# 仮想デスクトップの高拡張性、高可用性アーキテクチャのベンチマーク

Ziya Aral
Jonathan Ely
*DataCore Software*

## 要約

このレポートでは、ストレージインフラを含めたトータルのハードウェアコストをデスクトップ1台あたり32.41ドルまで削減する仮想デスクトップ（VD）の構成について記述します。32.41ドルという数字は、デュアルノード、クロスミラー、高可用性ストレージを使用した構成で達成しています。以前に発行されたレポートでは、VDI単体でのストレージインフラコストは、仮想マシン1台あたり50～数百ドルになると結論付けており、これと比較するだけでも今回の数字の持つ重要性が分かります。今回のレポートでは、ストレージハードウェアのコストは取るに足らないほどの金額になります。

さらに重要な点として、このレポートで取り上げる構成では、220台のVDをシンプルかつ低コストなサーバペアで実行しています。主な改善は、冗長ストレージ仮想システムをVDと同じハードウェアプラットフォームに配置することで、多くの仮想サーバプラットフォームと数千台ものVDに対して、高価な独立型ストレージコントローラを使用しなくて済むようにした点です。以前のレポートでは、数千台のVDを使用することで、これらの高価なコントローラのコストを相殺すると記述していました。しかし、このような構成では、規模が縮小するほどVDあたりのハードウェアコストが大きく増大します。実際は、小さい規模のVDI構成こそが実用的なのです。

一方、今回紹介する構成は、数千台のVDへと直線的に拡張することも可能なため、コストを最適化するための「スイートスポット」を模索した結果としての拡張した構成を排除できます。

最後に、この構成はVSIベンチマークを使用し、DataCoreのSANmelodyソフトウェアとMicrosoftのHyper-V仮想プラットフォームを基本としています。通常は、VDIのサイジングにはVMwareのESXプラットフォームを使用するため、このようなベンチマークでHyper-Vを使用することは珍しいと言えますが、DataCoreではESXでも同じような結果になると予測しており、そちらの結果が出たら改めて報告します。

## 問題

最近のハイパーバイザ（仮想化サーバーOS）は、ストレージエリアネットワーク（SAN）を使用して仮想マシンとホストのネットワーク用のストレージインフラを提供します。このような共有インフラは、ポータビリティ、プロビジョニング能力、およびフレキシビリティを仮想マシンに提供します。仮想マシンの数が増えると、このようなネットワークに対するニーズも比例して増えます。仮想マシンが増えるほど、ネットワーク停止の影響は1～2台には留まらず、数10台以上にもおよぶため、これらの動作を継続させるための高可用性構成が必要となるのです。仮想マシンが普及するに従って、バックアップやデータの移行といった日々の管理作業もSANのストレージ管理機能に頼るようになります。仮想マシンの数が10倍、100倍になると、VDとSANとの連携が必須になります。

VDを実装する際に問題となるのは、通常SANは大規模でコストの高いストレージコントローラと複雑な外部ストレージネットワークで実装されているという点です。これらのハードウェアは必要な拡張性を実現するには利点がありますが、VDIを実装するには非常に大規模以上でないとコストが掛かりすぎるという問題があります。この問題を払拭するため、ハードウェアベンダは通常千台から数千台のVDでベンチマークを実施しています。

仮想デスクトップはまだ登場したばかりで多くの企業は、この技術がもたらす潜在的な利点を理解しつつも、パイロットプログラムを導入したり、もしくは既存の組織にVDI実装を適合させるための試みを行ったりしているにすぎません。何しろ、パン屋に例えるなら、VDI実装の規模が数千台のVDになるとしたら、まだ味が分からないにも関わらず一斤だけ買うのではなく、店ごと買い占めるさせるようなものです。

他に良い方法も見つからないので、ユーザは最適な数のVDを稼働させる為に、非常に高価な最低コストと複雑な本格的SANを購入するしかないのです。しかも、デスクトップ1台あたりのコストは、「旧式の」個別デスクトップよりも高くついてしまいます。このことは多くの体験者がブログに書いているように、新しい「コスト節約」技術への入り口なのです。

## DataCore

このベンチマークは、DataCoreのソフトウェアベースストレージ仮想化環境を使用して、仮想デスクトップをホスティングするサーバーハイパーバイザと同じプラットフォームで実行されます。DataCoreのソフトウェアは、高可用性SANのフル機能実装であり、仮想マシンがホスティングするインフラに必要なすべての機能と、標準より高い性能特性を備えています。さらに、このソフトウェアは、このアプリケーションに推奨されるいくつかの特性を備えています。ソフトウェアには移植性があるため、Hyper-V やESXでVMとして実行したり、Hyper-Vに加えてWindows Server 2008のOSレベル（Hyper-Vの親パーティション）で実行したりできます。つまり、このソフトウェアは一般のハードウェアリソースで稼働するだけではなく、多くの最も重要なSANの相互接続性およびストレージ構造（ブロックキャッシュなど）を一般の機器上でローカルに、あるいはバーチャルに実装することができます。このような方式では、下記の「性能の考察」で述べるように、一般のハードウェアで稼働することによる「コスト」メリットが、顧客の近くに居て提案するメリットを上回っています。

さらに、ここで解説する方式は、スケーリングを透過的に行うことで全体の構成を簡素化します。新しいVDIホストが追加されるたびに、追加のストレージリソースが必要になります。また、スケーリングは拡大と縮小の両方向で可能であるため、適切なインフラ規模を設定する際の難しい問題を排除します。

ただし、これは決して、ソフトウェアベースのコントローラがハードウェアベースのコントローラに「喧嘩を売る」のとは違います。DataCoreでは、このようなハードウェアコントローラも併用しています。ユーザがこれらの機器を必要とした場合には、何も問題はありません。しかし、VDIアーキテクチャの基盤にこのような機器が必要とされない場合（コストや機能が見合わない場合）は、ハードウェアからはVDIアーキテクチャにとって足かせにしかなりません。

## ベンチマーク

使用するベンチマークは、Login ConsultantsのVirtual Session Indexer（VSI）Pro 2.1です。VSIは、このような種類のワークロードとして標準になりつつあります。
一部のストレージベンダが採用している「セルフサービス方式」のベンチマークや、トレースデータと比べると、VSIはI/O重視の傾向が強いようです（特に書き込み）。
VSIは、個別のデスクトップで予測されるより大きなワーキングセットを作成します。その理由は、VSIがエミュレートしたデスクトップ環境のすべての要素をすべてのVDで実行するためです。また、VD間のアプリケーション同期は最小限に抑えられています。

全体としては、再現性の高いベンチマーク環境を構築するためのシンプルなワークロードの標準化だと言えます。ただし、VSIは保守的すぎる面もあり、多くの独自のワークロードと比較して問題は多少残ります。

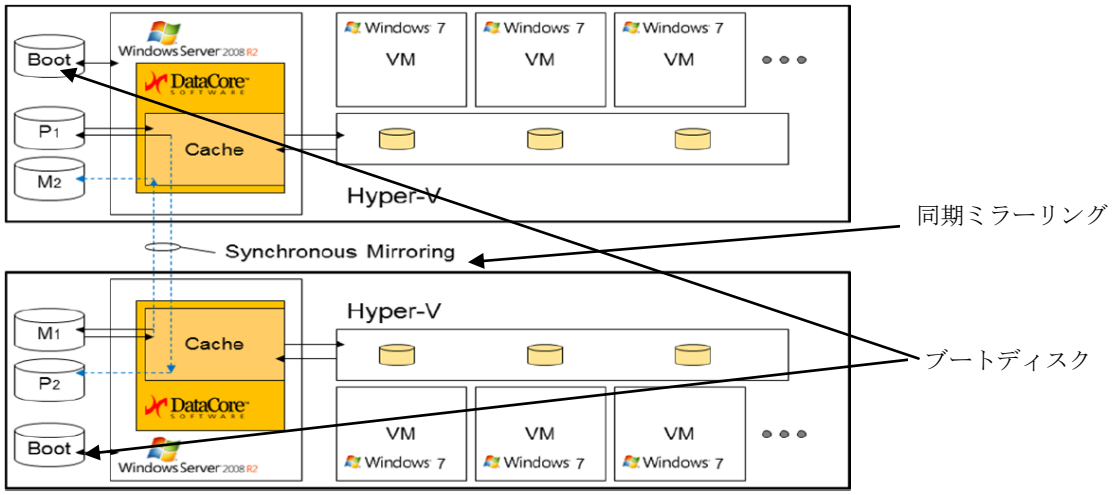
## 構成

このベンチマークで使用する構成は、2つの「ホワイトボックス」サーバノードから構成されます。それぞれのノードは、ASUS Z8PE-D18デュアルLGAマザーボード、Intel Xeon E5640 Westmere 2.66GHz Quad Core CPU×2、そして80GBのSATAブート用ドライブから構成されます。サーバ1台あたりのコスト合計は2114.14ドルで、これにはDRAMとアプリケーション用ストレージは含まれません。各ノードには64GBのDDR3 PC3-10600 DRAM（1056.24ドル）と4台のアプリケーション用ドライブ（SAMSUNG HD103SJ SATAディスクドライブ×2とWestern Digital VelociRaptor WD3000HLFS SATAディスクドライブ×2、合計で393.60ドル）が搭載されており、ハードウェアコスト合計は各サーバで3564.68ドル、合計で7,129.36ドルです。

2台のSATAドライブは、SANmelodyのストレージプールとして構成しています。この構成では、SANmelodyはOSレベルで動作します。SANmelodyがOSレベルで動作しても仮想マシンとして動作しても大きな違いはありませんが、今回はベンチマークの可視性を高めるためにOSレベルでの動作を採用しています。
構成されたストレージプールは、VDの「ゴールデンイメージ」をホスティングします。ゴールデンイメージは、DataCoreのスナップショット機能によって「スナップ」され、書き込み可能なそれぞれの「スナップ」がVDに対してブートLUNとして提供されます。複数のゴールデンイメージを作成することによって性能が向上するかどうかという実験も行われましたが、性能向上には貢献しなかったばかりか、複数のソースイメージが存在することで、DataCoreのブロックキャッシュにワーキングセットを格納するためのメモリ量が増えてしまいました。

最後に、オリジナルイメージとスナップショットを（iSCSIによるDataCoreの同期ミラーを使用して）セカンダリサーバにミラーリングすることで高可用性を実現します。この構成はストレージ障害に強いだけではなく、さらに重大な障害が発生した場合は、VDを他のサーバから再起動できるようにします。2台のSATAドライブは、他のノードからミラートラフィックを受け取るためのセカンダリストレージプールとして構成されます。

上記の構成は、いくつかのシンプルなスクリプトによって実装されます。このスクリプトは、各サーバで110台のデスクトップを起動して、構成全体をVSIベンチマークに渡します。
今は、どのクラスのコンピュータであっても、ほとんどの機能が半導体シリコン（チップ）で提供され、同じクラスのコンピュータなら性能はほぼ変わりません。この構成を選んだ理由は、「最小公倍数」での基本リファレンスを構築するためです。サーバプラットフォームの選択とコストはさまざまであり、各社の製品から最適なフットプリント、消費電力、そして特定の「エンタープライズ機能」を選ぶことができます。しかしながら、これらのバリエーションは性能には大きく影響せず、ここで選んだ基本リファレンスのコストを少し引き上げるにすぎません。



## 結果

上記の構成では、VSI 2.0.1ベンチマークを使用して220台のデスクトップをホスティングできました。詳細とベンチマーク構成は付録に記載してあります。

## 性能の考察

**共存 –** このベンチマーク結果の大きな改善点は、DataCoreのストレージ仮想化システムを、仮想デスクトップをホスティングしたのと同じハードウェアプラットフォームで使用したことです。従来の見解なら、DataCoreソフトウェアの存在が、ハードウェアプラットフォームでのリソース（メモリやCPUなど）の不足に輪を掛けたと結論付けたでしょうが、多くの実験により、それが正反対であることが証明されています。各ケースでは、このような共存構成のほうが、外部ストレージを使用した構成よりも性能が優れていたのです。

理由は簡単に説明できます。VDアプリケーションはI/Oインテンシブではなく、非常に少ないCPUサイクルでDataCoreによって簡単に処理できることが証明されています。外部チャネルトラフィックの多くを不要にしたことで、CPU要求量が減ったのです。また、ブロックキャッシュのレイテンシはほとんど無いに等しく、チャネルオーバヘッドが消え、I/Oレイテンシが無くなりました。SANmelodyを常駐させることは、その分以上の見返りをもたらし、能力やポータビリティを損なうこともありません。

メモリの観点からも同様の利点があります。VDは比較的小さいワーキングセットを使用し、読み出し重視になります。すべての読み出しは単独のソースボリュームを対象とし、それは各VDの差動システムボリュームの基本となるため、グループ全体は非常に小さいキャッシュサイズで効率よくキャッシュ処理が可能になります。キャッシュサイズを大きくすると、利点が損なわれます。

最後に、これらの構成で予測できる利点の1つは、「自己調整方式」であるということです。VDのグループを追加すると、それらをサポートするためのストレージインフラも追加されます。一般的なSANは、要件が開放的で、独立して編成され、複雑なストレージネットワークとして構築されることが多いのですが、VDIアプリケーションの場合は要件が分かっているために、ストレージインフラを透過的にすることができます。

**ディスク –** 書き込みキャッシュ処理を効率的に行うことにより、要求されるディスク性能を大幅に抑えることができます。このベンチマークでは4台のSATAドライブを使用し、2つの独立したDataCore 2ディスクプールとして構成しました。1つめのプールはプライマリストレージ、2つめのプールは代替ノードのミラーリングに使用されています。7,200 rpmと10,000 rpmのドライブが混在していますが、Samsungs製品と同じような特性を持つ7,200 rpmのSATAドライブを4台（あるいは3台）使用した場合でも、結果は同じようになるでしょう。

この違いの重要性については、ベースライン構成の構築に関する部分で説明します。このベンチマークでは、簡単に構成できる標準SATA機器に限定しました。ファイバ チャネルスピンドル、高速デバイス、ハイブリッドディスク、そしてSSDは、実際の世界ではそれぞれに活躍の場があり、一般的なSATA機器の数倍の性能をもたらすことが知られています。それでも、これらの機器をベースラインベンチマークに組み込んだとしても、他のストレージアレイを中心に構築したVDIと同じ程度の効果しか得られないでしょう。このような機器を組み込んだ最適化は単に複雑さを増すだけで、本末転倒だと言えます。比較が難しくなるだけでなく、VDI要件自体が最も重要な領域に、特定のハードウェアアーキテクチャが割り込むことになります。

**メモリ –** DRAMのコストは、この構成のトータルプラットフォームコストの大きな部分を占めます。各VDのメモリサイズを小さくする場合は、各プラットフォームでサポートされるVDの数も減らす必要があります。難しい点は、この業界ではDRAMの集積度とコスト構造が常に変化しているということです。そのため、DRAMの最適化は回避したいところです。

この問題をさらに難しくするのは、MicrosoftとVMwareの両社が、各仮想マシンに割り当てられたメモリの「オーバロード」を可能にするハイパーバイザ機能を提供していることです。この機能により、「マシン」間でグローバル仮想メモリが構築されます（具体的な実装メカニズムは製品によって異なります）。

最後に、DataCoreのVSIベンチマークの途中で発見されたことですが、メモリ割り当て量が非常に少ないVD（単体のPCですら起動できないほどの量）でさえ、この構成では正常に実行できました。調べたところ、DataCoreブロックキャッシュの性能によってページング階層が構築され、それによって「スラッシング」を防止したわけではありませんが、ワーキングセットサイズが一致したときにはスラッシングを吸収する効果があったのです。

最終的には、ここで解説した実装では「オーバロード」も「高速スラッシング」も使用しませんでした。その理由は、すでに公開済みのベンチマークデータとの公正な比較を行うためと、独自性が強く、VDIのサイジングに依存しない方式による最適化を避けるためです。つまり、ここで使用した従来の割り当て方式は「抜け道」を避けたということです。

実際に使用したDataCoreキャッシュサイズは4GBでした。キャッシュサイズを大きくすれば結果は多少は良くなりますが、「収穫逓減」のポイントが4GBでした。Windows Server 2008でこのメモリを差し引いた残りのDRAMは、従来の固定割り当て方式によって常駐するVD数で単純に分割されました。

**ソフトウェアコスト –** DataCoreはハードウェアコントローラの代わりにソフトウェアベースのストレージ仮想化システムを使用したのだから、ソフトウェアのコストも含めなければ「同じ土俵での」比較にはならない、という主張もあるでしょう。しかしながら、それにはそれで問題があります。すでにベンチマークされたハードウェアコントローラのほとんどは最小限の構成のものばかりです。SANmelodyの機能セットの大半は、レポートで使用された構成では利用できないか、または利用するためには追加コストが必要になります。

その点を考慮した上で、ソフトウェアのトータルコストは1ノードあたり3848ドル、VD 1台あた34.98でした。このコストには、SANmelody環境とHyper-VベースのVDをホスティングしたMicrosoft Server 2008オペレーティングシステムと、DataCoreのSANmelodyも含まれています。

インフラソフトウェアおよびハードウェアを含め、OSやVDアプリケーションを除いた、ベンチマーク構成の完全なコストは、デスクトップ1台あたり67.39ドルになります。

**スケーリング –** すでに述べたように、本当に難しいのは、VDを「数千台」まで拡張することよりも、むしろ実用的な規模までスケールダウンすることです。言うまでも無く、ローエンドでコストが大幅に増えたり、あるいは移植性、可用性、およびデータ冗長性を保証してくれるSANの特性が損なわれたりすることなく、実用的な構成が実現できなければなりません。そうでなければ、VDIがもたらす恩恵が意味を持たなくなってしまうからです。ここで述べるDataCore構成は、VDとスレーブインフラストラクチャを同時にホスティングするペアサーバアプローチによって、220台のデスクトップでインフラストラクチャコストを上手に抑えてくれます。

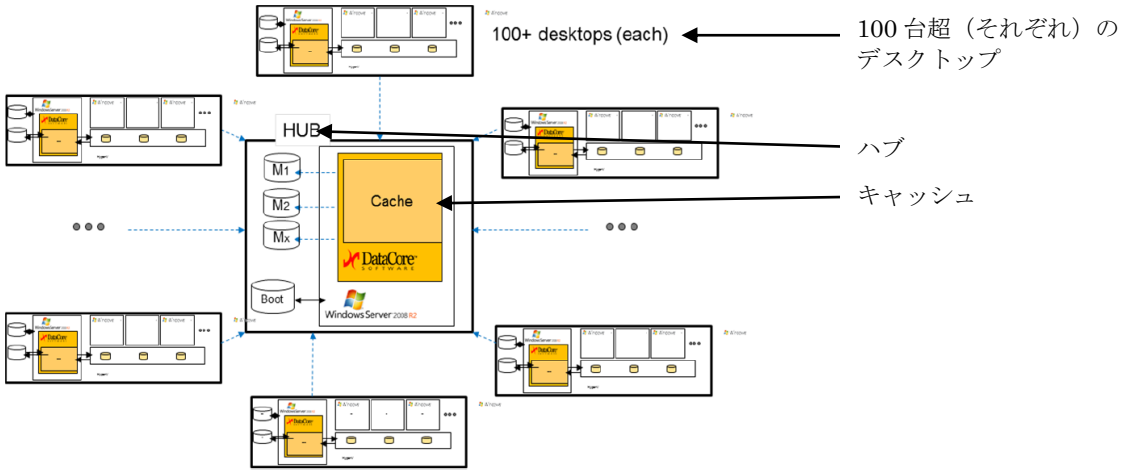
では、拡張についてはどうでしょうか。「数千台」ものデスクトップまでスケールアップする必要がある場合には、どのようにすればよいのでしょうか。そしてさらに重要なのは、コストが大幅に変動したり、構成の再設定が非常に複雑になったりせずにこのようなスケールアップが可能だろうか、という点です。

大規模な構成を実現するには、「スター」と呼ばれるDataCoreトポロジを使用します。スタートポロジでは、3番目のノードを追加して、スターセンター（ハブ）を構成します。そして、その他のノードをスターに追加します。スターの各ノードは以前と同様に機能しますが、それぞれは自身の内容を（お互いにではなく）ハブノードにミラーリングします。このようにして、ハブが構成全体

を1点で管理するようになります。

この方式では、スターにポイントとして追加される各サーバは、自身のストレージ装置とインフラ、そしてVDI環境を持ち込むことになります。この方式は、ハブがセカンダリミラートラフィックを受け入れられなくなるまで、自己調整方式で拡張が可能です。もう1つの利点として、VDIの欠点であるマイナスの現象（ブートストームなど）が、非常に小さく管理可能なモジュールに押さえ込まれ、トポロジ全体に影響することはありません。

シミュレーションと初期実験では、シンプルなスタートポロジによって、性能（およびVDあたりのコスト）を数千台のデスクトップまで拡張できることが示されました。DataCoreでは、これらの発見点を支持する結果が準備できしだい公開する予定です。




## 結論と今後の方向性

このレポートでは。DataCoreのSANmelody、MicrosoftのServer 2008/Hyper-V、そして標準的なサーバハードウェアをベースとしたVDIのアーキテクチャについて記述しました。ベンチマークツールとしては、Login ConsultantのVirtual Session Indexer（VSI）を使用しました。

全体として、今回のベンチマーク対象となったアーキテクチャの結果は、他のストレージベンダが公開している結果と異なっています。VDI構成をストレージコントローラや既存のSANアーキテクチャを中心に構築するのではなく、VDIサーバ自身を中心としてSANを構築しました。特に、VDIサーバは仮想サーバとしても機能します。その結果として生じるローカリティにより、他社が公開している結果と比べて10倍もの価格対性能比を達成しつつ、VD数を大きく削減して（220台）、この経済性を実現しています。アーキテクチャ全体は、最も単純な「最小公倍数」のハードウェアによって構成されています。そでもなお、環境自体は「自己調整方式」であり、規模が拡張されても同じような評価を維持できるうえ、ハードウェア障害、再構成、および「ブートストーム」などのVDIの異常にも強いモジュール式となっています。

今後の方向性はどうなるのでしょうか。

いくつかの最適化がすぐに思い浮かびます。DataCoreのラボでは、すでに同じサーバプラットフォームでVDの密度を50%以上も高め、VD数として175台を増やすことに成功しています。この結果は、メモリの調整、ハードウェアの強化、および各種のチューニング「トリック」によって達成さ

れています。ただし、コストや複雑さが増大するため、価格対性能比については、このレポートで紹介した結果からほとんど向上してません。また残念なことに、古い機器を利用した「オールドチューナー」の場合は、実用的なVDI構成の限界を超えるには今回の構成で十分に最適化しきれており、低レベルの最適化を加えても、ごくわずかな増分的な変化しかもたらされない、という結論に達しています。

ハイレベルの最適化はどうでしょうか。ホストOS、ハイパーバイザ、および仮想化サーバを（VDと一緒に）共存させれば、統合が強化されるかも知れないという考えがあります。しかしながら、現時点では独立性の高い構成のほうが「最適化」の観点では優位性に勝るという事実があります。

現在のコンピューティング業界の大きな流れの1つであるクラウドコンピューティングでは、VDIが1つの要素となっています。どちらの技術も、非常に多くの同じクラスの「マシン」を集めるため、リソース要件が同様で、事前に予測可能な多くのプラットフォームが約束されます。つまり異種環境で要件が未知なのではなく、環境が同種で要件が既知になるということです。コンピュータの世界では、「既知」は事前に構成可能であることを意味します。

この結果、さらに上のレベルの仮想化と、VDIに取り組むとすぐに直面する事実、つまりどれも同じような大量のコンポーネントマシンを個別に管理しなければならないという問題の「分割統治」が可能になります。
数百、数千もの仮想マシンを個別に管理するのではなく、これらを任意のグループに分けることを考えてみましょう。このサブユニットは「仮想データセンタ」と呼ばれ、DataCoreでは「ハイブ」と呼んでいます。それぞれのハイブ全体を1つのユニットとして扱えたらどうなるでしょうか。この場合の「ハイブ」は、VDのグループ（50～70台）から構成され、DataCoreの仮想化サーバとローカルサービスを提供するための他のサーバにより、すべて仮想マシンとして実装されて、互いに対話できるように構成されます。しかも、基礎となるハードウェア環境についての知識は不要です。構築されたハイブ全体は、外界への統一「ポート」として機能する仮想化サーバによって1つのユニットとして管理、移動、および制御が可能です。環境を管理する際には、「2つのハイブをここ、4つをあそこに配置して・・・」と言うように、容量をおおまかに見積もるだけで済みます。

以上

## 付録

ベンチマークテストと構成

この構成における仮想デスクトップの最大数を決定するために使用した性能ベンチマークツールは、Login ConsultantsのVirtual Session Indexer（VSI）rev 2.1.2です。Login VSIは、各デスクトップでユーザのワークロード（Outlook、Internet Explorer、Excelなど）をシミュレートします。各デスクトップは、共有フォルダに性能情報を出力し、ベンチマークの完了後にこの情報を解析します。

シミュレートされたユーザワークロードは、VDに対してリモートデスクトップを開くことで開始されます。ここではVSI Launcherプログラムをデフォルト構成で使用してセッションを開き、60秒ごとに1台のデスクトップを起動します。ユーザワークロードは、すべてのデスクトップセッションが開いて最後のセッションがループを完了するまで、連続した12分間のループで実行されます。

すべてのVDがシミュレーションユーザワークロードを同時に実行して共有フォルダにデータを出力すると、VSI Analysisツールが実行されます。このツールは、データをExcelで集計して、ユーザエクスペリエンスの満足度を表す応答時間をグラフとテーブルで示します。

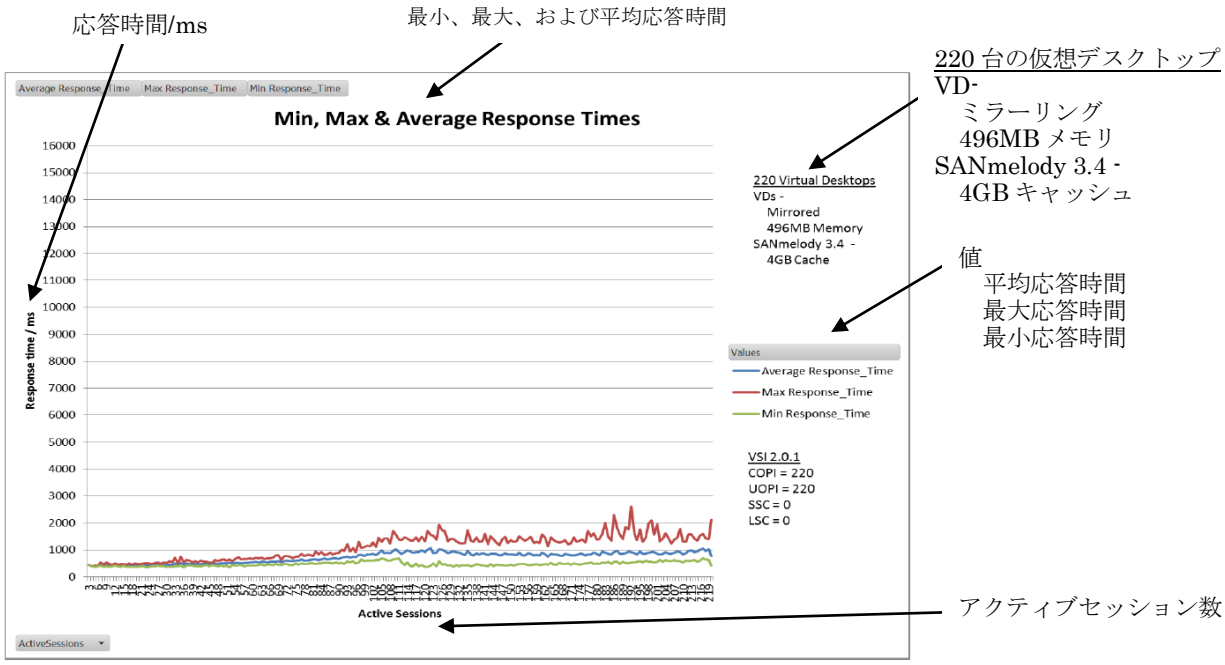
構成情報:

仮想デスクトップ -

Microsoft Windows 7 Enterprise x86 – デフォルトのVSI最適化と、Virtual Reality Checkのレポート『Project VRC: Phase III』（http://www.projectvrc.nl/）で推奨される最適化で構成

ベンチマーク –

Login Consultants VSI 2.1.2（http://www.loginconsultants.com/）
VSIの推奨に従い、7つのVSI Launcherプログラムを使用しました。すべてのLauncherプログラムは、VDが動作するホストとは独立したHyper-Vホストで仮想マシンとして動作します。これらの仮想マシンのOSはMicrosoft Server 2008 R2です。




220台の仮想デスクトップが正常に動作していることを示すVSIベンチマーク解析