

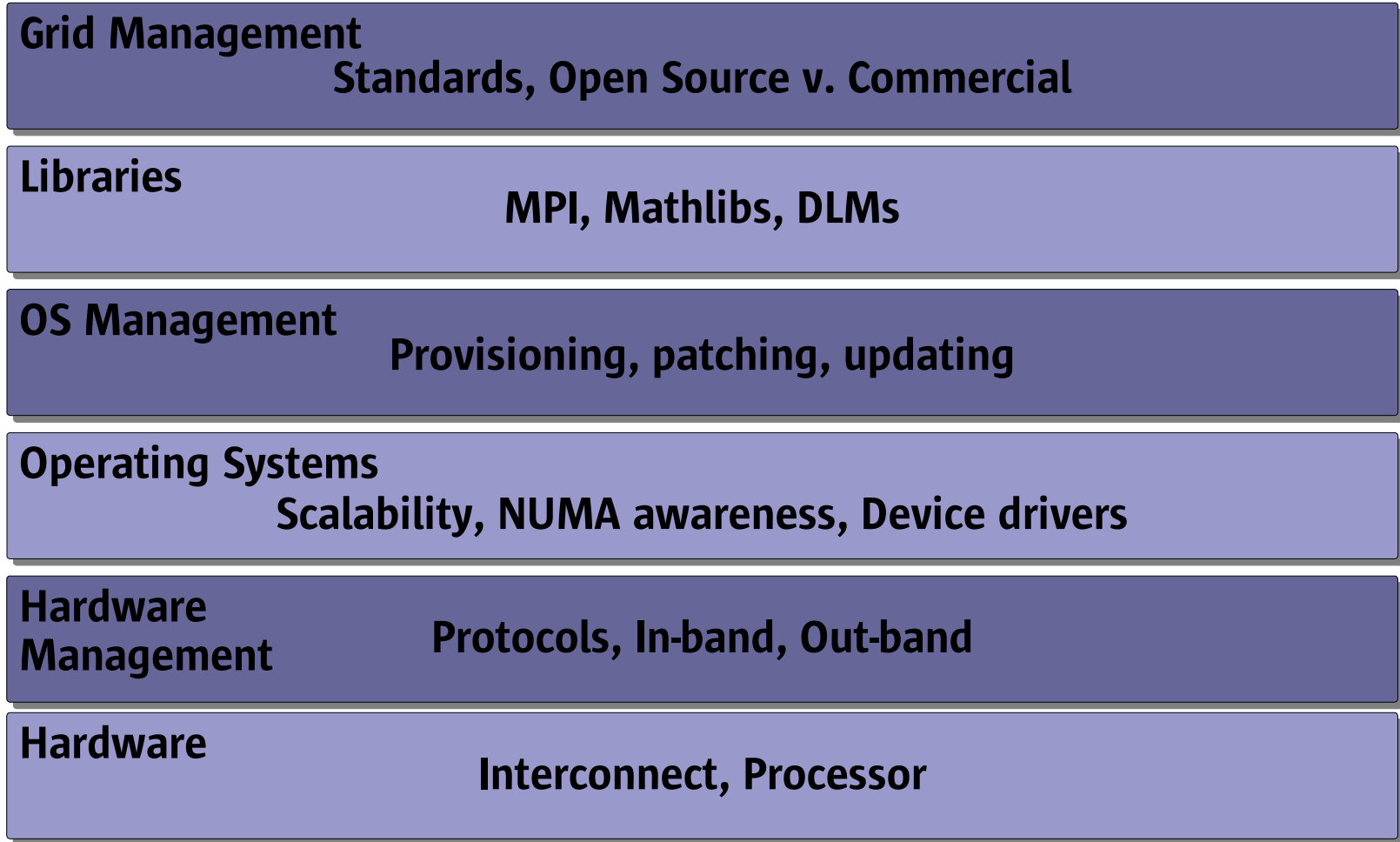
Cluster Lifecycle Management

Carlo Nardone
Technical Systems Ambassador
GSO – Client Solutions



The Stack

Applications



Cluster Management (1/2)

- Need for scalable and reproducible tools
 - often overlooked!
- Initial set-up (provisioning) is tightly linked with on-going management
 - HW faults, software patches and updates, new HW ...
- Effective Open Source tools from the Beowulf community
 - Rocks vs Oscar
- Proprietary tools specialized for vendor HW
 - Sun Control Station 2.x
 - Sun Control Station 3.x (Project HotDog)

Cluster Management (2/2)

- There are NO homogeneous clusters
 - role differences
 - (sometimes unexpected) small low-level differences
 - standard image-based approaches are ineffective
- System software consistency approaches
 - Diskless architecture: scalability issues
 - Disk imaging: OSCAR (and many others)
 - Description-based: Sun Jumpstart first, then RedHat Kickstart, SuSE YaST, NPACI Rocks
- Monitoring, logging, config mgmt
 - Ganglia monitoring (ganglia.sourceforge.net)
- Account, workload, upgrades mgmt
 - Workload mgmt: OpenPBS, Condor, GridEngine

SDSC ROCKstar: #201 in < 2 h!



SDSC “ROCKstar”: A 128 node/256 CPU x86 cluster, #201 in Top 500
Built live on the SC2003 show floor in 100 minutes, adding a new
server every 30 seconds. 200 jobs were submitted to the queue
and the cluster began to execute 15 minutes into the build cycle.

Rocks and OSCAR

- NPACI Rocks
 - Uses descriptions to build “appliances”
 - Description (configs and RPM packages) organised in graphs and stored in MySQL DB
 - RedHat-based (depends on Anaconda and Kickstart)
 - Very fast start-up of bare metal HW
 - Configurable descriptions (XML), HW independent
 - www.rocksclusters.org
 - Commercial support from Scalable Systems Ltd (Singapore)

Rocks and OSCAR

- NPACI Rocks
 - Uses descriptions to build “appliances”
 - Description (configs and RPM packages) organised in graphs and stored in MySQL DB
 - RedHat-based (depends on Anaconda and Kickstart)
 - Very fast start-up of bare metal HW
 - Configurable descriptions (XML), HW independent
 - www.rocksclusters.org
 - Commercial support from Scalable Systems Ltd (Singapore)
- OSCAR (Open Source Cluster Application Resource)
 - Uses disk images
 - Support multiple distro
 - Config wizard to create client images
 - Support cluster nodes with no installed OS
 - oscar.openclustergroup.org

Sun Control Station 2.x (1/4)

- Lifecycle model
 - Install, initialize, provision
 - Maintenance: monitor, configure, patches
 - Issue resolution: back-up, recovery, restore
 - Migration, remove, reallocate
- Modular software architecture
 - Ability to write your own control module (e.g. GridEngine console)
 - Integrate in-house scripts in SCS framework
- Support for several flavours of RedHat and SuSE, Solaris 9 (SPARC and x86)
- (v2.2) Imminent support for AMD HW

Sun Control Station 2.x (2/4)

- Major capabilities
 - Software management
 - Image management/deployment (no Jumpstart support)
 - Inventory management
 - Health and Performance Monitoring
 - Lights Out Management (LOM), IPMI-based (RedHat only)
 - Cluster Grid management
- Core capabilities
 - Group operations (manage volume servers in groups)
 - Hierarchical architecture
 - Simple, clean browser-based UI
 - Secure communications - 128 bit encryption btw servers
 - Extensible and flexible platform, with developm. guide
 - Interface with system management products

Sun Control Station 2.x (3/4)

- Software Management
 - Package management
 - Volume software push, patching, analysis and updating to multiple servers
 - Software pushing to push an update onto a server
 - Leveraging capabilities in Sun CNS Org
- Remote image deployment
 - Manage a defined set of packages or system images for rapid setup and provisioning of servers
 - Deploy on multiple systems simultaneously and quickly
 - Industry standard methodologies
 - User interface to Kickstart, AutoYast and Jumpstart
 - Rapid and simple industry standard functionality
 - Used in conjunction with LOM

Sun Control Station 2.x (4/4)

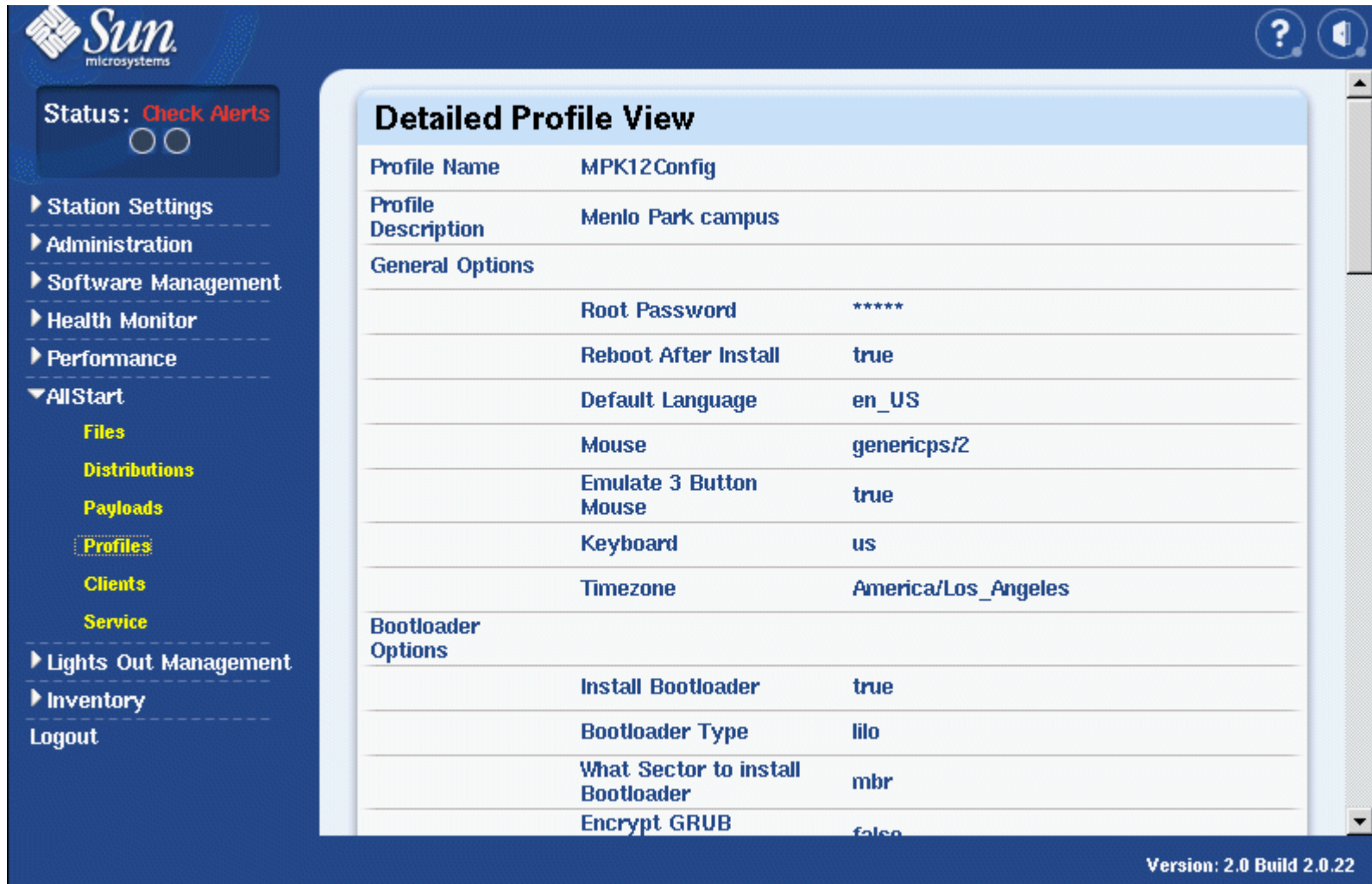
- Inventory information
 - Managed server details, OS, Hardware profiles
 - Health and Performance Monitoring
 - Snapshot of system characteristics
 - Memory, swap space, CPU utilization
 - Lights Out Management
 - Uses Intel's IPMI interface (IPMItool and OpenIPMI)
 - Reset, Power On/Off, CPU, Fan speeds, Temperature
- Grid engine management
 - Uses N1GE 6.0 to launch grid capabilities from SCS user interface
 - Modular addition to the SCS framework

Screenshot: Administration

The screenshot displays the Sun microsystems Administration web interface. On the left is a navigation menu with options like Station Settings, Administration (with sub-items: Hosts, Modules, Groups, Tasks, Schedule), Software Management, Health Monitor, Performance, AllStart, Lights Out Management, Inventory, and Logout. A status box at the top left shows 'Status: Check Alerts' with a red indicator. The main content area is titled 'Managed Hosts' and shows a table of 13 items. The table has columns for Name, Type, Description, and Email. Below the table are buttons for 'Modules', 'Group', 'Modify', 'Remove', and 'Add'. The version '2.0 Build 2.0.20' is shown at the bottom right.

Name	Type	Description	Email
192.18.87.24	LX50		
10.6.75.170	Netra T1120-1125 [S9]		
10.6.75.160	Netra X1 [S8]		
10.6.73.43	Qube 3		
10.6.73.47	Qube 3		
10.6.73.49	RaQ 550		
10.6.73.31	RaQ XTR		
10.6.73.30	RaQ XTR		

Screenshot: Allstart



The screenshot shows the Sun Allstart web interface. On the left is a navigation menu with the Sun Microsystems logo at the top. Below the logo, there is a 'Status: Check Alerts' section with two circular indicators. The navigation menu includes: Station Settings, Administration, Software Management, Health Monitor, Performance, AllStart (expanded), Files, Distributions, Payloads, Profiles (highlighted with a dashed border), Clients, Service, Lights Out Management, Inventory, and Logout. The main content area is titled 'Detailed Profile View' and displays configuration details for a profile named 'MPK12Config' at the 'Menlo Park campus'. The configuration is organized into sections: General Options and Bootloader Options. The version number 'Version: 2.0 Build 2.0.22' is visible in the bottom right corner.

Detailed Profile View	
Profile Name	MPK12Config
Profile Description	Menlo Park campus
General Options	
Root Password	*****
Reboot After Install	true
Default Language	en_US
Mouse	genericps/2
Emulate 3 Button Mouse	true
Keyboard	us
Timezone	America/Los_Angeles
Bootloader Options	
Install Bootloader	true
Bootloader Type	lilo
What Sector to install Bootloader	mbr
Encrypt GRUB	false

Screenshot: LOM

The screenshot displays the Sun Lights Out Management (LOM) interface. On the left is a navigation sidebar with the Sun logo and a 'Status: Check Alerts' indicator. The main content area is titled 'Lights Out Management Power Menu' and shows a table of system configurations. One item, '10.6.75.125 x86 Red Hat EL AS 2.1', is selected. Below the table are buttons for 'Power On', 'Power Off', 'Reset', and 'Identify'. The version '2.0 Build 2.0.22' is shown at the bottom right.

Status: Check Alerts

- Station Settings
- Administration
- Software Management
- Health Monitor
- Performance
- AllStart
- Lights Out Management
 - Power**
 - Sensors/SEL
- Grid Engine
- Inventory
- Logout

Lights Out Management Power Menu

Select all Deselect all 1 of 4 items selected

Name	Type	Description	Lom Configured
Linux x86			
Raq XTR			
10.6.75.125	x86 Red Hat EL AS 2.1		

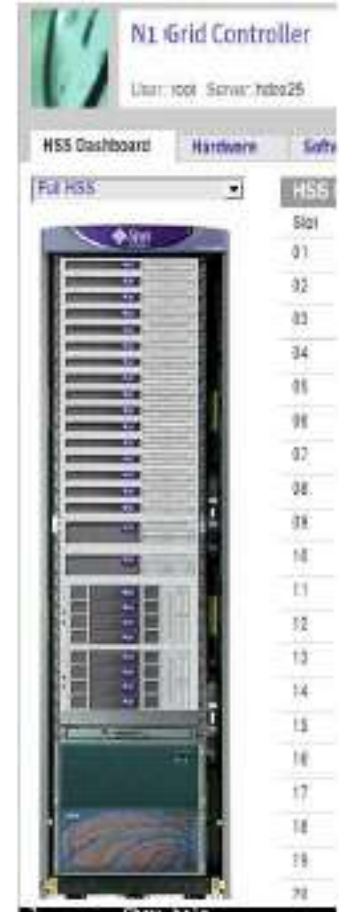
Power On Power Off Reset Identify

Version: 2.0 Build 2.0.22

Sun Control Station 3.x

- Project HotDog
 - NAS/SAN based bootup and deployment
 - New user interface
 - Enhanced software management and analysis
 - Support for Galaxy, Andromeda systems
 - Expansion packs to integrate N1SPS
 - Network Discovery & OS Boot
 - Completely scriptable CLI

- Support for
 - Sun Fire V20z and V40z
 - Upcoming Sun Opteron-based systems
 - IP, NAS and FC/SAN connection
 - InfiniBand bridges (IP, SDP, uDAPL, RDMA protocols)



Thanks!

carlo.nardone@sun.com

cell: 335 5828197

The Network is the Computer

