

Resilience

障害を前提としたAWSアーキテクチャの設計と実装



Resilience (レジリエンス) の定義

回復力と弾力性

ワークロードがインフラストラクチャやサービスの中断、動的な需要の変化に対応し、復旧する能力。

AWSにおける定義

障害を完全に防ぐことではなく、障害が発生した際の中断時間を最小限に抑え、期待通りのサービスレベルを維持し続けること。

「絶対に壊れないシステム」 は存在しない

前提



ハードウェアの故障、ネットワーク切断、バグなど、あらゆるコンポーネントは最終的に故障する。

マインドセットの転換



「壊れないこと」から「壊れても自動復旧（Auto-Recovery）すること」へのシフト。

影響最小化



単一障害点（SPOF）を排除し、障害時の影響範囲（Blast Radius）を局所化するアーキテクチャ設計が必須。

障害の発生を前提とした対策

1

自動復旧の組み込み

人手による介入を待たず、システム自身が障害を検知し、健全なリソースへ切り替えるメカニズム。

2

分散と冗長化

マルチAZ、マルチリージョンを活用したトラフィックの分散とフェイルオーバー。

3

継続的なテスト

意図的に障害を注入し、システムの回復力を鍛え上げる（Chaos Engineering）。



Resilienceにおける責任共有モデル



AWS Resilience Hub: アプリケーション回復力の評価



レジリエンススコア

アプリケーションの回復力を
定量的なスコアとして算出。

目的: アプリケーションのレジリエンスを定義、検証、追跡するための単一の場所を提供。



RTO/RPO評価

顧客が定義した目標復旧時間(RTO)と目標復旧時点(RPO)の要件を満たしているかを検証。



推奨事項の提示



スコア改善のためのアーキテクチャ改善案、SOP（標準作業手順書）、アラーム設定を自動提示。





潜在するリスク: グレー障害と大規模障害

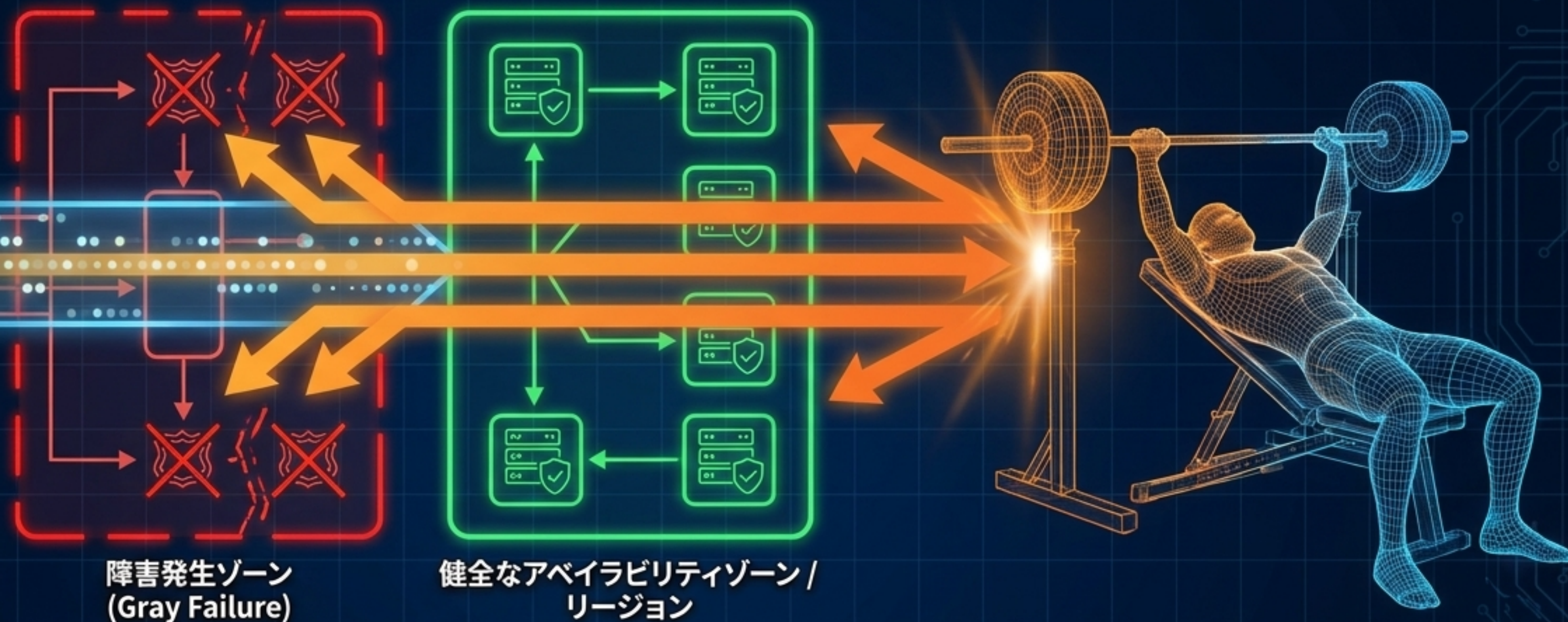
グレー障害の脅威

完全なダウンではなく、レイテンシーの悪化や一部のエラー頻発など、標準的なヘルスチェックでは「正常」と誤検知されやすい障害。

検知の遅延と課題

自動ルーティング（Route 53の標準ヘルスチェック等）だけでは、このようなグレー障害の検知と回避が遅れ、ユーザーへの影響が長引くケースが存在する。

Amazon Route 53 ARC による確実な復旧



Routing Control

複雑なヘルスチェックに依存せず、手動/自動のトリガーにより、オン・オフのスイッチ感覚でトラフィック全体を健全な環境へフェイルオーバー。

Zonal Shift

単一のAZでグレー障害等が発生した際、トラフィックを一時的に別の健全なAZへ即座に退避させる機能。

効果

影響範囲の特定が困難な障害時でも、迅速かつ確実なユーザー影響の最小化を実現。

まとめ 1: ビジネス要件との合意



システムダウン時の 許容値を定義する

- Resilienceを設計する第一歩は、技術的アプローチの前に「ビジネス上の合意」を形成すること。
- 「システムが落ちた際を想定して、どれだけリカバリーができるか」を明確化。
- 許容できる最大のダウンタイム (RTO) と、データ損失の範囲 (RPO) を顧客と綿密に相談し、不可欠な要件として合意する。

まとめ 2: AWSインフラ基盤のフル活用



強固な土台のポテンシャルを引き出す

- AWS側が提供する Resilience（グローバルネットワーク、AZ設計などの強固な土台）は既にそこに存在している。
- この Resilience 'OF' the Cloud を前提とし、AWS Resilience Hub や ARC を駆使して、そのポテンシャルを最大限に活かせるアーキテクチャ (Resilience 'IN' the Cloud) を構築すること。
- これこそが、アーキテクトである私たちの使命である。

ありがとうございました。