GROUP ASSIGNMENT

TECHNOLOGY PARK MALAYSIA

AICT005-4-1- DAS

DATABASE SYSTEMS

UCDF2005(1) ICT(DI)

DAS Assignment Part 1 (Group 2)

Case Study: APU's E-Bookstore

Group Members:

1. Chan Ming Li TP060774
2. Leanne Ooi Xin Ru TP059138
3. Lim Jia Yong TP059192

Hand out Date: 23rd of March 2021

Hand in Date: 19th of May 2021

Weightage: 40%

Lecturer Name: Ms. Seetha Letchumy

Table of Contents

1.0 Introduction....3
2.0 Database and Database Management System....4
2.1 Disadvantages of File-based System....4
2.2 Advantages and Functions of Database and DBMS....6
2.3 Relation of File-based System, Database and DBMS to Case Study....9
3.0 Business Rules and Normalization....11
3.1 Business Rules....11
3.2 Normalization....12
4.0 Entity Relationship Diagram....17
5.0 Conclusion....18
References....19

1.0 Introduction

In light of the increasing population of *Asia Pacific University* (APU), the University has taken initiative to implement a new and more efficient database for *APU's e-bookstore.* This benefit both students and staffs from having to seek for books in the city due to limitations of books in the current bookstore. In order to design a new database system for *APU's e-bookstore*, detailed research has been done and systematic database design has been carried out to meet these expectations.

This report includes the pros and cons of a file-based system as well as database and database management system (DBMS), alongside with a list of business rules and implementation of normalization according to case study. Design of entity relationship diagram is drawn with all entities, attributes, and relationships clearly illustrated in the diagram with the use of Chen's notation.

2.0 Database and Database Management System

2.1 Disadvantages of File-based System

One of the well-known ways of storing useful data and information on a computer today is by keeping them on a few different files, with each file being logically related to the data stored within. While it is widely known that this way of storing information, also known as file-based systems, are much more useful in terms of efficiency when compared to manual filing systems, it is unfortunately also true that file-based systems do have some limitations and disadvantages that we should be aware of (BCCampus, 2011).

Data redundancy is one of the many disadvantages present when using a file-based system. For instance, since files which are not logically related are normally stored separately in an organisation, every department would have their own set of files and these files are most likely to be created and managed by a different person as well. This will potentially result in the same field of data being stored in multiple different files as illustrated in *Figure 2.1*, where there are duplications of customer file, stock file, order file in various places. It would be a tedious process if one of these fields needs to be updated or deleted, since it would mean that the update or deletion must be done multiple times in each set of files that happen to have that same field. Ultimately, this brings to duplication of effort to maintain these files (Watt, 2012).

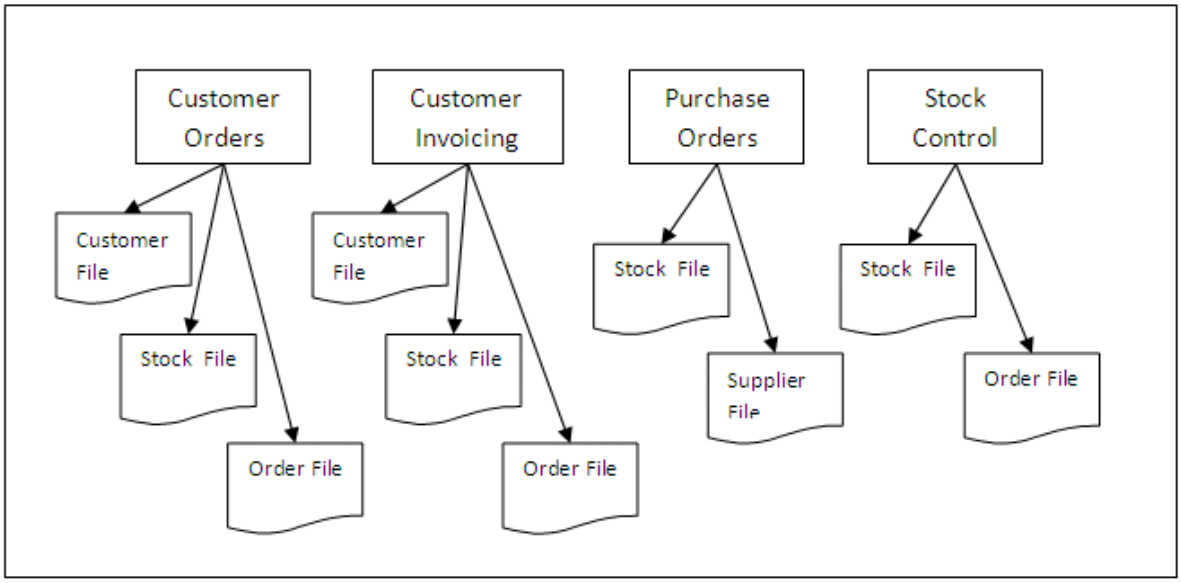

Figure 2.1: Example of Data Redundancy
Source: (Anon., 2021)

File-based system also tends to be highly time consuming to maintain. With the example provided earlier where modifications of files have to be done in several places, the process of maintaining files becomes inefficient and time consuming. The matter would be worse if these modifications are to produce errors, which are highly possible and would then require additional time from the user to resolve the error.

Besides, data redundancy often leads to inconsistency of data and integrity problems as well (Watt, 2012). For instance, while data have been updated in a different file, the current file that the user is reviewing may not happen to be the latest updated version. This results in the occurrence of data anomaly, in which two different versions of data are stored in different files, which may later lead to confusion if it is not fixed immediately. It would be disastrous if these data have to be updated daily as it maximizes the chances of error, where two files of the same field happen to store different data while ideally, they should be storing identical data.

Data isolation also tends to be a problem for file-based system as it is difficult for users from a different department to retrieve necessary data and information stored in other files. Since these data are scattered in various places, it would be a tedious process for one to look through every file randomly just to retrieve some data and information. Not only is this process highly time consuming, but it also causes difficulty for users to share useful information through multiple different applications (Tutorialink, 2017).

Furthermore, file-based systems also caused some problems in relation to data security. Proper security systems are not provided in file-based system which may cause unauthorized access of data from third parties. People with some knowledge on this matter would be able to easily change, modify or delete vital data from the file (BCCampus, 2011). Obviously, it is not a good practice if crucial files are being vulnerable to security problems such as unauthorized access and unauthorized use.

All in all, while file-based systems are more efficient than manual file systems, it is also time consuming and effort demanding for one to maintain the files and to ensure that they are all up to date and well maintained. It may also be disastrous if it is not maintained in a proper way especially when these files store huge amounts of crucial data.

2.2 Advantages and Functions of Database and DBMS

Today, huge amounts of data are generated from online searches, social media, and many other areas. A collection of logically related data, otherwise known as a database, are stored, processed, and managed by organization using a database management system (DBMS). By definition, a DBMS is a packaged software which eases the process of managing and retrieving information from databases (Anon., 2020). While databases are used to store data, database management system are used to monitor databases and provide various functions which helps in increasing the efficiency and effectiveness of data retrieval, data updating, and data manipulation. Users can access databases through several communication pathways of the network environment. Web browsers such as *Mozilla Firefox*, and *Internet Explorer* are commonly used to gain access into the database. *Oracle*, *MySQL*, and *SQL Server* are some of the popular database management systems used by many organizations for different purposes.

Database and DBMS are generally preferred in an organisation as opposed to a file processing system as the former has many advantages over the latter. First and foremost, data sharing among multiple users is improved in a DBMS, allowing users with access privileges to manipulate and monitor data easily. Subsequently, data can be shared across multiple departments in an organization which again, increases the effectiveness in managing and manipulating data such as customer information. For instance, multiple users are allowed to manipulate and update data residing in the database during data sharing, thereby allowing the database to always stay up to date. Therefore, better, and more accurate decisions can be made in the organisation, and data inconsistency can be avoided as well with the utilization of updated database (Pedamkar, n.d.).

In addition, data security is also enhanced when utilizing databases and DBMS. As the amount of users' access to database increases, undoubtedly the risk of data breaches increases as well (Castro, 2018). This risk can be reduced in DBMS as DBMS implemented some policies to ensure the security of the databases. For instance, DBMS authenticates the identity of users and only allows users with access privileges to access the database in order to ensure the integrity of the data. Firewalls and encryption systems, as illustrated in *Figure 2.2a*, are also implemented in DBMS to avoid possible data breaches and malicious attacks from outsiders who intend to steal crucial data. Hence, databases are much more secure when compared to file-based system.

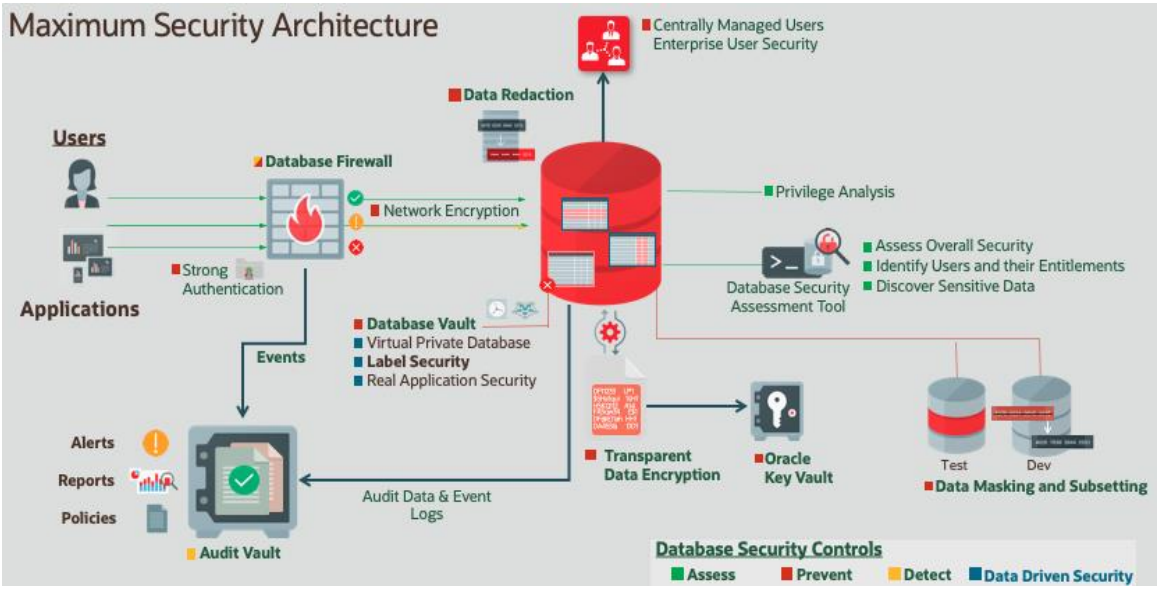

Figure 2.2a: Example of Oracle's Database Security Architecture
Source: (Cahill, 2020)

Besides, databases and DBMS have crash recovery mechanism which automatically backs up and recovers data in case of the occurrence of unfavourable circumstances (Castro, 2018). These data are usually backed up in another server, thus reducing the risk of data loss. This is vital for organizations as they may face difficulties if substantial amount of crucial data is lost. In a traditional file processing system, data backup is not impossible, but it does consume a lot of time and resources. In short, databases and DBMS speeds up and eases the process of backup and recovery, thereby saving users from suffering major data loss if anything were to happen. As an illustration, if the server crashed or is attacked by malicious attacks, the restoration of database to the previous condition can be done easily.

Moreover, the various functions provided by DBMS contributes hugely to the wide usage of DBMS by many organizations. For instance, data dictionary management is capable of storing and keeping track of relationships between each data element as well as providing metadata and some other details of these elements. This example is clearly illustrated in *Figure 2.2b* where details of each field has been provided. This is especially useful as it helps programmers to avoid complex codes. Since databases are usually large and contains many tables and constraints, data dictionary is useful for assisting users to identify different tables by providing details in it. In short, data abstraction provided by DBMS will help users to have a more organized database structure (Thiru, n.d.).

Field Name	Data type	Field Length	Constraint	Description
Client_id	Int	10	Primary key	Client id, Auto generated
Client_name	Varchar	20	Not null	Name of client
Password	Varchar2	30	Not null	Login Password for client
Contact_no	Int	15	Not null	Landline or mobile number
Email_id	Varchar2	30	Not null	Any email id
Max_Users	Int	10	Not null	Maximum number of users
Current_users	Int	10	Not null	Currently present user

Figure 2.2b: Example of Data Dictionary
Source: (Gubicak, 2016)

Furthermore, data storage management is also helpful while users are using DBMS to manage data. The storage can be used to store various handy information such as data, data entry forms, reports, procedural codes, data validation rules, and many more. This helps to increase database performance tuning, or in other words, the efficiency of database. Consequently, retrieval of data, query from database, generation of report will be timelier. Next, majority of the DBMS also supports structured query language (SQL) to query and access data. By utilizing SQL, users can specify what they want without mentioning how it can be done, which leads to users' convenience when accessing data (Thakur, 2021).

Data transformation and presentation is another crucial function supported by DBMS. DBMS will transform the inserted data to a different data structure or format which is required to be stored in the database (Thakur, 2021). Conversely, physically retrieved data can be transformed accordingly to meet the logical expectations of users as well. Hence, data is well-organized and formatted accurately to ensure integrity of data in the database.

All in all, databases and DBMS have many advantages and functions which will increase the efficiency and effectiveness of users on retrieving, monitoring, and managing data.

2.3 Relation of File-based System, Database and DBMS to Case Study

Asia Pacific University's (APU) e-Bookstore will be utilizing database system to carry out various data managing activities such as storing data on new book arrivals, tracking customer's membership, storing member's feedback, etc.

Since file-based system has the tendency of promoting data redundancy, this will cause having a poor access time and an increasing storage cost (Singh, 2015). As an example, each department of the e-bookstore has their own files and each of the members of the department has a copy of the same file which contains identical data. If there is an arrival of a new book, the bookstore manager will have to update each file in the inventory department. This will be a tedious process and might lead to a mistake in the date entry or a typo (Alley, 2018). It will be very challenging for the manager to play his or her part in maintaining the data and its accuracy. Furthermore, with the repetition of data, additional storage space will be needed, and it will not be beneficial financially to the e-bookstore as storage cost will increase.

Next, it is known that data redundancy will lead to data inconsistency and reduction of the value of data stored. Since every file requires constant updates, it is possible that the file being displayed to the members has not been updated yet. This makes purchasing a book more difficult for the bookstore's members. Moreover, the information that members are getting from the e-bookstore's website might not be timely and as useful as it should be, thus degrading the integrity of the data in the files.

The file-based system is also known to provide improper data security. Data in the system are crucial assets to *APU's e-bookstore,* this includes data regarding publishers, stocks, and member. It will be a massive issue if the data were manipulated or eliminated. Lastly, as data isolation is very common in a file-based system, all data will be stored in various files and those files might be in different formats (Singh, 2015). If the manager wishes to generate a report that consists of member's information, such as historical purchases, transactions, and feedback, it will be strenuous to retrieve the needed data.

On the other hand, database and DBMS allows data sharing which will be a prime advantage to the manager of *APU's e-Bookstore*. In a database and DBMS, data are stored in one or more servers in a network (Pedamkar, n.d.). This will improve the efficiency, thus reducing the time

needed to share data among departments or co-workers. Furthermore, this data sharing feature comes with a data protection feature which requires accessor's verifications. This brings us to the data security of database and DBMS. Database and DBMS has a mechanism that defines access privileges for each user (Pedamkar, n.d.). Therefore, only authorized users such as managers can access the data and make changes to the files. This will enhance the safety of the data stored. Next, since the e-bookstore is created to satisfy the needs and wants of *APU* population, naturally it is expected that a large scale of data will be stored in the database system. The crash recovery mechanism in the database and DBMS will be of good use to backup and provide smooth recovery of data when there is a system malfunction or system crash. This will reduce the challenges in obtaining data after the occurrence of an unfavorable event.

DBMS has a data dictionary management function as well. This function avoids data duplications, promotes data integrity, and improves data consistency. Since this function adopts consistent data elements and terminology, the manager of the e-bookstore will be able to analyze data efficiently and that will lead to making professional decisions (Agustiyadi, 2013). Besides, insertion of data can also be done with ease as data dictionary management makes programming easier and less time consuming (Agustiyadi, 2013).

The storing of e-bookstore's data is much neater and more organized with the presence of data storage management feature in DBMS as data will be stored in the same server regardless of data type. Thus, the process of obtaining data will be easier. Lastly, DBMS has a data transformation and presentation function. This function helps to convert data according to the requirements and needs of users. For instance, customers who wish to purchase any products will get the display of the *APU e-Bookstore* website. Similarly, when they wish to rate books that they have purchased, they will be able to see a feedback form. This function gives the database system a more flexible and versatile touch.

In conclusion, a database and DBMS is a more suitable choice for *APU's e-Bookstore*. This decision will benefit stakeholders of the e-bookstore.

3.0 Business Rules and Normalization

3.1 Business Rules

1. Every publisher has a publisher ID (PK), a name, an address, and a phone number.
2. Every member has a member ID (PK), a name [First name, Last Name], gender, an email, an address, and a phone number.
3. Each book has an ISBN (PK), title, synopsis, publish date, price, author, stock quantity, and genre.
4. Each order has an order ID (PK), order date, and delivery status.
5. The order quantity of each book is recorded as well.
6. Each cart has a cart ID (PK), quantity, and cost of book can be derived accordingly.
7. Each sale consists of a sale ID (PK), sale date, and delivery status.
8. The sale quantity of each book also is recorded.
9. Each feedback has a feedback ID (PK), a rating (ranging from 0 to 10, where 0 = terrible, 10 = masterpiece) and one optional comment.
10. Each publisher may send zero or many books, and each book is sent by exactly one publisher. 1:M
11. Each order contains one or more books, and each book may appear in zero or many orders. M: N
12. Each publisher receives zero or many orders, each order is received by exactly one publisher. 1:M
13. Each member has zero or many cart items, and each cart item is associated with exactly one user. 1:M
14. Each cart item contains exactly one type of book, and each book may appear in zero or many carts. M: 1
15. Each member can make zero or many sales by purchasing books, each sale is made by exactly one member. 1:M
16. Each sale contains one or many books, and each book may appear in zero or many sales. M: N
17. Each member can give zero or many feedbacks, and each feedback is written by exactly one member. 1:M
18. Each feedback belongs to one book, each book can have zero or many feedbacks. M:1

3.2 Normalization

APU's e-Bookstore sells book to members of the bookstore, details of each sale are maintained and recorded in a database. The table shown below *(Table 3.2)* includes some of the information stored on *APU's e-Bookstore* system. Each sale is made by one of the bookstore's member, and each member can purchase multiple books as per their need. Members' name and address are assumed to be not unique.

Table 3.2: APU's e-Bookstore Table

SaleID	***SaleDate***	***MemberID***	***Name***	***Address***	***ISBN***	***BookTitle***	***Genre***	***Price (RM)***	***Quantity***
S0001	3/1/2021	M0001	Harry Potter	7, Jalan Maju, Bangsar Utama, Kuala Lumpur.	978-1-1797-0381-7	Copper Heart	Fantasy	27.50	1
					978-2-5318-4227-4	Compass	Adventure	35.00	2
S0002	15/1/2021	M0015	Hermione Granger	28, Jalan Tan Ming, Taman Puncak Jalil, Seri Kembangan.	978-3-2533-1195-6	Vanishing Man	Horror	59.99	2
					978-2-8894-9691-4	Artificial Puppet	Comedy	40.00	1
					978-1-8941-5788-9	The Passengers	Romance	37.00	3

SaleID	***SaleDate***	***MemberID***	***Name***	***Address***	***ISBN***	***BookTitle***	***Genre***	***Price (RM)***	***Quantity***
S0003	2/2/2021	M0002	Ronald Weasley	34, Jalan Sultan Ismail, Kuala Lumpur.	978-7-4866-6264-9	Black Demon	Fantasy	29.90	4
S0004	18/2/2021	M0007	Draco Malfoy	15, Jalan Maharajalela, Taman Ast, Seremban.	978-7-2196-5078-3	Storm of Kings	Sci-fi	17.00	2
					978-1-1797-0381-7	Copper Heart	Fantasy	27.50	1
S0005	25/2/2021	M0012	Woody Hard	50, Jalan Desa Bakti, Taman Desa, Kuala Lumpur.	978-1-0528-8971-3	Rogue's Duel	Fantasy	55.00	1

Unnormalized Form (UNF)

Sale (SaleID, SaleDate, MemberID, Name, Address, (ISBN, BookTitle, Genre, Price, Quantity))

The table shown above can be divided into two groups, namely non-repeating groups, and repeating groups. Non-repeating groups are attributes in this table which do not have more than one value in a cell such as SaleID, SaleDate, MemberID, Name, and Address. Repeating groups on the other hand, are attributes with more than one value in a cell. ISBN, BookTitle, Genre, price, and quantity are examples of repeating groups in this table. Since this table contains repeating groups, it is considered as an unnormalized table.

First Normal Form (1NF)

Sale2 (SaleID, ISBN, SaleDate, MemberID, Name, Address, BookTitle, Genre, Price, Quantity)

To convert unnormalized form to first normal form, repeating groups must not be present in the relation. This can be done by 'flattening' the table, which is a technique whereby rows which previously contained repeating groups are broken down into more rows to ensure each cell contains only one value. Empty cells as the result of this process should be filled with appropriate values as well. A composite key is required now to uniquely identify each row. In this case, SaleID and ISBN will make up the composite key since the combination of both primary keys can uniquely identify each record. Each table cell now consists of exactly one value and a composite key can be used to identify them; therefore, first normal form is achieved.

Functional Dependencies

SaleID → SaleDate, MemberID, Name, Address (partial dependency)

ISBN → BookTitle, Genre, Price (partial dependency)

MemberID → Name, Address (transitive dependency)

Before the conversion of first normal form to second and third normal form, relationship between attributes, also known as functional dependencies, such as partial dependency and transitive dependency must be identified first. Partial dependency can be found when a non-primary-key attribute is functionally dependent on a part of primary key. For instance, non-key attributes such as SaleDate, MemberID, Name and Address can be determined by SaleID which

is part of the primary key. Hence, the relationship between them is considered as partial dependency.

Besides, transitive dependency can be found when a non-primary-key attribute is indirectly dependent on a primary key. For example, SaleID, which is a part of primary key, can determine MemberID which is a non-primary-key attribute. On the other hand, MemberID can determine other non-primary-key attributes such as 'Name' and 'Address'. Therefore, it can be concluded that Name and Address are indirectly or transitively dependent on SaleID. Their relationship is known as transitive dependency.

Second Normal Form (2NF)

SaleDetails (SaleID, SaleDate, MemberID, Name, Address)

Book (ISBN, BookTitle, Genre, Price)

Sale3 (SaleID, ISBN, Quantity)

After identifying all functional dependencies in the table, second normal form (2NF) is carried out to further reduce data redundancies. To be in 2NF, all relations must satisfy conditions in first normal form and all partial dependencies should be eliminated, in other words, all non-primary-keys attribute must be functionally dependent on the primary key and not portion of primary key. To achieve this, table names such as SaleDetails, Book and Sale3 are created and written separately. Previous determinant and dependents included in the partial dependency will be attributes of the newly created tables, respectively. In addition, determinants will be the primary key of its own table. Thus, most of the anomalies have been eliminated and 2NF has been achieved.

Third Normal Form (3NF)

Member (MemberID, Name, Address)

SaleDetails (SaleID, SaleDate, MemberID)

Book (ISBN, BookTitle, Genre, Price)

Sale3 (SaleID, ISBN, Quantity)

To further reduce data anomalies and redundancies, third normal form (3NF) is carried out. This stage requires all transitive dependencies to be removed. In order to remove the transitive dependency relationship between MemberID, Name and Address, an additional table, Member,

is created where these transitively dependent attributes can be placed together. MemberID will be the primary key here since it is the determinant of the previous transitive dependency relationship. Dependents of the transitive dependency such as Name and Address will be attributes of the Member table. Since the table is now in 2NF, and there are no transitive dependencies left between attributes, this table has now achieved 3NF.

4.0 Entity Relationship Diagram

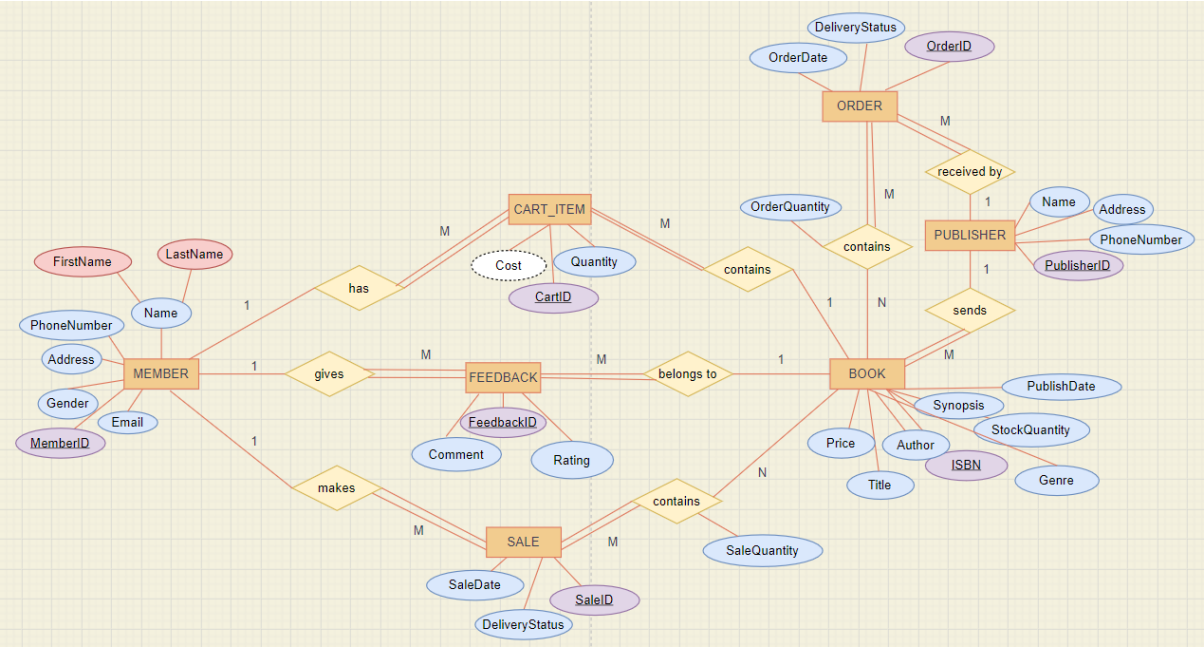

Figure 4.0: ERD for APU's E-Bookstore

5.0 Conclusion

In a nutshell, maintaining a file-based system poses many disadvantages such as data redundancy, highly time consuming, problems related to data inconsistency and data integrity, data isolation, security vulnerabilities, etc. Nonetheless, these disadvantages can be avoided with database and database management system (DBMS). Database and Database Management System (DBMS) brings us several advantageous functionalities such as data sharing, data security, data dictionary management, data storage management, data transformation and presentation, crash recovery mechanism and many more. Users are able to increase their efficiency and effectiveness on their tasks due to the convenience brought by database and DBMS. Hence, database and DBMS is a sensible option for monitoring, retrieving and manipulating data in order to have an organized and neat data storage. When envisioning file-based system as the database system for the *APU's e-bookstore*, it was clear that this system is not suitable for the e-bookstore's system as opposed to a database. There will be an increase in workload and complications. On the other hand, database, and database management system (DBMS) makes every operation more efficient, allowing all data to be more organized and useful. It is also a safer and secure option. Thus, database and database management system (DBMS) is more compatible as the database system of *APU's e-bookstore* compared to file-based system. Next, there is a total of 16 detailed business rules, and an Entity Relationship Diagram (ERD) was designed based on them. In addition, a relational schema was created according to the ERD. Lastly, normalization up to three normal form (3NF) was carried out with an instance of the *APU's e-bookstore* table. With these carefully planned designs, *APU's e-bookstore* system is ready to be implemented.

References

- Agustiyadi, 2013. *Wordpress.* [Online]
  Available at: https://agustiyadi.wordpress.com/2013/06/21/the-benefits-of-a-data-dictionary/
  [Accessed 27 March 2021].
- Alley, G., 2018. *Alooma.* [Online]
  Available at: https://www.alooma.com/blog/what-is-data-redundancy
  [Accessed 27 March 2020].
- Anon., 2020. *Database Management Systems (DBMS) explained - IONOS.* [Online]
  Available at: https://www.ionos.com/digitalguide/hosting/technical-matters/database-management-systems-dbms/
  [Accessed 26 3 2021].
- Anon., 2021. *Traditional file-based approach.* [Online]
  Available at: https://www.cs.uct.ac.za/mit_notes/database/htmls/chp01.html
  [Accessed 24 March 2021].
- BCCampus, 2011. *Database Management System.* [Online]
  Available at: https://opentextbc.ca/
  [Accessed 24 March 2021].
- Cahill, S., 2020. *Oracle Cloud Security.* [Online]
  Available at: https://blogs.oracle.com/cloudsecurity/oracles-maximum-security-architecture-for-database-security
  [Accessed 26 3 2021].
- Castro, K., 2018. *Advantages of Database Management System.* [Online]
  Available at: https://www.tutorialspoint.com/Advantages-of-Database-Management-System
  [Accessed 26 3 2021].
- Gubicak, A., 2016. *What is a SQL Server Data Dictionary and why would I want to create one?.* [Online]
  Available at: https://www.sqlshack.com/sql-server-data-dictionary-want-create-one/
  [Accessed 26 3 2021].

- Pedamkar, P., n.d. *Educba.* [Online]
  Available at: https://www.educba.com/advantages-of-dbms/
  [Accessed 27 March 2021].
- Pedamkar, P., n.d. *EDUCBA.* [Online]
  Available at: https://www.educba.com/advantages-of-dbms/
  [Accessed 26 3 2021].
- Singh, C., 2015. *Beginners Book.* [Online]
  Available at: https://beginnersbook.com/2015/04/dbms-vs-file-system/
  [Accessed 27 March 2021].
- Thakur, D., 2021. *What are the Functions and Service of DBMS.* [Online]
  Available at: https://ecomputernotes.com/fundamental/what-is-a-database/functions-and-service-of-dbms
  [Accessed 26 3 2021].
- Thiru, n.d. *Functions of Database Management System (DBMS).* [Online]
  Available at: http://www.myreadingroom.co.in/notes-and-studymaterial/65-dbms/465-functions-of-dbms.html
  [Accessed 26 3 2021].
- Tutorialink, 2017. *Advantage and Disadvantages of File-oriented system.* [Online]
  Available at: https://tutorialink.com/dbms/advantage-and-disadvantages-of-file-oriented-system.dbms
  [Accessed 13 April 2021].
- Watt, A., 2012. *Chapter 1 Before the Advent of Database Systems.* [Online]
  Available at: https://opentextbc.ca/dbdesign01/chapter/chapter-1-before-the-advent-of-database-systems/
  [Accessed 24 March 2021].