WSG Log Analysis Review

I've reviewed ten different log analysis packages based on the following criteria:

**Analysis priorities** (must include ...)
- Traffic patterns (daily, hourly, etc)
- Most visited urls
- Referrer information
- Search phrases
- Visits/Unique Visitor counts
- Visit length
- Most downloaded files
- Breakdown by average visitor (paths)
- Easy to drill down for specific url reports
- Attractive interface for sharing with clients

**Technical priorities**
- Stores aggregate summary data about logs so we don't need to keep them around after processing
- Flexible access control, easy to share reports
- Easy to install and maintain
- Reasonable support available

Each package can perform most of the analysis tasks listed. They vary in how well the information is presented and to what degree you can refine results. Not all of the packages are able to present information about common site navigation paths and I'll note that in their description.

The packages function in one of two ways. Either they are hosted locally and use server log data directly or they are hosted remotely and rely on a javascript embedded in pages to be monitored:

**Log data**

**Pros**
- We can control what log data gets fed into the system (slice and dice with regular expressions, reprocess for specialized reports).

**Cons**
- We must control what log data gets fed into the system (push the logs to where the tool can find them, synchronize log analysis with log rotation).
- Not all of the log-based analysis tools aggregate data from previously analyzed log files, so you must be careful that they don't see bits of the same data twice.
- We've got to install and maintain the system.
Javascript Monitoring
Pros
• All we need to do to track access to a page is add a snippet of javascript.
• Once the pages to be monitored are set up system administration should be minimal.

Cons
• We've got to modify every page we want monitored. We could modify Reason to automatically add the javascript (which would only cover Reason pages), or write an apache module to do the same for everything on the server (potentially dangerous and slow).
• Bandwidth charges – the javascript hits the service provider’s server and they charge based on the number of hits.

Summary
I've listed a capsule review of each package below and you can access demo instances of several at http://ventnor.its.carleton.edu/stats.html. Jaye or I can provide access to WebTrends if you'd like to have a look. After reviewing everything, I think NetTracker is probably our best option. If offers a reporting flexibility and attractive interface similar to WebTrends and could be applied to more sites at considerably less cost. We would need to maintain it, and it’s worth considering whether it may be overkill or possibly missing something I’ve overlooked. Please have a look and hopefully we can talk about them as a group to decide which way to go. If you have any questions or want to see something I haven’t put online please let me know.

The Contenders:

WebTrends OnDemand
• http://www.webtrends.com
• Commercial
• $35/month base cost.
• We pay for bandwidth @ $0.65/10k hits (most of the cost).
• $10 for every additional profile.
• We've paid ~$4200 in the last year.
• This uses the javascript/remotely hosted model. We've got three profiles configured, but I haven't surveyed which pages are included in each.
• We're likely paying double bandwidth charges for the internal/external and external only profiles.
• WebTrends has all the analysis tools you could want in an attractive user interface and costs accordingly.
• There is a locally hosted version available ($895 + $161/year for support, saved bandwidth costs), but it's only available for Windows.
• The query interface is a bit clunky, but could probably be used in place of the profiles when we consolidate www and apps.
• This is an excellent product, but adding profiles means adding bandwidth/cost.
NetTracker Professional

- http://www.sane.com
- Commercial
- Purchase $1395 for 5 profiles or $1995 for 10
- Support (telephone M-F 9-6, all available updates) $350/year for 5 profiles or $495/year for 10. Drops to $270 and $390 per year respectively after the first year.
- This is hosted locally and works directly on log files.
- NetTracker is on par with WebTrends for analysis features and an attractive interface.
- It's easy to filter results, limiting them to specific hosts, directories, etc.
- NetTracker keeps track of which records it has already seen so we shouldn't have to worry about reprocessing entries.
- It's possible to create archive reports to save log data storage space, but it's no longer possible to apply filters without restoring the original logs.

NetTracker wants a dedicated server, but I think we could comfortably host it on Chicago or Ventnor. Log processing would occur in the early morning when load is low and the report review web interface isn't particularly taxing.

At about half the cost of WebTrends in the first year, and considerably less each year after I think this would make a good alternative to our current logging software.

Webalizer

- http://www.mrunix.net/webalizer/
- Free
- We're already using it, the infrastructure is in place.
- It's packaged by RedHat, so updates are more or less automatic.
- It summarizes the logs files in it's own database, so we can do away with them, but it doesn't keep the same level of detail as WebTrends/NetTracker.
- The reporting options are not on par with WebTrends and NetTracker. It doesn't provide path tracking.
- You can't really dig down into the results beyond what it's already displaying without pre-processing the logs.
- The existing infrastructure could use some refinement. It's not filtering for search bots, the regex's used to filter the log files may be causing miss-counts.
- It needs to be extended to other hosts (www, acad, people, orgs, go, blogs).
- I think this would probably be fine if it were just us looking at the results. We can manually slice the log files to generate specific reports if need be.
- We would need to keep a years worth of original log files to do specialized reporting.
Summary Plus
- [http://summary.net/](http://summary.net/)
- $249
- Lots of available reports, possible to drill down via search
- This uses local log files, and has a fairly small install footprint.
- It maintains its own database of compressed log info, so we can do away with the originals.
- It's a budget NetTracker/WebTrends, with many of the same features, but lacking the refinement and slick UI (graphs without labels, bad navigation).
- If we wanted to eliminate 'big-ticket' log analysis tools from our budget, this would probably be a reasonable option.

Out of the running:

Visitors
- [http://www.hping.org/visitors/](http://www.hping.org/visitors/)
- Free
- This uses local log files and we would need to eliminate duplicate records from being processed. It does show path information and generates a nifty graphical visitor flowchart.
- It's just a static report and doesn't permit you to dig down further but it's very fast and covers all the basics.
- It's neat to see, but aside from possibly using it to generate the path flowcharts I don't think we'll have a use for it.

Analog
- [http://www.analog.cx](http://www.analog.cx)
- Free
- This uses local log files and we would need to eliminate duplicate records from being processed.
- One of the original log analysis tools, included for completeness.
- It has the same faults as Visitors, and is not packaged by RedHat.

AWStats
- Free
- This uses local log files and we would need to eliminate duplicate records from being processed.
- More detail than Webalizer, not packaged by RedHat.
- Recent security flaws.
ClickTracks
- http://www.clicktracks.com/
- $495
- Provides a unique view: statistics overlaying the original page interface. You see activity as you navigate the site itself.
- It's something of an executive toy, would be hard to automate and share results.

Urchin
- http://www.urchin.com/
- $199/mo for 100,000 hits, $99 per additional 1m hits (similar to WebTrends)
- Embedded javascript required.
- Not possible to demo without a giving them a credit card and the service is in flux since being purchased by Google.

Mint
- http://www.haveamint.com/
- $30 per site
- Slick web based ui
- No real demo to speak of (they provide a video), can't evaluate further.