NORMALIZATION

Normalization is a database design technique

- We have achieved normalization already by using **ERD modelling.**
- However, we can also approach Normalization **without** having to do any modelling

- These are the main stages of normalization

Stage of Normalization	Abbreviation	Description
First normal form	1NF	Remove Repeating groups
Second normal form	2NF	Remove Part Key Dependencies
Third normal form	3NF	Remove Non Key Dependencies

There are also some additional but seldom used stages of normalization

Stage of Normalization	Abbreviation	Description
Boyce-Codd normal form	BCNF	Remove Non Key Determinants
Fourth normal form	4NF	Remove Multivalued Dependencies
Fifth normal form	5NF	Remove Join Dependencies

Zero Normal Form (0NF)

This is an example of an **unnormalized** table (0NF)
- There are multiple values in a single cell.
- These are known as **repeating groups**.

StuID	Student Name	Major	SubCode	Subject Title	Convenor	Convenor Location	Grade
38214	Bright	IS	IS350	Database	Codd	B104	HD
			IS465	SysAnalysis	Kemp	B213	D
69173	Smith	PM	IS467	SysAnalysis	Kemp	B213	N
			PM300	Prod Mgt	Lewis	D317	D
			QM440	Op Res	Kemp	B213	C

To achieve 1NF, ensure no cell has multiple values

First Normal Form (1NF)

This relation is **First Normal Form** - 1NF.

This relation is also known as **The Universal Relation.**

All **repeating groups** have been **removed**.

- All data begins as **one** large relation (or table)

- Each cell in a row has a maximum of 1 value

StuId	Student Name	Course	SubCode	SubjectTitle	Convenor Name	Convenor Location	Grade
38214	BRIGHT	BIS	IS350	Database Mgt	CODD	B104	HD
38214	BRIGHT	BIS	IS465	Bus Intelligence	KEMP	B213	D
69173	SMITH	BCom	IS465	Bus Intelligence	KEMP	B213	N
69173	SMITH	BCom	PM300	Project Mgt	LEWIS	D317	D
69173	SMITH	BCom	QM440	Statistics	KEMP	B213	C

Identification of Primary Key

This step is often the most difficult.

- Identify the Primary Key of this **Universal Relation**?
- The PK **must** be determined at the outset and **before** normalization can continue.
- The more columns you have, the **harder** this step becomes.

Stuld	Student Name	Course	SubCode	SubjectTitle	Convenor Name	Convenor Location	Grade
38214	BRIGHT	BIS	IS350	Database Mgt	CODD	B104	HD
38214	BRIGHT	BIS	IS465	Bus Intelligence	KEMP	B213	D
69173	SMITH	BCom	IS465	Bus Intelligence	KEMP	B213	N
69173	SMITH	BCom	PM300	Project Mgt	LEWIS	D317	D
69173	SMITH	BCom	QM440	Statistics	KEMP	B213	C

What is the Primary Key of the relation?
- *Use the fewest columns possible*

Identification of Primary Key

StuId and SubCode uniquely identify each row

- *While **additional** columns may also uniquely identify every row, you must **reduce** to number of columns to the **fewest possible***

- The key StuId & SubCode are known as the **Universal Key**

StuId	Student Name	Course	SubCode	SubjectTitle	Convenor Name	Convenor Location	Grade
38214	BRIGHT	BIS	IS350	Database Mgt	CODD	B104	HD
38214	BRIGHT	BIS	IS465	Bus Intelligence	KEMP	B213	D
69173	SMITH	BCom	IS465	Bus Intelligence	KEMP	B213	N
69173	SMITH	BCom	PM300	Project Mgt	LEWIS	D317	D
69173	SMITH	BCom	QM440	Statistics	KEMP	B213	C

Second Normal Form (2NF)

To achieve 2nd Normal Form **remove functional dependencies.**

What is a **functional dependency?**

- An FD is where one attribute can be **determined** by another attribute.

    e.g. StuId **determines** the Studentname
        StuId **identifies** Studentname
        StuId is the **key** for Studentname

If in doubt, ask yourself:
-   "Is there only ONE student name for that StuId ?"
-   If your answer is "Yes" then there is a functional dependency

Remove PART-KEY dependencies (1NF -> 2NF)

The **Universal Key** of this Universal Relation is <u>StuId, SubCode</u>

- • A Part Key dependency is where a column in the Universal Relation is dependent on **only part** of the Universal Key.

- In this example the Universal Key is made up of **two** columns.
  StuId & SubCode

- Is there any column that is dependent on **only StuId** ?
  Yes
  - • **Studentname** is **dependent** on StuId
    (Studentname has no dependency with SubCode)

This is called a **Part-Key Dependency because**
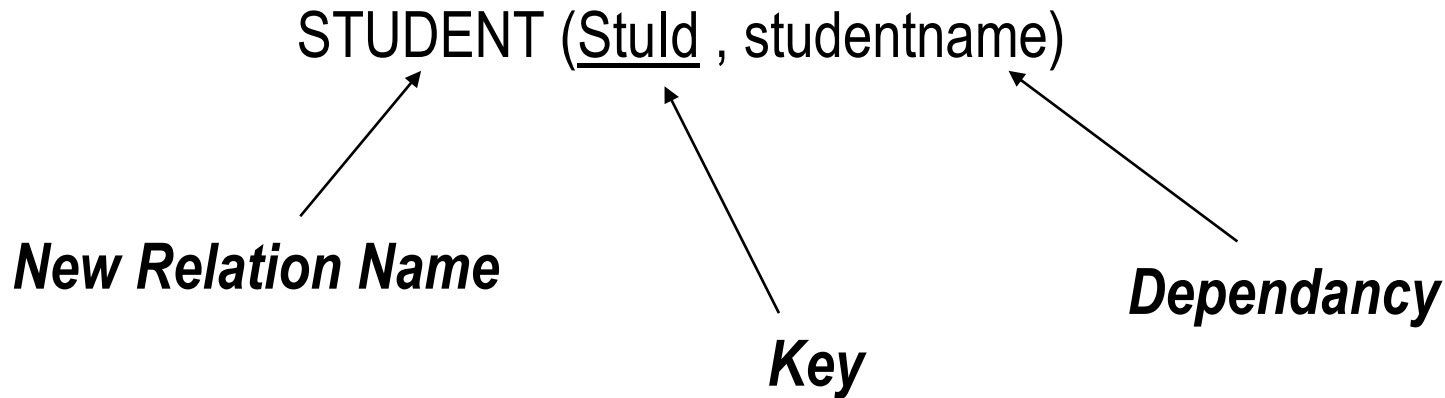- • **Studentname** is dependent on **part** of the universal key

Remove PART-KEY dependencies (1NF -> 2NF)

Remove the **part-key** dependency

This requires you to create a new relation (table) that will store
  - The part-key value (called the determinate)
  - The dependency
  - You need to create a name for this new relation (table)

STUDENT (StuId , studentname)

New Relation Name

Key

Dependancy

Remove PART-KEY dependencies (1NF -> 2NF)

Are there any other **dependencies** based on StuId. Yes
- **Course** is also dependent on StuId

- Add the dependant column the to Student relation (table)

STUDENT (StuId , studentname, course)

What have we done?

Using an ERD perspective, we have created a **'strong entity'** type relation from the original Universal Relation

Remove PART-KEY dependencies (1NF -> 2NF)

What have we done?

Using an ERD perspective, we have created a '**strong entity**' type relation from the original Universal Relation

Universal Relation

StuId	SubCode	Subject Title	Convenor Name	Convenor Locaton	Grade
38214	IS350	DATABASE	CODD	B104	HD
38214	IS465	SysAnalysis	KEMP	B213	D
69173	IS465	SysAnalysis	KEMP	B213	N
69173	PM300	PROD MGT	LEWIS	D317	D
69173	QM440	OP RES	KEMP	B213	C

Student Relation

StuId	Student Name	Course
38214	BRIGHT	IS
69173	SMITH	PM

- Note. The column StuId exists in <u>both</u> relations.
- The StuId is the Universal Relation is a Foreign Key value that references Student

Remove PART-KEY dependencies

- Now consider the other column of the Universal Key
- Is there apart key dependencies based on SubCode?   Yes

SubCode   determines / identifies / is the key of   SubjectTitle

If fact SubCode also determines / identifies / is the key of
- ConvenorName
- ConvenorLocation

So, create a new relation named subject

SUBJECT (SubCode ,SubjectTitle, Convenorname, Convenorlocation)

Remove PART-KEY dependencies

All Part-Key dependencies have been **removed** from the Universal Relation
Note. StuId & SubCode in the UR are FKs referring to 'strong' relations.
We have achieved 2NF (second normal form)

Universal Relation

StuId	SubCode	Grade
38214	IS350	HD
38214	IS465	D
69173	IS465	N
69173	PM300	D
69173	QM440	C

Subject Relation

SubCode	Subject Title	Convenor Name	Convenor Location
IS350	DATABASE	CODD	B104
IS465	SysAnalysis	KEMP	B213
PM300	PROD MGT	LEWIS	D317
QM440	OP RES	KEMP	B213

Student Relation

StuId	Student Name	Course
38214	BRIGHT	BIS
69173	SMITH	Bcom

Foreign Keys

Removal of Non-Key Dependencies (2NF -> 3NF)

The next step is to remove any Non-Key dependencies from the newly created tables.

- This is where:
  - A column has a **dependency** on another column
  - But that other column **is NOT part of the Universal Key**

Consider the Subject Relation.

- Are there any attributes in the tables that are **not dependant** on the Key?

Convenor Location:

- Is **not** dependent on SubCode
- **Is** dependant on Convenor Name

Subject Relation

SubCode	Subject Title	Convenor Name	Convenor Location
IS350	DATABASE	CODD	B104
IS465	SysAnalysis	KEMP	B213
PM300	PROD MGT	LEWIS	D317
QM440	OP RES	KEMP	B213

SWIN BUR NE
SWINBURNE UNIVERSITY OF TECHNOLOGY

Removal of Non-Key Dependencies

- Create a new table named Convenor

- Place the key and any **dependencies into this new relation (table)**

CONVENOR(<u>ConvenorName,</u> ConvenorLocation)

By Removing all **Non-Key** dependencies
- **We have moved from 2NF to 3NF**

The database is now NORMALIZED

Removal of Non-Key Dependencies

Removal of NonKey Dependencies

2NF → **3NF**

Note:

The column ConvenorName remains in <u>both</u> relations

<u>SubCode</u>	Subject Title	Convenor Name
IS350	DATABASE	CODD
IS465	SysAnalysis	KEMP
PM300	PROD MGT	LEWIS
QM440	OP RES	KEMP

Foreign Key

<u>CONVENORNAME</u>	CONVENOR LOCATION
CODD	B104
KEMP	B213
LEWIS	D317

Third Normal Form   3NF

Recap: Remove any part key dependency based on StuId

StuId	StuId	Student Name	Course	SubCode	SubjectTitle	Convenor Name	Convenor Location	Grade
38214	38214	BRIGHT	BIS	IS350	Database Mgt	CODD	B104	HD
38214	38214	BRIGHT	BIS	IS465	Bus Intelligence	KEMP	B213	D
69173	69173	SMITH	BCom	IS465	Bus Intelligence	KEMP	B213	N
69173	69173	SMITH	BCom	PM300	Project Mgt	LEWIS	D317	D
69173	69173	SMITH	BCom	QM440	Statistics	KEMP	B213	C

Third Normal Form   3NF

Recap: Remove part key dependencies based on SubCode

StuId	SubCode	SubjectTitle	Convenor Name	Convenor Location	Grade
38214	IS350	Database Mgt	CODD	B104	HD
38214	IS465	Bus Intelligence	KEMP	B213	D
69173	IS465	Bus Intelligence	KEMP	B213	N
69173	PM300	Project Mgt	LEWIS	D317	D
69173	QM440	Statistics	KEMP	B213	C

StuId	Student Name	Course
38214	BRIGHT	BIS
38214	BRIGHT	BIS
69173	SMITH	BCom
69173	SMITH	BCom
69173	SMITH	BCom

Third Normal Form   3NF

Recap: Remove non key dependency on convenor name

StuId	SubCode	Grade
38214	IS350	HD
38214	IS465	D
69173	IS465	N
69173	PM300	D
69173	QM440	C

SubCode	SubjectTitle	Convenor Name	Convenor Location
IS350	Database Mgt	CODD	B104
IS465	Bus Intelligence	KEMP	B213
IS465	Bus Intelligence	KEMP	B213
PM300	Project Mgt	LEWIS	D317
QM440	Statistics	KEMP	B213

StuId	Student Name	Course
38214	BRIGHT	BIS
38214	BRIGHT	BIS
69173	SMITH	BCom
69173	SMITH	BCom
69173	SMITH	BCom

Third Normal Form   3NF

All relations are now in 3NF

StuId	SubCode	Grade
38214	IS350	HD
38214	IS465	D
69173	IS465	N
69173	PM300	D
69173	QM440	C

Convenor Name	CONVENOR LOCATION
CODD	B104
KEMP	B213
LEWIS	D317

StuId	Student Name	Course
38214	BRIGHT	BIS
69173	SMITH	Bcom

SubCode	Subject Title	Convenor Name
IS350	DATABASE	CODD
IS465	SysAnalysis	KEMP
PM300	PROD MGT	LEWIS
QM440	OP RES	KEMP

- - - Foreign Keys