Guide de l'utilisateur des plans de dimensionnement

AWS Auto Scaling

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

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

Copyright © 2024 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 ... 2
Disponibilité par région ... 3
Tarification ... 3
Fonctionnement des plans de dimensionnement ... 5
Bonnes pratiques ... 8
Autres considérations ... 9
Éviter l'erreur ActiveWithProblems ... 10
Démarrer ... 11
Étape 1 : Rechercher vos ressources évolutives ... 12
Prérequis ... 12
Ajout de votre groupe Auto Scaling à votre nouveau plan de dimensionnement ... 12
En savoir plus sur la découverte de vos ressources évolutives ... 14
Étape 2 : Spécifier la stratégie de dimensionnement ... 16
Étape 3 : Configurer les paramètres avancés (facultatif) ... 19
Paramètres généraux ... 19
Paramètres de mise à l'échelle dynamique ... 22
Paramètres de mise à l'échelle prédictive ... 23
Étape 4 : Créer votre plan de dimensionnement ... 24
(Facultatif) Affichage des informations de mise à l'échelle pour une ressource ... 25
Étape 5 : nettoyer ... 27
Suppression de votre groupe Auto Scaling ... 28
Étape 6 : étapes suivantes ... 28
Sécurité ... 30
Points de terminaison d’un VPC (AWS PrivateLink) ... 30
Créer un point de terminaison de VPC d'interface pour les plans de dimensionnement ... 31
Création d'une stratégie de point de terminaison de VPC pour le dimensionnement des plans ... 31
Migration des points de terminaison ... 32
Protection des données ... 33
Gestion des identités et des accès ... 34

Contrôle d’accès 35
Fonctionnement des plans de dimensionnement avec IAM 35
Rôles liés à un service 39
Exemples de politiques basées sur l’identité 41
Validation de conformité 48
Sécurité de l’infrastructure 49
Quotas 50
Informations connexes 52
Historique de document 55
..... lviii

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 terminez les instances EC2 en augmentant ou 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 Spot— Lancez ou terminez des instances EC2 en augmentant ou en diminuant la capacité cible d'un parc Spot.

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 changements de trafic ou de 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 de faire évoluer plus rapidement votre capacité Amazon EC2 lorsque des pics se produisent régulièrement.

> **Important**
>
> Si vous utilisez des plans de mise à l'échelle uniquement la mise à l'échelle prédictive, nous vous recommandons fortement de paramétrer plutôt les stratégies de mise à l'échelle prédictive directement sur vos groupes Auto Scaling. Cette option récemment introduite offre plus de fonctionnalités, telles que l'utilisation d'agrégations de mesures pour créer de nouvelles mesures personnalisées ou conserver des données de mesure historiques dans les déploiements bleu/vert. Pour plus d'informations, consultez Mise à l'échelle prédictive pour Amazon EC2 Auto Scaling dans le Guide de l'utilisateur Amazon EC2 Auto Scaling.

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 AWS Tools for Windows PowerShell pour démarrer. Pour plus d'informations, consultez le Guide de référence des cmdlets AWS Tools for PowerShell.
- AWS SDK — 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 plus d'informations, consultez Kits SDK AWS.
- 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 Points de AWS Auto Scaling terminaison et quotas dans les points de Références générales AWS .

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 les plans de dimensionnement pour configurer un ensemble d'instructions pour dimensionner vos ressources. Si vous utilisez AWS CloudFormation ou que vous ajoutez des balises à des ressources évolutives, vous pouvez configurer des plans de dimensionnement pour différents ensembles de ressources, par application. LeAWS Auto Scalingconsole fournit des recommandations pour les stratégies de dimensionnement personnalisées 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 dimensionnement indique à AWS Auto Scaling la façon dont vous souhaitez optimiser l'utilisation des ressources dans votre plan de dimensionnement. 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 ?

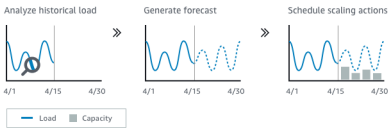
La mise à l'échelle dynamique crée des stratégies de dimensionnement de suivi de la cible pour les ressources dans votre plan de dimensionnement. Ces stratégies de dimensionnement ajustent la capacité des ressources en réponse à des modifications en direct de l'utilisation des ressources. L'objectif 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.



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 effectue des prévisions de demandes à venir 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 ajoutent et suppriment proactivement des capacités de ressources pour refléter la charge prévue. A l'heure prévue, AWS Auto Scaling met à jour la capacité minimale de la ressource avec la valeur spécifiée par l'action de dimensionnement planifiée au moment prévu. 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.

> **Note**
>
> Vous pouvez désormais utiliser les stratégies de dimensionnement prédictive des groupes Auto Scaling à la place. Pour plus d'informations, consultez Mise à l'échelle prédictive pour Amazon EC2 Auto Scaling dans le Guide de l'utilisateur Amazon EC2 Auto Scaling.

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 les données de mesure CloudWatch pour les instances EC2 à une fréquence d'une minute, car cela garantit une réponse plus rapide aux changements 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. Les instances EC2 sont activées pour une surveillance basique par défaut, ce qui signifie que les données de métrique sont disponibles à des intervalles de 5 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 de plus amples informations, veuillez consulter Configuration de la surveillance pour des 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 de plus amples informations, veuillez consulter Surveillance des métriques CloudWatch pour vos instances et groupes Auto Scaling dans le Guide de l'utilisateur 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. Les instances Amazon EC2 à capacité extensible, à savoir les instances T3 et T2, sont conçues pour offrir un niveau de performances d'UC de base, avec la possibilité d'atteindre un niveau supérieur lorsque votre application 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 comme étant illimitées (`unlimited`), consultez Utilisation d'un groupe Auto Scaling pour lancer une instance à capacité extensible en mode Illimité dans le Guide de l'utilisateur Amazon EC2 pour les instances Linux.

Autres considérations

> **Note**
>
> Il existe une version plus récente du dimensionnement prédictif, publiée en mai 2021. Certaines fonctionnalités introduites dans cette version ne sont pas disponibles dans les plans de dimensionnement et vous devez utiliser une stratégie de dimensionnement prédictive définie directement dans le groupe Auto Scaling pour accéder à ces fonctionnalités. Pour plus d'informations, consultez Mise à l'échelle prédictive pour Amazon EC2 Auto Scaling dans le Guide de l'utilisateur Amazon EC2 Auto Scaling.

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 mise à l'échelle 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 une action de mise à l'échelle planifiée qui chevauche une action de mise à l'échelle 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'erreur ActiveWithProblems

Une erreur « ActiveWithProblems » peut se produire lorsqu'un plan de dimensionnement est créé ou que des ressources sont ajoutées 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 stratégies de dimensionnement existantes et les remplacer par des stratégies de dimensionnement Suivi de la cible créées à partir de la console AWS Auto Scaling. 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 ressources d'UC que les instances EC2 d'un groupe Auto Scaling sont censées utiliser par rapport à leur capacité d'UC 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. En le sachant, vous pourrez configurer une fenêtre pour chaque instance EC2 à préparer après le lancement, ce qui nous évitera de lancer un nouveau serveur 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 l'AWS Auto Scalingconsole. S'il s'agit de votre premier plan de dimensionnement, nous vous recommandons de créer un exemple de plan de dimensionnement en utilisant un groupe Amazon EC2 Auto Scaling.

Prérequis

Pour vous entraîner à utiliser un plan de dimensionnement, créez un groupe Auto Scaling. Démarrage d'au moins une instance Amazon EC2 dans le groupe Auto Scaling. Pour plus d'informations, consultez Mise en route avec Amazon EC2 Auto Scaling dans le Guide de l'utilisateur Amazon EC2 Auto Scaling.

Utilisez un groupe Auto Scaling dont les mesures CloudWatch sont activées pour avoir des données de capacité sur les graphiques disponibles lorsque vous terminez l’assistant Créer un plan de dimensionnement. Pour plus d'informations, consultez Activation des métriques de groupe Auto Scaling dans le Guide de l'utilisateur Amazon EC2 Auto Scaling.

Générez une certaine charge pendant quelques jours ou plus pour que les données de mesure CloudWatch soient disponibles pour la fonction de dimensionnement 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 dimensionnement.

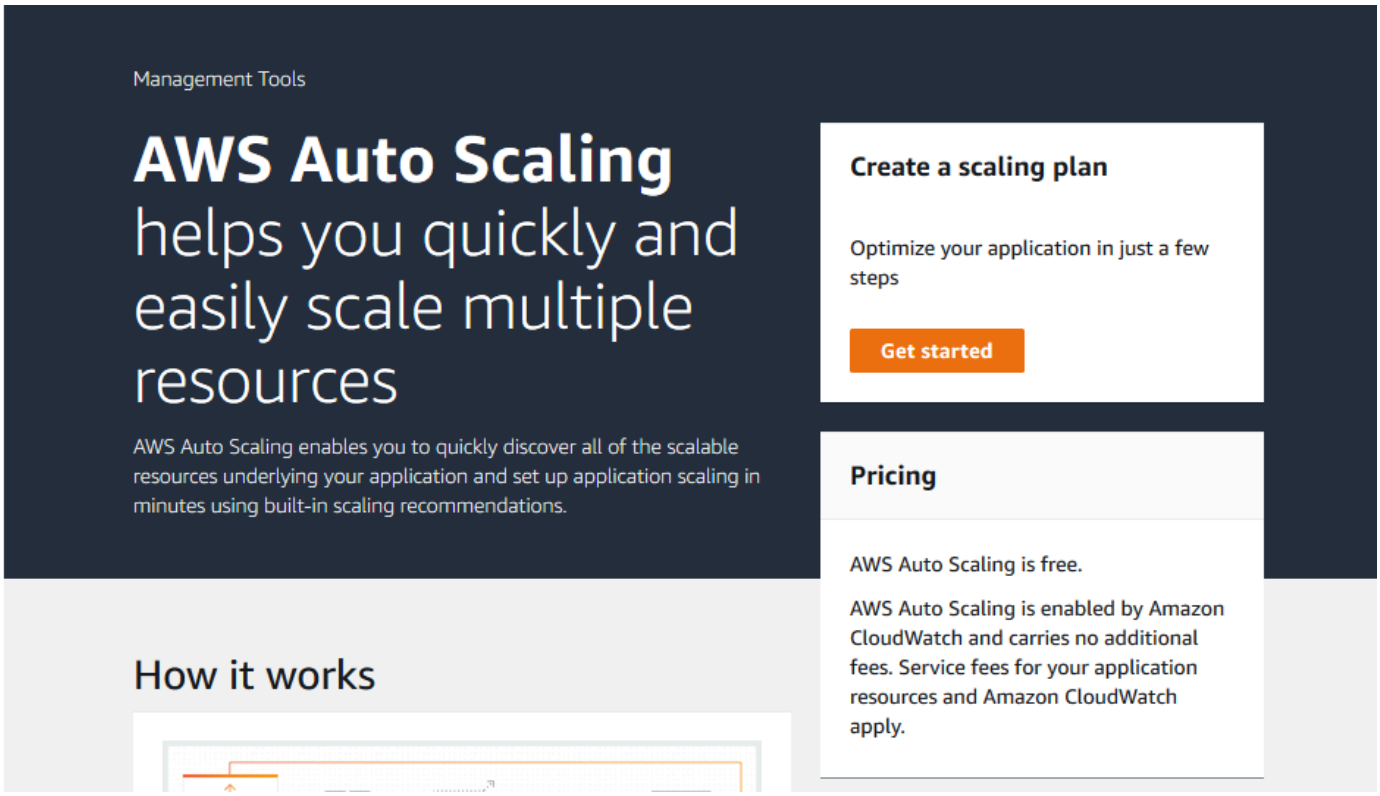
Ajout de votre groupe Auto Scaling à votre nouveau plan de dimensionnement

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 instance EC2, 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

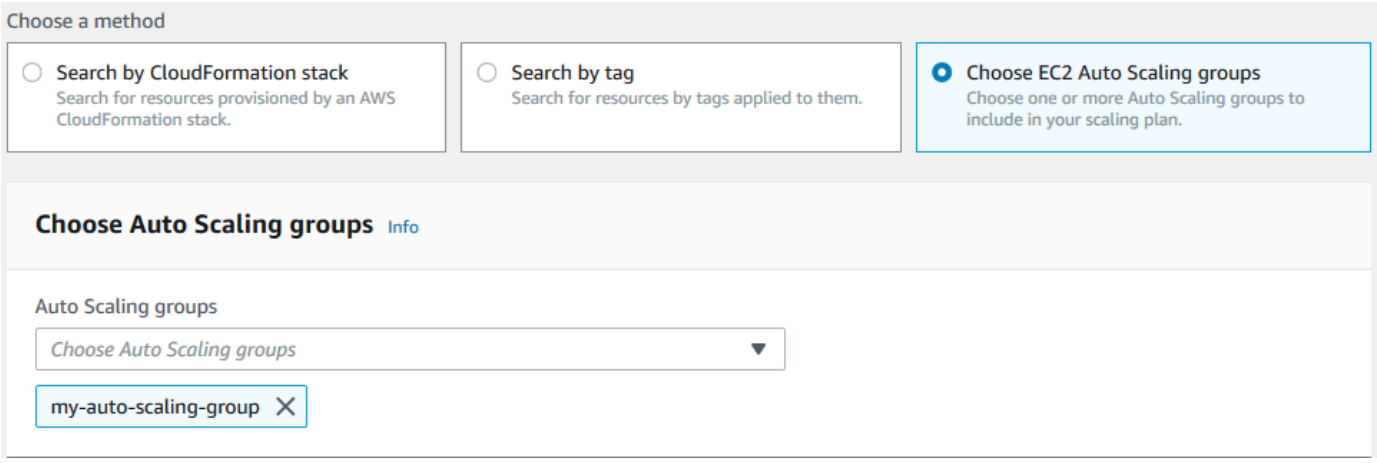
1. Ouvrez la console AWS Auto Scaling à l'adresse https://console.aws.amazon.com/awsautoscaling/.
2. Dans la barre de navigation située en haut de l'écran, choisissez la même région que celle utilisée lorsque vous avez créé votre groupe Auto Scaling.
3. Depuis la page de bienvenue, sélectionnez Get started (Mise en route).



4. Dans la pageRechercher des ressources évolutives, effectuez l'une des actions suivantes :
   - Choisissez Search by CloudFormation stack (Recherche par pile CloudFormation), et ensuite sélectionnez la pile AWS CloudFormation 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 EC2 Auto Scaling groups (Choisir les groupes Auto Scaling EC2), et ensuite sélectionnez un ou plusieurs groupes Auto Scaling.

> **Note**
>
> Pour un didacticiel d'introduction, choisissezChoisissez les groupes EC2 Auto Scaling, 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 dimensionnement et que vous souhaitez en créer d'autres, veuillez voir les scénarios d'utilisation d'une pile CloudFormation ou d'un ensemble de balises. Lorsque vous créez votre plan de dimensionnement à l'aide de la console, cette section peut vous être utile pour choisir entre l'option Search by CloudFormation stack (Recherche par pile CloudFormation) ou Search by tag (Rechercher par balise) afin de découvrir vos ressources évolutives.

Lorsque vous choisissez l'option Search by CloudFormation stack (Recherche par pile CloudFormation) ou Search by tag (Rechercher par balise), le fait de choisir Next (Suivant) a pour effet d'associer les ressources évolutives à la pile ou de rendre l'ensemble des balises accessible au plan de dimensionnement. Vous pouvez choisir les ressources à inclure ou exclure au fur et à mesure que vous définissez votre plan de dimensionnement.

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

Lorsque vous utilisez CloudFormation, vous utilisez des piles pour allouer 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 console AWS Auto Scaling, vous pouvez sélectionnez une pile existante pour y rechercher des ressources pouvant être configurées pour une scalabilité automatique. AWS Auto Scaling recherche uniquement les ressources qui sont définies dans la pile sélectionnée. La recherche ne traverse pas les piles imbriquées.

Pour que vos services ECS puissent être découverts dans une pile CloudFormation, la console AWS Auto Scaling doit savoir quel cluster ECS exécute le service. Pour cela, vos services ECS doivent être dans la même pile CloudFormation 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 sur CloudFormation, consultez Qu'est-ce qu'AWS CloudFormation ? dans le Guide de l'utilisateur AWS CloudFormation.

Découverte de ressources évolutives à l'aide de balises

Les balises fournissent des métadonnées qui peuvent être utilisées pour découvrir des ressources évolutives associées dans la console AWS Auto Scaling, 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 comment baliser des clusters Aurora dans le Guide de l'utilisateur Amazon Aurora.
- Découvrez comment balise Groupes Auto Scaling dans le Guide de l'utilisateur Amazon EC2 Auto Scaling.
- Découvrez comment balise des ressources DynamoDB dans le Amazon DynamoDB Developer Guide.
- Pour en savoir plus sur les bonnes pratiques en matière de balisage, consultez Balisage des ressources AWS dans la Références générales AWS.

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 qui est la plus couramment utilisée pour déterminer la quantité de ressource en cours d'utilisation à tout moment. 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 :

- Optimize for availability (Optimiser pour la disponibilité) – AWS Auto Scaling ajuste automatiquement les ressources pour maintenir leur utilisation à 40 %. Cette option est utile lorsque votre application a des besoins de mise à l'échelle urgents et parfois imprévisibles.
- Balance availability and cost (Équilibrer la disponibilité et les coûts) – AWS Auto Scaling ajuste automatiquement les ressources pour maintenir leur utilisation à 50 %. Cette option vous permet de conserver une haute disponibilité, tout en réduisant les coûts.
- Optimize for cost (Optimiser le coût) – AWS Auto Scaling ajuste automatiquement les ressources pour maintenir leur utilisation 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 instances Amazon EC2 en fonction de la quantité de ressources d'UC utilisée en

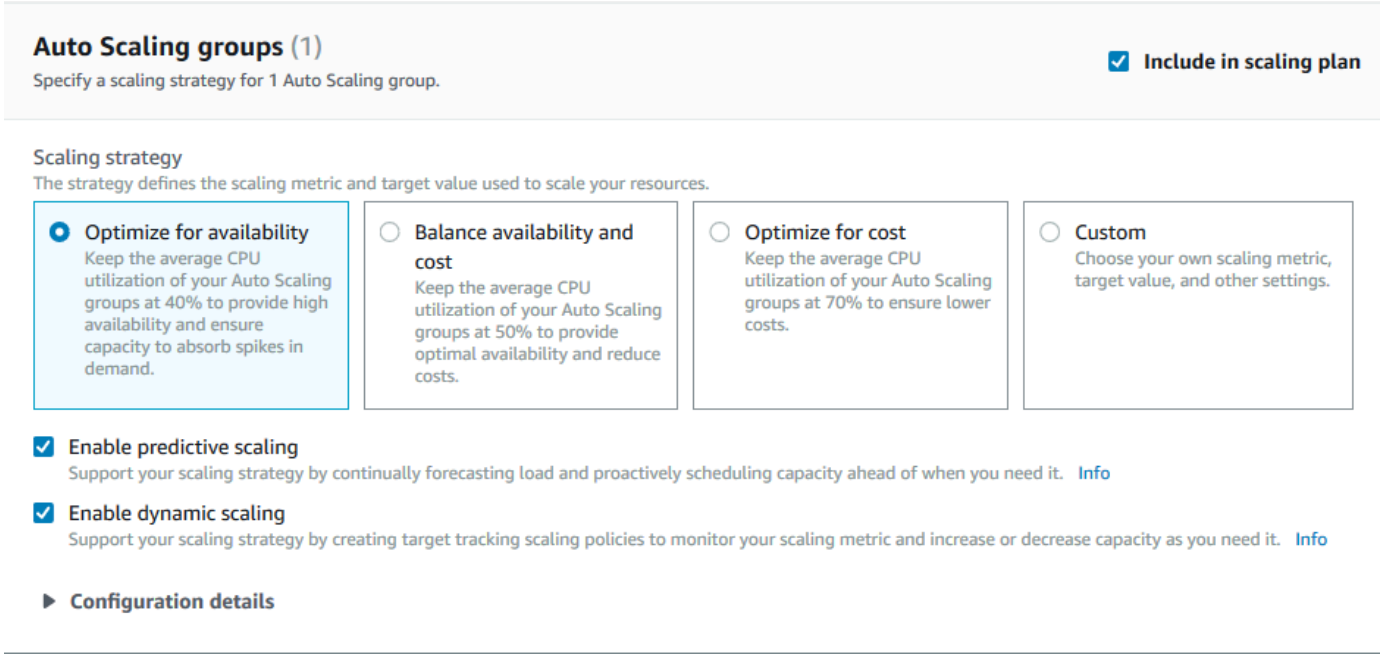
moyenne pour toutes les instances du groupe. Vous pouvez choisir d'optimiser l'utilisation en termes de disponibilité, de coûts ou d'une combinaison des deux, en modifiant la stratégie de dimensionnement.

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.
      - Choisissez Replace external scaling policies si vous souhaitez indiquer que AWS Auto Scaling peut supprimer des stratégies de dimensionnement précédemment créées de l'extérieur du plan de dimensionnement (par exemple, à partir d'autres consoles) et les remplacer par les nouvelles stratégies de dimensionnement de suivi de cible 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.
   - Minimum capacity (Capacité minimale) – Spécifie la capacité minimale de la ressource. AWS Auto Scaling veille à ce que la ressource ne descende jamais en dessous de cette taille.

- Maximum capacity (Capacité maximale) – Spécifie la capacité maximale de la ressource. AWS Auto Scaling veille à ce que la 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 vous permet d'effectuer un choix parmi un grand nombre de métriques différentes. Vous pouvez même publier vos propres métriques.

JSON permet de spécifier une métrique CloudWatch personnalisée. Avant de suivre ces instructions, nous vous recommandons de vous familiariser avec le Guide de l'utilisateur Amazon CloudWatch.

Pour spécifier une métrique personnalisée, vous devez créer une charge utile au format JSON en utilisant un ensemble de paramètres obligatoires de modèle. Vous devez ajouter les valeurs pour chaque paramètre 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 stratégie de dimensionnement de suivi de la cible créée par AWS Auto Scaling.

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.)

- Instance warmup (Préparation d'instance) – [groupes Auto Scaling uniquement] Contrôle le délai écoulé avant qu'une instance nouvellement lancée commence à contribuer aux métriques CloudWatch. Pour plus d'informations, consultez Préparation d'instance dans le Guide de l'utilisateur Amazon EC2 Auto Scaling.

Paramètres de mise à l'échelle prédictive

Si votre ressource est un groupe Auto Scaling, utilisez la procédure ci-dessous pour afficher et personnaliser les paramètres utilisés par AWS Auto Scaling 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).

- Enforce the maximum capacity setting (Appliquer le paramètre de capacité maximale) – AWS Auto Scaling ne peut pas mettre à l'échelle la capacité de ressource au-delà de la capacité maximale. La capacité maximale est appliquée comme une limite stricte.
- Set the maximum capacity to equal forecast capacity (Définir la capacité maximale pour obtenir une capacité égale à celle des prévisions) – AWS Auto Scaling peut mettre à l'échelle la capacité de ressource au-delà de la capacité maximale jusqu'à la capacité de prévision, mais sans la dépasser.
- Increase maximum capacity above forecast capacity (Augmenter la capacité maximale au-delà de la capacité des prévisions) – AWS Auto Scaling peut mettre à l'échelle la capacité de ressource au-delà de la capacité maximale selon une valeur de 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.

Avec la mise à l'échelle prédictive, AWS Auto Scaling analyse l'historique d'une métrique de charge spécifiée dans les 14 derniers jours (minimum de 24 heures de données requises) pour générer une prévision des 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 illustrant l'historique récent des métriques de CloudWatch.
- Graphiques de mise à l'échelle prédictive montrant les prévisions de charge et les prévisions de capacité en fonction des données provenant d'AWS Auto Scaling.
- Un tableau qui répertorie toutes les actions de mise à l'échelle prédictive planifiées pour la ressource.

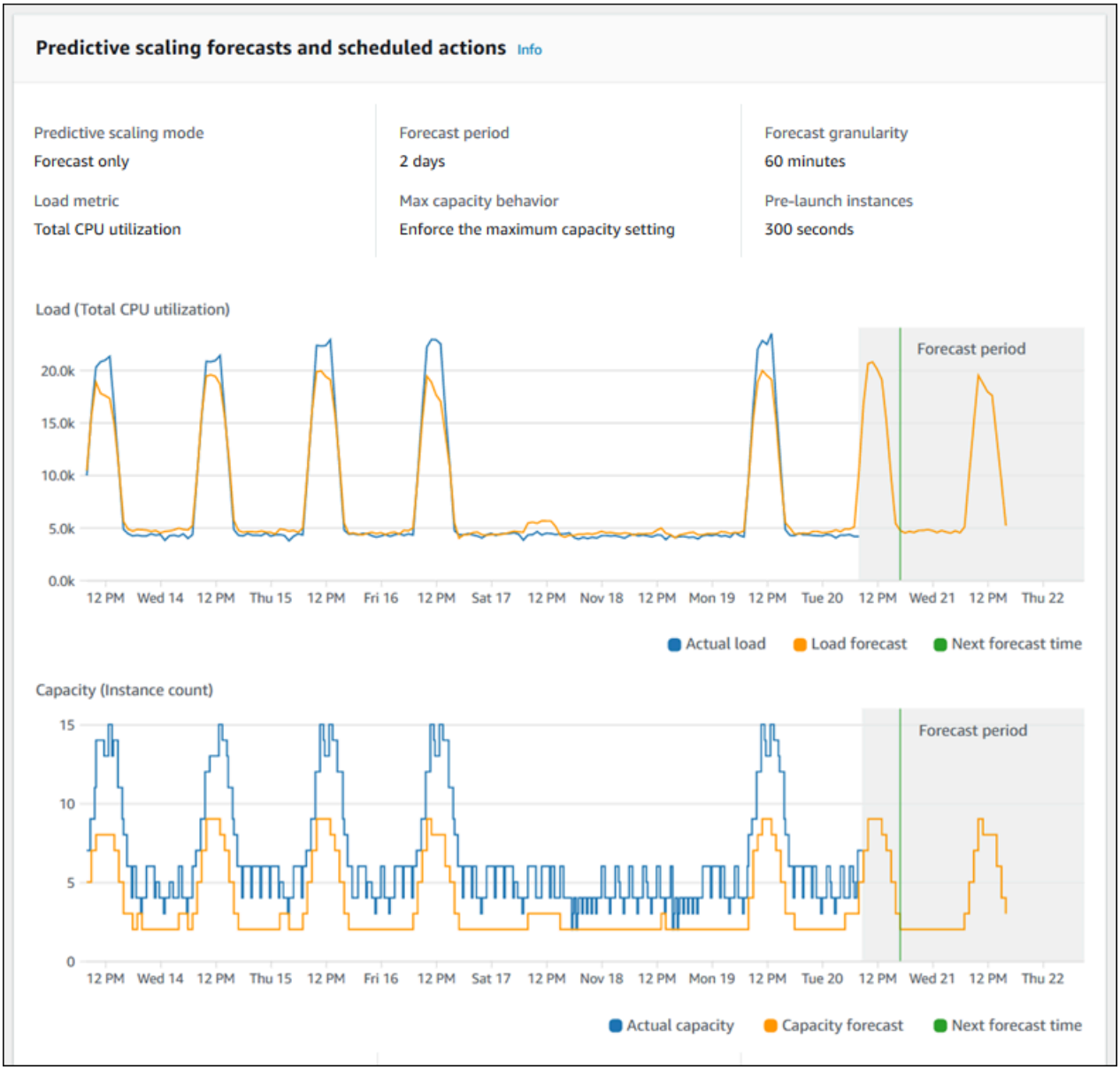
Pour afficher des informations de mise à l'échelle pour une ressource

1. Ouvrez la console AWS Auto Scaling 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 console AWS Auto Scaling pour tous les groupes Auto Scaling qui sont 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 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, AWS Auto Scaling fournit des graphiques basés sur les dernières données réelles. 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 supprime les politiques de suivi des objectifs et d'échelonnement, les alarmes CloudWatch associées et les actions de mise à l'échelle prédictive créées par AWS Auto Scaling en votre nom.

La suppression d'un plan de dimensionnement ne supprime pas votre pile AWS CloudFormation, votre groupe Auto Scaling ou d'autres ressources évolutives.

Pour supprimer un plan de dimensionnement

1. Ouvrez la console AWS Auto Scaling 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 n'engendre des frais Amazon EC2, vous devez également supprimer le groupe Auto Scaling que vous avez créé pour ce tutoriel.

Pour obtenir des instructions détaillées, consultez Suppression de votre groupe Auto Scalingdans le Guide de l'utilisateur 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 modèle AWS CloudFormation est un fichier texte au format JSON ou YAML qui décrit l'infrastructure Amazon Web Services nécessaire à l'exécution d'une application ou d'un service, ainsi que les interconnexions entre les composants de l'infrastructure. Avec AWS CloudFormation,

vous déployez et gérez une collection de ressources associée en tant que pile. AWS CloudFormation est disponible sans frais supplémentaires, vous ne payez que les ressources AWS nécessaires à l'exécution de vos applications. Les ressources peuvent être toute ressource AWS que vous définissez dans le modèle. Pour plus d'informations, consultez Concepts AWS CloudFormation dans le Guide de l'utilisateur AWS CloudFormation.

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 section AWS::AutoScalingPlans::ScalingPlan de la documentation de référence du modèle AWS CloudFormation. 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.

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

- Plans de dimensionnement et points de terminaison d'un VPC (AWS PrivateLink)
- Protection des données
- Gestion des identités et des accès pour les plans de dimensionnement
- Validation de la conformité
- Sécurité de l’infrastructure

Plans de dimensionnement et points de terminaison d'un VPC (AWS PrivateLink)

Vous pouvez améliorer le niveau de sécurité de votre VPC en le configurant de manière AWS Auto Scaling à utiliser un point de terminaison VPC d'interface. Les points de terminaison d'interface

sont alimentés par AWS PrivateLink une technologie qui vous permet d'accéder de manière privée aux AWS Auto Scaling API en limitant tout le trafic réseau entre votre VPC AWS Auto Scaling et le réseau. AWS Avec les points de terminaison d'interface, vous n'avez pas non plus besoin d'une passerelle Internet, d'un appareil NAT ou d'une passerelle privée virtuelle.

Vous n'êtes pas obligé de le configurer AWS PrivateLink, mais c'est recommandé. Pour plus d'informations sur les points AWS PrivateLink de terminaison VPC, consultez Qu'est-ce que c'est ? AWS PrivateLink dans le 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 des SDK AWS CLI et. Si ce n'est pas déjà fait, installez la dernière version AWS CLI et les SDK pour utiliser le nouveau point de terminaison. Pour le mettre à jour AWS CLI, consultez la section Installation du pip AWS CLI d'utilisation dans le guide de l'AWS Command Line Interface utilisateur. Pour plus d'informations sur les AWS SDK, 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 point de terminaison `autoscaling.region.amazonaws.com` en tant que point de terminaison de VPC d'interface privée, consultez Amazon EC2 Auto Scaling et points de terminaison de VPC d'interface 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 connaître les différents kits SDK utilisés pour appeler les API, consultez la documentation du kit SDK concerné pour savoir comment diriger les demandes vers un point de terminaison spécifique. Pour de plus amples informations, veuillez consulter Outils pour Amazon Web Services.

Protection des données

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 en savoir plus 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.
- Utilisez le protocole SSL/TLS pour communiquer avec les ressources. AWS Nous exigeons TLS 1.2 et recommandons TLS 1.3.
- Configurez l'API et la journalisation de l'activité des utilisateurs avec AWS CloudTrail.
- 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 140-2 pour accéder AWS via une interface de ligne de commande ou une API, utilisez un point de terminaison FIPS. Pour en savoir plus sur les points de terminaison FIPS (Federal Information Processing Standard) disponibles, consultez Federal Information Processing Standard (FIPS) 140-2 (Normes de traitement de l'information fédérale).

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 Name (Nom). Cela inclut lorsque vous travaillez avec AWS Auto Scaling ou d'autres Services AWS utilisateurs de la console, de l'API ou AWS des SDK. AWS CLI Toutes les données que vous saisissez 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 dimensionnement

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 (ACL)
- 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 stratégie 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, vous ne pouvez utiliser aucun nom ARN (Amazon Resource Name) dans une stratégie 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 (ACL)

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

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 dont il a besoin 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 plus d’informations, consultez 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 Rôles liés à un service pour Amazon EC2 Auto Scaling dans le guide de l'utilisateur 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, sélectionnez 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.

Fonctions du service

AWS Auto Scaling n'a aucun rôle de service pour les plans de mise à l'échelle.

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 plus d'informations, consultez 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_EC2 AutoScaling pour appeler les actions suivantes en votre nom lorsque vous activez le dimensionnement prédictif :

- `cloudwatch:GetMetricData`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:DescribeScheduledActions`
- `autoscaling:BatchPutScheduledUpdateGroupAction`
- `autoscaling:BatchDeleteScheduledAction`

AWSServiceRoleForAutoScalingPlans_EC2 AutoScaling fait confiance au `autoscaling-plans.amazonaws.com` service pour assumer le rôle.

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 Création d'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 commande create-service-linked-roleCLI 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, consultez Modification d'un rôle lié à un service dans le Guide de l'utilisateur IAM.

Suppression du 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, consultez Suppression d'un rôle lié à un service dans le Guide de l'utilisateur IAM.

Une fois que vous avez supprimé les AWSServiceRoleForAutoScalingPlans_EC2 rôles AutoScaling liés aux services, 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 Régions AWS cas où des plans de scalabilité 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 de plus amples informations, consultez Politiques gérées AWS ou Politiques gérées AWS pour les activités professionnelles dans le Guide de l'utilisateur IAM.
- Accorder 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 de IAM pour appliquer des autorisations, consultez Politiques et autorisations dans IAM dans le Guide de l'utilisateur IAM.
- Utiliser 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 IAM Access Analyzer pour valider vos politiques IAM afin de garantir des autorisations sécurisées et fonctionnelles - IAM Access Analyzer 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 de plus amples informations, consultez Validation de politique 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 le MFA lorsque des opérations d'API sont appelées, ajoutez des conditions MFA à vos politiques. Pour de plus amples informations, consultez Configuration de l'accès aux API protégé par 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.

```
{
    "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. L'ensemble d'autorisations supplémentaires requis pour utiliser la AWS Auto Scaling console est répertorié 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.

```
{
    "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 à partir 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 parcs d'instances Spot à un plan de dimensionnement, les utilisateurs doivent disposer des autorisations suivantes à partir d'Amazon EC2 et 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.

```
{
    "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 plus d’informations, consultez Rôle lié à un service de dimensionnement prédictif.

Validation de la conformité

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 :

- Guides de démarrage rapide sur la sécurité et la conformité : ces guides de déploiement abordent les considérations architecturales et indiquent les étapes à suivre pour déployer des environnements de base axés sur AWS la sécurité et la conformité.
- Architecture axée sur la sécurité et la conformité HIPAA sur Amazon Web Services : ce livre blanc décrit comment les entreprises peuvent créer des applications AWS conformes à la loi HIPAA.

> **Note**
>
> Tous ne Services AWS sont pas éligibles à la loi HIPAA. Pour plus d’informations, consultez le HIPAA Eligible Services Reference.

- 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 de nombreux 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.
- 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

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

Votre Compte AWS dispose des quotas de service suivants (précédemment désignés sous le nom de limites) liés aux plans de dimensionnement.

Pour demander une augmentation des limites, utilisez le formulaire Auto Scaling limits (Limites d'Auto Scaling). Veillez à spécifiez le type de ressource dans votre demande d'augmentation, par exemple Amazon EC2 Auto Scaling, Amazon ECS ou DynamoDB.

Quotas par défaut par région et par compte

Élément	Par défaut
Nombre maximal de ressources évolutives par type de ressource	Les quotas varient en fonction du type de ressource. Amazon DynamoDB : 3 000 Groupes Amazon EC2 Auto Scaling : 200 Tous les autres types de ressources : 500
Nombre maximal de plans de dimensionnement	100
Nombre maximal d'instructions de dimensionnement par plan de dimensionnement	500
Nombre maximal de configurations de suivi de cible selon l'instruction de dimensionnement	10

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 chute et que la capacité actuelle diminue, AWS Auto Scaling peut réaliser à nouveau une montée en puissance. 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.

Informations connexes

Les ressources supplémentaires suivantes peuvent s'avérer utiles lors de l'utilisation de ce service.

Ressource	Description
points de terminaison et quotas AWS Auto Scaling dans Références générales AWS	Informations sur les régions et les points de terminaison AWS Auto Scaling.
Référence API AWS Auto Scaling	La documentation de chaque opération d'API présente la syntaxe de la demande, les paramètres de la demande et les éléments de réponse et fournit des liens vers des rubriques de référence du kit SDK spécifiques au langage.
autoscaling-plans dans Référence des commandes AWS CLI	Descriptions des commandes AWS CLI que vous pouvez utiliser pour travailler avec des plans de dimensionnement.
Guide de l'utilisateur AWS CloudFormation	La référence AWS : :AutoScalingPlans : :ScalingPlan dans la ressource Guide de l'utilisateur AWS CloudFormation vous permet de créer, de modéliser et de gérer vos plans de dimensionnement sans actions manuelles.
Page produit	La page Web principale pour obtenir des informations sur AWS Auto Scaling.
AWS re:Post	service géré par AWS de questions et réponses (Q&A) proposant des réponses participatives et révisées par des experts à vos questions techniques.
Guide de l'utilisateur AWS CloudTrail	Apprenez à utiliserAWS CloudTrailpourenregistrer les appels effectués vers leAWS Auto ScalingAPIet stockez-les dans

Ressource	Description
	Amazon S3. Vous pouvez utiliser ces journaux CloudTrail afin de déterminer les appels qui ont été effectués, l'adresse IP source d'où provient les appels, qui les a effectué, quand ils ont été effectués, et ainsi de suite.
Stratégies de dimensionnement Suivi de la cible pour Amazon EC2 Auto Scaling dans le Guide de l'utilisateur d'Amazon EC2 Auto Scaling	Découvrez comment dimensionner vos groupes Auto Scaling à l'aide des stratégies de dimensionnement suivi de la cible.
Stratégies de dimensionnement Suivi de la cible pour Application Auto Scaling dans le Guide de l'utilisateur Application Auto Scaling.	Découvrez comment mettre à l'échelle vos autres ressources évolutives, telles que les index et tables DynamoDB et les services Amazon ECS, à l'aide de stratégies de dimensionnement suivi de la cible.
Dimensionnement prédictif pour Amazon EC2 Auto Scaling dans le Guide de l'utilisateur Amazon EC2 Auto Scaling	Il s'agit d'une version plus récente du dimensionnement prédictif, publiée en mai 2021. Certaines fonctionnalités introduites dans cette version ne sont pas disponibles dans les plans de dimensionnement et vous devez utiliser une stratégie de dimensionnement prédictive définie directement dans le groupe Auto Scaling pour accéder à ces fonctionn alités.

Les ressources générales suivantes sont disponibles pour vous aider à en savoir plus sur AWS.

- Formations et ateliers : liens vers des formations spécialisées et basées sur les rôles, ainsi que des ateliers d'autoformation pour améliorer vos compétences AWS et acquérir une expérience pratique.
- Centre pour développeurs AWS : parcourez des didacticiels, téléchargez des outils et découvrez les événements pour les développeurs AWS.
- Outils pour développeur AWS : liens vers des outils pour développeur, kits SDK, boîtes à outils IDE et outils de ligne de commande pour développer et gérer des applications AWS.

- Centre de ressources pour la mise en route : découvrez comment configurer votre Compte AWS, rejoindre la communauté AWS et lancer votre première application.
- Didacticiels de prise en main : suivez des didacticiels étape par étape pour lancer votre première application sur AWS.
- Livres blancs AWS : liens vers une liste complète des livres blancs techniques AWS couvrant des sujets tels que l'architecture, la sécurité et l'économie, créés par des architectes de solutions AWS ou d'autres experts techniques.
- AWS SupportCentre – Hub pour la création et la gestion de vos cas AWS Support. Inclut également des liens vers d'autres ressources utiles, telles que des forums, des FAQ techniques, l'état de santé d'un service et AWS Trusted Advisor.
- AWS Support : principale page web d'informations à propos d'AWS Support, un canal d'assistance technique individuelle rapide pour vous aider à développer et à exécuter des applications dans le cloud.
- Contactez-nous : point de contact central pour toute question relative à la facturation AWS, à votre compte, aux événements, à des abus ou à d'autres problèmes.
- AWSConditions d'utilisation du site : informations détaillées sur nos droits d'auteur et notre marque, sur votre compte, votre licence et votre accès au site, et sur d'autres sujets.

Historique de document pour les plans de dimensionnement

Le tableau ci-dessous décrit les ajouts majeurs apportés à la documentation AWS Auto Scaling. Pour recevoir les notifications sur les mises à jour de cette documentation, vous pouvez vous abonner au Flux RSS.

Modification	Description	Date
Nouveau chapitre « Sécurité »	Un nouveau chapitre Sécurité dans le Guide de l'utilisa teur AWS Auto Scaling vous aide à comprendre comment appliquer le modèle de responsabilité partagée lorsque vous utilisez AWS Auto Scaling. Dans le cadre de cette mise à jour, le chapitre « Authentification et contrôle d'accès » du guide de l'utilisa teur a été remplacé par une nouvelle section plus utile, Gestion des identités et des accès pour AWS Auto Scaling.	12 mars 2020
Prise en charge des points de terminaison Amazon VPC	Vous pouvez maintenant établir une connexion privée entre votre VPC et AWS Auto Scaling. Pour des considéra tions et des instructions sur la migration, consultez Plans de dimensionnement et points de terminaison de VPC d'interfa ce.	22 novembre 2019
Prise en charge de l'augment ation de la capacité maximale au-delà de la capacité de	Ajoute une prise en charge dans la console pour autoriser le plan de dimensionnement	9 mars 2019

prévision, ainsi que des modifications apportées au guide	à augmenter la capacité maximale au-delà de la capacité de prévision selon une valeur de tampon spécifiée. Pour plus d'informations, consultez Paramètres de dimensionnement prédictif dans le Guide de l'utilisateur AWS Auto Scaling. Cette version comprend également plusieurs sections remaniées dans le tutoriel Mise en route avec AWS Auto Scaling.	
Dimensionnement prédictif et améliorations	Vous pouvez désormais utiliser le dimensionnement prédictif pour dimensionner proactivement 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. Pour plus d'informations, consultez Démarrez avec AWS Auto Scaling.	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. Pour plus d'informations, consultez Démarrez avec AWS Auto Scaling.	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	Première version 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.