

ホワイトペーパー

FUJITSU Server PRIMERGY

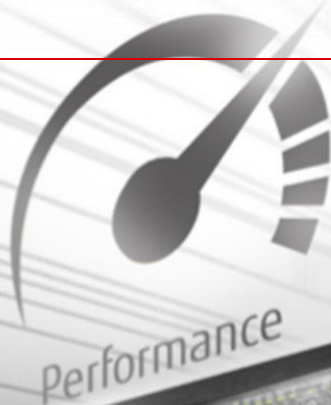
Xeon E5-2600 v4 (Broadwell-EP) 搭載システムのメモリパフォーマンス

Xeon E5-2600 v4 (Broadwell-EP) 搭載 FUJITSU Server PRIMERGY モデルでも、QuickPath インターコネクト (QPI) のメモリアーキテクチャーの改善で、パフォーマンスが目覚しく向上します。これは 5 世代のシステムで証明されています。このホワイトペーパーでは、アーキテクチャーの重要な特徴と最近の改善点について説明し、それが商用アプリケーションのパフォーマンスに与える影響を数量化します。

バージョン

1.0

2016-03-31



目次

ドキュメントの履歴	2
はじめに	3
メモリアーキテクチャー	5
DIMM スロットとメモリコントローラー	5
DDR4 トピックと使用可能な DIMM タイプ	8
メモリ周波数の定義	10
BIOS パラメーター	12
Memory Configuration (メモリ構成) のメモリパラメーター	12
CPU Configuration (CPU 構成) のメモリパラメーター	12
パフォーマンスを考慮したメモリ構成	14
パフォーマンスモード構成	14
独立モード構成	15
対称型メモリ構成	16
メモリパフォーマンスに対する定量的影響	17
測定ツール	18
STREAM ベンチマーク	18
SPECint_rate_base2006 ベンチマーク	18
メモリチャネルへのインターリーブ	19
メモリ周波数	20
DIMM タイプの影響	21
キャッシュコヒーレンスプロトコルの最適化	23
リモートメモリへのアクセス	23
冗長性を考慮した際のメモリパフォーマンス	24
関連資料	25
お問い合わせ先	25

ドキュメントの履歴

バージョン 1.0 (2016 年 3 月 31 日)

初版

はじめに

現在のデュアルソケット PRIMERGY サーバに搭載されている Intel Xeon E5-2600 v4 (Broadwell-EP) プロセッサは最新の 14 nm 製造プロセスを使用して生産されており、旧世代の Haswell-EP (22 nm 製造プロセスを使用して生産) と比較して、基本的にパフォーマンスが向上しています。マイクロアーキテクチャー、チップセットおよび Grantley-EP プラットフォームは、旧世代から引き継がれています。

新世代のパフォーマンスは、ほとんどの負荷シナリオで旧世代に比べて 20~30% の向上を果たしています。この成果の大きな要因は、プロセッサあたりの最大コア数が 18 から 22 に増えたことによります。メモリシステムにも、新世代のパフォーマンスの向上に寄与する新しい機能があります。

Haswell-EP 世代では、DDR4 メモリテクノロジーが導入されました。現行の Broadwell-EP システムもこのテクノロジーを使用していますが、最大メモリ周波数は Haswell-EP が 2133 MHz であったのに対して、Broadwell-EP システムは 2400 MHz をサポートしており、これが新たな特徴となっています。QuickPath インターコネクト (QPI) リンクの最大周波数には変更はありません。Haswell-EP および Broadwell-EP は、いずれも最大 9.6 GT/s をサポートします。

メモリシステムのその他の新機能は、旧世代で導入されたキャッシュコヒーレンスプロトコルを選択できる機能に関連しています。Haswell-EP *Early Snoop*、*Home Snoop*、および *Cluster-on-die* の 3 バージョンについては既に周知されていますが、これらに、*Home Snoop with Directory and OSB* オプションが加わりました。同時に、旧世代では *Early Snoop* がデフォルトであったのに対して、このオプションは新たな BIOS のデフォルトとなりました。これらのバージョンは、ローカルおよびリモートメモリアクセスに関するレイテンシと帯域幅のトレードオフの扱いが異なります。ほとんどのアプリケーションでは、デフォルト設定から外れた設定にする必要はありません。微小なパフォーマンスの違いと対応するテストにもこだわる場合は、このトピックをさらに深く掘り下げることも有用です。

それ以外の点では、実績のある旧世代の QPI ベースのメモリアーキテクチャーの基本機能が維持されています。プロセッサには、オンチップメモリコントローラが搭載されており、各プロセッサはそれぞれに割り当てられているメモリモジュールのグループを制御します。このローカルメモリアクセスのパフォーマンスは、非常に優れています。このプロセッサは、片方向シリアル 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にメモリの内容を要求します。リモートアクセスのパフォーマンスは、さほど高くありません。ローカルメモリとリモートメモリアクセスを区別するこのアーキテクチャーは、NUMA (Non-Uniform Memory Access : 非均等型メモリアクセス) タイプのアーキテクチャーです。

ただし、直前の世代である Haswell-EP [L3] のメモリシステム機能について言えば、多くがそのまま引き継がれています。各プロセッサに 4 つのメモリチャネルが存在し、DIMM スロットは各チャネルに 3 本あります。そのため、プロセッサあたりの最大 DIMM 数は 12 のままです。新しい点は、前述のとおり、最大メモリ周波数が 2133 MHz から 2400 MHz に増えたことです。その結果、最も基本的なメモリパフォーマンスの指標である最大メモリ帯域幅が、デュアルソケットサーバで 10% 以上も向上し、130 GB/s となりました。

このドキュメントでは、一方で最新のサーバ世代の新しいメモリシステム機能について見ていきます。もう一方で、これまでの号と同様に、強力なシステムを構成する上で不可欠な QPI ベースのメモリアーキテクチャーの基本的な知識について説明します。ここでは、次の点を取り上げます。

- NUMA アーキテクチャーであるため、両プロセッサのメモリを可能な限り同等の構成にする必要があります。これは、各プロセッサが原則としてそのローカルメモリ上で動作するためです。
- メモリアクセスを並列化し、さらに高速化するために、物理アドレス空間の隣接する領域をメモリシステムの複数のコンポーネントに分散させます。これは技術用語で *インターリーブ* と呼ばれます。インターリーブは 2 つの次元で行われます。まず、プロセッサあたり 4 つのメモリチャネルが横方向に存在します。各プロセッサ上の PRIMERGY コンフィギュレータを同じタイプの 4 枚 1 組の DIMM によるパフォーマンスモード構成にすることで、この方向への最適なインターリーブを実現しています。また、個々のメモリチャネルの中でもインターリーブを実現しています。このための決定的なメモリリソースが、いわゆるランク数です。ランク数は、DIMM の下位構造で、ここに DRAM (Dynamic Random Access Memory : ダイナミックランダムアクセスメモリ) チップのグループが統合されています。個々のメモリアクセスでは、常にこのようなグループを参照します。
- メモリ周波数はパフォーマンスに影響を与えます。プロセッサタイプ、DIMM タイプ、メモリ容量、および BIOS 設定に応じて、2400、2133、1866 MHz のいずれかになります。メモリ容量を非常に大きくすると、メモリ周波数が制限されます。このため、パフォーマンスと容量のバランスを取る必要があります。

このホワイトペーパーでは、影響を与える要因を取り上げ、数量化しています。数量化には、STREAM と SPECint_rate_base2006 のベンチマークを使用します。STREAM でメモリ帯域幅を測定します。SPECint_rate_base2006 は、商用アプリケーションのパフォーマンスのモデルとして使用されます。

測定結果では、プロセッサのパフォーマンスごとの影響をパーセンテージで示します。構成プロセッサモデルが強力であるほど、本書で取り上げているメモリ構成の問題について十分に考慮する必要があります。

ミラーリングやランクスペアリングなど、冗長性を考慮する場合のメモリパフォーマンスについては、本書の最後にまとめています。

メモリアーキテクチャー

ここでは、5 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。2 つ目のセクションでは、使用可能な DIMM タイプを示します。続く 3 つ目のセクションでは、有効なメモリ周波数への影響について説明します。4 つ目のセクションでは、メモリシステムに影響を与える BIOS パラメーターについて説明します。最後のセクションでは、メモリパフォーマンスを最適化した DIMM 構成例のリストを示します。

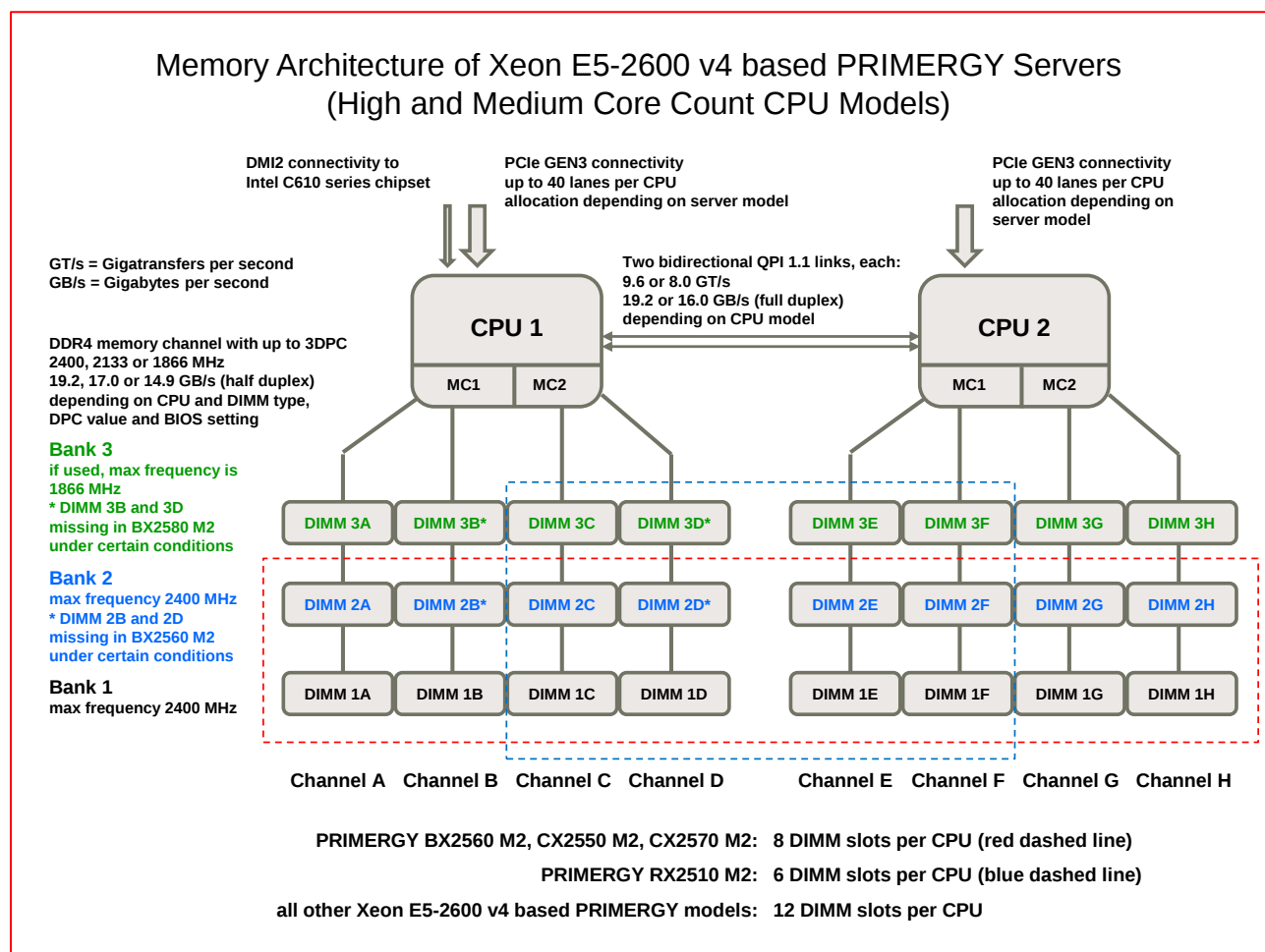
DIMM スロットとメモリコントローラー

次の図は、メモリシステムの構造を示します。2 つの図の細かな違いについて説明する前に、まず重要な要素が共通している点に注目してみましょう。

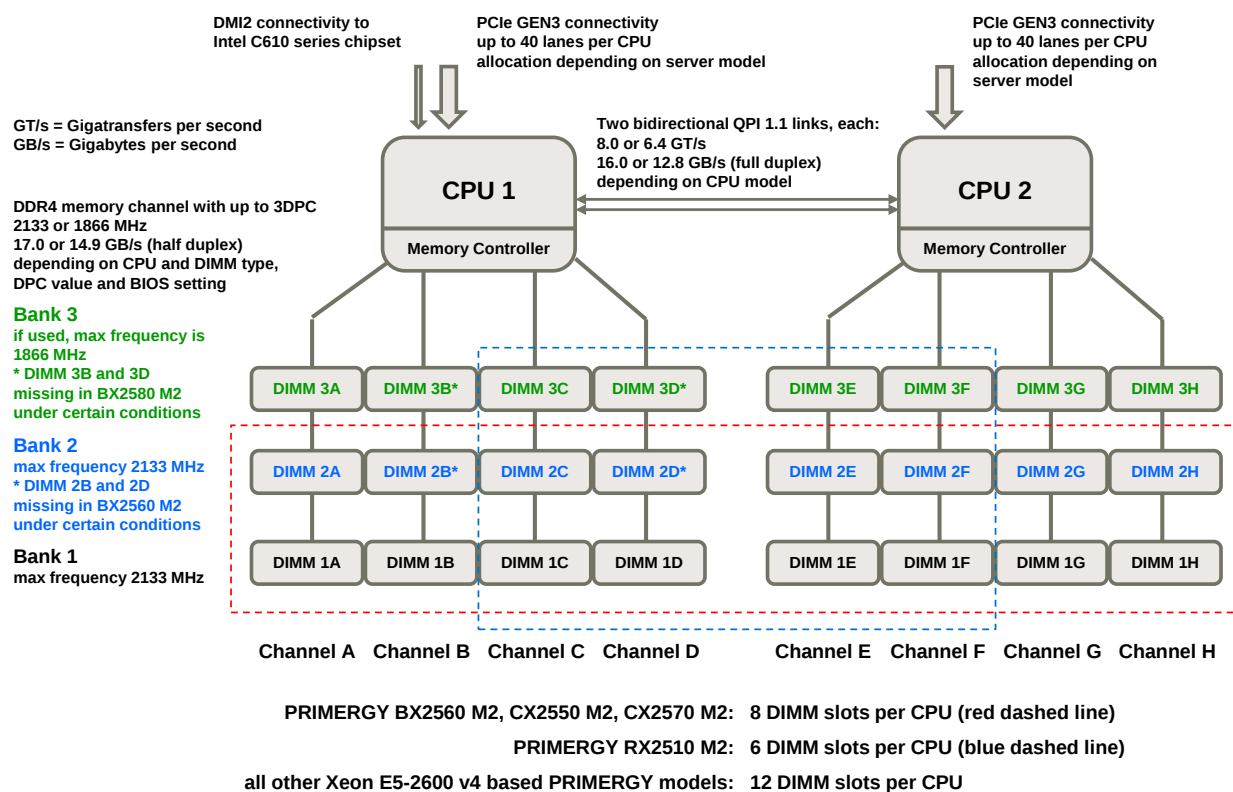
Xeon E5-2600 v4 搭載 PRIMERGY サーバは、通常、プロセッサあたり 12 本の DIMM スロットを装備しています。例外は、PRIMERGY BX2560 M2 モデル、CX2550 M2 モデル、および CX2570 M2 モデルで、フォームファクターが高密度であるため、スロット数は 8 となっています。もう 1 つの例外は、PRIMERGY RX2510 M2 モデルです。このシステムは、コスト面に基づいて最適化されているため、プロセッサあたりの利用可能なメモリチャンネル 4 つのうち 2 つが存在しません。その結果として、プロセッサあたり 6 DIMM スロットとなっています。

ブレードサーバ PRIMERGY BX2560 M2 および BX2580 M2 には 3 つ目の例外がある可能性があります。高パフォーマンスプロセッサのグループ（詳細はコンフィギュレータをご確認ください）には CPU 1 用により大きなヒートシンクが必要となり、その結果として 2 DIMM スロットは省略されます。

図でリソースのメモリチャンネルと QPI リンクを見ると、周波数と帯域幅の間に接続があり、これはそれぞれの実質データパス幅によるものです。DDR4 メモリチャンネルでは 64 ビット、QPI リンクでは 16 ビットです。双方向 QPI リンクの場合は、帯域幅が各方向で有効であるため、全二重方式と呼ばれています。メモリチャンネルでは、リード/ライトアクセスがデータパスを共有しなければならないため、半二重方式と呼ばれます。



Memory Architecture of Xeon E5-2600 v4 based PRIMERGY Servers (Low Core Count CPU Models)



周波数と帯域幅についての詳細では、それぞれの図に表示されるプロセッサクラスも考慮されます。

1つのプロセッサには、通常4つのメモリチャンネルが存在します。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。この値は、DPC（チャンネルあたりのDIMM数）と呼ばれます。以降、この用語を使用します。チャンネルの構成が異なる場合、発生する最大DPC値は、メモリ構成が周波数に与える影響の決定的な要因になります。

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4つまたは2つ（PRIMERGY RX2510 M2）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、サーバのシステムボード上で、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、チャンネル全体で最適なインターリーブが得られます。インターリーブは、メモリパフォーマンスに影響を与える主要な要因です。

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。最大構成がない場合、空のCPUソケットに割り当てられたスロットは使用できません。

プロセッサあたりのメモリチャンネル数は、すべてのXeon E5-2600 v4プロセッサファミリーで共通です。ただし、このセクションに2つの図があるのは、プロセッサあたりのメモリコントローラー数に違いがあるためです。

Xeon E5プロセッサファミリーでは、進化的な開発が続いています。NehalemからSandy Bridgeまでは、デュアルソケットサーバのEPプロセッサには1つのメモリコントローラーしかありませんでした。Ivy Bridgeでは、2つの最も強力なプロセッサモデルに（非常に少数のグループですが）、初めて2つのメモリコントローラーが搭載されました。HaswellとBroadwellでは、この機能は拡張され、プロセッサコア数が中位から上位のモデルのすべてのプロセッサに搭載されています。これはそれぞれのプロセッサファミリーの約半分を占めます。

正確な分類については、次の図の「ダイ設計」列を参照してください。

- HCC（高コア数）および MCC（中コア数）プロセッサには、2つのメモリコントローラーが搭載されています（1つ目の図）。
- LCC（低コア数）プロセッサには1つのコントローラーが搭載されています（2つ目の図）。

プロセッサ（システムリリース以降）									
プロセッサ	コア数	スレッド数	ダイ設計	キャッシュ [MB]	QPI スピード [GT/s]	公称周波数 [GHz]	最大 ターボ 周波数 [GHz]	最大 メモリ 周波数 [MHz]	TDP [W]
Xeon E5-2623 v4	4	8	LCC	10	8.00	2.60	3.20	2133	85
Xeon E5-2637 v4	4	8	LCC	15	9.60	3.50	3.70	2400	135
Xeon E5-2603 v4	6	6	LCC	15	6.40	1.70	該当なし	1866	85
Xeon E5-2643 v4	6	12	LCC	20	9.60	3.40	3.70	2400	135
Xeon E5-2609 v4	8	8	LCC	20	6.40	1.70	該当なし	1866	85
Xeon E5-2620 v4	8	16	LCC	20	8.00	2.10	3.00	2133	85
Xeon E5-2667 v4	8	16	LCC	25	9.60	3.20	3.60	2400	135
Xeon E5-2630L v4	10	20	LCC	25	8.00	1.80	2.90	2133	55
Xeon E5-2630 v4	10	20	LCC	25	8.00	2.20	3.10	2133	85
Xeon E5-2640 v4	10	20	LCC	25	8.00	2.40	3.40	2133	90
Xeon E5-2650 v4	12	24	MCC	30	9.60	2.20	2.90	2400	105
Xeon E5-2650L v4	14	28	MCC	35	9.60	1.70	2.50	2400	65
Xeon E5-2660 v4	14	28	MCC	35	9.60	2.00	3.20	2400	105
Xeon E5-2680 v4	14	28	MCC	35	9.60	2.40	3.30	2400	120
Xeon E5-2683 v4	16	32	HCC	40	9.60	2.10	3.00	2400	120
Xeon E5-2690 v4	16	32	MCC	40	9.60	2.60	3.50	2400	135
Xeon E5-2697A v4	16	32	HCC	40	9.60	2.60	3.60	2400	145
Xeon E5-2695 v4	18	36	HCC	45	9.60	2.10	3.30	2400	120
Xeon E5-2697 v4	18	36	HCC	45	9.60	2.30	3.60	2400	145
Xeon E5-2698 v4	20	40	HCC	50	9.60	2.20	3.60	2400	135
Xeon E5-2699 v4	22	44	HCC	55	9.60	2.20	3.60	2400	145

HCC と MCC の違いは、チップ内で構成されるプロセッサコアのトポロジーに依存します。コアと L3 キャッシュ共有は、縦横に配置されています。HCC モデルには 4 列、MCC には 3 列、LCC には 2 列あります。LCC は全コアがリング状に相互接続され、そこにメモリコントローラーも接続されています。HCC と MCC では相互接続されたリングが 2 つあり、コントローラーが 1 つずつ接続されています。

「ダイ設計」の分類は、基本的にはプロセッサモデルのパフォーマンスレベルの分類にすぎません。定量的なメモリパフォーマンステストは、トピックに応じて HCC、MCC、LCC のプロセッサクラスの分類に従って別々に実施しました。表の右から 2 番目の列については、サポートされているメモリ周波数による分類を使用しました。

DDR4 トピックと使用可能な DIMM タイプ

Broadwell-EP 搭載の PRIMERGY サーバには DDR4 SDRAM メモリモジュールが使用されています。旧世代の Haswell-EP で、DDR3 から DDR4 への移行が行われました。DDR3 と DDR4 の仕様を含む JEDEC (Joint Electron Device Engineering Council : 電子機器技術評議会) 規格では、メモリーメーカーとシステムメーカーの橋渡しとなるインターフェースを定義しています。

DDR4 テクノロジーは、現在も比較的新しく、DDR3 と比べて重要な違いがあります。DDR3 から DDR4 への移行は進化という点では自然なことであり、1 回限りのパフォーマンス向上にはとどまりませんでした。

- DDR4 では DIMM あたりのピン数が増えたため、DDR3 と DDR4 の DIMM ソケットに互換性はありません。古い DDR3 メモリモジュールを DDR4 ベースのシステムで使用することはできません。
- DDR4 では、3200 MHz までのメモリ周波数をサポートします。この周波数範囲は、今後、数世代のサーバで使用されることとなります。DDR3 ベースのサーバ世代と同様に、周波数は 266 MHz ずつ継続的に上昇していきます。Broadwell-EP の場合、これは 2400 MHz に到達しています。Haswell-EP 搭載システムは、最大 2133 MHz をサポートしていました。
- DDR4 の重要なメリットは、DIMM がわずか 1.2 V で動作する点です。DDR3 では、1.5 V または 1.35 V (低電圧版) でした。これは、データ転送速度が同じ場合、約 30% の消費電力の節約に相当します。
- DDR3 テクノロジーの最初のフェーズのときと同様に、現在のところ DDR4 に低電圧版はありません。したがって、メモリシステムに関する限り、BIOS におけるパフォーマンスと消費電力の構成トレードオフもほとんど関係ありません。それでも、これらのトレードオフは、プロセッサに対して重要な役割を果たしています。

次の表に従い、Xeon E5-2600 v4 搭載 PRIMERGY サーバのメモリ構成に DIMM を検討しました。DIMM には、レジスタード (RDIMM)、ロードリデュースド (LRDIMM) があります。それぞれのケースで、混在構成は表の 3 つのセクション内でのみ可能です。つまり、タイプ x4 の RDIMM とタイプ x8 の RDIMM の間のみで可能です。主 LRDIMM については、RDIMM とは混在できません。

DIMM タイプ	制御	最大周波数 (MHz)	電圧	ランク数	容量	SDDC	GB あたりの相対価格
8GB (1x8GB) 1Rx4 DDR4-2400 R ECC	Registered	2400	1.2	1	8 GB	ある	1.2
16GB (1x16GB) 1Rx4 DDR4-2400 R ECC	Registered	2400	1.2	1	16 GB	ある	¹⁾
16GB (1x16GB) 2Rx4 DDR4-2400 R ECC	Registered	2400	1.2	2	16 GB	ある	1.0
32GB (1x32GB) 2Rx4 DDR4-2400 R ECC	Registered	2400	1.2	2	32 GB	ある	1.2
8GB (1x8GB) 2Rx8 DDR4-2400 R ECC	Registered	2400	1.2	2	8 GB	ない	¹⁾
16GB (1x16GB) 2Rx8 DDR4-2400 R ECC	Registered	2400	1.2	2	16 GB	ない	1.1
64GB (1x64GB) 4Rx4 DDR4-2400 LR ECC	ロードリデュースド	2400	1.2	4	64 GB	ある	3.0

¹⁾ 本ドキュメントの公開時点ではまだ販売されていません

どの DIMM タイプでも、データは 64 ビット単位で転送されます。これは、DDR-SDRAM メモリテクノロジーの特徴です。64 ビットの帯域幅のメモリ領域は、DRAM チップのグループから DIMM 上に設定されます。この個々のチップが 4 ビットまたは 8 ビットを受け持ちます（タイプ名のコード x4 または x8 を参照してください）。このようなチップグループをランクと呼びます。表に示すように、1 ランク、2 ランク、または 4 ランクの DIMM タイプがあります。4 ランクの DIMM のメリットは、最大容量である点にありますが、同時に DDR4 の仕様はメモリチャネルあたり最大 8 ランクしかサポートしません。メモリチャネルあたりの利用可能なランク数は、パフォーマンスに一定の影響を及ぼします。これについては後述します。

そのことを踏まえると、2 つの DIMM タイプの重要な特徴は、次のようになります。

- RDIMM : 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）。メモリチャネルの負荷が軽減されることで、最大 3 DPC（チャネルあたりの DIMM）での構成が可能になります。
- LRDIMM : 制御コマンドとは別に、データ自体も DIMM 上のコンポーネントにバッファーされます。さらに、この DIMM タイプの「ランク乗算」機能により、いくつかの物理ランクを仮想ランクにマップできます。したがって、メモリコントローラーは仮想ランクを監視するだけです。この機能は、メモリチャネル内の物理ランクの数が 8 を超える場合に有効になります。

DIMM の x4 または x8 構造は、エラー（修正できるものと修正できないものを含めて）の検出度に影響します。これが、タイプ x4 の RDIMM をタイプ x8 の RDIMM と混在させることができない理由です。SDDC（Single Device Data Correction、表の最後から 2 列目を参照）は、x4 モジュールに限定された拡張 ECC 機能であり、DRAM チップ全体の障害を補うことができます。

RDIMM または LRDIMM のうち、どのタイプグループが望ましいかは、通常、必要なメモリ容量によって決まります。周波数とランク数のパフォーマンスへの影響は、2 タイプとも同じです。こうした影響は、タイプとは関係がありません。タイプごとにパフォーマンスに影響が出ることもありますが、非常に小さいため、たいていの場合は無視できます。タイプごとの影響として、ここで例を 2 つ挙げておきます。ただし、大きな影響ではないため、システムによる定量的評価には表れません。

- DIMM にコンポーネントを追加したために、DIMM タイプの RDIMM と LRDIMM の複雑さが増し、アクセス遅延がわずかに数ナノ秒ほど長くなります。
- LRDIMM を使った構成でメモリチャネルあたりの物理ランク数が 8 を超える場合、ランク乗算を行うと、最大メモリ帯域幅とアプリケーションパフォーマンスがわずかに（RDIMM を使った構成と比べて 5 %未満）減少します。

特定の構成の有効な周波数は、一連の影響に左右されます。DIMM タイプの表内の最大周波数は、こうした有効な周波数の上限を表しているにすぎません。

表の最終列は、各 DIMM の価格を相対比で示しています。この価格は、2016 年 4 月現在の PRIMERGY RX2540 M2 の料金表を元としています。ここでは 16 GB の 2Rx4 RDIMM を基準とし（1.0 として強調表示）、GB あたりの価格比を示します。本ドキュメントシリーズの以前の各号と比較すると、相対メモリ価格が常に変化してきたことが分かります。

PRIMERGY モデルによっては、一部の 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。

メモリ周波数の定義

メモリの周波数には、2400、2133、および 1866 MHz の 3 種類があります。システムに電源が入ると、周波数が BIOS によって定義され、プロセッサごとではなくシステムごとに適用されます。まず、定義上、構成プロセッサモデルが非常に重要になります。

このセクションでは、Xeon E5-2600 v4 モデルを、次の表（前に示した表と同じもの）の右から 2 番目の列に従って分類することをお勧めします。この列は、サポートされる最大メモリ周波数を示しています。また、この分類では QPI クロックレートに従った周波数を対象にしています。

プロセッサ（システムリリース以降）									
プロセッサ	コア数	スレッド数	ダイ設計	キャッシュ [MB]	QPI スピード [GT/s]	公称周波数 [GHz]	最大 ターボ 周波数 [GHz]	最大 メモリ 周波数 [MHz]	TDP [W]
Xeon E5-2623 v4	4	8	LCC	10	8.00	2.60	3.20	2133	85
Xeon E5-2637 v4	4	8	LCC	15	9.60	3.50	3.70	2400	135
Xeon E5-2603 v4	6	6	LCC	15	6.40	1.70	該当なし	1866	85
Xeon E5-2643 v4	6	12	LCC	20	9.60	3.40	3.70	2400	135
Xeon E5-2609 v4	8	8	LCC	20	6.40	1.70	該当なし	1866	85
Xeon E5-2620 v4	8	16	LCC	20	8.00	2.10	3.00	2133	85
Xeon E5-2667 v4	8	16	LCC	25	9.60	3.20	3.60	2400	135
Xeon E5-2630L v4	10	20	LCC	25	8.00	1.80	2.90	2133	55
Xeon E5-2630 v4	10	20	LCC	25	8.00	2.20	3.10	2133	85
Xeon E5-2640 v4	10	20	LCC	25	8.00	2.40	3.40	2133	90
Xeon E5-2650 v4	12	24	MCC	30	9.60	2.20	2.90	2400	105
Xeon E5-2650L v4	14	28	MCC	35	9.60	1.70	2.50	2400	65
Xeon E5-2660 v4	14	28	MCC	35	9.60	2.00	3.20	2400	105
Xeon E5-2680 v4	14	28	MCC	35	9.60	2.40	3.30	2400	120
Xeon E5-2683 v4	16	32	HCC	40	9.60	2.10	3.00	2400	120
Xeon E5-2690 v4	16	32	MCC	40	9.60	2.60	3.50	2400	135
Xeon E5-2697A v4	16	32	HCC	40	9.60	2.60	3.60	2400	145
Xeon E5-2695 v4	18	36	HCC	45	9.60	2.10	3.30	2400	120
Xeon E5-2697 v4	18	36	HCC	45	9.60	2.30	3.60	2400	145
Xeon E5-2698 v4	20	40	HCC	50	9.60	2.20	3.60	2400	135
Xeon E5-2699 v4	22	44	HCC	55	9.60	2.20	3.60	2400	145

DIMM タイプおよびメモリ構成の DPC 値によっても、周波数が制限されます。プロセッサタイプ、DIMM タイプ、および DPC 値はメモリ周波数に大きな影響を及ぼします。これを BIOS で無効にすることはできません。ただし、BIOS パラメーターの DDR Performance (DDR パフォーマンス) を使用することで、限定的ですがパフォーマンスと消費電力のどちらを優先させるかを選択できます（詳細は後述）。パフォーマンスを選択した場合、有効なメモリ周波数は次の表のようになります。これはデフォルトの BIOS 設定です。

DDR Performance = Performance optimized (パフォーマンスに最適化) (デフォルト)						
CPU タイプ	RDIMM			LRDIMM		
	1DPC	2DPC	3DPC	1DPC	2DPC	3DPC
DDR4-2400	2400	2400	1866	2400	2400	1866
DDR4-2133	2133	2133	1866	2133	2133	1866
DDR4-1866	1866	1866	1866	1866	1866	1866

PRIMERGY BX2560 M2、CX2550 M2、CX2570 M2 には 3DPC 構成のロットがないので 3DPC 列は無関係です。

RDIMM に関して、2DPC で 2400 MHz、3DPC で 1866 MHz がサポートされるのは（赤で表示）、Xeon E5-2600 v4 搭載 PRIMERGY サーバの特別な機能です。Intel 仕様では、これらのケースでは 2133 MHz（2DPC）および 1600 MHz（3DPC）を考慮しています。PRIMERGY の各モデルでは、BIOS バージョンの最新のコンフィギュレータとデータシートが常に決定的な要素となります。周波数を示すだけであれば不要ですが、この特別な機能を参照できるように、この表では RDIMM と LRDIMM を分けて記載しています。前述のように、現在のところ DDR4 メモリモジュールに低電圧版はありません。DDR4 モジュールは常に 1.2 V 電圧で動作します。これは、低電圧版の 1.35 V DDR3 より低い値です。本ドキュメントシリーズの以前の号で紹介していた「DDR Performance = Low-voltage optimized（低電圧に最適化）」の設定は、Xeon E5-2600 v4 搭載 PRIMERGY サーバにはありません。

ただし、メモリ周波数を下げることでわずかに消費電力を節約できます。しかし、メモリモジュールの消費電力は主に電圧の影響を受ける点に注意してください。メモリ周波数を下げるとシステムパフォーマンスも低下するため（本ドキュメントの第 2 部で説明）、次の表に従って設定を行う際は、ある程度の注意を払うことをお勧めします。注意を払うとは、本稼働の前に影響をテストするということです。

DDR Performance = Energy optimized (消費電力に最適化)						
CPU タイプ	RDIMM			LRDIMM		
	1DPC	2DPC	3DPC	1DPC	2DPC	3DPC
DDR4-2400	1866	1866	1866	1866	1866	1866
DDR4-2133	1866	1866	1866	1866	1866	1866
DDR4-1866	1866	1866	1866	1866	1866	1866

BIOS パラメーター

前のセクションでは、BIOS パラメーター DDR Performance を見ましたが、ここでは、メモリシステムに影響を与える他の BIOS オプションを見ていきます。このパラメーターは、Advanced (詳細) の下のサブメニュー、Memory Configuration (メモリ構成) と CPU Configuration (CPU 構成) にあります。

Memory Configuration (メモリ構成) のメモリパラメーター

次の 4 つのパラメーターがあります。それぞれ下線付きのオプションがデフォルトです。

- Memory Mode : Normal / Mirroring / Sparing
- NUMA : Disabled (無効) / Enabled (有効)
- DDR Performance : Performance optimized / Energy optimized
- Patrol Scrub : Disabled / Enabled

最初のパラメーター Memory Mode (メモリモード) は冗長性機能を扱います。これらは、RAS (Reliability: 信頼性、Availability: 可用性、Serviceability: サービス性) 機能の一部であり、メモリをミラーリングするか (ミラーリング)、メモリエラーが頻繁に発生する場合に DIMM ランクのレベルでメモリスペアを有効化すること (スペアリング) で、耐故障能力を向上させます。System Architect で構成中にこれらの機能が要求された場合、工場で適切なデフォルト設定が行われます。それ以外の場合、このパラメーターは Normal (通常の冗長性なし) に設定されます。冗長性機能がシステムパフォーマンスに与える影響に関する数値を後で示します。

物理アドレス空間をローカルメモリのセグメントから構築するか、またオペレーティングシステムに構造を通知するかを NUMA パラメーターで定義します。デフォルト設定は *Enabled* です。明確な理由がないこれを限り変更しないでください。このトピックの数量的な面については、後述します。

3 番目のパラメーター DDR Performance は、メモリ周波数に関係しています。これについては、最後のセクションで説明します。

Patrol Scrub パラメーターは、あらかじめ *Enabled* に設定されています。メインメモリは、24 時間サイクルで修正可能なエラーの検索が実行され、必要に応じて修正が開始されます。これにより、自動修正が不可能になるようなメモリエラーの累積を防ぎます (対応するレジスターでカウントされます)。感度の高いパフォーマンス指標がある場合は、この機能を一時的に無効にすることもできます。ただし、パフォーマンスに及ぶ効果を実証するのは難しい場合があります。

CPU Configuration (CPU 構成) のメモリパラメーター

サブメニューの CPU Configuration (CPU 構成) は扱う範囲が広いため、詳細については『[Xeon E5-2600 v4 搭載システムのための BIOS 最適化](#)』[L6] で解説します。次の 3 つのパラメーターがメモリシステムにとっては特に重要です。下線付きのオプションがデフォルト設定です。

- COD Enable (COD 有効化) : Disabled (無効) / Enabled (有効) / Auto (自動)
- Early Snoop (アーリースヌープ) : Disabled (無効) / Enabled (有効) / Auto (自動)
- Home Snoop Dir OSB (ホームスヌープ、ディレクトリ OSB) : Disabled (無効) / Enabled (有効) / Auto (自動)

これらのパラメーターは、キャッシュコヒーレンスプロトコルの 4 つのバージョン (*Home Snoop with Directory and OSB*、*Cluster-on-die*、*Home Snoop*、および *Early Snoop*) を選択するために使用されます。このようなプロトコルによって、共有アドレスエリアを持つマルチプロセッサシステムで、複数のプロセッサキャッシュで同一のメモリアドレスがキャッシュされることによる不整合の発生を防ぐことができます。

Xeon E5-2600 v4 搭載システムの場合、デフォルト設定は *Home Snoop with Directory and OSB* となり、相応の理由がない限り変更することはできません。プロトコルの最適化で難しいのは、アプリケーションのパフォーマンスに対する影響が注意深いテストでしか判明しない点です。後述のガイダンスによる単なる定量的な比較 (概算) は十分ではありません。

そのようなテストでデフォルト設定から変更する場合は、以下のプランに従って、3 つの BIOS パラメーターを設定し、プロトコルバージョンを変更します。

プロトコル	COD Enable	Early Snoop	Home Snoop Dir OSB
Home Snoop Dir OSB	Auto	Auto	Auto
Cluster-on-die	Enabled	Auto	Auto
Home Snoop	Disabled	Disabled	Disabled
Early Snoop	Disabled	Enabled	Disabled

基本的にキャッシュコヒーレンスプロトコルに関係するのは、システムの L3 キャッシュと、対応するアドレスを担当するメモリコントローラーです。正確に言うと、これらは、キャッシングエージェント (CA) またはホームエージェント (HA) と呼ばれるサブコンポーネントで、QPI ネットワークを介してチップ内でメッセージを交換します。関係する可能性のあるキャッシングエージェントを含むメッセージタイプをスヌープと言います。以上を前提に 4 つのプロトコルを説明すると、次のようになります。

プロトコル	説明
Home Snoop Dir OSB	担当 HA によるホームスヌープ。この HA はメモリ内拡張ディレクトリを持ちます。予備的なスヌープ (OSB = Opportunistic Snoop Broadcast) によって補完され、使用可能な QPI 帯域幅、およびその他のヒューリスティクスに依存します。
Cluster-on-die	2 つのメモリコントローラーを持つプロセッサモデル専用。各プロセッサに 2 つの NUMA ノードがあり、プロセッサコアの半分、L3 キャッシュの半分、メモリコントローラー 1 つが、各 NUMA ノードに割り当てられます。メモリ内ディレクトリおよびオンダイディレクトリキャッシュを持つ HA によるスヌープ。卓越した NUMA 機能で、負荷に対応します。
Home Snoop	アクティベートする CA は、特にスヌープを送信する担当 HA とやり取りします。多数のプロセッサを搭載するシステムの正統な動作です。QPI ネットワークが解放されますが、レイテンシが増大します。
Early Snoop	アクティベートする CA による予備的なスヌープ。可能な限り最も早い時点で実行されます。数個のプロセッサを搭載するシステムの正統な動作です。QPI ネットワークの負荷が高く、レイテンシは短くなります。

COD (cluster-on-die) の説明で述べた 2 つのメモリコントローラーを搭載するプロセッサモデルへの制限は、プロセッサコア数に応じた前述の HCC、MCC、LCC の分類に基づいています。COD は、HCC/MCC クラスのプロセッサ専用です。プロセッサの表は、前のセクションで表示済みのため、ここでは示していません。

Broadwell-EP 搭載 PRIMERGY サーバに関して実行される標準ベンチマークの多くで COD を構成しましたが、その他のプロトコルが使用されることもあります。ただし、BIOS のデフォルト設定とは別の設定を未テストのまま使用することはお勧めしません。

最後に、4 つのプロトコルがメモリパフォーマンスのレイテンシと帯域幅にどのように影響するかについて、定性的に示しました。ローカルメモリとリモートメモリのアクセスの区別を行う必要があります。レイテンシは低い値ほど優れており、帯域幅は高い値ほど優れています。メモリのレイテンシと帯域幅に関するアプリケーション要件が判明している場合は、次の表を参考に最適なプロトコルを見つけてください。

スヌープモードの相対的なパフォーマンス				
パフォーマンスの評価指標	Home Snoop Dir OSB	COD (HCC および MCC CPU のみ)	Home Snoop	Early Snoop
L3 キャッシュヒットレイテンシ	低	最低	低	低
ローカルメモリレイテンシ	低	最低	高	中
リモートメモリレイテンシ	低	低 - 高	低	最低
ローカルメモリ帯域幅	高	高	高	低
リモートメモリ帯域幅	高	中	高	低 - 中

パフォーマンスを考慮したメモリ構成

要するに、メモリパフォーマンスに影響するのは主にメモリ周波数であることが証明されています。構成のメモリ周波数に影響するいくつかの依存関係が定義されています。各ユーザーは、自分の環境のメモリ周波数を把握しておくべきです。

さらに、メモリパフォーマンスにも影響する構成機能がいくつかあります。ランク数、冗長機能の有効化、NUMA 機能の無効化、キャッシュコヒーレンスプロトコルのオプションなど、個々の DIMM タイプの特別な機能です。本ドキュメントの第 2 部では、これらのトピックのテスト結果を報告しますが、ほとんどの顧客環境では大きな影響はないと思われるため、多くのユーザーは読み飛ばしても構いません。

パフォーマンスモード構成

常に観察すべき 2 つ目の要因は、DIMM 配置の影響です。最小構成（構成プロセッサあたり 8 GB DIMM 1 枚）から最大構成（複数の 64 GB DIMM からなるフル構成）の間には、いくつかのメモリパフォーマンスについての理想的な構成があります。次の表に、特に興味深い構成を挙げています（すべての構成を網羅している訳ではありません）。

これらの構成で、プロセッサあたり全部で 4 本のメモリチャネルという点は同じです。バンク単位の構成では、同タイプの DIMM 4 枚セットを使用しています。これにより、メモリアクセスは、これらのメモリスistemリソースに均等に分散されます。技術的に言えば、メモリチャネル経由で最適な 4-WAY インターリーブが実現します。PRIMERGY コンフィギュレータでは、これをパフォーマンスモード構成と呼んでいます。

パフォーマンスモード構成 Xeon E5-2600 v4 搭載 PRIMERGY サーバ (第 3 バンクにメモリが割り当てられている構成は、プロセッサあたり 12 枚の DIMM を搭載できる PRIMERGY モデルでのみ可能)							
1 CPU システム	2 CPU システム	DIMM タイプ	DIMM サイズ (GB) バンク 1	DIMM サイズ (GB) バンク 2	DIMM サイズ (GB) バンク 3	CPU あたり最大 MHz	注
32 GB	64 GB	DDR4-2400 R	8			2400	2Rx8 版も使用可能 (++)
64 GB	128 GB	DDR4-2400 R	16			2400	
96 GB	192 GB	DDR4-2400 R	16	8		2400	混在構成 (-)
128 GB	256 GB	DDR4-2400 R	16	16		2400	4-WAY ランクインターリーブ (++)
192 GB	384 GB	DDR4-2400 R	16	16	16	1866	周波数低下の 3DPC (-)
		DDR4-2400 R	32	16		2400	混在構成 (-)
256 GB	512 GB	DDR4-2400 R	32	32		2400	4-WAY ランクインターリーブ (++)
320 GB	640 GB	DDR4-2400 R	32	32	16	1866	混在構成 (-)
384 GB	768 GB	DDR4-2400 R	32	32	32	1866	周波数低下の 3DPC (-)
512 GB	1024 GB	DDR4-2400 LR	64	64		2400	(++)
768 GB	1536 GB	DDR4-2400 LR	64	64	64	1866	最大構成

表は左端の総メモリ容量に従って構成されています。総容量は、1 つまたは 2 つのプロセッサ構成で定義されています。2 プロセッサ構成の場合、どちらのプロセッサについてもメモリ構成は同じであるという想定です。右の列は、使用した DIMM タイプです。RDIMM または LRDIMM テクノロジー、GB DIMM サイズが決定要因です。DIMM サイズがバンク単位で表記されているのは、パフォーマンスモード構成を使用するため、DIMM を 4 枚 1 組で（バンク単位で）構成するからです。

表の最小構成で 2 つのプロセッサに対して 64 GB となっているのは、各プロセッサに 8 GB DIMM を 4 枚（32 GB）使用しているためです。

パフォーマンスモード構成では、バンクあたり 4 枚の同一の DIMM をセットで使用する必要がありますが、次の制限を満たしていれば、別のバンクで異なる DIMM サイズを使用することもできます。

- RDIMM と LRDIMM を併用することはできません。
- タイプ x4 と x8 の RDIMM を併用することはできません。
- 構成はバンク 1 から 3 に DIMM サイズを小さくしながら進めます。大きいモジュールほど先に取り付けます。

最後の列は注意書きです。たとえば、混在構成では、決定的なパフォーマンス要因の 4-WAY チャネルインターリーブがあっても、単一 DIMM タイプ構成と比べると若干パフォーマンスが低下することが示されています。これは、個々のメモリチャネル内での複雑なアドレッシングに起因します。

当然ながら、表には、Xeon E5-2600 v4 搭載 PRIMERGY サーバで実施した標準ベンチマークの構成も含まれています。これらには、注の列に ++ が付けられています。

表の右から 2 列目は、それぞれの構成で達成可能な最大メモリ周波数を示しています。ただし、その値に達するかどうかは使用するプロセッサモデルにも依存します。この列では、メモリ容量とメモリパフォーマンス（周波数）のトレードオフも示されています。

独立モード構成

これには、パフォーマンスモードや冗長モードには含まれない構成がすべて含まれます。RDIMM と LRDIMM、x4 と x8 RDIMM の「混在禁止」ルール以外の制約はありません。

プロセッサあたりの DIMM が 4 つ未満の構成、つまりパフォーマンスモード構成に必要な最小数未満の構成にも特に注意する必要があります。このような構成にする理由は、節電や、必要なメモリ容量が少ないためです。節約が DIMM 数の最小化で実現する場合もあります。この後で紹介する、3 つ以下のメモリチャネルを使った構成がシステムパフォーマンスに及ぼす影響を示した定量的評価から、次のような事項が推奨されます。

- LCC プロセッサクラス（低コア数）については、プロセッサあたり DIMM 1 枚だけ（最小構成）での動作はお勧めしません。プロセッサあたりの DIMM 数を 2 または 3 とすると、パフォーマンスと電力消費量のバランスが取れた結果になります。
- HCC（高コア数）および MCC（中コア数）プロセッサクラスでは、プロセッサあたり 1 枚または 3 枚の DIMM での動作はお勧めしません。プロセッサあたりの DIMM 数を 2 とすると、パフォーマンスと電力消費量のバランスが取れた結果になります。

推奨されない構成は、全体（プロセッサあたり 1 DIMM）または部分（HCC および MCC プロセッサの場合、プロセッサあたり 3 DIMM）ということになります。以下に示すように、メモリチャネルを介した 1-WAY インターリーブには、商用アプリケーションのパフォーマンスが最大 30 %低下するというデメリットがあります。この HCC と LCC プロセッサに DIMM 3 枚を使用する構成では、2 つのメモリコントローラーに 3 枚の DIMM を均等に配分できません。

対称型メモリ構成

最後のこのセクションでは、すべてのプロセッサのメモリを可能な限り同等に構成し、BIOS の *NUMA = enabled* のデフォルト設定を確たる理由なく変更するべきではないということに再度焦点を当てます。このように考慮されるのは、QPI ベースのアーキテクチャーを持つシステムのみです。

工場でのプレインストールでは、このような状況が当然考慮されています。要求されたメモリモジュールは、プロセッサ全体に可能な限り均等に分散されます。

こうした手法と、関連するオペレーティングシステムによって、ローカルのハイパフォーマンスメモリで可能な限りアプリケーションを実行する前提条件が整備されます。プロセッサコアのメモリアクセスは、通常、各プロセッサに直接割り当てられた DIMM モジュールに対して行われます。

これにどのようなパフォーマンス上のメリットがあるのかを見積もるため、2-WAY サーバのメモリが対称型に構成されているものの、BIOS オプションが *NUMA = disabled* に設定されている場合の測定結果を以下のリストに示します。統計上、各 2 次メモリアクセスは、リモートメモリに対して行われます。アプリケーションが 100 % リモートメモリによって実行される非対称型メモリ構成、または片側メモリ構成では、50 % / 50 % の場合の 2 倍パフォーマンスが低下するものとして見積もる必要があります。

たとえば、1 つ目のプロセッサに 8 枚の DIMM、2 つ目のプロセッサに 4 枚の DIMM の構成は、パフォーマンスモードの基準を満たします。なぜなら、プロセッサごとのメモリチャネルが同じように処理されるからです。これは PRIMERGY の注文および構成プロセスでの考え方です。ただし、このような構成は推奨されません。

メモリパフォーマンスに対する定量的影響

メモリシステムの機能とその定性的情報を説明した後は、メモリ構成に関するパフォーマンスの向上と低下について説明します。その準備として、最初のセクションでは、メモリパフォーマンスの特徴を表すための使用する 2 つのベンチマークについて説明します。

その後、すでに説明した特徴であるメモリチャネルのインターリーブ、メモリ周波数、DIMM タイプの影響、およびキャッシュコヒーレンスプロトコルについて、その影響の大きさの順に説明します。最後に、NUMA = disabled で、冗長性を考慮する場合のメモリパフォーマンスについて測定します。

1 つの例外として、定量的テストは、Xeon E5-2600 v4 プロセッサファミリーのパフォーマンスクラス別に実施しました。通常はプロセッサコア数による HCC、MCC、LCC の分類を使用していますが、メモリ周波数をテストする際は、DDR4 の周波数分類の方が適切です。例外：各タイプの DIMM の測定は、タイプ HCC のプロセッサ 1 つのみを使用して行われました。

測定は、Linux オペレーティングシステムの 2 つのプロセッサを搭載した PRIMERGY RX2530 M2 で行われました。次の表は、測定対象となった構成、特にプロセッサクラスの代表的な構成を示します。

SUT (System Under Test : テスト対象システム)	
ハードウェア	
モデル	PRIMERGY RX2530 M2
プロセッサ	Xeon E5-2698 v4 (HCC, DDR4-2400) × 2 Xeon E5-2660 v4 (MCC, DDR4-2400) × 2 Xeon E5-2630 v4 (LCC, DDR4-2133) × 2 Xeon E5-2609 v4 (MCC, DDR4-1866) × 2
メモリタイプ	8GB (1x8GB) 1Rx4 DDR4-2400 R ECC 8GB (1x8GB) 2Rx8 DDR4-2400 R ECC 16GB (1x16GB) 2Rx4 DDR4-2400 R ECC 16GB (1x16GB) 2Rx8 DDR4-2400 R ECC 16GB (1x16GB) 1Rx4 DDR4-2400 R ECC 32GB (1x32GB) 2Rx4 DDR4-2400 R ECC 64GB (1x64GB) 4Rx4 DDR4-2400 LR ECC
ディスクサブシステム	HD SATA 6G 1TB 5.4Krpm 1 台 (SATA/SAS 用オンボードコントローラーを介して)
ソフトウェア	
BIOS	1.4.0
オペレーティングシステム	Red Hat Enterprise Linux Server リリース 6.7

16 GB 2Rx4 RDIMM は、通常は以下に説明するテストセットに使用されました。メモリチャネル間でのインターリーブのテストには 32 GB 2Rx4 RDIMM がテストケースとして使用されました。テストに必要なメインメモリの最小容量を実現するため、プロセッサあたりの DIMM は 1 つまたは 2 つのみとしました。表に示したすべての DIMM は、DIMM タイプの影響についてのテストセットのみで使用されました。

次の表は、相対的なパフォーマンスを示します。理想的なメモリ条件下での STREAM および SPECint_rate_base2006 のベンチマークの絶対測定値は（通常、次の表で 1.0 の値に相当）、Xeon E5-2600 v4 搭載 PRIMERGY サーバの個別のパフォーマンスレポートに記載されています。

このテストから得られた 1 つの重要な結果を、最初に明らかにしておきましょう。使用するプロセッサモデルが強力であるほど、パフォーマンスへの影響が顕著になり、構成の詳細をより慎重に検討する必要があります。HCC クラスの最も強力で高価なプロセッサについて不可欠な考慮事項も、LCC クラスについては、多くの場合無視できます。

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いました。

STREAM ベンチマーク

STREAM ベンチマーク（開発者：John McCalpin 氏）[関連資料 4] 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 ベンチマーク

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。

メモリチャネルへのインターリーブ

インターリーブは、最初のブロックは最初のチャネルに、2 番目のブロックは 2 番目のチャネルにという具合に、プロセッサ単位で 4 つのメモリチャ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ネルに分散されます。このようなパフォーマンスの向上は、並列化によるものです。チャネルインターリーブのブロックサイズは、64 バイトというキャッシュラインサイズに基づいています。キャッシュラインサイズは、プロセッサの観点におけるメモリアクセスの単位です。

次の表は、パフォーマンスモードでのメモリ構成によって実現される理想的な 4-WAY インターリーブを行わない場合のパフォーマンスの低下を示しています。この表では、すでに取り上げた、プロセッサが強力になるほどパフォーマンスへの影響が顕著になるという事実が示されています。

ベンチマーク	CPU 種類	4-way	3-way	2-way	1-way
STREAM	高コア数	1.00		0.52	0.26
	中コア数	1.00		0.52	0.26
	低コア数	1.00	0.84	0.59	0.30
SPECint_rate_base2006	高コア数	1.00		0.87	0.66
	中コア数	1.00		0.93	0.77
	低コア数	1.00	0.99	0.96	0.82

このテスト（および同じ分類の後出のテスト）に使用したプロセッサモデルは、Xeon E5-2698 v4 (HCC)、Xeon E5-2660 v4 (MCC)、および Xeon E5-2630 v4 (LCC) です。使用した DIMM タイプは 16 GB 2Rx4 RDIMM です。1DPC 構成で使用しています。

SPECint_rate_base2006 に関する評価は、商用アプリケーションのパフォーマンスに関するものです。STREAM で示されているように、メモリ帯域幅の関係は、特に HPC (High-Performance Computing : 高性能コンピューティング) 環境では、特定のアプリケーション領域において除外できない極端なケースとして理解する必要があります。ただしこうした動作は、ほとんどの商用のワークロードでは見られません。STREAM および SPECint_rate_base2006 に関する解釈の質は、このセクションで取り上げているパフォーマンス面だけでなく、以降のすべてのセクションにも当てはまります。

2-WAY および 3-WAY インターリーブのパフォーマンスの低下が穏やかなのは、それなりの理由があります。必要となるメモリ容量が少ないか、または低消費電力のために DIMM 数が最小限に抑えられるためです。1-WAY インターリーブは推奨できません。これは厳密に言うとインターリーブではなく、分類上そのようにと呼ばれているだけです。この場合、プロセッサとメモリシステムの潜在的なパフォーマンスのバランスが取れていません。

HCC および MCC クラスのプロセッサは、2 つのメモリコントローラーを使った構成により、3-WAY インターリーブはサポートしていません。そのため、これらのプロセッサモデルでプロセッサあたり 3 つの DIM を構成した場合にどうなるか、がポイントとなります。

これらは、物理アドレス領域を、インターリーブが異なる複数のセグメントに分割する必要があるケースの例です。分割が必要なその他の例としては、メモリチャネルあたりの部分容量（チャネルあたりの GB）が異なる構成が挙げられます。このような構成は、サイズの異なる DIMM を使った場合や、同サイズの DIMM を 5 つ以上使った場合に生じることがあります。すべての例に共通するのは、メモリチャネル間を切り替える方法で標準化アドレス領域セグメントを設定することができない点です。切り替えは常に「均等に」行われます。このようなケースでは、既存の DIMM をグループ化することにより、可能なかぎり高いインターリーブでセグメントが生成されます。次の表は、その例を 2 つ示しています。表の左の列は、CPU あたり 4 チャネルのそれぞれに接続した DIMM の枚数を表しています。

アプリケーションのメモリパフォーマンスは、アプリケーションにメモリが提供されるセグメントによって異なる可能性があります。パフォーマンスが特に重要なケースでは、分割の必要な構成は避けた方がいい場合もあります。

分割が必要な DIMM 構成 の例 (CPU あたり)	アドレス領域セグメント	サイズ/インターリーブ
1-1-1-0 (HCC および MCC)	1-0-1-0	アドレス領域の 66 %/2WAY
	0-1-0-0	アドレス領域の 33%/1WAY
2-1-1-1	1-1-1-1	アドレス領域の 80%/4WAY
	1-0-0-0	アドレス領域の 20%/1WAY

メモリ周波数

有効なメモリ周波数への影響は、これまでのセクションで詳しく説明しています。大メモリ構成（3DPC 構成）および節電（BIOS パラメーター DDR Performance で管理）は、実効周波数がそのプロセッサタイプでサポートされる最大周波数より低くなってしまう原因になります。

次の表は、影響を比較し、バランスを取る際に役立ちます。最初の表の数値は、一連のすべての測定に共通の最小メモリ周波数 1866 MHz に基づきます。2 番目の表は、同じ情報を異なる観点から捉えたものです。数値は、理想的なケース、つまりそのプロセッサクラスで最大限の周波数に基づいています。

ベンチマーク	CPU 種類	1866 MHz	2133 MHz	2400 MHz
STREAM	DDR4-2400	1.00	1.11	1.21
	DDR4-2133	1.00	1.07	
	DDR4-1866	1.00		
SPECint_rate_base2006	DDR4-2400	1.00	1.01	1.02
	DDR4-2133	1.00	1.01	
	DDR4-1866	1.00		

ベンチマーク	CPU 種類	1866 MHz	2133 MHz	2400 MHz
STREAM	DDR4-2400	0.83	0.92	1.00
	DDR4-2133	0.93	1.00	
	DDR4-1866	1.00		
SPECint_rate_base2006	DDR4-2400	0.98	0.99	1.00
	DDR4-2133	0.99	1.00	
	DDR4-1866	1.00		

このテストで使用したプロセッサモデルは、Xeon E5-2660 v4（DDR4-2400）、Xeon E5-2630 v4（DDR4-2133）、および Xeon E5-2609 v4（DDR4-1866）です。使用した DIMM タイプは 16 GB 2Rx4 RDIMM です。1DPC 構成で使用しています。

BIOS で「DDR Performance = *Energy optimized*（電力に最適化）」と設定すると、周波数は常に 1866 MHz になります。ただし、消費電力は DIMM の電圧の影響が大きく、メモリ周波数の影響は小さいため、得られる節電効果はわずかです。新しい DDR4 モジュールの電圧は常に 1.2 V なので、最低でも 1.35 V あった DDR3 世代と比べると節電効果があります。これが、*Energy optimized*（電力に最適化）という設定をお勧めしない理由です。

メモリ周波数の低下がメモリ容量に関係している場合、万全を期するために 1 つの問題について説明する必要があります。例えば、I/O 速度など、メモリ容量はアプリケーションのパフォーマンスに対して暗黙の影響を及ぼします。こういった影響は、当然ながらこのセクションで基にしているテストでは考慮していません。表で比較すると、メモリ周波数の違いだけがパフォーマンスに影響を及ぼしています。

DIMM タイプの影響

Xeon E5-2600 v4 搭載 PRIMERGY サーバの一般公開時には、7 タイプの DIMM が計画されています。ただし、特定のサーバの例外および特性については、それぞれのコンフィギュレータを参照してください。

次の表に、これらの DIMM タイプの同一条件下でのパフォーマンスの違いを示します。

- すべての測定は、HCC プロセッサモデルの Xeon E5-2698 v4 を使用して実施されました。この強力なプロセッサモデルを使用することで、パフォーマンスの違いが非常に明確になります。比較的微細な相違の測定であるため、一連の測定は例外的にこのモデルのみを使用して実施されました。
- これらの測定で、すべてのメモリチャンネルが同じ構成であった、すなわちパフォーマンスモード構成での比較が行われたことは明らかです。取り付けられた DIMM の数は、1DPC の測定では 8、2DPC の測定では 16 でした。
- すべての測定は、メモリ周波数 2400 MHz で一様に実施されました。この一般条件のため、検討することができたのは 1DPC 構成および 2DPC 構成のみでした。
- この表では、現時点で最善のメモリパフォーマンスを提供する 16 GB 2Rx4 RDIMM を使用した 2DPC 構成（太字で強調表示）を基準としています。実現されるメモリ容量が十分である場合に限り、ベンチマークではこの DIMM が推奨されます。

DIMM タイプ	構成	STREAM	SPECint_rate_base2006
8GB (1x8GB) 1Rx4 DDR4-2400 R ECC	1DPC	0.87	0.97
	2DPC	0.98	0.99
8GB (1x8GB) 2Rx8 DDR4-2400 R ECC	1DPC	0.98	0.99
	2DPC	1.00	1.00
16GB (1x16GB) 1Rx4 DDR4-2400 R ECC	1DPC	0.86	0.96
	2DPC	0.97	0.99
16GB (1x16GB) 2Rx4 DDR4-2400 R ECC	1DPC	0.98	0.99
	2DPC	1.00	1.00
16GB (1x16GB) 2Rx8 DDR4-2400 R ECC	1DPC	0.96	0.99
	2DPC	0.99	0.99
32GB (1x32GB) 2Rx4 DDR4-2400 R ECC	1DPC	0.97	0.99
	2DPC	0.99	0.99
64GB (1x64GB) 4Rx4 DDR4-2400 LR ECC	1DPC	0.95	0.98
	2DPC	0.91	0.98

ここで示されているパフォーマンスの違いは、インターリーブの形式が別ということが主な原因です。物理アドレス空間のセットアップ時にメモリリソースを切り替える方法は、メモリチャンネルでのインターリーブからチャンネルのランクでのインターリーブまで継続できます。

ランクインターリーブは、アドレスビットを経由して直接制御されます。3-WAY ケースの確立のためにチャンネルインターリーブで実行されるビット演算は行われません。この理由から、2 のべき乗でのインターリーブのみが問題となります。つまり、2-WAY、4-WAY または 8-WAY のランクインターリーブのみが存在します。メモリチャンネルでの奇数のランク数は、常に 1-WAY インターリーブとなりますが、これは分類上そのように呼ばれているだけです。1-WAY の場合、ランクは次のランクに変更される前にすべて利用されます。

メモリチャンネルあたりのランク数は、構成の DIMM タイプおよび DPC 値に従います。たとえば、表に示されているデュアルランク DIMM を使用した 1DPC 構成の場合は、2-WAY のランクインターリーブ、2DPC 構成の場合は 4-WAY のインターリーブが許容されます。

ランクインターリーブの粒度は、チャンネルでのインターリーブより大きくなります。チャンネルでのインターリーブは 64 バイトキャッシュラインサイズに使用されています。ランクインターリーブは、オペレーティングシステムの 4 KB ページサイズに向かい、DRAM メモリの物理特性に関係します。メモリセルは、大ま

かに言って 2 つの次元で行われます。行（ページとも呼ばれる）が開かれ、列項目が読み取られます。ページが開いている間、より大幅に低いレイテンシで他の列の値を読み取ることもできます。さらに大まかなランクインターリーブは、この機能に最適化されます。

2-WAY および 4-WAY ランクインターリーブは、非常に優れたメモリパフォーマンスを実現します。最終的なパフォーマンスを考えた場合、4-WAY インターリーブにのみわずかなメリットがありますが、通常は無視できる程度です。

たとえば、この表で最も目立つパフォーマンス低下は 1DPC での 8 GB 1Rx4 RDIMM ですが、これはランクインターリーブが欠落していることから説明が付きまします。シングルランク DIMM を使用した 1DPC 構成を除き、このケースは、たとえば、1 番目のバンクに 16 GB 2Rx4 RDIMM、2 番目のバンクに 8 GB 1Rx4 RDIMM を使用する混在構成でも発生する可能性があります。このような欠落、または 1-WAY ランクインターリーブといったケースでは、ある程度のパフォーマンス低下に注意する必要があります。特にパフォーマンスが重視され、強力なプロセッサモデルを使用している状況では、これを回避する必要があります。

結果の表には、ランクインターリーブによる大きな影響のほか、その他の微細な影響もいくつか含まれます。たとえば、メモリチャネルに 4 つを超えるランクがあるため、DRAM をリフレッシュするために実行されるランクごとのオーバーヘッドが、否定的な意味で目立つようになります。このリフレッシュは、すべてのランクで共有される、メモリチャネルのアドレス行ごとの一定の基本負荷を表します。これにより、2DPC 構成の結果が対応する 1DPC 構成よりも悪くなる唯一のケースである、前述した 4Rx4 LRDIMM との関係が説明されます。

キャッシュコヒーレンスプロトコルの最適化

キャッシュコヒーレンスのプロトコルの 3 つのバージョンである *Early Snoop*、*Home Snoop*、*Cluster-on-die* を選択できる機能は、旧世代の Haswell-EP の新機能に含まれていました。Haswell-EP のデフォルトは *Early Snoop* でした。Broadwell-EP にはさらに 4 番目のオプションである *Home Snoop with Directory and OSB* があり、これが新しいデフォルトとなります。詳細については、メモリシステムの BIOS オプションに関するセクションを参照してください。

次の表は、本ドキュメントで実施した 2 つの負荷（ベンチマーク）の効果を示しています。この表では、新しい BIOS デフォルトである *Home Snoop with Directory and OSB* のパフォーマンスを基準としています。

測定は、16 GB 2Rx4 RDIMM を使用した 1DPC 構成で行われました。

表には、数パーセントの範囲でパフォーマンスに影響が出ることが示されています。この表を評価する際は、テストセットアップ中の注意深いプロセスバインディングにより、両ベンチマークが極めて NUMA と相性が良いテストになっている点を考慮してください。したがって、商用アプリケーションパフォーマンスへの SPECint_rate_base2006 のモデル特性の適用は、この段階では限定的なものとする必要があります。

メモリコントローラーが 1 つのタイプ LCC のプロセッサは、*Cluster-on-die* プロトコルをサポートしないことを覚えておいてください。

ベンチマーク	CPU 種類	Home Snoop Dir OSB	Cluster-on-die	Home Snoop	Early Snoop
STREAM	高コア数	1.00	1.02	0.99	0.87
	中コア数	1.00	1.02	0.99	0.94
	低コア数	1.00		0.82	0.80
SPECint_rate_base2006	高コア数	1.00	1.02	0.98	0.98
	中コア数	1.00	1.01	0.98	0.99
	低コア数	1.00		0.98	1.00

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（プロセッサが自身のメモリチャネルの DIMM モジュールにアクセスする）。隣接するプロセッサのモジュールはまったくアクセスされないか、まれに QPI リンクを経由してアクセスされるのみです。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。

次の表は、プロセッサタイプ別の最大メモリ周波数を下回る 16 GB 2Rx4 RDIMM での 4-WAY ランクインターリーブパフォーマンスモードという理想的なメモリ構成でありながら、BIOS 設定が「NUMA = disabled」に設定されている場合の影響を示しています。各 2 次メモリアクセスがリモート DIMM、つまり隣接プロセッサに割り当てられた DIMM に対して行われ、データが QPI リンク経由で迂回する必要があるために、パフォーマンスの低下が発生します。

ベンチマーク	CPU 種類	NUMA = enabled	NUMA = disabled
STREAM	高コア数	1.00	0.78
	中コア数	1.00	0.77
	低コア数	1.00	0.80
SPECint_rate_base2006	高コア数	1.00	0.93
	中コア数	1.00	0.94
	低コア数	1.00	0.94

NUMA = disabled では、プロセッサの詳細メッシュ切り替えによって物理アドレス空間を設定しています。この切り替えは、両プロセッサが同一のメモリ容量であることを前提にしています。こうした一般的な状態が存在しない場合、アドレス空間はソケット間インターリーブが許可される主要部分と、プロセッサ - ローカルの残りの部分に分割されます。

システムソフトウェアまたはシステム関連ソフトウェアで NUMA がサポートされていないか、または十分でないため、NUMA = disabled に関する測定は、設定が推奨される例外的なケースとして、狭い範囲で実施しました。上記の測定はすべて、大部分あるいはすべてのアクセスがリモートメモリに対して行われる場合の影響を見積もる場合に役立ちます。この状況は、プロセッサごとの構成メモリ容量が大幅に異なる場合に発生します。ローカルアクセスと比較したパフォーマンスの低下は、表に示した低下分の最大 2 倍になることがあります。

この表は、低下率がこれまでの表ほど規則的ではないことを示しています。プロセッサが強力であればあるほど低下率が高くなる、という経験則は、ここでは必ずしも当てはまりません。これは、プロセッサタイプの QPI 周波数が 6.4~9.6 GT/s と異なることの影響に起因します。QPI リンクは、NUMA サポートを無効にした場合の帯域幅測定のボトルネックです。

冗長性を考慮した際のメモリパフォーマンス

Xeon E5-2600 v4 搭載 PRIMERGY サーバには、冗長オプションが 2 つあります。

ミラーリングでは 1 つのプロセッサの全 4 メモリチャネルが同一に構成されますが（パフォーマンスモードの構成ルールと一致します）、2 つのチャネルが他の 2 つのチャネルをミラーリングします。オペレーティングシステムは、実際に構成されているメモリの 50 % を利用できます。

スペアリング、厳密に言うとランクスペアリングの場合、メモリチャネルあたり 1 つのランクは未使用で、蓄積されるメモリエラーによってアクティブランクが動作を停止する場合の予防措置として予約されています。オペレーティングシステムが利用できる実質メモリ容量は、この場合 DIMM タイプおよび DPC 値によって異なります。正確な計算やスペアリングモード DIMM 構成の一般的な条件については、各 PRIMERGY モデルのコンフィギュレータを参照してください。

次の表では、それぞれのケースで最大メモリ周波数を下回る 16 GB 2Rx4 RDIMM でのパフォーマンスモード 2DPC 構成という理想的なメモリ構成でありながら、冗長オプションが有効化されている場合の影響を示しています。表の各列は、BIOS パラメーター Memory Mode（メモリモード）のオプションに対応しています。

ミラーリング下で発生する低下は、2-WAY チャネルインターリーブによるものよりも小さくなります。これは、ミラーの半分がどちらも読み取りアクセスで使用できるためです。

スペアリング下での関係は、DIMM タイプの影響についての前述のセクションで説明したランクインターリーブから導かれます。通常、予約ランクはメモリチャネルあたりの奇数のアクティブランク数になり、1-WAY ランクインターリーブになります。スペアリング列は、4-WAY ランクインターリーブと 1-WAY ランクインターリーブ間の違いによるものです。

ベンチマーク	CPU 種類	ノーマル	ミラーリング	スペアリング
STREAM	高コア数	1.00	0.68	0.87
	中コア数	1.00	0.69	0.90
	低コア数	1.00	0.77	0.93
SPECint_rate_base2006	高コア数	1.00	0.96	0.97
	中コア数	1.00	0.99	0.98
	低コア数	1.00	0.99	0.99

関連資料

PRIMERGY サーバ

[関連資料 1] <http://jp.fujitsu.com/primergy>

メモリパフォーマンス

[関連資料 2] このホワイトペーパー :

 <http://docs.ts.fujitsu.com/dl.aspx?id=8f372445-ee63-4369-8683-da9557673357>

 <http://docs.ts.fujitsu.com/dl.aspx?id=5c5111c5-7d23-4c2c-aa3f-88c7053fe41a>

 <http://docs.ts.fujitsu.com/dl.aspx?id=3ce313c6-6713-4350-880c-16959489a510>

[関連資料 3] Xeon E5-2600 v3 (Haswell-EP) 搭載システムのメモリパフォーマンス

<http://docs.ts.fujitsu.com/dl.aspx?id=342460c7-d7ec-4a7c-ade8-c33016154215>

ベンチマーク

[関連資料 4] STREAM

<http://www.cs.virginia.edu/stream/>

[関連資料 5] SPECcpu2006

<http://docs.ts.fujitsu.com/dl.aspx?id=00b0bf10-8f75-435f-bb9b-3eceb5ce0157>

BIOS 設定

[関連資料 6] Xeon E5-2600 v4 搭載システムのための BIOS 最適化

<http://docs.ts.fujitsu.com/dl.aspx?id=6caa932b-38c5-4088-8c13-7dcc9366f6a0>

PRIMERGY のパフォーマンス

[関連資料 7] <http://jp.fujitsu.com/platform/server/primergy/performance/>

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>