AWS Black Belt Online Seminar

Amazon EMR
コスト最適化編

川村　誠

Solutions Architect

2024/03

アジェンダ

- Amazon EMR の基礎
- やっておくと良いこと
  - 進化するプラットフォームの利用
  - クラスターのステートレス化
- コスト最適化ベストプラクティス
  - Auto scaling
  - スポットインスタンスの活用
  - クラスター運用の自動化
- まとめ

Amazon EMR の基礎

Amazon EMR

Apache Spark, Presto, Trino, Hadoop, Hive, Hbase, Flink などのオープンソースフレームワークを使用したビッグデータ分析

パフォーマンスが最適化されたランタイムを利用可能

Spark、Hive、Presto、Flinkなどの一般的なフレームワーク向けにパフォーマンスが最適化されたランタイムで、100% オープンソース API互換性を実現

最新の OSS 機能を利用可能

オープンソースでリリースされる新しい機能が60日以内に利用可能になる

ビッグデータ分析に最適なコストパフォーマンス

Amazon EC2 スポット、Amazon EMR マネージドスケーリング、および 1 秒単位の請求を使用してコストを削減

EMR を使用するだけでコストを下げられる

パフォーマンスの向上
= 処理時間の短縮
= コスト削減

コスト最適化を実現する機能を利用可能

- ワークロードに合わせてクラスターの規模を柔軟に変化させる/不要な計算リソースを削減する
- インスタンスのコスト自体を下げる

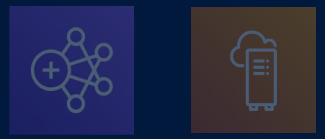
EMR Deployment Options

Amazon EMR on Amazon EC2

ワークロードに対して最高のコストパフォーマンスを発揮するインスタンスを選択可能

Amazon EMR on Amazon EKS

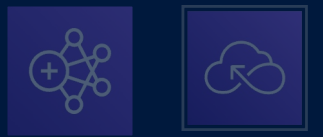
EKS での Apache Spark ジョブの プロビジョニング、管理、スケーリングを自動化



Amazon EMR on AWS Outposts

クラウドの場合と同様に、オンプレミス環境で EMR をセットアップ、管理、スケーリング可能

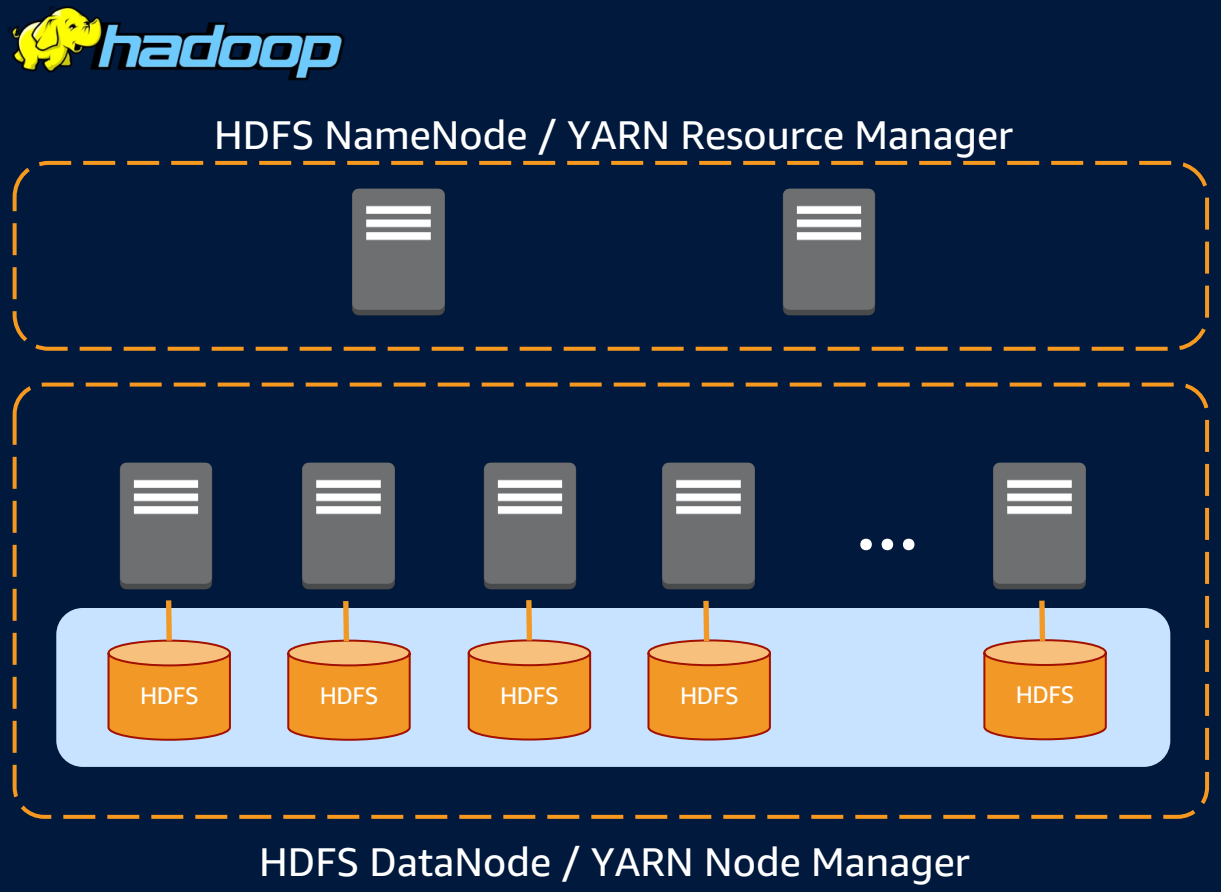
今回の最適化対象



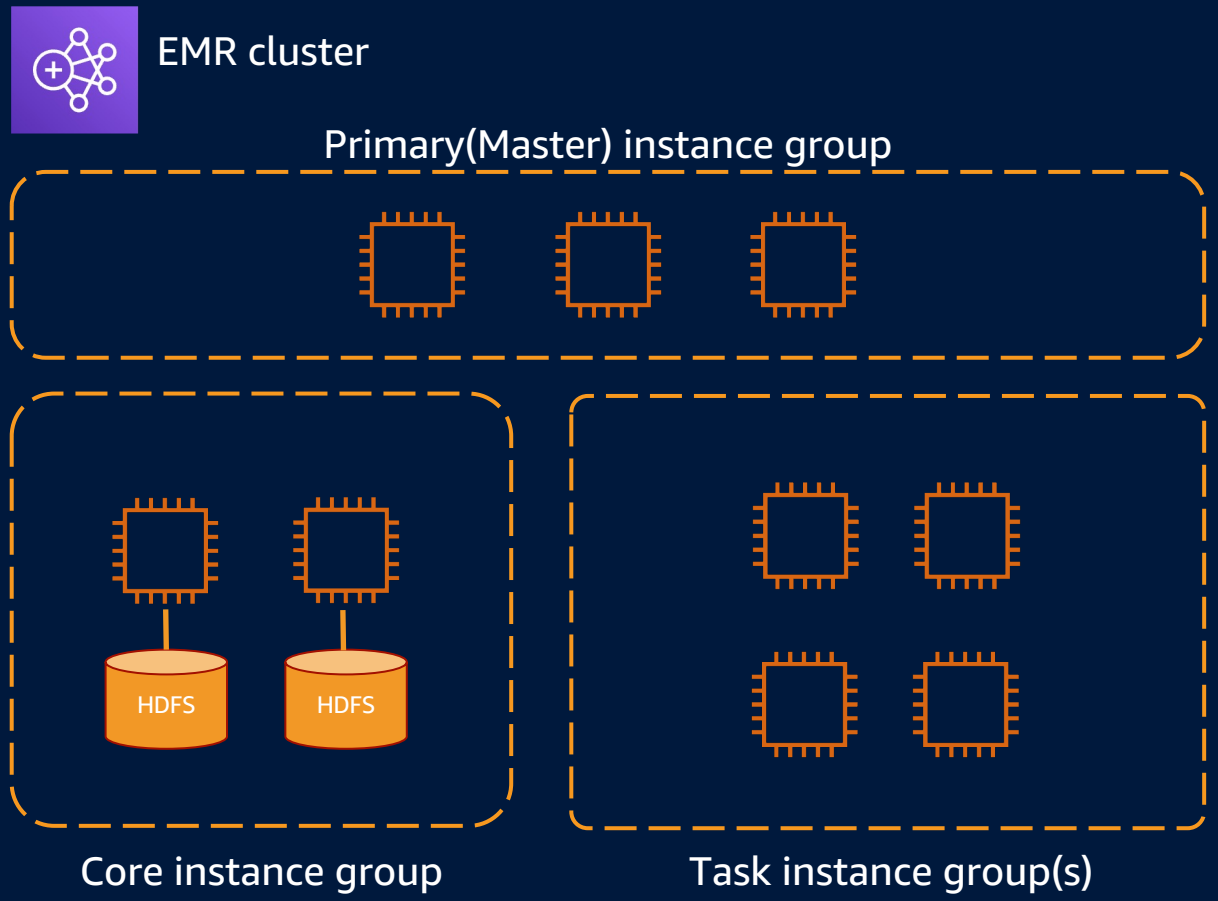
Amazon EMR Serverless

クラスターの管理や運用を行わずに、ペタバイト規模のデータ分析を実行可能

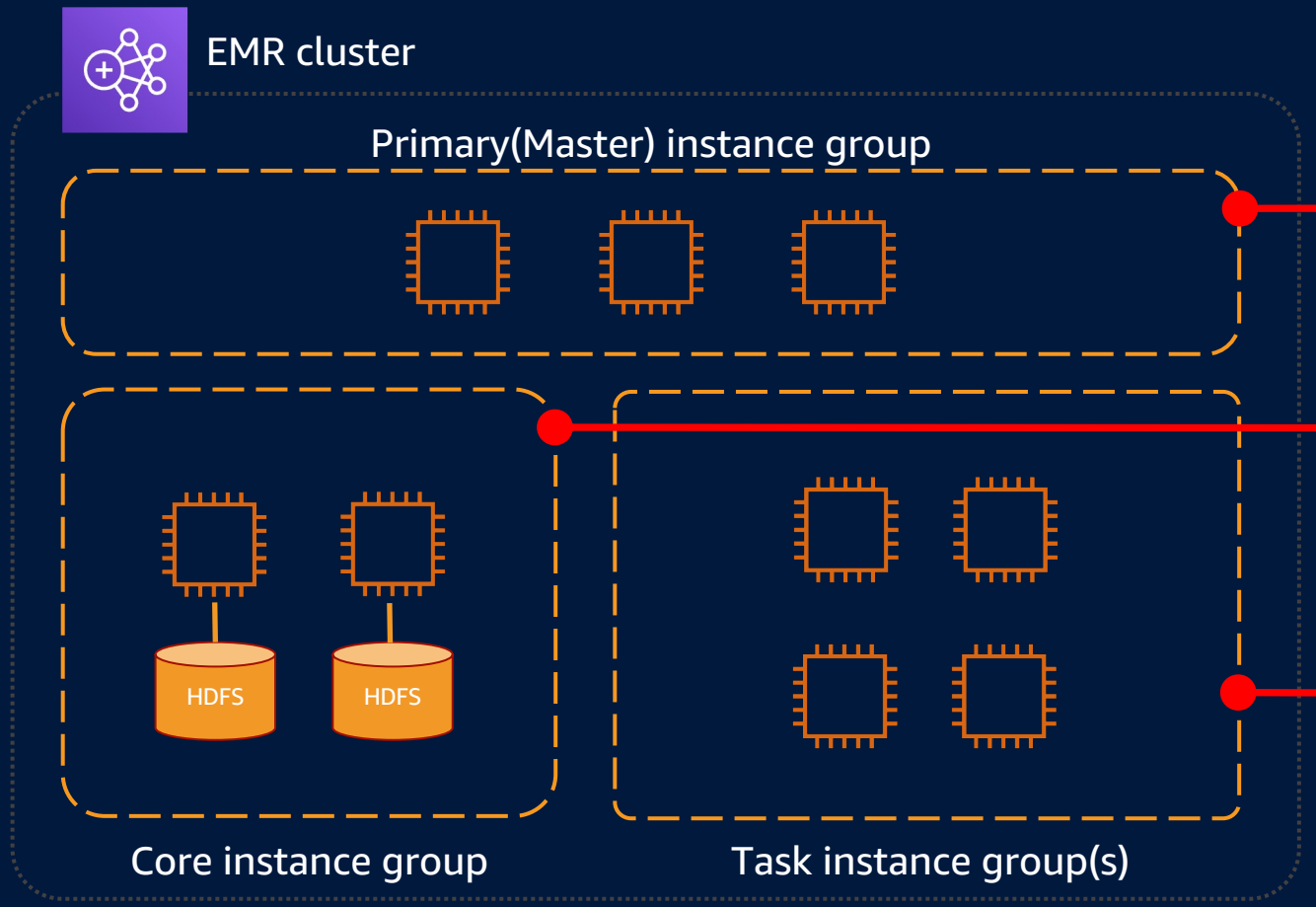
Hadoop クラスター



EMR クラスター



EMR クラスター: 弾力性を実現するノード構成



マスターインスタンスグループ

- マスターノードは稼働し続ける必要がある
- HA構成可能

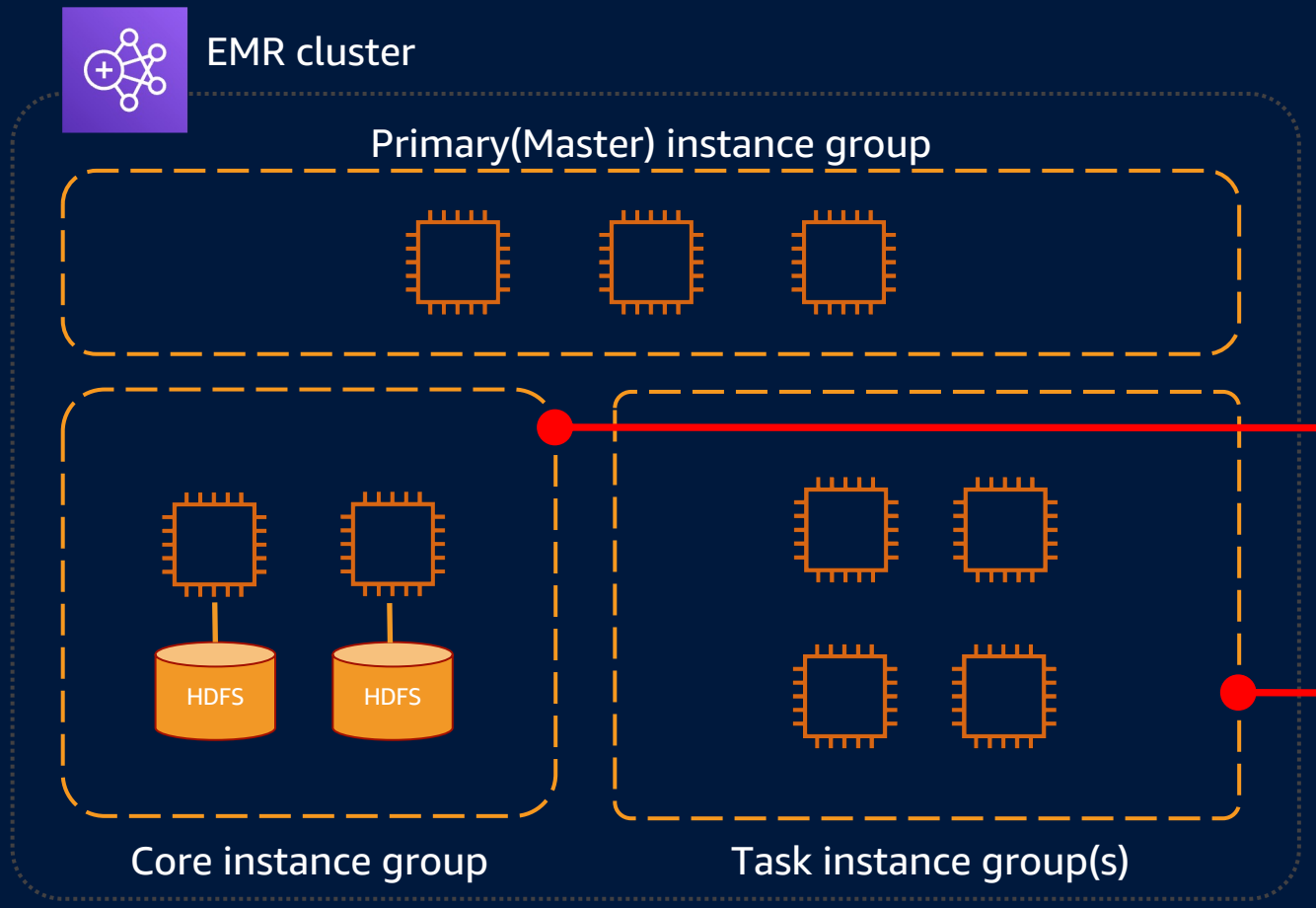
コアインスタンスグループ

- コアノードは HDFS を持つ worker ノード群

タスクインスタンスグループ

- 複数のタスクインスタンスグループを構成可能
- タスクノードは HDFS のない worker ノード群

EMR クラスター: 弾力性を実現するノード構成



コアインスタンスグループ

- コアノードは Graceful に追加および削除が可能

タスクインスタンスグループ

- タスクノードのロストは、クラスターに対する問題なし

EMR on EC2 の利用料金

- Amazon EC2(以下、EC2)インスタンスタイプ毎に決められた Amazon EMR と EC2 インスタンスの利用料金(いずれも 1 秒ごとに課金、最小課金時間は 1 分)
- アタッチされた Amazon Elastic Block Store (Amazon EBS) で発生する料金 (ストレージ、IOPS、スループット)
- Amazon S3 や AWS Glue Data Catalog、AWS Key Management Serviceなど、お使いのアプリケーションが AWS の他のサービスを使用する場合、状況により追加料金が発生

※ EC2 料金にはオンデマンド、1 年間および 3 年間のリザーブドインスタンス、Capacity Savings Plans、スポットインスタンスなど、さまざまなオプションが用意されています

https://aws.amazon.com/jp/emr/pricing/

やっておくと良いこと

進化するプラットフォームの利用

最適化された Spark / Presto ランタイム

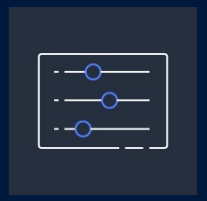
TPC-DS ベースの 3 TB ベンチマークの結果、標準の Apache Spark 3.0 よりも 3 倍 以上高速

TPC-DS ベースの 3 TB ベンチマークの結果、標準の Presto 0.238 よりも 2.6 倍 以上高速

オープンソース API に 100% 準拠しているため、アプリケーションを EMR に簡単に移行可能

パフォーマンス向上はデフォルトで有効

Dynamic-sized executors



Adaptive join selection

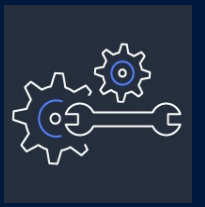
Dynamic pruning of data columns

Operator optimization

Early worker allocation

Intelligent filtering

Parallel/async initialization



Redundant scan elimination

Data pre-fetch

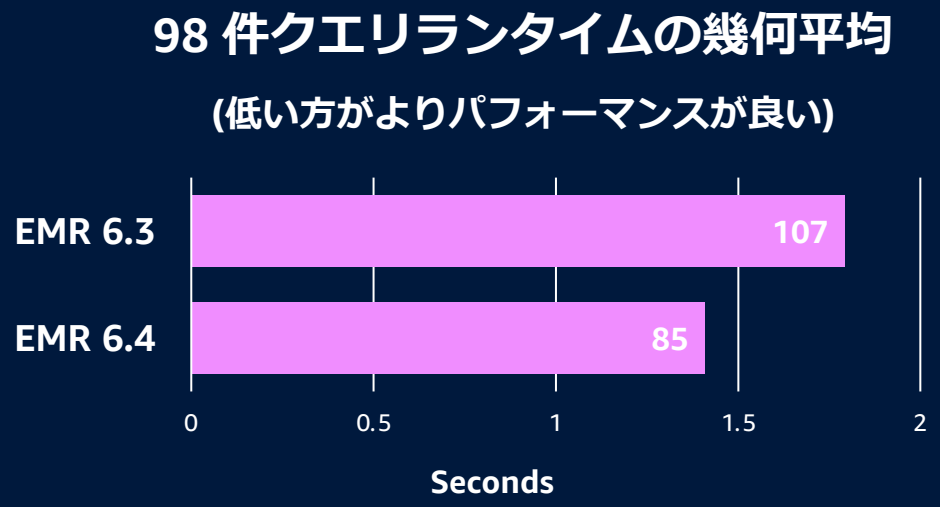
Broadcast join w/o statistics

Stats inference

Optimized metadata fetch

Amazon EMR runtime for Apache Hive

EMR 6.4 の Apache Hive 3.1.2 で 1.25 倍速いパフォーマンス



Apache Hive 3.1.2 on EMR 6.4 vs. EMR 6.3*

**Based on TPC-DS 3 TB benchmarking running 16 node M5.8xlarge cluster*

パフォーマンスが最適化された Apache Hive ランタイムを利用可能

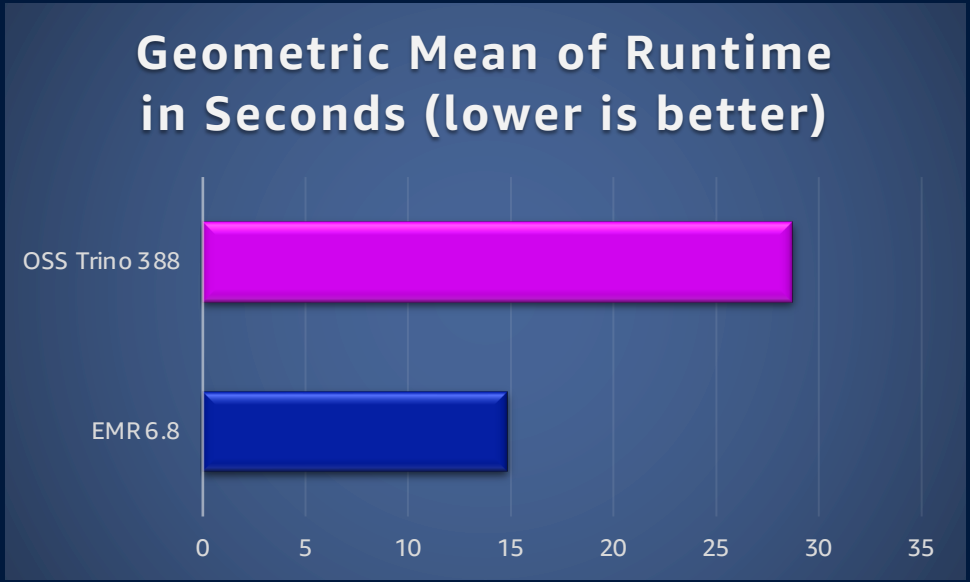
ベストパフォーマンス

- **幾何平均で1.25倍高速**
- **個々のクエリが最大2倍向上**
- **AWS Glue データカタログのクエリプランニング時間を短縮**
- **Amazon S3 からの ORC データのクエリ実行時間を改善**

オープンソースの Apache Hive API に 100% 準拠

Amazon EMR runtime vs OSS Trino

EMR 6.8.0 の Trino 388 で最大 3.1 倍速いパフォーマンス



パフォーマンスが最適化された Trino ランタイムを利用可能

ベストパフォーマンス

- **幾何平均で最大 3.1 倍高速**
- **クエリ実行時間が最大 4.2 倍高速**

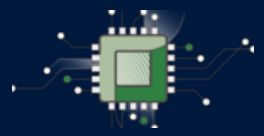
Trino 388 on EMR 6.8.0

*Based on TPC-DS 3TB Benchmarking running 6 node C5.9XL cluster and EMR 6.8.0 running Trino 388

ワークロードに最適なインスタンスタイプの利用

柔軟なコンピューティング能力

汎用



M6 Family
M5 Family
M4 Family

バッチ処理

コンピュート

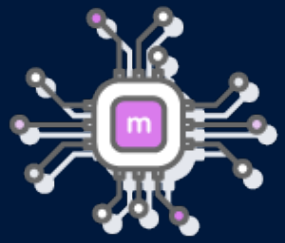
P3 Family
P2 Family

G4 Family
G3 Family

C6 Family
C5 Family
C4 Family

高速コンピューティング
/機械学習

メモリ



R6 Family
R5 Family
R4 Family

対話型の分析

ストレージ

D2 Family
I3 Family

大容量の HDFS

AWS Graviton インスタンスの活用: AWS Graviton2

インスタンスファミリーの中で最高のパフォーマンスを提供

M5(Intel アーキテクチャ汎用) と M6g(Graviton2: Arm アーキテクチャ汎用) 、2つのインスタンスファミリーで構成した EMR (5.30.1) クラスターをTPC-DS ベースの 3 TB ベンチマークで比較

12%～16% の パフォーマンス向上

20% コスト削減

最大 30% のコスト パフォーマンス向上

Resource: https://aws.amazon.com/jp/blogs/big-data/amazon-emr-now-provides-up-to-30-lower-cost-and-up-to-15-improved-performance-for-spark-workloads-on-graviton2-based-instances/

AWS Graviton インスタンスの活用: AWS Graviton3

インスタンスファミリーの中で最高のパフォーマンスを提供

Graviton インスタンスファミリー C6g(Graviton2) と C7g(Graviton3)、それぞれで構成した EMR (6.9.0) クラスターをTPC-DS ベースの 3 TB ベンチマークで比較 (Spark/Trino)

13.65～18.73% の パフォーマンス向上

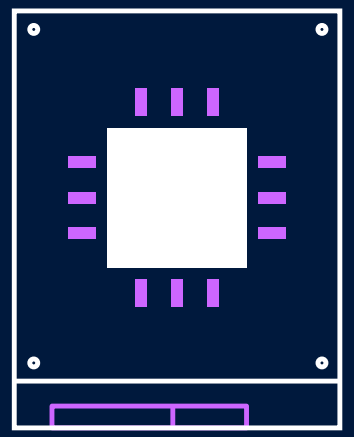
7.93～13.35% コスト削減

7-13% のコスト パフォーマンス向上

Resource: https://aws.amazon.com/jp/blogs/big-data/amazon-emr-launches-support-for-amazon-ec2-c7g-graviton3-instances-to-improve-cost-performance-for-spark-workloads-by-7-13/

Amazon EBS GP3 の活用

コストパフォーマンスを最大化



gp3

General purpose SSD

リレーショナルおよび非リレーショナルデータベース、エンタープライズアプリケーション、コンテナ化されたワークロード、ビッグデータ、ファイルシステム、メディアワークフローに最適なストレージ

3,000 ベースライン IOPS (GP2 では 3 IOP/GiB) 、および、 最大スループット 1,000 MiB/秒 (GP2 では 250 MiB/秒)で、容量とは別に IOPS とスループットをプロビジョニング可能

月額 0.096 USD/GB*、以前の GP2 ボリュームよりも最大 20% 低いストレージ価格

*東京リージョンでの価格

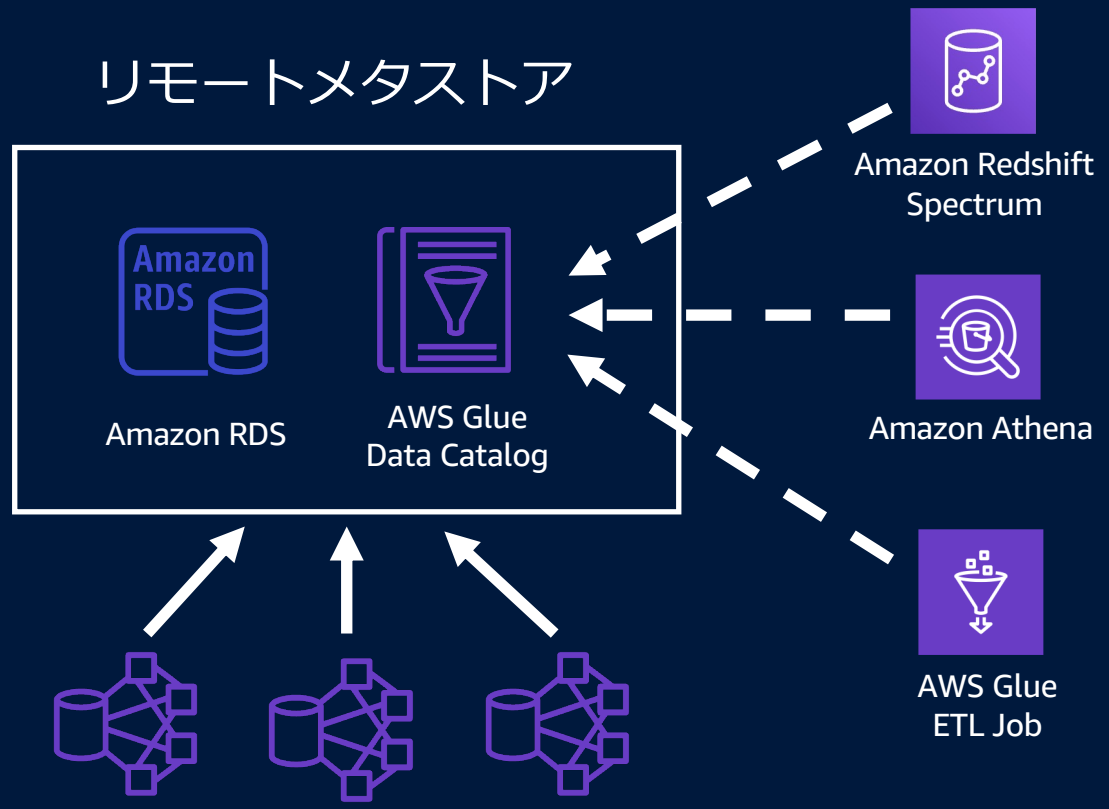
Resource: https://aws.amazon.com/jp/about-aws/whats-new/2020/12/introducing-new-amazon-ebs-general-purpose-volumes-gp3/

クラスターのステートレス化

クラスターのステートレス化：メタデータ

リモートメタストアの利用

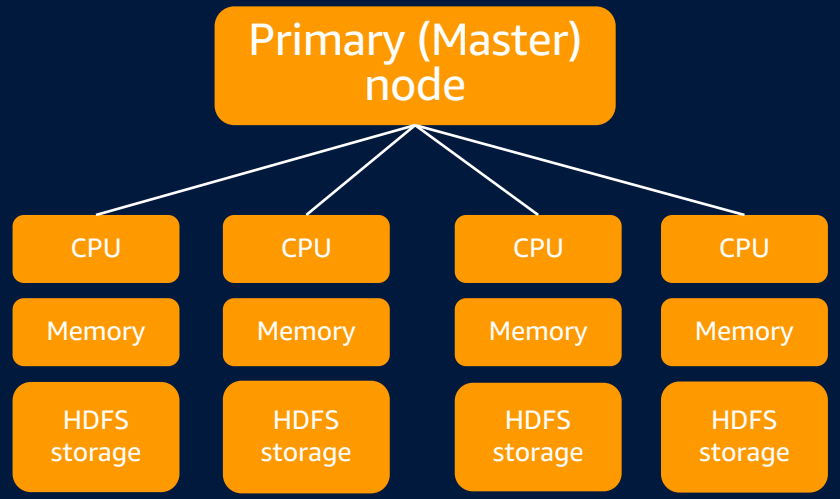
- メタストアをクラスタ外に保持
- 起動時間が短縮され、コストが削減される



クラスターのステートレス化：ストレージ

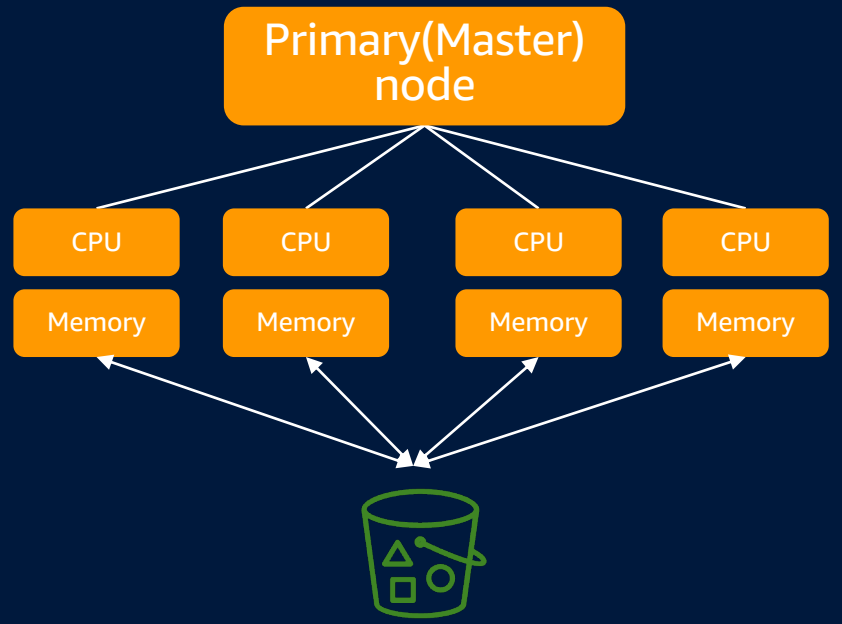
EMR File System (EMRFS) の利用

HDFS を利用する場合



- HDFS の基本戦略ではデータの冗長度は 3
- ストレージコストがデータ量の3倍かかる

EMRFS を利用する場合



- EMRFS を利用すると HDFS のインタフェースから透過的に S3 上のデータを参照可能
- データの耐久性は S3 にオフロード

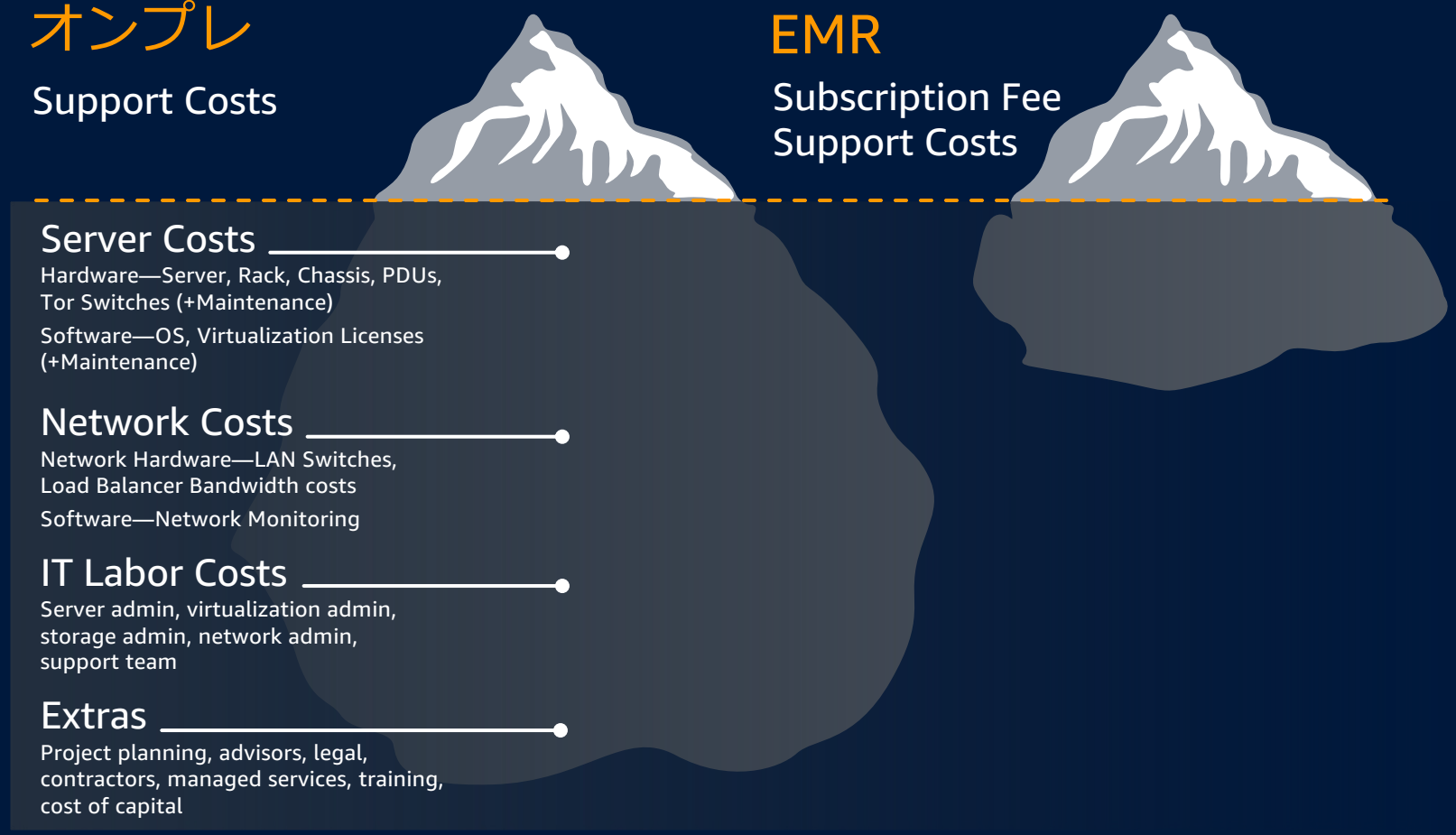
一時的なクラスター:

コンピュートとストレージの分離で低コストを実現

コスト最適化ベストプラクティス

低コスト

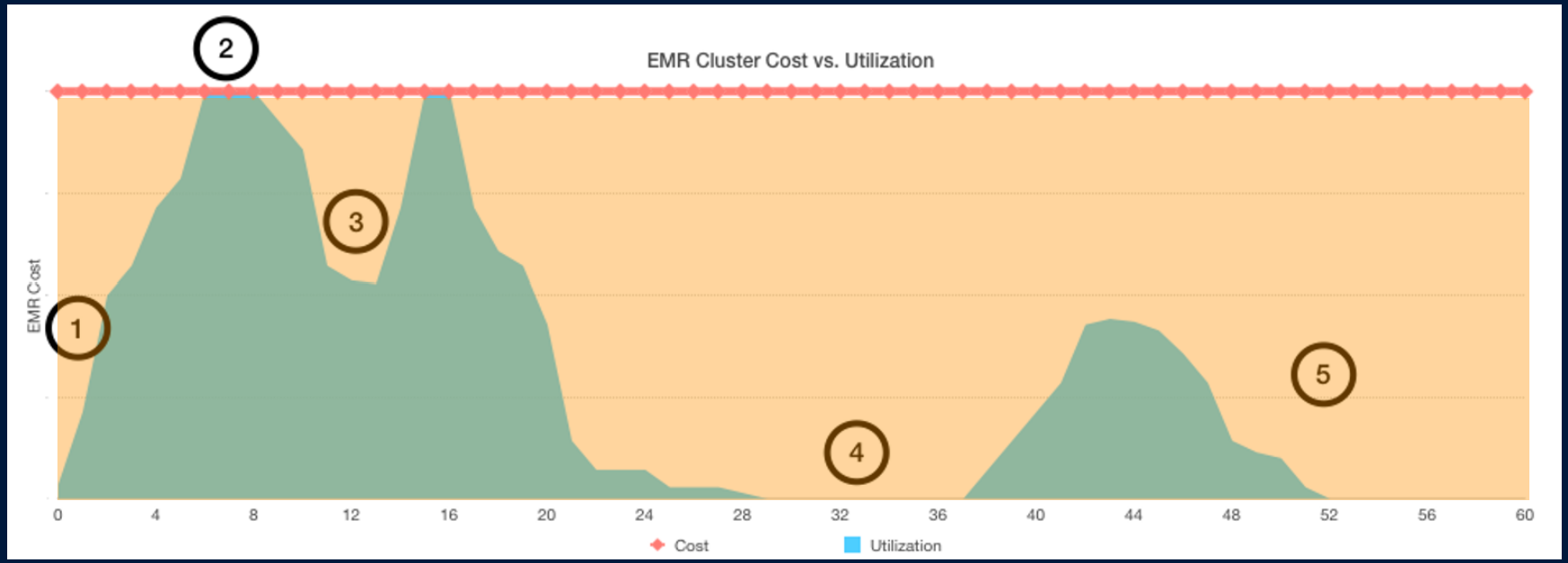
低い TCO (Total Cost of Ownership)



- Hadoop クラスターの管理とサポートにかかる管理時間の短縮
- 前払いコストなし : ハードウェアの取得、設置
- オペレーションコストの節約 : データ・センターのスペース、電力、冷却
- ビジネス価値 : 遅延コスト、リスクプレミアム、競争力、ガバナンスなど

ワークロード例

1. 変化をともなうワークロード
2. クラスターのキャパシティを超えるピークワークロード
3. ピークの繰り返し
4. アイドルタイム
5. 複数ワークロードの実行



Auto Scaling

マネージドスケーリング

クラスタサイズを調整してコストを自動的に削減

継続的に改善されるアルゴリズムで完全にマネージド化されたエクスペリエンスを提供

マネージドスケーリングで高解像度メトリクスを実現

最小/最大コスト制約設定のみで利用可能

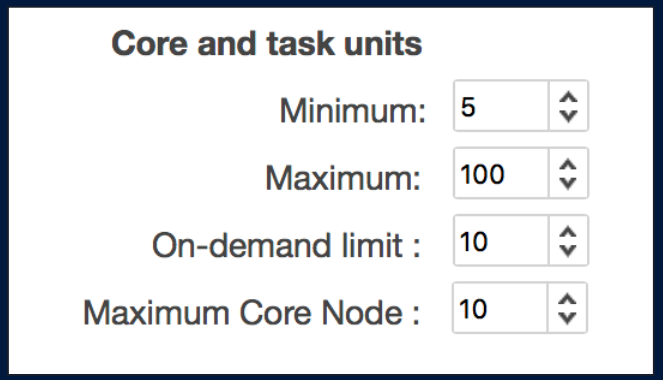
Auto Scaling よりもデータポイントが多く、反応時間が短い

コストを20% ～ 60%節約可能

マネージドスケーリング

クラスタサイズを調整してコストを自動的に削減

- 下記パラメータを設定するだけで利用可能
  - 最小 – クラスターの最小ユニット数
  - 最大 – クラスターの最大ユニット数
  - オンデマンド制限 – オンデマンドユニット数の上限
  - 最大コアノード数 – コアノードユニット数の上限

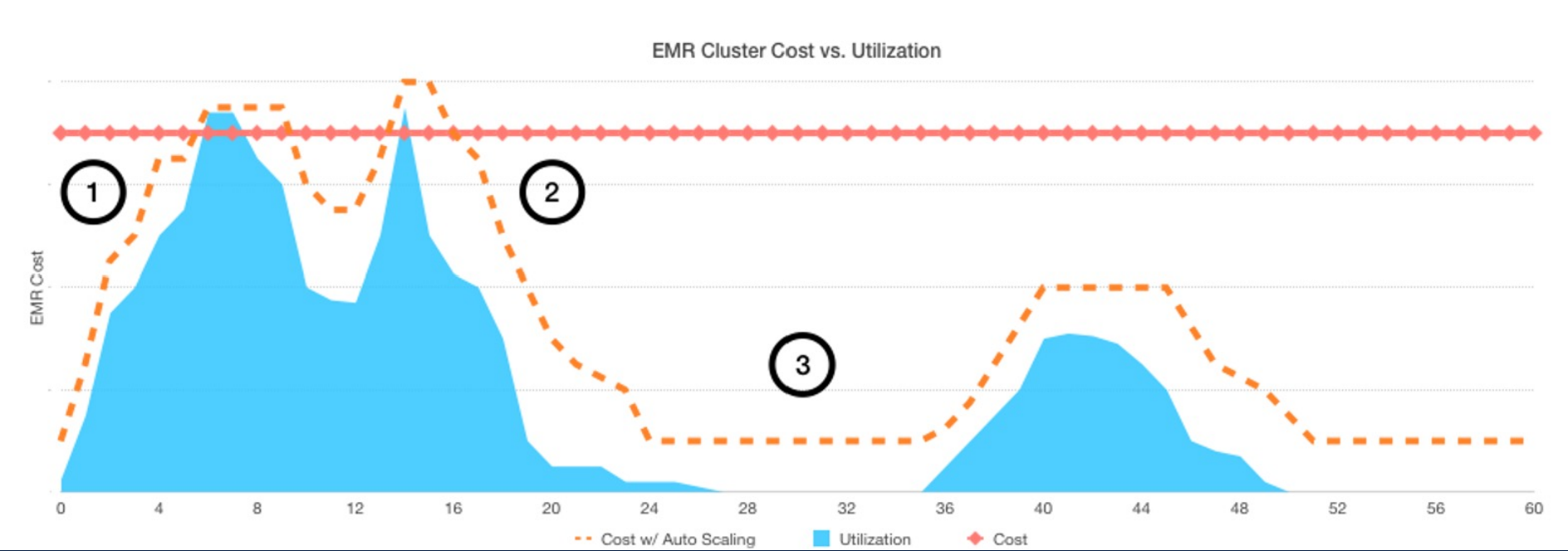


- 制限事項
  - Spark、Hadoop、Hive、Flinkなどの YARN アプリケーションでのみ機能
  - EMR-6.0.0 を除く、EMR-5.30.0 以降の EMR で利用可能

https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-managed-scaling.html

ワークロード例：弾力性によるコスト最適化

1. ワークロードに対してクラスターをスケールイン/アウト
2. ピークワークロードに対処可能
3. アイドルリソースの削減

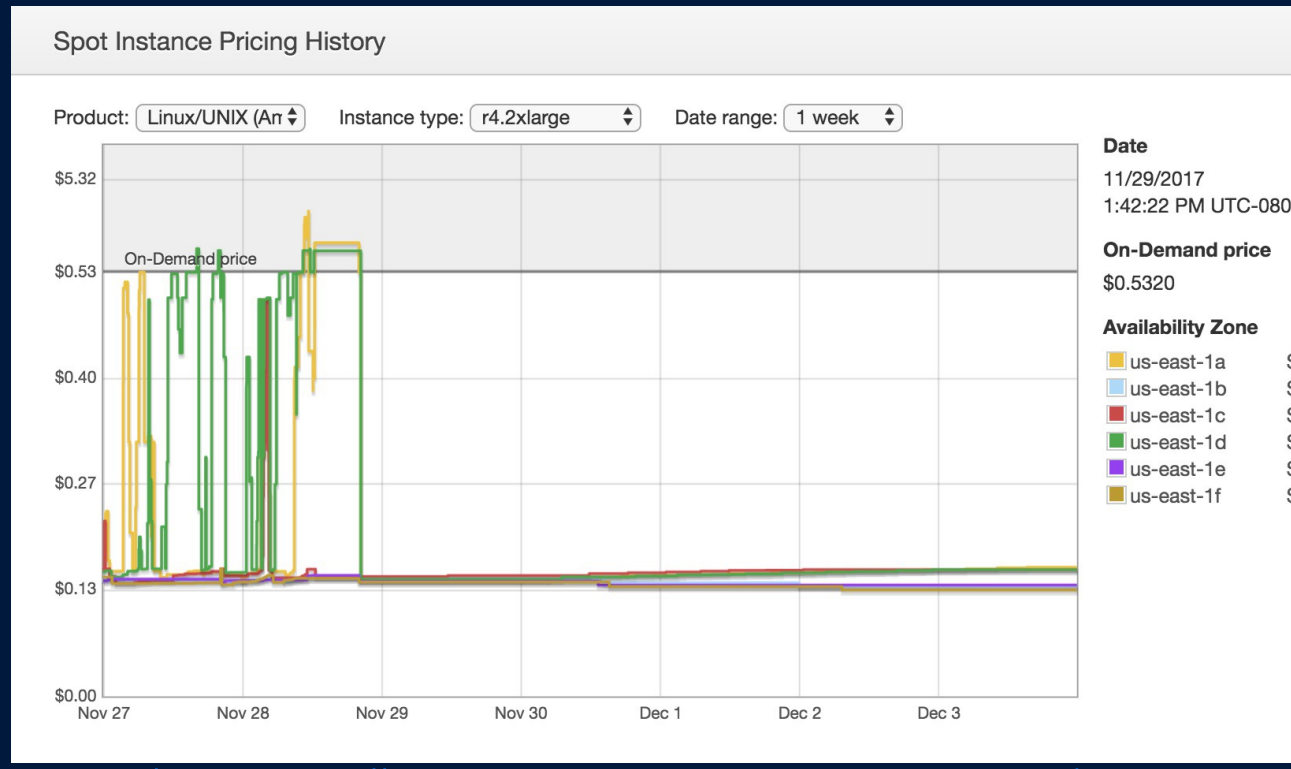


スポットインスタンスの活用

スポットインスタンス

より少ないコストでコンピューティングを加速

- Amazon EC2 の予備容量から
  オンデマンド(OD)価格の
  最大 90 %割引でインスタンスを
  利用可能
- 料金は長期的な需要と供給に
  基づいて決められ、スムーズ
  に変化
- 中断は Amazon EC2 が OD 用の
  キャパシティを必要とする場合
  にのみ入札なしで発生



2017年11月
以前(入札有)

現在の
料金モデル

スポットインスタンスで得られる EMR のメリット

計算処理を加速

スポットインスタンスを利用するとオンデマンドインスタンスを利用する料金で、よりたくさんのインスタンスを実行可能

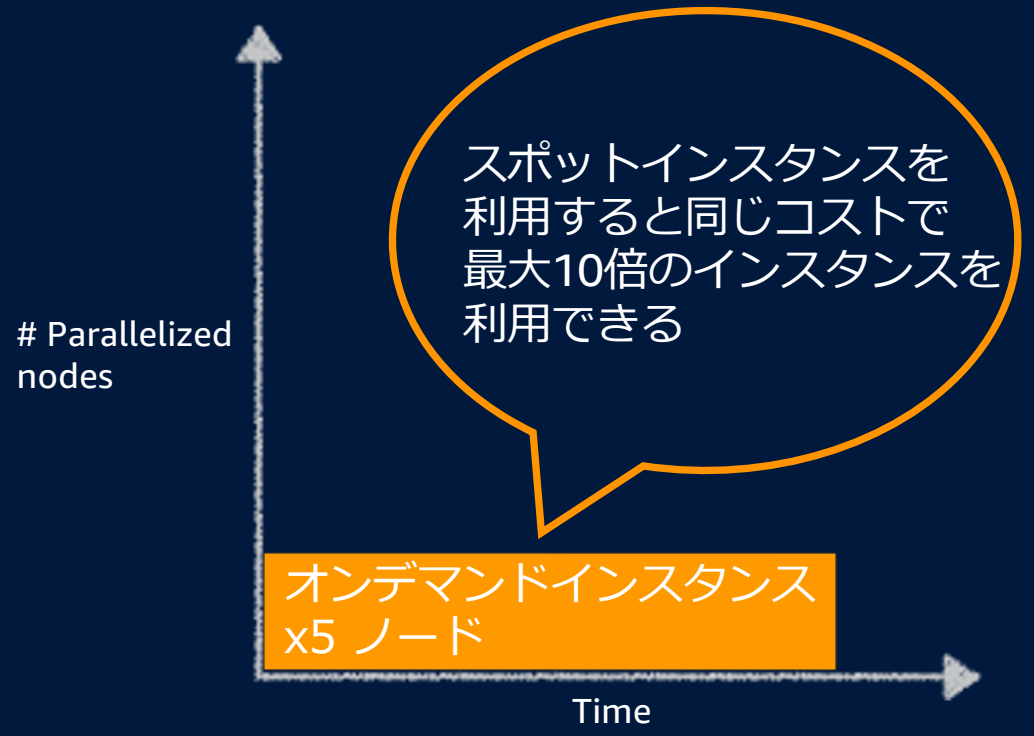
さらなるコスト削減

オンデマンド価格の最大 90 % 割引でインスタンスを利用可能

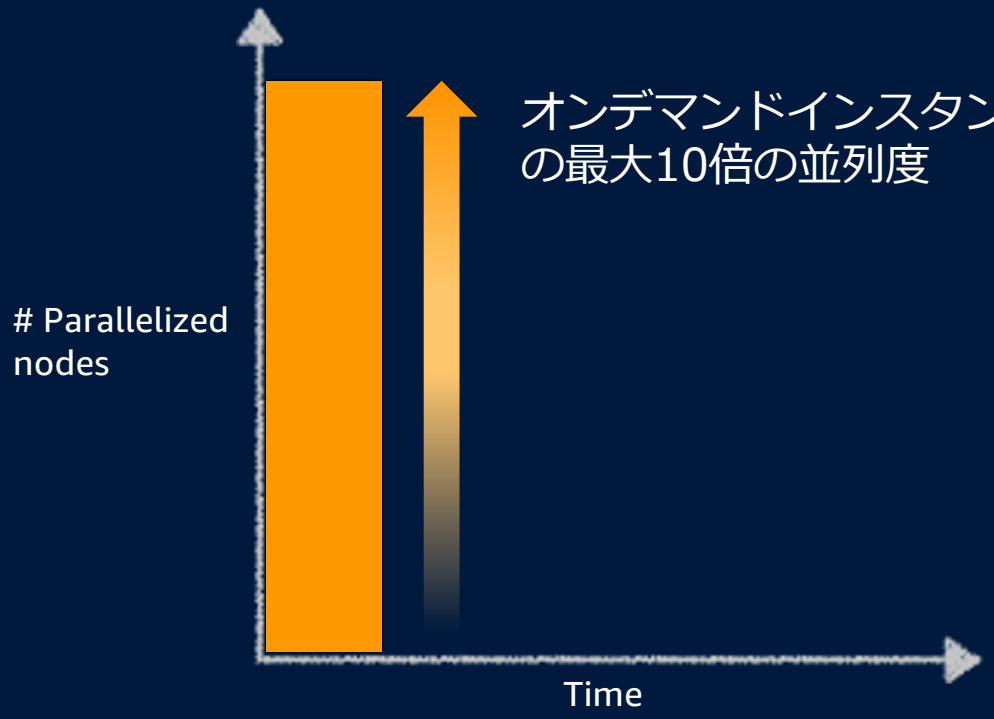
スケールに合わせて構築

ワークロードの並列度を上げ、ジョブの実行時間を短縮できる

スポットインスタンス利用でワークロードの並列度を上げ、ジョブの実行時間を短縮できる



ジョブ実行時間: **10 時間**



ジョブ実行時間: **1 時間**

費用対効果の具体例

10 node cluster running for 14 hours
Cost = 1.0 * 10 * 14 = $140

費用対効果の具体例

Add 10 more nodes on Spot

費用対効果の具体例

20 node cluster running for 7 hours

Cost1 = 1.0 * 10 * 7 = $70
(オンデマンド)

Cost2 = 0.5 * 10 * 7 = $35
(スポット)

Total = Cost1+Cost2 = $105

費用対効果の具体例

費用対効果:

50 % 処理時間削減 (14 → 7)

25% コスト削減 (140 → 105)

オンデマンドインスタンスとスポットインスタンスで柔軟なサービスレベルを定義できる

予測可能なコストでサービスレベルにあったクラスターを構築可能

サービスレベルを超える予測不可能なワークロードに対するリソースコストをスポットインスタンスで極小化可能

コアノードにオンデマンドインスタンスを利用

EC2 の標準利用料でコストを計算可能

タスクノードにスポットインスタンスを利用

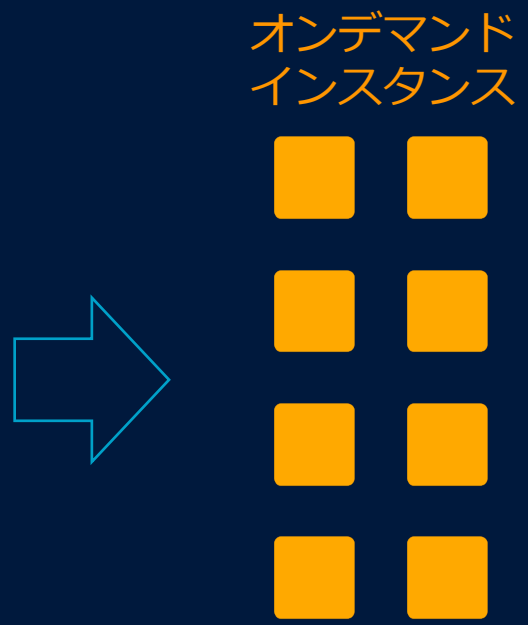
オンデマンド価格の最大 90 % 割引価格

オンデマンドインスタンスのコストを最適化する

予測可能なコストでサービスレベルにあったクラスターを構築可能

コアノードにオンデマンドインスタンスを利用

EC2 の標準利用料でコストを計算可能



ワークロードの稼働時間に合わせてコスト最適化を検討可能

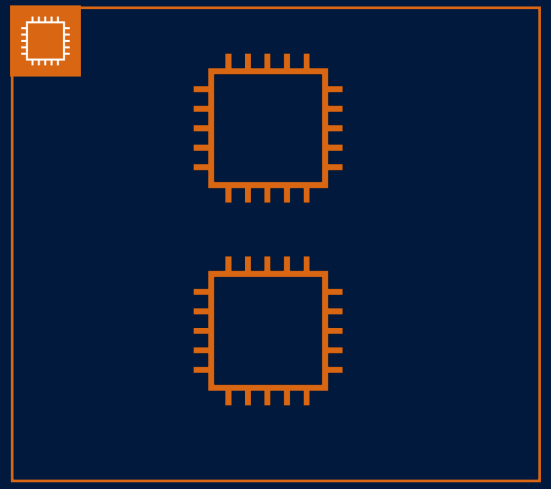
- 永続的に稼働するワークロードの場合、Reserved Instance (RI) を活用することで最大 70 % の EC2 利用料を削減可能！
- 永続的では無い場合、稼働時間で EC2 のコストを見積り、RI 利用時のコストと比較し、コストが低い選択肢を採用する

フレキシブルインスタンスフリート

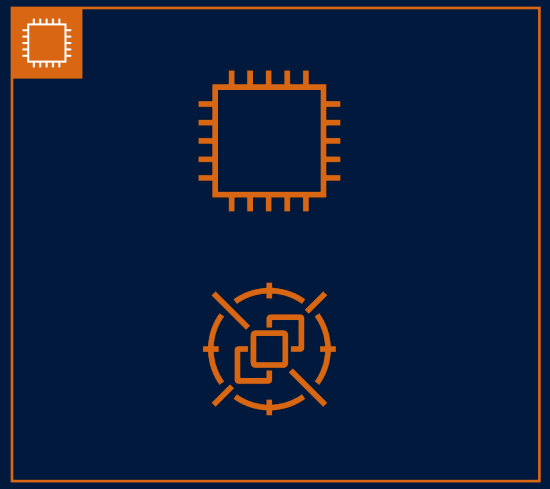
フレキシブルインスタンスフリート

EMR でスポットインスタンスを使いこなすための機能

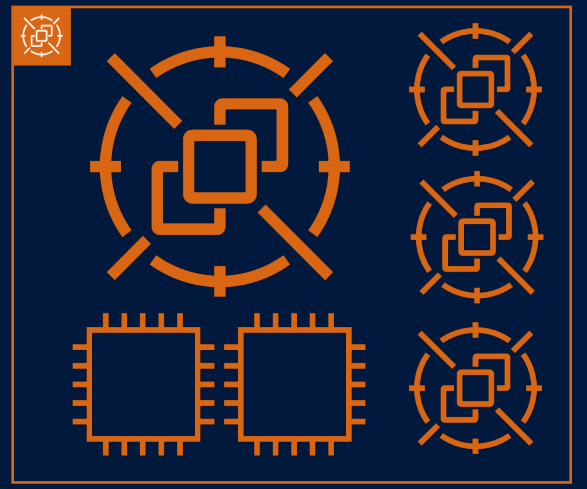
Primary(Master) nodes



Core instances



Task instances



- 指定した複数のインスタンスタイプリストから、スポット / オンデマンドでインスタンスをプロビジョニング可能とする機能
  (以下、インスタンスフリート)
- 容量/価格に基づいて最適なアベイラビリティゾーンを EMR が選択

EMR のインスタンスフリートに最適なスポットインスタンス

- ✓ ノードは、オンデマンドインスタンスとスポットインスタンスが混在するように設定可能
- ✓ スポットインスタンスが中断されると、フリート内の別のインスタンスを追加することで設定されたリソースを維持するように機能する

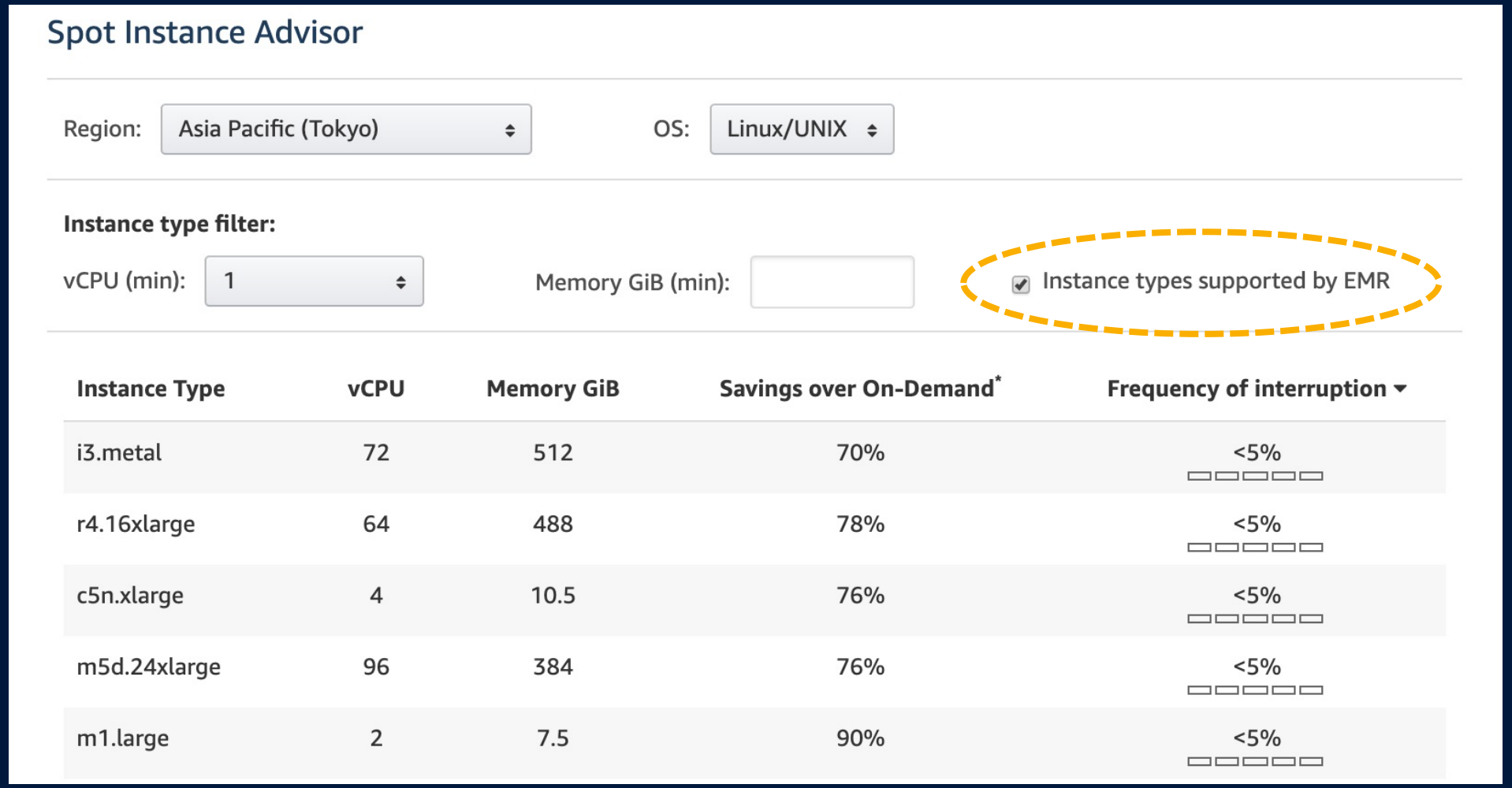
割り当て戦略

スポットインスタンスの中断発生率を制御する

- ✓ 起動するスポットインスタンスに対して、4つの選択肢からスポット戦略を選択可能
  - ► 料金キャパシティ最適化 (推奨)
  - ► キャパシティ最適化
  - ► 最低料金
  - ► すべてのプール間で分散
- ✓ オンデマンドインスタンスは最低料金戦略に基づく

(Tips)
フリートに指定可能なインスタンスタイプ数は最大 5 種類ですが、割り当て戦略を利用する場合、API/CLI からの実行で、各種フリートに最大 30種類のインスタンスタイプ を設定可能

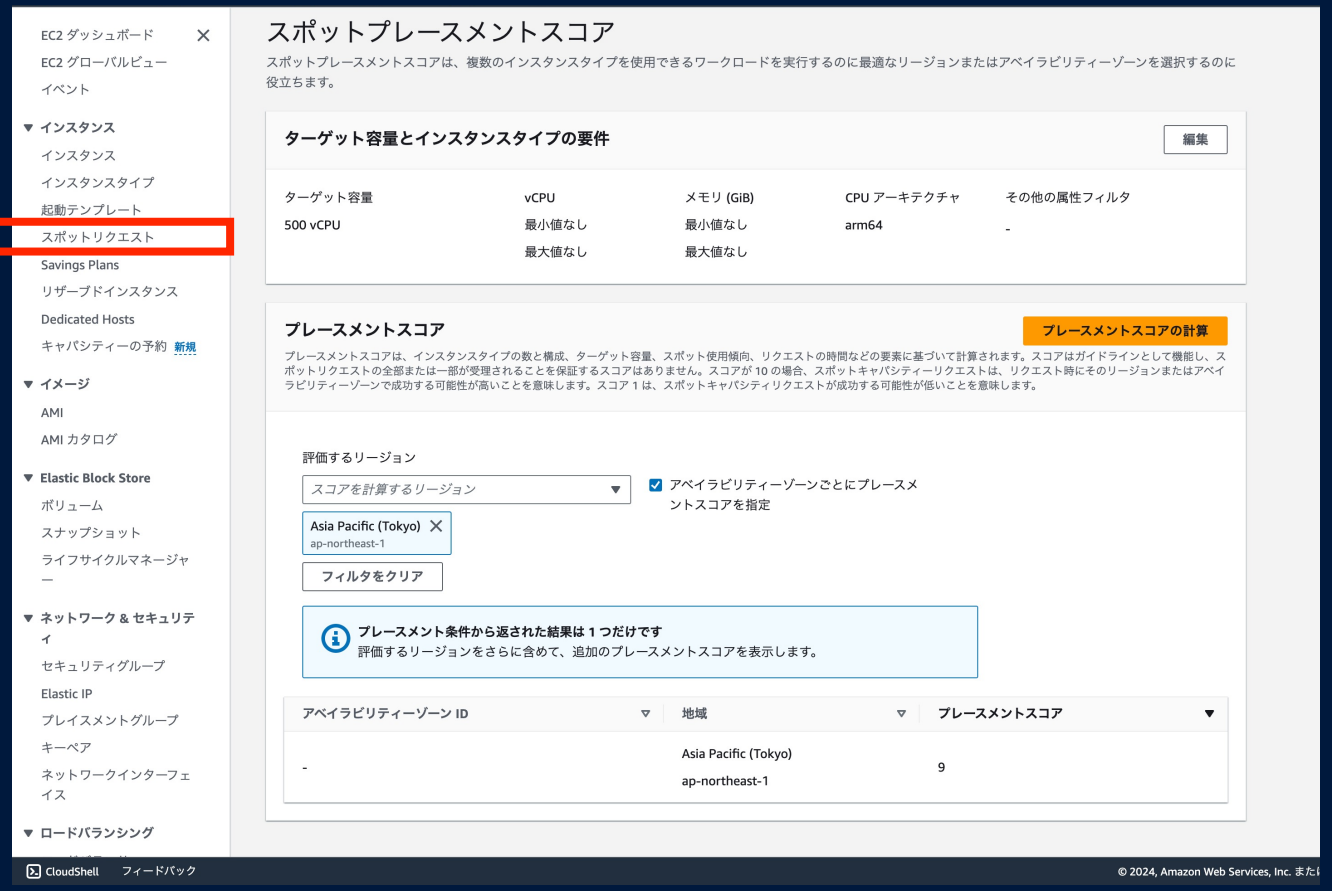
Spot Instance Advisor の活用



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

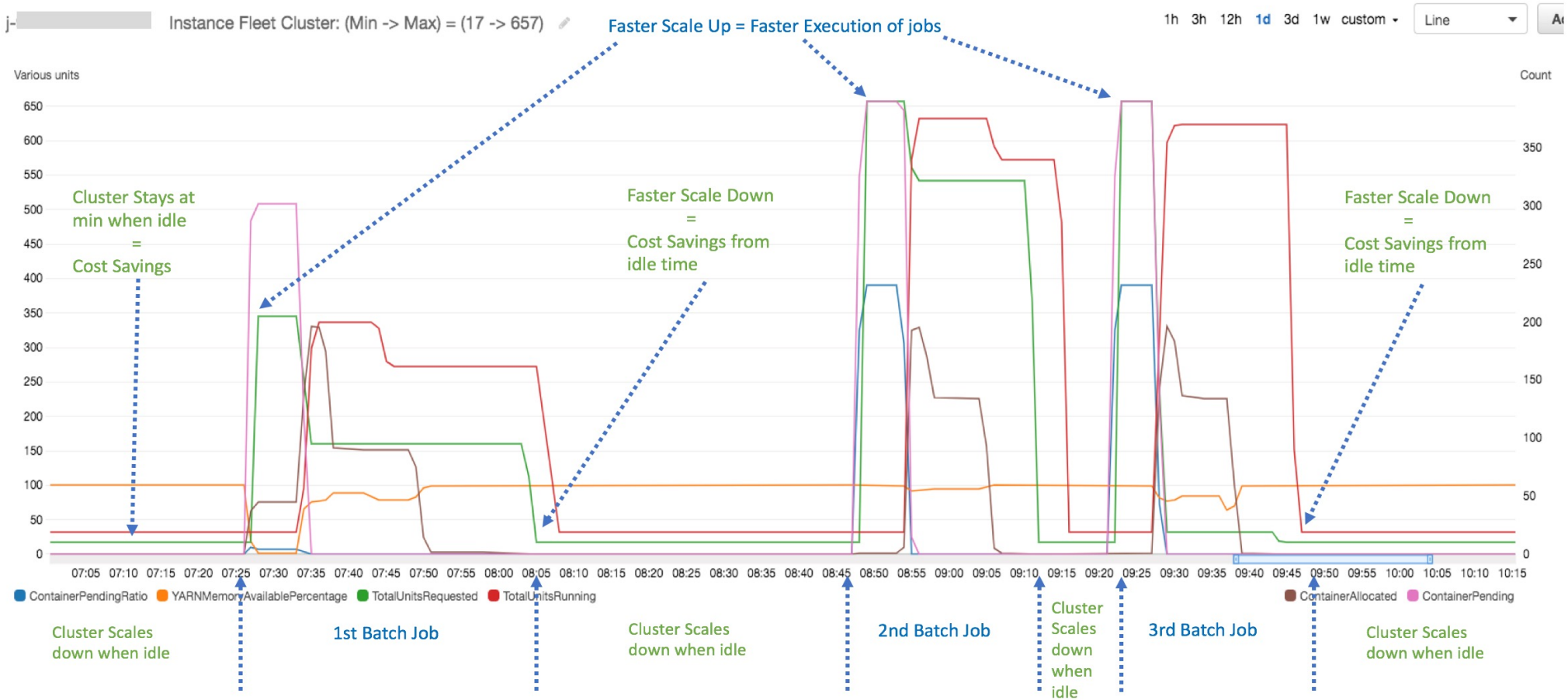
Amazon EC2 スポット配置スコアの活用

- 特定の時点でスポットインスタンスを起動する際に、どのリージョンまたはAZ が基準として最も適しているかを示す機能
- EC2 の容量は刻々と変化するため、結果はリクエストごとに異なる
- スポットインスタンスの利用体験が最高になるように、お客様がインスタンスタイプと AZ の最適な組み合わせを選択できるよう支援するための機能



https://ap-northeast-1.console.aws.amazon.com/ec2/home?region=ap-northeast-1#SpotPlacementScore:

マネージドスケーリング – インスタンスフリート



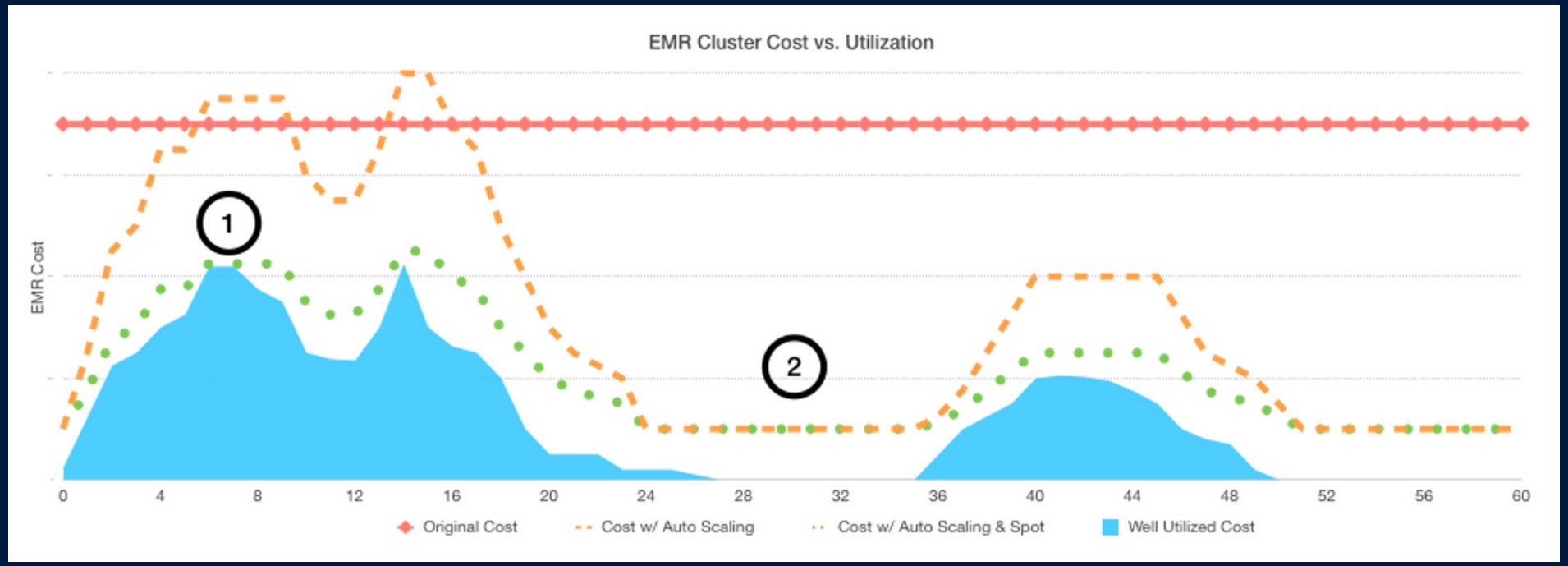
さまざまなシナリオでオンデマンドインスタンスとスポットインスタンスを組み合わせて活用する

Scenario	Master node	Core nodes	Task nodes
長時間稼働クラスターとデータウェアハウス	On-demand	On-demand or instance-fleet mix	Spot or instance-fleet mix
コスト重視ワークロード	Spot	Spot	Spot
データクリティカルなワークロード	On-demand	On-demand	Spot or instance-fleet mix
アプリケーションテスト	Spot	Spot	Spot

Ref: ガイドラインとベストプラクティス
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html

ワークロード例：
スポットインスタンスによるコスト最適化

スポットインスタンスを活用することでコストを更に削減することが可能



クラスター運用の自動化

クラスター運用自動化の選択肢

Step API を利用した
クラスターの設定とジョブ
の送信

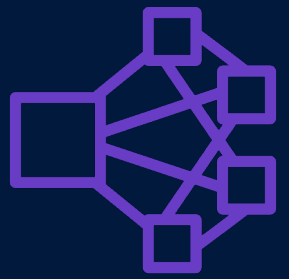
Amazon EMR
Step API

Amazon Managed Workflows
for Apache Airflow(MWAA) /
AWS Lambda / AWS Step Functions
から EMR Step APIをコール、
または、クラスター内の
アプリケーションに直接コール

AWS Lambda

MWAA

AWS Step Functions



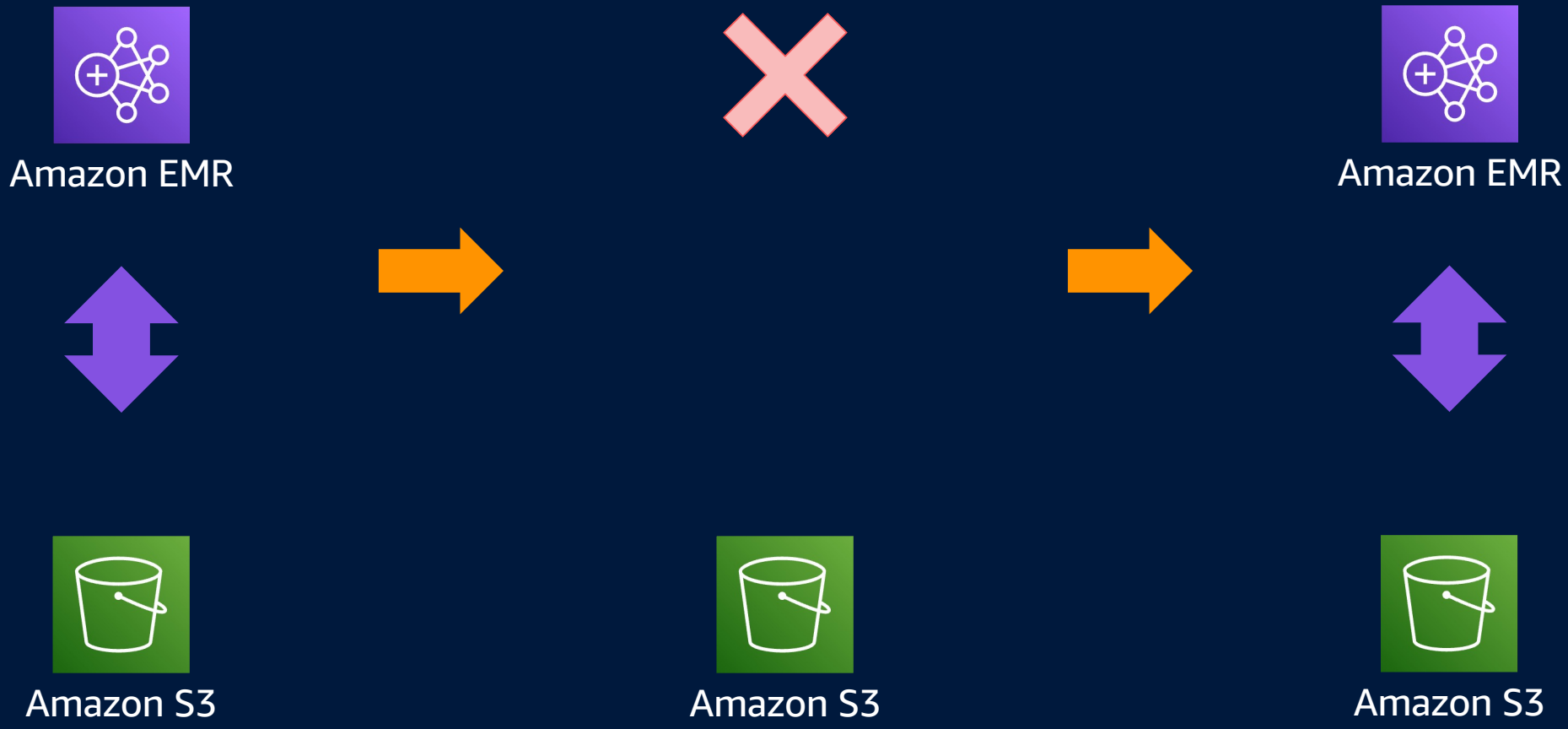
Amazon EMR

ジョブ送信のスケジューリングや、
複雑なワークフローを定義するため、
OSS スケジューラを利用

Airflow、Luigi、Digdag、その他
任意のスケジューラー on EC2

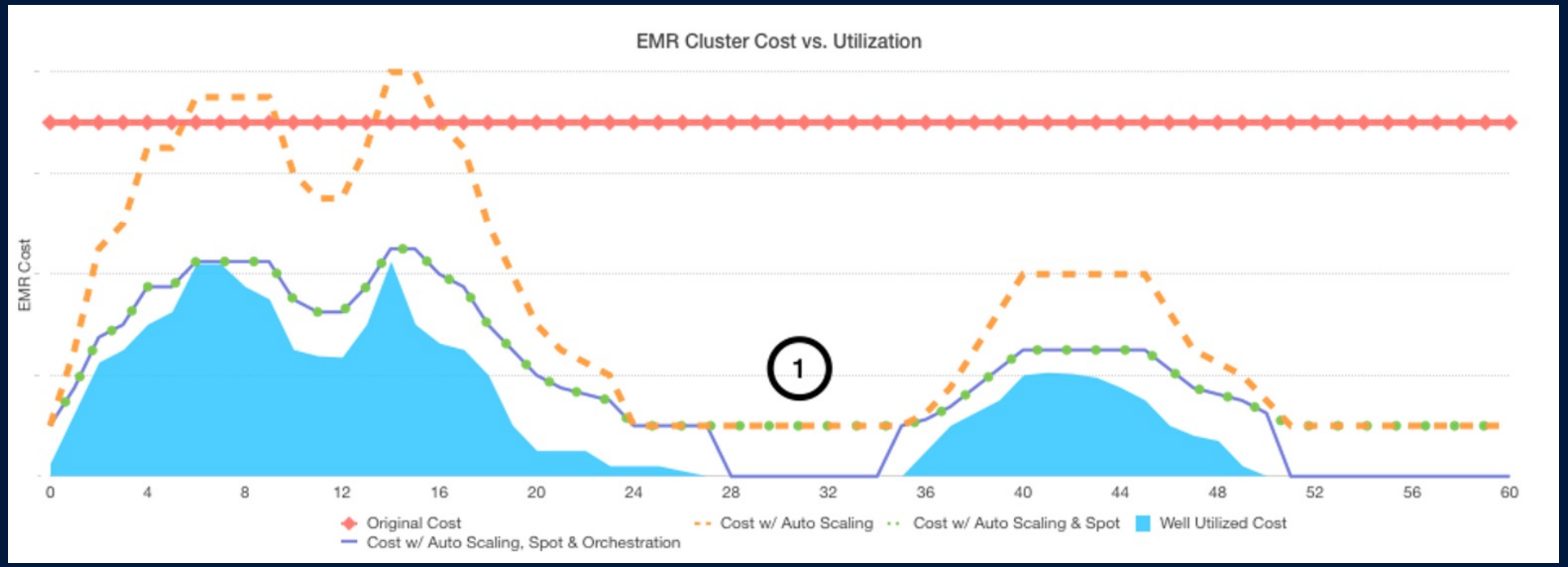
一時的なクラスター:

コンピュートとストレージの分離で低コストを実現



ワークロード例：クラスター運用の自動化

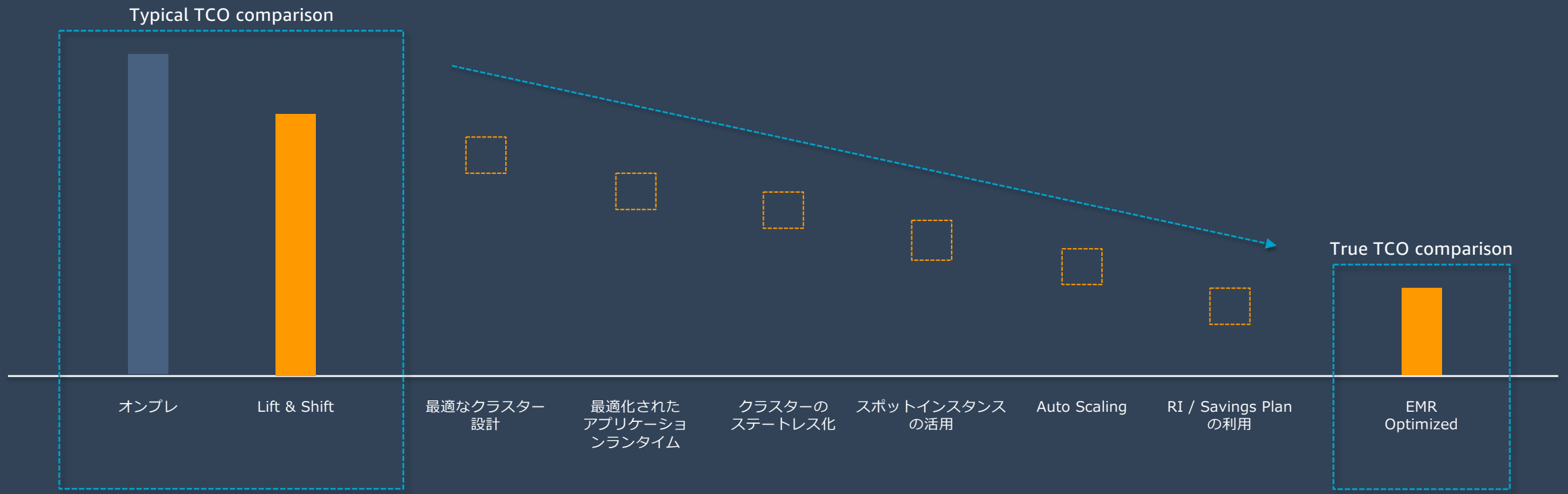
ワークロードが発生しない場合にクラスターを停止することでコストを削減



まとめ

まとめ

EMR はワークロードに対し、クラスターのコストパフォーマンス最適化を実現するための多くのオプションを提供し、進化を継続中



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!