File Data Reduction

Patrick Stanko

SNIA Emerald™ Training

SNIA Emerald™ Power Efficiency Measurement Specification

Version 3.0

February-March 2018

Data Reduction of sFlow and power meter

SNIA GSI | GREEN STORAGE

- Required data reduction steps
  - Extract the data from text or CSV file from sFlowtool
    - Cut and past the IF_(index) from the text file
    - Extract the data from the CSV file (need to use –4 -L)
  - Convert sFlow data into MiB/s with time stamps
  - Identify the workload start and stop time stamps
    - Can use tag2014
  - Combine the sFlow and power data
  - Use linear interpolation on sFlow data to match power meter
  - Generate 30 Metric points (MiB/s/W)

Validation and Metric Generation

- Stability Validation
  - test using the moving average
  - test using the least fit squares
- Generate Metric
  - average performance/average power across the 300 second period

Generate the MiB/s from sFlow data

- Align UTC time stamps with in bytes and out bytes counters
- Formulate the in byte/sec and out byte/sec
  - Difference between counters/ difference between sample UTC times
- Add in byte/sec and out byte/sec for total bytes/sec
- Divide by 1048576 (MiB) to derive MiB/s

$$\frac{Counter_n - Countern\_1}{UTC_n - UTC_{n\_1}}$$

	A	B	C	D	E	F	G	H
1	unixSecondsUTC	In Bytes	Out Bytes		In Byte/sec	Out Byte/Sec	Total Byte/sec	Total MiB/s
2	1519773493	629701134054 47	93468768627092		(B3-B2)/(A3-A2)	(C3-C2)/(A3-A2)	E3+F3	G3/1048576
3	1519773503	629701134189 72	93468768650965		1353	2387	3740	0.004
4	1519773513	629701134207 00	93468768660429		173	946	1119	0.001
5	1519773523	629701134222 36	93468768673146		154	1272	1425	0.001

Match sFlow time stamp to Power meter

- Need to convert unixSecondsUTC time into MS Excel time stamp
  - Find the number of days since January 1, 1970
  - Adjust to match power supply time zone

	A	B	C
1	unixSecondsUTC	UTC Time	MST Time
2		(((A3/60)/60)/24)+DATE(1970,1,1)	B3-(7/24)
3	1519773493	2/27/2018 23:18:13	2/27/2018 16:18:13
4	1519773503	2/27/2018 23:18:23	2/27/2018 16:18:23
5	1519773513	2/27/2018 23:18:33	2/27/2018 16:18:33
6	1519773523	2/27/2018 23:18:43	2/27/2018 16:18:43

Combine sFlow and Power Meter Data

SNIA GSI | GREEN STORAGE

- Combine the sFlow and Power Meter Data in spread sheet
  - Use linear interpolation to align performance data to the power meter data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Sflow Time	sFlow Hour	sFlow Minute	sFlow Second	MiB/s									K3-D3	E2+((D3+N3-D2)/(D3-D2))*(E3-E2)
2	7:30:27 PM	19	30	27	510.3698		Sample	Power Time	Power Hour	Power Minute	Power Second	Watts		Difference in Seconds	Linear Interpolation MiB/S
3	7:30:37 PM	19	30	37	463.9902		0	7:30:33 PM	19	30	33	524		-4	482.542
4	7:30:47 PM	19	30	47	523.0991		1	7:30:43 PM	19	30	43	519		-4	499.456
5	7:30:58 PM	19	30	58	450.7144		2	7:30:53 PM	19	30	53	530		-5	483.617

Generate the metric samples

- Take the aligned performance data divided by power
- Generate 30 samples
  - 10 second sample rate over 300 seconds

	A	B	C	D	E	F
1	Sample	Power Time	Watts	Linear Interpolation MiB/s		SNIA Emerald File IO Metric MiB/s/W
2						D3/C3
3	0	7:30:33 PM	524	483		0.920
4	1	7:30:43 PM	519	499		0.963
5	2	7:30:53 PM	530	484		0.913

Validate using move average

- Need to generate Sample 0 (S0)
  - The average across all 30 samples
- Generate the moving average across the 30 sample
  - Sn = Mn*w+(1-w)*Sn-1
- Verify that all 30 points are with in ±5% of So

	A	B	C	D	E
			Sn	Moving Average	% Change
1	Sample	Metric	Average(1:30)	B4*w+(1-w)*D3	(Sn-S0)/S0
2	0	0.920	0.874		
3	1	0.963		0.883	1.02%
4	2	0.913		0.886	1.36%

Validate using least squares linear fit

- Find the Slop of M
  - $Slope(M) = (\sum_{n=1}^{k} M_{N+n}(12n - 6K - 6))/(K(K-1)(K+1)$
- Find the Int(M)
  - $Int(M) = (\sum_{n=1}^{K} M_{N+n})K \ - Slope(M)(K+1)/2$
- Find Y sample 30 and sample 1
  - $Y{=}n{*}Slope(M){+}Int(M)$

	A	B	C	D	E	F	G	H
			Slope Term	Slope M	Int	Y30	Y1	Difference
1	Sample	Metric	B3*(12*A3-6*30-6)/(30*(30-1)*(30+1))	sum(c3:C33)	AVERAGE(B4:B33)-(D3*((30+1)/2))	30*D3+E3	1*D3+E3	(F3-G3)/G3
2	0	0.920		0.000507286	0.866	0.881	0.867	1.70%
3	1	0.963	-0.006					
4	2	0.913	-0.005					

Compute metric

- Calculate the metric using the 30 points
  - Average(MiB/s)/Average(Power)

	A	B	C	D	E
1					Metric
2	Sample	Power Time	Linear Interpolation MiB/s	Watts	Average(C3:C33)/Average(D3:D33)
3	0	7:30:33 PM	483	524.45	0.875
4	1	7:30:43 PM	499	518.7	
5	2	7:30:53 PM	484	529.95	
6	3	7:31:03 PM	459	535.6	