Normalization

The Raw Database

A database that is not normalized may include data that is **contained in one or more different tables for no apparent reason**. This could be bad for **security reasons**, **disk space usage, speed of queries**, **efficiency of database updates**, and, maybe most importantly, **data integrity**.

Normalizing a Database

Normalization is a process by which data structures in a relational database are as efficient as possible, including the **elimination of redundancy**, the **minimization of the use of null values** and the **prevention of the loss** of information. The actual guidelines of normalization, called **Normal Forms**.

Aims of Normalization:

- **Greater overall database** organization

• Normalization ensures that the database is **structured** in the **best possible way**.

• To achieve control over **data redundancy**. There should

be **no unnecessary duplication** of data in different tables.

• To **ensure data consistency**.

• To allow data in different tables can be used in **complex queries**.

Candidate Key

• **A Candidate Key** is an attribute (possibly composite) that can be used to uniquely identify each tuple in a relation.

• A relation may have **more than one candidate key**

• If so, one candidate key is nominated as the **primary key**

Stages of Normalization (Normal Forms)

- **First Normal Form (1NF)**
- **Second Normal Form (2NF)**
- **Third Normal Form (3NF)**
- **Boyce-Codd Normal Form (BCNF)**
- **Fourth Normal Form (4NF)**
- **Fifth Normal Form (5NF)**

First Normal Form (1NF)

- The official qualifications for 1NF are:
- Each **attribute name** must be **unique**.
- Each **attribute value** must be **single**.
- Each **row must** be **unique**.
- There is **no repeating groups**.
- **Additional**:
    - Choose a **primary key**.
- **Reminder**:
    - A primary key **is unique, not null, unchanged.**
    - A primary key **can be either an attribute or combined attributes.**

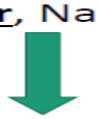
- **Non-Normalized Table: (Student Table)**

Number	Name	Classes
001231	William Hartnell	Information Systems, Systems Analysis, Data Communications
001232	Patrick Troughton	Systems Analysis, Data Communications
001233	Jon Pertwee	OO Programming, Systems Analysis, Data Communications
001234	Tom Baker	Systems Analysis, Data Communications

- To convert data for **unnormalized form to 1NF,**

simply **convert any repeated attributes** into part of the candidate key:

- STUDENT(**Number**, Name, Classes)



- STUDENT(**Number**, Name, **Classes**)

Number	Name	Classes
001231	William Hartnell	Information Systems
001231	William Hartnell	Systems Analysis
001231	William Hartnell	Data Communications
001232	Patrick Troughton	Systems Analysis
001232	Patrick Troughton	Data Communications
001233	Jon Pertwee	OO Programming
001233	Jon Pertwee	Systems Analysis
001233	Jon Pertwee	Data Communications
001234	Tom Baker	Systems Analysis
001234	Tom Baker	Data Communications

Second Normal Form(2NF)

• A table is in the second normal form **if:**

1. It's in the **first normal form**

2**. No column** that is **not part of the primary key** is dependent **only** a **portion** of the **primary key.**

* The official qualifications for 2NF are:

- A table is already in **1NF**.
- **All non-key attributes** are **fully dependent** on the primary key.
- **All partial dependencies** are removed to place in another table.

RefNo	Name	Address	Status	AccNo
345	C.J. Date	23, High Street	Business	120768
345	C.J. Date	23, High Street	Business	348973
543	F.D. Rolland	45, The Ash	Domestic	987654
675	D.R. Howe	17, Low Street	Business	745363
675	D.R. Howe	17, Low Street	Business	678453
675	D.R. Howe	17, Low Street	Business	348973

Functional Dependency

- It is clear that :

RefNo -> Name, Address, Status

or, most correctly,

AccNo, RefNo -> Name, Address, Status

RefNo	Name	Address	Status	AccNo
345	C.J. Date	23, High Street	Business	120768
345	C.J. Date	23, High Street	Business	348973
543	F.D. Rolland	45, The Ash	Domestic	987654
675	D.R. Howe	17, Low Street	Business	745363
675	D.R. Howe	17, Low Street	Business	678453
675	D.R. Howe	17, Low Street	Business	348973

RefNo	AccNo
345	120768
345	348973
543	987654
675	745363
675	678453
675	348973

RefNo	Name	Address	Status
345	C.J. Date	23, High Street	Business
543	F.D. Rolland	45, The Ash	Domestic
675	D.R. Howe	17, Low Street	Business

Third Normal Form(3NF)

• A table is in the third normal form if:

1. **It is the second normal form.**

2. There are **no non-key columns dependent on other non-key columns that could not act as the primary key.**

* The official qualifications for 3NF are:

  • A table is already in **2NF**.

  • Nonprimary key attributes **do not depend on other nonprimary key attributes**. (i.e**. no transitive dependencies**)

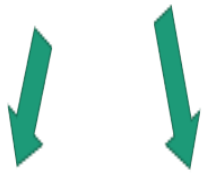
  • **All transitive dependencies are removed to place in another table.**

Branch Name	Address	Manager No	Acc No	Balance	Type
Rathmines	15 Upr Rathmimes Road	1234	1205	-£123.45	'D'
Rathmines	15 Upr Rathmimes Road	1234	6784	£67.54	'C'
Rathmines	15 Upr Rathmimes Road	1234	9843	£43.43	'C'
Dame St.	1 Dame Street	1101	5422	£34.50	'C'
Dame St.	1 Dame Street	1101	0998	£666.66	'D'

Functional Dependency:

BranchName -> Address, ManagerNo
AccNO -> Balance, Type, BranchName, Address, ManagerNo

Branch Name	Address	Manager No	Acc No	Balance	Type
Rathmines	15 Upr Rathmimes Road	1234	1205	-£123.45	'D'
Rathmines	15 Upr Rathmimes Road	1234	6784	£67.54	'C'
Rathmines	15 Upr Rathmimes Road	1234	9843	£43.43	'C'
Dame St.	1 Dame Street	1101	5422	£34.50	'C'
Dame St.	1 Dame Street	1101	0998	£666.66	'D'



Acc No	Balance	Type	Branch Name
1205	-£123.45	'D'	Rathmines
6784	£67.54	'C'	Rathmines
9843	£43.43	'C'	Rathmines
5422	£34.50	'C'	Dame St.
0998	£666.66	'D'	Dame St.

Branch Name	Address	Manager No
Rathmines	15 Upr Rathmimes Road	1234
Dame St.	1 Dame Street	1101

Thanks