



Amazon EC2 Auto Scaling 入門編

AWS Black Belt Online Seminar

滝口 開資 (はるよし)
シニアソリューションアーキテクト
EC2 フレキシブルコンピュートスペシャリスト
2023/04

AWS Black Belt Online Seminarとは

- ・ 「サービス別」「ソリューション別」「業種別」などのテーマに分け、アマゾン ウェブ サービス ジャパン合同会社が提供するオンラインセミナーシリーズです
- ・ AWS の技術担当者が、AWS の各サービスやソリューションについてテーマごとに動画を公開します
- ・ 動画を一時停止・スキップすることで、興味がある分野・項目だけの聴講も可能、スキマ時間の学習にもお役立ていただけます
- ・ 以下の URL より、過去のセミナー含めた資料などをダウンロードすることができます
 - ・ <https://aws.amazon.com/jp/aws-jp-introduction/aws-jp-webinar-service-cut/>
 - ・ <https://www.youtube.com/playlist?list=PLzWGOASvSx6FIwIC2X1nObr1KcMCBBlqY>

内容についての注意点

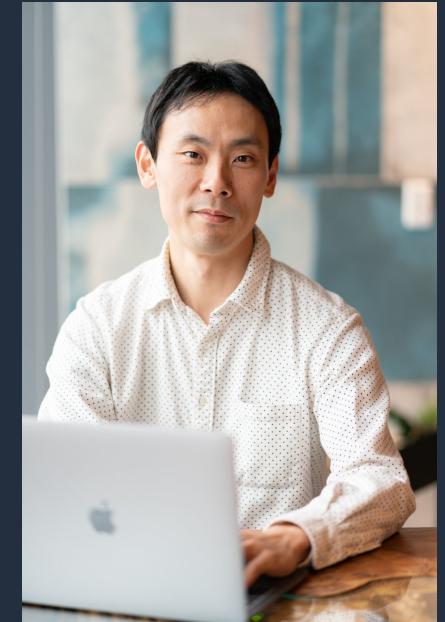
- ・ 本資料では 2023 年 4 月時点のサービス内容および価格についてご説明しています。最新の情報は AWS 公式ウェブサイト (<https://aws.amazon.com/>) にてご確認ください
- ・ 資料作成には十分注意しておりますが、資料内の価格と AWS 公式ウェブサイト記載の価格に相違があった場合、AWS 公式ウェブサイトの価格を優先とさせていただきます
- ・ 価格は税抜表記となっています。日本居住者のお客様には別途消費税をご請求させていただきます

自己紹介

名前：滝口 開資 (はるよし)

所属：アマゾンウェブサービスジャパン合同会社 コンピュート事業本部
シニアソリューションアーキテクト
EC2 フレキシブルコンピュートスペシャリスト

経歴：銀行様担当メインフレーム SE (外資ベンダー)
→クラウドサポートエンジニア (AWS)
→クラウドサポートチームリード (AWS)
→ソリューションアーキテクト (AWS)



好きなAWSサービス：Amazon EC2 Auto Scaling, AWSサポート

本セミナーの対象者

AWS 基盤環境のインフラを担当されている方

EC2 インスタンスの自動スケールの基礎を知りたい方

本セミナーの前提知識：Amazon EC2 の概要

→ Blackbelt Amazon EC2 入門を参照してください。

動画：<https://www.youtube.com/watch?v=1ALvDtb2ziM>

資料：https://pages.awscloud.com/rs/112-TZM-766/images/202111_AWS_Black_Belt_AWS_EC2_introduction.pdf

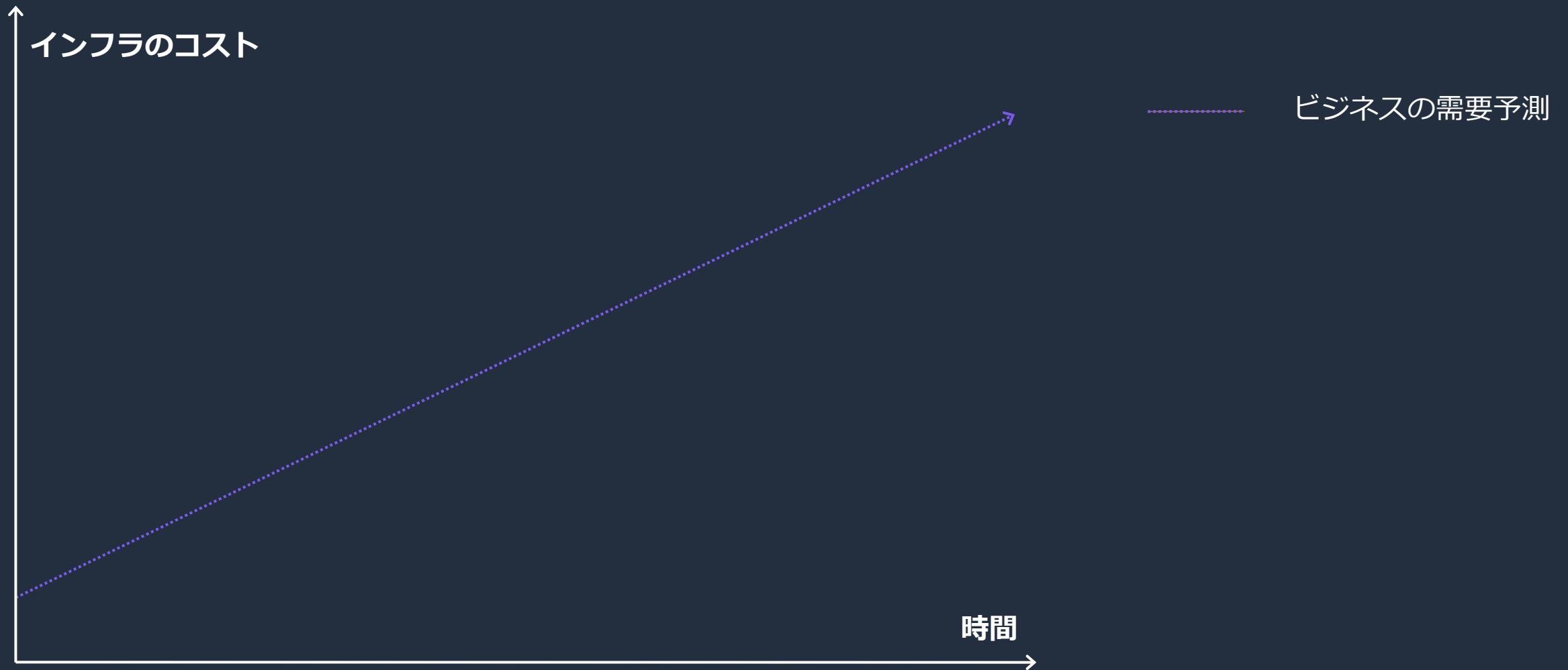
アジェンダ

- Auto Scaling サービスのコンセプト
- EC2 Auto Scaling の動作原理
- EC2 Auto Scaling を使ってみる

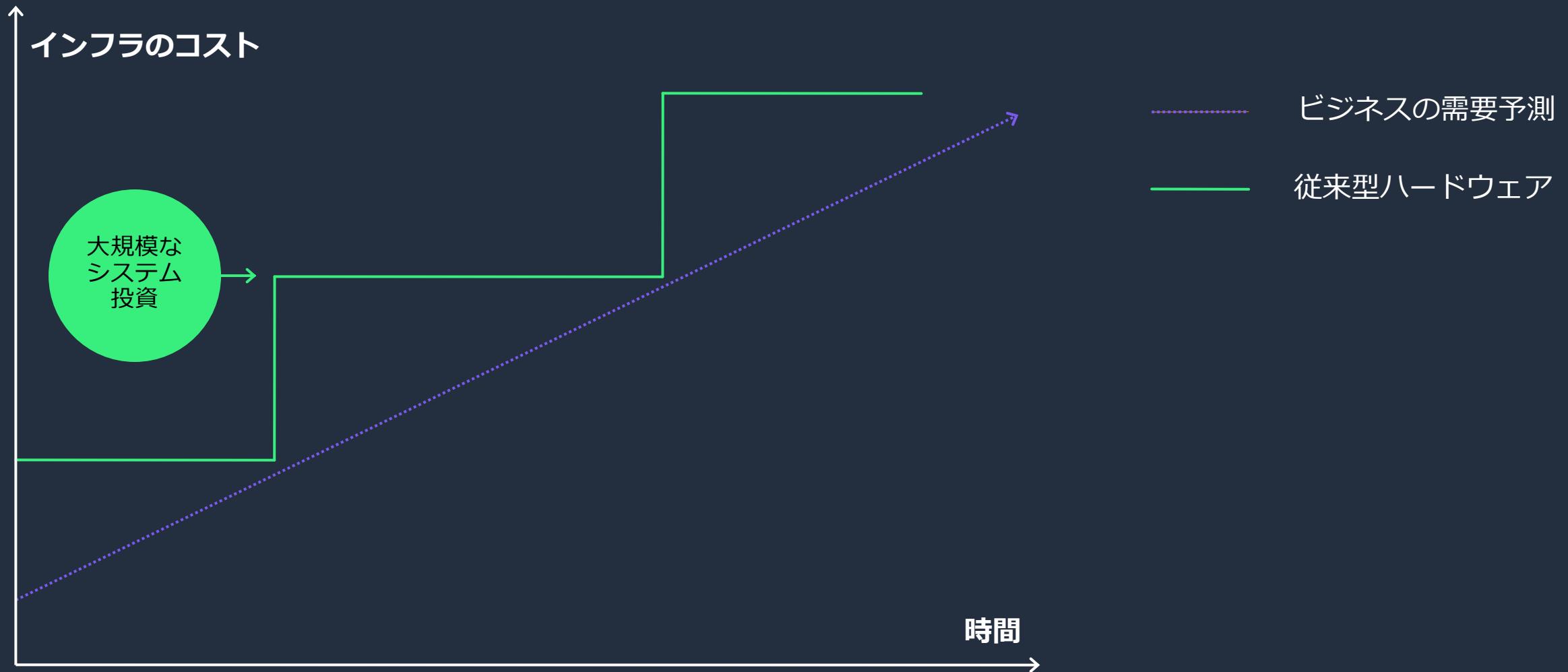
Auto Scaling サービスの コンセプト



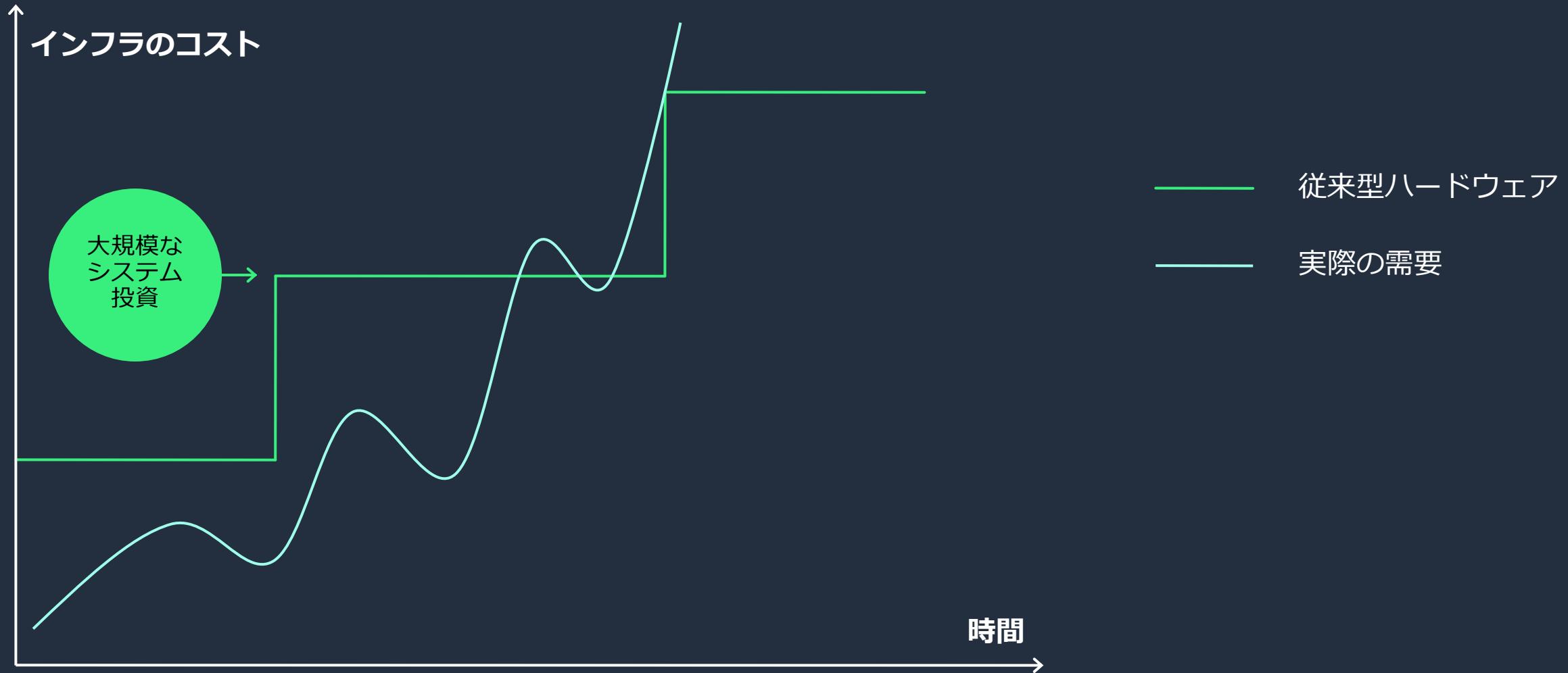
ビジネス需要に応じたキャパシティ準備



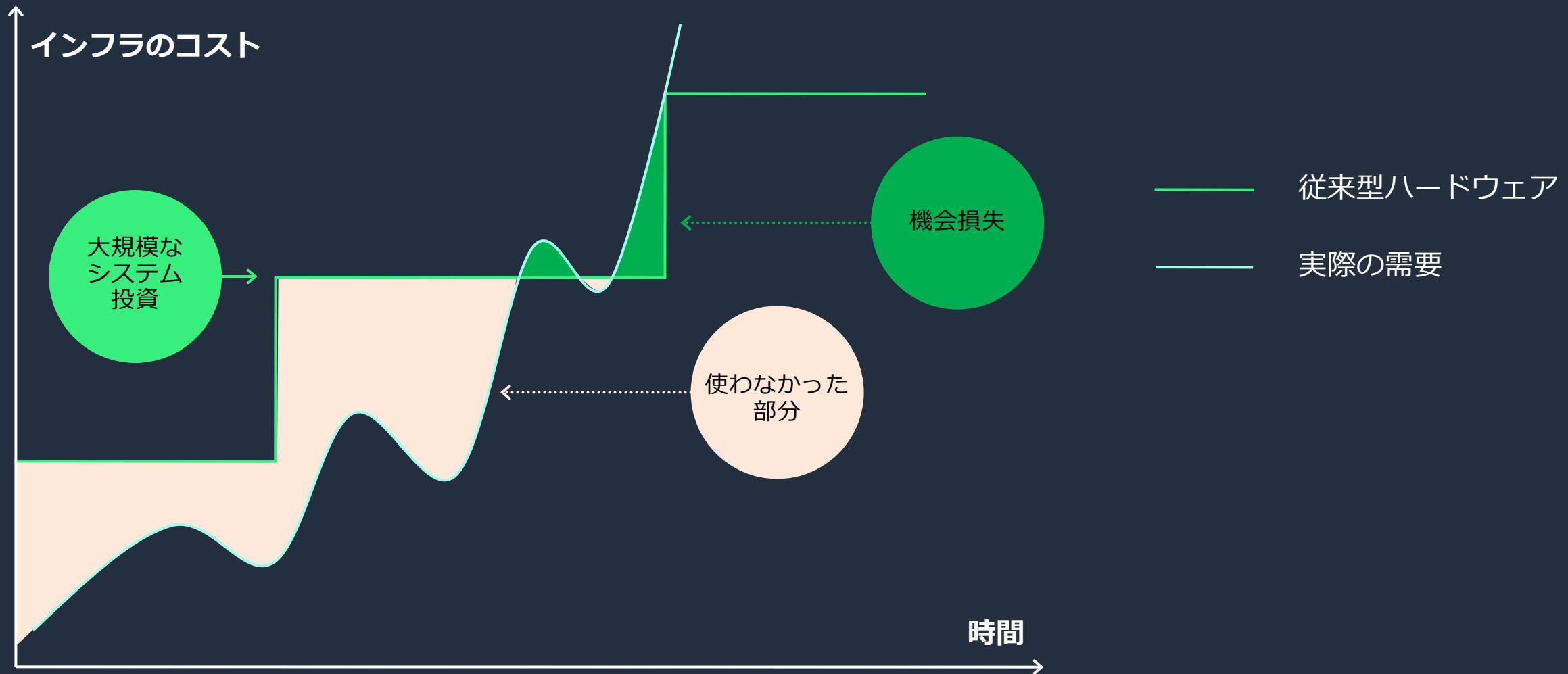
ビジネス需要に応じたキャパシティ準備



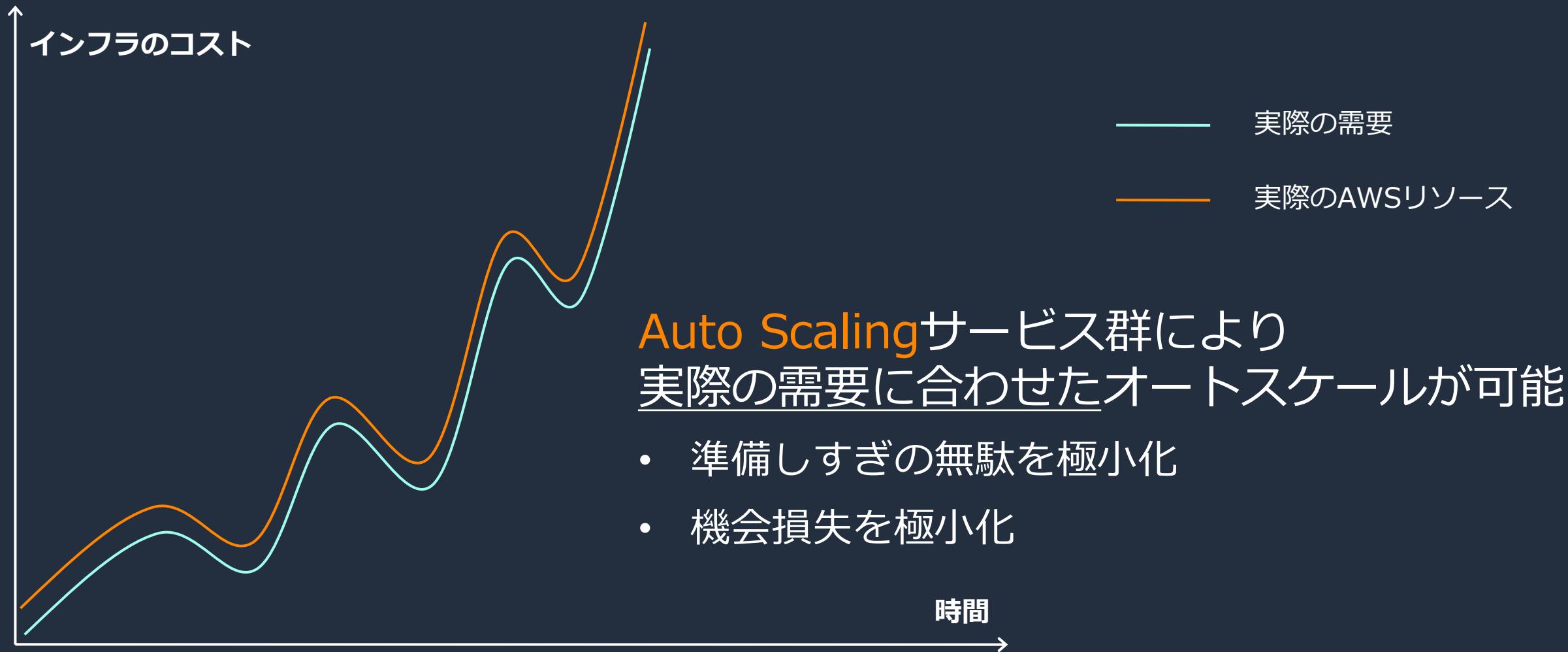
ビジネス需要に応じたキャパシティ準備



ビジネス需要に応じたキャパシティ準備



ビジネス需要に応じたキャパシティ準備



EC2 Auto Scaling の動作原理



EC2 Auto Scalingの動作原理

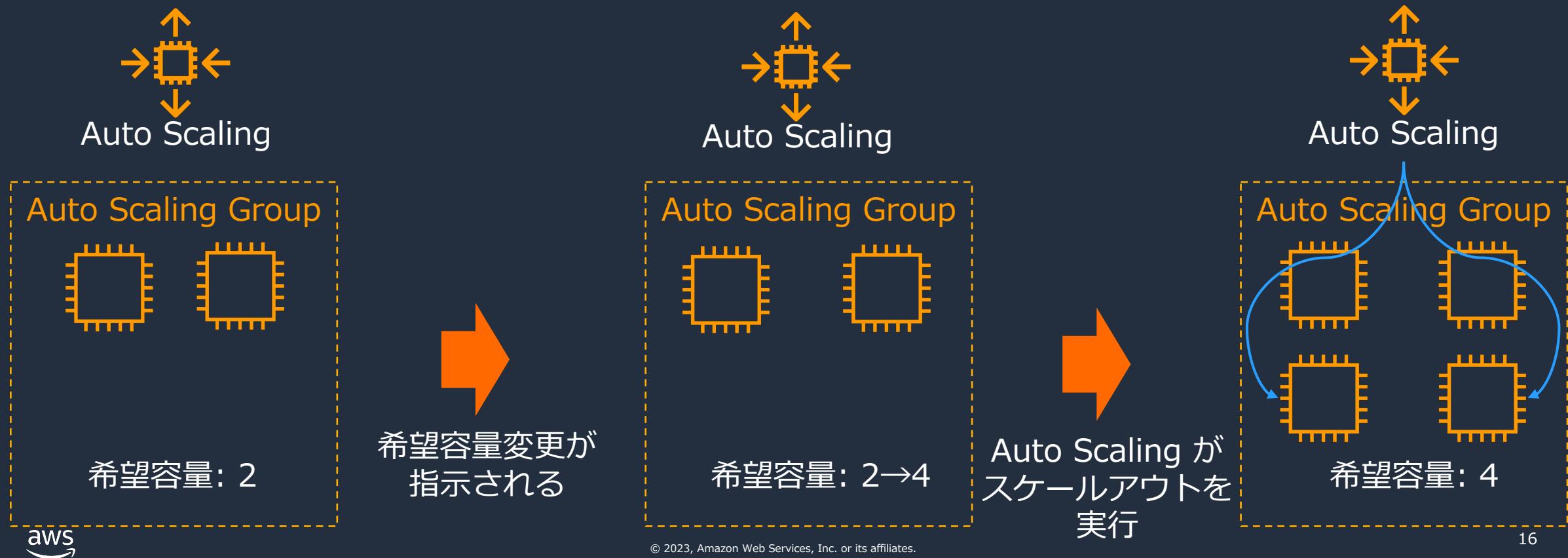
- 指定キャパシティの維持(台数維持) + 条件に応じた自動スケール
- インスタンスの分散(AZ間リバランス)

EC2 Auto Scalingの動作原理

- 指定キャパシティの維持(台数維持) + 条件に応じた自動スケール
- インスタンスの分散(AZ間リバランス)

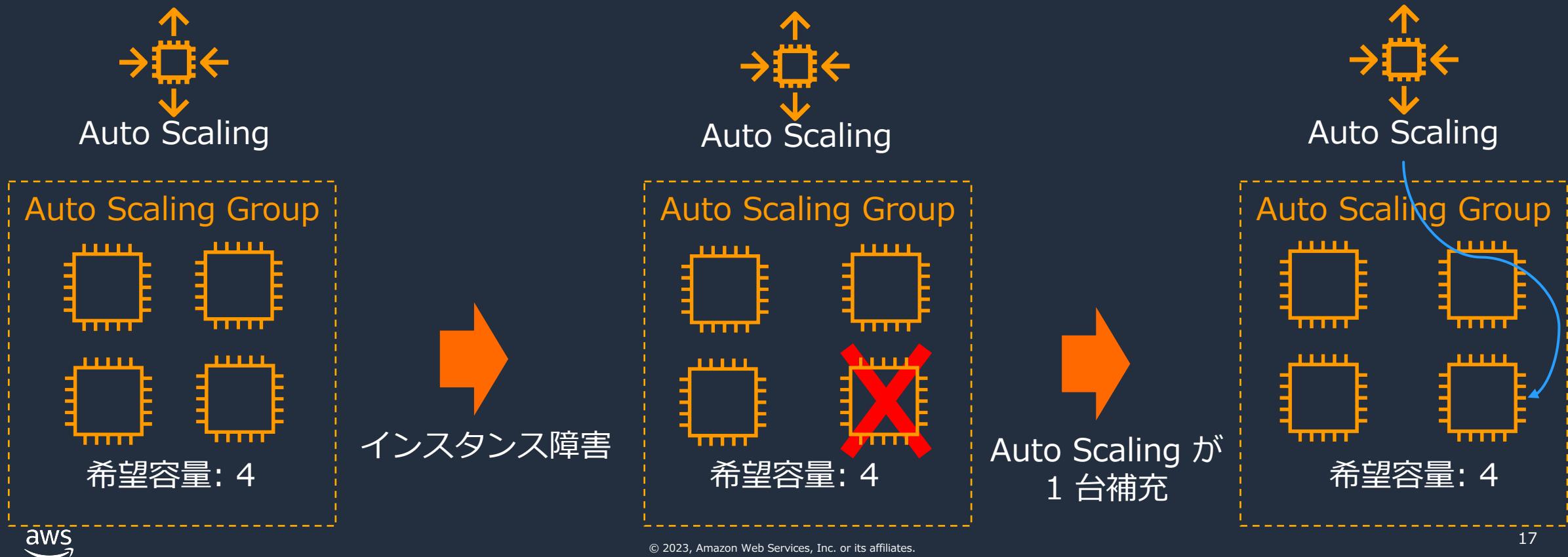
指定キャパシティの維持

EC2 Auto Scalingの動作原理(1)：指定された数の EC2 インスタンスを維持する



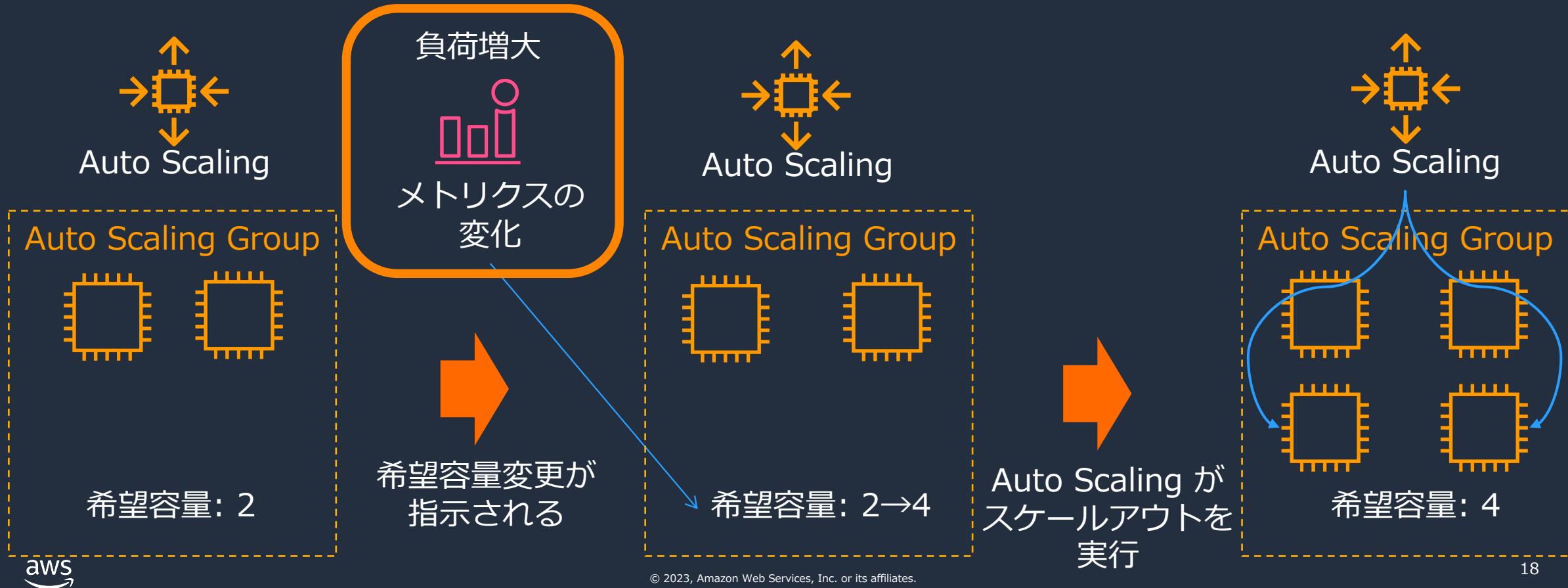
指定キャパシティの維持

EC2 Auto Scalingの動作原理(1)：指定された数の EC2 インスタンスを維持する



条件に応じた自動スケール

自動スケール：負荷に応じて自動的に増減してくれる仕組み

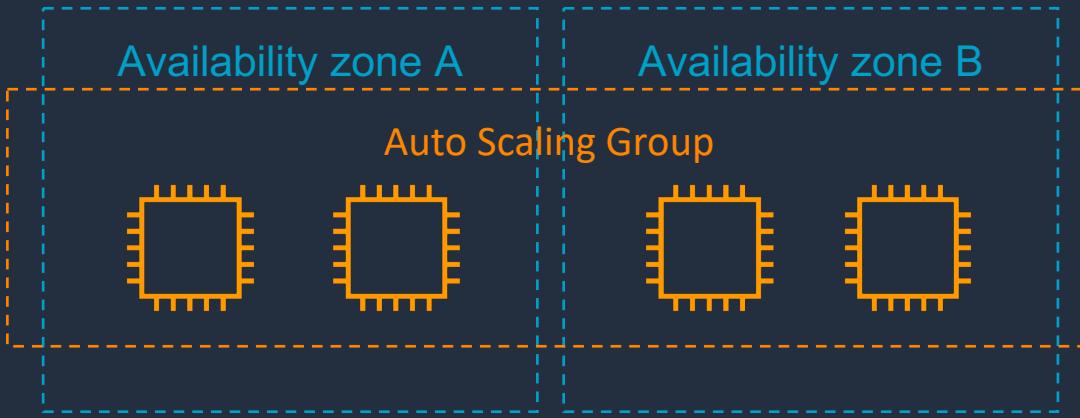


EC2 Auto Scalingの動作原理

- ・ 指定キャパシティの維持(台数維持) + 条件に応じた自動スケール
- ・ インスタンスの分散(AZ間リバランス)

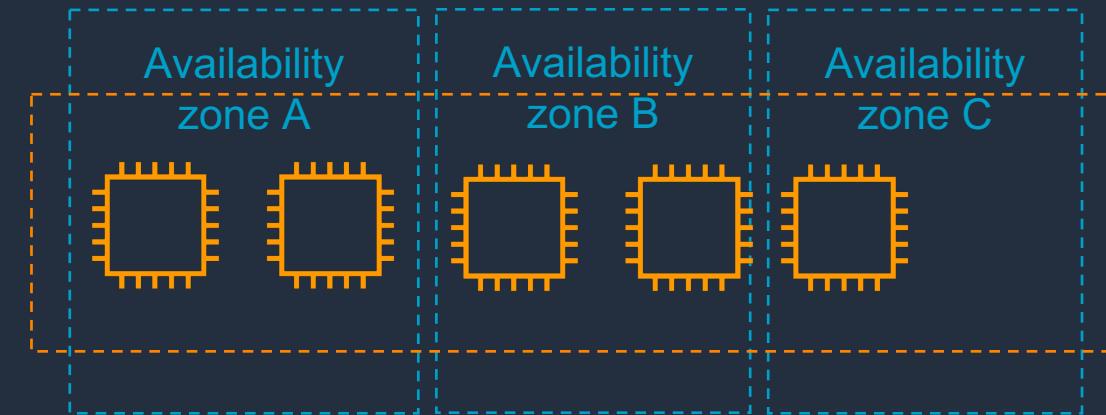
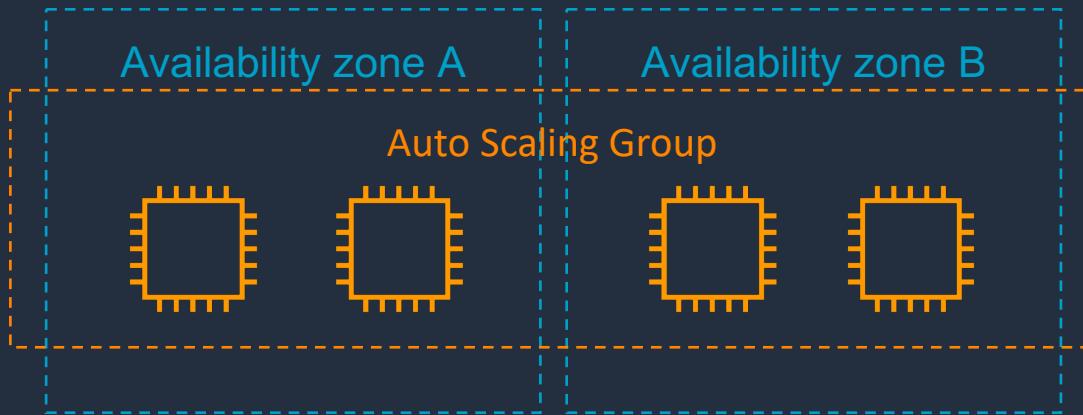
インスタンスの分散

EC2 Auto Scalingの動作原理(2)：使用できるアベイラビリティゾーンの間で均等にインスタンスを配置しようとする



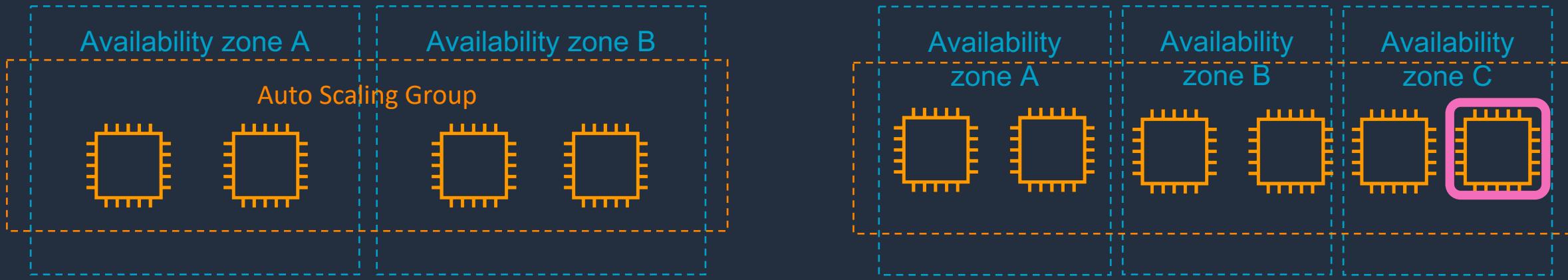
インスタンスの分散

EC2 Auto Scalingの動作原理(2)：使用できるアベイラビリティゾーンの間で均等にインスタンスを配置しようとする



インスタンスの分散

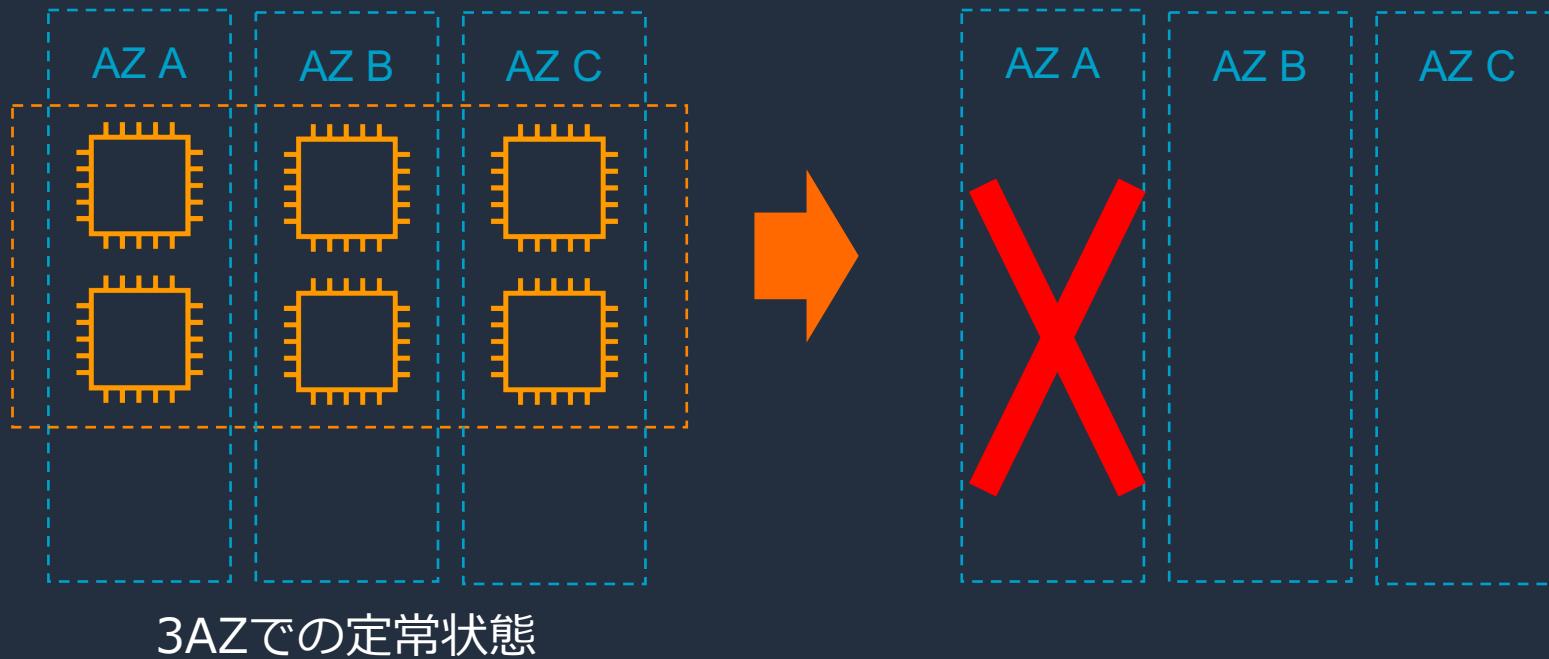
EC2 Auto Scalingの動作原理(2)：使用できるアベイラビリティゾーンの間で均等にインスタンスを配置しようとする



- スケールアウトするとき、インスタンス数が最も少ないアベイラビリティゾーンに新規起動
- これに失敗する場合、別のアベイラビリティゾーンが選択される

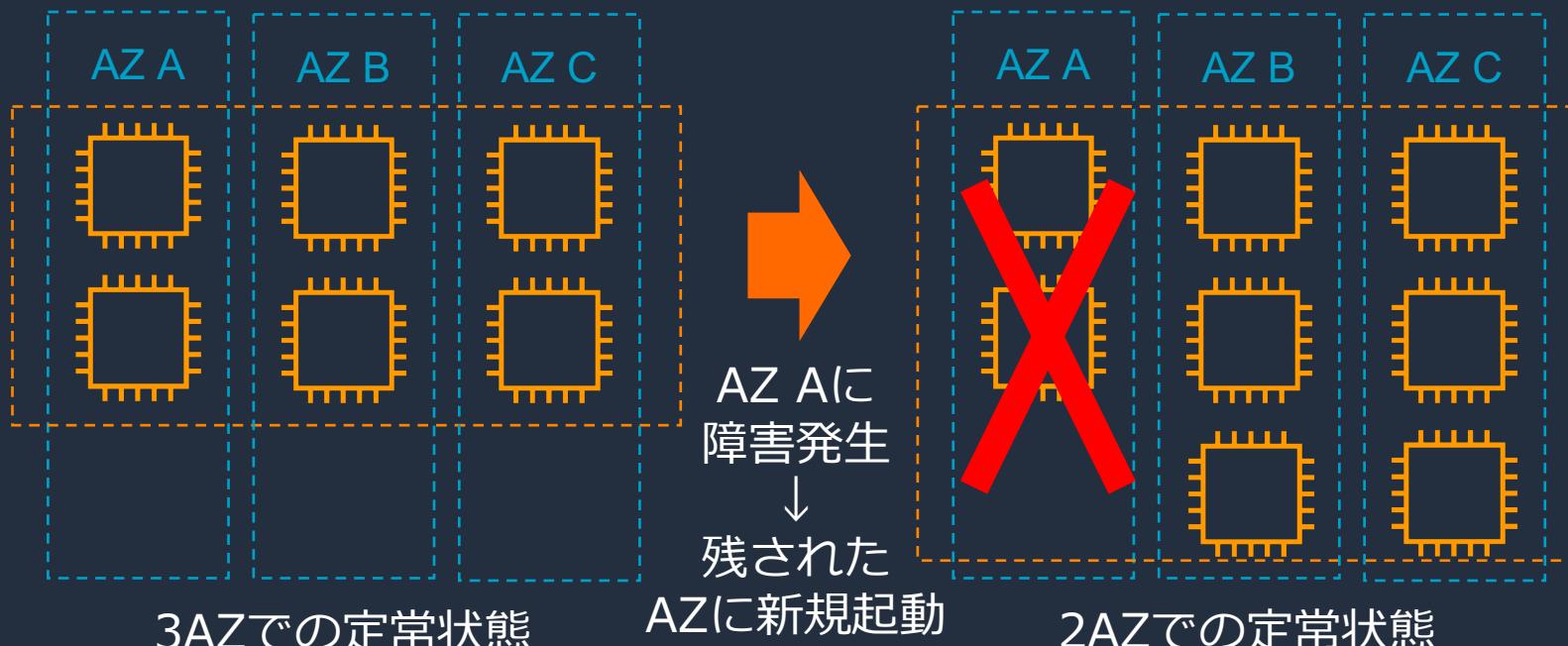
よくある質問：アベイラビリティゾーン (AZ) 障害時の動作

- 3AZ 構成で 6 台を起動しているとき、1AZ が障害になつたらどのような動作になるか？



よくある質問：アベイラビリティゾーン (AZ) 障害時の動作

- 3AZ 構成で 6 台を起動しているとき、1AZ が障害になつたらどのような動作になるか？
 - 残された AZ の間で均等に希望容量を維持する
 - サブネット定義が残っている限り、障害になつた AZ での起動をゆるやかに試みる

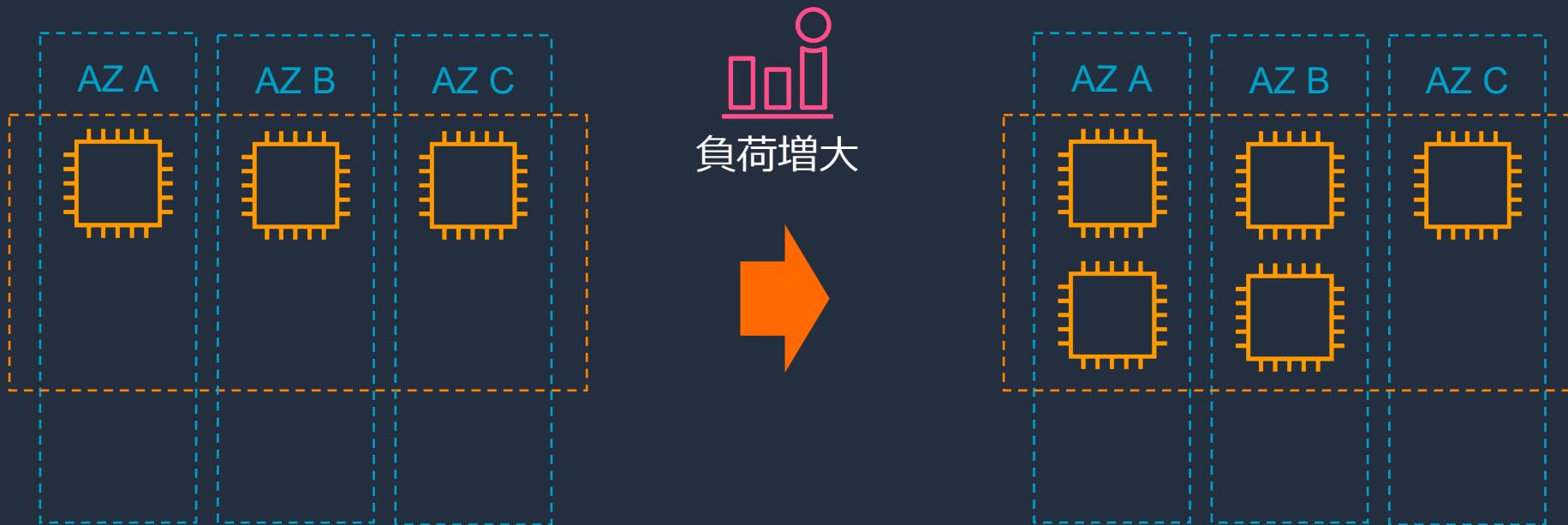


EC2 Auto Scaling を使ってみる



実現する構成のイメージ図

- ・3つのアベイラビリティーゾーン (AZ) 構成
- ・合計 3 台のオンデマンドインスタンスを起動する
- ・インスタンスタイプは t3.micro を選択する
- ・全体の CPU 使用率が 50% となるよう自動スケールさせる



EC2 Auto Scaling を使ってみる

- ・起動テンプレートの準備
- ・Auto Scaling グループの作成
- ・自動スケールの設定 (スケーリングポリシーの作成)

EC2 Auto Scaling を使ってみる

- 起動テンプレートの準備
- Auto Scaling グループの作成
- 自動スケールの設定 (スケーリングポリシーの作成)

起動テンプレートの準備 (1)

▼ インスタンス

インスタンス

インスタンスタイプ

起動テンプレート

スポットリクエスト

Savings Plans

リザーブドインスタンス

起動テンプレートを使用して、インスタンスの作成の自動化、アクセス権限ポリシーの簡素化、および組織全体のベストプラクティスの強化を行うことができます。起動パラメータをテンプレートに保存して、オンデマンドでの起動や、EC2 Auto Scaling や EC2 フリートなどのマネージドサービスで使用できます。新しい起動テンプレートのバージョンを作成することで、起動パラメータを簡単に更新できます。

新しい起動テンプレート

起動テンプレートを作成

- 「起動テンプレートを作成」を押す

起動テンプレートの準備 (2)

EC2 > 起動テンプレート > 起動テンプレートを作成

起動テンプレートを作成

起動テンプレートを作成することで、後で再利用、共有、起動できる保存済みインスタンス設定を作成できます。テンプレートには複数のバージョンを含めることができます。

起動テンプレート名と説明

起動テンプレート名 - 必須

このアカウントに固有である必要があります。最大 128 文字です。スペースや「&」、「*」、「@」などの特殊文字は使用できません。

テンプレートバージョンの説明

最大 255 文字

Auto Scaling のガイダンス 情報
EC2 Auto Scaling でこのテンプレートを使用する場合は、これを選択します
 EC2 Auto Scaling で使用できるテンプレートをセットアップする際に役立つガイダンスを提供

▶ テンプレートタグ
▶ ソーステンプレート

起動テンプレートのコンテンツ

起動テンプレートの詳細を以下で指定します。フィールドを空白のままにすると、フィールドが起動テンプレートに含まれません。

▼ アプリケーションおよび OS イメージ (Amazon マシンイメージ) **必須 情報**

AMI は、インスタンスの起動に必要なソフトウェア設定 (オペレーティングシステム、アプリケーションなど) を含むテンプレートです。お探しのものが以下に表示されない場合は、AMI を検索または参照してください。

- 起動テンプレートの名前を入力
- 「EC2 Auto Scaling で使用できるテンプレートをセットアップする際に役立つガイダンスを提供」にチェック
- これ以降の設定項目に「必須」表示が追加されるようになる



起動テンプレートの準備（3）



- AMIを選択

起動テンプレートの準備 (4)



- 「起動テンプレートを作成」を押す

EC2 Auto Scaling を使ってみる

- ・起動テンプレートの準備
- ・Auto Scaling グループの作成
- ・自動スケールの設定 (スケーリングポリシーの作成)

Auto Scaling グループの作成 (1)

The screenshot shows the AWS CloudFormation console with the following interface elements:

- Sidebar:** A navigation menu on the left side with sections: Elastic Block Store, ネットワーク & セキュリティ, ロードバランシング, and Auto Scaling. The "Auto Scaling" section is highlighted with a purple box, and "Auto Scaling グループ" is selected.
- Main Content Area:** A large central area with the heading "Amazon EC2 Auto Scaling は、アプリケーションの可用性を維持するために役立ちます". Below it is a description: "Auto Scaling グループは、オートスケーリングとフリート管理機能を可能にする Amazon EC2 インスタンスのコレクションです。これらの機能は、アプリケーションの正常性と可用性を維持するために役立ちます。".
- Call-to-Action:** A button labeled "Auto Scaling グループを作成する" located in a white box with a purple border.
- Side Panels:** Two panels on the right side:
 - 仕組み:** An icon showing two boxes connected by arrows, labeled "Auto Scaling group".
 - 料金:** A description of the costs associated with Auto Scaling, mentioning Amazon EC2, CloudWatch, and other AWS services.

- 「Auto Scaling グループを作成する」を押す

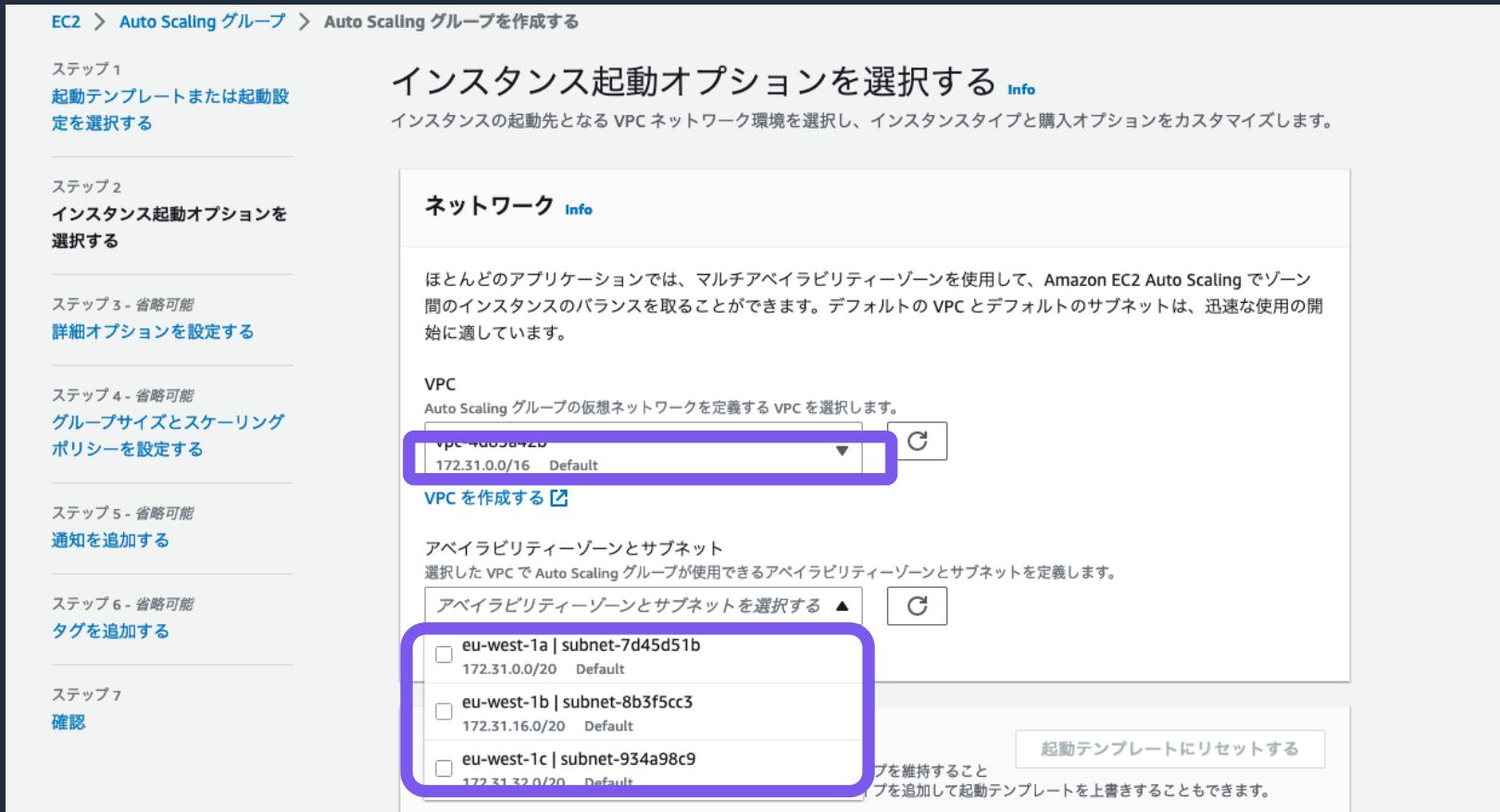
Auto Scaling グループの作成 (2)

The screenshot shows the 'Auto Scaling グループを作成する' (Create Auto Scaling Group) wizard. On the left, a sidebar lists steps 1 through 7. Step 1 is currently selected, titled '起動テンプレートまたは起動設定を選択する'. The main area displays the '起動テンプレートまたは起動設定を選択する' section. It includes a '名前' (Name) input field containing 'my_first_asg', which is highlighted with a purple rectangle. Below it is a note: '現在のリージョンにあるこのアカウントに固有で、255 文字以内にする必要があります。' (Must be unique within the account in the current region and must be 255 characters or less). To the right, there's a link '起動設定に切り替える' (Switch to launch configuration). At the bottom, there's a dropdown menu '起動テンプレートを選択する' with a search bar '起動テンプレートを検索する' and a list item 'my_first_launch_template'.

- Auto Scaling グループ名を入力
- 先ほど作成した起動テンプレート名を選択

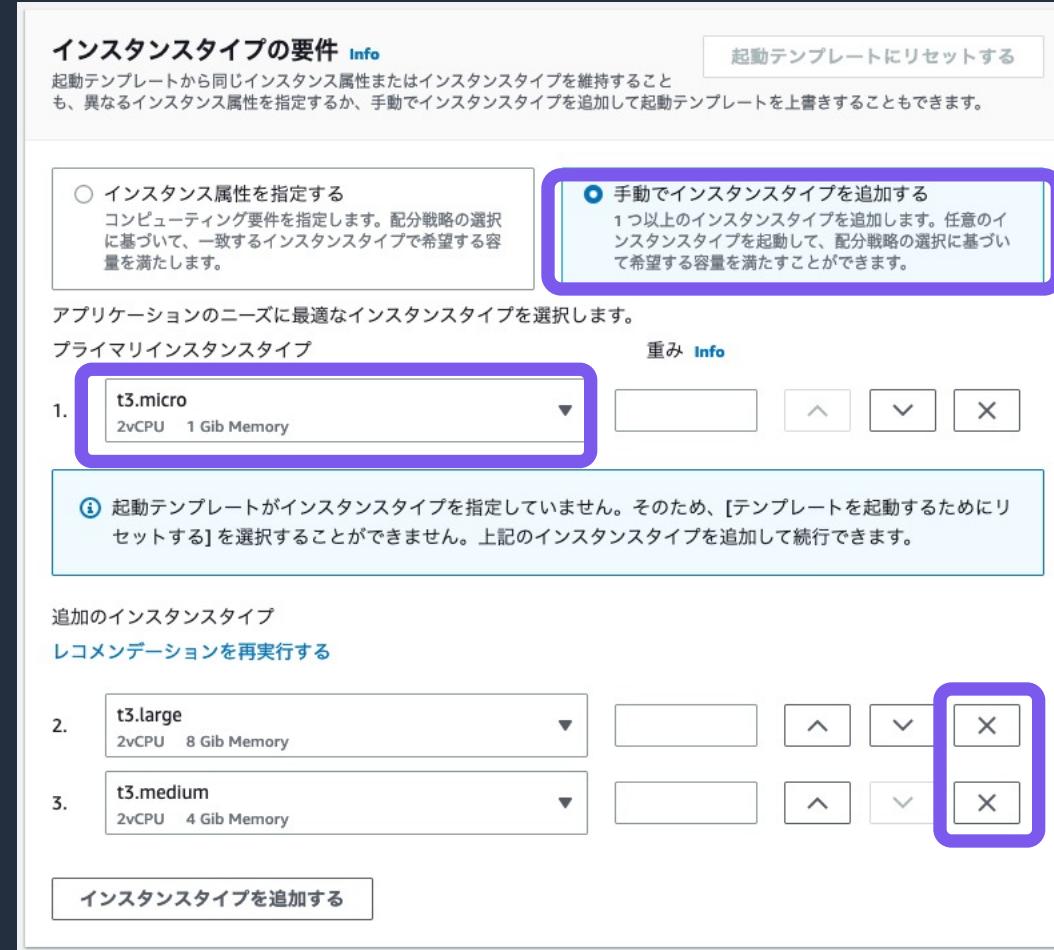
Auto Scaling グループの作成 (3)

ヒント
特別な理由がない限り
そのリージョンで使用
できるすべての
アベイラビリティー^{ゾーン}が含まれる
ようにしましょう



- VPC を選択
- その VPC 内に定義されたサブネットを選択

Auto Scaling グループの作成 (4)



ヒント

複数インスタンスタイプの活用について
は後続の「複数のインスタンスタイプと購入オプションの活用編」で解説します

- 「手動でインスタンスタイプを追加」を選択
- 要件に合ラインスタンスタイプを選択
- 自動推奨されるインスタンスタイプを使用しない場合は削除

Auto Scaling グループの作成 (5)

インスタンスの購入オプション [Info](#)

インスタンスの分散
耐障害性のあるワークロードを低成本で実行するには、スポットインスタンスとなるインスタンスの割合を定義します。スポットインスタンスは、AWS が 2 分前に通知することで変更できるオンデマンド料金に比べて大幅な割引を提供する予備の EC2 容量です。

100 % オンデマンド

0 % スポット

オンデマンドベース容量を含める
パーセンテージでスケールする前に、Auto Scaling グループがそのベース部分のために使用するオンデマンド容量を指定します。最大グループサイズはこの値まで増加します(減少することはできません)。

0 オンデマンドインスタンス

配分戦略 [Info](#)

オンデマンド配分戦略
オンデマンドインスタンスの起動時に適用する配分戦略を選択します。

高い優先順位で設定済み
上記で設定したインスタンスタイプの優先順位に基づいて、オンデマンドインスタンスをリクエストします。この戦略は、属性ベースのインスタンスタイプの選択では使用できません。

最低料金
アベイラビリティゾーン内で最低料金のプールからオンデマンドインスタンスをリクエストします。

① [インスタンスタイプの要件] のインスタンスタイプが 1 つのみであるため、このセクションは使用できません。スポットインスタンスまたはオンデマンドインスタンスで配分戦略を使用するには、少なくとも 2 つのインスタンスタイプまたは一連のインスタンス属性を指定する必要があります。

キャンセル スキップして確認 戻る 次へ

- 残りの項目はそのままにし、「次へ」を押す

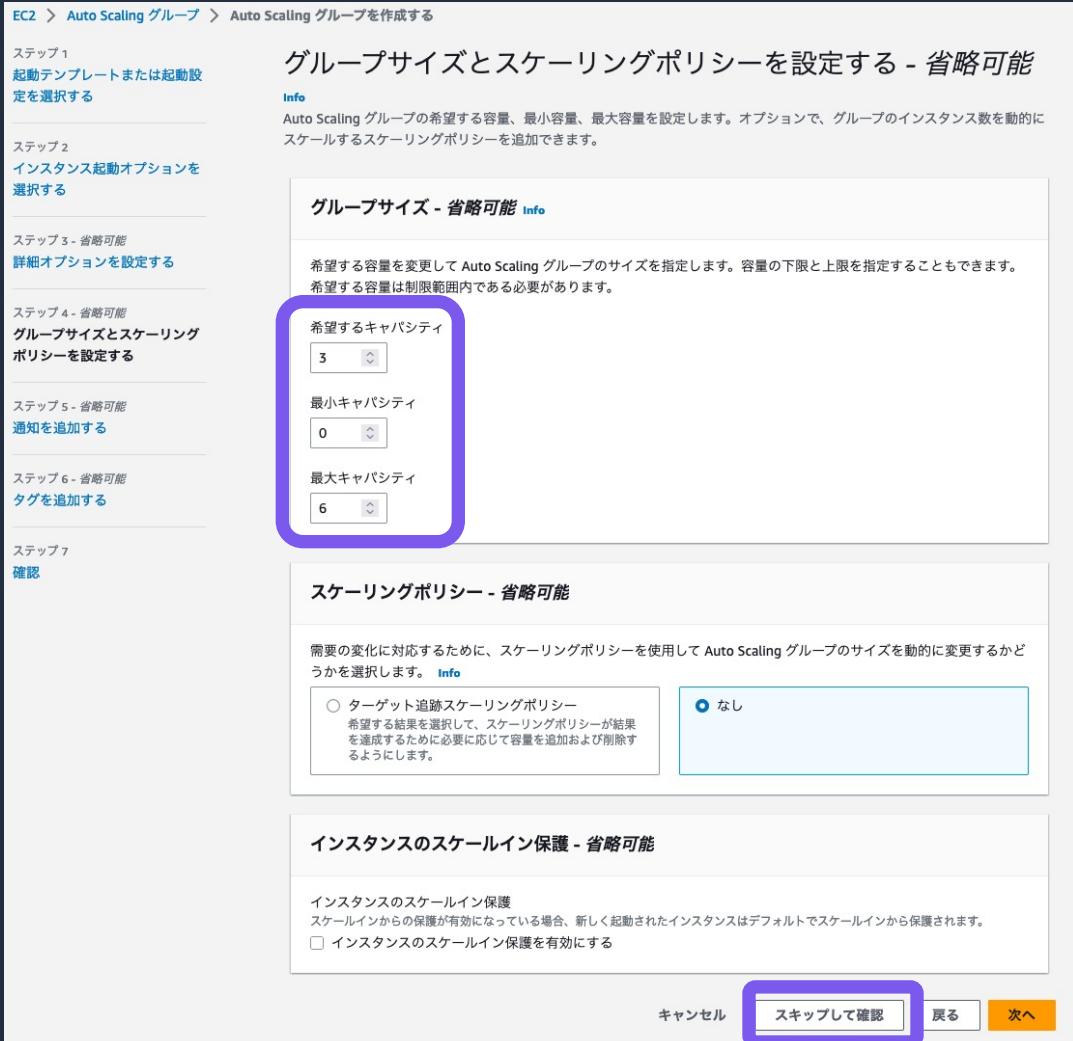
Auto Scaling グループの作成 (6)



- 残りの項目はそのままにし、「次へ」を押す

- グループメトリクスを有効にするのがオススメ。ASG内の台数などをCloudWatchから確認できる

Auto Scaling グループの作成 (7)



- 最初は 3 台起動する
- 0 - 6 台の間でスケールできるようにする

- 残りの項目はそのままにし、「スキップして確認」を押す

Auto Scaling グループの作成 (8)

EC2 > Auto Scaling グループ > Auto Scaling グループを作成する

確認 [Info](#)

ステップ 1
起動テンプレートまたは起動設定を選択する

ステップ 2
インスタンス起動オプションを選択する

ステップ 3 - 省略可能
[詳細オプションを設定する](#)

ステップ 4 - 省略可能
グループサイズとスケーリングポリシーを設定する

ステップ 5 - 省略可能
通知を追加する

ステップ 6 - 省略可能
[タグを追加する](#)

ステップ 7
確認

確認 Info

ステップ 1: 起動テンプレートまたは起動設定を選択する

編集

グループの詳細

Auto Scaling グループ名
my_first_asg

起動テンプレート

起動テンプレート	バージョン	説明
my_first_launch_template	Default	lt-0e5c87a9fe91678ef

ステップ 2: インスタンスの起動オプションを選択する

編集

ネットワーク

ネットワーク

ステップ 6: タグを追加する

編集

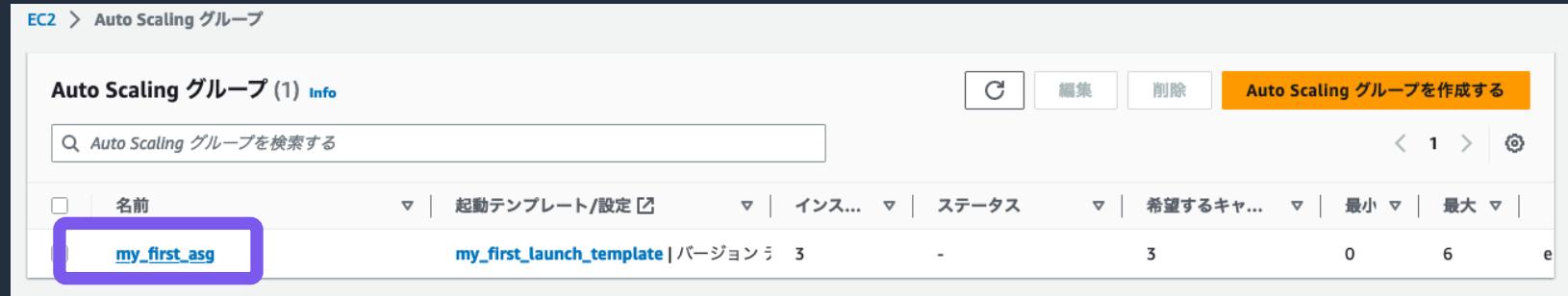
タグ (0)

キー	▼	値	▼	新しいインスタンスをタグ付けする	▼
タグなし					

キャンセル 戻る [Auto Scaling グループを作成する](#)

- 内容を確認し、「Auto Scaling グループを作成する」を押す

Auto Scaling グループの作成確認と詳細画面への遷移



The screenshot shows the AWS Auto Scaling Groups page. At the top, there is a breadcrumb navigation: EC2 > Auto Scaling グループ. Below the breadcrumb, a search bar contains the placeholder text "Auto Scaling グループを検索する". To the right of the search bar are buttons for "C" (Create), "編集" (Edit), "削除" (Delete), and "Auto Scaling グループを作成する" (Create Auto Scaling Group). A yellow box highlights the "Auto Scaling グループを作成する" button. Below these buttons is a pagination control with arrows and the number "1".

The main table lists one Auto Scaling group:

名前	起動テンプレート/設定	インス...	ステータス	希望するキャ...	最小	最大
my_first_asg	my_first_launch_template バージョン 3	-	3	0	6	

- 作成された Auto Scaling グループ名をクリック

EC2 インスタンス起動状況の確認

The screenshot shows the AWS Auto Scaling Groups console. The navigation path is EC2 > Auto Scaling グループ > my_first_asg. The main tab 'my_first_asg' is selected. Below it, the 'Activities' tab is highlighted with a purple rectangle. The 'Activities' section displays the following details:

グループの詳細		
Auto Scaling グループ名 my_first_asg	希望するキャパシティ 3	ステータス -
作成日 Sun Mar 19 2023 23:13:48 GMT+0900 (Japan Standard Time)	最小キャパシティ 0	
	最大キャパシティ 6	

Below this, there is a section titled '起動テンプレート'.

- 「アクティビティ」をクリック

EC2 インスタンス起動状況の確認

The screenshot shows the AWS Auto Scaling Groups console for the group 'my_first_asg'. The 'Activity' tab is selected. The 'Successful' status column is highlighted with a purple box. The table lists three successful launches of EC2 instances:

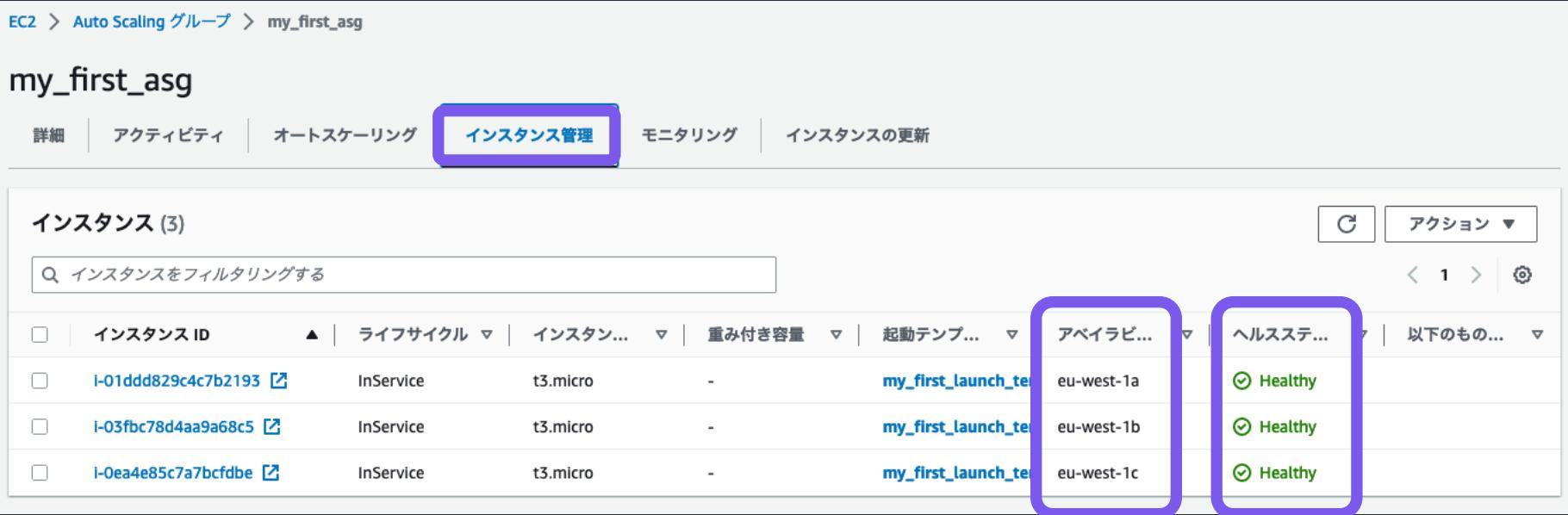
ステータス	説明	原因	開始時刻	終了時刻
Successful	Launching a new EC2 Instance: i-0ea4e85c7a7bcfdbe	At 2023-03-19T14:13:48Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 3. At 2023-03-19T14:13:50Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 3.	2023 March 19, 11:13:54 PM +09:00	2023 March 19, 11:14:00 PM +09:00
Successful	Launching a new EC2 Instance: i-01ddd829c4c7b2193	At 2023-03-19T14:13:48Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 3. At 2023-03-19T14:13:50Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 3.	2023 March 19, 11:13:54 PM +09:00	2023 March 19, 11:14:00 PM +09:00
Successful	Launching a new EC2 Instance: i-03fb78d4aa9a68c5	At 2023-03-19T14:13:48Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 3. At 2023-03-19T14:13:50Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 3.	2023 March 19, 11:13:54 PM +09:00	2023 March 19, 11:14:00 PM +09:00

ヒント

Auto Scaling グループのトラブルシューティングには「アクティビティ」タブから動作記録を確認しましょう

- ステータス列を確認
- エラーが発生している場合、Auto Scaling ドキュメントのトラブルシューティング節を参照して対処する
 - https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/CHAP_Troubleshooting.html

EC2 インスタンス起動状況の確認



The screenshot shows the AWS EC2 Auto Scaling Groups console for the group 'my_first_asg'. The 'Instances Management' tab is selected. The table displays three instances:

インスタンス ID	ライフサイクル	インスタンス型	重み付き容量	起動テンプレート	アベイラビリティゾーン	ヘルステータス
i-01ddd829c4c7b2193	InService	t3.micro	-	my_first_launch_template	eu-west-1a	Healthy
i-03fbc78d4aa9a68c5	InService	t3.micro	-	my_first_launch_template	eu-west-1b	Healthy
i-0ea4e85c7a7bcfdbe	InService	t3.micro	-	my_first_launch_template	eu-west-1c	Healthy

Three instances are highlighted with purple boxes: eu-west-1a, eu-west-1b, and eu-west-1c. All three instances are listed as 'Healthy'.

- それぞれのインスタンスがどのアベイラビリティーゾーンに起動されたかを確認
- ヘルステータスが Healthy であることを確認

EC2 Auto Scaling を使ってみる

- ・起動テンプレートの準備
- ・Auto Scaling グループの作成
- ・自動スケールの設定 (スケーリングポリシーの作成)

スケーリングポリシーの作成



- 「動的スケーリングポリシーを作成する」を押す

スケーリングポリシーの作成 - ターゲット追跡スケーリング

EC2 > Auto Scaling グループ > my_first_asg

動的スケーリングポリシーを作成する

ポリシータイプ
ターゲット追跡スケーリング

スケーリングポリシー名
Target Tracking Policy

メトリクスタイプ
平均 CPU 使用率

ターゲット値
50

インスタンスには以下のものが必要です
300 メトリクスに含める前にウォームアップする秒数

スケールインを無効にしてスケールアウトポリシーのみを作成する

キャンセル 作成

スケーリングポリシーの作成 - ターゲット追跡スケーリング

EC2 > Auto Scaling グループ > my_first_asg

動的スケーリングポリシーを作成する

ポリシータイプ
ターゲット追跡スケーリング

スケーリングポリシー名
Target Tracking Policy

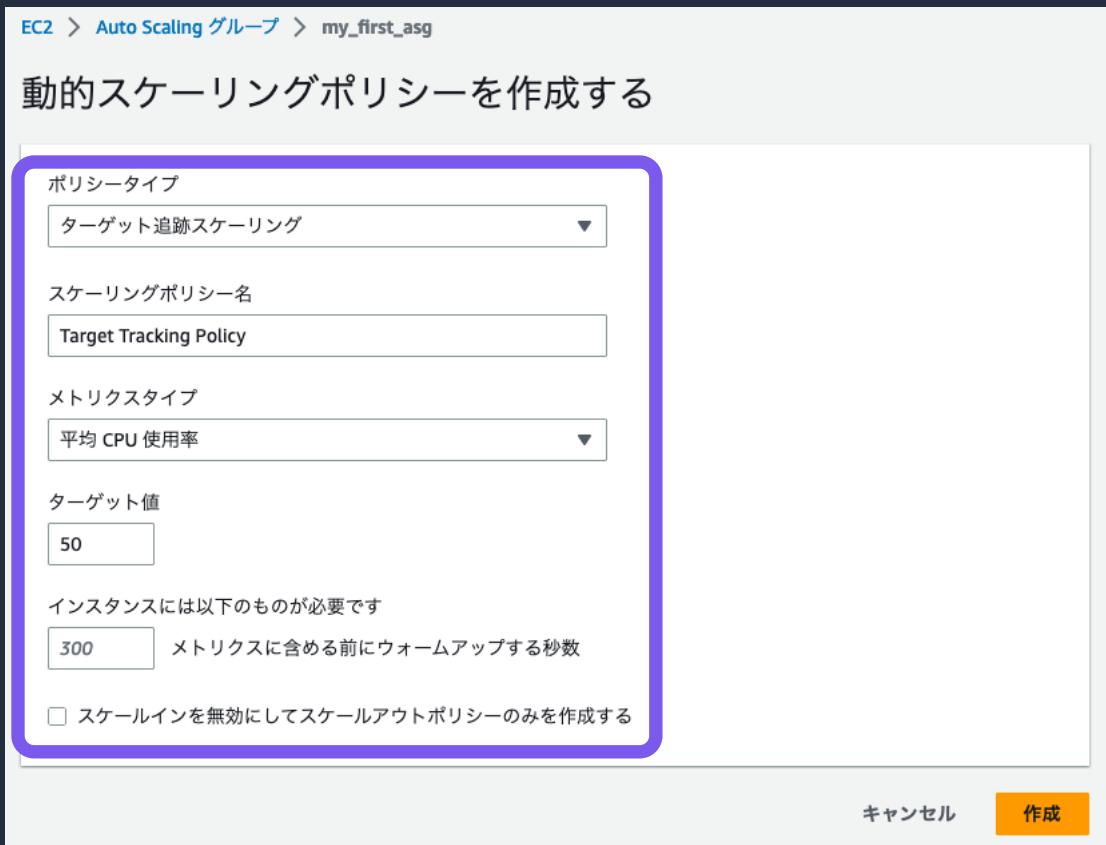
メトリクスタイプ
平均 CPU 使用率

ターゲット値
50

インスタンスには以下のものが必要です
300 メトリクスに含める前にウォームアップする秒数

スケールインを無効にしてスケールアウトポリシーのみを作成する

キャンセル 作成



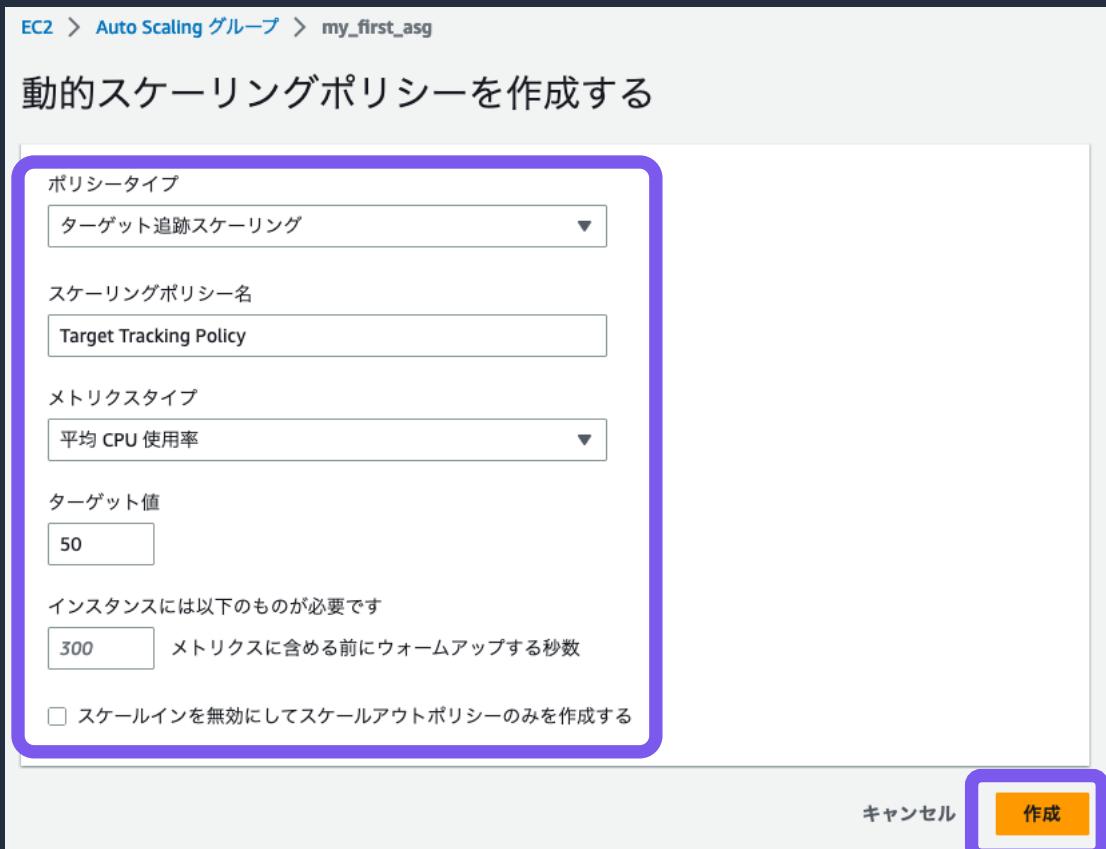
ヒント

スケーリングポリシーについては後続の「スケーリング編」で解説します



- この Auto Scaling グループではCPU 使用率を 50% に保ってほしい、と指定
- 自動スケールのためのCW Alarmが2本作られる
 - 負荷が上がり、50%を超えた期間がしばらく続くとスケールアウト（台数増加）
 - 負荷が下がり、50%以下の期間がある程度続くとスケールイン（台数減少）
- https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/as-scaling-target-tracking.html

スケーリングポリシーの作成 - ターゲット追跡スケーリング

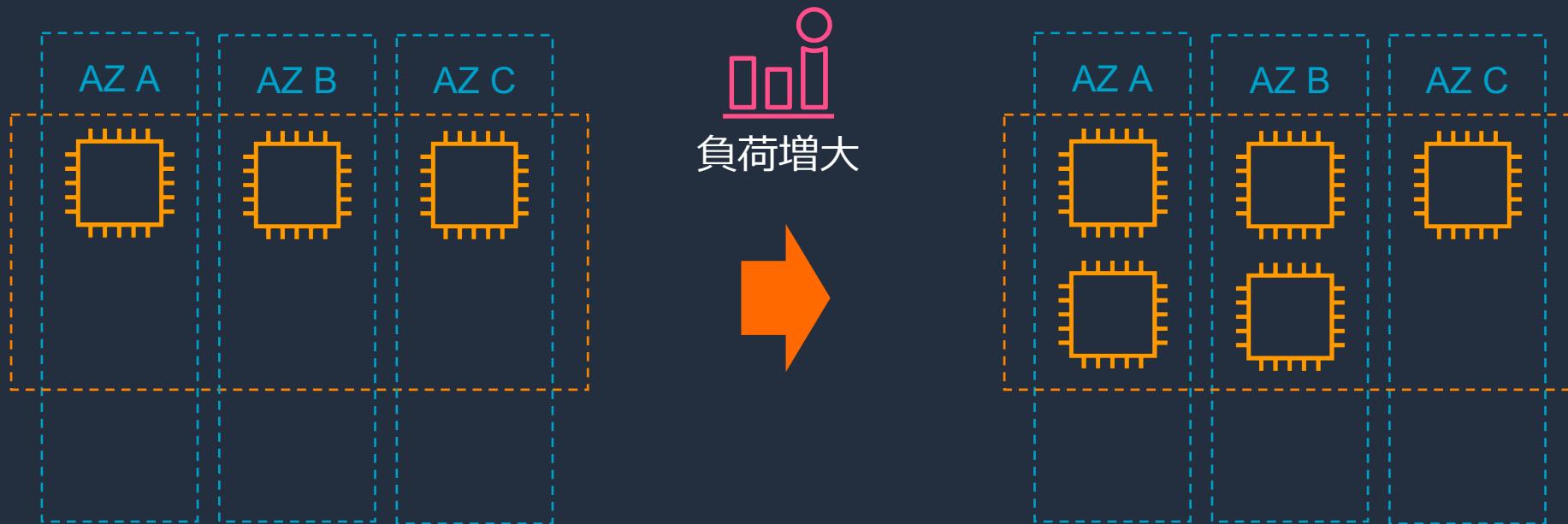


- この Auto Scaling グループではCPU 使用率を 50% に保ってほしい、と指定
- 自動スケールのためのCW Alarmが2本作られる
 - 負荷が上がり、50%を超えた期間がしばらく続くとスケールアウト（台数増加）
 - 負荷が下がり、50%以下の期間がある程度続くとスケールイン（台数減少）
- https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/as-scaling-target-tracking.html

- 内容を確認し、「作成」を押す

実現する構成のイメージ図 (再掲)

- ・3つのアベイラビリティーゾーン (AZ) 構成
- ・合計 3 台のオンデマンドインスタンスを起動する
- ・インスタンスタイプは t3.micro を選択する
- ・全体の CPU 使用率が 50% となるよう自動スケールさせる



おわりに



今回お話しした内容

- Auto Scaling サービスのコンセプト
- EC2 Auto Scaling の動作原理
- EC2 Auto Scaling を使ってみる
- 実際の負荷をかけるテストを体験できます。「AWS Hands-on for Beginners - Amazon EC2 Auto Scaling スケーリング基礎編」をご覧ください
https://pages.awscloud.com/JAPAN-event-OE-Hands-on-for-Beginners-Auto_Scaling-2022-reg-event.html

本資料に関するお問い合わせ・ご感想

技術的な内容に関しましては、有料のAWSサポート窓口へ
お問い合わせください

<https://aws.amazon.com/jp/premiumsupport/>

料金面でのお問い合わせに関しましては、カスタマーサポート窓口へ
お問い合わせください（マネジメントコンソールへのログインが必要です）

<https://console.aws.amazon.com/support/home#/case/create?issueType=customer-service>

具体的な案件に対する構成相談は、後述する個別相談会をご活用ください



ご感想はTwitterへ！ハッシュタグは以下をご利用ください
#awsblackbelt



その他コンテンツのご紹介

ウェビナーなど、AWSのイベントスケジュールをご参照いただけます

<https://aws.amazon.com/jp/events/>

ハンズオンコンテンツ

<https://aws.amazon.com/jp/aws-jp-introduction/aws-jp-webinar-hands-on/>

AWS 個別相談会

AWSのソリューションアーキテクトと直接会話いただけます

<https://pages.awscloud.com/JAPAN-event-SP-Weekly-Sales-Consulting-Seminar-2021-reg-event.html>



Thank you!