web-generator-toolkit2
Documentation
jfbrady@admin.nv.gov

Overview

The **web-generator-toolkit2.pl** Perl script adds **user request concurrency** calculations to the web-generator-toolkit repository data analysis software. That is, it provides the statistical count and proportion of time there are 0, 1, 2, etc. user requests in the target system and it does this for specific request types as well as the total number of requests outstanding.

The proportion of time information correlates with the state probabilities produced by the QueState Repository software. These state probabilities may be a useful modeling substitute for user request concurrency statistics in the early stages of design or when load testing data is not available.

This document is focused on the concurrency calculations and assumes the reader is familiar with the information contained in the web-generator-toolkit repository. The JMeter output file used as an example in this repository's demo directory is the 2000_AggRpt_120_1199.csv file produced in the select directory of the web-generator-toolkit demo for the "2000" test run.

Key changes and additions to the old web-generator-toolkit environment are:

Change in data analysis Perl script name:

1. perf_arr_rt_jmeter_stats.pl -> web_generator_toolkit2.pl (old -> new)

Additions to output:

2. "concur" file in statistics directory – statistical count
3. "concur" histogram directory – proportion of time
4. "concur" graphs directory – statistical count column charts

The Perl script is in the bin directory and the additions are produced by the demo.

Package Directory Structure

The bin directory contains the computer program, the demo directory provides an example test run, and the jmeter directory includes the load testing script used to produce the data to be analyzed.

1. bin
    a. web_generator_toolkit2.pl – data analysis script
    b. Graph.pm – needs GD.pm
2. demo
    a. Input file - 2000_AggRpt_120_1199.csv
    b. Program run file - run_web_generator_toolkit2.pl
3. jmeter
    a. GOV_Test.jmx – jmeter script which matches demo layout.
    b. DataSet directory - CountyID.csv

Package Setup

The environmental variable **WEB_GEN_TOOLKIT**2 needs to be created pointing to the **bin** directory. Use "set" for Windows and "export" for Unix / Linux.

Script Execution

The script is executed by running the "run_web-generator-toolkit2.pl Perl script in the demo directory. That script execution produces a "2000_AggRpt_120_1199" directory which contains all the output.

Output Directory - 2000_AggRpt_120_1199

The sub-directories are:

1. **statistics** – arrivals, response times, concurrency counts, and web page sizes (csv).
2. **histograms** – arrival, response time, and concurrency proportion histograms (csv).
3. **graphs** – column charts for a myriad of variables including concurrency counts (png).

statistics (always produced)

The statistics directory contains five files in csv format. The 2000 test run results for all five are shown in Figure 1 below. The new "concur" file provides a statistical count of concurrent events for individual web pages as well as the total.

The five statistics files.

1. demo_2000_AggRpt_120_1199_20120315_agg.csv
2. demo_2000_AggRpt_120_1200_20120315_arr.csv
3. demo_2000_AggRpt_120_1200_20120315_byte.csv
4. **demo_2000_AggRpt_120_1200_20120315_concur.csv**
5. demo_2000_AggRpt_120_1200_20120315_rt_1st.csv

1. Response Time, %Error, Bandwidth

Aggregate Stats [Response Time(ms)-%Err-BW] - demo_2000_AggRpt_120_1199 Thursday 03/15/2012													
label	n	tps	median	mean	sdev	cv	p90	p95	p99	min	max	%error	KB/sec
010_Home	24384	20.34	9	31.6	161.33	5.11	20	62	516	5	3616	0	365.9
012_Home_jpg	24950	20.81	20	42.23	173.89	4.12	41	65	301	13	3864	0	2953.03
020_Dept	12489	10.42	7	34.79	184.64	5.31	19	61	1412	4	3227	0	277.46
022_Dept_jpg	12533	10.45	5	30.07	183.2	6.09	14	33	1298	3	3430	0	329.73
030_Demographics	12424	10.36	7	30.04	164.54	5.48	17	53	1220	4	3217	0	173.43
040_Statistics	6204	5.17	622	638.7	171.31	0.27	674	768	1038	334	4060	0	570.05
Total	92984	77.55	11	74.97	228.82	3.05	88	617	1048	3	4060	0	4669.51

2. Inter-arrival Times

Inter-arrival Summary Statistics (ms) - demo_2000_AggRpt_120_1199 Thursday 03/15/2012											
label	n	tps	median	mean	sdev	cv	p90	p95	p99	min	max
010_Home	24383	20.34	34	49.17	49.55	1.01	113	148	228	0	533
012_Home_jpg	24949	20.81	33	48.06	48.66	1.01	111	145	227	0	527
020_Dept	12488	10.42	67	95.99	94.58	0.99	219	285	432	0	903
022_Dept_jpg	12532	10.45	66	95.67	95.74	1	221	288	437	0	871
030_Demographics	12423	10.36	67	96.51	97.87	1.01	224	295	450	0	990
040_Statistics	6203	5.17	133	193.24	196.31	1.02	445	579	910	0	1847
Total	92983	77.55	9	12.89	13.22	1.03	30	39	59	0	341

3. Web Page Size

Web Event Size Summary Statistics (bytes) - demo_2000_AggRpt_120_1199 Thursday 03/15/2012											
label	n	tps	median	mean	sdev	cv	p90	p95	p99	min	max
010_Home	24384	20.34	17991	17991	0	0	17991	17991	17991	17991	17991
012_Home_jpg	24950	20.81	141907	141907	0	0	141907	141907	141907	141907	141907
020_Dept	12489	10.42	26632	26632	0	0	26632	26632	26632	26632	26632
022_Dept_jpg	12533	10.45	31541	31541	0	0	31541	31541	31541	31541	31541
030_Demographics	12424	10.36	16752	16735.56	58.06	0	16760	16773	16773	16512	16773
040_Statistics	6204	5.17	110193	110193	0	0	110193	110193	110193	110193	110193
Total	92984	77.55	26632	60211.89	54220.05	0.9	141907	141907	141907	16512	141907

4. Concurrent Events

Concurrent Event Statistics (Count) - demo_2000_AggRpt_120_1199 Thursday 03/15/2012											
label	**n**	**tps**	**median**	**mean**	**sdev**	**cv**	**p90**	**p95**	**p99**	**min**	**max**
010_Home	**47751**	**39.83**	**1**	**1.12**	**1.13**	**1.01**	**2**	**3**	**5**	**0**	**10**
012_Home_jpg	**48798**	**40.7**	**1**	**1.37**	**1.25**	**0.91**	**3**	**4**	**5**	**0**	**13**
020_Dept	**24719**	**20.62**	**1**	**0.85**	**0.85**	**1**	**2**	**2**	**4**	**0**	**7**
022_Dept_jpg	**24825**	**20.71**	**1**	**0.81**	**0.77**	**0.96**	**2**	**2**	**3**	**0**	**6**
030_Demographics	**24560**	**20.49**	**1**	**0.8**	**0.81**	**1.02**	**2**	**2**	**3**	**0**	**7**
040_Statistics	**12331**	**10.28**	**4**	**3.83**	**1.95**	**0.51**	**6**	**7**	**9**	**0**	**13**
Total	**171699**	**143.2**	**5**	**6.21**	**3.66**	**0.59**	**11**	**13**	**19**	**0**	**33**

5. Response Time to 1st Byte

Response Time 1st Byte Summary Statistics (ms) - demo_2000_AggRpt_120_1199 Thursday 03/15/2012											
label	n	tps	median	mean	sdev	cv	p90	p95	p99	min	max
010_Home	24384	20.34	7	16.25	80.87	4.98	17	43	175	4	3020
012_Home_jpg	24950	20.81	3	8.92	91.59	10.26	9	19	69	1	3016
020_Dept	12489	10.42	5	13.63	81.04	5.94	14	38	166	2	3009
022_Dept_jpg	12533	10.45	3	8.35	81.78	9.79	10	20	73	1	3021
030_Demographics	12424	10.36	5	14.88	88.91	5.98	14	37	165	3	3010
040_Statistics	6204	5.17	606	608.76	97.31	0.16	640	698	892	317	3836
Total	92984	77.55	6	52.22	172.02	3.29	44	593	629	1	3836

Figure 1: Statistical Reports

histograms (-d option)

The histograms directory produces three types of histograms, arrival, concurrency, and response time. The arrival and response time information is setup slightly differently than it is in the original web-generator-toolkit and the concurrency reports are new with web-generator-toolkit2. The sub-directories below the histograms directory are:

1. arr - arrival
2. concur - concurrency
3. rt - response time

Figure 2 is a segment of the “Total” histogram csv file for concurrency.

Concurrent Events (Count) {Total} - demo_2000_AggRpt_120_1199 Thursday 03/15/2012											
	----------Statistics----------										
Statistic	mean	sdev	var	max	p_sum	freq_sum					
Value	5.812479	3.653162	13.3456	33	1	171699					
State	0	1	2	3	4	5	6	7	8	9	10
Prob	0.011811	0.043135	0.08711	0.131772	0.147837	0.138664	0.112959	0.085051	0.061406	0.044809	0.033877
Freq	1091	4991	11664	19359	24583	24918	21655	16669	12091	8725	6538

Figure 2: Partial contents of the demo_2000_AggRpt_120_1199_Total_concur.csv file

This file segment provides proportion of time, “Prob”, and frequency count, “Freq”, numbers through State 10 as well as statistics, “Value”, based on the proportionality information.

graphs (-g option)

The graphs directory contains column charts in png format for each web page and total. These graphs are in the following sub-directories: agg, agg_rt, arr, byte, **concur**, rt_1st. The concur directory is an addition to this list with web-generator-toolkit2 and Figure 3 is an example of these concurrency graphs.

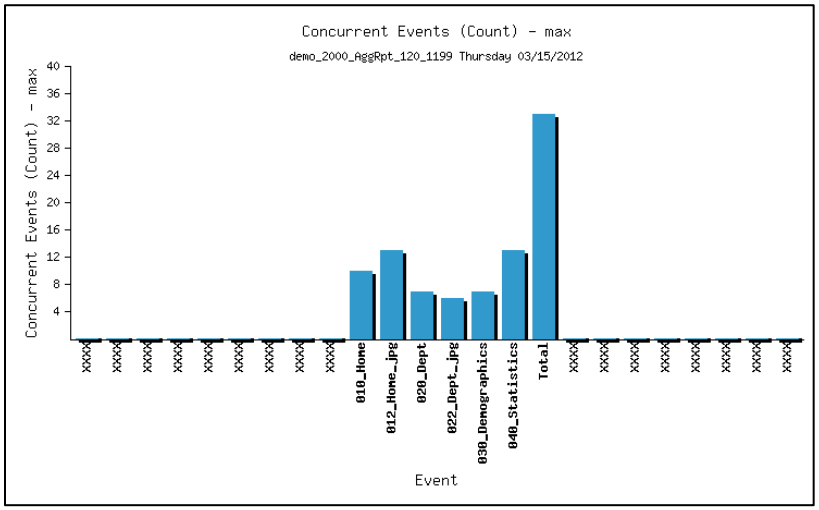

Figure 3: Max Concurrent Events (demo_2000_AggRpt_120_1199_20120315_concur_11_max.png)

The vertical axis numbers associated with each column in Figure 3 match the data contained in the "max" column of the Figure 1 "4. Concurrent Events" table. For example, Figure1 and Figure 3 both have a value of 33 for "Total".

Output Summary

Figure 4 is extracted from the QueState_doc.xlsx spreadsheet contained in this and the QueState repository. It shows the concurrency proportions produced by the web-generator-toolkit2 software as a think black line and the state probabilities calculated for the QueState queueing models in thin lines of various colors.

Table A in Figure 4 is created from the Figure 2 histogram information and Table B is QueState model input based on Table A. Table C shows the first twenty-five 2000_Agg proportion of time decimal fractions along with a comparable set of state probabilities for the eight queueing models.

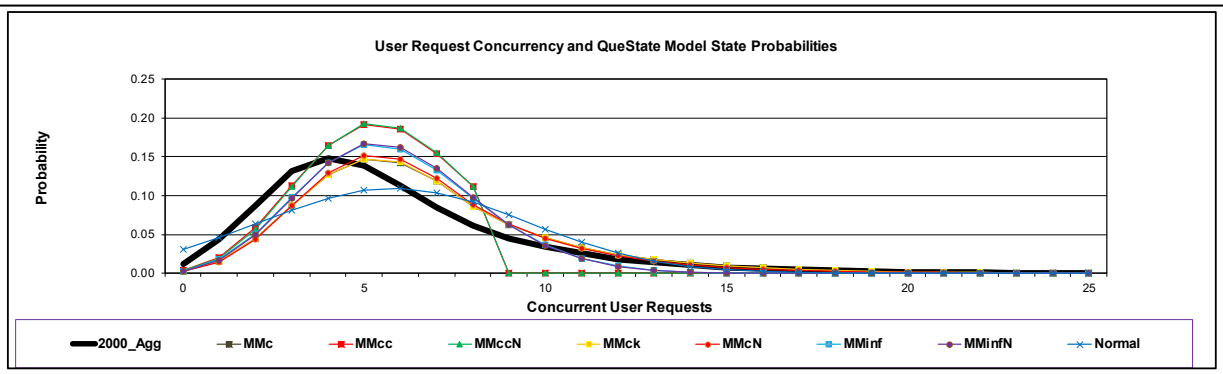

Table A	User Request Concurrency					
	Concurrent Requests					
Test Run	mean	sdev	var	max	p_sum	f_sum
2000_Agg	5.81	3.65	13.35	33	1.00	171699

Table B	QueState Model Input				lamda				
Model	Servers	N	Qsize	p[show]	lamda	mu	/mu	var	sdev
MMc	8	-	-	25	77.55	13.34	5.81	-	-
MMcc	8	-	-	-	77.55	13.34	5.81	-	-
MMccN	8	200	-	-	77.55	13.34	5.81	-	-
MMck	8	-	25	-	77.55	13.34	5.81	-	-
MMcN	8	200	-	-	77.55	13.34	5.81	-	-
MMinf	8	-	-	25	77.55	13.34	5.81	-	-
MMinfN	8	200	-	-	77.55	13.34	5.81	-	-
Normal	8	-	-	25	77.55	13.34	5.81	13.35	3.65

Table C	State Probabilities																									
StateProb	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Total
2000_Agg	0.0118	0.0431	0.0871	0.1318	0.1478	0.1387	0.1130	0.0851	0.0614	0.0448	0.0339	0.0263	0.0182	0.0144	0.0109	0.0078	0.0062	0.0051	0.0037	0.0025	0.0018	0.0014	0.0010	0.0007	0.0006	0.0003	0.9994
MMc	0.0027	0.0155	0.0450	0.0871	0.1266	0.1472	0.1426	0.1184	0.0861	0.0625	0.0455	0.0330	0.0240	0.0174	0.0127	0.0092	0.0067	0.0049	0.0035	0.0026	0.0019	0.0014	0.0010	0.0007	0.0005	0.0004	0.9990
MMcc	0.0035	0.0201	0.0583	0.1130	0.1642	0.1909	0.1849	0.1536	0.1116	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
MMccN	0.0032	0.0189	0.0562	0.1111	0.1639	0.1923	0.1871	0.1552	0.1121	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
MMck	0.0027	0.0155	0.0450	0.0872	0.1267	0.1474	0.1428	0.1186	0.0862	0.0626	0.0455	0.0331	0.0240	0.0175	0.0127	0.0092	0.0067	0.0049	0.0035	0.0026	0.0019	0.0014	0.0010	0.0007	0.0005	0.0004	1.0000
MMcN	0.0025	0.0148	0.0441	0.0872	0.1286	0.1509	0.1468	0.1218	0.0880	0.0632	0.0452	0.0321	0.0227	0.0160	0.0112	0.0078	0.0054	0.0037	0.0025	0.0017	0.0012	0.0008	0.0005	0.0004	0.0002	0.0002	0.9997
MMinf	0.0030	0.0174	0.0505	0.0978	0.1422	0.1653	0.1601	0.1330	0.0966	0.0624	0.0363	0.0192	0.0093	0.0042	0.0017	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
MMinfN	0.0027	0.0164	0.0489	0.0966	0.1424	0.1671	0.1626	0.1349	0.0974	0.0622	0.0356	0.0184	0.0087	0.0038	0.0015	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Normal	0.0307	0.0458	0.0633	0.0812	0.0966	0.1066	0.1092	0.1037	0.0913	0.0747	0.0566	0.0398	0.0260	0.0157	0.0088	0.0046	0.0022	0.0010	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.9586

Figure 4: User Request Concurrency (2000_Agg) and QueState Model State Probabilities

The model numbers that are most closely aligned with the web-generator-toolkit2 results may be useful in estimating concurrency for similar projects.

Summary

The web-generator-toolkit2.pl Perl script enhances the functionality of the original web-generator-toolkit by adding user request concurrency analysis.

This script possesses two pieces of functionality not supported by most load testing data analysis tools:

1. Inter-arrival Time "cv" statistics for assessing the level of load tool request independence.
2. User request concurrency calculations for estimating target system resource use maximums.

A deeper understanding of how the full set of web-generator-toolkit2 features improve the quality of performance engineering data analysis can be gained by reviewing three of the references [2], [4], and [5]. The concurrency analysis provided here expands upon the ideas contained in these three documents by applying the same load testing data they used as an illustrative base.

Any ideas that increase the utility of these tools is strongly encouraged. One thought that has been successfully implemented is the processing of production system log files containing timestamp and latency information.

References

[1.] A.O. Allen, “Probability, Statistics, and Queueing Theory”, Academic Press, Inc., Orlando, Florida, 1978.

[2.] J. F. Brady, “When Load Testing Large User Population Web Applications the Devil Is In the (Virtual) User Details,” CMG *Proceedings* 2012, http://www.perfdynamics.com/Classes/Materials/Brady-CMG12.pdf

[3.] J. F. Brady, “It’s Time to Retire Our 1970’s User Demand Model for Transaction Computing and Adopt One That Reflects Modern Web User Traffic Flow,” CMG *Proceedings*, 2014, https://jamesbrady.academia.edu/research

[4.] J. F. Brady and N. J. Gunther, “How to Emulate Web Traffic Using Standard Load Testing Tools,” CMG *Proceedings*, 2016, https://arxiv.org/abs/1607.05356

[5.] J. F. Brady, “Is Your Load Generator Launching Web Requests In Bunches?,” CMG *Proceedings* 2019, https://arxiv.org/abs/1809.10663

[6.] R. B. Cooper, “Introduction to Queueing Theory”, Elsevier Science Publishing Co., Inc, New York, N.Y., (1984).

[7.] W.C. Giffin, “Queueing: Basic Theory and Applications”, Grid, Inc, Columbus, Ohio, 1978.

[8.] L. Kleinrock, “Queueing Systems Volume 1 and 2”, John Wiley & Sons, New York, N.Y., (1975).

Acknowledgement

User request concurrency proportion of time Perl script logic developed by Xiaosong Lou.

Glossary

Terms in the following list are defined and referenced within the context of the web-generator-toolkit2.

cv: coefficient of variation equal to sdev/mean

inter-arrival time: time between requests

KB/sec: kilobytes/second

mean: the average or arithmetic mean

sdev: standard deviation

tps: transactions per second