

ホワイトペーパー

Fujitsu PRIMERGY サーバ

RAID コントローラーのパフォーマンス 2013

本書は、Fujitsu PRIMERGY サーバのディスク I/O パフォーマンスの担当者を対象としています。内蔵ディスクサブシステムに適用する各種 RAID コントローラーのオプションやアプリケーション分野について、パフォーマンスの観点から理解するための情報を提供しています。推奨するコントローラを選択およびパラメータ設定は、データの安全性やパフォーマンスに対する要件、およびサーバ構成により異なります。本書では、2013 年の PRIMERGY システムに搭載可能なコントローラーを紹介します。

バージョン

2.0a

2013-08-01



目次

ドキュメントの履歴	2
はじめに	3
PRIMERGY 用 RAID コントローラー : 基本	4
RAID コントローラーの概要	4
コントローラーインターフェースとそのスループットの限界	7
電源障害に対するコントローラーキャッシュの保護	10
パフォーマンス関連の設定	11
MegaRAID アドバンスドソフトウェアオプション	13
オンボードコントローラーの特性	14
測定内容	15
測定方法	15
測定環境	16
コントローラーの比較	18
RAID 1 (SATA ハードディスク 2 台)	19
RAID 1 (SAS ハードディスク 2 台)	21
RAID 0 および 10 (SATA ハードディスク 4 台で構成)	24
RAID 0 および 10 (ハードディスク 5~8 台で構成)	27
RAID 0 および 10 (9 台以上のハードディスクで構成)	35
RAID 5	42
低負荷レベル	46
MegaRAID アドバンスドソフトウェアオプション	47
CacheCade 2.0	47
FastPath	50
結論	52
関連資料	54
お問い合わせ先	54

ドキュメントの履歴

バージョン 1.0a

バージョン 2.0

- ドキュメントのタイトルを 2013 年用に変更
- PCIe Gen3 をサポートする 2013 年に導入されたコントローラーの追加
- MegaRAID アドバンスドソフトウェアオプションの追加
- キャッシュ設定の推奨事項の変更にに基づき、RAID 1 構成でコントローラーを比較した場合の測定結果を更新
- マイナー修正

バージョン 2.0a

- マイナー修正

はじめに

ハードディスクは、サーバ環境においてセキュリティ上の要素であると同時に、パフォーマンスを左右する重要なコンポーネントでもあります。そのため、ハードディスクがシステムのボトルネックとならないように、インテリジェントな方法で複数のハードディスクのパフォーマンスを統括することが重要です。また同時に、特定のコンポーネントで障害が発生した場合、他のコンポーネントで補う必要があります。複数台のハードディスクを組み合わせることで、いずれかのハードディスクが故障したときに、他のハードディスクでそれを補うという方法があります。これは RAID (Redundant Array of Independent Disks : 独立ディスク冗長配列) と呼ばれる方法です。RAID の構成には、通常、特別な RAID コントローラーを使用します。

PRIMERGY サーバは、種々の RAID コントローラーとハードディスク構成によるさまざまな内部構成で利用できます。PRIMERGY ファミリーのすべてのサーバに標準で提供される「モジュラー RAID」コンセプトは、RAID コントローラーファミリーと富士通製 RAID Manager ソフトウェア「ServerView RAID Manager」で構成されています。幅広い RAID ソリューションが用意されているため、ユーザーは特定のアプリケーションシナリオに合った適切なコントローラーを選択できます。ディスクサブシステムのパフォーマンスは、コントローラー、選択したハードディスク、および RAID レベルの機能によって決まります。

これまでの PRIMERGY ホワイトペーパーシリーズでは、「モジュラー RAID」のパフォーマンスに関して、次のようにあらゆる側面を取り上げてきました。

- ディスク I/O パフォーマンスの概要については、『[ディスク I/O パフォーマンスの基本](#)』を参照してください。
- 現在 PRIMERGY に対応している各種ハードディスクと、さまざまな状況でのそれぞれのパフォーマンスについては、『[単一ディスクのパフォーマンス](#)』を参照してください。
- 現在 PRIMERGY に対応しているすべての RAID コントローラーとそれぞれのパフォーマンスについては、本書『[RAID コントローラーのパフォーマンス 2013](#)』を参照してください。

PRIMERGY サーバの内蔵ディスクサブシステムをサイジングする際は、最初に適切なハードディスクタイプを選択し、次に、所要の RAID レベルを確立するために必要なハードディスクの数を経験則から見積もります。接続するハードディスクの数およびテクノロジーと、所要の RAID レベルにより、RAID コントローラーが決まります。ディスクサブシステムを的確にサイジングするには、数年かかるかもしれません。

しかし、SSD (Solid State Drives : ソリッドステートディスク) などのストレージメディアやサーバの内部インターフェースは日々技術的に進化しており、それにより増大した要件が、サイジングしたディスクサブシステムで満たされなくなります。あるいは、実環境のサーバ構成では、アプリケーションシナリオの変化により、ハードディスクの数は十分でも期待どおりのディスク I/O パフォーマンスが得られないこともあります。このような場合、RAID コントローラーがパフォーマンスに与える影響をさらに詳しく検証すると効果的です。適切なコントローラーを選択すること、あるいは単にコントローラーを正しく構成することが、最高のパフォーマンスを得るための必須条件になることがあります。

本書の目的は、上記の課題を解決することです。最初に、PRIMERGY システムに搭載可能な内蔵 RAID コントローラーの概要を説明します。次に、関連するコントローラーインターフェースのスループットの限界について、パフォーマンスの観点から説明します。さらに、測定の背景を簡単に説明した後、さまざまな RAID コントローラーを各 RAID レベルおよび異なるアプリケーションシナリオで測定し、測定結果を元に比較します。

従来、「ハードディスク」や「ハードディスクドライブ (HDD)」という用語は、直接アドレス指定できる、硬質で磁気コーティングされた、回転式のデジタル非揮発性ストレージメディアを指していました。現在では技術の進歩により、新しい形の「ハードディスク」がストレージメディアとして登場しています。これらはサーバに対して従来と同じインターフェースを持ち、サーバからはハードディスクと同様に扱われます。代表的な例が SSD です。これは可動部分のない電子ストレージメディアですが、通常、ハードディスクとして扱われます。本書では、総称として「ハードディスク」という用語を使用し、区別する場合に「SSD」や「HDD」という用語を使用します。

本書では、ハードディスクの容量を示す場合は 10 のべき乗 ($1 \text{ TB} = 10^{12}$ バイト)、その他の容量やファイルサイズ、ブロックサイズ、スループットを示す場合は 2 のべき乗 ($1 \text{ MB/s} = 2^{20}$ バイト/s) で表記しています。

PRIMERGY 用 RAID コントローラー : 基本

ここでは、まず PRIMERGY サーバに搭載可能な RAID コントローラーとそれらの基本的な機能について説明します。さらに、各コントローラーとサーバ内インターフェースの組み合わせにより得られるスループットの限界について詳しく見ていきます。次に、コントローラーで選択できる設定について解説し、最後にオンボードコントローラーの特性について述べます。

RAID コントローラーの概要

次の表は、利用可能な RAID コントローラーの機能性に関する最も重要なデータをまとめたものです。

本書では、コントローラーの名称を簡略化するため、表の「略称」列に記載した略称を使用します（例：LSI2008）。これらの略称はコントローラーの基本チップ名に基づいており、必要に応じてキャッシュサイズも併記しています。

コントローラー名	略称	FF	キャッシュ	周波数	対応インターフェース		最大ディスク数	RAID レベル	BBU / FBU
Intel C200 (オンボード SATA) 上の LSI SW RAID	Cougar Point	I	-	3G	SATA 3G		4	JBOD、0、1、10	-/-
Intel C600 (オンボード SATA) 上の LSI SW RAID	Patsburg A	I	-	3G	SATA 3G		4	0、1、10	-/-
Intel C600 (オンボード SAS) 上の LSI SW RAID	Patsburg B	I	-	3G	SATA 3G SAS 3G		4	0、1、10	-/-
RAID 0/1 SAS based on LSI MegaRAID 4Port	LSI1064	P	-	3G	SATA 3G SAS 3G	PCIe 1.0 x4	4	0、1、1E	-/-
RAID 0/1 SAS based on LSI MegaRAID 8Port	LSI1068	P	-	3G	SATA 3G SAS 3G	PCIe 1.0 x4	8	0、1、1E	-/-
RAID Ctrl SAS 6G 0/1 (D2607)	LSI2008	P	-	6G	SATA 3G/6G SAS 3G/6G	PCIe 2.0 x8	8	0、1、1E、10	-/-
RAID Ctrl SAS 6G 5/6 512 MB (D2616)	LSI2108	P	512 MB	6G	SATA 3G/6G SAS 3G/6G	PCIe 2.0 x8	8	0、1、5、6、10、50、60	✓/-
PY SAS RAID Mezz Card 6Gb		M							
RAID Ctrl SAS 6G 5/6 1GB (D3116)	LSI2208-1G-2.0	P	1 GB	6G	SATA 3G/6G SAS 3G/6G	PCIe 2.0 x8	8	0、1、1E、5、6、10、50、60	-/✓
RAID Ctrl SAS 6G 5/6 1GB (D3116C)	LSI2208-1G-3.0	P	1 GB	6G	SATA 3G/6G SAS 3G/6G	PCIe 3.0 x8	8	0、1、1E、5、6、10、50、60	-/✓
PY SAS RAID HDD Module ¹⁾²⁾	LSI2208-512-2.0 ¹⁾²⁾	M	512 MB	6G	SATA 3G/6G SAS 3G/6G	PCIe 2.0 x8	8	0、1、1E、5、6、10、50、60 ²⁾	-/✓
PY SAS RAID HDD Module 3.0 ¹⁾²⁾	LSI2208-512-3.0 ¹⁾²⁾	M	512 MB	6G	SATA 3G/6G SAS 3G/6G	PCIe 3.0 x8	8	0、1、1E、5、6、10、50、60 ²⁾	-/✓
PY SAS RAID HDD Module w/o cache ¹⁾	LSI2208-Lite-2.0 ¹⁾	M	-	6G	SATA 3G/6G SAS 3G/6G	PCIe 2.0 x8	8	0、1	-/-
PY SAS RAID HDD Module w/o cache 3.0 ¹⁾	LSI2208-Lite-3.0 ¹⁾	M	-	6G	SATA 3G/6G SAS 3G/6G	PCIe 3.0 x8	8	0、1	-/-

¹⁾ これらのコントローラーは PRIMERGY BX920 S3 にのみ搭載可能です。

²⁾ PRIMERGY SX980 ストレージブレードなどより多くのディスクスロットを持つ機器を接続する場合、LSI2208-512 コントローラーは RAID 0 および RAID 1 だけでなくすべての RAID レベルを使用できます。

列「FF」はフォームファクターを表し、「I」は「統合型」、「P」は「PCIe スロット」、「M」は「メザニンカード」を指します。「最大 ディスク数」列は、PRIMERGY サーバの RAID 管理コンセプトに基づき、コントローラーで直接操作できるハードディスクの最大数を示します。この情報により、コントローラーが理論的なボトルネックになるかどうか判断できます。一部の PRIMERGY モデルでは、特定のコントローラーモデルに「エクスパンダー」（SAS 規格で定義されている専用コンポーネント）を接続することで、ハードディスクの最大数をさらに増やすことができます。この場合、エクスパンダーは既存のポートの帯域幅を拡大することはできませんが、接続されているすべてのハードディスクでその帯域幅を利用できるようにします。

「RAID 0/1」 SAS-RAID コントローラーでは、LSI2008 コントローラーが LSI1068 コントローラーの 6G 後継となり、LSI2208 コントローラーが LSI2108 コントローラーの 6G 後継になります。いずれも「RAID 5/6」 SAS-RAID コントローラーファミリーの製品です。

LSI2208 コントローラーには、キャッシュサイズの違いにより、略称 LSI2208-1G および LSI2208-512 の 2 モデルがあります。以降で LSI2208 とだけ記載している場合は、両方のキャッシュサイズを指します。

2013 年には、PCIe Gen3 をサポートする新しいチップが、3 つの LSI2208 コントローラーのグループ向けに導入されました。本書では、略称の末尾に 2.0 および 3.0 を付けて区別しています。略称の末尾にこれらの数字がない場合は、コントローラーの両方のバージョンを指します。

ディスクサブシステムのパフォーマンスの評価では、プロセッサパフォーマンスとメモリ構成は、今日のシステムでは、ほとんどの場合大きな要因ではありません。通常、考えられるボトルネックはサーバシステムの CPU やメモリにではなく、ハードディスクや RAID コントローラーに影響を及ぼします。したがって、ハードディスクの拡張性の違いのためすべての PRIMERGY ですべての構成を実現できない場合でも、使用する PRIMERGY のモデルに関係なく、さまざまな RAID コントローラーを比較できます。

次の表は、本書の作成時に各 PRIMERGY システムでリリースしている、または近い将来リリースされるハードディスク接続用の RAID コントローラー（過去にリリースされたものも含む）と、各モデルでサポートする最大ハードディスク数をまとめたものです。PRIMERGY 構成バージョンとコントローラーの可能な組み合わせについては、システムのコンフィギュレータを参照してください。

PRIMERGY	最大 ディスク数	エクスパンダー	オンボードコントローラー			PCIe インターフェースを搭載したコントローラー						
			CougarPoint	Patsburg A	Patsburg B	LSI1064	LSI1068	LSI2008	LSI2108	LSI2208-1GB	LSI2208-512	LSI2208-Lite
BX920 S3	2			✓	✓						✓	✓
BX924 S3	2			✓	✓				✓			
CX250 S1	8	-/✓		✓					✓			
CX270 S1	24	-/✓		✓				✓	✓			
RX100 S7	4		✓			✓		✓	✓			
RX100 S7p	4		✓			✓		✓	✓	✓		
RX200 S7	8			✓	✓			✓	✓	✓		
RX300 S7	16	-/✓		✓	✓			✓	✓	✓		
RX350 S7	24	-/✓		✓	✓			✓	✓	✓		
RX500 S7	8			✓	✓			✓		✓		
RX600 S6	8							✓	✓	✓		
RX900 S2	8								✓			
SX940 S1	4					✓		✓	✓			
SX960 S1	10	✓							✓	✓		
TX100 S3	4		✓						✓			
TX100 S3p	4		✓						✓	✓		
TX120 S3	4		✓			✓		✓	✓			
TX120 S3p	4		✓			✓		✓	✓	✓		
TX140 S1	8		✓				✓	✓	✓			
TX140 S1p	8		✓				✓	✓	✓	✓		
TX300 S7	24	-/✓		✓	✓			✓	✓	✓		

LSI2108 コントローラーは、サーバブレード PRIMERGY BX924 S3 のメザニンカードとして使用できます。サーバブレード PRIMERGY BX920 S3 の場合は、LSI2208 コントローラーをメザニンカードとして使用できます。

本書では、同サーバブレードの内蔵ハードディスクに接続された上記メザニンカードについてのみ調査します。SAS スイッチによってこれらのメザニンカードに接続される PRIMERGY SX980 ストレージブレードのパフォーマンスについては、PRIMERGY SX980 のパフォーマンスレポートで取り上げています。

コントローラーインターフェースとそのスループットの限界

RAID コントローラーには、ハードディスクに接続するインターフェースと、マザーボード上のチップセットに接続するインターフェースが必要です。ハードディスクとのインターフェースは、一般に SAS または SATA です。チップセットとのインターフェースは一般に PCIe ですが、統合型オンボードコントローラーの場合は、マザーボードの内部バスインターフェースを使用します。SAS、SATA、PCIe のスループットの上限は次のとおりです。

SAS および SATA

「Serial Attached SCSI」 (SAS) および「Serial Advanced Technology Attachment」 (SATA) はシリアルインターフェースであり、データスループットは周波数に依存します。これらのインターフェースは、ハードディスク、光ドライブ、テープドライブなどの非揮発性ストレージメディアを接続するのに使用されます。

タイプ	周波数	理論スループット	実効スループット (85 %)
SAS 3G / SATA 3G	3000 MHz	286 MB/s	243 MB/s
SAS 6G / SATA 6G	6000 MHz	572 MB/s	486 MB/s

周波数は、コントローラーやハードディスクの名前に含まれる 3G または 6G という略語で確認できます。または、SAS のバージョン番号で確認できます。バージョン番号が 1.0 であれば 3G、2.0 であれば 6G です。

理論的に達成可能なスループットは、1 Hz あたり 1 ビットから、いわゆる 8b/10b コーディングによるシリアル転送の冗長性 20 % を引いて計算します。実際に達成可能なスループットは、この値に 0.85 を掛けることで見積もれます。この 85 % という値は、長年観測してきたさまざまなコンポーネントの値から算出された平均経験値です。

端末間を接続するすべてのコンポーネントは、同じバージョンの SAS または SATA プロトコルを使用する必要があります。このコンポーネントには、ハードディスクのほかに、使用する可能性のあるコントローラーおよびエクспанダーも含まれます。バージョンが異なるコンポーネントを混在させると、すべてのコンポーネントでサポートされる最も高いパフォーマンス基準が自動的に選択されるため、周波数が低くなる可能性があります。この点では、上位のプロトコルには下位互換性があります。

SATA 対応の各ポートは、通常それぞれハードディスクに接続されます。一方 SAS の場合は、SAS ケーブル 4 本をひとまとめにすることが多く、「x4 SAS」または「x4 ワイドポート」と呼ばれます。これにより、バックプレーン経由で最大 4 台の SAS ハードディスクを直接接続できます。x4 SAS のスループットは SAS 接続を個別に使用した場合の 4 倍になります。これは SATA の場合でも同様です。

インターフェース	接続	周波数	理論スループット	実効スループット (85 %)
SAS 3G / SATA 3G	x4 × 1	3000 MHz	1144 MB/s	973 MB/s
SAS 3G / SATA 3G	x4 × 2	3000 MHz	2289 MB/s	1945 MB/s
SAS 6G / SATA 6G	x4 × 1	6000 MHz	2289 MB/s	1945 MB/s
SAS 6G / SATA 6G	x4 × 2	6000 MHz	4578 MB/s	3890 MB/s

一部の PRIMERGY モデルは、コントローラーのハードディスク接続数より多くのハードディスクを接続できます。この場合、接続可能なハードディスクの数は、エクспанダーを使用して拡張します。すでに述べたように、エクспанダーはデータフローを分散するだけで、スループットを増加させるわけではありません。

SAS プロトコルは、周波数が同じか、より低い SATA プロトコルも転送（トンネリング）できるように定義されています。これにより、どちらのバージョンの SAS コントローラーでも、SATA ハードディスクと通信できます。逆に、SATA インターフェース経由で SAS ハードディスクを接続することはできません。

PCIe

PCIe は、コントローラーとマザーボード間のシリアルインターフェースでもあります。コネクタには、数種類の帯域（レーン数）があります。通常は x4（4 レーン）と x8（8 レーン）ですが、重要なのは電氣的に使用される実際のレーンの数（以降、「PCIe 有効帯域」と呼びます）です。レーンのスループットは周波数によっても変化します。

インターフェース	接続	周波数	理論スループット	実効スループット (90 %)
PCIe.1.0、PCIe Gen1	x4	2500 MHz	954 MB/s	858 MB/s
PCIe.1.0、PCIe Gen1	x8	2500 MHz	1907 MB/s	1716 MB/s
PCIe.2.0、PCIe Gen2	x4	5000 MHz	1907 MB/s	1716 MB/s
PCIe.2.0、PCIe Gen2	x8	5000 MHz	3815 MB/s	3433 MB/s
PCIe 3.0、PCIe Gen3	x4	8000 MHz	3756 MB/s	3380 MB/s
PCIe 3.0、PCIe Gen3	x8	8000 MHz	7512 MB/s	6761 MB/s

PCIe 1.0 は「PCIe Gen1」、PCIe 2.0 は「PCIe Gen2」、PCIe 3.0 は「PCIe Gen3」とも表記します。

理論的に達成可能なスループットは、PCIe 1.0 および 2.0 の場合、1 Hz あたり 1 ビットに接続数（x4 または x8）を掛けたものから、いわゆる 8b/10b コーディングによるシリアル転送の冗長性 20 %を引いて計算します。PCIe 3.0 の場合、128b/130b コーディングによる冗長性 1.54 %を引きます。実際に達成可能なスループットは、この値に 0.90 を掛けることで見積もれます。この 90 %という値は、長年測定してきたさまざまなコンポーネントの値から算出した、経験的な平均値です。

2010 年の導入世代以降のすべての PRIMERGY サーバ（つまり、PRIMERGY RX300 S5 以降）は PCIe 2.0 をサポートしており、2012 年の導入世代以降のすべての PRIMERGY サーバ（つまり、PRIMERGY RX300 S7 以降）は、PCIe 3.0 をサポートします。バージョンが異なるコンポーネントを混在させると、すべてのコンポーネントでサポートされる最も高い周波数が選択されます。

RAID コントローラーへの適用

次の表は、すべての RAID コントローラーのパフォーマンスを判断するためのデータです。前述の「[SAS および SATA](#)」と「[PCIe](#)」で示したスループットの限界を、ここに記載しています。各ケースのスループットの限界で重要なものを太字で示しています。

コントローラーの略称	CPU コアの数	キャッシュメモリタイプ	ディスク接続の数	ディスクインターフェースのスループットの限界	PCIe バージョン	有効 PCIe 幅	PCIe インターフェースのスループットの限界
Cougar Point			SATA 3G × 4	973 MB/s			
Patsburg A			SATA 3G × 4	973 MB/s			
Patsburg B			SAS 3G × 4	973 MB/s			
LSI1064	1		SAS 3G × 4	973 MB/s	1.0	x4	858 MB/s
LSI1068	1		SAS 3G × 8	1945 MB/s	1.0	x4	858 MB/s
LSI2008	1		SAS 6G × 8	3890 MB/s	2.0	x4	1716 MB/s
	1					x8	3433 MB/s
LSI2108	1	DDR2 / 800 MHz	SAS 6G × 8	3890 MB/s	2.0	x4	1716 MB/s
	1					x8	3433 MB/s
LSI2208-2.0	2	DDR3 / 1333 MHz	SAS 6G × 8	3890 MB/s	2.0	x4	1716 MB/s
						x8	3433 MB/s
LSI2208-3.0	2	DDR3 / 1333 MHz	SAS 6G × 8	3890 MB/s	3.0	x4	3380 MB/s
				3890 MB/s		x8	6761 MB/s

ほとんどの場合、スループットの限界はボトルネックにはなりません。従来型ハードディスクヘランダムアクセスを行うアプリケーションシナリオは高いスループットを達成できませんが、実際には特に優れたパフォーマンスを達成します。

完全なシーケンシャルアクセスで、多くのブロックサイズが 64 KB であるアプリケーションの限界は、PCIe 1.0、x4 で 5 台の従来型ハードディスクを使用した場合にのみ発生します。PCIe 2.0、x8 の場合は、17 台の従来型のハードディスクを使用した場合にのみ発生します。同じブロックサイズを持つ最近の SAS-SSD に対してシーケンシャルアクセスまたはランダムアクセスを行う場合、PCIe 2.0、x8 における限界は、ライト比率に基づき 9~15 台の SSD を使用したときに発生します。

「ディスクインターフェースのスループットの限界」列のスループット値は、そのままコントローラーとハードディスクの間の接続に適用されます。この SAS/SATA インターフェースを介したスループットは、RAID 0 構成の場合のみ、アプリケーション側から見たスループットと同一になります。その他の RAID レベルの場合、SAS/SATA インターフェースを介したスループットは、アプリケーション側から見ると、スループットに特定の係数を乗じた値になります。この係数は常に 1 以上であり、RAID レベルやアクセスパターンの複数の特徴によって異なります。そのため実際のスループットの限界は、この係数により、「ディスクインターフェースのスループットの限界」列の値よりも常に小さい値になります。

電源障害に対するコントローラーキャッシュの保護

電源障害に備えてキャッシュのデータをバックアップするために、現在 PRIMERGY サーバには次の 2 つの選択肢が用意されています。

バッテリーバックアップユニット (BBU)

バッテリーバックアップユニット (BBU) を使用した従来の方法です。この場合、電源障害発生時には、揮発性キャッシュメモリの電源が再充電可能バッテリー (アキュムレーター) を使用してバックアップされます。バッテリーの容量には上限があり、物理的、化学的プロセスによって時間とともに減少していくので、キャッシュメモリに電源を供給できるのは限られた時間だけです。この時間は一定ではなく、BBU の寿命、充電状態、温度といった影響をおよぼす複数の要因によって変わります。そのため、BBU による保証には制限があると言えます。

電源が復旧してサーバに電源が入るとすぐに、RAID コントローラーはキャッシュメモリコンテンツの操作を続行し、データの一貫性は保持されます。

フラッシュバックアップユニット (FBU)

より新しい技術を基盤としたフラッシュバックアップユニット (FBU) を使用すると、電源障害発生時にはデータがキャッシュメモリに保存されるのではなく、キャッシュメモリコンテンツが不揮発性フラッシュメモリにコピーされます。このフラッシュメモリには、データを必要な期間保存しておくことができます。つまり、BBU の問題であった保存期間が、FBU では解消されます。

電源障害発生時に、キャッシュメモリからフラッシュメモリにコピーするために必要となるエネルギーは、スーパーコンデンサから供給されます。

電源が復旧してサーバに電源が入るとすぐに、キャッシュコンテンツはフラッシュメモリから元のキャッシュメモリへ書き込まれます。RAID コントローラーは再度動作可能になり、データの一貫性は保持されます。

パフォーマンス関連の設定

RAID コントローラーが最高のパフォーマンスを発揮するには、各 RAID アレイのパラメータ設定を正確に行う必要があります。設定できるパラメータの数は、コントローラーによって異なります。RAID コントローラーとハードディスクの設定を簡単かつ確実に行うため、PRIMERGY サーバ向けに提供されている RAID-Manager ソフトウェア「ServerView RAID」の使用を推奨します。あらかじめ定義されている「Performance」モードまたは「Data Protection」モードを使用すると、特定のアプリケーションに合わせたコントローラーおよびハードディスクのキャッシュ設定を一括処理できます。「Performance」モードでは、ほとんどのアプリケーションシナリオに対応した、最高のパフォーマンス設定を行えます。

「Performance」モードでは、既存のコントローラーとハードディスクのすべてのキャッシュが有効になります。このため、このモードでは、電源障害が発生した場合に備え、バッテリバックアップユニット（BBU）またはフラッシュバックアップユニット（FBU）を使用して RAID コントローラーのキャッシュをデータ損失から守る必要があります。さらに、ハードディスクのキャッシュも、無停電電源装置（UPS）を使用して保護する必要があります。

特殊なケースでは、「Performance」モードのパラメーター設定を標準から変更したほうが効果的な場合もあります。変更が効果的な場合については、「[コントローラーの比較](#)」の該当する箇所を参照してください。

「ServerView RAID」ソフトウェアのキャッシュ設定には、次の RAID コントローラーおよびハードディスクの設定オプションがあります（設定できるオプションは、コントローラーにより異なります）。最初の 3 つの設定オプションは RAID コントローラーを制御し、最後のオプションは RAID アレイのハードディスクを制御します。すべてのパラメーターは、RAID アレイごとに個別に設定することができます。

Read mode

「Read mode」パラメーターでは、リード時のキャッシュの動作を変えることができます。設定できるオプションは、「No read-ahead」と「Read-ahead」の 2 種類です。「No read-ahead」に設定した場合は、リード時のキャッシュは行われません。「Read-ahead」に設定すると、コントローラーはリード時にキャッシュを行います。

「Read-ahead」の場合、オンボードコントローラー（Patsburg A など）は通常、ブロックを先読みします。キャッシュ付きの PCIe コントローラーでこのオプションを選択すると、さらに異なる動作が見られます。リクエストされたブロックは継続的に分析され、シーケンシャルリードアクセスが行われるかどうかを確認されます。コントローラーがこのようなアクセスを検出すると、リクエストされたブロックに加えて、キャッシュ内の後続ブロックのリードも開始します。これは、それらのブロックが次にリクエストされることを予想しての動作です。現在のオプション「Read-ahead」は、適応型（Adaptive）とも言えます。このオプションは、従来の 2 つのオプション「Read-ahead」と「Adaptive」を統合したものです。

Write mode

「Write mode」は、ライトリクエスト時のコントローラーキャッシュの動作を制御する設定オプションです。ライトキャッシュの設定には、「Write-through」、「Write-back」、「Always Write-back (independent of BBU state)」という 3 つのオプションがあります。「Write-through」オプションでは、コントローラーからの各ライトリクエストは、ハードディスクから応答があった時点で初めて完了済みとしてレポートされます。

「Write-back」および「Always Write-back」オプションでは、リクエストはコントローラーキャッシュにキャッシュされ、完了済みとしてただちにアプリケーションにレスポンスが返されます。リクエストが実際にハードディスクに転送されるのはその後になります。この方法により、コントローラーのリソース利用が最適化され、ライトリクエストの処理が速くなり、スループットが向上します。電源障害には、オプションの BBU/FBU により対応できるので、コントローラーキャッシュのデータ整合性が保証されます。「Always Write-back」オプションでは、キャッシュへの書き込みが常時確保されます。BBU/FBU が動作していない場合にも使用されます。これに対し、「Write-back」オプションでは、コントローラーキャッシュが BBU/FBU によって保護されていない場合、自動的に「Write-through」に切り替わります。

Cache mode

「Cache mode」パラメーター（「I/O キャッシュ」という場合もあります）は、リード時のコントローラーキャッシュの動作に影響を与えます。「Direct」オプションに設定した場合は、データはハードディスクから直接読み取られます。「Cached」に設定した場合は、まずコントローラーキャッシュ内に、ハードデ

ディスクから先読みされたリードリクエストを満たすデータがないか、検索が行われます。この設定では、以降のリードリクエストで利用できるように、すべてのデータがコントローラーキャッシュに書き込まれます。

Disk cache mode

設定できる値は「enabled」および「disabled」です。ほとんどの場合、ディスクキャッシュを有効（enabled）にすると、ライトアクセスのスループットが向上します。システムが UPS で保護されている場合は、パフォーマンスの向上のためにディスクキャッシュを有効にすることを推奨します。

各コントローラーで選択できる設定オプションは次のとおりです。

コントローラーの略称	Read mode	Write mode	Cache mode
オンボードコントローラー Cougar Point	✓		
オンボードコントローラー Patsburg A	✓		
オンボードコントローラー Patsburg B	✓		
LSI1064 / LSI 1068			
LSI2008			
LSI2108	✓	✓	✓
LSI2208-1GB	✓	✓	✓
LSI2208-512	✓	✓	✓
LSI2208-Lite			

最後に、ServerView RAID の「Data Protection」および「Performance」モードでの標準設定を、次の表にまとめます。なお、コントローラーキャッシュ付きのコントローラーを設定する場合、BBU/FBU の有無によっても異なりますが、選択された RAID レベルは無関係です。

コントローラーの略称	BBU/FBU?	Data Protection				Performance			
		Read mode	Write mode	Cache mode	Disk cache	Read mode	Write mode	Cache mode	Disk cache
オンボードコントローラー Cougar Point		Read-ahead			オフ	Read-ahead			オン
オンボードコントローラー Patsburg A		Read-ahead			オフ	Read-ahead			オン
オンボードコントローラー Patsburg B		Read-ahead			オフ	Read-ahead			オン
LSI1064 / LSI1068					オフ				オン
LSI2008					オフ				オン
LSI2108		Read-ahead	Write-through	Direct	オフ	Read-ahead	Always Write-back	Direct	オン
	✓	Read-ahead	Write-back	Direct	オフ	Read-ahead	Write-back	Direct	オン
LSI2208-1GB / LSI2208-512		Read-ahead	Write-through	Direct	オフ	Read-ahead	Always Write-back	Direct	オン
	✓	Read-ahead	Write-back	Direct	オフ	Read-ahead	Write-back	Direct	オン
LSI2208-Lite					オフ				オン

MegaRAID アドバンスソフトウェアオプション

MegaRAID アドバンスソフトウェアオプションには、CacheCade 2.0 と FastPath があります。ここでは、パフォーマンスを向上させるための、個別に検討すべき 2 つの異なるアプローチを取り上げます。

MegaRAID アドバンスソフトウェアオプションは、LSI2208-1GB および LSI2208-512 コントローラーで使用できます。これらをアクティブ化するには、専用のライセンスキーが必要です。MegaRAID アドバンスソフトウェアオプションの最小構成では 2 台のハードディスクを使用しますが、ブレードサーバの場合には LSI2208-512 コントローラーと PRIMERGY SX980 ストレージブレードを組み合わせる必要があるため、通常はより多くのハードディスクが使用されます。

CacheCade 2.0

CacheCade 2.0 はホットスポットデータ用の読み取りおよび読み取り／書き込みキャッシュソフトウェアであり、SSD を高パフォーマンスな不揮発性の動的コントローラーキャッシュプールとして使用することにより、HDD アレイ使用時のアクセス時間を大幅に削減します。これにより、HDD および SSD のハードディスクタイプの 2 つのメリットである、大容量とアクセス時間の短縮を同時に実現することができます。

前提条件と制限

重要な前提条件の 1 つに、いわゆる IO アクセスのアプリケーションホットスポットがあります。データは HDD で構成される RAID アレイ上に存在し、ランダムにアクセスされます。アクセス時間を短縮するには、CacheCade 2.0 キャッシュとして設定され、HDD-RAID アレイに割り当てられた SSD が必要です。これらの HDD と SSD は同じコントローラーに接続する必要があります。CacheCade 2.0 キャッシュに設定できるのは RAID 0、RAID 1、および RAID 1E であり、使用可能な最大キャッシュ容量は 512 GB です。パラメーター設定「Write-through」および「Write-back」では、キャッシュを読み取り専用キャッシュとして使用するか、書き込み／読み取りキャッシュとして使用するかを定義します。

設定例

CacheCade 2.0 の典型的な設定例としては、4 台の SATA-6G-HDD で構成されたタイプ RAID 0 の HDD アレイと、CacheCade 2.0 キャッシュとしての 2 台の SAS-6G-SSD から成るタイプ RAID 1 の SSD アレイなどが挙げられます。これらの設定については、本書で個別に取り上げ、さらに CacheCade 2.0 と組み合わせる詳細に説明します。

FastPath

FastPath は SSD で構成される RAID アレイの高パフォーマンス IO アクセラレーターです。この最適化された LSI MegaRAID テクノロジーにより、SATA-6G-SSD または SAS-6G-SSD が使用された場合に、ランダムアクセスにより IO 負荷が高いアプリケーションのパフォーマンスを明確に向上させることができます。

前提条件と制限

SAS-／SATA-6G および PCIe のスループットの限界にはまだ達しておらず、使用できる SSD の数が十分であっても、高パフォーマンスの 6G-SSD を構成する RAID アレイのトランザクションレートは、小さなブロックへのランダムアクセスが行われる場合に制限されることがあります。このような場合、FastPath を使用することによってトランザクションレートを大幅に向上させることができます。

FastPath を有効に使用できる IO の範囲は、コントローラーにつき約 200000 IO/s からです。「[測定環境](#)」には、8 KB ブロックサイズ、リードの割合 67 % の SAS-6G-SSD の最大トランザクションレートは 14184 IO/s と記載されています。そのため、限界は約 14 台の SSD を使用した場合に始まると判断できます。

オンボードコントローラーの特性

一部の PRIMERGY モデルには、現在、CougarPoint、Patsburg A、Patsburg B のいずれかのオンボードコントローラーが搭載されています。これらのオンボードコントローラーは、最大 4 台の SATA ハードディスクを運用できる、シンプルで低価格なエントリーレベルのソリューションです。また、これらのオンボードコントローラーは、RAID (0、1、10) 機能をサポートすることで一般的な RAID レベルに幅広く対応し、PCIe スロットを占有しません。

本書では、PCIe インターフェース経由で接続されるすべてのコントローラーは、マザーボードに組み込まれている場合でも、オンボードコントローラーとして扱いません（「[RAID コントローラーの概要](#)」の表を参照）。

オンボードコントローラーは、ファームウェア/ドライバベースのソフトウェア RAID ソリューションとして実装されます。通常は、マザーボードのチップセットの 1 つである「サウスブリッジ」チップに組み込まれます。サーバ起動時は、RAID アレイへのアクセスは、ファームウェアから行われます。

オンボードコントローラー自体は CPU を持ちませんが、RAID 機能を実行するためにサーバシステムの CPU を使用します。プロセッサパフォーマンスの使用率は、新しいサーバになるほど重要性が低下します。

CougarPoint

オンボードコントローラー CougarPoint は、BIOS でさまざまなモードに設定できる、純粋な SATA コントローラーです。このタイプのコントローラーを効果的に使用するには「RAID」モードが適していますが、ここでは SATA コントローラーのすべてのモードについて説明します。プロセッサパフォーマンスの使用率は、新しいサーバになるほど重要性が低下します。次の 3 つのモードがあります。

- | | |
|------|---|
| RAID | 柔軟性に優れた推奨モードです。SATA-HDD を非 RAID から RAID 構成にスムーズに移行できる唯一のモードです。NCQ や「ホットスワップ」を含む、SATA のすべての機能をサポートします。PRIMERGY サーバのコントローラー BIOS には、サポートする RAID レベルに対応した「LSI Logic Embedded MegaRAID」というファームウェアが組み込まれています。起動フェーズで RAID アレイが有効になるのはこのモードのみです。また、RAID-Manager ソフトウェアの「ServerView RAID」でコントローラーとハードディスクを認識して管理できるのもこのモードのみです。専用のドライバが必要です。 |
| AHCI | AHCI (Advanced Host Controller Interface) は、メーカー共通の SATA コントローラーのインターフェース規格です。NCQ と「ホットスワップ」をサポートしています。AHCI の場合も、オペレーティングシステムで専用のドライバが必要です。 |
| IDE | このモードでは、SATA ポートがオペレーティングシステムで認識されるようになります。NCQ はサポートしません。また、適切な SATA ドライバが必要です。「ServerStart DVD」に各種オペレーティングシステム用のドライバが収録されています。 |

Patsburg A と Patsburg B

2012 年にリリースされたサーバ世代の C600 チップセットで、2 つ以上のプロセッサソケット (PRIMERGY RX300 S7 など) が付属しています。完全に新しいアーキテクチャーで、初めて SAS/SATA コントローラーを統合しました。このコントローラーは、標準で SATA モードで動作します (Patsburg A)。別売の「SAS 有効化キー」を利用すると、より高価値の SAS/SATA バージョン (Patsburg B) にアップグレードできます。

SATA 設定オプション (「AHCI」を含む) については、Patsburg A/B を搭載した多数のサーバの BIOS でプロビジョニングされていますが、通常は光ドライブの SATA 接続のみを指し、ここで取り上げている RAID 構成では使用されません。

測定内容

ここまでは、各種コントローラーを紹介し、その技術的特徴について説明してきました。次の「[コントローラーの比較](#)」では、さまざまなアプリケーションシナリオでのコントローラーについて、測定結果に基づいて説明します。そのため、まず測定方法と測定環境について簡単に説明します。

測定方法とディスク I/O パフォーマンスの基本については、ホワイトペーパー『[ディスク I/O パフォーマンスの基本](#)』を参照してください。

測定方法

PRIMERGY サーバのディスクサブシステムのパフォーマンス測定は、実際のアプリケーションシナリオでのディスクアクセスを、規定に基づいてモデル化して行います。

規定する項目は次のとおりです。

- ランダムアクセス／シーケンシャルアクセスの比率
- リードアクセス／ライトアクセスの比率
- ブロックサイズ (KB)
- 同時アクセス数 (未処理 I/O の数)

規定した値の組み合わせを「負荷プロファイル」と呼びます。次の 5 つの標準負荷プロファイルは、典型的なアプリケーションシナリオに相当します。

標準負荷プロファイル	アクセス	アクセスの種類		ブロックサイズ [KB]	アプリケーション
		リード	ライト		
ファイルコピー	ランダム	50 %	50 %	64	ファイルのコピー
ファイルサーバ	ランダム	67 %	33 %	64	ファイルサーバ
データベース	ランダム	67 %	33 %	8	データベース (データ転送) メールサーバ
ストリーミング	シーケンシャル	100 %	0 %	64	データベース (ログファイル)、データバックアップ、ビデオストリーミング (一部)
リストア	シーケンシャル	0 %	100 %	64	ファイルのリストア

異なる負荷密度で同時にアクセスするアプリケーションをモデル化するため、「未処理 I/O の数」を 1、3、8 から 512 まで増やしていきます (8 以降は 2 の累乗で加算していきます)。

本書の測定は、これらの標準負荷プロファイルに基づいて行いました。

主な測定結果は次のとおりです。

- スループット [MB/s] 1 秒あたりのデータ転送量 (メガバイト単位)
- トランザクション [I/O/s] 1 秒あたりの I/O 処理数
- レイテンシー [ms] 平均応答時間 (ミリ秒単位)

通常、シーケンシャルな負荷プロファイルでは「データスループット」が使用され、小規模なブロックサイズを使用するランダムな負荷プロファイルでは「トランザクションレート」が使用されます。スループットとトランザクションは互いに正比例の関係にあるので、次の計算式で相互に算出できます。

データスループット [MB/s]	= トランザクションレート [I/O/s] × ブロックサイズ [MB]
トランザクションレート [I/O/s]	= データスループット [MB/s] / ブロックサイズ [MB]

測定環境

本書で示すすべての測定は、次のハードウェアとソフトウェアのコンポーネントを使用して行いました。

SUT (System Under Test : テスト対象システム)	
モデル	PRIMERGY RX300 S7 PRIMERGY TX300 S7 PRIMERGY TX120 S3p PRIMERGY BX920 S3
オペレーティングシステム	Microsoft Windows Server 2008 Enterprise x64 Edition SP2
RAID Manager ソフトウェア	ServerView RAID Manager 5.5.2 ServerView RAID Manager 5.7.3
RAID アレイの初期化	RAID アレイは、測定前に 64 KB の基本ブロックサイズ（「ストライプサイズ」）で初期化
ファイルシステム	NTFS
測定ツール	Iometer 2006.07.27
測定データ	32 GB の測定ファイル（1~8 台のハードディスク用）、64 GB の測定ファイル（9~16 台のハードディスク用）、128 GB の測定ファイル（17 台以上のハードディスク用）
オンボード SATA コントローラー 「Cougar Point」	Intel BD82C202 PCH、（PRIMERGY TX120 S3p で使用） ドライバ名 : megasr1.sys、ドライババージョン : 15:00.0329.2012 BIOS バージョン : A. 10.03031333R SATA RAID モード
オンボード SATA コントローラー 「Patsburg A」	Intel C600 チップセット（PRIMERGY RX300 S7 で使用） ドライバ名 : megasr1.sys、ドライババージョン : 15:00.0329.2012 BIOS バージョン : SCU.11.03201000R
オンボード SAS コントローラー 「Patsburg B」	Intel C600 チップセット（PRIMERGY RX300 S7 で使用） ドライバ名 : megasr1.sys、ドライババージョン : 15:00.0329.2012 BIOS バージョン : SCU.11.03201000R SAS 有効化キー
コントローラー 「RAID 0/1 SAS based on LSI MegaRAID 4Port」 （LSI MegaRAID SAS 1064）	ドライバ名 : lsi_sas.sys、ドライババージョン : 1.34.03.00 ファームウェアバージョン : 1.33.00.00 BIOS バージョン : 6.36.00.00
コントローラー 「RAID 0/1 SAS based on LSI MegaRAID 8Port」 （LSI MegaRAID SAS 1068）	ドライバ名 : lsi_sas.sys、ドライババージョン : 1.34.03.00 ファームウェアバージョン : 1.33.00.00 BIOS バージョン : 6.36.00.00
コントローラー 「RAID Ctrl SAS 6G 0/1」 （LSI MegaRAID SAS 2008）	ドライバ名 : megasas2.sys、ドライババージョン : 5.2.103.64 ファームウェアパッケージ : 20.10.1-0075 ファームウェアバージョン : 2.120.224-1451
コントローラー 「RAID Ctrl SAS 5/6 512MB （D2616）」 （LSI MegaRAID SAS 2108）	ドライバ名 : megasas2.sys、ドライババージョン : 5.2.116.64 ファームウェアパッケージ : 12.12.0-0087、ファームウェアバージョン : 2.120.233-1471 コントローラーキャッシュ : 512 MB
コントローラー 「RAID Ctrl SAS 6G 5/6 1GB （D3116）」	ドライバ名 : megasas2.sys、ドライババージョン : 5.2.116.64 ファームウェアパッケージ : 23.2.1-0048、ファームウェアバージョン : 3.152.65-1626 コントローラーキャッシュ : 1 GB
コントローラー 「RAID Ctrl SAS 6G 5/6 1GB （D3116C）」	ドライバ名 : megasas2.sys、ドライババージョン : 6.505.5 ファームウェアパッケージ : 23.9.0-0023、ファームウェアバージョン : 3.220.75-2196 コントローラーキャッシュ : 1 GB
コントローラー 「PY SAS RAID HDD Module」	ドライバ名 : megasas2.sys、ドライババージョン : 5.2.116.64 ファームウェアパッケージ : 23.2.1-0049、ファームウェアバージョン : 3.152.75-1658

	コントローラーキャッシュ : 512 MB
コントローラー 「PY SAS RAID HDD Module w/o cache」	ドライバ名 : megasas2.sys、ドライババージョン : 5.2.116.64 ファームウェアパッケージ : 23.2.1-0049、ファームウェアバージョン : 3.152.74-1659
SATA-6G-HDD	Seagate ST91000640NS、2.5 インチ、1000 GB、7200 rpm、6 Gb/s
SAS-6G-HDD	Toshiba MK1401GRRB、2.5 インチ、146 GB、15000 rpm、6 Gb/s
SATA-6G-SSD	Seagate ST200FM0002、2.5 インチ、200 GB、6 Gb/s

ここで再び、コントローラーの比較に使用したハードディスクモデルとその基本的なパフォーマンス値を詳しくまとめ、次の表に示します。これらは、コントローラーで得たパフォーマンス値を理解するために重要です。高性能の SATA-6G および SAS-6G ハードディスクは、従来型のハードディスク（HDD）として各ケースで選択されています。また 200 GB SAS-6G-SSD は、SSD クラスとして表記しています。

表では、前述の「[測定方法](#)」で説明した 5 つの標準負荷プロファイルに従って、1 台のハードディスクで測定した場合の最大値を示しています。最適なパフォーマンスが得られるように、ハードディスクキャッシュはすべてのケースで有効になっています。

ハードディスクタイプ	省略名 (略称)	シーケンシャル最大 スループット [MB/s] 64 KB ブロックサイズ		ランダムアクセスの最大トランザク ションレート [IO/s]		
		リード	ライト	8 KB ブロックサイズ	64 KB ブロックサイズ	
				リードの割合 :	リードの割合 :	
				67 %	67 %	50 %
HDD SATA、6 Gb/s、2.5 インチ 1000 GB、7200 rpm ホットプラグ対応 カテゴリー : ビジネスクリティ カル (BC)	SATA-6G-HDD	108 MB/s	107 MB/s	372 IO/s	317 IO/s	314 IO/s
HDD SAS、6 Gb/s、2.5 インチ 146 GB、15000 rpm、 カテゴリー : エンタープライズ (EP)	SAS-6G-HDD	207 MB/s	204 MB/s	594 IO/s	506 IO/s	492 IO/s
SSD SAS、6 Gb/s、2.5 インチ 64 GB、 ホットプラグ対応、 カテゴリー : エンタープライズ (EP)	SAS-6G-SSD	339 MB/s	196 MB/s	14184 IO/s	3006 IO/s	2985 IO/s

ここに記載した情報、および PRIMERGY サーバ向けにリリースされたハードディスクの詳細については、ホワイトペーパー『[単一ディスクのパフォーマンス](#)』を参照してください。

コントローラーの比較

ここまでは、コントローラに関する重要な予備情報について解説しました。多くの場合、この情報により、所定のアプリケーションで選択すべきコントローラーを絞り込めます。しかし、コントローラーの想定用途に関する顧客情報がさらに追加されると、個々のコントローラーのパフォーマンスに関して、より詳細な情報が求められる可能性があります。このため、ここでは、さまざまな RAID レベル、アプリケーションシナリオ、負荷密度、ハードディスク数、およびハードディスクテクノロジーにおいて、コントローラーを比較し、測定結果を分析しながら解説します。比較は次のように分けて行われました。これらの比較は、それぞれ単独で参照できます。

- [RAID 1 \(SATA ハードディスク 2 台\)](#)
- [RAID 1 \(SAS ハードディスク 2 台\)](#)
- [RAID 0 および 10 \(SATA ハードディスク 4 台で構成\)](#)
- [RAID 0 および 10 \(ハードディスク 5~8 台で構成\)](#)
- [RAID 0 および 10 \(9 台以上のハードディスクで構成\)](#)
- [RAID 5](#)

比較に関する一般的な前提事項：

- 比較では、「[測定方法](#)」で説明した 5 つの負荷プロファイル（ファイルコピー、データベース、ファイルサーバ、ストリーミング、リストア）を主に使用しています。これにより、ランダムおよびシーケンシャルアプリケーションシナリオを十分にカバーできます。顧客の負荷プロファイルがこれと大幅に異なる場合は、ここでの解説は制限付きでの適用となります。
- ディスクサブシステムのパフォーマンスを測定するベンチマークでは、一般の規定に従い、ランダム負荷プロファイルではトランザクションレート（IO/s）を、シーケンシャル負荷プロファイルではスループット（MB/s）を使用しています。
- RAID レベルをサポートするすべてのコントローラーと、これまで解説してきたハードディスクタイプについて検討します。
- わかりやすくするため、以降のほとんどのグラフで、達成可能な最大値のみを示しています。これらは通常、ディスクサブシステムが高負荷密度の場合にのみ達成される値です。
- ハードディスクテクノロジーの代表として、「[測定環境](#)」で詳しく説明している 3 種類のハードディスク（SATA-6G-HDD、SAS-6G-HDD、SAS-6G-SSD）を使用しています。また、これらの重要なパフォーマンスデータも示します。以降の比較の一部では、達成したパフォーマンス値を、これらのハードディスクタイプのパフォーマンスデータに基づいて説明しています。
- 本書では、コントローラーで達成できる最高のパフォーマンスをテーマとしています。このため、「ServerView RAID」の設定は「Data Protection」モードではなく、「Performance」モード（キャッシュおよびハードディスクの設定は規定のまま）で測定しています。この設定は、顧客が電源障害に対して十分な保護対策を行っている場合に有効です。これらの設定を変更した方が効果的な場合は、ケースごとに明記します。
- 以降のコントローラーの比較では、従来のハードディスクと SSD を区別するため、従来のハードディスクを「HDD」と表記します。
- 最新の世代では、LSI2208 コントローラーのグループは PCIe 3.0 をサポートします。これは一般にスループットの限界が高いことを意味しており、多数のハードディスクを使用した構成によって達成されます。そのため、最大 8 台のハードディスクを使用した構成に関する説明は、PCIe 2.0 および 3.0 の世代にも同様に有効となります。したがって、該当のセクションにおいて、2 つの世代は分けて表記していません。PCIe 3.0 世代を別個に扱う意味があるのは、9 台以上のハードディスクを搭載した RAID 0 および RAID 10 構成と、24 台のハードディスクを搭載した RAID 5 構成、すなわち、LSI2208-1G-3.0 コントローラーを基盤としている場合だけです。

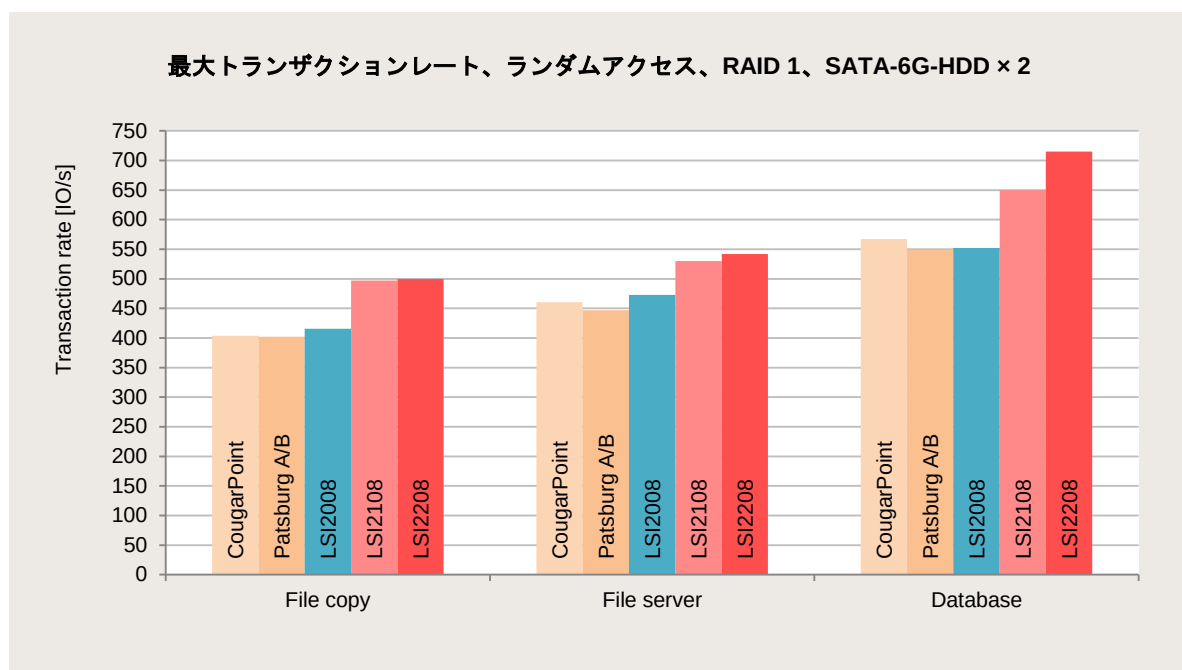
RAID 1 (SATA ハードディスク 2 台)

すべてのオンボードコントローラーと、多くの PCIe コントローラーとの比較はハードディスク 2 台以下でのみ可能なため、3 台以上のハードディスクを使用する場合とは分けて検証する必要があります。比較は SATA-6G-HDD を使用して行います。SATA-HDD の詳細は、「[測定環境](#)」を参照してください。この数の HDD ではコントローラーの能力は限界に達しないため、各負荷プロファイルでのパフォーマンス値は主に HDD によって決まります。

ランダムアクセス

RAID 1 (SATA-6G-HDD × 2 で構成)

次のグラフは、2 台の SATA-6G-HDD で RAID 1 を構成した場合におけるコントローラーの比較を示しています。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。このケースでは、コントローラーの違いによるパフォーマンスへの影響はほとんどありません。

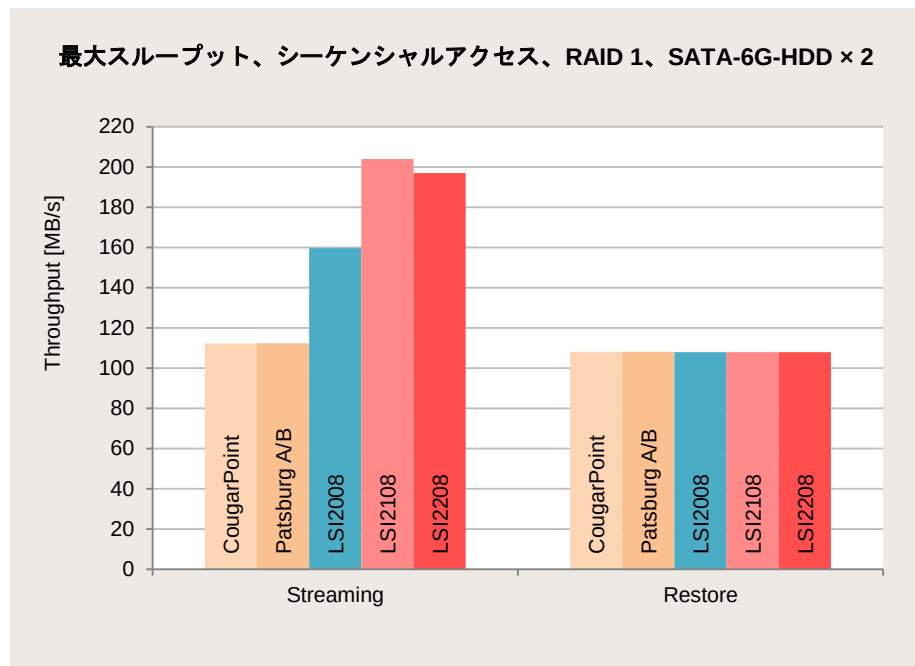


LSI2108 および LSI2208-1G コントローラーは、ここで総合的に最高のトランザクションレートを提供するコントローラーです。

シーケンシャルアクセス

RAID 1 (SATA-6G-HDD × 2 で構成)

次のグラフは、2 台の SATA-6G-HDD で RAID 1 を構成した場合におけるコントローラーの比較を示しています。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」(シーケンシャルアクセス、100 %リード、64 KB ブロックサイズ)と「リストア」(シーケンシャルアクセス、100 %ライト、64 KB ブロックサイズ)でのスループットを示しています。



一般的に、この RAID レベルでは、最大スループットは単一ハードディスクの限界に近いものになります。LSI2008、LSI2108、LSI2208 コントローラーは 2 台のハードディスクを高負荷密度で使用するため、リードでの最大スループットが高くなります。

RAID 1 (SAS ハードディスク 2 台)

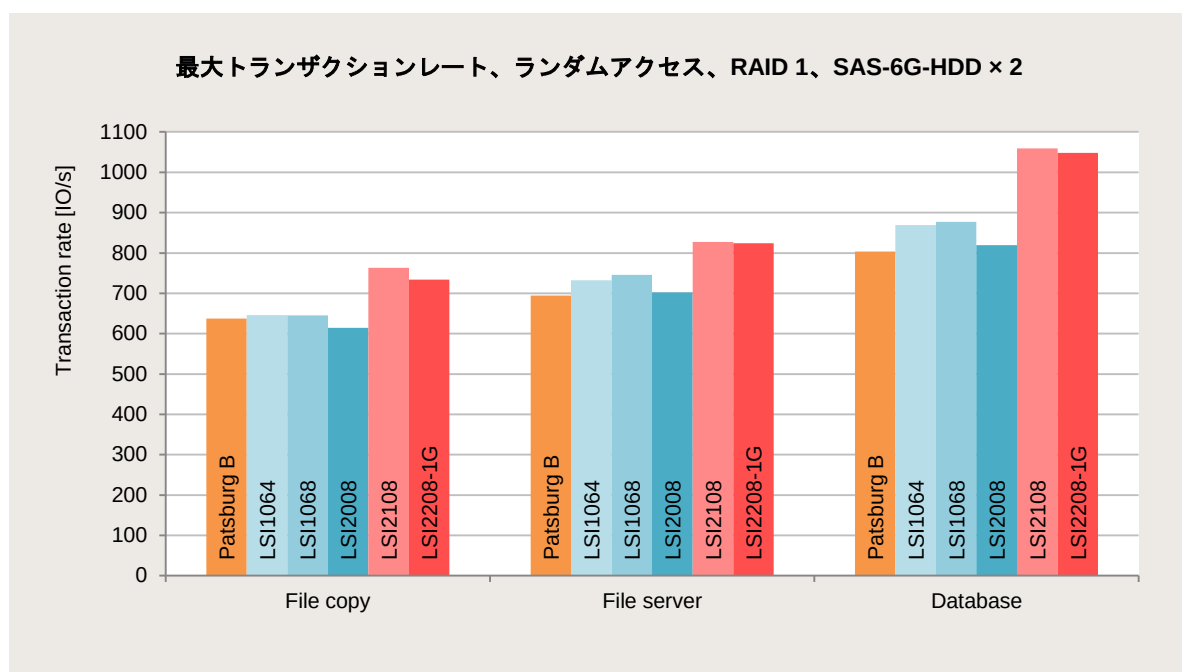
すべての PCIe コントローラーと、少なくとも 1 台のオンボードコントローラーとの比較はハードディスク 2 台以下でのみ可能なため、3 台以上のハードディスクを使用する場合とは分けて検証する必要があります。比較は SAS-6G-HDD と SAS-6G-SSD を使用して行います。これらの詳細は、「[測定環境](#)」を参照してください。

RAID 1 の場合、キャッシュ付きのコントローラーでは、通常、コントローラーキャッシュを無効にした方がパフォーマンスが向上します。

ランダムアクセス

RAID 1 (SAS-6G-HDD × 2 で構成)

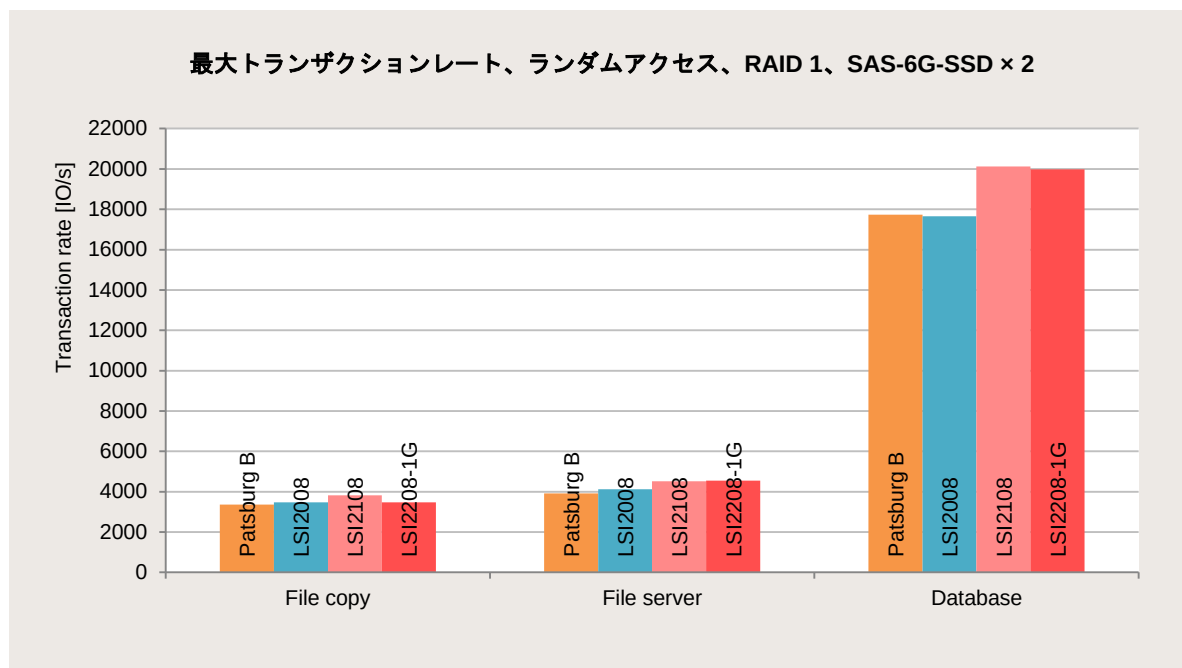
次のグラフは、2 台の SAS-6G-HDD で RAID 1 を構成した場合におけるコントローラーの比較を示しています。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



LSI2108 および LSI2208-1G コントローラーは、ここで総合的に最高のトランザクションレートを提供するコントローラーです。

RAID 1 (SAS-6G-SSD × 2 で構成)

次のグラフは、2 台の SAS-6G-HDD で RAID 1 を構成した場合におけるコントローラーの比較を示しています。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



LSI2108 および LSI2208-1G コントローラーは、ここで総合的に最高のトランザクションレートを提供するコントローラーです。

シーケンシャルアクセス

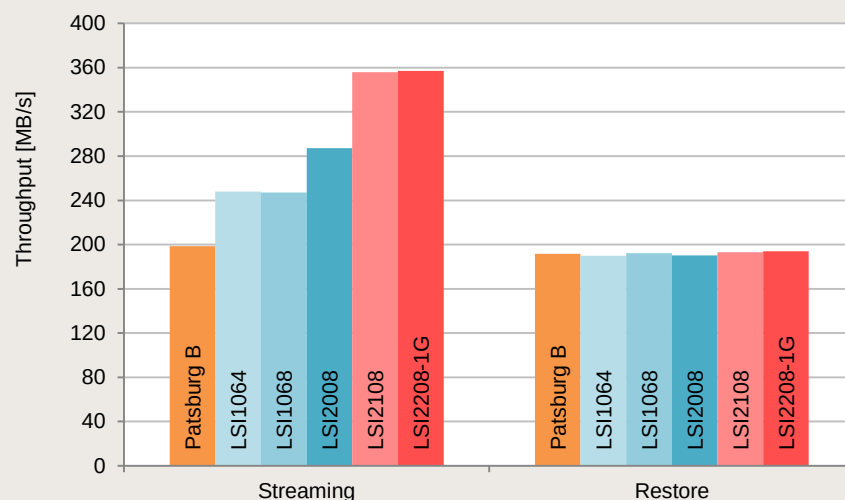
RAID 1 (SAS-6G-HDD × 2 で構成)

次のグラフは、2 台の SAS-6G-HDD で RAID 1 を構成した場合におけるコントローラーの比較を示しています。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

書き込みの場合、この RAID レベルでは、すべてのケースにおける最大スループットは単一ハードディスクの最大値に近いものになります。

読み取りの場合、負荷密度が高いとき、Patsburg B を除くすべてのコントローラーが両方の HDD を使用し、単一 HDD を使用した場合よりも高い読み取りスループットを達成しています。

最大スループット、シーケンシャルアクセス、RAID 1、SAS-6G-HDD × 2



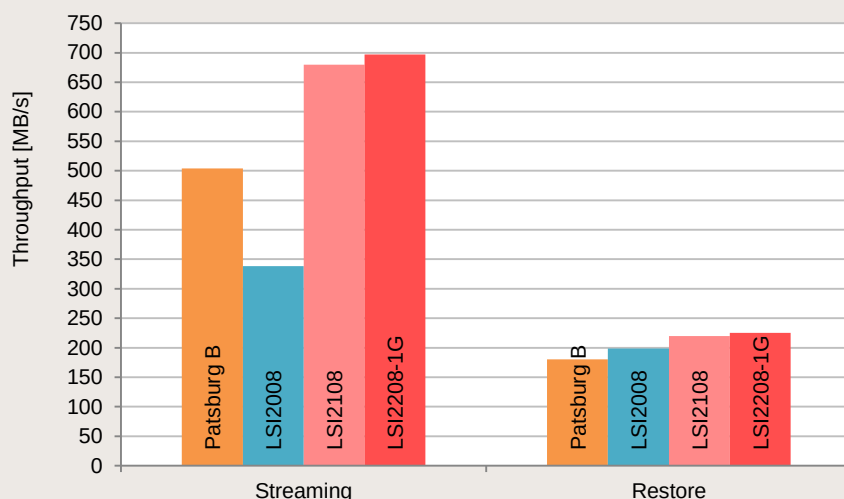
RAID 1 (SAS-6G-SSD × 2 で構成)

次のグラフは、2 台の SAS-6G-SSD で RAID 1 を構成した場合におけるコントローラーの比較を示しています。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

書き込みの場合、この RAID レベルでは、すべてのケースにおける最大スループットは単一ハードディスクの最大値に近いものになります。

読み取りの場合、負荷密度が高いとき、多くのコントローラーが両方の SSD を使用し、単一 SSD を使用した場合よりも高い読み取りスループットを達成しています。

最大スループット、シーケンシャルアクセス、RAID 1、SAS-6G-SSD × 2



Patsburg B コントローラーは標準負荷プロファイル「ストリーミング」で約 500 MB/s を達成しています。これは、SAS-3G 接続の実効スループットの限界の約 2 倍です。これはつまり、2 つの SAS-3G 接続がフルに使用されたことを示します。ここでは、LSI2108 および LSI2208 コントローラーによって最大シーケンシャルスループットが達成されました。2 つのコントローラーのスループットは、標準負荷プロファイル「ストリーミング」で、500 MB/s を大幅に上回っています。

RAID 0 および 10 (SATA ハードディスク 4 台で構成)

ここでは 4 台のハードディスクを使用して、「[RAID 1 \(SATA ハードディスク 2 台で構成\)](#)」で行われたコントローラーの比較を続けます。ハードディスクの数が増えると、各コントローラーの違いが明確になります。

RAID 0 構成の場合、キャッシュ付きのコントローラーでは通常、コントローラーキャッシュを無効にした方が最大パフォーマンスの値が大きくなります。このため、ここでは LSI2108 および LSI2208-1G コントローラーのトランザクションレートは、ServerView RAID の「Performance」モードの[標準設定](#)を次のように変更したうえで測定しています。

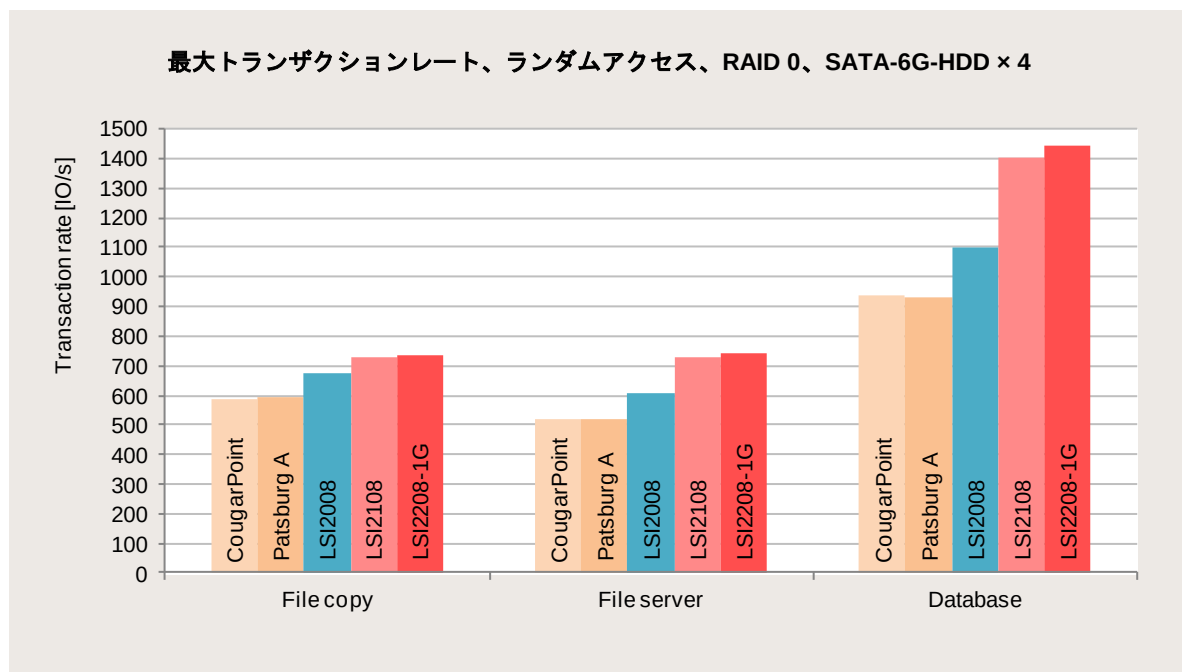
- Read mode : 「Read-ahead」から「No read-ahead」に変更
- Write mode : 「Always write-back」から「Write-through」に変更

RAID 10 構成では[標準設定](#)を使用しています。

ランダムアクセス

RAID 0 (SATA-6G-HDD × 4 で構成)

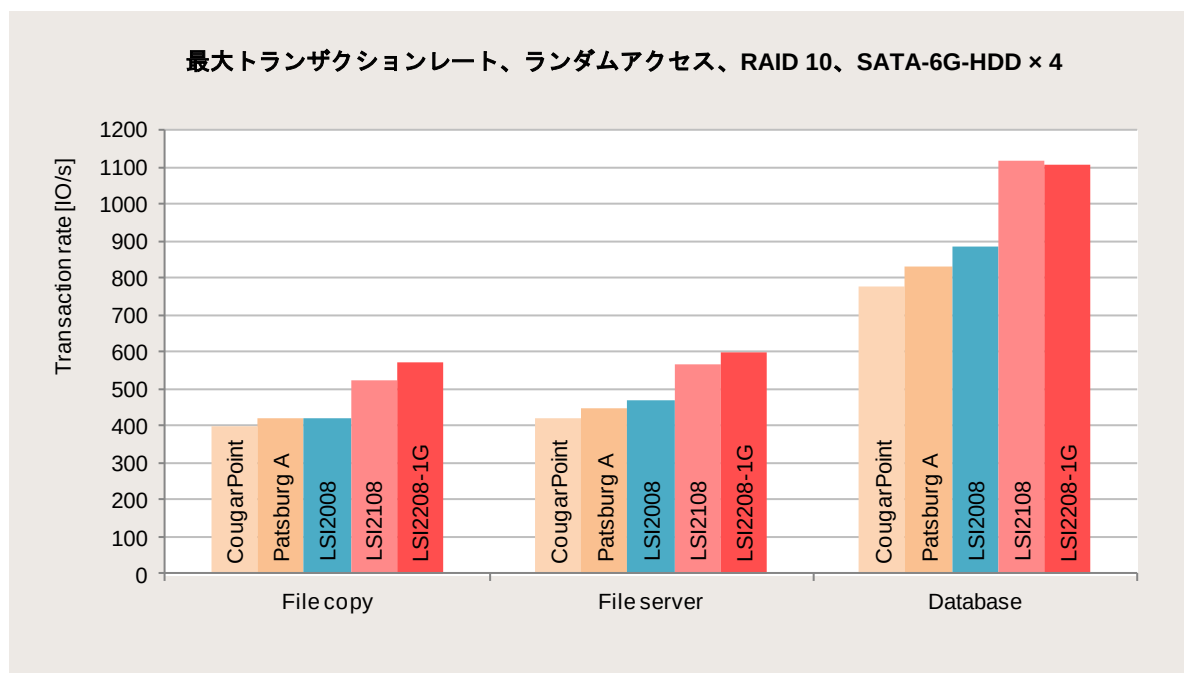
次のグラフは、ランダム負荷プロファイルでの RAID 0 のハードディスクアレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」(ランダムアクセス、50 %リード、64 KB ブロックサイズ)、「ファイルサーバ」(ランダムアクセス、67 %リード、64 KB ブロックサイズ)、「データベース」(ランダムアクセス、67 %リード、8 KB ブロックサイズ)でのトランザクションレートを示しています。



コントローラーの品質が高いと、トランザクションレートが高いことが明らかです。

RAID 10 (SATA-6G-HDD × 4 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 10 のハードディスクアレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



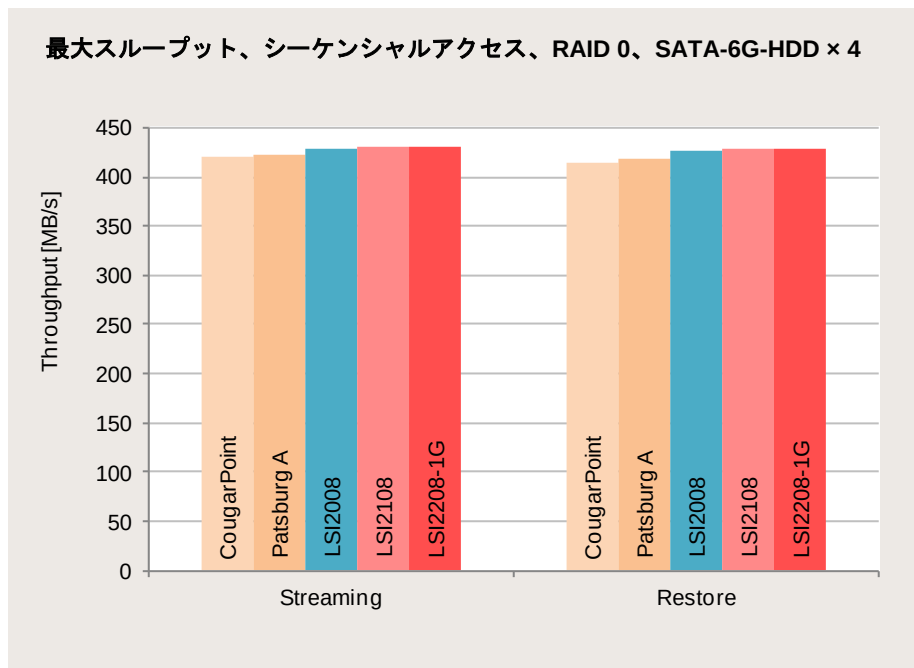
ここでも、コントローラーの品質が高いと、トランザクションレートが高いことが明らかです。

シーケンシャルアクセス

RAID 0 (SATA-6G-HDD × 4 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 0 のハードディスクアレイにおける最大スループットを示しています。このスループットはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

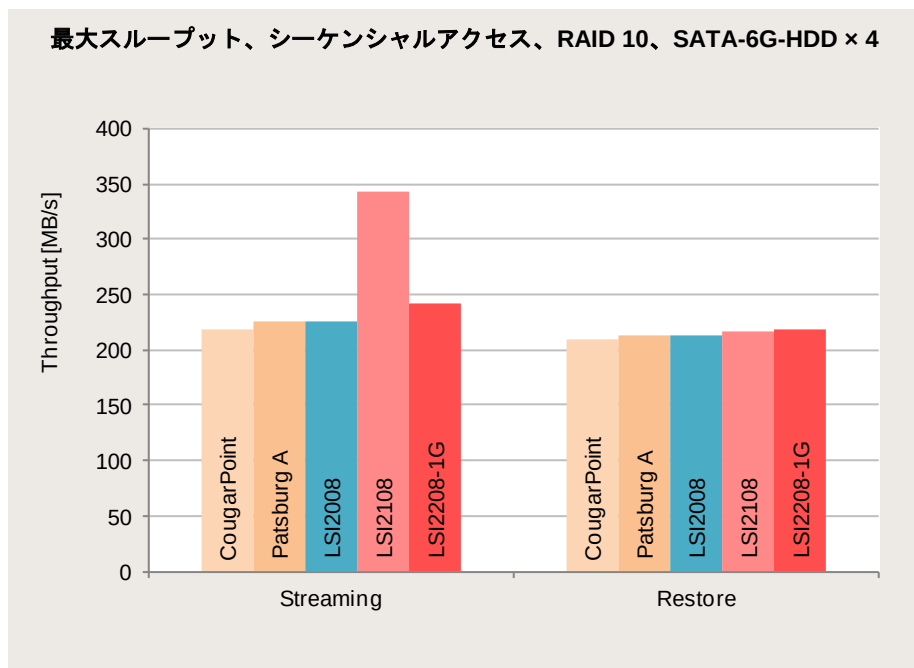
これらのケースでは、各コントローラーのパフォーマンスはほとんど同じです。



RAID 10 (SATA-6G-HDD × 4 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 0 のハードディスクアレイにおける最大スループットを示しています。このスループットはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

これらの多くのケースでは、各コントローラーのパフォーマンスは同じです。標準負荷プロファイル「ストリーミング」の場合、LSI2108 コントローラーだけが高いスループットを達成しています。LSI2108 コントローラーは同時に 2 台だけでなく、4 台の HDD から読み取りを行い、結果的に単一 HDD の 2 倍を超える値にまでスループットを大幅に向上させています。



RAID 0 および 10（ハードディスク 5～8 台で構成）

オンボードコントローラーは、現在の PRIMERGY サーバで 5 台以上のハードディスクを操作するには不十分です。このため、ここでは PCIe コントローラーのみを比較しています。2 つの PCIe コントローラー（LSI1068 および LSI2008）は最大 8 台のハードディスクに対応しているため、8 台のハードディスクが接続された場合について、すべてのコントローラーを比較します。また、ここで示されているパフォーマンス値は、典型的な中程度の数のハードディスクを使用した場合のパフォーマンス値の範囲をカバーしています。ここでは、より高いパフォーマンス要件の最大値が特に重要なので、高パフォーマンスの SAS-6G-HDD または SAS-6G-SSD による測定に基づいて解説しています。これらのハードディスクについての詳細は「[測定環境](#)」を参照してください。

LSI1068 コントローラーとは異なり、LSI1064 コントローラーはリリースされているすべての構成で最大 4 台のハードディスクしか接続できないため、ここでは扱いません。このコントローラーのパフォーマンス関連の詳細については、前述の「[RAID 1（SAS ハードディスク 2 台）](#)」を参照してください。

RAID 0 構成の場合、キャッシュ付きのコントローラーでは通常、コントローラーキャッシュを無効にした方が最大パフォーマンスの値が大きくなります。このため、ここでは LSI2108 および LSI2208-1G コントローラーのトランザクションレートは、ServerView RAID の「Performance」モードの[標準設定](#)を次のように変更したうえで測定しています。

- Read mode : 「Read-ahead」から「No read-ahead」に変更
- Write mode : 「Always write-back」から「Write-through」に変更

RAID 10 構成では[標準設定](#)を使用しています。

ランダムアクセス

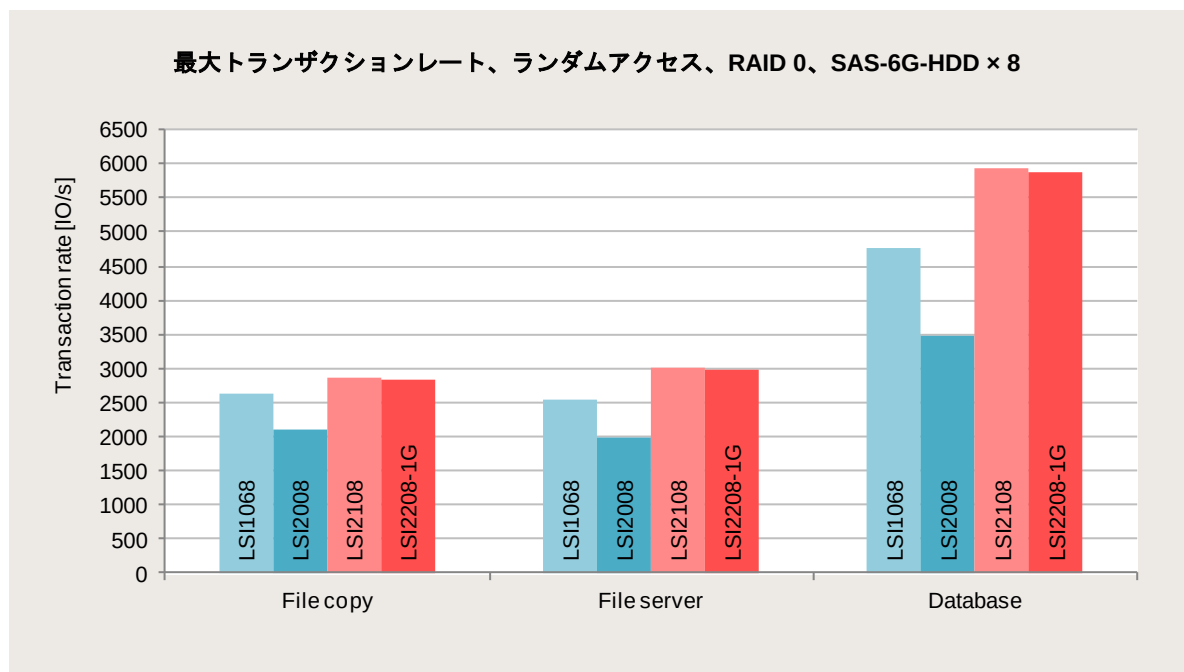
多数のハードディスクへのランダムアクセスでは、SSD の最大値が大幅に異なるため、HDD と SSD を区別して検討します。

HDD

以降では、コントローラーによる HDD へのランダムアクセスを比較しています。ここでは、各負荷プロファイルでのストレージメディアの最大トランザクションレートが、最も重要な制限要因になります。ただし、こうしたケースでのパフォーマンスは、コントローラーと完全に独立しているわけではありません。キャッシュなしのコントローラー（LSI1068 および LSI2008）では別の影響要因もあります。次に示す測定結果は 8 台の SAS-6G-HDD で得られたものですが、異なるタイプやハードディスク数（8 台以下）での最大トランザクションレートを見積もる場合にも使用できます。HDD へのランダムアクセスで得られるスループットは非常に低いため、コントローラーの PCIe または SAS インターフェースの限界には達しません。

RAID 0 (SAS-6G-HDD × 8 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 0 のハードディスクアレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



このグラフの各プロファイルの右側 2 列は、キャッシュ付きのコントローラー（LSI2108 および LSI2208-1G）を示しています。この 2 つのコントローラーは、各負荷プロファイルでこの RAID アレイの最大トランザクションレートをほぼ達成しています。残り 2 つのコントローラー（LSI1068 および LSI2008）のトランザクションレートは、それぞれが、その 60~90 % しか達成していません。トランザクションレートをより精密に分析すると、ランダム負荷プロファイルの場合、これら 2 つのコントローラーのパフォーマンス値は、負荷密度が低い、または中程度のキャッシュ付きコントローラー 2 つと同程度のパフォーマンス値を達成しています（LSI2008 の場合は同時アクセスの回数が約 32 回まで、LSI1068 の場合は 64 回まで）。同時アクセスの回数がそれ以上になると、キャッシュなしのコントローラー（LSI1068 および LSI2008）では、トランザクションレートが停滞します。

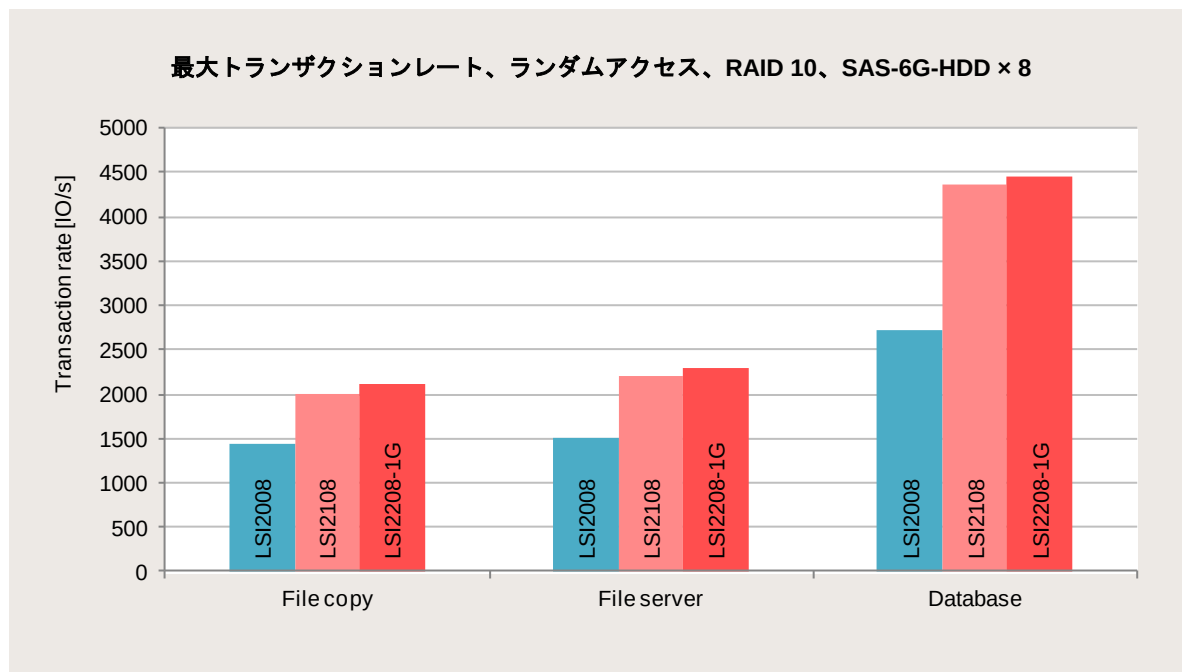
ここでは、8 台のハードディスクでランダムアクセスを行った場合を例に、2 つのコントローラーグループの主な違いを示しています。この違いの正確なパーセンテージは、RAID アレイのサイズとハードディスクのタイプによって異なります。このような RAID アレイ（HDD 8 台のキャッシュなしコントローラー）について、32 回以上の同時アクセスでもパフォーマンスの向上が見られる場合は、より少数のハードディスクで構成された RAID アレイでも同様の違いが生じます。「[RAID 1 \(SAS ハードディスク 2 台\)](#)」ではまだ測定結果に違いが見られませんが、「[RAID 0 および 10 \(SATA ハードディスク 4 台で構成\)](#)」の測定結果における違いは明らかです。

ここでの測定結果によると、3 つすべての負荷プロファイルにおいて、使用されている HDD タイプの最大トランザクションレートは比較的近いものの、ブロックサイズ 64 KB での負荷プロファイルは、ブロックサイズ 8 KB での負荷プロファイルの約半分のトランザクションレートしか達成できません。¹

¹ これは、RAID アレイが 64 KB の基本ブロックサイズ（「ストライプサイズ」）で初期化されるためです。「ストライプサイズ」より小さい 8 KB のブロックサイズでのランダム負荷プロファイルは、「HDD の数 × HDD パフォーマンス」という式で直接予測できます。RAID アレイを初期化したときと同じブロックサイズを負荷プロファイルで使用すると、約半分のトランザクションレートしか達成できません。この現象の詳細な理由については、ここでは触れません。

RAID 10 (SAS-6G-HDD × 8 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 10 のハードディスクアレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。LSI1068 コントローラーは RAID 10 をサポートしていないため、ここでは扱いません。



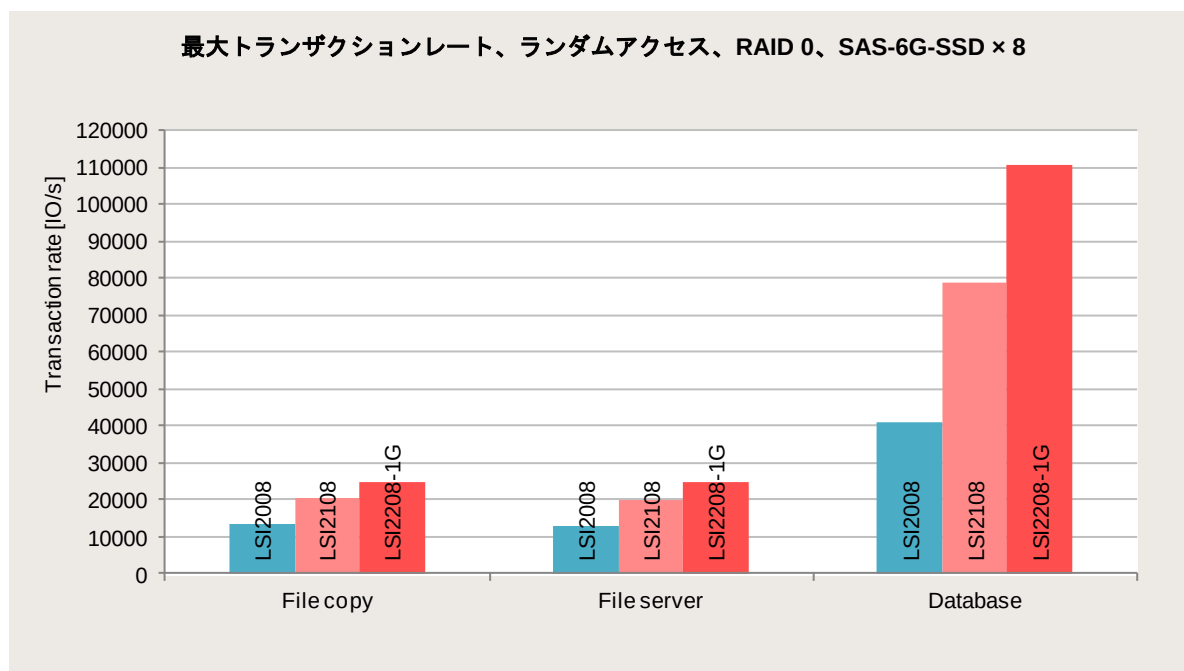
グラフは、RAID 0 と同様の傾向を示しています。つまり、各負荷プロファイルにおいて、キャッシュなしのコントローラー（LSI2008）が達成できるのは、キャッシュ付きのコントローラーのトランザクションレートの約 70 % です。また、RAID 10 においても、このようにパフォーマンスの差異が生じるのは、RAID アレイへ 32 回以上の同時アクセスを行った場合のみです。

SSD

通常、個別の HDD にランダムアクセスを行った場合の最大トランザクションレートは、ハードディスクキャッシュを有効にした場合でも 700 IO/s 未満ですが、SAS-6G-SSD の場合のトランザクションレートはその約 20 倍になります。このようにトランザクションレートが高いため、複数の SAS-6G-SSD でアレイを構成すると、ランダムアクセスで数百 MB/s のスループットを提供できます。つまり、コントローラーのリソースとインターフェースに HDD よりはるかに高い負荷がかかるため、コントローラー世代間で差が生じます。

RAID 0 (SAS-6G-SSD × 8 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 0 の SSD アレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



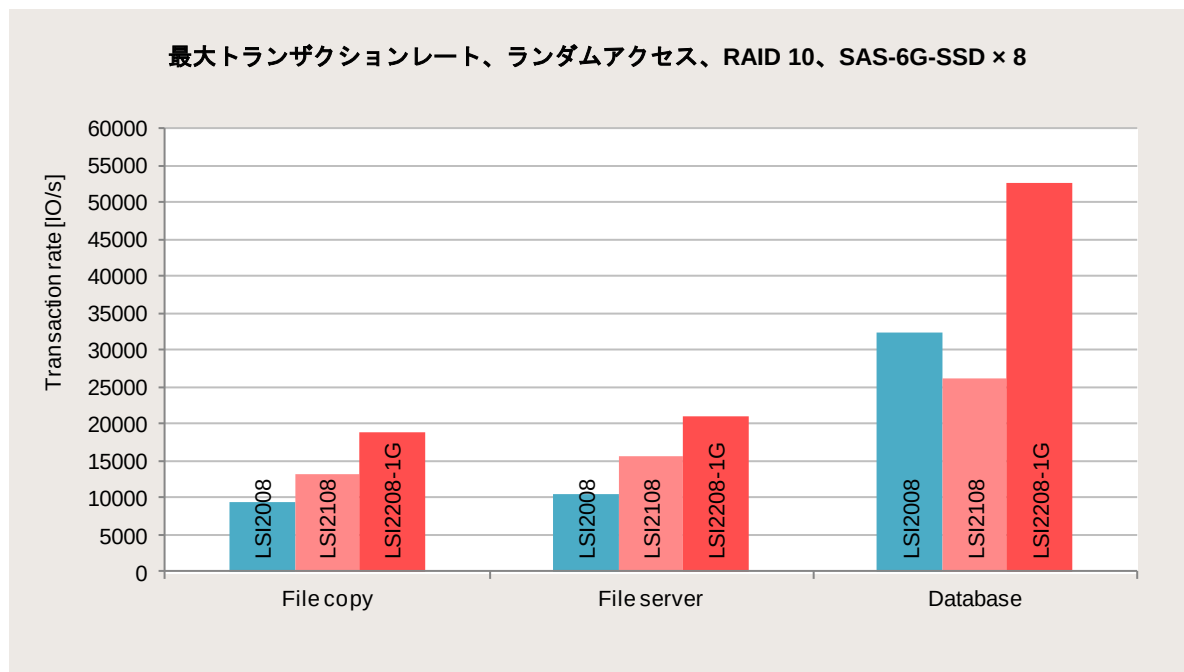
ここでは、LSI2208-1G コントローラーが最高のパフォーマンスを実現しています。

これらのトランザクションレートに関連するスループット値を理解すると興味深いことがわかります。トランザクションレートが低くても、64 KB ブロックサイズの 2 つの負荷プロファイルではスループットが高くなります。例えば、LSI2208-1G コントローラーは、「ファイルサーバ」負荷プロファイルで約 1563 MB/s のスループットを処理します。

次に思い浮かぶ基本的な疑問は、このような RAID アレイの効果的なトランザクションレートは、何台までの SSD に対応するかということです。SSD 8 台の場合、LSI2208-1G コントローラーは、3 つすべてのランダム負荷プロファイルにおいて、単一 SSD の約 8 倍のトランザクションレートを達成します。つまり、まだ過負荷になっていません。LSI2108 コントローラーは、ランダム負荷プロファイルによっては単一 SSD のトランザクションレートの約 5~8 倍を達成します。つまり、これはすでにやや過負荷になっていることを意味します。

RAID 10 (SAS-6G-SSD × 8 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 10 の SSD アレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



ここでも、LSI2208-1G コントローラーは非常に優れたパフォーマンスを実現しており、特にブロックが小さい場合にはその威力を発揮します。

シーケンシャルアクセス

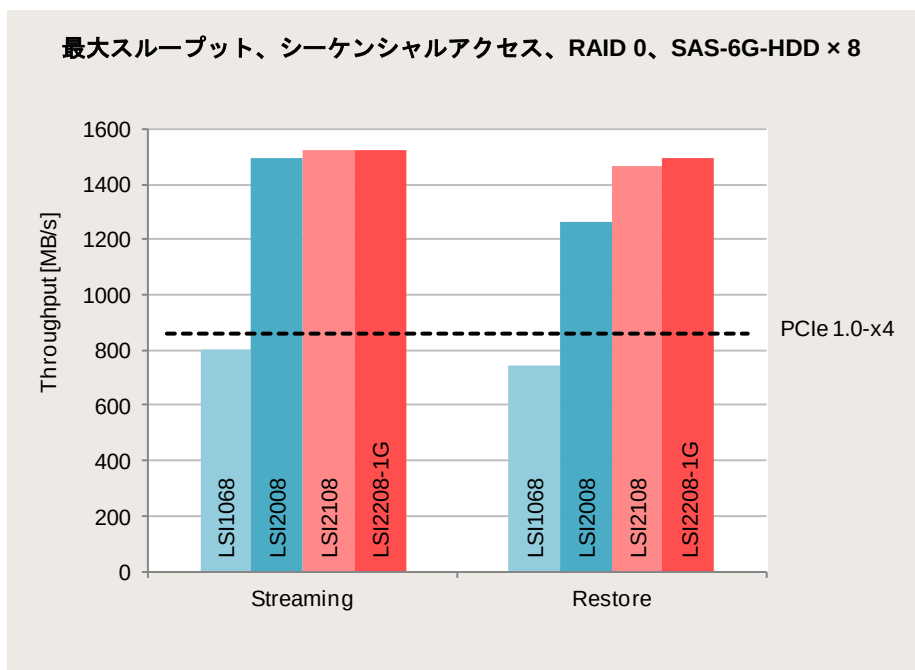
以降では、8 台のハードディスクでの測定に基づいて、シーケンシャルアクセスを行う場合のコントローラーの概要を説明します。この場合、HDD には SSD よりも 1 つ多くコントローラーがリリースされており、SSD の方が LSI2008 コントローラーのスループットの限界をよりわかりやすく説明できるので、HDD と SSD を区別して検討します。ここで予想される最大スループットに適切な値を掛けることで、他のハードディスクタイプや台数（8 台以下）で予想される最大スループットを計算できます。この方法で計算したスループットがコントローラーの限界値を超える場合は、コントローラーの限界値が有効になります。

HDD

RAID 0 (SAS-6G-HDD × 8 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 0 のハードディスクアレイにおけるスループットを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

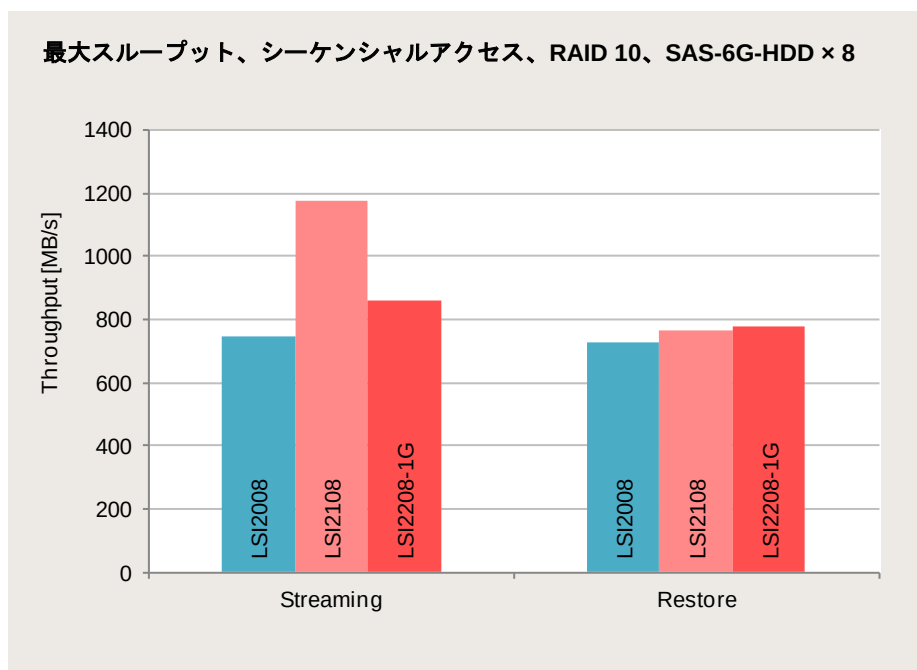
このグラフは、LSI1068 コントローラーは 800 MB/s 以上のシーケンシャルスループットには不十分であることを示しています。これは、PCIe インターフェースのスループットの限界（PCIe 1.0、x4、黒の破線）によるものです。そのため、高いスループットを達成するには、他の 3 つのコントローラーのうちいずれかを使用する必要があります。



RAID 10 (SAS-6G-HDD × 8 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 10 のハードディスクアレイにおけるスループットを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

シーケンシャルリードおよびライトについては、この RAID アレイに対して検討されている 3 つすべてのコントローラーが、単一 HDD の最大スループットの約 4 倍（つまり、この場合は約 800 MB/s）を達成または超過しています。

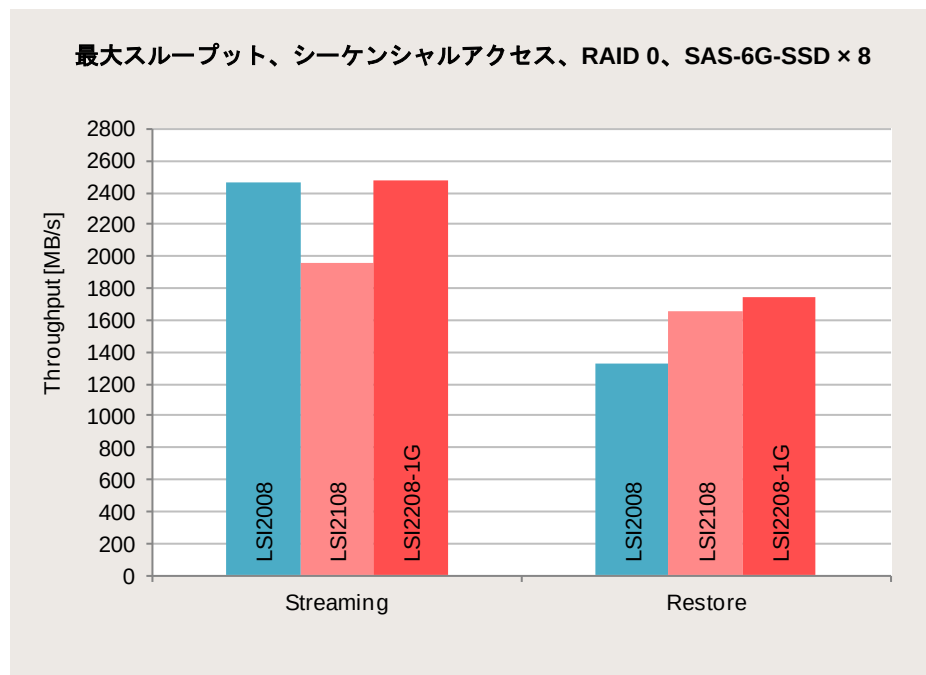


SSD

RAID 0 (SAS-6G-SSD × 8 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 0 のハードディスクアレイにおけるスループットを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

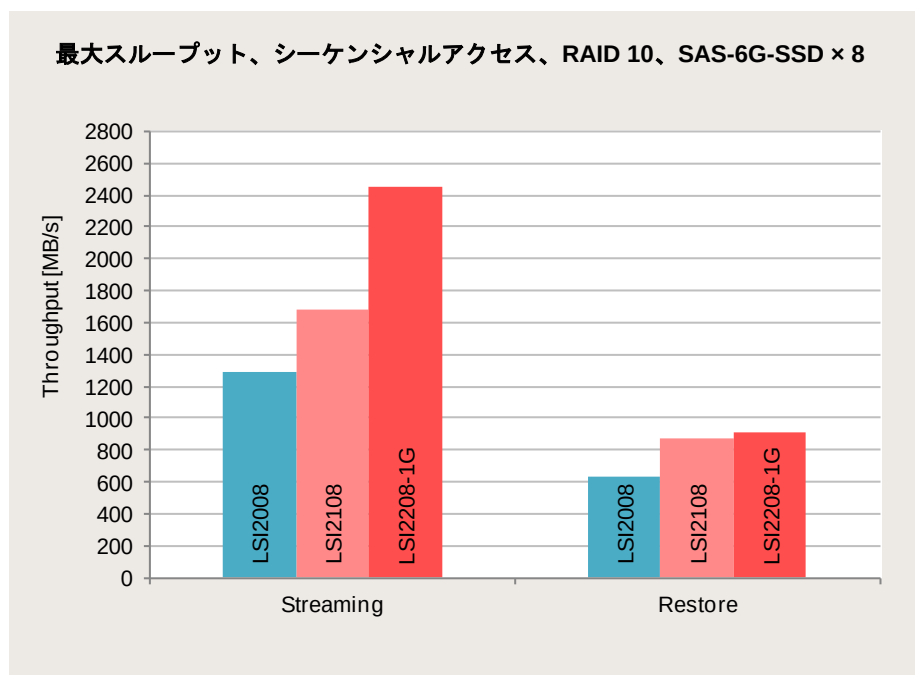
この場合は全体として、LSI2208-1G コントローラーが最高のパフォーマンスを実現しています。



RAID 10 (SAS-6G-SSD × 8 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 10 のハードディスクアレイにおけるスループットを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

この場合も、LSI2208-1G コントローラーが最高のパフォーマンスを実現しています。



RAID 0 および 10 (9 台以上のハードディスクで構成)

現在の PRIMERGY サーバでは、9 台以上のハードディスクの操作に、最高のパフォーマンスを実現する 3 つのコントローラー (LSI2108、LSI2208-1G-2.0、LSI2208-1G-3.0) を使用できます。これらのコントローラーでは、現在のところ、最大 24 台の内蔵ハードディスクによるサーバ構成が可能です (PRIMERGY TX300 S7 などを使用)。RAID 0 構成での比較には 24 台のハードディスクを使用し、RAID 10 構成での比較には 16 台のハードディスクを使用します (RAID 10 構成の場合、16 台が単一 RAID アレイの現在の上限です)。結果の値は、多数のハードディスクを使用した場合のパフォーマンス値の範囲をカバーしています。前の項と同様、説明には、高パフォーマンスの SAS-6G-HDD または SAS-6G-SSD を使用した測定が取り上げられています。これらのハードディスクについての詳細は「[測定環境](#)」を参照してください。

RAID 0 構成の場合、キャッシュ付きのコントローラーでは通常、コントローラーキャッシュを無効にした方が最大パフォーマンスの値が大きくなります。このため、ここでは LSI2108 および LSI2208-1G コントローラーのトランザクションレートは、ServerView RAID の「Performance」モードの[標準設定](#)を次のように変更したうえで測定しています。

- Read mode : 「Read-ahead」から「No read-ahead」に変更
- Write mode : 「Always write-back」から「Write-through」に変更

RAID 10 構成では[標準設定](#)を使用しています。

ランダムアクセス

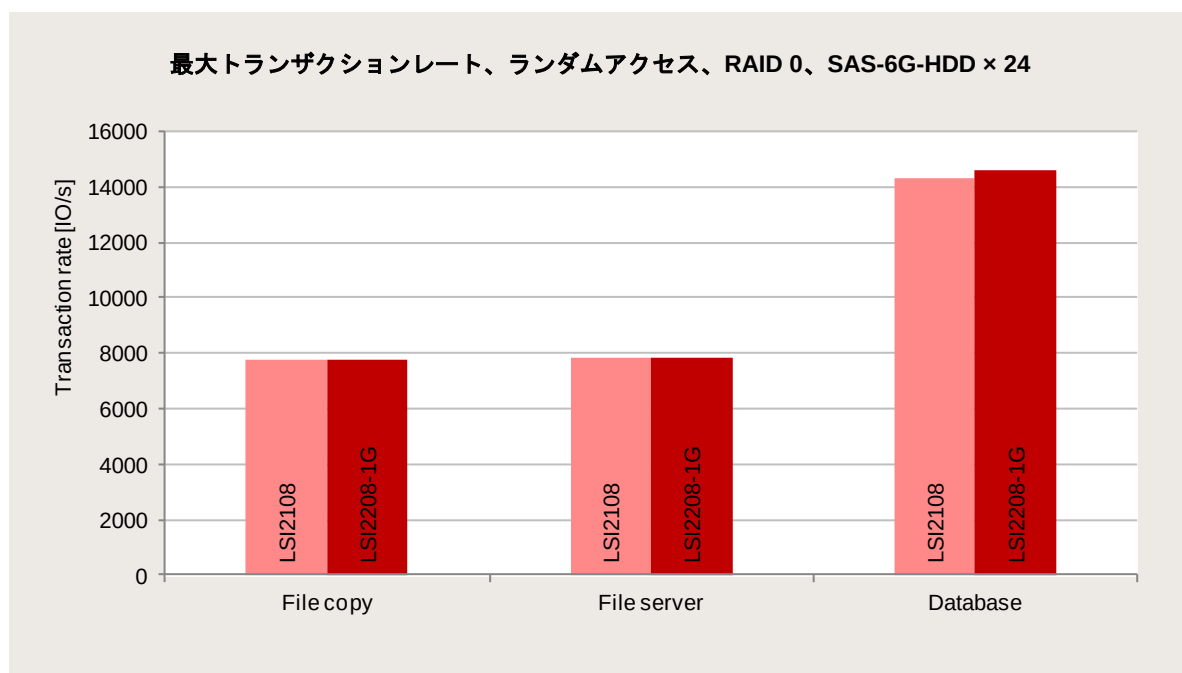
多数のハードディスクへのランダムアクセスでは、SSD の限界が大幅に異なるため、HDD と SSD を区別して検討します。

HDD

以降では、コントローラーによる HDD へのランダムアクセスを比較しています。ここでは、各負荷プロファイルでのストレージメディアの最大トランザクションレートが、最も重要な制限要因になります。次に示す測定結果は 24 台の SAS-6G-HDD で得られたものですが、異なるタイプやハードディスク数（24 台以下）での最大トランザクションレートを見積もる場合にも使用できます。HDD へのランダムアクセスで得られるスループットは非常に低いため、コントローラーの PCIe または SAS インターフェースの限界には達しません。

RAID 0 (SAS-6G-HDD × 24 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 0 のハードディスクアレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 %リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 %リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 %リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



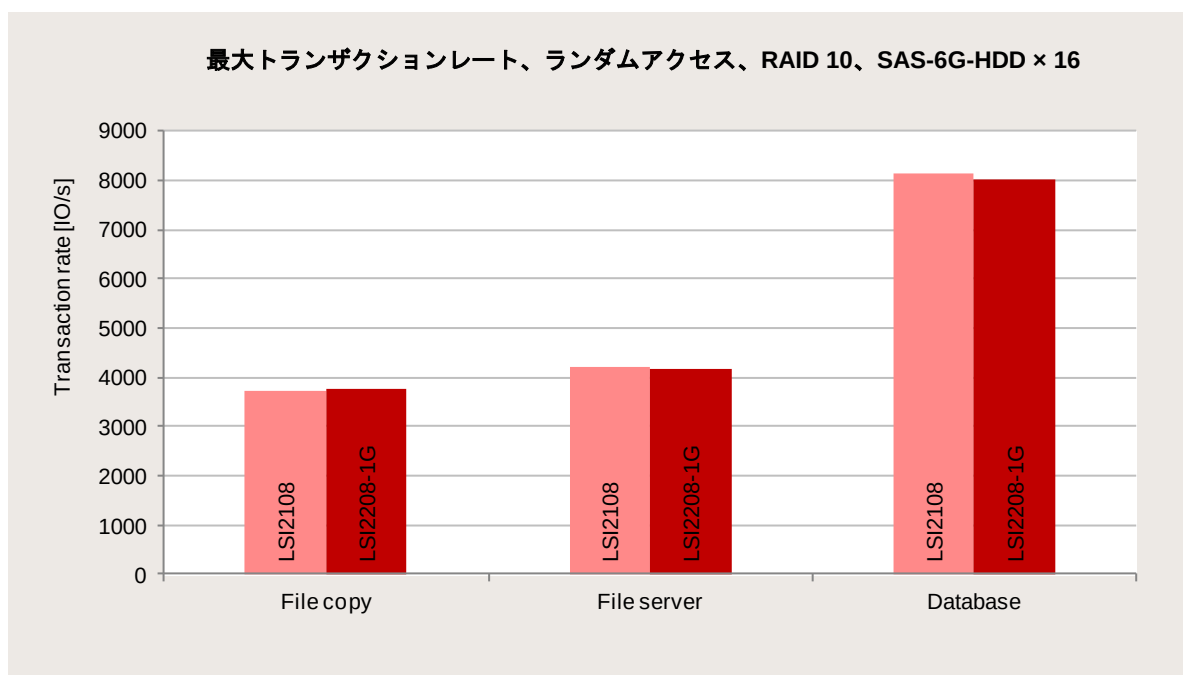
使用した SAS-6G ハードディスクの最大トランザクションレートは、「データベース」負荷プロファイルで 594 IO/s になります。したがって、このハードディスク 24 台で構成されるアレイは、理論上は最大で $594 \text{ IO/s} \times 24 = 14256 \text{ IO/s}$ のトランザクションレートを達成できることになります。「データベース」負荷プロファイルで測定した最大トランザクションレートは 14396 IO/s なので、この予測値が実測値に非常によく近似していることがわかります。3 つすべての負荷プロファイルにおいて、使用されている HDD タイプの最大トランザクションレートは比較的近いものの、ここに示す測定では、ブロックサイズが 64 KB の負荷プロファイルの場合、ブロックサイズが 8 KB の負荷プロファイルの約半分のトランザクションレートしか達成できません。²

² これは、RAID アレイが 64 KB の基本ブロックサイズ（「ストライプサイズ」）で初期化されるためです。「ストライプサイズ」より小さい 8 KB のブロックサイズでのランダム負荷プロファイルは、「HDD の数 × HDD パフォーマンス」という式で直接予測できます。RAID アレイを初期化したときと同じブロックサイズを負荷プロファイルで使用すると、約半分のトランザクションレートしか達成できません。この現象の詳細な理由については、ここでは触れません。

すべてのコントローラーは、各負荷プロファイルでこの RAID アレイの最大トランザクションレートをほぼ達成しています。

RAID 10 (SAS-6G-HDD × 16 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 10 のハードディスクアレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



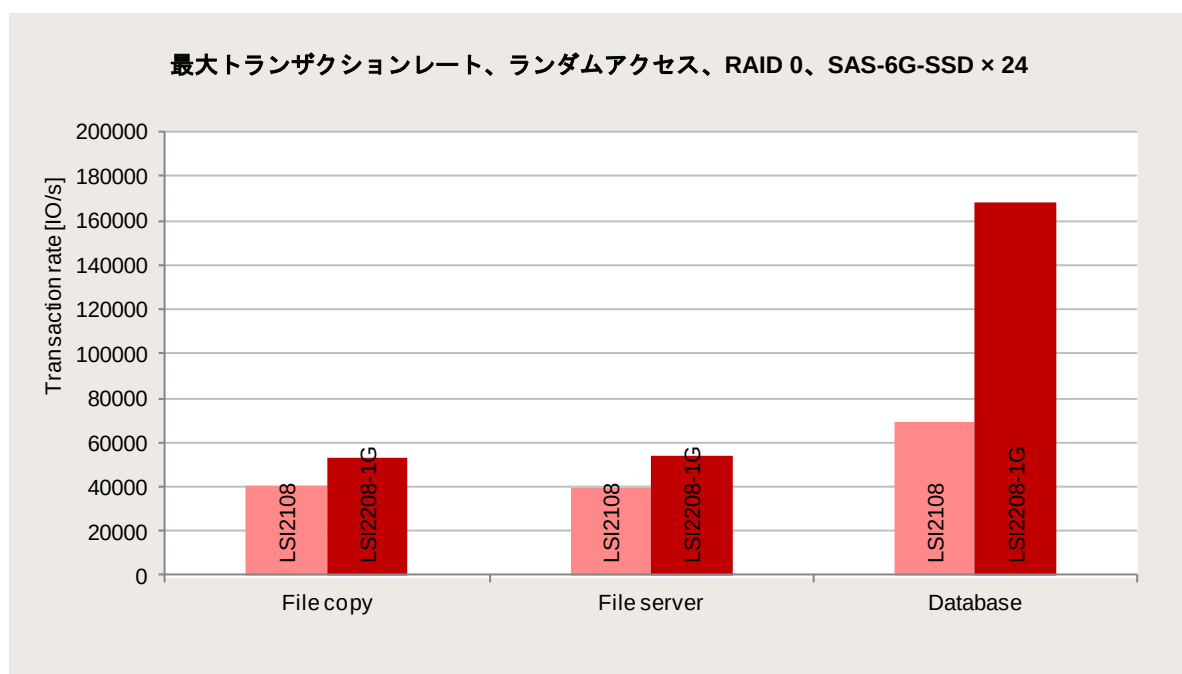
グラフは、RAID 0 と同様の傾向を示しています。

SSD

通常、個別の HDD にランダムアクセスを行った場合の最大トランザクションレートは、ハードディスクキャッシュを有効にした場合でも 700 IO/s 未満ですが、SAS-6G-SSD の場合のトランザクションレートはその約 20 倍になります。このようにトランザクションレートが高いため、複数の SAS-6G-SSD でアレイを構成すると、ランダムアクセスで数百 MB/s のスループットを提供できます。つまり、コントローラーのリソースとインターフェースに HDD よりはるかに高い負荷がかかるため、コントローラー世代間で差が生じます。

RAID 0 (SAS-6G-SSD × 24 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 0 の SATA-SSD アレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



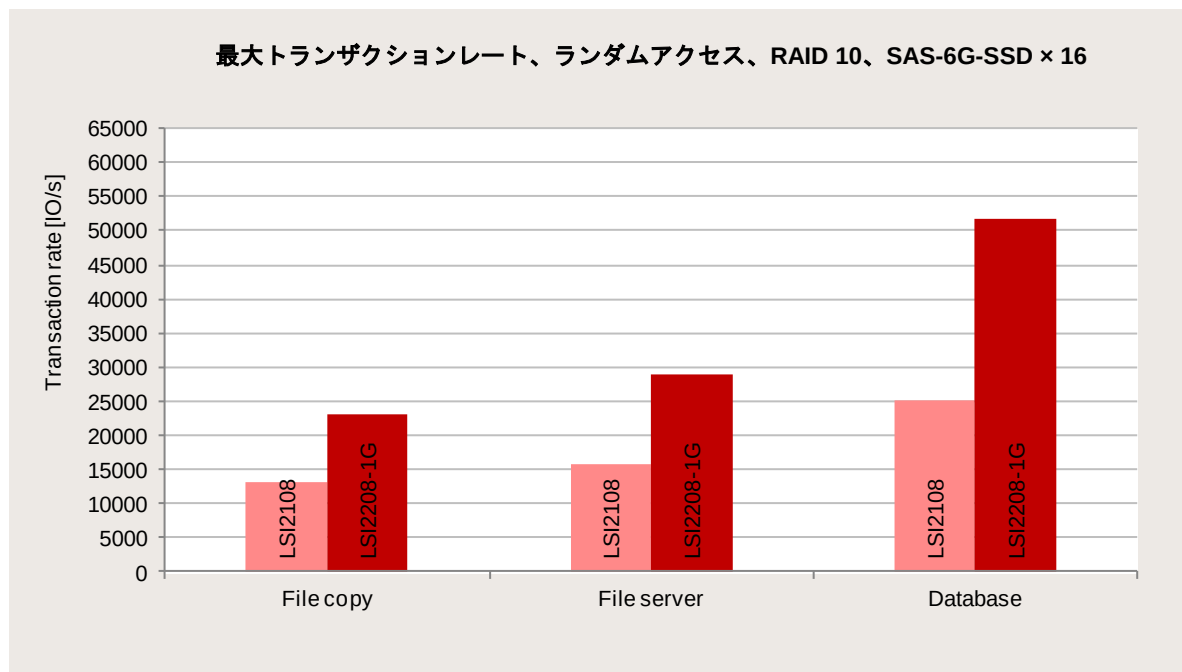
ここでは、LSI2208-1G コントローラーが最高のパフォーマンスを実現しています。ブロックサイズが小さい負荷プロファイル（「データベース」）の場合、LSI2208-1G コントローラーは LSI2108 コントローラーの約 2.5 倍のトランザクションレートを達成しています。

SAS-6G-SSD の数で表すと、例えば RAID 0 構成で LSI2208-1G コントローラーの速度性能を最大限活用するには、ランダム負荷プロファイルに応じて、13~17 台の SAS-6G-SSD が必要です。これにより、ランダム負荷プロファイルで 8 台の SSD をサポートする LSI2208-1G がまだ過負荷になっていないという「[RAID 0 および 10（ハードディスク 5~8 台で構成）](#)」における説明が裏付けられます。

これらのトランザクションレートを変換することで生成されるスループット値も重要です。トランザクションレートが低くても、64 KB ブロックサイズの 2 つの負荷プロファイルではスループットが高くなります。例えば、LSI2208-1G コントローラーは、「ファイルコピー」負荷プロファイルで約 3200 MB/s のスループットを処理します。この値は、100 % リード、100 % ライトの負荷プロファイルにおける、この RAID レベルでのコントローラーの 2 つの最大シーケンシャルスループットよりも高いものであり、注目に値します。この値は、SAS 接続を実際に双方向で使用しない限り、達成できない値です。

RAID 10 (SAS-6G-SSD × 16 で構成)

次のグラフは、ランダム負荷プロファイルでの RAID 0 の SSD アレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



ここでは、LSI2208-1G コントローラーが最高のパフォーマンスを実現しています。

8 台の SAS-6G-SSD で構成した RAID 10 の場合にも、類似したグラフが描かれます。ただし、SSD の数を 2 倍の 16 台にしても、トランザクションレートはわずかに増えません。そのため、RAID 10 構成におけるランダム負荷プロファイルで最高のパフォーマンスを達成するには、コントローラーごとに 11 台以上の SSD を接続しないことや、LSI2208-1G コントローラーを使用することが推奨されます。

シーケンシャルアクセス

多数のハードディスクへのシーケンシャルアクセスを検討する場合、通常は HDD と SSD を区別する必要はありません。2 つのハードディスクタイプの最大スループットはほとんど同じだからです。

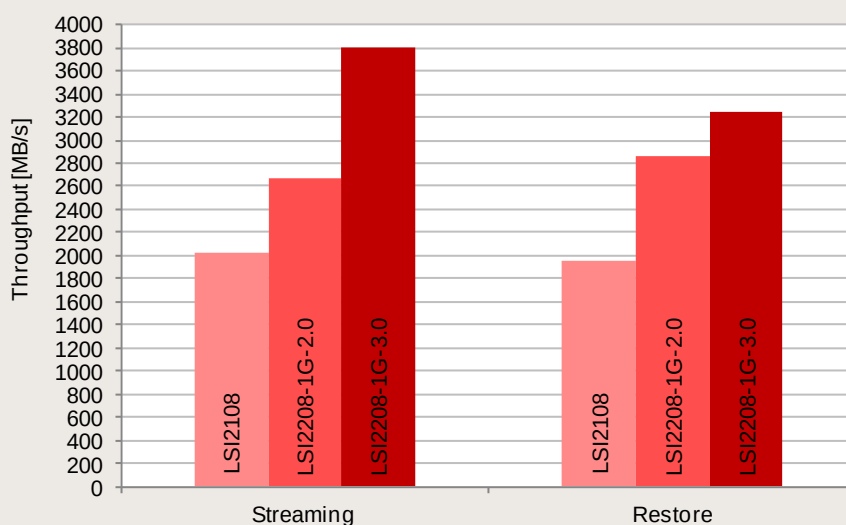
RAID 0 (SAS-6G-SSD × 24 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 0 のハードディスクアレイにおけるスループットを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

LSI2208-1G-3.0 コントローラーはここで最高のパフォーマンスを実現し、「ストリーミング」で PCIe Gen2 の上限を超えています。「リストア」の場合は、SSD アレイが限界となります。

このグラフを、8 台の SAS-6G-SSD で構成された RAID 0 の対応グラフと比較すると、LSI2108 コントローラーについては今回の結果とほぼ同じ値が達成されていたことがわかります。

最大スループット、シーケンシャルアクセス、RAID 0、SAS-6G-SSD × 24

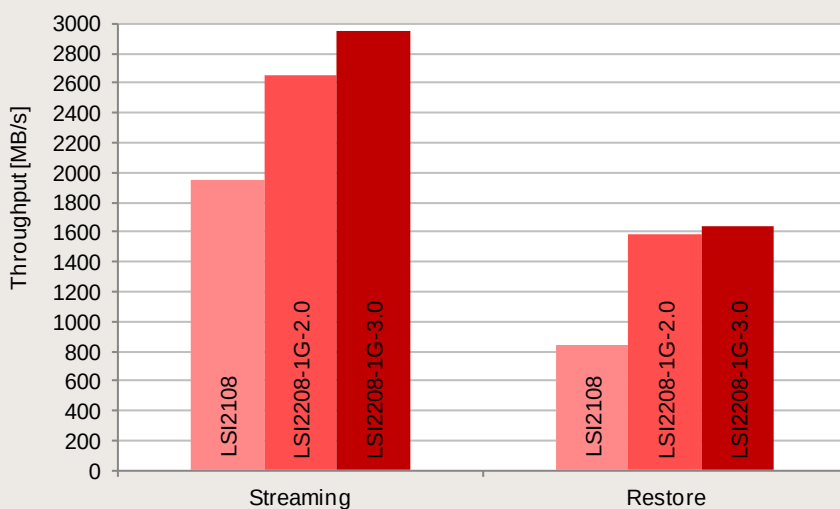


RAID 10 (SAS-6G-SSD × 16 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 10 のハードディスクアレイにおけるスループットを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。

LSI2208-1G-3.0 コントローラーは PCIe Gen3 の影響により、ここでも最高のパフォーマンスを実現しています。ただし、特に負荷プロファイル「リストア」の場合、高いスループットの限界を最大限に利用するには SSD 16 台では少なすぎるため、Gen2 との違いはそれほど大きくありません。

このグラフを、再び 8 台の SAS-6G-SSD で構成された RAID 0 の対応グラフと比較すると、LSI2108 コントローラーについては今回の結果とほぼ同じ値が達成されていたことがわかります。

最大スループット、シーケンシャルアクセス、RAID 10、SAS-6G-SSD × 16

RAID 5

RAID レベル 0、1、10 では、コントローラーの機能は、データブロックの論理ポジションを RAID アレイの物理ポジションへ割り当てる動作（「ストライピング」）や、必要に応じて、ブロック全体のミラー形式でのライト（「ミラーリング」）に限定できます。

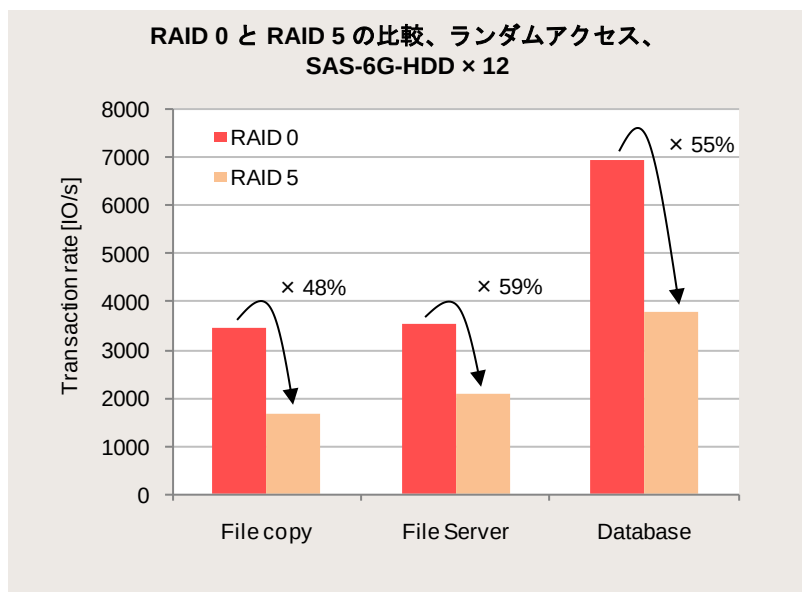
一方、RAID 5 のような RAID レベルでは、コントローラーは、非常に高い要件に対応する必要があります（特にライトについて）。こうした RAID レベルでは、パリティデータのブロックが必要であり、まず最初に実際のデータからパリティデータを計算しなければならないからです。

RAID 5 を標準でサポートするコントローラーは LSI2108 および LSI2208 のみなので、ここではこれらのコントローラーで測定しました。より高いパフォーマンス要件の最大値が特に重要なので、高パフォーマンスの SAS-6G-HDD または SAS-6G-SSD による測定に基づいて解説しています。これらのハードディスクについての詳細は「[測定環境](#)」を参照してください。

ランダムアクセス

RAID 5 のアプリケーションシナリオでは、ライトリクエストにより、RAID アレイへのアクセスが大幅に増加します。これはランダムアクセスに特に影響があります。この結果、同数のハードディスクによる RAID 0 構成と RAID 5 構成を比較した場合、アプリケーション側からみると、RAID 5 で達成可能な最大トランザクションレートが一定の割合まで低下します。次の例では RAID 5 でランダム負荷プロファイルを使用した場合の比較に限定してこれらの割合を示していますが、コントローラーによっても割合は異なります。

次のグラフは、最初に RAID 0 として構成され、その後に RAID 5 として構成された 12 台の SAS-6G-HDD アレイの最大測定トランザクションレートを示しています（どちらの RAID でも LSI2108 コントローラーと接続した場合）。



RAID 0 での測定値は、この HDD アレイの技術的な最大値をほぼ達成しています。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 %リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 %リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 %リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。RAID 5 での測定値は、各負荷プロファイルでそれぞれの RAID 0 の値の一定の割合（%）に留まっています。グラフの各矢印の上に、割合（%）を記載しています。

コントローラーにおいて、この割合は、50 %ライトの負荷プロファイル（「ファイルコピー」）で 48 %相当になります。また 33 %ライトの負荷プロファイル（「ファイルサーバ」と「データベース」）では、それぞれ 59 %と 55 %相当になります。ライトリクエストの増倍率に基づくと、これらの割合の理論値は、「ファイルコピー」では 40 %、「ファイルサーバ」および「データベース」では 50 %と推計できます。したがって、コントローラーは、すべてのケースで理論値よりも明らかに高い割合を達成しています。なお、この割合はコントローラキャッシュが有効の場合にのみ実現されます（標準では有効に設定されています）。これらのパーセンテージが高いほど、コントローラー設計が良いと言えます。

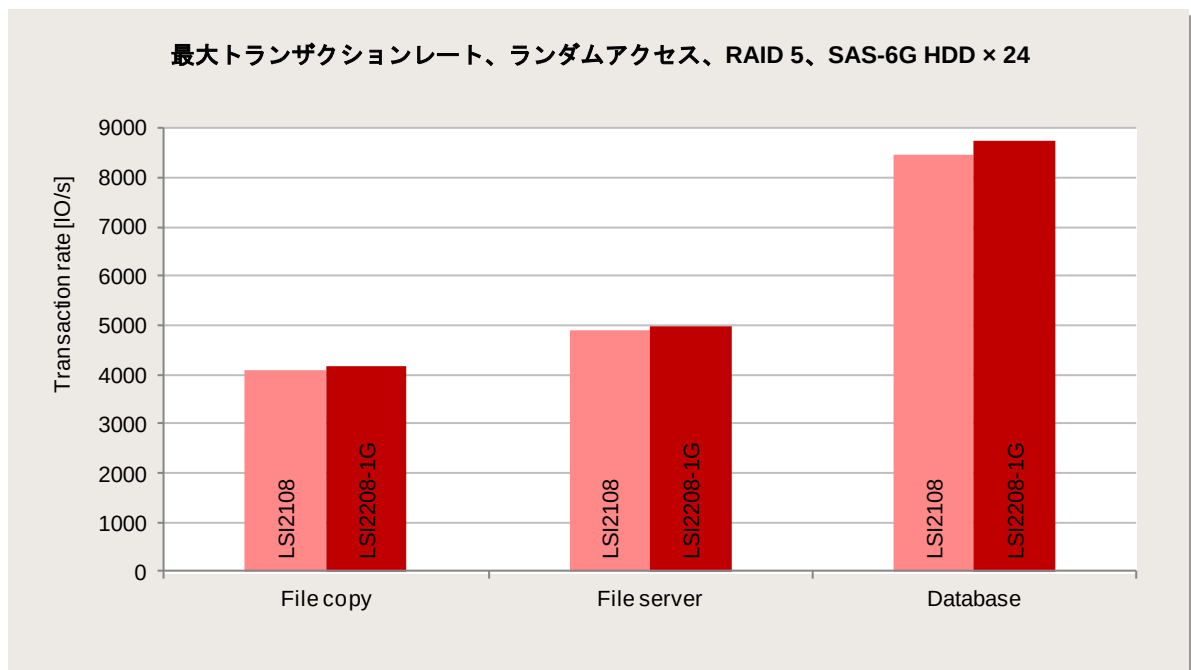
40 %と 50 %の理論的なパーセンテージの詳細な予測方法は、ここでは扱いません。

HDD

RAID 5 のランダム負荷プロファイルで達成できるパフォーマンス値について、まず HDD のケースで説明します。RAID レベル 0、1、10（パリティ計算なし）の場合と同様、RAID 5 においても HDD と SSD の最大パフォーマンスには大きな違いがあります。このため、HDD と SSD を分けて検討します。

RAID 5（SAS-6G-HDD × 24 で構成）

次のグラフは、ランダム負荷プロファイルでの RAID 5 のハードディスクアレイにおけるトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



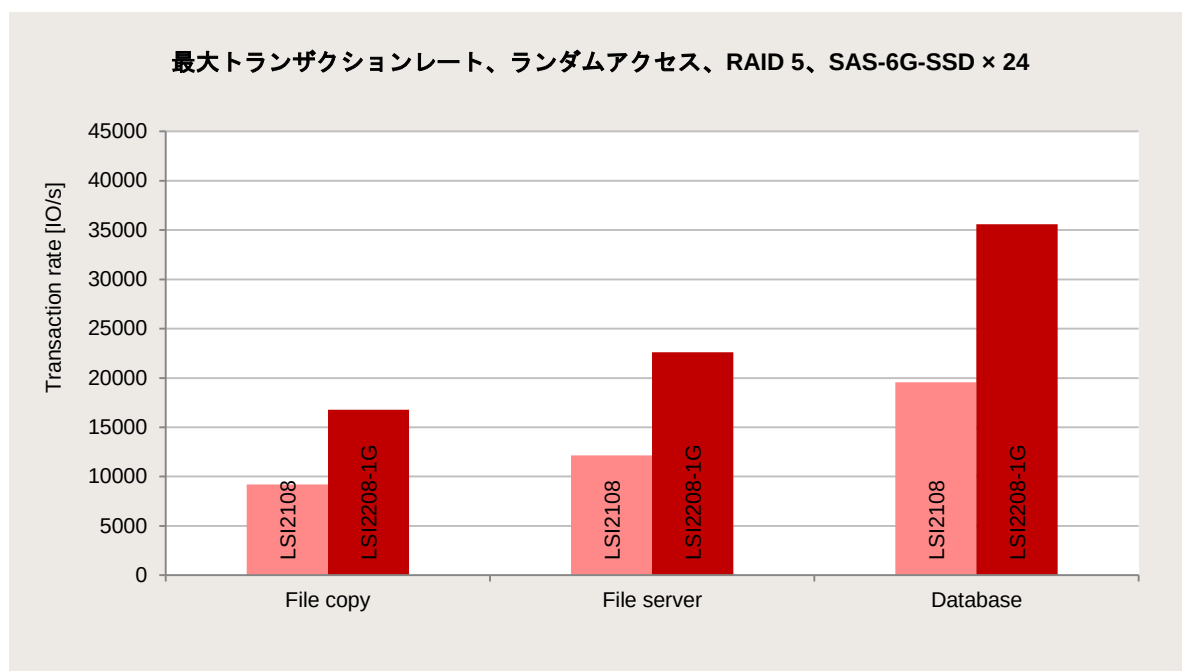
キャッシュが増加された最新コントローラー LSI2208-1G が、わずかに優れています。

SSD

ここでは、SAS-6G-SSD で達成できるトランザクションレートについて説明します（前述と同様、RAID 5 のランダム負荷プロファイルでのケースです）。前述の RAID 0、10 と同様に、RAID 5 でも SAS-6G-SSD の使用によりコントローラー間の差が明確になります。

RAID 5 (SAS-6G-SSD × 24 で構成)

次のグラフは、24 台の SAS-6G-SSD で構成したアレイを使用する各コントローラーの最大トランザクションレートをまとめたものです。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。



24 台の SSD で構成した RAID 0 のときと同様、このケースでも、旧世代から新世代への技術的進歩には目を見張るものがあります。標準負荷プロファイル「データベース」の場合、LSI2208-1G コントローラーは LSI2108 コントローラーの最大 2 倍のトランザクションレートを提供します。

SAS-6G-SSD の数で表すと、LSI2208-1G コントローラーは RAID 5 構成において、コントローラーでボトルネックを発生させずに、ランダム負荷プロファイルによって 6~16 台の SAS-6G-SSD にフル負荷を与えることができます。この推計は、「[測定環境](#)」で説明している SAS-6G-SSD 固有の最大値と、「[ランダムアクセス](#)」の冒頭に示した RAID 5 における最大トランザクションレートの低下の割合に基づいています。

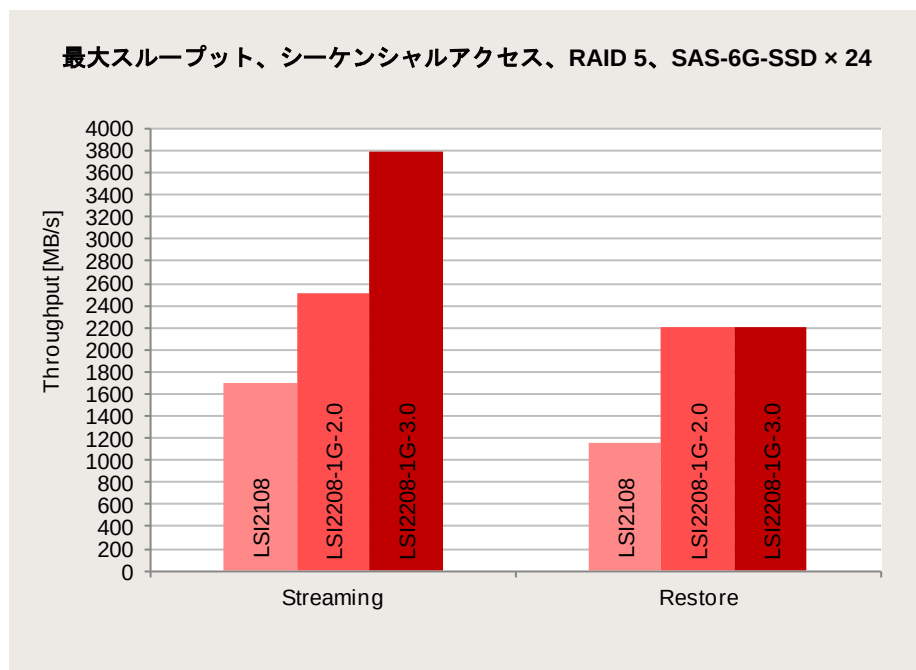
シーケンシャルアクセス

多数のハードディスクで RAID 5 を構成したときのシーケンシャルアクセスを検討する場合も、それぞれの最大値がほとんど同じなので、HDD と SSD を区別する必要はありません。以降では、台の SAS-6G-HDD での測定に基づいて、各コントローラーの概要を説明します。他のハードディスクタイプや台数（24 台以下）で予想される最大スループットを計算できます。予想されるスループットがコントローラーの限界値を超える場合は、コントローラーの限界値が有効になります。

RAID 5 の場合、N 台のハードディスクで構成したアレいの、論理的に関連付けられたブロックグループでは、常に (N-1) 個のブロックに実データ、1 個のブロックにパリティデータが格納されます。このため、このようなグループが完全にシーケンシャルで処理される場合、どのケースでも、1/ (N-1) の割合でパリティデータが含まれます。したがって、このようなシーケンシャル処理では、アプリケーションで使用するハードディスク 1 台でのスループットの (N-1) 倍を超える値は得られません。

RAID 5 (SAS-6G-SSD × 24 で構成)

次のグラフは、シーケンシャル負荷プロファイルでの RAID 5 のハードディスクアレイにおけるスループットを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 2 つのグループは、それぞれ標準負荷プロファイル「ストリーミング」（シーケンシャルアクセス、100 % リード、64 KB ブロックサイズ）と「リストア」（シーケンシャルアクセス、100 % ライト、64 KB ブロックサイズ）でのスループットを示しています。



ここでは、LSI2208-1G-3.0 コントローラーの優れたメリットが明確になっています。RAID 0 構成では、シーケンシャルリードで 3800 MB/s が達成されており、PCIe Gen2 の限界を明らかに超えています。

この RAID レベルのシーケンシャルライトでは、LSI2108 コントローラーは約 1200 MB/s、LSI2208-1G コントローラーは約 2200 MB/s の最大スループットを達成しています。RAID 5 構成の場合、コントローラーの速度は、比較的ゆがめられることなくパリティブロックの計算に反映されているので、このシーケンシャルライトの最大値は RAID コントローラーのパフォーマンスを示す重要なインジケータであると言えます。

すでに述べたように、3 ~ 24 台のハードディスク (N) を使用する場合、このシーケンシャル負荷プロファイルの予想されるスループットは、ハードディスク 1 台でのスループットの (N-1) 倍で計算できます。この場合、コントローラーの限界値を超えることはありません。「[測定環境](#)」の値から、本書の土台となっている SAS-6G-HDD および SAS-6G-SSD の検討結果として、例えば、負荷プロファイル「リストア」の場合、LSI2208-1G コントローラーはそれぞれ 7 台以上の HDD、8 台以上の SSD に使用するべきである、といったことがわかります。

低負荷レベル

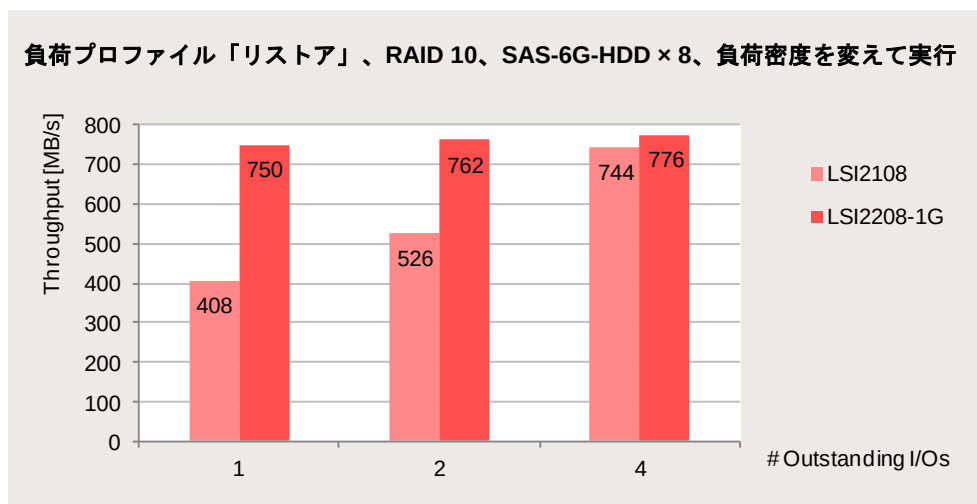
コントローラー間の差は、高負荷時の最大達成値だけでなく、低負荷時にも見られます。そのため、次の表では、同時アクセスなし（下記では「同時アクセス（未処理 I/O）× 1」と表記）で動作する、単一アプリケーションのシーケンシャルスループットについて、個々のコントローラの最大値をまとめました。これらの最大値は、コントローラーがサポートするすべての RAID レベルに適用されます。また、この値が超過されることはありません。これらのスループットは、RAID 0 でのみ得られます。RAID 0 以外の RAID レベルでは、多くの場合、スループットは最大値未満に留まります。

コントローラー	シーケンシャルスループットの最大値 同時アクセス（未処理 I/O）× 1、64 KB ブロックサイズ	
	100 %リード （負荷プロファイル「ストリーミング」）	100 %ライト （負荷プロファイル「リストア」）
LSI1068	230 MB/s	180 MB/s
LSI2008	310 MB/s	340 MB/s
LSI2108	670 MB/s	670 MB/s
LSI2208-1G	930 MB/s	890 MB/s

シーケンシャルリードでこれらの値を達成するためには、「Read-ahead」設定が前提条件となります（シーケンシャルライトの場合は「Write-back」設定）。これらの最大スループットはブロックサイズにも大きく依存しますが、サイズに関する表の値の相互関係は、他のブロックサイズの場合でも同様になります。

「同時アクセス（未処理 I/O）× 1」で 900 MB/s 以上のシーケンシャルスループットが得られる RAID アレイの場合、表の最後の 2 つのコントローラー間の差はさらに大きくなります。この場合、コントローラーの選択を誤ると、制限に影響が出ます。

このことについて、次の例で説明します。この例は、SAS-6G-HDD 8 台で構成した RAID 10 アレイに負荷プロファイル「リストア」（シーケンシャルアクセス、100 %ライト、64 KB ブロックサイズ）を適用した場合のスループットの測定に基づいています。比較は、LSI2108 および LSI2208-1G コントローラー間で、同時アクセスの数（「処理待ち IO 数」）を変えながら行いました。



このケースでは、同時アクセス（未処理 I/O）が 1 および 2 の場合、LSI2108 コントローラーは、LSI2208-1G コントローラーと同等のスループットを達成していないことが明確にわかります。LSI2208-1G コントローラーはすでに 750 MB/s のスループットを達成していますが、LSI2108 コントローラーのスループットはその半分を少し超える程度です。

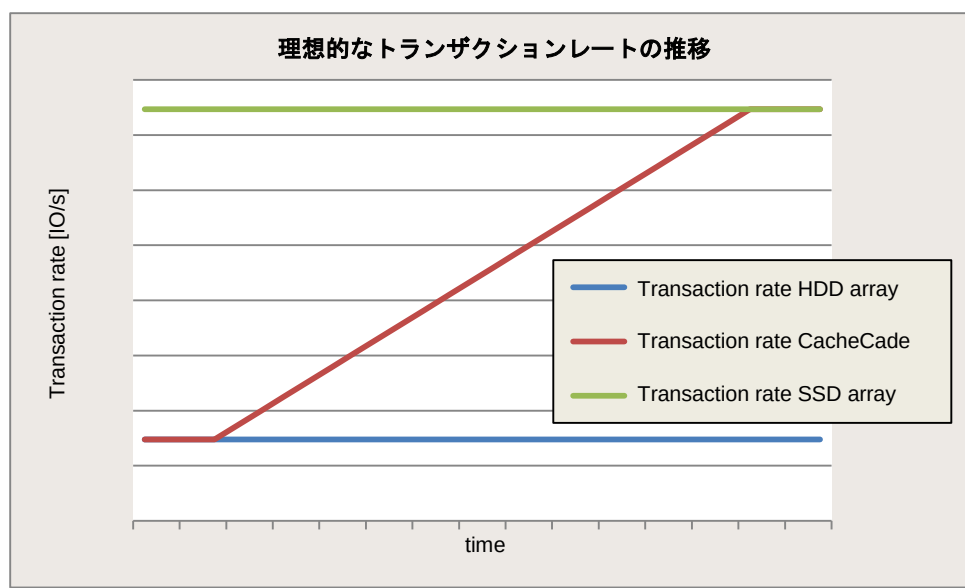
応答時間の観点からみると、低負荷密度では、LSI2208-1G コントローラーの応答時間は LSI2108 の約半分です。

MegaRAID アドバンスドソフトウェアオプション

CacheCade 2.0

CacheCade 2.0 を使用するかどうかを判断するには、ランダムアクセスが主流の HDD で構成される RAID アレイの検討から開始します。データの安全性が確保される RAID レベルを使用する場合、または大容量の SATA-HDD が使用される場合、トランザクションレートがそれほど高くないと想定できます。これを高くするために、SSD と CacheCade 2.0 を使用することができます。ただし、データへのアクセスに、ホットスポットが含まれていること、つまり、データの小さな一部分へのアクセスが頻繁に行われていることが前提条件となります。CacheCade 2.0 の目的は、このホットスポットを SSD に保存することで、アクセスの大半が行われる SSD の高いトランザクションレートを達成することです。

データへアクセスすると、トランザクションレートは純粋な HDD アレイのレベルから開始され、時間の経過とともに、純粋な SSD アレイのレベルを達成します。理想的なケースでは、次のようになります。



SSD と比較すると、HDD アレイのトランザクションレートは低くなっています。データへアクセスすると、ホットスポットは CacheCade 2.0 キャッシュ、つまり SSD に少しずつ保存されます。そして SSD の比較的高いトランザクションレートを達成することができます。SSD で構成された、対応するアレイのトランザクションレートを達成することが理想的です。

ホットスポット

すでに述べたように、CacheCade 2.0 の重要な前提条件は、ホットスポットへのランダムアクセスが行われることです。これを認識すること、およびホットスポットのサイズの見積もりが、CacheCade 2.0 キャッシュを計画する際の最初のステップの 1 つになります。ホットスポットへのアクセスだけでなく、その他のアクセスも可能ですが、ホットスポットへのアクセスの割合が 85 % 以上である必要があります。時間の経過とともにホットスポットは移動する可能性がありますが、この場合はトランザクションレートが低く設定されます。ホットスポットのサイズは大き過ぎではありません。また、CacheCade 2.0 キャッシュに収まるサイズである必要があります。ホットスポットのサイズが最大で 100 GB の場合には特に効率的にバッファできることや、キャッシュのサイズはホットスポットのサイズの 2 倍以上である必要があることがわかっています。

トランザクションレートの向上

CacheCade 2.0 に関するさらに重要な側面は、HDD アレイと SSD アレイのトランザクションレートの違いです。この違いが比較的大きい場合、CacheCade 2.0 を使用すると特に効果的です。一方、HDD の数が多い場合など、HDD アレイがすでに高いトランザクションレートを実現するように設計されている場合、違いは小さくなります。

「[RAID 0 \(SAS-6G-HDD × 24 で構成\)](#)」では、標準負荷プロファイル「データベース」で約 14000 IO/s を達成していることなどがわかります。これは、「[測定環境](#)」からわかるように、SAS-6G-SSD のトランザクションレートと非常に正確に対応します。そのため、このような構成では CacheCade 2.0 を使用するメリットを得ることができません。

どのようなケースでも、トランザクションレートの向上を期待できるかどうか、トランザクションレートのわずかな向上が要件に合うかどうかを検討する必要があります。

時間的要因

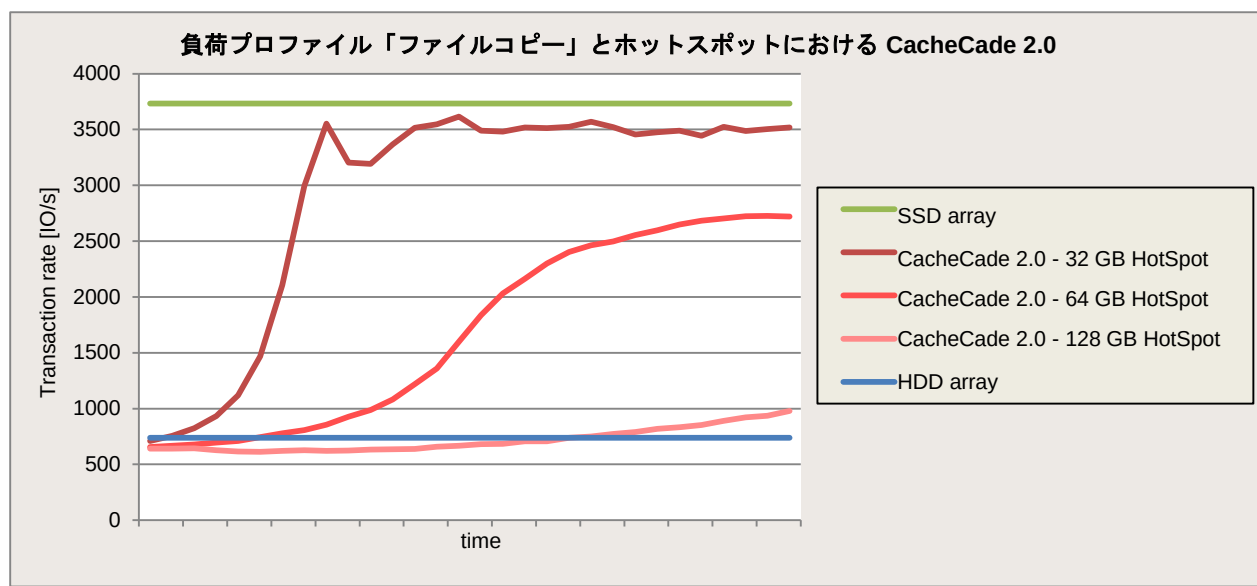
最終的には、時間的要因が重要になります。つまり、CacheCade 2.0 が高いトランザクションレート達成するまでの時間です。小さなホットスポットの場合、数分で達成することがあります。しかし、データ量が大量な場合や、トランザクションレートが低い HDD アレイの場合、数時間かかる可能性があります。

したがって、上記のグラフの転送時間は、複数の要因の影響を受けます。ホットスポットが大きいほど、CacheCade 2.0 キャッシュへの転送時間は長くなります。高パフォーマンス HDD アレイを使用すると、データはより短い時間で転送できます。さらに、アクセスの発生頻度、つまり、HDD アレイの最大トランザクションレートが使用されるかどうかにも影響を及ぼします。

また、どれだけの時間がかかるかは、ケースごとに、要件を考慮して検討する必要があります。

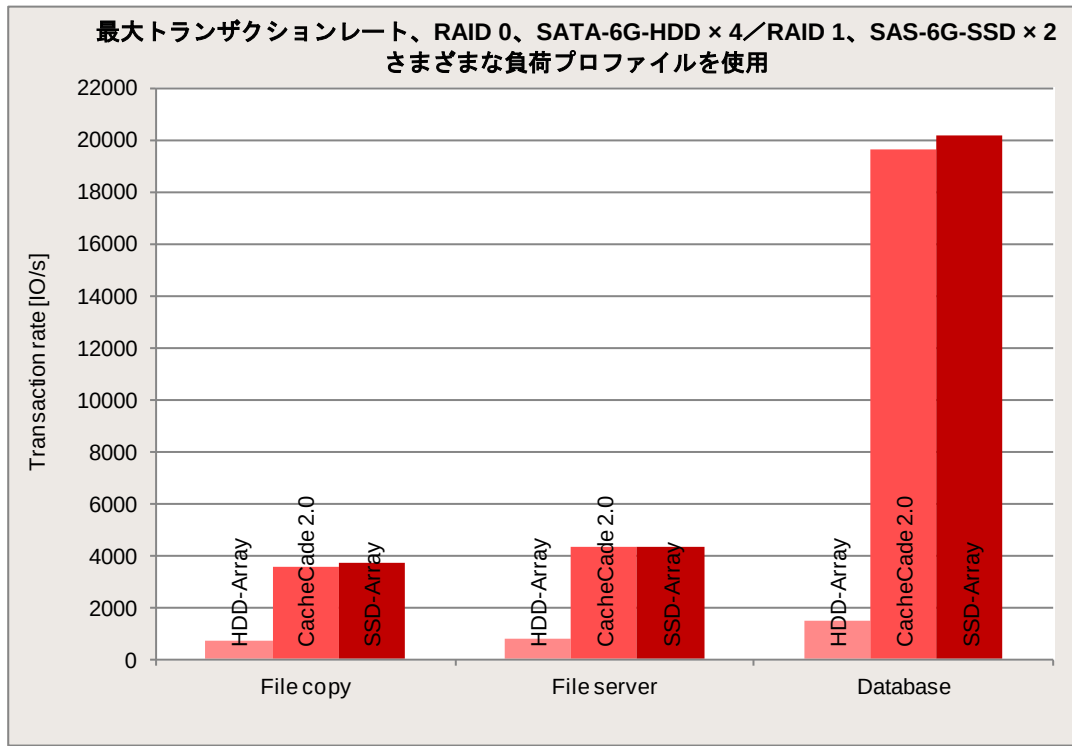
例 : SATA-HDD 4 台で構成される RAID 0 および SAS-SSD 2 台で構成される RAID 1

CacheCade 2.0 の典型的な構成例として、SATA-HDD 4 台で構成される RAID 0、および SAS-SSD 2 台で構成される RAID 1 があります。本書の「[RAID 0 および 10 \(SATA ハードディスク 4 台で構成\)](#)」および「[RAID 1 \(SAS ハードディスク 2 台\)](#)」では、この 2 つの構成とトランザクションレートについて詳細に説明しています。4 台の SATA-6G-HDD を使用すると、RAID 0 構成の合計ストレージ容量は約 4 TB になります。2 台の SAS-6G-SSD を使用した場合、RAID 1 構成の合計ストレージ容量は約 200 GB です。次のグラフは、32 GB、64 GB、128 GB のホットスポットへのアクセスが、CacheCade 2.0 を使用した負荷プロファイル「ファイルコピー」で発生した場合の、トランザクションレートの推移を示しています。



実際には、CacheCade 2.0 の一定のオーバーヘッドを考慮する必要があるため、SSD の最大トランザクションレートは達成されません。32 GB のホットスポットでは、非常に優れた結果となっています。64 GB のホットスポットの場合、トランザクションレートは HDD アレイよりも明らかに高くなっていますが、SSD のレベルには達していません。128 GB のホットスポットの場合、トランザクションレートの増加は非常に緩慢で、HDD アレイのトランザクションレートを超過するまでにはかなりの時間がかかります。

過去の測定では、負荷プロファイル「ファイルコピー」が代表例として選ばれました。ちょうど 32 GB のデータへアクセスしているという単純化した仮定に基づくと、「ファイルコピー」、「ファイルサーバ」、「データベース」の 3 つの標準負荷プロファイルの最大トランザクションレートは次のようになります。



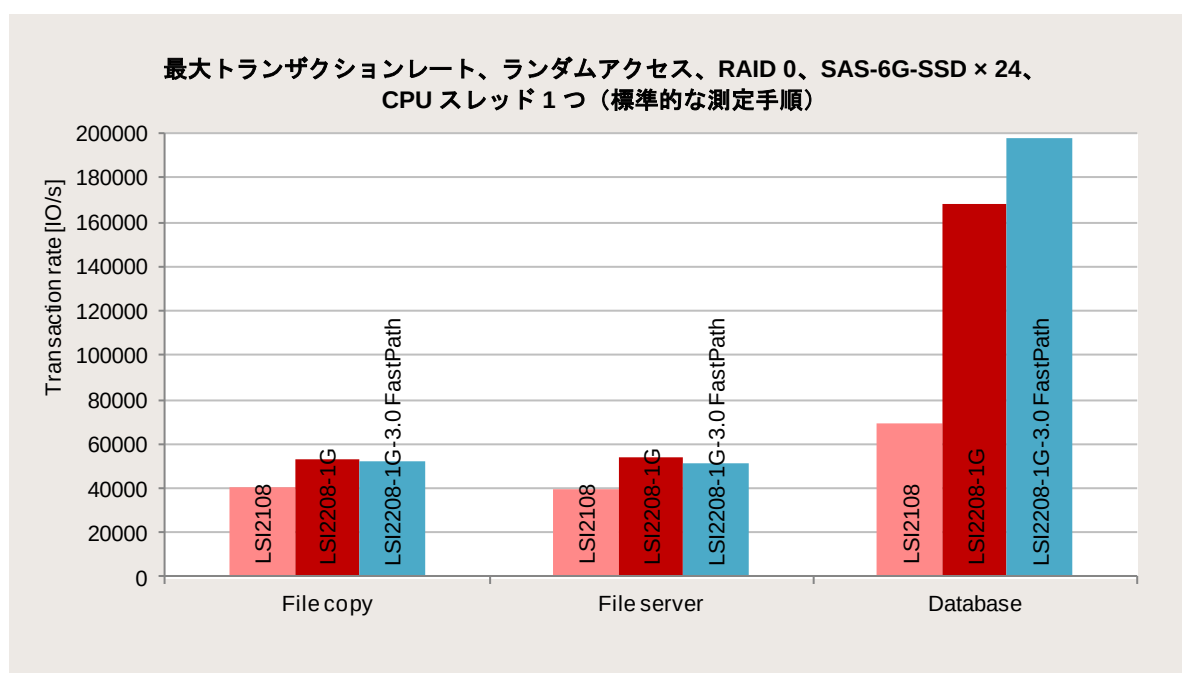
測定値は LSI2208-1G コントローラーを使用した場合のものです。HDD アレイは 4 台の SATA-6G-HDD で構成された RAID 0 であり、CacheCade 2.0 キャッシュは 2 台の SAS-6G-SSD で構成された RAID 1 となっています。CacheCade 2.0 キャッシュはリード/ライト用に構成されています。ケースごとに達成可能な最大トランザクションレートが示されています。データへのアクセスには、32 GB のホットスポットを仮定しています。

FastPath

多数のハードディスクを使用した例として、24 台の SAS-6G-SSD で構成された RAID 0 について詳細に検討し、「[RAID 0 および 10 \(9 台以上のハードディスクで構成\)](#)」に結果を示しました。ここでは、ランダムアクセスにおける最高のトランザクションレートが達成されており、FastPath との比較が適切となります。

RAID 0 (SAS-6G-SSD × 24 で構成) と FastPath

次のグラフは、RAID 0 構成である SSD アレイのランダム負荷プロファイルのトランザクションレートを示しています。このレートはさまざまなコントローラーで達成可能です。グラフの 3 つのグループは、それぞれ標準負荷プロファイル「ファイルコピー」（ランダムアクセス、50 % リード、64 KB ブロックサイズ）、「ファイルサーバ」（ランダムアクセス、67 % リード、64 KB ブロックサイズ）、「データベース」（ランダムアクセス、67 % リード、8 KB ブロックサイズ）でのトランザクションレートを示しています。ここには、「[コントローラーの比較](#)」で示した測定結果に基づいて、LSI2208-1G-3.0 コントローラーと FastPath を組み合わせて使用した結果を追加しています。



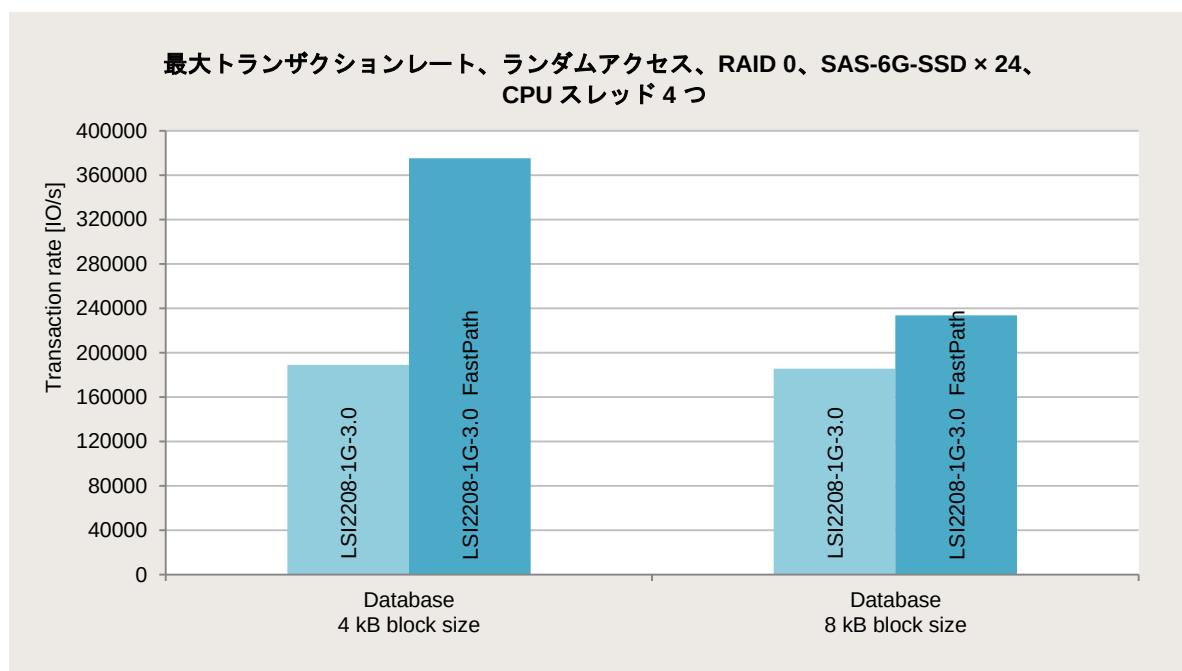
2 つの負荷プロファイル「ファイルコピー」および「ファイルサーバ」は、ブロックサイズ 64 KB のアプリケーションを示しています。すでに説明したとおり、LSI2208-1G コントローラーによるトランザクションレートは比較的低いにもかかわらず、3000 MB/s を超えるスループットレートが達成されています。これはパフォーマンスの限界を示しています。そのため、FastPath を使用してもトランザクションレートは向上しません。負荷プロファイル「データベース」は、ブロックサイズ 8 KB のアプリケーションを示しています。168000 IO/s のトランザクションレートで約 1310 MB/s のスループットが達成されており、まだ限界に近づいていません。標準的な測定手順により、FastPath を使用した結果、「データベース」負荷プロファイルではパフォーマンスの向上が見られます。「[測定方法](#)」で説明したとおり、同時にアクセスするアプリケーションをモデル化するため、「未処理 I/O の数」が使用されています。Iometer を使用した標準的な測定手順では、この数は、負荷を生成するただ 1 つの CPU スレッド内で、RAID アレイごとに変化します。トランザクションレートが約 200000 IO/s である場合、サーバの 1 つの CPU スレッドの使用率は 100 % です。実際のアプリケーションが同様に動作する場合、つまり、1 つの CPU スレッドで IO 負荷全体を処理する場合も、最大 200000 IO/s のトランザクションレートが達成され、FastPath を使用しても向上は見られません。

ただし、トランザクションレートが高いアプリケーションの場合、例えば標準データベースの場合には、通常、アクセスアクティビティは複数の CPU スレッドで行われます。「データベース」アプリケーションのモデル化ではこれに従い、Iometer 内で 4 つの CPU スレッドを使用します。ブロックサイズ 8 KB の補足として、ブロックサイズ 4 KB についても検討します。ブロックサイズが小さくなると、達成されるトランザクションレートは高くなります。ただし、4 KB よりも小さいブロックサイズはあまり一般的ではありません。

ServerView RAID の「Performance」モードの[標準設定](#)と比べて、ServerView RAID の指示により、ServerView RAID モードの「Fast Path optimum」モードを使用しています。これは現在、次のような設定になっています。

- Read Mode 「No read-ahead」
- Write Mode 「Write-through」
- Cache Mode 「Direct」
- Disk Cache 「Enabled」

次のグラフは、複数の CPU スレッドが使用されている場合に、負荷プロファイル「データベース」（4 KB および 8 KB ブロックサイズ）を実行したときの LSI2208-1G-3.0 コントローラーのトランザクションレートを示しています。



FastPath を使用しない場合の LSI2208-1G-3.0 コントローラーのトランザクションレートは、4 KB および 8 KB で 190000 IO/s に近い値になりました。前のグラフではたった 1 つの CPU スレッドで約 170000 IO/s をすでに達成しているので、システムの CPU ではなく、コントローラーが限界に影響を与えていることがわかります。これとは対照的に、4 KB ブロックサイズで FastPath を使用した場合には、360000 IO/s を超えるトランザクションレートが達成されています。8 KB ブロックサイズのアプリケーションの場合も、FastPath を使用して最大 240000 IO/s のトランザクションレートを達成しました。ただし、SAS 側のスループットは約 1900 MB/s です。

ここで示している「データベース」負荷プロファイルでは、ライト比率は 33 %でした。アプリケーションがより小さなライト比率を使用したり、1 KB または 2 KB ブロックサイズへのアクセスが発生することも十分考えられます。このような場合、絶対的なトランザクションレートも 400000 IO/s より高くなる可能性があります。

結論

PRIMERGY サーバは、「モジュラー RAID」のコンセプトによって、さまざまなアプリケーションシナリオの多様な要件を満たすことができます。

オンボードコントローラーは、RAID レベル 0、1、10 に対応する低価格なエントリーレベルの代替ソリューションです。PCIe スロットを 1 つ節約できますが、ハードディスクは 4 台に制限されます。プロセッサパフォーマンスの使用率は、新しいサーバになるほど重要性が低下します。

SAS/SATA 側では、現在のオンボードコントローラーは最大周波数 3G の標準をサポートします。最近の SAS-6G-SSD の場合、最適なスループットを達成するには、SAS-6G インターフェースを搭載したコントローラーも必要です。つまり、PRIMERGY BX920 S3 には「オンボード Intel C600 SAS RAID」コントローラーではなく、「PY SAS RAID HDD モジュール」コントローラー (LSI2208-512) などを使用すべきです。

キャッシュなしの「SAS 0/1」RAID コントローラーは、基本的な RAID ソリューションである RAID 0、1、1E、10 をそれぞれサポートし、これらの RAID レベルで高いパフォーマンスを発揮します。過去の世代の「LSI MegaRAID 8 ポートベースの RAID 0/1 SAS」コントローラー (LSI1068) だけは、サポートするシーケンシャルスループットが最大でわずか 800 MB/s 程度です。より高いスループットが必要な場合、適切な最新のコントローラーを使用する必要があります。

「SAS 5/6」RAID コントローラーは、現在のすべての標準 RAID ソリューション RAID 0、1、5、6、10、50、60 を提供します。これらのコントローラーにはコントローラーキャッシュが搭載され、オプションで BBU/FBU を使用したバックアップも可能です。キャッシュの使用を設定する多様なオプションにより、使用する RAID レベルに合わせた最適なパフォーマンスを柔軟に引き出すことができます。プロセッサパフォーマンスの使用率は、新しいサーバになるほど重要性が低下します。負荷密度の高い従来のハードディスクでランダムアクセスが発生する場合など、多くのアプリケーションシナリオで、これらのコントローラーは「SAS 0/1」RAID コントローラーよりも 70 %高いトランザクションレートを達成できます。

技術の進歩に従って、現在の「SAS 5/6」RAID コントローラーは 3 つのグループに分けられます。最初のグループは LSI SAS2108 チップを基盤とし、2 番目のグループは PCIe 2.0 向けの LSI SAS2208 チップを、3 番目のグループは PCIe 3.0 向けの LSI SAS2208 チップを基盤としています。最初のグループには「RAID Ctrl SAS 6G 5/6 512 MB (D2616)」コントローラーと「PY SAS RAID Mezz カード 6Gb」コントローラーが含まれます。2 番目のグループには、「RAID Ctrl SAS 6G 5/6 1GB (D3116)」コントローラー、「PY SAS RAID HDD モジュール」コントローラー、「PY SAS RAID HDD モジュール (キャッシュなし)」コントローラーが含まれます。3 番目のグループには、「RAID Ctrl SAS 6G 5/6 1GB (D3116C)」コントローラー、「PY SAS RAID HDD モジュール 3.0」コントローラー、「PY SAS RAID HDD モジュール (キャッシュなし) 3.0」コントローラーが含まれます。

ディスクサブシステムに負荷をかけるアプリケーションシナリオのほとんどは、ランダムリード/ライトアクセスが伴います。SSD を使用して非常に高い I/O レートを管理する場合は、コントローラーが最大トランザクションレートに大きく影響します。そのため、例えば「RAID Ctrl SAS 6G 5/6 1GB (D3116)」コントローラー (LSI2208-1G) は、最初のグループの最大トランザクションレートのほぼ 3 倍を部分的に達成します。したがって、「[コントローラーの比較](#)」のグラフを使用して、まだ「RAID Ctrl SAS 6G 5/6 512 MB (D2616)」コントローラー (LSI2108) で十分かどうかや、各負荷プロファイルや計画されている RAID アレイに応じて、1 つで 5 台以上の SSD を構成できる新しいコントローラーを用意する必要があるかどうかを確認する必要があります。

ハードディスクタイプにかかわらず、さまざまなコントローラーがそれぞれ、RAID レベルと負荷プロファイルに固有の最大シーケンシャルスループットを実現します。これらの最大値の一部は、「SAS 5/6」コントローラーの 3 つのグループごとに大幅に増加しています。例えば RAID 5 におけるシーケンシャルリードで、最初のグループは 1700 MB/s、2 番目のグループは 2500 MB/s を達成し、最終的には 3 番目のグループが 3800 MB/s を達成します。

計画されているアプリケーションシナリオで、単一のコントローラーで実現するトランザクションレートやスループットよりも高い値が必要とされる場合は、2 つのコントローラーを使用できます。このオプションを提供する PRIMERGY サーバは多数存在します (、PRIMERGY TX300 S7 など)。

より高速なコントローラーでは、シーケンシャルアクセスの負荷プロファイルで、同時アクセスが少ないケースで達成したスループットを達成できます。RAID アレイの効率が十分であれば、「RAID Ctrl SAS 6G

5/6 1GB (D3116) 」コントローラーを使用することにより、このようなアプリケーションで 900 MB/s を超えるリードおよびライトのスループットも実現可能です。これらの特殊なケースでは、旧世代のコントローラーと比較して、最大スループットを大幅に増加させることができます。

MegaRAID アドバンスドソフトウェアオプションは CacheCade 2.0 および FastPath という 2 つの機能で構成されていますが、これらは異なる最適化アプローチに従っているため、同じコントローラーで同時に使用することはできません。

CacheCade 2.0 は HDD アレイへのランダムアクセスを高速化しますが、一定の最小サイズが設定されています。1 台の HDD と 1 台の SSD という構成ではあまり意味がありません。一方で、大きな HDD アレイはそれだけで高いトランザクションレートを達成でき、CacheCade 2.0 キャッシュを使用するメリットは得られない点に注意が必要です。CacheCade 2.0 が特に有効なのは、32~64 GB のデータの小さな領域（ホットスポット）へアクセスする場合です。

CacheCade 2.0 とは対照的に、FastPath は、SSD アレイのようにトランザクションレートが高い場合に使用するように設計されています。ある構成で、200000 IO/s の最大トランザクションレートは超過していないが、SSD にはまだ適切なリザーブが存在し、SAS リソースが使い果たされていない場合、FastPath を使用してトランザクションレートを高めることができます。

コントローラーとハードディスクの構成には、PRIMERGY サーバに添付される RAID-Manager ソフトウェア「ServerView RAID」の使用を推奨します。このソフトウェアを使用すると、パフォーマンスやデータセキュリティに関する顧客要件に合わせて、コントローラーとハードディスクの設定を容易に調整できます。設定は、ほとんどのアプリケーションシナリオでコントローラーに依存せずに行えます。電源障害時の緩衝措置として BBU/FBU および UPS を使用している場合は、データの安全性を確保しながら最大パフォーマンスを得ることができます。

関連資料

PRIMERGY システム

<http://primergy.com/>

PRIMERGY のパフォーマンス

<http://www.fujitsu.com/fts/products/computing/servers/primergy/benchmarks/>

このホワイトペーパー :



<http://docs.ts.fujitsu.com/dl.aspx?id=e2489893-cab7-44f6-bff2-7aeea97c5aef>

<http://docs.ts.fujitsu.com/dl.aspx?id=71fac54b-7ec3-4b3f-b13d-f80fbb42d583>

<http://docs.ts.fujitsu.com/dl.aspx?id=e34159fa-0196-4a01-99ff-8792b5f644eb>

RAID のパフォーマンス (過去のホワイトペーパー、2011 年まで有効)

<http://docs.ts.fujitsu.com/dl.aspx?id=c852cc6b-e5af-46aa-ab60-4a5a154d42a8>

単一ディスクのパフォーマンス

<http://docs.ts.fujitsu.com/dl.aspx?id=de940140-2f25-4207-8862-563c4d91f30c>

ディスク I/O パフォーマンスの基本

<http://docs.ts.fujitsu.com/dl.aspx?id=35801735-a223-491a-a879-43f506444366>

Iometer についての情報

<http://www.iometer.org>

PC サーバ PRIMERGY (ブライマジー)

<http://jp.fujitsu.com/platform/server/primergy/>

お問い合わせ先

富士通

Web サイト : <http://jp.fujitsu.com/>

PRIMERGY のパフォーマンスとベンチマーク

<mailto:primergy.benchmark@ts.fujitsu.com>