
AWS Auto Scaling

スケーリングプラン



AWS Auto Scaling: スケーリングプラン

Copyright © 2023 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

スケーリングプランとは？	1
サポートされるリソース	1
スケーリングプランの特徴と利点	1
開始方法	2
スケーリングプランを使用	2
料金	2
スケーリングプランの仕組み	3
ベストプラクティス	5
その他の考慮事項	5
ActiveWithProblems エラーの回避	6
開始方法	7
ステップ 1: スケーラブルなリソースの検索	7
前提条件	7
Auto Scaling グループを新たなスケーリングプランに追加	8
スケーラブルなリソースの検出の詳細	9
ステップ 2: スケーリング戦略の指定	10
ステップ 3: 詳細設定を構成する (オプション)	12
全般設定	12
動的なスケーリング設定	13
予測スケーリング設定	14
ステップ 4: スケーリングプランを作成する	15
(オプション) リソースのスケーリング情報を表示する	15
ステップ 5: クリーンアップ	17
Auto Scaling グループの削除	17
ステップ 6: 次のステップ	17
セキュリティ	19
VPC エンドポイント (AWS PrivateLink)	19
スケーリングプラン用インターフェイス VPC エンドポイントの作成	20
スケーリングプラン用の VPC エンドポイント ポリシーを作成	20
エンドポイントの移行	20
データ保護	21
アイデンティティとアクセス管理	22
アクセスコントロール	22
スケーリングプランが IAM と動作する仕組み	22
サービスにリンクされたロール	25
アイデンティティベースポリシーの例	26
コンプライアンス検証	30
インフラストラクチャセキュリティ	31
クォータ	32
リソース	33
ドキュメント履歴	34

スケーリングプランとは？

スケーリングプランを使用して関連したスケーラブル リソースの Auto Scaling をわずか数分で構成できます。例えば、タグを使用して、生産、テスト、開発などのカテゴリでリソースをグループ化できます。次に、各カテゴリに属するスケーラブル リソースのスケーリングプランを検索して設定できます。または、クラウド インフラストラクチャに AWS CloudFormation が含まれている場合、リソースのコレクションを作成するために使用するスタックテンプレートを定義できます。次に、各スタックに属するスケーラブル リソースのスケーリングプランを作成します。

サポートされるリソース

AWS Auto Scaling は、次のサービスとリソースのスケーリングプランの使用をサポートしています。

- Amazon Aurora – Aurora DB クラスター用にプロビジョニングされた Aurora リードレプリカの数を増減します。
- Amazon EC2 Auto Scaling – Amazon EC2 Auto Scaling は、Auto Scaling グループの容量を希望に応じて増減して、EC2 インスタンスを起動または終了します。
- Amazon Elastic Container Service – Amazon ECS のタスク数を増減します。
- Amazon DynamoDB – DynamoDB テーブルまたはグローバル セカンダリ インデックス用にプロビジョニングされた読み取りと書き込み容量を増減します。
- スポットフリート – スポットフリートのターゲット容量を増減して、EC2 インスタンスを起動または終了します。

スケーリングプランの特徴と利点

スケーリングプランには次の特徴と利点があります。

- リソース検出 – AWS Auto Scaling は、スケーリング可能なリソースをアプリケーション内で見つけるため、リソースを自動的に検出します。
- 動的スケーリング – スケーリングプランは Amazon EC2 Auto Scaling と Application Auto Scaling サービスを使用して、スケーラブルなリソース容量を調整することにより、通信量やワークロードの変化に対処します。動的スケーリング メトリクスは標準使用率またはスループット メトリクス、あるいはカスタム メトリクスです。
- 組み込み型スケーリングの推奨事項 – AWS Auto Scaling は、スケーリング戦略とパフォーマンス、コスト、または両者のバランスを最適化するために適用できる推奨事項を示します。
- 予測スケーリング – スケーリングプランは、Auto Scaling グループの予測スケーリングもサポートします。これにより、定期的にスパイクが発生している場合、Amazon EC2 の容量を迅速に拡張できます。

Important

Auto Scaling グループの予測スケーリングを設定するときのみにスケーリングプランを使用している場合は、代わりに Auto Scaling グループの予測スケーリングポリシーを使用することを強くお勧めします。最近導入されたこのオプションは、メトリクス集計を使用して新しいカスタムメトリクスを作成やブルー/グリーン展開全体のメトリクスデータ履歴を保持など、強化された機能を提供します。詳細については、Amazon EC2 Auto Scaling ユーザーガイドの「[Amazon EC2 Auto Scaling の予測スケーリング](#)」を参照してください。

開始方法

次のリソースを使ってスケーリングプラン作成に活用します。

- [スケーリングプランの仕組み \(p. 3\)](#)
- [スケーリングプランのベストプラクティス \(p. 5\)](#)
- [スケーリングプランの使用開始 \(p. 7\)](#)

スケーリングプランを使用

次のインターフェイスのいずれかを使用してスケーリングプランの作成、アクセス、管理することができます。

- AWS Management Console – スケーリングプランにアクセスする際に使用するウェブインターフェイスを提供します。AWS account にサインアップ済みの場合は、ナビゲーションバーの検索ボックスを使用して AWS Auto Scaling を検索し、AWS Auto Scaling を選択してから、AWS Management Console にサインアップすると、スケーリングプランにアクセスできます。
- AWS Command Line Interface (AWS CLI) – さまざまな AWS services のコマンドを提供し、Windows、macOS、および Linux でサポートされています。開始するには、「[AWS Command Line Interface ユーザーガイド](#)」を参照してください。詳細については、AWS CLI Command Reference の「[application-autoscaling](#)」を参照してください。
- AWS Tools for Windows PowerShell – PowerShell 環境でスクリプトを作成するユーザー向けに、さまざまな AWS 製品用のコマンドを提供します。開始するには、「[AWS Tools for Windows PowerShell ユーザーガイド](#)」を参照してください。詳細については、「[AWS Tools for PowerShell Cmdlet Reference](#)」を参照してください。
- AWS SDK – 言語固有の API オペレーションを提供し、署名の計算、リクエストの再試行処理、エラー処理など、接続のさまざまな詳細に対処します。詳細については、[AWS SDK](#) をご参照ください。
- クエリ API – HTTPS リクエストを使用して呼び出す低レベル API アクションを提供します。クエリ API の使用は、AWS services にアクセスする最も直接的な方法です。ただし、リクエストに署名するハッシュの生成やエラー処理など、アプリケーションが低レベルな作業を処理することを要求します。詳細については、「[AWS Auto Scaling API リファレンス](#)」を参照してください。
- AWS CloudFormation – CloudFormation テンプレートを使用したスケーリングプランの作成をサポートします。詳細については、AWS CloudFormation ユーザーガイドの [AWS::AutoScalingPlans::ScalingPlan](#) リファレンスを参照してください。

AWS service にプログラムで接続するため、エンドポイントを使用します。AWS Auto Scaling へのコールのエンドポイントについては、AWS General Reference の「[AWS Auto Scaling エンドポイントとクォータ](#)」を参照してください。このページには、スケーリングプランのリージョン別利用可能性についても表示されます。

料金

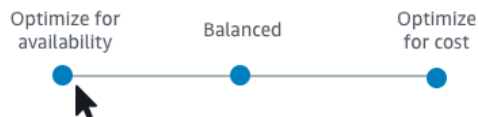
すべてのスケーリングプラン機能が使用可能です。この機能は、CloudWatch および使用されている他の AWS Cloud リソースのサービス料金以外の追加料金なしで提供されます。

スケーリングプランの仕組み

AWS Auto Scaling は、リソースをスケーリングする一連の指示を設定するため、スケーリングプランの使用を可能にします。AWS CloudFormation を使用またはスケーラブルなリソースにタグを追加する場合、アプリケーションごとに異なる一連のリソース用にスケーリングプランを設定できます。AWS Auto Scaling のコンソールは、各リソースに合わせてカスタマイズされたスケーリング戦略に関する推奨事項を提供します。スケーリングプランを作成した後、同プランは動的スケーリングと予測スケーリング方法を組み合わせてスケーリング戦略をサポートします。

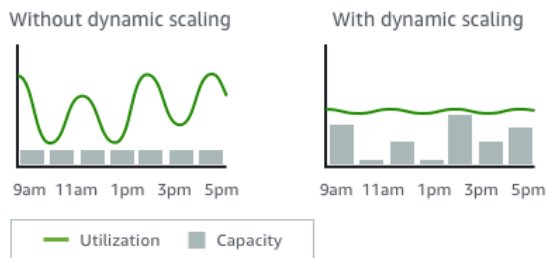
スケーリング戦略とは

スケーリング戦略では、スケーリングプランでのリソース使用率の最適化方法を AWS Auto Scaling に指示します。可用性優先、コスト優先、または両方のバランスを取って最適化できます。または、定義したメトリクスとしきい値に従って、独自のカスタム戦略を作成することもできます。リソースまたはリソースタイプごとに、別の戦略を設定できます。



動的スケーリングとは

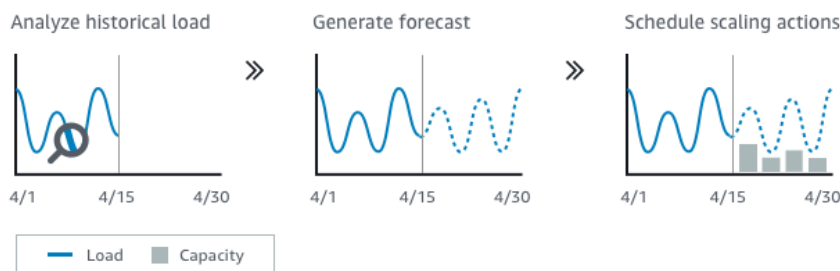
動的スケーリングは、スケーリング計画で、リソースに対するターゲットの追跡スケーリングポリシーを作成します。これらのスケーリングポリシーにより、リソース使用率のリアルタイムの変化に応じてリソースキャパシティが調整されます。これは、スケーリング戦略で指定されたターゲット値に維持するために十分なキャパシティを提供するためです。これはサーモスタットが家の温度を維持する方法に似ています。温度を選択すれば、後はサーモスタットがすべてを実行します。



例えば、スケーリングプランを設定し、Amazon Elastic Container Service (Amazon ECS) サービスが CPU の 75% で実行するタスク数を維持できます。サービスの CPU 使用率が 75% を上回る場合 (サービスに割り当てられた CPU が 75% を上回って使用されること)、スケーリング ポリシーが増加したロードに対処してサービスに別なタスクを追加します。

予測スケーリングとは

予測スケーリングは機械学習によって各リソースのワークロード履歴を分析し、今後のワークロードを定期的に予測します。これは、天気予報の仕組みに似ています。予測スケーリングでは予測を使用して、スケジュールされたスケーリングアクションが生成され、アプリケーションが必要とする前にリソースのキャパシティが準備されます。予測スケーリングは動的なスケーリングのように機能し、スケーリング戦略によって指定されたターゲット値に使用率を維持します。



たとえば、予測スケーリングを有効にしてスケーリング戦略を設定し、Auto Scaling グループの平均 CPU 使用率を 50 パーセントに維持できます。お客様の予測は、毎日午前 8 時に通信量のスパイクが発生が想定されます。スケーリングプランによって今後のスケジュールされたスケーリングアクションが作成され、Auto Scaling グループがトラフィックを処理する準備が事前に行われます。これにより、アプリケーションのパフォーマンスを一定に保ち、可能な限り常に 50 パーセントに近い一定のリソース使用率を維持するために必要なキャパシティを確保します。

予測スケーリングを理解するキーコンセプトは次の通りです。

- **負荷予測：** AWS Auto Scaling は指定された負荷メトリクスの最大 14 日間の履歴を分析して、次の 2 日の今後の需要を予測します。このデータは 1 時間間隔で利用可能であり、毎日更新されます。
- **スケジュールされたスケーリング アクション：** AWS Auto Scaling は、ロード予測に合わせて積極的に容量を増減するスケーリング アクションをスケジュールします。スケジュールされた時刻になると、AWS Auto Scaling は最小容量をスケジュールされたスケーリング アクションで指定された値に基づいて更新します。これは、リソースの使用率をスケーリング戦略で指定されたターゲット値に維持するためです。アプリケーションが予測よりも多くの容量を必要とする場合は、動的スケーリングを使用して追加のキャパシティを追加できます。
- **最大容量動作：** Auto Scaling の最小と最大容量制限は各リソースに適用されます。ただし、予測容量が最大容量を上回った場合、アプリケーションが最大容量を超えて増加できるかどうかについて制御できません。

Note

これで Auto Scaling グループの予測スケーリング ポリシーを代わりに使用できるようになりました。詳細については、Amazon EC2 Auto Scaling ユーザーガイドの「[Amazon EC2 Auto Scaling の予測スケーリング](#)」を参照してください。

スケーリングプランのベストプラクティス

以下のベストプラクティスは、スケーリングプランを最大限に活用するために役立ちます。

- 起動テンプレートまたは起動設定を作成する際、詳細モニタリングを有効にして EC2 インスタンスの CloudWatch メトリクスデータを 1 分間隔の頻度で取得できます。これにより、ロード変化に対する応答が速くなります。5 分間隔のメトリクスをスケーリングすると、応答時間と古いメトリクスデータのスケールが遅くなる可能性があります。デフォルトで EC2 インスタンスは基本モニタリングのために有効になっています。これにより、インスタンスのメトリクスデータは 5 分間隔で利用できます。別途料金で詳細モニタリングを有効にして、1 分間隔でインスタンスのメトリクスデータを取得することができます。詳細については、Amazon EC2 Auto Scaling ユーザーガイドの「[Auto Scaling インスタンスのモニタリングを設定する](#)」を参照してください。
- また、Auto Scaling グループのメトリクスを有効にすることをお勧めします。それ以外の場合は、実際のキャパシティデータは、スケーリングプランの作成ウィザードの完了時に利用できるキャパシティ予測グラフには表示されません。詳細については、Amazon EC2 Auto Scaling ユーザーガイドの「[Auto Scaling グループとインスタンスの CloudWatch メトリクスを監視する](#)」を参照してください。
- Auto Scaling グループで使用されているインスタンスタイプを確認し、バーストパフォーマンスインスタンスのタイプを使用する際には注意してください。T3 インスタンスおよび T2 インスタンスなど、バーストパフォーマンスを含む Amazon EC2 インスタンスは、ベースラインレベルの CPU パフォーマンスを実現するとともに、ワークロードの必要に応じて高いレベルまでバーストする機能を実現できるように設計されています。スケーリングプランで指定されたターゲット使用率により、ベースラインを超過することによって CPU クレジットが不足してパフォーマンスが制限されるリスクがあります。詳細については、「[バーストパフォーマンスインスタンスの CPU クレジットおよびベースラインパフォーマンス](#)」を参照してください。これらのインスタンスを unlimited として設定する場合、Linux インスタンス用 Amazon EC2 ユーザーガイドの「[Auto Scaling グループを使用してバーストパフォーマンスインスタンスを無制限で起動する](#)」を参照してください。

その他の考慮事項

Note

これで Auto Scaling グループの予測スケーリング ポリシーを代わりに使用できるようになりました。詳細については、Amazon EC2 Auto Scaling ユーザーガイドの「[Amazon EC2 Auto Scaling の予測スケーリング](#)」を参照してください。

以下の点を常に考慮してください。

- 予測スケーリングはロード予測によって今後の容量をスケジューリングします。予測の品質はロードのサイクル内容や学習予測モデルによって異なります。予測の品質や、予測で作成されたスケーリングアクションを評価するには、予測スケーリングを予測のみのモードで実行します。[Predictive scaling mode] (予測スケーリングモード) は、スケーリングプランの作成時に [Forecast only] (予測のみ) に設定し、予測の品質の評価終了後に [Forecast and scale] (予測とスケール) に変更することができます。詳細については、[予測スケーリング設定 \(p. 14\)](#) および [予測のモニタリングと評価 \(p. 15\)](#) を参照してください。
- 予測スケーリングに別のメトリクスを指定する場合は、スケーリングメトリクスおよび負荷メトリクスが緊密に相関していることを確認する必要があります。メトリクス値は Auto Scaling グループのインスタンス数に比例して増減する必要があります。これにより、メトリクスデータを使用して比例的にインスタンス数をスケールアウトまたはスケールインできます。たとえば、負荷メトリクスはリクエストの合計数であり、スケーリングメトリクスは CPU 使用率の平均です。キャパシティが変更されない

- 限り、リクエストの合計数が 50 パーセント増加すると、CPU 使用率の平均も 50 パーセント増加します。
- スケーリングプランを作成する前に、作成元のコンソールにアクセスして、不要になったスケジュール済みのスケールアクションを削除する必要があります。AWS Auto Scaling では、スケジュールされた既存のスケールアクションと重なる予測スケールアクションは作成されません。
 - 最小キャパシティと最大キャパシティに関するカスタマイズ済み設定や、動的スケールリングに使用するその他の設定は、他のコンソールに表示されます。ただし、他のコンソールからの更新は、スケールリングプランに送信されないため、スケールリングプランの作成後は、他のコンソールからこれらの設定を変更しないことをお勧めします。
 - スケーリングプランには複数のサービスのリソースを含めることができますが、リソースは一度に 1 つのスケールリングプランにしか存在できません。

ActiveWithProblems エラーの回避

スケールリングプランの作成時やスケールリングプランへのリソースの追加時、「ActiveWithProblems」エラーが発生することがあります。スケールリングプランがアクティブであっても、1 つまたは複数のリソースのスケールリング設定を適用できなかった場合に、このエラーが発生します。

通常、このエラーが発生するのは、リソースにすでにスケールリングポリシーがあるか、Auto Scaling グループが予測スケールリングの最小要件を満たしていないためです。

いずれかのリソースにさまざまなサービスコンソールからのスケールリングポリシーがすでにある場合、AWS Auto Scaling によってこれらの他のスケールリングポリシーが上書きされたり、デフォルトで新しいポリシーが作成されたりしません。オプションで、既存のスケールリングポリシーを削除し、AWS Auto Scaling コンソールから作成されたターゲット追跡スケールリングポリシーに置き換えることができます。そのためには、スケールリングポリシーを上書きする各リソースの [Replace external scaling policies (外部スケールリングポリシーを置き換え)] 設定を有効にします。

予測スケールリングでは、新しい Auto Scaling グループを作成してから 24 時間待ってスケールリングを設定することをお勧めします。最初の予測を生成するには、最低 24 時間の履歴データが必要です。グループの履歴データが 24 時間未満で、予測スケールリングが有効になっている場合、グループで必要なデータ量が収集された後の次の予測期間に達するまで、スケールリングプランで予測を生成することはできません。ただし、24 時間の履歴データが利用可能になり次第、予測プロセスを再開するように、スケールリングプランを編集して保存することもできます。

スケーリングプランの使用開始

アプリケーションで使用するスケーリングプランは AWS Cloud 内で運用されるため、作成前にアプリケーション内容を徹底的に確認します。以下の情報を記録します。

- 他のコンソールから作成された既存のスケーリングポリシーがあるかどうか。スケーリングプランを作成するときに、既存のスケーリングポリシーを置き換えることも、そのポリシーを保持することもできます（値を変更することは許可されません）。
- リソース全体に基づいたアプリケーション内のスケーラブルなリソースそれぞれのターゲット使用率。例えば、使用可能な CPU と比較して、Auto Scaling グループの EC2 インスタンスで使用すると予想される CPU の容量です。または、プロビジョニングされたスルーputモデルを使用する DynamoDB のようなサービスの場合、使用可能なスルーputと比較して、テーブルまたはインデックスが使用すると予想される読み取りおよび書き込みアクティビティの量です。つまり、プロビジョニングされた容量に対する消費の割合です。ターゲット使用率は、スケーリングプランを作成した後にいつでも変更できます。
- サーバーの起動と設定に要する時間。これを知ることによって、起動後にウォームアップするために各 EC2 インスタンスのウィンドウを設定するうえで参考になり、以前のサーバーが起動中に新たなサーバーが起動されることを防止します。
- メトリクス履歴が予測スケーリングに使用できるほど十分に長期か（新規作成した Auto Scaling グループを使用する場合）。一般的に、履歴データが 14 日以上あればより正確な予測に変換されます。最小は 24 時間です。

アプリケーションの理解が進むにつれて、スケーリングプランをより効率的なものにできます。

以下のタスクはスケーリングプランに慣れるうえで参考になります。単一の Auto Scaling グループのスケーリングプランを作成し、予測スケーリングと動的スケーリングを有効にします。

タスク

- [ステップ 1：スケーラブルなリソースの検索 \(p. 7\)](#)
- [ステップ 2：スケーリング戦略の指定 \(p. 10\)](#)
- [ステップ 3：詳細設定を構成する \(オプション\) \(p. 12\)](#)
- [ステップ 4：スケーリングプランを作成する \(p. 15\)](#)
- [ステップ 5：クリーンアップ \(p. 17\)](#)
- [ステップ 6：次のステップ \(p. 17\)](#)

ステップ 1：スケーラブルなリソースの検索

このセクションは AWS Auto Scaling コンソール内でスケーリングプラン作成における実践的内容の紹介が含まれます。初めてスケーリングプランを体験する場合、Amazon EC2 Auto Scaling グループを使用して見本のスケーリングプランを作成することから始めることをお勧めします。

前提条件

スケーリングプランを使う練習をする場合、Auto Scaling グループを作成します。Auto Scaling グループ内で少なくとも 1 つの Amazon EC2 インスタンスを起動します。詳細については、「Amazon EC2 Auto Scaling ユーザーガイド」の「[Amazon EC2 Auto Scaling の開始方法](#)」を参照してください。

スケーリングプランの作成ウィザードを完了した後、CloudWatch メトリクスを有効にした Auto Scaling グループを使用して、利用可能なグラフに容量データを表示します。詳細については、Amazon EC2 Auto Scaling ユーザーガイドの「[Auto Scaling グループのメトリクスを有効にする](#)」を参照してください。

可能な場合、予測スケーリング機能用に CloudWatch メトリクス データを利用可能にするために、数日分のロードを生成します。

スケーリングプランを使用するために必要な許可があることを確認してください。詳細については、「[スケーリングプラン用の Identity and Access Management \(p. 22\)](#)」を参照してください。

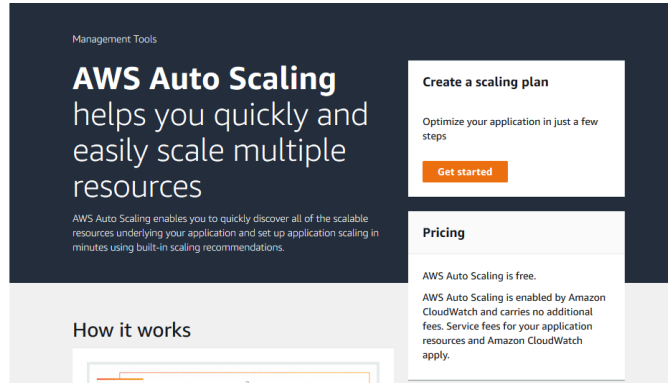
Auto Scaling グループを新たなスケーリングプランに追加

コンソールからスケーリングプランを作成する場合、最初のステップとしてスケーラブル リソースを見つけるうえで役立ちます。開始する前に、以下の要件を満たしていることをご確認ください。

- 前のセクションで説明された通り、Auto Scaling グループを作成し、少なくとも 1 つの EC2 インスタンスを起動しました。
- 作成した Auto Scaling グループは少なくとも 24 時間は存在しています。

スケーリングプランの作成開始

1. AWS Auto Scaling コンソール (<https://console.aws.amazon.com/autoscaling/>) を開きます。
2. 画面上部のナビゲーションバーで、Auto Scaling グループの作成時に使用したのと同じリージョンを選択します。
3. ウェルカムページで、[Get started (使用開始)] を選択します。



4. スケーラブルリソースの検索ページで以下のいずれかの操作を実行します。
 - CloudFormation スタックで検索を選択し、使用する AWS CloudFormation のスタックを選択します。
 - [Search by tag] (タグで検索) を選択します。さらに、タグごとに [Key] (キー) からタグキーを選択し、[Value] (値) からタグ値を選択します。タグを追加するには、[別の列を追加] を選択します。タグを削除するには、[削除] を選択します。
 - [Choose EC2 Auto Scaling groups] (EC2 Auto Scaling グループの選択) を選択し、1 つ以上の Auto Scaling グループを選択します。

Note

入門チュートリアルを開始する場合、[Choose EC2 Auto Scaling groups] (EC2 Auto Scaling グループの選択) を選択し、作成した Auto Scaling グループを選択します。

Choose a method

Search by CloudFormation stack
Search for resources provisioned by an AWS CloudFormation stack.

Search by tag
Search for resources by tags applied to them.

Choose EC2 Auto Scaling groups
Choose one or more Auto Scaling groups to include in your scaling plan.

Choose Auto Scaling groups [Info](#)

Auto Scaling groups

Choose Auto Scaling groups

my-auto-scaling-group X

5. スケーリングプランの作成手順を続行する場合、[Next] (次へ) を選択します。

スケーラブルなリソースの検出の詳細

見本のスケーリングプランをすでに作成して、別なものを作成したい場合、CloudFormation スタックまたは一連のタグを使用するシナリオについて以下の通りに詳しく説明します。このセクションを使用して、コンソールを使用してスケーリングプランを作成するときに、[CloudFormation スタックによる検索] または [タグによる検索] オプションを選択して、スケーラブルなリソースを検出するかどうかを決定できます。

[Create Scaling Plan] (スケーリングプラン作成) ウィザードのステップ 1 で [Search by CloudFormation stack] (CloudFormation スタックで検索) または [Search by tag] (タグで検索) オプションを選択した場合、スタックまたは一連のタグに関連付けられたスケーラブルなリソースが、スケーリングプランで使用できるようになります。スケーリングプランを定義する際、以下のうち、どのリソースを含めるかまたは除外するかを選択できます。

CloudFormation スタックを使用したスケーラブルなリソースの検出

CloudFormation を使用する場合は、スタックを使用してリソースをプロビジョニングします。スタック内のすべてのリソースはスタックのテンプレートで定義されます。スケーリングプランにより、スタックの上にオーケストレーションレイヤーが追加され、複数のリソースのスケーリングを設定しやすくなります。スケーリングプランを使用しない場合は、スケーラブルな各リソースのスケーリングを個別に設定する必要があります。つまり、リソースのプロビジョニングとポリシーのスケーリングの順序を把握し、これらの依存関係がどのように機能するかの微妙な点を理解することが求められます。

AWS Auto Scaling コンソールで、既存のスタックを選択して、オートスケーリング用に設定できるリソースをスキャンできます。AWS Auto Scaling は、選択したスタックで定義されているリソースのみを検索します。ネストされたスタックは通過しません。

CloudFormation スタックで ECS サービスを検出できるようにするには、サービスを実行中の ECS クラスターが AWS Auto Scaling によって認識されている必要があります。そのためには、サービスを実行している ECS クラスターと同じ CloudFormation スタックに ECS サービスが存在している必要があります。それ以外の場合は、デフォルトクラスターの一部である必要があります。正しく識別されるためには、ECS サービス名がこれらの ECS クラスター間で一意であることも必要です。

CloudFormation の詳細については、AWS CloudFormation ユーザーガイドの「[AWS CloudFormation とは](#)」を参照してください。

タグを使用したスケーラブルなリソースの検出

タグは、タグフィルタを使用して、AWS Auto Scaling コンソールで関連するスケーラブルなリソースを検出するために使用できるメタデータを提供します。

タグを使用して、以下のいずれかのリソースを見つけます。

- Aurora DB クラスター

- Auto Scaling グループ
- DynamoDB テーブルとグローバルセカンダリインデックス

複数のタグで検索する場合、各リソースは、リストされたすべてのタグを検出する必要があります。

タグ付けの詳細については、次のドキュメントをご参照ください。

- [Aurora クラスターにタグ付けする方法](#)については、Amazon Aurora ユーザーガイドを参照してください。
- [Auto Scaling グループにタグ付けする方法](#)については、Amazon EC2 Auto Scaling ユーザーガイドを参照してください。
- [DynamoDB リソースにタグ付けする方法](#)については、Amazon DynamoDB 開発者ガイドを参照してください。
- [AWS リソースのタグ付け](#)のベストプラクティスの詳細については、AWS General Referenceを参照してください。

ステップ 2 : スケーリング戦略の指定

前のステップで検出したリソースに対してスケーリング戦略を指定するには、次の手順に従います。

AWS Auto Scaling では、任意の時点で使用されているリソースの量を確認するために最も一般的に使用されるメトリクスがリソースのタイプごとに選択されます。このメトリクスに基づいてアプリケーションのパフォーマンスを最適化する最も適切なスケーリング戦略を選択します。動的なスケーリング機能と予測スケーリング機能を有効にすると、スケーリング戦略は両者間で共有されます。詳細については、「[スケーリングプランの仕組み \(p. 3\)](#)」を参照してください。

使用できるスケーリング戦略は以下のとおりです。

- 可用性を考えた最適化 — AWS Auto Scaling は、リソースを自動的にスケーリングして、リソースの使用率を 40% に維持します。このオプションは、アプリケーションに緊急で予測不可能なスケーリングのニーズがある場合に役立ちます。
- 可用性とコストのバランス — AWS Auto Scaling は、リソースを自動的にスケーリングして、リソースの使用率を 50% に維持します。このオプションは、コストを削減しながら高可用性を維持するのに役立ちます。
- コストに合わせて最適化 — AWS Auto Scaling は、リソースを自動的にスケーリングして、リソースの使用率を 70% に維持します。このオプションは、予期しない需要の変化があったがアプリケーションでバッファ容量の減少に対応できる場合に、コストを削減する上で役立ちます。

たとえば、スケーリングプランで、グループ内のすべてのインスタンスでの CPU の平均使用量に基づき Amazon EC2 インスタンスを追加または削除するように Auto Scaling グループを設定します。スケーリング戦略を変更することで、可用性がコスト、またはその両方に合わせて使用率を最適化します。

また、既存の戦略がニーズに合わない場合、カスタム戦略を設定することもできます。カスタム戦略を設定した場合は、ターゲット使用率の値を変更するか、各メトリクスを選択します。または、両方行うこともできます。

Important

入門チュートリアルの場合、次の手順の最初のステップのみ実行して次へを選択して進みます。

スケーリング戦略の指定

1. [スケーリング戦略の指定] ページの [スケーリングプランの詳細] で、[名前] にスケーリングプランの名前を入力します。スケーリングプランの名前は、地域における一連のスケーリングプランの中で独

特である必要があります。最大 128 文字を使用することができますが、パイプ「|」、フォワードスラッシュ「/」、コロン「:」を含めることはできません。

2. 含まれるすべてのリソースは、リソースタイプ別でリスト化されています。Auto Scaling グループの場合、次の作業を実行します。

Auto Scaling groups (1) Include in scaling plan

Specify a scaling strategy for 1 Auto Scaling group.

Scaling strategy
The strategy defines the scaling metric and target value used to scale your resources.

<input checked="" type="radio"/> Optimize for availability Keep the average CPU utilization of your Auto Scaling groups at 40% to provide high availability and ensure capacity to absorb spikes in demand.	<input type="radio"/> Balance availability and cost Keep the average CPU utilization of your Auto Scaling groups at 50% to provide optimal availability and reduce costs.	<input type="radio"/> Optimize for cost Keep the average CPU utilization of your Auto Scaling groups at 70% to ensure lower costs.	<input type="radio"/> Custom Choose your own scaling metric, target value, and other settings.
---	---	--	--

Enable predictive scaling
Support your scaling strategy by continually forecasting load and proactively scheduling capacity ahead of when you need it. [Info](#)

Enable dynamic scaling
Support your scaling strategy by creating target tracking scaling policies to monitor your scaling metric and increase or decrease capacity as you need it. [Info](#)

► **Configuration details**

- a. デフォルトのスケーリング戦略とメトリクスをする場合、このステップをスキップします。代わりに別のスケーリング戦略またはメトリクスを使用する場合、次のステップを実行します：
 - i. スケーリング戦略の場合、適したスケーリング戦略を選択します。

入門チュートリアルの場合、[Optimize for availability] (利用可能性に向けた最適化) を選択することを必ず確認します。これによって、Auto Scaling グループの平均 CPU 使用率が 40% に維持されるよう指定されます。
 - ii. [Custom] (カスタム) を選択する場合、[Configuration details] (設定の詳細) を展開して適したメトリクスとターゲット値を選択します。
 - [スケーリングメトリクス] で、目的のスケーリングメトリクスを選択します。
 - ターゲット値の場合、任意の 1 分間隔のターゲット使用率やターゲットスループットなど、適したターゲット値を選択します。
 - [Auto Scaling groups only] (Auto Scaling グループのみ) 負荷メトリクスの場合、予測スケーリングに使用する負荷メトリクスを選択します。
 - AWS Auto Scaling が、スケーリングプラン外部 (他のコンソールなど) で以前作成されたスケーリングポリシーを削除して、スケーリングプランによって作成された新しいターゲット追跡スケーリングポリシーで置き換え可能であることを指定するため、[Replace external scaling policies] (外部スケーリング ポリシーの置き換え) を選択します。
 - b. (オプション) デフォルトでは、Auto Scaling グループの予測スケーリングが有効になります。Auto Scaling グループで予測スケーリングをオフにする場合、[Enable predictive scaling] (予測スケーリングを有効化) をクリアします。
 - c. (オプション) デフォルトでは、リソースタイプごとに動的スケーリングが有効化されます。リソースタイプの動的スケーリングをオフにする場合、[Enable dynamic scaling] (動的スケーリングを有効化) を消します。
 - d. (オプション) デフォルトでは、複数のスケーラブルなリソースの検出元となるアプリケーションソースを指定すると、スケーリングプランにはリソースタイプがすべて自動的に含まれます。スケーリングプランからリソースのタイプを省略するには、[スケーリングプランに含める] をオフにします。
3. (オプション) 別なリソースタイプのスケーリング戦略を指定する場合、前述のステップを繰り返します。
 4. 完了したら[Next] (次へ) を選択して、スケーリングプランの作成手順を続行します。

ステップ 3：詳細設定を構成する (オプション)

リソースタイプごとに使用するスケーリング戦略を指定したので、[詳細設定の設定] ステップで、リソース単位でデフォルト設定をカスタマイズできます。リソースタイプごとに、カスタマイズ可能な設定のグループが複数あります。ただし、ほとんどの場合、慎重に調整する必要のある最小と最大容量のそれぞれの値を除き、デフォルト設定は効率化が向上しています。

デフォルト設定を維持する場合は、この手順をスキップします。これらの設定はスケーリングプランを編集することでいつでも変更できます。

Important

入門チュートリアルで、Auto Scaling グループの最大容量を更新し、予測のみモードで予測スケーリングを有効にするためにいくつかの変更を加えましょう。チュートリアルの設定をすべてカスタマイズする必要はありませんが、各セクションの設定についても簡単に調べてみましょう。

全般設定

この手順を使用して、前の手順で指定した設定を表示し、リソースごとにカスタマイズします。各リソースの最小容量と最大容量をカスタマイズすることもできます。

全般設定を表示およびカスタマイズするには

1. [詳細設定の設定] ページで、いずれかのセクション見出しの左側にある矢印を選択して、セクションを展開します。チュートリアルの [Auto Scaling グループ] セクションを展開します。
2. 表示されたテーブルから、このチュートリアルで使用している Auto Scaling グループを選択します。
3. [スケーリングプランに含める] オプションは選択したままにします。このオプションが選択されていない場合、リソースはスケーリングプランから除外されます。リソースが 1 つ以上含まれていない場合は、スケーリングプランを作成できません。
4. ビューを展開して、[全般設定] セクションの詳細を表示するには、セクション見出しの左側にある矢印を選択します。
5. 次のいずれかの項目から選択できます。このチュートリアルの [最大キャパシティ] で、現在の値ではなく、3 を入力します。
 - スケーリング戦略 — 可用性、コスト、または両方のバランスを取って最適化できます。カスタムの戦略を指定することもできます。
 - Enable dynamic scaling (動的スケーリングを有効にする) — この設定をオフにすると、選択したリソースでターゲット追跡スケーリング設定を使用してスケーリングすることはできません。
 - 予測スケーリングの有効化 — [Auto Scaling グループのみ] この設定をオフにすると、選択されたグループで予測スケーリングを使用してスケーリングすることはできません。
 - スケーリングメトリクス — 使用するスケーリングメトリクスを指定します。[カスタム] を選択した場合は、コンソールで利用できる事前定義メトリクスの代わりにカスタマイズされたメトリクスを指定することもできます。詳細については、このセクションの次のトピックを参照してください。
 - ターゲット値 — 使用するターゲット使用率の値を指定します。
 - 負荷メトリクス — [Auto Scaling グループのみ] 使用する負荷メトリクスを指定します。[カスタム] を選択した場合は、コンソールで利用できる事前定義メトリクスの代わりにカスタマイズされたメトリクスを指定することもできます。詳細については、このセクションの次のトピックを参照してください。
 - 最小キャパシティ — リソースの最小キャパシティを指定します。AWS Auto Scaling のリソースがこのサイズよりも下回ることはありません。
 - 最大キャパシティ — リソースの最大キャパシティを指定します。AWS Auto Scaling のリソースがこのサイズを上回ることはありません。

Note

予測スケーリングを使用する場合は、予測キャパシティに基づいて使用する別の最大キャパシティ動作を必要に応じて選択できます。この設定は、[予測スケーリング設定] セクションにあります。

カスタムメトリクス

AWS Auto Scaling は、自動スケーリングで最も一般的に使用されるメトリクスを提供します。ただし、必要に応じて、コンソールにあるメトリクスではなく別のメトリクスからデータを取得することもできます。Amazon CloudWatch には CloudWatch から選択できるさまざまなメトリクスが多数用意されており、独自のメトリクスを発行することもできます。

CloudWatch のカスタムメトリクスを指定するには、JSON を使用します。これらの手順を実行する前に、[Amazon CloudWatch ユーザーガイド](#) を理解しておくことをお勧めします。

カスタマイズされたメトリクスを指定するには、テンプレートの一連の必須パラメータを使用して JSON 形式のペイロードを作成する必要があります。CloudWatch のパラメータごとに値を追加します。このテンプレートは、スケーリングプランの詳細設定の [スケーリングメトリクス] および [ロードのメトリクス] のカスタムオプションの一部として提供されています。

JSON は 2 つの方法でデータを表します。

- オブジェクトは、順序が設定されていない一連の名前と値のペアです。オブジェクトは左中括弧 ({) と右中括弧 (}) で囲んで定義します。それぞれの名前と値のペアは名前で始まり、続けてコロン、その後には値が続きます。名前と値のペアはカンマで区切ります。
- 配列は、順序が設定された一連の値です。配列は左角括弧 ([) と右角括弧 (]) で囲んで定義します。配列の項目はカンマで区切ります。

以下は、各パラメータのサンプル値を持つ JSON テンプレートの例です。

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

詳細については、AWS Auto Scaling API リファレンスの「[CustomizedLoadMetricSpecification](#)」と「[CustomizedLoadMetricSpecification](#)」を参照してください。

動的なスケーリング設定

この手順を使用して、AWS Auto Scaling で作成されたターゲット追跡スケーリングポリシーの設定を表示およびカスタマイズします。

動的なスケーリング設定を表示およびカスタマイズするには

1. ビューを展開して、[動的なスケーリング設定] セクションの詳細を表示するには、セクション見出しの左側にある矢印を選択します。
2. 次の項目から選択できます。ただし、このチュートリアルではデフォルトの設定を使用します。

- 外部のスケーリングポリシーの置き換え — この設定をオフにすると、このスケーリングプラン以外で作成されたスケーリングポリシーが維持され、新しいものは作成されません。
- スケールインの無効化 — この設定をオフにすると、指定したメトリクスがターゲット値を下回ったときに、リソースの現在のキャパシティを減らすように自動スケールインが有効になります。
- クールダウン—スケールアウトおよびスケールインのクールダウン期間を作成します。前回のスケーリングアクティビティが有効になるまでスケーリングポリシーが待機する時間をクールダウン期間と呼びます。詳細については、Application Auto Scaling ユーザーガイドの「[クールダウン期間](#)」を参照してください。(この設定は、リソースが Auto Scaling グループの場合は表示されません。)
- インスタンスのウォームアップ — [Auto Scaling グループのみ] 新しく起動されたインスタンスで CloudWatch メトリクスへの反映が開始するまでの時間を制御します。詳細については、Amazon EC2 Auto Scaling ユーザーガイドの「[インスタンスのウォームアップ](#)」を参照してください。

予測スケーリング設定

リソースが Auto Scaling グループの場合は、この手順を使用して予測スケーリング用に AWS Auto Scaling で使用する設定を表示してカスタマイズすることができます。

予測スケーリング設定を表示およびカスタマイズするには

1. ビューを展開して、[予測スケーリング設定] セクションの詳細を表示するには、セクション見出しの左側にある矢印を選択します。
2. 次の項目から選択できます。このチュートリアルでは、[Predictive scaling mode](予測スケーリングモード) を [Forecast only] (予測のみ) に変更します。
 - 予測スケーリングモード—スケールモードを指定します。デフォルトは[Forecast and scale](予測とスケール)です。[Forecast only] (予測のみ) に変更すると、スケーリングプランは将来のキャパシティを予測しますが、スケーリングアクションは適用されません。
 - インスタンスの事前起動 — スケールアウト時にスケーリングアクションを早めに実行するように調整します。たとえば、予測が午前 10 時 00 分にキャパシティを追加する予定の場合、バッファ時間は 5 分 (300 秒) です。したがって、対応するスケーリングアクションの実行時刻は、午前 9 時 55 分になります。これは Auto Scaling グループに対して有益です。これらのグループでは、インスタンスが起動してから実行されるまでに数分かかる場合があるためです。実際の時間は、インスタンスのサイズや、完了する起動スクリプトがあるかどうかなど、いくつかの要因によって異なります。デフォルトは 300 秒です。
 - 最大キャパシティの動作 — 予測キャパシティが現在指定されている最大キャパシティに近づいたかそれを超えたときに、選択されたリソースが最大キャパシティ以上にスケールアップできるかどうかを制御します。デフォルトでは、最大キャパシティの設定が適用されます。
 - 最大キャパシティ設定の適用 — AWS Auto Scaling では最大キャパシティより高いリソースキャパシティをスケーリングできません。最大容量はハード制限として適用されます。
 - 最大キャパシティを予測キャパシティに設定 — AWS Auto Scaling では、予測キャパシティを超えないように最大キャパシティよりも大きいリソースキャパシティを拡張できます。
 - 予測キャパシティより高く最大キャパシティを設定 — AWS Auto Scaling では、リソースキャパシティを最大キャパシティよりも指定されたバッファ値分高くスケールすることができます。これにより、予期しないトラフィックが発生した場合に、ターゲットの追跡スケーリングポリシーに容量が追加されます。
 - 最大キャパシティの動作のバッファ — [予測キャパシティより高く最大キャパシティを設定] を選択した場合は、予測キャパシティが最大キャパシティに近づいたか、それを超えたときに使用するキャパシティバッファのサイズを選択します。この値は予測容量に対する割合として指定されます。たとえば、10% のバッファでは、予測キャパシティが 50、最大キャパシティが 40 の場合、有効な最大キャパシティは 55 です。
3. 設定のカスタマイズが完了したら、[次へ] を選択します。

Note

任意の変更を元に戻すには、元に戻すリソースを選び、[元に戻す] を選択します。これにより、選択されたリソースが、スケーリング計画内の最新の既知の状態にリセットされます。

ステップ 4：スケーリングプランを作成する

[確認と作成] ページでスケーリングプランの詳細を確認し、[スケーリングプランの作成] を選択します。スケーリングプランのステータスを示すページに進みます。スケーリングプランでは、リソースが更新されている間、作成が完了するまで少し時間がかかることがあります。

予測スケーリングでは、AWS Auto Scaling は、今後 2 日間の予測を生成するために、過去 14 日間の指定された負荷メトリクスの履歴を分析します (24 時間以上のデータが必要です)。その後、スケーリングアクションをスケジュールして、予測期間内の各時間の予測と一致するリソース容量を調整します。

スケーリングプランの作成が完了したら、スケーリングプラン画面から名前を選択してスケーリングの詳細を表示します。

(オプション) リソースのスケーリング情報を表示する

この手順を使用して、リソース用に作成されたスケーリング情報を表示します。

データは次の方法で表示されます。

- CloudWatch からの最新のメトリクス履歴データを示すグラフ。
- AWS Auto Scaling のデータに基づく負荷予測とキャパシティ予測を示す予測スケーリンググラフ。
- リソースに対してスケジュールされているすべての予測スケーリングアクションを一覧表示するテーブル。

リソースのスケーリング情報を表示するには

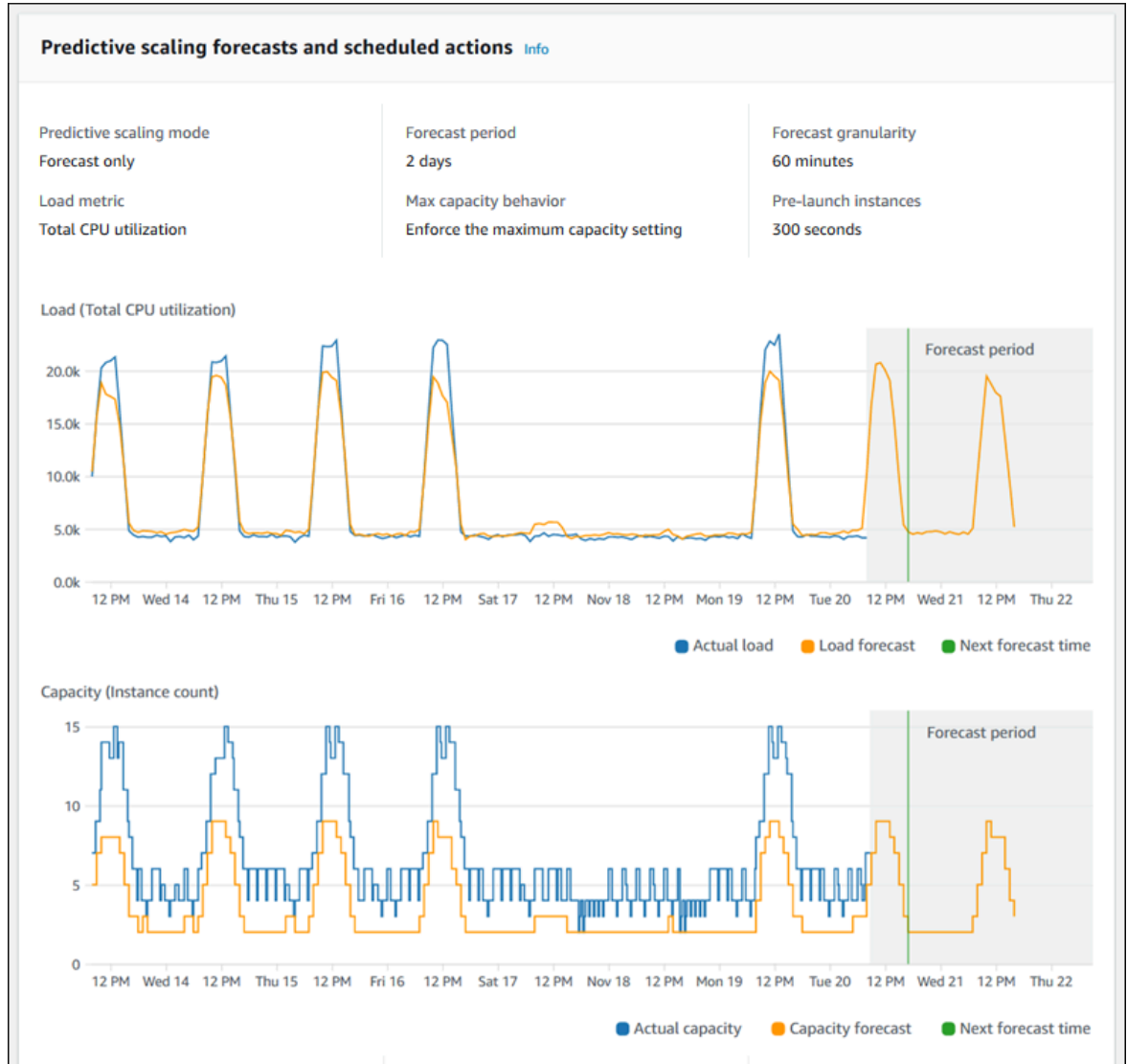
1. AWS Auto Scaling コンソール (<https://console.aws.amazon.com/autoscaling/>) を開きます。
2. [スケーリングプラン] ページで、スケーリング計画を選択します。
3. [スケーリングプランの詳細] ページで、表示するリソースを選択します。

予測のモニタリングと評価

スケーリングプランが立ち上がって実行されると、負荷予測、キャパシティ予測、およびスケーリングアクションをモニタリングして、予測スケーリングのパフォーマンスを調べることができます。このデータはすべて、予測スケーリングが有効になっているすべての Auto Scaling グループについて、AWS Auto Scaling コンソールで利用できます。スケーリングプランでは、最初の予測を行うために少なくとも 24 時間の履歴負荷データが必要であることを注意してください。

次の例では、各グラフの左側に履歴パターンを示しています。右側には、スケーリングプランによって生成された予測期間の予測が表示されています。実際の値と予測値 (青とオレンジ色) がプロットされています。

AWS Auto Scaling スケーリングプラン
(オプション) リソースのスケーリング情報を表示する



AWS Auto Scaling は自動的にデータから学習します。まず、負荷予測を行います。次に、キャパシティー予測の計算により、アプリケーションをサポートするために必要な最小インスタンス数が決定されます。AWS Auto Scaling は、キャパシティー予測に基づき、予測された負荷の変化の前に、Auto Scaling グループをスケールするスケーリングアクションをスケジュールします。動的スケーリングが有効になっている場合 (推奨されます)、Auto Scaling グループはインスタンスのグループの現在の使用状況に基づいて、キャパシティーをスケールアウト (またはキャパシティーを削除) できます。

予測スケーリングのパフォーマンスを評価する場合、時間の経過に伴って、実際の値と予測値がどれだけ一致するかをモニタリングします。スケーリングプランを作成するときは、最新の実際のデータに基づいたグラフが AWS Auto Scaling によって提供されます。また、今後 48 時間の初期予測も提供されます。ただし、スケーリングプランを作成するときに、実際のデータを比較する予測データはほとんどありません。スケーリングプランで、いくつかの期間の予測値を取得するまで待ってから、実際の値に対して履歴の予測値を比較します。毎日の予測を数日分取得すると、実際の値と比較する予測値のサンプルが数多く得られます。

日常的に発生するパターンについては、スケーリングプランの作成と予測効果の評価の間の期間は、数日という短い期間にすることができます。ただし、この期間の長さは、最新のパターンの変更に基いて予測を評価するには不十分です。例えば、前の週に新しいマーケティングキャンペーンを開始した Auto Scaling グループの予測を調べるとします。このキャンペーンにより、毎週同じ 2 日間にウェブトラ

フィックが大きく増加します。このような状況では、グループで 1 週間または 2 週間分の新しいデータを収集するのを待ってから、予測効果を評価することをお勧めします。メトリクスデータの収集を開始したばかりの新しい Auto Scaling グループにも、この同じ推奨事項が適用されます。

適切な期間にわたりモニタリングしても実際の値と予測値が一致しない場合は、負荷メトリクスの選択についても検討する必要があります。負荷メトリクスが有効であるためには、Auto Scaling グループのすべてのインスタンスの合計負荷について、信頼できる正確な計測が得られる必要があります。負荷メトリクスは、予測スケーリングの核となります。最適でない負荷メトリクスを選択した場合、予測スケーリングによる正確な負荷とキャパシティーの予測、および Auto Scaling グループのキャパシティーの正しい調整のスケジュールができなくなる可能性があります。

ステップ 5 : クリーンアップ

開始方法のチュートリアルを完了したら、スケーリングプランを維持するように選択できます。ただし、スケーリングプランをアクティブに使用しない場合は、アカウントで不要な料金が発生しないように削除を検討する必要があります。

スケーリングプランを削除すると、ターゲット追跡スケーリングポリシー、関連する CloudWatch アラーム、お客様に代わって AWS Auto Scaling が作成した予測スケーリングアクションが削除されます。

スケーリングプランを削除しても、AWS CloudFormation スタック、Auto Scaling グループなどのスケラブルなリソースは削除されません。

スケーリング計画を削除するには

1. AWS Auto Scaling コンソール (<https://console.aws.amazon.com/autoscaling/>) を開きます。
2. [スケーリングプラン] ページで、このチュートリアル用に作成したスケーリングプランを選択し、[削除] を選択します。
3. 確認を求めるメッセージが表示されたら、[削除] を選択します。

スケーリングプランを削除した後、リソースは元の容量に戻りません。たとえば、Auto Scaling グループが 10 インスタンスにスケールされていた場合にこのスケーリングプランを削除すると、グループはスケーリングプランが削除された後も 10 インスタンスにスケールされます。特定のリソースのキャパシティーを更新するには、各サービスのコンソールにアクセスします。

Auto Scaling グループの削除

アカウントに対して Amazon EC2 の料金が発生するのを防ぐには、このチュートリアルで作成した Auto Scaling グループも削除する必要があります。

ステップごとの手順については、Amazon EC2 Auto Scaling ユーザーガイドの「[Auto Scaling グループの削除](#)」を参照してください。

ステップ 6 : 次のステップ

スケーリングプランとその特徴の一部を理解したら、AWS CloudFormation を使用して独自のスケーリングプランのテンプレート作成に挑戦してみましょう。

AWS CloudFormation テンプレートは、アプリケーションまたはサービスの実行に必要な Amazon Web Services インフラストラクチャをインフラストラクチャコンポーネント間の相互接続と共に定義する、JSON または YAML 形式のテキストファイルです。AWS CloudFormation では、関連するリソースのコレクションをスタックとしてデプロイおよび管理します。AWS CloudFormation は追加料金なしで利用できます。アプリケーションを実行するために必要な AWS リソース料金のお支払いのみで済みます。

リソースはテンプレート内で定義した AWS のリソースから構成される場合があります。詳細については、AWS CloudFormation ユーザーガイドの「[AWS CloudFormation の概念](#)」を参照してください。

AWS CloudFormation ユーザーガイドでは、すぐに使用できる簡単なテンプレートが提供されています。サンプルテンプレートは、AWS CloudFormation テンプレートリファレンスドキュメントの [AWS::AutoScalingPlans::ScalingPlan](#) セクションで使用可能です。サンプルテンプレートは、1 つの Auto Scaling グループのスケールリングプランを作成し、予測スケールリングおよび動的スケールリングを有効にします。

詳細については、「AWS CloudFormation ユーザーガイド」の「[AWS CloudFormation の開始方法](#)」を参照してください。

スケーリングプランのセキュリティ

AWS では、クラウドのセキュリティが最優先事項です。AWS の顧客は、セキュリティを最も重視する組織の要件を満たすように構築されたデータセンターとネットワークアーキテクチャから利点を得られます。

セキュリティは、AWS とお客様の間の責任共有です。[責任共有モデル](#) では、これをクラウドのセキュリティおよびクラウド内のセキュリティとして説明しています。

- クラウドのセキュリティ – AWS は、AWS クラウド内で AWS サービスを実行するインフラストラクチャを保護する責任を担います。また、AWS は、使用するサービスを安全に提供します。[AWS コンプライアンスプログラム](#)の一環として、サードパーティーの監査が定期的にセキュリティの有効性をテストおよび検証しています。AWS Auto Scaling に適用するコンプライアンスプログラムの詳細については、[コンプライアンスプログラムによる対象範囲内の AWS のサービス](#)を参照してください。
- クラウド内のセキュリティ – お客様の責任はお客様が使用する AWS のサービスによって決まります。また、お客様は、お客様のデータの機密性、企業の要件、および適用可能な法律および規制などの他の要因についても責任を担います。

このドキュメントは、スケーリングプランを使用する際に責任共有モデルの適用方法、並びにスケーリングプランへのアクセスを管理する方法を理解するうえで役立ちます。

トピック

- [スケーリングプランとインターフェイス VPC エンドポイント \(AWS PrivateLink\) \(p. 19\)](#)
- [データ保護 \(p. 21\)](#)
- [スケーリングプラン用の Identity and Access Management \(p. 22\)](#)
- [コンプライアンス検証 \(p. 30\)](#)
- [インフラストラクチャセキュリティ \(p. 31\)](#)

スケーリングプランとインターフェイス VPC エンドポイント (AWS PrivateLink)

インターフェイス VPC エンドポイントを使用するように AWS Auto Scaling を設定することで、VPC のセキュリティ体制を強化できます。インターフェイスエンドポイントは、AWS PrivateLink を使用します。これは、VPC および AWS Auto Scaling と AWS ネットワーク間のすべてのネットワークトラフィックを制限することで、プライベートに AWS Auto Scaling API にアクセスできるようにするテクノロジーです。インターフェイスエンドポイントでは、インターネットゲートウェイ、NAT デバイス、または仮想プライベートゲートウェイも必要ありません。

AWS PrivateLink の設定は要件ではありませんが、推奨されます。AWS PrivateLink と VPC エンドポイントの詳細については、AWS PrivateLink Guide の「[AWS PrivateLink とは](#)」を参照してください。

トピック

- [スケーリングプラン用インターフェイス VPC エンドポイントの作成 \(p. 20\)](#)
- [スケーリングプラン用の VPC エンドポイント ポリシーを作成 \(p. 20\)](#)
- [エンドポイントの移行 \(p. 20\)](#)

スケーリングプラン用インターフェイス VPC エンドポイントの作成

次のサービス名を使用して、AWS Auto Scaling スケーリングプランのエンドポイントを作成します:

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

詳細については、AWS PrivateLink Guideの「[Access an AWS service using an interface VPC endpoint](#)」(インターフェイス VPC エンドポイントを使用して AWS のサービスにアクセスする)を参照してください。

その他の設定を変更する必要はありません。AWS Auto Scaling API は、サービスエンドポイントまたはプライベート インターフェイス VPC エンドポイントのうち使用中のいずれかを使って、他の AWS services サービスをコールします。

スケーリングプラン用の VPC エンドポイント ポリシーを作成

AWS Auto Scaling API へのアクセスをコントロールするために VPC エンドポイントにポリシーをアタッチすることができます。このポリシーでは以下の内容を指定します。

- アクションを実行できるプリンシパル。
- 実行可能なアクション。
- このアクションを実行できるリソース。

以下の例では、エンドポイントを介してスケーリングポリシーを削除するためのアクセス許可を全員に対して拒否する VPC エンドポイントポリシーを示しています。このポリシー例では、他のすべてのアクションを実行するアクセス許可も全員に付与しています。

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

詳細については、AWS PrivateLink Guideの「[VPC エンドポイントポリシー](#)」を参照してください。

エンドポイントの移行

2019 年 11 月 22 日、AWS Auto Scaling API へのコール用に新たなデフォルト DNS ホスト名とエンドポイントとして `autoscaling-plans.region.amazonaws.com` を導入しました。新しいエンドポイントには、AWS CLI および SDK の最新リリースとの互換性があります。まだインストールしていない場合

は、最新の AWS CLI と SDK をインストールして、新しいエンドポイントを使用してください。AWS CLI を更新するには、AWS Command Line Interface ユーザーガイドの「[pip を使用した AWS CLI のインストール](#)」を参照してください。AWS SDK については、「[Amazon Web Services での構築ツール](#)」を参照してください。

Important

下位互換性のために、既存の `autoscaling.region.amazonaws.com` エンドポイントは AWS Auto Scaling API コールで引き続きサポートされます。 `autoscaling.region.amazonaws.com` エンドポイントをプライベートインターフェイス VPC エンドポイントとしてセットアップするには、Amazon EC2 Auto Scaling ユーザーガイドの「[Amazon EC2 Auto Scaling とインターフェイス VPC エンドポイント](#)」を参照してください。

CLI または AWS Auto Scaling API の使用時に呼び出すエンドポイント

AWS Auto Scaling の現在のリリースでは、AWS Auto Scaling API コールは、`autoscaling.region.amazonaws.com` ではなく `autoscaling-plans.region.amazonaws.com` エンドポイントを自動的に参照します。

各コマンドでパラメータ `--endpoint-url https://autoscaling-plans.region.amazonaws.com` を使用してエンドポイントを指定することにより、CLI で新しいエンドポイントを呼び出すことができます。

推奨されませんが、各コマンドでパラメータ `--endpoint-url https://autoscaling.region.amazonaws.com` を使用してエンドポイントを指定することにより、CLI で古いエンドポイントを呼び出すこともできます。

API コールに使用されるさまざまな SDK については、目的の SDK のドキュメントで特定のエンドポイントへのリクエストの送信方法を参照してください。詳細については、[Tools for Amazon Web Services](#) を参照してください。

データ保護

AWS [責任共有モデル](#)は、AWS Auto Scaling のデータ保護に適用されます。このモデルで説明したように、AWS は、すべての AWS クラウド クラウドを実行するグローバルインフラストラクチャを保護します。お客様は、このインフラストラクチャでホストされているコンテンツに対する管理を維持する責任があります。このコンテンツには、使用する AWS サービスのセキュリティ設定および管理タスクが含まれます。データプライバシーの詳細については、「[データプライバシーのよくある質問](#)」を参照してください。欧州でのデータ保護の詳細については、AWS セキュリティブログの「[AWS の責任共有モデルと GDPR](#)」のブログ記事を参照してください。

データ保護の目的で、AWS アカウントの認証情報を保護し、個々のユーザーアカウントを AWS Identity and Access Management (IAM) で設定することをお勧めします。この方法により、それぞれの職務を遂行するために必要なアクセス許可のみを各ユーザーに付与できます。また、以下の方法でデータを保護することをお勧めします。

- 各アカウントで多要素認証 (MFA) を使用します。
- SSL/TLS を使用して AWS リソースと通信します。TLS 1.2 以降が推奨されています。
- AWS CloudTrail で API とユーザーアクティビティログをセットアップします。
- AWS 暗号化ソリューションを、AWS サービス内のすべてのデフォルトのセキュリティ管理と一緒に使用します。
- Amazon Macie などの高度なマネージドセキュリティサービスを使用します。これにより、Amazon S3 に保存される個人データの検出と保護が支援されます。
- コマンドラインインターフェイスまたは API を使用して AWS にアクセスするときに FIPS 140-2 検証済みの暗号化モジュールが必要な場合は、FIPS エンドポイントを使用します。使用可能な FIPS エンドポイントの詳細については、「[連邦情報処理規格 \(FIPS\) 140-2](#)」を参照してください。

顧客の E メールアドレスなどの機密情報や重要情報は、タグや自由形式のフィールド ([名前] フィールドなど) に配置しないことを強くお勧めします。これは、コンソール、API、AWS CLI、または AWS で AWS Auto Scaling または他の AWS サービスを使用する場合も同様です。タグ、または名前に使用される自由形式のフィールドに入力するデータは、請求や診断ログに使用される場合があります。外部サーバーへの URL を指定する場合は、そのサーバーへのリクエストを検証するための認証情報を URL に含めないことを強くお勧めします。

スケーリングプラン用の Identity and Access Management

AWS Identity and Access Management (IAM) は、管理者が AWS リソースへのアクセスを安全にコントロールするために役立つ AWS のサービスです。IAM 管理者は、AWS Auto Scaling リソースを使用するために認証 (サインイン) および承認 (アクセス許可を持つ) される者を制御します。IAM は、追加料金なしで使用できる AWS のサービスです。

スケーリングプランを使用する前に、アカウントにサインインするための AWS account と特定のユーザー認証情報が必要です。詳細については、「[IAM ユーザーガイド](#)」を参照してください。

また、API を介してスケーリングプランを使用する場合は、Query (HTTPS) インターフェイスを介して直接、または [SDK](#)、[AWS Command Line Interface](#)、または [AWS Tools for Windows PowerShell](#) を介して間接的に、AWS アクセスキーを取得する必要があります。AWS アクセスキーは、アクセスキー ID とシークレットアクセスキーで構成されます。AWS アクセスキーの取得に関する詳細については、AWS General Reference の「[AWS セキュリティ認証情報](#)」を参照してください。

アクセスコントロール

有効な認証情報があればリクエストを認証できますが、アクセス許可が付与されている場合を除き、スケーリングプランの作成やアクセスはできません。たとえば、スケーリングプランの作成、予測スケーリングの設定などのアクセス許可が必要です。

以下のセクションでは、IAM 管理者が IAM の使用により、スケーリングプランを使用できるユーザーを制御することによってスケーリングプランを保護する方法について詳しく説明します。

トピック

- [スケーリングプランが IAM と動作する仕組み \(p. 22\)](#)
- [サービスにリンクされたロールの予測スケーリング \(p. 25\)](#)
- [スケーリングプランの ID に基づいたポリシーの例 \(p. 26\)](#)

スケーリングプランが IAM と動作する仕組み

IAM を使用して AWS Auto Scaling スケーリングプランの作成、アクセス、および管理を行う前に、スケーリングプランで使用できる IAM の機能について理解しておく必要があります。

トピック

- [アイデンティティベースのポリシー \(p. 23\)](#)
- [リソースベースのポリシー \(p. 24\)](#)
- [アクセスコントロールリスト \(ACL\) \(p. 24\)](#)
- [タグに基づく認可 \(p. 24\)](#)
- [IAM ロール \(p. 24\)](#)

アイデンティティベースのポリシー

IAM のアイデンティティベースのポリシーでは、許可または拒否するアクションとリソース、およびアクションが許可または拒否される条件を指定できます。スケーリング プランは、特定のアクション、リソース、条件キーをサポートします。JSON ポリシーで使用するすべての要素については、IAM ユーザーガイドの「[IAM JSON ポリシー要素のリファレンス](#)」を参照してください。

アクション

Administrators can use AWS JSON policies to specify who has access to what. That is, which principal can perform actions on what resources, and under what conditions.

JSON ポリシーの Action 要素は、ポリシー内のアクセスを許可または拒否するために使用できるアクションを記述します。ポリシーアクションの名前は通常、関連する AWS API オペレーションと同じです。ただし、一致する API オペレーションを持たないアクセス許可のみのアクションなど、いくつかの例外があります。また、ポリシーに複数のアクションが必要なオペレーションもあります。これらの追加アクションは、従属アクションと呼ばれます。

このアクションは、関連付けられたオペレーションを実行するためのアクセス許可を付与するポリシーで使用されます。

IAM ポリシーステートメントのスケーリングプランアクションは、アクションの前に `autoscaling-plans:` プレフィックスを使用します。ポリシーステートメントには、Action または NotAction 要素を含める必要があります。スケーリングプランには、このサービスで実行できるタスクを記述する独自の一連のアクションがあります。

1 つのステートメントで複数のアクションを指定するには、次の例のようにカンマで区切ります。

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

ワイルドカード (*) を使用して複数のアクションを指定することができます。例えば、Describe という単語で始まるすべてのアクションを指定するには、次のアクションを含めます。

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

ポリシーステートメントに使用できるスケーリングプラン アクションの全体リストを閲覧する場合、サービス認証リファレンス内の「[AWS Auto Scaling 用のアクション、リソース、条件キー](#)」をご参照ください。

リソース

Resource 要素は、アクションが適用されるオブジェクトを指定します。

スケーリングプランは、IAM ポリシーステートメントの Resource 要素として使用できるサービス定義のリソースはありません。そのため、IAM ポリシーで使用する Amazon リソースネーム (ARN) はありません。スケーリングプラン アクションへのアクセスをコントロールする場合、IAM ポリシーを記述する際にリソースとして必ず「*」(アスタリスク)を使用します。

条件キー

Condition 要素 (または Condition ブロック) を使用すると、ステートメントが有効な条件を指定できます。例えば、特定の日付の後のみ適用されるポリシーが必要になる場合があります。条件を表すには、あらかじめ定義された条件キーを使用します。

スケーリングプランはサービス固有の条件キーがありませんが、一部のグローバル条件キーの使用に対応しています。すべての AWS グローバル条件キーを確認するには、IAM ユーザーガイドの「[AWS グローバル条件コンテキストキー](#)」を参照してください。

Condition 要素はオプションです。

例

スケーリングプラン用アイデンティティベースのポリシーの例については、[スケーリングプランの ID に基づいたポリシーの例 \(p. 26\)](#) をご参照ください。

リソースベースのポリシー

Amazon Simple Storage Service など、他の Amazon Web Services では、リソースベースのアクセス権限ポリシーがサポートされています。例えば、ポリシーを S3 バケットにアタッチして、そのバケットに対するアクセス許可を管理できます。

スケーリングプランはリソースベースのポリシーをサポートしません。

アクセスコントロールリスト (ACL)

スケーリングプランはアクセスコントロールリスト (ACL) をサポートしません。

タグに基づく認可

スケーリングプランにタグ付けできません。また、タグ付けできるサービス定義のリソースはありません。したがって、タグ付けしたリソースに基づくアクセスコントロールをサポートしません。

スケーリングプランには、タグに基づくアクセス制御をサポートするタグ可能なリソース (Auto Scaling グループなど) が含まれている場合があります。詳細については、その AWS service のドキュメントを参照してください。

IAM ロール

[IAM ロール](#) は AWS account 内のエンティティで、特定のアクセス許可を持っています。

一時的な認証情報を使用する

一時的な認証情報を使用して、フェデレーションでサインイン、IAM ロールを引き受ける、またはクロスアカウントロールを引き受けることができます。一時的なセキュリティ認証情報を取得するには、[AssumeRole](#) または [GetFederationToken](#) などの AWS STS API オペレーションを呼び出します。

スケーリングプランは一時的な認証情報の使用をサポートします。

スケーリングプランのサービスにリンクされたロール

AWS Auto Scaling は、ユーザーに代わって AWS の他のサービス呼び出すために必要なアクセス許可を持つサービスにリンクされたロールを使用します。必要なアクセス許可を手動で追加する必要がないため、サービスにリンクされたロールはスケーリングプラン設定を簡素化します。詳細については、IAM ユーザーガイドの「[サービスリンクロールの使用](#)」を参照してください。

AWS Auto Scaling は複数タイプのサービスにリンクされたロールを使用し、スケーリングプランを使用する際にお客様に代わって他の AWS services をコールします:

- 予測スケーリングのサービスにリンクされたロール — AWS Auto Scaling に対して CloudWatch の履歴メトリクスデータへのアクセスを許可します。また、ロードと容量予測に基づいて Auto Scaling グループ向けのスケジュールされたアクションを作成することもできます。詳細については、「[サービスにリンクされたロールの予測スケーリング \(p. 25\)](#)」を参照してください。
- Amazon EC2 Auto Scaling のサービスにリンクされたロール — AWS Auto Scaling に対して Auto Scaling グループのターゲット追跡スケーリングポリシーへのアクセスと管理を許可します。詳細については、Amazon EC2 Auto Scaling ユーザーガイドの「[Amazon EC2 Auto Scaling のサービスにリンクされたロール](#)」を参照してください。

- Application Auto Scaling のサービスにリンクされたロール — に対して他のスケーラブルリソースのターゲット追跡スケーリングポリシーへのアクセスと管理を許可します。サービスリンクロールはサービスごとに 1 つあります。詳細については、Application Auto Scaling ユーザーガイドの「[Application Auto Scaling 用のサービスリンクロール](#)」を参照してください。

次の手順を使用して、サービスにリンクされたロールがすでにアカウントにあるかどうかを判断できます。

サービスにリンクされたロールがすでに存在するかどうかを確認するには

1. IAM コンソール (<https://console.aws.amazon.com/iam/>) を開きます。
2. ナビゲーションペインで [ロール] を選択します。
3. リストから `AWSServiceRole` を検索してアカウントに存在するサービスにリンクされたロールを見つけます。目的の、サービスにリンクされたロールの名前を探します。

サービスロール

AWS Auto Scaling はスケーリングプランのサービスロールはありません。

サービスにリンクされたロールの予測スケーリング

AWS Auto Scaling は、スケーリングプランを使用する際にお客様に代わって他の AWS サービスを呼び出すために必要な許可を得るうえでサービスにリンクされたロールを使用します。詳細については、「[スケーリングプランのサービスにリンクされたロール \(p. 24\)](#)」を参照してください。

以下のセクションでは、予測スケーリングにおけるサービスにリンクされたロールの作成と管理する方法について説明します。まず、サービスにリンクされたロールを作成、編集、または削除するためのアクセス許可を設定して IAM エンティティ (ユーザー、グループ、ロールなど) に付与します。

サービスにリンクされたロールによって付与されるアクセス許可

AWS Auto Scaling は、予測スケーリングを有効にした際、`AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` という名前のサービスにリンクされたロールを使用して、お客様に代わって次のアクションを呼び出します。

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

`AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` はロールを実行するうえで `autoscaling-plans.amazonaws.com` のサービスを信頼しています。

サービスにリンクされたロールを作成する (自動)

`AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` のロールを手動で作成する必要はありません。アカウントにスケーリングプランを作成して予測スケーリングを有効にすると、AWS によってこのロールが作成されます。

AWS がお客様に代わってサービスにリンクされたロールを作成するには、必要なアクセス許可がお客様に付与されていない必要があります。詳細については、IAM ユーザーガイドの「[サービスにリンクされたロールのアクセス許可](#)」を参照してください。

サービスにリンクされたロールを作成する (手動)

サービスにリンクされたロールを手動で作成するには、IAM コンソール、IAM CLI、または IAM API を使用できます。詳細については、IAM ユーザーガイドの「サービスにリンクされたロールの作成」を参照してください。

サービスリンクロールの作成 (AWS CLI)

次の [create-service-linked-role](#) の CLI コマンドを使用して サービスにリンクされたロールを作成します。

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

サービスにリンクされたロールを編集する

IAM を使って `AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` の説明を編集できます。詳細については、IAM ユーザーガイドの「[サービスにリンクされたロールの編集](#)」を参照してください。

サービスにリンクされたロールを削除する

スケーリングプランを使用する必要がなくなった場合、`AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` を削除することをお勧めします。

AWS account にあるスケーリングプランのうち、予測スケーリングが有効になっているものをすべて削除した後、サービスにリンクされたロールを削除することができます。これにより、スケーリングプランのアクセス許可を誤って削除することを防止します。

IAM コンソール、IAM CLI、または IAM API を使用して、サービスにリンクされたロールを削除することができます。詳細については、IAM ユーザーガイドの「[サービスにリンクされたロールの削除](#)」を参照してください。

`AWSServiceRoleForAutoScalingPlans_EC2AutoScaling` のサービスにリンクされたロールを削除した後、予測スケーリングを有効にしたスケーリングプランを作成する場合、AWS Auto Scaling は再度ロールを作成します。

サポートされるリージョン

AWS Auto Scaling は、スケーリングプランが利用可能な AWS Regions のすべてに該当するサービスにリンクされたロールの使用をサポートします。スケーリングプランの地域別の利用については、AWS General Reference の「[AWS Auto Scaling エンドポイントとクォータ](#)」を参照してください。

スケーリングプランの ID に基づいたポリシーの例

デフォルトでは、新しい IAM ユーザーには、何かを実施するためのアクセス権限がありません。IAM 管理者は、スケーリングプランを使用するための許可をエンドユーザーに付与する IAM ポリシーを作成して割り当てる必要があります。

JSON ポリシードキュメントのこれらの例を使用して、IAM ポリシーを作成する方法については、IAM ユーザーガイドの「[\[JSON\] タブでのポリシーの作成](#)」を参照してください。

トピック

- [ポリシーのベストプラクティス \(p. 27\)](#)
- [スケーリングプランの作成をユーザーに許可する \(p. 27\)](#)
- [予測スケーリングの有効化をユーザーに許可する \(p. 28\)](#)
- [追加に必要な許可 \(p. 28\)](#)

- [サービスリンクロールの作成に必要な許可 \(p. 30\)](#)

ポリシーのベストプラクティス

アイデンティティベースのポリシーは、そのユーザーがアカウント内で、AWS Auto Scaling リソースを作成、アクセス、または削除できるかどうかを決定します。これらのアクションを実行すると、AWS accountに追加料金が発生する可能性があります。アイデンティティベースのポリシーを作成または編集するときは、以下のガイドラインと推奨事項に従います。

- AWS 管理ポリシーから始めて、最小権限権限に移行する – ユーザーとワークロードに権限を付与し始めるには、多くの一般的なユースケースで使用する権限を付与する AWS 管理ポリシーを使用してください。AWS accountで利用できます。ユースケースに特化した AWS カスタマー管理ポリシーを定義して、権限をさらに削減することをお勧めします。詳細については、IAM ユーザーガイドの「[AWS 管理ポリシー](#)」または「[AWS ジョブ機能の管理ポリシー](#)」を参照してください。
- 最小権限の許可を適用する – IAM ポリシーを使用して許可を設定する場合は、タスクを実行するために必要な許可のみを付与してください。そのためには、特定の条件下で特定のリソースに対して実行できるアクションを定義します (最小権限の許可とも呼ばれます)。IAM を使用して許可を適用する方法の詳細については、IAM ユーザーガイドの「[IAM のポリシーとアクセス許可](#)」を参照してください。
- IAM ポリシー内の条件を使用してアクセスをさらに制限する – ポリシーに条件を追加して、アクションとリソースへのアクセスを制限できます。たとえば、すべてのリクエストを SSL を使用して送信する必要があることを指定するポリシー条件を作成できます。AWS CloudFormation など特定の AWS service を介してサービスアクションが使用される場合は、条件を使用してサービスアクションへのアクセスを付与することもできます。ポリシー要素の詳細については、IAM ユーザーガイドの「[IAM JSON ポリシー要素: 条件](#)」を参照してください。
- IAM Access Analyzer を使用して IAM ポリシーを検証し安全で機能的なアクセス許可を確認する – IAM Access Analyzer は新規および既存のポリシーを検証して、ポリシーが IAM ポリシー言語 (JSON) および IAM のベストプラクティスに準拠していることを確認します。IAM Access Analyzer には、安全で機能的なポリシーの作成に役立つ 100 を超えるポリシーチェックと実用的な推奨事項が用意されています。詳細については、IAM ユーザーガイドの「[IAM Access Analyzer ポリシーの検証](#)」を参照してください。
- 多要素認証 (MFA) を要求する – AWS accountに IAM usersまたはroot userを必要とするシナリオがある場合は、セキュリティを強化するために MFA を有効にしてください。API オペレーションが呼び出されるときに MFA を要求するには、ポリシーに MFA 条件を追加します。詳細については、IAM ユーザーガイドの「[MFA 保護 API アクセスの設定](#)」を参照してください。

IAM のベストプラクティスの詳細については、IAM ユーザーガイドの「[IAM でのセキュリティのベストプラクティス](#)」を参照してください。

スケーリングプランの作成をユーザーに許可する

次に、スケーリングプランの作成をユーザーに許可するアクセス許可ポリシーの例を示します。

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

```
} ]  
}
```

スケーリングプランを使用するには、アカウントで特定のリソースを使用できる追加の許可がエンドユーザーに付与されている必要があります。これらのアクセス許可については、「[追加に必要な許可 \(p. 28\)](#)」を参照してください。

各コンソールユーザーには、自分のアカウントのスケラブルなリソースを検出するアクセス許可や、AWS Auto Scaling コンソールから CloudWatch メトリクスデータのグラフを表示するアクセス許可も必要です。AWS Auto Scaling コンソールを使用して作業するのに必要な追加のアクセス許可は以下に記載されています。

- `cloudformation:ListStacks` : スタックのリストを取得します。
- `tag:GetTagKeys` : 特定のタグキーを含むスケラブルなリソースを検索します。
- `tag:GetTagValues` : 特定のタグ値を含むリソースを検索します。
- `autoscaling:DescribeTags` : 特定のタグを含む Auto Scaling グループを検索します。
- `cloudwatch:GetMetricData` : メトリクスグラフでデータを表示する。

予測スケーリングの有効化をユーザーに許可する

次に、予測スケーリングの有効化をユーザーに許可するアクセス許可ポリシーの例を示します。これらのアクセス許可は、Auto Scaling グループのスケーリングのために設定されたスケーリングプランの機能を拡張します。

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

追加に必要な許可

スケーリングプランのアクセス許可を付与する際、ユーザーがどのリソースを対象に許可を取得するか決定する必要があります。サポートするシナリオに応じて、IAM ポリシーステートメントの Action 要素で次のアクションを指定できます。

Auto Scaling グループ

スケーリングプランに Auto Scaling グループを追加するには、ユーザーに Amazon EC2 Auto Scaling から次のアクセス許可が必要です。

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

- `autoscaling:DeletePolicy`

ECS サービス

スケーリングプランに ECS サービスを追加するには、ユーザーに Amazon ECS と Application Auto Scaling から次のアクセス許可が必要です。

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

スポットフリート

スポットフリートをスケーリングプランに追加するには、ユーザーに Amazon EC2 と Application Auto Scaling から次のアクセス許可が必要です。

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

DynamoDB テーブルまたはグローバルインデックス

スケーリングプランに DynamoDB テーブルまたはグローバルインデックスを追加するには、ユーザーに DynamoDB と Application Auto Scaling から次のアクセス許可が必要です。

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

Aurora DB クラスター

Aurora DB クラスターをスケーリングプランに追加するには、ユーザーに Amazon Aurora と Application Auto Scaling から次のアクセス許可が必要です。

- `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

サービスリンクロールの作成に必要な許可

AWS Auto Scaling は、AWS account 内のユーザーが予測スケーリングを有効にしたスケーリングプランを初めて作成する際、サービスにリンクされたロールを作成するためのアクセス許可が必要です。サービスにリンクされたロールがまだ存在しない場合は、AWS Auto Scaling によってアカウント内に作成されます。サービスにリンクされたロールにより、ユーザーに代わって他のサービスを呼び出すために必要なアクセス許可が AWS Auto Scaling に付与されます。

この自動ロール作成を成功させるには、ユーザーには iam:CreateServiceLinkedRole アクションへのアクセス権限が必要です。

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

次の例は、サービスにリンクされたロールの作成をユーザーに許可する許可ポリシーを示したものです。

```
{
  "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"
        }
      }
    }
  ]
}
```

詳細については、「[サービスにリンクされたロールの予測スケーリング \(p. 25\)](#)」を参照してください。

コンプライアンス検証

AWS Auto Scaling のサービスや他の AWS のサービスが、特定のコンプライアンスプログラムの対象であるかどうかを確認するには、「[コンプライアンスプログラムによる対象範囲内の AWS のサービス](#)」を参照してください。一般的な情報については、「[AWS コンプライアンスプログラム](#)」を参照してください。

サードパーティーの監査レポートをダウンロードするには、AWS Artifact を使用します。詳細については、「」の「[AWS Artifact にレポートをダウンロードする](#)」を参照してください。

AWS サービスを使用する際のお客様のコンプライアンス責任は、データの機密性、企業のコンプライアンス目的、適用法規によって決まります。AWS ではコンプライアンスに役立つ以下のリソースを用意しています。

- [セキュリティおよびコンプライアンスのクイックスタートガイド](#) – これらのデプロイガイドでは、アーキテクチャ上の考慮事項について説明し、セキュリティとコンプライアンスに重点を置いたベースライン環境を AWS でデプロイするための手順を示します。
- [HIPAA のセキュリティとコンプライアンスに関するホワイトペーパーを作成する](#) – このホワイトペーパーでは、企業が AWS を使用して HIPAA 準拠のアプリケーションを作成する方法について説明します。

Note

すべてのサービスが HIPAA に準拠しているわけではありません。

- [AWS コンプライアンスのリソース](#) – このワークブックおよびガイドのコレクションは、お客様の業界や場所に適用される場合があります。
- AWS Config 開発者ガイドの「[ルールでのリソースの評価](#)」 – AWS Config サービスでは、リソース設定が社内のプラクティス、業界のガイドライン、規制にどの程度適合しているかを評価します。
- [AWS Security Hub](#) – この AWS サービスでは、AWS 内のセキュリティ状態を包括的に表示しており、セキュリティ業界の標準およびベストプラクティスへの準拠を確認するのに役立ちます。
- [AWS Audit Manager](#) – この AWS サービスでは、AWS の使用状況を継続的に監査し、リスクの管理方法と、規制や業界標準へのコンプライアンスの管理方法を簡素化できます。

インフラストラクチャセキュリティ

マネージドサービスとしての AWS Auto Scaling は、ホワイトペーパー「[Amazon Web Services: セキュリティプロセスの概要](#)」に記載されている AWS グローバルネットワークセキュリティの手順で保護されています。

AWS の発行済み API コールを使用して、ネットワーク経由で AWS Auto Scaling にアクセスします。クライアントで Transport Layer Security (TLS) 1.0 以降がサポートされている必要があります。TLS 1.2 以降が推奨されています。また、Ephemeral Diffie-Hellman (DHE) や Elliptic Curve Ephemeral Diffie-Hellman (ECDHE) などの Perfect Forward Secrecy (PFS) を使用した暗号スイートもクライアントでサポートされている必要があります。これらのモードは、Java 7 以降など、最近のほとんどのシステムでサポートされています。

また、リクエストは、アクセスキー ID と、IAM プリンシパルに関連付けられているシークレットのアクセスキーを使用して署名する必要があります。または、[AWS Security Token Service \(AWS STS\)](#) を使用して、一時的なセキュリティ認証情報を生成し、リクエストに署名することもできます。

スケーリングプランのクォータ

お客様の AWS account はスケーリングプランに関連する次のクォータ (過去に制限と呼ばれていました) があります。

引き上げをリクエストするには、[Auto Scaling の制限フォーム](#)を使用します。引き上げをリクエストするリソースのタイプ (Amazon EC2 Auto Scaling、Amazon ECS、DynamoDB など) を指定していることを確認します。

各アカウントのリージョンあたりのデフォルトクォータ

項目	デフォルト
リソースタイプごとにスケーラブルなリソースの最大数	クォータは、リソースタイプに応じて異なります。 Amazon DynamoDB: 3000 Amazon EC2 Auto Scaling グループ: 200 その他のすべてのリソースタイプ: 500
スケーリングプランの最大数	100
スケーリングプランあたりのスケーリング命令の最大数	500
スケーリング命令あたりのターゲット追跡設定の最大数	10

ワークロードをスケールアウトする際は、サービスのクォータを念頭に置いてください。例えば、サービスで許可されるキャパシティユニットの最大数に達すると、スケールアウトは停止します。需要が低下し、現在の容量が減少した場合、AWS Auto Scaling は再びスケールアウトできます。このサービスクォータの制限に再度到達しないようにするには、引き上げをリクエストできます。各サービスには、リソースの最大容量に対する独自のデフォルトのクォータがあります。他の Amazon Web Services のデフォルトクォータについては、Amazon Web Services 全般的なリファレンスの「[サービスエンドポイントとクォータ](#)」を参照してください。

スケーリングプランのリソース

次の追加リソースはスケーリングプランを利用する際に参考になります。

- Amazon EC2 Auto Scaling グループを自動的にスケーリングするときに[ターゲット追跡](#)を使用する方法。
- DynamoDB インデックス、テーブル、Amazon ECS サービスなど、Amazon EC2 を超えたスケーラブルリソースの自動スケーリング向けの[ターゲット追跡](#)を使用する方法。
- Amazon S3 を使用して [API へのコールをログに記録AWS Auto Scaling](#)し、AWS CloudTrail に保存する方法。これらの CloudTrail ログを使用して、行われた呼び出し、呼び出し元のソース IP アドレス、呼び出し元、呼び出し時間などを判断できます。

Amazon Web Services の詳細を参照するために、以下の追加リソースを利用できます。

- [クラスとワークショップ](#) – AWS に関するスキルを磨き、実践的経験を積むために役立つセルフペースラボに加えて、職務別のコースおよび特別コースへのリンクです。
- [AWS 開発者用ツール](#) – AWS アプリケーションの開発と管理のための開発者ツール、SDK、IDE ツールキット、およびコマンドラインツールへのリンクです。
- [AWS ホワイトペーパー](#) – アーキテクチャ、セキュリティ、エコノミクスなどのトピックをカバーし、AWS のソリューションアーキテクトや他の技術エキスパートによって書かれた、技術的な AWS ホワイトペーパーの包括的なリストへのリンクです。
- [AWS Support センター](#) – AWS Support のケースを作成して管理するためのハブです。フォーラム、技術上のよくある質問、サービス状態ステータス、AWS Trusted Advisor などの便利なリソースへのリンクも含まれています。
- [AWS Support](#) – AWS Support に関する情報のメインウェブページです。クラウド上のアプリケーションの構築および実行を支援するために 1 対 1 での迅速な対応を行うサポートチャネルとして機能します。
- [お問い合わせ](#) – AWS の請求、アカウント、イベント、不正使用、およびその他の問題に関する問い合わせ先です。
- [AWS サイトの利用規約](#) – 当社の著作権、商標、お客様のアカウント、ライセンス、サイトへのアクセス、その他のトピックに関する詳細情報。

スケーリングプランのドキュメント履歴

次の表は、AWS Auto Scaling ドキュメントへの重要な追加項目をまとめたものです。このドキュメントの更新に関する通知については、RSS フィードにサブスクライブできます。

変更	説明	日付
新しい「セキュリティ」章 (p. 34)	AWS Auto Scaling ユーザーガイドの新しい セキュリティ の章では、AWS Auto Scaling を使用する際に 責任共有モデル を適用する方法を説明しています。この更新の一環として、ユーザーガイドの「認証とアクセスコントロール」の章が、新しくより役立つセクション「 AWS Auto Scaling の Identity and Access Management 」に置き換えられました。	March 12, 2020
Amazon VPC エンドポイントのサポート (p. 34)	これで、VPC と AWS Auto Scaling との間でプライベート接続を確立できます。移行の考慮事項と手順については、「 スケーリングプランとインターフェイス VPC エンドポイント 」を参照してください。	November 22, 2019
予測キャパシティを超える最大キャパシティの増加のサポート、およびガイドの変更 (p. 34)	スケーリングプランで最大キャパシティを予測キャパシティよりも指定されたバッファ値分増やすことを許可するコンソールサポートを追加します。詳細については、AWS Auto Scaling ユーザーガイドの「 予測スケーリング設定 」を参照してください。このリリースでは、「 AWS Auto Scaling の開始方法 」チュートリアル複数のセクションが修正されています。	March 9, 2019
予測スケーリングと機能強化 (p. 34)	予測スケーリングを使用して Amazon EC2 Auto Scaling グループを事前にスケールできるようになりました。このリリースでは、スケーリングプラン以外 (他のコンソールなど) で作成されたスケーリングポリシーの置き換えと、プランの動的スケーリング機能を有効にするかどうかの制御に関するサポートも追加されています。詳細については、「 AWS	November 20, 2018

Auto Scaling の使用開始 を参照してください。		
カスタムリソース設定のサポート (p. 34)	各リソースおよび複数のリソースの設定を同時にカスタマイズできるサポートが追加されました。詳細については、「 AWS Auto Scaling の使用開始 」を参照してください。	October 9, 2018
アプリケーションソースとしてのタグ (p. 34)	このリリースでは、アプリケーションソースとして一連のタグを指定できるようになりました。	April 23, 2018
新しいサービス (p. 34)	AWS Auto Scaling の初回リリース。	January 16, 2018