web-generator-toolkit
Documentation
jfbrady@admin.nv.gov

Overview
The web-generator-toolkit is a set of Perl scripts which process Apache JMeter load testing tool output files to produce graphs, histograms, and statistics. This is the set of tools referred to in the paper included in this repository titled "How to Emulate Web Traffic Using Standard Load Testing Tools" that is used to analyze the results of the case study. This tool set produces the statistics necessary to confirm test data adherence to that paper's Principle B.

> **Principle B**
>
> Verify that the requests generated by applying Principle A (i.e. large N with low rate 1/Z) closely approximate a Poisson process by measuring the coefficient of variation of the inter-arrival periods and demonstrating that CoV ~ 1 within acceptable measurement error.

A CoV computation example is contained in the gray block of the **Inter-arrival Times** statistics report on page 4 under the "cv" heading. CoV is computed by web page and total in this report with the values in the "Total" row matching corresponding entries in the paper's Table 3.

Package Deployment
The web-generator-toolkit runs in either Windows (zip) or Unix/Linux (tar.gz).

Package Directory Structure
The home directory is named web-generator-toolkit and contains sub-directories and files as follows:

1. bin directory
    a. perf_jmeter_records_select.pl – removes beginning and ending data records from test file
    b. perf_arr_rt_jmeter_stats.pl – main data analysis script
    c. Graph_jfb.pm – Graph.pm by Matt Kruse modified to support png images (needs GD.pm)
2. demo directory
    a. Seven 25-minute JMeter output files of increasing load, e.g., 1830_AggRpt.csv, where the four digit number indicates the test start time in a 24-hour clock.
    b. run_perf_jmeter_records_select.pl
    c. select directory
        i. run_perf_arr_rt_jmeter_stats.pl
3. jmeter directory
    a. GOV_Test.jmx – dummy jmeter script which matches demo layout.
    b. DataSet directory
        i. CountyID.csv
    c. RunTable directory
        i. 1830_AggRpt.csv

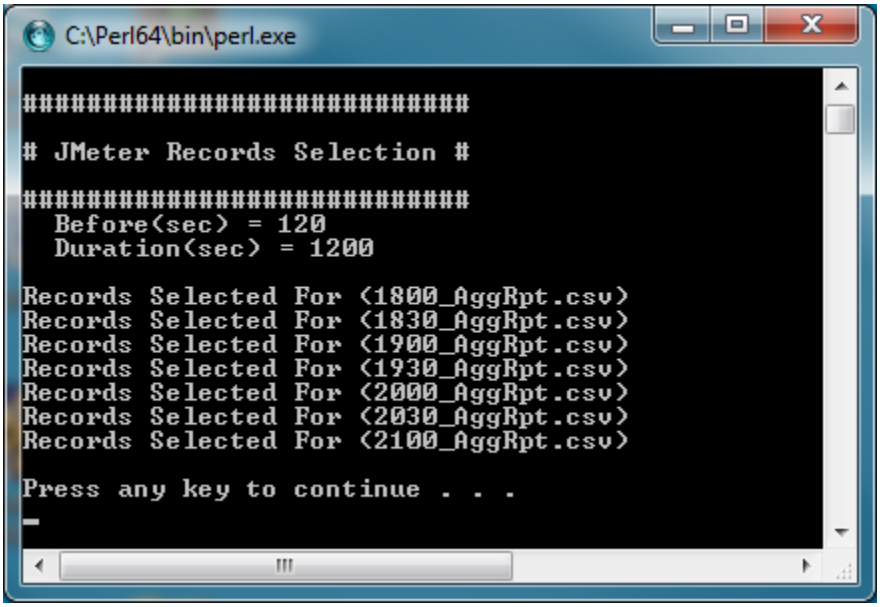
Package Setup
The environmental variable **WEB_GEN_TOOLKIT** must to be created pointing to the **bin** directory. Use "set" for Windows and "export" for Unix / Linux.

Script Execution
After package deployment and setup go to the demo directory and execute the run_perf_jmeter_records_select.pl Perl script. This script runs the perf_jmeter_records_select.pl script and deposits "trimmed" data files in the select directory. Once the "trimmed" data files are created go to the select directory and execute the run_perf_arr_rt_jmeter_stats.pl script.

1. cd demo -> <execute> run_perf_jmeter_records_select.pl
2. cd select -> <execute> run_perf_arr_rt_jmeter_stats.pl

<execute> demo / run_perf_jmeter_records_select.pl

As shown below, seven trimmed files are created in the select directory with a single execution. These files are produced by removing the first 120 seconds and retaining the next 1200 seconds of each file.

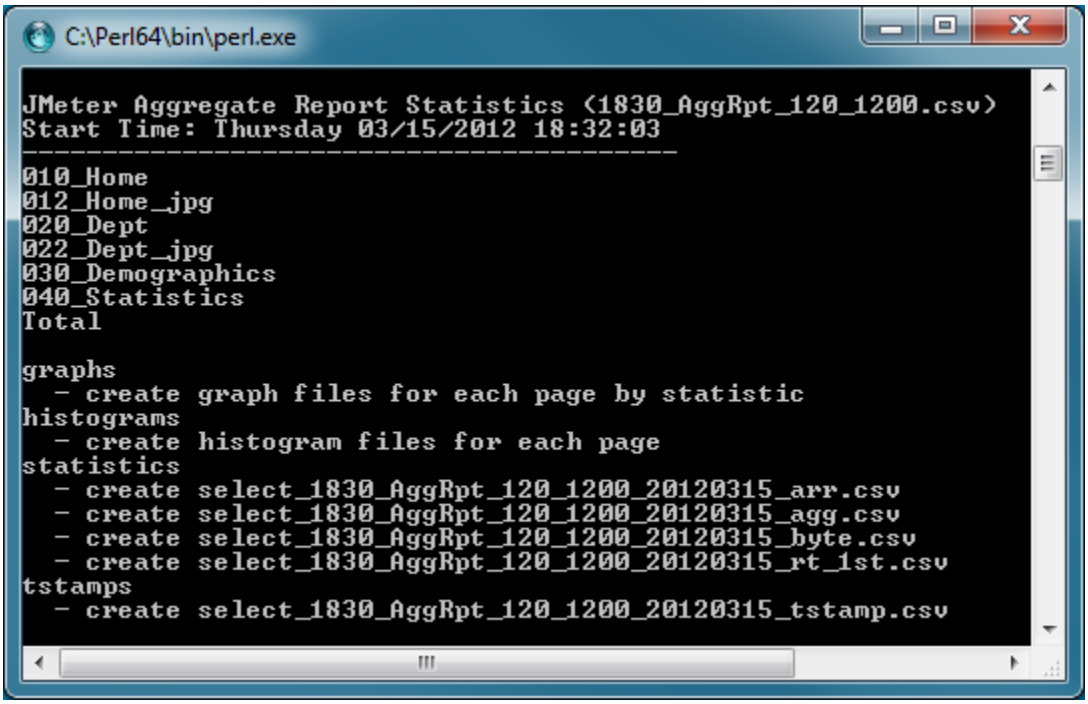


Program execution options:

1. -a add web page response time to timestamp
2. -b subtract web page response time from timestamp
3. -h display help text
4. -l put output file in home directory (default is select)

<execute> select / run_perf_ arr_rt_jmeter_stats.pl

The seven files created in the select directory are processed in a single execution. The screen output for processing trimmed file 1830_AggRpt_120_1200.csv is below.



An output directory is produced for each trimmed file that has the same name less the ".csv" file extension, e.g., 1830_AggRpt_120_1200. There are four sub-directories created under each of these test run directories. Example output for the 1830 test run is as follows.

Output Directory - 1830_AggRpt_120_1200

The sub-directories are:

1. **graphs** – bar charts by web page and total for a myriad of variables (png).
2. **histograms** – frequency counts by time interval for inter-arrival and response times (csv).
3. **statistics** – web page and total for arrivals, response times, and web page sizes (csv).
4. **tstamp** – input file with added lead column containing readable timestamps (csv).

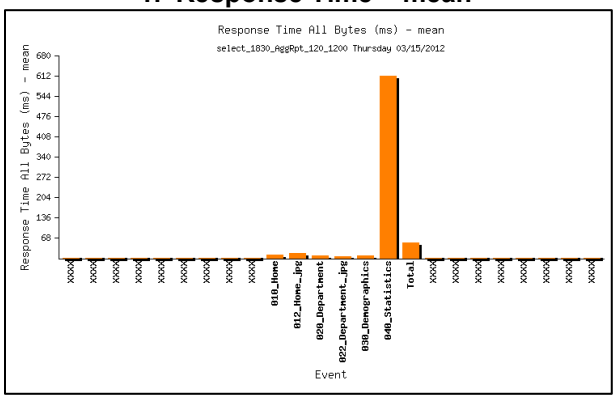
graphs (-g option)

Bar charts in png format across web pages and total.

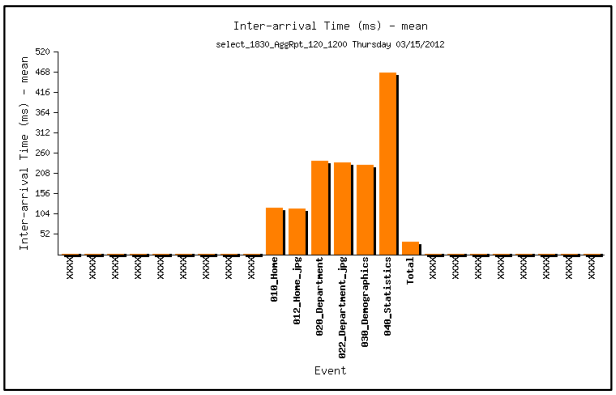
Two of the many png files are:

1. "select_1830_AggRpt_120_1200_20120315_agg_rt_04_mean.png"
2. "select_1830_AggRpt_120_1200_20120315_arr_04_mean.png".

1. Response Time – mean



2. Inter-arrival Time - mean



histograms (-d option)

Inter-arrival and response time histograms by web page and total.

Associated Options:
-a first histogram cell size (default is 10 milliseconds)
-b second histogram cell size (default is 100 milliseconds)
-c maximum number of histogram cells (default is 10000)

The two "Total" histogram csv files with 10 millisecond cell intervals are:

1. "select_1830_AggRpt_120_1200_Total_arr_0010.csv"
2. "select_1830_AggRpt_120_1200_Total_rt_0010.csv"

Below are the first twenty five rows of each of these files spreadsheet formatted for readability.

1. Inter-arrival Time - Total

Inter-arrival Time Histogran For {Total} - select_1830_AggRpt_120_1200 Thursday 03/15/2012	
Milliseconds	Count
0	690
10	10078
20	7510
30	5296
40	3909
50	2898
60	2093
70	1521
80	1077
90	833
100	614
110	423
120	334
130	229
140	151
150	102
160	79
170	65
180	45
190	31
200	21
210	19
220	12
230	13
240	10

2. Response Time - Total

Response Time Histogran For {Total} - select_1830_AggRpt_120_1200 Thursday 03/15/2012	
Milliseconds	Count
0	0
10	20594
20	12581
30	1737
40	310
50	79
60	38
70	31
80	21
90	10
100	12
110	5
120	6
130	3
140	3
150	6
160	5
170	1
180	3
190	4
200	2
210	5
220	5
230	2
240	3

Statistics (always produced)

The statistics directory and four statistics files in csv format are always produced. The 1830 test results are shown below in a spreadsheet layout. Note the gray “cv” column in the **Inter-arrival Times** table mentioned at the beginning of this document.

Associated Options:
-r set dispersion statistic to cv for agg, byte, rt_1st (default is vmr)
-y set arr, rt, byte probability levels (default is 90_95_99)

The four statistics files.

1. select_1830_AggRpt_120_1200_20120315_agg.csv
2. select_1830_AggRpt_120_1200_20120315_arr.csv
3. select_1830_AggRpt_120_1200_20120315_byte.csv
4. select_1830_AggRpt_120_1200_20120315_rt_1st.csv

1. Response Time, %Error, Bandwidth

Aggregate Stats [Response Time(ms)-%Err-BW] - select_1830_AggRpt_120_1200 Thursday 03/15/2012													
label	n	tps	median	mean	sdev	cv	p90	p95	p99	min	max	%error	KB/sec
010_Home	10027	8.36	9	10.99	26.15	2.38	13	16	30	5	1429	0	150.34
012_Home_jpg	10206	8.51	15	18.39	34.49	1.88	25	28	45	13	2634	0	1207.11
020_Dept	4976	4.15	7	9.82	54.17	5.52	11	13	24	5	3196	0	110.46
022_Dept_jpg	5069	4.22	5	6.47	34.73	5.37	7	9	15	4	1420	0	133.24
030_Demographics	5221	4.36	7	9.18	31.32	3.41	11	13	24	4	1608	0	72.9
040_Statistics	2574	2.15	616	614.1	38.89	0.06	636	644	695	333	1246	0	236.48
Total	38073	31.73	10	52.74	155.39	2.95	25	607	633	4	3196	0	1910.16

2. Inter-arrival Times

Inter-arrival Summary Statistics (ms) - select_1830_AggRpt_120_1200 Thursday 03/15/2012											
label	n	tps	median	mean	sdev	cv	p90	p95	p99	min	max
010_Home	10026	8.36	82	119.68	119.18	**1**	278	358	555	0	1201
012_Home_jpg	10205	8.51	81	117.57	116.97	**0.99**	270	352	527	0	1145
020_Dept	4975	4.15	172	241.15	232.47	**0.96**	555	710	1038	0	2285
022_Dept_jpg	5068	4.22	166	236.76	235.35	**0.99**	546	712	1080	0	2143
030_Demographics	5220	4.36	161	229.59	229.04	**1**	519	662	1071	0	2208
040_Statistics	2573	2.15	332	465.92	458.19	**0.98**	1062	1419	2054	0	3660
Total	38072	31.73	22	31.52	31.59	**1**	73	94	143	0	435

3. Web Page Size

Web Page Size Summary Statistics (bytes) - select_1830_AggRpt_120_1200 Thursday 03/15/2012											
label	n	tps	median	mean	sdev	cv	p90	p95	p99	min	max
010_Home	10027	8.36	17991	17991	0	0	17991	17991	17991	17991	17991
012_Home_jpg	10206	8.51	141907	141907	0	0	141907	141907	141907	141907	141907
020_Dept	4976	4.15	26632	26632	0	0	26632	26632	26632	26632	26632
022_Dept_jpg	5069	4.22	31541	31541	0	0	31541	31541	31541	31541	31541
030_Demographics	5221	4.36	16752	16734.03	60.3	0	16760	16773	16773	16512	16773
040_Statistics	2574	2.15	110193	110193	0	0	110193	110193	110193	110193	110193
Total	38073	31.73	26632	60202.92	54254.6	0.9	141907	141907	141907	16512	141907

4. Response Time to 1st Byte

Response Time 1st Byte Summary Statistics (ms) - select_1830_AggRpt_120_1200 Thursday 03/15/2012											
label	n	tps	median	mean	sdev	cv	p90	p95	p99	min	max
010_Home	10027	8.36	7	8.28	12.42	1.5	10	13	27	4	505
012_Home_jpg	10206	8.51	2	3.33	5.27	1.58	5	7	16	1	255
020_Dept	4976	4.15	4	6.18	43.34	7.01	7	10	19	3	3004
022_Dept_jpg	5069	4.22	2	3.11	4.97	1.6	4	6	12	1	184
030_Demographics	5221	4.36	5	6.37	10.11	1.59	8	10	21	3	319
040_Statistics	2574	2.15	601	598.23	35.76	0.06	617	624	674	318	971
Total	38073	31.73	5	45.61	150.15	3.29	12	592	614	1	3004

tstamp (-t option):

This option produces a trimmed file with added lead column containing readable timestamps (csv). This directory and file are intended to provide readable time stamps for diagnostic purposes.

t Option Arguments:
-t cp input files to tstamp dir(3) + column YYYYMMDDhhmmss(1) or hh:mm:ss(2)

The “select_1830_AggRpt_120_1200_20120315_tstamp.csv” file saved as a spreadsheet.

tstamp file - YYYYMMDDhhmmss

Aggregate Report Transaction Records - select_1830_AggRpt_120_1200 Thursday 03/15/2012										
YYYYMMDDhhmmss	TimeStamp_ms	Rpage_ms	WebPageName	ResponseCode	ResponseMessage	UserThread	DataType	Success	Bytes	Rbyte1_ms
20120315183203	1331861523116	9	010_Home	200	OK	Thread Group 1-97	text	TRUE	17991	7
20120315183203	1331861523145	9	010_Home	200	OK	Thread Group 1-127	text	TRUE	17991	7
20120315183203	1331861523160	5	022_Dept_jpg	200	OK	Thread Group 1-198	bin	TRUE	31541	3
20120315183203	1331861523166	9	020_Dept	200	OK	Thread Group 1-8	text	TRUE	26632	6
20120315183203	1331861523167	25	012_Home_jpg	200	OK	Thread Group 1-179	bin	TRUE	141907	2
20120315183203	1331861523169	26	012_Home_jpg	200	OK	Thread Group 1-87	bin	TRUE	141907	7
20120315183203	1331861523213	5	022_Dept_jpg	200	OK	Thread Group 1-110	bin	TRUE	31541	3
20120315183203	1331861523283	14	012_Home_jpg	200	OK	Thread Group 1-80	bin	TRUE	141907	3
20120315183203	1331861523306	15	012_Home_jpg	200	OK	Thread Group 1-52	bin	TRUE	141907	3
20120315183203	1331861523330	5	022_Dept_jpg	200	OK	Thread Group 1-95	bin	TRUE	31541	3
20120315183203	1331861523355	7	020_Dept	200	OK	Thread Group 1-168	text	TRUE	26632	5
20120315183203	1331861523466	15	012_Home_jpg	200	OK	Thread Group 1-29	bin	TRUE	141907	3
20120315183203	1331861523595	10	010_Home	200	OK	Thread Group 1-31	text	TRUE	17991	8
20120315183203	1331861523620	637	040_Statistics	200	OK	Thread Group 1-72	text	TRUE	110193	610

Additional Script Options:

Three additional script execution options are:
-e create failure file + column YYYYMMDDhhmmss(1) or hh:mm:ss(2)
-f exclude failure records from analysis
-h display help text

Directory - jmeter

The items in the JMeter directory are intended to provide the setup information required for JMeter to produce output files with the correct properties and format for the perf_arr_rt_jmeter_stats.pl script.

From a properties perspective be sure that sampleresult.timestamp.start=true in the jmeter.properties file to insure the timestamps generated are user request, not response, timestamps. This is the default setting so no action is normally required.

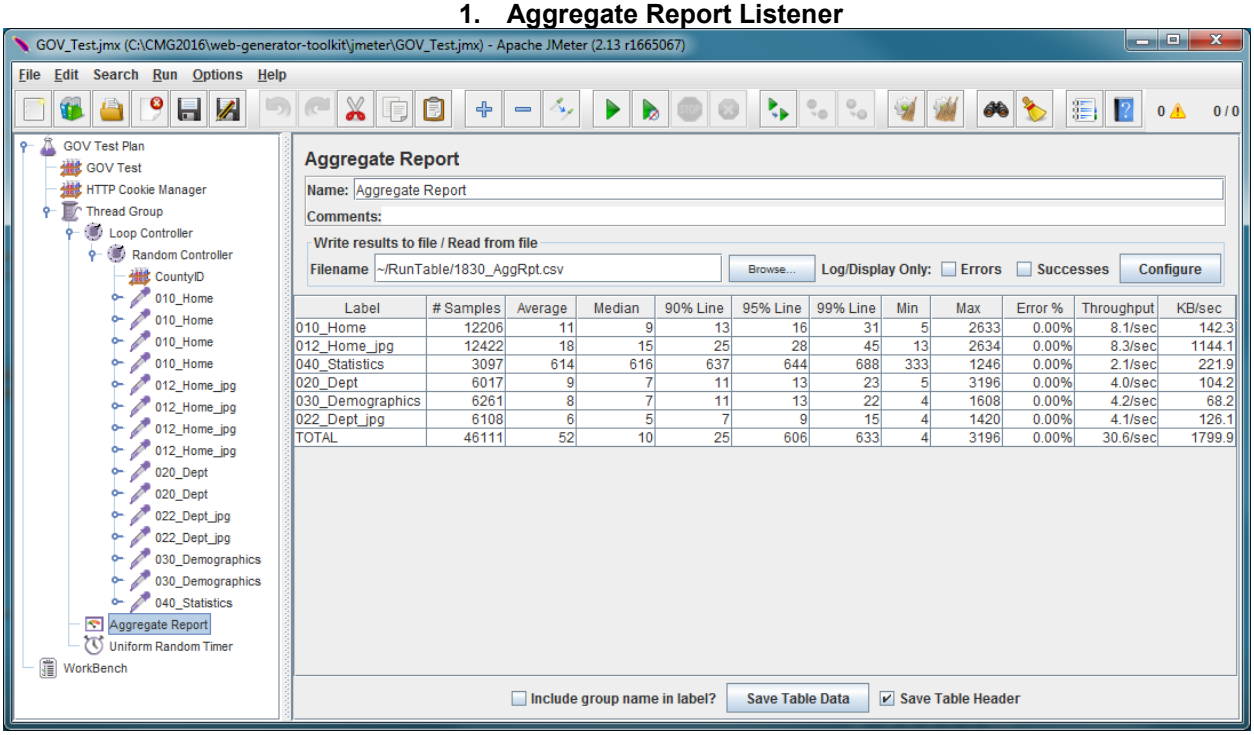
Proper data collection and formatting is accomplished by implement JMeter's Aggregate Report Listener with the correct configuration settings. The GOV_Test.jmx" file is intended to facilitate that process with the following screens as a guide.

Three screens are shown below:

1. Aggregate Report Listener – main Aggregate Report screen
2. Aggregate Report Listener Configuration – displayed when pressing the Configure button in 1.
3. Aggregate Report Listener Output File – file produced in the RunTable directory. See Filename entry in 1, i.e., ~/RunTable/1830_AggRpt.csv.

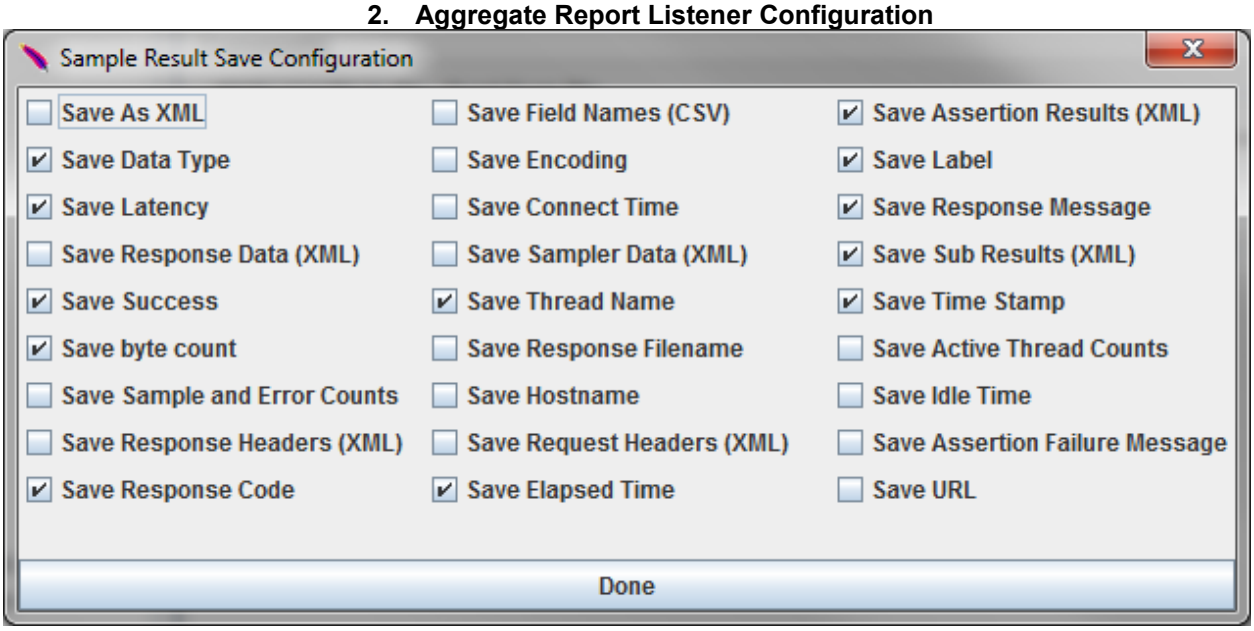
The Labels in 1 start with a 3 digit number for post- test row sorting purposes. Rows in 1 are populated in first occurrence order during the test and the perf_arr_rt_jmeter_stats.pl script puts them in sorted order when generating the four reports in its statistics directory.

1. Aggregate Report Listener



The boxes checked in 2 produce the correct output for the perf_arr_rt_jmeter_stats.pl script.

2. Aggregate Report Listener Configuration



This table shows the first fifteen records of the 1830_AggRpt.csv JMeter output file. A title row has been added and the records are being displayed in spreadsheet format for readability purposes.

3. Aggregate Report Listener Output File - 1830_AggRpt,csv

TimeStamp_ms	Rpage_ms	WebPageName	ResponseCode	ResponseMessage	UserThread	DataType	Success	Bytes	Rbyte1_ms
1331861403100	18	010_Home	200	OK	Thread Group 1-6	text	TRUE	18078	13
1331861405554	16	010_Home	200	OK	Thread Group 1-2	text	TRUE	18078	11
1331861407598	132	012_Home_jpg	200	OK	Thread Group 1-8	bin	TRUE	141994	69
1331861408175	612	040_Statistics	200	OK	Thread Group 1-9	text	TRUE	110280	594
1331861408757	79	012_Home_jpg	200	OK	Thread Group 1-4	bin	TRUE	141994	43
1331861408947	75	012_Home_jpg	200	OK	Thread Group 1-7	bin	TRUE	141994	62
1331861409146	10	020_Dept	200	OK	Thread Group 1-12	text	TRUE	26719	6
1331861409280	10	030_Demographics	200	OK	Thread Group 1-4	text	TRUE	16757	5
1331861409284	16	012_Home_jpg	200	OK	Thread Group 1-6	bin	TRUE	141907	8
1331861409324	44	012_Home_jpg	200	OK	Thread Group 1-3	bin	TRUE	141994	33
1331861410168	13	030_Demographics	200	OK	Thread Group 1-13	text	TRUE	16838	9
1331861411230	15	012_Home_jpg	200	OK	Thread Group 1-1	bin	TRUE	141994	5
1331861411974	11	020_Dept	200	OK	Thread Group 1-6	text	TRUE	26632	6
1331861412068	12	010_Home	200	OK	Thread Group 1-6	text	TRUE	17991	8
1331861412257	11	020_Dept	200	OK	Thread Group 1-7	text	TRUE	26632	6

Summary

As mentioned in the beginning of this document “How to Emulate Web Traffic Using Standard Load Testing Tools” is the conceptual basis for this toolkit. The reader is encouraged to become familiar with the ideas expressed in that paper to properly implement these tools. The illustrations in Section 4, “Applying the Methodology”, are all produced with the data and scripts contained in this package.

The web-generator-toolkit Perl scripts are set up for JMeter Aggregate Report Listener output file processing but can potentially interface with other load testing tools if the data record formats match. Since the programs are written in Perl, the scripts can be altered as well. The Perl code is heavily commented so the needed modification should be relatively easy to perform. Any enhancements that increase the utility of these tools are strongly encouraged.

Glossary

Terms in the following list are defined and referenced within the context of the web-generator-toolkit and its interface to the JMeter load tool.

Bytes: web page size in bytes (JMeter - Receive Traffic).

cv: coefficient of variation equal to sdev/mean (Script Statistics Report)

Inter-arrival Time: time between requests (JMeter - Difference in two sequential TimeStamp_ms values)

KB/sec: kilobytes/second (JMeter - Receive Traffic).

mean: the average or arithmetic mean (Script Statistics Report)

N: number of sources making requests (JMeter - Threads).

Rbyte1_ms: milliseconds taken to obtain first byte of the web page (JMeter – Aggregate Report)

Rpage_ms: milliseconds taken to render full web page (JMeter – Aggregate Report)

sdev: standard deviation or square root of the second moment about the mean (Script Statistics Report)

TimeStamp_ms: Unix time expressed in milliseconds (JMeter – Aggregate Report)

tps: transactions per second (Script Statistics Report)

vmr: variance to mean ratio equal to $sdev^2$/mean (Script Statistics Report)

Z: think time (JMeter - Timer)