



夏の AMAZON EC2 祭り 2023 最新インスタンス活用編

まずは基本のおさらい！ Amazon EC2 の概要と インスタンスタイプの選び方

Daisuke Miyamoto

Sr. Specialist Solutions Architect, Compute/HPC

Amazon Web Services Japan G.K.

2023/07/20

宮本 大輔, Ph.D.

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
シニア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



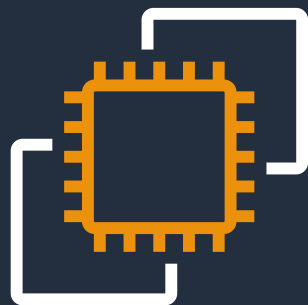
夏の Amazon EC2 祭り！

最新 CPU の話から実際の活用まで
多くのセッションがありますが
その前に基本のおさらいから入りましょう！

Amazon Elastic Compute Cloud (EC2)

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※の従量課金で利用可能
- 独自の Nitro System により仮想化の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



AMAZON EC2



Amazon EC2 の成り立ち

2006年に Beta 提供開始

選択可能なスペックは一種類

AWS News Blog

Amazon EC2 Beta

by Jeff Barr | on 25 AUG 2006 | [Permalink](#) | [Share](#)

▶ 0:00 / 0:00



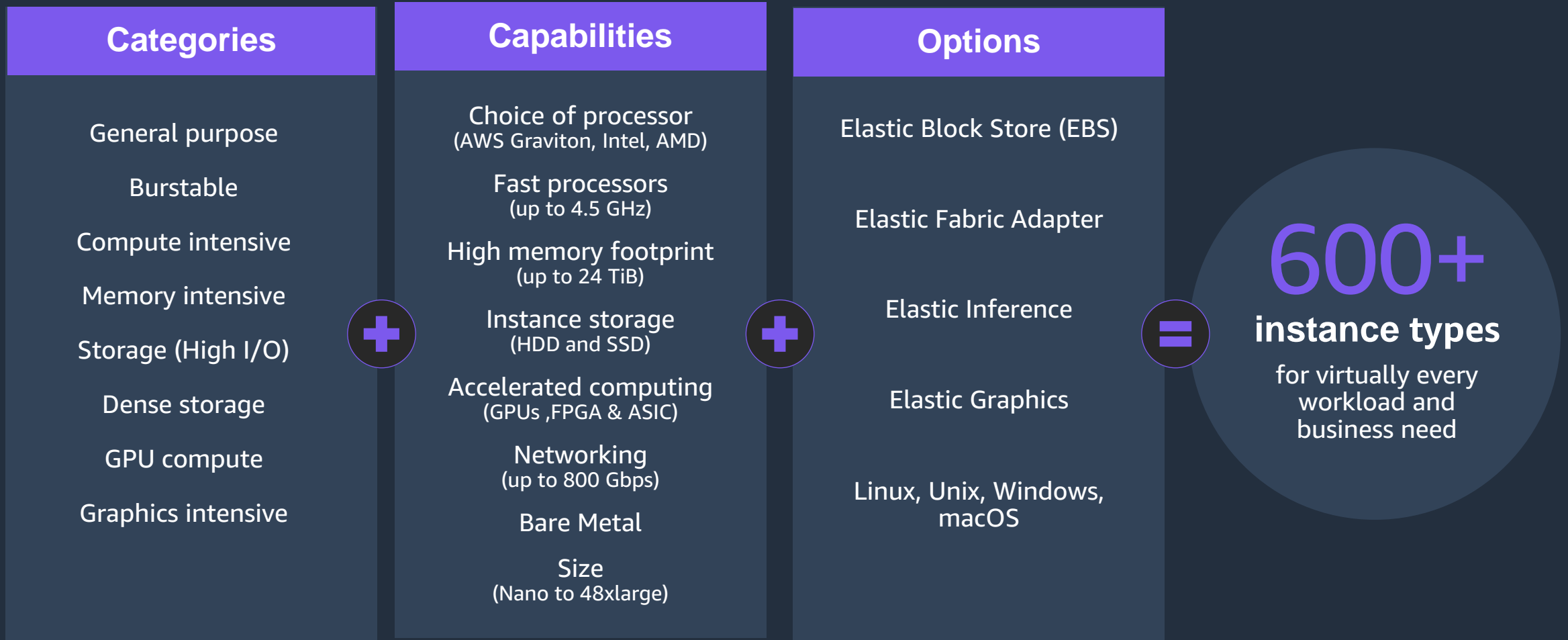
Voiced by Amazon Polly

Innovation never takes a break, and neither do I. From the steaming hot beaches of Cabo San Lucas I would like to tell you about the Amazon Elastic Compute Cloud, or Amazon EC2, now open for limited beta testing, with more beta slots to open soon.

Amazon EC2 gives you access to a virtual computing environment. Your applications run on a “virtual CPU”, the equivalent of a 1.7 GHz Xeon processor, 1.75 GB of RAM, 160 GB of local disk and 250 Mb/second of network bandwidth. You pay just 10 cents per clock hour (billed to your Amazon Web Services account), and you can get as many virtual CPUs as you need. You can learn more on the [EC2 Detail Page](#). We built Amazon EC2 using a virtual machine monitor by the name of [Xen](#).



Amazon EC2 インスタンスタイプの選択肢



沢山あるインスタンスタイプから
どうやって適したものを
選んだらいいんだろう？



本日のテーマ

Amazon EC2 のスペックを定義するインスタンスタイプには
どのような種類があって、どう選択したらよいのか

本日の内容

- インスタンスタイプの読み方
- インスタンスファミリーの種類と選択
- インスタンスタイプ最適化のポイント

インスタンスタイプの読み方

インスタンスタイプの読み方

インスタンス世代

c7gn.xlarge

インスタンス
ファミリー

追加機能

インスタンス
サイズ

インスタンスタイプ

https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/UserGuide/instance-types.html



インスタンスファミリー (追加機能)

c7gn.xlarge

インスタンス世代

インスタンスサイズ

Amazon EC2 インスタンスファミリー

メモリ・I/O・CPUクロック重視、GPU・FPGA搭載などの種類を表す多くの種類があるが、5つのカテゴリに分類することが可能

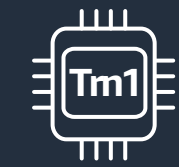
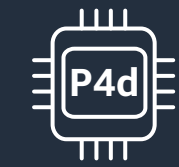
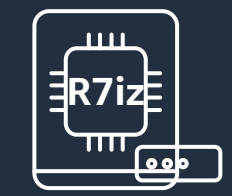
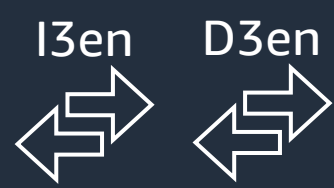
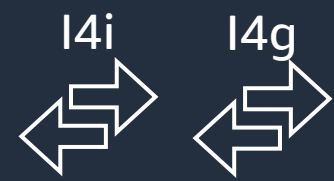
汎用

コンピューティング最適化

ストレージ最適化

メモリ最適化

高速コンピューティング (アクセラレータ)



インスタンスファミリー (追加機能)
 c7gn.xlarge
 インスタンス世代 インスタンスサイズ

Amazon EC2 インスタンス世代

同じインスタンスファミリーでも世代が進むにつれ数字が大きくなる
 世代が新しい方が高性能でコストパフォーマンスも良い

例)コンピューティング最適化インスタンスの変遷

旧世代

新世代

	CC1	CC2	C3	C4	C5	C6i
最大vCPU	16	32	32	36	72	128
最大RAM (GiB)	23	60.5	60	60	144	256
CPU	Xeon X5570 (Nehalem)	Xeon E5-2670 (Sandy Bridge)	Xeon E5-2680v2 (Ivy Bridge)	Xeon E5-2666v3 (Haswell)	Xeon Platinum 8000 (Skylake or Cascade Lake)	Xeon Platinum 8375C (Ice Lake)
最大 Network	10 Gbps	10 Gbps	10 Gbps	10 Gbps	25 Gbps	50 Gbps
Launch Date	Jul, 2010	Nov, 2011	Nov, 2013	Jan, 2015	Nov, 2017	Oct, 2021



Amazon EC2 インスタンスの追加機能 (CPU タイプ)

第 6 世代からは全てのインスタンスタイプで CPU タイプを表す記号 (i / a / g) が付与

Intel Xeon 搭載



AMD EPYC 搭載



AWS Graviton 搭載



Amazon EC2 インスタンスの追加機能

d: 標準インスタンスに対して内蔵ストレージ(インスタンスストア)付加



n: 標準インスタンスに対してネットワークを強化



z: 標準インスタンスに対して高い CPU クロック



b: 標準インスタンスに対して EBS 帯域を強化



インスタンスファミリー (追加機能)

c7gn.xlarge

インスタンス世代

インスタンスサイズ

Amazon EC2 インスタンスサイズ

インスタンスの大きさを表す

large → xlarge → 2xlarge と進むに従って、
vCPU 数だけでなくメモリ・ネットワーク帯域も増加する

例) C5インスタンスのサイズとキャパシティ

タイプ	vCPU	メモリ	EBS帯域幅	NW帯域幅
c5.large	2	4GiB	最大2.25Gbps	最大10Gbps
c5.xlarge	4	8GiB	最大2.25Gbps	最大10Gbps
c5.2xlarge	8	16GiB	最大2.25Gbps	最大10Gbps
c5.4xlarge	16	32GiB	2.25Gbps	最大10Gbps
c5.9xlarge	36	72GiB	4.5Gbps	10Gbps
c5.18xlarge	72	144GiB	9Gbps	25Gbps



インスタンスタイプの読み方（再掲）

インスタンス世代

c7gn.xlarge

インスタンス
ファミリー

追加機能

インスタンス
サイズ

インスタンスタイプ

https://docs.aws.amazon.com/ja_jp/AWSEC2/latest/UserGuide/instance-types.html



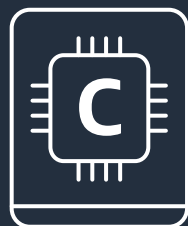
インスタンスファミリーの種類と選択

Mファミリーを基準に、ワークロードに合わせて選択

バーストパフォーマンス



コンピュート最適化
2 GiB/vCPU



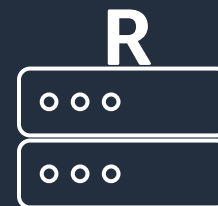
メモリは少なくてもよい

汎用
4 GiB/vCPU



より大きいメモリ量

メモリ最適化
8 GiB/vCPU

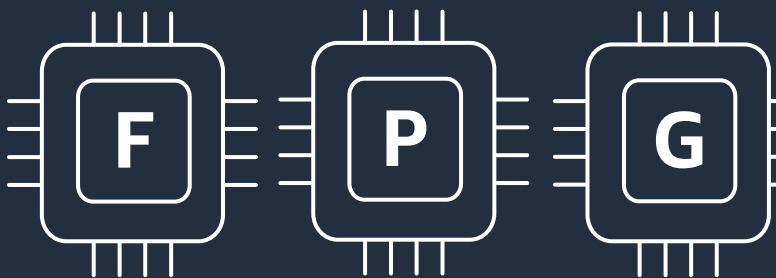


更に大きいメモリ量

メモリ最適化



High Memory



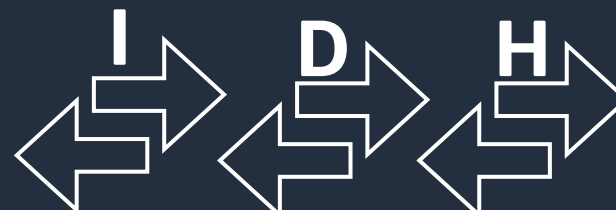
アクセラレーテッド

アクセラレータ
利用

インスタンスストア付加

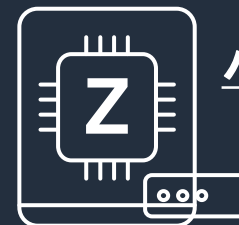


高速大容量
ストレージ



ストレージ最適化

ハイクロック



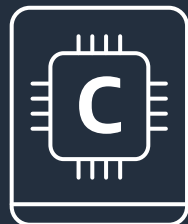
定常的なCPU性能は不要

高いCPUコアあたりの性能



メモリ要件に合わせて C / M / R の中から選択

コンピューター最適化
2 GiB/vCPU



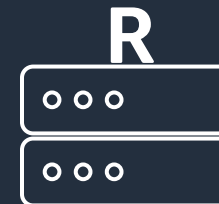
←
メモリは少なくてもよい

汎用
4 GiB/vCPU



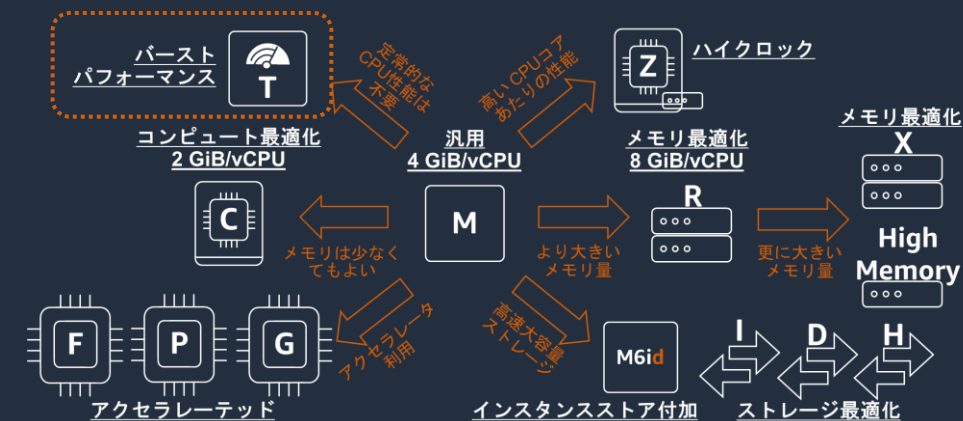
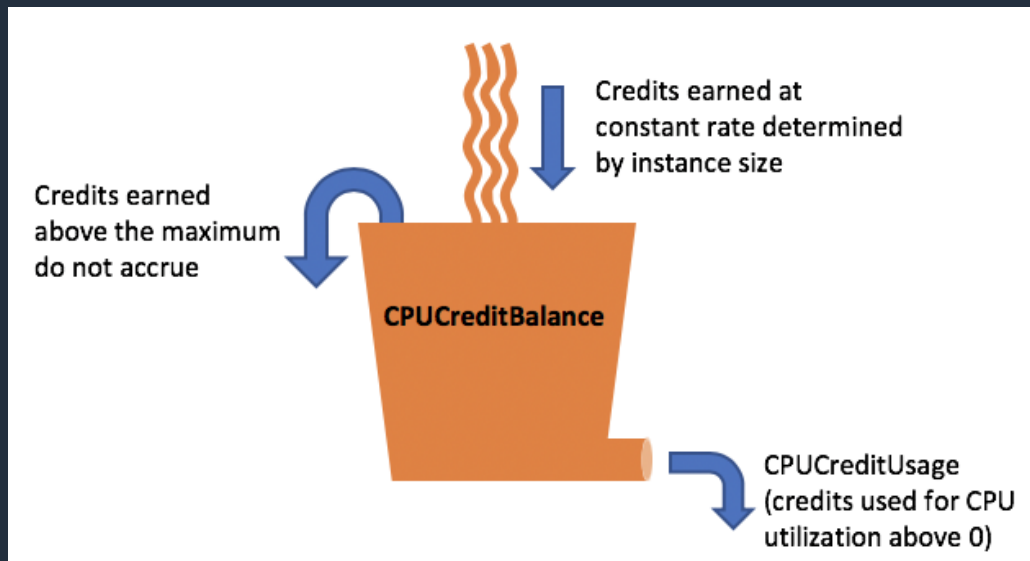
→
より大きいメモリ量

メモリ最適化
8 GiB/vCPU



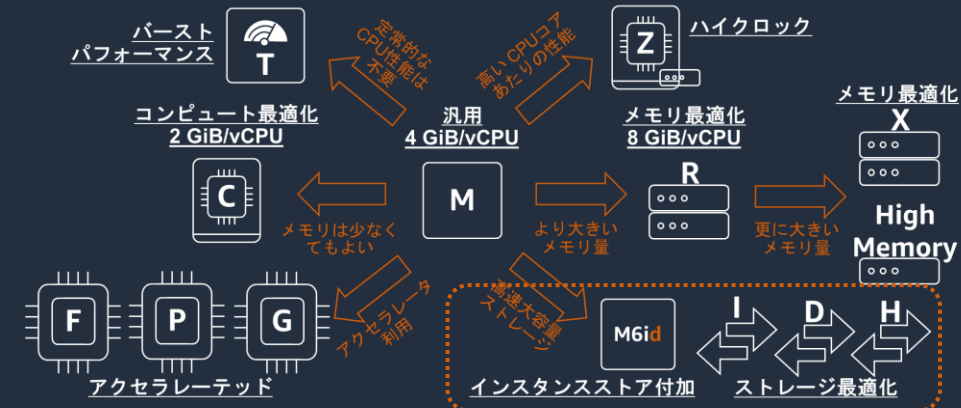
バーストパフォーマンス

- クレジット方式でバーストする CPU 性能を提供
- 通常 CPU 使用率が低く瞬間的に負荷がかかるワークロードに適するが定常的に負荷がかかるワークロードには向かない
- T4g / T3 / T3a 等



ストレージ最適化 / インスタンスストア付加

- ホストコンピュータ上に接続された「インスタンスストア」を提供し、高速なディスクアクセスが可能
- インスタンスストアはインスタンス停止時には消去されるため注意
- インスタンスストアは **I / H / D** といったストレージ最適化ファミリーのほか、M6id 等 **d** が付いたインスタンスタイプでも利用可能
- I4i/Im4gn/Is4gen では、新たに設計されたインスタンスストアである Nitro SSD を搭載し、特にレイテンシ変動を低減



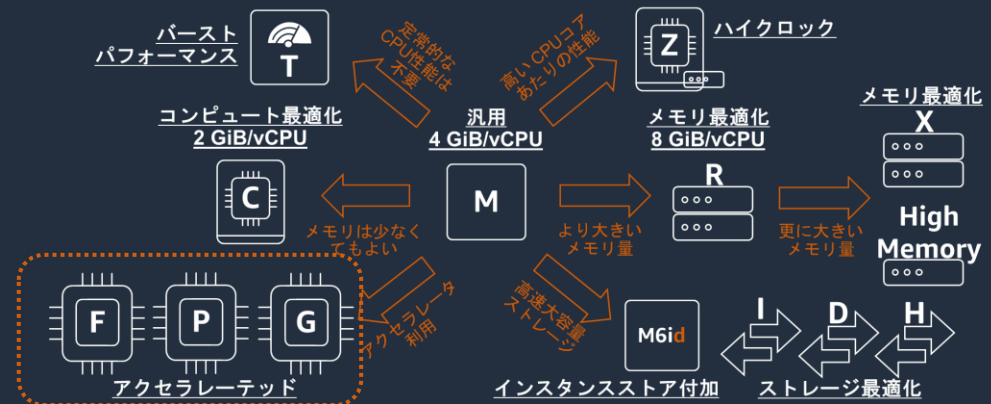
高速コンピューティング (1)

ワークロード例：機械学習（トレーニング）

- P5 (Preview) : NVIDIA H100 Tensor Core GPU
- P4d: NVIDIA A100 Tensor Core GPU
- Trn1: AWS Trainium
- DL1: Intel Habana Labs Gaudi アクセラレーター

ワークロード例：機械学習（推論）

- G5: NVIDIA A10G Tensor Core GPU
- Inf2: AWS Inferentia2



夏の Amazon EC2 祭り 2023 最新インスタンス活用 - 機械学習編 (8月4日)

https://pages.awscloud.com/eib-compute-230804-reg.html?trk=aws_event_page



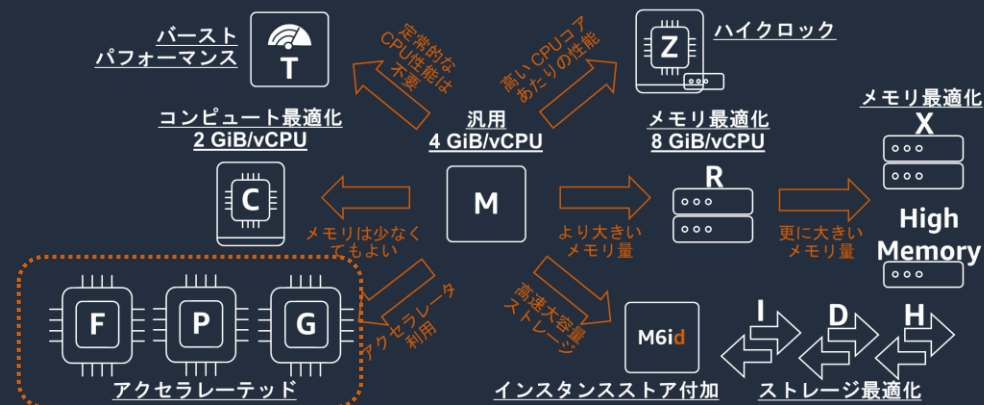
高速コンピューティング (2)

ワークロード例 : グラフィックス

- G5g: NVIDIA T4G Tensor Core GPU
- G4ad: AMD Radeon Pro V520 GPU

ワークロード例 : 動画処理

- VT1: Xilinx U30 メディアアクセラレーターカード
- F1: Xilinx Virtex UltraScale+ VU9P FPGA



その他特殊なインスタンスタイプ

- Mac インスタンス
macOS 専用のインスタンス
ベアメタルのみ提供
- HPC 最適化インスタンス
High Performance Computing (HPC) に
特化したインスタンス

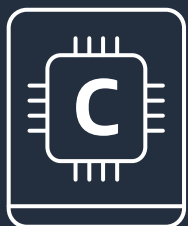


Mファミリーを基準に、ワークロードに合わせて選択

バーストパフォーマンス



コンピュート最適化
2 GiB/vCPU



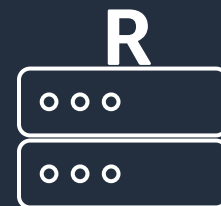
メモリは少なくてもよい

汎用
4 GiB/vCPU



より大きいメモリ量

メモリ最適化
8 GiB/vCPU

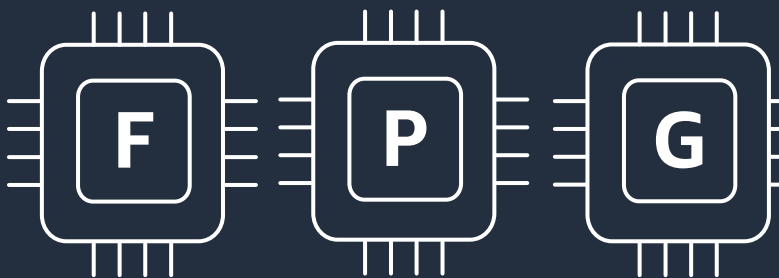


更に大きいメモリ量

メモリ最適化



High Memory



アクセラレーテッド

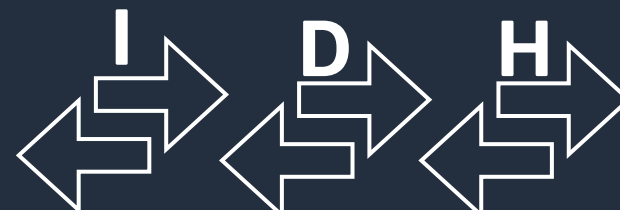
アクセラレータ
利用

インスタンスストア付加



高速大容量
ストレージ

ストレージ最適化



インスタンスタイプ最適化のポイント

「推測するな計測せよ」

実ワークロードでは想定外の部分に時間がかかっていたり
本番規模のデータ・負荷でないと露見しないボトルネックも多い

→ クラウド上では **様々なインスタンスタイプで実際に検証を行う** ことが可能なため、できるだけ早い段階で本番相当のデータ・負荷で評価を行うことを推奨

Point:

- インスタンスタイプごとにコスト・パフォーマンス評価を行う
- 最終的な判断には、リージョンや購入オプション等の観点も重要

継続的な改善

ワークロード利用のされ方や、新しいインスタンスタイプの登場等外的環境は変化し続ける

→ 継続的なモニタリング・見直し が重要

Point:

- 新しい世代のインスタンスタイプはコストパフォーマンスで優れることが多い
- Amazon CloudWatch メトリクスによる主要な指標の継続的な監視
- AWS Compute Optimizer による推奨インスタンス情報も活用

新しいインスタンスタイプ

- C7g/M7g/R7g インスタンス
AWS Graviton3 プロセッサ搭載
- R7iz インスタンス (Preview)
第 4 世代 Intel Xeon スケーラブルプロセッサ (Sapphire Rapids) 搭載
全コアターボ周波数が最大 3.9 GHz
- M7a インスタンス (Preview)
第 4 世代 AMD EPYC プロセッサ (Genoa) 搭載

発表済みの第 7 世代インスタンスは全て DDR5 メモリを搭載しており
メモリバンド幅が向上

本セッションのまとめ

- Amazon EC2 は Beta 提供を開始した 2006 年より
カスタマーの要望をもとに継続的な改善を繰り返しており
現在では 600 を超えるインスタンスタイプが利用可能
- c7gn.xlarge といったインスタンスタイプの表記から大まかな
スペックについて理解することができる
- インスタンスファミリーを選ぶ際は M 系を中心に必要な特性に応じて選択
- 定期的な計測と見直しによる継続的な改善が重要

**今回のイベントが皆様の利用している EC2 インスタンスの
見直しやコスト最適化のきっかけになれば幸いです**



Thank you!