

Database Software

Terry Marris September 2009

6 Principles

We have seen how to create, query and report on a simple database. Now we look at some theory.

6.1 Database Basics

Examples of well-known databases include:

- criminal records - the national DNA Database holds samples of DNA taken at crime scenes and from individuals in police custody. It helps to:
 - identify offenders
 - make arrests
 - secure convictions
 - provide leads for investigations
- car records - Driver Vehicle Licensing Agency (DVLA) database records motor vehicles and their registered keepers. It helps to:
 - record tax paid
 - register new and imported vehicles
 - record sold, scrapped and exported vehicles
 - allocate number plates and registration marks
 - fight crime involving vehicles
- bank accounts - keeps track of financial transactions (money going in and money going out) of individuals and company's bank accounts. It helps to:
 - automate speedy money transfers
 - minimise errors
 - detect fraud
- employee records - keeps track of a company's employees. It helps to:
 - identify current employees and where they are located
 - pay the right employee the right salary each month
 - allocate employees to training programs
- student records - keeps records of students. It helps to:
 - identify each student
 - list the courses each student has enrolled on
 - record their achievements

A database is a structured collection of data, together with a set of tools that can retrieve the data on demand and for different purposes.

The advantages and disadvantages of a computerised database include:

Advantages	Disadvantages
<ol style="list-style-type: none"> 1. speed - you can quickly find what you want 2. accuracy - tasks are automated, duplicated data is minimised 3. efficiency - maximum work done for minimum cost - you do not need to pay an army of workers 4. updates to records, and additions, easily done 5. security - could be made difficult (but not impossible) to steal data 6. shared access - easily shared among those who need to use it, all at the same time 	<ol style="list-style-type: none"> 1. complexity - large databases with many linked tables can be horribly complicated 2. costs money e.g. for updating hardware with more storage capacity, training staff and maintaining security 3. risk of system failure - if the computer system goes down nobody can do any work

The advantages and disadvantages of a paper-based database include:

Advantages	Disadvantages
<ol style="list-style-type: none"> 1. sweet, simple and small 2. hard copy - you can immediately see what is going on 	<ol style="list-style-type: none"> 1. size - cannot handle large volumes of data 2. accuracy - users (people) make mistakes. Updating is a pain - you have to scratch out unwanted data before you replace it - messy 3. speed - cannot always easily find the information you want 4. compiling reports - difficult and time consuming 5. updates - inserting, amending or deleting records is difficult 6. storage - takes up space e.g. filing cabinets, shoe boxes, ..., to store the paper records 7. security - you have fire risks - paper burns, you have unauthorised people reading what is written - not nice, you have no automatic and reliable audit trail - somebody can change the data and you not know about it, paper gets lost or miss-placed 8. shared access - impossible. Only one person at a time may view the same record

Database objects include:

Database Object	Example																					
tables - contain data in rows (records) and columns (fields)	<table border="1"> <thead> <tr> <th>Name</th> <th>Date of Birth</th> <th>Gender</th> </tr> </thead> <tbody> <tr> <td>Pearl Button</td> <td>16 Nov 1996</td> <td>female</td> </tr> <tr> <td>Jo King</td> <td>30 Apr 1998</td> <td>male</td> </tr> <tr> <td>Barry Cade</td> <td>4 May 1995</td> <td>male</td> </tr> <tr> <td>Carrie Oakey</td> <td>25 Sep 1999</td> <td>female</td> </tr> <tr> <td>Priti Manek</td> <td>13 Feb 1999</td> <td>female</td> </tr> <tr> <td>Tim Burr</td> <td>7 Mar 1997</td> <td>male</td> </tr> </tbody> </table>	Name	Date of Birth	Gender	Pearl Button	16 Nov 1996	female	Jo King	30 Apr 1998	male	Barry Cade	4 May 1995	male	Carrie Oakey	25 Sep 1999	female	Priti Manek	13 Feb 1999	female	Tim Burr	7 Mar 1997	male
Name	Date of Birth	Gender																				
Pearl Button	16 Nov 1996	female																				
Jo King	30 Apr 1998	male																				
Barry Cade	4 May 1995	male																				
Carrie Oakey	25 Sep 1999	female																				
Priti Manek	13 Feb 1999	female																				
Tim Burr	7 Mar 1997	male																				
queries - looks for information	who has a birthday in September? Carrie Oakey																					
reports - looks for information, result can be formatted and printed	list of all females in date of birth order Pearl Button Priti Manek Carrie Oakey																					
forms - used to get input from the user	to add a new record <table border="1"> <tbody> <tr> <td>Name?</td> <td>Terry Bull</td> </tr> <tr> <td>Date of Birth?</td> <td>5 June 1996</td> </tr> <tr> <td>Gender?</td> <td>male</td> </tr> </tbody> </table>	Name?	Terry Bull	Date of Birth?	5 June 1996	Gender?	male															
Name?	Terry Bull																					
Date of Birth?	5 June 1996																					
Gender?	male																					

A data type is a set of similar values. Date types include:

Data Type	Description	Example
Text	letters of the alphabet, digits, punctuation symbols	221B Baker Street
Number	whole numbers (1, 2, 3, ...) or floating point numbers (numbers with a decimal point e.g. 3.1416). You do sums with numbers.	123
Date/Time	day, month, year. Time in 24-hour clock notation	4 Sep 2009, 16:00
Currency	Money	£350.75

Size of a Database One way of estimating the size of a database is to count the number of records it contains.

Name	Date of Birth	Gender
Pearl Button	16 Nov 1996	female
Jo King	30 Apr 1998	male
Barry Cade	4 May 1995	male
Carrie Oakey	25 Sep 1999	female
Priti Manek	13 Feb 1999	female
Tim Burr	7 Mar 1997	male

You can see with your own eyeballs that, in this example, the size is 6 records.

A more usual way of estimating size is to count the maximum number of characters that a record can have, then multiplying that by the number of records. Suppose we allow:

- 25 characters for a name
- 11 characters for a date of birth (a space is also a character to be counted)
- 6 characters for a gender

that is 42 characters for one record

that is 42 x 6 characters for six records
or 252 characters altogether (what is 42 x 6?)

Storage measurements include:

byte 8 bits, stores just one character
kilobyte (Kb) 1000 characters
megabyte (Mb) 1 million characters

gigabyte (Gb) 1 billion characters

terabyte (Tb) 1 thousand billion characters

a home computer might have
80Mb hard disk space
a small business with 5
computers might need 160Mb for
each computer
a small business with ten
computers on a network might
need 500 Gb
a large network with 50
computers might need a Tbs of
storage

6.2 Database Structures

A table is a fundamental element of an Access database. Any number of tables can be linked, and this can lead to large, complex structures. But we shall KISS - keep it small and simple. A table is organised into rows and columns.

A row is known as a record.

A column is known as a field.

A primary key uniquely identifies a record. Two records may have the same name or the same date of birth or the same gender for example. But no two records may have the same primary key. A primary key might be a record number, a student number, a national insurance number, a car registration number, an e-mail address.

Record Number	Name	Date of Birth	Gender
1	Pearl Button	16 Nov 1996	female
2	Jo King	30 Apr 1998	male
3	Barry Cade	4 May 1995	male
4	Carrie Oakey	25 Sep 1999	female
5	Priti Manek	13 Feb 1999	female
6	Tim Burr	7 Mar 1997	male
7	Tim Burr	15 May 1996	male

This kind of organisation helps to find data quickly. For example, to find all people over the age of 21 you just have to look down the Date of Birth column.

6.3 Data Integrity

Data integrity is about data being consistent and correct.

Data accuracy means that data remains unchanged when it is copied, moved or printed.

Data consistency means ensuring that if you have two copies of the same item of data, perhaps in different tables, both copies get updated at the same time.

Validation means ensuring that data input conforms to pre-defined standards. For example:

- a person's age input is between 0 and 120
- a person's gender is either male or female
- a student's id number has six digits
- a person's e-mail address is not blank
- a person's post code has the format two letters, one or two digits, space, one digit, two letters e.g. LE1 3WA, LE18 5HP

Lookup lists, default values and upper and lower limits help ensure data input is valid.

A lookup list presents a list of options - you choose one and the appropriate value is input.

A default value is the usual value. For example, if you were inputting students' addresses, the default city might be Leicester since nearly every student lives in Leicester. Of course, you can change this value if you need to.

As an example of a limit, a student's date of birth must be at least 18 years before 1 September of this year.

We have looked at look at some example databases, discussed some advantages and disadvantages of computerised databases, described some database objects, explored some data types, seen how to determine the size of a database, looked at a simple database structure, and examined data integrity.

Next we ... who knows! Validation in more detail?

Bibliography

www.homeoffice.gov.uk/science-research/using-science/dna-database accessed Aug 2009
http://www.dft.gov.uk/dvla/vehicles.aspx accessed August 2009