Service Level Management in the Data Center

By Edward Wustenhoff –Sun Professional

Sun BluePrints™ OnLine - March 2002

http://www.sun.com/blueprints

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

Part No.: 816-4528-10
Revision 04 March 2002

Please
Recycle

Adobe PostScript

Service Level Management in the Data Center

> "In today's world, where ideas are increasingly displacing the physical in the production of economic value, competition for reputation becomes a significant driving force, propelling our economy forward. Manufactured goods often can be evaluated before the completion of a transaction. Service providers, on the other hand, usually can offer only their reputations."

—Alan Greenspan, Chairman of the Board of Governors, Federal Reserve System

In today's service driven networks, the ability to deliver according to pre-defined agreements becomes increasingly a competitive requirement. Aside from being able to deliver highly available, reliably performing systems, just being able to deliver your promise is key to success. This is why an effective and efficient Service Level Management (SLM) system is important.

This article describes what SLM entails and what advantages it brings to Internet data center (IDC) providers. It also describes recommendations to successfully implement an SLM system.

An important element of the success of SLM is a sound Service Level Agreement (SLA). More details about this topic are discussed in the article *Service Level Agreement in the Data Center.*

Service Level Management

Ultimately, the goal of good SLM is to avoid what is quoted by `nextslm.org` as a widely anticipated pitfall:

> "The IT industry has a history of over-promising and under-delivering. This history repeats itself in the applications service provider (ASP) model, with promises of absolute (99.99999%) reliability, global availability and rock bottom costs."

"`nextslm.org`" (`http://www.nextslm.org/`) is an online learning community, dedicated to providing clear, concise answers about SLM. This community is sponsored by BMC, PriceWaterhouseCoopers, and Sun Microsystems.

SLM is the process that enables Internet data center (IDC) providers to deliver according to the SLAs that describe the expected performance of a service provider. It involves people, process and technology at all key areas of data center operations. The following paragraphs describe these areas in more detail.

Operational Aspects

FIGURE 1 is graphical overview of the operational management aspects commonly found in an IDC environment. This model is based on the Information Technology (IT) management framework as it is defined by Sun Professional Services of the Americas. The main objective is to show where SLM fits into the overall data center management challenge.

FIGURE 1 shows the organization of the key operational management aspects by function. The three faces of the cube show the pillars of a successful IT management model—people, process and tools. Within each face are more details of the contents.

People

This area includes all aspects of organization and human resources. It addresses all challenges in this space as they relate to the other faces.

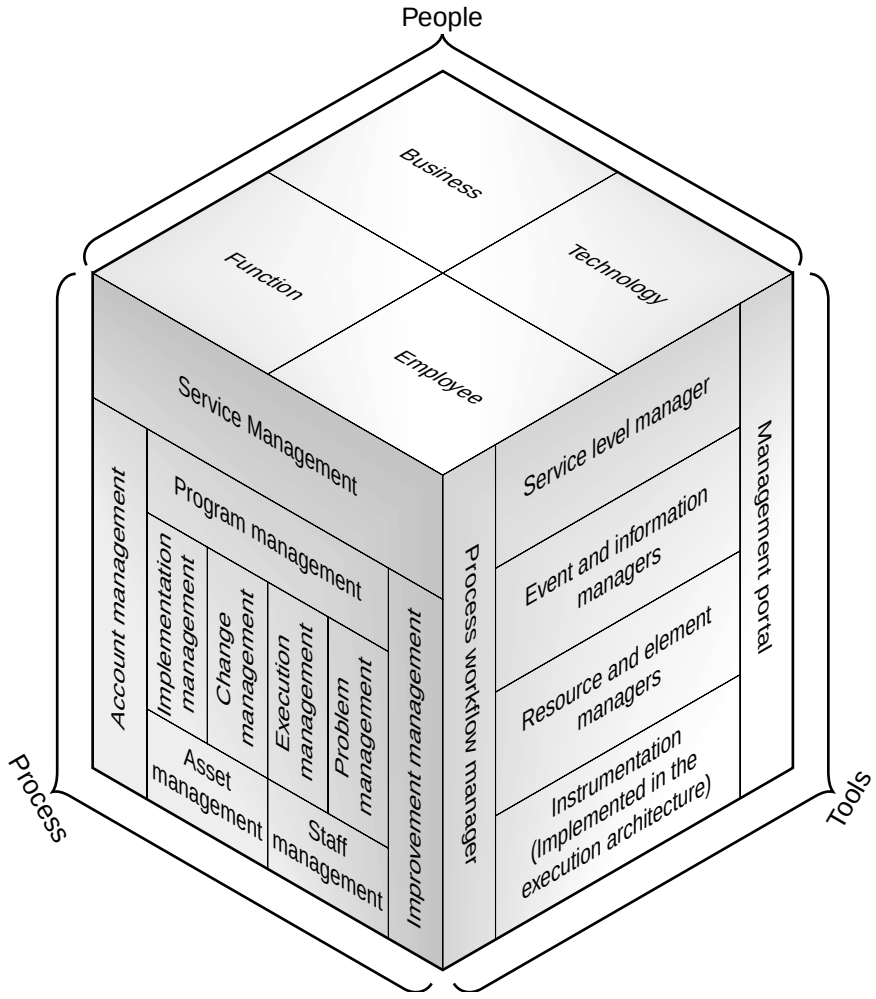

FIGURE 1 Overview of Operational Aspects in a Data Center Environment

Instrumentation

Process

This face shows the processes that are identified by the Sun Ready methodology as key to successful data center management. Somewhat counter intuitively, the SLM process is part of the Account Management process. This is because it is here that Service Performance is measured against customer expectations.

Tools

This face shows the main building blocks of a data center management tools framework. It follows a multitiered approach where agents reside on the managed components and integrate into higher layers of abstraction. Around it are the entry and exit points of the automation—a management portal to provide visibility into the information and a process workflow manager to facilitate the hand-off into the management processes face. The SLM is the highest level of abstraction in this model and refers to the tools that facilitate SLM.

For any category to function successfully, all three dimensions of people, process and technology must be addressed properly; a successful approach to SLM does this. This section focuses more on the technology aspects of the SLM process.

The IDC provider must deal with the following challenges while maintaining a high Quality of Service (QoS):

- Manage multiple network entry points provided by different vendors.
- Manage multiple, different infrastructure technologies like wireless, computer systems, databases and so forth with their associated vendors.
- Manage frequent changes in platform technology, functionality and scale.
- Manage frequent changes in service technology, functionality and scale.
- Provision over multiple service vendors to be able to manage and provide “single sign-on.”
- Manage geographically dispersed consumers and services.

The common thread between these challenges is “How do you keep your hands around all these different and dispersed components of your service?” A good approach to SLM provides the means to identify the issues and resolve them, while enabling a continuous quality improvement process.

The Process

The main purpose of SLM is to be able to measure the data center performance against pre-defined goals. These goals should be derived from the SLA between the consumer and the service provider. These goals, in turn, drive the service level objectives (SLOs) in the data center. In addition, it is important to be able to report to customers that promises were met. Having the information that shows the performance against the goals supports the continuous improvement process.

To facilitate the continuous improvement of quality, a good approach to the SLM system uses a circular process because it allows for adjustments as time progresses. The key challenge is to have a realistic starting point when a new service is

provided. Because the SLA must be able to be met by the infrastructure capabilities of the IDC, it is important to include the design of an SLM solution during the IDC architectural phase and revisit this design as changes to the architecture are incorporated.

Customers expect the goals, for example, transaction response times and so forth, to be addressed in the SLA. Managing against an SLO also forces the system to be able to report on performance on higher layers, which is what customers experience typically. Managing against an SLO also forces IDC providers to manage components that are not under their control. A third party service provider is a good example, but the network components in the Internet that separate the consumer from the service provider and are largely unknown entities for the service provider are examples, too. The SLM process helps determine where the boundaries are and what can be expected when these boundaries are crossed.

FIGURE 2 is a graphical representation of a SLM process that focuses on the relationship between the SLA and the performance metrics collected. This example is specifically designed to facilitate the tools selection process.

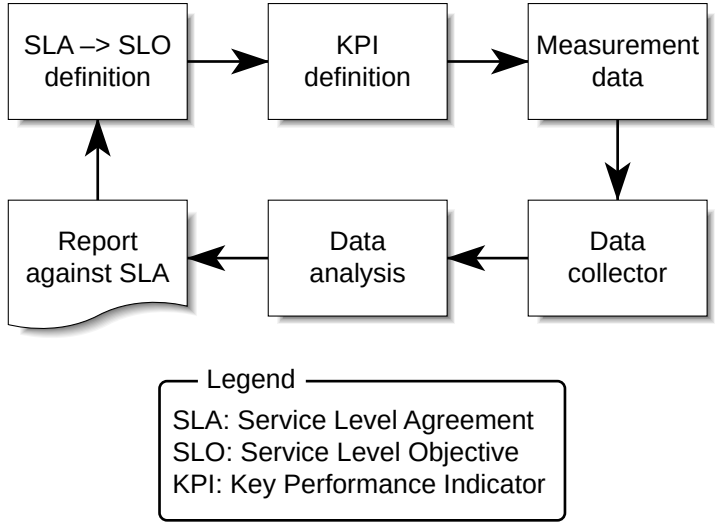

FIGURE 2 SLM Process

The major change is the decision to have SLAs drive the operational measurements of success. Often the service levels delivered are derived from the capabilities of the infrastructure. The business case should drive the QoS, which in turn drives the infrastructure requirements. FIGURE 2 illustrates this point.

The Technology

Aside from providing the proper information to compare performance against the SLOs, the SLM solution must also include information and/or tools that facilitate troubleshooting, capacity planning and modeling. Modeling is very important in the frequently changing Internet age because time to market is paramount; the implementation of *new* technologies and services does not have time for trending solutions. Troubleshooting and trending solutions are commonly available in the more traditional system and network management tools. A good modeler is often hard to find.

To understand some of the technology challenges better, the following paragraphs provide a more technical and detailed description of managed components in a data center infrastructure.

The technology components in a service provider's data center architecture consist of five platform layers as defined by the SunPS℠ 3DM program and somewhat expanded to reflect a data center environment more than a development environment:

- Business application or service
- Application infrastructure (application servers, Lightweight Directory Access Protocol (LDAP), relational data base management system (RDBMS) and so forth.)
- Computing and storage platform (Sun Fire™ 15K server, Solaris™ operating environment, disks, storage area networks (SANs) and so forth).
- Network infrastructure
- Facilities infrastructure (air conditioner, power, access and so forth)

To instrument this infrastructure, small programs that monitor specific components to collect information and even sometimes act on certain events must be inserted. These *agents* or *element managers* run at each platform layer. On network devices we often see remote monitoring (RMON) probes and Simple Network Management Protocol (SNMP) Mibs. On systems, we see vendor specific agents like the Sun™ Management Center software that not only monitor, but often also help manage the system. On the application infrastructure layer, we see application agents like BMC's Knowledge Modules and sometimes custom agents (for custom applications) and, finally, at the services layer, we find aggregator tools that can often test the health of the service and correlate events from the other layers. An example of such a tool is MicroMuse's Netcool, but we often find custom agents too. All of the information is typically aggregated at different stages, often first locally then network wide, and ultimately service wide. A network operations center (NOC) monitors all of the agents and facilitates Help desk and customer care activities in addition to its normal operational tasks.

FIGURE 3 is a graphical representation of such an SLM infrastructure.

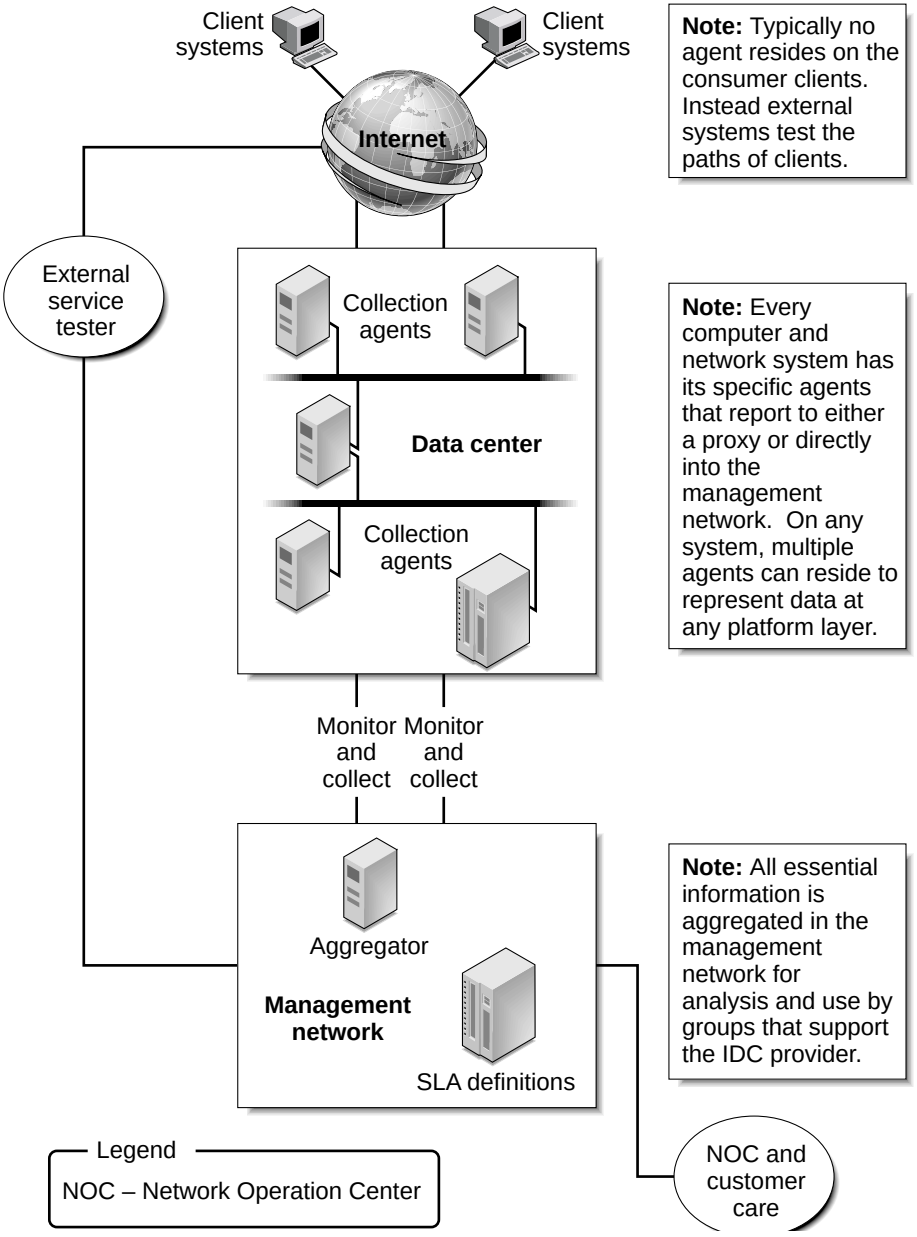

FIGURE 3 SLM Infrastructure

The SLA definitions play an integral part in the process. Every monitor agent (network, system and application layer) measures itself against metrics derived from the SLOs to determine whether they are working within the correct parameters.

An example is an SLA requirement that states that a certain transaction can take no longer than 10 seconds before its result is presented to the consumer. Consequently, it has been established that the complete path cannot take longer than 10 seconds. This requirement drives the service level tester to test for this behavior, the application monitor to ensure it gets a return from the system within the time limits that allow the SLO to be met, and the system monitor to guarantee that all related I/O responses from the database happen in time, and that the network can transport all data from the back end system to the consumer at sufficient speed.

FIGURE 3 shows an example of how these elements determine how much time can be spent per area that drives what thresholds are set and what metrics are collected.

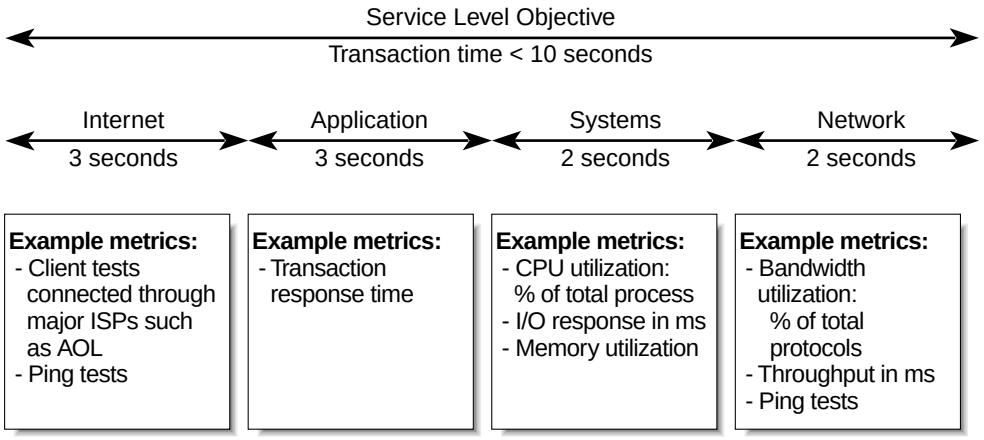

FIGURE 4 Time Scale of Thresholds and Metrics of a Service Level Objective

With this picture of what is involved in managing service levels we can consider the benefits of making the investment to set up all of this infrastructure.

Benefits of Service Level Management

As Alan Greenspan pointed out in the introductory quote, the challenge service providers face is the fact that their products cannot be touched and tested extensively before one buys in. The only thing they can offer beforehand is their reputations, which is directly related to the QoS that is delivered to existing consumers. To manage QoS, an SLM system is crucial.

SLM Provides Customer Focus and Management

Reputation is based entirely on the perception of others. Psychology 1.1 advises us that the first impression is the benchmark for all future experiences in the human mind. It is, therefore, extremely important for Internet service providers (ISPs) to set the proper expectations from the beginning and measure against their ability to deliver on these expectations. An SLM system enables that by continuously monitoring and improving the QoS at all layers.

An SLM system facilitates the focus and direction of internal teams by forcing the different groups (such as the network and system specialists) to prove they conform to the requirements of their internal SLA, which is directly derived from the requirements set forth in the customer's SLA. Both the Help desk team and the network operations team are measured because they affect the customer's perception. Therefore, SLM enables the data center to do what the following sections describe.

SLM Continually Improves RAS

RAS means reliability, availability and serviceability. The ability to show statistics that prove the performance in each of these areas allows the IDC to adjust (and preferably raise) the bar on a regular basis to their ISPs.

By adjusting SLAs to reflect higher standards (that are either made possible due to technology enhancements or are demanded by market dynamics), continuous improvement is driven into the service as a natural progression.

The reports that are generated at the systems, network and applications layers show the responsible vendors or technology teams whether they are meeting their SLOs. This information drives better results by accountability. Better results from vendors and internal teams mean a stronger service offering. These reports drive all parties involved to strive for higher quality standards at lower prices.

SLM is an Early Warning System

The collected information about the infrastructure enables the service provider to see trends and adapt accordingly.

Having this type of information improves the probability of success when introducing new services. By using this information with a modeling tool and combining it with business trends and marketing plans, you can make some predictions around future volumes and load.

The use of a modeling tool is very compelling. However, there are some inherent challenges with this approach. It requires a high level of expertise and knowledge about the modeled technology. It is, therefore, an expensive proposition. Success is not guaranteed because of the number of variables. One mistake can have a great impact on the result. The “garbage-in, garbage-out” paradigm is very applicable here.

Our experience advises us to use a modeling tool with the vendor’s guidance to predict what systems should be ordered to accommodate the changes in usage. These systems should be set up in a test laboratory and rigorously tested using known usage patterns. This testing improves the probability of getting appropriately sized systems in the laboratory and minimizes the costs due to corrections of anticipated load.

When security statements are part of an SLA, the process is monitored and measured; therefore, security attacks can be uncovered early.

Recommendations

Our experience shows that a good approach is to start with conservative promises in the SLA that have a high probability of being met by the initial infrastructure. This approach allows time to tune the SLM.

The technical manager must consider the design of the SLM system during the planning and architecture phases. Overhead, network availability, agent definition and so forth can only be done effectively during the early stages. If you do it later, it becomes a bolt-on solution with its inherent agility issues.

We consider 5 to 10 percent CPU utilization overhead acceptable for a robust SLM solution. Naturally, business requirements will influence the actual results. Make sure the infrastructure capacity planning sessions always include the SLM overhead.

Our experience shows that if you need more than four or five tools to build the SLM infrastructure, you should seriously revisit the tool selection. The biggest risk associated with a SLM system is over-instrumenting the infrastructure. Typically there is one tool per layer—network, system and application. Frequently the latter has custom agents because of the uniqueness of most applications. In addition, an aggregator tool should collect cross-layer statistics and correlate the information between SLA definitions and agent information. Each tool should have a modeler/trend analysis and reporting component.

For security, availability and performance reasons, it is imperative that the SLM system have its own network infrastructure. As a rule of thumb, no public traffic should be mixed with private traffic. Aside from the security implications, performance can be affected by data collection and analysis. In addition, it is good to have an alternate path to the system in case the primary path to service is down.

For similar reasons, all systems should write log information to a different, separate server or servers. This improves the ability to secure them because, in that case, a potential hacker has to be able to break the security of one additional server to erase his traces. It also will help off-load SLM overhead when log analysis is being done.

Make sure that the key performance indicators (KPIs) in the SLA can be translated easily into measurable metrics of components. This translation is done more easily by being as specific as possible in the definition of the SLA. More details can be found in the article on SLAs.

When defining the SLA metrics per component in the IDC, it is important to realize that the availability of the service is the multiplication of the availability of all components that comprise the service (for example 99.9 percent x 99.9 percent = 99.800 percent). This means that each component has to have a higher availability than defined in the SLA for the overall service.

Most components have no documented or published availability numbers; therefore the only way to obtain these numbers is by actually measuring availability after the fact. However, for most hardware some statistics are available under Non Disclosure. These numbers can be used as a guideline for the initial availability commitments.

TABLE 1 lists the aspects that should be considered when looking for management tools. Each aspect needs to be considered at all four layers—the lower platform (facilities, network, systems, storage and so forth), the upper platform (application infrastructure services like LDAP, domain name service (DNS), network file system (NFS) and the actual service or application layer.

Depending on the requirements, some aspects are more relevant than others. It is based on the ISO defined fault, configuration, accounting, performance and security (FCAPS) model and encompasses the operational management aspects.

Using TABLE 1 on page 12, you can assess what aspects at what layers should be addressed based on the known business requirements.

TABLE 1 Management Tool Aspects

	Fault	Configuration	Accounting	Performance	Security
Service/Business application					
Application Infrastructure					
Compute and Storage platform					
Network					
Facilities					

Summary

This article has defined and examined the SLM process. A carefully defined SLM solution is critical to be able to measure a data center infrastructure against SLAs and their associated SLOs.

An SLM system touches on all aspects necessary to organize, optimize, protect and control a data center infrastructure.

The SLM process includes the gathering of metrics regarding all aspects (FCAPS) at all platform layers to evaluate compliance and to facilitate the continuous improvement process. SLM is driven by the SLA.

By following our recommendations, the most common mistakes can be avoided.

A good approach to SLM will avoid what is quoted by `nextslm.org` as a widely anticipated pitfall—"*The IT industry has a history of over-promising and under-delivering. This history repeats itself in the ASP model, with promises of absolute (99.99999%) reliability, global availability and rock bottom costs.*"—by setting realistic expectations, defined in mutually agreed upon agreements, and providing the process to show compliance and improvement when necessary.

Author's Biography

EDWARD WUSTENHOFF
Chief IT Consultant
Sun Professional Services

Edward has a total of 16 years experience in networked computer systems and data center management, including the latest internet technologies. The past 7 years were at Sun where he has been exposed to most Sun products and technologies.

Edward is currently a Chief IT Consultant in Sun Professional Services of the America's at Sun Microsystems. In one of his previous roles at Sun he managed the Enterprise Management Practice where he advised SUN's customers about best practices, tools selection and deployment strategies.

Previous projects included architecture and design of networks and computer systems, in addition to assessing customer environments and suggesting improvements.

Last year he was very active in wireless portal architecture and IDC management.