2.5.3 データセンタースケール・コンピューティング

（1）研究開発領域の定義

Google、Amazon、FacebookなどのIT巨大企業はサービスを提供するために巨大なデータセンターを運用している。数十万台、あるいは数百万台のサーバーを設置し、それをつなぐネットワーク、ストレージなどが数百メートル四方の広さの建屋に設置されている。このような大規模データセンターにおいては、従来のラックスケールの考え方だけでは不十分である。物理的な制約も考えて、サーバー、ネットワーク、ストレージの配置や処理方式を最適化しなければならない。本研究開発領域においては、こういった大規模データセンターに向けた研究開発課題を俯瞰する。

（2）キーワード

データスケール・コンピューティング、メモリー・セントリック・コンピューティング、Hyper Converged Infrastructure、AIアクセラレーター、デジタルアニーラー、高速不揮発性メモリー、高速インターコネクト

（3）研究開発領域の概要

［本領域の意義］

情報検索やECサイト、SNSなどのサービスがグローバルに拡大するとともに、スマートフォンとクラウドコンピューティングの融合によるさまざまなサービスの利活用、IoTを代表とする膨大なデータの取得と処理などによって、データセンターで処理するデータ量は近年ますます増加している。これまでのラックスケールでの計算実行から、データセンタースケールでの計算の実行へと移りつつある。また、膨大なデータを対象とした情報処理では、計算（プロセッシング）よりもデータの記憶・移動に要する時間がボトルネックとなる。そのため、プロセッサーを中心とするコンピューティング（プロセッサー・セントリック・コンピューティング）だけでなく、メモリーを中心とするコンピューティング（メモリー・セントリック・コンピューティング、あるいはデータ・セントリック・コンピューティング）が必要になる。磁気メモリー（MRAM）、抵抗変化型メモリー（ReRAM）、相変化メモリー（PCRAM）といった高速な不揮発性メモリーや、光通信技術などの登場により、従来のFlash/HDDデバイスや電気通信と比較してデータを高速に転送・記憶することが可能になりつつある。このような、膨大な数のサーバー・メモリー・ストレージ・インターコネクトで構成されるデータセンターが今後のクラウドコンピューティングにおいて重要な役割を果たす。

［研究開発の動向］

ムーア則の終焉を迎え、コンピューティングデバイスでは各種の処理に特化したアクセラレーターの開発と実適用が進んでいる。 AI技術の発展とともに適用例が増えているGPUが代表的であるが、ネットワークやストレージにおいてもアクセラレーター適用の研究開発が進んでいる。このようなコンピューティングデバイスだけでなく、データの蓄積・供給側であるメモリーでも特徴的なデバイスの開発が進められている。2015年にIntelとMicronが共同でDRAMより大容量な128Gbitの3D-Xpointメモリーを発表[1)]、2018年5月にはIntelがDDR4スロットに挿せる「インテル® Optane™ DCパーシステント・メモリー」を発表した[2)]。2019年には各社からこのメモリーを搭載するサーバーが発表され、これまでのコンピューターアーキテクチャーには存在しなかった不揮発なメモリーレイヤーを活用する研究開発が進められている。例えば、SAPはこの大容量不揮発性メモリーを自社のインメモリーDB製品に適用し、大容量と不揮発性を生かしデータ処理

能力と可用性の劇的な向上を実現している[3]。米国の次世代スパコンに向けて開発中である分散ストレージDAOS[4]は、Optane DCメモリーとNVMeデバイスを利用することで、開発中ながら世界最速のIO性能を実現している。また、DRAMの大容量化の鈍化を解決するため、Optane DCメモリーを単なる大容量メモリーとして利用し、アプリケーションに合わせて高速なDRAMと低速であるが大容量なOptane DCメモリーを適切に組み合わせることで高速大容量にする取り組みが進んでいる[5]。SamsungやKioxiaからは、アクセス時間が10μs以下の極めて高速であるが小容量なFlashメモリー技術[6]と、それを採用したSSDが発表されるなど、微細化による汎用品の性能・容量向上技術から、次の時代に向けた目的特化の技術の開発が加速している。

これら特徴的なコンピューティングデバイスとメモリーデバイスを活用する新しいアーキテクチャーの取り組みとして、HPEのThe machine[7]が挙げられる。大容量な不揮発性メモリーを中心に、高速ネットワークで各種の特化型プロセッサーと接続することで、データ移動を抑えたデータセントリックのコンピューティングを目指している。2020年9月に発表されたIBM社のPower10プロセッサーでは、クラスター内のノードのメモリーをどのノードからもアクセスできる機能を搭載することで、最大2PBの巨大なメモリーを中心とするコンピューティングを実現している[8]。

データセントリックの概念はビッグデータ処理に特化したソフト技術によりApache Hadoop[9]でも一部実現されている。Hadoopでは、大量のデータを複数ノードで構成された分散ファイルシステムに格納し、極力データのあるノードで処理を行う。既存アーキテクチャーをベースに、ストレージとコンピューターノードがネットワークで分断されることなく、データのあるところで処理するデータセントリックな構成となっている。また、従来は専用ハードで提供されていたネットワークやストレージ機能をソフトウェアで実現するSDx（Software Defined X）化が進展し、汎用サーバー上のソフトウェアとして統合型基盤を実現できるようになってきた。事前にネットワーク機能やストレージ機能をインストールしておくことで、設置や増設時の設定作業を排したHyper Converged Infrastructure（HCI）[10]が注目されている。HCIは簡単利用が注目され普及してきたが、HCIではローカルディスクを利用するため、仮想マシンやコンテナをデータが保存されたノードに配置することで、データ移動を最小限に抑え、優れたパフォーマンスを実現するアーキテクチャーとなっている。

既存の多くのデータセンターは仮想化基盤上で運用されるが、これは性能が同じである大量のコンピューティングリソースから、ユーザーが求めるリソースを切り出し提供する技術である。ドメイン指向コンピューティングでは、様々なコンピューティングリソースやメモリーリソース中からアプリケーションに最適なハードを選択し再構成する技術が必要となる[11]。Intelのラックスケールデザイン[12]では、サーバーを構成するプロセッサーやメモリー、ストレージなどを分解し、サーバー単位ではなくラック単位でコンピューターを構築することにより、アプリに応じて最適な構成をとることが可能となっている。製品化の例として、HPEの「HPE Synergy」[13]やLiquid[14]、DriveScale[15]の製品などが挙げられる。これらはコンピューター、ストレージがリソースプール化されており、各ワークロードに合わせて再構成できる。いずれもラックスケールでの例であるが、データセンタースケールに拡張することでさらなるリソースの最適化が可能となる。

また、リアルタイム性を要求するシステムにおいては、データセンター側までデータを転送する時間が許容できず、エッジ側での処理が必要な場合もあり、データセンター側とエッジ側の連携アーキテクチャーが重要な課題である[16]。大手パブリッククラウドベンダーは、パブリッククラウドのインフラとサービスをオンプレミスの施設で提供することにより、エッジ側での低レイテンシー処理を含めて、データセンターを超えたコンピューティングを実現する取り組みを始めている[17]。

（4）注目動向

［新展開・技術トピックス］

次世代メモリーの実用化とプログラミングモデルの整備

2015年にIntelとMicronが共同でDRAMより大容量な128Gbitの3D-Xpointメモリーを発表した。学会では各種の次世代不揮発性メモリーが議論されてきたが、3D-XPointは商用化された最初の大容量不揮発性メモリー技術である。ダイあたりの容量は128GbitとDRAMより1桁大きく、2015年当時のFlashメモリーに近い容量である。高速SSDとして製品化され、FlashメモリーベースのSSDより1桁高速な性能を実現している[18)]。また、2018年5月にはDDR4スロットに挿せる「Optane DC」不揮発性メモリーを発表し、主記憶としての不揮発性メモリーを実現した。従来、主記憶であるDRAMは揮発性であることを前提にプログラミングされており、主記憶が不揮発化することを想定していない。ストレージの業界団体であるStorage Networking Industry Association（SNIA）は、不揮発性メモリーに対するプログラミングモデルを標準化し、Persistent Memory Development Kit（PMDK）[19)] と呼ばれる不揮発性メモリーを活用するためのライブラリーが開発されている。その中で、2020年からはRemote Persistent Memory Access（RPMA）[20)] が提供されており、データセンター用ネットワークを介した高速なサーバー間不揮発メモリーアクセスが実現されている。また、高速ネットワークと不揮発メモリーを組み合わせることにより、異なるサーバー上の記憶装置にデータを記録するために要する時間を従来のシステムに比べて劇的に短縮する取り組み[21)] も始まっている。さらに、Hypervisor、OS、ファイルシステムでもバイトアクセス可能な不揮発性メモリーを活用する対応[22)]が進められ、新しい時代のメモリーデバイスの活用に向けて環境の開発が加速している。

Compute Express Link：CXL

2019年3月に大手コンピューター関連企業9社（Alibaba, Cisco, Dell EMC, Facebook, Google, HPE, Huawei, Intel, Microsoft）によって設立されたコンソーシアムで、データセンター向けにCPUとメモリー、GPU、FPGAといったアクセラレーター間を高速で通信するインターコネクトの実現を目指している[23)]。同様の規格としては、CCIX[24)] が2016年に設立され2018年に基本仕様1.0がリリースされている。HPCの主要会議であるSuper Computing 2019（SC19）でIntelから米国次世代スパコンに向けCXLに対応したGPUが発表されているが、PCI Express 5.0（PCIe Gen5）をベースとするため、製品化は2021年以降とみられている。CXLではCPU、アクセラレーターに向けて低レイテンシーでキャッシュコヒーレントなインターコネクトを実現するため、メモリーを中心としたコンピューティングに必要な材料が揃いつつあると言える。CXLコンソーシアムの策定と同時にCXL仕様1.0が公開され、事前に十分に仕様が検討されていたことがわかる。

Open Compute Project（OCP）

OCPとは，2011年にFacebookの呼びかけにより発足したデータセンターのファシリティーからサーバーに至るまでの管理仕様を公開したプロジェクト[25)]。現在は、MicrosoftとIntelを加えた3社が牽引し、200社弱のベンダーが参画、データセンターの標準化を目指している。ハードウェア仕様の標準化からスタートし、2017年のSummitでは、ソフトウェアやファームウェアの標準化にも着手している。 MicrosoftはProject Olympusというサーバー仕様を公開、データセンターの管理粒度をサーバーからラックにしようとしている。Intelは、自社のIntel Rack scale designを標準化すべくOCPに参画している。

［注目すべき国内外のプロジェクト］

日本ではAI応用などに向けた新しいコンピューティングの国家プロジェクトとして、2018年度からJST CREST「Society5.0を支える革新的コンピューティング技術」、JSTさきがけ「革新的コンピューティング技術の開拓」、NEDO「高効率・高速処理を可能とするAIチップ・次世代コンピューティングの技術開発事業」が実施されている。さらに2020年度からJSTCREST・さきがけ「情報担体を活用した集積デバイス・システム」が始まっている。しかし、カバーすべき領域がエッジ、データセンター、インターコネクトのハード・ソフトと広範であるため、例えば本研究開発領域であるデータセンター向けの研究開発が十分にカバーされているとは言い難く、より一層の強化が必要である。

米国においては、The National Computing Initiative（NSCI）が2015年の大統領令によって発足し、ハイ・パフォーマンス・コンピューティングでの米国のリーダーシップを実現しようとしている。 NSCIは複数の省庁にまたがる統一的な戦略であり、産業、アカデミアの連携のもとに活動を進めている。 National Science Foundation（NSF）においても、NSCI@NSFと題して、HPCの研究開発と配備に対してファンディングを行っている。 Energy-Efficient Computing：from Devices to Architectures（E2CDA）というプログラムでは、材料からデバイス、コンピューティングアーキテクチャーなどの研究に対して、2016年から18年に合計で2千2百万ドル、Scalable Parallelism in the Extreme（SPX）においても2千万ドルの研究資金を提供している。また、Petascale Computing Resource Allocations（PRAC）においては、実際にスーパーコンピューティング環境を構築するために、2013年から2018年までに3億3千万ドルが投資されている。ほかにも、サイエンスのためのコンピューティング、教育、量子コンピューター、産学連携など多彩なプログラムが用意されている。

これまでコンピューティングの性能向上を支えてきたムーアの法則の終焉を迎え、コンピューティングを根底からすべて考え直そうという動きがIEEEのRebooting Computing Initiativeである。材料、デバイス、システム、アーキテクチャー、言語など多くの領域の専門家が集い、コンピューティングの将来を模索している。

欧州においてはHORIZON2020において、Advanced Computingというトピックを設定し、2014年から研究提案を募集し、現在30以上のプログラムが実行されている。低電力プロセッサー、フォグコンピューティング、ディープラーニングの学習アルゴリズムなど多岐にわたる研究が推進されている。2015年から2017年の間に開始されたプロジェクトへの投資金額は総額1億5千万ユーロに達する。2021年から開始されるHorizon Europeにおいても、Digital and Industryというクラスターの中にAdvanced Computing and Big Dataという領域を設け、ハイ・パフォーマンス・コンピューティング、ビッグデータおよびICTにおける低炭素化などに関する研究開発が予定されている。

（5）科学技術的課題

AIアクセラレーター、高速な不揮発性メモリー、高速インターコネクトなど、AIを使ったリアルタイム処理などを差別化するハードウェア技術の研究を強力に推進することは当然、必要である。ただし、こうしたハードウェアのキラー技術の能力を最大限に発揮させるためには、個別のハードウェアの研究開発だけでは不十分であり、ハードウェアを融合したサーバー・ストレージを構築するためのOS、デバイスドライバ、分散処理・データベース等のミドルウェアなど、ハードウェア・ソフトウェアを統合・全体最適化したコンピューティング技術の研究開発が必須になる。ハードウェア・ソフトウェアの統合は1つの企業、単独の研究機関・大学では行うことは難しく、分野のレイヤーをまたいだ複数の産学の連携が必要になる。

（6）その他の課題

（5）で記載のように、ハードウェア・ソフトウェアを統合・全体最適化したコンピューティング技術の研究開発では、分野のレイヤーをまたいだ複数の産学の連携が必要になる。また、日本の産業界の状況に鑑みても、遠い将来の技術・コアとなる事業以外の技術の研究開発に投資を行うことが難しくなっている。いわゆる中央研究所が減ってきていることはその一つの象徴ではないか。そのため、日本企業が将来にわたって競争力を保つためには、外部の研究機関・大学を「基礎研究や異分野融合のための研究開発リソース」として活用することが極めて重要になると考えられる。

（7）国際比較

国・地域	フェーズ	現状	トレンド	各国の状況、評価の際に参考にした根拠など
日本	基礎研究	△	↘	・一部の有力な研究者がトップカンファレンスで発表しているが、アーキテクチャー分野での存在感は低い。
	応用研究・開発	○	→	・スパコン富岳が各種ベンチマークで世界1位を獲得したが、クラウドコンピューティングなどで存在感はなく米国企業に依存
米国	基礎研究	◎	→	・産業界と大学が協力し、アーキテクチャーのトップカンファレンスでの発表件数は首位。コンピューティングの主要な部品ベンダーを擁しており圧倒的。
	応用研究・開発	◎	→	・Amazon、Google、Facebook、Microsoftなどメガデータセンターを運用するプラットフォーマーが主導。最新開発成果は自社で利用した後に発表するほどの開発力を持ち、自社サービスを強化するハード、ソフトを強化。
欧州	基礎研究	○	→	・SAPのファウンダーが創設したHPIや、Tableauに買収されたHyperを開発したミュンヘン工科大など、DB分野では数多くの成果をトップカンファレンスで発表。
	応用研究・開発	○	→	・Horizon EuropeにおいてCloud ComputingのWork Programmeを実行するなど、具体的な応用に向けた研究開発が進められている。
中国	基礎研究	○	↗	・論文数は米国を抜いて首位。半導体のトップ国際会議であるISSCCでも4年間で採択論文が1件から15件に急増。
	応用研究・開発	◎	↗	・Baiduは2018年にAIチップである「Kunlun」、Alibabaは「Ali-NPU」を開発するなどAI分野で攻勢。データセンターの規模でも、Alibabaクラウドは AWS、AZUREに次ぐ第3位のシェア。
韓国	基礎研究	○	→	・Samsungなどメモリーベンダの存在感が大きく、大学でもメモリーにかかわる研究テーマが多い。メモリー技術に関しては、トップカンファレンスで発表する力を持つ。
	応用研究・開発	○	→	・電子政府の普及など、実用化が進んでいる。

（註1）フェーズ

基礎研究：大学・国研などでの基礎研究の範囲

応用研究・開発：技術開発（プロトタイプの開発含む）の範囲

（註2）現状　※日本の現状を基準にした評価ではなく、CRDS の調査・見解による評価

◎：特に顕著な活動・成果が見えている　　○：顕著な活動・成果が見えている

△：顕著な活動・成果が見えていない　　×：特筆すべき活動・成果が見えていない

（註3）トレンド　※ここ1～2年の研究開発水準の変化

↗：上昇傾向、→：現状維持、↘：下降傾向

参考文献

1）Intel News Release July 28, 2015, Intel and Micron Produce Breakthrough Memory Technology, https://newsroom.intel.com/news-releases/intel-and-micron-produce-breakthrough-memory-technology/#gs.gq890Fs.

2）Intel News Release May 30, 2018,“Reimagining the Data Center Memory and Storage Hierarchy”, https://newsroom.intel.com/editorials/re-architecting-data-center- memory-storage-hierarchy/#gs.234zB98.

3）SAP HANA Blog,“Continuous and Customer-Driven Innovation with SAP HANA”,https://blogs.saphana.com/2018/06/05/continuous-and-customer-driven-innovation- with-sap-hana/.

4）DAOS Storage Stack, https://github.com/daos-stack

5）Intel® Optane™ Persistent Memory Product Brief, https://www.intel.co.jp/content/www/jp/ja/products/docs/memory-storage/optane-persistent-memory/optane-dc-persistent-memory-brief.html

6）Wooseong Cheong et al.“A Flash Memory Controller for 15us Ultra-Low-Latency SSD Using High-Speed 3D NAND Flash with 3us Read Time”International Solid-State Circuits Conference（ISSCC）, 2018.

7）Hewlett Packard Enterprise Press release May 16, 2017,“HPE Unveils Computer Built for the Era of Big Data”, https://news.hpe.com/a-new-computer-built-for-the- big-data-era/.

8）IBM　NewsRoom, https://newsroom.ibm.com/2020-08-17-IBM-Reveals-Next-Generation-IBM-POWER10-Processor

9）Apache Hadoop, https://hadoop.apache.org/

10）Takashi Miyoshi, Kazuichi Oe, Jun Tanaka, Tsuyoshi Yamamoto and Hiroyuki Yamashima, New System Architecture for Next-Generation Green Data Centers：Mangrove, FUJITSU SCIENTIFIC & TECHNICAL JOURNAL Vol.48, No.2.

11）IDC Japan, 国内コンバージドシステム市場予測を発表, Jun.2020, https://www.idc.com/getdoc.jsp?containerId=prJPJ46448520

12）Intel Rack Scale Design, https://www.intel.com/content/www/us/en/architecture-and-technology/rack-scale-design-overview.html.

13）Hewlett Packard Enterprise, HPE Synergy, https://www.hpe.com/jp/ja/integrated- systems/

synergy.html.
14）Liquid Composable Infrastructure, https://www.liqid.com/
15）DriveScale Composable Platform,https://drivescale.com/
16）David Reinsel, John Gantz and John Rydning, The Digitization of the World From Edge to Core, IDC White Paper, 2018.
17）AWS outpost, https://aws.amazon.com/outposts/
18）F.T. Hady, A. Foong, B. Veal, and D. Williams,“Platform Storage PerformanceWith 3D XPoint Technology,”Proceedings of the IEEE, vol.105, no.9, pp.1822–1833, 2017.
19）Andy Rudoff, Persistent Memory Programming：The Current State and Future Direction, SNIA Persistent Memory Summit 2019.
20）RPMA：Remote Persistent Memory Access, https://github.com/pmem/rpma
21）Hiroki Ohtsuji, Takuya Okamoto Erika Hayashi and Eiji Yoshida, Low-overhead Remote Persistence for Scalable Low-latency File Systems, ISC High Performance 2020 Digital, 2020
22）Mark Carlson, Persistent Memory what developers need to know, SNIA Storage Developer Conference（SDC）, 2018.
23）CXL：Compute Express Link, https://www.computeexpresslink.org/
24）CCIX, https://www.ccixconsortium.com/
25）Open Compute Project, https://www.opencompute.org/