

Two Case Studies of the Application of Dynamic Modeling Techniques in Performance Assessment and Prediction of Complex Shared Storage Architectures

Marti Bancroft, Phillip L. (Rocky) Snyder, and
Mark Woodyard

Sun Microsystems, Palo Alto, California, USA

HPC Technology Engineering

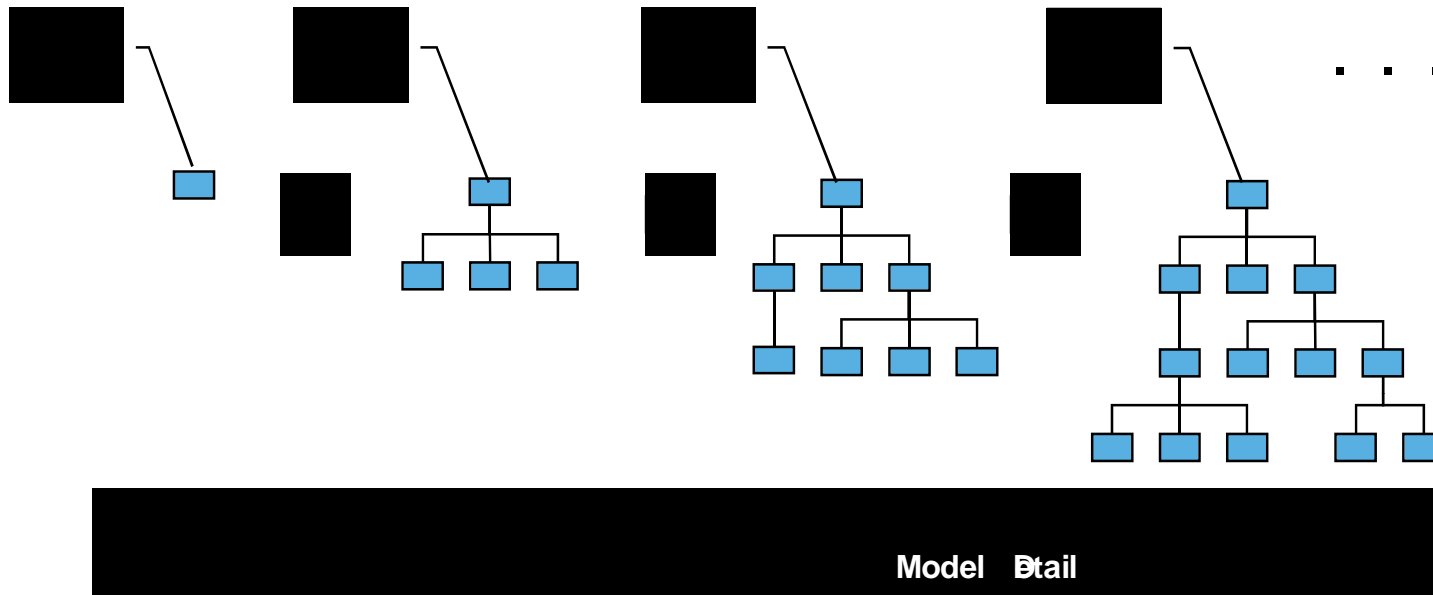
HPC Horizontal Solutions

Outline

- Role of dynamic models in predicting the behavior of large systems
- Systems being studied
- Test data which invalidated our models
- Next steps

Overview of Findings

- Oops! (no, not Object Oriented Programming System, but... new data invalidated our models - an oops!)
- Customer calls this “frictionless I/O”
- Key question: how far from ‘local’ will this be valid?
- Key question: at what load will this return to “normal”? I.e. serious reduction in efficiency...



As model detail increases,
 Model behavior more closely resembles the true system
 behavior

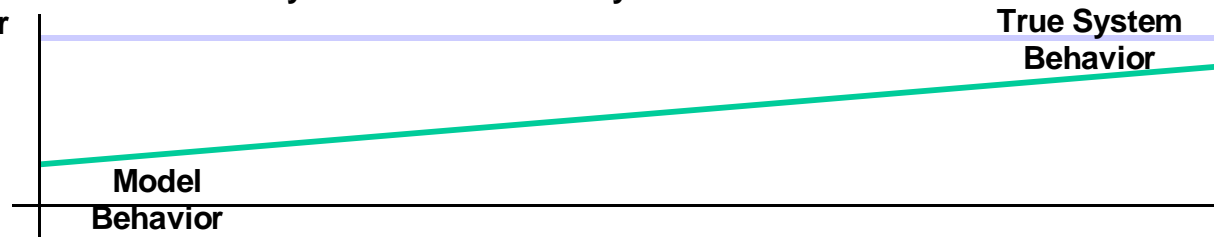


Figure 3 Evolution of model detail over time

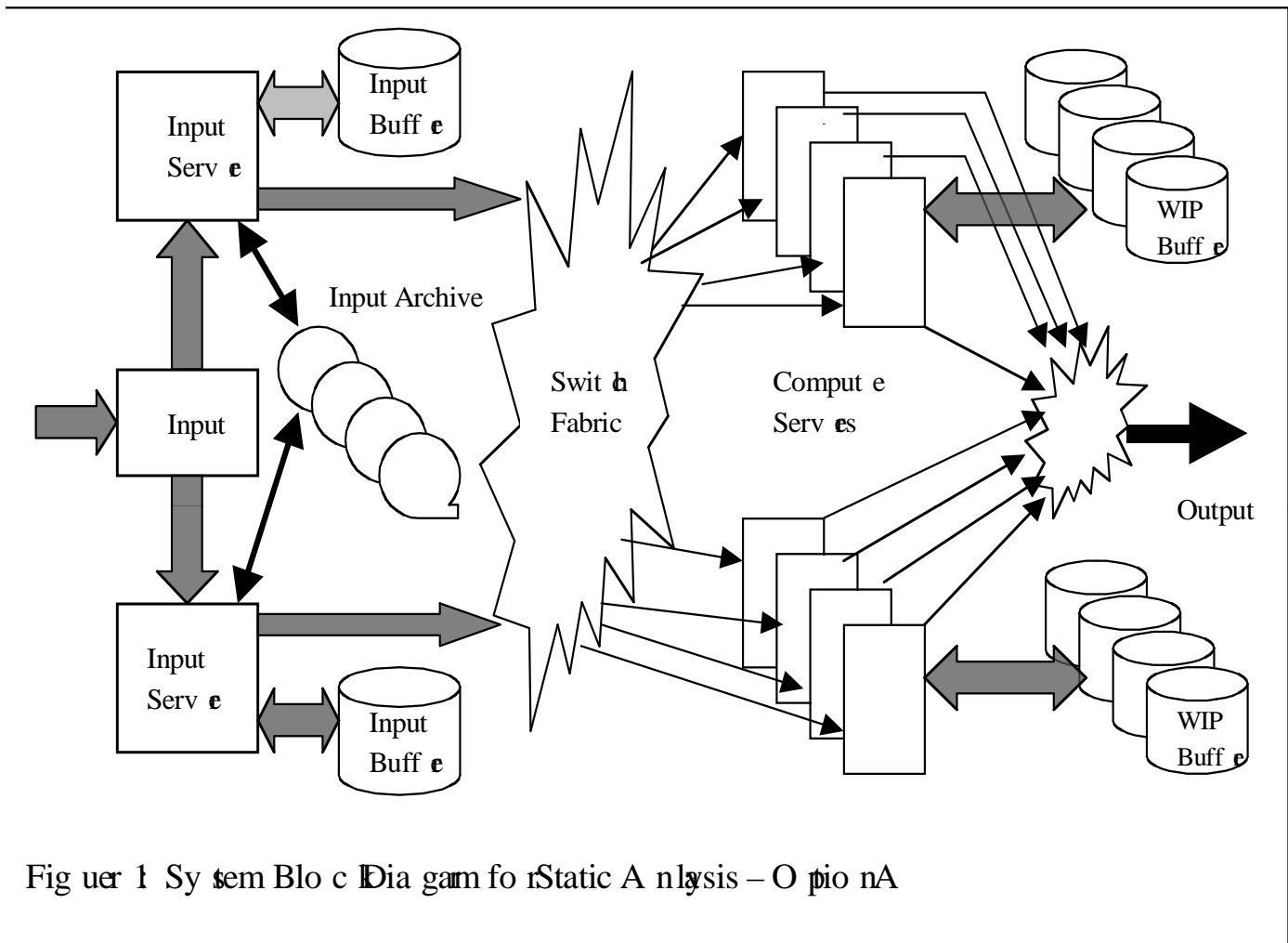


Figure 1 System Block Diagram for Static Analysis - OptioNA

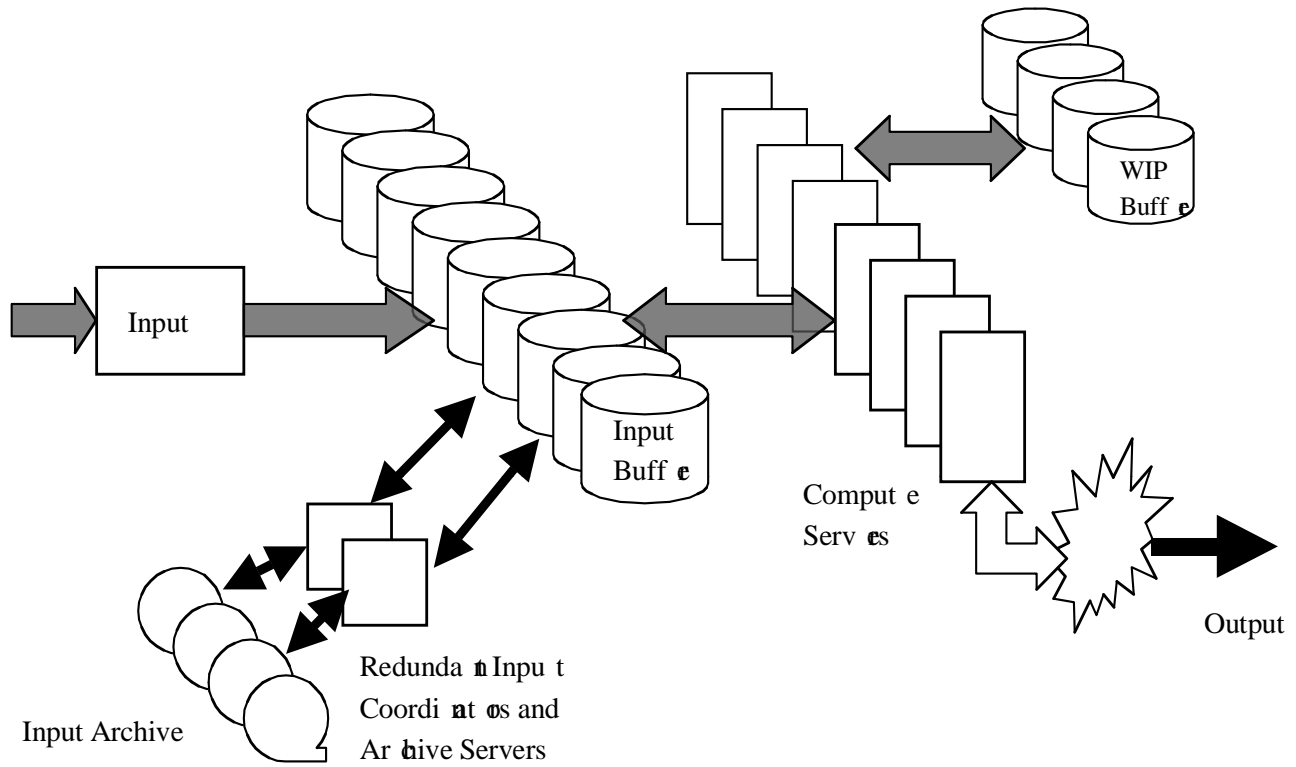
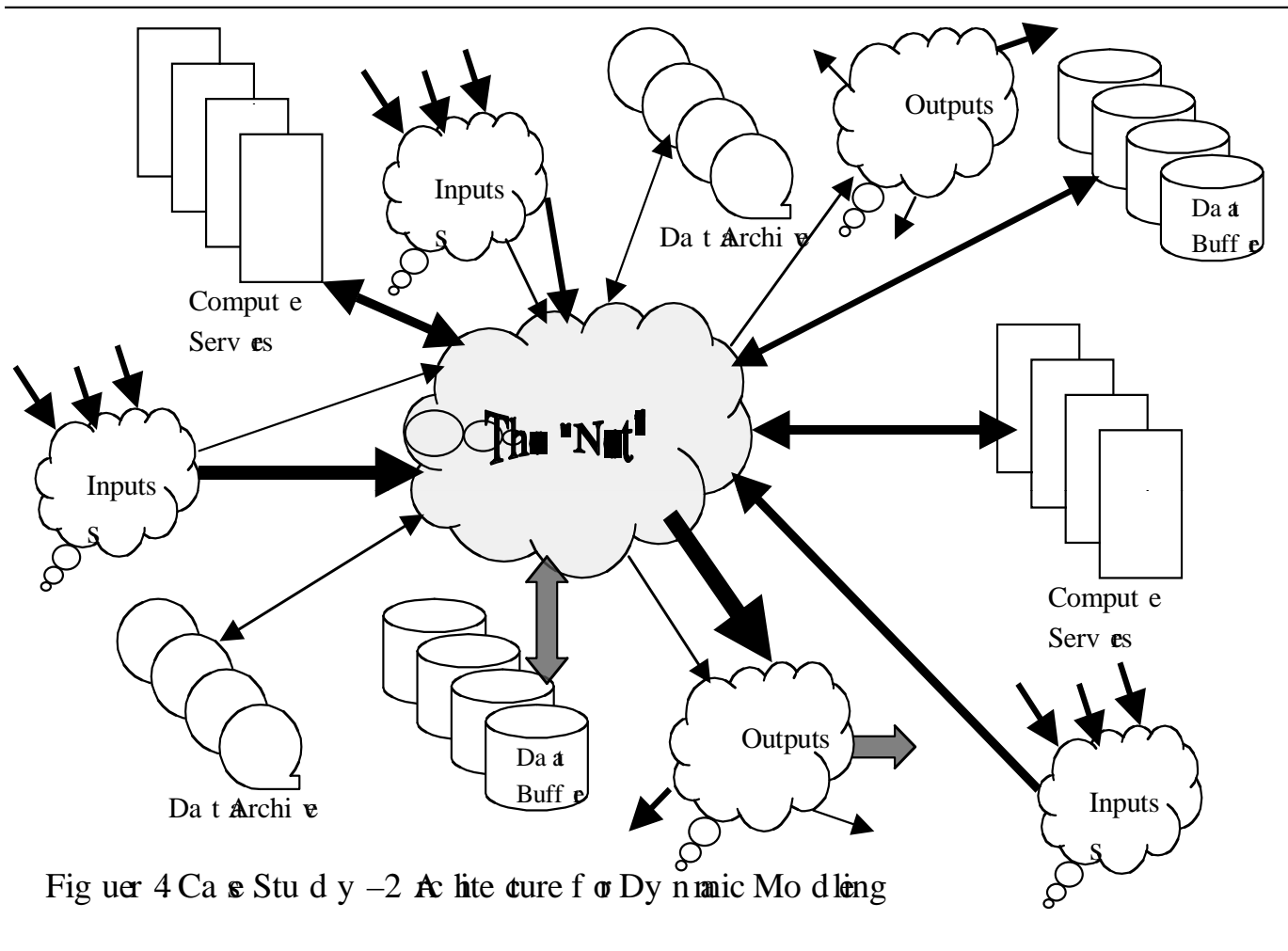


Figure 2 System Block Diagram for Static Analysis - Option B



SANergy client writes to 105GB file, preallocated, sustaining more than 1 Gigabyte per second, using 12 threads (LSC's recommended minimum) (1008 MB/s shown) (QFS 3.5.0-28b)

The screenshot shows a Netscape browser window displaying the Tivoli SANergy configuration page. The browser title is "Netscape: bangor Tivoli SANergy Unix Configuration To". The address bar shows "file:/tmp/SANergyConfig.htm". The page features a header with the Tivoli SANergy logo and version information: "Version: 2.2.1.3 Eval Registration String 294 days".

The main content area is divided into several sections:

- FusedMount:** A table listing mounted volumes. The "/san" volume is checked and mounted to "buddy:/san".
- Server:** Lists servers: "bangor.vold(pid286)", "stager/export/share", and "buddy:/san".
- Performance Statistics:** A table showing reads and writes in bytes and MB/sec.
- General:** Configuration options for tracing, hyper-extend, fuse exclude, and minread/minwrite.
- Cache:** Configuration options for mode, line size, process, exclude, and #maps.
- Control:** Buttons for "Apply", "Cancel", and "Defaults".

Reads (bytes)	
Fused	0
Cached	0

Writes (bytes)	
Fused	127,987,089,408
Cached	0

MB/sec	
Reads	0.00
Writes	1008.00

Buttons: [Clear](#) [Refresh OFF](#) [ZOOM Stats](#)

SANergy client
 writes to 105GB
 file, 18 cpus,
 preallocated,
 sustaining more
 than 1 Gigabyte
 per second,
 using 12 threads
 (LSC's
 recommended
 minimum) (1008
 MB/s on screen)

IOSTAT:

device	extended device statistics									tty		cpu			
	r/s	w/s	kr/s	kw/s	wait	actv	svc_t	%w	%b	tin	tout	us	sy	wt	id
sd0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	0	0	652	5	3	91	0
ssd0	0.0	161.6	0.0	82738.4	0.0	10.0	61.8	0	100						
ssd1	0.0	161.6	0.0	82738.4	0.0	9.1	56.5	0	100						
ssd2	0.0	161.2	0.0	82533.6	0.0	9.4	58.1	0	100						
ssd3	0.0	161.8	0.0	82840.8	0.0	9.3	57.6	0	100						
ssd4	0.0	162.0	0.0	82943.2	0.0	9.3	57.6	0	100						
ssd5	0.0	161.8	0.0	82840.8	0.0	8.4	52.1	0	100						
ssd6	0.0	161.8	0.0	82840.8	0.0	9.0	55.5	0	100						
ssd7	0.0	162.2	0.0	83045.6	0.0	8.8	54.6	0	100						
ssd8	0.0	161.8	0.0	82840.8	0.0	8.7	53.5	0	100						
ssd9	0.0	161.4	0.0	82636.1	0.0	9.4	58.4	0	100						
ssd10	0.0	161.6	0.0	82738.5	0.0	8.9	55.2	0	100						
ssd11	0.0	162.0	0.0	82943.3	0.0	8.4	51.6	0	100						

SANergy client
writes to 105GB
file, preallocated,
6 threads (0.5 per
T3 brick) (918
MB/s shown)

Netscape: bangor Tivoli SANergy Unix Configuration To

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop

Bookmarks Location: file:/tmp/SANergyConfig.htm What's Related

Members WebMail Connections BizJournal SmartUpdate Mktplace

Tivoli SANergy
Version: 2.2.1.3 Eval Registration String 294 days [Registration String...](#)

FusedMount **Server** [Help](#)

<input type="checkbox"/> /vol	bangor:vold(pid286)
<input type="checkbox"/> /import/share	stager:/export/share
<input checked="" type="checkbox"/> /san	buddy:/san

[Set!](#) [Refresh](#)

[Perf Tests](#) [Xterm...](#) [Ownership...](#)

Clear	Refresh OFF
Reads (bytes)	
Fused	220,194,669,568
Cached	0
Writes (bytes)	
Fused	317,221,502,976
Cached	0
MB/sec	
Reads	0.00
Writes	918.00

[ZOOM Stats](#)

General **Cache** **Control**

Tracing Off [EventLog Help](#)

Hyper-extend Excl

Fuse Exclude MinRead

MinWrite

Mode [Help](#)

Line size (KB) Total process (MB)

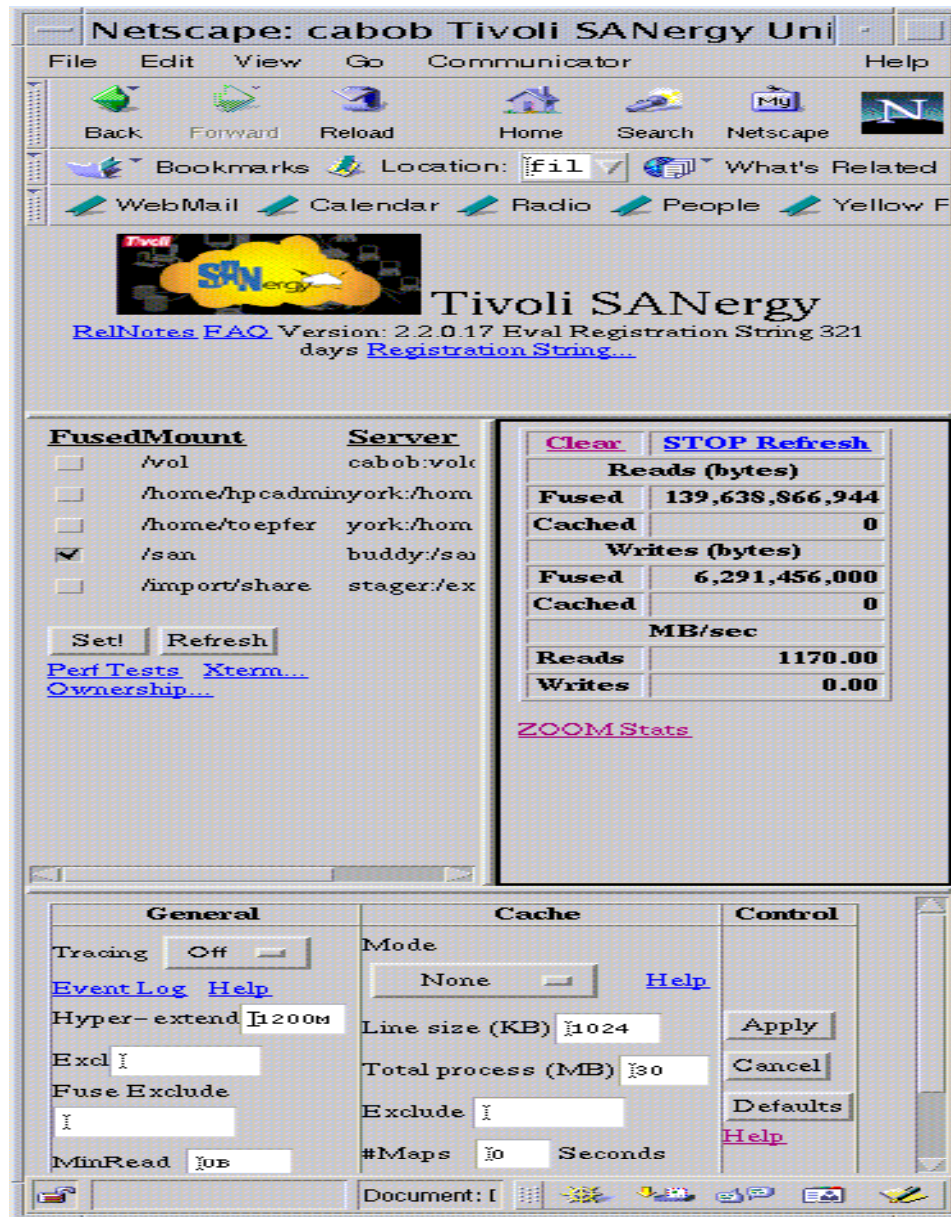
Exclude

#Maps seconds

Use comma-separated tails for exclusions, example: *.gif,.db

Document: Done.

SANergy client reads from 105GB file, 6 threads (0.5 per T3 brick) - using prerelease version at customer demos - shown bursting at 1170 MB/s



SANergy client writes to 105GB file, 12 threads but without file preallocation, 1GB extents, showing almost 1 Gigabyte per second (999 MB/second)

The screenshot shows a Netscape browser window titled "Netscape: bangor Tivoli SANergy Unix Configuration To". The address bar shows "file:/tmp/SANergyConfig.htm". The page content includes the Tivoli SANergy logo and version information: "Version: 2.2.1.3 Eval Registration String 294 days".

FusedMount

	Server	Help
<input type="checkbox"/> /vol	bangor.vold(pid286)	
<input type="checkbox"/> /import/share	stager:/export/share	
<input checked="" type="checkbox"/> /san	buddy:/san	

Buttons: Set, Refresh

[Perf Tests](#) [Xterm...](#) [Ownership...](#)

Performance Statistics

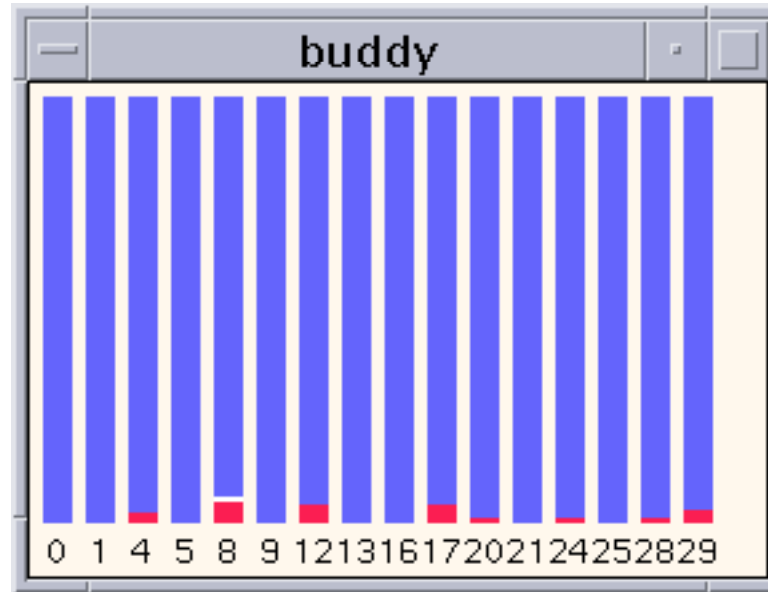
Clear		Refresh OFF	
Reads (bytes)			
Fused			0
Cached			0
Writes (bytes)			
Fused		13,872,660,480	
Cached		0	
MB/sec			
Reads		0.00	
Writes		999.00	

[ZOOM Stats](#)

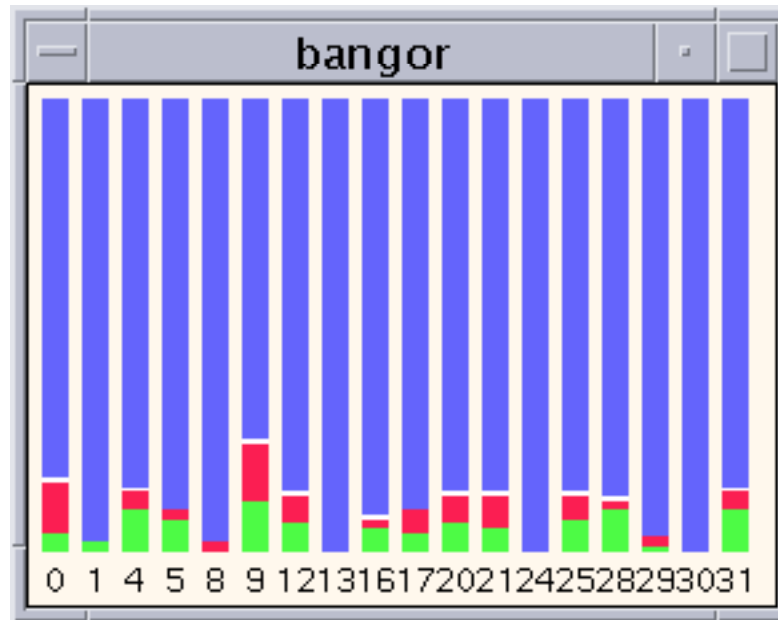
Configuration Options

General	Cache	Control
Tracing: Off <input type="checkbox"/> Event Log Help	Mode: None <input type="text"/> Help	Apply Cancel Defaults Help
Hyper-extend: 1024M <input type="text"/> Excl: <input type="text"/>	Line size (KB): 1024 <input type="text"/> Total process (MB): 30 <input type="text"/>	
Fuse Exclude: <input type="text"/> MinRead: 10B <input type="text"/>	Exclude: <input type="text"/>	
MinWrite: 10B <input type="text"/>	#Maps: 10 <input type="text"/> seconds: 30 <input type="text"/>	
Use comma-separated tails for exclusions, example: *.gif,.db*		

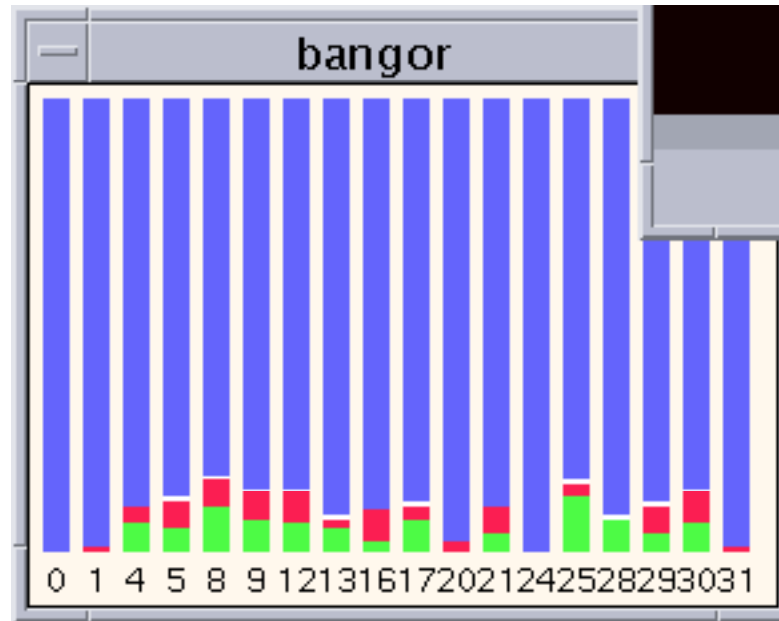
SANergy client
writes to 105GB
file, 12 threads
but without file
preallocation,
1GB extents,
CPU load on
metadata
controller during
999 MB/s stream



SANergy client
writes to 105GB
file, 12 threads
but without file
preallocation,
1GB extents,
CPU load on
client during 999
MB/s stream



SANergy client
writes to 105GB
file, 12 threads
with file
preallocation,
CPU load on
client during > 1
GB/second
stream



SANergy 2.2.0.17 CLIENT AND QFS 3.5.0.22 NATIVE COMBINATION TESTING USING VERY LARGE FILES

Host Name	Test Used	Block/ File Size	Threads	Sustained	Aggregate
SANergy client reads mcbob- base In e	jdv tto/s an on 12 T 3s (1 fi le per proc ess)	6M B/24 GB	R =12 (q= 24)	R - 106. 6 - 1108.66 MB/s	N/A(t ests alone)
		6M B/ 105G B	R =12 (q= 24) R =32 (q= 64)	R - 091. 6 - 1104.90 MB/S R - 112. 8 MB/s	N/A(t ests alone)
		3M B/ 105G B	R =12 (q= 24)	R - 072. 6 - 1106.29 MB/s	N/A(t ests alone)
SANergy client reads mcbob- base In e	same	6M B/ 105G B	R =32 (q= 64)	R - 103. 9 MB/s	N/A(t ests alone)
SANergy client reads mcbob and cabob	jdv tto/s an on 12 T 3s (1 fi le per proc ess)	6M B/ 105G B	R =32 (q= 64)	R =539.97 + 695. 3 R =492.11 + 617.34	1235.27 MB/s 1109.45 MB/s
SANergy client reads mcbob and cabob; QFS native writes on buddy	jdv tto/s an on 12 T 3s (1 fi le per proc ess)	6M B/ 105G B	W = 32 (q=64) R =12 (q= 24)	W = 408.56 MB/s R =392.46 MB/s R =468.32 MB/s	1269.34 MB/s
SANergy client reads -baseline	dv tto /san on 12 T 3s	6M B/ 105G B	R =6 (q=12)	c bob: R= 1108. 2 MB/s; c bob: R =1100.33 MB/s	N/A(t ests alone)
QFS baseline writes	dv tto /san on 12 T 3s	6M B/ 105G B	W = 12 (q=24)	buddy: W = 798. 9, 795. 9, 793.27 MB/s	N/A(t ests alone)
SANergy client reads mcbob and cabob	dv tto /san on 12 T 3s	6M B/ 105G B	R =6 (q=12)	R =520.55 + 520. 6 MB/s	1041.21 MB/s

SANergy 2.2.0.17 CLIENT AND QFS 3.5.0.22 NATIVE COMBINATION TESTING USING VERY LARGE FILES

Host Name	Test Used	Block/ File Size	Threads	Sustained	Aggregate
SANergy client realms on c_hoa and ca_hob; QFS native writes on buddy	dv_tto /san on 12_T3s	6M_B 105G_B	W= 2 (q=24) R=6 (q=12)	W= 40.48 MB/s R=331.11 + 332.6 MB/s	1113.75 M_Bs
QFS baseline realms from one file system, write to second (1 process)	jdvt from /ca_pure (6 x T3s) to /san (12 x T3)	6M_B 105G_B	R=32, W= 2 (q=64)	Sustained write rate to /san = 518.2 - 526.25 MB/s	N/A (tests alone)
SANergy client realms on c_hoa and ca_hob	dv_tto /san on 12_T3s	6M_B 105G_B	R=6 (q=12) (4 processes, 4 files)	R=216.34 + 217.9 + 216.4 + 216.88 MB/s	867.6 MB/s
SANergy client realms on c_hoa and ca_hob; QFS native writes on buddy	dv_tto /san (12x_T3s) = writes; 2 realms from each client	6M_B 105G_B	W= 2 (q=24) R=6 (q=12)	W= 279.59 MB/s R=184.48 + 183.2 + 182.9 + 183.32	1013.55 M_Bs

METADATA LAN LOADING EFFECT OF RATES

(SANergy 22.0.17 and QFS 3.5.0.22)

Host Name	Test Used	Block/ File Size	Threads	Sustained	Aggregate
SANergy client reads on c:\boa-baseline-unbade\met\da\p\ah	dv\to\san on 12 T3s (1 file)	6MB/ 105GB	R=6 (q=12)	R = 109.9 MB/s	N/A (tests alone)
SANergy client reads on c:\bob-base\in-e-unbade\met\da\p\ah	same	6MB/ 105GB	R=6 (q=12)	R = 110.3 MB/s	N/A (tests alone)
QFS baseline read from onefile system, write to seond (1 process)	jdvtf from /ca\pure (6 x T3s) to /san (12 x T3s)	6MB/ 105GB	R=32, W=32 (q=64)	Sustained = effective write rate to /san = 518.2 - 531.12 MB/s	N/A (tests alone)
SANergy client reads on c:\boa and \bob; QFS native writes on buddy from /ca\pure to /san-base\in-e-unload\ad\me\ad\da\p\ah	dv\tf from /san on 12 T3s (2 files); jdvtf from /ca\pure (6 x T3s) to /san (12 x T3s)	6MB/ 105GB	R=32, W=32 (q=64) R=6 (q=12)	W = 501.60 MB/s R = 330.88 + 331.47 MB/s	116395 MB/s
SANergy client read on c:\boa-heavyload\me\ad\da\p\ah	dv\to\san on 12 T3s	6MB/ 105GB	R=6 (q=12)	R = 785.24 MB/s	N/A (tests alone)
SANergy client reads on c:\boa and \bob; QFS native writes on buddy from /ca\pure to /san-base\in-e-heavyload\me\ad\da\p\ah	dv\tf from /san on 12 T3s (2 files); jdvtf from /ca\pure (6 x T3s) to /san (12 x T3s)	6MB/ 105GB	R=32, W=32 (q=64) R=6 (q=12)	W = 491.50 MB/s sustained R = 306.63 + 291.6 MB/s	108951 MB/s