

**INFORMATION TECHNOLOGY PROGRAMME**

**COMP504 Fundamentals of Computer Databases**

**Assignment 2, Semester 1 2019**

**Due:** 4 p.m. Friday 12th April 2019

**Submission**: Via Moodle. If unable to submit via Moodle, email to [saghira@ais.ac.nz](mailto:saghira@ais.ac.nz). Only electronic submissions will be marked. Hard-copy submissions WILL NOT BE MARKED. All submissions will be checked for plagiarism using Turnitin.

*Late assignments will be penalized 5% per day (or part of the day) for four days and will receive a mark of zero after four days.*

There is a total of 100 marks for this assignment. This is an individual assignment; group submissions are not allowed.

# This assignment is worth 15% of the total course marks.

## Assignment Case Study:

Universities and degree-level institutions required their staff to be research-active. The standard of research activity is publishing papers in peer-reviewed journals as well as publishing books with various publishers, and academic staff who publish often in these journals are highly regarded.

The Northshore University wants to construct a database to track journal publications as well as published books. The purpose is to keep the record of the research activities of their staff. This requires storing data about books and journals, the papers published in them, and the co-authors of each paper.

A journal is a regular publication that is produced by a publisher. Every journal in the world has a unique ISSN that identifies it. Journals also have names, but those names can change from time to time: for example, *Journal of Neuroinformatics* became *Journal of Neuroinformatics and Computations* from 1999. It is necessary to track the name of a journal over time, so that papers can be properly referenced. Journals also have values called *impact factors* that measure how influential the journal is. The impact factor is re-calculated every year, but previous impact factors must also be recorded. A journal publishes papers in one or more field of study, where each field is named (for example Computer Science, Ecology, Finance etc) and has an associated description.

Journals publish papers: A paper has a title, one or more authors, and a unique Digital Object Identifier (DOI). Papers are published once, in a particular issue and volume of one journal, and typically span several pages (usually expressed as e.g.: pp 325-342).

Similarly, a book is published with a publisher. A book has a title, one or more authors, edition number, edition year, and a unique ISBN number (of two different kinds).

The author(s) of a paper or book are researchers. A researcher has a given name, family name, and title (such as Doctor or Professor). A researcher’s name is not necessarily unique - there is more than one academic named James Sharp, for example. A researcher can be associated with several institutions (not just Northshore University), and each institution has a name (which is not necessarily unique) and is in a particular country. There can be several contact details recorded for each researcher. Researchers can also have multiple fields of study. A researcher can publish papers or books or both.

Your task is to design a data model for this scenario that supports all of the above operations and recording of data.

*The lecturer – Saghir Ahmad - will be the point of contact for any queries from students concerning any business rules or requirements that are not shown in the above scenario. If there is any ambiguity in what is required, ask the client’s representative! Remember, good software needs good design. Good design comes from asking the client good questions!*

## Requirements (Due 4.00 p.m. Friday 12th April 2019, Week 11)

For this part of the project you will be implementing and testing a database using MS SQL Server. This will involve translating a logical database design into SQL DDL, compiling the SQL into a database using MS SQL Server, and populating it with appropriate test data. You will have to write and show the results of some specific queries.

The ERD you must work from is available from Moodle as *UniDatabaseDesign.pdf*. And the second additional document *UniDataDictionary.docx* contains a list of attributes and attribute data types for each entity. A database will be created for you on the SQL server. This database will be named ResearchOutputs-XXXXXXXX, where XXXXXXXX is your student ID number.

You must perform the following tasks:

1. Write SQL code that implements the supplied ERD. This code must: **(40 marks)**
   1. Define the tables described in the ERD
   2. Define the columns specified for each table in the ERD
   3. Define the primary keys for each table
   4. Define the foreign keys for each table that has them
   5. Define the cascade update and cascade delete rules where appropriate
   6. Define data integrity constraints where appropriate. This means that primary and foreign keys must be enforced, that not null columns must be defined where appropriate, and that logical ranges for values must be checked.
2. Create suitable test data for your database. In this case, “suitable” means that each table must have at least one row in it and you must have sufficient data so that each query in part 3 below, that asks for a list of results, must have more than one row in the result. Use SQL INSERT statements to populate your database with this test data. You must also test any cascade deletes and cascade updates that you have defined.

**(10 marks)**

1. Create and execute queries that show the following: **(10 marks)**
   1. A list of the given names, family names, and titles of all researchers in one field of study of your choosing.
   2. A list of all paper titles and DOI where at least one of the authors are not researchers at Northshore University.
   3. A list of the paper titles published in one field of study of your choosing.
   4. The average impact factor of journals in which papers were published by Northshore University researchers.
2. Create a test data set that tests each and every data integrity constraint in the database. This means that you must create test data that the DBMS will reject as invalid data. Every constraint defined in step 1 above must be tested. You will need to define a separate INSERT statement for each constraint. **(20 marks)**
3. Define views that implement each of the queries specified in step 3 above. Demonstrate these views with queries. **(5 marks)**
4. Write a trigger that will generate a unique researcher ID when a new researcher is inserted. A researcher ID is generated as the family name of the researcher, and a counter that starts at one for the first researcher with that family name and increments by one for each following researcher. For example, the first researcher in table RESEARCHER would have the ID “WATTS01”, the second “WATTS02” and so on.

**(10 marks)**

1. Test the trigger. This means showing the contents of RESEARCHER before the insert, the INSERT command inserting another researcher with the same family name, and the contents of RESEARCHER after the trigger has fired. **(5 marks)**

**Submission Instructions:**

Your report must be a single Word document and must contain the following:

1. SQL code that implements the supplied ERD according to (1)
2. SQL code that inserts the test data into the database according to (2)
3. SQL code for each of the queries in (3) and the results of executing these queries
4. The constraint test data set in (4) and the results of each test command
5. SQL code for each of the views in (5) and the output of querying each one
6. SQL code for the trigger in (6)
7. The results of testing the trigger in (7)