NORMALIZATION

LOGICAL DATA ANALYSIS

- *Functional Dependency*
- *Process of Normalization*
- *First Normal Form*
- *Second Normal Form*
- *Third Normal Form*

What Is Normalization ?

- The process of identifying the correct location of each attribute and the correct structure of the relations in the system is called normalization

- It is the process by which incorrectly constructed relations are decomposed into multiple correctly constructed relations

Normalization

- You normalize a database in order to ensure data consistency and stability, to minimize data redundancy,

- Redundant data can lead to -
  - Inconsistency in data retrieval
  - Errors while updating data tables

Normal Forms

- Dr. Codd originally defined three levels of normalisation.

- These three levels were called *first normal form*, *second normal form* and the *third normal form* respectively.

Normal Forms

- Normalization is usually discussed in terms of forms.
- Normal forms are table structures with minimum redundancy
  - ✓*First Normal Form (1st NF)*
  - ✓*Second Normal Form (2nd NF)*
  - ✓*Third Normal Form (3rd NF)*
  - ✓*Boyce-C odd Normal Form*

Functional Dependency

- Functional dependency is a relationship between attributes (fields).
- Because of functional dependency, given the value of one attribute, it is possible to look up the value of another, attribute.
- In general, attribute Y is functionally dependent on attribute X, if the value of X determines the value of Y

First Normal Form

- This form is represented as 1NF. To put data into 1 NF all the repeating information has to be removed.

First Normal Form - definition

- A table is said to be in its first normal form if and only if all the underlying domains contain values that are not decomposable any further.

First Normal Form

Mem-Name	Contact	Book Code	Cat Code	Issue Date	Return D
JOE	4142319	BO020	SCIENCE	18-3-98	18-4-98
SAM	8900909	BO189	ASTRO	18-3-98	18-4-98
		BO090		183-98	18-4-98
		BO091		18-3-98	18-4-98
PAUL	421 9498	B0656	ROMANC	19-3-98	19-4-98
JENNY	567 0967	BOl98	ADVENT	20-3-98	20-4-98
		B0212		20-3-98	20-4-98
		B0400		20-3-98	20-4-98
		B0555		20-3-98	20-4-98
PETER	328 9565	BOOO I	CLASSIC	21-3-98	21-4-98
	Table 1 - Unnormalised data				

First Normal Form

Mem-Cd	Mem-Name	Contact	Book Code	Cat Code	Issue Dt	Return Dt
MOO I	JOE	4142319	BO020	SCIENCE	18-3-98	18-4-98
MOO8	SAM	8900909	BOl89	ASTRO	18-3-98	18-4-98
MOO8	SAM	8900909	BO090	ASTRO	18-3-98	18-4-98
MOO8	SAM	890 Q909	BO091	ASTRO	18-3-98	18-4-98
MO67	PAUL	421 9498	B0656	ROMANC	19-3-98	19-4-98
MOl23	JENNY	567 0967	BO198	ADVENT	20-3-98	20-4-98
MOI23	JENNY	567 0967	B0212	ADVENT	20-3-98	20-4-98
MOl23	JENNY	567 0967	B0400	ADVENT	20-3-98	20-4-98
MOl23	JENNY	567 0967	BO555	ADVENT	20-3-98	20-4-98
MO880	PETER	328 9565	QOOOT	CLASSIC	21-3-98	21-4-98

First Normal Form

- The un-normalized data in the first table has been listed together as shown in table 1.
- A key that will uniquely identify each record should be assigned to the table.
- This key has to be unique because it should be capable of extracting information from any row. In our case, a unique way of identifying a row would be a combination of member code + book code. This will be the primary key.

Problem with First Normal Form

- Several problems can be identified with this table with respect to inserting, updating and deleting data. Let's understand what these problems are:
- ***Inserting data***: A new book in the library cannot be entered until it is issued to some one.
- ***Updating data***: Any change in one of the records will lead to changing more than one record.
- ***Deleting records***: After a member returns a book, that particular record will be deleted. This will also result in loss of information about the book and its category.

Second Normal Form

- To bring a table to its second normal form, data that is dependent on part of the key should be separated.
- To be in 2NF, a relation must be already in 1NF and its non-key attributes must be fully functionally dependent on the primary key
- All functional dependencies must be noted and each non-key attribute must be functionally dependent on all of the attributes that make up the primary key

Second Normal Form

Member Code	Member Name	Contact
MOO I	JOE	4142319
MOOS	SAM	8900909
MO67	PAUL	421 9498
MOl23	JENNY	567 0967
MOSSO	PETER	328 9565

Table 3 - Member Table

The primary key in table 3 is the member code.

Second Normal Form

Book Code	Member Code	Category-Code	Issue Date	Return Date
BO020	MOO I	SCIENCE	18-3-98	18-4-98
BO189	MOOS	ASTRO	18-3-98	18-4-98
BO090	MOOS	ASTRO	18-3-98	18-4-98
BO091	MOOS	ASTRO	18-3-98	18-4-98
B0656	MO67	ROMANC	19-3-98	19-4-98
BO198	MOl23	ADVENT	20-3-98	20-4-98
B0212	MOl23	ADVENT	20-3-98	20-4-98
B0400	MOl23	ADVENT	20-3-98	20-4-98
B0555	MOl23	ADVENT	20-3-98	20-4-98
B0001	MO880	CLASSIC	21-3-98	21-4-98

Table 4 - Issue Table

Second Normal Form - Definition

- A table is in its second normal form when it is already in its first normal form and every field which is not a key is fully dependent on the primary key.
- To convert a table to 2NF, follow these steps:
  - Find and remove fields that are related to only a part of the key
  - Group the removed items in another table
  - Assign the new table with a key that is a part of the old composite key

Third Normal Form

- A table is said to be in its third normal form and when it is in its 2NF and every field which is not a key is functionally dependent on just the primary key.
- To be in the 3NF, a relation must be in 2NF and each non-key attribute must depend only on the Primary Key and on the entire Primary Key.

Third Normal Form

- The member table, table 3 remains as it is:

Member Code	Member Name	Contact
MOO I	JOE	4142319
MOOS	SAM	8900909
MO67	PAUL	421 9498
MOl23	JENNY	567 0967
MOSSO	PETER	328 9565

Third Normal Form

- The Issue Table is further decomposed:

Member Code	Book Code	Issue Date	Return Date
MOO I	BO020	18-3-98	18-4-98
MOOS	BO189	18-3-98	18-4-98
MOOS	BO090	18-3-98	18-4-98
MOOS	BO091	18-3-98	18-4-98
MO67	B0656	19-3-98	19-4-98
MOl23	BO 198	20-3-98	20-4-98
MOl23	B0212	20-3-98	20-4-98
MOl23	B0400	20-3-98	20-4-98
MOl23	B0555	20-3-98	20-4-98
MOSSO	BOOO 1	21-3-98	21-4-98

Table 5 - Issue Table

Third Normal Form

- The table shown below is the Book Table.

Book Code	Category-Code
BO020	SCIENCE
BO189	ASTRO
BO090	ASTRO
BOO91	ASTRO
B0656	ROMANC
BO198	ADVENT
B0212	ADVENT
BO4PO	ADVENT
B0555	ADVENT
B0001	CLASSIC

Table 6 – Book Table

KEY CONCEPTS

- Relation, table, entity
- Tuple, row, record
- Attribute, column, field
- Correlation between the formal names of Tables, Rows, and Columns in Relational Theory and their more common counterparts:

Formal Name	*Common Name*	*Also Known As*
Relation	**Table**	**Entity**
Tuple	**Row**	**Record**
Attribute	**Column**	**Field**

SIX PROPERTIES OF A RELATIONAL TABLE

Property 1: Entries in columns are single-valued.

Property 2: Entries in columns are of the same kind.

Property 3: Each row is unique.

Property 4: Sequence of columns is insignificant.

Property 5: Sequence of rows is insignificant.

Property 6: Each column has a unique name.