

Some notes on databases

John R Hudson

6th May 2024

Definition

A database contains ordered lists of entities with some of their attributes; it does not have to be complete; it can be divided into sections; it can be indexed. A card index, a dictionary and a telephone directory are all examples of databases. Computers offer many types of database, among them, card index, hierarchical, key pair and relational table while some computer programs, such as email clients and accounting programs, have databases in which to store relevant data.

Types

Card index

A card index, or flat file, database looks like a card index (Figure 1). You can look at one card, or 'record,' at a time and it is usually set out in an attractive way like a visiting card or a business card. All simple computer databases use the card index principle.

Hierarchical database

Hierarchical databases are used where a number of attributes can be linked hierarchically to an entity, for example, the data relating to a passenger who has booked a seat on an aeroplane or a coach (figure 2). This structure can allow rapid retrieval of data but relies on one to many relationships.

A variation on the hierarchical database is a network database which permits many to many relationships; while not a perfect example of a network database, figure 3 illustrates using several network-like databases to show the relationships between the elements in the HTML specification (Hickson, 2023).

Key pair

Key pair databases are used where the relationships between entities are not essential; they consist of a variable, the key, and its value. They are widely used for settings on computers and smartphones and in a number of programming languages. In dBASE memory variables are key pairs and are used to allow manipulation of data, particularly where there are multiple stages to the manipulation, both for speed and to avoid constantly accessing the data table.

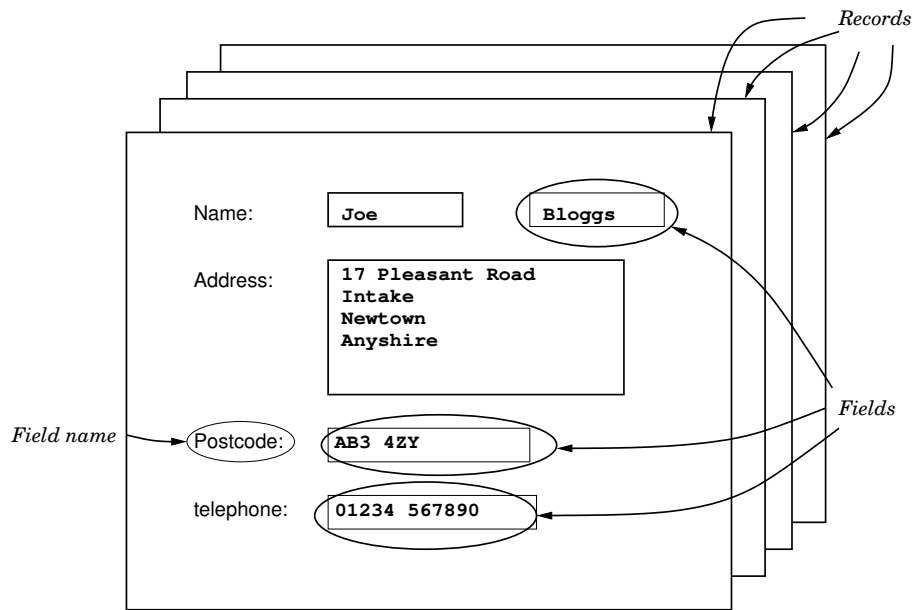


Figure 1: A card index database

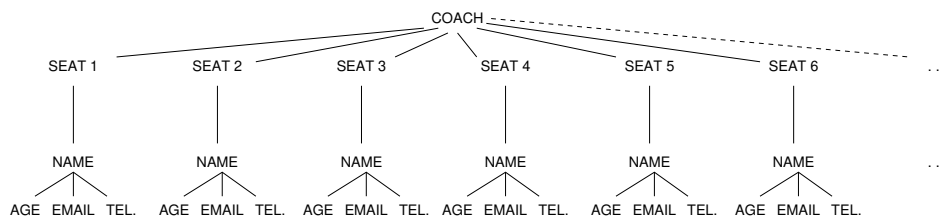


Figure 2: A hierarchical database

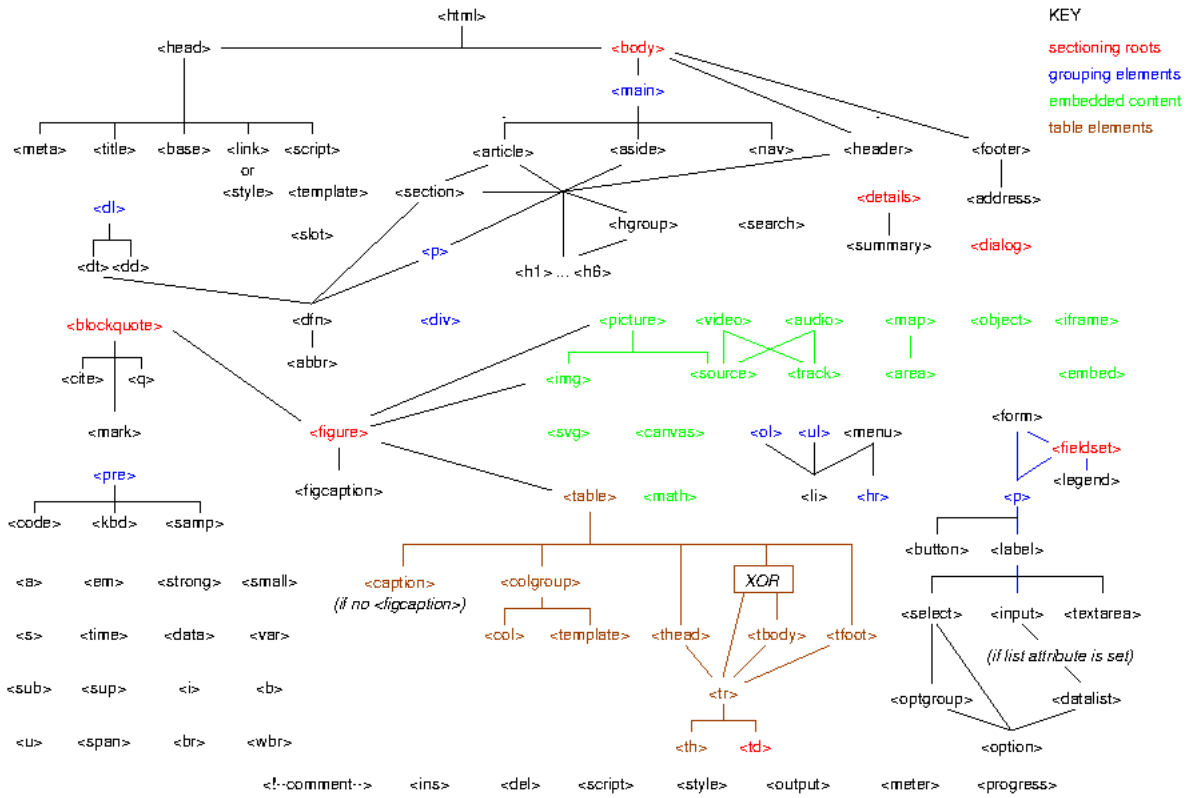


Figure 3: A network database

[Valkey](#) uses the same principles and can be used as a standalone database or, more commonly, for high speed manipulation of data held in another database.

Relational table

Relational table databases are now the most popular type of personal computer and smartphone database with [SQLite](#) on smartphones the most popular database ahead of [MariaDB](#), the fork of [MySQL](#), on personal computers, the latter two often the underlying database on websites. All can be disguised to look like a simple card index. When you see them in all their glory they can look like a very big spreadsheet (Figure 4).

	Field name	Field			
	FORENAME	SURNAME	ADDRESS1	ADDRESS2	ADDRESS3
	Joe	Bloggs	17 Pleasant Street	Intake	Newtown
	Mary	Smith	2 Eden Drive	Intake	Newtown
	Steve	Jones	85 Paradise Drive		Oldtown
Record	Tom	Cobley	The Cottage	Marsh	Oldtown
	Harry	Hawk	4 Marsh Road	Intake	Newtown
Record number	6	Daniel	7 Forest Drive		Oldtown

Empty field

Figure 4: A relational database

Components

Entities: items about which data is to be stored

Attributes: data relating to each entity; among the most common types of data are:

Alphanumeric a type of data using any printable character — text, numerals, punctuation, etc. — always used for text and also for numerals like telephone numbers on which you are not going to do any mathematical operations

Date a type of data in which only dates are used — you can do mathematical operations on date data like adding so many days to a date in order to find a new date

Logical or Boolean, a type of data which can only be ‘T(rue)’ or ‘F(alse)’, ‘Y(es)’ or ‘N(o)’

Numeric a type of data in which only numbers are used — you can do mathematical operations on numeric data

Fields: ‘containers’ to hold entities and attributes

Field name the name of a particular ‘container’

Field size the length in bytes, characters or digits, including any decimal points, of a ‘container’; not all database programs require you to specify the length of each field; note that, with the adoption of utf-8, the length of a field in bytes may be different from the length of a field in characters or digits

Field type the type of data which the field can hold — alphanumeric, date, logical, numeric, etc.; some databases do not enforce this rigorously and allow different types of data to be stored in a field

Indexes: a means of pretending that a table has been sorted when it hasn’t; because sorting large tables takes time, it is usually easier to create an index of the entries in one field, or, in some databases, entries in more than one field, and consult that to find particular records; also, by having indexes on different fields, you can pretend that a table has been sorted one way and then another without ever changing the order of any of the records in the table

A database may not have an index particularly if it is small or there is another way of finding items quickly, for example, by searching for the subject or sender of an email or a supplier in an accounting program.

Key field a field you cannot leave blank, whether because it is the field on which an index has been created or is the one around which a relational table is organised (used in some relational table database programs but not all)

Logical and physical views

Physical view: the way data is stored in the database, normally using one of the main types of database

Logical view: the way data is viewed on the screen; this should normally reflect the user’s needs rather than the type of database; for example,

- with hierarchical and relational databases only part of a record may be seen whether for ease of use or because the user only has permission to view or edit certain fields
- with an 'indexed' field records may be seen in a different order from their physical order.

Common terms

Ascending the normal order of a sort — see Sort

Character a type of data — the same as Alphanumeric

Column all instances of a particular field in a relational table database

Descending the order of a sort used when you want the most recent date or the highest figure in a numeric field to come first in the database — see Sort

File all the records in one card index or one relational table (n.b. large relational table databases have more than one relational table, each in its own file); key pairs are normally held in memory having been drawn from a database or program where they are normally stored but can be stored in a file as, for example in dBASE, to preserve a record of all the memory variables that were in use at the end of a session so that the user can load them at the start of the next session and carry on where they left off

First normal form the first step in normalising a database — see Normalisation

Normalisation the process of ensuring that the tables in a multiple table relational table database are consistent

Record one each of every field in a database relating to one entity in the database; in a relational table database this is a row

Row see Record

Search engine a program that carries out searches on data

Sort alter the order of the records in a file so that one field is in alphabetic, date or numeric order throughout. After a sort, each record normally has a new record number. There are two types of sort — *ascending* in which the order of alphabetic fields goes space, A-z or the order of numeric fields goes minus values through 0 (zero) upwards and *descending* in which the order is reversed. Descending is usually only used when you want the most recent date in a date field or the largest figure in a numeric field to appear in the first record of a file.

Sort order all computers think that A-Z comes before a-z; so many databases will put MacShane before Macintosh. Some databases are 'fixed' to use the order AaBbCcDd ... so that this does not happen. It is usually obvious which order your database uses and more advanced databases allow you to choose the order.

Table a single file of entities and their attributes in a relational table database

Text a type of data — the same as Alphanumeric

Wildcard a character you can use instead of one or more specific characters to identify a group of similar entries in a field.

Some notes on searching

Though advanced search engines can find ‘MacIntosh’ when you ask for ‘Macintosh’, most are very fussy about what they will find.

The simplest find is a perfect match or ‘=’. You can also find anything which is not a perfect match or ‘not equal’ but different search engines use different symbols for expressing this — so check the manual or online help.

You can find a group of records with similar entries in a particular field by using the wildcards *, and ?; for example, if you wanted to find everyone whose surname started with "O", you could ask for ‘SURNAME="O'*"’ and it would come up with O’Donnell, O’Donovan, etc. (if they were in that field of the database).

Some databases also allow you to search for a string anywhere in a field; for example, you might want to find everyone called ‘McDonald’ and ‘MacDonald.’ In this case you could enter a query such as ‘SURNAME LIKE "%Donald"’.

You can use the ‘<’ and ‘>’ signs because every character has its own number — the space has the lowest number followed by the numerals, followed by the upper case letters followed by the lower case letters. So anything less than M (<"M") means space, all the numerals and A–L; anything more than M (>"M") means N–Z and a–z. Note that, when searching on an alphanumeric (or character or text) field, you normally have to put " " round whatever you are searching on.

In a small database you can usually use just the first letter, or the first few letters, of the word you are searching for. In a large database, make indexes and use them when searching.

In a simple database, you can only search on one field at a time, for example, ‘SURNAME="Jones"’. In more advanced databases, you can index on more than one field and search on several fields at once, like ‘SURNAME="Jones" and FORENAME="Davy"’.

References

Hickson, I. (editor). (2023, 11 September). HTML: living standard. <https://html.spec.whatwg.org/>.

The document is licensed under the
Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International

