Engineer

ダウンタイム

🛟 利用者が必要なサービスを使えない時間

どのようなものか

ダウンタイムとは、サービスを必要とする人が、そのサービスを使えない時間です。停止や利用不能とも呼びます。反対は稼働時間で、サービスが稼働している時間の割合が可用性です。通常、一か月や一年を対象に、パーセントで表します。

分かりやすいのは、完全な停止です。サイトを読み込めない、APIが接続を拒む、データベースが止まっている、といった状態です。よくあるのは、もっと複雑な状態です。商品一覧は使えるのに購入手続きがタイムアウトする、ログイン画面は開くのにすべてのパスワードを拒む、APIが三回に一回しか応答しない。その経路が必要な人にとっては、どれもダウンタイムです。トップページを調べるヘルスチェックが正常と報告していても変わりません。役立つ定義は、プロセスが動いているかではなく、利用者が何をできるかを基準にします。

ダウンタイムには二種類あります。予期しない停止は、誰も選ばなかった停止です。ディスクの故障、不具合のあるデプロイ、証明書の期限切れ、先に止まった依存先などが原因です。計画停止は、事前に決めます。通常は保守時間帯を設け、トラフィックを処理しながらは変更できないものを、更新、移行、移設します。利用者へ予告できる分、受け入れてもらいやすくなりますが、契約で明示的に除外していない限り、可用性の計算には含めます。

HTTPには、停止時の状態を伝える語彙があります。RFC 9110は、自分の側でエラーが起きた、または要求を実行できないと分かっているサーバーのために、5xxクラスのステータスコードを定義します。停止時には、そのうち三つが別々の意味を持ちます。502 Bad Gatewayは、プロキシが背後のサーバーから無効な応答を受けたことを示します。504 Gateway Timeoutは、応答を長く待ちすぎたことを示します。意図して返すのが、503 Service Unavailableです。仕様は、一時的な過負荷や計画保守に結び付け、いつ戻ればよいかをサーバーが伝えられるようにしています。

```
HTTP/1.1 503 Service Unavailable
Retry-After: 120
```

これは、ブラウザー、APIクライアント、検索エンジンのクローラーへ、一時的な状態であり、二分間待つとよいと伝えます。壊れたページを成功コードで返すのとは、まったく異なるメッセージです。

概念を判断に生かす

「ダウンタイムとは？」を実際の判断に生かしましょう。その概念を、開発または購入するシステムに結び付けて考えるお手伝いをします。 障害が小さいうちに知らせる監視とアラートを整え、安定した状態をどれだけ維持できているかを数値で示します。

サービスを見る お問い合わせ →

用語の背後にある前提を検証する

解説とリンク先の実績を参考に、プロジェクトについての質問を整理してください。ツールを選ぶ前に、前提を一緒に検討できます。 社内初のオブザーバビリティダッシュボードを構築 — 社内システムの状態を初めて可視化

社内初のオブザーバビリティダッシュボードを構築 お問い合わせ →

なぜ重要なのか

可用性の目標は、九を並べて表します。九がいくつ並ぶかは、許される停止時間に換算できます。365日、つまり8,760時間の一年では、各目標が許す停止時間は次のとおりです。

- 99%：年間87.6時間、約三日半。
- 99.9%：年間8.76時間、または30日間の一か月で43.2分。
- 99.95%：年間4.38時間、または30日間の一か月で21.6分。
- 99.99%：年間52.56分。
- 99.999%：年間約5.26分。

九が一つ増えるたびに、許容時間は十分の一になり、実現する費用は急増します。99.9%なら、担当者が呼び起こされ、ログインし、原因を突き止めて直すところまで、月間の許容時間に収められます。99.99%では、一年全体でも一時間未満です。一件のインシデントで使い切るかもしれず、復旧を自動化する必要があります。99.999%では、通常の再起動さえ、利用者が気付かないように行う必要があります。

依存関係の可用性は掛け合わせます。ロードバランサー、アプリ、データベースを必要とするリクエストを考えます。それぞれの可用性が99.9%で、障害が互いに独立して起きるなら、三つとも動いているときだけ処理できます。0.999^3、約

99.7%となり、年間の停止は8.76時間ではなく約26時間です。サービスの可用性が、必要なものをつないだ鎖全体の可用性を超えるには、鎖の一部を取り除くか、複製して冗長化する必要があります。

計画は、事業の側から始めます。NISTの連邦情報システム向け緊急時対応計画ガイド、SP 800-34は、障害が起きる前に決めるべき三つの値を挙げます。最大許容停止時間（MTD）は、システムの所有者が受け入れる停止時間の総量です。目標復旧時間（RTO）は、影響が許容できなくなるまでシステムを停止させておける時間で、MTDより短くなければなりません。目標復旧時点（RPO）は、最後のバックアップから復元するとき、どこまで過去の時点のデータへ戻ることが許されるかを示します。停止時間ではなく、許容するデータ損失を測ります。この三つを明記すれば、バックアップの予定やフェイルオーバーの設計は、勘ではなく、決めた問いへの答えになります。

実際の活用例

中国、米国、ドイツに顧客を持つIPTV・OTT配信プラットフォームでは、どの時間も誰かにとってのゴールデンタイムでした。保守に使える静かな時間はありませんでした。約千台のサーバーと、相当数の顧客所有システムを、常時サポートしました。仕事の多くは、システムの状態が悪化し始めたことに、停止へ至る前に気付くことでした。生配信では、停止すると誰かが見ているテレビの映像が止まるためです。プラットフォームは、世界の視聴者に対して継続的な可用性を保ちました。

中規模の企業では、十台のPostgreSQL・SQL Serverデータベースサーバーを、稼働率99.9%で運用しました。上の計算では、一台ごとに年間8.76時間の停止を許す値です。その数字を支えたのは、`pg_dump`とSQL Server Agentジョブによる自動バックアップ、保持ポリシーとオフサイト保存、PrometheusとGrafanaによる監視、そして時点復元とフェイルオーバーのために文書化した災害復旧手順でした。

成り行きで設定が増えていた三十台のUbuntuサーバーでは、バックアップと、実際に試した復元手順で、災害復旧の仕組みを作り直しました。ネットワーク設定は、全インスタンスで共通の、セキュリティを強化した基準構成にしました。復旧経路を必要になる前に試すことで、停止のリスクを減らしました。

問題が起きるところ

監視が、停止している側にあります。 同じサーバーや同じデータセンターで動くヘルスチェックは、正常と報告し続けます。その一方で、利用者は期限切れのTLS証明書、不正なDNSレコード、宛先のないロードバランサーに遭遇します。外部の複数の場所から、実際の利用者と同じ経路を確かめ、その結果で通知します。

エラーページが200を返します。 親切なエラーページを成功ステータスコードで表示するフレームワークは、監視や、再試行するはずのクライアントから停止を隠します。クローラーには、本来のページの代わりにエラーページを登録させかねません。エラーにはエラーのステータスを付け、計画保守には503とRetry-Afterヘッダーを返すことが推奨されます。

バックアップを一度も復元していません。 バックアップが証明するのは、バイト列をコピーしたことだけです。復元を実行し、時間を測り、データを照合するまでは、復旧時間は推測です。紙に書いたRTOより、はるかに長いこともよくあります。定期的に別環境へ復元し、所要時間を記録します。フェイルオーバーも同じです。一度も引き継いでいない待機系は、期待にすぎません。平時にフェイルオーバーを練習すれば、インシデントが起きる前に問題を見つけやすくなります。

停止を防ぐはずのリリースが、停止を起こします。 新版の準備が整う前に旧版を止めるデプロイは、毎回、空白時間を作ります。大きなテーブルを排他ロックの下で書き換えるスキーマ変更は、終了するまで、そのテーブルへの全クエリーを止めます。新しいインスタンスを展開し、準備完了チェックを通してから、古いインスタンスを止めます。スキーマ変更は、先に拡張し、後から縮小する二段階に分け、その間は新旧両方のコードが動くようにします。Kubernetesではpod disruption budgetを設定し、ノードのドレインで全レプリカが同時に止まらないようにします。

参考資料

- RFC 9110：HTTPの意味論、15.6節 — 503とRetry-Afterによる待機時間の目安を含む、5xxステータスコードです。
- NIST SP 800-34改訂1：連邦情報システム向け緊急時対応計画ガイド — 最大許容停止時間、RTO、RPOの定義です。
- PostgreSQL：高可用性、負荷分散、レプリケーション — データベース自身のマニュアルによる、待機サーバーとフェイルオーバーの説明です。

- PostgreSQL：継続的アーカイブと時点復元 — 時点復元に必要な、ベースバックアップとWALアーカイブです。
- Kubernetes：中断 — 自発的な中断と、それを制限するPodDisruptionBudgetです。

用語集

実際の制約に合わせて選ぶ

デプロイの流れ、運用上のリスク、緊急対応の担当者をお知らせください。検査、復旧手段、明確な責任の所在を伴う変更の範囲を定められます。

料金を見る お問い合わせ →

用語を実務に結び付ける

「ダウンタイムとは？」がお客様の組織にどう当てはまるかをご相談ください。現在のシステムの具体例が一つあれば、話を始められます。

サービスを見る

welcome@engineer.company +45 50 10 61 84 デンマーク、コペンハーゲン

このページを持ち歩く

連絡先カード

心を込めて丁寧に制作： D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

著作権 © 2025 – 現在 · Engineer ApS · すべての権利を留保します。

印刷元： https://engineer.company/ja/glossary/downtime/