



# Amazon EC2 スポットインスタンスの基礎

## 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

すでにAmazon EC2 オンデマンドインスタンスを利用して、スポットインスタンスの概要や特徴などについて知りたい方

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

# アジェンダ

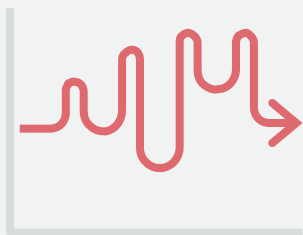
- EC2 スポットインスタンス概要
- スポットインスタンスの概念
- スポットインスタンスの価格と上限価格
- スポットインスタンスの起動方法
- お客様事例 - みらい翻訳様

# EC2 スポットインスタンス概要

# Amazon EC2 の購入オプション

## オンデマンド インスタンス

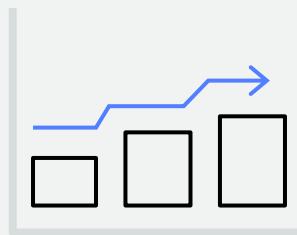
長期コミット不要  
使用分のみ支払い  
(秒単位/時間単位)  
Amazon EC2 の定価



スパイクするような  
ワークロード

## リザーブドインスタンス / Savings Plans

1年 / 3年の長期コミットに応じた  
**大幅なディスカウント**価格  
Savings Plans はリザーブドインスタ  
ンスの後継でより優れた柔軟性を提供



一定の負荷の見通しが  
あり長期コミット  
可能なワークロード

## スポット インスタンス

Amazon EC2 の空きキャ  
パシティを活用して**最大  
90% 値引き** (インスタ  
ンス中断あり)



中断に強く  
様々なインスタンス  
タイプを活用可能なワー  
クロード



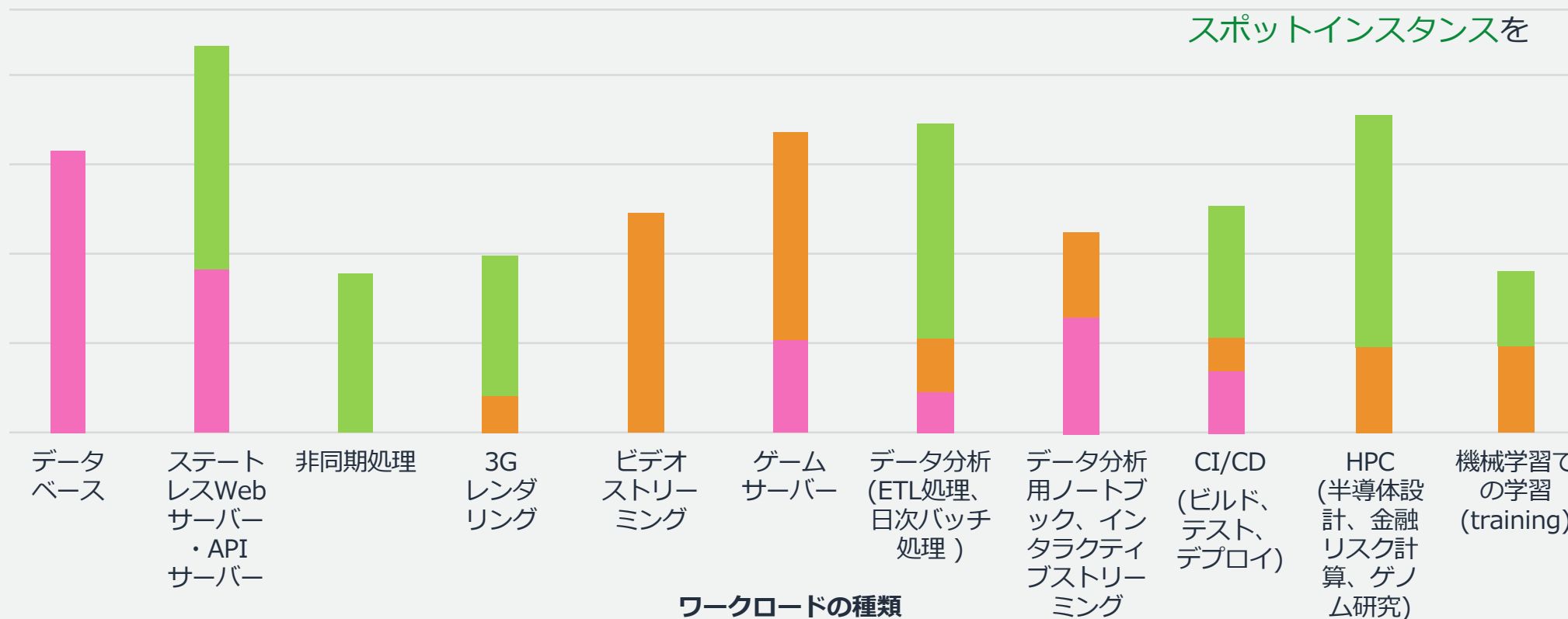
# ワークロード別 購入オプションの選び方の例

長く運用してきたもの、  
安定稼働しているものには  
RIやSavings Plansを

新しいものやステータフルな  
もの、スパイクするものには  
オンデマンドインスタンスを

複数のインスタンスタイプを活用  
できるもの、中断や一時的な障害に  
耐えられるものには  
スポットインスタンスを

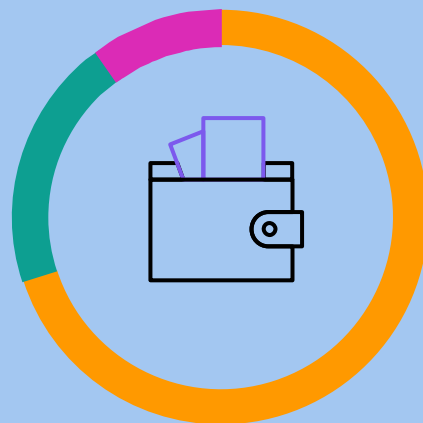
コンピューティングリソースの使用量



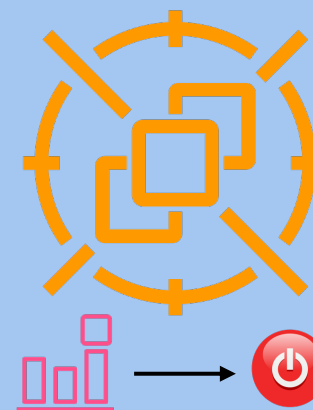
# スポットインスタンスとは



オンデマンドインスタンスや  
Savings Plans と  
共通のハードウェア基盤で提供

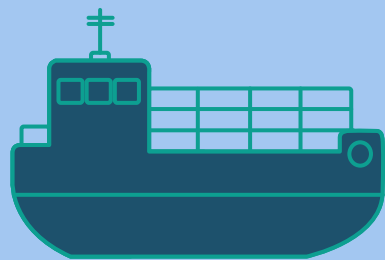


スポットインスタンスの  
価格は緩やかに変化

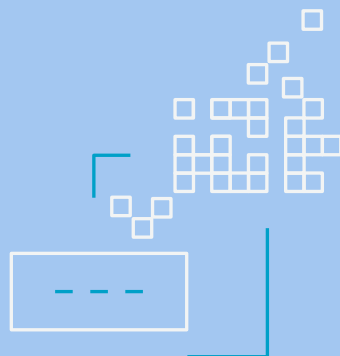


EC2 サービスの空き  
キャパシティに基づいて提供  
起動しない可能性や中断の可能性

# スポットインスタンスと親和性の高いワークロード



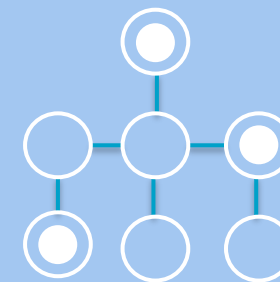
コンテナ



データ分析



CI/CD



並列分散バッチ処理



ウェブサービス



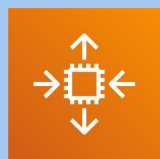
機械学習



高性能技術計算 (HPC)

キーワードは  
**「失敗に強い」**

# スポットインスタンスと連携できる サービスとソフトウェア



EC2 Auto  
Scaling



Amazon Elastic  
Container  
Service



Amazon Elastic  
Kubernetes  
Service



AWS  
Fargate



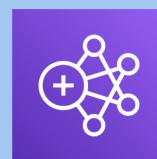
AWS  
Elastic  
Beanstalk



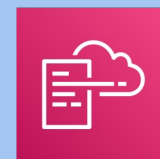
AWS  
Batch



Amazon  
SageMaker



Amazon  
EMR



AWS  
CloudFormatio  
n



AWS  
Thinkbox  
Deadline



Jenkins



HashiCorp  
Terraform



databricks™

cloudera



kubernetes



Bamboo



HashiCorp  
Packer



Qubole



MESOS



docker

# スポットインスタンスの特徴

## • 起動と中断

- EC2 サービスに空きキャパシティがあれば起動できる
- EC2 サービスに空きキャパシティがなければ起動できない
- EC2 サービスに空きキャパシティが不足してくると中断される

## • 価格

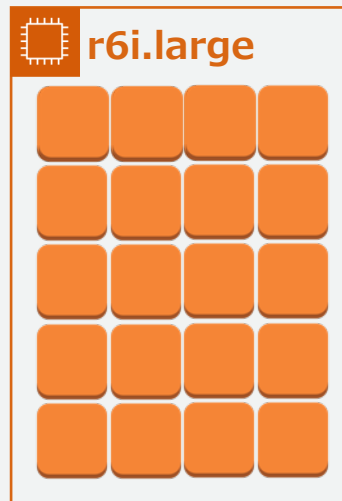
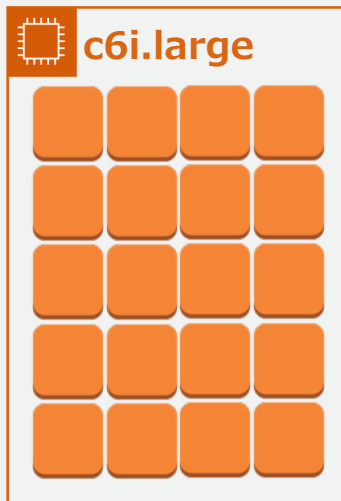
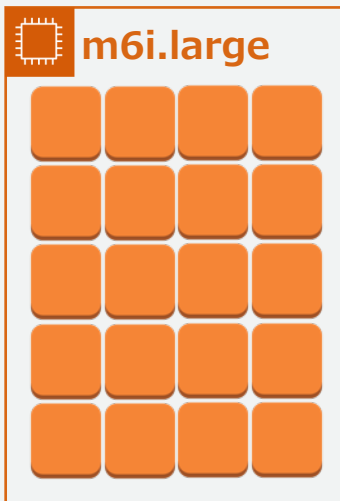
- スポットプールごとの需要と供給に応じて、徐々に調整される
- 上限価格を指定可能だが、設定不要
  - デフォルト値のオンデマンド価格を推奨

# スポットインスタンスの概念

# スポットインスタンスの概念 (1)

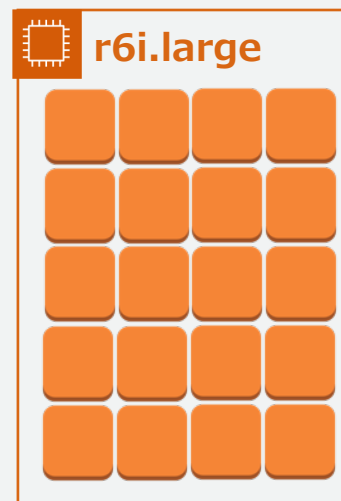
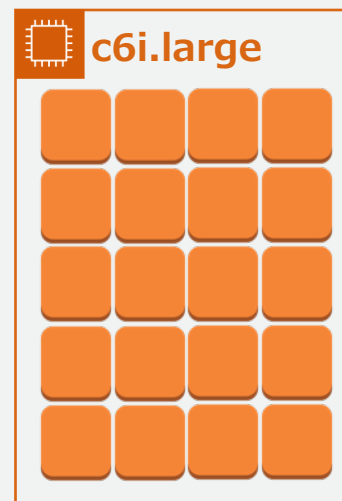
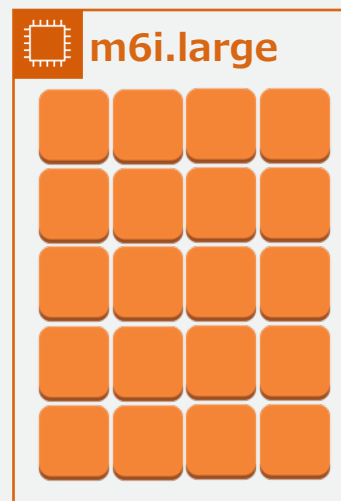
東京リージョン

ap-northeast-1a



...

ap-northeast-1c



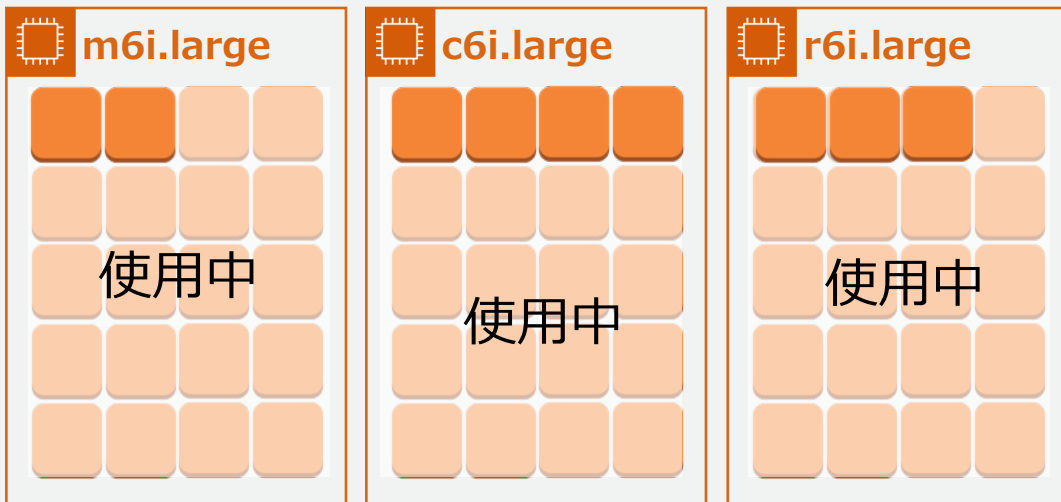
...

- 濃オレンジ色：未使用の空きキャパシティ

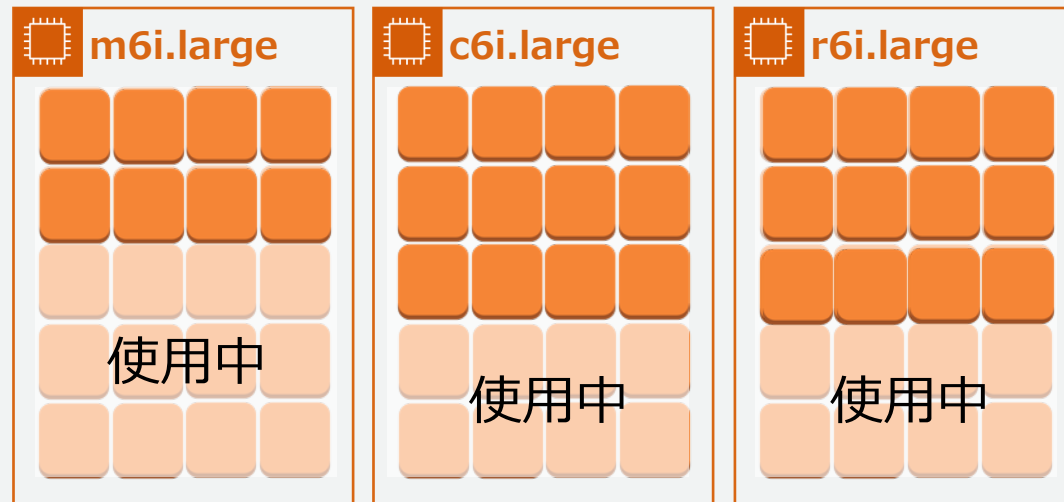
# スポットインスタンスの概念 (2)

東京リージョン

ap-northeast-1a



ap-northeast-1c



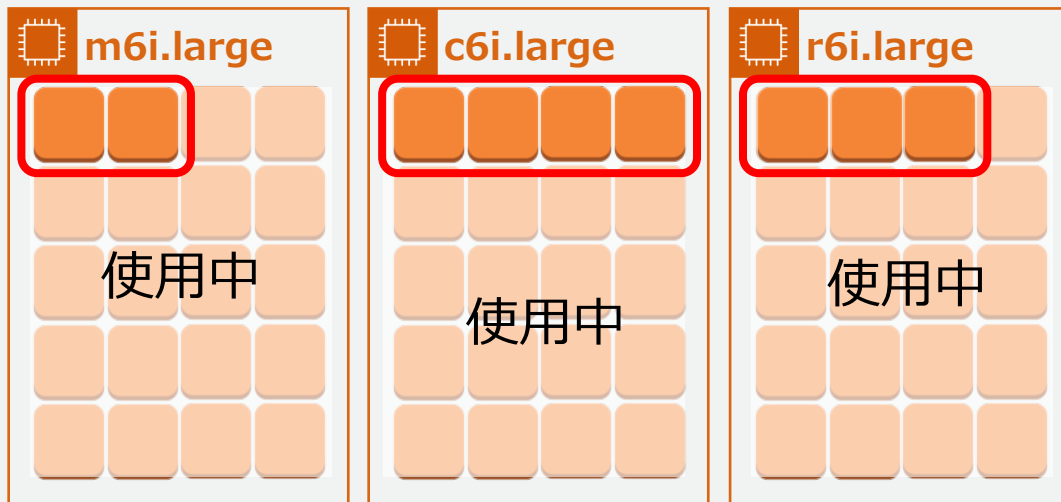
- 薄オレンジ色：お客様または AWS のマネージドサービスなどによって使用



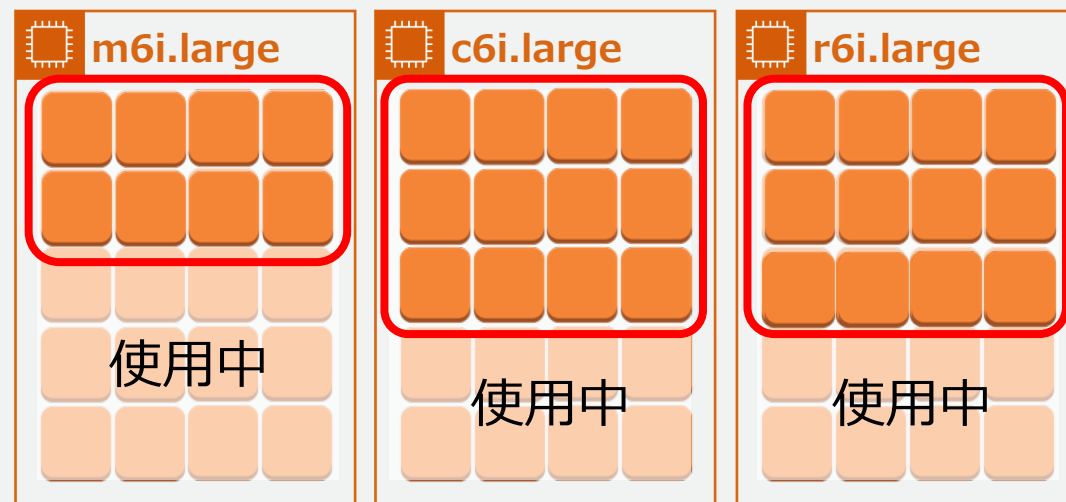
# スポットインスタンスの概念 (3)

東京リージョン

ap-northeast-1a



ap-northeast-1c

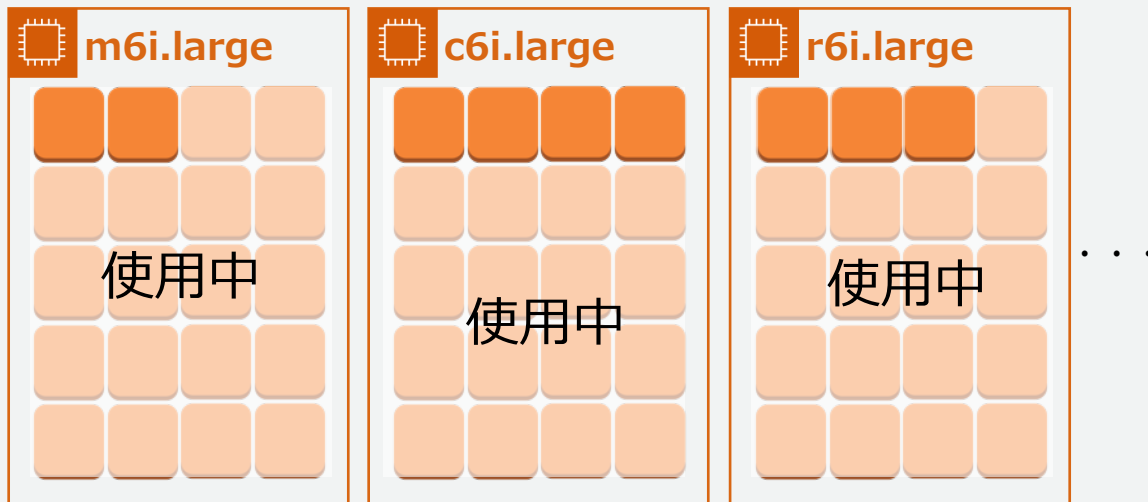


- スポットプール  
リージョン、アベイラビリティゾーン (AZ)、インスタンスタイプごとに独立した空きキャパシティ
- スポットインスタンスはスポットプールから起動される

# スポットインスタンスの概念 (4)

東京リージョン

ap-northeast-1a



ap-northeast-1a ゾーンには  
どのインスタンスタイプ  
(スポットプール)にも  
空きキャパシティがある



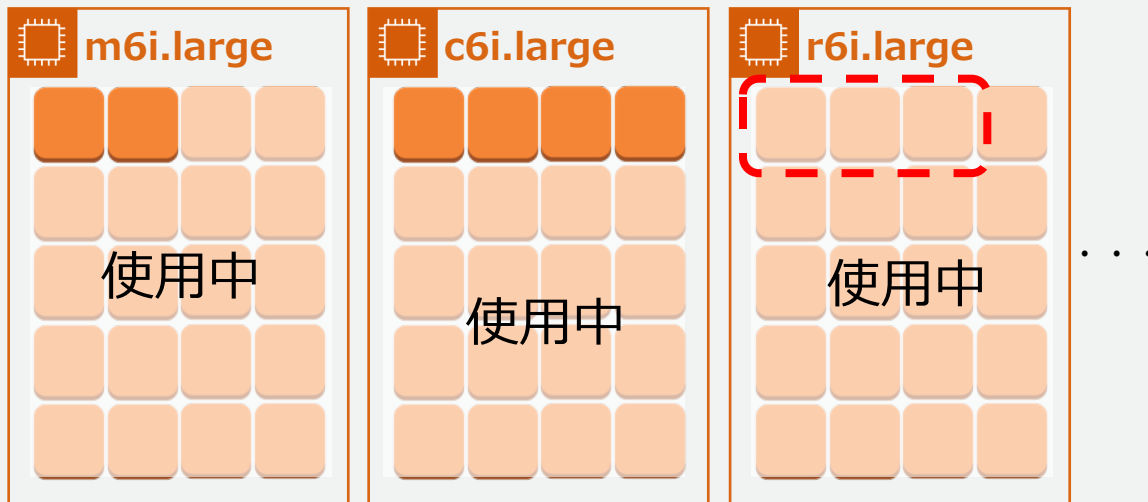
スポットインスタンスを起動できる

スポットプールの空きキャパシティがある → 起動できる

# スポットインスタンスの概念 (5)

東京リージョン

ap-northeast-1a



ap-northeast-1a ゾーンで  
r6i.large の空きキャパシティが不足



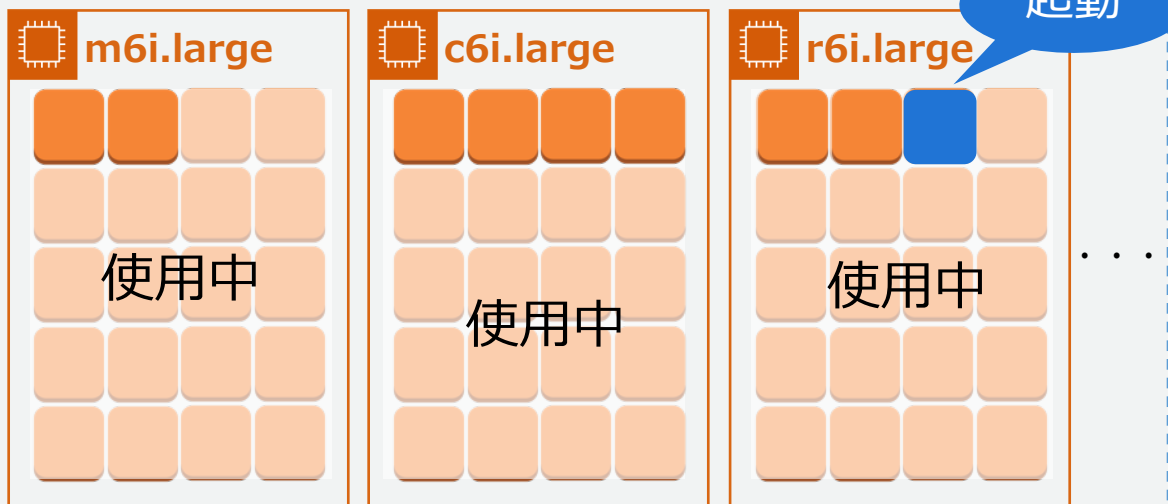
r6i.large のスポットインスタンスを  
起動できない

スポットプールに空きキャパシティがない → 起動できない

# スポットインスタンスの概念 (6)

東京リージョン

ap-northeast-1a



ap-northeast-1a ゾーンでは  
どのインスタンスタイプも  
空きキャパシティがある



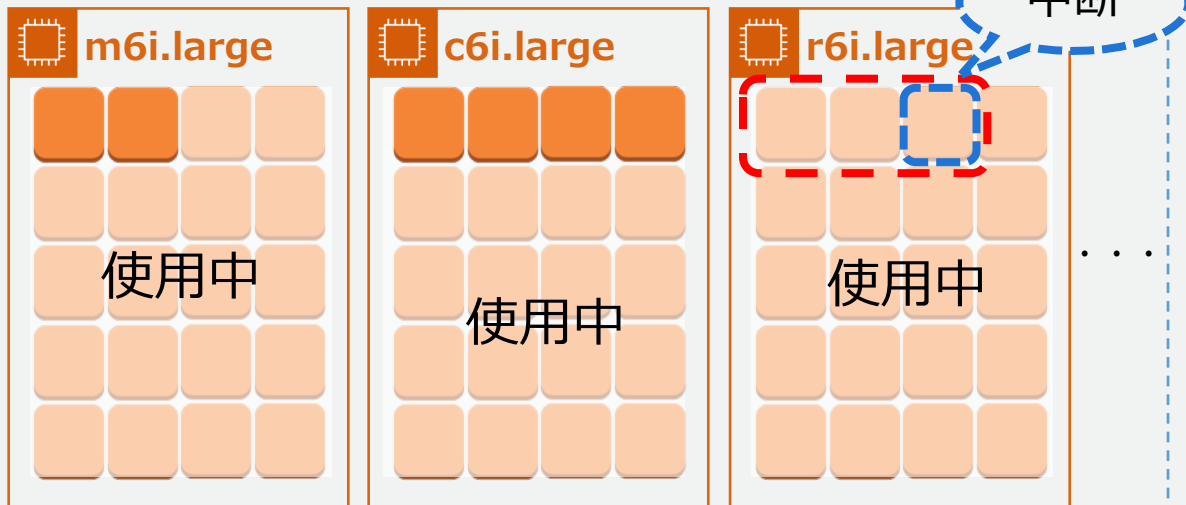
スポットインスタンスを起動できる

スポットプールに空きキャパシティがある → 起動できる

# スポットインスタンスの概念 (7)

東京リージョン

ap-northeast-1a



ap-northeast-1a ゾーンで  
r6i.large の空きキャパシティが不足



スポットインスタンスは中断される

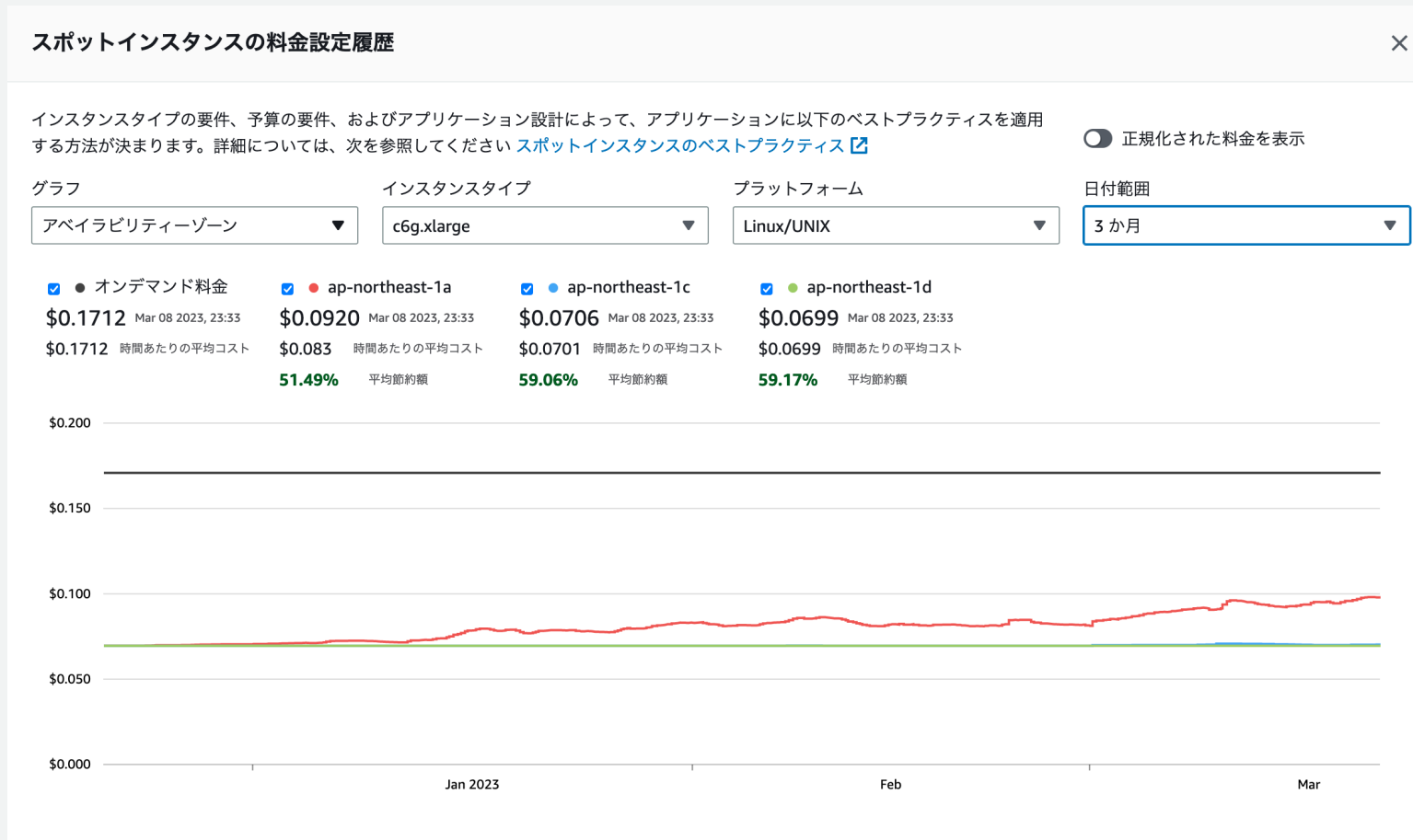
スポットプールに空きキャパシティがない

→ 起動中のスポットインスタンスは中断される (デフォルトでは終了)

# スポットインスタンスの価格

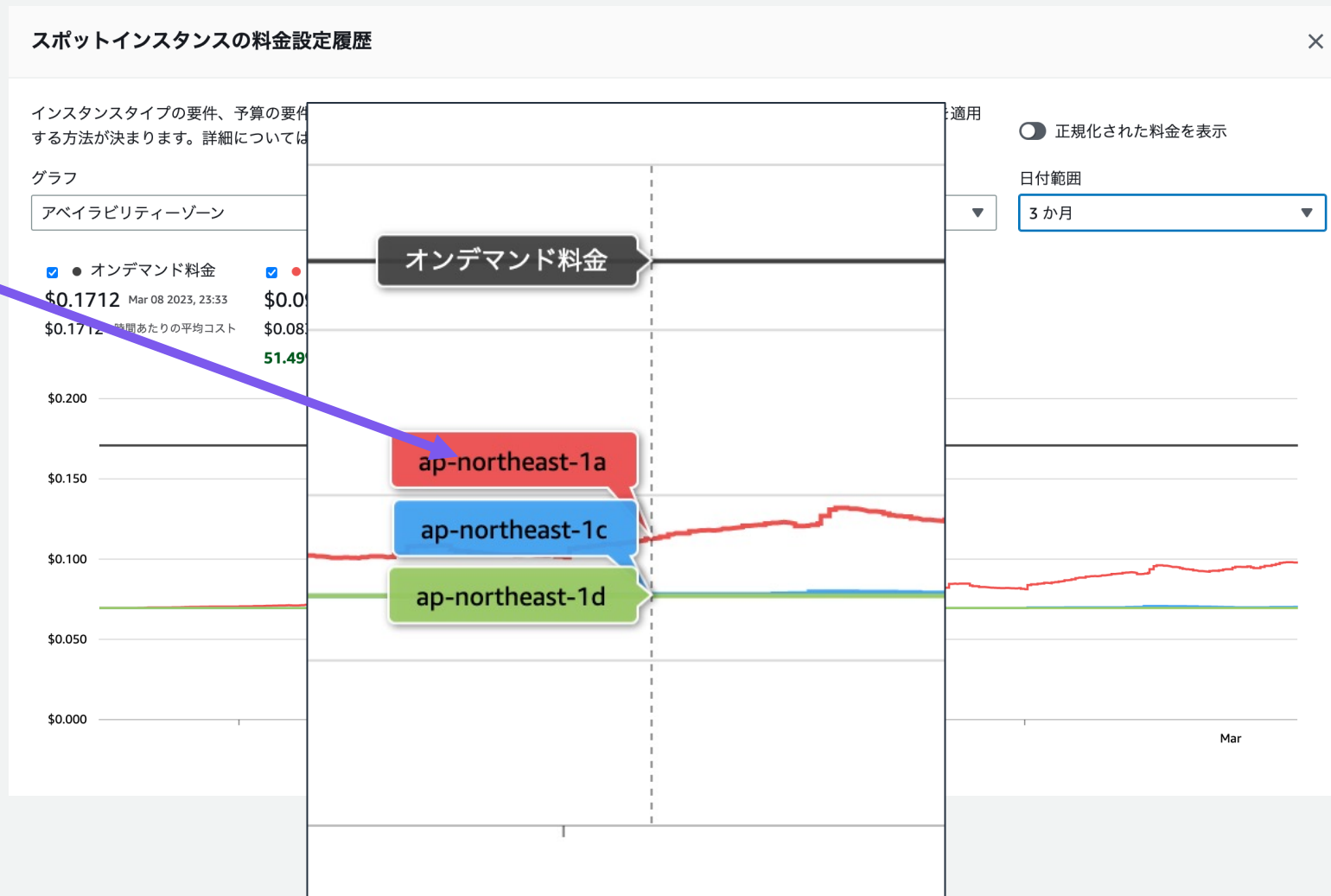
# スポットインスタンスの料金設定履歴

- EC2マネジメントコンソール「スポットリクエスト」 → 「料金設定履歴」



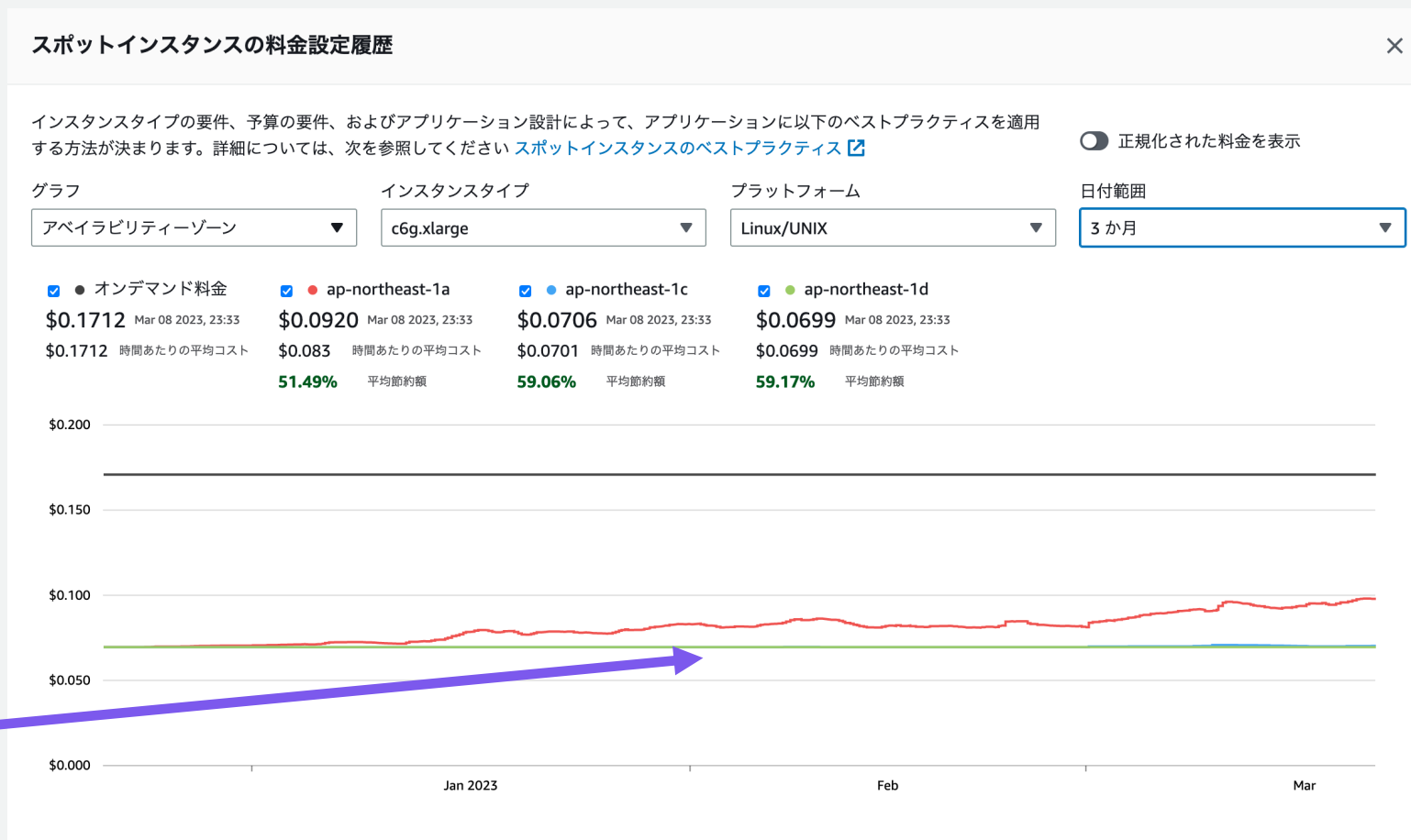
# スポットインスタンス価格の特徴 1

アベイラビリティゾーンごとに価格が異なる





# スポットインスタンス価格の特徴 2



緩やかに  
変化

# 上限価格

## スポットインスタンスの料金設定履歴

インスタンスタイプの要件、予算の要件、およびアプリケーション設計によって、アプリケーションに以下のベストプラクティスを適用する方法が決まります。詳細については、次を参照してください [スポットインスタンスのベストプラクティス](#)

正規化された料金を表示

グラフ

インスタンスタイプ

プラットフォーム

日付範囲

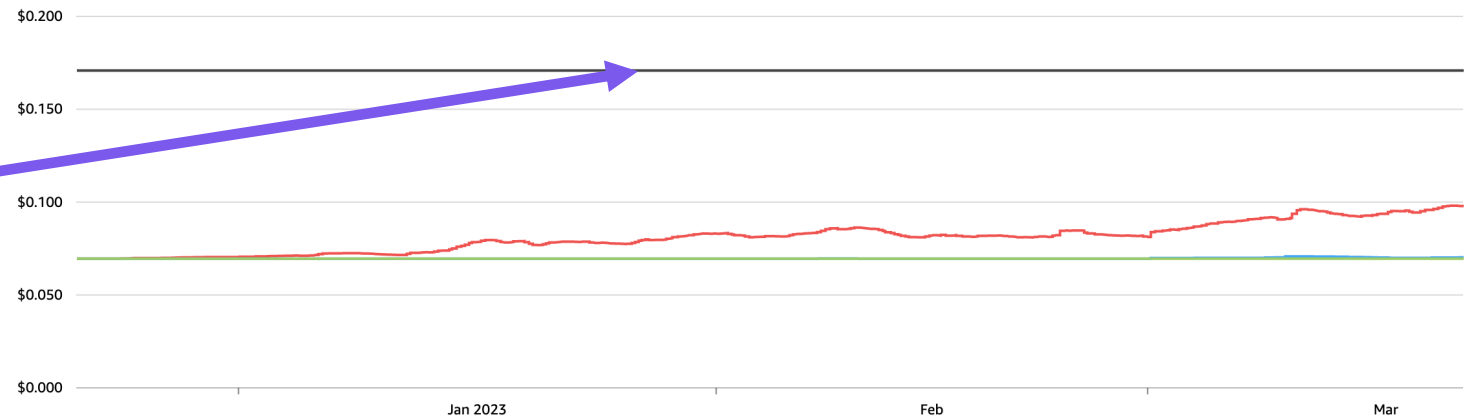
アベイラビリティゾーン

c6g.xlarge

Linux/UNIX

3 か月

<input checked="" type="checkbox"/> オンデマンド料金	<input checked="" type="checkbox"/> ap-northeast-1a	<input checked="" type="checkbox"/> ap-northeast-1c	<input checked="" type="checkbox"/> ap-northeast-1d
\$0.1712 Mar 08 2023, 23:33	\$0.0920 Mar 08 2023, 23:33	\$0.0706 Mar 08 2023, 23:33	\$0.0699 Mar 08 2023, 23:33
\$0.1712 時間あたりの平均コスト	\$0.083 時間あたりの平均コスト	\$0.0701 時間あたりの平均コスト	\$0.0699 時間あたりの平均コスト
	51.49% 平均節約額	59.06% 平均節約額	59.17% 平均節約額



上限価格

==

ここまで支払ってよい価格

# 上限価格

## スポットインスタンスの料金設定履歴

インスタンスタイプの要件、予算の要件、およびアプリケーション設計によって、アプリケーションに以下のベストプラクティスを適用する方法が決まります。詳細については、次を参照してください [スポットインスタンスのベストプラクティス](#)

正規化された料金を表示

グラフ

アベイラビリティゾーン ▼

インスタンスタイプ

c6g.xlarge ▼

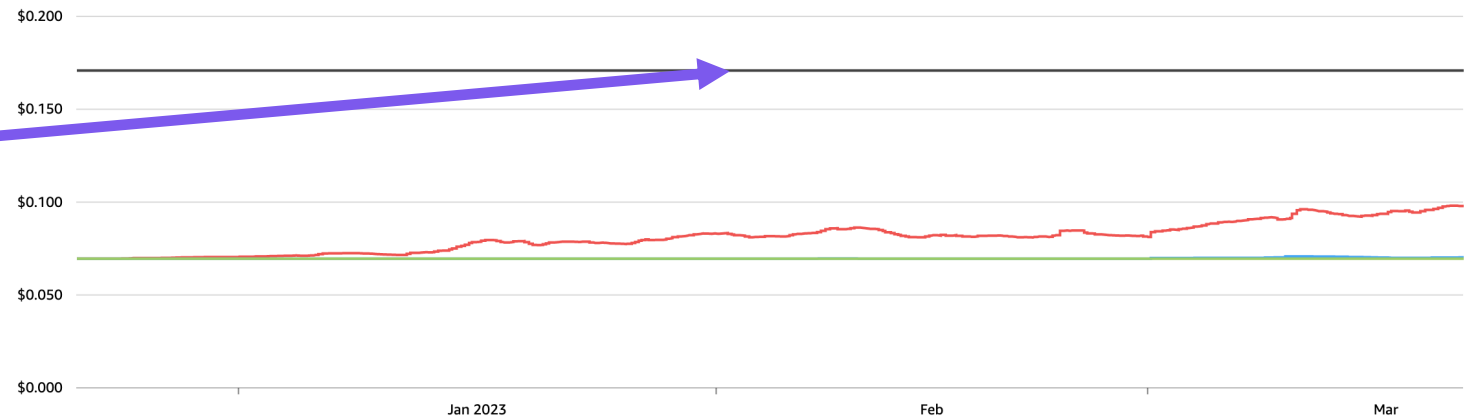
プラットフォーム

Linux/UNIX ▼

日付範囲

3 か月 ▼

<input checked="" type="checkbox"/> ● オンデマンド料金	<input checked="" type="checkbox"/> ● ap-northeast-1a	<input checked="" type="checkbox"/> ● ap-northeast-1c	<input checked="" type="checkbox"/> ● ap-northeast-1d
\$0.1712 Mar 08 2023, 23:33	\$0.0920 Mar 08 2023, 23:33	\$0.0706 Mar 08 2023, 23:33	\$0.0699 Mar 08 2023, 23:33
\$0.1712 時間あたりの平均コスト	\$0.083 時間あたりの平均コスト	\$0.0701 時間あたりの平均コスト	\$0.0699 時間あたりの平均コスト
	51.49% 平均節約額	59.06% 平均節約額	59.17% 平均節約額



上限価格のデフォルト値  
==  
オンデマンド料金

# 上限価格

## スポットインスタンスの料金設定履歴

インスタンスタイプの要件、予算の要件、およびアプリケーション設計によって、アプリケーションに以下のベストプラクティスを適用する方法が決まります。詳細については、次を参照してください [スポットインスタンスのベストプラクティス](#)

正規化された料金を表示

グラフ

インスタンスタイプ

プラットフォーム

日付範囲

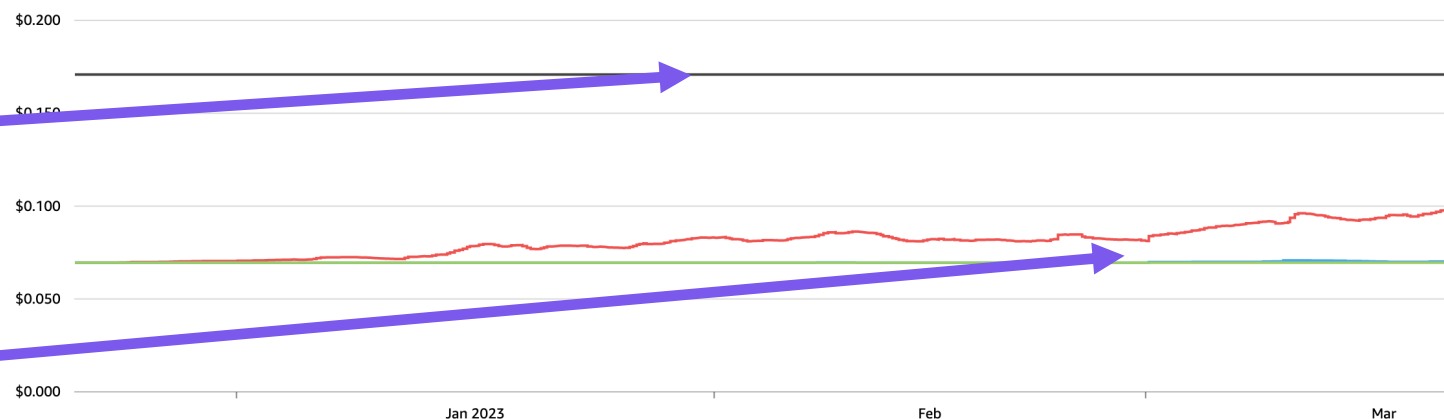
アベイラビリティゾーン

c6g.xlarge

Linux/UNIX

3 か月

<input checked="" type="checkbox"/> オンデマンド料金	<input checked="" type="checkbox"/> ap-northeast-1a	<input checked="" type="checkbox"/> ap-northeast-1c	<input checked="" type="checkbox"/> ap-northeast-1d
\$0.1712 Mar 08 2023, 23:33	\$0.0920 Mar 08 2023, 23:33	\$0.0706 Mar 08 2023, 23:33	\$0.0699 Mar 08 2023, 23:33
\$0.1712 時間あたりの平均コスト	\$0.083 時間あたりの平均コスト	\$0.0701 時間あたりの平均コスト	\$0.0699 時間あたりの平均コスト
	51.49% 平均節約額	59.06% 平均節約額	59.17% 平均節約額

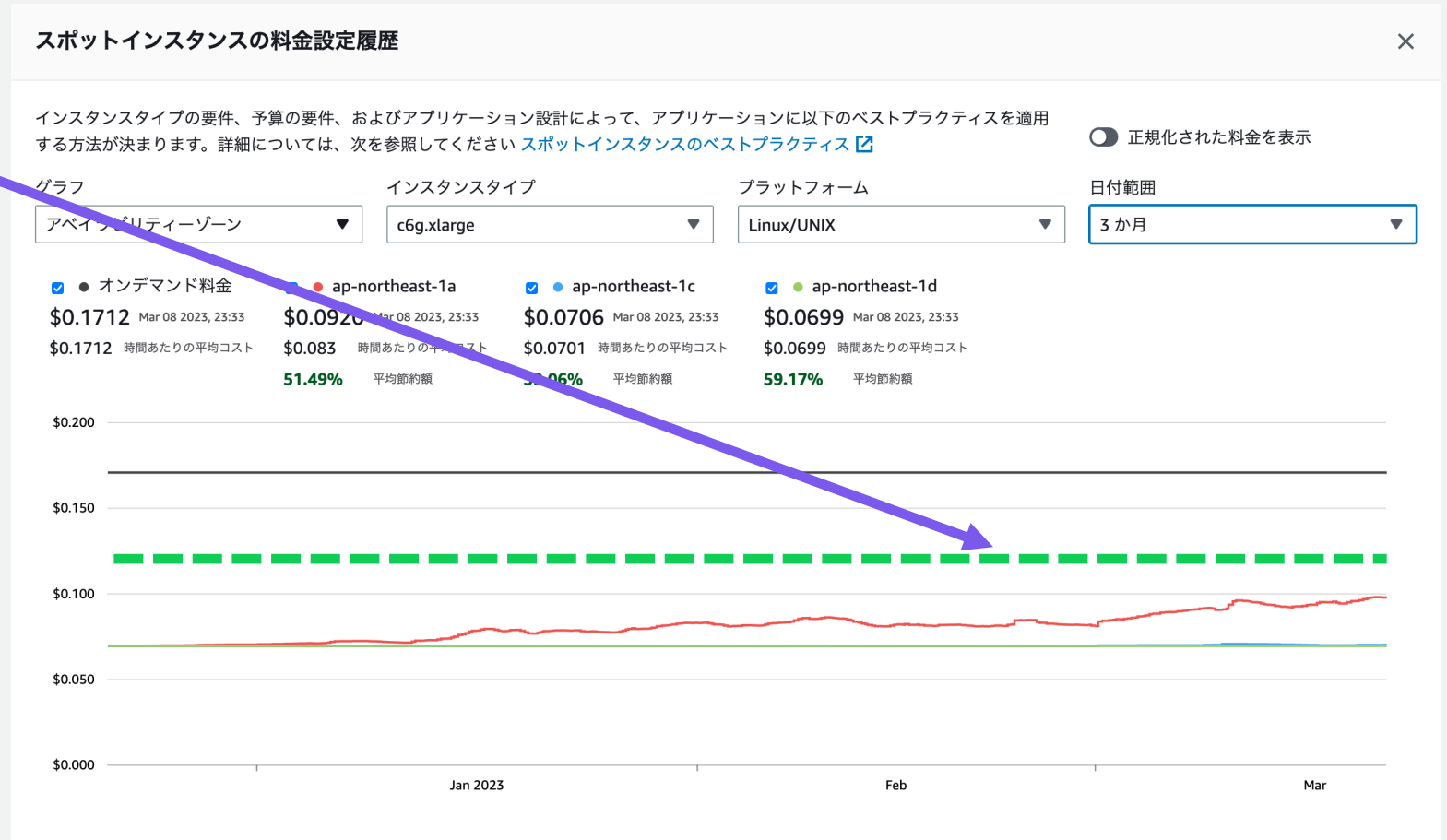


上限価格のデフォルト値  
==  
オンデマンド料金

実際に支払う  
スポット料金

# 上限価格

上限価格に任意の  
価格を指定できる。  
しかし！

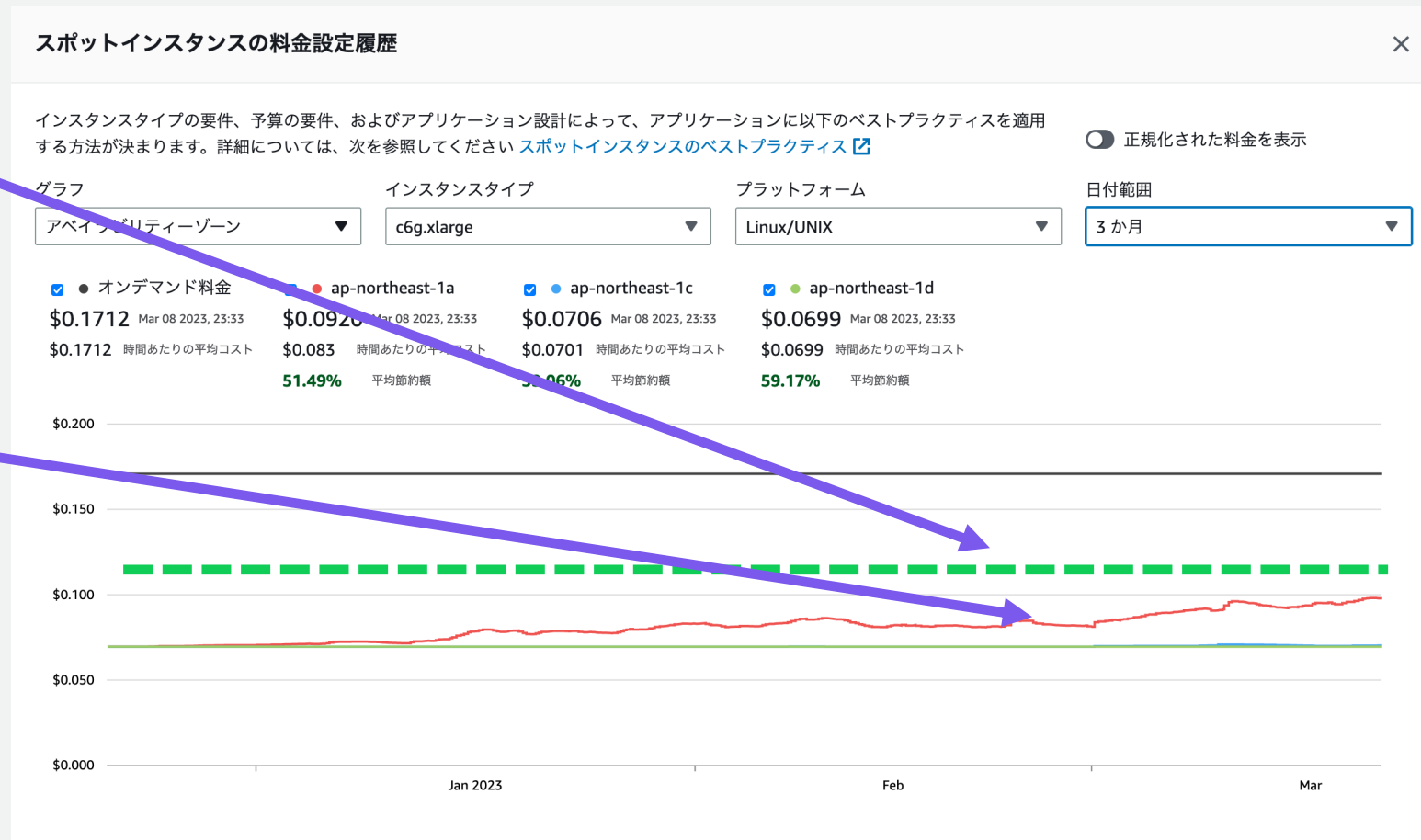


# 上限価格

上限価格に任意の  
価格を指定できる。  
しかし！

スポット料金 > 上限価格  
の場合

- 起動できません
- 中断されます



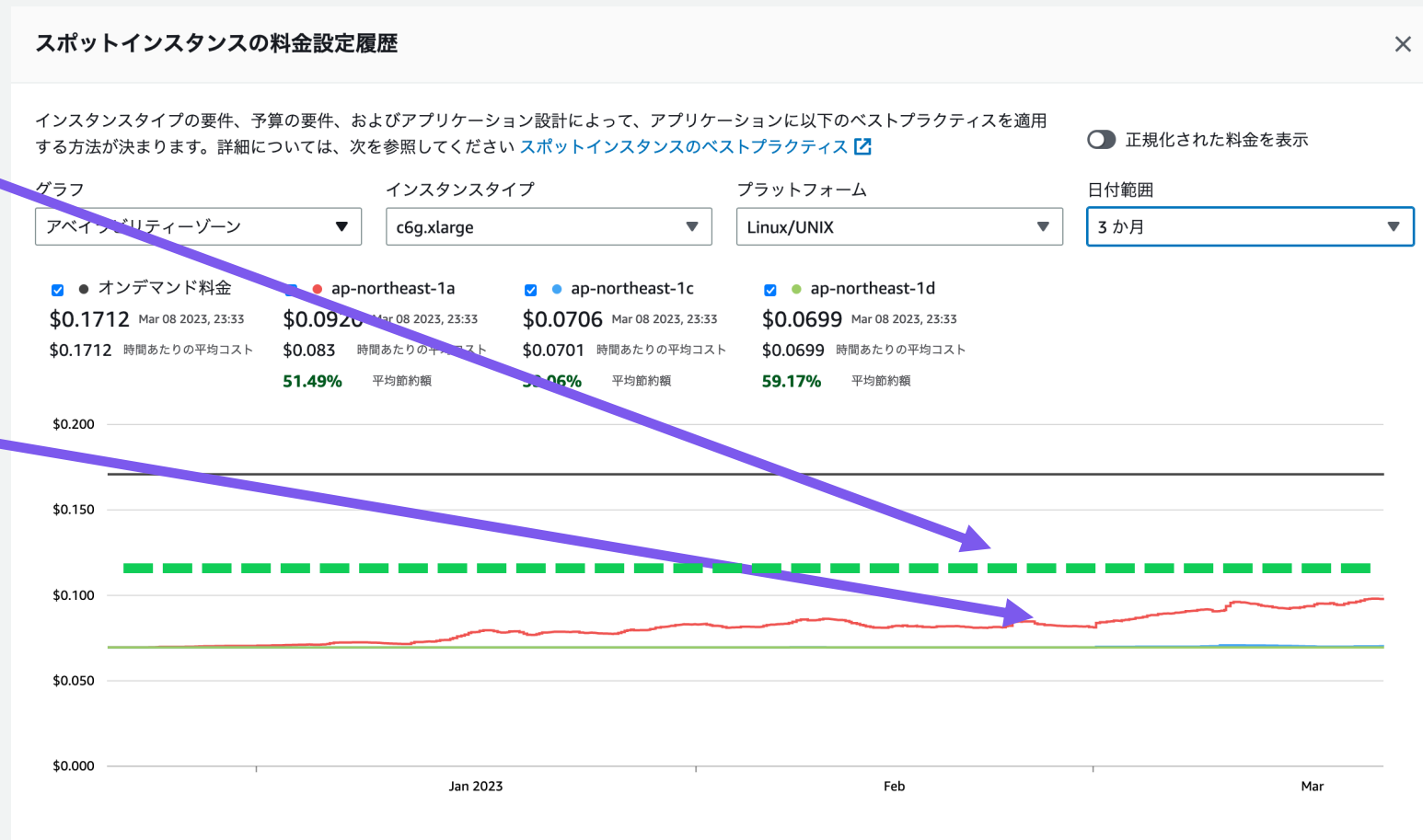
# 上限価格

上限価格に任意の  
価格を指定できる。  
しかし！

スポット料金 > 上限価格  
の場合

- 起動できません
- 中断されます

**上限価格は指定せず、  
デフォルト値の  
オンデマンド価格の採用を！**



# クイズ

Q.

上限価格にオンデマンド価格の10倍を指定します。  
スポットインスタンスは起動しやすくなりますか？



# クイズ

Q.

上限価格にオンデマンド価格の10倍を指定します。  
スポットインスタンスは起動しやすくなりますか？

A.

いいえ

起動のしやすさはEC2 サービスの需要、空きキャパシティだけで決まります

# スポットインスタンスアドバイザー

- 過去 30 日間の履歴データに基づいた、オンデマンドと比較した費用削減率と中断頻度を表示

スポットインスタンスアドバイザー

リージョン: アジアパシフィック (東京) OS: Linux/UNIX

インスタンスタイプフィルタ:

vCPU (分) 1      メモリ GiB (分) : 0       EMR でサポートされるインスタンスタイプ

インスタンスタイプ	vCPU	メモリ GiB	オンデマンドと比較した費用削減*	中断の頻度▼
c5.12xlarge	48	96	69%	<5% □□□□□
c5.xlarge	4	8	64%	<5% □□□□□
m5d.xlarge	4	16	75%	<5% □□□□□
m5.2xlarge	8	32	71%	<5% □□□□□
r5d.xlarge	4	32	79%	<5% □□□□□
i3en.2xlarge	8	64	70%	<5% □□□□□
c5n.2xlarge	8	21	71%	<5% □□□□□
c5d.large	2	4	73%	<5% □□□□□

**ただし、過去の中断頻度は将来の中断頻度を予測するためのものではないことに注意！**

<https://aws.amazon.com/jp/ec2/spot/instance-advisor/>

# スポットインスタンスの特徴

- 起動と中断

- EC2 サービスに空きキャパシティがあれば起動できる
- EC2 サービスに空きキャパシティがなければ起動できない
- EC2 サービスに空きキャパシティが不足してくると中断される

- 価格

- スポットプールごとの需要と供給に応じて、徐々に調整される
- 上限価格を指定可能だが、設定不要
  - デフォルト値のオンデマンド価格を推奨

# スポットインスタンスの起動方法

# スポットインスタンスの起動方法

	リクエスト単位	起動タイプ	容量変更	リクエストタイプと永続性
スポットインスタンス	インスタンス台数	インスタンス	×	one-time (sync) / persistent (sync)
スポットフリート (非推奨)	インスタンス台数 / vCPU数	フリート (インスタンス群)	○	request (async) / maintain (async)
EC2フリート	インスタンス台数 / vCPU数	フリート (インスタンス群)	○	instant (sync) / request (async) / maintain (async)
EC2 Auto Scaling	インスタンス台数 / vCPU数	フリート (インスタンス群)	○	n/a

# スポットインスタンスの起動

### ▼ 高度な詳細 [情報](#)

#### 購入オプション [情報](#)

スポットインスタンスをリクエスト  
オンデマンド料金を上限とするスポット料金でスポットインスタンスをリクエスト

[カスタマイズ](#)

#### ドメイン結合ディレクトリ [情報](#)

選択

[新しいディレクトリの作成](#)

#### IAM インスタンスプロフィール [情報](#)

選択

[新しい IAM プロファイルの作成](#)

#### ホスト名のタイプ [情報](#)

IP 名

#### DNS ホスト名 [情報](#)

IP 名 IPv4 (A レコード) DNS リクエストを有効化

リソースベースの IPv4 (A レコード) DNS リクエストを有効化

リソースベースの IPv6 (AAAA レコード) DNS リクエストを有効化

### ▼ 概要

インスタンス数 [情報](#)

1

[ソフトウェアイメージ \(AMI\)](#)

Amazon Linux 2023 AMI 2023.0.2...[続きを読む](#)  
ami-00cd61829b9bb9eca

[垂直サーバータイプ \(インスタンスタイプ\)](#)

m6g.medium

[ファイアウォール \(セキュリティグループ\)](#)

新しいセキュリティグループ

[ストレージ \(ボリューム\)](#)

1 ボリューム - 128 GiB

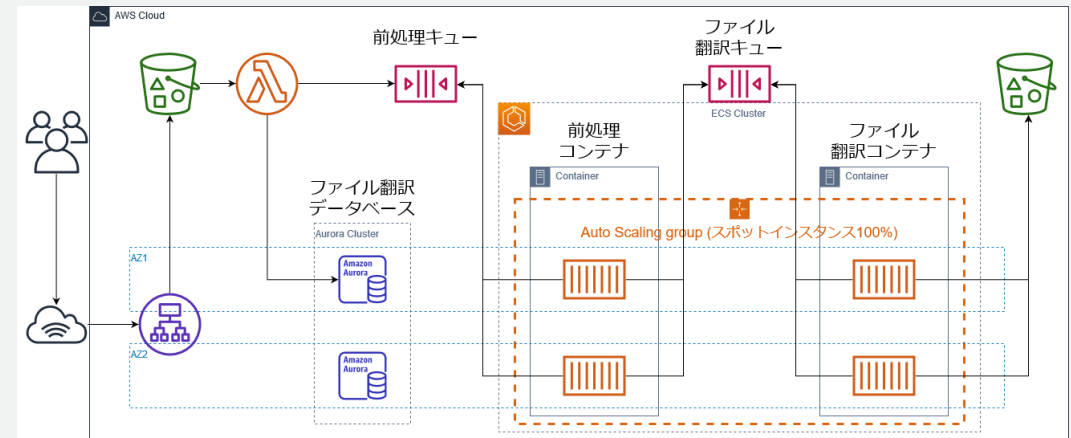
[キャンセル](#) [インスタンスを起動](#)

# お客様事例

# 株式会社みらい翻訳様

TOEIC960点レベルの日英翻訳、人手レベルの日中翻訳等の高精度な機械翻訳サービスをAWSで実現。  
高い可用性、性能、セキュリティを保ちつつ、EC2 スポットインスタンスの活用で費用を最適化。

- 日英・日中、ファイル・テキスト等、多様な形式の翻訳を行う推論環境をAWSで実現
- スポットインスタンス比率を100%にするため、従来のアプリケーションをコンテナ化。スポットインスタンスの中断に強いアーキテクチャを構築
- Amazon EC2 Auto ScalingやAmazon ECS、各種サーバーレスサービスを活用。コンテナ、EC2インスタンス、アベイラビリティゾーン障害も考慮したサービス継続性の高い環境を最適な費用で実現



## ステップ1: ユーザーからのリクエスト

- 翻訳対象ファイルをS3バケットに配置
- 配置をトリガーにLambda関数実行
- ファイル翻訳データベースに管理レコードを登録
- 翻訳対象の文章を作業メッセージとして前処理キューにエンキュー(登録)

## ステップ2: 前処理

- 前処理キューからメッセージ取得
- 一定時間の可視性タイムアウトを付与
- 前処理を実行
- 対象メッセージをデキュー(削除)
- 翻訳キューにエンキュー

## ステップ3: 翻訳

- 翻訳キューからメッセージ取得
- 一定時間の可視性タイムアウトを付与
- 翻訳処理を実行
- 対象メッセージをデキュー
- 翻訳済みファイルをS3バケットに配置

## ポイント

- 中断への対処: ワークロードの再開

SQSメッセージに可視性タイムアウトを指定。スポットインスタンスの中断等の理由で処理完了できなかった場合にも、タイムアウト経過後に再び可視となるため次に手の空いたコンテナタスクが自動的にリトライできる

- ワークロードに応じたキャパシティのスケール

各キューのメッセージ量に応じ、ECS、ASGでコンテナレベル・EC2インスタンスレベルのスケールアウト・スケールインを実現している



Amazon EC2 だけじゃない！最高のコスト効率を手に入れるための  
スポットインスタンス使いこなし術（本事例は後半に紹介されています）

<https://resources.awscloud.com/aws-summit-online-japan-2020-on-demand-aws-sessions-2-45123/aws-27-aws-summit-online-2020-720p>

# 参考URL

- Amazon EC2 スポットインスタンス
  - <https://aws.amazon.com/jp/ec2/spot/>
- 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](https://aws.amazon.com/jp/ec2/faqs/#Spot_Instances)
- インスタンスアドバイザー
  - <https://aws.amazon.com/jp/ec2/spot/instance-advisor/>
- Amazon EC2 ユーザーガイド – Spot Instances
  - [https://docs.aws.amazon.com/ja\\_jp/AWSEC2/latest/UserGuide/using-spot-instances.html](https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/UserGuide/using-spot-instances.html)
  - [https://docs.aws.amazon.com/ja\\_jp/AWSEC2/latest/WindowsGuide/using-spot-instances.html](https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/WindowsGuide/using-spot-instances.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!