

データを企業競争力につなげるには「3つのハイブリッド化」対応が必要だ

ここ数年で企業の本格的なデータ活用手段が広がり、データの構造も保管場所もワークロードもまったく違うデータを同じように取り扱えることが企業競争力の源泉となる時代になった。一方で、既存のIT資産を持つ企業ほどデータ活用環境の整備に手間取っており、データにまつわる「3つのハイブリッド化」への対応に苦慮している。これに対する有効な解決策はあるのだろうか。

※当記事は、Web「ビジネス+IT」に掲載されたものです。

■ データ活用の現状は10年前と大きく変わった

ビッグデータという言葉が言われるようになってからもう10年になる。データがビジネスにとって重要なことは引き続き変わらないが、データ活用の現状は一昔前とは違う。企業は今までなかった課題で苦慮している。

それが、データ活用環境における「3つのハイブリッド化」だ。(図1)

1つ目は、データ構造のハイブリッド化である。以前はリレーショナルデータベース(RDB)で扱いやすい構造化データがエンタープライズシステムの主役だった。

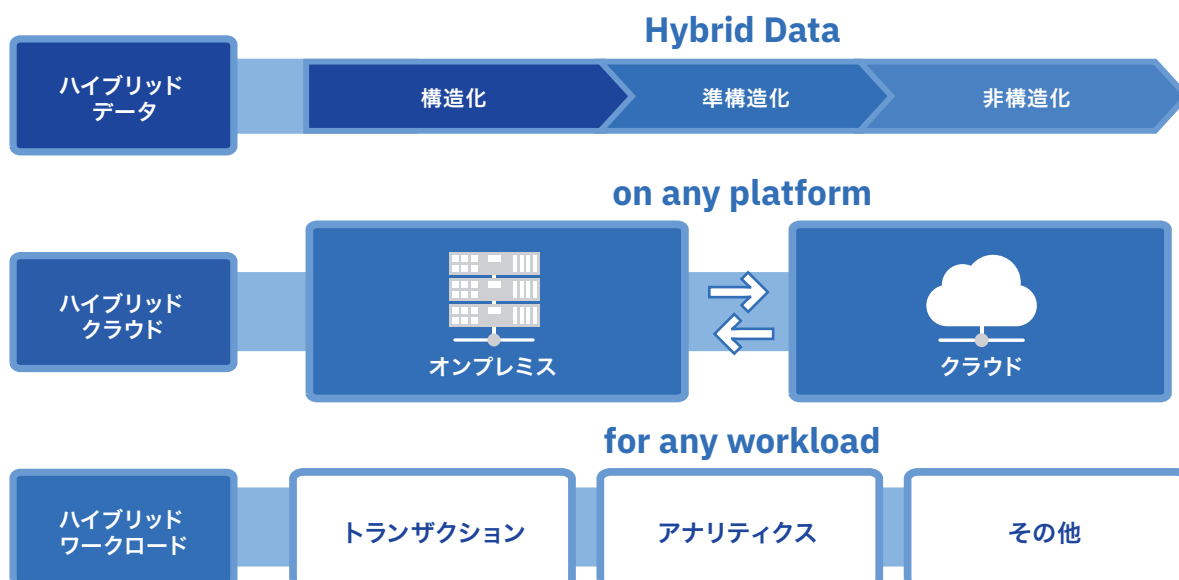
ところが現在は、準構造化データ(XML、JSON等)や非構造化データ(テキスト、音声、センサーデータ等)といった以前は捨てていたデータに、企業の競争力を生む価値があるとされるようになった。

もちろん構造化データに価値があることは変わらないが、企業は双方のデータを扱えなければならなくなった。これが1つ目のハイブリッドである。

2つ目は、データ保管場所のハイブリッド化だ。クラウドの台頭により、データをクラウドに保管する企業が増えている。すべてをクラウド化するべきかどうかは有識者の間でも見解が分かれるところだが、いずれにせよオンプレミスとクラウドの双方を連携させなければならない。

3つ目はワークロードのハイブリッド化である。処理するデータ量は少ないが高いスループットが求められるトランザクション処理と、莫大なデータ量の中からインサイトを抽出するアナリティクス処理では、アプリケーションを支えるアーキテクチャが異なる。

図1 企業がデータ活用で直面する「3つのハイブリッド化」



上記3つの観点に対して、実はそれぞれに最適化されたデータベースソリューションは存在している。しかし、各ニーズに対応したソリューションを採用する動きが進んでしまった結果、サイロ化に陥っており、ビジネスに役立つ横断型のデータベースとはほど遠い状況になってしまっている。

スタートアップなら自社にとってベストな環境を新たに構築すればいいが、既存のIT資産を抱えている企業は連携方法も複雑化し、検討にも構築にも時間がかかる。

では、こうしたデータ活用で直面する「3つのハイブリッド化」に対する有効な対応策は存在するのだろうか。

■ データの適材適所に関する3つの問題

ここで思いつくのが、従来から言われていたデータの「適材適所」というアプローチだろう。「確かにこれは理想ですが、ずっと理想を追い求めるくり返しでした」と語るのは、データ・アーキテチャーとビジネス活用の両方に詳しい日本アイ・ビー・エムの苧阪(おさか)浩輔氏だ。

たとえば、Hadoopだけでデータレイクを構築することを例に考えてみよう。大きく3つの問題が発生することになると苧阪氏は指摘する。

1つ目は、コストや組織の問題。データを1箇所に集めるためにはコストがかかるし、またデータを提出することに抵抗がある組織もある。どの組織がコストを負担するのかという問題もある。

2つ目は、技術力の問題だ。一般的にHadoopを含めたOSSをビ

ジネス用途に耐え得る品質レベルで使いこなすには、かなりの知識と技術力、そして受け入れ側企業の体制が必要となる。また、エンタープライズ用途では保守やサポートは他社にお願いせざるをえない場合も少なくなく、そうなればTCO(総コスト)につながることになるだろう。

3つ目は、データガバナンスの問題。単純にデータを集めても、それぞれのデータが何なのか、どこから来たのか、どの程度正しいものなのかが分からなくなりがちだ。

「小さな組織や部門といった単位であれば、データレイクを物理的に1箇所に集めるメリットはあります。しかし、ある程度の規模になると、一箇所に集めるデメリットのほうが大きくなるケースも増えてきます」(苧阪氏)

だとすれば、既存資産に関してはできる限り現状を維持したまま、それらに統一的にアクセスする手段が望ましいことになる。

Common SQL Engineがデータ構造のハイブリッド化を解決する

こうした課題への一つの答えがIBMの掲げる「Common SQL Engine」という概念だ。(図2)IBMは2017年7月に自社が開発・展開しているオンプレ・クラウドひっくるめたデータベース製品群を「Db2」とリブランドした。その狙いは、IBMデータベースを利用するあらゆるシーン、たとえばオンプレやクラウド、アプリケーションや手組みソフトウェア、OLTPワークロードかOLAPワークロードかに関わらず、そのすべての共通エンジンをDb2とする

図2 Common SQL Engineの考え方



ためである。すなわち、IBMのデータベースを使う場合、構造化/非構造化、オンプレ/クラウド、OLTP/OLAPに関わらずすべてがDb2エンジンで処理されるわけだ。

さらに、このDb2がゲートウェイになって、マルチベンダー、マルチリソース、マルチプレースでアクセスできるデータベースを実現しようとする取り組みがCommon SQL Engineだ。

Common SQL Engineにより、どんなデータでも、どのデータベースでも、どこにあるデータでもアクセスすることができるようになる。ここで重要なのがフェデレーション機能で、Db2ではJSON、HadoopなどさまざまなNoSQLに透過的にアクセスできるほか、Oracle DatabaseやMicrosoft SQL Server、PostgreSQL、MySQLなど他ベンダーやオープンソースのRDBにもアクセス可能だ。

特にOracleに関しては、最適なアクセスパスを選択するなどの高速化と、PL/SQLなどもそのまま実行できるなど高い互換性を実現している。

つまりDb2がフロントにあれば、あらゆるデータ構造が取り扱えるということであり、データ構造のハイブリッド化を解決できる。

さらに苧阪氏は運用負荷やコストの軽減にもつながると次のようにアピールする。

「HadoopやNoSQLのクエリーが書ける技術者は少ないですが、SQL文を書ける技術者は数多くいます。Common SQL EngineのあらゆるデータソースへSQLでアクセスできれば、技術者不足を解消できますし、既存資産の活用も楽になります」(苧阪氏)

コストの面で特に注目したいのがライセンス体系だ。仮想コア数のみ、国内外すべてのクラウドに適用といった保管場所の移動を考慮したものとなっている。

膨大なトランザクションをさばく 企業に導入済み

さらにワークロードのハイブリッド化にも対応する。

たとえば、Apache Sparkで機械学習環境を構築するには、Hadoopのアクセス・クエリーを使う必要があったが、これは

OSSに不慣れな技術者には難易度が高い。一方で、IBMが提供するDb2互換のBig SQLを利用すれば、ANSI SQLと高い互換性のあるSQLでアクセスできるようになり、機械学習環境の構築が極めて容易になる。

この新しいHadoop/SparkとBig SQLの組み合わせは、すでに国内大手流通業で実際に導入された事例があるという。

「そのお客さまがこの方式を採用したのは、従来の非構造化データへのバッチ処理に加えて、膨大なデータが必要なアナリティクス処理と使い慣れたSQL処理を、同一のプラットフォームで処理したかったからです。このニーズにDb2はピッタリと当てはまりました」(苧阪氏)

なお、同じプラットフォーム上で、トランザクションとアナリティクスの両立を可能にしているのがDb2 BLUアクセラレーションという機能だ。

従来のRDBでは、行単位でデータをストアしているが、BLUアクセラレーションでは、カラム単位でまとめてストアする。検索処理では不要なカラムがたくさんあるが、不要な分のカラム・データへのディスクIOの発生やキャッシュ・アクセスをさせず、かつ高度なデータ圧縮アルゴリズムを組み合わせることで、大幅な高速化を実現している。

実測データでは、アナリティクスのスピードが35倍から73倍改善し、一部のクエリーでは1,400倍以上の改善が見られた。またデータベースサイズが15分の1になったという報告もある。

アナリティクスを高速化すると同時に、RDBなのでトランザクション処理にも対応できる。つまり、ワークロードのハイブリッド化にも対応できているということだ。これはDb2の大きな強みだといえるだろう。現在IBMでは、このBLUテクノロジーを土台に、今以上のトランザクションを高速処理できる改良に取り組んでおり、実際のオペレーショナル・アナリティクス処理基盤にも適応できるようになっているという。

「Db2は昔からITエンジニアに余計な手間をかけさせないことを開発思想にしてきました。テーブル情報を見てDBエンジンが自動的にチューニングする機能がバージョン8から搭載されているのがその一例でしょう。BLUアクセラレーションも以前から提供していますが、その延長線上で出てきた技術なのです」(苧阪氏)。

■ DB2からDb2へリブランディングの狙い

本記事を読みながら気づいた方もいたかもしれませんが、IBMはデータベース製品を「DB2」から「Db2」にリブランディングしている。

DB2ファミリー製品をすべて「Db2」に改称したほか、「dashDB for Analytics」を「Db2 Warehouse on Cloud」にするなど製品名も変更した。

「『DB2』は長い伝統があるため、オンプレミスのDBというイメージが強くあると思います。しかし、今回のリブランディングでは、Common SQL Engineの考え方に基づき、オープンで、かつクラウド上のDBもアプライアンス製品もすべて同じエンジンを使うというIBMの意志の表れでもあります」(苧阪氏)

ITインフラを支えるこれらの技術はなかなか目立たないが、データ活用で課題に直面する企業ニーズに間違いなく応えるものだ。ここ数年、Db2のシェアは伸長傾向にあるが、顧客課題の解決にフォーカスした地道な技術イノベーションを実現していることが評価されているからかもしれない。

詳細はこちら

<http://ibm.biz/hybridmanagement>



日本アイ・ビー・エム株式会社

©Copyright IBM Japan, Ltd. 2018

〒103-8510 東京都中央区日本橋箱崎町19-21

このカタログの情報は2018年4月現在のものです。仕様は予告なく変更される場合があります。記載の事例は特定のお客様に関するものであり、全ての場合において同等の効果が得られることを意味するものではありません。効果はお客様の環境その他の要因によって異なります。製品、サービスなどの詳細については、弊社もしくはビジネス・パートナーの営業担当員にご相談ください。IBM、IBMロゴおよびibm.comは、世界の多くの国で登録されたInternational Business Machines Corp.の商標です。他の製品名およびサービス名等は、それぞれIBMまたは各社の商標である場合があります。現時点でのIBM商標リストについてはwww.ibm.com/legal/copytrade.shtmlをご覧ください。