

新たにビジネスアプローチを生む Hadoop基盤構築を通して、



大規模分散処理技術を実装するHadoopを使ったビッグデータ解析基盤の構築に取り組む、三菱UFJインフォメーションテクノロジー。金融取引に関するさまざまなデータを用い、新しい角度からの顧客分析に役立てるなど、新たなエンタープライズシステムを生み出し始めている。

世に先駆けて、2009年から

Hadoop関連技術に取り組む

三菱UFJインフォメーションテクノロジーは、三菱UFJフィナンシャル・グループをIT面から支えるプロフェッショナル集団。セキュリティ基準を満たすメールの共同利用やインターネット接続サービスなど、グループ企業の円滑なビジネス推進をインフラ側から支えている。

さらに、金融×ITで培ったモノづくりのスピリットを生かし、さまざまなR&Dにも挑む。その一つが、ビッグデータ分野だ。

三菱UFJインフォメーションテクノロジーには、最新技術について業務適用の可能性を探る「実験室=ラボ」的な役割を持つ部署がある。それがITプロデュース部だ。SOA(サービス指向アーキテクチャ)の活用や、開発言語にRubyを導入するなど、さまざまなオープンソース・ソフトウェアの活用にも意欲的だ。大規模分散処理技術を実装するHadoopを使ったビッグデータ解析基盤の構築に早くから取り組み、グループ企業向けに提案・導入を進める先導的な役割を担う。

Hadoop技術検証が始まったのは、2009年から。まだビッグデータがバズ

ワードとして広く知られる以前だ。

最初にテーマにしたのは、バッチ処理業務の高速化。「夜間のバッチ処理が朝までに終了しない」という状況があり、Hadoopを使って高速化できれば面白いのではないか、という動機からでした。Hadoopを用いて業務を対象としたHadoop環境は、スタート時点での規模は8node^{*1}、格納可能なデータ量は2TB程度であり、それほどビッグなデータとは言えません。

「たしかにバッチ処理そのものは高速化されたのですが、業務システムからHadoopにデータをもってくるところに時間が要し、トータルでいうとトントンくらいの効果しか出ませんでした。Hadoop上で業務システムを動かすわけではないので、その改善のためにには、業務データそのものの作り直しが必要ですが、はたしてそこまでやる必要性があるだろうかと。高速処理だけではなく、運用時間やコストを含めたトータルで効果を生み出さなければ、新たに取り組む意味がありません」。Hadoopのバッチ処理業務への適用は、いったん保留という形になった。

新技術の有効性を検証するのも、この部署の仕事の一つであり、それで一つの役目は終わったようにみえた。ただ、土佐氏はそこで諦めなかつた。

「そのうち、世の中でマーケティングの分野からビッグデータの先進事例が出てくるようになりました。しかし、製造業や

は、手元のログデータなどを用いて、テスト的解析を行うようになり、2011年に入ると、業務バッチを実際にHadoop環境で処理するという想定でのトライアルが始まった。

流通業ではなく、金融分野においてHadoopが生かされる分野は何だろうと模索しました。

発想の転換があったきっかけは、顧客へのヒアリングだった。

「同じバッチ処理といつても、日々発生する事務処理のバッチではなく、アクセスログなどの非構造データを構造化するための処理もあるはず。これまで処理が面倒であり顧みられなかつた非構造データの処理が簡単にできるのであれば、マーケティングやリスク評価といった業務に生かせるのではないか。ヒアリングに行けば行くほどマーケティングなどの部門で需要があることがわかりました」。

現在は、顧客が今まで使ってこなかつた非構造データを分析して提供サービスを向上させる取り組みが始まっている。まだトライアルプロジェクトの段階ながら、「ユーザー自身が驚くほどの効果が上がっている」という。

トライアルを通して、単なるWEBフレームワークのログからマーケティングとして必要な導線情報を抽出して、分析しやすい形に整理するノウハウも培われてきた。「ログから情報を抽出して、導線の形に変換するプログラムなども開発しています。この導線を抽出するところでHadoopのパワーがすぐ生きてくると実感しています」とのことだ。

R & D部門だからこそ、 新技術を共に楽しむ関係が必要

Hadoop環境構築に関して、同社と日立ソリューションズとの関わりは2009年から始まった。情報が少ない中を共に悪戦苦闘してきた間柄だ。2011年後半からは、Hadoopの商用ディストリビューションであるCloudera Enterpriseの導入を進めており、この環境の上でトライアルプロジェクトが動いている。

また、土佐氏が企画、主催するAsakusa Framework勉強会では、日立ソリューションズが初回の開催会場を提供したり、発表者として参加するなど、幅広い連携が生まれている。

「基盤技術について高いスキルと経験を持っており、新技術の評価にも積極的。私の方からの“無茶ぶり”にもすぐに反応していただけて、感謝しています。私たちが金融系のIT企業ということで、どうしても検証しつくされた技術を手堅く提供するというスタンスのSlerが多いのですが、ITプロデュース部はR&D部門。私自身が新しい技術導入に意欲的なので、お互い楽しんでチャレンジしてくれるようなSlerのほうがよい。日立ソリューションズは、パートナーとしては最適だと考えています」と評価する。

開発者の関心を よりビジネス領域に広げる

Hadoopを使ったビッグデータ解析に関する、共に苦闘しながらビジネスへの導入を進めてきた5年間。そこで得た知見をベースに、ビッグデータの将来について、土佐氏はこう言う。

「ビッグデータ解析は、非常に基盤的な技術なのに、ビジネスそのものを変えてしまうほどの影響力があります。それだけの可能性を秘めているものの、基盤技術とビジネスを繋げるのはそう簡単ではない。システム側のエンジニアだけで発想すると、単なる技術開発で終わってしまう。やはり、ビジネスの“現場感覚”を持つている人のアイデアを広く集めないと、ビッグデータは宝の持ち腐れになってしまいます」。

言いかえれば、エンジニアに求められるスキルをより広く、高くということだろう。「基盤技術だけと付き合って終わりではなく、さらに自分の領域を広げて、データサイエンティストをめざしていくようなアプローチの必要性」を、ビッグデータは求めているのだ。

そうしたスキルを磨くため、ITプロデュース部のメンバーは、最近、世界中の統計家やデータ分析家がその最適モデルを競い合うサイト“Kaggle”への関心を強めている。ある課題では世界中の投稿者のなかでトップ・クォーター(上位25%)にランクインすることもあった。

「私たち自身の関心やスキルのレイ

Teppi Tosa



三菱UFJインフォメーションテクノロジー株式会社
ITプロデュース部
マネージャー

土佐 鉄平 氏

ヤーを縦に広げていくような取り組みは、これからますます大切になると思います」。

また、最新技術に意欲的な土佐氏が注目しているのが“Spark”だ。繰り返し計算処理の技術であり、計算処理の非常に多い金融系の分析に効果を発揮すると見ている。現在、金融業界ならではのビッグデータへのアプローチ確立をめざして、同技術について日立ソリューションズと検証を始めようとしている。

【Hadoop】

大容量のデータを多数のサーバーに分散し、並列処理するためのオープンソースのプラットフォーム。Googleが検索エンジン用に開発したバッチ処理システムを基に開発された。膨大な量のデータを短時間で処理することが可能なため、ビッグデータ分析の仕組みを構築する上で注目されている。

【Asakusa Framework】

Hadoop上で大規模な基幹バッチ処理を行うためのフレームワーク。基幹バッチシステムに必要な開発環境・実行環境・運用環境を実装しているため、Hadoopを意識せずに複雑な業務処理の開発が可能になる。

【Spark】

Hadoopと同じく、オープンソースの分散データ処理プラットフォーム。インメモリー処理を主体とするSparkは、機械学習やグラフ計算のように繰り返しの計算が多い処理をHadoopよりも高速で実行できる。

導入ソリューション概要

【ジャンル】	【品名】
Hadoop	Cloudera Enterprise
Asakusa Framework	金融