

Fishworks

Brendan Gregg Cindi McGuire

Sun Microsystems



Fishworks is the name of an engineering team at Sun Microsystems

- FISH: "Fully Integrated Software and Hardware" a suitable acronym to describe our strategy
- Our goal to provide a unified management framework for appliances built on Solaris



Learn it. Live it. Make IT real.



Fishworks Overview

- Fully Integrated Software and Hardware
- Unified User Interface
- Turning Solaris into an appliance
- Example: NAS appliance





What Does it Take to Build an Appliance?

- Solid OS foundation
 - Key Solaris 10 building blocks:

SMF (Service Management Facility) FMA (Fault Management Architecture)

DTrace (Dynamic Tracing)

Networking

Security

- Common user interface
- Integrated higher-level management and configuration tasks with OS







Unified User Interface

- One User Interface to rule them all
 - BUI: Browser User Interface
 - CLI: Command Line Interface
- This is possible in the confines of an appliance
 - A special-purpose server confined to a limited set of configuration and management tasks





BUI: Browser User Interface

- Consistent look and feel
- As fast as possible
- Usability no special OS knowledge required
- Value add a real BUI (not a CLI wrapper)
 - Pie charts, traffic lights, plots, dialogs, navigation, ...
- Status updated live no need to refresh
- Not a(nother) skin speaks to akd, which speaks to OS
- Communication secured over HTTPS
- Extensive test framework
 - Required writing a JavaScript CLI





BUI Examples

Masthead:



Lists:

Filesystems LUNs 25 Total			াব বা 1-20 ♦♦ ▶। 🔍
NAME .	SIZE	MOUNTPOINT	
admin / accounts	18K	/export/accounts	
admin / exports	18K	/export/exports	
admin / primary	18K	/export/primary	
⊕ admin / traffic I	18K	/export/traffic	e ti
admin / workflow	18K	/export/workflow	





Dashboard:







CLI: Command Line Interface

- Mirror BUI functionality as much as possible
- Standard framework a tree of contexts
- Usability
 - Help for every context
 - Tab-completion ++
- Rich scripting environment
 - Stripped-down JavaScript
 - SSH keys can be added for automated scripts from a different host







CLI Example

vimba:> tree

```
+---> configuration
| |
+---> net
| | |
+---> datalinks
| |
+---> devices
| |
+---> interfaces
| |
+---> services
```

vimba:> configuration net interfaces select e1000gtab e1000g0 e1000g1 vimba:> configuration net interfaces select e1000g1 vimba:configuration net interfaces e1000g1> set v4dhcp=tab false true



Learn it. Live it. Make IT real.



CLI Scripting Example

% ssh root@vimba << EOF configuration net interfaces select e1000g1 show EOF

Properties:

```
<state> = up
        class = ip
        label = Untitled Interface
        admin = true
        links = nge0
dhcp_clientid =
        dhcp_hostname =
        dhcp_primary = false
        v4addrs = 192.168.2.124/22
        v4dhcp = true
        v6addrs =
        v6dhcp = false
```





Solaris Server Configuration

For example...

• NFS

/etc/default/nfs

/var/svc/log/network-nfs-server:default.log

• DNS

/etc/resolv.conf, /etc/nsswitch.conf

/var/svc/log/network-dns-client:default.log

Networking

ifconfig, dladm, netstat, route, routeadm
/etc/inet/hosts, /etc/inet/ipnodes, /etc/hostname.*
/var/adm/messages, /var/svc/log/*

Consider NIS, LDAP, FTP, Apache, iSCSI, etc...





Fishworks Server Configuration

Sun,			200	\mathcal{P}	Supe	r-User@vimbaLOG	OUT HELP
Storage 7110	Configuration	Maintena	ince	Shares	St	atus /	Analytics
0 53	SERVICES	STORAGE	NETWORK	CLUSTER	USERS	PREFERENCES	ALERTS
Services	• NES					Propertie	
Active Directory	Ch dh					Propertie	S LOgs
 CIFS 	◆ 7 O 2006-2-15 02:07:52 Online					REVERI	APPLT
DNS							
• FTP		Minimum supp	orted version	NESV3			
HTTP		Maximum supp	orted version				
Identity Mapping		waximum supp	oneu version	NF5V4 ¥			
IPMP		Maximum # of se	erver threads	500			
iscsi			Grace period	90 seconds			
LDAP	Use DNS d	lomain as NFSv4 ide	entity domain				
NDMP	Us	e custom NFSv4 ide	entity domain	domain			
NFS	100			1			
NIS							
Dhana Marra							
Phone Home Pouting							
SNMP							
SSH							
System Identity							
Virus Scan							





Fishworks Server Configuration

vimba:> configuration services nfs
vimba:configuration services nfs> show
Properties:

<status> = online
version_min = 3
version_max = 4
nfsd_servers = 500
grace_period = 90
mapid_dns = true
mapid_domain = domain

vimba:configuration services nfs> set grace_period=30
 grace_period = 30 (uncommitted)
vimba:configuration services nfs> commit
vimba:configuration services nfs> get grace_period
 grace_period = 30







Solaris Server Status

For example...

- Hardware
 - fmadm faulty
- Services

svcs (if the service is in SMF, otherwise application specific commands and log files must be used to determine service status)

• Consider older Solaris (and other OSes):

```
ps -ef, iostat -En, netstat -i
```

/var/adm/messages, /var/log/*







Sun.				90	Super-User@lo	ader LOGOU	T HELP
Storage 7110	Configuration	Mainten	ance	Shares	Status	Ana	alytics
€€ ()				HARDWARE	SYSTEM PI	ROBLEMS	LOGS
⊜ ● 0810QAS004 I	※ 69 (J)	🥥 Dis	ks 🛛 🥥 Slots	⊖CPU ⊖Mem	nory 🥥 Fans	💊 PSUs	⊖ SP
	TOP BACK FRONT	ID	MANUFACTURI	ER MODEL	SIZ	E TYPE	
		HDD 0	SEAGATE	ST914602SS	UN146G 137	G System	⊘∦
		😜 HDD 1	SEAGATE	ST914602SS	UN146G 137	G System	Ø ∦:
		🥥 HDD 2	SEAGATE	ST914602SS	UN146G 137	G Data	⊘∦
		🥥 HDD 3	SEAGATE	ST914602SS	UN146G 137	G Data	Ø 🔆
		🥥 HDD 4	SEAGATE	ST914602SS	UN146G 137	G Data	Ø 🔆
		🥥 HDD 5	SEAGATE	ST914602SS	UN146G 137	G Data	Ø 🔆
		HDD 6	SEAGATE	ST914602SS	UN146G 137	G Spare	*
		GHDD 7	SEAGATE	ST914602SS	UN146G 137	G Data	Ø ∦
		HDD 8	SEAGATE	ST914602SS	UN146G 137	G Data	⊘∦
		HDD 9	SEAGATE	ST914602SS	UN146G 137	G Spare	*
		HDD 10	SEAGATE	ST914602SS	UN146G 137	G Data	Ø ∦
		HDD 11	SEAGATE	ST914602SS	UN146G 137	G Data	Ø ∦
		HDD 12	SEAGATE	ST914602SS	UN146G 137	G Data	Ø ∦
		HDD 13	SEAGATE	ST914602SS	UN146G 137	G Data	Ø ∦
		🥥 HDD 14	SEAGATE	ST914602SS	UN146G 137	G Data	⊘∦
		HDD 15	SEAGATE	ST914602SS	UN146G 137	G Data	Ø ∦







tarpon:> maintenance hardware show

Chassis:

	NAME	STATE	MANUFACTURER MODEL
chassis-000	0839QCJ01A	ok	Sun Microsystems, Inc
cpu-000	CPU 0	ok	AMD Quad-Core AMD Op
cpu-001	CPU 1	ok	AMD Quad-Core AMD Op
cpu-002	CPU 2	ok	AMD Quad-Core AMD Op
cpu-003	CPU 3	ok	AMD Quad-Core AMD Op
disk-000	HDD 0	ok	STEC MACH8 IOPS
disk-001	HDD 1	ok	STEC MACH8 IOPS
disk-002	HDD 2	absent	
disk-003	HDD 3	absent	
disk-004	HDD 4	absent	
disk-005	HDD 5	absent	
disk-006	HDD 6	ok	HITACHI HTE5450SASUN500G
disk-007	HDD 7	ok	HITACHI HTE5450SASUN500G
fan-000	FT O	ok	unknown ASY, FAN, BOARD, H2



. . .



SUN, microsystems				Super-Oser@koi LOGOUT HEL
Storage 7410	Configuration	Maintenance	Shares S	Status Analytics
ر ال	NETWORK	SERVICES CLUSTER	USERS PREFEREN	CES ALERTS STORAG
	_			
Services				
Active Directory	Data Caminaa			
	Data Services		et 1	-
FTP	● NFS	Online	2008-10-8 04:45:31	
HTTP	iscsi	Online	2008-10-6 09:37:05)
Identity Mapping	CIFS	Disabled	2008-10-6 09:34:56 🖅 🕻)
IPMP 🐨	● FTP	Disabled	2008-10-6 09:34:59 🜮 🕻)
iSCSI	HTTP	Disabled	2008-10-6 09:34:57 5)
LDAP	NDMP	Online	2008-10-8 04:45:41 🗲 🕻)
NDMP	Virus Scan	Disabled	2008-10-6 09:35:05 💔 🕻	
NFS				
NIS	Directory Services			
NTP		Disabled	2008-10-8 04:45:20	2
Phone Home		Disabled	2008-10-8 04:45:20)
Routing		Disabled	2008-10-8 04:45:25 49	
Service Tags	Active Directory	Online	2008-10-8 09:35:08 49 0	1
	 Identity Mapping 	Unime	2008-10-8 04:45:31 🗸 😋	/
System Identity				
Virus Scan	System Settings			_
	🔍 DNS	Online	2008-10-8 04:45:42 🚺 🖯)
		Online	2008-10-8 04:45:28 🗲 🖯)
	NTP	Disabled	2008-10-6 09:34:55 🛷 😃)
	Routing	Online	2008-10-6 10:25:14 🗲 🖯)
	Phone Home	Disabled	2008-10-9 11:57:58)
	SNMP	Disabled	2008-10-6 09:34:57 5)
	Service Tags	Online	2008-10-8 04:45:41 🗲 🖯)
	System Identity	Online	2008-10-8 04:45:30 🚯 🕻)
	-,,			
	Remote Access			
	Nelliole Access		C	





vimba:> configuration services show
Services:

ad => disabled cifs => disabled dns => online ftp => disabled http => disabled identity => online idmap => online ipmp => online iscsi => online ldap => disabled ndmp => online nfs => online nis => disabled ntp => disabled

&





Solaris Server Performance Observability For example...

• CPU

vmstat, mpstat, prstat, dtrace

• Memory

vmstat, prstat

• Disk I/O

iostat, dtrace

• Network I/O

netstat, dladm, nicstat, nx.se, dtrace

• NFS

nfsstat, dtrace





Fishworks Server Performance Observability







Fishworks Server Performance Observability

Ok, that's a bit hard to do in the CLI. This is one of the few differences between BUI and CLI functionality.

But while the graphs aren't available, the data is:

vimba:> status	activi	ity show	
Activity:			
CPU	10	%util	Sunny
Disk	2	ops/sec	Sunny
iscsi	0	ops/sec	Sunny
NDMP	0	bytes/sec	Sunny
NFSv3	0	ops/sec	Sunny
NFSv4	0	ops/sec	Sunny
Network	3K	bytes/sec	Sunny
CIFS	0	ops/sec	Sunny

And individual statistics (datasets) ...





Fishworks Server Performance Observability

vimba:> analytics datasets
vimba:analytics datasets> show
Datasets:

DATASET	STATE	INCORE	ONDISK	NAME	
dataset-000	active	893K	342K	arc.acces	sses[hit/miss]
dataset-001	active	270K	83.1K	cpu.utili	ization
dataset-002	active	748K	280K	cpu.utili	<pre>ization[mode]</pre>
હ					
vimba: analy	tics data	asets> s	select d	lataset-00)6 read 5
DATE/TIME			≷UTIL	%UTIL	BREAKDOWN
2006-2-15 1	5:56:55		7	6	kernel
				1	user
2006-2-15 1	5:56:56		7	6	kernel
				1	user
2006-2-15 1	5:56:57		29	17	user
				12	kernel







Missing Piece

That looks great but how do we link our new Unified User Interfaces with the core OS services in Solaris?





Learn it. Live it. Make IT real.



Fishworks Unified Management

- Appliance Kit Daemon (akd)
 - Not a(nother) wrapper around the Solaris CLIs
 - Tightly integrated with the Solaris OS libraries to provide appliance abstractions for:
 - Storage: ZFS, NDMP
 - Protocols: iSCSI, NFS, CIFS, HTTP, FTP, WebDAV
 - Networking: ifconfig, routing, IPMP
 - Security: OpenSSL, ssh
 - RAS: fmd, libtopo, IPMI, SMBIOS, SNMP
 - Service management: SMF
 - Observation: DTrace, kstats





Fishworks Unified Management

- Additional features added to support appliance-specific tasks
- Clustering
- Software upgrade/rollback
- Integrated phone home, service tag, and audit capabilities
- Roles and authorizations
- Secure communication channel for BUI and CLI
- Customers interact with the BUI or CLI, akd interacts with Solaris







Putting it All Together



Common BUI, CLI, and test framework to drive management software: JavaScript

Standard protocol for communication: XML-RPC

Common control point (akd) to OS libraries

Enhance OS to leverage appliance hardware: clustering and ZFS L2ARC

Hardware supported by FMA





SMF: Service Management Facility

- Service abstraction for a running application, device state or set of other services
- SMF(5) provides a common infrastructure for service:
 - Configuration
 - Fault monitoring
 - Restart
 - Observability
- All appliance applications and facilities run under the SMF







FMA: Fault Management Architecture

- Appliance software and hardware errors reported to fmd(1M)
 - CPU/Memory, PCI-Express, HBA controllers, fans, power supplies, and disks
 - Appliance kit software instrumented for FMA
- Faults and defects reported using the Sun Fault Messaging Standard with problem resolution at http://www.sun.com/msg
- Guided FRU replacement made possible by FMA topology libraries
- IPMI, SMART, and other sensor data collected and reported to fmd(1M)
- Configurable SNMP traps and alerts





DTrace

- Analytics uses DTrace (and Kstat) to visualize statistics in real-time
 - Not just bolting on a GUI, but rethinking how to visualize performance – and investigating what new features GUIs make possible
- Statistics can be archived and saved forever
 - Investigate performance issues after the event
- Analytics can answer high level questions:

"What clients are making NFS requests?"

"What CIFS files are being accessed?"

"How long are disk operations taking?"





DTrace: Analytics

Demonstrating how GUIs can add value





Learn it. Live it. Make IT real.



A Word about the Solaris Shell

• The appliance is entirely manageable from the BUI and CLI: no Solaris shell access required. For example:

 $if config \rightarrow \texttt{buri:} > \texttt{configuration net}$

route \rightarrow **buri:**> configuration services routing

ping/nslookup (builtins)

```
buri:> ping kipper
buri:> nslookup 192.168.2.104
```

- akd manages resources such as ZFS, use of the original zpool/zfs commands can easily create issues that are extremely difficult to troubleshoot
- The Solaris shell is available for trained Sun Service staff to use only if absolutely necessary.







Example: NAS appliance

- Features from Solaris 10:
 - Enterprise-class scalability, RAS, and performance
 - IPv4 and IPv6 networking, LACP, IPMP, VLANs, ...
 - NFSv3, v4, FTP, HTTP, WebDAV, iSCSI, and now CIFS
 - Scalability of all key subsystems to 64 cores and beyond
 - Unique innovations: ZFS, DTrace, FMA, SMF, ...
- Features added/enhanced for this appliance:
 - ZFS: L2ARC, log devices, RAID-Z DP
 - Integration with Solaris CIFS and Windows Identities
 - Clustering

...



Example: NAS appliance

A tour of the interface and features





Learn it. Live it. Make IT real.





Thank you

CEC2008

Learn it. Live it. Make || real.

November 9-14, 2008 Las Vegas, NV