

さまざまな課題を克服してビッグデータを有効活用

ビッグデータの活用には、「容量」「種類」「スピード」といったさまざまな課題があります。本稿では、それぞれの課題に着目し、海外でのお客様の取り組み事例をご紹介します。

1つ目は、「容量」と「種類」に対応した風力タービン設置計画のためのシステムを採用されたことにより、データの解析時間を大幅短縮されたお客様の事例です。次は「スピード」に着目し、リアルタイムな判断・予測が要求される新生児ICU（集中治療室）での取り組み内容を、最後は、「容量」に対応した大量データへのIOパフォーマンスに着目したSAP用ストレージにより、お客様へのサービス向上を図る事例となっています。ぜひ実例からその活用方法やテクノロジーの使い分けを感じ取ってください。

<掲載したお客様>

・Vestas Wind Systems A/S社

膨大なデータを解析し、風力発電所設置を最適化

参考) <http://www.ibm.com/jp/press/2011/10/2502.html>

・オンタリオ工科大学

新生児の測定データをリアルタイムに分析・活用

参考) http://www.ibm.com/innovation/jp/technologies/bigdata2/medical_front.shtml

・Gerber Scientific社

新しいストレージにより、処理時間を大幅に短縮

参考) <http://www.ibm.com/systems/jp/storage/case/xiv/gerberscientific.pdf>

<記事作成協力>

日本アイ・ビー・エム株式会社
ソフトウェア事業 インフォメーション・マネージメント事業部
ビッグデータタイガー・テクニカルリード
土屋 敦

日本アイ・ビー・エム株式会社
システム製品事業 マーケティング&ソリューション
マーケット・マネジメント
風口 悦子

膨大なデータを解析し、風力発電所設置を最適化

デンマークを本拠地とする Vestas Wind System A/S (以下、Vestas 社) は、気象情報、地理空間データ、衛星写真、森林伐採マップなどのモデルを利用して、膨大な構造化、非構造化データを解析。風力発電タービンの設置場所の最適化を図りました。このビッグデータを処理するエンジンには IBM InfoSphere BigInsights (以下、InfoSphere BigInsights) と高性能な IBM システムを導入。今まで数週間を要していた解析時間の大幅短縮に成功しました。

設置場所の分析により風力発電の効率を最大化

Vestas 社は、デンマーク、ランダースに本拠地を置く、風力発電システムのグローバル・リーダーで、これまでに 67 カ国で 44,500 機を超える風力発電タービンを供給してきました。欧州連合では 2020 年までにすべてのエネルギーの 20% を再生エネルギーで賄うことが決定されており、風力発電がエネルギー源の中心となっています。風力発電の成功事例の増加に伴い、風力発電プラントを短期間で設置できるテクノロジーの需要が増加する中、Vestas 社が再生エネルギー・ビジネスでナンバーワンとなるには、最新テクノロジーを活用する必要があります。しかし単に強い風が吹く場所に風力発電タービンを設置するだけでは、効率的に発電を行い、ビジネスを最大化できるとは限りません。従って、タービンの設置場所を分析することが非常に重要であり、難しい課題でした。風は変動が大きいエネルギー源の代表であるといえます。そうした問題を踏まえつつお客様に対して、風力こそが最も信頼性が高いエネルギー源であることを訴求しなければなりません。この高い信頼性を証明するためには、膨大なデータの分析が必須であり、よりきめ細やかな情報が必要となっていました。

3週間程度要していた解析時間を約15分に短縮

このような再生エネルギーの代表である風力発電を実現するための風力発電タービンの効率的で迅速な提供を実現するためのシステムとして、Vestas 社は、InfoSphere BigInsights、IBM System x (以下、System x)、IBM System x iDataPlex (以下、iDataPlex) から構成されるスーパーコンピューター「Firestorm」を活用しました。現在、Vestas 社が所有する雨、風をはじめとする気象データ、衛星写真、森林伐採マップなどのデータ・ライブラリーは、およそ 18PB (ペタバイト) から 24PB になります。これ



は HD 高画質のビデオに換算すると約 70 年間分ものデータに相当します。このような大量なデータも、効率よく分析ができなければまったく意味がありません。そのため、Vestas 社では、構造化データ、非構造化データを問わずに格納、分析することができる InfoSphere BigInsights ソリューションを導入しお客様に分析結果を提供することで、お客様にとってのビジネス意思決定のスピードを迅速化することに成功しました。その結果、これまでに風力発電タービンの最も適した設置場所を特定するための解析にかかる所要時間は、約 3 週間から 15 分ほどに大きく短縮されました。この解析では、地球表面を 1 キロメートル四方のグリッドに分割し、各グリッドに 150 以上のパラメーターを設定するという膨大な環境情報を取り扱います。このシステムは、15,000 本のケーブルで接続された System x と iDataPlex サーバーで構成されており、InfoSphere BigInsights のストレージ・サイズは、現状 2.8PB になります。このサイズは将来的には 20PB になると計画されています。

風力発電の市場開拓を促進

Vestas Technology R&D のプラント・サイティング・アンド・フォーキャスティング担当バイス・プレジデントであるラース・クリスチャン・クリステンセン氏は、今後のソリューションへの期待について次のように述べています。

「Vestas 社は風力タービンの運用において 10 年の実績を持っており、わたしたちのお客様はタービンを設置する前に、期待できる発電量や ROI (Return On Investment) に関する情報を求めています。IBM のソフトウェアとシステムを導入することでこれらの情報を速やかにお客様に提示し、新規の風力発電市場を開拓するとともに、お客様の再生可能エネルギー関連の目標達成に寄与したいと考えています」

オンタリオ工科大学 (University of Ontario Institute of Technology)

新生児の測定データをリアルタイムに分析・活用

カナダを本拠地とするオンタリオ工科大学 (University of Ontario Institute of Technology : 以下、UOIT) は、IBM InfoSphere Streams (以下、InfoSphere Streams) を活用して、新生児の状態急変をリアルタイムに把握・予測するソリューションを構築しました。新生児のモニタリング機器から生成される、毎秒1,006に及ぶすべてレート異なる継続的なデータをリアルタイムに記録し、データ・マイニングのために蓄積。新生児医療における先進的なリアルタイム・データ分析を実現しました。

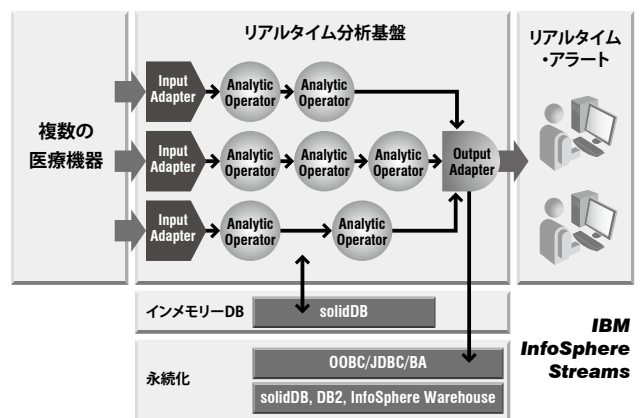
新生児医学の診療をテクノロジー面でサポート

UOIT では、Hospital for Sick Kids プログラムにおいて、新生児医療における先進的な医療改革を実践しました。出産予定日より最大 17 週早く生まれてきた新生児の微妙な状態の変化をリアルタイムに検知できるよう、心拍数や呼吸などの医学的データ・ストリームを絶え間なく取り込み監視しています。

新生児医療において広く使われている分析ツールは、データベース内のデータ使用だけでなく、幾多の数値、英数字、音声、ビデオ、心電図などの構造化データや非構造化データに対して強力な分析を実行したり、複雑な相関を見つけることに制約がありました。そのため UOIT は、新生児医学の臨床医をテクノロジー面でサポートすること、つまり、現状予測の実現可能性を秘めたリアルタイム・ソリューションの構築を研究活動としていました。看護スタッフや、新生児生理学者によって監視されるデータは、16 ストリームに及ぶことがあります。これらは、新生児のモニタリング機器からすべて異なるレートで生成されます。これらのデータをリアルタイムに記録し、データ・マイニングのために蓄積し、新生児医療の質の向上に利用することは先進的な医療改革でした。

危険な兆候を最大24時間前に検知可能に

UOIT が新生児学の領域において試みた方法は、InfoSphere Streams、IBM DB2 (以下、DB2) そして IBM solidDB (以下、solidDB) を統合させた環境を構築することでした。この環境によって、集中新生児ケアユニットの医療機器からのデータをリアルタイムに記録することが可能になっただけでなく、医療情報データを DB2 に永続化することも可能になりました。後には個別患者のためだけでなく、臨床研究にも利用することが可能となりました。最大 16 の異なる複数の身体機能を観察することによって、問題が発生していることを医師に警告することができます。これにより、生命にかかわる感染症などを最大 24 時間前に検知できる可能性を実現しました。



さらに、毎秒 1,006 に及ぶ測定値が DB2 に蓄積され、新生児医学の臨床医はこれを利用して分析を行うことができます。この値は、一人の患者当たり 1日 9,000 万の数値に換算されます。この膨大な量のデータを相関させることが可能な先進的な分析パラダイムの適応を InfoSphere Streams が実現したのです。ここで重要なのはリアルタイムの観察・分析だけでなく、今まで分からなかった相関関係などのデータを蓄積・分析することで新たな予測・発見が可能になるということです。これによって、新たに判明したことをリアルタイム観察にフィードバックすることが可能になります。

新生児医療研究として継続的に運用

臨床実験の最大の目的は集中治療室 (以下、ICU) における情報利用の在り方を模索することにあります。使用している医療機器は非常に高度化されたものですが、医療の意思決定という観点では従来通りのリアクティブな検診だといえます。看護スタッフが、あらゆる医療モニターからの数値を観察・記録したデータを医師が診ることで、何か異常がないかを確認しています。そこで、ビジネス・インテリジェンスやデータ・ウェアハウス分析の知識をリアルタイム分析と合わせて、新生児医療の世界に適応することを目指して今回の臨床実験が実施されました。実際に、2009 年に Hospital for Sick Kids プログラムで実装されてから、新生児医療研究として継続運用がなされています。このように、医療はエビデンス・ベースであることが求められるため、その実証を行うためのデータ臨床実験が継続的に行われています。

最近の例としては、米国ロードアイランド州の病院と提携し、データを送信し監視するクラウド版システムの取り組みが挙げられます。遠隔医療の実現に発展させる可能性もあり、UOIT では情報マネジメントを有効活用した先進的な臨床実験を続けています。

新しいストレージにより、処理時間を大幅に短縮

Gerber Scientific 社（以下、Gerber 社）は、急速に増大し続けるデータ、トランザクションに対応し、既存の SAP 環境で十分なパフォーマンスを得るために IT 基盤を刷新。その条件としては、最低でも 1 万 IOPS（1 秒当たりの入出力数）と 1.6TB（テラバイト）を超える Oracle データベースの複数のインスタンスをサポートする能力を備えたものでなければなりません。これらの目標を達成するためのストレージ基盤として、IBM XIV Storage System（以下、XIV）を採用しパフォーマンスと容量不足の問題を解決すると同時に、ビッグデータ時代の将来の拡張にも対応することでサービス・レベルを向上し、企業競争力を高めています。

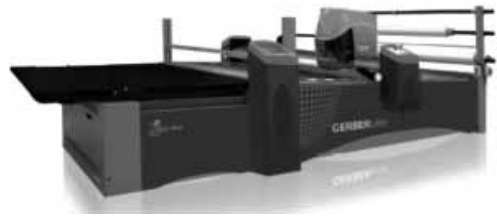
ビッグデータ時代、SAP本番環境が抱える課題

米国コネチカット州サウスウインザーを本拠地とする Gerber 社は、看板製作、特殊グラフィックス、眼鏡レンズ加工、アパレル、フレキシブル材料、印刷、包装をはじめとする幅広い業界向けの革新的なエンド・ツー・エンド顧客ソリューションにおける世界的リーダーです。同社のソフトウェア、コンピューター化された製造システム、消耗品、およびサービスは、顧客サポートと柔軟性を最大限に高めるために完全に統合化されています。また、同社のグローバルな組織は、財務会計・管理および販売・流通プロセスを担う SAP ERP 5.0 を中心に構築された IT インフラストラクチャーによって支えられています。Gerber 社は、IBM Power アーキテクチャーと、他ベンダー製の独自技術のストレージ・デバイスをベースとするストレージ・エリア・ネットワーク（SAN）を組み合わせて SAP を運用していました。しかし、SAN システムとストレージ機器の老朽化が進む一方、これまでにない規模で急速に増加するデータに対応するために、パフォーマンスのボトルネックや容量不足が目立ち始めていたことから IT 基盤の更新の必要性が生じていました。

全世界のユーザーに対するサービスの劇的な向上

既存の SAP 本番環境で十分なパフォーマンスを得るためには、新しいインフラは最低でも 1 万 IOPS と 1.6TB を超える Oracle データベースの複数インスタンスをサポートする能力を備えたものでなければなりません。Gerber 社は、パフォーマンスの向上に加え、信頼性やスケラビリティを低下させることなく、物理的な設置面積が小さく、しかも消費電力と冷却要件が低いストレージ・システムに統合することで運用コストの削減を図ることも目指していました。

さまざまなベンダー製品を評価した結果、XIV のパフォーマンスが、現在の SAP ERP ワークロードを効率よく処理するために必要な速度の 550% に相当し、既存ストレージ・アレイよ



り 1,200% も高速であり、同社の一般的なビジネス・プロセスの多くについて、パフォーマンスと世界中のエンド・ユーザーに対するサービスの劇的な向上につながると判断。採用を決定しました。

例えば、ある SAP バッチ・プロセスの処理時間は、従来の 8 時間からわずか 15 分に短縮されました。また、この IOPS 値は、エンド・ユーザーに対する応答時間が、大多数の SAP トランザクションにおいて 1 秒以下に短縮することも意味しています。PoC（Proof of Concept）の一環として、Gerber 社はさまざまなフェイルオーバー・シナリオのテストも行いましたが、いずれも障害やパフォーマンス低下は一切発生せず、満足できる結果が得られました。しかもプロジェクト開始から本番稼働までのプロセス全体の所要期間は、60 日未満でした。こうして、ダウンタイムやユーザーに対するサービスの中断を最小限にとどめながら、スムーズにビッグデータを処理する環境を手に入れました。

TCOの削減と将来の企業成長への対応の両立

XIV はチューニングレスで高いアクセス性能を維持し、性能を劣化させない上、高い可用性でアプリケーション稼働時間の最大化を支援するのに加え、モジュールの追加だけで必要なときに容量と性能をリニアに拡張することができます。さらに既存ストレージからのデータ移行作業も含む管理作業を大幅に簡素化することができるなど、TCO 削減にも大きく貢献します。

Gerber 社でも、すでにさまざまな効果が得られています。超並列型の XIV アーキテクチャーにより、SAP 環境のバックアップ所要時間が半分以下に短縮されました。その結果、複雑なタスク・スケジューリングの必要性が少なくなり、IT スタッフの負担軽減につながったほか、全体的なシステム・パフォーマンスを 1 日 24 時間週 7 日という高いレベルに維持することが可能になりました。従来のストレージ・システムの 60% 程度のコストで 2 倍のストレージ容量を確保し、設備投資を抑制しながら、現在の SAP ERP 本番環境の要件が 10,000 IOPS 程度であることを踏まえ、将来の成長に十分に対応できる余裕を確保し、サービス・レベルを劇的に向上しています。