

分散 KVS 技術“okuyama”を検索エンジンに最適化

観測史上最速を誇る SaaS 型検索エンジン

“sui-sei”とは？

急激に増加するデータの 1 つに SNS にも代表されるライフログデータがあります。単純なデータ構造で半永久的に増え続けるライフログのようなデータの管理を容易にすることを目的として、株式会社 神戸デジタル・ラボ(以下、KDL)は 2010 年に分散 KVS 技術“okuyama”を独自開発しました。

一方で、KDL ではこれまで構築を中心に Web システムに関わるノウハウを長年に渡って培ってきました。

KDL が独自開発した技術と KDL の経験を組み合わせで生まれた “sui-sei”は、高速性を追求して実現された KDL 独自の次世代検索技術です。

検索エンジンの重要性

EC サイトにおいて、ユーザの離脱を回避するために不可欠な要素はレスポンスの速さです。例えば、EC サイト内検索において、ユーザが商品一覧を得るために、何秒待ってくれるでしょう？ Web ページの表示速度が 1 秒遅くなることで、CVR が 7%低下、顧客満足度が 16%も低下するという調査結果¹が報告されています。また、Web サイトの表示速度に不満を感じたユーザのうち、実に 79%はそのサイトでショッピングをしない、といった調査結果²も報告されています。これらの数値を見ても、いかに高速性が消費者にとって重要であるかお分かりになるでしょう。さらに、EC サイトの構成要素を調べてみると、約 60%のページで検索機能が動作しており、EC サイトで検索機能が占める割合が非常に高い事が分かります。

しかし、EC サイトを運営されている企業様で検索機能を独自に開発している企業様はほとんどいません。実際には、昨今の EC サイトでは検索機能にサードパーティが提供する検索エンジンを利用する傾向が強くなってきています。サードパーティの検索エンジンを利用するメリットは、初期費用が安価に抑えられる、最新の技術が利用できる、等が挙げられます。

それでは、今後さらに重要性を増すであろう検索エンジンが EC サイトへもたらす効果の一例をご紹介します。

● 大量な商品情報への利用

EC サイトの普及に伴い、各企業は売上 UP のために CVR を重要な指標として、その向上に取り組んでいます。例えば、詳細な商品情報を掲載する事で、ユーザへより丁寧に商品を説明し、購入意欲を促すような施策があります。また、ユーザ端末の多様化によって、各端末で最適な表示を実現するために、端末毎に商品データを用意する例も少なくありません。このように昨今の EC サイトでは、表示される情報量もその普及および利用形態の多様化とともに増加しています。情報量が多くなれば Web ページの表示速度は遅くなります。サイト運営者には、Web ページの表示遅延を避けることと、ユーザへの丁寧な説明をすることを両立するといった非常に高度な運営スキルが求められます。しかし、高速性に優れた検索エンジンを導入し、検索実行速度が高速化され

ば、サイト運営者は余計なストレスを感じることなく、ユーザに丁寧な説明をするためのコンテンツ作りに集中できるようになります。このように検索速度が高速化されることで、ユーザの商品ページの表示を待つといったストレスを解消するほか、サイト運営者にとっても本来の業務に集中できるといった効果も期待できます。

● 豊富な検索軸による効率的な検索への利用

EC サイトにおいて、ユーザが自社の商品に関心を持ったとしても、それが見つけられなければ購入に繋がらないため、検索は非常に重要な機能です。そのため、サイト運営者は、キーワードの他にナビゲーションを豊富に用意し、いかにサイト訪問者が要望に合った商品を見つけ易いようにするか工夫しています。具体的には、サイト運営者はユーザが色、サイズ、商品カテゴリ等を選び、キーワードなしでも効率良く商品に辿りつけられるように豊富で分かり易い検索軸を設定します。しかし、検索軸を増やすことによって検索の実行速度が劣化するといった弊害が生じます。そのため、サイト運営者は検索の実行速度に影響がない範囲で検索軸を決定する事を強いられており、ユーザの利便性は最大化されていないのが現状です。高速性、拡張性に優れた検索エンジンを導入すれば、サイト運営者は、速度の問題に悩まされることなく豊富な検索軸を用意する事ができ、ユーザへ高い利便性を提供することができるようになります。

● セールへの利用

EC サイトでは、実店舗と同様にセールが実施されています。EC サイトの場合、実店舗と異なりセール期間中は際限なくアクセスが集中します。また、実店舗では店舗に入りきらないユーザは行列に並ぶことで入店を待つことができますが、EC サイトでは不可能です。そのため、EC サイトにはそれを捌くことができる性能が求められ、それを担保するために現実にはセール期間中だけシステム保守体制を強化するといった運用がされています。従来の検索技術では、アクセス集中による高負荷分散が十分に考慮されていなかったため、セール期間中のアクセス集中によって EC サイトにアクセスできないといった問題が頻発しているためです。しかし、安定性に優れた検索エンジンを導入すると、高負荷状態でも平常時と変わらない安定性が提供されるため、保守体制を強化することなく、セールを迎える事ができます。このようにセー

ル対策をする必要がなくなれば、これまでセール対応として投資していた費用を本来の EC サイト改善に費やす事ができ、さらなる売上 UP に繋がっていくことも期待できます。

“sui-sei”の開発背景

EC サイトにおける商品データのような単純なデータ構造は、okuyama が管理できるデータ構造と非常に高い親和性があります。okuyama を利用した検索機能を KDL のクライアントである子供服 EC サイト運営企業様に導入したところ、検索の高速化と負荷分散の効果もあって前年対比約 3 倍の売上客数を獲得しました。KDL では、この実績を基に okuyama の構造を見直し、高速性を維持しながら拡張性を兼ね備えた検索エンジン“sui-sei”を開発しました。

SaaS 型検索エンジン “sui-sei”

KDL は、okuyama の持つ高速性、高負荷分散が検索機能として最大限発揮されるよう開発された okuyama4SS テクノロジーを搭載した“sui-sei”を開発しました。(図 1) okuyama4SS テクノロジーは、検索エンジンとしての okuyama の効率的な利用のほか、キャッシング機構や従来の KVS 技術では実現できなかった検索条件の複数指定やソート種類の指定、データ取得範囲の指定などの検索機能を実現しました。

“sui-sei”の開発には、長年 EC サイトの構築・運用を手がけ、okuyama の特性を熟知した技術者が複数名携わっており、“sui-sei”開発後も継続して okuyama のパフォーマンス・チューニングを実施しております。

● 高速性

“sui-sei”は、検索対象データ全てをメモリ上に展開し、検索要求に合致するデータをメモリ上に展開されたデータから探し出すことで、高速レスポンスを実現しています。具体的には、“sui-sei”は検索対象データ登録時に検索用インデックス情報を作成し、検索対象データと共に okuyama の管理するメモリ領域に格納します。格納した okuyama 管理メモリ上から WebAP の利用するメモリ上のキャッシュへ検索対象データと検索用インデックスを予めコピーしておくことで、“sui-sei”は検索時にディスクアクセスを伴わない高速かつ安定した検索を実現しています。

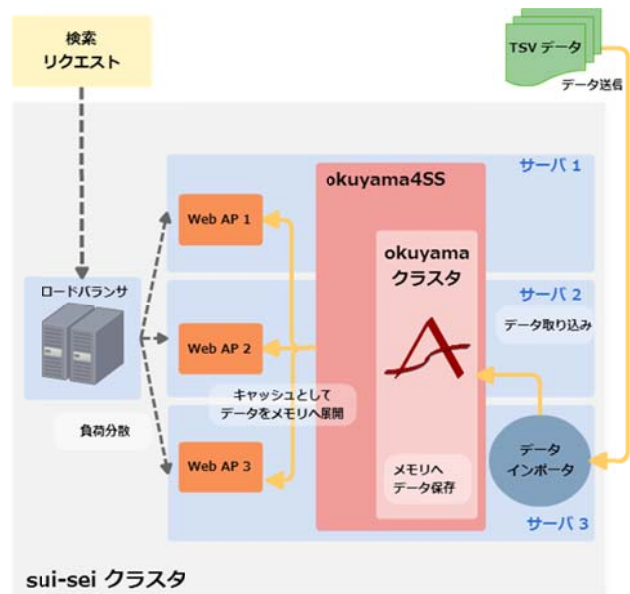


図 1. “sui-sei”の基本構成

● 拡張性

あらかじめ高負荷が予想される場合には、“sui-sei”クラスタに okuyama と Web AP をインストールしたサーバを追加することで、“sui-sei”はより多くの検索リクエストを処理可能となります。クラスタ内のサーバ数を変化させることで、“sui-sei”はスケラブルに負荷に対応します。

● 安定性

“sui-sei”が検索用インデックス情報と検索対象データを全てメモリ上に展開しておくことのメリットは、1 回の検索が高速化されることだけではありません。検索に必要な処理を全てメモリ上で行われるため、大量の検索リクエストが発生した場合にデータ取得に時間がかかりボトルネックとなっていたディスクアクセスの発生を避けられ、常にほぼ一定の検索速度を実現することが可能になりました。“sui-sei”は、セールなど大量のアクセスが発生するシーンにおいても安定した検索性能を発揮します。

● 可用性

“sui-sei”は 3 台のサーバ(インスタンス)で構成されており、データを格納する okuyama を 3 台にまたがってクラスタを構成する事で、メモリ上に保存された検索対象データは完全に冗長化されます。“sui-sei”クラスタ 3 台のうち 2 台が停止した場合でも、okuyama から検索対象データと検索用インデックスの取得が可能です。

検索機能の高速化に拘った理由

“sui-sei”の実力

“sui-sei”の性能検証に利用した環境の性能は表 1 に示すとおりです。

表 1. 性能検証の環境

	CPU
サーバ CPU	Xeon 2.90GHz
サーバメモリ	16 GB
通信回線	1GB 共有線

検証に用いたサンプルデータ仕様を表 2 に示します。

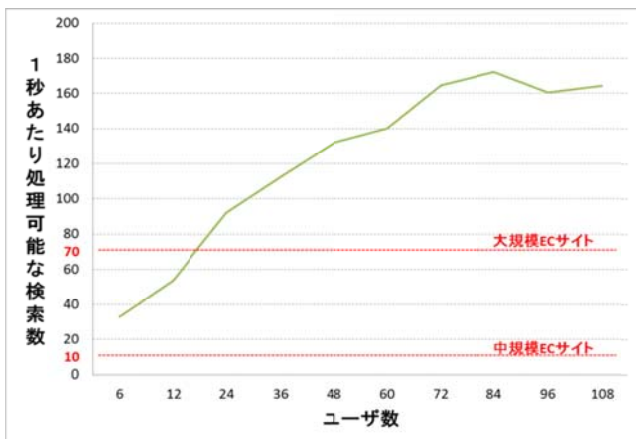
表 2. サンプルデータ仕様

商品 1 点あたりのデータ数	30
データ長	10 Bytes
インデックスデータ長	100 Bytes
商品点数	50,000 点

“sui-sei”の実力を測る検索実行速度の基準値として、現行の EC サイトでの実績からその規模に応じて中規模 EC サイト:10 検索/秒³、大規模 EC サイト:70 検索/秒⁴としました。

まず、同時アクセスユーザの増加に伴う処理可能な検索数の変化を検証しました。各ユーザは時間的に順次連続して検索リクエストを送信します。また、アクセスユーザ数は疑似的に増加させています。

中規模、大規模それぞれの基準値を圧倒的に上回るパフォーマンスであることがわかります。(図 2)



高負荷状態(108 ユーザが同時アクセス)が続いた場合で

も“sui-sei”はその高速性がほぼ劣化することなく、安定して稼働していることがわかります。

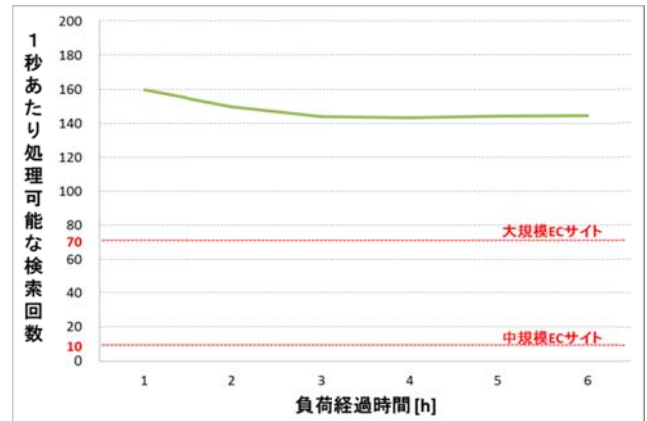


図 3. クラウド/SaaS での高負荷時の性能測定結果

“sui-sei”はクラウド/SaaS 製品として開発されましたが、オンプレミスへの導入も可能です。

検索エンジンの高速化を可能とした okuyama の特長

okuyama の主な特長は以下のとおりです。

● インメモリ・データグリッド

“sui-sei”ではインポートしたデータのメモリ展開に okuyama を利用することで、複数のサーバの持つメモリを 1 台のサーバのメモリであるかのように扱えるインメモリ・データグリッドを実現しています。インメモリ・データグリッドによって、データと負荷を分散することで高速なデータアクセスが可能なことに加えて、データはグリッド内で冗長化して保存されるため、グリッド内の特定のサーバが停止したことによるサービス全停止を防ぐことが可能です。また okuyama はクラスタへの動的なサーバ追加が可能なため、“sui-sei”の高いスケールビリティを実現しています。

市場への貢献

ソフトウェアのクラウド化によって拡張性はますます向上しています。今回は拡張性を維持しつつ、高速化を実現する検索エンジンを紹介しました。今後、ECサイトを初めとするWebシステムのリアルタイム化は加速する事が予想されます。その時、今以上に利便性が問われることは間違いないでしょう。そのためにも、KDL は日々お客様と接する中で、顧客インサイトを掴み、それを満たすような技術開発に取り組み、市場の活性化のために貢献してまいります。

関連情報

okuyama については、

<http://okuyama-project.com/ja/index.html> を参照ください。

³ 売上規模 5 億～30 億の中規模 EC サイトの秒間当たりの最大平均検索回数(当社実績値より)

⁴ 売上規模 30 億以上の大規模 EC サイトの秒間当たりの最大平均検索回数(当社実績値より)

¹ 2009, 米 Aberdeen Group,

<http://www.gomez.com/wp-content/downloads/aberdeen-gomez-best-in-class.pdf>

² 2009, 米 Forrester Research,

http://www.akamai.com/html/about/press/releases/2009/press_091409.html

本記事で紹介した性能に関するテストや評価は、特定のコンピュータ・システム、コンポーネント、またはそれらを組み合わせて行ったものであり、このテストによる KDL 社製品の性能の概算値を表しているものです。システム・ハードウェア、ソフトウェアの設計、構成の違いにより、実際の性能は掲載された性能テストや評価とは異なる場合があります。システムやコンポーネントの購入を検討される場合は、他の情報も参考にして、パフォーマンスを総合的に評価する事をお勧めします。

本記事の情報を参考に製品を購入され損害が生じても、弊社は何ら責任を負いかねますこと、予めご了承ください。

“sui-sei”の性能評価についてさらに詳しい情報をお知りになりたい場合は、下記の連絡先までご連絡ください。

“sui-sei”開発チーム sui-sei@kdl.co.jp

株式会社 神戸デジタル・ラボ 先端技術開発事業部
〒650-0034 神戸市中央区京町 72 番地 新クレセントビル
<http://www.kdl.co.jp/>

Copyright © 2013. KobeDigitalLabo Inc.,Ltd All rights reserved.