

FIT5195 Assignment 1 - Summer Semester B/2021 (Weight: 30%)
Transformation stage (20%) - Monday, 25 January 2021, 11:55pm
Data analytic stage (10%) - Friday, 5 February 2021, 11:55pm

General Information and Submission

- This is an individual assignment.
- *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
- *Assignment FAQ*: There is an [Assignment Frequently Asked Questions page](#) set up for the Assignment 1 on EdStem Forum.

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 E/R diagram of MonEquip's existing operational database is shown below:

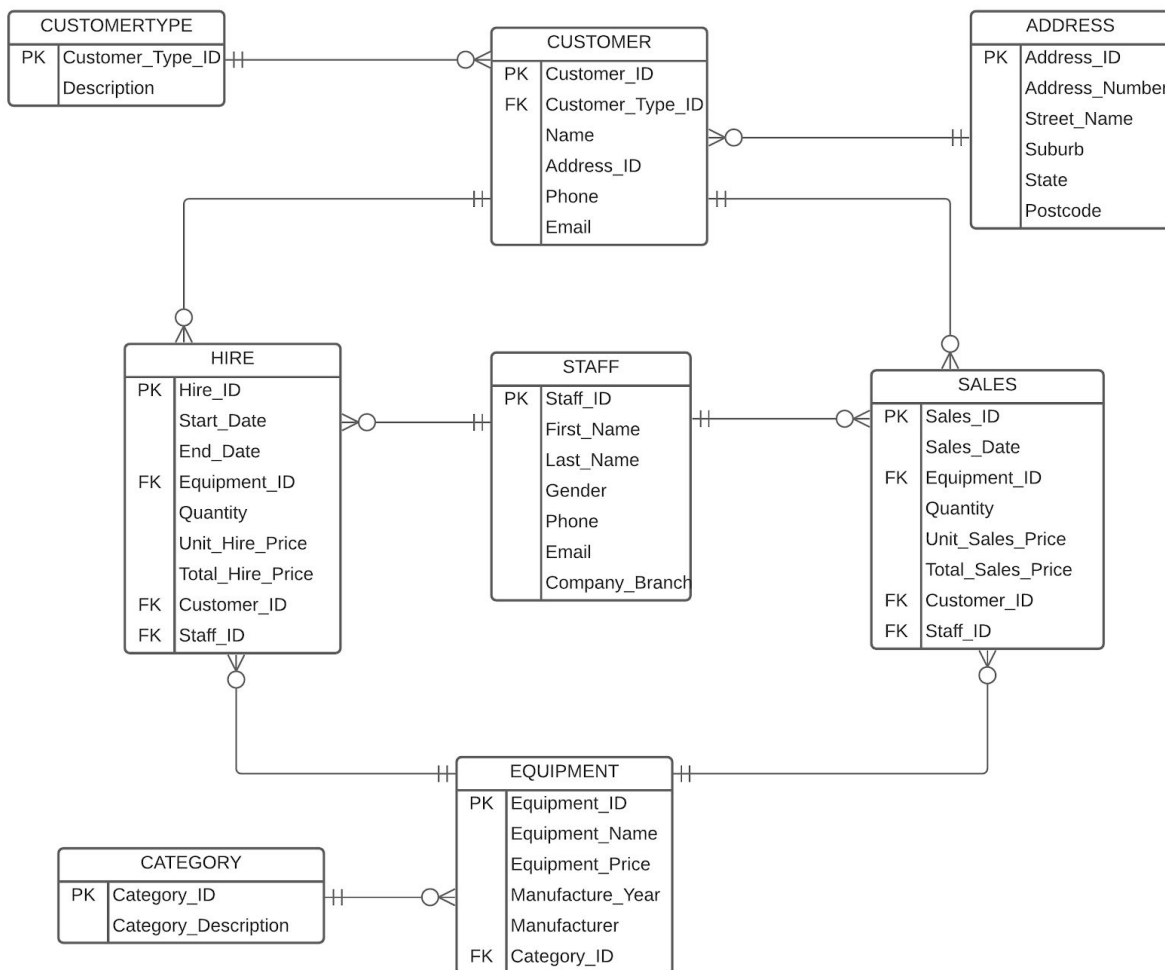


Figure 1: E/R diagram of MonEquip's existing operational database

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>;

A. Transformation Stage (20%) - due Monday, 25 January 2021, 11:55pm

The first stage of this assignment is divided into **TWO** 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 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

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

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?

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 outputs of this task are:

- c) Two versions of star/snowflake schema diagrams,
- d) A short explanation of the difference among the two versions of star/snowflake schema.

2. **Implement version 1 star/snowflake schema using SQL.**

You are required to implement the star/snowflake schema for version 1 that you have drawn in Task 1. This implies that you need to create the different fact and dimension tables for version 1 in SQL. 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) 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 only show the first part of the data.

B. Data Analytic Stage (10%) - due Friday, 5 February 2021, 11:55pm

The second stage of this assignment is divided into **TWO** main tasks:

3. Create the following reports using OLAP queries.

You are required to generate the reports using data warehouse **version-1 (Level 2)**, that you have implemented in Task 2. For each report, you ought to produce the SQL command and sample report output.

a. Simple reports:

Produce **two** reports. Each report contains two attributes from two different dimensions, and one fact measurement.

For the report itself, the first report must be about **Top *n*** and the second report is **Top *n*%**.

The outputs of this task are:

- (a) The query questions written in English,
- (b) Your explanation on why such a query is necessary or useful for the management,
- (c) The SQL commands, and
- (d) The screenshots of the query results (or part of the query results), including all attribute names.

b. Reports with proper sub-totals:

Produce **four** reports. These reports must include subtotals, using the Cube or Roll-up or Partial Cube/Roll-up operators

REPORT 3 and REPORT 4: What are the subtotals and total sales from each branch, customer type, and time period (month)? (You must use the Cube and Partial Cube operator)

REPORT 5 and REPORT 6: Produce two other subtotals reports that are useful for management using Roll-up and Partial Roll-up

The outputs of this task are:

- (a) The query questions written in English,
- (b) Your explanation on why such a query is necessary or useful for the management,
- (c) The SQL commands that include subtotals, using the Cube or Roll-up or Partial Cube/Roll-up operators, and
- (d) The screenshots of the query results (or part of the query results).

c. Reports with moving and cumulative aggregates:

Produce **two** reports containings moving and cumulative aggregates.

REPORT 7: What are the total revenue for hiring equipment and cumulative total hiring revenue for Site Equipment in each year?

REPORT 8: Produce one other moving/cumulative aggregate report that is useful for management.

The outputs of this task are:

- (a) The query questions written in English,
- (b) Your explanation on why such a query is necessary or useful for the management,
- (c) The SQL commands that contains moving and cumulative aggregates, and
- (d) The screenshots of the query results (or part of the query results).

d. Reports with Partitions:

Produce **one** report that contains partitions.

REPORT 9: Show ranking of each equipment category based on the monthly total number of equipment hired and the ranking of each branch based on the monthly total number of equipment hired.

The outputs of this task are:

- (a) The SQL commands that contain partitions, and
- (b) The screenshots of the query results (or part of the query results), including all attribute names.

4. Business Intelligence (BI) Reports.

Choose any **three** reports from Task 3, and change the presentation of these reports by representing these in a graph format. This new presentation should be more appealing to the management. You can use any graph software to show the graph reports. Additionally, in these new reports, you might want to include some selection buttons (for illustrative purposes), which may give users options on what criteria to choose, so that the graph report will be more dynamic.

Assignment Submission

There are two submission deadlines for this assignment.

Assignment	Weight	Assessment	Deadline
Transformation stage	20%	Task 1 and 2: Data cleaning, Star schema design and implementation	Monday, 25 January 2021, 11:55pm
Data analytic stage	10%	Task 3 and 4: OLAP and BI reports	Friday, 5 February 2021, 11:55pm

Submission Checklist

A. Transformation Stage (20%) - due Monday, 25 January 2021, 11:55pm

1. One **combined pdf file** containing all tasks mentioned above:
 - ☐ Cover page
 - ☐ A signed coversheet
 - ☐ Task 1 (outputs a, b, c, d)
 - ☐ Task 2 (outputs a, b)
2. **.sql files** from the following tasks:
 - ☐ Task 1 (SQL command as required by output *a*)
 - ☐ Task 2 Implement Star Schema (SQL command as required by output *a*)

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. 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 incur mark penalties.

B. Data Analytic Stage (10%) - due Friday, 5 February 2021, 11:55pm

1. One **combined pdf file** containing all tasks mentioned above:
 - ☐ Cover page
 - ☐ A signed coversheet
 - ☐ Task 3 Simple Reports (outputs a, b, c, d)
 - ☐ Task 3 More Reports with Subtotals (outputs a, b, c, d)
 - ☐ Task 3 Reports with Moving and Cumulative Aggregates (outputs a, b, c, d)
 - ☐ Task 3 Reports with Partitions (outputs a, b)
 - ☐ Task 4 (three graphs)
2. **.sql files** from the following tasks:
 - ☐ Task 3 Simple Reports (SQL command as required by output *c*)

- ☐ Task 3 More Reports with Subtotals (SQL command as required by output *c*)
- ☐ Task 3 Reports with Moving and Cumulative Aggregates (SQL command as required by output *c*)
- ☐ Task 3 Reports with Partitions (SQL command as required by output *a*)

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. 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 incur mark penalties.

THE END