

Real Estate Reporting v.1 Setup and Configuration Guide

Table of Contents

Overview.....	2
Run Load Request Log Shell Script.....	2
1. Run Analog Script.....	2
2. Run Rotate Logs Script.....	2
3. Run Load Request Log Perl Script	2
Perl Library Dependencies to Running loadRequestLog.pl.....	2
3a. Determine Whether ORACLE_HOME Is Set.....	3
3b. Run Script	3
Modify Load Request Log Variables.....	3
4. Determine Data Load Successful.....	3
Examples of Common Errors Running Scripts.....	4
Run Recover Perl Script for Fatal Errors	4
Support Information.....	5
Appendix A. Analog Configuration File.....	6

Overview

Reports are essential in determining the benefits of web traffic and the impact your online listings have on your business. However, they must be properly set up and configured to get the best results. Your System Administrator can set up and configure the appropriate scripts.

This guide describes in detail the scripts used to take your web data and prepare it to be input correctly into the database. A cron job runs the scripts daily.

Run Load Request Log Shell Script

This shell script (`/opt/MDW/bin/re_report/runLoadRequestLog.sh`) handles the entire job of preparing and inputting the web data into the database. It calls all of the necessary scripts in the following order:

1. Run Analog Script

This shell script (`/usr/local/bin/re_report/runAnalog.sh`) analyzes the web access logs, ignores unnecessary information, and outputs to a different log file in the format specified in the following configuration file:

(`/usr/local/bin/re_report/conf/stats.conf`) (See Appendix A. Analog Configuration File.)

The paths of the logs are as follows:

```
analog: /usr/local/analog
access log: /usr/local/apache/logs/<access_log>
output log file:
/usr/local/bin/re_report/data/<current_month>/<date>.txt.
(For example, a specific output file name is:
/usr/local/bin/re_report/data/Sep/09122003.txt)
```

The `runLoadRequestLog.sh` script then calls the rotate logs Perl script.

2. Run Rotate Logs Script

This Perl script (`/home/scripts/rotatelogs.pl`) rotates the web access logs and gzips them.

The `runLoadRequestLog.sh` script then calls the next script to run.

3. Run Load Request Log Perl Script

This Perl script (`loadRequestLog.pl`) parses the log file and loads it into the `mdw_log_report` schema for reporting.

Perl Library Dependencies to Running loadRequestLog.pl

The following Perl libraries must be installed to run this script:

- DBI
- Oraperl
- IO::File
- IO::Socket

Note: ORACLE_HOME must be set to run this script.

3a. Determine Whether ORACLE_HOME Is Set

To determine if ORACLE_HOME is set:

1. Enter on the command line:
echo \$ORACLE_HOME (If set, the path is returned to the screen; if not, the screen is blank.)
2. If **not** set, enter on the command line:
export ORACLE_HOME=/full/path/to/ORACLE_HOME

Note: The above export command can be added to the shell initialization scripts (.bash_profile, .profile, or .bashrc) to ensure ORACLE_HOME is always set when the System Administrator logs in. Also, /full/path/to is a placeholder for the actual path.

3b. Run Script

To run the script, use the following command line:

```
/opt/MDW/bin/re_report/runLoadRequestLog.sh [*<filename>] [debug] [*<Date>]
```

Note: The asterisk (*) and angle brackets (<>) indicate this command requires you to perform a function, in this case enter a file name or insert a date, instead of using the argument name [debug].

The command takes **0, 1, 2,** or **3** [optional] arguments:

- If **0**, argument: file(s) = \$configPath/<current month>/<date>.txt (The file name becomes the last modified file, and this file name must start with a number [0-9].)
- If **1**, argument: file = the optional argument entered.
- If the [debug] argument is specified, debug is turned on, and SQL statements display on your screen. (They can also be diverted to a file.)
- If the [<Date>] argument is specified, then it is the date the logs are inserted into the database. (Date format must be MM/DD/YYYY.)

Modify Load Request Log Variables

The loadRequestLog.pl Perl script contains the following three variables that can be changed as needed:

- \$configPath = "/full/path/re_report.cfg"
(Full path of configuration file; contain paths to where the input files are)
- \$errorLog = "/full/path/re_report.err.log"
(Full path of error log)
- \$theemails = "elena@morris.com, misselenax@yahoo.com"
(If any error occurs, a message is sent to these email addresses.)

4. Determine Data Load Successful

If the [debug] argument is specified, as data is loaded into the database, process messages appear on your screen. The last message displayed should be "DATA LOAD SUCCESSFUL," indicating the error log has been deleted and the data load is successful.

If the error log is not deleted, an email is sent to the System Administrator with the location of the error log that has been renamed to /full/path/re_report.err.date. He may log in to the server and review this file, fix the errors, and run the shell script again. If there are *new* errors, they are entered in the same file separated by the old errors with five equal signs (=====). Once errors are corrected, or if they are not relevant, the error log may be deleted.

Note: /full/path/ is a placeholder for the actual path and date is a placeholder for the date the job is run.

If the error log contains fatal errors (Connection interrupted; load didn't finish, etc.), run the Recover Perl script (see "Run Recover Perl Script for Fatal Errors" section) to roll back the partial data load.

Examples of Common Errors Running Scripts

While these scripts usually run smoothly, occasionally errors may occur. If you have specified the full path for the `$errorLog` variable, the following common errors may be found in the error log:

- No ORACLE_HOME environment variable set
(Enter the set command to bring up ORACLE_HOME.)
- Could not open config file
(This a permissions error; either modify your permissions or sign on as someone with proper authority.)
- Could not open input file /opt/MDW/bin/re_report/Sep/: No such file or directory at /opt/MDW/bin/re_report/loadRequestLog.pl line 127, line 2
(The file doesn't exist under the specified path; verify path.)
- ERROR[49]: No info found for ad ID 100641698. Skipping record...
(While this error is primarily informational, you may go to your DBA and ask why this ad is not in the database. If it is expired, the templates can be modified not to call it.)

Run Recover Perl Script for Fatal Errors

This Perl script (`/opt/MDW/bin/re_report/recover.pl`) can be used when fatal errors occur and you need to undo or start over. This script has the same dependencies as the `runLoadRequestLog` script.

The command takes **1** [optional] argument - [`<Date>`]:

- If **0**, argument: file(s) becomes the current date logs.
- If **1**, argument: file is the date the logs are inserted into the database. (Date format must be MM/DD/YYYY.)

Warning: This script should be used sparingly (use primarily for fatal errors). Also, avoid using the [`<Date>`] argument to ensure you do not roll back successful data loads.

Support Information

Morris Digital Works (MDW), a division of Morris Communications Co. (MCC), provides tools, technologies, consulting, and Web development services to Morris newspapers and external clients. MDW award-winning technologies include: world-class hosting facilities, robust content management software, high performance application tools, site enhancement tools, and comprehensive classified and display classified technology. Founded in 1995, MDW has over 100 employees with offices in Topeka, KS, Joplin, MO, New York, NY, and is headquartered in Augusta, GA. An additional 250 MCC employees also participate in our Internet business and report directly to newspapers, magazines, book publishing, and other internal organizations.

If you are experiencing problems with any Morris Digital Works product, please contact customer support at (706) 828-2955.

Appendix A. Analog Configuration File

The `stats.conf` file specifies the format the Analog program must use to convert the web access logs to another log file. It contains the following information:

```
## Analog Conf - Reporter Version 1

OUTPUT COMPUTER
COMPSEP |

LOGFILE -

## What to include in the request report.
REQINCLUDE PAGES

## Which files are included and excluded from the report. (FILEEXCLUDE
Overrides FILEINCLUDE).
FILEINCLUDE /*
FILEEXCLUDE /RealMedia/*

GENERAL OFF      #General Summary
MONTHLY OFF      #Monthly Report
WEEKLY OFF       #Weekly Report
FULLDAILY OFF   #Daily Report
DAILY OFF        #Daily Summary
HOURLY OFF       #Hourly Summary
HOST OFF         #Host Report
REQUEST ON       #Request Report
DIRECTORY OFF    #Directory Report

## What type of file to count as a "page".
PAGEINCLUDE *.gif
FILEINCLUDE *.gif

REQCOLS R
DIRCOLS R

DIRFLOOR 0r
REFFLOOR 1r
REQFLOOR 1r

WARNINGS ON
DEBUG OFF
```