

ディープラーニング活用でわかった、 IoT/ビッグデータ/AI 連携時代の IT インフラ

WHITE PAPER

IoT/ビッグデータ/AI が連携し新たな価値を生み出し始めています。クラウドリアンは、AI（人工知能）として注目を集めるディープラーニングを使い、ビデオカメラで撮影した映像から走行車種をリアルタイムに自動認識する技術を開発し、道路沿いに設置するデジタルサイネージにターゲット広告を表示する実証実験に着手しました。この開発経験から、IoT/ビッグデータ/AI が連携するソリューションやビジネスを持続的に維持し、成長させるためには、IT インフラが重要な役割を果たすことがわかっています。本ホワイトペーパーでは、IoT/ビッグデータ/AI 連携が本格化する時代を先取りして、まさに今から準備すべき IT インフラについて考えてみます。



IoT（モノのインターネット）は、インターネット経由で世界中のデバイスやセンサーからデータを集め、いわゆるビッグデータを生み出します。このビッグデータから、OSS（オープンソースソフトウェア）として登場し身近になった AI を使い、新たな知見を得られるようになってきました。

その一方で、従来型の IT インフラは、IoT、ビッグデータ、AI 登場以前の IT 環境への対応を前提としています。それは、インターネット接続に対し堅牢に防護され、ビッグデータの処理と保存に多額の投資を覚悟せざるを得ず、データ急増に伴う増設・置換も容易ではありません。特に、AI の学習過程では数多くの試行錯誤が必要となるものの、従来型の IT インフラでは迅速かつ柔軟な対応が難しいのが現実です。

IoT/ビッグデータ/AI 連携時代の IT インフラとは、どのような要件を備えているべきなのでしょう。

IoT/ビッグデータ/AI の連携により、それぞれの課題を解決

これからは、インターネットビジネスだけではなく、あらゆる業種において IoT/ビッグデータ/AI が連携し、新たな価値を創出するといった事例が増えてきます。それは、連携によりそれぞれが抱えてきた課題を解決できるようになってきたからです。

■ IoT が直面する大量データ

モノのインターネットというコンセプト自体は新しいものではありませんが、特に最近、IoT について見聞きすることが増えています。IoT に最適なデータ通信網やサービスが整い、デバイスやセンサーが小型化・多様化し、インダストリー 4.0 のような産官での業界育成の動きなどが、その普及を促進しているからです。

IoT の展開初期や、取扱量が少なければ、従来型の IT インフラで対応できるかもしれませんが、しかし、一旦始めれば、24 時間 365 日、世界中にばらまかれる IoT デバイスが大量のデータを集め続けます。ビジネスが軌道に乗ればすぐに、これらのビッグデータを経済的、効率的に扱い、単に稼働状況の「見える化」に留まらない、新たな価値を生み出すためにどうすべきか？という課題に直面することになります。

■ ビッグデータ分析には専門家が必要？

この課題への技術的な対応は、すでに始まっています。Hadoop や Spark などの OSS（オープンソースソフトウェア）の分散処理技術や、大量データを経済的に保存できるオブジェクトストレージを使うことです。

残る課題は、そのビッグデータから何をえられるのかという点です。分類・整理のうえ、分析して新たな知見を得たいとなれば、データサイエンティストなどの高度な専門性が求められます。

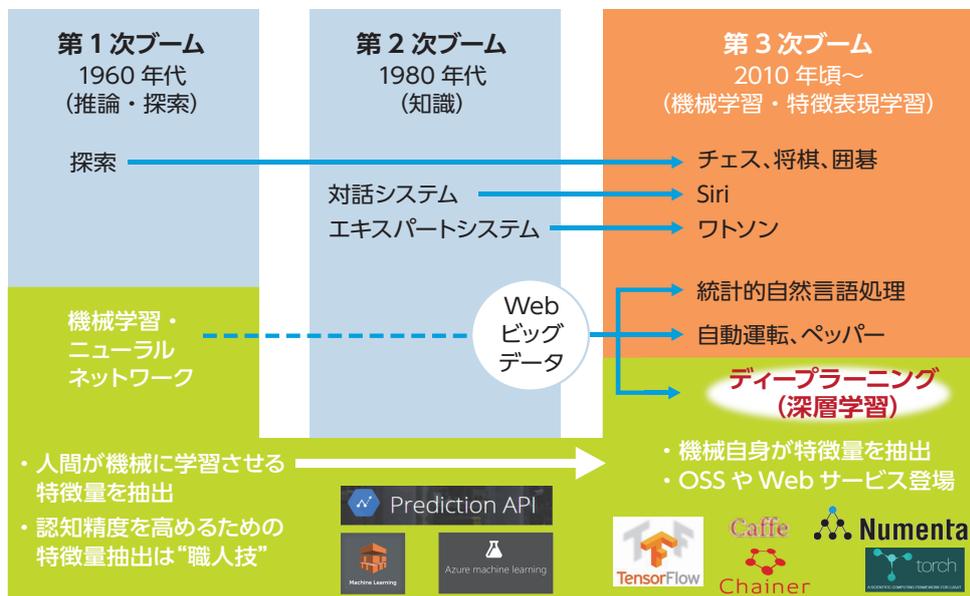
■ AI によりビッグデータの分類・整理、分析を自動化

このビッグデータの分類・整理、そして分析といった課題の解決には、OSS として登場し、身近になった AI を使うことが、ひとつの方法です。

AI は、現在、第 3 次ブームを迎えています。その背景には、人間の棋士に勝ったことで話題になったグーグルの AI 「アルファ碁」にも使われたディープラーニングの登場があります。

このディープラーニングは、第 1 次ブームからある機械学習が進化したものです。機械学習では、人間が機械に学習させる特徴量を定める必要がありました。この特徴量の抽出が認識精度を左右するため、職人技がいられていました。一方のディープラーニングでは、機械自身が大量のサンプルから特徴量を抽出します。そのため、認識精度を高める過程の多くの部分を機械に任せることができます。

AI(人工知能)は第3次ブームといわれる



出典：「人工知能は人間を超えるか」松尾 豊著、角川選書刊の分類を参照

IoT/ビッグデータ/AIが連携する「走行車種リアルタイム自動判別」

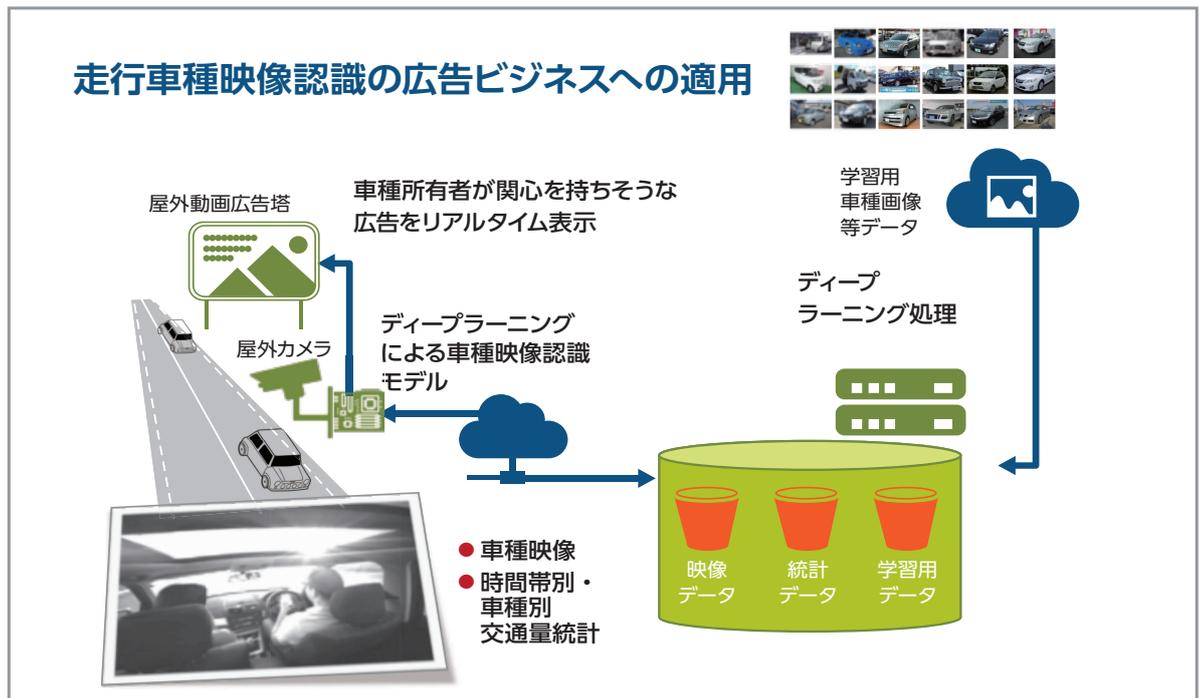
クラウドファンでは、ビデオカメラで撮影した映像から走行車種を自動判別し、道路沿いに設置されたデジタルサイネージに、車種の運転者が関心を持ちそうな広告をリアルタイムに表示する実証実験に着手しました。これまでは、車の運転者の関心とは無関係に、年間走行量や時間帯に基づき屋外広告を流していましたが、この取り組みによりターゲット広告を表示できるようになります。この車種を自動判別する画像認識技術はディープラーニングを使い開発しています。

ただし、クラウドファンにおける開発経験からわかるのは、このディープラーニングを使えば、簡単に高い認識率を得られるかということ、それほど単純ではないということです。学習を何度も繰り返すことにより、より誤りの少ない特徴を抽出し、その認識率を高めなければなりません。

特に、学習方法によって結果が大きく変わります。学習させる画像の種類、その枚数、学習速度、そして各種パラメータ等を調整した、いくつもの異なる方法です。学習方法が悪いと、いくら学習を繰り返しても認識率はあがりません。良い結果を得るためには、何度も試行錯誤を繰り返すことが必要です。そういった試行錯誤を柔軟に、迅速に実行できる IT インフラでない場合には、開発のスピードに大きな影響を与えてしまいます。

また、学習過程の画像は多くあれば多くあるほど認識精度があがります。正しく車種を認識していることを確認するためには、実際の映像を保存し、レビューし、修正するサイクルも必要になります。さらに、長期間に亘り統計データを蓄積することも重要です。いずれも大量のデータです。それらを経済的に長期間に亘り保存しておける IT インフラが求められます。

IoT/ビッグデータ/AIが連携するソリューションを提供するには、このような多種多様で膨大なデータ量を迅速に処理できる IT インフラが必要であることは当然ですが、その他にも事業部門やアプリ開発部門から、IT インフラ部門には様々な期待が寄せられます。



IT インフラへの期待と悩み

たとえば、IoT/ビッグデータ/AI連携ソリューションを企画開発する事業部門や開発部門と、ITインフラ部門の間では、以下のようなやりとりがなされることでしょう。

【企画開発】「世界展開も含め、多くの拠点で実施したい」

【ITインフラ】「初期から大規模システムを調達するのではなく、小規模から段階的に進めたい。各拠点における実施作業も標準化したい」

【企画開発】「IoTはインターネット経由で直接データの読み書きをしたい」

【ITインフラ】「ファイルアクセスプロトコルではなく、HTTP/HTTPSに対応したストレージを使いたい。暗号化も担保しておきたい。」

【企画開発】「AIの認識精度を高めるために、『開発しながら試す』を繰り返したい」

【ITインフラ】「その都度、ITインフラの調達や設定に人手も時間もかかる。ITインフラ側でもDevOpsの仕組みを整えたい」

【企画開発】「パブリッククラウドも使い開発、運用したい」

【ITインフラ】「パブリッククラウドを透過的に使い、社内システムとパブリッククラウドがサイロになるのは避けたい。」

こういった部門間のやり取りは、IoT/ビッグデータ/AI連携ソリューションに限ったことではありません。新事業展開を検討する都度、繰り返し行われているはずです。そして、これからますます増えることになるでしょう。このような議論に時間を使わずに済むITインフラを求める声は大きくなるばかりです。

IoT/ビッグデータ/AI連携時代のITインフラに求められる要件

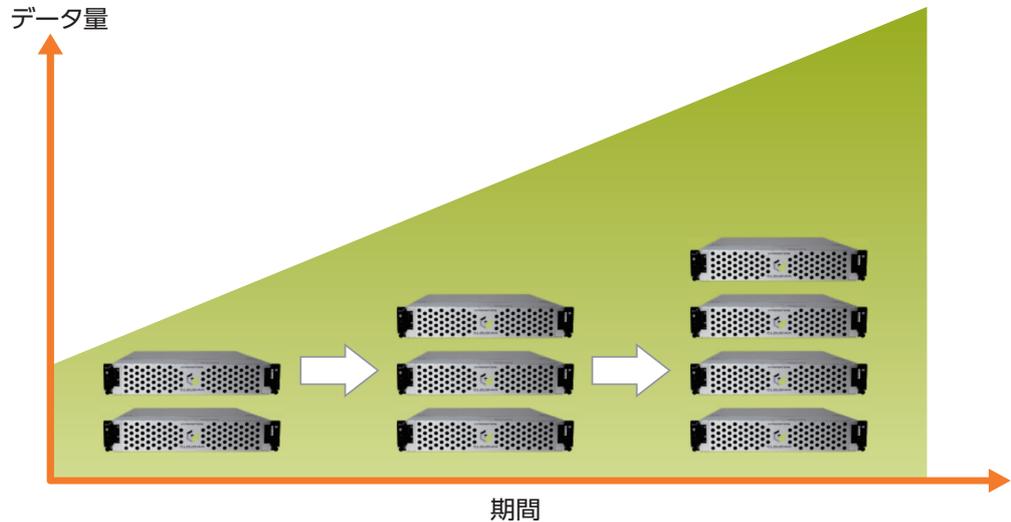
それでは、IoT/ビッグデータ/AI時代に求められるITインフラは、どうあるべきなのでしょうか。代表的な要件を考えてみましょう。

■ スケールアウト

小規模からスタートし、データが増えれば、その都度、設備を追加することで全体のストレージ容量を拡張（スケールアウト）できるITインフラであるべきです。従来型のストレージ装置のように、容量が足りなくなれば、置換・増設をしてデータの入れ替えを繰り返すのでは持続的な展開は難しいと言わざるを得ません。

小規模からスタートしスケールアウト

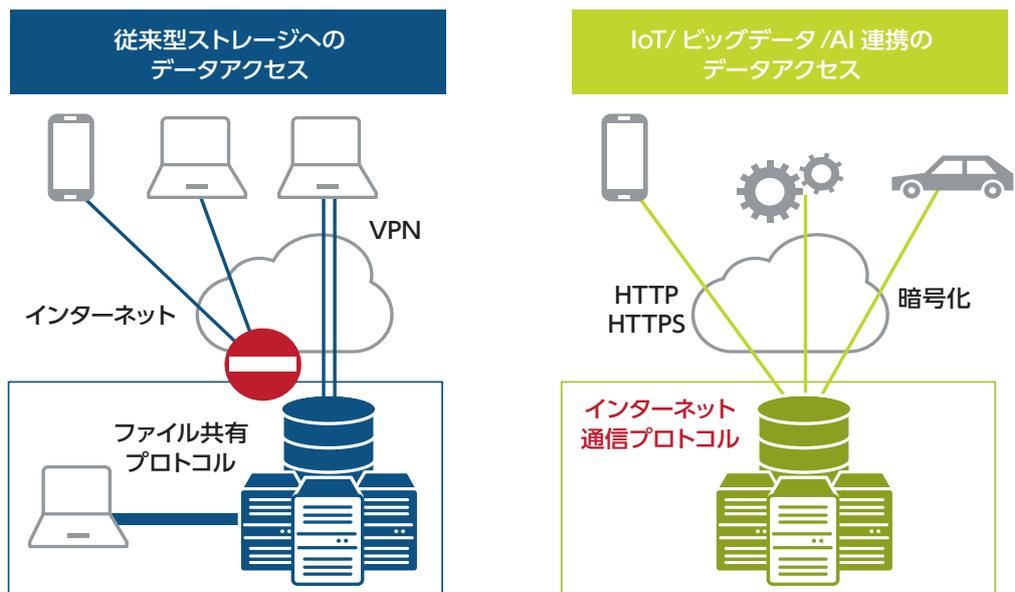
小規模からスタートし、データ量が増えれば、設備を追加するだけで、自動的に全体のストレージ容量を拡張



■ インターネットからの直接アクセス

HTTPS(SSL 通信) のようなインターネット通信プロトコルでデータの読み書きができ、クライアントとサーバー間、またサーバー間での通信暗号化にも対応し、インターネット経由でも安心してデータを送ることができる IT インフラであるべきです。従来型 IT インフラのファイルアクセスプロトコルは、外部から安全に隔離された企業構内で、かつ近距離で使われることを前提としており、インターネットを経由したデータの読み書きには適していません。

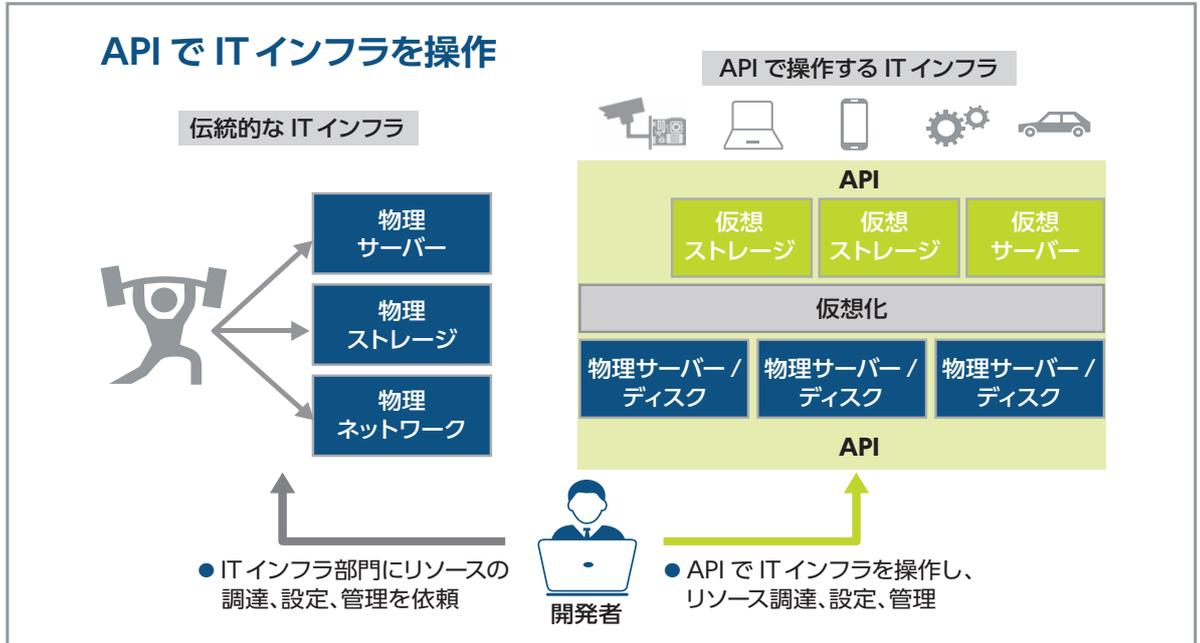
インターネットから直接アクセス



■ API で操作

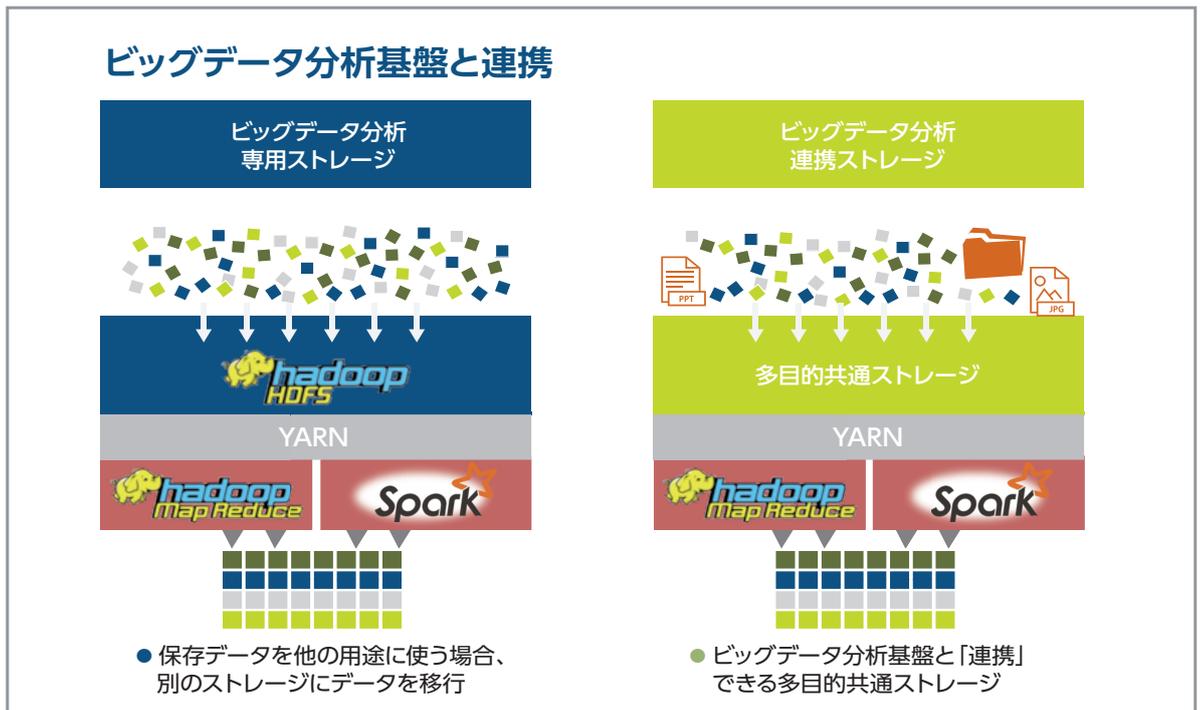
パブリッククラウドサービスのように、IT インフラには API (Application Programming Interface) を備え、アプリケーションの開発者自身がリソースを調達、設定、操作できる IT インフラであるべきです。AI で試行錯誤する際でも、この API を使ってストレージのリソースを調達できるので、IT インフラ担当者は、その都度の力作業から解放されます。従来型の IT インフラは、物理サーバー、物理ストレージ、物理ネットワークです。サーバーは仮想化

されはじめましたが、未だストレージは物理装置のままであるケースが多いと言えます。この場合、開発プロジェクトが立ち上がると IT インフラ側では、ストレージ装置を調達したり、ボリューム割当等の設定作業といった、人手と時間のかかる作業をします。これでは IT インフラが開発速度を遅らせてしまいます。



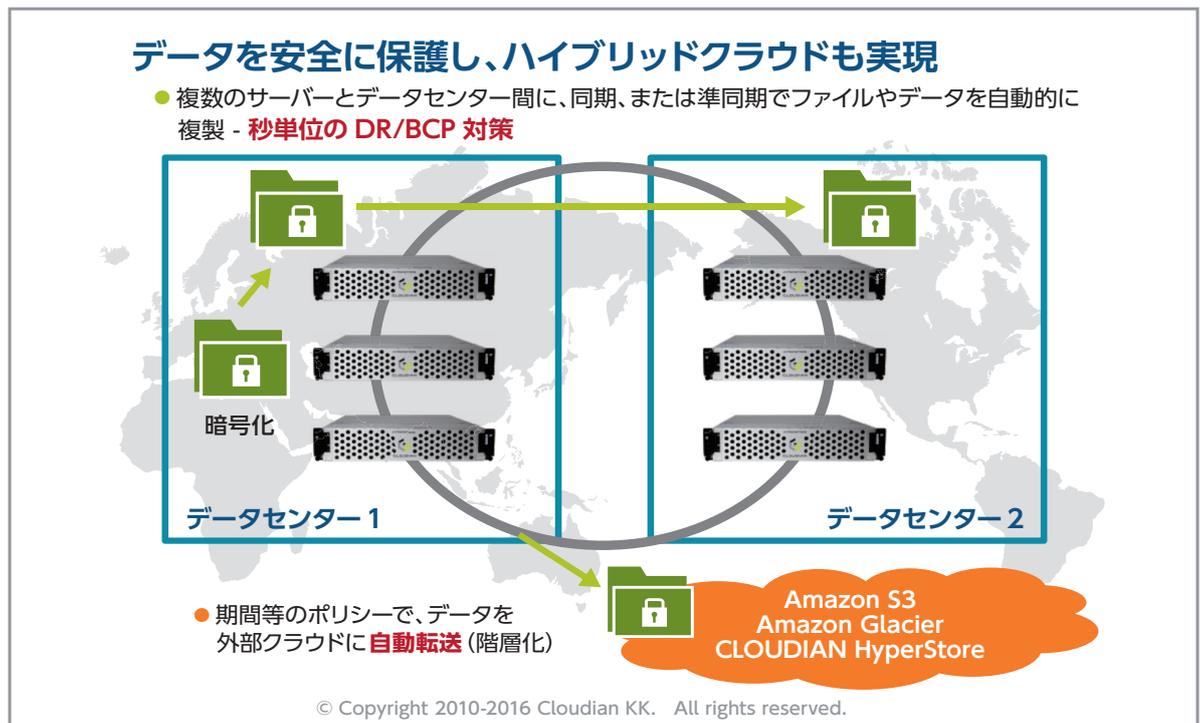
■ ビッグデータ分析基盤との連携

Hadoop や Spark といったビッグデータ分析基盤と連携できる IT インフラであるべきです。Hadoop の HDFS (Hadoop Data File System) は、Hadoop MapReduce 処理をするためのデータ格納場所です。Hadoop MapReduce 処理以外の目的で HDFS に格納されたデータを使うためには、他のストレージへのデータ移行が必要になります。つまり、余分なストレージリソースの用意が必要で、ストレージの管理対象が増えてしまいます。



■ DR と BCP 対策

当然のことながら、DR (Disaster Recovery:災害復旧) や BCP (Business Continuity Plan:事業継続計画) のために、複数のデータセンター間が連携してデータ保護し、経済的に対応できる IT インフラであるべきです。また、経済性や冗長性を高めるため、パブリッククラウドと連携できることも必要です。従来型の IT インフラでは、バックアップ用のソフトウェアやオプションサービスの追加購入が一般的であり、運用面での負担も大きいのが現状です。



■ 多目的に使える

エンタープライズ IT で普通に使うアプリケーションとともに、あらゆるデータの保管ができ、多目的に使える IT インフラであるべきです。独自インターフェースを備えた IT インフラの場合、アプリケーション改修等に余計なリソースが必要になります。

CLOUDIAN HyperStore について

クラウドファンが開発提供する「CLOUDIAN HyperStore」は、ここまで解説した IoT/ビッグデータ/AI 連携時代に求められる全ての要件を満たすソフトウェア定義のオブジェクトストレージ製品です。汎用的で、経済的なサーバーをハードウェアに使うため、高価な専用ストレージ装置と比べ、低コストで大容量のデータを扱うことができます。ソフトウェア製品だけではなく、汎用的なサーバーとセットにしたアプライアンス製品としても提供しています。

特に「CLOUDIAN HyperStore」は、Amazon Web Services (AWS) のクラウドストレージサービス、Amazon S3 と同等のストレージシステムをオンプレミス、すなわち企業が管理するプライベート/オンプレミス環境（データセンター等）に構築できるという点に大きな特長があります。

スケールメリットがあり、経済的な料金で利用できる Amazon S3 のようなパブリッククラウドは、多くの企業で利用されはじめています。しかし、外部にデータを預けられない事情があったり、一部のデータ保存だけにしか使えない、または既に使っているものの、データ量が増えて利用料が高額といったように、パブリッククラウドを適切な方法で利用したいとの声を聞くことも増えています。

「CLOUDIAN HyperStore」は、この Amazon S3 がアプリケーションに提供するインターフェースである S3 API に完全準拠しており、Amazon S3 に対応するアプリ側を改修することなく活用できます。Amazon S3 は、クラウド

ドストレージの事実上の標準であり、Amazon S3 を利用するアプリケーション、ツール、製品、サービスは数百種類もあると言われています。そして、現在企業で一般的に利用されるアプリケーションの多くが S3 API に対応しています。

CLOUDIAN HyperStore の特長

Amazon S3 と同等のストレージシステムを構築し、運用できる

S3 API 対応アプリ等を改修せずに活用できる



そのため、「CLOUDIAN HyperStore」でオンプレミスに Amazon S3 と同等のオブジェクトストレージ環境を構築運用すれば、社内の利用者にパブリッククラウドの利便性を提供できるだけでなく、ひとつのアプリケーションを使いながら、オンプレミスとパブリッククラウドをハイブリッドに使い分けることができます。

なお、日本生まれの「CLOUDIAN HyperStore」は、ニフティ様、NTT コミュニケーションズ様、NTT 東日本様を始め、世界でも特に品質を重視することで有名なお客様に長年にわたりご利用いただいている日本品質の製品です。海外では日本よりもはるかに多くのお客様が利用されており、日本流のきめ細かな製品サポートにも高い評価を得ています。

さいごに

従来、ビッグデータはインターネット企業のものでしたが、IoT により実世界のあらゆる出来事がビッグデータ化されていきます。これからの時代は、インターネット企業だけでなく、あらゆる業種において IoT / ビッグデータ / AI 連携への対応が求められるでしょう。事業部門や開発部門では、すでに検討をはじめています。やがてすぐに、本日で紹介したような期待が IT インフラ部門に寄せられることになるはずで

スピードが求められる時代です。周りからの声が大きくなってから IT インフラの検討を始めるのでは遅すぎます。時代とニーズを先取りし、いまから準備しておくことが重要です。