INF-21306

Data Management

Period 1, Academic Year 2018-2019

Course Guide

Contact person
Ioannis N. Athanasiadis
ioannis.athanasiadis@wur.nl

Language	English
Credits	6
Period	P1 2018-2019
Time and venue	**First lecture:** **September 3rd, at 12:00, Room C0217 (Forum)** See detailed schedule at https://ssc.wur.nl/Handbook/CourseSchedule/INF-21306
Exam	**Three Quizzes** and **Groupwork report** Quiz 1: 2018-09-07 at 12:00 Room C1005 (Orion) Quiz 2: 2018-09-14 at 12:00 Room C1005 (Orion) Quiz 3: 2018-09-21 at 12:00 Room C1005 (Orion) Groupwork Report and Database: 2018-10-15 at 12:00 via the Upload Link: https://www.dropbox.com/request/g6yjIyjhn9cv77MhhExn
Re-exam	Quiz Retake: 2018-09-28 at 12:00, Room C0075 (Leeuwenborch)
Always verify dates/times/rooms here:	https://ssc.wur.nl/Handbook/CourseSchedule/INF-21306
Lecturer(s)	Gert Jan Hofstede (on sabbatical leave) Ioannis N. Athanasiadis Cagatay Catal
Examiner(s)	Ioannis N. Athanasiadis Gert Jan Hofstede (on sabbatical leave)
Secretariat	Ilona van Veen/Claudia Ravestein Leeuwenborch building (6th floor) tel: 0317-84157 email: office.inf@wur.nl

Profile of the course

The digital footprint of Life, Environmental, Social, Food and Nutrition Sciences is increasing at unprecedented speed, bringing opportunities for data-driven innovation in both research and industry. To be successful, scientists need to master both domain expertise and data science skills to solve practical problems.

Aim: to become competent as a manager of data and a worker with data models and relational databases.

Students that successfully participated in the course, should be able to design and use their own relational databases and communicate effectively on topics related to data management and database modelling. Graduates are conversant where data models and databases are concerned.

This course is tailored to MSc students from any WU curriculum, or BSc student in the Data Science Minor.

Course contents

This course covers database design and the use of databases in applications, with a focus on applications in the Wageningen domains. Topics include the relational model, database design principles, the structured query language (SQL), including temporal and spatial queries. Data lifecycle topics and contemporary issues for data scientists and practitioners are also introduced, i.e. big data, FAIR principles, data governance, licensing, privacy, blockchains.

The course includes extensive practical work in the design, construction and use of databases in the students' field of study. Practical work involves MySQL and Microsoft Access.

Assumed prerequisite knowledge

Fluency in English, at ease with computers, basic mathematics.

Learning outcomes

After successful completion of this course students are expected to be able to:

1. demonstrate a managerial perspective on an organization's memory;
2. explain key concepts of data modelling and databases (i.e. entities, relationships, primary and foreign keys);
3. interpret data model diagrams using different notations (E-R diagrams);
4. compile database queries with SQL; including nested subqueries, arithmetic, logical and spatial operations;
5. analyse a realistic data problem and propose a data design solution;
6. design and implement a database for a problem in their field of study;
7. debate some of the contemporary challenges in data management, as FAIR principles, security, big data, blockchains, privacy, licensing, etc.

Course materials and resources

- Textbook: R. Watson (2017) Data Management 6th edition. e-book sold on Amazon. https://www.amazon.com/dp/B00E8HS8N2
  A Website for additional material and answers to selected exercises is being maintained by the author: http://www.richardtwatson.com/dm6e/.
- The course makes extensive use of Blackboard: the course Blackboard offers all support material: study guide, explanatory texts, self tests, sample exam questions, sample project results of prior teams, links and facilitates the group work.
- Computer programs: MS Access, MySQL Workbench.

Educational (=teaching and learning) activities

The course operates in two “modes”, each covering three weeks.

In the **first three weeks** students work mostly individually on understanding and practicing the concepts. Activities include studying the book, database design and SQL tutorials with homework assignments for the book, lectures, plenary Q/A sessions.
We work with the 'flipped classroom' model. It means that we expect students to come to the lecture to ask about material you have already studied, not to be passively fed new material. We will also explain, but assuming you have already seen the material. We will further explain this in the first lecture, and we are confident that you will like it, as the great majority of past students.
The Blackboard website is the guide to your learning process. The main topics of the course are structured in "Weeks". Each week has its own goals, and you should achieve those before moving to the next one.

The general rule is: check the Blackboard before the book; and study the book before the lecture.

Mondays and Wednesdays, we have lectures. On Blackboard there are instructions on the book chapters you *have to* study before coming to the lecture.
Tuesdays and Thursdays, we have scheduled the computer rooms, for working individually on the exercises and the tutorials, and in teams for the groupwork. On Blackboard there are instructions exercises you need to do individually to master the skills of the week. There are also instructions about the group activities per each week. There are teachers and teaching assistants in the computer rooms, do not hesitate to ask. We are there to help you learn!
Fridays we have Quizzes, to make sure that you have mastered the material of the week, and you can move on. You will receive feedback on the Quizzes by following Monday.

At the end of week three, you should have learned key concepts related to data management, how to design and use relational databases, practiced with SQL, and be a member of a team and found a project to work on.

The **second three weeks** the course switches in “groupwork mode”. The focus is to put in practice your data management skills in a team assignment. Also, it involves lectures on advanced topics as per the program found on Blackboard.

The students pick a practical assignment to design and build a database for a real client. Each team of 4-5 students has a dedicated supervisor (teaching assistant). Teams should contact the supervisor to report their progress and get feedback **at least twice a week**.
Each team member needs to present the group work to other teams in an interactive session scheduled for in Week 6, in order to provide and receive feedback. Each team will deliver a final report and a working database, after incorporating feedback from the presentations. The report should incorporate at least one section inspired from the “advanced topics” lectures.

During Weeks 4-6:
Mondays and Wednesdays we have lectures on more advanced topics. These include study material beyond the book and can be found on the Blackboard.

Tuesdays, Thursdays and Fridays we have computer rooms scheduled for working on your group assignments. While teams are not obliged to be in the computer rooms, there will be teaching assistants on site to help you with your group work.

Assessment strategy

The course is assessed in two complementary ways:
- three (3) individual closed book quizzes (50%, needs a minimum mark of 5.0 to pass);
- team assignment (50%, needs a minimum mark of 5.5 to pass).

Quizzes are scheduled in the three first weeks.
There is one quiz re-examination in week 4 for cases of force majeure.

Quizzes cover the material of the Week the correspond to. i.e. Quiz 1 assesses the learning goals and the material of Week 1. Quizzes may include different types of questions including multiple choice questions, questions to fill in the blanks, or short questions. There are points assigned to each question.

The assessment of the team assignment is based on the report and the database that was developed. The grading involves six elements, that are combined altogether into one final team mark: 1) data modelling, 2) SQL, 3) independence, and effort, 4) data dictionary, 5) report quality, 6) advanced topics. The report should justify design choices, include at least one section debating on advanced topics, and include a reflection on the challenges that you faced during the team work.

The final mark is the average of the individual quizzes and the team assignment grades.

- Note that you not need to score above 5.0 in all three quizzes, but to have an average above 5.0, i.e. you can pass the course even if you fail one quiz.
- Note that the Quiz retake is only for cases of force majeure, not for improving your grade.

The table below links learning outcomes to assessment.

Learning outcomes	**Quizzes**	**Group work**
1. demonstrate a managerial perspective on an organization's memory;	X	
2. explain key concepts of data modelling and databases (i.e. entities, relationships, primary and foreign keys);	X	
3. interpret data model diagrams using different notations (E-R diagrams);	X	
4. compile database queries with SQL; including nested subqueries, arithmetic, logical and spatial operations;	X	x
5. analyse a realistic data problem and propose a data design solution;	x	X
6. design and implement a database for a problem in their field of study;		X
7. debate some of the contemporary challenges in data management, as FAIR principles, security, big data, blockchains, privacy, licensing, etc.		X

The principal themes and schedule

The table below outlines the principle themes of the course and the schedule

week	theme	In lecture rooms	In computer rooms and/or out of class	
			Self study	Team assignment
1	Introduction	2 lectures 1 Quiz	- Study Book Ch 1-4 - Exercises Ch 1-3 - Tutorial MS Access - Self Test	Create team, brainstorm
2	Database design and simple SQL	2 lectures 1 Quiz	- Study Book Ch 4-7, Ref 1 - Exercises Ch 4-7 - Tutorial MySQL - Self study module on client-server databases - Self Test	Pick topic, contact potential clients
3	Advanced SQL and design choices	2 lectures 1 Quiz	- Study Book Ch 8-10 and Playbook - Exercises - Self Test	Write proposal
4	Advanced topics: - Intellectual property - Privacy - Open data - FAIR principles	2 lectures Quiz retake	- Study online material related to IP, Privacy, Open Data, FAIR	Design database, queries
5	Advanced topics - Big Data - Data Mining	2 lectures	- Study online material related to Big data and Data mining	Build database, insert data, start report
6	Final line	presentation session		Test, present to other teams, finalize
7	Submit report			Monday 12:00 deadline

Course schedule

A detailed time-schedule of activities are classrooms is available at:
https://ssc.wur.nl/Handbook/CourseSchedule/INF-21306