

ビッグデータの戦略的活用を支援する vRAMcloud

Cloud-based Solution for Utilizing Big Data

村谷 宏明
Muratani Hiroaki
石合 秀喜
Ishiai Hideki

勝瑞 雅也
Shozui Masaya
高田 真
Takada Makoto

アクセスログやソーシャルメディアへの書き込みなど、間断なく流れ込んでくる膨大な情報を集計・分析し、企業の競争力強化につなげたいというビッグデータの活用ニーズが高まっている。

大量かつ多種多様で、爆発的に増加するという特徴を持つこれらのデータを活用するため、日立グループは、ビッグデータの戦略的活用を支援する「vRAMcloud」を提供している。vRAMcloudは、分析力、超並列高速処理、自律運用という三つの要件、すなわち「Power」・「Speed」・「Cost」を備えており、ビッグデータを活用した新たな価値創造を支援するソリューションである。

1. はじめに

近年、企業では、これまで扱ってきた業務データだけでなく、「ビッグデータ」をいかに活用できるかが、業績を左右する一つの重要な要素になっている。

ビッグデータは、従来のRDB (Relational Database) システムでは効率的な処理ができないデータと概括することができ、以下の特徴がある。

- (1) データが大量である。
- (2) 多種多様な種類がある。
- (3) 爆発的に増加する。

ビッグデータの例としては、Facebook^{※1)} や Twitter^{※2)} をはじめとする SNS (Social Networking Service) への書き込み、RFID (Radio-frequency Identification) タグや各種センサー、携帯電話やスマートフォンに組み込まれた GPS (Global Positioning System) データなどが代表的である。

ここでは、日立グループがビッグデータの戦略的活用を支援するソリューションとして開発した「vRAMcloud」の特徴と適用事例について述べる。

※1) Facebook, Facebookロゴは、Facebook, Inc.の登録商標である。

※2) Twitterは、Twitter, Inc.の登録商標である。

2. ビッグデータ活用を推進するための施策

ビッグデータから価値を見出して競争力につなげることの重要性は、多くの企業で認識され始めている。しかし、ビッグデータの活用モデルは企業によってさまざまであり、定石が存在するわけではない。ビッグデータ活用の目的や方法が確立している企業は少ないのが現状である。

この課題に対し、日立グループは、ビッグデータ活用を推進するための施策として以下の3点が重要であると考えている。

- (1) トライアルアンドエラーで仮説検証を繰り返し、最適な分析モデルを導き出すこと
- (2) ビジネスチャンスを逃さないために、大量データの高度な分析結果を短時間で得られること
- (3) データ量が爆発的に増加した場合に、低コストで対応できること

3. vRAMcloudの「Power」・「Speed」・「Cost」

vRAMcloudが備えている「Power」・「Speed」・「Cost」について述べる。

3.1 vRAMcloudの構成要素と特徴

vRAMcloudは、KVS (Key-value Store) 型の分散キャッシュ製品であるVMware^{※3)} vFabric^{※3)} GemFire^{※3)} (以下、GemFireと記す。)を中心に、ストリームデータ処理製品uCosminexus Stream Data Platform(以下、uCSDPと記す。)やグリッドバッチ製品uCosminexus Grid Processing Serverなど、およびそれらの製品を連携・活用するフレームワークから構成されている。

ビッグデータ活用を推進するための施策と、vRAMcloud

※3) VMware, vFabric, GemFire は、VMware, Inc.の米国および各国での登録商標または商標である。

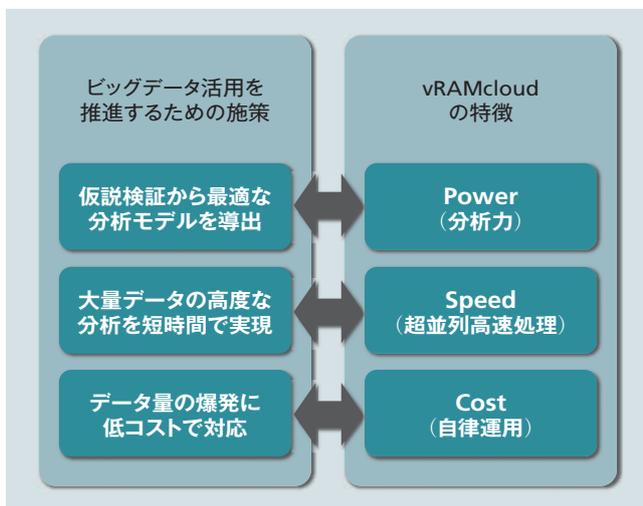


図1 | ビッグデータ活用を推進するための施策と「vRAMcloud」の特徴
ビッグデータ活用を推進するために重要となる三つの施策をvRAMcloudで実現する。

の特徴である「Power」・「Speed」・「Cost」の関連を図1に示す。

「Power」は、仮説検証から最適なモデルを検討できる処理エンジンを充実させ、データの特性や量に応じて適切な処理エンジンを適用できることである。

「Speed」は、インメモリ処理と超並列処理により、オンライン処理、バッチ処理を高速で行えることである。さらに、データ構造を柔軟に変更でき、ビジネス環境の変化に早期に対応し、サービスインを短期間で実施できる。

「Cost」は、初期コスト、運用コストを最小化できることである。「Power」・「Speed」の改善によって1台当たりの処理能力が劇的に向上し、大規模なシステム構成が不要になる。また、自律的にリソースを拡張するためピーク時に合わせたサイジングをする必要がなく、初期コストを削減できる。システムを停止せずに自律的にリソース拡張できるため、データ量の爆発にも低い運用コストで対応可能である。

3.2 vRAMcloudの「Power」

vRAMcloud は、大量で多種多様なデータに対して適材適所で適用できる分析エンジンを提供する。データの特性や分析の要件によって分析エンジンを組み合わせ、データ抽出、リアルタイム分析、テキスト解析、バッチ分析などを実現する。

分析エンジンを適材適所で適用する事例について、SNS分析システム、ログ収集・分析システムを例に述べる。

3.2.1 SNS分析システム

SNS分析システムは、SNSに書き込まれたテキストの内容を解析し、単語の関連性分析やコメントの評価分析を行うシステムである。このシステムの特徴は、SNSを対

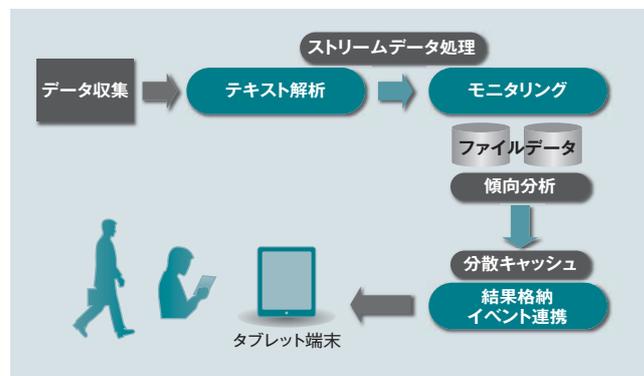


図2 | SNS (Social Networking Service) 分析システムの概要
テキスト解析、モニタリングをストリーム処理で、結果格納、イベント連携を分散キャッシュでそれぞれ実現する。

象にするため、大量なデータや多種多様なデータを扱うことである。

システム概要を図2に示す。SNSに書き込まれた文字データを収集し、テキスト解析エンジンを用いて分析する。単語の出現数やポジティブ/ネガティブ分析結果の集計、特定キーワードのモニタリングにストリームデータ処理エンジンを使用している。また、分析結果の格納やタブレット端末へのイベント連携には、分散キャッシュエンジン（超並列エンジン）を使用する。これら複数のエンジンを組み合わせることでSNS分析システムを実現している。

3.2.2 ログ収集・分析システム

ログ収集・分析システムは、ログをリアルタイムに収集・分析するシステムである。ログは、障害発生時やキャンペーンなどのイベント発生時には急激に出力量が増加するが、障害原因の調査やアクセス分析に必要なログをロスしないように保存する必要がある。また、障害メッセージの傾向分析やアクセスランキング作成をリアルタイムで行うことができる。このシステムの特徴は、高頻度に発生するログを格納し、リアルタイムに分析することである。

システム概要を図3に示す。障害発生時などにシステムから大量に発生するログを漏れなく格納するため、分散

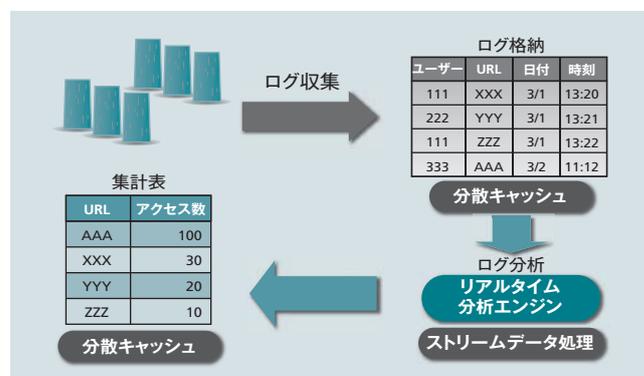


図3 | ログ収集・分析システムの概要
ログの格納に分散キャッシュを、分析にリアルタイム分析エンジンをそれぞれ適用している。

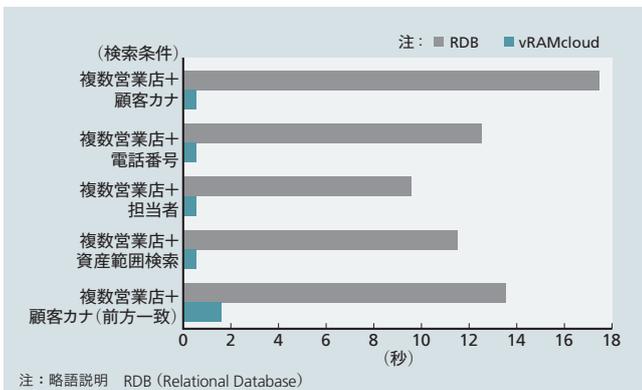


図4 | CRM (Customer Relationship Management) システム適用時の検索性能比較

RDBシステムとvRAMcloudの検索性能比較を示す。vRAMcloud は、RDBに比べて最大10倍の検索性能を実現している。

キャッシュエンジンを用いている。また、ログの傾向をリアルタイムに把握するためにストリームデータ処理エンジンを適用し、ログ収集・分析を実現している。

3.3 vRAMcloudの「Speed」

vRAMcloudは、オンライン処理のレスポンスとバッチ処理の高速化を、インメモリ処理と超並列処理によって実現する。

高速検索エンジン、分散キャッシュを利用してオンライン処理を高速化した事例を図4に示す。口座数1,000万件を超えるCRM (Customer Relationship Management) システムをモデルに検証した。

同図のRDB検索時間は、複数営業店を対象に、検索条件別に最大限チューニングした結果である。顧客数が少ない営業店1店舗での検索は、3秒以下という性能が出ていた。しかし、検索対象が大規模店や複数の営業店の場合で検索母数が増加すると、チューニングしても検索処理に10秒以上を要し、性能を満たすことができなかった。

一方、vRAMcloudでは、超並列処理が可能な分散キャッシュエンジンを適用しているため、複数営業店を対象とした検索母数が多い検索でも、検索処理を1秒未満で実現できた。この事例で適用している高速検索エンジンは、KVSが不得手とする範囲検索やあいまい検索を高速に行うエンジンである。

高速検索エンジンにより、顧客カナ前方一致検索や資産範囲検索といったCRMシステムの主要な検索処理において、最大10倍以上の高速検索が可能となった。

バッチ処理を高速化した事例を図5に示す。この事例は、800万件のデータ集計バッチ処理を超並列処理で行ったものであり、実機で28秒という、従来の360秒と比べて10倍以上の高速化を達成した。超並列処理によって並列実行が可能のため、さらに多重度を増加させれば、50



図5 | グリッドバッチ適用による処理時間

グリッドバッチの実行多重度を増やすことで、処理時間を10倍以上高速化した。

倍以上の高速化を実現できる見込みである。

検索結果、分析・集計結果を高速に得られるため、条件を変えながら仮説検証を繰り返すことができ、新たな価値発見につながる。

vRAMcloudは、サービスインまでの期間短縮も実現できる。従来のRDBでは、既存のアプリケーションへの影響調査、新サービスリリースに関する既存システムのデグレードテストが必要であった。vRAMcloudではオブジェクト指向の開発に加え、オブジェクトの世代管理をフレームワークで実現するため、デグレードテストを最小限にでき、大幅な期間短縮が可能になる。アプリケーション自体の開発は、新規フレームワークを活用することにより、通常のスラッシュ開発の場合に比べて $\frac{1}{2}$ 程度の工数・期間で実現できた事例もある。

対象システムとその特性、およびvRAMcloudで提供しているエンジンの対比を表1に示す。SNS分析やログ収集・分析では、テキスト解析、分散キャッシュ(GemFire)、リアルタイム分析エンジン(uCSDP)を利用した。CRMシステムや大量データを対象にしたバッチ処理は、高速検索によるデータ特定の高速化、グリッドバッチを適用した超並列処理により、「Speed」が向上して処理の高速化が可能となる。

表1 | システムの特性からのエンジン選択

SNS分析システム、ログ収集・分析システム、CRMシステム、バッチの特性から、最適なエンジンを選択する。

特性	エンジン	エンジン				
		GemFire	uCSDP	高速検索	グリッドバッチ	テキスト解析
SNS分析	大量データ	○				
	多種多様		○			○
ログ収集・分析	高頻度データ	○				
	リアルタイム分析		○			
CRM	大量データ	○				
	高速検索			○		
バッチ	大量データ	○			○	
...						

注：略語説明 uCSDP (uCosminexus Stream Data Platform)

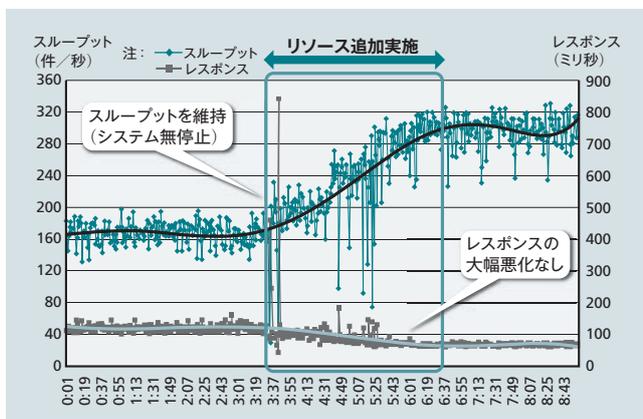


図6 | オンライン中の動的リソース追加

枠内でリソースを追加する。スループット、レスポンスを悪化させずにオンライン中のリソース追加が可能である。

このように、データ利活用目的により、適材適所で最適なエンジンを適用できる。

3.4 vRAMcloudの「Cost」

vRAMcloudでは、「Power」・「Speed」が改善するため、マシン1台当たりの処理能力が向上する。そのため、小規模なシステム構成でも大量のデータを処理することができ、初期コストを削減できる。また、動的なスケールアウトなどの自律運用により、TCO (Total Cost of Ownership) に占める割合が高い運用コストの低減も可能となる。

処理量の急激な増加に対して動的にリソースを追加し、処理能力を拡張した事例を図6に示す。爆発的なデータの増加に対して、システムを停止することなく、スループットを増加させ、レスポンスを低下させずにリソースを追加できる。リソース追加時にレスポンスの悪化がないため、システムを利用しているエンドユーザーに対して、画面操作速度の低下を感じさせない。

リソースの拡張時には、データのリバランスも自動的に行うため、RDBで実現する場合に発生するデータ分散設計やチューニング、システムの停止・再起動が不要であり、運用コストの削減が可能である。

4. おわりに

ここでは、日立グループがビッグデータの戦略的活用を支援するソリューションとして開発した「vRAMcloud」の特徴と適用事例について述べた。

vRAMcloudが提供する「Power」・「Speed」・「Cost」という特徴により、ビッグデータの戦略的活用を推進できる。

実世界から間断なく流れ込んでくる膨大な情報をリアルタイムにさまざまな角度から分析すれば、より迅速な意思決定やリスク管理、現場状況の把握などにつながる。また、バッチ処理の短時間化によるオンラインサービス時間の増

加などで、RDBコストの大幅削減だけでなく、顧客満足の向上やビジネスチャンスの増大にも寄与する。

現在、あらゆる業務分野において、このようなスピードを付加価値とした新ビジネスの創出が強く求められている中、vRAMcloudはこれらの課題を解決できる有効なソリューションである。

vRAMcloudは、小規模なシステム構成で新サービスを提供でき、事業規模の拡大とともにリソースを追加できる。サービスインまでの時間を短縮し、また、小規模でサービスを開始できるため、仮説検証を繰り返すことで「Small Start, Grow Big」が実現する。

vRAMcloudは金融機関以外のさまざまな業種にも適用可能であることから、適用業種の拡大も視野に入れ、開発を進めていく。さらに、分析エンジン強化のため、多様なデータから将来予測に基づく意思決定やビジネスの最適化までを支援するBA (Business Analytics) 機能の開発に取り組んでいく。

執筆者紹介



村谷 宏明

1988年日立製作所入社、情報・通信システム社 金融システム事業部 金融システム第三本部 第一部 所属
現在、vRAMcloudソリューションの企画に従事



勝瑞 雅也

1991年日立製作所入社、情報・通信システム社 金融システム事業部 金融システム第三本部 第一部 所属
現在、vRAMcloudソリューションの企画に従事



石合 秀喜

1987年日立製作所入社、情報・通信システム社 金融システム事業部 金融システム第三本部 第一部 所属
現在、vRAMcloudソリューションの開発に従事



高田 真

2000年日立製作所入社、情報・通信システム社 金融システム事業部 金融システム第三本部 第一部 所属
現在、vRAMcloudソリューションの開発に従事