
Databases

Overview

Definition

- A database is really any collection of related data.
- It could be a spreadsheet, a contact list or a full-blown relational database system like Oracle or Sql server.

Uses of Databases

- Storing business transaction data
- Storing scientific data
- Storing and managing medical data
- Content management for web pages
- Managing and tracking back-end transactions such as virtual machines in the cloud
- Storing and managing metadata

Types of Databases

- Even a rolodex is a database, though we usually think of electronic, computer based databases.
- A rolodex was a way of keeping addresses and phone numbers



Flat Files

Flat files are just files of data with each piece of data separated by a delimiter such as a comma, or given a fixed width.

```
|LastName,FirstName,EAddress,City,Zipcode,phone,Email,HireDate
Jordan,Mary,2002 South Mercer Street,Seattle,98100,2065558828
Danner,Thomas,100 Boardwalk South,Seattle,98100,2065551001,Da
Terrance,Sarah,202 Rt 3,Kent,98211,3605550128,Terrance.Sarah@
Camlin,Lenny,Somewhere Ave,Tukwilla,98101,2535551010,Camlin.L
Sanderson,Lewis,101 Elsewhere Avenue,Seattle,98100,2065550987
Able,Tina,1000 West Blv,Tukwilla,98101,3605552020,Able.Tina@m
Evans,Karl,11 North Hill street,Tukwilla,98101,2535558998,Eva
Zukof,Bob,WaterFront Blvd.,Bellevue,98301,3605552435,Zukof.Bo
Martinez,Patrick,EastLake Blvd.,Seattle,98100,2065553679,Mart
Patterson,Lisa,1010 Binary Drive,Kent,98211,3605551100,Patter
Franklin,Bernice,222 ITC Road,Seattle,98100,3605552221,Frankl
Kim,Susan,111 Martin Luther King Way,Seattle,98100,2065550742
```

Flat Files Continued

Flat files are useful for transferring data from one database system to another or for storing simple data that a program may need to operate.

A special type of flat file database is a spreadsheet. Spreadsheet programs like Excel are still flat file because they store their data in a single file, but they have many other features for analysing, sorting and querying the data.



Excel

Excel ribbon interface showing the HOME tab. The ribbon includes tabs: FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA, REVIEW, VIEW, LOAD TEST, and TI. The HOME tab is active, displaying groups: Clipboard (Paste, Cut, Copy, Format Painter), Font (Font face: Calibri, Size: 11, Bold, Italic, Underline, Text color, Background color), Alignment (Text alignment, Orientation, Merge & Center), and Numbers (Number format: General). The formula bar shows the active cell D22.

	A	B	C	D	E	F	G	H	I
1	STATION	STATION_NAME	ELEVATION	LATITUDE	LONGITUDE	DATE	TPCP	EMXT	EMNT
2	COOP:310301	ASHEVILLE NC US	682.1	35.5954	-82.5568	201001	413	59	10
3	COOP:310301	ASHEVILLE NC US	682.1	35.5954	-82.5568	201002	328	61	17
4	COOP:310301	ASHEVILLE NC US	682.1	35.5954	-82.5568	201003	286	73	25
5	COOP:310301	ASHEVILLE NC US	682.1	35.5954	-82.5568	201004	149	89	37
6	COOP:310301	ASHEVILLE NC US	682.1	35.5954	-82.5568	201005	265	85	40
7	COOP:310301	ASHEVILLE NC US	682.1	35.5954	-82.5568	201006	272	95	57
8	COOP:310301	ASHEVILLE NC US	682.1	35.5954	-82.5568	201007	229	95	62
9	COOP:310301	ASHEVILLE NC US	682.1	35.5954	-82.5568	201008	294	95	60
10	COOP:310301	ASHEVILLE NC US	682.1	35.59528	-82.5567	201009	325	90	48
11	COOP:310301	ASHEVILLE NC US	682.1	35.59528	-82.5567	201010	99	82	33
12	COOP:310301	ASHEVILLE NC US	682.1	35.59528	-82.5567	201011	348	72	23
13	COOP:310301	ASHEVILLE NC US	682.1	35.59528	-82.5567	201012	156	55	8

Excel

AutoSave Off | JacquieMusicVideo.xlsx | Search | JACQUELINE George

File Home Insert Draw Page Layout Formulas Data Review View Help

Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A A⁺ B I U Merge & Center General \$ % 0.00 0.00 Conditional Formatting Format as Table Cell Styles Insert Delete Format Sort & Filter Find & Select

E1 : X ✓ f_x Format

	A	B	C	D	E	F	G	H	I
1		First Name	Last Name / Band	Title	Format				
2	Music		10,000 Maniacs	Few and Far Between	CD				
3			10,000 Maniacs	In My Tribe	CD				
4			10,000 Maniacs	MTV Unplugged	CD				
5			10,000 Maniacs	Our Time in Eden	CD				
6			ABBA	The Best of ABBA	CD				
7			Acoustic Alchemy	Reference Point	CD				
8			Across the Universe	Across the Universe Soundtrack	Digital				
9		Bryan	Adams	MTV Unplugged	CD				
10		Ryan	Adams	Gold	CD				
11		Ryan	Adams	Heartbreaker	CD				
12		Ryan	Adams	Love is Hell	CD				
13			Adele	21	CD				
14			A-Ha	Scoundrel Days	CD				
15			Air Supply	Lost in Love	Vinyl				
16		William	Albright	The Complete Rags of Scott Joplin	Digital				
17		Herb	Alpert & Lani Hall	Steppin' Out	CD				
18		Herb	Alpert & the Tijuana Brass	Going Places	Vinyl				
19		Herb	Alpert & the Tijuana Brass	Sergio Mendes & Brazil '66	Vinyl				
20		Herb	Alpert & the Tijuana Brass	SRO	Vinyl				
21		Herb	Alpert & the Tijuana Brass	The Beats of the Brass	Vinyl				
22		Herb	Alpert & the Tijuana Brass	What Now My Love	Vinyl				
23			Amelle	Soundtrack to the Motion Picture	CD				
24			America	History: America's Greatest Hits	CD				
25			American Beauty	Soundtrack to the Motion Picture	CD				
26		The	Angel Brothers	Forbidden Fruit	Digital				
27		Darol	Anger	Diary of a Fiddler	CD				

Music Audiobooks Classical Film/Video

Excel

JacqueMusicVideo2.csv - Notepad

File Edit Format View Help

```
,First Name,Last Name / Band,Title,Format
Music,, "10,000 Maniacs",Few and Far Between,CD
,, "10,000 Maniacs",In My Tribe,CD
,, "10,000 Maniacs",MTV Unplugged,CD
,, "10,000 Maniacs",Our Time in Eden,CD
,,ABBA,The Best of ABBA,CD
,,Acoustic Alchemy,Reference Point,CD
,,Across the Universe,Across the Universe Soundtrack,Digital
,Bryan,Adams,MTV Unplugged,CD
,Ryan,Adams,Gold,CD
,Ryan,Adams,Heartbreaker,CD
,Ryan,Adams,Love is Hell,CD
,,Adele,21,CD
,,A-Ha,Soundrel Days,CD
,,Air Supply,Lost in Love,Vinyl
,William ,Albright,The Complete Rags of Scott Joplin,Digital
,Herb,Alpert & Lani Hall,Steppin' Out,CD
,Herb,Alpert & the Tijuana Brass,Going Places,Vinyl
,Herb,Alpert & the Tijuana Brass,Sergio Mendes & Brazil '66,Vinyl
,Herb,Alpert & the Tijuana Brass,SRO,Vinyl
,Herb,Alpert & the Tijuana Brass,The Beats of the Brass,Vinyl
,Herb,Alpert & the Tijuana Brass,What Now My Love,Vinyl
,,Amelie,Soundtrack to the Motion Picture,CD
,,America,History: America's Greatest Hits,CD
,,American Beauty,Soundtrack to the Motion Picture,CD
```

Advantages/Disadvantages

Flat files are lightweight making them easy to use and transport.

They are generally just text files and don't require any special software.

They are generally humanly readable.

No data type enforcement

Redundancy--data is often repeated in multiple places.

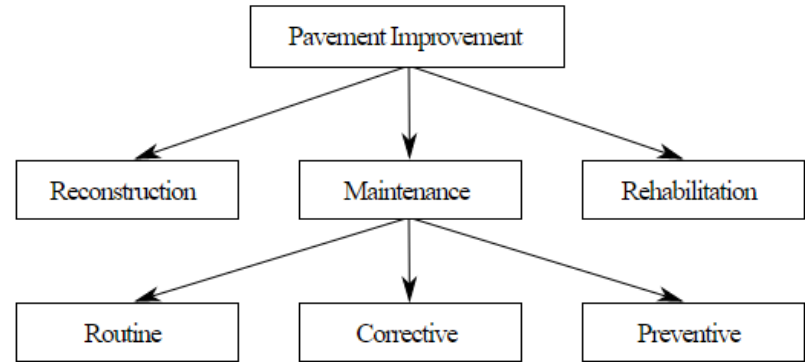
Hard to query data, especially across multiple files.

Hierarchical Databases

Hierarchical database store their data in a tree structure that resembles a file system. There is a root or trunk and then branches of data. Branches can have sub branches (Just as folders can contain folders.)

Hierarchical database were used a lot in the early days of computerized databases, especially by large institutions such as banks and insurance companies. Many still maintain their large hierarchical databases simply because they have so much money and data invested in them.

Hierarchical Model



Advantages/Disadvantages

Easy to Navigate, just like navigating file folders

Relatively fast

Redundancy--the need to store the same information in more than one place

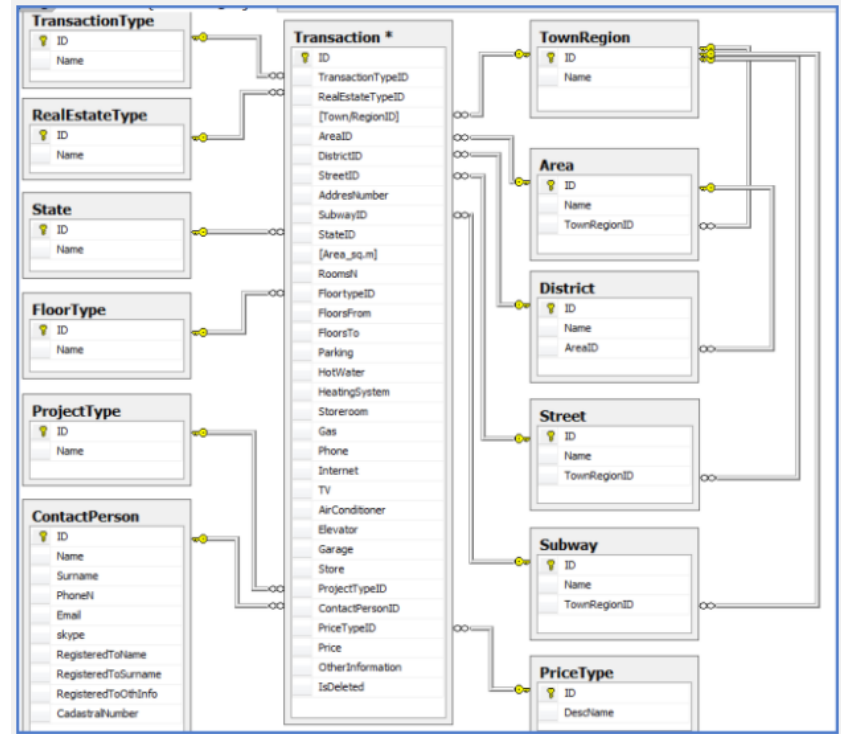
Difficult to compare data across folders

Relational Databases

Relational databases originated in the 1970's.

For the last 40+ years they have dominated the database market.

Relational databases divide the data into separate tables. Information is only entered in one place. Tables relate to each other by means of keys. The key field in one table (the primary key) is repeated in another table to make the connection between the data.



Relational Databases

Relational databases were developed to solve problems that plagued early database designs such as the hierarchical and network designs

These problems included

- Redundancy--data repeated in more than one place
- Difficulties in comparing and relating data from different nodes
- Scalability

Types of Relationships

There are only three types of relationships between entities in a relational database:

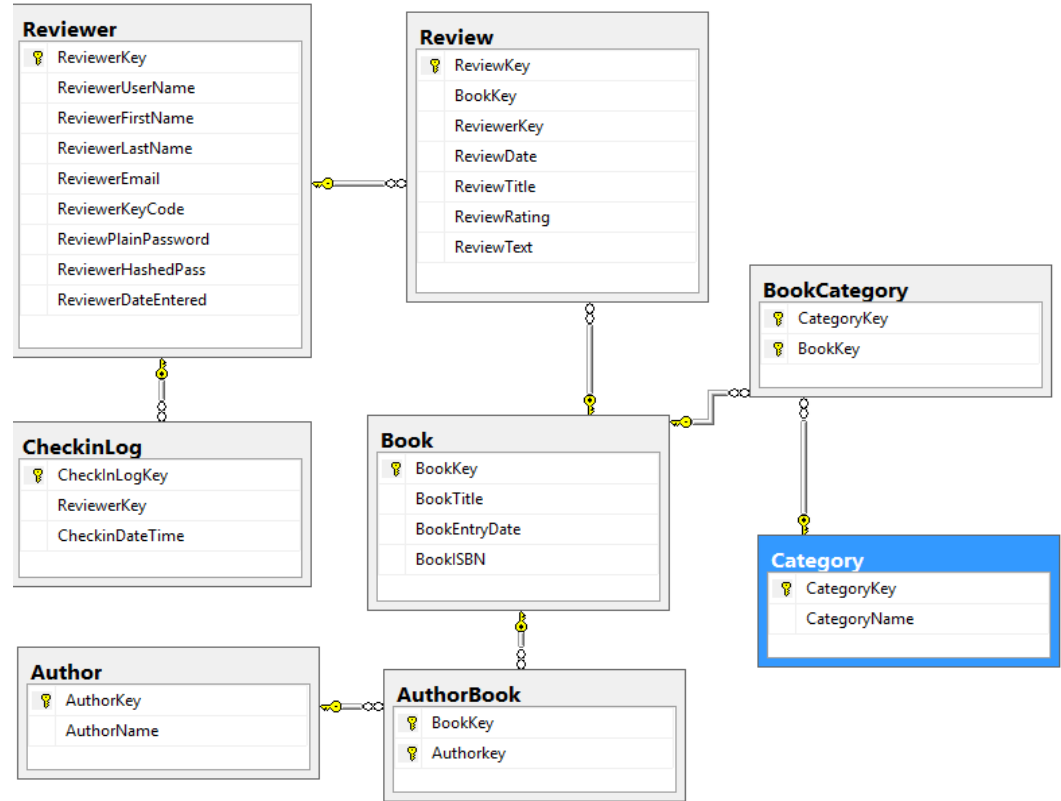
1. **One to one:** This means for every record in Table1, There is exactly one related record in Table2. This is a rare relationship.
2. **One to Many:** For every record in Table1, there can be zero to any number of related records in Table2. This is the relationship of 99% of all tables in a relational database.
3. **Many to Many:** This says that Table1 can have many related records in Table2, and Table2 can have many related records in table one. This relationship cannot be resolved by a relational database. You must make a Linking table that converts the many to many relationship into two one to many relationships.

Database Design

All relational databases regardless of the product share the same design principles

- Everything must be broken into **entities**(tables)
- Entities consist of **attributes** (columns) that describe that entity
- Every entity should have a **primary key** that uniquely identifies each row
- That primary key can be repeated in other tables, showing a relationship between the tables. A primary key repeated in another table is called a **Foreign Key**
- **Foreign keys** are the only redundancy (repeated values) allowed in the relational database

Relational Diagram



Relational Tables

Note the Primary key for Author
Is related to the author key in
BookAuthor. The primary Key
For Book is related to the BookKey
In AuthorBook

100 %

Results Messages

	AuthorKey	AuthorName
1	1	David Foster Wallace
2	2	Michel Foucault
3	3	Jon Duckett
4	4	Brian Green
5	5	Pablo Neruda
6	6	Klm Stanley Robinson
7	7	Don Delilo
8	8	Gilles Delueze

	BookKey	Authorkey
1	1	8
2	1	9
3	2	7
4	3	8
5	4	4
6	5	5
7	6	6
8	7	1

	BookKey	BookTitle	BookEntryDate	BookISBN
1	1	What is Philosophy	2015-10-11 16:02:13.417	0231079893
2	2	Underworld	2015-10-11 16:02:13.417	0965664120
3	3	JavaScript and J...	2015-10-11 16:02:13.417	97811185...
4	4	The elegant Univ...	2015-10-11 16:02:13.417	0965088806
5	5	All the Odes	2015-10-11 16:02:13.417	97803741...
6	6	Shaman	2015-10-11 16:02:13.417	97818414...

Advantages/Disadvantages

Relational databases offer relational integrity, meaning data is only ever entered and updated in one place. This prevents many errors.

Also relational database are easy to query. Any piece of data can be related to any other piece of data.

The primary drawback is complexity. Relational database divide data into multiple tables which are not necessarily humanly readable.

It is fairly easy to make a bad database that will not fulfill its business requirements.