

# Guia de uso de SLOs para executivos

Por que os novos paradigmas de experiência do cliente exigem que os executivos pensem nos objetivos de nível de serviço?



# Resumo executivo

As aplicações modernas estão mais complexas e distribuídas do que nunca. E o desempenho dessas aplicações ficou ainda mais crítico para as empresas. Por isso, os executivos estão sob pressão para demonstrar o impacto comercial da organização. Felizmente, as mesmas tecnologias usadas para modernizar e gerenciar aplicações criam oportunidades para vincular diretamente o impacto da TI aos negócios. As equipes de TI e plataforma encontraram novas maneiras de medir a integridade dos ambientes, definir as expectativas das aplicações e conectar os esforços ao contexto de negócios, ou seja, aos objetivos de nível de serviço. Como as aplicações modernas são críticas para os negócios, os executivos precisam orientar a estratégia da empresa para definir SLOs considerando o impacto comercial e a experiência do cliente.

**Neste documento, vamos analisar abordagens e armadilhas comuns ao definir e implementar SLOs e fornecer orientações claras sobre as melhores práticas de seu uso para gerar resultados de negócios superiores.**

## Contexto

Hoje, “a satisfação do usuário final com o desempenho e a confiabilidade da aplicação é fundamental para operações de negócios digitais bem-sucedidas”.<sup>1</sup> No entanto, oferecer uma ótima experiência ao usuário final é um desafio porque as aplicações modernas, assim como as plataformas e infraestruturas nas quais elas são executadas, estão mais complexas e distribuídas do que nunca. Conforme observado no relatório State of the Multicloud de 2021, a complexidade foi citada como um dos principais desafios para atingir as metas de negócios.<sup>2</sup>

Além do desenvolvimento de aplicações modernas, “a mudança para uma infraestrutura ágil com TI híbrida, multivem e contêineres” fez as equipes de TI terem muito trabalho para entender a imensa quantidade de dados e “colocou em dúvida a viabilidade das ferramentas tradicionais de monitoramento de infraestrutura”.<sup>3</sup>

Em resposta a esses desafios, as equipes de TI e plataforma encontraram novas maneiras de medir a integridade dos ambientes, definir as expectativas das aplicações e conectar os esforços aos efeitos na organização. Além disso, como as expectativas dos usuários finais estão cada vez mais altas, os executivos precisam garantir que as empresas usem métricas relevantes para avaliar o desempenho das aplicações e o impacto nos negócios.

---

<sup>1</sup> Fonte: Worldwide Application Performance Management Software Market Shares, 2020 | Junho de 2021, IDC #US47989021

<sup>2</sup> Fonte: 2021 Turbonomic State of Multicloud Report, 2020 CNCF Survey

<sup>3</sup> Fonte: Monitoring and Observability for Modern Services and Infrastructure | Junho de 2020, Gartner G00720854

# A abordagem atual: trabalhosa e reativa

Uma abordagem comum adotada pelas equipes de TI e plataforma para medir a integridade dos ambientes é a identificação e configuração de indicadores de nível de serviço, acordos de nível de serviço e objetivos de nível de serviço.

- **Um indicador de nível de serviço (SLI) é a métrica usada pelas equipes para medir o desempenho de um serviço em relação à meta.**
- **Um acordo de nível de serviço (SLA) é uma promessa de disponibilidade que as equipes de TI e plataforma fazem aos clientes ou usuários finais.**
- **Já um objetivo de nível de serviço (SLO) é o objetivo ou a meta de nível de serviço, medido com um SLI, que as equipes se comprometem a alcançar.**

Os SLOs são especialmente importantes porque definem critérios e expectativas claras para o desempenho das aplicações de uma empresa.

Atualmente, a maioria das empresas segue um processo de configuração de SLO manual que é demorado e trabalhoso. As equipes de TI precisam primeiro determinar quais serviços de uma aplicação afetam diretamente os negócios e a experiência do usuário final. O próximo passo é identificar quais métricas devem ser usadas como indicadores de nível de serviço. Algumas métricas de SLI comuns são disponibilidade, latência e rendimento da transação. Depois que os SLIs apropriados são escolhidos, as equipes de TI precisam determinar a meta de SLO para essa métrica e um período de medição específico. Em seguida, elas criam orçamentos de erro para cada SLO e vinculam essas informações a um sistema de alerta baseado em limite. Muitas empresas adotam essa abordagem baseada em limite porque é humanamente impossível monitorar o desempenho das aplicações 24 horas por dia, 7 dias por semana, 365 dias por ano.

Infelizmente, definir limites de SLO não resolve os problemas de desempenho. Essa estratégia de TI é ineficaz porque é muito reativa para a infraestrutura ágil usada atualmente, que tem aplicações executadas em ambientes multinuvem ou contêineres. Se algo der errado com um serviço e um limite ou alerta for acionado, a degradação do desempenho já ocorreu, o que resulta em uma experiência ruim para o usuário final.

As equipes de TI tentaram melhorar o sistema de limites/alertas implementando o Horizontal Pod Autoscaling (HPA) nas aplicações executadas em ambientes containerizados. No entanto, o HPA também não consegue garantir uma ótima experiência para o usuário final e evitar a degradação do desempenho. Como acontece no processo de configuração do SLO, para definir o ajuste automático de escala horizontal com o objetivo de atender às demandas de recursos, as equipes de TI precisam identificar as métricas que melhor representam as necessidades de recursos, configurar metas, definir limites e fazer testes. Esse processo precisa ser repetido para cada serviço de uma aplicação. Como algumas aplicações têm centenas de serviços diferentes, a implementação do HPA em escala é muito difícil. Além disso, diferentes políticas de HPA ainda dependem de limites e não se correlacionam nem estão em conformidade umas com as outras. Isso significa que o ajuste de escala de um serviço pode afetar outro negativamente. Por fim, esse não é um processo que possa ser feito uma única vez. As políticas de ajuste de escala do HPA precisam ser reconfiguradas e monitoradas continuamente para serem eficientes.



# Como pensar em SLOs...

## **Defina SLIs e SLOs levando em consideração a experiência do cliente.**

De acordo com especialistas do setor, “o foco no gerenciamento de sistemas e aplicações para otimizar a experiência do usuário final é uma grande prioridade porque **um bom desempenho e 100% de tempo de atividade são apostas para o sucesso dos negócios digitais**”.<sup>4</sup> Como o bom desempenho e o tempo de atividade agora são apostas importantes, as empresas não podem perder tempo coletando dados que não sejam indicadores diretos da experiência do usuário final ao avaliar o desempenho de uma aplicação.

Coletar os dados corretos não é uma tarefa fácil. Por exemplo, embora a disponibilidade seja uma métrica comumente usada, ela não é uma extensão direta do desempenho. Uma aplicação pode estar disponível, mas ser afetada por gargalos de recursos e degradação do desempenho. As empresas podem adotar diversas métricas para medir o desempenho. Por isso, é importante que as equipes de TI e das linhas de negócios cheguem a um consenso sobre os dados que devem ser coletados e usados nos relatórios. Como a Gartner observa: “Escolher SLIs representativos e relevantes é fundamental. Na maioria dos casos, é improvável que uma métrica baseada em infraestrutura (“memória disponível” ou “percentual de servidores de processamento livres”) seja relevante porque geralmente não é algo que preocupe os usuários de um serviço. Escolha SLIs que sejam medições diretas da experiência dos usuários com os seus serviços”.<sup>5</sup>

Em outras palavras, os SLOs devem ser usados como uma ferramenta que mostra para as empresas se as aplicações estão atendendo às necessidades dos negócios. Se as aplicações não estiverem funcionando como esperado, as empresas com um SLO significativo definido no nível certo da aplicação vão saber exatamente quais ações terão mais impacto para restaurar o desempenho. Cada empresa vai precisar usar métricas específicas para as linhas de negócios. No entanto, como as organizações estão substituindo arquiteturas de aplicações monolíticas por opções mais modernas e distribuídas, os indicadores de desempenho tradicionais, como uso excessivo da memória e da CPU, são menos úteis. As empresas precisam definir SLIs/SLOs com métricas que sejam relevantes para os negócios, como tempo de resposta e rendimento da transação. Essas métricas são uma medida mais direta de desempenho porque, por exemplo, ao definir SLOs para o rendimento da transação, as equipes de TI podem saber exatamente como as solicitações estão sendo atendidas em cada VM ou pod individual. Essa métrica é uma avaliação mais direta do desempenho do que definir SLOs com base em métricas de utilização e associar essas informações de forma não estruturada a métricas como tempo de resposta e rendimento.

<sup>4</sup> Fonte: Worldwide Application Performance Management Software Market Shares, 2020 | Junho de 2021, IDC #US47989021

<sup>5</sup> Solution Path for Modern Infrastructure and Application Monitoring | Junho de 2019, Gartner

Embora o tempo de resposta e o rendimento da transação sejam algumas das formas mais comuns de avaliar a experiência do cliente e o desempenho da aplicação, são métricas que não funcionam para todas as empresas. Por exemplo, uma provedora de serviços de infraestrutura de desktop virtual não precisa definir SLOs para rendimento da transação, mas para tíquetes de suporte. Os tíquetes de suporte seriam uma métrica mais importante para um SLO porque são uma medição direta do desempenho dos desktops virtuais. Podemos concluir que é preciso definir SLIs/SLOs específicos para cada empresa e linha de negócios.

### **É preciso fazer uma análise contínua das mudanças na demanda por aplicações.**

As aplicações e as infraestruturas modernas são flexíveis, e a demanda por recursos é dinâmica. Devido a essas mudanças constantes, é muito difícil entender as relações entre diferentes fontes e tipos de dados. Geralmente, as empresas usam ferramentas específicas para monitorar as camadas do stack nas várias equipes enquanto tentam resolver diferentes lados do mesmo problema. Essa falta de coordenação resulta em muitas pistas falsas da causa de um problema. Além disso, os dados precisam ser coletados de forma contínua. Caso contrário, as equipes de TI e a plataforma precisarão adivinhar quais dados são necessários para resolver um problema. Como observado pela Gartner no relatório *Monitoring and Observability de 2020*, “a coleta de dados sobre possíveis causas deve ser feita de forma contínua porque é inútil fazer isso depois da ocorrência do problema”.<sup>6</sup>

Para evitar esse dilema, as empresas precisam de um repositório central que possa agregar e correlacionar dados de todas as camadas de um stack de aplicações. Com esse sistema, as empresas podem reunir e analisar dados de forma contínua com base nos SLOs. Por meio dessa análise contínua, as empresas podem contextualizar as informações e identificar qual objetivo devem atingir para melhorar o desempenho das aplicações. Se o objetivo não for alcançado, as empresas vão saber quais são os problemas que precisam ser corrigidos nas camadas subjacentes do stack de aplicações. A implementação de sistemas que façam essa análise contínua é a única maneira de acompanhar as necessidades dinâmicas de recursos das aplicações e infraestruturas modernas.



### **Automatize o uso de recursos dinâmicos.**

A natureza dinâmica das aplicações modernas e a infraestrutura na qual elas são executadas exigem análise contínua. Os executivos e as empresas também devem implementar a automação contínua se quiserem aproveitar os SLOs e criar um sistema preventivo para gerenciar as aplicações. Em um relatório recente, a IDC apontou que as empresas precisam “considerar o papel da automação nos recursos dos produtos” se quiserem manter a competitividade.<sup>7</sup>

<sup>6</sup> Fonte: *Monitoring and Observability for Modern Services and Infrastructure* | Junho de 2020, Gartner G00720854

<sup>7</sup> Fonte: *Worldwide Application Performance Management Software Market Shares, 2020* | Junho de 2021, IDC #US47989021

O desenvolvimento da infraestrutura moderna prometia resiliência e flexibilidade para as aplicações, mas muitos negócios estão com dificuldade em manter o desempenho delas. A implementação da automação no gerenciamento das aplicações modernas é essencial para obter os benefícios da resiliência e da flexibilidade. Sem automação, as resoluções possíveis por meio da configuração dos SLOs corretos e da coleta contínua de dados não são eficazes. As equipes de TI e plataforma não podem garantir o desempenho das aplicações sem automatizar o fornecimento dinâmico de recursos: quando um alerta é acionado e uma decisão de recurso é executada manualmente, a degradação do desempenho já ocorreu.

## O problema dessa abordagem é que ela se baseia no fato de que um problema já ocorreu.

A automação pode ser mais do que uma reação a um evento acionado por um limite. Com a coleta contínua de dados e a identificação dos principais SLIs e SLOs da linha de negócios, as empresas podem fazer o software gerar decisões úteis que devem ser automatizadas. Se, como empresa, você tomar as medidas necessárias e se comprometer com a automação, poderá criar um ambiente realmente flexível que gerencie proativamente como as aplicações recebem recursos, garantindo o desempenho sem interrupções. Esse tipo de automação exige um sistema inteligente que analise um ambiente em constante mudança e automatize a cadeia de decisões necessária para os problemas serem resolvidos *antes* que ocorra a degradação do desempenho. É impossível conseguir esse tipo de desempenho e flexibilidade com um processo que usa planilhas e alertas.

Para se afastar das abordagens reativas, é preciso haver adesão total dos stakeholders nas várias equipes que gerenciam as aplicações e nas equipes que gerenciam a infraestrutura. Os proprietários de aplicações e produtos geralmente hesitam em abrir mão do controle para adotar a automação. A relutância se deve à falta de confiança. No entanto, esse problema pode ser superado. A automação exige uma transformação cultural da sua empresa de TI. Para ter flexibilidade e resiliência, as empresas precisam confiar nas ações que serão automatizadas. Com SLOs relevantes centrados nos negócios e nas aplicações e vinculados ao fornecimento dinâmico de recursos na plataforma e na infraestrutura, é mais fácil para os proprietários de produtos e aplicações confiarem na automação e se sentirem confortáveis com a expectativa de que ela será totalmente implementada nos fluxos de trabalho de negócios.

## Pontos importantes

As aplicações vão ficar cada vez mais complexas porque as empresas estão adotando infraestruturas ágeis, como TI híbrida, plataformas de contêineres e multinuvem. Considerando essas inovações, as organizações que vão prosperar são aquelas que atenderem às expectativas de experiência do cliente e desempenho de aplicações. Estas são as melhores práticas: definir SLOs levando em consideração a experiência do cliente, analisar continuamente as mudanças na demanda por aplicações e automatizar o fornecimento dinâmico de recursos. Para obter sucesso nos negócios digitais, as empresas devem tomar as medidas necessárias para implementar a automação que forneça recursos dinamicamente às aplicações a fim de atender às mudanças na demanda e nos SLOs de negócios.

# Garanta automaticamente os SLOs das aplicações com o Turbonomic.

**O Turbonomic transforma dados em ação, implementando a automação que evita riscos no desempenho das aplicações e melhora a flexibilidade.**

A modernização das aplicações e da infraestrutura de missão crítica é um investimento com inúmeros benefícios. No entanto, para obter flexibilidade, resiliência e velocidade de lançamento no mercado, você precisa de um software que analise continuamente seu ambiente e execute a decisão certa de fornecimento de recursos no momento adequado para garantir o desempenho da aplicação. Com o Turbonomic, você pode correlacionar o tempo de resposta da aplicação, o rendimento da transação ou outros SLIs/SLOs com o fornecimento dinâmico de recursos. Como a demanda flutua, o fornecimento dinâmico de recursos do Turbonomic garante o desempenho contínuo da aplicação.


**O HPA não é suficiente.** O Turbonomic usa análise de stack completa e de cima para baixo para garantir seus SLOs de forma dinâmica. Você define seus SLOs, e nosso software baseado em IA garante que a plataforma e a infraestrutura subjacente forneçam os recursos necessários onde suas aplicações forem executadas.

**Faça a integração com os fluxos de trabalho de negócios.** Com a integração do Turbonomic com Webhooks, você pode injetar facilmente ações do Turbonomic nos ciclos de vida das aplicações, DevOps e pipelines de infraestrutura, fluxos de trabalho de aprovação e auditoria e processos de comunicação.

**Minimize o trabalho manual.** Dev, DevOps e SREs não precisam definir limites, restrições ou políticas de ajuste de escala automático. O software toma as decisões de fornecimento de recursos certas, fornecendo ações que você pode automatizar.

**Não gaste demais na capacidade.** Não há necessidade de depender da equipe de desenvolvimento para tomar decisões de recursos. Ela geralmente superprovisiona apenas por segurança. Nosso software determina os recursos necessários com base na demanda da aplicação.

**Planeje o crescimento de forma rápida e fácil.** Simule a integração de novos serviços com nosso software. Determine exatamente do que você precisa para viabilizar um novo crescimento.



Experimente  
o Turbonomic  
hoje mesmo

[turbonomic.com/try-SLO](https://turbonomic.com/try-SLO)

© Copyright IBM Corporation 2022

**IBM Brasil Ltda**

Rua Tutóia, 1157

CEP 04007-900

São Paulo, SP

Produzido nos Estados Unidos da América

Janeiro de 2022

IBM e o logotipo IBM são marcas comerciais ou marcas registradas da International Business Machines Corporation nos Estados Unidos e/ou em outros países. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atualizada das marcas comerciais da IBM está disponível em [ibm.com/trademark](http://ibm.com/trademark).

Este documento é atual na data de sua publicação inicial e pode ser alterado pela IBM a qualquer momento. Nem todas as ofertas estão disponíveis em todos os países onde a IBM opera.

Os dados de desempenho e os exemplos de clientes citados são apresentados apenas para fins ilustrativos. Os resultados reais de desempenho podem variar de acordo com configurações e condições operacionais específicas.

AS INFORMAÇÕES NESTE DOCUMENTO SÃO FORNECIDAS “TAIS COMO ESTÃO”, SEM GARANTIA EXPRESSA OU IMPLÍCITA DE, ENTRE OUTRAS, COMERCIALIZIDADE, ADEQUAÇÃO A UM DETERMINADO FIM OU DE NÃO INFRAÇÃO.

Os produtos IBM são garantidos de acordo com os termos e condições dos acordos sob os quais eles são fornecidos.

