

大容量・高信頼グリッドストレージ iStorage HSシリーズ(HYDRAsTOR)

川名部 正純・吉村 茂・宇高 淳也
吉岡 弘・水町 弘明・加藤 貢

要 旨

経営情報のバックアップや取引先とのメールアーカイブなど、長期保管が求められる企業データが急増するなか、大容量データを安全かつ容易に格納したいというニーズが、ますます高まっています。「HYDRAsTOR」は、革新的なグリッドアーキテクチャの採用により、高い性能と拡張性、信頼性、及び運用管理の省力化を実現し、この要望に応えるグリッドストレージであり、ビッグデータの格納に適しています。本稿では、HYDRAsTORに採用されている技術の概要及び特長と、活用方法について紹介します。

キーワード

●ビッグデータ ●グリッドストレージ ●重複排除 ●スケーラビリティ
●仮想化 ●高信頼性

1. はじめに

NECでは、増大している企業データ、メールアーカイブ、画像データなどのビッグデータを高い信頼性で格納し、拡張できる国産ベンダ唯一のスケールアウト・ストレージ基盤「HYDRAsTOR」を提供しています。

HYDRAsTORは、性能を業界最高レベルまで拡張可能です。北米の著名な評論家で、TechTargetのStorage Media Group編集主幹のW. Curtis Preston氏も、HYDRAsTORを最も高速な重複排除ストレージとして紹介しています¹⁾。また、容量についても、最小構成の71テラバイト (TB) から最大構成の6.3ペタバイト (PB) まで、必要に応じてスケーラブルに拡張することができます (20倍圧縮時の論理容量)。

2. 概要

HYDRAsTORは、グリッドアーキテクチャを採用した大容量、高信頼なストレージ基盤であり、ビッグデータの格納に適した次のような特長を持っています。

(1) 性能・容量の動的拡張と自動最適化

グリッドアーキテクチャにより、ノード追加で性能と容量を動的に拡張することができます。容量も、日々増大していくユーザーデータにスケーラブルに対応可能です。

また、複数ノードに分散されたデータは仮想化され、単一のストレージプールとして見えるため、大容量のデータ格納庫として扱うことができます。

(2) 高効率なデータ格納による高コストパフォーマンス

HYDRAsTOR独自の重複排除技術 (DataRedux) により、データの格納効率を格段に高め、格納するデータ量に対する実際の物理ディスク容量が大幅に縮小されます。このため、大容量のデータを低コストで格納しておくことが可能になります。

(3) RAIDを超えた高信頼性

HYDRAsTORは、複数のノードにデータの分散冗長配置を行います。指定により、2~6個のパリティを付加しますが、3個のパリティでは、3つのデータブロックを同時に失っても元のデータを復元することが可能です。このときの信頼性は、一般的に最大2つのHDDの同時故障にまで耐えることで知られるRAID6よりも優れており、大容量のデータを高信頼に格納しておくことが可能になります。

(4) ノード交換による長期データ保存

HYDRAsTORの自動最適化を利用すると、データ移行を行わなくても、データを保持したまま古いノードを新しいノードに置き換えることが可能です。このノード交換を計画的に行い、古いノードを新しいノードに順次交換していくことにより、データ移行を行わずに長期間大容量のデータを保存することができます。

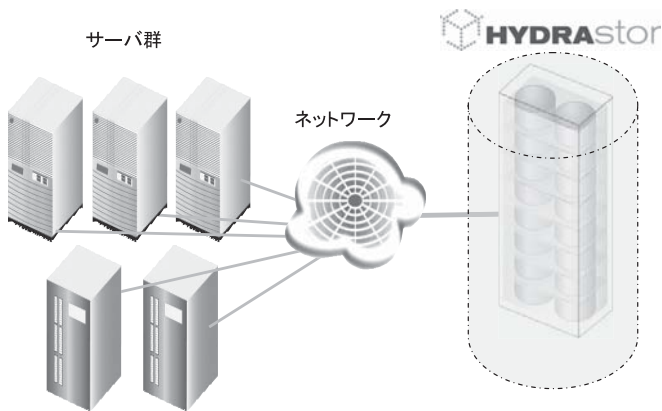


図1 システム構成イメージ

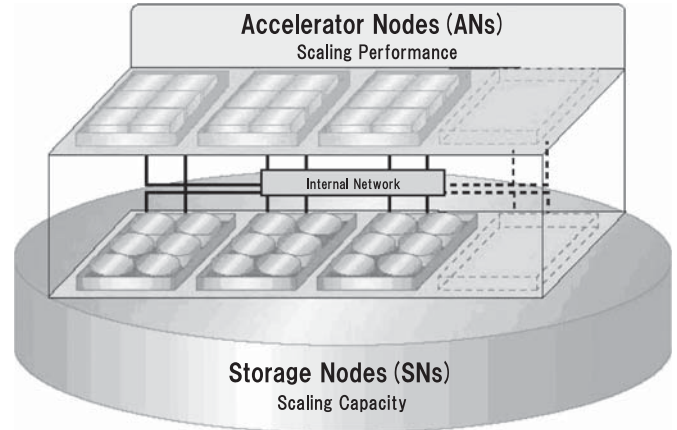


図2 グリッドストレージアーキテクチャ

なお、データアクセスプロトコルとしては、ネットワーク共有プロトコルとして一般に広く普及しているNFS (Network File System) /CIFS (Common Internet File System) /OST (OpenStorage Technology) をサポートしており、ネットワーク共有サーバとして、UNIXやLinux、Windowsなど幅広いプラットフォームから利用することができます。図1はHYDRAsTORを用いたシステムの構成イメージを示しています。

第3章では、これらの特長を具体的に紹介します。

3. HYDRAsTORの特長

3.1 性能、容量の動的拡張と自動最適化

HYDRAsTORは、データリクエストの処理を行うアクセラレータノードと、データブロックを実際に保存するストレージノードの2種類のノードから構成されます(図2)。アクセラレータノード、ストレージノードを追加することにより、性能と容量を動的に拡張することが可能です(図3)。

データ書き込み速度は1ノードあたり最高750MB/sであり、外部ネットワーク環境の条件が許す範囲でリニアに向上させることができます。容量については、71TBから6.3PBまで拡張できます(20倍圧縮時の論理容量)。ノード追加は、データ格納位置を考慮することなく、必要に応じて実施可能です。更に、自動最適構成技術(Dynamic Topology)により、追加ノードは自動的にシステム内に認識され、ボトルネックが発生しないよう自律的に最適な構成にデータが分散再配置されます。

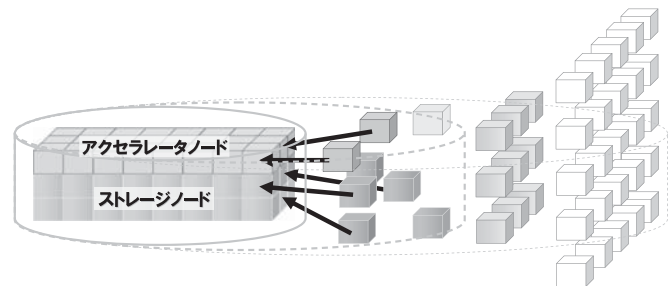


図3 自動最適構成 (Dynamic Topology)

この自動最適構成技術により、従来では極めて複雑であった次のような運用管理が簡便となり、管理コストを大幅に削減できます。

- ・ 保存データ量の増大に伴う容量の拡張
- ・ データ転送量の増大に伴う性能の拡張
- ・ 性能ボトルネックの改善
- ・ 障害発生時のノード交換

3.2 高効率なデータ格納によるコスト削減

HYDRAsTOR独自の重複排除技術DataReduxは、書き込まれるデータの重複をチェックし、既にストレージ内に書き込み済みのデータと重複するデータを新たに書き込まないことによって、データの格納効率を格段に高め、高性能と高コストパフォーマンスを実現しています。

重複排除 (DataRedux)

知的な可変長分割により、重複部分が最大となるよう分割

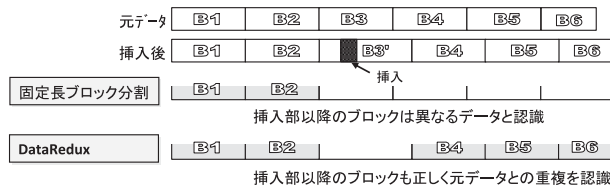


図4 重複排除 (DataRedux)

DataReduxは、既存データとの重複を最大限に検出するよう、知的にデータを可変長に分割します。これにより、固定長のデータ分割手法では検出できなかったデータの重複を最大限に検出することが可能となります (図4)。

この重複排除技術により、ディスクへのデータ転送量と、保存するデータに対する実際の物理ディスク容量が大幅に縮小されます。このため、日々のディスクへのデータ書き込みを高速かつ低コストで実現することが可能です。また、この技術を遠隔レプリケーションにも適用することにより、転送データを更に圧縮することが可能になります。これにより、遠隔サイトへのデータ転送量を飛躍的に小さくでき、帯域の狭い低速回線を用いた遠隔レプリケーションを実現できます。

3.3 RAIDを超えた高信頼性

前節で説明した重複排除技術を用いると、1つのデータブロックを複数のデータが共有します。このとき、1つのデータブロックが万一消失した場合の影響は、そのデータブロックを参照しているすべてのデータに及ぶため、影響が広範囲にわたる恐れがあります。そこで、HYDRAStorでは、従来のRAID (Redundant Array of Independent Disks) よりも更に強固な信頼性を実現するため、データの分散冗長配置 (Distributed Resilient Data) を行います。保存するデータブロックを更に分割し、冗長コードを付加して複数のストレージノードへ分散格納することにより信頼性を向上させています。

図5は、元のデータブロックを9つに分割し、3つの冗長コードを付加する場合の例です。この例では、1から12までの分割データ (フラグメント) が4台のストレージノードに分散して配置されています。このとき、12のフラグメントのうち同時に3つまで失っても、元のデータを復元することが可能で

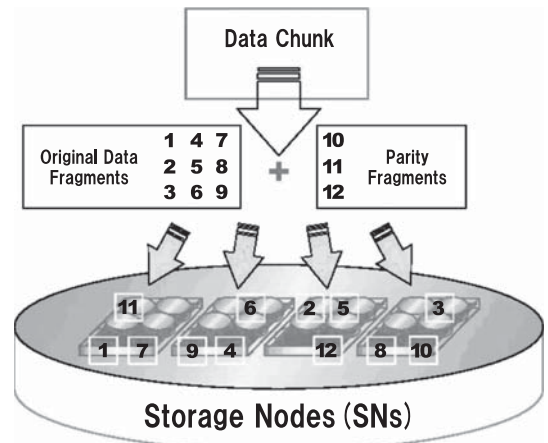


図5 データの分散冗長配置

す。このときの信頼性は、一般的に最大2つのHDDの同時故障にまで耐えることで知られるRAID6よりも優れています。更にこの冗長度は、保存データの重要度などに応じて自由に設定することができ、管理者は柔軟にシステムを構築、管理することが可能です。

HYDRAStorは万一の故障の際、故障部分を自動的に検出し、バックグラウンドで再構成の処理を実施します。そのため、通常、管理者に必要とされる面倒な管理業務を必要としません。更に、この再構成は十分な処理能力を持った複数のストレージノードで、他に実行中の処理を妨げるようなオーバーヘッドをかけずに処理されます。

図6は、故障により失われたフラグメント2、5、12が、即座に検知されて自動的に他のストレージノードへ再構成・再配置される様子を示しています。

この分散冗長配置技術により、既存ディスクストレージ製品を大幅に上回る信頼性と、障害発生時の管理コストの削減を実現しています。

3.4 ノード交換による長期データ保存

HYDRAStorの自動最適構成技術を利用すると、一括データ移行を行わなくても、データを保持したまま古いノードを新しいノードに交換することが可能です。

- ・新しいノードをHYDRAStorに追加すると、他のノードから新しいノードへデータが再配置されます。

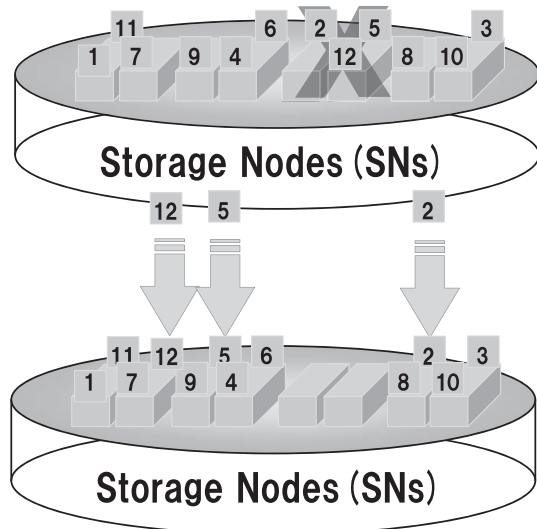


図6 自律データ復旧

- 次に、古くなったノードを構成から削除すると、このノードに保存されていたデータは、他の残ったノードに、システム全体がバランスを取れるよう自動的に再配置されます。

このノード交換を計画的に行い、順次古いノードを新しいノードに交換していくことにより、データ移行を行わずとも、システムを徐々に新しいハードウェアに置換していくことが可能となり、長期間大容量のデータを保管することができます。

4. ユースケース

次に、大量の映像・画像データを扱うお客様の導入事例を紹介します。HYDRAsstor導入前は、日々増加するコンテンツをサーバに蓄積することができなかったため、過去のコンテンツについては、テープで倉庫に保管していました。このため、過去のコンテンツを利用するには時間が掛かり、有効活用できないという課題がありました（図7）。

この解決策として、大容量コンテンツをHYDRAsstorに一時保存し、利用時にはネットワーク経由でコンテンツを転送することで、迅速な映像配信・編集を実現しました（図8）。なお、長期間参照されていないコンテンツの保管先としてテープ装置も導入しています。

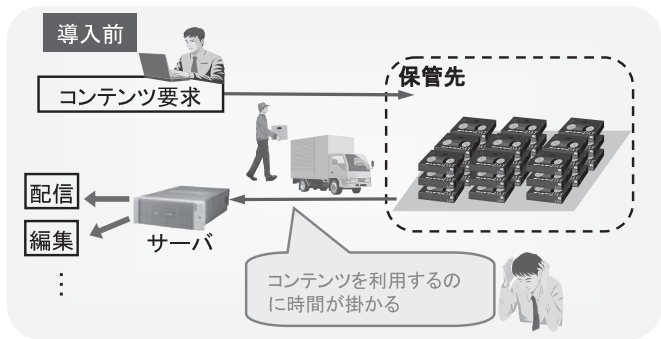


図7 HYDRAsstor導入前のイメージ

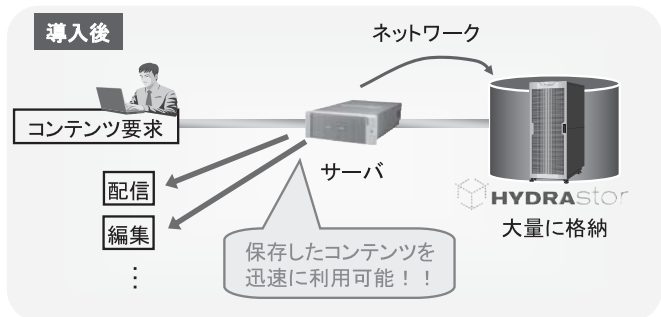


図8 HYDRAsstor導入後のイメージ

5. おわりに

ベースサーバの高性能化、HDDの大容量化に対応し、HYDRAsstorは今後もビッグデータ格納用のストレージ基盤として、進化を続けていきます。

*UNIXは、The Open Groupの米国およびその他の国における登録商標です。

*Linuxは、Linus Torvalds氏の日本およびその他の国における登録商標または商標です。

*Windowsは、米国Microsoft Corporationの米国およびその他の国における登録商標です。

参考文献

1) W. Curtis Preston "Target deduplication appliance performance comparison," <http://www.backupcentral.com/mr-backup-blog-mainmenu-47/13-mr-backup-blog/348-target-deduplication-appliance-performance-comparison.html>, 2010.10

執筆者プロフィール

川名部 正純
ITソフトウェア事業本部
第一ITソフトウェア事業部
マネージャー

宇高 淳也
ITソフトウェア事業本部
第一ITソフトウェア事業部
マネージャー

水町 弘明
ITソフトウェア事業本部
第一ITソフトウェア事業部
エグゼクティブエキスパート

吉村 茂
ITソフトウェア事業本部
第一ITソフトウェア事業部
マネージャー

吉岡 弘
ITソフトウェア事業本部
第一ITソフトウェア事業部
シニアマネージャー

加藤 貢
ITソフトウェア事業本部
第一ITソフトウェア事業部
事業部長代理

関連URL

iStorage HSシリーズ:
<http://www.nec.co.jp/products/istorage/product/backup/hs/index.shtml>

NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報(日本語)

NEC Technical Journal(英語)

Vol.65 No.2 ビッグデータ活用を支える 基盤技術・ソリューション特集

ビッグデータ活用を支える基盤技術・ソリューション特集よせて
ビッグデータを価値に変えるNECのITインフラ

◇ 特集論文

データ管理/処理基盤

超高速データ分析プラットフォーム [InfoFrame DWH Appliance]
SDN 技術で通信フローを制御する [UNIVERGE PF シリーズ]
大量データをリアルタイムに処理する [InfoFrame Table Access Method]
大量データを高速に処理する [InfoFrame DataBooster]
ビッグデータの活用最適なスケールアウト型新データベース [InfoFrame Relational Store]
高い信頼性と拡張性を実現した Express5800/ スケーラブル HA サーバ
大規模データ処理に対する OSS Hadoop の活用
大容量・高信頼グリッドストレージ iStorage HS シリーズ (HYDRAStor)

データ分析基盤

ファイルサーバのデータ整理・活用を支援する [Information Assessment System]
超大規模バイオメトリック認証システムとその実現
WebSAMの分析技術と応用例～インバリエント分析の特長と適用領域～

データ収集基盤

スマートな社会を実現する M2M とビッグデータ
微小な振動を検知する超高感度振動センサ技術開発とその応用

ビッグデータ処理を支える先進技術

多次元範囲検索を可能とするキーバリューストア [MD-HBase]
高倍率・高精細を実現する事例ベースの学習型超解像方式
ビッグデータ活用のためのテキスト分析技術
ビッグデータ時代の最先端データマイニング
ジオタグ付きデータをクラウドでスケラブルに処理するジオフェンシングシステム
柔軟性と高性能を備えたビッグデータ・ストリーム分析プラットフォーム [Blockmon] とその使用事例

◇ 普通論文

地デジ TV を活用した「まちづくりコミュニティ形成支援システム」

◇ NEC Information

NEWS

スケールアウト型新データベース [InfoFrame Relational Store] が 2 つの賞を受賞



Vol.65 No.2
(2012年9月)

特集TOP