TINTRÍ

Next-generation Tintri VMstore™

TECHNICAL WHITE PAPER

Vince Guan – Architect
Ed Lee – Architect
Pratap Singh – MTS

なぜVMに特化したストレージなのか?

仮想化技術により、データセンターは大きな変化を遂げました。しかしながら、バーチャルマシン（VMs）のストレージはボトルネックを引き起こしてしまいます。VMは汎用的なハードウェアリソースの集合体上で稼働させることができ、そのCPUやメモリの使用状況を簡単にモニタリングしたり、変更したりすることができます。

ところが、従来ストレージにおいては問題が発生します。仮想化技術が登場する以前に開発された従来のストレージは、個別の設定が必要です。このミスマッチがバーチャルアプリケーションを物理的な領域へと引き戻してしまうのです。

Tintriはこれらの制限を超越したソリューションを提案します。VM専用に開発され、VMストレージの問題に注力したTintri VMstoreはその他の仮想インフラと同等レベルの管理を可能にします。仮想環境に特化して開発されたTintri VMstoreは、市場に出ている最も画期的な仮想マシン用（VMware専用）ストレージソリューションです。

主な拡張

今回の機能拡張によって、さらに導入の幅が広がりました。堅牢性を高めるデュアルコントローラをはじめ、Tintriは業界初の２つの機能を追加しました。ボトルネックの可視化機能とVMのオートアライメント機能です。これら２つの機能は、VMに特化したTintriのファイルシステムを拡張したもので、仮想環境には必ず存在する致命的な問題を解消してくれる機能です。

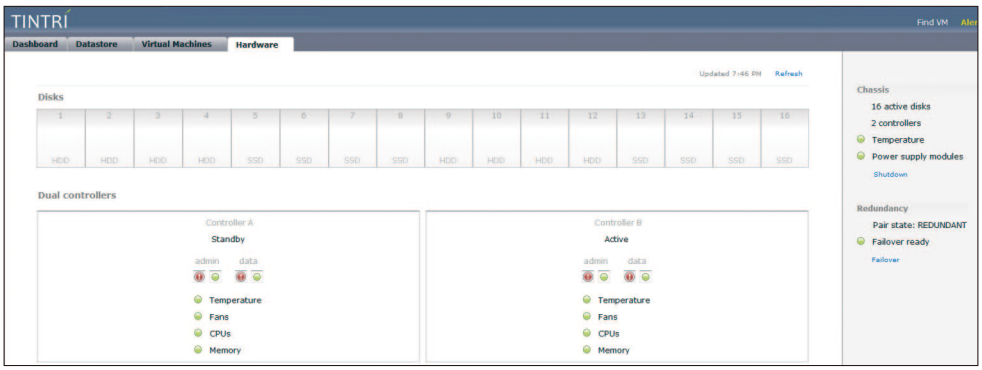
冗長用デュアルコントローラ

Tintri VMstore T540は、企業のIT環境が求める高い信頼性を提供するために冗長機能を搭載したデュアルコントローラのシステムです。この機能はシングル構成時のコントローラ障害を回避し、障害や管理イベントに影響されることなくVMstoreからのVM運用の継続を可能にします。Tintriのデュアルコントローラシステムは、シングルのコントローラシステムより高い冗長性を実現するために開発されました。

基本機能に追加されたT540のデュアルコントローラ機能は、Tintriのファイルシステムを完全に独立した２個のコンピュータシステムとして冗長性を提供します。どんな内的・外的なハードウェア障害であっても、片方のシステムが動かなくなると、もう片方のシステムが代わりに起動します。例えば、ファーストコントローラのネットワーク接続に問題が発生したり、ハードウェア障害が発生したりすると（例えば、メモリのECCエラーやディスクコントローラのエラー）、セカンドコントローラが処理を引き継ぎ、VMsへのアクセスを維持します。このハードウェア冗長により、どのようなコンポーネント障害が起きてもファイルシステムはサービスを提供し続けることができ、大幅に冗長性を高めることができるのです。

堅牢なアクティブ／スタンバイ型

T540はアクティブ／スタンバイ型を採用し、VMsは１つのコントローラからのみアクティブなサービスが提供されます。どちらがいつアクティブになっても対応できるように、データおよびステータス情報はコントローラ間で共有されます。この構造により、両コントローラが同時に同負荷をかけられることを回避し、両コントローラ上で同じソフトウェアが稼働してしまう状況を避けることができます。これにより、連動した障害を回避し、システム全体の信頼性が高まります。これに対し、アクティブ／アクティブ型はより複雑で、障害を起こしやすいフェイルオーバーと切り戻しが発生します。さらに、アクティブ／アクティブ型は潜在的かつ致命的なパフォーマンスの劇的な低下を避けるため、より頻繁な管理・運用を必要とし、セカンドコントローラの 50～70%のパフォーマンスを必要とする場合があります。



シンプルになった運用

Tintriのアクティブ／スタンバイ型の冗長機能はアップグレードや運用を容易にし、システムを止めることなくソフトウェアアップグレードを実行できるようになりました。アップグレード作業では、まずスタンバイのコントローラがアップデートされます。アップデートが完了すると、スタンバイコントローラがアクティブコントローラの処理を引き受け、同時にアクティブだったコントローラがアップデートされます。これにより、VMsへの移行作業がスムーズに行われます。運用管理者は、GUI上からどちらのコントローラをアクティブにするかを選択するだけでフェイルオーバーを実行でき、アクティブからスタンバイ（またはその逆）の処理をいつでも行うことができます。これにより、IT管理者は企業環境においても問題なく定期的なメンテナンスのスケジュールを立てられるようになります。

ボトルネックの可視化機能

管理者にとって、ストレージのパフォーマンス問題の原因を探るのはとても骨の折れる仕事です。ユーザーにVMが遅いと言われ、ストレージに問題があると推測したとしても、問題のVMが他のVMsと多数のLUNを共有し、そのLUNが複数のLUNで構成されたRAIDアレイの一部だったとしたら、どうやって問題を特定すればいいのでしょうか？ 残念ながら、従来のアレイではVM毎の統計情報を取得することはできません。場合によっては、問題はストレージにあるのではなく、ESXホストかストレージネットワーク、もしくはユーザーのアプリケーションに存在しているかもしれません。

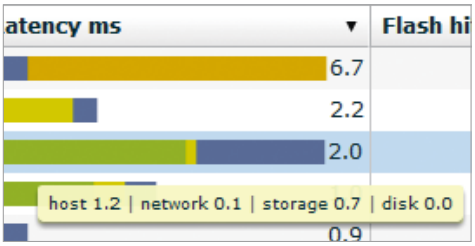
背景

パフォーマンスのボトルネックの原因を特定するには大変な時間と労力がかかり、時として結論が出ない作業です。何度も情報を収集し、分析し、仮説を立て、検証を行う必要があります。大企業においては、こういった作業は複数の社員や部署と協力しなければいけないケースがほとんどで、場合によっては作業日数も数日ではなく数週間かかることもあります。こういった作業は「責任のなすりつけあい（たらいまわし）」になってしまうこともあります。

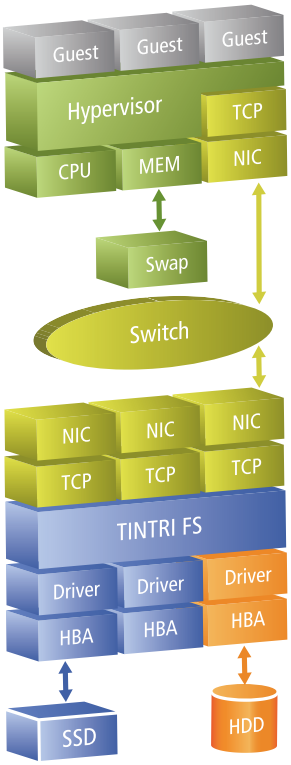
Tintriのボトルネックの可視化機能

幸い、Tintriの新機能、ボトルネックの可視化機能は問題の切り分け作業を自動的に行ってくれます。システムに格納されたVMとvDisk毎に、TintriはゲストOSからTintriマシンのディスクまでのEnd to Endのレイテンシーの詳細を表示します。どのVMやvDiskも、ESXホスト、ネットワーク、Tintriのファイルシステム、またはディスクアクセスでどれだけのレイテンシーが発生しているのか一目で確認することができます。さらに、これらの過去の履歴は自動的に保存され、グラフ化されるので、どのVMに対しても過去７日間のどの時点でボトルネックが発生したのかを見ることができます。

この可視化機能は、VM毎のハイパーバイザーのレイテンシー情報とTintri VMstoreが各VM（次頁図参照）のストレージ情報を自動的に収集して連携させたものを表示しています。ハイパーバイザーのレイテンシーは標準のvCenter APIを利用して取得されます。ネットワーク、ファイルシステム、そしてディスクのレイテンシーはTintri VMstoreによって提供され、各I/Oリクエストに対応するVMが識別されます。



VM毎の統計情報

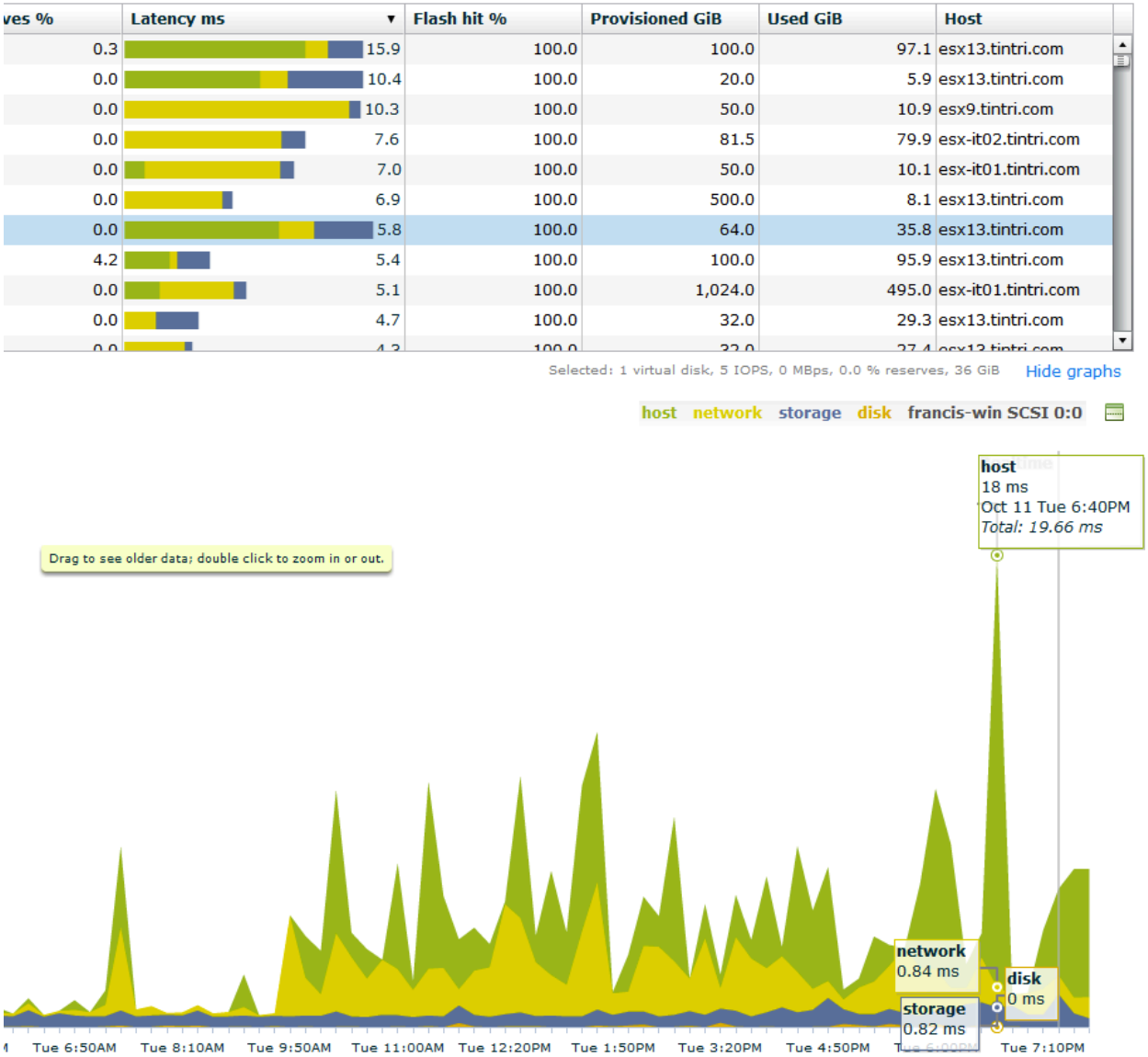


- ホスト ＝ ハイパーバイザー＋CPU Ready＋Swap Wait
- ネットワーク ＝ TCP/IP＋NIC＋スイッチ
- ストレージ ＝ Tintri FS＋ドライバ＋SSD
- ディスク ＝ ドライバ＋HDD

ホスト	ハイパーバイザーのオーバーヘッドによる遅延、低CPUリザーブによるVMsの稼働不可、低メモリによるスワッピングが含まれます。
ネットワーク	ハイパーバイザーによる遅延、TCP/IPスタックとNIC、ネットワークスイッチの遅延が含まれます。
ストレージ	リクエストを処理するTintriファイルシステム全てのオーバーヘッド、SSDアクセスのドライバとHBAオーバーヘッドが含まれます。
ディスク	HDDアクセスのドライバとHBAレイテンシー、HDDリクエスト処理時間が含まれます。

Tintriは第一世代のシステム（T445）の時から、左記に挙げたレイテンシー情報をモニタリングし、活用してきました。この情報は現場のパフォーマンス問題における修復優先順位の区分けにとても役立ちます。そのため、この情報を直接GUIに表示させることにしたのです。Tintriの初期の導入時期からこの情報を活用したところ、現場で遭遇するパフォーマンス問題の多くは、ストレージではなくハイパーバイザーもしくはネットワークで多発していることが分かりました。

ユーザーのリクエストに応え、Tintriはレイテンシーの統計情報を直感的なGUIで分かりやすく公開しています。曖昧な計測値や時間のかかる調査業務から情報を導き出そうとするより、遥かに短い時間でボトルネックを見つけることができます。



VMアライメント

VMアライメントはどのVM管理者にとっても「やることリスト」に書かれている項目ですが、実際に取り組むにはとても気の重い仕事です。主幹業務にも仮想化が導入され始めている現在となっては、さらに取り組みにくい問題です。ミスアラインされたVMsはI/Oリクエストを増大させ、ストレージアレイに余計なIOPSを発生させます。VMsが少なく、小規模な環境ではI/Oリクエストが増えてもあまり影響はありません。しかし、環境が拡大するにつれて影響は大きくなり、シングルのアレイで数百のVMsをサポートする場合にはなおさらです。パフォーマンスに与える影響は10～30%と推測されます。

背景

全てのゲストOSは物理的なチャンク（まとまり）でディスクにデータを書き込みます。ストレージアレイも物理的なチャンクまたはブロックでデータを処理しています。VMが作成されると、ゲストOSが物理的なデータブロックをディスクに書き込みますが、ゲストOSとストレージにおけるブロックの領域は必ずしも自動的にアライン（整列）されるわけではありません。

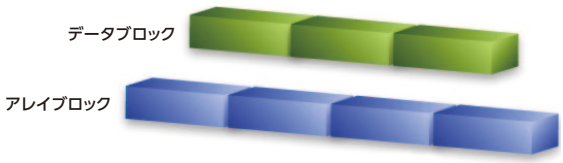


図1：ミスアラインブロック

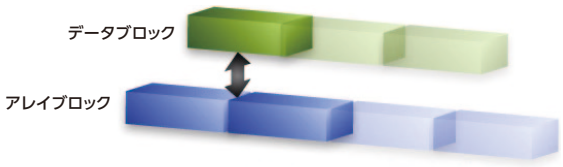


図2：ミスアラインによるオーバーヘッド

ブロックが整列されていないと、ゲストOSからのI/Oリクエストは2つのストレージブロックにまたがり、I/Oの処理が増えてしまいます。図1と図2に、不揃いなレイアウトとI/Oの影響を図解しています。

この問題は以前からある問題で、ベンダー（ヴイエムウェア、EMC、NetApp、Microsoft）や著名なブロガーの間で広く議論されてきました。インターネットで検索すると、アライメント問題については数多くのスレッド情報を見つけることができます。これらの議論は根が深く、VMアライメントが非常に深刻な問題であることが分かります。

アライメントとバーチャルマシン

仮想環境では、VMの状態はデータストア上に格納された仮想ディスクとして表示されています。このデータストアはストレージアレイの場合もあれば、各仮想サーバに接続されたローカルディスクの場合もあります。データストア上のストレージレイアウトはブロック単位で構成されています。

VMは現状を記録するために単体の、もしくは複数の仮想ディスクを作成するゲストOSを稼働させます。一般的には、ゲストOSはよくあるパーティションレイアウト（例えばマスターブートレコード（MBR））を使って各仮想ディスクのレイアウトを決定します。MBRには、サイズとロケーション情報を含んだ各仮想ディスクの細分化情報が記録されています。Windows Server 2008とWindows 7を除き、ゲストOSのファイルシステム（NTFS、EXT3など）によって形成されたブロックと、それに呼応するデータストアのブロックとレイアウトが揃うことはありません。

ゲストOSによって不揃いになったI/Oはデータストアに要求されるI/Oの量を増加させます。多数のVMsが1つのデータストアを利用しているため、増幅分が少ない場合でも不揃いなI/Oによってデータストアのリソースはいとも簡単に食いつぶされてしまいます。これは頻繁に指摘される問題ですが、これを解消しようとする管理者はあまり多くありません。例えば、Tintri社の環境を調査したところ、半分近くのVMsは不揃いのままでした。その後、すぐにアライメントを行い、パフォーマンスが30%向上しました。

VMアライメントのすすめ

では、なぜVMsは不揃いになってしまうのでしょうか？ 管理者の注意不足や、問題の認識不足が原因ではありません。様々なツールを使えばVMsを整列させ、不必要なI/Oリクエストを削減することはできます。VMsを整列しなければいけない理由とその方法を詳細に解説したブログ、ホワイトペーパー、記事などが多く存在します。

しかし、多くの管理者が認識している通り、VMのアライメントは手動で行わなければならない作業なのです。

TintriのVMオートアライメント

VMに特化したTintriのファイルシステムは、各仮想ディスクを「理解（・把握）」しています。基本的な機能を拡張させた第二世代のTintri VMstoreはVMのオートアライメント機能を搭載しています。Tintriは各ゲストを型通りに整列させる手法を取らず、動的にゲストレイアウトに適応する手法を採用しているため、ゲストOSの視点からは変更は見えません。

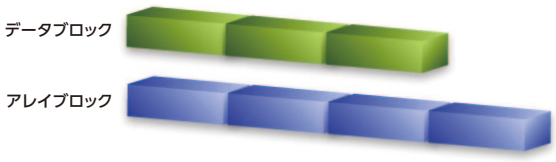


図3：アラインされたブロック

Tintri VMstoreは、VMsが移行・クローン・生成されると同時に、サービスを停止させることなく全てのVMsを自動的に整列させます。VM管理者は、これでストレージに関連する不可解な業務を完全に排除することができ、システムを停止させることなく、手動で作業をすることもなく、少なくとも10%から最大30%のパフォーマンスを向上させる楽しみを味わうことができるのです。

図4：アライメントによりオーバーヘッドが発生しない

まとめ

仮想化環境の進化において、ストレージはユーザーが一番初めに遭遇する問題です。堅牢性の高いデュアルコントローラを搭載したTintriの第二世代のシステムは、仮想環境における大きなストレージ問題を解消する画期的な新機能を搭載しました。ボトルネックの可視化機能とVMのオートアライメント機能は、TintriのVMに特化したファイルシステムを拡張させた機能です。Tintri VMstoreは、お客様のコンピュータインフラストラクチャーのさらなる仮想化を進めるため、複雑さ・パフォーマンス・コストの問題を解決してくれる革新的なソリューションです。

発売元

ノックス株式会社
www.nox.co.jp

本　　社 〒152-0023 東京都目黒区八雲2-23-13　Tel. 03-5731-5551　Fax. 03-5731-5552
西日本支社 〒533-0033 大阪市東淀川区東中島1-17-5 Tel. 06-4809-5544　Fax. 06-4809-5547

●本製品に関するお問い合わせ：営業本部
●メールでのお問い合わせ：tintri@nox.co.jp

お問い合わせ先

201112-A