

jh230066

空閑洋平（東京大学情報基盤センター）

次世代学術情報基盤に向けた基盤ソフトウェアの実践的な研究・開発・評価

課題分野（2）「データ科学・データ利活用分野」、一般共同研究課題（継続）、カテゴリA（HPCI資源を利用する研究課題）

研究代表者：空閑 洋平（東京大学 情報基盤センター）

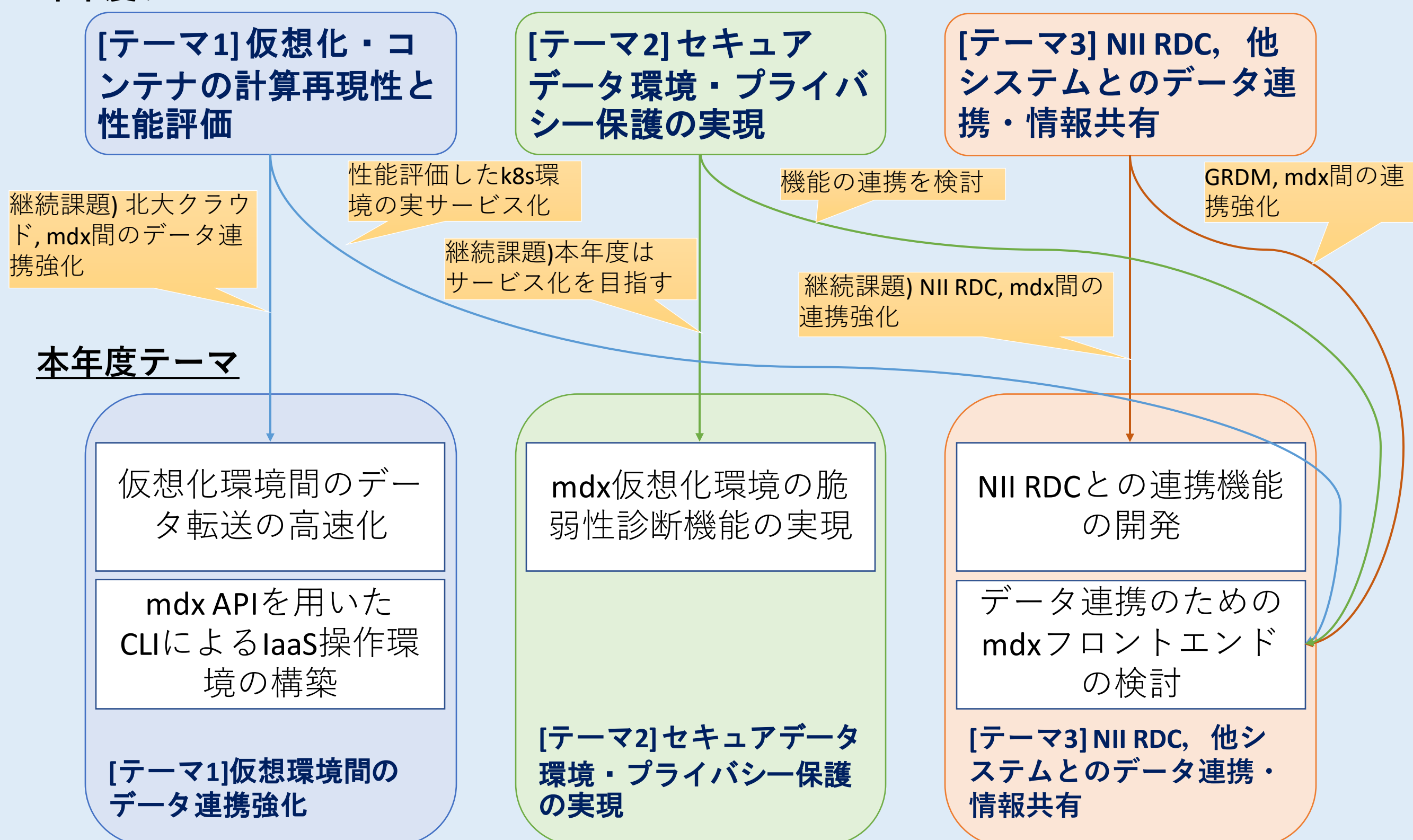
副代表者：中村 遼（東京大学 情報基盤センター）

共同研究者：杉木 章義（北海道大学 情報基盤センター）、田浦 健次郎、埴 敏博、宮本 大輔（東京大学 情報基盤センター）、合田 憲人、竹房 あつ子、藤原 一毅（国立情報学研究所 アーキテクチャ科学研究系）、大江 和一（国立情報学研究所 クラウド基盤研究開発センター）

研究目的

2021年度から本格稼働したNII Research Data Cloud (RDC) と高性能仮想化基盤mdx、情報基盤センターやHPCIの大型計算機群との連携を深化させ、次世代の計算科学・データ科学・データ駆動科学・データ利活用分野を支える学術情報基盤に向けた**基盤ソフトウェアの研究・開発・評価**を行う。以下の3点を中心に、学術基盤の実環境へのデプロイ、活用を目指した実践的な研究開発を行う。

昨年度テーマ



・[テーマ1] 仮想環境間のデータ連携強化

北大クラウドやmdx 環境を利用し、クラスタ・クラウド間のデータ連携の高速化を目指す。また、新規課題として mdx のIaaS 機能呼び出しをプログラマブルにするために、mdx 機能操作が可能な CLI コマンドの開発を実施する

・[テーマ2] セキュアデータ環境の構築とプライバシー保護 の実現

昨年度検討を開始した mdx の脆弱性診断機能のサービス化を目指す

・[テーマ3] NII RDC, 他システムとのデータ連携・情報共有

NII RDC および NII GRDM とmdx 間連携強化のために、mdx に Web フロントエンド機能の追加を検討する

研究の意義

データ科学の計算基盤に対する要求

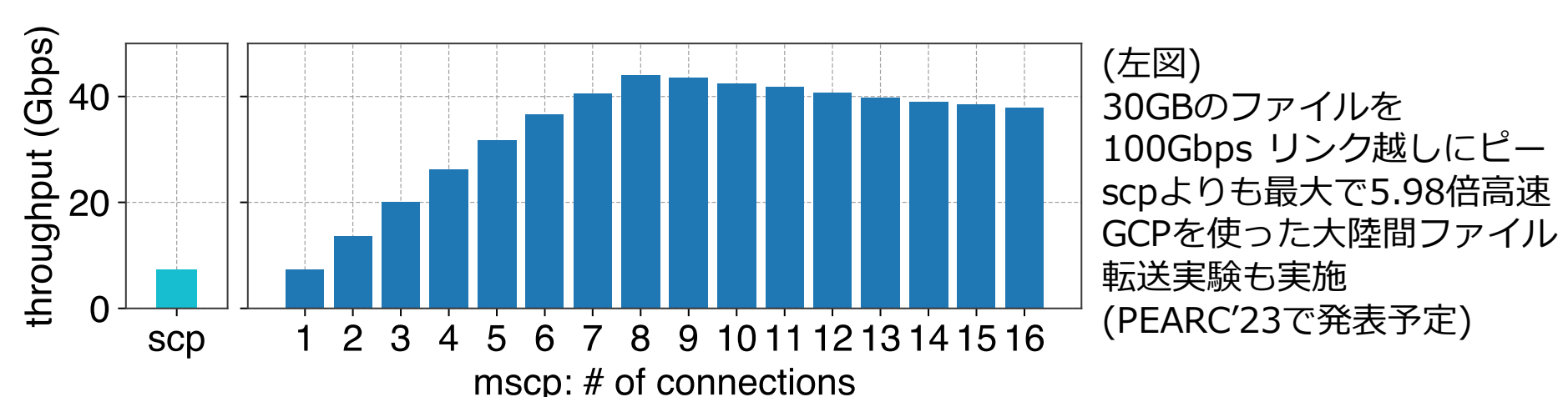
- データの収集, 蓄積, 管理, 活用のための基盤としての需要
- データのセキュリティを確保することの重要性
- 様々なアプリケーションの形態

IaaSクラウドとしての機能・拡張性・セキュリティ **×** HPC分野の研究基盤

仮想環境間の高速かつ手軽なデータ転送

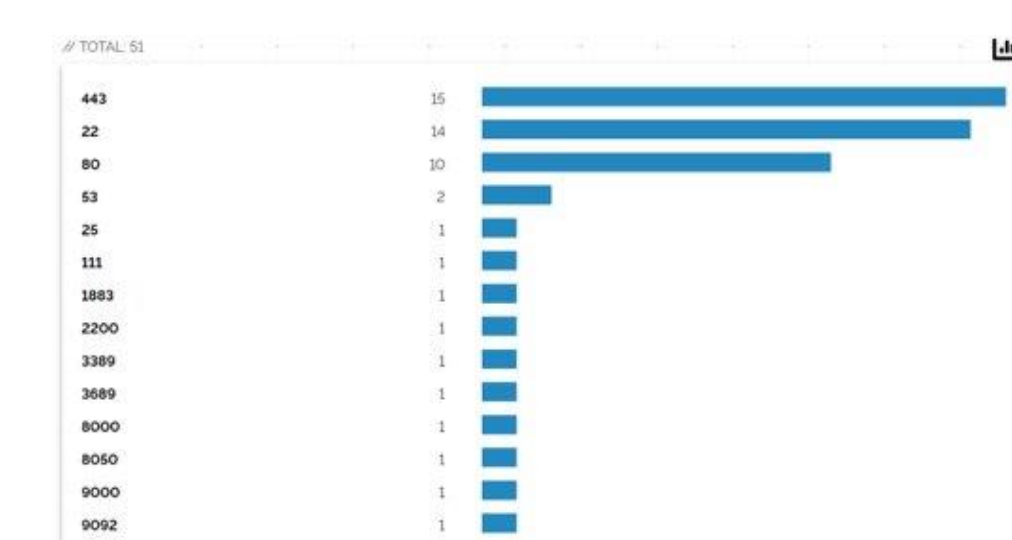
- mdxのようなIaaSクラウドで高速ファイル転送ツールを使うにはユーザがツールを入手し、インストールし、設定して、時にはチューニングする必要がある手間がかかる。
- mscp: 通常のsshサーバへ複数のSSHコネクションをはって高速なファイル転送を実現する新しいファイル転送ツール

```
$ scp /var/ram/30g.img 10.0.0.1:/var/ram/30g.img 100% 30GB 845.0MB/s 00:36
$ mscp /var/ram/30g.img 10.0.0.1:/var/ram/30g.img 100% 30GB/30GB 3.9GB/s 00:08
```

OSSとして公開: <https://github.com/upa/mscp>

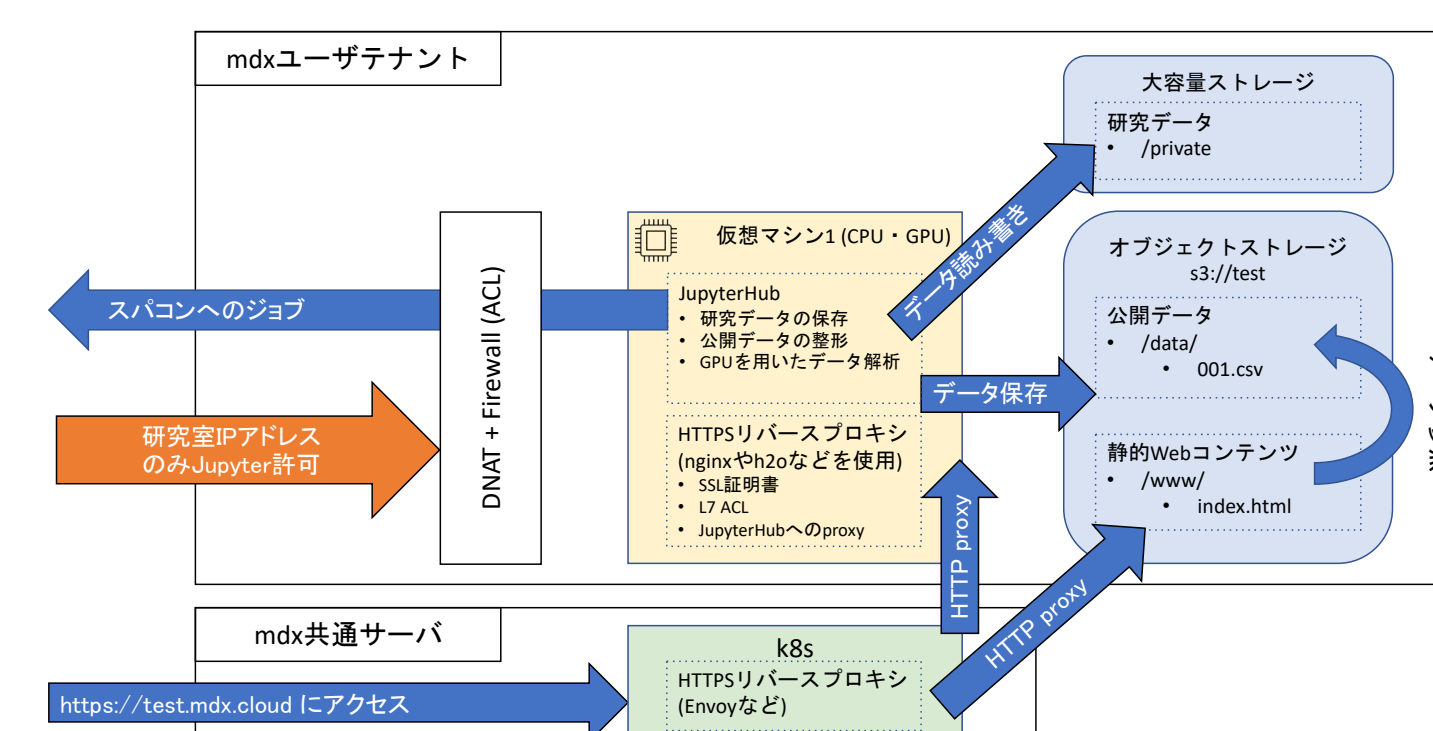
セキュアデータ環境の構築

- mdx脆弱性検査の実施について mdx-CSIRTを設置し、データ活用社会創成プラットフォーム共同研究基盤のテナントの状況調査として、脆弱性検査を定期的に行う
- 検査対象 2023年度中にmdxにおいてDNATによるグローバルIPアドレスを利用している全てのVMに対して脆弱性検査を行う
- 検査及び通知について 脆弱性検査には Greenbone Vulnerability Management を用い、既知の脆弱性情報に基づいた診断を行う



RDC/GRDMとのデータ連携 / mdxフロントエンド

- RDC-mdx間の連携強化 NII RDC と mdx 間、および mdx と基盤センタースパコンとの連携、およびそれらの活用事例を検討
- mdxfrontの導入を検討 mdx 上でのオープンデータ公開基盤の構築を想定し、構築を容易にするための mdx の拡張機能を検討する。mdx側でWebサーバ、ドメイン名、SSL 証明書の構築・運用することで、データ公開基盤の構築を支援



最近発表した成果

- [1] Ryo Nakamura and Yohei Kuga, "Multi-threaded scp: Easy and Fast File Transfer over SSH", ACM PEARC, 2023. (SCA併設Data mover Challengeに応募予定)
- [2] 中村 遼, 空閑 洋平, "複数コネクションを用いる高速なscpの実装", IPSJ 研究報告システムソフトウェアとオペレーティング・システム, 2023.
- [3] 杉木 章義, 他, "データ活用社会創成に向けた基盤ソフトウェア環境の構築", 学術情報処理研究, 2022. (招待)
- [4] 空閑洋平, 中村遼, "遠隔会議システムの計測データを用いた広域ネットワーク品質計測", IPSJ IOTSシンポジウム, 2022. (藤村記念ベストプラクティス賞受賞)
- [5] Toyotaro Suzumura, Akiyoshi Sugiki et al., "mdx: A Cloud Platform for Supporting Data Science and Cross-Disciplinary Research Collaborations", IEEE CBDCOM, 2022
- [6] 大江, 竹房, 丹生, 埴, 工藤, 合田, "クラウド環境構築システムVCPのmdxへの適用とOSS化に向けた試作", 情報処理学会研究報告, 2022.