FIT3003 Major Assignment - Sem 2 2023 (Weight: 25%)
Due date: Week 11, Wednesday 11 October 2023, 11:55PM

Learning Outcomes:

LO1. Design multi-dimensional databases and data warehouses.
LO2. Use fact and dimensional modelling.
LO3. Explain the roles of data warehousing architecture and the concepts of granularity in data warehousing.

A. General Information and Submission

- This is a group assignment. One group consists of two 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
- *Assignment Coversheet*: You will need to sign the assignment coversheet
- *Contribution form:* The contribution form needs to be completed by all members and signed (e-signature is acceptable) as an agreement between members.
- *Assignment FAQ*: There is a Major Assignment FAQ page setup on EdStem Forum.

B. Problem Description – MonEquip Data Warehouse

Monash Equipment Centre (MonEquip) is an Australian-based company that provides a wide range of equipment for civil infrastructure and commercial construction projects.
The company has two main business functions:

- selling equipment along with excellent after-sales service,
- providing equipment for hire while ensuring a seamless hiring experience.

MonEquip has different branches throughout the state of Victoria, in which the management team frequently generates reports to keep track of the business (e.g. calculating the revenues from hiring and selling equipment). The reports are then used for forecasting various trends and making predictions about the business.

MonEquip currently has an existing operational database which maintains and stores all of the information of their business transactions (e.g. product hiring, product sales, equipment, etc.) required for the management's daily operation. The operational database records the transaction **from April 2018 to December 2020**. However, since the staff at MonEquip has limited database knowledge and the operational database is quite large, the management team has decided to hire you as a Data Warehouse Engineer to design, develop, and quickly generate Business Intelligence reports from a new Data Warehouse.

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

Table Name	Attributes and Data Types		Notes
ADDRESS	ADDRESS_ID	NUMBER	This table contains customer address information
	STREET_NUMBER	NUMBER	
	STREET_NAME	VARCHAR2	
	SUBURB	VARCHAR2	
	STATE	VARCHAR2	
	POSTCODE	NUMBER	
CATEGORY	CATEGORY_ID	NUMBER	This table contains category information
	CATEGORY_DESCRIPTION	VARCHAR2	
CUSTOMER	CUSTOMER_ID	NUMBER	This table contains customer information
	CUSTOMER_TYPE_ID	NUMBER	
	NAME	VARCHAR2	
	GENDER	VARCHAR2	
	ADDRESS_ID	NUMBER	
	PHONE	VARCHAR2	
	EMAIL	VARCHAR2	
CUSTOMERTYPE	CUSTOMER_TYPE_ID	NUMBER	This table contains customer type information
	DESCRIPTION	VARCHAR2	
EQUIPMENT	EQUIPMENT_ID	NUMBER	This table contains equipment information
	EQUIPMENT_NAME	VARCHAR2	
	EQUIPMENT_PRICE	NUMBER	
	MANUFACTURE_YEAR	NUMBER	

	MANUFACTURER	VARCHAR2	
	CATETOGORY_ID	NUMBER	
HIRE	HIRE_ID	NUMBER	This table contains hire information
	START_DATE	DATE	
	END_DATE	DATE	
	EQUIPMENT_ID	NUMBER	
	QUANTITY	NUMBER	
	UNIT_HIRE_PRICE	NUMBER	
	TOTAL_HIRE_PRICE	NUMBER	
	CUSTOMER_ID	NUMBER	
	STAFF_ID	NUMBER	
SALES	SALES_ID	NUMBER	This table contains sales information
	SALES_DATE	DATE	
	EQUIPMENT_ID	NUMBER	
	QUANTITY	NUMBER	
	UNIT_SALES_PRICE	NUMBER	
	TOTAL_SALES_PRICE	NUMBER	
	CUSTOMER_ID	NUMBER	
	STAFF_ID	NUMBER	
STAFF	STAFF_ID	NUMBER	This table contains the information of staff
	FIRST_NAME	VARCHAR2	
	LAST_NAME	VARCHAR2	
	GENDER	VARCHAR2	
	PHONE	VARCHAR2	
	EMAIL	VARCHAR2	

	COMPANY_BRANCH	VARCHAR2	

For the daily operation, MonEquip normally purchases equipment from its suppliers. The company then resells the equipment to customers while also providing a hiring service. When an equipment is purchased, the price of the equipment is kept in the Equipment table as EquipmentPrice. Meanwhile, the hiring rate is calculated as follows:

- The unit hire price is the hiring rate per day.
- If the customer returns the equipment within the same day, they only need to pay for 50% of the unit hire price.
- Otherwise, Total hire price is calculated as *(End Date - Start Date)* * *UnitHirePrice* * *Quantity*.

As you are required to design a Data Warehouse for **MonEquip**, the Operational Database tables can be accessed through the MonEquip account. You can, for example, execute the following query:

select * from MonEquip.<table_name>;

C. Tasks

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

1. **Design a data warehouse for the above MonEquip database.**
   You are required to create a data warehouse for the MonEquip database.
   The management is especially interested in the following fact measures:
   - Average hire price
   - Average sales price
   - Total revenue for equipment hiring
   - Total revenue for sales
   - Number of equipment hired
   - Number of equipment sold

   The following shows some possible dimension attributes that you may need in your data warehouse:
   - Time: Month, Year
   - Season [Australian Season: Summer, Winter, Autumn, Spring]
   - Customer type
   - Category
   - Company branch

- Sales price scale: low sales <$5,000; medium sales between $5,000 and $10,000; high sales > $10,000

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 drawn an E/R diagram for the operational database.

  The output of this task is:

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

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

  The star schema for this data warehouse may contain multi-fact(s). 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:

  - What was the total sales revenue in January, 2020?
  - How many pieces of equipment were sold in Winter, 2018?
  - How many equipment were hired by business customers?
  - What was the total hire revenue in Clayton branch?
  - How many trailers were hired by individual customers in Summer?
  - What is the average sale revenue for Lighting equipment in 2019?
  - What is the average hire revenue for Vehicles by individual customers?
  - How much sales revenue was generated from a high sale in Summer?

  Please note that the questions provided are to help identify fact measures and dimensions only, they do not provide all the possible connections between dimensions and fact tables.
  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 Version-1.

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

The star/snowflake schema of both versions you created might contain **bridge table** and **temporal dimensions**. If you are using a bridge table, make sure to include weight factor and list aggregate attributes. you can use different temporal data warehousing techniques for the temporal dimensions, if there are any, you must provide the reasons for your choice(s).

The outputs of this task are:

b) Two versions of star/snowflake schema diagrams,

c) An explanation of the difference among SCD types 1, 2, 3, 4, and 6. Explain the reasons for the choice of SCD type(s) for any temporal dimensions in your star schema, if there are any,

d) An explanation of the difference among the two versions of star/snowflake schema.

Note: *The above explanation must be consistent with your star schema diagram and based on the assignment scenario. And, maximum 300 words for each explanation.*

2. Data cleaning stage.

Before you start implementing the data warehouse, you have to ensure that you 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.

If you have done the data cleaning process, explain the strategies you used in this process. The outputs of this task are:

a) SQL statements to explore the operational database, and SQL statements of the data cleaning,

b) Screenshot of data *before* and *after* data cleaning

3. Implement two versions of the 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. That is, you need to create the different fact and dimension tables for the 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 suffix the version number to differentiate them. For example, “MonEquip_Face_V1” for version-1 and “MonEquip_Fact_V2” for version-2. If the dimension is the same between the two versions, you do not need to create them twice.

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 implementation and 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 only show the first part of the data.
Note: *The SQL statements for both levels of star schema must be presented in the PDF file.*

D. Checkpoints

Checkpoint	Weight	Assessment	Due date
Checkpoint 1	1%	ER Diagram Star schema Version 1 Group contribution check	Week 8 (during lab)
Checkpoint 2	1%	Data cleaning Group contribution check	Week 9 (during lab)
Checkpoint 3	1%	Star schema Version 2 Group contribution check	Week 10 (during lab)

The Checkpoints will only be assessed during your allocated lab. Your group is required to complete the assessment for each checkpoint in order to obtain the allocated mark. There are associated mark penalties for not meeting the checkpoint assessment on time to a satisfactory state.

The member contribution will be checked by your allocated tutor regularly. Your contribution will be based on team member reviews if you have been absent from one of the checkpoint assessments, and the Major Assignment grade will be adjusted accordingly.

Note that the Final Report and Implementation are worth 22%.

E. Submission Checklist

1. One **combined pdf file** containing all tasks mentioned above:

- □ Cover page
- □ A signed coversheet
- □ A contribution declaration form:

  Each student must state the parts of the assignment that they completed. 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 completed:

  1. Adam: list the parts that Adam did

  2. Ben: list the parts that Ben did
- □ Task C.1 (outputs a, b, c, d)
- □ Task C.2 (outputs a, b)
- □ Task C.3 (outputs a, b, c)

2. **.sql files** from the following tasks:
   - □ Task C.2 Data cleaning (SQL command as required by output a)
   - □ Task C.3 Implement Star Schema (SQL command as required by output a and b)

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

3. Zip all the SQL files from #2 above and name the ZiIP folders as MA_SQL.zip.

Submission Method:

1. Upload the PDF file from Checklist #1 and the ZIP file from Checklist #3 to Moodle by the due date: **Wednesday, 11 October 2023, 11:55 pm**

- The submission of this assignment must be in the form of **a single PDF file AND a single ZIP file**. No other forms will be accepted.
- One member of your group can upload the submission. However, **please note that all group members must click the submit button and accept the submission statement** (failure to do 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.

2. Penalty for late submission: 10% deduction for each day, including weekends

3. Submission cut-off time: **Wednesday, 18 October 2023, 11:55 pm**. The submission link will be unavailable after this time.

Late Penalty:

Late assignments submitted without an approved extension will be **penalised at the rate of 10% per day (including weekends and public holidays)**. Assignments submitted more than seven days after the due date will receive **a mark of zero** for that assignment and may **not receive any feedback**.

Please note (late penalty and extension):

1. An inability to manage your time or computing resources will not be accepted as a valid excuse. (Several assignments being due at the same time are a fact of university life.)
2. Group issues, 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.

Use of Generative AI Tools

You must not use generative artificial intelligence (AI) to generate any materials or content in relation to this assessment task.

Getting help and support:

What can you get help for?

- Consultations with the Teaching Team Talk to the Teaching Team: https://lms.monash.edu/course/view.php?id=135574§ion=2
- English language skills Talk to English Connect: https://www.monash.edu/english-connect
- Study skills Talk to a learning skills advisor: https://www.monash.edu/library/skills/contacts
- Counselling Talk to a counsellor: https://www.monash.edu/health/counselling/appointments

Extensions:

If you are experiencing difficulties that you think will impact your ability to meet this deadline, you may apply for a short extension of two calendar days.

The extension application can be found on *Moodle > Assessments > Extensions and Special Consideration* for an Extension. You must apply for a short extension as soon as possible, but no later than 11:55 PM on the day of the assessment due on 11 October 2023.

Special Consideration:

Students should carefully read the Special Consideration website, especially the details about the formal documentation required.

All special consideration requests should be made using the Special Consideration Application.

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

Plagiarism and Collusion:

Monash University is committed to upholding standards and academic integrity and honesty. Please take the time to view these links.
Academic Integrity
Student Academic Integrity Policy
Test your knowledge, collusion (FIT No Collusion Module)

All the best for your Major Assignment!