



デジタル時代のマルチクラウド構成にも最適

Amazon S3とAPI互換でハイブリッドに 構成できるIBM Cloud Object Storageとは？

NASでは不得手な大容量データの格納先として注目を集める「オブジェクトストレージ」。企業で利用する場合は何を選択すべきなのだろうか。

ビッグデータの保存先として、またクラウドネイティブなストレージとして、「オブジェクトストレージ」が注目を集めている。既に数々のサービスや製品が展開されているが、2015年にIBMが買収した「Cleversafe」を元にしてIBMクラウドからも提供している「IBM Cloud Object Storage」は、有力な選択肢となるはずだ。日本IBMクラウド事業統括 Cleversafe 営業部 担当部長 コンサルティング セールス スペシャリスト 田中克哉氏は「独自のデータ分散技術『Secure Slice (セキュア・スライス)』を使った高い信頼性や、パブリック/デディケイテッド (専有) / オンプレミスの3つの提供モデルで利用できる柔軟性といった特長に加え、データ暗号化などのエンタープライズ用途では欠かせない機能を標準で持っている。また、独自にデータ消失への対策が施されており、仮に物理的なデバイス盗難があった場合でも情報を保護できるため、機密性の高いデータの保存場所としても期待されている」と説明する。

加えて、オブジェクトストレージのデファクトスタンダードともいえる「Amazon Simple Storage Service (Amazon S3)」や「OpenStack Swift」のAPIとも高い互換性を持つことから、Amazon S3にも精通している日本IBMクラウド事業統括 担当部長 コンサルティング・アーキテクト 平山 毅氏は、「Amazon S3のAPIを組み込んだクラウドネイティブアプリケーションやサードパーティーツールに対しても、認証情報とエンドポイントを変えるだけでそのまま使えるため、Amazon S3やOpenStack Swiftとマルチクラウドを構成したり、移行したりしやすい。最近では、特定のクラウドベンダーのオブジェクトストレージが障害になる際のリスク対策として複数のクラウドのオブジェクトストレージを組み合わせるアーキテクチャも注目されているため、その場合にもIBM Cloud Object Storageは、最適な選択肢になります。また、デジタル時代のラージオブジェクトにも対応し、プライベートなエンドポイントや1つのバケットに対するマルチリージョン、などが標準で対応されており、ハイブリッドクラウド環境でも同じアーキテクチャで構成できる点などが、Amazon S3に対しての差別化ポイント

トになります」と説明する。

つまり、セキュリティや運用負荷の課題、アプリケーション開発コストの課題、無停止での運用、ラージオブジェクト、プライベートエンドポイント、専有モデル、API互換、マルチリージョン、ハイブリッドクラウド、といったエンタープライズならではの要件をことごとくクリアしているのが、本稿で紹介するIBM Cloud Object Storageなのだ。以降では、機能詳細やその特長を解説していく。

そもそも「オブジェクトストレージ」はなぜ必要なのか？

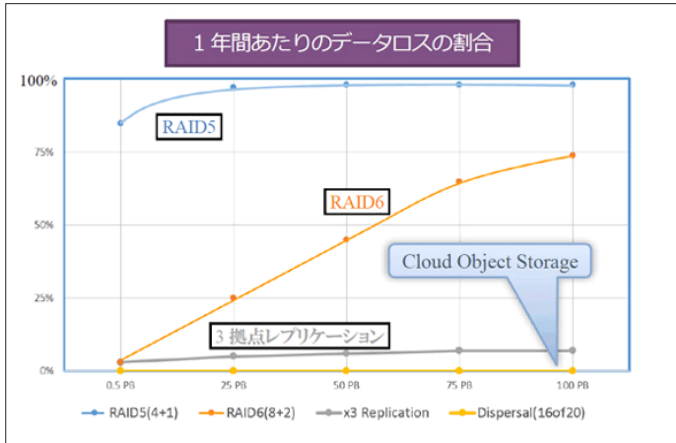
従来型のストレージ (やファイルシステム) がデータを階層で管理するのに対し、オブジェクトストレージは、一般的にはデータを入れるエリアである「バケット」と実ファイルである「オブジェクト」の2つから構成され、全てのデータをフラットに扱う。バケットとオブジェクトの組み合わせで生成される一意の「データID (キー)」があれば、長いパスをたどらなくてもすぐに目的のデータを取り出すことができる。こうした特性は、データが大きくなると非常に有効だ。というのも、ストレージがペタバイトを超えると、従来の代表的な信頼性を保つ技術であったRAIDによるデータ保護が限界に至るのだ。

例えばRAID 5ならパリティを使い「エラーがあれば復旧する」ことでデータの整合性を保つ。しかし、ペタバイトを超える環境では「常にどこかでエラー処理が動いている状態」となりがちで、実運用に耐える性能が出しにくくなるのだ。また、災害対策も含めて、データを保護するには、例えば、別拠点に対してストレージレプリケーションを行う方式も考えられるが、当然保存するデータ量は、拠点数が増える分だけ増えてしまう。オブジェクトストレージの場合、RAIDなどの構成が不要で、その代わりに複数のノードにデータを分散して格納することで冗長化し、データを保護する。この仕組みであれば、RAIDと比べてもデータ消失のリスクは少ない。かつRAIDやストレージに対する管理工数も不要にできる。そこで、ビッグデータの保存先として期待されているのがオブジェクトスト

IBM 平山氏が解説する Amazon S3 API 互換と差別化ポイント (出典: 日本 IBM)

The screenshot shows the IBM Cloud Object Storage management interface. At the top, it displays 'Account ID - S3 API Type' and '0 バイト Account Usage'. Under 'アカウントの資格情報', there are fields for 'Access Key ID' and 'Secret Access Key', which are circled in red. A red annotation says 'Amazon S3 と同じ認証キー'. Below, under 'Authentication Endpoints', there are two sections: 'PUBLIC' and 'PRIVATE'. The 'PRIVATE' section lists endpoints for 'us-geo', 'Dallas', 'Washington', and 'San Jose'. The 'Dallas' endpoint is circled in red, with a red annotation saying '個別の エンドポイント'. A red annotation at the bottom left says 'プライベートIP が返る エンドポイント'.

RAID 5 ではペタバイト級のデータ量になるとほぼ確実にデータロスを起こす（出典：日本 IBM）



日本 IBM クラウド事業統括 担当部長
コンサルティング・アーキテクト
平山 毅氏

レージというわけだ。RAID などの仕組みを使わずに内部でデータの整合性をとる技術が搭載され、ペタバイトを超えても安定して利用できる。

もう1つ、今やすっかり普及したクラウドもオブジェクトストレージの勢いを後押ししている。クラウドのメリットはサイジングが不要で柔軟に拡張できる点にあるが、従来型のファイルシステムを使って運用してしまうと、せっかく拡張して

も、その都度ディスクの割り当て操作が必要になってしまい、クラウドのスケールメリットを最大限に生かせない。そのため、クラウドではファイルの保存場所としてオブジェクトストレージを採用することが多くなっている。平山氏は「特に、近年では IaaS を使ったシステムだけではなく、PaaS やサーバーレスアーキテクチャを採用したクラウドネイティブアプリケーションも急増しています。PaaS を使ったアプリケーションでは当たり前ですが、アプリケー

ションからは OS が見えません。OS のファイルシステムにファイルを配置することができないため、必然的にファイルの格納先として、クラウドで提供されているオブジェクトストレージが選択されていくのです」と話す。

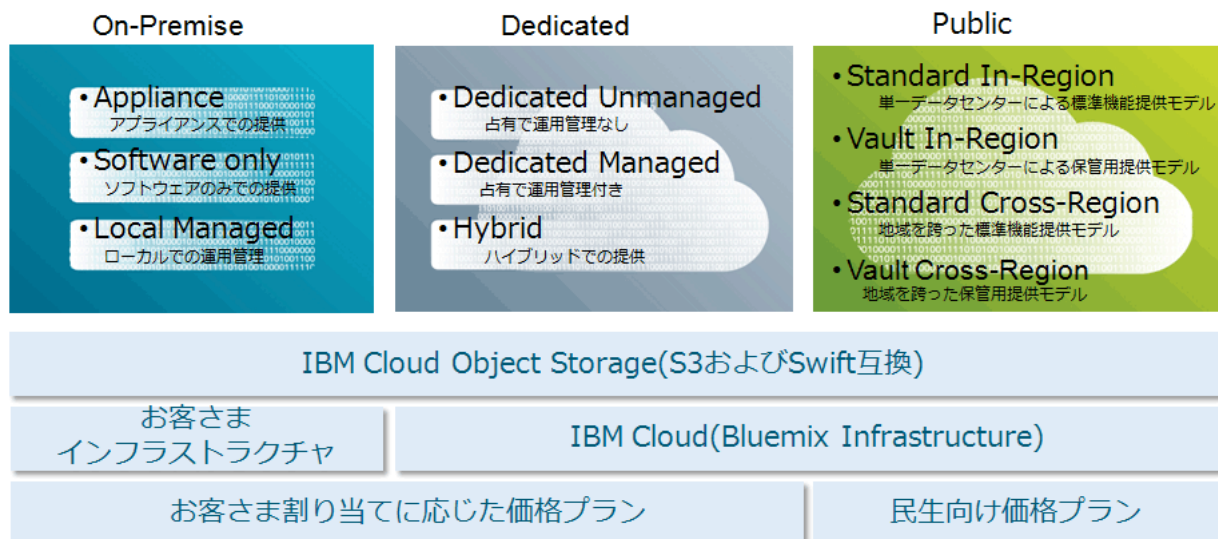
エクサバイト級データでの安定した運用実績、デファクトのAPI仕様と互換性を重視

IBM Cloud Object Storage は 2015 年に IBM が買収した「Cleversafe」がベースになっているため、既にエクサバイト級の環境での稼働実績がある他、例えば「ゲノム解析データ」や「高画質の長時間動画」のように巨大なファイルサイズであっても性能に影響なく扱える点が特長だ。アナリティクスやコグニティブの世界では、1 つ当たりのファイルサイズが以前よりも大きくなることが考えられるため、将来性を考えた場合にも優位といえる。

さらに、このストレージを操作するアプリケーションの開発ではオープンな API を提供していることにより、エコシステムを活用できる。具体的には Amazon S3、オープンソースの IaaS 基盤ソフトウェア「OpenStack」のストレージコンポーネント「Swift」の API と互換性があるため、平山氏によれば「API が Amazon S3 や OpenStack Swift に対応していますから、ファイル管理やバックアップなど、世の中に出ているソリューションの多くをそのまま、認証情報とエンドポイントを変えるだけでアプリケーションの改修なしに利用できます。『AWS CLI (Command Line Interface)』もそのまま使えます」と話す。

IBM Cloud Object Storage は、パブリッククラウドだけでなく、企業の要望に応じてデディケイテッド（専有）やオンプレミスでも利用できる点も大きな特長だ。一般的なクラウドオブジェクトストレージはクラウド環境でハードウェアリソース共有モデルでの提供に限られるが、IBM Cloud Object Storage ならば従量課金制のリソース共有型パブリッククラウドだけでなく、IBM クラウドのデータセンター内でネットワークやハードウェアリソースを完全に専有することも、オンプレミスの自社データセンターで構築する

IBM Cloud Object Storage の提供オプション（出典：日本 IBM）



こともできる。ハードウェア専有で利用できるのは Bluemix を持つ IBM ならではの強みだろう。「オンプレミスとクラウドを組み合わせたハイブリッド環境も構築できますし、一般的なパブリッククラウドでは難しい、海外 3 つのリージョンに分散した構成も可能です。『オブジェクトストレージは使いたいが、機密性の高いデータだから絶対にインターネット経由では使いたくない』『ハードウェアは専有したい、でも S3 対応のソリューションを使いたい』というニーズは多く、引き合いも多数あります」（平山氏）

容量やアクセス数が大きくなっても 性能が落ちず、フレキシブルな対応が可能

さらに、専有環境だからこそできるのが、パフォーマンス対応の設定だ。オブジェクトが大きくなると使い方によって書き込みが多いものもあれば、読み込みが多いものもあり、また、同時アクセス数もパフォーマンスに左右する。

「共有型ではセルフチューニングが働くので、ある程度まではシステム任せでもパフォーマンスは落ちませんが、バケット当たりアクセス数やオブジェクトサイズごとに性能限界があります。一方のデディケイテッドはリソースを専有できるため、チューニングの余地が大きいのです。ベースとなっている Cleversafe ストレージのセルフチューニング機能が優れていることに加え、『書き込み (PUT) だけを早くする』『読み込み (GET) だけを早くする』あるいは『ピークに耐えられるようにリソース配分を調整する』『暗号化方式の選択を変える』など細かな調整が可能です」（平山氏）という。

データロス「120億年に1回」 情報伝播アルゴリズムの堅牢さ

IBM Cloud Object Storage の信頼性を支える技術が「IDA (Information Dispersed Algorithm、情報伝播アルゴリズム)」だ。IDA ではまず、信頼性の基準となる 3 つの数字を設定する。標準的なケースでは「12/7/8」と設定する。この場合は、データを 7 分割し、5 つのイレイジャーコード（消失訂正符号）を付加する（合計 12）。このうち「7 つ」にアクセスできればデータを読み込むことができ、「8 つ」にアクセスできればデータを書き込める。

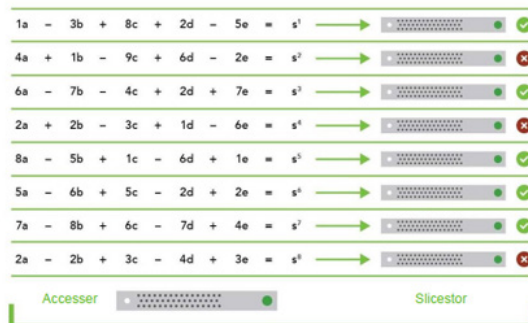
「このデータを 12 台に分散させると、仮に 4 台にアクセスでき



日本 IBM クラウド事業統括 Cleversafe 営業部 担当部長 コンサルティング セールス スペシャリスト 田中克哉氏

IDA (Information Dispersed Algorithm、情報伝播アルゴリズム) の概念 (出典: 日本 IBM)

情報分散の背後にある数学



k 個の変数を解決するために
知る必要があるのは k 個の結果です
そのため、(n-k) 個の結果を失っても構いません
ディスク上と送信中のデータ・オーバーヘッドは (nk) です

ない状況が発生したとしても問題なく読み込み/書き込みができます。この場合、データロスは約 12 億年に 1 回しか起こらない計算です」（田中氏）

3拠点レプリケーションよりも 少ないデータ量で冗長性を維持

さらに、この 12 のストレージを 3 拠点に分散させれば、従来の 3 拠点レプリケーションよりも圧倒的に少ないデータ量で十分な信頼性を保てるのだ。例えば、1 ペタバイトのデータを冗長化して保護する際でも、必要なストレージ量は 1.7 ペタバイト程度と、圧倒的なコスト削減になる。また、データの分割は暗号化してから行うため物理的な盗難によるリスクもない。

「オブジェクトストレージは後から暗号化する製品が大半で、先に暗号化しているものは他にない。IBM Cloud Object Storage は万が一、物理的にディスクを抜かれてしまうことがあってもデータを復元できず、安全と言い切れる」（田中氏）という。さらに、「IBM Cloud Object Storage では、API によるオブジェクト操作、オブジェクトのデータ管理も暗号化されているため、クラウド提供でも安全」（平山氏）だ。

リージョンを超えて自動的にデータを分散、 リージョン間通信は無料

IBM Cloud Object Storage ではリージョンを超えた 3 拠点に自動でデータを分散できる。アジア太平洋地域ならば、東京、香港、ソウルの 3 拠点に、ヨーロッパではフランクフルト、ロンドン、アムステルダム、といった具合だ。この仕組みを応用することで、グローバルシステムの構築も現実的になるという。

一般的なクラウドオブジェクトストレージサービスでは、1 つのバケット内のオブジェクトに対して、リージョンをまたぐデータの分散には対応していないため、リージョンを超えた冗長化を行うにはオブジェクトをバケット間でコピーしたりレプリケーションしたりする必要があり、この場合はデータ更新にタイムラグが発生する。全

く同じタイミングでは「東京では新しいデータ、ロンドンでは古いデータ」となる可能性が高く、バケットが違うため別のオブジェクトとして認識され、別のデータ ID (キー) として扱われる。これに対して IBM Cloud Object Storage では、リージョン間も同じバケット内の同じオブジェクトとして分散される仕組みのため、データ ID (キー) が変わらずに同期できる。リージョンが離れていても見掛け上は 1 つのストレージプールとして扱えるため、グローバル分散したアプリケーションからオブジェクトストレージに対しての論理的なアクセス先を変えずに分散データストアにアクセスできるのだ。

「クラウドでグローバル分散システムを構築する際、フロント部分のアプリケーションロジックをグローバル分散させるのは比較的容易ですが、難しいのはバックエンド部分のデータストアをいかに分散させるかです。IBM Cloud Object Storage ならば、グローバルで 1 つのストレージプールを提供でき、さらにリージョン間をつなぐ国際専用線は無料。その中でも性能を求めて専有環境でチューニングしたり、共有型でコストを抑えたりと、細かく組み合わせることもできます。巨大ファイルやエクサバイト級のデータを運用できる上に、ここまで幅広い提供モデルがあるオブジェクトストレージは他にはありません」(平山氏)

150ペタバイトのデータを3人で管理、サービス7年間無停止を実現している事例

ここで 1 つ事例を紹介しよう。米国の無料オンラインアルバムサービスで急成長を遂げている「Shutterfly」は、オンプレミス環境に IBM Cloud Object Storage を導入している。そのデータ量はペタバイトを超え、1 分間に 5 万枚もの写真がアップされ続けているが、この環境をわずか 3 人の管理者で運用している。

「これだけのデータ量で運用した場合、通常であれば毎日のよう

にどこかが故障し、日々ハードウェアメンテナンスが必要なところ。しかし、オブジェクトストレージであれば自動的にエラーノードを省いて再構成するリビルダーという優れた機能があります。このため、ハードウェアのメンテナンスは週に 1 回まとめて実施する、というような効率の良い運用体制が実現できるため、3 人でも十分に運用可能です」(平山氏)

ちなみに同サービスは IBM Cloud Object Storage の高い信頼性を前提に、驚くべきことに 7 年間無停止でサービスを提供し続けており、その間サービスを止めることなくデータセンターの引越しも終えたという。

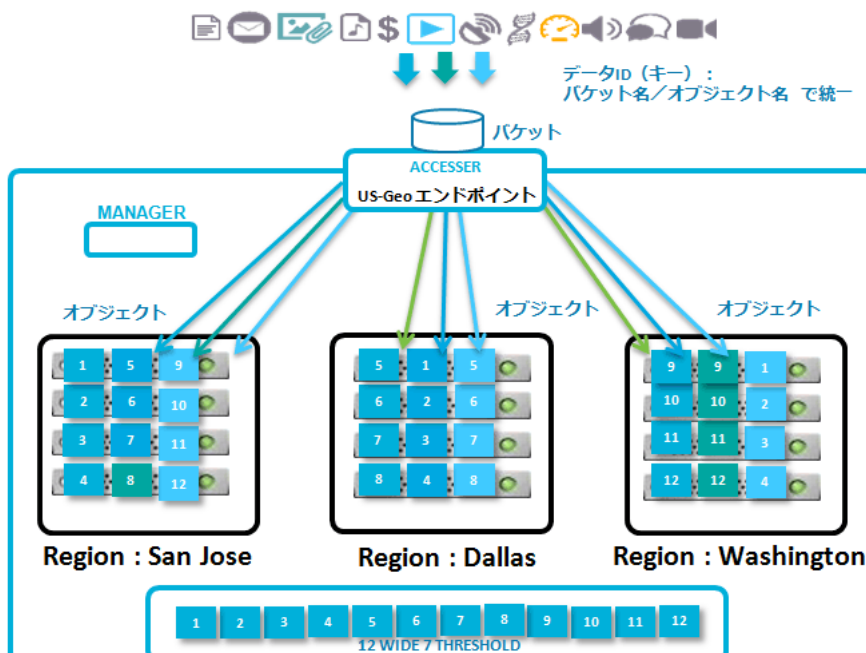
テープライブラリからアクティブアーカイブへ

このようにデータファイルの保存先として、あるいはバックアップなどの用途に用いられるオブジェクトストレージだが、これを「アクティブアーカイブ」と捉え、これまでテープ媒体などに保存していた膨大なデータをオブジェクトストレージに移行するケースも増えている。テープの保存・運用コストを考えたら、クラウドで月額費用が掛かったとしてもトータルコストは削減できる。「IBM Cloud Object Storage には、アーカイブ用途のストレージ『Amazon Glacier』の『Vault (ポルト:アーカイブ格納コンテナ)』に相当する機能も有しており、アーカイブに特化した安価なオブジェクトストレージとしても併用することができます」(平山氏)

コグニティブシステムで重要な役割を果たすオブジェクトストレージ

IBM はこのようなデータを眠らせておくのではなく、このようなダークデータを解析・分析に活用することも視野に入れている。つまり、IBM Cloud Object Storage は IBM の人工知能エンジン

IBM Cloud Object Storage のグローバル分散 (出典: 日本 IBM)



「Watson」との関係で可能性を広げていく。

「WatsonをはじめとするAIは学習の基となるデータがなければ何もできません。十分に学習させるには相応のデータが必要ですが、コグニティブシステムの世界では、画像、音声、動画などの非構造化データが重要な役割を担っています。これらのデータは、ファイルサイズが大きく、基本的にはWatsonに学習させるためにデータを投入するというシンプルな操作が中心となり、Watsonはクラウドを中心に提供されています。そのため、そのデータの置き場としてはオブジェクトストレージが最適なのです」（平山氏）

写真や動画に「誰が映っているか」を認識する、あるいは製品画像から不具合を見つける、といった仕組みは決して夢物語ではない。身近なところでは、ビジネスドキュメントの学習も考えられるだろう。オブジェクトストレージに保存されたMicrosoft ExcelやMicrosoft Wordのファイル、PDFファイルなどが格納されている

「Box Platform」や「Salesforce.com」などのデータをWatsonで解析する取り組みも既に進められているという。オブジェクトストレージを単なる「既存ストレージからの置き換え」「データをためるだけ」のサービスとして考えていると、その可能性を狭める結果になりかねない。先々のデータ活用まで見据えた戦略として考えたいところだ。

ちなみに、IBM Cloud Object Storageはキャンペーンを実施しており、USリージョン限定だが25GBまで無料で使えるという。これだけの容量があれば、自社の用途に適しているかを十分判断できるはず。まずは試してみるのもいいだろう。クラウドネイティブやビッグデータを活用したシステムが主流になる中で、重要な役割を果たすオブジェクトストレージにおいて、エンタープライズ向けの要件を充足しているIBM Cloud Object Storageは注目だ。

● IBM クラウドに関する詳細はこちら

- ・ IBM Cloud Object Storage
<https://www.ibm.com/cloud-computing/jp-ja/products/storage/object-storage/>
- ・ IBM Cloud Object Storage テクニカルドキュメント
<https://ibm-public-cos.github.io/crs-docs/>
- ・ IBM Cloud
<https://www.ibm.com/cloud-computing/jp-ja/>

日本アイ・ビー・エム株式会社

電話：0120-550-210

Eメール：cloudedm@jp.ibm.com