



## Level 7: MSc in Data Science

# Advanced Databases Assignment

### Assignment Information

<b>Assessment Title :</b> Advanced Databases Assignment		
<b>Module Title :</b> Advanced Databases for MSc		
<b>Module CRN Code :</b> 33386, 34902	<b>Level:</b> 7	<b>Semester:</b> 2
<b>Programme Code(s):</b> G430 M0052	<b>Issue date<sup>1</sup>:</b>	week 5 (01/03/2021)
<b>Weighting :</b> 100 % of the total module mark	<b>Submission date<sup>2</sup>:</b>	11/05/2021 16:00
<b>Assessor(s):</b> Prof. Mo Saraee, Safwan Umer, Dr Charith Silva, Peter Tracey	<b>Return date<sup>3</sup>:</b>	The marks will be available in a maximum of 4 weeks after submission.

## **Learning Outcomes of this Assessment**

The learning outcomes covered by this assignment are:

- Provide a broad overview of the general field of 'database systems' and to develop specialised knowledge in areas that demonstrate the interaction and synergy between ongoing research and practical deployment of this field of study.

## **Key Skills to be Assessed**

This assignment assesses your skills in:

- Designing and implementing a database system
- Conveying your knowledge of database systems
- Use of T-SQL
- Knowledge on ethical, legal and data protection issues in the database systems

## **Recommended Reading**

The module notes are enough literature for completing this assignment successfully.

## **Equipment and Facilities to be Used**

- The assignment must be implemented using T-SQL, and any data should be stored in a Microsoft SQL server database. SQL statements should be stored in plain text files. Your submission must run on a computer with Microsoft SQL Server 2012, 2014, 2016 or 2017.

## **Workload**

For the successful completion of this assignment, a total of 75 hours should be budgeted.

## Task 1 (40 marks):

Imagine you are employed as a database developer for an international charity organisation (the client). The client is currently in the process of developing a reporting tool called **Child Well-Being Monitor** to analyse the child poverty data. A software developer will develop the functionality of the front-end reporting tool. Child Well-Being Monitor is a simple and low-cost web-based reporting solution which can be used to analyse child poverty. It allows users to extract and present data in charts, tables, and other visualisations so users can find useful information. It can also allow you to build paginated reports ideal for printing. The purpose of this tool is to translate data into actionable information to understand child poverty in lower-income countries.

The client is initially planning to import data from the Young Lives project (<https://www.younglives.org.uk/>) and build several reports to understand Child poverty in lower-income countries. The Young Lives survey is an innovative long-term project investigating the changing nature of childhood poverty in four developing countries. The purpose of the project is to improve understanding of the causes and consequences of childhood poverty and examine how policies affect children's well-being, in order to inform the development of future policy and to target child welfare interventions more effectively.

The study is being conducted in Ethiopia, India, Peru and Vietnam. These countries were selected because they reflect a range of cultural, geographical and social contexts and experience a variety of issues facing the developing world, high debt burden, emergence from conflict, and vulnerability to environmental conditions such as drought and flood.

The Young Lives study aims to track the lives of 12,000 children over 15 years, surveyed once every 3-4 years. Round 1 of Young Lives surveyed two groups of children in each country, at 1 year old and 5 years old. Round 2 returned to the same children who were then aged 5 and 12 years old. Round 3 surveyed the same children again at aged 7-8 years and 14-15 years, Round 4 surveyed them at 12 and 19 years old, and Round 5 surveyed them at 15 and 22 years old. It is sponsored by the Department for International Development and conducted by the Young Lives team based at the University of Oxford.

Link for the Metadata and Datasets:

<https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000060#!/access>

Students are required to register and apply for a password with the UK Data Service and sign a confidentiality agreement before they can access the data.

The client requires your knowledge and expertise to design, implement and test a database system for this reporting system. So, that software developer can build a front end of the reporting system.

You will provide a report detailing your suggested schema and advice. In addition to the database schema, the client also requires working T-SQL statements, with meaningful comments to achieve the following:

- Insert the full dataset from the Young Lives data repository to the local SQL Server database.
- Create all the tables using T-SQL.
- Create various summarised reports that allow you to measure the child poverty in Ethiopia, India, Peru and Vietnam (minimum 5 reports).
- Include filtering, sorting and grouping functionality in the reports.
- Include various search facilities in the Child Well-Being Monitor.
- Include concurrency techniques in the statements to ensure that the database remains in a consistent state when multiple users access it simultaneously.

### Extra features to be implemented

To get more than a “Satisfactory” mark, a number of extra features should be implemented. Features include, but are not limited to:

- Use of views
- Use of stored procedures
- Use of system functions and user-defined functions
- Consider data protection issues as part of the design and implementation of your solution.

## Task 2 (30 marks):

The client is planning to initiate a project helping children from low-income families in Vietnam. They require to analysis the inequality in the education sector in Vietnam. Therefore, they require to analysis and generate an overview report using Young Lives: School Survey, Vietnam, 2016-2017 dataset. A school survey was introduced into Young Lives Project in 2010, following the third round of the household survey, in order to capture detailed information about children's experiences of schooling and to improve our understanding of the relationships between learning outcomes, and children's home backgrounds, gender, work, schools, teachers and class and school peer-groups.

The survey consists of three main elements: a child questionnaire, a household questionnaire and a community questionnaire. It covers a range of topics such as household composition, livelihood and assets, household expenditure, child health and access to essential services, and education. This is supplemented with additional questions that cover caregiver perceptions, attitudes, and aspirations for their child and the family.

The client requires your knowledge and expertise to design, implement and test a simple reporting system for this project using SQL Server and Microsoft Excel. Use Microsoft Excel as a presentation layer. You will provide a report detailing your approach and proposed frond-end design in Microsoft excel. You will also provide your SQL statements with appropriate comments and a full backup of your database. In addition to the front-end design, the client also requires working T-SQL statements, with meaningful comments to achieve the following:

- Insert the full dataset from the School Survey, Vietnam, 2016-2017 dataset to local SQL Server database.
- Create all the tables and Views using T-SQL.
- Create various summarised reports in Excel, which project participants can use to measure and understand the inequality in the education sector in Vietnam (minimum 3 reports).
- Include filtering, sorting and grouping functionality in the reports.
- Include various charts in the reports.
- Consider data protection issues as part of the design and implementation of your solution.

### Task 3 (30 marks):

Imagine you are employed as a Research Assistance by the University of Salford for an academic research project. The objective of the project is to develop a reporting tool called **Crime Profiler** to assist in the development of computational criminology as a part of the advancement of the theory and method in criminological research. As a part of the project, you have been given a task to build a report containing Lower Layer Super Output Areas (LSOAs) wise crime report with local population data in **Greater Manchester** between Jan 2017 and Dec 2018. Data can be downloaded from below links.

Crimes data: <https://data.police.uk/data/>

LSOA population data :

<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates>

The principal investigator of the research project requires your knowledge and expertise to design and implement a simple reporting system using SQL Server and Microsoft Excel. Use Microsoft Excel as a presentation layer. You should provide a report detailing your approach and proposed front-end design. In addition to the front-end design, the principal investigator also requires working T-SQL statements, Excel files, QGIS Maps and Screen Shots (where appropriate) to achieve the following:

- Insert required Crimes data from the Police data repository to a local SQL Server database.
- Insert required Lower Layer Super Output Areas (LSOAs) data to a local SQL Server database.
- Create all the tables and Views using T-SQL.
- Create a summarised LSOAs wise crime report with local population data in **Greater Manchester** between Jan 2017 and Dec 2018.
- Include filtering, sorting and grouping functionality in the Excel report.

### Extra features to be implemented

To get more than a “Satisfactory” mark, two extra features should be implemented.

- Visualise the Vehicle crime in **Greater Manchester** with QGIS using MSSQL Connector. Use OpenStreetMap as OpenLayers plugin.
- Visualise the Anti-social behaviours crimes in **Salford** with QGIS using MSSQL Connector. Use Google Satellite map as OpenLayers plugin.

## Report

A 3000 word report that documents your database and reporting solutions.

Additional consultancy advice to each task will award marks above the “Satisfactory” grade. This could include but is not limited to:

- Advice on data integrity and concurrency
- Advice on database security
- Advice on database backup and restore
- When working with data, you are expected to be aware of the ethical and legal issues surrounding your data. Provide advice on those ethical and legal issues.
- Advice on introducing business intelligence techniques or data science to each task.

## Requirements / Marking Scheme

Requirement	Assessment Method	Weight (%)
Database schema design (Task1), Front-End Design (Task 2 and Task 3)	Report	25%
SQL statements that support the client requirements	Report	50%
Report	Report	25%

## Notes

- The assignment must be completed on your own.  
Any code snippets that are not directly written by you (e.g. used from a tutorial) must be referenced as such within your code. You must directly

comment the code to explain its source. Failure to reference code that is not yours will be treated as plagiarism.

- The assignment must be completed on time  
If you submit work late, it will be marked according to the University's late submission policy

## Unfair means

The University has strict policies on unfair means. It is your responsibility to ensure that you both understand these and adhere to them in the production of your assignment. Any submitted works with such content identifiable will be penalised in accordance with the University of Salford regulations

[http://www.governance.salford.ac.uk/page/academic\\_handbook](http://www.governance.salford.ac.uk/page/academic_handbook)

## Submission

Your submission should be a single **ZIP file** uploaded via Blackboard. The file should be named as:

<<Your surname>>\_<<Your name>>.zip — for example, **Smith\_John.zip**.

**All items in the zip file should also be prepended by your surname. (Ensure you replace “Smith” by your surname in the names below).**

The following items must be included in your submission:

1. A file named **Smith\_database.bak** containing a full SQL backup of your database.
2. A folder named **Smith\_statements** containing the plain text files of your SQL statements. The SQL statements must feature comments detailing their intended purpose.
3. A PDF document named **Smith\_report.pdf** containing your report (Include all 3 tasks in one report).

It is assumed that you will also address any social/legal and ethical issues surrounding the design and implementation of the project, such as copyright, references, and web law.



## Assessment Criteria

The following assessment criteria are provided as a guide to the criteria that you need to satisfy in order to get a grade within each of the following ranges.

### **Extremely poor (0-9)**

- Totally inadequate demonstration of required knowledge.
- Not able to apply the practical and analytical skills from their programmes.
- No appropriate design methodology.
- No demonstration of analysis evaluation or synthesis.
- No evidence of the ability to self-manage a significant piece of work and critical self-evaluation of the process.
- Little academic value; presentation is extremely poor; work has no structure or clarity; extremely poor use of language; no references; no attempt to provide evidence of sources used.

### **Very Poor (10-19)**

- Virtually no relevant knowledge demonstrated.
- Fails to adequately apply the practical and analytical skills from their programme.
- Very poor use of design methodology.
- No meaningful analysis or evaluation or synthesis.
- Unable to self-manage a significant piece of work and to identify appropriate issues for critical self-evaluation of the process for reflection.
- Academic arguments presented are inappropriate or very poorly linked; presentation is very poor; work has little discernible structure or clarity; very poor use of language; lack of ability to source adequate material; very poor referencing.

### **Poor (20-29)**

- Inconsistent or inaccurate knowledge.
- Limited and inappropriate and inaccurate application of the practical and analytical skills from their programme.
- Poor use of methodology.
- Descriptive, occasional attempts to analysis or evaluate material but lacks critical approach to evaluation or synthesis.
- Identifies issues for reflection but lacks evidence of reflective processes.
- Some but inconsistent ability to self-manage a significant piece of work or critical self-evaluation of the process.
- Confusion or weakness in academic argument; presentation is poor; work is disorganised and lacks clarity; poor use of language; poor use of reference material; inappropriate or out dated sources with numerous referencing errors.

### **Inadequate (30-39)**

- Limited evidence of knowledge.
- Inappropriate application of the practical and analytical skills from their programme.
- Unsatisfactory design methodology.
- Mainly descriptive evidence of analysis, inconsistent critical approach, little evaluation or synthesis.
- Follows processes of reflection but fails to demonstrate insight; lacks coherence in the self-management of a significant piece of work.

- Presentation is unsatisfactory; work is limited in terms of structure, coherence or clarity; limitations in academic style; unsatisfactory referencing with errors; limited ability to support content with relevant sources.

#### **Unsatisfactory (40-49)**

- Basic knowledge with occasional inaccuracies.
- Appropriate yet basic application of the practical and analytical skills from their programme.
- Superficial depth or limited breadth, but an overall adequate identification of design methodology.
- Critical analysis evident, with some evaluation and synthesis, although limited evidence of reflection.
- Some evidence of an ability to self-manage a significant piece of work and critical self-evaluation of the process.
- Some appropriate academic argument although not well applied and lacking in clarity; presentation of work is adequate in terms of structure, coherence, clarity and academic style; some inconsistencies; some grammar and syntax errors which detract from the content; narrow range of sources; referencing in presented work is adequate with some inconsistencies or inaccuracies; over utilises secondary sources; references used are inappropriate in terms of currency.

#### **Satisfactory (50-59)**

- Mostly accurate knowledge with satisfactory depth and breadth of knowledge.
- Solid application of the practical and analytical skills from their programme
- Fair use of design methodology.
- Sound critical analysis and evaluation or synthesis.
- Demonstrates basic ability of synthesise information in order to formulate appropriate questions and conclusions; reflective process is utilised, with insight demonstrating planning for future practice; shows the ability to self-manage a significant piece of work and critical self-evaluation of the process.
- Relevant academic argument: presentation of work is fair in terms of structure coherence, clarity and academic style; some inconsistencies in grammar and syntax; fair range of sources identified with appropriate referencing and few inaccuracies; appropriate use of primary and secondary sources.

#### **Good (60-69)**

- Consistently relevant accurate knowledge with good depth and breadth.
- Clear and relevant application of the practical and analytical skills from their programme.
- Good use of design methodology.
- Clear, in depth critical analysis, evaluation and academic argument with synthesis of different ideas and perspectives.
- Utilises reflection to develop self and practice; aware of the influence of varied perspectives and time frames; demonstrates an ability to self-manage a significant piece of work and critical self-evaluation of the process.
- Presentation of work is well organised with good use of language to express ideas or argument; very few inconsistencies in grammar and syntax good; good range of sources; well referenced with very few inaccuracies; good use of primary and secondary sources.

**Very Good (70-79)**

- Comprehensive knowledge demonstrating very good depth and breadth.
- Clear insight into links between the practical and analytical skills from their programme.
- Strong use of design methodology.
- Very good analysis and synthesis of material with evidence of critical and independent thought.
- Demonstrates ability to transfer knowledge between different contexts appropriately; balanced and mature approach to reflection used to enhance practice and performance; clear ability to self-manage a significant piece of work and critical self-evaluation of the process.
- Presentation is of a very good standard, demonstrating a scholarly style. Very good grammar and syntax. Clear evidence of referencing to a wide range of primary and secondary sources which are used effectively in supporting the work.

**Excellent (80-89)**

- Excellent depth of knowledge in a variety of contexts.
- Coherent and systematic application of the practical and analytical skills from their programme.
- Excellent use of design methodology.
- Excellent critical analysis and synthesis.
- Integrates the complexity of a range of knowledge and excellent understanding of its relevance; confident in their ability to self-manage a significant piece of work and critical self-evaluation of the process
- Arguments handled skillfully with imaginative interpretation of material; presentation is excellent, well-structured and logical; demonstrates a scholarly style; excellent grammar and syntax.

**Outstanding (90-100)**

- Outstanding knowledge.
- Exceptional application of the practical and analytical skills from their programme.
- Excellent professional execution of design methodology.
- Outstanding critical analysis and synthesis.
- Excels in self-managing a significant piece of work and critical self-evaluation of the process show an aptitude to formulate new questions, ideas or challenges.
- Incorporates evidence of original thinking; presentation is outstanding demonstrating a fluent academic style.

School of Computing, Science & Engineering -  
**Assessment Grid**

	90-100% Outstanding	80-89% Excellent	70-79% Very Good	60-69% Good	50-59% Satisfactory	40-49% Unsatisfactory	30-39% Inadequate	20-29% Poor	10-19% Very Poor	0-9% Extremely Poor
<b>Database schema &amp; Front-End Design (25%)</b>	Schema is developed in a flexible manner that caters for clients current and possible future requirements. Contains a relevant stored procedure, function and trigger.	Schema fulfils clients requirements. Contains a relevant: stored procedure, function and trigger.	Schema fulfils clients requirements. Contains at least two of either a relevant: stored procedure, function or trigger.	Schema fulfils clients requirements. Contains at least one relevant: stored procedure, function or trigger.	Schema fulfils all of the clients current requirements.	Schema fulfils three of the clients requirements.	Schema fulfils two of the clients requirements.	Schema fulfils one of the clients requirements.	Schema partially fulfils one of the clients requirements.	Schema does not fulfil the client requirements.
<b>SQL statements (50%)</b>	Statements fulfil all of the clients requirements and demonstrates correct use of multiple concurrency control techniques and SQL temporary objects . A stored procedure, trigger and function is used correctly.	Statements fulfil all of the clients requirements and demonstrate correct use of a concurrency control technique. A stored procedure, trigger and function is used correctly.	Statements fulfil all of the clients requirements and demonstrate correct use of a concurrency control technique. At least two of either a stored procedure, trigger or function is correctly used.	Statements fulfil all of the clients requirements and demonstrate correct use of a concurrency control technique. At least one of either a stored procedure, trigger or function is correctly used.	Statements fulfil all of the clients requirements and demonstrate correct use of a concurrency control technique.	Statements fulfil three of the clients requirements and demonstrate an attempt at the use of a concurrency control technique.	Statements fulfil two of the clients requirements.	Statements fulfil one of the clients requirements.	Statements partially fulfil one of the clients requirements.	Statements do not fulfil the clients requirements.

School of Computing, Science & Engineering -  
**Assessment Grid**

	90-100% Outstanding	80-89% Excellent	70-79% Very Good	60-69% Good	50-59% Satisfactory	40-49% Unsatisfactory	30-39% Inadequate	20-29% Poor	10-19% Very Poor	0-9% Extremely Poor
<b>Report (25%)</b>	Detailed and in-depth explanation of schema design with rational. Discussion with notable and quality references and examples. Including advice on <b>all extra</b> consultancy tasks.	Detailed and in-depth explanation of schema design with rational. Discussion with notable and quality references and examples. Including advice on at least <b>one extra</b> consultancy tasks.	Detailed and in-depth explanation of schema design with rational. Discussion with notable and quality references and examples.	Good explanation of schema design with rational. Discussion with notable and quality references and examples.	Basic explanation of schema design. Discussion with examples.	Explanation of schema design. Discussion with examples.	Minimal explanation of schema design. Discussion with examples.	Minimal explanation of schema design and limited discussion.	Minimal explanation of schema design.	Totally inadequate report.