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

宮本 大輔
技術統括本部 Compute/HPC スペシャリストソリューションアーキテクト
アマゾン ウェブ サービス ジャパン合同会社

宮本 大輔

アマゾン ウェブ サービス ジャパン 合同会社
技術統括本部
Compute/HPC スペシャリストソリューションアーキテクト

製薬・金融・気象といった分野を中心に
AWS 上で大規模な計算を行われるお客様の技術支援を担当

本セッションについて

本セッションのねらい

- **Arm アーキテクチャを採用した AWS Graviton プロセッサの特徴や移行方法を理解し Amazon EC2 や AWS Graviton 対応マネージドサービスのコスト最適化を実施できるようになる**

本セッションの対象参加者

- Amazon EC2 や AWS Graviton 対応マネージドサービスを利用中で、コスト最適化に取り組みたい方
- AWS Graviton について聞いたことはあるが、Arm アーキテクチャへの移行に不安がある方
- 上記課題を持つお客様にシステム提案を行う立場の方

本セッションでは説明しないこと

- Amazon EC2 や各種マネージドサービス自体の詳細

Agenda

- Amazon EC2 のコスト最適化手法
- AWS Graviton2 の概要とエコシステム
- アプリケーションの AWS Graviton2 への移行
- AWS におけるチップ開発の歴史と AWS Graviton3

AWS Graviton2 はなぜ高コストパフォーマンスなのか
Arm アーキテクチャにどのように移行するのかをご紹介

Agenda

- **Amazon EC2 のコスト最適化手法**
- AWS Graviton2 の概要とエコシステム
- アプリケーションの AWS Graviton2 への移行
- AWS におけるチップ開発の歴史と AWS Graviton3

Amazon EC2 (Elastic Compute Cloud)

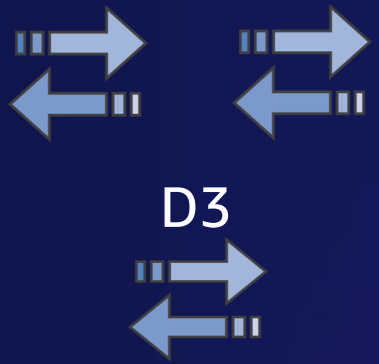
- 必要なときに必要な計算リソースを確保可能な仮想サーバサービス
- 数分で起動し、秒単位の従量課金（一部タイプについては1時間単位）
- 独自の仮想化基盤（Nitro System）により仮想化のオーバヘッドを極小化
- ワークロードに応じて 475 種類を超えるインスタンスタイプから選択可能

インスタンスタイプ一覧と分類

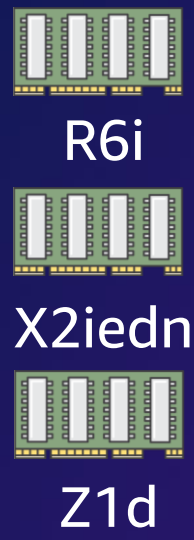
汎用

コンピューティング最適化

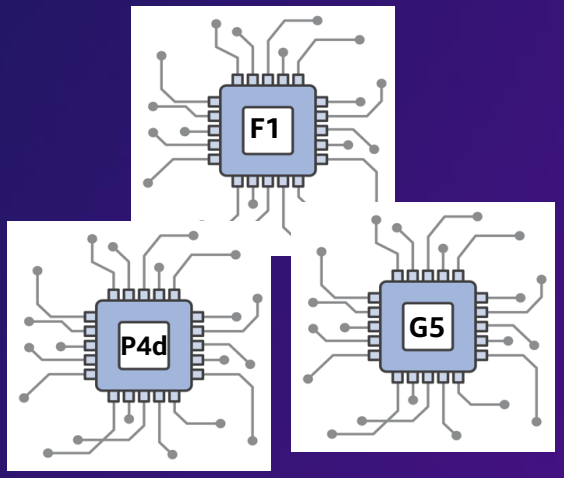
ストレージ・IO最適化



メモリ最適化



GPU ・FPGA アクセラレーテッド



Amazon EC2 のコスト最適化手法例

インスタンス台数の最適化

- Amazon EC2 Auto Scaling による動的な台数変化も活用し、負荷に適合した台数を使用

購入オプションの最適化

- 用途に応じて、オンデマンドだけでなくスポットインスタンスや
  リザーブドインスタンス/Savings Plans も活用

インスタンスタイプの最適化

- 汎用（M）、コンピュート最適化（C）、メモリ最適化（R/X）等から適切なものを選択
- **CPU の検討（Intel Xeon、AMD EPYC、AWS Graviton）**

本セッションでは「AWS Graviton 搭載インスタンスの活用」に
フォーカスして紹介
実際には上記項目を多面的に検討することが重要

コスト最適化については AWS-52 「ここがすごいよ Amazon EC2」も併せてご参照ください

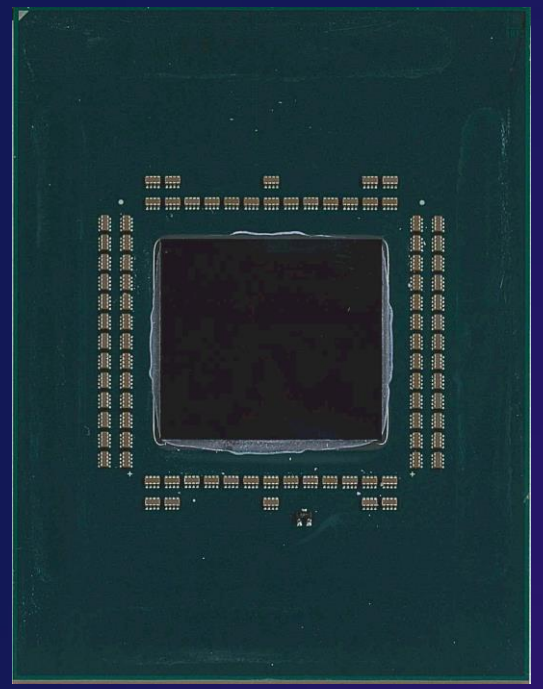
Agenda

- Amazon EC2 のコスト最適化手法
- **AWS Graviton2 の概要とエコシステム**
- アプリケーションの AWS Graviton2 への移行
- AWS におけるチップ開発の歴史と AWS Graviton3

AWS Graviton2 プロセッサ

AWS が独自に設計した **Arm アーキテクチャ** CPU

- Arm のサーバ向けコアである Arm Neoverse N1 を採用
- 1 チップに 64 物理コアを搭載
- 7 nm プロセスルール、300億トランジスタ



AWS Graviton2 は
なぜコストパフォーマンスが高いのか？

コストが低い + パフォーマンスが高い

Graviton2 はなぜコストパフォーマンスが高いのか

コスト：

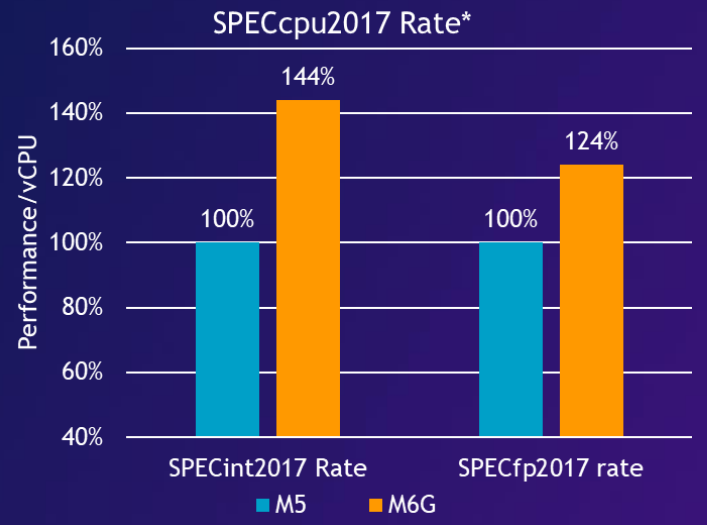
x86 系インスタンスの同サイズと比較した場合、**約 20 % 安価** な価格設定

東京リージョンでの例： m5.16xlarge: 3.968 USD/hour
m6g.16xlarge: 3.168 USD/hour

パフォーマンス：

x86 系インスタンスでは 2 vCPU = 1 物理コア だが、
Arm 系インスタンスは 1 vCPU = 1 物理コア であり、
同インスタンスサイズでは **2 倍の物理コアが利用可能**

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



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

Webサーバ、コンテナ基盤、BigData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 **約 40 % のコストパフォーマンス向上**

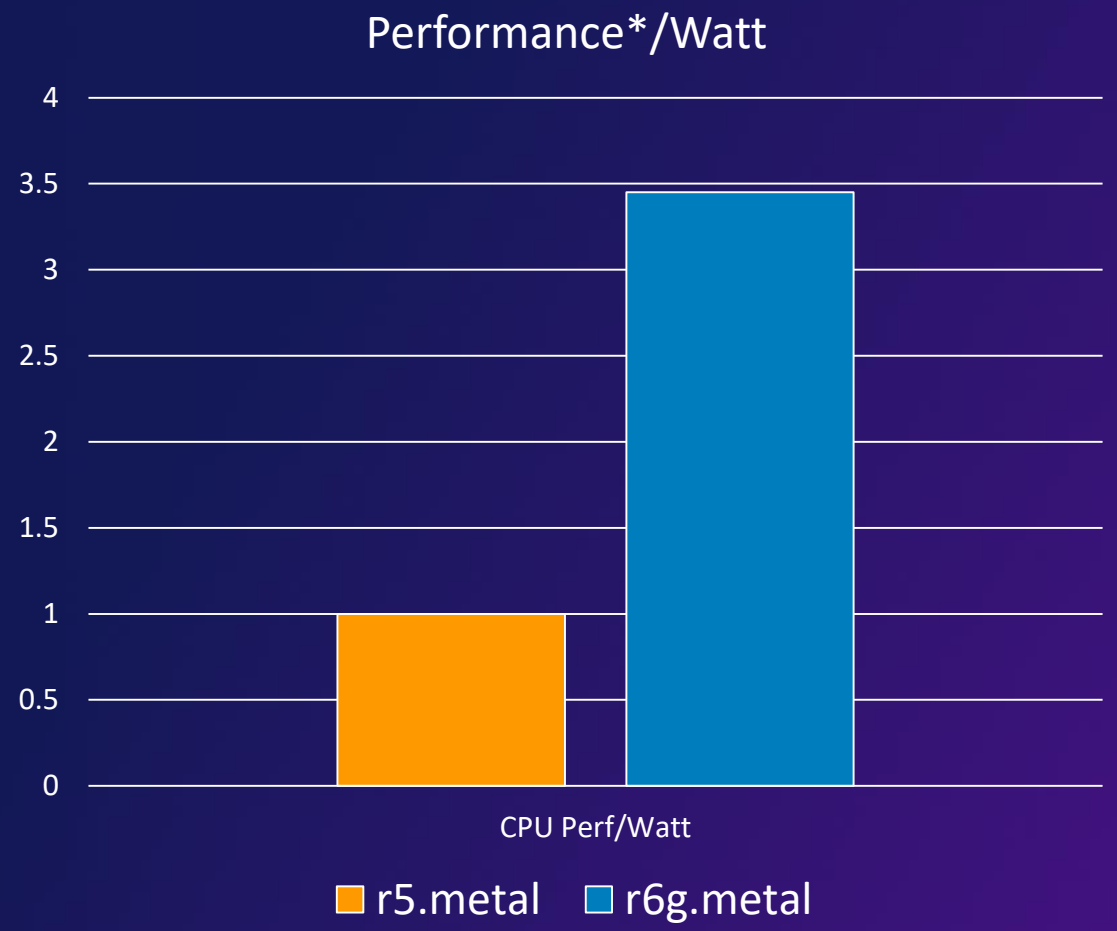
AWS Graviton2 の消費電力効率

Graviton2 は同等の x86 系インスタンスと比較して高い消費電力効率

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



AWS Graviton2 を活用いただいているカスタマー

https://aws.amazon.com/ec2/graviton/

Amazon Prime Day での AWS Graviton2 活用

Amazon.com は、Datapath と呼ばれる構造化データのクエリ・ルックアップ・結合を担当する内部サービスを AWS Graviton2 に移植した

- 53,200 台を超える Graviton2 搭載インスタンスを使用
- 同等の第 5 世代 x86 ベースインスタンスと比較した場合 AWS Graviton2 の **コストパフォーマンスは最大 40% 優れていた**
- AWS Graviton2 の高い電力効率は、気候変動に対処するための目標達成に役立つ

参考：その他の指標による Amazon Prime Day の規模

Amazon CloudFront 1 分あたり 2 億 9,000 万件を超える HTTP リクエストのピーク負荷を処理し、合計 6,000 億件を超える HTTP リクエストを処理

Amazon Elastic Block Store 159 ペタバイトのストレージを追加し、1 日あたり 614 ペタバイトを転送

Amazon Aurora 3,715 インスタンスが 2,330 億件のトランザクションを処理し、1,595 テラバイトのデータを格納し、615 テラバイトのデータを転送

https://aws.amazon.com/jp/blogs/news/prime-day-2021-two-chart-topping-days/

日本でも様々なカスタマーが Graviton2 を活用中

CyberAgent is a Japanese internet media-services company operating a wide variety of web services in support of media, Internet advertising and game businesses.

> "The CyberAgent Service Reliability Group performs verification of each web service are migrating multiple services from existing x86-based EC2 instances to AWS Graviton2-based instances. The dot money by Ameba award point exchange service was the first service to migrate. Adopting Graviton2 delivered total cost reductions of 50-60% from a combination of reduced development fleet size, and production service optimization using smaller instances that reduce costs while maintaining service performance. Based on this positive experience we look forward to adopting Graviton2 for additional services to drive additional cost savings."

- Takuya Hasegawa, Engineer, Service Reliability Group, CyberAgent, inc.

> "NAVITIME is the leading provider of navigation technology and services in Japan. We completed the deployment of 4 services, the common API, map data, tile map, and full-text search distribution services, running on x86 based instances to Graviton2-based M6g and C6g instances within just 3 months for the purpose of cost reduction. These services are running on Java8 and C/C++ applications with Amazon EKS/ECS. Migration to M6g instance provided up to 15% higher throughput at 20% lower cost. We are investigating additional workloads to migrate to Graviton2 instances."

Katsuhide Kayashima, senior engineer, NAVITIME JAPAN Co., Ltd.

Supership enables enterprise digital transformation with one-stop solution for collecting, analyzing data scattered throughout the company, and automating them with AI technology.

> "We had two major challenges in business - to reduce large number of lightweight web application infrastructure cost and to improve mass data processing performance. For lightweight web applications, we could see ~40% higher request-per-second performance at lower cost on m6g.large comparing to m5.large.As mass data processing performance, we tested Elasticsearch7.3/OpenJDK11 workload on c6g.xlarge up to c6g.8xlarge and confirmed ~40% better price performance compared to C5 instances. It was easy to transition most of our workloads from x86 to Arm-based AWS Graviton2. Most of processing systems could be built without any special effort."

- Yutaka Nakano, Data Solution Studio Engineering Group II - Supership Inc.

> "Nulab is a software development company by and for creators. Our products — a project management tool called Backlog, an online diagramming tool called Cacoo, and a team chat tool called Typetalk — help teams around the world enjoy and simplify the creative process of bringing their ideas to life. Nulab migrated all of the Typetalk workloads from Amazon EC2 M5 to AWS Graviton2-powered M6g instances. As a result, the overall response time of production environment improved by up to 30%, and our annual EC2 costs were reduced by about 30%. The migration was quick, and Graviton2 exceeded our expectations. The Nulab team is excited about migrating even more workloads to Graviton2-based instances for price performance gains."

- Hisatomo Futahashi, Site Reliability Engineer, Nulab Inc.

fluct is the largest ad tech company operating SSP (Supply-Side Platform) in Japan.

"We adopted AWS Graviton2-based C6gd instances for our Supply-Side Platform's ad server based on CentOS 8, Go, and Erlang/OTP. It took about a month for validation and we realized 60% better performance over comparable x86-based instances. Most of our Graviton2-based instances are running on EC2 Spot and our total cost savings with Graviton2-based instances are around 40%-50%."

Takuya Nishigori, Principal Engineer, fluct, Inc.

https://aws.amazon.com/jp/ec2/graviton/customers/?nc1=h_ls

現在利用しているものと同様の環境を AWS Graviton2 でも利用できるのか？

x86 系とほぼ同等の多様なインスタンスタイプに加え 多くの OS、コンテナ環境、パートナーソリューション マネージドサービスが利用可能

Amazon EC2 AWS Graviton2 搭載インスタンス

第5世代の x86 インスタンスと比較し最大 40 % のコストパフォーマンス向上
多くのインスタンスファミリーがローンチ済み

M6g	T4g	C6g	R6g	X2gd	C6gn
汎用	汎用 （ブーストタイプ）	コンピュート最適化	メモリ最適化	メモリ最適化 （より多くのメモリ搭載）	高速ネットワーク搭載

Local NVMe-based SSD storage 搭載:
general purpose (M6gd), compute-optimized (C6gd), and memory-optimized (R6gd & X2gd)

ベアメタルタイプも提供中:
(M6g.metal, M6gd.metal, C6g.metal, C6gd.metal, R6g.metal, R6gd.metal, X2gd.metal)

AWS Graviton2 Instance 対応状況（収録時）
・東京リージョン: 3AZ対応済み (※C6gnは2AZのみ, X2gdインスタンスは未対応)
・大阪リージョン: 未対応

Amazon EC2 G5g インスタンス

GPU を搭載した初の AWS Graviton2 ベースインスタンス

- NVIDIA T4 Tensor コア GPU を搭載
- コストパフォーンマンスの高い GPU インスタンスであり
  ゲームストリーミング・グラフィックス処理
  機械学習における推論・自動運転シミュレーションに最適
- Arm ベースの Android ゲームをネイティブ実行可能
  クロスコンパイルやエミュレーションを必要としない
- Android ゲームのストリーミングにおいて
  コストパフォーマンスを最大 30% 向上

※ 東京リージョンでも利用可能

https://aws.amazon.com/jp/blogs/news/new-amazon-ec2-g5g-instances-powered-by-aws-graviton2-processors-and-nvidia-t4g-tensor-core-gpus/

Amazon EC2 Im4gn/Is4gen インスタンス

高速なローカルストレージ搭載の AWS Graviton2 インスタンス

第 2 世代 AWS Nitro SSD の採用により I3 インスタンスと比較して 75 % 低いレイテンシ変動

- I3 インスタンスと比較して
  コストパフォーマンスが 最大 40% 向上
- NoSQL データベース、データの索引・検索、
  データ分析に最適

- I3en インスタンスと比較してストレージの
  TB あたりのコストが最大 15% 安価
- ストリーム処理、リアルタイムデータベース、
  ログ解析に最適

https://aws.amazon.com/jp/blogs/news/new-storage-optimized-amazon-ec2-instances-im4gn-and-is4gen-powered-by-aws-graviton2-processors/
https://aws.amazon.com/jp/blogs/news/aws-nitro-ssd-high-performance-storage-for-your-i-o-intensive-applications/

AWS Graviton 対応 OS・コンテナ環境

Operating Systems

- Amazon Linux 2
- Red Hat Enterprise Linux 8.2+
- SLES 15 SP2+
- 18.04LTS, 20.04LTS
- CentOS
- Debian
- fedora
- FreeBSD
- NetBSD

Containers

- Amazon Elastic Container Service (Amazon ECS)
- AWS Fargate
- Amazon Elastic Kubernetes Service (Amazon EKS)
- Docker
- Kubernetes

AWS Graviton Ready Partners

多くのパートナーソリューションが AWS Graviton をサポート済み

aqua
aviatrix
Buildkite
CANONICAL
circleci
Cribl
DATADOG
docker
dynatrace
elastic
Genymotion Device image
GitLab
GRANULATE
honeycomb.io
InterSystems Creative data technology
Lacework
New Relic
Qualys.
Red Hat
SUSE
Starburst
tenable
TREND MICRO

https://aws.amazon.com/jp/ec2/graviton/partners/

AWS Graviton2 対応済みマネージドサービス

様々なマネージドサービスが AWS Graviton2 に対応済み

マネージドサービスにおいても **AWS Graviton2 は優れたコストパフォーマンスを提供**

マネージドサービスでは、**計算基盤が x86 であるか Arm であるかを意識する必要性が低く**、
移行・移植の手間を削減（パフォーマンス検証は必要）

Databases

Amazon DocumentDB

Amazon Aurora

Amazon RDS
(Open-Source DBs)

Amazon ElastiCache

Amazon MemoryDB

NEW!

Amazon Neptune

Analytics

Amazon OpenSearch

Amazon EMR

AWS Lambda

Compute

NEW!

AWS Fargate

NEW!

AWS Elastic Beanstalk

Storage

NEW!

Amazon FSx for Lustre, Open ZFS

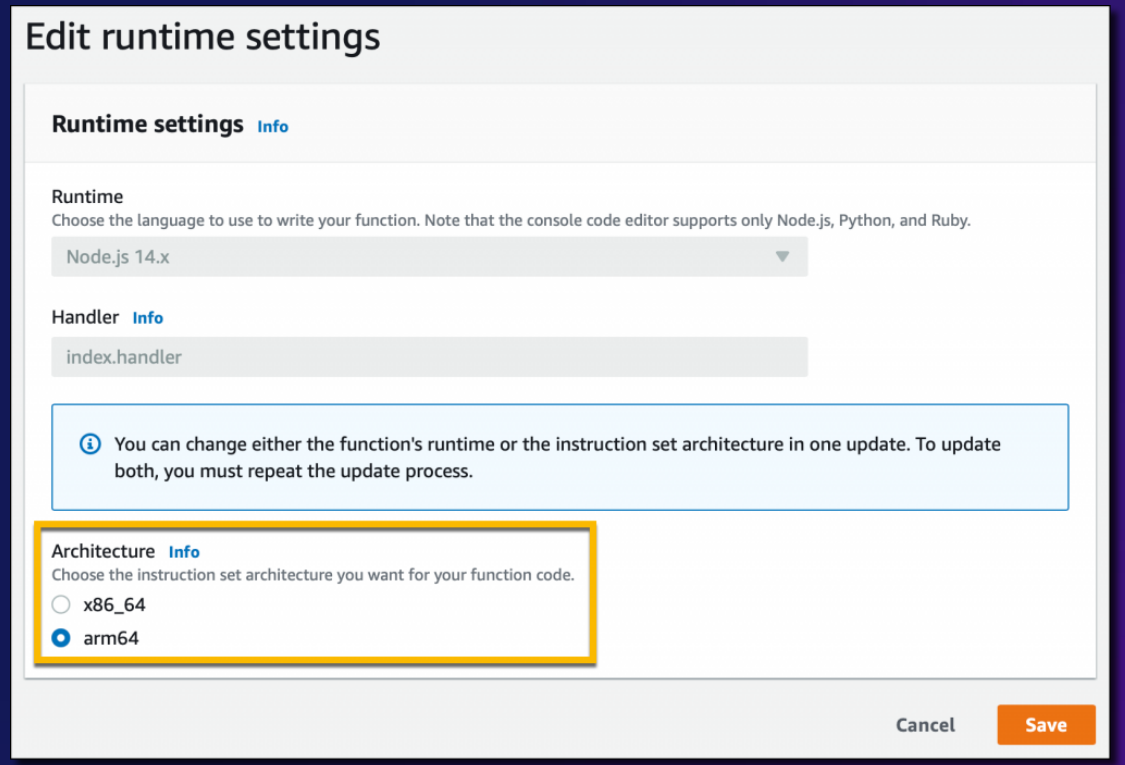
https://github.com/aws/aws-graviton-getting-started/blob/main/managed_services.md

AWS Lambda での AWS Graviton2 活用方法

arm64 アーキテクチャの選択により最大34%のコストパフォーマンス向上

- 20% 安価な時間当たり料金
- 最大 19 % 優れたパフォーマンス
- バイナリ依存のない関数であれば、
  スイッチを切り替えるように移行が可能

AWS Lambda



https://aws.amazon.com/jp/blogs/news/aws-lambda-functions-powered-by-aws-graviton2-processor-run-your-functions-on-arm-and-get-up-to-34-better-price-performance/

Agenda

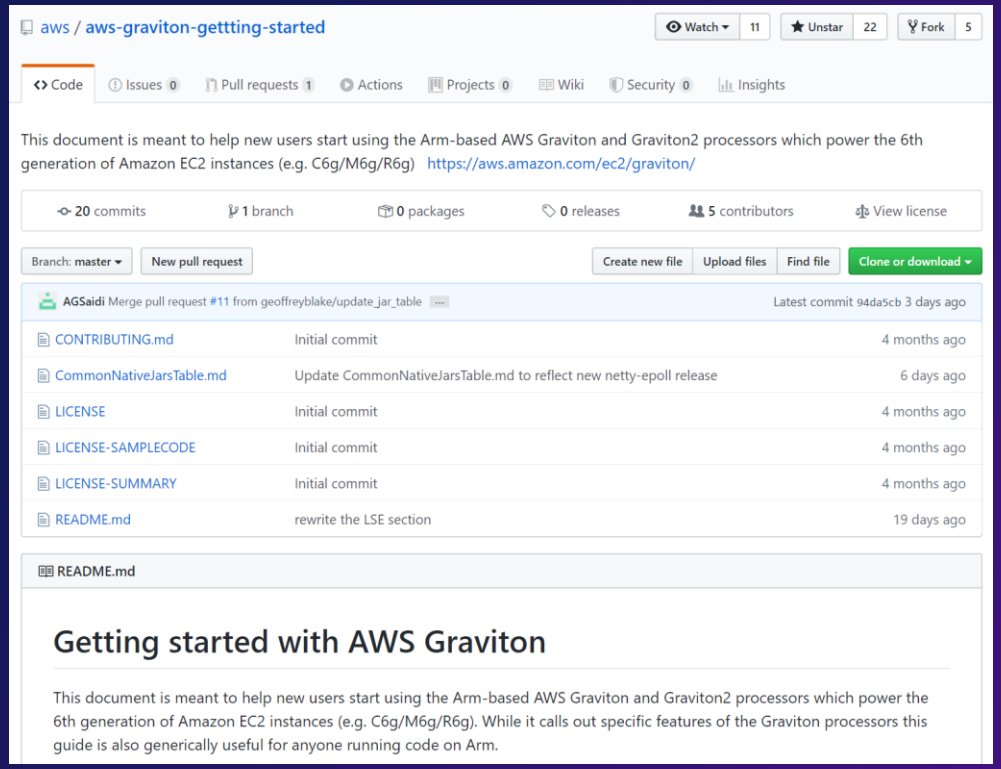
- Amazon EC2 のコスト最適化手法
- AWS Graviton2 の概要とエコシステム
- **アプリケーションの AWS Graviton2 への移行**
- AWS におけるチップ開発の歴史と AWS Graviton3

AWS Graviton getting-started

AWS Graviton/Graviton2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-gettting-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..



Arm アーキテクチャである AWS Graviton への移行は大変？

プログラミング言語環境を含む
様々な OSS アプリケーションが対応済み
多くのカスタマーから想像よりずっと簡単だったとの声

ワークロードタイプ別 移行ガイド 1

ワークロードの種類別に複数の移行パターンが存在

Case1: Graviton2 をサポートしているマネージドサービスを利用している場合

- 多くの場合、**タイプの変更のみ**で移行可能
- パフォーマンステストは実施する必要がある

Case2: アプリケーションが yum/apt 等のパッケージマネージャで取得できる場合

- まずはパッケージマネージャからインストールを行い、評価を実施
- 想定されるパフォーマンスが得られない場合、ソースコードからのコンパイルにより性能向上する場合も

Case3: Python や Ruby 、 Java 等でアプリケーションが記述されている場合

- 基本的には**そのまま利用する事が可能**だがパフォーマンステストは行う必要がある
- **最新の言語環境**（OpenJDK11/Amazon Corretto11 等）で性能向上がみられるケースもあるため、利用推奨
- JNI や Python-native モジュール等、**ネイティブバイナリが使用されている場合は別途対応**が必要

ワークロードタイプ別 移行ガイド 2

Case4: C/C++、FORTRAN など、コンパイルを行う必要がある場合

- **getting-started を熟読**し、推奨環境・コンパイルオプションを使用（後述）
- Arm Neoverse N1 向けに最適化された数学ライブラリである **Arm Performance Libraries も無償で提供**されている (BLAS, LAPACK, FFT 等をサポート)

Case5: コンテナを利用している場合

- Case2/3/4 と並行して検討
- Graviton 上でもコンテナは利用可能だが、**アーキテクチャごとに Docker イメージを作成**する必要がある（後述）

Case6: Microsoft Windows Server 上で動作するアプリケーションの場合

- Microsoft Windows Server には Arm 版が存在せず AWS Graviton2 では利用できない
- Linux への移植を検討、また .NET core 環境であれば Linux 上でも動作

C/C++ におけるコンパイル時の注意点

GCCのバージョン

- GCC 9以降を推奨

GCC における Graviton2 向け推奨コンパイルオプション

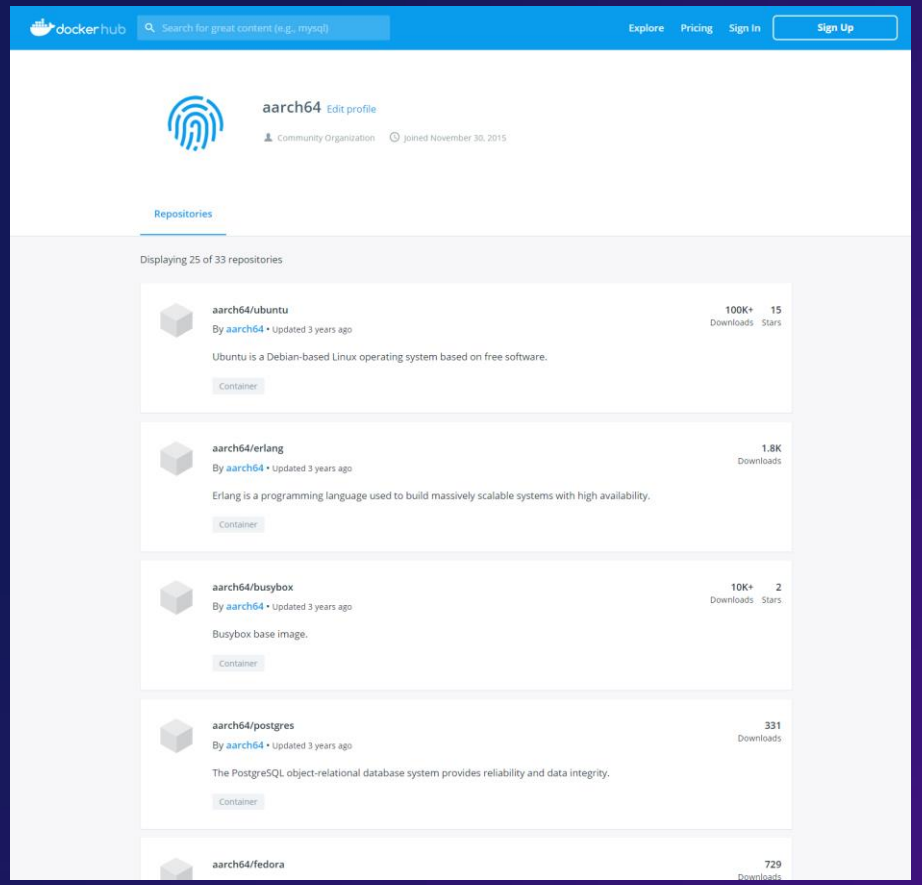
- `-march=armv8.2-a+fp16+rcpc+dotprod+crypto`

Large-System Extensions の活用

- Graviton2 では、LSE (Large-System Extensions) をサポートしており POSIX thread における高速なスレッド間同期などを提供
- 現在は Ubuntu 20.04 等で LSE に対応した libc6-lse ライブラリが提供されており上記コンパイルオプションを使用し適切にリンクを行うことで利用可能
- HPCアプリケーション等での OpenMP 利用時にも効果が大きい

AWS Graviton2 でのコンテナ利用の基本

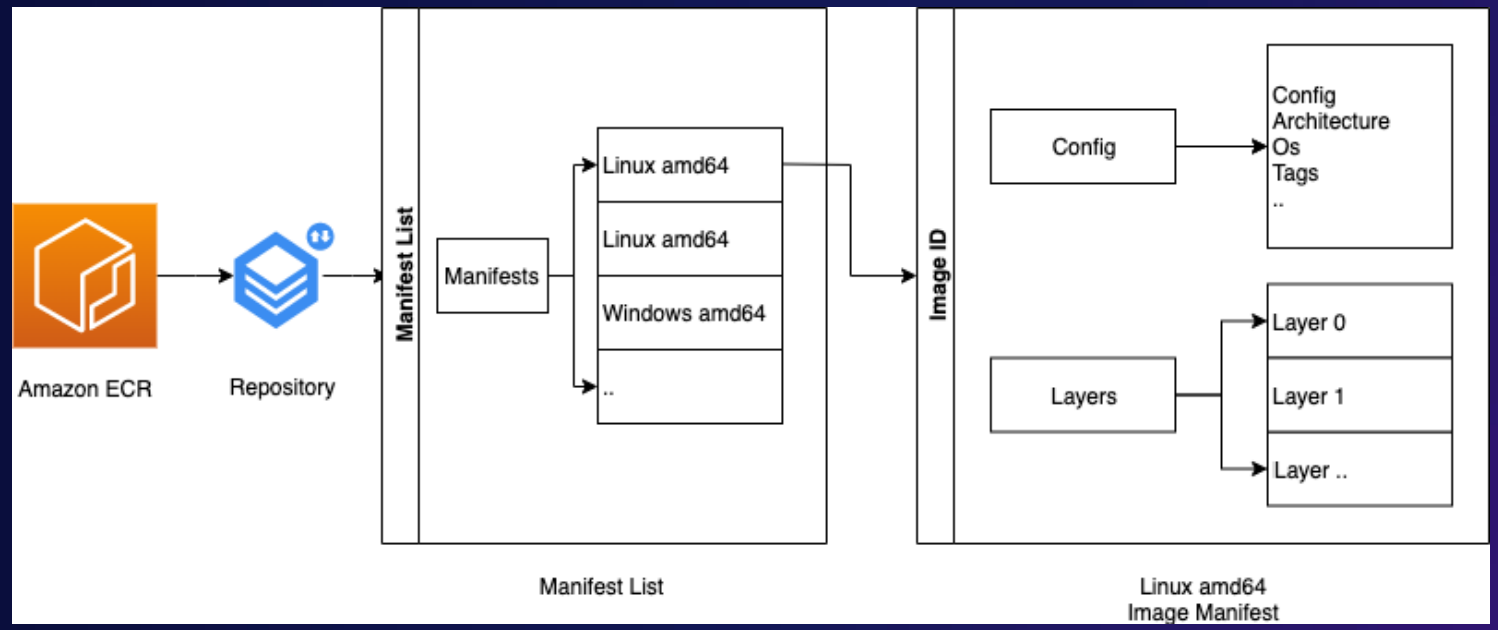
- Graviton2 上で Docker を利用することは可能（イメージの作成・実行等）
- ただし、Image は、aarch64 (arm64) と amd64 (x86_64) で互換性が無く アーキテクチャごとに Image を作成する必要がある
- DockerHub 上では、主要な Image について aarch64 版が公開済み



https://hub.docker.com/u/aarch64/

マルチアーキテクチャ対応 Docker Image

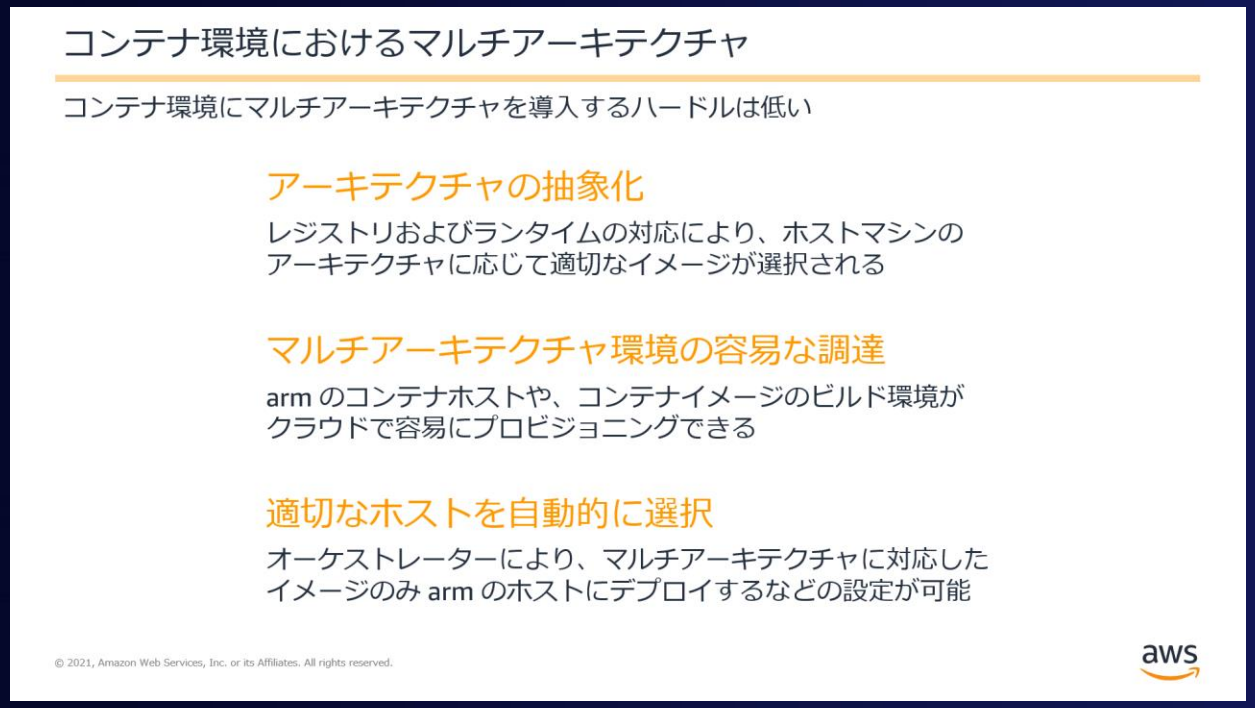
- **イメージ作成**：Graviton 上での build だけでなく、Docker の buildx サブコマンドにより、QEMU を使用して、amd64 環境でも aarch64 用イメージの作成が可能
- **レポジトリ**：Docker の Manifest 機能を使用することで、単一のレポジトリで複数のアーキテクチャをホストする事が可能（Amazon ECR も対応済み）

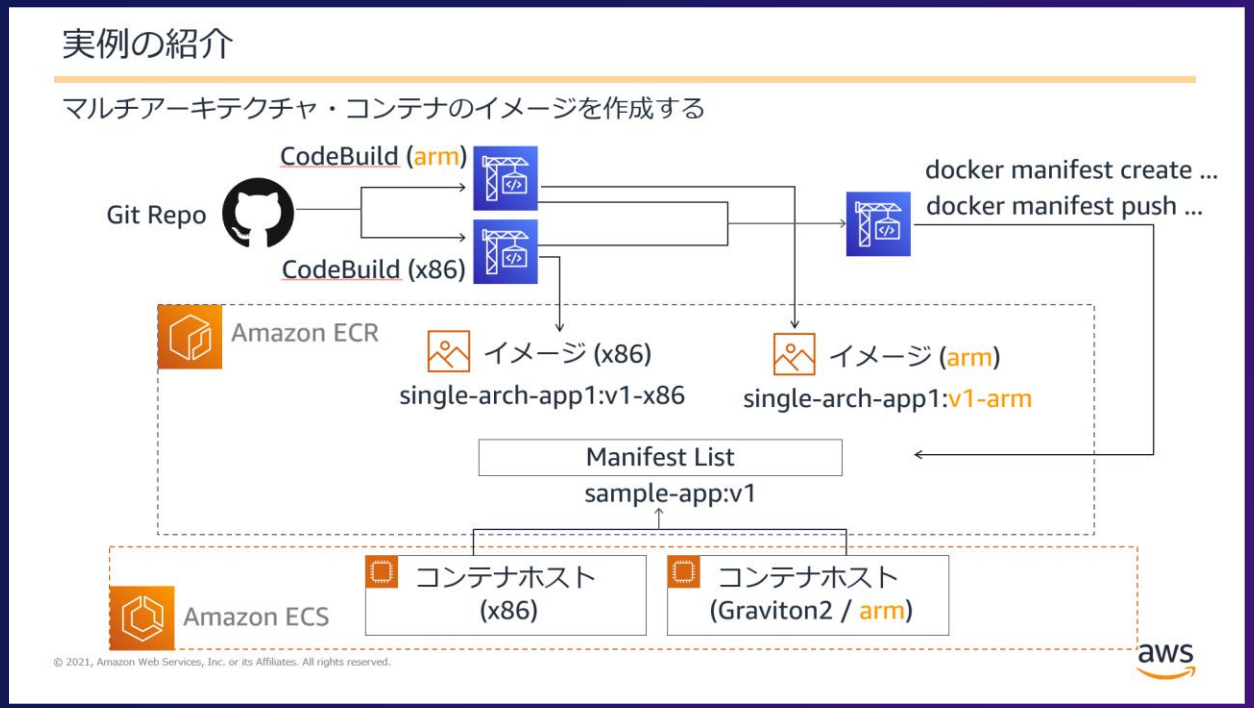


https://aws.amazon.com/jp/blogs/containers/introducing-multi-architecture-container-images-for-amazon-ecr/

マルチアーキテクチャでのコンテナ活用

[AWS Black Belt Online Seminar] CON437 AWS Graviton2 でマルチアーキテクチャのデリバリーパイプラインを作成する
では、マルチアーキテクチャでのコンテナビルド環境についてご紹介





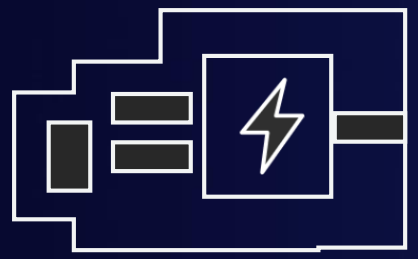
https://aws.amazon.com/jp/blogs/news/aws-black-belt-online-seminar-con437-containers-delivery-graviton2/

Agenda

- Amazon EC2 のコスト最適化手法
- AWS Graviton2 の概要とエコシステム
- アプリケーションの AWS Graviton2 への移行
- **AWS におけるチップ開発の歴史と AWS Graviton3**

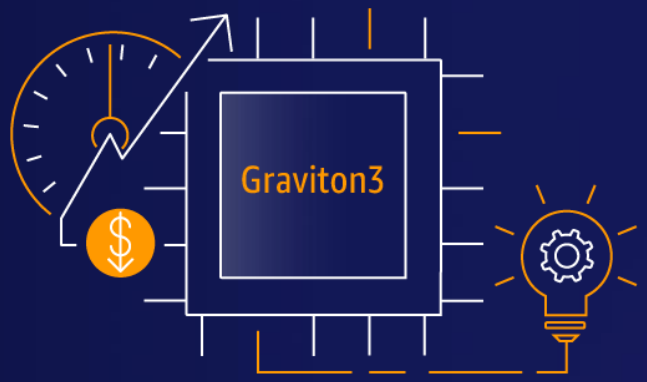
AWS におけるシリコンイノベーション

AWS ではイノベーションのため、独自にチップ設計を実施



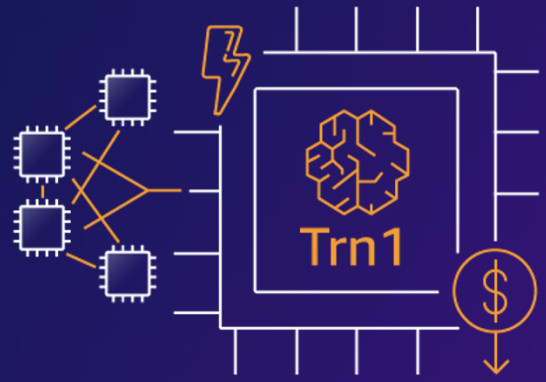
AWS Nitro System

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

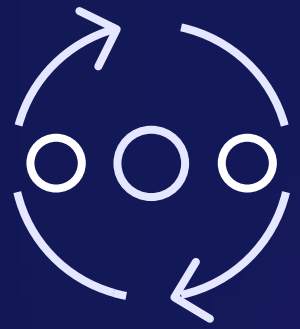
なぜ独自チップを作るのか？

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

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



スピード

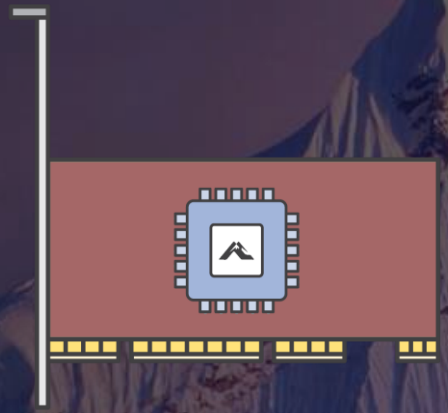
製品の仕様化から導入
までエンドツーエンド
の開発プロセス

イノベーション

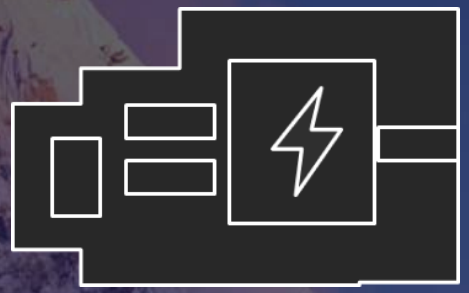
より多くの価値を創造
エンドツーエンドでの
最適化

Annapurna Labs によるチップ開発の歴史

I/O Accelerator Card



Nitro Card

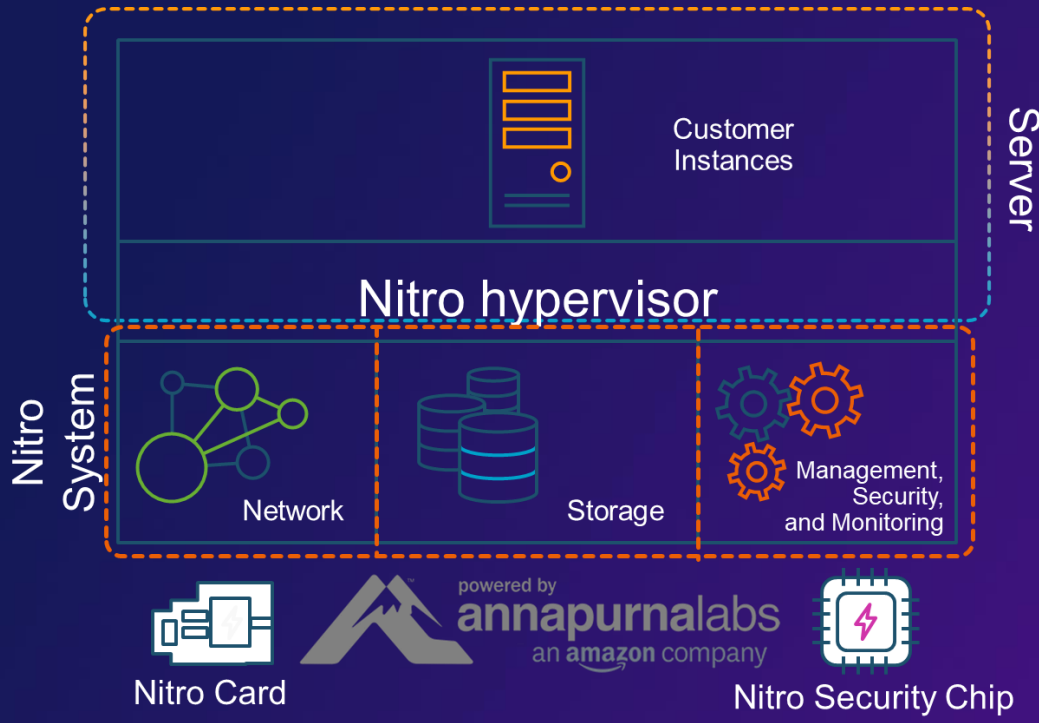


7+ years of innovation with Annapurna Labs

EC2 のシステム基盤 (AWS Nitro System)

独自のハードウェア/Hypervisor により最適化された性能を提供

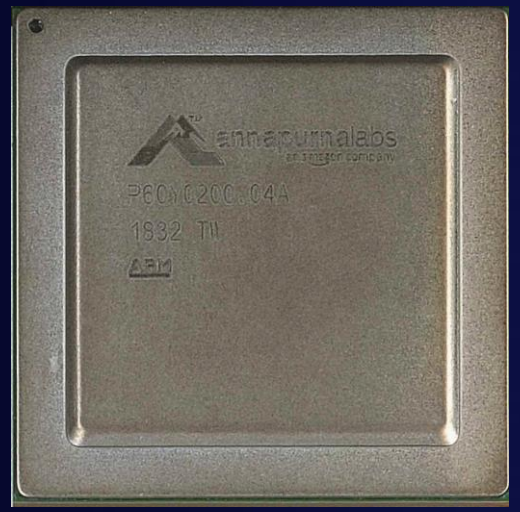
- 2017年11月に発表
- C5、M5、R5 の世代から対応
- 独自のハードウェアや
  KVM ベース の Hypervisor により
  仮想化オーバーヘッドを低減
- 幅広いインスタンスでベアメタルタイプを提供



Annapurna Labs によるチップ開発の経験から Arm アーキテクチャの採用によって よりコストパフォーマンスに優れた選択肢を カスタマーに提供できるのではないか

第一世代 AWS Graviton プロセッサ

2018 年 11 月に発表

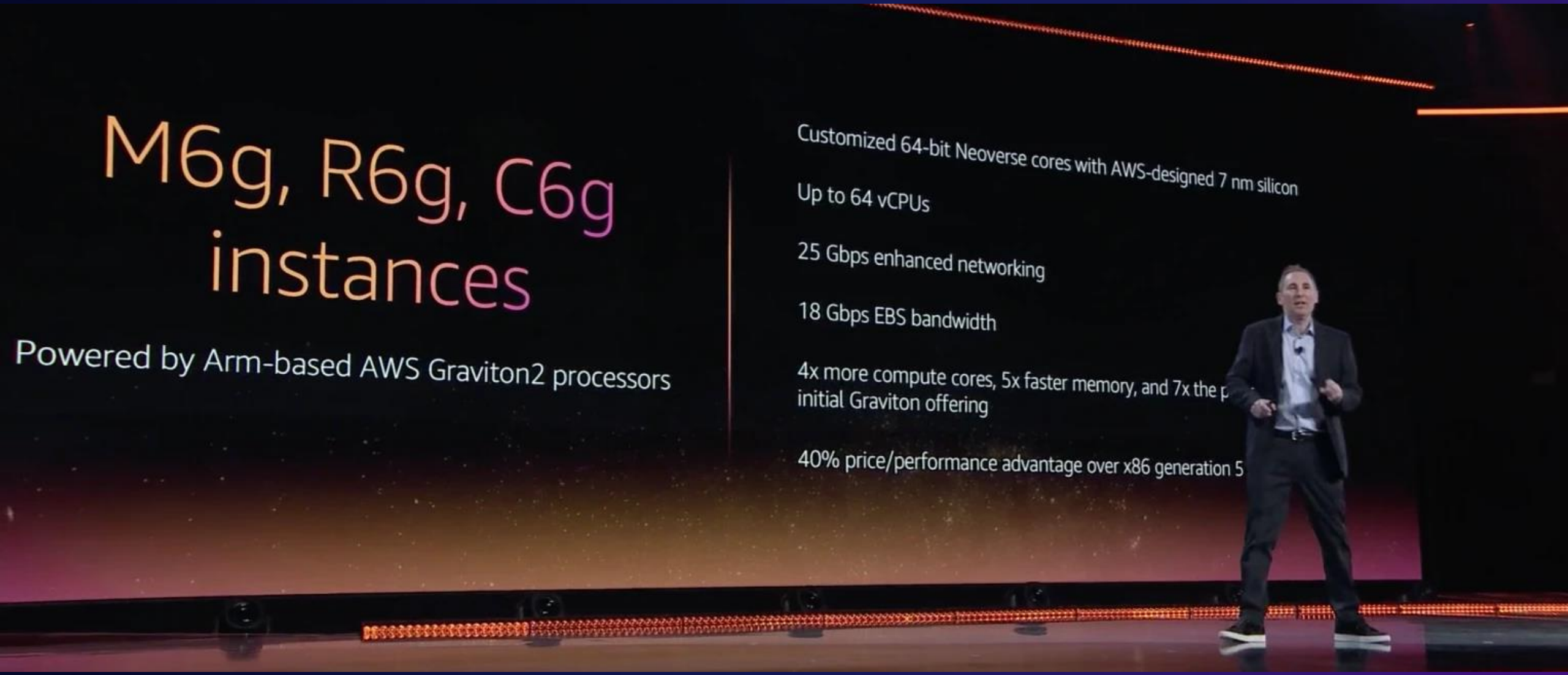


64-bit Arm コア採用 AWS カスタムチップ
16 物理コア搭載

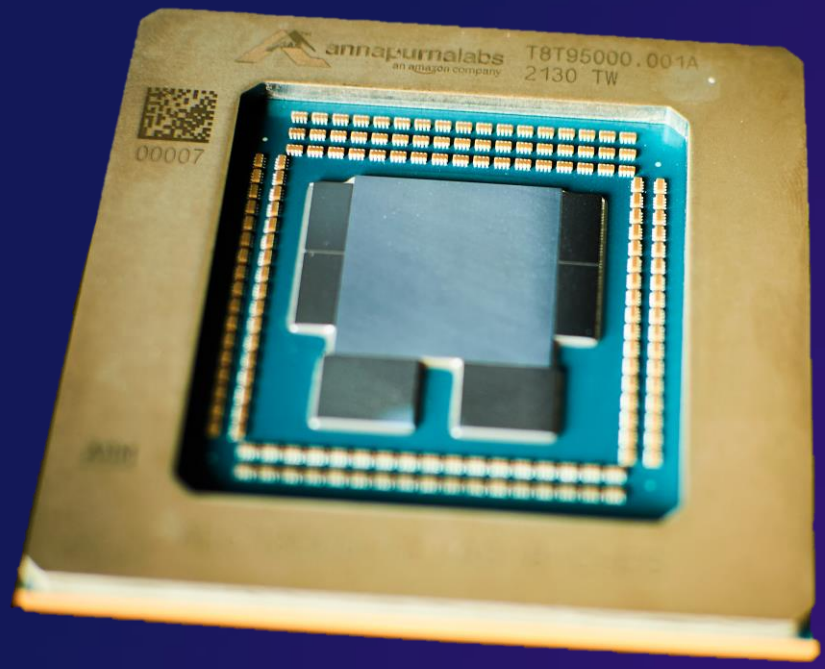
A1 インスタンスとしてローンチ

顧客からのフィードバックのもと
Arm エコシステムの醸成に貢献

re:Invent 2019 Graviton2 搭載 M6g, R6g, C6g 発表



AWS Graviton3



re:Invent 2021 で発表

- 第三世代 AWS Graviton プロセッサ
- 55 billion トランジスタ
- AWS として初の DDR5 メモリ採用により Graviton2 と比較して メモリバンド幅が +50 % 向上

AWS Graviton2 と比較してコストパフォーマンスが +25 % 向上

US East (N. Virginia) 及び US West (Oregon) にて一般提供開始済み
https://aws.amazon.com/blogs/aws/new-amazon-ec2-c7g-instances-powered-by-aws-graviton3-processors/

（参考）Graviton3 CPU enhancements

PtrAuth
RNG Inst

AWS Graviton2

4–8 wide Fetch

4 wide Decode

8 wide issue

AWS Graviton3

8 wide Fetch

5–8 wide Decode

15 wide issue & 2x larger instruction window

bfloat16
256b SVE

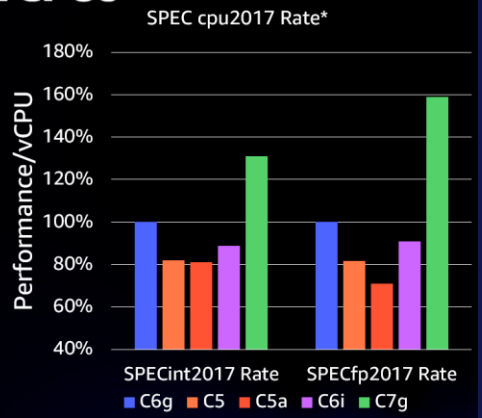
2x Mem ops
enhanced
prefetching

~2x
TLS

AWS Graviton3 ベンチマーク

SPEC cpu2017 rate – All vCPUs

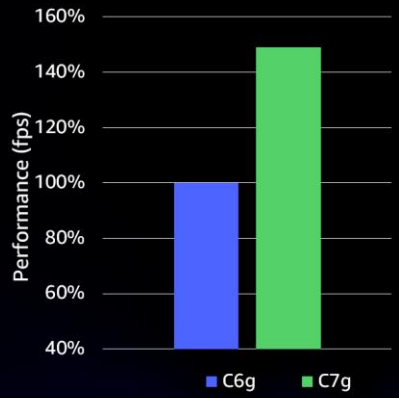
- Runs concurrently on all vCPUs
- Closer to real performance as all vCPUs are typically busy



All SPEC scores estimates, compiled with gcc v10 -Ofast –mtune=native –march=neoverse-n1 (Graviton) –march=native (x86), run on largest size for each instance type tested.

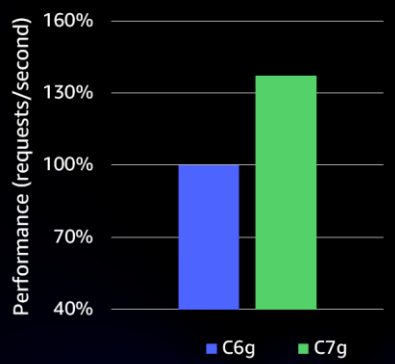
Video encoding with x264

- Video accounts for over 60% of downstream traffic on the internet
- Encoding reduces the bandwidth to deliver and store video content
- Used libx264 to encode uncompressed 1080p to h.264
- 49% more frames per second



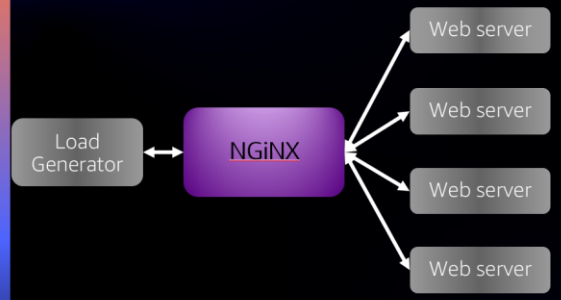
x264 (ae03d92b), 2xl instance size, veryslow preset, input uncompressed 1080p50, output encoded h.264 1080p50

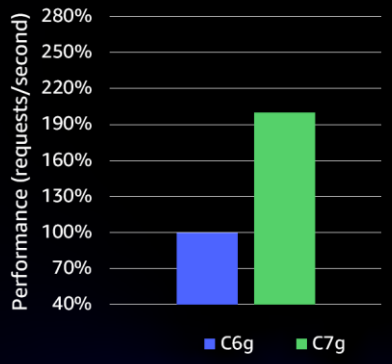
Node.JS applications



Node.JS 16.7.0, AcmeAir test application with one process per vCPU, JMeter load generator on c6g.4xlarge in a cluster placement group with NGiNX as reverse proxy, HTTP connections, connection count varied to control load

Load balancing with NGiNX





NGINX 1.18.0, 128 GET/POST payloads, all HTTPS connections, AES128-GCM-SHA256 encryption, OpenSSL 1.1.1f, 2xl system-under-test, load generator c5n.9xl, 4 c5.4xl as backend servers; all servers run in a cluster placement group

https://www.youtube.com/watch?v=WDKwwFQKfSI

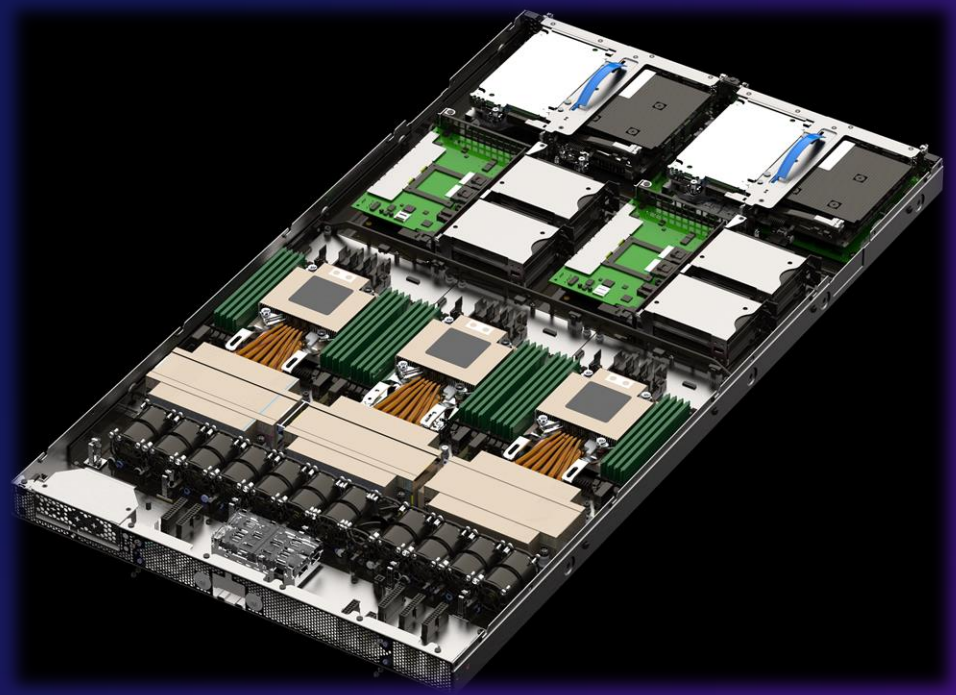
（参考）AWS Graviton3 ラック構成

チップ・パッケージ・マザーボードのデザインを
最適化し、1 サーバあたり 3 ソケット構成が可能に

これまでと比較してソケット/ラック が +50% 向上

ラックの電力要求については他のベンダーと同様

Nitro Card が 3 ソケット構成にも対応



AWS Graviton3 カスタマーフィードバック

“We found Graviton3-based C7g instances deliver 20-80% higher performance vs. Graviton2 based C6g instances, while also reducing tail latencies by as much as 35%. ”

“We were able to run 30% fewer instances of C7g than C6g serving the same workload, and with 30% reduced latency.”

“They are suitable for even the most demanding latency sensitive workloads while providing significant price performance benefits.”

“We have now found Graviton3 C7g instances to be 40% faster than the Graviton2 C6gn instances for those same simulations.“

https://aws.amazon.com/ec2/graviton/

まとめ

- Amazon EC2 には様々なコスト最適化手法があり
  AWS Graviton の活用もその中の一つ
- AWS Graviton2 は同等の x86 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
- 様々なマネージドサービス、プログラミング言語環境、コンテナ環境が対応済み
- AWS では、独自ハードウェアの設計・開発に長期的な投資を実施
  AWS Graviton3 は AWS Graviton2 と比較して +25% のコストパフォーマンス向上

本セッションをきっかけに、AWS Graviton プロセッサの活用による
コスト最適化に取り組んでみませんか？

AWS Graviton 参考 URL

- AWS Graviton プロセッサ
  https://aws.amazon.com/jp/ec2/graviton/
- AWS Graviton Getting Started
  https://github.com/aws/aws-graviton-gettting-started
- White Paper: AWS Graviton2 for ISVs
  https://docs.aws.amazon.com/whitepapers/latest/aws-graviton2-for-isv/welcome.html

関連セッション

- [AWS-52] ここがすごいよ Amazon EC2

Thank you!

Daisuke Miyamoto