第3世代 JAXAスーパーコンピュータシステムの紹介

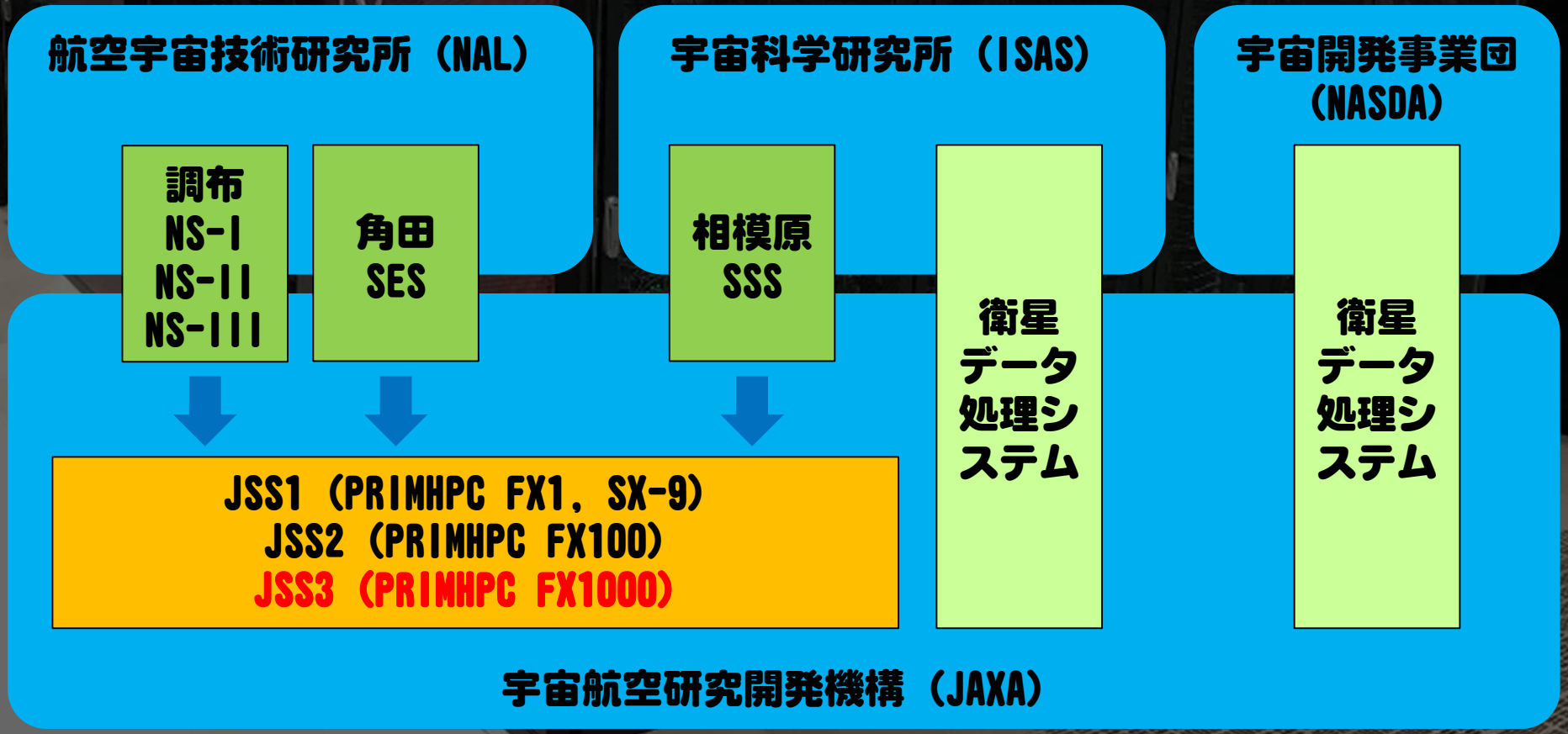
高木　亮治
宇宙航空研究開発機構
宇宙科学研究所

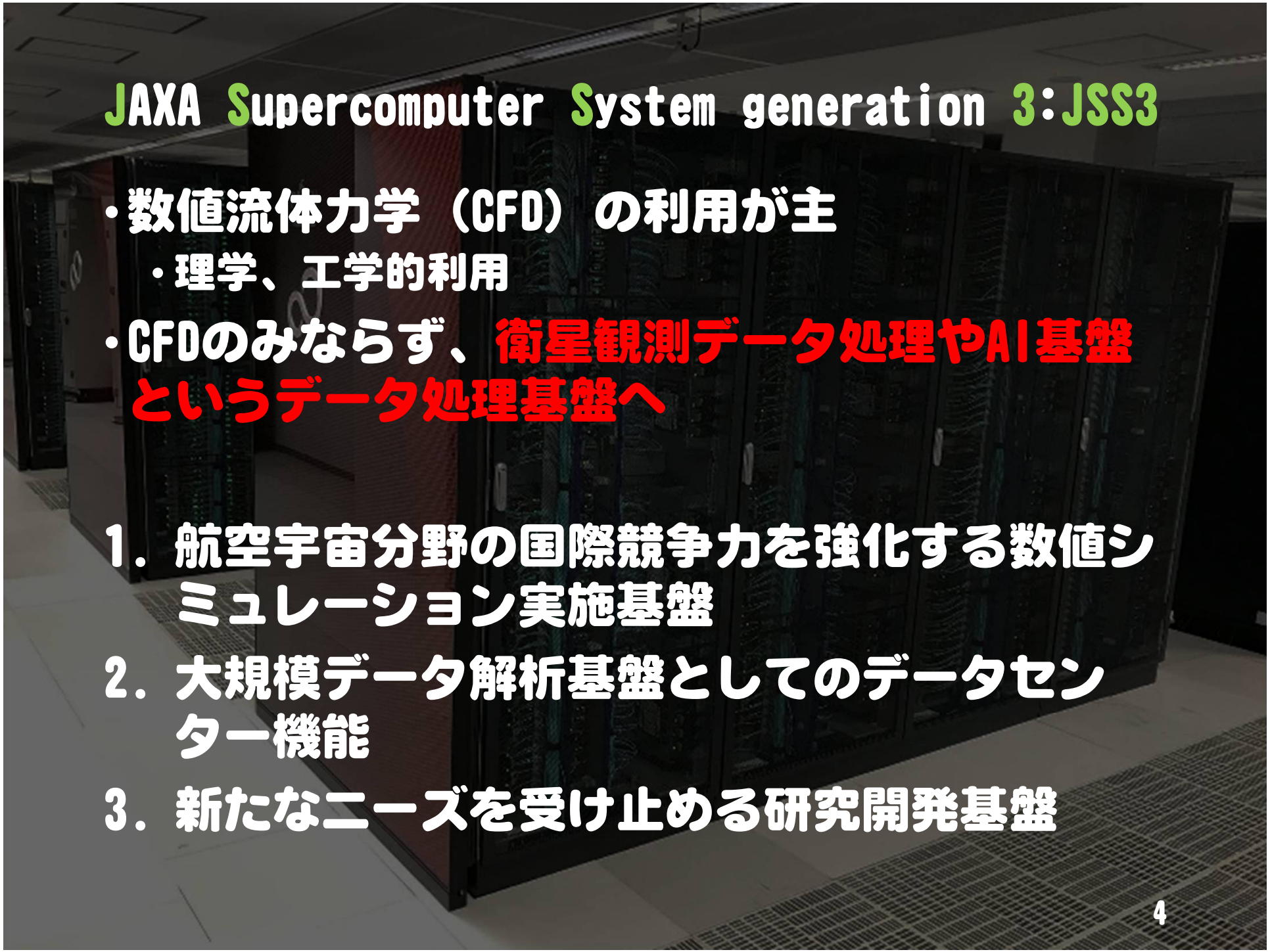
内容

- 第3世代JAXAスーパーコンピュータシステムの概要
  - 位置付け
  - H/W（各サブシステム）
  - S/W（OS、言語環境）
- 今回の「推し」
  - 可視化
  - 仮想化技術
  - 止まらない
- その他諸々
  - 外部からの利用制度
  - 大規模チャレンジ
- 利用事例（TOKI-SORA）
  - メモリ性能の実力
  - 高速化のポイント
- まとめ

JAXAスーパーコンピューター

・JAXA統合によりJSS（JAXA Supercomputer System）が誕生





JAXA Supercomputer System generation 3:JSS3

- 数値流体力学（CFD）の利用が主
  - 理学、工学的利用
- CFDのみならず、**衛星観測データ処理やAI基盤というデータ処理基盤へ**

1. 航空宇宙分野の国際競争力を強化する数値シミュレーション実施基盤
2. 大規模データ解析基盤としてのデータセンター機能
3. 新たなニーズを受け止める研究開発基盤

JAXAスーパーコンピュータシステム JSS3

稼働開始
2020年12月1日

【コンピュータ基盤】TOKI: TOkyo and ibaraKI

調布航空宇宙センター

TOKI-SORA: HPCシステム
PRIMEHPC FX1000
ノード数：5,760 ノード（15ラック）
総理論演算性能：19.4 PFLOPS
総主記憶容量：180 TiB（32 GiB/ノード）

TOKI-RURI: 汎用システム
総理論演算性能：1.24 PFLOPS
総主記憶容量：104 TiB
- ST: PRIMERGY RX2540 M5 x 375ノード（192 GiB/ノード, Quadro x 1 基）
- GP: PRIMERGY CX2570 M5 x 32ノード（384 GiB/ノード, Tesla V100 x 4 基）
- XM: PRIMERGY RX2540 M5 x 2 ノード（DCPMM 6.0 TiB/ノード, Quadro x 1基）
- LM: PRIMERGY RX2540 M5 x 7ノード（DCPMM 1.5 TiB/ノード, Quadro x 1基）

TOKI-FS: ファイルシステム
ファイルシステム：FEFS
オールフラッシュ NVMe ストレージ: 10PB
ハードディスクドライブ ストレージ: 40PB

TOKI-LI: ログインシステム
PRIMERGY RX2540 M5 x 最大14ノード（384 GiB/ノード, Quadro x 1基）

運用管理システム

12 Tbps 相互接続網（InfiniBand） 45.7 Tbps 20.8 Tbps 2.8 Tbps

360 Gbps 高速 Ethernet バックボーン 416 Gbps 280 Gbps

10 Gbps

筑波宇宙センター

40 Gbps 49 Gbps 高速 Ethernet バックボーン

400 Gbps 400 Gbps 5.2 Tbps 相互接続網（InfiniBand）

TOKI-TFS: 筑波ファイルシステム
ファイルシステム：FEFS, 総実効容量：0.4PB

TOKI-TLI: 筑波ログインシステム
PRIMERGY RX2540 M5 x 2ノード（384 GiB/ノード, Quadro x 1基）

筑波運用管理制御システム

TOKI-TRURI: 筑波汎用システム
総理論演算性能：145 TFLOPS
総主記憶容量：10.8 TiB
- TST: PRIMERGY RX2540 M5 x 46 ノード（192 GiB/ノード, Quadro x 1基）
- TGP: PRIMERGY CX2570 M5 x 2 ノード（384 GiB/ノード, Tesla V100 x 4基）
- TLM: PRIMERGY RX2540 M5 x 1 ノード（DCPMM 1.5 TiB/ノード, Quadro x 1基）

80 Gbps

【アーカイバ基盤】J-SPACE
ディスクキャッシュ容量：3PB
テープ容量：70PB

Powered by HPSS

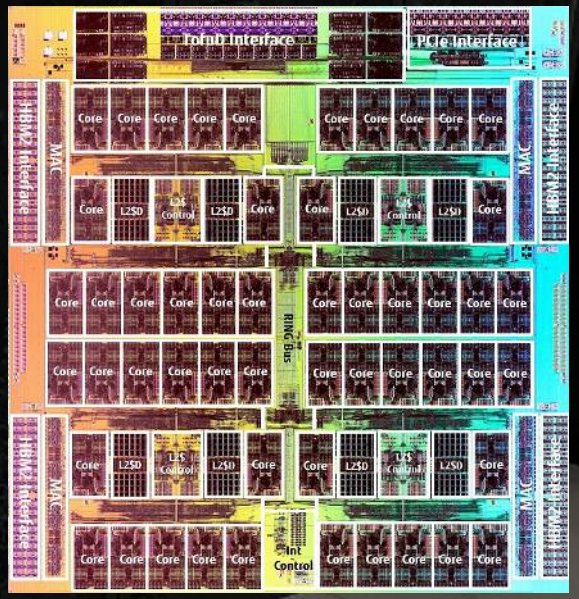
＊調布航空宇宙センター内に設置

©スーパーコンピュータ活用課

TOKI-SORA（PRIMEHPC FX1000）

SORA: Supercomputer for earth Observation, Rockets, and Aeronautics

- TOKI-SORA(時空)：HPCシステム
  - 総理論演算性能：19.4PFLOPS（5,760ノード）「富岳」の1/27.6
  - 総主記憶容量：180TiB
  - CPU：A64FX（Armv8.2+SVE）
    - 3.3792TFLOPS（倍精度演算）
    - 2.2GHz、48コア+アシスタントコア
    - 12コアがL2$を共有（CMG）
  - メモリ：HBM2
    - 32GiB、1,024GB/s
  - インタコネクト：TofuインターコネクトD
    - 40.8GB/s×双方向



TOKI-RURI（PRIMERGY RX2540/CX2570）

RURI: all-RoUnd Role Infrastructure

・TOKI-RURI：汎用システム（Intel x86系CPU）

拠点	名称	ノード数	CPU（18コア）	メモリ	GPU
調布	ST	375	Xeon Gold 6240 2.6GHz×2	192GiB	NVIDIA Quadro P4000
	GP	32		384GiB	NVIDIA Tesla V100×4
	LM	7		1.5TiB(DCPMM)	NVIDIA Quadro P4000
	XM	2	Xeon Gold 6240L 2.6GHz×2	6TiB(DCPMM)	NVIDIA Quadro P4000
筑波	TST	46	Xeon Gold 6240 2.6GHz×2	192GiB	NVIDIA Quadro P4000
	TGP	2		384GiB	NVIDIA Tesla V100×4
	TLM	1		1.5TiB(DCPMM)	NVIDIA Quadro P4000

Intel Xeon Gold 6240（2.6GHz）：2.9952TFLOPS
DDR4：2,933MHz，32GiB，281.5GB/s，Intel Optane DCPMM
NVIDIA Tesla V100：5,120 CUDAコア，32GiB，62.8TFLOPS(倍精度)
NVIDIA Quadro P4000：1,792 CUDAコア，8GiB

TOKI-FS/J-SPACE

・TOKI-FS：ファイルシステム（FEFS+DDN）

拠点	領域	path	容量	用途
調布	HOME	/home	500TB	ソース、実行モジュールなど小規模ファイル
	DATA	/data	40PB	大容量ファイル
		/ltmp		一時ファイル（60日間未参照未更新で削除）
	SSD	/ssd	10PB	高速なI/Oが必要なファイル
筑波	HOME	/home	400 TB	ソース、実行モジュールなど小規模ファイル
		/data		大容量ファイル
		/ltmp		一時ファイル（60日間未参照未更新で削除）

・J-SPACE：階層型ストレージシステム（IBM HPSS）

・ディスクキャッシュ：3PB

・テープ：67.9 PB

ソフトウェア

	TOKI-SORA	TOKI-LI/TLI	TOKI-RURI/TRURI
OS	RHEL 8.2	RHEL 7.7	CentOS 7.7
コンパイラ	• TCS開発環境 • GCC	• TCS開発環境（クロス） • GCC	• Intel Parallel Studis XE Cluster Edistion • NVIDIA HPC SDK (CUDA,OpenACC) • GCC
プロファイラ	• TCS開発環境 • Gprof	• TCS開発環境	• Intel Vtune Amplifier XE • NVIDIA HPC SDK(Nsight) • Gprof
デバッガ	• TCS開発環境 • GDB	• TCS開発環境	• IDB • NVIDIA HPC SDK(cuda-gdb) • GDB
MPI	• TCS開発環境	N/A	• Intel MPIライブラリ • NVIDIA HPC SDK (Open MPI) • Open MPI

TOKI-SORA：言語環境

- Fortran：
  - Fortran 77，90，95，2003，2008，2018（の一部）
- C言語：
  - GNU C，C89，C99，C11
- C++言語：
  - GNU C++，C++03，11，14，17
- OpenMP：
  - Version 4.5，5.0（の一部）
- MPI：
  - Version 3.1，4.0（の一部）

C/C++のモード：
- trad：従来コンパイラ
- clang：Clang/LLVM

↓

OSSへの対応
（PyTorch，TensorFlow）

今回の「推し」

- 可視化
- 仮想化
- 止まらない

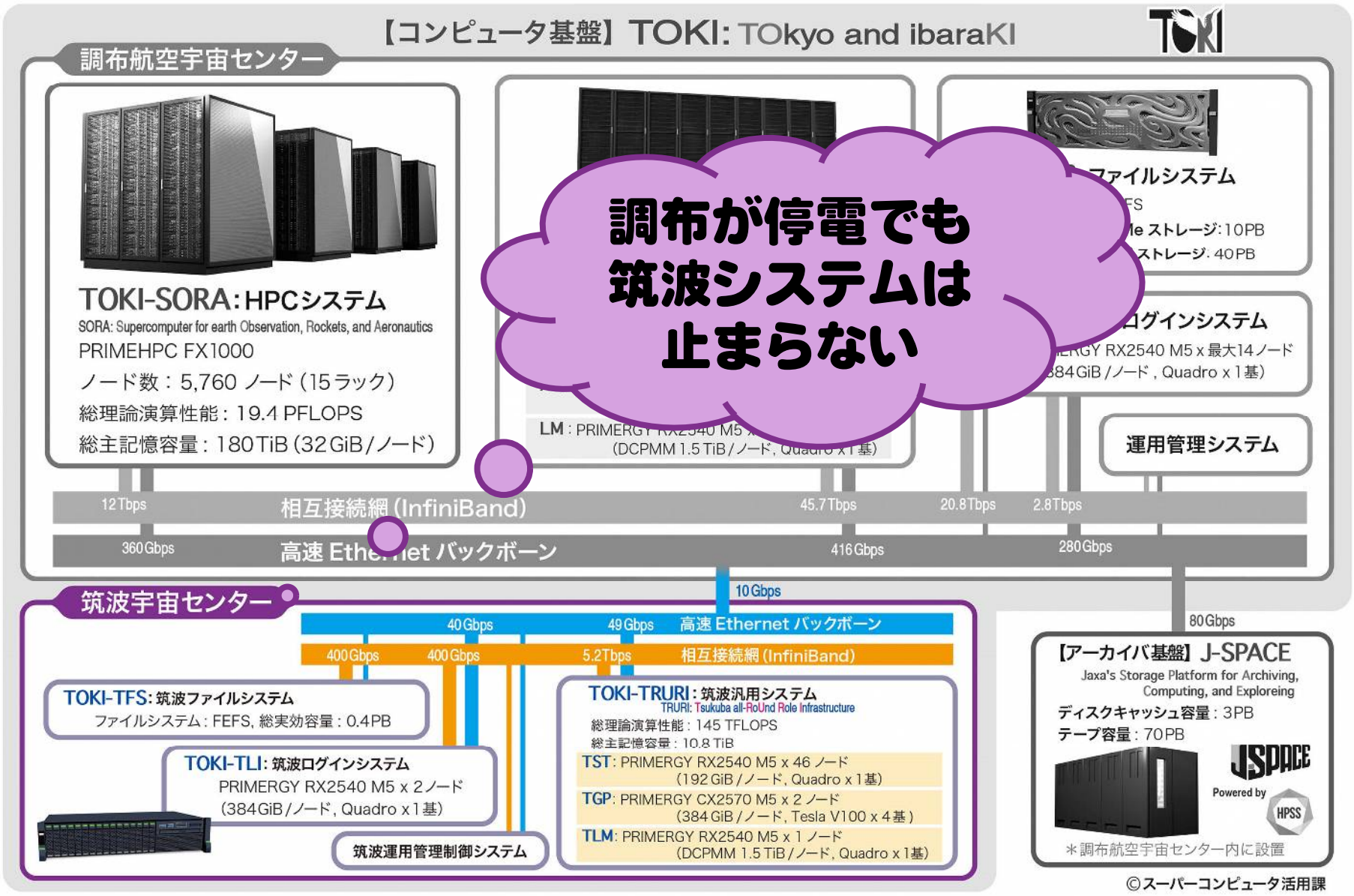
可視化

- In-situ/in-transit可視化：シミュレーションをしながら可視化を行う。
  - 従来：シミュレーション→結果データ→可視化→画像、動画、etc….
  - 課題：データ（ファイル）I/Oが困難
    - データ容量、データI/O速度
    - 演算性能とI/O性能の乖離
- 処理形態の異なるプロセスの協調処理
  - ソルバー：ノンストップ処理
  - 可視化：バッチ処理 or インタラクティブ処理
- ADIOS2などのミドルウェア
  - ソルバープロセスと可視化プロセスでデータをメモリ上で共有

仮想化

- 所謂コンテナ（Docker，Singularity）の提供
  - ユーザーが自由にインストール・利用できる実行環境を提供する仮想化技術
- JSS3ではSingularityを提供
  - ユーザー権限で実行されるのでroot権限が不要
- 利用イメージ
  - 取得（singularityコンテナイメージに変換）
    - 公開されているコンテナの利用
      - Docker hub：https://hub.docker.com/
    - JSS3に用意されたNGCのコンテナ
      - NVIDIAが提供するML、DL向けのGPU対応コンテナ
  - 変更（sandboxに変換、OSSインストールなど、sandboxからイメージファイルへ変換）
  - 実行（ジョブ投入、singularity run）

JAXAスーパーコンピュータシステム JSS3

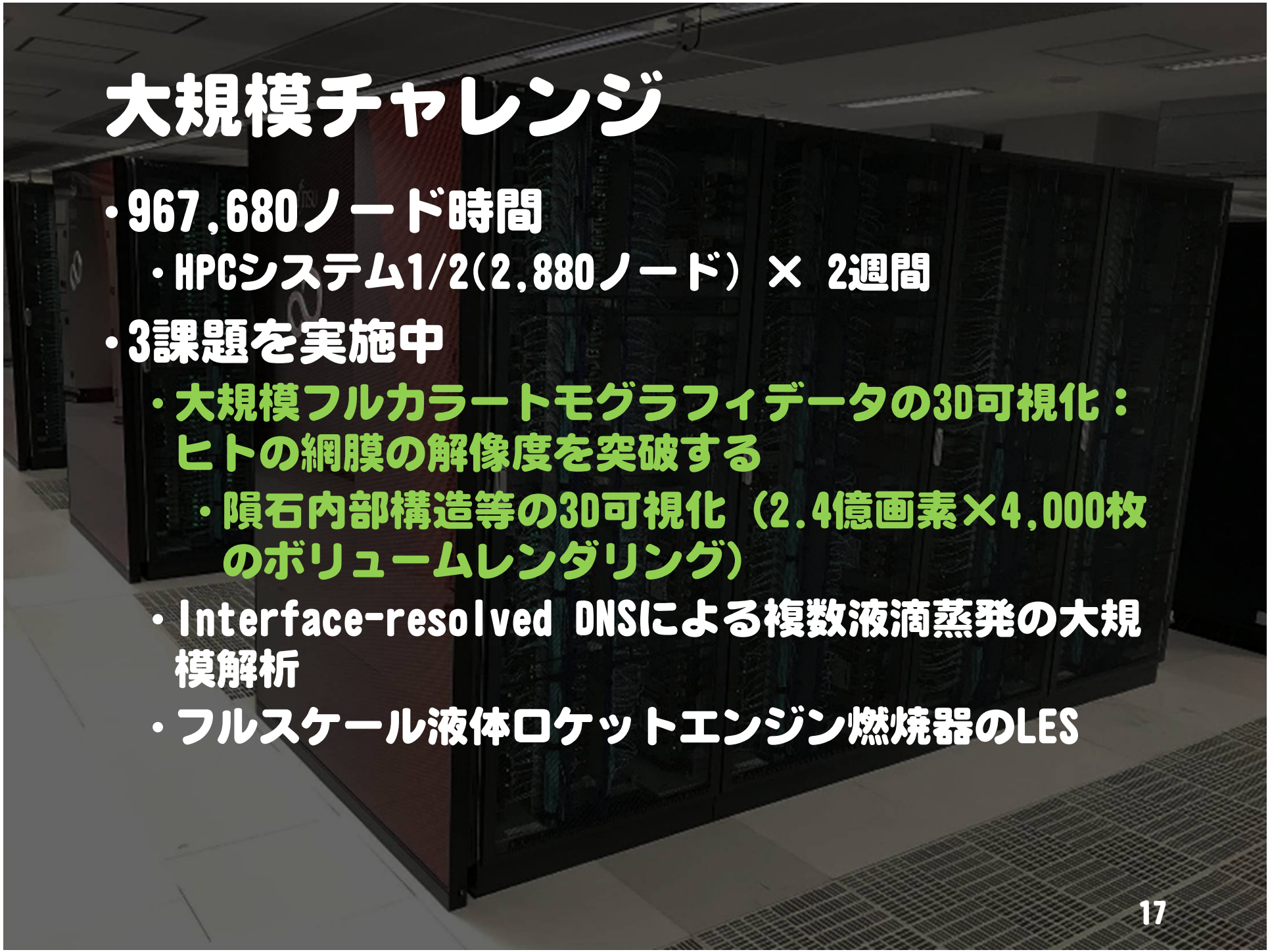


その他諸々

- JAXA外部からの利用
- 大規模チャレンジ

JAXA外部からの利用

- JSS大学共同利用
  - 全国大学共同利用研究の一貫として、宇宙科学研究所が行っている飛翔体(科学衛星・ロケット・大気球)プロジェクト等と密接に関連する宇宙科学の研究課題について、スパコンシステムを利用する制度。
- 設備供用
  - JAXAが保有する試験設備等を広く機構外の方に有償でご利用いただく制度。JSS3資源をJAXA外部の申請者に有償で貸し出し、プロダクション実行も可能。
    - 「トライアルユース」制度も設けており、有償利用の前に利用可否を判断するために無償で試用可能。
- 一般利用
  - JAXA内で研究する学生、派遣・請負・委託・共同研究相手方



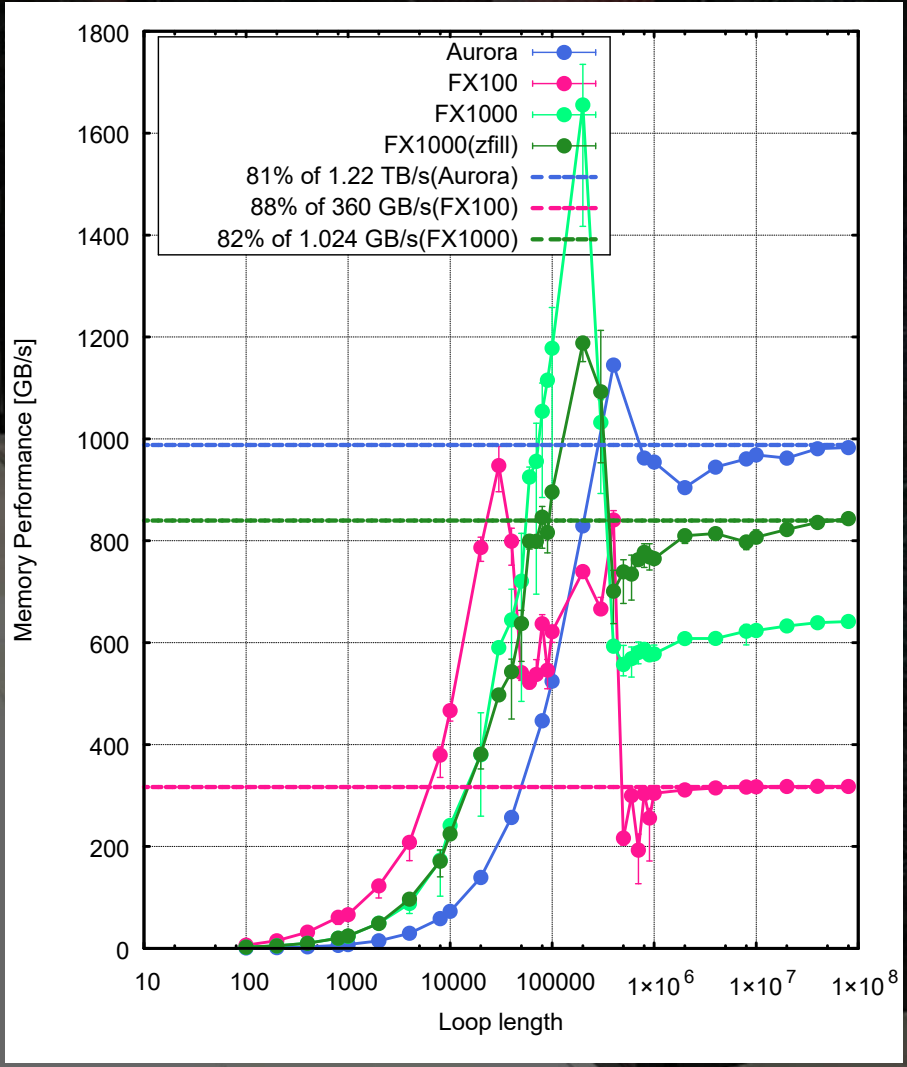
大規模チャレンジ

- 967,680ノード時間
  - HPCシステム1/2(2,880ノード）× 2週間
- 3課題を実施中
  - 大規模フルカラートモグラフィデータの3D可視化：ヒトの網膜の解像度を突破する
    - 隕石内部構造等の3D可視化（2.4億画素×4,000枚のボリュームレンダリング）
  - Interface-resolved DNSによる複数液滴蒸発の大規模解析
  - フルスケール液体ロケットエンジン燃焼器のLES

利用事例（TOKI-SORA）

- メモリ性能の実力
- 高速化のポイント

メモリ性能の実力

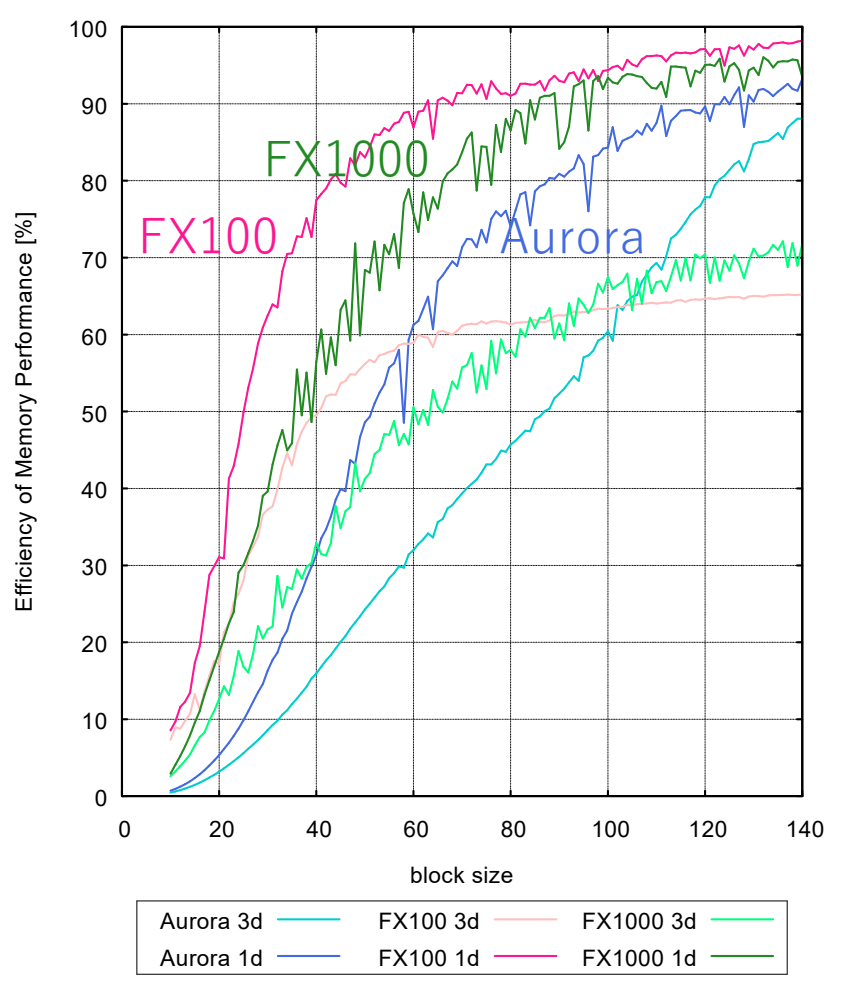


- TRIAD： メモリ性能のBM

```
do n=1,nmax
   a(n) = b(n) + S × c(n)
enddo
```

  - B/F = 24
  - 演算効率： 1.3%（FX1000）
- 理論メモリ性能（B/F）：
  - FX100: 0.48GB/s（0.43）
  - FX1000： 1.0TB/s（0.30）
  - Aurora： 1.2TB/s（0.50）
- メモリの実力値： 最大で80%程度

メモリ性能の実力



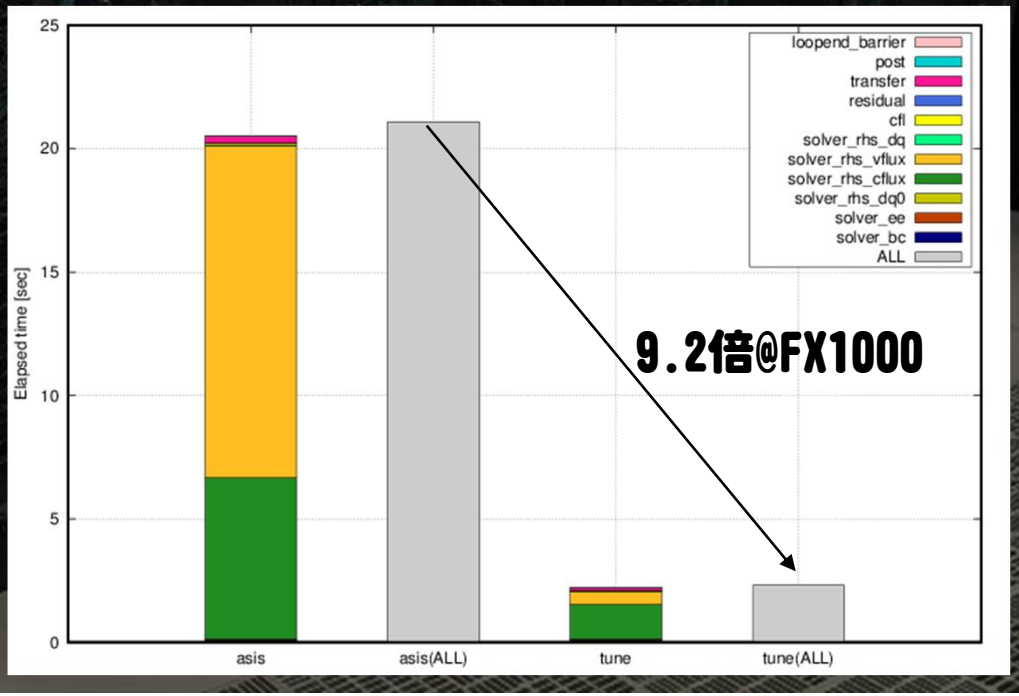
- MBTRIAD： キャッシュの影響を排除し純粋にメモリ性能の評価

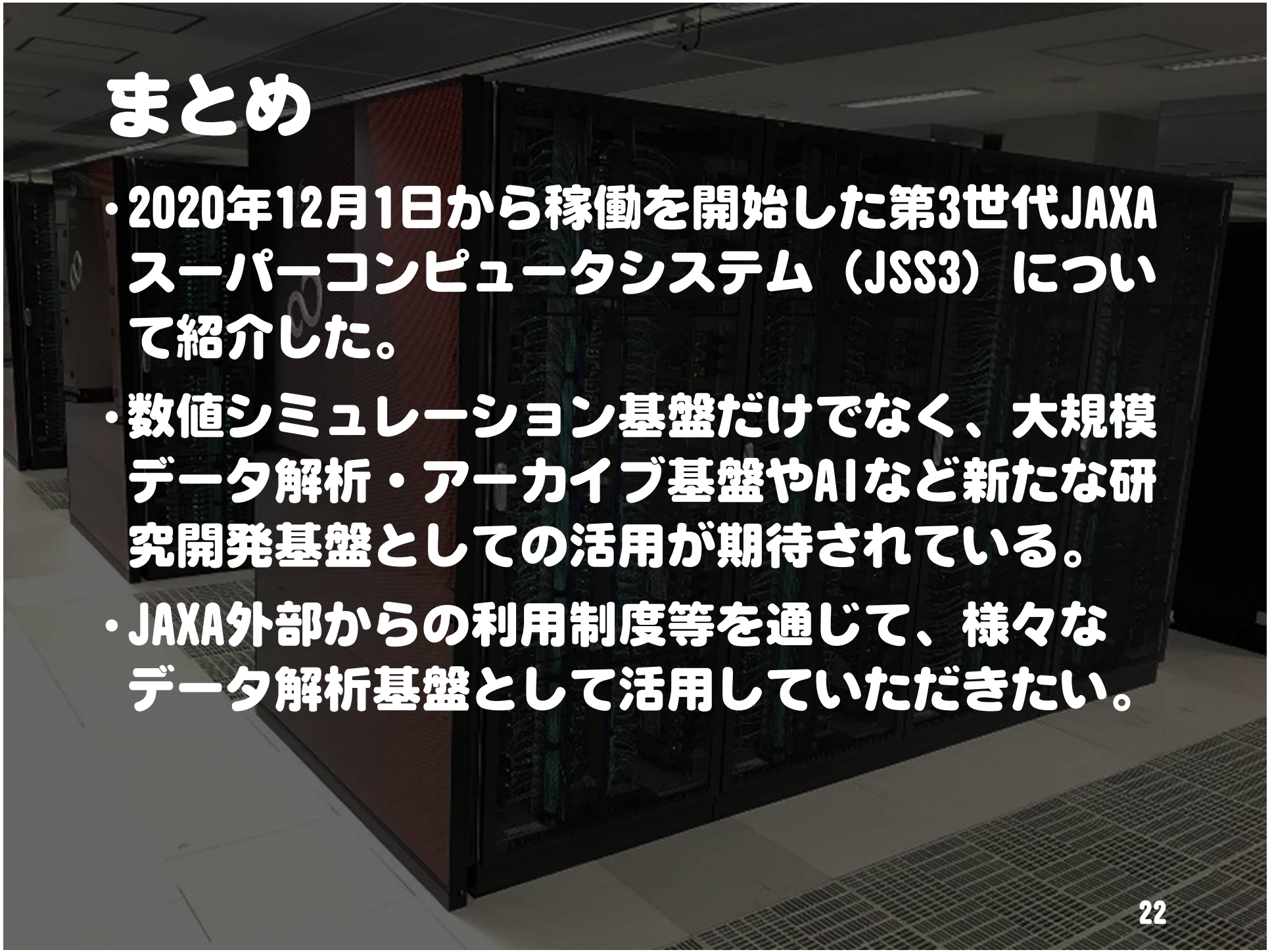
```
do n=1,NB
do k=1,N
do j=1,N
do i=1,N
  blk(n)%a(i,j,k) =
blk(n)%b(i,j,k) + S ×
blk(n)%c(i,j,k)
enddo; enddo; enddo; enddo
```

  - N: block size, NB: # of blocks
- ループ長の立ち上がりに注目
  - ある程度のループ長が必要

プログラム高速化のポイント

- 並列化
  - プロセス並列：MPI
  - スレッド並列：OpenMP、OpenACC（GPU）
- 演算機能の活用
  - SIMD、SWP、コンパイラの最適化
- メモリ性能
  - 連続アクセス
  - 長いループ長





まとめ

- 2020年12月1日から稼働を開始した第3世代JAXAスーパーコンピュータシステム（JSS3）について紹介した。
- 数値シミュレーション基盤だけでなく、大規模データ解析・アーカイブ基盤やAIなど新たな研究開発基盤としての活用が期待されている。
- JAXA外部からの利用制度等を通じて、様々なデータ解析基盤として活用していただきたい。