

教育 研究分野のストレージソリューションに期待するもの

～ データマネジメントを意識したストレージソリューションWG 成果報告 ～

中京大学情報理工学部

磯 直行

fmiso@sist.chukyo-u.ac.jp

[Abstract]

2001 年度から「ネットワーク時代の統合ストレージマネージメントWG」、2003 年度から「ストレージを中心としたシステムマネージメントWG」の活動が始まり、続いて 2005 年度から「データマネジメントを意識したストレージソリューションWG」が開始された。「データマネジメントを意識したストレージソリューションWG」では、会員機関で運用されている大規模ストレージシステムで増大しつつある多様な非構造型データについて、その特性を分類し、その効果的な管理方法とストレージソリューションの明確化について検討したのでここに報告する。特に、大学・研究所といった教育・研究分野で使用される大規模ストレージシステムにおいて、既存または今後期待する技術について述べる。

[Keyword]

ストレージソリューション、構造型 非構造型データ 仮想化、8 つのカテゴリ ガイドライン

1. はじめに

近年、実験・観測機器等の性能向上により科学技術計算の入力データ量が增大している。また、大学等の教育機関でもマルチメディアデータを教材として取り扱うようになり、やはりそのデータ量は増大している。加えて、研究・教育活動に Web システムや電子メール等がなくてはならない存在となり、多種多様なデータを取り扱う機会が増え、ユーザの周辺にはますますデジタルデータが氾濫していく。

これまで研究所や大学では比較的大規模なストレージシステムを用意し、それらをユーザが共通利用することにより高速かつ高可用のシステムを維持してきた。しかし、このような時代背景から最近では単にデータを保存する機能に注目するだけではストレージシステムの管理は十分ではなくなりつつある。

従来取り扱ってきたデータは、その目的や方法を明確にした上で形式を定めたデータ、いわゆる「構造型データ」であることが多かった。例えば、リレーショナルデータベース(RDB)では、データをさらに細分化するためのレコードをあらかじめ設定し、それに対してインデックスを作成する等の処理を可能にしている。このように、データに対してある一定の内部構造を付加することにより効率の良いデータ処理を行うことができた。

しかし、最近ではデータの種類や大きさについてそれらの間の関係が不明確な「非構造型データ」が増えている。例えば、電子メールのようなデータはストレージシステムから見ると単なるファイルであり、内容に関わらずすべて同一のものとして扱われる非構造型データである。

一方、大学や研究所では、測定・観測データ等の学術データを長期保存するためにデータアーカイブが行われているが、単にデータを保存しておくだけでは不十分である。必要なときに必要なデータを適切な時間内で取り出せることが望ましい。しかし、ユーザとしてはファイルに対して発生時期、場所、種類等のデータの意味を考えながら使用したいにも関わらず、実際には短いファイル名だけを頼りにそれらを探索することくらいしかできない。ファイルを大量に保存しておいても、目的のファイルを発見するまでに多くの時間を費やすか、途中で探索をあきらめることになる。当然ながら、管理者は各ユーザが考えるファイル分類についてすべてのパターンを考慮した運用はできないため、現状のストレージ

システムではユーザ全員の要望を満足させる状況にはなっていない。

今後、このような非構造型データのファイルに対して、ストレージシステム内で効率良く管理し、ユーザが望むファイルをすぐに提供できるデータマネジメントシステムが必要になってくるであろう。そのためには、プラットフォームに依存しない仮想化による処理環境がストレージシステムに要求されるに違いない。また、ユーザにとって意味のあるファイル群を1つにまとめて管理するコンテンツ管理ミドルウェアも必要になるだろう。

そこで、サイエンティフィック・システム研究会システム技術分科会では、まず、ストレージシステムをデータの相互活用を目的とした統合ストレージシステムとして位置づけることを試みた。それが2001年度に始まった「ネットワーク時代の統合ストレージマネジメントWG」である。当時ホットな話題であったSAN/NASの技術動向調査と大学・研究機関への導入モデルを検討した。次に、それらのストレージシステム全体を効率良くマネジメントする方法について2003年度から「ストレージを中心としたシステムマネジメントWG」で議論され、システム管理者が設計段階および運用段階で考慮すべき項目について整理した。同時に、設計時および運用時に考慮しなければならないポリシーを設定し「ストレージシステムポリシー評価ワークシート」としてまとめ、具体的に会員機関の設備でその評価も行った。そして、具体的な調査事例をもとに、ストレージシステムをどのようなアプリケーションやミドルウェアで管理すればより良いソリューションが得られるのかを検討するために、2006年2月に「データマネジメントを意識したストレージソリューションWG」が活動を開始した。このWGでは、大学や研究所で氾濫するデータの種別を洗い出し、分類を行ったうえで、そのストレージシステムをどのようなアプリケーションやミドルウェアを利用すると管理効果が上がるのかということに着目し検討を重ね、ストレージソリューションに必要な要件を明確化する検討を行った。

以下では「データマネジメントを意識したストレージソリューションWG」の活動について述べる。特に本WGで注目した仮想化と8つのカテゴリについて述べ、最後に、データの相互活用を意識した統合ストレージシステムとして今後のストレージソリューションに期待することについて述べる。

2. WGの活動

本WGの活動の主旨は、「大学・研究所における代表的な業務データを洗い出し、分類を行った上で、それらをどのようなアプリケーションやミドルウェアで管理しているか、または管理すべきかを検討するとともに、大学・研究所が目指すデータマネジメントを意識したストレージソリューションの明確化とその対応を検討する」ことである。この主旨に基づいて、本WGは年4回のペースで会合を開催した。まず、会員機関の研究所・大学におけるデータ特性の分析、および研究開発系の構造型データと非構造型データの整理と検討を行った。その結果として、各業務におけるストレージシステムに求められるソリューションが明らかになった。

3. Sub WG 活動

本WGでは、Sub WGが特に活発に行われた。前WGである「ストレージを中心としたシステムマネジメントWG」の成果物として得た「ストレージシステムマネジメントポリシー評価ワークシート」を再検討し、より実践的に利用できる「ストレージシステム設計・導入にあたってのガイドライン」として改良した。

ガイドラインの作成については、Sub WG活動とし、WGの通常の会合とは別に活動を行った。

Sub WGの活動の結果として、2007年9月に「ストレージシステム設計・導入にあたってのガイドライン」が完成した。このガイドラインは、ストレージシステムの導入および運用段階で留意すべき事項について、まず各項目の技術解説を行った後、その実現手段だけでなく、ストレージ管理者が業務上すぐにも利用可能な具体的な製品事例についても情報提供を行っている。このガイドラインによって、経験の少ない若い管理者でもその機関に対してバランスの良いストレージソリューションの選択ができるようになったと思われる。

4.8 つのカテゴリ

大学・研究所における代表的な業務データの整理・分析を行うにあたり、データマネジメントを意識した視点から現状のストレージシステムの課題・要望を分析するために、本 WG ではストレージソリューションを構成する代表的な 8 つのカテゴリを定めた。これは、IDC (International Data Corporation 社) が定義しているストレージソリューションのカテゴリに基づいている。

1. データプロテクション
2. アーカイブ・HSM (階層ストレージ管理)
3. ストレージ管理
4. データアクセス管理 (ストレージインフラストラクチャ)
5. デバイス管理
6. データレプリケーション
7. ファイルシステム
8. その他 (圧縮、暗号化など)

4.1. データプロテクション

データプロテクションは、格納されたデータがそのままの状態参照できることを保証することである。ストレージシステムの大容量化、および格納されるデータ特性や種類の多様化により、ディスク全体のバックアップをはじめ、個別データのアーカイブによる選択的保護、さらに更新されたデータのみを自動検出して保護するまでに発展している。RAID をはじめ、一般家庭にもその技術の普及が始まっている。

4.2. アーカイブ・HSM (階層ストレージ管理)

バックアップがデータの復元手続きを前提とするデータ保護技術であることに比べ、アーカイブは保護機能が特別に整備されている領域にデータを格納することで、処理手続きの後のデータ保護技術である。HSM はファイルシステム空間の拡張を実現する技術である。

4.3. ストレージ管理

大規模化、複雑化するストレージシステムをサーバに接続し、ファイルサーバ等としてアクセス可能にするための論理的な記憶空間 (ストレージ空間やファイルシステム空間) アクセスポス、可用性レベルなどのサービスを提供する仕組みがストレージ管理である。

4.4. データアクセス管理 (ストレージインフラストラクチャ)

データアクセス管理は、可用性維持と性能確保のために複数のネットワークパスを制御し、ストレージやファイルサーバの記憶空間とアプリケーションを処理するサーバとの接続を制御する。QoS によるデータ転送帯域保証が行われることが多い。

4.5. デバイス管理

ストレージシステムは個々のデバイスであるディスクドライブ、コントローラ、ネットワークで構成される。複雑化するストレージ関連装置を構成しているモジュールの動作状況を監視し、システムとしての可用性を維持するために構成定義の変更、装置の追加・交換、データ移行の容易化等が必要である。ここまでに説明したストレージ管理、データアクセス管理、デバイス管理の関係を図 1 に示す。

4.6. データレプリケーション

データレプリケーション (データの複製) は、データアクセスを容易にし、処理性能を向上させるためにデータを必要とするシステムに近接させることと、データの保護レベルを向上させることを目的として行われる。キャンパスが離れている大学等では、学生が行き来して講義を受講する等の場面で有効

な機能である。

4.7. ファイルシステム

ファイルシステムは、ストレージ記憶空間をアプリケーションソフトウェアから POSIX などの標準インタフェースでアクセスさせるためのインタフェースソフトウェアである。通常はサーバ OS のコンポーネントとして内蔵されている。

4.8. その他（圧縮、暗号化など）

上記のほかに、データマネジメントレベルでのデータ圧縮が考えられる。ファイルの同一性検査の仕組みを発展させ、ファイル内の同一ブロックを検出し、それらを共用ブロックとして管理すればストレージに保存されているデータ量を圧縮することが可能である。特に非構造型データに対してファイルの重複、データの部分的な重複が多く含まれることを利用して、重複部分を排除して保存する De-dup 技術が注目を集めている。

また、情報のアクセス権限管理と暗号化機能を組み合わせ、データアクセス管理に適用することでストレージのセキュリティ管理機能を強化できる。

このほか、データの暗号化はネットワーク管理やファイル管理においてすでに実装されている技術であり、一般ユーザのプライバシーとシステム管理者などによる内部監査や情報統制にまで対応する必要がある。

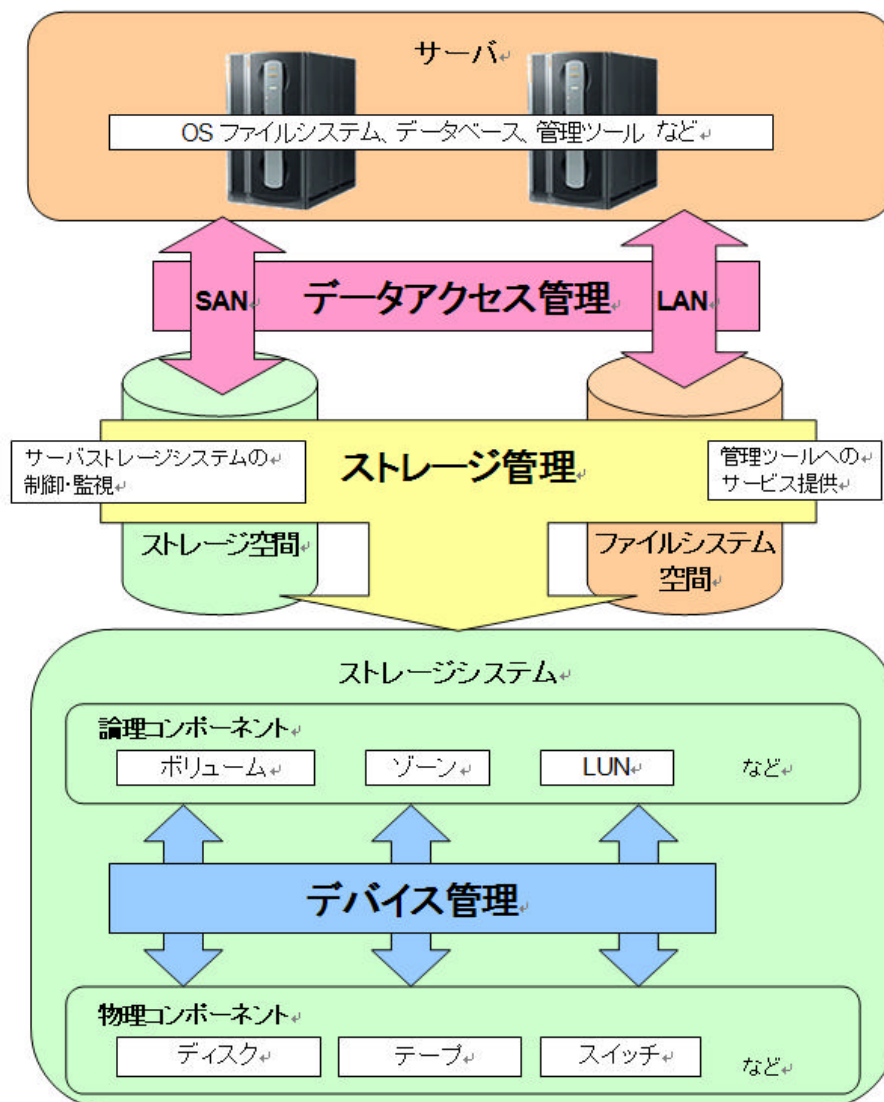


図1：ストレージ管理、データアクセス管理、デバイス管理

5. 会員機関における8つのカテゴリの重要性

各会員機関に対して、以上に説明した8つのカテゴリについて、どれを重要視してストレージシステムの管理運用を行っているかを調査した。カテゴリごとに最重要課題を で(5点)、重要課題を で(3点)、将来対応すべき課題を で(1点)、対応不要は - で(0点)として選択し重要度に関する順位付けを行った。その結果を表1に示す。

この結果からわかることは、まず、どの会員機関も「データアクセス管理」を最重要と回答していることである。エンドユーザにとって、データアクセス速度は直接的に影響のある部分であり、データアクセス管理に関するサービスの向上が望まれている。

また、管理者の視点からは、管理負荷の軽減に関する項目が上位に列挙されている。システムの複雑化が進む中で、装置の状況認識の即時性の向上が望まれている。このほか、マルチベンダによるシステム構築の結果、ログや解析データが装置ごとに異なり、その統合管理もストレージソリューションに期待されている。

表1：課題の重要度一覧

		JAXA	天文台	中京大	JAIST	理研	得点
1	(4)データアクセス管理						25
2	(2)アーカイブ・HSM						19
3	(1)データプロテクション						17
4	(7)ファイルシステム						15
5	(5)デバイス管理						15
6	(3)ストレージ管理						15
7	(6)データレプリケーション			-			6
8	(8)その他(圧縮、暗号化など)		-	-	-		2

6. ストレージソリューションに期待するもの

前節まで述べたように、まずユーザの観点から期待したいものは、利用者の利便性の向上であることがわかる。アクセススピードの確保をはじめ、ユーザが直接的に体感できものについては、今後もオペレーションの簡素化、ハードウェア・ソフトウェア機能の最大活用が望まれる。また、大学のように離れたキャンパス間を学生が行き来するような場合、遠隔地のストレージシステムとの連携を強化し QoS を維持することにより地域差がなくなり、ユーザは良いサービスを受けることができる。

次に、ストレージ管理者の立場として期待したいものは、管理コストの削減である。特に、システム更新時や、大学で卒業生を送り新入生を迎える時にはデータ移行作業が発生する。マルチベンダにより構築された各種ストレージシステムの統合的な情報把握ができると、効率の良い移行作業が可能である。また、バックアップやリカバリ処理の時間短縮、バックアップの信頼性向上も望まれている。もちろん、限られたコストの中でシステムに最大パフォーマンスを発揮させるよう考慮しなければならない。例えば、ストレージシステムの価格低下の実現や、省電力・省スペースが実現できるとなお良いだろう。

そして、非構造型データの効率的な管理やコンテンツ管理機能の向上にも大きな期待を寄せたい。本WGでは、単なる製品レベルの性能向上では対応は難しく、今後開発されるハードウェア・ミドルウェア・OS などの総合的な技術が必要になるだろうとの結論に至った。今後も引き続き使い勝手がよく、柔軟な管理が可能で、コストパフォーマンスに優れた仮想化技術や、コンテンツ管理ミドルウェアの実現に期待したい。

7. おわりに

2001年度から始まった一連のストレージ系WGについて、データマネジメントを意識したストレージ

ソリューションWG」の成果を報告した。ストレージシステムが多面で使われるようになり、これまでのように単に高速、大容量の観点だけでなく、データの特長も含めた運用が求められているのが現状である。普段ストレージシステムは裏方のシステムのように思えるが、技術的にはシステム全体の性能を決定してしまうこともある重要なものである。データマネジメントも意識しなければならない時代になり仕組みが複雑になっていくのは確実だが、今後もストレージシステムの進化に期待したい。