

ビッグデータを活用する情報センター向け技術

坂下 善隆

株式会社富士通研究所 フェロー

[アブストラクト]

情報センターにおける技術の観点から、ビッグデータに関する富士通および富士通研究所の取組みをご紹介します。昨今のクラウド技術の発展により、ビッグデータを従来より安価に高速に処理することが可能になった。今まで利用してこなかったデータの活用や、複合するデータを効率的に活用できるようになり、新たなビジネス開拓や知識活用による作業の効率化、高度化が可能になりつつある。この分野の技術に関連する高速・大量処理のためのソフトウェア技術、ストレージアクセス技術、散在するデータを高速に集める通信技術について紹介する。また、いくつかの適用ケースを紹介する。

[キーワード]

ビッグデータ、情報センター、データ活用技術、ストレージアクセス、高速データ転送

1. はじめに

クラウドの進展に伴い、その安価なデータサービスを背景に、1昨年からはビッグデータに対するニーズが拡大している。従来の数種類のデータの大量処理ではなく、多様なデータを組み合わせ、より深い分析が技術的に可能になったことも背景である。ビッグデータの特徴として、「Volume」「Variety」「Velocity」と言う切り口で、その活用を分類でき、いくつかの事例が出てきている。一方先行商談を見ると、従来からの基幹システムのデータを基にプロセス革新を目指すケース、情報系システムの戦略的活用を深化させる目的の適用、新ビジネスへのチャレンジを目指す活用の3つのケースが均等して発生している。この状況において、富士通研究所におけるビッグデータの取組状況を中心に紹介し、一部、富士通の製品化状況を紹介する。

2. 分析・活用技術

①分析シナリオを自動推薦

分析の専門家が創出した「分析シナリオ(データの組み合わせ、分析プロセス)を蓄積し再利用を行う。蓄積された分析シナリオからも目的にあったシナリオを自動推薦する。

②ビッグデータ統合開発環境

ビッグデータではバッチ処理やリアルタイム処理など異なるデータ処理タイプを使いこなす必要がある。このような開発実行環境を統合する機能を開発し、開発工数を従来の8週間から1.5週間に短縮させた。

3. データ処理技術

①高性能・高信頼 Hadoop 統合商品

ここでは、富士通が2012年4月に発表した「Hadoop ソフトウェアサポートサービス」を支える製品を紹介する。大量バッチ処理ではオープンソースの「Hadoop」が広く利用されており、それを前提とした、使い勝手や性能向上を図る機能を各ベンダーが提供するようになっている。富士通も、並列分散処理による高速化や独自ファイルシステムを提供することにより、既存システム連携やデータバックアップの効率性を高めている。

②複合イベント処理(CEP)

センサーやスマホなど複数の情報発生源から絶え間なく送られてくる時系列データを、リアルタイムに処理をする。リアルに役立つ情報を、設定された「ルール」に基づき抽出し、役に立つ情報を適切に知らせる。この処理の高速化技術が進められており、今までできなかった応用面の創出が期待されている。

③インクリメンタル方式

バッチ処理方式とリアルタイム処理の中間の容量を対象に、準リアルタイム処理を狙った技術である。メモリを活用しつつ、大容量ディスクのアクセス回数を減らして高速処理を実現するものであり、時間単位で処理をしていたものを分単位で処理できるようにすることで、大量データ処理の応用分野を拡大させる。

④自律負荷分散ストレージ

センセーショナルな事件が起きた時、Webサーバなどでは特定データに対する負荷集中による性能低下がしばしば発生する。負荷集中が起きているデータのレプリカ数を動的に増加させ、アクセスを分散させる技術である。アクセス集中度分析機構を使いアクセス頻度を検知しその頻度に応じて増加レプリカ数を決定させる。

⑤高速ファイル転送

設計図や観測データなどをの大量データを遠隔地に送りたいという要望に応える技術と製品である。TCPファイル転送やUDPファイル転送の課題を解決する技術を製品化し、本年8月より順次、各OS版を提供開始している。

[参考文献]

- (1) データの急激な増加 出典:IDC「IDC Predictions 2012: Competing for 2020」、(#231720、2011/12)
- (2) ビッグデータ市場規模の拡大 出典:IDC「Worldwide Big Data Technology and Services 2012-2015 Forecast」(#233485、2012/3)