Table of contents

Front matter

List of figures

List of abbreviations

CK	Candidate Key
PK	Primary Key
FK	Foreign Key
SK	Super Key
C.K	Composite Key

Chapter 1

Introduction

This assignment depends on normalization ,one of the major concepts in Database Systems .This assignment introduce you with use of normalization , its different forms , when not to use this major concept , when to use this concept and one practical task related to library management system .I tried my best to complete this task ,I know there would be lot of mistakes but I burnt midnight oil to complete it . I hope my instructor will like it

Note: Author made this assignment by herself

When we talked about file based system we said that it has decentralized approach .To prevent this Database system was introduced.Obviously Database system prevents redundancy and repetition of data .For this normalization plays key role .Every table in database has to be in normal form for better performance . What if we don't use normalization?

- Having repeated data not only makes process slow but can cause trouble during later part of transaction .
- Huge data sets in database without any purpose is wastage of time

1.1 What is ”Normalization”?

Normalization is a technique of organizing data in Database system.It is a systematic approach of decomposing data to eliminate data redundancy . Following are the purpose of normalization

- Improves data integrity
- Reducing the redundancy of data
- To ensure the data dependency makes some logical sense

Here are some forms of normalization .

1.2 What happen if we have redundant data?

Anomalies are basically technical problems we face if we have redundant data , When we want to insert ,delete or update values .It is of three types:

Insertion anomaly : Suppose a new position in a company Miss Hadia is selected , but the department has not been for her. In that case if you want to update her information , we need set the department information as null .Similarly if we have to insert data of 1000 employees of similar situation , department information will be repeated for all those employees

Update Anomaly: Let Miss Hadia left the company .In that case all the records will have to be updated and if by mistake we miss any data record, it will leads to data inconsistency

Deletion Anomaly: Let two pieces of information are kept together (employee and department),hence at the end of financial year if employee records are deleted we will also loose department information

These are the problems occur while managing the data

1.3 Types of Normalization

1.3.1 First Normal Form

In this form , we tackle problem of atomicity ie values in the table should not be divided in other words single cell cannot hold multiple values

Functions

- It removes repeated groups from the table
- It creates separate table for each set of the data
- It identifies each set of data with **Primary key** to understand this in a better way

Book ID	Author
2001,2002	DATABASE SYSTEMS , ALGORITHMS
2001	DATABASE SYSTEMS

Table 1.1: Un normalized table

Book ID(PK)	Author
2001	DATABASE SYSTEMS
2002	ALGORITHMS

Table 1.2: Normalized table

1.3.2 Second Normal Form

The table is put in 2NF if it fulfils following conditions

- It has to be in 1NF
- Table should not contain partial dependency (ie non prime depends on only one part of the composite key)

customer ID	Store ID	Location
1001	S001	Faisalabad
1001	S003	Abbaspur
1002	S001	Faisalabad
1003	S002	Kotli

Table 1.3: Un normalized table

customer ID	Store ID
1001	S001
1001	S003
1002	S001
1003	S002

Table 1.4: Normalized table 1

Store ID	Location
S001	Faisalabad
S003	Abbaspur
S002	Kotli

Table 1.5: Normalized table 2

Composite key: Customer ID, Store ID

Prime Attributes:Customer ID, Store ID

Non-primary key: Location

1.3.3 Third Normal Form

It is used to reduce duplication data and ensure referential integrity .It contains following rules:

- The table has to be in 2NF.
- No non-prime attribute is transitively dependent on any non-prime attribute which depends on another non prime attribute .For Example ,In maths we say: if a=b and b=c then a is also equal to c. .This should not be happened.
- Non-prime attribute must depend on the prime attribute .

  Purpose

- Eliminate undesirable data anomalies
- Reduce the need for restructuring over time.
- To make data model more informative.

Name	Physname
Gore.A	Scholl F
Bush.G	Seuss
Cheney.J	Spock

Table 1.6: Un normalized table

PatID	Name	Physname
54321	Gore.A	Scholl F
54322	Bush.G	Seuss
54333	Cheney.J	Spock

Table 1.7: normalized table

1.3.4 Boyce Codd Normal Form

Also known as 3.5 normal form , higher form of 3NF.

Conditions:

- Table has to satisfy 3NF .
- In Boyce Codd Normal Form , if every functional dependency that is A implies B then A has to be super key of that table .Super key (Group of single or multiple key which identifies row in a table

Student ID	course	Teacher
S101	DATABASE SYSTEMS	Mr.Mehtab Mushtaq
S103	LEARN PYTHON	Miss Umera
S102	ALGORITHM	Mr.Nabeel Ali
S103	LEARN PYTHON	Miss Umera

Table 1.8: Un normalized table

Student ID	course
S101	DATABASE SYSTEMS
S103	LEARN PYTHON
S102	ALGORITHM

Table 1.9: normalized table 1

course	Teacher
DATABASE SYSTEMS	Mr.Mehtab Mushtaq
LEARN PYTHON	Miss Umera
ALGORITHM	Mr.Nabeel Ali

Table 1.10: normalized table 2

1.3.5 Fourth Normal Form

This is implemented after Boyce Codd Normal Form . It contains following conditions :

- It should satisfy BCNF.
- It should not have multi valued and dependency

Example : A person has two phone numbers and two emails in table we will write it as

Name	Phone	Email
Asma	$phone_1$	$email_1$
Asma	$phone_1$	$email_2$
Asma	$phone_2$	$email_1$
Asma	$phone_2$	$email_2$

Table 1.11: Example table of 4NM

which is a bad design . For this two separate tables are created.

Name	Phone	Email
Asma	$phone_1$	$email_1$
Asma	$phone_1$	$email_2$

Table 1.12: Normalized table of 4NM

Name	Phone	Email
Asma	phone_2	email_1
Asma	phone_2	email_2

Table 1.13: Example table of 4NM

1.3.6 Fifth Normal Form

Conditions

- The table should be in 4NF
- It should not have join dependency. If the relation has a join dependency then it can be decomposed into smaller relations such that one can rejoin these relations to reproduce the original relation.
- This problem is faced when we rejoin smaller tables

1.3.7 Sixth Normal Form

- It is irreducible normal form .This is last form after this data cannot be further normalized .
- Generalization Multi level Dependencies are eliminated
- Each non-prime key attribute depends on entire primary key
- each table ha primary key

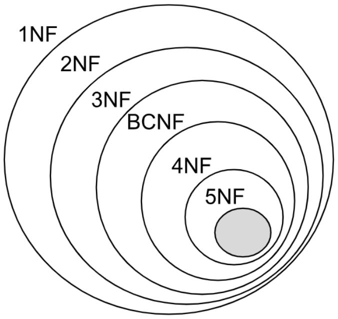


Figure 1.1: Levels of normalization

1.4 When Not to normalize?

There are situations where normalization might not be the best approach.

In online analytical processing system (OLAP) Denormalized tables are used .For example The manager wants to take a large number of backdated historical data.He wants to analyze the data and he wants to do some forecasting on the data .This means he wants to do some Kinds of sales management on the data .In these kinds of heavy operations done on the data such as selecting the data ,creating a complicated report of data etc

Consider a dataset containing financial transactions, such as stock prices or trading volumes. In this context, the original scale of the data is crucial for understanding the magnitude of the transactions. Normalizing this data could make it difficult to interpret the results, as the normalized values would lose their connection to the original currency or units.

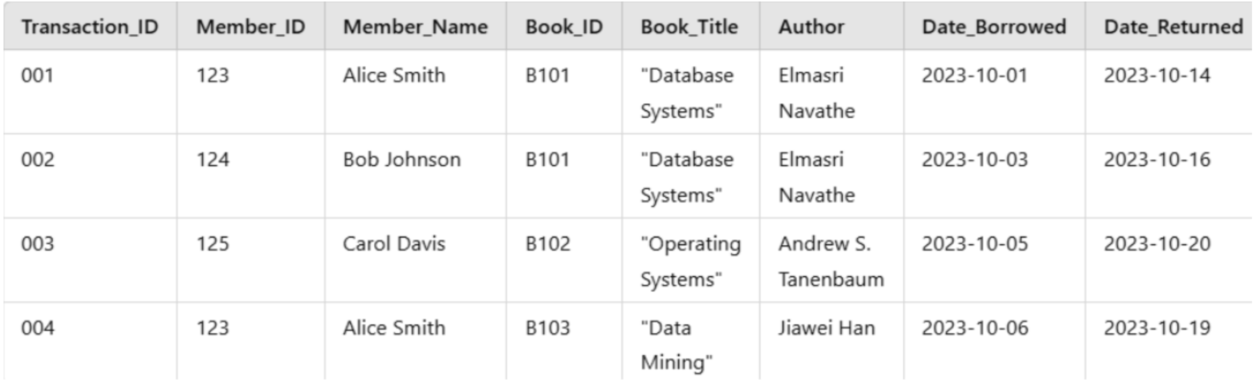
Example: Stock Price Analysis

Chapter 2

Practical Task

Scenario:

You’ve been hired by a local library to help them redesign their database system. The library currently stores book lending records in a flat, unstructured table which is inefficient, and they face redundancy and update issues. Below is a sample of their d



Transaction_ID	Member_ID	Member_Name	Book_ID	Book_Title	Author	Date_Borrowed	Date_Returned
001	123	Alice Smith	B101	"Database Systems"	Elmasri Navathe	2023-10-01	2023-10-14
002	124	Bob Johnson	B101	"Database Systems"	Elmasri Navathe	2023-10-03	2023-10-16
003	125	Carol Davis	B102	"Operating Systems"	Andrew S. Tanenbaum	2023-10-05	2023-10-20
004	123	Alice Smith	B103	"Data Mining"	Jiawei Han	2023-10-06	2023-10-19

Figure 2.1: Task Table

Step 1:Analyzing problems

Unnormalized Database Problems: Identify at least three major problems in the current library database table (hint: look for redundancy, anomalies, or inefficiencies). Explain why these are problematic.

Solution

- **Data Redundancy**
  Data redundancy occurs when same piece of information is stored multiple times .Here we can see member id and his name is repeated in 1st and last column . Book id ,name and its author is repeated in first and second column .
- **Deletion Anomaly**
  If we delete the row ,Useful data will be also deleted.
- **Update Anomaly**
  Update anomalies occur when changes to data require multiple updates across different records
- **Insertion Anomaly**
  Insertion anomaly occur when we want to add row in the table in this case some extra cells will be written NULL
- **Why it is problematic?**
  1. Storage inefficiencies
  2. Update Anomalies
  3. Increased Complexity
  4. Loss of valuable data
  5. Compromised data integrity.
  6. Reduced historical insight
  7. Data Inconsistency
  8. Increased Risk of errors
  9. Maintenance overhead

  Step 2:Applying Normalization

    - First we have to resolve issue of repetition of data within a single cell .Here a table is already in First Normal Form because they have primary keys so we will move toward next Normal form.
    - Here Transaction is , member id and book id are primary keys.
    - In second normal form we already discussed that table should be in first condition must me satisfied that is table should be in FNM .
      Second thing we will see that table should not contain partial dependency .Here there is no composite key , so we will split tables:

member ID (PK)	member name
123	Alice Smith
124	Bob Johnson
125	Carol Devil

Table 2.1: Member table

Book ID(PK)	Book Title	Author
B101	”Database systems”	Elmasri Navathe
B102	”Operating systems”	Andrew.S Tanenbaum
B103	” Data Mining”	Jiawei Han

Table 2.2: Book table

Transaction ID(PK)	Borrow date	return date
001	2023-10-1	2023-10-14
002	2023-10-3	2023-10-18
003	2023-10-5	2023-10-30
004	2023-10-6	2023-10-19

Table 2.3: Split table Transactions

- Now we will move our tables toward third normal form where transitive dependency will be removed .There is no problem in table .All the non-prime keys are fully dependent on their primary keys. So our final table is:

Transaction ID(PK)	Member id(FK)	Book Id (FK)	Borrow date	return date
001	123	B101	2023-10-1	2023-10-14
002	123	B101	2023-10-3	2023-10-18
003	124	B102	2023-10-5	2023-10-30
004	125	B103	2023-10-6	2023-10-19

Table 2.4: Transaction table

Transaction ID(PK+FK)	Member id(FK)	Book Id (FK)
001	123	B101
002	123	B101
003	124	B102
004	125	B103

Table 2.5: Relationship table

Identified keys

- Members Table: Member ID (Primary key of member table)

- Books Table: Book ID (Primary key of book table)

- Transactions Table: Transaction ID (Primary Key), Member ID and Book ID (foreign keys)

Step:3 Final ERD

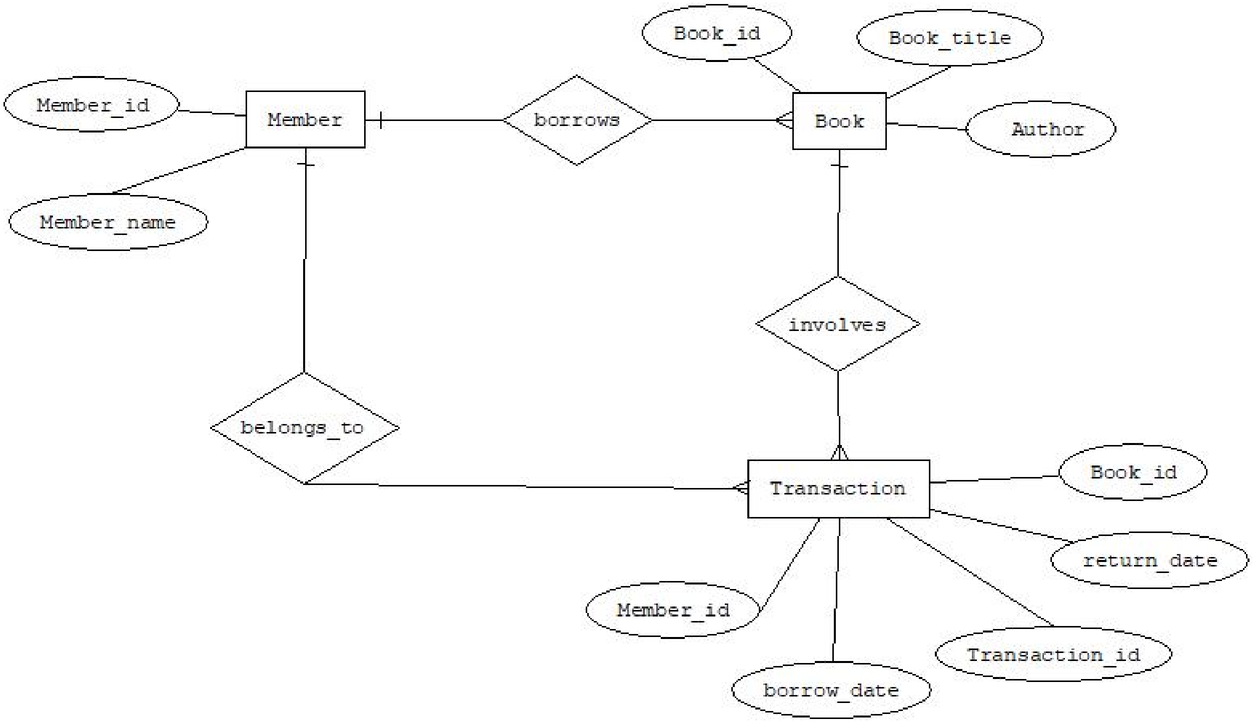


Figure 2.2: Final ERD

Logic used for transaction

- The Member ID and Book ID are used to link the transaction to the corresponding member and book.
- The Borrow Date and Return Date are used to track the duration of the transaction.

Chapter 3

Results

3.1 Summary

This assignment highlighted forms of normalization . Here i explained various points related to normalization , anomalies , , data redundancy , inconsistency , its role in DBS and how to solve out these problems .

3.2 Conclusion

Normalization is one of the major concepts in Database Systems . It reduces inconsistency , Data redundancy and increase efficiency , prevent storage problems and obviously time saving .In second scenario I normalized a table keeping rules in mind and at the end i made an ER-Diagram of final table.

Chapter 4

References

- https://youtu.be/rBPQ5fg$_{k}iY?si = cu_{h}ddl_{n}37rFVgf$
- When not to normalize https://youtu.be/TV4yxwCXJ24?si=xP5QL6hRsQ2GD$_{L}D$
- Database system project of university of Sargodha on library management system https://www.scribd.com/document/326829849/Database-System-for-Library-Management-s
- Rules of Normalization explained by Kolade Chris https://www.freecodecamp.org/news/database-normalization-1nf-2nf-3nf-table-examples/
- A you tube video by Studytonight with Abhishek https://youtube.com/playlist?list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3si=Hm2ojJciuHDQL0xC
- Chapter 14 of Database System sixth edition by Thomas M.connolly and Carolyne E.Begg