Writing SLAs for the Cloud

Your roadmap for creating a successful cloud service-level agreement (SLA)

In this e-guide

In this e-guide:

Service-level agreements are supposed to help keep customers and providers happy, but writing one for the cloud means you must set boundaries to account for more players and possibilities.

This 3-part e-guide aims to help you understand, plan, and create effective cloud service-level agreements (SLAs).

First, Tom Nolle, president of CIMI Corp. and communications expert, explains 3 steps to help you craft complete cloud SLAs.

Then, B2B technology specialist and writer Crystal Bedell explains how to differentiate between cloud and hybrid cloud services.

Finally, delve into the world of public cloud SLAs, and discover the critical mistake companies often make when buying cloud services.

Dodge the traps in creating an SLA

Tom Nolle, President, CIMI Corporation

Most cloud computing users think service-level agreements are important for the cloud, but they usually don't have a handle on how to enforce them. Without proper consideration and tools, even a good service-level agreement (SLA) could fail if you don't know it's being violated or why.

One of the challenges with a cloud computing SLA is that the experience delivered by a cloud application is the sum of the performance of three or more entities. Figuring out which one might be causing a problem can be a challenge, so the first task in creating an SLA decision framework for the cloud is to develop a simple entity map that shows who provides each portion of a cloud service and where their portion transitions into another's area.

A typical cloud application starts with user-owned facilities such as a mobile device or an entire company network. From this user-supplied piece, the cloud application connects through a wide area network, usually the Internet, to the cloud provider's infrastructure. Some users employ a virtual private network (VPN) for cloud access from fixed sites, and others may have more than one cloud provider, so it's possible that there will be more than the three standard responsibility zones in your own cloud.

Cloud applications generate workflows that move across these zones. You'll want to understand exactly how that movement takes place for each type of cloud application you run. It should be possible for you to identify, based on the name of an application, how work flows to fulfill users' needs. That workflow is the basis for your SLA decisions.

To establish good SLA management and policy decisions, you need to measure the behavior of each of the players in your cloud. You should always start with measuring response time and then measure conditions at the Home Editor's Note Dodge the Traps in Creating an SLA Responsibility in the Era of Hybrid Clouds Keep Applications Flowing Smoothly With a Cloud SLA 4 All in the Cloud Together getting started boundary points of different zones.

End-to-end response time measurement is best done at the user connection point so you can read the full response time. In some cases, this means building response time monitoring into the application itself; however, the device's TCP/IP software often provides some of that data through a management interface.

For zone-boundary monitoring, some form of traffic or protocol monitoring is hard to beat. These tools put probes, software tools or hardware elements in the network at various places, and they allow a central management console to view the traffic flow using deep packet inspection to sort out applications.

One big mistake users make at this point is to focus on monitoring without knowing what's good or bad. A network management system (NMS) may collect data in a repository naturally (OpenNMS does this, for example). This data collection allows you to run queries to analyze performance and conditions over time and set baselines for normal behavior as well as thresholds for what you'd consider to be SLA violations. If your management system doesn't provide a repository, you'll want to add network analytics tools to gather and correlate management data and set your performance baselines.

Network analytics can be a strong foundation for decisions concerning cloud SLAs. Make sure the tool can add cloud performance data obtained from the cloud management system APIs to network data obtained from your own NMS. If you have a VPN or a hybrid cloud with a large data center component, it's even smart to first look at tools from your primary network vendor. These are always helpful in maintaining your own IT and network infrastructure performance and will also help manage cloud SLA decisions.

It all comes down to how SLA errors are detected. A good system has three inputs in the detection process. One is subjective user reports of poor performance, the second is an end-to-end response time problem for one or more applications and the third is a report of a specific problem at a zone boundary. In all cases, you should first assess the impact of the problem and then target possible contributors to it.

Your workflow-zone map will let you see whether there's a general problem with several applications at a zone boundary or with only one application. In the former case you probably are experiencing a network or cloud infrastructure problem and in the second a cloud boundary. Your traffic probes will most likely identify either of these faults.

If there's a problem, the remediation should be treated as a small project, with a project manager and a fixed set of tasks that are usually called the escalation procedure. Some users even employ simple software project management or fault tracking tools to monitor the process of cloud SLA issues from their detection to their resolution. Fault tracking tools intended for software projects can sometimes be used, but some network analytics tools include at least an option for fault tracking.

Taking an organized approach to a cloud computing SLA and the decisions that come out of its enforcement is critical if the SLA is to be successful. If you start your deliberations with plans to support SLA decisions, you'll have a better experience overall.

↘ **Next article**

Responsibility in the era of hybrid clouds

Crystal Bedell, Technical Writer, SearchSoftwareQuality

Most end users don't care whether their IT services are hosted in a public, private or hybrid cloud unless a service goes down or data is lost. Then the pressure is on IT. As IT departments adopt hybrid clouds, writing service-level agreements (SLAs) becomes increasingly difficult. But with a bit of foresight and understanding, IT can weave together a cloud SLA that sets up service standards and meets compliance, security and governance requirements.

"A hybrid cloud means a combination of services—some public, some private—and they work in unison as though they are a single system," said Judith Hurwitz, CEO and founder of technology consultancy Hurwitz & Associates. "The issue is that these are not static systems. Any one of the components could be the source of problems. Furthermore, your weakest link one day might not be your weakest link the next day. It's very fluid," Hurwitz explained.

It's essential that IT departments understand who owns all the parts of the stack in the new hybrid environment. "As you introduce multiple players, you have to be very [conscious] of who has what responsibility. When developing an SLA, you have to have a clear understanding of ownership

and support at multiple levels of the stack," said Tim Burke, CEO of managed service provider Quest Media and Supplies.

As an example, Burke said if you're using compute and storage from Amazon Web Services, then Amazon is managing those stacks and their security. You may think that all you have to do is connect them. "But by introducing the connectivity between private and public cloud, you've also introduced the transport issue, so you need to know who is responsible for that," Burke said.

The problem is that cloud providers have a demarcation point where their responsibility ends and the user's begins, according to Abner Germanow, senior director of enterprise marketing for New Relic, a software analytics company. "Depending on the cloud service, [the demarcation point] might be the OS, the platform as a service or a container. It can be one of any number of layers, but it always exists," he continued.

In addition to understanding this demarcation line, IT organizations need to understand the service levels a cloud provider delivers. "[Cloud services] provide an excellent platform based on the standards they've defined. If that works for you, that's great. But don't impart other attributes [on them]. It's not fair for them," Burke said. "Most public cloud providers focus on scaling, so sitting down and [spending] a lot of time with you to map out new service levels is outside their business model."

Hurwitz agreed, but she also pointed out one possible exception. "If you're a multi-gazillion dollar company and you've signed up for a special service, maybe then you will have clout and they'll do something special for you. It's all a matter of scale and importance," she said.

But for the large majority of businesses that aren't multi-gazillion dollar companies, Hurwitz offered this advice: "You have to decide what tolerance you have for unpredictability. If the services fail in any way or slow down, it will impact revenue and your ability to do business."

On a similar note, companies should also consider which applications are the most essential to their business model. For example, if a company communicates with customers only via email, then it's a requisite application. "You might have to pay a little more or look at a different strategy because you know if that service goes down for 10 seconds, you're hosed," Hurwitz said.

The organization also must determine what areas are most critical, such as performance and security requirements, and what backup and disaster recovery strategies are in place. "All of these become issues you have to figure out," Hurwitz explained, "[and then] you have to work on bringing elements together so that it gives you the service level that you require."

Germanow said there may be times when your SLA with end users is more stringent than your cloud service-level agreement. "You might have an

application architecture that fails over with multiple providers or otherwise triggers different availability scenarios so that the customer experience stays solid on your service-level agreement. [Meanwhile] the line you have with the various cloud providers might be much looser because you can fail over from service provider to service provider," he explained. "Understanding all that is pretty critical."

↘ **Next article**

SearchCloud
Applications

TechTarget

⚑ Keep applications flowing smoothly with a cloud SLA

Tom Nolle, President, CIMI Corporation

Evaluating or writing a service-level agreement for your cloud service is a lot more complicated than writing one for simple connection services, such as a virtual private network (VPN). To get cloud computing service-level agreements (SLAs) right, you need to understand the elements of a cloud experience and who actually provides them. Public cloud computing services offer incredible agility and efficiency within their scope, but just how broad that scope is depends on service cost, as well as availability and performance.

The most critical mistake buyers of cloud services make with SLAs is forgetting that all applications are really workflows. A request is passed over a network connection from a user to an application, often one made up of multiple components. That request can then cause work to flow to other components—within the cloud or back into the data center—and to databases that could be located inside or outside the cloud. Eventually, the response is returned to the user.

SLAs aren't useful if they focus on only one piece of this, such as the part concerning public cloud hosting. If any part of this workflow is interrupted, the application fails. If performance falls anywhere along the flow, the application experience is affected. It does no good to tighten the boundaries of performance or availability in the cloud when everywhere else is only loosely guaranteed.

Another problem with evaluating cloud SLAs is the failure to get guarantees from the relevant players. The cloud workflow usually involves at least three players—the worker's own local networking, the network provider that gets workers to the cloud and the cloud provider. It may also involve your company's data center—for network and hosting—and a different network provider with a cloud-to-data-center connection. Providers usually can't write or accept SLAs that apply to the handling of workflow pieces they're not involved with. You'll need to either get them to agree to be a prime contractor, for which they'll charge a fee, or you'll have to write an SLA for each player involved.

The network connections are usually the biggest problem in SLAs because, in most cases, the cloud provider doesn't offer the network services, except within the cloud itself. You'll need to write an SLA for network services if you want to maintain stringent service levels. As a result, you should first see if your cloud provider will offer a VPN or whether they work with VPN providers you could use. In many cases, you'll still need to use the Internet to

get users connected, but a VPN will give you a solid network boundary where you can expect certain amenities.

Border crossings in hybrid clouds also create SLA issues. Workflows follow paths determined by application and business logic, and if these paths make multiple and variable crossings between the data center and the cloud, the performance and availability risk will rise. Your cloud provider can't hit a moving target of workflow patterns when guaranteeing performance or availability, so try to ensure that you don't introduce significant variables to the workflow where you expect firm guarantees. If you do, you'll have to write a very detailed and complex SLA to address all of the variables, and many providers simply won't accept it.

The final issue in SLAs is detection of violations, as well as the penalties and processes of remediation. It's very unlikely that either you or your cloud provider — or other network partners — will accept an SLA based on one party's measurement of conditions. Good SLAs define a point of measurement at a boundary where each party can make independent measurements for verification. Your own SLA should identify those points, the measurements to be taken and the conditions that will be considered a violation.

Availability and performance violations for packet network and cloud services are usually based on a fairly long reporting period, such as outages per week or month. It's best to have downtime-per-interval agreements

rather than simple fault counts because the latter won't account for the mean time-to-repair. Response-time SLAs are harder to write due to the difficulty in measuring those times correctly. If you include response times in your SLA, take time to specify exactly how both parties are going to measure them.

Remediation or penalties are always a sticky point. Many users think they can arrange to be compensated for a business loss if an SLA is violated. That's extremely rare and expensive; you'd be better off engineering your applications for high availability than trying to negotiate such an agreement.

Users report that the most helpful penalty in an SLA is an escalation clause. If an SLA failure occurs, the provider's operations center should be notified. If there's no resolution in a specified time or if the event frequency exceeds a threshold, then a notification should be sent up the provider's management chain — with the higher level becoming responsible for checking the situation and personally contacting you with status updates and remedies. This guarantees senior management will be involved in your problems, which reduces the chances you'll have issues in the first place.

Financial penalties in SLAs should be for the cost of service during the outage period with perhaps a rebate of service costs over the entire measurement interval if the outage is severe. Your chances of getting such a penalty clause will depend on the size of your contract and your potential as a future customer.

🚩 About SearchCloudApplications

SearchCloudApplications.com is dedicated to serving the information needs of organizations which have, or are planning to add, business applications in the cloud. The site spans cloud delivery models and challenges, with articles, tips, tutorials and other resources covering Software as a Service (SaaS), Platform as a Service (PaaS), integration and cloud development.

IT, business professionals, and developers will find unbiased information across cloud topics including technology strategy, Java, vendor selection and database management, architectural considerations, deployment strategies, mobile applications, development and big data strategies, and more.

For further reading, visit us at http://SearchCloudApplications.com

Images; Fotalia