Assignment 1
Posted Date: May 19, 2023
Submission Due: Jun 2, 2023 (11:59 pm)
Late assignments will not be accepted and will result in a 0 on the assignment

This assignment is created only for academic purposes and for the CSCI 5408 course. The objective is to learn - "how to design a good database". This assignment specification document should not be used to define any real facts or figures or for any commercial purpose.

Objective: This assignment covers two learning objectives (Lo).

- ***Lo#1:*** Solve two data management problems, by building two separate databases following a systematic approach. Understand the problems, build data models, and map the model to a physical design – To achieve this task, you need to perform conceptual design, logical design, and physical design in a sequence.
- ***Lo#2:*** Perform a comparative analysis of ERDs that are generated by two different approaches. ERD created manually from problem definition should be compared with ERD generated using reverse engineering in MySQL Workbench. Expectation is – writing a summary of your understanding and capturing differences of two ERDs in a tabular format.

Plagiarism Policy:

- This assignment is an individual task. Collaboration of any type amounts to a violation of the academic integrity policy and will be reported to the AIO.
- Content cannot be copied verbatim from any source(s). Please understand the concept and write in your own words. In addition, cite the actual source. Failing to do so will be considered as plagiarism and/or cheating.
- The Dalhousie Academic Integrity policy applies to all material submitted as part of this course. Please understand the policy, which is available at: https://www.dal.ca/dept/university_secretariat/academic-integrity.html

Assignment Rubric - based on the discussion board rubric (McKinney, 2018)

	Excellent (25%)	Proficient (15%)	Marginal (5%)	Unacceptable (0%)
Completeness including Citation	All required tasks are completed	Submission highlights tasks completion. However, missed some tasks in between, which created a disconnection	Some tasks are completed, which are disjoint in nature.	Incorrect and irrelevant
Correctness	All parts of the given tasks are correct	Most of the given tasks are correct However, some portions need minor modifications	Most of the given tasks are incorrect. The submission requires major modifications.	Incorrect and unacceptable
Novelty	The submission contains novel contribution in key segments, which is a clear indication of application knowledge	The submission lacks novel contributions. There are some evidences of novelty, however, it is not significant	The submission does not contain novel contributions. However, there is an evidence of some effort	There is no novelty
Clarity	The written or graphical materials, and developed applications provide a clear picture of the concept, and highlights the clarity	The written or graphical materials and developed applications do not show clear picture of the concept. There is room for improvement	The written or graphical materials, and developed applications fail to prove the clarity. Background knowledge is needed	Failed to prove the clarity. Need proper background knowledge to perform the tasks

Citation: McKinney, B. (2018). The impact of program-wide discussion board grading rubrics on students' and faculty satisfaction. Online Learning, 22(2), 289-299.

Explanation of the rubric: Suppose you received different grades in Clarity for the 2 problems
Problem #1: 25% in clarity
Problem #2: 15% in clarity
Then your overall grade for the clarity will be avg of (25+15) % = 20%

Problem 1: Data Modelling related problem. Objective is Lo#1

Follow all the steps in the given order.
For you, some steps could be optional and depends on the solution you create.

Hypothetical Problem Scenario: A new business "Hello12 Management team" is trying to establish a "Bed & Breakfast" in Halifax region. They need your help in building their backend information system. They are new in Nova Scotia, and does not have much knowledge about the surroundings or policies etc. They just gave you the following abstract requirements and ideas.

"

- Visit multiple bed and breakfast or hotel websites and understand the concepts
- Once you get the business domain knowledge start building a conceptual design for the team
- We can only tell you that for our Bed & Breakfast, we want to offer different types of breakfast items, there will be different types of rooms, and we do not have swimming pool or gymnasium. However, we can give coupons for those, so that boarders can access those facilities outside.

– Hello12 Management team"

Step 1. Identify at least 8 websites related to hotels or lodging facilities, which are useful for the given requirements, and create a table with a column containing websites URLs, and another column highlighting what information you gathered from those websites. **You do not have to run any automated crawler or web scrapping tool. This is just a manual process, so that you get the business domain knowledge**

Step 2. Identify at least 9 unique entities (sub-types are not counted as separate entities for this problem; Only the supertype will be considered as 1 entity), which must contain at least 1 weak entity. Write your assumption in a short paragraph explaining why you considered these entities, and how do you justify your strong entities as strong and weak entities as weak.

Step 3. Now using Chen model create an ERD or EERD. Your ERD/EERD should highlight all the possible attributes for the entities. In addition, highlight what types of attributes are these, e.g. primary key, partial key, multivalued attribute etc. **Note**: Your ERD should contain cardinality information, and it should be built using a standard data modelling tool, such as Erwin, Visio, draw.io etc.

Step 4. Identify if there are any design issues in your ERD. If you find any, then write a short paragraph about the issues, and fix it by generating a refined ERD/EERD
By the end of Step 4, you will be completing a conceptual model. Now, it is time to move to Logical mapping.

Step 5. Consider each entity from your data model as a table. Now, create a tabular structure with the attributes only (no need to consider any data points; just attributes are needed), and write a short paragraph and/or couple of bullet points to highlight the dependencies that exist in your logical model

Step 6. If required, perform some level of normalization (1NF->2NF->3NF), and write your assumptions on why you considered this normalization.

Step 7. Now, create a database named as "BedBreakfast", and create empty tables based on your final logical model by writing DDL SQL query in MySQL Workbench. This will give you the physical model.

Step 8. Export the SQL Dump of your MySQL Database and submit it as part of your assignment.

Problem 2: Data Modelling related problem. Objective is Lo#1 and Lo#2

Visit the website https://parks.novascotia.ca/ and any links that is mentioned in that site, which can give appropriate information on Nova Scotia Parks. The province is trying to build an information system to capture all the key information related to the parks that are operating in the province. Your initial task is to identify the key entities and the relationships, so that at next phase of the project, Nova Scotia can decide on how to create the database.

Hint: This website is used only to gather the information, you are not building another website, so website menu items do not necessarily represent entities or attributes. You are visiting the website to build the domain knowledge

Therefore, at this stage of the project, the province is expecting you to provide a correct and flexible data modelling, which is free from any of the design flaws (e.g., absence of capturing historical data, chasm trap, and fan-trap etc.)

Follow all the steps in the given order. For you, some steps could be optional and depends on the solution you create. Note: **This task does not require any web scrapping or automated information extraction from websites**

Step 1. Visit https://parks.novascotia.ca/ and sub-pages within dal.ca domain. **Note**: You do not have to consider other provinces or the entire country. The scope is only parks of Nova Scotia.
Step 2. Identify at least 8unique entities (sub-types are not counted as separate entities for this problem; Only the supertype will be considered as 1 entity). Write your assumption in a short paragraph explaining why you considered these entities
Step 3. Now using Crow's foot model create an ERD or EERD. Your ERD/EERD should highlight all the possible attributes for the entities. In addition, highlight what types of attributes are these, e.g. primary key, partial key, multivalued attribute etc. **Note**: Your ERD should contain cardinality information, and it should be built using a standard data modelling tool, such as Erwin, Visio, draw.io etc.
Step 4. Identify if there are any design issues in your ERD. If you find any, then write a short paragraph about the issues, and fix it by generating a refined ERD/EERD
By the end of Step 4, you will be completing a conceptual model. Now, it is time to move to Logical mapping.
Step 5. Consider each entity from your data model as a table. Since you have designed a crow's foot model, which has foreign key reference, and every entity is considered as a table, therefore, your design qualifies as a logical model. You do not have to perform any operation at this stage. Exception: If you perform any normalization, then you might need to decompose the structure, and you will need to provide the justification.
Step 6. Now, create a database, and empty tables by writing DDL SQL query in MySQL Workbench. This will give you the physical model.
Step 7. Once the physical model is built, export the SQL Dump of the database and table structure
Step 8. In MySQL Workbench create an ERD by performing reverse engineering. Write your observations on the similarities and differences you found between the ERD generated by you (manual), and ERD generated by MySQL Workbench (automatic).

Submission Guidelines:

1. All written reports, screenshots, images, SQL Dump, tables etc. must be added in a folder, and compress it with **.ZIP** format only. Use proper naming convention.
2. If not mentioned by TAs, then please rename the .zip file with your B00xxxxx_FnameLname_A1
3. Check the next point "**Suggestions**" for quality improvement and time management.

Suggestions:

Better Quality: To obtain good grades, you should follow the points given below:

- Try to understand the assignment requirement and follow all the steps required.
- Do not miss **adding citations**. If you write a single sentence taking the idea from somewhere else, then give credit to the author. Therefore, provide citation for any report you write, or any code you implement
- When you add citation, make sure to add it in a **standard format and uniform format**. E.g. if I refer 3 sources for writing a report, then I must cite the 3 sources in same format. One source in MLA, two sources in APA citation format will be a mismatch. Therefore, follow any one standard citation format
- Make sure to provide **inline citations** within report, and programming code
- Any image/picture/flowchart/diagram you add, make sure to provide a caption and a number for that image. It should be placed at the bottom of the image. E.g. "**Fig 1: Weekly time management chart for CSCI 5408**"
- Any table you add, must have a number and caption. This should be added on top of the table. E.g. "**Tab1: Table highlights the requirements in an ordered format**"
- To get points for novelty, you need to do something additional, which is not part of the requirement, e.g. writing an excellent report, adding flowcharts, more entities, checking all border line cases etc.

Time Management: Follow proper time management to reduce stress, and last-minute preparations. I am suggesting you follow the pie chart, which will require you to spend 5 hours in a week outside the classroom time for this course.

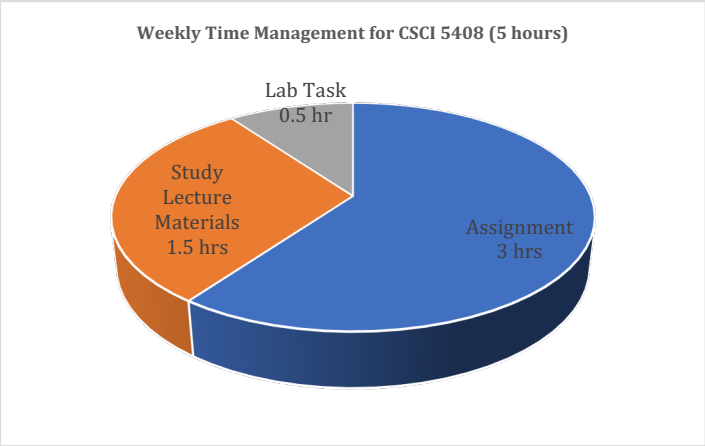

Fig 1: Weekly time management chart for CSCI 5408