



国立大学法人北海道大学情報基盤センター
 国立大学法人東北大学サイバーサイエンスセンター
 国立大学法人筑波大学人工知能科学センター
 国立大学法人東京大学情報基盤センター
 大学共同利用機関法人情報・システム研究機構国立情報学研究所
 国立大学法人東京工業大学学術国際情報センター
 国立大学法人東海国立大学機構名古屋大学情報基盤センター
 国立大学法人京都大学学術情報メディアセンター
 国立大学法人大阪大学サイバーメディアセンター
 国立大学法人九州大学情報基盤研究開発センター

2021年3月9日

データ活用社会創成プラットフォーム mdx を導入

—9大学2研究機関が共同運営しデータ活用の産学官連携・社会実装・研究を推進—

ポイント

- ◆9大学2研究所が連合して共同運営する、データ活用にフォーカスした高性能仮想化環境
- ◆データの入出力等の様々な目的に使用できる汎用ノード群と、機械学習等の演算を高速に処理できる演算加速ノード群を有し、Society 5.0（注1）実現に貢献する
- ◆インテルディープラーニングブーストテクノロジーを有するインテル社製「第3世代 Xeon スケーラブルプロセッサ（開発コード名 Ice Lake）」816基、NVIDIA社の最新GPUである「NVIDIA A100 Tensor コア」を320基搭載

概要

国立大学法人北海道大学情報基盤センター（センター長：棟朝雅晴）、国立大学法人東北大学サイバーサイエンスセンター（センター長：菅沼拓夫）、国立大学法人筑波大

学人工知能科学センター（センター長：櫻井鉄也）、国立大学法人東京大学情報基盤センター（センター長：田浦健次朗）、大学共同利用機関法人情報・システム研究機構国立情報学研究所（所長：喜連川優）、国立大学法人東京工業大学学術国際情報センター（センター長：伊東利哉）、国立大学法人東海国立大学機構名古屋大学情報基盤センター（センター長：森健策）、国立大学法人京都大学学術情報メディアセンター（センター長：岡部寿男）、国立大学法人大阪大学サイバーメディアセンター（センター長：下條真司）、国立大学法人九州大学情報基盤研究開発センター（センター長：小野謙二）は、国立研究開発法人産業技術総合研究所とともにデータ活用に関する研究、産学官連携、社会実装の全国での展開を支援するためのプラットフォームであるデータ活用社会創成プラットフォーム「mdx」を、2021年3月8日に千葉県柏市の東京大学柏Ⅱキャンパスに導入しました。

mdx は、高性能な計算機と大容量のストレージを備え、国立情報学研究所が運用する学術情報ネットワーク SINET5（2022年度から次期システムに更新予定）と連携することで、広域からのデータ収集機能と、データ集積・処理機能を、企業や自治体との共同研究も含めた全国の大学・公的研究機関が関与する様々なデータ活用の取組に提供し、さらにはデータ活用のコミュニティを形成して分野・セクタを横断した連携を触媒するハブとなることを目指します。

mdx は、仮想化技術を用いて利用者に広域網とストレージ、計算機等からなる IT 環境を提供します。利用者は、mdx と SINET を用いて広域でデータを収集・集積・解析する情報基盤を容易に構築し、あたかも専用の情報基盤のように使用できます。

また、簡単にデータ分析や共有などの利活用を行えるインタフェースやサービスも提供していく予定です。これによりブラウザだけでも高度なデータ解析を行えるようになります。情報技術に詳しくない利用者も、容易に大量のデータの高度な分析が可能になり、既存の情報学の枠を超え、社会課題に対応してデータを活用する新たな応用と研究領域の創生につながります。

内容

1. 背景と目的

サイバー空間とフィジカル空間を高度に融合させた、経済発展と社会的課題の解決を両立する人間中心の社会 Society 5.0 の実現には、サイバー空間とフィジカル空間の間でのリアルタイムで安定かつ安全なデータのやり取りと、サイバー空間を支える高性能な情報インフラが必要です。このようなインフラを、規模の大小を問わず企業や自治体、団体等が容易に利用できるようにすることが、だれもがデータ利活用の恩恵を享受できるインクルーシブな社会の実現につながります。

また、データを最大限に利活用するためには、複数の分野、複数のセクタの連携が重要です。しかし、従来は分野ごとにデータを共有・活用するプラットフォームを構築す

ることが多く、分野を超えたデータ活用につなげるのは困難でした。mdx は分野共通のプラットフォームであり、分野ごとの知見や文化に加えて分野を超えたデータ科学や情報科学の知見、様々なデータやソフトウェアなどを迅速かつ効率的に連携させたデータ活用の実現を目的としています。

mdx は、仮想化技術を用いて、他と分離されたネットワーク、計算機、ストレージからなる情報処理環境（テナント）を複数のプロジェクトに提供します。広域ネットワークと連携し、利用者のリクエストに応じて短期間で広域ネットワーク、計算機、ストレージなどから構成される広域にまたがるテナントをプロジェクトに割り当てます。利用する個々のプロジェクトは、テナントをあたかもそのプロジェクト専用のインフラが整備されたかのように使うことができます。テナントはハードウェアの構成を変更することなく速やかに構築できるため、様々なデータ活用用途で、高性能なインフラを用いた速やかな PoC（Proof of Concept）が可能となります。

mdx は、国立情報学研究所が運営する日本全国の大学・研究機関等が利用する学術専用の情報通信ネットワーク SINET と 100 Gbps 以上の帯域で接続し、SINET を介して SINET 広域データ収集基盤（モバイル SINET）とも接続され、センサや外部データアーカイブなどとの間でリアルタイムデータを広帯域でセキュアなネットワークを介して入出力できます。テナントは常に利用できるため、スパコンで一般的なバッチ処理環境とは異なりいつでもデータを入出力し、処理することができます。多数のセンサからのリアルタイムデータを受け取る必要がある IoT や、観測データの即時活用などの用途での利用が想定されます。

また、東京大学柏 II キャンパス内の、産業技術総合研究所「AI 橋渡しクラウド:ABCI」や東京大学情報基盤センターの次期スパコンシステム「Wisteria/BDEC-01 システム」とは直接接続され、データを共有しながらバッチ処理による効率的で高性能な計算処理を行うことも可能です。

セキュリティの面では、個人情報等の非常に高いセキュリティレベルが要求される用途での使用も想定し、通常の仮想環境よりもより強固なセキュリティを提供するテナント間の「強い分離」機能を提供します。

2. 詳細

新規導入される mdx システムは、高性能 CPU を備えた汎用ノード群と、CPU に加えて高性能 GPU を備えた演算加速ノード群を有します。ストレージは、高速内部ストレージ、大容量内部ストレージ、共有ストレージからなり、内部ストレージはテナントの一部として用います。富士通株式会社がシステムの構築を行いました。

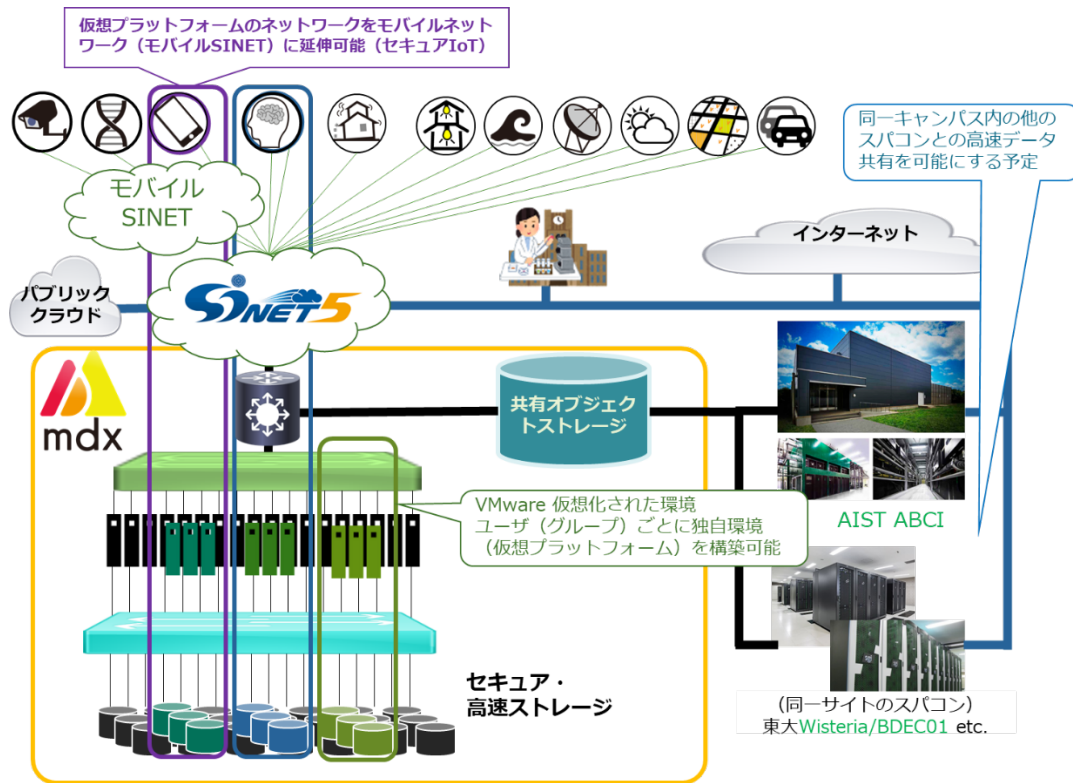


図1 mdx の概要

汎用ノード群は「FUJITSU Server PRIMERGY CX2550」の次期モデル 368 ノードで構成され、ピーク性能 2.1 ペタフロップス、総主記憶容量 94 テビバイトです (注 2)。演算加速ノード群は「FUJITSU Server PRIMERGY GX2570」の次期モデル 40 ノードで構成され、ピーク性能 6.4 ペタフロップス、総主記憶容量 10 テビバイトです。また、合計 27 ペタバイトのストレージシステムを備えています。

利用者が mdx を使用する際には、必要なネットワーク、計算、ストレージ資源を要求します。要求に応じて mdx 上にテナントが構築されます。利用者からみるとテナントは自分専用の環境に見えます。

テナントは SINET に延長可能で、例えば SINET とつながっている研究機関であれば、研究室等の計算機と mdx 上の計算・ストレージ資源がインターネットを介さずに直結した環境を構築できます。

2.1. mdx と SINET

学術情報ネットワーク SINET は、日本全国の大学、研究機関等の学術情報基盤として、国立情報学研究所 (NII) が構築、運用している情報通信ネットワークです。SINET では様々なサービスが提供されており、その中に L2/L3 VPN サービスや、VPN ベースのモバイルサービスがあります。これらの SINET が提供する VPN を用いると、大学や研究機

関の実験装置や計算機、モバイルでつながる広域に分散したセンサなどを mdx 上に構築するテナントの一部とすることができます。テナント内の通信はインターネットと隔離されているので、センサが第三者によって攻撃されるなどのインシデントを防ぐことができます。また、SINET は、全都道府県を 100 Gbps 以上の帯域でつなぐなど、非常に広帯域なネットワークで、VPN では通信帯域を確保する QoS サービスも利用できます。このため、mdx のテナントは国内のほとんどの地域で安全で安定した通信を享受できます。また、SINET は、米国、欧州、アジアにも 100Gbps の接続を有しており、これを利用して mdx を用いた国際共同研究等が加速することも期待できます。

SINET は、2022 年 4 月に現在の SINET5 から SINET6 に移行する予定です。SINET6 では、ほとんどの都道府県が 400 Gbps 以上でつながり、さらに SINET につなぐことができる接続点も増えるため、より一層高性能なテナントの構築が容易になることが期待できます。

2.2. 高いレベルのセキュリティ

mdx 上のテナントでは医療データや個人データなどの機微な情報を扱うことも想定されます。そこで mdx では二段階の分離機能を提供しています。通常分離では、従来のクラウドと同様に、ソフトウェア機能による分離（注 3）が提供されます。「強い分離」では、ソフトウェアによる分離に加えて、ネットワークの設定によっても分離されます。

強い分離のテナントが使用する計算機はネットワーク上でも他の計算機から通信できないように設定されます。また、ストレージも同様に分離されるため、分離に使われるソフトウェアに未知の脆弱性がある場合でも、テナントの分離が破られることはありません。

3. 運用

mdx は 2021 年 3 月 9 日に稼働を開始し、2022 年度より正式運用を行う予定です。全国の大学、公的研究機関およびこれらの機関との共同研究を行う自治体、企業、団体等の利用を見込んでいます。運用は 9 大学 2 研究機関が共同で行います。利用に関するご相談もこれらの大学・研究機関で受け付ける予定です。

4. 問合せ先

東京大学情報システム部情報戦略課総務チーム（情報基盤センター事務担当）

TEL : 03-5841-2710 E-mail : itc-press[at]itc.u-tokyo.ac.jp

5. 用語解説

(注1) Society 5.0

Society 5.0とは、サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させた、経済発展と社会的課題の解決を両立する人間中心の社会（Society）です。狩猟社会（Society 1.0）、農耕社会（Society 2.0）、工業社会（Society 3.0）、情報社会（Society 4.0）に続く、新たな社会を指すもので、第5期科学技術基本計画において我が国が目指すべき未来社会の姿として提唱されました。Society 5.0の実現にはIoT（Internet of Things）、ロボット、AI（人工知能）、ビッグデータといった社会の在り方に影響を及ぼすデジタル革新・イノベーションが不可欠です。

https://www8.cao.go.jp/cstp/society5_0/

(注2) フロップス（FLOPS）、テビバイト（TiB）

計算機の処理性能の指標としてFLOPS（Floating-point Operations Per Second）、すなわち1秒間に実行可能な浮動小数点数演算回数（実数演算回数）が用いられます。PFLOPS（Peta FLOPS）= 10^{15} FLOPSです。また、テビバイトとは記憶容量を表す単位です。1テビバイトは2の40乗バイトです。

(注3) ソフトウェア機能による分離

ここでは、サーバ仮想化を支えるソフトウェアであるハイパーバイザによってVM（仮想マシン）相互間の通信を制限することで実現する分離を指しています。