

IBM、NVLink付POWER8 3システムを発表

John Russel



2017年に出荷する予定の次世代POWER9チップの詳細が明らかになってからそれ程時間が経っていないが、IBM

は、三つの新しいPOWER8ベースのLinuxサーバーとNVLinkインターコネクトを内蔵したPOWER8チップの新しいバージョンを発表した。そのサーバーのひとつは、Power S822LC for HPCであり、新しいチップ(NVLink内蔵POWER8)は、NVIDIAの最新の最高性能GPUであるP100 Pascal GPUと通信する。

他の二つのサーバーは、Power S821LCとPower S822LC for Big Dataであり、PCIeインターフェース経由でGPUアクセラレーション技術(K80またはP100)を活用し、フラッシュストレージおよびFPGAで使用するためのIBMのCAPI(Coherent Accelerator Processor Interface)を有している。すべての3つのサーバーは、IBMのLinuxのラインに標準的な2ソケットの追加になる。

IBMのHigh Performance Computing and Analytics部門のVPであるSumit Gupta氏は、これらの新しいサーバーの導入は、新しいパラダイムであるアクセラレーション・コンピューティングや主要な市場促進要因である認知コンピューティング(特筆される)とビッグデータ分析に向けたIBMのビジョンに対する継続的なコミットメントの証しとして見られるべき

です、と述べている(HPCwireの記事「[Think Fast: IBM Talks Acceleration in HPC and the Enterprise](#)」を参照)。

また、注目すべきは、新しいシステムがパートナーによって製造されていることだ。「これらのOpenPOWERシステムの3つすべては、アクセラレーション機能から設計と製造にいたるまで、OpenPOWERパートナーの強みと専門知識を活用しています。オープンな精神に基づき、これらのシステムを製造している私たちの業界パートナー、OpenPOWERパートナーでもあるWistron社およびSupermicro社は、彼らの顧客に対してOpenPOWERエコシステムを増殖させるために、POWERベースのサーバーを提供することになると期待している。」とGupta氏は述べている。

シリコンレベルで埋め込まれたNVLinkを備えたS822LC for HPCは同等のx86ベースのシステムよりも「5倍は速くデータを転送できる」とIBMは言っている。発表に際しGupta氏が書いている詳細なブログによれば、これによりGPUの積極的な使用に際しプログラミングレベルの障壁を大幅に低減することができる。IBMによると、これらのサーバーはNVLinkとともに提供される初めてのPOWER8ベースのシステムだ。

「CPUからGPUへデータを移動する際、それがボトルネックになっていました。何故ならほとんどのシステムで、PCIeという細いパイプを通過しているためです。NVLinkによって、GPUがはそのリンクの相手であるCPUと1TBの半分までのアクセスが可能になりました。」

と、Gupta氏はHPCWireに語った。NVLinkは両プロセッサ間のデータ転送を改善し、それは基本的にプログラムすることが容易になる。

「アプリケーションが起動されると、すべてのデータは、システムメモリ上にあります。あなたは、GPUへデータを移動しなればなりません。」

IBM New LC Server Specification

IBM

S822LC for High Performance Computing



System Details

- 2-socket, 2U
- Up to 20 cores (2.86-3.26Ghz)
- 1 TB Memory (32 DIMMs)
- 230GB/sec memory bandwidth
- 2x SFF (HDD/SSD), SATA
- Up to 4 integrated NVIDIA Pascal GPUs
- 3 PCIe slots, 3 CAPI enabled, IB Add-in
- Air or water cooled

S822LC for Big Data



System Details

- 2-socket, 2U
- Up to 20 cores (2.9-3.3Ghz)
- 512 GB Memory (16 DIMMs)
- 115GB/sec memory bandwidth
- 12 SFF/LFF (HDD/SSD) 96 TB storage
- 5 PCIe slots, 4 CAPI enabled
- 2 NVIDIA K80 GPU capable

S821LC



System Details

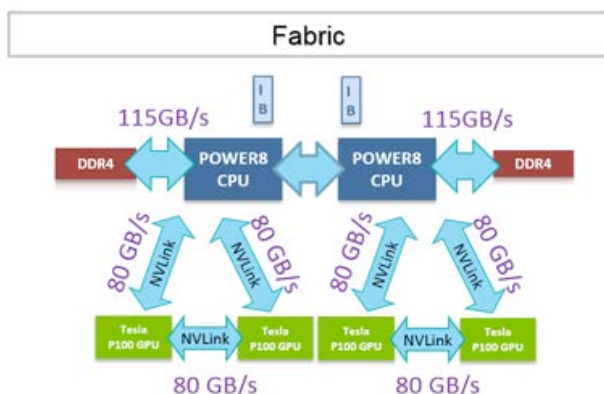
- 2 socket, 1U
- Up to 20 cores (2.09-2.32Ghz)
- 512 GB Memory (16 DIMMs)
- 115 GB/sec memory bandwidth
- 4 SFF/LFF (HDD/SSD), 32 TB Storage
- 4 PCIe slots, 3 CAPI enabled
- 1 NVIDIA K80 GPU capable

NVLink は、3 つのを行います。我々はプロセッサ間に太いパイプを用意したので転送性能が向上します。それにより、より小さな機能を GPU に移動することができます。GPU 側でのデータ管理を減らす事ができるので、アクセラレータ側のプログラミングが簡単になります。」

NVLink を搭載した新しい POWER8 プロセッサは 10 のコアを備え、最大 3.26 GHz で動作する。このサーバーで使われている POWER8 プロセッサは、x86 CPU よりも高いメモリバンド幅、115 ギガバイト/秒を有しており、ソケットあたりのシステムメモリを 0.5 テラバイトまで持つことができる。POWER8 プロセッサでは、コア当たり大きなキャッシュがあり、これが高速なコアに結合され、高いメモリ帯域幅がアプリケーションの性能とスループットの向上をもたらしている。

新しい NVIDIA の Tesla P100 GPU アクセラレータは浮動小数点性能を向上させ、半精度で 21 テラフロップス、単精度で 10.6 テラフロップス、倍精度で 5.3 テラフロップスを実現している。アクセラレータは 16GB の HBM2 積層メモリを持ち、GPU との間で 720GB /秒メモリ帯域幅を提供している。

NVLink 結合を行う NVIDIA の Tesla P100 GPU の SXM2 フォームファクタは、PCI-E 系のシステムよりも 14%以上の演算性能を提供する。



Gupta 氏は自身のブログに、アプリケーション・アクセラレーションにおける NVLink の三つの主要な利点を記載している。

性能： NVLink を搭載した新しい POWER8 プロセッサと新しい Tesla P100 GPU は 4 つの NVLink インターフェースを持ち、「他のシステムで使用される、PCIe x16 Gen3 の接続よりも 5 倍の高速通信を」可能にする。これにより、GPU との細い PCIe データパイプの限界を克服して、より高速なデータ交換とアプリケーションの性能向上を可能にする。

プログラマビリティ： CUDA 8 ソフトウェアと Tesla P100 のページマイグレーション・エンジンによって、CPU と GPU のメモリ間で統合メモリ空間が提供される。NVLink と相まって、統合メモリ空間は GPU アクセラレータのプログラミングが開発者にとってより容易になる。アプリケーション開発者は、データ管理に苦勞すること無く、順次 GPU に CPU の機能を移動させることにより、加速化を行うことができる。

加速化するアプリケーションの多様化： NVLink は、CPU と GPU 間の通信時間を短縮するので、加速のために GPU に転送するデータが小さい事を可能にします。これは、多くのアプリケーションが、GPU で加速できることを意味します。

IBM は発表を行う際に、「IBM は Intel よりも良い」という言い方を徐々に増やしている。その広範なアプリケーションのターゲットは、ビッグデータと分析だけでなく、ディープラーニングや認知コンピューティングを包含している。

「データ分析、データベース分野そして機械学習とディープラーニングさらに人工知能を含めた HPC 分野において、我々は POWER8 の市場での優位性を感じています。」と、Gupta 氏は語っている。「何故なら我々は速いコアを持っており、例えばデータベースでは Intel ベースのシステムに比べて私たちはより高い性能を見えています。『kinetica』のような GPU により加速化されたディープラーニングや機械学習向けのデータベースアプリケーションにとって、CPU と GPU 間の NVLink 高速データ接続は価値ある事です。」

x86 の市場シェア獲得で苦戦を強いられた IBM は、その成功へ転ずる為にハイパースケーラーに浸透するための努力を過去に強調していました (HPCwire の記事「[Handicapping IBM/OpenPOWER's Odds for Success](#)」を参照)。

IBM の発表によると、「世界最大のインターネット・サービス・プロバイダのひとつであり、中国に拠点を置く Tencent における初期テストの結果、新しい Power S822LC for Big Data の大規模クラスターが以前の x86 ベースのシステムと比較して、データ集約型のワークロードで 3 倍の性能を達成した。それ以上に、この結果はサーバーの数を三分の二に削減して達成されたのです。少数のサーバーを使用して、より高い性能を実現できる大幅なコストメリットを考えて、同社は現在、ビッグデータのワークロードのために、そのハイパースケールデータセンターに新しい LC サーバーを統合しようとしています。」



Sumit Gupta、IBM

Gupta 氏は、POWER9 の発表にもかかわらず POWER8 ベースの新しいサーバーに対する大きな需要があると主張している。「いくつかの企業、研究機関そして政府機関は初期のシステムを事前にテストして発注しています。その中でも製品の出荷を先頭で待っているは多国籍小売販売巨大企業と米司法省のエネルギーのオークリッジ国立研究所 (ORNL) です。」と、IBM は発表している。

ORNL は、内蔵 NVLink インターフェース技術を活用してアプリケーションを最適化するための開発プラットフォームとしてこの新しいシステムを使用する。さらにこのシステムは、IBM が 2017 年に提供する POWER9 チップベースのオークリッジ国立研究所の次世代スーパーコンピュータ「Summit」に必要なアプリケーションを開発するための初期テストベッドの役割を務めることになる。OLCF プロジェクトディレクターである Arthur S. (Buddy) Bland 氏は、プレスリリースで次の様に述べている。「GPU の長年のユーザーとして、私たちはこのシステムによりアプリケーションの性能が向上し、ユーザーが素晴らしい成果を容易に達成できるようになるであろうと信じている。」

エコシステムを構築することはむづかしい問題だ。IBM と OpenPOWER にとっては、たくさんの多様なピースが、必要に応じて一箇所に集結する必要がある。結果は時が教えてくれるだろう。

Gupta 氏のブログへのリンク：

<http://www.ibm.com/blogs/systems/ibm-nvidia-present-nvlink-server-youve-waiting>

お問い合わせ

Web フォーム：ibm.biz/firstpower

お電話：IBM アクセスセンター 0120-550-210

出典：HPCwire Japan 2016 年 11 月 24 日掲載記事「IBM、NVLink 付 POWER8 と 3 システムを発表」<http://www.hpcwire.jp/archives/11605>

画像の出典：International Business Machines Corp.

©米国 Tabor Communications, Inc.および株式会社イージーマネージメント

本記事の情報は 2016 年 11 月現在のものです。仕様は予告なく変更される場合があります。記載の事例は特定のお客様に関するものであり、全ての場合において同等の効果が得られることを意味するものではありません。効果はお客様の環境その他の要因によって異なります。製品、サービスなどの詳細については、弊社もしくはビジネス・パートナーの営業担当員にご相談ください。

IBM、IBM ロゴ、ibm.com および Power、Power Systems、POWER8 は、世界の多くの国で登録された International Business Machines Corp.の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM 商標リストについては www.ibm.com/legal/copytrade.shtml をご覧ください。