

●伊藤忠テクノソリューションズ

人材を結集して専門部隊を設立 顧客のビッグデータ活用をトータル支援

2011年12月には全社横断組織の「ビッグデータ・ビジネス・タスクフォース」を設置するなど、早期からビッグデータ・ビジネスを推進している伊藤忠テクノソリューションズ(以下、CTC)。同社は、長年のBIソリューション提供やインフラ運用業務で培った知見や技術力、豊富な製品ラインナップなどを背景に、ビッグデータ・アナリティクスソリューション分野においても大きな存在感を示している。2014年4月には専門の「ビッグデータ・アナリティクス部」を設立。顧客のビッグデータ活用をトータルに支援する体制を強化した。本記事では、CTCのビッグデータ・アナリティクスソリューションの取り組みについて紹介する。

コンサルから製品、開発まで 一気通貫のソリューションを提供

伊藤忠テクノソリューションズ(以下、CTC)は2014年4月、同社のITサービス事業グループソリューション事業推進本部配下に、ビッグデータ・アナリティクスソリューションを提供・支援する専門部署「ビッグデータ・アナリティクス部」(図1)を新設した。これまで社内さまざまな部署に分散していたビッグデータ・アナリティクスやそれに関連する製品の専門家を結集し、顧客のビッグデータ活用をトータルに支援する体制を作り上げることを

狙ったものである。

ビッグデータ・アナリティクス部は、大規模データ分散処理基盤の「Apache Hadoop」(以下、Hadoop)関連のサービスや製品

を主管する「ビッグデータプラットフォーム課」と、BIツールやデータウェアハウス(DWH)などのデータ分析用製品を主管する「ビジネスアナリティクス課」、情報活用やBI活用についてのコンサルティングを提供する「アナリティクスコンサルティング課」の3課で構成され



伊藤忠テクノソリューションズ(株) IT サービス事業グループ
ソリューション事業推進本部 ビッグデータ・アナリティクス部
ビッグデータプラットフォーム課

[左から]梅川 真人氏、課長 小林 範昭氏

る。「ビッグデータ・アナリティクスに関連するインフラ、分析ツール、コンサルティングを同じ部署で取り扱うようになったことで、上流から下流までのトータルソリューションを提供しやすくなりました。当社内の業務がスムーズになったというだけでなく、さまざまなご要望に迅速かつ柔軟に対応できる体制になったことで、お客様のビッグデータ活用をこれまでより効率的に支援できると考えています。」(CTC ビッグデータ・アナリティクス部 課長 小林 範昭氏)

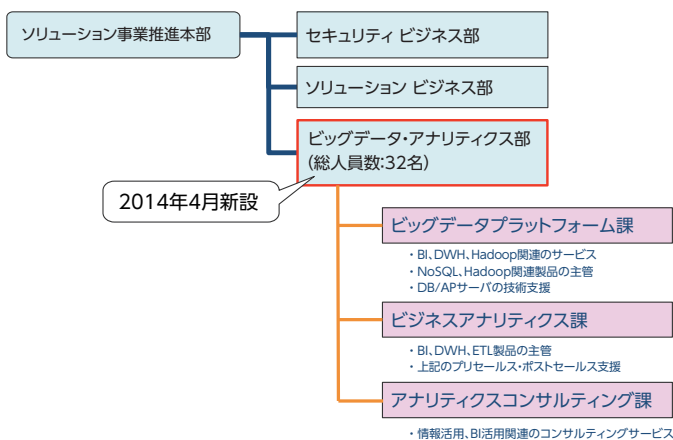


図1 2014年4月にビッグデータ専門の部署を新設

長年にわたるBI/DWHの経験と豊富な製品ラインナップが武器

CTCは、BIがブームとなった1994年頃からデータ・アナリティクス/ビジネス・アナリティクス領域に注力してきた。長年にわたって培った同社の経験は、ビッグデータ・アナリティクスにおいても大きな武器である。

また、豊富な製品ラインナップも同社の強みである。取扱製品は年々拡充している(図2)。例えば、2007年には米 Netezza 社(2010年に米 IBM 社に買収)のDWH製品「NETEZZA」の販売代理店となったほか、2010年3月には日本ネティーズとSAPジャパンと共同で開発したBIソリューション(BIアプライアンス)「C-BIA」の販売を開始した。

2014年10月時点のビッグデータ・アナリティクス関連のCTC製品ラインナップは図3に挙げた通りである。保守まで一元的に提供している製品も多い。

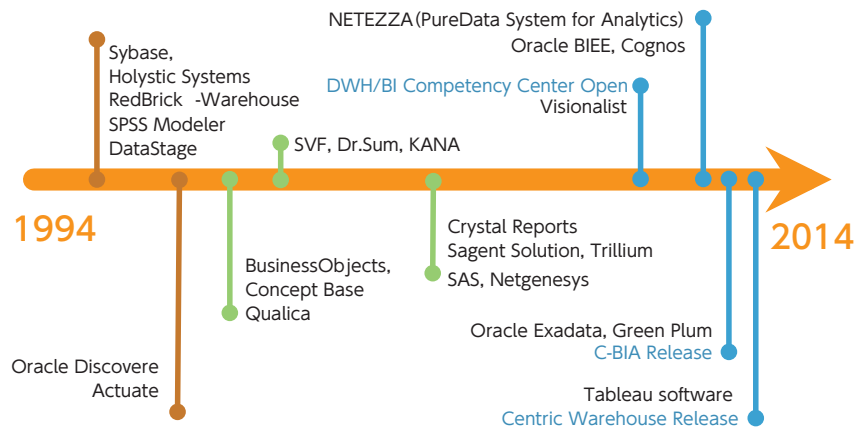


図2 CTCのアナリティクス領域での取扱製品タイムライン

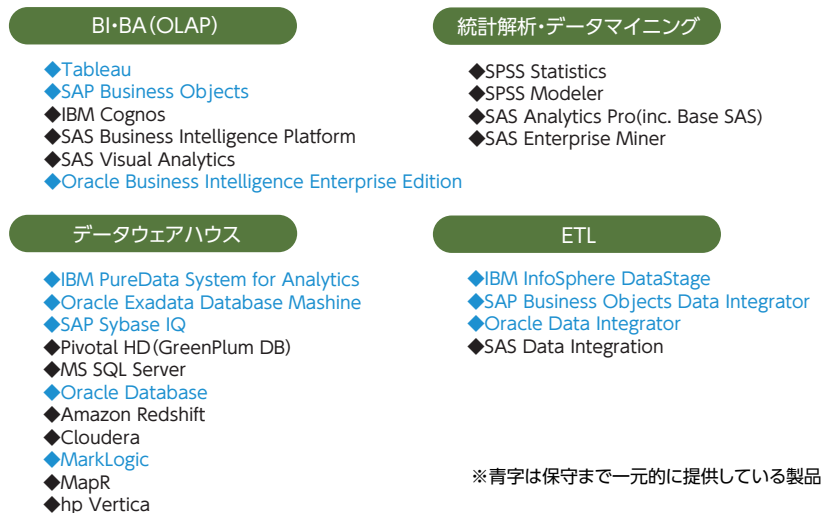


図3 幅広い製品ラインナップと充実した保守体制

日本最大級の検証環境「Big Data Processing Lab」

CTCは2005年より、マルチベンダー環境でさまざまな検証を行える総合検証施設「Technical Solution Center (以下、TSC)」を運営している。TSCのサーバエリアは約500台のサーバで構成され、ストレージエリアの容量は合計約1PB(1000TB)と日本最大級の規模を誇る(数値はいずれも2014年4月時点のもの)。2012年2月には、

3つのポイント

- ①ベンダーを限定せずソフト・ハードを選定
- ②BAの枠を超えた検証環境の利用
- ⑤専任エンジニアのノウハウ、スキルを活用

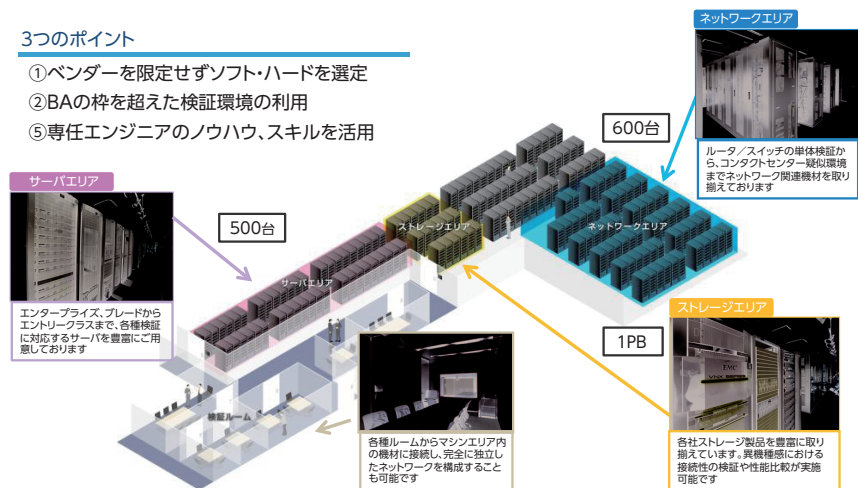


図4 検証環境「Big Data Processing Lab (BPL)」

ビッグデータ活用を支援するサービスを「CTC BD-Navi(ビッグデータ・ナビゲーションサービス)」として体系化。データ分析・活用のコンサルティングからシステム基盤の導入支援まで、全部で8つのサービスを提供します

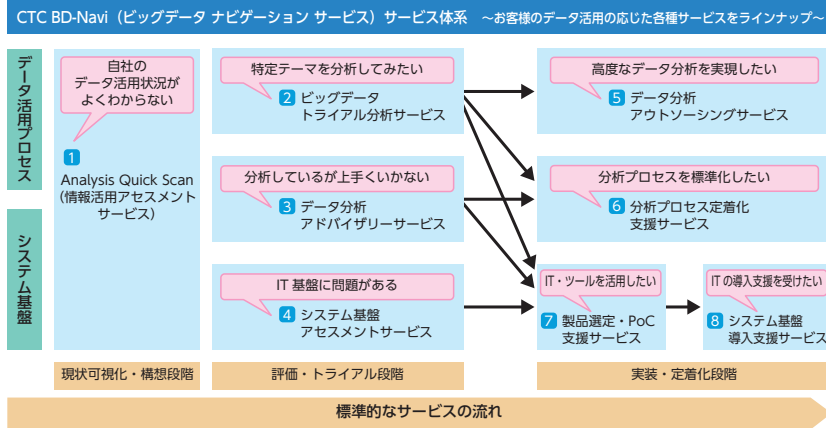


図5 ビッグデータ活用支援サービス「CTC BD-Navi」の概要

このTSC内にビッグデータ関連の製品や技術の検証を目的とする検証環境「Big Data Processing Lab (以下、BPL)」を開設した。同社の顧客は、導入しようと考えている製品や技術の動作をBPLで事前に検証できる。「ビッグデータ・アナリティクスでは多数のソフトウェアとハードウェアを組み合わせるシステムを構築します。組み合わせ方や処理するデータによっては、お客様の望む処理速度や機能を提供できないこともあります。例えば、HadoopとRDBMSを連携させるソフトウェアは多数ありますが、あるソフトウェアはデータ量が一定量を超えるとクエリが失敗することが分かっています。BPLを利用することで、システムが必要な能力を備えているかどうかをお客様自身が試していただくことができます。もちろん、私たちのエンジニアが検証を支援することや、検証を代行することも可能です。」(小林氏)

データ活用支援サービス群を「CTC BD-Navi」として体系化

ビッグデータを活用して新たなビジネス価値を創造したいと考える顧客が増えている。しかし、顧客自身が「何をどうしたい」というゴールを明確に意識しているケースはそれほど多くないという。そこでCTCでは、ビッグデータ活用を支援する代表的な8つのサービスを体系化し、どの場合にどのサービスを利用するのが適切かをナビゲートする「CTC BD-Navi (ビッグデータ・ナビゲーションサービス)」(図5)の提供を2014年4月に開始した。

CTC BD-Naviではサービスへの「入口」となる課題や目的を8つ用意し、それぞれに対して利用すべきサービスを規定している。例えば、「IT基盤に問題がある」という課題を持つユーザーに対しては、システム基盤を調査する「システム基盤アセスメントサービス」の利用を勧める。

これにより、漠然とした問題意識しか持っていない顧客でも、どの支援サービスをまず利用すべきかが分かる。さらにCTC BD-Naviでは、標準的なサービスの流れに沿って次に利用すべき支援サービスも提示する。例えば、「システム基盤アセスメントサービス」に続いては「製品選定・PoC支援サービス」→「システム基盤導入支援サービス」の流れでサービスを利用すると良いことが分かる。「お客様のニーズを引き出して、ゴールを明確化するのは本来コンサルティングの役割です。当社でもそうしたコンサルティングサービスを提供しています。しかし、コストや手間の問題から、必ずしもすべてのお客様がコンサルティングを受けられるわけではありません。そこで簡易コンサルティングサービスの的な位置付けでCTC BD-Naviの提供を始めました。これによってお客様のイメージをある程度はっきりさせることができ、双方の誤解なく作業をスムーズに進めることができます。」(CTC ビッグデータ・アナリティクス部 梅川 真人氏)

Hadoopに処理をオフロードして高性能システムを安価に構築

ビッグデータ・アナリティクス向けシステム基盤は複数ある。そのうちCTCが現在注力しているものの1つがHadoopである。

Hadoopの特長は主に2つある。一つは、サーバ数を増やすだけで性能を高められるスケールアウトである。もう一つは、データを比較的低

コストかつ安全に保持できるストレージとしての性質を持つことである。これらの特長を活かせば、高性能なデータ分析基盤を低コストに構築できる。2014年3月にCTCが発表したマイクロアの広告配信システムにおけるHadoop導入事例(図6)も、そうした活用例の一つだ。

マイクロアでは、サイト利用者の行動履歴データを蓄積・解析し、それに基づいて効果が高いと推測できる広告を配信するシステムを構築している。2011年の構築当初は、比較的小規模構成のDWHアプライアンスに行動履歴データを格納して分析していた。また、ビジネス用途の定型的な集計処理も同アプライアンスで実施していた。

その後、スマートフォンの爆発的な普及や同社の海外展開などを受けてデータ量が増え、記録容量や処理性能の不足が目立ってきた。そこで同社は2012年により大規模、高性能なDWHアプライアンスを追加した。しかし、それは焼け石に水で、すぐに容量と性能に限界が見えてきた。さらに大型のアプライアンスを導入すれば問題は解決できるが、それには多額の費用がかかる。

そこで同社はコストパフォーマンスの高いHadoopを導入し、定型的な集計クエリをHadoopに任せる(オフロードする)方式の検討を始めた(図6の③)。この方式では、DWHアプライアンスには整形・集計後のデータだけが格納され、クエリも分析クエリだけを担当するので、容量・性能共に余裕が生まれる

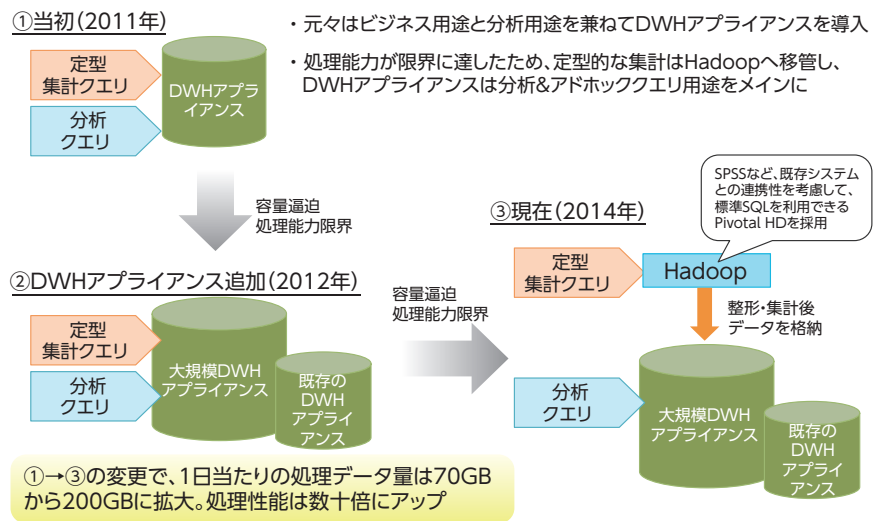


図6 マイクロアにおけるデータ分析基盤導入事例

ことになる。

ところが実際にテストしてみると想定よりも性能が悪かった。CTCに性能改善を依頼したところ、原因がSPSSなどの既存システムとの連携に使っていたSQLクエリエンジン「Hive」にあることが分かった。「調査の結果、Hiveで処理させていたクエリに時間がかかっている事が分かりました。」(梅川氏)

CTCではHiveとOSS版Hadoopの組み合わせの代わりに、高性能なSQLクエリエンジン(HAWQ)を備える商用Hadoopディストリビューション「Pivotal HD」の採用を決定した。これにより、性能は20~70倍程度向上したという。「Hadoop向けのSQLクエリエンジンは複数存在していますが、Pivotal HDの方が、エンタープライズ用途でより安心してSQLクエリを実行できると考えています。また、Pivotal HDではSPSSとの連携方法が確立されていたのも

評価ポイントでした。」(梅川氏)

Pivotal HD(Hadoop)を採用したマイクロアの新システムは2014年3月から稼働している。度重なる改修によって、当初は1日当たり70GBしか処理できなかったデータが、200Gバイトにまで増えたという。「比較的lowコストにスケールアウト性を実現できるHadoopはオフロード用途に最適です。多量のデータを格納・分析しなければならない通信事業者でもオフロード目的のHadoop採用が増えています。我々はそうしたニーズに対応すべく、さまざまな製品や実装の検証を続けています。世界的にHadoopが注目される中、Hadoop上で稼働するソフトは群雄割拠の様相を呈しています。大規模な検証環境を持つ私たちにとって、こうした状況は歓迎すべきことです。良いものがあれば積極的に採用していきたいと考えています。」(梅川氏)