BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE

ADVANCED DATABASE DESIGNS AND MANAGEMENT - CS413

2 HOURS 30 MINUTES

JUN 2023

INSTRUCTIONS TO CANDIDATES

Answer **ALL** Questions.
Total marks are 100

Question 1

a) Prove the Armstrong's union rule. **[5]**

b) Let us assume a database for Web pages that keep data related to courses taught at a university. We consider the following part of the database schema:

webpage(webID, webTitle, url, hits)
courseware(cID, cDescription, ftpLocation, category)
ftpLink(webID, cID)

The relation *webpage* stores data about web pages. Each web page is described by a unique identifier *webID*, a title *webTitle*, its *url*, and the number of *hits* to it. The relation *courseware* contains data about the course material of a course. Each courseware record stores a unique identifier *cID* to identify a courseware file, a description *cDescription* of the courseware file, a directory *ftplocation* from where the file can be downloaded, and the *category* of courseware ('P' = publication, 'N' = notes, 'D' = document, 'E' = executable). The relation *ftpLink* stores for each Web page the ftp links to multiple courseware files and for each courseware file the different Web pages to which it is linked.

Write relational algebra for the following queries

i. Find web ids and web titles of web pages which provide courseware with the cID of "COP5725".

ii. Find web ids and web titles of web pages which ONLY provide courseware with the cID of "COP 5725".

iii. Find the url of web pages which have the number of hits between [5,000, 10,000).

iv. Find web ids and web titles of web pages which provide course materials with the cID of "COP5725" and the category of publication.

[4X3]

Question 2

a) A small Bank has two branches, one located in York and one located in Leeds. Currently the banks centralised database is managed at its HQ in London, where it keeps data about its customer accounts. Other uses include collecting data for the generation of reports and to monitor the services that customers use. Applications have been installed at the two branches to access the centralised database via a fast communication network for accessing the data they need. There is also a communication link between the two branches, which is currently used only when one of the main links to the London headquarters fail. For simplicity the centralised database system holds a single Customer table (see figure below), where data about customer accounts are kept. The columns of the Customer table are the account number, the customer's name, the branch where the account is kept and its current balance. The bank has been asked to split its branches so that they can support customers when there is failure of the centralised database or the communication system is slow or breaks down.

Figure 1 Examples of a customer relation held at HQ

Account_Number	Customer_Name	Branch	Balance
20012	Brown	York	1000.11
20454	Patel	Leeds	-250.89
61344	Smith	Leeds	58.99
n75630	Gray	York	2956.93
30435	Green	Leeds	-33.62
96567	Richards	York	45.76
01232	Akhtar	York	43.92

Describe **three** ways to distribute data across the three sites. Comment on the pros and cons of each option. **[15]**

b) Describe any **three** transparency features that exist in a distributed database environment. **[9]**

Question 3

The Tables (1A, 1B, 1C, 1D, 1E) are currently held on a database at the single site called CTR. Following reorganization, it is intended to distribute the journals held at the central database across 3 new branch libraries located at remote sites called GTR; UTC and TWE. The central library becomes the HQ (an administrative centre) meaning that it no longer keeps or loans out any journals itself. Instead, journals are made available for loan to borrowers registered at any of the 3 new sites.

Table 1A (Journal)

JOURNAL

JournalID	JournalName
3215	Database Weekly
3216	Database Journal
3217	Oracle News
3218	ACM TODS

Table 1B (Article)

ARTICLE

ArticleID	ArticleTitle	AuthorID	JournalID
3215	ObjectOriented Analysis	23	3216
2409	Oracle indexing	18	3217
1398	DBA performance tools	23	3216
1289	Pioneers of Databases	23	3215
2554	Query optimisation	67	3216
1678	Daplex	18	3218
4561	Niam Fact model	18	3218

Table 1C (Author)

AUTHOR

AuthorID	AuthorFName	AuthorSname
23	Norman	Gray
18	Carlos	Santos
67	Modrich	Neuman

Table 1D (Loan)

LOAN

ArticleID	BorrowerID	LoanDate	ReturnDate
2409	43	3/1/15	4/2/15
1398	43	3/1/15	24/1/15
1289	17	6/2/15	8/2/15
2554	26	1/2/15	12/2/15
2409	43	14/2/15	
2554	52	14/2/15	

Table 1E (Borrower)

BORROWER

BorrowerID	BorrowerFname	BorrowerLname	BorrowerTelNo
A52	Jane	Green	0156387562
43	Fred	Briggs	01985-86722
17	Henry	Dhura	01582-74238
26	Jonas	Smith	01933-632001

a) Describe **three** different proposals for data distribution of the central database (CTR). [9]

Hint: Show the distribution/replication of table fragments/partitions and explain any trade-off and pros/cons you think are relevant.

b) Consider the following relations containing airline flight information:

Flights (flno: integer, from: string, to: string,
distance: integer, departs: time, arrives: time)
Aircraft (aid: integer, aname: string, cruisingrange: integer)
Certified (eid: integer, aid: integer)
Employees (eid: integer, ename: string, salary: integer)

Note that the Employees relation describes pilots and other kinds of employees as well; every pilot is certified for some aircraft (otherwise, he or she would not qualify as a pilot), and only pilots are certified to fly.

Write the following queries in relational algebra:

i. Find the **eid** of pilots certified for some Boeing aircraft. [2]

ii. Find the names of pilots certified for some Boeing aircraft. [3]

iii. Identify the flights that can be piloted by every pilot whose salary is more than $100,000. [3]

c) Write the above queries in **(b.)** in tuple relational calculus. [6]

Question 4

a) Discuss the key characteristics of a data warehouse and how it differs in content, structure and function from an on-line transaction processing (OLTP) database. You should support your discussion with suitable diagrams and examples. [10]

b) Describe any **three** types of knowledge produced from data mining giving an example of each. [6]

Question 5

a. Micro loans are small loans, which are beginning to gain popularity especially among borrowers in developing countries. The idea is to bring venture lenders together using information technology. Typically, the loans will be used to finance startup or development of the borrower's company, so that there is a realistic chance for repayment. The money in a loan can, unlike traditional loans, come from many lenders. In this problem, you must create an E-R model that describes the information necessary to manage micro loans. The following information form the basis for creating the model:

- Each borrower and lender must be registered with information about name and address.
- A loan starts with a loan request, which contains information about when the loan should at latest be granted, the total amount being discussed (US-dollars), and how long the payback period is. Also, a description is included of how the money will be used. The rent on the payment is calculated in the loan amount, which is to say, the full amount is not paid.
- Lenders can commit to an optional portion of the total amount of a loan request.
- When the commitments for the loan request covers the requested amount, the request is converted to a loan. If not enough commitments can be reached, the loan request is cancelled. A borrower can have more than one request, and more than one loan at a time, but can at most make one request per day.
- The loan is paid through an "intermediary", typically a local department of a charity, who has a name and an address.
- The borrower chooses when he or she will make a payment. Every payment must be registered in the database with an amount and a date (at most one payment per

loan per day). The lenders share the repayment based on how large a part of the loan they are responsible for.

- If the loan is not repaid before the agreed upon deadline, a new date is agreed. The database must not delete the old deadline, but save the history (the deadline can be overridden multiple times).
- Each lender can for each burrower save a "trust", which is a number between 0 and 100 that determines the lender's evaluation of the risk of lending money to that person. The number must only be saved for the borrowers, for whom there has been made such an evaluation.

Make an E-R model for the data described above. If you make any assumptions about data that is not shown in the problem, they must be described. **[20]**

***** End of Exam *****