

Lesson 8 Outline - Databases

- I. Applications of Database
 - A. Traditional uses
 - 1. Product information retrieval
 - 2. Customer support center (aka knowledge database)
 - 3. Product reviews
 - 4. Data mining (gathering data/statistics from large repositories)
 - 5. Surveys/testing
 - B. eCommerce
 - 1. Product inventory/catalog
 - 2. Order logs
 - 3. Customer accounts
 - 4. Booking systems
 - C. Cutting edge interactivity
 - 1. “my” sites – i.e. customize your website based on customer preferences
 - 2. Suggestive selling (recommend items, etc)
 - 3. Tracking systems

- II. Types of Databases (much of this is a summary of Chapter 2 in *Programming the Perl DBI* by Alligator Descartes & Tim Bunce – worth checking out from your library!)
 - A. Flat file
 - 1. Typically consists of a plain text file (or a group of files), viewable by a regular text editor
 - 2. Data is written in a standard format such as a pipe-delimited string with one record per line
 - 3. Pros
 - a. Fast to implement for small, well defined data sets
 - b. Code required is simple
 - c. Total control over data file formats
 - d. Easy to read datafile (via plain text editor)
 - 4. Cons
 - a. Scalability problems – algorithms to search/edit the datafile are sequential, which means the time required increases in direct proportion to the size of the dataset
 - b. Concurrent access to the database (i.e. two people searching/editing the file at the same time) must be implemented manually by using file locking (ex. the perl flock() function)
 - c. To grant access to a CGI script, the datafile must (usually, based on the web user ID) be world-writeable (be sure to move it outside of the web directory for increased security!)
 - B. DBM file
 - 1. A programming layer that allows you to store information in an “indexed” format – i.e. a pair of strings, one being the “key” and one being the “value”. This is the same concept as a perl hash.
 - 2. Stored as binary files (which means they are not readable via a text editor)
 - 3. There are several flavors of DBM files (ndbm, db, gdbm, sdbm, odbm) available in perl
 - 4. The DBM file on the server is associated with a perl hash variable
 - 5. Pros
 - a. More efficient data retrieval by using the key
 - b. Easy to program since the interface is just a perl hash variable (all file-related operations are abstracted away from the programmer)
 - 6. Cons
 - a. Manual file locking is still required
 - b. Can be tricky to store complex data sets

- c. You can only search the database by the single key – this means you can't (easily) do sophisticated queries like “show me all products that cost less than \$2.00” without doing a sequential search through all entries in the database. This gets you back to the scalability problems inherent in flat-file databases
 - d. To grant access to a CGI script, the datafile must (usually, based on the web user ID) be world-writable (be sure to move it outside of the web directory for increased security!)
- C. Relational databases – SQL (Structured Query Language)
- 1. Uses tables consisting of columns and rows to store the data.
 - 2. Efficiently supports complex searches of large datasets
 - 3. Datafiles are not viewable via the command line interface; you must be inside the database application to view or edit the data
 - 4. Pros
 - a. No longer restricted by sequential searches; SQL can search on any column in any table
 - b. No file I/O is required to search or edit the dataset, instead a direct connection to the SQL software is used (this is much faster than file I/O)
 - c. No world-writable/readable datafiles exist on the ISP
 - d. Concurrent access to the database is handled by the application (although there are limits on how many concurrent accesses can be allowed)
 - 5. Cons
 - a. Design of the database is crucial and non-trivial; efficiency is directly related to the way the tables have been designed
 - b. Requires that you have the SQL software available on your ISP
 - c. Data is not viewable from any outside means, which can be harder to debug
 - d. The software required is significantly more complex than for flat-file or DBM file databases
 - 6. There are many flavors of SQL databases. They range from free open-source programs to commercial applications costing hundreds of thousands of dollars. The functionality also varies. Some of the more popular ones are listed below.
 - a. mSQL
 - b. MySQL
 - c. Microsoft Access
 - d. Microsoft SQL Server
 - e. Oracle
 - f. Sybase

III. Interfacing to SQL Databases from the WWW

- A. Publish-only access can be achieved from within your database program
- B. Client-side interactivity
 - 1. Scripting languages – JavaScript, VBScript
 - 2. DORMAT – Java, ActiveX
 - 3. Requires that browser supports the desired language
- C. Server-side interactivity
 - 1. Traditional CGI languages
 - a. Perl – DBM is built-in; DBI module facilitates SQL access
 - b. ASP – a natural with Access (both are Microsoft, NT only)
 - c. PHP – usually paired with MySQL (both are open source, very portable)
 - 2. Server API (NSAPI, ISAPI)
 - 3. More compatible with the general browser
- D. Deciding on a platform depends on many criteria
 - 1. Availability of both database and language (check with your ISP!)
 - 2. Features required (See below)
 - 3. Compatibility with other systems

One of the additional resources listed on the website is a database comparison wizard (“crash-me”) from TCX. Here are some interesting numbers comparing the more popular databases for the WWW. This information is taken from <http://www.mysql.com/information/crash-me.php>. The full comparison also includes specific details about SQL feature implementation.

Other limits					
	MySQL	Access	Microsoft SQL Server	mSQL	Oracle
Simultaneous connections	101	64	1000	200	41
Value of logical operation (1=1)	1	-1	not supported	not supported	not supported
Columns in table	2364	255	1024	75	1000
max table row length (without blobs)	65534	2025	8036	19125	255000
table row length with nulls (without blobs)	65501	2025	8036	19125	255000
query size	1048574	16777216	16777216	131069	16777216

This information is provided by TCX so one can get the real limitations from the database server (not the information from sales managers!). Hopefully the above information will make it easier for you to find a database server that has the functionality you need and that you can rely on!