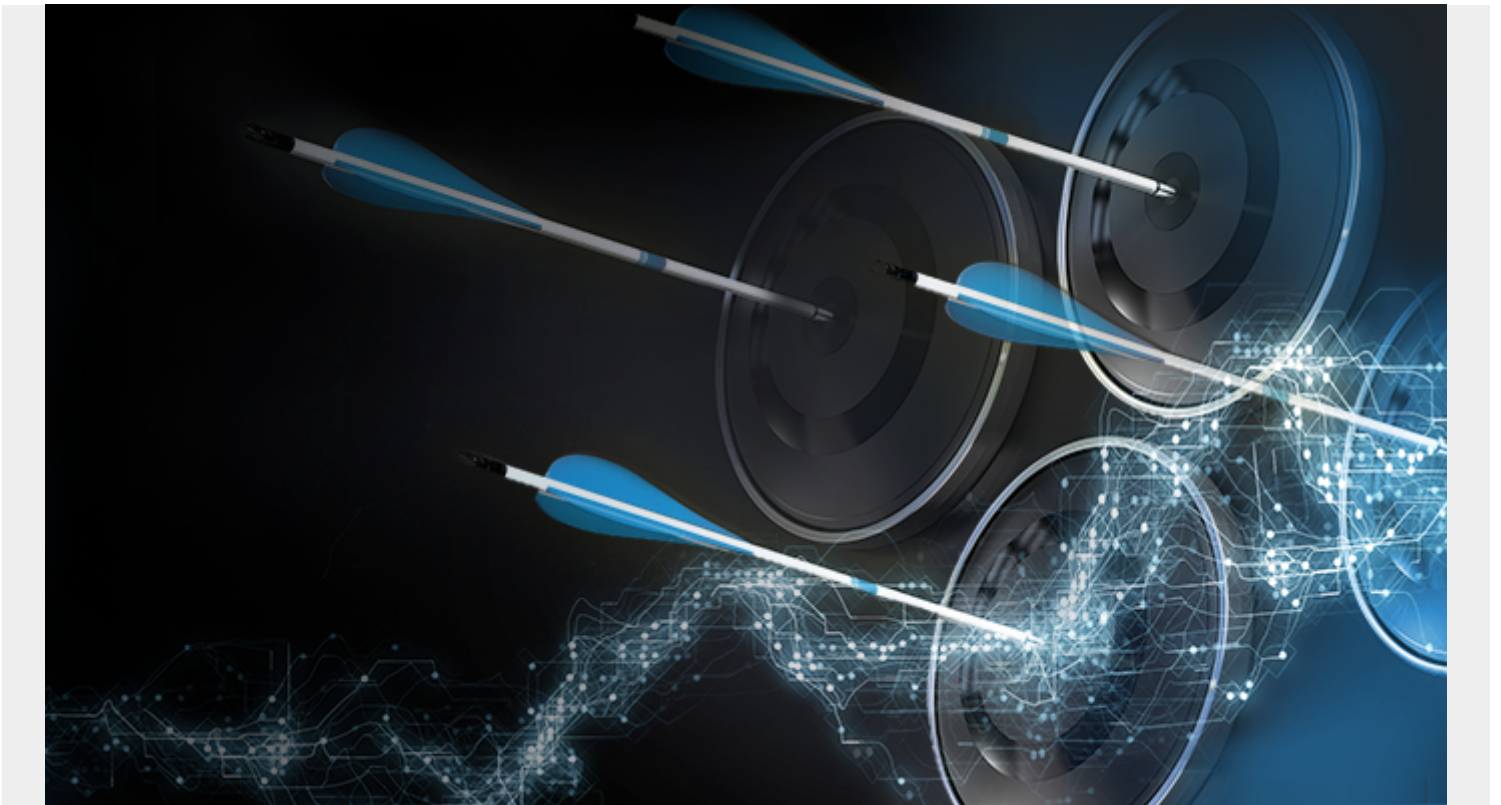


SERVICE LEVEL OBJECTIVES (SLOS) EXPLAINED



When IT workloads are moved from on-premise IT infrastructure to off-site cloud data centers, it's important to ensure that service levels are consistent with your business requirements. The parameters and metrics defining service levels for each element of the cloud solution should meet those requirements and the service provider should maintain high performance, [security](#) and compliance standards at an affordable cost structure. In order to meet these goals, organizations must understand, measure and evaluate the service behavior based on well-defined objectives. The topic of Service Level Agreements (SLAs) is widely covered to discuss the responsibilities of cloud service providers. However, a better understanding of Service Level Objectives (SLOs) can help evaluate measurements of service performance that actually matter to end-users.

Service Level Objective (SLO) serves as a benchmark for indicators, parameters or metrics defined with specific service level targets. The objectives may be an optimal range or value for each service function or process that constitutes a cloud service. The SLOs can also be referred to as measurable characteristics of an SLA, such as Quality of Service (QoS) aspects that are achievable, measurable, meaningful and acceptable for both service providers and customers. An SLO is agreed within the SLA as an obligation, with validity under specific conditions (such as time period) and is expressed within an SLA document.

While the end-goal of defining effective SLOs is to deliver reliable services to end-users, the cost and complexity of getting closer to a 100 percent reliability increases exponentially. Every component of the cloud service causes different impact on the service performance as perceived by customers. For instance, an app may require responsiveness at a specific performance level

beyond which customers can no longer feel a difference. The measure of responsiveness and app performance may be defined through numerical indicators such as request latency, batch throughput or failures per seconds, among other metrics. These indicators describe the service level at any moment in time and must be analyzed over a longer time period to understand the overall performance in context of the agreed SLA contract or availability requirements. Mathematically, the SLO analysis involves aggregating the service level indicator performance over long time and comparing the result with a numerical target for system availability.

An SLO is not intended to define the best performance level but a range of best possible and least acceptable performance standards. Imagine a scenario where a cloud service purchased with an SLA of 99 percent uptime, translating into 7.31 hours of downtime per month. Several months pass by and the systems are maintained at the upper bound of the SLOs delivering 99.9 percent of uptime, or less than one hour of downtime per month.

Suddenly when the systems do go down for several continuous hours, end-users are unpleasantly surprised for the service to perform below their usual expectations. At the same time, the service provider may not be obligated to offer support if it doesn't intend or commit to delivering the best possible SLO.

Since SLO involves the measurement of several quantifiable metrics describing reliability of the system, you should carefully understand the [difference between the terms Reliability and Availability](#) as we explained in a recent BMC blog post. In practice, the SLOs are defined by the lowest acceptable reliability standards.

In summary, Service Level Objectives describe how good was the service reliability during a specific proportion of the time, based on the measurements of specific service level indicators. The following best practices can help you achieve these goals:

- Identify the right metrics and indicators that could be measured to accurately describe system reliability as perceived, expected and required by your organization and end-users.
- The SLO should be well understood by the technical team and organizational leaders. Organizations should devise SLOs based on the business requirements as well as the technical capacity and expertise available to the organization.
- The technical team and business stakeholders should be on the same page in terms of SLO targets. If engineers cannot deliver on the SLO targets, the organization risks failure to comply with its SLAs to customers.
- The SLOs should cover each logical component of the system independently. Every system component may impact or contribute to the overall system differently. Therefore, it's important to define optimum SLOs for every system component based on cost, complexity and other associated business and technical challenges.
- Several service level indicators should be measured collectively to evaluate single SLO target. For instance, the latency, errors and other QoS metrics may be required to evaluate a complete system performance with respect to specific objectives.
- The SLOs should be well documented and communicated between all stakeholders. This information is often critical for technical teams or business leaders to make relevant decisions.
- Service providers may need to prioritize SLOs for different customers. Paying customers with stringent availability requirements may require a higher SLO baseline than freemium users.
- Consider SLOs as an ongoing commitment to deliver optimum system performance across various service level indicators. SLOs evolve over time and cannot be considered as static

targets. IT workloads and end-user expectations change on a continuous basis. An SLO designed for the workload requirements at present may not be equally valid for its future performance requirements.

- Keep SLOs simple, few and avoid absolute numbers that are unrealistic. You may set an internal SLO that acts as a safety margin or buffer to deliver a lower SLO target agreed with the end-users.

It may not be possible to meet service level objectives 100 percent of the time. Cloud service providers need to innovate, add features and update systems, which may involve temporary downtime across several data center instances. Consider this as a tradeoff to achieve SLOs to deliver better services, which may not be possible without the inherent downtime.

Service Level Objectives are all about the targets beyond which the service level is not acceptable to customers and end-users. Setting expectations realistic and not overachieving the targets may be the first step toward delivering a cloud service with acceptable end-user experience at an affordable budget.