



HPC on AWSの概要と最新動向

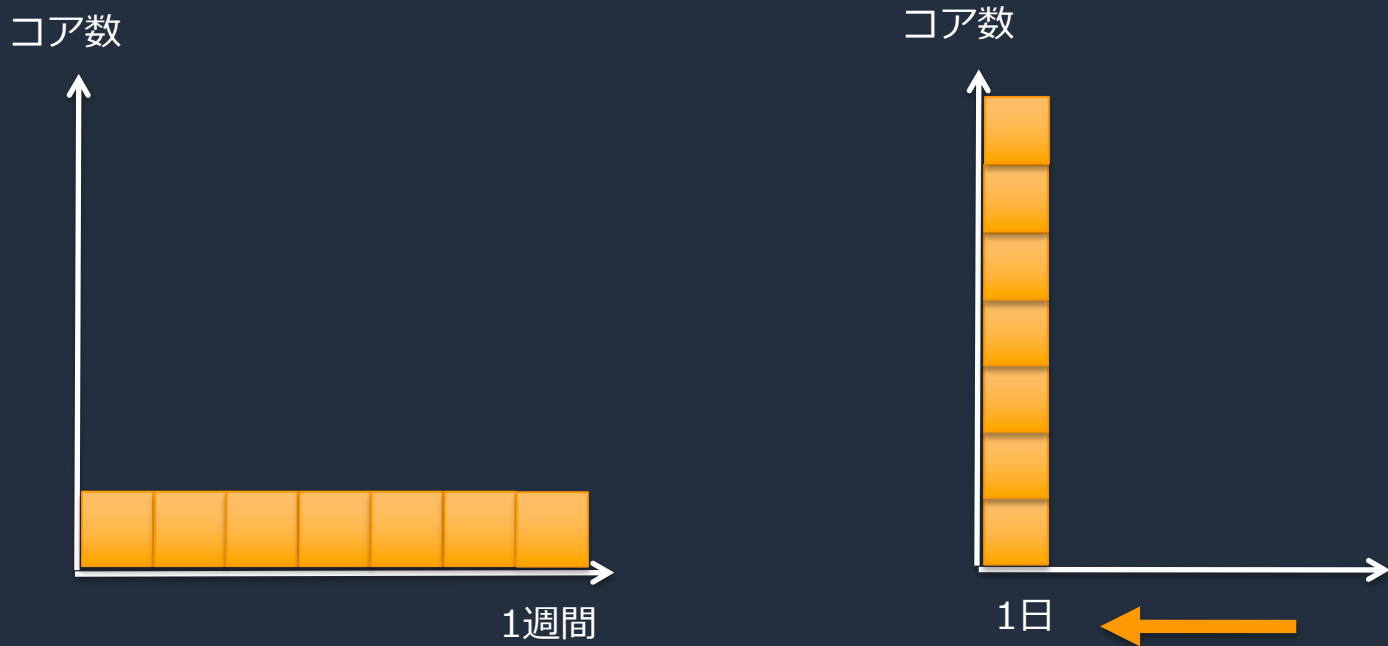
アマゾン ウェブ サービス ジャパン合同会社
ソリューションアーキテクト
吉廣 理

Agenda

- HPC on AWSの特徴
- AWSのHPC関連サービス概要
- HPCオーケストレーションサービス
- HPC on AWSアップデート
- まとめ

HPC on AWSの特徴

スケーラビリティの活用による計算時間短縮



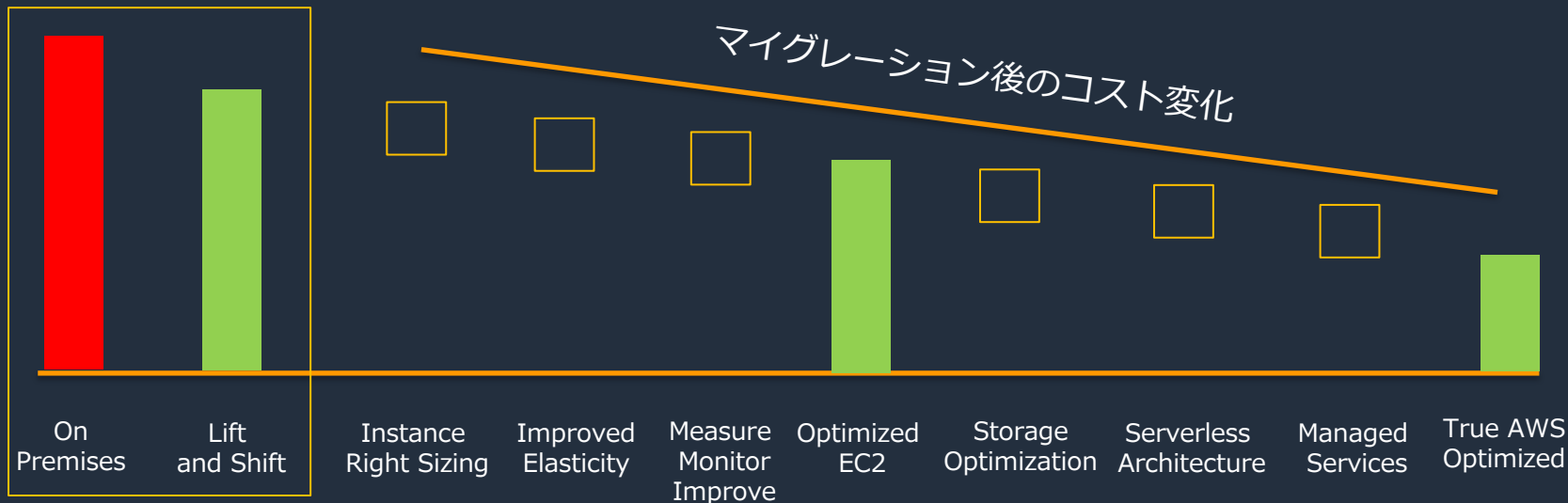
従来は手持ちの限られたリソースで、逐次処理していたジョブも
AWSなら必要な台数、インスタンスを起動して、一斉処理。
しかも費用は「時間×台数」なのでどちらも同じ。

継続的な改善によるコスト最適化

クラウド活用では、最初から最適を目指すよりも
小さな改善のサイクルを継続的に実施していくことが重要

マイグレーション後の構成の見直しや新しいサービスの活用

典型的なTCO比較



豊富なHPC関連サービス

クラウド上で高性能かつ使いやすいHPC環境を実現するAWSのHPC関連サービス

コンピューート

Amazon EC2



用途に応じて多様なインスタンスを利用可能な仮想サーバサービス



NVIDIA H100
GPU搭載



200 Gbps の
ネットワーク帯域



4.5GHzで動作する
高性能CPU搭載

スポットインスタンスの活用で大幅なコスト減も可能

ネットワーク

Placement Group

EC2インスタンスの基盤上の配置を制御してネットワークを高速化

Elastic Fabric Adapter

libfabric 対応の専用ネットワークにより MPI 利用のアプリケーション等を高速化

ストレージ

FSx for Lustre

S3連携する高速な並列ストレージをフルマネージドで提供

管理自動化

AWS Batch



スケーラブルなバッチコンピューティングジョブをフルマネージドで管理

AWS ParallelCluster



AWS上に Linux/HPC クラスタ環境を数ステップで構築、Slurmなどのスケジューラに対応

可視化

NICE-DCV

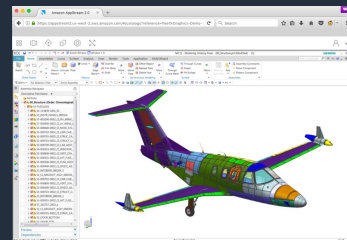


GPUアクセラレーションに対応し、インタラクティブなアプリケーションに適したVDI環境

Amazon AppStream 2.0



フルマネージドのアプリケーションストリーミングサービス



AWSでHPCを実行するメリット

大規模なインフラストラクチャによって
オンプレミスでは不可能なスケーラビリティと俊敏性を
実現できる

サーバーの長い調達サイクルや大きな投資なしで即座に
最新のテクノロジーを利用できる

柔軟な設定オプションにより、リソースの選択を
すばやく反復処理して、コストの最適化を実現できる



ROI の改善



結果達成までの
時間を短縮

AWSのHPC関連サービス概要

Amazon EC2で選択できるCPU&アクセラレーター

コンピュータ	ストレージ
ネットワーク	その他

CPU



INTEL Xeon Scalable
processors



AMD EPYC processor



AWS Graviton
Processor

Accelerator



H100 and T4
Tensor Core
GPUs



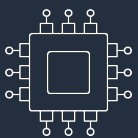
FPGAs for
custom
hardware
acceleration



AWS
Trainium
Inferentia



Radeon Pro
V520 GPU



Amazon EC2 (Elastic Compute Cloud)

コンピュータ	ストレージ
ネットワーク	その他

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 AWS Nitro System により、仮想化オーバーヘッドを極小化
- ワークロードに応じて様々なインスタンスタイプを選択可能

高性能計算向けインスタンスタイプの例

高性能 CPU の選択肢

アクセラレータの選択肢



Intel Xeon processor
(x86_64 arch)

AMD EPYC processor
(x86_64 arch)

AWS Graviton Processor
(64-bit Arm arch)

NVIDIA GPU

Xilinx FPGA

C6i インスタンス

Ice Lake

最大時全コア 3.5 GHz 駆動

R7iz インスタンス※

Sapphire Rapids

最大全コア 3.9 GHz 駆動

C6a インスタンス

EPYC Milan

最大 3.3 GHz 駆動

Hpc6a インスタンス

EPYC Milan

HPC特化

C7g インスタンス

64bit Arm Neoverse V1ベース
AWS Graviton3 CPU 搭載

P4d インスタンス

A100 GPU 搭載

P4de インスタンス※

A100 (80GB版) GPU 搭載

G5 インスタンス

A10G GPU 搭載

P5 インスタンス

H100 GPU 搭載

F1 インスタンス

Virtex UltraScale+
VU9P 搭載

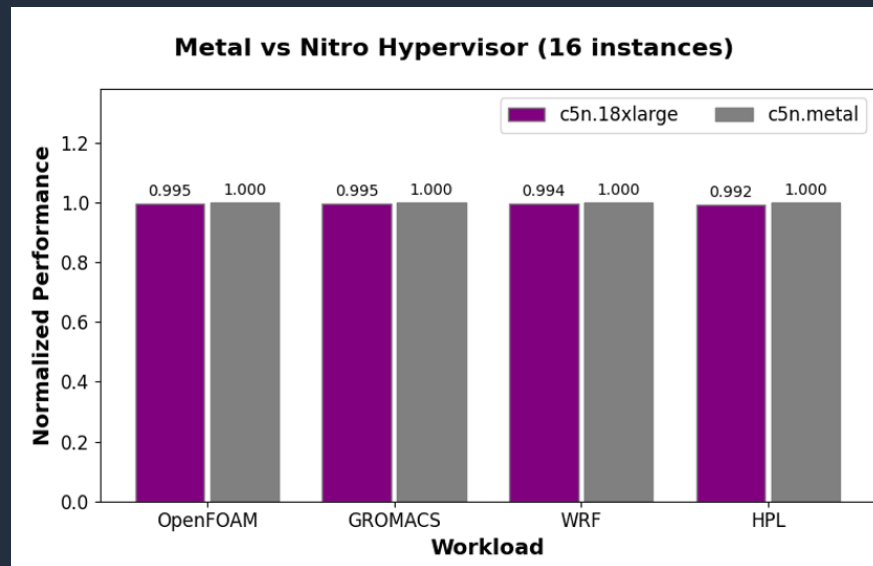
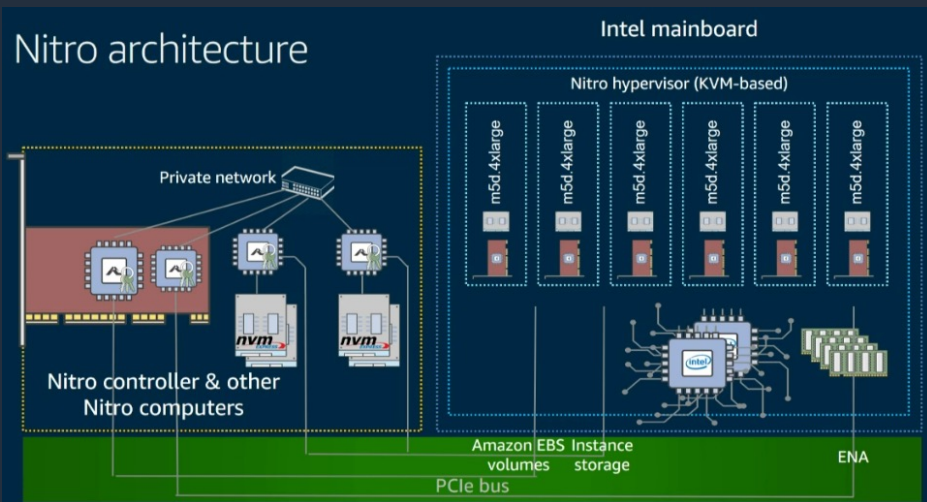
※ 2023年8月現在プレビュー提供中



EC2 の仮想化基盤である AWS Nitro System

コンピュータ	ストレージ
ネットワーク	その他

2017年に EC2 の新しい仮想化基盤である Nitro System を発表
仮想化のために必要な処理を、**外部の専用ハードウェアにオフロード**
仮想化オーバーヘッドによるパフォーマンス影響を極小化、セキュリティの向上



<https://aws.amazon.com/blogs/hpc/bare-metal-performance-with-the-aws-nitro-system/>



EFA: Elastic Fabric Adapter

コンピュート	ストレージ
ネットワーク	その他

MPI/NCCL で利用可能な低レイテンシネットワークアダプタ

c5n.18xlarge, c6gn.16xlarge, p4dn.24xlarge 等の EFA 対応インスタンスで利用可能

OpenMPI, Intel MPI, MVAPICH2 が対応

通信プロトコルとして、TCP/IP ではなく、独自の SRD (Scalable Reliable Datagram) を使用

インスタンス間通信速度

C6in

c6in.32xlarge
200 Gbps

C7gn

c7gn.16xlarge
200 Gbps

P4d

p4d.24xlarge
400 Gbps
(GPUDirect RDMA対応)

計算ノード間の通信レイテンシが重要となる処理例



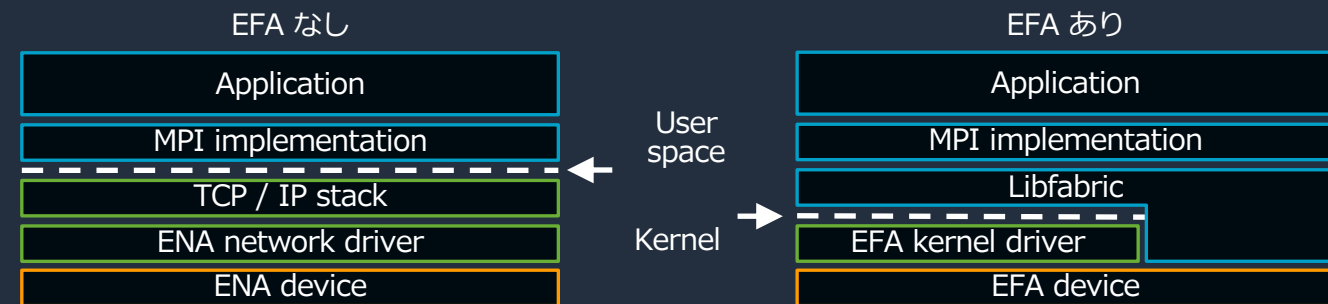
流体解析



地震シミュレーション



気象予測



Amazon FSx for Lustre

コンピュータ	ストレージ
ネットワーク	その他

HPC向け高速ファイルシステムのフルマネージドサービス



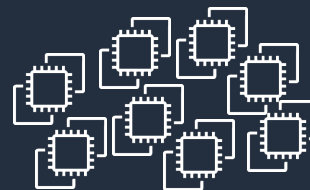
スケラブルで
高いパフォーマンス

並列ファイルシステム



毎秒数百 GB 単位のスループットと
数百万単位 IOPS までスケール
SSDベースでミリ秒未満、HDDベ
ースで1桁ミリ秒の低レイテンシー

複数のストレージオプション



数千のコンピューティング
インスタンスからの
同時アクセスをサポート

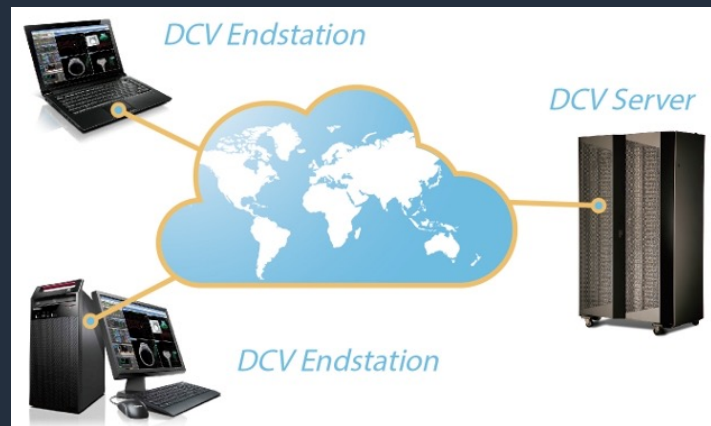
HDDタイプではストレージの 1 TB あたり 12 MB/秒~40 MB/秒 のスループット
SSDタイプではストレージの 1 TB あたり 50 MB/秒~ 200MB/秒 のスループットを提供

<https://aws.amazon.com/jp/fsx/lustre/>



クラウド上での可視化処理を実現する高性能VDIツール

- 高性能なNICE DCVプロトコルにより、クラウド上のデスクトップ画面を高速にストリーミング、スムーズな画面描画による快適なVDI環境を実現
- Windows、Linuxの両方に対応
- GPUインスタンスと組合せて使用することで、サーバー側に搭載されたGPUによる高度な3Dグラフィックス描画の他、H.264ハードウェアエンコードによりCPU負荷無く画面のストリーミングを実現 (non-GPU環境ではCPU処理で動作)
- HTML5クライアントサポートによりブラウザでのリモートアクセスにも対応
- Amazon EC2で利用する場合は追加のライセンスコスト無しで利用可能



<https://aws.amazon.com/hpc/dcv/>

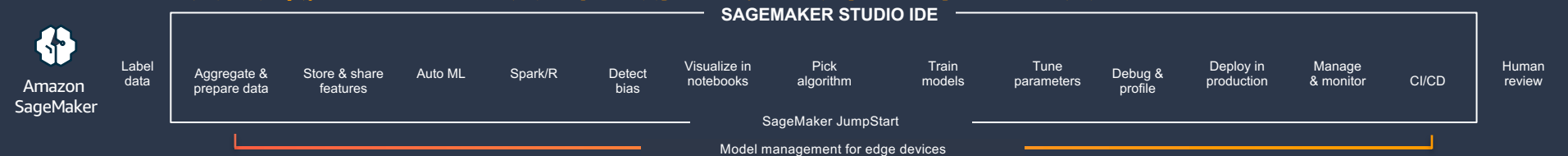
AWS の AI/機械学習関連サービス

コンピュー	ストレージ
ネットワーク	その他

AI サービス: 機械学習の深い知識なしに利用可能

HEALTH AI Amazon HealthLake Amazon Transcribe Medical Amazon Comprehend Medical			INDUSTRIAL AI AWS Panorama + Appliance Amazon Monitron Amazon Lookout for Equipment Amazon Lookout for Vision			ANOMALY DETECTION Amazon Lookout for Metrics		CODE AND DEVOPS Amazon DevOps Guru Amazon CodeGuru		
VISION Amazon Rekognition	SPEECH Amazon Polly Amazon Transcribe <small>+Medical</small>		TEXT Amazon Comprehend <small>+Medical</small> Amazon Translate Amazon Textract		SEARCH Amazon Kendra	CHATBOTS Amazon Lex	PERSONALIZATION Amazon Personalize	FORECASTING Amazon Forecast	FRAUD Amazon Fraud Detector	CONTACT CENTERS Contact Lens Voice ID <small>For Amazon Connect</small>

ML サービス: 機械学習のプロセス全体を効率化するマネージドサービス



ML フレームワークとインフラストラクチャ: 機械学習の環境を自在に構築して利用

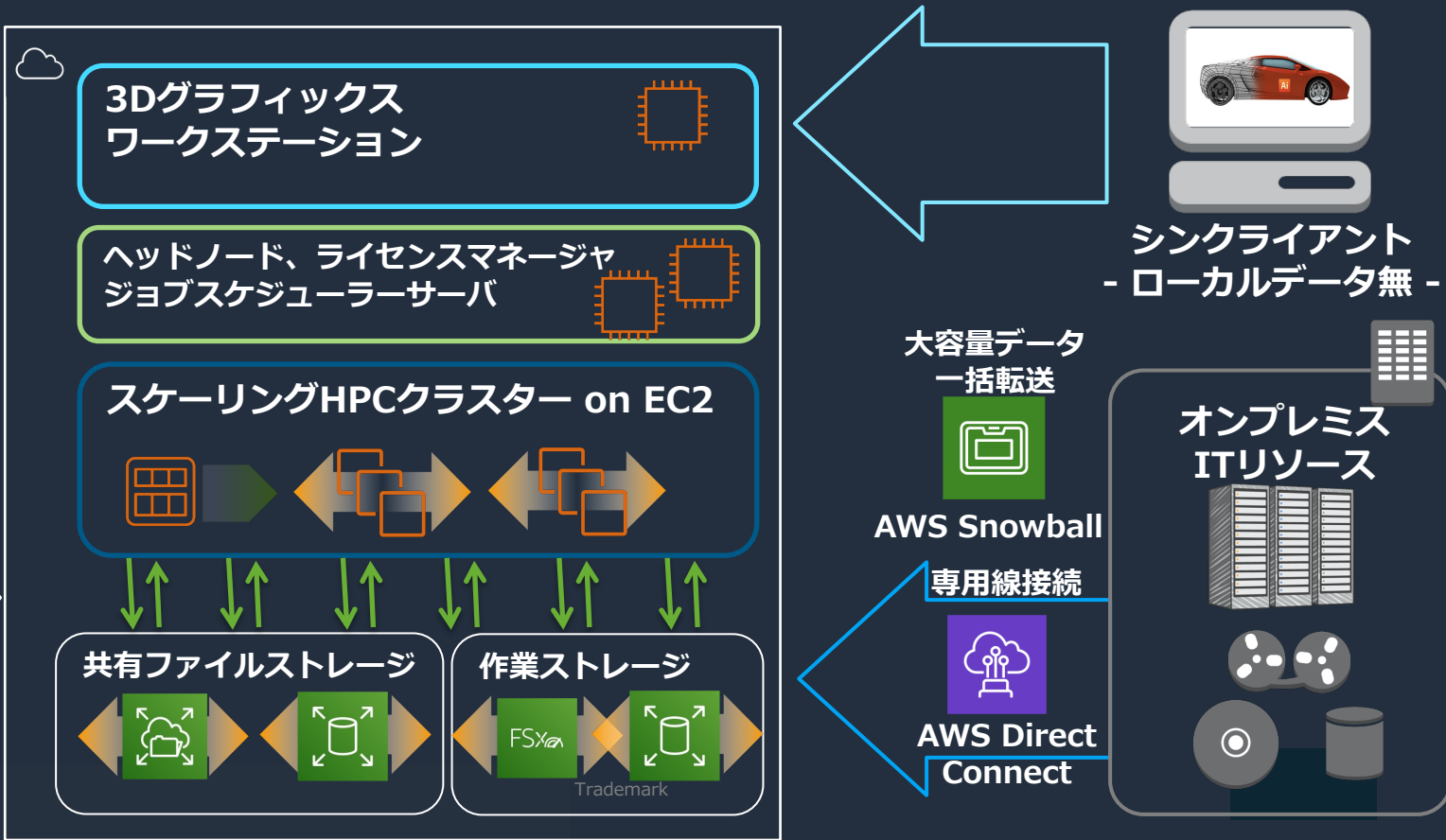
TensorFlow	mxnet	GLUON	Keras	Deep Learning AMIs & Containers	GPUs & CPUs	Elastic Inference	Trainium	Inferentia	FPGA
PyTorch	light.learn	mistral	DeepGraphLibrary						



HPCオーケストレーションサービス

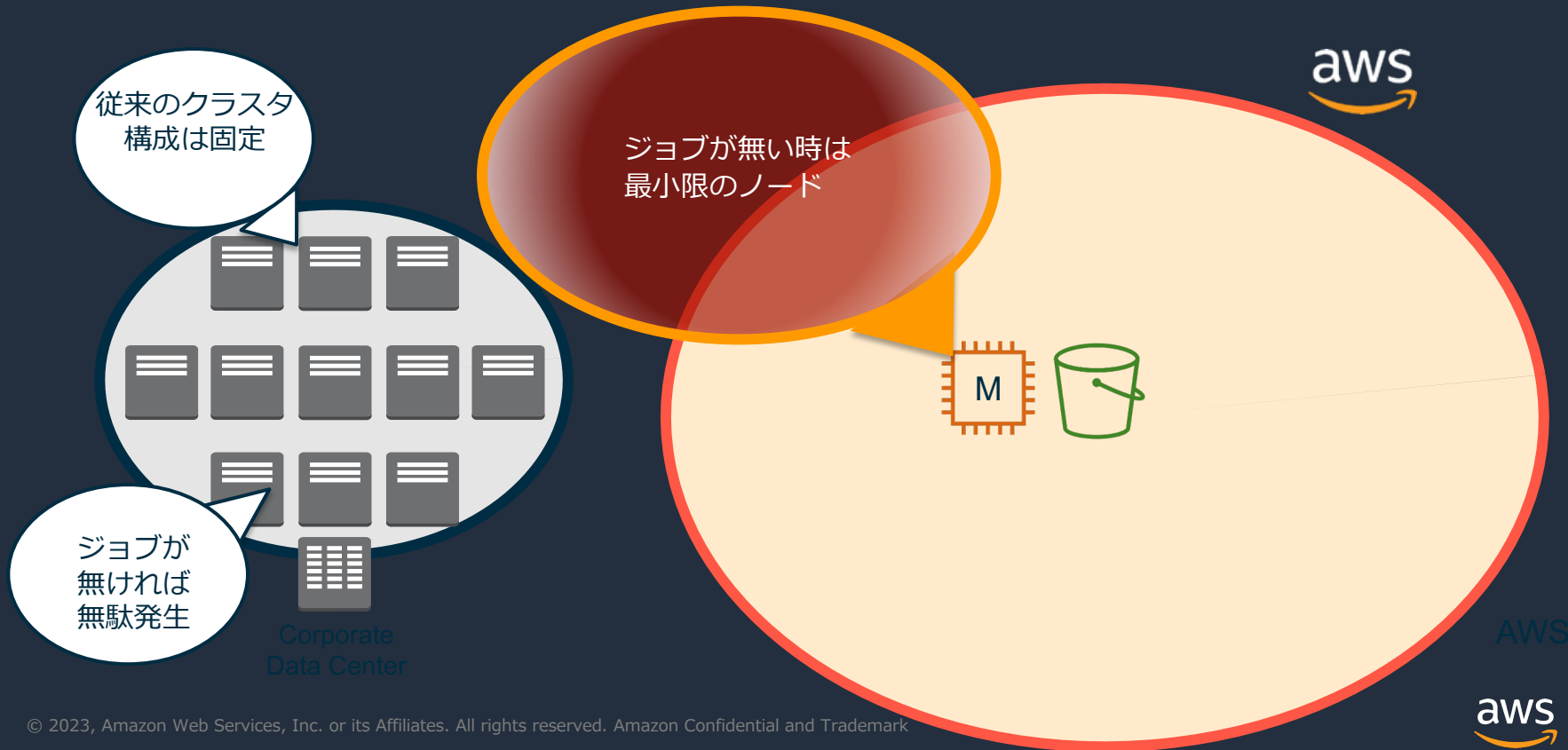
クラウドHPCのリファレンスアーキテクチャ

クラウドHPCの環境→常時システムをスクラップ&ビルドしてベストコンディションで利用



クラウドHPCのあるべき姿

処理にあわせてオートスケールさせることで常にシステムを最適な状態に保つ



クラウドHPCのあるべき姿

処理にあわせてオートスケールさせることで常にシステムを最適な状態に保つ

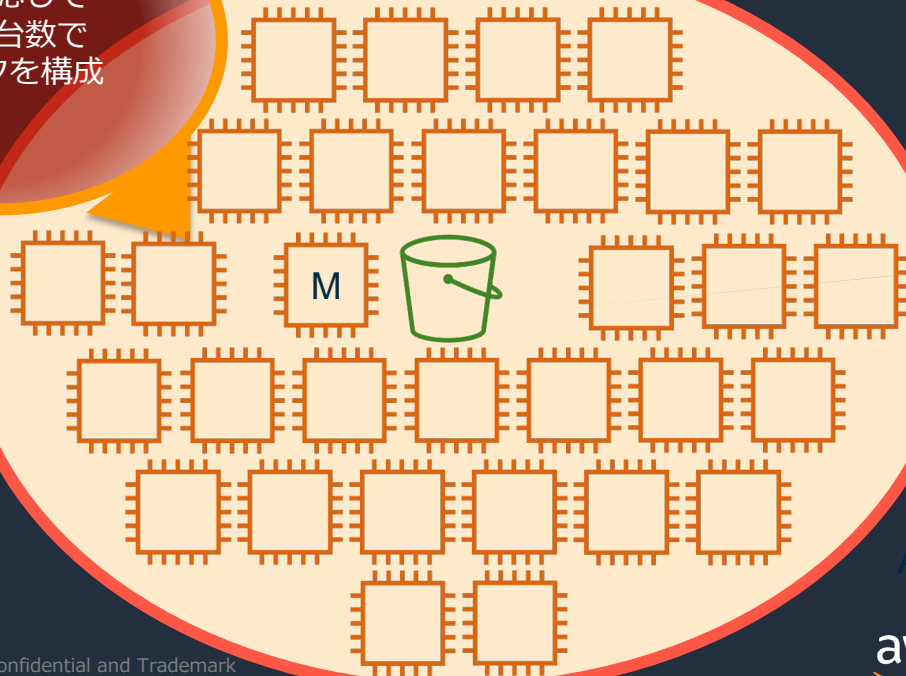
従来のクラスタ
構成は固定



ジョブが
無ければ
無駄発生

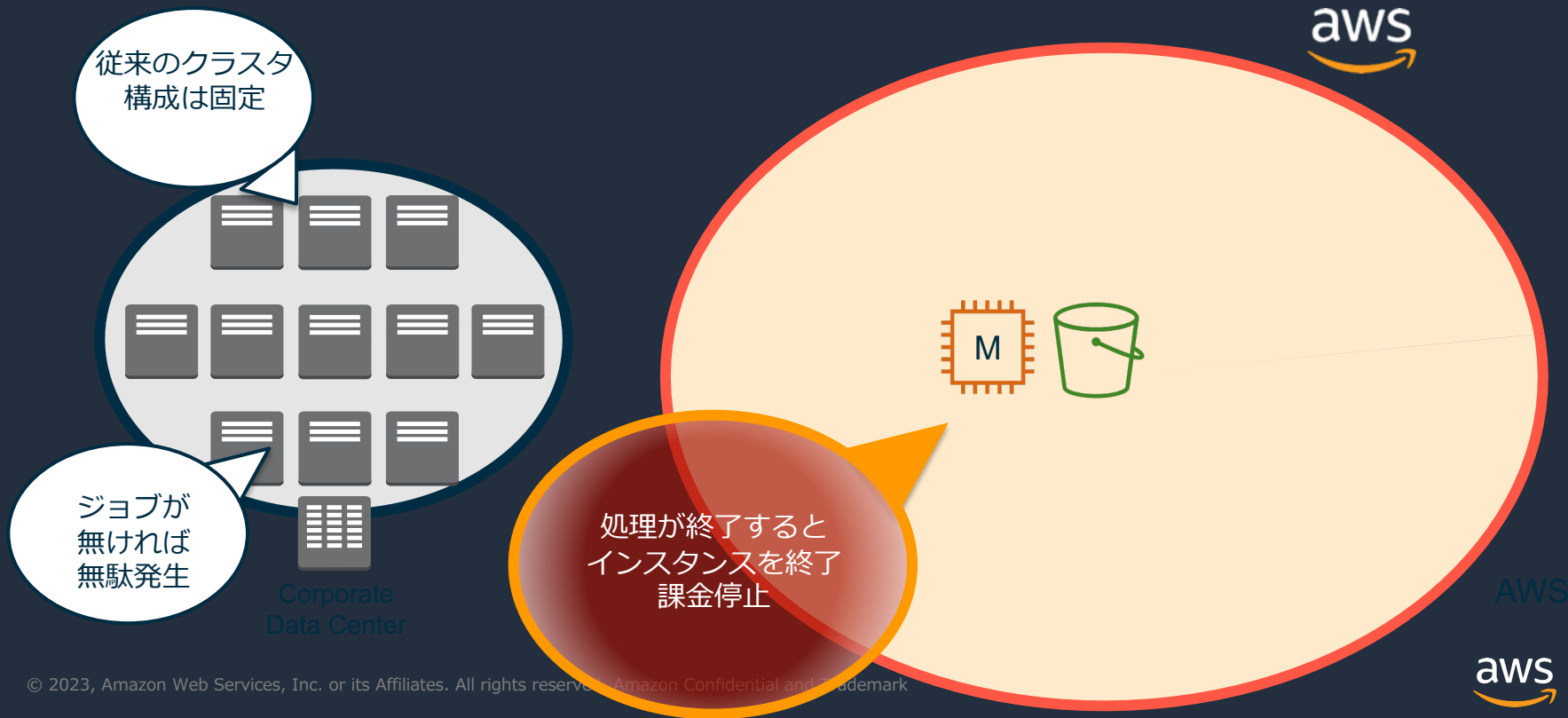
Corporate
Data Center

必要に応じて
必要な台数で
クラスタを構成



クラウドHPCのあるべき姿

処理にあわせてオートスケールさせることで常にシステムを最適な状態に保つ



AWS上にオートスケールするHPCクラスタを構成する方法1

AWSのバッチコンピューティングサービスを利用する



AWS ParallelCluster

- 一般的なジョブスケジューラの **Slurm** に対応し、オンプレミスのHPC利用者が馴染みやすい環境
- 単一のジョブが大量の CPU core を使用する密結合ワークロードに向いている
- **Pros:** 既存の HPC クラスタからの移行が容易
- **Cons:** 単一のAZとなるため、可用性確保には工夫が必要



AWS Batch

- フルマネージドの**独自ジョブスケジューラ**
- 少数の CPU core を使用するジョブを大量に実行する疎結合・High Throughput Computing ワークロードに向いている
- **Pros:** フルマネージドサービスでありコンテナさえ用意すれば計算基盤の管理は不要
- **Pros:** 複数 AZ や複数インスタンスタイプの利用が可能であり、可用性を確保しやすい
- **Cons:** コンテナ化やスケジューラ対応が必要

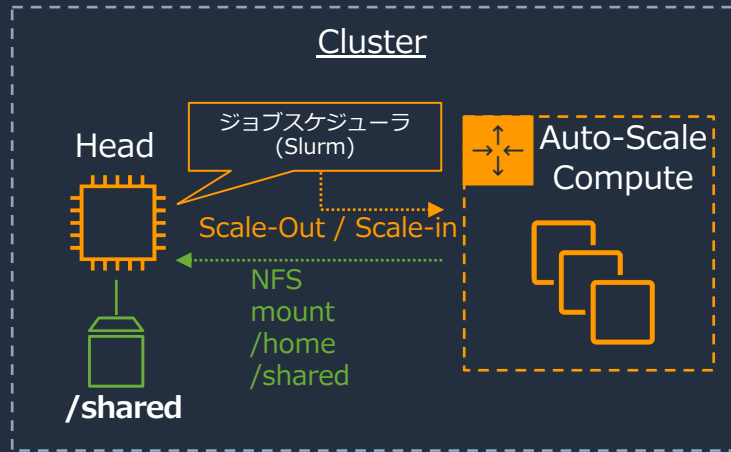
AWS ParallelCluster の概要

数コマンド操作でジョブ投入に応じて自動でスケールするクラスタを
AWS 上に構築可能な AWS 公式のオープンソースソフトウェア

AWS ParallelCluster の特徴

- 既存の HPC 向けスケジューラ (Slurm) と連動しスケラブルな計算環境を作成
- MPI/NCCL 環境がセットアップ済み
- x86/Arm 両方の環境に対応
- 使用するOSやネットワーク環境、ストレージ構成などを柔軟にカスタマイズ可能
- コマンド / グラフィカルのインターフェース
- オープンソースソースのプロジェクトであり、誰でもソースコードを入手可能

<https://github.com/aws/aws-parallelcluster>

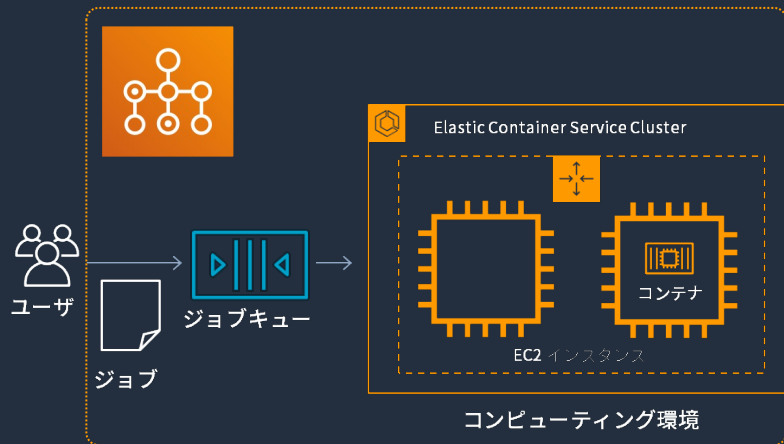


AWS Batch の概要



大規模バッチ処理のため環境をフルマネージドで提供

- AWS Batch がインスタンスの起動や停止を行うため、スケジューラや計算ノードなどの **管理が不要**
- ジョブは **Docker コンテナイメージ** を元に作成し、自動でスケールするコンピューティング環境で実行する
- コンピューティング環境ではインスタンスタイプや vCPU 数、スポットインスタンス利用有無などを任意に指定可能



コンテナイメージを用意するだけでスケラブルな大規模バッチ処理環境が得られる

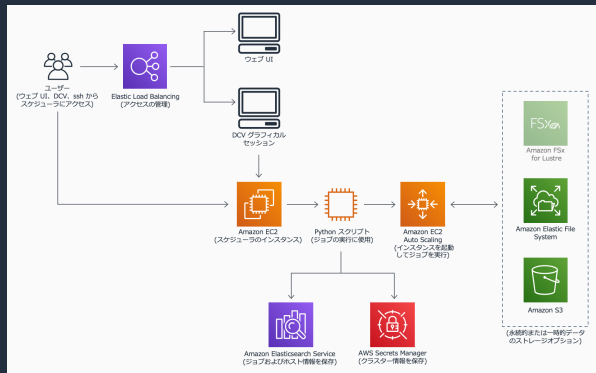
AWS上にオートスケールするHPCクラスタを構成する方法2

AWS Solutionsを活用する

AWS上でのHPCリファレンスデザインとして、AWS Solutions: Scale-out Computing on AWS が公開中。CloudFormationテンプレートで簡単に展開が可能

<主な機能>

- Web フロントエンド
- NICE-DCV 用ジョブ作成による可視化、Pre/Post 処理
- LDAPによるユーザー管理
- ジョブスケジューラ(PBS)からAWSのリソースを指定する方法
- AWS Budget とジョブスケジューラを組み合わせた予算管理
- Amazon Elasticsearch Service を使ったコスト可視化
- Amazon SES連携によるジョブのメール通知



Scale-out Computing on AWS をそのまま使用することも可能だが実装のアイデア・サンプルコードとして用いることで要件に応じたクラスタ環境のリファレンスデザインとしても利用可能

<https://aws.amazon.com/jp/solutions/scale-out-computing-on-aws/>



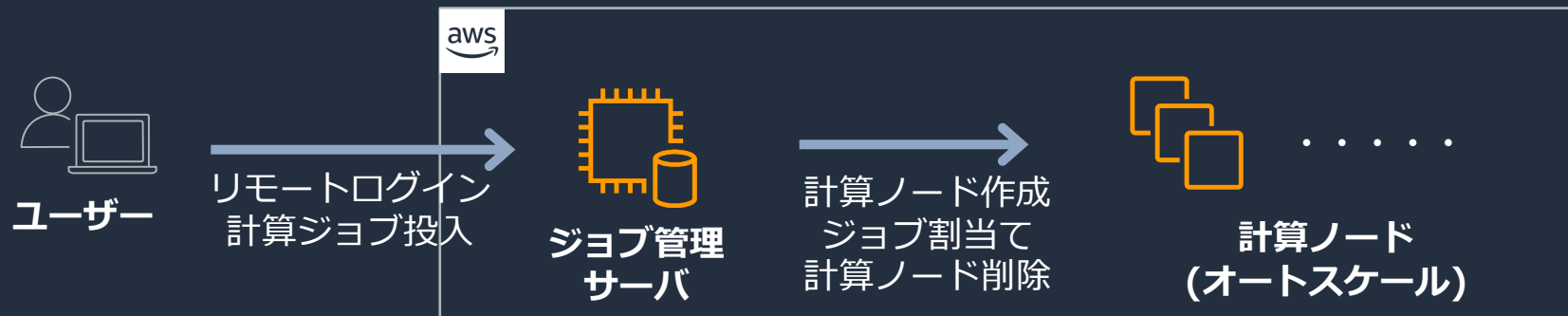
AWS上にオートスケールするHPCクラスタを構成する方法3

汎用ジョブスケジューラの機能を利用する

汎用のジョブスケジューラが持つ、プラグインやカスタムスクリプトを使用して、AWSにバーストしたり、動的に変化するHPCクラスタ環境を管理する

AWS上に動的リソースを管理する機能を持たせられるジョブスケジューラの例

- IBM LSF – resource connector
- Univa (現Altair) Grid Engine – Navops
- SchedMD slurm – power save logic



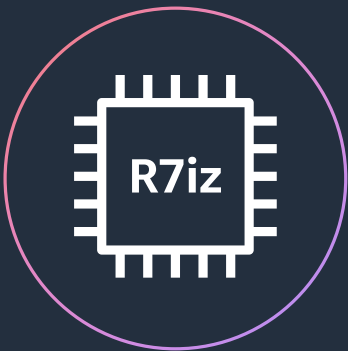
HPC on AWS 最近のアップデート

Amazon EC2 R7iz インスタンス

PREVIEW

高クロック周波数のCPUを搭載したメモリ最適化インスタンス

- 第4世代 Intel Xeon Scalable Processor(Sapphire Rapids) を搭載し、最大 3.9GHz で動作
- 最大 128vCPU、1,024GiB の DDR5 メモリ。最大 50Gbps のネットワーク帯域
- 高クロック周波数の CPU により、半導体設計ワークロード、金融リスク分析など疎結合タイプのジョブを高速に実行



	z1d.12xlarge	m5zn.12xlarge	x2iezn.12xlarge	r7iz.32xlarge
vCPU数	48	48	48	128
最大クロック周波数	4.0 GHz	4.5 GHz	4.5 GHz	3.9 GHz
RAM	384 GiB (8GB/vCPU)	192 GiB (4GB/vCPU)	1536 GiB (32GB/vCPU)	1024 GiB (8GB/vCPU)
ネットワーク帯域幅	25 Gbps	100 Gbps	100 Gbps	50 Gbps
EFA 対応	なし	あり	なし	なし
価格※	\$ 4.464 USD/h	\$3.9641 USD/h	\$10.008 USD/h	TBD



Amazon EC2 HPC6a インスタンス

Amazon EC2で実行されるHPCワークロードに対して最高の価格パフォーマンスを提供するように設計



AMD

AMD 第3世代 EPYC Milanプロセッサを 96コア、および 384GB のメモリを搭載

同等の AmazonEC2 x86 ベースのコンピューティング最適化インスタンスよりも最大 65% 優れた価格パフォーマンス、**1物理コアあたり \$0.03/ 時間***の低コストで提供

100 Gbps のネットワーク帯域幅、EFA 対応

気象、金融、流体力学、ヘルスケアおよびライフサイエンスなど大規模な計算ワークロードに最適

オハイオ、ストックホルム、GovCloud (米国西部および東部) で提供

* Ohio リージョンにおける 2023 年 7 月時点のオンデマンド価格

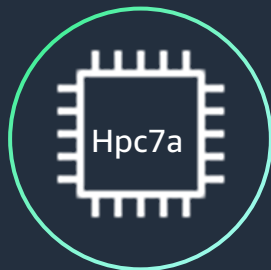
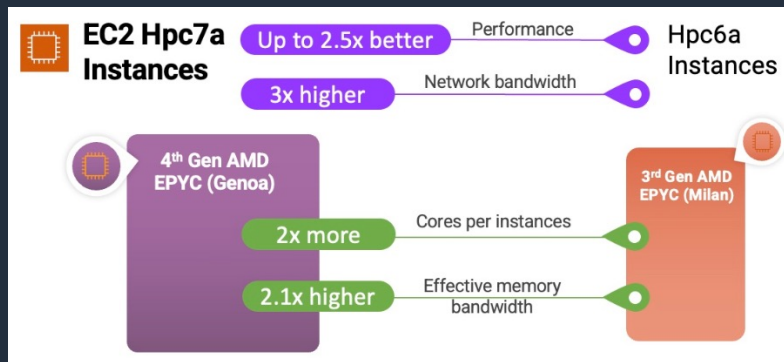
<https://aws.amazon.com/jp/blogs/news/new-amazon-ec2-hpc6a-instance-optimized-for-high-performance-computing/>



Amazon EC2 HPC7a インスタンス

AVAILABLE

- AMD 第4世代 EPYC Genoaプロセッサを最大192コア、および768GBのDDR5メモリを搭載。300GbpsのEFAネットワークに対応
- 前世代のHPC6aインスタンスと比較して、コアとメモリの搭載量を最大2倍、有効メモリバンド幅が2.1倍
- 気象予測、流体解析等のワークロードに最適
- 1物理コアあたり \$0.0375/時間*
- Ohio, Ireland, GovCloud リージョンにて2023年8月より一般提供開始



AMD

Instance	vCPU	memory	EFA Network BW	On-Demand Hourly rate
hpc7a.12xlarge	24	768GB	Up to 300Gbps	\$7.20
hpc7a.24xlarge	48			
Hpc7a.48xlarge	96			
hpc7a.96xlarge	192			

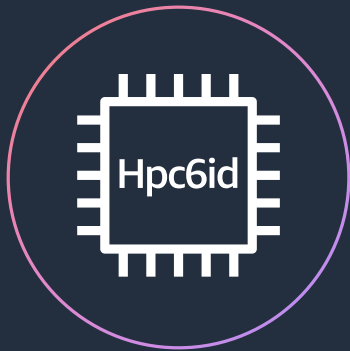
<https://aws.amazon.com/blogs/aws/new-amazon-ec2-hpc7a-instances-powered-by-4th-gen-amd-epyc-processors-optimized-for-high-performance-computing/>



Amazon EC2 HPC6id インスタンス

AVAILABLE

メモリ・データ集約的な HPC ワークロードに適したインスタンス

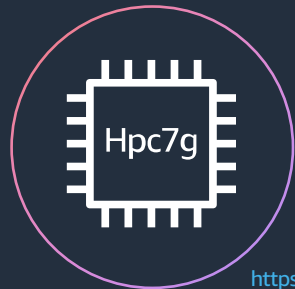
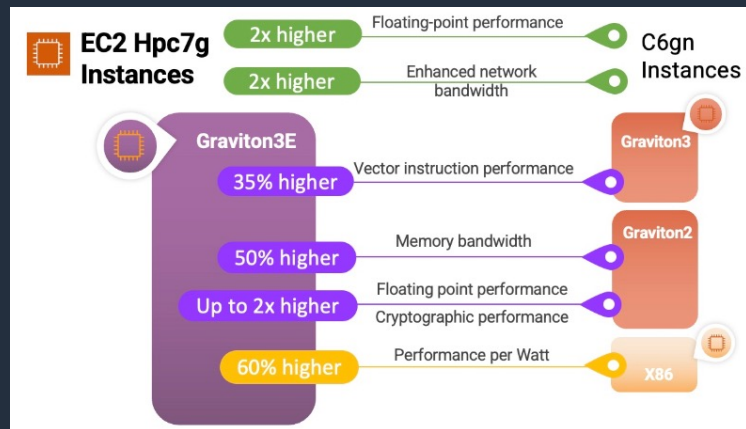


- 第三世代 Intel Xeon Scalable Processor (Ice Lake) を搭載し、最大 3.5GHz で動作。64 物理コア、1024GiB メモリ、15.2TB の SSD を搭載。200Gbps の EFA ネットワークに対応
- 有限要素法解析のような広いメモリバンド幅が求められるワークロードに最適
- 1物理コアあたり \$0.089/時間*
- オハイオ、フランクフルト、GovCloud (米国西部) で一般利用開始

Amazon EC2 HPC7g インスタンス

AVAILABLE

- AWS **Graviton3E** 搭載、最新の Nitro v5 プラットフォームによる第二世代 EFA 採用し **200 Gbps のネットワーク帯域**
- 前世代のC6gn インスタンスと比較して**2倍の浮動小数点演算性能**、同等のx86インスタンスと比較して最大**60%高いエネルギーあたりの計算性能**
- 気象予測、流体解析、金融リスク計算等のワークロードに最適
- N.Virginia リージョンにて2023年6月より一般提供開始



Instance	vCPU	memory	CPU Freq	EFA Network	On-Demand Hourly rate
hpc7g.4xlarge	16	128GB	2.6GHz	200Gbps	\$1.6832
hpc7g.8xlarge	32				
hpc7g.16xlarge	64				

<https://aws.amazon.com/blogs/aws/new-amazon-ec2-hpc7g-instances-powered-by-aws-graviton3e-processors-optimized-for-high-performance-computing-workloads/>



まとめ

まとめ

- HPC on AWSではAWSクラウドが持つ多くのサービスとスケーラビリティを組み合わせることで利用効率を最適化できる
- AWSでは高性能なHPC環境を実現するコンピューティングサービスと、その利用を支援する周辺サービスも拡充している
- サービスは日々進化しており、稼働中のシステムにも最新サービスを適用させることで最新環境の維持とコスト低減が行える

ご清聴ありがとうございました