



Sun Fire™ 15K/12K Server Preferred Practices

Lee Lustig, Sun Professional Services

Sun BluePrints™ OnLine—July 2003



<http://www.sun.com/blueprints>

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95045 U.S.A.
650 960-1300

Part No. 817-3140-10
Revision 1.0, 7/7/03
Edition: July 2003

Copyright 2003 Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, California 95045 U.S.A. All rights reserved.

Sun Microsystems, Inc. has intellectual property rights relating to technology described in this document. In particular, and without limitation, these intellectual property rights may include one or more patents or pending patent applications in the U.S. or other countries.

This product or document is protected by copyright and distributed under licenses restricting its use, copying, distribution, and decompilation. No part of this product or document may be reproduced in any form by any means without prior written authorization of Sun and its licensors, if any. Third-party software, including font technology, is copyrighted and licensed from Sun suppliers.

Parts of the product may be derived from Berkeley BSD systems, licensed from the University of California. UNIX is a registered trademark in the United States and other countries, exclusively licensed through X/Open Company, Ltd.

Sun, Sun Microsystems, the Sun logo, Sun BluePrints, JumpStart, Java, JVM, OpenBoot, Solaris, Solstice DiskSuite, Sun Cluster, Sun Fire, Sun Enterprise, and Sun StorEdge are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the US and other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc.

The OPEN LOOK and Sun™ Graphical User Interface was developed by Sun Microsystems, Inc. for its users and licensees. Sun acknowledges the pioneering efforts of Xerox in researching and developing the concept of visual or graphical user interfaces for the computer industry. Sun holds a non-exclusive license from Xerox to the Xerox Graphical User Interface, which license also covers Sun's licensees who implement OPEN LOOK GUIs and otherwise comply with Sun's written license agreements.

U.S. Government Rights—Commercial use. Government users are subject to the Sun Microsystems, Inc. standard license agreement and applicable provisions of the FAR and its supplements.

DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

Copyright 2003 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, Californie 95045 Etats-Unis. Tous droits réservés.

Ce produit ou document est protégé par un copyright et distribué avec des licences qui en restreignent l'utilisation, la copie, la distribution, et la décompilation. Aucune partie de ce produit ou document ne peut être reproduite sous aucune forme, par quelque moyen que ce soit, sans l'autorisation préalable et écrite de Sun et de ses bailleurs de licence, s'il y en a. Le logiciel détenu par des tiers, et qui comprend la technologie relative aux polices de caractères, est protégé par un copyright et licencié par des fournisseurs de Sun.

Des parties de ce produit pourront être dérivées des systèmes Berkeley BSD licenciés par l'Université de Californie. UNIX est une marque enregistrée aux Etats-Unis et dans d'autres pays et licenciée exclusivement par X/Open Company Ltd.

Sun, Sun Microsystems, le logo Sun, Sun BluePrints, JumpStart, Java, JVM, OpenBoot, Solaris, Solstice DiskSuite, Sun Cluster, Sun Fire, Sun Enterprise, et Sun StorEdge sont des marques de fabrique ou des marques déposées, ou marques de service, de Sun Microsystems, Inc. aux Etats-Unis et dans d'autres pays. Toutes les marques SPARC sont utilisées sous licence et sont des marques de fabrique ou des marques déposées de SPARC International, Inc. aux Etats-Unis et dans d'autres pays. Les produits portant les marques SPARC sont basés sur une architecture développée par Sun Microsystems, Inc.

L'interface d'utilisation graphique OPEN LOOK et Sun™ a été développée par Sun Microsystems, Inc. pour ses utilisateurs et licenciés. Sun reconnaît les efforts de pionniers de Xerox pour la recherche et le développement du concept des interfaces d'utilisation visuelle ou graphique pour l'industrie de l'informatique. Sun détient une licence non exclusive de Xerox sur l'interface d'utilisation graphique Xerox, cette licence couvrant également les licenciés de Sun qui mettent en place l'interface d'utilisation graphique OPEN LOOK et qui en outre se conforment aux licences écrites de Sun.

CETTE PUBLICATION EST FOURNIE "EN L'ETAT" ET AUCUNE GARANTIE, EXPRESSE OU IMPLICITE, N'EST ACCORDEE, Y COMPRIS DES GARANTIES CONCERNANT LA VALEUR MARCHANDE, L'APTITUDE DE LA PUBLICATION A REPENDRE A UNE UTILISATION PARTICULIERE, OU LE FAIT QU'ELLE NE SOIT PAS CONTREFAISANTE DE PRODUIT DE TIERS. CE DENI DE GARANTIE NE S'APPLIQUERAIT PAS, DANS LA MESURE OU IL SERAIT TENU JURIDIQUEMENT NUL ET NON AVENU.



Please
Recycle



Adobe PostScript

Sun Fire™ 15K/12K Server Preferred Practices

In many ways, the Sun Fire™ 15K/12K server is a microcosm of today's computing world. It encompasses many of the most diverse and powerful technologies available today. In addition to being the highest capacity and highest performance Sun Fire servers currently released, these servers have the ability to run many diverse, mission-critical applications in a highly consolidated configuration. To successfully manage and maintain this environment, it is important to follow well established and tested practices and guidelines. Following these guidelines and practices can improve the success of an implementation.

Many documents about configuring Sun Fire servers have been written at Sun Microsystems. The information in those documents is consolidated to derive a set of preferred practices you can quickly reference while planning an implementation. This article addresses preferred practices at a high level, referencing supporting documentation when a more in-depth technical discussion is warranted.

This article categorizes preferred practices in the following sections:

- “Recommendations for Applying Preferred Practices” on page 2
- “Principals of Mission-Critical Implementations” on page 2
- “Physical Environment” on page 3
- “Internal Network Planning” on page 6
- “External Network Planning” on page 10
- “System Controller Configuration” on page 12
- “Platform and Domain Administration” on page 16
- “Security” on page 19
- “Error Analysis and Diagnosis” on page 21
- “Platform and Domain Configuration” on page 22
- “Dynamic Reconfiguration” on page 27

Recommendations for Applying Preferred Practices

Use the preferred practices and guidelines presented in this article whenever Sun Fire 15K/12K servers are initially configured, or when additions and changes are being made. Examples of situations when you should use these practices include creating domains, adding resources (physically or with dynamic reconfiguration), adding applications to existing domains, setting up new platforms, and updating software and hardware.

Do not lose sight of the fact that the technical implementation is only a small part of the overall success of a project that utilizes Sun Fire 15K/12K servers. Establishing a process that addresses all phases of the project, and using that process consistently, is very important. This process should include phases such as establishing business requirements, application selection criteria, project planning, business continuity, documentation, implementation procedures, testing procedures, and sustaining services. Not all successful implementations hinge on technical expertise; adhering to processes and procedures is just as important.

Note – Many of the preferred practices and guidelines that are presented in this article apply to the majority of cases. Not all recommendations apply to every circumstance. As Sun Microsystems Inc. releases new technologies and updates, these preferred practices might become outdated. Therefore, keeping up-to-date on the latest software and hardware releases and documentation is essential.

Principals of Mission-Critical Implementations

As a general rule, Sun Fire 15K/12K servers are designed to run large, mission-critical applications that require very high levels of availability, recoverability, serviceability, and manageability. Keep in mind that even the most advanced and well tested technologies can and will fail, whether the cause is from software, hardware, or operational procedures.

Consider the following basic design principals when using Sun Fire 15K/12K servers to achieve a successful implementation.

- Configure these servers such that there is no single point of failure, and implement simple and fast recovery procedures in case built-in redundancy fails. For example, design components with redundancy, and design procedures that support a simple and fast recovery in the event that redundancy and failovers do not work.
- Design the platform infrastructure with the mind-set that maintenance windows might be non-existent or very minimal. Therefore, design the platform so maintenance and servicing can be done when systems are online.
- Put in place low-level infrastructure and services that have very high availability (for example, 99.99 percent). This includes local area networks (LANs), wide area networks (WANs), naming and directory services, network file services, power, cooling, and storage area networks (SANs).
- Select standard solutions for backup, recovery, storage, SAN, and network. Avoid having too many one-off cases for applications when building the platform infrastructure. Minimize the number of builds, platforms, versions, and tools used.
- Choose technologies that are well established in the industry and well supported by their vendors. Avoid in-house customized solutions that must be maintained internally by the system administrators and IT staff. Avoid bleeding edge technologies that do not have a track record.
- Choose technologies and designs that are well understood and manageable by your IT staff. Design architectures that are simple, and reduce complexity whenever possible.
- Invest in areas that have a big effect on availability and quality of service. For example, invest in operation management, testing procedures, on-site spares, high-end service contracts, IT staff training, change management, monitoring tools, and performance tools.
- Implement security procedures and utilize security hardening toolkits to protect the servers' network access points and services. Test security around the Sun Fire platforms with intrusion detection software and audits. Use established products and procedures for implementing security.
- Develop, maintain, and test online operational procedures, and implement monitoring software from well established vendors.

Physical Environment

Although the physical environment is well documented in the Sun Fire 15K/12K server's product documentation, it is worth reviewing some important points regarding this subject.

The data center floor must be designed and laid out to meet the physical dimensions and weight of the Sun Fire 15K/12K server frame. It is important to fully understand all the requirements documented in the *Sun Fire 15K Site Planning Guide* and to have Sun Support Services review the proposed site to make sure it meets the requirements documented in the planning guide before the physical installation is scheduled.

The first task when implementing the physical set up of a Sun Fire 15K/12K server is to layout a diagram and a set of specifications. This diagram should include the newly purchased components of the architecture and their physical requirements (such as access routes, cooling, power, grounding, floor strength, space, and cabling). Be sure to include not only the Sun Fire 15K/12K server frames, but also the racks that support the servers used for boot arrays. The next task is to start the installation with Sun Services. This task should be scheduled after the physical environment has been fully installed and tested.

Planning for Cabling

When planning for cabling, account for all components and add extra cables to account for redundancy in case of failure. This will alleviate some of the headaches that might occur later if there are cable failures or when expansion of the system components is required. Cabling should include all network, communication, power, and I/O for all racks. All cables should be labeled clearly and installed with a cable binder to keep them well organized. Cables should be installed with the adequate amount of slack to prevent them from being damaged, especially when installing fiber cables. Cables should be placed in positions that allow ample room and easy access for removing and servicing internal components of the platform.

In addition, decide on a naming standard for labeling all cables early in the implementation planning phase of the project. A detailed cabling worksheet or spreadsheet is a good document to develop to help in the design and planning of this activity. The document should include the type (fiber or cooper), connection type (FC or SC), quantities, and lengths of the cables and what racks they connect to. There might be restrictions placed on the type of cabling used for certain types of components, so follow specific product documentation and specifications closely.

Ensure that adequate clearance is maintained between the Sun Fire 15K/12K frames and other equipment. Reference the *Sun Fire 15K Site Planning Guide* for these specifications. Make sure there is adequate room for servicing, when needed. Remember, it is not recommended or supported, that you install or place any non Sun Fire 15K components in the Sun Fire 15K cabinets.

Planning for Cooling

As with other physical components in a Sun Fire 15K/12K frame, the cooling system is monitored and adjusted by system controllers. If system controllers are offline or if fans have failed, the remaining fans will be run at maximum speed to make sure they compensate for the failure. Regardless of the ability of the server to compensate for airflow and fan redundancy, you must plan for and maintain adequate facility resources. If the cooling environment is not maintained, the server could shut down automatically. The cooling required for each frame depends on the internal configuration of components within a frame. Because the internal server configuration affects the power consumption and the heat produced for each frame, it is important that you carefully calculate the required cooling using the spreadsheet within the *Sun Fire 15K Site Planning Guide*. To satisfy all types of conditions, install cooling capacity for the maximum anticipated configuration so that upgrades of the cooling system will not be required as internal components of the server or additional servers are added later.

We recommend that you install Sun Fire 15K/12K frames on a raised-floor configuration with perforated tiles. The Sun Fire 15K/12K uses air flow from the bottom of the frame to the top of the frame, internally, and the raised floor and perforated tiles can generate the proper air flow. For more information about the environmental specifications for air flow and temperature, refer to the *Sun Fire 15K Site Planning Guide*.

In addition, fill any open slots on the Sun Fire 15K frame with filler panels to ensure proper air flow and cooling. The design of the Sun Fire 15K allows you to replace components such as air filters without any service interruption, and these should be checked and replaced regularly.

Configuring Power Supplies

You can configure Sun Fire 15K/12K servers with up to six hot-swappable single phase AC power supplies, and twelve independent power cords. We recommend that you install and test all six power supplies and twelve power cords before placing a server into service. This will alleviate extra maintenance and testing later if the server is expanded and requires the extra power.

Because Sun Fire 15K/12K servers support a dual-power grid design and n+1 power supplies, 100 percent redundancy can always be achieved. Therefore, in a fully loaded frame, these servers can operate with full functionality with less than the maximum of six power supplies. Nonetheless, we recommend that you turn on all power supplies to utilize the extra redundancy. There is a power planning worksheet in the *Sun Fire 15K Site Planning Guide* to help you calculate the specific power requirements for individual installations. For power configuration drawings of both

single- and dual-redundant requirements, as well as voltage and circuit requirements, reference the *Sun Fire 15K System Site Planning Guide*, Chapter 3 “Facility and System Requirements.”

Within each AC power supply, there are redundant and equal halves for connecting two AC input cords. To alleviate the possibility of incurring repeated power interruptions and fluctuations, which, in turn, could cause a higher rate of component failures, ensure that the power feeds being used come from separate dedicated distribution panels with a constant and stable UPS source.

Installing Sun Fire 15K/12K servers requires additional mirrored domain boot devices such as the Sun StorEdge™ S1 array. These arrays are typically located in separate 72-inch racks, which also require dual redundant circuits. The power requirements for racks should be documented in the rack manufacturer’s data specs.

All Sun Fire 15K/12K servers are properly grounded through the power cables and additional grounding cables are not provided with the system. Although not required, adding extra grounding might improve the ability to dissipate current more efficiently. It is important to check that the data center's grounding rails or grounding potential is the same as the grounding potential of the server's power cables. For more information about these procedures, refer to the *Sun Fire 15K Installation and De-Installation Guide* and the *Sun Fire 15K Site Planning Guide*.

Internal Network Planning

The Sun Fire 15K/12K servers have two internal networks within the frame of each server. These are called the management network (MAN) I1 and I2 networks. The MAN I1 provides private point-to-point network connections between the system controller and each domain. The MAN I2 network provides the point-to-point connections between the primary and secondary system controllers. The overview of the Sun Fire 15K/12K management networks are shown in FIGURE 1 on page 7. The main thing to remember about these networks is that they are dedicated and serve specific functions for internal communications. They are not general purpose networks, therefore, no external packets should be routed across them.

The connections that make up the MAN I1, as built from the factory, are designed to provide maximum availability, redundancy, and security for system controllers to accomplish domain management. This environment has redundant system controllers with dedicated separate point-to-point connections to each system board on the frame. If there is a problem with any of the components that make up the MAN I1 network, the network will automatically switch to the redundant system without any service interruption to domains running applications.

The Sun Fire 15K/12K servers are designed with twenty internal network connections that make up the MAN I1 and I2 networks, as shown in FIGURE 2 on page 8. Eighteen of these connections are connected between system controllers and domain expander boards, and two are dedicated for connections between the system controllers. For a detailed description of the system controller’s internal network configurations, see the *SMS Administrators Guide*, Chapter 6 “Domain Services.”

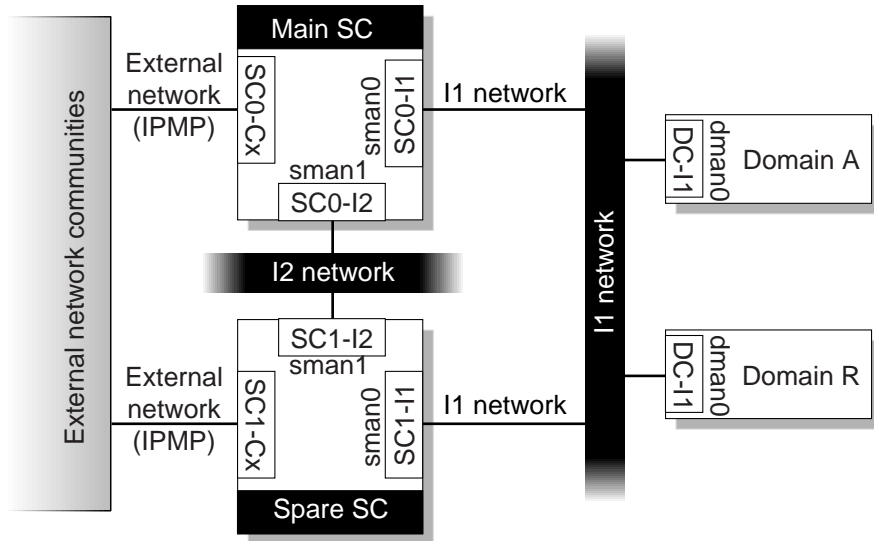


FIGURE 1 Management Network Overview

I1 Network

The MAN I1 network is a private network, not a general purpose network. The network provides functions such as domain consoles, message logging, dynamic reconfiguration, booting from the network, and network time synchronization (NTP). Internal to the dman processes (management network driver) is the ability of the I1 network to detect failures and provide path switch-over in the event of a failure. No packets addressed to one domain can be routed along the network connection between the system controller and another domain. Access to MAN is restricted to the system controller and the domains, and this configuration must not be changed. MAN software also enforces domain isolation of network traffic on the I1 network. Similar software operates on the domain side.

The MAN I1 network is designed to be a completely separate and dedicated point-to-point network for communicating between system controllers and domains on one Sun Fire 15K/12K frame. The I1 network is connected in a point-to-point fashion to network interfaces located on each of the 18 expander I/O slots. The I1 is a 100

megabit, half-duplex configuration that uses internal hubs to connect to each I/O board. Using this design, the number of point-to-point Ethernet links between a system controller and a given domain is based on the number of I/O boards configured in that domain. Each NIC from the system controller connects to a NIC on the I/O board through the hub. The NIC is an internal part of the I/O board, and is not a separate adapter card.

Access to MAN is restricted to the system controller and the domains, and this configuration must not be changed. MAN software also enforces domain isolation of network traffic on the I1 network. Similar software operates on the domain side. For added security purposes the address resolution protocol (ARP) can be disabled on this network to remove the ability of ARP spoofing attacks and other IP attacks on this network. The procedure is well documented in the Sun BluePrints OnLine article “Securing the Sun Fire 15K/12K Domains.”

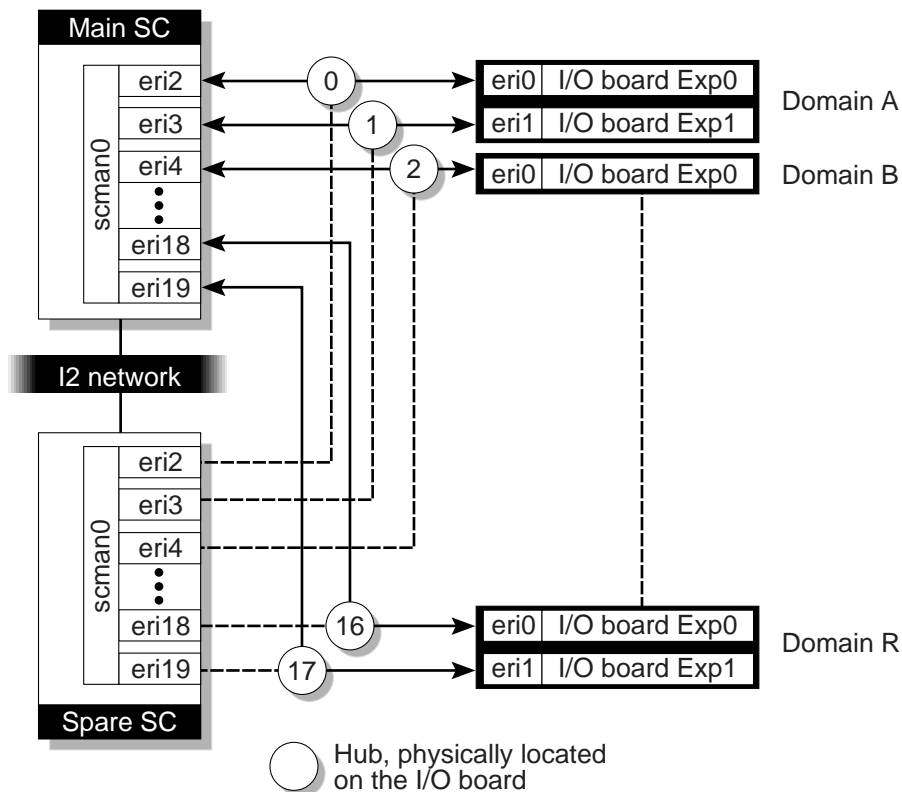


FIGURE 2 MAN I1 Network Overview

I2 Network

The MAN I2 internal network consists of the two system controllers internal redundant NICs. The MAN I2 network, shown in FIGURE 3, is a private system controller-to-system controller network, which is entirely separate from the MAN I1 network. It is used for heartbeat communication between system controllers to initiate failover, when needed, as well as data synchronization between system controllers.

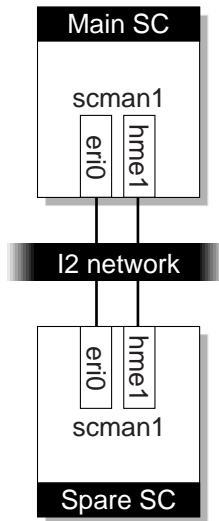


FIGURE 3 I2 Network Detail

The I2 network is configured for failover using the System Management Services (SMS) software configuration tool. The `scman` pings the other system controller every 10 seconds and checks for activity every 30 seconds. If no activity is found, the `scman` initiates a failover. Even if the entire I2 network fails, the failover mechanism can still take place using the high availability static rapid access memory (HASRAM). Because system controllers are some of the most important components of the platform, they should be among the first components tested before completing additional installation tasks. This testing can be done manually by forcing a system controller to failover using SMS commands and verifying that the secondary controller has full control of the platform.

The virtual network adapter on the system controller presents itself as a standard network adapter. It can be managed and administered just like any other network adapter (for example, `qfe` and `hme`). The usual system administration tools, such as

`ndd`, `netstat`, and `ifconfig`, can be used to manage the virtual network adapter. Certain operations with these tools (for example, changing the Ethernet address) should be disallowed, for security reasons.

MAN operates and is managed as an IP network with special characteristics. For example, IP forwarding is disallowed by the MAN software. As such, the MAN operation is the same as any other IP network, with the noted exception documented above.

External Network Planning

External network design and planning is completely separate from the MAN internal I1 and I2 networks. The external application networks should be separate and secure VLANs, as required by the application architecture. Additionally, you should configure managing and monitoring network segments using secure, non-routable IP addresses although, in many cases, this is not typical. These network segments should be located on a switch that is separate from the larger core switch VLANs used for applications. The managing and monitoring network is typically used for system administrators and monitoring agents such as Sun™ Management Center (SunMC). This network should connect through the SunMC server and terminal concentrator, as shown in FIGURE 4 on page 11, and should be secured with the appropriate minimization and hardening.

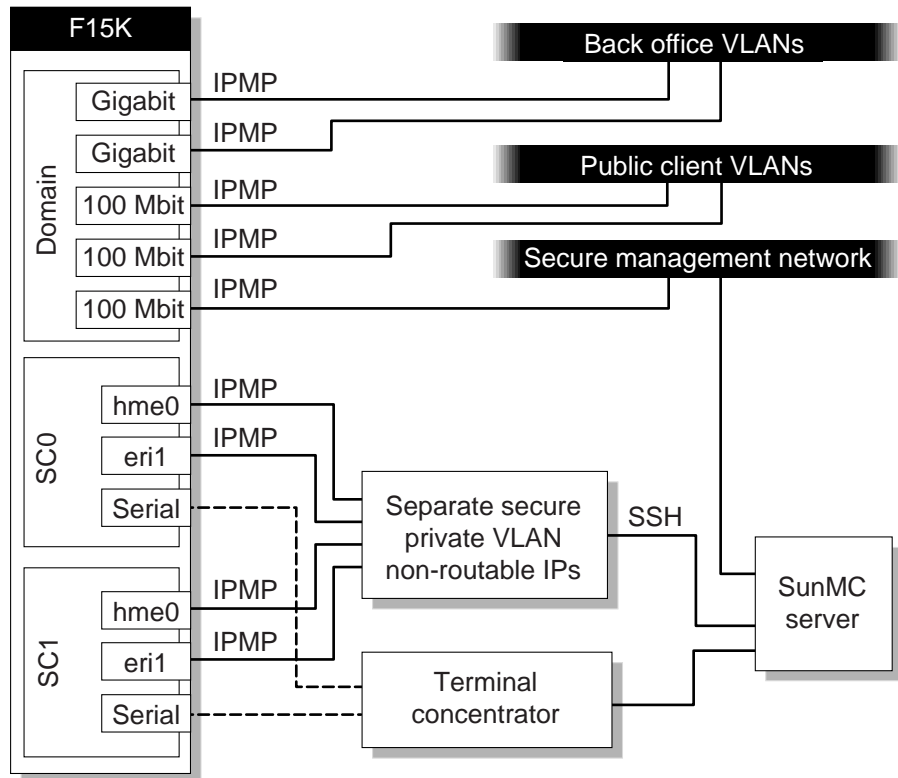


FIGURE 4 External Network Overview

Configuring an External System Controller Network

Each system controller comes with two external network 10/100 ports. To ensure maximum redundancy and quick failover, we recommend that you configure both the `hme0` and the `eri1` public interfaces on both system controllers and include both of the interfaces in an IPMP configuration.

With this type of configuration, seven public IP addresses will be required; two for each system controller's `hme0` and `eri1` (for a total of four), one for each system controller for local failover (for a total of two), and one floating or community IP address. To do this, use the `smsconfig` command on each of the controllers and then reboot. Then, test each of the interfaces to ensure they failover. Depending on the requirements of each site, a default router might be required when all domains on a subnet are secured, which can affect the way IPMP test partners work. Refer to

the appropriate IPMP procedures for configuration instructions regarding installations that require a default router (available at <http://sunsolve.sun.com>). This applies to IPMP configurations on the domains, as well. For specific commands and setup instructions refer to the *System Management Services Installation Guide* and Release Notes for the version of SMS software you are using on the system controllers. FIGURE 5 shows how the system controllers can be implemented with IPMP.

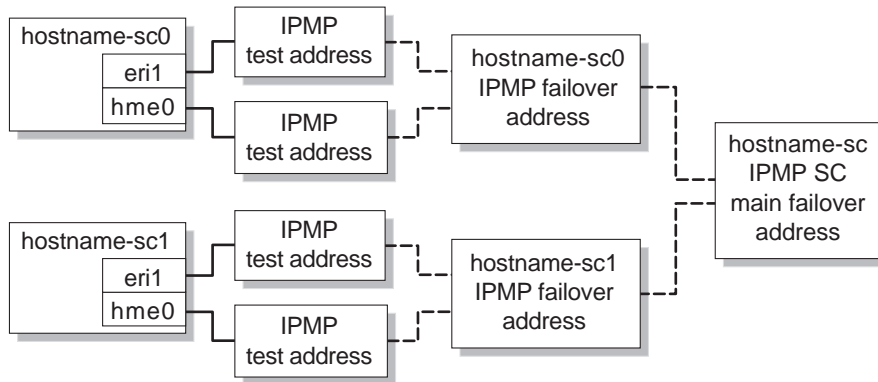


FIGURE 5 Implementing System Controllers With IPMP

For security reasons, separate the system controller's two external network connections `hme0` and `eri1` from all other networks. We recommend that you place this network on separate network switches and that you do not use VLANs on a larger shared core switch. If separate switches are used, monitor them as you monitor other critical network components.

System Controller Configuration

The system controller configuration consists of the system controller disk, console, OpenBoot™ parameters, SMS-SVC user configuration, network, operating system, NTP, and SMS software. Many improvements have been made in SMS software version 1.3, including `ssh` support for file propagation, enhanced access control, and `ssh` configuration through the `smsconfig` tool. In addition to security improvements, SMS software version 1.3 fixes many bugs, improves failover functions, and utilizes the HASRAM (memory chips that keep static information even if power is lost) as a backup to the I2 network. Therefore, we recommend that you use SMS software version 1.3 instead of version 1.2.

Configuring Boot Disks

Each system controller is configured from the factory with the operating system and Solaris™ Volume Manager software (formerly, Solstice DiskSuite™ software) for mirroring boot drives. We highly recommend that you do not change this pre-installed configuration or replace it with other volume management products.

The initial configuration uses the first slice of the boot disk to install the root file system and adds the second slice only for swap space. In this configuration, the fourth and fifth slices of the boot disk contain the Solaris Volume Manager state databases. Sun Services completes the configuration of the boot disk configuration on site, using the Sun Services Engineering scripts (EIS Enterprise Installation Standards) to sync and attach Solaris Volume Manager metadevices. If a site has an established JumpStart™ server, this should be used to initially boot and install the domain's operating system. If this is not the case, a file system can be configured on slice seven for initial installations using a Flash archive. Initial communication to the system controller boot server by the domains can be done through the I1 internal network.

System controllers have two internal 18-gigabyte disk drives that are mirrored using Solaris Volume Manager software. The EIS CD contains a script, `SF15k-sc-bootdisks-start.sh`, that assumes the disk is already formatted with the default partitioning. The script sets up Solaris Volume Manager slices and state databases, and the script `SF15k-sc-bootdisks-finish.sh` completes the process by syncing and attaching the metadevices. The following table shows the default system controller configuration.

TABLE 0-1 Default System Controller Configuration

Partition	Size MB	Mount Point	Description
0	8192	/	Root file system
1	2048		Primary swap
4	11.5		SVM state DB
5	11.5		SVM state DB
7	8192	/export/install	Optional install images

Configuring NTP

Network Time Protocol, generally referred to as NTP, is designed to synchronize the time and date of a client to a time server. If the site has established an NTP configuration to a stratum primary and secondary server, this can be used for the domain and system controller clients. If this is not the case, the system controller can

be configured as the NTP server and the domain configured as the NTP client. Before running any applications on the domains, the NTP configuration should be completed and tested.

To initially configure the time on the system controller, you can use the `setdate` command in the system controller `sms-svc` account `/opt/SUNWSMS/bin/setdate`. The spare system controller will synchronize its time internally using the SMS processes. After setting the system controller's date and time, you can set up the domains to use NTP from the system controllers for synchronization of their clocks.

Initially, the recommendation was that the system controller should not use NTP to set its own clock because no offset adjustments will be made, and the virtual TOD values stored on the domain could get skewed. Many installations, however, require their domain times to be synchronized with a time server other than the system controller. Therefore, the system controller and the spare system controller must be able to synchronize their clocks using NTP to another time server, as well. For a detailed procedure for configuring the main system controller, spare system controller, and the domains as NTP clients, refer to the Sun BluePrints OnLine article "Using NTP on the Sun Fire 15K/12K Servers."

Configuring System Controller `sms-svc` User

The `sms-svc` user account should not be configured to have platform administrative privileges as well as domain administrative privileges for each domain. The use of the `sms-svc` account should be discouraged because it is a shared account and its use makes accountability difficult. You should delegate the `sms-svc` functionality for the platform and domains to specific system administrators and then lock the `sms-svc` account. Additionally, it is not a good practice to assign the system controller's root user account the same SMS privileges as the `sms-svc` user account. Any configuration on the platform should be done by the system administrator logging into the account with platform privileges for consistency. The EIS CD provides a script that configures the `sms-svc` user account as described here, and it should run on each system controller.

Managing System Controller Failover

System controller failover is managed by the daemons running on the primary and the secondary system controller. These daemons communicate across the private network that is built into the Sun Fire 15K/12K server frames called the I2 network. It is a preferred practice, and it is crucial to the highly available environment, for the Sun Fire 15K/12K servers to always have failover enabled, and for the data between the two system controllers to always be in sync. Both system controllers should always be running the same versions of OS and SMS, and they should be maintained

at the same patch level. Schedule periodic system controller failover tests during off-hour maintenance windows, especially after any changes are made. Tests should include using the `SMS set failover force` command as well as the `halt` command for each controller's logon session.

Managing the System Controller Operating System

The system controller operating system and the SMS software are vital to the availability of the Sun Fire 15K/12K servers. A complete and current backup image of the system controller boot disk can prevent unnecessary downtime and possible platform outages. You should create this backup image at regular intervals and after applying any OS or SMS patch updates. SMS software provides the `smsbackup` and `smsrestore` utilities to preserve the current SMS software configuration state and history data. It is imperative that you execute an `smsbackup` after each system configuration change and before applying a new release of SMS software. You can automate this process through the configuration as a `cron` job, being sure not to store the `smsbackup` images on the system controllers themselves.

SMS software patches are considered mandatory and might require change control planning to maintain current levels.

Open System Controller (Non-SMS Qualified Software)

System controllers are dedicated components used only for controlling and monitoring the Sun Fire 15K/12K platforms, and they must run at peak performance. The system controllers are used to manage environmental changes and must react in a timely manner to critical events to prevent hardware failures. For this reason, we recommend that you do not install non-SMS software on system controllers.

If it is absolutely necessary for you to install non-SMS software on system controllers, there is a procedure for verification. The Open System process requires a specific setup and validation. Examples of non-SMS software used on the system controllers could include `arpwatch` (monitors modifications to the `arp` table), backup agents, and monitoring agents. It is the customer's (system administrator's) responsibility to ensure that adequate system controller resources are available to support all SMS functions and all other software packages that are being installed on the system controller.

The `SUNWexplo` package is typically the only additional software package that is added to system controllers. The Sun™ Explorer software scripts are used for acquiring static system configuration information for servicing and supportability. These scripts do not impede the performance or affect the functions of the system controllers.

Platform and Domain Administration

We recommend that you have Sun Support Services perform the initial installation and test of the Sun Fire 15K/12K hardware and software. The installation will use the Enterprise Installation Standards (EIS) CD installation checklist and scripts. The EIS CD is an internal Sun toolkit designed to help Sun Services engineers perform consistent installations that can be supported easily. The EIS CD is available only to Sun Service engineers, however, you can validate that the installation used the EIS methodology by checking the EIS log files in the `/var/sun` directory. These log files show the installation date, EIS version used, and the actions selected during the setup. This installation should be followed by a Sun Professional Services Application Readiness Service (ARS) to test, document, and prepare the platform and domains for mission critical application deployment. The Sun Fire 15K/12K servers operational runbook, testing procedures, and build specifications are included with this service.

Sizing and Allocating Memory

Properly sizing and testing applications before deploying them in the production environment should identify the appropriate memory resource requirements. Internally, Sun Microsystems has many resources dedicated to helping customers properly size specific applications prior to deployment, including Sun™ ONE applications, SAP, PeopleSoft, and ORACLE®, to name a few. Remember, with dynamic reconfiguration (DR), memory can be added online; however, this should not be used as a remedy for poor planning and poor application sizing up front.

You can adjust the Solaris `/etc/system` parameters to optimize domains that require large memory models. These parameters are documented in other Sun BluePrints OnLine publications such as “Configuring and Tuning Databases on the Solaris Platform.” Before using the parameters in the production domain, fully test all modifications to the Solaris `/etc/system` parameters on a separate non-production domain, and install the test domain with all target applications and cluster software replicated as closely to the production domain as is possible. In addition, ensure that frequent checks and monitoring software are in place to detect

memory shortages, leaks, and excessive scanning of memory pages. This is especially important in domains that run many concurrent database instances, which is now commonly done on Sun Fire 15K/12K platforms.

From a hardware perspective, memory DIMMS within each domain of the platform should have a consistent footprint. Therefore, try to design domains with CPU/Memory boards that contain, for example, all 1-gigabyte DIMMS or all 256-megabyte DIMMS. This also eases the planning of resources and DR operations in a consolidated system.

Configuring the Dump Device

The dump device is used when the Solaris OE panics and writes an image of the system memory (physical memory, not virtual or swapped-out memory) to the designated dump device. Usually, this device is designated as the primary swap device. When the system is rebooted after the panic, the operating system performs a savecore dump and then writes the contents to a file on a mounted file system in the `/var/crash/'uname -n'` directory. To manage and change the dump device configuration, use the `dumpadm` command.

There might be other configuration exceptions to consider if the system is using VERITAS Volume Manager with root encapsulation or if the system is being mirrored with Solaris Volume Manager software. In these cases, refer to the latest documentation on <http://sunsolve.sun.com> for detailed configurations. The main consideration is to make the dump device big enough to hold the entire core dump. This can vary by the size of the memory that is configured, but to play it safe, configure the dump device at least 4 gigabytes. Because no hard rules exist, test it to see if it creates usable savecore files. You can do this by setting the `autoboot` parameter to true and using the `sync` command at the OpenBoot prompt. Then, check the `savecore vmcore.x` and `unix.x` files. If space in the file system is a concern for storing the `vmcore.x` and `unix.x` files, you can change the location of both the dump device and the `savecore` directory with the `dumpadm` command.

Configuring Swap

In most instances, application requirements dictate swap space allocation. Some applications, such as SAP, require large amounts of swap space. Sometimes, more space is allocated than is actually needed because of overly conservative recommendations by the application vendor. From a performance perspective, swap space makes a difference only when you are short of physical memory. If the system is not paging, it makes no difference how much swap space you allocate. However, the consequences of running out of swap when a system needs it can have extremely detrimental effects on availability. This is especially true with large domains running many applications or database instances. Therefore, we recommend that you allocate

a lot more space for swap than you expect to need to cope with usage peaks. In addition, when you add applications to existing Sun Fire 15K/12K production domains, review and modify the swap space accordingly. Systems with Oracle 9i databases and Solaris 8 OE that will use dynamic reconfiguration (DR) might also require additional swap.

In the absence of any vendor application-specific swap space requirements, the current practice is to configure swap to be equal to the amount of configurable memory on each CPU/Memory board, and to double that the amount for domains that are doing DR operations. (Note that these are the minimum estimations, not absolute recommendations.) For example, if the Sun Fire 15K/12K domains have two CPU/Memory boards, each configured with 8 gigabytes of memory (16 gigabytes, total, for the domain), then the swap should be 16 gigabytes. If DR is being considered in the domain, configure the swap to be twice the maximum amount of configurable memory on one CPU/memory board or, in this example, 32 gigabytes for the domain. The DR detach operation uses swap space for draining the pagable memory contained on the board that is being removed from the domain.

Keeping OS Software Up-to-Date

Because Sun Fire 15K/12K applications generally require large memory models and high performance, we recommend that you use the most recent Solaris OE version and update that the application will support to take advantage of this. If the application will support it, we recommend the use of Solaris 9 OE to obtain the best possible performance.

In most cases, system kernel parameters are dependent on the application requirements, and any changes should be well tested before you deploy them in a production environment. Some Sun Fire 15K/12K system parameters are required for performing DR operations on domains. These parameters are described in “Dynamic Reconfiguration” on page 27.

Keeping Firmware Current

The firmware levels on all system boards should be identical to each other. Always use the latest firmware available for the version of SMS being used, even if it means you have to downgrade a board. Each CPU/Memory board contains two `fproms` that must be maintained. Because the CPU/Memory boards on the Sun Fire 15K/12K, the Sun Fire 6800/4810/4800, and the Sun Fire 1280 servers are the same, they should be flashed to the same levels if there are plans to swap or use common components for maintenance purposes. Additionally, the system controller's OBP and POST must be `flash` updated to the same levels.

Security

This section provides information about securing Sun Fire 15K/12K domains, system controllers, and any external networks running Sun Fire 15K/12K servers. This topic has been covered in great detail in many other Sun BluePrints OnLine articles and these references are provided accordingly.

Securing System Controllers

The Sun Fire 15K/12K system controller are vital and critical components of the platform. They are the central control points for all Sun Fire 15K/12K management activities and, therefore, require the highest level of security hardening. Because all domains can be brought down if system controller procedures or access is compromised, only certain administrators should have access to the system controllers. To protect system controllers, restrict access to the system controllers as much as possible, using access control configurations. At some sites, administrators need to administer all domains, which is okay as long as privileges and access is given only to administrators who really need them. We recommend provisioning a separate and dedicated network for the system controllers, which only administrators can access.

All administrators should implement the recommendations documented in the Sun BluePrints OnLine article, “Securing the Sun Fire 15K Controller.” The topics in this document include the Solaris Security Toolkit, supportability, SMS software, setting up administration accounts, system controller network interfaces, Solaris™ Secure Shell (SSH) configuration, and a summary of security recommendations.

Securing Domains

Each domain of the Sun Fire 15K/12K servers run a separate instance of the Solaris OE. Therefore, the rules for security apply as they would for securing other Sun servers. The tools and processes for securing a Solaris OE should include, but should not be limited to, firewall solutions, Solaris Security Toolkit, data encryption, role-based access control, SSH, and IPsec. The specifics for securing the Solaris OE are beyond the scope of this document but can be referenced in the Sun BluePrints OnLine article “Securing the Sun Fire 15K/12K Domains.”

The Sun Fire 15K/12K servers should be configured to provide separate isolated domains where users accessing one domain should not be allowed to access other domains and the system controller. Therefore, the security model you implement should prevent users from gaining access to domains, unless they are permitted to

access them. This can be accomplished using multilayered access control. When creating domains, ensure that domain administrators have access only to administer their specific domains and that they do not have access to the entire platform. The Sun Fire 15K/12K server platforms have the ability to grant access control at multiple levels by assigning certain administrators non-root user IDs with different UNIX® groups. This gives these administrators certain SMS capabilities, but does not give them all capabilities. These administration groups can be broken down as follows:

- **Platform administrator group.** Hardware administration, platform configuration, environmental status, and power management, but no access to individual domains.
- **Platform operator group.** Platform status and power only.
- **Platform service group.** Platform operator, plus limited platform configuration privileges.
- **Domain administrator groups.** Can manage only their respective domains, but they have no access or control of the platform.
- **Domain operator group.** Can manage only power and domain board configuration for their respective domains, but have no other domain control capability.
- **System controller root user.** Has root access to the system controller and associated functions.

We recommend that you implement a secure shell tool as a means to encrypt data being communicated to and from the system controllers and any administration servers and networks. Solaris 9 OE includes Solaris Secure Shell as a supported utility, or SSH can be obtained as freeware or a commercial product. An SSH implementation can be used to prevent security risks such as password theft and session intrusion, and can replace risky UNIX security commands such as `rlogin`, `rsh`, `rcp`, `ftp`, and `telnet`. For information about implementing and configuring OpenSSH in the Sun Fire 15K environment, reference the Sun BluePrints OnLine article, “Building and Deploying OpenSSH for the Solaris Environment.”

Securing Networks

In addition to hardening the Solaris OE against security attacks, it is important to secure the Sun Fire 15K/12K server’s external network. At a high level, this includes developing company-wide network security policies, security assessments, and audits, and identifying how to fix known security risks and vulnerabilities. This should also include, but should not be limited to, applications, firewalls, switches, servers, intrusion detection, and authentication. For information about this topic, reference the documents listed in “Related Resources” on page 31.

Error Analysis and Diagnosis

The Sun Fire 15K/12K servers have many tools and utilities designed to help monitor, isolate, and report faults and errors.

Detecting ECC Errors

Occasionally, Sun Fire 15K/12K processors experience and detect a correctable memory error called an error correction code (ECC). When the server detects an ECC, it tries to correct the error, logs the error to its fault status register, and continues operation. Additionally, the Solaris OE logs these errors to the error log and to the system console as part of the error reporting process. Several errors might be logged in this process for just one correctable error event. Recognizing and understanding these types of errors can be complex and is best left for Sun Support Services engineers.

For information about memory error handling and a summary of asynchronous fault tags (AFTs), reference the Sun white paper “Soft Memory Errors and Their Effect on Sun Fire System,” available at <http://www.sun.com/products-n-solutions/hardware/docs/pdf/816-5053-10.pdf>

Selecting Diagnostic Tools

To capture information used in support and problem resolution, implement a consistent process and set of tools on all Sun Fire 15K/12K server domains and system controllers. We recommend that you install the Sun Explorer Data Collection tool on all domains and run it after every major-change event. It is best to run the Sun Explorer scripts when the system is fully up, but not under a heavy load. Optionally, you can install the Sun Explorer tool and run it from an NFS server. The Sun Explorer tool changes frequently, so stay up-to-date on the latest versions. The tool is available for download from the Sun support portal at <http://sunsolve.sun.com>. You can also use the output data as an optional follow-up check of the installation using the Sun Services RAS profile.

Isolating Faulty Components

The Sun Fire 15K/12K servers have several mechanisms for identifying and isolating components that are faulty. These mechanisms include `redx`, POST, `dsmd` dumps, automatic system recovery (ASR), and log files. The `redx` program (sometimes

referred to as “red cross” or “red ex”) is used for debugging and maintenance purposes. The `redx` program is normally reserved for Sun Services engineers performing low-level hardware and firmware diagnostics. You can run the `redx` program offline on a separate Solaris workstation to look at `Dstop` (domain stop) dump files to investigate fatal errors, such as CPU internal failures.

The `dsmd` (domain status monitoring daemon) dumps are files created by the SMS domain status monitoring daemon. If the daemon detects a component fault error, contact Sun Services before attempting any physical hardware replacement. You can use the SMS software commands to isolate and place components offline until they can be replaced by Sun Services.

In some situations, the ASR feature of Sun Fire 15K/12K servers will automatically detect a fault error in a component, and the SMS process will place the component into the black list file. The failed component will then be deconfigured from the system by the POST on the next reboot.

Platform and Domain Configuration

By design, the Sun Fire 15K/12K servers are well suited for running many applications that scale horizontally, as well as vertically. However, due to the very high capacity and efficient footprint of these servers, they are best suited for vertically scalable applications with dense consolidation. This type of scalability allows the platform and the application to expand by adding additional resources in same domain, or across multiple domains. Additionally, the application can take advantage of load balancing using the internal multitasking and threading attributes of the operating system and application. Either a single instance or a single domain can be made larger, or multiple instances of the applications can be installed on a single domain to accommodate additional users and functionality. Server resource management tools (such as Solaris™ Resource Manager) can be added to control and allocate system resources such as CPUs, processes, and memory.

Configuring Domains for Redundancy

Internally, most of the Sun Fire 15K/12K server components are designed with built-in redundancy and online recoverability. The main reasons for configuring separate domains are fault containment, security isolation, and workload separation. The Sun Fire 15K/12K servers are designed so that the physical location or proximity of domain hardware components is not relevant. This means that CPU/Memory boards (slot 0 boards), or I/O assemblies (slot 1 boards) can be located anywhere on the frame in their respective slots and will still be part of the same domain.

The minimum configuration for any Sun Fire 15K/12K domain is one CPU/Memory board, one gigabyte of memory, an hsPCI assembly, access to the backplane through an expander board, one network-capable PCI card, and a local boot disk subsystem. However, in most production environments, these minimum requirements would produce an unacceptable configuration due to the lack of redundant components. Therefore, we recommend that you configure all mission-critical production domains with redundant expander boards, CPU/Memory boards, and I/O boards.

All mission-critical domains should have a minimum of two CPU/Memory boards, two expander boards, and two I/O boards. Additionally, each mission-critical domain should have a minimum of two network cards, two boot I/O paths, and two data storage HBAs, each installed in separate expanders. Depending on the chosen cluster product, you should also implement dynamic multipathing and failover technologies on both the network and data storage. For example, you might use the Sun StorEdge™ Traffic Manager software, VERITAS DMP, or the Hitachi DLM for storage path failover. You might use Sun IP Multipathing (IPMP) or VERITAS Cluster for the network NIC failover. If the domain is configured with multipathed technologies, such as DMP or STMS, all primary paths should be located on one I/O assembly tray, and the secondary paths should be located on another I/O assembly tray. Having multiple paths for all I/O devices on separate I/O assemblies enhances the ability for performing dynamic reconfiguration operations.

Applying Naming Standards

Naming platforms and domains is a very important task, though often overlooked. Domain and platform names should be designed to minimize complexity so that system administrators are not easily confused about which domain they are working on. Also, a naming scheme should be developed so that possible intruders cannot easily identify mission-critical production domains and platforms. Because making name changes after the domains are placed into production can be a major headache, complete this task and establish a naming standard early in the planning phase of the project.

Patching Domains

Unlike with system controllers, where keeping up-to-date on the latest recommended patches is required, domains are generally managed differently with regard to patches. A well thought out patch management strategy is important for all domains, and the strategy you choose might depend on factors such as application requirements, outage windows, and testing. When patching domains, remember to validate that the proposed patches will actually fix the problem you are having, ensure that the proposed patch will not cause other problems even though it

will fix something else, and ensure that the patch will not affect the system's performance. There are many patch management tools and utilities from Sun Microsystems that make this task easier, including `patchdiag`, Solaris™ Patch Manager, PatchCheck, PatchPro Expert, signed patches, Live Upgrade, and SunMC Change Manager. Live Upgrade can be used to test domain's new patches, and if necessary, provide the ability to quickly fall back to the operating system's state as it was before the patch was applied.

Splitting Expander Boards

Internal to the Sun Fire 15K/12K servers, expander boards are used for connecting the CPU/Memory boards and I/O boards to the Sun Fire interconnect ports on the backplane. The Sun Fire 15K/12K server configurations have the option to “split” expander boards. This means that two different boards connected to the same expander can be assigned to separate domains. This allows an I/O board to be assigned to another domain separate from the CPU/Memory board installed on that same expander. Because some domains require more I/O than CPU, the CPU/memory board and its resources could be assigned for Domain A, and the I/O board and its resources could be assigned for Domain B. In effect, this gives the Sun Fire 15K/12K servers enhanced flexibility. In turn, this added flexibility brings some availability and performance drawbacks. Because a “split expander” board is a shared physical resource among two domains, a failure of this board set affects both domains. Additionally, there is a performance penalty for configuring a split expander configuration. It is best to fully evaluate the advantages and disadvantages of the added flexibility before using this feature.

Domain Boot Devices

Although it is possible to boot domains over the network or from the DVD drive, we recommend that you do not do so on live production domains. Booting over the network is acceptable for initial installations and troubleshooting, but for mission-critical production domains, you should configure separate mirrored boot devices (RAID1). Mirrored boot devices should be configured on either the Sun StorEdge S1 or the Sun StorEdge T3 arrays, which are the only currently supported Sun Fire 15K/12K server boot devices. The arrays should be connected with redundant paths on separate I/O boards and I/O assemblies, and should be configured with either Solaris Volume Manager or VERITAS Volume Manager software. The boot arrays should also be installed across separate data center racks with separate power sources for added availability. When mirroring the boot devices, new `devalias` names should clearly identify the primary and secondary boot drive at the OBP prompt. For example, “bootdisk” for the primary and “mirrordisk” for the secondary. Be sure to document all procedures for boot disk recovery in an online runbook.

We recommend that you install the S1 boot array in the first PCI card slot that is accessed by the OpenBoot PROM (OBP) probe list, which is generally the lowest numbered I/O board slot in the domain. This will guarantee that the device paths and the configuration to the `/etc/path_to_inst` will not change if a subsequent boot with the reconfigure option (`boot-r` or `reboot -- -r`) is used after additional components are added to the domain.

The configuration of domain operating system disks varies from site to site. The key thing to remember is to build the operating system configuration so that it can be easily upgraded with a minimum number of file systems and is consistent across domains. Also utilize technologies such as Solaris Volume Manager or VERITAS Volume Manager, as well as Flash Archive, JumpStart, and Live Upgrade technologies. The simpler the configuration is normally, the better when configuring OS disks.

Monitoring the Server

We recommend that you monitor the Sun Fire 15K/12K servers with Sun Management Center software agents. Although SunMC is the recommended tool for first-level monitoring of the Sun Fire 15K/12K server platforms, it should not be considered as the only tool that will fulfill the requirements for a large scale enterprise monitoring effort. You could also consider using the Sun SRS Net Connect tool. This tool is a collection of system management services designed to enable you to securely monitor hardware, alarming, system performance, and trend reporting through internet access. You can easily integrate SunMC with other supporting enterprise-wide management and monitoring platforms, such as Tivoli and BMC, through supplied management information base (MIB) files for SNMP. Although it is not a monitoring tool, SunMC Change Manager can also be integrated into the system to track system changes and patches, and to provide resource provisioning capabilities.

The SunMC server component should be installed on a separate Sun Enterprise™ or Sun Fire class-two processor server, with a minimum of two network interfaces. Additionally, this server should be security hardened. For each agent, there is a set of statuses and rules that define the alarm conditions that can be generated and tuned by that agent. This information is documented in the *Sun Management Center Supplement for the Sun Fire 15K Systems* (Part # 816-2701-10), which is available at <http://docs.sun.com>.

The following list summarizes the Sun Fire 15K/12K server categories that should be monitored by SunMC:

- Platform configuration reader (environmental platform agent)
- Domain configuration reader (domain configuration agent)
- System controller configuration reader (system controller environment and configuration agent)

- Platform/domain state management module (manage and manually monitor the platform and domain states)
- Dynamic reconfiguration module (manage and manually monitor the device and dynamic attach points and availability)
- System controller monitoring module
- Agent statistics

Performing Online Maintenance Testing

Regularly test the following technologies and procedures to make sure they are functioning properly. Some of these tests might not be needed, depending on the particular environment. Additionally, some tests might be required only after changes are made to specific system components or operating system configuration changes. Each site should develop and execute a test plan specific to its needs.

- Replacing CPUs online
- Adding CPUs online
- Adding memory online
- Replacing memory online
- Replacing I/O cards online
- Adding I/O cards online (reference supported configurations)
- Performing a live upgrade of the OS (Using Live Upgrade)
- Performing hot patching
- Performing system controller failover
- Testing cluster failover
- Deleting and adding boards to domains (if these functions are used on a regular basis by the environment)
- Booting domains over the network from the system controller or from a JumpStart server
- Booting domains and system controllers from secondary mirrored boot devices
- Testing IPMP, DMP, DLM, or STMS failover (whichever applies)
- Security

Dynamic Reconfiguration

Dynamic reconfiguration (DR) is a powerful tool for allocating and de-allocating resources to and from a domain with minimal interruption. Included in the features of DR is the ability to add or delete CPU/Memory boards within a running domain, as well as move them between running domains. You can also hot swap CPU/Memory and PCI boards using DR. DR operations support both hsPCI I/O assemblies and MaxCPU boards, as well as CPU/Memory boards. DR is also very useful when servicing failed components in the system because they can be dynamically removed from the running domain.

Some applications are better suited to maximize the effectiveness of DR than others. To fully understand the impact of a DR operation on a domain, detailed knowledge of the domain's configuration and application workload is critical. This section provides some preferred practices that you should consider when designing and deploying a domain that will utilize DR. For more information about DR, reference the online Sun documentation sets for DR, which include the *Dynamic Reconfiguration Users Guide*, *Dynamic Reconfiguration Installation Guide*, Release Notes, and the *SMS Users Guide*. This documentation can be found at:

<http://docs.sun.com>, <http://www.sun.com/products-n-solutions/hardware/docs>, and <http://sunsolve.sun.com>

Before you use DR, you need to validate some basic tasks. Make sure that certain components are up-to-date and available. This includes the proper Solaris OE version and update, SMS version and patches, CPU memory Fcode versions, and SunMC version. Some general guidelines can be implemented into the design of the Sun Fire 15K/12K server to enhance the ability for successful DR operations. This should include spreading memory and CPUs evenly across all boards in the frame and configuring applications with the mind-set that both the OS and the application will need to be made inactive for a short period of time to complete the detach process.

Additionally, when designing Sun Fire 15K/12K domains for DR, consider that domains with only one CPU/Memory board cannot be detached. Boards with only one path to the boot array cannot be detached. All critical resources that will be detached must be redundant and must have alternate paths, and most importantly, the domain that is detaching a board must be able to tolerate less processors and memory.

On a tactical level, the following tasks must be completed and verified.

- Ensure that the board is flashed to the correct LPOST version, check on <http://sunsolve.sun.com> for the latest firmware versions.

- When adding a component you must be able to verify that the firmware level on the domain matches the firmware level on the component that is being added to the domain.
- Are there any bound processes to the CPU/memory board that is being detached? (The `psbind` command can be used to detect these processes.)
- Are there any unsupported or third party adapters in the board that is being used in the DR operation?
- Are there any ‘real-time’ processes running such as NTP on the domain that is detaching a CPU/memory board?
- Are there available boards equipped in the domain to receive the permanent memory from the board that is being detached?
- The proper Solaris `/etc/system` parameters must be in place to enable caged mode for the kernel and to allow for memory to be moved.
- The status of the component must show that it is available. To check the status of the component, use the `showboards` command.
- When removing a component from a domain, check to see if the board contains permanent memory (nonpageable memory such as the kernel) by using the `cfgadm` command.
- Memory interleaving must also be set on the boards that are being detached.
- Once the DR operation has completed successfully, check that the operating system has detected the change by using common UNIX commands such as `psrinfo` and `prtconf | grep -i memory`. Also use the SMS commands such as `showplatform` and `showdevices`. All the DR messages will be logged in the `/var/adm/messages` file for the domain involved in the DR operations.
- Is the domain involved in the DR operations running a database that contains ISM segments?

Using Dynamic Reconfiguration With Oracle Databases

Many applications that run on the Sun Fire 15K/12K servers use a database (primarily, an Oracle database). When using DR to move system board resources on a domain running Oracle, domains must be configured with sufficient physical memory and swap space to contain the memory-resident components of both the new (attached) database image and the old (detached) images. If sufficient physical memory is not in place at the time of the DR operation, the operating system will return an error. This occurs because the Solaris OE cannot obtain enough free memory within the domain to move the intimate shared memory (ISM) segments out of the CPU/memory board where memory is being drained. In addition to moving ISM segments, the domains must also be able to store the kernel, Oracle instances, and shared memory segments. Therefore, when designing a domain for

DR, make sure that the smallest set of system boards to be used for a domain running an instance of Oracle can contain the kernel, Oracle shadow processes, ISM segments, and shared memory (SGA) segments.

Oracle databases use intimate shared memory, this allows a database to share data within the shared memory space between processes. It also locks shared memory used by the database so it will not be swapped out. Therefore, because ISM segments cannot be paged out during the drain operations, they must be relocated to other physical memory on a remaining system board. You can identify the location of ISM segments using the `cfgsadm` command, where they are reported as permanent memory. Sufficient memory for the ISM segments must remain in the domain after the CPU/Memory board is detached for the DR operation to be successful. After you relocate ISM, Oracle will lock pages involved in I/O operations, making them inaccessible to a DR operation until the I/O completes. Be aware this will slow down the progress of a DR operation on a heavily loaded system.

The combination of Oracle 9i and the Solaris 8 OE or the Solaris 9 OE has a unique feature when used in combination with DR. When memory is added by a DR operation, Oracle instances can dynamically recognize additional memory within the domain without having to restart the Oracle instance. This is accomplished by the Oracle 9i feature called Dynamic SGA and the added support within the Solaris 8 OE for dynamic intimate shared memory (DISM). The key point is that, depending on the Solaris and Oracle versions being used, you might not have to restart some applications before they will recognize the added resources provided by the DR operations.

Implementing Dynamic Reconfiguration Procedures

When detaching components from a domain, you must consider the implications to the application. Before attempting the DR operation, all DR activities should be fully tested with standard testing procedures and should be well documented. Be sure to configure the test domain with the same technologies and applications used in the production environment to simulate the production domain as closely as possible. It is critical that you strictly adhere to testing, because the detach operation must pause the operating system and drain memory pages being used by the operating system and applications, thereby temporarily suspending CPUs, processes, and devices. The testing must show that the pausing and draining operation during detach does not adversely affect the applications and operating system on the domain. The testing must also validate that the applications running in the domain where a board is being removed can run effectively with less memory, CPU resources, or I/O bandwidth. Special provisions might be required for DR operations with VERITAS Cluster Server and Sun Cluster™ software, and you should test these procedures as well. This might mean that domains that are part of an active cluster will need to be taken out of the cluster during the DR operation. Also, testing should validate that

the boards being added to a domain are reliable. Therefore, the domain used for testing DR operations should test the exact set of boards being proposed in the DR attach operation to the production domain.

About the Author

Lee M. Lustig has been an IT architect with Sun Professional Services for the past seven years. He has over 12 years experience helping customers architect and implement high-end servers and mission-critical applications. Lee has been involved in many large Sun Fire 15K server and Sun Enterprise 10K server implementations for Sun Professional Services.

References

Beloro, Jason. "Using NTP on the Sun Fire 15K/12K Server," *Sun BluePrints OnLine*, June, 2003

To access this article online, go to <http://www.sun.com/solutions/blueprints/0603/817-2979.pdf>

Noordergraaf, Alex and Nimeh, Dina. "Securing the Sun Fire 15K Controller," *Sun BluePrints OnLine*, February, 2003

To access this article online, go to <http://www.sun.com/solutions/blueprints/0203/817-1358.pdf>

Noordergraaf, Alex and Nimeh, Dina. "Securing the Sun Fire 15K/12K Domains," *Sun BluePrints OnLine*, February, 2003

To access this article online, go to <http://www.sun.com/solutions/blueprints/0203/817-1357.pdf>

Packer, Allan. "Configuring and Tuning Databases on the Solaris Platform," *Sun Microsystems Press* ISBN# 0-13-083417-2,

To access this book online, go to <http://www.sun.com/books/catalog/packer/>

Reid, Jason. "Building OpenSSH—Tools and Tradeoffs," *Sun BluePrints OnLine*, January, 2003

To access this article online, go to <http://www.sun.com/solutions/blueprints/0103/817-1307.pdf>

"Sun Management Center 3.0 Supplement for Sun Fire 15K Systems,"

Part # 816-2701-10. To access this paper, go to <http://docs.sun.com>

SMS Administrators Guide. Sun Product Documentation.

Soft Memory Errors and Their Effect on Sun Fire Systems, Sun Microsystems Press

To access this book online, go to <http://www.sun.com/products-n-solutions/hardware/docs/pdf/816-5053-10.pdf>

Sun Fire 15K Installation and De-Installation Guide. Sun Product Documentation.

Sun Fire 15K Site Planning Guide. Sun Product Documentation.

System Management Services Installation Guide. Sun Product Documentation.

Related Resources

Domain Security:

Enterprise Security Solaris Operating Environment. ISBN #0-13-100092-6. Upper Saddle River: Prentice Hall.

To access this book online, go to <http://www.sun.com/books/catalog/noord2/>

Solaris Security Toolkit (JASS) and supporting documentation

To access this toolkit and all supporting documentation online, go to <http://www.sun.com/software/security/jass/>

Noordergraaf, Alex. "Solaris Operating Environment Minimization for Security: A Simple, Reproducible and Secure Application Installation Methodology,"

Sun BluePrints OnLine, November, 2000

To access this article online, go to <http://www.sun.com/solutions/blueprints/1100/minimize-updt1.pdf>

Several articles on Solaris Operating System security, *Sun BluePrints OnLine*

To access these articles online, go to <http://www.sun.com/security/blueprints>

Network Security:

Andert, Donna; Wakefield, Robin; and Weise, Joel. "Trust Modeling for Security Architecture Development," *Sun BluePrints OnLine*, December, 2002

To access this article online, go to <http://www.sun.com/solutions/blueprints/1202/817-0775.pdf>

Noordergraaf, Alex. "Building Secure N-Tier Environments," *Sun BluePrints OnLine*, month year,

To access this article online, go to <http://www.sun.com/solutions/blueprints/1000/ntier-security.pdf>

Weise, Joel and Martin, Charles. "Developing a Security Policy," *Sun BluePrints OnLine*, December, 2001

To access this article online, go to <http://www.sun.com/solutions/blueprints/1201/secpolicy.pdf>

Ordering Sun Documents

The SunDocsSM program provides more than 250 manuals from Sun Microsystems, Inc. If you live in the United States, Canada, Europe, or Japan, you can purchase documentation sets or individual manuals through this program.

Accessing Sun Documentation Online

The docs.sun.com web site enables you to access Sun technical documentation online. You can browse the docs.sun.com archive or search for a specific book title or subject. The URL is <http://docs.sun.com/>

To reference Sun BluePrints OnLine articles, visit the Sun BluePrints OnLine Web site at: <http://www.sun.com/blueprints/online.html>