

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**

**Assignment Project Exam Help**

SEMESTER 1 EXAMINATIONS 2020/21

**<https://tutorcs.com>**

**Advanced Databases**

**WeChat: cstutorcs**

Dr. Pierpaolo Dondio

Dr. Deirdre Lillis

Ms. Pauline Martin – DT211C

Dr. Martin Crane – DT228/282

Duration: 3 Hours

Exam script available at 2:00pm on date of the exam.

All exams submissions should be uploaded before 5:00pm on the date of the exam.

Uploading open till 6:00pm

Answer question 1 and 2, and then choose two other questions

### SUBMISSION INSTRUCTIONS

*Submit a word or text file with your answers. Diagram can be drawn with any software, including saving Powerpoint slides as picture and include in the file. You can also draw your diagrams on papers and take and include in the Word document a picture of them.*

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

### Question 1 [15 marks]

Insert in a (2,3) b-tree the following index values:

50, 60, 70, 40, 30, 20, 10, 80, 90, 100

[15 marks]

### Question 2 [15 marks]

You are requested to explain the pros and cons of different RAID systems. You have three possible choices:

1. RAID 10
  2. RAID 5
  3. RAID 0
- a. Write a report where you describe the pros and cons of each technology, and for each choice you are requested to describe a use case. Discuss the different storage efficiency, writing and reading performance and fault-tolerance level of each RAID system.

- b. RAID 10 and RAID 0+1 are often presented as similar technologies. Could you find a difference between the two configurations in terms of fault tolerance? Provide an example.

<https://tutorcs.com>

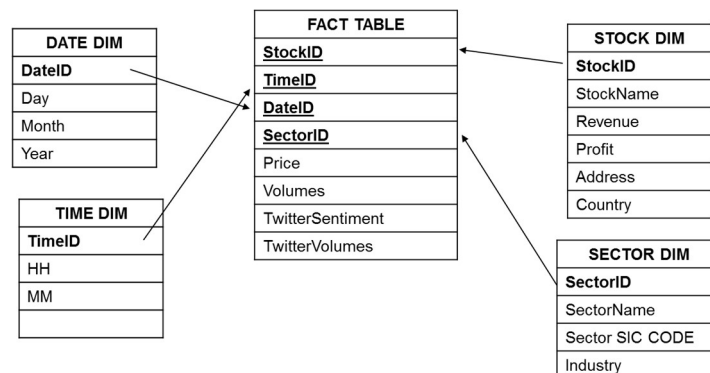
[3 marks]

**WeChat: cstutorcs**

### Question 3 – Dimensional Model

[35 marks]

- a. Why can the date (or time) dimension be pre-populated in a dimensional model?  
[5 marks]
- b. Consider the dimensional schema below:



*HH* represents hours, *MM* minutes, *price* is the stock price, *volumes* the stock volumes, *Twittersentiment* the sentiment of tweets, *TwitterVolumes* the number of tweets. The

Stock dimension stores the information about a specific stock and the Sector dimension the information about the sector of each stock.

i. What is the grain of the model?

[2 marks]

ii. Is it possible to analyse the sentiment of individual tweets with the above dimensional model? Justify your answer.

[3 marks]

c. A travel agency portal is selling holiday packages online. Its database is composed of the following tables (primary keys are underlined):

**CUSTOMER** (C\_ID, C\_name, C\_Address, Phone, Country\_Code [FK])

**HOLIDAY\_PACKAGES** (H\_Pck\_ID, tour\_operator\_id [FK], Description, from\_date, to\_date)

**HOLIDAY\_PCK\_HOTELS** (H\_Pck\_ID [FK], Hotel\_ID [FK], number\_of\_nights)

**HOTELS** (Hotel\_ID, Hotel\_name, H\_Address, H\_Category, Country\_Code [FK])

**COUNTRY** (Country\_Code, Country\_Name)

**BOOKING** (C\_ID [FK], H\_Pck\_ID [FK], number\_of\_persons)

**TOUR\_OPERATOR** (tour\_operator\_id, tour\_op\_name, tour\_op\_address, Country\_Code [FK])

**PRICES** (H\_Pck\_ID [FK], date, holiday\_package\_price)

Each holiday package is offered by a tour operator. A single package consists of a number of hotels and a number of nights in each hotel. A hotel can be offered in multiple holiday packages. The table HOTELS stores the information about each individual hotel.

When a customer books a holiday package she/he specifies the number of persons. The price of a holiday package changes – potentially every day (that is why the date is also part of the primary key of the table PRICES). All prices for each holiday package are for a single person.

Produce a star schema for the above ER diagram. The diagram should support the following queries and reports:

- (i) A weekly report showing the total revenue for each holiday package.
- (ii) A weekly report showing the distribution of customers by country
- (iii) Show the list of Tour Operators that sold more than “X” number of holiday packages in each quarter
- (iv) Show the average number of holiday packages sold for each tour operator in each country
- (v) Show the list of customers that did not buy any holiday package in each quarter
- (vi) Show the name of the bestselling holiday package each week
- (vii) Show the total number of nights booked in each hotel by each quarter.

Justify all your design choices. If a field in the fact or dimension table is not in the ER diagram, explain how to derive it and why.

[15 marks]

d. Using your dimensional model, how would you write the SQL query at 3c (iii)?  
[5 marks]

e. Using your dimensional model, how would you write the SQL query at 3c (vii)?  
[5 marks]

#### Question 4

[35 marks in total]

You are required to design a database for a recommendation system for movies. The database stores a list of users and a list of movies. Users can rate movies with a score from 1 to 5. Users can also add a review to each movie. Users can become friends of other users. The friendship relation is symmetrical. Each user is identified by a unique username, a password, gender and age. Users are not required to rate movies or write reviews, these are optional activities. Each movie is identified by a unique movie\_id, a title and a genre. For each review the text and the review date are stored. The database has to store the information about the rating each user had assigned to movies and the *friendship* relation between users.

You are required to:

a. Provide a relational schema to show how to implement the above database. Provide tables, fields and show primary and foreign keys  
[5 marks]

b. Write an SQL query to get the list of the users that are friends of username "John".  
[5 marks]

c. Write an SQL query to get the list of friends of "John" and the friends of friends of "John".  
[5 marks]

d. Provide a *json* structure(s) to store the same information provided in the relational model into *mongodb* collection(s).  
[5 marks]

e. Compare the two data models: which one is easier to query? Is standard SQL a sustainable way to perform such queries? What could be a better solution?  
[5 marks]

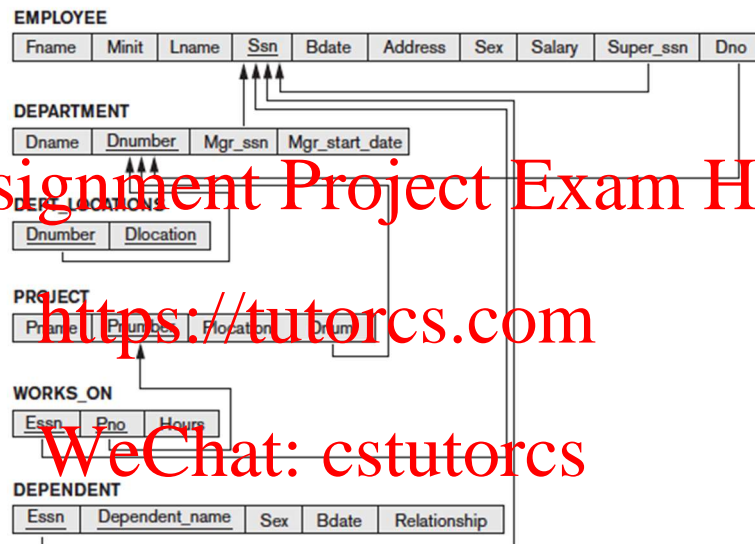
f. Show how the same information would be stored in a graph database  
[5 marks]

g. You need to choose among the three technologies used above (relational model, *mongodb* document-based model or graph database). Provide a use case where the best choice to implement the movie recommendation database would be a graph database, and a use case where the best choice would be a relational database. Justify your answer.  
[5 marks]

### Question 5

[35 marks in total]

The following ER diagrams refers to a database of a company. The database has a list of employees (table *Employees*) with primary key SSN. An employee is supervised by one other employees (field *super\_ssn*) The database has also a list of *Dependent* of each employees. For each employee's dependent, the name, sex, birthdate and type of relationship (such as son, daughter, wife, husband and so forth) is stored. There is a list of *Departments*, identified by a department number (*Dnumber*). Each department has a manager and each manager manages a set of employees. The table *Dept. locations* stores the location of each department. The database contains a list of projects (table *Project*). The table project is in a many-to-many relationship with the table *employee*. The bridge table *Works\_on* stores how many hours each employee spent on each project.



- a. Show how the same information can be stored in a document-based NOSQL database like MongoDB.

[10 marks]

- b. Show how the same information can be stored in a graph database. Draw a graph-like structure showing nodes, links and their attributes.

[10 marks]

- c. Suppose you want to implement the above database. You have been asked to recommend the best technology among relational databases, nosql databases and graph databases. Write a recommendation explaining the pro and cons of each choice and your suggestion. Justify your recommendations.

[15 marks]