CEE 412/CET 522 Transportation Data Management and Visualization

Project 1 | Winter 2020

Objectives

1. Become familiar with the database design concepts and procedure;
2. Learn how to apply the E/R diagram and the relational model for database design; and
3. Use the SQL Server to create and manage your database.

Description

Suppose you are hired to design a freeway rear-end accident database for a safety analysis project. The objective of the research project is to identify causal factors for two-vehicle rear-end accidents on freeways (I-5, I-90, I-405, and SR-520).

The occurrence of a rear-end accident typically involves factors of road, driver, and vehicle categories. Here are some example factors of our interest from each category (note that these are not required fields in your final database):

Road: Type of Road, Road Surface Condition (rain, snow, dry, etc), Road Surface Type (concrete, asphalt, etc), Speed Limit, Lighting Condition, Number of Lanes, Annual Averaged Daily Traffic (AADT), etc.

Vehicle: Type of Vehicle, Vehicle Condition, Number of People in the Vehicle, Movement, etc.

Driver: Age, Gender, Sobriety Condition, etc.

For a two-vehicle rear-end accident, there should be exactly one leading vehicle and one following vehicle involved. Several attributes, such as accident case number (should be unique), occurrence time, severity, etc, are needed to describe a rear-end accident.

To design the database, you need to determine how many entity sets are needed, how these entity sets are related, and which attributes should be included for each entity set. Please start your design with E/R diagrams. Then convert your E/R diagrams into relational schemas. Once your schema design is complete, you may find the provided data files (P1_Data.xlsx) useful to populate your tables. Considering the time limit for this project, keep the scope of your project under control. It is NOT necessary to include all of the provided data in your final database.

Please keep in mind that although attributes such as driver's license, driver's SSN, vehicle license plate number, etc., are important, you will never obtain such data due to privacy issues. Therefore, please replace them with public-accessible information in your design whenever possible. If the drivers entity set is weak without these attributes, you may want to combine it with other entity sets.

Database Requirements

The database created in this project should be suitable for managing data over an extended period of time, and in the future may include roadway segments, accidents, and vehicles that are not included in the files provided. The database should be normalized (minimum third normal form) with appropriate primary keys, foreign keys, and other constraints that you deem appropriate. In designing the database, you should consider the potential analysis functions you want to support, such as aggregating accidents by road segments and surface/lighting conditions, as well as retrieving individual records by a variety of conditions including but not limited to weather, AADT, and driver/vehicle characteristics. Note that not all of the columns in the data provided will be

of interest in your database. Note also that, because this database is for rear-end accidents, we are only interested in records describing accidents with 2 vehicles.

Data

Three data tables (Accident, Road, Vehicle) are provided in the data file (P1_Data.xlsx) for this project (available for download on Canvas), each table in a separate worksheet. All of them were from the true databases maintained by the Highway Safety Information System (HSIS) at the University of North Carolina. Here is a brief description about the three data tables:

- Accident: contains reported freeway accident data in Washington State in 2002.

- Road: contains road geometry, lighting, AADT, and speed limit information for each roadway section of Washington freeways.

- Vehicle: contains data of vehicles and their drivers involved in the reported accidents.

For details of the variables used in the provided files, please refer to the dictionary file (P1_Dictionary.pdf). It is very possible that some attributes in your design are not included in any of the data files.

Database Management System

We will use Microsoft SQL Server as the database management system for this course. Each of you should have an account in the Microsoft SQL Server. Your username and password to access the SQL Server are the same as those used for in-class exercises.

Products and Requirements

You need to deliver two products: the project report and the rear-end accident database.

Your project report must include:

1. Assumptions, if any, made in your design;

2. E/R diagrams. Please do not forget to identify entity keys with underlined texts;

3. Relational schemas converted from your E/R diagrams;

4. If any combinations or splits were made to the relations, please justify;

5. Make a simple user manual that explains attributes for each relation and applied constraints;

6. If some attributes in your database were not available from the given files, please specify how you handled the inputs for these attributes;

7. Your suggestions, if any, on rear-end accident data collection and management;

8. Names and locations of your rear-end accident database files;

9. Summary of your achievements in this project. Clearly specify whether your team has met all the requirements. If you failed to finish some of the requirements, please tell me the reason.

10. Please specify duties of each team member in this project.

How to submit:

- Your project report should be submitted electronically through Canvas.

- Your rear-end accident database tables should be placed in your team database at the SQL Server for this course, so that I can access and check them easily. Make sure you save the tables before the project deadline.

Hints

Design considerations:

- Simplicity and applicability are very important features of a good database design. Please take data availability into account in your E/R diagram design and do not include unrealistic attributes, such as a driver's SSN, name, etc.

- A well designed database should maximally eliminate data redundancy.

- Do not get bogged down by the volume of data that I have provided, you need only to consider what you need to represent in the database, use the provided data dictionary to find out what is available, and design a relational schema that represents the overlap of the two.

Two methods for inputting data to your relations:

- Manipulate data in an Excel worksheet and then convert the Excel worksheet into a SQL Server database table, or

- Convert the provided files into SQL Server database tables and then use SQL commands to project and select tuples for your designed tables (preferred).

Recommended steps:

1. Discuss the characteristics of two-vehicle rear-end accidents, the data available in the provided files, and summarize related factors and their relationships.
2. Identify entities and attributes of each entity.
3. Draw the E/R diagrams.
4. Check data availability and logical relationships of the entities.
5. Modify the E/R diagrams if necessary.
6. Convert the E/R diagrams into relational schemas.
7. Check data redundancy in each relation and modify your design if necessary.
8. Design an ER diagram representing your final relational schema
9. Populate designed relations (you can choose to use either of the two methods listed above).
10. Specify the primary key for each designed relation in your database.
11. Specify Foreign Key (FK) constraints.
12. Compose a simple user manual that explains attributes for each relation and applied constraints.
13. Test your database and complete the report for this project.