Guide de l'utilisateur des plans de dimensionnement

AWS Auto Scaling

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Auto Scaling: Guide de l'utilisateur des plans de dimensionnement

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Les marques et la présentation commerciale d'Amazon ne peuvent être utilisées en relation avec un produit ou un service qui n'est pas d'Amazon, d'une manière susceptible de créer une confusion parmi les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d’Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

Table of Contents

Qu'est-ce qu'un plan de dimensionnement ? ... 1
Ressources prises en charge ... 1
Caractéristiques et avantages du plan de mise à l'échelle ... 1
Comment démarrer ... 2
Travailler avec les plans de dimensionnement ... 3
Disponibilité par région ... 3
Tarification ... 4
Fonctionnement des plans de dimensionnement ... 5
Bonnes pratiques ... 8
Autres considérations ... 9
Éviter l' ActiveWithProblems erreur ... 10
Premiers pas ... 12
Étape 1 : Rechercher vos ressources évolutives ... 13
Prérequis ... 13
Ajout de votre groupe Auto Scaling à votre nouveau plan de dimensionnement ... 13
En savoir plus sur la découverte de vos ressources évolutives ... 15
Étape 2 : Spécifier la stratégie de dimensionnement ... 17
Étape 3 : Configurer les paramètres avancés (facultatif) ... 20
Paramètres généraux ... 20
Paramètres de mise à l'échelle dynamique ... 23
Paramètres de mise à l'échelle prédictive ... 23
Étape 4 : Créer votre plan de dimensionnement ... 25
(Facultatif) Affichage des informations de mise à l'échelle pour une ressource ... 25
Étape 5 : nettoyer ... 28
Suppression de votre groupe Auto Scaling ... 29
Étape 6 : étapes suivantes ... 29
Migrez votre plan de mise à l'échelle ... 31
Étape 1 : passez en revue votre configuration existante ... 31
Différences entre les plans de dimensionnement et les politiques de dimensionnement ... 32
Étape 2 : créer des politiques de dimensionnement prédictives ... 32
Étape 3 : Passez en revue les prévisions générées par les politiques de dimensionnement prédictif ... 38
Étape 4 : préparer la suppression du plan de dimensionnement ... 39
Étape 5 : Supprimer le plan de dimensionnement ... 40

Étape 6 : réactiver la mise à l'échelle dynamique 42
Créez des politiques de dimensionnement pour le suivi des cibles pour les groupes Auto Scaling 42
Créez des politiques de dimensionnement pour le suivi des cibles pour d'autres ressources évolutives 44
Étape 7 : réactiver le dimensionnement prédictif 46
Référence Amazon EC2 Auto Scaling pour la migration des politiques de dimensionnement du suivi des cibles 47
Référence Application Auto Scaling pour la migration des politiques de dimensionnement du suivi des cibles 49
Informations supplémentaires 51
Sécurité 52
AWS PrivateLink 52
Créer un point de terminaison de VPC d'interface pour les plans de dimensionnement 53
Création d'une stratégie de point de terminaison de VPC pour le dimensionnement des plans 53
Migration des points de terminaison 54
Protection des données 55
Gestion des identités et des accès 56
Contrôle d’accès 57
Fonctionnement des plans de dimensionnement avec IAM 57
Rôles liés à un service 61
Exemples de politiques basées sur l’identité 63
Validation de conformité 70
Sécurité de l’infrastructure 71
Quotas 73
Historique de la documentation 75
..... lxxviii

Qu'est-ce qu'un plan de dimensionnement ?

Utilisez unplan de dimensionnementpour configurer la mise à l'échelle automatique pour les ressources évolutives associées ou associées en quelques minutes. Par exemple, vous pouvez utiliser des balises pour regrouper des ressources dans des catégories telles que la production, les tests ou le développement. Vous pouvez ensuite rechercher et configurer des plans de dimensionnement pour des ressources évolutives appartenant à chaque catégorie. Ou, si votre infrastructure cloud l'inclut AWS CloudFormation, vous pouvez définir des modèles de stack à utiliser pour créer des collections de ressources. Créez ensuite un plan de dimensionnement évolutif qui appartiennent à chaque pile.

Ressources prises en charge

AWS Auto Scaling prend en charge l'utilisation de plans de mise à l'échelle pour les services et ressources suivants :

- Amazon Aurora— Augmentez ou diminuez le nombre de réplicas en lecture Aurora qui sont provisionné pour un cluster de bases de données Aurora.
- Amazon EC2 Auto Scaling — Lancez ou mettez fin à EC2 des instances en augmentant ou en diminuant la capacité souhaitée d'un groupe Auto Scaling.
- Amazon Elastic Container Service— Augmentez ou diminuez le nombre de tâches souhaitées dans Amazon ECS.
- Amazon DynamoDB— Augmentez ou diminuez la capacité en lecture et écriture provisionnée d'une table DynamoDB ou d'un index secondaire global.
- Parc d'instances : lancez ou mettez fin à EC2 des instances en augmentant ou en diminuant la capacité cible d'un parc d'instances.

Caractéristiques et avantages du plan de mise à l'échelle

Les plans de dimensionnement offrent les fonctions et avantages suivants :

- Découverte des ressources : AWS Auto Scaling permet une découverte automatique des ressources pour vous aider à trouver des ressources évolutives dans votre application.
- Dimensionnement dynamique : les plans de dimensionnement utilisent les services Amazon EC2 Auto Scaling et Application Auto Scaling pour ajuster la capacité des ressources évolutives afin

de gérer les modifications du trafic ou de la charge de travail. Les mesures de dimensionnement dynamique peuvent être des mesures d'utilisation standard ou de débit, ou des mesures personnalisées.

- Recommandations de dimensionnement intégrées– AWS Auto Scaling fournit des stratégies de dimensionnement avec des recommandations que vous pouvez utiliser pour optimiser les performances, les coûts ou un équilibre entre les deux.
- Dimensionnement prédictif— Les plans de dimensionnement prennent également en charge la mise à l'échelle prédictive pour les groupes Auto Scaling. Cela permet d'augmenter plus rapidement la EC2 capacité de votre Amazon en cas de pics réguliers.

> **Important**
>
> Si vous utilisez des plans de dimensionnement uniquement pour le dimensionnement prédictif, nous vous recommandons vivement de définir des politiques de dimensionnement prédictif directement sur vos ressources Auto Scaling. Cette option offre davantage de fonctionnalités, telles que l'utilisation d'agrégations de métriques pour créer de nouvelles métriques personnalisées ou conserver les données métriques historiques entre les blue/green déploiements. Pour plus d'informations sur Amazon EC2 Auto Scaling, consultez la section Predictive scaling for Amazon EC2 Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling. Pour plus d'informations sur Application Auto Scaling, consultez la section Predictive Scaling for Application Auto Scaling dans le Guide de l'utilisateur d'Application Auto Scaling.
> Pour obtenir un guide sur la migration des plans de dimensionnement vers les politiques de dimensionnement prédictif d'Amazon EC2 Auto Scaling, consultezMigrez votre plan de mise à l'échelle.

Comment démarrer

Aidez-vous des ressources suivantes pour créer et utiliser les plans de dimensionnement :

- Fonctionnement des plans de dimensionnement
- Bonnes pratiques pour les plans de dimensionnement
- Démarrez avec les plans de dimensionnement

Travailler avec les plans de dimensionnement

Vous pouvez créer vos plans de dimensionnement, y accéder et les gérer à l'aide des interfaces suivantes :

- AWS Management Console - offre une interface Web que vous pouvez utiliser pour accéder à votre . Si vous vous êtes inscrit à un Compte AWS, vous pouvez accéder à vos plans de dimensionnement en vous connectant au AWS Management Console, en utilisant le champ de recherche de la barre de navigation pour rechercher AWS Auto Scaling, puis en choisissant AWS Auto Scaling.
- AWS Command Line Interface (AWS CLI) — Fournit des commandes pour un large éventail de Services AWS, et est compatible avec Windows, macOS et Linux. Consultez le Guide de l'utilisateur AWS Command Line Interface pour démarrer. Pour plus d'informations, consultez autoscaling-plans dans AWS CLI Références des commandes.
- AWS Tools for Windows PowerShell— Fournit des commandes pour un large éventail de AWS produits pour ceux qui écrivent des scripts dans l' PowerShell environnement. Consultez le Guide de l'utilisateur Outils AWS pour PowerShell pour démarrer. Pour plus d'informations, consultez le Guide de référence des cmdlets Outils AWS pour PowerShell.
- AWS SDKs— Fournit des opérations d'API spécifiques au langage et prend en charge de nombreux détails de connexion, tels que le calcul des signatures, la gestion des nouvelles tentatives de demande et la gestion des erreurs. Pour de plus amples informations, veuillez consulter AWS SDKs.
- API HTTPS : Fournit des actions d'API de bas niveau appelées à l'aide de demandes HTTPS. Pour plus d'informations, consultez la page Référence de l'API AWS Auto Scaling.
- AWS CloudFormation— Permet de créer des plans de mise à l'échelle à l'aide CloudFormation de modèles. Pour plus d'informations, consultez la AWS::AutoScalingPlans::ScalingPlanréférence dans le guide de AWS CloudFormation l'utilisateur.

Disponibilité par région

L' AWS Auto Scaling API est disponible en plusieurs versions Régions AWS et fournit un point de terminaison pour chacune de ces régions. Pour obtenir la liste de toutes les régions et points de terminaison dans lesquels l'API est actuellement disponible, consultez la section AWS Auto Scaling Points de Références générales AWSterminaison et quotas dans Services .

Tarification

Toutes les fonctionnalités du plan de dimensionnement sont activées pour votre utilisation. Les fonctionnalités sont fournies sans frais supplémentaires au-delà des frais de service CloudWatch et des autres AWS Cloud ressources que vous utilisez.

> **Note**
>
> La fonction de mise à l'échelle prédictive repose sur l' CloudWatch GetMetricDataopération de collecte de données métriques historiques pour la prévision des capacités, ce qui entraîne des coûts. Toutefois, si vous activez le dimensionnement prédictif avec une politique de dimensionnement Amazon EC2 Auto Scaling au lieu d'un plan de dimensionnement, les appels vers`GetMetricData`.

Fonctionnement des plans de dimensionnement

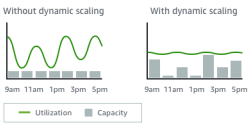
AWS Auto Scaling vous permet d'utiliser des plans de dimensionnement pour configurer un ensemble d'instructions pour dimensionner vos ressources. Si vous utilisez des ressources évolutives AWS CloudFormation ou si vous y ajoutez des balises, vous pouvez configurer des plans de dimensionnement pour différents ensembles de ressources, par application. La AWS Auto Scaling console fournit des recommandations pour les stratégies de dimensionnement adaptées à chaque ressource. Une fois que vous avez créé votre plan de dimensionnement, combine les méthodes de dimensionnement dynamique et de dimensionnement prédictif pour prendre en charge votre stratégie de dimensionnement.

Qu'est-ce qu'une stratégie de dimensionnement ?

La stratégie de mise à l'échelle indique AWS Auto Scaling comment optimiser l'utilisation des ressources dans le cadre de votre plan de mise à l'échelle. Vous pouvez optimiser la disponibilité, la rentabilité, ou un équilibre des deux. Sinon, vous pouvez également créer votre propre stratégie personnalisée, selon les métriques et les seuils que vous définissez. Vous pouvez définir des stratégies distinctes pour chaque ressource ou type de ressource.

Qu'est-ce que le dimensionnement dynamique ?

La mise à l'échelle dynamique crée des stratégies de dimensionnement de suivi de la cible pour les ressources dans votre plan de dimensionnement. Ces stratégies de dimensionnement ajustent la capacité des ressources en réponse à des modifications en direct de l'utilisation des ressources. L'objectif est de fournir suffisamment de capacité pour maintenir l'utilisation des ressources à la valeur cible spécifiée par la stratégie de dimensionnement. Cette option est similaire à la façon dont votre thermostat maintient la température de votre domicile. Choisissez la température et le thermostat se charge du reste.



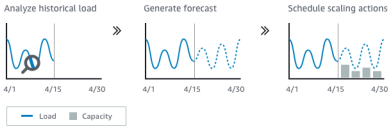
Par exemple, vous pouvez configurer votre plan de dimensionnement afin de maintenir le nombre de tâches que votre service Amazon Elastic Container Service (Amazon ECS) exécute à 75 % de l'UC. Lorsque l'utilisation de l'UC de votre service dépasse les 75 % (ce qui signifie que plus de 75 % de

l'UC qui est réservée au service est utilisé), cela déclenche votre stratégie de dimensionnement pour ajouter une autre tâche à votre service afin de faire face à l'accroissement de la charge.

Qu'est-ce que le dimensionnement prédictif ?

Le dimensionnement prédictif utilise l'apprentissage automatique (machine learning) pour analyser la charge historique de travail et prévoit régulièrement la charge future. Ceci est similaire au fonctionnement des prévisions météorologiques. En utilisant la prévision, le dimensionnement prédictif génère des actions de dimensionnement planifiées pour garantir que la capacité de ressources soit disponible avant que votre application en ait besoin. Comme le dimensionnement dynamique, le dimensionnement prédictif vise à maintenir l'utilisation à la valeur cible spécifiée par la stratégie de dimensionnement.

> **Important**
>
> Si vous utilisez des plans de dimensionnement uniquement pour le dimensionnement prédictif, nous vous recommandons vivement de définir des politiques de dimensionnement prédictif directement sur vos ressources Auto Scaling. Cette option offre davantage de fonctionnalités, telles que l'utilisation d'agrégations de métriques pour créer de nouvelles métriques personnalisées ou conserver les données métriques historiques sur les déploiements bleu/vert. Pour plus d'informations sur Amazon EC2 Auto Scaling, consultez la section Predictive scaling for Amazon EC2 Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling. Pour plus d'informations sur Application Auto Scaling, consultez la section Predictive Scaling for Application Auto Scaling dans le Guide de l'utilisateur d'Application Auto Scaling.
> Pour obtenir un guide sur la migration des plans de dimensionnement vers les politiques de dimensionnement prédictif d'Amazon EC2 Auto Scaling, consultezMigrez votre plan de mise à l'échelle.



Par exemple, vous pouvez activer le dimensionnement prédictif et configurer votre stratégie de dimensionnement afin de conserver l'utilisation moyenne des ressources d'UC de votre groupe Auto Scaling à 50 %. Votre prévision implique des pics de trafic se produisant tous les jours à 8 heures du matin. Votre plan de dimensionnement crée les futures actions de mise à l'échelle planifiées

pour préparer votre groupe Auto Scaling à gérer ce trafic à l'avance. Cela permet de conserver une performance constante de l'application, dans le but de toujours disposer de la capacité requise pour maintenir l'utilisation des ressources aussi proche de 50 % que possible à tout moment.

Voici les concepts clés pour comprendre la mise à l'échelle prédictive :

- Prévision de charge : AWS Auto Scaling analyse jusqu'à 14 jours d'historique pour une métrique de charge spécifiée et prévoit la demande future pour les deux prochains jours. Ces données sont disponibles en intervalles d'une heure et sont mises à jour quotidiennement.
- Actions de dimensionnement planifiées : AWS Auto Scaling planifie les actions de dimensionnement qui augmentent et diminuent la capacité de manière proactive afin de correspondre aux prévisions de charge. À l'heure planifiée, AWS Auto Scaling met à jour la capacité minimale avec la valeur spécifiée par l'action de dimensionnement planifiée. L'objectif est de maintenir l'utilisation des ressources à la valeur cible spécifiée par la stratégie de dimensionnement. Le dimensionnement dynamique est disponible pour ajouter des capacités supplémentaires si votre application nécessite plus de capacité que ce qui est prévu.
- Comportement de capacité maximale: Les limites de capacité minimale et maximale pour la mise à l'échelle automatique s'appliquent à chaque ressource. Cependant, vous pouvez contrôler la capacité de votre application à ajouter des ressources au-delà de sa capacité maximale lorsque la prévision de la capacité est supérieure à la capacité maximale.

Bonnes pratiques pour les plans de dimensionnement

Les bonnes pratiques suivantes peuvent vous aider à tirer le meilleur des plans de dimensionnement

- Lorsque vous créez un modèle de lancement ou une configuration de lancement, activez la surveillance détaillée pour obtenir des données CloudWatch métriques pour les EC2 instances à une fréquence d'une minute, afin de garantir une réponse plus rapide aux modifications de charge. Un dimensionnement sur des métriques à une fréquence de 5 minutes peut entraîner un temps de réponse plus lent et un dimensionnement sur des données de métrique obsolètes. Par défaut, les EC2 instances sont activées pour la surveillance de base, ce qui signifie que les données métriques des instances sont disponibles toutes les cinq minutes. En payant un petit supplément, vous pouvez activer la surveillance détaillée de votre instance pour obtenir les données de métrique toutes les minutes. Pour plus d'informations, consultez Configurer la surveillance pour les instances Auto Scaling dans le guide de l'utilisateur Amazon EC2 Auto Scaling.
- Nous vous recommandons également d'activer les métriques de groupe Auto Scaling. Sinon, les données de capacité réelles ne seront pas affichées dans les graphiques de prévision de capacité disponibles lorsque l'assistant de création de plan de dimensionnement a terminé. Pour plus d'informations, consultez la section Surveillance CloudWatch des métriques pour vos groupes et instances Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling.
- Vérifiez le type d'instance utilisé par votre groupe Auto Scaling et méfiez-vous de l'utilisation d'un type d'instance de performances burstable. EC2 Les instances Amazon offrant des performances exceptionnelles, telles que les instances T3 et T2, sont conçues pour fournir un niveau de référence en matière de performances du processeur, capable de passer à un niveau supérieur lorsque votre charge de travail l'exige. En fonction de l'utilisation cible spécifiée par le plan de dimensionnement, vous pouvez courir le risque de dépasser le niveau de base, puis de manquer de crédits UC, ce qui limite les performances. Pour plus d'informations, consultez Crédits UC et performances de référence pour les instances à capacité extensible. Pour configurer ces instances en tant que `unlimited` telles, consultez la section Using an Auto Scaling group to launch a burstable performance instance Unlimited dans le guide de l' EC2 utilisateur Amazon.

Autres considérations

> **Important**
>
> Si vous utilisez des plans de dimensionnement uniquement pour le dimensionnement prédictif, nous vous recommandons vivement de définir des politiques de dimensionnement prédictif directement sur vos ressources Auto Scaling. Cette option offre davantage de fonctionnalités, telles que l'utilisation d'agrégations de métriques pour créer de nouvelles métriques personnalisées ou conserver les données métriques historiques sur les déploiements bleu/vert. Pour plus d'informations sur Amazon EC2 Auto Scaling, consultez la section Predictive scaling for Amazon EC2 Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling. Pour plus d'informations sur Application Auto Scaling, consultez la section Predictive Scaling for Application Auto Scaling dans le Guide de l'utilisateur d'Application Auto Scaling.
>
> Pour obtenir un guide sur la migration des plans de dimensionnement vers les politiques de dimensionnement prédictif d'Amazon EC2 Auto Scaling, consultezMigrez votre plan de mise à l'échelle.

Gardez les points suivants supplémentaires à l'esprit :

- La mise à l'échelle prédictive (ou dimensionnement prédictif) utilise des prévisions de charge de travail afin de planifier la capacité future. La qualité des prévisions varie en fonction du caractère cyclique de la charge de travail et de l'applicabilité du modèle de prévisions formé. Le dimensionnement prédictif peut être exécuté en mode prévisionnel uniquement afin d'évaluer la qualité des prévisions et les actions de mise à l'échelle créées par les prévisions. Vous pouvez définir le mode de mise à l'échelle prédictive sur Forecast only (Prévision uniquement) lorsque vous créez le plan de dimensionnement, puis le modifier ensuite en Forecast and scale (Prévision et mise à l'échelle) lorsque vous avez fini d'évaluer la qualité des prévisions. Pour plus d'informations, consultez Paramètres de mise à l'échelle prédictive et Surveillance et évaluation des prévisions.
- Si vous choisissez de spécifier différentes métriques de dimensionnement prédictif, vous devez vous assurer que la métrique de dimensionnement et celle de charge sont étroitement liées. La valeur de métrique doit augmenter et diminuer en proportion du nombre d'instances présentes dans le groupe Auto Scaling. Cela garantit que les données de la métrique peuvent être utilisées afin d'augmenter ou réduire proportionnellement le nombre d'instances. Par exemple, la métrique de charge est le nombre total de demandes et la métrique de dimensionnement est l'utilisation

moyenne de l'UC. Si le nombre total de demandes augmente de 50 %, l'utilisation moyenne de l'UC doit également augmenter de 50 %, à condition que la capacité reste inchangée.

- Avant de créer votre plan de dimensionnement, vous devez supprimer toutes les actions de dimensionnement précédemment planifiées dont vous n'avez plus besoin en accédant aux consoles à partir desquelles elles ont été créées. AWS Auto Scaling ne crée pas d'action de dimensionnement prédictif qui chevauche une action de dimensionnement planifiée existante.
- Vos paramètres personnalisés pour une capacité minimale et maximale, ainsi que d'autres paramètres utilisés pour le dimensionnement dynamique, s'affichent dans d'autres consoles. Cependant, après avoir créé un plan de dimensionnement, nous vous recommandons de ne pas modifier ces paramètres à partir d'autres consoles, car votre plan de dimensionnement ne reçoit pas les mises à jour d'autres consoles.
- Votre plan de dimensionnement peut contenir des ressources de plusieurs services, mais chaque ressource ne peut être que dans un seul plan de dimensionnement à la fois.

Éviter l' ActiveWithProblems erreur

Une erreur « ActiveWithProblems » peut se produire lors de la création d'un plan de dimensionnement ou lors de l'ajout de ressources à un plan de dimensionnement. L'erreur se produit lorsque le plan de dimensionnement est actif, mais que la configuration de dimensionnement pour une ou plusieurs ressources n'a pas pu être appliquée.

Habituellement, cela se produit parce qu'une ressource comporte déjà une stratégie de dimensionnement ou qu'un groupe Auto Scaling ne répond pas aux exigences minimales du dimensionnement prédictif.

Si l'une de vos ressources comporte déjà des stratégies de dimensionnement issues de diverses consoles de services, AWS Auto Scaling ne remplace pas ces autres stratégies de dimensionnement ou n'en créer de nouvelles par défaut. Vous pouvez éventuellement supprimer les politiques de dimensionnement existantes et les remplacer par des politiques de dimensionnement de suivi des cibles créées à partir de la AWS Auto Scaling console. Pour ce faire, activez le paramètre Replace external scaling policies (Remplacer des stratégies de dimensionnement externes) pour chaque ressource qui comportant des stratégies de dimensionnement à remplacer.

Dans le cas du dimensionnement prédictif, nous vous recommandons d'attendre 24 heures après la création d'un groupe Auto Scaling avant de configurer cette fonction. Vous devez au minimum disposer de 24 heures de données historiques pour générer la prévision initiale. Si le groupe dispose de moins de 24 heures de données historiques et que la mise à l'échelle prédictive est activée, le

plan de dimensionnement ne peut pas générer une prévision avant la prochaine période de prévision une fois que le groupe a collecté le volume de données requis. Toutefois, vous pouvez également modifier et enregistrer le plan de dimensionnement pour redémarrer le processus de prévision dès que les 24 heures de données sont disponibles.

Démarrez avec les plans de dimensionnement

Avant de créer un plan de dimensionnement à utiliser avec votre application, étudiez minutieusement votre application lorsqu'elle fonctionne dans le cloud AWS Cloud. Notez les informations suivantes :

- Indique si vous disposez de stratégies de dimensionnement existantes créées à partir d'autres consoles. Au moment de créer votre plan de dimensionnement, vous pouvez remplacer les stratégies de dimensionnement existantes ou les conserver (sans être autorisé à apporter des modifications à leurs valeurs).
- L'utilisation cible la plus logique pour chaque ressource évolutive dans votre application en fonction de la ressource dans son ensemble. Par exemple, la quantité de processeur que les EC2 instances d'un groupe Auto Scaling sont censées utiliser par rapport à leur processeur disponible. Ou, pour un service comme DynamoDB qui utilise un modèle de débit alloué, la quantité d'activité de lecture et d'écriture qu'une table ou un index est censé utiliser par rapport au débit disponible. En d'autres termes, le rapport entre la capacité consommée et la capacité alloué. Vous pouvez modifier l'utilisation cible à tout moment après avoir créé votre plan de dimensionnement.
- Combien de temps prennent le lancement et la configuration d'un serveur. Cela vous permet de configurer une fenêtre pour que chaque EC2 instance se réchauffe après le lancement afin de garantir qu'un nouveau serveur ne soit pas lancé alors que le précédent est toujours en cours de lancement.
- L'historique des métriques est-il suffisamment long pour être utilisé avec le dimensionnement prédictif (si vous utilisez des groupes Auto Scaling nouvellement créés) ? En général, le fait d'avoir plus de 14 jours entiers de données historiques se traduit en des prévisions plus précises. Le minimum est de 24 heures.

Mieux vous comprenez votre application, plus votre plan de dimensionnement est efficace.

Les tâches suivantes vous aideront à vous familiariser avec les plans de dimensionnement. Vous allez créer un plan de dimensionnement pour un groupe Auto Scaling unique et permet un dimensionnement prédictif et dynamique.

Tâches

- Étape 1 : Rechercher vos ressources évolutives
- Étape 2 : Spécifier la stratégie de dimensionnement
- Étape 3 : Configurer les paramètres avancés (facultatif)

- Étape 4 : Créer votre plan de dimensionnement
- Étape 5 : nettoyer
- Étape 6 : étapes suivantes

Étape 1 : Rechercher vos ressources évolutives

Cette section comprend une présentation pratique de la création de plans de dimensionnement dans la console AWS Auto Scaling . S'il s'agit de votre premier plan de dimensionnement, nous vous recommandons de commencer par créer un exemple de plan de dimensionnement à l'aide d'un groupe Amazon EC2 Auto Scaling.

Prérequis

Pour vous entraîner à utiliser un plan de dimensionnement, créez un groupe Auto Scaling. Lancez au moins une EC2 instance Amazon dans le groupe Auto Scaling. Pour plus d'informations, consultez Get started with Amazon EC2 Auto Scaling dans le guide de l'utilisateur Amazon EC2 Auto Scaling.

Utilisez un groupe Auto Scaling dont CloudWatch les métriques sont activées pour que les données de capacité apparaissent sur les graphiques disponibles lorsque vous avez terminé l'assistant de création d'un plan de dimensionnement. Pour plus d'informations, consultez la section Surveillance CloudWatch des métriques pour vos groupes et instances Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling.

Générez une certaine charge pendant quelques jours ou plus afin de disposer de données CloudWatch métriques pour la fonction de mise à l'échelle prédictive, si possible.

Assurez-vous de disposer des autorisations requises pour travailler avec les plans de dimensionnement. Pour de plus amples informations, veuillez consulter Gestion des identités et des accès pour les plans de mise à l'échelle.

Ajout de votre groupe Auto Scaling à votre nouveau plan de dimensionnement

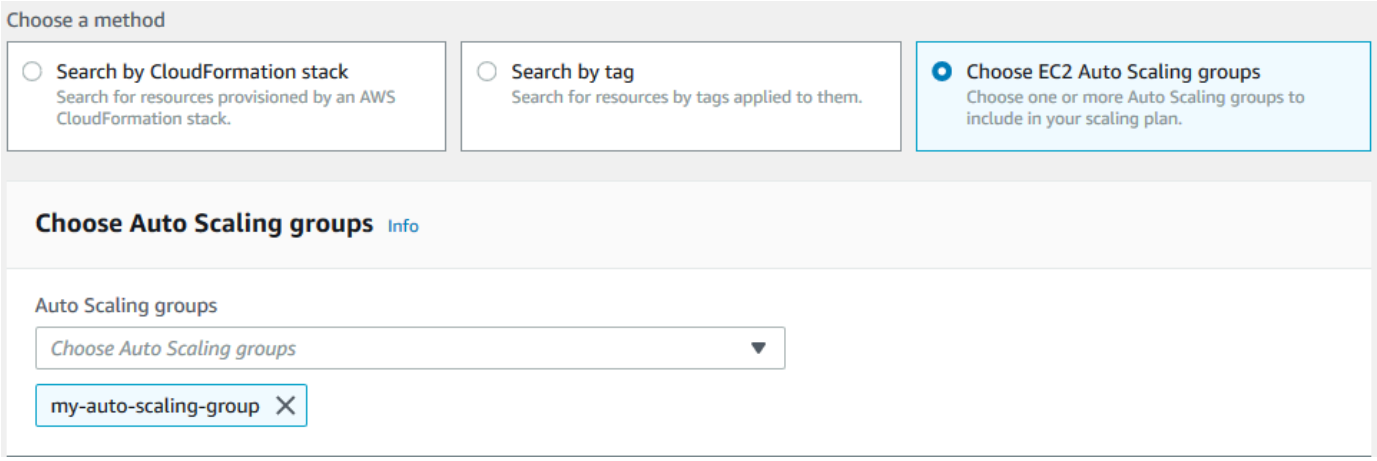
Lorsque vous créez un plan de dimensionnement à partir de la console, il vous aide dans un premier temps à trouver vos ressources évolutives. Avant de poursuivre, confirmez que vous respectez les exigences suivantes :

- Vous avez créé un groupe Auto Scaling et lancé au moins une EC2 instance, comme décrit dans la section précédente.
- Le groupe Auto Scaling que vous avez créé existe depuis au moins 24 heures.

Pour commencer la création d’un plan de dimensionnement

1. Ouvrez la AWS Auto Scaling console à l'adresse https://console.aws.amazon.com/awsautoscaling/.
2. Dans la barre de navigation située en haut de l'écran, choisissez la même région que celle utilisée lorsque vous avez créé votre groupe Auto Scaling.
3. Depuis la page de bienvenue, sélectionnez Get started (Mise en route).
4. Dans la pageRechercher des ressources évolutives, effectuez l'une des actions suivantes :
   - Choisissez Rechercher par CloudFormation pile, puis choisissez la AWS CloudFormation pile à utiliser.
   - ChoisissezRechercher par balise. Ensuite, pour chaque balise, sélectionnez une clé de balise dans Key (Clé) et des valeurs de balise dans Value (Valeur). Pour ajouter des balises, choisissez Add another row (Ajouter une autre ligne). Pour supprimer des balises, choisissez Remove (Supprimer).
   - Choisissez Choose EC2 Auto Scaling groups, puis choisissez un ou plusieurs groupes Auto Scaling.

   > **Note**
   >
   > Pour un didacticiel d'introduction, choisissez Choose EC2 Auto Scaling groups, puis choisissez le groupe Auto Scaling que vous avez créé.



5. ChoisissezSuivantpour poursuivre le processus de création d'un plan de dimensionnement.

En savoir plus sur la découverte de vos ressources évolutives

Si vous avez déjà créé un exemple de plan de mise à l'échelle et que vous souhaitez en créer d'autres, consultez les scénarios suivants pour utiliser une CloudFormation pile ou un ensemble de balises de manière plus détaillée. Vous pouvez utiliser cette section pour décider de choisir l'option Rechercher par CloudFormation pile ou Rechercher par tag pour découvrir vos ressources évolutives lorsque vous utilisez la console pour créer votre plan de dimensionnement.

Lorsque vous choisissez l'option Rechercher par CloudFormation pile ou Rechercher par balise à l'étape 1 de l'assistant de création d'un plan de dimensionnement, les ressources évolutives associées à la pile ou à l'ensemble de balises sont mises à la disposition du plan de dimensionnement. Vous pouvez choisir les ressources à inclure ou exclure au fur et à mesure que vous définissez votre plan de dimensionnement.

Découvrir des ressources évolutives à l'aide d'une CloudFormation pile

Lorsque vous utilisez CloudFormation, vous travaillez avec des piles pour provisionner des ressources. Toutes les ressources d'une pile sont définies par son modèle. Votre plan de dimensionnement ajoute une couche d'orchestration au-dessus de la pile, ce qui facilite la configuration de la mise à l'échelle pour plusieurs ressources. Sans plan de dimensionnement, vous devez configurer la mise à l'échelle pour chaque ressource scalable individuellement. Cela implique de déterminer l'ordre de mise en service des ressources et des stratégies de dimensionnement, et de comprendre les subtilités du fonctionnement de ces dépendances.

Dans la AWS Auto Scaling console, vous pouvez sélectionner une pile existante pour la scanner à la recherche de ressources pouvant être configurées pour un dimensionnement automatique. AWS Auto Scaling ne trouve que les ressources définies dans la pile sélectionnée. La recherche ne traverse pas les piles imbriquées.

Pour que vos services ECS soient détectables dans une CloudFormation pile, la AWS Auto Scaling console doit savoir quel cluster ECS exécute le service. Cela nécessite que vos services ECS soient dans la même CloudFormation pile que le cluster ECS qui exécute le service. Sinon, ils doivent faire partie du cluster par défaut. Pour être identifié correctement, le nom du service ECS doit également être unique dans chacun de ces clusters ECS.

Pour plus d'informations CloudFormation, voir Qu'est-ce que c'est AWS CloudFormation ? dans le guide de AWS CloudFormation l'utilisateur.

Découverte de ressources évolutives à l'aide de balises

Les balises fournissent des métadonnées qui peuvent être utilisées pour découvrir les ressources évolutives associées dans la AWS Auto Scaling console, à l'aide de filtres de balises.

Les balises permettent de rechercher les ressources suivantes :

- Clusters DB Aurora
- Groupes Auto Scaling
- Tables DynamoDB et index secondaires globaux

Si votre recherche porte sur plusieurs balises, chaque ressource doit posséder toutes les balises répertoriées afin d'être détectée.

Pour plus d'information sur les balises, consultez la documentation suivante.

- Découvrez commentbaliser des clusters Auroradans leGuide de l'utilisateur Amazon Aurora.
- Découvrez comment étiqueter les groupes Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling.
- Découvrez commentbalise des ressources DynamoDBdans leAmazon DynamoDB Developer Guide.

Étape 2 : Spécifier la stratégie de dimensionnement

Utilisez la procédure suivante pour spécifier les stratégies de dimensionnement pour les ressources que vous avez trouvées à l'étape précédente.

Pour chaque type de ressource, AWS Auto Scaling choisit la métrique la plus couramment utilisée pour déterminer la quantité de ressource utilisée à un moment donné. Vous choisissez la stratégie de dimensionnement la plus appropriée pour optimiser les performances de votre application en fonction de cette métrique. Lorsque vous activez la fonction de mise à l'échelle dynamique et la fonction de mise à l'échelle prédictive, la stratégie de dimensionnement est partagée par ces fonctions. Pour de plus amples informations, veuillez consulter Fonctionnement des plans de dimensionnement.

Les stratégies de dimensionnement suivantes sont disponibles :

- Optimisation de la disponibilité :AWS Auto Scaling redimensionne automatiquement les ressources pour maintenir le taux d'utilisation des ressources à 40 %. Cette option est utile lorsque votre application a des besoins de mise à l'échelle urgents et parfois imprévisibles.
- Équilibrez la disponibilité et AWS Auto Scaling le coût : redimensionnez automatiquement les ressources pour maintenir le taux d'utilisation des ressources à 50 %. Cette option vous permet de conserver une haute disponibilité, tout en réduisant les coûts.
- Optimisation en fonction des coûts :AWS Auto Scaling redimensionne automatiquement les ressources pour maintenir le taux d'utilisation des ressources à 70 %. Cette option est utile pour diminuer les coûts si votre application peut gérer le fait de disposer d'une capacité tampon réduite en cas de changements imprévus de la demande.

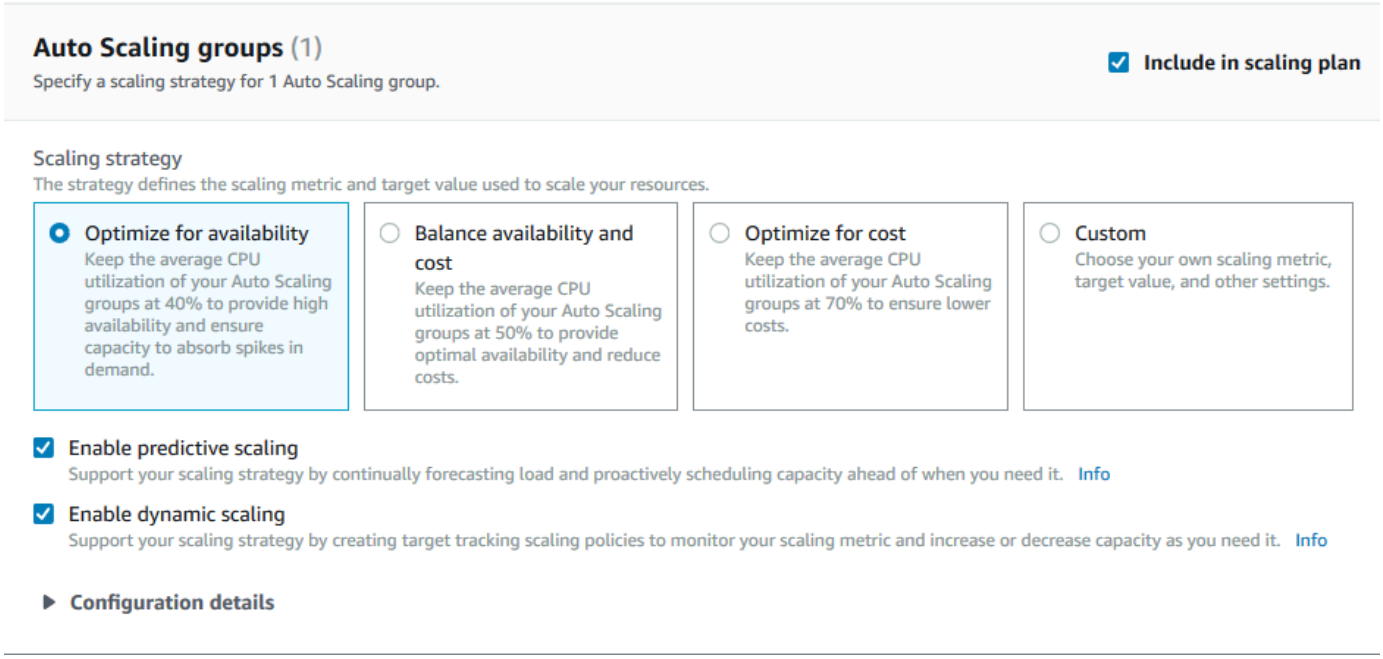
Par exemple, le plan de dimensionnement configure votre groupe Auto Scaling pour ajouter ou supprimer des EC2 instances Amazon en fonction de la quantité de processeur utilisée en moyenne pour toutes les instances du groupe. Vous pouvez choisir d'optimiser l'utilisation en termes de disponibilité, de coûts ou d'une combinaison des deux, en modifiant la stratégie de dimensionnement.

Vous pouvez également configurer une stratégie personnalisée si une stratégie prête à l'emploi ne répond pas à vos besoins. Avec une stratégie personnalisée, vous pouvez modifier la valeur d'utilisation cible, choisir une autre métrique, ou les deux.

> **Important**
>
> Dans la cadre du didacticiel pour débutant, effectuez uniquement la première étape de la procédure suivante, puis choisissez Next (Suivant) pour continuer.

Pour spécifier les stratégies de dimensionnement

1. Sur la page Specify scaling strategy (Spécifier la stratégie de dimensionnement), pour Scaling plan details (Détails du plan de dimensionnement), Name (Nom), entrez un nom pour votre plan de dimensionnement. Le nom de votre plan de dimensionnement doit être unique parmi tous vos plans de dimensionnement pour la région. Il peut comporter 128 caractères maximum et il ne doit pas contenir des barres obliques «/» ou des deux-points « : ».
2. Toutes les ressources incluses sont répertoriées par type de ressource. PourGroupes Auto Scaling, procédez comme suit :



   a. Ignorez cette étape pour utiliser la stratégie et les mesures de dimensionnement par défaut. Pour utiliser plutôt une stratégie ou des mesures de dimensionnement différentes, procédez comme suit :

      i. PourStratégie de dimensionnement, choisissez la stratégie de dimensionnement souhaitée.

Pour le didacticiel d'introduction, veillez à choisirOptimiser pour la disponibilité. Cela spécifie que l'utilisation moyenne du processeur de votre groupe Auto Scaling sera maintenue à 40 %.

ii. Si vous avez choisiPersonnalisée, développez laDétails de configurationpour choisir les mesures et la valeur cible souhaitées.

- Pour Scaling metric (Métrique de dimensionnement), choisissez la métrique de dimensionnement souhaitée.
- PourValeur cible, choisissez la valeur cible souhaitée, telle que l'utilisation cible ou le débit cible pendant n'importe quel intervalle d'une minute.
- Pour Load metric (Métrique de charge) [groupes Auto Scaling uniquement], choisissez la métrique de charge appropriée à utiliser pour le dimensionnement prédictif.
- Sélectionnez Remplacer les politiques de dimensionnement externes pour spécifier qui AWS Auto Scaling peut supprimer les politiques de dimensionnement créées précédemment en dehors du plan de dimensionnement (par exemple depuis d'autres consoles) et les remplacer par de nouvelles politiques de dimensionnement de suivi des cibles créées par le plan de dimensionnement.

b. (Facultatif) Le dimensionnement prédictif est activé par défaut pour vos groupes Auto Scaling. Pour désactiver le dimensionnement prédictif pour vos groupes Auto Scaling, désactivez Enable predictive scaling (Activer le dimensionnement prédictif).

c. (Facultatif) Le dimensionnement dynamique est activé par défaut pour chaque type de ressource. Pour désactiver la mise à l'échelle dynamique pour un type de ressource, désactivez Enable dynamic scaling (Activer le dimensionnement dynamique).

d. (Facultatif) Par défaut, lorsque vous spécifiez une application source à partir de laquelle plusieurs ressources évolutives sont découvertes, tous les types de ressource sont automatiquement inclus dans votre plan de dimensionnement. Pour omettre un type de ressource de votre plan de dimensionnement, désélectionnez Include in scaling plan (Inclure dans le plan de dimensionnement).

3. (Facultatif) Pour spécifier une stratégie de dimensionnement pour un autre type de ressource, répétez les étapes précédentes.

4. Lorsque vous avez terminé, choisissezSuivantpour poursuivre le processus de création d'un plan de dimensionnement.

Étape 3 : Configurer les paramètres avancés (facultatif)

Maintenant que vous avez spécifié la stratégie de dimensionnement à utiliser pour chaque type de ressource, vous pouvez choisir de personnaliser l'un des paramètres par défaut ressource par ressource en utilisant l'étape Configurer les paramètres avancés. Pour chaque type de ressource, vous pouvez personnaliser plusieurs groupes de paramètres. Toutefois, dans la plupart des cas, les paramètres par défaut devraient être parfaits, à l'exception éventuelle des valeurs pour la capacité minimale et maximale qui doivent être soigneusement ajustées.

Ignorez cette procédure si vous souhaitez conserver les paramètres par défaut. Ces paramètres peuvent être modifiés à tout moment en modifiant le plan de dimensionnement.

> **Important**
>
> Dans la cadre du tutoriel pour débutant, nous allons effectuer quelques modifications pour mettre à jour la capacité maximale de votre groupe Auto Scaling et activer le dimensionnement prédictif en mode prévisionnel uniquement. Même si vous n'avez pas besoin de personnaliser tous les paramètres pour ce didacticiel, nous allons également brièvement examiner les paramètres dans chaque section.

Paramètres généraux

Utilisez cette procédure pour afficher et personnaliser les paramètres que vous avez spécifiés à l'étape précédente, ressource par ressource. Vous pouvez également personnaliser la capacité minimale et maximale pour chaque ressource.

Pour afficher et personnaliser les paramètres généraux

1. Sur la page Configure advanced settings (Configurer les paramètres avancés), choisissez la flèche située à gauche de l'un des en-têtes de section pour développer cette section. Pour le tutoriel, développez la section Auto Scaling groups (Groupes Auto Scaling).
2. Dans le tableau qui s'affiche, choisissez le groupe Auto Scaling que vous utilisez dans ce tutoriel.
3. Laissez l'option Include in scaling plan (Inclure dans le plan de dimensionnement) sélectionnée. Si cette option n'est pas sélectionnée, la ressource est omise du plan de dimensionnement. Si vous n'incluez pas au moins une ressource, le plan de dimensionnement ne peut pas être créé.
4. Pour développer la vue et voir les détails de la section Paramètres généraux, sélectionnez la flèche à gauche de l'en-tête de section.

5. Vous pouvez effectuer des choix pour les éléments suivants. Pour ce didacticiel, recherchez le paramètre Maximum capacity (Capacité maximale) et saisissez la valeur 3 à la place de la valeur actuelle.

   - Scaling strategy (Stratégie de dimensionnement) – Vous permet d'optimiser la disponibilité, le coût, ou un équilibre des deux, ou de spécifier une stratégie personnalisée.
   - Enable dynamic scaling (Activer le dimensionnement dynamique) – Si ce paramètre est désactivé, la ressource sélectionnée ne peut pas être mise à l'échelle selon une configuration de suivi des objectifs et de l'échelonnement.
   - Enable predictive scaling (Activer le dimensionnement prédictif) – [Groupes Auto Scaling uniquement] Si ce paramètre est désactivé, le groupe sélectionné ne peut pas être mis à l'échelle en utilisant le dimensionnement prédictif.
   - Scaling metric (Métrique de mise à l'échelle) – Spécifie la métrique de dimensionnement à utiliser. Si vous choisissez Custom (Personnalisé), vous pouvez spécifier une métrique personnalisée à utiliser à la place des métriques prédéfinies disponibles dans la console. Pour plus d'informations, consultez la rubrique suivante de cette section.
   - Target value (Valeur cible) – Spécifie la valeur d'utilisation cible à utiliser.
   - Load metric (Métrique de charge) – [groupes Auto Scaling uniquement] Spécifie la métrique de charge à utiliser. Si vous choisissez Custom (Personnalisé), vous pouvez spécifier une métrique personnalisée à utiliser à la place des métriques prédéfinies disponibles dans la console. Pour plus d'informations, consultez la rubrique suivante de cette section.
   - Capacité minimale — Spécifie la capacité minimale de la ressource. AWS Auto Scaling garantit que votre ressource ne descend jamais en dessous de cette taille.
   - Capacité maximale — Spécifie la capacité maximale de la ressource. AWS Auto Scaling garantit que votre ressource ne dépasse jamais cette taille.

   > **Note**
   >
   > Lorsque vous utilisez la mise à l'échelle prédictive, vous pouvez, si vous le souhaitez, choisir un autre comportement de capacité maximale à utiliser en fonction de la capacité de prévision. Ce paramètre se trouve dans la section Predictive scaling settings (Paramètres de mise à l'échelle prédictive).

Métriques personnalisées

AWS Auto Scaling fournit les métriques les plus couramment utilisées pour le dimensionnement automatique. Toutefois, selon vos besoins, vous pouvez préférer obtenir les données de différentes métriques au lieu de celles proposées dans la console. Amazon CloudWatch propose de nombreux indicateurs parmi lesquels choisir. CloudWatch vous permet également de publier vos propres statistiques.

Vous utilisez le JSON pour spécifier une métrique CloudWatch personnalisée. Avant de suivre ces instructions, nous vous recommandons de vous familiariser avec le guide de CloudWatch l'utilisateur Amazon.

Pour spécifier une métrique personnalisée, vous devez créer une charge utile au format JSON en utilisant un ensemble de paramètres obligatoires de modèle. Vous ajoutez les valeurs de chaque paramètre à partir de CloudWatch. Nous fournissons le modèle dans le cadre des options personnalisées pour Scaling metric (Métrique de mise à l'échelle) et Load metric (Métrique de charge) dans les paramètres avancés de votre plan de dimensionnement.

JSON représente les données de deux manières :

- Un objet, qui est une collection désordonnée de paires nom-valeur. Un objet est défini entre des crochets de gauche ({) et de droite (}). Chaque paire nom-valeur commence par le nom, suivi de deux points, suivi de la valeur. Les paires nom-valeur sont séparées par des virgules.
- Un éventail, qui est un ensemble ordonné de valeurs. Un éventail est défini entre des parenthèses carrées de gauche ([) et de droite (]). Les éléments de l'éventail sont séparés par des virgules.

Voici un exemple du modèle JSON avec des exemples de valeurs pour chaque paramètre :

```
{
  "MetricName": "MyBackendCPU",
  "Namespace": "MyNamespace",
  "Dimensions": [
    {
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }
  ],
  "Statistic": "Sum"
}
```

Pour plus d'informations, consultez Spécification de métrique de dimensionnement personnalisée et Spécification de métrique de charge personnalisée dans Référence d'API AWS Auto Scaling .

Paramètres de mise à l'échelle dynamique

Utilisez cette procédure pour afficher et personnaliser les paramètres de la politique de dimensionnement du suivi des cibles AWS Auto Scaling créée.

Pour afficher et personnaliser les paramètres de dimensionnement dynamique

1. Pour développer la vue et voir les détails de la section Paramètres de mise à l'échelle dynamique, sélectionnez la flèche à gauche de l'en-tête de section.
2. Vous pouvez effectuer des choix pour les éléments suivants. Cependant, les paramètres par défaut conviennent pour ce didacticiel.
   - Replace external scaling policies (Remplacer les stratégies de dimensionnement externes) – Si ce paramètre est désactivé, il conserve tout de même les stratégies de dimensionnement existantes créées en dehors du plan de dimensionnement et n'en crée pas de nouvelles.
   - Disable scale-in (Désactiver la diminution en charge) – Si ce paramètre est désactivé, la diminution en charge automatique pour réduire la capacité actuelle de la ressource est autorisée lorsque la métrique spécifiée est en dessous de la valeur cible.
   - Cooldown (Stabilisation) – Crée des temps de stabilisation de montée ou de diminution en charge. Le temps de stabilisation représente le temps passé par une stratégie de dimensionnement à attendre qu'une activité de mise à l'échelle précédente prenne effet. Pour plus d'informations, consultez Temps de stabilisation dans le Guide de l'utilisateur Application Auto Scaling. (Ce paramètre ne s'affiche pas si la ressource est un groupe Auto Scaling.)
   - Préchauffage de l'instance — [Groupes Auto Scaling uniquement] Contrôle le temps qui s'écoule avant qu'une instance nouvellement lancée ne commence à contribuer aux métriques. CloudWatch Pour plus d'informations, consultez Instance warmup dans le manuel Amazon EC2 Auto Scaling User Guide.

Paramètres de mise à l'échelle prédictive

Si votre ressource est un groupe Auto Scaling, utilisez cette procédure pour afficher et personnaliser les paramètres AWS Auto Scaling utilisés pour le dimensionnement prédictif.

Pour afficher et personnaliser les paramètres de dimensionnement prédictive

1. Pour développer la vue et voir les détails de la section Paramètres de mise à l'échelle prédictive, sélectionnez la flèche à gauche de l'en-tête de section.
2. Vous pouvez effectuer des choix pour les éléments suivants. Pour ce didacticiel, remplacez le mode Predictive scaling mode (Mise à l'échelle prédictive) par le mode Forecast only (Prévision uniquement).
   - Predictive scaling mode (Mode de mise à l'échelle prédictive) – Spécifie le mode de dimensionnement. La valeur par défaut est Forecast and scale (Prévision et mise à l'échelle). Si vous le remplacez par Forecast only (Prévision uniquement), le plan de dimensionnement peut prévoir une capacité future, mais il n'applique pas les actions de mise à l'échelle.
   - Pre-launch instances (Instances de pré-lancement) – Ajuste les actions de mise à l'échelle pour qu'elles s'exécutent plus tôt lors d'une montée en puissance. Par exemple, les prévisions indiquent qu'il faut ajouter de la capacité à 10 h 00 et que le temps de tampon est de 5 minutes (300 secondes). L'heure d'exécution de l'action de dimensionnement correspondante est ensuite à 9 h 55. C'est utile pour les groupes Auto Scaling, dans lesquels il peut s'écouler quelques minutes entre le lancement d'une instance et sa mise en service. Le temps réel peut varier, car il dépend de plusieurs facteurs, comme la taille de l'instance et si des scripts de démarrage doivent être terminés. La durée par défaut est 300 secondes.
   - Max capacity behavior (Comportement de capacité maximale) – Contrôle si la ressource sélectionnée peut augmenter l'échelle au-delà de la capacité maximale lorsque la capacité de prévision est proche de la capacité maximale actuellement spécifiée ou la dépasse. La valeur par défaut est Enforce the maximum capacity setting (Appliquer le paramètre de capacité maximale).
     - Appliquez le paramètre de capacité maximale : vous AWS Auto Scaling ne pouvez pas augmenter la capacité des ressources au-delà de la capacité maximale. La capacité maximale est appliquée comme une limite stricte.
     - Définissez la capacité maximale pour qu'elle soit égale à la capacité prévue :AWS Auto Scaling vous pouvez augmenter la capacité des ressources par rapport à la capacité maximale pour qu'elle soit égale mais pas supérieure à la capacité prévue.
     - Augmenter la capacité maximale par rapport à la capacité prévue :AWS Auto Scaling permet d'augmenter la capacité des ressources au-delà de la capacité maximale en fonction d'une valeur tampon spécifiée. L'objectif est d'accorder à la politique de suivi des objectifs et d'échelonnement une capacité supplémentaire en cas de trafic imprévu.

- Max capacity behavior buffer (Tampon avec comportement de capacité maximale) – Si vous avez choisi Increase maximum capacity above forecast capacity (Augmenter la capacité maximale au-delà de la capacité des prévisions), choisissez la taille du tampon de capacité à utiliser lorsque la capacité de prévision est proche de la capacité maximale ou la dépasse. La valeur est spécifiée sous la forme d'un pourcentage par rapport à la capacité prévue. Par exemple, avec un tampon de 10 %, si la capacité de prévision est 50 et que la capacité maximale est 40, la capacité maximale effective est 55.

3. Une fois que vous avez terminé de personnaliser les paramètres, cliquez sur Suivant.

> **Note**
>
> Pour annuler vos modifications, sélectionnez les ressources et choisissez Revenir au paramètre initial. Ceci réinitialise les ressources sélectionnées à leur dernier état connu au sein du plan de mise à l'échelle.

Étape 4 : Créer votre plan de dimensionnement

Sur la page Vérifier et créer, examinez les détails de votre plan de mise à l'échelle et choisissez Créer un plan de mise à l'échelle. Vous êtes dirigé vers la page qui affiche le statut de votre plan de dimensionnement. L'achèvement de la création du plan de dimensionnement peut prendre quelques instants pendant que vos ressources sont mises à jour.

Grâce à la mise à l'échelle prédictive, AWS Auto Scaling analyse l'historique de la métrique de charge spécifiée au cours des 14 derniers jours (un minimum de 24 heures de données est requis) afin de générer une prévision pour les deux jours à venir. Il planifie ensuite des actions de dimensionnement pour faire correspondre la capacité de ressource à la prévision pour chaque heure de la période de prévision.

Une fois la création du plan de dimensionnement terminée, affichez ses détails en choisissant son nom dans l'écran Scaling plans (Plans de dimensionnement).

(Facultatif) Affichage des informations de mise à l'échelle pour une ressource

Utilisez cette procédure pour afficher les informations de mise à l'échelle créées pour une ressource.

Les données sont présentées de la manière suivante :

- Graphiques montrant les données récentes de l'historique des mesures provenant de CloudWatch.
- Graphiques de mise à l'échelle prédictive montrant les prévisions de charge et les prévisions de capacité basées sur les données de AWS Auto Scaling.
- Un tableau qui répertorie toutes les actions de mise à l'échelle prédictive planifiées pour la ressource.

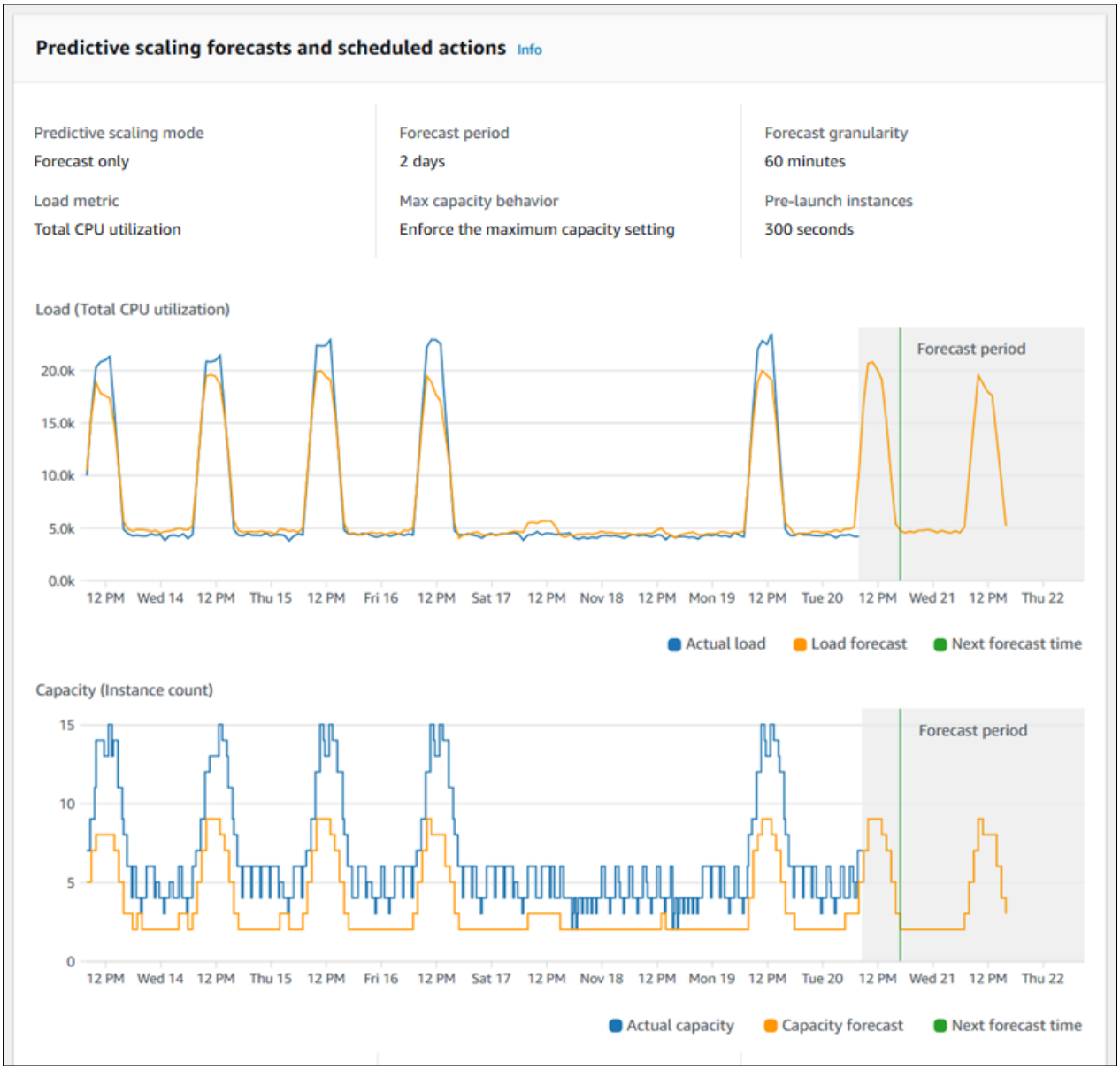
Pour afficher des informations de mise à l'échelle pour une ressource

1. Ouvrez la AWS Auto Scaling console à l'adresse https://console.aws.amazon.com/awsautoscaling/.
2. Sur la page Plans de dimensionnement, choisissez le plan de dimensionnement.
3. Sur la page Détails du plan de dimensionnement, choisissez la ressource à afficher.

Surveillance et évaluation des prévisions

Lorsque votre plan de dimensionnement est opérationnel, vous pouvez surveiller la prévision de charge, la capacité de prévision et les actions de mise à l'échelle pour examiner les performances de la mise à l'échelle prédictive. Toutes ces données sont disponibles dans la AWS Auto Scaling console pour tous les groupes Auto Scaling activés pour le dimensionnement prédictif. Gardez à l'esprit que votre plan de dimensionnement a besoin d'au moins 24 heures de données de charge historiques pour effectuer la prévision initiale.

Dans l'exemple suivant, la partie gauche de chaque graphique illustre un modèle d'historique. La partie droite présente la prévision générée par le plan de mise à l'échelle pour la période de prévisions. Les valeurs réelles et prévisionnelles (en bleu et orange) sont comparées.



AWS Auto Scaling apprend automatiquement à partir de vos données. Il commence par effectuer une prévision de charge. Un calcul prévisionnel de capacité détermine ensuite le nombre minimum d'instances requises pour prendre en charge de l'application. En fonction de la prévision de capacité, AWS Auto Scaling planifie les actions de mise à l'échelle qui mettent à l'échelle le groupe Auto Scaling avant les changements de charge prévus. Si le dimensionnement dynamique est activé (recommandé), le groupe Auto Scaling peut ajouter de la capacité supplémentaire (ou en supprimer) en fonction de l'utilisation actuelle du groupe d'instances.

Lorsque vous évaluez la façon dont la mise à l'échelle prédictive s'effectue, surveillez la manière dont les valeurs réelles et prévisionnelles correspondent au fil du temps. Lorsque vous créez un plan de mise à l'échelle, il AWS Auto Scaling fournit des graphiques basés sur les données réelles les plus

récentes. Il fournit également une prévision initiale pour les prochaines 48 heures. Toutefois, lorsque le plan de mise à l'échelle est créé, il y a très peu de données prévisionnelles auxquelles comparer les données réelles. Patientez jusqu'à ce que le plan de mise à l'échelle ait obtenu des valeurs prévisionnelles pour quelques périodes avant de comparer les valeurs prévisionnelles historiques avec les valeurs réelles. Après quelques jours de prévisions quotidiennes, vous disposerez d'un échantillon plus important de valeurs prévisionnelles auquel comparer les valeurs réelles.

Pour les modèles qui se produisent sur une base quotidienne, l'intervalle de temps entre la création de votre plan de mise à l'échelle et l'évaluation de l'efficacité de la prévision peut être de quelques jours seulement. Cependant, cette durée est insuffisante pour évaluer la prévision basée sur une modification de modèle récente. Par exemple, supposons que vous examinez les prévisions pour un groupe Auto Scaling qui a lancé une nouvelle campagne de marketing une semaine auparavant. La campagne augmente considérablement votre trafic Web pour les deux mêmes jours de chaque semaine. Dans des situations comme celle-ci, nous vous recommandons d'attendre que le groupe ait collecté des données pendant une ou deux semaines avant d'évaluer l'efficacité de la prévision. La même recommandation s'applique pour un nouveau groupe Auto Scaling qui vient seulement de commencer à recueillir des données de métriques.

Si les valeurs réelles et prévisionnelles ne correspondent pas après une surveillance sur une période plus longue, vous devez également réfléchir au choix de la métrique de charge. Pour être efficace, la métrique de charge doit représenter une mesure fiable et précise de la charge totale sur toutes les instances du groupe Auto Scaling. La métrique de charge est essentielle pour la mise à l'échelle prédictive. Si vous choisissez une métrique de charge non optimale, cela peut empêcher le dimensionnement prédictif d'établir des prévisions précises de charge et de capacité et de planifier des ajustements de capacité adaptés pour votre groupe Auto Scaling.

Étape 5 : nettoyer

Une fois que vous avez terminé le tutoriel de mise en route, vous pouvez choisir de conserver votre plan de dimensionnement. Toutefois, si vous n'utilisez pas activement votre plan de dimensionnement, nous vous conseillons de le supprimer afin que votre compte ne génère pas de frais inutiles.

La suppression d'un plan de dimensionnement entraîne la suppression des politiques de dimensionnement du suivi des cibles, des CloudWatch alarmes associées et des actions de dimensionnement prédictives AWS Auto Scaling créées en votre nom.

La suppression d'un plan de dimensionnement n'entraîne pas la suppression de votre AWS CloudFormation stack, de votre groupe Auto Scaling ou d'autres ressources évolutives.

Pour supprimer un plan de dimensionnement

1. Ouvrez la AWS Auto Scaling console à l'adresse https://console.aws.amazon.com/awsautoscaling/.
2. Sur la page Scaling plans (Plans de dimensionnement), sélectionnez le plan de dimensionnement que vous avez créé pour ce didacticiel, puis choisissez Delete (Supprimer).
3. Lorsque vous êtes invité à confirmer l'opération, choisissez Supprimer.

Après avoir supprimé votre plan de mise à l'échelle, vos ressources ne reviennent pas à leur capacité originale. Par exemple, si votre groupe Auto Scaling est dimensionné à 10 instances au moment où vous supprimez le plan de dimensionnement, il sera toujours dimensionné à 10 instances après la suppression. Vous pouvez mettre à jour la capacité de ressources spécifiques en accédant à la console pour chaque service.

Suppression de votre groupe Auto Scaling

Pour éviter que votre compte ne soit débité par Amazon EC2 , vous devez également supprimer le groupe Auto Scaling que vous avez créé pour ce didacticiel.

Pour step-by-step obtenir des instructions, consultez Supprimer votre groupe Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling.

Étape 6 : étapes suivantes

Maintenant que vous êtes familiarisé avec les plans de dimensionnement et certaines de ses fonctionnalités, vous pouvez essayer de créer votre propre modèle de plan de dimensionnement à l'aide de AWS CloudFormation.

Un AWS CloudFormation modèle est un fichier texte au format JSON ou YAML qui décrit l'infrastructure Amazon Web Services nécessaire pour exécuter une application ou un service, ainsi que les interconnexions entre les composants de l'infrastructure. Avec AWS CloudFormation, vous déployez et gérez un ensemble de ressources associé sous forme de pile. AWS CloudFormation est disponible sans frais supplémentaires et vous ne payez que pour les AWS ressources nécessaires à l'exécution de vos applications. Les ressources peuvent consister en n'importe quelle AWS ressource que vous définissez dans le modèle. Pour plus d'informations, reportez-vous à la section AWS CloudFormation Fonctionnement du guide de AWS CloudFormation l'utilisateur.

Dans le Guide de l'utilisateur AWS CloudFormation , nous fournissons un modèle simple pour vous aider à démarrer. L'exemple de modèle est disponible à titre d'exemple dans la AWS::AutoScalingPlans::ScalingPlansection de la documentation de référence du AWS CloudFormation modèle. L'exemple de modèle crée un plan de dimensionnement pour un groupe Auto Scaling unique et permet un dimensionnement prédictif et dynamique.

Pour plus d'informations, consultez Mise en route avec AWS CloudFormation dans le Guide de l'utilisateur AWS CloudFormation .

Migrez votre plan de mise à l'échelle

Vous pouvez passer d'un plan de dimensionnement aux politiques de dimensionnement Amazon EC2 Auto Scaling et Application Auto Scaling.

Processus de migration

- Étape 1 : passez en revue votre configuration existante
- Étape 2 : créer des politiques de dimensionnement prédictives
- Étape 3 : Passez en revue les prévisions générées par les politiques de dimensionnement prédictif
- Étape 4 : préparer la suppression du plan de dimensionnement
- Étape 5 : Supprimer le plan de dimensionnement
- Étape 6 : réactiver la mise à l'échelle dynamique
- Étape 7 : réactiver le dimensionnement prédictif
- Référence Amazon EC2 Auto Scaling pour la migration des politiques de dimensionnement du suivi des cibles
- Référence Application Auto Scaling pour la migration des politiques de dimensionnement du suivi des cibles
- Informations supplémentaires

> **Important**
>
> Pour migrer un plan de dimensionnement, vous devez effectuer plusieurs étapes dans l'ordre exact. Lorsque vous migrez votre plan de dimensionnement, ne le mettez pas à jour, car cela perturberait l'ordre des opérations et risquerait de provoquer un comportement indésirable.

Étape 1 : passez en revue votre configuration existante

Pour déterminer les paramètres de mise à l'échelle que vous devez modifier, utilisez la describe-scaling-planscommande.

```
aws autoscaling-plans describe-scaling-plans \
  --scaling-plan-names my-scaling-plan
```

Notez les éléments que vous souhaitez conserver dans le plan de mise à l'échelle existant, notamment les éléments suivants :

- **`MinCapacity`**— La capacité minimale de la ressource évolutive.
- **`MaxCapacity`**— La capacité maximale de la ressource évolutive.
- **`PredefinedLoadMetricType`**— Une métrique de charge pour une mise à l'échelle prédictive.
- **`PredefinedScalingMetricType`**— Une métrique de mise à l'échelle pour le suivi des cibles (dynamique) et la mise à l'échelle prédictive.
- **`TargetValue`**— La valeur cible de la métrique de mise à l'échelle.

Différences entre les plans de dimensionnement et les politiques de dimensionnement

Il existe quelques différences importantes entre les plans de dimensionnement et les politiques de dimensionnement :

- Une politique de dimensionnement ne peut permettre qu'un seul type de dimensionnement : le dimensionnement par suivi des cibles ou le dimensionnement prédictif. Pour utiliser les deux méthodes de dimensionnement, vous devez créer des politiques distinctes.
- De même, vous devez définir la métrique de dimensionnement pour le dimensionnement prédictif et la métrique de dimensionnement pour le suivi des cibles séparément dans le cadre de leurs politiques respectives.

Étape 2 : créer des politiques de dimensionnement prédictives

Si vous n'utilisez pas le dimensionnement prédictif, passez directement àÉtape 4 : préparer la suppression du plan de dimensionnement.

Pour avoir le temps d'évaluer les prévisions, nous vous recommandons de créer des politiques de dimensionnement prédictif avant les autres politiques de dimensionnement.

Pour tous les groupes Auto Scaling dotés d'une spécification de métrique de charge existante, procédez comme suit pour en faire une politique de dimensionnement prédictif basée sur Amazon EC2 Auto Scaling.

Pour créer des politiques de dimensionnement prédictives

1. Dans un fichier JSON, définissez une `MetricSpecifications` structure comme indiqué dans l'exemple suivant :

```
{
  "MetricSpecifications":[
    {
      ...
    }
  ]
}
```

2. Dans la `MetricSpecifications` structure, pour chaque métrique de charge de votre plan de dimensionnement, créez un `PredefinedLoadMetricSpecification` ou `CustomizedLoadMetricSpecification` en utilisant les paramètres équivalents du plan de dimensionnement.

   Voici des exemples de structure de la section métrique de charge.

   With predefined metrics

```
{
  "MetricSpecifications":[
    {
      "PredefinedLoadMetricSpecification":{
        "PredefinedMetricType":"ASGTotalCPUUtilization"
      },
      ...
    }
  ]
}
```

   Pour plus d'informations, consultez PredictiveScalingPredefinedLoadMetricle manuel Amazon EC2 Auto Scaling API Reference.

   With custom metrics

```
{
  "MetricSpecifications":[
    {
      "CustomizedLoadMetricSpecification":{
```

```
        "MetricDataQueries":[
          {
            "Id":"load_metric",
            "MetricStat":{
              "Metric":{
                "MetricName":"MyLoadMetric",
                "Namespace":"MyNameSpace",
                "Dimensions":[
                  {
                    "Name":"MyOptionalMetricDimensionName",
                    "Value":"MyOptionalMetricDimensionValue"
                  }
                ]
              },
              "Stat":"Sum"
            }
          }
        ]
      },
      ...
    }
  ]
}
```

Pour plus d'informations, consultez PredictiveScalingCustomizedLoadMetricle manuel Amazon EC2 Auto Scaling API Reference.

3. Ajoutez la spécification de la métrique de mise à l'échelle à `MetricSpecifications` et définissez une valeur cible.

   Vous trouverez ci-dessous des exemples de structure des sections relatives à la métrique de mise à l'échelle et à la valeur cible.

   With predefined metrics

```
{
  "MetricSpecifications":[
    {
      "PredefinedLoadMetricSpecification":{
        "PredefinedMetricType":"ASGTotalCPUUtilization"
      },
      "PredefinedScalingMetricSpecification":{
        "PredefinedMetricType":"ASGCPUUtilization"
```

```
      },
      "TargetValue":50
    }
  ],
  ...
}
```

Pour plus d'informations, consultez PredictiveScalingPredefinedScalingMetricle manuel Amazon EC2 Auto Scaling API Reference.

With custom metrics

```
{
  "MetricSpecifications":[
    {
      "CustomizedLoadMetricSpecification":{
        "MetricDataQueries":[
          {
            "Id":"load_metric",
            "MetricStat":{
              "Metric":{
                "MetricName":"MyLoadMetric",
                "Namespace":"MyNameSpace",
                "Dimensions":[
                  {
                    "Name":"MyOptionalMetricDimensionName",
                    "Value":"MyOptionalMetricDimensionValue"
                  }
                ]
              },
              "Stat":"Sum"
            }
          }
        ]
      },
      "CustomizedScalingMetricSpecification":{
        "MetricDataQueries":[
          {
            "Id":"scaling_metric",
            "MetricStat":{
              "Metric":{
                "MetricName":"MyUtilizationMetric",
                "Namespace":"MyNameSpace",
```

```
                "Dimensions":[
                  {
                    "Name":"MyOptionalMetricDimensionName",
                    "Value":"MyOptionalMetricDimensionValue"
                  }
                ]
              },
              "Stat":"Average"
            }
          }
        ]
      },
      "TargetValue":50
    }
  ],
  ...
}
```

Pour plus d'informations, consultez PredictiveScalingCustomizedScalingMetricle manuel Amazon EC2 Auto Scaling API Reference.

4. Pour les prévisions uniquement, ajoutez la propriété `Mode` avec une valeur de`ForecastOnly`. Une fois que vous avez terminé la migration de la mise à l'échelle prédictive et que vous vous êtes assuré que les prévisions sont précises et fiables, vous pouvez modifier le mode pour autoriser la mise à l'échelle. Pour de plus amples informations, veuillez consulter Étape 7 : réactiver le dimensionnement prédictif.

```
{
  "MetricSpecifications":[
  ...
  ],
  "Mode":"ForecastOnly",
  ...
}
```

Pour plus d'informations, consultez PredictiveScalingConfigurationle manuel Amazon EC2 Auto Scaling API Reference.

5. Si la **`ScheduledActionBufferTime`** propriété est présente dans votre plan de dimensionnement, copiez sa valeur dans la `SchedulingBufferTime` propriété dans votre politique de dimensionnement prédictif.

```
{
  "MetricSpecifications":[
  ...
  ],
  "Mode":"ForecastOnly",
  "SchedulingBufferTime":300,
  ...
}
```

Pour plus d'informations, consultez PredictiveScalingConfigurationle manuel Amazon EC2 Auto Scaling API Reference.

6. Si les **`PredictiveScalingMaxCapacityBuffer`** propriétés **`PredictiveScalingMaxCapacityBehavior`** et sont présentes dans votre plan de dimensionnement, vous pouvez configurer les `MaxCapacityBuffer` propriétés `MaxCapacityBreachBehavior` et dans votre politique de dimensionnement prédictif. Ces propriétés définissent ce qui doit se passer si la capacité prévisionnelle approche ou dépasse la capacité maximale spécifiée pour le groupe Auto Scaling.

> **Warning**
>
> Si vous définissez cette `MaxCapacityBreachBehavior` propriété sur`IncreaseMaxCapacity`, davantage d'instances pourraient être lancées que prévu, à moins que vous ne surveilliez et ne gériez l'augmentation de la capacité maximale. La capacité maximale accrue devient la nouvelle capacité maximale normale pour le groupe Auto Scaling jusqu'à ce que vous la mettiez à jour manuellement. La capacité maximale ne revient pas automatiquement à la capacité maximale initiale.

```
{
  "MetricSpecifications":[
  ...
  ],
  "Mode":"ForecastOnly",
  "SchedulingBufferTime":300,
  "MaxCapacityBreachBehavior": "IncreaseMaxCapacity",
  "MaxCapacityBuffer": 10
}
```

Pour plus d'informations, consultez PredictiveScalingConfigurationle manuel Amazon EC2 Auto Scaling API Reference.

7. Enregistrez le fichier JSON sous un nom unique. Notez le nom du fichier. Vous en aurez besoin à l'étape suivante et de nouveau à la fin de la procédure de migration lorsque vous réactivez vos politiques de dimensionnement prédictif. Pour de plus amples informations, veuillez consulter Étape 7 : réactiver le dimensionnement prédictif.

8. Après avoir enregistré votre fichier JSON, exécutez la put-scaling-policycommande. Dans l'exemple suivant, remplacez chacune *user input placeholder* par vos propres informations.

```
aws autoscaling put-scaling-policy --policy-name my-predictive-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \
  --predictive-scaling-configuration file://my-predictive-scaling-config.json
```

Si elle aboutit, cette commande renvoie l'Amazon Resource Name (ARN) de la stratégie.

```
{
  "PolicyARN": "arn:aws:autoscaling:region:account-id:scalingPolicy:2f4f5048-
d8a8-4d14-b13a-d1905620f345:autoScalingGroupName/my-asg:policyName/my-predictive-
scaling-policy",
  "Alarms": []
}
```

9. Répétez ces étapes pour chaque spécification de métrique de charge que vous migrez vers une politique de dimensionnement prédictif basée sur Amazon EC2 Auto Scaling.

Étape 3 : Passez en revue les prévisions générées par les politiques de dimensionnement prédictif

Si vous n'utilisez pas le dimensionnement prédictif, ignorez la procédure suivante.

Une prévision est disponible peu après la création d'une politique de dimensionnement prédictif. Une fois qu'Amazon EC2 Auto Scaling a généré les prévisions, vous pouvez consulter les prévisions relatives à la politique via la console Amazon EC2 Auto Scaling et les ajuster si nécessaire.

Pour consulter les prévisions relatives à une politique de dimensionnement prédictive

1. Ouvrez la EC2 console Amazon à l'adresse https://console.aws.amazon.com/ec2/.
2. Dans le volet de navigation, choisissez Auto Scaling Groups, puis choisissez le nom de votre groupe Auto Scaling dans la liste.
3. Dans l'onglet Mise à l'échelle automatique, dans Politiques de dimensionnement prédictif, choisissez votre stratégie.
4. Dans la section Surveillance, vous pouvez afficher les prévisions passées et futures de votre politique concernant la charge et la capacité par rapport aux valeurs réelles.

   Pour plus d'informations, consultez les graphiques de surveillance du dimensionnement prédictif dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling.
5. Répétez ces étapes pour chaque politique de dimensionnement prédictif que vous avez créée.

Étape 4 : préparer la suppression du plan de dimensionnement

Pour toutes les ressources dotées d'une configuration de dimensionnement de suivi des cibles existante, procédez comme suit pour collecter les informations supplémentaires dont vous avez besoin dans le plan de dimensionnement avant de le supprimer.

Pour décrire les informations relatives à la politique de dimensionnement issues du plan de dimensionnement, utilisez la describe-scaling-plan-resourcescommande. Dans l'exemple de commande suivant, remplacez-le *my-scaling-plan* par vos propres informations.

```
aws autoscaling-plans describe-scaling-plan-resources \
  --scaling-plan-name my-scaling-plan \
  --scaling-plan-version 1
```

Passez en revue le résultat et confirmez que vous souhaitez migrer les politiques de dimensionnement décrites. Utilisez ces informations pour créer de nouvelles politiques de dimensionnement du suivi des cibles basées sur Amazon EC2 Auto Scaling et Application Auto Scaling dans. Étape 6 : réactiver la mise à l'échelle dynamique

Étape 5 : Supprimer le plan de dimensionnement

Avant de créer de nouvelles politiques de dimensionnement pour le suivi des cibles, vous devez supprimer le plan de dimensionnement afin de supprimer les politiques de dimensionnement qu'il a créées.

Pour supprimer votre plan de dimensionnement, utilisez la delete-scaling-plancommande. Dans l'exemple de commande suivant, remplacez-le *my-scaling-plan* par vos propres informations.

```
aws autoscaling-plans delete-scaling-plan \
  --scaling-plan-name my-scaling-plan \
  --scaling-plan-version 1
```

Après avoir supprimé le plan de dimensionnement, le dimensionnement dynamique est désactivé. Ainsi, en cas d'augmentation soudaine du trafic ou de la charge de travail, la capacité disponible pour chaque ressource évolutive n'augmentera pas d'elle-même. Par mesure de précaution, vous souhaiterez peut-être augmenter manuellement la capacité de vos ressources évolutives à court terme.

Pour augmenter la capacité d'un groupe Auto Scaling

1. Ouvrez la EC2 console Amazon à l'adresse https://console.aws.amazon.com/ec2/.
2. Dans le volet de navigation, choisissez Auto Scaling Groups, puis choisissez le nom de votre groupe Auto Scaling dans la liste.
3. Sous l'onglet Details (Détails) choisissez Group details (Détails du groupe), Edit (Modifier).
4. Pour Capacité souhaitée, augmentez la capacité souhaitée.
5. Lorsque vous avez terminé, choisissez Mettre à jour.

Pour ajouter un réplica Aurora à un cluster de bases de données

1. Ouvrez la console Amazon RDS à l'adresse https://console.aws.amazon.com/rds/.
2. Dans le volet de navigation, choisissez Databases, puis sélectionnez votre cluster de bases de données.
3. Assurez-vous que le cluster et l'instance principale ont l'état Disponible.
4. Choisissez Actions, puis Ajouter un lecteur.
5. Sur la page Ajouter un lecteur, spécifiez les options pour votre nouvelle réplique d'Aurora.

6. Choisissez Ajouter un lecteur.

Pour augmenter la capacité de lecture et d'écriture allouée à une table DynamoDB ou à un index secondaire global

1. Ouvrez la console DynamoDB à l'adresse. https://console.aws.amazon.com/dynamodb/
2. Dans le volet de navigation, choisissez Tables, puis choisissez le nom de votre table dans la liste.
3. Dans l'onglet Paramètres supplémentaires, choisissez Capacité de lecture/écriture, Modifier.
4. Sur la page Modifier la capacité de lecture/écriture, pour Capacité de lecture, Unités de capacité allouées, augmentez la capacité de lecture allouée de la table.
5. (Facultatif) Si vous souhaitez que vos index secondaires globaux utilisent les mêmes paramètres de capacité de lecture que la table de base, cochez la case Utiliser les mêmes paramètres de capacité de lecture pour tous les index secondaires globaux.
6. Pour la capacité d'écriture, unités de capacité allouées, augmentez la capacité d'écriture allouée de la table.
7. (Facultatif) Si vous souhaitez que vos index secondaires globaux utilisent les mêmes paramètres de capacité d'écriture que la table de base, cochez la case Utiliser les mêmes paramètres de capacité d'écriture pour tous les index secondaires globaux.
8. Si vous n'avez pas coché les cases à cocher aux étapes 5 ou 7, faites défiler la page vers le bas pour mettre à jour la capacité de lecture et d'écriture de tous les index secondaires globaux.
9. Choisissez Enregistrer les modifications pour continuer.

Pour augmenter le nombre de tâches en cours d'exécution pour votre service Amazon ECS

1. Ouvrez la console à la https://console.aws.amazon.com/ecs/version 2.
2. Dans le volet de navigation, choisissez Clusters, puis choisissez le nom de votre cluster dans la liste.
3. Dans la section Services, cochez la case à côté du service, puis choisissez Mettre à jour.
4. Pour Tâches souhaitées, entrez le nombre de tâches que vous souhaitez exécuter pour le service.
5. Choisissez Mettre à jour.

Pour augmenter la capacité d'une flotte Spot

1. Ouvrez la EC2 console Amazon à l'adresse https://console.aws.amazon.com/ec2/.
2. Dans le volet de navigation, choisissez Spot Requests, puis sélectionnez votre demande Spot Fleet.
3. Choisissez Actions, Modify target capacity (Modifier la capacité cible).
4. Dans Modifier la capacité cible, entrez la nouvelle capacité cible et la partie Instance à la demande.
5. Sélectionnez Envoyer.

Étape 6 : réactiver la mise à l'échelle dynamique

Réactivez le dimensionnement dynamique en créant des politiques de dimensionnement pour le suivi des cibles.

Lorsque vous créez une politique de dimensionnement du suivi des cibles pour un groupe Auto Scaling, vous l'ajoutez directement au groupe. Lorsque vous créez une politique de dimensionnement de suivi des cibles pour d'autres ressources évolutives, vous enregistrez d'abord la ressource en tant que cible évolutive, puis vous ajoutez une politique de dimensionnement de suivi des cibles à la cible évolutive.

Rubriques

- Créez des politiques de dimensionnement pour le suivi des cibles pour les groupes Auto Scaling
- Créez des politiques de dimensionnement pour le suivi des cibles pour d'autres ressources évolutives

Créez des politiques de dimensionnement pour le suivi des cibles pour les groupes Auto Scaling

Pour créer des politiques de dimensionnement pour le suivi des cibles pour les groupes Auto Scaling

1. Dans un fichier JSON, créez un `PredefinedMetricSpecification` ou `CustomizedMetricSpecification` en utilisant les paramètres équivalents du plan de dimensionnement.

Voici des exemples de configuration de suivi des cibles. Dans ces exemples, remplacez chacun *user input placeholder* par vos propres informations.

With predefined metrics

```
{
  "TargetValue": 50.0,
  "PredefinedMetricSpecification":
    {
      "PredefinedMetricType": "ASGAverageCPUUtilization"
    }
}
```

Pour plus d'informations, consultez PredefinedMetricSpecificationle manuel Amazon EC2 Auto Scaling API Reference.

With custom metrics

```
{
  "TargetValue": 100.0,
  "CustomizedMetricSpecification": {
    "MetricName": "MyBacklogPerInstance",
    "Namespace": "MyNamespace",
    "Dimensions": [{
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }],
    "Statistic": "Average",
    "Unit": "None"
  }
}
```

Pour plus d'informations, consultez CustomizedMetricSpecificationle manuel Amazon EC2 Auto Scaling API Reference.

2. Pour créer votre politique de dimensionnement, utilisez la put-scaling-policycommande, ainsi que le fichier JSON que vous avez créé à l'étape précédente. Dans l'exemple suivant, remplacez chacune *user input placeholder* par vos propres informations.

```
aws autoscaling put-scaling-policy --policy-name my-target-tracking-scaling-policy
 \
```

```
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \
  --target-tracking-configuration file://config.json
```

3. Répétez ce processus pour chaque politique de dimensionnement basée sur un plan de dimensionnement que vous migrez vers une politique de dimensionnement de suivi des cibles basée sur Amazon EC2 Auto Scaling.

Créez des politiques de dimensionnement pour le suivi des cibles pour d'autres ressources évolutives

Créez ensuite des politiques de dimensionnement du suivi des cibles pour d'autres ressources évolutives en effectuant les tâches de configuration suivantes.

- Enregistrez une cible évolutive pour le dimensionnement automatique avec le service Application Auto Scaling.
- Ajoutez une politique de suivi des cibles et de mise à l'échelle sur la cible évolutive.

Pour créer des politiques de dimensionnement pour le suivi des cibles pour d'autres ressources évolutives

1. Utilisez la register-scalable-targetcommande pour enregistrer la ressource en tant que cible évolutive et définir les limites de dimensionnement pour la politique de dimensionnement.

   Dans l'exemple suivant, remplacez chacune *`user input placeholder`* par vos propres informations. Pour les options de commande, fournissez les informations suivantes :

   - `--service-namespace`— Un espace de noms pour le service cible (par exemple,**`ecs`**). Pour obtenir les espaces de noms des services, consultez la RegisterScalableTargetréférence.
   - `--scalable-dimension`— Dimension évolutive associée à la ressource cible (par exemple,**`ecs:service:DesiredCount`**). Pour obtenir des dimensions évolutives, consultez la RegisterScalableTargetréférence.
   - `--resource-id`— Un identifiant de ressource pour la ressource cible (par exemple,**`service/`*`my-cluster`*`/`*`my-service`***). Pour plus d'informations sur la syntaxe et des exemples de ressources spécifiques IDs, consultez la RegisterScalableTargetréférence.

```
aws application-autoscaling register-scalable-target --service-namespace namespace
 \
```

```
  --scalable-dimension dimension \
  --resource-id identifier \
  --min-capacity 1 --max-capacity 10
```

En cas de réussite, cette commande renvoie l'ARN de la cible évolutive.

```
{
    "ScalableTargetARN": "arn:aws:application-autoscaling:region:account-
id:scalable-target/1234abcd56ab78cd901ef1234567890ab123"
}
```

2. Dans un fichier JSON, créez un `PredefinedMetricSpecification` ou `CustomizedMetricSpecification` en utilisant les paramètres équivalents du plan de dimensionnement.

   Voici des exemples de configuration de suivi des cibles.

   With predefined metrics

   ```
   {
     "TargetValue": 70.0,
     "PredefinedMetricSpecification":
       {
         "PredefinedMetricType": "ECSServiceAverageCPUUtilization"
       }
   }
   ```

   Pour plus d'informations, reportez-vous PredefinedMetricSpecificationà la section Application Auto Scaling API Reference.

   With custom metrics

   ```
   {
     "TargetValue": 70.0,
     "CustomizedMetricSpecification": {
       "MetricName": "MyUtilizationMetric",
       "Namespace": "MyNamespace",
       "Dimensions": [{
         "Name": "MyOptionalMetricDimensionName",
         "Value": "MyOptionalMetricDimensionValue"
       }],
       "Statistic": "Average",
   ```

```
    "Unit": "Percent"
  }
}
```

Pour plus d'informations, reportez-vous CustomizedMetricSpecificationà la section Application Auto Scaling API Reference.

3. Pour créer votre politique de dimensionnement, utilisez la put-scaling-policycommande, ainsi que le fichier JSON que vous avez créé à l'étape précédente.

```
aws application-autoscaling put-scaling-policy --service-namespace namespace \
  --scalable-dimension dimension \
  --resource-id identifier \
  --policy-name my-target-tracking-scaling-policy --policy-
type TargetTrackingScaling \
  --target-tracking-scaling-policy-configuration file://config.json
```

4. Répétez ce processus pour chaque politique de dimensionnement basée sur un plan de dimensionnement que vous migrez vers une politique de dimensionnement de suivi des cibles basée sur Application Auto Scaling.

Étape 7 : réactiver le dimensionnement prédictif

Si vous n'utilisez pas le dimensionnement prédictif, ignorez cette étape.

Réactivez la mise à l'échelle prédictive en passant à la mise à l'échelle prédictive aux prévisions et à l'échelle.

Pour effectuer cette modification, mettez à jour les fichiers JSON que vous avez créés Étape 2 : créer des politiques de dimensionnement prédictives et modifiez la valeur de l'`Mode`option `ForecastAndScale` comme dans l'exemple suivant :

```
"Mode":"ForecastAndScale"
```

Mettez ensuite à jour chaque politique de dimensionnement prédictif à l'aide de la put-scaling-policycommande. Dans cet exemple, remplacez chacune *`user input placeholder`* par vos propres informations.

```
aws autoscaling put-scaling-policy --policy-name my-predictive-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \
```

```
  --predictive-scaling-configuration file://my-predictive-scaling-config.json
```

Vous pouvez également effectuer cette modification depuis la console Amazon EC2 Auto Scaling en activant le paramètre Scale basé sur les prévisions. Pour plus d'informations, consultez Predictive Scaling for Amazon EC2 Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling.

Référence Amazon EC2 Auto Scaling pour la migration des politiques de dimensionnement du suivi des cibles

À des fins de référence, le tableau suivant répertorie toutes les propriétés de configuration de suivi des cibles dans le plan de dimensionnement avec leur propriété correspondante dans le fonctionnement de l'`PutScalingPolicy`API Amazon EC2 Auto Scaling.

Propriété source du plan de dimensionnement	Propriété cible EC2 d'Amazon Auto Scaling
`PolicyName`	`PolicyName`
`PolicyType`	`PolicyType`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Dimensions.Name`	`TargetTrackingConfiguration.CustomizedMetricSpecification.Dimensions.Name`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Dimensions.Value`	`TargetTrackingConfiguration.CustomizedMetricSpecification.Dimensions.Value`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.MetricName`	`TargetTrackingConfiguration.CustomizedMetricSpecification.MetricName`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Namespace`	`TargetTrackingConfiguration.CustomizedMetricSpecification.Namespace`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Statistic`	`TargetTrackingConfiguration.CustomizedMetricSpecification.Statistic`

Propriété source du plan de dimensionnement	Propriété cible EC2 d'Amazon Auto Scaling
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Unit`	`TargetTrackingConfiguration.CustomizedMetricSpecification.Unit`
`TargetTrackingConfiguration.DisableScaleIn`	`TargetTrackingConfiguration.DisableScaleIn`
`TargetTrackingConfiguration.EstimatedInstanceWarmup`	`TargetTrackingConfiguration.EstimatedInstanceWarmup` [1]
`TargetTrackingConfiguration.PredefinedScalingMetricSpecification.PredefinedScalingMetricType`	`TargetTrackingConfiguration.PredefinedMetricSpecification.PredefinedMetricType`
`TargetTrackingConfiguration.PredefinedScalingMetricSpecification.ResourceLabel`	`TargetTrackingConfiguration.PredefinedMetricSpecification.ResourceLabel`
`TargetTrackingConfiguration.ScaleInCooldown`	`Not available`
`TargetTrackingConfiguration.ScaleOutCooldown`	`Not available`
`TargetTrackingConfiguration.TargetValue`	`TargetTrackingConfiguration.TargetValue`

[1] Le préchauffage des instances est une fonctionnalité destinée aux groupes Auto Scaling qui permet de s'assurer que les instances nouvellement lancées sont prêtes à recevoir du trafic avant de contribuer leurs données d'utilisation à la métrique de dimensionnement. Alors que les instances sont encore en train de s'échauffer, Amazon EC2 Auto Scaling ralentit le processus d'ajout ou de suppression d'instances au groupe. Au lieu de spécifier un temps de préchauffage pour une politique de dimensionnement, nous vous recommandons d'utiliser le paramètre de préchauffage d'instance par défaut de votre groupe Auto Scaling afin de garantir que tous les lancements d'instances utilisent le même temps de préchauffage d'instance. Pour plus d'informations, consultez la section Définir le

préchauffage d'instance par défaut pour un groupe Auto Scaling dans le manuel Amazon EC2 Auto Scaling User Guide.

Référence Application Auto Scaling pour la migration des politiques de dimensionnement du suivi des cibles

À des fins de référence, le tableau suivant répertorie toutes les propriétés de configuration de suivi des cibles dans le plan de dimensionnement avec leur propriété correspondante dans le fonctionnement de l'`PutScalingPolicy`API Application Auto Scaling.

Propriété source du plan de dimensionnement	Propriété cible Application Auto Scaling
`PolicyName`	`PolicyName`
`PolicyType`	`PolicyType`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Dimensions.Name`	`TargetTrackingScalingPolicyConfiguration.CustomizedMetricSpecification.Dimensions.Name`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Dimensions.Value`	`TargetTrackingScalingPolicyConfiguration.CustomizedMetricSpecification.Dimensions.Value`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.MetricName`	`TargetTrackingScalingPolicyConfiguration.CustomizedMetricSpecification.MetricName`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Namespace`	`TargetTrackingScalingPolicyConfiguration.CustomizedMetricSpecification.Namespace`
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Statistic`	`TargetTrackingScalingPolicyConfiguration.CustomizedMetricSpecification.Statistic`

Propriété source du plan de dimensionnement	Propriété cible Application Auto Scaling
`TargetTrackingConfiguration.CustomizedScalingMetricSpecification.Unit`	`TargetTrackingScalingPolicyConfiguration.CustomizedMetricSpecification.Unit`
`TargetTrackingConfiguration.DisableScaleIn`	`TargetTrackingScalingPolicyConfiguration.DisableScaleIn`
`TargetTrackingConfiguration.EstimatedInstanceWarmup`	`Not available`
`TargetTrackingConfiguration.PredefinedScalingMetricSpecification.PredefinedScalingMetricType`	`TargetTrackingScalingPolicyConfiguration.PredefinedMetricSpecification.PredefinedMetricType`
`TargetTrackingConfiguration.PredefinedScalingMetricSpecification.ResourceLabel`	`TargetTrackingScalingPolicyConfiguration.PredefinedMetricSpecification.ResourceLabel`
`TargetTrackingConfiguration.ScaleInCooldown` [1]	`TargetTrackingScalingPolicyConfiguration.ScaleInCooldown`
`TargetTrackingConfiguration.ScaleOutCooldown` [1]	`TargetTrackingScalingPolicyConfiguration.ScaleOutCooldown`
`TargetTrackingConfiguration.TargetValue`	`TargetTrackingScalingPolicyConfiguration.TargetValue`

[1] Application Auto Scaling utilise des périodes de recharge pour ralentir le dimensionnement lorsque votre ressource évolutive augmente (augmente la capacité) ou augmente (réduit la capacité). Pour plus d'informations, consultez la section Définir les périodes de refroidissement dans le Guide de l'utilisateur d'Application Auto Scaling.

Informations supplémentaires

Pour savoir comment créer de nouvelles politiques de dimensionnement prédictif à partir de la console, consultez la rubrique suivante :

- Amazon EC2 Auto Scaling — Créez une politique de dimensionnement prédictif dans le guide de l'utilisateur Amazon EC2 Auto Scaling.

Pour savoir comment créer de nouvelles politiques de dimensionnement du suivi des cibles à l'aide de la console, consultez les rubriques suivantes :

- Amazon Aurora — Utilisation d'Amazon Aurora Auto Scaling avec des répliques Aurora dans le guide de l'utilisateur Amazon RDS.
- DynamoDB — Utilisation de la mise à l'échelle automatique avec AWS Management Console DynamoDB décrite dans le manuel du développeur Amazon DynamoDB.
- Amazon EC2 Auto Scaling — Créez une politique de dimensionnement du suivi des cibles dans le guide de l'utilisateur Amazon EC2 Auto Scaling.
- Amazon ECS — Mise à jour d'un service à l'aide de la console décrite dans le manuel Amazon Elastic Container Service Developer Guide.
- Spot Fleet — Faites évoluer Spot Fleet à l'aide d'une politique de suivi ciblée décrite dans le guide de EC2 l'utilisateur Amazon.

Sécurité du plan de dimensionnement

La sécurité du cloud AWS est la priorité absolue. En tant que AWS client, vous bénéficiez d'un centre de données et d'une architecture réseau conçus pour répondre aux exigences des entreprises les plus sensibles en matière de sécurité.

La sécurité est une responsabilité partagée entre vous AWS et vous. Le modèle de responsabilité partagée décrit cela comme la sécurité du cloud et la sécurité dans le cloud :

- Sécurité du cloud : AWS est chargée de protéger l'infrastructure qui exécute les AWS services dans le AWS cloud. AWS vous fournit également des services que vous pouvez utiliser en toute sécurité. Des auditeurs tiers testent et vérifient régulièrement l'efficacité de notre sécurité dans le AWS cadre des programmes de de). Pour en savoir plus sur les programmes de conformité qui s'appliquent à AWS Auto Scaling, consultez la section AWS services concernés par programme de conformité et .
- Sécurité dans le cloud — Votre responsabilité est déterminée par le AWS service que vous utilisez. Vous êtes également responsable d'autres facteurs, y compris la sensibilité de vos données, les exigences de votre entreprise, ainsi que la législation et la réglementation applicables.

Cette documentation vous aide à comprendre comment appliquer le modèle de responsabilité partagée lors de l'utilisation de plans de dimensionnement. Vous aide également à comprendre comment gérer l'accès à des plans de dimensionnement.

Rubriques

- Accédez aux plans de dimensionnement à l'aide des points de terminaison VPC de l'interface
- Protection des données pour les plans de mise à l'échelle
- Gestion des identités et des accès pour les plans de mise à l'échelle
- Validation de conformité pour les plans de dimensionnement
- Sécurité de l'infrastructure pour les plans d'évolutivité

Accédez aux plans de dimensionnement à l'aide des points de terminaison VPC de l'interface

Vous pouvez l'utiliser AWS PrivateLink pour créer une connexion privée entre votre VPC et. AWS Auto Scaling Vous pouvez y accéder AWS Auto Scaling comme s'il se trouvait dans votre VPC, sans

utiliser de passerelle Internet, de périphérique NAT, de connexion VPN ou AWS Direct Connect de connexion. Les instances de votre VPC n'ont pas besoin d'adresses IP publiques pour y accéder. AWS Auto Scaling

Vous établissez cette connexion privée en créant un point de terminaison d'interface optimisé par AWS PrivateLink. Nous créons une interface réseau de point de terminaison dans chaque sous-réseau que vous activez pour le point de terminaison d'interface. Il s'agit d'interfaces réseau gérées par le demandeur qui servent de point d'entrée pour le trafic destiné à AWS Auto Scaling.

Pour plus d'informations, consultez la section Accès Services AWS par AWS PrivateLink le biais du AWS PrivateLink guide.

Rubriques

- Créer un point de terminaison de VPC d'interface pour les plans de dimensionnement
- Création d'une stratégie de point de terminaison de VPC pour le dimensionnement des plans
- Migration des points de terminaison

Créer un point de terminaison de VPC d'interface pour les plans de dimensionnement

Créez un point de terminaison pour AWS Auto Scaling dimensionner les plans en utilisant le nom de service suivant :

```
com.amazonaws.region.autoscaling-plans
```

Pour plus d'informations, consultez la section Accès à un AWS service à l'aide d'un point de terminaison VPC d'interface dans le AWS PrivateLink Guide.

Il n'est pas nécessaire de modifier les autres paramètres. AWS Auto Scaling L'API appelle d'autres utilisateurs Services AWS à l'aide de points de terminaison de service ou de points de terminaison VPC d'interface privée, selon ceux utilisés.

Création d'une stratégie de point de terminaison de VPC pour le dimensionnement des plans

Vous pouvez associer une politique à votre point de terminaison VPC pour contrôler l'accès à l' AWS Auto Scaling API. La politique spécifie :

- Le principal qui peut exécuter des actions.
- Les actions qui peuvent être effectuées.
- La ressource sur laquelle les actions peuvent être effectuées.

L'exemple suivant montre une stratégie de point de terminaison VPC qui refuse à tout le monde l'autorisation de supprimer une stratégie de dimensionnement via le point de terminaison. L'exemple de politique accorde également à tout le monde l'autorisation d'effectuer toutes les autres actions.

```
{
   "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
            "Action": "autoscaling-plans:DeleteScalingPlan",
            "Effect": "Deny",
            "Resource": "*",
            "Principal": "*"
        }
    ]
}
```

Pour plus d'informations, consultez Stratégies de point de terminaison d'un VPC dans le AWS PrivateLink Guide.

Migration des points de terminaison

Le 22 novembre 2019, nous avons introduit `autoscaling-plans.region.amazonaws.com` le nouveau nom d'hôte et point de terminaison DNS par défaut pour les appels à l' AWS Auto Scaling API. Le nouveau point de terminaison est compatible avec la dernière version du AWS CLI et SDKs. Si ce n'est pas déjà fait, installez la dernière version AWS CLI et utilisez SDKs le nouveau point de terminaison. Pour le mettre à jour AWS CLI, reportez-vous à la section Installation ou mise à jour du AWS CLI Guide de AWS Command Line Interface l'utilisateur. Pour plus d'informations à ce sujet AWS SDKs, consultez la section Outils pour Amazon Web Services.

> **Important**
>
> Pour des raisons de rétrocompatibilité, le point de `autoscaling.region.amazonaws.com` terminaison existant continuera d'être pris en charge pour les appels à l' AWS Auto Scaling API. Pour configurer le `autoscaling.region.amazonaws.com` point de terminaison en tant que point de terminaison VPC à interface privée, consultez Amazon EC2 Auto Scaling et interfacez les points de terminaison VPC dans le guide de l'utilisateur Amazon EC2 Auto Scaling.

Point de terminaison à appeler lors de l'utilisation de la CLI ou de l' AWS Auto Scaling API

Dans la version actuelle de AWS Auto Scaling, vos appels à l' AWS Auto Scaling API sont automatiquement dirigés vers le `autoscaling-plans.region.amazonaws.com` point de terminaison au lieu de`autoscaling.region.amazonaws.com`.

Vous pouvez appeler le nouveau point de terminaison dans l'interface de ligne de commande en utilisant le paramètre suivant avec chaque commande pour préciser le point de terminaison : `--endpoint-url https://autoscaling-plans.region.amazonaws.com`.

Bien que ce ne soit pas conseillé, vous pouvez également appeler l'ancien point de terminaison dans l'interface de ligne de commande en utilisant le paramètre suivant avec chaque commande pour préciser le point de terminaison : `--endpoint-url https://autoscaling.region.amazonaws.com`.

Pour les différents types SDKs utilisés pour appeler le APIs, consultez la documentation du SDK qui vous intéresse pour savoir comment diriger les demandes vers un point de terminaison spécifique. Pour plus d'informations, consultez Outils pour Amazon Web Services.

Protection des données pour les plans de mise à l'échelle

Le modèle de responsabilité AWS partagée de s'applique à la protection des données dans AWS Auto Scaling. Comme décrit dans ce modèle, AWS est chargé de protéger l'infrastructure mondiale qui gère tous les AWS Cloud. La gestion du contrôle de votre contenu hébergé sur cette infrastructure relève de votre responsabilité. Vous êtes également responsable des tâches de configuration et de gestion de la sécurité des Services AWS que vous utilisez. Pour plus d'informations sur la confidentialité des données, consultez Questions fréquentes (FAQ) sur la confidentialité des données. Pour en savoir plus sur la protection des données en Europe, consultez

le billet de blog Modèle de responsabilité partagée AWS et RGPD (Règlement général sur la protection des données) sur le Blog de sécuritéAWS .

À des fins de protection des données, nous vous recommandons de protéger les Compte AWS informations d'identification et de configurer les utilisateurs individuels avec AWS IAM Identity Center ou AWS Identity and Access Management (IAM). Ainsi, chaque utilisateur se voit attribuer uniquement les autorisations nécessaires pour exécuter ses tâches. Nous vous recommandons également de sécuriser vos données comme indiqué ci-dessous :

- Utilisez l'authentification multifactorielle (MFA) avec chaque compte.
- SSL/TLS À utiliser pour communiquer avec AWS les ressources. Nous exigeons TLS 1.2 et recommandons TLS 1.3.
- Configurez l'API et la journalisation de l'activité des utilisateurs avec AWS CloudTrail. Pour plus d'informations sur l'utilisation des CloudTrail sentiers pour capturer AWS des activités, consultez la section Utilisation des CloudTrail sentiers dans le guide de AWS CloudTrail l'utilisateur.
- Utilisez des solutions de AWS chiffrement, ainsi que tous les contrôles de sécurité par défaut qu'ils contiennent Services AWS.
- Utilisez des services de sécurité gérés avancés tels qu'Amazon Macie, qui contribuent à la découverte et à la sécurisation des données sensibles stockées dans Amazon S3.
- Si vous avez besoin de modules cryptographiques validés par la norme FIPS 140-3 pour accéder AWS via une interface de ligne de commande ou une API, utilisez un point de terminaison FIPS. Pour plus d'informations sur les points de terminaison FIPS disponibles, consultez Norme FIPS (Federal Information Processing Standard) 140-3.

Nous vous recommandons fortement de ne jamais placer d'informations confidentielles ou sensibles, telles que les adresses e-mail de vos clients, dans des balises ou des champs de texte libre tels que le champ Nom. Cela inclut lorsque vous travaillez avec AWS Auto Scaling ou d'autres Services AWS utilisateurs de la console, de l'API ou AWS SDKs. AWS CLI Toutes les données que vous entrez dans des balises ou des champs de texte de forme libre utilisés pour les noms peuvent être utilisées à des fins de facturation ou dans les journaux de diagnostic. Si vous fournissez une adresse URL à un serveur externe, nous vous recommandons fortement de ne pas inclure d'informations d'identification dans l'adresse URL permettant de valider votre demande adressée à ce serveur.

Gestion des identités et des accès pour les plans de mise à l'échelle

AWS Identity and Access Management (IAM) est un outil Service AWS qui permet à un administrateur de contrôler en toute sécurité l'accès aux AWS ressources. Les administrateurs IAM contrôlent qui peut être authentifié (connecté) et autorisé (autorisé) à utiliser AWS Auto Scaling les ressources. IAM est un Service AWS outil que vous pouvez utiliser sans frais supplémentaires.

Pour une documentation IAM complète, consultez le Guide de l'utilisateur IAM.

Contrôle d’accès

Vous pouvez avoir des informations d'identification valides pour authentifier vos demandes, mais à moins d'avoir les autorisations requises, vous ne pouvez pas créer des plans de dimensionnement ni à y accéder. Par exemple, vous devez disposer des autorisations nécessaires pour créer des plans de dimensionnement, configurer le dimensionnement prédictif, etc.

Les sections suivantes fournissent des détails sur la façon dont un administrateur IAM peut utiliser IAM pour contribuer à sécuriser vos plans de dimensionnement en contrôlant qui peut travailler avec les plans de dimensionnement.

Rubriques

- Fonctionnement des plans de dimensionnement avec IAM
- Rôle lié à un service de dimensionnement prédictif
- Exemples de stratégies basées sur l'identité pour les plans de dimensionnement

Fonctionnement des plans de dimensionnement avec IAM

Avant d'utiliser IAM pour déterminer qui peut créer, accéder et gérer des plans de AWS Auto Scaling dimensionnement, vous devez comprendre quelles fonctionnalités IAM peuvent être utilisées avec les plans de dimensionnement.

Rubriques

- Politiques basées sur l’identité
- Politiques basées sur les ressources
- Listes de contrôle d'accès (ACLs)
- Autorisation basée sur les balises
- Rôles IAM

Politiques basées sur l'identité

Avec les politiques IAM basées sur l'identité, vous pouvez spécifier des actions et ressources autorisées ou refusées et les conditions dans lesquelles les actions sont autorisées ou refusées. Plans de dimensionnement prend en charge des actions, des ressources et des clés de conditions spécifiques. Pour en savoir plus sur tous les éléments que vous utilisez dans une politique JSON, consultez Références des éléments de politique JSON IAM dans le Guide de l'utilisateur IAM.

Actions

Les administrateurs peuvent utiliser les politiques AWS JSON pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions sur quelles ressources et dans quelles conditions.

L'élément `Action` d'une politique JSON décrit les actions que vous pouvez utiliser pour autoriser ou refuser l'accès à une politique. Les actions de stratégie portent généralement le même nom que l'opération AWS d'API associée. Il existe quelques exceptions, telles que les actions avec autorisations uniquement qui n'ont pas d'opération API correspondante. Certaines opérations nécessitent également plusieurs actions dans une politique. Ces actions supplémentaires sont nommées actions dépendantes.

Intégration d'actions dans une politique afin d'accorder l'autorisation d'exécuter les opérations associées.

Les actions des plans de dimensionnement dans les instructions de stratégie IAM utilisent le préfixe suivant avant l'action : `autoscaling-plans:`. Les déclarations de politique doivent inclure un élément `Action` ou `NotAction`. Les plans de dimensionnement ont leurs propres ensembles d'actions qui décrivent les tâches que vous pouvez effectuer avec ce service.

Pour préciser plusieurs actions dans une seule déclaration, séparez-les par des virgules comme l'indique l'exemple suivant.

```
"Action": [
      "autoscaling-plans:DescribeScalingPlans",
      "autoscaling-plans:DescribeScalingPlanResources"
```

Vous pouvez aussi préciser plusieurs actions à l'aide de caractères génériques (*). Par exemple, pour spécifier toutes les actions qui commencent par le mot `Describe`, incluez l'action suivante.

```
"Action": "autoscaling-plans:Describe*"
```

Pour afficher une liste complète des actions de plan de dimensionnement qui peuvent être utilisées dans les déclarations de stratégie, consultezActions, ressources et clés de condition pour AWS Auto Scalingdans leRéférence de l'autorisation de service.

Ressources

L'élément `Resource` précise les objets auxquels l'action s'applique.

Les plans de dimensionnement ne comportent aucune ressource définie par le service pouvant être utilisée en tant qu'élément `Resource` d'une déclaration de stratégie IAM;. Par conséquent, il n'existe aucun Amazon Resource Names (ARNs) à utiliser dans une politique IAM. Pour contrôler l'accès aux plans de dimensionnement, utilisez toujours un astérisque (*) comme ressource lorsque vous écrivez une stratégie IAM.

Clés de condition

L'élément `Condition` (ou le bloc `Condition`) vous permet de préciser des conditions lorsqu'une instruction est appliquée. Par exemple, il est possible d'appliquer une stratégie après seulement une date spécifique. Pour exprimer des conditions, utilisez des clés de condition prédéfinies.

Les plans de dimensionnement ne fournissent pas de clés de condition spécifiques au service, mais prennent en charge l'utilisation de certaines clés de condition globales. Pour voir toutes les clés de condition AWS globales, voir les clés de contexte de condition AWS globales dans le guide de l'utilisateur IAM.

L'élément `Condition` est facultatif.

Exemples

Pour voir des exemples de politiques basées sur l'identité, veuillez consulter Exemples de stratégies basées sur l'identité pour les plans de dimensionnement.

Politiques basées sur les ressources

D'autres services Amazon Web Services, tels qu'Amazon Simple Storage Service, prennent en charge les politiques d'autorisation basées sur une ressource. Par exemple, vous pouvez attacher une politique d'autorisation à un compartiment S3 pour gérer les autorisations d'accès à ce compartiment.

Les plans de dimensionnement et ne prennent pas en charge les stratégies basées sur les ressources.

Listes de contrôle d'accès (ACLs)

Les plans de dimensionnement ne prennent pas en charge les listes de contrôle d'accès (ACLs).

Autorisation basée sur les balises

Les plans de dimensionnement ne peuvent pas être balisés. Ils n'ont également aucune ressource définie par le service qui peut être balisée. Par conséquent, il ne prend pas en charge le contrôle d'accès basé sur les balises.

Les plans de dimensionnement peuvent contenir des ressources balisables, telles que des groupes Auto Scaling, qui prennent en charge le contrôle d'accès en fonction des balises. Pour plus d'informations, consultez la documentation relative au Service AWS.

Rôles IAM

Un rôle IAM est une entité au sein de votre Compte AWS qui dispose d'autorisations spécifiques.

Utilisation d'informations d'identification temporaires

Vous pouvez utiliser des informations d'identification temporaires pour vous connecter à l'aide de la fédération pour endosser un rôle IAM ou bien un rôle entre comptes. Vous obtenez des informations d'identification de sécurité temporaires en appelant des opérations d' AWS STS API telles que AssumeRoleou GetFederationToken.

Les plans de dimensionnement prennent en charge l'utilisation d'informations d'identification temporaires.

Rôles liés à un service pour les plans de dimensionnement

AWS Auto Scaling utilise des rôles liés à un service pour obtenir les autorisations nécessaires pour appeler d'autres AWS services en votre nom. Un rôle lié à un service simplifie la configuration des plans de dimensionnement, car vous n'avez pas besoin d'ajouter manuellement les autorisations requises. Pour plus d’informations, consultez Utilisation des rôles liés à un service dans le Guide de l’utilisateur IAM.

AWS Auto Scaling utilise quelques types de rôles liés à un service pour appeler d'autres personnes en votre Services AWS nom lorsque vous travaillez avec un plan de dimensionnement :

- Rôle lié au service de dimensionnement prédictif : permet d'accéder AWS Auto Scaling aux données métriques historiques à partir de. CloudWatch Permet également de créer des actions

planifiées pour les groupes Auto Scaling en fonction d'une prévision de charge et d'une prédiction de capacité. Pour de plus amples informations, veuillez consulter Rôle lié à un service de dimensionnement prédictif.

- Rôle lié au service Amazon EC2 Auto Scaling : permet d'accéder AWS Auto Scaling aux politiques de dimensionnement du suivi des cibles et de les gérer pour les groupes Auto Scaling. Pour plus d'informations, consultez la section Rôles liés aux services pour Amazon EC2 Auto Scaling dans le guide de l'utilisateur d'Amazon EC2 Auto Scaling.
- Rôle lié au service Application Auto Scaling : permet d'accéder AWS Auto Scaling aux politiques de dimensionnement du suivi des cibles et de les gérer pour d'autres ressources évolutives. Il y a un rôle lié à un service pour chaque service. Pour plus d'informations, consultez Rôles liés aux services pour Application Auto Scaling dans le Guide de l'utilisateur Application Auto Scaling.

Vous pouvez utiliser la procédure suivante pour déterminer si votre compte possède déjà un rôle lié au service.

Pour déterminer si un rôle lié au service existe déjà

1. Ouvrez la console IAM à l'adresse https://console.aws.amazon.com/iam/.
2. Dans le panneau de navigation, choisissez Roles (Rôles).
3. Recherchez `AWSServiceRole` dans la liste pour trouver les rôles liés à un service qui existent dans votre compte. Recherchez le nom du rôle lié à un service que vous souhaitez vérifier.

Rôles de service

AWS Auto Scaling n'a aucun rôle de service pour les plans de dimensionnement.

Rôle lié à un service de dimensionnement prédictif

AWS Auto Scaling utilise des rôles liés à un service pour obtenir les autorisations nécessaires pour appeler d'autres personnes en votre AWS nom lorsque vous travaillez avec un plan de dimensionnement. Pour de plus amples informations, veuillez consulter Rôles liés à un service pour les plans de dimensionnement.

Les sections suivantes décrivent comment créer et gérer le rôle lié à un service pour le dimensionnement prédictif. Commencez par configurer des autorisations pour permettre à une entité IAM (comme un utilisateur, un groupe ou un rôle) de créer, modifier ou supprimer un rôle lié à un service.

Autorisations accordées par le rôle lié à un service

AWS Auto Scaling utilise le rôle lié au service nommé AWSServiceRoleForAutoScalingPlans_EC2AutoScaling pour appeler d'autres AWS services en votre nom lorsque vous activez le dimensionnement prédictif.

AWSServiceRoleForAutoScalingPlans_EC2AutoScalingfait confiance au `autoscaling-plans.amazonaws.com` service pour assumer le rôle.

Ce rôle lié à un service utilise la politique gérée. AWSAutoScalingPlansEC2AutoScalingPolicy Pour voir les autorisations de cette stratégie, consultez AWSAutoScalingPlansEC2AutoScalingPolicy dans le AWS Guide de référence des stratégies gérées par.

Création du rôle lié à un service (automatique)

Il n'est pas nécessaire de créer manuellement le AutoScaling rôle AWSServiceRoleForAutoScalingPlans_EC2. AWS crée ce rôle pour vous lorsque vous créez un plan de dimensionnement dans votre compte et que vous activez le dimensionnement prédictif.

AWS Pour créer un rôle lié à un service en votre nom, vous devez disposer des autorisations requises. Pour plus d’informations, consultez Autorisations de rôles liés à un service dans le Guide de l’utilisateur IAM.

Création du rôle lié à un service (manuelle)

Pour créer manuellement le rôle lié à un service, vous pouvez utiliser la console IAM, la CLI IAM ou l'API IAM. Pour plus d'informations, consultez la section Créer un rôle lié à un service dans le guide de l'utilisateur IAM.

Pour créer un rôle lié à un service (AWS CLI)

Utilisez la create-service-linked-rolecommande suivante pour créer le rôle lié au service.

```
aws iam create-service-linked-role --aws-service-name autoscaling-plans.amazonaws.com
```

Modifier le rôle lié à un service

Vous pouvez modifier la description de AWSServiceRoleForAutoScalingPlans_EC2 à l'AutoScalingaide d'IAM. Pour plus d'informations, voir Modifier la description d'un rôle lié à un service dans le Guide de l'utilisateur IAM.

Supprimer le rôle lié à un service

Si vous n'avez plus besoin d'utiliser de plans de dimensionnement, nous vous recommandons d'en supprimer AWSServiceRoleForAutoScalingPlans_EC2 AutoScaling.

Vous pouvez supprimer un rôle lié à un service uniquement après la suppression de tous les plans de dimensionnement dans votre Compte AWS dont le dimensionnement prédictif est activée. Ainsi, vous ne pouvez pas involontairement supprimer les autorisations d'accès aux plans de dimensionnement.

Vous pouvez utiliser la console IAM, la CLI IAM ou l'API IAM pour supprimer des rôles liés à un service. Pour plus d'informations, voir Supprimer un rôle lié à un service dans le Guide de l'utilisateur IAM.

Après avoir supprimé le rôle AWSServiceRoleForAutoScalingPlans_EC2AutoScaling lié à un service, AWS Auto Scaling créez-le à nouveau si vous créez un plan de dimensionnement avec le dimensionnement prédictif activé.

Régions prises en charge

AWS Auto Scaling prend en charge l'utilisation de rôles liés aux services dans tous les domaines Régions AWS où des plans de dimensionnement sont disponibles. Pour plus d'informations sur la disponibilité régionale des plans de dimensionnement, voir Points de terminaison et quotas AWS Auto Scaling dans Références générales AWS.

Exemples de stratégies basées sur l'identité pour les plans de dimensionnement

Par défaut, un nouvel utilisateur IAM ne dispose d'aucune autorisation. Un administrateur IAM doit créer et assigner des politiques IAM qui accordent à une identité IAM (utilisateur ou rôle, par exemple) l'autorisation de travailler avec les plans de mise à l'échelle.

Pour savoir comment créer une stratégie IAM à partir de ces exemples de documents de stratégie JSON, consultez Création de stratégies dans l'onglet JSON dans le Guide de l'utilisateur IAM.

Rubriques

- Bonnes pratiques en matière de politiques
- Autorisation des utilisateurs à créer des plans de dimensionnement
- Autorisation des utilisateurs à activer le dimensionnement prédictif
- Autorisations supplémentaires requises

- Autorisations requises pour créer un rôle lié à un service

Bonnes pratiques en matière de politiques

Les politiques basées sur l'identité déterminent si quelqu'un peut créer, accéder ou supprimer AWS Auto Scaling des ressources dans votre compte. Ces actions peuvent entraîner des frais pour votre Compte AWS. Lorsque vous créez ou modifiez des politiques basées sur l'identité, suivez ces instructions et recommandations :

- Commencez AWS par les politiques gérées et passez aux autorisations du moindre privilège : pour commencer à accorder des autorisations à vos utilisateurs et à vos charges de travail, utilisez les politiques AWS gérées qui accordent des autorisations pour de nombreux cas d'utilisation courants. Ils sont disponibles dans votre Compte AWS. Nous vous recommandons de réduire davantage les autorisations en définissant des politiques gérées par les AWS clients spécifiques à vos cas d'utilisation. Pour plus d'informations, consultez politiques gérées par AWS ou politiques gérées par AWS pour les activités professionnelles dans le Guide de l'utilisateur IAM.
- Accordez les autorisations de moindre privilège : lorsque vous définissez des autorisations avec des politiques IAM, accordez uniquement les autorisations nécessaires à l'exécution d'une seule tâche. Pour ce faire, vous définissez les actions qui peuvent être entreprises sur des ressources spécifiques dans des conditions spécifiques, également appelées autorisations de moindre privilège. Pour plus d'informations sur l'utilisation d'IAM pour appliquer des autorisations, consultez politiques et autorisations dans IAM dans le Guide de l'utilisateur IAM.
- Utilisez des conditions dans les politiques IAM pour restreindre davantage l'accès : vous pouvez ajouter une condition à vos politiques afin de limiter l'accès aux actions et aux ressources. Par exemple, vous pouvez écrire une condition de politique pour spécifier que toutes les demandes doivent être envoyées via SSL. Vous pouvez également utiliser des conditions pour accorder l'accès aux actions de service si elles sont utilisées par le biais d'un service spécifique Service AWS, tel que AWS CloudFormation. Pour plus d'informations, consultez Conditions pour éléments de politique JSON IAM dans le Guide de l'utilisateur IAM.
- Utilisez l'Analyseur d'accès IAM pour valider vos politiques IAM afin de garantir des autorisations sécurisées et fonctionnelles : l'Analyseur d'accès IAM valide les politiques nouvelles et existantes de manière à ce que les politiques IAM respectent le langage de politique IAM (JSON) et les bonnes pratiques IAM. IAM Access Analyzer fournit plus de 100 vérifications de politiques et des recommandations exploitables pour vous aider à créer des politiques sécurisées et fonctionnelles. Pour plus d'informations, consultez Validation de politiques avec IAM Access Analyzer dans le Guide de l'utilisateur IAM.

- Exiger l'authentification multifactorielle (MFA) : si vous avez un scénario qui nécessite des utilisateurs IAM ou un utilisateur root, activez l'authentification MFA pour une sécurité accrue. Compte AWS Pour exiger la MFA lorsque des opérations d'API sont appelées, ajoutez des conditions MFA à vos politiques. Pour plus d'informations, consultez Sécurisation de l'accès aux API avec MFA dans le Guide de l'utilisateur IAM.

Pour plus d'informations sur les bonnes pratiques dans IAM, consultez Bonnes pratiques de sécurité dans IAM dans le Guide de l'utilisateur IAM.

Autorisation des utilisateurs à créer des plans de dimensionnement

Voici un exemple de politique basée sur l'identité qui accorde des permissions de créer des plans de mise à l'échelle.

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
              "autoscaling-plans:*",
              "cloudwatch:PutMetricAlarm",
              "cloudwatch:DeleteAlarms",
              "cloudwatch:DescribeAlarms",
              "cloudformation:ListStackResources"
            ],
            "Resource": "*"
        }
    ]
}
```

Pour travailler avec un plan de dimensionnement, les utilisateurs finaux doivent disposer d'autorisations supplémentaires leur permettant de travailler avec des ressources spécifiques dans leur compte. Ces autorisations sont répertoriées dans Autorisations supplémentaires requises.

Chaque utilisateur de console a également besoin d'autorisations lui permettant de découvrir les ressources évolutives de son compte et de visualiser des graphiques de données CloudWatch

métriques à partir de la AWS Auto Scaling console. Les autorisations supplémentaires requises pour utiliser la AWS Auto Scaling console sont répertoriées ci-dessous :

- `cloudformation:ListStacks` : pour répertorier les piles.
- `tag:GetTagKeys`: pour rechercher des ressources évolutives qui contiennent certaines clés de balise.
- `tag:GetTagValues` : pour rechercher des ressources qui contiennent certaines valeurs de balise.
- `autoscaling:DescribeTags` : pour rechercher des groupes Auto Scaling qui contiennent certaines balises.
- `cloudwatch:GetMetricData` : pour afficher des données dans des graphiques de métriques.

Autorisation des utilisateurs à activer le dimensionnement prédictif

Voici un exemple de politique basée sur l'identité qui accorde des permissions d'activer la mise à l'échelle prédictive. Ces autorisations étendent les fonctionnalités des plans de dimensionnement configurés pour mettre à échelle les groupes Auto Scaling.

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
              "cloudwatch:GetMetricData",
              "autoscaling:DescribeAutoScalingGroups",
              "autoscaling:DescribeScheduledActions",
              "autoscaling:BatchPutScheduledUpdateGroupAction",
              "autoscaling:BatchDeleteScheduledAction"
            ],
            "Resource": "*"
        }
    ]
}
```

Autorisations supplémentaires requises

Pour configurer correctement les plans de dimensionnements, les utilisateurs finaux doivent disposer des autorisations pour chaque service cible pour lequel ils vont configurer la mise à l'échelle. Pour accorder les autorisations minimales requises pour utiliser les services cibles, lisez les informations dans cette section et spécifiez les actions pertinentes dans l'élément `Action` d'une déclaration de politique IAM.

Groupes Auto Scaling

Pour ajouter des groupes Auto Scaling à un plan de dimensionnement, les utilisateurs doivent disposer des autorisations suivantes de la part d'Amazon EC2 Auto Scaling :

- `autoscaling:UpdateAutoScalingGroup`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:PutScalingPolicy`
- `autoscaling:DescribePolicies`
- `autoscaling:DeletePolicy`

Services ECS

Pour ajouter des services ECS à un plan de dimensionnement, les utilisateurs doivent disposer des autorisations suivantes à partir d'Amazon ECS et Application Auto Scaling :

- `ecs:DescribeServices`
- `ecs:UpdateService`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling:DeleteScalingPolicy`

Parc d’instances Spot

Pour ajouter des flottes de spots à un plan de dimensionnement, les utilisateurs doivent disposer des autorisations suivantes auprès d'Amazon EC2 et d'Application Auto Scaling :

- `ec2:DescribeSpotFleetRequests`
- `ec2:ModifySpotFleetRequest`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling:DeleteScalingPolicy`

Tables et index globaux DynamoDB

Pour ajouter des tables ou des index globaux DynamoDB à un plan de dimensionnement, les utilisateurs doivent disposer des autorisations suivantes à partir de DynamoDB et Application Auto Scaling :

- `dynamodb:DescribeTable`
- `dynamodb:UpdateTable`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling:DeleteScalingPolicy`

Clusters DB Aurora

Pour ajouter des clusters de base de données Aurora à un plan de dimensionnement, les utilisateurs doivent disposer des autorisations suivantes à partir d'Amazon Aurora et Application Auto Scaling :

- `rds:AddTagsToResource`

- `rds:CreateDBInstance`
- `rds:DeleteDBInstance`
- `rds:DescribeDBClusters`
- `rds:DescribeDBInstances`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling:DeleteScalingPolicy`

Autorisations requises pour créer un rôle lié à un service

AWS Auto Scaling nécessite des autorisations pour créer un rôle lié à un service la première fois qu'un de vos utilisateurs Compte AWS crée un plan de dimensionnement avec le dimensionnement prédictif activé. Si le rôle lié au service n'existe pas déjà, AWS Auto Scaling créez-le dans votre compte. Le rôle lié au service accorde des autorisations lui AWS Auto Scaling permettant d'appeler d'autres services en votre nom.

Pour que cette création de rôle automatique aboutisse, les utilisateurs doivent disposer des autorisations nécessaires pour l'action `iam:CreateServiceLinkedRole`.

```
"Action": "iam:CreateServiceLinkedRole"
```

L'exemple suivant illustre une politique basée sur l’identité qui permet des autorisations pour créer un rôle lié à un service.

JSON

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
```

```
            "Resource": "arn:aws:iam::*:role/aws-service-role/autoscaling-
plans.amazonaws.com/AWSServiceRoleForAutoScalingPlans_EC2AutoScaling",
            "Condition": {
                "StringLike": {
                    "iam:AWSServiceName":"autoscaling-plans.amazonaws.com"
                }
            }
        }
    ]
}
```

Pour de plus amples informations, veuillez consulter Rôle lié à un service de dimensionnement prédictif.

Validation de conformité pour les plans de dimensionnement

Pour savoir si un programme Services AWS de conformité Service AWS s'inscrit dans le champ d'application de programmes de conformité spécifiques, consultez Services AWS la section de conformité et sélectionnez le programme de conformité qui vous intéresse. Pour des informations générales, voir Programmes de AWS conformité Programmes AWS de .

Vous pouvez télécharger des rapports d'audit tiers à l'aide de AWS Artifact. Pour plus d'informations, voir Téléchargement de rapports dans AWS Artifact .

Votre responsabilité en matière de conformité lors de l'utilisation Services AWS est déterminée par la sensibilité de vos données, les objectifs de conformité de votre entreprise et les lois et réglementations applicables. AWS fournit les ressources suivantes pour faciliter la mise en conformité :

- Conformité et gouvernance de la sécurité : ces guides de mise en œuvre de solutions traitent des considérations architecturales et fournissent les étapes à suivre afin de déployer des fonctionnalités de sécurité et de conformité.
- Référence des services éligibles HIPAA : liste les services éligibles HIPAA. Tous ne Services AWS sont pas éligibles à la loi HIPAA.
- AWS Ressources de https://aws.amazon.com/compliance/resources/ de conformité — Cette collection de classeurs et de guides peut s'appliquer à votre secteur d'activité et à votre région.
- AWS Guides de conformité destinés aux clients — Comprenez le modèle de responsabilité partagée sous l'angle de la conformité. Les guides résument les meilleures pratiques en matière

de sécurisation Services AWS et décrivent les directives relatives aux contrôles de sécurité dans plusieurs cadres (notamment le National Institute of Standards and Technology (NIST), le Payment Card Industry Security Standards Council (PCI) et l'Organisation internationale de normalisation (ISO)).

- Évaluation des ressources à l'aide des règles du guide du AWS Config développeur : le AWS Config service évalue dans quelle mesure les configurations de vos ressources sont conformes aux pratiques internes, aux directives du secteur et aux réglementations.
- AWS Security Hub— Cela Service AWS fournit une vue complète de votre état de sécurité interne AWS. Security Hub utilise des contrôles de sécurité pour évaluer vos ressources AWS et vérifier votre conformité par rapport aux normes et aux bonnes pratiques du secteur de la sécurité. Pour obtenir la liste des services et des contrôles pris en charge, consultez Référence des contrôles Security Hub.
- Amazon GuardDuty — Cela Service AWS détecte les menaces potentielles qui pèsent sur vos charges de travail Comptes AWS, vos conteneurs et vos données en surveillant votre environnement pour détecter toute activité suspecte et malveillante. GuardDuty peut vous aider à répondre à diverses exigences de conformité, telles que la norme PCI DSS, en répondant aux exigences de détection des intrusions imposées par certains cadres de conformité.
- AWS Audit Manager— Cela vous Service AWS permet d'auditer en permanence votre AWS utilisation afin de simplifier la gestion des risques et la conformité aux réglementations et aux normes du secteur.

Sécurité de l'infrastructure pour les plans d'évolutivité

En tant que service géré, AWS Auto Scaling il est protégé par la sécurité du réseau AWS mondial. Pour plus d'informations sur les services AWS de sécurité et sur la manière dont AWS l'infrastructure est protégée, consultez la section Sécurité du AWS cloud. Pour concevoir votre AWS environnement en utilisant les meilleures pratiques en matière de sécurité de l'infrastructure, consultez la section Protection de l'infrastructure dans le cadre AWS bien architecturé du pilier de sécurité.

Vous utilisez des appels d'API AWS publiés pour accéder AWS Auto Scaling via le réseau. Les clients doivent prendre en charge les éléments suivants :

- Protocole TLS (Transport Layer Security). Nous exigeons TLS 1.2 et recommandons TLS 1.3.
- Ses suites de chiffrement PFS (Perfect Forward Secrecy) comme DHE (Ephemeral Diffie-Hellman) ou ECDHE (Elliptic Curve Ephemeral Diffie-Hellman). La plupart des systèmes modernes tels que Java 7 et les versions ultérieures prennent en charge ces modes.

En outre, les demandes doivent être signées à l’aide d’un ID de clé d’accès et d’une clé d’accès secrète associée à un principal IAM. Vous pouvez également utiliser AWS Security Token Service (AWS STS) pour générer des informations d’identification de sécurité temporaires et signer les demandes.

Quotas pour vos plans de dimensionnement

Vous Compte AWS disposez des quotas par défaut (précédemment appelés limites) liés aux plans de dimensionnement. Sauf indication contraire, chaque quota est spécifique à la région. Vous pouvez demander des augmentations pour certains quotas, et d'autres quotas ne peuvent pas être augmentés.

Pour afficher les quotas pour Application Auto Scaling, ouvrez la console Service Quotas. Dans le volet de navigation, choisissez Services AWSet sélectionnez AWS Auto Scaling Plans.

Pour demander une augmentation de quota, consultez Demande d'augmentation de quota dans le Guide de l'utilisateur Service Quotas.

Vous Compte AWS disposez des quotas suivants relatifs aux plans de dimensionnement.

Nom	Par défaut	Ajustable
Ressources évolutives par type de ressource	Amazon DynamoDB : 3 000 | Groupes Amazon EC2 Auto Scaling : 200 | Tous les autres types de ressources : 500	Oui
Plans de dimensionnement	100	Oui
Instructions de dimensionnement par plan de mise à l'échelle	500	Non
Configurations de suivi de la cible selon l'instruction de dimension nement	10	Non

Gardez à l'esprit les quotas de service lorsque vous mettez à l'échelle vos charges de travail. Par exemple, lorsque vous atteignez le nombre maximal d'unités de capacité autorisé par un service, la montée en puissance cesse. Si la demande baisse et que la capacité actuelle diminue, AWS Auto Scaling vous pouvez à nouveau évoluer. Pour éviter d'atteindre à nouveau cette limite de quota de service, vous pouvez demander une augmentation. Chaque service a ses propres quotas par défaut pour la capacité maximale de la ressource. Pour plus d'informations sur les quotas par défaut

pour les autres services Amazon Web Services, consultez Service endpoints and quotas (Points de terminaison et quotas de service) dans le Référence générale d'Amazon Web Services.

Historique de document pour les plans de dimensionnement

Le tableau suivant décrit les ajouts importants à la AWS Auto Scaling documentation. Pour recevoir les notifications sur les mises à jour de cette documentation, vous pouvez vous abonner au Flux RSS.

Modification	Description	Date
Nouveau contenu pour la migration AWS Auto Scaling vers des options alternatives	Vous pouvez désormais AWS Auto Scaling migrer depuis Amazon EC2 Auto Scaling Predictive Scaling, qui offre davantage de fonctionnalités. Pour plus d'informations, consultez Migrer votre plan de dimensionnement.	5 avril 2024
Nouveau contenu de sécurité	Nous avons publié un chapitre sur la sécurité mis à jour. Dans le cadre de cette mise à jour, nous avons remplacé « Authentification et contrôle d'accès » par la gestion des identités et des accès pour AWS Auto Scaling.	12 mars 2020
Prise en charge des points de terminaison Amazon VPC	Vous pouvez désormais établir une connexion privée entre votre VPC et. AWS Auto Scaling Pour des considérations et des instructions sur la migration, consultez Plans de dimensionnement et points de terminaison de VPC d'interface.	22 novembre 2019

Support pour augmenter la capacité maximale par rapport à la capacité prévue	Ajoute une prise en charge dans la console pour autoriser le plan de dimensionnement à augmenter la capacité maximale au-delà de la capacité de prévision selon une valeur de tampon spécifiée. Pour plus d'informations, consultez la section Paramètres de mise à l'échelle prédictive.	9 mars 2019
Dimensionnement prédictif et améliorations	Vous pouvez désormais utiliser le dimensionnement prédictif pour dimensionner de manière proactive vos groupes Amazon EC2 Auto Scaling. Cette mise à jour ajoute également la prise en charge pour le remplacement des stratégies de dimensionnement créées à l'extérieur du plan de dimensionnement (par exemple, depuis d'autres consoles) et pour contrôler si vous activez la fonctionnalité de dimensionnement dynamique de votre plan.	20 novembre 2018
Prise en charge des paramètres de ressources	Ajout de la prise en charge afin de personnaliser plusieurs paramètres pour chaque ressource individuelle ou multiple au même moment.	9 octobre 2018

Balises comme source d'application	Cette version prend désormais en charge la spécification d'un ensemble de balises comme source d'application.	23 avril 2018
Nouveau service	Version initiale de AWS Auto Scaling.	16 janvier 2018

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.