FIT5195 Major Assignment - Sem 1/2021 (Weight = 25%)
Due date: Week 9, Wednesday 5-May-2021, 11:55pm

A. General Information and Submission

- This is a group assignment. One group consists of 2 students from the same tutorial. You need to register your group composition through the form as soon as possible.
- *Submission method*: Submission is online through Moodle
- *Penalty for late submission*: 10% deduction for each day
- *Oracle account details*: You will need to supply with this assignment an Oracle username and password, used for this assignment.
- *Assignment Coversheet*: You will need to sign the assignment coversheet
- *Contribution Form*: The contribution needs to be completed by all members and please sign (e-signature is acceptable) the form as an agreement between members.
- *Assignment FAQ*: There is an Major Assignment FAQ page set up for the Major Assignment on EdStem Forum.

B. Problem Description

MonExplore is a not-for-profit health and education centre. MonExplore provides different services and events for students, staff, and community to support their health and wellness. MonExplore focuses on different topics or aspects of interests for students, staff, and communities, including Networking, Health & Lifestyle, and Spirituality. The centre has different programs in each topic, and each program will be organized several times a year depending on the demand and resources available. For example, under the Health & Lifestyle topic, one of the programs provided is 'Optimize your brain'. This program is run twice a year so there will be a March event and a September event for this program. Each event will run for eight sessions but the participants will only need to register for the event, not the individual session.

When people come to MonExplore, they will first fill in a survey to record their interests in different topics. They will then subscribe to the programs that they are interested in. When events are organised for the program that they are subscribed to, they will receive the information. When they register, they will need to provide how many people they want to register for as the system accepts group registrations. After registration, they will come to the event. However, not all people registered will come to the event, thereby, their attendance is recorded as well. During the event, the participants might donate to MonExplore since it is a not-for-profit centre and volunteer-led. MonExplore has volunteers who help to follow-up with the participants by providing support and answering questions, if required.

MonExplore currently has an existing operational database which maintains and stores all of the transaction information required for the management's daily operation. However, since the staff at MonExplore has limited database knowledge and the operational database is quite large, MonExplore has decided to hire your team of Data Warehouse Engineers to design, develop, and quickly generate reports from a Data Warehouse.

MonExplore's operational database tables can be found at MonExplore. You can, for example, execute the following query:

select * from MonExplore.<table_name>;

The data definition of each table in MonExplore is as follows:

Table Name	Attributes and Data Types		Notes
TOPIC	TOPIC_ID	NUMBER	This table stores the topic information.
	TOPIC_DESC	VARCHAR	
PROGRAM	PROGRAM_ID	NUMBER	This table stores the program information. One topic can have different programs but one program only focuses on one topic.
	PROGRAM_NAME	VARCHAR	
	DETAILS	VARCHAR	
	PROGRAM_FEE	NUMBER	
	PROGRAM_LENGTH	VARCHAR	
	FREQUENCY	VARCHAR	
	TOPIC_ID	NUMBER	
EVENT	EVENT_ID	NUMBER	This table stores the event information. One program can be offered several times, each time is considered as one event.
	START_DATE	DATE&TIME	
	END_DATE	DATE&TIME	
	EVENT_SIZE	NUMBER	
	LOCATION	VARCHAR	
	TOTAL_COST	NUMBER	
	PROGRAM_ID	NUMBER	
MEDIA_CHANNEL	MEDIA_ID	NUMBER	This table stores the media channel information. MonExplore can use
	DESCRIPTION	VARCHAR	
	COST	VARCHAR	

			different channels to promote their events.
EVENT_MARKETING	MEDIA_ID	NUMBER	This table stores the channel MonExplore used to promote an event and the cost for that particular event.
	EVENT_ID	NUMBER	
	COST	NUMBER	
PERSON	PERSON_ID	NUMBER	This table stores the information of the people coming to MonExplore, both volunteers and participants.
	NAME	VARCHAR	
	AGE	NUMBER	
	CONTACT_NO	NUMBER	
	EMAIL	VARCHAR	
	ADDRESS_ID	NUMBER	
	JOB	VARCHAR	
	MARITAL_STATUS	VARCHAR	
	GENDER	CHAR	
ADDRESS	ADDRESS_ID	NUMBER	This table stores the information of the people's address.
	STREET_NO	VARCHAR	
	STREET_NAME	VARCHAR	
	SUBURB	VARCHAR	
	STATE	VARCHAR	
	POSTCODE	NUMBER	
VOLUNTEER	PERSON_ID	NUMBER	This table stores the information of the volunteers, brief information about their background, and their volunteer time.
	DESCRIPTION	VARCHAR	
	START_DATE	DATE	
	END_DATE	DATE	
PARTICIPANT	PERSON_ID	NUMBER	This table stores the information of the participant, when
	1ST_DATE	DATE	

	REASON	VARCHAR	they first came to MonExplore, and who/how they knew about the centre.
FOLLOW_UP	VOLUNTEER_ID	NUMBER	This table stores the information of the following up with participants of the volunteer. Volunteer_ID and Participant_ID are the same with Person_ID.
	PARTICIPANT_ID	NUMBER	
	NOTE	VARCHAR	
PERSON_INTEREST	PERSON_ID	NUMBER	This table stores the information of the topics that a person is interested in.
	TOPIC_ID	NUMBER	
SUBSCRIPTION	SUBSCRIPTION_ID	NUMBER	This table stores the information of the program that a person subscribed to.
	PROGRAM_ID	NUMBER	
	PERSON_ID	NUMBER	
	SUBSCRIBED_DATE	DATE	
ATTENDANCE	ATTENDANCE_ID	NUMBER	The table stores the information of attendance.
	PERSON_ID	NUMBER	
	EVENT_ID	NUMBER	
	ATTENDED_DATE	DATE	
	DONATION_AMOUNT	NUMBER	
	NUMBER_OF_PEOPLE_ATTENDED	NUMBER	
REGISTRATION	REGISTRATION_ID	NUMBER	This table stores the information of the registration.
	EVENT_ID	NUMBER	

	PERSON_ID	NUMBER	
	MEDIA_ID	NUMBER	
	REGISTERED_DATE	DATE	
	NUMBER_OF_PEOPLE_REGISTERED	NUMBER	

C. Tasks

The assignment is divided into **FOUR** main tasks:

1. **Design a data warehouse for the above MonExplore database.**
   You are required to create a data warehouse for the MonExplore database.
   The management is especially interested in the following fact measures:
   - Number of people interested
   - Number of people subscribed
   - Number of people registered
   - Number of people attended
   - Total donation

   The following show some possible dimension attributes that you should need in your data warehouse:
   - Month, year
   - Participants' location
   - Demographic information: Age group [Child: 0-16 years old; Young adults: 17-30 years old, Middle-aged adults: 31-45 years old, Old-aged adults: Over 45 years old]; Marital status; Occupation [Student, Staff, Community]
   - Program
   - Topic
   - Event size: small event <= 10 people, medium event between 11 and 30 people, and large event > 30 people
   - Media
   - Program length: short < less than three sessions, medium event between three to six sessions, and long event > six sessions

   For each attribute, you may apply your own design decisions on specifying a range or a group, but make sure to specify them in your submission.

- **Preparation stage.**
  Before you start designing the data warehouse, you have to ensure that you have explored the operational database and have done sufficient data cleaning. Once you have done the data cleaning process, you are required to explain what strategies you have taken to explore and clean the data.

  The outputs of this task are:

  a) The E/R diagram of the operational database,

  b) If you have done the data cleaning process, explain the strategies you used in this process (you need to show the SQL to explore the operational database, and SQL of the data cleaning, as well as the screenshot of data *before* and *after* data cleaning),

- **Designing the data warehouse by drawing star/snowflake schema.**

  The star schema for this data warehouse contains multi-facts. You need to identify the fact measures, dimensions, and attributes of the star/snowflake schema. The following queries might help you to identify the fact measures and dimensions:

  - How many people subscribed to MonExplore's programs in January, 2020?
  - How many students attended the Dinner with Doctor program?
  - What is the most popular state that the participants came from in 2018?
  - What are the top 3 topics that married people are interested in?
  - How much money was donated according to different event sizes?
  - How many people attended long programs in 2019?
  - What is the most popular media channel that was given by the participants when they registered?
  - Which program is most interesting to young adults?

  You should pay attention to the granularity of your fact tables. You are required to create **two versions** of star/snowflake based on different levels of aggregation.
  The two versions of the star/snowflake represent different levels of aggregation. Version-1 should be in the highest level of aggregation. Version-2 should be in level 0, which means no aggregation. To make it simple, you can assume that the highest aggregation for this assignment is Level-2.

Version Name	Level
Version-1	High aggregation (Level 2)
Version-2	No aggregation (Level 0)

  The star/snowflake schema of both versions you created might contain **Bridge Table** and **Temporal**. If needed, you can use different temporal data warehousing techniques for the temporal dimension and provide the reasons of your choice.

  The outputs of this task are:

c) Two versions of star/snowflake schema diagrams,

d) The reasons of the choice of SCD type for temporal dimension,

e) A short explanation of the difference among the two versions of star/snowflake schema.

2. **Implement the two versions star/snowflake schema using SQL.**
You are required to implement the star/snowflake schema for the two versions that you have drawn in Task 1. This implies that you need to create the different fact and dimension tables for two versions in SQL, and populate these tables accordingly.

When naming the fact tables and dimension tables, you are required to give the identical name for the two versions and end with the version number to differentiate them. For example, "MonExplore_fact_v1" for version-1 and "MonExplore_fact_v2" for version-2.

The output is a series of SQL statements to perform this task. You will also need to show that this task has been carried out successfully.

If your account is full, you will need to drop all of the tables that you have previously created during the tutorials.

The outputs of this task are:

a) SQL statements (e.g. create table, insert into, etc) to create the star/snowflake schema Version-1

b) SQL statements (e.g. create table, insert into, etc) to create the star/snowflake schema Version-2

c) Screenshots of the tables that you have created; this includes the contents of each table that you have created. If the table is very big, you can show only the first part of the data.

D. Submission Checklist

1. One **combined pdf file** containing all tasks mentioned above:
   - □ Cover page
   - □ A signed coversheet
   - □ Details of your ORACLE accounts
   - □ A contribution declaration form:

     Each student must state the parts of the assignment that he/she did. An example is as follows:

     Percentage of contribution:
     1. Name: Adam, ID: 210008, Contribution: 60%
     2. Name: Ben, ID: 230933, Contribution: 40%

List of parts that each student did:
1. Adam: list the parts that Adam did
2. Ben: list the parts that Ben did

- □ Task C.1 (outputs a, b, c, d, e)
- □ Task C.2 (outputs a, b, c)

2. **.sql files** for the following task:
   - □ Task C.2 Implement Star Schemas (SQL command as required by output a and b)

   All of the above SQL files must be runnable in Oracle.

3. Zip all the files above (pdf from #1 above, and SQL files from #2 above), and upload this zip file to Moodle. One member of your group can upload the submission, however, **please note that all members of the group must click the submit button and accept the submission statement** (failure to so will mean your assignment will not be submitted and will incur late penalties).

   You must ensure that you have all the files listed in this checklist before submitting your assignment to Moodle. Failure to submit a complete list of files will lead to mark penalties.

Submission Method:

1. ZIP the folder **MajorAssignment_YourGroupNo.zip**. This must be a ZIP file and **not** other types of compressed folder. The zip file should contain the prescribed files as listed in the Submission Checklist.
2. Upload your zip file on Moodle by the due date: **Wednesday, 5 May 2021, 11:55pm**.
3. Submission Cut-off time: **Wednesday, 12 May 2021, 11:55 pm** (Submission link will be unavailable after the cut-off date).

Extensions / Special Considerations

If due to some circumstances that you need to apply for extension for the assignment, you would need to apply **no later than two University working days after the due date** of this assignment. All applications for special consideration must be made and submitted online by filling the application form available on the Special Consideration website and providing the supporting documents as required.

Please do not assume that submission of a Special Consideration application guarantees that it will be granted – you must receive confirmation that it has been granted.

Late assignments submitted without an approved extension may be accepted (up to a maximum of **seven days**) with the approval of the lecturer, but will **be penalised at the rate of 10% per day (including weekends)**. Assignments submitted more than seven days after the due date will receive **a mark of zero** for that assignment and may **not receive feedback.**

Please Note:

1. Your inability to manage your time or computing resources will not be accepted as a valid excuse. (Several assignments falling due at the same time are a fact of university life.)
2. Hardware failures, whether of personal or University equipment, are not normally recognised as valid excuses. Failure to back-up assignment files is also not recognised as a valid excuse.

All the best for your Major Assignment!