5DV120 Databasteknik Spring 2015
Database System Principles

1 Course Staff

Instructor in charge:	Name:	Stephen J. Hegner
(kursansvarig)	Office:	C444 MIT-huset
	Telephone:	090 786 5760
	E-mail:	`hegner@cs.umu.se`
	URL:	`http://www.cs.umu.se/~hegner`
	Office hours:	1015-1100 on days with a scheduled lecture
Instructor:	Name:	Johan Mollevik
	Office:	B444 MIT-huset
	Telephone:	090 786 6701
	E-mail:	`mollevik@cs.umu.se`
	URL:	`http://www.cs.umu.se/~mollevik`
	Office hours:	1600-1700 on March 31 and April 1

- Johan Mollevik will handle the introductory lecture constituting a brief review of SQL and how to use the PostgreSQL system. He will also have consulting hours during the week following this lecture for that material only. He will have no further duties for the course.
- Stephen J. Hegner is the main instructor for the course and will have sole responsibility for the course beginning with Week 15 (April 06), including the grading of all obligatory exercises.

2 Course Language

The official language of the course is English. The lectures will be given in English, and since one of the goals of this course is to develop skills in presenting technical ideas in English, all written work must be submitted in English as well. Furthermore, the questions on the final examination will be written in English, and answers must be written in English. For the final examination, it will be permitted to use an XX-English / English-XX dictionary, where XX is a natural language of the student's choice.

Questions may be asked during the lectures in either English or Swedish. However, as there will likely be students in the class who do not speak Swedish, questions posed in Swedish will have to be translated by the instructor into English. It is therefore preferable to ask questions in English whenever possible.

3 Course Literature

The main textbook for this offering of the course is the following.

Version: 12 April 2015

- Abraham Silberschatz, Henry F. Korth, and S. Sudarsham, *Database System Concepts*, Sixth Edition, McGraw-Hill, 2011; ISBN: 978-0-07-128959-7 (international paper edition).

The sixth edition of this book is also available in hardcover, ISBN: 978-0-07-352332-3. The contents of these two version are the same; only the type of binding and the price are different.

The sixth edition of this book has been re-organized substantially from earlier ones, and includes important new material as well. Since the course will follow the new organization, use of earlier editions is not recommended.

The following textbook has been used in the introductory course 5DV119, Introduction to Database Management, and is useful as a secondary reference:

- Ramez Elmasri and Shamkant B. Navathe, *Database Systems: Models, Languages, Design, and Application Programming*, Sixth Edition, Addison-Wesley, 2011; ISBN: 978-0-13-214498-8.

While it is also a comprehensive textbook and a very good introduction to the subject, its technical level on some topics is not adequate for 5DV120. Thus, it is not recommended as the primary technical reading for this course.

In addition to the course textbook, there will be relatively detailed overhead slides. These materials will be available for download on the course Web page.

4 Course Content and Outline

The official kursplan is available in Swedish here and in English here. A more offering-specific outline is shown below. The numbers shown in the single rectangular brackets (i.e., [..]) identify chapters and sections in the sixth edition of the textbook. The numbers in angle brackets $\langle .. \rangle$ indicate the approximate number of 45-minute lecture periods which will be devoted to the topic.

- Reasonably detailed overhead slides will be available for many topics. The authoritative source for relevant (i.e., possible examination) material is the course lectures and these slides. In many cases, material not covered in the textbook may nonetheless be covered in lecture presentations.

- The number of 45-minute lecture "hours" to be devoted to each topic is approximate, and in particular is rounded to the nearest integer. Adjustments will be made as the course progresses, and so the table below should not be used a definitive guide to which topics will be covered on which days.

- The slides are *not* a self-contained written record of all that will be covered during the class meetings. Students are responsible for all material which is covered in the course, regardless of whether or not it is found in the slides.

0. DBMS usage at the Department of Computing Science ⟨1⟩
1. Introduction to the course proper ⟨1⟩
2. Data and query languages
   - 2.1 Review of the relational model and SQL [2-4] ⟨1⟩
   - 2.2 Selected topics in advanced SQL [5.3-5.6,20.2] ⟨3⟩
   - 2.3 XML and XQuery [23] ⟨3⟩
3. Systems aspects
   - 3.1 Physical database design [10-11] ⟨4⟩
   - 3.2 Transaction models and concurrency control [14-15,10.8] ⟨6⟩
   - 3.3 Recovery from errors and from failure [16] ⟨4⟩
   - 3.4 Review of the relational algebra [6.1] ⟨1⟩
   - 3.5 Query processing and optimization [12-13] ⟨3⟩

5 Course Materials Outline

5.1 Textbook Materials Outline

The following is a list of those chapters and sections of the main textbook which will be covered in the course. For each chapter or section, a symbol is given which indicates the nature of coverage in the course. The meaning of these symbols is provided in the table below.

✔	Material will be covered in the course.
✘	Material will not be covered in the course.
✙	Review material, prior knowledge is expected.
✣	Material will be covered partially or selectively.

Note the following:

- If an entire chapter is covered, no section-by-section breakdown is given.
- Entries have not been provided for sections entitled “Summary” or the like.
- In general, omitted items will not be covered. However, the possibility that some covered material may appear in an omitted chapter or section remains. In all cases, the lectures and course notes should be taken to be the definitive statement for the course material.

1. Introduction ✙

5DV120, Syllabus, page 4

2. Introduction to the Relational Model ✙
3. Introduction to SQL Constraints ✙
4. Intermediate SQL ✙
5. Advanced SQL
    - 5.1 Accessing SQL from a Programming Language ✙
    - 5.2 Functions and Procedures ✣
    - 5.3 Triggers ✔
    - 5.4 Recursive Queries ✔
    - 5.5 Advanced Aggregation Features ✔
    - 5.6 OLAP ✔
6. Formal Relational Query Languages ✙
7. Database Design and the E-R Model ✙
8. Relational Database Design ✙

10 Storage and File Structure

- 10.1 Overview of Physical Storage Media ✔
- 10.2 Magnetic Disk and Flash Storage ✔
- 10.3 RAID ✔
- 10.4 Tertiary Storage ✣
- 10.5 File Organization ✔
- 10.6 Organization of Records in Files ✔
- 10.7 Data Dictionary Storage ✔
- 10.8 Database Buffer ✔

11 Indexing and Hashing

- 11.1 Basic Concepts ✔
- 11.2 Ordered Indices ✔
- 11.3 B^+-Tree Index Files ✔
- 11.4 B^+-Tree Extensions ✔
- 11.5 Multiple-Key Access ✔

Version: 12 April 2015

- 11.6 Static Hashing ✙
- 11.7 Dynamic Hashing ✔
- 11.8 Comparison of Ordered Indexing and Hashing ✔
- 11.9 Bitmap Indices ✣
- 11.10 Index Definition in SQL ✣

12 Query Processing

- 12.1 Overview ✔
- 12.2 Measures of Query Costs ✔
- 12.3 Selection Operation ✔
- 12.4 Sorting ✔
- 12.5 Join Operation ✔
- 12.6 Other Operations ✔
- 12.7 Evaluation of Expressions ✔

13 Query Optimization

- 13.1 Overview ✔
- 13.2 Transformations of Relational Expressions ✔
- 13.3 Estimating Statistics of Expression Results ✔
- 13.4 Choice of Evaluation Plans ✔
- 13.5 Materialized Views ✔
- 13.6 Advanced Topics in Query Optimization ✣

14 Transactions

- 14.1 Transaction Concept ✔
- 14.2 A Simple Transaction Model ✔
- 14.3 Storage Structure ✔
- 14.4 Transaction Atomicity and Durability ✔
- 14.5 Transaction Isolation ✔
- 14.6 Serializability ✔
- 14.7 Transaction Isolation and Serializability ✔
- 14.8 Transaction Isolation Levels ✔

Version: 12 April 2015

- 14.9 Implementation of Isolation Levels ✔
- 14.10 Transactions as SQL Statements ✔

15 Concurrency Control

- 15.1 Lock-Based Protocols ✔
- 15.2 Deadlock Handling ✔
- 15.3 Multiple Granularity ✔
- 15.4 Timestamp-Based Protocols ✣
- 15.5 Validation-Based Protocols ✣
- 15.6 Multiversion Schemes ✔
- 15.7 Snapshot Isolation ✔
- 15.8 Insertion Operations, Delete Operations, and Predicate Reads ✣
- 15.9 Weak Levels of Consistency in Practice ✔
- 15.10 Concurrency in Index Structures ✣

16 Recovery System

- 16.1 Failure Classification ✔
- 16.2 Storage ✔
- 16.3 Recovery and Atomicity ✔
- 16.4 Recovery Algorithm ✔
- 16.5 Buffer Management ✔
- 16.6 Failure with Loss of Nonvolatile Storage ✔
- 16.7 Early Lock Release and Logical Undo Operations ✣
- 16.8 ARIES ✣
- 16.9 Remote Backup Systems ✘

20 Data Warehousing and Mining

- 20.1 Decision Support Systems ✣
- 20.2 Data Warehousing ✔
- 20.3 Data Mining ✣
- 20.4 Classification ✘
- 20.5 Association Rules ✣

Version: 12 April 2015

20.6 Other Types of Associations ✣

23 XML

23.1 Motivation ✔
23.2 Structure of XML Data ✔
23.3 XML Document Schema ✔
23.4 Querying and Transformation ✔
23.5 Application Program Interfaces to XML ✘
23.6 Storage of XML Data ✘
23.7 XML Applications ✘

27 PostgreSQL

27.1 Introduction ✣
27.2 User Interfaces ✣
27.3 SQL Variations and Extensions ✣
27.4 Transaction Management in PostgreSQL ✔
27.5 Storage and Indexing ✔
27.6 Query Processing and Optimization ✔
27.7 System Architecture ✔

5.2 Online Materials

The Web site for the course is located at:

`http://www.cs.umu.se/kurser/5DV120/V15/index.html`

The following materials may be found on these pages.

1. This syllabus, in both PDF and HTML.
2. The lecture slides for the course.
3. Descriptions of the obligatory exercises.
4. Some miscellaneous exercises which are not obligatory, but which will be discussed in class.
5. SQL and XML code associated with the slides and exercises.

6. Information on the database system *PostgreSQL* as well as others.
7. Some useful links for *SQL* and *XML*.
8. Miscellaneous links to database-related things.
9. Some official documents required by the Department of Computing Science.

The Web site for the textbook is located at:

http://www.db-book.com

The following materials may be found there.

1. An errata sheet. (Since the number of printing errors is substantial, it is important to download this sheet and mark the appropriate corrections in your copy of the book.)
2. Lecture slides prepared by the authors. (These are not the same as those used in the course, but they are nevertheless provide a useful overview of the material.)
3. SQL code to generate the *University* database schema, as well as the associated instance, which are used in the textbook.
4. Other useful things not listed here.

6 Laboratory Schedule and Computer Resources

There is no official laboratory booking for the course, nor any in-laboratory instruction. In general, when not reserved by a course, the computer laboratories of the department are open for use by students for their coursework.

The relational database system *PostgreSQL* (and possibly *MySQL* as well), will be used in this course, as will the XML database system *eXist-db*.

7 Course Schedule

The table below identifies the course meeting times and places, together with the nature of the meeting. The key "L" denotes a lecture, "R" a review session, and "E" an examination booking.

For each lecture, the topics to be covered are identified via the outline header number of Section 4 of this syllabus. So, for example, on April 23 the topics of 3.1 (physical database design) will be covered. This is only an approximate assignment of meeting times to topics, and it may be altered as the course progresses.

Rooms whose identifiers begin with the letter *M* are located in MIT-huset, rooms whose identifiers begin with an *N* are located in Naturvetarhuset, and rooms beginning with *UB* are

near the University Library. The examination rooms (skrivsalar) are in the building known as Östra paviljongen.

Meetings scheduled for 0815-0955 will consist of two 45-minute sessions, 0815-0900 and 0910-0955, with a ten-minute break between them.

Week	Day	Date	Time	Type	Room	Topics	Instructor
13	Wed	Mar 25	0910-0955	L	UB333	0	Mollevik
15	Thu	Apr 09	0815-0955	L	N300	1, 2.1	Hegner
16	Mon	Apr 13	0815-0955	L	N300	2.2	Hegner
16	Thu	Apr 16	0815-0955	L	N300	2.2, 2.3	Hegner
17	Mon	Apr 20	0815-0955	L	N300	2.3	Hegner
17	Thu	Apr 23	0815-0955	L	N300	3.1	Hegner
18	Mon	Apr 27	0815-0955	L	MC413	3.1	Hegner
18	Thu	Apr 30	0815-0955	L	MC413	3.2	Hegner
19	Mon	May 04	0815-0955	L	MC413	3.2	Hegner
19	Thu	May 07	0815-0955	L	MC413	3.2	Hegner
20	Mon	May 11	0815-0955	L	MC413	3.3	Hegner
20	Wed	May 13	0815-0955	L	MC413	3.3	Hegner
21	Mon	May 18	0815-0955	L	MC413	3.4, 3.5	Hegner
21	Thu	May 21	0815-0955	L	MC413	3.5	Hegner
22	Mon	May 25	0815-0955	L	MC413	TBA/Review	Hegner
23	Tue	Jun 02	0900-1500	E	Skrivsal 6	Final examination	
35	Fri	Aug 28	0900-1500	E	Skrivsal 8	Final examination	
02	TBA	Jan ??	TBA	E	Skrivsal ?	Final examination	

8 Prerequisites

The formal requirements are listed in the course plan, which may be found using the link in Sec. 4 above. They include the following.

1. A knowledge of basic database management. This requirement is met by the formal prerequisite of the course *Introduktion till Datbashantering* (Introduction to Database Management).

2. A knowledge of programming in C in the Unix/Linux environment. This requirement is met by the formal prerequisite of the course *Systemnäraprogrammering* (Systems Programming).

3. A broad background in computer science consisting of at least 60 credits (one full year), or at least two years of study in a related discipline.

There are also implicit prerequisites, which are prerequisites of the courses given above.

Version: 12 April 2015

1. A thorough knowledge of data structures and algorithms, as presented in the course *Datastrukturer och algoritmer* (Data Structures and Algorithms).
2. A knowledge of discrete mathematics and the formal foundations of computer science. This requirement is met by the course *Diskret matematik* (Discrete Mathematics).

9 Grading System

This course has two parts (*moment* in Swedish), a conceptual part (*teoridelen* in Swedish) and an exercise part (*laborationsmoment* in Swedish).

The only possible grades for the exercise part are S (Satisfactory; G=*Godkänd* in Swedish) and U (Unsatisfactory, *Underkänd* in Swedish). The grade on this part will be determined entirely by five obligatory exercises. Each will be graded as S (Satisfactory) or U (Unsatisfactory). To obtain the grade of S for the exercise part of the course, it is both necessary and sufficient to obtain the grade of S on all five obligatory submissions. In addition, for each obligatory exercise except for the first, it will be possible to earn a maximum of 50 quality points. These points will be assigned based upon the overall quality and correctness of the work. Thus, a maximum of 200 points may be earned on the four obligatory submissions. The first obligatory exercise carries no points.

The examination will have a total of 1000 points.

The final point total F for the course is computed as

$$F = \max(E, \quad 0.8 \times E + L)$$

with E the number points earned on the examination and L the number of points earned on the obligatory exercises. The final grade on the conceptual part of the course is computed as follows.

Number of points	Grade
$F \geq 800$	5 (med beröm godkänd – excellent)
$650 \leq F < 800$	4 (icke utan beröm godkänd – very good)
$500 \leq F < 650$	3 (godkänd – satisfactory)
$F < 500$	U (underkänd – unsatisfactory)

In addition, to pass the course, a minimum of 500 points on the examination is necessary, regardless of how many points are earned on the exercises. Thus, exercise points can only be used to increase the grade from 3 to 4, or from 4 to 5. They cannot be used to rescue a performance of less than 50% on the examination.

10 Obligatory Work

The course includes five obligatory written and computer exercises, numbered 0-4. The rules governing these submissions are as follows.

- Points will be assigned to the submitted solutions for each exercise except Exercise 0. These points will be based upon the quality of the solutions, as well as their timeliness.
- Exercises will furthermore be graded as satisfactory or unsatisfactory. To receive a passing grade in the course, the student must have submitted a satisfactory solution to each exercise.
- Each exercise will have a due date. For each working day or fraction thereof that the submission is late, five points will be subtracted from the grade (except for Exercise 0). (The grade may never be less than zero, of course.)
- Exercises marked as unsatisfactory may be resubmitted, in order that a satisfactory evaluation be obtained. However, resubmitted work will never receive any additional points.
- Each exercise, except Exercise 0, is worth 50 points. There are no points for Exercise 0.

10.1 General Remarks on the Obligatory Work

- The obligatory exercises and project may be completed in groups, roughly as described in the documents *Riktlinjer vid labgenomförande (Policy for Obligatory Exercises)* and *Hederskodex (Honor Code)*. More details will be provided later, when the descriptions of these exercises are distributed.
- The obligatory exercises may be submitted jointly, with up to five persons submitting one solution. However, once a solution is submitted, only those named on the submission will receive credit for it. Partners in solution may not be added after the initial submission.
- Grading resources are limited. Therefore, work submitted well after the deadline may not be graded until the next examination. If you must submit work late, discuss this with the course assistant.
- All obligatory exercises and projects must be submitted for grading on or before the third and final examination (during January 2016). No exercises will be accepted after that date.

10.2 Deadlines for Submission of Obligatory Work

- The deadlines for exercise submissions are as follows.

Exercise	Deadline	Deadline Type
Exercise 0	April 07, 2015 at 08:00	Initial
Exercise 1	April 23, 2015 at 14:00	Initial
Exercise 2	April 30, 2015 at 14:00	Initial
Exercise 3	May 07, 2015 at 14:00	Initial
Exercise 4	May 21, 2015 at 14:00	Initial
All	August 28, 2015 at 08:00	Second
All	January ??, 2016 at 08:00	Final

- The initial deadline for each exercise is the time at which points begin to be deducted for lateness.
- An exercise solution which is submitted in advance of its initial deadline will be graded in a timely fashion, and students will be given the opportunity to improve unsatisfactory submissions and to resubmit them for an additional evaluation.
- For each obligatory exercise, there will be a resubmission deadline, defined to be five working days after the graded initial submission is returned. The resubmissions will be graded within ten working days of the resubmission deadline. This applies only to resubmissions of initial submissions which were received no more than four days after the deadline. For initial submissions which are more than four days late, there are no guarantees.
- Re-grading of submissions, other than for first resubmissions which are received no more than four working days after the initial deadline, is entirely at the discretion of the grader. The only exception is that all remaining submission will be graded within ten working days of the second and final resubmission dates.
- The final resubmission date coincides with the date of the third examination, which has not yet been announced.
- With the exception of extenuating circumstances (which will generally require written documentation), no submissions will be accepted after the third and final examination.

10.3 Obligatory Work Completed in Previous Years

- Both points and a satisfactory grade for corresponding obligatory exercises which received the grade of satisfactory may be carried forward from the previous offering (Spring 2014), but not further back in time. The student must request the transfer of points explicitly; it will not be done automatically.
- Points for an unsatisfactory exercise may not be carried forward from a previous offering.

5DV120, Syllabus, page 13

- An exercise which is carried forward from the previous year may not be shared with group members unless those members were also in the group which completed the exercise in the previous offering.
- Once a solution for a given exercise is submitted, a carry forward from the the previous year for that exercise cannot be obtained, either for points or for a satisfactory rating.
- There was no exercise in 2014 corresponding to Exercise 0 for 2015.