

INTERNET

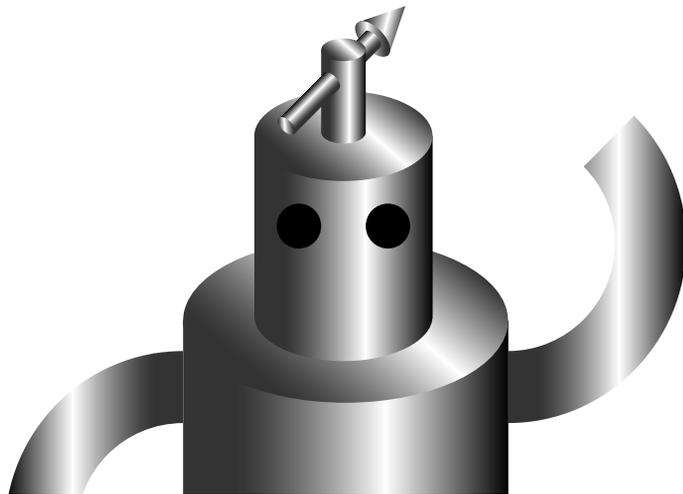
● インターネット最新テクノロジー：第43回

大量のアクセスからサーバーを守る

ロードバランシング技術

インターネット利用者の増加により、コンテンツを提供するサーバーへの負荷が急速に増大している。こうした状況ではサーバーマシンを高性能なものに置きかえるよりも、複数台のサーバーに処理を分散させる「ロードバランシング技術」が必要になってくる。インフラのブロードバンド化やストリーミングコンテンツの普及を目前に控え、注目を集めているロードバランシング技術を解説する。

上谷 — F5 ネットワークスジャパン(株)



パフォーマンスの向上と ノンストップ運用の実現

昨年12月、ブッシュ、ゴア両氏の大統領選の集計結果が注目を集め、米最高裁判所のウェブサイトはリンク寸前の状態となった。こうした事態を受けて最高裁のサイトが導入したのが、「ロードバランサー」と呼ばれる装置を用いて14台のサーバーで分散処理させる方式だった。

このように、複数台のサーバーを使ってサイトの運営を分散処理させるのが、ロードバランシングと呼ばれる技術である。ロードバランシングの目的は、アクセス数の増大に対するパフォーマンスの向上とノンストップ運用にある(図1)。

現在ではサーバーの性能も向上しているため、HTML ファイルを返すだけであれば1秒あたり数百リクエスト程度の処理は可能だ。しかし、CGIやJava、SSLといった負荷の

かかる処理では、1台のサーバーで1秒あたり数十リクエストといったところまで処理能力が落ちてしまう。こうした場合にはさらに高性能なサーバーマシンを導入するよりも、ロードバランサーを用いて複数台のサーバーで運用するほうが、サーバーの台数に比例した処理能力の向上が期待できる。

また、サーバーマシンに障害が発生した場合や、OSをバージョンアップする際など、サーバーを止めた状態でのメンテナンスが必要になることも多い。こうした場合でも、ロードバランサーを導入すると、アクセスできないサーバーにはリクエストを割り振らず、稼働しているサーバーだけに割り振るので、サイトのノンストップ運用が可能となる。

ロードバランサーの 基本的な動作概念

図2は、ロードバランサーを使用したサイトの基本的な構成図である。ロードバランサーは基本的にはルーターとサーバーの間に設置される。ロードバランサーを使った場合、ユーザーからはサーバーは1台に見える。この仮想サーバーのIPアドレスはロードバランサーに割り当てられ、ロードバランサーはNATを使って実際のサーバーにアクセスを割り振っていく。サーバーからの応答も同様にロードバランサーがNATによって変換し、ユーザーにデータを送信する。したがって、実際のサーバー側にはプライベートアドレスを割り当てるのが一般的だ。

またNATだけでなく、後述するようにHTTPヘッダーなどの情報によってアクセスを割り振るサーバーを決定する場合には、途中まではロードバランサーがHTTPサーバーの代理をするようになる。

各サーバーへのバランシング方式

各サーバーにどのようにアクセスを割り振っていくかには、338ページの図3に示すよう

いくつかの方式がある。もっとも単純なのが「ラウンドロビン方式」で、各サーバーに順番にアクセスを割り振っていく。この方式は各サーバーが同程度の性能である場合には有効だが、サーバーの性能にばらつきがある場合には性能の低いサーバーがボトルネックになってしまう可能性がある。こうした場合には、各サーバーの処理性能を考慮して、あらかじめ各サーバーにどのくらいの割合でアクセスを割り振るかを決定しておく「重みつきラウンドロビン方式」が有効となる。

さらに効率よくアクセスを割り振るには、各サーバーの状態に応じてそのつど最適なサーバーを選択する「ダイナミックモード」と呼ばれる方式を用いる。具体的には、アクセスがあった時点で処理している接続数をもっとも少ないサーバーにリクエストを割り振る方式や、サーバーからのレスポンスのスピードを計測してもっとも速いサーバーにアクセスを割り振る方式などがある。

さらに高度な方式としては、サーバーのCPUやメモリーの使用率をリアルタイムで監視する方式や、これらの要素を総合的に判断して割り振る方式などもある。また、あらかじめ各サーバーに優先順位を設定し、優先順位の高いものが使用できないときのみ次の優先順位のサーバーにリクエストを割り振る方式もある。これは、メインのサーバーが一杯のときに「ただいま混みあっております」といったメッセージを表示させる場合にも有効となる。

このほかの方式としては、HTTPヘッダーの要素により割り振り先を定義する方式がある。たとえばユーザーからリクエストのあったURLにより、拡張子がcgiならAのサーバー、その他であればBのサーバーといった使い方ができる。また、ユーザーの使用しているブラウザを示す「User-Agent」というヘッダーの違いにより、通常のサーバーと携帯端末用のサーバーに割り振るといった動作にも適している。

サーバーの状態を診断するヘルスチェック

ロードバランサーにより複数台のサーバーで運用していると、1台のサーバーがトラブルを起こしただけでは障害の発生に気が付きにくい。そこで、ロードバランサーは各サーバーがきちんと動作しているか、ヘルスチェックとよばれるテストを定期的に行うのが一般的である。もちろん、このヘルスチェックにパスしないサーバーには、ユーザーからのアクセスを割り振らないように制御することになる。

ヘルスチェックの方式としては、Pingチェ

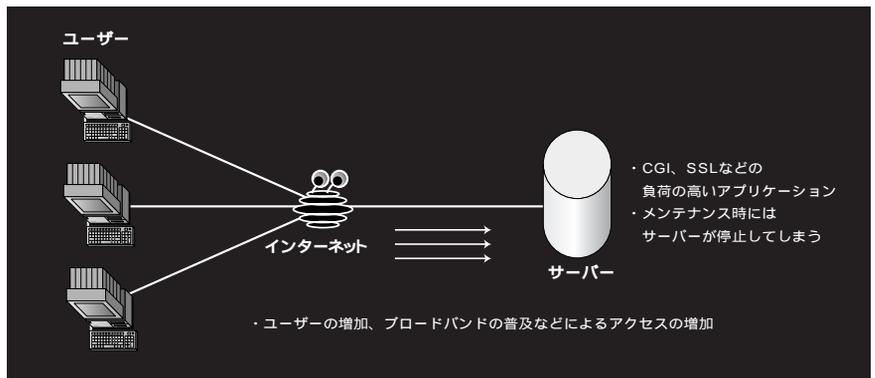


図1 インターネットサーバーの課題

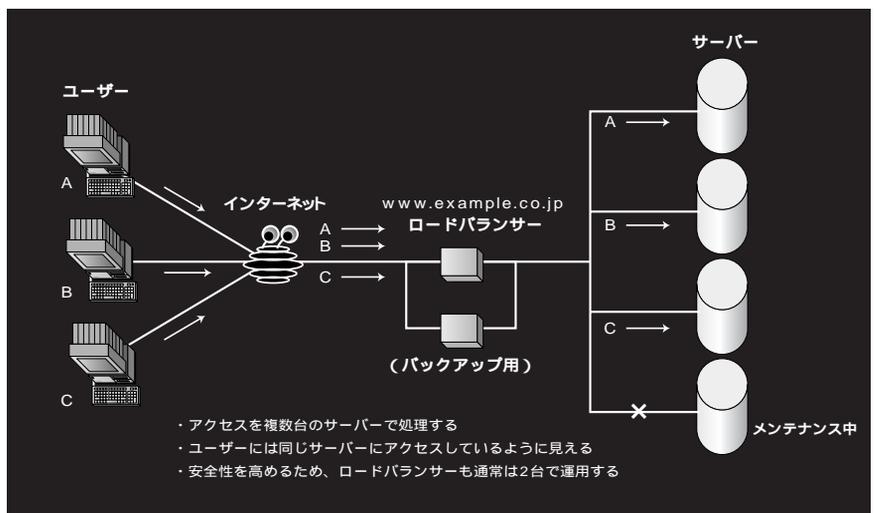


図2 ロードバランサーの設置例

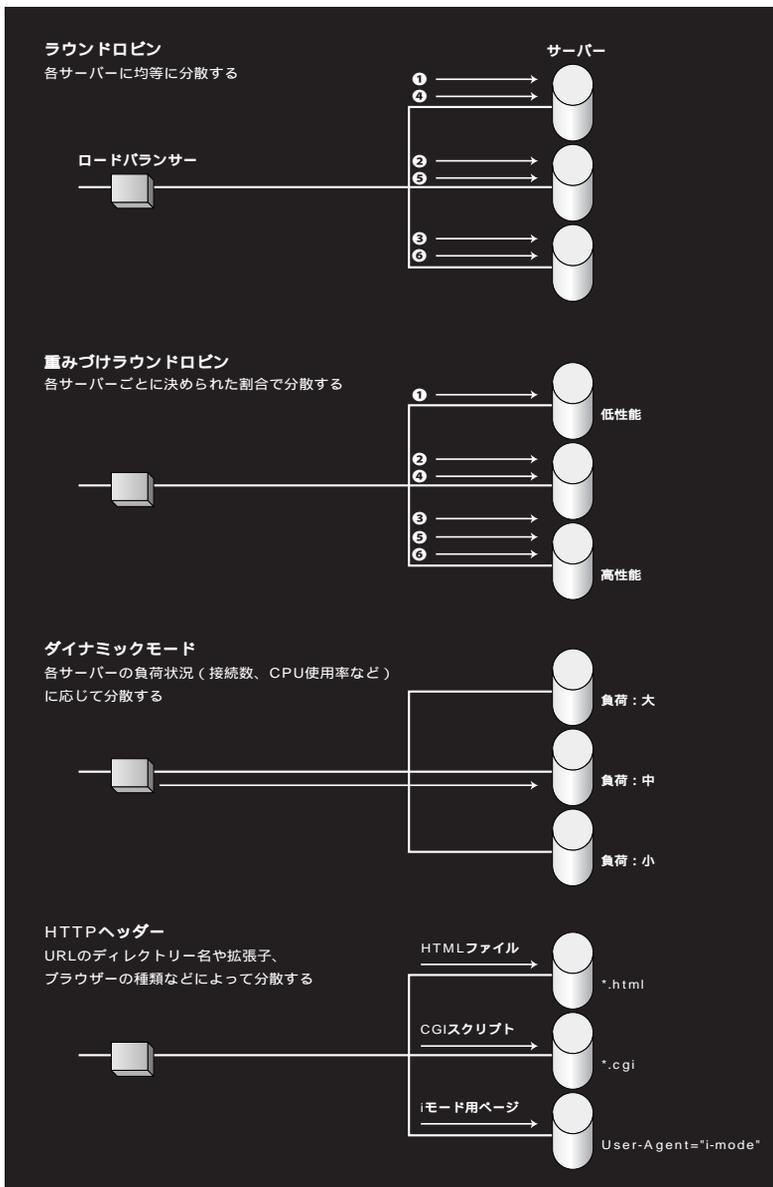


図3 負荷分散方式の種類

ックやPortチェックのほか、ウェブサーバーが正しく動作しているかを確認するコンテンツチェックや、アプリケーションが正しく動作しているかを確認するアプリケーションチェックなどがある（図4）。

アプリケーションチェックでは、ロードバランサーが仮想端末となり、RADIUSやIMAPなどのリクエストを出して、応答形式が正しいかをチェックする。さらに高度なチェックとしては、ロードバランサーがサーバーに対してアクセスのシミュレーションを行い、検索や注文、クレジット決済といった一連の流れがすべて正常に行われているかといったことをチェックする方法もある。

このほかには、各サーバーの処理数の上限を設定し、あるしきい値を超えた場合はそのサーバーをダウンしたものとみなし、しきい値以下になるまではそのサーバーにはリクエストを割り振らない方式もある。

同じサーバーで処理しつづける パーシステンス機能

ロードバランサーの目的はユーザーからのアクセスを複数台のサーバーに分散処理させることだが、サーバーによってはこうした処理により誤動作する可能性もある。たとえばショッピングサイトなどでは、商品名や価格などは共通のデータベースで管理していても、ユーザーの買い物途中の買い物カゴの中身の情報は、各サーバーがそれぞれ一時的に記憶しているのが一般的だ。したがって、こうした場合には、アクセスごとにユーザーが実際に接続するサーバーが変わってしまうと、誤動作の原因となってしまう。少なくともユーザーがすべての買い物を終えるまでは、1台のサーバーで処理しつづける必要がある（図5）。

こうした処理を実現するのが、同一ユーザーのリクエストを同一サーバーに割り振りつづける「パーシステンス」と呼ばれる機能だ。パーシステンスを実現する手法としては、①

ユーザーのIP アドレスが同一の場合には同一サーバーに接続する、② SSLのセッションIDが同一の場合には同一サーバーに接続する、③ cookieにサーバーIDを埋め込みロードバランサーがそれに従いリクエストを割り振る、といった方式がある。ロードバランサーにはSSLの処理を専用に行うSSLアクセラレーターカードなどのオプションが用意されている場合も多い。これによりロードバランサー側でもSSL通信の内容に応じた処理が可能になるほか、サイト全体としてもSSLに必要な処理が軽減されるといったメリットがある。

一方、iモード端末のように、ユーザーのIPアドレスやSSLのセッションIDが同一ユーザーからのアクセスであっても毎回変化する可能性もある。こうした場合には、URL内にサーバーIDを埋め込むという方式が、現在多くのサイトで使われている。ロードバランサーではHTTPヘッダーによるアクセスの割り振りと組み合わせて、同一のIDの場合には同一のサーバーに割り振るといった設定をすることになる。

また、メンテナンスを行いたいサーバーには新規のアクセスは割り振らず、現在処理中のパーシステンスによるアクセスだけを割り振り、すべての処理が終了した段階でメンテナンスを開始するといった、メンテナンスモードルールと呼ばれる機能も重要となる。

ブロードバンド時代に必須のロードバランシング技術

サーバー側からより積極的にロードバランサーをコントロールする方式としては、SOAP/XMLとよばれるプロトコルを利用する方式が発表されている。これは、Javaやネットワーク管理製品といったアプリケーションからロードバランサーの動作を動的に変更できるようにする技術である。ストリームサーバーへの同時接続数、回線の使用率などといった無数の選択肢から任意のものを管理者が選択し、それを判断基準としてユーザーからのリ

クエストを最適なサーバーに割り振れるようになる。

このほか、同一LAN内のサーバーだけでなく、複数地点のデータセンターで処理を行う広域分散方式なども、ユーザーに対していかに高速にレスポンスするかという点できわめて重要な要素となってきている。

今後、ロードバランシング技術には単純に負荷を分散させるだけでなく、このような複雑な処理も要求されるようになる。ともあれ、ブロードバンド時代を迎え、ユーザーからのアクセスは増大する一方で、さらにはウェブアプリケーションなどの機能の重要性もますます高まっている。こうしたアクセスを確実に処理するためにも、ロードバランシング技術の進化が必要であるのは間違いないところだ。

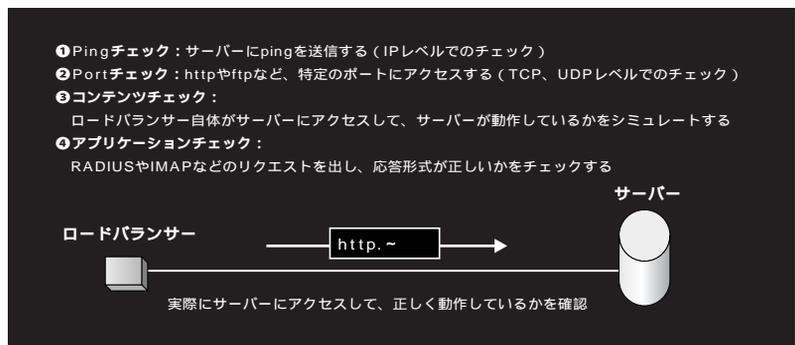
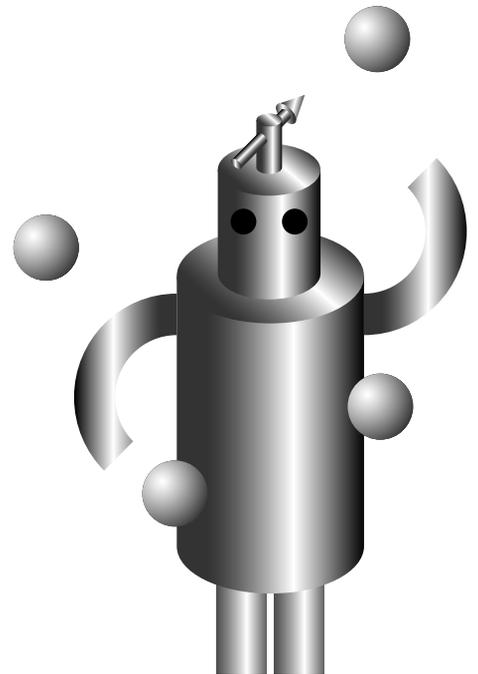


図4 ヘルスチェックの方式

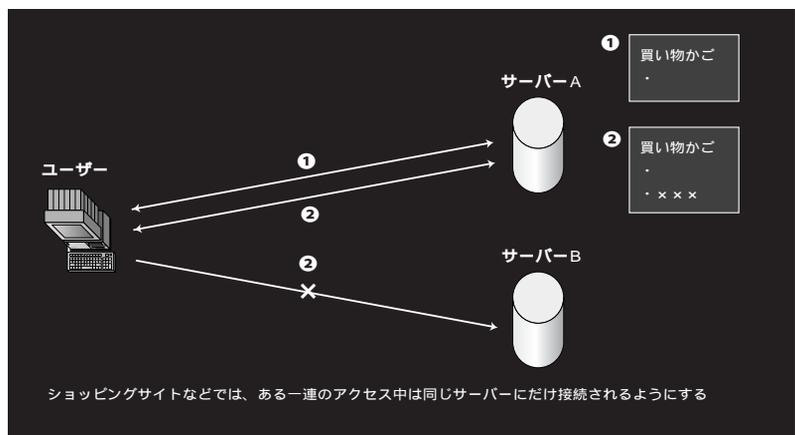


図5 パーシステンス機能





[インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

株式会社インプレスR&D

All-in-One INTERNET magazine 編集部

im-info@impress.co.jp