



MIC Associates, Inc.

NAS vs. SANから 統合、融合へ

ここに記載された内容は更新される可能性があります。この文書に記載されている内容はこの文書の発行時点におけるエムアイシー・アソシエーツ株式会社の見解を述べたものです。エムアイシー・アソシエーツ株式会社が、この文書に記載された内容の実現に関して確約するものではありません。また発行日以降については、この文書に記載された内容の正確さは保証しません。

この文書は情報の提供のみを目的としており、明示的または黙示的に関わらず、この文書の内容についてエムアイシー・アソシエーツ株式会社はいかなる保証をするものでもありません。

エムアイシー・アソシエーツ株式会社は、本書に記載してあるすべて、または、一部の記載内容に関し、許可なく転載、または、引用することを禁じます。

バージョン	作成日付	旧バージョンからの 変更点	総ページ数
3.0	2009/05/22	全面改訂	8
2.0	2009/3/23	全面改訂	5
1.0	2009/01/14	初版	2

本書作成、編集、管理



エムアイシー・アソシエーツ株式会社
〒103-0004 東京都中央区東日本橋3-
12-12

櫻正宗東日本橋ビル9F
Tel. 03-5614-3757 Fax. 03-5614-3752

目次

はじめに	1
NAS vs.SANの議論からNASとSANの統合へ	1
SANとNASの優位点を統合	2
NASとSANの融合	2
1.NASヘッド	2
2.NASヘッドのフェールオーバーとロードバランス	3
3.ストレージの増設を保証するシリアルインタフェース	4
4.ボリューム容量の追加	4
NAS性能の問題点	5
NAS以外のソリューションのご紹介	5

はじめに

古くからユーザや業界でしばしば議論されてきたのが「SAN vs. NAS」というテーマです。筆者がこのテーマのパネルディスカッションを初めて聞いたのが1997年のラスベガスでのコンベンションですから、いかに長い間議論されてきたかということです。当時はファイバチャネルの第一世代の1Gb FCを持つストレージや、スイッチがリリースされ始め、漸くSANについての理解が進み、実際にSANの導入が始まるようになっていた時期です。一方、NASは丁度この頃Network Attached Storageと言う名前で、NetApp、IBM等が製品を発表したところで、このコンベンションでもこれらの企業がハイエンド製品を発表していました。Unix系のNFSと、Microsoft系のCIFSが実装されたNASは、複数のネットワーククライアントからのファイルレベルでの共有を可能にし、ワークグループの生産性向上に多に貢献しました。一方、SANに関しては当時はパラレルインタフェースからシリアルインタフェースへ、ハードウェアのリソース共有(ブロックデバイスとしてのストレージの共有)という観点で議論がなされ、ストレージに保存されたデータの共有というポイントからは距離が有ったように思われます。いずれにしてもNASはファイル共有のソリューションで、SANはストレージ共有のソリューションという点で話がされ、議論がなかなか噛み合なかったことを記憶しています。

さて、あれから既に10年以上が経過し、現在のSAN vs NASというテーマの議論には若干の変化が起きている様に思われます。

NAS vs.SANの議論からNASとSANの統合へ

1994年頃に紹介された初期のNASは一方側面にSCSIのポートを持ち、反対の側面にはEthernetのRJ-45の口を持つ今で言うエンベット型の装置でした。それまでは、高いサーバ用コンピュータとサーバソフトを購入し、SCSIのインタフェースに高価なSCSI RAID装置を接続しないとファイルサーバを構築できませんでした。その為、各自がフロッピーや、MOでデータをバックアップしたり、データ交換を行っていました。この様なNASが出て来る事で、ファイルサーバのコストを大幅に低減することができ、ユーザに多に受け入れられました。

その後、Unix系のNFSや、Windows系でSMB/CIFSプロトコルを使用できるようになり、ビジネス環境で広く使用されるようになったため、ネットワークストレージとしてのNASは企業部門、中小企業に急激に普及しました。また、1992年に設立されたNetApp社はWAFLという独自のファイルシステムで、NASに特化したファイルサーバを出荷し急激な成長をし、ストレージ業界においてNASの存在感をますます高めることになりました。一方、SANに関しては、1988年からストレージ用シリアルインターフェースとして標準化の作業が始まり、1994年に漸くSCSIの1規格として承認されました。それ以降、FC-IAという業界団体によって各種FC関連製品の互換性テスト等が行われ漸く2000年には2Gb FCが本格的に使用されると同時にSANストレージばかりではなく、SANスイッチ、SFP、HBA等々のユーザ価格も実用的な金額になってきました。この間はNASもSANもそれぞれの技術の優位性を主張し、ネットワークという言葉キーワードに各種の議論が展開されていました。

さて、昨年、某大手コンサルタント会社が非定型のデータ量が今後3年間で現在の6倍程度に増加するという予想を発表しました。この爆発的に増大するデータの保存先はファイルサーバ、または、NASに保存されるケースが多いと考えられ、今後年率73%という高い率でNASの容量が増大すると予想しています。このようなデータ量の急激な増加に伴って、「NAS vs. SAN」の議論は「NASとSANの統合」に変わってきています。SANはサーバに直接接続されたストレージに対してアクセスするのにに対し、NASはネットワーク経由でストレージに対してアクセスします。二つのアーキテクチャーの最大の違いはファイル共有の可否と、IO性能と言えます。

SANとNASの優位点を統合

IO性能に関しては、SANはファイバチャネルというSCSIプロトコルをそのままサポートするストレージ専用のインタフェースを経由し、一回のリードライトコマンドで最大2GBまでの大量のデータを4Gb/秒で転送することができます。更に、ファイバチャネルに代表されるシリアルインタフェースはアーキテクチャーの本来の仕様として活線挿抜がサポートされています。それゆえ、SAN用ストレージ機器はシステムの稼動中でもストレージの追加ができるようになってきました。SANのユーザは現在の爆発的なデータ量の増大に対して初期投資を抑制し、必要に応じてストレージの追加で対応することが可能になりました。

一方、NASではEthernet上のTCP/IPプロトコルによって、ファイバチャネルより遥かに小さいパケットサイズで転送される上、大半のネットワークは1Gb/秒の性能ですので、データを転送するインタフェースとしては、SANに使用されるファイバチャネルの方が優位だといえます。更に、NASの方はTCP/IPのプロトコルを使用するため、上位のTCP層でデータコレクションを行い、コンピュータシステムに大きなオーバーヘッドが生じますが、SANの場合はHBAというハードウェアレベルでデータのチェックや修正を行うことができ、データ転送のオーバーヘッドは僅かです。ネットワークにはその公開性のために、クライアントコンピュータのNIC性能や、ネットワークメディアや、スイッチの品質によって、全体のサービス性能が左右される傾向にあります。1Gb Ethernetポートを仮想化させ、あたかも1つのポートとして見せながら、集めたネットワークの分だけバンド帯域を増加させるチーミングや、ボンディング等と呼ばれる技術、10Gb Ethernetなどにより、NASの性能上のボトルネックは改善されつつあります。また、NASは、SANでは実現するのが難しい、OSが異なる複数のクライアントコンピュータからのストレージ共有が可能です。特にWindows用CIFSでは、NASに保存されたファイルに同時にアクセスすることができますので、非定型データの高い可用性が必要なワークグループにとっては有用なストレージです。

SANに於いては、そこに保存されたデータをサーバで共有することは簡単にはできません。ファイル共有を保障するファイルシステムやミドルウェアを使用しない限り、1台以上のサーバから同時にあるファイルのデータを更新すると、このデータは消失することは必定です。

以上の様に、NASはNASの優位性が、また、SANはSANの優位性があり、長くこの議論がされてきました。しかし、近年非構造化データとよばれる画像、映像コンテンツや、メール等のデータが急激に増大したことで、NASはストレージの信頼性と拡張性を求められ、SANは複数のサーバによる真のストレージ共有、則ち、NASで実現しているファイルレベルの共有が求められるようになってきました。

このようにして、SANとNASは、「NAS vs. SAN」の議論から「NAS とSANの統合」に移行し、さらに「融合」が語られるようになったのです。

NASとSANの融合

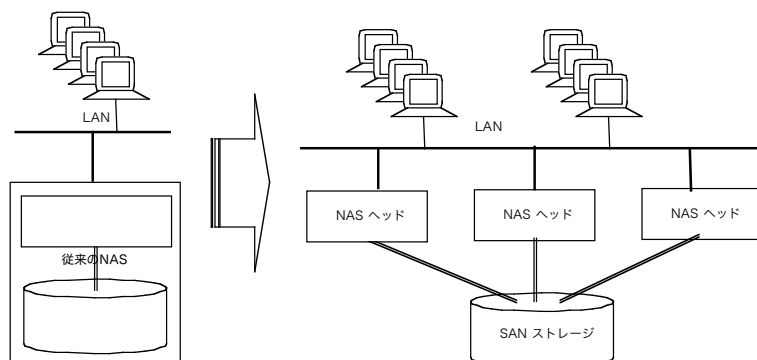
1.NASヘッド

先にも述べましたが、NASはネットワークに接続したファイルサービスに特化したサーバです。今日の非構造化データの急激な増大により、その重要性はますます増大する傾向にあります。

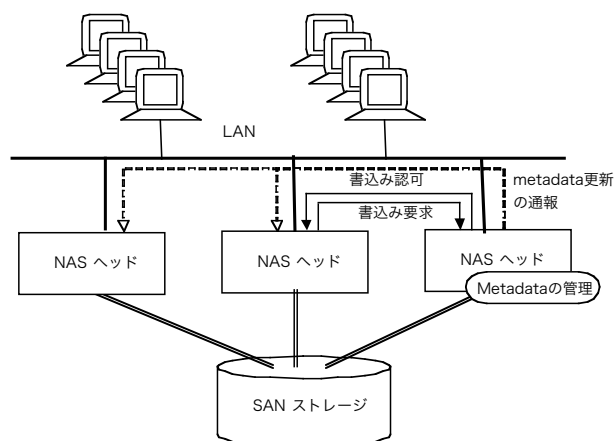
ストレージサーバとしてNASに求められていることは、24/7のサービス継続、ネットワークの影響によるクライアントサービスが劣化しない対策、無停止のストレージスペース拡張です。こうしたユーザの要求を実現する方法として、「NASとSANの融合」という概念が必要とされるようになりました。

従来ストレージと一体化されたNASという概念から、ストレージを複数のNASヘッドと呼ばれるサーバで共有、NASのサービスを効率良く提供します。万一、NASヘッドに障害が発生した場合は、代替のNASヘッドにサービスがフェールオーバーし、クライアントからのファイルアクセスが1台のNASヘッドに集中しないよう、複数のNASヘッド間で負荷を分散させるシステム構成が可能となってきました。こうして、NASヘッドは1サーバとしてSANに参加する様になり、「NASとSANの融合」が図られるようになっ

てきました。しかし、複数のNASヘッドがSANに参加し、クライアントからのファイルアクセスを可能とすることは、それほど簡単ではありません。複数のNASヘッドが同一のSANに参加するには、アクセスの排他制御をする必要があります。従来、シングルヘッドのNASによるCIFSのサービスの場合、複数のクライアントによるファイル共有は、VFSのインタフェースを介して簡単に行われてきました。しかし、SANを複数のNASヘッドで共有し、その内部に保存されたファイルに対しNASヘッドを介して、複数のクライアントで共有することは簡単ではありません。何故なら、同一のファイルに対するNASヘッドからの同時書き込みが発生した場合、直ちにデータ消失に繋がるからです。そこで、この問題を解決することで、初めてSANとNASの融合が実現すると言えます。NASの機能を維持しながら、ストレージに対するファイルレベルの共有の保証する技術がSAN側に求められるのです。



一般的にストレージのデータアロケーションテーブルを記録しているデータをメタデータといいます。複数のNASヘッドでストレージを共有するにはこのメタデータを専用のサーバ、または、一般のSANのメンバーで一元的に管理します。SANに参加する総てのNASヘッドはこのメタデータ情報の同期を取って更新して行く必要があります。このようなファイルシステムを分散ファイルシステムといいます。ストレージに書き込まれるファイルのアロケーションテーブルを特定のNASヘッドで管理し、それぞれのファイルに対する書き込み、読み込み要求に対し、許可を与える方法を排他制御といい、複数のNASヘッドからの同時アクセスに対し、常に唯一1ファイル、1書き込みという制御をすることで、データを破壊することなくサービスを継続することができます。



2.NASヘッドのフェールオーバーとロードバランス

万が一NASヘッド内のNICに障害が発生した場合や、NASヘッド本体に障害が発生した場合、そのサービスを継続するためにパスを他のNASヘッドにフェールオーバーすることも必要になります。このモードをサーバフェールオーバーといいます。フェールオーバーを実現するためにはパスの仮想化や、NASヘッドのフェールオーバーメカニズムが必要になります。複数のNASヘッドのIPアドレスを1つに仮想化し、NASヘッド間の負荷を適度に分散させるネットワークとNASヘッドの間にロードバランシングを設けることにより実現することができます。

一方、SAN側においては、同一ファイルシステムへのアクセスの制御を行う特定のサーバに障害が発生した場合、このサービスが継続して提供されていることを監視している他のサーバがこの特定のサーバに障害が発生したと判断し、そのSAN共有の制御を引継ぎ、ネットワーク全体に対するサービスを継続します。NASヘッドのアクティブ・アクティブ構成のフェールオーバーメカニズムを持ったクラスター型のサーバシステムの場合は、殆どの場合、ネットワーククライアントの負荷に応じて、負荷分散するロードバラン

シング機能も同時に実現することができます。弊社で提供していますmetaSANもこのクラスターファイルシステムの一つで、各アローケーションブロックテーブルをファイルシステムとは別途に持ち、SANのストレージに対し、複数のコンピュータからの同時アクセスを可能にします。

3.ストレージの増設を保証するシリアルインタフェース

SANストレージのハードウェアに関してですが、ユーザデータは急激に増大してまいりますので、必要に応じてストレージ容量を追加させる必要があります。従来のNASストレージでは簡単にはストレージ追加に対応することができないばかりではなく、増設の際に長時間のサービス停止をよぎなくされます。それは内部のバスが活栓挿抜ができる仕様になっていないためですが、近年のエントリークラス以上のストレージ製品はSCSIのプロトコルをサポートするFCやSAS等のシリアルインタフェースになり、新たにバスにログインするデバイスがある場合、総ての参加デバイスのノードテーブルを再登録し、バスのパワーサイクル無しに新たな構成に更新することができます。このことは、SANストレージを追加する上で非常に重要です。NASヘッドは共有するSANのストレージを必要に応じて追加し、新しい容量をユーザに開放する必要があります。

4.ボリューム容量の追加

新たに増設されたストレージリソースを新規ボリュームとして従来のストレージに追加されただけではストレージサーバに対するユーザ要求を十分満たしたとはいえません。ユーザは自分のデータが増えるに従って以前から割り当てられたストレージ領域を増やしてもらいたいという欲求をもちます。また、一つ一つの新規ボリュームに新たなアクセス認証の設定をしたり、その他の管理設定をすることは容易ではありません。既存に設定され、ユーザに公開されているボリュームを必要に応じて増設できることが求められます。SANストレージをそのまま増設して、容量を増やす方法と、ファイルアクセスデバイスを集合化させてファイルサーバの容量を増加させる方法があります。今回は、ブロックアクセスデバイス(SANストレージ)を前提にご紹介します。

最近リリースされているOSにはファイルシステムを作成する前に、物理ボリュームを集めて仮想的にボリュームを仮想的に大きなボリュームにするという概念があります。これは物理的に初期化されたストレージデバイスをボリュームグループという仮想的なストレージ領域に入れ、このストレージ領域を一つ以上の論理ストレージ領域としてファイルシステムで初期化する方法です。さらに、このストレージ領域は新たに追加されたストレージデバイスを新たなストレージ資源として、既存の論理ストレージ領域に加えることができます。最近のOSでLVM(Logical Volume Manager)とよばれるストレージ管理マネージャーを持つオープン系のOSには、こうしたボリュームグループ、LVMといったファイルシステム構築の環境が用意されるようになってきました。例えば、RedHat AS5に実装されているGFSファイルシステムはこのボリュームグループおよび、LVMをサポートしています。GFSのファイルシステムを作成するまでに、物理デバイス->ボリュームグループ->LVMでGFSファイルシステムを作成というプロセスを経て作成したボリュームがマウントされます。このボリュームグループは複数のストレージの境界を論理的に結合させる一種のスパニングにあたりますが、必要に応じて物理デバイスをこのボリュームグループに加え、さらにLVMで追加されたストレージスペースを既存のボリュームの増設リソースとして追加して行くことができます。

実際に弊社ではLinux環境でXFSのファイルシステムを使用し、当初14TBのストレージをVGに登録し、LVMでXFSのボリュームにした後、段階的に物理ボリュームをVG加え、最終的に70TBまでの単一の論理ボリュームとしストレージサーバとして公開し、問題無く稼動することを確認しています。

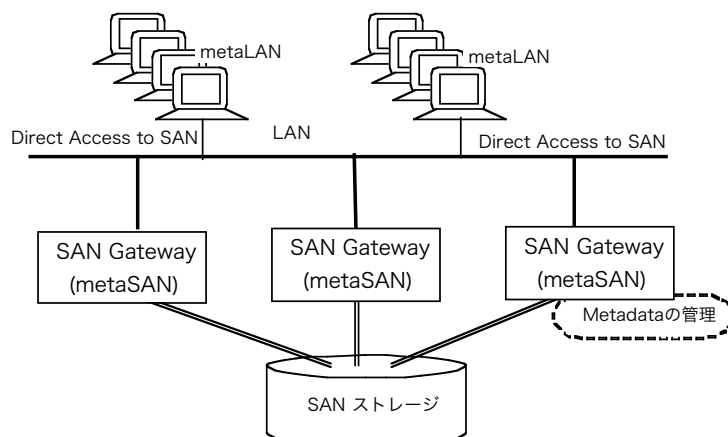
NAS性能の問題点

ここまで「NASとSANの融合」というテーマで話を進めてまいりましたが、従来のブロックアクセス vs. ファイルアクセスという議論から、いかに有効にデータを保存し、利用するかという基本的な目標の為に、NASとSANは融合してきたということが云えると思います。ただ、SANの場合には圧倒的に高速なアクセス性能と大容量データを一瞬で転送することができる大容量バンド幅がありますが、NASのファイルレベルアクセスにはTCPを介してデータにアクセスし、転送するという問題が潜んでいます。通常NASの場合、Sambaを介したCIFSやNFSプロトコルで複数ネットワーククライアントがNASのボリューム内のファイルを同時共有することになります。しかし、ネットワークを介してNASへのアクセスが集中したり、大容量のファイル転送がおこなわれた場合に、ネットワークに大きな負荷が掛かります。例えば、昼休みを終え、多くの作業者が同時にファイルサーバにアクセスする状況を考えてみましょう。ネットワークでは輻輳(ふくそう)という問題が起ります。これは、ネットワーク上で多重化したデータ転送により、TCPでのパケットロス等が発生、ネットワークの転送性能が著しく劣化する現象です。作業開始直後のユーザはその時点ではネットワーク、サーバを独占しますので、データ転送は大きなパケットで実行されます。しかし、ファイルサーバにアクセスする作業が増えるに従い、データ転送にストレスを感じる様になり、やがてデータの転送をひたすら待つ非効率的な時間を過ごすこととなります。通常、1Gb Ethernetでは帯域の80%程度までデータ転送を行うことができますが、一旦、この輻輳が発生すると帯域の10%程度までデータ転送レートが落ちる場合もあります。通常、各NICのドライバには輻輳制御機能が実装されています。この機能によりネットワークが混雑していると判断するとデータ転送を送信側で絞ってしまい、結果的にはNASの性能が大幅に劣化することを制御することができなくなります。また、このような状況がネットワークで発生すると、スイッチングハブなどのフロー制御機能により、一時的にデータ転送が停止する場合があります。

このような問題はネットワークの問題ですので、NASの側でネットワークのボトルネックをできるだけ解消しておくことが一つの対策にはなります。NASのネットワークポートを仮想化しネットワークのバンド帯域を太くする方法や、10Gb Ether NICを使用し、大幅にボトルネックを解消する方法もあります。しかし、1Gbのクライアント側では常に輻輳、フロー制御の問題が潜在的に存在します。この問題を解消するには新たにInfiniBand等でデータ転送をより効率化させるために使用されるプロトコルとしてRDMA等の新しい性能向上の対策が期待されます。

NAS以外のソリューションのご紹介

弊社が取り扱うSANファイル共有ソフト metaSANは先にご紹介しました様に、SANボリュームを複数のサーバからファイルレベルの共有を可能にするミドルウェアです。このmetaSANの姉妹製品でmetaLANというネットワーククライアントに使用するミドルウェアがあります。このmetaLANはネットワーククライアントがmetaSANのインストールしたSANに直接接続されたサーバをSANへのゲートウェイとして、SANのボリュームをあたかもローカルのボリュームの様にマウントが可能になります。このmetaLANには、ネットワーク上の輻輳の問題を最小限にする技術が多く取り入れられています。共有ボリュームに対するクライアントアクセスが増加した場合でも、各クライアントに対するネットワーク経由のデータアクセス性能、デー



タ転送性能に通常のCIFS、NFSベースのファイルアクセスより劣化しません。このことは、大容量コンテンツを保存し、複数のクライアントが同時にアクセスする必要がある、デジタルコンテンツの取込み、編集、エンコード、デリバリーといったワークフローにとって劣化の少ないファイルアクセスを可能にするソリューションとして有効です。