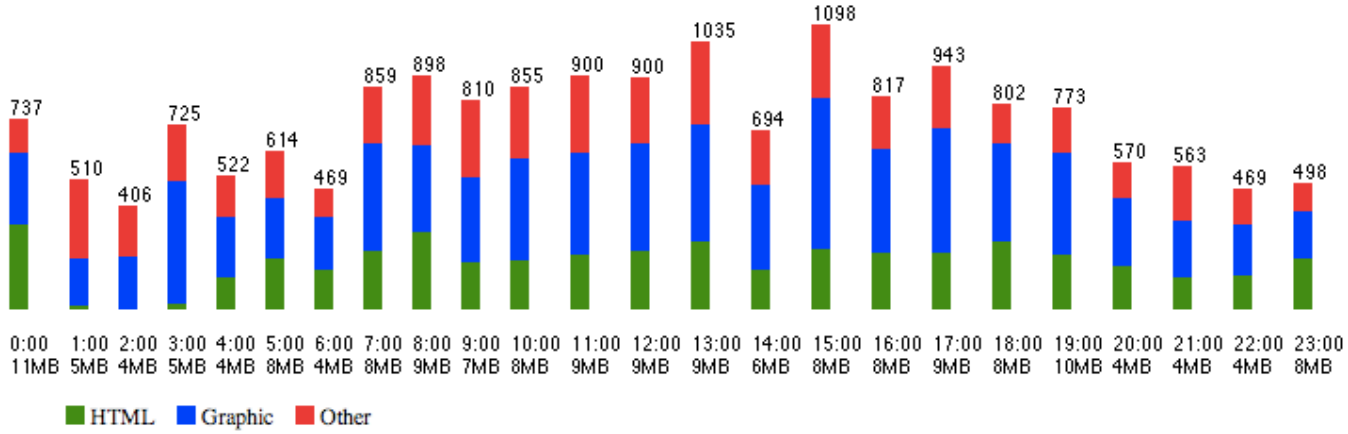
Super Webtrax

Tom Van Vleck

15. Hits by access time



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.*

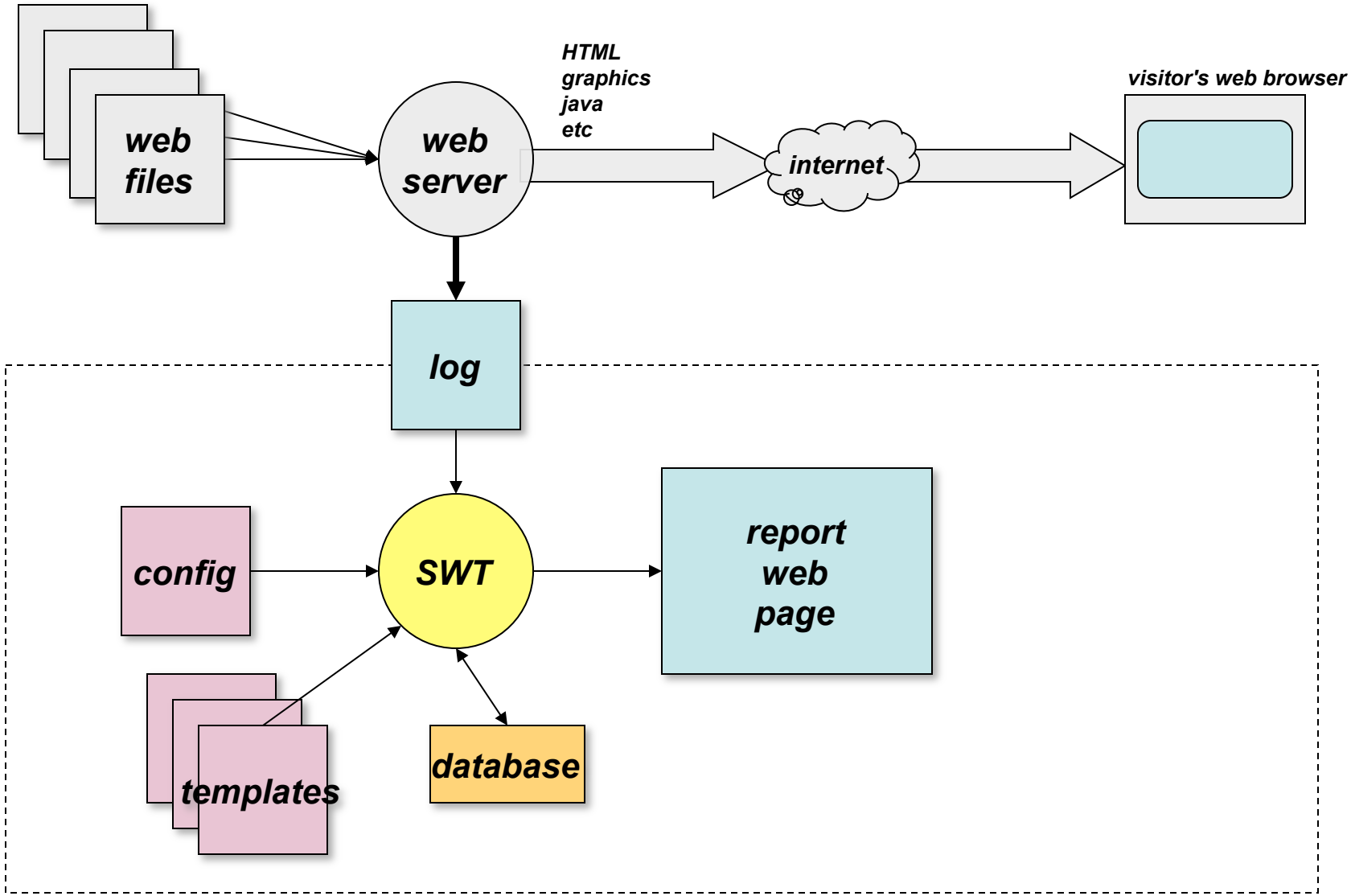
What It Doesn't Do

- ***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***

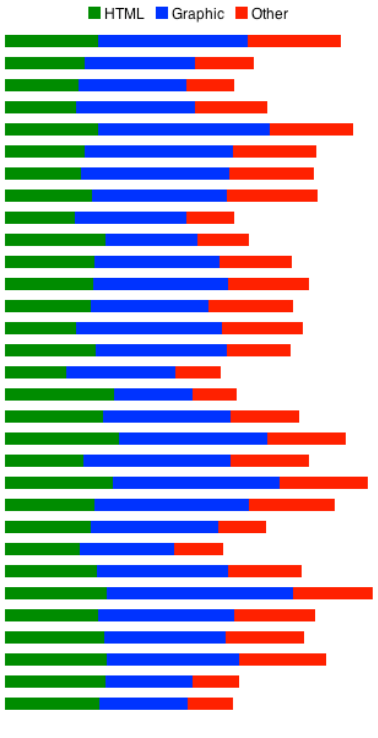
Report Sections (1)

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

1. Month Summary: 2124869 hits, 28715 MB since 2006-05-03 00:02:08

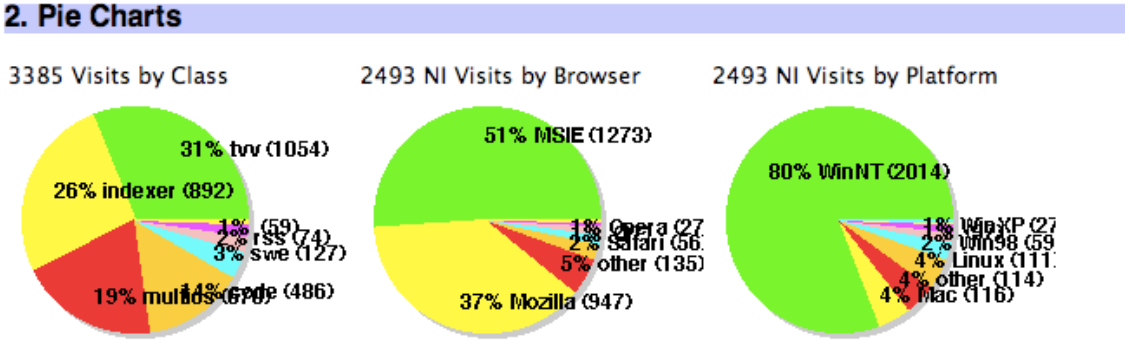
Super Webtrax version 11 2006-08-24 10:04

From	To	Visits	Mb	Hits	Pages
2006-09-18 00:01 Mon	2006-09-19 00:00	3385	182	16628	4654
2006-09-17 00:01 Sun	2006-09-18 00:00	2671	198	12281	3945
2006-09-16 00:01 Sat	2006-09-17 00:00	2423	161	11366	3671
2006-09-15 00:01 Fri	2006-09-16 00:01	2980	158	12948	3524
2006-09-14 00:01 Thu	2006-09-15 00:00	3168	214	17221	4629
2006-09-13 00:01 Wed	2006-09-14 00:01	3033	176	15405	3939
2006-09-12 00:01 Tue	2006-09-13 00:00	2969	173	15278	3789
2006-09-11 00:01 Mon	2006-09-12 00:00	2888	205	15436	4305
2006-09-10 00:01 Sun	2006-09-11 00:00	2347	132	11315	3458
2006-09-09 00:01 Sat	2006-09-10 00:00	2176	165	12071	5017
2006-09-08 00:01 Fri	2006-09-09 00:00	2469	249	14197	4451
2006-09-07 00:01 Thu	2006-09-08 00:00	2803	189	15003	4414
2006-09-06 00:01 Wed	2006-09-06 23:59	2802	196	14223	4273
2006-09-05 00:02 Tue	2006-09-06 00:00	2933	164	14723	3513
2006-09-04 00:01 Mon	2006-09-05 00:00	2683	199	14143	4477
2006-09-03 00:01 Sun	2006-09-04 00:00	2246	129	10653	3024
2006-09-02 00:01 Sat	2006-09-03 00:00	2194	372	11452	5402
2006-09-01 00:01 Fri	2006-09-02 00:00	2654	258	14535	4844
2006-08-31 00:01 Thu	2006-09-01 00:00	2881	328	16866	5645
2006-08-30 00:01 Wed	2006-08-31 00:00	2859	171	15011	3904
2006-08-29 00:01 Tue	2006-08-30 00:01	3124	211	17912	5348
2006-08-28 00:01 Mon	2006-08-29 00:01	2986	181	16308	4429
2006-08-27 00:01 Sun	2006-08-28 00:00	2316	157	12912	4265
2006-08-26 00:01 Sat	2006-08-27 00:00	2256	143	10768	3723
2006-08-25 00:02 Fri	2006-08-26 00:01	2749	193	14690	4543
2006-08-24 00:01 Thu	2006-08-25 00:00	2877	238	18143	5046
2006-08-23 00:01 Wed	2006-08-24 00:01	2839	190	15334	4625
2006-08-22 00:01 Tue	2006-08-23 00:00	2720	216	14766	4907
2006-08-21 00:01 Mon	2006-08-22 00:00	3043	198	15860	5025
2006-08-20 00:05 Sun	2006-08-21 00:00	2126	240	11564	4991
2006-08-19 00:01 Sat	2006-08-20 00:00	2114	198	11293	4669
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*



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*

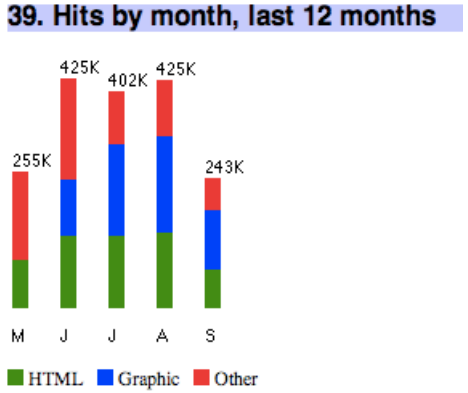
Report Sections (6)

- ***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&msgNum=00000A00:0017cVMs000027bN&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***



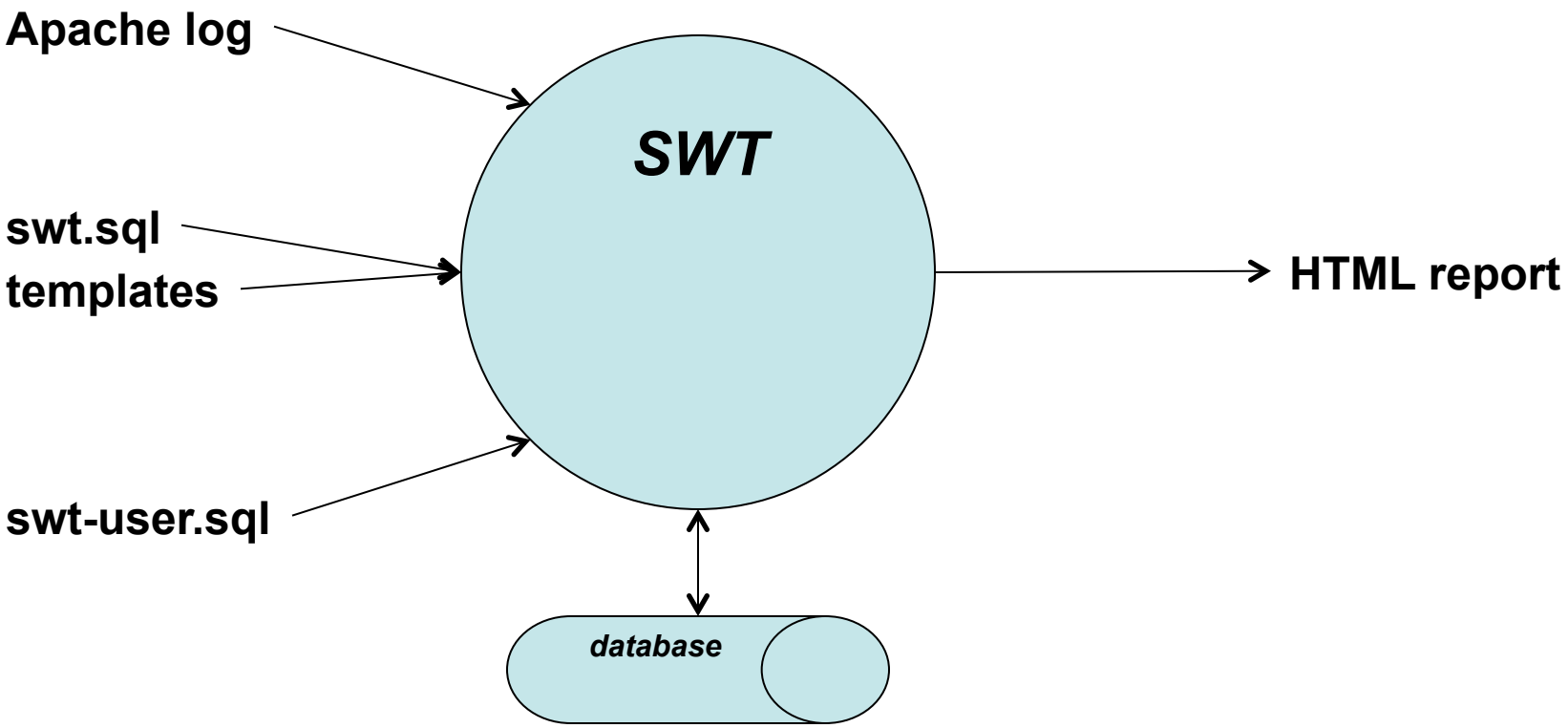
Installation

- *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
SWT
HTML report
database

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