Chapter Ten

Log Analysis

One of the benefits of utilizing CGI in your Web site is the ability of a CGI script to gather and store information about your customers and what they have been doing. Two particularly important types of information are who accessed your site and when and what types of errors, if any, they found.

The Web store application attempts to document this information in two files located in the **Admin_files** subdirectory: **error.log** and **access.log**. These two files are used by the application to store information that might be useful for you in debugging or upgrading the Web store.

For example, by analyzing the access log, you might discover that you tend to be hit at certain peak hours and from certain locations. With this information you might adjust your backup schedules or the content of your pages, or even streamline server speed by getting together with your sysadmin to compare peak statistics. Similarly, by analyzing the error log, you can quickly determine what types of errors may be occurring and/or if you are being hacked and by whom.

This chapter will take a brief look at the log files and discuss how you can utilize the log analysis script to use them to the best of your advantage.

The Access Log

The **access.log** file generates server-known statistics about every customer who accesses your store. Specifically, it appends the current time and date to the list of HTTP environment variables known by the server.

The environment variables are shown in Table 10.1:

Table 10.1 Environment Variables

GATEWAY_INTERFACE	(What type of gateway interface you are using, probably CGI)
DOCUMENT_ROOT	(The directory that your Web Server sees as root)
REMOTE_ADDR	(IP address of your customer)
SERVER_PROTOCOL	(Protocol used by the server)
REQUEST_METHOD	(GET or POST)
REMOTE_HOST	(Domain name of your customer)
QUERY_STRING	(The information coming in as form data)
HTTP_USER_AGENT	(The type of browser your customer uses)
PATH	(The Web server's known paths)
TZ	(Time Zone)
HTTP_CONNECTION	(Type of HTTP Connection)
HTTP_ACCEPT	(Types of Documents allowed by your server)
SCRIPT_NAME	(Location of **web_store.cgi**)
SERVER_NAME	(Your web server's URL)
SERVER_PORT	(The port your Web server is running on)
HTTP_HOST	(The URL of your host)

You can use these variables or others that your specific server is defining to learn more about the use of your store. Although many of the environment variables will be only valuable for specific uses, several can be picked out as broadly useful. These include **DOCUMENT_ROOT**, **REMOTE_ADDR**, **REMOTE_HOST**, **HTTP_USER_AGENT**, and **SCRIPT_NAME**. Further, if your server is enabled with its own authentication environment, you can gather more detailed information about your users through the **REMOTE_USER** variable.

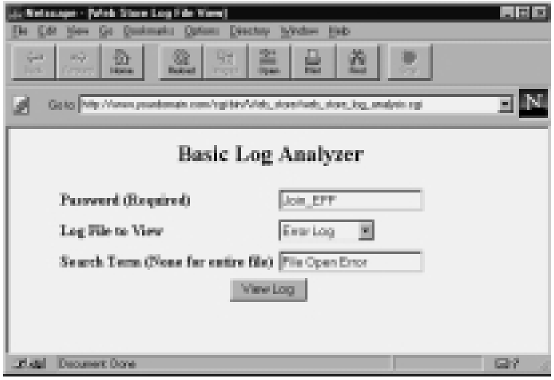
The Error Log

The error log also collects information about the Web server environment and the date and time of the error. Plus, it records the line number that the error occurred on and the type of error. There are currently three types of errors documented by this script:

- **File Open Errors** occur when the script has trouble opening a requested file. This will typically happen if you have set the wrong permissions for files that the Web store application needs, or if you have incorrectly specified a path. Thus, once you have correctly installed the application on your server, you should not get this error. If you do, it almost definitely means that you have modified the permissions accidentally. You should execute **web_store_check_setup.cgi** to determine where the problem is.
- **Page Load Warnings** occur when a customer has attempted to call a page that is restricted. As noted in Chapter 2, the Setup file defines only a limited set of acceptable file extensions to display. If the customer attempts to read a file other than one specified there, the script will die. These are interesting errors to view because they will alert you when someone is trying to manipulate the script for fun or for malice.
- **Randomizing errors** occur when the script has a problem creating a unique cart id for new customers. It is most likely that this error will occur because you have set the wrong permissions for the **User_carts** subdirectory. It is also possible that your server does not support the Perl **rand** function. If this is the case, you should disable this routine by commenting out the code.

The Log Analysis Script

For quick analysis of the log files, we have created the log analysis script which allows you to do simple keyword searching of the log file so that you can compare rows in the log database against each other for similarities. Figure 10.1 shows the Log Analysis query form.



http://www.yourdomain.com/cgi-bin/Web_store/web_store_log_analysis.cgi

Basic Log Analyzer

Password (Required)	Join_EFF
Log File to View	Error Log
Search Term (None for entire file)	File Open Error

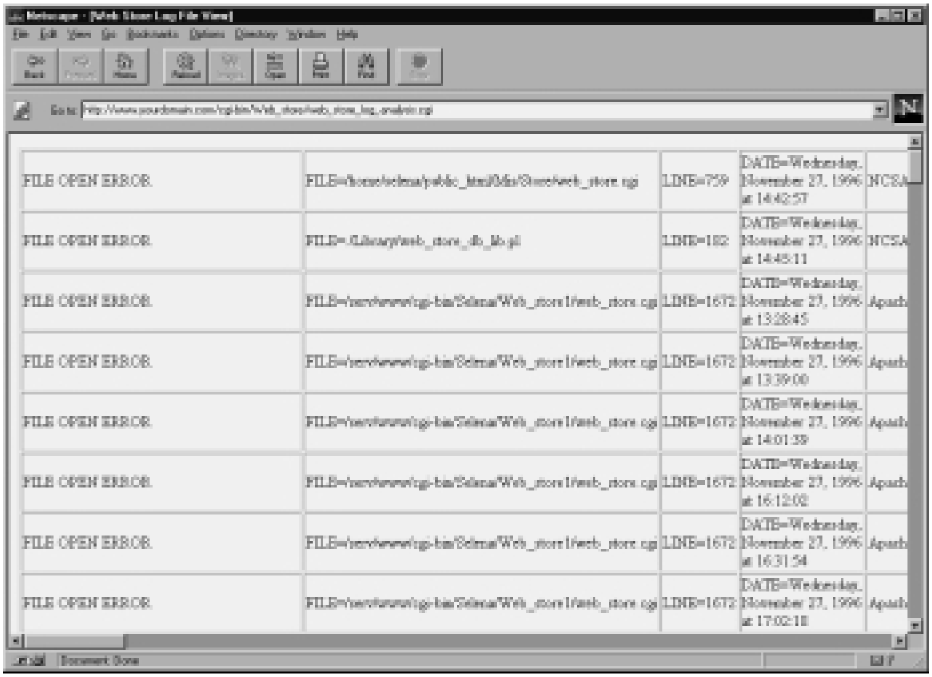
View Log

Figure 10.1 Log Analysis query form.

Rows that satisfy your search criteria are returned as tabular-view database rows with all the fields represented in table cells. Figure 10.2 shows a sample output from a nonkeyword filtered request.

For example, you may want to know how many hits you get from a certain IP address or how many hits you get on a single day. Similarly, you may want to see all the occurrences of Page Load Warnings and determine if they are from a sole source.

To use the log analysis script, you must point your Web browser to the script, enter your desired password, select a log file to review, and a keyword with which to filter.



Netscape - [Web Store Log File View]
File Edit View Go Bookmarks Options Directory Window Help
Back
Forward
Home
Reload
Images
Open
Print
Find
Stop
Go to: http://www.yourdomain.com/cgi-bin/Web_store/web_store_log_analysis.cgi
Document: Done

FILE OPEN ERROR.	FILE=/home/selena/public_html/Sdis/Store/web_store.cgi	LINE=759	DATE=Wednesday, November 27, 1996 at 14:42:57	NCSA
FILE OPEN ERROR.	FILE=./Library/web_store_db_lib.pl	LINE=182	DATE=Wednesday, November 27, 1996 at 14:45:11	NCSA
FILE OPEN ERROR.	FILE=/usr/www/cgi-bin/Selena/Web_store1/web_store.cgi	LINE=1672	DATE=Wednesday, November 27, 1996 at 13:28:45	Apach
FILE OPEN ERROR.	FILE=/usr/www/cgi-bin/Selena/Web_store1/web_store.cgi	LINE=1672	DATE=Wednesday, November 27, 1996 at 13:39:00	Apach
FILE OPEN ERROR.	FILE=/usr/www/cgi-bin/Selena/Web_store1/web_store.cgi	LINE=1672	DATE=Wednesday, November 27, 1996 at 14:01:39	Apach
FILE OPEN ERROR.	FILE=/usr/www/cgi-bin/Selena/Web_store1/web_store.cgi	LINE=1672	DATE=Wednesday, November 27, 1996 at 16:12:02	Apach
FILE OPEN ERROR.	FILE=/usr/www/cgi-bin/Selena/Web_store1/web_store.cgi	LINE=1672	DATE=Wednesday, November 27, 1996 at 16:31:54	Apach
FILE OPEN ERROR.	FILE=/usr/www/cgi-bin/Selena/Web_store1/web_store.cgi	LINE=1672	DATE=Wednesday, November 27, 1996 at 17:02:18	Apach

Figure 10.2 Raw dump of log file.