

ISYS1055 Database Concepts

Assessment 4: Database Design Project



Assessment type: Take-home assessment

Word limit: N/A



Draft Due Date: 19 October at 23:59 (Melbourne Time) –Week 12 (otherwise 2 marks deduction)

Final Due Date: 2 November at 23:59 (Melbourne Time) – Week 14



Weighting: 30% (30 Marks)

Overview

This is a practical and real-world project that puts the knowledge you gained into practice. You are required to investigate and understand a **publicly available dataset**, design a **conceptual model** for storing the dataset in a **relational database**, apply **normalisation** techniques to improve the model, **build the database** according to your design and import the data into your database, and develop **SQL queries** in response to a set of requirements.

The objective of this assignment is to reinforce what you have learned in the whole course. Specifically, it involves how to build a simple application that connects to a database backend, running a simple relational schema.

Part A: Understanding the Data (0 Marks, Preliminary Work)

Part B: Designing the Database (12%)

- Task B.1 Produce an ER diagram for a relational database that will be able to store the given dataset.

Part C: Creating the Database and Importing Data (12%)

- Task C.1 Produce one SQL script file
- Task C.2 Create a database file and import the given dataset into your database.

Part D: Data Retrieval (6%)

- Task D.1-D.2 Produce one SQL query file that includes two SQL queries, produce a PDF file that includes the interpretation on the client's requirements and the running result screenshot of two queries.

Assessment criteria

This assessment will measure your ability to:

- Analyse the requirements outlined in the problem description
- Develop a conceptual model for the design of a database backend required for the system
- Use an industry-standard ER modelling tool to draw the ER model
- Use 7-step mapping process to create relational database schema
- Use normalisation process to evaluate the schema and make sure that all the relations are at least 3NF
- Create tables on SQLite Studio and populate them with data available from the specified sources.
- Write SQL statements required for CRUD (create, read, update and delete) operations on the database you built
- Develop your knowledge further to represent data in a meaningful way using data visualisation.

Course learning outcomes

This assessment is relevant to the following course learning outcomes:

CLO1	Describe the underlying theoretical basis of the relational database model and apply the theories into practice;
CLO2	Explain the main concepts for data modelling and characteristics of database systems.
CLO3	Develop a sound database design using conceptual modeling mechanisms such as entity-relationship diagrams.
CLO4	Develop a database based on a sound database design;
CLO5	Apply SQL as a programming language to define database schemas, update database contents, and to extract data from databases for specific users' information needs.
CLO6	Create then populate a normalised database based on a publicly available dataset, and visualise the results of queries against the dataset.

Assessment details

Part A: Understanding the Data

In this assignment, we are working with the publicly available dataset: **A Global Database of COVID-19 Vaccinations**. Further details about this dataset are available in the article available through the following URL: <https://www.nature.com/articles/s41562-021-01122-8>. The abstract of the article is as follows.

An effective rollout of vaccinations against COVID-19 offers the most promising prospect of bringing the pandemic to an end. We present the Our World in Data COVID-19 vaccination dataset, a global public dataset that **tracks the scale and rate of the vaccine rollout across the world**. This dataset is updated regularly and includes data on the **total number of vaccinations administered, first and second doses administered, daily vaccination rates** and **population-adjusted coverage for all countries** for which data are available (169 countries as of 7 April 2021). It will be maintained as the global vaccination campaign continues to progress. This resource aids policymakers and researchers in understanding the rate of current and potential vaccine rollout; the interactions with non-vaccination policy responses; the potential impact of vaccinations on pandemic outcomes such as transmission, morbidity and mortality; and global inequalities in vaccine access.

A live version of the vaccination dataset and documentation are available in a public GitHub repository at <https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations>. These data can be downloaded in CSV and JSON formats.

For the purposes of completing this assignment, we are only using the following files. **You are required to review and analyse the dataset available in these files**. You will find that reviewing the rest of the files, even if not listed below, will help you to form a better understanding about the big picture.

	FILE NAME	DESCRIPTION
1	locations.csv	Country names and the type of vaccines administered . Each line represents the last observation in a specific country. Refer to README.md for the details.
2	us_state_vaccinations.csv	History of observations for various locations in the US.
3	vaccinations-by-age-group.csv	History of observations for vaccinations of various age groups in each country.
4	vaccinations-by-manufacturer.csv	History of observations for various types of vaccines used in each country.
5	vaccinations.csv	Country-by-country data on global COVID-19 vaccinations. Each line represents an observation date. Refer to README.md for the details.
6	country_data/Australia.csv	Daily observations of vaccination in Australia.
7	country_data/India.csv	Daily observations of vaccination in India.
8	country_data/United States.csv	Daily observations of vaccination in US.
9	country_data/Argentina.csv	Daily observations of vaccination in Argentina.
10	country_data/Canada.csv	Daily observations of vaccination in Canada

Table 1: List of data files

To complete the tasks in the following sections, you are required to review and analyse the dataset that is available in the named files (For example: Identify the main entities and relationships present in the data across all files, Note the granularity of data in each file (e.g., country-day, US state-day, country-age group-day), or Observe potential data inconsistencies or areas requiring normalisation.).

Part B: Designing the Database (12%)

Task B.1 Produce an ER diagram for a relational database that will be able to store the given dataset.

It is important to note that the given CSV files are not necessarily representing a good design for a relational database. It is your task to design a database that will adhere to good design principles that were taught throughout the course. This means your database schema will not match the structure of the CSV files and, therefore, you will require to manipulate the structure of the dataset (and not the data itself) to import it into your database. Importing the data is required to complete Task C.2.

The ER diagram must be produced by [Lucidchart](#) similar to the exercises that were completed in the course. UML notation is expected and using other notations will not be acceptable. Including a high-quality image representing your model is important, which can be achieved using Export function of Lucidchart.

You are also required to transform the ER diagram into a database schema that will be used in the next part of the assignment.

Creating a good database design typically involves some database normalisation activities. You should document your normalisation activities and support them with good reasoning. This typically involves explaining what the initial design was, what the problem was, and what changes have been made to rectify the issue.

The expected outcome of completing this task is one PDF file named Model.pdf containing the following sections.

1. Database ER diagram and, if needed, a reasonable set of assumptions.
2. Explanation of normalisation challenges and the resulting changes.
3. Database schema.

You should present your ER diagram, then explain how you applied the 7-step mapping to get an initial schema. Then, document the normalisation steps taken from this initial schema to reach the final schema.

Part C: Creating the Database and Importing Data (12%)

Task C.1 Produce one SQL script file named Database.sql. This script file requires all the SQL statements necessary to create all the database relations and their corresponding integrity constraints as per your proposed design in Part B. The script file must run without any errors in SQLite Studio and contain necessary commenting to separate various relations (for example, one comment to explain each table). Note that this script is not supposed to store any data into the relations.

The expected outcome of completing this task is one script file with the specific name of Database.sql.

Task C.2 Create a database file named Vaccinations.db and import the given dataset into your database.

To complete this task, you may need to change the format of the CSV files to match the attributes of your designed database. You can use a spreadsheet editor such as Microsoft Excel.

The next step is to *import* the spreadsheets into the database you create in SQLite Studio. To complete this task, use the menu option *Tools – Import in SQLite*.

The expected outcome of completing this task is one database file named Vaccinations.db, which must contain all the data that is stored in the CSV files named in Table 1.

Part D: Data Retrieval and Interpretation (6%)

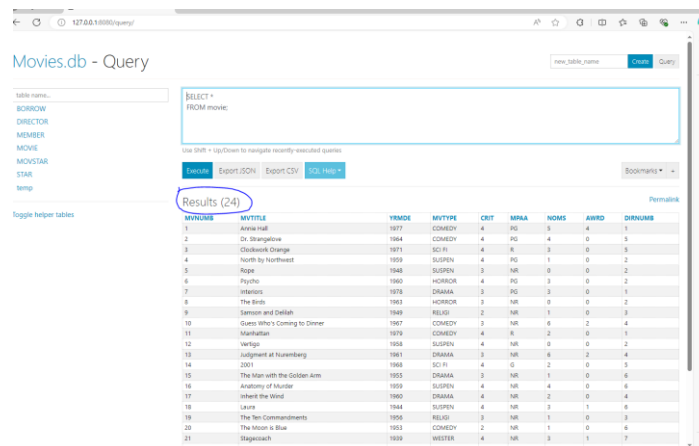
Now that you have created and populated a database, it is time to create some queries to investigate the data in various ways. In addition to writing the required queries, you are also asked to interpret the client's requirement for query generation.

Each query must consist of one SQL statement. It would be acceptable to use several nested queries, combine several SELECT statements with various operators etc. However, it would not be acceptable to have multiple and separated queries for each task (or to use views).

Before you write each query, you need to give interpretation on the client's requirement. After you have written each query, you are expected to inject your query into webpage, run the query on SQLite_Web.

The expected outcome of completing this task is as follows.

1. One SQL script file named Queries.sql containing *all* the queries developed for the tasks in this section. It is important that you add comment lines to separate the queries and indicate which task they belong to. Note that valid SQL comments must not generate errors in SQLite Studio. The marker of your work will use this file to execute and test your queries.
2. A PDF file named Queries.pdf containing the following elements for each task.
 - a. The interpretation on the client's requirements
 - b. The SQL query.
 - c. A snapshot of your query conducted on SQLite_Web. The snapshot must also show the total number of results retrieved by the query. A sample snapshot is provided below for your reference. (you may only show the top ranked rows in the snapshot)



ID	TITLE	YEAR	GENRE	RATING	DURATION
1	Annie Hall	1977	COMEDY	4	95
2	Dr. Strangelove	1964	COMEDY	4	95
3	Dr. Strangelove	1964	COMEDY	4	95
4	North by Northwest	1959	SUSPEN	4	95
5	Psycho	1960	HORROR	4	95
6	Psycho	1960	HORROR	4	95
7	Intolerance	1916	DRAMA	3	95
8	The Birds	1963	HORROR	3	95
9	Sansone and Delilah	1949	RELIG	2	95
10	Guero Who's Coming to Dinner	1967	COMEDY	3	95
11	Marathon	1976	COMEDY	4	95
12	Vertigo	1958	SUSPEN	4	95
13	Judgment at Nuremberg	1961	DRAMA	3	95
14	2001	1968	SCI FI	4	95
15	The Man with the Golden Arm	1955	DRAMA	3	95
16	Anatomy of Murder	1959	SUSPEN	4	95
17	Unholy the Wind	1960	DRAMA	4	95
18	Leaves	1964	SUSPEN	4	95
19	The Ten Commandments	1956	RELIG	3	95
20	The Moon is Blue	1953	COMEDY	2	95
21	Stagecoach	1939	WESTERN	4	95
22	Rear Window	1954	SUSPEN	4	95

Figure 1: Sample results snapshot with total rows

List of Tasks

Task D.1 (3 marks)

Penny: Good afternoon, I am Penny. May I help you, Sir?

Nick: Yes please. I understand that Australia government paid great effort to the vaccination during COVID-19 epidemic. Do you know the total number of its vaccines administered in each observation month recorded in the dataset from the first observation until the end of June 2021? I mean the sum of daily vaccinations within each month.

Penny: Just a second, I'll check it in the database

Year (OY)	Month (OM)	Monthly Vaccinations
-----------	------------	----------------------

Figure 2: Column Headers in the Result Set for Task D.1

Task D.2 (3 marks)

Nick: Could you please compare six countries including Australia, India, US, Argentina, Canada and Greece, in the speed of their monthly vaccine administration for all the observation dates? I would like to know the total number change trend of people *fully vaccinated* for each of these countries.

Penny: Just a second, I'll check it in the database

Year (OY)	Month (OM)	Australia	India	US	Argentina	Canada	Greece
-----------	------------	-----------	-------	----	-----------	--------	--------

Figure 3: Column Headers in the Result Set for Task D.2

Hint: For the first month in the dataset, there is no previous month (VOM1). You should display 0. Let the total number of people full vaccinated for a country in OM2 be VOM2, that in the previous month OM1 be VOM1. Then, the total number change trend of people *fully vaccinated* for this country in the current month (OM2) is computed by (VOM2-VOM1). If one country has no data for a month, please just omit the row.

Timeline:

Consider the following timeline to help manage your time to successfully complete this Project:

Timeline	Activities
Week 9 Part A	<ul style="list-style-type: none"> Review this assignment, the assessment criteria & rubrics. Familiarise yourself with the Harvard Referencing style. Go to the website provided in the instructions for Part A to familiarise yourself with the dataset you are working with. - "A Global Database of COVID-19 Vaccinations". Read the article for further details on this dataset. Watch video/attend live session on how to get started, develop the ER model, create the relations, format the CSV file using Microsoft Excel and import the data into the database.
Week 10-11 Part B	<ul style="list-style-type: none"> Work on Part B of the assessment. Get up to date with Lucidchart to produce the ER diagram. Create a PDF and create sections as mentioned in the assessment instructions for Part 2.
Week 12 Part C	<ul style="list-style-type: none"> Work on Part C of the assessment + submit draft (part A + B + partial C). Create relations, format CSV files according to database design and import data.
Week 13 Part D	<ul style="list-style-type: none"> Work on Part D of the assessment. Download a data visualisation tool that you will be using e.g. Excel, Google Charts, Tableau.
Week 14	<ul style="list-style-type: none"> Revise / fine-tune your assessment including the files required for submission.

Assessment Criteria

Your report will be assessed on the following criteria:

- This assessment will measure your ability to:
- Analyse the requirements outlined in the problem description
- Develop a conceptual model for the design of a database backend required for the system
- Use an industry-standard ER modelling tool to draw the ER model
- Use the 7-step mapping process to create relational database schema
- Use the normalisation process to evaluate the schema and make sure that all the relations are at least 3NF
- Create tables on SQLite Studio and populate them with data available from the specified sources
- Write SQL statements for CRUD (create, read, update, delete) operations on the database you built
- Develop your knowledge further to represent data in a meaningful way using data visualisation

[Learn how to use rubrics for assessment in Canvas.](#)

Submission Format

You are required to submit the files with the exact names as below.

1. Model.pdf
2. Database.sql
3. Vaccinations.db
4. Queries.sql
5. Queries.pdf

In the previous sections of the assignment, the expected content of each of the files is explained in detail.

Referencing guidelines

Use [RMIT Harvard](#) referencing style for this assessment.

You must acknowledge all the courses of information you have used in your assessments.

Refer to the [RMIT Easy Cite](#) referencing tools to see examples and tips on how to reference in the appropriated style. You can also refer to the library referencing page for more tools such as EndNote, referencing tutorials and referencing guides for printing.

Academic integrity and plagiarism

Academic integrity - When you submit work, it must be your own. Learn [how to avoid plagiarism and cheating](#)

Academic integrity is about honest presentation of your academic work. It means acknowledging the work of others while developing your own insights, knowledge, and ideas.

You should take extreme care that you have:

- Acknowledged words, data, diagrams, models, frameworks and/or ideas of others you have quoted (i.e., directly copied), summarised, paraphrased, discussed, or mentioned in your assessment through the appropriate referencing methods.
- Provided a reference list of the publication details so your reader can locate the source if necessary. This includes material taken from Internet sites.
- Please note that this assignment is an **INDIVIDUAL** assignment. Students sharing answers or using AI tools will be considered plagiarism and may incur serious consequences.

If you do not acknowledge the sources of your material, you may be accused of plagiarism because you have passed off the work and ideas of another person without appropriate referencing, as if they were your own.

RMIT University treats plagiarism as a very serious offence constituting misconduct.

Plagiarism covers a variety of inappropriate behaviours, including:

- Failure to properly document a source
- Copyright material from the internet or databases
- Collusion between students

For further information on our policies and procedures, please refer to the [University website](#).

Assessment declaration

When you submit work electronically, you agree to the [assessment declaration](#).

Study support

Visit the 'Setting up for Success' module in the course to get help with referencing, writing skills, study skills, finding information, group work, and more. You can also get connected to one-on-one support with an Academic Skills Advisor, a librarian, or a peer mentor.