Service Level
Agreement in the Data Center

By Edward Wustenhoff –Sun Professional Services

Sun BluePrints™ OnLine - April 2002

Please
Recycle

Adobe PostScript™

Building a Service Level Agreement in the Data Center

As we discussed in the *Service Level Management in the Data Center* article, the ability to deliver according to pre-defined agreements increasingly becomes 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. Key to the success of such a system is a sound Service Level Agreement (SLA).

This article describes what role an SLA plays in the Internet Data Center (IDC) and how it helps assure that one's reputation stays intact. It also includes sample agreements that can be used as templates.

Due to time sensitivity, the necessary localization and customization legal aspects are not included in this article.

Service Level Agreements

A good SLA helps the IDC promise what is possible to deliver and deliver what is promised.

In this article, we will establish what an SLA is and provide two sample agreements and one example of how an agreement can map to key performance indicators. A section on why we believe SLAs are so important and what we believe are the essential benefits of managing against one follows these sections.

What a Service Level Agreement Is

An SLA sets the expectations between the consumer and provider. It helps define the relationship between the two parties. It is the cornerstone of how the service provider sets and maintains commitments to the service consumer.

A good SLA addresses five key aspects:

- What the provider is promising.
- How the provider will deliver on those promises.
- Who will measure delivery, and how.
- What happens if the provider fails to deliver as promised.
- How the SLA will change over time.

In the definition of an SLA, realistic and measurable commitments are important. Performing as promised is important, but swift and well communicated resolution of issues is even more important.

The challenge for a new service and its associated SLA is that there is a direct relationship between the architecture and what the maximum levels of availability are. Thus, an SLA cannot be created in a vacuum. An SLA must be defined with the infrastructure in mind.

An exponential relationship exists between the levels of availability and the related cost. Some customers need higher levels of availability and are willing to pay more. Therefore, having different SLAs with different associated costs is a common approach.

The following section contains an example template of an SLA to show all of the important components one wants to address in such a document.

The SLA example in TABLE 1 is based on the short form template that is available from nextslm.org (http://www.nextslm.org/). An online learning community, nextslm.org is dedicated to providing clear, concise answers about SLA. They are sponsored by BMC, PriceWaterhouseCoopers and Sun Microsystems, Inc.

This template is customized and augmented to reflect a service provider model. By replacing the *italicized* text with specific service aspects the template can be customized to reflect a specific service offer.

TABLE 1   SLA Template

The *insert service name* is used by *insert customer name* to *insert description of the service capability.* The Internet Service Provider (ISP*)* guarantees that:
The *service name* will be available *insert percentage* of the time from *insert normal hours of operation* including hours and days of the week. Any individual outage in excess of *insert time period* or sum of outages exceeding *insert time period* per month will constitute a violation.
Insert percentage of *service name transactions* will exhibit *insert value* seconds or less response time, defined as the interval from the time the user sends a transaction to the time a visual confirmation of transaction completion is received. Missing the metric for business transactions measured over any business week will constitute a violation.
The IDC Customer Care team will respond to service incidents that affect multiple users (typically more than 10) within *insert time period,* resolve the problem within *insert time period,* and update status every *insert time period.* Missing any of these metrics on an incident will constitute a violation.
The IDC Customer Care team will respond to service incidents that affect individual users within *insert time period,* resolve the problem within *insert time period,* and update status every *insert time period.* Missing any of these metrics on an incident will constitute a violation.
The IDC Customer Care team will respond to non-critical inquiries within *insert time period,* deliver an answer within *insert time period,* and update status every *insert time period.* Missing any of these metrics on an incident will constitute a violation. A non-critical inquiry is defined as a request for information that has no impact on the service quality if not answered or acted upon promptly.
The external availability measurements are done by *insert test company name* and reported on a monthly basis to *insert customer name.* The internal processes are measured and reported by the ISP to the *insert customer name* on a monthly basis. This service includes incident reporting.

TABLE 2 shows the number of violations and associated penalty on a monthly basis.

TABLE 2   Monthly Violations and Associated Penalties

Number of violations	Penalty
1>5	*Insert penalty.* Typically a reduction in fees.
5>10	*Insert penalty.* Typically a reduction in fees plus some additional compensation and a corrective action plan.
10>	*Insert penalty.* Typically a reduction in fees plus some additional compensation and a corrective action plan.

As services and technologies change, the SLA may change to reflect the improvements and/or changes. This SLA will be reviewed every six months and updated as necessary. When updates are deemed necessary, the *customer* will be asked to review and approve the changes.

Other areas that must be defined in an SLA are details on how the measurements are done, what usage limitation the service has with regard to number of concurrent users and so forth, and details on how and who receives reports and how conflicts are arbitrated. Because these topics are unique in each contract, they are not included in the preceding example.

This SLA is a "short form" SLA to illustrate essential aspects between a consumer and provider in an ASP context. Internal SLAs, between operations support groups in the IDC, for example, are different and often contain more details and specifications.

The main reason for this difference is that internal SLAs are driven by budget constraints and the business management's view of IT, while external SLAs are driven by revenue, cost and earnings.

The following template provides a general description of an internal SLA as well as the owners approval and review process, and a definition of the terms used in the document. It is another example of an internal SLA from nextslm.org.

TABLE 3   Internal SLA Template

1.0 Statement of Intent	This section states the objectives of the document.
1.1 Approvals	All parties must agree on the SLA. This section contains a list of who approved the SLA.
1.2 Review Dates	This section contains the track record of the SLA reviews.
1.3 Time and Percent Conventions	This section contains the descriptions of what time conventions and metrics are being used.
2.0 About the Service	This section introduces the service addressed by this SLA.
2.1 Description	This section describes the service in detail.
2.2 User Environment	This section describes the architecture and technologies that are used by the consumers of the service.
3.0 About Service Availability	This section introduces the availability concepts used in this SLA.
3.1 Normal Service Availability Schedule	This section describes what is considered *normal* service availability.
3.2 Scheduled Events That Impact Service Availability	This section describes what scheduled outages are to be expected,

TABLE 3   Internal SLA Template *(Continued)*

3.3 Non-emergency Enhancements	This section describes the process that inserts enhancements into the infrastructure.
3.4 Change Process	This section describes the complete process of how changes are introduced in the service., including the associated availability impact.
3.5 Requests for New Users	This section describes the provisioning process of new users/customers.
4.0 About Service Measures	This section contains a detailed description of how the service availability is measured and reported.

How an SLA Maps to Key Performance Indicators

TABLE 4 uses the first example SLA and shows, in the right column, what key performance indicators result from the stated commitment. These indicators. in turn, drive what performance data and metrics are collected by the SLM process. A separate article regarding SLM describes how this information interacts with the SLM system. The performance indicators in TABLE 4 drive the internal SLAs and their associated metrics.

FIGURE 1 shows how different internal groups (networking, systems and applications) must commit to certain transaction response times to achieve the promised levels of service to the consumer.

TABLE 4   SLA Key Performance Indicators

| Commitment	Key Performance Indicator
The *service name* will be available *insert percentage* of the time *from insert normal hours of operation* including hours and days of the week. Any individual outage in excess *of insert time period* or sum of outages exceeding *insert time period* per month constitutes a violation.	Service Availability as a Percentage of Normal Business hours. Note–We must measure overall service availability. The maximum threshold is the maximum outage per incident and/or total sum of outage per month.
Insert percentage of *service name transactions* will exhibit *insert value* seconds or less response time, defined as the interval from the time the user sends a transaction to the time a visual confirmation of transaction completion is received. Missing the metric for business transactions measured over any business week constitutes a violation.	Percentage of transaction response times more than x seconds. Note–We must measure transaction times against a threshold of x seconds and measure the number of slower transactions as a percentage of the total.
The IDC Customer Care team will respond to service incidents that affect multiple users within *insert time period*, resolve the problem within *insert time period*, and update status every *insert time period*. Missing any of these metrics on an incident constitutes a violation.	Service Incident (affecting multiple users) Response times, resolution times and status updates. Note–We must be able to create incident reports that track actions with timestamps. These are measured against the time thresholds.
The IDC Customer Care team will respond to service incidents that affect individual users within *insert time period*, resolve the problem within *insert time period*, and update status every *insert time period*. Missing any of these metrics on an incident constitutes a violation.	Service Incident (affecting single users) Response times, resolution times and status updates. Note–We must be able to create incident reports that track actions with timestamps. These are measured against the time thresholds.
The IDC Customer Care team will respond to non-critical inquiries within *insert time period*, deliver an answer within *insert time period*, and update status every *insert time period*. Missing any of these metrics on an incident constitutes a violation.	Inquiry response times, answer times and status updates. Note–We must be able to create incident reports that track actions with timestamps. These are measured against the time thresholds.
The external availability measurements are done by *insert test company name* and reported on a monthly basis to *customer name*. The internal processes are measured and reported by the ISP to the *customer name* on a monthly basis.	This is an example of an external management service that is managed by its own SLA between the IDC provider and the external ISP that supports this commitment to the service consumer.
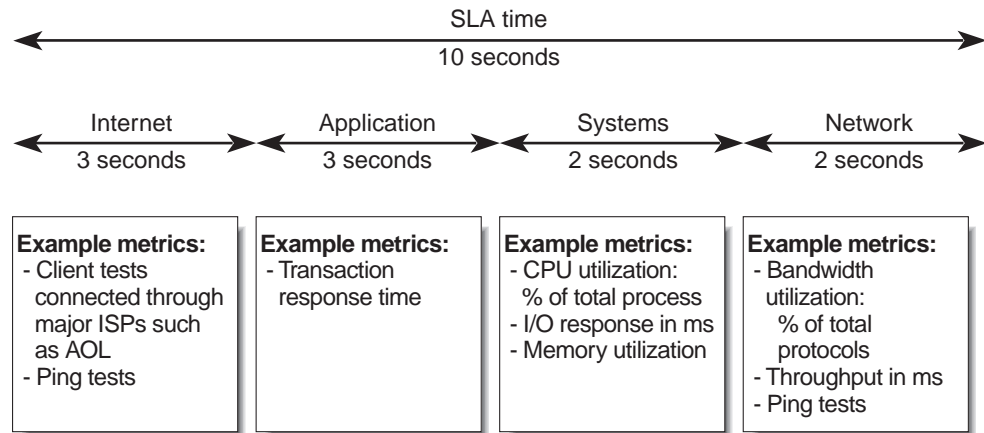
FIGURE 1   Transaction Response Times for Promised Service Levels

Why a Service Level Agreement is Important

A good SLA is important because it sets boundaries and expectations for the following aspects of data center service provisioning.

■ Customer commitments. Clearly defined promises reduce the chances of disappointing a customer. These promises also help to stay focused on customer requirements and assure that the internal processes follow the right direction.

■ Key performance indicators for the customer service. By having these indicators established, it is easy to understand how they can be integrated in a quality improvement process (like Six Sigma). By doing so, improved customer satisfaction stays a clear objective.

■ Key performance indicators for the internal organizations. An SLA drives internal processes by setting a clear, measurable standard of performance. Consequently, internal objectives become clearer and easier to measure.

■ The price of non-conformance. If the SLA has penalties (something that many IDC providers prefer to avoid but should not) non-performance can be costly. However, by having penalties defined, the customer understands that the IDC provider truly believes in its ability to achieve the set performance levels. It makes the relationship clear and positive.

Benefits of Measuring Against an SLA

This chapter describes why measuring against SLAs is so important to a successful data center service and a healthy infrastructure.

Key to success for a data center service is the ability to perform according to predefined standards. Performance is summarized as the ability to meet (or preferably exceed) the customer's expectations. The first step is to set realistic expectations for both sides of the contract. An SLA is a great vehicle to communicate the expectations and create a level of trust by adding conditions and penalties when the promises are not met.

The SLA defines a clear relationship between the customer and the provider by setting boundaries, conditions, penalties and expectations.

Because an SLA links the customer requirements to infrastructure requirements, it creates the ability to link service levels to service cost and, as a result, profitable pricing can be set. Moreover, by spending wisely on well defined requirements rather than rules of thumb or gut feeling, more efficient cost management can be achieved.

The capability to segment service offerings with different pricing for different service levels benefits both the vendor and the customer. The vendor widens its target market by being able to customize its services and the customer only pays for what it needs.

Having the ability to measure against key performance indicators facilitates the continuous quality improvement process. Being able to raise the performance bar at a steady pace helps the data center service provider to remain competitive. By tying the problem resolution process to an SLA, a service performance problem becomes an opportunity to structurally improve overall service quality and customer satisfaction, as opposed to just resolving the symptoms of the real issue.

An SLA sets the standards to which the IDC service provider committed. As a result, a set of common and/or global parameters is derived to which all organizational groups must be managed and measured. Requirements for success of the business are now directly translated to measurable requirements for the technical teams.

The SLA drives the definition of Key Performance Indicators (KPIs) at the service, application, system and network level. Defining these KPIs facilitates the proper tools selection, process definitions and skills (people, process and technology) for an organization.

Knowing what to measure eliminates redundant data collection, which reduces the total overhead of the SLM system on the service infrastructure. A good SLM system collects data only once.

Recommendations

The following is a set of suggestions based on our experience in writing and managing against SLAs.

1. Keep the agreements simple, measurable and realistic.

   This technique creates clarity and keeps the process simpler. Just think of all the hours attorneys will charge to unfold a complex agreement. Being measurable also improves the ability to perform according to the agreement. Realistic goals can be achieved and the penalties can be limited.

2. Bring business managers and technology managers together during the definition of the customer SLAs.

   By doing so, promises made in the SLA will be able to be met by the underlying technology. The challenge is that the initial SLA should be the basis for internal technology choices, and the underlying technology should help define the limitations of the initial promises in the SLA. This is a "catch 22". The best way to break the catch is to ensure that technology and business managers come together.

3. Map the commitments to KPIs and have these indicators drive the performance metrics of the business units.

Summary

In this article, we explained what an SLA should look like and how it supports the SLM process.

We also gave examples to show the key elements in an SLA:

- What the provider is promising.
- How the provider will deliver on those promises.
- Who will measure delivery, and how.
- What happens if the provider fails to deliver as promised.
- How the SLA will change over time

We have shown two templates to illustrate how these principles can be translated in real examples. Remember that every business is unique and that one should be flexible in applying these principles.

We mapped the commitments in an SLA to KPIs to help facilitate the performance metrics definition. These KPIs drive the internal goals and even influence the tool selection of the SLM system as explained in the *Service Level Management in the Data Center* article.

We have described the importance of SLAs to avoid the most common pitfall as quoted by `nextslm.org`: *"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."*

The benefits of a good SLA are many, but the most important ones are:

- Sets clear customer relationships
- Sets goals for internal organizations
- Sets a framework for continuous quality improvement
- Drives the SLM system architecture

We ended with some suggestions to assure the creation of good SLAs. The most important one is to keep the SLA simple, measurable and realistic.

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 seven years were at Sun where he became familiar with most Sun products and technologies.

Edward is currently a Chief IT Consultant in Sun Professional Services of the America's at Sun Microsystems, Inc. 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.

Currently, Edward is very active in wireless portal architecture and IDC management.