Chapter 3

Demystifying Service-Level Agreements and Avoiding the “Gotchas”

By Michael J. Dunne[1]

I. Introduction

This chapter describes the importance of services level agreements to both the users of cloud services and to the providers of those services and, using the three most common service levels or “metrics,” suggests and explains approaches to address many of the issues that arise in reviewing and negotiating services level agreements. This chapter ends with a list of general suggestions to avoid typical “gotchas” in the review and negotiation of service level agreements.

II. Cloud Services Service-Level Agreements in General

Cloud services are a continuing and often time-sensitive type of service for which a simple warranty of performance in substantial accordance with the documentation and a corresponding obligation to reperform unsatisfactory services do not always provide appropriate protection for the client utilizing the cloud services (the Client). Although warranties regarding certain aspects of cloud services are appropriate, when it comes to performance or nonperformance of the cloud services, covenants regarding the level of performance with corresponding remedies provisions offer important and necessary protections when properly drafted, implemented, monitored, and enforced.

The below discussion is applicable mainly to “private” cloud services arrangements in which the Client has the ability to require and negotiate performance standards. It should also prove helpful in evaluating performance standards, if any, provided in the terms and conditions of “public” cloud services. Additionally, as discussed at the end of this chapter, the below discussion should be useful with those “public” cloud arrangements where there is an ability to obtain deal-specific performance standards.

Performance covenants also offer an excellent method for the provider of services (the SP) and its Client to work together as a team to ensure a win-win cloud services relationship. Unfortunately, taking a “team” or “win-win” approach to service-level agreements (SLAs) is the exception among SPs and their Clients. In most circumstances, neither the SP nor the Client takes the time or makes the effort to realize the value of such an approach. Instead, they default into taking the same adversarial approach to SLAs as they take to negotiating other arms-length agreements between them—that is, the SP strives to

[1] Michael J. Dunne is a partner with Day Pitney LLP in the Mergers, Acquisitions, and Joint Venture Practice Group; the Technology, Telecommunications, and Outsourcing Practice Group; and the Emerging Companies and Venture Finance Practice Group. He is a frequent lecturer and has written numerous articles regarding technology, outsourcing, and mergers and acquisitions topics.

structure each SLA to limit any possibility that the SLA won't be met and ensure that the Client will not have any meaningful remedy if the SLA is not met. In that regard, see the service-level objectives (SLO) discussion below. Clients then respond adversarially by attempting to maximize their ability to claim a SLA failure and obtain painful remedies for the SP, believing that such a threat will force the SP to stay focused on providing top-level services to the Client. Given that the adversarial approach to SLAs is the most common approach taken by SPs, their Clients, and their respective counsels, the below analysis is premised on that approach.

SLAs, which are covenants governing the level of service performance by SPs, may be referred to by other terms such as "performance standards" or "performance measures." Regardless of the moniker, properly drafted SLAs provide a means of measuring the SP's performance against agreed-upon expectations and moving forward together when such performance does not meet those expectations, without requiring the Client to declare a breach and seek standard contract remedies for breach. A concomitant benefit for the SP is a reduced likelihood that a Client experiencing difficulties with the SP's cloud services will allege a breach, the mere allegation of which could adversely affect the SP in a variety of ways.

The importance of SLAs for Clients has not been lost on regulators. For example, in the financial services industry, principal federal regulators have consistently advised banks and other financial institutions to ensure they consider and, where appropriate, enter into SLAs with their service providers.[2]

Properly drafted SLAs will, at a minimum, include four separate elements: (1) measurable standards of the promised services (often referred to as the "metrics"); (2) how the metrics or the performance of the metrics will be measured, including, where appropriate, over what time periods; (3) measuring and reporting responsibilities; and (4) ramifications or remedies. Each element must be carefully reviewed, considered, and drafted to ensure effective SLAs. The omission of an important metric, lack of clarity in the definition of a metric or how it is measured, or the failure to include meaningful remedies are mistakes that could undermine the usefulness of the SLAs and lead to dissatisfaction in the cloud service relationship.

Equally important from the SP's perspective are the exceptions to or exclusions from the SLAs. It is in the SP's interest to exclude events, conditions, and circumstances that may adversely affect its performance, such as the performance of third parties. The question for the Client is whether the risk and ramifications of any such event, condition, or circumstance should be borne by the SP or by the Client. Therefore, from the Client's perspective, the exceptions and exclusions must be carefully reviewed, considered, and drafted to ensure not only that the exception does not become the rule and are appropriately limited, but that the risk-shifting is both appropriate and the rules clearly stated so as

[2] *See* Office of the Comptroller of the Currency, OCC Bulletin 2013-29, Description: Risk Management Guidance, Oct. 30, 2013, at 8, *available at* https://www.occ.gov/news-issuances/bulletins/2013/bulletin-2013-29.html; Federal Financial Institutions Examination Council: Information Technology Examination Handbook Sept. 2016, at 12–15, *available at* https://www.ffiec.gov/press/PDF/FFIEC_IT_Handbook_Information_Security_Booklet.pdf; Board of Governors of the Federal Reserve System, SR 13-19/CA 13-21, Guidance on Managing Outsourcing Risk, Dec. 5, 2013, at 6, *available at* https://www.federalreserve.gov/supervisionreg/srletters/sr1319a1.pdf; Federal Deposit Insurance Corporation, FIL-44-2008, Third-Party Risk Guidance for Managing Third-Party Risk, June 6, 2008, at 7, *available at* https://www.fdic.gov/news/news/financial/2008/fil08044a.html.

not to be overly broad or open to dispute. For example, if a third party is retained by the SP as a subcontractor to provide a portion of the cloud service, should the SP be allowed to exclude from its failure to perform the cloud service any and all such failures caused by such third-party's failure to perform? This is a commonly negotiated issue, the outcome of which will depend on the facts at hand and the bargaining power of the parties, and is discussed more fully below.

III. SLOs—Not the Same as SLAs

Before delving more deeply into service-level agreements, a note of caution: many SPs will offer service-level objectives (SLOs) to Clients in lieu of SLAs. Many SPs will even argue that there is no substantive difference between SLOs and SLAs. In fact, from the Client's perspective, SLOs are vastly different from and, in many ways, vastly inferior to true SLAs. Both will include metrics and descriptions of levels of performance of those metrics. Both may also include obligations to monitor and report. There may even be ramifications for the SP if the services fall below the SLO objectives, such as the obligation to conduct a root-cause analysis to determine how the failure occurred and to take corrective action based on that analysis. However, as the name implies, SLO metrics are only *objectives*. There is no assurance, covenant, or promise that the "objectives" of any metric will be met, and no meaningful repercussions for the SP or agreed-upon remedies for the Client if the objective is not met—it's just an "objective," a "target," after all, not an agreed level of performance with an agreed-upon remedy.

IV. Metrics—The Core of SLAs

Metrics form the core of SLAs. Consequently, the first step in reviewing SLAs proposed by an SP, or in preparing SLAs, is to determine what metrics should be included. The proper metrics will depend upon the services to be provided by the SP and the aspects of those services that are measurable.

However, the determination of what metrics to include must be made in consultation with the Client and, in particular, the individuals who will use or in some manner rely on the proposed services. To be useful and meaningful for the Client, the SLAs must focus on the business objectives and needs of the Client in obtaining the cloud services. The individuals who will use or otherwise rely on the services will be in the best position to define and explain to counsel those objectives and needs and the promised aspects of the SP's services that were key to the Client's decision to engage the SP. Those individuals will also know how the degradation of various aspects of the SP's services will affect the Client and its business, and of those aspects, the ones that are measurable in a meaningful way. Stated another way, the business users will be best positioned to know which failures in the SP's services will cause the most pain and disruption for the Client, and therefore which metrics should be included in the SLAs.

Equally important will be discussions with and assistance from the Client's chief information or chief technology officer and the applicable member(s) of such officer's team. Often an organization's technical team has significant experience working with and negotiating SLAs and can use that experience to help achieve appropriate SLAs. Additionally, an organization's technical staff can be key to explaining to the organization's business users the importance and impact of appropriate SLAs.

A. Defining and Measuring Metrics—Some Typical Metrics and Some Typical "Gotchas" to Avoid

Unfortunately, no standard set of metrics exists that should be included in all cloud services agreements. As noted above, metrics should be tailored to the particular transaction, to the particular objectives and needs of the Client, and of course to the particular services. However, three fairly common metrics for cloud services can be referred to as (1) "availability"; (2) "responsiveness" or "response times"; and (3) "incident response and resolution." These three metrics can also serve as the basis for describing how SLAs should be structured and how to avoid traps for the unwary.

1. Availability

The concept of the "availability" service level is at first blush simple and understandable: What percentage of the time is a cloud service available to the Client? However, an "availability" service level can be and often must be a definition-driven covenant based on the cloud services to be provided. At a minimum, the parties must agree upon: (1) what "available" means; and (2) what time periods are covered. Within those two points, other parameters must be addressed, including exceptions or exclusions.

Depending upon the nature of the service, the SP will likely wish to limit the time during each day that the availability of the service is monitored for purposes of the availability service level. For example, if the users of the cloud services are employees of the Client who will only use the services during normal business hours, then the Client may agree to limit the "availability" metric to the Client's normal business hours on days that it is open for business. However, if the cloud services will support or provide services used on an external-facing website intended to be used by the Client's customers, then having an availability metric of 24 hours per day, seven days a week may be critical to the Client.

As noted above, carefully defining "available" is critical. To simply state that the SP's services will be available 24 hours per day may sound fine; however, in actual practice, such a statement should be seen as ambiguous. What if all but one function of 10 total functions of the cloud service are accessible and may be used by the Client's employees or customers? Is that "available" for purposes of measuring the availability service level? What if that one function is a key function from the Client's perspective that must be functioning for the other nine to be of any use to the Client's employees or customers? What if the Client's employees can access and use all 10 functions, but it takes multiple attempts to enter data and have the data properly processed by some or all of the functions? The point is that for those and other reasons, it is best that the parties define exactly what they mean by "available." A Client would be best served by its counsel conferring with the Client's business users to understand how to define "available" with respect to the particular cloud service to be provided.

Counsel should closely review any calculation proposed by the SP (and independently check such calculations proposed by the SP) to determine with the Client how the calculation will work in practice and whether the results are acceptable to the Client.

A typical approach used by SPs is to lengthen the measurement period over which its performance is measured. Consider that a 10-hour problem spread over a longer period has a lesser proportional or percentage effect on a performance measure than the same problem occurring within a shorter period. Consequently, be wary of measurement periods greater than one month. Another favorite "gotcha" is to take advantage of a limited period during which the cloud service must be available. For example, if the measurement period for the availability service level will be tested as a percentage over each calendar month, and the SP and Client have agreed that the critical period of availability is a set period each day, such as 6:00 a.m. to 10:00 p.m. (the Critical Hours), there are numerous ways for the SP to manipulate the calculation.

First, the calculation for "availability" can be set so a 10-hour problem period during the Critical Hours is deducted from the numerator but not the denominator, while any problem period outside the Critical Hours is not deducted because the Client is not using the service during those hours. This may appear appropriate. However, if the denominator is set at the total hours in the month, such approach will give the SP a "free pass" on the hours between 10:00 p.m. and 6:00 a.m. Instead, the calculation should compare apples-to-apples. Thus, the denominator in the above example would be reduced to the number of hours in the applicable 6:00 a.m. to 10:00 p.m. period during the given month, and the numerator would be the number of hours of availability during the hours of 6:00 a.m. to 10:00 p.m. in such month. There are other tactics to manipulate time in the given example, but the above gives a fair idea of why the Client's counsel must carefully review and perhaps work through a few examples of the proposed service-level calculations.

As noted above, exclusions or exceptions are the SP's means of avoiding responsibility for what might otherwise be failures to comply with SLAs. Some SPs even refer to them as "performance exclusions." Below is an example of how an SP may define performance exclusions:

> SP will not be responsible for, and may exclude from the calculation of compliance with any performance metric, any failure to meet the performance metric if and to the extent that such failure to meet a performance metric is related to or caused by (any of the following being referred to as a "Performance Exclusion"):
>
> (i) prescheduled downtime, downtime during maintenance windows, or downtime during any preventative maintenance, provided advance notice has been given to Client for such downtime;
> (ii) acts or omissions of Client or third-party providers;
> (iii) an event, condition, or other circumstance beyond the reasonable control of SP; or
> (iv) failure of the data communications carrier lines between Client and SP's System.

Again, at first blush, the above may appear reasonable and acceptable. However, even without knowing the exact cloud services to be provided, considered more closely, the above will be seen as ambiguous and heavily in favor of the SP. For example, in subsection (iii) above, it would be better to rely upon a clear, agreed-upon definition of force majeure. The cloud services agreement between the SP and the Client should have a force majeure provision that, among other things, provides some parameters (think restrictions and obligations) around when the SP may rely on an event truly beyond its control to excuse its performance under the cloud services agreement. The better approach, therefore, would be to replace subsection (iii) above with something like "the occurrence of a Force Majeure Event (as defined in the Services Agreement)." The force majeure exculpatory provision in a SLA should also include a right to terminate if the excused performance extends past a set point (e.g., 20 days).

Similarly, consider subsection (ii) in the above sample: "acts or omissions of Client or third-party providers." Subsection (ii) addresses and treats in the same manner two distinct types of actors: first, the Client, and second, "third-party providers." Again, from the Client's perspective, a strong argument may be made that the acts or omissions of third parties, especially those retained by the SP, should be run through the test of the definition of the term "Force Majeure Event." The type of service provided by the third-party provider may affect the strength of the argument for the application of a Force Majeure test. If the third-party service provider is providing a service that falls well within the service expected directly from the SP, the argument for Force Majeure treatment is greatly enhanced. A second way of addressing the point without relying on the Force Majeure reference is to simply revise the wording to focus on which party retains the third-party provider by adding the word "its" or "Client's" before "third-party providers," for example, so that the provision would read: "acts or omissions of Client or Client's third-party providers."

One can make a similar argument with respect to subsection (iv) in the above sample. The argument would be that the exclusion in subsection (iv) is acceptable only if the SP is not responsible for such communication carrier lines. The Client's argument would be that, if the SP is responsible for such carrier lines, then the adverse effect on any SLA caused by a failure of such lines should only be excluded from the measurement of the SLA's performance, if such failure falls within a Force Majeure exclusion. Again, in most cases, a failure to perform does not fall within a Force Majeure exclusion unless it meets certain express conditions.

Next consider subsection (i) above, which provides various exclusions from SLA performance measurements for various types of maintenance services. Addressing the effects of maintenance services on SLA obligations and measurements can, at times, be difficult. The difficulties often stem from the fact that the SP wants to maximize its ability to schedule and perform maintenance, whereas the Client wants to minimize any adverse effects the performance of maintenance may have on its operations or business. As a general principle, from the Client's perspective, maintenance that impinges upon the Client's business operations or needs should not be excluded from the measurement of any SLA. A Client may, however, agree to certain exceptions to that general principle. For example, a Client may be willing to allow its use of the SP's service to be interrupted during the Client's normal working hours solely for the purpose of installing a critical security patch, and to agree that the time to install the patch would not be included with the measurement of any SLA.

Keeping these general principles and exceptions in mind while reviewing subsection (i) in the above sample, the Client would benefit from defining the terms "prescheduled downtime," "downtime," "maintenance windows," and "preventative maintenance." The intent would be to define those terms such that their meanings and application within the cloud services agreement are clear, and not subject to unilateral interpretation by the SP or its service department in a manner that could impinge upon the Client's operations. For example, the Client would want to consider defining "prescheduled downtime" and ensure that such time is excluded from SLA performance measures only if it falls outside of the time periods during which use of the services is critical to the Client's operations.

As shown in the above example, one component of the computation of availability is often the time that the relevant service or system is scheduled to be down and unavailable or not fully available. The goal for the Client is to ensure that the definition of scheduled downtime does not for all practicable purposes make the "availability" metric meaningless, which can occur in many ways within the definition. Given that Scheduled Downtime reflects time that the service or system can be unavailable or not fully available without negatively affecting the SP's performance under the Availability service level, it is important to address restrictions on what may fall within the definition of "Scheduled Downtime." For example, if Scheduled Downtime is excluded from the measurement of SLAs, it would be helpful to the Client to limit (1) the hours during which Scheduled Downtime may occur, (2) how long any one specific period of Scheduled Downtime may occur, and (3) how frequently Scheduled Downtime may occur. The below sample clause[3] provides examples of some of the restrictions that should be considered in defining Schedule Downtime:

> "**Scheduled Downtime**" means any scheduled outage in Availability of which SP notifies the Client at least X (x) business days in advance, provided that such scheduled outage (a) lasts no longer than ___ (_) hour(s); (b) is scheduled between the hours of [X] a.m. and [X] a.m., [TIME ZONE/LOCATION] Time; and (c) occurs no more frequently than [X] per [week] [month]. [Service Provider may request Client's approval for extensions of Scheduled Downtime above one (1) hour [, which approval may [be granted in Client's sole discretion] [not be unreasonably withheld or delayed].]

As noted above, SLAs must be viewed in light of the cloud services provided. If the parties agreed upon a "Critical Hours" approach to measuring the availability service level, then the Client should seek to limit the "maintenance windows" and any other maintenance (prescheduled, emergency, etc.) that may affect availability to the period outside the Critical Hours. In that case, the above "Performance Exclusions" may be reduced to something like "SP shall be excused for a Service Level Failure to the extent the Service Level Failure is caused by a Force Majeure Event or caused solely by Client's acts or omissions."

2. Responsiveness/Response Times

An important aspect of performance for certain cloud services is responsiveness or response time. The concern is perhaps best understood in terms of the answers to the following questions: When a user is logging onto the cloud service, how much time will pass from the time the user has clicked "enter" with the user name and password inserted before the user is actually logged on and the service is available? Or, after logging on, how much time will pass from the time the user makes an inquiry of the cloud service until the user has the response displayed on the user's screen? Will the cloud service's response

[3] The sample provisions in this chapter are not recommended provisions. They are provided only for purposes of explaining or demonstrating points made within the discussion. Each SLA must be tailored for the specifics of the particular cloud service(s) provided.

speed meet the needs and expectations of the user, or will it be frustrating or useless to the user because of its delays?

Responsiveness can be affected, however, by many forces, some within the control of the SP, some within the control of the Client, and some within neither party's control. An example of an event that could fall within any one of those three categories is the telecommunications connection. The nature of the event may also affect responsiveness. For example, will the telecommunications connection be a dedicated line, a VPN, a plain-old internet browser connection by the user, or some combination of the above depending on the specific aspect of the cloud service accessed/provided?

Again, the nature of the cloud service must be understood and considered in connection with a response time service level. SPs that provide only cloud storage may be unwilling to provide any assurance on responsiveness or any meaningful assurances. Their position is often that they are not providing any processing and that availability is the only appropriate metric. If, however, the SP is providing more than just storage, such as the underlying application, or if the cloud storage provider is also responsible for the telecommunications connection, then that SP should be more willing to provide an appropriate response time service level.

Similar to the availability metric, a service level for response times will normally be based on some type of average percentage (e.g., 99.99 percent) over a set period of time (e.g., a calendar month) and may be measured 24 × 7 × 365 or only during Critical Hours on defined business days. Depending upon the nature of the cloud service, however, there may be other aspects of responsiveness that are critical to the Client that may be measured, such as ensuring that no response takes longer than a specified time during the Critical Hours. In that regard, consider the importance of responsiveness to a high-speed securities or commodities trading platform where the phrase "he who hesitates is lost" has real meaning. In that industry, seconds can make the difference between making and missing a trade, and any trade can be significant. Consequently, there could be a need for multiple response time metrics for the same cloud service.

Below is a sample response time metric offered by a SP that was providing its software as a cloud service (SaaS). The "metric" is the defined term "Response Time." The other terms are used to measure the metric.

> "**Average Response Time**" means the average Response Time of the Services during the Responsive Hours, calculated over the course of a calendar month.
>
> "**Response Time**" means the number of seconds required for the Services to fully render the initial login page, log into the application, and fully render the End User's account home page.
>
> "**Responsive Hours**" means the hours between 6 a.m. Eastern Time and 11 p.m. Eastern Time.

Recall that one of the four principal elements of effective SLAs is "measuring and reporting responsibilities." Measuring "response time" can be complicated, and the Client must understand the process that will be used and who will be responsible for monitoring, measuring, and reporting (i.e., the Client or the SP).

Consider how "Response Time" would be (or could be) measured under the above sample provision if the cloud service at issue was an internet banking web site for retail customers of a bank. The Client (i.e., the bank) will be concerned with its customers' experience with the bank's website. The website's responsiveness to the customer's actions will be a big part of that experience. For example, the bank will be concerned with how long it takes the website to display the customer's initial page after the customer has gone through the log-on process and submitted its access information and password. The bank will also be concerned with how long it takes for the website to display the customer's checking or other account information after the customer clicks on the icon for the appropriate account. It is important to understand how those response times are measured and by whom.

Such a situation does not lend itself to direct measurements of the actual experience of each customer. You could not, for example, expect each customer to make the measurements, nor could you expect the bank's customers to allow the bank or the SP to place measuring and reporting software on the customers' computers. Additionally, in such situations, multiple parties are responsible for the telecommunication connections that could affect responsiveness. Nevertheless, responsiveness is a key metric for such websites, and the bank will want some way to measure the metric and its customer's experience. Often in such circumstances where measuring actual performance is not practical or not fully within the SP's control, the SP will propose use of a software tool that, in essence, simulates the customer's actions, such as logging on, and measures the corresponding Response Time.

The measurements resulting from the tool can be used for purposes of the service level. Typically, to avoid adversely affecting Response Time due to network latency, the software tool will be located within the same data center as the servers hosting the Client application or website. Such an approach may be acceptable to a Client if the functioning of the tool is understood and other approaches are not practicable.

3. Incident Response and Resolution

With any cloud service, as with any service based on computers and software, there will be problems or failures to perform as promised, and a properly drafted cloud services agreement should include support and maintenance obligations to remediate such problems or failures. However, an obligation to fix a service deficiency may be insufficient if not coupled with a timeliness obligation. An incident response and resolution service level focuses on the timeliness obligation.

The timeliness obligation can be broken down into its four component aspects: (1) the time until the initial response or acknowledgement from the SP that it is aware a problem has occurred; (2) the effort that will be exerted to fix the problem; (3) the time until an acceptable work-around is provided; and (4) the time until a final resolution or fix is provided.

Not all problems are equal. Some are more catastrophic for the Client than others. In that regard, "severity levels" are often defined on a scale from the most severe (e.g., "Severity Level 1" or "Sev 1") to least troublesome (e.g., "Severity Level 4" or "Sev 4"), with different levels of obligations for the above four components. As the time periods to respond (think time between problem discovery and problem resolution) become longer for each lower level of severity, the severity level definitions can be critical to the success or failure of the Incident Response and Resolution service levels from the Client's perspective.

Equally important can be which party determines the severity level assigned to each problem. Often when pressed, the SP will suggest that the assignment be as "mutually agreed," and many Clients consider "mutually" an acceptable compromise. Unfortunately, when it comes to agreeing upon whether a problem is addressed as a Severity Level 1 or Severity Level 2 incident, many of the advantages of having the problem treated as a Severity Level 1 will have been lost by the time the parties are able to "mutually agree."

Additionally, any stalemate on the decision will favor the SP unless addressed in some manner in the SLA. It is best to avoid the "mutually agreed" approach and allow either party the right to initially designate the severity level, but to give the Client a right to override and make the definitive assignment of severity level based upon the impact to the Client. The logic to such an override right is that the problem is affecting the Client's business, and if there is an ambiguity in the definition of the two levels, the Client should make the business decision as to how much effort must be exerted by the SP to remediate the situation. Protections can be provided to the SP against a Client that constantly cries wolf or that the sky is falling by designating everything as a Severity 1.

The first and probably most important protection is clear and sufficiently detailed definitions for the top severity levels inclusive of the aspects that separate those levels. Such clarity and detailed definitions help avoid disputes and help keep the Client honest in its designation of severity levels for the issues that may arise. As a second protection, establish a standard (e.g., three disagreements on Severity 1 designations by the Client within one month) that allows the SP to bring in an agreed independent third party to review those disagreements. If the independent third party agrees with the SP's lower severity level designation on those issues, then there can be remedies for the SP. Such remedies could include the Client paying for the independent third-party review, and paying the SP at its standard time and material rates with a premium for the services provided in responding to the underlying issues.

Below is a sample incident response and resolution metric service level. The below does not include the definitions of the severity levels, which are necessarily service-specific.

> When Errors are reported, SP will [make commercially reasonable efforts to promptly] meet the applicable acknowledgement and status update requirements as set forth in the table below, whether SP is responsible for the Error or not. For Errors for which SP is responsible, SP shall also meet the resolution requirements set forth in the table below.

Incident Severity Level	Acknowledgement/ Initial Response	Status Update Frequency	Target Resolution
SL-1	30 Minutes	Hourly	2 Hours
SL-2	2 Hours	Every 3 Hours	1 Bus. Day
SL-3	1 Bus. Day	Once Per Bus. Day	5 Bus. Days
SL-4	2 Bus. Days	Once Per Month	Next Release

Using the above as an example, the Client must, at a minimum, (1) define all terms such as "Error"; (2) have an understanding or definition of the Errors for which the SP is responsible; (3) include a detailed description of how the Client reports Errors; and (4) clarify the meaning of "Target" and whether "Resolution" involves a permanent fix or a mere work-around. In addition, consider how the phrase in brackets above waters down the SP's obligations.

B. Ramifications or Remedies—Not Having Specified SLA Remedies Is a Major "Gotcha"

As important as proper metrics and measuring for effective SLAs are the "remedies" for failure. Without specified remedies tied to SLA failures, the Client is, *at best*, back to being required to declare a breach and seek standard contract remedies. *At worst*, the Client has, in essence, lost all effective remedies.

In arguing against providing specific remedies, especially against a right to terminate, SPs and their counsel will often state that the remedy is to terminate the services agreement for breach. However, accepting such a statement may be stepping into one of the worst "gotchas" in SLA negotiations.

In order to understand why accepting the breach approach may be a major "gotcha" for the Client, the terms and conditions of the applicable service levels and of the right to terminate must be reviewed and considered. For example, if the service levels are drafted with wording such as that contained in the brackets in the last sample, the Client may not even have a right to claim a breach, and will hence have no remedy for the SP's failure to meet its service levels. The SP will simply argue that all it was required to do was "make commercially reasonable efforts to" meet the metrics, and given that it had made those efforts, there was no breach.

If the SLA is more specific regarding the SP's obligation to perform at the level or levels set forth in the SLA, such that failure to perform at the level set forth in the SLA is a clear breach of the SP's obligations, then the Client may still be left with no satisfactory or truly effective remedy.

First, consider whether it is clear under the cloud services agreement that a failure to meet a service level is a breach that would allow for remedies. For example, must a breach be "material" for the Client to have remedies? Second, if a service level contains various levels, at what level would the failure be "material?" For example, if there are four severity levels of incidents and the SP constantly fails to meet the response times for Severity 3 incidents, but generally meets the response times for Severity 1 and 2 incidents, is there a "material" breach? Clearly, without more, the use of the term "material" could have a chilling effect on the Client's willingness to exercise any right based on a "breach" under certain service levels. The "more" that would be needed is to ensure a direct coordination between the SLA provisions and what constitutes a material breach under the breach termination provision. A simple fix is to set forth expressly the SLA failures that will be considered a "material breach entitling the Client to the remedies provided by" the breach termination provision.

A second problem with accepting the breach termination approach often suggested by cloud providers is that beach termination provisions usually include a notice and cure period. Again, if the breach termination approach is accepted by the Client, counsel must ensure the SLA provisions and the breach termination provisions are properly coordinated. Consider, for example, if the cure period in the breach termination provision should be omitted for SLA failure because the SLA provisions by their very nature or express terms already provide a type of cure period. For example, SLA provisions often contain the concept that a SLA failure has not occurred until the SP's performance has been below a minimum acceptable level for multiple months or quarters. Allowing a "cure period" for unacceptable SLA performance could easily result in never ending alternating periods of failed performance following by a cure period of minimally satisfactory performance, followed by a period of failed performance leading to another cure period. If the Client is willing to accept the termination for breach remedy, then the applicable provisions must be coordinated to avoid such a result.

Even if the Client is comfortable that a SLA failure is a "breach" that entitles it to exercise its remedies for breach without delay of a cure period or otherwise, such a remedy may not be one that the Client wants or is willing to exercise. In most cloud service agreements, the remedy for any material breach is the right to terminate the cloud services agreement or the service, subject to a right to cure. Of course, for most cloud service agreements, what the Client really wants is proper performance, not a right to terminate and incur the time and expense of switching to a new cloud service. Consequently, the Client could find itself in a situation where performance is unsatisfactory in some measurable manner, but the Client has no effective remedy because, for practical purposes, it cannot terminate the relationship.

For cloud services agreements, SLAs should include two basic types of remedies. First are remedies that provide meaningful incentives for the SP to perform at the desired level. Second is a right to terminate when the performance becomes so poor that the pain of transitioning the services to a new SP or back to the Client is exceeded by the pain of continuing with the present SP.

Specific "incentive" remedies are typically some type of financial credit against fees due, often with an increase in the percentage of the credit for increased levels of failures. The general thought is that the financial credits to the Client will help motivate the SP to perform at an acceptable level to avoid the obligation to provide such credits.

In certain circumstances, especially long-term arrangements, it may be advantageous to the Client to agree to a method for the SP to gain back credits. For example, enhanced performance by the SP with no SLA failures over a set period may negate the SP's obligation to provide the Client with the service level credit previously earned. Similarly, in long-term agreements, it may be equally or more important to the Client that the SP promptly conduct a thorough analysis of the cause of the service level failure, take steps to minimize the likelihood of repeated failures, and report to the Client the results of such analysis and the steps undertaken.

If, however, the Client does want to terminate the cloud service agreement or cloud services because of a SLA failure or chronic SLA failures, relying on the standard contractual right to terminate may be ineffective for the Client. As noted above, such right is often contingent upon the provision of notice of breach and an opportunity to cure. From Client's position, that is unsatisfactory because it allows the SP to "cure" after already failing on numerous occasions or in a material manner. In many cases it would, in essence, be placing a right to cure on top of a prior right to cure. Such a right to cure before termination could result in rolling service-level defaults (i.e., chronic poor performance over several months) separated only by the rolling cure periods (e.g., 30 days) during which the SP performs at the minimum acceptable level.

Below is a sample termination provision for “chronic” service level failures:

> Without limiting Client’s remedies under the Agreement or applicable law, if a Service Level Failure occurs: (i) in any three (3) consecutive months or (ii) in any five (5) separate months in any rolling twelve (12) month period, Client shall have the right to terminate the Agreement at any time thereafter upon prior written notice to SP (which notice shall reference this Section and shall describe such failures) without any penalty or liability, and shall receive a prorated refund of all amounts prepaid by Client and unearned by SP as of the date of termination.

In such a provision, counsel must ensure the phrase “Service Level Failure” was properly and clearly defined. Note also that the provision includes a clause to ensure the right to terminate is not interpreted as an exclusive remedy.

V. General Suggestions to Avoid Typical “Gotchas”

1. A major mistake in preparing a SLA is a lack of focus on the business objectives of the Client in retaining the cloud service. The Client is relying upon the SP to meet its business needs. Ensure those business needs and objectives are well understood, the expectations around them are set, and they are properly addressed within the service levels. That will require the descriptions of the services (which will likely be outside the SLA provisions) to be clearly written with the appropriate detail to meet the Client’s business objectives.
2. Another major mistake is to place the Client in a position where it has no ability to adapt to changes in its business objectives and needs with respect to the cloud services. Especially in long-term arrangements with SPs, counsel for the Client should consult the business side about the importance of and need to require periodic (e.g., annual) reviews by the Client and the SP of the services and the associated service levels, including the process to revise, add to, and replace the service levels based on the SP’s performance to that point and changes in technology and the relevant industry. This could be tied into a requirement in the main agreement for the SP to continuously improve its cloud services and the performance of its cloud services.
3. Similarly, in long-term agreements there should be a requirement to include new service levels for any new service that may be added to the cloud services agreement. The cloud services agreement or SLA provisions should include an understanding of minimum service level requirements, including credits or liquidated damages, and a right to terminate for chronic failures and/or severe underperformance.
4. All SLAs should include a savings clause to ensure that remedies available with respect to the performance standards are *not* the exclusive remedies available to the Client for SLA failures. Counsel for the Client should ensure that the payment of credits will not limit the Client’s right to recover other damages and losses, whether pursuant to other provisions in the agreement or applicable law or equitable remedies.

5. Clarify that the credits or liquidated damages are not "penalties," given that provisions in commercial contracts that are viewed as penalties under the law are often unenforceable. Rather, such credits should be intended and seen as a genuine estimate of reduced value to the Client resulting from the SP's failure to meet the agreed service levels or performance measures in that such reduced value to the Client is difficult or impossible to calculate in advance.

6. Whether the SLA has a specific termination right, counsel for the Client should seek to ensure that nothing in the SLA provisions (the credits or otherwise) will be deemed to limit or obviate the Client's right to terminate the cloud services agreement or seek and obtain remedies under other portions of the Agreement or applicable law, even if the SP issues the appropriate service-level fee credits to the Client.

7. Keep the service levels simple and clear. Avoid SLAs with complicated and intertwined provisions, such as weighting provisions. "Weighting" provisions look at numerous metrics and their possible failures or reduced levels of performance and give different "weights" to the different metrics based on the concept that some metrics are more important than others. A numerical value based on the weighting is then calculated for each metric, which numeric values are then added together to determine whether there has been a service level failure. Such weighting provisions can become so complicated it is nearly impossible to understand how they will function. If you believe examples are needed to ensure the metrics or measurements are understood, then the SLA is likely too complicated and must be rewritten so it is clear. Remember, the service levels must be focused on the Client's objectives and needs, and those are usually easy to state and define. The users of the cloud services are unlikely to find it useful, or in any way meaningful, to have complicated, weighted service levels. They have specific objectives and goals they need met. Those objectives and needs are what the metrics should be measuring. In most instances, complicated metrics serve only the goal of the SP to avoid SLA failures rather than focusing on ensuring the service meets the needs of the business user.

8. Consider including a requirement for periodic meetings (monthly, quarterly) to review performance of the cloud service and an escalation process to address problems in managing the SLAs.

9. Consider placing the service levels in a separate document or attachment to the services agreement. Such an arrangement makes the service levels accessible and more easily used by the users of the cloud service and by those managing the ongoing relationship with the SP.

10. Finally, in reviewing and drafting SLAs, keep in mind the differences among "public" and "private" cloud services delivery models. A public cloud service assumes a shared service platform for all of the SP's customers, whereas a private cloud service assumes a dedicated service platform for each separate SP customer. Typically, this means that the Client will have greater flexibility establishing the Client's specific service levels in the private cloud environment than in the public cloud environment, but don't let SPs tell you that the SP has no ability to negotiate regarding SLAs for pubic cloud services. SPs often offer fairly modest

public cloud service levels with no (or minimal) service level credits or remedies. Experience shows, however, that many SPs are willing to offer enhanced service levels and service level credits as an incentive to win business or as part of a higher cost support package. Always remember that you will never get enhanced service levels or remedies if you don't ask!

VI. Conclusion

Properly structured service level agreements can be and often are the key to ensuring a successful cloud service arrangement for both the Client and the SP. Consequently, it is well worth the time and effort to carefully work through with your client and negotiate the appropriate and well-defined metrics, service levels, monitoring, and remedies.