

AWSクラウド演習

AWSクラウド演習授業資料



ELB(ELASTIC LOAD BALANCING)とは

- ELB(Elastic Load Balancing)とは

アプリケーションへの受信トラフィックを自動的に分散するサービス。EC2、コンテナ、IPアドレスなどのターゲットへのトラフィックを分散することができます。

- ELBの種類

ELBにはALB(ApplicationLB)、NLB(Network LB)、CLB(Classic LB)、GLB(Gateway LB)の4種類あります。

ALB・・・HTTP、HTTPS、Webソケットの高速なロードバランシングができます。レイヤ7で動作。

NLB・・・TCP、レイヤ4レベルで動作します。1秒間に数千万個のリクエストの処理ができます。

GLB・・・ネットワークファイアウォール用、ファイアウォールのルートを管理します。

CLB・・・旧世代のLB、HTTP、HTTPS、TCPに対応します。

ユースケース

- ELBは次のようなユースケースで使用されます。

単一アクセスポイント・・・Webサーバへのアクセスを保護します。

アプリケーション環境のデカップリング・・・環境を疎結合にします。

高可用性と耐障害性を実現(複数のAZに分散することで実現)。

伸縮性とスケラビリティ(拡張性)の向上。

- その他のユースケース

ALB・・・ポートによりパスを変更する。ターゲットへのルーティングルールを設定など。

パスベース、ホストベースでの分散することができる。

NLB・・・極度のパフォーマンスが求められるアプリなど。パケットのヘッダ情報により処理を変えるなど。IPアドレスを指定して負荷分散することができる。

ELBにかかわる用語

- ターゲットグループ

ELBを使用するインスタンスやIPアドレスが該当する。

- ヘルスチェック

インスタンスが正常に動作しているかのチェックする機能。EC2の振り分けを行う。

- AMI(Amazon Machine Image)

EC2インスタンスのテンプレート(OS、アプリケーションサーバ、アプリケーションを含む)。EC2のバックアップにも使用されます。

ELBの機能

- クロスゾーン負荷分散

ALBで有効な機能、ゾーンごとに均等に負荷分散される。

- スティックセッション

同じクライアント端末からのリクエストを同じEC2インスタンスへに送信する機能。

- Connection Draining

インスタンスが停止などの異常が発生した場合、数秒間は通信を切らずに一定時間待つ機能。

※処理中のリクエストが終了するまで待機する。

AUTO SCALING

- Auto Scalingとは

ルールに従い、EC2インスタンスを増減することができるサービス。処理の集中時にEC2インスタンスを増やし処理を行うことができます。複数のAZ(アベイラビリティゾーン)で作成することができます。

- Auto Scalingの3つのコンポーネント

起動テンプレート・・・何を(AMI、インスタンスタイプ、セキュリティグループ、ロールなど)を設定。

Auto Scalingグループ・・・どこで(VPC、サブネット、ELB、最小最大インスタンスなど)を設定。

Auto Scalingポリシー・・・いつ(スケジュール、オンデマンド、スケールイン/アウトポリシーなど)を設定。

用語・設計の考え方

- スケールイン/スケールアウト

スケールイン・・・インスタンスが削除されることです。

スケールアウト・・・インスタンスが追加されることです。

- 設計について

設計するうえで次のようなことを考慮します。単一障害点(Single Point of Failure)を作成しない。

DBはRDSなどはマルチAZを利用します。Webサーバはロードバランサーを利用して冗長構成で作成します。