

## The Automated Process

This note describes the automated process of updating daily the Deep Creek Answers website.

Updating the website involves a number of steps, quite a few actually.

There are basically three activities that must be performed to update the web pages daily:

1. **Scraping** - Extract information from website(s) on a regular basis during the day
  1. Collect lake level and generator status data from the [deepcreekhydro](#) website on a regular basis. We started (and are continuing) with collecting (scraping) that data every 10 minutes. It turns out that these items are updated by Brookfield every 5 minutes.
  2. Extract the desired quantities and store the results in a file
2. **Preprocessing** - Process all of the data into meaningful web pages and data files
  1. Correct the extracted data for “glitches”
  2. Collect data from other sites that are already available in daily series
  3. Process ALL data and make the appropriate web pages and data sets
3. **Uploading** - Upload all of the information onto the web server
  1. The constructed web pages that contain the “Highcharts JS” scripts and data
  2. The constructed web pages and images that are created with R
  3. The php file that updates the randomly selected image and link that’s on the home page

All of these processes currently run on a 2 x 3 GHz Quad-Core Intel Xeon Mac with 16GB of memory. The process described under 1 and 2 about take about 1 minute; item 1 is run every 10 minutes, and takes just a few seconds.

There are a variety of ways that the above can be accomplished. The tools that have been used are those that the author has learned how to use and is comfortable with.

The above mentioned three activities make use of the following technologies:

**iCal** - a free Apple application that schedules events. It is used to trigger the application that controls the performance of various tasks, here Automator.

**Automator** - a native Apple application that automates all kinds of tasks (more on what tasks later). It is basically used to execute shell scripts and to move files.

**Shell Scripts** - Shell scripting is wildly useful and a powerful way to manipulate a lot of files and to automate behind-the-scenes tasks in Mac OS X. Shell scripts contain command-line entries to execute Ruby and R scripts and to move files on the local machine.

**Ruby** - A very popular scripting language that is used to manipulate data records, write web pages and perform mathematical calculations. Execution of Ruby scripts is done via shell script(s).

**R** - Another fairly popular functional programming language primarily used for statistical analyses and spatial data (involving latitudes and longitudes) manipulations. It has a very strong graphical component that makes the production of beautiful charts very easy. It does have a very steep learning curve. Execution of R scripts is done via shell script(s).

**Highcharts JS** - is a library of functions written in JavaScript (JS) to make it intuitive to create creative and beautiful charts. It is used to display the data that is regularly collected from various websites. Ruby scripts use the syntax of Highcharts JS to create the web pages that contain Highcharts JS calls.

First there is a Ruby script, "**lake\_level.rb**," that is put in motion every 10 minutes to retrieve the data for water level and generator status from the deepcreekhydro.com website, from the box with information in the upper right hand corner. The retrieved data is stored in a file called "hydroreadingsRAW.txt" The script that does this is called "lake\_level.rb". The script then extracts the relevant information and appends it to the "hydroreadings.txt". A snippet of the file is shown in the figure below:

Tue Jan 01 17:19:45 -0500 2013	ON	2458.19	1/1/2013 14:00	ON	1/1/2013 17:18
Tue Jan 01 17:29:45 -0500 2013	ON	2458.18	1/1/2013 14:00	ON	1/1/2013 17:28
Tue Jan 01 17:39:45 -0500 2013	ON	2458.18	1/1/2013 14:00	ON	1/1/2013 17:39
Tue Jan 01 17:49:45 -0500 2013	ON	2458.17	1/1/2013 14:00	ON	1/1/2013 17:49
Tue Jan 01 17:59:45 -0500 2013	ON	2458.17	1/1/2013 14:00	ON	1/1/2013 17:58
Tue Jan 01 18:09:45 -0500 2013	ON	2458.16	1/1/2013 14:00	ON	1/1/2013 18:08
Tue Jan 01 18:19:45 -0500 2013	ON	2458.15	1/1/2013 14:00	ON	1/1/2013 18:18

Via "/library/LaunchDeamons/com.pitiur.convert5am.plist" the file "**convert5am.sh**" is executed every morning at 5am. It is an executable shell script that pulls together all scripts and apps that have to be executed on the local computer in order to generate the web pages with the appropriate data. Its content is shown below:

```
1. #!/bin/bash
2. cd /Users/frostypiet/ruby/
3. ruby weather_gca.rb
4. ruby WeatherAnalysis.rb
5. ruby Convert5am.rb
6. ruby usgs_data.rb
7. ruby usgs_data_analysis.rb
8. R < HomePageImages.R --no-save
```

"**weather\_gca.rb**" retrieves the weather parameters at Garrett County Airport of the last 24 hrs.

"**WeatherAnalysis.rb**" processes the weather data and creates the webpages containing the Highchart JS instructions for displaying the data.

"**Convert5am.rb**" is a Ruby script that processes the retrieved data to fix any glitches that may be present, and they do creep in quite regularly, with cause(s) unknown.

"**usgs\_data.rb**" retrieves the data from USGS stream gages.

"**usgs\_data\_analysis.rb**" processes the stream gage data and creates the webpages containing the Highchart JS instructions for displaying the data.

**“HomePageImages.R”** is an R script that randomly rotates the image on the home page linking to a current item of interest.

The processes described above produce a set of web pages, containing mostly data, and updates the website with them.

The whole scheme scheme became into existence in April of 2012. The early parts were done in December 2011.

PLV: 1/19/2013