FIT5195 S1 2020 – Take Home Test (10%)
Due date: Week 6, Friday 1-May-2020, 11:55pm

This Take Home Test consists of **two** Case Studies. You are required to answer all questions.

Case Study 1: Government Hospital Data Warehouse (50 marks)

Currently, the Victorian government has a simple database system implemented to store the collective data of all government-run hospitals. The database system includes information about their services, staff, and patients. The simple database system consists of the following tables and attributes:
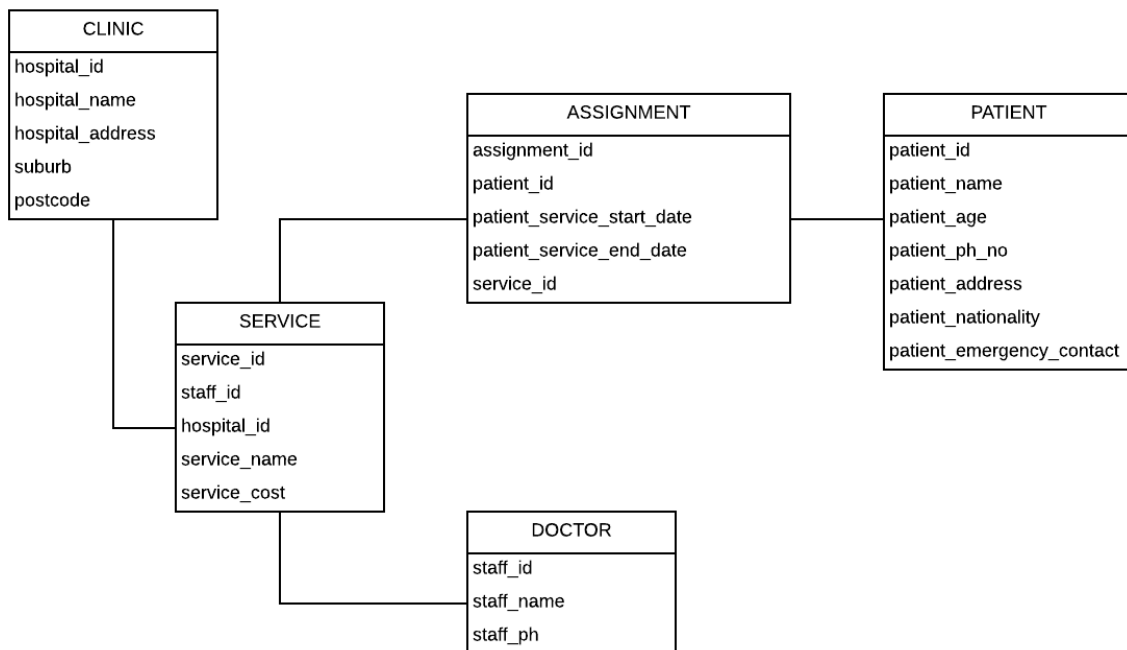

Figure 1 : A simplified structure of the Government Hospital's current E/R Diagram

Due to the nationwide impact of a new pandemic, the Premier of Victoria (i.e. the head of government in the Australian state of Victoria) requires a data warehouse for analyzing trends and forecasting purposes. The goal of the analyzing is to identify which frequently used services, popular hospital locations, etc.

You are required to design a small Data Warehouse to keep track of the statistics. The Premier is particularly interested in analyzing the *total number (population quantity) of patients and total service charged by services used, time periods, age groups, and hospital locations*.

The Premier also wants to be able to drill down based on the *cost of services range* (low price < $20, medium price between $20 and $50, and high price >$50).

Tasks:

[1]. Create the operational database based on the above E/R Diagram, and then populate with sufficient numbers of records in each table. It is suggested that each table would have 5-20

records depending on the use of that table in the operational database. You need to create a script file containing the CREATE TABLE and INSERT INTO commands.
You can follow the list of services from this link:
https://www.monash.edu/health/medical/services.

[2]. Then develop a GOVERNMENT HOSPITAL star schema. Identify the fact table, dimensions and attributes required to support the schema. If the star schema consists of a Bridge Table, you have to also include the Weight Factor and List Aggregate.
The result of this task is a star schema diagram. You can use any drawing tool, such as Lucidchart, to draw the star schema.

[3]. Validate your star schema using the Two-Column Table Methodology. You are required to illustrate some two-column tables for this task based on your star schema design.

[4]. Write the SQL commands to create the fact and dimension tables. You need to create a script file containing the appropriate SQL commands to create the fact and dimension tables.

[5]. Write the SQL commands to answer the following queries: (you need to make sure that there are records in your fact and dimensions tables. For each of the following queries, write the SQL and show the results):
a) Show the total number of patients making appointments during Winter.
b) Show the total service charged for each service cost type.
c) Show the total number of patients by each age group (infant <1, children <18, adult 18+, senior 65+) in April 2020.
d) Show the total service charged for general medical consultations in each suburb.

Submission checklist for Case Study 1:

 A. A **pdf file** or **word document** containing:
  i. The SQL commands to create tables and insert records to the operational database – ***Task 1***
  ii. The star schema diagram – ***Task 2***
  iii. The Two-Column Table Methodology illustration – ***Task 3***
  iv. The SQL commands to create the dimension and fact tables, as well as the contents of these tables – ***Task 4***
  v. The SQL commands to answer the queries in ***Task 5*** and the query results
 B. The **.sql files** containing:
  i. The SQL commands to create the operational database.
  ii. The SQL commands to create the data warehouse.
  iii. The SQL commands to answer the queries in *Task 5*.

  (Notes: The marker of this test will simply run this file. So you need to make sure that the SQL commands are written correctly in the .sql file. Pay a particular attention to the quote, especially if you cut and paste from a Word file).

 C. A **folder**:
  i. Save all the files from [A] and the SQL files from [B] in one folder.
  ii. Name the folder **CaseStudy1_yourStudentID**.

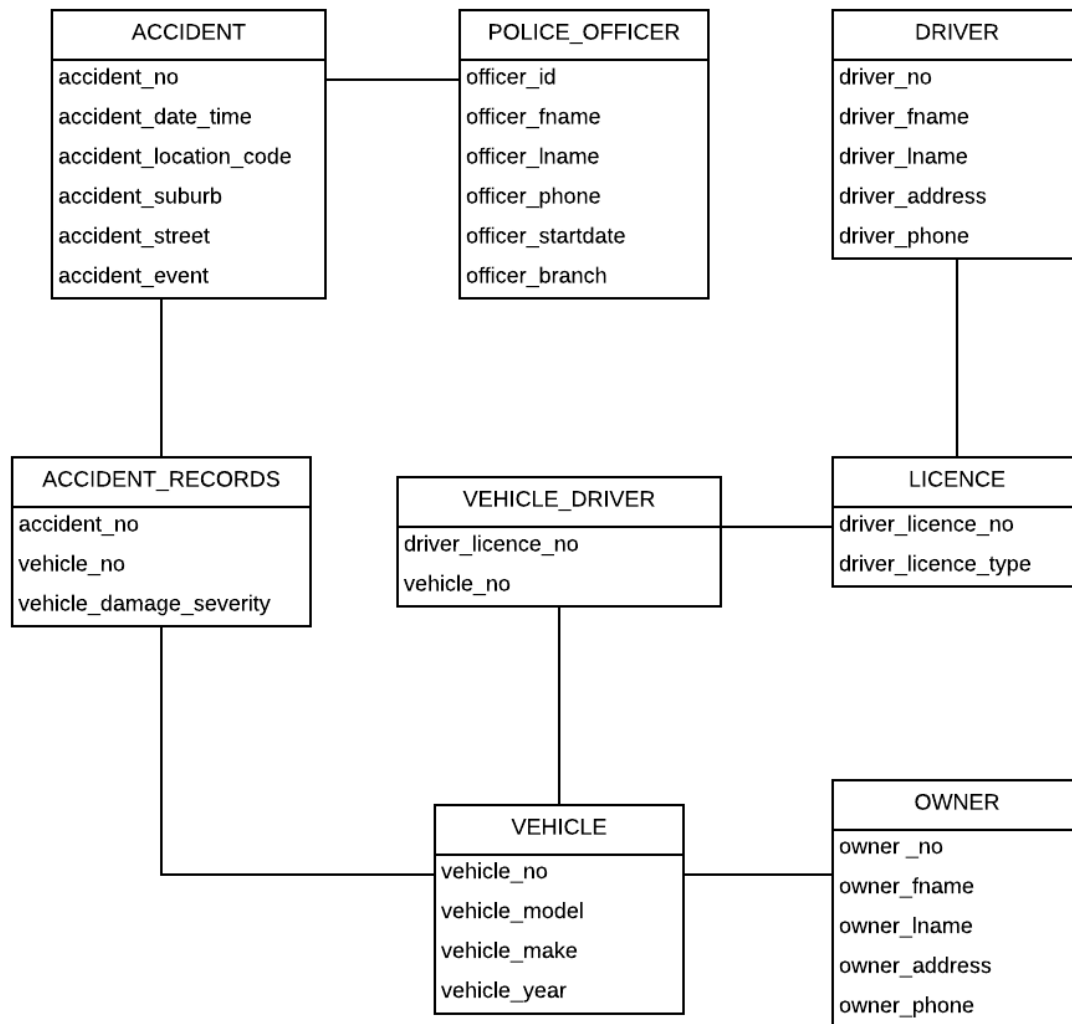# Case Study 2: Accident Records Data Warehouse (50 marks)

Figure 2 : A simplified structure of the Accident Record's current database system

The above figure displays how the Victorian Roads and Safety (VicR&S) Department stores accident related data. When there are any road accidents, the details about the accidents such as the location, date, time and event (e.g. what exactly happened during the accident) are recorded. For each accident, the information (e.g. name, contact number, employment start date, current employment branch) about one police officer who is in charge of overseeing the events of the accident (e.g finding out who/what caused the accident) are stored. In addition, to identify a particular accident, each accident is given a unique accident number.

On the other hand, one accident may be caused by several vehicles (e.g. a collision between 2 vehicles). All vehicles have owners who may or may not be the same as the vehicle driver during the accident. Therefore, the VicR&S database stores information about vehicles involved in the accident (i.e. vehicle number, vehicle model, vehicle make and vehicle manufacturer), the information about the owners of the vehicles (i.e. owner name, owner address, owner contact number) and the information about the drivers at the time of the accident (driver name, driver address, driver contact number, driver licence no, and when the driver started to drive the vehicle). Furthermore, for each accident, the severity of damage incurred by each vehicle is recorded.

Currently, to reduce the number of road accidents VicR&S department is interested in the following:
- The total number of accidents happening by different locations and by different lighting periods (daytime: 6AM - 5:59PM and nighttime 6PM - 5:59AM).
- The total number of accidents by each vehicle model.
- The number of vehicles involved in every accident event on different locations.
- The number of accidents taken care of different police officer branches.

Tasks:

[1]. Develop an ACCIDENT RECORDS star schema. Identify the fact table, dimensions and attributes required to support the schema. If the star schema consists of a Bridge Table, you have to also include the Weight Factor and List Aggregate.
The result of this task is a star schema diagram. You can use any drawing tool, such as Lucidchart, to draw the star schema.

[2]. Validate your star schema using the Two-Column Table Methodology. You are required to illustrate some two-column tables for this task based on your star schema design.

[3]. Write the SQL commands to create the fact and dimension tables. You need to create a script file containing the appropriate SQL commands to create the fact and dimension tables. The operational tables are accessible from the **`ACCIDENT`** account. The result of this task is the SQL commands. You will also need to show the contents of the tables that you have created.

[4]. Write the SQL commands to answer the following queries: (you need to make sure that there are records in your fact and dimensions tables. For each of the following queries, write the SQL and show the results):
a) Show the total number of accidents happening by different locations and by different lighting periods (daytime: 6AM - 5:59PM and nighttime 6PM - 5:59AM).
b) Show the total number of accidents by each vehicle model.
c) Show the number of vehicles involved in every accident event on different locations.
d) Show the number of accidents taken care of by different police officer branches.

[5]. You need to come up with additional two more questions and answer these questions using the SQL commands. Also explain the reason for why the management would like to have such information.

Submission checklist for Case Study 2:

A. A **pdf file** or **word document** containing:
   i. The star schema diagram – *Task 1*
   ii. The Two-Column Table Methodology illustration – *Task 2*
   iii. The SQL commands to create the dimension and fact tables, as well as the contents of these tables – *Task 3*
   iv. The SQL commands to answer the queries in *Task 4* and the query results
   v. The additional two questions with the SQL commands, query result, and your explanation – *Task 5*
B. The **.sql files** containing:
   i. The SQL commands to create the data warehouse.
   ii. The SQL commands to answer the queries in *Task 4 and Task 5*.

(Notes: The marker of this test will simply run this file. So you need to make sure that the SQL commands are written correctly in the .sql file. Pay a particular attention to the quote, especially if you cut and paste from a Word file).

C. A **folder**:
    i.   Save all the files from [A] and the SQL files from [B] in one folder.
    ii.   Name the folder **CaseStudy2_yourStudentID**.

Submission Method:

1. A **zip file**:
   a. Save the CaseStudy1_yourStudentID and CaseStudy2_yourStudentID folders into one folder named **FIT5195_Test_yourname_studentID**.
   b. ZIP the folder **FIT5195_Test_yourname_studentID.zip**. This must be a ZIP file and not other types of compressed folder. The zip file should contain the prescribed folder structure as follows:
      - FIT5195_Test_yourname_studentID/
        - CaseStudy1_yourStudentID/
          - Files for tasks
        - CaseStudy2_yourStudentID/
          - Files for tasks

2. Upload your zip file on Moodle by the due date: Friday 1-May-2020, 11:55pm (No late submission is accepted).

END OF TAKE HOME TEST