Super Webtrax

Tom Van Vleck

15. Hits by access time

<table>
<thead>
<tr>
<th>Time</th>
<th>HTML</th>
<th>Graphic</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00</td>
<td>737</td>
<td>510</td>
<td>406</td>
</tr>
<tr>
<td>01:00</td>
<td>725</td>
<td>522</td>
<td>418</td>
</tr>
<tr>
<td>02:00</td>
<td>694</td>
<td>439</td>
<td>650</td>
</tr>
<tr>
<td>03:00</td>
<td>893</td>
<td>953</td>
<td>589</td>
</tr>
<tr>
<td>04:00</td>
<td>880</td>
<td>900</td>
<td>500</td>
</tr>
<tr>
<td>05:00</td>
<td>1035</td>
<td>908</td>
<td></td>
</tr>
<tr>
<td>06:00</td>
<td>1108</td>
<td>817</td>
<td>943</td>
</tr>
<tr>
<td>07:00</td>
<td>862</td>
<td>817</td>
<td></td>
</tr>
<tr>
<td>08:00</td>
<td>773</td>
<td>862</td>
<td></td>
</tr>
<tr>
<td>09:00</td>
<td>654</td>
<td>817</td>
<td></td>
</tr>
<tr>
<td>10:00</td>
<td>530</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>11:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>12:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>13:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>14:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>15:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>16:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>17:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>18:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>19:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>20:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>21:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>22:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
<tr>
<td>23:00</td>
<td>493</td>
<td>817</td>
<td>469</td>
</tr>
</tbody>
</table>

07/22/15
What Super Webtrax Does

• Reads a daily log from a web server.
• Produces a web site report in HTML.
  – Multiple report sections (45)
  – Many options
How I Use It

• I look at the report every day for
  - Signs of problems with the site or ISP
  - Signs of attacks or misuse
  - Level of traffic and resources
  - Popularity of pages

• Some sections provide more information if I see something interesting.
What It Doesn't Do

• **Real time analysis**
  – Even if I had realtime log access, I don't have time to pore over them.

• **Summaries by week or month**
  – I'm not interested in this.
  – Would not be difficult to add.

• **Ability to drill down on reports, e.g. show all sessions from a particular referrer**
  – Would require interactive queries to the database.
  – Substantial rework of interface.
  – I run queries by hand for rare cases.
  – Queries against more than one day would need a huge database.

• **Handle very busy sites**
  – Would need dedicated resources for > 500K hits/day.
  – Reverse DNS queries are slow, would have to reprogram.
How it Works

• Web servers write a log entry every time they send a page to a user.
• Once a day, Super Webtrax loads web server logs into MySQL.
• SWT expands templates that query MySQL to produce HTML reports with graphs and tables.
• Configuration comes from SQL tables.
Super Webtrax
What You Need

- Analyzes NCSA Combined Format logs
- Uses MySQL 4.1 or later
- Uses Perl and shell scripting
- Runs on Unix or Mac OS X
- Can use free geolocation data from MaxMind
History

• 1995: Webtrax by John Callender
  – Perl
  – e-mail report

• 1996-2005: Webtrax by THVV
  – Perl
  – HTML report, graphical, multiple sections
  – Java pie charts

• 2006: Super Webtrax by THVV
  – Perl, MySQL
  – Many more report sections and charts
  – JavaScript pie charts, multiple chart views
Processing

- Run daily (cron)
- Need not run on web hosting server
- Output can be put on any web location
- One MySQL database for each log stream
Input Logs

- **NCSA Combined format**
  - Containing referrer and user agent

- **Program `combinelogs` can**
  - Merge multiple logs, add file prefix

- **Program `logextractor` can**
  - Extract one day’s usage from a running log
  - Look up domain names from IP
  - Look up geographical location from IP
Output Report Structure

• Navigation Bar at top and bottom
• User supplied preamble and postamble
  – HTML, can be output of local program
• Sections toggle between
  – Short report
  – Long report

When you click the [+]

Auxiliary Reports

- Last 7 days of important visits
- Input for dashboard report
- Input for GraphViz
- Others as defined by user
Report Sections (1)

- **Bar chart: Month Summary**
  - *Highest numbers red, lowest blue*

1. Month Summary: 2124869 hits, lowest red blue

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Visits</th>
<th>Mb</th>
<th>Hits</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-09-18</td>
<td>2006-09-19</td>
<td>3355</td>
<td>152</td>
<td>10528</td>
<td>6034</td>
</tr>
<tr>
<td>2006-09-19</td>
<td>2006-09-20</td>
<td>2671</td>
<td>118</td>
<td>12281</td>
<td>3948</td>
</tr>
<tr>
<td>2006-09-20</td>
<td>2006-09-21</td>
<td>2423</td>
<td>161</td>
<td>11366</td>
<td>3671</td>
</tr>
<tr>
<td>2006-09-21</td>
<td>2006-09-22</td>
<td>2980</td>
<td>138</td>
<td>12948</td>
<td>3524</td>
</tr>
<tr>
<td>2006-09-22</td>
<td>2006-09-23</td>
<td>3168</td>
<td>214</td>
<td>12221</td>
<td>4627</td>
</tr>
<tr>
<td>2006-09-23</td>
<td>2006-09-24</td>
<td>3033</td>
<td>176</td>
<td>15405</td>
<td>3993</td>
</tr>
<tr>
<td>2006-09-24</td>
<td>2006-09-25</td>
<td>2969</td>
<td>172</td>
<td>15278</td>
<td>2789</td>
</tr>
<tr>
<td>2006-09-25</td>
<td>2006-09-26</td>
<td>2822</td>
<td>205</td>
<td>15436</td>
<td>4305</td>
</tr>
<tr>
<td>2006-09-26</td>
<td>2006-09-27</td>
<td>3497</td>
<td>112</td>
<td>13123</td>
<td>3485</td>
</tr>
<tr>
<td>2006-09-27</td>
<td>2006-09-28</td>
<td>2176</td>
<td>165</td>
<td>12071</td>
<td>5012</td>
</tr>
<tr>
<td>2006-09-28</td>
<td>2006-09-30</td>
<td>2469</td>
<td>299</td>
<td>14197</td>
<td>4451</td>
</tr>
<tr>
<td>2006-09-30</td>
<td>2006-09-01</td>
<td>2803</td>
<td>189</td>
<td>15003</td>
<td>4414</td>
</tr>
<tr>
<td>2006-09-01</td>
<td>2006-09-02</td>
<td>2982</td>
<td>196</td>
<td>14223</td>
<td>4273</td>
</tr>
<tr>
<td>2006-09-02</td>
<td>2006-09-03</td>
<td>2533</td>
<td>164</td>
<td>14272</td>
<td>5513</td>
</tr>
<tr>
<td>2006-09-03</td>
<td>2006-09-04</td>
<td>2683</td>
<td>199</td>
<td>14143</td>
<td>4477</td>
</tr>
<tr>
<td>2006-09-04</td>
<td>2006-09-05</td>
<td>2246</td>
<td>129</td>
<td>10653</td>
<td>9023</td>
</tr>
<tr>
<td>2006-09-05</td>
<td>2006-09-06</td>
<td>2194</td>
<td>372</td>
<td>11452</td>
<td>5402</td>
</tr>
<tr>
<td>2006-09-06</td>
<td>2006-09-07</td>
<td>2654</td>
<td>258</td>
<td>14555</td>
<td>4344</td>
</tr>
<tr>
<td>2006-09-07</td>
<td>2006-09-08</td>
<td>2081</td>
<td>528</td>
<td>10966</td>
<td>5642</td>
</tr>
<tr>
<td>2006-09-08</td>
<td>2006-09-09</td>
<td>2939</td>
<td>171</td>
<td>15011</td>
<td>3904</td>
</tr>
<tr>
<td>2006-09-09</td>
<td>2006-09-10</td>
<td>3124</td>
<td>211</td>
<td>17912</td>
<td>5345</td>
</tr>
<tr>
<td>2006-09-10</td>
<td>2006-09-11</td>
<td>2958</td>
<td>181</td>
<td>10309</td>
<td>4429</td>
</tr>
<tr>
<td>2006-09-11</td>
<td>2006-09-12</td>
<td>2316</td>
<td>157</td>
<td>12012</td>
<td>4205</td>
</tr>
<tr>
<td>2006-09-12</td>
<td>2006-09-13</td>
<td>2256</td>
<td>143</td>
<td>10308</td>
<td>3723</td>
</tr>
<tr>
<td>2006-09-13</td>
<td>2006-09-14</td>
<td>2249</td>
<td>193</td>
<td>14696</td>
<td>4543</td>
</tr>
<tr>
<td>2006-09-14</td>
<td>2006-09-15</td>
<td>2877</td>
<td>238</td>
<td>18843</td>
<td>9640</td>
</tr>
<tr>
<td>2006-09-15</td>
<td>2006-09-16</td>
<td>2839</td>
<td>190</td>
<td>15334</td>
<td>4625</td>
</tr>
<tr>
<td>2006-09-16</td>
<td>2006-09-17</td>
<td>2720</td>
<td>216</td>
<td>14766</td>
<td>4907</td>
</tr>
<tr>
<td>2006-09-17</td>
<td>2006-09-18</td>
<td>2445</td>
<td>198</td>
<td>15096</td>
<td>5025</td>
</tr>
<tr>
<td>2006-09-18</td>
<td>2006-09-19</td>
<td>2126</td>
<td>240</td>
<td>11364</td>
<td>4991</td>
</tr>
<tr>
<td>2006-09-19</td>
<td>2006-09-20</td>
<td>2114</td>
<td>198</td>
<td>11293</td>
<td>4609</td>
</tr>
</tbody>
</table>

31 days Avg 2700 199 14203 4401
Report Sections (2)

- **Pie Charts**
  - Hits by File Type
  - MB by File Type
  - Hits by TLD
  - Visits by TLD
  - MB by TLD
  - Visits by Hit Source
  - Visits by Class
  - Visits by Browser excluding indexers
  - Visits by Platform excluding indexers
  - Visits by Continent excluding indexers

<table>
<thead>
<tr>
<th>3385 Visits by Class</th>
<th>2493 NI Visits by Browser</th>
<th>2493 NI Visits by Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31% brw (1054)</td>
<td>51% MIE (1273)</td>
<td>90% WinNT (2014)</td>
</tr>
<tr>
<td>26% Indexer (832)</td>
<td>37% Mozilla (947)</td>
<td></td>
</tr>
<tr>
<td>19% mult (670)</td>
<td>1% Opera (27)</td>
<td></td>
</tr>
<tr>
<td>1% Ecel (1)</td>
<td>1% Windows xp (27)</td>
<td></td>
</tr>
<tr>
<td>1% swe (127)</td>
<td>2% Linx (111)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1% other (135)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1% Mac (116)</td>
</tr>
</tbody>
</table>
Report Sections (3)

- Table: Analysis
- Bar chart: HTML pages
- Bar chart: Graphic files
- Bar chart: CSS files
- Bar chart: Flash files
- Bar chart: Files Downloaded
- Bar chart: Sound files
- Bar chart: XML files
- Bar chart: Java Class files
- Bar chart: Source files
- Bar chart: Other files
- List: Files not found

You can turn reports off if you don’t need them.
Report Sections (4)

- Bar chart: Forbidden transactions
- Bar chart: Illegal referrers
- Vertical Bar chart: Hits by access time
- Bar chart: Visits by duration
- Bar chart: Visits by number of hits
- Bar chart: Visits by number of page views
- Bar chart: Visits by Top-level Domain
- Bar chart: Visits by Domain
- Bar chart: Visits by Second level Domain
- Bar chart: Visits by Third level Domain
- Bar chart: Visits by Authenticated User
Report Sections (5)

- Bar chart: Visits by Class
- Bar chart: Visits by Browser
- Bar chart: Hits by Query
- Bar chart: Visits by Search Engine
- Bar chart: Files crawled by Google
- Bar chart: Hits by Referrer
- Bar chart: Number of Hits by file size
- Bar chart: Hits by Local Query
- Bar chart: Repeated hits by Domain
- Bar chart: Attacks on the site (CGI Attacks, Hack Probes, Excess use)
- Bar chart: Transactions by server return code
- Bar chart: Transactions by protocol verb
Report Sections (6)

• Visit Detail Report
  – Time
  – Visitor’s domain (new domains in blue)
  – HTML pages (colored depending on user defined filename)
  – Query used to find page (green)
  – Time between pages
  – Total hits and KB, Browser
  – Visit class (user defined)
  – Authentication ID; authenticated sessions highlighted in yellow
• User defines which visits are “interesting.” Can view only interesting visits.

11:52 201-016-239-042.static.ctbctelecom.com.br -- index.html 0:21, services.html 0:17, index.html [13, 57 KB, MSIE 6.0; Windows NT 5.1] {techtalk}
11:53 adsl-dyn84.91-127-243.t-com.sk -- thvv/threeq.html (images.google.sk: comix) [4, 43 KB, Firefox; Windows NT 5.1] {thvv}
12:04 209.212.4.130[us].-- thvv/private/computer-advice.html (webmailbb.juno: folder=Inbox&amp;msgNum=00000A00:0017cVMs000027bN&amp;bl) [5, 53 KB, MSIE 7.0; Windows NT 5.1] {thvv:hths}
Report Sections (7)

- Bar chart: Year Non-search Hits by Referrer
- Bar chart: Year Hits by Query
- Bar chart: Year hits by domain
- Bar chart: Domains by days since last visit
- Bar chart: Year hits on HTML Pages
- Vertical Bar chart: Hits by month
- Bar chart: Hits by year
Installation

- Hard for beginner, assumes Unix skills
- Install MySQL
- Install Perl and extensions
- Create database
- Install Super Webtrax
- Run "configure"
  - Answer questions
  - Can re-run
- Run "install"
Extensibility

• Create a new report
  – Define SQL queries
  – Create new template file
  – Define parameters

• Add configuration values to `swt_user.sql`

• Add to `swt`

• The template
  – Fetches the queries
  – Sets up headings
  – Executes the queries generating HTML lines
  – Output is often formatted as an HTML table
  – Generates short and long report panes

• Example: Funnel Report
  – for an electronic commerce client
  – summarized when visitors exited shopping sessions
Data flow

Apache log

swt.sql
templates

swt-user.sql

database

HTML report

SWT
Perl programs

• logvisits.pl
  – Reads Apache log
  – Writes log2db.sql which creates **hits** table

• visitdata.pl
  – reads hits table
  – writes visits.sql which creates **visits** table

• wordlist.pl

• expandfile
  – expands templates, reads database, writes reports

• printvisitdetail.pl
  – reads **hits x visits**, generates report section