
第 12 世代 Dell PowerEdge

サーバのメモリ

本書は、第 12 世代の Dell PowerEdge サーバが提供するメモリオプションと、その選択に応じた既存環境への影響について解説するテクニカル概要書です。

ESG メモリエンジニアリング
ESG アドバンスドエンジニアリング
ESG マーケティング



**本書は、情報提供のみを目的に執筆されており、誤字脱字や技術上の誤りには責任を負いません。
本書の内容は執筆時現在のものであり、明示的、暗示的を問わず、いかなる内容も保証いたしません。**

© 2012 Dell Inc. ©2012 デル株式会社 All rights reserved. (著作権所有)

デルとその関連会社は、誤字、脱字、誤植や、図、写真の誤りや不備について一切の責任を負いません。

Dell、DELL のロゴマーク、PowerEdge は、米国 Dell Inc. の商標です。Intel、インテル、Xeon は、アメリカ合衆国およびその他の国におけるインテルコーポレーションおよび子会社の登録商標または商標です。本書では、マークや名前を届け出た実在のもの、もしくは、その製品のいずれかを参照するため、その他の商標、商号を使用している可能性があります。デルは、その他のマークや名称について、商標上の利権に対する要求に一切に応じません。

2012 年 2 月 | Rev 1.0

目次

はじめに.....	4
メモリサブシステム：機能と仕組み.....	4
標準化.....	5
信頼性.....	5
電力.....	6
DIMM 編成.....	8
チャンネル、スロット、ランク：メモリサブシステムの構築方法.....	9
チャンネル、スロット、ランク：メモリ速度への影響.....	10
第 12 世代の PowerEdge サーバ：メモリの特長と強み.....	11
新機能と強化機能.....	11
メモリ編成：インテル® Xeon® E5 プロセッサ.....	11
メモリチャンネルの使用とペア構成.....	12
BIOS とシステム構成.....	13
RAS (信頼性、可用性、保守性) 機能.....	13
DIMM.....	14
メモリサブシステム.....	14
新規、または、機能強化されたテクノロジー.....	14
BIOS 対応.....	15
サポート対象のメモリ.....	15
まとめ.....	16

表

表 1. 各状態における DIMM の消費電力 (相対比較).....	7
表 2. チャンネルあたりの DIMM 数.....	9
表 3. DIMM の種類と負荷に応じて変わる PowerEdge のメモリスピード.....	10
表 4. 世代間の機能比較.....	11
表 5. RAS 機能.....	13
表 6. PowerEdge R720 に 256GB メモリを設置するときのサンプル構成.....	16
表 7. メモリー覧.....	16

図

図 1. PowerEdge サーバの世代別メモリ比較.....	12
----------------------------------	----

はじめに

第 12 世代の Dell™ PowerEdge™ サーバは、メモリサブシステムの能力が旧世代サーバに比べて飛躍的に向上しています。このテクニカル概要書は、最新世代の PowerEdge サーバに搭載されたメモリサブシステムが、旧世代からどのような進化を遂げたのか説明します。

メモリサブシステム：機能と仕組み

注：このセクションでは、DDR3 メモリの基礎知識をご紹介します。既に DDR3 の概要についてご存知の方は、割愛していただいて構いません。

どんなコンピューティング環境でも、プロセッサは、データの集合を操作し、演算/判断を行い、そこから結果を出力するという一連のタスクを実行します。これらの要素 (タスクの命令群と関連データ) を長期保存するときはディスクストレージサブシステムを使用しますが、プロセッサの処理中は、シリコンテクノロジーに基づく高速なランダムアクセスメモリ (RAM) が使用され、これが高速・高信頼性の短期保存先となります。この保存 (ストレージ) に使われる基本コンポーネントが、DRAM (Dynamic Random-Access Memory) です。

それぞれの DRAM には、保存できるデータビット数に上限があります。PowerEdge サーバで典型的に使用される DRAM は、1 つあたり 20~40 億ビット、つまり、2~4 ギガビット (Gb) を格納することができます。これは膨大なビット数に感じるかもしれませんが、最近では、512 ギガバイト (GB) 以上のメモリに対応できるサーバも登場しています。1 バイト = 8 ビットなので、512GB のメモリを提供するには、2 ギガビットの DRAM デバイスが 2,048 個以上必要です。

これほど多くの DRAM を 1 台のサーバ内に取り入れるには、パッケージングにも考慮しなければなりません。そこで、最大 72 個の DRAM デバイスが取り付けられるモジュール (基板) を利用し、この基板上に DRAM パッケージをマウントしています。これらのモジュールは、2 つ (デュアル) のピン配列が、シングルワイドコネクタ上に取り付けられているため、そのコネクタピンの形状から DIMM (Dual In-line Memory Modules、デュアルインラインメモリモジュール) と呼ばれています。一連の DRAM は、モジュールの片面、または、両面にマウントされます。一枚あたり最大 72 個の DRAM を搭載できる DIMM であれば、わずか 32 枚で 512GB のメモリが提供できます。

一枚のメモリモジュール上に多数の DRAM が搭載されるため、DIMM は同時に複数の DRAM にアクセスすることになり、実際、1 回のメモリサイクルでアクセスされるデータは、64 ビット単位となります。これらの 64 ビット (= 8 バイト) は、DRAM の密度に応じて 8 個または 16 個の DRAM から提供されます。

前述のとおり、今日のサーバメモリは、2、または、4 ギガビット密度の DRAM を使用しています。一枚あたりの容量が 2~32 GB の DIMM を作成するには、目的の DIMM 容量を達成するのに必要な幅 (Width) と深さ (Depth) を持った、一定密度の DRAM を適切に配置する必要があります。

たとえば、2GB の DIMM を製造すると仮定しましょう。この場合、2 ギガビット DRAM なら、8 個もあれば十分です。DIMM のワード幅は 64 ビット (8 バイト) であるため、1 バイトあたり 1 個の DRAM を使用とした場合、幅が 8 ビットで、深さが 256 メガワードの DRAM が必要になります ($8 \times 256M = 2Gb$)。このようなデバイスを 8 個搭載すれば、目的の 2GB DIMM が完成します。それでは次に、4 ギガビット DRAM を使ってみたらどうなるでしょうか？ この場合、たとえば幅が 16 ビットで、深さが 256 メガワードの DRAM を使うことが考えられます。このような 16 ビット幅の DRAM を使うのであれば、2GB の容量を達成するのに 4 個しか必要ありません。ただし、詳細は後ほど説明しますが、PowerEdge サーバの DIMM に搭載できる DRAM は、8 ビット幅が最大となります。

今日のサーバでは、メモリのサイズや容量だけでなく、アクセス速度も重視されます。デルの次世代サーバは、最大 1600 MT/s のスピードで動作する、極めて高速なメモリも提供しています (MT/s = MegaTransfers per Second、1 秒あたりのメガ転送量、または、1 秒あたりのメガビット数。Mbps で表わすことも)。

標準化

ここまで高速なメモリアクセスを実現するには、綿密に定義されたアクセス手法が必要です。さらに、これらのシステムで多数の DIMM が使用される現実を考えると、これらのメモリモジュール (と、その上に搭載される DRAM) には、業界標準規格が絶対に欠かせません。そこで、現世代のメモリは主に、JEDEC¹ DDR3² と呼ばれる一連の標準仕様に従っています。これらの仕様には、第 12 世代の Dell PowerEdge サーバで使用されているメモリの電気インタフェース、パッケージング、操作プロトコルなどが定義されています。

標準化は、複数のサプライヤによるメモリモジュールの安定供給と相互運用性を促し、また、標準化によってメモリが量産できるようになるため、メモリの低価格化にも貢献します。第 12 世代の PowerEdge サーバは、DDR3 と DDR3L の両標準に基づいています。両者の詳細は後述しますが、ここで簡単に説明しておく、両標準とも 240 ピンの DIMM を定義しており、第 12 世代のメモリサブシステムもこれに基づいています。この DIMM のデータバス幅は 64 データビットで、これに 8 ビットの ECC (Error-Correcting Code、誤り訂正符号) が加わります。

信頼性

第 12 世代の Dell PowerEdge サーバと、これらのサーバ上で実行される基幹系アプリケーションは、大量のメモリを使用するため、メモリの信頼性は最優先課題となります。DRAM の半導体製造プロセステクノロジーは、30 ナノメートル台にまで細密化されており、現行サーバの製品ライフ中には、さらに小さな 20 ナノメートル台への進化も十分に見込めますが、その一方で、無作為に発生するビットエラーの影響も受けやすくなります。これらのビットエラーは、アルファ粒子やスプリアス電気といった、多くの自然現象が原因です。

第 12 世代の PowerEdge サーバに搭載されるメモリデータビット数は膨大になることから、その分、エラー発生の可能性も測定し得るほど高くなるため、これらを検出し、可能であれば、修正もできるような仕組みが必要です。8 ビットの ECC データを追加しているのも、このためです。

64 ビットのデータごとに 8 ビットの ECC を付加すると、どの 1 ビットにエラーが起きても検出でき、さらに、修正することもできます。これらの冗長ビットは、エラーをリアルタイムに修正していくので、メモリの信頼性と可用性が飛躍的に高まります。

さらに、サーバは、これらの冗長ビットを通して、複数のビットエラーを検出することもできます。複数ビットエラーの場合、修正まではできませんが、少なくとも不適切なデータの使用は避けられます。これらの 8 ビットが追加されることで、DIMM のワード幅は、72 ビット = 9 バイトになります。したがって、この DIMM に使用される DRAM の個数は、72 の約数でなければなりません。ここで、先に例示した 16 ビット幅の 4 ギガビット DRAM を思い出してください。これを使用して 2GB の DIMM を製造しようとする、72 ビットのデータをカバーするのに 5 個の DRAM が必要となり、この条件に当てはまりません (5 は 72 の約数ではない)。

16 ビット幅の DRAM を 5 個使うと、 $16 \times 5 = 80$ ビットとなり、8 ビットの無駄が出てしまいます。このような理由から、第 12 世代サーバのメモリは、8 ビット幅または 4 ビットワード幅の DRAM 構成となっており、これは、DDR3 標準で定義された仕様にも適合します。

¹ JEDEC : Joint Electron Devices Engineering Council (電子機器技術評議会) の略

² DDR3 : Double Data Rate Type 3 (ダブルデータレートタイプ 3) の略。SDRAM (synchronous dynamic random access memory、同期 DRAM) の一種

電力

メモリモジュールも、実質、他のあらゆる電子部品と同様、電力を消費します。メモリが消費する電力量は一般に、メモリが採用している半導体テクノロジー、メモリの動作電圧、メモリ速度、メモリ利用率、モジュールに適用されている電力管理機能などによって決まります。

モジュールの消費電力について詳細を知るには、まず、DRAM (動的ランダムアクセスメモリ) セルの基本的な仕組みを知っておく必要があります。動的メモリセルの大きなメリットとは、トランジスタとコンデンサを組み合わせた高密度ストレージによって、データをビットごとに格納できることです。コンデンサは実際ストレージの役割を果たしており、電子を選択的に格納することで、メモリセルに保存された 0 か 1 のデータを表します。一方、トランジスタは、コンデンサの読み書き手段を提供します。セルに書き込むときは、トランジスタがオンになり、電子の電荷を送り込んでコンデンサを充電するか (1 を書き込むとき)、または、コンデンサを放電させます (0 を書き込むとき)。

このメカニズムは、高密度をサポートするものの、メモリ設計に 1 つの要件を課します。データの保存に使われるコンデンサは完全ではなく、時間が経つと電子が流れ出てしまいます。このため、各セルに保存されたデータを維持するには、電子が流出し切る前にコンデンサを再充電しなければなりません。これを「リフレッシュサイクル」と呼び、周期的な充電を設計に組み込む必要があります。リフレッシュは通常、64 ミリ秒間隔で実行されており、人間の常識から考えるとかなり短時間ですが、標準のメモリアクセス速度から見れば長時間です (1333 MT/s のメモリを 80% のバンド幅利用率で使用すると、このサイクル中に、約 550MB ものデータが読み取れます)。

第 12 世代の PowerEdge サーバ用メモリには動作状態が数レベルあり、いずれも、ある程度の電力を消費します。たとえば、バースト書き込みやバースト読み取りといった状態では電力が最も消費され、メモリアイドルなどの状態では、電力消費が少なくなります。メモリサブシステム全体の総電力とは、様々な状態遷移を経ながら、一定時間に消費した電力の合計となります。後ほど詳しく説明しますが、メモリの動的な性質とサブシステムレベルでの管理機能が、節電に大きく影響します。

本セクションの始めにも軽く触れましたが、メモリの消費電力を左右する要因は、次のとおりです。

- メモリに使われている半導体テクノロジー：主に、トランジスタの物理的な大きさ
 - 第 12 世代の Dell PowerEdge サーバは、利用できる最新のメモリテクノロジーを使用しています。つまり、デルは、原則的に、全サプライヤの中から最も省電力なデバイスを提供しています。
- メモリの動作電圧
 - デルが提供しているメモリの大半は、1.35 ボルトか 1.5 ボルトの動作電圧をサポートする DDR3L 標準に基づきます。第 12 世代の Dell PowerEdge サーバは、可能な限り最低の電圧で動作しながら、お客様の性能ニーズを満たすことができます。
- メモリの動作速度
 - Dell PowerEdge サーバで提供される大半のメモリは、たとえ最小限の電圧でも 1333 MT/s で動作することができます。デルは、最高性能を求めめるお客様に対し、1600 MT/s 対応の DIMM も幅広く用意しています。
- メモリ利用率、または、転送用チャネルバンド幅の利用率
 - これは、サーバで実行しているアプリケーションによって大きく変わりますが、デルなら各種の DIMM が包括的に揃うため、電力と容量の兼ね合いを図りながら、最もニーズに合致したメモリが選択できます。
- メモリサブシステムの電力管理機能
 - 第 12 世代の Dell PowerEdge サーバは、全体的な電力を節約するため、業界で最も先進的とされる電源管理アルゴリズムを複数採用しています。

上記のうち最初の 3 つは、システムの電力消費に関わるもので、その省電力効果は、設計にどれだけ時間と労力をかけ、良いものを追求できるかにかかっています。よく考えられて設計されたメモリサブシステムは、最小電力を維持しながら最大速度で動作できますし、供給元から最高品質の部品を調達する力は、この能力を安定させることができます。メモリ利用率は、お客様が使用するアプリケーションに応じて変わるものの、デルはデバイスを包括的に取り揃えているため、お客様固有の環境に応じてメモリ利用率を最適化することができます。

最後の電力管理機能についてですが、もしここで任意のメモリサブシステムを調査したら、特に電力管理が最も重視される大規模システムでは、DIMM あたりのメモリ利用率が 50% 未満になることは想像に難くありません。これは単に、チャンネルあたり 2 つ以上の DIMM が配置されていても、一度にアクセスできるのは一枚の DIMM に限られるからです。このとき、残りの DIMM はアイドル状態にあるか、リフレッシュされます。したがって、アイドルまたはリフレッシュ時の DIMM をできる限り最小の電力状態に置くことができれば、大きな節電になります。

幸い、DDR3 標準が定めた動作状態を適切に実装すれば、アイドル時の大幅な電力削減が可能です。表 1 は、典型的な DIMM (8GB、1333MT/s) が消費する電力の割合を示したものです (メモリがフルスピードで動作し、バンド幅利用率が 80% のときの消費電力を基準値の 100% とし、これを基に相対比較したもの)。この値は、概要レベルのテスト結果であり、考え得るすべての電源状態を完璧に網羅した分析結果ではありません。

表 1. 各状態における DIMM の消費電力 (相対比較)

状態	消費電力の相対値
フルアクセス、バンド幅利用率 = 80%	100%
アクティブアイドル	63%
セルフリフレッシュ/S3	7%

表 1 からわかるとおり、メモリが単純にアイドル状態にあるときの消費電力は、稼働時に比べればかなり低いものの、セルフリフレッシュ/S3 と呼ばれる状態に比べると、9 倍近くも電力を消費しています。このセルフリフレッシュとは、どのような状態なのでしょう？

64 ミリ秒ごとにメモリセルをリフレッシュする、という要件を満たすには、このリフレッシュサイクルを実行する何らかの仕組みが必要です。旧世代のサーバでは、「アクティブアイドル」状態 (システムがメモリにコマンドを送るために必要な状態) のときに、システムのメモリコントローラがリフレッシュを実行していました。一方 DDR3 では、セルフリフレッシュ状態に置かれているメモリ自身が、内部のリフレッシュサイクルを実行できるので、大幅な節電になります。

残念ながら、セルフリフレッシュ状態に出入りする状態遷移を管理するには、精密なタイミング調整が必要なため、IT 業界はつい最近まで堅牢な実装方式を実用化できずにいました。この件はまさに、デルが秀逸な設計にこだわる適例の 1 つです。デルは、お客様に真の価値をもたらすため、第 12 世代の PowerEdge サーバ設計に多大な投資を注いできました。そしてデルは、業界で最も包括的なテスト&検証手法を開発し、高信頼性かつ安全なメモリ電力管理方式を実現するに至ったのです。

DIMM 編成

DRAM の編成と、一枚の DIMM 上に最大 72 個の DRAM を搭載する概念については、先に説明したとおりです。ここでは、メモリをさらに深く理解するため、システムレベルでどうメモリを構成し、なぜメモリの組み合わせに制約が生じるのか説明していきます。

前回例示した 2GB DIMM に話題を戻すと、この DIMM は、8 個の DRAM (256M x 8) に、1 つの ECC データビットを追加して、合計 9 個で編成できたことを思い出してください。それでは、同じ DRAM を使って 4GB の容量を持つ DIMM を製造するにはどうしたら良いのでしょうか？この場合、単純にもう一列の DRAM 群を追加し、DIMM に合計 18 個の DRAM を搭載することになります。しかし、今や、列数は 2 つになってしまいました。

2 列あるのに、どうしたら特定の列にアクセスできるようになるのでしょうか？

JEDEC DDR3 DIMM 仕様では、一連の選択信号を定義しており、これを使えば、DIMM 上でどちらの列にアクセスするのか判断できます。これらの信号は、S_n[0:3] と呼ばれ、既にお察しかもしれませんが、このような信号が 4 つあります。ということは、DIMM 上に DRAM 列が最大 4 列搭載できるということでしょうか？はい、確かに可能です。しかし、3 列以上になると、DIMM の種類が限定されてしまいます。

JEDEC 用語では、DRAM の列のことを「ランク」と呼びます。最初に例示した DIMM は「2GB DIMM 1Rx8」と表記され、これは、1 ランク、8 ビット幅の編成を示します。4GB モジュールを製造するとしたら、2Rx8 という編成が考えられます。さらに、同じ 2Gb DRAM を使用して 4Rx8 編成の 8GB モジュールを作ること也不可能ではありません。しかし、前述のとおり、この実装法には制限があり、さらに他の要因も複雑に絡み合ってくるため、これより良い実装法が別途見つかる可能性もあります。

モジュール上の DRAM 数が増えると、シグナルロード (信号の負荷) も比例して増えていきます。1 列 8 個の DRAM であれば、ロードはそれほど重くありません。各 DRAM は、それぞれ自分自身のデータ I/O ラインと連結されていますが、クロック、アドレス、コマンドラインは全 DRAM で共有します。たとえば、2 ランクメモリの場合、クロック、アドレス、コマンドラインは 16 ロードとなり、データ I/O ラインは、2 つ (各列から 1 つずつ) の DRAM から共有されます。

さらにランク数を増やしていくと、各信号のロードが倍増していき、いずれ、電氣的に処理不能な構成に陥り、信号の整合性が著しく低下します。DDR3 標準では、三種類の DIMM を定義することで、このようなロード状況に対応しています。

- **Unbuffered DIMM (バッファなし DIMM) :** UDIMM と呼ばれ、アドレス、制御、クロック、I/O ラインもデータ I/O ラインもバッファされません。
- **Registered DIMM (レジスタ付き DIMM) :** RDIMM と呼ばれ、アドレス、制御、クロックラインはバッファされますが、データ I/O ラインはバッファされません。
- **Load reduced DIMM (低負荷 DIMM) :** LRDIMM と呼ばれ、アドレス、制御、クロックラインとデータ I/O ラインの両方をバッファします。

UDIMM は、最初に挙げたシンプルなメモリ例 - 一枚のモジュール上に 8 個または 16 個の DRAM を搭載し、DRAM の信号ピンがモジュールコネクタに直結しているもの - に似ています。この実装は単純ですが、メモリバス上に最も重い信号負荷をもたらします。

RDIMM は、アドレス、制御、クロックラインのバッファリングとレジスタリングを行うため、システムメモリバスからこれらの負荷を取り除くことができ、これらの負荷を、DIMM 自身に備わる高機能な信号経路側に隔離することができます。これらの信号をバッファに入れるときは、レジスタに多少のレイテンシが生じるものの、一旦、転送が始まれば、パイプライン方式のメモリアクセスによって、フルスピードが達成されます。この場合、システムメモリバスから見えるロード数は、レジスタの 1 つのみとなるため、前例 (ロード数 = 8 個または 16 個) よりはるかに信号負荷が軽くなります。しかし、RDIMM では、依然、データ I/O 信号がコネクタと各 DRAM 間に直接接続されたままです (たとえば 4 ランク DIMM の場合、最大 4 ロードになる)。

そこで登場したのが、第 12 世代の PowerEdge サーバでもご利用いただける新テクノロジー、LRDIMM です。LRDIMM は、バッファデバイスを使用して、DRAM とシステムメモリバス間のデータ I/O 信号をバッファリングします。

その結果、たとえ 4 ランクの DIMM でも、システムメモリバスから見えるロードは、すべての DIMM コネクタピン上でたったの 1 つです。将来、メモリが拡張されると、この数は増えることが予想されます。

チャンネル、スロット、ランク：メモリサブシステムの構築方法

これまでのところ DIMM と DRAM の概念について説明してきましたが、それでは PowerEdge サーバは、これらを組み合わせてどのようにメモリサブシステムを構築しているのでしょうか？

プロセッサがメモリにアクセスするには、プロセッサのメモリコントローラとメモリモジュール間を結ぶ、少なくとも一本のメモリチャンネルが必要です。一本のメモリチャンネルには、DDR3 DIMM コネクタを含むスロットが 1 つ以上あり、このスロット数を増やせば、システムにより多くのメモリが追加できます。

しかし、チャンネルに設置できるスロット数には限りがあります。PowerEdge サーバの場合、一本のチャンネルに設置できる最大スロット数は 3 個です。幸い Dell PowerEdge サーバのハイエンドモデルでは、プロセッサから 4 本のメモリチャンネルが提供されるので、プロセッサあたり最大 12 スロットが提供できます。したがって、4 プロセッサ対応の Dell PowerEdge サーバには、48 個のメモリスロットが搭載されています。

3 つのスロットを備えた各種の DIMM に話を戻すと、次に検討するのは、チャンネルにどれくらいのメモリを追加できるか、ということです。DIMM は種類に応じて負荷や選択肢が異なりますし、どのチャンネルも、DIMM の種類別にサポートできる DIMM 数の上限が決まっています。表 2 は、PowerEdge サーバがサポートする DIMM タイプごとに、対応する最大 DIMM 数をまとめたものです。

表 2. チャンネルあたりの DIMM 数

DIMM の種類	ランク数 (DIMM あたり)	チャンネルあたりの 最大 DIMM 数	備考
UDIMM	1、または、2	2	
RDIMM (1 または 2R)	1、または、2	3	
RDIMM (4R)	4	2	チャンネルあたり 8 ランクに制限
LRDIMM	最大 4	3	

システムに構成できる最大容量を判断するには、まず、ランク、DRAM 編成、DIMM 設計についてもう少し詳しく知る必要があります。

本書の冒頭で、DIMM は最大 72 個の DRAM デバイスを保持することができ、DRAM は 4 ビットまたは 8 ビット幅で編成できることを説明しました。仮に 4 ビット幅のデバイスを使用する場合、最大 72 個の DRAM を使用する 4R DIMM では、ランクあたり 18 デバイスが配置されます。4 ギガビット DRAM の使用時は、DIMM 一枚につき最大 32GB の容量が提供可能です (付加されている ECC ビットは、容量に含めません)。ここで注意が必要なのは、8 ビット幅の DRAM を 72 個使用すると、8 ランクが必要となり、上表の最大ランク数を超えてしまうという点です。

32GB のメモリは、上表にある 4 ランクが可能なため、32GB RDIMM を使用した場合はチャンネルあたり二枚だけ設置できますが、32GB LRDIMM を使用すれば、チャンネルあたり三枚まで設置可能です。プロセッサあたり 4 チャンネルが提供できるシステムでは、1 基のプロセッサにつき 384GB が提供可能となるため、4 ソケットを備える特定の Dell PowerEdge サーバでは、合計 1.5TB になります。

ここで一言付け加えておくと、小さな DIMM 回路基板上 (30 x 133.35 mm = 1.18 x 5.25 インチ) に 72 個の DRAM を搭載するには、特別なパッケージングが必要です。一枚の基板上に搭載する DRAM 数が 36 個を超えるときは、1 つの SMD (Surface Mount Device、表面実装部品) 上に、2 個の DRAM から成る DDP (Dual Die Package、デュアルダイパッケージ) が配置され、デバイスピンに組み込まれた 2 つのチップセレクションが 2 つのダイのうち 1 つを選択します。

同じシステム内に、種類の異なる DIMM を混在させることはできないため、ご注意ください。同じシステム内に設置する DIMM はすべて、同じ種類で統一する必要があります。

チャンネル、スロット、ランク：メモリ速度への影響

第 12 世代の Dell PowerEdge サーバは、1600 MT/s の対応を含め、メモリサブシステムの性能をまったく新しい次元にまで高めました。さらに、デルサーバのメモリが期待通りのスピードを達成できるよう、綿密な設計と徹底した検証も行っています。

もう既にお気付きかもしれませんが、チャンネル内の負荷はメモリ速度に多大な影響を与える要因となっており、表 3 から、メモリのロード状況によって達成可能な最大スピードが変わることがわかります。

表 3. DIMM の種類と負荷に応じて変わる PowerEdge のメモリスピード

DIMM の種類	DIMM のランク	DIMM の定格電圧、速度	第 12 世代メモリの参照ポイント (ソケット R、3DPC プラットフォーム)					
			1 DPC		2 DPC		3 DPC	
			1.35 V	1.5 V	1.35 V	1.5 V	1.35 V	1.5 V
RDIMM	SR/DR	DDR3 (1.5V)、1600 MT/s		1600		1600		1066
RDIMM	SR/DR	DDR3L (1.35V/1.5V)、1333 MT/s	1333	1333	1333	1333		1066
UDIMM	SR/DR	DDR3L (1.35V/1.5V)、1333 MT/s	1066	1333	1066	1333		
RDIMM	QR	DDR3L (1.35V/1.5V)、1066/1333 MT/s	800	1066	800	800		
LRDIMM	QR	DDR3L (1.35V/1.5V)、1333 MT/s	1333	1333	1333	1333	1066	1066

- 緑の四角：性能、または、ワット性能比 (performance/performance per watt) 用のデフォルト
- 白い四角 (1.35 V の方)：選択可能だが、速度は低下
- 白い四角 (1.5 V の方)：BIOS からカスタム構成することで選択可能

DPC：DIMM per Chanel の略。チャンネルあたりの DIMM 数

SR：Single Rank の略。シングルランク (1 ランク)

DR：Dual Rank の略。デュアルランク (2 ランク)

QR：Quad Rank の略。クアッドランク (4 ランク)

チャンネル内に三枚目の DIMM をインストールすると、速度が大幅に低下することが表 3 から確認できます。これは負荷が原因であるため避けることはできませんが、性能を重視するお客様は、第 12 世代の Dell PowerEdge サーバの場合、チャンネルあたり二枚の DIMM を搭載することで、優れたワット性能比を達成します。

第 12 世代の PowerEdge サーバ: メモリの特長と強み

ここでは、第 12 世代 Dell PowerEdge サーバに搭載されているメモリの特長と強みについて詳しく見て行きます。

新機能と強化機能

第 12 世代では、旧世代よりオプションが拡充されており、容量や周波数の選択肢が増えています。さらに、選択できる DIMM の種類と構成方法の幅も増えたことから、信頼性/可用性/保守性 (RAS)、性能、電力、コスト間の兼ね合いが、きめ細かく調整できるようになりました。前世代の PowerEdge サーバで使われていた一部のメモリには相互運用性がありますが、UDIMM と RDIMM に限られます。表 4 は、第 11 世代 PowerEdge サーバと第 12 世代 PowerEdge サーバのメモリ比較表です。

表 4. 世代間の機能比較

項目	第 11 世代	第 12 世代
DIMM の種類	UDIMM ECC RDIMM	UDIMM ECC RDIMM LRDIMM
転送速度	800 MT/s 1066 MT/s 1333 MT/s	800 MT/s 1066 MT/s 1333 MT/s 1600 MT/s
電圧	1.5 V (DDR3) 1.35V (DDR3L)	1.5 V (DDR3) 1.35V (DDR3L)

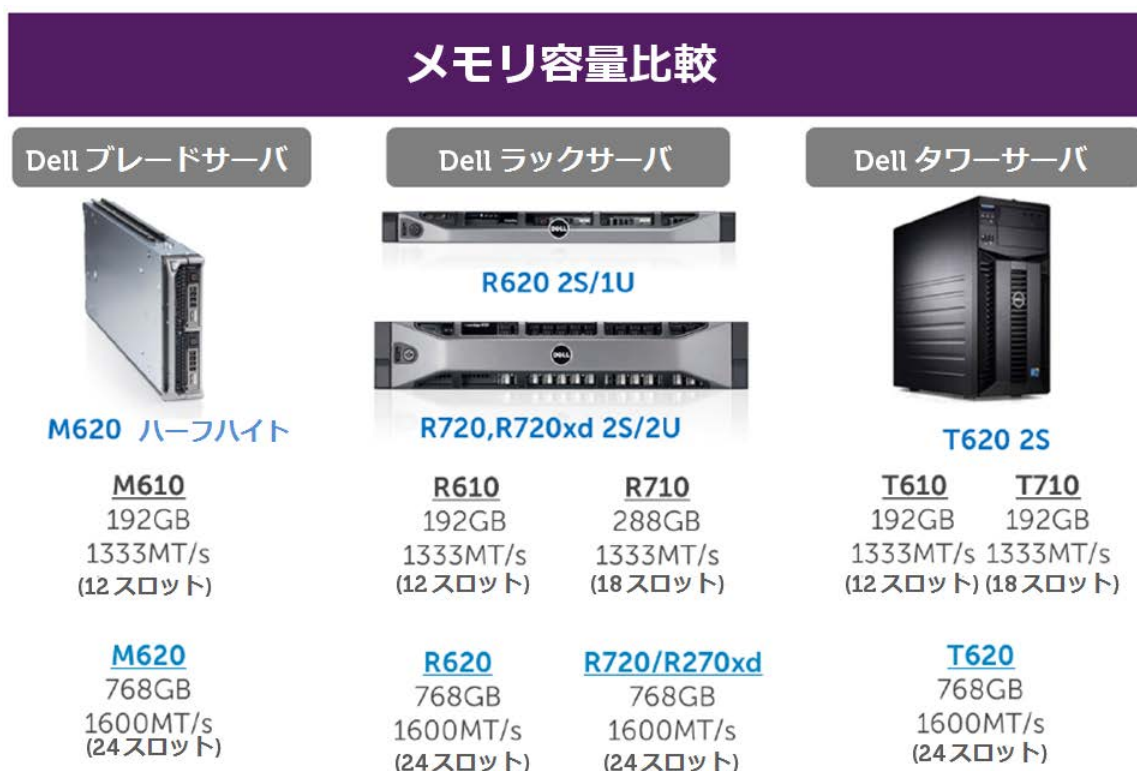
メモリ編成: インテル® Xeon® E5 プロセッサ

インテル® Xeon® プロセッサ E5-2600 および E5-4600 製品ファミリをサポートする第 12 世代の Dell PowerEdge サーバは、ラック型、タワー型、ブレードを網羅する幅広い製品構成となっており、2 プロセッサまたは 4 プロセッサ構成が可能です。このため、メモリ容量も幅広くお選びいただけます。

これらのシステムの基盤を成すのは、プロセッサあたり 4 本のメモリチャネルを提供し、チャネルあたり最大 3 個の DIMM スロットをサポートするインテル® Xeon® E5 プロセッサです。

図 1 は、第 11 世代と第 12 世代の PowerEdge サーバを、メモリの観点から比較しています。

図 1. PowerEdge サーバの世代別メモリ比較



1つの例外を除き、これらの全プラットフォームは、同じメモリ容量セットを提供します。その例外とは、PowerEdge M620 です。このシステムは極めて小型なため、最速のプロセッサを使用するときは、メモリ構成に一部制限が生じます。M620 ブレードサーバで 115W または 135W TDP (サーマルデザインポイント) プロセッサを使用すると、そのヒートシンクが DIMM スロットの障害物となってしまうため、各プロセッサあたり使用できる DIMM スロット数が 2 個ずつ減り、その結果、利用できる合計スロット数も、最大数の 24 個より少なくなります。構成に応じた正確なスロット数は、表 6 をご覧ください。

メモリチャネルの使用とペア構成

4本のメモリチャネルを提供するインテル® Xeon® E5 プロセッサには、4個のメモリコントローラがあり、それぞれ独立して動作することも、また、特定のモードでは、二個一組のペアで動作することもできます。それぞれのチャネルには、0~3の番号が付けられており、ペアモードで動作するときは、0と1、および、2と3という組み合わせになります。この組み合わせは、チャネルの物理的な位置によるもので、チャネル0と1はプロセッサパッケージの片側に、また、チャネル2と3は反対側に配置されています。

実生活でも「一長一短」と言うように、ある部分を最適化すれば、別の部分にトレードオフが生じます。たとえば、パフォーマンスを最適化すると、今まで信頼性を重視していた控え目なメモリチャネル設定が、論理的に影響を受ける可能性があります。詳細は後述しますが、具体的には各チャネルが個々に独立して動作し、全チャネルでメモリモジュールを同一に構成していれば、システムから最高レベルの性能を引き出すことができます。しかし、エラーチェックと修復機能を最重視する場合は、適切な RAS モードとメモリチャネルのペアリングを選択するようお勧めします。

第 12 世代の Dell PowerEdge サーバは、このような幅広い選択肢に完全対応できる柔軟性があり、第 12 世代のメモリがこれらのオプションを強力に支えます。

BIOS とシステム構成

第 12 世代の Dell PowerEdge サーバは、システムプロファイルを通して、サーバの動作条件を容易に設定できます。具体的には、BIOS セットアップから、希望する構成に合致したプロファイル - 性能最優先 (Highest Performance)、ワット性能比最優先 (Highest Performance per Watt)、高密度構成 (High-Density Configuration)、または、カスタム構成 (Custom Configuration) - を選択するだけです。

最初の 2 つのプロファイルは読んで字の如くですが、その次の高密度構成は、性能よりも信頼性と対応能力を重視する設定となっており、システムを意図的に「性能控え目」で動作させます。これはメモリサブシステムのデフォルトに最も影響を与える設定で、クロックの抑制、より高い電圧での動作 (対応に余力を持たせるため)、各種の安全措置 (例：通常の 2 倍の頻度でリフレッシュ) などが選択されます。この結果、たとえ、メモリを最大構成したプラットフォームでも、信頼性が最大限に高まります。

もちろんカスタム構成なら、システム設定を自在に制御できるので、お客様固有のニーズに合わせた調整が可能です。

RAS (信頼性、可用性、保守性) 機能

現世代のシステムは、前世代から多くの機能強化が加えられましたが、中でも新しい RAS 機能の数々は、新プラットフォームの信頼性と可用性をかつてないレベルにまで押し上げています。これらの新しい RAS 機能を語るだけで一冊のホワイトペーパーが書けてしまいますが、ここでは、利用できる機能のほんの一部をご紹介しますと思います。表 5 は、メモリ特有の RAS 機能をまとめたものです。

表 5. RAS 機能

カテゴリ	RAS 機能
DIMM	ECC
	レジスタ/PLL
メモリサブシステム	SDDC
	ミラーリング
	ランクスペアリング
	デマンド/パトロールスクラビング
第 12 世代の新機能/ アドバンスド機能	メモリバッファ (LRDIMM)
	DIMM SPD エラーログ
	修正可能エラーのスレッショルド
	メモリページのリタイア
	高密度プロファイル
BIOS 対応	汚染データの拡散防止
	MCA リカバリ
	デバイスのタグ付け

DIMM

ECC メモリには多大なメリットがありますが、RDIMM や LRDIMM に備わるレジスタバッファによって、メリットがもう 1 つ加わります。それは、アドレスとコマンドの両方に対するパリティチェック能力であり、UDIMM にはこの能力がありません。ECC は誤ったアドレスやコマンドを検出することができ、メモリデータが壊れていることをシステムに伝えます。他の RAS 動作の設定内容にもよりますが、このような状況をトラップすると、再試行するか、別の DIMM/ランクで処理することで、エラーから回復できる可能性があります。

メモリサブシステム

SDDC (Single Device Data Correction)

SDDC (Single Device Data Correction) は、1 つの x4 または x8 DRAM からワード単位で流される一連のデータを受け取ると、バーストモードでアクセスできるように、チャンネル内でひと塊のデータ (バンドル) に編成し直します。万が一、バンドル内の一部または全部のビットに破損が生じたときは、修正処理が起動されます。x4 の場合は、どのような状況でも修正されますが、x8 の場合は、システムがアドバンスト ECC (ロックステップ) モードのときのみ修正されます。

第 12 世代 Dell PowerEdge サーバのメモリは、単なる前世代からの強化にとどまらず、メモリ製品構成がより包括的になっており、また、アドバンスト ECC モードがアクティブなときは、メモリチャンネルロスを回避するアーキテクチャも実現されています。

ミラーリング

メモリミラーリングは、ペアのチャンネル間でメモリをコピーすることで、100% の冗長性を提供します。第 12 世代のメモリミラーリングは、旧世代のサーバで発生していたメモリチャンネルの完全損失に陥ることはありません。

ランクスペアリング

ランクスペアリングとは、チャンネル内に少なくとも 4 つのランクか、2 つの DIMM が残されている限り、メモリの 1 ランクをスペアとして予約しておく機能です。メモリコントローラは、修正可能なエラーが頻繁に発生するようになったランクを、スペアランクに移行します。OS やアプリケーションからは、このスペアランクが見えません。本機能は、前世代のサーバでもご利用いただけます。

デマンドパトロールスクラビング

メモリサブシステムは、読み取り中に見つかったエラーに正しい情報を書き込む、自動訂正に対応します。

新規、または、機能強化されたテクノロジー

メモリバッファ (LRDIMM)

RDIMM 内のレジスタがパリティチェックを提供するように、LRDIMM もアドレスおよびコマンドラインにパリティチェックを提供します。さらに LRDIMM は、データ I/O ラインもバッファリングできるので、負荷のかかったデータパスが DIMM モジュール側に隔離され、システムの堅牢性が一層高まります。

DIMM SPD エラーログ

それぞれの DIMM には、自身に関する情報が書き込まれた SPD (Serial Presence Detect) と呼ばれるメモリが搭載されており、システムはこれを見て DIMM の特性を知ることができます。この不揮発性フラッシュメモリは、小さな指定エリアに少量の OEM データを保存しています。第 12 世代の PowerEdge サーバは、このスペースを有効活用して、メモリエラー発生時に収集された重要な情報を記録するようにしたため、障害発生時の分析時間が大幅に節約できます。第 11 世代の PowerEdge サーバでは、同種の機能を限定的に提供していました。

修正可能エラーのスレッシュホールド

第 12 世代サーバの BIOS では、修正可能なエラーの追跡機能がより堅牢になりました。各 DIMM は、修正可能エラーの頻度と場所が追跡できるように、この種の記録をメモリ内に残しています。修正可能エラーの発生率が急増しているときは、故障の前兆を示していることがあるため、メモリサブシステムがこれらのエラーログを受け取ると、システム管理サブシステムにエスカレーションします。

メモリページのリタイア

これは、第 12 世代の PowerEdge サーバで初めて導入された、デル独自のまったく新しい機能です。この機能は、稼働中のハイパーバイザーと連携しながらメモリ障害を監視し、特定のエリア (メモリページなど) の修正可能エラー発生数が特定の閾値 (スレッシュホールド) を超えると、問題のページをリタイアさせてシステムから効果的に取り除き、今後、そのエリアにアクセスしないようにします。このイベントは、ログにも書き込まれるため、次の定期メンテナンス時に当該メモリを交換することができます。

高密度プロファイル

前述のとおり、このプロファイルを利用すると、より安全圏内の動作パラメータを使ってシステムが構成されるため、システム全体の可用性と堅牢性が高まります。

汚染データの拡散防止

これはインテル[®] チップセット内に組み込まれている機能ですが、Dell BIOS と統合されています。メモリ読み取り中に修正不能エラーが検出されると、一旦保留状態に置いておき、プロセッサがこの破損データ値を実際に使用するときが来たら、レポートします。キャッシュされたデータが破損してもシステムには影響がないため、システム全体の性能が向上します。

ちなみに、エラー報告を先延ばしにするこのメカニズムは、「Error Containment Bit」 (エラーコンテインメントビット) を使用しますが、これは通称「Poisoned」 (毒入り) データと呼ばれます。

BIOS 対応

MCA (マシンチェックアーキテクチャ) リカバリ

エラー復旧メカニズムの一種である MCA (Machine Check Architecture、マシンチェックアーキテクチャ) リカバリは、BIOS が様々なマシンチェック状況をインテリジェントに処理して、可能なときは復旧を試みます。たとえ復旧が不可能でも、障害イベントの記録がログに残ります。最悪の状況ではシステムが自動シャットダウンし、再起動します。

デバイスのタグ付け

DRAM の 1 つにエラーが発生し始め、障害が差し迫っていると見なされると、システムは、そのデータ DRAM の内容を、普段 ECC ビットの維持に使われている DRAM に移動することができます。この措置により、ECC (エラーの検査と訂正) 能力は落ちますが、障害を未然に防ぐことができます。ハードエラーとスタックビットを排除するこの機能は、DIMM を交換するまでの暫定措置として非常に有効です。

サポート対象のメモリ

今回、メモリで最も変わったことの 1 つは、テクノロジーとは直接関係がなく、主にデルの運営方針に関わることです。具体的に言うと、お客様へのメモリ提供体制が変わりました。前世代ではメモリ構成を固定しており、システムレベルで特定の種類と容量を決めていました。しかし、第 12 世代のサーバでは、DIMM を包括的に取り揃え、選択肢が広がったため、お客様固有のニーズに合わせてシステムを構成できます。表 6 は、PowerEdge R720 サーバ内に 256GB のメモリを設置するときの構成例です。

表 6. PowerEdge R720 に 256GB メモリを設置するときのサンプル構成

枚数	容量	種類	特長
16	16GB	4Rx4 1066 RDIMM	最も低コストだが、動作スピードは最も低速 (800)
16	16GB	2Rx4 1333 RDIMM	コストは標準的、動作スピードは良好 (1333)
16	16GB	2Rx4 1600 RDIMM	コストは上記より若干高めだが、動作スピードは高速 (1600)
8	32GB	4Rx4 1333 RDIMM	高価で動作スピードは妥当 (1066) だが、スケーラビリティは良好
8	32GB	4Rx4 1333 LRDIMM	最も高価、良好な動作スピード (1333)、優れたスケーラビリティ

表 7 は、第 12 世代の Dell PowerEdge サーバがサポートするメモリー一覧です。

表 7. メモリー一覧

DIMM 速度	DIMM の種類	DIMM の容量 (GB)	チャンネルあたりのランク数	データ幅	サポートする SDDC	DIMM 電圧 (V)
1600	RDIMM	2	1	x8	アドバンスド ECC	1.5
1333	RDIMM	2	1	x8	アドバンスド ECC	1.35
1333	UDIMM	2	1	x8	アドバンスド ECC	1.35
1600	RDIMM	4	2	x8	アドバンスド ECC	1.5
1333	RDIMM	4	2	x8	アドバンスド ECC	1.35
1333	RDIMM	4	1	x4	全モード	1.35
1333	UDIMM	4	2	x8	アドバンスド ECC	1.35
1600	RDIMM	8	2	x4	全モード	1.5
1333	RDIMM	8	2	x4	全モード	1.35
1600	RDIMM	16	2	x4	全モード	1.5
1333	RDIMM	16	2	x4	全モード	1.35
1333	LRDIMM	32	4	x4	全モード	1.35
1333	RDIMM	32	4	x4	全モード	1.35

まとめ

デルは、Dell PowerEdge サーバのお客様に最先端メモリテクノロジーをお届けすることを重視しています。第 12 世代 PowerEdge サーバでは、メモリ製品構成に 32GB RDIMM が加わったため、メモリ容量がさらに拡張できます。また、1600MT/s DIMM も加わったことから、メモリレイテンシの削減とメモリバンド幅の向上が達成でき、ひいては、アプリケーション性能の向上も期待できます。第 12 世代の Dell PowerEdge サーバに投じられた強力なメモリ戦略によって、お客様は、ワークロードニーズを満たし、実業務アプリケーションに対応する最新テクノロジーとアドバンスド機能が得られます。