



Amazon ECS Deep Dive

~ Capacity Providers ~

アマゾン ウェブ サービス ジャパン株式会社
落水 恭介

2021.09.07



本日は話すこと

- Amazon ECS での Auto Scaling とは
 - コンテナインスタンス (EC2 インスタンスレイヤ)
 - ECS サービス (タスクレイヤ)
- ECS Cluster Auto Scaling
 - ECS Cluster Auto Scaling とは
 - Capacity Providers の概要

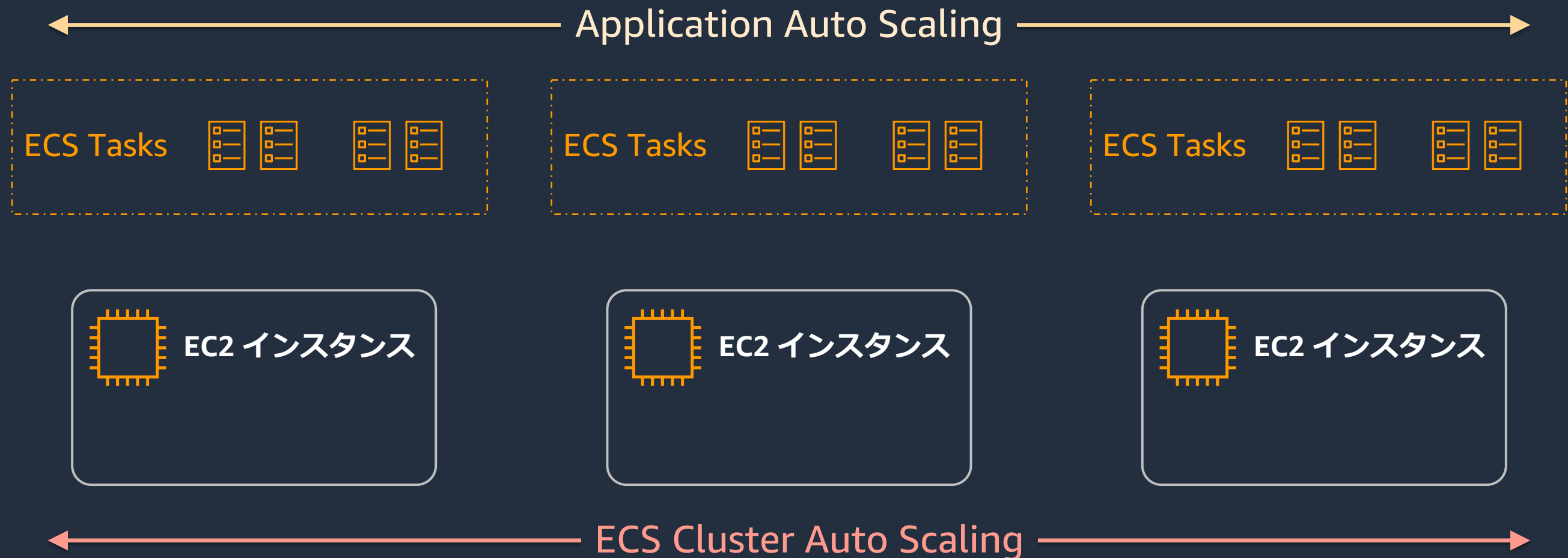
ECS Cluster Auto Scaling



ECS Cluster Auto Scaling とは

ECS が Auto Scaling グループのスケールインおよびスケールアウトアクションを管理

➤ **Capacity Providers** を利用することで実現



Capacity Providers とは

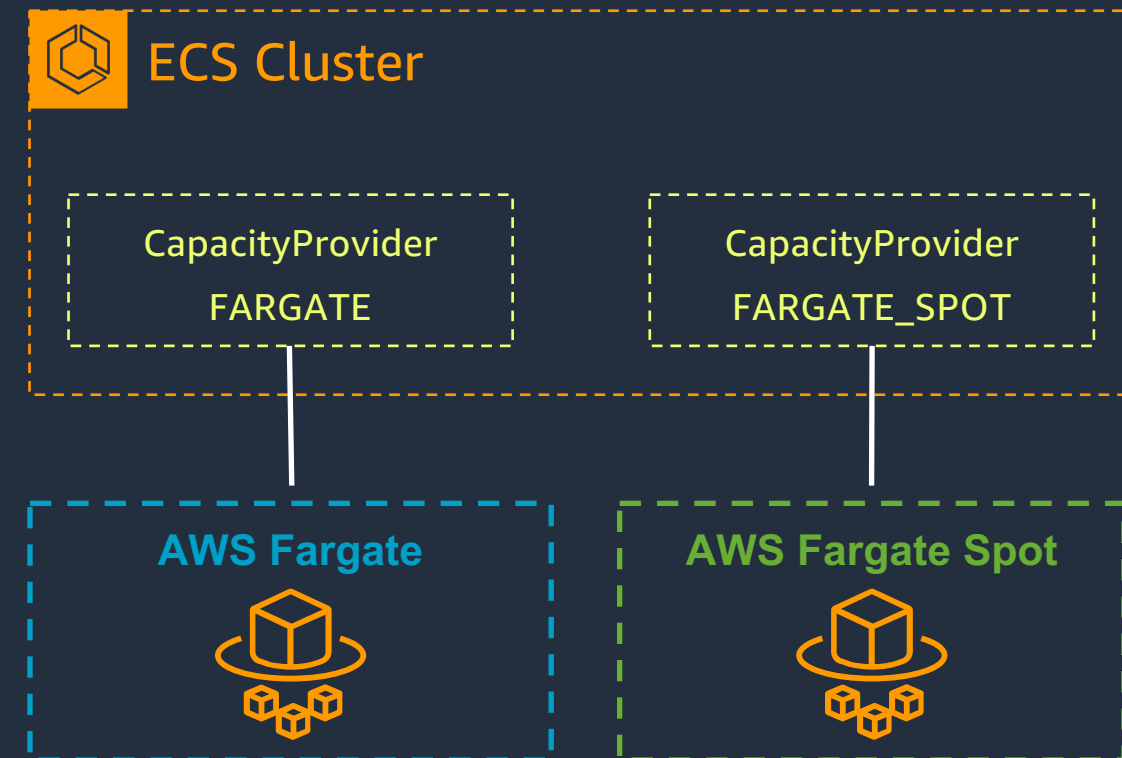
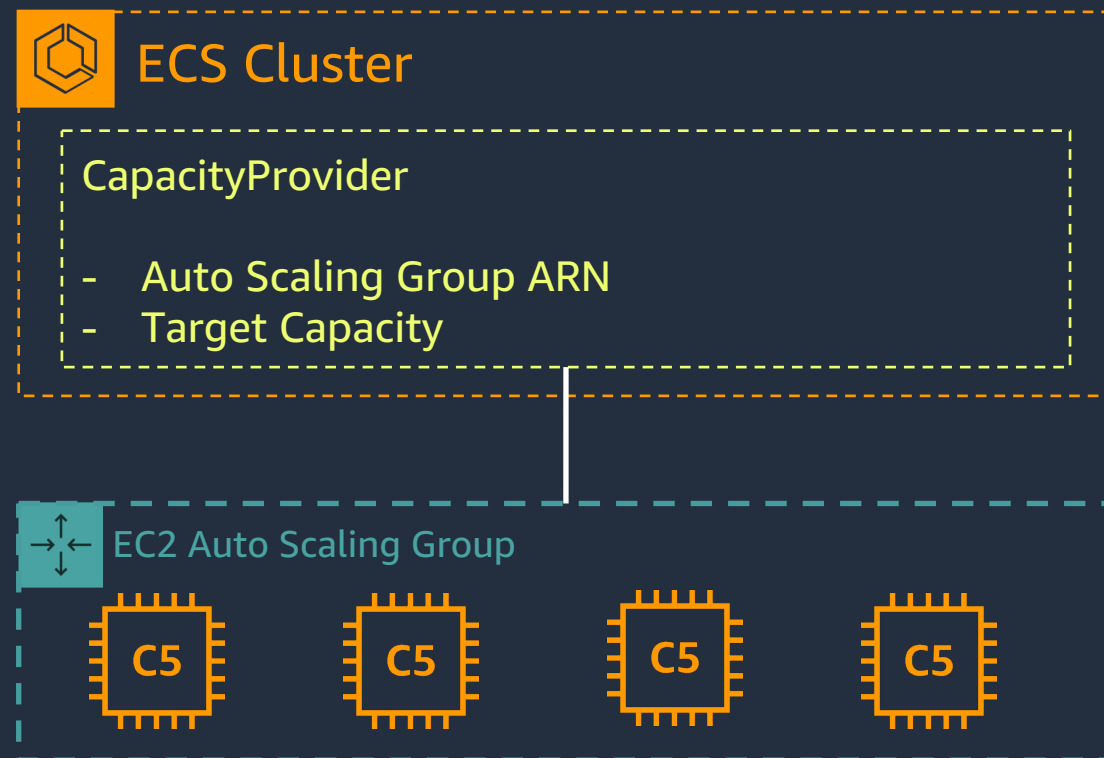
- ECS クラスタ内のタスクで使用するインフラストラクチャを管理
- 以下のような概念、機能を持つ
 - Capacity Provider
 - Capacity Provider Strategy
 - Capacity Provider Reservation
 - Managed Scaling
 - Managed Termination Protection

Capacity Providers とは

- ECS クラスタ内のタスクで使用するインフラストラクチャを管理
- 以下のような概念、機能を持つ
 - **Capacity Provider**
 - Capacity Provider Strategy
 - Capacity Provider Reservation
 - Managed Scaling
 - Managed Termination Protection

Capacity Provider

- ECS とコンテナ実行基盤 (Auto Scaling グループ / Fargate) のインターフェイス
- Auto Scaling グループや Fargate を 1 つの Capacity Provider として定義
 - 定義した Capacity Provider を ECS クラスタに紐付けて利用
- Fargate の Capacity Provider (FARGATE / FARGATE_SPOT) は事前定義済み



Capacity Providers とは

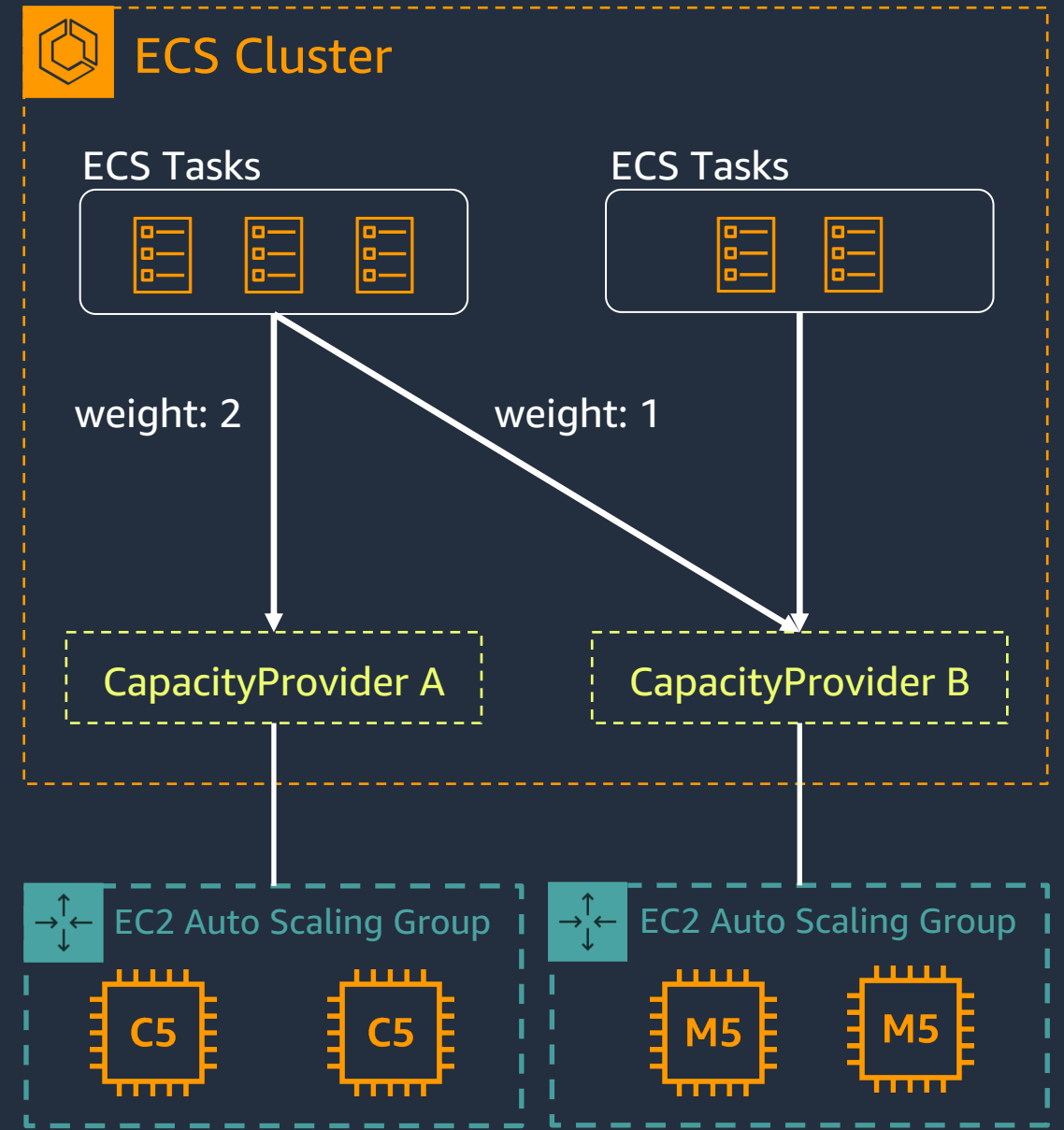
- ECS クラスタ内のタスクで使用するインフラストラクチャを管理
- 以下のような概念、機能を持つ
 - Capacity Provider
 - **Capacity Provider Strategy**
 - Capacity Provider Reservation
 - Managed Scaling
 - Managed Termination Protection

Capacity Provider Strategy

- タスクをどの Capacity Provider に配置するかを決定する「戦略」
- タスクの実行や ECS サービスを作成する際に指定する
- base と weight を使用して、複数の Capacity Provider にまたがって配置することも可能

Default Capacity Provider Strategy

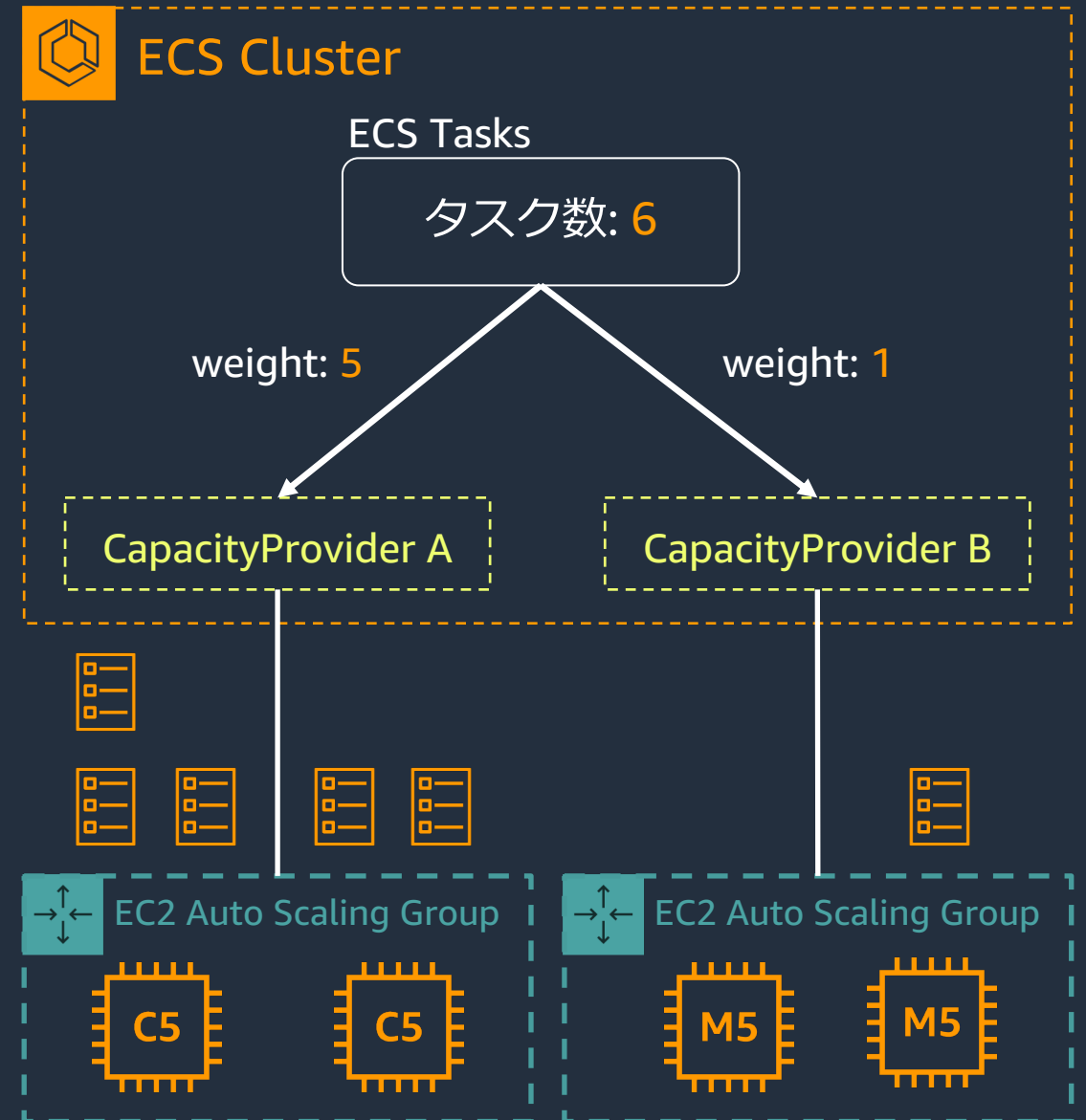
- ECS クラスタがデフォルトで利用する Capacity Provider Strategy
- Capacity Provider Strategy を指定せずにタスクの実行や ECS サービスを作成した場合に利用される



Capacity Provider Strategy

- **weight**
 - 実行するタスクの総数に対する指定した Capacity Provider での相対的な割合
- **base**
 - その Capacity Provider での最小実行タスク数

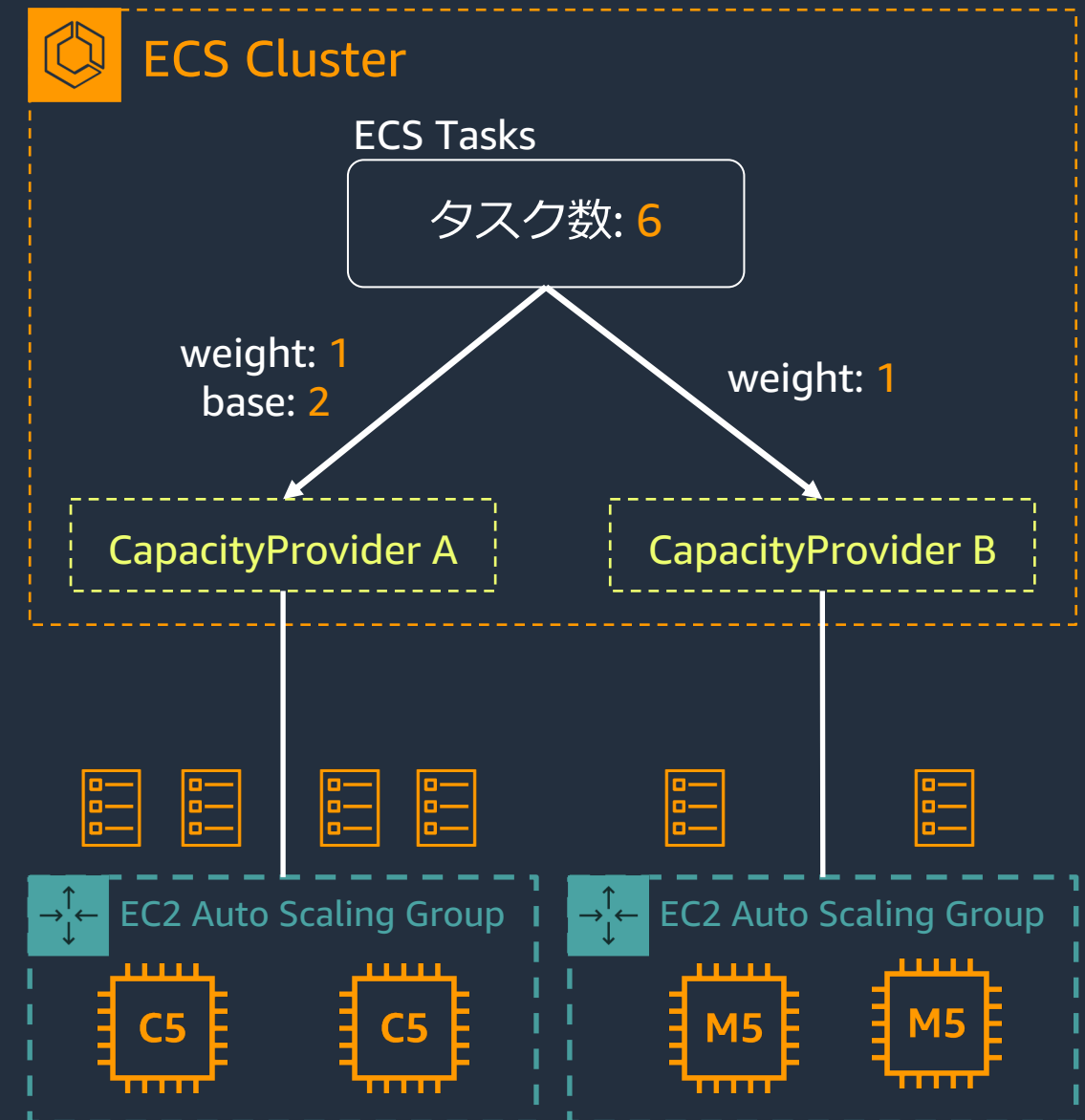
```
$ aws ecs run-task  
  --cluster MyEcsCluster  
  --task-definition task-def-family:revision  
  --count 6  
  --capacity-provider-strategy  
    capacityProvider=CapacityProviderA,weight=5  
    capacityProvider=CapacityProviderB,weight=1
```



Capacity Provider Strategy

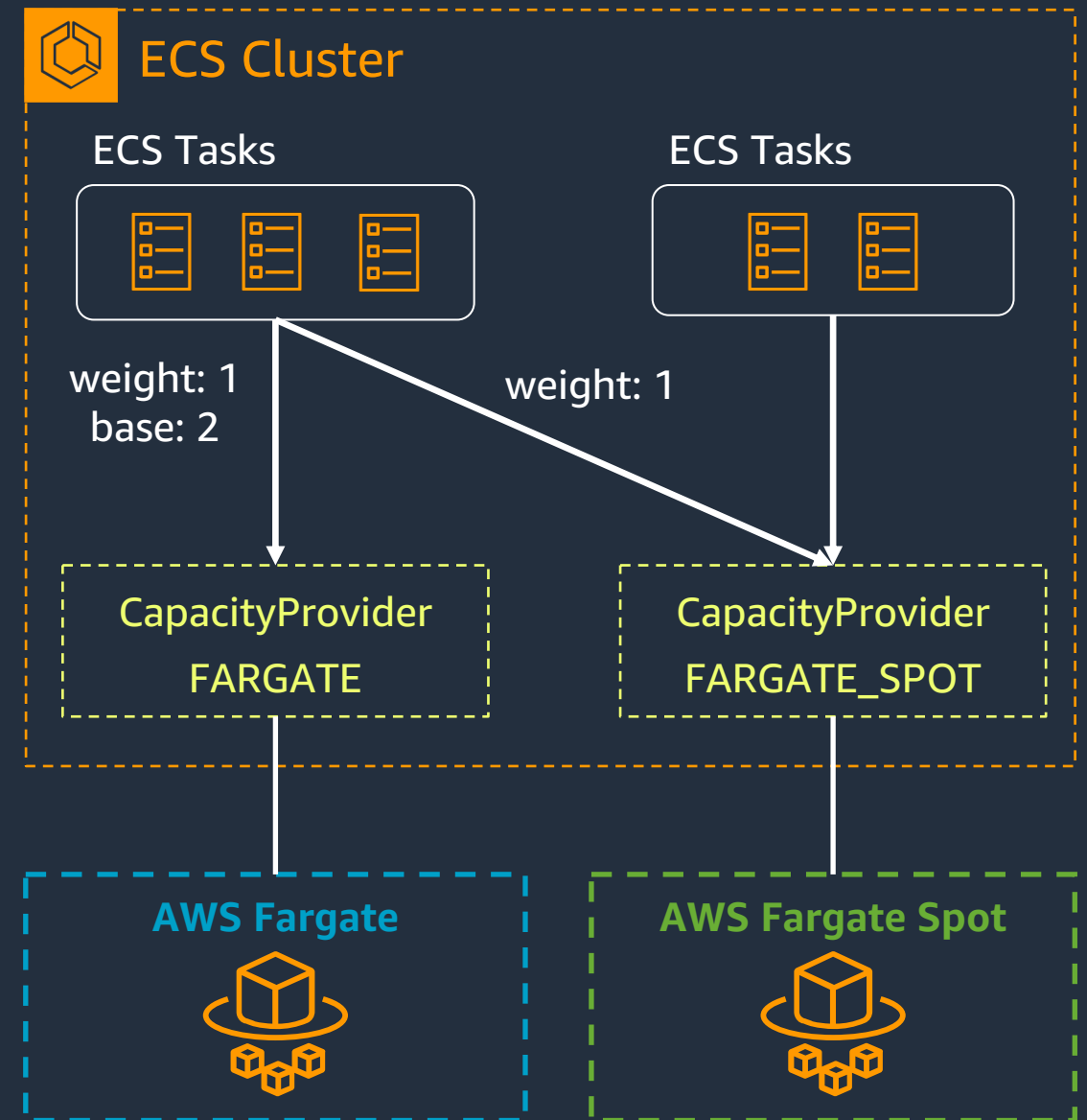
- **weight**
 - 実行するタスクの総数に対する指定した Capacity Provider での相対的な割合
- **base**
 - その Capacity Provider での最小実行タスク数

```
$ aws ecs run-task
--cluster MyEcsCluster
--task-definition task-def-family:revision
--count 6
--capacity-provider-strategy
  capacityProvider=CapacityProviderA,weight=1,base=2
  capacityProvider=CapacityProviderB,weight=1
```



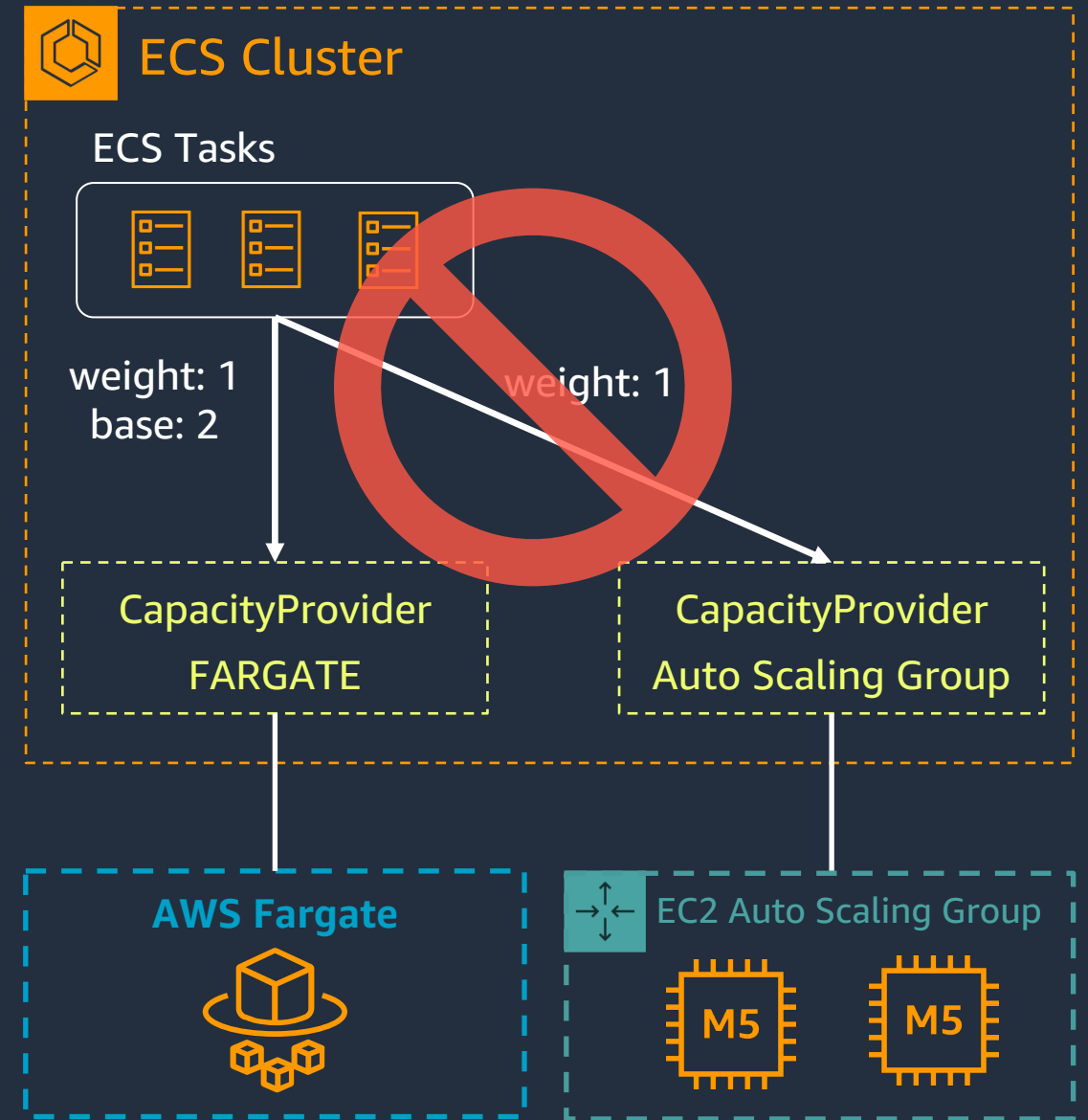
Capacity Provider Strategy (Fargate の場合)

- Fargate の場合も基本的な考え方は同じ
- Fargate と Fargate Spot は別の Capacity Provider で扱う
 - Fargate -> FARGATE
 - Fargate Spot -> FARGATE_SPOT
- FARGATE / FARGATE_SPOT は事前定義済みのため、ECS クラスタに紐づけるだけで利用可能



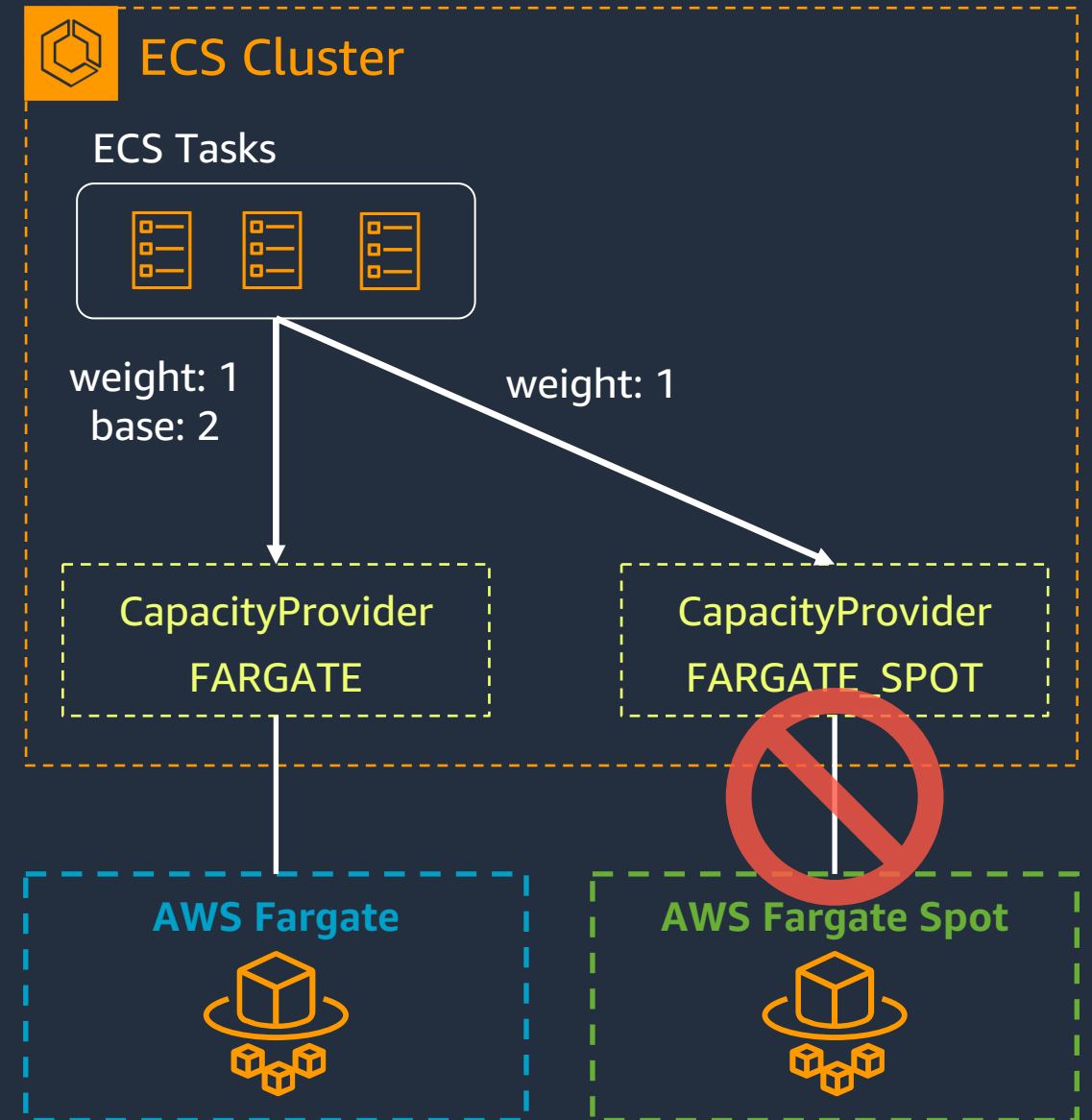
Capacity Provider Strategy の考慮事項 (1)

1. Fargate の Capacity Provider と Auto Scaling グループの Capacity Provider を混在させた Strategy を設定することはできない



Capacity Provider Strategy の考慮事項 (2)

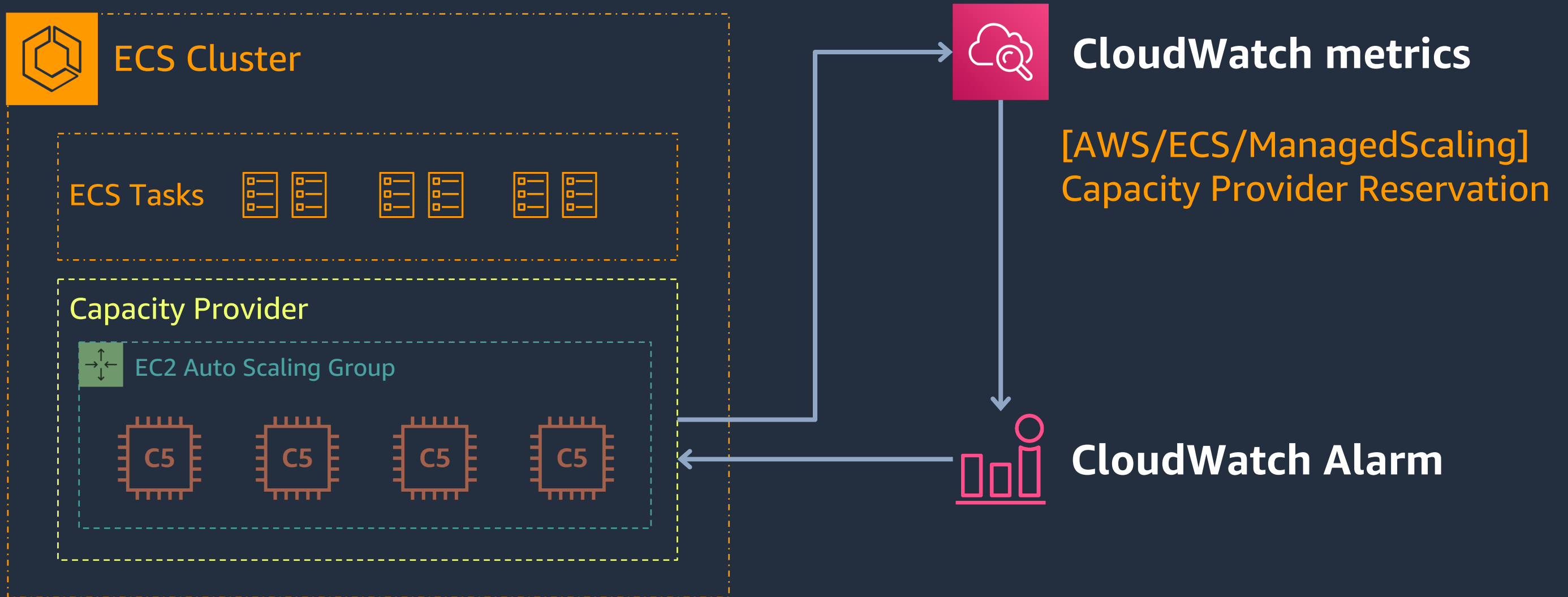
2. 特定の Capacity Provider でタスクが実行できない場合に、別の Capacity Provider にフォールバックすることはできない



Capacity Providers とは

- ECS クラスタ内のタスクで使用するインフラストラクチャを管理
- 以下のような概念、機能を持つ
 - Capacity Provider
 - Capacity Provider Strategy
 - Capacity Provider Reservation
 - Managed Scaling
 - Managed Termination Protection

Capacity Provider Reservation



Capacity Provider Reservation

CapacityProviderReservation = $M / N \times 100$ で導出される値

- M: 実行しようとしているタスクに必要な Auto Scaling グループ内のインスタンス数
- N: 既に実行されている Auto Scaling グループ内の現在のインスタンス数

このメトリクスを利用したスケーリングポリシーで Cluster Auto Scaling を実現している

- Target Capacity = 100 → 既存リソースを使い切るまでスケールしない
- Target Capacity < 100 → 早めにスケールアウトし、急なスパイクにも対応

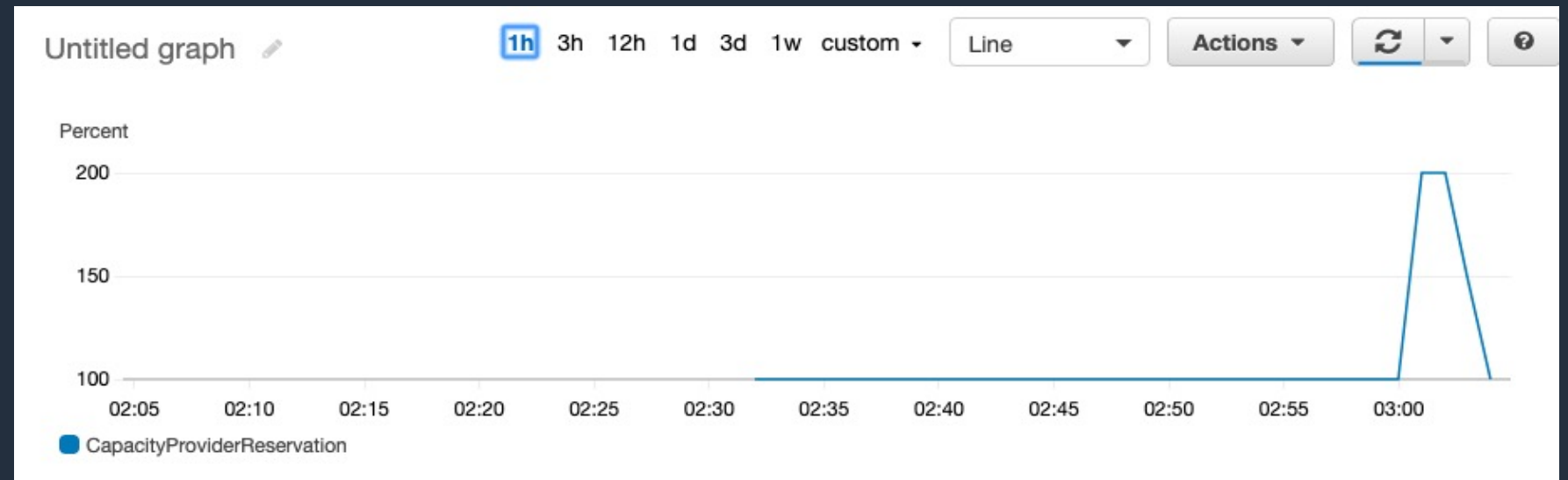
キャパシティプロバイダーを作成

クラスター名

キャパシティプロバイダー名*

マネージドスケーリング 有効にする 無効にする

ターゲットキャパシティ %



Capacity Providers とは

- ECS クラスタ内のタスクで使用するインフラストラクチャを管理
- 以下のような概念、機能を持つ
 - Capacity Provider
 - Capacity Provider Strategy
 - Capacity Provider Reservation
 - **Managed Scaling**
 - Managed Termination Protection

Managed Scaling

キャパシティープロバイダーを作成

クラスター名

キャパシティープロバイダー名*

キャパシティープロバイダーの名前です。アルファベット (大文字と小文字)、数字、アンダースコア、ハイフンを含めて最大 255 文字を使用できます。この名前には、「aws」、「ecs」、または「fargate」のプレフィックスを付けることはできません。

Auto Scaling グループ

希望する Auto Scaling グループ ARN を手動で入力します

キャパシティープロバイダーを作成する前に Auto Scaling グループを作成する必要があります。
[Amazon EC2 コンソール](#)を使用して Auto Scaling グループを作成します。

マネージドスケーリング 有効にする 無効にする ⓘ

ターゲットキャパシティー % ⓘ

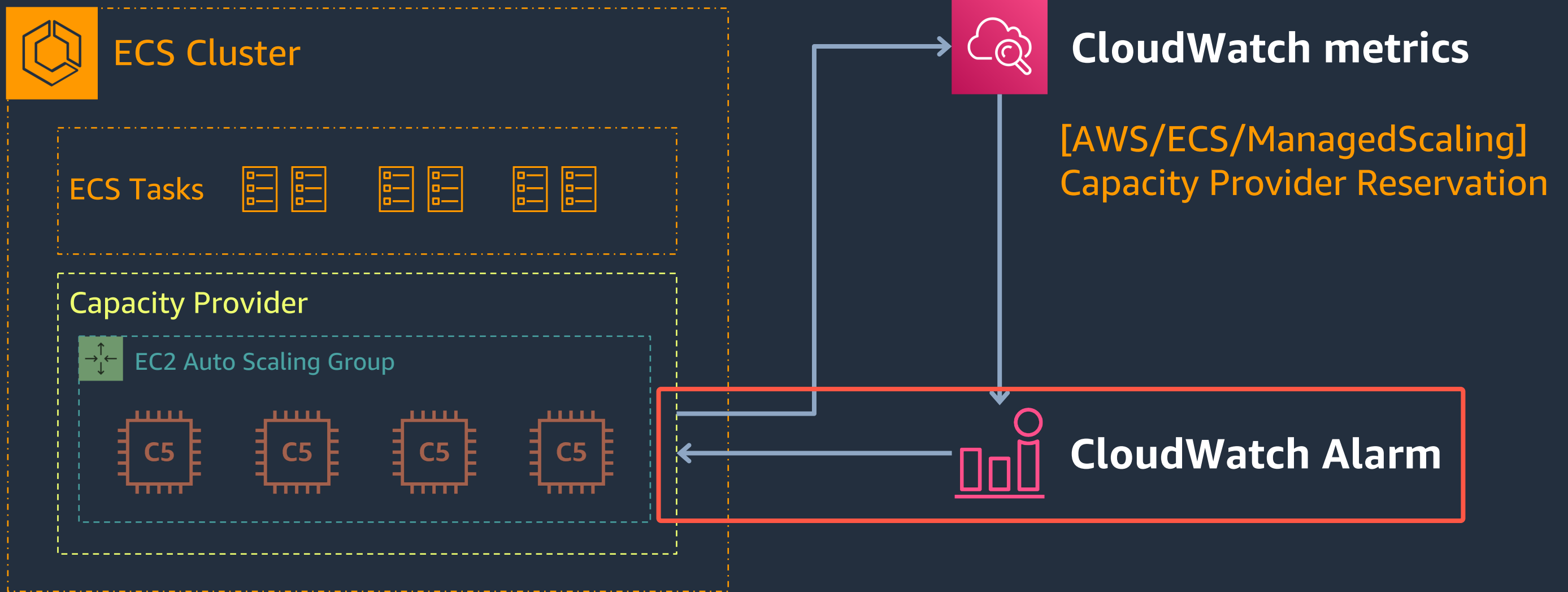
マネージドターミネーション保護 有効にする 無効にする ⓘ

*必須

- Auto Scaling グループの Capacity Provider ではマネージドスケーリングを有効にできる
- マネージドスケーリングが有効な場合、ECS が Auto Scaling グループのスケールインおよびスケールアウトのアクションを管理



Managed Scaling



Capacity Providers とは

- ECS クラスタ内のタスクで使用するインフラストラクチャを管理
- 以下のような概念、機能を持つ
 - Capacity Provider
 - Capacity Provider Strategy
 - Capacity Provider Reservation
 - Managed Scaling
 - **Managed Termination Protection**

Managed Termination Protection

キャパシティープロバイダーを作成

クラスター名

キャパシティープロバイダー名*

キャパシティープロバイダーの名前です。アルファベット (大文字と小文字)、数字、アンダースコア、ハイフンを含めて最大 255 文字を使用できます。この名前には、「aws」、「ecs」、または「fargate」のプレフィックスを付けることはできません。

Auto Scaling グループ

希望する Auto Scaling グループ ARN を手動で入力します

キャパシティープロバイダーを作成する前に Auto Scaling グループを作成する必要があります。
[Amazon EC2 コンソール](#)を使用して Auto Scaling グループを作成します。

マネージドスケーリング 有効にする 無効にする ⓘ

ターゲットキャパシティー % ⓘ

マネージドターミネーション保護 有効にする 無効にする ⓘ

*必須

マネージドターミネーション保護が有効な場合、ECS はタスクを実行中の EC2 インスタンスが終了しないように Auto Scaling グループのスケールイン保護を制御する



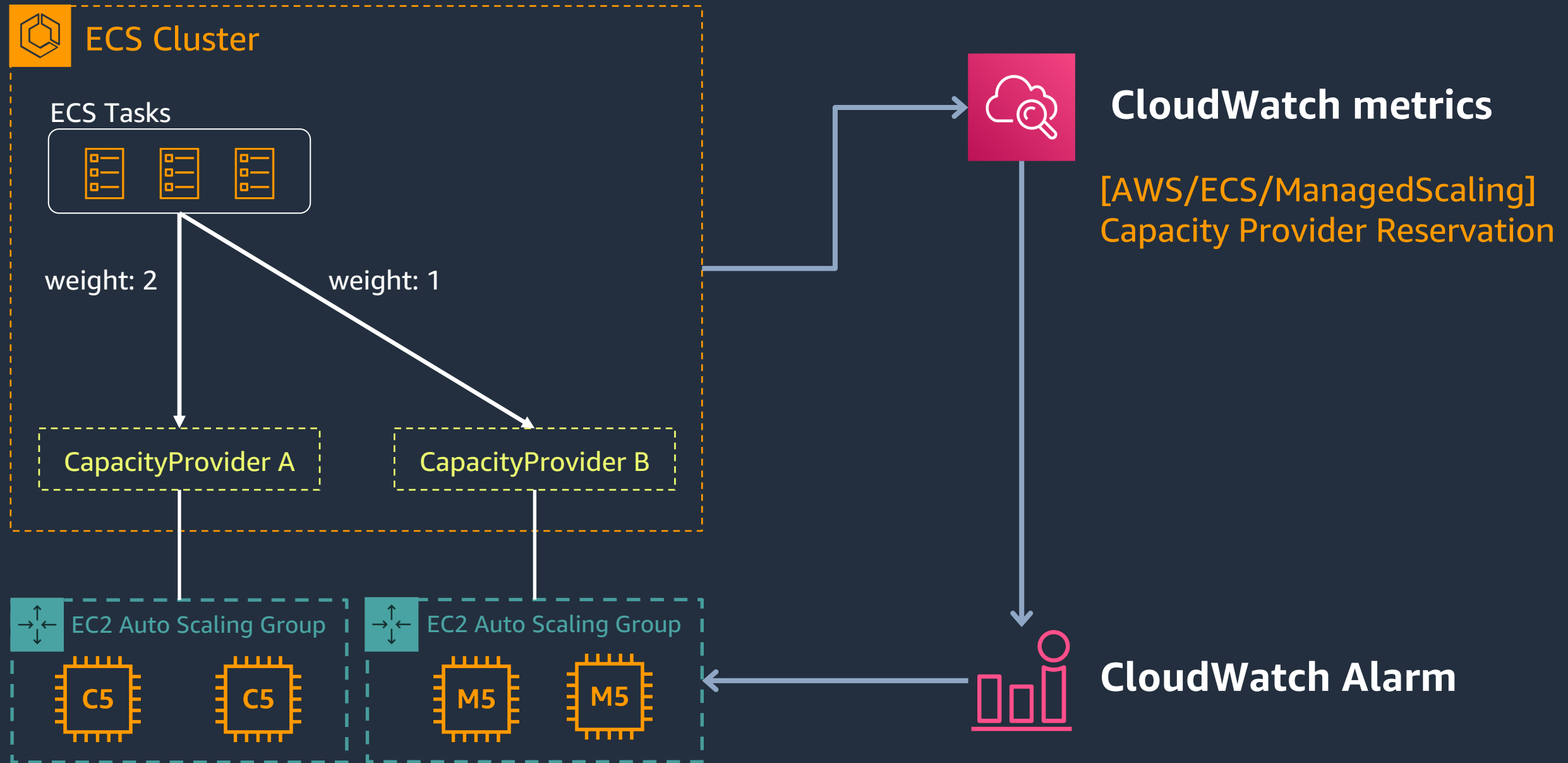
Managed Termination Protection



まとめ



まとめ



まとめ

- ECS on EC2 では、EC2 レイヤと ECS サービスレイヤの Auto Scaling を設定する必要がある
- ECS Cluster Auto Scaling により EC2 レイヤのマネージドスケーリングが可能に
- Capacity Providers という機能により ECS Cluster Auto Scaling を実現
 - Capacity Provider
 - Capacity Provider Strategy
 - Capacity Provider Reservation
 - Managed Scaling
 - Managed Termination Protection

Thank you!