



このコンテンツは公開から3年以上経過しており内容が古い可能性があります
最新情報については[サービス別資料](#)もしくはサービスのドキュメントをご確認ください

[AWS Black Belt Online Seminar]

Amazon EC2スポットインスタンス

サービスカットシリーズ

滝口 開資

ソリューションアーキテクト

EC2スポットインスタンススペシャリスト

2019-03-06

AWS 公式 Webinar

<https://amzn.to/JPWebinar>



過去資料

<https://amzn.to/JPArchive>



自己紹介

滝口 開資 (はるよし)

ソリューションアーキテクト / EC2スポットインスタンススペシャリスト
日本市場でのEC2スポットインスタンス製品担当

普段の業務

EC2スポットインスタンスを活用するお客様を
技術面からサポート

好きなAWSサービス

- Amazon EC2 Auto Scaling
- AWS Support



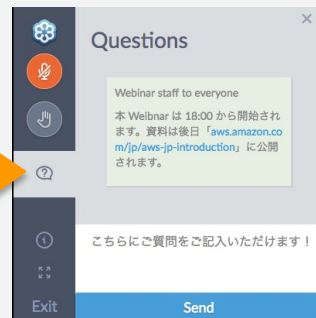
AWS Black Belt Online Seminar とは

「サービス別」「ソリューション別」「業種別」のそれぞれのテーマに分かれて、アマゾンウェブサービス ジャパン株式会社が主催するオンラインセミナーシリーズです。

質問を投げることができます！

- 書き込んだ質問は、主催者にしか見えません
- 今後のロードマップに関するご質問は
お答えできませんのでご了承下さい

- ① 吹き出しをクリック
- ② 質問を入力
- ③ Sendをクリック



Twitter 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2019年3月6日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(<http://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が東京リージョンを使用する場合、別途消費税をご請求させていただきます。
- AWS does not offer binding price quotes. AWS pricing is publicly available and is subject to change in accordance with the AWS Customer Agreement available at <http://aws.amazon.com/agreement/>. Any pricing information included in this document is provided only as an estimate of usage charges for AWS services based on certain information that you have provided. Monthly charges will be based on your actual use of AWS services, and may vary from the estimates provided.

今日わかるようになること

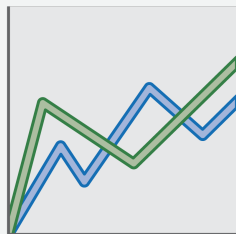
- スポットインスタンスは最大()割引。()が発生する
- アンチパターン
 - ミツ()ョンクリティカル
 - ステ()トフル
- 中断通知はインスタンスメタデータ、もしくはCloudWatch ()で取得
- 使いこなすために
 - ステ()トレス：状態を持たせない
 - ()開可能
 - ()結合
 - 分散：複()のインスタンスタイプを活用
 - (入札でなく)「上()価格」にはオンデマンド価格を指定

はじめに - スポットインスタンスの概要

はじめに - Amazon EC2の購入オプション

オンデマンドインスタンス

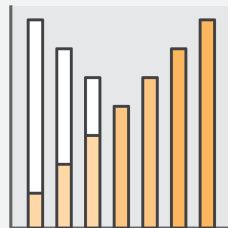
長期コミット無し、使用分への支払い(秒単位/時間単位)。Amazon EC2の定価



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

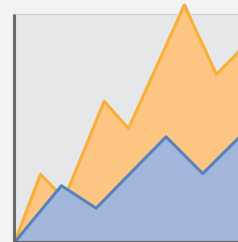
1年/3年の長期コミットをする代わりに大幅なディスカウント価格



一定の負荷の見通しがあるワークロード

スポットインスタンス

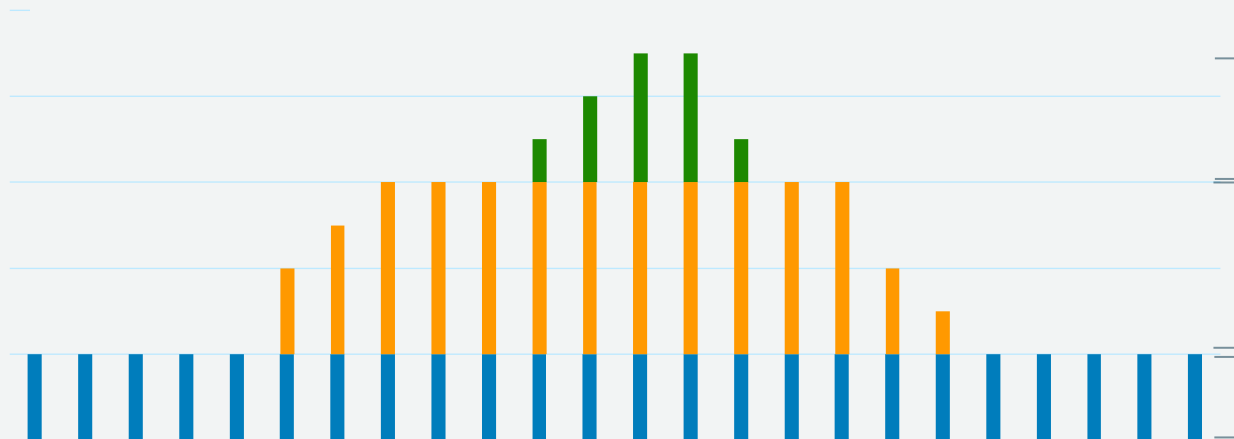
Amazon EC2の空きキャパシティを活用し、最大90%値引き



中断に強く、かつ様々なインスタンスタイプを活用できるワークロード

EC2インスタンスとしての性能に違いはない

EC2購入オプションをどう組み合わせるか？



新規のワークロード、あるいはステートフルなワークロードには
オンデマンドインスタンス

中断に強く、スケールするワークロードには
スポットインスタンス

一定の負荷が見込めるワークロードには
リザーブドインスタンス

Amazon EC2スポットインスタンスの特徴

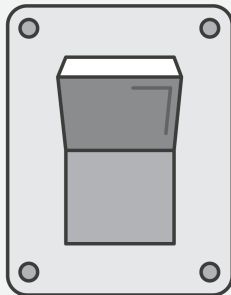
低コスト

- 最大9割引
- 中断に強いワークロードに最適 - ビッグデータ、コンテナ、HPC/グリッド、レンダリング



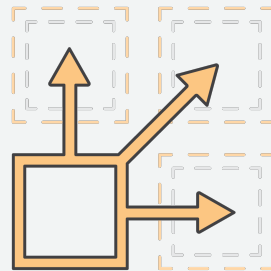
容易な起動方法

- 一度のリクエストで1,000台以上を起動可能
- マネジメントコンソール
- 用途に応じたAPIをご提供



柔軟な選択肢

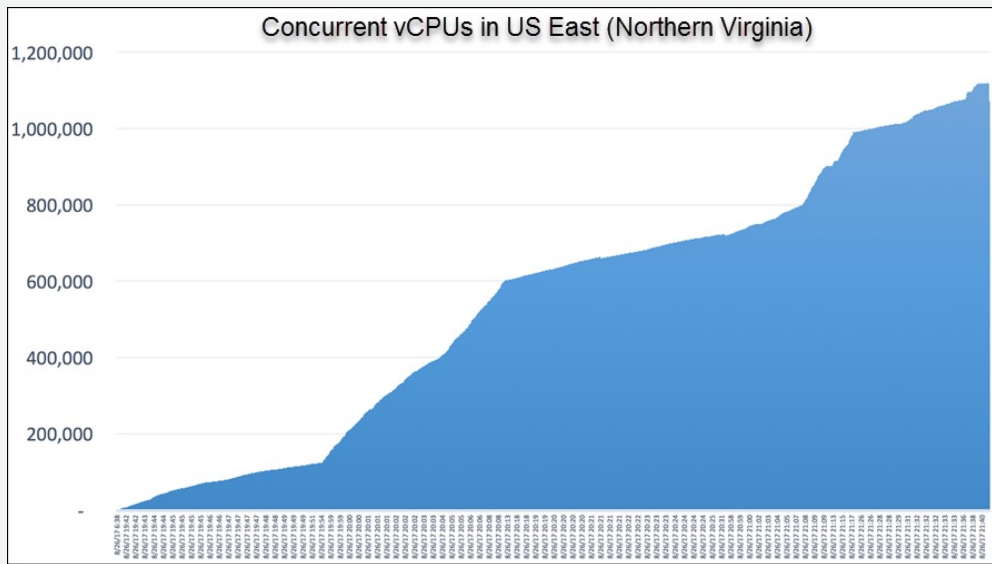
- 一度のリクエストでオンデマンド、リザーブド、スポットをまとめて指定可能
- 複数のインスタンスタイプを指定可能



スポットインスタンスの価値と 活用事例

クレムゾン大学 - 110万コアを活用した事例

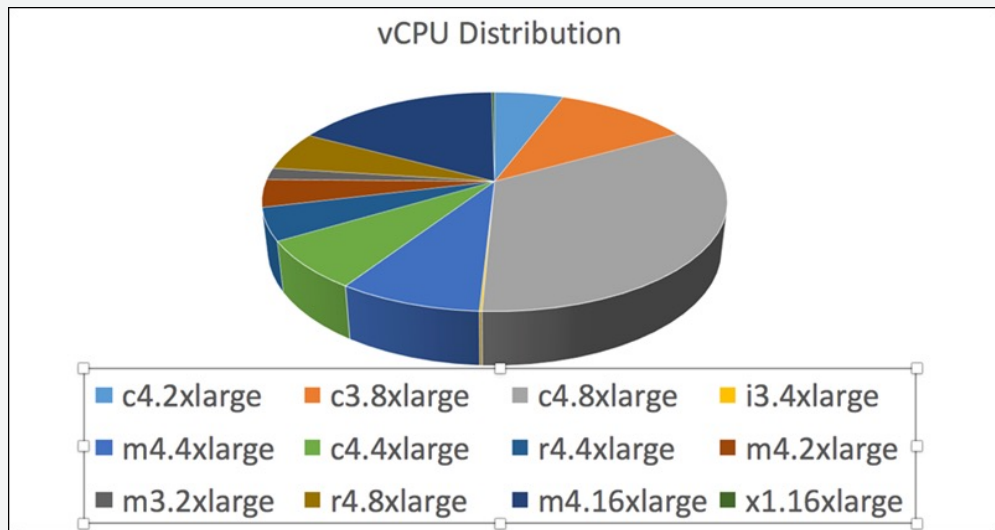
- トピック推定の研究(自然言語処理)
 - 文書集合を元にトピックを推定できるように学習
 - 学習プロセスをAWSで並列化して実行
- スポットインスタンスを活用
 - 2017年8月26日、同時稼働vCPU数が110万コアを達成(右図)
 - スーパーコンピューターも含めた中で世界最大規模のクラスター
 - 米国東部(バージニア北部)リージョンのみ使用



<https://aws.amazon.com/blogs/aws/natural-language-processing-at-clemson-university-1-1-million-vcpus-ec2-spot-instances/>

クレムゾン大学 - 110万コアを活用した事例

- トピック推定の研究(自然言語処理)
 - 文書集合を元にトピックを推定できるように学習
 - 学習プロセスをAWSで並列化して実行
- スポットインスタンスを活用
 - 2017年8月26日、同時稼働vCPU数が110万コアを達成(右図)
 - スーパーコンピューターも含めた中で世界最大規模のクラスタ
 - 米国東部(バージニア北部)リージョンのみ使用



<https://aws.amazon.com/blogs/aws/natural-language-processing-at-clemson-university-1-1-million-vcpus-ec2-spot-instances/>

様々なインスタンスタイプを存分に活用

スポットインスタンスの活用事例

ビッグデータ

コンテナ & test/dev

HPC & グリッド

ウェブサービス



FINRA はEC2スポットインスタンスを使うことで、オンプレ比で最大50%のコスト削減と柔軟性/拡張性を実現し、数ヶ月の処理を数日にまで短縮。

Yelp はEC2スポットインスタンスで毎日数百万のテストを実施。テスト実行時間を2日から30分に短縮し、費用も大幅に削減。

TLG Aerospace はEC2スポットインスタンスを流体(CFD)シミュレーションで利用し75%のコスト削減を実現。コスト削減は顧客へのメリットにつながった。

AdRoll はEC2スポットインスタンスを使うことで、世界中のお客様へシームレスなスケールのインフラを提供でき、固定費を75%、運用費を83%削減。

スポットインスタンスのアンチパターン

- 高可用性を求められるミッションクリティカルなワークロード
 - データベース(オンライントランザクション)
- チェックポイントを設けにくいステートフルなワークロード
 - マルチプレイヤーのゲームステージ

連携するAWSサービス、OSS, パートナー製品



Auto Scaling



AWS Batch



Amazon
EMR



Amazon
ECS



AWS
OpsWorks



AWS
CloudFormation



AWS Data
Pipeline



AWS Thinkbox
Deadline



Apache
MESOS™



Jenkins



cloudera

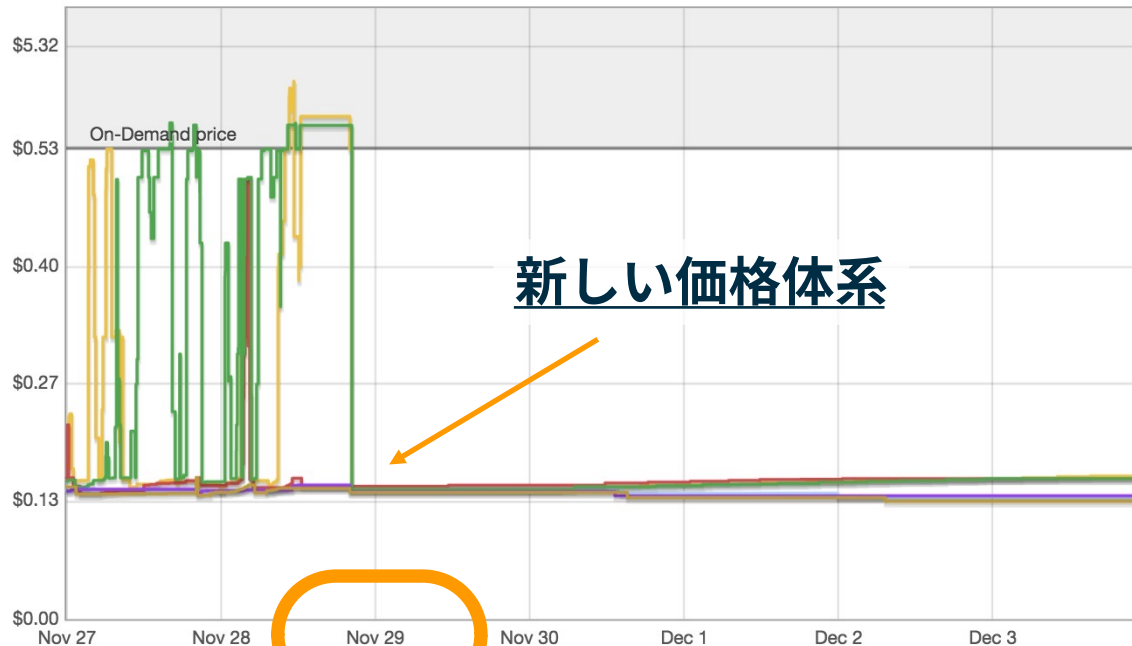


スポットインスタンスのしくみ とルール

スポット価格安定化 - 2017年11月

Spot Instance Pricing History

Product: Linux/UNIX (Am) Instance type: r4.2xlarge Date range: 1 week



Date

11/29/2017
1:42:22 PM UTC-0800

On-Demand price

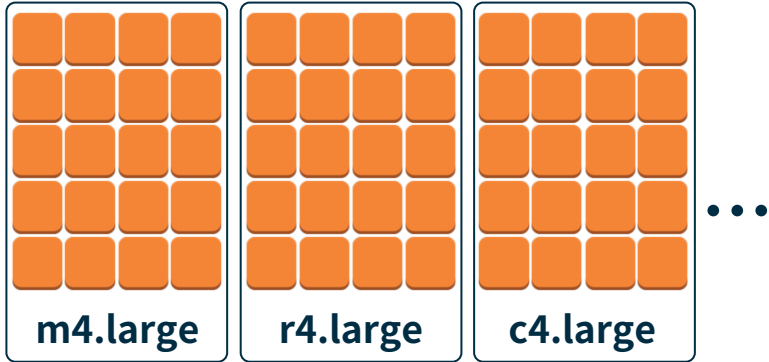
\$0.5320

Availability Zone

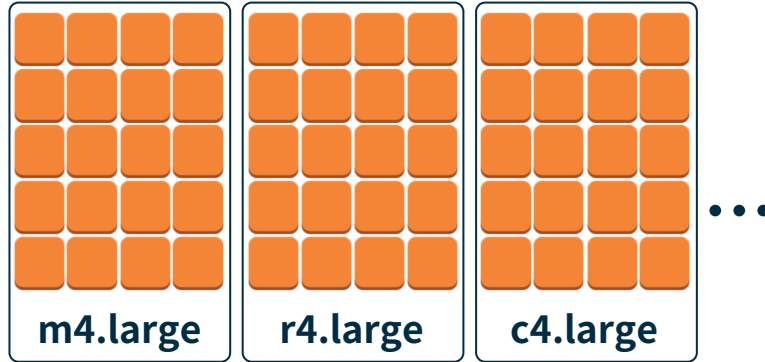
Availability Zone	Price
us-east-1a	\$0.1470
us-east-1b	\$0.1486
us-east-1c	\$0.1515
us-east-1d	\$0.1470
us-east-1e	\$0.1445
us-east-1f	\$0.1434

空きキャパシティとスポット価格(1)

AWS



ap-northeast-1a

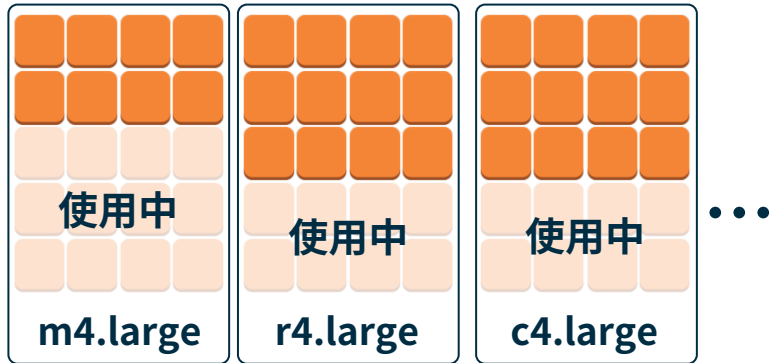


ap-northeast-1c

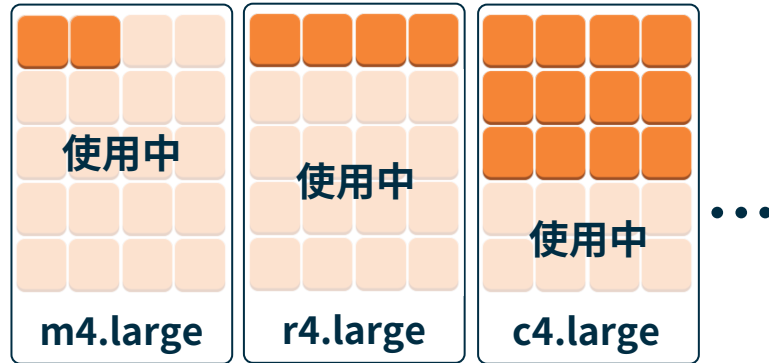
(Tokyo Region)

空きキャパシティとスポット価格(2)

AWS



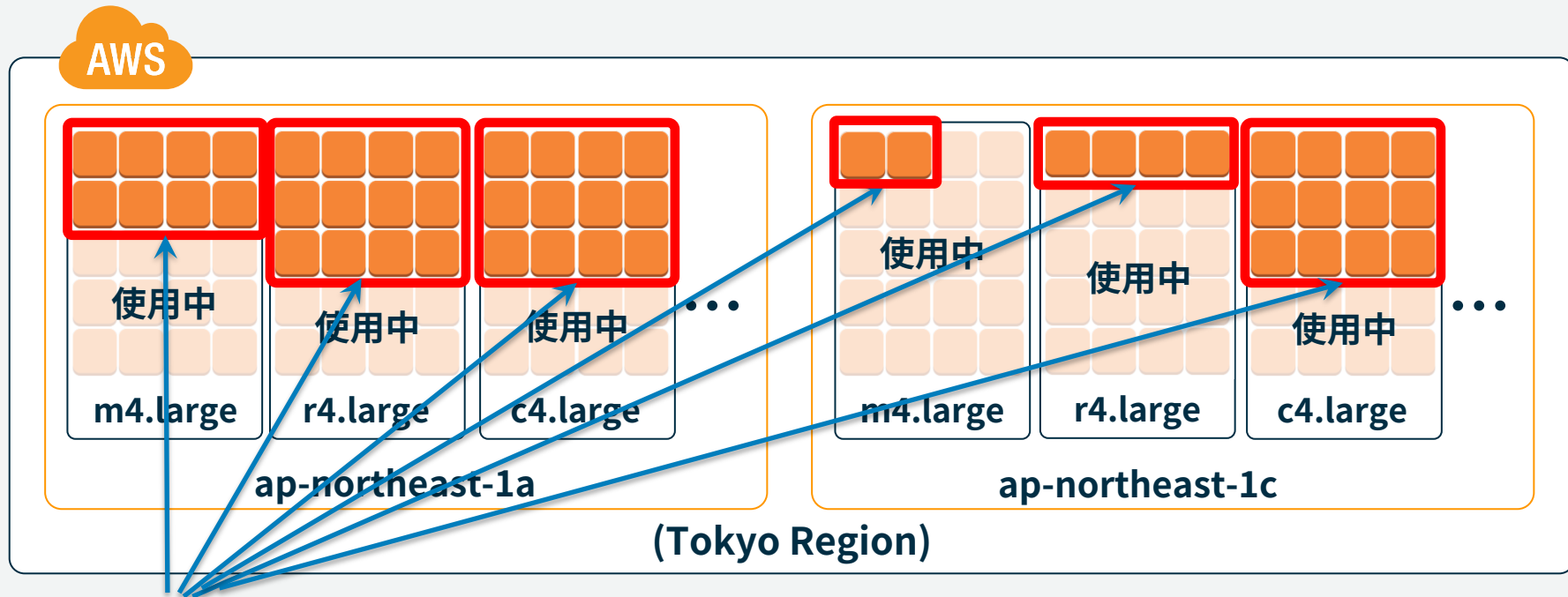
ap-northeast-1a



ap-northeast-1c

(Tokyo Region)

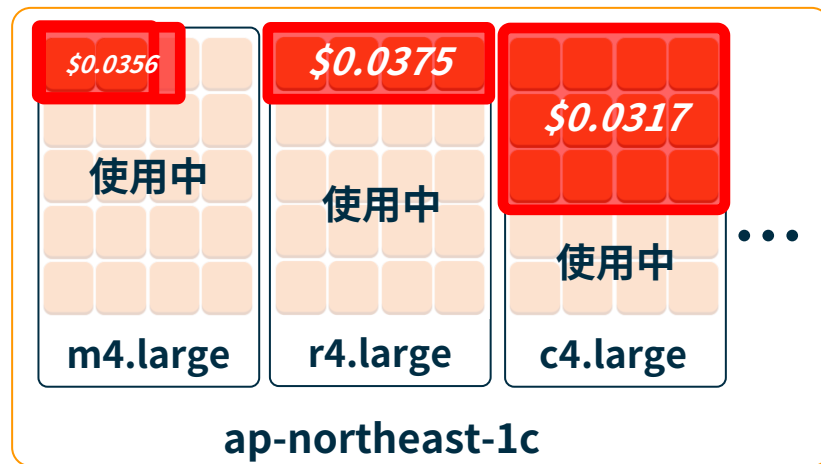
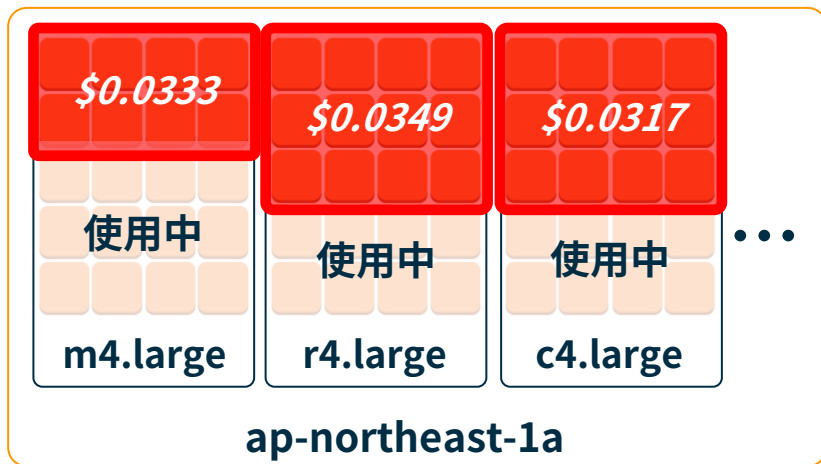
空きキャパシティとスポット価格(3)



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

空きキャパシティとスポット価格(4)

AWS



(Tokyo Region)

スポット価格…その時点のスポットインスタンス価格。スポットプールごとの需要と供給で決まる

スポットインスタンスの起動条件(1)

AWS

オンデマンド価格
\$0.1260

\$0.0333	\$0.0349	\$0.0317
使用中	使用中	使用中
m4.large	r4.large	c4.large

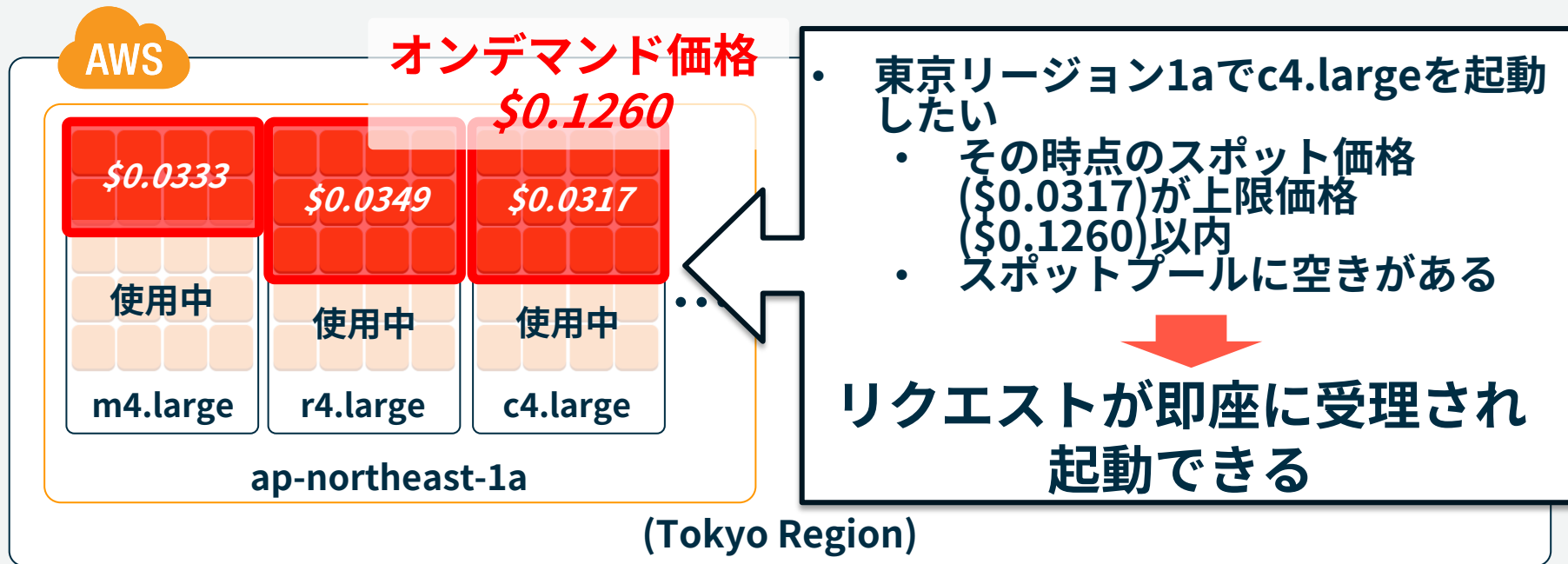
ap-northeast-1a

(Tokyo Region)

- 東京リージョン1aでc4.largeを起動したい
- リクエスト時に「上限価格」を指定：最大\$0.1260まで支払っても良い

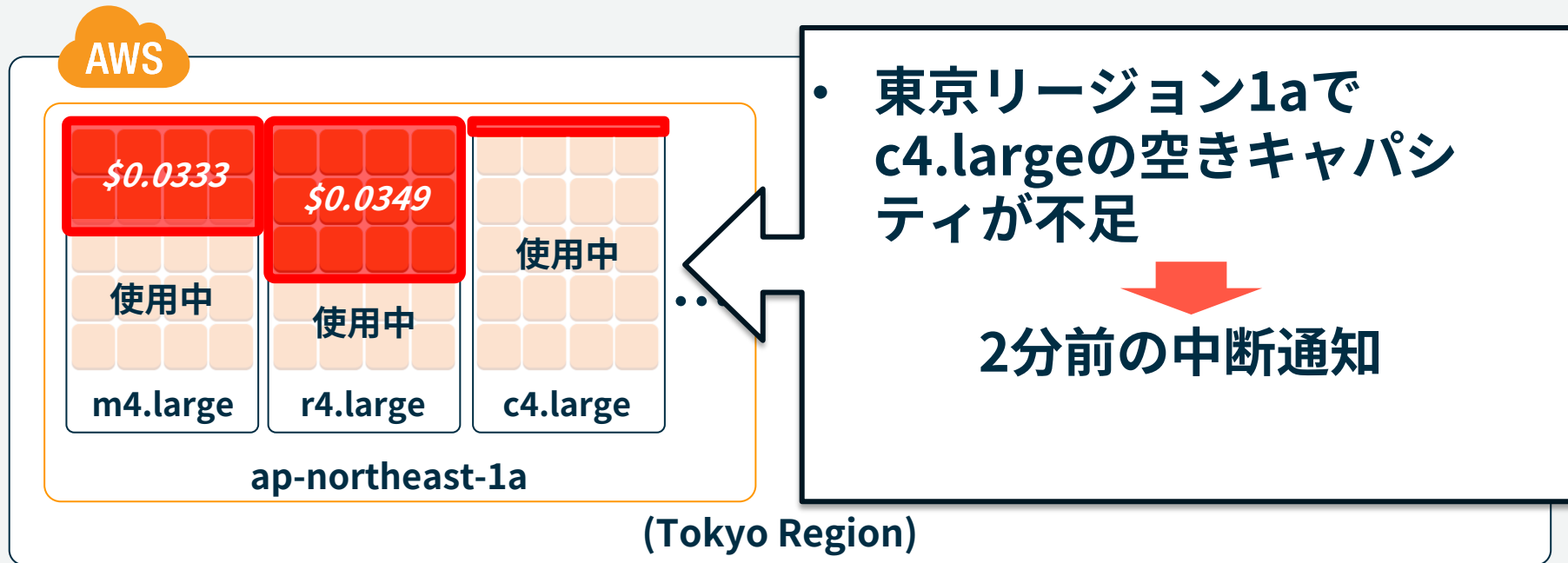
上限価格…スポットインスタンスに支払っても良いと思う最大料金。デフォルトはオンデマンドインスタンス価格 (旧「入札価格」に相当する概念)

スポットインスタンスの起動条件(2)



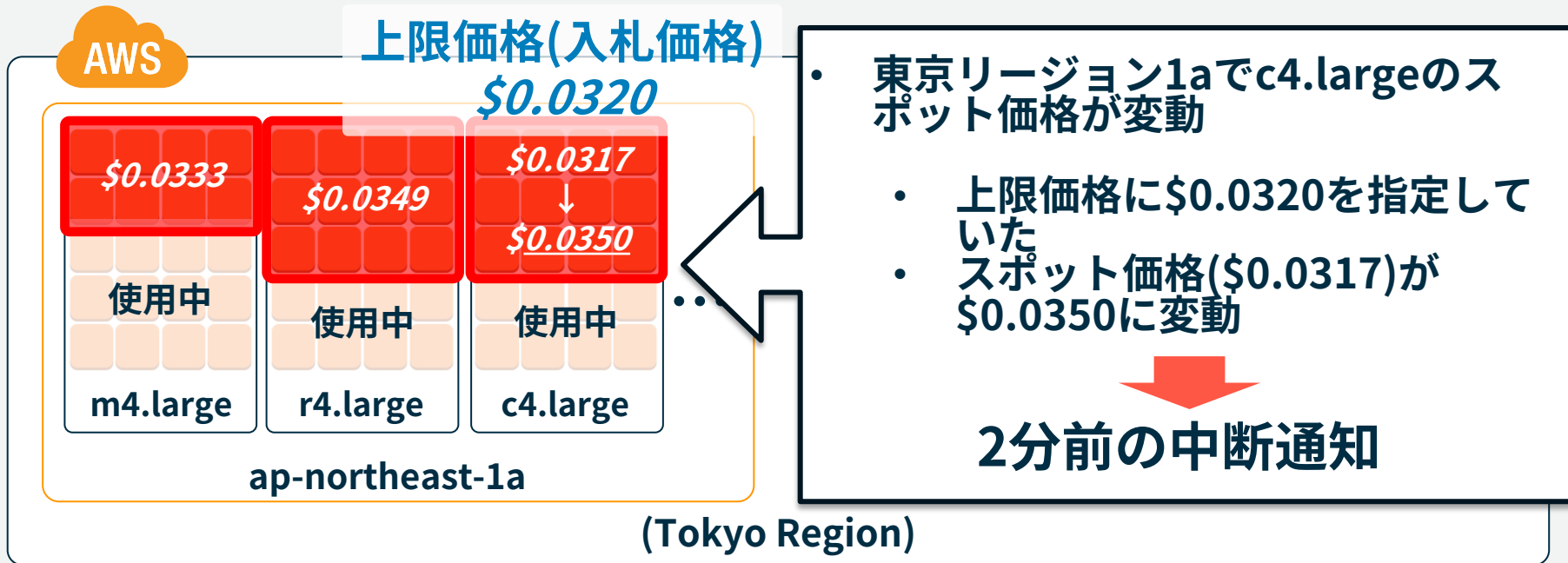
上限価格がスポット価格を上回り、スポットプールに空きがある場合に希望したスポットインスタンスを使用できる

スポットインスタンスの中断(1)



中断理由(1)…Amazon EC2のキャパシティ要件(オンデマンドインスタンスの需要増加など)

スポットインスタンスの中断(2)



中断理由(2)…スポット価格が変動した結果、設定していた上限料金を上回った

スポットインスタンスのシンプルなルール

- スポットインスタンスの価格は長期供給と需要に基づいて徐々に調整される
- スポットインスタンスは AWS によって中断されることがあり、その際には 2 分前までに通知される
 - Amazon EC2の空きキャパシティが使用できなくなったとき
 - 指定した価格(「上限価格」)をスポットインスタンス価格が上回ったとき

中断通知

中断通知

- スポットインスタンスが中断する場合、2分前に通知
- どうやって受け取るか？
 - インスタンスメタデータ
 - CloudWatch Events

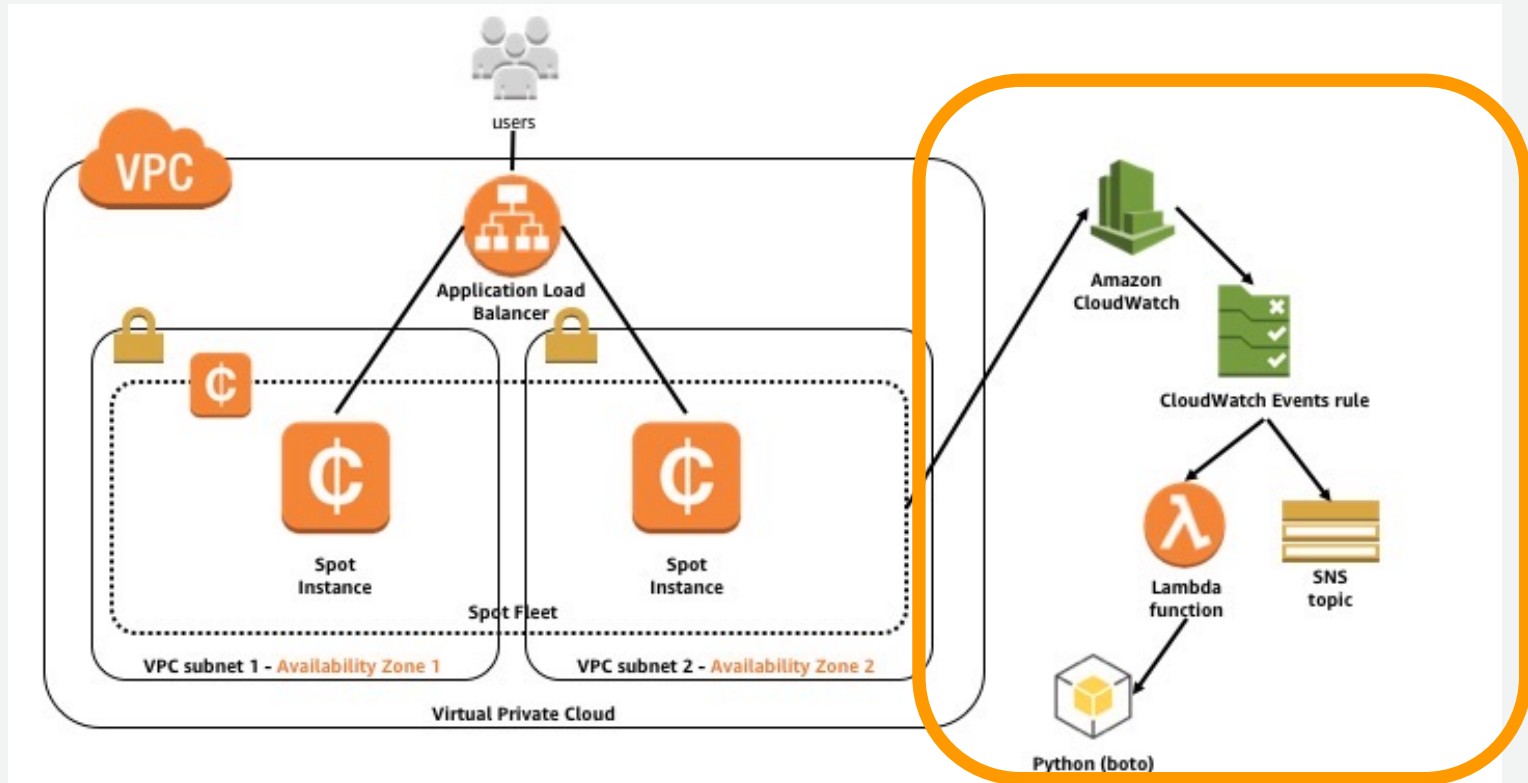
中断通知の受信 - インスタンスメタデータから

- **インスタンスメタデータサービス**
 - 実行中のインスタンスに関する情報・属性を提供
 - <http://169.254.169.254/latest/meta-data/> 以下に格納される
 - https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/UserGuide/ec2-instance-metadata.html
- **中断通知**
 - <http://169.254.169.254/latest/meta-data/spot/instance-action> に格納される
 - 普段は空
 - 5秒おきのチェックを推奨
 - https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/UserGuide/spot-interruptions.html

中断通知の例

```
{  
  "version": "0",  
  "id": "12345678-1234-1234-1234-123456789012",  
  "detail-type": "EC2 Spot Instance Interruption Warning",  
  "source": "aws.ec2",  
  "account": "123456789012",  
  "time": "yyyy-mm-ddT hh:mm:ssZ",  
  "region": "us-east-2",  
  "resources": ["arn:aws:ec2:us-east-2:123456789012:instance/i-1234567890abcdef0"],  
  "detail": {  
    "instance-id": "i-1234567890abcdef0",  
    "instance-action": "action"  
  }  
}
```

中断通知の受信 - CloudWatch Eventsから



<https://aws.amazon.com/blogs/compute/taking-advantage-of-amazon-ec2-spot-instance-interruption-notices/>

中断通知の受信 - CloudWatch Eventsから

ステップ 1: ルールの作成

AWS 環境で発生するイベントに基づいてターゲットを呼び出すためのルールを作成します。

イベントソース

イベントパターンを構築またはカスタマイズするか、スケジュールを設定してターゲットを呼び出します。

イベントパターン ⓘ スケジュール ⓘ

サービス別のイベントに一致するイベントパターンの構築

サービス名

EC2

イベントタイプ

EC2 Spot Instance Interruption Warning

▼ イベントパターンのプレビュー

```
{
  "source": [
    "aws.ec2"
  ],
  "detail-type": [
    "EC2 Spot Instance Interruption Warning"
  ]
}
```

すべてのイベント

EC2 Instance State-change Notification

EBS Volume Notification

EBS Snapshot Notification

EC2 Spot Instance Interruption Warning

AWS API Call via CloudTrail

クリップボードにコピー 編集

リクエスト種別と料金

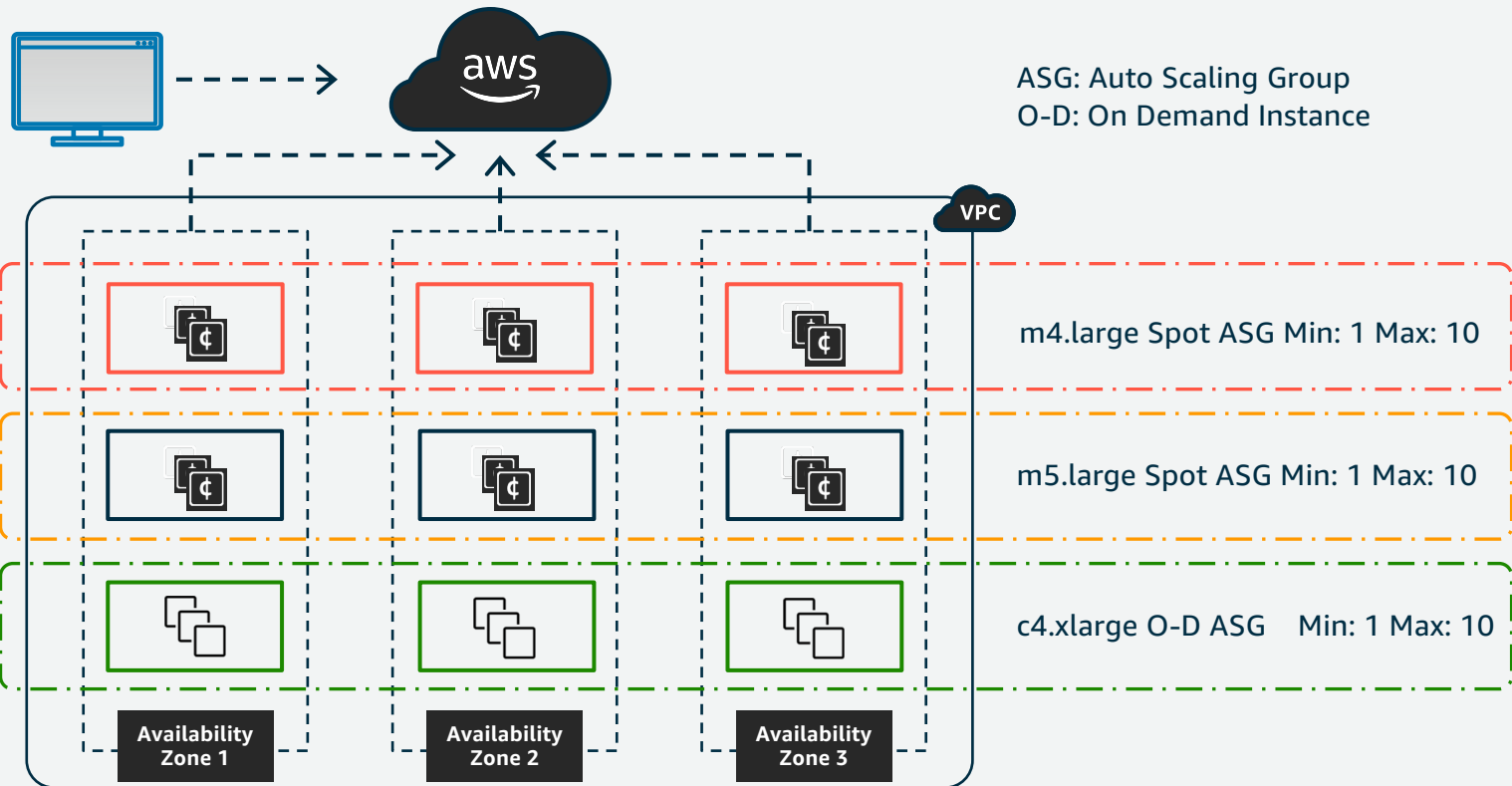
スポットリクエストの種類

	起動用AWS CLI (aws ec2 ...)	リクエスト単位	リクエストタイプ	永続性	容量変更	対応するマネジメントコンソール (EC2)
スポットインスタンス	run-instances (request-spot-instances)	インスタンス台数	インスタンス	one-time (sync) / persistent (sync)	×	「インスタンス」
スポットフリート	request-spot-fleet	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)	○	「スポットリクエスト」
EC2フリート	create-fleet	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)	○	「Auto Scalingグループ」 (New!)

ワンタイムリクエスト：one-time, request, instant

永続リクエスト：persistent, maintain

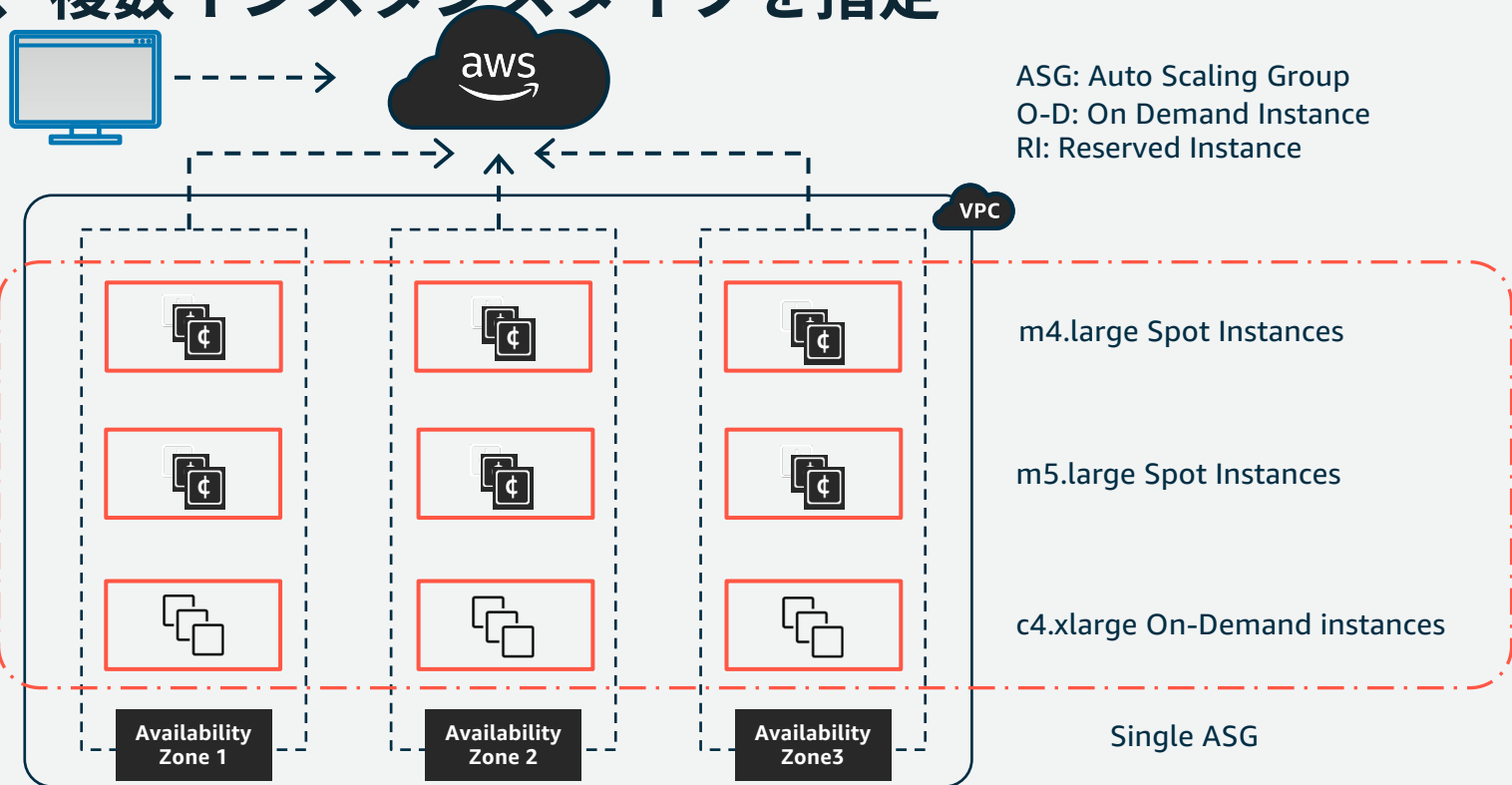
Auto ScalingとEC2フリートの統合：統合以前



以前の設計

- オンデマンド用 ASGとスポット用 ASG を別に用意
- 使用するインスタンスタイプごとに ASG を用意

Auto ScalingとEC2フリートの統合：単一ASGにO-D/RI/スポット、複数インスタンスタイプを指定



現在の設計：単一のASGに複数購入オプションや複数のインスタンスタイプを指定できる

起動画面

スポットリクエストの種類

	起動用AWS CLI (aws ec2 ...)	リクエスト単位	リクエストタイプ	永続性	容量変更	対応するマネジメントコンソール (EC2)
スポットインスタンス	run-instances (request-spot-instances)	インスタンス台数	インスタンス	one-time (sync) / persistent (sync)	×	「インスタンス」
スポットフリート	request-spot-fleet	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)	○	「スポットリクエスト」
EC2フリート	create-fleet	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)	○	「Auto Scalingグループ」 (New!)

ワンタイムリクエスト：one-time, request, instant

永続リクエスト：persistent, maintain

スポットインスタンスの起動

1. AMI の選択
2. インスタンスタイプの選択
3. インスタンスの設定
4. ストレージの追加
5. タグの追加
6. セキュリティグループの設定
7. 確認

手順 3: インスタンスの詳細の設定

要件に合わせてインスタンスを設定します。同じ AMI からの複数インスタンス作成や、より低料金を実現するためのスポットインスタンスのリクエスト、インスタンスへのアクセス管理ロール割り当てなどを行うことができます。

インスタンス数 ⓘ Auto Scaling グループに作成する ⓘ

購入のオプション ⓘ スポットインスタンスのリクエスト

現在の価格 ⓘ	アベイラビリティゾーン	現在の価格
	ap-northeast-1a	\$0.0333
	ap-northeast-1c	\$0.0333
	ap-northeast-1d	\$0.0362

最高価格 ⓘ \$ 例: 0.045 = 4.5 セント/時間 (オプション)

[永続的リクエスト] ⓘ [永続的リクエスト]

作成グループ ⓘ (省略可能)

リクエストの有効期間の開始 ⓘ 任意の時刻 [編集](#)

リクエスト有効期間の終了 ⓘ 任意の時刻 [編集](#)

スポットフリートの起動(1)

EC2 ダッシュボード

イベント

タグ

レポート

制限

インスタンス

インスタンス

起動テンプレート

スポットリクエスト

リザーブドインスタンス

専用ホスト

スケジュール済みインスタンス

スポットインスタンスのリクエスト アクション ▾ 価格設定履歴 Savings Summary

リクエストタイプ: all ▼ 状態: all ▼ キーワードによる検索

リクエスト ID リクエストタ... インスタンスタ... 状態 容量 ステータス

現在、このリージョンにスポットリクエストはありません。

EC2 スポットインスタンスを初めて使用する場合は、「開始方法」ページにアクセスしてください。

スポットインスタンスを起動するには、スポットインスタンスのリクエスト ボタンをクリックします。

スポットインスタンスのリクエスト

詳細を表示するには、上記から 1 つのスポットリクエストを選択します。

スポットフリートの起動(2)

EC2 > スポットリクエスト > スポットインスタンスのリクエスト

スポットインスタンスのリクエスト

アプリケーションまたはタスクのニーズについてお知らせください

ジョブに最も適切なコンピューティング性能を AWS が確認できるように、アプリケーションまたはタスクのニーズに最も一致するものを選択します。

Load balancing

workloads

Launch instances of the same size, in any Availability Zone. Good for running web services.

Flexible workloads

Launch instances of any size, in any Availability Zone. Good for running batch and CI/CD jobs.

Big data workloads

Launch instances of any size, in a single Availability Zone. Good for MapReduce jobs.

Defined duration

workloads

Launch instances into a Spot block for 1 to 6 hours.

1 時間

ワークロードを選ぶと、スポットフリートに使用するインスタンスタイプの種類を提案してくれる

- Load balancing: **近いサイズのインスタンスタイプから選択**
- Flexible workloads: **サイズの異なるインスタンスタイプから選択**

スポットフリーストの起動(3)

インスタンスの設定

起動テンプレートを使用して迅速にインスタンス起動パラメータを設定します (一部のパラメータは変更できません)。複数のアベイラビリティゾーンを含めることで、スポット容量へのアクセスを強化します。

起動テンプレート

なし (デフォルトパーシ | 起動テンプレートの作成

AMI ⓘ

Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type (カスタム AMI を使う

最小コンピュータユニット 仕様として インスタンスタイプとして

vCPU 2 メモリ (GiB) 7

ネットワーク ⓘ

起動するスポットインスタンスが最低限満たす仕様を指定する

- 「仕様として」：vCPUとメモリを指定
- 「インスタンスタイプとして」：起動したいインスタンスタイプの一例を指定

スポットフリートの起動(4)

必要な容量をお知らせください

起動するターゲット容量 (インスタンス数または vCPU 数) を設定します。起動テンプレートを指定した場合、ターゲット容量の一部をオンデマンドとして割り当てることができます。オンデマンドインスタンスの数は常に保持されますが、スポットインスタンスはスケールできます。

合計ターゲット容量 ⓘ

1

インスタンス

オプションのオンデマンド部分 [詳細はこちら](#)

ら ↗

0

インスタンス

起動テンプレートを指定するリクエストのみがオンデマンドの対象です

ターゲット容量を維持する

中断動作 ⓘ

終了

- スポットフリートに起動するスポットインスタンスの容量を、台数もしくはvCPUで指定する
- スポットフリートの一部をオンデマンドインスタンスで構成できる
- 「ターゲット容量を維持する」にチェックを入れると、中断などで指定容量を下回った場合にスポットフリートが自動で新しいインスタンスを起動する

スポットフリートの起動(5)

フリートリクエストの設定

推奨事項の適用

フリートリクエスト

Amazon EC2 はこれらのインスタンスタイプからターゲット容量をリクエストします。指定したインスタンスタイプが多いほど、ターゲット容量を満たす可能性が高まります。

インスタンスタイプ	vCPUs	Memory (GiB)	スポット価格	オンデマンド料金の削減
t2.large	2 vCPUs	8GiB	\$0.0278/hr	70 %
m3.large	2 vCPUs	7.5GiB	\$0.0307/hr	77 %
m4.large	2 vCPUs	8GiB	\$0.0322/hr	68 %
r4.large	2 vCPUs	15.25GiB	\$0.0337/hr	75 %
m5.large	2 vCPUs	8GiB	\$0.0338/hr	65 %

フリー트의強度: 強度が十分

フリートには、ターゲット容量リクエストを満たす十分なインスタンスプールが含まれています。
5 instance types x 4 Availability Zones = 20 instance pools

- 指定した値をもとにEC2スポットサービスがおすすめのインスタンスタイプを選定した結果が表示される
- この結果をカスタマイズしたい場合は「推奨事項の適用」チェックを外す

スポットフリートの起動(6)

フリートリクエストの概要

合計ターゲット容量 インスタンス (maintain capacity)	インスタンスの設定 mylaunchtemplate, v.1 2 vCPU, 7 GiB (min) 4 Availability Zones	フリートの強度 強度が十分 20 instance pools	推定料金 ~\$0/hr ターゲット容量で 71% 削減額 オンデマンドと比較
--------------------------------------------	--------------------------------------------------------------------------------------------	----------------------------------------------	------------------------------------------------------------

[JSON 設定](#) [キャンセル](#) [作成](#)

- フリートの強度が「強度が十分」であることを確認する
- フリート全体の削減額の見通しが表示される
- 問題なければ「作成」

EC2 Auto Scaling Groupの作成(1)

起動テンプレートの提供が開始されました。
EC2 Auto Scaling コンソールで、EC2 起動テンプレートのフルサポートが開始されました。新しい Auto Scaling グループには起動テンプレートを使用することをお勧めします。起動テンプレートにより、Amazon EC2 の最新機能を活用することができます。Auto Scaling グループを作成して開始するか、[詳細はこちら](#)を参照してください。

Auto Scaling へようこそ

Auto Scaling を使用すると、Amazon EC2 キャパシティの自動的な管理、適切な数のアプリケーションインスタンスの維持、インスタンスの正常なグループの運用、必要に応じたスケールリングを行うことができます。
[詳細はこちら](#)

Auto Scaling グループの作成

注意: 別のリージョンで Auto Scaling グループを作成するには、ナビゲーションバーからリージョンを選択します。

追加情報

- 入門ガイド
- ドキュメント
- すべての EC2 リソース
- フォーラム
- 料金
- お問い合わせ

- EC2フリートと統合した新しいEC2 Auto Scalingの機能を使うには「起動テンプレート」を用いる必要がある

EC2 Auto Scaling Groupの作成(2)

Auto Scaling グループの作成

キャンセルして終了

このウィザードを終了して Auto Scaling グループを作成します。最初に、起動設定または起動テンプレートを選択して、インスタンスの起動に Auto Scaling グループが使用するパラメータを指定します。

起動設定

必要な Amazon EC2 の機能をサポートしている場合は、引き続き起動設定を使用できます。詳細はこちら [🔗](#)

起動テンプレート **新規**

起動テンプレートにより、1つの種類のインスタンスを起動するか、複数のインスタンスタイプと購入オプションの組み合わせを起動するかのオプションを利用できます。起動テンプレートには Amazon EC2 の最新機能が含まれていて、更新とバージョンアップができます。詳細はこちら [🔗](#)
[新しい起動テンプレートの作成](#) [🔗](#)



🔍 起動テンプレートのフィルタリング... X

◀ < 1 から 1 / 1 の テンプレートの 起動 > ▶

名前	起動テンプレート ID	デフォルトバージョン	最新バージョン	作成時刻	作成者
<input type="checkbox"/>	mylaunchtemplate1-0ce01f06f248ac920	1	2	Wed Feb 21 15:25:30 GMT+900 2018	arn:aws:sts::234992788963:assu

EC2 Auto Scaling Groupの作成(3)

フリートの構築

起動テンプレートに従う
起動テンプレートにより、インスタンスタイプと購入オプション(オンデマンドまたはスポット)が決まります。

購入オプションとインスタンスを組み合わせる
オンデマンドインスタンスとスポットインスタンスの組み合わせにより複数のインスタンスタイプを選択します。スポットインスタンスは、利用できる最も安い料金で自動的に起動されます。

インスタンスタイプ

許容できるインスタンスタイプをフリートに追加します。順序を変更し、オンデマンドインスタンスの起動の優先度を設定します。この順序によるスポットインスタンスへの影響はあります。

m4.large (2vCPU, 8GiB)	
c4.large (2vCPU, 3.75GiB)	
インスタンスタイプの追加	

- 「購入オプションとインスタンスを組み合わせる」を選択し、要件に合うインスタンスタイプを複数選択する
- 起動テンプレートに指定しておくことも可能

EC2 Auto Scaling Groupの作成(4)

インスタンスの分散 ⓘ	<input type="checkbox"/> 次のデフォルト設定を使用し、すぐに開始します。
オンデマンドの割り当て戦略 ⓘ	優先順位付け
最大スポット料金 ⓘ	<input checked="" type="radio"/> デフォルトを使用 (推奨) デフォルトでは現在のスポット料金が使用されますが、オンデマンド価格に上限が設定されます。 <input type="radio"/> 上限価格を設定 (1 インスタンス/時間あたり)
スポットの配分戦略 ⓘ	スポットインスタンスを <input type="text" value="2"/> アベイラビリティゾーンごとに最も価格の安いインスタンスタイプ間で多様化する
オプションのオンデマンドベース ⓘ	最初のインスタンスを <input type="text" value="0"/> オンデマンドとして指定します
ベースを超えるオンデマンド割合 ⓘ	<input type="text" value="70"/> % オンデマンドおよび 30% スポット
グループサイズ ⓘ	開始時 <input type="text" value="1"/> インスタンス

• 台数の考え方

- 1. 全体数は「グループサイズ」
- 2. そこから「オプションのオンデマンドベース」の台数をオンデマンドで起動
- 3. 残りを「ベースを超えるオンデマンド割合」にしたがって分配

• 台数の考え方の例

- 「グループサイズ」：12
- 「オプションのオンデマンドベース」：2
- 「ベースを超えるオンデマンド割合」：30:70
- 結果
 - オンデマンド2台+3台
 - スポット7台

<参考> EC2 Auto ScalingとSpot Fleetの比較 (1)

機能	EC2 Auto Scaling	Spot Fleet
希望容量の維持	○	○
自動スケーリング	○	○
複数AZの指定	○	○
複数インスタンスタイプの指定	○	○
オンデマンドインスタンスの混在	○ [1] 1. 全体数はTotal target capacity 2. そこから"Optional On-demand portion"の台数をオンデマンドで起動 3. 残りをスポットで起動	○ [2] 1. 全体数はDesired Capacity 2. そこから"On-Demand base"の台数をオンデマンドで起動 3. 残りをOD:Spot Ratioにしたがって分配
ELBからのトラフィック受信	○	○
Launch Template \$latest, \$defaultでの参照	○	×

[1] https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/UserGuide/spot-fleet.html#on-demand-in-spot

[2] https://docs.aws.amazon.com/ja_jp/autoscaling/ec2/userguide/AutoScalingGroup.html#asg-purchase-options

<参考> EC2 Auto ScalingとSpot Fleetの比較 (2) - 独自機能

EC2 Auto Scaling

- 終了ポリシー (Termination Policy)
- インスタンスの保護 (Termination Protection)
- ライフサイクルフック
- 一時的なインスタンスの削除 (Entering Standby)
- スケーリングプロセスの中断 (Suspending Processes)
- アクティビティの再分散 (Rebalancing activities)

Spot Fleet

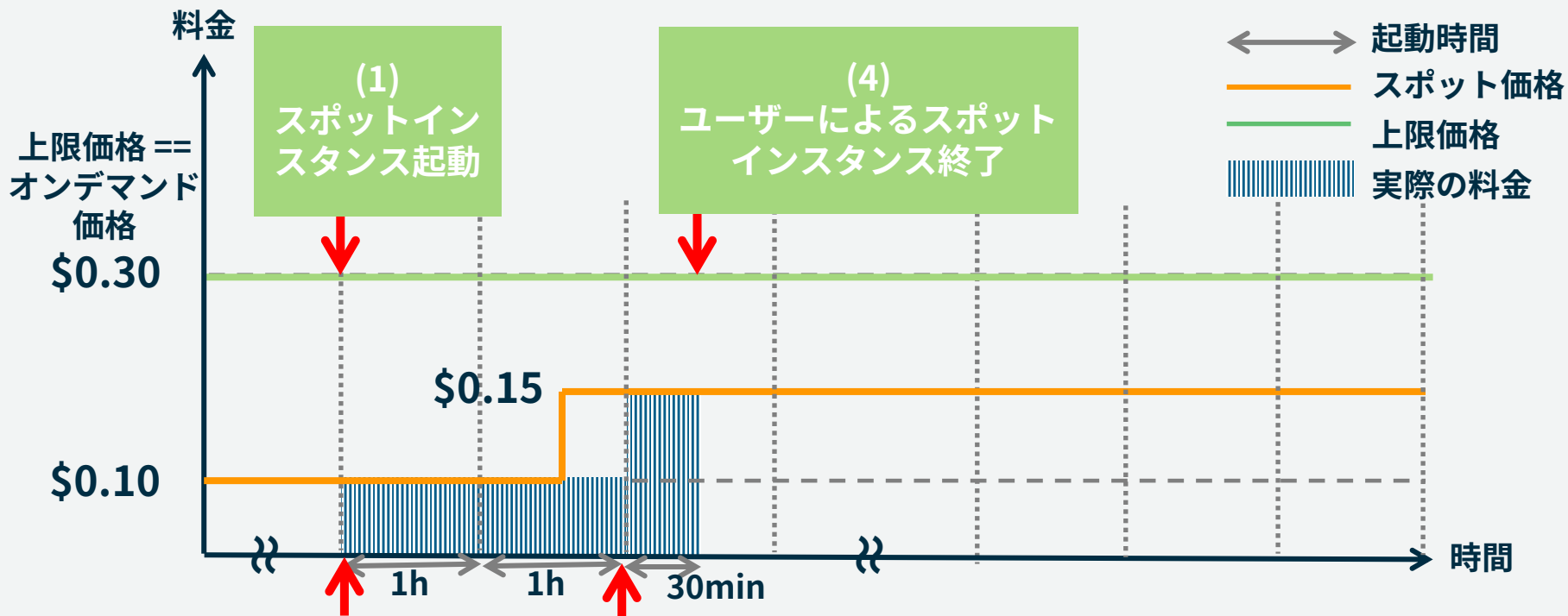
- スポットコンソール関連
 - 複数インスタンスタイプの提示
 - 推定料金(リクエスト時)および削減額(表示時)の提示

スポットインスタンス料金のルール

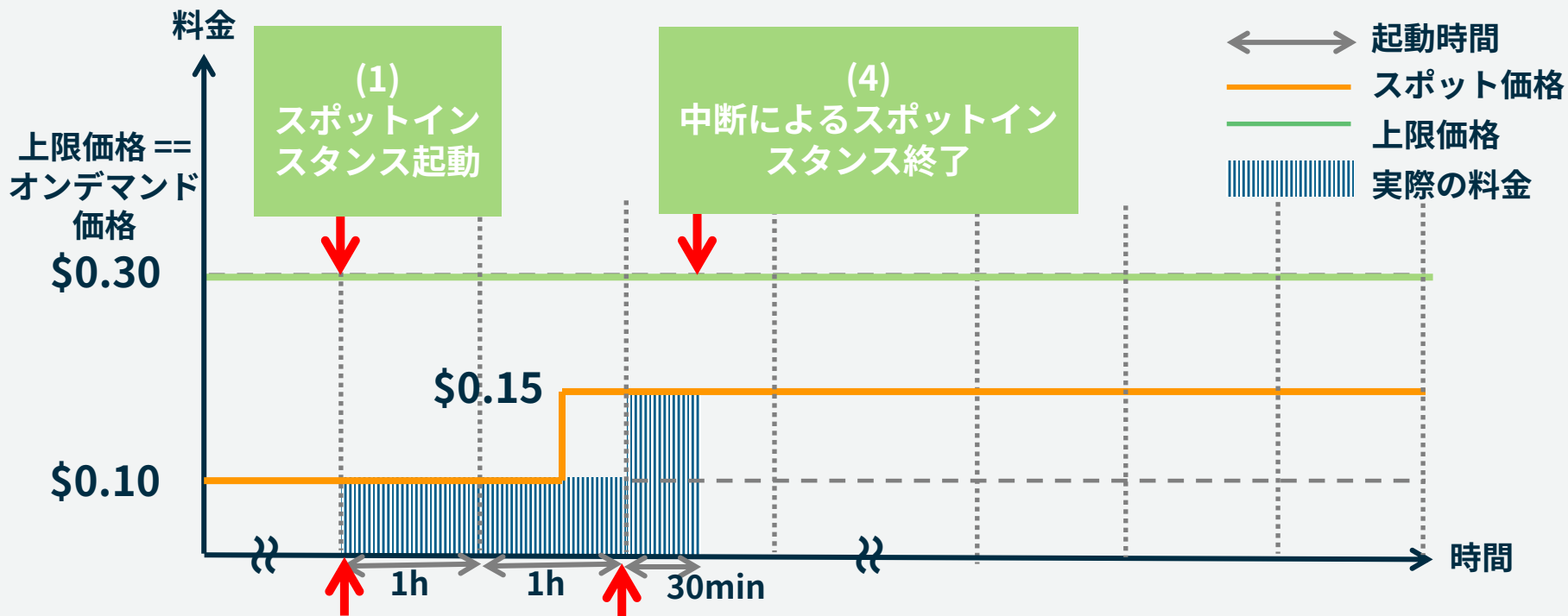
- 起動したタイミングのスポットインスタンス料金を選択される。稼働中はその料金が適用される
- 起動後、1時間単位(「インスタンス時間」)で料金変動の有無をチェック。変動していた場合はそのインスタンス時間から新しい料金が適用される

	中断 – 起動後1時間以内	中断 – 起動後1時間後	ユーザーによる停止・終了
秒単位支払い	料金発生なし	秒単位支払い	秒単位支払い
時間単位支払い	料金発生なし	時間単位支払い、 最後の1時間分は 料金発生なし	時間単位支払い

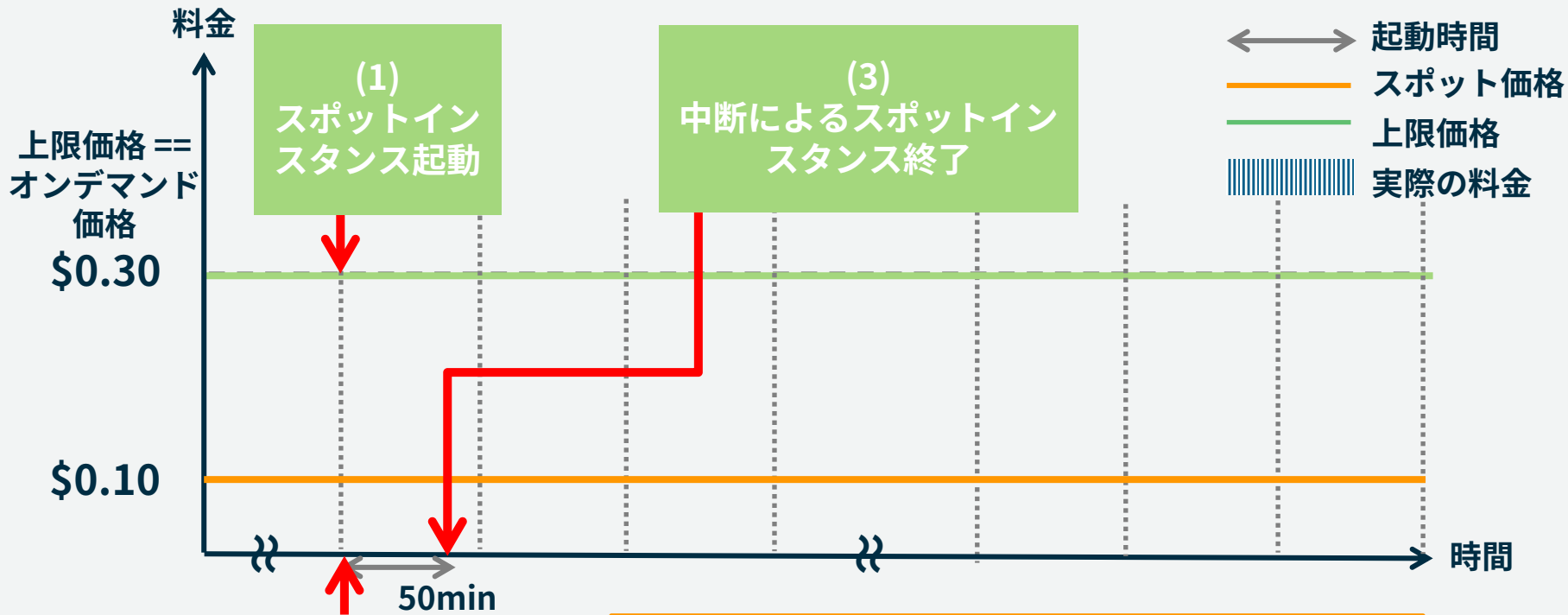
スポットインスタンス料金のルール (秒単位支払い)



スポットインスタンス料金のルール (秒単位支払い)

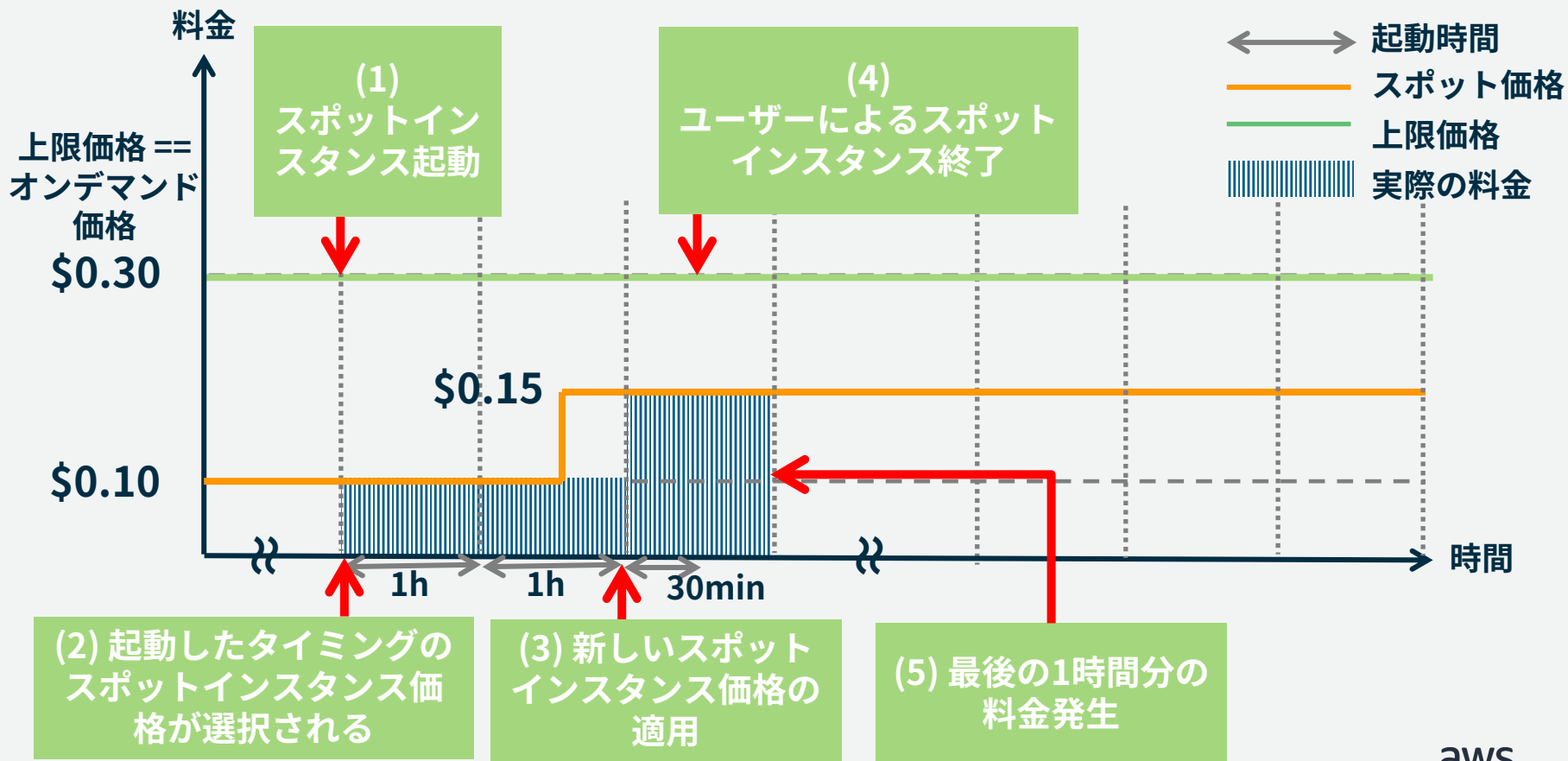


スポットインスタンス料金のルール (秒単位支払い)

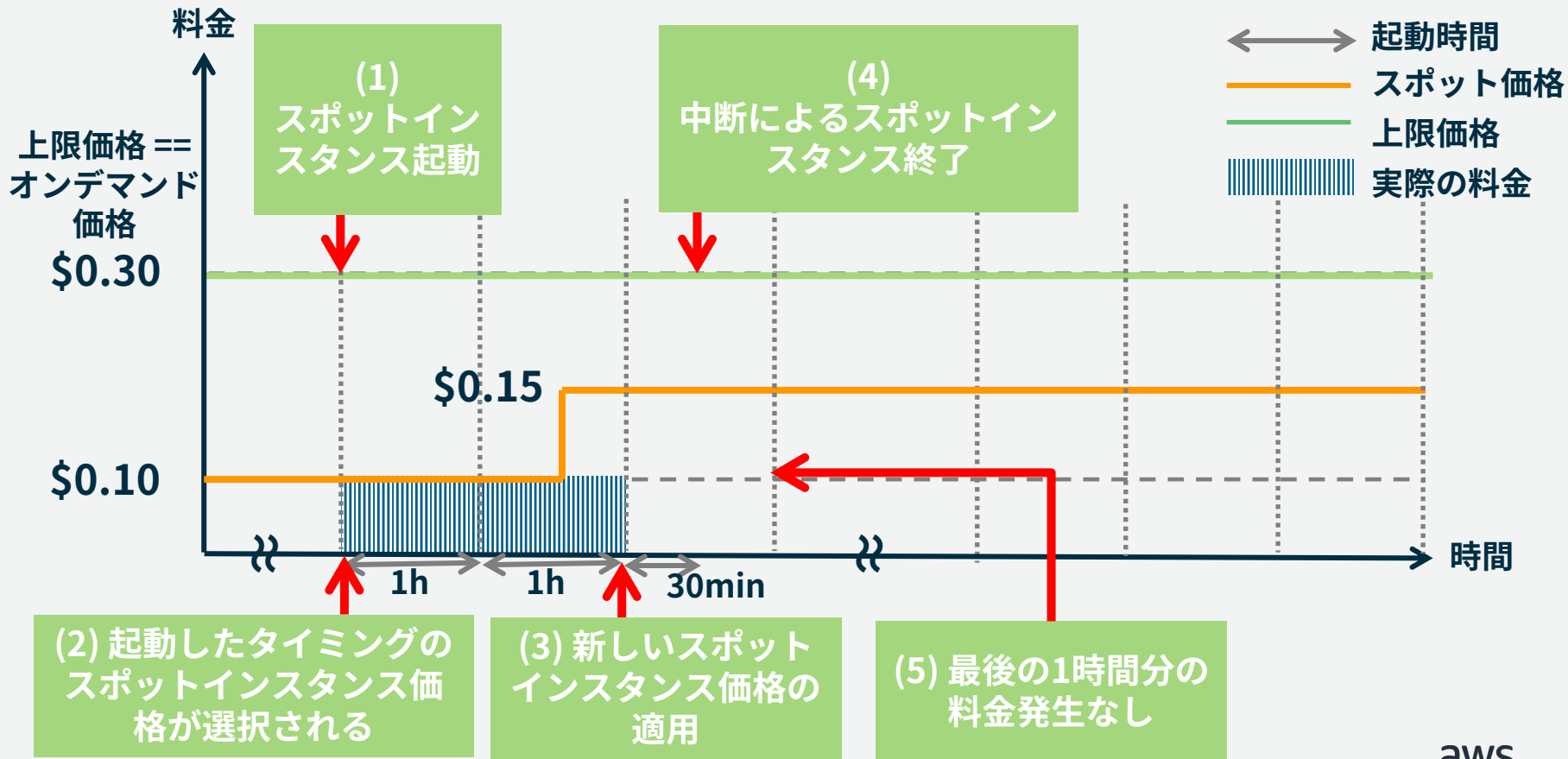


起動後1時間以内の中断には
料金発生なし

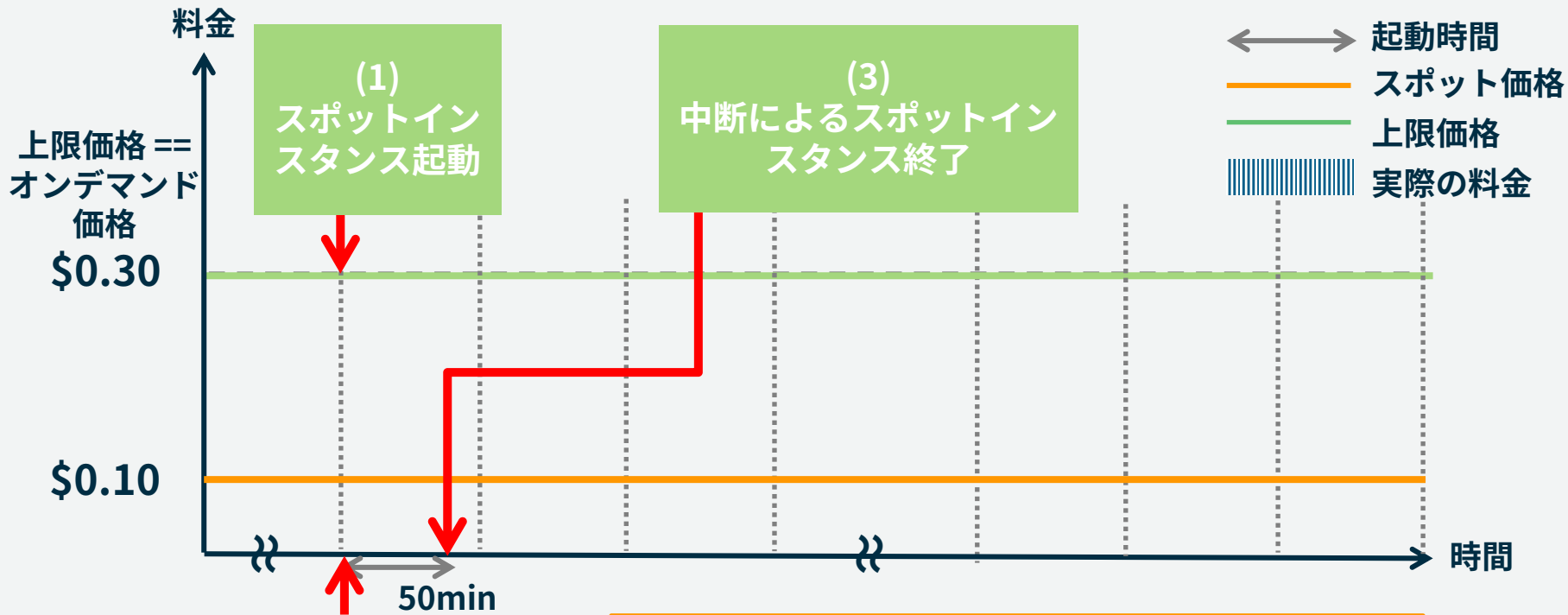
スポットインスタンス料金のルール (時間単位支払い)



スポットインスタンス料金のルール (時間単位支払い)



スポットインスタンス料金のルール (時間単位支払い)



起動後1時間以内の中断には
料金発生なし

最新のスポットインスタンス価格の確認方法

- 料金 - Amazon EC2 スポットインスタンス - Amazon EC2 | AWS
<https://aws.amazon.com/jp/ec2/spot/pricing/>

aws 日本担当チームへお問い合わせ サポート 日本語 ▼ アカウント ▼ 今すぐ AWS アカウント作成 »

製品 ソリューション 料金 ドキュメント 学習 パートナー AWS Marketplace その他 Q

製品 & サービス

- スポットインスタンス >

関連リンク

- Amazon EC2 スポットインスタンス
- Amazon EC2 リザーブドインスタンス
- Amazon EC2 Dedicated Hosts
- Amazon EC2 ハードウェア専有インスタンス
- Amazon EC2 Elastic GPU
- Windows インスタンス
- VMware Cloud on AWS
- Systems Manager
- Server Migration Services
- Application Discovery

スポットインスタンス価格

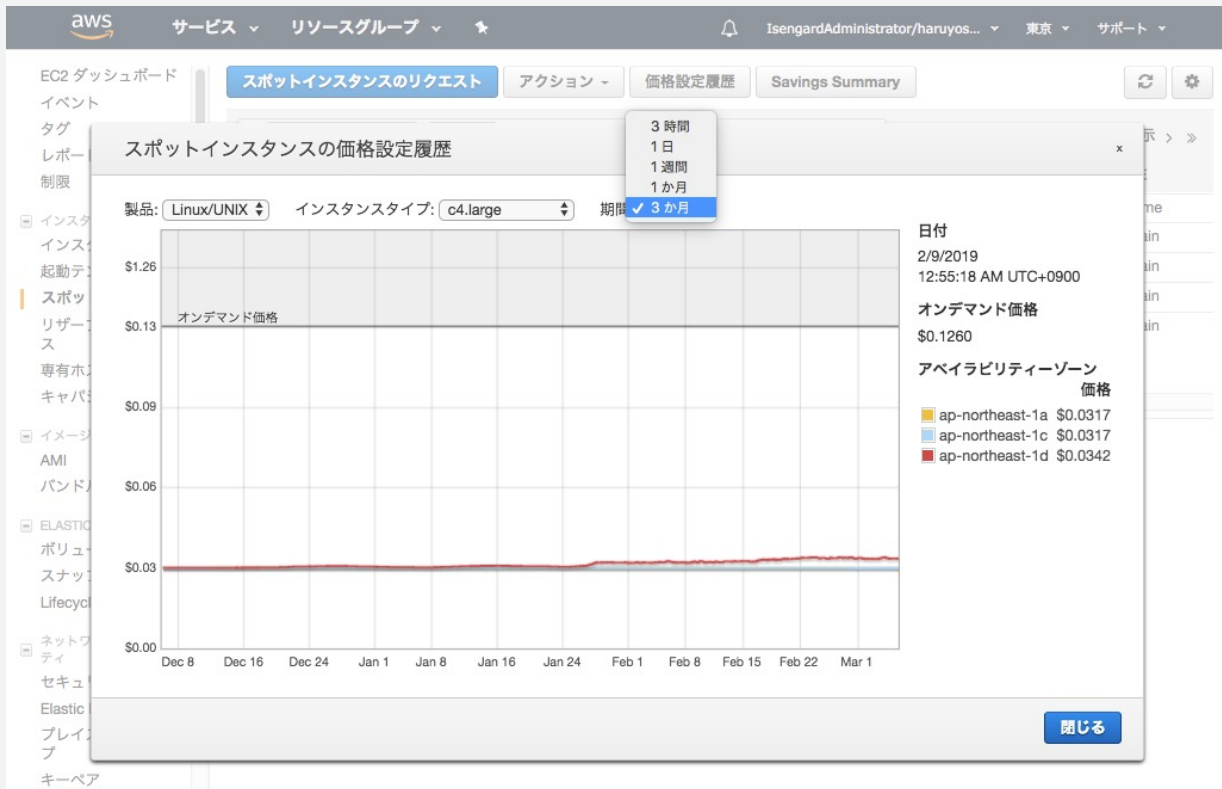
スポットインスタンス Linux 用指定継続時間 Windows 用指定継続時間

リージョン: アジアパシフィック (東京) ▼

	Linux/UNIX 料金	Windows 料金
一般的な目的 - 現行世代		
a1.medium	該当なし*	該当なし*
a1.large	該当なし*	該当なし*
a1.xlarge	該当なし*	該当なし*
a1.2xlarge	該当なし*	該当なし*
a1.4xlarge	該当なし*	該当なし*
t2.micro	\$0.0046 /1 時間	\$0.0092 /1 時間

スポットインスタンス価格履歴の確認方法(1)

EC2スポットインスタンスマネジメントコンソール「価格設定履歴」



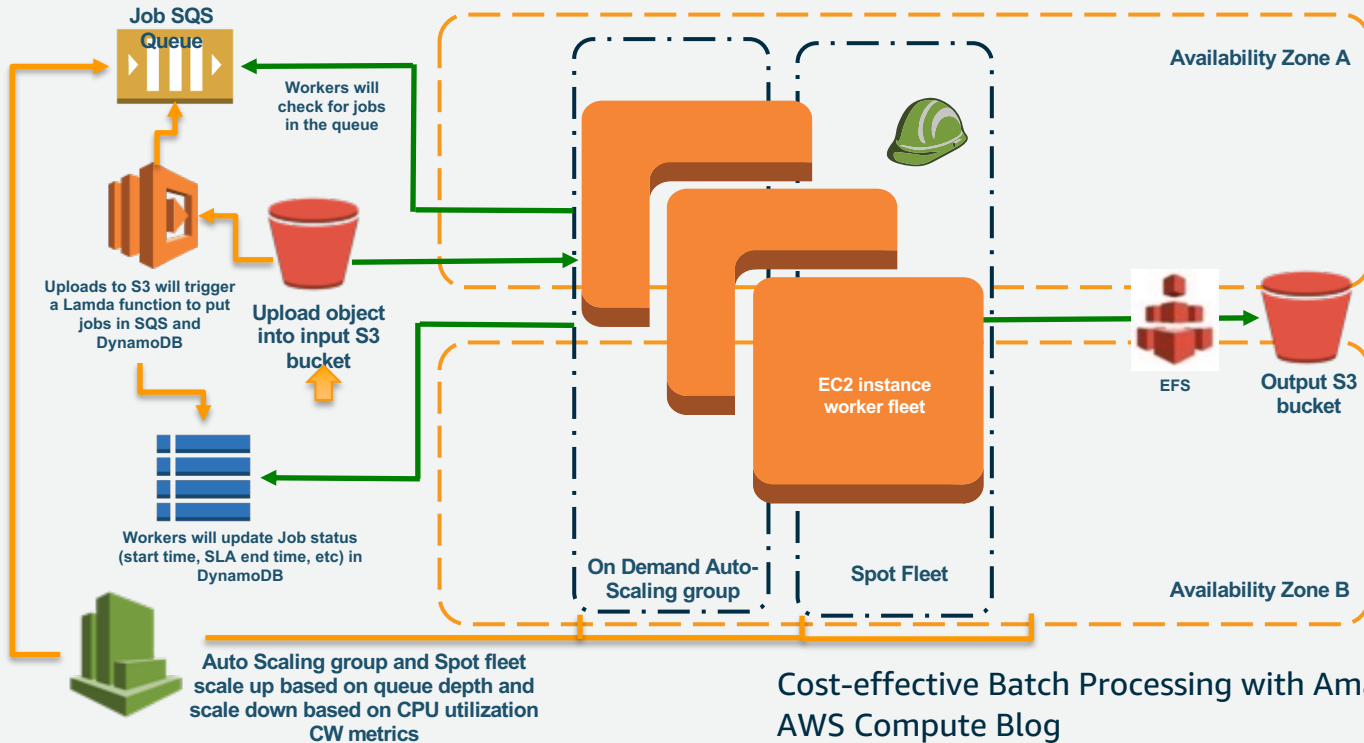
スポットインスタンス価格履歴の確認方法(2)

AWS CLI: describe-spot-price-history

```
$ aws ec2 describe-spot-price-history --instance-types m3.medium
{
  "SpotPriceHistory": [
    {
      "Timestamp": "2019-03-04T21:36:38.000Z",
      "AvailabilityZone": "ap-northeast-1a",
      "SpotPrice": "0.109600",
      "ProductDescription": "SUSE Linux",
      "InstanceType": "m3.medium"
    },
    {
      "Timestamp": "2019-03-04T21:36:38.000Z",
      "AvailabilityZone": "ap-northeast-1c",
      "SpotPrice": "0.109600",
      "ProductDescription": "SUSE Linux",
      "InstanceType": "m3.medium"
    }
  ],
  :
}
```

<参考> サンプルアーキテク チャ：ワークロード事例

スポットインスタンスによるバッチ処理(旧来の方法)



Cost-effective Batch Processing with Amazon EC2 Spot |
AWS Compute Blog
<https://aws.amazon.com/blogs/compute/cost-effective-batch-processing-with-amazon-ec2-spot/>



AWS Batchによるバッチ処理(推奨)



submit-job

list-jobs
describe-jobs
cancel-jobs
terminate-jobs



AWS Batch
JobQueue

JobDefinitio
n

事前に定義

Compute Environment

jobディスパッチ



Autoscalingする
ECS環境

スポットインスタンスを
選択可能



事前に作成・保存

Docker Image



ECR/Docker
Registry

データ読み書き



S3/EFS/NFS

Amazon Newservice

Dashboard

Jobs status

Submit job

Queue name (priority)	Submitted	Pending	Runnable	Starting	Running	Failed	Succeeded
test-queue (500)	1	1	1	1	1	1	1
production-queue (1000)	1231	942	12	104020	57	17	8742

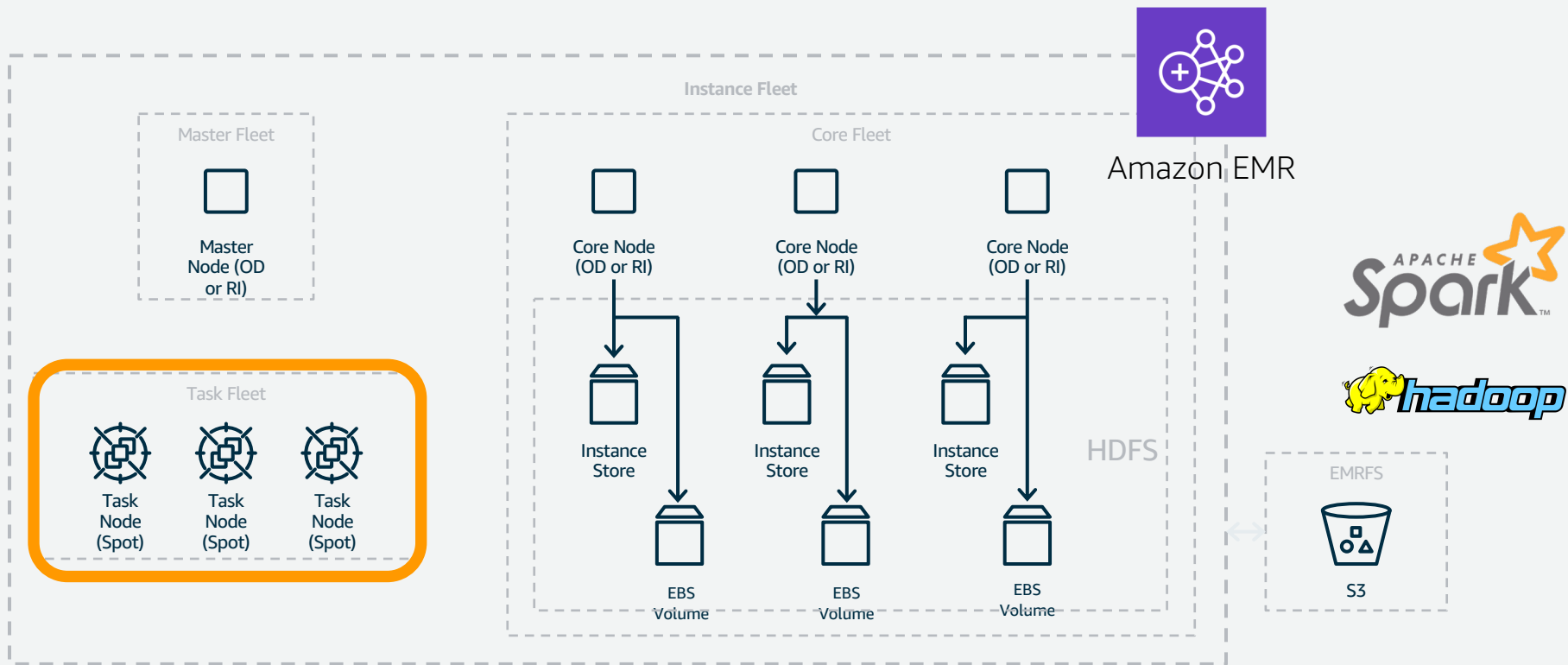
Job queues

Name	Priority	Pending job count	Running job count	Registered instance count	Total vCPUs
test-queue	500	1	1	4	4
production-queue	1000	942	104202	20	96

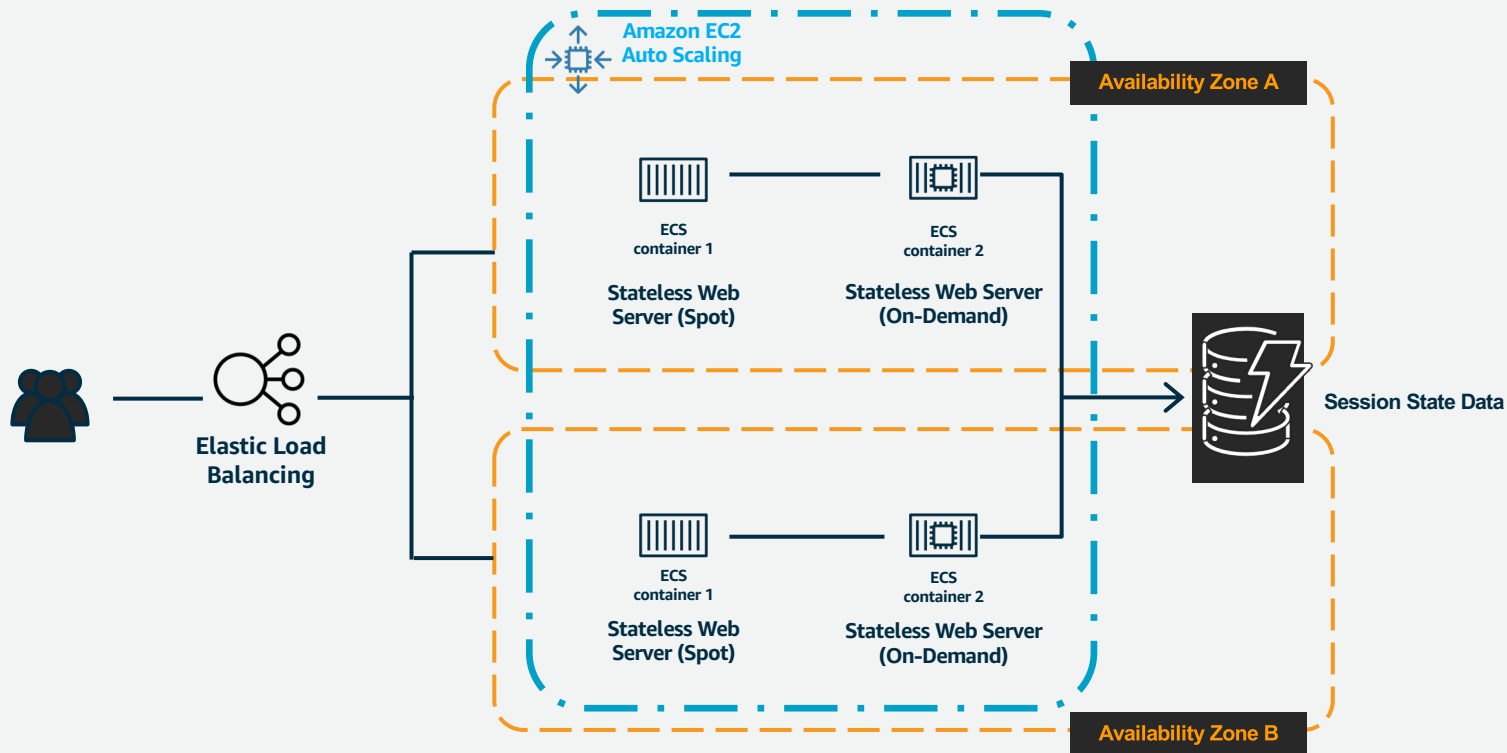
Compute environments

Name	Type	Running job count	Desired vCPUs	Registered instance count
spot-env	managed	1002301	1600	100
production	unmanaged	12	1000	100

ビッグデータ領域のリファレンスアーキテクチャ



コンテナを活用したステートレスウェブアプリケーション



<https://github.com/aws-labs/ec2-spot-labs/tree/master/workshops/ec2-spot-fleet-web-app>

<参考> ECSクラスタでのスポットインスタンスの活用時の 考慮点：スケールアウト編

- スケールアウト時の考慮：ECSクラスタに追加したい
 - ECSコンテナエージェント設定ファイルにクラスタ名を書いておく (/etc/ecs/ecs.config)
 - ECS対応AMIをベースにするのがオススメ

https://docs.aws.amazon.com/ja_jp/AmazonECS/latest/developerguide/ecs-agent-install.html

https://docs.aws.amazon.com/ja_jp/AmazonECS/latest/developerguide/example-user-data-scripts.html

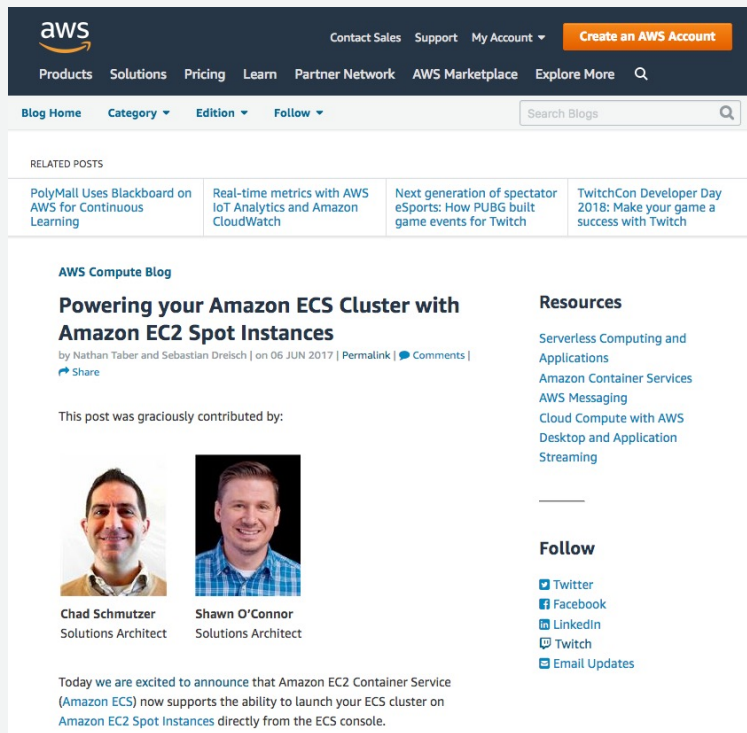
<参考> ECSクラスタでのスポットインスタンスの活用時の 考慮点：スケールイン(中断)編

- 中断時の考慮：サービス影響の極少化
 - コンテナインスタンスのステータスをDrainingにする
 - (オプション)SNSトピックに通知
- いつトリガーしてくれるのか？
→中断通知を受け取る必要がある

https://docs.aws.amazon.com/ja_jp/AmazonECS/latest/developerguide/container-instance-draining.html

<参考> ECSクラスタでのスポットインスタンスの活用時の考慮点：リファレンスアーキテクチャ

- CloudFormationテンプレートの提供
 - コンテナエージェントの設定：コンテナへの自動参加
 - 中断通知のモニタリング
 - 中断通知を受けてのDraining設定
 - SNSトピックからEメール通知



The screenshot shows the AWS Compute Blog page for the article "Powering your Amazon ECS Cluster with Amazon EC2 Spot Instances". The article is by Nathan Taber and Sebastian Dreisch, dated 06 JUN 2017. It features two authors: Chad Schmutzer and Shawn O'Connor, both Solutions Architects. The article text states: "Today we are excited to announce that Amazon EC2 Container Service (Amazon ECS) now supports the ability to launch your ECS cluster on Amazon EC2 Spot Instances directly from the ECS console." The page also includes a "Resources" section with links to "Serverless Computing and Applications", "Amazon Container Services", "AWS Messaging", "Cloud Compute with AWS Desktop and Application Streaming", and a "Follow" section with social media links for Twitter, Facebook, LinkedIn, Twitch, and Email Updates.

<https://aws.amazon.com/blogs/compute/powering-your-amazon-ecs-cluster-with-amazon-ec2-spot-instances/>

スポットインスタンスを 使いこなすために

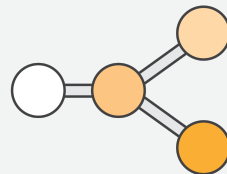
スポットインスタンスを使いこなす4+1原則



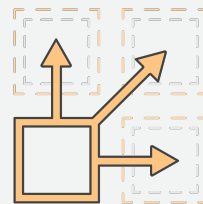
ステートレス：
状態を持たせない



**再開可能なワーク
ロード：**
安全な再開



疎結合：
周辺影響の極小化

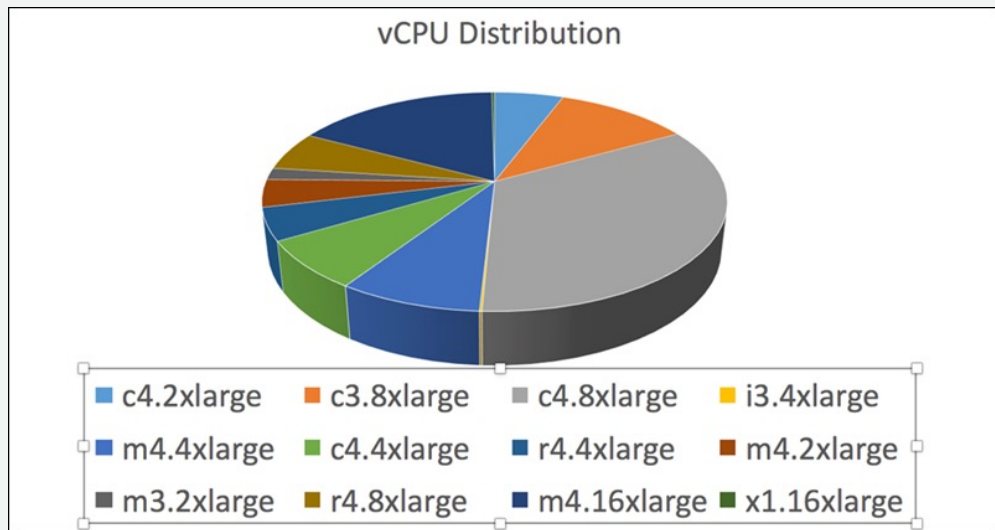


分散：
複数アベイラビリ
ティゾーンと複数
インスタンスタイ
プの活用

「上限価格」にはオンデマンドインスタンス価格を

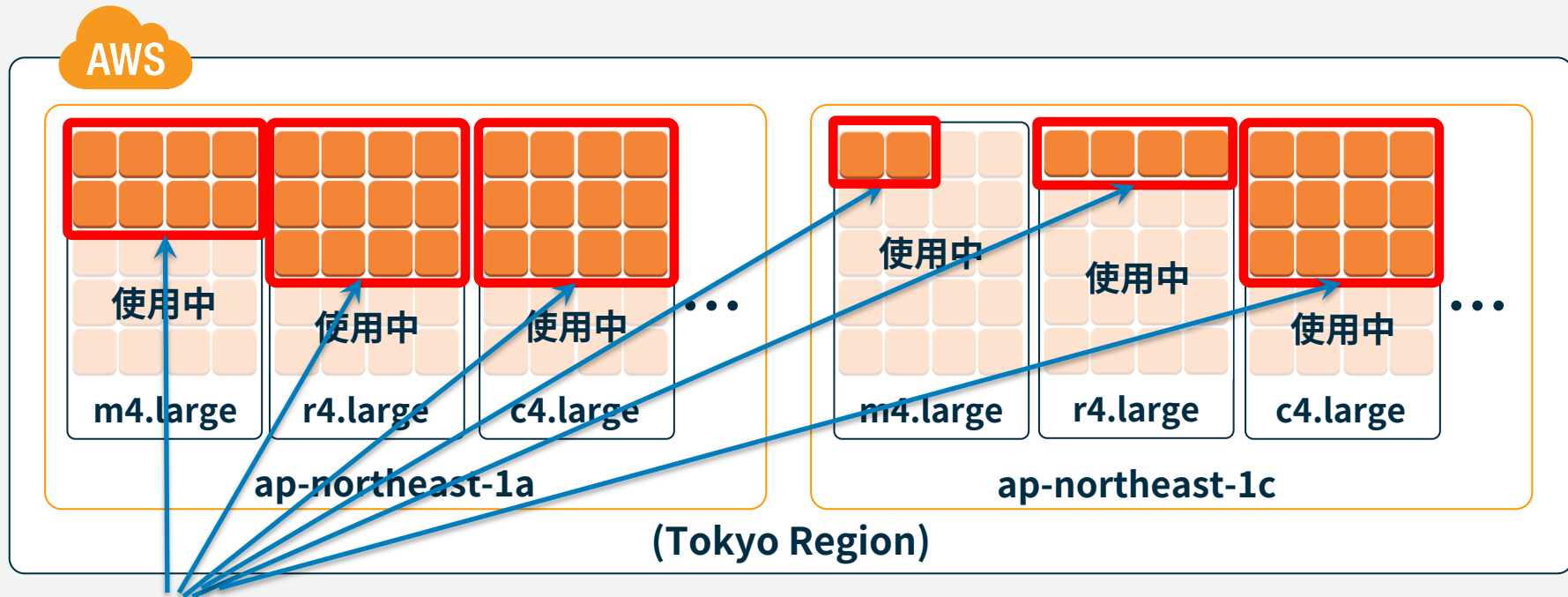
スポットインスタンスを使いこなす4+1原則 - 分散

- トピック推定の研究(自然言語処理)
 - 文書集合を元にトピックを推定できるように学習
 - 学習プロセスをAWSで並列化して実行
- スポットインスタンスを活用
 - 2017年8月26日、同時稼働vCPU数が110万コアを達成(右図)
 - スーパーコンピューターも含めた中で世界最大規模のクラスタ
 - 米国東部(バージニア北部)リージョンのみ使用



<https://aws.amazon.com/blogs/aws/natural-language-processing-at-clemson-university-1-1-million-vcpus-ec2-spot-instances/>

スポットインスタンスを使いこなす4+1原則 - 分散



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

さらにスポットインスタンスを知るために

「これからはじめる AWS Auto Scaling 2019年版」 -外道父の匠

外道父の匠

全ては息子のために



目次 Linux MySQL BigData Cloud Hardware Infrastructure 育成

← RDSのGeneralLogをバケットキャプチャに切り替える

これからはじめる AWS Auto Scaling 2019年版

124 users

81 ツイート 検索 いいね! 21 シェア

昨年は内部的なことを多くやっていたり、10年ぶりに格ゲー復帰したりで、なんとなく無沙汰になります。が、[元同僚エンジニアに名指しされたり](#)、[パルが年末にアドベで連続更新](#)してるのを見て、俺も頑張らなきゃなと思い直した次第であります。



外道父
ドリコムで主にインフラエンジニアをやっている何でも屋。

Twitter, Facebook

著書

たのしい
インフラの
歩き方

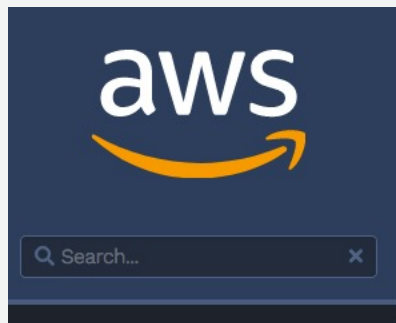
青藤雄介



「外道父 2019」
で検索!

これからはじめる AWS Auto Scaling 2019年版 | 外道父の匠 <http://blog.father.gedow.net/2019/02/05/aws-auto-scaling-2019/>

EC2 Spot Instances Workshops



- Launching EC2 Spot Instances

Overview

Launching EC2 Spot instances via EC2 Auto Scaling Group

Launching an EC2 Spot Instance via the RunInstances API ✓

Launching EC2 Spot Instances via Spot Fleet

Launching EC2 Spot Instances via an EC2 Fleet ✓

Amazon EC2 Spot Workshops > Launching EC2 Spot Insta...

LAUNCHING EC2 SPOT INSTANCES

- Overview
- Launching EC2 Spot instances via EC2 Auto Scaling Group
- Launching an EC2 Spot Instance via the RunInstances API
- Launching EC2 Spot Instances via Spot Fleet
- Launching EC2 Spot Instances via an EC2 Fleet
- Finding Running Spot Instances
- Clean Up
- Final Thoughts

<https://ec2spotworkshops.com/>

今日わかるようになること(再掲)

- スポットインスタンスは最大()割引。()が発生する
- アンチパターン
 - ミツ()ョンクリティカル
 - ステ()トフル
- 中断通知はインスタンスメタデータ、もしくはCloudWatch ()で取得
- 使いこなすために
 - ステ()トレス：状態を持たせない
 - ()開可能
 - ()結合
 - 分散：複()のインスタンスタイプを活用
 - (入札でなく)「上()価格」にはオンデマンド価格を指定

今日わかるようになること(回答例)

- スポットインスタンスは最大9割引。中断が発生する
- アンチパターン
 - ミッションクリティカル
 - ステートフル
- 中断通知はインスタンスメタデータ、もしくはCloudWatch Eventsで取得
- 使いこなすために
 - ステートレス : 状態を持たせない
 - 再開可能
 - 疎結合
 - 分散 : 複数のインスタンスの活用
 - (入札でなく)「上限価格」にはオンデマンド価格を指定

Q&A

お答えできなかったご質問については

AWS Japan Blog 「<https://aws.amazon.com/jp/blogs/news/>」にて

資料公開と併せて、後日掲載します。

3月のBlack Belt Online Seminar 配信予定

<https://amzn.to/JPWebinar>

3/5 (火) 12:00-13:00 Amazon EC2

3/6 (水) 18:00-19:00 Amazon EC2スポットインスタンス

3/12 (火) 12:00-13:00 AWS Well-Architected Framework によるコスト最適化

3/13 (水) 18:00-19:00 Amazon VPC

3/19 (火) 12:00-13:00 Amazon FSx for Windows File Server/Lustre

3/20 (水) 18:00-19:00 Amazon EBS

3/26 (火) 12:00-13:00 Amazon CloudWatch



AWS の日本語資料の場所「AWS 資料」で検索

AWS クラウドサービス活用資料集トップ

アマゾン ウェブ サービス (AWS) は安全なクラウドサービスプラットフォームで、ビジネスのスケールと成長をサポートする処理能力、データベースストレージ、およびその他多種多様な機能を提供します。お客様は必要なサービスを選択し、必要な分だけご利用いただけます。それらを活用するために役立つ日本語資料、動画コンテンツを多数ご提供しております。(本サイトは主に、AWS Webinar で使用した資料およびオンデマンドセミナー情報を掲載していません。)

AWS Webinar お申込 »

AWS 初心者向け »

サービス別資料 »

<https://amzn.to/JPArchive>

[申込受付中！] AWS Innovate オンラインカンファレンス



INNOVATE
ONLINE CONFERENCE

2019年4月8日(月) ~ 5月7日(火) 開催

今すぐ無料参加申込み >>

<https://amzn.to/AWSInnovateJP>

AWS Well-Architected 個別技術相談会

毎週”W-A個別技術相談会”を実施中

- AWSのソリューションアーキテクト(SA)に
対策などを相談することも可能

• 申込みはイベント告知サイトから

(<https://aws.amazon.com/jp/about-aws/events/>)

AWS イベント で[検索]



ご視聴ありがとうございました

AWS 公式 Webinar

<https://amzn.to/JPWebinar>



過去資料

<https://amzn.to/JPArchive>

