

---

---

---

---

---



# AWSにおけるコンピューティングサービス

**コンピューティングサービス**は、アプリケーションを稼働させるインフラストラクチャサービスで、システムアーキテクチャの中核を担う。各サービスの機能、設計時に意識すべきことやコストの考え方をしっかり理解する必要がある。

- EC2(Amazon Elastic Compute Cloud)
- ECS(Amazon Elastic Container Service)
- Lambda(AWS Lambda)

**EC2**は、仮想サーバーを提供するコンピューティングサービスである。

必要な数だけにサーバーを立てることが出来る、いわゆるIaaS型のサービス。

Elastic Load Balancing(ELB)やAuto Scalingといったサービスと組み合わせることで、負荷に応じて動的にサーバーの台数を変更するクラウドらしい使い方も出来る。

**ECS**は、Dockerコンテナの実行環境を提供するサービスである。

このサービスが登場するまでは、AWSでDockerを利用するには、EC2上にコンテナ管理用のソフトウェアを導入する必要があった。ECSはサービスとしてDocker環境を提供してくれるため、利用者が設定・構築する項目を減らすことが出来る。

**Lambda**はサーバーを用意しなくてもプログラムを実行できる環境を提供するサービスである。

サーバーを用いないアーキテクチャ、すなわち**サーバーレスアーキテクチャ**の中心と言えるサービスで、拡張性やコスト効率の面でメリットがある。

サーバーのセットアップやメンテナンスの必要がないため、アプリケーションの開発に集中できる。

これらのサービスはどれが優れている、劣っていると言うものではなく、機能要件や非機能要件に応じて適切なものを選択する、もしくは組み合わせて利用するものである。各サービスの特徴を覚えるだけでなく、各サービスのユースケースを理解し、「過去に関わった案件をもう一度やるならどう設計するか？」といったことを考えてみると良い。

## EC2

オンプレミスな環境でサーバーを用意する場合、OSのインストールはもちろん、サーバーの調達、ラックの増設、ネットワークや電源の管理といった様々な作業が必要になる。サーバーの増設にはリードタイムがかかるため、新しいサービスを構築するときには時間をかけて見積もりを行い、ひつようにおう必要に応じて余裕率を掛けたスペックの環境を用意していた。

**Amazon Elastic Compute Cloud**(以下**EC2**)は仮想サーバーを提供するコンピューティングサービスである。

インスタンスという単位でサーバーが管理され、何度かボタンをクリックするだけで、あるいはCLIからコマンドを一つ叩くだけで、新しいインスタンスを作ることが出来る。

そのため、オンプレミス環境に比べ、サーバー調達のリードタイムを非常に短くできる。

サービスリリース前に最低限の見積もりは必要だが、リリース後に想定以上のペースで人気が出たとしても、インスタンスの数を増やす、インスタンスの性能を上げる調整をすれば対応できる。

このように、EC2を用いることでインフラリソースを柔軟に最適化することができる。

そのため、そのため、事前の見積りに時間をかけるのではなく、リソースまでの時間をいかに短くするか、リリースした後のトライアンドエラーや改善活動をいかに素早く回すかといった、ビジネス的に価値を生む行為に注力できる。また、カスタマー向けサイトだけでなく、社内の基幹システムでもAWSを利用することが増えてきている。

EC2を起動するときは、元となるイメージを選んでインスタンスを作成する。

このイメージのことをAmazon Machine Imageと呼ぶ。AMIと略される。

AMIにはAmazon Linux AMIやRed Hat Enterprise linux、Microsoft Windows ServerといったAWSが標準で提供しているものや、各ベンダーがサービスをプリインストールしたAMIがある。

また、各利用者がインスタンスの断面をAMIとしてバックアップすることも可能。

AMIを有効活用することで構築の時間を短縮したり、同じインスタンスを簡単に増やしたりすることが出来る。

## EC2における性能の考え方

EC2では**インスタンスタイプ**という形で、インスタンスのスペックを選択することが出来る。

インスタンスタイプは「m5.large」や「p3.large」といった形で表記される。

先頭の「m」や「p」はインスタンスファミリーを表し、何に最適化しているインスタンスタイプかを意味する。

たとえば、コンピューティングに最適化したインスタンスタイプは「c」から始まる「c5」や「c4」タイプ、

メモリに比重を置くインスタンスタイプは「r」から始まる「r5」や「r4」といった具合になる。

インスタンスファミリーの後ろの数字は世代を表し、大きいものが最新となる。

一般的に世代が新しいものの方がスペックが良かったり、安価だったりする。

「xlarge」や「8xlarge」の部分はインスタンスサイズを表し、大きいものほどスペックが高いインスタンスタイプになる。

基本的には、インスタンスサイズが倍になるとスペックも倍になる。

また、インスタンスの性能を決める他の重要な要因として、ディスク機能であるEBS(Elastic Block Store)がある。

EC2では通常の通信で使用するネットワーク帯域と、EBSとのやり取りで利用する帯域を共有している。

そのため、ディスクI/O、外部とのリクエストとともに多く発生する場合、帯域が足りなくなってしまうことがある。

このようなときに利用できるオプションが、**EBS最適化インスタンス**である。

このオプションを有効にすると、通常のネットワーク帯域とは別に、EBS用の帯域が確保される。

そのため、ディスクI/Oが増えても、外部との通信に影響が出なくなる。

EBS最適化インスタンスはある程度大きめのインスタンスタイプでしか利用できないため、公式ドキュメントで

オプションを利用できるかどうか確認することが重要。

## EC2における費用の考え方

EC2は、インスタンスを使った分だけ課金される重量課金型のサービスである。

EC2のコストは下記によって決まる。

- ・ インスタンスがRunning状態だった時間
- ・ Runnning状態だったインスタンスのインスタンスタイプ、AMI、起動リージョン

インスタンスには、起動中(Running)、停止中(Stopped)、削除済み(Terminated)の3つの状態がある。

EC2では起動しているインスタンスのみが課金対象となるので、一時的に停止中ステータスにしたインスタンスや削除したインスタンスは課金対象にならない。ただし、停止中のインスタンスでもEBSの費用はかかることに注意。

起動しているインスタンスは、インスタンスタイプに応じて課金される。

この費用はリージョンやインスタンスのAMIによっても異なる。



## スポットインスタンスとリザーブドインスタンス

EC2の価格は、**オンデマンドインスタンス**という通常の利用形態のものとなる。

EC2にはこれ以外にスポットインスタンスとリザーブドインスタンスという利用オプションがある。

**スポットインスタンス**は、AWSが余らせているEC2リソースを入札形式で安く利用する方式である。

例えば、m4.largeはオンデマンドで利用すると時間当たり0.129USDの利用料がかかる。

もし、m4.largeのリソースが余っているときに0.095USDという価格でスポット入札に成功すると、この値段でインスタンスを利用できる。

ただし、他の利用者からm4.largeの利用リクエストが増え、余剰なリソースがなくなってしまうと

インスタンスが自動的に中断される。この制約を許容できる場合、たとえば開発用の環境や機械学習のデータ学習のために一時的にスペックの大きいインスタンスを使いたいといった用途であれば、相性のよいオプションと言える。

**リザーブドインスタンス(RI)**は、長時間の利用を約束することで割引を受けられるオプションである。

例えば、m4.largeタイプの1年間のスタンダードRIを購入すると、37%も費用を削減できる。

サービスをリリースしてからしばらく経ち、インスタンスタイプを固定できると判断できた時点でRIの購入を検討するとよい。

## インスタンスの分類と用途

ソリューションアーキテクトが行う設計業務の中には、アプリケーションの用途に応じて適切なインスタンスタイプを選ぶというものがある。

そのためには、分類方法とインスタンスファミリーごとの主要な5つの用途を把握しておく必要がある。

- ・ 汎用
- ・ コンピューティング最適化
- ・ メモリ最適化
- ・ 高速コンピューティング
- ・ ストレージ最適化

**汎用**は、一番利用の範囲が広くCPUとメモリのバランス型である。M5やT3などが該当する。

**コンピューティング最適化**はCPUの性能が高いものである。C5やC系統のもの。

**メモリ最適化**は、メモリ容量が大きいものである。R5やX1など複数の系列があるが、一般的な利用でのメインはR系となる。

**高速コンピューティング**は、GPUなどCPU以外の計算リソースが強化されてる。

かなり細分化されているが、画像処理用のP系と機械学習用のG系をまず押さえる。

**ストレージ最適化**は、これもかなり細分化されているが、HDDを利用するとD2とSSDを利用するI3を覚える。

## Saving Plansとスケジュールされたリザーブドインスタンス

リザーブドインスタンスと比較して、より柔軟な形で割引を享受できるプランとして2019年11月から**Saving Plans**というサービスが提供されている。

リザーブドインスタンスとの違いは、インスタンスの利用者ではなく、EC2インスタンスの利用額に対してコミットする点。

Saving Plansには、**Compute Saving Plans**と**EC2 Instance Saving Plans**の2種類があり、

EC2 Instance Saving Plansは従来のリザーブドインスタンスと同じようにリージョンやインスタンスファミリーを指定して購入する。

Compute Saving Plansそれとは関係なく、全てのEC2インスタンスを対象に単位時間あたりの利用料を指定して購入する。

またリザーブドインスタンスの買い方のパターンとして、**スケジュールされたリザーブドインスタンス**というものがある。

これは、1年間のうちに毎日、毎週、毎月の一定時間のみ使うというパターンで、平日日中のみ使うという場合などに利用料を削減できる可能性がある。