Fujitsu Server PRIMERGY
Xeon スケーラブル・プロセッサ（Ice Lake）搭載
システムのメモリパフォーマンス

第 3 世代 Xeon スケーラブル・プロセッサ（Ice Lake）搭載 Fujitsu Server PRIMERGY では、プロセッサコア数の増加、マイクロアーキテクチャーの進化、および、メモリアーキテクチャーの改善で、パフォーマンスが大幅に向上しています。このホワイトペーパーでは、メモリ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定量化します。

バージョン
1.1
2023-10-03

目次

はじめに

第 3 世代 Xeon スケーラブル・プロセッサ（Ice Lake）は、これまでの Xeon スケーラブル・プロセッサ（Skylake-SP、Cascade Lake-SP）世代の特長を受け継ぐとともに、Intel の最新 10nm 製造プロセスを使用することで、前世代のプロセッサから大きな性能向上を実現しています。プロセッサの最上位モデルでは、多くのシナリオで前世代の Cascade Lake-SP プロセッサに比べて 40～50 % の性能向上を果たしています。この成果の大きな要因として、プロセッサあたりのコア数が最大 28 から最大 40 に増えたこと、より進化したマイクロアーキテクチャーを採用したことが挙げられます。

メモリアーキテクチャーの観点でも、プロセッサは大幅に進化しました。キャッシュの増量に加えて、プロセッサには、前世代の倍の 4 つのオンチップメモリコントローラが搭載されました。また、メモリチャネルも、前世代の 6 つから 8 つに増加しました。最大メモリ周波数は、前世代のシステムが 2933 MHz であったのに対して、Ice Lake 世代では 3200 MHz をサポートします。これらにより、最大メモリ搭載時のピークメモリ帯域はプロセッサあたり 205 GB/s となりました。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。また、大容量の 256GB 3DS RDIMM のサポートにより、1 プロセッサあたり 4 TB のメモリを搭載可能です[1]。

このプロセッサが、隣接プロセッサのメモリ（リモートメモリ）の内容を要求するとき、Ultra Path Interconnect （UPI）リンクを使用します。リモートメモリへのアクセスのパフォーマンスは、ローカルメモリアクセスに比べると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の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。この プロセッサ間接続の速度は、前世代の 10.4 GT/s から 11.2 GT/s に引き上げられました。

Ice Lake 世代では、SNC（Sub NUMA Clustering）と呼ばれるプロセッサ内のクラスタリングに関するオプションに加えて、新しく UMA-Based Clustering と呼ばれるオプションが追加されました。これらのオプションは、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が、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以外、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。

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 UPI 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、各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できるようにするためです。
- 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 8 つのメモリチャネルが横方向に存在します。各プロセッサのメモリ搭載数を 8 の倍数とすることで、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、3200、2933、2667、2400、1867 MHz のいずれかになります。

このホワイトペーパーでは、メモリ性能に影響を与える要因を取り上げ、定量化しています。定量化には、STREAM と SPECrate2017 Integer 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。SPECrate2017 Integer 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比で示します。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。

---

[1] 最大搭載メモリ容量は、機種、プロセッサのタイプによって異なります。

ミラーリングや ADDDC 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

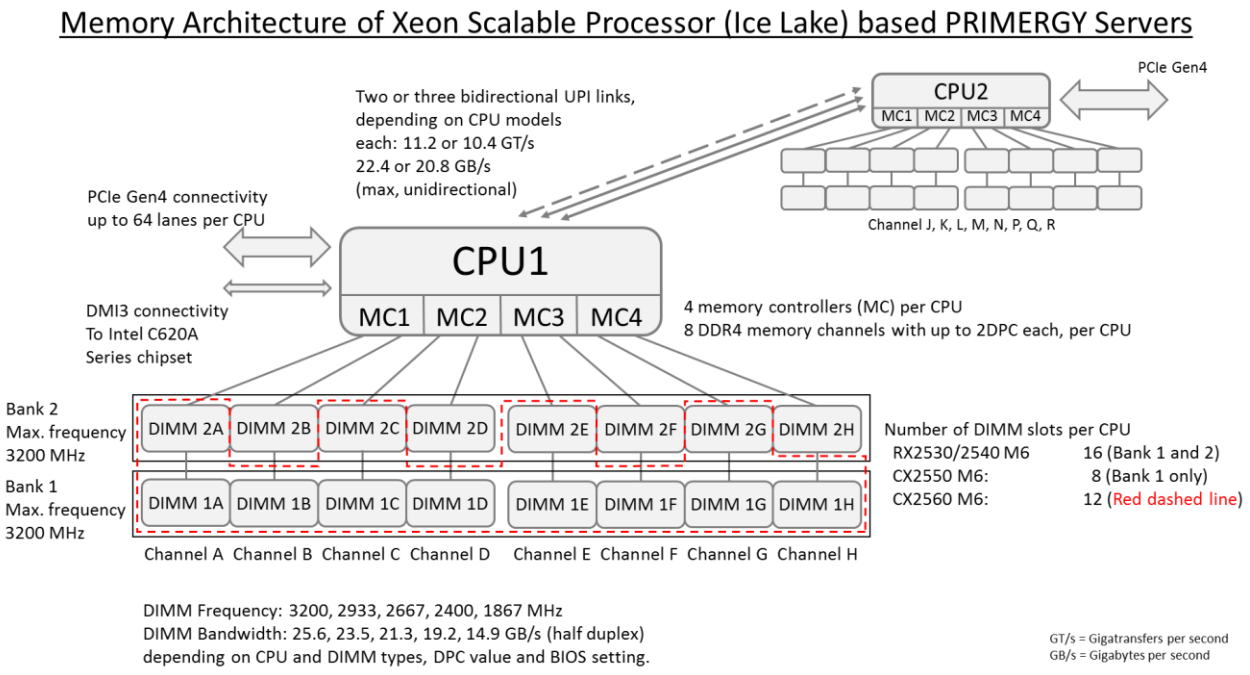
ここでは、5 部構成でメモリシステムの概要を説明します。まずブロック図で、利用可能な DIMM スロットの配置を説明します。2 つ目のセクションでは、使用可能な DIMM タイプを示します。続く 3 つ目のセクションでは、有効なメモリ周波数への影響について説明します。4 つ目のセクションでは、メモリシステムに影響を与える BIOS パラメーターについて説明します。最後のセクションでは、メモリパフォーマンスを最適化した DIMM 構成例のリストを示します。

DIMM スロットとメモリコントローラ

次の図は、第 3 世代 Xeon スケーラブル・プロセッサ（Ice Lake）搭載 PRIMERGY サーバでのメモリシステムの構造を示します。

Ice Lake 搭載 PRIMERGY サーバは、プロセッサあたり 16 本の DIMM スロットを装備しています。データパス幅は、DDR4 メモリチャネルでは 64 ビット、UPI リンクでは 20 ビットです。双方向である UPI リンクの場合、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。

前世代（Cascade Lake-SP、Skylake-SP）では、UPI リンクの速度は最大 10.4 GT/s でしたが、Ice Lake では最大 11.2 GT/s に向上しました。また、プロセッサ間の UPI リンクの本数は、2 ソケットの前世代 RX サーバの場合 2 本でしたが、Ice Lake 世代では最大 3 本に増加しました。これらにより、データベース処理のようなプロセッサ間にまたがるメモリアクセスが多いアプリケーションの性能向上が期待できます。



1 つのプロセッサには、4 つのメモリコントローラと 8 つのメモリチャネルが存在します。前世代のプロセッサは、2 つのメモリコントローラ、6 つのメモリチャネルでした。3200 MHz の DDR4 メモリの採用と相まって、メモリ帯域幅は理論値で 45 % 向上しています。

Broadwell-EP 世代ではチャネルあたりの DIMM 枚数である DPC（以降、この用語を使用します）の値が変わると、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した。しかし、この世代の Xeon スケーラブル・プロセッサ搭載 PRIMERGY サーバでは DPC によってメモリ周波数が低下しないようになりました。

以降では、「メモリバンク」という用語も使用します。図では、複数のチャネルに分配されている 8 つの DIMM のグループが、1 つのバンクを形成しています。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。プロセッサの搭載が最大構成でない場合、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。

プロセッサの正確な分類については、次の表を参照してください。

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	キャッシュ [MB]	UPI 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 Platinum 8380	40	80	60	11.2	2.3	3.4	3200	270
Xeon Platinum 8368Q	38	76	57	11.2	2.6	3.7	3200	270
Xeon Platinum 8368	38	76	57	11.2	2.4	3.4	3200	270
Xeon Platinum 8362	32	64	48	11.2	2.8	3.6	3200	265
Xeon Platinum 8360Y	36	72	54	11.2	2.4	3.5	3200	250
Xeon Platinum 8358P	32	64	48	11.2	2.6	3.4	3200	240
Xeon Platinum 8358	32	64	48	11.2	2.6	3.4	3200	250
Xeon Platinum 8352Y	32	64	48	11.2	2.2	3.4	3200	205
Xeon Platinum 8352V	36	72	54	11.2	2.1	3.5	2933	195
Xeon Gold 6354	18	36	39	11.2	3.0	3.6	3200	205
Xeon Gold 6348	28	56	42	11.2	2.6	3.5	3200	235
Xeon Gold 6346	16	32	36	11.2	3.1	3.6	3200	205
Xeon Gold 6342	24	48	36	11.2	2.8	3.5	3200	230
Xeon Gold 6338T	24	48	36	11.2	2.1	3.4	2933	165
Xeon Gold 6338	32	64	48	11.2	2.0	3.2	3200	205
Xeon Gold 6336Y	24	48	36	11.2	2.4	3.6	3200	185
Xeon Gold 6334	8	16	18	11.2	3.6	3.7	3200	165
Xeon Gold 6330N	28	56	42	11.2	2.2	3.4	2667	165
Xeon Gold 6330	28	56	42	11.2	2.0	3.1	2933	205
Xeon Gold 6326	16	32	24	11.2	2.9	3.5	3200	185
Xeon Gold 6314U	32	64	48	11.2	2.3	3.4	3200	205
Xeon Gold 6312U	24	48	36	11.2	2.4	3.6	2933	185
Xeon Gold 5320	26	52	39	11.2	2.2	3.4	2933	185
Xeon Gold 5318Y	24	48	36	11.2	2.1	3.4	2933	165
Xeon Gold 5318S	24	48	36	11.2	2.1	3.4	2933	165
Xeon Gold 5317	12	24	18	11.2	3.0	3.6	2933	150
Xeon Gold 5315Y	8	16	12	11.2	3.2	3.6	2933	140
Xeon Silver 4316	20	40	30	10.4	2.3	3.4	2667	150
Xeon Silver 4314	16	32	24	10.4	2.4	3.4	2667	135
Xeon Silver 4310	12	24	18	10.4	2.1	3.3	2667	120
Xeon Silver 4309Y	8	16	12	10.4	2.8	3.6	2667	105

定量的なメモリパフォーマンステストは、トピックに応じて表の右から 2 番目の列に記載したサポートされているメモリ周波数による分類を使用して、別々に実施しました。

DDR4 トピックと使用可能な *DIMM* タイプ

Ice Lake 搭載の PRIMERGY サーバには、これまでの Xeon スケーラブル・プロセッサ搭載 PRIMERGY サーバと同様に、DDR4 SDRAM メモリモジュールが使用されています。Ice Lake 搭載システムでは、以下の改善がなされました。

- DDR4 では、3200 MHz までのメモリ周波数をサポートします。Ice Lake 搭載システムの場合、最大メモリ周波数は 3200 MHz に到達しました。前世代の Cascade Lake-SP 搭載システムは、最大 2933 MHz をサポートしています。
- Ice Lake 搭載システムでは、256GB 3DS RDIMM を使用することで、1 ソケットあたり最大 4TB の DRAM を搭載可能です[2]。Cascade Lake-SP 搭載システムでは、最大 1.5TB の DRAM を搭載可能でした。

次の表に、Ice Lake 搭載 PRIMERGY サーバでサポートする DIMM を示します[3]。DIMM には、Registered DIMM（RDIMM）、Load Reduced DIMM（LRDIMM）、3DS Registered DIMM （3DS RDIMM） があります。混在構成は表の 4 つのセクション内でのみ可能です。すなわち、RDIMM x4、RDIMM x8、LRDIMM および 3DS RDIMM の混在はできません。

DIMM タイプ	制御	最大メモリ周波数 [MHz]	電圧 [V]	ランク数	容量	GB あたりの相対価格
8GB (1x8GB) 1Rx8 DDR4-3200 R ECC	Registered	3200	1.2	1	8 GB	0.92
16GB (1x16GB) 2Rx8 DDR4-3200 R ECC	Registered	3200	1.2	1	16 GB	0.98
16GB (1x16GB) 1Rx4 DDR4-3200 R ECC	Registered	3200	1.2	2	16 GB	0.98
32GB (1x32GB) 2Rx4 DDR4-3200 R ECC	Registered	3200	1.2	2	32 GB	**1.00**
64GB (1x64GB) 2Rx4 DDR4-3200 R ECC	Registered	3200	1.2	2	64 GB	1.00
64GB (1x64GB) 4Rx4 DDR4-3200 LR ECC	Load Reduced	3200	1.2	4	64 GB	1.34
128GB (1x128GB) 4Rx4 DDR4-3200 LR ECC	Load Reduced	3200	1.2	4	128 GB	1.34
128GB (1x128GB) 4Rx4 DDR4-3200 3DS R ECC	3DS Registered	3200	1.2	4	128 GB	1.10
256GB (1x256GB) 8Rx4 DDR4-3200 3DS R ECC	3DS Registered	3200	1.2	8	256 GB	1.10

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 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、4 ランク、または 8 ランクの DIMM タイプがあります。8 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、3 つの DIMM タイプの重要な特徴は、次のようになります。

---

[2] 最大搭載メモリ容量は、機種、プロセッサのタイプによって異なります。

[3] 機種により搭載可能な DIMM が異なります。詳細はシステム構成図を参照してください。

- RDIMM：メモリコントローラの制御コマンドは、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 2 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。
- 3DS RDIMM：Three Dimensional Stack （3DS） という規格に基づき、シリコン貫通電極(Through Silicon Via) 技術により複数枚のシリコン・ダイを積層させた RDIMM です。マスターと呼ばれる 1 枚のダイだけが外部と信号をやり取りし、それ以外のダイはスレーブとしてマスターとだけ信号をやり取りするアーキテクチャーを採用しており、大容量化や高速化が可能になります。

RDIMM、LRDIMM、または 3DS RDIMM のうち、どのタイプが望ましいかは、通常、必要なメモリ容量によって決まります。ただし、LRDIMM、3DS RDIMM には、若干の性能オーバーヘッドがあります。

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、2021 年 8 月現在の PRIMERGY RX2540 M6 の料金表を元にしています。ここでは 32 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。

なお、価格は、販売地域によって異なる場合があります。また、販売地域によっては、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、3200、2933、2667、2400 および 1867 MHz の 5 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 スケーラブル・プロセッサ 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。

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	キャッシュ [MB]	UPI 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 Platinum 8380	40	80	60	11.2	2.3	3.4	3200	270
Xeon Platinum 8368Q	38	76	57	11.2	2.6	3.7	3200	270
Xeon Platinum 8368	38	76	57	11.2	2.4	3.4	3200	270
Xeon Platinum 8362	32	64	48	11.2	2.8	3.6	3200	265
Xeon Platinum 8360Y	36	72	54	11.2	2.4	3.5	3200	250
Xeon Platinum 8358P	32	64	48	11.2	2.6	3.4	3200	240
Xeon Platinum 8358	32	64	48	11.2	2.6	3.4	3200	250
Xeon Platinum 8352Y	32	64	48	11.2	2.2	3.4	3200	205
Xeon Platinum 8352V	36	72	54	11.2	2.1	3.5	2933	195
Xeon Gold 6354	18	36	39	11.2	3.0	3.6	3200	205
Xeon Gold 6348	28	56	42	11.2	2.6	3.5	3200	235
Xeon Gold 6346	16	32	36	11.2	3.1	3.6	3200	205
Xeon Gold 6342	24	48	36	11.2	2.8	3.5	3200	230
Xeon Gold 6338T	24	48	36	11.2	2.1	3.4	2933	165
Xeon Gold 6338	32	64	48	11.2	2.0	3.2	3200	205
Xeon Gold 6336Y	24	48	36	11.2	2.4	3.6	3200	185
Xeon Gold 6334	8	16	18	11.2	3.6	3.7	3200	165
Xeon Gold 6330N	28	56	42	11.2	2.2	3.4	2667	165
Xeon Gold 6330	28	56	42	11.2	2.0	3.1	2933	205
Xeon Gold 6326	16	32	24	11.2	2.9	3.5	3200	185
Xeon Gold 6314U	32	64	48	11.2	2.3	3.4	3200	205
Xeon Gold 6312U	24	48	36	11.2	2.4	3.6	2933	185
Xeon Gold 5320	26	52	39	11.2	2.2	3.4	2933	185
Xeon Gold 5318Y	24	48	36	11.2	2.1	3.4	2933	165
Xeon Gold 5318S	24	48	36	11.2	2.1	3.4	2933	165
Xeon Gold 5317	12	24	18	11.2	3.0	3.6	2933	150
Xeon Gold 5315Y	8	16	12	11.2	3.2	3.6	2933	140
Xeon Silver 4316	20	40	30	10.4	2.3	3.4	2667	150
Xeon Silver 4314	16	32	24	10.4	2.4	3.4	2667	135
Xeon Silver 4310	12	24	18	10.4	2.1	3.3	2667	120
Xeon Silver 4309Y	8	16	12	10.4	2.8	3.6	2667	105

Ice Lake ではメモリ構成の DPC 値はメモリ周波数に影響を及ぼしませんが、プロセッサタイプ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 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 （性能に最適化、デフォルト）						
プロセッサタイプ	RDIMM		LRDIMM		3DS RDIMM	
	1DPC	2DPC	1DPC	2DPC	1DPC	2DPC
DDR4-3200	3200	3200	3200	3200	3200	3200
DDR4-2933	2933	2933	2933	2933	2933	2933
DDR4-2667	2667	2667	2667	2667	2667	2667

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。

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が、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）						
プロセッサタイプ	RDIMM		LRDIMM		3DS RDIMM	
	1DPC	2DPC	1DPC	2DPC	1DPC	2DPC
DDR4-3200	1867	1867	1867	1867	1867	1867
DDR4-2933	1867	1867	1867	1867	1867	1867
DDR4-2667	1867	1867	1867	1867	1867	1867

BIOS パラメーターの DDR Performance で Power balanced（消費電力との均衡）を選択すると、パフォーマンスと消費電力とでバランスをとった設定となります。

DDR Performance = Power balanced（消費電力との均衡）						
プロセッサタイプ	RDIMM		LRDIMM		3DS RDIMM	
	1DPC	2DPC	1DPC	2DPC	1DPC	2DPC
DDR4-3200	2400	2400	2400	2400	2400	2400
DDR4-2933	2400	2400	2400	2400	2400	2400
DDR4-2667	2400	2400	2400	2400	2400	2400

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 の下のサブメニュー、Memory Configuration にあります。

Memory Configuration のメモリパラメーター

次の 10 個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：Independent／Mirroring／Address Range Mirroring
- Partial Cache Line Sparing (PCLS)：Disabled／Enabled
- ADDDC Sparing：Disabled／Enabled
- NUMA：Disabled／Enabled
- Virtual NUMA：Disabled／Enabled
- DDR Performance：Performance optimized／Energy optimized／Power balanced
- PPR Type：Hard PPR／Soft PPR／PPR Disabled
- Patrol Scrub：Disabled／Enabled
- SNC(Sub NUMA)：Disabled／Enabled
- UMA-Based Clustering：Hemisphere／Disabled

最初の 3 つのパラメーター、Memory Mode、Partial Cache Line Sparing（PCLS）、ADDDC Sparing（ADDDC：Adaptive Double Device Data Correction）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）機能の一部です。

Memory Mode は、メモリのデータを複製するか（ミラーリング）を指定します。Mirroring を指定するとミラーリングが有効となりますが、メモリ容量は半分になります。Address Range Mirroring は、システムメモリの一部をミラーリングします。これには、オペレーティングシステムのサポートが必要です。

Partial Cache Line Sparing は、第 3 世代 Xeon スケーラブル・プロセッサ（Ice Lake）で新しく提供される機能です。UEFI ファームウェアが永続的な 1 ビットエラーを検出すると、プロセッサのメモリコントローラに用意された予備領域を使用して 1 ニブル（4 ビットサイズ）分のデータを置きかえます。64 バイトのキャッシュライン単位で 1 ビットエラーまで訂正でき、メモリチャネル当たり 16 箇所の訂正が可能です。

ADDDC Sparing は、メモリエラーが頻繁に発生する場合に DIMM ランクまたはバンクのレベルで予備領域を有効化すること（スペアリング）で、耐故障能力を向上させます。2 つの DRAM デバイスのエラーに対してエラー訂正を行うことができます。

Memory Mode、ADDDC Sparing が利用できる構成については制限があります。これらについては、システム構成図を参照してください。

これらの機能が要求される場合、工場での出荷時には適切なデフォルト設定が行われます。それ以外の場合、これらのパラメーターは Independent（通常の冗長性なし）、および、 Disabled（無効化）に設定されます。これらの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。

4 番目のパラメーター NUMA は、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。

5 番目のパラメーター Virtual NUMA は、64 個を超える論理 CPU を持つプロセッサを Windows で使用する場合に使用します。Windows が論理 CPU を管理するために用いるプロセッサ・グループは、64 論理 CPU が上限のため、それを超える論理 CPU は別のプロセッサ・グループとして管理されます。その結果、プロセッサ・グループの大きさが不均一となることで、性能

面で不利となります。Virtual NUMA を有効にすることで、プロセッサは 2 つの同サイズの仮想的な NUMA ノードに分割して使用されます。後述の SNC と似ていますが、SNC の持つ性能向上の効果はありません。

6 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、直前のセクションで説明しました。

7 番目の PPR Type は、DDR4 の機能である Post Package Repair（PPR、ポストパッケージリペアー）を扱います。PPR は、システム起動時に故障したメモリセルを DRAM チップ内の予備領域に置き換えます。Soft PPR を設定すると、この置き換えはシステムの電源オフやリセットで失われます。Hard PPR を設定すると、置き換えは恒久的に保持されます。PPR Disabled の場合、置き換えは行われません。

8 番目のパラメーターPatrol Scrub パラメーターは、Enabled に設定すると、メインメモリに対し修正可能なエラーの検索が定期的に実行され、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、この機能に影響をうける可能性があります。ただし、パフォーマンスに及ぶ影響を実証するのは難しい場合があります。

最後の 2 つのパラメーターは、プロセッサ内のクラスタリングに関する設定です。

SNC(Sub NUMA) は、プロセッサ内でコア、L3 キャッシュ、メモリコントローラをクラスタに分割するためのパラメーターです。Enabled に設定すると、これらのリソースがプロセッサ内の 2 つのクラスタのどちらか一つにくくりつけられます。このクラスタは、オペレーティングシステムからは一つの NUMA ドメインとして扱われます。SNC はデフォルトでは Disabled です。この場合、プロセッサは UMA（Uniform Memory Access：均等型メモリアクセス）である 1 つのクラスタとして扱われます。

SNC により、NUMA ノード内のコアから L3 キャッシュやメモリへのアクセスは、そのレイテンシが改善します。ローカルメモリレイテンシを最小化、ローカルメモリ帯域を最大化することができるため、NUMA 最適化されたアプリケーションにおいて特に推奨されます。

新しく追加された UMA-Based Clustering パラメーターは、UMA 構成でのキャッシュコヒーレンシーの動作を変えます。これを Enabled にすると、L3 キャッシュとメモリコントローラは、お互いの近さに基づいて 2 つの領域に分割されます。コアは分割されません。このモードは Hemisphere と呼ばれ、デフォルトの設定となっています。このモードでは、L3 キャッシュとメモリ間の距離が短くなり、レイテンシが改善します。

SNC や Hemisphere を利用できる構成については制限があります。DIMM スロットとメモリコントローラのセクションに記載の図にあるチャネル A、B、C、D の DIMM 配置とチャネル E、F、G、H の DIMM 配置は対称でなければなりません。SNC ではさらに、チャネル A、B の DIMM 配置とチャネル C、D の DIMM 配置も対称でなければなりません。チャネル E、F とチャネル G、H についても同様です。これら以外の制限については、システム構成図を参照してください。

パフォーマンスを考慮したメモリ構成

メモリパフォーマンスにはメモリ周波数と使用するメモリチャネル数が大きく影響します。メモリ周波数は、搭載するプロセッサの種類に依存するため、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。また、Xeon スケーラブル・プロセッサはプロセッサあたり全部で 8 本のメモリチャネルがあり、高いメモリパフォーマンスを実現するためには、可能な限り多くのメモリチャネルに DIMM を配置する必要があります。

さらに、メモリパフォーマンスに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 機能の無効化などの機能です。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。

パフォーマンスモード構成

常に注意すべき 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 256 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 8 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 8 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、これらのメモリシステム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 8 WAY インターリーブが実現します。本書では、これをパフォーマンスモード構成と呼んでいます。

Xeon スケーラブル・プロセッサ（Ice Lake）ファミリー搭載 PRIMERGY サーバのパフォーマンスモード構成						
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	最大メモリ周波数 [MHz]	注
64 GB	128 GB	DDR4-3200 R	8		3200	
128 GB	256 GB	DDR4-3200 R	16		3200	（++）
192 GB	384 GB	DDR4-3200 R	16	8	3200	混在構成（-）
256 GB	512 GB	DDR4-3200 R	16	16	3200	
256 GB	512 GB	DDR4-3200 R	32		3200	
384 GB	768 GB	DDR4-3200 R	32	16	3200	混在構成（-）
512 GB	1024 GB	DDR4-3200 R	32	32	3200	（++）
512 GB	1024 GB	DDR4-3200 R	64		3200	
768 GB	1536 GB	DDR4-3200 R	64	32	3200	混在構成（-）
1024 GB	2048 GB	DDR4-3200 R DDR4-3200 LR	64	64	3200	（++）
2048 GB	4096 GB	DDR4-3200 LR DDR4-3200 3DS R	128	128	3200	
4096 GB	8192 GB	DDR4-3200 3DS R	256	256	3200	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。ど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。次の列は、使用した DIMM タイプです。RDIMM、LRDIMM、 または 3DS RDIMM テクノロジーが決定要因です。その次の列で 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 8 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 1 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 8 枚（64 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 8 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、LRDIMM、3DS RDIMM 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 2 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 8 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 スケーラブル・プロセッサファミリー搭載 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依存します。

インディペンデントモード構成

これには、パフォーマンスモードには含まれない構成がすべて含まれます。以下のルールがありますが、詳細については、各機種のシステム構成図を参照ください。

- RDIMM、LRDIMM、3DS RDIMM 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 2 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。
- 1 プロセッサ に搭載できる DIMM 枚数は 1、2、4、6、8、12、または 16 枚に制限されます。

プロセッサあたりの DIMM 枚数 が 8 の倍数にならない構成、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注意する必要があります。省電力や、必要なメモリ容量が少ないといった理由のためにこのような構成にすることがあります。DIMM 枚数の最小化で費用削減が実現できる場合もあります。この後で紹介する、メモリチャネルへのインターリーブ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。

- プロセッサあたりの DIMM 枚数を 4、6 または 8 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。1、2DIMM 構成での動作はお勧めしません。

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、UPI 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後述しています。統計上、メモリアクセスの 2 回に 1 回は、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、ローカルメモリとリモートメモリが 50 %／50 %の割合で実行される場合の 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。

なお、1 つ目のプロセッサに 16 枚の DIMM、2 つ目のプロセッサに 8 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。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、メモリ構成に関係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 の場合と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。

Xeon スケーラブル・プロセッサでは、プロセッサタイプによりサポートする最大メモリ周波数が異なります。そのため、一部の例外を除いて、定量的テストは、プロセッサがサポートする最大メモリ周波数を基準にプロセッサを選択し実施しました。

測定は、Linux オペレーティングシステムが動作する 2 つのプロセッサを搭載した PRIMERGY RX2540 M6 で行いました。次の表は、定量化テストの構成、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 RX2540 M6
プロセッサ	Xeon Platinum 8360Y (36 コア、2.4GHz、DDR4-3200) × 2 Xeon Gold 5320 (26 コア、2.2GHz、DDR4-2933) × 2 Xeon Silver 4316 (20 コア、2.3GHz、DDR4-2667) × 2
メモリタイプ	8GB (1x8GB) 1Rx8 DDR4-3200 R ECC 16GB (1x16GB) 1Rx4 DDR4-3200 R ECC 16GB (1x16GB) 2Rx8 DDR4-3200 R ECC 32GB (1x32GB) 2Rx4 DDR4-3200 R ECC 64GB (1x64GB) 2Rx4 DDR4-3200 R ECC 64GB (1x64GB) 4Rx4 DDR4-3200 LR ECC 128GB (1x128GB) 4Rx4 DDR4-3200 LR ECC 256GB (1x256GB) 8Rx4 DDR4-3200 3DS R ECC
ディスクサブシステム	SAS 12G SSD 400GB 1 台（SAS RAID コントローラを介して）
ソフトウェア	
BIOS	R1.6.0
オペレーティングシステム	Red Hat Enterprise Linux 8.2

以下に説明するテストセットには通常 64 GB 2Rx4 RDIMM、または 32 GB 2Rx4 RDIMM が使用されました。表に示したその他のすべての 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。

以降の表は、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 SPECrate2017 Integer のベンチマークの絶対測定値は（通常、表の 1.0 の値に相当）、Xeon スケーラブル・プロセッサ搭載 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。

測定ツール

測定は、STREAM および SPECrate2017 Integer 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）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、コンパイラー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全般的な予測を立てることは非常に困難です。

SPECrate2017 Integer ベンチマーク

SPECrate2017 Integer 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 SPEC CPU2017 の一部です。SPEC CPU2017 は、システムのプロセッサ、メモリおよびコンパイラー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。

SPEC CPU2017 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、10 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、10 または 13 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、コンパイラー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、コンパイラー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 SPEC CPU2017 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 SPECrate2017 Integer でスループットを測定しました。

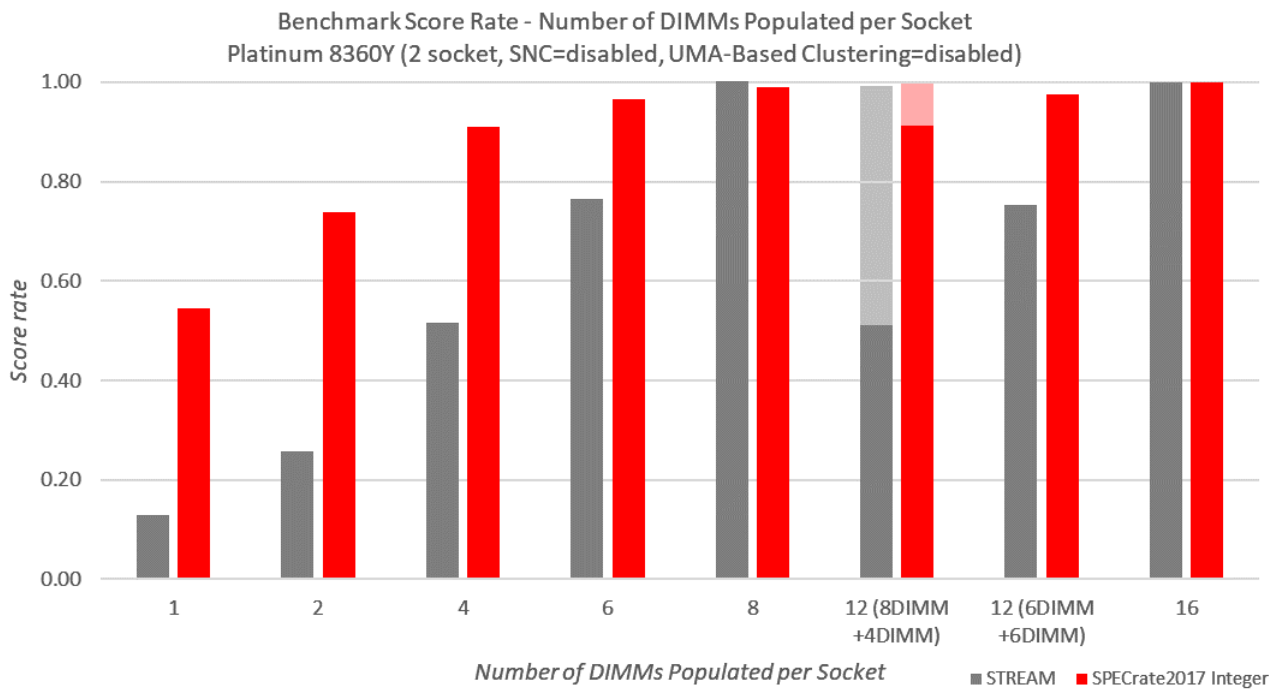
本来のルールに準拠した測定では 3 回の実行が必要であり、各ベンチマークに対して中央値の結果が評価されます。しかし、ここで説明している技術調査では、このルールに準拠していません。効率化のために、測定は 1 回にしています。

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 8 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、局所性原理に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。

次の図は、同種類の DIMM をプロセッサあたり 8 枚単位で搭載せず理想的な 8 WAY インターリーブを行わない場合に、DIMM 枚数が 16 の場合を 1 としたパフォーマンスの比率を示しています。Ice Lake 搭載 PRIMERGY サーバでは、1 プロセッサに搭載できる DIMM の枚数は 1、2、4、6、8、12、16 枚のみに制限されます。ここでは、SNC=Disabled かつ UMA-Based Clustering=Disabled の設定での結果だけを記載しています。SNC=Enabled の場合、SNC=Disabled かつ UMA-Based Clustering=hemisphere の場合の結果は、設定可能な DIMM の構成が限られること、および、性能比の傾向はほぼ同じであったことから、省略しています。

特にメモリのスループットを測定する Stream の指標において、顕著な低下が見られます。DIMM 枚数が 8 枚以下の場合には、DIMM 枚数の増加に応じてパフォーマンスが向上しています。



1 プロセッサに搭載する DIMM の枚数が 12 枚の場合は注意が必要です。1 番目のバンクに 8 枚、2 番目のバンクに 4 枚の DIMM を搭載する場合、前者の DIMM が構成する物理アドレス領域は 8 チャネルインターリーブ構成となり、後者の DIMM が構成する物理アドレス領域は 4 チャネルインターリーブとなります。そのため、アプリケーションがどの領域の上で動作するかで性能が変わります。

上記グラフで 12 DIMM（8DIMM＋4DIMM）での結果は、薄い色と濃い色の 2 つの棒グラフで示されています。薄い色の棒グラフは、8 チャネルインターリーブの領域を使用して測定した結果、濃い色の棒グラフは、4 チャネルインターリーブの領域を使用しての測定結果となります。一方、12 DIMM （6DIMM＋6DIMM）、すなわち、1 番目のバンク、2 番目のバンクとも 6 枚の DIMM を搭載する場合には、6 チャネルインターリーブの領域のみとなり、性能の違いは生じません。このように、搭載する DIMM の構成によっては、インターリーブによる性能向上が十分に得られない場合があります。

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 Platinum 8360Y です。使用した DIMM タイプは 64 GB 2Rx4 RDIMM です。

SPECrate2017 Integer 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 SPECrate2017 Integer 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。

パフォーマンスの低下が穏やかな 4 WAY、6 WAY インターリーブを選ぶ場合は、それなりの理由があるかもしれません。つまり、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 枚数が最小限に抑えられるような場合です。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パフォーマンスのバランスが取れていません。

また、12DIMM 構成での測定結果が示すように、不均等な DIMM の構成は、安定して性能を最大限に引き出す観点からは推奨できません。

メモリ周波数

Xeon スケーラブル・プロセッサ搭載 PRIMERGY サーバでは、DPC によってメモリ周波数は変化しません。プロセッサの種類と BIOS パラメーターがメモリ周波数に影響を与えます。節電設定（BIOS パラメーター DDR Performance で管理）は、実効メモリ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 1867 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	プロセッサ種類	最大メモリ周波数	1867 MHz	2400 MHz	2667 MHz	2933 MHz	3200 MHz
STREAM	Platinum 8360Y	3200 MHz	1.00	1.24			1.60
	Gold 5320	2933 MHz	1.00	1.24		1.47	
	Silver 4316	2667 MHz	1.00	1.24	1.33		
SPECrate2017 Integer	Platinum 8360Y	3200 MHz	1.00	1.04			1.07
	Gold 5320	2933 MHz	1.00	1.03		1.04	
	Silver 4316	2667 MHz	1.00	1.02	1.02		

ベンチマーク	プロセッサ種類	最大メモリ周波数	1867 MHz	2400 MHz	2667 MHz	2933 MHz	3200 MHz
STREAM	Platinum 8360Y	3200 MHz	0.62	0.77			1.00
	Gold 5320	2933 MHz	0.68	0.84		1.00	
	Silver 4316	2667 MHz	0.75	0.93	1.00		
SPECrate2017 Integer	Platinum 8360Y	3200 MHz	0.94	0.98			1.00
	Gold 5320	2933 MHz	0.96	0.99		1.00	
	Silver 4316	2667 MHz	0.98	1.00	1.00		

このテストで使用したプロセッサモデルは、Xeon Platinum 8360Y（DDR4-3200）、Xeon Gold 5320（DDR4-2933）、Xeon Silver 4316（DDR4-2667）です。使用した DIMM タイプは 64 GB 2Rx4 RDIMM（Platinum 8360Y 使用時）、32 GB 2Rx4 RDIMM（Gold 5320、Silver 4316 使用時）です。2DPC 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 1867 MHz になります。また「DDR Performance = Power balanced（消費電力との均衡）」と設定すると、周波数は 2400 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、電圧が常に 1.2 V である DDR4 モジュールの場合、得られる節電効果はわずかです。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。

DIMM タイプの影響

Ice Lake 搭載 PRIMERGY サーバの一般公開時には、最大 9 タイプの DIMM がサポートされています。ただし、特定のサーバでの利用可能な構成については、それぞれのシステム構成図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。

- 測定は、Xeon Platinum 8360Y を使用して実施しました。
- これらの測定では、すべてのメモリチャンネルが同じ構成、すなわちパフォーマンスモード構成での比較が行われました。取り付けられた DIMM の数は、1DPC の測定では 16、2DPC の測定では 32 でした。
- すべての測定は、メモリ周波数 3200 MHz で一様に実施されました。

この表では、現時点で最善のパフォーマンスを提供するであろう 64 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	SPECrate2017 Integer
8GB (1x8GB) 1Rx8 DDR4-3200 R ECC	1DPC	0.85	0.94
	2DPC	0.85	0.97
16GB (1x16GB) 2Rx8 DDR4-3200 R ECC	1DPC	1.01	0.98
	2DPC	0.99	1.00
16GB (1x16GB) 1Rx4 DDR4-3200 R ECC	1DPC	0.91	0.96
	2DPC	0.94	0.98
32GB (1x32GB) 2Rx4 DDR4-3200 R ECC	1DPC	1.02	0.99
	2DPC	0.98	1.00
64GB (1x64GB) 2Rx4 DDR4-3200 R ECC	1DPC	1.01	0.99
	2DPC	**1.00**	**1.00**
64GB (1x64GB) 4Rx4 DDR4-3200 LR ECC	1DPC	0.97	1.00
	2DPC	0.87	0.99
128GB (1x128GB) 4Rx4 DDR4-3200 LR ECC	1DPC	0.97	1.00
	2DPC	0.86	0.99
256GB (1x256GB) 8Rx4 DDR4-3200 3DS R ECC	1DPC	0.91	0.98
	2DPC	0.84*	0.97*

（*：推定値）

ここで示されているパフォーマンスの違いは、ランクインターリーブ数が異なることが主な原因です。ランクインターリーブ数はメモリチャネルあたりのランク数に等しく、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

メモリチャ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、チャネルでのインターリーブより大きくなります。チャ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 つの次元で配置されます。アクセスの際は行

（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。商用アプリケーションでのパフォーマンスを考えた場合、4 WAY インターリーブに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。一方で、8 GB 1Rx8 RDIMM や 16 GB 1Rx4 RDIMM の 1DPC 構成、すなわち、1 WAY ランクインターリーブでは、パフォーマンス低下が目立ちます。

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、悪い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に相当します。これにより、前述した 4Rx4 LRDIMM において、2DPC 構成の結果が対応する 1DPC 構成の結果よりも悪くなるケースとの関係が説明できます。またリフレッシュの影響は大容量の DIMM ほど目立つようになります。

プロセッサ内クラスタリングの最適化

これまでの Xeon スケーラブル・プロセッサでは、SNC(Sub NUMA) を有効にした場合、無効にした場合（ここでは UMA と呼びます） の 2 つの選択肢がありましたが、Ice Lake プロセッサでは、これに加えて、Hemisphere と呼ばれるモードを選択できます。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。

測定は、64 GB 2Rx4 RDIMM（Platinum 8360Y 使用時）、32 GB 2Rx4 RDIMM（Gold 5320、Silver 4316 使用時）を使用した 2DPC 構成で行われました。

表には、約 1 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 SPECrate2017 Integer 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。

ベンチマーク	プロセッサ種類	SNC	Hemisphere (デフォルト)	UMA
STREAM	Platinum 8360Y	1.01	1.00	1.00
	Gold 5320	1.01	1.00	0.99
	Silver 4316	1.01	1.00	1.00
SPECrate2017 Integer	Platinum 8360Y	1.01	1.00	1.00
	Gold 5320	1.01	1.00	1.00
	Silver 4316	1.00	1.00	1.00

これらのクラスタリングオプションでの BIOS 設定値は以下の通りです。

クラスタリングオプション	SNC(Sub NUMA)	UMA-Based Clustering
SNC	Enabled	- [4]
Hemisphere	Disabled	Hemisphere
UMA	Disabled	Disabled

[4] SNC(Sub NUMA) を Enabled に設定すると、この項目は設定できなくなります。

リモートメモリへのアクセス

前述の STREAM および SPECrate2017 Integer 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 UPI 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され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、最大メモリ周波数で動作し 64 GB 2Rx4 RDIMM、または 32 GB 2Rx4 RDIMM での 8 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。統計的にメモリアクセスの半数がリモート 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 UPI 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。特に、UPI リンクが 2 本で UPI の動作周波数が低い Silver 4316 プロセッサでは、STREAM のような帯域幅が重視されるベンチマークで顕著に性能が低下しています。

ベンチマーク	プロセッサ種類	UPI 周波数	NUMA = Enabled	NUMA = Disabled
STREAM	Platinum 8360Y	11.2 GT/s	1.00	0.60
	Gold 5320	11.2 GT/s	1.00	0.65
	Silver 4316	10.4 GT/s	1.00	0.45
SPECrate2017 Integer	Platinum 8360Y	11.2 GT/s	1.00	0.92
	Gold 5320	11.2 GT/s	1.00	0.93
	Silver 4316	10.4 GT/s	1.00	0.92

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。

NUMA = Disabled に関する測定は、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に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。

冗長性、信頼性を考慮した際のメモリパフォーマンス

ここでは、冗長性、信頼性のオプションの性能への影響を評価します。

ミラーリングではプロセッサの 1 つのメモリコントローラ内で 2 つのメモリチャネルの間でミラーが構成されます。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。

ADDDC スペアリングの場合、DIMM の予備領域を使用して 故障 DRAM セルを置きかえるため、容量の減少はありません。

次の表では、それぞれのケースで 64 GB 2Rx4 RDIMM、または 32 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、および、ADDDC Sparing 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性能低下は、デフォルト時の性能の 50 % 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。ADDDC スペアリングでは、機能を有効にすることによる若干の性能低下がみられます。

ベンチマーク	プロセッサ種類	デフォルト	ミラーリング	ADDDC スペアリング
STREAM	Platinum 8360Y	1.00	0.71	1.00
	Gold 5320	1.00	0.69	0.97
	Silver 4316	1.00	0.72	0.98
SPECrate2017 Integer	Platinum 8360Y	1.00	0.97	0.99
	Gold 5320	1.00	0.98	0.98
	Silver 4316	1.00	0.99	0.99

各冗長性オプションでの BIOS 設定値は以下の通りです。

冗長性オプション	Memory Mode	ADDDC Sparing
デフォルト	Independent	Disabled
ミラーリング	Mirroring	Disabled
ADDDC スペアリング	Independent	Enabled

関連資料

PRIMERGY サーバ

https://www.fujitsu.com/jp/products/computing/servers/primergy/

メモリパフォーマンス

このホワイトペーパー

https://docs.ts.fujitsu.com/dl.aspx?id=1930d389-7521-4c85-bcf9-86a71a14a7c3

https://docs.ts.fujitsu.com/dl.aspx?id=74e595de-344f-47a6-8995-fe340733dfbb

過去のホワイトペーパー

Xeon スケーラブル・プロセッサ（Skylake-SP）搭載システムのメモリパフォーマンス

https://docs.ts.fujitsu.com/dl.aspx?id=0a62cad1-469d-4b70-a028-ba6f8b74c7f1

Xeon スケーラブル・プロセッサ（Cascade Lake-SP）搭載システムのメモリパフォーマンス

https://docs.ts.fujitsu.com/dl.aspx?id=ade521ff-45c7-408d-9b36-a88b248497ca

ベンチマーク

STREAM

https://www.cs.virginia.edu/stream/

SPEC CPU2017

https://www.spec.org/osg/cpu2017/

ベンチマークの概要 SPECcpu2017

https://docs.ts.fujitsu.com/dl.aspx?id=0f641c7e-bb5e-45e4-854f-cdd31faf5343

BIOS 設定

Xeon スケーラブル・プロセッサ 搭載システムのための BIOS 最適化

https://docs.ts.fujitsu.com/dl.aspx?id=696984c2-7a49-4b64-ba34-77888c8a68d6

PRIMERGY のパフォーマンス

https://jp.fujitsu.com/platform/server/primergy/performance/

文書変更履歴

版数	日付	説明
1.1	2023-10-03	新 Visual Identity フォーマットに変更 マイナーな修正
1.0	2021-11-05	初版

お問い合わせ先

富士通株式会社

Web サイト: https://global.fujitsu/ja-jp/

PRIMERGY のパフォーマンスとベンチマーク

mailto:fj-benchmark@dl.jp.fujitsu.com

