

ビッグデータ時代のクラウドを支える ストレージソリューション

Storage Solution for Cloud Computing

島田 朗伸

Shimada Akinobu

山本 康友

Yamamoto Yasutomo

阿部 哲也

Abe Tetsuya

喜多村 将之

Kitamura Masayuki

ビッグデータ時代が到来し、仮想化技術を駆使して多様かつ大量のデータを蓄積、検索、瞬時に分析し、新たな価値創造を支援するIT環境が求められている。日立グループはユニファイドストレージコンセプトの下、仮想化技術によってITインフラを統合し、データ形式を問わずにまとめてストレージシステムに格納することでビッグデータの有効活用を支援するユニファイドストレージHitachi Unified Storage 100シリーズを2012年4月にリリースした。

さらに、散在・増加し続けるファイルデータの効率的な蓄積・管理を支援するファイルストレージの提供や、データ量の肥大、処理時間の増大といったバックアップ運用の課題を解決する、重複排除機能を備えたバックアップ用途向けファイルストレージの製品化などにより、データセンター運用の効率化を支援していく。

1. はじめに

近年、クラウドコンピューティングや多機能情報端末の急速な普及、ストレージやサーバなどのIT (Information Technology) プラットフォームの高性能化を背景に、流通、医療、金融、行政、さらにはエネルギーや水処理、交通分野をはじめとする社会インフラなど、さまざまな分野で発生するビッグデータの収集、蓄積、検索、処理、分析による新たな価値創造への機運が高まっている。

企業内のデータもますます大容量化、かつ多様化しており、それらを一元的、効率的に管理し活用することが、企業の成長戦略において重要な課題になってきた。

こうした背景を受け、企業や社会などに存在するさまざまなデータをその種類によらず、まとめて格納できるITシステムが期待され、また機種、設置場所などに依存せず一元的に運用し管理することが求められている。

ここでは、きたるべきビッグデータ時代へ向けた新たな日立ストレージソリューションについて述べる。

2. ビッグデータ活用を支えるクラウド

数年前までは、各企業や組織が保有しているデータ量はペタバイト (PB) 規模だったが、情報量は爆発的に増大し、ビッグデータと言われる巨大なデータの集まりにより、今ではその1,000倍のエクサバイト (EB) も珍しくなくなっている。また、今後さらに1,000倍のゼタバイト (ZB) という容量になる時代を迎えると言われている。

このビッグデータの利活用により、時々刻々と変化するビジネスの状況や社会環境にタイムリーに対応し、企業の経営判断やさまざまなオペレーションをダイナミックに修正/変更したり、今まで気がつかなかった切り口でビジネスモデルを見直すといったことが可能となる。このような背景から、ビッグデータを活用する情報システムとしてのクラウド環境が、ますます注目されている。

2.1 三つのクラウド戦略

日立グループは、ビッグデータ利活用に向けたクラウドIT環境の対応として、(1) 収集・蓄積、(2) 検索・可視化、(3) 分析・予測の三つのステップがあると考える(図1参照)。

それぞれのステップにおいて、データの扱い方や処理のしかたが従来と大きく変わり、各ステップの要件を満たすクラウド構築が必要になる。

大量のデータを効率的に保管できる「インフラクラウド」(収集・蓄積)と、さまざまなファイルやコンテンツを一元管理して簡単に参照できるようにする「コンテンツクラウド」(検索・可視化)、そしてそれらのファイル、コンテンツをさまざまな切り口で分析したり、別のアプリケーションから簡単に活用できるようにする「インフォメーションクラウド」(分析・予測)の三つのクラウドを実現し、ビッグデータ利活用の支援のためのプラットフォーム、ソリューション、サービスの提供をめざしている。

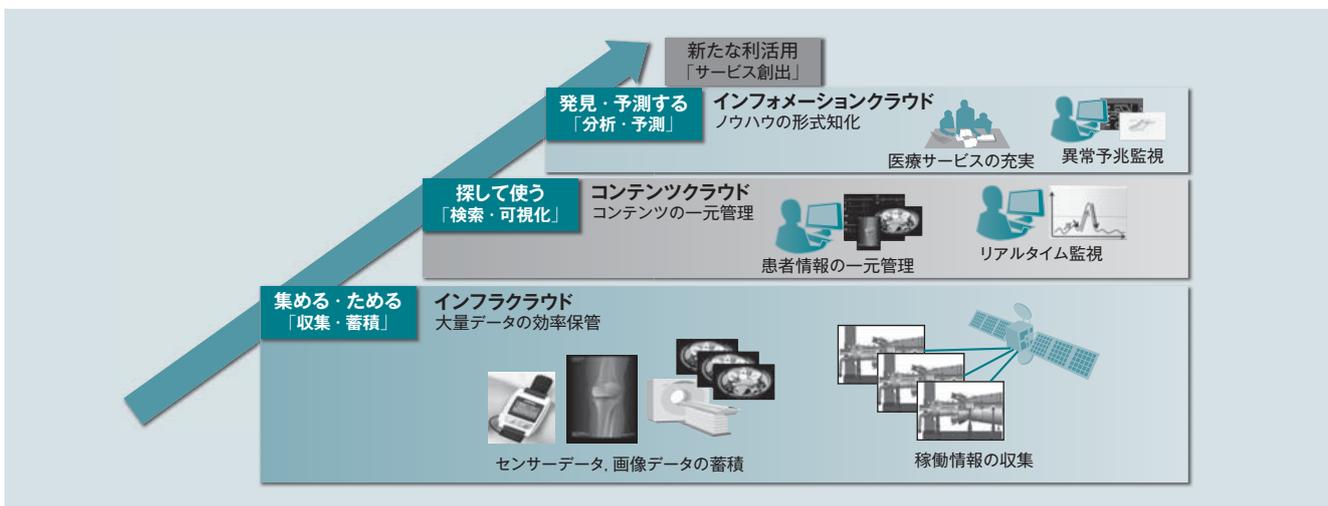


図1 | 日立の三つのクラウド戦略

ビッグデータの利活用をめざし、三つのステップ(収集・蓄積、検索・可視化、分析・予測)で支援する。データの扱い方や処理のしかたが異なり各ステップの要件を満たすクラウド構築が必要となる。

2.2 「One Platform for All Data」

日立グループは、ITプラットフォーム事業ビジョン「One Platform for All Data」を基に、高効率のクラウドIT環境を包括的に支援する基盤の実現をめざしている。特に、拡大するコンテンツクラウドのさまざまなデータへの対応とそれを支えるインフラクラウドとの連携が重要である。このため、ビッグデータの利活用を支えるクラウドの基盤となるストレージプラットフォームの強化を図った。

3. 日立ユニファイドストレージ

3.1 日立ユニファイドストレージコンセプト

クラウドでは、データが高信頼に保存されることが、従来以上に重要となる。加えてコンテンツクラウドシステム、インフラクラウドシステムが要求するストレージ仮想化レイヤ、およびストレージプラットフォームへの要件には以下のようなものがある。

- (1) 爆発的に増加し続ける多様なデータを、効率よく蓄積・管理・活用できること
- (2) 蓄積されている大量または大容量のデータに、必要なときに迅速にアクセスできること
- (3) 大量に蓄積されているデータの保管コスト(スペース、電力、容量)を最適化できること

このような要件を満たすストレージインフラ実現に向け、2012年4月に日立ユニファイドストレージコンセプトを定めた(図2参照)。このコンセプトは、特長の異なる複数種類のストレージ(ユニファイドストレージ、ファイルストレージ、バックアップ向けストレージなど)を一元管理することにより、あらゆる種類のデータ(データベース、メール、映像など)を蓄積するストレージインフラの管理運用の効率化・最適化をめざすものである。これ

により、先に挙げた三つの要件を満たすストレージシステムを実現する。次に、日立ユニファイドストレージコンセプトの中核となるユニファイドストレージについて述べる。

3.2 ユニファイドストレージ

Hitachi Unified Storage 100シリーズは、Hitachi Adaptable Modular Storage 2000シリーズの後継製品として市場に投入されたが、ブロックデータ(以下、ブロックと記す。)やファイルデータ(以下、ファイルと記す。)といったさまざまな種類のデータを、ひとつのシステムにまとめて格納することで、ビッグデータの迅速な利活用を支援する。

ブロック向けの高信頼、高性能なストレージとして利用ができるほか、ファイルモジュールを追加することにより、ブロック、ファイルなどさまざまな種類のデータを1システムにまとめて格納できる。ブロック/ファイル双方のアクセスプロトコルに対応し、データ種類の異なるアプ

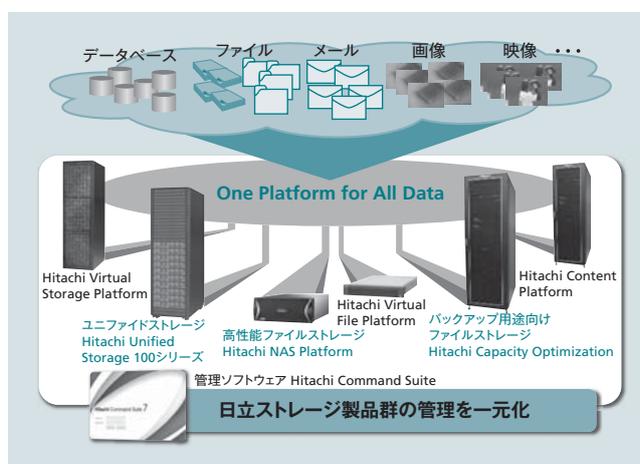


図2 | 日立ユニファイドストレージコンセプト

管理ソフトウェアHitachi Command Suiteがブロックデータ、ファイルデータを問わずストレージ製品群の統一的な管理を実現する。

リケーションからのアクセスを1システムに集約して管理できるため、アプリケーションごとにストレージ装置を用意し煩雑な運用をする必要がなく、設備投資や運用管理コストも低くなる（図3参照）。

また、Hitachi Unified Storage 100シリーズでは標準搭載するHitachi Command Suiteのストレージハードウェア管理ソフトウェアHitachi Device Managerにより、ブロック/ファイル双方のストレージ容量や利用状況の確認、割り当てを一つの画面から一元的に行うことができ、管理者の日々の運用負担を軽減する。また、ボリューム容量仮想化機能Hitachi Dynamic Provisioningにより、ブロック/ファイル双方を共通のストレージプールで格納し、ストレージ容量の効率的な利用が可能となった。

従来モデル比2倍^{※1)}となる最大2.8 PBまでストレージ容量を拡張でき、大容量データの格納ができるほか、データ処理性能も従来モデル比最大3倍^{※1)}に強化し、増え続けるビッグデータの迅速な利活用を可能としている。

また、ストレージ階層仮想化機能Hitachi Dynamic Tiering^{※2)}を適用し、ストレージ階層を隠蔽しながら効率的な利用を実現する。業務アプリケーションから読み書きされるデータをアクセス頻度に応じ、自動的にSSD (Solid State Drive) やSAS [Serial Attached SCSI (Small Computer System Interface)] HDD (Hard Disk Drive)、ニアラインSAS HDDなどの適切なストレージ階層に割り当てる。

エンタープライズディスクアレイHitachi Virtual Storage Platformで用いられている技術を、ミッドレンジストレージにも適用し、必要な性能を維持しながらTCO (Total Cost of Ownership) の削減に貢献する。

3.3 ファイルデータの効率的利活用の基盤を提供する ファイルストレージ

メールや書類などといった非構造化データを中心に、企業内で利用されるファイルデータが、多数の拠点や部門などに散在し、日々増加している。この散在・増加し続けるファイルデータの、効率的な蓄積・管理に対する課題を解決するものとして、仮想ファイルプラットフォームHitachi Virtual File Platformを提供している。拠点・部門に設置された本製品がクラウド環境の入り口 (Cloud on-Ramp) として機能し、拠点データのデータセンターへの自動集約・一元管理を実現する。さらにファイル仮想化機能により、拠点側のユーザーはデータが拠点側・データセンター側のどちらにあるかを意識せずに利用することができる。これにより、拠点ごとの運用管理や、データ増大に伴う煩雑なストレージ装置・容量増設/データ管理といった従来型NAS(Network Attached Storage)の課題を解消し、TCO削減やROI (Return on Investment: 投資対効果) 向上などを実現する。

また、日々増加するファイルデータを迅速に利活用する、高性能なファイルサービスへのニーズに対応するものとして、ハイエンドクラスのファイルストレージHitachi NAS Platformを提供している。本製品は、Hitachi Virtual File Platformの上位に位置づけられるファイルストレージ製品で、通常、ソフトウェアで実現する通信プロトコルの制御やファイルシステム機能をハードウェアとして実装することで、科学技術計算、医療・ライフサイエンスやエンターテインメント分野などで求められる大容量データの高速な処理を実現する。

両ファイルストレージ製品とも、日立ユニファイドストレージコンセプトに基づいており、Hitachi Unified Storage 100シリーズ、およびストレージ装置の一元的管理を実現するストレージ管理ソフトウェアHitachi Command Suite

※1) Hitachi Unified Storage 150とHitachi Adaptable Modular Storage 2500との比較。
※2) 2012年末までに提供予定。

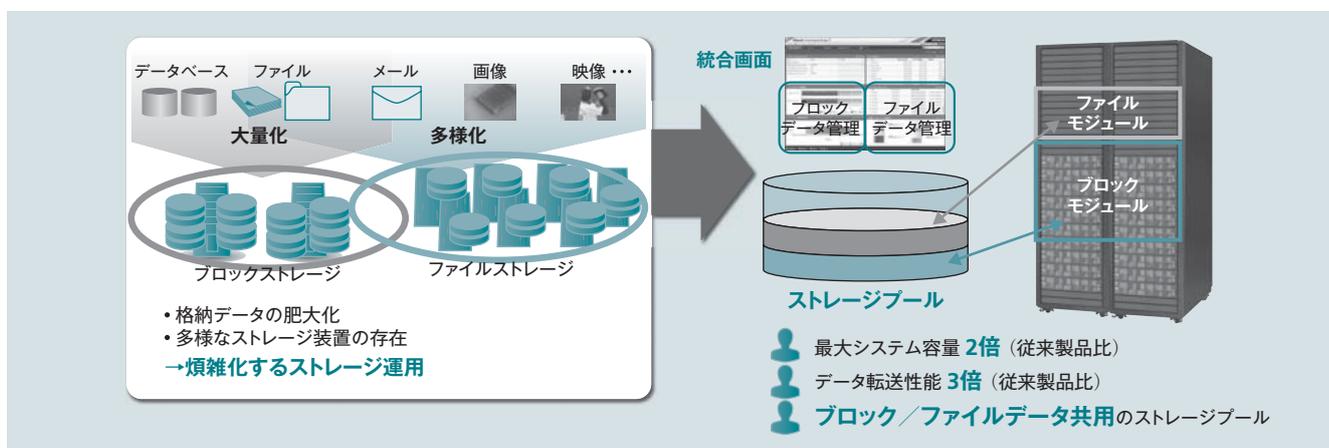


図3 | Hitachi Unified Storage 100シリーズ
ファイルデータ、ブロックデータを集約し、管理ソフトウェアによって共用のストレージプールで効率的な管理が可能である。

と組み合わせたストレージソリューションとして、ストレージシステムの運用・管理を容易化し、増え続けるコンテンツ管理のより効率的な運用を支援する。

3.4 企業データのバックアップ効率化を支援するバックアップ用途向けストレージ

企業やデータセンターでは、格納に必要なストレージ容量を抑えて、より経済的に、バックアップを行うことが一層重要な課題となっている。このバックアップ用ストレージ容量削減の課題に対応するため、バックアップ用途向けファイルストレージHitachi Capacity Optimizationを新たに開発・製品化した。本製品は、バックアップサーバによって格納されたファイルデータどうしを比較し、同一内容のデータ部分を検出・削除する重複排除機能により、バックアップデータ量の大幅な削減を可能とする。これにより、多世代のバックアップデータに対して特に大きな効果を発揮する(図4参照)。

また、ファイルデータを細かく分割して重複部分を検出し、一部分が異なるデータどうしであっても共通している要素を見つけ出して重複排除する方式を採用した。さらに日立独自の重複排除技術の採用により、バックアップの対象となるデータの形式や内容に合わせて、複数の重複排除方法から最適な方法を自動で選択・実行することで高い重複排除率を実現し、バックアップデータ量を約 $\frac{1}{20}$ ^{※3)}に削減することを可能とした。

Hitachi Capacity Optimizationは、バックアップデータを格納するストレージ部としてHitachi Unified Storage

100シリーズを活用しており、最大物理容量384 TB、最大論理容量7.6 PB^{※4)}と大量データのバックアップニーズへの対応が可能である。

さらに、LAN (Local Area Network) によってバックアップサーバに接続できるため特別な環境を必要としない、重複排除処理を開始する時間を任意に指定可能とし、バックアップと重複排除の処理が重ならないようにスケジュールすることでバックアップや重複排除の処理効率を最大限に高めることができるなど、容易なシステム導入や効率的な運用が可能となる点も特長としている。

4. おわりに

ここでは、きたるべきビッグデータ時代へ向けた新たな日立ストレージソリューションについて述べた。

情報社会におけるデータ量の増加と技術の進歩とともに、その情報から先読みして未来をひらく新たな社会イノベーションが数々生まれることが期待される。

今や社会インフラとなったストレージにより、世界の英知・智慧・存在の記録を永く蓄積・継承し、アクセス可能にするだけでなく、さらに新たな価値を見つける宝庫として、社会の進歩に貢献していきたい。

※3) 多世代バックアップ運用時での重複排除の効果。対象データの種類と内在比率、およびバックアップ運用方法によって変動する。

※4) バックアップ容量を $\frac{1}{20}$ に削減した場合の、重複排除前のデータ容量。

執筆者紹介



島田 朗伸

1989年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 事業統括本部 企画本部 所属
現在、ITプラットフォーム全体の戦略策定の取りまとめに従事



山本 康友

1992年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 事業統括本部 ストレージ企画部 所属
現在、ストレージ製品企画策定の取りまとめに従事



阿部 哲也

1993年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 事業統括本部 ストレージ企画部 所属
現在、ミッドレンジストレージの企画業務に従事



喜多村 将之

1993年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 事業統括本部 ストレージ企画部 所属
現在、ファイルストレージの企画業務に従事

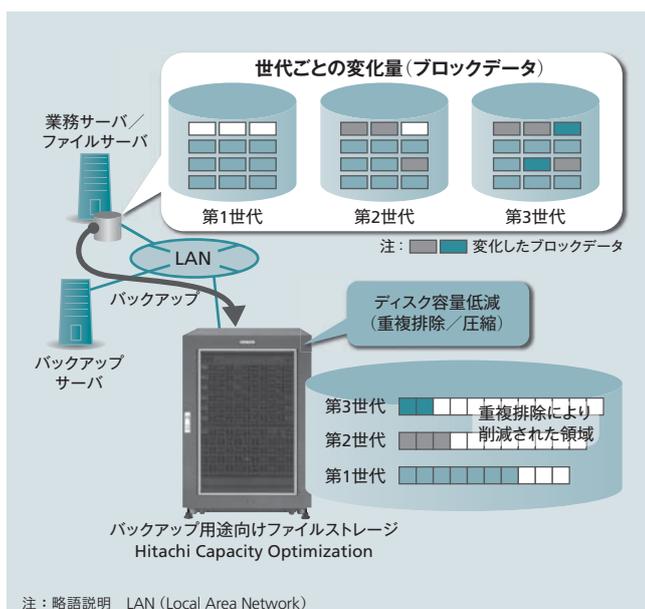


図4 | Hitachi Capacity Optimizationの適用システム構成

バックアップサーバなどから書き込まれたファイルデータを、Hitachi Capacity Optimizationの重複排除エンジンがバックグラウンドで重複排除・圧縮し、バックアップデータ容量を削減する。