Programme Code: DT211C, DT228, DT282
Module Code: CMPU 4003
CRN: 30088, 22415, 31081

TECHNOLOGICAL UNIVERSITY DUBLIN
KEVIN STREET CAMPUS

BSc. (Honours) Degree in Computer Science (Infrastructure)

BSc. (Honours) Degree in Computer Science

BSc. (Honours) Degree in Computer Science (International)

Year 4

SEMESTER 1 EXAMINATIONS 2019/20

Advanced Databases

Dr. Pierpaolo Dondio
Dr. Deirdre Lillis
Ms. Pauline Martin – DT211C
Dr. Martin Crane – DT228/282

Two Hours

Answer **four** questions out of **five**.
Each question carries equal marks.

Question 1.   **Indexes**   **[25 marks in total]**

a. What is a Bitmap Index in Oracle?

[4 marks]

b. Consider a database storing information about all the students in Ireland. Would you define a Bitmap Index on the field "Student Name"? Justify your answer.

[3 marks]

c.

    (i) Insert the following data in a 2-3 B-tree index.

        5, 6, 7, 4, 3, 2, 1, 8, 9, 10

[10 marks]

    (ii) Suppose your database is receiving search queries for each number from 1 to 10, each query with the same frequency. What is the average number of nodes that has to be visited to answer those queries?

[4 marks]

    (iii) Suppose you are not using an index, but your data are stored in an unsorted table in the order shown above. What is the average number of records that has to be visited to answer the same queries?

[4 marks]

Question 2.   **[25 marks in total]**

a. Describe a situation in which triggers can be used to support the correct functioning of your database and the consistency of your data.

[4 marks]

b. Consider the following tables. The table MARKS stores the marks each student got in each exam (identified by the CourseID). The table STUDENT_AVERAGE stores, for each student, the number of exams the student attempted and the average mark of the student in all her/his exams. Some sample data are displayed on the next page.
The STUDENT_AVERAGE table is pre-populated at the start of each academic year with the ids of all the students registered. The field *number of exams* and *average* are set to zero.
The table STUDENT_AVERAGE is updated by a trigger called EXAM_AVERAGE defined on the table MARKS. The trigger automatically updates the student's average and the number of exams. Exams results can also be deleted from the table MARKS.
You are required to provide an implementation of the trigger EXAM_AVERAGE.

MARKS		
StudentID	CourseID	Exam Mark
1	1	56
1	2	64
1	3	60
2	1	75
2	3	61

STUDENT_AVERAGE		
StudentID	Number of Exams	Average
1	3	60
2	2	68

[15 marks]

c. Describe a situation where a graph database could be a better choice than a relational database. Justify your answer.

[3 marks]

d. Describe a situation where a document-based database like MongoDB could be a better choice than a relational database. Justify your answer.

[3 marks]

Question 3.   **[25 marks in total]**

a. Explain the CAP theorem for distributed databases, providing examples of one system that drops Consistency, a system that drops Availability and a system that drops Partition-tolerance

[9 marks]

b.
Describe the ACID properties of a relational database.   [3]
Describe the BASE properties of a NOSQL database.   [3]
Discuss the difference between the two approaches and for which application the BASE approach is more suitable   [3]
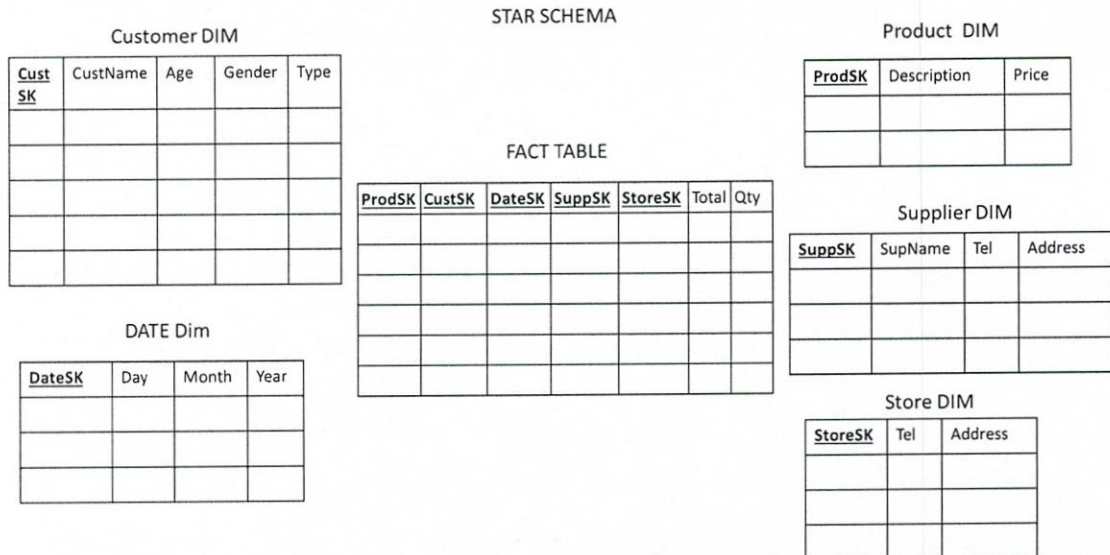
[9 marks in total, 3 for each question]

c. What are the advantages of a dimensional model over an ER diagram? What are the strengths of an **ER** Diagram?

[7 marks]

Question 4. [25 marks in total]

a. Consider the following star schema, storing information about the sales of a chain of stores. The dimensional model stores information about the quantity and the total price of a product sold to each customer in a specific store. For each product the supplier is also provided. The granularity of the DATE dimension is a single day. The field TOTAL in the FACT TABLE is expressed in euro.

STAR SCHEMA

Customer DIM

Cust SK	CustName	Age	Gender	Type

Product DIM

ProdSK	Description	Price

FACT TABLE

ProdSK	CustSK	DateSK	SuppSK	StoreSK	Total	Qty

Supplier DIM

SuppSK	SupName	Tel	Address

DATE Dim

DateSK	Day	Month	Year

Store DIM

StoreSK	Tel	Address

(i) Provide an SQL query to display the names of the top 10 best-selling products by store in the month of February 2012.

[5 marks]

(ii) You are asked to re-design the above dimensional schema, since there is no need to store information about individual customers and only monthly reports are required. Provide a new dimensional model to satisfy these new specifications.

[5 marks]

(iii) Is the new dimensional model bigger or smaller than the previous? Justify your answer

[2 marks]

(iv) After you have re-desinged the dimensional model, what is the meaning of the numbers in the field Total and Qty. of the new FACT TABLE ?

[2 marks]

b. Consider the star schema of question a).

(i) How would you store the same information in a document-based NoSQL database (such as MongoDB)?

[6 marks]

(ii) Discuss the advantages and limitations of the two implementations?

[5 marks]

Question 5.   **[25 marks in total]**

a. The School of Computer Science needs to store information about the modules offered, the modules that are pre-requisites for each module and the timetable of each module. For each module, the timetable contains the time of each lecture and the lecturer room. Sample data is given in the table below:

M_ID	ModuleName	Pre-requisites	Day	Time	RoomID	Campus
1	Databases	None	Wed	10-12	B101	City
1	Databases	None	Mon	12-14	B105	City
2	Adv. DB	1	Fri	14-17	K113	Grangegorman
3	Python	None	Tue	14-16	K114	Grangegorman
3	Python	None	Fri	9-11	B105	City
4	Data Mining	2,3	Thu	10-12	B105	City
4	Data Mining	2,3	Tue	16-18	B105	City
4	Data Mining	2,3	Thu	14-16	B101	City

The field M_ID is unique and it represents the module ID. The room is also unique across multiple campuses. You are required to:

(v) Show how the above information can be stored in a normalized relational database, by providing the list of tables and for each table the list of fields, the table primary key and foreign key(s).

[9 marks]

(iii) Show how the above information can be stored in a GRAPH database, showing nodes, relations and their attributes.

[9 marks]

c. Describe the concept of transactional, periodic and accumulating fact table and discuss the differences between the three approaches in terms of storage, frequency of update and type of update needed by each approach.

[7 marks]