Final Term Exam CSCC43 H3Y

Summer 2021

Duration: 12:00 to 3:00

Total Points: [150]

Question 0. [0 mark]

Write and then sign the declaration below. When you take a photo of it, place your picture ID beside and hide any sensitive information. Note you may have multiple questions on one page and in one photo.

"I have not consulted any resources including but not limited to classmates, tutors, textbooks, webpages, cheat sheets."

Signature: _____ Student Number: _____

There are total 8 questions with multiple subparts for each questions.

You can submit single PDF for each question (total 8 pdfs), where each PDF will contain answer for single question.

Refer to this table to for each question and number of subparts for that question

Questions	**Number of sub questions**	**Marks**
Question 1 multiple choice	15	15
Question 2 Relational Algebra Schema I	5	15
Question 3 Short answers	2	5
Question 4 Relational Algebra Schema II	6	15
Question 5 SQL Queries DML	8	30
Question 6 SQL DDL	3	10
Question 7 ER Diagrams	3	20
Question 8 Normal forms	5	40

Question 1 Warm-up multiple choice questions [15 marks]

Descriptions: For following 15 questions, mention the correct option and upload a single file that contains answers for all the 15 questions. Make sure to number your answers correctly.

Question 1.1

What is an update anomaly? Choose one of the following:

(a) One a transaction reads an element that was updated by an earlier, uncommitted transaction.

(b) The application wants to update a foreign key to a new value that does not exists in the referenced relation.

(c) The same information is stored redundantly in the database, and only some, but not all copies are updated.

Question 1.2

State whether this statement is true or false:

Every relational schema in SQL is in 1st normal form

Question 1.3

State whether this statement is true or false:

Every XML data is in 1st normal form.

Question 1.4

Which of the following statements best describes the main reason for representing a relational database in 1st normal form?

(a) To achieve physical data independence.

(b) To remove data anomalies (insertion, update, deletion anomalies).

(c) To save space on disk.

Question 1.5

Which of the following statements best describes the main reason for representing a relational database in BCNF?

(a) To achieve physical data independence.

(b) To remove data anomalies (insertion, update, deletion anomalies).

(c) To save space on disk.

Question 1.6

Which guarantee can be relaxed for the following use case? (Partition tolerance, Availability, Consistency)

"Data is too big to be served on a single server and is not modified after creation."

(a) Partition tolerance

(b) Availability

(c) Consistency

Question 1.7

Which data model is the best for the use case: "A system needs to reflect a change immediately"

(a) BASE

(b) ACID

Question 1.8

Advantages of NoSQL are ____.

(a) It supports query language.

(b) It provides fast performance.

(c) It provides horizontal scalability.

(d) All of the above

Question 1.9

What is the correct sequence to create a database connection?

i. Import JDBC packages.
ii. Open a connection to the database.
iii. Load and register the JDBC driver.
iv. Execute the statement object and return a query resultset.
v. Create a statement object to perform a query.
vi. Close the resultset and statement objects.
vii. Process the resultset.
viii. Close the connection.

(a) i, ii, iii, v, iv, vii, viii, vi
(b) i, iii, ii, v, iv, vii, vi, viii
(c) ii, i, iii, iv, viii, vii, v, vi
(d) i, iii, ii, iv, v, vi, vii, viii

Question 1.10

Parameterized queries can be executed by?

(a) ParameterizedStatement
(b) PreparedStatement
(c) CallableStatement and Parameterized Statement
(d) All kinds of Statements

Question 1.11

State whether this statement is true or false:
$(X^+)^+ = X^+$

Question 1.12

State whether this statement is true or false:
BCNF is neither the third nor the fourth normal form.

Question 1.13

Assume that a relation R has following properties, what is the normal form of R?
Properties: Has no multi-valued attributes, has no partial key dependencies, has attributes wih atomic domains, has transitive dependencies

(a) 1NF, 2NF, 3NF
(b) 1NF, 2NF
(c) BCNF
(d) 1NF, 2NF, 3NF, BCNF

Question 1.14

A relation in which every non-key attribute is fully functionally dependent on the primary key and which has no transitive dependencies, is said to be in _____

(a) BCNF
(b) 1NF
(c) 2NF
(d) 3 NF

Question 1.15

Consider following four relational schemas. For each schema, all non-trivial functional dependencies are listed. The underlined attributes are the primary keys.

Assume that the attributes in each schema are the only attributes for that schema.

Schema I

Field 'courses' is a set-valued attribute containing the set of courses a student has registered for:

Non-trivial functional dependency:

rollno → courses

Schema II

Non-trivial functional dependency:

rollno, courseid → email

email → rollno

Schema III

Non-trivial functional dependency:

rollno, courseid → marks, grade

marks → grade

Schema IV

Non-trivial functional dependency:

rollno, courseid → credit

courseid → credit

Which one of the schemas above is in 3NF but not in BCNF?

(a) Schema I

(b) Schema II

(c) Schema III

(d) Schema IV

Question 2 Relational Algebra Schema I [15 marks]

Description: Consider following schema and answer queries in relational algebra expressions.
Upload a single file for all the sub-questions for Question 2.

Consider a database with the following schema:
Students (name, age, gender)
Visits_often (name, café)
Student_drinks (name, coffee)
Serves (café, coffee, price)

Question 2.1 [3 marks]
Find all the Cafés that are visited by only female students or by only male students.

Question 2.2 [3 marks]
Find the names of all students who go to *only* those Cafés serving at least one coffee they drink.

Question 2.3 [3 marks]
Find the names of all students who visit *every* single Cafés serving at least one coffee they drink.

Question 2.4 [3 marks]
Find the Cafés serving the cheapest vanilla latte. In the case of ties, return all of the cheapest-vanilla latte.

Question 2.5 [3 marks]

For each student, find the names of all coffees student drinks that are not served by any Cafés they visit often. Return all such student name and coffee pairs.

Question 3 Relational Algebra short answers [5 marks]

Descriptions: Short answers based on Relational Algebra. Upload a single file by appropriately naming two in Question 3.

Question 3.1 [2.5 marks]

Suppose you are given schema which has two relations A(P, Q) and B(Q, R).
Without assuming any information about the keys of these relations, consider following expressions in RA:

a. $\pi_{P,R}\left(A \bowtie_{\sigma_{Q=1}} \mathrm{B}\right)$
b. $\pi_P\left(\sigma_{Q=1}A\right) \times \pi_R\left(\sigma_{Q=1}B\right)$
c. $\pi_{P,R}\left(\pi_Q A \times \sigma_{Q=1}B\right)$

Out of three, there are two expressions that produce the same answer on all the databases, making them equivalent. One produces different answer.
Identify one expression that produces different answer.
To justify your answer, give an instance of a simple database that produces different answer.

Question 3.2 [2.5 marks]

Consider a relation $Emp(countryID, name, high, low)$ that records historical high and low employment percentages for various countries from 1920 to 2020. Employment percentage is measured every month for a country and highest and lowest percent is noted and entered in the Emp database for each year.

Countries have names, but they are identified by countryID.
Consider the following query.

$$R1(cID, h) := \pi_{countryID,high} Emp$$
$$R2(cID, l) := \pi_{countryID,low} Emp$$
$$R3\ (cID) := \pi_{cID}\left(R1 \bowtie_{h<high} \mathrm{Emp}\right)$$
$$R4\ (cID) := \pi_{cID}(R2 \bowtie_{l>low} \mathrm{Emp})$$
$$R5\ (cID) := \pi_{countryID} Emp - R3$$
$$R6\ (cID) := \pi_{countryID} Emp - R4$$
$$Final\_ans := \pi_{name}(Emp \bowtie (R5 \cup R6))$$

State in single English phrase what the $Final\_ans$ computes.

Question 4 Relational algebra Schema II [15 marks]

Description: Consider following schema and answer queries and constraints in relational algebra expressions.
Upload a single file for all the sub-questions for Question 4.
Assume that the database schema consists of four relations, whose schemas are:

Furniture(<u>ID</u>, company, type_furniture)
chair(<u>ID</u>, material, weight, premium, cost)
table(<u>ID</u>, material, weight, premium, length, cost)
cabinet(<u>ID</u>, color, type, cost)

A sample snapshot of these relations is given below. **Your answer should work for arbitrary data**, not just the data of these figures.

Furniture

ID	company	type_furniture
101	wellBuilt	chair
301	modestComfort	cabinet
201	IkeaStore	table

Chair

ID	material	weight	premium	cost
101	pine	25	adjustableHeight	150
102	mahogany	30	wheels	200
103	birch	32	metalFrame	190

Table

ID	material	weight	premium	length	cost
201	pine	30	wheels	114	150
202	Ash	35	adjustableHeight	150	200
203	Fir	42	metalFrame	100	190

cabinet

ID	color	type	cost
301	Red	wallMounted	600
302	Brown	Wide	550
303	Black	Slim	330

Question 4.1 [2 marks]
Find those companies that produce at-least two different chairs or tables with weight of at-least 25 kgs.

Question 4.2 [2 marks]
Find those companies that sell all the chairs and red cabinets.

Question 4.3 [2 marks]
Find those companies whose tables have all the weights that Ikea's tables have.

Question 4.4 [2 marks]
Express the constraint: A chair with a weight less than 10 kg must not sell for more than 20 CAD.

Question 4.5 [2 marks]
Express the constraint: No chair manufacturing company may also make tables.

Question 4.6 [5 marks]
Express the constraint A company manufacturing a chair must also make a table with weight at least the highest weight chair they make.

Question 5 SQL Queries DML [30 marks]

Description: Consider following schema and answer queries in SQL syntax.

Upload a single file for all the sub-questions for Question 5.

The following relations keep track of train routes information:

Trains(tid: integer, tname: string, fuelrange: integer)
Routes(rno: integer, departs: time, arrives: time, from: string, to: string, distance: integer, price: real)
Employees(eid: integer, ename: string, salary: integer)
Licensed(eid: integer, tid: integer)

Note:

- *fuelrange* defines the maximum distance a train can travel at a given speed without stopping for refueling.
- *Employees* relation describes motorman and other kinds of employees as well; every motorman is licensed for some trains, and only motormen are allowed to operate a train.

Question 5.1 [2 marks]

Find the *tids* of all trains that can be used on routes from Toronto to Jonquière without stopping.

Question 5.2 [2 marks]

For each motorman who is certified for more than three trains, find the *eid* and the maximum *fuelrange* of the train for which she or he is certified.

Question 5.3: [4 marks]

Print the names of motormen who can operate trains with fuelrange greater than 300 miles but are not certified on any Bullet Trains.

Note: You can assume the names of Bullet trains are of this format: "BulletHikari", "BulletKodama", etc.

Question 5.3 [4 marks]

Find the non-stop paths that can be operated by every motorman who makes more than 120k.

Question 5.4 [6 marks]

Find out the names and salary of all employees who are not motormen and whose salary is more than the average salary for motormen.

Question 5.5 [6 marks]

Consider a scenario where a passenger wants to travel from Toronto to Boston with no more than two changes of trains. List the choice of departure times from Toronto if the customer wants to arrive in Boston by 3 p.m.

Question 5.6 [6 marks]

Find the names of the all the trains with fuelrange over 300 miles and the average salary of all motormen licensed for such trains.

Question 6 SQL Queries DDL [10 marks]

Description: Consider following schema and answer DDL questions in SQL syntax.

Upload a single file for all the sub-questions for Question 6.

Consider the following relational schema and briefly answer the questions that follow

Worker (wid: integer, wname: string, age: integer, salary: real)
Works(wid: integer, cid: integer, time: integer)
Company (cid: integer, budget: real, supervisorid: integer)
Note:
A worker can work in more than one company and time attribute in Works shows the percentage of time that a given worker works in a given company.

Question 6.1 [2 marks]
Define a table constraint on Worker relation that will ensure that every worker makes at least 15k.
Begin your answer using CREATE TABLE statement that describes the schema for worker relation. Make sure your answer reflects all the information required to define Worker relation.

Question 6.2 [2 marks]
Define a table constraint on Company that will ensure that all the supervisors have age ≥ 30.
Begin your answer using CREATE TABLE statement that describes the schema for Company relation.
Make sure your answer reflects all the information required to define Company relation.

Question 6.3 [6 marks]
Write SQL statements to delete all information about workers whose salaries exceed that of the supervisor of one or more companies that they work in.
Be sure to ensure that all the relevant integrity constraints defined in other tables are satisfied after your updates.

Question 7 Database Design Theory [20 marks]

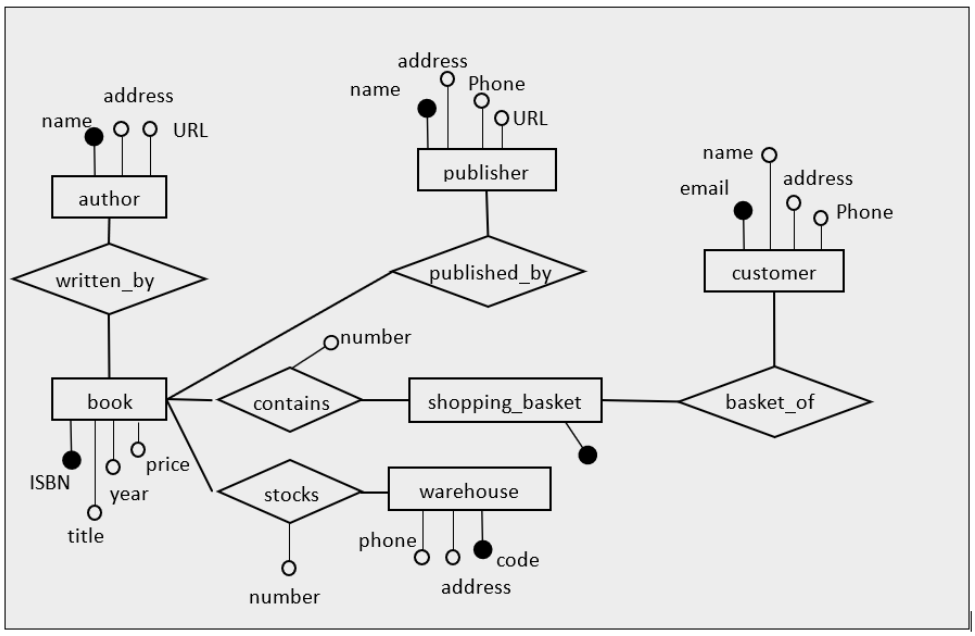
Description: Answer following questions based on Database design theory. Create a single PDF for all sub questions

Upload a single file for all the sub-questions for Question 7.

Question 7.1 [5 marks]
Consider the database of an online bookstore.
Suppose the bookstore decides to sell books in two more formats: audio book, and video(movie) of the book along with its regular paper version. Audio will store information such as length of the audio, video stores information such as size of the video, format of the video, etc., and paper version of the bookstores number of pages, etc.
Each version of the book might be present in one or more formats and sell with different prices.

- Draw the part of ER diagram that models this addition.
- Assuming that there are only 15 copies of each book in any of the three formats combined, and a customer can only buy at most 15 items, mention cardinalities only for concerned relation and entities.

Note: You only need to draw the addition/change that you will make to the existing ER diagram. You need not draw the entire ER diagram again.



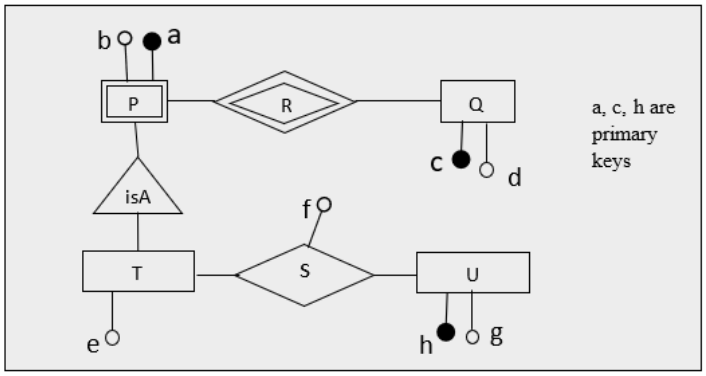
Question 7.2 [10 marks]

Construct an E-R diagram for a vehicle insurance company whose customers own one or more vehicles each. Each vehicle has associated with it zero to any number of recorded accidents. Each insurance policy covers one or more vehicles and has one or more premium payments associated with it. Each payment is for a particular period of time, and has an associated due date, and the date when the payment was received. Design an ER diagram for such a vehicle insurance company. Identify the cardinality constraints.
Explain if there is any weak entity in your ER diagram.

Construct appropriate relation schemas for vehicle insurance company and identify foreign keys.

Question 7.3 [5 marks]

Write the schema of the relations corresponding to all the entities and relationships in this diagram. You need to underline the keys of each relation, and mention if there is a foreign key referencing to another relation.
e.g. your answer should look like (this is not a real answer)
1. R1(m, n, o, p), where p references to attribute p of relation R2.
2. ....



Question 8 Normal forms [40 marks]

Description: Answer following questions based on Normalization. Create a single PDF for all sub questions

Question 8.1 [10 marks]

Santa Claus has many helpers in stores that deliver gifts, to talk to boys and girls about what they would like. These helpers have been using spreadsheets as crude databases to keep track of these visits. The spreadsheet entries have the following schema:

Visitor(name, age, address, email, item, color, weight)
From looking at the data, we've discovered the following functional dependencies between columns (attributes):

name → age	item → weight
name → address	email → name

(a) List all the superkeys of this relation and indicate which ones are also keys. Justify your answer in terms of functional dependencies and closures.

(b) Decompose the Visitor(name, age, address, email, item, color, weight) relation into BCNF using the functional dependency information from part (a).
If the relation is already in BCNF explain why. If one or more decomposition steps are needed, be sure to identify the functional dependency and closures used at each step so we can follow your work.

Question 8.2 [5 marks]

Consider following relation R(A, B, C, D, E) that satisfies the following functional dependencies:

$ABC \rightarrow D$
$E \rightarrow B$
$AD \rightarrow C$

Decompose the schema in BCNF. Show all your steps. Show your work for partial credit. Your answer should consist of a list of table names, attributes, projected dependencies, and an indication of the keys in each table.

Question 8.3 [5 marks]

Consider the following relational schema and set of functional dependencies.

R(A,B,C,D,E,F,G) with functional dependencies:
$A \rightarrow D$
$D \rightarrow C$
$F \rightarrow EG$
$DC \rightarrow BF$

Decompose R into BCNF. Show your work for partial credit. Your answer should consist of a list of table names, attributes, projected dependencies, and an indication of the keys in each table.

Question 8.4 [10 marks]

Consider a relation Stocks(B, O, I, S, Q, D), whose attributes may be thought of informally as broker, office of the broker, investor, stock, quantity of the stock owned by the investor and dividend of the stock. Let the set of FDs for Stocks be

$S \rightarrow D$
$I \rightarrow B$
$IS \rightarrow Q$
$B \rightarrow O$

a) What are the candidate keys for Stocks?
b) Are the given FDs their own minimal basis? Justify your answer.
c) Use the 3NF decomposition algorithm to find a lossless-join, dependency preserving decomposition of R into 3NF relations. Are any of the relations not in BCNF?

Question 8.5 [10 marks]

For the given relation R(A,B,C,D,E) and the given set of FDs
$F : \{A \rightarrow D, BC \rightarrow AD, C \rightarrow B, E \rightarrow A, E \rightarrow D\}$.
Find the minimal basis of F.
Show your work in steps.