CLOUDFLARE

ホワイトペーパー

負荷分散のベストプラクティスでWebパフォーマンスと信頼性を最適化

概要

企業は毎年、Webサイトの表示の遅さやダウンタイムによって数百万ドルを失っています。その大半は逸失収益です。サイトやアプリが遅かったり使えなかったりすると、社内の生産性にも悪影響を及ぼし、検索順位も下がります。こうしたパフォーマンスと信頼性の問題には、以下のようなさまざまな要因があります：

- サーバーの過負荷や不健全性
- エンドユーザーとサーバーの地理的距離
- DNS解決の遅さ
- 分散型サービス妨害（DDoS）攻撃
- Web訪問者がインターネットへのアクセスに使うデバイスのタイプ

ロードバランサーは、Webトラフィックをサーバーネットワーク全体に均一に分散することによって、遅延と可用性の問題を軽減し、サーバーが過負荷に陥らないように、そしてサーバーが1台ダウンしてもWebアセットの可用性が保たれるようにします。企業は、従来はデータセンターに物理的ロードバランサーをデプロイしていましたが、コンピューティングがクラウドへ移行するにつれて、柔軟性が高く安価で使いやすいクラウド型の負荷分散ソリューションに魅力を感じるようになっています。

しかし、クラウド型の負荷分散ソリューションならどれも同じかというと、そうではありません。堅牢なソリューションはグローバルコンテンツ配信ネットワーク（CDN）と統合して、グローバルジオロケーションに基づくルーティング、DDoS攻撃時の回復、レイヤー3とレイヤー4の負荷分散、分析、ほぼリアルタイムのフェイルオーバーといった機能を提供します。現在ほとんどの企業が持つマルチクラウドやハイブリッドクラウドのデータ環境にも、シームレスに統合します。

遅延とダウンタイムの原因

遅延とダウンタイムはビジネスに大きな悪影響をもたらします。企業で発生する遅延やダウンタイムにはさまざまな原因が考えられます。

サーバーの負荷分散が不均一

過剰使用されているサーバーでは、リクエスト間で限りあるリソースの取り合いが起こるため処理速度が落ちます。サーバーが過負荷になると、Webサイトやアプリケーションのパフォーマンスが低下したり、完全に使えない状態に陥ったりする可能性があります。

効果的な負荷分散はワークロードをサーバーネットワーク全体に均一に分散し、パフォーマンスを大幅に向上させることができます。例えば、あるSaaS企業の顧客は世界のさまざまな地域で遅延に悩まされていましたが、Cloudflare LoadBalancingをデプロイしたとたんに遅延が改善され、ページ読み込み時間が2〜3秒短縮されました。[1]

地理的距離

世界のインターネット普及率は急上昇しています。**2023年1月には世界人口の64.4%が接続しており、うち100万人以上は2022年に初めて接続したユーザーでした。**[2]

インターネットのグローバル化は、ネットワークのパフォーマンスに複数の影響を与えています。アクティブユーザー数が増えるにつれてユーザーあたりの使用可能帯域幅が減り、遅延が発生します。

ここ数年は、リモートワークの普及によってユーザーの分散化も進みました。かつては企業環境内のEast-Westトラフィックと考えられていたものが、現在はリモートユーザーへの通信がインターネット経由で行われ、North-Southトラフィックになっています。この変化によってインフラにさらなる負荷がかかり、トラフィックが移動するユーザー・サーバー間の往復距離が長くなって遅延が増大しているのです。[3]

サイトとアプリの複雑さ

インターネットは複数の段階を経て進化し、イテレーションのたびにWebサイトやアプリケーションが複雑化しています。最新のWebサイトは情報量がかつてなく多く、総ページサイズが2011年以降着実に大きくなっています。[4]

ビデオ会議、オンラインゲーム、その他類似のオンラインサービスも、Webサイトとアプリケーションのサイズ増大と複雑化につながっています。それらのアプリケーションは帯域幅の使用量が多い上に遅延の影響を受けやすいため、企業のネットワークやインフラにさらなる負荷がかかり、プレッシャーが強まります。

デバイスタイプ

Webトラフィックの60%以上はモバイル端末からのもので、[5] モバイルユーザーの約半分はアプリが2秒以内に応答することを期待しています。[6] Webサイトとアプリケーションをモバイル端末用に設計し、最適化する必要があります。

5Gモバイルネットワークが出現しても、モバイルユーザーにとっては高速で無制限のネットワークアクセスが保証されるわけではありません。顧客のコンバージョン率は、コンテンツをモバイル端末へすばやく配信できるかどうかにかかっています。

遅いDNS解決

DNSリゾルバーはドメイン名をIPアドレスに変換し、Webアセットを求めるリクエストのルーティングに必要な情報をコンピューターに提供します。DNS解決はオンラインリソースにアクセスするための重要な第一歩であり、その最適化はパフォーマンスの最大化に欠かせません。

解決の速度を上げるための最適化がどのDNSリゾルバーでも行われているわけではありません。多くのDNSプロバイダーでは、各DNSクエリーの解決に20〜120ミリ秒かかります。[7] 最速のDNSプロバイダーは20ミリ秒以内でクエリーを解決します。例えば、Cloudflare DNSは平均8.92ミリ秒で解決します。[8]

これらの数値は大差ないように思えるかもしれませんが、1ページのレンダリングに複数のHTTPリクエストやDNSリクエストが必要な場合があることを考慮することが重要です。例えば**平均的なWebページの場合、デスクトップで71件のHTTPリクエスト、モバイルで66件のリクエストが行われます。**[9] それらのリクエストの中には同一ドメインに対するものもあるでしょうが、遅延は一意のDNSリクエストが送信されるたびに発生します。

サーバーの健全性

サーバー障害の原因はさまざまです。サーバーが落ちると、そのサーバーでホストしているアプリケーションやWebページをユーザーが利用できなくなる場合があります。

また、ユーザーが消費する動画コンテンツが増えており、今日の世界では、何かがバズると大量のトラフィックが集中してサービスが応答しなくなる可能性があります。正当なトラフィックでDDoS攻撃時のようにアプリケーションがダウンしないよう保護するには、ITインフラストラクチャに負荷分散ソリューションを追加し、冗長性を持たせることが肝要です。

負荷分散ソリューションはサーバーの健全性を監視して、アプリケーションを利用可能な状態に保つ必要があります。そうしなければ、問題が起きているサーバーにトラフィックがうっかりルーティングされて、ユーザーが大幅な遅延や障害を体験することになるかもしれません。

サイバー攻撃

分散サービス妨害（DDoS）攻撃は、オンラインサービスの健全性と可用性を大いに脅かします。DDoS攻撃はWebサーバーに大量のスパムリクエストを送り付けて正当なトラフィックを圧倒しますし、他の攻撃を隠すための煙幕となる可能性があります。

安全が確保されていないIoT（モノのインターネット）デバイスが増えて、それらがよくDDoSボットネットマルウェアの標的にされることも、DDoS攻撃増加の一因になっています。**2022年の第4四半期には、DDoS攻撃が前年比79%増となりました。**[10]

遅延とダウンタイムのコスト

ネットワークの遅延とサイトの読み込み時間は、カスタマーエクスペリエンスとコンバージョン率に大きな影響を与えます。実際、わずか100ミリ秒の遅れでも消費者行動に測定可能な影響を与えます。

遅延はビジネスにさまざまな悪影響を及ぼしかねません。遅延とダウンタイムの一般的なコストには、以下のようなものがあります：

- **逸失収益：**企業はWebサイトを通じて顧客とつながり、サービスを提供することが多くなっています。ダウンタイムや遅延によって顧客が企業のWebサイトにアクセスできなかったり、ページの読み込みが遅いためにカートを放棄してしまったりして、セールスの機会が失われる可能性があります。
- **顧客の解約：**ページの読み込みが遅いと売り上げが失われます。**1秒で読み込めるページのコンバージョン率は、5秒のページの3倍です。**[11]
- **生産性の低下：**内部アプリケーションの遅延とダウンタイムは、従業員の生産性にも影響を及ぼします。例えば**米国の平均的従業員は、アプリケーションを1分間使うごとに約1秒の待ち時間を過ごしています。**[12] これは、年間で4日以上を無駄にしている計算になります。
- **ブランド認知度：**Googleでは、デスクトップとモバイルの両方の検索でランキング要素にページ速度を採用しています。[13] ページの読み込み速度が遅いと、ブランド認知度に悪影響を及ぼす可能性があります。
- **法規制コンプライアンス：**オンラインサービスのプロバイダーは、可用性とアップタイムを含めたサービスレベル契約（SLA）に拘束される場合が多く、ダウンタイムや遅延は罰金やことによると訴訟の対象になりかねません。

ビジネスにとってダウンタイムは高くつきます。ダウンタイムの平均コストは1分あたり9000ドルですが、[14] これは業界や事業規模によって変わってきます。例えば、Facebookは14時間のサービス障害により推定9000万ドルの損失を被りました。1分あたりのコストにすると10万7000ドルに上ります。[15]

負荷分散を理解

遅延やダウンタイムは、企業にとっては大きなコストを伴います。ロードバランサーはオリジンサーバーのネットワークとインターネットの間に位置するサービスで、複数サーバーへの均一分散によってこのコストを軽減する効果があります。それにより、トラフィック急増時でも個々のサーバーは過負荷に陥らず、アプリケーションの信頼性、効率性、応答性が保証されます。

ロードバランサーが必要な理由

エンドユーザーがWebページを訪問する際は、オリジンサーバーがこのリクエストを受信して応答します。リクエストを処理し、求められたコンテンツを収集してユーザーのブラウザーへ送信し、そのブラウザーがレンダリングするという流れです。

1台のオリジンサーバーが扱えるリクエストの数は、物理的インフラストラクチャとコードの複雑さによります。しかし、Webサイトが受信するリクエストの数が、最高のハードウェアと最高パフォーマンスのWebアプリケーションの許容範囲を上回る可能性があります。その場合、リクエストは待ち行列に入れられて遅延増大の原因となるか、破棄されます。

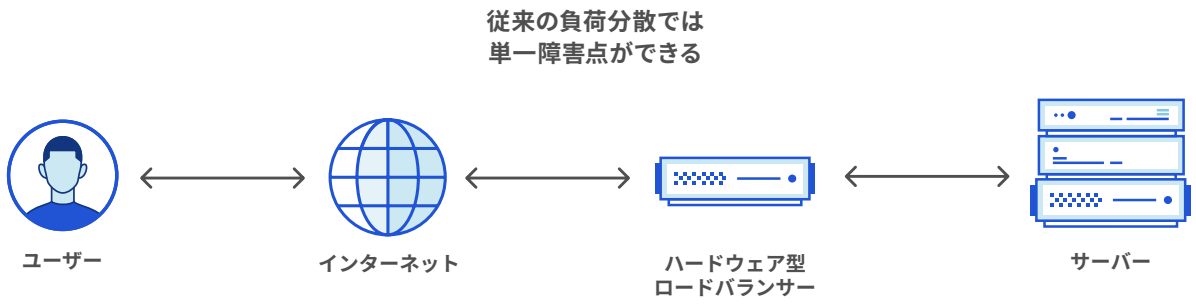
ロードバランサーを使えば、個々のサーバーでこのような問題が起こるのを回避できます。ロードバランサーはエンドユーザーとオリジンサーバークラスターの間に位置し、負荷をサーバープール全体に均一に分散します。各サーバーの負荷を減らすことによって、Webサイトのパフォーマンスと耐障害性を高めるのです。

従来のロードバランサー

従来、ロードバランサーはオンプレミスのデータセンターにデプロイされていました。専用ハードウェアで実装されるのが一般的でしたが、仮想化されたロードバランサーも利用可能でした。耐障害性を確保するために、一対のロードバランサーをデプロイして、プライマリロードバランサーに障害が発生した時にバックアップシステムが稼働するようにしておくのが一般的でした。

こうした従来のハードウェア型ロードバランサーにはかなりの制約がありました。以下のような問題点があったのです：

- **初期費用：**ロードバランサー機器を購入して設置しなければなりません。出費がかさむ可能性があり、しかもすべて前払いです。
- **拡張可能性：**ハードウェア型のソリューションは最大容量が決まっており、トラフィックが異常に急増した時はロードバランサーがボトルネックになる場合があります。帯域幅を広げる必要があるため、既存ソリューションを新たなハードウェアで強化または代替しなければなりません。
- **地理的制約：**ロードバランサー機器は、企業が物理的ハードウェアを設置できるデータセンターにしかデプロイできません。そのため、管理できるのはオンプレミスアプリケーションへのトラフィックのみで、クラウドベースのアプリケーションへのトラフィックは管理できません。
- **スキルギャップ：**社内ロードバランサーは概して、社内の人材が設定して運用します。企業は、必要なスキルセットを持つ人材の募集と確保に苦労するかもしれません。
- **柔軟性の欠如：**ハードウェアのロードバランサーは、企業の物理的ネットワークインフラストラクチャに接続された機器であり、要件の変化に適応しにくくなっています。



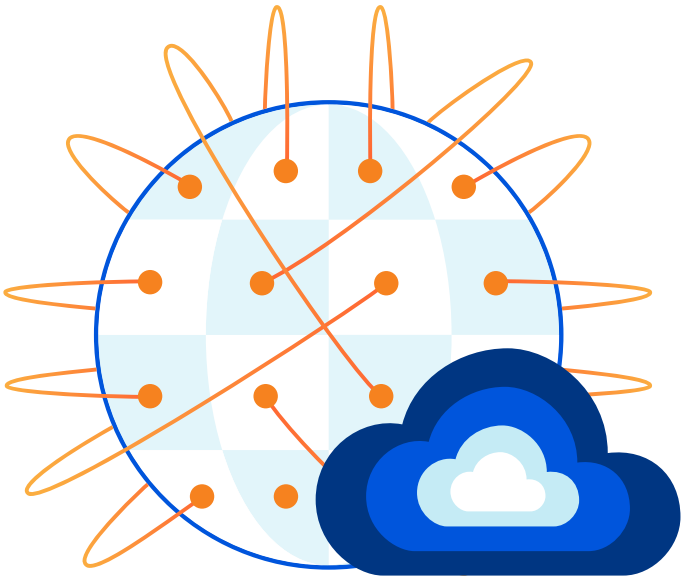
次世代のクラウド型ロードバランサー

大部分の企業はクラウドへ急速に移行しています。企業の87%はマルチクラウドインフラストラクチャを持ち、72%はパブリッククラウドとプライベートクラウドの両方を取り入れたハイブリッドクラウド環境になっています。[16] 企業アプリケーションも、もはやハードウェアロードバランサーの背後に置けないものが増えてきています。

堅牢なクラウド型スタンドアロンロードバランサーは、ハイブリッド環境で従来のハードウェア型デバイスと一緒に使えますし、パブリッククラウドにネイティブなロードバランサーとも併用できます。スタンドアロンロードバランサーは、企業のハードウェア型ロードバランサーやパブリッククラウドネイティブのロードバランサーの上に位置する、クラウドに依存しないニュートラルなレイヤーです。企業は、すべてのトラフィックを流すためのプライマリプロバイダーを選びます。ロードバランサーが障害を検知すると、バックアップのプロバイダーまたはリージョンへ自動的にトラフィックをルーティングします。企業がパブリッククラウドや自社インフラで障害や断続的なネットワーク接続を経験すると、クラウド型のスタンドアロンロードバランサーが健全なプロバイダーまたはサーバーへ自動的にフェイルオーバーします。

仮想化されたロードバランサーをクラウドにデプロイすれば、それらのアプリケーションへのトラフィックを管理することができます。クラウド型ロードバランサーには、以下を含むさまざまなメリットがあります：

- **実質的に無制限の拡張可能性：**クラウドロードバランサーには、クラウドの柔軟性と拡張可能性という利点があります。容量を必要に応じてすばやく追加でき、企業Webアプリケーションへのトラフィック急増に対応できます。
- **従量課金制でコスト削減：**クラウドロードバランサーは、サービスベースモデルで提供されるのが一般的です。企業は過大な機器を購入せずに済み、利用した容量に対してだけ支払いをすればよいのです。
- **より広い地理的範囲をカバー：**クラウドロードバランサーはグローバルなプレゼンスのあるネットワークで稼働するのが理想的で、それにより、アプリケーションがどこにあろうとも近くに配置することができます。



- **設定と管理が容易：**ロードバランサーがサービスとして提供されている場合は、サービスプロバイダーが設定と管理の大方を行います。これにより、企業のオーバーヘッドが削減でき、専門の人材を確保する必要性も減ります。
- **柔軟性：**クラウド型のスタンドアロンロードバランサーは、新しい環境で稼働するアプリケーションをサポートするための再設定や移動が簡単にできます。そのため、企業は変化にすばやく適応することができ、ベンダーロックインも回避できます。
- **耐障害性：**クラウド型ロードバランサーは、クラウドに内蔵された耐障害性と可用性保証を活かすことができます。これにより、ロードバランサーの背後にあるアプリケーションが障害時にオフラインになるリスクが軽減されます。
- **機能の統合：**クラウド型のソリューションであれば、Load Balancingのオンボーディング後に必要に応じてWebアプリケーションファイアウォール（WAF）、ボット管理などのモジュールを簡単に追加できます。ハードウェアソリューションは概して、アップグレードの際にハードウェアデバイス全部を入れ替えたり、物理的なモジュールやブレードを追加したりする必要があります。企業はそうした変更をするために、保守のためのダウンタイムの予定を組まなければならず、その間は顧客が無防備となりビジネスに悪影響を及ぼしかねません。

クラウドベースの負荷分散ソリューションを評価する際に確認すべき点

クラウド型の負荷分散ソリューションは、企業が遅延やダウンタイムを減らしてビジネスへの影響を軽減するのに役立ちます。以下では、負荷分散ソリューションを評価する際に注目すべき機能について説明します。

グローバルコンテンツ配信ネットワーク（CDN）との統合

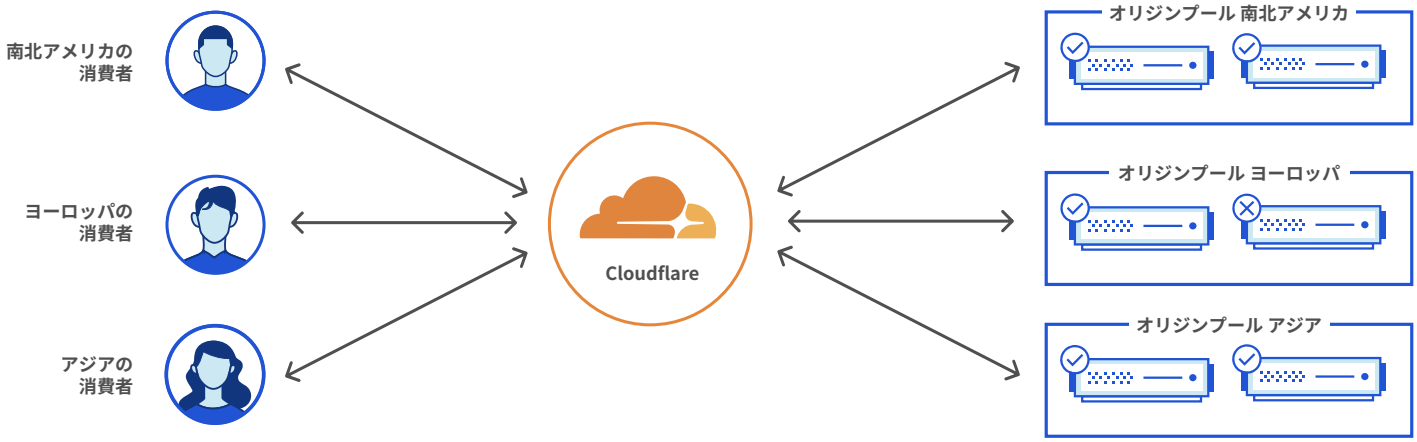
ロードバランサーとCDNはいずれも、遅延を減らして可用性を高めるために設計されたソリューションです。CDNは静的コンテンツをネットワークエッジでキャッシュし、リクエストと応答が移動する距離を短縮します。さらに、分散されたCDNサーバーからコンテンツを配信することでオリジンサーバーの負荷を軽減します。

負荷分散とCDNの統合でコンテンツの配信を最適化します。ロードバランサーは、CDNクラスターとオリジンサーバーにリクエストを分散して、パフォーマンスを最適化し、帯域幅の消費量を最低限に抑えます。

グローバルジオロケーションに基づくルーティング

サーバーとエンドユーザーの地理的距離は、リクエストと応答の遅延を大きく左右します。ロードバランサーは、最寄りの利用可能なインフラストラクチャへトラフィックをルーティングするはずですので、移動距離を最短化できます。例えば英国のトラフィックは、ニューヨークではなくロンドンのデータセンターへ向かいます。

また、ロードバランサーは最適化された高速DNSルックアップを実現します。例えば、DNSクエリーは最寄りの健全なDNSサーバーへ送られ、DNSルックアップによる遅延を最小化します。



アプリケーションの配信とセキュリティを統合

ロードバランサーとCDNネットワークは、さまざまなセキュリティの懸念事項を念頭に設計されていなければなりません。例えば、DDoS攻撃はサーバーの健全性と可用性にとって大きな脅威となります。そのため、CDNネットワークは最大級のDDoS攻撃にも耐えられるように拡張され、保護されなければならないのです。

ロードバランサーとCDNに関するもう1つの大きな懸念は、プライバシー標準とセキュリティ標準のコンプライアンスです。例えば、ロードバランサーは、顧客データを暗号化しWebトラフィックを認証するためのTLS/SSLの使用をサポートしていなければなりません。

レイヤー3 & 4の負荷分散機能

DDoS攻撃はOSI参照モデルの複数レイヤーで起こる可能性があります。帯域幅消費型のDDoS攻撃は、さまざまなサービスを実装するポートに大量のトラフィックを送り付けてWebサーバーを圧倒します。例えばDDoS攻撃は、SMTPポートを標的にしてメールサービスを中断することもありますし、カスタムゲーミングプロトコルやその他のオンラインサービスの実装に使われるカスタムポートを狙うこともあります。ロードバランサーは、レイヤー3/4へのDDoS攻撃から保護されていなくてはならず、攻撃中も通常のサービスを維持するのに十分な容量がなくてはなりません。

ほぼリアルタイムのフェイルオーバー

クラウド型ロードバランサーはパブリックDNSに依存する場合が多いですが、パブリックDNSは変更の反映が遅く、問題発生時のフェイルオーバーが遅れてしまいます。フェイルオーバーが数秒で発動するためには、ロードバランサーがTTLの短いDNSリゾルバーを使っていなければなりません。

マルチクラウドとハイブリッドクラウドのサポート

多くの企業は、マルチクラウドまたはハイブリッドクラウドの環境です。ベンダーロックインを回避し、簡素化し、マルチクラウドやハイブリッドの環境での設定ミスを最小限に抑えるために、負荷分散ソリューションがオンプレミスでもどんなパブリッククラウドでも機能するニュートラルなレイヤーであるようにします。

ベンダー非依存型のロードバランサーは、クラウドベンダーネイティブのロードバランサーや従来型のハードウェア機器にとって代わるものではありませんが、それらと連携してマルチクラウドインフラストラクチャがスムーズに機能するようにします。

自動化とDevOpsサポート

ロードバランサーは、サーバークラスターにリクエストを分散するように設計されています。アジャイルとDevOpsのプロセスやクラウドコンピューティングの出現によって、企業アプリケーションのインフラは常に変化する移動標的になっているかもしれません。

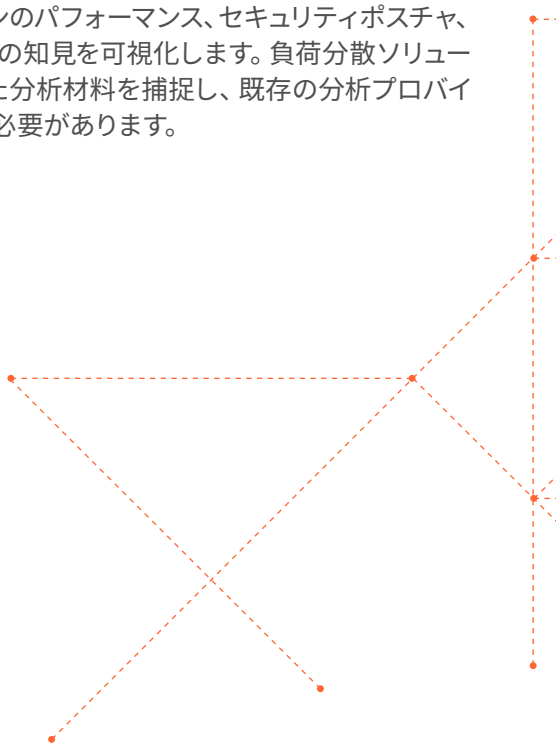
定義と設定変更の実装を人間のオペレーターに依存することは、可用性とパフォーマンスに対する大きなリスクになります。ロードバランサーには自動化とDevOpsのサポートを組み込み、企業のITインフラの進化に伴って変更をすばやく大規模に行えるようにしなければなりません。

使いやすさ

負荷分散ソリューションの設定と管理は、熟練者が時間をかけて行うリソース集約的な作業になる可能性があります。良いクラウド型ロードバランサーは、数分で設定してセットアップでき、管理の手間も最低限で済むものです。グラフィカルUIや強力なAPIをサポートし、ビジネスニーズの変化をサポートするために簡単に再設定できるソリューションでなければなりません。

詳細なAnalytics

ロードバランサーがエンドユーザーとアプリケーションの間に置かれているのは、実用的なビジネスインテリジェンスを収集するためです。ロードバランサーは、顧客の挙動、アプリケーションのパフォーマンス、セキュリティポスチャ、その他の運用上の知見を可視化します。負荷分散ソリューションはそうした分析材料を捕捉し、既存の分析プロバイダーと統合する必要があります。

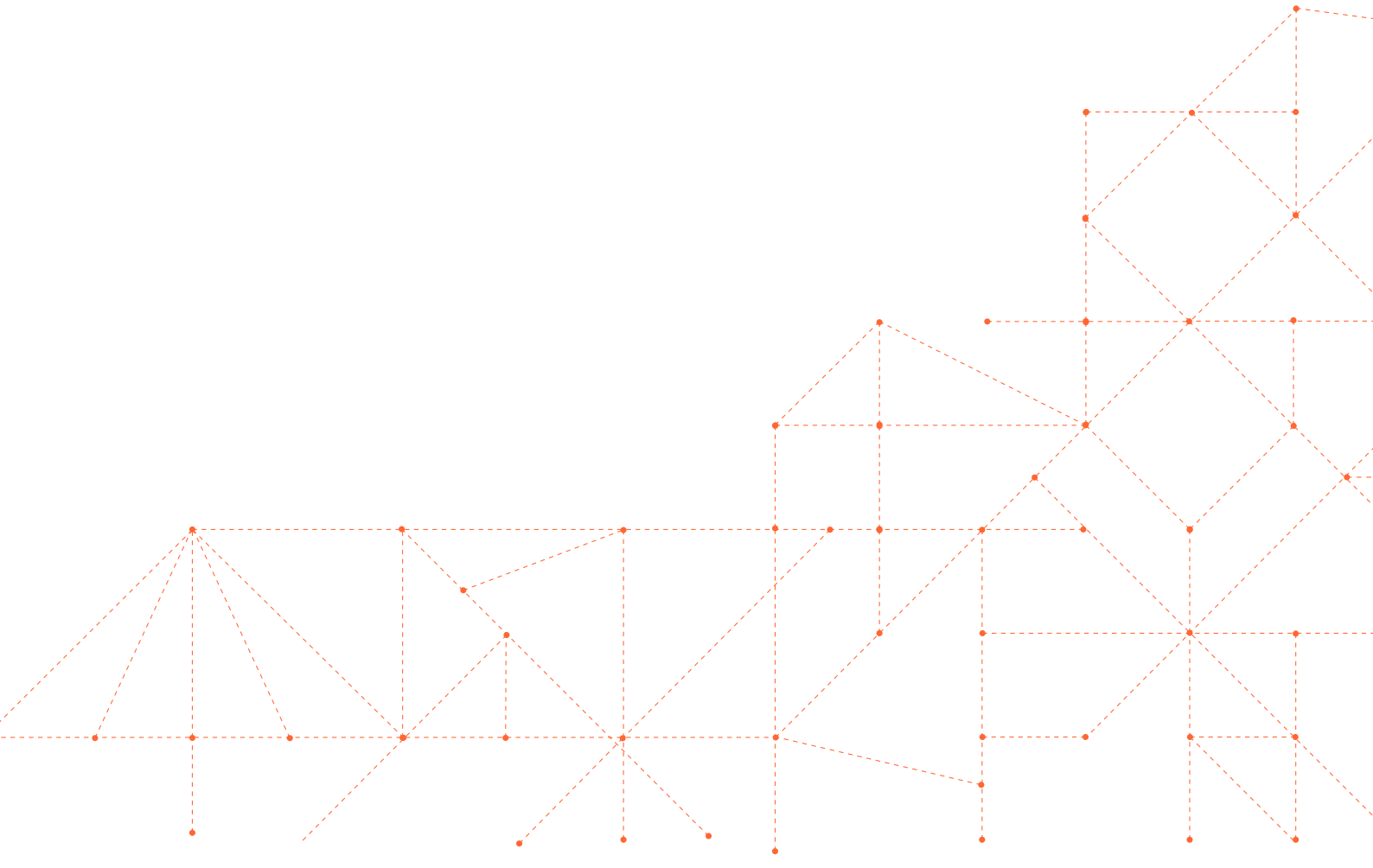


まとめ

最新のWebサイトやアプリケーションは、ロードバランサーが無ければ正しく機能せず、一貫してオンライン状態を維持することもできません。堅牢なクラウド型ロードバランサーは、従来のハードウェア型のソリューションより遥かに良い選択肢です。

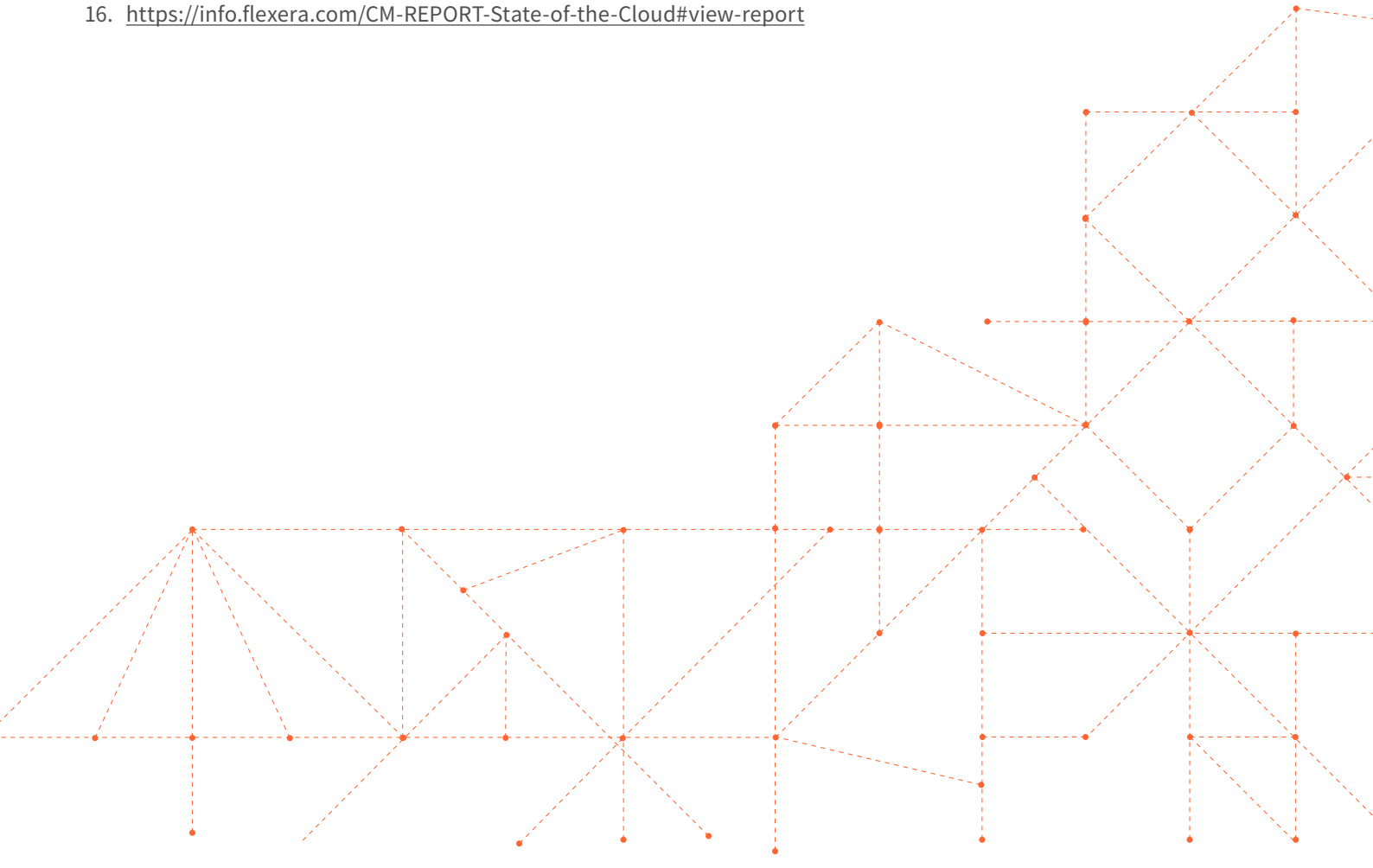
クラウド型のスタンドアロンロードバランサーは、比較的安価で使いやすく拡張可能な上、従来のハードウェア型ロードバランサーやパブリッククラウドプロバイダーが提供する独自ソリューションを補強して、Webアセットが常に可用性と高パフォーマンスを保てるようにします。

Cloudflareのグローバルネットワークと高パフォーマンスなCDNは、可用性の最大化と遅延の最小化に役立ちます。Cloudflare Load Balancingの詳細をご確認ください。



参考文献

1. https://www.cloudflare.com/case-studies/crisp/
2. https://datareportal.com/global-digital-overview
3. https://www.techwalla.com/articles/network-latency-milliseconds-per-mile
4. https://httparchive.org/reports/state-of-the-web#bytesTotal
5. https://gs.statcounter.com/platform-market-share/desktop-mobile/worldwide/#yearly-2011-2022
6. https://techbeacon.com/sites/default/files/gated_asset/mobile-app-user-survey-failing-meet-user-expectations.pdf
7. https://sematext.com/glossary/dns-lookup-time/
8. https://www.dnsperf.com/
9. https://httparchive.org/reports/state-of-the-web#reqTotal
10. https://blog.cloudflare.com/ja-jp/ddos-threat-report-2022-q4-ja-jp/
11. https://www.portent.com/blog/analytics/research-site-speed-hurting-everyones-revenue.htm
12. https://www.apmdigest.com/the-impact-of-app-performance-on-productivity
13. https://developers.google.com/search/blog/2018/01/using-page-speed-in-mobile-search
14. https://www.vertiv.com/globalassets/documents/reports/2016-cost-of-data-center-outages-11-11_51190_1.pdf
15. https://www.ccn.com/facebooks-blackout-90-million-lost-revenue/
16. https://info.flexera.com/CM-REPORT-State-of-the-Cloud#view-report



CLOUDFLARE

enterprise@cloudflare.com | www.cloudflare.com

REV:BDES/4505.2023APR27