

ホワイトペーパー

FUJITSU PRIMERGY サーバ

XEON 5600 (WESTMERE-EP) 搭載システムの メモリパフォーマンス

Xeon 5500 (Nehalem-EP) では、メインメモリへの接続に、FSB (Front Side Bus : フロントサイドバス) ではなく QPI (QuickPath Interconnect : QuickPath インターコネクト) を使用するというパラダイム転換がありました。Xeon 5600 (Westmere-EP) を搭載する PRIMERGY 2 ソケットモデルは、このパラダイム転換以降の第 2 世代です。この新しいアーキテクチャーで、可能な限り強力なシステムを構成しようとする場合、いくつかの新しいパラメーターについて考慮する必要があります。考慮する主な内容とは、800、1066、および 1333 MHz という異なるメモリ周波数がある点と、各プロセッサの 3 つのメモリチャンネルにできるだけ同じ容量になるようメモリモジュールを搭載するという点です。本書では、これらの要因がパフォーマンスに及ぼす効果を説明し、強力かつ低コストな構成の定義に役立つ情報を提供します。



概要

Xeon 5500 (Nehalem-EP) では、メインメモリへの接続に、FSB (Front Side Bus : フロントサイドバス) ではなく QPI (QuickPath Interconnect : QuickPath インターコネクト) を使用するというパラダイム転換がありました。Xeon 5600 (Westmere-EP) を搭載する PRIMERGY 2 ソケットモデルは、このパラダイム転換以降の第 2 世代です。強力なシステムの構成には、次の特徴があります。

- NUMA アーキテクチャー。
- 周波数は 1333、1066、または 800 MHz です。
- インターリーブは、3WAY、2WAY、1WAY のいずれか。これは、各プロセッサに 3 つずつあるメモリチャンネルへの DIMM の配分によって異なります。

NUMA アーキテクチャーでは、DIMM スロットは各プロセッサに直接割り当てられます (ローカルメモリ)。そのため、両方のプロセッサに DIMM を取り付ける必要があります、さらに DIMM の配分を対称型にするのが理想的です。メモリの容量が大きい PRIMERGY BX920 S2 では、パフォーマンスの低下が 2~3 %程度で抑えられる非対称型の構成が必要です。

BIOS のデフォルト値は、NUMA Optimization が Enabled に設定されているので、変更しないでください。

強力な Xeon 5600 モデルでの最大メモリ周波数は 1333 MHz で、パワーの低いモデルでは 1066 MHz です。周波数を 1066 または 800 MHz にダウングレードする理由は、メモリ容量が大きいこと (特に 16 GB や 32 GB の DIMM を使用する場合) と、チャンネルあたり 2 つの DIMM 構成で省エネルギーの 1.35 V 低電圧 (LV) 運用を行うためです。ダウングレードするとパフォーマンスが最大 5 %低下しますが (商用アプリケーションの平均値)、通常は問題ありません。

PRIMERGY コンフィギュレータの「パフォーマンスモード」メモリオプションにより、システム内すべてのメモリチャンネルを同一の構成にできるので、最適な 3WAY インターリーブになります。16、32、64 GB などの従来のメモリサイズは、パフォーマンスモードでは実装できないので (必要な DIMM の数が 6 の倍数でないため)、2WAY インターリーブになります。これにより、パフォーマンスが 1~5 %低下しますが、通常は問題ありません。

各プロセッサに DIMM を 1 枚だけ装着する構成 (1WAY インターリーブ) は使用しないでください。その理由は、最もパワーの低いプロセッサでなくても、パフォーマンスが約 20 %低下するためです。

冗長性を考慮したメモリパフォーマンスでは、DIMM スペアリングは、パフォーマンスが 1~5 %低下します。ミラーリング中は、約 10 %のパフォーマンス低下に対してフェールセーフの利点をとるかどうかを比較検討する必要があります。

さらに、基本的なルールも関連します。つまり、プロセッサが強力なほど、メモリパラメーターの影響は大きくなります。

ドキュメントの履歴

バージョン 1.0 (2010 年 5 月 7 日)

初版

バージョン 1.0 (2010 年 8 月 23 日)

PRIMERGY CX120 S1 を追加

バージョン 1.2 (2010 年 12 月 14 日)

PRIMERGY CX122 S1 を追加

バージョン 2.0 (2011 年 6 月 6 日)

新しい Westmere-EP CPU モデル (Westmere-EP Refresh) を導入
32 GB DIMM を導入

はじめに

PRIMERGY 2 ソケット現行モデル（ラックサーバ、タワーサーバ、ブレードサーバ、およびクラウドサーバ）に搭載される Intel Xeon 5600 (Westmere-EP) プロセッサの主要な革新的テクノロジーは、32 nm（ナノメートル）製造プロセスです。製造プロセスが 45 nm だった先行世代 Xeon 5500 (Nehalem-EP) と比較すると、このテクノロジーではプロセッサあたり最大 6 コアのサポートが可能になり、L3 キャッシュは 8 MB から 12 MB に増加しました。その結果、パフォーマンスは約 40 %向上しました。

Xeon 5500 および Xeon 5600 プロセッサ搭載モデルは、同じ Intel QPI (QuickPath Interconnect : QuickPath インターコネクト) ベースのマイクロアーキテクチャーが採用されています。このアーキテクチャーにより、プロセッサとシステムの他のコンポーネント、特にメインメモリとの接続性が飛躍的に向上し、以前のアーキテクチャーの 2 倍近いシステムパフォーマンスを実現しました。Intel Pentium Pro プロセッサ（1995 年）以来使用されてきた FSB (Front Side Bus : フロントサイドバス) テクノロジーは、チップセット内の FSB あたり必要なピン数を例にしても、複雑さにおいて限界に達していました。QPI では、システムアーキテクチャーにおける SMP (Symmetric Multiprocessing : 対称型マルチプロセッシング) から NUMA (Non-Uniform Memory Access : 非均等型メモリアクセス) へのパラダイム転換が実現されています。本書では、最も強力なシステムを実現するメモリ構成という観点から QPI アーキテクチャーのパフォーマンスの特徴を説明します。その上で、Xeon 5600 (Westmere-EP) を搭載した世代独自の特徴について触れていきます。Xeon 5600 (Westmere-EP) と Xeon 5500 (Nehalem-EP) には、さまざまな細かい相違点があります。

QPI は、プロセッサモデルに応じて、6.4、5.9、または 4.8 GT/s (ギガトランスファー/秒) の単方向のシリアルリンク経由で、プロセッサ同士およびプロセッサと I/O を管理するチップセット間を接続します。Xeon 5500 および Xeon 5600 シリーズのプロセッサには、メインメモリを接続するメモリコントローラーが内蔵されており、割り当てられたメモリモジュールを各プロセッサが直接制御します。プロセッサは、QPI リンク経由でメモリの内容を隣接プロセッサに提供し、同時に隣接プロセッサからの情報を要求します。

プロセッサとメモリが直接接続されているため、メモリパフォーマンスの向上が期待できますが、ローカル要求とリモート要求のパフォーマンスが異なるため、このアーキテクチャーは NUMA に分類するのが妥当です。オペレーティングシステムは、物理メモリの割り当て時と、プロセスのスケジューリング時に、NUMA を考慮します。メモリの総容量は、2 つのプロセッサにできるだけ等しく分散させる必要があります。

このルールを前提とすると、メモリシステム機能から生じる他の多様な項目を考慮する必要があります。これによりメモリは、1333、1066、または 800 MHz のクロック速度を提供するようになります。各構成の実効値は、プロセッサの種類、使用される DIMM の種類、および各プロセッサに 3 つあるメモリチャンネルに対するメモリ容量の配分によって決定されます。プロセッサごとだけでなく、チャンネルごとでも DIMM の数が対称であることが理想的です。その結果、DIMM の数は、6 の倍数が推奨されます（3 チャンネルのプロセッサが 2 基）。このガイドラインに従うと、8、16、32、64 および 128 GB のメモリを構成する場合の従来のマトリクスは適切ではありません。しかし、お客様がこれらのメモリサイズを指定した場合、パフォーマンスにどのような影響が生じるのでしょうか。

本書では、まず Xeon 5600 搭載 PRIMERGY サーバのメモリアーキテクチャーの概要を説明します。その後で、実際のアプローチについて説明します。ここでは、構成を決定するために必要な情報をまとめた、パフォーマンスを考慮したメモリ構成表を用意しています。なお、これらの表では、システムおよび CPU は指定されており、その上で特定のメモリ容量（または大まかなメモリ構成）を満たす最適な構成を検討することを想定しています。そのため、ほとんどの場合は、これらの表の内容を検討するだけで、適切な構成を定義できます。続いて、STREAM および SPECint_rate_base2006 ベンチマークテストの結果に基づいて、推奨する構成の背景を説明します。このセクションの説明は、要求されるメモリ容量がパフォーマンスを考慮したメモリ構成表に示されておらず、個別に構成を定義しなければならない場合にお読みください。

この複雑な問題に関しては、次の項目が適用されます。最初は多くの要因によってパフォーマンスが左右されるように見える場合でも、ベストプラクティスに関するさまざまなルールに従うことにより、強力なシステムを迅速に構成できます。コスト面に基づいてバランスのとれたソリューションを検討してみると、多くの場合でコストを無限にかけることは可能ですが、平均 5 %未満のわずかなパフォーマンスの向上しか図れません。限りなくコストをかけるべきか、それが常に必要かを検討する場合には、ある程度の注意深さが必要です。同様に、プロジェクトの背景に関する知識も必要になります。実稼動システムを購入する場合は、ベンチマークテストの扱いが異なる場合があるので注意してください。

メモリアーキテクチャー

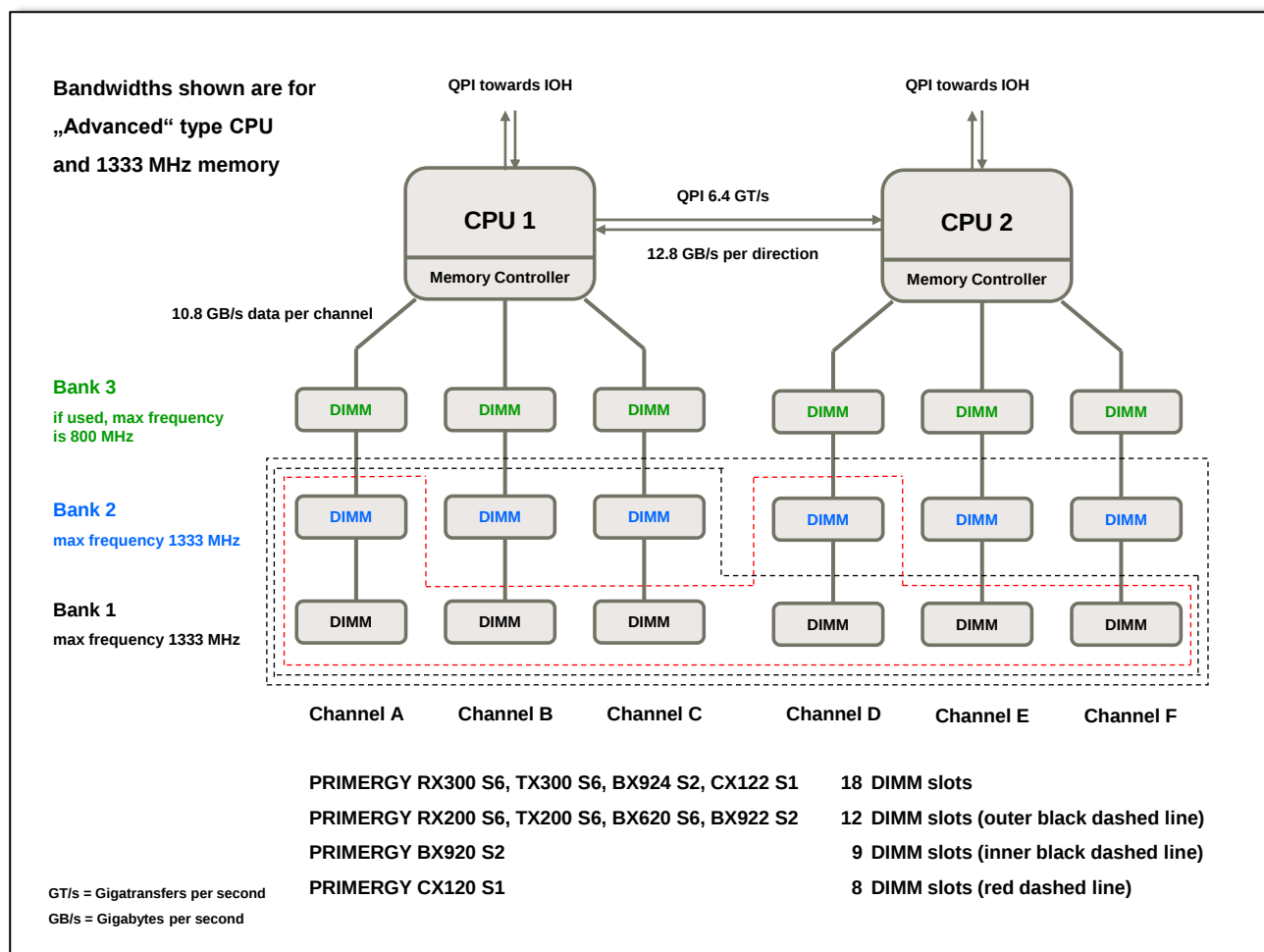
ここでは、3 部構成でメモリシステムの概要を説明します。まず、ブロック図で利用可能な DIMM スロットの配置を説明します。次に、4 つのメモリ構成モードについて説明します。これらのモードは、PRIMERGY コンフィギュレータでも参照できます。最後に、利用可能な DIMM のタイプを説明します。

DIMM スロット

次の図は、メモリシステムの構造を示します。PRIMERGY の各モデルは、DIMM スロット数とその配置について、次の 4 つのグループに分けられます。

- グループ 1 (18 スロット) : PRIMERGY RX300 S6、TX300 S6、BX924 S2、CX122 S1
- グループ 2 (12 スロット) : PRIMERGY RX200 S6、TX200 S6、BX620 S6、BX922 S2
- グループ 3 (9 スロット) : PRIMERGY BX920 S2
- グループ 4 (8 スロット) : PRIMERGY CX120 S1

なお、ここでの説明は、2011 年 5 月現在のシステムに基づきます。



1 つのプロセッサには、常に 3 つのメモリチャンネルが存在します。ただし、筐体内のスペースが決定的要因となる場合、チャンネルあたりの搭載可能な最大 DIMM 数に応じて、これら 4 つのモデルグループは変更されます。チャンネルあたりの DIMM 数が変わると、メモリ周波数に変化が生じ、さらにはメモリパフォーマンスに影響を与えます。チャンネルあたりのメモリ数は、DPC (DIMM per channel) と呼ばれます。以降、この用語を使用します。例えば、PRIMERGY RX300 S6 の 2DPC 構成では、チャンネルあたり 2 枚、合計 12 枚の DIMM が搭載されます。

DPC 値は、システム内のすべてのチャンネルで同じである必要はありません。メモリ構成を記述する場合は、次のような省略形を使用します。

2 - 2 - 2 / 1 - 1 - 1

上記の例では、1 つ目のプロセッサには各チャンネルに 2 枚ずつ、2 つ目のプロセッサには各チャンネルに 1 枚ずつのメモリモジュールを装着することを示します。

以降では、「メモリバンク」という用語も使用します。図では、複数のチャンネルに分配されている 3 つの DIMM のグループが、1 つのバンクを形成しています。図中の色 (黒、青、緑) は、サーバのメインボード上で、構成エラーを防ぐためにバンクに付けられたカラーマークに対応します。プロセッサあたりの利用可能なスロット経由で DIMM を分配する場合、バンク 1 から順に割り当てることにより、最適なインターリーブが得られます。インターリーブは、メモリのパフォーマンスに最も影響を与えます。これについては、後ほど説明します。

DIMM スロットを使用するためには、対応するプロセッサを搭載する必要があります。1 基のプロセッサのみで運用する場合、空のソケットに割り当てられた DIMM スロットは使用できません。

4 つのメモリ構成モード

メモリ構成を定義する際には、パフォーマンス以外にも考慮すべき事項があります。考慮すべき事項は、RAS (Reliability (信頼性)、Availability (可用性)、Serviceability (サービス性)) という略語で表されます。メモリシステムには、特に高度な RAS 要件を求めるお客様向けのオプションが用意されています。これは、以下に示す 4 つのメモリ構成モードのうち、最初の 2 つに該当します。これら 2 つのモードは、必要に応じて、BIOS で指定します。指定しない場合は、実際の DIMM 構成によって、パフォーマンスモードか独立チャンネルモードかが判定されます。適正な DIMM が適切に配置されていれば、自動的にパフォーマンスモードに設定されます。

- **スペアチャンネルモード** : 各バンクは、空の状態、または 3 枚の DIMM (同一タイプ、同一容量) で構成します。チャンネル A と B (または D と E) の DIMM のみが使用されます。チャンネル C (または F) には、DIMM に障害が生じた場合のスペアが装着されます。このモードは、BIOS で設定してください。
- **ミラーチャンネルモード** : 各バンクでチャンネル A と B (または D と E) のみが使用されます。これらのチャンネルは、同じタイプの DIMM で構成する必要があります。チャンネル C (または F) には、DIMM を搭載しません。ハードウェアは、オペレーティングシステムやアプリケーションに対し、透過的にメモリの内容をミラーリングします。実際には、構成したメモリ容量の半分を使用できます。DIMM が 1 枚故障しても、システムのダウンタイムは生じません。このモードは、BIOS で設定してください。
- **パフォーマンスモード** : 各バンクは、空の状態、または 3 枚の DIMM (同一タイプ、同一容量) で構成します。この構成では、3 つのメモリチャンネル経由で、最適なインターリーブが実現されます。
- **独立チャンネルモード** : 上記 3 つ以外の構成が、このカテゴリに分類されます。各スロットには、後述する DIMM のいずれも割り当て可能です。ただし、Unbuffered モジュールと Registered モジュールを混在させることはできません。

最初の 2 つのモードは、Xeon 5600 搭載のサーバモデルによっては未サポートの場合があります。

利用可能なメモリのタイプ

本書で取り上げた PRIMERGY モデルの構成を検討する場合は、次の表に示す DIMM を使用します。メモリモジュールには、ECC 保護の DDR3 メモリモジュールが使用されます。モジュールには Registered (RDIMM) と Unbuffered (UDIMM) があります。RDIMM と UDIMM を組み合わせた構成はできません。UDIMM は構造がシンプルなため、最大容量は RDIMM より小さいです。価格およびエネルギー消費に関する限り、シンプルな構造の方に利点があります。UDIMM には次の特長があります。

- UDIMM は、1DPC および 2DPC でのみ構成可能です。

DIMM には、1.5 V で動作する標準モジュールのほかに、1.35 V で動作する、エネルギー効率に優れた低電圧 (LV) モジュール (UDIMM 用および RDIMM 用) があります。1.5 V モジュールと 1.35 V モジュールを組み合わせることは可能ですが、推奨できません。1.5 V と 1.35 V のモジュールを混在させた場合は、すべてのモジュールが 1.5 V で動作します。モジュールが 1.35 V で動作するときは、次の一般的条件が適用されます。

- 1.35 V の使用は、1DPC および 2DPC 構成のみで可能です。
- 2DPC 構成の 1.35 V モジュールでは、最大メモリ周波数は 1066 MHz に制限されます。

低電圧版メモリ (LV-DIMM) の構成が、これらの一般的条件に反した場合、モジュールは 1.5 V で動作します。例えば、3DPC 構成を使用したり、BIOS オプションの [Memory Speed] をデフォルトの [Auto] ではなく [Performance] に設定して強制的に 1333 MHz を使用したりした場合です。

クアドランク (QR) の 16 GB および 32 GB の DIMM にもこの特長が当てはまります。この DIMM を使用すると、最大のメモリ構成を実現できます。

- QR DIMM は、1DPC および 2DPC でのみ構成可能です。
- QR DIMM は PC3-8500 構成であるため、最大メモリ周波数は 1066 MHz に制限されます。
- QR DIMM の 2DPC 構成のメモリ周波数は 800 MHz です。

メモリ周波数のこれらの一般的条件は、BIOS で無効にすることはできません。同様に、3DPC 構成のシステムを起動することはできません。

表の最終列は、各 DIMM の価格を相対比で示しています。各 DIMM の価格は、2011 年 5 月現在の PRIMERGY RX300 S6 の料金表を使用しています。ここでは、4 GB のサイズの Registered PC3-10600 DIMM を基準とし (1 として強調表示)、GB あたりの相対価格を示します。DDR3 メモリモジュールの導入以降、相対価格の状況は絶えず変化しています。RDIMM と比較し UDIMM のコストが低いことは、変わりません。一方、8 GB および 16 GB の RDIMM は、以前ほど割高ではありません。より高価だった LV 版でも、同じ現象がみられます。これらに代わり、4 Gbit テクノロジーに基づいた新しい 32 GB メモリが、より高価なメモリとして加わりました。

タイプ			制御	最大 MHz	ランク	容量	GB あたりの価格比
UDIMM	DDR3-1333 PC3-10600		Unbuffered	1333	2	2 GB	0.7
UDIMM	DDR3-1333 PC3-10600	LV	Unbuffered	1333	2	2 GB	0.9
RDIMM	DDR3-1333 PC3-10600		Registered	1333	1	2 GB	1.1
RDIMM	DDR3-1333 PC3-10600		Registered	1333	1 または 2	4 GB	1
RDIMM	DDR3-1333 PC3-10600	LV	Registered	1333	1 または 2	4 GB	1.0
RDIMM	DDR3-1333 PC3-10600		Registered	1333	2	8 GB	0.9
RDIMM	DDR3-1333 PC3-10600	LV	Registered	1333	2	8 GB	0.9
RDIMM	DDR3-1066 PC3-8500		Registered	1066	4	16 GB	1.1
RDIMM	DDR3-1066 PC3-8500		Registered	1066	4	32 GB	3.5

4GB の RDIMM は、物流と在庫の状況に応じて、1 ランクまたは 2 ランクのモジュールが提供されます。「ランク」という用語については、「[パフォーマンスの二次的影響](#)」を参照してください。

PRIMERGY モデルによっては、特に新しい 32 GB DIMM では、特定の DIMM タイプを利用できない場合があります。常に最新のコンフィギュレータを参照してください。また、販売地域によっても、利用できない DIMM タイプがあります。

パフォーマンスを考慮したメモリ構成

次の 2 つの表に、パフォーマンスを考慮したメモリ構成例を示します。最初の表に示した構成は、システム内のすべてのメモリチャネルにメモリが均等に配分されているため、「理想的」といえます。これらの構成は、パフォーマンスモードに対応します。

2 番目の表は、8、16、32 GB など、従来のシステムアーキテクチャの「古典的」構成について示しています。これらの構成は、理想的な構成と詳細に比較すると、容量の違いがテスト結果に影響を与えない限り、パフォーマンスが 1~5 %低下します。しかし、このマイナス面は、ほとんどの用途で問題になりません（以下で説明するとおり、この差異の原因は、従来のサイズで 2WAY インターリーブが使用されていることにあります。一方、理想的な構成では、3WAY インターリーブです）。

表 1 : 理想的なメモリサイズ

容量	タイプ	モジュールサイズ (GB)	構成	MHz (最大) 1.5 V	MHz (最大) 1.35 V	備考	RX/TX300 S6 BX924 S2	RX/TX200 S6 BX620 / BX922	BX920 S2	CX120 S1	CX122 S1
12 GB	UDIMM	2	1-1-1 / 1-1-1	1333	1333	RDIMM の価格優位性および LV オプションとの比較	✓	✓	✓		✓
24 GB	UDIMM	2	2-2-2 / 2-2-2	1333	1066	RDIMM の価格優位性および LV オプションとの比較	✓	✓			
	RDIMM	4	1-1-1 / 1-1-1	1333	1333	BX920 S2 に搭載可能	✓	✓	✓	✓	✓
36 GB	RDIMM	4 および 2	2-2-2 / 2-2-2	1333	該当せず	1 番目のバンク 4 GB DIMM 2 番目のバンク 2 GB DIMM	✓	✓			
48 GB	RDIMM	8	1-1-1 / 1-1-1	1333	1333		✓	✓	✓	✓	✓
60 GB	RDIMM	8 および 2	2-2-2 / 2-2-2	1333	該当せず	1 番目のバンク 8 GB DIMM 2 番目のバンク 2 GB DIMM	✓	✓			
72 GB	RDIMM	8 および 4	2-2-2 / 2-2-2	1333	1066	1333 MHz が可能	✓	✓			
	RDIMM	4	3-3-3 / 3-3-3	800	該当せず	CX122 S1 に搭載可能	✓				✓
84 GB	RDIMM	8、4 および 2	3-3-3 / 3-3-3	800	該当せず	1 番目のバンク 8 GB DIMM 2 番目のバンク 4 GB DIMM 3 番目のバンク 2 GB DIMM	✓				
96 GB	RDIMM	8	2-2-2 / 2-2-2	1333	1066	1333 MHz が可能	✓	✓			✓
	RDIMM	16	1-1-1 / 1-1-1	1066	該当せず	BX920 S2 に搭載可能	✓	✓	✓		
108 GB	RDIMM	16 および 2	2-2-2 / 2-2-2	800	該当せず	1 番目のバンク 16 GB DIMM 2 番目のバンク 2 GB DIMM	✓	✓			

容量	タイプ	モジュールサイズ (GB)	構成	MHz (最大) 1.5 V	MHz (最大) 1.35 V	備考	RX/TX300 S6 BX924 S2	RX/TX200 S6 BX620 / BX922	BX920 S2	CX120 S1	CX122 S1
120 GB	RDIMM	8 および 4	3-3-3 / 3-3-3	800	該当せず	1 番目と 2 番目のバンク 8 GB DIMM 3 番目のバンク 4 GB DIMM	✓				
	RDIMM	16 および 4	2-2-2 / 2-2-2	800	該当せず	1 番目のバンク 16 GB DIMM 2 番目のバンク 4 GB DIMM	✓	✓			
144 GB	RDIMM	8	3-3-3 / 3-3-3	800	該当せず	CX122 S1 に搭載可能	✓				✓
	RDIMM	16 および 8	2-2-2 / 2-2-2	800	該当せず	1 番目のバンク 16 GB DIMM 2 番目のバンク 8 GB DIMM	✓	✓			
192 GB	RDIMM	16	2-2-2 / 2-2-2	800	該当せず		✓	✓			
	RDIMM	32	1-1-1 / 1-1-1	1066	該当せず	1066 MHz が可能 BX920 S2 に搭載可能	✓*	✓*	✓*		
204 GB	RDIMM	32 および 2	2-2-2 / 2-2-2	800	該当せず	1 番目のバンク 32 GB DIMM 2 番目のバンク 2 GB DIMM	✓*	✓*			
216 GB	RDIMM	32 および 4	2-2-2 / 2-2-2	800	該当せず	1 番目のバンク 32 GB DIMM 2 番目のバンク 4 GB DIMM	✓*	✓*			
240 GB	RDIMM	32 および 8	2-2-2 / 2-2-2	800	該当せず	1 番目のバンク 32 GB DIMM 2 番目のバンク 8 GB DIMM	✓*	✓*			
288 GB	RDIMM	32 および 16	2-2-2 / 2-2-2	800	該当せず	1 番目のバンク 32 GB DIMM 2 番目のバンク 16 GB DIMM	✓*	✓*			
384 GB	RDIMM	32	2-2-2 / 2-2-2	800	該当せず		✓*	✓*			

✓*: 32 GB DIMM の導入は、PRIMERGY RX300 S6、TX300 S6、RX200 S6、BX920 S2 のシステムのみで段階的に行われています。最新のコンフィギュレータを参照してください。

表 2 : 従来のメモリサイズ

容量	タイプ	モジュール サイズ (GB)	構成	MHz (最大) 1.5 V	MHz (最大) 1.35 V	備考	RX/TX300 S6 BX924 S2	RX/TX200 S6 BX620 / BX922	BX920 S2	CX120 S1
8 GB	UDIMM	2	1-1-0 / 1-1-0	1333	1333	RDIMM の価格優位性および LV オプションとの比較	✓	✓	✓	✓
16 GB	UDIMM	2	2-1-1 / 2-1-1	1333	1066	RDIMM の価格優位性および LV オプションとの比較	✓	✓		✓
	RDIMM	4	1-1-0 / 1-1-0	1333	1333	BX920 S2 に搭載可能	✓	✓	✓	
32 GB	RDIMM	8	1-1-0 / 1-1-0	1333	1333		✓	✓	✓	
64 GB	RDIMM	8	2-1-1 / 2-1-1	1333	1066	1333 MHz が可能 CX120 S1 に搭載可能	✓	✓		✓
	RDIMM	16	1-1-0 / 1-1-0	1066	該当せず	BX920 S2 に搭載可能	✓	✓	✓	
128 GB	RDIMM	16	2-1-1 / 2-1-1	800	該当せず	32 GB より価格的に優位	✓	✓		
	RDIMM	32	1-1-0 / 1-1-0	1066	該当せず	BX920 S2 に搭載可能	✓*	✓*	✓*	
256 GB	RDIMM	32	2-1-1 / 2-1-1	800	該当せず		✓*	✓*		

上記 2 つの表は、メモリが両方のソケットに対称的に配分されており、NUMA について最適です。非対称メモリ構成については、後述する表に示します。

これらの表には、各構成において可能なメモリ周波数の最大値を示しています。1.5 V と 1.35 V の運用では、後者が低電圧メモリ (LV-DIMM) を使用する場合のみ、違いが生じます。低電圧メモリ (LV-DIMM) を使用しない場合、この列は「該当せず」となります。表に記載された特長や DPC 値、および DIMM タイプの他に、プロセッサのタイプも有効周波数の決定要因となります。Xeon 5600 世代の強力なプロセッサは最大 1333 MHz、それよりパワーの低いプロセッサは最大 1066 MHz のメモリ周波数をサポートします。利用可能なすべての Xeon モデルの一覧と分類は、後述のセクションを参照してください。メモリの有効な周波数は、表とプロセッサのクラスに基づいた最小値です。

表の右 4 列または 5 列は、それぞれの構成が可能な PRIMERGY モデルを示します。

ここで扱っていない構成に対するメモリ構成については、「[メモリパフォーマンスに対する影響](#)」を参照してください。

必要なメモリ容量は仮定条件です。例えば、I/O 速度など、アプリケーションのパフォーマンスに対する暗黙の影響は、ここでは無視しています。

非対称型メモリ構成

すべてのシステムが、すべての構成バージョンで対称型メモリ構成を形成できるわけではありません。フォームファクターによっては、形成できない構成もあります。「[メモリアーキテクチャー](#)」の図は、PRIMERGY BX920 S2 の DIMM スロットの非対称型配置を示しています。最初のソケットに 2 つのメモリバンクがあり、2 番目のソケットに 1 つのメモリバンクがあります。NUMA の推奨は、両方のソケット経由でメモリを対称的に配分することです。推奨構成を考慮すると、この非対称配置には別の側面があります。

PRIMERGY BX920 S2 は、スロットは非対称ですが、192 GB の容量までは合計メモリ容量を 2 等分することができます。このような構成は、NUMA に最適です。これらの構成は、前述の表「理想的なメモリサイズ」と「従来のメモリサイズ」で示しています。

次の表の「構成」欄では、左側のメモリ数は右側より大きくなっています。超過分は、総容量の 1/4 ~ 1/3 です。最大超過分の半分、つまり 1/8 ~ 1/6 には、QPI リンク経由の「リモート」アクセスが行われます（統計的に確認）。このような軽度の非対称型の場合、対称型と比較して 2 ~ 3 % のパフォーマンスの低下分を計算に入れる必要があります。大容量の共有メモリセグメントを持つデータベースなど、いずれにしてもリモートアクセスが避けられない作業負荷の場合、パフォーマンス上のマイナス影響はありません。このことは、PRIMERGY BX920 S1 で Windows Server 2008 と SQL Server 2008 を使用して行われた OLTP2 評価 [関連資料 4] で確認されました。

PRIMERGY BX920 S2 表 3 : 非対称型構成						
容量	タイプ	モジュール サイズ (GB)	構成	MHz (最大) 1.5 V	MHz (最大) 1.35 V	備考
36 GB	RDIMM	4	2-2-2 / 1-1-1	1333	1066	
72 GB	RDIMM	8	2-2-2 / 1-1-1	1333	1066	
128 GB	RDIMM	16	2-2-1 / 1-1-1	800	該当せず	32 GB DIMM より価格的に優位
144 GB	RDIMM	16	2-2-2 / 1-1-1	800	該当せず	
256 GB	RDIMM	32	2-2-1 / 1-1-1	800	該当せず	
288 GB	RDIMM	32	2-2-2 / 1-1-1	800	該当せず	

メモリパフォーマンスに対する影響

ここでは、RAM のパフォーマンスに影響を与える要因について説明します。まず、本書の基準となったテストで、メモリパフォーマンスをどのように測定し、それらのデータをどのように判断したかについて説明します。

測定ツール

測定は、STREAM および SPECint_rate_base2006 ベンチマークを使用して行われました。

STREAM ベンチマーク

STREAM ベンチマーク（開発者：John McCalpin 氏）[関連資料 3] は、メモリのスループットを測定するツールです。このベンチマークは、double 型データの大規模な配列でコピーおよび算術演算を実行して、Copy、Scale、Add、Triad の 4 種類のアクセスの結果を提供します。Copy 以外のアクセスタイプには、算術演算が含まれています。結果は、常に GB/s 単位のスループットで示されます。一般に、Triad の値が最もよく引用されます。以降、STREAM のベンチマークの測定値は、Triad アクセスでの値であり、単位は GB/s です。

STREAM は、サーバのメモリ帯域幅を測定するための業界標準で、シンプルな方法を使用してメモリシステムに大規模な負荷を与えることができます。特にこのベンチマークは、複雑な構成でのメモリパフォーマンスに対する影響を調査する場合に適しています。STREAM は、構成によるメモリへの影響とそれによって生じるパフォーマンスへの影響（低下または向上）を示します。後述する STREAM ベンチマークに関する値は、パフォーマンスへの影響度を示しています。

アプリケーションのパフォーマンスに対するメモリの影響は、各アクセスの遅延時間とアプリケーションが必要とする帯域幅に区別されます。メモリ帯域幅が増加すると遅延時間は増加するため、両者は関連しています。並列メモリアクセスによって遅延時間が相殺される度合いは、アプリケーションや、コンパイラによって作成されたマシンコードの質にも依存します。このため、すべてのアプリケーションシナリオでの一般的な予測を立てることは非常に困難です。

SPECint_rate_base2006

SPECint_rate_base2006 ベンチマークは、商用アプリケーションパフォーマンスのモデルとして追加されました。これは、Standard Performance Evaluation Corporation (SPEC) の SPECcpu2006 [関連資料 5] の一部です。SPECcpu2006 は、システムのプロセッサ、メモリおよびコンパイラを評価するための業界標準です。大量の測定結果が公開され、販売プロジェクトおよび技術調査に使用されているため、サーバ分野で最も重要なベンチマークとなっています。

SPECcpu2006 は、大量の整数演算および浮動小数点演算を使用する独立した 2 つのテストセットで構成されています。整数演算部分は商用アプリケーションに相当し、12 種類のベンチマークから構成されます。浮動小数点演算部分は科学アプリケーションに相当し、17 種類のベンチマークで構成されます。いずれの場合も、ベンチマークの実行結果は、個々の結果の幾何平均です。

さらに、それぞれのテストセットには、単体実行時の処理性能を評価する速度測定と、並行処理の性能を評価するスループット測定があります。多数のプロセッサコアとハードウェアスレッドを持つサーバにとっては、後者が重要です。

また、測定の種類により、コンパイラに許可される最適化が異なります。ピーク値の測定では、各ベンチマークを個別に最適化できますが、ベース値の測定では、コンパイラフラグがすべてのベンチマークで同一である必要があり、特定の最適化は許可されません。

以上が SPECint_rate_base2006 の概要です。PRIMERGY サーバでは商用アプリケーションの使用が主流であるため、整数演算を使用するテストセットである SPECint_rate_base2006 でスループットを測定しました。

本来のルールに準拠した測定では 3 回の実行が必要であり、各ベンチマークに対して平均の結果が評価されます。しかし、ここで説明している技術調査では、このルールに準拠していません。効率化のために、測定は 1 回にしています。

パフォーマンスに影響を与える主な要因

ここでは、メモリパフォーマンスに対する 2 つの主な影響要因である、周波数とインターリーブについて説明します。どちらもそれぞれ 800、1066、1333 MHz の周波数と 1WAY、2WAY、3WAY のインターリーブの 3 つのオプションがあります。最適値である 1333 MHz および 3WAY に反対する理由については、前のセクションで説明しました。非常に大きいメモリ構成またはエネルギー節約はメモリ周波数の低下を招き、お客様は 16、32、64 GB などの従来のメモリサイズを使用するのに 2WAY インターリーブを要求されるためです。

メモリ構成を計画する際には、まず、これらのパラメーターの計画を立てる必要があります。

このセクションの最後では、冗長性（スペアリングおよびミラーリング）を考慮した際のメモリパフォーマンスについて説明します。

メモリの有効な周波数

システムの電源を入れたときに BIOS によって判断される有効な周波数は、次の 3 つの要素によって決まります。

- プロセッサのタイプ。プロセッサは、次の表に従って分類されます。注目すべき特徴の列は、灰色でマークされています。強力なモデルは最大 1333 MHz、それよりパワーの低いモデルは最大 1066 MHz をサポートします。
- DIMM のタイプ。UDIMM および RDIMM は通常、最大 1333 MHz をサポートしています。最大 1066 MHz のみをサポートする、クアッドランク（QR）16 GB および 32 GB RDIMM は例外です。
- DPC 値（チャンネルあたりの DIMM）。1DPC 構成と 3DPC 構成の場合は単純で、1DPC は 1333 MHz をサポートしますが、3DPC では常に 800 MHz に制限されています。ここでもう一度確認しておきますが、UDIMM、1.35 V 低電圧（LV）運用、またはクアッドランク（QR）モジュールを使って 3DPC 構成を実現することはできません。2DPC 構成の場合、1066 MHz で動作する LV モジュールと 800 MHz で動作するクアッドランク（QR）モジュール、およびこれらのモジュールを組み合わせた構成を除き、通常は周波数 1333 MHz がサポートされます。6 個のチャンネルを同じ構成にしていない場合は、最も大きい DPC 値によって判断されます。

クラス	Xeon タイプ	コア数	GHz	L3 キャッ シュ (MB)	QPI (GT/s)	最大 メモリ (MHz)	TDP (ワット)
Advanced	X5690	6	3.46	12	6.4	1333	130
	X5687	4	3.60	12	6.4	1333	130
	X5672	4	3.20	12	6.4	1333	95
	X5675	6	3.06	12	6.4	1333	95
	X5660	6	2.80	12	6.4	1333	95
	X5650	6	2.66	12	6.4	1333	95
Standard	E5649	6	2.53	12	5.9	1333	80
	E5645	6	2.40	12	5.9	1333	80
	X5647	4	2.93	12	5.9	1066	130
	E5620	4	2.40	12	5.9	1066	80
Low Power	L5640	6	2.26	12	5.9	1333	60
	L5630	4	2.13	12	5.9	1066	40
	L5609	4	1.86	12	4.8	1066	40
Basic	E5607	4	2.26	4	4.8	1066	80
	E5606	4	2.13	4	4.8	1066	80
	E5603	4	1.60	4	4.8	1066	80

できるだけ高い周波数を使用した方が効果的です。ただし、3 つの要素の中で最も低い値が使用されます。つまり、構成の周波数は 3 つの要素の中の最も低い値によって決定されます。また、周波数は、プロセッサ単位ではなく、システムの標準値として定義されます。

次に、このメカニズムについて、PRIMERGY RX300 S6 にタイプ Xeon E5620 プロセッサと 4 GB の RDIMM × 18 枚を搭載する場合を例に説明します。このプロセッサは 1066 MHz をサポートし、さらにこの DIMM タイプは 1333 MHz をサポートしますが、3DPC 構成なので、周波数は 800 MHz に制限されます。したがって、有効周波数は 800 MHz になります。1 番目のバンクに 8 GB モジュール、2 番目のバンクに 4 GB モジュールを使用し、3 番目のバンクを空のままにしても、同じ 72 GB のメモリ構成を実現できます。この 2DPC 構成では、プロセッサは 1066 MHz をサポートします。

次の表は、メモリ周波数の違いがアプリケーションパフォーマンスに及ぼす影響を示しています。測定は、STREAM（赤：各セルの上の値）および SPECint_rate_base2006（緑：各セルの下の値）のベンチマークを使用して行われました。緑の値は、商用アプリケーションで代表的な値です。表の最後から 2 行目は、Xeon E5620 プロセッサで説明した例の決定的なデータです。メモリ周波数によるパフォーマンスの違いは 3 % です。

STREAM は、メモリ帯域幅と同意語です。ここに示された差異は、例外的なケースのアプリケーションによってのみ達成される上限値を表しています。SPECint_rate_base2006 の 12 のコンポーネントには、例えば、パフォーマンスの差異の上限値（この例では 20 %）に実際に達した場合など、STREAM のように動作するテストケース（libquantum）などが含まれます。

この表では、Xeon 5600 モデルを、最大メモリ周波数 1333 MHz と 1066 MHz に対応する 2 クラスだけではなく、4 つのクラスに分類しています。このセクションのベースとなっている分析で、この分類が推奨されたためです。プロセッサコア周波数に応じた区別は必要ありませんが、少なくとも QPI 周波数は、プロセッサコアとメモリシステム間の相互作用に影響します。

アプリケーションのパフォーマンスに対するメモリ周波数の影響が小さくなっていることが、明確にわかります。プロセッサモデルが強力なほど、影響も大きくなります。この見解については、インターリーブに関する次のセクションでも説明します。つまり、ここではこれが重要なポイントです。

この表は、相対的なパフォーマンスを示しています。STREAM および SPECint_rate_base2006 ベンチマークの絶対値（表の 1.00 基準点に対応）については、それぞれの PRIMERGY システムのパフォーマンスレポートを参照してください。パフォーマンスレポートの測定では、最適なメモリ構成が使用されています。

Relative Performance for Different Memory Speeds

Memory Bandwidth (STREAM)

Commercial Application Performance (SPECint_rate_base2006)

QPI	Max Mem MHz	CPU Models	Effective Memory MHz		
			1333	1066	800
			Max Performance	Energy Efficiency	Max Capacity
6.4	1333	X5690 X5687 X5675 X5660 X5650	1.00	0.84	0.62
			1.00	0.98	0.91
5.9	1333	E5649 E5645 L5640	1.00	0.94	0.72
			1.00	1.00	0.95
5.9	1066	X5647 E5620 L5630	N/A	1.00	0.80
				1.00	0.97
4.8	1066	E5607 E5606 E5603 L5609	N/A	1.00	0.95
				1.00	0.99

インターリーブ

インターリーブは、最初のブロックは最初のチャンネルに、2 番目のブロックは 2 番目のチャンネルにという具合に、プロセッサ単位で 3 つのメモリチャンネルを交互に利用するように物理アドレス領域を設定する手法です。メモリアクセスは、局所性原理より主に隣接するメモリ領域に行われ、結果としてすべてのチャンネルに分散されます。これは、並列化による性能向上の効果があります。さらに、遅延も目立たなくなります。これは、アクティブ（開いている）メモリページを変更する前に DRAM メモリの物理特性に従って確認する必要があります。

次の図に、前述のメモリ周波数よりも優れた効果をもたらすインターリーブの影響を示します。理想的な状況は 3WAY インターリーブであり、3 つのすべてのチャンネルを同一に構成すると実現されます。メモリ構成オプションのパフォーマンスモードは、このシナリオに基づいています。「[パフォーマンスを考慮したメモリ構成](#)」の 1 つ目の表（「理想的なメモリサイズ」）に一覧表示された構成は、すべて 3WAY インターリーブです。

「推奨」はされていますが、16、32、64 GB などの従来のメモリ構成を要求された場合など、理想的な状況の周波数を達成できないこともあります。そのため、「[パフォーマンスを考慮したメモリ構成](#)」の 2 つ目の表（「従来のメモリサイズ」）で示された構成が発生します。これらはすべて、2WAY インターリーブです。表を詳細に見てみると、2WAY インターリーブには 2 つのスキーマがあることがわかります。1 つ目のスキーマは、プロセッサの 3 番目のメモリチャンネルが使用されず、残りの 2 つが同一容量で構成されていることを示しています。2 つ目のスキーマでは 3 つのチャンネルすべてが使用されていますが、チャンネルごとに容量が異なるため、3WAY インターリーブは使用できません。2WAY インターリーブが発生する仕組みの詳細については、このセクションの最後で説明します。

この表の構造は、前のセクションと同じです。STREAM（赤：各セルの上の値）および SPECint_rate_base2006（緑：各セルの下の値）の負荷プロファイルを考慮し、緑の値は商用アプリケーションの平均値で、赤の値は例外的な状況で達成される極値を示しています。2WAY インターリーブでは、プロセッサモデルに応じてパフォーマンスが平均 1~5 %低下しますが、これは通常問題ではありません。

Relative Performance for Different Interleaving Levels

Memory Bandwidth (STREAM)

Commercial Application Performance (SPECint_rate_base2006)

QPI	Max Mem MHz	CPU Models	Effective Interleaving		
			3-way	2-way	1-way
			Max Performance	Classical Memory Capacities	Discouraged
6.4	1333	X5690 X5687 X5675 X5660 X5650	1.00 1.00	0.70 0.95	0.39 0.76
5.9	1333	E5649 E5645 L5640	1.00 1.00	0.77 0.97	0.43 0.82
5.9	1066	X5647 E5620 L5630	1.00 1.00	0.71 0.97	0.39 0.84
4.8	1066	E5607 E5606 E5603 L5609	1.00 1.00	0.83 0.99	0.45 0.88

1WAY インターリーブ構成は使用しないでください。これは分類上 1WAY と呼ばれているだけで、実際には非インターリーブ構成です。パフォーマンスの低下が想定されます。これは、プロセッサのパフォーマンス能力に密接に関係しません。この判断から、必要に応じて、最もパワーが低く最もコスト効率が高いプロセッサを除外することもできます。例えば、お客様の要求で、プロセッサごとに DIMM が 1 つしかない最低限度までメモリ構成を下げる場合などです。

プロセッサのパフォーマンスに応じた分類も、先に説明した重要ポイントと同様で、プロセッサモデルが強力なほど影響が大きくなります。

インターリーブは、周波数と同様、システムの電源を入れたときに BIOS によって定義されます。チャンネルあたりの GB 数が同じ場合、3 つの構成済みのチャンネルの 3WAY インターリーブが可能です。また、1 つのチャンネルを使用せずに、2 つのチャンネルで 2WAY インターリーブが可能です。DPC 値が異なっても、異なるサイズの DIMM を使用することで、インターリーブに最適な状況を作ることができます。チャンネルあたりの総 GB は決定的な要素です。

チャンネルあたりの総 GB が異なる場合、物理メモリは異なるインターリーブの領域に分割されます。これは、1WAY インターリーブの領域を回避することを目的としています。このため BIOS は、次に示すように、4 GB DIMM (例えば、32 GB の総容量にするのに必要) を 2 つの 2WAY に分割します。

2 - 1 - 1 / 2 - 1 - 1

次のように 2 つの 2WAY に分割します。

1 - 1 - 0 / 1 - 1 - 0 (メモリ容量の 50 %)

2WAY インターリーブ

1 - 0 - 1 / 1 - 0 - 1 (50 %)

2WAY インターリーブ

次のように、不均一にならないようにします。

1 - 1 - 1 / 1 - 1 - 1 (75 %)

3WAY インターリーブ

1 - 0 - 0 / 1 - 0 - 0 (25 %)

1WAY インターリーブ

冗長性を考慮した際のメモリパフォーマンス

インターリーブに関するセクションに続いて、冗長性を考慮した際のメモリパフォーマンスに関して説明します。DIMM スペアリングに必要なテストは、先ほど説明した 2WAY 構成の 1 つ目のスキーマに分類されるので、プロセッサの 3 つ目のメモリチャネルは使用されません。これらの構成で提供されるスペアリングモジュールの領域は、パフォーマンスに影響を与えません。したがって、次の表は、ミラーリングに関する一番右の列を除き、前の表と同じです。

しかし、ミラーリングを 1WAY インターリーブと同等に見なすことはできません。ミラーリングでは、各プロセッサにある最初の 2 つのメモリチャネルは同一に構成され、3 つ目のチャネルは空です。オペレーティングシステムは、最初のチャネル、つまり実際の構成の半分だけに対応したアドレス領域を認識します。ハードウェアは、書き込みプロセスごとに、1 つ目のチャネルが 2 つ目のチャネルに自動的にミラーリングされるようにします。ただし、読み取り中はミラーリングも使用できるため、ミラーリングの際のパフォーマンスは、1WAY インターリーブと 2WAY インターリーブの中間にあります。

パフォーマンスに対するスペアリングの影響は、2WAY インターリーブのように、プロセッサモデルに応じてパフォーマンスが平均 1~5 %低下しますが、これは通常問題ではありません。ミラーリング中は、約 10 %のパフォーマンス低下に対してフェールセーフの利点をとるかどうかを比較検討する必要があります。

Relative Performance for Redundant Configurations

Memory Bandwidth (STREAM)

Commercial Application Performance (SPECint_rate_base2006)

QPI	Max Mem MHz	CPU Models	Redundancy		
			Disabled ¹	Sparing	Mirroring
6.4	1333	X5690 X5687 X5675 X5660 X5650	1.00 1.00	0.70 0.95	0.57 0.87
5.9	1333	E5649 E5645 L5640	1.00 1.00	0.77 0.97	0.60 0.91
5.9	1066	X5647 E5620 L5630	1.00 1.00	0.71 0.97	0.57 0.92
4.8	1066	E5607 E5606 E5603 L5609	1.00 1.00	0.83 0.99	0.59 0.95

¹ Redundancy disabled and all three memory channels per CPU populated

パフォーマンスの二次的影響

ここまでで説明したトピックでは、測定を注意深く行った場合に、アプリケーションパフォーマンスでこれらの影響が認識できるようになると想定しています。以降のトピックでは、測定ツールを使用してパフォーマンスへの影響を実証していきます。ただし、実際のアプリケーションパフォーマンスに影響を与えるとは限りません。

UDIMM と RDIMM

Registered DIMM (RDIMM) モジュールとは別に、Unbuffered DIMM (UDIMM) モジュールも、次の表に従って利用できます。よりシンプルな UDIMM 構成は、安価で、エネルギー使用量も若干少なく済みます。これらの理由により、必要なメモリ容量に対応できる場合は、推奨できる構成と言えます。

タイプ			制御	最大 MHz	ランク	容量	GB あたりの価格比
UDIMM	DDR3-1333 PC3-10600		Unbuffered	1333	2	2 GB	0.7
UDIMM	DDR3-1333 PC3-10600	LV	Unbuffered	1333	2	2 GB	0.9
RDIMM	DDR3-1333 PC3-10600		Registered	1333	1	2 GB	1.1
RDIMM	DDR3-1333 PC3-10600		Registered	1333	1 または 2	4 GB	1
RDIMM	DDR3-1333 PC3-10600	LV	Registered	1333	1 または 2	4 GB	1.0
RDIMM	DDR3-1333 PC3-10600		Registered	1333	2	8 GB	0.9
RDIMM	DDR3-1333 PC3-10600	LV	Registered	1333	2	8 GB	0.9
RDIMM	DDR3-1066 PC3-8500		Registered	1066	4	16 GB	1.1
RDIMM	DDR3-1066 PC3-8500		Registered	1066	4	32 GB	3.5

RDIMM と UDIMM を組み合わせることはできません。

RDIMM では、メモリコントローラーの制御コマンドは、DIMM 上の独自のコンポーネントにあるレジスター内でバッファーされます（これが名前の由来です）。これにより、メモリチャネルが解放され、UDIMM では不可能な 3DPC が可能になります。その逆も同様で、UDIMM での 2DPC 構成では（1DPC と比較して）負荷が大きくなり、1N ではなく、2N 周波数での DIMM アドレッシングが必要になります。制御コマンドは、メモリチャネルのすべての 2 次クロックでのみ可能だからです。この結果、UDIMM での 2DPC 構成の最大メモリ帯域幅は、RDIMM よりも 5 %程度低下します。

商用アプリケーションのパフォーマンスの場合、この影響は無視できます。

ランク数

最後の表に、1 ランク、2 ランク、または 4 ランクが可能なメモリモジュールを示します。これは、64 ビットの帯域幅のメモリ領域を同時に読み取り、または書き込む DRAM チップのグループが 1 つだけの DIMM が存在することを示します。各チップは、4 ビットまたは 8 ビットを受け持ちます。または、そのようなグループが 2 つまたは 4 つあります。ただし、DIMM アドレスとデータラインは、両方のグループで共通です。つまり、1 つのグループのみが一度にアクティブになることができます。表に示されるように、2 ランクおよび 4 ランク DIMM のメリットとして、まず容量の大きさが挙げられます。

2 ランクおよび 4 ランクモジュールの 2 番目のメリットは、前述した物理的な理由です。メモリセルは、2 次元に配置されています。ラインが開かれ、このラインで列項目が読み取られます。ライン（一般的にはページと呼ばれる）が開いている間は、大幅に少ない待機時間で他の列値を読み取ることができます。この待機時間の違いが、メモリコントローラーの最適化を促します。最適化では、「開いている」メモリページに関する保留中の命令が再割り当てされます。2 ランクおよび 4 ランクのモジュールでは、開いているページにアクセスする確率が高まります。

これは、次の表に従って STREAM でメモリ帯域幅を測定しているときに確認できます。

CPU	RAM				帯域幅 (GB/s)
	タイプ	容量	ランク数	構成	
X5690	RDIMM 1333 MHz	8 GB	2	1 - 1 - 1 / 1 - 1 - 1	41.6
X5690	RDIMM 1333 MHz	2 GB	1	1 - 1 - 1 / 1 - 1 - 1	35.5

DPC 値が高い構成では、チャンネルあたりのランク数が奇数の場合に、同様の効果が得られます。2 ランクおよび 4 ランクのモジュールを使用している場合、このような状況は起こりません。2 GB モジュールを使用した構成で、チャンネルあたりのランク数が奇数の場合に実際のパフォーマンスが 1 ~ 2 %低下すること、2 ランクの UDIMM モジュールが推奨される理由の 1 つです。

リモートメモリへのアクセス

前述の STREAM および SPECint_rate_base2006 ベンチマークを使ったテストでは、ローカルメモリのみが対象になっていました (CPU が自身のメモリチャンネルの DIMM モジュールにアクセスする)。隣接する CPU のモジュールには、QPI リンク経由ではアクセスできません。実際のアプリケーションにおいて、オペレーティングシステムやシステムソフトウェアの NUMA サポートによってアクセスできるメモリの大半がローカルメモリである限り、この状況は代表的なものであると言えます。

次の表では、STREAM や、商用アプリケーションを代表するさまざまな標準ベンチマークとは逆のケースにおける効果を示しています。ここでは、明示的にプロセスをバインドする方法で、強制的にリモートメモリを使用しました。この表では、測定結果の低下を比率 (%) で示しています。

ベンチマーク	リモートメモリの強制使用による影響
STREAM Triad	-49 %
SPECint_rate_base2006	-13 %
SPECint_rate2006	-14 %
SPECjbb2005	-20 %

STREAM では、プロセッサ間を接続する QPI リンクの帯域幅が、結果を左右するボトルネックになります。他のベンチマークに見られる低下の原因は、主に個々のアクセスの待機時間が約 50 %増加したことによります。これらの結果は、リモートメモリを使用すると、商用アプリケーションで 10 ~ 20 %パフォーマンスが低下することを意味しています。

実用的でないように見えるこれらの測定結果も、BIOS で NUMA サポートを無効にした場合の影響を評価する際には有用になります。この場合、物理アドレス領域は両方のプロセッサのメモリモジュールを経由して詳細メッシュインターリーブで設定します。これにより、アプリケーションによるアクセスの 50 %はローカルメモリに、50 %はリモートメモリに対して行われます。商用アプリケーションの場合、予測される低下範囲は 5 ~ 10 %です。これは、NUMA サポートを無効にした場合の効果とほぼ同等です。

さらに、非対称型メモリ構成による効果も測定できます。非対称型メモリ構成については、前述の PRIMERGY BX920 S2 の例を参照してください。非対称構成での 2 ~ 3 %の低下予測は、前述したリモートアクセスのみによる低下と、こうした状況は最大でアクセスの 1/6 で発生するという統計的な考察に基づいています。

関連資料

[関連資料 1] PRIMERGY システム<http://ts.fujitsu.com/primergy>**[関連資料 2] PRIMERGY のパフォーマンス**http://ts.fujitsu.com/products/standard_servers/primergy_bov.html**[関連資料 3] STREAM ベンチマーク**<http://www.cs.virginia.edu/stream/>**[関連資料 4] OLTP-2 ベンチマーク**<http://docs.ts.fujitsu.com/dl.aspx?id=9775e8b9-d222-49db-98b1-4796fbc6d67a>**[関連資料 5] ベンチマークの概要 SPECcpu2006**<http://docs.ts.fujitsu.com/dl.aspx?id=00b0bf10-8f75-435f-bb9b-3eceb5ce0157>**[関連資料 6] PC サーバ PRIMERGY (プライマジー)**<http://primeserver.fujitsu.com/primergy>

お問い合わせ先

富士通テクノロジー・ソリューションズWeb サイト : <http://ts.fujitsu.com>**PRIMERGY のパフォーマンスとベンチマーク**<mailto:primergy.benchmark@ts.fujitsu.com>

知的所有権を含むすべての権利は弊社に帰属します。製品データは変更される場合があります、またデータの配布は状況に応じます。データおよび図の完全性、事実性、または正確性について、弊社は一切の責任を負いません。本書に記載されている記号表示は、それぞれのメーカーの商標または著作権である場合があります。第三者が各自の目的でこれらを使用した場合、当該所有者の権利を侵害することがあります。詳細については、http://ts.fujitsu.com/terms_of_use.html を参照してください。

2011-06-06 WW JA

Copyright © Fujitsu Technology Solutions GmbH 2011