

ビッグデータ利活用を支える オープンミドルウェア

Open Middleware for Big Data Utilization

田中 慎一
Tanaka Shinichi

稲場 淳二
Inaba Junji

梅田 多一
Umeda Taichi

社会の実態を表す生データ、いわゆるビッグデータを解析して新たな「知識」や「気づき」を得ることがITの進歩によって可能となってきた。ビッグデータの利活用では、データの特性や用途に応じて処理の要件が異なる。日立グループは、データ活用シーンを「リアルタイム処理」、「蓄積・検索」、「集計・分析」に分け、各活用要件に応じたデータ処理基盤として、ストリームデータ処理基盤、インメモリデータグリッド、高速データアクセス基盤、およびグリッドバッチソリューションをオープンミドルウェアとして提供し、ビッグデータ利活用を支えていく。

1. はじめに

社会の実態を表す生データを活用し、未来の価値創造を支えるために、データの特性や用途に応じたデータ処理基盤が必要となる。日立グループは、データ利活用の課題とその解決技術を、現場の事実から未来の業務に不可欠な情報を生成する技術「Field to Future Technology」として体系化し、この技術を実装したミドルウェア製品やサービスの研究開発ならびに機能強化を推進している。

ここでは、ビッグデータ利活用を支えるオープンミドルウェアについて述べる。

2. ビッグデータ利活用を支えるオープンミドルウェア

2.1 ビッグデータのライフサイクル

ビッグデータ利活用においては、実世界で発生する大量のデータを「収集」、「リアルタイム監視」、「蓄積」、「抽出」、「分析」から業務にフィードバックする一連のプロセスが重要な鍵を握る（図1参照）。

特に、データを収集し、整理・分析するだけにとどまらず、「分析」によって得られた「知識」や「気づき」を、業務活動、情報提供としてデータ配信し、業務改善としてのフィードバックを形成することがデータ利活用による価値

創生に最も重要なポイントである。

2.2 ビッグデータ活用要件

先に説明した一連のプロセスを、データ活用要件として三つのデータ活用シーンに整理する。一つ目は、多様な場で提供されるサービス群を通じて発生する大量のデータをリアルタイムに監視して即時処理する要件、二つ目は、収集したデータを効率的に蓄積して必要なデータを必要なときに素早く容易に検索する要件、そして三つ目は、抽出したさまざまなデータを使用して集計や分析をする要件である（図2参照）。

そうして得られた集計・分析結果から、「知識」や「気づき」を社会の現場にフィードバックすることが可能となり、リアルタイムに今の状況が分かり、自分に合ったサービスが得られることになる。

2.3 ビッグデータ利活用を支えるオープンミドルウェア

ビッグデータを有効に活用するためには、活用要件に応じたデータ処理基盤が不可欠である。日立グループは、

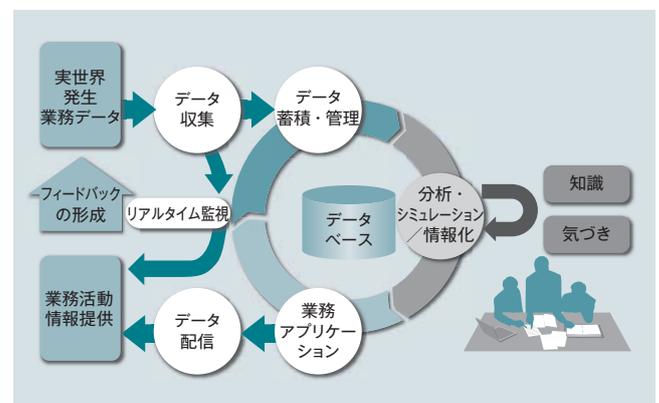


図1 | ビッグデータのライフサイクル

データ発生から知識創造に至るデータ処理のプロセスを示す。

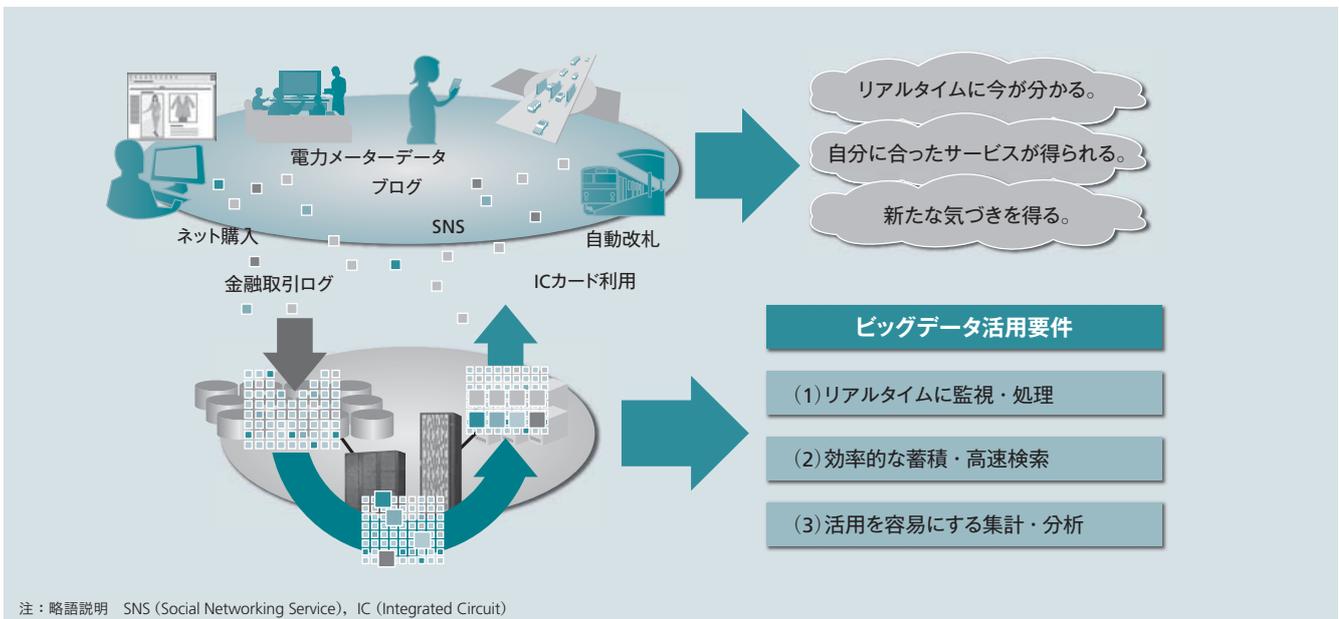


図2 | ビッグデータ活用要件

リアルタイム処理、蓄積・検索、集計・分析がビッグデータ活用の3要件である。

「Field to Future Technology」技術群をソフトウェアとして実装し、オープンミドルウェアとして実現している（図3参照）。

大量の情報をリアルタイム処理する要件に対して、発生する情報を即応的に処理することが可能なストリームデータ処理基盤 uCosminexus Stream Data Platform と、大量のデータを高速に処理するインメモリデータグリッドを構築可能な uCosminexus Elastic Application Data store を提供している。

また、収集したデータを効率的に蓄積し、必要なデータ

を必要とときに、素早く、容易に検索する要件に対しては、高速データアクセス基盤 Hitachi Advanced Data Binder プラットフォームを提供している。

さらに蓄積したこれらのデータを使用して、集計・分析する要件に対しては、COBOL (Common Business Oriented Language) プログラムなどの既存資産を活用したバッチ高速化ソリューションとしてグリッドバッチソリューションを提供している。

新たにラインアップに加えたインメモリデータグリッドと高速データアクセス基盤を次に述べる。

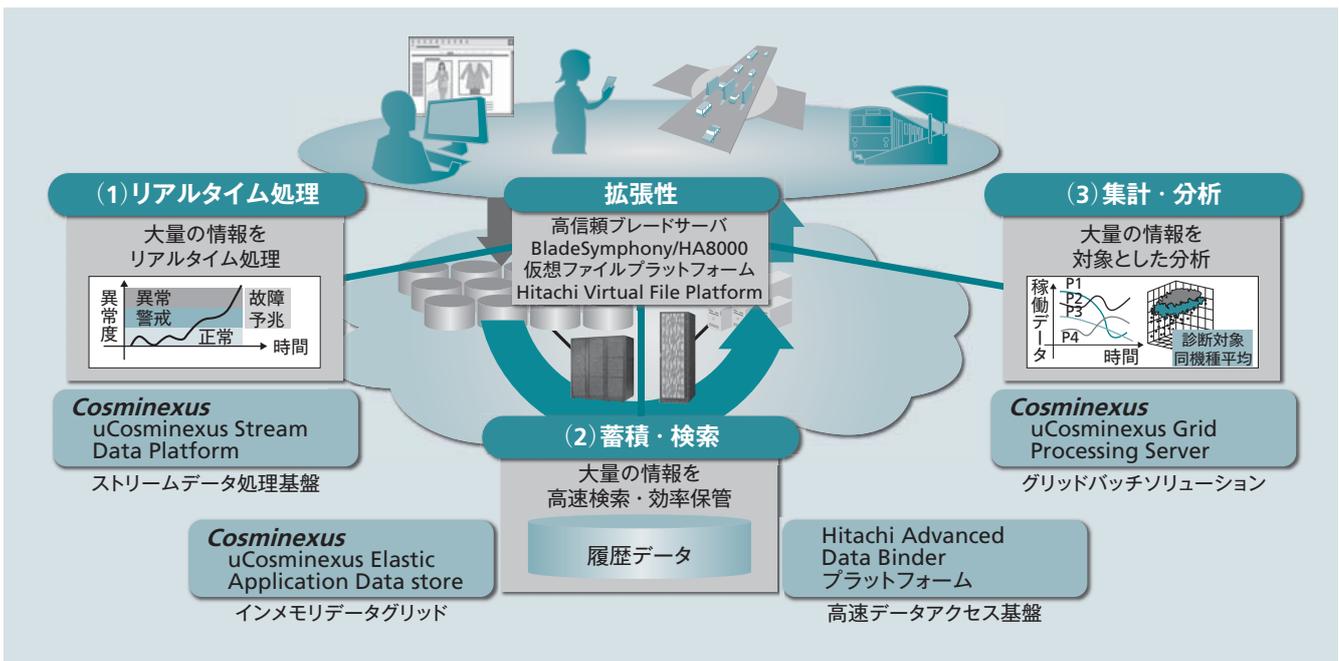


図3 | ビッグデータ利活用を支えるオープンミドルウェア

ビッグデータ活用要件に対応したオープンミドルウェア群を示す。

3. インメモリデータグリッド

3.1 インメモリデータグリッドの特長

ビッグデータ解析では、大量のリクエストを高速に処理することが重要である。それを扱うための基盤として、まず入り口でデータを受け取るためのミドルウェアが uCosminexus Elastic Application Data store (以下、EADSと記す。)である。EADSは、すべてインメモリでデータを処理する仕組みとなっているほか、複数のノード(構成要素)にデータを分散して配置することが可能(データ並列化)である。アプリケーション側はノードの位置を意識せずに KVS(Key-value Store)としてアクセスできる(データ仮想化)。

データの分散は、複数のノードにレコードを分散させる方法として広く使われている Consistent Hashing^{※1)} によって行う。信頼性を高めるためにレプリケーション(複製)の仕組みを備えており、マスタ(親)として書き込まれたノードのほか、別のノードにもスレーブ(子)として同じデータを書き込む。これにより、いずれかのノードに障害が発生しても、別のノードを使って処理を継続することが可能となる。なおEADSはデータの一貫性を確保するため、データの読み込みは必ずマスタを利用する。

分散アーキテクチャでは、いずれかのノードがダウンした際にそれを検知するための仕組みが重要である。ノードダウンの検知はハートビートと呼ばれる応答信号を使って各サーバの生死確認を行う。いずれかのサーバからハートビートがなくなった場合、生存しているサーバが確認のために接続し、それでも応答がなければノードダウンの可能

性があると判断する。多数決によってダウンした可能性が高いと判断すれば、そのノードはクラスタから切り離される。

3.2 インメモリデータグリッドの適用事例

EADSの適用シーンとして、データのバッファリング(一時保存)およびフィルタリングとしての利用が考えられる(図4参照)。書き込み処理をRDBMS(Relational Database Management System)に直接行うのではなく、EADSをバッファとして利用することでRDBMSのパフォーマンス不足によるスループット(単位時間あたりの処理能力)の低下を防ぐ。また全データをCEP(Complex Event Processing)に直接流すのではなく、EADSをフィルタとして利用することで、より効率的な解析処理を実現する。スマートメーターなど各種センサーから送られてくるビッグデータをリアルタイムに蓄積、そして解析を行うといった用途でEADSは使われていくのではないかと考える。

4. 高速データアクセス基盤

東京大学と共同で推進している超高速データベースエンジンの研究開発^{※2)} 成果を製品化し、ビッグデータ利活用におけるデータの検索および分析を高速に行う高速データ

※1) キーのハッシュ値によって配置先ノードを決定し、複数ノードにデータを分散させる方法。

※2) これらの技術には内閣府が創設した最先端研究開発支援プログラムで採択された「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的サービスの実証・評価」(東京大学、日立製作所)で技術開発された成果が反映されている。

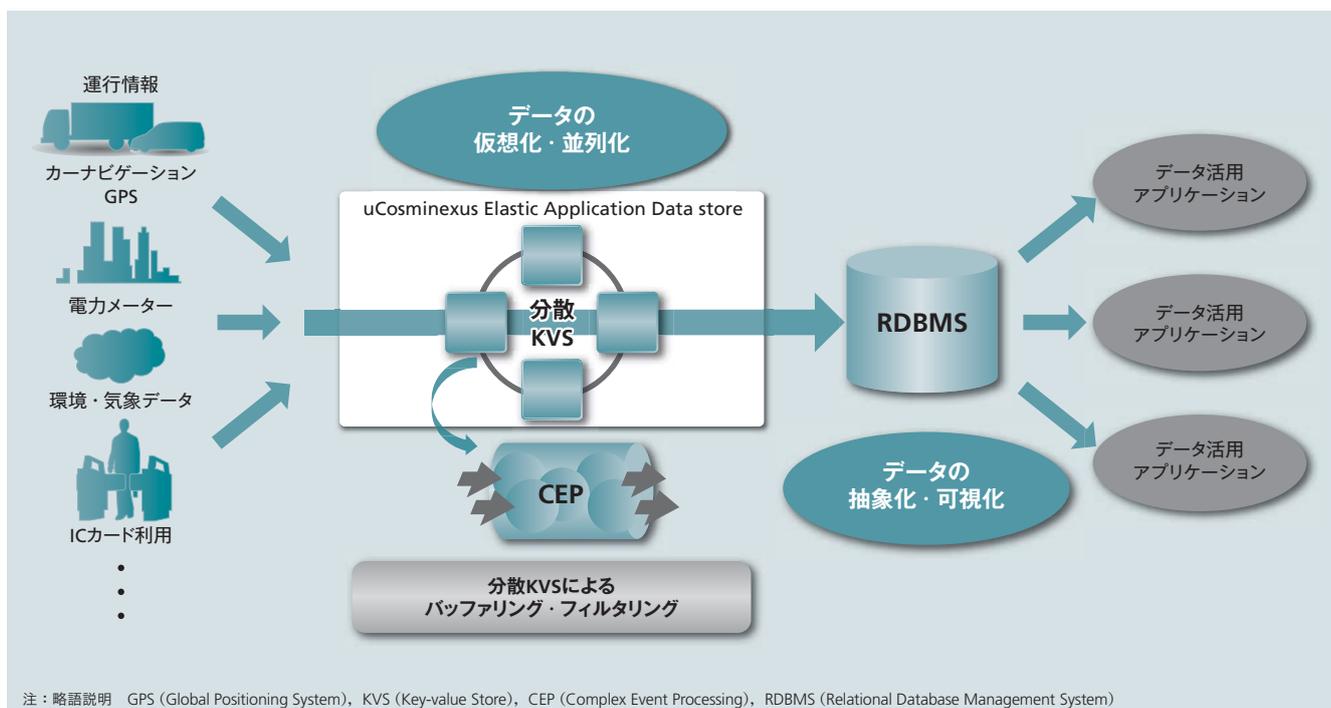


図4 | インメモリデータグリッド uCosminexus Elastic Application Data store の適用シーン

発生するRFID(Radio-frequency Identification)、スマートメーターを蓄積するためのバッファリングあるいはフィルタリングへの適用を想定している。

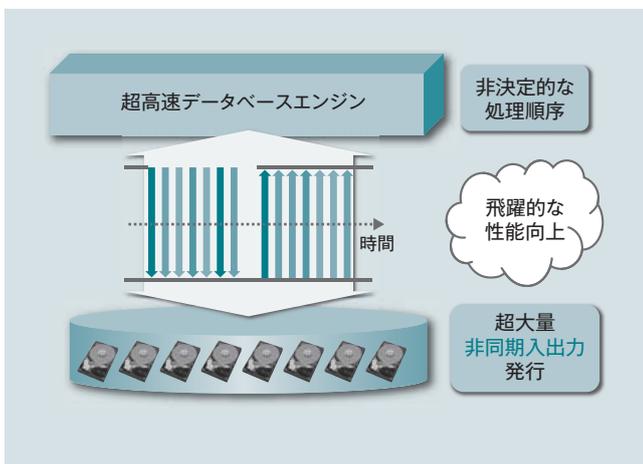


図5 | 超高速データベースエンジンのデータ処理方式
データ入出力要求の発生順序とは無関係な順序で非同期にデータを処理する「非順序型実行原理」を採用している。

アクセス基盤を製品化した。

4.1 超高速データベースエンジンの特長

超高速データベースエンジンは、データ入出力要求の発生順序とは無関係な順序で非同期にデータを処理する「非順序型実行原理」に基づき、データ入出力を行う（図5参照）。

これにより、サーバのマルチコアプロセッサおよびストレージシステムの利用効率を最大限に高めることができ、高速なデータ抽出および検索を可能とする。また、標準SQL (Structured Query Language) に対応することで顧客の既存の分析ノウハウを活用できる。

4.2 高速データアクセス基盤

ユーザーのビッグデータ分析ニーズに対して超高速データベースエンジンの性能を最大限に発揮するため、このデータベースエンジンと、高信頼・高性能な日立のサーバおよびストレージ製品とを組み合わせた事前検証済みのベストプラクティスモデルを高速データアクセス基盤として製品化した。

第一弾として、超高速データベースエンジン、および日立アドバンストサーバHA8000シリーズと、データアクセス性能に優れたSSD (Solid State Drive) を搭載した日立のストレージ製品とを組み合わせた構成のHitachi Advanced Data Binderプラットフォームを2012年6月から

販売開始した。

なお、今後、この製品を基に、顧客およびパートナー企業との共同でのデータ利活用のPoC (Proof of Concept) 実施などを通じてベストプラクティスモデルを拡充するとともに、関連する各種ソリューション、サービスを展開する予定である。

5. おわりに

ここでは、ビッグデータ利活用を支えるオープンミドルウェアについて述べた。

人の暮らしや、モノの動き・変化に伴って、現在、私たちの周囲では、大量の「情報」が発信されている。そうしたビッグデータを活用するためには、「リアルタイム処理」、「蓄積・検索」、「集計・分析」するデータ処理基盤が必要と考えている。日立グループは、これらオープンミドルウェアを継続的に研究開発ならびに機能強化を推進することで、ビッグデータ利活用を支え、さまざまな分野の顧客と、ビジネスやサービスへの応用について協創し、さらに安全・安心で快適なサービスの実現をめざしていく。

参考文献など

- 1) ビッグデータ利活用, <http://www.hitachi.co.jp/bigdata/>
- 2) 超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会的サービスの実証・評価, <http://www.tkl.iis.u-tokyo.ac.jp/FIRST/index.html>

執筆者紹介



中 慎一

1990年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 開発統括本部 ソフトウェア本部 ビッグデータソリューション部 所属
現在、ビッグデータ・スマート情報事業ソリューションの開発に従事



稲場 淳二

1990年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 事業統括本部 PFビジネス本部 ITマネジメントソフトビジネス推進部 ビッグデータビジネス推進室 所属
現在、ビッグデータビジネスの事業企画に従事



梅田 多一

2000年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 開発統括本部 ソフトウェア本部 第2AP基盤ソフト設計部 所属
現在、インメモリデータグリッド製品の開発に従事