Database basics

The modern world relies on **data**

Businesses, shops, delivery companies, schools, sports teams – all of these use data

Modern life runs using data

Database basics

Data needs to be organised

We do this in **databases**

A **database** is a structured (organised) collection (or set) of **data**

It's data that's organised in a designed way so that it can be used easily

Data can be added, amended, deleted and retrieved from a database. We do all of this using a **query language**

Database basics

Key definition:

A **database** is a structured (organised) collection (or set) of **data**

NOT information

Database basics

Data needs to be organised.

We do this in **databases**.

When you think of a database, think of a set of files in a filing cabinet

Each file contains information about a person or thing (an **entity**)

They are organised so it's easy to find things quickly

Database basics

Data in databases is stored in **tables**

Simple databases might have just one table. Complex ones might have lots

A table stores data about one type of thing - products, people, songs, films etc… Individual things are called **entities**

Marg	Giraffe	12	F	4.7
Molly	Giraffe	6	F	4.4
Mike	Giraffe	14	M	5.1
Sammy	Giraffe	1	M	2.4
Rex	Lion	4	M	1.8
Leah	Lion	5	F	1.3
Luna	Lion	2	F	0.8
Andrea	Aardvark	7	F	0.54
Jawa	Python	3	M	3.2
Petra	Python	2	F	2.8

Marg	Giraffe	12	F	4.7
Molly	Giraffe	6	F	4.4
Mike	Giraffe	14	M	5.1
Sammy	Giraffe	1	M	2.4
Rex	Lion	4	M	1.8
Leah	Lion	5	F	1.3
Luna	Lion	2	F	0.8
Andrea	Aardvark	7	F	0.54
Jawa	Python	3	M	3.2
Petra	Python	2	F	2.8

name	species	age	gender	height
Marg	Giraffe	12	F	4.7
Molly	Giraffe	6	F	4.4
Mike	Giraffe	14	M	5.1
Sammy	Giraffe	1	M	2.4
Rex	Lion	4	M	1.8
Leah	Lion	5	F	1.3
Luna	Lion	2	F	0.8
Andrea	Aardvark	7	F	0.54
Jawa	Python	3	M	3.2
Petra	Python	2	F	2.8

name	species	age	gender	height
Marg	Giraffe	12	F	4.7
Molly	Giraffe	6	F	4.4
Mike	Giraffe	14	M	5.1
Sammy	Giraffe	1	M	2.4
Rex	Lion	4	M	1.8
Leah	Lion	5	F	1.3
Jawa	Python	3	M	3.2
Petra	Python	2	F	2.8

Each column stores data on one feature of the animal. In the database, a column is called a **field**

name	species	age	gender	height
Marg	Giraffe	12	F	4.7
olly	Giraffe	6	F	4.4
ike	Giraffe	14	M	5.1
ammy	Giraffe	1	M	2.4
ex	Lion	4	M	1.8
Leah	Lion	5	F	1.3
Luna	Lion	2	F	0.8
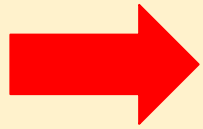
The top row is the **field names** (or **attributes**) – we call this the **identifier**

Each field uses the same **data type** – string, integer, floating point number, Boolean etc…

Data types can also include date and currency – both of which are really useful

name	species	age	gender	height
Marg	Giraffe	12	F	4.7
Molly	Giraffe	6	F	4.4
Mike	Giraffe	14	M	5.1
Sammy	Giraffe	1	M	2.4
Rex	Lion	4	M	1.8
Andrea	Aardvark	7	F	0.54
Jawa	Python	3	M	3.2
Petra	Python	2	F	2.8

Each row stores the data about one animal. If I add an animal, I need to add a row at the bottom

The data in each row is called a **record**

Database basics

- **Table** – a set of data organised together for a group of entities of the same type

- **Field** – an individual piece of data about a thing (entity) – a column

- **Field names / attributes** – the name (identifier) given to each field

- **Record** – a number of related data items about the same thing (entity) – a row
  Each record in a data table has the same set of fields

name	species	age	gender	height
Marg	Giraffe	12	F	4.7
Molly	Giraffe	6	F	4.4
Molly	Giraffe	5	F	4.2
Mike	Giraffe	14	M	5.1
Sammy	Giraffe	1	M	2.4

A new giraffe is joining my zoo. Her name is also Molly

What problem does this cause?

How can I resolve the problem?

Andrea	Aardvark	7	F	0.54
Jawa	Python	3	M	3.2
Petra	Python	2	F	2.8

name	species	age	gender	height
Marg	Giraffe	12	F	4.7
Molly	Giraffe	6	F	4.4
Molly	Giraffe	5	F	4.2
Mike	Giraffe	14	M	5.1

How can I be certain I'm going to get the right giraffe called Molly?

What is I also have an aardvark called Molly?

Names are difficult as they aren't unique

Andrea	Aardvark	7	F	0.54
Jawa	Python	3	M	3.2
Petra	Python	2	F	2.8

id	name	species	age	gender	height
001	Marg	Giraffe	12	F	4.7
002	Molly	Giraffe	6	F	4.4
003	Molly	Giraffe	5	F	4.2
004	Mike	Giraffe	14	M	5.1
005	Sammy	Giraffe	1	M	2.4
006	Rex	Lion	4	M	1.8

By adding an ID field, I can identify each **record** with a **unique identifier**

This is a really important idea in database development. Each record **must** have a unique identifier. We call this the **Primary Key**

| 011 | Petra | Python | 2 | F | 2.8 |

Database basics

- **Primary key** -– a field which uniquely identifies a record

  It is a **unique identifier**

  It is often an ID code of some kind

  You can use the primary key to get all the data related to a record (e.g id 003)

  It is important that the primary key is **unique** so that the right data is identified

id	name	species	age	gender	height
001	Marg	Giraffe	12	F	4.7
002	Molly	Giraffe	6	F	4.4
003	Molly	Giraffe	5	F	4.2
004	Mike	Giraffe	14	M	5.1
005	Sammy	Giraffe	1	M	2.4
006	Rex	Lion	4	M	1.8
007	Leah	Lion	5	F	1.3
008	Luna	Lion	2	F	0.8
009	Andrea	Aardvark	7	F	0.54
010	Jawa	Python	3	M	3.2
011	Petra	Python	2	F	2.8