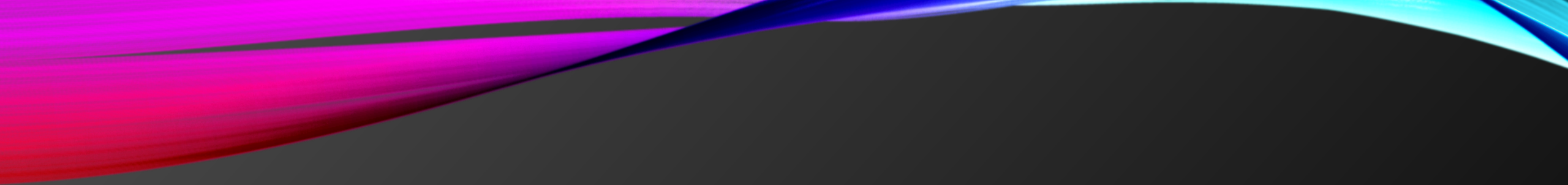




DATABASE DESIGN

Julie Allen 2024



DATABASE NORMALISATION

Can you remember some of the issues that are likely to arise from storing information like this?

The Sock Company

Name	Address	Date	Product	Qty	Price
Matthew Bates	17 Grasmere Street	12/02/21	Motorbike Socks	2	20.00
Jonathan Edwards	401 Ullswater Road	13/02/21	Unicorn Socks	1	12.50
Halima Kure	297 Tam Hows Street	14/02/21	Glittery Socks	1	15.99
John Hrycak	12 Windermere Cl	16/02/21	Superman Socks	1	8.50
Edwards, Jonathan	401 Ullswater Road	18/02/21	Fluffy Puppy Socks	2	50.00
Oliver Kerr	45 Rydal Rd	19/02/21	Superman Socks	3	25.50
Julie allen	96 Derwent Road	20/02/21	Ninja Turtles Socks	1	7.95
J. Edwards	401 Ullswater Road	27/02/21	Ninja Turtles Socks	1	7.95
C. Fasching	126 Ennerdale Avenue	01/03/21	Glittery Socks	1	15.99
John King	1 Palace Road	02/03/21	Tartan Socks	1	12.50
Jonathan Edwards	401 Ullswater Road	04/03/21	Fluffy Bunny Socks	3	75.00
Folayo Aina	123 Buttermere Street	05/03/21	Unicorn Socks	1	12.50
j. Edwards	401 Ullswater Road	08/03/21	Cute Cat Socks	1	15.99

Data Redundancy: the same piece of information is being stored in more than one place at a time.

Supplier No	Supplier Name	Supplier City	PNo	Part Descript	Date Shipped	Qty Ordered
S1	Jenkins	London	P1	Blue Pen	28 Jan	20
S1	Jenkins	London	P2	Pink Pen	28 Jan	200
S1	Jenkins	London	P3	Small KeyRing	17 Jan	300
S1	Jenkins	London	P4	Small Eraser	14 Jan	10
S2	Smith & Son	Preston	P5	Large KeyRing	1 Feb	300
S2	Smith & Son	Preston	P6	Novelty	1 Feb	50
S2	Smith & Son	Preston	P1	Blue Pen	14 Jan	200
S3	Jenkins	Preston	P2	Pink Pen	7 Jan	10
S4	Fisher	London	P2	Pink Pen	25 Jan	20
S4	Fisher	London	P4	Small Eraser	24 Jan	30

THE PROBLEM OF DATA REDUNDANCY

- Wasted storage
- Data Inconsistencies
- Poor performance
- Security risks
- Etc.
- Data anomalies:
 - Insert
 - Delete
 - Update

Insert Anomaly

The insertion of data is not possible without adding other unavailable data.

In the Sock example it is not possible to add a customer record until the customer has placed an order:

Name	Address	Date	Product	Qty	Price
Matthew Bates	17 Grasmere Street	12/02/21	Motorbike Socks	2	20.00
Jonathan Edwards	401 Ullswater Road	13/02/21	Unicorn Socks	1	12.50
Halima Kure	297 Tarn Hows Street	14/02/21	Glittery Socks	1	15.99
John Hrycak	12 Windermere Cl				

Delete Anomaly

Deleting data means we delete other data that we might want to keep.

If we delete the sale record for Matthew Bates we lose his name and address details and potentially the product details. This is because he has only placed one order:

Name	Address	Date	Product	Qty	Price
Matthew Bates	17 Grasmere Street	12/02/21	Motorbike Socks	2	20.00
Jonathan Edwards	401 Ullswater Road	13/02/21	Unicorn Socks	1	12.50
Halima Kure	297 Tam Hows Street	14/02/21	Glittery Socks	1	15.99
John Hrycak	12 Windermere Cl	16/02/21	Superman Socks	1	8.50

Update Anomaly

If the same data is stored in multiple places and we don't update all of them the data becomes inconsistent and loses its integrity.

If we update Jonathan Edwards address only on one record, we then have no idea which is his correct address:

Name	Address	Date	Product	Qty	Price
Jonathan Edwards	401 Ullswater Road	13/02/21	Unicorn Socks	1	12.50
Edwards, Jonathan	401 Ullswater Road	18/02/21	Fluffy Puppy Socks	2	50.00
J. Edwards	402 Ullswater Road	27/02/21	Ninja Turtles Socks	1	7.95
Jonathan Edwards	401 Ullswater Road	04/03/21	Fluffy Bunny Socks	3	75.00
j. Edwards	401 Ullswater Road	08/03/21	Cute Cat Socks	1	15.99

So, we saw that the answer is to arrange the data across several tables that are linked via primary and foreign keys...

Customer

CustomerID	Name	Address	Phone
101	Matthew Bates	17 Grasmere Street	0565780129
102	Jonathan Edwards	401 Ullswater Road	0899665490
103	Halima Kure	297 Tam Hows Street	0786520134
104	John Hrycak	12 Windermere Close	0678259018
106	Oliver Kerr	45 Rydal Road	0967821904
107	Chiara Fasching	126 Ennerdale Avenue	0698430289
109	Folayo Aina	123 Buttermere Street	0896733456
110	John King	1 Palace Road	0926781111
113	Julie Allen	96 Derwent Road	0769844457

Order

OrderID	Date	CustomerID	ProductID	Qty
1001	12/02/21	101	3245	2
1002	13/02/21	102	3249	1
1003	14/02/21	103	3248	1
1004	18/02/21	102	3251	2
1005	01/03/21	109	3246	1
1006	02/03/21	110	3245	1
1007	04/03/21	102	3250	3

Product

ProductID	Product	QtyInStock	Price
3245	Motorbike Socks	180	10.00
3246	Superman Socks	358	8.50
3247	Ninja Turtle Socks	21	7.95
3248	Glittery Socks	45	15.99
3249	Unicorn Socks	70	12.50
3250	Fluffy Bunny Socks	1000	15.99
3251	Fluffy Puppy Socks	290	15.99

It is a relatively easy process to design a database and to decide which entities/tables we will need...

Student

Student ID	Name
101	Zayne
102	Sarah
103	Michael

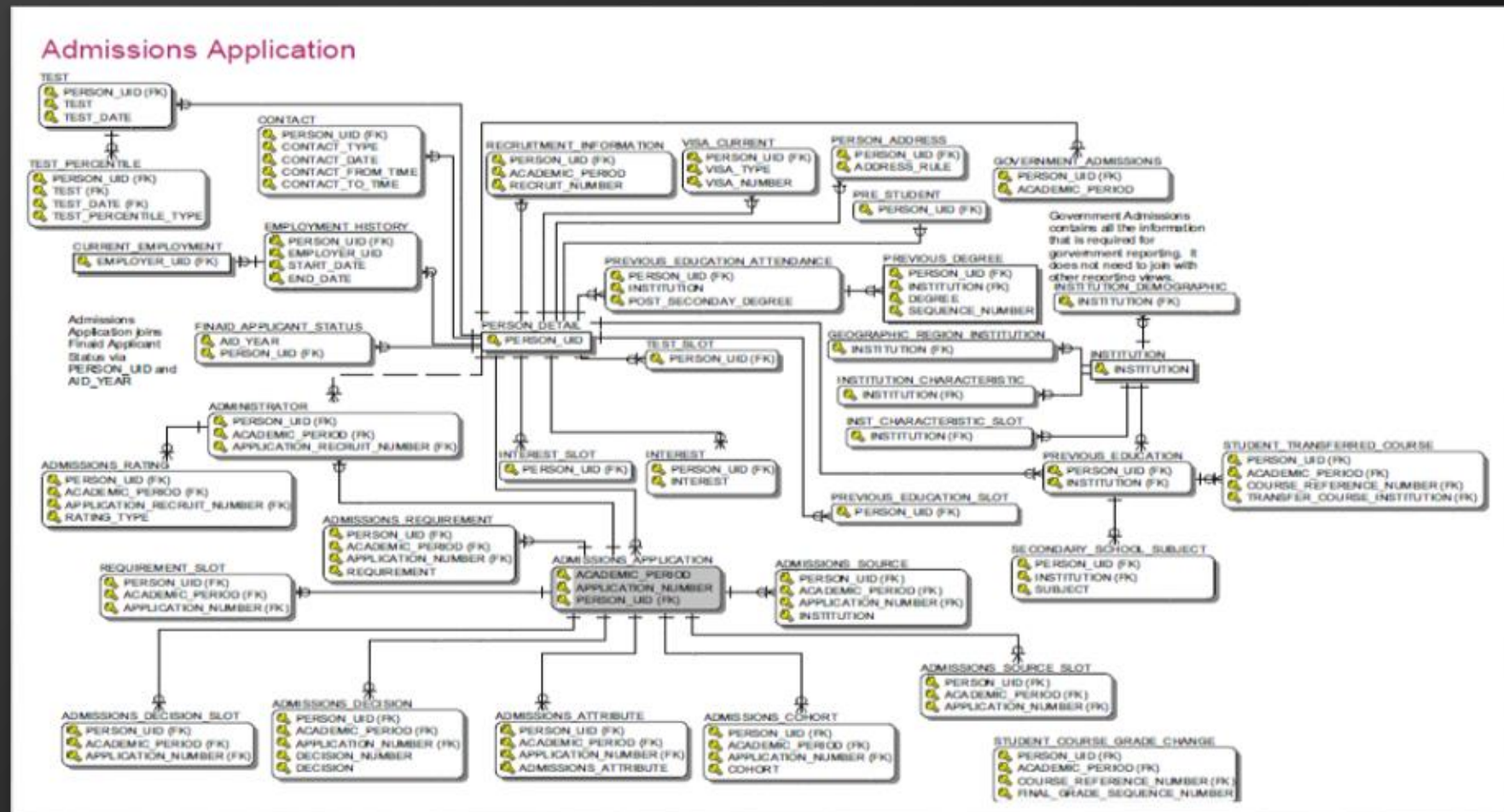
Student_Module

Student ID	Module ID
101	CO1409
101	CO1507
101	CO1301
102	CO1409
102	CO1301
Etc.	Etc.

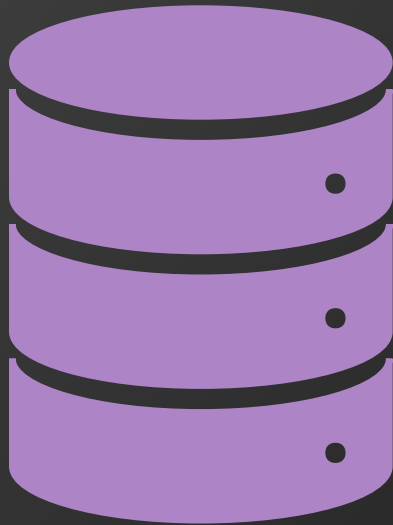
Module

Module ID	Name
CO1409	Programming
CO1605	Databases
CO1507	Networking
CO1301	Games

However, real systems can get very complex!



As such, there is a more formal database design process we can follow...



Database Normalisation is the process of structuring a relational database in accordance with a series of so-called normal forms in order to reduce data redundancy and improve data integrity.

OBJECTIVES OF NORMALISATION

- To put data into a form that conforms to relational principles; e.g. single valued columns; each relation represents one 'real world' entity, etc.
- To avoid redundancy by storing each 'fact' within the database only once.
- To put the data into a form that is more able to accommodate change.
- To avoid 'anomalies': insert, update, delete.
- To facilitate the enforcement of constraints on the data.

NORMALISATION PROCESS

NORMALISATION involves the checking that tables conform to certain rules, and if not, re-organising the data until it does. This requires making new tables to place the data drawn from the original tables.

NORMALISATION is a multi-stage process, of which the result of each is called a NORMAL FORM. Each successive stage produces a greater degree of normalisation.

There are a total of 7 normal forms

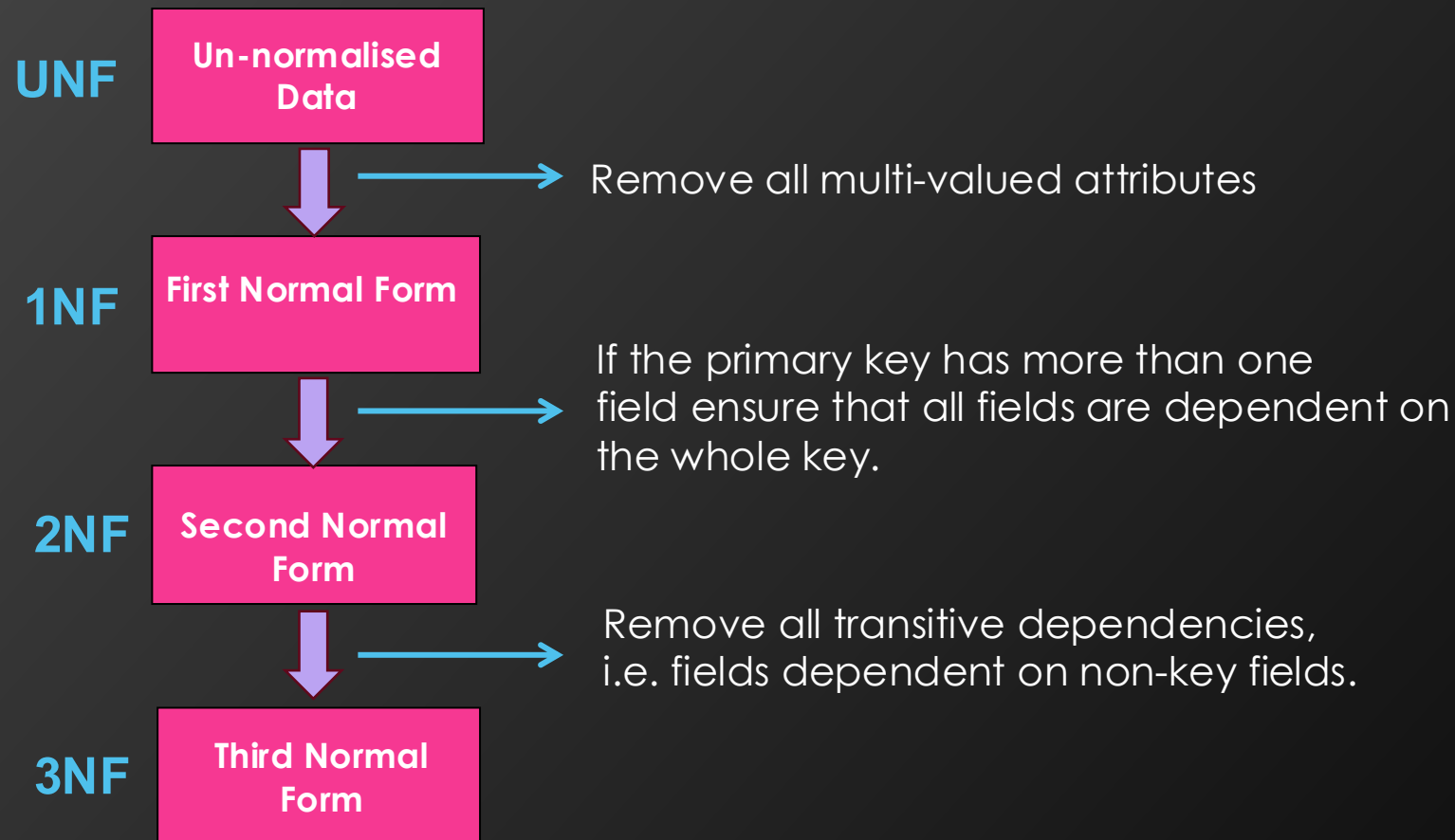
First, Second and Third Normal Forms (1NF, 2NF, 3NF)

Boyce-Codd Normal Form (BCNF)

Fourth Normal Form (4NF)

Fifth Normal Form (5NF) and Domain-Key Normal Form (DK/NF).

NORMALISATION SUMMARY



Normalisation proceeds in a series of stages.

Tests are applied at each stage and if the design does not agree with the conditions, it is modified.

THIS WEEK – FIRST NORMAL FORM

1NF Conditions:

1. There are no repeating groups
2. All data values are atomic
3. Each field attribute has a unique name
4. Each table has a primary key

1NF: EXAMPLE



Circus Skills – Sample Data

Trainee Number	Title	First Name	Surname	Date of Birth	Address Line 1	Town/City	County	Postcode	Tel No	Module Number	Module Name	Level	Module Duration (days)	Cost	Assessment Date	Result
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	250	Intermediate Static Trapeze	2	24	£50	21/02/2012	Distinction
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	101	Learning to balance	1	15	£35	10/07/2012	Distinction
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	102	Fire Eating for beginners	1	12	£30	11/07/2012	Distinction
1223	Miss	Clare	Croston	15/08/1998	15 Sure Street	Leyland	Lancashire	PR5 9GW	01772 678453	100	Learning to juggle	1	12	£25	14/07/2012	Pass
1222	Mrs	Lauren	Fazackerley	13/08/1978	21 Abbotts Road	London	Greater London	L56 7GH	0171 834 6337	200	Acrobatic Balancing	2	18	£48	22/07/2012	Pass
1223	Miss	Clare	Croston	15/08/1998	15 Sure Street	Leyland	Lancashire	PR5 9GW	01772 678453	100	Learning to juggle	1	12	£25	23/07/2012	Fail
1222	Mrs	Lauren	Fazackerley	13/08/1978	21 Abbotts Road	London	Greater London	L56 7GH	0171 834 6337	252	Intermediate fire eating	2	18	£45	29/07/2012	Merit
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	210	Flying Trapeze	2	20	£58	25/08/2012	Pass
1259	Miss	Patricia	Smithies	14/09/1966	27 Croston Road	Preston	Lancashire	PR5 3JL	01772 567685	330	Advanced Juggling	3	24	£75	25/08/2012	Merit
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	252	Intermediate fire eating	2	18	£45	01/09/2012	Pass
1254	Mr	Steven	Buchanan	13/08/1986	14 Garrett Hill	London	Greater London	SW1 8JR	0171 5554848	252	Intermediate fire eating	2	18	£45	20/09/2012	Pass
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	330	Advanced Juggling	3	24	£75	30/11/2012	Distinction

NORMALISE THE CIRCUS TO 1NF



1. Make a list of attributes.
2. Establish what each attribute means.
3. Ascertain Candidate Keys.
4. Pick a Primary key from the Candidate keys.

Trainee Number	Title	First Name	Surname	Date of Birth	Address Line 1	Town/City	County	Postcode	Tel No	Module Number	Module Name	Level	Module Duration (days)	Cost	Assessment Date	Result
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	250	Intermediate Static Trapeze	2	24	£50	21/02/2012	Distinction
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	101	Learning to balance	1	15	£35	10/07/2012	Distinction
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	102	Fire Eating for beginners	1	12	£30	11/07/2012	Distinction
1223	Miss	Clare	Croston	15/08/1998	15 Sure Street	Leyland	Lancashire	PR5 9GW	01772 678453	100	Learning to juggle	1	12	£25	14/07/2012	Pass
1222	Mrs	Lauren	Fazackerley	13/08/1978	21 Abbotts Road	London	Greater London	L56 7GH	0171 834 6337	200	Acrobatic Balancing	2	18	£48	22/07/2012	Pass
1223	Miss	Clare	Croston	15/08/1998	15 Sure Street	Leyland	Lancashire	PR5 9GW	01772 678453	100	Learning to juggle	1	12	£25	23/07/2012	Fail
1222	Mrs	Lauren	Fazackerley	13/08/1978	21 Abbotts Road	London	Greater London	L56 7GH	0171 834 6337	252	Intermediate fire eating	2	18	£45	29/07/2012	Merit
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	210	Flying Trapeze	2	20	£58	25/08/2012	Pass
1259	Miss	Patricia	Smithies	14/09/1966	27 Croston Road	Preston	Lancashire	PR5 3JL	01772 567685	330	Advanced Juggling	3	24	£75	25/08/2012	Merit
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	252	Intermediate fire eating	2	18	£45	01/09/2012	Pass
1254	Mr	Steven	Buchanan	13/08/1986	14 Garrett Hill	London	Greater London	SW1 8JR	0171 5554848	252	Intermediate fire eating	2	18	£45	20/09/2012	Pass
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	330	Advanced Juggling	3	24	£75	30/11/2012	Distinction

ATTRIBUTES:



Attributes
(taken from the column headers)

Trainee Number

Title

First name

Surname

Date of Birth

Address

Tel No

Module Number

Module Name

Level

Module Duration

Cost

Assessment Date

Result

Trainee Number	Title	First Name	Surname	Date of Birth	Address Line 1	Town/City	County	Postcode	Tel No	Module Number	Module Name	Level	Module Duration (days)	Cost	Assessment Date	Result
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	250	Intermediate Static Trapeze	2	24	£50	21/02/2012	Distinction
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	101	Learning to balance	1	15	£35	10/07/2012	Distinction
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	102	Fire Eating for beginners	1	12	£30	11/07/2012	Distinction
1223	Miss	Clare	Croston	15/08/1998	15 Sure Street	Leyland	Lancashire	PR5 9GW	01772 678453	100	Learning to juggle	1	12	£25	14/07/2012	Pass
1222	Mrs	Lauren	Fazackerley	13/08/1978	21 Abbotts Road	London	Greater London	L56 7GH	0171 834 6337	200	Acrobatic Balancing	2	18	£48	22/07/2012	Pass
1223	Miss	Clare	Croston	15/08/1998	15 Sure Street	Leyland	Lancashire	PR5 9GW	01772 678453	100	Learning to juggle	1	12	£25	23/07/2012	Fail
1222	Mrs	Lauren	Fazackerley	13/08/1978	21 Abbotts Road	London	Greater London	L56 7GH	0171 834 6337	252	Intermediate fire eating	2	18	£45	29/07/2012	Merit
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	210	Flying Trapeze	2	20	£58	25/08/2012	Pass
1259	Miss	Patricia	Smithies	14/09/1966	27 Croston Road	Preston	Lancashire	PR5 3JL	01772 567685	330	Advanced Juggling	3	24	£75	25/08/2012	Merit
1234	Rev	Brian	Broadgate	09/08/1980	Town Lane	Manchester	Greater Manchester	ML6 7FJ	01252 727 850	252	Intermediate fire eating	2	18	£45	01/09/2012	Pass
1254	Mr	Steven	Buchanan	13/08/1986	14 Garrett Hill	London	Greater London	SW1 8JR	0171 5554848	252	Intermediate fire eating	2	18	£45	20/09/2012	Pass
1235	Mr	Steve	Caswell	12/09/1985	32 Chestnut Close	Ramsgate	Cumbria	OL5 5DQ	01772 763425	330	Advanced Juggling	3	24	£75	30/11/2012	Distinction

Candidate Keys: Trainee Number, Module Number

1: Identify any repeating groups...

If Trainee number is central to the system (and becomes the primary key), which attributes would be repeated?

Attributes

Trainee Number

Title

First Name

Surname

Date of Birth

Address

Tel No

Module Number

Module Name

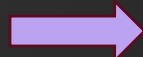
Level

Duration

Cost

Assessment date

Result



Each Trainee number relates to one:

Title

First Name

Surname

Date of Birth

Address

Tel No

Repeating Groups:

But each Trainee Number could have more than one module:

Module Number

Module Name

Level

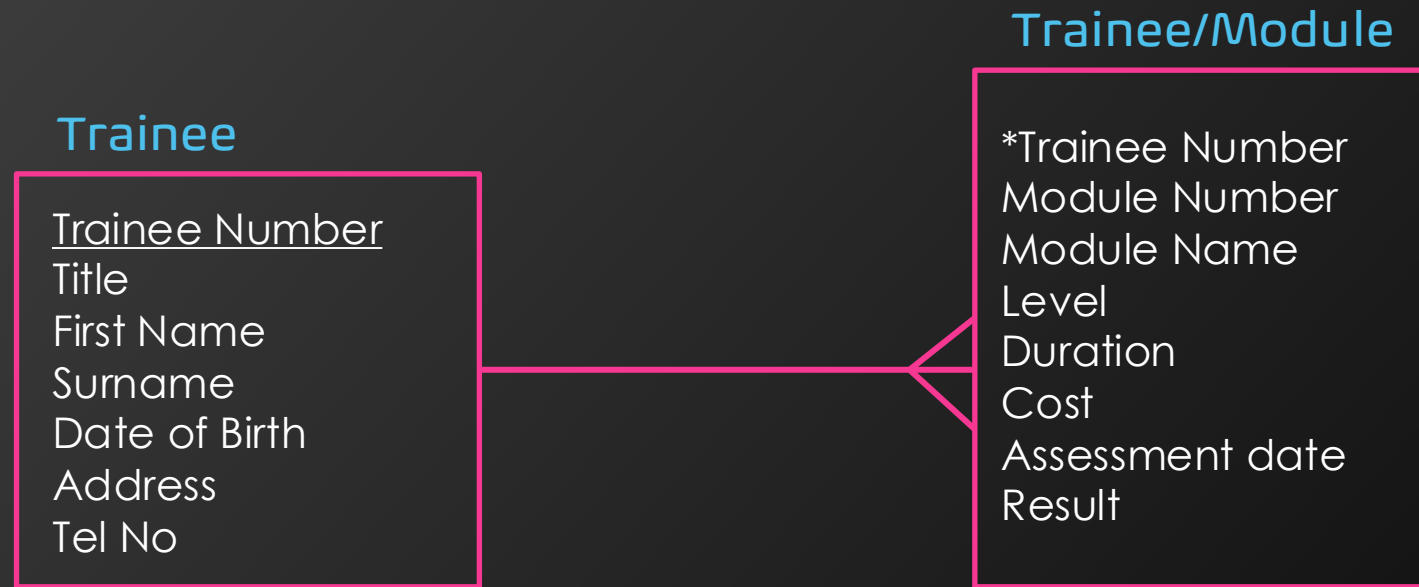
Duration

Cost

Assessment date

Result

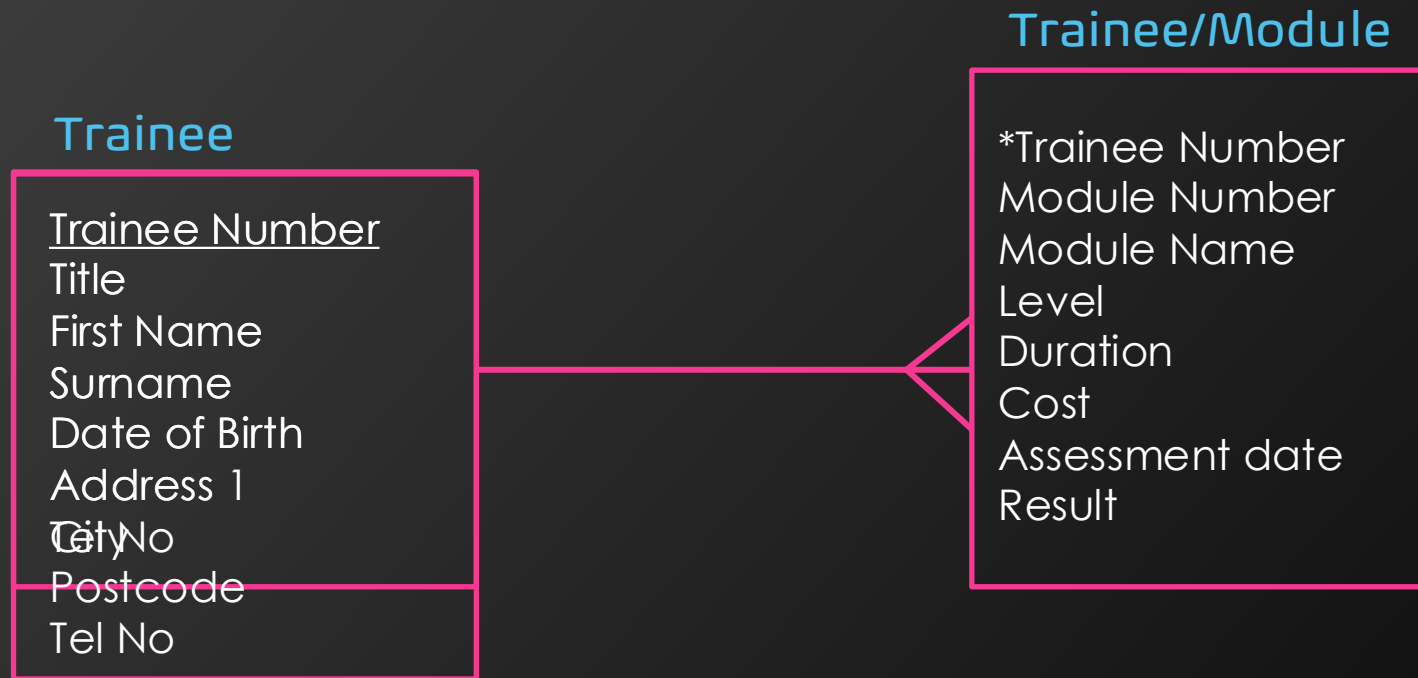
The repeating groups are then taken out and made into a new table.
We now have what resembles two tables - TRAINEE and TRAINEE/MODULE.



We also take a copy of the PRIMARY KEY (Trainee Number) which becomes a FOREIGN KEY in the new table.

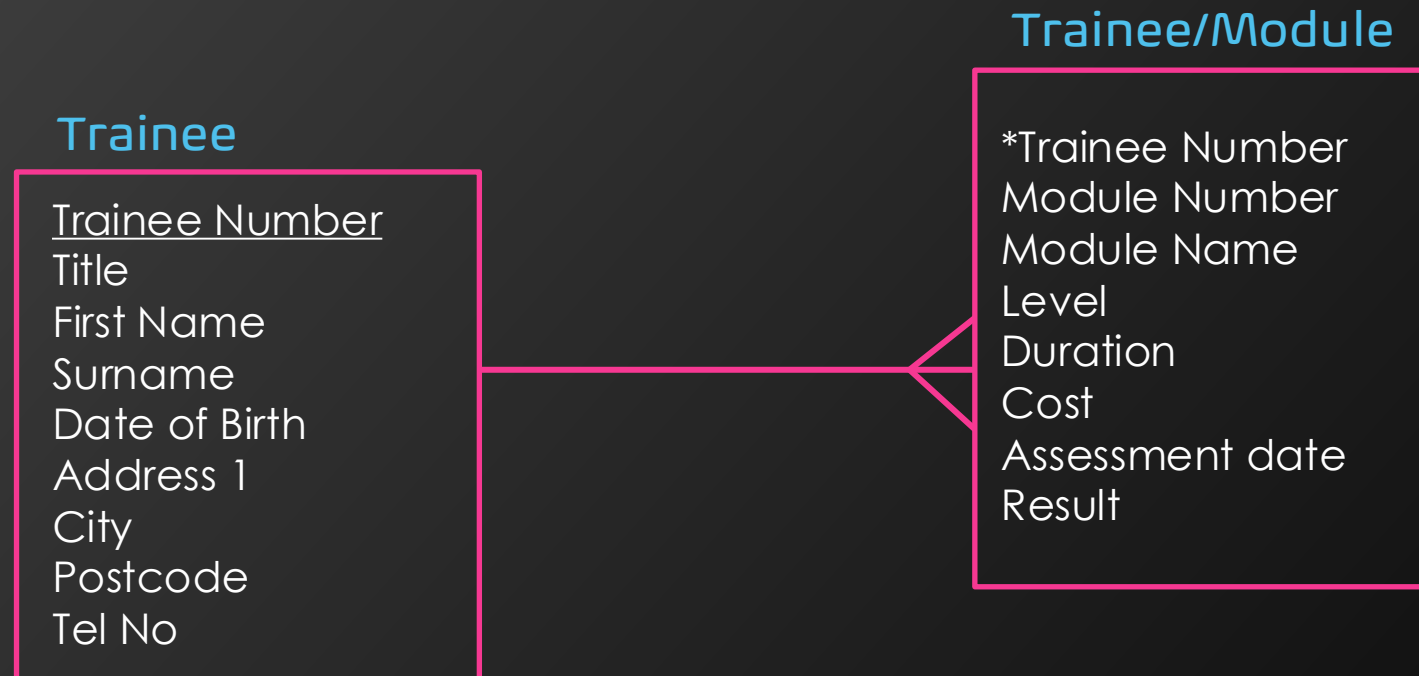
NOTE: Primary Keys are denoted by an underline. Foreign Keys are denoted by the * symbol.

2: Ensure all data values are atomic...



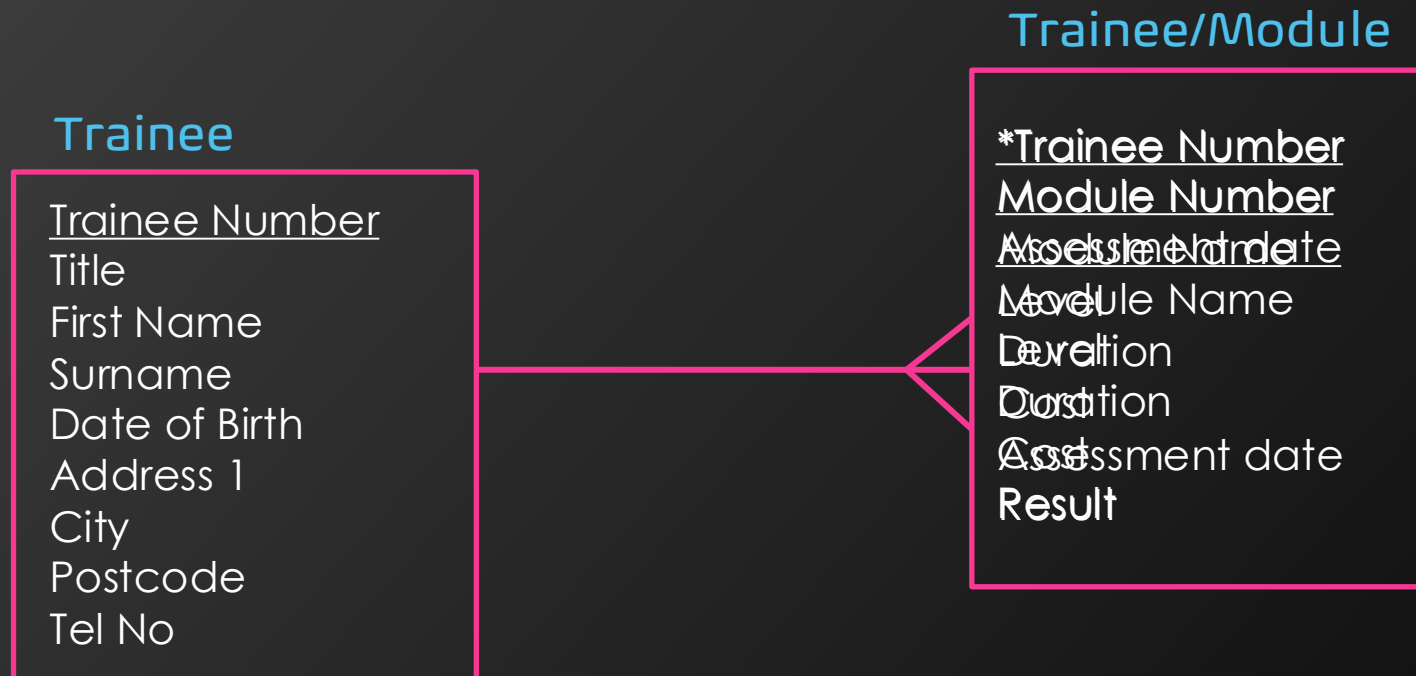
The ADDRESS field is currently NOT atomic as it contains values for the street, city, and postcode. We need to separate these into separate fields.

3: Each field/attribute has a unique name...



Our attributes pass this test.

4: Each table has a Primary Key...



In order to ensure that every record in the TRAINEE/MODULE table is unique we need to create a COMPOSITE Primary Key of fields Trainee Number, Module Number, and Assessment date (assuming that a trainee can only have one assessment for a module on one date)

1NF: EXAMPLE

TESCO

CHORLEY BUCKSHAW 0345 0269632

	£
CAT FOOD *	4.00
CAT FOOD *	4.00
CTRY LIFE SPRD	2.90
CTRY LIFE SPRD	2.90
HAM SLICES	1.60
SAUSAGES	1.10
MANUAL WEIGHT	
BANANAS LOOSE	
1.280 kg @ £0.68/ kg	0.87

SUB-TOTAL 17.37

MULTIBUY SAVINGS
SPREAD 2 FOR £4 -1.80
PET FOOD 2F£6 -2.00
TOTAL SAVINGS -3.80

TOTAL TO PAY 13.57
VISA DEBIT SALE 13.57

AID : A5018000081040
NUMBER : *****0014 ICC
PAN SEQ NO : 00
AUTH CODE : 13698
MERCHANT : 08406808
START : 07/13 EXPIRY : 06/16
Cardholder PIN Verified
CHANGE DUE 0.00

CLUBCARD STATEMENT

CLUBCARD NUMBER 60400403403765903*
POINTS THIS VISIT 13
TOTAL UP TO 02/02/14 68
TOTAL INCLUDES :
GREEN CLUBCARD POINTS 0

How did we do?
Visit www.tescocomments.com and
tell us about your shopping trip

3/02/14 8:19 6250 073 9073 9691

TESCO

CHORLEY BUCKSHAW 0345 0269632

	£
CAT FOOD *	4.00
CAT FOOD *	4.00
CTRY LIFE SPRD	2.90
CTRY LIFE SPRD	2.90
HAM SLICES	1.60
SAUSAGES	1.10
MANUAL WEIGHT	
BANANAS LOOSE	
1.280 kg @ £0.68/ kg	0.87
SUB-TOTAL	17.37

MULTIBUY SAVINGS	
SPREAD 2 FOR £4	-1.80
PET FOOD 2FE6	-2.00
TOTAL SAVINGS	-3.80

TOTAL TO PAY	13.57
VISA DEBIT SALE	13.57

AID : 5000000081040	
NUMBER : *****0014	ICC
PAN SEQ NO : 00	
AUTH CODE : 13698	
MERCHANT : 08400800	
START : 07/13 EXPIRY : 06/16	
Cardholder PIN Verified	
CHANGE DUE	0.00

CLUBCARD STATEMENT	
CLUBCARD NUMBER 60400401407755903*	
POINTS THIS VISIT	13
TOTAL UP TO 02/02/14	68
TOTAL INCLUDES :	
GREEN CLUBCARD POINTS	0

How did we do?
Visit www.tescocomments.com and
tell us about your shopping trip

3/02/14 8:19 6250 073 9073 9691

Attributes (with extras added to clarify)

Receipt ID

Date

Time

Store ID

Store Address line1, line2, etc.

Store Telephone No.

Till ID

Till Location

Cashier ID

Product ID

Product Description

Product Weight

Product Price

Clubcard No.

Clubcard Points Earned This Visit

Customer First name

Customer Surname

Customer Address line1, line2, etc.

Clubcard Total Points

Which bits make
up the receipt ID?

Not Included:

Total to Pay: probably just a
calculated field

Change Due: Another
calculated field

Credit Card Detail: not stored

2: Identify repeating groups...

If Receipt ID becomes the primary key,
which attributes would be repeated?

Receipt ID
Date
Time
Store ID
Store Address line1, line2, etc.
Store Telephone No.
Till ID
Till Location
Cashier ID
Product ID
Product Description
Product Weight
Product Price
Clubcard No.
Clubcard Points Earned This Visit
Customer First name
Customer Surname
Customer Address line1, line2, etc.
Clubcard Total Points



Each Receipt ID relates to one:
Date
Time
Store ID
Store Address line1, line2, etc.
Store Telephone No.
Till ID
Till Location
Cashier ID
Clubcard No.
Clubcard Points Earned This Visit
Customer First name
Customer Surname
Customer Address line1, line2, etc.
Clubcard Total Points

Repeating Groups:

But each Receipt ID is likely to have more than one product:
Product ID
Product Description
Product Weight
Product Price



Receipt
<u>Receipt ID</u>
Date
Time
Store ID
Store Address line1, line 2, etc.
Store telephone number
Till ID
Till location
Cashier ID
Cashier FirstName
Cashier Surname
Clubcard No
Clubcard points earned this visit
Customer FirstName
Customer Surname
Customer Address line 1, line 2 etc.
Clubcard total points

Items Bought
<u>*Receipt ID</u>
<u>Product ID</u>
Product Description
Product Weight
Product Price

- 2: Ensure all values are atomic
- 3: All attributes have a unique name
- 4: All tables have a primary key

It could be argued that the Receipt ID alone does not uniquely identify the receipt as different stores may use the same receipt numbers. In this case we may need to use the whole barcode (date, time, store, till, cashier, receipt ID)

SUMMARY

- Normalisation is a set of procedures that aid in the design of databases.
- Normalisation is used as part of the database design process and complements the ERD design technique.
- Normalisation removes unnecessary redundancy from a database.
- Normalisation can identify potential anomalies in the database.
- Normalisation is a multi-stage process where the product of each stage is referred to as a Normal Form.
- Normal Forms 1NF, 2NF, 3NF and Boyce-Codd are the most significant.

RESOURCES

- Richie, C. (2008). *Database Principles and Design*. 3rd Ed. Thompson Learning
- [Normalisation 1NF: First Normal Form Example \(youtube.com\)](#)