2.5.4 データ処理基盤

（1）研究開発領域の定義

本領域は、多数の計算機あるいはメニーコアを搭載する等のハイエンドな計算機を利用することで、大規模なデータ（ビッグデータ）に対する処理を効率的に実行する基盤的ソフトウェア技術を確立する領域である。主な要素技術はクラウド環境で利用されている分散並列型の大規模データ処理であり、代表的なデータ処理の例としてデータベースの検索処理・機械学習・データマイニングが挙げられる。

（2）キーワード

クラウドコンピューティング、並列分散データ処理、データベース、機械学習、データマイニング、最適化、分類、クラスタリング

（3）研究開発領域の概要

［本領域の意義］

ビッグデータが急速に増大する一方で、ムーア則の終焉により1CPU当たりの計算量に上限があるため、計算機を並列分散化してデータを処理する技術が必要不可欠である。特に、近年普及しているクラウド環境において、ビッグデータの並列分散処理基盤が活用されている。このような処理基盤の具体例として、分散ファイルシステム、分散データベースシステム、Sparkなどの分散処理基盤、Serverless computing、機械学習のワークフロー全体をサポートする機械学習基盤が挙げられる。これらにおける主な技術課題としては、高速化・効率化によるクラウド環境における処理コスト削減、およびクラウドユーザーによる応用プログラム（機械学習やデータ分析）の開発コストの削減が挙げられる。

本領域の市場規模の観点について述べる。調査会社であるIDC Japanの2020年6月の報告[1)]によれば、国内における2019年のBDA（Big Data and Analytics）テクノロジー／サービス市場は、前年比10.0%増の高い成長率を維持し、市場規模は1兆799億5,100万円となったとされている。また、同市場においても新型コロナウイルス感染症（COVID-19）流行の影響のため2020年、2021年は一時的に成長が鈍化するが、企業のデータ分析需要は継続していて2024年の市場規模は2019の二倍近くの1兆8,765億7,400万円になると予想されている。このようなビッグデータテクノロジー／サービス市場を牽引するのは、大規模データを対象とした機械学習・データマイニング処理などの技術である。また世界のクラウドビジネスの観点では、調査会社のCanalysの2020年7月のレポート[2)]によれば、クラウドビジネスは31%と高い年成長率を示しており、2020年第二四半期の時点で346億ドルの市場規模に達したと報告されている。また市場全体はAmazon AWSが31%、Microsoft Azureが20%、Google GCPが6%、Alibabaが5%、その他が37%を占めると報告されている。

［研究開発の動向］

ビッグデータの並列分散処理基盤の経緯を説明する。古くは1980年代に分散データベースが登場し、1990年のインターネットの普及に伴い検索エンジンのデータ管理が分散化され、2000年代にはGoogleの初期の分散処理基盤である分散ファイルシステムGFS、分散処理システムMapReduce、分散テーブルBigTableが登場し、同時期にweb系の企業を中心としてAmazon Dynamo、Yahoo PNUTSや、Googleに対抗する形でオープンソースプロジェクトとしてHadoopプロジェクトが2006年に登場した。これらのシス

テムでは、分散データベースにおけるSQL処理機能および分散トランザクション処理機能を簡略化することで、1000台規模の廉価サーバーで高スケールな分散処理を実現していた。2010年代には主記憶の大容量化に伴いHadoopの後継としてSparkプロジェクトが2014年に開始され、そのサブシステムとしてSQL処理、Streamingデータ処理、機械学習処理、グラフデータ処理の機能が開発された。

また、業界を牽引するGoogleからは、上記のBigTableでは機能が制限されていたSQL処理と分散トランザクション処理をサポートするF1, Spannerが発表され、また機械学習処理基盤のTensorFlowが2015年に登場した。 Googleのクラウド環境においてはSQL処理が可能なBigQueryが2011年にリリースされ、機械学習に関してはColaboratoryが2018年にリリースされた。 Colaboratoryは、AIプログラムの標準的な開発環境であるJupiter notebookとクラウド環境をシームレスに連携するとともに、機械学習のワークフロー全体を支援する。このように2010年代の1つ目の動向として、機械学習が多くの応用分野に急速に普及した影響を受けて、機械学習の開発環境とクラウド環境の連携が目覚ましく進化を遂げたことが挙げられる。

一方でクラウド環境の進化としては、従来のSaaS、PaaS、IaaSが発展し、仮想計算機のインスタンスを意識することなくクラウド資源を簡単に利用できるサービスであるServerless computingが2015年前後に登場し、FaaS（Function as a service）として現在では多くの企業で利用されるサービスに成長してきている。更には、AmazonのElastic Fabric AdapterのようにHPC（high performance computing）で利用されていた高速ネットワークの機能をクラウド環境で利用できるようになってきている。このように現在のクラウド環境における2つ目の動向としては、用途に応じて多種多様なサービスが提供されていることが挙げられる。その一方で、それぞれのクラウドサービス毎に得意・不得意があるため、ユーザーは開発対象の応用プログラムの特性に応じて、適切なクラウドサービスを選択して利用しなければならないという状況にある。

これらのビッグデータの並列分散処理基盤に関する研究開発の動向に関しては、1）ビッグデータ処理・管理の要素技術に関する研究開発、2）クラウド環境におけるビッグデータ処理基盤に関する研究開発、3）機械学習のワークフローを支援する研究開発、に大別することができる。

ビッグデータ処理・管理の要素技術

ビッグデータを処理・管理する基盤としてデータベース管理システムを中心とした研究開発が挙げられる。データベース管理システムにおけるコア技術としては、クエリーワークロード処理の高速化およびトランザクション処理の高速化が挙げられる。クエリーワークロード処理の高速化に関しては、コストを最小化する最適化問題あるいはインデックス構造を用いた高速化のアプローチがある。いずれのケースでも近年では機械学習などの数理科学の知見を活用する動向が多く見られる。また、機械学習が得意とする予測技術をデータベースに応用してクエリーワークロードの予測をすることで将来にわたってワークロードコストを削減する研究が取り組まれている。一方、トランザクション処理の高速化に関しては、今後普及が見込まれる不揮発性メモリーや1000コアレベルの並列処理を想定したトランザクション処理の高速化に関する研究や、応用プログラムのロジックを活用することで従来の直列化可能性とは異なる方法で一貫性を保証する高スケールなトランザクション処理を実現するための研究が進められている。

クラウド環境におけるビッグデータ処理基盤

上述したビッグデータ処理・管理の要素技術はクラウド環境におけるビッグデータ処理基盤にも適用可能であるが、クラウド環境特有の性質に関して、サーバー数を動的に制御することでクラウド環境のSLO（Service-level objective）に関する性能保証を行う研究（プロビジョニング）、CPUやメモリーなどのハー

ドウェア資源に対する競合の制御（過度の資源利用により性能が急激に劣化することを避ける）、大規模なクエリーワークロードに対する効率的な近似最適化に関する研究などが行われている。また、近年普及しているServerless computingに関する取り組みとしては、現状のServerless computing環境が有する制約条件（メモリー量および処理の実行時間に上限がある、statelessな処理、すなわち内部状態を保持せず、入力によってのみ出力が決定する。同じ入力に対しては、同じ出力が得られるような処理しか実行できない、入出力時のIOコストが大きい）を課題に挙げる研究がおこなわれている。具体的には、現状のServerless computing環境を効率的に活用するアプローチと、現状のServerless computingの課題を解消する新たなServerless computing基盤を考案するというアプローチに分類することができる。

機械学習のワークフロー支援

機械学習のワークフローは、GoogleのColaboratoryやAmazonのSageMakerなどで支援されており、クラウド環境と連携した統合的な開発および運用環境として利用されている。具体的には、教師データの準備、データクリーニング、notebookによるプログラム開発、モデル評価、特徴量選択・ハイパーパラメーターのチューニング、運用時の性能モニタリング等のツール群から構成されている。教師データの準備に関しては、教師データを自動生成する研究が注目されている。データクリーニングに関してはデータの加工や異常値の除去などを自動化する研究が取り組まれている。特徴量選択・ハイパーパラメーターのチューニングについてはAutoMLやベイズ最適化が代表的である。更には、大規模データ処理と機械学習処理を独立に処理するのではなく、これらを横断して分析処理工程全体を最適化する研究や、エッジコンピューティングによってエッジ側でも機械学習を行う研究が行われている。

（4）注目動向

［新展開・技術トピックス］

ビッグデータ処理・管理の要素技術

画像認識や自然言語処理の分野と同様に、データベース、あるいはビッグデータ処理・管理の領域においても、数理科学の知見（線形計画問題、深層学習、強化学習等）を活用したコスト最適化技術の研究が盛んにおこなわれている。具体的には、クエリーコストの予測問題、最適なスキーマ設計の問題、最適なクエリーの実行プラン選択、インデックスなどの最適なデータ構造の自動設計などに対して、機械学習の技術が適用されてきており、AIとシステムとデータ処理分野に横断の国際会議（MLSys：2018年～）やワークショップ（AIDB：2019年～、aiDM：2018年～、DISPA：2020年～）が開催されて注目を集めている。

クラウド環境におけるビッグデータ処理基盤

クラウド環境におけるビッグデータの並列分散処理に関しては、市場ニーズが高いと同時に技術的に難易度が高いため今後の更なる技術開発が必要である。技術的な難しさの根源は、1）大量かつ多様なユーザーは異なるリクエスト（クエリー）をビッグデータ処理基盤で実行すること、2）ユーザーのリクエストは時間と共に変化すること、3）CPUやメモリーなどのハードウェア資源に対する競合あるいはデータ資源に対する競合によってリクエストの処理性能が極端に低下すること、等が挙げられる。1）に関してはスケーラブルなクエリー処理の共有化技術（インデックス化や部分クエリーの実体化）、2）に関してはユーザーリクエストの変化に追従するワークロード処理の最適化技術、3）に関してはデータ資源に対する競合の予測と回避に関する研究が取り組まれている。

Serverless computingに関しては、現状のServerless computingの制約として1）実行時間に上限が決められていること、2）関数の入出力時のIOコストが大きいこと、3）関数間でダイレクトな通信ができないこと、4）GPUなどの特別なハードウェアが利用できないこと、が指摘されていて[3]、これらの課題を克服する新たなServerless computing基盤の研究がなされている。

機械学習のワークフロー支援

機械学習の適用を妨げている大きな問題の1つに教師データの準備がある。この問題に対して、教師データを自動生成して機械学習を促進する取り組みや、Jupiter notebookにより利用されているデータと分析処理をパターン化し、分析対象となるデータを推薦する研究がなされている。また、データ分析においてはデータクリーニングが作業の大半の時間を占めることが知られており、データクリーニングの自動化の研究が取り組まれている。具体的には、データクリーニング操作に基づいて自動的にプログラムを合成する研究や、データに含まれる異常値を自動的に特定して除去する研究が取り組まれている。更に、機械学習工程における有意な特徴量を選択することで事前データ処理において不要な特徴量に関する処理を除去する研究など、機械学習の工程と事前のデータ処理の工程を横断する最適化の研究が盛んである。また、エッジコンピューティングと融合してエッジ側で機械学習の一部を実施する研究（学習モデルの小型化、FPGAによる実装、クラウド側との連携アーキテクチャー）の研究が取り組まれている。

［注目すべき国内外のプロジェクト］

ビッグデータ処理・管理の要素技術

数理科学の知見（線形計画問題、深層学習、強化学習等）を活用したビッグデータ処理・管理の高速化は多数の大学で研究がなされている。代表的なプロジェクトしては、マサチューセッツ工科大学とGoogleが共同で取り組んでいるSageDBプロジェクトがあり、学習型のデータ構造（Btree、Bloom filter、多次元インデックス）、学習型のクエリー実行計画（ソート処理、スケジューリング）、クエリー最適化（クエリー結果件数の予測）に渡って体系的にデータ処理を学習型に置き換える取り組みがある[4), 5)]。カリフォルニア大学バークレー校のRISELabでは、自己教師あり学習を適用して、クエリー結果サイズの予測およびJoinクエリーの最適化の課題に取り組んでいる[6), 7)]。

高速トランザクション処理に関しては、ミュンヘン工科大学、スイス連邦工科大学、Oracleなどが代表的な研究プロジェクトを実施している。ミュンヘン工科大学では、これまで主記憶データベースエンジンであるHyperを起業後（Tableauに買収された）、現在はフラッシュメモリーと主記憶を連携して高速にトランザクション処理可能なUmbraを研究開発している[8), 9)]。

クラウド環境におけるビッグデータ処理基盤

この領域における代表的なプロジェクトとして、マイクロソフトではBing、Office、Windows、Skype、Xboxなどの各種サービスのデータ分析を行っており、この大規模なワークロードを最適化するプラットフォームとしてPeregrineの研究開発を行っている。本プロジェクトでは、上記の実ワークロードの分析クエリーの傾向に基づいて、学習型のクエリー最適化、ハードウェアリソース量に応じたクエリープランニング、大量クエリーのキャッシュや同時実行処理、自動性能チューニングの技術に関して取り組んでいて、多数の研究発表がなされている[10)]。また、シカゴ大のCrocodileDBプロジェクトでは[11)]、データの更新とクエリーの将来予測を行い、この予測に基づいてクエリー処理をpushベースで処理する（事前にクエリーを実行した結果を生

成して漸進的更新する）か、あるいはpullベースで処理する（クエリーを随時実行する）かを決定するデータベースエンジンの研究を行っている。

Serverless computingに関する代表的なプロジェクトの1つとしてカリフォルニア大学バークレー校のHydroプロジェクト[12)]では、レプリカ間の逐次的な更新同期を必要とせずに分散処理の一貫性を保証する高スケールなKey value store（分散データベースの一種）のAnna、および高スケール性を保ちながらstatefulな操作、すなわちシステムが内部状態を保持し、同じ入力でもそれまでの実行状況によっては違う出力を与えるような処理を実現したFaaS基盤であるCloudburstの研究開発を行っている。

機械学習のワークフロー支援

昨今の機械学習の開発環境はワークフロー全体を支援するツール群から構成されており、ワークフローの各ステップ毎に研究開発が進められている。弱教師データを自動生成する研究プロジェクトとしてSnokelが挙げられる。 Snokelプロジェクト[13)]で取り組まれている技術として、ドメイン知識を用いてルールベースで弱教師データを自動生成する機能、学習ベースのデータオーギュメンテーション機能（既存のデータに基づいてデータを変換して教師データを増やす）、学習が進みにくい教師データのサブセットを特定する機能がある。データクリーニングの操作履歴から自動的にクリーニングのプログラムを生成する研究については、スタンフォード大のDataWranglerプロジェクト[14)]やマイクロソフト社でのプログラム合成の研究プロジェクト[15)]が挙げられる。前者はTrifactaの社名で起業に成功しており、後者はExcelの機能として商用化されている。特徴量選択・ハイパーパラメーターのチューニングに関しては、GoogleおよびOracleはAutoMLを採用している。分散アプリケーションを対象にしたプロジェクトとしてはカリフォルニア大学バークレー校のRayプロジェクト[16)]が、クラウド環境と連携した分散環境での高スケールな強化学習機能とハイパーパラメーターのチューニング機能の研究を行っている。

国内の研究プロジェクトしては、Preferred Networksが提供しているハイパーパラメーターのチューニングツールであるOputuna[17)]は広く利用されており、東京工業大学の篠田浩一教授が研究代表であるJST CREST「社会インフラ映像処理のための高速・省資源深層学習アルゴリズム基盤」[18)]ではアルゴリズムとシステムアーキテクチャーを同時に最適化する枠組みの構築に取り組んでいる。また、産業技術総合研究所は我が国における産学共同のAI研究開発を加速するプラットフォームであるAI橋渡しクラウド（AI Bridging Cloud Infrastructure, ABCI）を運用している。 ABCIは世界トップレベルの計算処理能力とデータ処理能力を持ち、AIとビッグデータのアルゴリズム、ソフトウェア、応用開発に用いられている。2020年11月には理化学研究所の「富嶽」とともに、大規模機械学習処理のベンチマークであるMLPerfHPCにおいて世界最高レベルの性能を達成した[19)]。

（5）科学技術的課題

ビッグデータ処理・管理の要素技術

機械学習などの数理科学の知見を活用した取り組みは学習に時間を要するため、データベースの更新に応じた学習モデルの最新化が難しいという課題がある。このため、現状では研究されている技術の応用可能な範囲が限定されてしまっている。今後は学習そのものの高速化が重要な課題になると考えられる。また、最新ハードウェアを利用した研究については、ハードウェアの成熟およびコンパイラーを含む開発環境の進化に伴い、ビッグデータ処理・管理での活用が更に促進するものと考えられる。応用プログラムのロジックを活用することで従来の直列化可能性とは異なる方法で一貫性を保証する研究は[20)]、過去30年に渡るトランザク

ション処理の概念を変えるものであり、現在は未成熟ではあるが大きな可能性を秘めている。

クラウド環境におけるビッグデータ処理基盤

この領域でもコスト最小化問題を機械学習によって解くというアプローチがなされているが、特にハードウェアなどの資源競合によって引き起こされる急激な処理性能低下によって、クラウドサービスにおけるSLO保証が困難である問題は大きな技術課題として残っている。ハードウェアリソースを動的に制御することでリアルタイム性を保証する技術は必要不可欠であり、今後の研究の進化が期待される。

機械学習のワークフロー支援

ビッグデータを処理する観点あるいはデータ管理の側面から機械学習を捉えた場合、モデルの再利用の促進が大きな課題になると考えられる。ベンチマークデータとしての教師データはアーカイブ等で共有が進んでいるが、学習モデルの共有およびモデルをアンサンブルして統合的に再利用する取り組みが重要な課題であると考えられる。具体的には、共有化された膨大な学習モデル集合の中から適用先に適した学習モデルを選別するモデル検索技術、検索した学習モデルを転用するための転移学習の技術、転移学習後の再学習において破壊的忘却を避ける技術などが重要な課題であると考えられる。

（6）その他の課題

クラウドベンダーによるロックインが大きな課題の1つとして挙げられる。上述の通りコロナの影響下にあってもクラウド業界は大きく市場規模を拡大しており、Amazon、Microsoft、Google各社ともAIを中心とした開発環境の充実化を急速に進めている。現在、誰でも簡単に機械学習の応用プログラムを開発できて、クラウドの膨大な計算資源を簡単に利用可能であるため膨大な処理を短時間で実行することも簡単にできる環境が提供されている。その一方で、クラウド環境は各社ごとに異なるインターフェースになってしまっているため、利用者は開発したプログラムを他のクラウド環境やオンプレミス環境に移行することが困難な状況になってしまっている。特に大規模データを一度特定のクラウド環境に置いてしまうと、データの送信に膨大な時間を要するため他のクラウド環境に移行することは困難である。今後の方策として、複数のクラウド環境を自由に切り替えできるよう、クラウド環境におけるインターフェースの標準化、あるいは複数のクラウド環境を統合的に利用できるインタークラウドをサポートする開発環境の研究開発が求められると考えられる。

このような課題に関しては、JST CREST「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」の合田チームらによって開発されたVirtual Cloud Provider（VCP）基盤ミドルウェアが、クラウドプロバイダ毎に異なるサービスを抽象化し、クラウド上の計算資源の選択と確保、ネットワーク設定、ソフトウェアの配備等の処理をユーザーに替わって自動的に実行することを提供しており今後の普及が期待される[21)]。また、学習したモデルをクラウド間で移行可能なモデル変換の機能も必要であると考えられる。

（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	○	→	・機械学習やデータマイニング系の最難関会議では、日本の大学および企業からコンスタントに発表されている（KDD2020では日本からの投稿は世界4位）。データベースやシステム系の最難関会議でもVLDB2020では日本からの投稿は世界5位と増加している。
	応用研究・開発	○	→	・クラウド環境の浸透に伴い、多くの大企業およびスタートアップ企業がAIを活用したサービスを開始している。スタートアップ企業としては、Preferred networks、ティアフォー、Yahoo、サイバーエージェントなどが応用研究で目立っている。
米国	基礎研究	◎	→	・米国の大学・企業における基礎研究レベルは高く、データベース系および機械学習系の両面で世界をリードしており、特にデータベース分野での最難会議の約6割弱は米国からの発表である。
	応用研究・開発	◎	→	・分散処理基盤技術の取り組みについては、Web系の企業がOSS開発を先導しており米国が圧倒している。 OSS開発コミュニティーは大学との連携が強い。大学発のスタートアップが企業に買収される例[22)]や、製品がクラウドで利用できる例などがある（Trifacta社の機能はGoogleのクラウド環境で利用可能）。
欧州	基礎研究	◎	→	・不揮発性メモリー・メニーコア・高速ネットワーク・GPUやFPGAなどの最新ハードウェアを活用したDBMSの研究開発が強い（ミュンヘン工科大、スイス連邦工科大学）。
	応用研究・開発	○	→	・SAPのHANAなどのカラム指向で主記憶型のDBMSやストリームデータ処理エンジンの取り組みが目立っている。
中国	基礎研究	◎	→	・大学が中心となって基礎研究において多く成果を挙げている。難関国際会議への採録も米国・欧州に次いで3番目に多い。アルゴリズム系の研究に加えてクラウド系の研究も急速に進んできている。 ・世界で活躍する研究者の採用が目立ち、また現地の優秀な学生を囲い込むためMicrosoft研究所などの外資研究機関の現地法人化が進んでいる。
	応用研究・開発	◎	→	・Alibabaが商業的に成功しており、Eコマース業界からクラウド業界へと進出を果たしている（世界4位のシェア）。 ・中国全体としてスタートアップ企業が好調である。
韓国	基礎研究	○	→	・最新ハードウェアを利用したDBMSの高速化などの高速DBMSの取り組みや、グラフエンジンの取り組みなどが目立っている。
	応用研究・開発	△	→	・SAPは韓国に支店を構え、韓国の大学と共同研究を行い、SAP HANA DBMSの高速化に取り組んでいる。

（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）国内BDAテクノロジー／サービス市場予測を発表, IDC, 10 Jun 2020. https://www.idc.com/getdoc.jsp?containerId=prJPJ46435220
2）https://www.canalys.com/newsroom/worldwide-cloud-infrastructure-services-Q2-2020
3）Joseph M. Hellerstein, Jose M. Faleiro, Joseph Gonzalez, Johann Schleier-Smith, Vikram Sreekanti, Alexey Tumanov, Chenggang Wu：Serverless Computing：One Step Forward, Two Steps Back, CIDR 2019.
4）SageDB：A Self-Assembling Database System, http://dsail.csail.mit.edu/index.php/projects/
5）Tim Kraska, Mohammad Alizadeh, Alex Beutel, Ed H. Chi, Ani Kristo, Guillaume Leclerc, Samuel Madden, Hongzi Mao, Vikram Nathan：SageDB：A Learned Database System, CIDR 2019.
6）Zongheng Yang：Self-supervised learning for cardinality estimation, AIDB 2020.
7）Zongheng Yang, Amog Kamsetty, Sifei Luan, Eric Liang, Yan Duan, Xi Chen, Ion Stoica：NeuroCard：One Cardinality Estimator for All Tables, PVLDB 2021.
8）UMBRA Flash-Based Storage + In-Memory Performance, https://umbra-db.com/
9）Thomas Neumann, Michael J. Freitag：Umbra：A Disk-Based System with In-Memory Performance, CIDR 2020.
10）Alekh Jindal, Hiren Patel, Abhishek Roy, Shi Qiao, Zhicheng Yin, Rathijit Sen, Subru Krishnan：Peregrine：Workload Optimization for Cloud Query Engines, SoCC 2019.
11）Zechao Shang, Xi Liang, Dixin Tang, Cong Ding, Aaron J. Elmore, Sanjay Krishnan, Michael J. Franklin：CrocodileDB：Efficient Database Execution through Intelligent Deferment. CIDR 2020.
12）The Hydro project, https://hydro-project.github.io/
13）Snorkel, Programmatically Build Training Data, https://www.snorkel.org/
14）DataWrangler, http://vis.stanford.edu/wrangler/
15）Sumit Gulwani, Programming by Input-Output Examples, ECML 2019.
16）RAY, Fast and Simple Distributed Computing, https://ray.io/
17）Optuna, Optimize your optimiaiton, https://optuna.org/
18）【篠田　浩一】社会インフラ映像処理のための高速・省資源深層学習アルゴリズム基盤, https://www.jst.go.jp/kisoken/crest/project/1111094/1111094_07.html
19）MLPerf , “MLPerf Training v0.7 Results”, https://mlperf.org/training-results-0-7
20）Coodination Avoidance, Encyclopedia of Big Data Technologies, 2019.
21）JST, ”インタークラウドを活用したアプリケーション中心型オーバーレイクラウド技術に関する研究”, https://www.jst.go.jp/kisoken/crest/project/45/15655350.html
22）Informatica Acquires GreenBay Technologies to Advance AI and Machine Learning Capabilities, https://www.informatica.com/about-us/news/news-releases/2020/08/20200818-informatica-acquires-greenbay-technologies-to-advance-ai-and-machine-learning-capabilities.html