

# Database Normalization

Slides are from below resources:

[https://en.wikipedia.org/wiki/Database\\_normalization](https://en.wikipedia.org/wiki/Database_normalization)

<https://learn.microsoft.com/en-us/office/troubleshoot/access/database-normalization-description>

# Reviewing database normalization

The process used to organise the database into tables in a way that eliminates redundant data and improves data integrity.

Reasons to normalise data:

- Reduce redundant data points
- Reduce quality issues
- Simplify your queries

# Deals with prevention of anomalies

- Insertion anomaly
- Deletion anomaly
- Update anomaly

# Deletion anomaly

- If I delete the orange row, I am going to loss Art.

Student #	Unit #	Student name	Address	Unit
2002001	100	Adam	Perth	Cyber Security
2002001	101	Adam	Perth	Cyber Security
2002002	101	Mark	NSW	Data analysis
2002003	100	Ben	VIC	Art
2002003	103	Ben	VIC	Physics

# Insertion anomaly

- If I want add new Unit for Ex database design. Other columns are null. That means I don't have information for other columns.

Student #	Unit #	Student name	Address	Unit
2002001	100	Adam	Perth	Cyber Security
2002001	101	Adam	Perth	Cyber Security
2002002	101	Mark	NSW	Data analysis
2002003	100	Ben	VIC	Art
2002003	103	Ben	VIC	Physics

# Update anomaly

- I want change the address of Student 2002001. So, I would have to find, because of this horribly design table, every instance of Adam and change their address.

Student #	Unit #	Student name	Address	Unit
2002001	100	Adam	Perth	Cyber Security
2002001	101	Adam	Perth	Cyber Security
2002002	101	Mark	NSW	Data analysis
2002003	100	Ben	VIC	Art
2002003	103	Ben	VIC	Physics

# Steps of Normalization

- Step 1: select the data source and convert it to an unnormalized table (UNF)
- Step 2: transform the unnormalized data into first normal form (1NF) : primary key and non-repeating fields
- Step 3: transform data in first normal form (1NF) into second normal form (2NF) : make sure our data is based on primary key
- Step 4: transform data in second normal form (2NF) into third normal form (3NF): all our data is based on primary key, No attribute in a table that do not depend on primary key.

# 1NF

- Have a primary key
- Each field should contain ONE value
- No repeating groups

Ex: when the company grows maybe one person contact increase as well.

Staff ID	Name	Address	Certificate ID	Staff ID	Certificate
2002001	Adam	Perth	1	2002001	Microsoft
2002002	Mark	NSW	2	2002002	A+
2002003	Ben	VIC	3	2002002	Excel
			4	2002003	B+
			5	2002003	Cyber
			6	2002003	Adv word

# 2NF

- Any non-primary key field should be dependent on entire primary key.
- Ex: two columns are PK for below table. Are non-primary key field should be dependent on entire primary key? Certificate price just depend on Certificate ID.

Staff ID	Certificate ID
2002001	Cert1
2002002	Cert2
2002001	Cert4
2002003	Cert3
2002003	Cert1
2002002	Cert5

Certificate ID	Certificate Price
Cert1	\$2000
Cert2	\$1000
Cert3	\$2000
Cert4	\$1500
Cert5	\$1500

# 3NF

- No non-key fields are dependent on another non-key field.
- Below table is on 1 and 2 NF but not 3NF. Why? Workshop ID is PK. occupancy is dependent on another non PK, Room. So, we should delete occupancy column.

Workshop ID	Certificate ID	Date	Room	occupancy
100	Cert1	03/04/25	A21	55
101	Cert2	01/11/25	B11	60
102	Cert1	04/8/25	S14	45
103	Cert3	03/04/25	A21	55
104	Cert3	16/06/25	B11	60