CSE 341

Database Systems, Algorithms and Applications

Spring 2020 (April 26, 2020)

CHECK ON PIAZZA FOR UPDATES DURING THE SEMESTER!!!!!!!

Instructor:

- Hank Korth, email: hfk@lehigh.edu, PA 414
- TA: Jacob Nelson, jjn217@lehigh.edu
- Graders:
  - Emily Mohrenweiser, eam220@lehigh.edu
  - Anmol Shresta, ans221@lehigh.edu
  - Jason Spector, jds221@lehigh.edu
  - Yuming (Marvin) Tian, yut222@lehigh.edu
  - Shengli (Shirley) Zhu, shz220@lehigh.edu

- For fastest response, use the Piazza site (details in main part of the syllabus) to post your questions. You should get an answer quickly from the instructor, the grader, or a fellow student. But (of course) contact us directly by email or in person regarding items of a personal matter.
- Office hours posted on Piazza: see resources/staff

ABET-format Syllabus
CSE 341 Database Systems , Algorithms, and Applications

Catalog description: Design of large databases; normalization; query languages (including SQL); Transaction-processing protocols; Query optimization; performance tuning; distributed systems. Not available to students who have credit for CSE 241 or IE 224. Prerequisites: CSE 17 or CSE 18 or consent of the instructor.

Credit hours: 3

Class/laboratory schedule: This class has two 75-minute meetings a week.

Instructor: Henry F. Korth

Required course for Computer Science and Business (alternative: CSE 241)
Elective course for Computer Engineering

Prerequisite: CSE 17

Textbooks/required materials: Silberschatz, Korth, Sudarshan, *Database System Concepts, 7th edition*, McGraw Hill 2020.

Course objectives:
Upon completing this course, students will: (*denotes primary objectives)

- Provide a strong foundation for further formal and informal study in Computer Science:
  - Principles of programming
  - Principles of data structures*
  - Principles of operating systems
- Provide a firm basis of science and mathematics
  - Perform experiments and analyze data
- Provide exposure to projects that have the elements of those the students will encounter on the job
  - Experience in a second language
- Integrate the business and computing science components of the program (for the CSB major only)
  - Apply computer technologies to the construction and control of business information systems
  - Design a database*

Relation to Student Outcomes (SOs) and Assessment: This course supports the following ABET criteria for computer science (1, 2, 7, 9, 11) and computer science and business (19, 21). It is an elective for computer engineering.

Criterion	Student Outcome	How Assessed
1	Write computer programs	JDBC homework and course project
2	Apply knowledge of data structures	Homework & test questions on B+-tree indexing
7	Apply principles of operating systems	Homework & test questions on buffer management and concurrent systems

9	Perform experiments and analyze data	Homework & test questions on OLAP and ranking queries
11	Write programs in a second computer language	Homework & test questions on SQL
19	Apply computer technologies to the construction and control of business information systems	Course project (design a database for an enterprise, implement design in Oracle, populate it with data, and build multiple end-user interfaces)
21	Design a database	Course project, and homework and test questions on entity-relationship design, and relational design.

Topics covered:

Note: The topic list below closely parallels that of CSE 241. The distinction in the courses is not mainly in the topics themselves but rather in the manner in which they are treated. CSE 341 incudes more formal rigor (mathematical foundations and proofs), more language details (formal database languages and query optimization) and more detailed coverage of how database systems interact with broader computing system functions such as those of the operating system.

- Relational algebra and calculus
- SQL: schema definition, queries, nested queries, updates, complex join expressions, integrity constraints, authorization, index creation, transactions
- Entity-relationship design
- JDBC, SQL functions, SQL stored procedures, map-reduce
- Triggers
- Recursive queries
- Ranking and OLAP queries
- Relational database design:
    - normalization and decomposition, losslessness, dependency preservation
    - proofs of properties of relational designs
    - soundness and completeness of axioms for functional dependencies
    - higher normal forms and other forms of redundancy
- Enterprise database implementation
- Storage management:
    - RAID disks
    - buffer management
    - index structures (both B+-tree and extendible hashing)
- Query processing algorithms (esp. join algorithms) and query optimization
- Transaction management:
    - ACID properties
    - Recovery:
        - write-ahead log, fault tolerance
        - operation logging and ARIES-style recovery
    - Concurrency:
        - Widely-used concepts: serializability, two-phase locking, snapshot isolation
        - Alternative approaches: timestamp ordering, validation
- Parallel and distributed databases (includes cloud)
    - Two- and three-phase commit
    - Distributed deadlock
    - Distributed query optimization
- Data mining and data warehouses
- Tuning and benchmarking

University-required syllabus statements:

- *If you have a disability for which you are or may be requesting accommodations, please contact both your instructor and the Office of Academic Support Services, Williams Hall, Suite 301 (610-758-4152) as early as possible in the semester. You must have documentation from the Academic Support Services office before accommodations can be granted.*

- *Lehigh University endorses The Principles of Our Equitable Community [http://www.lehigh.edu/~inprv/initiatives/PrinciplesEquity_Sheet_v2_032212.pdf]. We expect each member of this class to acknowledge and practice these Principles. Respect for each other and for differing viewpoints is a vital component of the learning environment inside and outside the classroom.*

Course Logistics

Text:

- Silberschatz, Korth, Sudarshan, *Database System Concepts, 7th edition,* McGraw Hill 2020.
- See **http://www.db-book.com** for useful materials, including errata, slides, and solutions to selected exercises.

Computer and Project:

- We will use the Oracle database system for projects. This year we are using Oracle 18c. Oracle expdp runs daily at 4am resulting in data being read-only from then until about 4:15am.
- The course project (to be distributed in week 3 of the semester) is to be done in Java. Students will be expected to submit a directory containing compilable Java source code (not an IDE project file) and a working jar file that was generated by compiling the submitted source code. All projects will be tested on the Sunlab machines using a standard-issue student account via terminal connection (ssh) with a default path and no IDE libraries. Students need Sunlab accounts to test their code. Any student who does not have a Sunlab account should contact us via email immediately; we'll need your Lehigh user id (e.g., xyz123@lehigh.edu).
- Please see a detailed statement of the strict project submission requirements in the project handout. Projects that do not run may get a zero on the executable part. The JDBC assignment will be structured as a test for the project submission rules.

Academic Integrity:

- I shall be submitting all code for the project to the Moss system to detect cases of suspected plagiarism. Moss cannot detect "outsourcing" of coding. Therefore, I shall also be selecting some students (based on a combination of Moss output and random selection) for meetings after the projects have been graded in which I shall discuss the specifics of the selected student's project with the student.
- For all exams and quizzes, backpacks, coats, etc. must be left at the front of the classroom. The will be no restroom breaks. Students needing an exception to the no-break rules must have the Student Health Center provide a statement to that effect.

CourseSite and Piazza: We shall be using both the Lehigh-supported CourseSite system and the externally-operated Piazza system. .

- Piazza will host our course discussions, assignments, and solutions to assignments.
  - You need to sign up for the Piazza site. You should already have received a welcome message from Piazza. If not, email hfk2@lehigh.edu to be added.
- CourseSite will host grade information. Note that the overall course grade provided by CourseSite will not always be accurate since newly posted assignments are treated as having a grade of zero until they are graded. To check your status in the course it is best to look at your grades and compute your grade against the breakdown listed on this syllabus.

Weather, Delayed Opening, and Illness:

- In the event of a late opening of the University due to weather, we shall start class at university opening time or normal class time, whichever comes later.

- If you are ill, your classmates will appreciate you not sharing your flu or virus. You are responsible for getting notes from a classmate. To avoid a zero on a pop-quiz for a day on which you are sick, be sure to email me *before* the time you would normally leave to go to class.

Grading:

- Late work policy: ***No late work will be accepted for credit without prior permission.*** Due dates will be stated on all assignments. If I erroneously set conflicting dates across Piazza, CourseSite, the syllabus,. and the assignment document, please inform me. Until any error is corrected the *earliest* date applies. Students are expected to be able to submit work correctly online and to back up their data. Therefore, "forgetting to click submit", "computer crashes", etc, are not acceptable lateness excuses. Note that online sites' clocks may not match yours perfectly, so don't wait until the last moment to submit. Absence from a scheduled quiz or exam due to illness requires a medical excuse from the Health Center.
- Religious holidays: If there are any scheduling issues something please inform us during the first week of class.
- Homework: 13 graded assignments: 16% (all 13 weighted equally). Homework 0 is an ungraded assignment to test account setup.
- Project: 33%: A full project document will be posted Jan 31. Checkpoints 1, 2, and 3 are worth 1 point each if completed on time. Remaining 30 points are assigned based on the completed project.
- Pop Quizzes 7%: Based on random choice, some classes will have a pop quiz. There will be no pop quiz in the first and the last week of the semester, nor on the two days when I will be away at a conference; for other classes the probability of a quiz is ⅓. This works to roughly 1 point per quiz. These quizzes will be simple questions from the prior class or the assigned reading and will take less than 5 minutes.
- Quizzes: 24%: 2 at 12% each. No makeup quizzes will be offered. In the event of a health-center approved medical excuse, the corresponding questions from the final will be used to generate a quiz grade.
- Final: 20%. date TBA, based on the registrar's exam schedule. Make-up exams will not be offered except when authorized by university policy. Do not make travel arrangements before the schedule is posted.

Schedule of Topics:

- The following schedule is tentative, but it is an accurate guide to content even if our pace slows or speeds up.
- There will be some topics in the assigned reading that are not covered in lecture, and vice versa. You are responsible for all material even if it is covered only in the reading or only in the lecture.
- DSC = *Database System Concepts,* the course text. All references are to the 7th edition.
- The powerpoint slides on the book web site follow the book organization. Since the book web site is public, the slides will not be posted on Piazza.

	Date	Topics	Book	Assignments
1	Tues Jan 21	Introduction, Overview, Relational algebra	DSC 1, 2.1-2.3, 2.5	Homework 0 (load SQL Developer; test account, set sunlab file protections)
2	Th Jan 23	Relational algebra	DSC 2.6	H0 due Homework 1 (relational algebra and calculus)
3	Tues Jan 28	Relational calculus	DSC 27.1-27.3	chapter 27 is available at db-book.com
4	Th Jan 30	Entity-relationship model	DSC 6.1 – 6.6	H1 due Homework 2 (ER Design)
5	Tues Feb 4	ER to relational, ER design issues; Extended ER project discussion;	DSC 6.7 -- 6.9, 6.11	Project handout
6	Th Feb 6	Oracle; SQL: basic schema definition, queries: joins, set ops, nulls, aggregate ops	DSC 3.1 – 3. 7	H2 due Homework 3 (SQL queries)
7	Tues Feb 11	SQL: nested queries, updates, complex join expressions, views, commit/rollback work	DSC 3.8, 3.9, 4.1 -- 4.3	ER review meetings. Details to be posted on Piazza.
8	Th Feb 13	SQL: schemas, integrity constraints, advanced schema definition, authorization, JDBC	DSC 4.4 – 4.7, 5.1.1,	H3 due Homework 4 (JDBC and more SQL queries) ER review meetings. Details to be posted on Piazza.
9	Tues Feb 18	JDBC PL/SQL: functions and stored procedures and triggers (special handout for PL/SQL; text covers SQL standard version)	DSC 5.2, 5.3	Instructor away at conference. Will give lecture via zoom. Details to be posted on Piazza. ER review meetings. Details to be posted on Piazza Homework 6 (PL/SQL, Triggers) [this is not an error. HW numbered in order of due date] Project checkpoint 1 deadline
10	Th Feb 20	Recursive queries, Ranking queries, OLAP	DSC 5.4 – 5.5, 11.1-11.3	Instructor away at conference. Will give lecture via zoom. Details to be posted on Piazza. H4 due Homework 5 (Ranking, OLAP), **due Mon Feb 24** so I can post answers before the quiz Homework 6 (PL/SQL) not due until after quiz
	Tues Feb 25	**Quiz 1, in class**		See Piazza for extra office hours You should practice PL/SQL prior to the quiz, but you can do your debugging after the quiz, since H6 is not due until Thurs Feb 27 RBC 184 9:20-10:35 RBC 241 (extra time students only) 8:40-9:20
11	Th Feb 27	Relational database design overview, keys, functional dependencies, losslessness, dependency preservation, BCNF, 3NF	DSC 7.1-7.3	H6 due H7 (inference with FDs, normalization)
12	Tues Mar 3	Inference with functional dependencies, normal form algorithms	DSC 7.4-7.5	
13	Th Mar 5	higher normal forms	DSC 7.6-7.7, 28	H7 due Homework 8 (proofs of theorems about FDs and normalization)

	Mar 9-13			**spring break** **good chance to work on project!** **(and don't forget H8!)**
14	Tues Mar 17	Other relational design issues Completeness of FD axioms	DSC 7.8-7.10	Completeness proof is not in book
15	Th Mar 19	Storage management: storage types, disk, SSD, RAID, buffer management;	DSC 12, 13.5	H8 due Homework 9 (storage and indexing) Project checkpoint 2
16	Tues Mar 24	B+-tree index and multi-key access Hash index structures Write-optimized indices	DSC 14.1-14.3, 14.5, 14.8	
17	Th Mar 26	Query processing	DSC 15	H9 due Homework 10 (query processing and optimization) Project checkpoint 3
18	Tues Mar 30	Query optimization	DSC 16	
19	Th Apr 2	Processing queries for data analytics and for streaming data	DSC 10.5, 10.6, 13.6 - 13.7, 11.4.4, 14.7	H10 due Project checkpoint 4
	Tues Apr 7	**Quiz 2, in class**		RBC 184 9:20-10:35 RBC 241 (extra time students only) 8:40-9:20
20	Th Apr 9	Transactions: serializability, isolation levels, two-phase locking, deadlock, phantoms	DSC 17, 18.1 (but not 18.1.5), 18.2	Homework 11 (concurrency)
21	Tues Apr 14	Multiple granularity, timestamps, validation	DSC 18.3 – 18.6	
22	Th Apr 16	multiversion, Snapshot Isolation, main-memory transactions	DSC 18.7, 18.8, 18.10	H11 due Homework 12 (snapshots and recovery) Project checkpoint 5
23	Tues Apr 21	Recovery and advanced recovery techniques	DSC 19	section 19.1 is just background reading
24	Th Apr 23	Parallel and distributed transactions, 2PC, 3PC, Paxos	DSC 20, 23.1, 23.2.1, 23.2.2, 23.4, 23.3.5, 23.3.6 23.8.1, 23.8.2, 23.4, 23.6.1	H12 due Homework 13 (distributed transactions) Ch 20 is background reading, not covered in lecture Wed Apr 22 11:55pm is the due date for the final ER diagram (see project handout)
25	Tues Apr 28	Parallel and distributed storage and sorting	DCS 21.1-21.4, 22.1, 22.2	MONDAY Apr 27 11:55pm is the due date for the final relational design and data (see project handout)
26	Th Apr 30	Query processing in distributed, parallel systems	DSC 22.3 - 22.5, 22.7	H13 due FRIDAY May 1 11:55pm is the due date for the project (see project handout and read submission rules carefully)

	TBD	final exam		no makeups for pre-purchased plane tickets. wait for exam schedule to be posted by the registrar!