An Information Technology Management Reference Architecture

Designed for the Sun Internet Data Center Mail and Messaging Reference Architecture

By Edward Wustenhoff

Chief IT—Consultant

`http://www.sun.com/blueprints`

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

Part No. 816-5134-10
Revision A June 2002

Copyright 2002 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, California, U.S.A. All rights reserved.

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.

Sun, Sun Microsystems, the Sun logo, Sun BluePrints, Sun Enterprise, Sun Fire, Sun Management Center, Sun StorEdge, SunPS, Netra, iForce, and Solaris are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries.

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.

RESTRICTED RIGHTS: Use, duplication, or disclosure by the U.S. Government is subject to restrictions of FAR 52.227-14(g)(2)(6/87) and FAR 52.227-19(6/87), or DFAR 252.227-7015(b)(6/95) and DFAR 227.7202-3(a).

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 2002 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, California, 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 déposée aux Etats-Unis et dans d'autres pays et licenciée exclusivement par X/Open Company, Ltd.

Sun, Sun Microsystems, the Sun logo, Sun BluePrints, Sun Enterprise, Sun Fire, Sun Management Center, Sun StorEdge, SunPS, Netra,iForce, et Solaris sont des met arques de fabrique ou des marques déposées, ou marques de service, de Sun Microsystems, Inc. aux Etats-Unis et dans d'autres pays.

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 REPONDRE 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™

Information Technology Management Reference Architecture

This article is the fourth in a series of articles by Edward Wustenhoff on the implementation of a data center management infrastructure. The focus of this article is on the infrastructure concepts and the details of the other requirements that drove the implementation of the management infrastructure.

This is a follow-up article on the "Management Framework Considerations" article published earlier by Edward Wustenhoff and the Sun BluePrints group. It describes the results of the process used to create an IT management reference architecture in the iForce℠ Ready Center (iFRC) that displays an IDC Mail and Messaging Architecture. The iFRC is a program at Sun that provides reference implementations and proof of concepts to assist our customers in preventing common pitfalls.

A fifth article, to be published in July 2002, will describe the details of the actual implementation.

Quality of Service (QoS) is an important competitive advantage in the Internet service provider (ISP) and large IDC market so Sun decided to develop a management infrastructure to highlight how such a service can be managed.

This management infrastructure provides six views into the data center.

- A Network Operations Center (NOC) view that shows all alerts in all components of the architecture. Each component has its own view.
- A Service Level Management view (SLM) view that reports the status of the messaging application against a predefined Service Level Agreement (SLA).
- An Applications Infrastructure view that reports on the status of the software components that support the application. Lightweight Directory Access Protocol (LDAP), Network Time Protocol (NTP) and domain name service (DNS) are included in this.

- A System Administration view that reports on the status and enables the administration of all computer system components.
- A Network Administration view that reports on the status and enables the administration of all network components.
- A Performance view that monitors the status of the system and networks and their ability to handle the processing loads.

It is important to realize that the implementation of this architecture was done within the capabilities of the iFRC. This means that the developers worked with the systems that were available. The result is a proof of concept that emphasizes functionality. Availability and sizing of the management platform was not the main focus. Management tools selection and their functions was done within similar constraints. The developers implemented tools that were known to provide the desired functionality and were available in the laboratory.

However, the implemented architecture follows Sun's tools framework design considerations as much as possible. The intention is to build out the functionality and its associated tools in subsequent releases of this architecture.

"Concepts" describes briefly the underlying design considerations and essential concepts. Its main purpose is to position what the developers think about when addressing data center management, and to create an overview that positions where this proof of concept fits.

"Concepts" addresses two main ideas—Service Level Management (SLM) and the IT management framework. Each concept drives certain implementation requirements, which are listed as key considerations.

The section on “Implementation Requirements” on page 9 contains a detailed description of additional requirements that drove the actual implementation of the management infrastructure.

A follow-on article will describe the details of the actual implementation.

Concepts

One of the main concepts to address is the principle of SLM. The following section briefly explains how the developers look at this process from an instrumentation perspective.

SLM Process

FIGURE 1 is a graphical representation of the main SLM process.

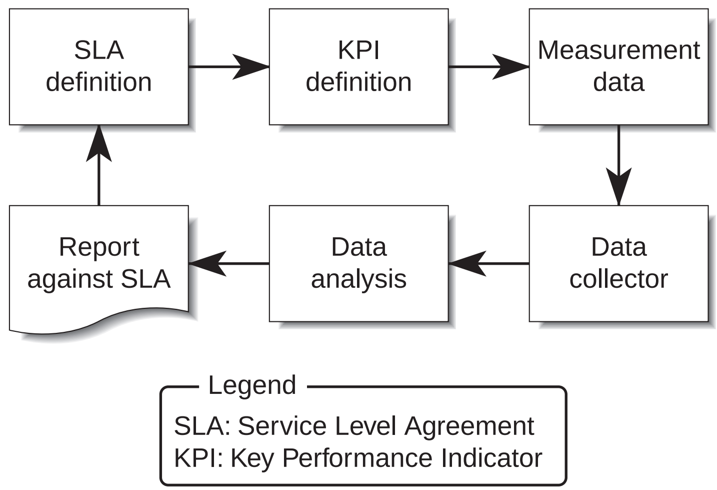

FIGURE 1 SLM Process Flow Diagram

This process shows that the SLA definition sets the key performance indicators (KPI), which in turn drive the measurement data or metrics. The measurement data metrics drive the choice of the data collectors (or agents), that, in turn, influence the data analysis tool. The process creates a report that measures the infrastructure's performance against the SLA. If the SLA objectives are not met, a choice remains—improve the performance or, if necessary, adjust the SLA.

Key Considerations for Service Level Management

This section describes the main SLM considerations that drove the management and operations (M&O) architecture. It points out the main considerations in the design process, not the details of SLM.

The description in this section follows the core SLM process.

SLA Definition

The following SLA sets the requirements for the SLM objectives. It is based on the `nextslm.org` sample template. Mainly, the modifications reflect an email focused ISP environment, because this is what the Sun IDC Mail and Messaging Architecture represents.

The Sun IDC Mail and Messaging Architecture is used by simulated email users to send and receive email. TABLE 1 lists the SLA objectives that the design of this architecture guarantees to meet.

TABLE 1 Service Level Agreement

<table>
<tr><td colspan="2">The send and receive email capability will be available 99.99 percent(*) of the time during normal hours of operation as described in the Test Plans. Any individual outage in excess of 1 minute or sum of outages exceeding 4 minutes per month will constitute a violation.</td></tr>
<tr><td colspan="2">99.99 percent of Sending and Receiving email transactions (using SMTP, IMAP and POP protocols) will exhibit 10 seconds or less response time. Response time is defined as the interval from the time the emulated user sends a transaction to the time a visual confirmation of transaction completion is received by the Micromuse probe. Missing this metric will constitute a violation.</td></tr>
<tr><td colspan="2">The Internet Data Center Customer Care team will respond to service incidents that affect multiple users within 1 hour, resolve the problem within 72 hours, and update status every 4 hours. Missing any of these metrics on an incident will constitute a violation. [Note: not currently implemented since these statistics will come from a process management tool like Remedy. The deployment of this is planned for the next phase].</td></tr>
<tr><td colspan="2">The Internet Data Center Customer Care team will respond to service incidents that affect individual users within 24 hours, resolve the problem within 3 business days, and update status every Day. Missing any of these metrics on an incident will constitute a violation. [Note: not currently implemented since these statistics will come from a process management tool like Remedy. The deployment of this is planned for the next phase].
The Internet Data Center Customer Care team will respond to non-critical inquiries within 24 hours, deliver an answer within 5 business days, and update status every other business day. Missing any of these metrics on an incident will constitute a violation. [Note: not currently implemented since these statistics will come from a process management tool like Remedy. The deployment of this is planned for the next phase].</td></tr>
<tr><td colspan="2">The following shows the number of violations and associated penalty on a monthly basis.</td></tr>
<tr><th>Number of violations</th><th>Penalty</th></tr>
<tr><td>1 > 5</td><td>1 day free service</td></tr>
<tr><td>5 >10</td><td>A free month of service</td></tr>
<tr><td>10 ></td><td>2 free months of service plus a corrective action plan that details the steps taken to correct the issues.</td></tr>
<tr><td colspan="2">As services and technologies change, the SLA can change to reflect the improvements and/or changes. This SLA will be reviewed every 12 months and updated as necessary. When updates are deemed necessary, the customer will be asked to review and approve the changes.</td></tr>
</table>

TABLE 1 contains a “short-form” SLA to illustrate the essential aspects between a consumer and the Service Provider in an IDC context. Internal SLAs, between operations support groups in the Internet data center (IDC), for example, are different and often contain more details and specifications.

TABLE 2 describes how the percentage relates to actual unavailability times. This table shows the impact of measuring against a “number of nines” of availability.

Note – One year equals 365.00 days.

TABLE 2 Actual Availability and Percentage of Downtime

Availability	Uptime (days/ year)	Downtime (days/year	Downtime (hours/ year)	Downtime (minutes/ year)	Downtime (minutes/ month)	Downtime (minutes/ week)	Downtime (minutes/ day
99.000 percent	361.35000	3.65000	87.60000	5256.00000	438.00000	101.07692	14.43956
99.900 percent	364.63500	0.36500	8.76000	525.60000	43.80000	10.10769	1.44396
99.990 percent	364.96350	0.03650	0.87600	52.56000	4.38000	1.01077	0.14440
99.999 percent	364.99635	0.00365	0.08760	5.25600	0.43800	0.10108	0.01444

Key Performance Indicators

Based on the SLA, the following parameters are the essential indicators that define the performance of the email service:

- Sending an email using the SMTP protocol
- Receiving an email using the POP3 protocol
- Receiving an email using the IMAP4 protocol

The system is deemed unavailable if the response time of any of the KPIs is larger than 1 minute. If the service is unavailable for more than 4.38 minutes per month or 1.01077 minutes per week, the SLA has not been met.

A response time larger than 10 seconds for any of the KPIs will constitute a service performance violation.

As a result the service will be tested by using a synthetic transaction that simulates these parameters. One message is delivered using SMTP that is consequently retrieved using IMAP4 and another message is delivered by using SMTP that is consequently retrieved using POP3. The results of this transaction is measured against the above defined metrics. The KPIs are measured and reported on a monthly basis.

IT Management Framework

This section contains a brief overview of what we define as management and how it relates to architecture and business practices. Its purpose is to show how things fit together.

FIGURE 2 shows the Enterprise Stack (E-Stack) reference model recently developed by SunPS[SM] Americas service. This figure shows how the business architecture environment, the physical execution architecture (middle box) and the M & O architecture environment relate to each other. Its main purpose is to show that both the business and physical architecture drive what happens in the IT environment.

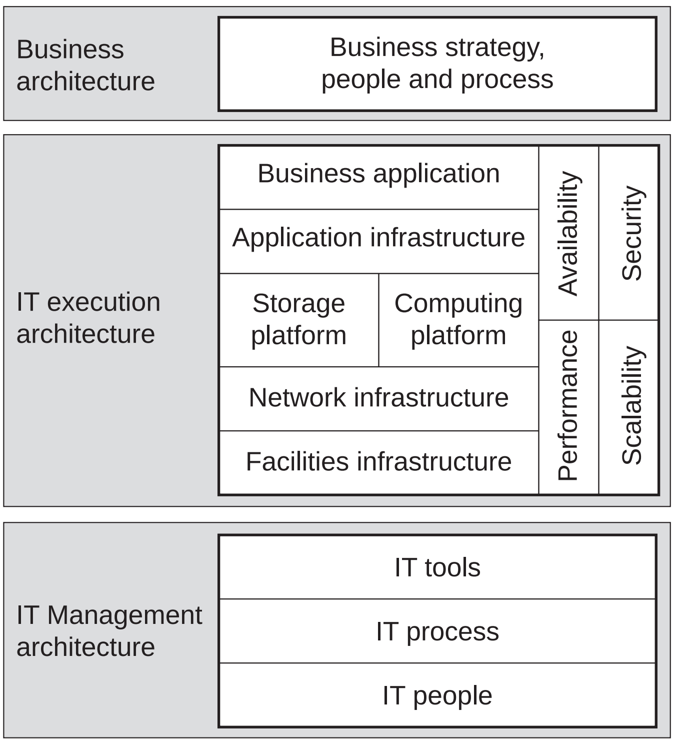

FIGURE 2 E-stack Reference Architecture

FIGURE 3 shows the details of each of the IT management architecture infrastructure.

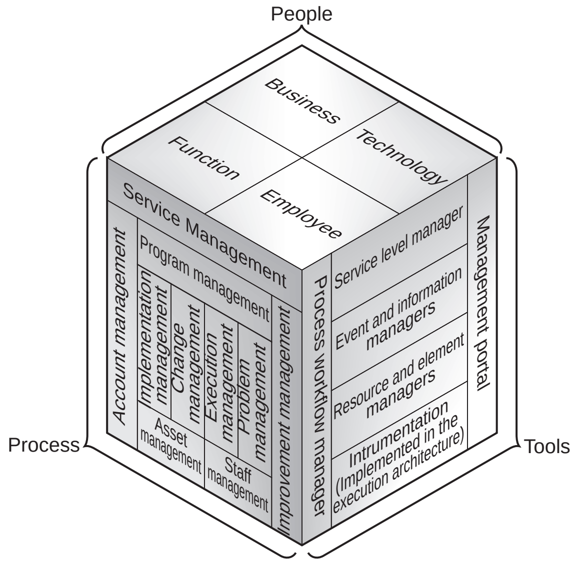

FIGURE 3 IT Management Architecture

The IT management infrastructure consists of three dimensions—people, process, and tools. The M&O architecture is based on this cube and focuses on the tools view of the IT management activities. The following paragraphs briefly describe each of the faces of this cube.

People

This dimension deals with all aspects of people management and organization. This cube shows four core aspects—business, function, technology and the employee. The M&O architecture does not intend to address this area. It is mentioned here to emphasize that it is a part of the complete IT management challenge.

Process

Good processes are essential to good IT management. This cube identifies the core processes. The M&O architecture tools are well equipped to support event management (as part of problem management) and SLA (as part of the account management process). The M&O architecture also implements some configuration management capabilities.

The M&O architecture design enables the deployment of any tool that supports any of the preceding processes as they affect the IDC infrastructure—at minimal overhead cost.

Tools

The major objective of this management architecture is to show a coherent set of tools. FIGURE 4 shows the structure of the deployed architecture.

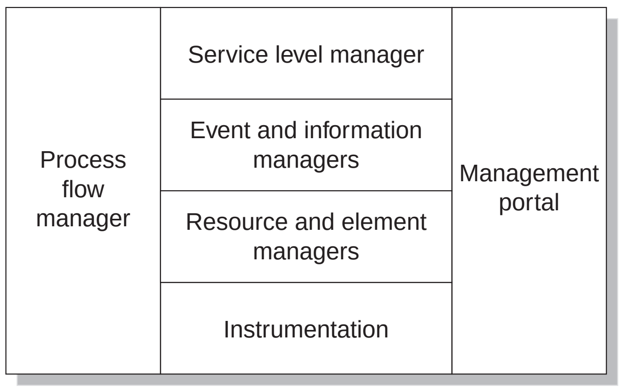

FIGURE 4 M&O Architecture Tools Dimension

The instrumentation layer is implemented as the agents that run on the monitored systems. The resource and element managers are the management servers that control the agents and the associated elements. This is where the primary automation activities between element and server take place. An example is the automated adjustment of swap space when a threshold is reached by the Sun™ Management Center (SunMC) software server process.

The event and information managers take input from the resource and element managers to correlate and consolidate the data into useful management information. This is also the conduit into the process flow management application.

The SLM (see“SLM Process” on page 3) operates at another level of abstraction. It tests the user experience and includes all components of the infrastructure that comprise the service. The developers defined the implementation requirements for this implementation in the SLA described earlier in this document.

The process flow manager facilitates the integration with the Process dimension of the cube. A good example is the Remedy (now Peregrin) tool that supports the Helpdesk processes. The management portal represents the interface to management reporting. This function is to display the information in the management data repository with any desired angle. In the deployed M&O infrastructure, the process flow manager and the management portal aspects have been included in the architectural considerations, but have not yet been implemented by tools.

Implementation Requirements

This section describes the additional requirements that went into the actual management and operations architecture. These requirements are in addition to the inherent requirements described in the previous sections.

Management At All Layers

TABLE 3 describes the aspects the M&O architecture manages per layer. Note that the layers follow the execution architecture cube as described in "IT Management Framework" on page 6. The execution architecture also implies that this requirement must be considered at all tiers (client through resource). The developers did not make this distinction because the complete managed architecture required the same visibility at all implemented tiers.

Facilities management is beyond the scope of this project. It is however an essential component of IT management.

TABLE 3 Managed Aspects By Layer

Layer	iForce Implementation	Fault	Configuration	Accounting	Performance	Security
Business application	IPlanet Message Server 5.10	Yes	Next phase	Next phase	Yes	Next phase
	Mail MultiPlexer (MMP)	Yes	Next phase	Yes	Next phase	Next phase
	Mail Transfer Agent (MTA)	Yes	Next phase	Yes	Next phase	Next phase
Application infrastructure	iPlanet Directory Server 4.13	Yes	Next phase	Yes	Next phase	Next phase
	DNS	Yes	Next phase	Yes	Next phase	Next phase
	Firewall	Yes	Next phase	Yes	Next phase	Next phase
	NTP	Yes	Next phase	Next phase	Next phase	Next phase
Computing and storage platform	Netra™ T1 server	Yes	Yes	Yes	Next phase	Next phase
	Netra 1405 server	Yes	Yes	Yes	Next phase	Next phase
	Sun Fire™ 6800 server	Yes	Yes	Yes	Next phase	Next phase
	Netra X1 server	Yes	Yes	Yes	Next phase	Next phase
	Sun StorEdge™ T3 array	Yes	Yes	Yes	Next phase	Next phase
	Sun Enteprise™ A1000 server	Yes	Yes	Yes	Next phase	Next phase
	Ancor Switches	Yes	Yes	Yes	Next phase	Next phase
	Sun™ Cluster 3.0 software	Yes	Yes	Yes	Next phase	Next phase
	Solaris™ OE	Yes	Yes	Yes	Next phase	Next phase
Facilities infrastructure	iForce lab	N/A	N/A	N/A	N/A	N/A

Yes: Management capability is included in current architecture.

Next phase: Management capability is to be included in a subsequent architecture.

N/A: Management capability is considered beyond the scope of this project.

Security management is planned for a subsequent phase. At this time the IDC Mail and Messaging Architecture facilitates security with a firewall complex at the entrance. A Security assessment will be scheduled to determine gaps and next steps. However, the tools currently deployed can facilitate security event management.

Accounting is currently under consideration because of its importance in an ISP/ ASP environment. However the scope of this effort is well beyond the current efforts and is still being defined.

The capacity planning aspect is an extension of performance management and affects multiple layers. It requires a complete process and the inclusion of variables that help anticipate future needs. The iFRC has done extensive sizing tests using the tools deployed in the M&O architecture. The results are published in the *Server Sizing Guide* related to the IDC Mail and Messaging Architecture project.

Performance Data Collection, Metrics and Thresholds

This section contains more details on the selected measurements and thresholds for performance tuning and capacity planning that are the main objectives of the IDC Mail and Messaging Reference Architecture project.

This section describes the baseline performance monitoring metrics as defined in the SunPS performance tuning and capacity planning methodology. Based on this information, the developers can identify and locate potential problems. In addition, provides the basic data to start the capacity planning process.

Every system's behavior depends on what application it is supporting. Therefore, you should do a detailed requirement analysis on a system and application pair basis. The information that follows, however, is a good baseline set of requirements.

The actually implementation of the concepts described in the preceding section was influenced by pragmatic constraints. The following are the major constraints considered.

- Cost
- Ease of deployment and availability
- Installation and configuration time

At all times, the important consideration was the ability of the tools to perform the required tasks. The following section lists those requirements.

Performance information is summarized in the following categories:

- CPU
- Disk

- Memory
- Network
- NFS (if applicable)
- Workloads (if possible)

The format and organization of information is identical for all six categories listed and presented in subsequent sections. Information for each category is tabulated under the three columns:

- Parameter

  Lists the parameters considered in the monitoring requirements for each of the six categories listed above.

- Description

  Describes the parameter.

- Expected Value

  Lists the acceptable value for each of the above parameters. If there is no threshold of acceptance for the parameter, this column will indicate "Relative or Informational".

Note – Some of the expected values listed may need adjustment based on individual characteristics. (For example, CPU utilization must be normalized for the number of CPUs or percent of disk full relative to the total size of available space.)

The following sections list the various requirements for performance monitoring

CPU Metrics

To determine system performance health, you should monitor the CPU parameters listed in TABLE 4. TABLE 4. lists the CPU-related metrics the monitoring tool must be able to collect.

TABLE 4 CPU Metrics

Parameter	Description	Expected Value
Percent CPU utilization	Total for all CPUs. Any utilization imbalance among CPUs identified.	<80 percent per CPU
User CPU	Percent CPU power spent on running user programs, libraries, and so forth should account for most of the CPU usage.	<90 percent
System CPU	Percent CPU power spent on executing system, kernel and administrative code (for example, device drivers, I/O handling and so forth).	<USR CPU
Run queue	Number of processes waiting to run on the CPU. UNIX uses the run queue to determine which process gets to use the CPU next. If the run queue exceeds two processes per processor, it may indicate a bottleneck.	< 2x CPU
Wait for I/O	Percent time CPU has to wait for disk to respond. High values could indicate a disk bottleneck (if disk busy and service times are high), otherwise could indicate a controller bottleneck.	<30 percent

References:

I/O Metrics

To determine system performance health, you should monitor the I/O related parameters listed in TABLE 5. TABLE 5 lists the I/O-related metrics the monitoring tool must be able to collect.

TABLE 5 I/O Metrics

Parameter	Description	Expected Value
Low activity disks	It is important to balance the load on disks. This list indicates disks with low or no activity that may be available for load balancing.	
Disk space used by files system	Indicates the file systems are running short of disk space.	<85 percent
Inode usage	Shows when space allocated for i-node entries is running short.	<20 percent
Percent busy (top 10)	Indicates the percent of time the disk is actually doing work. High values may indicate a disk or controller bottleneck.	<35 percent
Average service time	Indicates the time it takes the disk to complete a request. High values may indicate a disk or controller bottleneck. Note—Some lightly used disks may exhibit long service times. This is a well-known anomaly and should be taken into consideration during performance analysis.	<30ms
Queue length	Number of jobs waiting to be processed by the I/O system.	<1

References:

Memory Metrics

The memory metrics are divided into four subcategories:

- Paging
- Buffers
- Swap
- Kernel

Paging

Paging moves data or individual pages of a process between disk and memory. A high page-out rate (move to disk) could be due to heavy writing of data to disk and does not necessarily indicate a memory shortage. However, it is an important metric to collect. TABLE 6 lists the memory-paging metrics the monitoring tool must be able to collect.

TABLE 6 Memory Paging Metrics

Parameter	Description	Expected Value
Scan Rate	This parameter is used as a clear indication of memory shortage. A value higher than 320 per second may mean that the processes do not have enough memory in which to run.	<200 pages per second

References:

Buffers

In Solaris OE version 2 and above, cache buffers are used to cache inode, indirect block, and cylinder group information. A default value of percent of physical memory for buffers is generally considered too high for systems with large memory and can be reduced if the hit rates warrant it. TABLE 7 lists the buffer-related metrics te monitoring tool must be able to collect.

TABLE 7 Buffer Metrics

Parameter	Description	Expected Value
Percent Write Cache	System write percentage that is cached in buffers (instead of to disk)	>50 percent
Percent Read Cache	System read percentage that comes from cache buffers (instead of from disk).	100 percent

References:

Swap Areas

Swap areas should be distributed across many fast disks. Avoid placing them on disks used for OLTP databases. TABLE 8 lists the swap-related metrics the monitoring tool must be able to collect.

TABLE 8 Swap Area Metrics

Parameter	Parameter Description	Expected Value
Swap rate	Lack of memory can result in a whole process moving from memory to disk, called swap-out. This process should be very infrequent. Swap-ins indicate recalling of a swapped-out process, that is disk thrashing.	1 per day
Available swap	Low numbers indicate memory shortage and can cause processes to thrash to disk rather than perform their task.	32 Mbytes

References:

Kernel

This section lists the essential metrics to collect regarding the performance of processes in the kernel as they may indicate memory related issues. TABLE 9 lists the metrics the monitoring tool must be able to collect.

TABLE 9 Kernel Metrics

Parameter	Description	Expected Value
Memory failures	Memory failures indicate that permanent and huge kernel memory allocation failed. This metric is highly critical	0
File access	This category reflects the amount of activity spent in locating files through directory block reads, i-node searches and file system path searches. It is good for establishing baselines.	Relative

References:

Network Metrics

While these requirements do not focus on the network, Sun servers do provide some general statistics, derived from the network cards, which can indicate performance issues.

TABLE 10 lists the network metrics the monitoring tool must be able to collect.

TABLE 10

Parameter	Parameter Description	Expected Value
Collisions	Used as a measure of network congestion. Not relevant in switched segments. High values usually indicate mis-configured interfaces.	<15 percent
Errors in/out	This statistic usually indicates hardware/driver problems.	<0.02 percent
Bytes in/out	Used as a baseline.	Relative
No. of connections	(For example, `ftp`, `rlogin`, `telnet`). Used as a baseline.	Relative

References:

NFS Metrics

NFS is often the cause of performance issues. On systems that run NFS, the following parameters and values for this category should be monitored as they indicate potential performance issues. TABLE 11 lists the NFS metrics the monitoring tool must be able to collect.

TABLE 11 NFS Metrics

Parameter	Parameter Description	Expected Value
Calls	Used as a baseline.	Relative
Bad Calls	Used as a baseline.	Relative

References:

Workloads

This category of metrics characterizes workloads by the extent they use general resources of the server. Mainly, this category is used for capacity planning purposes but it can also be used to see where, from a business perspective, issues might come from. Defining a workload is a way of grouping resource usage to create a logical unit of work. For example, in one company the number of users in the sales department may be increasing by threefold, while marketing and finance, which also access the same server, may be expected to grow only twofold. In this case, you can define three workloads, each including users from a different department. In this way, resource usage by each department can be tracked and the appropriate factor for growth can be applied.

Or, it could be that a company that is running two applications on a single server is planning to increase the total number of users accessing application A by a factor of two and the total number of users accessing application B by a factor of three. To assess resource usage by each application, the processes associated with each application are included in a separate workload, and the appropriate growth factor can then be applied to each.

TABLE 12 lists the workload-related metrics the monitoring tool must be able to define and collect.

TABLE 12 Workload Metrics

Parameter	Description	Expected Value
Percent CPU per workload	Used as a baseline	Relative
Physical I/O per workload	Used as a baseline	Relative
No. of processes per workload	Used as a baseline.	Relative
Resident set size per workload	Memory occupied by each workload. Used as a baseline.	Relative

Tool Selection

This document described the SLM management concepts, implementation requirements and management metrics. TABLE 13 shows the tools that were selected as a result. The implementation of these will be discussed in the next article.

TABLE 13 Tools Distribution by Layer

Layer	iForce Implementation	Fault	Config	Accounting	Performance	Security
Business application	iPlanet Message Server 5.1	SunMC/ Micromuse	TBD	TBD	TeamQuest Micromuse	TBD
	MMP	SunMC/ Micromuse	TBD	TBD	TeamQuest Micromuse	TBD
	MTA	SunMC/ Micro-Muse	TBD	TBD	TeamQuest Micromuse	TBD
Application infrastructure	iPlanet Directory Server 4.13	Micromuse	TBD	TBD	TeamQuest Micromuse	TBD
	DNS	Micromuse	TBD	TBD	TeamQuest Micromuse	TBD
	Firewall	Micromuse	TBD	TBD	TeamQuest	TBD

TABLE 13 Tools Distribution by Layer

Layer	iForce Implementation	Fault	Config	Accounting	Performance	Security
	NTP	Micromuse	TBD	TBD	TeamQuest Micromuse	TBD
Computing and storage platform	Netra T1 server	SunMC	SunMC	TBD	TeamQuest	TBD
	Netra 1405 server	SunMC	SunMC	TBD	TeamQuest	TBD
	Sun Fire 6800 server	SunMC	SunMC	TBD	TeamQuest	TBD
	Netra X1 server	SunMC	SunMC	TBD	TeamQuest	TBD
	Sun StorEdge T3 Array	SunMC	SunMC	TBD	TeamQuest	TBD
	Sun Enterprise A1000	SunMC	N/A	TBD	TeamQuest	TBD
	Sun SAN Switches	SunMC	TBD	TBD	TeamQuest	TBD
	Sun Cluster 3.0 software	SunMC	SunMC	TBD	TeamQuest	TBD
	Solaris OE	SunMC	SunMC	TBD	TeamQuest	TBD
Network Infrastructure	Foundry NetIron	Foundry	Foundry	TBD	TeamQuest	TBD
	Foundry ServerIron	Foundry	Foundry	TBD	TeamQuest	TBD
	Foundry BigIron	Foundry	Foundry	TBD	TeamQuest	TBD
Facilities Infrastructure	iForce Lab	N/A	N/A	N/A	N/A	N/A