

# (COMP1044) Database And Interfaces CW2

### **Group 5**

ID	Last Name	First Name
20300517	Lim	Theresa Jiyi
20399870	Teo	Zhen Yi
20307595	Mikhail Fong Abdullah	Johan Ariff
20414672	Gnasegaran	Hemanth Avinash
20307063	Choy	Eu Gene

### **TASK A: Summary**

### 1. Why did we choose the datatypes of each field?

We had to choose the datatypes for each table as some of the given datatypes do not give purpose or are not efficient to us. Therefore, we had to choose the datatype based on their functionalities. These are the datatypes and their functionalities in their respective fields.

### Users

- user\_id, username, firstname, lastname, password.
- user \_id (Type: int) is the user information's primary key in the table which
  allocates all its other identifications together. All user\_id are in the form of
  numbers only, thus int type was chosen.
- username (Type: varchar) is a user's chosen name for their user account. The
  username can contain both numbers and characters so type varchar is
  appropriate.
- **firstname** (Type: varchar) and **lastname** (Type: varchar) are the user's first and last name for identification. Names are made of letters so varchar is used.
- password (Type: varchar) is a personal secret word for when a user wants to log into their account. A password can be anything including numbers and letters.
   Varchar is more appropriate for this field.

### **Borrow**

- borrow\_id, member\_id, due\_date, date\_borrow.
- borrow\_id (Type: int) is the borrowing identification number which also acts as a
  primary key in the table which allocates the member\_id, due\_date and
  date\_borrow. borrow\_id only consists of numbers so type int can be used.
- **member\_id** (Type: int) is the borrower's library member id. It is different than user\_id and only consists of numbers. Type int is suitable for this field.
- **due\_date** (Type: date) is the date when the borrower must return the book. Type date is used because when the book is due is based on date.
- date\_borow (Type: datetime) is the date when the user started borrowing the book. Type datetime is used to allow admin to check on whose shift was this book borrowed. Type date would only record the date but there may have more than one person on shift that day.

### **Borrowdetails**

- borrow\_details\_id, book\_id, borrow\_id, borrow\_status, date\_return.
- **borrow\_details\_id** (Type: int) is the primary key for all the borrowing details in the table. It is in the form of numbers only, so type int is chosen.
- book\_id (Type: int) is a foreign key and the identification number for a specific book. All book\_id are in numerical form only so type int is used.

- borrow\_id (Type: int) is a borrowing identification number for all and each borrowing record. It is also a foreign key. All borrow\_id is in numerical form so int was used.
- **borrow\_status** (Type: varchar) is the status of the borrowed book. The status is either "Pending" or "Returned". "Pending" is when the book is not yet returned to the library. "Returned" is when the book is already returned to the library. Thus type varchar is suitable for this field.
- date\_return (Type: datetime) is the date when the book is returned. date\_return
  can only exist if the borrow\_status is "Returned". Type datetime is used because
  it allows admin to check who was on shift when this book was returned.

### Member

- member\_id, year\_level, status, type, contact, address, gender, firstname, lastname.
- member\_id (Type: int) is the identification number for a library member. It is also a primary key in the Member table. It consists of numbers only, so type int is used.
- year\_level (Type: varchar) is the education level of a member in the campus. It
  could be "First Year", "Second Year", "Third Year", "Fourth Year" or Null Null
  means that the member is not a student. It records a string so type varchar is
  used.
- **status** (Type: varchar) is to show if a member is an "Active" member or a "Banned" member. It records a string so type varchar is used.
- **type** (Type: varchar) is to show if the member is a "Student" or a "Teacher". It records a string so type varchar is used.
- **contact** (Type: int) is the telephone number of a member. A phone number consists of numbers only so type int is used.
- **address** (Type: varchar) is the home address of a member. An address has both number and characters so type varchar is more suitable.
- **gender** (Type: varchar) is the gender of a member. Either "Male" or "Female". It records a string so type varchar is used.
- **firstname** (Type: varchar) and **lastname** (Type: varchar) are the user's first and last name for identification. Names are made of letters so varchar is used.

### Book

- book\_title, book\_id, status, date\_added, copyright\_year, isbn, book\_pub, book\_copies, author, category.
- **book\_title** (Type: varchar) is the title for a book in the library. The title may have numbers or letters so varchar is suitable.
- book\_id (Type: int) is the identification number for a specific book. Multiple
  copies of the same book have samebook\_id. The book\_id consists of numbers
  only so type int is used.

- **status** (Type: varchar) is the condition of a book. It can be "New", "Old", "Damage", "Lost" or "Archive". It records a string so type varchar is used.
- date\_added (Type: datetime) is the date when the book title is added to the virtual library. Type datetime is used because it is more specific and allows admin to track who added the book easier.
- **copyright\_year** (Type: year) is the date when the book is published. Only the year is recorded so type year is enough for this feild.
- **isbn** (International Standard Book Number) (Type: varchar) is a special identification number for a book that came from the publisher. Type varchar is used because an isbn may contain the letter X and numbers.
- book\_pub (Type: varchar) is the publisher of a book. The name of a publisher may contain numbers and characters so type varchar is suitable.
- **book\_copies** (Type: int) is the number of copies of a book in the library. This is a numerical value so type int is suitable.
- **author** (Type: varchar) is the author name of a book. Names are made of letters so varchar is used.
- category (Type: varchar) is the genre of a book. It can be "Periodical", "English", "Math", "Science", "Encyclopedia", "Filipiniana", "Newspaper", "General" or "References". It records a string so type varchar is used.

# 2. The reason behind the design of the tables and connections between the tables including PK and FK?

### Users

- The Users table is for a user to log into their account on the library's website. Thus it is not connected to any other tables.
- **user\_id** is the PK of the table because it is a unique identifier for the records in this table.

### **Borrow**

- The Borrow table is for a user to input borrow\_id into the website so that they can
  find the details of the borrowed book. The table has some brief information about
  a borrowed book like due\_date and date\_borrow. It is connected to the member
  table to record which member has borrowed a book from the library before.
- **borrow id** is the PK of the table because it is unique.
- **member\_id** is an FK in the table. Borrow table needs to be connected to member table because only Active members can borrow a book.

### **Borrowdetails**

- Borrowdetails table stores all the borrowing details of a book and acts as an
  extension of the Borrow table. It is connected to book table and borrow table so
  that admin can see what book was borrowed by who as well as important
  information like whether the books has been returned yet.
- **borrow\_details\_id** is the PK of the table to identify each record.
- book\_id and borrow\_id are the FK of the table. book\_id is an FK because it
  can help show the details of the book from the Book table. borrow\_id is an FK
  because it links a book from the Borrow table.

### Member

- Member is a table that contains all the details of a library member. This also acts
  as an extension of the Borrow table because the user/member is only eligible to
  borrow a book only if their status in the Member table is "Active" instead of
  "Banned".
- **member\_id** is the PK of the table to uniquely identify each record.

### Book

- Book is a table that contains all the details of a book in the library. This also acts as an extension to the Borrowdetails table because it can provide the details of a borrowed book.
- **book\_id** is the PK of the table to uniquely identify each record..

### 3. What kinds of errors did we find in the dataset?

From the given dataset, we have found errors in the member, type, book, borrow and category tables. Some errors we found were missing information, information with wrong data type, as well as redundancies in design of the tables in the dataset.

In the 'member' table, we found some errors in the 'year\_level' column. This is due to the data 'Faculty' being irrelevant to the data type in said column.

In the table 'type', we found the table rather redundant as there are only two columns 'id' and 'borrowertype' with each unique data. For example, we could use either column as the primary key.

In the table 'book', we found two columns with similar data. This would be an inefficient use of storage space. For example, the columns, 'book publisher' and 'publisher name' have the same data. Furthermore, we found that in the column 'date\_received' the data is empty.

In the borrow table, one of the given data is illogical. The date\_borrow is 20th March 2014 but due\_date is 3rd January 2014.

In the table 'category', there are two columns 'category\_id' and 'classname' each with unique data. Thus, we could use either as a primary key and it is rather inefficient thus we find that the table is rather impractical.

### 4. How did we find the completeness of the dataset?

From the given dataset, some tables are missing information. For example, an empty column can be noticed from table 'book'.

The relationship between Foreign keys and Primary keys between tables must be correct and both column data must be matched. When we use SQL JOIN, there will be no error.

Next, we can use ETL testing method to test and verify the completeness of the dataset. For example, we can use SQL query to check the number of records in the source file. We can compare the count of number of records by writing the following queries, such as

SELECT count(\*) FROM book; SELECT count(\*) FROM borrow;

For the next query, this will compare the distinct values and the count of rows for each distinct value, such as

SELECT book\_id, count(\*) FROM book GROUP BY book\_id;

Next, we can check for duplicate data, it involves validating the primary key and the unique key in a column or in combination of columns that should be unique, such as

SELECT book\_id, book\_title, author, count(1) FROM book GROUP BY book\_id, book\_title HAVING count(1)>1;

### 5. How did we address those problems?

In the table 'member', we have decided that Faculty does not fit under the column 'year\_level' and have replaced it with default type NULL because all Teacher type hold the same data and thus does not have any effect even with the removal of data Faculty. The user can still find members who are Faculty by searching for Teacher under 'type' column instead.

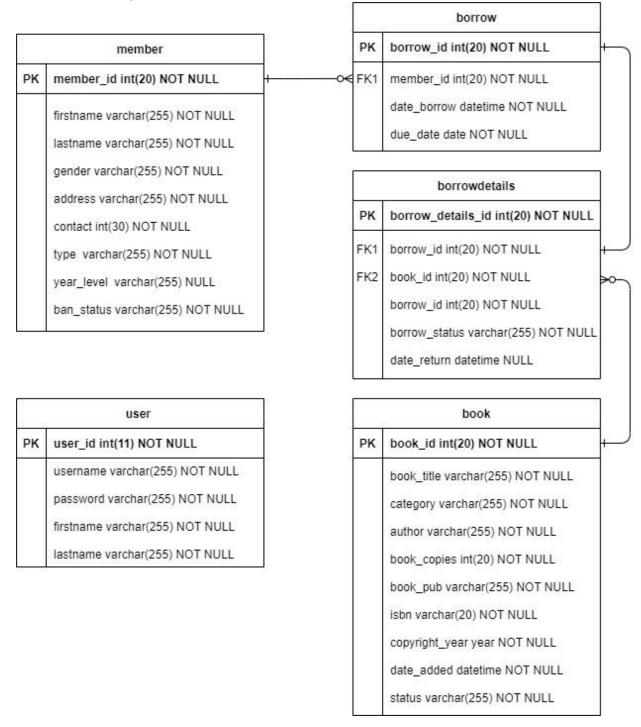
For the error of redundancy related to the 'type' table, we have decided to merge the column 'borrowertype' from 'type' table into the column 'type' in 'member' table. Therefore, we are able to eliminate the need for a separate table to store types. It would be easier for the user of the system to observe which type each member falls under when it is displayed in the table itself rather than having to look through another table.

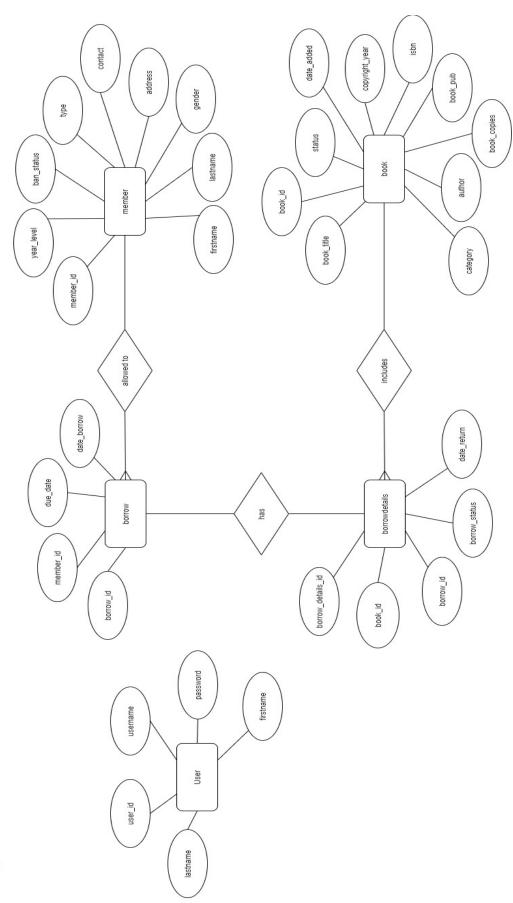
For the error related to similar data in 'book' table, we decided to transfer data that was missing in column 'book\_pub' from column 'publisher\_name'. For example, in row 14 and row 15, we copied the data from 'publisher\_name' to be added into 'book\_pub'. We then remove the column 'publisher\_name' because it conveys the same information as 'book\_pub'. We removed the column 'date\_received' since it does not contain any data and will not give an impact to other tables or columns.

Regarding the error in the date\_borrow and due\_date for table 'borrow', we edited the due date to 21st March 2014 which is the day after the date\_borrow in accordance to the pattern of other example data given.

We removed the need for a separate 'category' table by merging column 'classname' from 'category' table to column 'category\_id' in 'book' table. We are able to reduce the number of tables needed in this dataset as well as make it more convenient for the user to read the information.

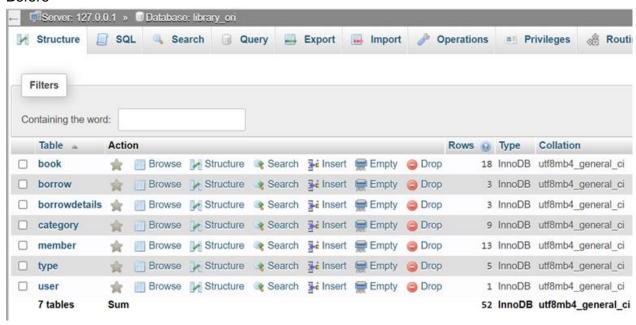
TASK B: ER Diagram 1



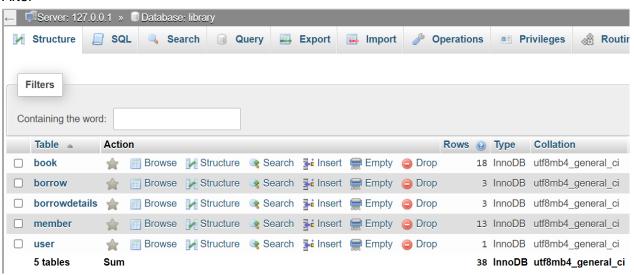


### **TASK B: Database Design**

### **Before**



### After



### **Book Table, Category Table ⇒ Book Table**

☐ 3 category

6 book\_pub

9 date\_added

☐ 7 isbn

☐ 10 status

☐ 5 book\_copies int(20)

8 copyright\_year year(4)

datetime

4 author

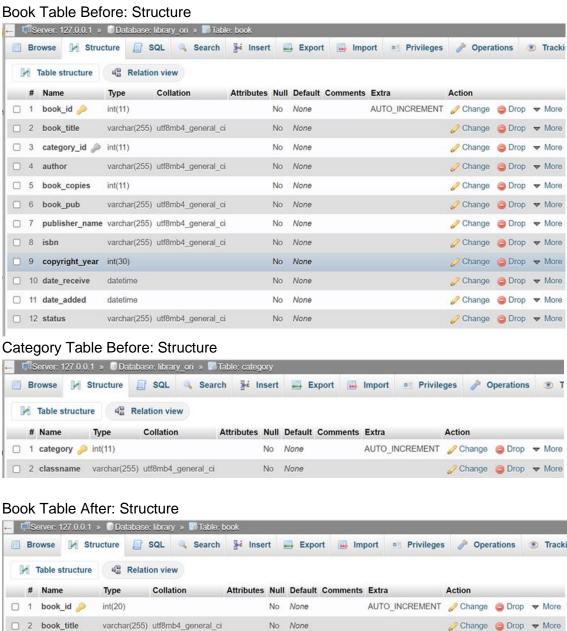
varchar(255) utf8mb4\_general\_ci

varchar(255) utf8mb4 general ci

varchar(255) utf8mb4\_general\_ci

varchar(20) utf8mb4 general ci

varchar(255) utf8mb4\_general\_ci



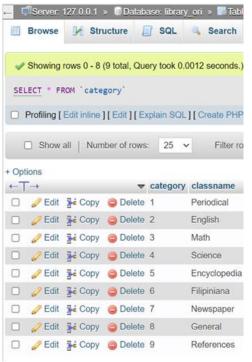
No None

⊘ Change 
⑤ Drop 
▼ More

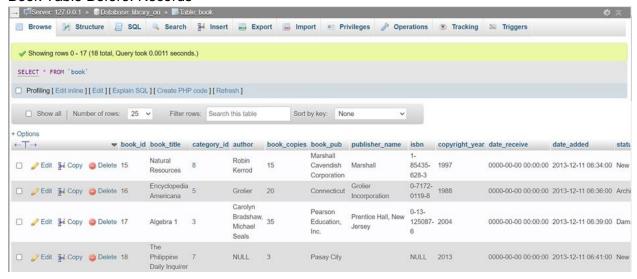
⊘ Change 
⑤ Drop 
▼ More ⊘ Change 
⑤ Drop 
▼ More

Ø Change 
 ⑤ Drop 
 ▼ More

### Category Table Before: Records

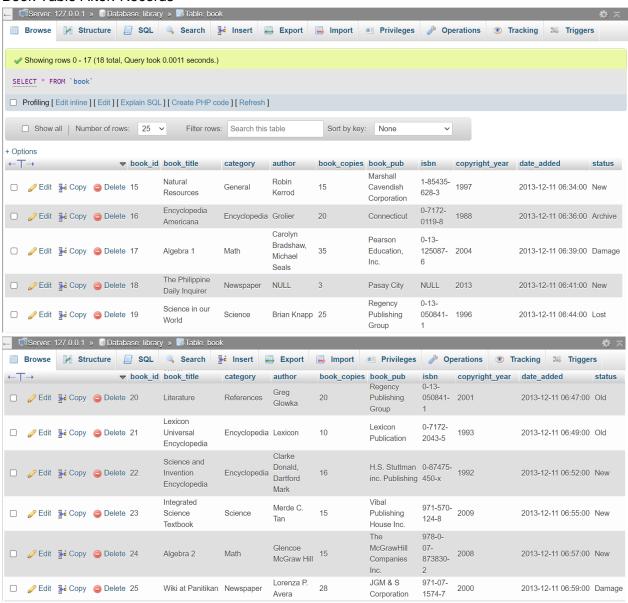


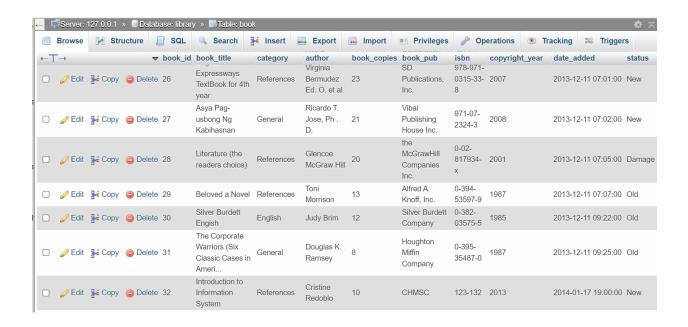
### Book Table Before: Records





### Book Table After: Records





### **Borrow Table**

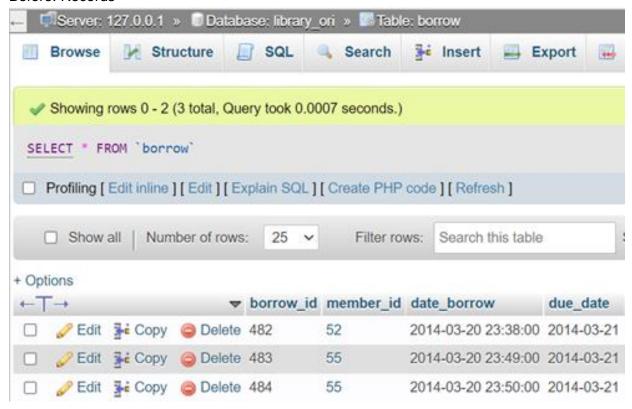
### Before: Structure



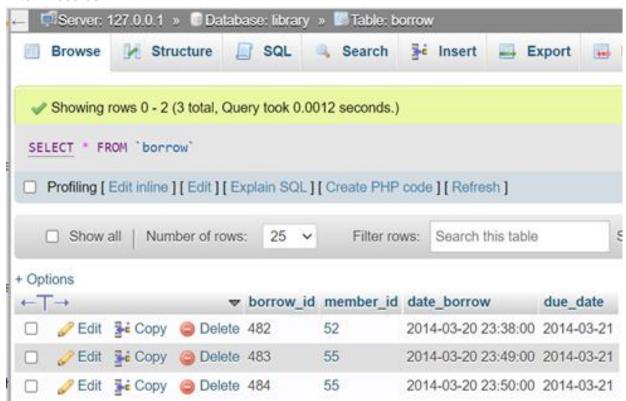
### After: Structure



### Before: Records



### After: Records

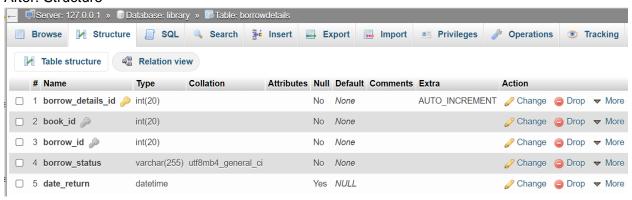


### **Borrowdetails Table**

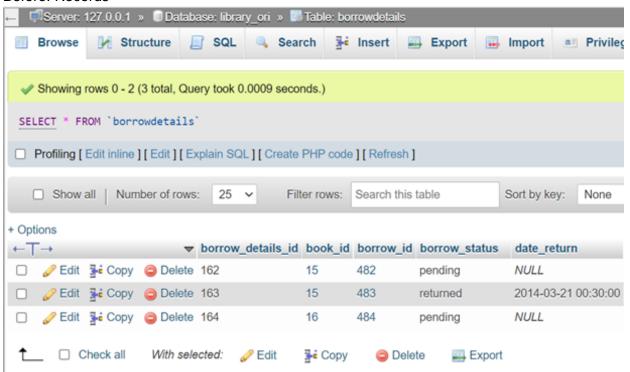
### Before: Structure



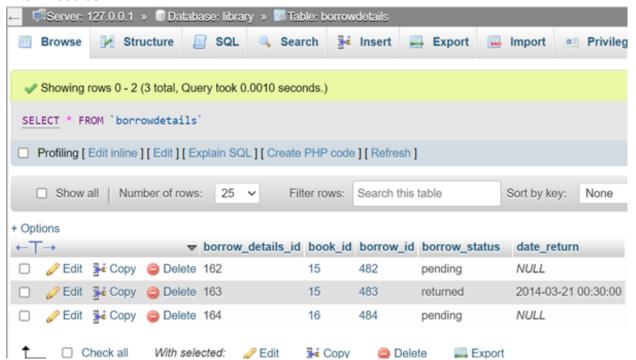
### After: Structure



### Before: Records

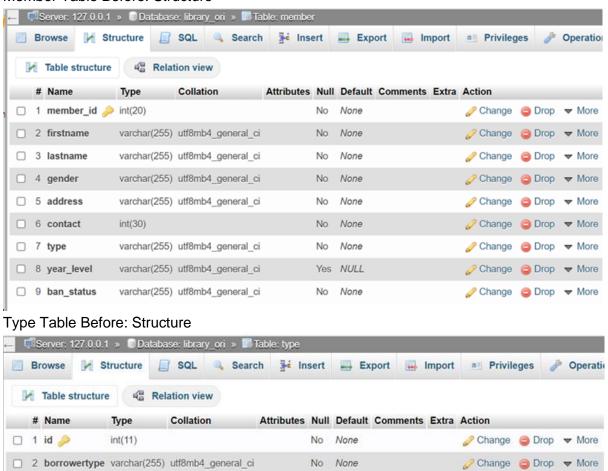


### After: Records

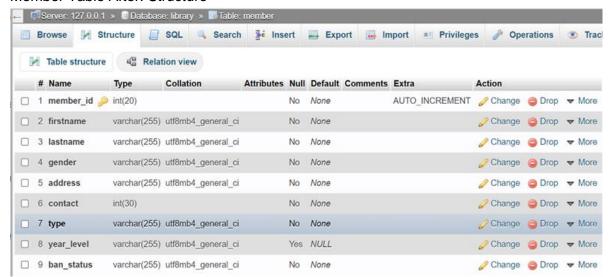


### Member Table, Type Table ⇒ Member Table

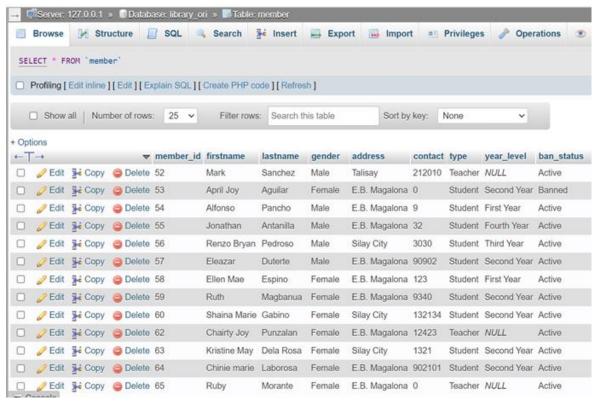
### Member Table Before: Structure



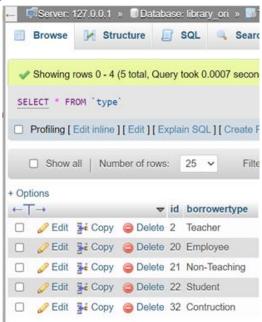
### Member Table After: Structure



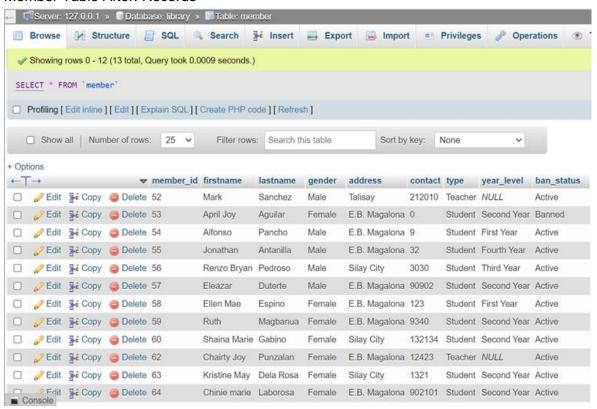
### Member Table Before: Records



Type Table Before: Records



### Member Table After: Records



### **User Table**

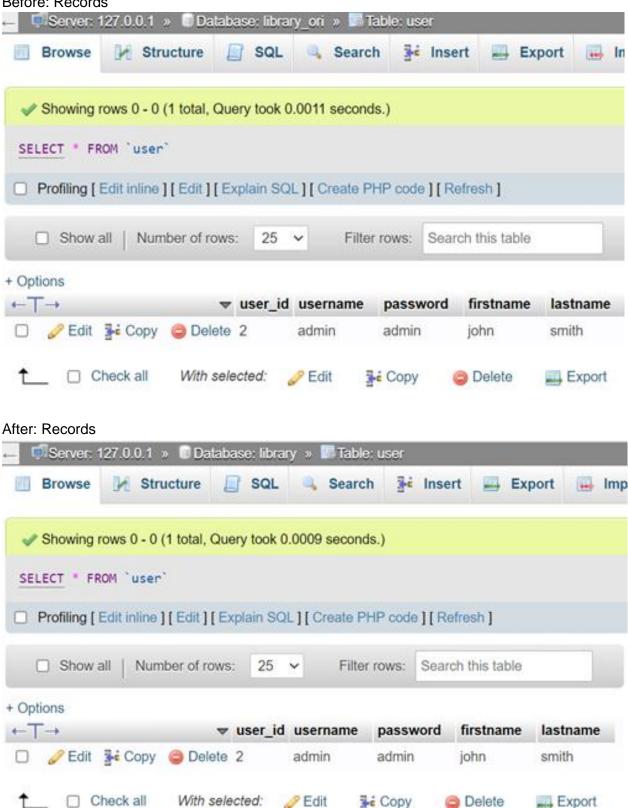
### Before: Structure



### After: Structure



### Before: Records



## TASK C: Github Repository

Link: https://github.com/zhenyiisme/COMP1044\_CW\_G5