

Executive Summary

データ分析の最新技術 – InterSystems DeepSee™

– リアルタイム多次元分析でビジネスインテリジェンスは新時代を迎える –

インターシステムズジャパン株式会社
代表取締役社長/日本統括責任者 植松裕史

従来型ビジネスインテリジェンス（BI）の仕組みでは、その構造上、リアルタイム分析を実現することはきわめて難しい。いくつか決定的な理由があるが、ひとつには従来の関係データベース（RDB）におけるデータの取り扱い方式があげられる。

RDBはその歴史上「固定長」データを主体にアーキテクチャが構築されている。実際にビジネスで発生するデータは必ずしも固定長ではなく、むしろ可変長データが圧倒的に多い。したがってRDBでは多くの場合可変長データを固定長データに変換するためのデータの正規化が要求される。データを多数の固定長データテーブルに作り直す膨大な作業だ。さらにRDBには「データベース管理システム（RDBMS）」がつきまとう。RDBMSとの連携には難解なSQL文によるアクセスが必要だ。処理効率を追求すればSQL構文が複雑化することは必定。複雑化したSQL文はさらに処理効率を落とす。そしてデータ分析には、SQL文（テーブルの行を読み取るためのSELECT文）により「分析キューブ（蓄積されたデータを多次元・多階層に編成したもの）」を作るために複数テーブルの結合処理（JOIN）を何度も繰り返すことになる。

「データクレンジング」の問題もある。RDBの構造では、元のトランザクションデータは後々分析目的に利用されるということを前提には処理されていない。元データの更新（フィールドの追加、削除、新しいテーブル追加など）や更新のタイミングに付随するデータ抽出のタイミングなどが、後の分析データ作成工程に必ず影響する。データごとに更新のタイミングも異なれば、アプリケーションの仕様変更などにも影響を受ける。端的に言えば、「後々分析する段階になってデータ仕様を変更するための膨大な作業が発生する」ということだ。

この複雑なデータベース処理に加え、膨大なI/O処理を経て「データウェアハウス」を構築し、同じく膨大なI/O処理を経てBIソフトウェアへ情報を渡す。結果、得られるデータは「周回遅れ」。ユーザーのデータ分析は後手に回り続ける。リアルタイム分析は不可能だ。

ビジネスプロセスは継続的な「改善」を以ってはじめてその効果をもたらす。「ビジネスプロセスリエンジニアリング（BPR）」「ビジネスアクティビティモニタリング（BAM）」なるものが常に叫ばれるのはそのためだ。

ビジネスプロセスとは「水の流れ」に似ている。どこかに「流れを妨げるもの」があれば、それを解決するには「妨げている様子」を「妨げが発生しているその時の水の流れ」を見なければならぬ。川下に行って、「どこで流れを妨げているのだろう？」と悩んでも何も見えない。ビジネスプロセスも「各プロセスの動いているその時」という時間軸で見なければ改善は図れない。「リアルタイム分析必要性」の意味はそこにある。

刻々と変化するビジネス。ビジネスプロセスを柔軟に組み替えて市場や競合に迅速に対応することが必要だ。製品別、市場別、担当者別、機能別、あるいは時間軸など、多次元で分析することも重要となる。ビジネスプロセスのどこの段階においても、その時点でのデータをリアルタイムにとらえて分析することで、はじめて生きたビジネスインテリジェンスとなる。

インターシステムズのもつ「トランザクショナルビットマップインデックス」技術（詳しくは「Cachéテクノロジーガイド」を参照）は、従来の「データウェアハウス」の基本データ構造である「スタースキーマ」（中央に分析対象、周辺に分析軸を置くデータ構造。線で結んだ形が星形に見えるスキーマであることからそう呼ばれる）をも超高速に分析できる高度な技術だ。InterSystems DeepSee™ は、InterSystems Caché®がもつ「トランザクショナルビットマップインデックス」技術により、先述の「分析キューブ」を動的に生成する速度を劇的に速め、高度圧縮技術によってトランザクション処理負荷を最小限におさえることができる。ゆえに、アプリケーションに組込む形で超高速多次元リアルタイム分析が実現できるのである。

DeepSee はCachéによるデータ管理が前提だが、RDBユーザーであってもその既存環境を崩すことなく、DeepSeeのコネクタ機能によりRDBとスムーズに連携しながら、Cachéの優位性である多次元データモデルを構築できる。「ぜい肉」の無い、「俊敏で筋肉質」な多次元データ分析が可能となる。既存環境を生かしながら、DeepSeeで高度ビジネスインテリジェンスを実現する。