



Motivation: Top 10 Software Companies Ranking 2016

Rank		Organization	Sales	FY	Market cap	Headquarters
1		Microsoft	$86.6	2017	$407	Redmond, WA, US
2		Oracle	$37.2	2017	$168.9	Redwood City, CA, US
3		SAP	$23.2	2017	$98.4	Walldorf, Germany
4		VMware	$6.7	2017	$24.7	Palo Alto, CA, US
5		Adobe Systems	$5	2017	$47.4	San Jose, CA, US
6		HCL Technologies	$5.9	2017	$17.9	Noida, UP, India
7		Fiserv	$5.3	2017	$21.9	Brookfield, WI, US
8		Salesforce.com	$6.7	2017	$51.9	San Francisco, CA, US
9		Symantec	$5.4	2017	$11.7	Mountain View, CA, US
10		Amadeus IT Holdings	$4.3	2017	$19.6	Madrid, Spain

https://en.wikipedia.org/wiki/List_of_the_largest_software_companies

Motivation: Co-founders of Google - Database Students

Larry (Lawrence) Page
Ph.D. Student
Computer Science Department
Stanford University

Member of Terry Winograd's Project on People, Computers, and Design.

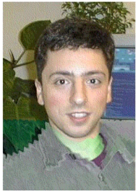
Currently working on Google, a search engine for the Web. Papers and a demo are available off this page.

email: page@cs.stanford.edu
office: Gates 360
phone: (650) 330-0100
mail: Department of Computer Science
Gates Building Room #360
Stanford University
Stanford, CA 94305-2140

`infolab.stanford.edu/~page/`

Sergey Brin's Home Page

Ph.D. student in Computer Science at Stanford - sergey@cs.stanford.edu



Research

Currently I am at Google.

In fall '98 I taught CS 349.

Data Mining

A major research interest is data mining and I run a meeting group here at Stanford. For more information take a look at the MIDAS home page or see the datamine maling list achive. Here are some recent publications:

- **Extracting Patterns and Relations from the World Wide Web**

`infolab.stanford.edu/~sergey/`

Motivation: The Business Value of Relational Databases

Data has evolved into the No 1 asset of organizations

- 90% of the world's data was produced in the prior two years
- Big data: It's so hot, so now, so wow
- 88% of Australasian businesses see data science as a key competitive advantage

Motivation: The Business Value of Relational Databases

Data has evolved into the No 1 asset of organizations

- 90% of the world's data was produced in the prior two years
- Big data: It's so hot, so now, so wow
- 88% of Australasian businesses see data science as a key competitive advantage

SQL vs NoSQL

- Oligopoly in SQL: Oracle, IBM, Microsoft, MySQL with high barriers to entry
- Lots of NoSQL players
- In NoSQL market: billion US dollars annually
- In SQL market: billion US dollars annually
- Oracle's license fee: $47,500 plus extras

Motivation: The Business Value of Relational Databases

Data has evolved into the No 1 asset of organizations

- 90% of the world's data was produced in the prior two years
- Big data: It's so hot, so now, so wow
- 88% of Australasian businesses see data science as a key competitive advantage

SQL vs NoSQL

- Oligopoly in SQL: Oracle, IBM, Microsoft, MySQL with high barriers to entry
- Lots of NoSQL players
- In NoSQL market: 3.5 billion US dollars annually
- In SQL market: billion US dollars annually
- Oracle's license fee: $47,500 plus extras

Motivation: The Business Value of Relational Databases

Data has evolved into the No 1 asset of organizations

- 90% of the world's data was produced in the prior two years
- Big data: It's so hot, so now, so wow
- 88% of Australasian businesses see data science as a key competitive advantage

SQL vs NoSQL

- Oligopoly in SQL: Oracle, IBM, Microsoft, MySQL with high barriers to entry
- Lots of NoSQL players
- In NoSQL market: 3.5 billion US dollars annually
- In SQL market: 40 billion US dollars annually
- Oracle's license fee: $47,500 plus extras

Motivation: The Business Value of Relational Databases

Data has evolved into the No 1 asset of organizations

- 90% of the world's data was produced in the prior two years
- Big data: It's so hot, so now, so wow
- 88% of Australasian businesses see data science as a key competitive advantage

SQL vs NoSQL

- Oligopoly in SQL: Oracle, IBM, Microsoft, MySQL with high barriers to entry
- Lots of NoSQL players
- In NoSQL market: 3.5 billion US dollars annually
- In SQL market: 40 billion US dollars annually
- Oracle's license fee: $47,500 plus extras

Database companies evaluate the new technology, determine its viability, and then integrate it into their product suite, giving the customer the best of all worlds.

Enjoy and interact