Harmonious Cloud
Hitachi vRAMcloud® Series

顧客情報を最大限生かす CRM ソリューションの実現

ビッグデータ超並列検索

ホワイトペーパー

2012 年 4 月 発行
Ver. 1.0
株式会社 日立製作所

＜本書で使用する用語＞

BCP ： Business Continuity Plan
CEP ： Complex Event Processing
CRM： Customer Relationship Management
DR ： Disaster Recovery
DWH： Data WareHouse
KVS ： Key-Value Store
LTV ： Life Time Value
M&A： Mergers and Acquisitions
RDB ： Relational DataBase
RDF ： Resource Description Framework
RSS ： RDF Site Summary（ or Really Simple Syndication）
SaaS ： Software as a Service
SCM ： Supply Chain Management
SFA ： Sales Force Automation
SQL ： Structured Query Language
TCO ： Total Cost of Ownership

＜留意事項＞

本ホワイトペーパーで記載する数値は、日立環境での計測データであり、諸条件により結果は異なります。

＜他社商標注記＞

・GemFire は、VMware,Inc.の米国およびその他の国における商標または登録商標です。日立製作所がソフトウェアを提供し、サポートを実施している製品です。
・その他記載されている会社名、製品名は、それぞれの会社の商標または登録商標です。

本ホワイトペーパーの想定読者

本資料は、(株)日立製作所　情報・通信システム社　金融システム事業部が提供するvRAMcloud®(※)を適用したソリューション「ビッグデータ超並列検索」についてまとめたものです。さまざまな企業、組織において情報システムの企画・構築に携わる方々を読者として想定しています。

(※)クラウド環境でビッグデータの利活用に最適な先端テクノロジーを提供するソリューション群。

概要

企業が扱う情報は、年々増加の一途をたどり、まさにビッグデータ時代が到来しています。しかし、ビックデータの内容は“玉石混交”のため、その中から本当に有効なデータを適切かつリアルタイムに把握できるかどうかが情報システムの重要な要件となっています。

本ホワイトペーパーは、増加し続けるビックデータを扱うために最適な先端テクノロジーを集結させたvRAMcloud®を適用することにより、

- コールセンタや対面販売などに代表される、〞お待たせできない“対応で活用できる高速顧客情報検索
- 購買履歴をベースに、クロスセル・アップセルなど攻めの営業活動につながる分析・シミュレーション
- お客様からの連絡情報をPush機能で渉外中の営業員に通達

が可能なCRMシステムについて、その特徴を記述します。

ビッグデータ超並列検索では、1,000万件の顧客情報を、あいまい検索・特定項目検索、範囲検索など、様々な検索条件の組み合わせで3秒以内に検索できることを実機検証しました。
コールセンタや対面販売など、お客様を決して待たせてはならない業務への適用が効果的です。

今後は、SFA機能、スマートフォン対応、Google Mapsとの連動などの機能拡充を計画しており、顧客管理を実施する際に最適なソリューションとして提供する予定です。

Contents

はじめに

1980 年代までの高度経済成長期は、モノやサービスを作れば作るほど売れた時代でした。しかし、バブル崩壊後“失われた 10 年”以降の売れない時代においても、企業には事業継続と拡大を追求し続けることが求められています。常に新たなマーケットが目の前に広がっていた時代とは異なり、既存顧客に対する継続的なアプローチ、新規顧客層への効率的なアプローチが、ますます重要な課題となってきています。

既存顧客へのアプローチは、お客様のライフサイクルを考慮した LTV(Life Time Value)の最大化に注力し、さらなる深耕、適切なサービスの提供により、企業と既存顧客間での Win-Win 関係を構築していくことが必要です。

一方、新規顧客層へのアプローチは、既得意顧客の属性（住所、年収、志向、趣味など）をベースに、類推機能（リコメンド機能）からご要望にフィットしそうな提案を行うことが期待され、新規顧客→既得意→ファン層へと、顧客ステージをシフトしていくことが求められています。

こうした背景のもと、本ホワイトペーパーでは、

- 既存顧客の深耕、適切なサービスの提供
- 営業活動の見える化（プロセス可視化）
- 新規顧客開拓へ向けたアプローチに活用できるセグメント化

など、従来技術よりも効率的に実現しながら、マス・マーケティングではない「個客」価値最大化に向けた顧客情報管理を可能とするビッグデータ超並列検索について記載します。

特に、コールセンタや対面販売など、クイックレスポンスが必要となる業務を想定し、高速検索の実現を中心とした実証実験結果をまとめています。

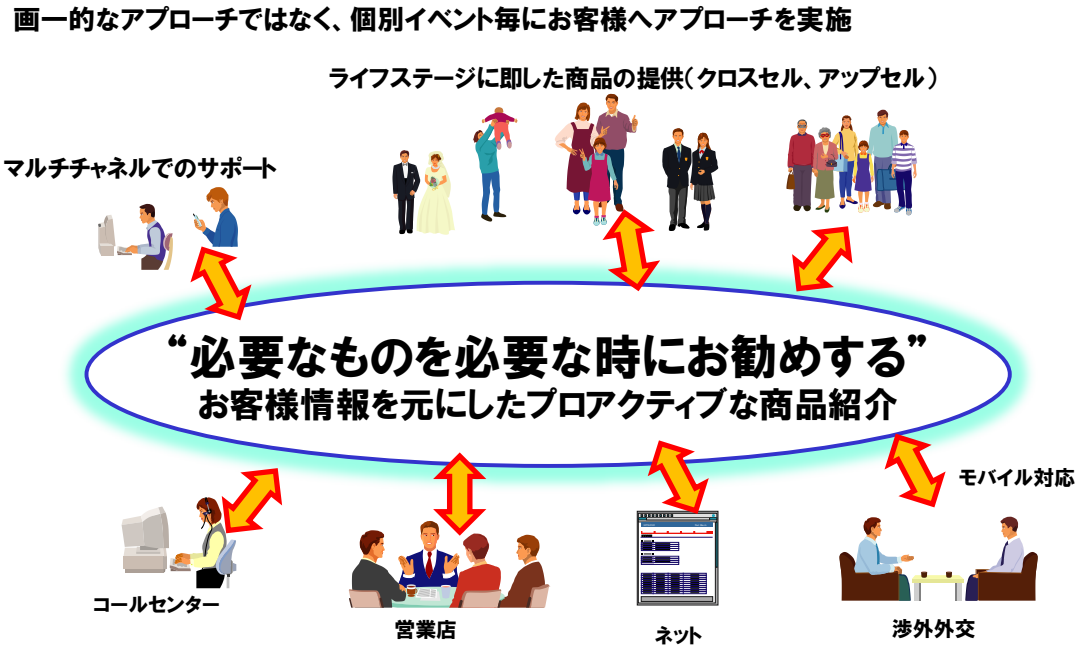

図 1　CRM コンセプト

CRM とは

CRM（Customer Relationship Management）とは、「情報システムを利用して企業が顧客と長期的な関係を築く手法」と定義されています。主に、営業員やその管理者が利用し、営業員が担当する既存顧客や新規に開拓する見込客を管理します。

近年の CRM システムには、SFA（Sales Force Automation）機能、分析機能などが追加され、いわゆる営業フロント業務に対して顧客情報を一元管理し、有効な営業施策の立案をサポートする機能が提供されています。

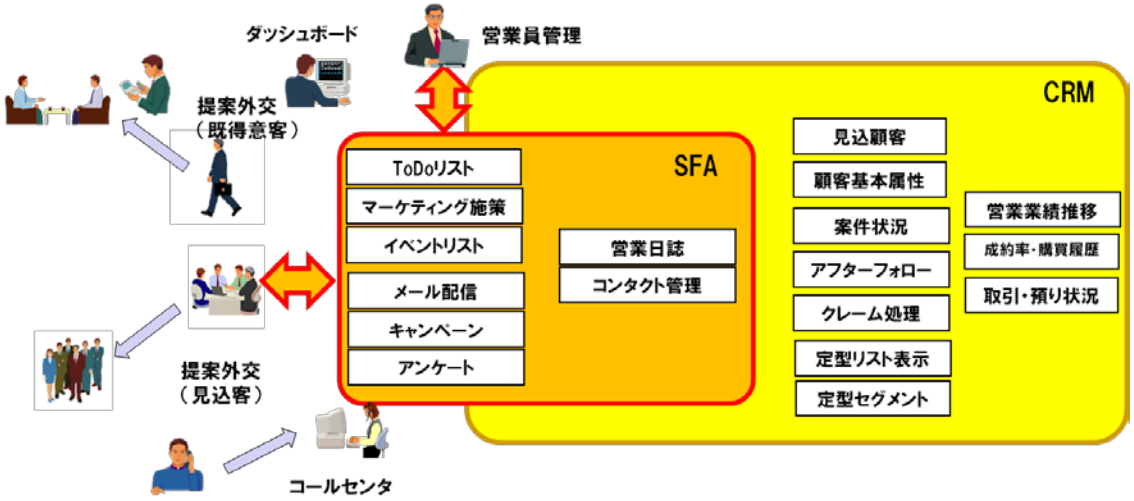

図 2　CRM/SFA で提供する機能

CRM 導入状況

CRM システムは、SFA との連携でマーケティング支援機能が追加され、既存顧客情報管理にとどまらず、見込客管理や商談管理などの顧客接点業務を一貫して支援できるようになってきています。

一方、CRM システムを顧客情報のマスタ化（統合）を主目的に導入したものの、有効活用ができていない状況も散見されます。例えば、顧客氏名や住所などの基本情報とともに、嗜好・志向のセグメント情報や、購買履歴データなどを膨大に保管してはいるものの、膨大すぎてデータの有効活用ができていないというケースです。そこで、これらの情報を分類・活用するために DWH(Data WareHouse)との連携や、ダッシュボード機能拡張が追加されるようになってきました。

また、CRM システム構築では近年、スクラッチ開発ではなくソフトウエアベンダが提供する CRM パッケージの適用が増えています。CRM パッケージは単体導入だけでなく、クラウド環境を活用した SaaS（Software as a Service)形態で提供されることも一般的となり、中小規模企業でも導入しやすくなっています。

パッケージ製品の限界

パッケージ製品は、どのベンダの製品も提供機能に差はなくなっていますが、総じてビッグデータをパッケージに取り込むバッチ処理、大量情報を対象にした検索処理で性能チューニングを必要としています。

また、ビッグデータの検索・登録に既存 RDB を適用すると、以下のような性能面や維持管理費用の観点から限界があることも指摘されています。

① データの分布状態に偏りがある場合、インデックスが十分に機能しない
② データ量の増減に対し、簡単にスケールアップ・スケールダウンできない
③ 検索を最適化するために作り込みが多い場合、バージョンアップやパッチの適用ができない

さらに、CRM パッケージの多くは欧米企業をモデルに作成されているため、パッケージ標準の組織テンプレートでは、日本の複雑な組織構造（兼務、統括など）が表現できず、カスタマイズが必要です。

CRM パッケージを導入する企業規模も、大規模になればなるほどカスタマイズ開発の割合が多くなる傾向があります。導入当初は問題がないものの、時間経過とともに処理件数が増加し、性能が出なくなるといったケースも少なくありません。常に一定の検索性能を保つためには、ハードウェアの増強などで追随する必要があり、ビッグデータ時代に対応するには、莫大なコスト追加を覚悟しなくてはなりません。

目指すべき方向性

日立が提供するビッグデータ超並列検索は、従来型のパッケージ製品では限界がある処理性能やスケーラビリティの向上を低コストで実現し、ビッグデータ時代に求められるシステム要件とお客様の期待にお応えします。

増加の一途をたどる顧客情報を利活用するためには、常に最適なパフォーマンスを実現できる CRM システムが必要です。お客様からの問合せ・クレームなどをコールセンタ・店頭などで受け付ける際、オペレータは、お客様の詳細な状況を確認しながら最適な対応を行う必要があります。そのためには、ビッグデータの中から当該顧客情報を瞬時に検索できる機能が必要です。

さらに、攻めの営業活動を行うには、ビジネストレンドに沿ったサービスの追加や企業独自の傾向分析などの機能が必要です。特に、新サービスの追加は、他社に先駆けてサービス提供を開始することがビジネス展開で優位に立つための重要な条件の 1 つになっています。

そのためには、検索性能が高速であるだけでなく、サービス追加に対して素早く、また低コストで対応できるシステムの採用が必要になります。

こうした要件を満たすには、様々なビッグデータを扱うために最適な先端テクノロジーを集結させ、高速検索機能、類推機能を備えた vRAMcloud®の適用が最善策と考えられます。

ビッグデータ超並列検索の狙い

vRAMcloud®の目的

分散処理技術、仮想化技術は、黎明期から成熟期へと移ってきています。

vRAMcloud®は、企業内で増加し続けるビックデータを扱うために最適な先端テクノロジーを集結させたソリューション群で、高速検索及びPush型の配信を分散キャッシュにより実現しています。また今後は、ダッシュボード機能を実現する予定です。

図 3　vRAMcloud ソリューション群

ビッグデータ超並列検索の狙いと実現の考え方

ビッグデータ超並列検索は、RDB では実現できなかったビッグデータに対する高速検索、動的再配置を実現し、将来の情報量増加、サービス追加、耐障害性にも対応できるスケーラビリティを実現しています。

CRM パッケージあるいはスクラッチによる開発は、RDB を前提として設計されています。近年のハードウエアスペック向上により、費用対効果が改善され、投資額を抑えながらメモリ DB 化できるようになりました。RDB における高速化の手法は、主としてアクセスするインデックス情報やテーブル情報をメモリに常駐させる一方、パーティション分割による分散アクセスで実現されます。しかし、急激なデータ増加に対してはパーティションの再設計が必要なほか、特定店舗の顧客数が多いなどデータに偏りがある場合、RDB では一定した性能が満たせません。

これに対し、分散キャッシュ技術を適用したビッグデータ超並列検索では、分散キャッシュによる高速検索を実現するだけでなく、データの増減が発生した場合、動的にスケールアウトができ、処理ノードの構成変更ができるため、データ量やデータの偏りに依存しない検索性能を確保できます。

また、ビッグデータ超並列検索では、安価なハードウェアを並列処理させることで性能を確保しているため、データが増加し、スケールアウトを実施した場合でもハードウェアのコスト増加を抑えることができます。

これにより、100 万件オーダのデータはもとより 1,000 万件を超えるデータでも一定性能を確保でき、"Small Start, Grow Big!"という要件を満たすことが可能となります。

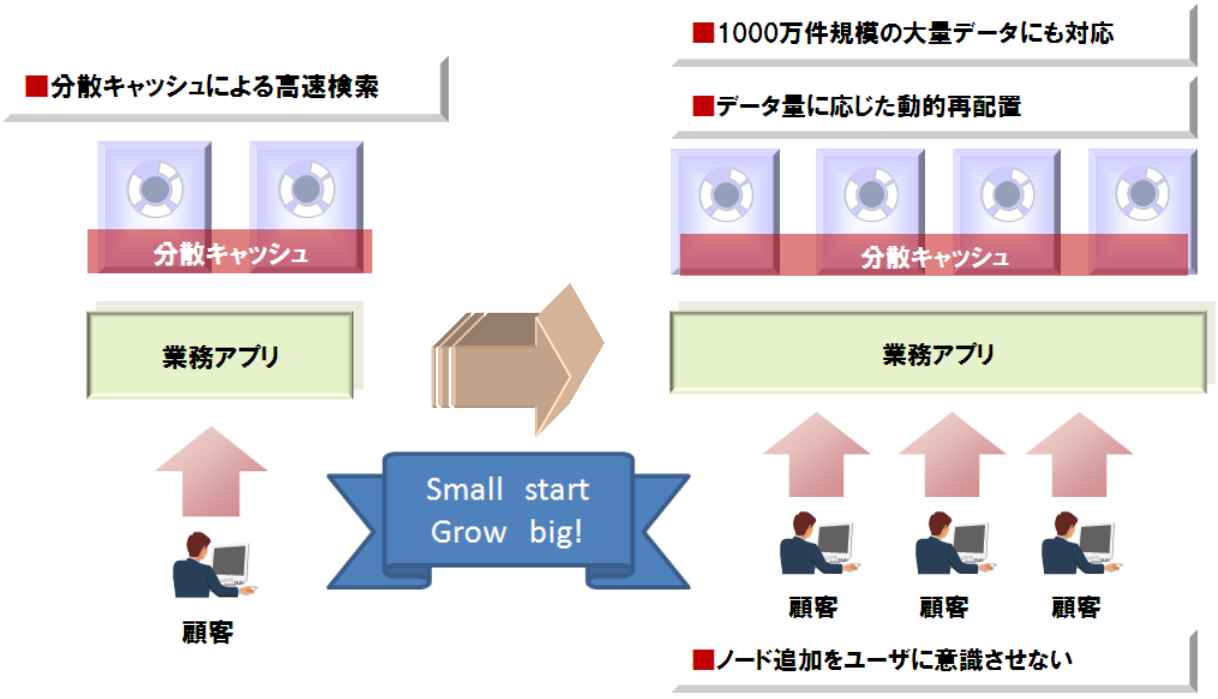

図 4　小規模導入から大規模へ拡張可能

高速検索

ビッグデータ超並列検索は、KVS（Key-Value Store）型の分散キャッシュ GemFire を使用しており、従来の RDB を前提とした場合と比べ、処理の高速化を実現しています。また、vRAMcloud®共通プラットフォームと組み合わせることで、従来 KVS の弱点といわれていた、あいまい検索・範囲検索・複合検索といった汎用検索においても、KVS の性能を活かした高速検索を実現しています。

データ検索をインメモリで並列実行するため、母数が 1,000 万件を超えるようなデータ検索において、RDB ではチューニングしても 10 秒以上かかった処理が、下記に示すように 3 秒以内で実現でき、RDB と比べ最大で 10 倍以上の高速検索が可能となります。

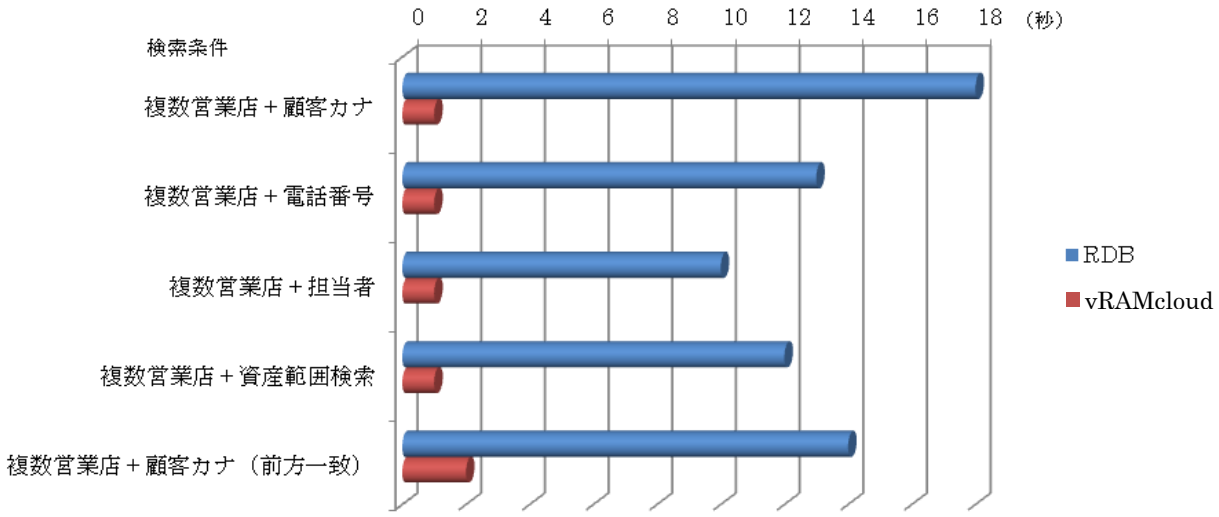

図 5　RDB と vRAMcloud ビッグデータ超並列検索ソリューション適用時の検索性能比較

スケーラビリティ

ビッグデータ超並列検索は、高いスケーラビリティ(拡張性・柔軟性)を確保したアーキテクチャとなっています。

業務要件の追加などに伴うデータ項目の変更に対して、データをオブジェクトとして管理していること、GemFire の機能であるオブジェクトのバージョニング機能を用いることで、システムを停止することなくデータ項目の変更が実施できるなど、柔軟に対応することが可能となります。

データ項目の変更を RDB で実装することを考えた場合、カラム追加やテーブル追加での対応になるため、システムを停止することが必須であることや、既存プログラムの影響有無把握、影響発生時の修正、変更対象外に影響がないことを確認するテストなどを考えると、柔軟な対応が困難だといえます。新サービス提供までに時間がかかりビジネスチャンスを逃すことにもつながりかねません。

ビッグデータ超並列検索では、動的にノードを追加しても、各ノードの負荷を平準化させるデータリバランス・アーキテクチャを採用しているため、急激なデータ増加にも柔軟に対応できます。

さらに、ノード追加・データリバランス中にもユーザからのリクエスト処理は継続して実行できるため、エンドユーザにノード追加を意識させることなく、処理能力の増強が行えます。このノード追加・データリバランスは 3 分程度で実施できるため、迅速な対応を行うことが可能です。

ノード追加・データリバランス実施時の検証結果として、処理ノードを 2 ノードから 4 ノードへ動的拡張した場合の結果を示します。

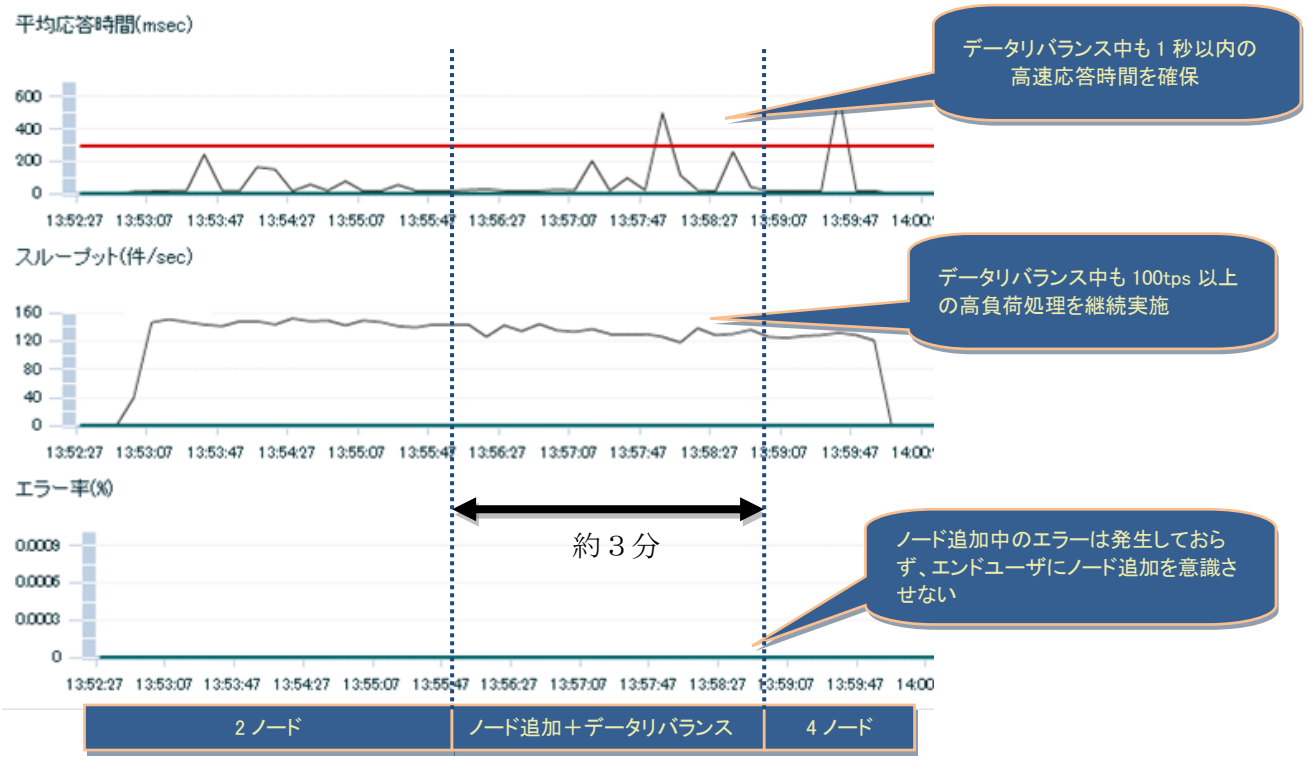

図 6　データリバランス状況とリバランス時のスループット遷移

上図に示す通り、ノード追加・データリバランス実施中においても、以下の特徴があります。

・平均応答時間　：特出した遅延がなく、応答性能を維持しながらノード追加可能

・スループット　：100tps 以上を維持しており、スループットを維持しながらノード追加可能

・エラー率　：0%＝ユーザにエラーリターンがなく、ユーザにノード追加を意識させない

また、ノード追加・データリバランスも 3 分程度で実施できており、迅速な対応が可能であるといえます。

また、RDB ではパーティション分割を行うことで、並列処理を実現していましたが、ある特定キーにデータが集中してしまう場合には、パーティション分割の効果が得られず、性能劣化が発生していました。ビッグデータ超並列検索では、同一キーであっても分割して配置することが可能で、特定キーにデータが集中した場合においても、並列処理を実現することで性能確保が可能なアーキテクチャとなっています。

現行システムからの移行

ビッグデータ超並列検索は、サービスレベルを一定に保ちながら移行コストを最小化し、将来の増強が容易に行える環境を実現することで TCO(Total Cost of Ownership)の最適化に寄与します。

ビッグデータ超並列検索は、既存のシステムをリプレースする際に

① アプリケーション移行：スクラッチで作成した既存システムを活用し、分散キャッシュ化する

② データ移行：パッケージ製品をリプレースする際、あるいは新規導入する際に、ビッグデータ超並列検索に利用できるように既存のデータを移行する

といった移行ソリューションを提供します。

アプリケーション移行

スクラッチ等で構築した CRM システムが既に稼働している場合、そのシステムは企業に最適な機能要件を備え、他社とは異なる顧客情報管理の強みが具現化されていると想定されます。

ビジネス形態の変化や M&A などによって、業務実態とシステムが乖離している可能性も考えられますが、企業の“強み”として実装された機能は、競争優位を保つためにも継続することが必要です。

このため、ビッグデータ超並列検索では、既存アプリケーションで存続させる業務ロジックを変更せず、DB I/O を変更するだけで移行できます。

既存システムの実装言語により移行方針は異なりますが、分散キャッシュを利用するために、通常 SQL として発行しているロジック（Pro*C などにより記載）を Java インタフェースに置き換えることになります。

置き換えについては、SQL 構文解析を実施しますが、ツールでの支援が可能なため、既存のアプリケーションをビッグデータ超並列検索に容易に移行することができます。

データ移行

分散キャッシュ上では、データ管理はオブジェクト単位で行います。このため、RDB でテーブルとして定義していた情報を、分散キャッシュ上にオブジェクトとして移行する必要があります。

テーブル構造からオブジェクト構造への変更は簡単です。オブジェクト構造の設計は必要となるものの、データ自体は一旦ファイルとして出力した既存情報を、分散キャッシュ上に一気に読み込むことが可能です。

分散キャッシュの情報を永続化（ディスクに保管）する場合、RDB をデータストアとして活用することも可能です。これに対応するテーブル設計は、既存システムのデータモデルを踏襲して実現できます。

実機検証結果

検証概要

CRM を例にビッグデータ超並列検索の性能について、実機検証を行いました。

具体的には、キャッシュサーバを 2 ノードから 4 ノードへと変化させ、クライアントからラッシュをかけることで、スループットやレイテンシの値を計測しました。

また、ラッシュ実行中の処理ノードの動的追加(スケールアウト)、データリバランスについても検証を実施しました。本検証では、下図のように営業準備に用いる顧客情報検索を対象に検証しています。

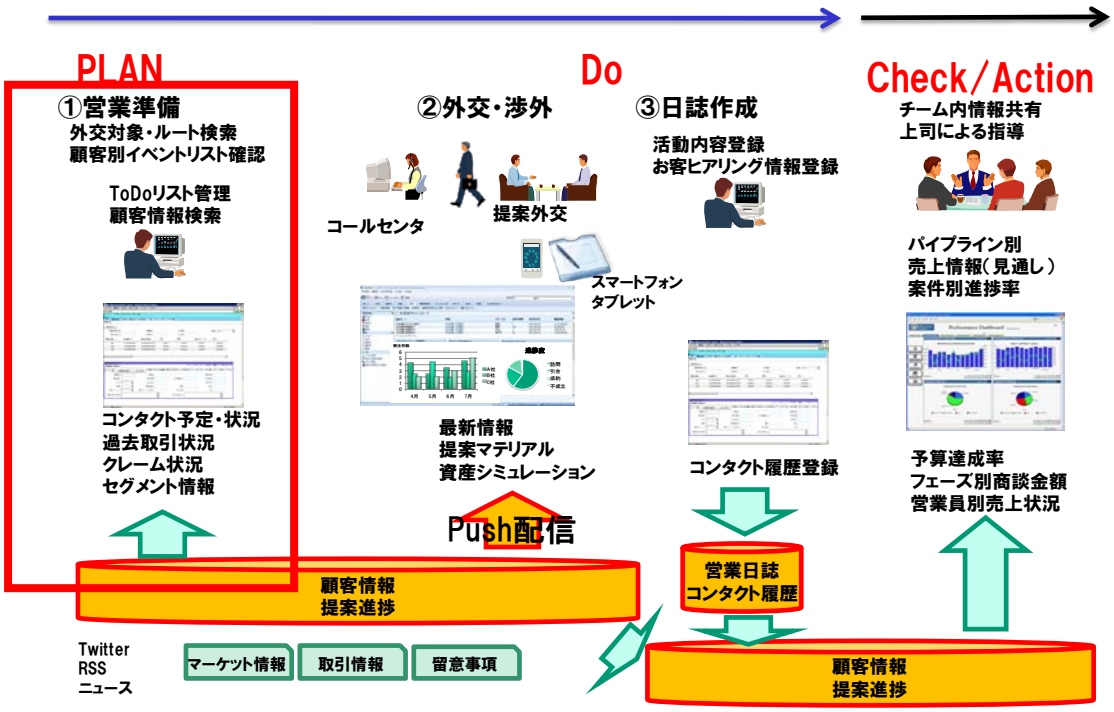

図 7　実機検証対象（赤枠）

検証環境

本検証における各マシンの役割およびサーバ環境を以下に示します。
各マシンの役割は、下記 2 種類です。

- 構成管理をするロケータ及びクライアントの役割
- Web/AP サーバ、キャッシュサーバの役割

各仮想マシン定義は 1 物理マシン上に 1 台としました。

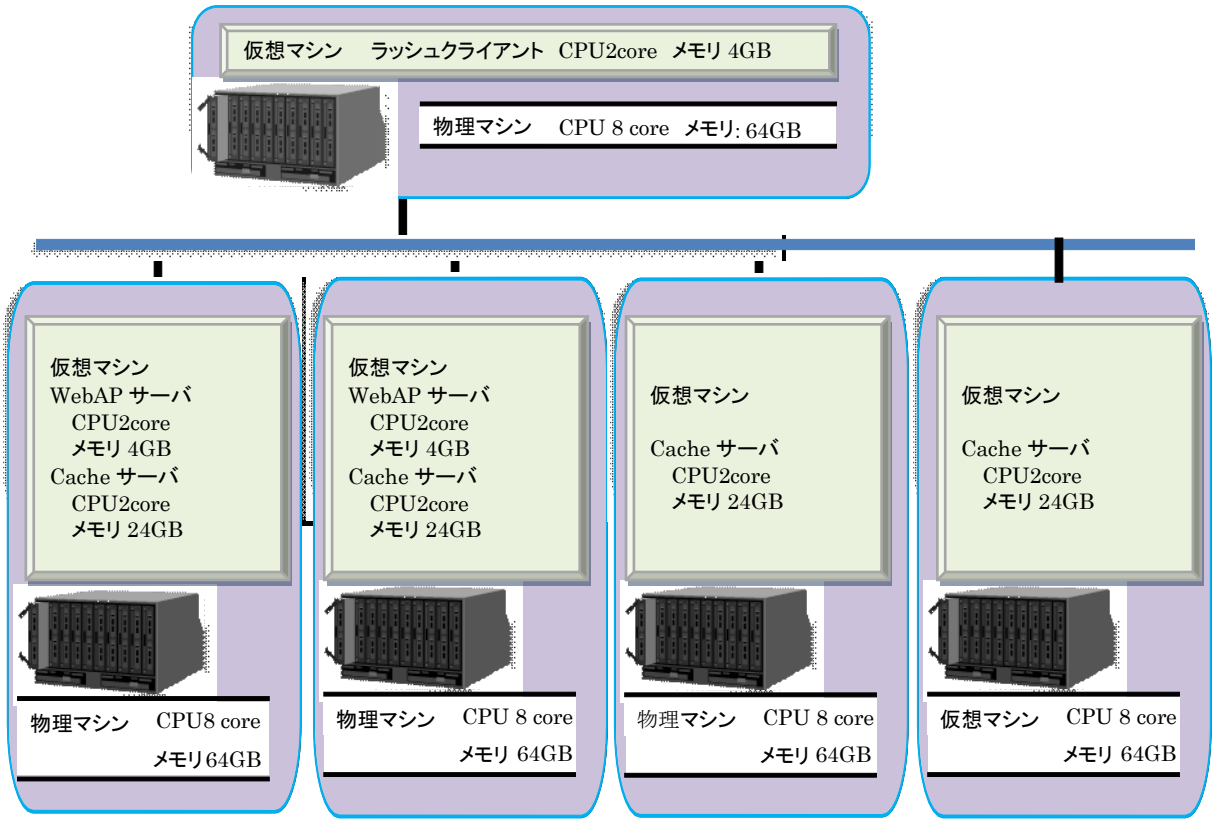

図 8　ハードウエア構成図

ハードウェア構成、ソフトウェア構成は以下の通りです。

表 1　ハードウエア構成一覧

#	項番	ハードウェア	員数	備考
1	ブレードサーバ	BladeSymphony320	5	
2	CPU	Xeon L5630(2.13GHz,4Core,12MB)	2	1 台あたり 8Core
3	メモリ	DDR3 8GB	3	1 台あたり 24GB
4	ディスク	147GB(SAS10,000rpm)	2	
5	ネットワーク	1Gbps	1	
6	共有ディスク	BR20 146GB(SAS 10,000rpm)	6	

ソフトウェア構成

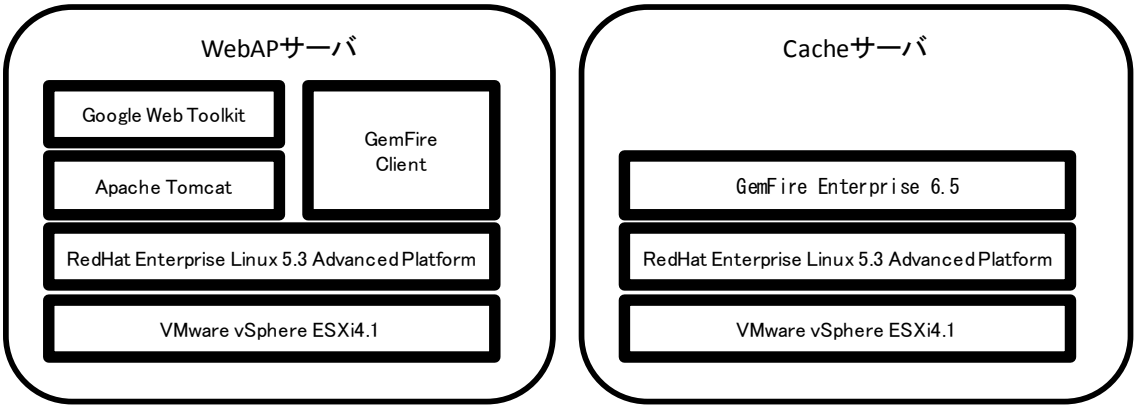

図 9　ソフトウエア構成図

検証方法

実証実験の測定方法として、Web ブラウザの代わりにオープンソースの負荷ツールである「Jmeter 2.4」を使用しました。検索条件を含んだ HTTP リクエストを送信することでシステムに負荷をかけ、スループット及びレイテンシを計測しました。なお、クライアントからのスレッド数は 12 に固定して検証しています。

また、実証実験実施中にノードの動的追加、データリバランスも行いました。

CRM システムの検索画面では、「営業店番号」や「口座番号」、「顧客名」などいくつもの検索項目が存在します。これらの項目につき、複数条件を組み合せた検索が可能です。また、検索の種類として「前方一致」「後方一致」「完全一致」「範囲検索」などの、あいまい検索が可能なため、検索条件を策定する際は、事前にこれらの組み合せを検討する必要があります。

本検証では、完全一致検索、中間一致検索、範囲検索、後方一致検索＋範囲検索、中間一致検索＋前方一致検索の 5 パターンの検索条件をランダムに実行しました。

検証結果

平均スループットはキャッシュサーバ 2 ノードで約 240tps、4 ノードで約 400tps となり、ノード数が増加するにしたがって、スループットが上昇しました。

1検索ごとのレイテンシは、2 ノードの時には約 30ms、4 ノード時には約 25ms であり、ノードの増加には影響がなく、非常に高速な結果が得られました。また、ラッシュ実施中のノードの動的追加（2 ノード→4 ノード）についてもノード追加後の構成での稼動が確認できました。

以上のことから、ビッグデータ超並列検索は負荷増加時にノードを動的に増加させることにより、スループット・レイテンシともに一定の性能を確保できると言えます。

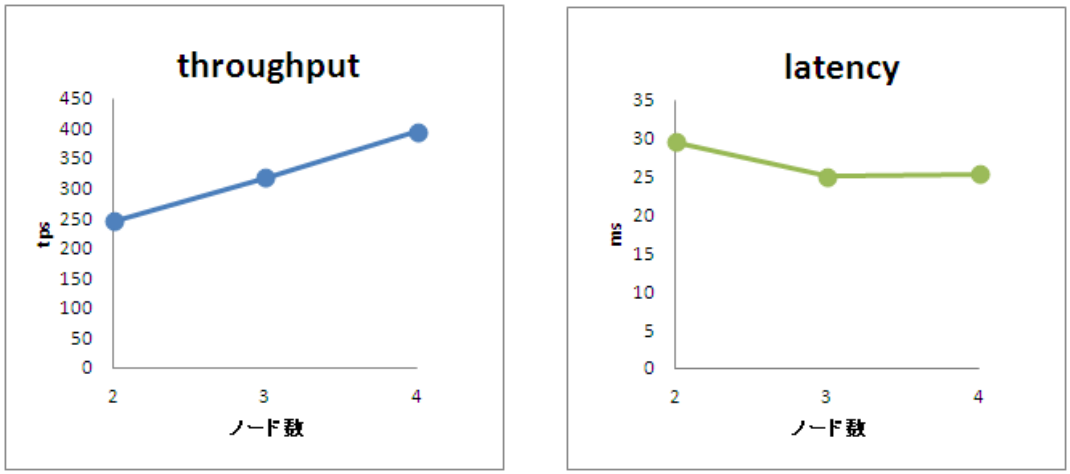

図 10　ノード数増加によるスループット、レイテンシーの変化

想定サービス提供形態(Easy To Start)

ビッグデータ超並列検索では、既存システムの移行、導入時のカスタマイズ、構築支援などをソリューションメニュー化しています。

動的にサーバ群を追加・削除できるという特性から、まずは部署内のような小規模トライアルで効果を確かめ、しだいに企業単位での大規模な利活用まで拡張していく、“Small Start, Grow Big!”を目指しています。導入時には、お客様要件に合わせたカスタマイズや、他システムとの連携インタフェース開発を行うSIサービスもオプションとしてご提供します。

導入に関しては、専門コンサルティング部隊を配しCRM導入に関する業務要件定義から、導入に必要となる技術的なコンサルテーションまでを行います。

下記は現在検討中のサービス提供モデルです。

ソリューション提供モデル

日立は、ビッグデータ超並列検索で実現しているCRM機能をご提供します。

ソリューションメニュー

日立からご提供するソリューションメニューは以下のようになります。

表 2　ソリューションメニュー一覧

#	メニュー	概要	費用	備考
1	導入診断 コンサルテーション	ソリューション導入可否、及び費用対効果の算定	個別見積	
2	導入コンサルテーション	ビッグデータ超並列検索導入支援(要件定義、カスタマイズ実施)	個別見積	
3	システムインテグレーション	ビッグデータ超並列検索と連携が必要な既存業務システム連携を含め、高品質システム開発を支援	個別見積	
4	既存システムマイグレーション	既存システムをビッグデータ超並列検索化するマイグレーション支援	個別見積	

まとめ及び今後の計画

vRAMcloud®は、データ量が増減しても一定の検索性能が保てるように、スケーラブルに対応できるため、今後のビッグデータ時代に適したソリューションと言えます。

まとめ

ビッグデータ超並列検索は、ビッグデータ時代に柔軟かつスケーラブルに対応できるソリューションであり、以下のような特徴を備えています。

- CRM のひな型を提供し、そのままでも利用可能
- 小規模から始められ、大規模までのスケーラビリティを確保できる
- インメモリ技術により、高速に情報連携でき、BCP 対応も標準装備（詳細はホワイトペーパー：M/F COBOL オンラインバッチ参照）

今後の計画

現在、ビッグデータ超並列検索は、顧客情報検索機能に特化したシステムとして、実機での性能評価を行っています。下図のシステムでは、見込客管理、顧客基本情報など CRM の基本となる機能、及び営業日誌、コンタクト管理の SFA の基本機能を実装済みです。

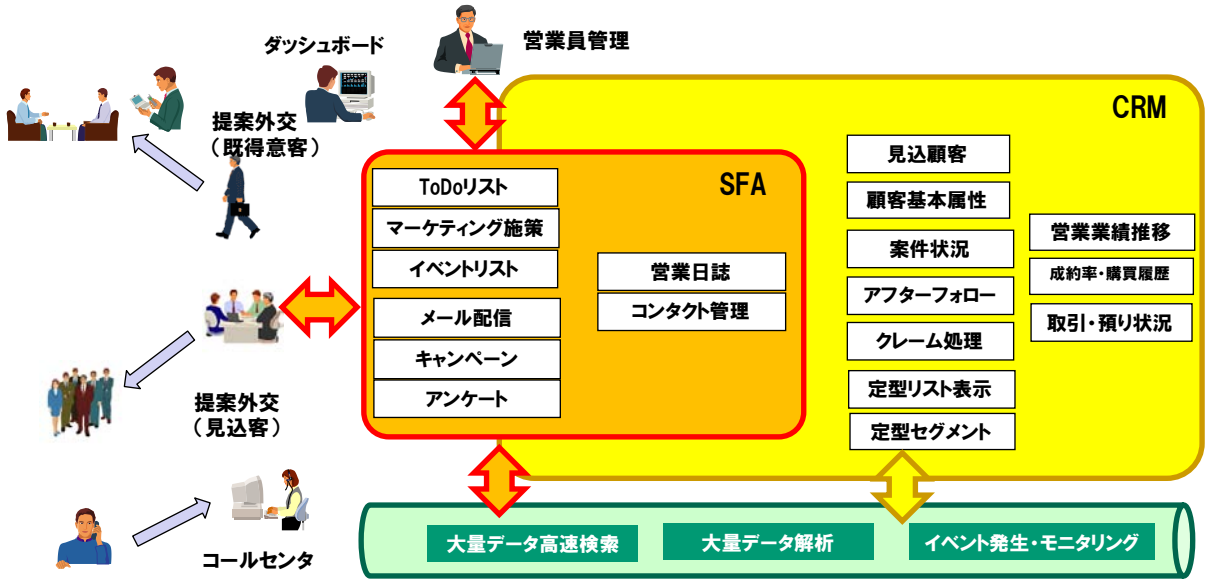

図 11　ビッグデータ超並列検索ソリューション将来像

今後の機能拡張として、下記の項目を予定しています。

✓ SFA 機能の拡充
ToDo リストの表示や、マーケティング施策（リコメンド情報）の生成、営業支援機能の充実

✓ スマートフォン対応
提案・外交に営業員が持参するスマートフォンに、Push 機能で顧客情報を提供する機能の実装

✓ カスタマイズ環境の提供
ビッグデータ超並列検索を導入する際のカスタマイズ環境・開発維持環境の提供

以上