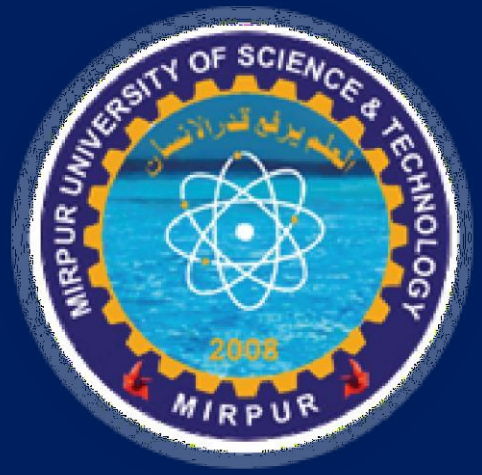


MUST

Wisdom & Virtue

MIRPUR UNIVERSITY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF SOFTWARE ENGINEERING

Business Intelligence

(Lecture # 3)
Role of Normalization in Business

Engr. Saman Fatima
(Lecturer)

Normalization

- The process of producing a simpler and more reliable database structure is called normalization.
- It is used to create a suitable set of relations for storing data.
- This process work through several stages known as normal forms.
- Each normal form has certain requirements and conditions.
- A large database defined as a single relation may result in data duplication. This repetition of data may result in:
- Making relations very large.
- It isn't easy to maintain and update data as it would involve searching many records in relation.
- Wastage and poor utilization of disk space and resources.
- The likelihood of errors and inconsistencies increases.

CONT!!!

1. So to handle these problems, we should analyze and decompose the relations with redundant data into smaller, simpler, and well-structured relations that are satisfy desirable properties.
2. Normalization is a process of decomposing the relations into relations with fewer attributes.

What is Normalization?

- Normalization is the process of organizing the data in the database.
- Normalization is used to minimize the redundancy from a relation or set of relations. It is also used to eliminate undesirable characteristics like Insertion, Update, and Deletion Anomalies.
- Normalization divides the larger table into smaller and links them using relationships.

CONT!!!

- The normal form is used to reduce redundancy from the database table.
- Why do we need Normalization?
- The main reason for normalizing the relations is removing these anomalies.
- Failure to eliminate anomalies leads to data redundancy and can cause data integrity and other problems as the database grows.
- Normalization consists of a series of guidelines that helps to guide you in creating a good database structure.

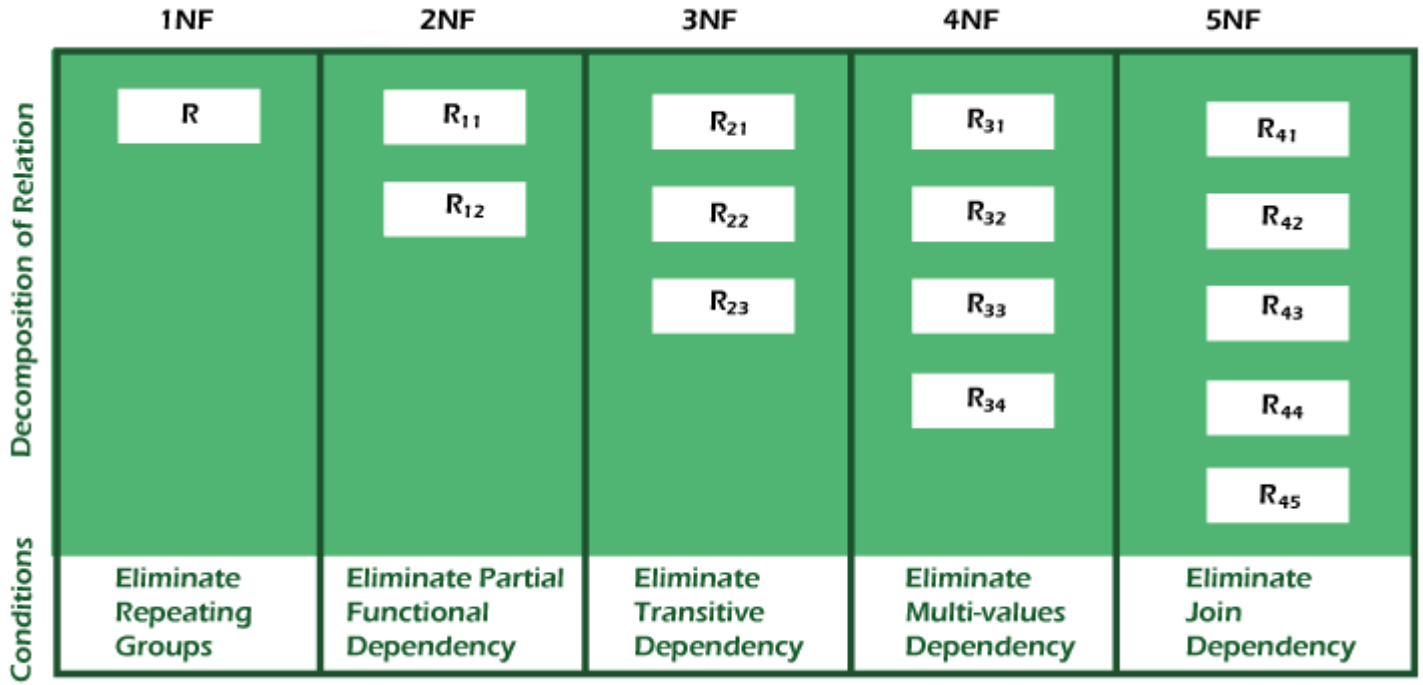
Data modification anomalies can be categorized into three types:

Data modification anomalies can be categorized into three types:

- **Insertion Anomaly:** Insertion Anomaly refers to when one cannot insert a new tuple into a relationship due to lack of data.
- **Deletion Anomaly:** The delete anomaly refers to the situation where the deletion of data results in the unintended loss of some other important data.
- **Updatation Anomaly:** The update anomaly is when an update of a single data value requires multiple rows of data to be updated.

Types of Normal Forms:

- Normalization works through a series of stages called Normal forms. The normal forms apply to individual relations. The relation is said to be in particular normal form if it satisfies constraints.
- **Following are the various types of Normal forms:**



Normal Form

Normal Form	Description
1NF	A relation is in 1NF if it contains an atomic value.
2NF	A relation will be in 2NF if it is in 1NF and all non-key attributes are fully functional dependent on the primary key.
3NF	A relation will be in 3NF if it is in 2NF and no transition dependency exists.
BCNF	A stronger definition of 3NF is known as Boyce Codd's normal form.
4NF	A relation will be in 4NF if it is in Boyce Codd's normal form and has no multi-valued dependency.
5NF	A relation is in 5NF. If it is in 4NF and does not contain any join dependency, joining should be lossless.

Advantages of Normalization

- Normalization helps to minimize data redundancy.
- Greater overall database organization.
- Data consistency within the database.
- Much more flexible database design.
- Enforces the concept of relational integrity.

Disadvantages of Normalization

- You cannot start building the database before knowing what the user needs.
- The performance degrades when normalizing the relations to higher normal forms, i.e., 4NF, 5NF.
- It is very time-consuming and difficult to normalize relations of a higher degree.
- Careless decomposition may lead to a bad database design, leading to serious problems.
- It states that an attribute of a table cannot hold multiple values. It must hold only single-valued attribute.
- First normal form disallows the multi-valued attribute, composite attribute, and their combinations.

Example:

1. **Example:** Relation EMPLOYEE is not in 1NF because of multi-valued attribute EMP_PHONE.

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385, 9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389, 8589830302	Punjab

The decomposition of the EMPLOYEE table into 1NF has been shown below:

CONT!!!

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385	UP
14	John	9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389	Punjab
12	Sam	8589830302	Punjab

2ND Example

Roll#	name	Course
1	FIZA	C/C++
2	SANA	JAVA
3	MUSA	C/DBMSM

Second Normal Form (2NF)

1. Second Normal Form (2NF)

- In the 2NF, relational must be in 1NF.
- In the second normal form, all non-key attributes are fully functional dependent on the primary key
- **Example:** Let's assume, a school can store the data of teachers and the subjects they teach. In a school, a teacher can teach more than one subject.

1. TEACHER table

TEACHER_ID	SUBJECT	TEACHER_AGE
25	Chemistry	30
25	Biology	30
47	English	35
83	Math	38
83	Computer	38

Second Normal Form (2NF)

In the given table, non-prime attribute TEACHER_AGE is dependent on TEACHER_ID which is a proper subset of a candidate key.
That's why it violates the rule for 2NF.
To convert the given table into 2NF, we decompose it into two tables:

TEACHER_ID	TEACHER_AGE
25	30
47	35
83	38

:

TEACHER_ID	SUBJECT
25	Chemistry
25	Biology
47	English
83	Math
83	Computer

2nd example

Pro-no	Emp-no	Proj-name	Emp-name	Job-class	Charge-per-hr	hours

Proj-no	Emp-no

Proj-no	Proj-name

Emp-no	Emp-name	Job-class	Charge-per-hr

Proj-no	Emp-name	hour

Third Normal Form (3NF)

- Third Normal Form (3NF)
- A relation will be in 3NF if it is in 2NF and not contain any transitive partial dependency.
- 3NF is used to reduce the data duplication. It is also used to achieve the data integrity.
- If there is no transitive dependency for non-prime attributes, then the relation must be in third normal form.
- What is Transitive Dependency
- When an indirect relationship causes functional dependency it is called Transitive Dependency.
- If P -> Q and Q -> R is true, then P-> R is a transitive dependency.
- To achieve 3NF, eliminate the Transitive Dependency.

3NF

Emp-no	Emp-name	Job-class	Charge-per-hr

Emp-no	Emmp-name	Job-class

Job-class	Charge-per-hr

Proj-no	Emp-no	hour

BCNF(3.5 NF)

- BCNF(3.5 NF)
- BCNF is the advance version of 3rd normal form
- BCNF IS ALSO KNOWN AS 3.5 Normal form
- To satisfy this normal form a table should be in 3NF
- To satisfy this normal form FD X--→Y so X should be super key.

Emp-Id	Emp-country	Emp-Dept	Dept-Type	Dept-No
10	Pakistan	Designing	D1	55
20	Pakistan	Testing	D2	58
30	Uk	Development	D3	52
40	Uk	QA	D4	85

CONT!!!

- In this table no transitive dependency
- EMP-ID-→EMP-COUNTRY
- EMP-DEPT-→DEPT-TYPE,DEPT-NO
- Table is not in BCNF because EMP-ID or EMP-Dept are not keys alone.
- 1st Table

Emp-Id	Emp-country
10	Pakistan
20	Pakistan
30	Uk
40	Uk

CONT!!!

- EMP-ID-→EMP-COUNTRY
- X-→Y and X is the super Key
- 2nd table

Emp-Dept	Dept-Type	Dept-No
Designing	D1	55
Testing	D2	58
Development	D3	52
QA	D4	85

CONT!!!

- EMP-DEPT-→DEPT-TYPE,DEPT-NO
- Mapping table
- Link table
- 3rd table
-

Emp-Id	Dept-No
10	55
20	58
30	52
40	85

(4 NF)

4th NF

Remove multivalued attributes

Name	mobile	Social network
Zaheer	Android/iphone	Twitter/facebook
jamal	Window/black berry	Insta/snapchat

(4NF)

A table with MVD

Name --→--→mobile

Name___>-→social network

Making 4th Normal form

Name	mobile	Social network
Zaheer	Android	Twitter
Zaheer	Iphone	Facebook
Zaheer	Iphone	Twitter
Zaheer	Android	Facebook
Jamal	Window	Insta
Jamal	Blackberry	Snapchat
Jamal	Window	Snapchat
jamal	blackberry	insta

4th NF

Name	Mobile
Zaheer	Android
Zaheer	Iphone
Jamal	Black berry
Jamal	window

name	Social network
Zaheer	Twitter
Zaheer	Facebook
Jamal	Insta
jamal	Snapchat

THANKS