Augmenter la résilience et améliorer l'expérience client en utilisant l'ingénierie du chaos sur AWS

AWS Conseils prescriptifs

AWS Conseils prescriptifs: Augmenter la résilience et améliorer l'expérience client en utilisant l'ingénierie du chaos sur AWS

Table of Contents

Augmenter la résilience et améliorer l'expérience client en utilisant l'ingénierie du chaos sur AWS

Laurent Domb, technologue en chef, finances fédérales, Amazon Web Services

Avril 2025 (historique du document)

L'ingénierie du chaos est la discipline qui consiste à expérimenter sur une application afin de renforcer la confiance dans la capacité de votre organisation et de votre application à résister à des conditions de production turbulentes. Il s'agit d'une approche proactive de la résilience, dont le but est de vérifier si votre application et votre organisation sont en mesure d'absorber les défaillances du service, de s'y adapter et, finalement, de s'en remettre en place en introduisant des défaillances contrôlées chez les personnes, les processus et les technologies. L'objectif est également d'identifier et d'éliminer les faiblesses avant qu'elles ne provoquent des interruptions ou d'autres perturbations de la production.

Chez Amazon, nous sommes conscients que les défaillances sont inévitables dans les systèmes distribués, à tel point que le fonctionnement malgré la présence de défaillances est un mode de fonctionnement normal. Les interactions entre les services étant vouées à l'échec, vous devez comprendre comment vos services réagissent en fonction des différents modes de défaillance et créer des services résistants aux principales vulnérabilités telles que les défaillances de dépendance, les tempêtes de nouvelles tentatives, les zones de disponibilité altérées et l'épuisement des ressources de l'hôte.

Prenons l'exemple d'une nouvelle tentative de tempête. Une défaillance localisée chez un client peut avoir un impact significatif sur plusieurs services. C'est ce que l'on appelle communément l'effet papillon. Une tempête de nouvelles tentatives est une manifestation de l'effet papillon : une dépendance défaillante incite les clients, et les clients de ces clients, à retenter l'opération qui a échoué, ce qui entraîne une croissance exponentielle du trafic. Les services sont surchargés car ils doivent répondre au trafic normal en plus de réessayer le trafic tout en gérant une dégradation des performances.

L'ingénierie du chaos est apparue en réponse à la complexité croissante des systèmes distribués. Il s'agit d'une approche multidisciplinaire qui combine les principes de la théorie du chaos, de la pensée systémique et de l'ingénierie pour concevoir et gérer des systèmes complexes résilients aux événements et aux comportements inattendus. À la base, l'ingénierie du chaos vise à comprendre et à gérer le comportement de systèmes complexes dans des conditions d'incertitude et d'imprévisibilité.

Il reconnaît que les approches traditionnelles de l'ingénierie, qui reposent sur la prévision et le contrôle des résultats, sont souvent insuffisantes pour faire face à la nature complexe et dynamique des systèmes distribués. À mesure que ces systèmes se développent, ils dépassent souvent le champ de compréhension de chaque individu.

L'ingénierie du chaos fournit des concepts, des techniques et des outils permettant d'injecter intentionnellement des défaillances dans les systèmes afin de découvrir les faiblesses avant qu'elles ne se manifestent en production. Cette approche proactive permet aux entreprises de s'assurer que leurs systèmes fonctionneront dans des conditions stressantes. Bien que l'ingénierie du chaos soit encore une pratique en évolution, elle représente une évolution fondamentale vers la conception, la gestion et l'exploitation de systèmes informatiques modernes afin d'être résilients face à la complexité et à l'interconnexion croissantes.

Les sections suivantes de ce guide présentent les avantages de l'ingénierie du chaos, expliquent comment mener des expériences d'ingénierie du chaos et décrivent les approches que vous pouvez adopter pour mettre en œuvre l'ingénierie du chaos à grande échelle dans votre organisation. Sont également inclus des exemples de documents de planification d'expériences et de résultats d'expériences que vous pouvez utiliser comme modèles pour vos expériences d'ingénierie du chaos.

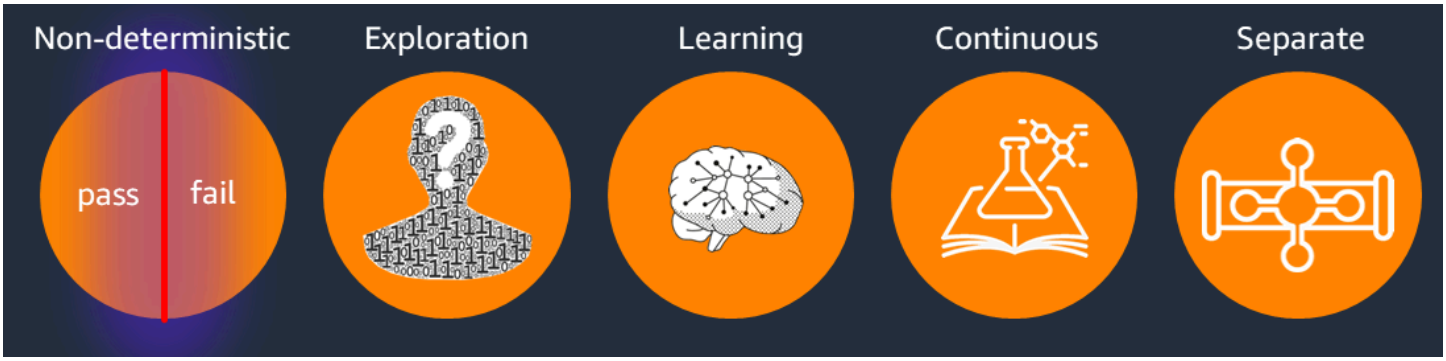
La section suivante explore en quoi les caractéristiques de l'ingénierie du chaos diffèrent des tests de résilience traditionnels tels que les tests unitaires, de fumée ou d'intégration.

Présentation de

Comparaison entre les tests de résilience et l'ingénierie du chaos

Les tests de résilience sont déterministes. En d'autres termes, il valide les caractéristiques connues des mécanismes de résilience, tels que les disjoncteurs, les nouvelles tentatives, les basculements ou les solutions de repli, qui ont été implémentés dans votre application. Cela confirme la manière dont ces composants de l'application absorbent les perturbations contrôlées avec un impact minimal, voire nul, sur les utilisateurs. Par conséquent, les tests de résilience se concentrent sur la validation des modes de défaillance connus qui sont injectés dans les composants de l'application dans le but de produire des pass/fail résultats. Vous devez effectuer des tests de résilience en continu dans le cadre de votre pipeline afin de vous assurer de ne pas introduire de régressions dans votre posture de résilience. Dans le cadre des tests de résilience, il est fréquent que vous n'exécutiez pas de tests sur des composants réels, mais sur des API fictives qui simulent un certain composant. Cette approche permet de tester des scénarios de défaillance de manière cohérente et reproductible dans un environnement contrôlé, ce qui la rend adaptée à l'intégration automatisée des pipelines et aux tests de régression.

En revanche, l'ingénierie du chaos n'est pas déterministe. En d'autres termes, il repose sur des hypothèses et vérifie votre modèle mental sur la manière dont l'application et ses dépendances (personnes, processus et technologie) absorbent, s'adaptent et finissent par se rétablir après des défaillances imprévues. Par conséquent, l'ingénierie du chaos se concentre sur la vérification de bout en bout des modes de défaillance inconnus, dans le but de détecter les défauts à un stade précoce et de les corriger avant qu'ils ne se transforment en événements de grande envergure. L'ingénierie du chaos favorise l'apprentissage continu et doit être mise en pratique par le biais d'un pipeline de chaos distinct ou d'expériences ad hoc qui vous permettent d'exécuter plusieurs expériences à tout moment sans bloquer la productivité de votre développeur lors du déploiement du code.



Le processus d'ingénierie du chaos commence souvent par une journée de jeu chaotique, un événement dédié au cours duquel les équipes injectent intentionnellement des défauts ou des défaillances contrôlés dans leur application. La journée de jeu est progressive : elle commence dans des environnements de niveau inférieur (tels que le développement ou les tests) et passe progressivement à des environnements de niveau supérieur (tels que la mise en scène et la préproduction) au fur et à mesure que la confiance augmente. En parcourant systématiquement ces environnements, les équipes peuvent vérifier que leurs systèmes tolèrent correctement les défauts injectés avant leur mise en production. Cette progression méthodique garantit qu'au moment où les expériences de chaos sont menées dans les environnements de production, les équipes ont acquis une confiance substantielle dans les capacités de résilience de leur système. Le processus du jour du match est une approche proactive visant à identifier les faiblesses et les vulnérabilités de l'architecture et des pratiques opérationnelles d'une application, tout en éliminant le stress lié à l'apprentissage lors d'une interruption de production imprévue.

La valeur de l'ingénierie du chaos

Les systèmes complexes sont omniprésents dans le monde d'aujourd'hui. Ils jouent un rôle essentiel dans de nombreux aspects de notre vie, des marchés financiers aux soins de santé. Nous nous attendons à ce que ces systèmes soient toujours opérationnels. Cependant, les systèmes complexes sont souvent vulnérables à des événements et à des comportements inattendus qui peuvent avoir des conséquences importantes. Organisations doivent planifier les perturbations au lieu de se demander si elles vont se produire. Ils peuvent le faire en appliquant des tests de scénarios à l'ensemble de leurs services commerciaux critiques ou stratégiques. C'est là que l'ingénierie du chaos entre en jeu.

L'ingénierie du chaos propose une approche pour gérer des systèmes complexes qui peut aider à atténuer les risques et à améliorer la résilience. Le processus de préparation des expériences sur le chaos oblige les équipes à élaborer des hypothèses sur le comportement de leur système,

ce qui leur permet de mieux comprendre comment les systèmes sont construits et comment ils fonctionnent. Cette préparation révèle souvent des lacunes mentales, des connaissances architecturales et des connaissances opérationnelles qui, autrement, ne seraient peut-être pas découvertes. En approfondissant la compréhension de la façon dont les systèmes complexes réagissent aux défaillances, l'ingénierie du chaos favorise une plus grande transparence et une plus grande responsabilité dans la conception et la gestion des systèmes. Plus votre organisation pratique fréquemment l'ingénierie du chaos, mieux elle est préparée sur le plan opérationnel. L'ingénierie du chaos vous aide à établir les meilleures pratiques pour concevoir des applications résilientes capables de résister aux défaillances des composants avec un impact minimal, voire nul, sur les utilisateurs. Cela garantit que les applications critiques fonctionnent dans les limites des niveaux de service et de tolérance aux impacts attendus, tout en améliorant continuellement les connaissances de l'équipe sur ses propres systèmes.

Se préparer à des conditions défavorables

Lorsque vous vous basez sur AWS, vous utilisez différents types de services, notamment des services zonaux tels qu'Amazon Elastic Compute Cloud (Amazon EC2), des services régionaux tels qu'Amazon Simple Storage Service (Amazon S3), des services mondiaux tels que Gestion des identités et des accès AWS (IAM), des services tiers en tant que service (SaaS) et des services sur site. Chaque type de service expose différents domaines de défaillance dont vous devez tenir compte. Comment vous préparez-vous à faire face à des événements auto-infligés ou à des événements provoqués par des tiers sur lesquels votre organisation n'a aucun contrôle ?

Pour mieux comprendre comment votre application peut réagir à des conditions défavorables, vous pouvez utiliser AWS Fault Injection Service (AWS FIS). AWS FIS est un service entièrement géré permettant d'exécuter des expériences d'injection de défauts de manière contrôlée. Vous pouvez utiliser ce service pour injecter des scénarios AWS fournis, tels que des interruptions de courant dans la zone de disponibilité et des problèmes de connectivité entre régions, ou pour créer vos propres expériences en regroupant une grande variété d'actions de défaillance fournies par le service. AWS FIS permet à vos équipes de s'entraîner en permanence et d'apprendre comment leur application réagirait aux défaillances courantes et corrigerait les défauts au fur et à mesure qu'elles les détectent.

Pratiquer l'ingénierie du chaos contrôlée

Les principes clés des expériences de chaos contrôlé sont les suivants :

- Commencez par un environnement similaire à votre environnement de production.

- Établissez une hypothèse et arrêtez les conditions de votre expérience.
- Commencez à petite échelle.
- Prenez le contrôle de vos expériences sur le chaos.
- Définissez l'étendue de l'impact.
- Connaissez votre base de service.
- Planifiez des expériences.
- Corrigez d'abord, puis expérimentez.
- Surveillez l'expérience de près.
- Tirez des leçons de vos résultats.
- Hiérarchisez les résultats, corrigez et vérifiez.
- Diffusez les connaissances acquises au sein de votre organisation.

Pour réussir à étendre l'ingénierie du chaos, vous devez mettre en œuvre des expériences sur le chaos de manière contrôlée. Lorsque vous l'utilisez AWS FIS, vous pouvez créer des conditions d'arrêt à l'aide des CloudWatchalarmes Amazon. Vous pouvez intégrer ces conditions dans un modèle d'expérience pour vous assurer que les expériences sont arrêtées en cas de dépassement des limites et ramenées à leur dernier état connu. AWS FIS fournit également des leviers de sécurité. Lorsque vous actionnez ces leviers, toutes les expériences en cours dans le compte sont arrêtées et AWS FIS annulées Région AWS, y compris les expériences multi-comptes, et empêche le démarrage de nouvelles expériences. Cela empêche l'injection de défauts pendant certaines périodes, telles que les heures de négociation, les événements commerciaux ou les lancements de produits, ou en réponse à des alarmes relatives à l'état d'une application. Le levier de sécurité reste enclenché jusqu'à ce qu'il soit débrayé manuellement.

Lorsque vous menez une expérience chaotique, vous devez définir des mesures de protection pour éviter les effets secondaires indésirables dans l'environnement, en particulier s'il est possible que l'expérience affecte les applications en production. Lorsque vous planifiez l'expérience, anticipez les effets indésirables qu'elle pourrait avoir sur les autres applications de l'environnement. Par exemple, d'autres applications peuvent recevoir des messages erronés de la part de l'application participant à l'expérience, recevoir des volumes de demandes élevés ou rencontrer des problèmes de ressources si elles partagent une infrastructure. Documentez ces risques et corrigez tout problème connu ou inacceptable avant de lancer l'expérience.

Débuter avec l'ingénierie du chaos

Avant de réaliser une expérience, nous vous recommandons de mettre en place quelques éléments essentiels pour tirer le meilleur parti de vos pratiques d'ingénierie du chaos. Ces éléments essentiels incluent :

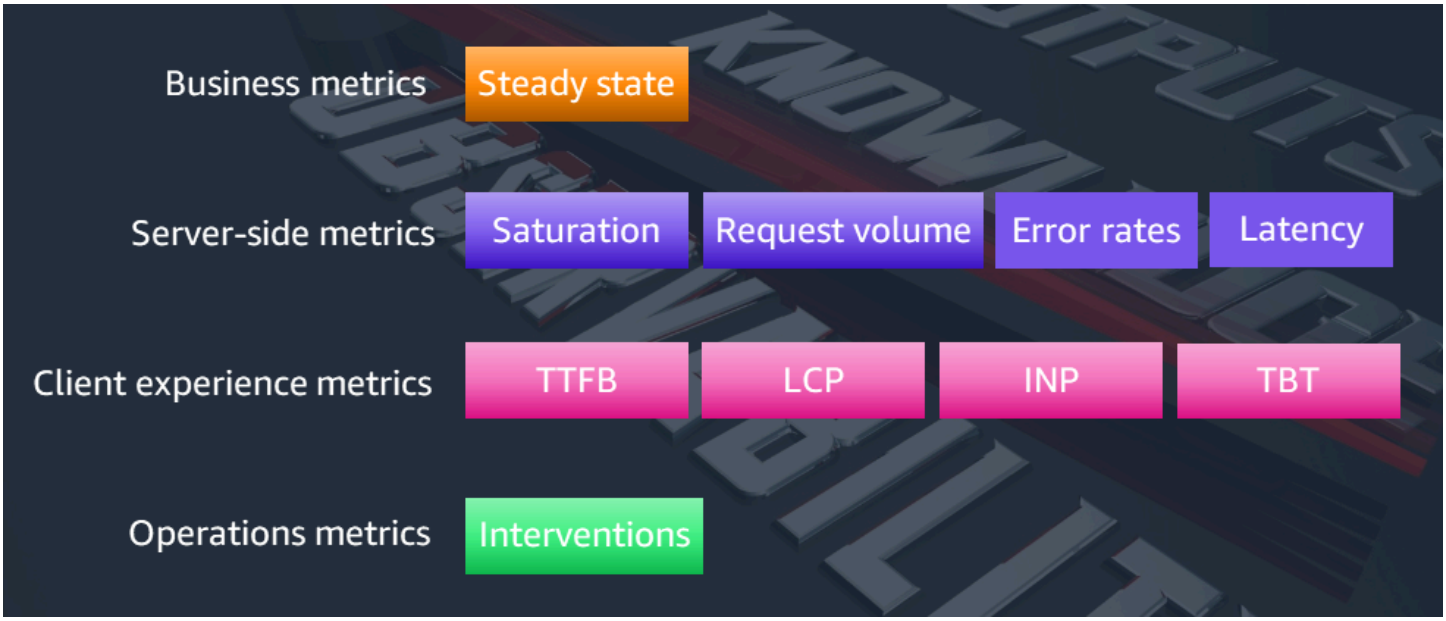
- Observabilité (métriques, journalisation, suivi des demandes)
- Une liste d'événements ou de failles du monde réel que vous aimeriez explorer
- Parrainage de la résilience organisationnelle grâce à l'adhésion des dirigeants
- Priorisation des résultats critiques, en fonction de l'impact commercial potentiel, par rapport aux nouvelles fonctionnalités découvertes lors d'expériences chaotiques

Observabilité pour les expériences sur le chaos

L'observabilité, qui comprend les métriques, la journalisation et le suivi des demandes, joue un rôle clé dans l'ingénierie du chaos. Lorsque vous réalisez un test, vous devez comprendre les indicateurs commerciaux, les indicateurs côté serveur, les indicateurs d'expérience client et les indicateurs opérationnels. Sans observabilité, vous ne serez pas en mesure de définir un comportement en régime permanent ou de créer une expérience significative pour vérifier si votre hypothèse concernant votre application est vraie.

Métriques

Le schéma suivant montre les types de métriques que vous pouvez suivre pour des expériences de chaos pour différents types d'applications.



- Indicateurs commerciaux — L'état d'équilibre indique le fonctionnement normal de votre système et est défini par vos indicateurs commerciaux. Il peut être représenté par des transactions par seconde (TPS), des flux de clics par seconde, des commandes par seconde ou une mesure similaire. Votre application présente un état d'équilibre lorsqu'elle fonctionne comme prévu. Vérifiez donc que votre application est saine avant de lancer des tests. L'état d'équilibre ne signifie pas nécessairement qu'il n'y aura aucun impact sur l'application en cas de panne, car un pourcentage de défauts peut se situer dans les limites acceptables. L'état d'équilibre est votre point de référence. Par exemple, l'état d'équilibre d'un système de paiement peut être défini comme le traitement de 300 TPS avec un taux de réussite de 99 % et un temps aller-retour de 500 ms. Visuellement, imaginez l'état d'équilibre comme un électrocardiogramme (ECG). Si l'état d'équilibre de votre système fluctue soudainement, vous savez qu'il y a un problème avec votre service.
- Server-side métriques — Pour comprendre les performances de vos ressources pendant le test, vous avez besoin d'informations sur leurs performances avant, pendant et après le test. Pour mesurer l'impact de vos ressources sur AWS, vous pouvez utiliser Amazon CloudWatch. CloudWatch est un service qui surveille les applications, répond aux changements de performances, optimise l'utilisation des ressources et fournit des informations sur l'état des opérations. Au cours de vos expériences, vous souhaiterez capturer des métriques côté serveur telles que la saturation, les volumes de demandes, les taux d'erreur et la latence.
- Indicateurs d'expérience client — Activé AWS, vous pouvez capturer des indicateurs réels des utilisateurs en utilisant CloudWatchRUM pour simuler les demandes des utilisateurs à l'aide d'outils tels que Locust, Grafana k6, Selenium ou Puppeteer. Les indicateurs réels des utilisateurs sont essentiels pour les organisations qui mènent des expériences d'ingénierie du chaos. En surveillant

l'impact des utilisateurs réels au cours d'une expérience, les équipes peuvent avoir une idée précise de la manière dont les pannes et les perturbations affecteront les clients en production. Les indicateurs clés de l'expérience client sont le délai jusqu'au premier octet (TTFB), le Largest Contentful Paint (LCP), l'interaction avec le prochain dessin (INP) et le temps de blocage total (TBT).

- Mesures opérationnelles : les interventions mesurent la capacité à atténuer les défaillances de manière automatisée, par exemple, la latence réussie des demandes des clients lors du redémarrage de pods, de tâches ou d'instances EC2 à l'aide de mécanismes tels que le contrôleur de réplication ou le dimensionnement automatique. La possibilité d'intervenir automatiquement en cas de panne est directement liée à une bonne expérience utilisateur. Il est essentiel de comprendre si ces mécanismes d'atténuation évoluent au fil du temps. En définissant des indicateurs pour les mesures d'atténuation automatisées réussies et échouées, vous créez des repères qui aident à identifier les régressions potentielles dans l'ensemble de votre système.

Logging

La journalisation centralisée est essentielle pour comprendre l'état des composants de votre application avant, pendant et après une expérience chaotique. Nous vous recommandons de collecter les journaux de tous les composants de votre application afin de créer une vue consolidée de ce que faisait chaque composant au moment de l'injection de l'expérience. Cela fournit une image claire du flux d'expériences de bout en bout.

Suivi des demandes

Le suivi des demandes vous permet d'observer le flux de chaque demande entre les composants de votre application afin de mieux comprendre l'impact de la panne injectée sur le système et ses dépendances. En suivant les demandes, vous pouvez voir comment la panne se propage entre les différents services et composants, afin de mieux évaluer l'ampleur de la perturbation. Pour suivre vos demandes AWS, vous pouvez utiliser AWS X-Ray.

Scénarios d'échec à intégrer dans les expériences de chaos

L'objectif de l'injection de défauts courants dans votre application est de comprendre comment l'application réagit à ces événements inattendus, afin de créer des mécanismes d'atténuation et de rendre votre système résilient face à de telles défaillances. En outre, vous devez utiliser l'ingénierie

du chaos pour rejouer les scénarios de défaillance historiques afin de vérifier que vos mécanismes d'atténuation fonctionnent toujours comme prévu et n'ont pas évolué au fil du temps.

Tenez compte des événements suivants lorsque vous planifiez vos expériences d'ingénierie du chaos.

Mode de défaillance	Description
Défaillance du serveur	Redémarrez les instances EC2, supprimez les pods Kubernetes ou les tâches Amazon Elastic Container Service (Amazon ECS) pour comprendre comment votre application réagit à de tels incidents.
erreurs d’API	Injectez AWS des erreurs dans vos propres API de service pour comprendre le comportement des applications.
Problèmes liés au réseau	Introduisez de la latence ou de la congestio n, ou bloquez les connexions pour imiter les problèmes de réseau réels.
AWS Atteinte à la zone de disponibilité	Reproduisez l'altération d'une zone de disponibi lité complète pour vérifier la reprise dans toutes les zones.
Région AWS troubles de la connectivité	Reproduisez une défaillance du réseau Régions AWS pour vérifier comment les ressources de la région secondaire réagissent à un tel événement.
Défaillances de base	Basculez sur des répliques de bases de données ou des données corrompues, ou rendez les instances de base de données inaccessibles, afin de comprendre l'impact sur votre application et vos stratégies de restaurat ion.

Mode de défaillance	Description
Pause de la réplication de la base de données et Amazon S3	Interrompez la réplication de la base de données ou d'Amazon Simple Storage Service (Amazon S3) entre les zones de disponibilité Régions AWS ou pour comprendre l'impact des applications en aval.
Dégradation du stockage	Suspendez I/O, supprimez les volumes Amazon Elastic Block Store (Amazon EBS) ou supprimez des fichiers pour vérifier la durabilité et la restauration des données.
Déficience de dépendance	Diminuez ou dégradez les performances des services en aval ou en amont dont vous dépendez, y compris les services tiers, afin de comprendre le flux de bout en bout et l'impact sur vos clients.
Hausses de trafic	Générez des pics de trafic utilisateur pour tester les fonctionnalités de dimensionnement automatique et découvrez comment le temps de démarrage à froid peut avoir un impact sur l'état général de votre application.
Épuisement des ressources	Maximisez le processeur, la mémoire et l'espace disque pour vérifier la dégradation progressive de votre application.
Défaillances en cascade	Initiez les défaillances principales qui se répercutent sur les applications et les composants en aval.
Déploiements incorrects	Déployez les modifications ou les configura tions problématiques pour vérifier les mécanismes de restauration.

Parrainage de la résilience organisationnelle

L'ingénierie du chaos apporte le plus de valeur lorsqu'elle est appliquée à l'ensemble de votre organisation. Nous vous recommandons de travailler avec un sponsor exécutif qui pourra vous aider à définir des objectifs de résilience au sein de votre organisation, à éliminer les craintes, les incertitudes et les doutes concernant le domaine, et à démarrer le processus de transformation afin que la résilience soit la responsabilité de tous.

Pour étayer l'analyse de rentabilisation de la création d'un cabinet d'ingénierie du chaos, associez des efforts d'ingénierie du chaos à vos projets commerciaux critiques. Faire de la résilience un atout et un moteur d'accélération vous aidera à suivre le succès au fil du temps. Commencez par le décompte des incidents critiques par mois ou par trimestre, le délai moyen de reprise et l'impact que ces incidents ont eu sur vos clients et votre entreprise. Fixez l'objectif avec vos équipes de réduire le nombre d'incidents sur une période de 6 à 12 mois à mesure que des améliorations seront apportées à vos piles d'applications en réponse aux expériences d'ingénierie chaotiques.

Mesurez si les incidents sont très répétitifs. Supposons, par exemple, qu'un certificat TLS expiré entraîne une interruption de service car les clients ne peuvent pas établir de connexion fiable. Si plusieurs incidents se produisent au cours d'une année en raison de l'expiration de plusieurs certificats TLS, vous pouvez tester l'expiration d'un certificat TLS et vérifier que vos équipes reçoivent des alertes ou sont en mesure d'atténuer automatiquement ces problèmes. Cela vous aidera à devenir résilient face à de tels défauts.

Pour suivre les progrès de l'ingénierie du chaos au fil du temps, capturez les indicateurs suivants afin de mettre en évidence la valeur de l'ingénierie du chaos tout au long du cycle de vie d'une application :

- Taux d'incidents réduit : suivez le nombre d'incidents de production au fil du temps et associez ce chiffre à l'adoption de l'ingénierie du chaos. On s'attend à ce que le taux d'incidents graves diminue.
- Temps moyen de résolution (MTTR) amélioré : calculez le temps moyen nécessaire pour résoudre les incidents et suivez ces données pour voir si l'ingénierie du chaos s'améliore au fil du temps.
- Disponibilité accrue des applications : utilisez des indicateurs de niveau de service pour montrer les améliorations de disponibilité à mesure que la résilience des applications augmente grâce à des expériences chaotiques.

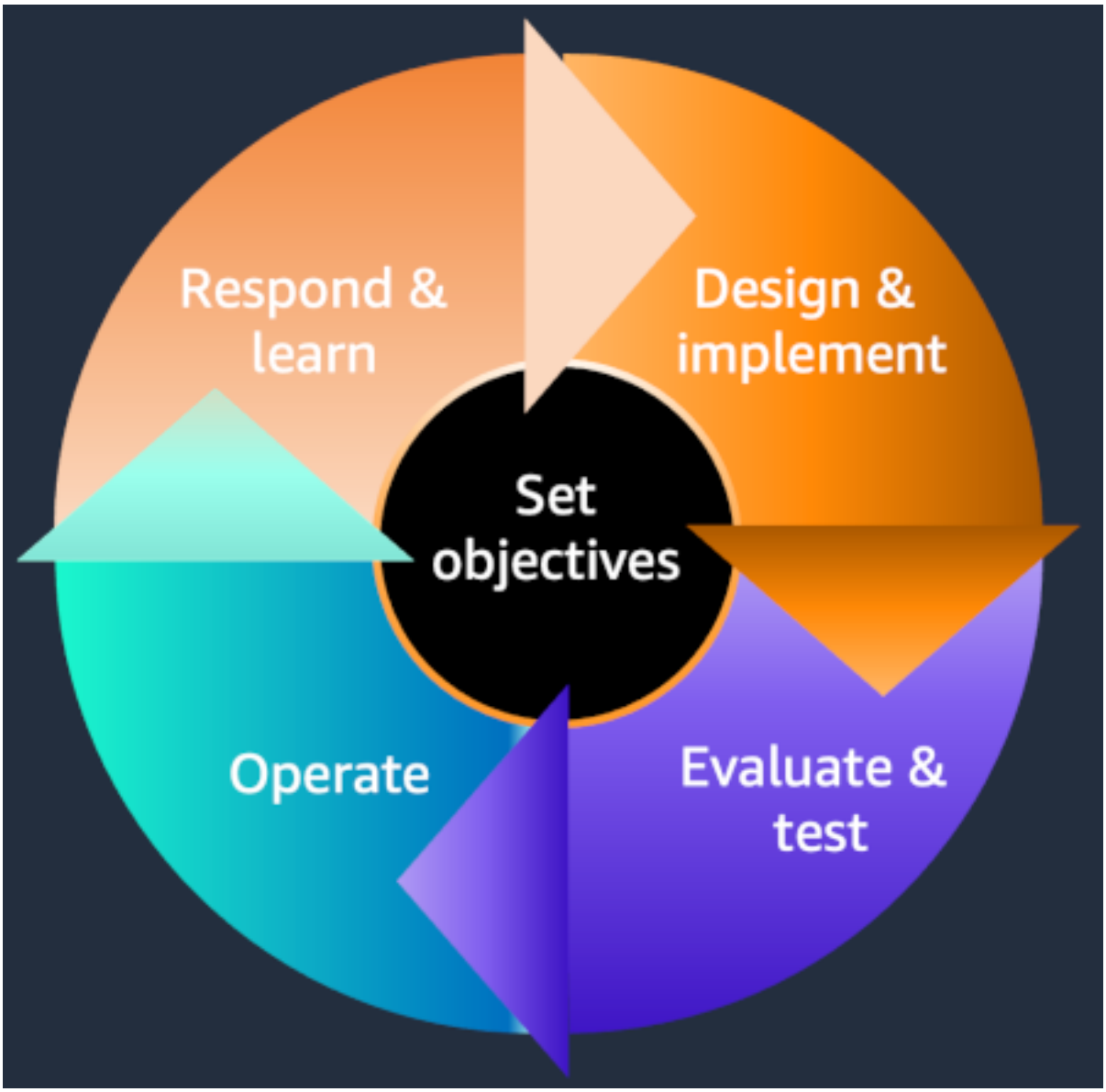
- Réduction des délais de commercialisation — L'ingénierie du chaos peut vous apporter la confiance nécessaire pour lancer des offres innovantes plus rapidement, car vous savez que vos applications sont résilientes. Suivez l'augmentation de la vitesse de lancement des produits.
- Réduction des coûts d'exploitation : quantifiez si les indicateurs tels que le bruit d'alerte, la charge opérationnelle et les efforts manuels nécessaires à la gestion des applications diminuent en raison de la mise en place de pratiques chaotiques.
- Renforcer la confiance : interrogez les développeurs, les ingénieurs de fiabilité des sites (SRE) et les autres membres du personnel technique pour déterminer si l'ingénierie du chaos a renforcé leur confiance dans la résilience des applications. Les perceptions sont importantes.
- Expériences client améliorées – Connectez l'ingénierie du chaos à des résultats positifs pour les clients, tels que la réduction des interruptions de service, des annulations et des pannes.

Prioriser les mesures correctives

Lorsque vous réalisez des expériences chaotiques, vous êtes susceptible d'identifier les domaines à améliorer dans lesquels l'application ne fonctionne pas comme prévu. La correction de ces éléments deviendra une tâche de votre carnet de travail qui devra être priorisée au même titre que d'autres tâches telles que le développement de fonctionnalités. Nous vous recommandons de prendre le temps de procéder à ces améliorations afin d'éviter de futures défaillances. Envisagez de hiérarchiser ces tâches d'apprentissage et de correction en fonction du niveau d'impact qu'elles peuvent avoir. Les résultats qui ont un impact direct sur la résilience ou la sécurité de votre application doivent avoir la priorité sur les nouvelles fonctionnalités, afin d'éviter tout impact sur les clients. Si l'équipe a du mal à prioriser les travaux de remédiation par rapport au développement de fonctionnalités, pensez à contacter votre sponsor exécutif pour vous assurer que les priorités sont définies en fonction de la tolérance au risque de l'entreprise.

Mise en œuvre de l'ingénierie du chaos sur AWS

L'ingénierie du chaos fait partie de la phase d'évaluation et de test du cycle de vie de la AWS résilience, comme illustré dans le schéma suivant. Les applications distribuées ne fonctionnent pas indépendamment des autres applications ou clients. Nous vous recommandons donc de passer en revue le cycle de vie complet de la résilience. Le changement est constant pour les applications distribuées à mesure que le réseau évolue, que les applications en amont et en aval subissent des changements et que l'utilisation des clients change au fil du temps.



Pour comprendre l'impact que ces modifications apportées à votre application peuvent avoir sur sa résilience, intégrez l'ingénierie du chaos à vos opérations quotidiennes. Vous pouvez implémenter des expériences de chaos de différentes manières :

- Ad hoc : vous pouvez réaliser des expériences de chaos sous forme d'expériences ponctuelles pour résoudre un problème ou une question spécifique.
- Chaos Game Days : il s'agit d'événements structurés et récurrents conçus pour vérifier la fiabilité et la résilience d'une application. Le but d'une journée de jeu sur le chaos est d'identifier les problèmes ou déficiences potentiels en matière de résilience chez les personnes, les processus et les technologies, et de mettre en pratique les processus et procédures permettant d'identifier, d'atténuer et de répondre aux incidents.
- Chaos Pipeline — L'intégration continue et la livraison continue (CI/CD) consistent à créer de nouvelles fonctionnalités et à les déployer en toute sécurité dans les environnements. Pour mettre en œuvre des expériences d'ingénierie du chaos, créez un pipeline de chaos distinct du CI/CD vôtre. Pour comprendre pourquoi, supposons que vous souhaitiez ajouter à votre CI/CD pipeline une seule expérience d'ingénierie du chaos qui augmente la perte de paquets pour les composants en aval. Cette expérience est exécutée 3 fois et prend 5 minutes pour se terminer à chaque fois. La perte de paquets augmente de 10 % à 20 % à 30 % à chaque exécution, et l'expérience dure 15 minutes au total. Si vous avez 100 déploiements parallèles, vous devrez attendre 1 500 minutes pour qu'une seule expérience soit terminée. Si vous avez 10 expériences à réaliser, l'impact pour vos développeurs serait insupportable. À grande échelle, l'ingénierie du chaos a besoin de son propre pipeline qui vous permet de réaliser des expériences en parallèle avec votre processus de cycle de vie de développement logiciel (SDLC).
- Déploiements Canary — Les îles Canaries fournissent un environnement de test pour les expériences de chaos. En dirigeant un faible pourcentage du trafic vers un service Canary ou en utilisant des méthodes telles que la mise en miroir du trafic ou le replay, vous pouvez vérifier la nouvelle infrastructure ou les modifications du code sans aucun impact sur votre système de production stable. Vous pouvez effectuer des tests sur le Canary et injecter des erreurs si nécessaire, car vous pouvez limiter l'impact sur l'utilisateur final.
- Expériences planifiées : vous pouvez planifier des expériences afin de vérifier les mécanismes de restauration prévisibles pour votre application. Utilisez des expériences planifiées pour rejouer des événements courants afin de comprendre comment vos systèmes peuvent se remettre d'événements tels que la mise hors service d'une instance EC2 associée à un groupe de dimensionnement automatique, le retrait d'un pod Kubernetes ou la suppression d'une tâche Amazon ECS.

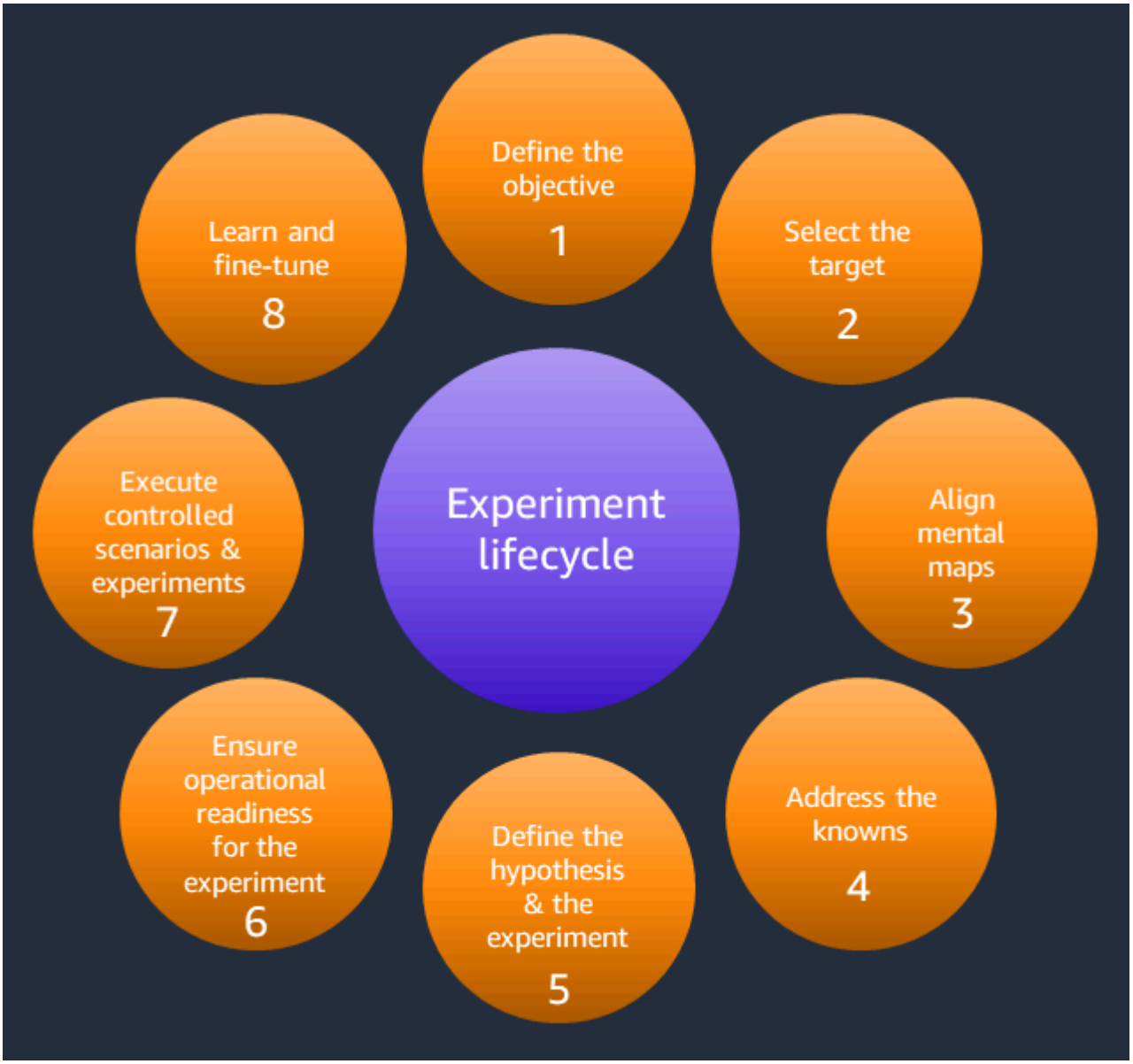
Cycle de vie continu des expériences d'ingénierie du chaos

Comme indiqué dans la section précédente, vous pouvez implémenter des expériences d'ingénierie du chaos de différentes manières. Dans tous les cas, pour créer une expérience de chaos significative, il est essentiel de comprendre l'application, les incidents historiques et les mesures correctives mises en œuvre, ainsi que de bien comprendre les domaines à étudier, tels que la résilience ou la sécurité. Votre connaissance de l'application vous aide à formuler une hypothèse sur les faiblesses potentielles de l'application et à comprendre comment elle détectera, corrigera et réparera une fois le défaut injecté.

Le cycle de vie d'une expérience sur le chaos comprend les étapes suivantes :

1. Définissez l'objectif de l'expérience.
2. Sélectionnez l'application cible.
3. Alignez les cartes mentales.
4. Résolvez les problèmes connus liés à votre application.
5. Définissez l'hypothèse et l'expérience.
6. Assurez-vous que l'expérience est prête à être opérationnelle.
7. Exécutez des scénarios et des expériences contrôlés.
8. Tirez des leçons de l'expérience et peaufinez celle-ci.

Ces étapes sont illustrées dans le schéma et décrites dans les sections suivantes.



Définissez les objectifs et les attentes

Avant chaque expérience, assurez-vous que vos objectifs et vos attentes sont spécifiques, mesurables, réalisables, pertinents et limités dans le temps. Définissez clairement ce qui suit :

- Identifiez les défaillances ou faiblesses potentielles des systèmes et des services, afin de comprendre leur impact sur les utilisateurs. Cela inclut l'identification des modes de défaillance possibles, tels que les pannes de serveur, les pannes réseau ou les bogues logiciels, et l'évaluation de leur impact potentiel sur les performances et la fiabilité globales du système.

- Quantifiez l'impact des défaillances en définissant des indicateurs de risque clés (KRI) sur vos systèmes et services. Cela inclut la mesure de l'effet des défaillances lorsque des indicateurs tels que la latence, le débit et les taux d'erreur s'écartent de leur état d'équilibre. En comprenant l'impact de tels écarts, vous pouvez hiérarchiser les efforts visant à atténuer les défaillances en fonction des risques commerciaux.
- Développez et vérifiez des stratégies pour atténuer ou prévenir les défaillances. Cela inclut l'identification de solutions potentielles, telles que la redondance, la correction d'erreurs ou les stratégies de repli, et le test de leur efficacité dans un environnement contrôlé. En vérifiant ces stratégies, vous pouvez vous assurer que vous êtes efficace pour prévenir ou atténuer les défaillances, et que vous pouvez les déployer dans vos systèmes de production en toute confiance.
- Améliorez les processus de réponse aux incidents et de reprise après sinistre. En reproduisant les défaillances dans un environnement contrôlé, vous pouvez tester les processus de réponse aux incidents, identifier les goulets d'étranglement ou les lacunes potentiels et affiner les procédures de reprise après sinistre. Cela permet de garantir que vous êtes prêt à réagir rapidement et efficacement en cas de panne imprévue.

Sélectionnez l'application cible

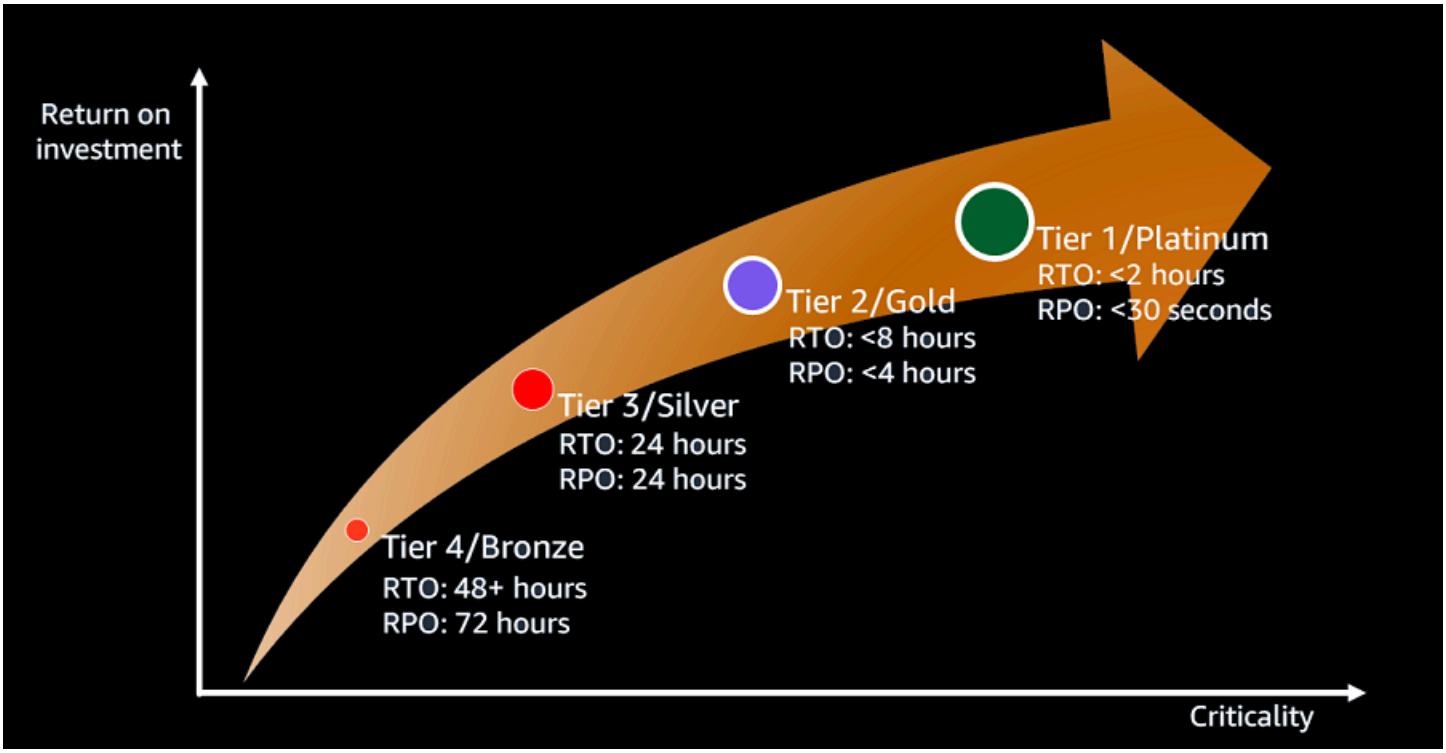
L'ingénierie du chaos est une technique puissante, mais elle nécessite une hiérarchisation réfléchie pour maximiser la valeur. Lorsque vous décidez où concentrer les efforts d'ingénierie du chaos, commencez par prendre en compte les services essentiels de votre entreprise. Demandez à vos équipes de suivre les étapes du cycle de vie du développement logiciel et de commencer par injecter des défauts dans les environnements de test. Business-critical les applications sont directement liées au chiffre d'affaires, à l'expérience client et aux activités principales. Les expériences chaotiques menées sur ces services peuvent révéler des vulnérabilités susceptibles d'avoir de graves répercussions sur l'entreprise, voire sur des marchés entiers, si elles ne sont pas corrigées. Par exemple, concentrez-vous d'abord sur les services destinés aux clients, tels que les systèmes de négociation ou les systèmes de commande. La priorisation de ces services centraux offre la meilleure protection par investissement de temps.

Après les services critiques, examinez les composants fondamentaux tels que les bases de données, les files d'attente de messages, les réseaux et les API de services partagés. Ils peuvent être utilisés en tant que composants ou services partagés au sein de votre organisation, de sorte que leur défaillance entraînera des problèmes généralisés. Confirmer la résilience des services d'infrastructure garantit qu'ils ne paralyseront pas les applications dépendantes situées au-dessus d'eux. Par exemple, une expérience d'ingénierie du chaos qui détruit un cluster Kafka en dit long sur

la tolérance aux pannes des applications en aval. Bien que l'infrastructure du système ne soit pas directement orientée vers le client, elle constitue une cible privilégiée de l'ingénierie du chaos.

N'oubliez pas de cartographier les lacunes mentales liées aux personnes, aux processus, aux informations sur les installations et aux dépendances avec des tiers, car elles peuvent provoquer des perturbations majeures si elles ne sont pas conformes aux objectifs de tolérance aux impacts de votre organisation. Pour plus d'informations sur la mesure du retour sur investissement de l'ingénierie du chaos, consultez la section Quantifier le retour sur investissement de l'ingénierie du chaos dans le document de stratégie Investir dans l'ingénierie du chaos en tant que nécessité stratégique.

Le schéma suivant montre le retour sur investissement lié à l'exécution d'expériences de chaos sur différents niveaux de services.



Aligner des cartes mentales (découverte d'applications)

Lorsque vous lancez des expériences ad hoc ou des journées de jeu, vous commencez le processus de découverte des applications en organisant une session de tableau blanc qui vise à cartographier les détails de votre application. (Si vous exécutez les expériences dans le pipeline du chaos, vous aurez déjà aligné cette carte mentale en définissant l'application cible.) Une bonne approche pour comprendre les lacunes mentales consiste à demander au membre le plus subalterne de l'équipe de dessiner d'abord un schéma de l'application, puis de demander aux membres du personnel plus

expérimentés d'y ajouter progressivement des éléments. Cela permettra de mettre en évidence toute lacune dans la compréhension entre les niveaux d'expérience.

Le diagramme doit décrire à la fois les dépendances directes en amont et en aval de l'application, ainsi que toutes les intégrations tierces critiques. Assurez-vous que le flux attendu d'une demande par le biais de l'application est aligné. Cartographiez les principaux flux de travail et les parcours des utilisateurs pour mieux comprendre comment les clients utilisent l'application. Envisagez d'utiliser un diagramme de séquence pour saisir ces informations.

Après cette session collaborative, l'équipe devrait disposer d'un modèle mental commun de l'application, de ses dépendances critiques et de ses capacités de surveillance, ainsi que d'une compréhension des risques pour prendre une décision éclairée de poursuivre ou d'annuler une expérience de chaos potentielle.

Résoudre les problèmes connus liés à votre application

Les expériences d'ingénierie du chaos sont conçues pour détecter de manière proactive les défauts d'une application. En injectant des défaillances telles que des augmentations de latence, des redémarrages de serveurs ou des pannes d'alimentation dans la zone de disponibilité, vous pouvez vérifier la capacité de votre application à tolérer des perturbations réalistes. Toutefois, ce processus suppose un niveau sous-jacent de stabilité et de santé dans l'application cible. Mener des expériences chaotiques sur une application déjà problématique risque de masquer des problèmes plus profonds.

Avant d'entreprendre une ingénierie du chaos, les équipes doivent résoudre tous les défauts, bogues et problèmes de performance connus de leur application.

Définir l'hypothèse et l'expérience

Les incidents passés qui ont perturbé votre application ou d'autres applications au sein de votre organisation peuvent constituer d'excellentes sources d'idées d'expériences chaotiques. Par exemple, les pannes précédentes ont-elles été déclenchées par des erreurs de configuration ou par l'absence de modèles de résilience ? L'examen de l'historique des incidents et la redécouverte des causes profondes de ces défaillances réelles par le biais d'expériences chaotiques constituent un moyen efficace de développer la résilience face à des problèmes similaires à l'avenir.

Une autre source précieuse de concepts d'expérimentation peut provenir directement des ingénieurs, des architectes et des opérateurs qui connaissent le mieux une application. Permettre aux membres

de l'équipe de soumettre des scénarios de défaillance hypothétiques qui, selon eux, pourraient perturber considérablement l'application vous permet de recueillir des idées basées sur des connaissances privilégiées. L'équipe chargée de l'application peut ensuite évaluer lequel de ces scénarios proposés est susceptible d'avoir l'impact potentiel le plus important ou d'exposer les plus grands risques inconnus. Cibler les expériences de chaos pour des scénarios aussi risqués et mal compris peut générer des enseignements importants et prévenir des problèmes à l'avenir.

Une troisième source d'idées provient de la modélisation de la résilience pour anticiper les conditions susceptibles d'entraîner des pertes commerciales identifiées. Certains exercices de modélisation de la résilience ont une approche basée sur les composants pour créer un modèle de résilience, tandis que d'autres ont une approche basée sur les systèmes. Une approche basée sur les composants pose la question suivante : « Que se passe-t-il lorsque le composant x est soumis à une charge extrême ou est défaillant ? » L'équipe qui développe le modèle de résilience émet ensuite des hypothèses sur l'effet d'un tel scénario sur l'application au sens large et identifie les contrôles de surveillance et de prévention actuellement en place pour détecter et atténuer les effets du scénario. Une approche basée sur les systèmes suit un processus descendant pour mettre en évidence un état indésirable de l'application, tel que « La vitrine Web affiche des niveaux d'inventaire périmés », et invite l'équipe chargée de l'application à anticiper la ou les conditions susceptibles de provoquer un tel comportement de l'application.

Garantir l'état de préparation opérationnelle pour l'expérience

Vous avez besoin d'indicateurs quantifiables pour mesurer l'impact des conditions défavorables sur l'application et son comportement, comme décrit précédemment dans la section consacrée à l'observabilité. Le fait de pouvoir mesurer le comportement de l'application vous permet de déterminer si les conditions défavorables ont eu un impact sur l'application et dans quelle mesure.

La meilleure façon de déterminer si votre application a un impact est de mesurer son état d'équilibre. Le régime permanent mesure à quoi ressemble le fonctionnement normal et s'aligne généralement sur les indicateurs commerciaux et d'expérience client pour une application donnée. Avant de passer à l'étape suivante, assurez-vous de disposer de l'observabilité nécessaire pour comprendre l'impact et de disposer de mécanismes de réduction au cas où l'expérience ne se déroulerait pas comme prévu.

Exécutez des expériences et des scénarios contrôlés

Chez AWS, nous vous déconseillons de mener une première expérience de chaos sur une application en cours de production. Le but d'une expérience de chaos est d'apprendre quelque chose de nouveau sur le comportement de l'application dans des conditions stressantes. Le comportement de l'application peut être imprévisible au cours de l'expérience, de sorte que la réalisation d'une expérience pour la première fois en production peut avoir un impact sur le client. Par conséquent, vous devez toujours exécuter une première expérience de chaos dans un environnement de niveau inférieur présentant un risque minimal d'affecter les utilisateurs du monde réel, puis parcourir vos environnements après avoir vérifié et être sûr que votre application peut absorber les actions injectées, s'y adapter et s'en remettre.

Planifiez minutieusement chaque expérience à l'aide d'un document qui capture les détails clés, similaire au document de planification de l'expérience fourni en annexe. Certains des domaines critiques à inclure sont la définition de l'état stationnaire, l'hypothèse et la méthode d'injection des défaillances. La planification, l'exécution et l'analyse d'une expérience de chaos peuvent être couvertes par un seul artefact.

Après avoir finalisé le plan écrit pour l'expérience, préparez le code nécessaire pour injecter les perturbations planifiées décrites dans le document.

Pour saisir l'impact potentiel au cours de l'expérience, assurez-vous que les mécanismes d'observabilité sont en place. Si vous ne disposez pas encore d'un moyen automatique pour capturer les résultats des expériences, tels que des rapports d' AWS FIS expérimentation, identifiez les membres de l'équipe qui prendront des notes pendant l'exécution, capturez des captures d'écran des tableaux de bord et dirigerez l'équipe tout au long de l'expérience.

Apprenez et peaufinez

Après chaque expérience, réunissez-vous en équipe pour passer en revue l'expérience du chaos et y réfléchir. Faites un effort conscient pour maintenir un état d'esprit irréprochable. Votre objectif doit être d'avoir un dialogue ouvert et constructif entièrement axé sur l'apprentissage maximal, et non sur le blâme.

Commencez par revoir la définition de l'état d'équilibre et l'hypothèse de l'expérience. L'application s'est-elle comportée comme prévu ? Y a-t-il eu des surprises qui ont invalidé les hypothèses ? Discutez des observations sur la façon dont l'application a réagi au cours de l'expérience, qu'elle soit

positive ou négative. Les données collectées (métriques, journaux, captures d'écran, etc.) devraient raconter exactement ce qui s'est passé.

Abordez cet examen des données avec curiosité plutôt qu'avec jugement, et identifiez les domaines dans lesquels des améliorations peuvent être apportées à la conception des applications, à la documentation, à la surveillance ou à d'autres capacités sur la base des enseignements tirés. Ces actions sont capturées sous forme de projets de suivi afin de rendre l'application plus résiliente.

Grâce à cette approche irréprochable, vous pouvez avoir des conversations franches sur ce qui n'a pas fonctionné et comment vous pouvez y remédier. Supposez que toutes les personnes impliquées ont une intention positive et ayez la certitude qu'elles ont travaillé pour obtenir de bons résultats. Votre objectif commun est la croissance et la progression de l'organisation grâce à l'apprentissage continu et à l'adaptation. Les évaluations d'expériences Chaos menées de manière constructive et irréprochable offrent à votre équipe un espace sûr où elle peut obtenir des informations précieuses qui rendront vos applications et votre organisation plus fiables et résilientes sur le long terme. L'accent reste mis sur les apprentissages, et non sur les personnes. Pour diffuser les enseignements au sein de vos équipes, publiez le rapport sur les résultats de l'expérience dans un endroit central et publiez les résultats afin que d'autres puissent en tirer des enseignements.

Étendre l'ingénierie du chaos à l'échelle de votre organisation

Au fur et à mesure que votre organisation adopte l'ingénierie du chaos, sa standardisation et sa mise en œuvre poseront des défis. Au cours des premiers stades de maturité, les différentes équipes sont susceptibles d'utiliser différents outils et variantes du processus d'ingénierie du chaos décrit dans les sections précédentes. Dans le même temps, certaines équipes peuvent ne pas donner la priorité à l'ingénierie du chaos ou ne pas l'adopter du tout, malgré ses avantages potentiels. Les sections suivantes fournissent des conseils sur la manière de surmonter ces défis.

Dans l'ensemble, votre approche de l'ingénierie du chaos doit être conçue pour trouver un équilibre entre le leadership centralisé et la participation décentralisée. Cet équilibre permet de garantir que l'ingénierie du chaos est intégrée au processus de développement et que les enseignements sont partagés au sein de votre organisation.

Mise en place d'une pratique d'ingénierie du chaos

La normalisation de la pratique de l'ingénierie du chaos peut accélérer son adoption. Le partage des enseignements tirés des expériences entre les équipes peut amplifier le retour sur investissement dans l'ingénierie du chaos.

Construisez un centre d'excellence centralisé ou réunissez un groupe d'experts en la matière dans le cadre de votre pratique d'ingénierie du chaos. En tant que petite fonction centralisée, cette équipe peut travailler avec les équipes de développement logiciel, d'infrastructure, de sécurité et commerciales et maintenir les normes utilisées par ces équipes. Par souci de simplicité, le centre d'excellence est appelé équipe de pratique centralisée, et les groupes qui appliquent l'ingénierie du chaos sont appelés équipes de pratique dans le reste de ce guide.

Rôle de l'équipe du cabinet centralisé

L'équipe du cabinet centralisé est chargée de développer et de mettre en œuvre des pratiques d'ingénierie du chaos au sein de l'organisation. Ils travaillent en étroite collaboration avec les équipes de praticiens pour les guider dans la conception et la réalisation d'expériences, et pour s'assurer que les expériences sont utiles à l'entreprise. L'équipe de pratique centralisée fournit également des conseils et un soutien aux équipes de développement, d'infrastructure et de sécurité pour les aider à intégrer l'ingénierie du chaos dans leurs processus de développement.

Les principales responsabilités d'une équipe d'ingénierie du chaos centralisée sont les suivantes :

- Habilitation — Une fonction centralisée d'ingénierie du chaos sert de facilitateur pour introduire la pratique de l'ingénierie du chaos par le biais de journées de jeu et d'ateliers. Ils guident les équipes dans le processus d'ingénierie du chaos, notamment en sélectionnant des scénarios de défaillance, en définissant des hypothèses et en produisant des rapports à partager avec l'ensemble de l'organisation. L'équipe de pratique centralisée devrait posséder du matériel de formation et s'efforcer d'améliorer les compétences des équipes de pratique dans leur utilisation de l'ingénierie du chaos.
- Conseil — L'équipe de pratique centralisée peut également jouer un rôle consultatif pour superviser les expériences menées par les équipes de praticiens. Leur expérience et leurs connaissances peuvent garantir que les expériences apportent de la valeur à l'entreprise et sont menées de manière sûre. De même, l'équipe peut superviser l'exécution et le compte rendu d'une expérience afin de guider les personnes novices dans le domaine de l'ingénierie du chaos.
- Marketing et suivi de la valeur — La communication de la valeur commerciale de l'ingénierie du chaos est essentielle au succès d'un tel programme. Chaque équipe participant à des expériences d'ingénierie du chaos doit collecter des données issues des expériences menées dans l'ensemble de l'entreprise et démontrer la valeur de l'investissement de l'organisation dans l'ingénierie du chaos. Cela inclut la quantification et la célébration du nombre d'incidents évités au cours de chaque expérience, du temps d'arrêt qui aurait été encouru en cas d'échec de l'expérience et de l'impact global sur l'entreprise si les scénarios de défaillance s'étaient produits en production. En collectant et en centralisant ces données auprès de toutes les équipes, et en les rendant disponibles dans l'ensemble de l'organisation, l'équipe du cabinet centralisé peut suivre et influencer la valeur dérivée de l'adoption de l'ingénierie du chaos dans l'ensemble de l'organisation.
- Normes — L'équipe de pratique centralisée doit posséder et maintenir le processus de réalisation des expériences de chaos, les modèles pour la planification et les rapports sur les expériences, ainsi que les outils utilisés pour mener les expériences.

  L'équipe centrale doit posséder et gérer les modèles de planification des expériences, les modèles de rapports d'expériences, la documentation des processus et le matériel d'habilitation. La documentation sur les meilleures pratiques et les supports d'habilitation fournissent des conseils aux équipes de praticiens sur des sujets tels que les garde-fous qu'elles peuvent utiliser pour limiter l'impact d'une expérience, le moment où mener une expérience en production et la manière de faire évoluer leur utilisation de l'ingénierie du chaos au fil du temps. Pour des exemples de modèles et de sorties, consultez l'annexe.

L'équipe de pratique centralisée doit également être responsable du processus de réalisation d'une expérience, y compris les communications et l'escalade, ainsi que du moment et de la manière de communiquer avec les autres équipes de l'organisation avant ou pendant une expérience. Le processus doit également indiquer quand des garde-corps sont nécessaires.

L'équipe de pratique centralisée doit également sélectionner et posséder les principaux outils pour mener des expériences sur le chaos (par exemple, des outils tels que AWS FIS). La sélection et la mise en œuvre d'outils supplémentaires, tels que les outils de génération de charge, devraient être laissées à la discrétion des équipes de pratique. Les équipes de praticiens devraient être en mesure d'adapter l'ensemble du processus et des outils pour répondre au mieux à leurs besoins.

Rôle des équipes d'entraînement

L'équipe centralisée est chargée de piloter la stratégie globale d'ingénierie du chaos, tandis que les équipes de praticiens participent au processus et sont responsables du développement et de l'exécution des expériences. Cela permet de garantir que les expériences sont pertinentes pour chaque produit ou service spécifique, et que les enseignements sont exploitables et peuvent être appliqués pour améliorer la fiabilité et la résilience du produit. L'équipe du cabinet centralisé agit en tant que mentor et responsable des normes et des processus d'ingénierie du chaos de l'organisation. Cependant, afin d'éviter que l'équipe centralisée ne devienne un goulot d'étranglement, les équipes de pratique individuelles devront s'inspirer de la pratique centrale pour réaliser elles-mêmes des expériences de chaos.

Création d'une communauté de pratique

En plus de créer une équipe centralisée, nous vous recommandons de créer une communauté informelle de praticiens intéressés par l'ingénierie du chaos. Cette communauté fournit une plate-forme pour partager les connaissances, les meilleures pratiques et les expériences entre les équipes de pratique et l'ensemble de l'organisation.

La communauté de pratique peut être gérée par l'équipe centralisée du cabinet d'ingénierie du chaos, mais n'importe quel membre de l'organisation peut devenir membre de la communauté. L'équipe centralisée peut tirer parti de la communauté de pratique pour diffuser des mises à jour et générer des enseignements, et pour recueillir les commentaires des équipes de pratique qui utilisent les normes et les processus gérés par l'équipe centralisée. La communauté agira comme une boucle de rétroaction pour informer l'équipe centralisée de l'efficacité des pratiques d'ingénierie du chaos au

sein des équipes de pratique. L'équipe de pratique centralisée peut ensuite ajuster sa documentation et ses artefacts de support afin de soutenir au mieux les équipes produit.

Intégrer l'ingénierie du chaos à votre résilience opérationnelle

Une expérience de chaos est un investissement de la part de votre entreprise pour prévenir les incidents de production. Il sera nécessaire de déterminer où l'entreprise peut tirer le meilleur parti de cet investissement. L'organisation peut travailler avec l'équipe centralisée du cabinet d'ingénierie du chaos pour mettre à jour ses normes et déterminer quels produits sont suffisamment critiques pour nécessiter une expérimentation du chaos.

Processus de développement des systèmes

L'ingénierie du chaos et les expériences de chaos doivent être effectuées à plusieurs reprises dans le cadre du cycle de vie d'une application. Tout comme les équipes effectuent régulièrement des tests de reprise après sinistre, elles devraient mener des expériences de chaos et des journées de jeu de manière continue et périodique tout au long de l'année. Cette approche améliore la façon dont une organisation anticipe, observe et répond aux incidents.

Conclusion

La discipline de l'ingénierie du chaos a beaucoup évolué au cours de la dernière décennie. Il a été adopté dans de nombreux secteurs et a aidé les entreprises à créer des services résilients et à accroître la satisfaction des clients. (Pour des exemples de la manière dont les organisations ont mis en œuvre ces pratiques, voir Chaos Engineering Stories.) L'ingénierie du chaos permet aux entreprises de réduire les risques liés à leurs applications critiques en injectant des défaillances contrôlées à tous les niveaux de leur pile d'applications, y compris les services des fournisseurs de cloud. La capacité d'influencer l'ensemble d'une pile d'applications de manière contrôlée permet une résilience continue, une amélioration de l'excellence opérationnelle, de l'observabilité et une architecture axée sur la restauration sans le stress des interruptions de production. Cette approche conduit également à de meilleures pratiques de test de résilience au sein de l'organisation. Pour commencer à adopter l'ingénierie du chaos, organisez une journée ou un atelier de jeu sur le chaos afin de démontrer la valeur que l'ingénierie du chaos peut apporter à votre organisation.

Ressources

AWS FIS implémentations de modèles de défaillance :

- AWS Fault Injection Service À utiliser pour démontrer la résilience des applications multirégions et multi-AZ (AWS article de blog)
- AWS Le simulateur d'injection de défauts prend en charge les expériences d'ingénierie du chaos sur les pods Amazon EKS (article de AWS blog)
- Annonce des nouvelles fonctionnalités du simulateur d'injection de AWS défauts pour les charges de travail Amazon ECS (article de AWS blog)
- Améliorez la résilience des applications grâce aux mesures de volume Amazon EBS et au simulateur d'injection de AWS défauts (AWS article de blog)
- Ingénierie du chaos utilisant le simulateur d'injection de AWS défauts dans un AWS environnement multi-comptes (article de AWS blog)
- Expériences de chaos sur Amazon RDS à l'aide du simulateur d'injection de AWS défauts (article de AWS blog)
- Création d'applications sans serveur résilientes grâce à l'ingénierie du chaos (article de AWS blog)
- Utiliser le FIS pour interrompre une instance ponctuelle (AWS atelier)
- Automatiser l'ingénierie du chaos dans vos pipelines de livraison (article AWS communautaire)

Autres ressources :

- Investir dans l'ingénierie du chaos en tant que nécessité stratégique
- Histoires d'ingénierie du chaos
- CloudWatchDocumentation Amazon
- Documentation Amazon CloudWatch RUM
- Documentation AWS X-Ray
- Documentation AWS FIS

Annexe : Exemples de documents

Les exemples de documents fournis dans cette section utilisent un site d'adoption d'animaux de compagnie fictif (PetSite) comme application cible pour une expérience chaotique.

- Document de planification de l'expérience
- Document sur les résultats de l'expérience

Document de planification de l'expérience

État d'équilibre

Nom du processus	Site d'adoption d'animaux
Architecture physique	(Lien vers le schéma d'architecture.)
Architecture logique	(Lien vers le diagramme logique.)
Définir l'état d'équilibre	Le temps de chargement moyen des pages, mesuré à l'aide de Largest Contentful Paint (LCP), pour le site d'adoption d'animaux de compagnie est de 2,5 secondes ou moins avec une latence de 99 percentiles (P99) de 4 secondes ou moins avec une base de référence de 5 000 utilisateurs simultanés.
Indicateurs de l'état d'équilibre	Métrique LCP capturée par l'ensemble de la base d'utilisateurs et métriques de référence (latence, débit, taux d'erreur, saturation).
Observabilité en régime permanent	Le LCP sera capturé par le navigateur de l'utilisateur, envoyé à Amazon CloudWatch et inspecté avec CloudWatch RUM. Sur une période de 60 secondes, le temps moyen et le temps LCP P99 seront agrégés pour toutes les demandes au cours de cette période. Top-level

Nom du processus	Site d'adoption d'animaux
	les indicateurs dorés sont capturés en utilisant CloudWatch.
Procédé pour atteindre l'état d'équilibre	Grafana K6 sera utilisé pour créer une charge qui simule les niveaux de trafic de production normaux d'environ 5 000 utilisateurs simultané s.

Exigences en matière d'observabilité

Les équipes devraient être en mesure de consulter les informations suivantes :

- État stable : qu'est-ce qui sera observé pour vérifier que l'application se déroule dans des conditions normales ?
- Condition de défaillance : Comment la condition de défaillance apparaîtra-t-elle sur le tableau de bord ? Par exemple :
  - Alarmes qui devraient être déclenchées
  - Journaux qui devraient être générés
- Impact de la défaillance : que faut-il observer pour visualiser les composants susceptibles d'être affectés (étendue de l'impact) ?
- Rétablissement : Comment le rétablissement sera-t-il visualisé et mesuré pour capturer le MTTR ?
- Débogage : détails de résolution des problèmes liés aux échecs des expériences.

Le tableau suivant fournit des suggestions et des exemples pour un tableau des exigences d'observabilité. Vous devez définir ce qui doit être observé en fonction de votre expérience spécifique.

Ce qui doit être observé	Lien vers l'outil d'observabilité	Ce qui est observé
Source d'entrée	Tableau de bord Grafana K6	• Nombre de conteneurs en circulation

Ce qui doit être observé	Lien vers l'outil d'observabilité	Ce qui est observé
		• Demandes par seconde
État général de l'application	• Tableau de CloudWatch bord d'adoption d'animaux • Tableau de bord de l'expérience utilisateur sur l'adoption d'animaux de compagnie (RUM	• Nombre de nœuds sains Amazon EKS • Utilisation du processeur du nœud Amazon EKS
État du flux de travail	Tableau de CloudWatch bord d'adoption d'animaux	Temps LCP, indicateurs de référence
Suivis	Tableau de X-Ray bord d'adoption d'animaux	• Latence des demandes • Request Count • Nombre de défaillances
Journaux	CloudWatch Journaux d'adoption d'animaux	Toute erreur rencontrée par les pods sera transmise à CloudWatch Logs.

Définition de l'expérience

Nom de l'expérience	Stress du PetSite processeur Amazon EKS
Code source de l'expérience	(Lien vers le référentiel des sources d'expériences.)
Description de l'expérience	Cette expérience explore l'impact d'une augmentation de l'utilisation du processeur dans le module d' PetSite application sur l'expérience client globale. En injectant le stress du processeur dans chaque PetSite module en

Nom de l'expérience	Stress du PetSite processeur Amazon EKS
	cours d'exécution, nous serons en mesure de comprendre s'il y a un impact sur les clients et l'ampleur de cet impact.
Exigences ou paramètres de l'expérience	Charge de l'application : moyenne de productio n Sélecteur d'étiquettes pour capsules : `app=petsite`
Durée de l'expérience	10 minutes
Environnement	Environnement de test alpha
Expérimentez les ressources cibles	PetSite modules d'application
Base de référence d'expérimentation introduite via l'outil de génération de charge	• 54 % des demandes ont un LCP inférieur à 2,5 secondes. • 46 % des demandes ont un LCP inférieur à 4 secondes. • Aucune erreur n'est constatée.
Condition de marche arrière	Aucune

Hypothèse

Et si	Impact	Récupération
Qu'arriverait-il à l'état d'équilib re si les modules PetSite d'application utilisaient ou provoquaient une utilisation du processeur supérieure à 60 % pendant 10 minutes dans des	Les temps LCP resteront inférieurs à 2,5 secondes pour les utilisateurs P50 avec un P99 de 4 secondes ou moins. Le consommateur doit être en	Détection : • Le stress du processeu r sera détecté par les alarmes configurées dans CloudWatch.

Et si	Impact	Récupération
conditions de trafic normales en production ?	mesure de charger la page de PetSite destination.	• Les métriques LCP généreront également des alarmes en cas de dégradation de l'expérience utilisateur. Self-healing: • La nature distribuée de l'architecture des microservices signifie que de nombreuses instances de pods s'exécutent dans plusieurs zones de disponibilité. • Le plan de contrôle du cluster EKS éloignera le trafic des pods concernés et lancera de nouveaux pods sur les nœuds de travail. Rétablissement : Lorsque l'utilisation du processeur revient à la normale, le LCP devrait se rétablir automatiquement.

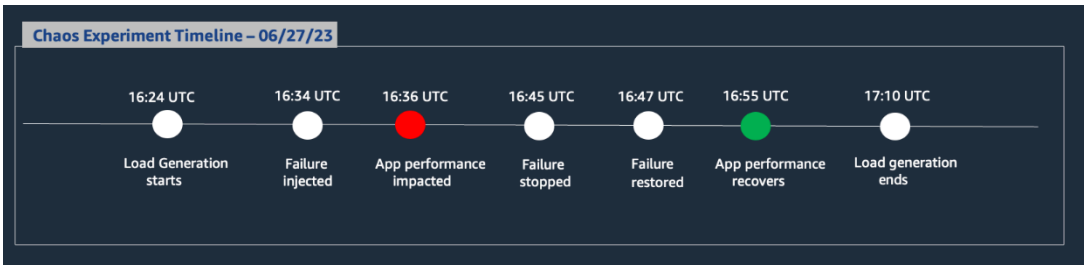
Processus d'expérimentation

Adaptez l'exemple de processus étape par étape suivant à votre expérience spécifique :

1. Validez l'accès et les fonctionnalités de tous les Amazon CloudWatch, CloudWatch RUM et AWS X-Ray tableaux de bord.
2. Validez l'état de l'environnement de l'application :
   a. Vérifiez que le cluster EKS est sain à l'aide du CloudWatch tableau de bord.
   b. Consultez le déploiement de l'application de test pour un site d'adoption d'animaux de compagnie à l'adresse URL d'exemple.
3. Initiez une charge pour atteindre un état stable :
   a. Vérifiez que le générateur de charge est en cours d'exécution et envoie 5 000 demandes par seconde.
   b. Attendez 5 minutes pour que l'application atteigne son état d'équilibre.
   c. Vérifiez l'état d'équilibre de l'application à l'aide du tableau de bord CloudWatch RUM.
4. Initier une panne (expérience) :
   a. Ouvrez la AWS FIS console.
   b. Lancez l'expérience Pet-Adoption-Pod-Stress.
   c. Vérifiez que le test est en cours d'exécution dans la console.
5. Observez l'impact du défaut sur votre application :
   a. Capturez des captures d'écran du CloudWatch RUM et CloudWatch des tableaux de bord, et notez les points de données anormaux.
   b. Une fois l'expérience terminée AWS FIS, prenez des captures d'écran supplémentaires pour enregistrer si l'application revient à l'état d'équilibre en l'absence de stress, et notez toute anomalie dans les points de données.
   c. Si l'état d'équilibre ne revient pas, prenez des mesures pour récupérer l'application et enregistrez les étapes effectuées.
6. Vérifiez que l'environnement est revenu à la normale :
   - Passez en revue tous les indicateurs relatifs à l'activité, à l'expérience utilisateur, aux applications et à l'infrastructure pour vérifier que le système est revenu à un état connu. Prenez des captures d'écran du tableau de bord si cela est utile

Chronologie de l'expérience

Assurez-vous de saisir la chronologie de l'expérience de bout en bout, en commençant par la génération de charge, l'injection du défaut, l'observation de l'impact et la restauration de l'application, jusqu'à l'arrêt de la génération de charge. Ceci est illustré dans l'exemple suivant.



Résultats de l'expérience

ID d'exécution de l'expérience	Résultats de l'expérience
`PET-ADOPT-EXP-23`	(Lien vers les résultats de l'expérience.)

Défauts identifiés

- Le cluster Kubernetes n'a pas détecté l'altération du processeur des PetSite pods. Il n'a donc pas planifié de déploiements supplémentaires.
- Aucune augmentation des taux d'erreur 4XX ou 5XX n'a été observée à la suite de cette expérience.
- Nous devons ajuster le bilan de santé du pod pour tenir compte de l'impact sur le LCP en cas de contraintes de ressources.

Document sur les résultats de l'expérience

Configuration

Documentez les configurations spécifiques à l'expérience. Par exemple :

- La génération de charge est configurée pour simuler 5 000 utilisateurs émettant un total de 85 demandes par seconde.

Conditions préalables

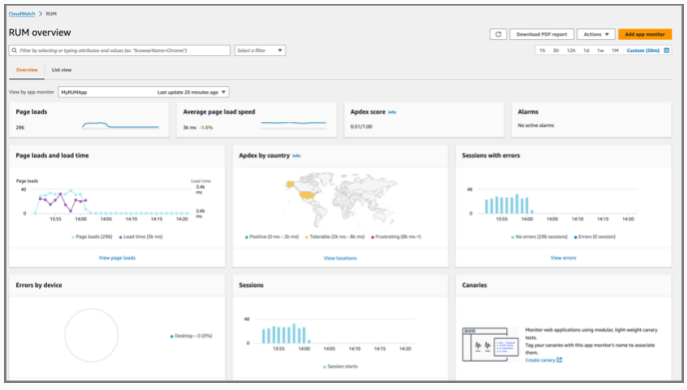
- Nous avons vérifié que le site d'adoption d'animaux de compagnie fonctionnait dans l'environnement de test alpha.
- Vérifiez que le modèle d'expérience a été configuré pour appliquer un stress du processeur aux modules PetSite d'application qui s'exécutent dans le cluster EKS. Les modules d'application ont été identifiés par l'étiquette Kubernetes. `app=petsite`
- Il a été confirmé que Load était en cours d'exécution et générait 85 demandes par seconde.

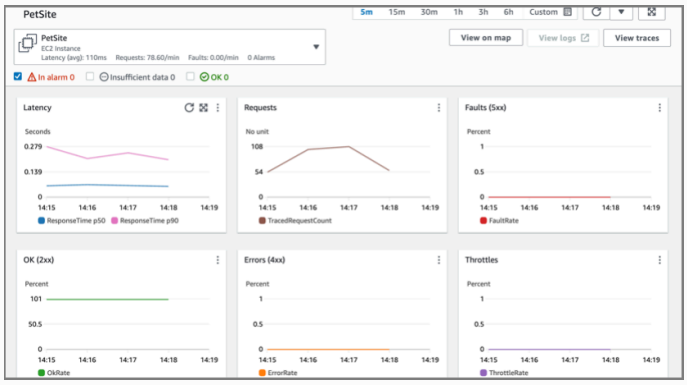
État d'équilibre

Documentez les mesures prises pour atteindre l'état stable et la façon dont vous l'avez vérifié. Par exemple :

Pour le déploiement test d'un site d'adoption d'animaux de compagnie, une charge de 85 RPS est générée pour simuler un état d'équilibre. Le CloudWatch RUM et CloudWatch les tableaux de bord ont été examinés pour vérifier que toutes les métriques commerciales et applicatives se situaient dans les plages normales avant l'exécution de l'expérience.

Données d'observabilité :

Expected	Observé
• Le LCP est inférieur à 4 secondes pour le P99 de requêtes. • La latence de réponse est inférieure à 500 ms. • Il n'y a aucune erreur 4XX ou 5XX.	

Expected	Observé
	

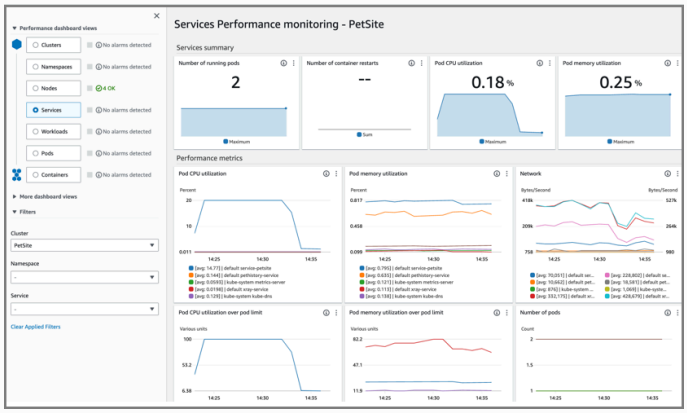
Injection de pannes

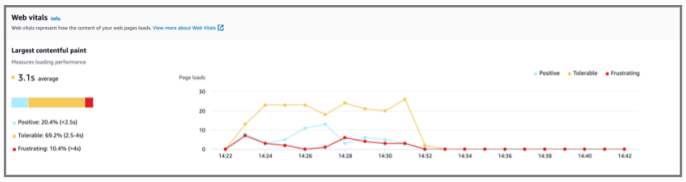
AWS FIS a été utilisé pour injecter des défauts en utilisant le modèle d'expérience (fournir le lien). L'expérience était programmée pour une durée de 10 minutes, et une annulation était configurée si les nœuds de travail subissaient un stress du processeur supérieur à 60 %.

Observation des défauts

Le CloudWatch RUM et CloudWatch les tableaux de bord ont été revus pour suivre l'état d'équilibre de l'application (défini à l'aide de métriques LCP). Les captures d'écran ont été capturées dans le tableau suivant.

Données d'observabilité :

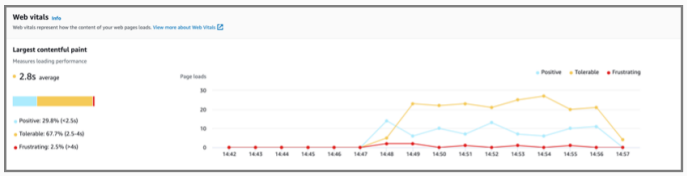
Expected	Observé
• Le LCP doit rester inférieur à 4 secondes pour le P99. • Le temps de réponse doit rester inférieur à 500 ms. • Aucune erreur 4XX ou 5XX ne doit être rencontrée.	

Expected	Observé
	

Récupération

Une fois le stress éliminé (l' AWS FIS expérience est terminée et le stress du processeur a été éliminé des modules), l'application devrait retrouver son état d'équilibre normal. Aucune intervention manuelle ne devrait être requise.

Données d'observabilité :

Expected	Observé (capture d'écran)
Le LCP P99 doit être inférieur à 4 secondes avec une moyenne inférieure à 2,5 secondes.	

Historique du document

Le tableau suivant décrit les modifications importantes apportées à ce guide. Pour être averti des mises à jour à venir, abonnez-vous à un fil RSS.

Modification	Description	Date
Publication initiale	—	4 avril 2025

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.