

# 企業が活用するオブジェクトストレージ入門

## 急増するデータの悩み、クラウドの IT 技術で低予算に解決

オフィスツールの高性能化とモバイル化、IoT や AI の活用によるビッグデータ対応、DR と BCP。企業活動におけるデータ量は増加する一方で、多くの企業を悩ませています。データ量が増えれば IT 予算も増えるという時代ではありません。従来の延長線にある仕組みとは異なる、新たな低コストの解決策が求められています。これを解決するひとつの方法が、クラウドの IT 技術を企業においても活用することです。

本ホワイトペーパーでは、クラウドのストレージ基盤に活用される「オブジェクトストレージ」の概要と、企業における使いどころ、そして国内外で豊富な商用実績をもつ SDS (Software Defined Storage:ソフトウェア定義ストレージ) のスケールアウト型オブジェクトストレージ製品「Cloudian HyperStore」についてご紹介します。

### 1 企業が抱える急増するデータの悩み、3つの要因

企業活動においてデータが増え続けることから生じる悩みは尽きることがありません。最近では、次の3つの要因がデータを急増させていると言えるでしょう。

#### (1) オフィスツールの高性能化とモバイル化

各種オフィスツールが高性能化し、私たちが日常見かけるオフィス資料には、いわゆる「ファイルサイズが重い」写真や画像が当たり前に使われるようになりました。また、タブレットやスマートフォンの普及により、移動時間や出張先での作業も可能になり、作成資料の数や、そのファイルをメールでやり取りする機会が増えているのではないのでしょうか。これを裏付けるように、多くの専門アナリストが年率 50%以上でデータ量が増加していると述べています。

#### (2) IoT / ビッグデータ / AI の活用

Hadoop や Spark といった OSS (オープンソースソフトウェア) のビッグデータ処理ツールを使ったビッグデータ分析が身近になっています。ビッグデータは主にインターネットという仮想の世界で収集され活用されてきました。しかし、モニターカメラやセンサーが大量のデータを集める IoT (モノのインターネット) の本格的な普及が始まり、あらゆる業種で現実の世界の出来事をビッグデータ化して活用し始めています。そして、AI (人工知能) を使い、

そのビッグデータを分類・整理し、分析する動きも始まっています。今後ライフサイクルに沿った経年変化の分析なども行えば、たとえば、今後 10 年間分のデータをオンラインで活用できる必要性が生じます。

#### (3) DR (災害復旧対策) に対する意識の高まり

東日本大震災以降、複数箇所にデータを保管する重要性が強く認識されました。一箇所だけでは事業継続性の観点からリスクが高いという経験を経て、複数地域のデータセンターにデータを複製しておくことが求められています。これを単純に考えれば保管するデータ量が倍になるということです。

これらの要因により、企業の IT 部門はひっ迫するストレージ容量、増え続けるファイル数、長期化するデータ保管期間への対応、データ量増加に伴うシステムの性能劣化等に対処しなければなりません。これは、ストレージ装置の増設・置換、データ移行作業や、バックアップ、アーカイブといったデータ管理コストの増大につながります。データ量が増えれば IT 予算も増えるという時代ではありません。従来の延長線にある仕組みとは異なる、新たな低コストの解決策が求められています。

### 2 急増しているのは「非構造化データ」

これらの課題解決にあたり着目すべきは、いま急増しているデータは、「非構造化データ」であるという点です。

この「非構造化データ」とは、一般的なリレーショナルデータベースで扱うデータとは異なり、あらかじめデータの構造が定められていないデータです。それは、オフィス文書、写真、動画、画像、音楽、

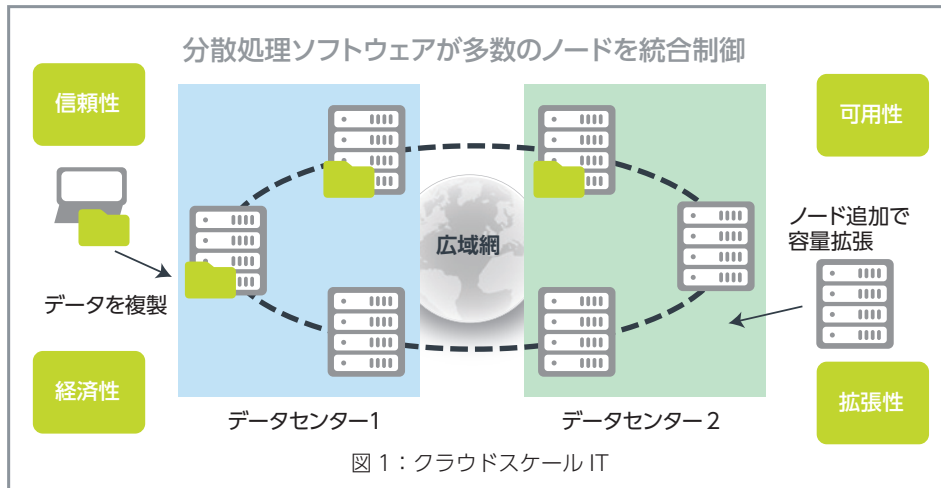
ログなど多種多様です。これらの多くが書き換えや更新の頻度が少ない、若しくは更新されることが無いデータです。写真が良い例でしょう。デジタルカメラやカメラ内蔵スマートフォンで多くの人が膨大な数の写真を撮りますが、一旦撮影し保存した写真が更新されることはありません。新しい写真が次々と追加されるだけです。

## 3 大量の「非構造化データ」を扱うのがクラウド

この「非構造化データ」を大量に扱っているのがクラウドです。たとえば、インターネットサービスの SNS は、何十万人、何百万人という利用者が、文字情報を交換し、写真を共有し、動画や音楽をアップロードしています。利用者がストレスを感じないように、速く処理することも求められます。格安であったり無償の広告モデルで提供されているサービスも多く、サービス運営は低コストであ

ることが絶対条件となっています。

Google や Amazon といった本物のクラウドでは、多種多様で膨大なデータを迅速に低コストで処理する IT 技術が活用されています。これを、本ホワイトペーパーでは「クラウドスケール IT」と呼ぶこととします。



## 4 「クラウドスケール IT」で活用されるストレージ基盤がオブジェクトストレージ

この「クラウドスケール IT」のストレージ基盤として活用されているのがオブジェクトストレージです。

呼ばれる大規模なインターネットサービスでもオブジェクトストレージが活用されています。

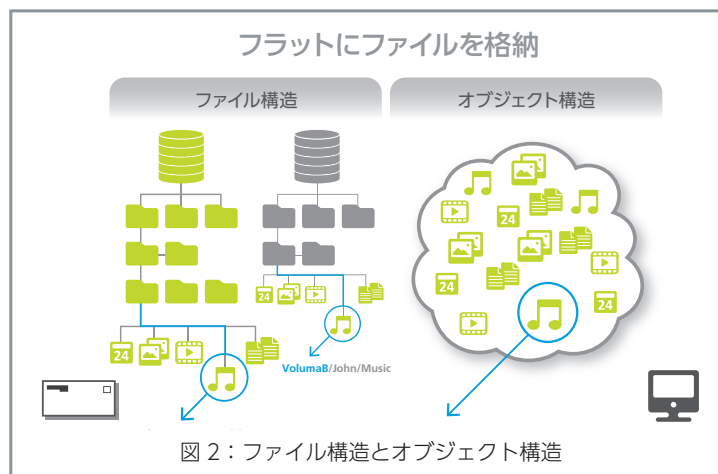
オブジェクトストレージは、Amazon では「Amazon S3」、ニフティでは「ニフティクラウドストレージ」、NTT コミュニケーションズでは「Biz ホスティング クラウド・エヌ オブジェクトストレージ」といったクラウドストレージサービスとして提供されています。また、写真等のオンラインストレージサービスを提供する NTT 東日本の「フレッツ・あずけ〜る」のような、最近では「Web スケール」とも

このオブジェクトストレージでは、オフィス文章、写真、動画等のファイルを「オブジェクト」と呼びます。そして、これらのオブジェクト間の関係がフラットであり、階層構造がないという特長があります。この構造ゆえに、クラウドで求められる可用性や拡張性といった要件が実現できます。

## 5 クラウド環境における階層構造の課題とは

私たちが日常使うファイルシステムは、図 2 の左側に示すようにディレクトリとフォルダーといった階層があります。この構造は、ある

意味、人間の直感に合っています。つまり、地球があり、日本があり、東京があり、渋谷があるという流れで目的地に到達できるた



め、たいへんにわかりやすいと言えるでしょう。

しかし、この構造の場合、ファイル数が増え、それを格納する物理的なハードウェアの容量が一杯になると困ったことが起こります。たとえば、別のハードウェアにファイルを移し替えるとなると、ファイルのありかを示す情報が完全に変わってしまいます。東京から埼玉に引っ越せば住所が変わってしまうことと同じです。そのたびに、住所変更のお知らせをしなければなりません。

## 6 フラットな構造のオブジェクトストレージ

オブジェクトストレージは、図 2 の右側に示すように階層構造がありません。オブジェクト毎に ID (識別子) が与えられますが、この ID には「東京」といった場所情報が含まれていません。つまり、データがどのハードウェアに格納されていても、どこに引っ越しをしても ID は同じということです。

ファイル構造を住所情報が必要な手紙にたとえるならば、オブジェクト構造は電子メールのようなものです。たとえば、東京から埼玉に引っ越しすると、住所情報を含む手紙の場合には宛先不明になるかもしれませんが、電子メールであれば住所変更しても、なんら影響なくメールが届きます。

## 7 属性情報 (メタデータ) の役割

オブジェクトストレージのもうひとつの大きな特長に、オブジェクト毎にデータを効率的に管理するための属性情報 (メタデータ) をカスタマイズして付けることができるという点があります。

オブジェクトに属性情報を付けることで、一定期間経過後、オブジェ

クラウドのような環境では、何万人という利用者が使い、急激にデータ量が増えるといったことが当たり前におこります。その一方で、データをストレージする物理的なハードウェアの容量には制限があります。容量が一杯になると、利用者に住所変更のお知らせをすることは運用者にとって大きな手間がかかります。利用者にとっても、ある日、ファイルの格納場所がわからなくなり、アクセスできないといったことも起きてしまうかもしれません。

このオブジェクト ID は、一般的にインターネットで使われる URL が付与されることが多く、相手にこの URL を知らせるだけでオブジェクトを読み出すこともでき、インターネットで便利に使うことができます。

また、オブジェクト単位で複製し信頼性を高めたり、物理的なサーバーを追加するだけでシステム全体の容量を拡張したり、複数データセンター間でオブジェクトをやり取りしたりということが簡単にできるようになります。一方、クラウドのような多数のサーバーが並ぶ環境のなかで、階層構造を維持することの難しさは直感的に理解していただけたと思います。

クトを消去したり、アクセスできる利用者を制限するといったことができます。Cloudian HyperStore の場合、格納したオブジェクトを Amazon S3 や超低価格のアーカイブストレージ Glacier に自動転送するといった機能にも活用しています。

## 8 オブジェクトストレージでデータ急増の悩みを解決

現在、一般的に企業 IT で活用されているストレージは、SAN と呼ばれるブロックストレージ、NAS と呼ばれるファイルストレージです。これらのストレージは、処理性能が高く、迅速な更新処理が求められるデータベースのデータ等を格納するために活用されています。ただし、これらのストレージは企業構内において近距離の高速なネットワークで接続されることを前提としています。インターネット経由や、多地点の複数データセンター間でのデータのやり取りに適しているとは言えません。オブジェクトストレージは、その逆です。性能はこれらのストレージに比べ落ちますが、更新頻度の少ない「非構造化データ」をクラウドのように広域網を経由する環境でやり取りすることに適しています。

さて、いま急増しているのは「非構造化データ」であると説明しました。この「非構造化データ」のうち、更新の必要のないデータはコールドデータ、頻度の少ないデータはクールデータとも呼ばれ、データ全体の 70% を占めるとも言われています。

考えてみれば、これらの特徴をもつ「非構造化データ」をストレージするためには、高性能であることは必要ありません。むしろ、多種多様で膨大なデータを経済的に処理し、拡張性が高いことが求められます。つまり、「クラウドスケール IT」のオブジェクトストレージが適しているのです。

これまで、オブジェクトストレージは大量データを対象とし、コンテンツを膨大に抱えるメディア業界や政府、研究等の専門的な領域で利用され、一般企業にとって身近な製品とは言えませんでした。そのため、SAN や NAS といった従来型のストレージ装置しか選択肢がなく、コールドデータやクールデータのストレージにも高性能の装置を使い続けてきたというのが現実です。しかし、いまや、これから紹介する Cloudian HyperStore のように手軽に利用できるオブジェクトストレージ製品が登場してきました。低予算で企業が抱える急増データの課題を解決できるのです。

## 9 オブジェクトストレージの使いどころ

### (1) オンライン活用するデータ

ここまで説明してきたように、更新頻度が少ない、若しくは更新することが無いものの、オンラインで活用するデータは、オブジェクトストレージの使いどころです。典型的な例がバックアップデータです。企業 IT システムのデータには必ずバックアップを用意しますが、必要な際にはオンラインですぐに読み出すことが求められます。しかし、その一方で、万が一のためのバックアップにコストをかけたくないというのが企業の本音です。

また、企業にはオフィス資料、イベントやプロジェクトの記録写真、機械による画像記録、建築物・機械・装置等の設計図面、各種実験等の動画記録等々、オンライン活用が求められる膨大な数のファイルがあります。こういったデータの保管にオブジェクトストレージを活用することで IT システム全体のコスト低減につながります。

特に最近では、モバイル環境やブロードバンド環境が整ったことにより、外出先、自宅、遠隔オフィス等、広域から社内ファイルシステムにオンラインでアクセスする機会が増えています。従来のファイルシステムは広域、分散環境への対応は得意ではありません。このようなファイル共有・同期には、オブジェクトストレージが適しています。

### (2) 安全に大量に保管するデータ

コンプライアンスの要請や証拠書類として保存が義務付けられているデータ量は年々累積し増加してゆきます。これらのデータはテープに移し倉庫で長期間保管されることが一般的です。しかし、テープ保管は、いざ必要な時には保管倉庫から運搬するといった人手もかかり、すぐに読み出すことができません。テープ保管に合わせてオブジェクトストレージを活用することで、これらのデータ

をオンラインで簡単に活用できるようになります。

特に Cloudian HyperStore であれば、複数データセンターにデータ複製を行い、自動的に複数地域に分散し保管できるため、遠隔の倉庫にテープ運搬するような作業も不要です。データを安全に大量に保管しながらも必要な時にすぐ使えるアクティブなアーカイブにオブジェクトストレージは最適です。

### (3) ビッグデータ分析のための大量データ

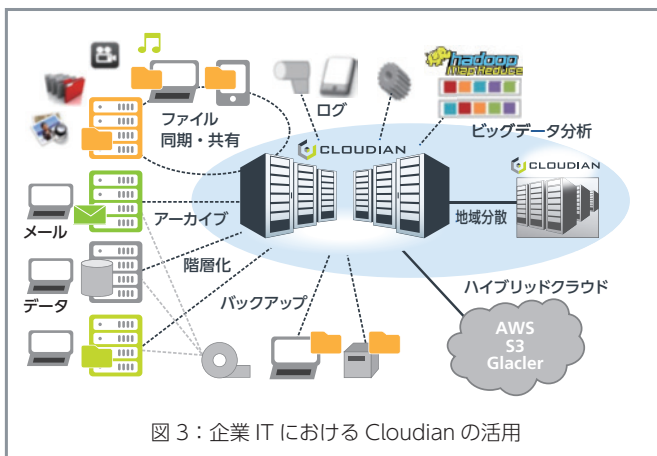
ビッグデータ分析は、その名のとおりに膨大な量のデータを分析することです。Hadoop のようなビッグデータ処理ツールが身近になったことに加え、センサーの小型化やモバイルネットワークの普及に伴い、これまで収集することが難しかった人間・装置・車等機械の移動、ビルや橋梁といった構造物の変化などのログデータまでもが分析の対象となりはじめています。

この膨大な量のデータ解析には、一時的に多数のコンピューティングリソースが必要となるため、パブリッククラウドを活用する事例も増えています。その一方で、データを外部に預けられないという事情がある企業も多く、その場合にはデータは、自社管理できる企業構内やデータセンターに保存しておくという対応も増えてきそうです。

これらのデータは 1 年で捨てられるわけではなく、経年変化をみるために貯め続けられ、将来に巨りどれだけ増えてゆくのか予測するのも難しいとも言えます。こういった大量データの保存には、拡張性高く、経済的なオブジェクトストレージが活用されることになるはずで

## 10 企業におけるオブジェクトストレージの活用パターン

一般の企業 IT 部門が企業構内（オンプレミス）でオブジェクトストレージを活用するパターンを紹介したのが図 3 です。以下、この図 3 を説明してゆきます。



**ログのストレージ**：監視カメラの映像、行動履歴、通信履歴、操作履歴、アクセス履歴等のログは膨大なデータ量になってゆきます。また、ログは書き換えたり更新されたりするものではなく、大量に貯めておく一方で、必要な際にはオンラインですぐに活用することが求められます。こういったデータを高価で高性能なストレージ装置に保存しておく必要はありません。まさに高い拡張性と経済性を実現するオブジェクトストレージが適しています。

**ビッグデータ分析データ**：ビッグデータ分析に利用されるオープンソースの Hadoop MapReduce (マップリデュース) は S3 インターフェースを備えています。そのため、Hadoop 専用の分散ファイルシステムである HDFS (Hadoop Distributed File System) ではなく、オブジェクトストレージにデータを直接保存してしまうという方法があります。こうすることで、Hadoop MapReduce 以外の目的にデータを使う場合にも、HDFS から他のストレージに移



し替える手間や、別途ストレージを準備する必要もなくなります。つまり、ビッグデータ分析のためだけの特別なストレージ領域が不要となり、企業の IT コスト効率を高めることができます。

**ファイル同期・共有**：ファイル同期・共有アプリケーションのファイルを格納するストレージにオブジェクトストレージは適しています。ファイル量が増加し、ストレージ装置の物理的な容量が一杯になり、高性能・大容量の装置に取り換えたり、そのストレージ装置の障害に備えたバックアップ用システムの心配をする必要ありません。

企業構内で利用されることを前提としている SAN や NAS といったストレージ装置とは異なり、オブジェクトストレージは、遅延が生じるモバイルやインターネットといった広域網経由でファイルにアクセスすることに適しています。オブジェクト ID が URL であれば、インターネット経由でファイルをやり取りすることも簡単にすることができます。

**バックアップ**：バックアップとは、ファイルサーバーに障害等があっても重要なファイルが消失しないよう複製データを別の装置やテープに保管しておくことです。このバックアップは、企業で導入検討される際、収入増につながるものではなく、万が一のためであり、そこにかかるコストは最小限にしたいというのが本音でしょう。実際のところ、高価なストレージ装置をバックアップデータのストレージに使用するケースが多く、ミラーリングのオプションソフト購入が必要なケースもあります。

一般的に企業では長期間のバックアップはテープに保存しています。ただし、そのテープを遠隔地の倉庫等に保管してしまえば、読み出しに手間と時間がかかってしまい、簡単には活用できない、いわゆる死蔵データとなってしまいます。そのため、テープ保管しながらも、オンラインで利用する可能性のあるバックアップデータはオブジェクトストレージを使い、オンラインで活用できるようにしておくことが望ましいと言えます。

**アーカイブ**：アーカイブとは、データを整理し、かつ改ざんできないように長期間安全に保存しておくことです。利用例では Eメールのアーカイブが代表的です。多くの企業では各社員あたりのメールボックスサイズに制限を設け、その制限サイズになると、社員が自らローカルの PC や外付けハードディスク等にダウンロードし保管しているという話を聞きます。メールボックスはデータ量が増えようと性能が劣化します。その一方でメールボックス拡張にはコ

ストがかかるため、社員にサイズ制限を強いることにより対応しているのが現実です。

今後、メール数が減るとは想定しにくく、社員の手元で保管される過去のメール量は累積し増える一方でしょう。社員にこのような手間をかけさせたり、その手元に大量の過去データがあるという状況は、企業 IT 部門としても決して好ましい状況とは考えていないはずで、S3 対応のアーカイブ用アプリケーションを使い、適切なタイミングで自動的にオブジェクトストレージにアーカイブしておけば、社員や企業 IT 部門の手間を軽減できるはずで

**階層化**：ストレージにおける階層化とは、データの種類や利用頻度に応じて複数のストレージ装置に階層を設けることです。つまり、一定期間が経過したり、利用されることが少ないデータは、高価で高性能ストレージから比較的安価なストレージ装置に移行するという具合に、ストレージ装置を使い分けるということです。

たとえば、大量の「非構造化データ」はオブジェクトストレージに格納することで、ストレージシステム全体としてのコスト最適化をはかることができます。

**地域分散**：東日本大震災以降、地域・広域分散した複数データセンターにデータを複製し安全に保管する必要性が強く認識されています。従来型のストレージ装置では、複数データセンターに対応する場合、同じ装置やシステムを 2 重に用意し、非同期、手動でデータ複製を行うことが一般的です。そのため、あるタイミングでは、複数データセンターの一方には、最新データが保存されているものの、もう一方では、そのデータが保存されていないというケースも起こります。複数データセンター間で、同期・非同期を選択しデータ複製を行うことができるオブジェクトストレージ製品であれば、ネットワーク経由で広域に複製データを分散配置しておけます。

**ハイブリッドクラウド**：パブリッククラウドサービスは、経済的で拡張性が高く、Web サービスや IoT で収集する大量のデータを保存しておくことに適しています。しかし、企業のデータを保存するとすると、内規で外部にデータを預けることができない、または読み出しに時間がかかるため速い処理を求めるデータは保存できない、データ量が多く利用料が高額になってしまうなどの課題もあります。

そのため、パブリッククラウドと同等の環境を企業 IT の管理下にあるデータセンターに構築すれば、パブリックとプライベートをデータの目的や種類に応じて使い分けることができます。

## 11 オブジェクトストレージの代表的なソフトウェア製品「Cloudian HyperStore」

Cloudian HyperStore は国内外の商用クラウドストレージサービスのストレージ基盤に活用されている、SDS (Software Defined Storage: ソフトウェア定義ストレージ) のスケールアウト型オブジェクトストレージ製品です。

汎用的な IA サーバーをハードウェアとし、OS には Linux を使います。汎用的なサーバーの価格は、年々高性能化し高集積化しながら、その値段を下げており、1 台あたり数十万円でも数十テラといった内蔵ディスクを搭載しています。このようなハードウェアを使うことで低コストのオブジェクトストレージシステムを構築することができます。

### (1) Amazon S3 の API に完全準拠

この Cloudian HyperStore の重要な特長は、クラウドストレージの事実上の標準となっている、Amazon S3 の API に完全準拠しているということです。

他製品でも S3 API を提供しているという製品はありますが、データの読み書きといった基本的な API を提供しているだけです。Cloudian HyperStore は、バージョンング、マルチパートアップロードといった高度な操作ができる API までも実装しており、質量ともに準拠しているという点が大きく異なります。

特に質という点では、数多くの Amazon S3 を利用しているアプリケーションや製品との相互接続検証がされ、そのいずれもが

Cloudian HyperStore にデータの保存先を向けるだけで簡単に接続できています。

このような S3 対応のエンタープライズ向け製品は数多くあり、すでに企業で広く利用されています。つまり、現在企業が利用しているアプリケーションを取り換えたり、改修することなく、簡単に Cloudian HyperStore にデータを格納することができるのです。

### (2) クラウドサービスからエンタープライズ IT まで、広い活用範囲

さらに重要な Cloudian HyperStore の特長は、非常に活用範囲が広いという点です。他製品には 1PB (ペタバイト) 規模のデータ量からの利用を前提としている製品もありますが、Cloudian HyperStore は汎用サーバー 2 台規模から利用開始することができます。このように企業 IT のデータ規模からスモールスタートできる一方で、すでに多くの国内外商用クラウドストレージサービスに採用されているように、サーバー台数を追加するだけで大規模なクラウドサービスにまで拡張することができます。

統計、課金、利用者、利用量等管理機能もパッケージされており、これらの管理機能を別途開発したり調達したりという必要無しにターンキーで利用開始できるという点も強みです。サービスプロバイダーが短期間でクラウドストレージサービスを開始できることはもちろんですが、企業利用においても部門予算管理等のために、別途アプリケーションを開発する必要もありません。

## 12 まとめ

Google や Amazon といった世界を代表するクラウドは、いま企業 IT が直面しはじめた膨大な「非構造化データ」に早い時期から対処してきました。この本物のクラウドが活用する IT 技術、すなわち「クラウドスケール IT」のストレージ基盤がオブジェクトストレージです。なかでも、Cloudian HyperStore は、アクセスキー

とシークレットキーを入力しエンドポイントを向けるだけで接続できる、Amazon S3 API 完全準拠のオブジェクトストレージ製品です。急増するデータの課題を解決するために、大規模クラウドからエンタープライズ IT まで幅広いシーンで活用いただけます。

### クラウドファンについて

日本と米国を開発拠点とするクラウドファンは、パブリッククラウド、プライベートクラウド、オンプレミス環境でハイブリッドに活用できる SDS (Software Defined Storage: ソフトウェア定義ストレージ) である「Cloudian HyperStore」をソフトウェア製品及びアプライアンス製品により提供しています。国内外大手プロバイダー、エンタープライズが採用する Cloudian HyperStore は、複数データセンター間を含み、データ複製・分散配置によるデータ保護をしながら、汎用サーバ 2 台からペタバイト超級にまで経済的に、柔軟にスケールアウトします。統計・課金・管理機能も実装済みであり短期間に利用開始できます。



クラウドファン株式会社

[www.cloudian.jp](http://www.cloudian.jp) | [info@cloudian.com](mailto:info@cloudian.com)