AWS Black Belt Online Seminar

Amazon EC2 Auto Scaling

スケーリングポリシーとおすすめ機能編

滝口 開資 (はるよし)

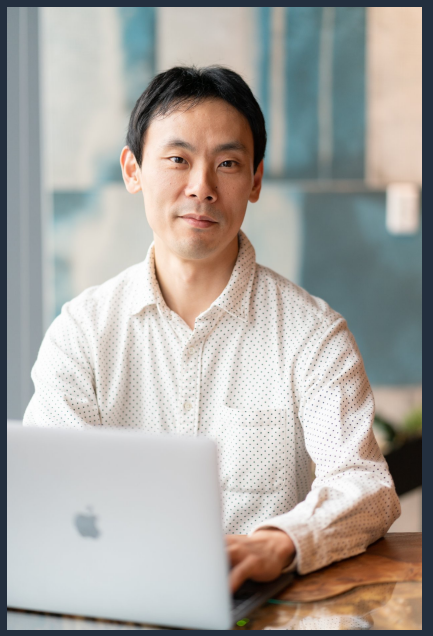
シニアソリューションアーキテクト
EC2 フレキシブルコンピュートスペシャリスト
2023/10

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 インスタンスを自動スケールさせる際に必要となる基礎知識を知りたい方

本セミナーの前提知識

- Black Belt Online Seminar Amazon EC2 入門
- Black Belt Online Seminar Amazon EC2 Auto Scaling 入門編
- Black Belt Online Seminar Amazon EC2 Auto Scaling 複数のインスタンスタイプと購入オプションの活用編

ヒント

AWS Black Belt コンピュートシリーズのあるきかた | Amazon Web Services ブログ –
https://aws.amazon.com/jp/blogs/news/aws-black-belt-compute-series/

アジェンダ

- Auto Scaling サービス群の整理

- EC2 Auto Scaling おすすめ機能
  - ワンタッチでできる自動スケール設定
  - ライフサイクルフック
  - インスタンスリフレッシュ
  - ウォームプール

- インスタンスの置き換えに関するよくある質問集

Auto Scaling サービス群の整理

3 つの Auto Scaling サービス群

- Amazon EC2 Auto Scaling
  - EC2 インスタンスの自動スケール機能を提供
- Application Auto Scaling
  - EC2 インスタンス以外のリソースにも自動スケーリングを提供
    - ECS サービスタスク、スポットフリート、AppStream 2.0 フリート、DynamoDB テーブル、Aurora レプリカ、ElastiCache for Redis レプリケーショングループ、SageMaker エンドポイントバリアント、Lambda 関数プロビジョン済み同時実行数、カスタムリソースなど

- AWS Auto Scaling
  - 2種類の Auto Scaling のスケーリングを設定・管理するためのワンストップサービス

Auto Scaling サービス群

- Amazon EC2 Auto Scaling
  - EC2 インスタンスの自動スケール機能を提供
- Application Auto Scaling
  - EC2 インスタンス以外のリソースにも自動スケーリングを提供
    - ECS クラスター、スポットフリート、EMR クラスター、AppStream 2.0 フリート、DynamoDB テーブル、Aurora レプリカ、SageMaker エンドポイントバリアント、Lambda 関数プロビジョン済み同時実行数、カスタムリソースなど

- ~~AWS Auto Scaling~~ (新規投資予定なし)
  - ~~2種類の Auto Scaling のスケーリングを設定・管理するためのワンストップサービス~~

- 新規適用は非推奨。各サービスの Application Auto Scaling 機能を活用してください
- バグフィックスは引き続き提供されます

EC2 Auto Scaling のおすすめ機能

EC2 Auto Scaling のおすすめ機能

- ワンタッチでできる自動スケール設定
- ライフサイクルフック
- インスタンスリフレッシュ
- ウォームプール

EC2 Auto Scaling のおすすめ機能

- ワンタッチでできる自動スケール設定
- ライフサイクルフック
- インスタンスリフレッシュ
- ウォームプール

ワンタッチでできる自動スケール設定

- ターゲット追跡スケーリング + 予測スケーリングの組み合わせ
- スケジュールスケーリングも組み合わせられます

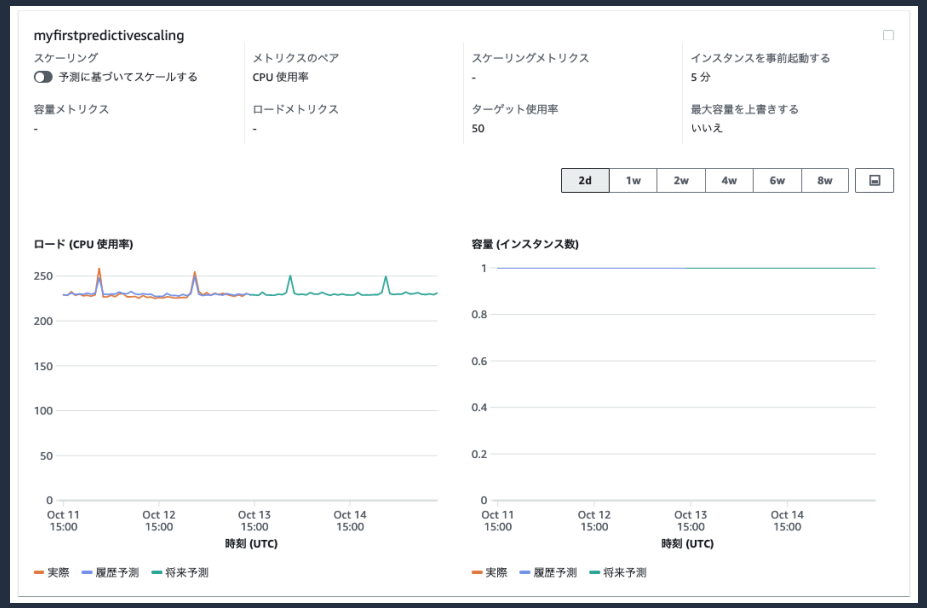
ターゲット追跡スケーリング

- 1つのメトリクスに対し、単に目標値を指定するのみで良い
  - CPUUtilizationを50%に維持して欲しい、ただこれだけ

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/as-scaling-target-tracking.html

予測スケーリング

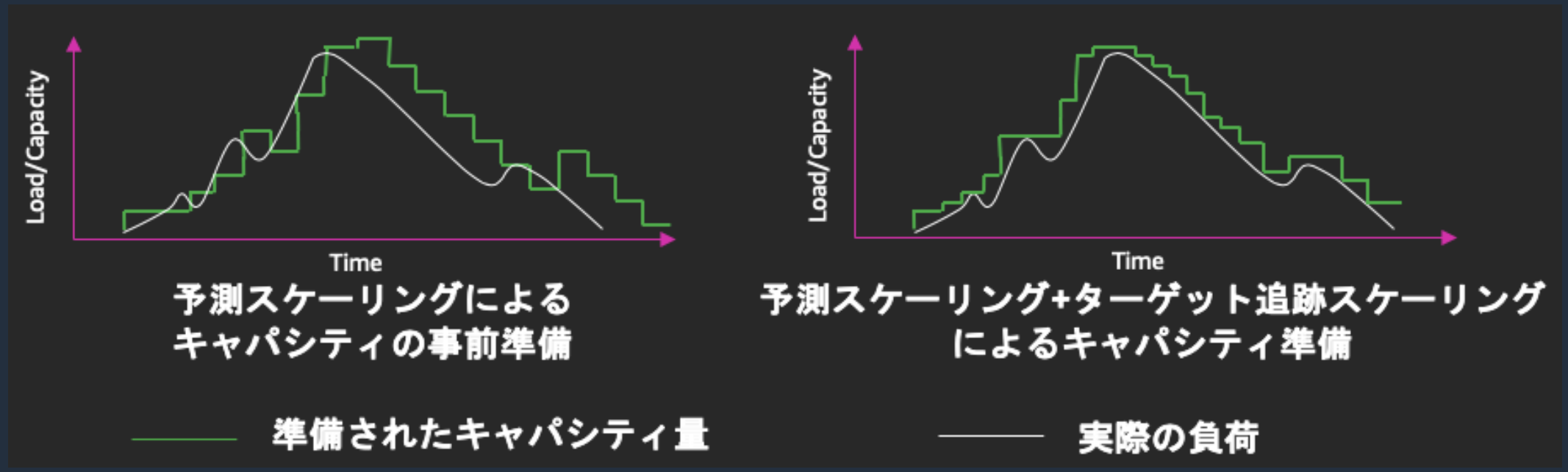
- 2週間分のメトリクスを分析し、次の2日の今後の需要を予測
  - 最短で24時間分のメトリクスデータから始められる
- 予測データに基づいてキャパシティの増減がスケジュールされる



https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/ec2-auto-scaling-predictive-scaling.html

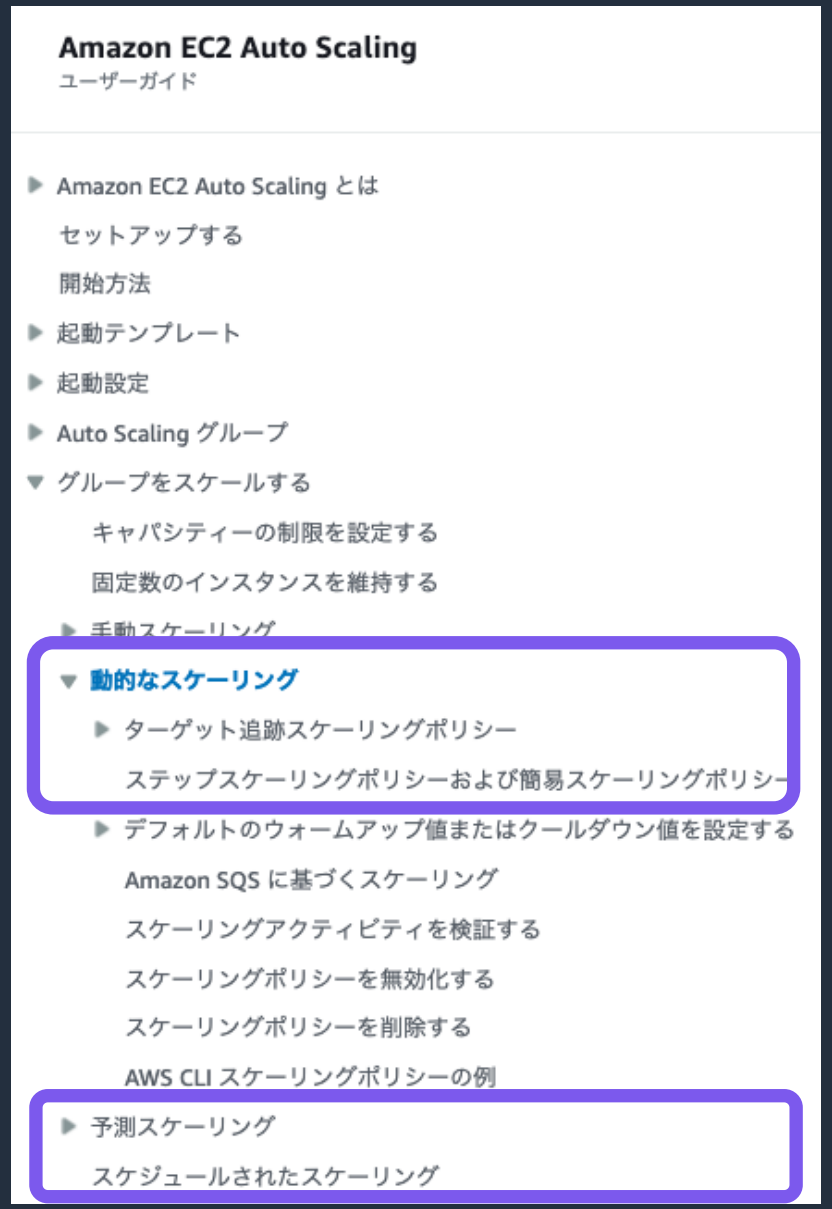
ターゲットトラッキング + 予測スケーリングの組み合わせ

1. 大まかなキャパシティ増減は予測スケーリングに任せ、
   前もってスケールしておく
2. 実際の負荷に対して不足した分をターゲット追跡で補充する
3. さらにスケジュールスケーリングを組み合わせることもできる



参考：スケーリングポリシーの整理

- 動的なスケーリング
  - 簡易スケーリング
  - ステップスケーリング
  - ターゲット追跡スケーリング
- 予測スケーリング
- スケジュールスケーリング



参考：スケーリングポリシーの整理

- 動的なスケーリング
  - 簡易スケーリング
  - ステップスケーリング
  - ターゲット追跡スケーリング
- 予測スケーリング
- スケジュールスケーリング

- 簡易スケーリングポリシーは互換性維持のために残されている。新規で作成する必要はない
- ユースケースによって、きめ細やかなスケール条件を指定できるステップスケーリングポリシーを採用する場合がある。ただし無理に使う必要はない

- 2023年のおすすめはターゲット追跡スケーリング + 予測スケーリング + スケジュールスケーリングの組み合わせ。最小の手間で最大の効果を

EC2 Auto Scaling のおすすめ機能

- ワンタッチでできる自動スケール設定
- ライフサイクルフック
- インスタンスリフレッシュ
- ウォームプール

ライフサイクルフック

- インスタンスの起動時や終了時に何かしたい、を実現する仕組み
- 起動時ライフサイクルフックが有効な場面
  - 例：ELBに登録される前にインスタンス上の様々な準備が正しく完了していることを確認したい
- 終了時ライフサイクルフックが有効な場面
  - 例：スケールインが発生するとき、アプリケーションを安全に終了させてからのインスタンス削除を保証したい

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/lifecycle-hooks.html

インスタンスリフレッシュ

- Auto Scaling グループ内のインスタンスを自動的に更新してくれる仕組み
  - AMI更新時などの場面で、手動で入れ替える必要がなくなった
- 入れ替えは一定割合のインスタンスが稼働中(Healthy)であることを保ちながら実施される
  - デフォルトは90%

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/asg-instance-refresh.html

ウォームプール

- 起動に長い時間のかかるインスタンスを事前起動できる仕組み
  - 事前に起動されたインスタンスが「ウォームプール」にStopped状態で保持
    - 事前起動することで時間を稼ぐ
    - 発生する費用はEBSボリュームとEIPのみ
  - スケールアウトが発生するとウォームプールから開始(Start)される
    - ゼロから起動(Launch)するよりも格段に速い
- 制約
  - スポットインスタンスを含むASG, 複数インスタンスタイプを指定したASGにはウォームプールを追加できない

スケール速度に問題がないケースでは
無理にウォームプールを使う必要はない

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/ec2-auto-scaling-warm-pools.html

インスタンスの置き換えに関するよくある質問集

こんなときどうする？

- 正常に動作しないインスタンスを自動的に置き換えたい
  - →ヘルスチェックを活用

- 特に指定しない場合、EC2 ヘルスチェックが有効になっている
  - 2/2 以外のステータスが続くとAuto Scaling サービスが置き換える
- ELB 配下の Auto Scaling グループの場合、ELB ヘルスチェックを有効にする
  - EC2 ヘルスチェックに加え、ELB からのヘルスチェックに応答しない場合の速やかな入れ替えが可能になる

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/ec2-auto-scaling-health-checks.html

こんなときどうする？

- スケールイン・スケールアウトを繰り返してしまい、いつまでたってもインスタンスが追加されない
    - →「ヘルスチェックの猶予期間」の設定を見直す
- ヘルスチェックの猶予期間：起動したばかりでヘルスチェックに応答できないインスタンスを保護する期間
    - デフォルトは 5 分 (300 秒)
        - もしこれよりインスタンスの準備に時間がかかるとすると、ヘルスチェックがそのインスタンスを置き換えてしまう。そのままでは希望容量を満たさないので再びスケールアウトが試行され、インスタンスの起動と削除が繰り返される
    - ELB ヘルスチェックに、 ユーザーデータなどで指示した S3 からのコンテンツ配備や DB 接続などを前提としたアプリケーションのパスを指定している場合に有効なときがある

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/health-check-grace-period.html

こんなときどうする？

- 特定のインスタンスをスケールインから保護したい
    - →インスタンスの保護

- ASG単位、もしくはインスタンス単位で設定。スケールインされなくなる
- 次の条件からは保護できないことに注意
    - 手動でのインスタンス削除(Terminate)
    - ヘルスチェックによる置き換え
    - スポットインスタンスの中断
- すべてのインスタンスが終了保護された状態でスケールインイベントが発生した場合、希望容量だけが減少し、スケールイン(インスタンス削除)は行われない

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/ec2-auto-scaling-instance-protection.html

こんなときどうする？

- 一時的にスケールインやスケールアウトを止めたい
    - →スケーリングプロセスの中断
- 一時的にスケール動作を停止できる
- ASG単位で設定
- 中断できるプロセス一覧：Launch, Terminate, AddToLoadBalancer, AlarmNotification, AZRebalance, HealthCheck, ReplaceUnhealthy, ScheduledActions
- 使いどころ：機能テストなど、一時的にAuto Scalingグループの特定プロセスの動作を止めてテスト条件を整えたい場合
    - LaunchとTerminateの両方のプロセスを中断することで、「何もしない」Auto Scalingグループを作り出せる
- 動作のおかしいインスタンスがいるのでスケールイン・スケールアウトを止めたい
    - →プロセスの中断ではなく次の項目を参照

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/as-suspend-resume-processes.html

こんなときどうする？

- 特定のインスタンスを Auto Scaling グループから外したい
    - →スタンバイ、もしくはデタッチ
- スタンバイ(「一時的なインスタンスの削除」)
    - インスタンス単位で設定
    - そのインスタンスは Auto Scaling グループにいながら「スタンバイ」状態に入る
        - 具体的にはそのインスタンスはELBから登録解除され、ヘルスチェック対象から外され、その Auto Scaling グループの希望容量は1つ減少する
    - その間にインスタンスのトラブルシューティングなどを行う
- デタッチ
    - インスタンス単位で設定
    - そのインスタンスはその Auto Scaling グループのメンバーから外れる
    - スタンバイと実質的な効果は同一。インスタンスはそのまま Running 状態で保持される。ただしデタッチの場合、Auto Scaling グループとして与えていたタグも除去される
    - 作業後、そのまま終了予定であればデタッチが適する

https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/as-enter-exit-standby.html
https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/detach-instance-asg.html

おわりに

今回お話しした内容

- Auto Scaling サービス群の整理
- EC2 Auto Scaling おすすめ機能
  - ワンタッチでできる自動スケール設定
  - ライフサイクルフック
  - インスタンスリフレッシュ
  - ウォームプール
- インスタンスの置き換えに関するよくある質問集

ヒント

AWS Black Belt コンピュートシリーズのあるきかた | Amazon Web Services ブログ ー
https://aws.amazon.com/jp/blogs/news/aws-black-belt-compute-series/

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
- 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

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

内容についての注意点

- 本資料では資料作成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。AWS のサービスは常にアップデートを続けているため、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
- 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
- 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)

Thank you!