AHSANULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY
Department: Computer Science and Engineering
Program: Bachelor of Science in Computer Science and Engineering
Semester Final Examination: Fall 2020
Year: 3rd   **Semester: 1st**
Course Number: CSE3103
Course Name: Database

Time: 2 (Two) hours   **Full Marks: 50**

Use single answer script

Instructions:	i)	Answer script should be hand written and should be written in A4 white paper. You must submit the hard copy of this answer script to the Department when the university reopens.
	ii)	You must write the following information at the top page of each answer script: **Department:** **Program:** **Course no:** **Course Title:** **Examination:** **Semester (Session):** **Student ID:** **Signature and Date:**
	iii)	Write down Student ID, Course number, and put your signature on top of every single page of the answer script.
	iv)	Write down page number at the bottom of every page of the answer script.
	v)	Upload the scan copy of your answer script in PDF format through provided **google form** at the respective course site (i.e., **google classroom)** using institutional email within the allocated time. Uploading clear and readable scan copy (uncorrupted) is your responsibility and you must cover all the pages of your answer script. However, for clear and readable scan copy of the answer script student should use only one side of a page for answering the questions.
	vi)	You must avoid **plagiarism,** maintain **academic integrity and ethics**. You are not allowed to take any help from another individual and if taken so can result in stern disciplinary actions from the university authority.
	vii)	Marks allotted are indicated in the **right margin.**
	viii)	Assume any reasonable data if needed.
	ix)	Symbols and characters have their usual meaning.
	x)	Before uploading, rename the PDF file as **CourseNo_StudentID.pdf** e.g., CSE3103_180204001.pdf
	xi)	The answer script (**one single pdf file**) must be uploaded at designated location in the provided **google form link** available in the google classroom.

There are 6 (Six) Questions. Answer any 4 (Four).

Question 1. [Marks: 12.5]

a)	Consider the following relation:	**[8]**

Articles (ID, title, journal, issue, year, startpage, endpage, TR_ID)

It contains information on articles published in scientific journals. Each article has a unique ID, a title and information on the journal where it got published (name of the journal, issue number, year and pages). Also, if the results of an article are previously published in an internal paper known as a 'technical report' (TR), the id of this technical report can be specified. The following is an instance of the relation.

ID	title	journal	issue	year	startpage	endpage	TR-ID
42	Cuckoo Hashing	JAlg	51	2004	121	133	87
33	Deterministic Dictionaries	JAlg	41	2001	69	85	62
33	Deterministic Dictionaries	JAlg	41	2001	69	85	56
39	Dictionaries in less space	SICOMP	31	2001	111	133	47
57	P vs NP resolved	JACM	51	2008	1	3	99
77	What Gödel missed	SICOMP	51	2008	1	5	98
78	What Gödel missed	Nature	2222	2008	22	22	98

Answer the following questions based on the above relation.
   i.   Write down at least two valid functional dependencies found in Articles.
   ii. Perform normalization up to 3NF over the given relation and state the resulting relations. Justify the steps you have taken in detail.

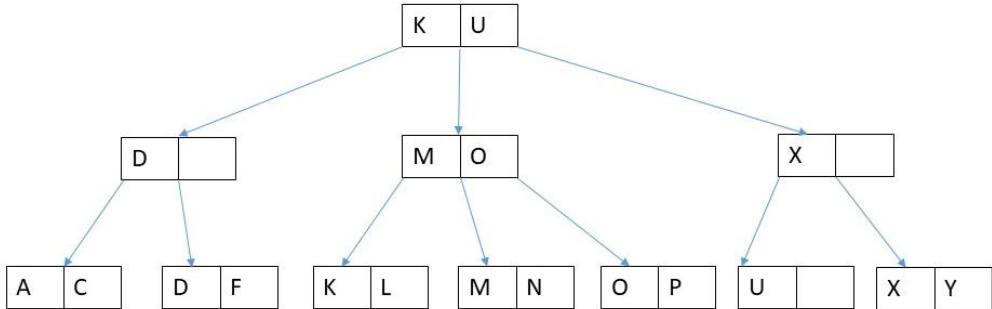| b)	Assume that you are the owner of a very popular online grocery store. Every day, your store has thousands of transactions and therefore, inserting a lot of data into your database. Finding some useful information out of this massive amount of data is a time-consuming task. How can you improve the performance of your database so that the searching would be faster? Justify your answer briefly by using relevant terms.	**[4.5]**

Question 2. [Marks: 12.5]

a)	Consider the following scenario that contains information on a hotel and its guests. Your job is to build an ERD model based on this scenario. Please note that the information is not complete, and you are free to make any reasonable assumptions. Your assumptions must be included in your answer. Guests can reserve a room or multiple rooms in a hotel. A guest is described by a unique identifier, name, age, gender, and phone number. A room is characterized by a room number, room type (e.g., double, single), and the room rate per night. While making a reservation, the guest should provide the expected check-in and check-out date.	**[8]**

Each reservation must be associated with payments. A guest would pay an amount in advance to confirm a room reservation, and the guest should pay the payment in full during check-out. Payment information must include a unique payment number, payment date and payment method (e.g., card, cash). Payments are made for a reservation and it existentially depends on the reservation.

A staff of the hotel will be assigned to each reservation till the guest check-out thereby ensuring happy and satisfied guests. Staff should have a unique identifier, name, age, phone number and number of working hours per week. At the time of check-out, the guest can provide feedback to the hotel management. Feedback must have a feedback identifier, feedback comment and customer satisfaction rating from 1 to 10.

b) Let L_1 and L_2 be two Instructions of transactions T_1 and T_2 respectively. Some item Q is accessed at the same time which consider the following schedule **[4.5]**

T_1	T_2	Result
$L_1 = Read(Q)$	$L_2 = Read(Q)$	No conflict L_1 and L_2
$L_1 = Read(Q)$	$L_2 = Write(Q)$	conflict L_1 and L_2
$L_1 = Write(Q)$	$L_2 = Read(Q)$	conflict L_1 and L_2
$L_1 = Write(Q)$	$L_2 = Write(Q)$	conflict L_1 and L_2

Is Q conflict concurrency control? Justify your answer by simplified a precedence rule for the above schedule.

Question 3. [Marks: 12.5]

a) Consider the following table *Student_Hobbies*. Suppose you want to create an index on the *student_phone* attribute. What type of index it should be? **Construct** and **draw** the index file pointing to the data file. **[8]**

Student_Hobbies

student_id	student_name	student_hobby	student_phone
2017-1-60-001	Alice	Cycling	01711987654
2017-1-60-001	Alice	Swimming	01711987654
2017-1-60-001	Alice	Cycling	01911600000
2017-1-60-001	Alice	Swimming	01911600000
2017-1-60-002	Bob	Cooking	01825654987
2017-1-60-002	Bob	Gardening	01825654987

b) Assume, transferring a block of query requires T_t seconds and seeking a block of query requires T_s seconds. To complete a query operation B block with S seeks, we require the query transferring cost formula to be $B * T_t + S * T_s$. In which step this formula can be used in the query processing system? Draw a diagram of the query processing system and explain. **[4.5]**

Question 4. [Marks: 12.5]		
a)	Consider the following scenario that contains information on researchers, academic institutions, and collaborations among researchers. Your job is to build an ERD model based on this scenario. Please note that the information is not complete, and you are free to make any reasonable assumptions. Your assumptions must be included in your answer. Researchers work in institutions. An institution can employ many researchers (possibly 0). Each researcher is described by a unique identifier, name, and year of birth. Each institution is characterized by a unique identifier, its name, country, and inauguration year. The database must keep the employment history of a researcher such as start date and end date of a placement in an institution. Moreover, each institution has many faculties (e.g., faculty of science and engineering, faculty of business, faculty of social relations, etc.). A faculty is described by the name and address and it is existentially dependent on the institution. A researcher is designated as the Chairperson of a faculty. A faculty also conducts many research projects in which several researchers can contribute. Each research project has a title, budget and, duration. Researchers may publish multiple papers as the outcome of a research project. A research paper is characterized by its digital object identifier (DOI), title and year of publication. A paper can include many researchers as its authors.	**[8]**
b)	Suppose you are selling RAID solutions to different customers. Let, customer A wants to set his/her system in a way that the storage devices are first to replicate then strip with the architecture. Customer B is looking for a global solution of storage devices which will be distributed in different geographical locations. Which RAID version will you suggest for customer A and customer B? Which version have some additive performance?	**[4.5]**

Question 5. [Marks: 12.5]		

a)	There are 3 relations in a football database schema. The relations are given below.
	      game (<u>id</u>, mdate, stadium, team1, team2)
	      eteam (<u>id</u>, teamname, coach)
	      goal (<u>matchid</u>, teamid, player, <u>gtime</u>)
	Construct Relational Algebra Expressions for the following queries based on the 'Football' database schema as given. Read the instruction in the parenthesis () carefully.
	
	  i. Find the output relation after applying full outer join in between *eteam* and *goal* relations.
	  ii. Find the players who scored goals in all matches of Germany (teamid: GER).
	  iii. Find the player who scored the slowest goal.
	  iv. Find the number of matches played at each stadium.
		[8]

b)	**Write** statements in SQL to create the relations as shown in the 'Football' Database given in Question 5(a).
	
	    i. Write a DDL statement using appropriate data type and define primary key and foreign key constraints for each relation.
	
	    ii. Add a new attribute in *game* relation that holds the number of audiences who enjoyed the match from the stadium. Choose appropriate attribute name and the datatype while writing the statement.
		[4.5]

Question 6. [Marks: 12.5]

a)	Consider the following B+ tree where branching factor m is three (03).
	
	 Perform the following operation separately and show the tree.
	
	    i) Insert the values "**E**" and **"Z"** sequentially in the given tree by following the rules in below:
	
	             LEFT Value < Variable Value
	             RIGHT Value >= Variable Value
	
	    ii) Delete the values **"M"** and **"U"** after the insertion operation is completed.
		[8]

b)	Write down the difference between Homogeneous and Heterogeneous distributed databases.	**[4.5]**