

# Guide sur les SLO destiné aux dirigeants

Pourquoi les nouveaux paradigmes de l'expérience client exigent que les responsables se penchent sur les objectifs de niveau de service



# Synthèse

Les applications modernes sont plus complexes et distribuées que jamais. Et leurs performances sont plus que jamais cruciales pour les entreprises. En conséquence, les dirigeants doivent constamment prouver l'impact métier de leur entreprise. Heureusement, les technologies utilisées pour moderniser et gérer les applications offrent également l'opportunité de démontrer l'impact de l'informatique sur l'entreprise. Les équipes en charge de l'informatique et des plateformes ont trouvé de nouveaux moyens de mesurer l'état de santé de leurs environnements, de fixer des objectifs pour leurs applications et d'aligner leurs efforts sur le contexte commercial ; autrement dit, de définir des objectifs de niveau de service (SLO). Compte tenu de l'importance vitale des applications modernes pour les entreprises, les responsables doivent pousser leur organisation à penser leurs SLO au regard de l'impact métier et de l'expérience client.

**Dans ce document, nous examinons les approches et les pièges habituels de la définition et de la mise en place des SLO. Nous fournirons des recommandations claires sur les meilleures pratiques permettant aux SLO de générer de meilleurs résultats commerciaux.**

## Contexte

Aujourd'hui, « la satisfaction de l'utilisateur final à l'égard des performances et de la fiabilité des applications est essentielle pour la réussite des opérations numériques de l'entreprise. »<sup>1</sup> Offrir une excellente expérience à l'utilisateur final est cependant un défi dans la mesure où les plateformes et les infrastructures dont dépendent les applications modernes sont plus complexes et distribuées que jamais. En fait, comme souligné dans notre rapport sur l'État du multicloud en 2021, la complexité est considérée comme l'un des principaux obstacles à la réalisation des objectifs de l'entreprise.<sup>2</sup>

En plus du développement d'applications modernes, « le passage à une infrastructure agile comprenant l'informatique hybride, le multicloud et les conteneurs » a compliqué le travail des équipes informatiques chargées de donner du sens aux immenses quantités de données et « a remis en question la pertinence des outils traditionnels de surveillance de l'infrastructure. »<sup>3</sup>

Pour relever ces défis, les équipes en charge de l'informatique et des plateformes ont trouvé de nouveaux moyens de mesurer l'état de santé de leurs environnements, de fixer des objectifs pour leurs applications et d'aligner leurs efforts sur l'impact commercial. Et, compte tenu des attentes toujours plus grandes des utilisateurs finaux, les responsables doivent absolument s'assurer que leur entreprise utilise des indicateurs pertinents pour mesurer les performances de leurs applications et leur impact sur l'entreprise.

---

1 Source : Worldwide Application Performance Management Software Market Shares, 2020 | Juin 2021, IDC #US47989021

2 Source : Rapport sur l'état du multicloud en 2021 d'IBM Turbonomic, enquête CNCF 2020

3 Source : Monitoring and Observability for Modern Services and Infrastructure | Juin 2020, Gartner G00720854

# L'approche actuelle : laborieuse et réactive

L'approche classique des équipes en charge de l'informatique et des plateformes pour mesurer la santé de leurs environnements consiste à identifier et à configurer des indicateurs de niveau de service, des accords de niveau de service et des objectifs de niveau de service.

- **Un indicateur de niveau de service (SLI) est une ou plusieurs mesures permettant aux équipes de mesurer la performance d'un service par rapport à un objectif.**
- **Un accord de niveau de service (SLA) est une promesse de disponibilité faite par les équipes en charge de l'informatique et des plateformes aux clients ou aux utilisateurs finaux.**
- **Un objectif de niveau de service (SLO) est une cible ou un niveau de service, mesuré à l'aide d'un SLI, que les équipes s'engagent à atteindre.**

Les SLO sont particulièrement importants car ils définissent les critères et fixent des attentes claires sur le niveau de performance des applications d'une entreprise.

Aujourd'hui, la plupart des entreprises passent par un processus manuel de configuration des SLO, ce qui est long et laborieux. Les équipes informatiques doivent d'abord identifier quels services d'une application ont un impact direct sur l'activité et sur l'expérience de l'utilisateur final. Ensuite, elles doivent définir les critères à utiliser comme indicateurs de niveau de service. La disponibilité, la latence et le débit des transactions figurent parmi les mesures SLI classiques. Une fois les SLI sélectionnés, les équipes informatiques doivent déterminer leur SLO pour cet indicateur et une période de mesure spécifique. Une fois ce processus terminé, les équipes informatiques doivent créer des budgets d'erreurs pour chaque SLO et les relier à un système d'alerte basé sur un seuil. De nombreuses entreprises adoptent cette approche basée sur les seuils parce qu'il est humainement impossible de surveiller les performances d'une application en permanence.

Malheureusement, définir des seuils de SLO ne résout pas le problème des performances. Cette stratégie informatique est inefficace car elle est trop « réactive » pour l'infrastructure agile moderne, qui contient des applications fonctionnant dans des environnements multicloud ou conteneurisés. Si un service est défaillant et qu'un seuil d'alerte est franchi, ses performances sont déjà dégradées, ce qui engendre une mauvaise expérience pour l'utilisateur final.

Les équipes informatiques ont tenté d'améliorer leur système de seuils et d'alertes en utilisant la mise à l'échelle automatique horizontale des pods (HPA) pour les applications fonctionnant dans des environnements conteneurisés. Cependant, l'HPA ne permet pas de garantir une expérience optimale pour l'utilisateur final ou d'empêcher la dégradation des performances. À l'instar du processus de configuration des SLO, afin de configurer la mise à l'échelle automatique horizontale en vue de satisfaire les demandes en ressources, les équipes informatiques doivent identifier les indicateurs qui représentent le mieux les besoins en ressources, configurer les objectifs, définir des seuils, puis effectuer des tests. Ce travail doit être répété pour chaque service d'une application. Certaines applications comportant des centaines de services différents, la mise en œuvre globale de l'HPA est très compliquée. En outre, les différentes politiques d'HPA reposent sur des seuils et ne sont pas corrélées ou conformes les unes aux autres, ce qui signifie que l'octroi de ressources supplémentaires à un service peut avoir des répercussions négatives sur un autre. Enfin, il ne s'agit pas d'un travail ponctuel et les politiques d'HPA nécessitent une reconfiguration et une surveillance continues pour être efficaces.



# Comment penser les SLO...

## Définir les SLI et SLO en termes d'expérience client.

Selon les experts, « se concentrer sur la gestion des systèmes et des applications afin d'optimiser l'expérience de l'utilisateur final est une priorité majeure, **car des performances rapides et un temps de disponibilité de 100 % sont les clés de la réussite des entreprises dans le numérique.** »<sup>4</sup> La rapidité des performances et la disponibilité étant désormais des enjeux majeurs, les entreprises ne peuvent pas perdre de temps à collecter des données qui ne sont pas directement indicatives de l'expérience de l'utilisateur final lors de l'évaluation des performances des applications.

Collecter les bonnes données n'est pas simple. Par exemple, bien que la disponibilité soit une mesure couramment utilisée, elle n'est pas forcément révélatrice des performances, car une application peut être disponible tout en subissant des goulots d'étranglement au niveau des ressources et une dégradation de ses performances. Il existe de nombreux indicateurs que les entreprises peuvent utiliser pour mesurer les performances. Il est donc important que l'informatique et les équipes concernées s'accordent sur les données à collecter et à communiquer. Comme le dit Gartner : « Il est crucial de choisir des SLI représentatifs et pertinents. Dans la plupart des cas, un indicateur basé sur l'infrastructure (la mémoire disponible ou le pourcentage de nœuds de travail libres) n'est pas pertinent, car les utilisateurs d'un service s'y intéressent peu. Choisissez des SLI qui sont des mesures directes de l'expérience que les utilisateurs ont avec vos services. »<sup>5</sup>

Autrement dit, les SLO doivent servir à indiquer aux entreprises si leurs applications font ce qu'elles sont supposées faire pour leur activité. Si les applications ne sont pas performantes, les entreprises qui ont un SLO pertinent et bien défini pour l'application savent exactement quelles actions seront les plus efficaces pour rétablir les performances. Les entreprises doivent mesurer des indicateurs différents qui sont propres à leur secteur d'activité. Cependant, plus elles abandonnent les architectures applicatives monolithiques pour se moderniser et devenir plus distribuées, moins les indicateurs de performance traditionnels, comme l'utilisation élevée de la mémoire et du processeur, sont utiles. Les entreprises doivent désormais chercher à définir leurs SLI/SLO à l'aide d'indicateurs permettent, comme le temps de réponse et le débit des transactions. Ces indicateurs constituent une mesure plus fidèle des performances car, par exemple, en définissant des SLO pour le débit des transactions, les équipes informatiques peuvent savoir exactement comment les requêtes sont traitées dans chaque pod ou machine virtuelle. Cet indicateur est une évaluation plus pertinente des performances que la définition de SLO reposant sur des mesures de l'utilisation vaguement associées à des mesures de type temps de réponse ou débit.

4 Source : Worldwide Application Performance Management Software Market Shares, 2020 | Juin 2021, IDC #US47989021

5 Solution Path for Modern Infrastructure and Application Monitoring | Juin 2019, Gartner

Même si le temps de réponse et le débit des transactions font partie des critères d'évaluation les plus courants de l'expérience client et des performances applicatives, ils ne conviennent pas à toutes les entreprises. Par exemple, un fournisseur d'infrastructure de bureaux virtuels (VDI) ne doit éviter les SLO basés sur le débit des transactions et opter plutôt pour des SLO basés sur le nombre de tickets de support. Le nombre de tickets de support est un indicateur plus pertinent car c'est une mesure directe des performances des bureaux virtuels. En définitive, les SLI/SLO doivent être définis spécifiquement pour chaque entreprise, selon son secteur d'activité.

### **Il est nécessaire d'effectuer une analyse continue des demandes applicatives changeantes.**

Les applications et infrastructures modernes sont élastiques et la demande en ressources est dynamique. En raison de ces variations permanentes, il est très difficile de comprendre les relations entre les différentes sources et les différents types de données. Souvent, les entreprises utilisent différents outils pour surveiller différentes couches de la pile à travers différentes équipes, tout en essayant de résoudre différents aspects du même problème. Ce manque de coordination engendre des erreurs lors de l'identification de la cause profonde d'un problème. De plus, les données doivent être collectées en permanence, sinon les équipes informatiques et les équipes chargées des plateformes doivent deviner les données dont elles ont besoin pour résoudre un problème. Comme le note Gartner dans son rapport Monitoring and Observability 2020 : « la collecte des données pour identifier les causes potentielles doit se faire en continu, car en l'activant en réponse à un symptôme, on peut passer entièrement à côté de la cause. »<sup>6</sup>

Pour éviter ce problème, les entreprises ont besoin d'un référentiel central capable d'agréger et de corréler les données provenant de chaque couche de leur pile d'applications. Avec ce système, elles peuvent rassembler leurs données collectées en permanence et les intégrer dans leurs SLO. Grâce à cette analyse continue, les entreprises sont à même de tout contextualiser et de savoir quel objectif elles doivent atteindre pour que leurs applications soient performantes. Et si un objectif n'est pas atteint, elles savent quels sont les problèmes à corriger dans les sous-couches de la pile d'applications. En fin de compte, ce type de système, capable d'effectuer une analyse continue, est le seul moyen de répondre aux besoins en ressources dynamiques des applications et des infrastructures modernes.



### **Automatiser l'allocation dynamique de ressources.**

Tout comme la nature dynamique des applications modernes et l'infrastructure sur laquelle elles s'exécutent exigent une analyse continue, les dirigeants et leurs entreprises doivent également chercher à mettre en œuvre une automatisation continue pour atteindre leurs SLO et créer un système préventif pour la gestion de leurs applications. Dans un récent rapport d'analyse, IDC indique que la compétitivité future des entreprises dépend de leur approche quant au « rôle de l'automatisation dans les capacités des produits. »<sup>7</sup>

<sup>6</sup> Source : Monitoring and Observability for Modern Services and Infrastructure | Juin 2020, Gartner G00720854

<sup>7</sup> Source : Worldwide Application Performance Management Software Market Shares, 2020 | Juin 2021, IDC #US47989021

Le développement d'une infrastructure moderne promettait une résilience et une élasticité futures des applications, mais de nombreuses entreprises ont aujourd'hui du mal à maintenir les performances de leurs applications. L'automatisation de la gestion des applications modernes est essentielle pour profiter des avantages de la résilience et de l'élasticité. Sans automatisation, on ne peut tirer parti des avantages offerts par la configuration des bons SLO et la collecte continue des données. Les équipes chargées de l'informatique et des plateformes ne peuvent pas garantir les performances applicatives sans automatiser l'allocation dynamique de ressources. Entre le déclenchement d'une alerte et l'allocation manuelle des ressources, la dégradation des performances a déjà eu lieu.

## En fin de compte, le problème avec cette approche est qu'elle repose sur le fait qu'un problème s'est déjà produit.

L'automatisation va plus loin qu'une simple réaction à un événement déclenché par le franchissement d'un seuil. En s'appuyant sur la collecte continue de données et en identifiant les principaux SLI et SLO pour leur secteur d'activité, les entreprises peuvent faire en sorte que les logiciels génèrent des décisions exploitables et exécutées automatiquement. Si votre entreprise prend les mesures nécessaires et s'engage dans l'automatisation, elle crée un environnement véritablement élastique qui gère de manière proactive les ressources de vos applications et garantit des performances stables. Ce type d'automatisation nécessite un système intelligent qui analyse un environnement changeant de manière dynamique et qui automatise la chaîne de décisions nécessaire pour traiter les problèmes **avant** que la dégradation des performances ne survienne. Il est impossible d'atteindre ce niveau de performance et d'élasticité avec un processus basé sur des feuilles de calcul et des alertes.

Pour s'éloigner d'une approche réactive, il faut une adhésion totale des parties prenantes des différentes équipes en charge des applications et de l'infrastructure. Les chefs d'applications et de produits hésitent souvent à céder le contrôle de leurs applications à l'automatisation. Cette réticence est due à un manque de confiance à l'égard l'automatisation. Mais elle peut être surmontée. L'automatisation exige un changement culturel de votre équipe informatique. Pour atteindre une véritable élasticité et résilience, les entreprises doivent pouvoir faire confiance aux actions qui seront automatisées. Avec des SLO pertinents centrés sur l'entreprise et les applications, qui sont directement liés à l'allocation dynamique des ressources dans la plateforme et l'infrastructure, il est plus facile pour les chefs d'applications et de produits de faire confiance à l'automatisation et d'être à l'aise avec l'idée que l'automatisation sera entièrement mise en œuvre dans les flux de travaux de l'entreprise.

## Points à retenir

Au fur et à mesure que les entreprises adopteront des infrastructures agiles, telles que l'informatique hybride, les plateformes de conteneurs et le multicloud, les applications gagneront en complexité. Face à ces futurs développements, les entreprises qui répondront aux attentes accrues en matière d'expérience client et de performances applicatives sont celles qui prospéreront. La meilleure approche à suivre consiste à définir les SLO en termes d'expérience client, à analyser en permanence les demandes applicatives changeantes et à automatiser l'allocation dynamique de ressources. Afin de réussir à l'ère du numérique, les entreprises doivent prendre les mesures nécessaires pour mettre en œuvre une automatisation permettant d'allouer dynamiquement les ressources aux applications en vue de satisfaire les besoins changeants et les SLO.

# Atteinte automatique des SLO d'une application avec Turbonomic.

**Turbonomic transforme les données en action, en offrant une automatisation qui prévient les risques liés aux performances applicatives tout en maximisant l'élasticité.**

La modernisation des applications et des infrastructures critiques est un investissement qui offre de nombreux avantages. Mais pour profiter des avantages de l'élasticité, de la résilience et de la rapidité de commercialisation, vous avez besoin d'un logiciel qui analyse en permanence votre environnement et exécute la bonne décision d'allocation de ressources au bon moment afin de garantir les performances applicatives. Avec Turbonomic, vous pouvez corréliser le temps de réponse des applications, le débit des transactions ou d'autres SLI/SLO à l'allocation dynamique des ressources. À mesure que la demande fluctue, l'allocation dynamique des ressources de Turbonomic assure une performance continue des applications.

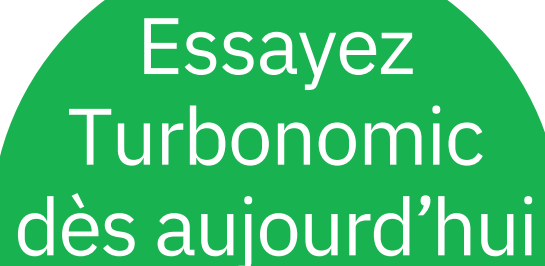
**L'HPA n'en est pas capable.** Turbonomic analyse la pile de bout en bout et en détail pour garantir dynamiquement vos SLO. Définissez vos SLO et notre logiciel basé sur l'IA veille à ce que la plateforme et l'infrastructure associées fournissent les ressources nécessaires pour atteindre vos SLO, quel que soit le lieu d'exécution de vos applications.

**Intégration transparente dans les flux de travaux de l'entreprise.** L'intégration de Turbonomic avec webhook permet d'injecter facilement des actions Turbonomic dans les cycles de vie des applications, dans les pipelines DevOps et d'infrastructure, dans les flux d'approbation et d'audit, et dans les processus de communication.

**Réduisez au minimum le travail manuel :** les développeurs, DevOps et SRE n'ont pas besoin de définir des seuils, des contraintes ou des politiques de mise à l'échelle automatique. Le logiciel prend des décisions en toute les bonnes décisions en matière de ressources à votre place, en proposant des actions que vous pouvez réellement automatiser.

**Évitez les dépenses excessives en capacité :** il n'est pas nécessaire de s'en remettre aux développeurs pour prendre des décisions en matière de ressources. Il est d'ailleurs courant que ceux-ci surprovisionnent, par prudence. Notre logiciel détermine exactement les ressources dont les services ont besoin, en fonction de la demande de l'application.

**Planifiez rapidement et facilement votre croissance :** simulez la mise en œuvre de nouveaux services avec notre logiciel. Déterminez exactement ce dont vous aurez besoin pour soutenir votre croissance.



Essayez  
Turbonomic  
dès aujourd'hui

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

© Copyright IBM Corporation 2022  
Compagnie IBM France  
17 avenue de l'Europe  
92275 Bois-Colombes Cedex

Produit aux États-Unis d'Amérique  
Janvier 2022

IBM et le logo IBM sont des marques commerciales ou des marques déposées d'International Business Machines Corporation, aux États-Unis et/ou dans d'autres pays. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés. Une liste actualisée des marques d'IBM est disponible sur [ibm.com/trademark](https://ibm.com/trademark).

L'information contenue dans ce document était à jour à la date de sa publication initiale et peut être modifiée sans préavis par IBM. Les offres mentionnées dans le présent document ne sont pas toutes disponibles dans tous les pays où la société IBM est présente.

Les données de performance et les exemples de client cités sont présentés à titre informatif uniquement. Les résultats de performances réels peuvent varier en fonction des configurations et des conditions de fonctionnement spécifiques.

LES INFORMATIONS CONTENUES DANS LE PRÉSENT DOCUMENT SONT FOURNIES « EN L'ÉTAT », SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DÉCLINE NOTAMMENT TOUTE RESPONSABILITÉ RELATIVE À CES INFORMATIONS EN CAS DE CONTREFAÇON AINSI QU'EN CAS DE DÉFAUT D'APTITUDE À L'EXÉCUTION D'UN TRAVAIL DONNÉ. Les produits IBM sont garantis conformément aux dispositions des contrats qui régissent leur utilisation.

