Unit 2 Creating Systems to Manage Information
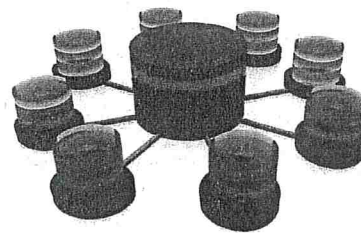
Normalisation - The Basics

Unit 2
Creating
Systems to
Manage
Information

What is Normalisation?

A technique of **organising** the **data** in a database into **multiple related entities** (tables) to <u>minimise</u> **Data Redundancy**

Unit 2 Creating Systems to Manage Information

How do we normalise a database?

1. The **aim is to reduce data redundancy (reduce duplicate data) to improve data integrity**

   a. Data Redundancy is the duplication of similar data in multiple places

2. **Divide the attributes (fields) into separate, logical entities (tables)**

   a. Each **entity (table)** will have a **primary key**

   b. All **attributes (fields)** in an entity (table) **must relate to the primary key**

   c. Entities will be linked with a **common key** (primary key from one table links to a foreign key in another table)

Unit 2 Creating Systems to Manage Information

Welcome to Herman's Hamster Club!

Members fill in this paper form

The data is then entered into a database

Herman's Hamster Club – Member Record

Member ID: _____003_____

Surname: _____Sholinski_____

Initial: _____G_____

House Number: __123__ Post Code: __Hj23 4BB__

Mobile Number: __07703334443__

Email Address: _____Geoff@gmail.com_____

Hamster Details

Hamster Name	Gender	Breed	Origin of Breed	Life Expectancy
Kebab	F	Chinese Dwarf	China	3 years
Chips	M	Chinese Dwarf	China	3 years
Chilli	F	Russian Dwarf	Russia	2 years

Unit 2 Creating Systems to Manage Information

The members' data has not been input into a database.

Can you see any duplicate data now?

UNF

M	li	LastNar	Hc	PostCo	Mobile	Email	Hamste	Ham	Breed	Origin	Li
001	L	Dolinski	41	HJ48 4JJ	07905555555	leedolinski@rodents.com	Stormy	M	Syrian	Syria	3
001	L	Dolinski	41	HJ48 4JJ	07905555555	leedolinski@rodents.com	Schuey	F	Chinese Dwarf	China	3
001	L	Dolinski	41	HJ48 4JJ	07905555555	leedolinski@rodents.com	Minnie	F	Syrian	Syria	3
002	S	Polinski	2a	HJ51 5PP	07804445554	SuperPolin@hotmail.com	Biscuit	F	Russian Dwarf	Russia	2
002	S	Polinski	2a	HJ51 5PP	07804445554	SuperPolin@hotmail.com	Flabs	M	Russian Dwarf	Russia	2
003	G	Sholinski	123	HJ23 4BB	07703334443	Geoff@gmail.com	Kebab	F	Chinese Dwarf	China	3
003	G	Sholinski	123	HJ23 4BB	07703334443	Geoff@gmail.com	Chips	M	Chinese Dwarf	China	3
003	G	Sholinski	123	Hj23 4BB	07703334443	Geoff@gmail.com	Chilli	F	Russian Dwarf	Russia	2

Customer and breed data has been duplicated

The is called un-normalised data (un-normalised form UNF)

UNF

M	li	LastNar	Hc	PostCo	Mobile	Email	Hamste	Ham	Breed	Origin	Lil
001	L	Dolinski	41	HJ48 4JJ	07905555555	leedolinski@rodents.com	Stormy	M	Syrian	Syria	3
001	L	Dolinski	41	HJ48 4JJ	07905555555	leedolinski@rodents.com	Schuey	F	Chinese Dwarf	China	3
001	L	Dolinski	41	HJ48 4JJ	07905555555	leedolinski@rodents.com	Minnie	F	Syrian	Syria	3
002	S	Polinski	2a	HJ51 5PP	07804445554	SuperPolin@hotmail.com	Biscuit	F	Russian Dwarf	Russia	2
002	S	Polinski	2a	HJ51 5PP	07804445554	SuperPolin@hotmail.com	Flabs	M	Russian Dwarf	Russia	2
003	G	Sholinski	123	HJ23 4BB	07703334443	Geoff@gmail.com	Kebab	F	Chinese Dwarf	China	3
003	G	Sholinski	123	HJ23 4BB	07703334443	Geoff@gmail.com	Chips	M	Chinese Dwarf	China	3
003	G	Sholinski	123	Hj23 4BB	07703334443	Geoff@gmail.com	Chilli	F	Russian Dwarf	Russia	2

What happens if we don't normalise?

Anomalies (errors) will occur

- Entering data multiple time will lead to **mistakes** and it **wastes time**
  - Every time we add a new hamster, we have to add the member's details AND the breed details
  - this will take a long time and we might make typos leading to **inconsistent data**
- If we delete a hamster's details we would also delete a member's details too
  - Losing important data is a massive risk

An attempt to normalise the database

| HamsterI	Membe	Hamste	Ha	Breed	Origin	Lifei
ham01	001	Schuey	F	Chinese Dwarf	China	3
ham02	001	Stormy	M	Syrian	Syria	3
ham03	001	Minnie	F	Syrian	Syria	3
ham04	002	Biscuit	F	Russian Dwarf	Russia	2
ham05	002	Flabs	M	Russian Dwarf	Russia	2
ham06	003	Kebab	F	Chinese Dwarf	China	3
ham07	003	Chips	M	Chinese Dwarf	China	3
ham08	003	Chilli	F	Russian Dwarf	Russia	2

I've divided the data into two tables

Is there any duplicate data left?

MemberID	Initia	LastName	HouseNo	PostCode	Mobile	Email
003	G	Sholinski	123	HJ23 4BB	07703334443	Geoff@gmail.
001	L	Dolinski	41	HJ48 4JJ	07905555555	leedolinski@r
002	S	Polinski	2a	HJ51 5PP	07804445554	SuperPolin@h

4

Your turn to attempt to normalise this database

- Draw a conceptual ERD:
- Show the **entities** (tables)
  - Do not add attributes (fields) in a conceptual ERD
- Show the **relationships** between the entities

| MemberID |
| Initial |
| LastName |
| HouseNo |
| PostCode |
| Mobile |
| Email |
| HamsterName |
| HamsterGender |
| Breed |
| Origin |
| LifeExpecYears |
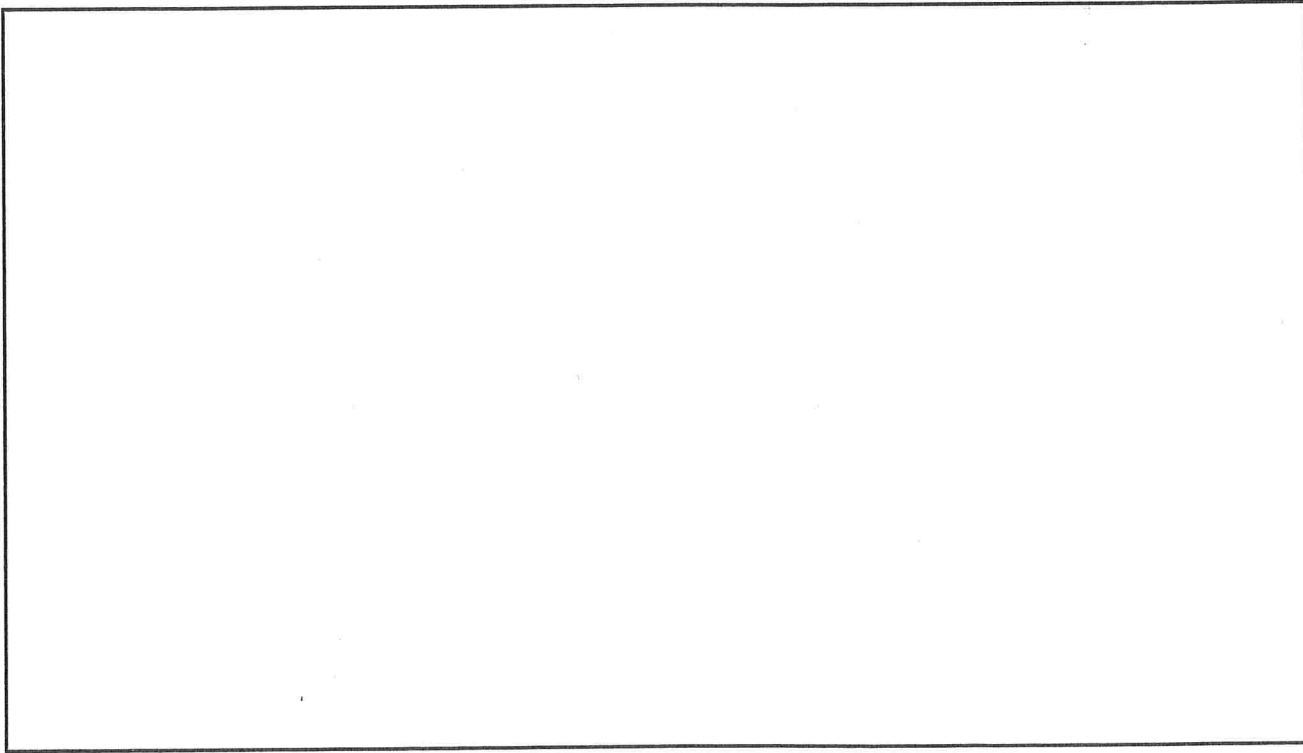
Now draw a Logical ERD

- Include the entity(table) names
- Include a primary key for each table
- Include foreign keys where necessary
- Put all the attributes (fields) into the correct entities
- Show the relationships

MemberID
Initial
LastName
HouseNo
PostCode
Mobile
Email
HamsterName
HamsterGender
Breed
Origin
LifeExpecYears
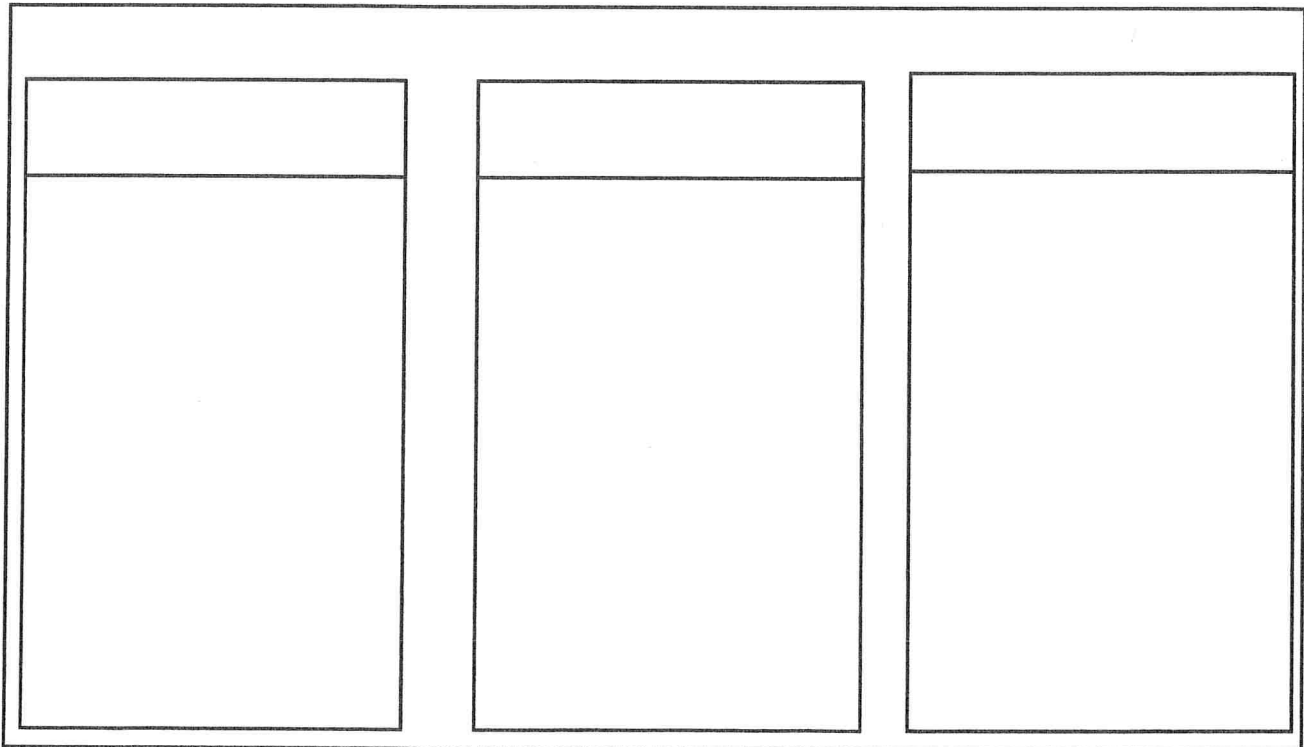
Congratulations!

- You have successfully normalised a database
  - **This is called third normal form (3NF)**
- You've removed duplicate data
  - **Data redundancy has been reduced** therefore you have **good data integrity**
- Created sensible **entities (tables)** and distributed **attributes (fields)** correctly
- **Primary keys** are shown in every entity and links between tables have been made with the addition of **foreign keys**