



Amazon EC2 スポットインスタンス活用のための 6つのベストプラクティスと実践例

AWS Black Belt Online Seminar

石神 靖弘

Solutions Architect

2023/03

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=PLzWGOASvSx6FIwIC2X1nObr1KcMCBBBlqY>

内容についての注意点

- 本資料では2023年3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おります。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：石神 靖弘（いしがみ やすひろ）

所属：技術統括本部
ソリューションアーキテクト



好きなAWSサービス：Amazon EC2

本セミナーの対象者

スポットインスタンスをすでに利用されている方

スポットインスタンスのベストプラクティスについて学びたい方

アジェンダ

1. スポットインスタンス活用のための6つのベストプラクティス

- 多様なインスタンスタイプを混ぜる
- 複数のアベイラビリティゾーンを使う
- スポットインスタンスの中断に備える
- Auto Scaling グループやフリート機能を使う
- price-capacity-optimized 配分戦略を使う
- キャパシティリバランシングを使う

2. スポットインスタンスの実践例

スポットインスタンス 活用のための 6つのベストプラクティス

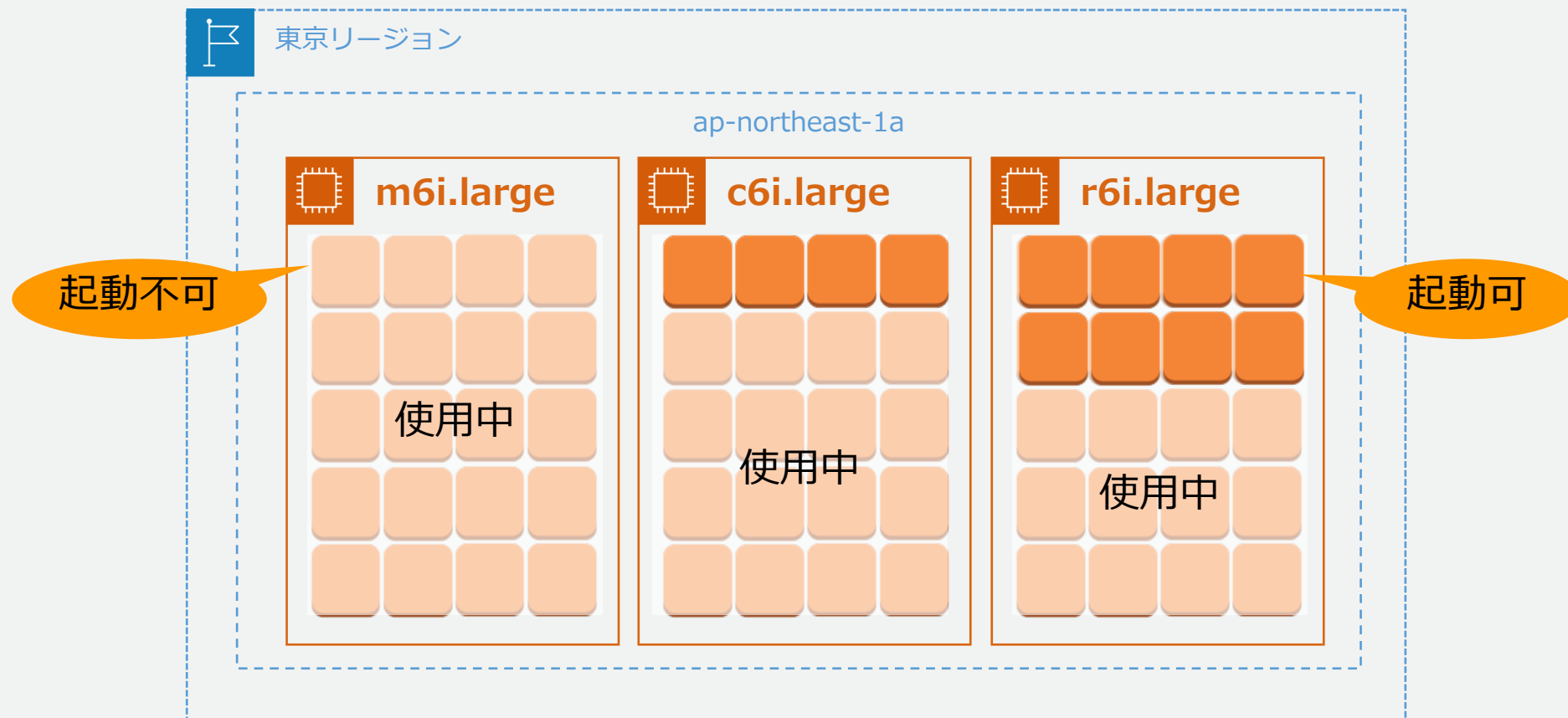
スポットインスタンス活用のための 6つのベストプラクティス

1. 多様なインスタンスタイプを混ぜる
2. 複数のアベイラビリティゾーンを使う
3. スポットインスタンスの中断に備える
4. Auto Scaling グループやフリート機能を使う
5. price-capacity-optimized 配分戦略を使う
6. キャパシティリバランシングを使う

1. 多様なインスタンスタイプを混ぜる

あらかじめ複数のインスタンスタイプを指定しておく

(m6i.large が足りない場合でも、r6i.large や c6i.large を起動できるためサービス影響を最小化)



1-1. 属性ベースのインスタンスタイプ自動選定 (Attribute-Based Instance Type Selection, ABIS)

インスタンスタイプの要件 Info テンプレートを起動するためにリセット

起動テンプレートから同じインスタンス属性またはインスタンスタイプを維持することも、異なるインスタンス属性を指定するか、手動でインスタンスタイプを追加して起動テンプレートを上書きすることもできます。

インスタンス属性を指定
コンピューティング要件を指定します。配分戦略の選択に基づいて、一致するインスタンスタイプで希望する容量を満たします。

手動でインスタンスタイプを追加
1つ以上のインスタンスタイプを追加します。任意のインスタンスタイプを起動して、配分戦略の選択に基づいて希望する容量を満たすことができます。

必須インスタンス属性
コンピューティング要件を仮想 CPU (vCPU) とメモリに入力します。

vCPUs
インスタンスあたりの vCPU の最小数と最大数を入力します。
0 最小 8 最大
 最小値なし 最大値なし

メモリ (GiB)
インスタンスあたりのメモリの最小量および最大量 (GiB) を入力します。
0 最小 最大
 最小値なし 最大値なし

追加のインスタンス属性 - 省略可能
インスタンス属性を追加して、希望する容量を満たすために使用できるインスタンスタイプをさらに制限します。
属性を選択 属性を追加

▼ **一致するインスタンスタイプをプレビュー (147)**
このリストには、コンピューティング要件に一致するすべてのインスタンスタイプが含まれます。Amazon EC2 は、これらのインスタンスタイプからプロビジョニングできます。希望する容量を満たすために使用される正確なインスタンスタイプは、選択する配分戦略と使用可能な容量によって異なります。

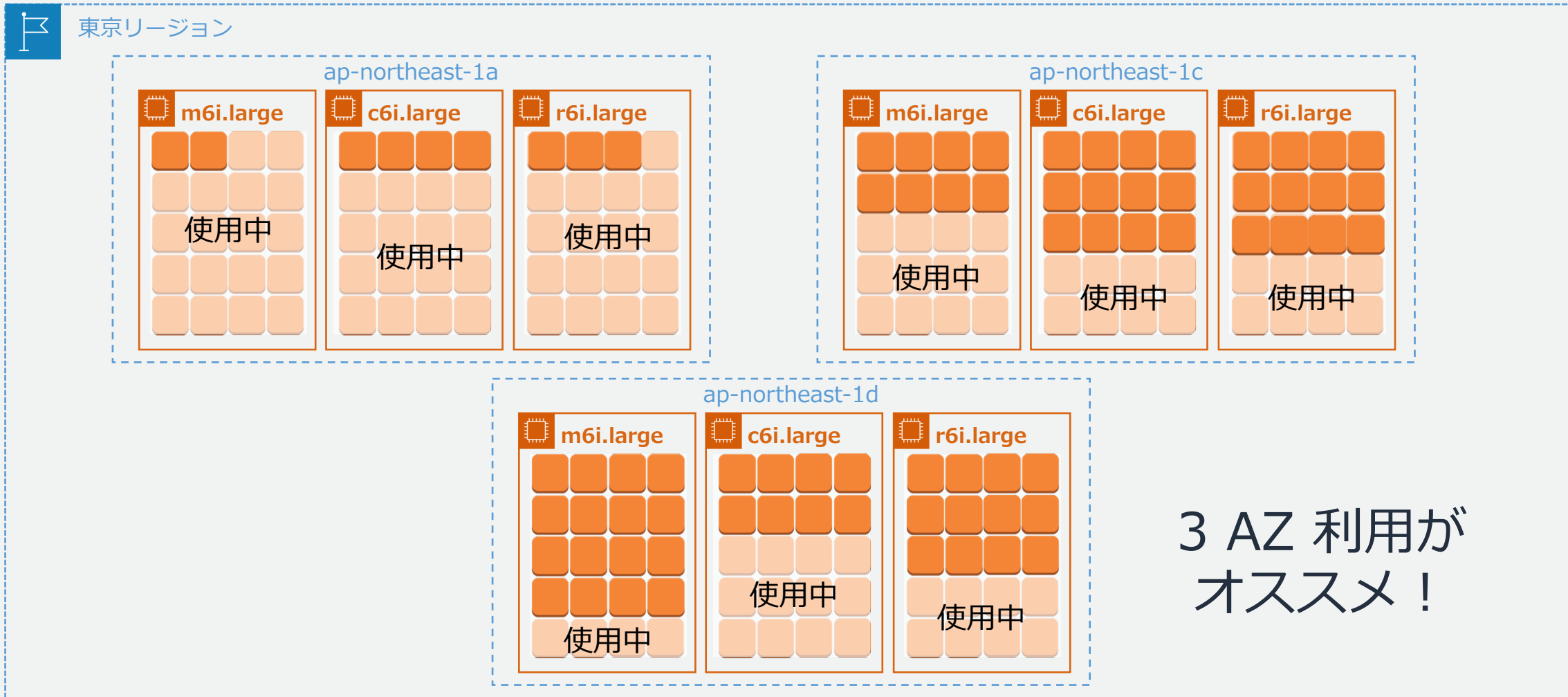
インスタンスタイプをフィルタリング 選択したインスタンスタイプを除外

<input type="checkbox"/>	インスタンスタイプ ▲	vCPUs ▼	メモリ (GiB) ▼
<input type="checkbox"/>	a1.2xlarge	8	16
<input type="checkbox"/>	a1.large	2	4
<input type="checkbox"/>	a1.medium	1	2
<input type="checkbox"/>	a1.xlarge	4	8

- vCPU 数やメモリ数量などの条件に応じて、該当するインスタンスタイプリストを表示
- スポットインスタンスの活用に必須であるインスタンスタイプの多様化を簡単に実現できる
- EC2 Auto Scaling、EC2 フリート、スポットフリートから利用可能

2. 複数のアベイラビリティゾーンを使う

複数のアベイラビリティゾーンを定義し、起動可能なスポットプールを増やす



3 AZ 利用が
オススメ！

ヒント：スポットプール数は最低20、可能なら30を目指す

- スポットプール数 = インスタンスタイプの数 x AZの数
 - インスタンスタイプ7種類 と 3AZ で21プール
- スポットインスタンスを本番適用する際に、
実用的なレベルでの起動のしやすさ・中断のしにくさを得られるのは20プールを超えるあたりからとなる

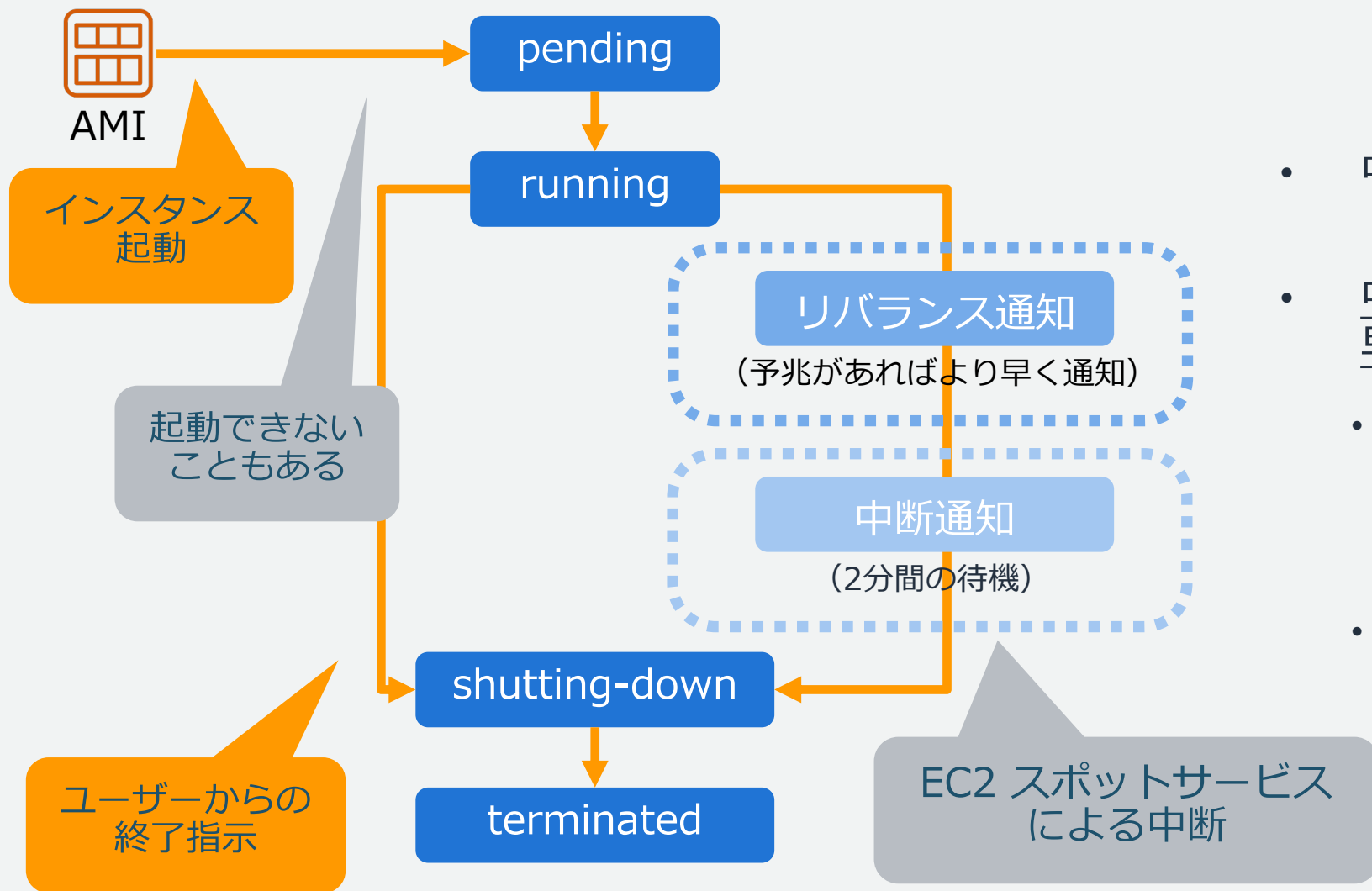
ヒント：スポットプール数は最低20、可能なら30を目指す

- DeNA様の事例では20プールまで増やしたところで、一部のスポットプールのキャパシティ不足の場合でも、スポットプールで起動されたスポットインスタンスにより、サービス提供を継続できるようになった

株式会社ディー・エヌ・エー様におけるEC2スポットインスタンスの大規模活用のための工夫とコンテナ技術を用いた設計例の紹介 | Amazon Web Services ブログ

<https://aws.amazon.com/jp/blogs/news/how-dena-successfully-applied-ec2-spot-on-production-and-reference-architecture-using-containers/>

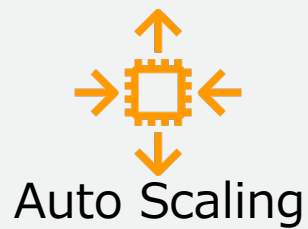
3. スポットインスタンスの中断に備える



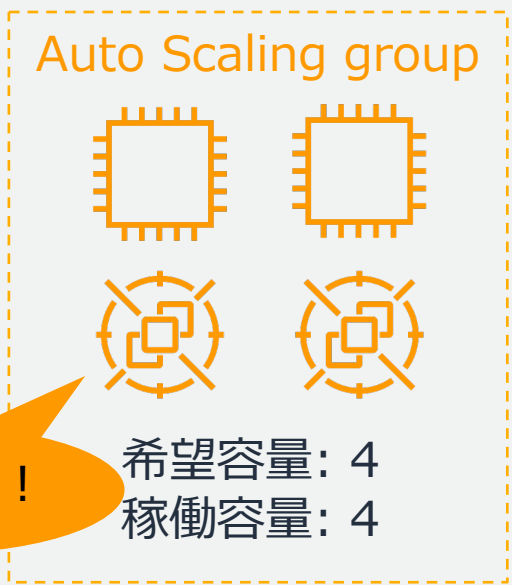
- 中断の発生する2分前に「中断通知」を送付
- 中断の予兆があれば、2分より早く「リバランス通知」を送付
- あるスポットインスタンスの中断リスクが高まってきたとき、EC2 サービスから「リバランス通知」が通知される
- 2分前の中断通知に先んじて提供されることが多い

4. Auto Scaling グループやフリート機能を使う

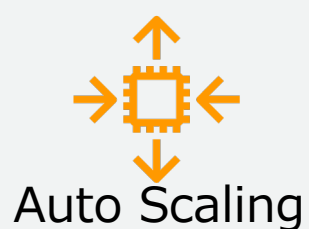
Auto Scaling グループ (ASG) が EC2 インスタンスの台数を自動維持
つまり、中断発生時の自動補充を任せられる



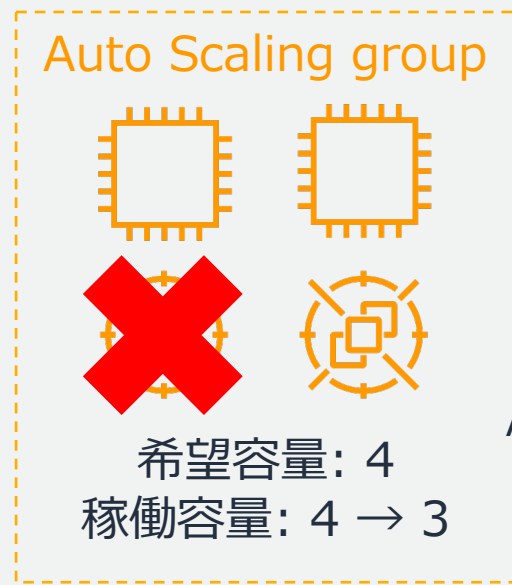
Auto Scaling



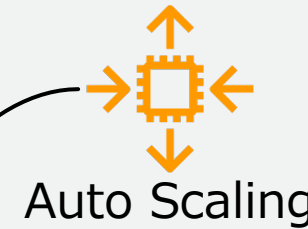
スポット
インスタンスの
中断が発生



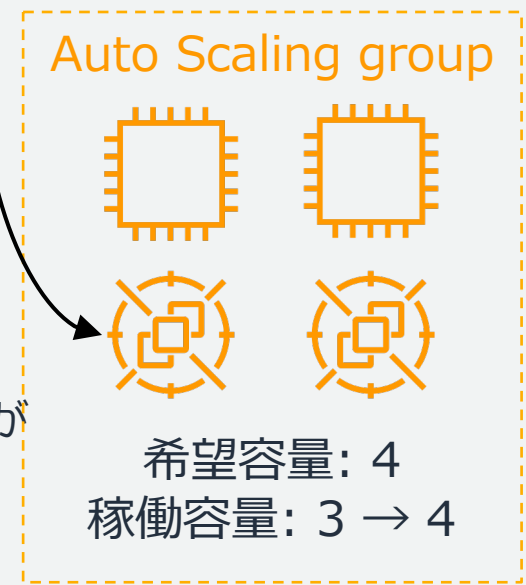
Auto Scaling



Auto Scaling が
自動補充



Auto Scaling



中断発生!



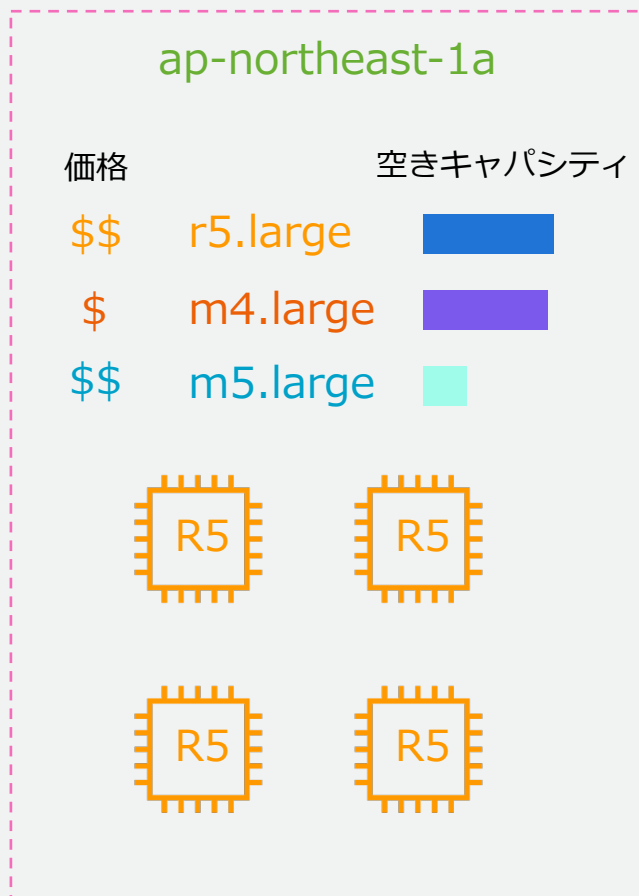
5. price-capacity-optimized 配分戦略を使う

主な配分戦略

- lowest-price
 - 価格が最安値のスポットプールを優先して選ぶ
 - 価格のみが考慮されるため、スポットインスタンスの中断がより頻繁に発生するケースがある
- capacity-optimized
 - 起動するスポットインスタンスを起動のしやすさのみで決定
 - 最も中断しにくいスポットプールから起動
- capacity-optimized-prioritized
 - capacity-optimizedに加えて、複数のスポットプール間で起動のしやすさが同等である場合、リスト上位のものを優先
- **price-capacity-optimized (推奨)**
 - 起動するスポットインスタンスを起動のしやすさとスポット価格の組み合わせで決定
 - なるべく価格が低く、かつ起動しやすいスポットプールから起動
 - 2022年11月に発表。費用と安定性のバランスを取れるためこのオプションを推奨

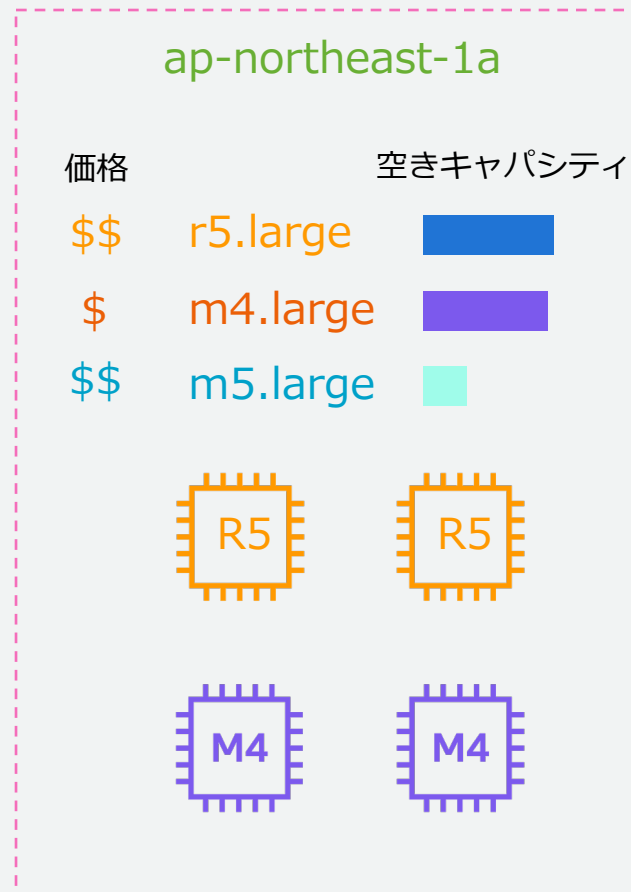
5. price-capacity-optimized 配分戦略を使う

capacity-optimized



起動のしやすさのみを考慮

price-capacity-optimized

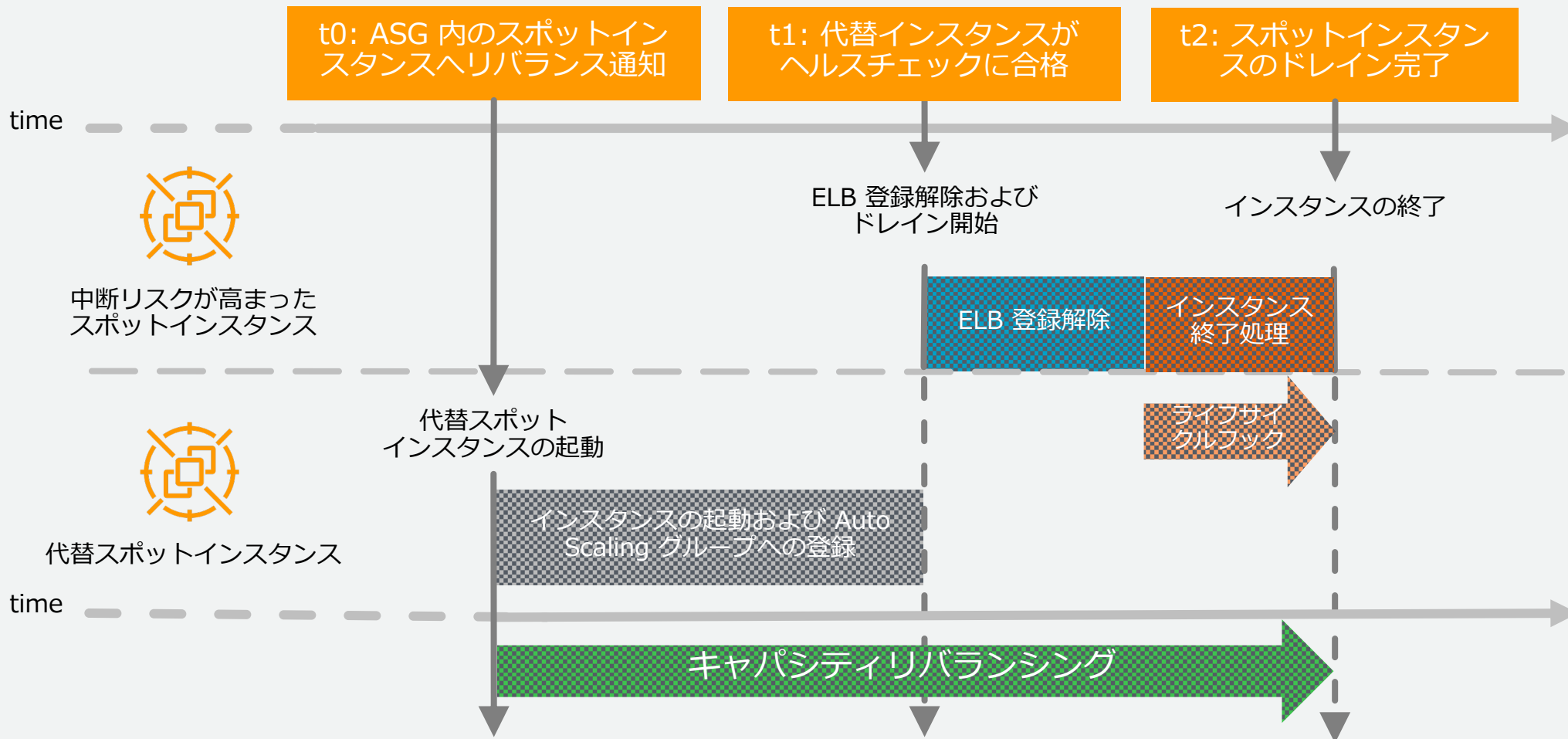


価格と起動のしやすさを考慮

空きキャパシティがほぼ同じであれば低価格プールからスポットインスタンスを起動

6. キャパシティリバランシングを使う

ASG やフリートで発生するスポットインスタンスの中断に対し、事前にスポットインスタンスを追加してキャパシティを維持する



ヒント：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
 - 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
 - キャパシティリバランシングによりスポットインスタンスの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インスタンスが削除されてほしい

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

ヒント：FISによる中断のシミュレーション

- 障害試験用のマネージドサービスである Fault Injection Simulator (FIS) から、スポットインスタンス中断を発生させられる
- リバランス通知と中断通知が送付されたのちに、実際の中断が発生する
- 単体インスタンスだけでなく、ASGを対象にできる

▼ 新しいアクション 保存 削除

名前

説明 - オプション

アクションタイプ
ターゲットで実行するアクションタイプを選択します。 [詳細はこちら](#)

次のあと開始 - オプション
このアクションの前に実行するアクションを選択します。それ以外の場合、このアクションは実験が開始するとすぐに実行されます。

aws:cloudwatch:assert-alarm-state
Asserts that the CloudWatch alarms are in the expected states.

aws:ec2:reboot-instances
Reboot the specified EC2 instances.

aws:ec2:send-spot-instance-interruptions
Interrupt the specified EC2 Spot instances.

aws:ec2:stop-instances
Stop the specified EC2 instances.

aws:ec2:terminate-instances
Terminate the specified EC2 instances.

aws:ecs:drain-container-instances
Drain percentage of underlying EC2 instances on an ECS cluster.

aws:eks:terminate-nodegroup-instances
Terminates a percentage of the underlying EC2 instances in an EKS cluster.

[こちら](#)

<https://docs.aws.amazon.com/fis/latest/userguide/fis-tutorial-spot-interruptions.html>

スポットインスタンスの実践例

スポットインスタンスはこの使い方に向いていますか？

1. インスタンスタイプにm6i.largeのみを指定したASGに対して6台のスポットインスタンスを起動
2. プロプライエタリなデータベースソフトウェアを2台のスポットインスタンス上でActive-Activeに構成
3. デザイナーのためのVDIデスクトップの用途としてGPUインスタンスをスポットインスタンスとして起動
4. オンメモリで60分以上掛かる計算を実施する計算ノードとしてc6i.4xlargeで10台のスポットインスタンスを起動

スポットインスタンスはこの使い方に向いていますか？

1. インスタンスタイプにm6i.largeのみを指定した ASGに対して6台のスポットインスタンスを起動

単一インスタンス
タイプは避けましょう

2. プロプライエタリなデータベースソフトウェアを
2台のスポットインスタンス上でActive-Activeに構成

ステートフル

3. デザイナーのためのVDIデスクトップの用途として
GPUインスタンスをスポットインスタンスとして起動

ステートフル

4. オンメモリで60分以上掛かる計算を実施する
計算ノードとしてc6i.4xlargeで10台の
スポットインスタンスを起動

ステートフル

スポットインスタンスはこの使い方に向いていますか？

1. インスタンスタイプにm6i.largeのみを指定した ASGに対して6台のスポットインスタンスを起動

単一インスタンス
タイプは避けましょう

2. プロプライエタリなデータベースソフトウェアを
2台のスポットインスタンス上でActive-Activeに構成

ステートフル

3. デザイナーのためのVDIデスクトップの用途として
GPUインスタンスをスポットインスタンスとして起動

ステートフル

4. オンメモリで60分以上掛かる計算を実施する
計算ノードとしてc6i.4xlargeで10台の
スポットインスタンスを起動

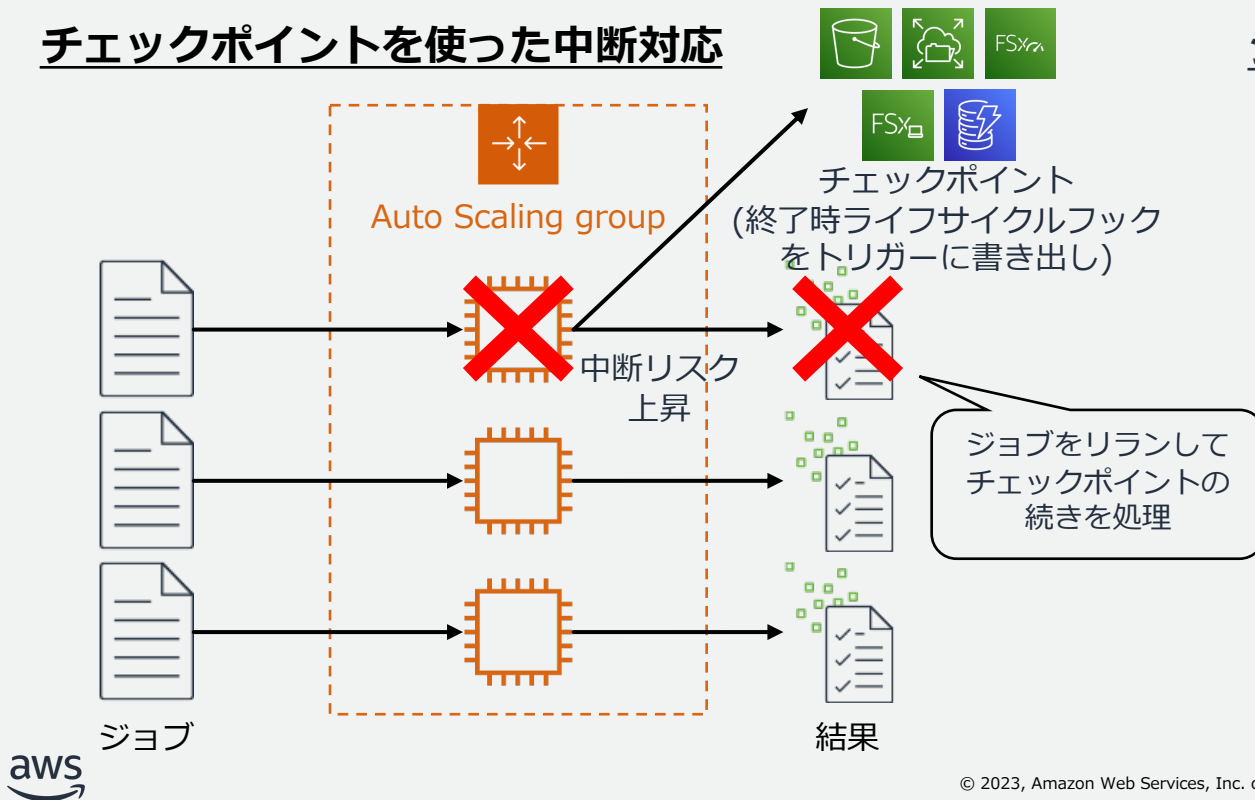
ステートフル

60分以上掛かる計算をスポットインスタンス向きに最適化する

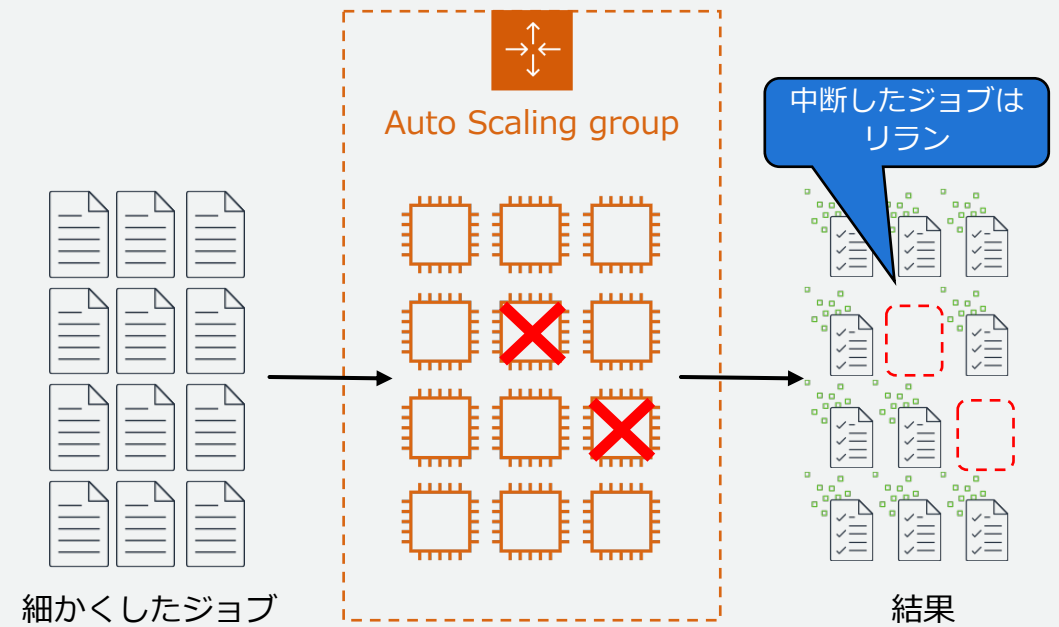
🔄 解決案

- 中断に備えるアクションを整える
- リトライに任せる

チェックポイントを使った中断対応



ジョブを細かくしてリトライに任せる対応



参考URL

- Amazon EC2 よくある質問 – スポットインスタンス
 - https://aws.amazon.com/jp/ec2/faqs/#Spot_Instances
- スポットインスタンスの配分戦略
 - https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/UserGuide/ec2-fleet-allocation-strategy.html
- EC2 スポットインスタンスの price-capacity-optimized 戦略のご紹介 | Amazon Web Services ブログ –
 - <https://aws.amazon.com/jp/blogs/news/introducing-price-capacity-optimized-allocation-strategy-for-ec2-spot-instances/>
- EC2 スポットを利用するうえでのベストプラクティス
 - https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/UserGuide/spot-best-practices.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!