

ここがすごいよ Amazon EC2

寺部 祐菜

技術統括本部 ソリューションアーキテクト
Amazon ウェブ サービス ジャパン合同会社

滝口 開資

コンピュート事業本部 シニアスペシャリスト
ソリューションアーキテクト
Amazon ウェブ サービス ジャパン合同会社

自己紹介

寺部祐菜（テラベ ユウナ）

エンタープライズ技術本部
製造グループ 組立製造ソリューション部
ソリューションアーキテクト



好きなAWSサービス
Amazon EC2

本セッションについて

本セッションの対象者：

- Amazon EC2 について入門的な内容は知っているが、もう一步深く理解したい方
- Nitro System について理解したい方
- Amazon EC2 をコスト効率よく使うオプションについて知りたい方

本セッションのゴール：

- Amazon EC2 の特徴と魅力を、
Nitro System、パフォーマンス、ハイブリッド環境、費用の最適化の観点で理解する。

本セッションでお話しないこと：

- Amazon EC2 の基礎、ご利用手順
- 各種インスタンスのご紹介

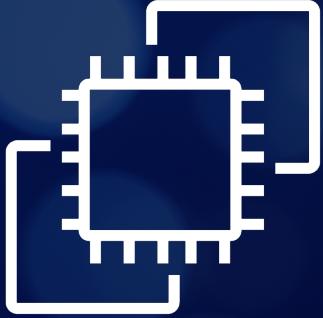
アジェンダ

- はじめに
- Amazon EC2 の特徴・魅力を伝える4つの柱
 - AWS Nitro System
 - パフォーマンス
 - ハイブリッド
 - 費用の最適化
- まとめ

はじめに



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Amazon EC2 は、**信頼性**が高く、**安全**で、
サイズ変更が可能なコンピューティングキャパシティと、
事実上あらゆるワークロードの要求に応える
柔軟なオプションを利用可能

幅広いインスタンスの選択肢

ファミリー
c6gn.8xlarge
世代 機能 サイズ



Amazon EC2 における CPU・アクセラレーター の選択肢

CPU



Intel Xeon Scalable
processor
(x86_64 アーキテクチャ)



AMD EPYC
processor
(x86_64アーキテクチャ)



AWS Graviton
Processor
(64-bit Armアーキテクチャ)

アクセラレーター



A100 , A10
and
T4 Tensor
Core GPUs



FPGAs for
custom
hardware
acceleration



Radeon Pro
V520 GPU



Habana-
based
instances



AWS
Trainium



AWS
Inferentia

Amazon EC2 で提供されるOS の選択肢



Windows



Mac OS



CentOS



※プレビュー

差別化につながらない業務を廃止し、 ビジネスの価値創出にフォーカスする

データセンターの
構築・管理に
より**重点を置かない**



ビジネスと研究構想に
より**重点を置く**

コンピューティングにおいて 幅広い選択肢と柔軟性が求められている



コストと
パフォーマンスの
最適化

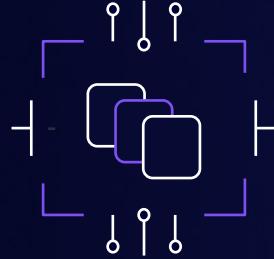


ワークフローと
ビジネスニーズに合わせて
インフラストラクチャを
カスタマイズ

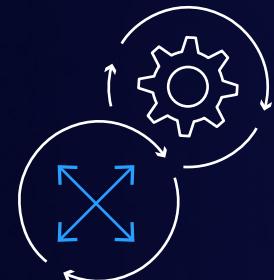


ワークフローを
独自の条件で実行

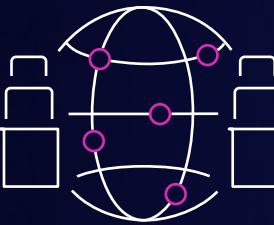
Amazon EC2 はハイパースケールのコンピューティングをより簡単に実現



あらゆるユースケースに最適な、最も幅広く、最も深いコンピューティングオプションのポートフォリオ



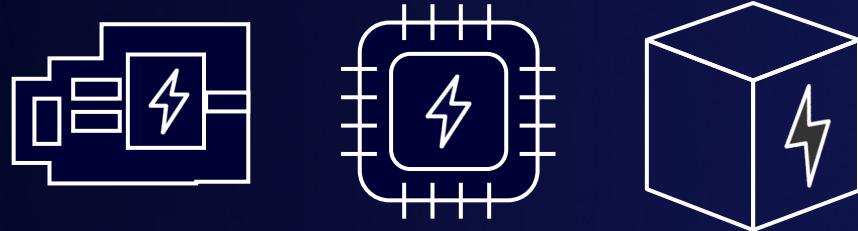
クラウドでコンピューティングキャパシティのサイズを変更できる、柔軟性の高い価格設定オプションで利用可能



サービス機能およびインスタンスファミリーの新しいアップデート、新しいエッジおよびハイブリッド環境を提供

Amazon EC2 の特徴・魅力を伝える 4つの柱

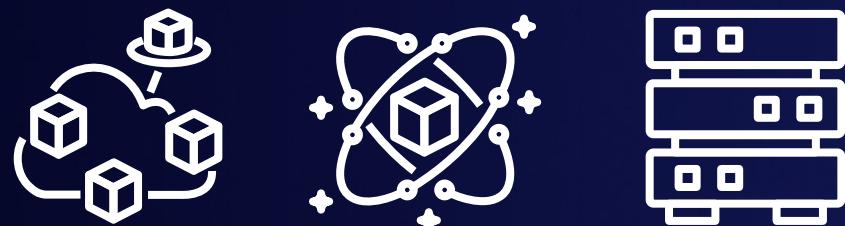
Nitro System



パフォーマンス



ハイブリッド



費用の最適化

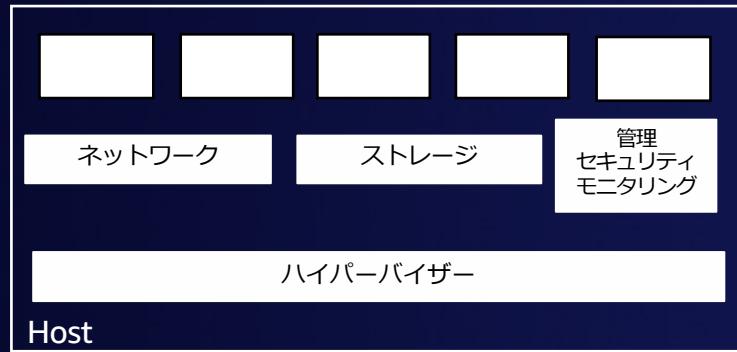


AWS Nitro System



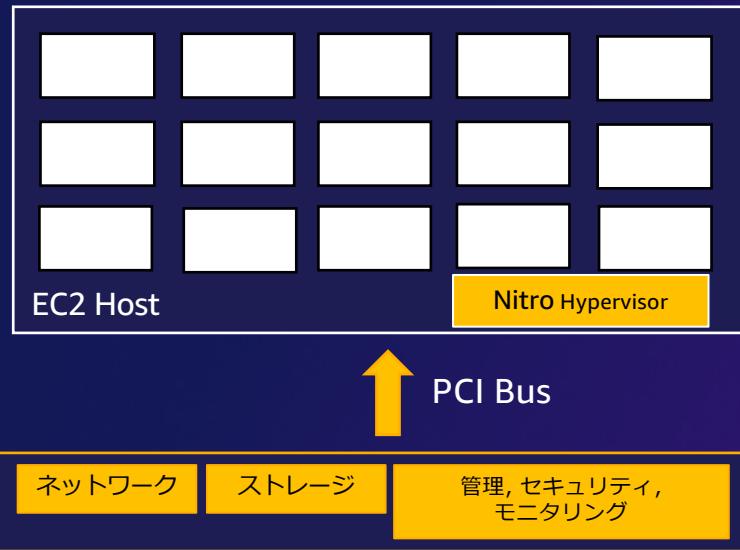
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 プラットフォームの歴史



従来の仮想化

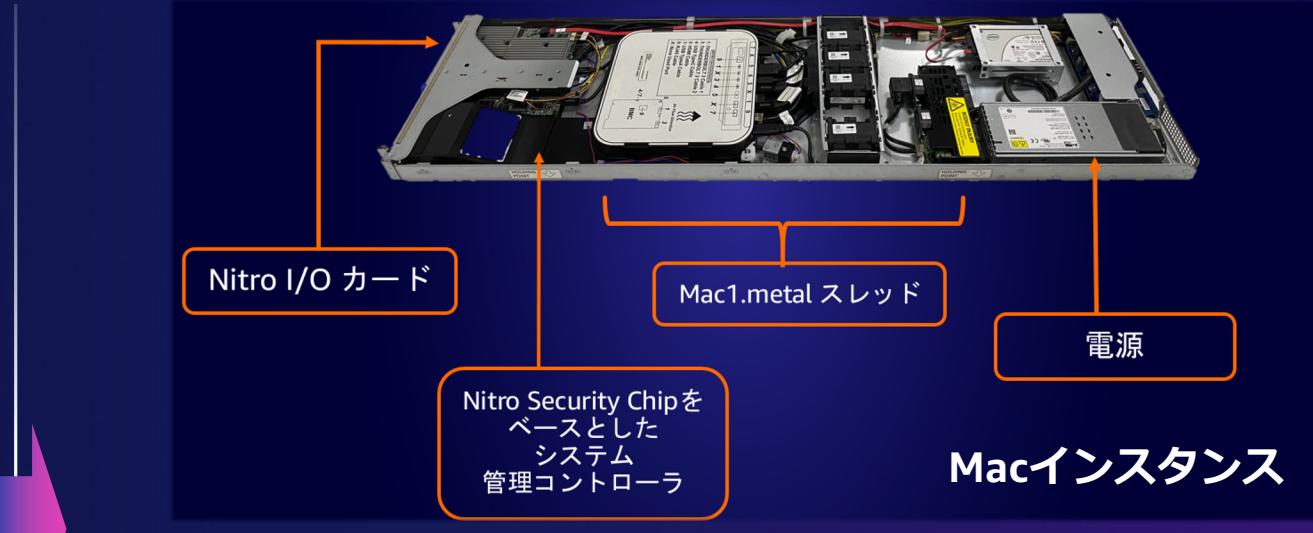
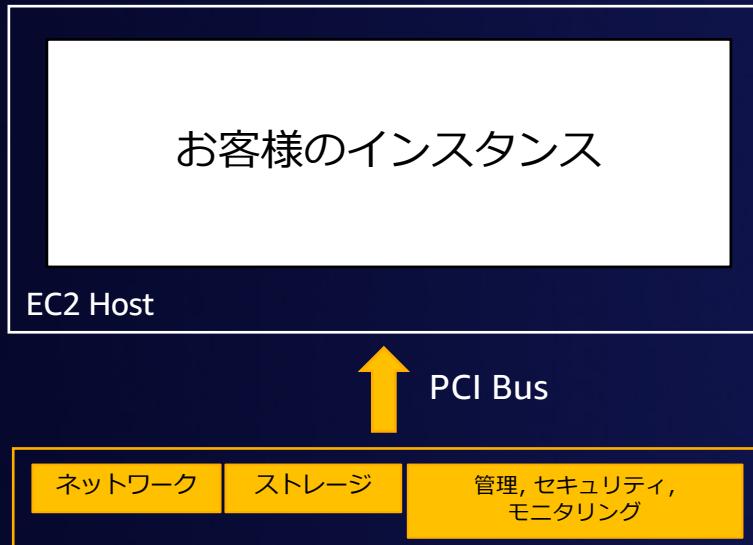
- ネットワーク、ストレージ、監視機能を備えた大規模な Trusted Computing Base (TCB) はすべて、お客様の VM を実行するのと同じハードウェア上で実行
- VM のパフォーマンス低下と利用可能なリソースの減少



EC2 仮想マシン

- Trusted computing base が大幅に削減
- ネットワーキング、ストレージ、モニタリングはすべて個別に分離された安全なハードウェア上で実行
- Amazon の従業員も含め、すべての管理アクセスを禁止
- ホスト上はハイパーバイザーのみ

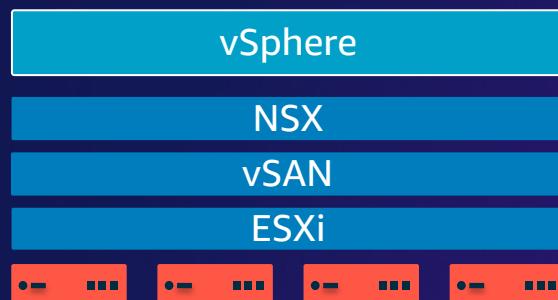
Nitro System 導入によるイノベーション



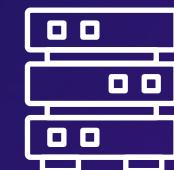
EC2 ベアメタルインスタンス

- ハイパーバイザーを廃止した Nitro のすべてのメリットを提供
- お客様は、物理ハードウェアリソースを使用するアプリケーションを AWS インフラストラクチャに直接デプロイ可能

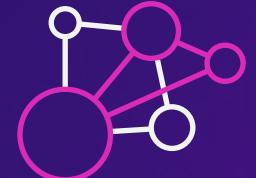
VMware Cloud on AWS
Software Defined Data Center



AWS OUTPOSTS



拡張
ネットワーキング



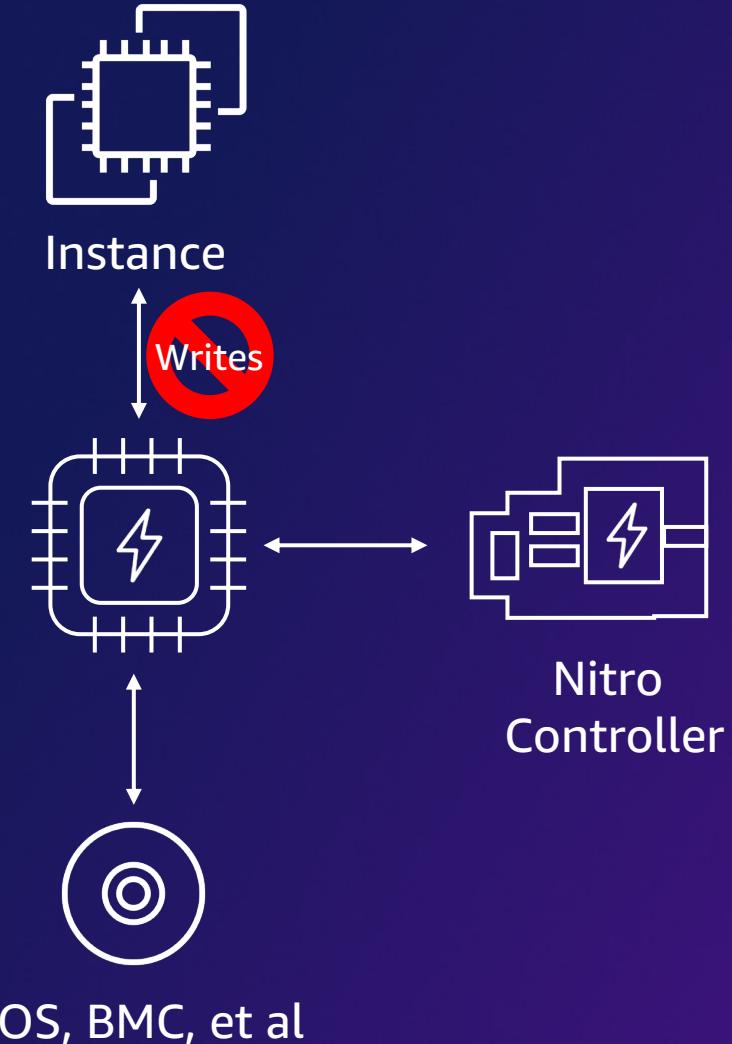
Nitro System 導入によるセキュリティの向上

Nitro Security Chips :
ハードウェアでの Root of Trust (信頼性の基点)

Nitro Security Chips にて、
全ての不揮発性ストレージへのI/Oをトラップしてコントロール

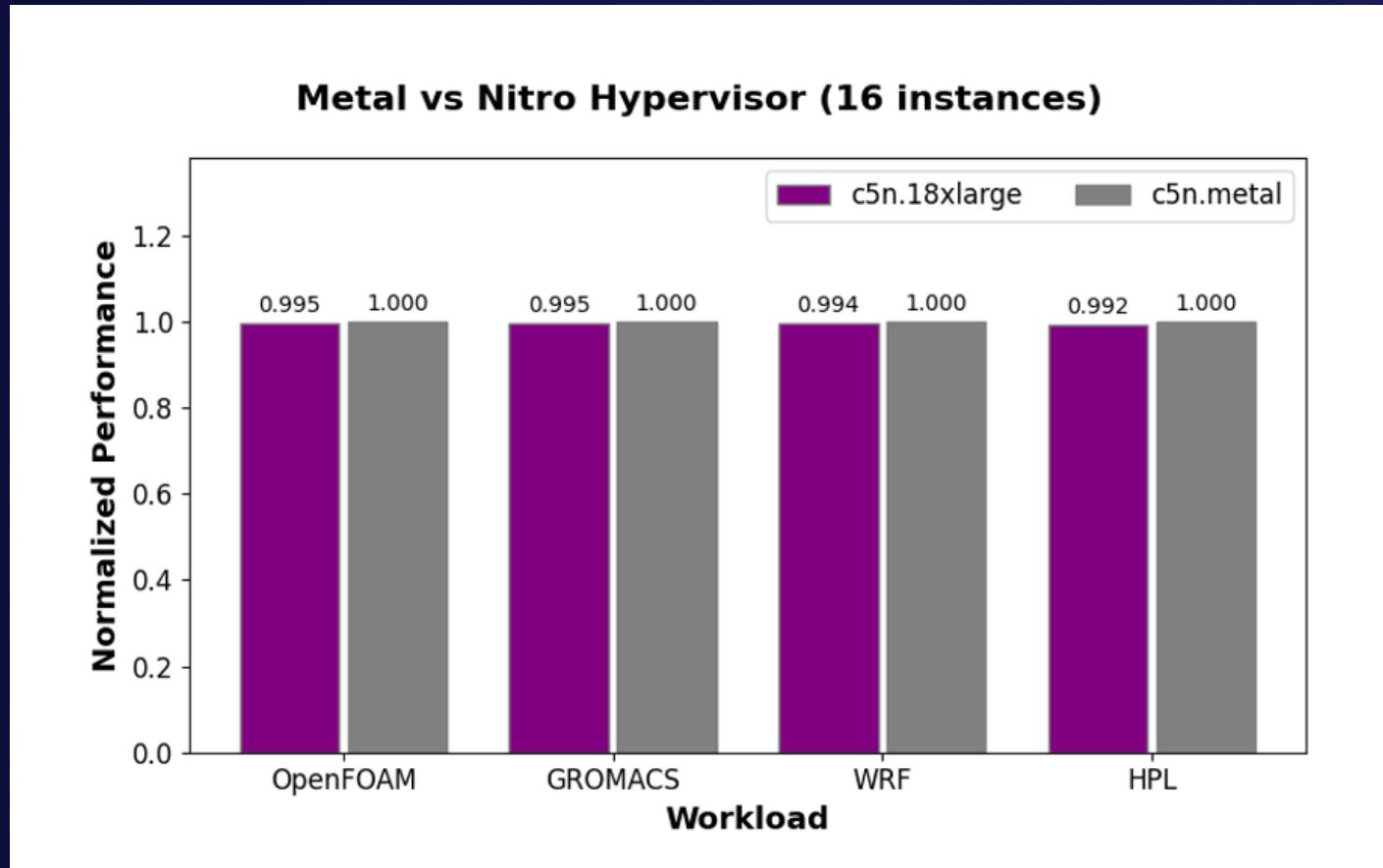
不揮発性ストレージの全ての書き込みを
ハードウェアでブロック

Nitro Controller からコントロール可能
システムブート時に
ストレージ上の情報が書き換わってないことを担保



Nitro System 導入によるパフォーマンスの向上

AWS独自の仮想化基盤 " Nitro System" により
EC2の仮想サーバはベアメタルとほぼ同等のパフォーマンスを実現



パフォーマンス

巨大かつ複雑な ハイパフォーマンスワークロードを AWS で実行

AWS infrastructure

事実上無制限のスケーラビリティと
幅広いクラウドベースの
サービスポートフォリオにアクセス

Elastic Fabric Adapter

最大 400 Gbps の
ネットワーク速度にアクセス

オンプレミスインフラストラクチャの限界を超えたイノベーション

AWS ParallelCluster

HPC 環境を効率的に構成・管理する

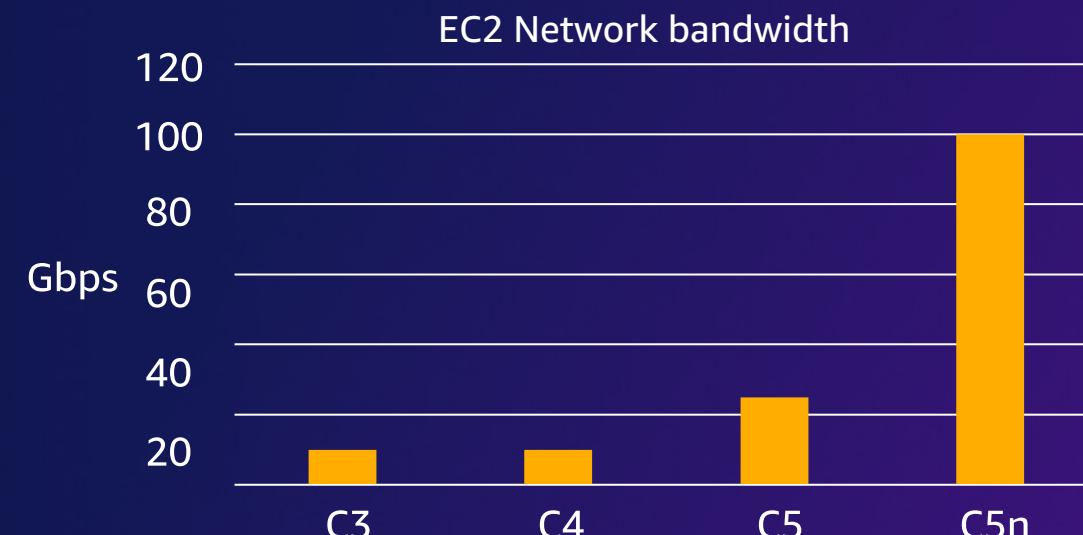
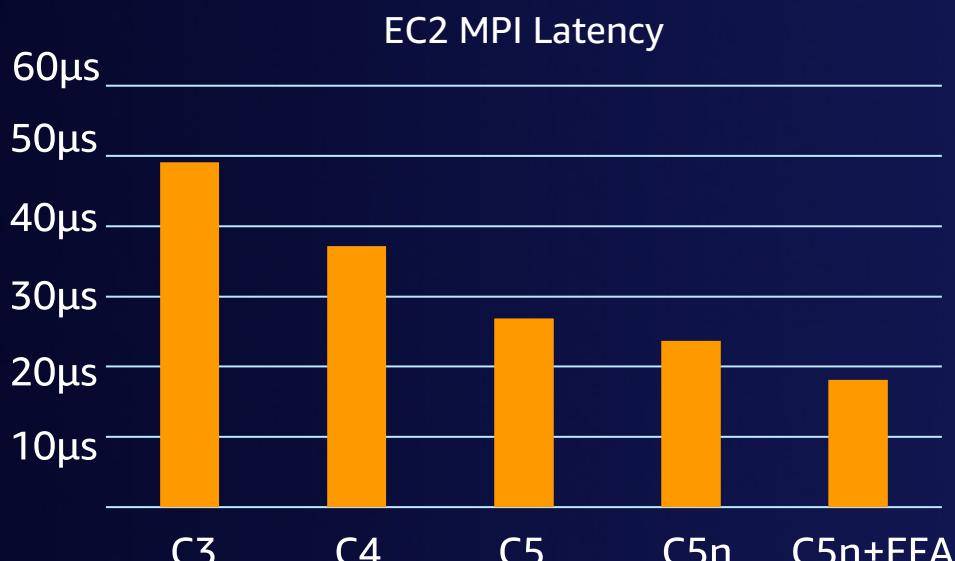
Amazon EC2

500 種類近くのインスタンスタイプから
選択して、価格パフォーマンスを最適化



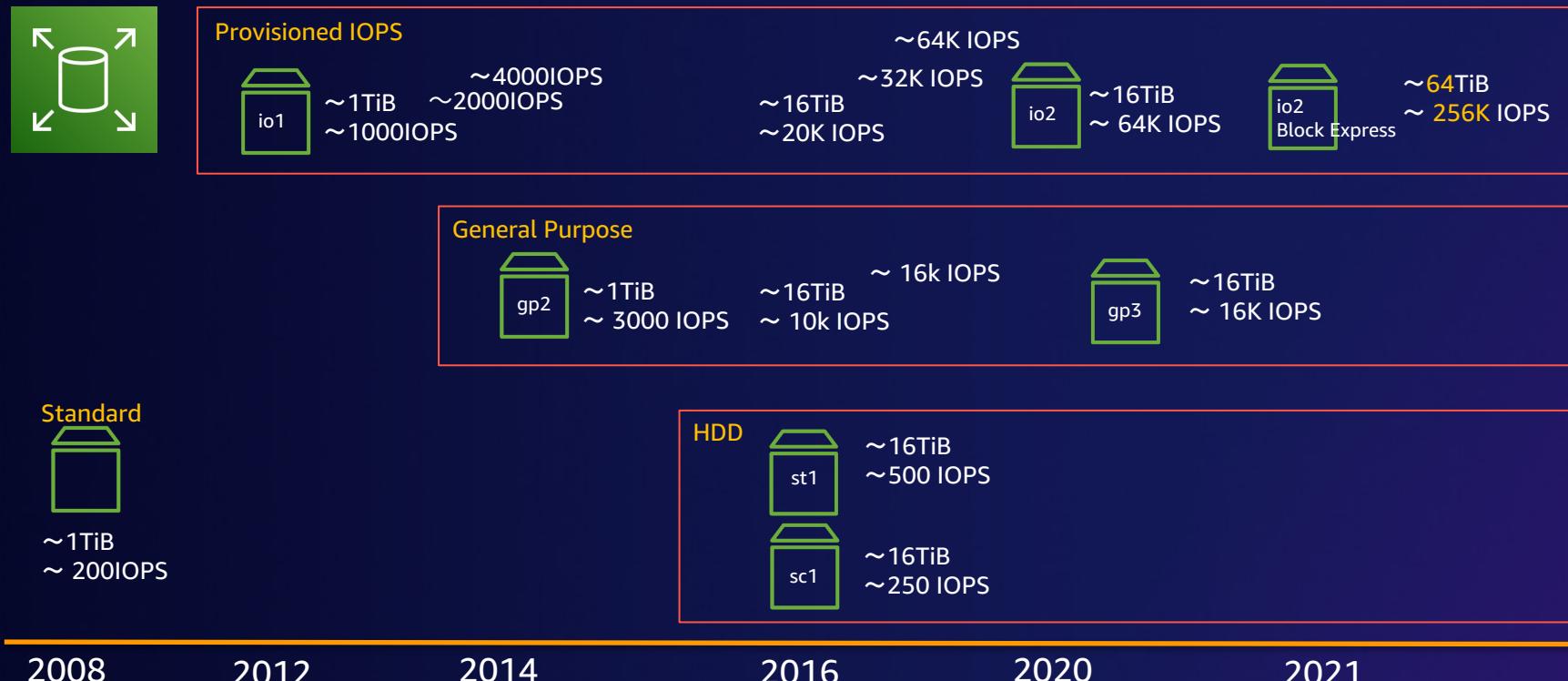
ネットワーク

- 拡張ネットワーキング (ENA) :
SR-IOV に対応し、仮想化オーバーヘッドを低減することで低レイテンシでの通信が可能
- Cluster Placement Group :
インスタンスの配置を最適化することで広帯域/低レイテンシ/フルバイセクション通信を実現
- Elastic Fabric Adapter (EFA) :
HPC向けに、MPI (Message Passing Interface) や NCCL (NVIDIA Collective Communications Library) などの libfabric 対応のアプリケーションでの通信をより低レイテンシ化

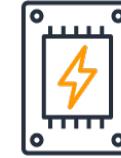


ストレージ

Elastic Block Store (EBS) : EFAと同じプロトコルを利用した **io2 Block Express** を2021年にリリース
インスタンスストレージ: AWSカスタム設計の **AWS Nitro SSD** を2021年末にリリース



NEW

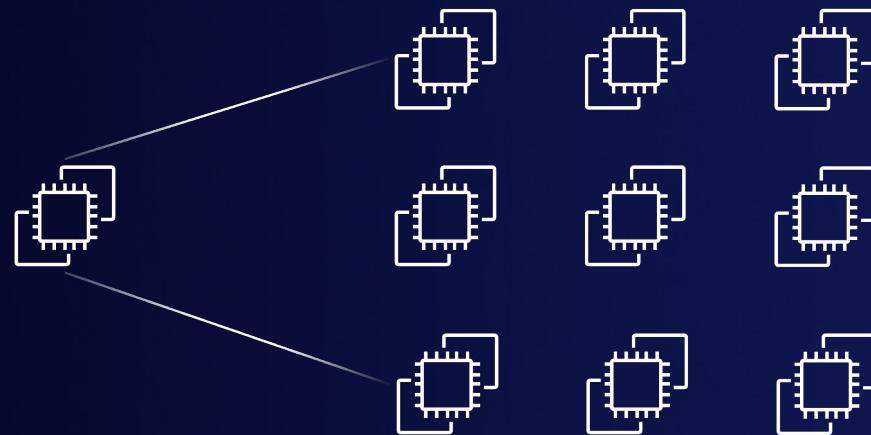


AWS Nitro SSD

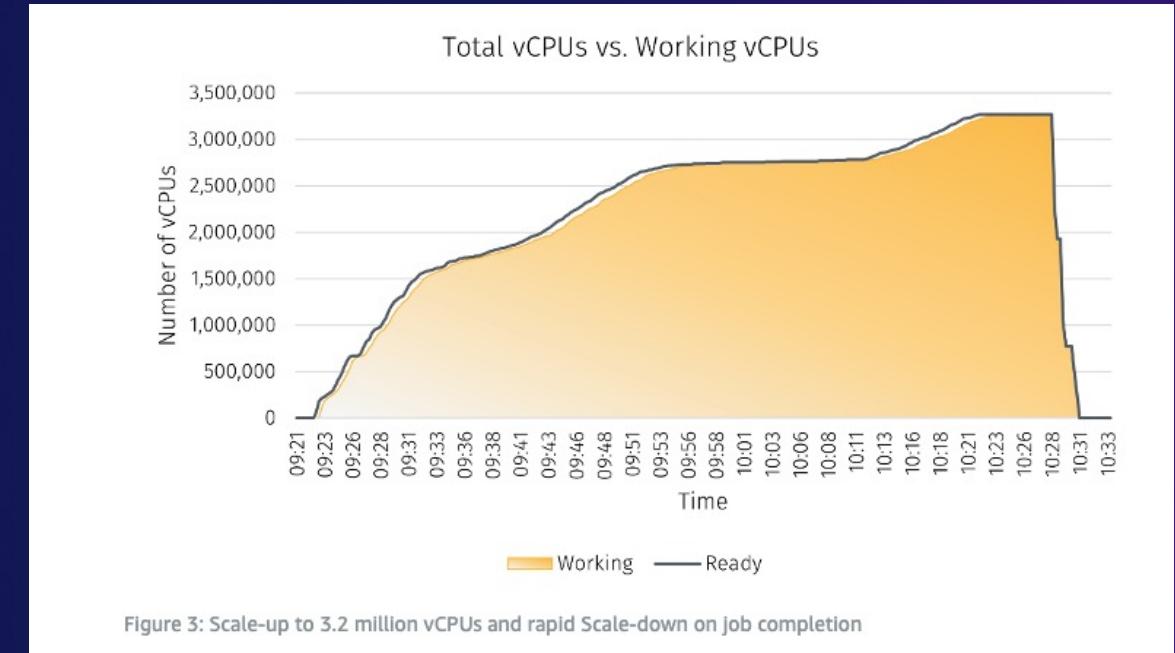
AWS がカスタム設計した
高性能、低レイテンシー SSD
商用 SSD と比較して
I/O レイテンシーを60%低減

スケーラビリティ

AWS のリソースを活用してハイパフォーマンスなワークロードを実現する
スケーラビリティをご提供
単一ワークロードでは最大300万 vCPU (複数リージョン活用)の事例も



必要に応じて必要な台数で処理の実行が可能



HPC on AWS でワークロードのニーズに合わせて インフラストラクチャを 動的にチューニング & スケーリング

最新テクノロジーへの
アクセス

パフォーマンスを最大化して
市場投入までの時間を短縮

運用効率の向上

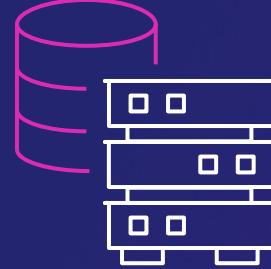


ハイブリッド

企業は急速にクラウドに移行している

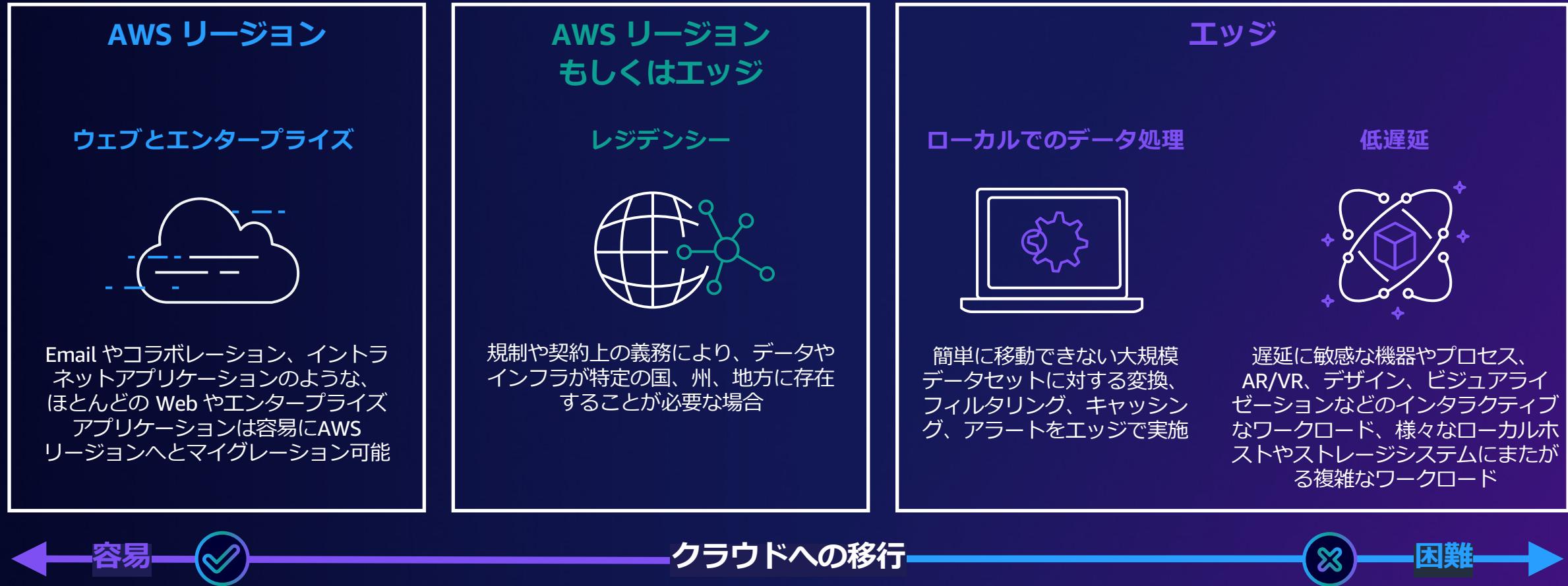


ほとんどのワークLOADは
クラウドに簡単に移行できるため、
組織はデジタルトランス
フォーメーションを加速できる

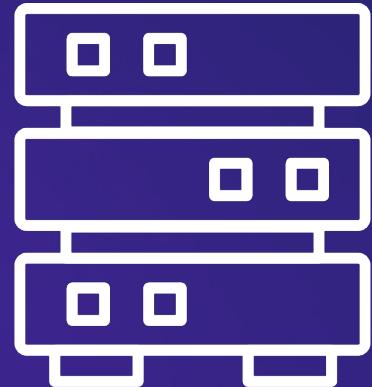


しかし、ワークLOADによっては
低レイテンシーまたはローカルな
データ処理を必要としたり、
データの常駐要件があるため、
オンプレミスやエッジノードに
残す必要があるものもある

AWSにおけるハイブリッド環境



AWSはハイブリッドコンピューティングと エッジコンピューティングを改革



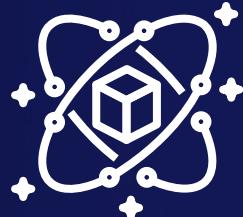
クラウド環境の拡張

AWS LOCAL ZONES



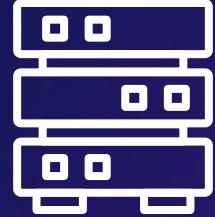
コンピュート、ストレージ、
ネットワーク、
その他一部サービスを
エンドユーザの近くに延伸

AWS WAVELENGTH



AWSインフラ・サービス・
APIをモバイルキャリアの
5Gネットワークに延伸

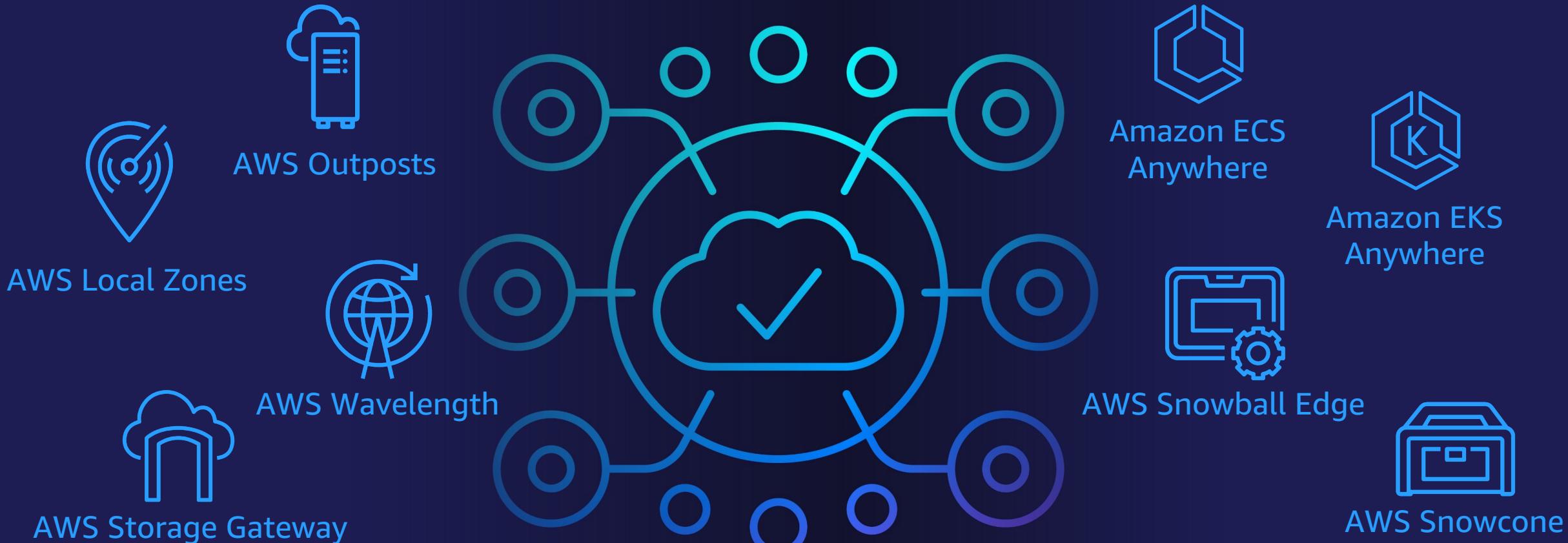
AWS OUTPOSTS



Nitro System で構成された
AWSデータセンタと同じ
インフラストラクチャを
オンプレミスに延伸

Hybrid Edge サービスの詳細 : <https://www.youtube.com/watch?v=nkIFJvXcJVA&t=11s>

AWSはお客様が必要とする場所ならどこでも同じインフラストラクチャ・サービス・ツールを提供



費用の最適化

自己紹介

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スポットインスタンススペシャリスト

普段の業務

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



費用の最適化 アジェンダ

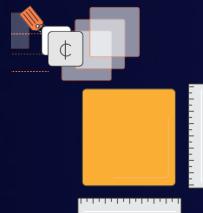
- Amazon EC2におけるコスト最適化の全体像
- Amazon EC2の4つの購入オプション
- EC2スポットインスタンスの活用

Amazon EC2のコスト最適化 全体概要

見積当初のインスタンス
サイズが大きすぎた



- 1 適切なインスタンス
の再選定



夜間帯はほとんど使われず
アイドル時間が長かった



- 2 未使用リソース停止
による規模の縮小



見積当初のディスクサイズが
大きすぎた



- 3 需要に応じた
スケジューリング



- 4 適切なストレージ
の再選定



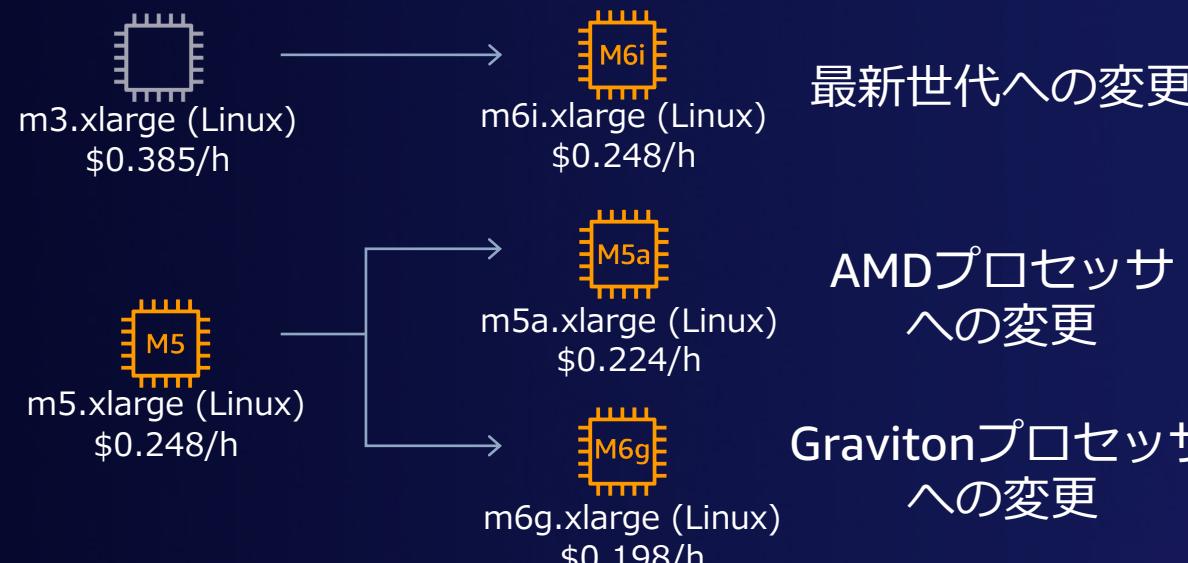
適切なインスタンス選定と無駄の検出

インスタンス再選定

- ✓ サイジング/ファミリ変更



- ✓ 最新世代化/CPU変更



[AWS-53]
Amazon EC2 のコストパフォーマンスを
最大 40 % 向上させる AWS Graviton に
よるコスト最適化入門
で詳しく解説いたします

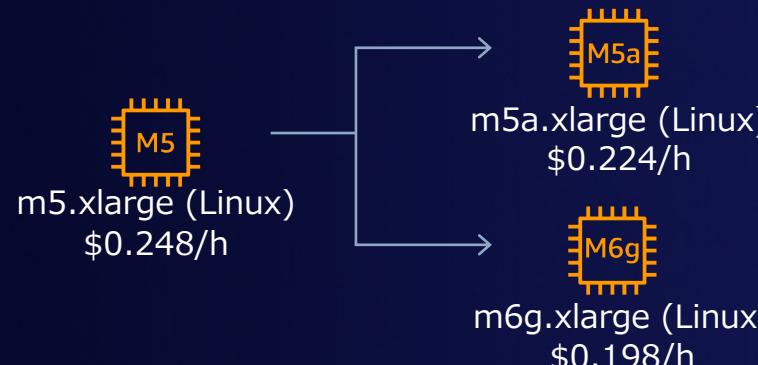
適切なインスタンス選定と無駄の検出

インスタンス再選定

- ✓ サイジング/ファミリ変更



- ✓ 最新世代化/CPU変更



購入オプション変更

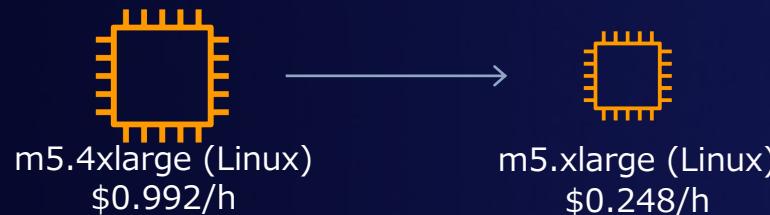
- ✓ リザーブドインスタンス (RI)・Savings Plans, スポットインスタンスに変更



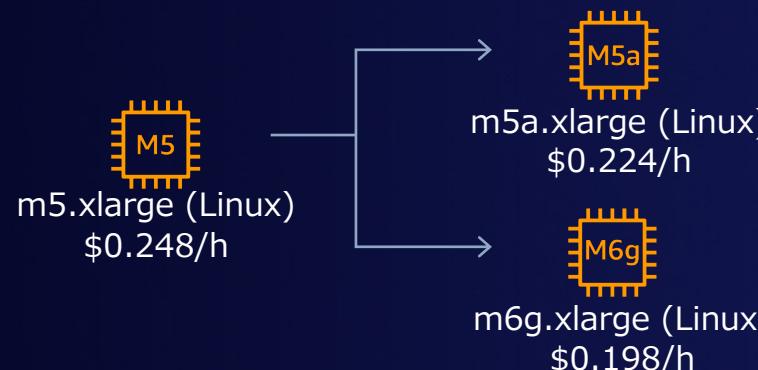
適切なインスタンス選定と無駄の検出

インスタンス再選定

- ✓ サイジング/ファミリ変更



- ✓ 最新世代化/CPU変更



購入オプション変更

- ✓ リザーブドインスタンス (RI)・Savings Plans, スポットインスタンスに変更

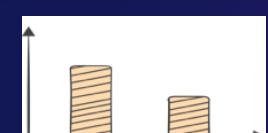
オンデマンド



RI/
Savings
Plans



スポット



検出

- ✓ AWS Cost Explorer の提供する「サイズの適正化に関する推奨事項」を活用

AWS Cost Explorer > サイズの適正化に関する推奨事項

サイズの適正化に関する推奨事項

推奨事項パラメータ

推奨事項の表示

- 同じインスタンスファミリー内
- インスタンスファミリー全体

タイプの検索中

- アイドル状態のインスタンス
- 低使用率のインスタンス

詳細オプション

- Savings Plans とリザーブドインスタンスを含める

推奨事項

最適化の機会	推定期別削減額	推定期別削減 (%)
1	\$21.59	51.06%

結果

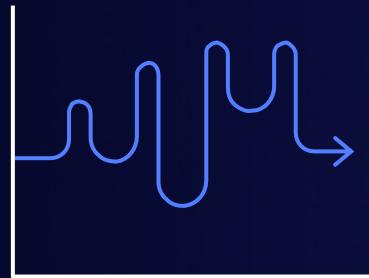
インスタンス ID	推定期別削減額	結果	理由の検索	インスタンスタイ
i-06da0321c83ec00b2	\$21.59/月	低使用率のインスタンス	CPUOverprovisioned 件、残り 4 件	c5.large



Amazon EC2の4つの購入オプション

オンデマンドインスタンス

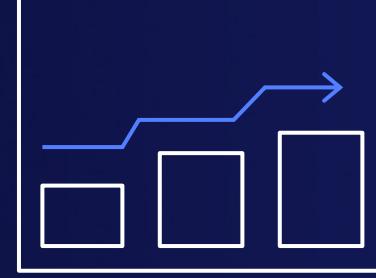
長期コミットなし、
使用分への支払い
(秒単位/時間単位)。
Amazon EC2の定価



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 (RI)

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



一定の負荷の見通しが
あり、長期コミット
できるワークロード

Savings Plans

RIと同等のディスカウン
トに加え、**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
可能なとき。
コンピュート能力を
柔軟に変更できる

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%の
値引き。中断あり



中断に強く、様々な
インスタンスタイプ
を活用できるワー
クロード

Savings Plansとは

コンピュートサービス(Amazon EC2など)の**定常的な利用**を
1年間または3年間**利用コミット契約**することで、
最大72%の割引を受けられる**割引プラン**

Savings Plansの2種類の料金プラン



Compute Savings Plans

優れた柔軟性。最大66%の費用削減
(コンバータブルRIと同等の割引率)



EC2 Instance Savings Plans

最安価格。リージョン内の個々の
インスタンスファミリーに対して
最大72%の費用削減
(スタンダードRIと同等の割引率)

- 柔軟に
変更可能**
- ✓ インスタンスファミリー - どれでも
 - ✓ リージョン - どこでも
 - ✓ OS - WindowsでもLinuxでも
 - ✓ テナント属性 - default(ハードウェア共有)
でもdedicated(ハードウェア専有)でも
 - ✓ コンピュート環境 - Amazon EC2でも
AWS FargateでもAWS Lambdaでも

- 柔軟に
変更可能**
- ✓ サイズ - m5.xlargeでも
m5.4xlargeでも
 - ✓ OS - Windowsを実行するm5.xlargeでも
Linuxを実行するm5.2xlargeでも
 - ✓ テナント属性 - default(ハードウェア共有)
でもdedicated(ハードウェア専有)でも

(参考) Savings Plansとリザーブドインスタンス(RI)との関係

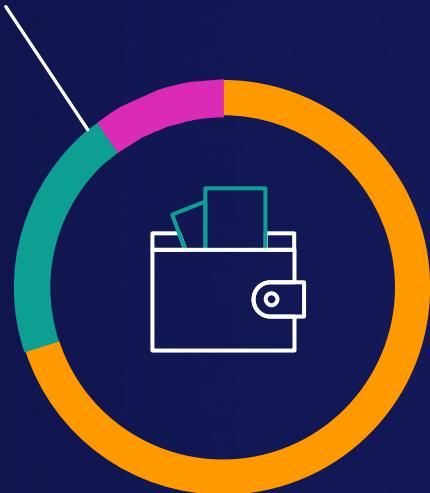
- RIとの違い
 - ゾーンRI(特定のアベイラビリティゾーンを指定)の「キャパシティ予約」機能はない。必要であれば別途「オンデマンドキャパシティー予約」を利用する
- 既存RIの契約がある場合
 - まずRIの契約が優先適用される
 - 次にRIのカバーしないオンデマンドインスタンスご利用分がSavings Plansの適用対象となる
- その他
 - RIと同様に、複数のSavings Plansを契約すること(買い増し)も可能
 - 現在対応しているサービスはAmazon EC2, AWS Fargate、AWS Lambda
 - Amazon RDSなど他サービスは従来通り、RIを活用してコスト最適化する必要がある

スポットインスタンスとは？

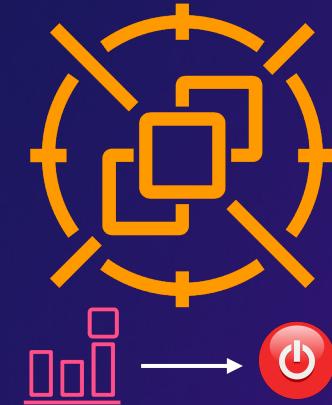
70–90% off



オンデマンドインスタンス
やSavings Plansと共に
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



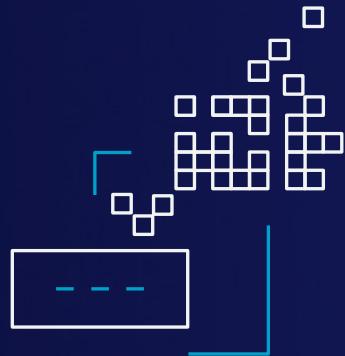
EC2サービスの空き
キャパシティに基づいて
提供される。起動しない
可能性や中断の可能性

サービスアップデート Compute 編：スポットインスタンスとAuto Scaling, フリート機能のアップデート
<https://www.youtube.com/watch?v=OMOL3xtAJ80>

スポットインスタンスと親和性の高いワークロード



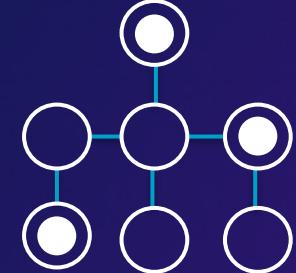
コンテナ



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
CloudFormation



AWS Thinkbox
Deadline



Jenkins

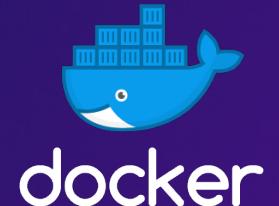


databricks™

cloudera

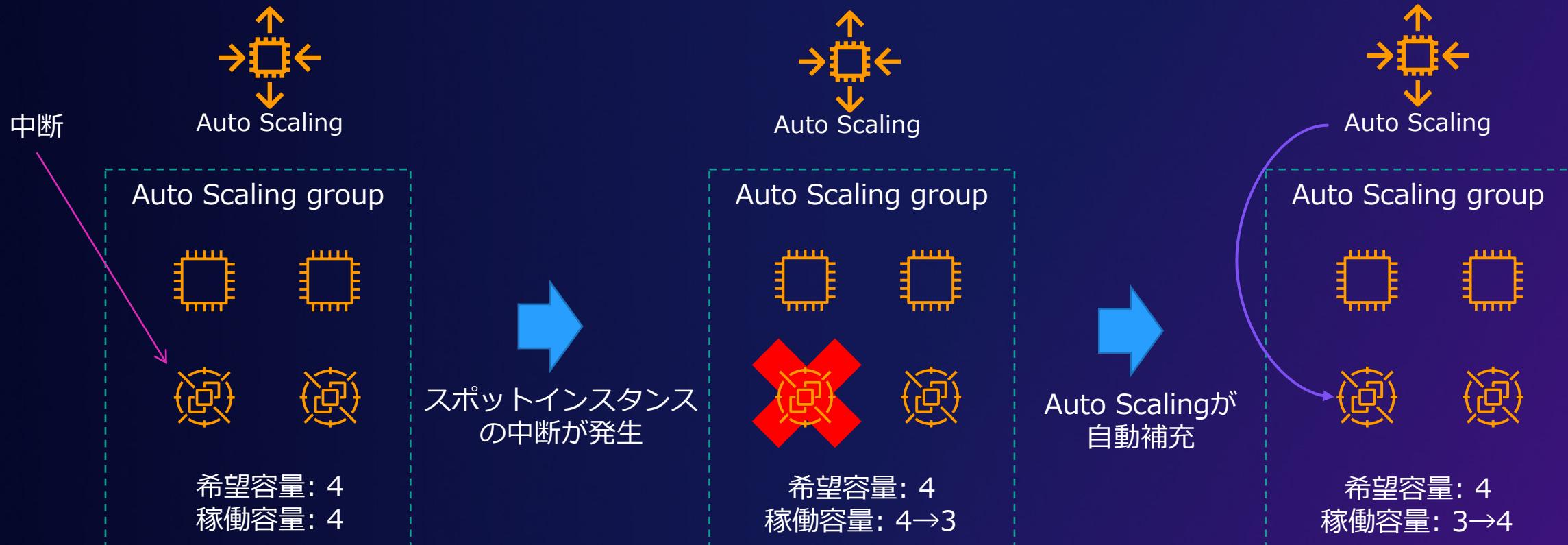


kubernetes



EC2 Auto Scalingグループとスポットインスタンス

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



スポットインスタンス活用のために知っておきたいASGの機能

- スポットインスタンスを起動できる可能性を高める
 - 複数のインスタンスタイプの指定
(ミックスインスタンスグループ)
 - 属性ベースのインスタンスタイプ自動選定
(Attribute-based Instance Type Selection)
- 中断しにくいスポットインスタンスを起動する
 - capacity-optimized配分戦略
- 中断する前に事前にスポットインスタンスを起動する
 - キャパシティリバランシング (Capacity Rebalancing)

スポットインスタンスの
可用性を向上

スポットインスタンスを
管理するASGでは
この3つの機能を有効化
しておくことを推奨

サービスアップデート Compute 編：スポットインスタンスとAuto Scaling, フリート機能のアップデート
<https://www.youtube.com/watch?v=OMOL3xtAJ80>

本日のまとめ

- Amazon EC2 の特徴・魅力を伝える4つの柱
 - 仮想化のオーバーヘッドを極小化したNitro System
 - 高性能科学技術計算をも支えるパフォーマンス
 - AWSインフラストラクチャーにとどまらない運用管理を実現するハイブリッドコンピューティングサービス
 - 豊富なコスト最適化ツールと購入オプションの選択肢、またコスト効率を最大化するスポットインスタンスとAuto Scalingサービスの親和性

Amazon EC2はこれからもユーザーのフィードバックをお聞きして進化を続けます

関連セッション一覧

- [AWS-53] Amazon EC2 のコストパフォーマンスを最大 40 % 向上させる AWS Graviton によるコスト最適化入門
- [AWS-05] リホストから始めるAWSへのサーバー移行
- [AWS-06] 企業内 Windows Server アプリケーション移行手法とツール

参考資料

- AWS Nitro Systemに於けるベアメタルインスタンスの性能
<https://aws.amazon.com/jp/blogs/news/bare-metal-performance-with-the-aws-nitro-system/>
- Running a 3.2M vCPU HPC Workload on AWS with YellowDog
<https://aws.amazon.com/jp/blogs/hpc/running-a-3-2m-vcpu-hpc-workload-on-aws-with-yellowdog/>
- サービスアップデート Hybrid Edge編
<https://www.youtube.com/watch?v=nkIFJvXcJVA&t=11s>
- サービスアップデート Compute 編：スポットインスタンスとAuto Scaling, フリート機能のアップデート
<https://www.youtube.com/watch?v=OMOL3xtAJ80&t=0s>

Thank you!



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.