



# コンテナ視点から見る Amazon EC2 のコスト最適化 とAWSのシリコンイノベーション

2023-05-25

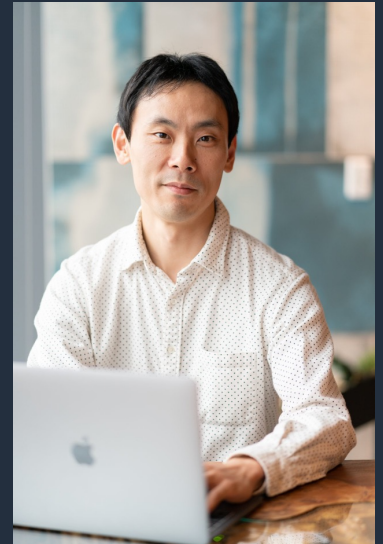
Haruyoshi Takiguchi

Senior Specialist Solutions Architect, Flexible Compute  
Amazon Web Services

# 自己紹介

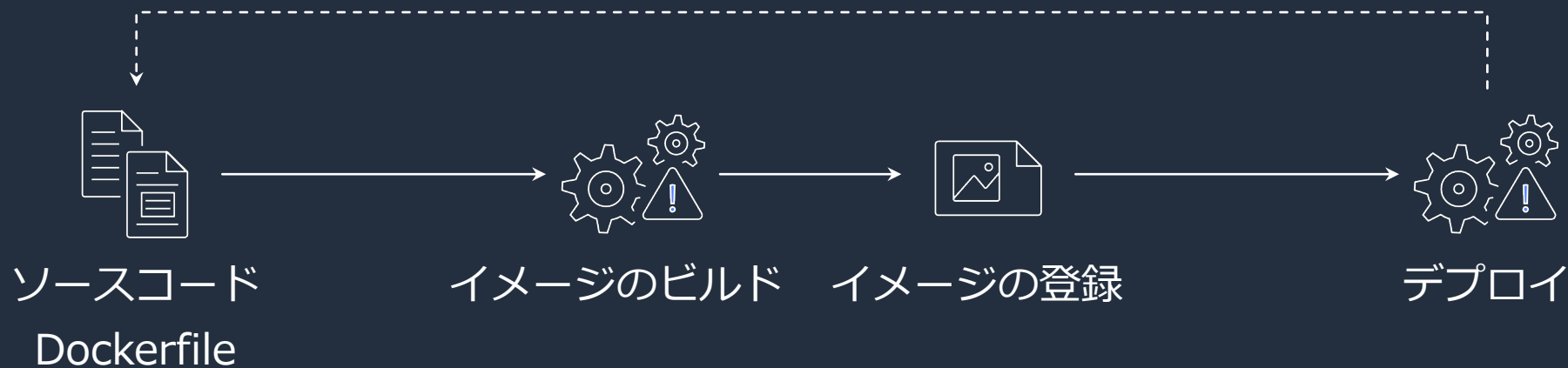
---

- 名前：滝口 開資 (はるよし)
- 所属：アマゾンウェブサービスジャパン合同会社 コンピュート事業本部  
シニアソリューションアーキテクト  
EC2 フレキシブルコンピューティングスペシャリスト
- 好きなAWSサービス：Amazon EC2 Auto Scaling, AWSサポート



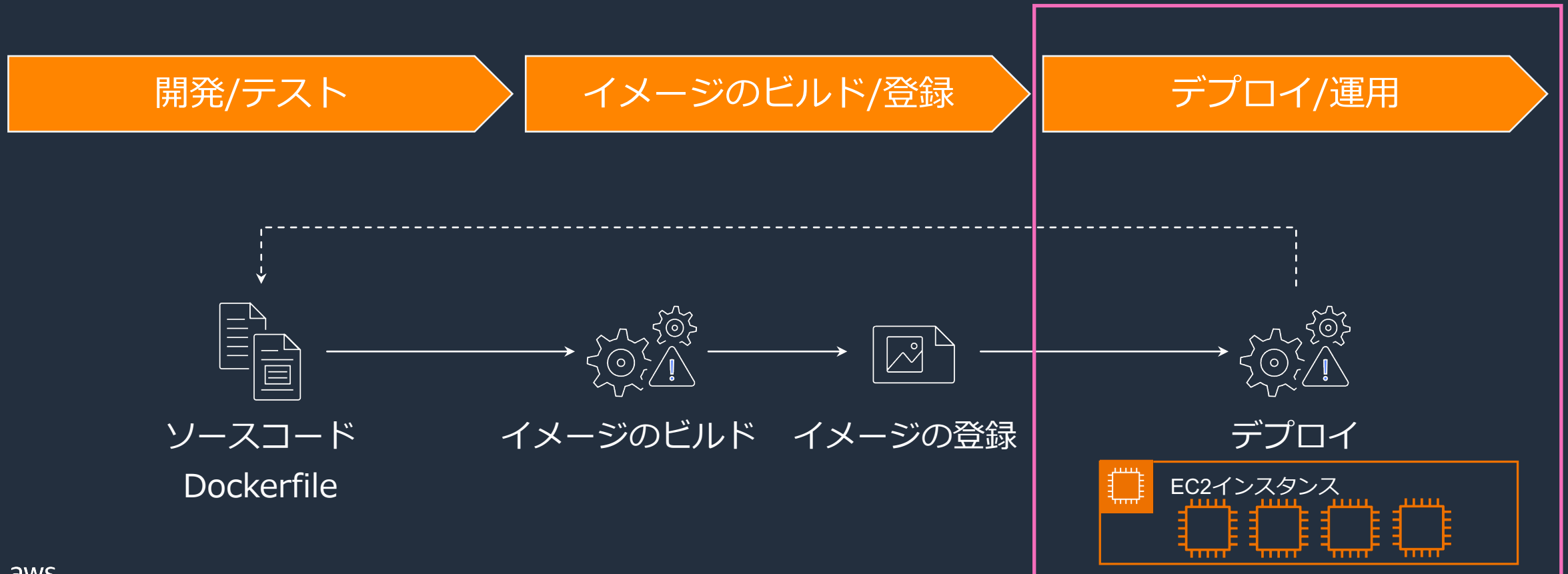
# 本日のセミナーで お話しすること

# コンテナにおけるソフトウェア開発ライフサイクル



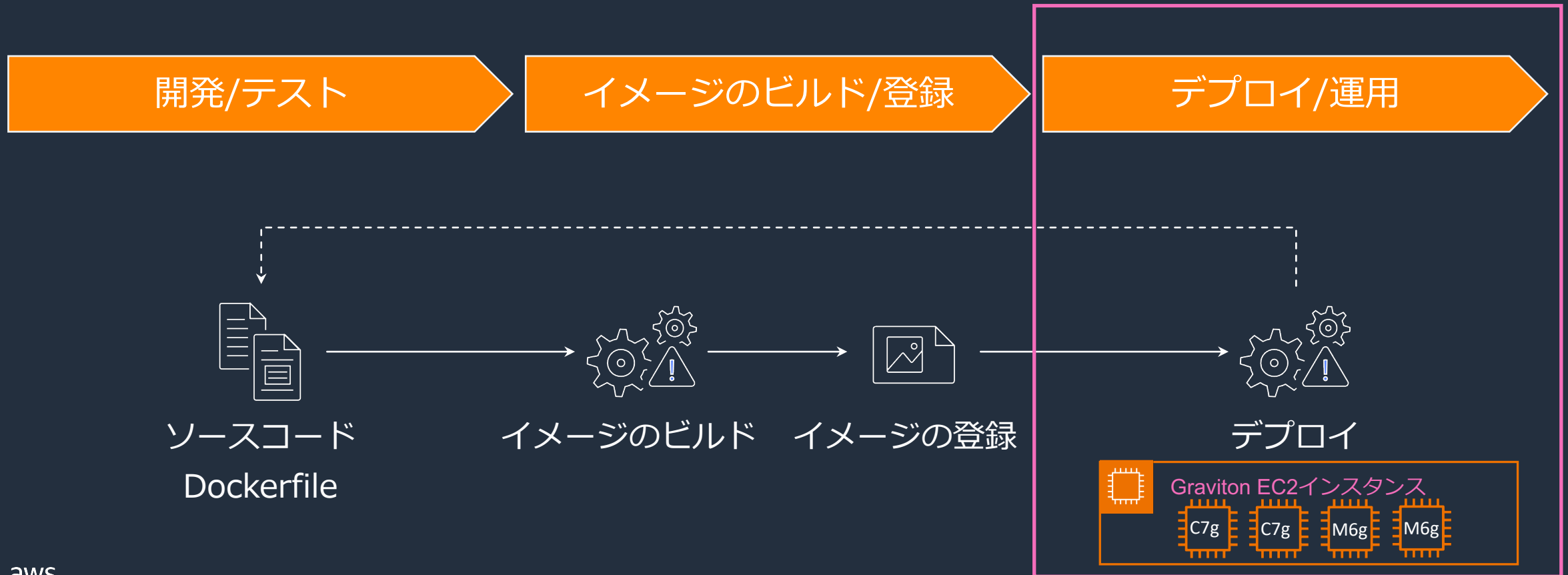
# コンテナにおけるソフトウェア開発ライフサイクル

コンテナアプリケーションの費用の大部分はAmazon EC2が占める  
(その他、Amazon EKSクラスター、Amazon ECRなどの料金)



# コンテナにおけるソフトウェア開発ライフサイクル

コンテナアプリケーションの費用の大部分はAmazon EC2が占める  
(その他、Amazon EKSクラスター、Amazon ECRなどの料金)

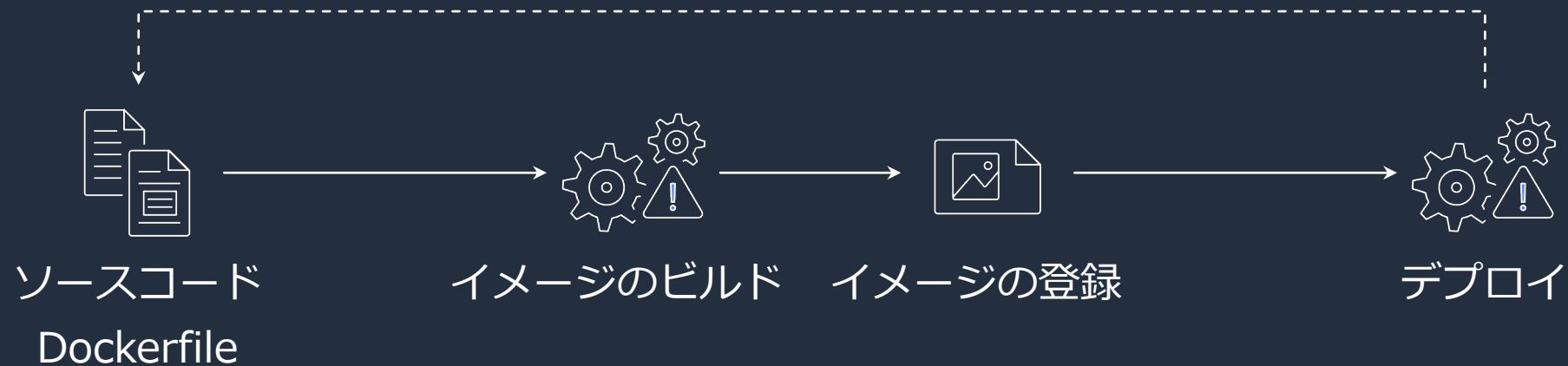


# コンテナにおけるAWS Gravitonの活用

アプリケーションソースコードへの考慮は？

コンテナイメージのビルドは？

Gravitonコンテナインスタンスの使用方法は？



# AWS Gravitonと シリコンイノベーション



# Amazon EC2 で選択できるプロセッサ

2006-



Intel® Xeon Scalable  
processors

**c6i.xlarge**

2018-



AMD EPYC  
processors

約10%コスト低減

**c6a.xlarge**

2018-



AWS Graviton  
processors

約40%コストパフォーマンス改善

**c7g.xlarge**

# Amazon EC2 で選択できるプロセッサ

2006-



Intel® Xeon Scalable  
processors

**c6i.xlarge**

x86アーキテクチャ

2018-



AMD EPYC  
processors

約10%コスト低減

**c6a.xlarge**

x86アーキテクチャ

2018-



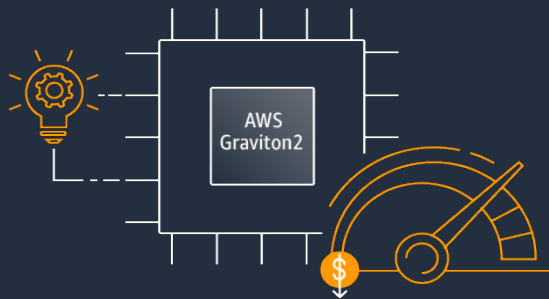
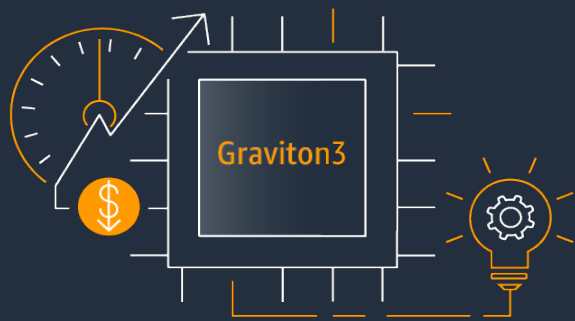
AWS Graviton  
processors

約40%コストパフォーマンス改善

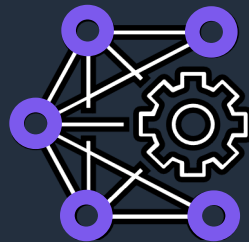
**c7g.xlarge**

Armアーキテクチャ

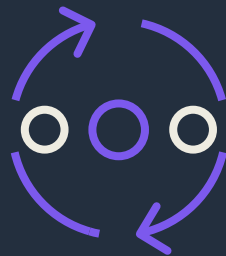
# AWS Gravitonプロセッサ



64 ビット Arm プロセッサコア搭載  
カスタム AWS シリコン



クラウドネイティブなワークロードに最適化



お客様に代わって迅速なイノベーション、ビルド、  
イテレートを実施

# AWS Graviton を選択する価値



幅広いワークロードで価格対性能比が最大 40% 向上



同等の x86 ベースのインスタンスよりも最大 20% 安価\*



同等の x86 ベースのインスタンスと比較して  
エネルギー効率が最大 60% 向上



# AWS Graviton の加速するイノベーション

AWS のワークロードに最適なコストパフォーマンスとエネルギー効率を実現

## AWS Graviton

2018



スケールアウトできるワークロードに大幅なコスト最適化をもたらした

## AWS Graviton2

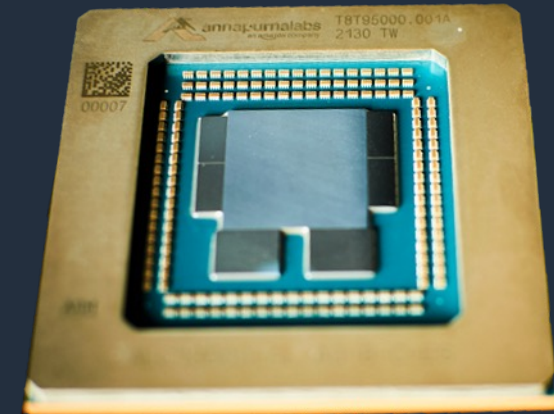
2019



同等のx86ベースのインスタンスと比べて価格パフォーマンスが最大 40% 向上し、幅広いワークロードに対応

## AWS Graviton3

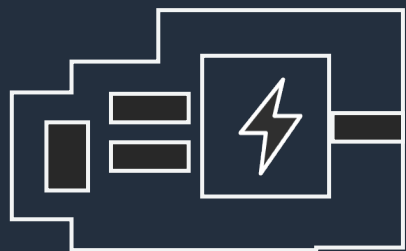
2021



計算量を必要とするワークロードに対し、Graviton2と比較して最大 25% のパフォーマンス向上

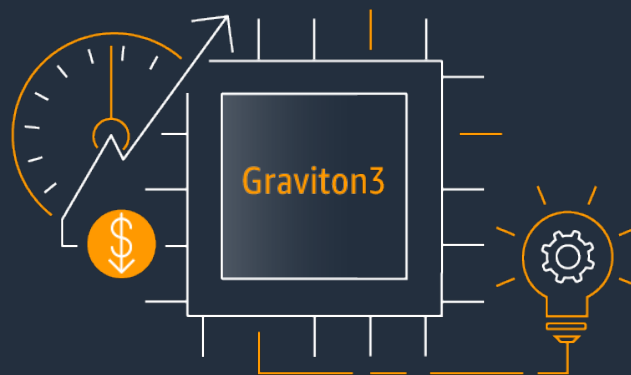
# AWSのシリコンイノベーション

- AWS では用途に合わせて様々な専用チップ開発を行っている



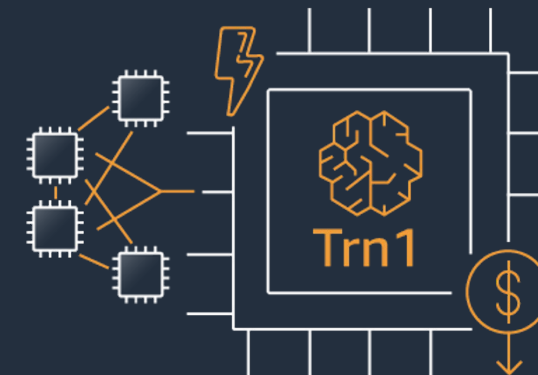
## AWS Nitro System

ハイパーバイザー, ネットワーク,  
ストレージ, SSD, セキュリティ



## AWS Graviton

パワフルかつ効率的な  
最新プロセッサ



## AWS Inferentia AWS Trainium

機械学習アクセラレーション

# なぜ自分たちでチップを作るのか？



## 最適化

AWS の仕様に合わせて  
ハードウェアを最適化  
高い電力効率



## スピード

製品の仕様化から導入  
までエンドツーエンド  
の開発プロセス



## 運用

信頼性・可用性  
動作監視・自己回復機能  
をチップレベルで実装



## イノベーション

より多くの価値を創造  
エンドツーエンドでの  
最適化

# AWS Graviton のパフォーマンス (1/2)

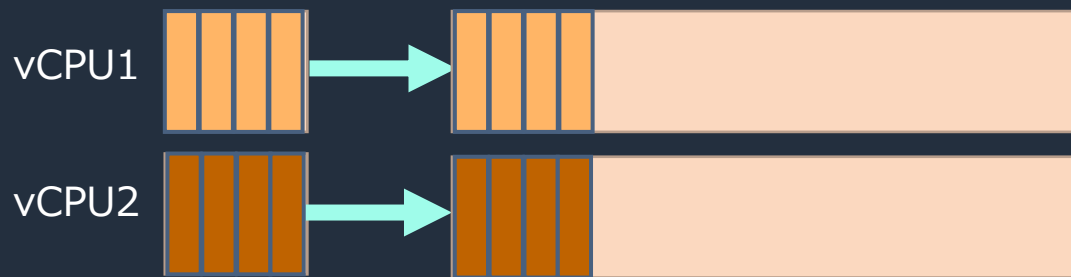
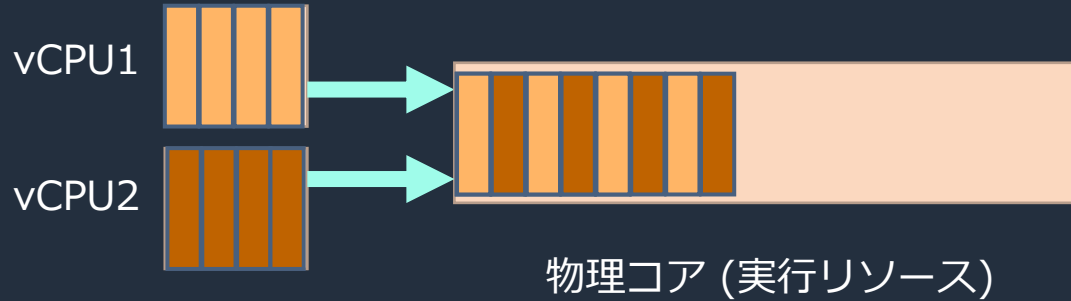
x86 系インスタンス: 2 vCPU = 1 物理コア

Arm 系インスタンス: 1 vCPU = 1 物理コア

例 : m5.16xlarge: 64 vCPU = 32 物理コア  
m6g.16xlarge: 64 vCPU = 64 物理コア

同インスタンスサイズでは **2 倍の物理コアが利用可能**  
→ **2 倍の物理コアでもコストは最大 20% 安価**

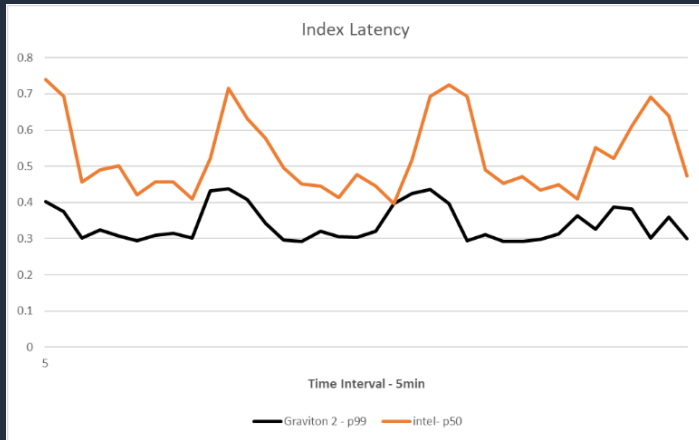
Graviton プロセッサ上の各 vCPU は、各々 1 つの物理コアにマッピングされる



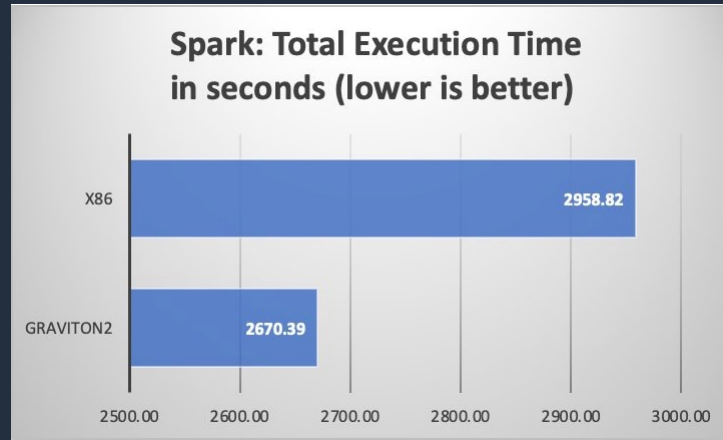
simultaneous multi-threading  
(SMT) なし



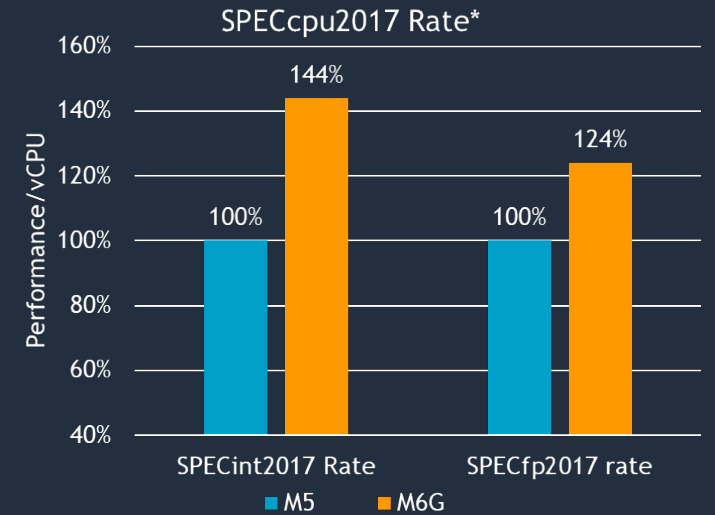
# AWS Graviton のパフォーマンス (2/2)



Amazon OpenSearch Service での比較



EMR Serverless での比較



\* All SPEC scores estimates, compiled with gcc v9 -O3 -march=native, run on largest single-socket size for each instance type tested.

SPECint, SPECfp での比較

Webサーバ、コンテナ基盤、ビッグデータ処理、HPC、機械学習等様々なワークロードで既存の x86 系インスタンスと比較して **最大 40 % のコストパフォーマンス向上**

<https://aws.amazon.com/jp/blogs/big-data/achieve-up-to-27-better-price-performance-for-spark-workloads-with-aws-graviton2-on-amazon-emr-serverless/>

<https://aws.amazon.com/jp/blogs/big-data/improved-performance-with-aws-graviton2-instances-on-amazon-opensearch-service/>

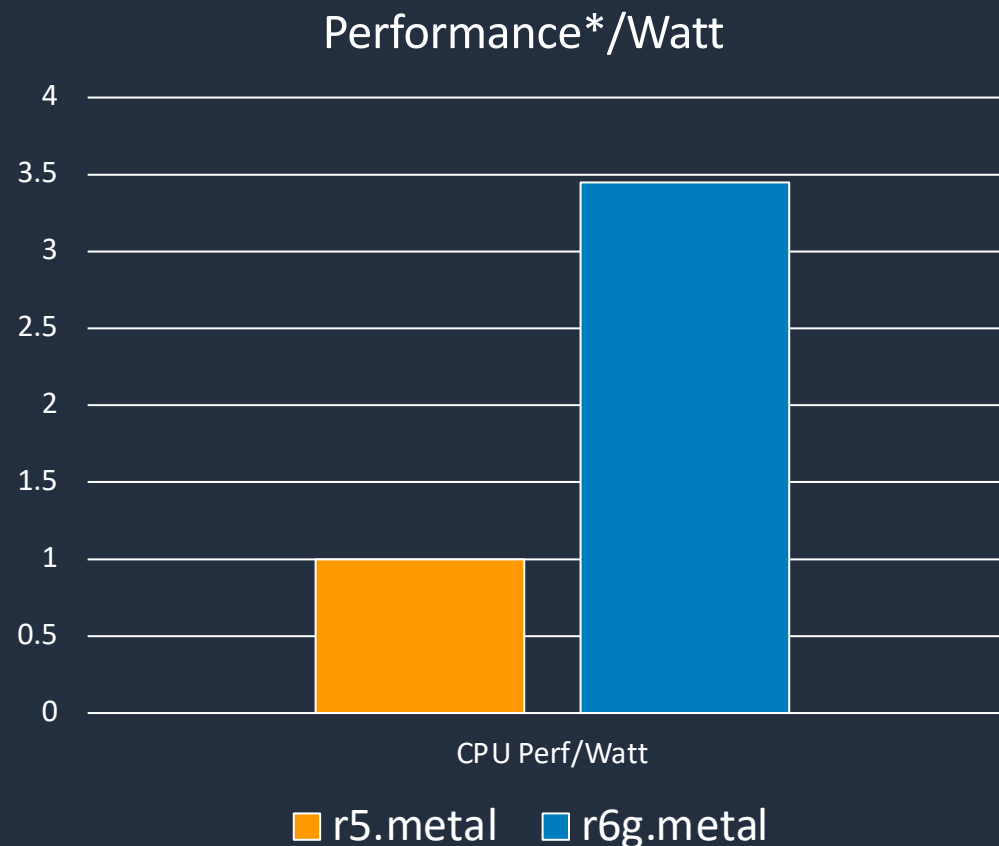


# サステナビリティ

- Graviton2 は同等の x86 系インスタンスと比較して **3.5倍** の高い消費電力効率
- Graviton3 は同等の x86 系インスタンスと比較して **60% 消費電力削減**

## 高い消費電力効率

- 安価なコスト
  - 高密度
  - 低いカーボンフットプリント
- より良いサービス提供に



ファミリー	Graviton	Graviton2	Graviton3	Graviton3E
一般用途向け General Purpose	A1	M6g, M6gd, T4g	M7g (New)	
コンピューティング最適化 Compute Optimized		C6g, C6gd, C6gn	C7g	C7gn (Preview)
メモリ最適化 Memory Optimized		R6g, R6gd, X2gd	R7g (New)	
高速コンピューティング Accelerated Computing		G5g (GPU 搭載)		
ストレージ最適化 Storage Optimized		Im4gn Is4gen		
HPC最適化 HPC Optimized				HPC7g (アナウンス)

橙字：東京・大阪対応

白字：東京対応・大阪未対応

紫字：東京・大阪未対応

# AWS Graviton 搭載 EC2 インスタンス

2023-05-25現在

ファミリー	Graviton	Graviton2	Graviton3	Graviton3E
一般用途向け General Purpose	A1	M6g, M6gd, T4g	M7g (New)	
コンピューティング最適化 Compute Optimized		C6g, C6gd, C6gn	C7g	C7gn (Preview)
メモリ最適化 Memory Optimized		R6g, R6gd, X2gd	R7g (New)	
高速コンピューティング Accelerated Computing		G5g (GPU 搭載)		
ストレージ最適化 Storage Optimized		Im4gn Is4gen		
HPC最適化 HPC Optimized				HPC7g (アナウンス)

橙字：東京・大阪対応

白字：東京対応・大阪未対応

紫字：東京・大阪未対応

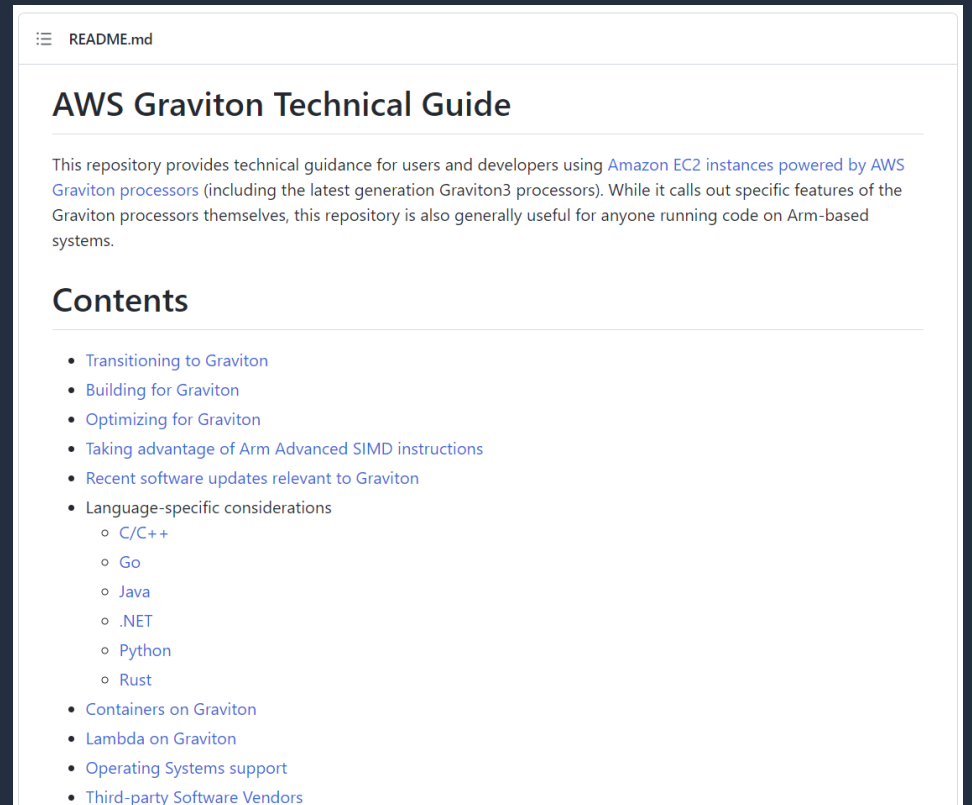


# AWS Graviton Technical Guide

## AWS Graviton 利用時に最初にご確認いただきたいドキュメント

<https://github.com/aws/aws-graviton-getting-started>

- プログラミング言語別考慮事項
  - 推奨コンパイルオプション
  - アーキテクチャ間の差異
- OS サポート情報
- 各種アプリケーションでの推奨バージョン、設定
- デバッグ・プロファイル
- etc …



The screenshot shows the README.md file for the AWS Graviton Technical Guide repository. The title is "AWS Graviton Technical Guide". The text describes the repository's purpose: providing technical guidance for users and developers using Amazon EC2 instances powered by AWS Graviton processors, including the latest generation Graviton3 processors. It also notes that the repository is useful for anyone running code on Arm-based systems. Below the text is a "Contents" section with a list of links to various topics:

- [Transitioning to Graviton](#)
- [Building for Graviton](#)
- [Optimizing for Graviton](#)
- [Taking advantage of Arm Advanced SIMD instructions](#)
- [Recent software updates relevant to Graviton](#)
- [Language-specific considerations](#)
  - [C/C++](#)
  - [Go](#)
  - [Java](#)
  - [.NET](#)
  - [Python](#)
  - [Rust](#)
- [Containers on Graviton](#)
- [Lambda on Graviton](#)
- [Operating Systems support](#)
- [Third-party Software Vendors](#)

# 補足 - スポットインスタンスとGravitonインスタンスの関係

EC2スポットインスタンス、Gravitonともにコスト効率化に有効な手段。  
どのように違うのか？

- EC2スポットインスタンス
  - Amazon EC2の購入オプション、最大で9割引
  - EC2サービスの空きキャパシティから起動するため、起動できない場合、また中断が発生することがある
  - 中断に強い仕組みを導入する必要がある
- Graviton搭載インスタンス
  - インスタンスファミリーのひとつ
  - オンデマンド価格が同等のx86ベースインスタンスより最大20%安い
  - 「中断に強い仕組み」の導入は不要
  - ArmベースのWindows Server製品が存在しないため、Windows環境では選択できない

# 補足 - スポットインスタンスとGravitonインスタンスの関係

EC2スポットインスタンス、Gravitonともにコスト効率化に有効な手段。  
どのように違うのか？

- EC2スポットインスタンス
  - Amazon EC2の購入オプション、最大で9割引
  - EC2サービスの空きキャパシティから起動するため、起動できない場合、また中断が発生することがある
  - 中断に強い仕組みを導入する必要がある
- Graviton搭載インスタンス
  - インスタンスファミリーのひとつ
  - オンデマンド価格が同等のx86ベースインスタンスより最大20%安い
  - 「中断に強い仕組み」の導入は不要
  - ArmベースのWindows Server製品が存在しないため、Windows環境では選択できない

Gravitonインスタンスはオンデマンドインスタンスで起動してよく、  
そのまま費用を下げるができる

# 関連リンク集

---

- AWS Gravitonについて
  - Black Belt オンラインセミナー AWS Graviton — <https://www.youtube.com/watch?v=MNHch4klkyo>
- EC2 スポットインスタンスについて
  - Black Belt オンラインセミナー Amazon EC2 スポットインスタンスの基礎 — <https://www.youtube.com/watch?v=3blgtNU-ZV0>
  - Black Belt オンラインセミナー 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 — <https://www.youtube.com/watch?v=HRaDSZ8N5TA>
- EC2 インスタンスのコスト最適化について
  - (要登録) AWS Summit Tokyo 2023 Amazon EC2のイロハとコスト最適化のキモ (AWS-01) — <https://jpsummit.awsevents.com/public/session/view/520>
- AWS Graviton Technical Guide — <https://github.com/aws/aws-graviton-getting-started>





Thank you!