

第1回SPring-8データワークショップ

「SPring-8データセンター構想によるデータ解析力の強化」
(2021年12月13日(月) 14:30～17:00)

Summary

城地 保昌

理研RSC

第1回SPring-8データワークショッププログラム

14:30~14:35 はじめに

矢橋 牧名 (理研)

(データセンター)

14:35~14:55 「SPring-8データセンター構想」

城地 保昌 (理研)

14:55~15:15 「富岳と大型施設との連携に向けて」

佐藤 賢斗 (理研)

(ビームライン)

15:15~15:30 「XAFS CTにおけるデータインフラの現状とニーズ」

関澤 央輝 (JASRI)

15:30~15:45 「イメージングビームラインにおけるデータインフラの現状とニーズ」

上杉 健太郎 (JASRI)

(サイエンスケース)

15:45~16:05 「オペランドXAFS分光イメージングにおけるデータ科学の必要性と期待」

唯 美津木 (名大)

16:05~16:25 「豊田中研の量子ビーム活用とSPring-8データセンターへの期待」

長井 康貴 (豊田中研)

16:25~16:45 ディスカッション

16:45~16:55 「研究データに関する国際的な流れ」

初井 宇記 (理研)

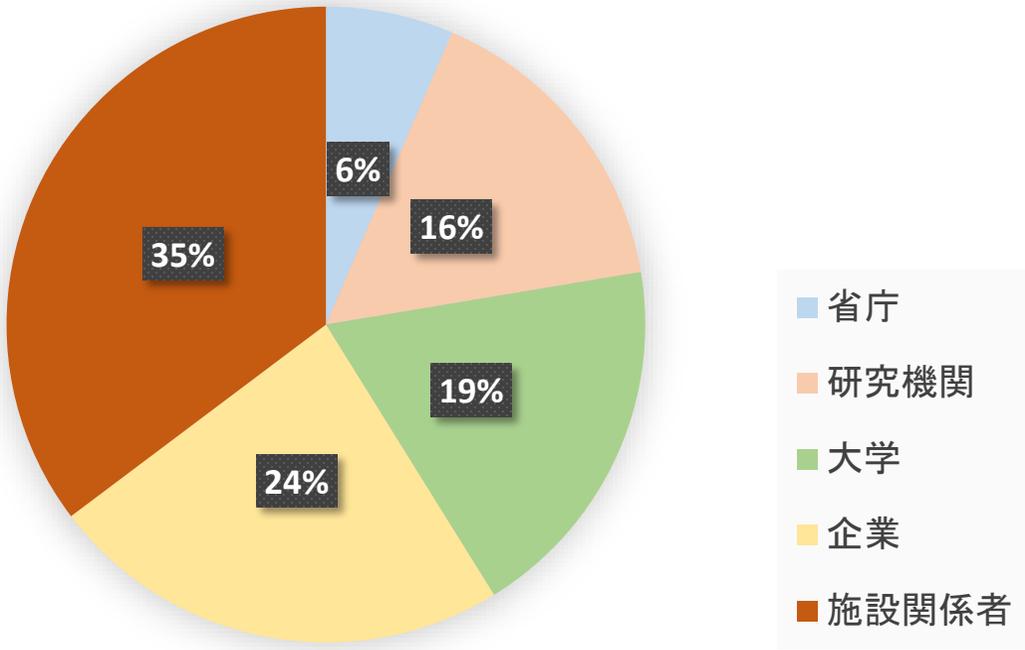
16:55~17:00 おわりに

矢橋 牧名 (理研)

第1回SPring-8データワークショップ Summary

- 講演スライドの公開HP
<https://dncom.spring8.or.jp/workshop/data-workshop/index.html>
- 最終的に講演者含めて170名の参加登録
 - ✓ 4分の1は企業関係者

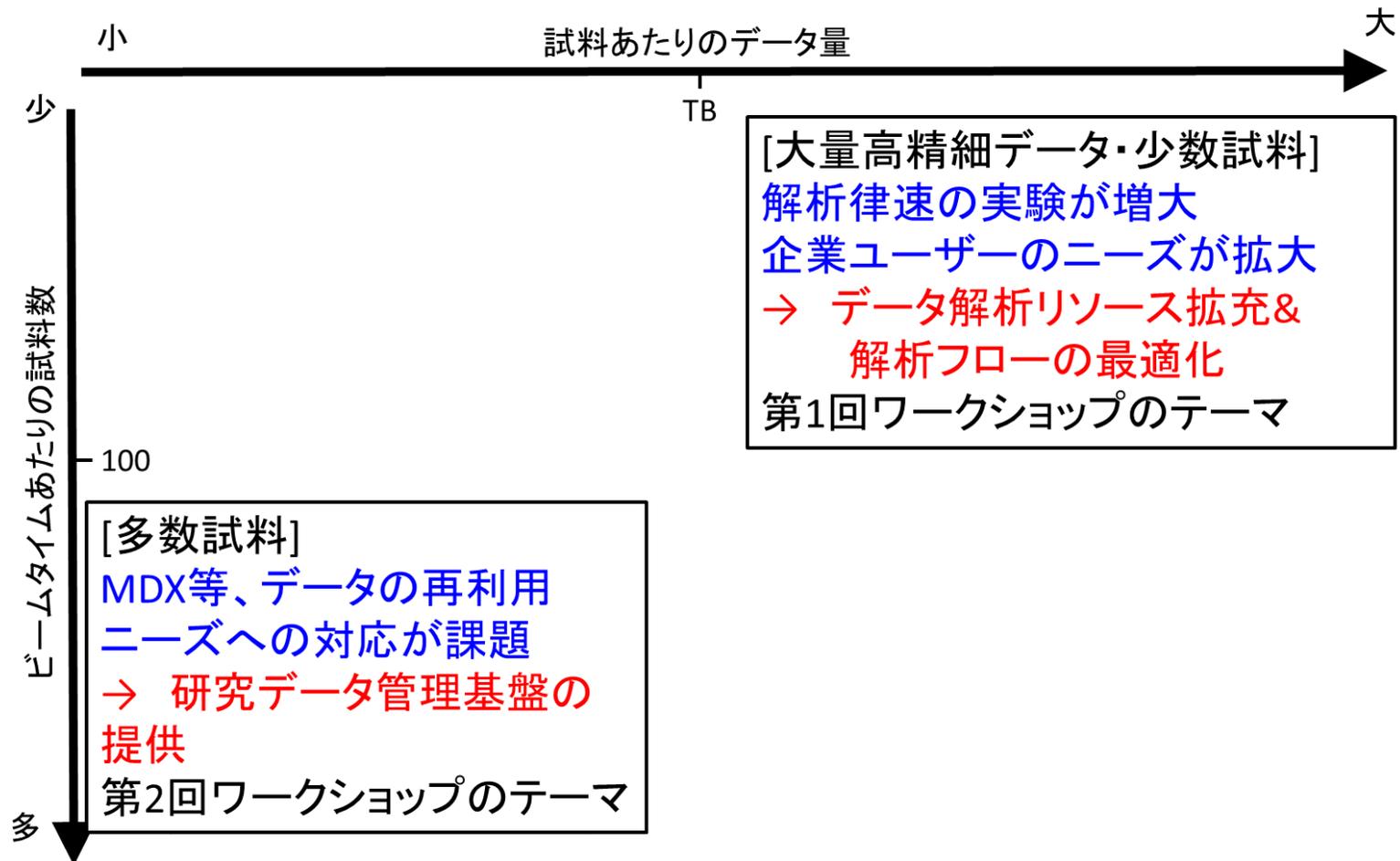
第1回SPring-8データワークショップ参加登録者



	人数
省庁	11
研究機関	27
大学	32
企業	40
施設関係者	60
合計	170

第1回Spring-8データワークショップ Summary

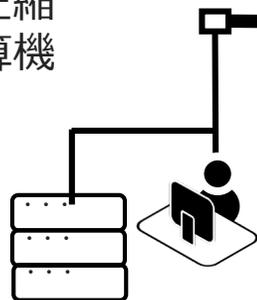
- 第1回は、試料あたりのデータ量がTBを超える実験で、データ解析が成果創出の律速となっている問題について議論
 - ✓ データセンター構想により、BL担当者や研究者側の負担となっていた解析リソース確保の問題を解消



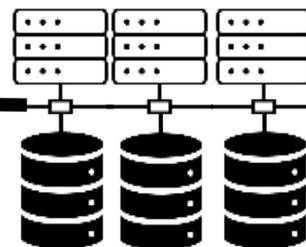
第1回SPring-8データワークショップ Summary



データ圧縮
専用計算機
BL 5本



基幹
ネットワーク
100 G



データセンター

[計算性能] 富岳の1/1000

CPU 5,120 core, 180 TFLOPS

GPU 16基, 320TFLOPS(FP32), 160TFLOPS(FP64)

[ストレージ容量] 富岳の1/8

ホットストレージ

> 10 PB (データ保持期間3年?)

コールドストレージ

小規模から始めて段階的に増強

[オンプレミスに構築する利点]

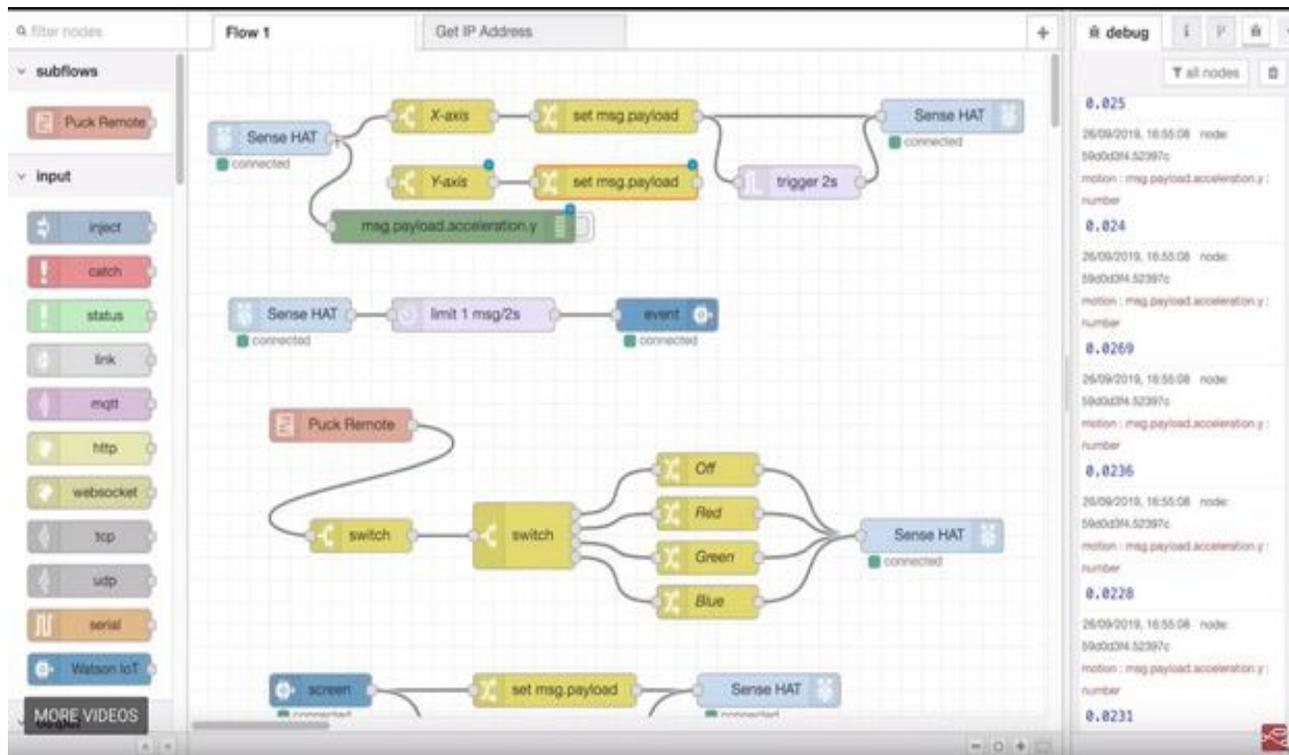
IO律速のデータ解析(実験中)に対応

計算ノードあたりのI/O性能は富岳より数倍速い

ユーザアクセスの利便性を向上させる機能を検討中

第1回Spring-8データワークショップ Summary

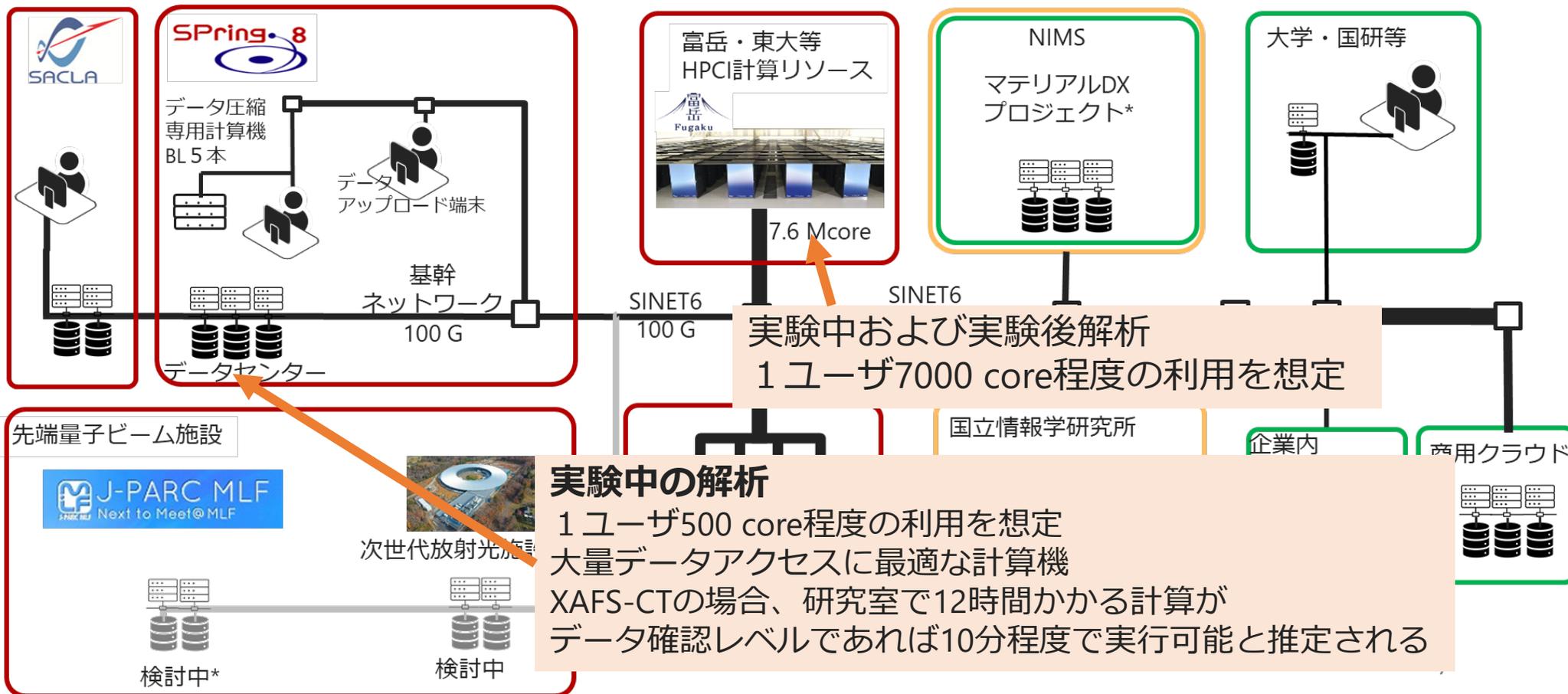
- BLでは測定速度向上や検出器の高精細化等により、今後データが増大していく(XAFS-CT: 10TB/day、X線CT: >300TB/year/BL)
 - ✓ ワークフローツール等を利用して典型的なデータ解析フローを作成し、研究グループのスキルによらず、簡単なパラメータ調整で、データ解析を実行できるようになると良い。



(Workflowツールによるノーコード開発の例)

第1回SPring-8データワークショップ Summary

- ✓ 最新の情報科学技術を導入し、計算量に応じてSPring-8データセンターとHPCI計算資源を使い分ける。
- ✓ データ転送のためのネットワーク高速化が重要。



第1回SPring-8データワークショップ Summary

- 質疑では、第2回のテーマである「データの流通、標準化」の関心が高かった。
 - ✓ 物質科学は、理想系での物性研究から、材料が機能している様子をそのまま可視化する研究へと進化してきており、データが多次元(空間・時間等)化し、データから意味を見出すためのデータ解析の重要度が高まっている。放射光ユーザーでない情報科学研究者と連携するための基盤整備が必要。
 - ✓ SPring-8のデータだけでなく、J-PARC等他施設のデータも同様に共有できる仕組みがあると良い。
 - ✓ 他施設も含めたデータ共有のために、データフォーマット内に格納する要素(メタデータ等)について放射光学会と協力しながら進めていきたい。
 - ✓ 大規模データを貯めるだけでなく、研究データをつなげて価値を生み出す環境整備が重要であり、そのための議論を継続して欲しい。

スケジュール

