

Databases vs. Spreadsheets

Think of a data base.. what are the fist things that come to mind?

How about your address book? your address book is probably digital and might look something like this

Address Book as a Spreadsheet

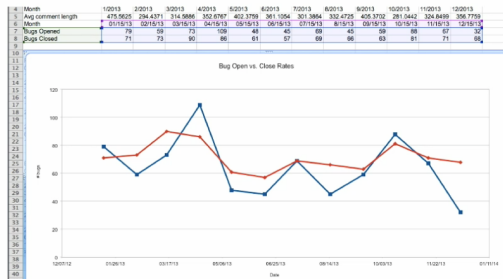
First Name	Last Name	Mobile #	Home #	Address	Birthday
Sheeri	Cabral	121-5555	555-1212	123 Main Street, Boston MA	Sept 17, 1978
Tony	Cabral	121-6666	555-1212	123 Main Street, Boston MA	Jan 27, 1975
John	Smith	653-1210	339-8482	500 College Ave, Boston MA	April 7, 1998
Paul	Thompson	586-7987	339-8482	500 College Ave, Boston MA	May 27, 1996
Camille	Durand	871-5719	485-2390	10 Wharf Road, Provincetown MA	Mar 23, 1970
Noelle	Durand	395-6161	485-2390	10 Wharf Road, Provincetown MA	July 6, 1960
Raj	Sharma	168-5223	626-5976	2 Harrison Street, San Francisco CA	Aug 31, 1980
Priya	Sharma	474-5391	626-5976	2 Harrison Street, San Francisco CA	Dec 4, 1979

Storing your address in a spread sheet is very convenient but it is important to know that a data base is not a spread sheet

Similarities

- Store data in columns and rows
- Calculate new data
- Multi-user

Both data base and spread sheets store data in rows and columns. Both can calculate data based on the existing data & both can allow more than one person to view and edit the data



A spread sheet may have a lot of fancy tools but at its heart it is an electronic ledger. It is a digital worksheet that can easily make nice graphs and charts.

Databases vs. Spreadsheets

Spreadsheets

- Data stored in a cell
- Calculation or value
- Contains formatting
- Can move, sort data

In a spread sheet data is stored in a cell, a cell can have raw data like the number 1 or the string “hello”.

A cell can also have a calculation and that calculation can be stand alone or refer to other cells. A cell also contains formatting information like the font size or highlighting. You can format the type of data you intend to see in a cell like specifying that numbers have two decimal places. You can manipulate the data in a spread sheet by visually moving or sorting The data.

Databases vs. Spreadsheets

Spreadsheets

- Data stored in a cell
- Calculation or value
- Contains formatting
- Can move, sort data

Databases

- Data stored in a record
- Values stored
- Formatting and calculations are applied later
- Sorting is a display issue

A Database on the other hand stores data in a record, and only stores values. Unlike a spreadsheet formatting information like font size highlighting, and even sort order are not stored in a Database record. Usually that information is completely separate. Part of a program that displays the data including sorting. By separating the data from the formatting databases allow you to manipulate data with worrying that you are changing the underlying Data.

Address Book as a Spreadsheet

First Name	Last Name	Mobile #	Home #	Address	Birthday
Sheeri	Cabral	121-5555	555-1212	123 Main Street, Boston MA	Sept 17, 1978
Tony	Cabral	121-6666	555-1212	123 Main Street, Boston MA	Jan 27, 1975
John	Smith	653-1210	339-8482	500 College Ave, Boston MA	April 7, 1998
Paul	Thompson	586-7987	339-8482	500 College Ave, Boston MA	May 27, 1996
Camille	Durand	871-5719	485-2390	10 Wharf Road, Provincetown MA	Mar 23, 1970
Noelle	Durand	395-6161	485-2390	10 Wharf Road, Provincetown MA	July 6, 1960
Raj	Sharma	168-5223	626-5976	2 Harrison Street, San Francisco CA	Aug 31, 1980
Priya	Sharma	474-5391	626-5976	2 Harrison Street, San Francisco CA	Dec 4, 1979

For example lets say that have our address book stored as a spread sheet. We just finished entering data and want to sort by last name so we can easily find an entry when we need it.

Address Book Sorted on Last Name

First Name	Last Name	Mobile #	Home #	Address	Birthday
Sheeri	Cabral	121-5555	555-1212	123 Main Street, Boston MA	Sept 17, 1978
Tony	Cabral	121-6666	555-1212	123 Main Street, Boston MA	Jan 27, 1975
John	Durand	653-1210	339-8482	500 College Ave, Boston MA	April 7, 1998
Paul	Durand	586-7987	339-8482	500 College Ave, Boston MA	May 27, 1996
Camille	Sharma	871-5719	485-2390	10 Wharf Road, Provincetown MA	Mar 23, 1970
Noelle	Sharma	395-6161	485-2390	10 Wharf Road, Provincetown MA	July 6, 1960
Raj	Smith	168-5223	626-5976	2 Harrison Street, San Francisco CA	Aug 31, 1980
Priya	Thompson	474-5391	626-5976	2 Harrison Street, San Francisco CA	Dec 4, 1979

So we sort by last name but we did not sort properly, the last names were sorted but not in the context of the rest of the row. So John Smith became John Durand and Noelle Durand became Noelle Sharma. Because the display information is separate from the data in a data base. Making a sorting mistake like this does not actually change the underlying data like it does in a spread sheet.

Data Type - Dates

- **Spreadsheets**

Can store non-dates

In a Data base you can give the data in a column a type, for example: You can say that information in the birthday column is a date. In a spread sheet we can put anything we want in a cell, a calculation or a word and the spread sheet allows that information in there.

Data Type - Dates

- **Spreadsheets**

Can store non-dates

- **Databases**

Date types

Non-dates cannot be stored

In a Data base if we define a column as having a date type it will not allow another type of data like a string.

Databases vs. Spreadsheets

Spreadsheets

- Every operation makes more data
- +1 to a column, in a new column

In a spreadsheet everything you do results in data, if you do a calculation in one cell, that calculation lives in another cell.

Databases vs. Spreadsheets

Spreadsheets

- Every operation makes more data
- +1 to a column, in a new column

Databases

- Data operations done after retrieval
- Calculations, formatting not stored
- Eliminate duplicate information

In a data base you can do a calculation on a record and just show the result without adding more data to the database. One of the biggest benefits of a database is you can eliminate duplicate information. For example

Address Book as a Spreadsheet

First Name	Last Name	Mobile #	Home #	Address	Birthday
Sheeri	Cabral	121-5555	555-1212	123 Main Street, Boston MA	Sept 17, 1978
Tony	Cabral	121-6666	555-1212	123 Main Street, Boston MA	Jan 27, 1975
John	Smith	653-1210	339-8482	500 College Ave, Boston MA	April 7, 1998
Paul	Thompson	586-7987	339-8482	500 College Ave, Boston MA	May 27, 1996
Camille	Durand	871-5719	485-2390	10 Wharf Road, Provincetown MA	Mar 23, 1970
Noelle	Durand	395-6161	485-2390	10 Wharf Road, Provincetown MA	July 6, 1960
Raj	Sharma	168-5223	626-5976	2 Harrison Street, San Francisco CA	Aug 31, 1980
Priya	Sharma	474-5391	626-5976	2 Harrison Street, San Francisco CA	Dec 4, 1979

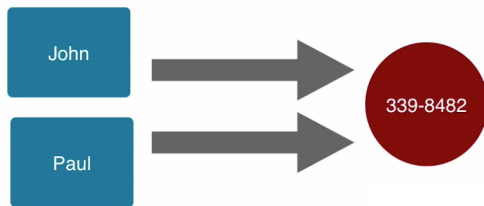
In Our address book we have eight people each of these peoples lives with another person. So the address and phone numbers are listed twice, what happens if we make a mistake in one of the cells.

Address Book as a Spreadsheet

First Name	Last Name	Mobile #	Home #	Address	Birthday
John	Smith	653-1210	339-8481	500 College Ave, Boston MA	April 7, 1998
Paul	Thompson	586-7987	339-8482	500 College Ave, Boston MA	May 27, 1996

In this example John and Paul are roommates living at the same address, but have two different home phone numbers. If you saw this you would know that one of the phone numbers is wrong but you would not know which of the numbers is wrong.

In a Database



In a Database it is very easy to put the names in one place and say the John and Paul both have the same home phone number. the same can be done with the address so you only have to edit one Database record. If a family moves and your address book is in a spreadsheet you have to update each address cell, because of this databases are more efficient at storing data.

Databases are More Efficient

Changing data is easier and less error prone in a database, spread sheets are great for ledger type activities and lists. Like showing bank transactions and calculating an account balance, they are also good for crating charts and graphs of data.

Databases are More Efficient

- Can give the data a type

Databases are good for storing lots of information, you can specify that the data is a certain type.

Databases are More Efficient

- Can give the data a type
- No redundancy for formatting differences
- Less error-prone

Databases separate the data from how it is displayed which can eliminate duplication and reduce data errors.

Databases are More Efficient

- Can give the data a type
- No redundancy for formatting differences
- Less error-prone
- Other users do not change formatting/sorting

Databases are also good at allowing more than one person at a time to look at and change the data without disturbing each other, learning a but database is worth the effort

Relational Databases vs. Flat Files

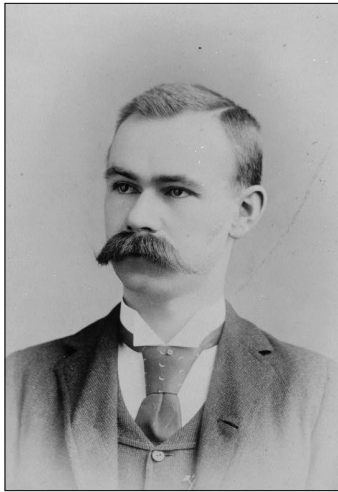
Flat File Model

	Route No.	Miles	Activity
Record 1	I-95	12	Overlay
Record 2	I-495	05	Patching
Record 3	SR-301	33	Crack seal

A flat file database is a database which is stored on its host computer system as an ordinary "flat file". To access the structure of the data and manipulate it, the file must be read in its entirety into the computer's memory. Upon completion of the database operations, the file is again written out in its entirety to the host's file system. In this stored mode the database is "flat", which means it has no structure for indexing and there are usually no structural relationships between the records. A flat file can be a plain text file or a binary file.

A flat file database is a database which is stored on its host computer system as an ordinary "flat file". To access the structure of the data and manipulate it, the file must be read in its entirety into the computer's memory. Upon completion of the database operations, the file is again written out in its entirety to the host's file system. In this stored mode the database is "flat", which means it has no structure for indexing and there are usually no structural relationships between the records. A flat file can be a plain text file or a binary file.

The term has generally implied a small, simple database. As computer memory has become cheaper, larger and more sophisticated databases can now be held in memory in their entirety for faster access, but these would not generally be referred to as flat-file databases.



Inventor of flat database

Herman Hollerith conceived the idea that data could be represented by holes punched in paper cards then tabulated by machine. He implemented this concept for the US Census Bureau; thus the 1890 United States Census processing created the first database—consisting of thousands of boxes full of punched cards.

Herman Hollerith conceived the idea that data could be represented by holes punched in paper cards then tabulated by machine. He implemented this concept for the US Census Bureau; thus the 1890 United States Census processing created the first database—consisting of thousands of boxes full of punched cards.

Hollerith's enterprise grew into the computer giant IBM, which dominated the data processing market for most of the 20th century. IBM's fixed-length field,

[illegible]



Similarities

- Show and manipulate data

There are some similarities between flat files and relational databases. Both can be manipulated to show information in a new form without changing the original data

Similarities

- Show and manipulate data
- Stored data does not change

Both allow you to extract data out of certain fields.

Similarities

- Show and manipulate data
- Stored data does not change
- Extract data from fields

Both allow you to extract data out of certain fields.

Similarities

- Show and manipulate data
- Stored data does not change
- Extract data from fields
- Matching and filtering

And both have powerful matching and filtering capabilities.

Similarities

- Show and manipulate data
- Stored data does not change
- Extract data from fields
- Matching and filtering
- Search and replace

Almost all text editors have some kind of global search and replace functionality and relational databases can do that too. Depending on your text program, you might have formatting capabilities like font size, and highlighting.

Maybe you have a particular phone number in big bold letters so the number is very visible for when you want to call it.

Text files

- No data types

A text file allows you to put anything into it. We can put a string, like Hello There, or we can put a number like 172, and the program won't complain, even if it's supposed to be the field for a date.

Text files

- No data types
- Field separation manually enforced

If we want to be able to use tools to manipulate the data, we might have some rules. If we are using a tab-delimited file, the rule is that the pieces of data are separated by tabs and that they appear in a a certain order.

There's also nothing stopping us from using a tab to separate the data on one line, but using a comma to separate the data on the next.

Text files

- No data types
- Field separation manually enforced

The program does not complain, but when we try to use tools to get the data we want, we won't be able to get our proper data. A database will separate the fields for us, a flat file does not. We have to enforce a structure if we want our text file to work in a certain way. Like a spreadsheet, text editors have display information stored with the data itself.

Text files

- No data types
- Field separation manually enforced
- For writing
- For layout
- For moving text around

Text editors are very good for writing text, laying out text on a page, and moving text around. You can cut and paste and easily move a word, paragraph or even several pages of text in a text editor. They are designed for storing, viewing and changing documents that consist mostly of text.

Text files

- No data types
- Field separation manually enforced
- For writing
- For layout
- For moving text around

A text or spreadsheet file is sometimes called a flat file especially when comparing to a relational database. This is because a relational database is build to highlight how some data relates to other data.

Databases vs. Spreadsheets

Flat Files

- Writing text
- Editing text
- Moving text around

Flat files are great for writing and editing text. They are also good for moving lots of text around.

Databases vs. Spreadsheets

Flat Files

- Writing text
- Editing text
- Moving text around

Databases

- Data types

Databases are good for storing information that you want to be a certain type.

Databases vs. Spreadsheets

Flat Files

- Writing text
- Editing text
- Moving text around

Databases

- Data types
- Data separate from display

Databases separate the data from how it is displayed

Databases vs. Spreadsheets

Flat Files

- Writing text
- Editing text
- Moving text around

Databases

- Data types
- Data separate from display
- Eliminate duplicate information
- Reduce data errors

Which can eliminate duplicate information and reduce data errors. Text / spreadsheet files are handy, but a database is a much more powerful tool when manipulating data.