Taking Magic and Mystery Out of End-to-End Service Levels

IS organizations must address business units' concerns with end-to-end availability and performance of applications used in business processes. This requires an expanded view from resource-specific SLAs to end-to-end SLAs.

Core Topics
Enterprise Management: Infrastructure and Application Management

Business Management of IT: Service Management Strategies

Key Issues
How will IS organizations use network and systems management to achieve enterprisewide service-level objectives?

How can enterprises achieve high availability and continuous operations for business-critical applications?

Strategic Planning Assumption
By 2005, 75 percent of large enterprises will be generating service-level reports that correlate business performance with end-to-end infrastructure service levels (0.7 probability).

Note 1
End-to-End Service Definition
The end-to-end service to be measured is typically defined jointly by the business units and the IS organization as part of the service-level management process. It includes a set of applications and underlying IT infrastructure components (e.g., network, server, OS, middleware, database and storage; whether insourced or outsourced) that are critical to a business process. Examples include all applications and infrastructure associated with an e-commerce service, a call center service, an enterprise resource planning service or a financial accounting service.

Today, most business processes are enabled by the IS organization using a set of applications on top of the IT infrastructure of network, server and storage components. To ensure that business process objectives are understood and achieved, the business unit and the IS organization should jointly define service-level agreements (SLAs). SLAs must be based on metrics that succeed in providing business value — taking the mystery out of technology details — while being able to be realistically measured by the IS organization — taking the magic out of monitoring.

Negotiating, documenting and communicating these SLAs foster greater business/IS alignment by setting expectations and directing people and technical resources toward meeting those expectations. Once metrics are defined, the IS organization is responsible for performing regular measurements and reporting service-level results to the business units.
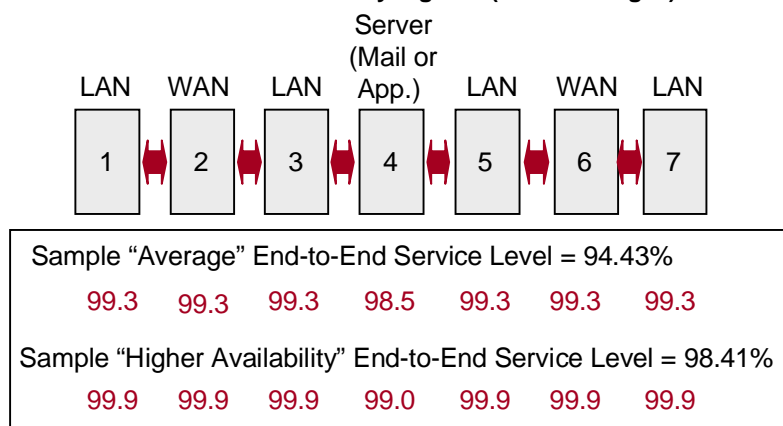
Measuring SLAs

Less than 20 percent of large enterprises measure end-to-end application service levels. Often, IS organizations set technical or resource-specific (e.g., network or server) availability objectives by which IS personnel are measured. Increasingly, IS managers are being pushed to measure business-centric SLAs for end-to-end services (see Note 1) that reflect end-user experiences, including service availability and performance. Yet, IS organizations have been hesitant to agree to such SLAs for distributed applications because of the difficulty in measuring end-to-end results. End-to-end service levels can be measured from the bottom up, summarizing individual IT component availability levels. They can also be measured from the top down, starting with end-to-end transaction response times.

Gartner

Bottom-Up SLA Measurement

To calculate end-to-end service availability from the bottom up —
that is, from the IT components that make up that service —
multiply the availability percentages of each of the components
together. For example, if each of seven components in an IT
service (see Figure 1) has an excellent 99.9 percent availability
level, multiplying 99.9 percent by 99.9 percent seven times yields
a less-than-excellent end-to-end service availability of just 98.41
percent — or more than 11 hours of downtime per month.
However, this calculation does not take into account when
redundant network links or backup servers are used to reduce or
negate the service impact of individual component downtime.

Figure 1
End-to-End Availability Figures (in Percentages)

LAN	WAN	LAN	Server (Mail or App.)	LAN	WAN	LAN
1	2	3	4	5	6	7

Sample "Average" End-to-End Service Level = 94.43%

99.3	99.3	99.3	98.5	99.3	99.3	99.3

Sample "Higher Availability" End-to-End Service Level = 98.41%

99.9	99.9	99.9	99.0	99.9	99.9	99.9

Source: Gartner Research

Top-Down SLA Measurement

Measuring end-to-end transaction response times from the top
down provides an alternative method of calculating availability
and performance, which more realistically portrays the end-user
experience. If the transaction goes through, the service is clearly
available, whether it is transiting primary or backup resources.
The response time provides the performance indicator, which
can be compared with previous results to show improvements or
detect service degradation. Real end-user transactions can be
monitored by instrumenting the end-user's system, or synthetic
transactions can be initiated that mimic the end-user experience
(see "End-to-End Application Response Times: Market Update,"
M-12-2080). Although top-down monitoring can be used to
communicate business-relevant service levels, it does not
mitigate the need for individual components to be monitored for
troubleshooting and problem resolution. It is the correlation of
these end-to-end service metrics with the individual components
that begins relating the IT infrastructure with the business
process.

Reporting SLAs

Service-level reporting will continue to be largely a manual process, supported with some automated tools. The reason for the large manual effort is primarily for impact analysis. Because an application is down does not mean that the business process is totally down or that end users are completely unproductive. The manual measure of downtime impact on end-user productivity requires input from the user base. The manual measure of downtime impact on business process revenue requires input from the business owner. By understanding people and business impact, downtime costs can be ascertained (see "Measuring End-to-End Application Service Availability," DF-13-1114).

Setting Targets to Achieve SLAs

As more business units begin to set SLAs with IS organizations, these measures must "trickle down" to IT performance objectives that correlate with the business SLAs. It is vital that IS organizations identify realistic targets internally and sum them up to ensure that business unit/IS organization SLAs are achievable. See Note 2 for examples of technical targets and business-relevant end-to-end service levels.

IS organizations must also factor in the performance targets of outsourced services when developing SLAs. Currently, many IS organizations selectively use outsourced services (e.g., WANs), yet the business requires overall end-to-end confidence in meeting availability and performance goals. As a result, it is the responsibility of the IS organization to manage the outsourced service relationship, including negotiating external SLAs to meet objectives. With many different components supported by various providers, the importance of setting realistic targets is essential to having successful end-to-end service levels.

Bottom Line: End-to-end SLAs need to provide meaningful measurements that are relevant to business units, and that connect to more-detailed metrics which enable the IS organization to manage components to meet the end-to-end goals. For end-to-end SLAs to be successful, each critical supporting component must be evaluated beforehand to have confidence the SLA can be achieved. Then the end-to-end service, as well as the critical components (whether insourced or outsourced), should be tracked and reported on an ongoing basis.