



MANCHESTER
1824

The University of Manchester

NORMALISATION

COMP23111 – Database Systems

OUTLINE

Normalisation – Definition

Normalisation – Advantages

First Normal Form

Second Normal Form

Third Normal Form

NORMALISATION

– DEFINITION

Normalisation is the process of **organizing** the data in the database, with a focus on **minimizing redundancy** in relations or set of relations.

FDs help identify redundancy in relations and their attributes.

Edgar F. Codd (19 August 1923 – 18 April 2003) was the computer scientist who invented the relational database model and normalisation (1970), while working for IBM.

Codd won the Turing Award in 1981 awarded by the Association for Computing Machinery.

NORMALISATION

– DEFINITION

6 Normal Forms (NFs) available.

We will focus on the first 3 (1NF, 2NF, 3NF) - most commonly used as they address the *3 data modification anomalies*:

1. **Insertion Anomaly** (**Omission** of data during insertion due to absence of other data.)
2. **Deletion Anomaly** (**Unintended loss** of data due to the deletion of other data.)
3. **Update Anomaly** (**Data inconsistency** that results from redundant data and partial updates.)

NORMALISATION

– DEFINITION

1. Insertion Anomaly (**Omission** of data during insertion due to absence of other data.)

2. Deletion Anomaly (**Unintended loss** of data due to the deletion of other data.)

3. Update Anomaly (**Data inconsistency** that results from redundant data and partial updates.)

Employee_ID	Name	Department	Scientific Group
1	Ann Smiths	Computer Science	AI
2	Mata Dobrevik	Psychology	CBT
2	Mata Dobrevik	Psychology	CBT
3	Jose Jimenez	Maths	Algorithms
3	Jose Jimenez	Maths	Algorithms

NORMALISATION – ADVANTAGES

1. Prevents data redundancy.
2. Offers data consistency.
3. Offers flexibility in design and data management.
4. Enforces relational (referential) integrity.

FIRST NORMAL FORM (1NF)

A table is in **1NF** if:

1. It contains **no repeating** groups.
2. The **values** in each column are **atomic** (cannot be divided so not multi-valued or composite attributes).

Customer_ID	Name	Tel. number
2812	Mary Smiths	7786554389, 7877346324
9322	John Caroll	7877346312
2324	Anna Ivanova	7823252311
7890	Pablo Fernadez	7765321561

Customer_ID	Name	Tel. number
2812	Mary Smiths	7877346324
2812	Mary Smiths	7786554389
9322	John Caroll	7877346312
2324	Anna Ivanova	7823252311
7890	Pablo Fernadez	7765321561

SECOND NORMAL FORM (2NF)

A table is in **2NF** if:

1. It is in 1NF

2. There are **no partial dependencies** (every non-key attribute is functionally dependent on the PK).

PD: if a proper subset of the PK determines a non-key attribute.

Employee_ID	Training Course	Fee
1	TC2	1200
2	TC1	1000
2	TC3	1500
3	TC4	1500
3	TC1	1000

Employee_ID	Training Course
1	TC2
2	TC1
2	TC3
3	TC4
3	TC1

Training Course	Fee
TC1	1000
TC2	1200
TC3	1500
TC4	1500

THIRD NORMAL FORM (3NF)

A table is in **3NF** if:

1. It is in 2NF
2. There are **no transitive dependencies** for non-key attributes.

TD: A, B and C are attributes of a relation such that if $A \rightarrow B$ and $B \rightarrow C$, then $A \rightarrow C$.

<u>Employee_ID</u>	Name	Birth_State	Birth_Country
1	Ann Smiths	Ohio	US
2	Mata Dobrevik	Ohio	US
3	Jose Jimenez	Texas	US

Employee_ID	Name	Birth_State
1	Ann Smiths	Ohio
2	Mata Dobrevik	Ohio
3	Jose Jimenez	Texas

State	Country
Ohio	US
Texas	US

$A \rightarrow B$ $C \rightarrow D$
 $A \rightarrow C$ $A \rightarrow D$