



Wisteria/BDEC-01

「計算・データ・学習」融合
スーパーコンピュータシステム



東京大学情報基盤センター

<https://www.cc.u-tokyo.ac.jp/supercomputer/wisteria/service/>

2024/12/13

- **東大情報基盤センターのスパコン概要**
- Wisteria/BDEC-01
- Ipomoea-01
- h3-Open-BDEC

FY11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Yayoi, Hitachi SR16000/M1
IBM Power-7
54.9 TFLOPS, 28.7 TB/sec.

Oakbridge-CX, Fujitsu
Intel Xeon Cascade Lake
6.61 PFLOPS, 385.2 TB/sec.

T2K Tokyo, Hitachi
AMD Opteron
140.0 TF, 38.1 TB/sec.

Oakforest-PACS (OFP) (JCAHPC)
Fujitsu, Intel Xeon Phi
25.0 PFLOPS, 8.51 PB/sec.

OFP-II
(JCAHPC)
100+ PF ?

Oakleaf-FX, Fujitsu PRIMEHPC FX10
SPARC64 IXfx
1.13 PFLOPS, 408.0 TB/sec.

 **Wisteria
BDEC-01**

BDEC: Wisteria/BDEC-01, Fujitsu
Fujitsu A64FX (Odyssey) +
Intel Xeon Ice Lake/NVIDIA A100
(Aquarius)
33.1 PFLOPS, 8.38 PB/sec.
**Integrated Supercomputer System for
Simulation, Data & Learning**

Oakbridge-FX, Fujitsu
SPARC64 IXfx
136.2 TFLOPS, 49.0 TB/sec.

Data Platform (mdx), Fujitsu
Intel Ice Lake/NVIDIA A100 etc.
8.50 PFLOPS, 0.665 PB/sec.

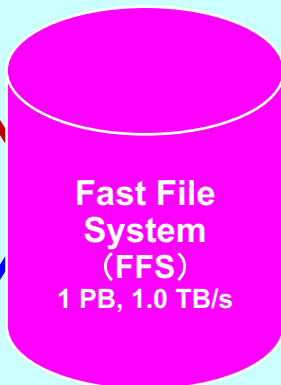
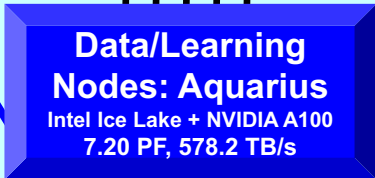
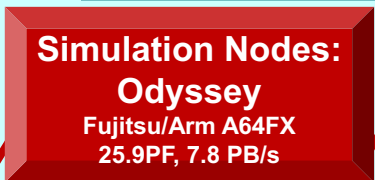
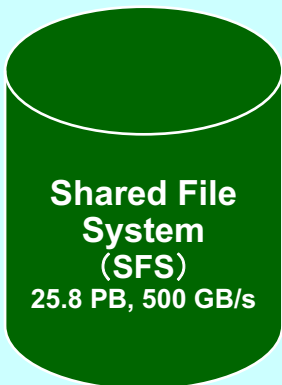
**東京大学情報基盤
センターのスパコン**
利用者2,600+名
55%は学外

Reedbush-U/H, HPE
Intel BDW + NVIDIA P100
1.93 PFLOPS, 258.6 TB/sec.

Reedbush-L, HPE
Intel BDW + NVIDIA P100
1.43 PFLOPS, 197.2 TB/sec.



Platform for Integration of (S+D+L)
Big Data & Extreme Computing

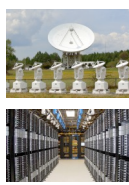


800 Gbps

External Resources



External Network



External Resources



Simulation Nodes
(Odyssey)

2021/12/13



Data/Learning Nodes
(Aquarius)



東京大学
THE UNIVERSITY OF TOKYO



東京大学情報基盤センター
INFORMATION TECHNOLOGY CENTER, THE UNIVERSITY OF TOKYO

Reedbush (HPE, Intel BDW + NVIDIA P100 (Pascal))

- データ解析・シミュレーション融合スーパーコンピュータ
- 2016年7月～2021年11月末(引退)
- 東大ITC初のGPUクラスタ, ピーク性能3.36 PF

Oakforest-PACS (OFP) (Fujitsu, Intel Xeon Phi (KNL))

- JCAHPC (筑波大CCS・東大ITC), 2016年10月～2022年3月末(予定)
- 25 PF, #39 in 58th TOP 500 (November 2021)

Oakbridge-CX (OBCX) (Fujitsu, Intel Xeon CLX)

- 2019年7月～2023年6月末(予定)
- 6.61 PF, #110 in 58th TOP500-June 2023 (Plan)



Wisteria/BDEC-01 (Fujitsu)

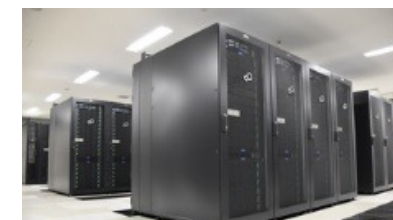
- シミュレーションノード群 (Odyssey) : A64FX (#17)
- データ・学習ノード群 (Aquarius) : Intel Icelake+NVIDIA A100 (#106)
- 33.1 PF, #13 in 57th TOP 500, 2021年5月14日運用開始
- 「計算・データ・学習(S+D+L)」融合のためのプラットフォーム
- 革新的ソフトウェア基盤「h3-Open-BDEC」
(科研費基盤(S) 2019年度～2023年度)



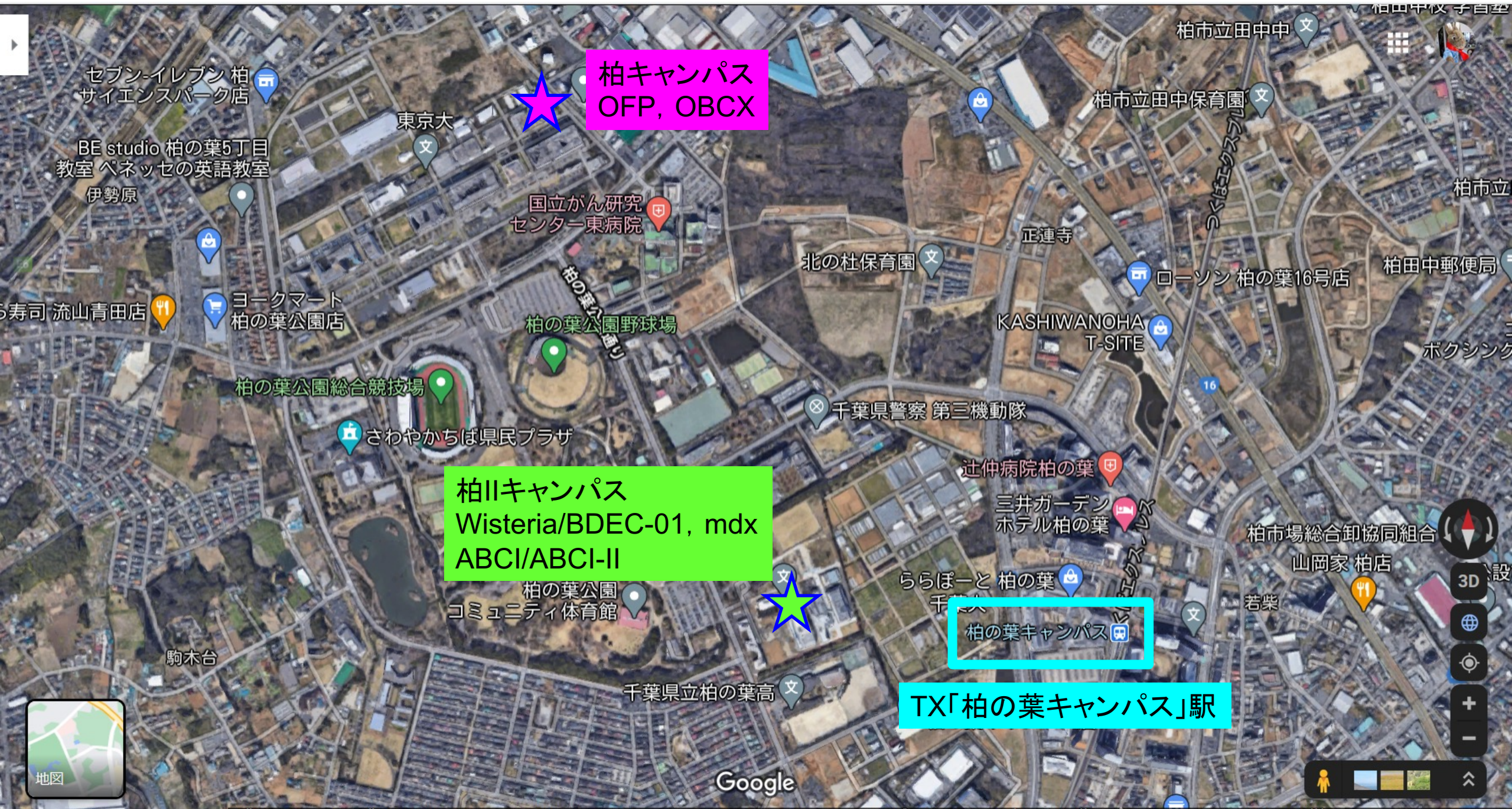
Reedbush



Oakforest-PACS



Oakbridge-CX

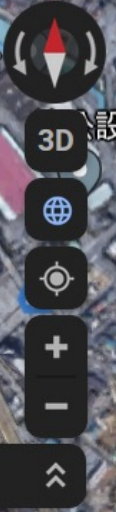


柏キャンパス
OFFP, OBCX

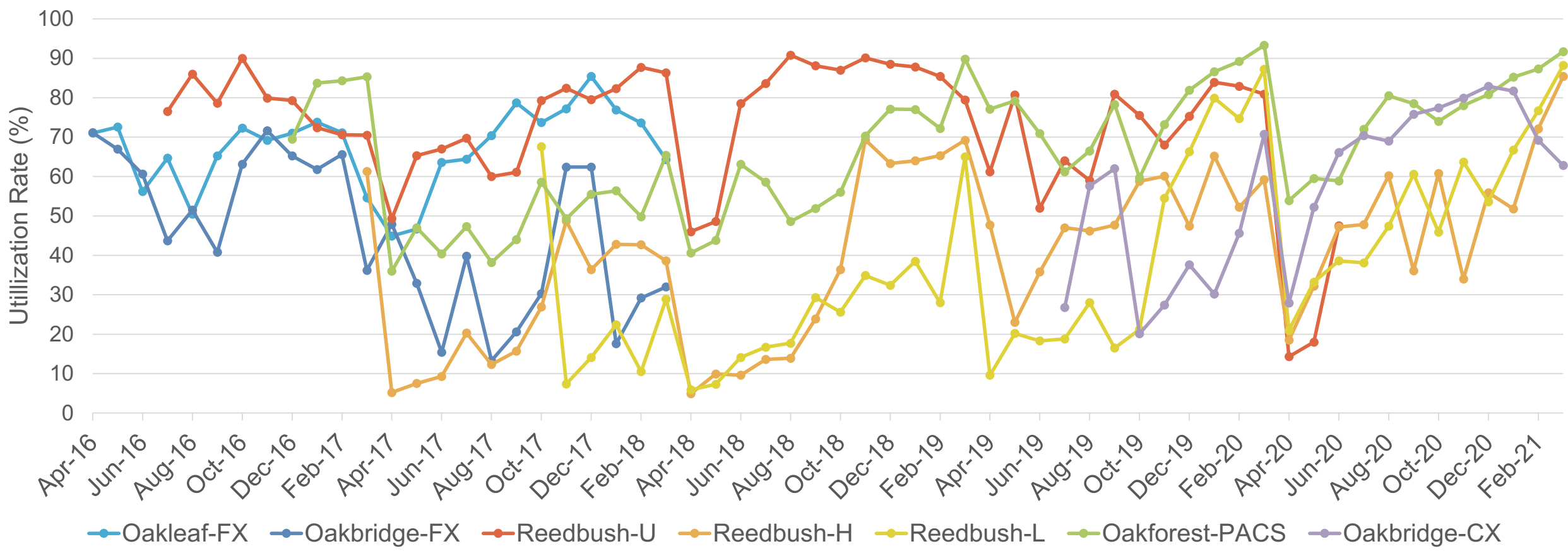
柏IIキャンパス
Wisteria/BDEC-01, mdx
ABCI/ABCI-II

柏の葉キャンパス

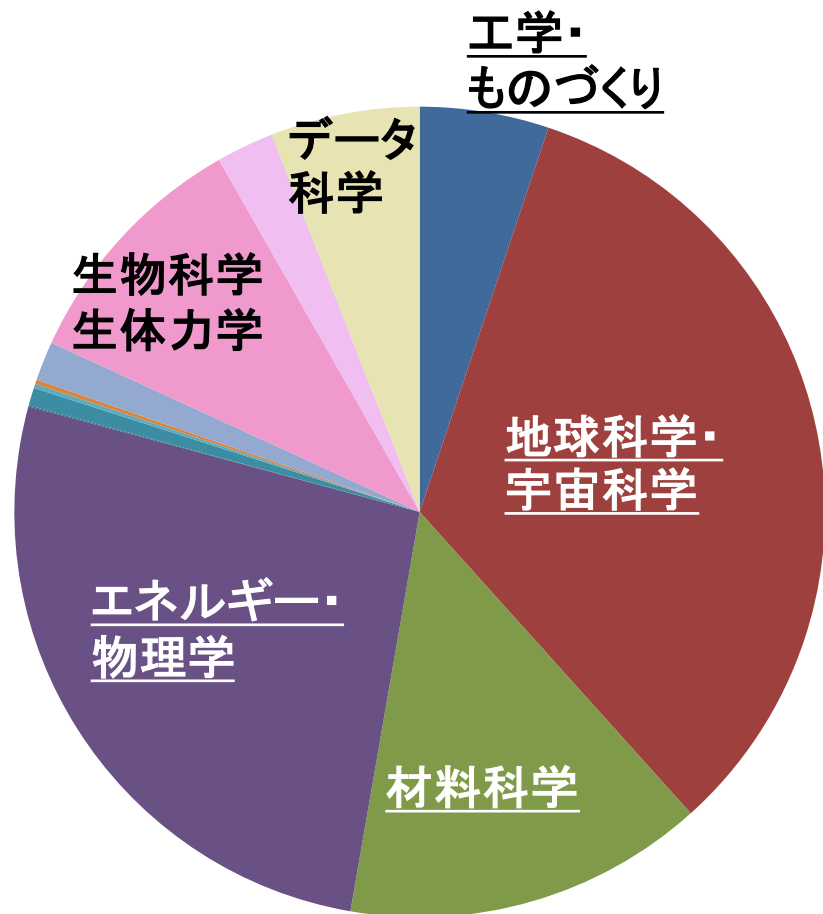
TX「柏の葉キャンパス」駅



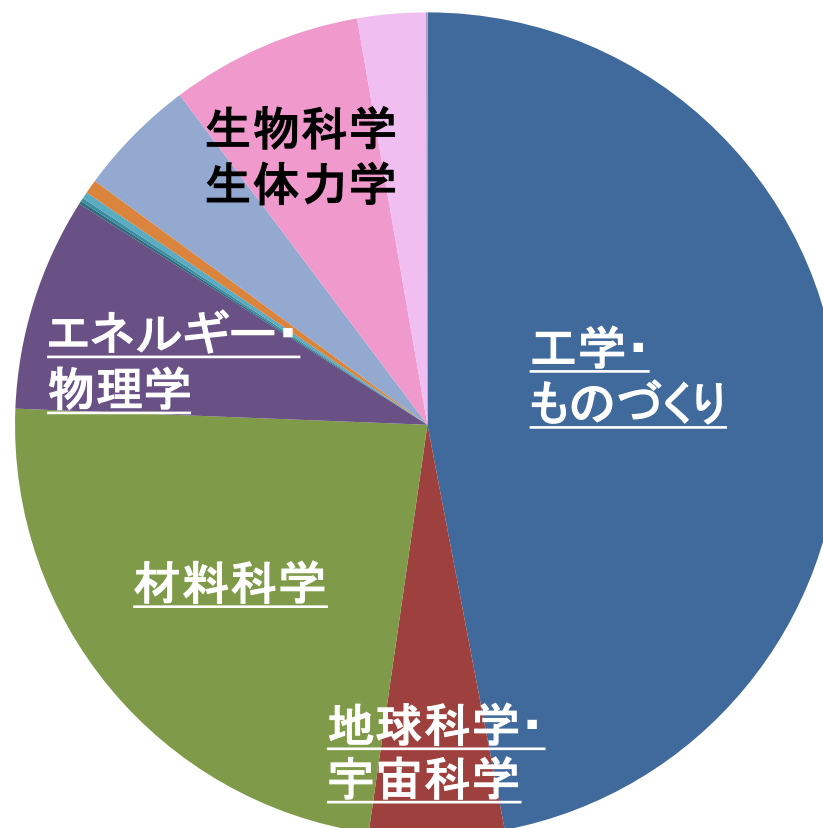
各システムの月平均利用率履歴



研究分野別利用CPU時間割合(2020年度)



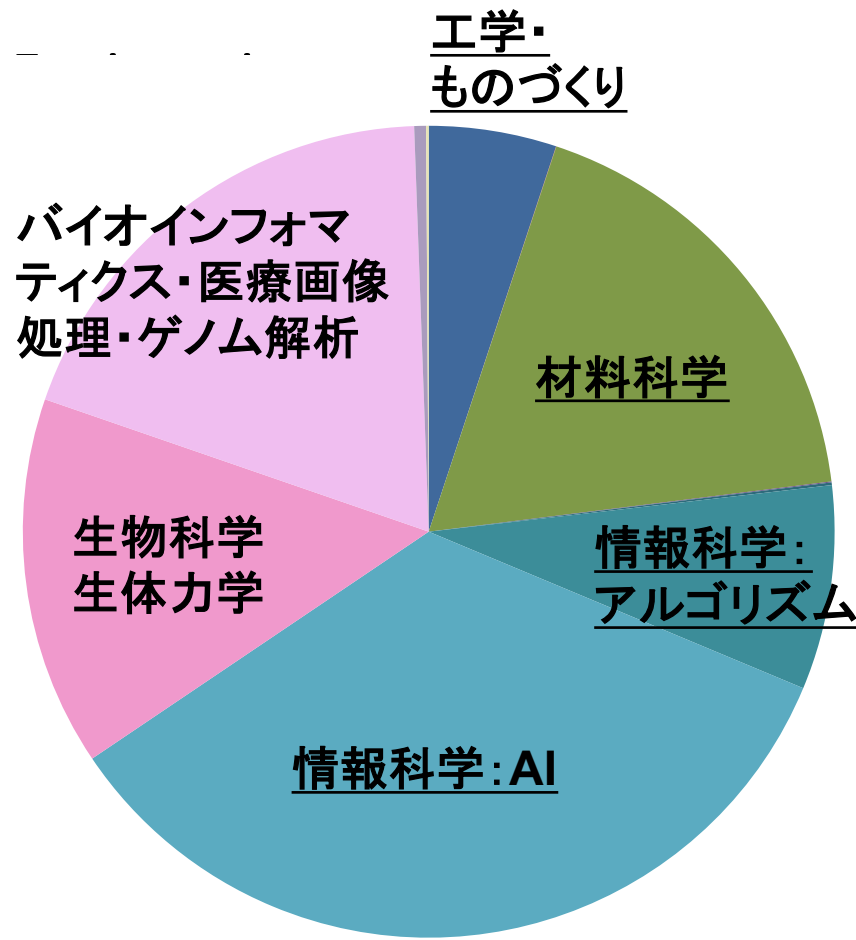
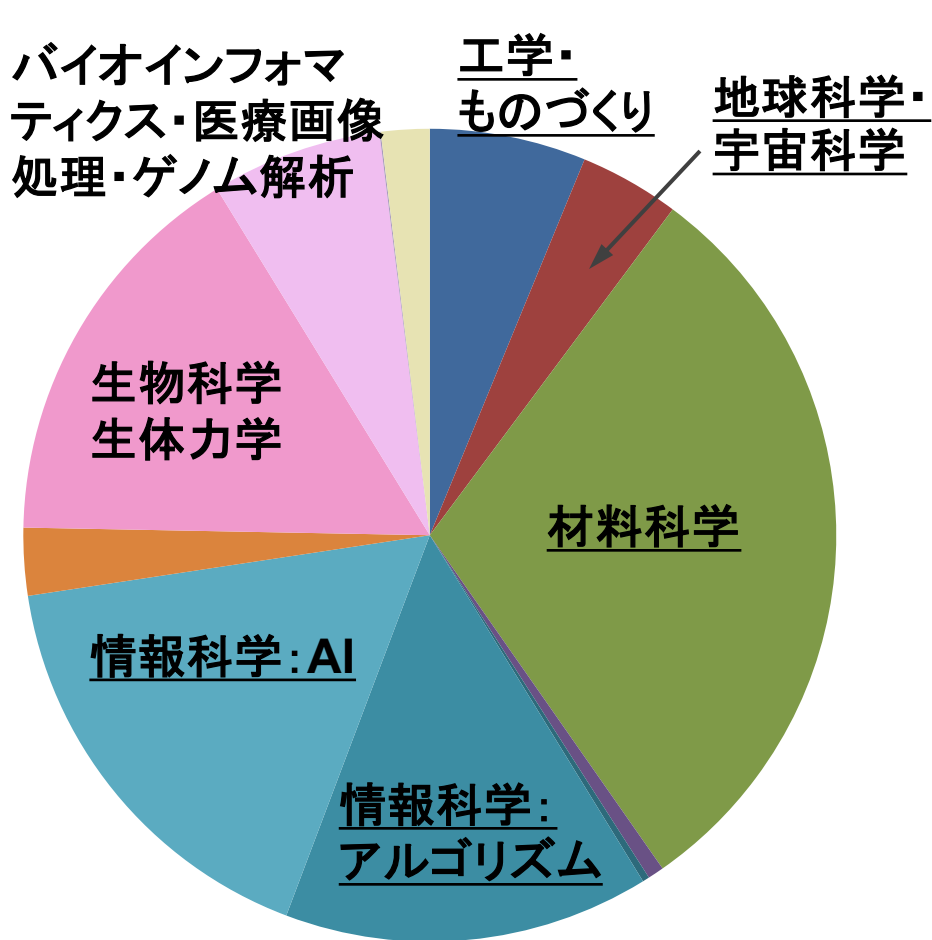
メニコアクラスタ
Intel Xeon Phi
(Oakforest-PACS)



マルチコアクラスタ
Intel CLX
(Oakbridge-CX)

- 工学・ものづくり
- 地球科学・宇宙科学
- 材料科学
- エネルギー・物理学
- 情報科学: システム
- 情報科学: アルゴリズム
- 情報科学: AI
- 教育
- 産業利用
- 生物科学・生体力学
- バイオインフォマティクス
- 社会科学・経済学
- データ科学・データ同化

研究分野別利用CPU時間割合(2020年度)



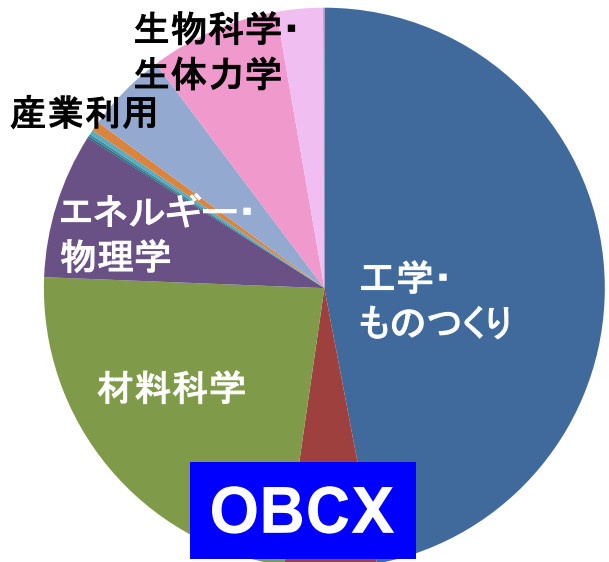
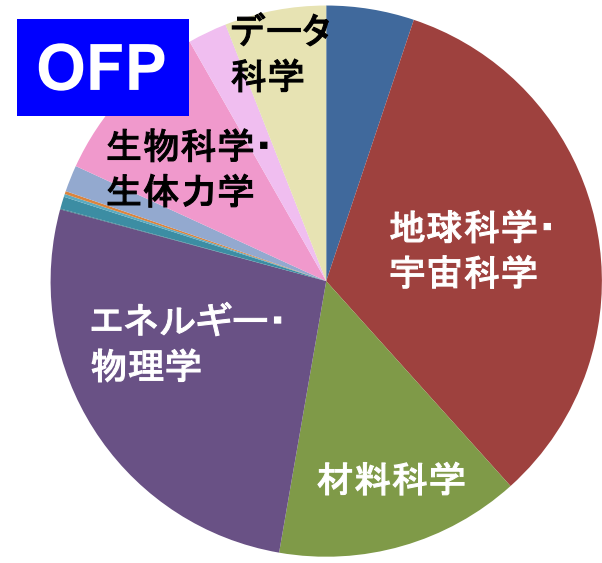
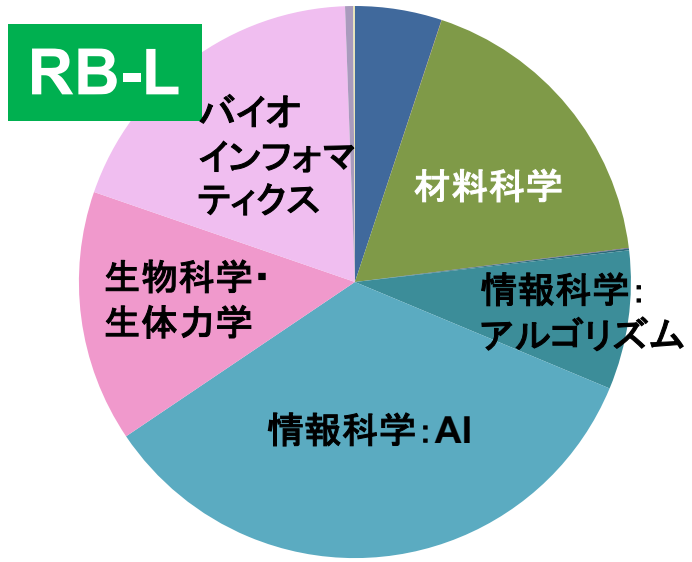
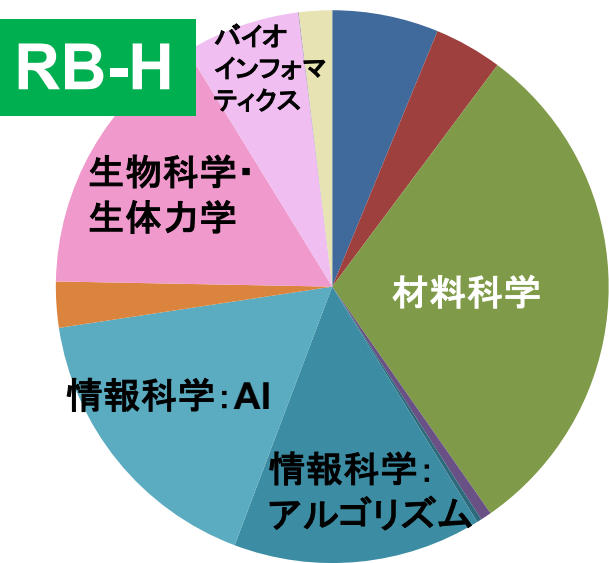
- 工学・ものづくり
- 地球科学・宇宙科学
- 材料科学
- エネルギー・物理学
- 情報科学: システム
- 情報科学: アルゴリズム
- 情報科学: AI
- 教育
- 産業利用
- 生物科学・生体力学
- バイオインフォマティクス
- 社会科学・経済学
- データ科学・データ同化

GPUクラスタ (2GPUs/node)
Intel BDW + NVIDIA P100
(Reedbush-H)

GPUクラスタ (4GPUs/node)
Intel BDW + NVIDIA P100
(Reedbush-L)

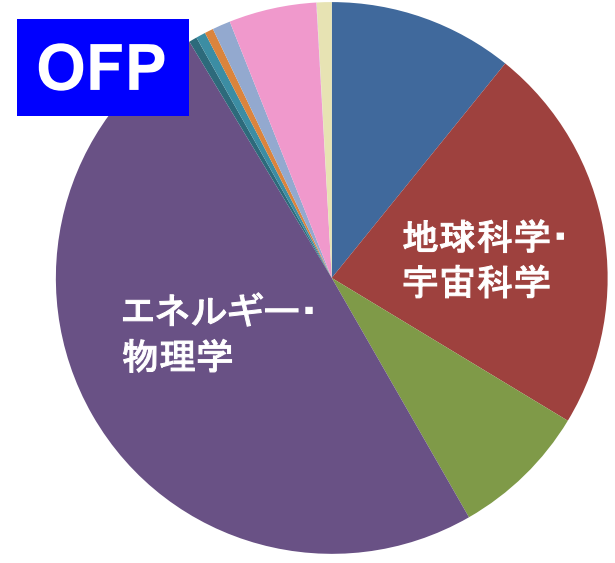
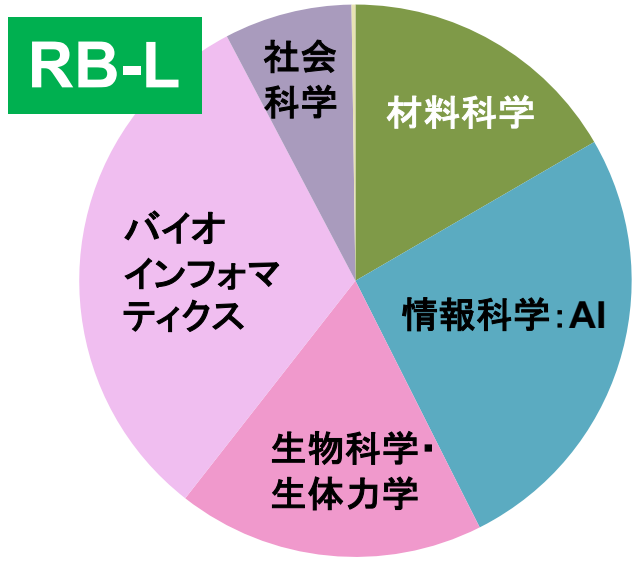
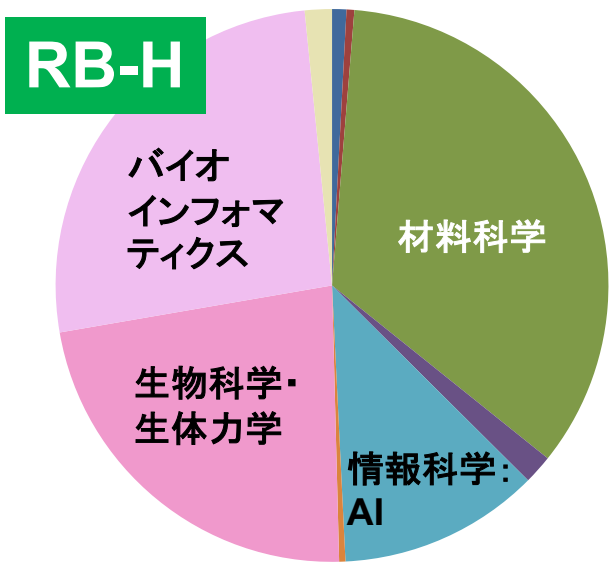
2020年度分野別 ■汎用CPU, ■GPU

- 工学・ものづくり
- 地球科学・宇宙科学
- 材料科学
- エネルギー・物理学
- 情報科学:システム
- 情報科学:アルゴリズム
- 情報科学:AI
- 教育
- 産業利用
- 生物科学・生体力学
- バイオインフォマティクス
- 社会科学・経済学
- データ科学・データ同化

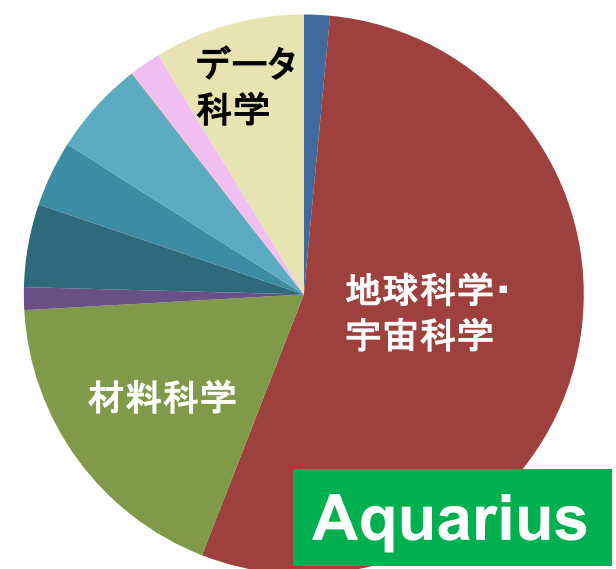
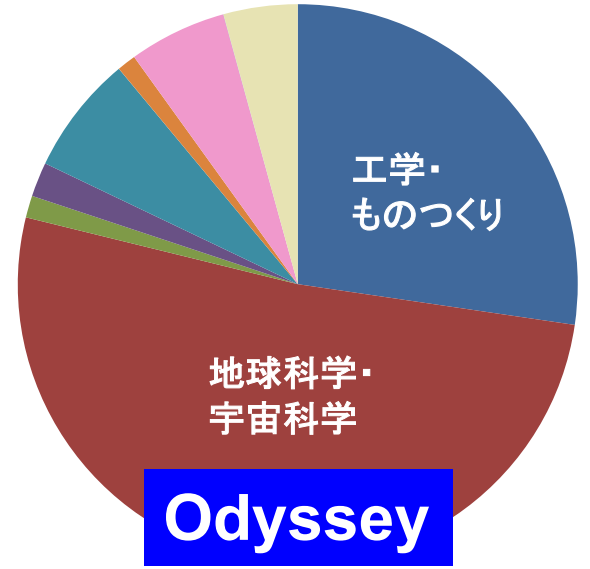
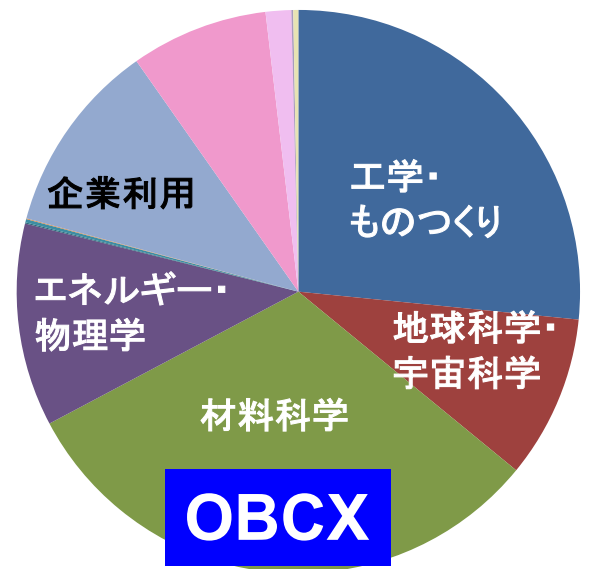


本年度分野別(9月末時点) ■汎用CPU, ■GPU

Odyssey, Aquariusは8月・9月分



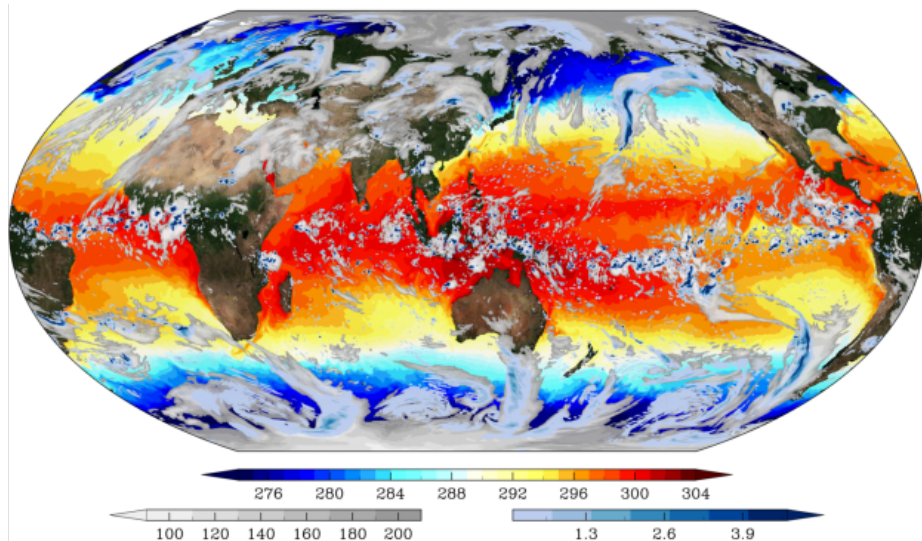
- 工学・ものづくり
- 地球科学・宇宙科学
- 材料科学
- エネルギー・物理学
- 情報科学: システム
- 情報科学: アルゴリズム
- 情報科学: AI
- 教育
- 産業利用
- 生物科学・生体力学
- バイオインフォマティクス
- 社会科学・経済学
- データ科学・データ同化



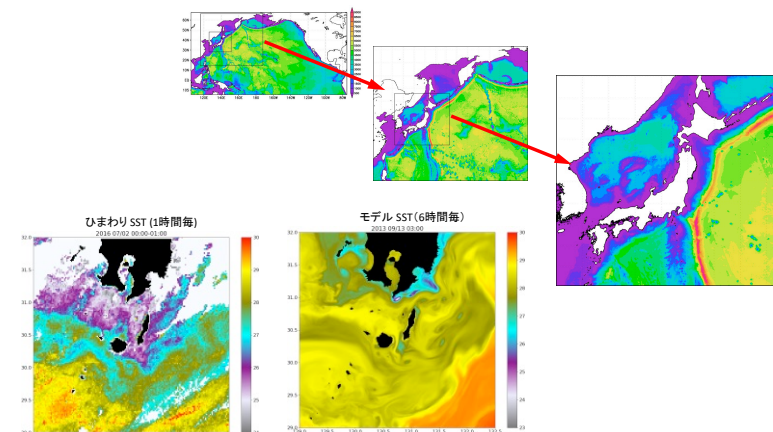
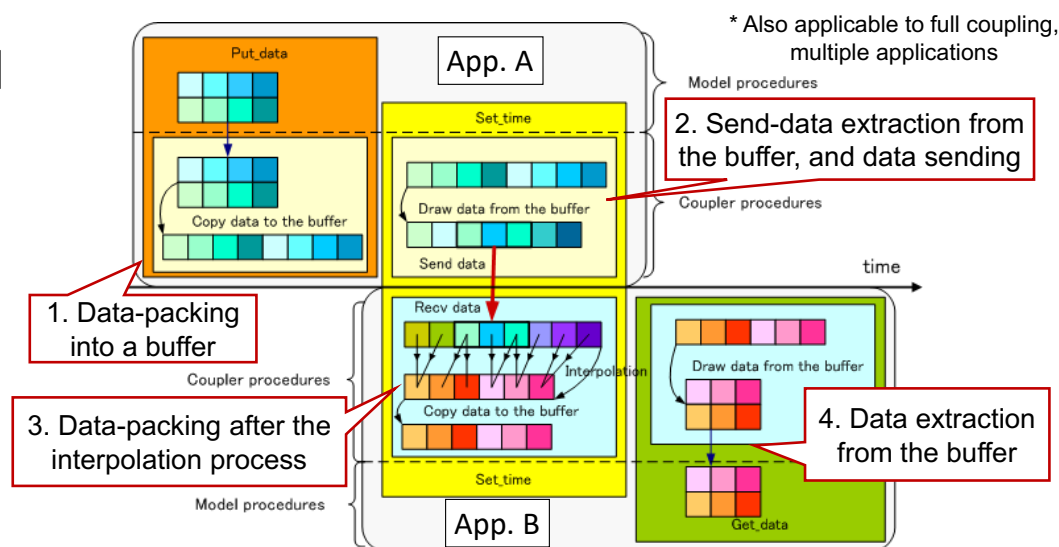
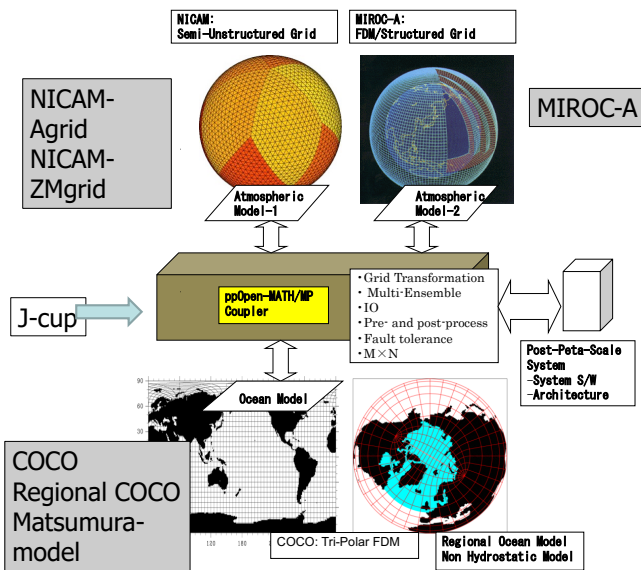
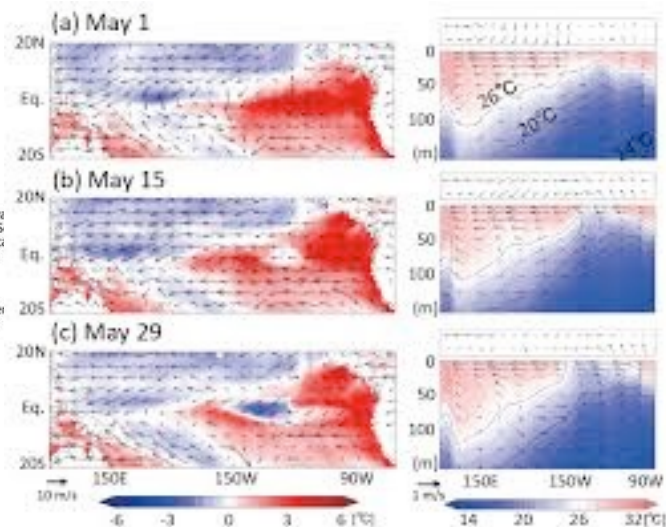
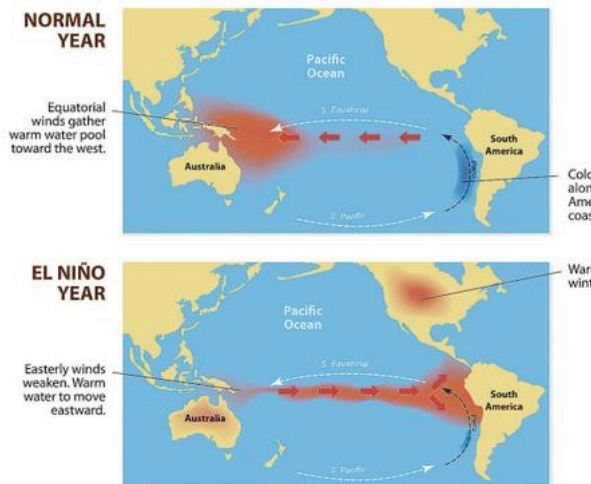
地球科学・宇宙科学分野ではOFP ⇒ Wisteria/BDEC-01 への移行が順調に進んでいる

全地球大気環境シミュレーション

東大大気海洋研究所, 東大理学系研究科等



THE EL NIÑO PHENOMENON

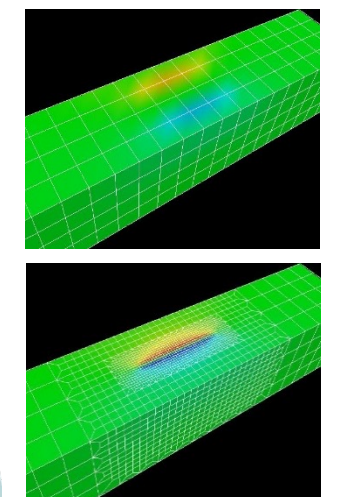
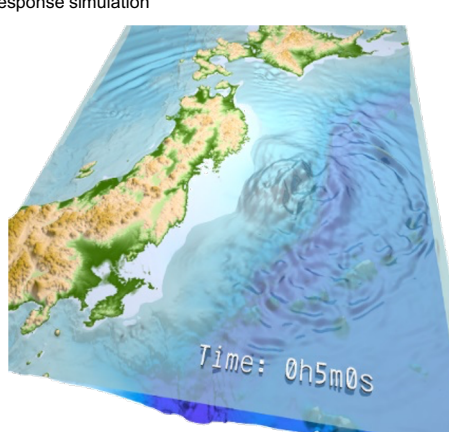
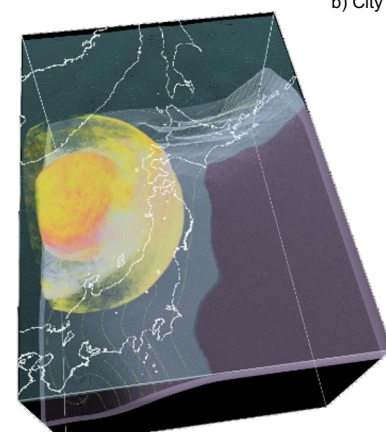
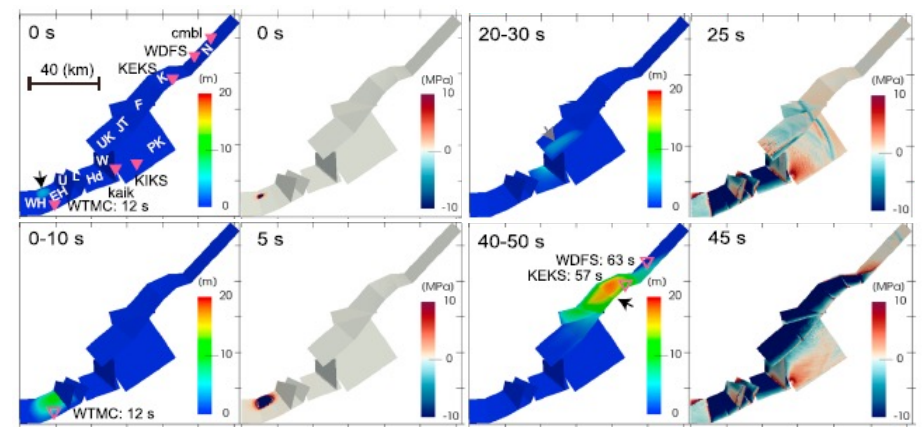
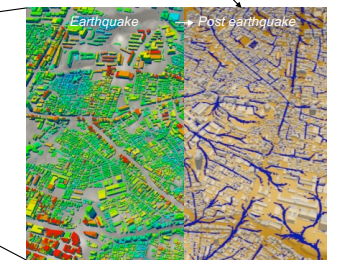
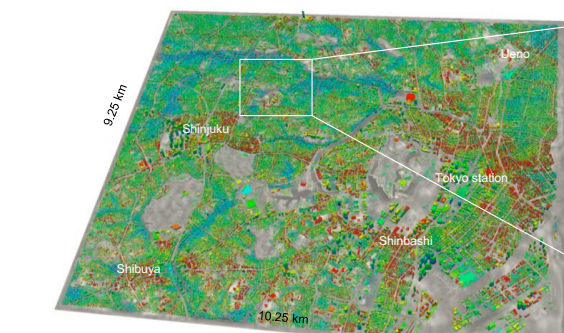
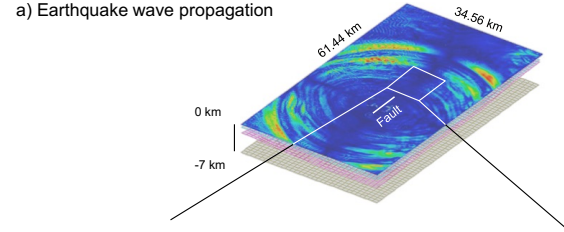
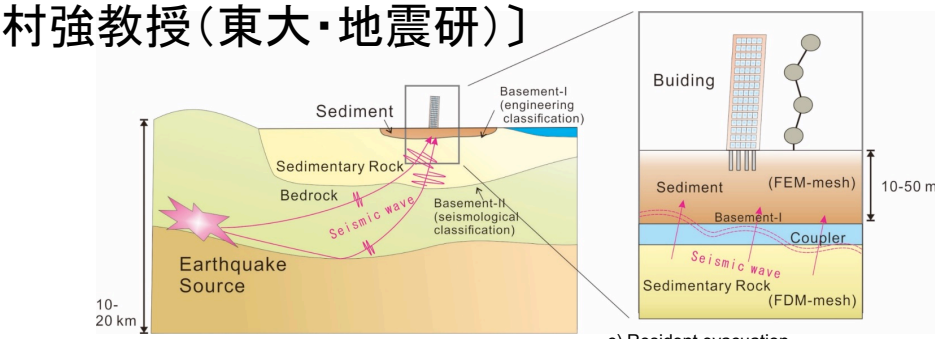
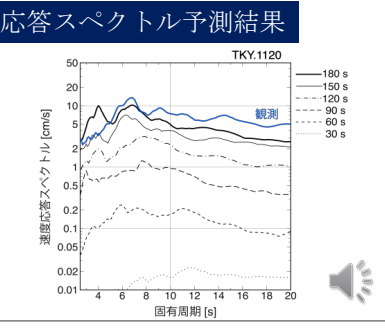
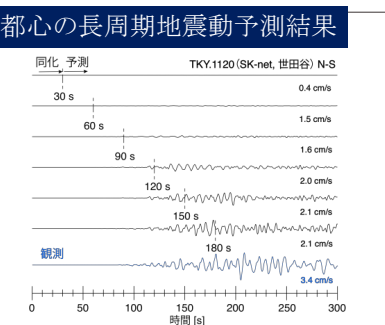
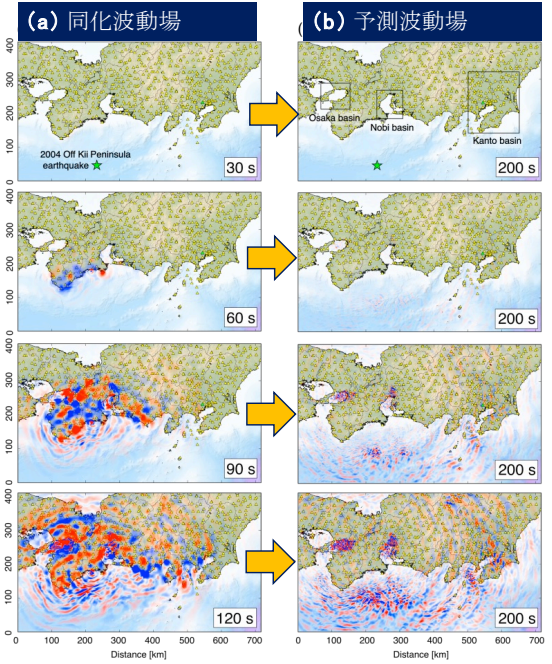
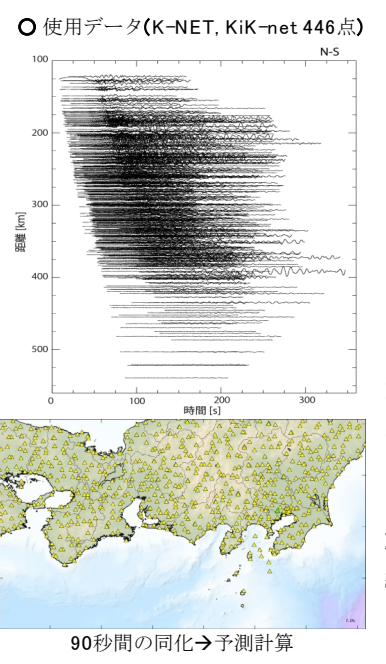


〔画像提供: 佐藤正樹教授・羽角博康教授(東大・大気海洋研)〕

地震シミュレーション・地殻変動

東大地震研究所, 東大理学系研究科等

[画像提供: 古村孝志教授・市村強教授(東大・地震研)]



[画像提供: 安藤亮輔准教授(東大・理学系)]

二酸化炭素地下貯留シミュレーション

大成建設, 理化学研究所等

[画像提供:
山本肇博士(大成建設)]

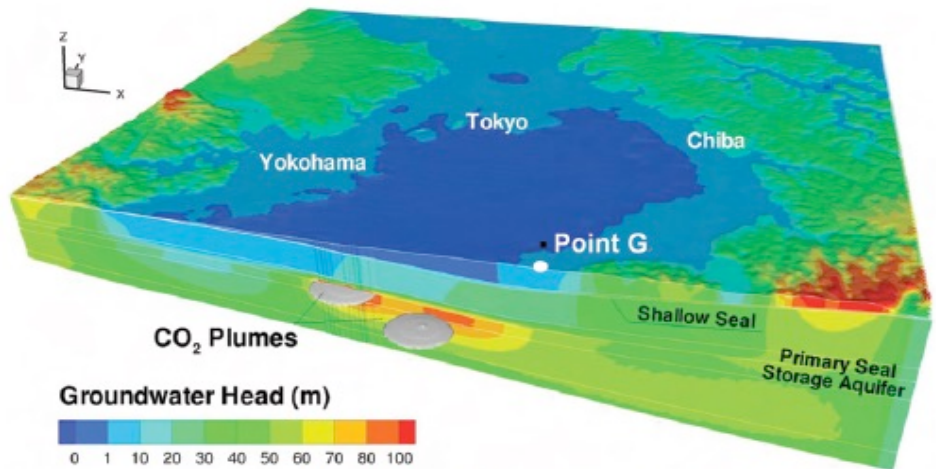
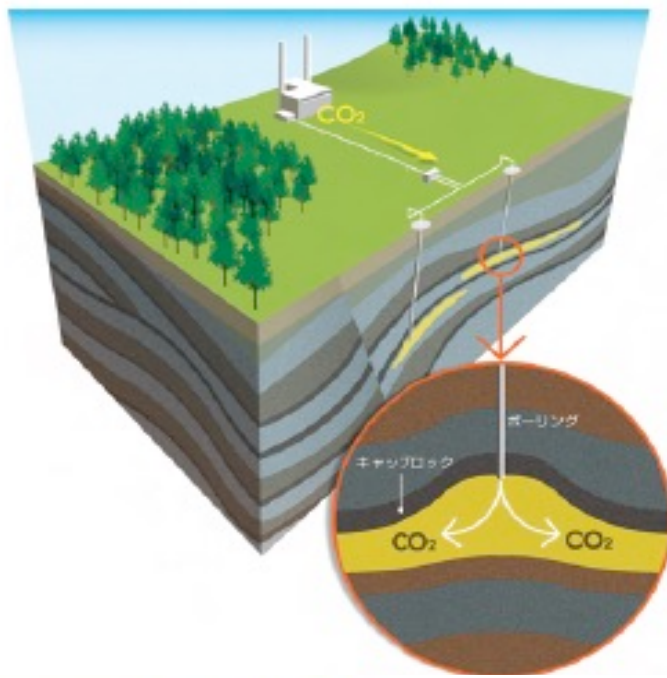
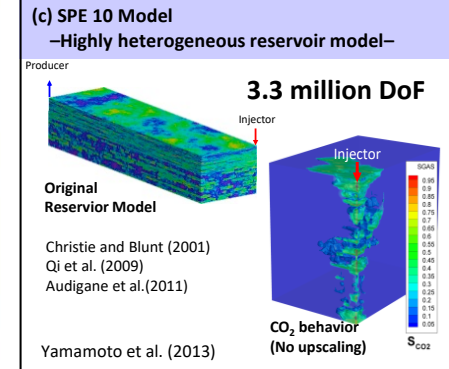
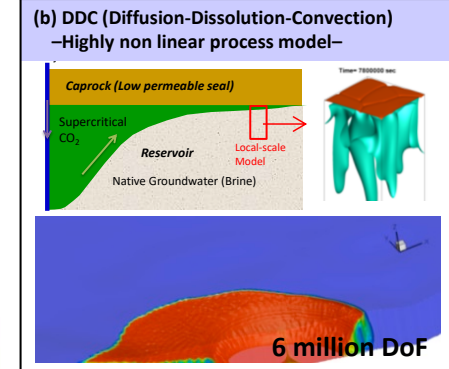
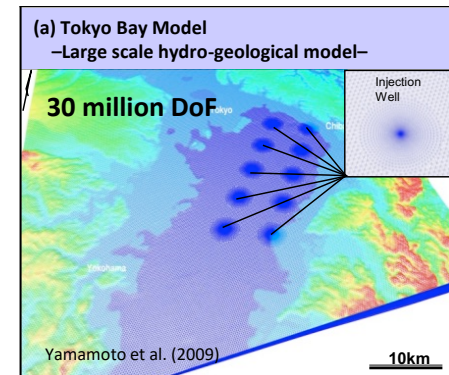
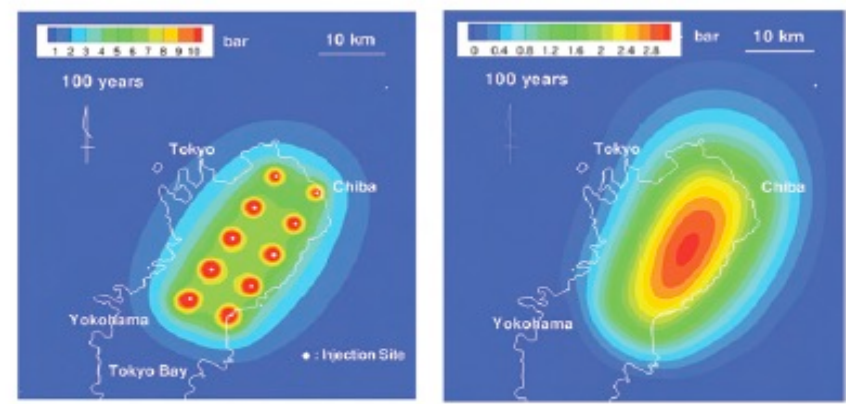
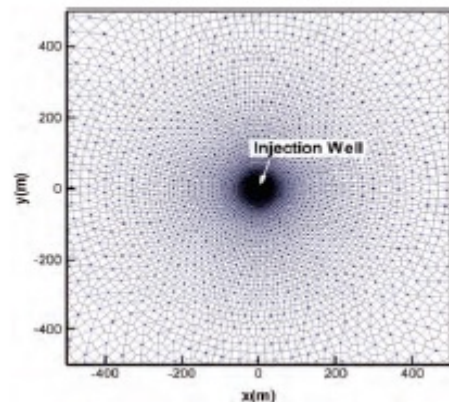
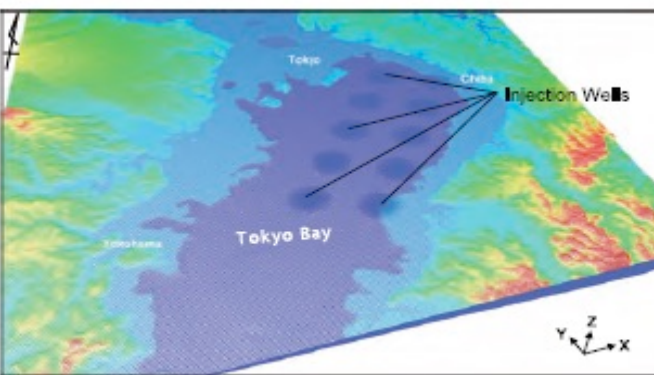


図-4 CO₂ 圧入後の地下水圧 (全水頭換算) の分布 (100 年後)

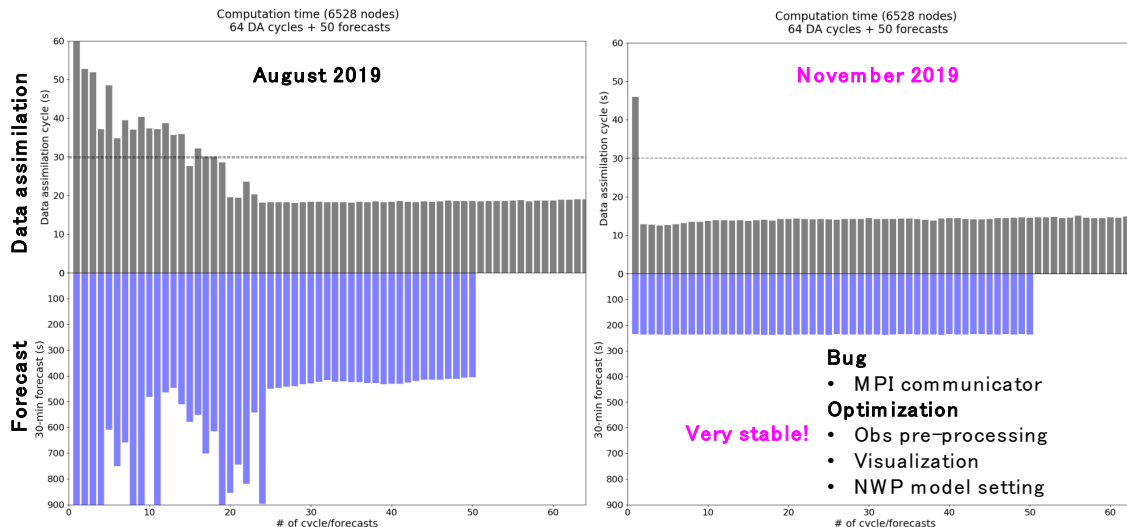
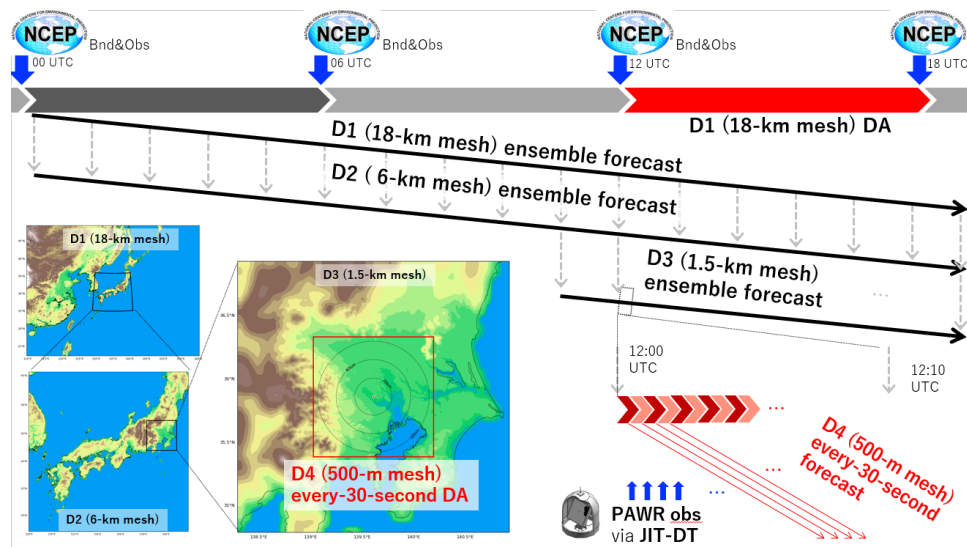


※DOF: degrees of freedom



(a) 深部遮蔽層下面 (b) 浅部遮蔽層下面
図-5 圧力上昇量の平面分布 (初期状態からの増分、圧入開始から 100 年後)

ゲリラ豪雨予測のリアルタイム実証実験 (理化学研究所)



全体のワークフロー

計算性能の向上。上段はデータ同化、下段は30分予報にかかった時間(秒)。(左)2019年8月、(右)2019年11月

15:30:00 15:40:00 15:50:00

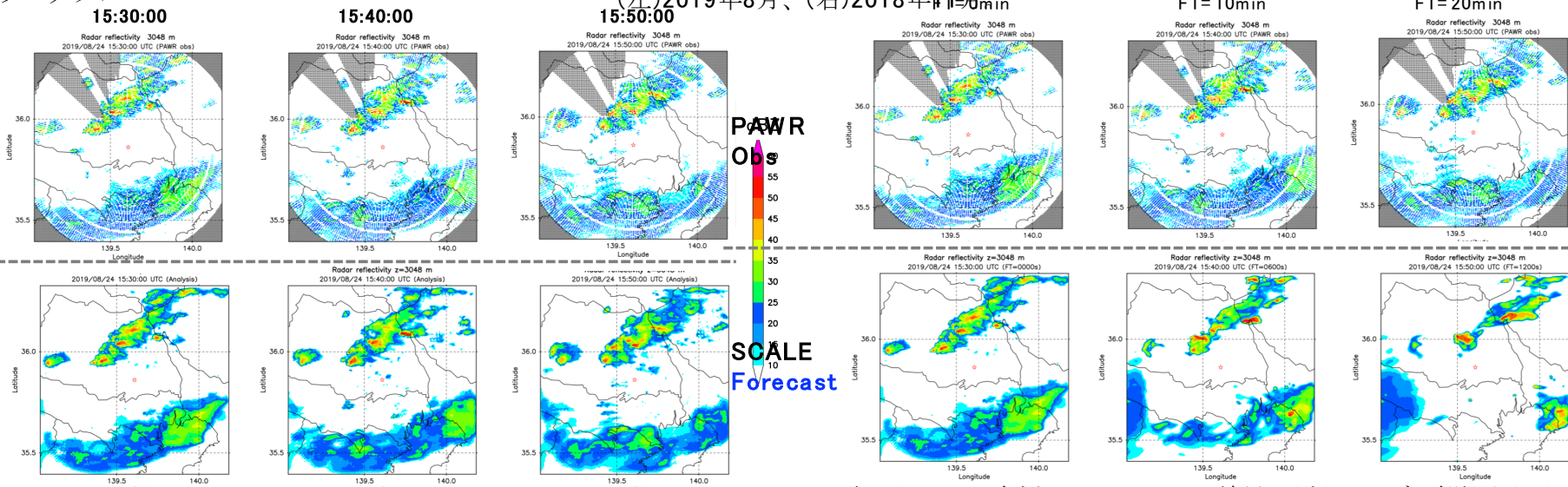
FT= 10min

FT= 20min



PAWR Obs

SCALE-LETKF Analysis



〔画像提供:三好建正博士 (理化学研究所)〕

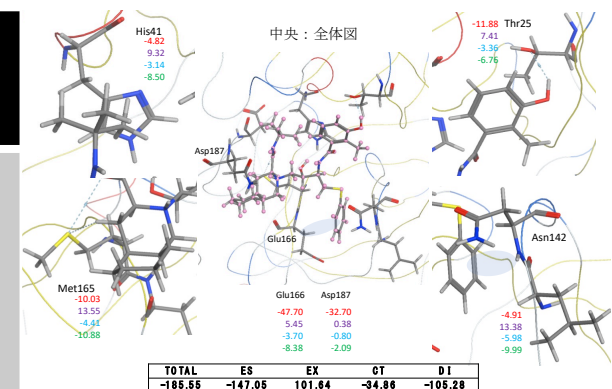
2019年8月24日の事例についてのテスト結果。(上)レーダー観測と(下)SCALE-LETKFによる解析で得られたレーダー反射強度(dBZ)を示す。

「COVID-19」対応HPCI臨時公募課題

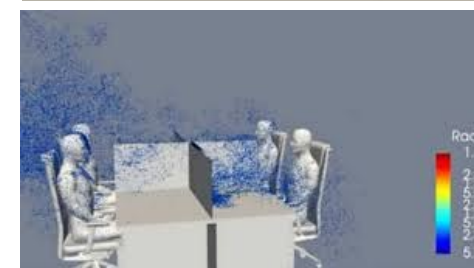
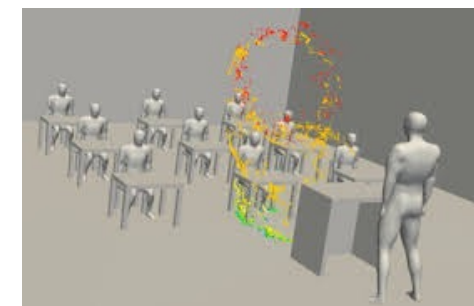
全14のうち6課題が東大システムを利用(2020年度)



課題名	代表者(所属)	使用システム
新型コロナウイルスの主要プロテアーゼに関するフラグメント分子軌道計算	望月 祐志 (立教大学)	Oakforest PACS
COVID-19治療の候補薬: chloroquine、hydroxychloroquine、azithromycinの催不整脈リスクの評価ならびにその低減策に関する研究	久田 俊明(株式会社UT-Heart研究所 / 東大)	
新型コロナウイルス表面のタンパク質動的構造予測	杉田 有治 (理化学研究所)	
計算機解析によるSARS-CoV-2増殖阻害化合物の探索	星野 忠次 (千葉大学)	Oakbridge CX
室内環境におけるウイルス飛沫感染の予測とその対策: 富岳大規模解析に向けたケーススタディ	坪倉 誠 (神戸大学)	
Spreading of polydisperse droplets in a turbulent puff of saturated exhaled air	Marco Edoardo Rosti (OIST)	



資料提供: 望月祐志教授(立教大学)



〔資料提供: 坪倉誠教授(神戸大学)〕

YouTubeチャンネルのご紹介



研究事例紹介や、セミナー・講習会の録画などをご覧になれます。

- 「東京大学情報基盤センター」チャンネル

<https://www.youtube.com/channel/UC2CHaGp1AO-vqRIV7wmU0-w>

- Wisteria/BDEC-01システム紹介

https://www.youtube.com/watch?v=SXjYtatzo-4&list=PLobjSv_ny85IW03OAPUJ9DWJoHhNiQgvY&index=3&t=104s

- 第10回JCAHPCセミナー

https://www.youtube.com/playlist?list=PLobjSv_ny85mfPTuCC2i7r_sPQYKZvy2e

- 柏キャンパス一般公開

https://www.youtube.com/playlist?list=PLobjSv_ny85kr1lg2m-bUiMC2a9W6k53u
<https://www.youtube.com/watch?v=q-0QtU7Ops4&t=116s>

- JCAHPCセミナー:「人類と地球を護るスーパーコンピューティング」

https://www.youtube.com/playlist?list=PLobjSv_ny85l-z-VJCy690ZjIAA04xCRA

- お試しアカウントつき講習会

https://www.youtube.com/playlist?list=PLobjSv_ny85kXY2Mtnhn1k7pM-epQaD2y

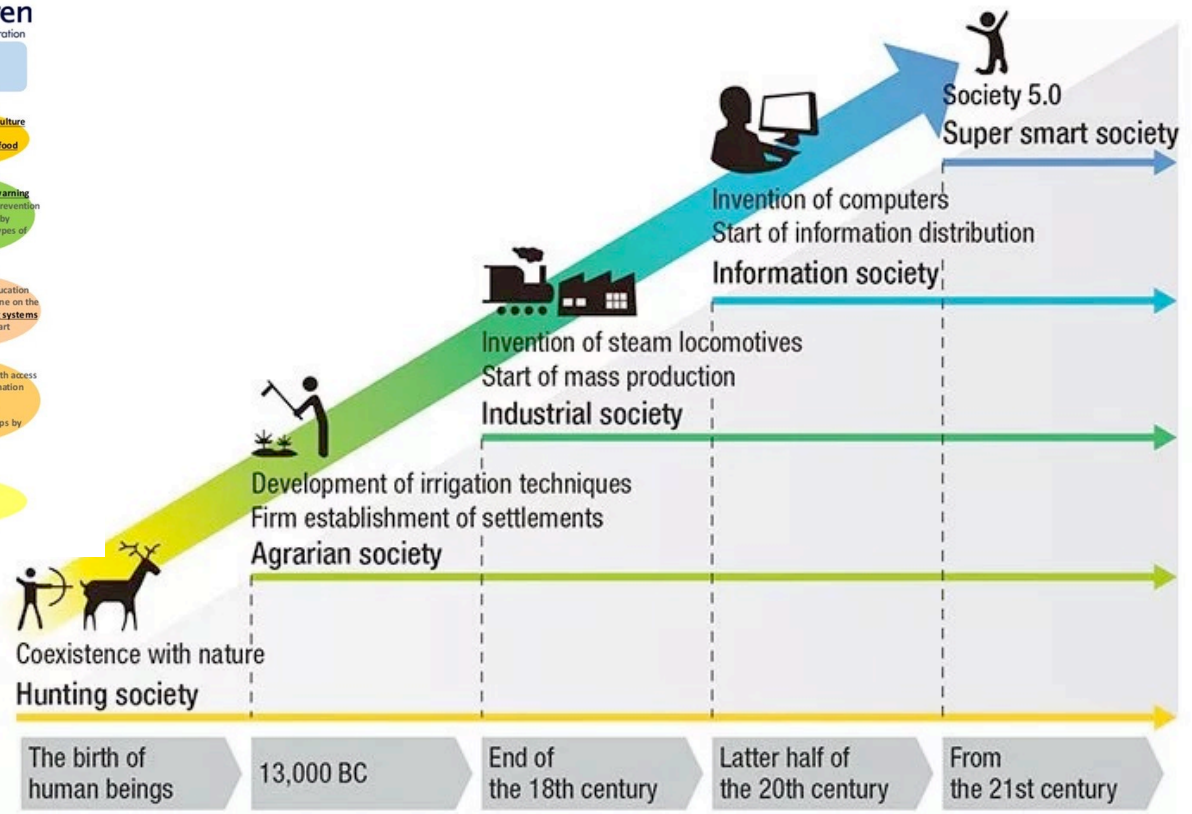
- 東大情報基盤センターのスパコン概要
- **Wisteria/BDEC-01**
- Ipomoea-01
- h3-Open-BDEC

Society 5.0: 日本が提唱する未来社会のコンセプト

デジタル革新・イノベーション（IoT, AI, ビッグデータ等）により、サイバー空間（仮想）とフィジカル空間（現実）を高度に融合させたシステムを構築し、経済発展と社会的課題の解決を両立する、超スマートな人間中心の社会



Economic and social innovation by deepening of Society 5.0

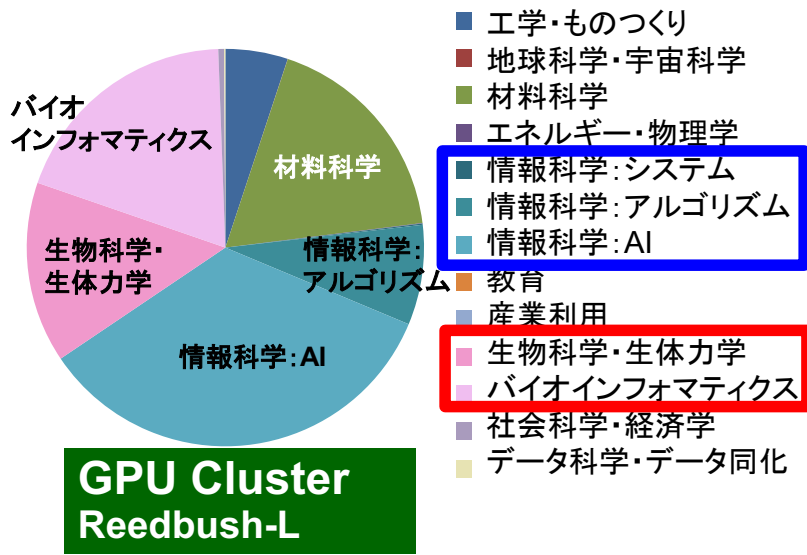
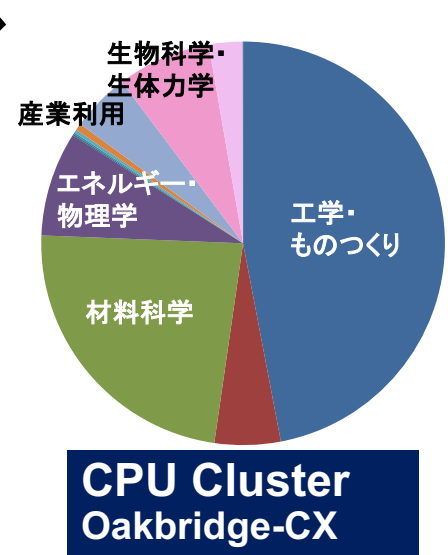


5.0: Super Smart
 4.0: 情報
 3.0: 工業
 2.0: 農耕
 1.0: 狩猟

Source: Prepared based on materials from the Japan Business Federation (Keidanren)

スーパーコンピューティングの今後

- ワークロードの多様化
 - 計算科学, 計算工学: Simulations
 - 大規模データ解析
 - AI, 機械学習



- 工学・ものづくり
- 地球科学・宇宙科学
- 材料科学
- エネルギー・物理学
- 情報科学: システム
- 情報科学: アルゴリズム
- 情報科学: AI
- 教育
- 産業利用
- 生物科学・生体力学
- バイオインフォマティクス
- 社会科学・経済学
- データ科学・データ同化

(シミュレーション(計算) + データ + 学習) 融合 ⇒ Society 5.0 実現に有効

- フィジカル空間とサイバー空間の融合
 - S: シミュレーション(計算) (Simulation)
 - D: データ (Data)
 - L: 学習 (Learning)
- Simulation + Data + Learning = S+D+L**



- 2021年春に柏IIキャンパスで始動
 - BDEC (Wisteria/BDEC-01): 賢いスパコン
 - Data Platform (mdx): Cloud的, よりフレキシブル

BDEC: S + D + L

mdx: s + D + L

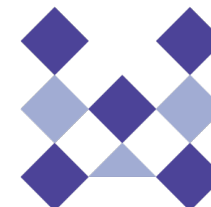
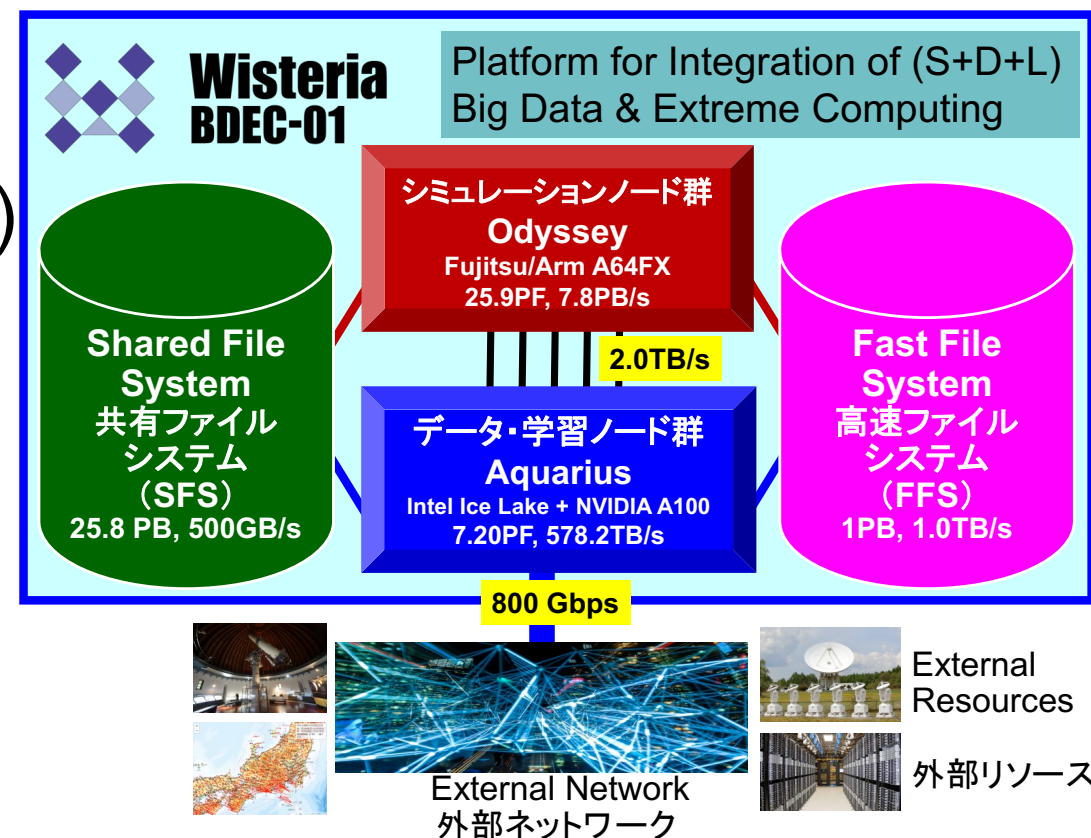
(シミュレーション(計算)+データ+学習)融合 (S+D+L)

- 東大情報基盤センターでは、2015年頃から「(S+D+L)融合」の重要性に注目し、それを実現するためのハードウェア、ソフトウェア、アプリケーション、アルゴリズムに関する研究開発を開始
 - BDEC計画(Big Data & Extreme Computing)
 - 「データ+学習」による「シミュレーション」
- 2021年5月に運用を開始した「Wisteria/BDEC-01」は「BDEC計画」の1号機
 - Reedbush, Oakbridge-CXは「BDEC」のプロトタイプと位置づけられる

Wisteria/BDEC-01

- 2021年5月14日運用開始予定
 - 東京大学柏Ⅱキャンパス
- 33.1 PF, 8.38 PB/sec., **富士通製**
 - ~4.5 MVA(空調込み), ~360m²
- Hierarchical, Hybrid, Heterogeneous (h3)
- **2種類のノード群**
 - シミュレーションノード群(S, SIM) : **Odyssey**
 - 従来のスパコン
 - **Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF**
 - 7,680ノード(368,640 コア), 20ラック, Tofu-D
 - データ・学習ノード群(D/L, DL) : **Aquarius**
 - データ解析, 機械学習
 - **Intel Xeon Ice Lake + NVIDIA A100, 7.2 PF**
 - 45ノード(Ice Lake:90基, A100:360基), IB-HDR
 - 一部は外部リソース(ストレージ, サーバー, センサーネットワーク他)に直接接続
 - ファイルシステム: 共有(大容量) + 高速

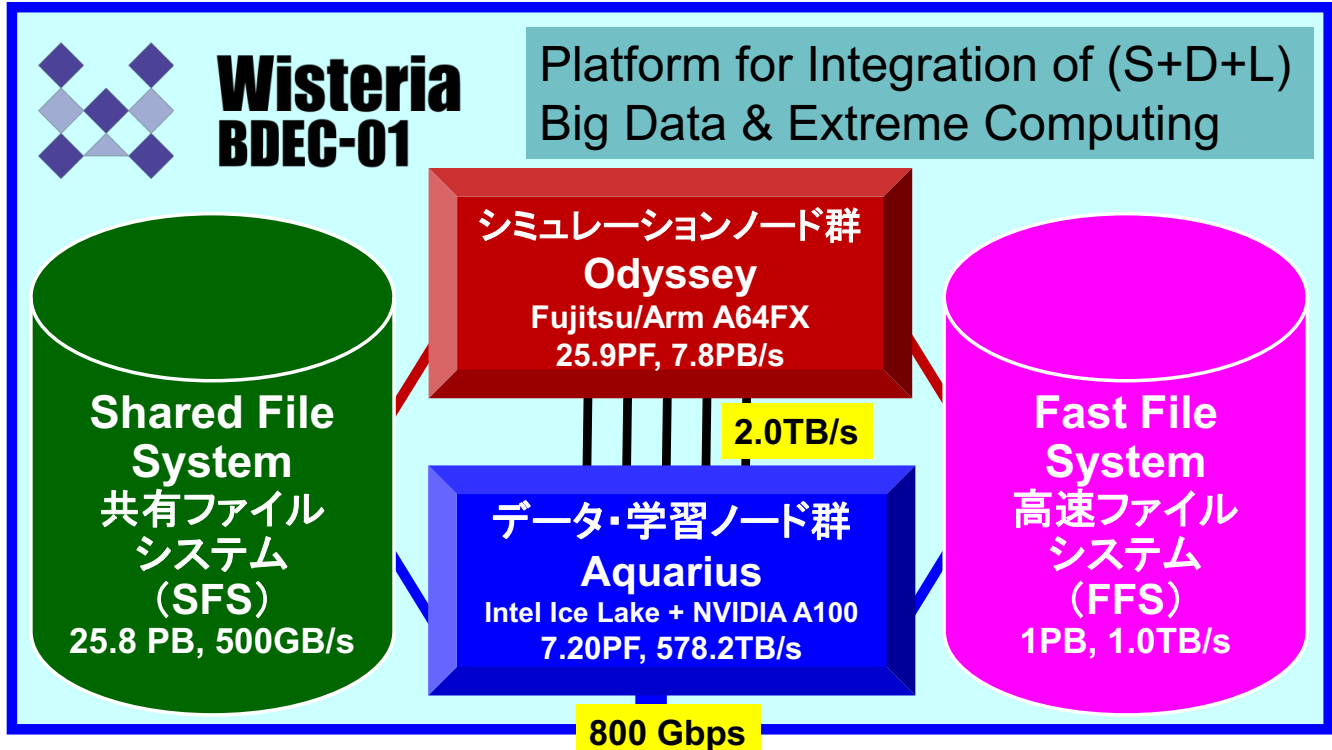
BDEC:「計算・データ・学習(S+D+L)」
融合のためのプラットフォーム
(Big Data & Extreme Computing)



Wisteria
BDEC-01



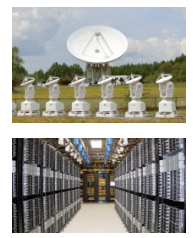
Wisteria/BDEC-01 (S+D+L) 融合プラットフォーム



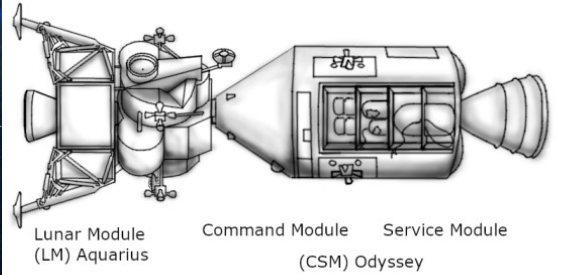
- Wisteria (紫藤)
 - 手賀沼(柏市)に伝わる「藤姫伝説」
- Odyssey
 - アポロ13号・司令船 (Command Module, CM) のコールサイン
- Aquarius
 - アポロ13号・月着陸船 (Lunar Module, LM) のコールサイン
- 人類と地球を護る



External Network
外部ネットワーク



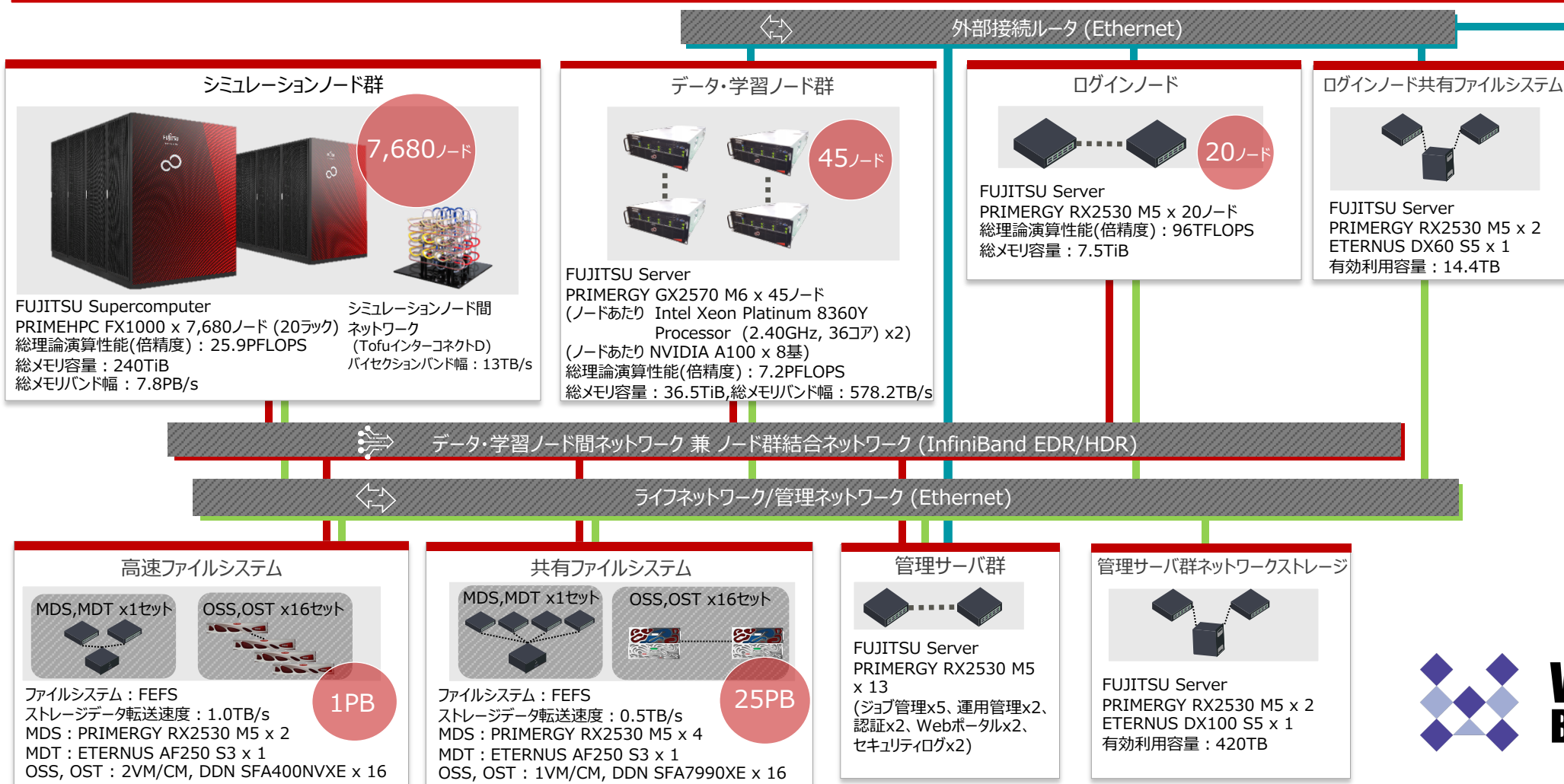
External Resources
外部リソース



システム構成図

シミュレーションノード : 7,680ノード (総理論演算性能 25.9 PFLOPS、総メモリバンド幅 7.8 PB/s)
データ・学習ノード : 45ノード (総理論演算性能 7.2 PFLOPS、総メモリバンド幅 578.2 TB/s)

外部ネットワーク
 ・柏キャンパス
 ・mdxシステム



*The Wisteria/BDEC-01 is a supercomputer system
operated by the Information Technology Center,
The University of Tokyo.*

項目		Wisteria-O (Odyssey)	Wisteria-A (Aquarius)
総理論演算性能		25.9 PFLOPS	7.2 PFLOPS
総ノード数		7,680	45
総主記憶容量		240.0 TiB	36.5 TiB
ネットワークトポロジー		6次元メッシュ / トーラス	Full-bisection Fat Tree
共有ファイルシステム	システム名	FEFS (Fujitsu Exabyte File System)	
	サーバ(OSS)	DDN SFA7990XE	
	サーバ(OSS)数	16	
	ストレージ容量	25.8 PB	
	ストレージデータ転送速度	504 GB/s	
高速ファイルシステム	システム名	FEFS (Fujitsu Exabyte File System)	
	サーバ(OSS)	DDN SFA400NVXE	
	サーバ(OSS)数	16	
	ストレージ容量	1.0 PB	
	ストレージデータ転送速度	1.0 TB/s	

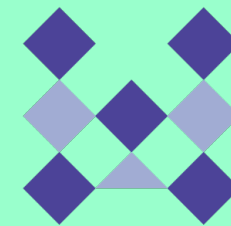
項目		Wisteria-O (Odyssey)	Wisteria-A (Aquarius)
マシン名		FUJITSU Supercomputer PRIMEHPC FX1000	FUJITSU Server PRIMERGY GX2570 M6
CPU	プロセッサ名	A64FX	Intel Xeon Platinum 8360Y (開発コード名: Ice Lake)
	プロセッサ数 (コア数)	1 (48+アシスタントコア2 or 4)	2 (36+36)
	周波数	2.2 GHz	2.4 GHz
	理論演算性能	3.3792 TFLOPS	5.53 TFLOPS
	メモリ容量	32 GB	512 GiB
	メモリ帯域幅	1,024 GB/s	409.6 GB/s
GPU	プロセッサ名	-	NVIDIA A100
	SM数 (単体)		108
	メモリ容量 (単体)		40 GB
	メモリ帯域幅 (単体)		1,555 GB/s
	理論演算性能 (単体)		19.5 TFLOPS
	搭載数		8
	CPU-GPU間接続		PCI Express Gen4 x 16レーン (1レーンあたり片方向32 GB/s)
	GPU間接続		NVLink x 12本 (1本あたり片方向25GB/s)
インターコネク		Tofuインターコネク	InfiniBand HDR(200Gbps) x 4

ソフトウェア群

項目	Wisteria-O (Odyssey)	Wisteria-A (Aquarius)
OS	Red Hat Enterprise Linux 8 (aarch64)	Red Hat Enterprise Linux 8 (x86_64)
コンパイラ	GNU コンパイラ	GNU コンパイラ
	富士通社製 コンパイラ (Fortran77/90/95/2003/2008、C、C++)	Intel コンパイラ(Fortran77/90/95/2003/2008、C、C++) NVIDIA HPC SDK (Fortran77/90/95/2003/2008、C、C++、OpenACC 2.7) NVIDIA CUDA SDK (CUDA C、CUDA C++)
メッセージ通信 ライブラリ	富士通社製MPI	Intel MPI、Open MPI

項目	Wisteria-O (Odyssey)	Wisteria-A (Aquarius)
ライブラリ	SuperLU、 SuperLU MT、 SuperLU DIST、 METIS、 MT-METIS、 ParMETIS、 Scotch、 PT-Scotch、 PETSc、 Trillinos、 FFTW、 GNU Scientific Library、 NetCDF、 Parallel netCDF、 HDF5、 Parallel HDF5、 CMake、 Miniconda、 Xabclib、 ppOpen-HPC、 MassiveThreads、 Boost C++、 mpiJava	
	富士通社製ライブラリ(BLAS、 CBLAS、 LAPACK、 ScaLAPACK)	Intel社製ライブラリ(MKL)(BLAS、 CBLAS、 LAPACK、 ScaLAPACK)、 cuBLAS、 cuSPARSE、 cuFFT、 MAGMA、 cuDNN、 NCCL
アプリケーション	OpenFOAM、 ABINIT-MP、 PHASE、 FrontFlow/blue、 FrontISTR、 REVOCAP-Coupler、 REVOCAP-Refiner、 OpenMX、 MODYLAS、 GROMACS、 BLAST、 R packages、 bioconductor、 BioPerl、 BioRuby、 BWA、 GATK、 SAMtools、 Quantum ESPRESSO、 Xcrypt、 ROOT、 Geant4、 LAMMPS、 CP2K、 NWChem、 DeepVariant、 Paraview、 VisIt、 POV-Ray、 TensorFlow、 Chainer、 PyTorch、 Keras、 Horovod、 MXNet	
		Theano
フリーソフトウェア	autoconf、 automake、 bash、 bzip2、 cvs、 emacs、 findutils、 gawk、 gdb、 make、 grep、 gnuplot、 gzip、 less、 m4、 python、 perl、 ruby、 screen、 sed、 subversion、 tar、 tcsh、 tcl、 vim、 zsh、 git など	
		Globus Toolkit、 Gfarm、 FUSE
コンテナ仮想化	Singularity Community Edition	

技術的な特徴など



Wisteria
BDEC-01

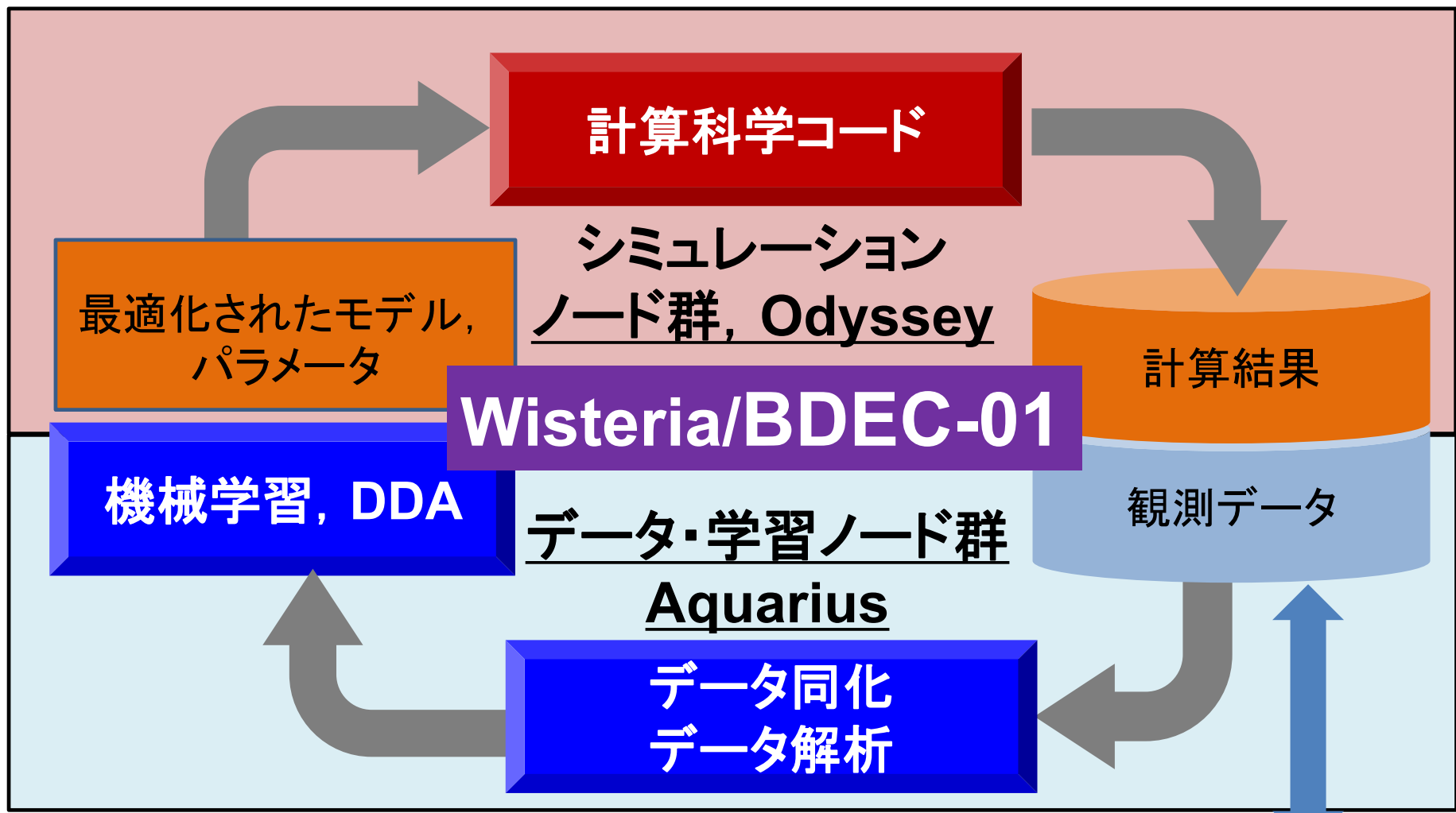
- Odyssey
 - SVE (Scalable Vector Extension)
 - Armv8-A命令セットアーキテクチャをスーパーコンピュータ向けに拡張
 - FP16
 - 機械学習・AIワークロードへの適用
- Aquarius
 - HPC・計算科学への適用
 - CPU: Intel Xeon Ice Lake
 - 3rd Generation Intel Xeon Scalable Processors
 - 推論, 単独での利用は難しいが
 - GPU: NVIDIA A100 Tensor Core
 - Tensor Core + Tensor Float [TF32]
- Odyssey-Aquarius
 - InfiniBand-EDR

Simulation Nodes
Odyssey
 25.9 PF, 7.8 PB/s

Fast File System (FFS)
 1.0 PB, 1.0 TB/s

Shared File System (SFS)
 25.8 PB, 0.50 TB/s

Data/Learning Nodes
Aquarius
 7.20 PF, 578.2 TB/s



Wisteria BDEC-01

サーバー
ストレージ
DB
センサー群
他

外部ネットワーク

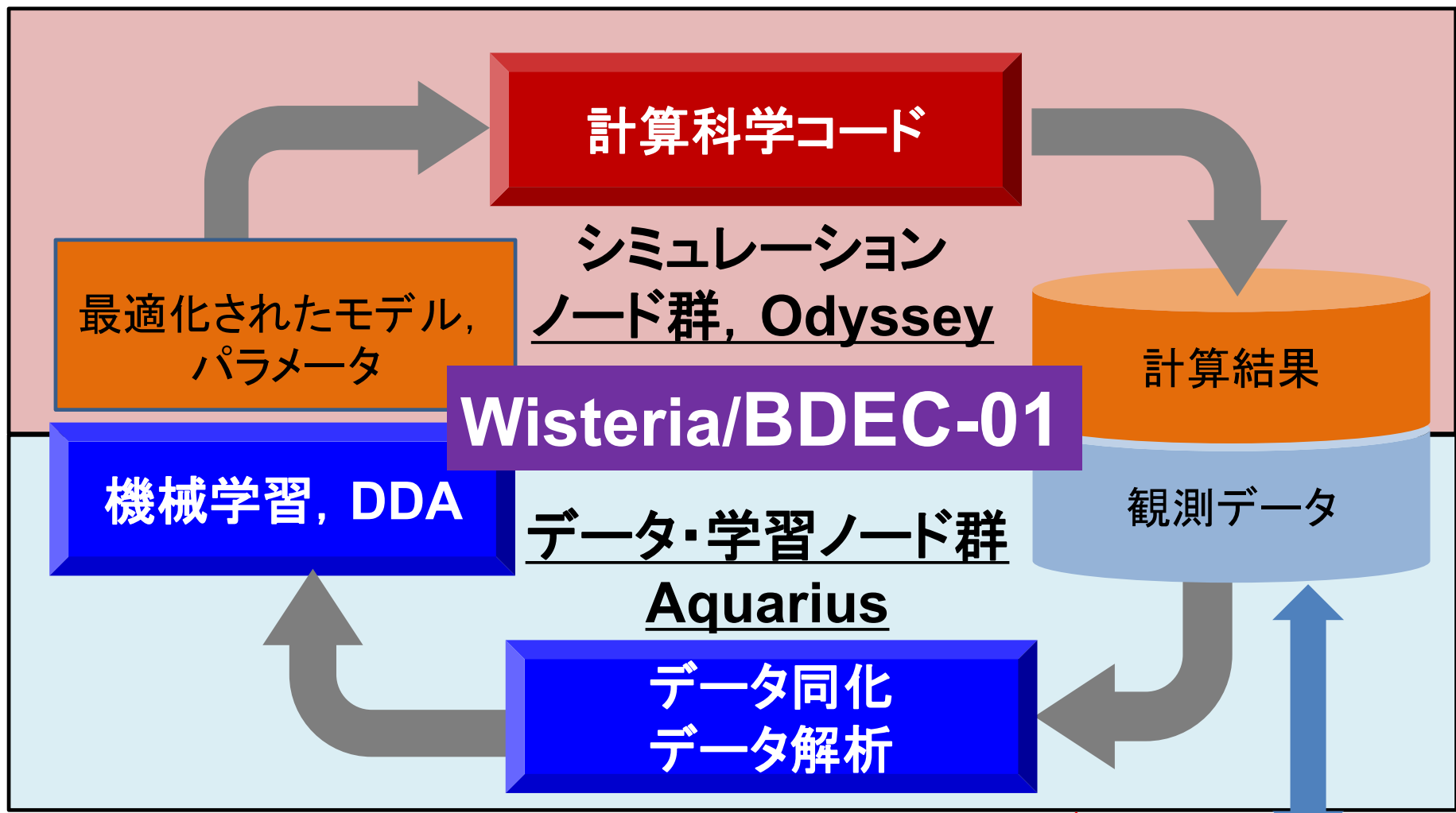
外部リソース

Simulation Nodes
Odyssey
 25.9 PF, 7.8 PB/s

Fast File System (FFS)
 1.0 PB, 1.0 TB/s

Shared File System (SFS)
 25.8 PB, 0.50 TB/s

Data/Learning Nodes
Aquarius
 7.20 PF, 578.2 TB/s



シミュレーションのためのモデル・パラメータのデータ解析, AI/機械学習による最適化 (S+D+L)



更に詳細な情報

- A64FX (富士通)
 - <https://www.fujitsu.com/jp/products/computing/servers/supercomputer/a64fx/>
 - https://old.hotchips.org/hc30/2conf/2.13_Fujitsu_HC30.Fujitsu.Yoshida.rev1.2.pdf
- FUJITSU PRIMEHPC FX1000
 - <https://www.fujitsu.com/jp/products/computing/servers/supercomputer/>
- 3rd Gen Intel Xeon Scalable
 - <https://www.intel.com/content/www/us/en/newsroom/news/3rd-gen-intel-xeon-scalable-video.html#gs.zb3u0m>
 - <https://www.intel.com/content/www/us/en/newsroom/news/3rd-gen-xeon-scalable-processors.html#gs.zb4d00>
 - https://www.hotchips.org/assets/program/conference/day1/HotChips2020_Server_Processors_Intel_Irma_ICX-CPU-final3.pdf
- NVIDIA A100 TENSORコア GPU
 - <https://www.nvidia.com/ja-jp/data-center/a100/>
 - https://www.hotchips.org/assets/program/conference/day1/HotChips2020_GPU_NVIDIA_Choquette_v01.pdf

The article on the Wisteria/BDEC-01 appears in *HPCwire*

<https://www.hpcwire.com/2021/02/25/japan-to-debut-integrated-fujitsu-hpc-ai-supercomputer-this-spring/>

The screenshot shows a web browser window displaying an article on the HPCwire website. The article title is "Japan to Debut Integrated Fujitsu HPC/AI Supercomputer This Spring" by Tiffany Trader, dated February 25, 2021. The article text describes the Wisteria/BDEC-01 supercomputer project at the University of Tokyo. The browser interface includes a search bar, social media sharing options, and a sidebar with various application icons.



参考リンク(ビデオ)

- Wisteria/BDEC-01利用説明会
 - <https://www.youtube.com/watch?v=1bbZVO6-UQg>
- h3-Open-BDEC:プロジェクトHP(工事中)
 - <http://nkl.cc.u-tokyo.ac.jp/h3-Open-BDEC/>
- Wisteria/BDEC-01 & h3-Open-BDEC紹介講演(日本語)
 - https://www.youtube.com/watch?v=CsJ_9aGNXCg
 - <https://www.pccluster.org/ja/event/pccc20/exhibition/itc-u-tokyo.html>
- Wisteria/BDEC-01 & h3-Open-BDEC紹介講演(英語)
 - <https://www.youtube.com/watch?v=jX51NF2LniE>



Wisteria
BDEC-01



GFLOPS (ピーク性能) 当たり利用負担 (円) : 電気代 GFLOPS/W (Green 500)

System	JPY/GFLOPS Small is Good	GFLOPS/W Large is Good
Oakleaf-FX/Oakbridge-FX (Fujitsu) (Fujitsu SPARC64 IXfx)	125	0.866
Reedbush-U (HPE) (Intel Xeon Broadwell (BDW))	61.9	2.310
Reedbush-H (HPE) (Intel BDW+NVIDIA P100x2/node)	15.9	8.575
Reedbush-L (HPE) (Intel BDW+NVIDIA P100x4/node)	13.4	10.167
Oakforest-PACS (Fujitsu) (Intel Xeon Phi/KNL)	16.5	4.986
Oakbridge-CX (Fujitsu) (Intel Xeon Cascade Lake)	20.7	5.076
Wisteria-Odyssey (Fujitsu/Arm A64FX)	17.8	15.07
Wisteria-Aquarius (Intel Xeon Ice Lake + NVIDIA A100x8)	9.00	24.06

November 2021 (SC21)の諸ランキング

Wisteria/BDEC-01のシミュレーションノード群 (Odyssey) とデータ・学習ノード群 (Aquarius) は別々に測定・申請

System	TOP500	Green500	HPCG	Graph500	HPL-AI
Oakforest-PACS	39	65	23	-	-
Oakbridge-CX	110	62	71	-	-
Wisteria/BDEC-01 (Odyssey)	17	27	9	3	9
Wisteria/BDEC-01 (Aquarius)	106	15	58	-	-

各種ベンチマーク(1/2)

- TOP 500 (Linpack, HPL)
 - 連立一次方程式ソルバー(直接法), 計算速度(FLOPS値)
 - 規則的な密行列: 連続メモリアクセス
 - 計算性能
- HPCG
 - 連立一次方程式ソルバー(反復法), 計算速度(FLOPS値)
 - 有限要素法から得られる疎行列(ゼロが多い)
 - 不連続メモリアクセス
 - 実アプリケーションに近い
 - メモリアクセス性能, 通信性能
- Green 500
 - HPL (TOP500) 実行時のFLOPS/W値

58th TOP500 List (Nov., 2021)

R_{\max} : Performance of Linpack (TFLOPS)

<http://www.top500.org/>

R_{peak} : Peak Performance (TFLOPS), Power: kW

	Site, Year	Computer, Vendor	Cores	R_{\max} (TFLOPS)	R_{peak} (TFLOPS)	Power (kW)
1	<u>Fugaku, 2020, Japan</u> R-CCS, RIKEN	Fujitsu PRIMEHPC FX1000, Fujitsu A64FX 48C 2.2GHz, Tofu-D	7,630,848	442,010 (= 442.0 PF)	537,212.0	29,899
2	<u>Summit, 2018, USA</u> DOE/SC/Oak Ridge National Laboratory	IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR InfiniBand	2,414,592	148,600	200,795	10,096
3	<u>Sierra, 2018, USA</u> DOE/NNSA/LLNL	IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR InfiniBand	1,572,480	94,640	125,712	7,438
4	<u>Sunway TaihuLight, 2016, China</u> National Supercomputing Center in Wuxi	Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway	10,649,600	93,015	125,436	15,371
5	<u>Perlmutter, 2021, USA</u> DOE/NERSC/LBNL	HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10	761,856	70,870	94,750	2,589
6	<u>Selene, 2020, USA</u> NVIDIA	NVIDIA DGX A100 SuperPOD, AMD EPYC 7742 64C 2.25GHz, NVIDIA GA100, Mellanox Infiniband HDR	555,520	63,460	79,215.0	2,646
7	<u>Tianhe-2A, 2018, China</u> National Super Computer Center in Guangzhou	TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000	4,981,760	61,445	100,679	18,482
8	<u>JUWELS Booster Module, 2020, Germany</u> Julich (FZJ)	Bull Sequana XH2000, AMD EPYC 7402 24c 2.8GHz, NVIDIA A100, Mellanox InfiniBand HDR	449,280	44,120	70,980	1,764
9	<u>HPC5, 2020, Italy</u> Eni S.p.A.	Dell C4140, Xeon Gold 6252 24c 2.1GHz, NVIDIA Volta GV100, Mellanox InfiniBand HDR	669,760	35,450	51,720	2,252
10	<u>Voyager-EUS2, 2021, USA</u> Azure East US 2	ND96amsr_A100_v4, AMD EPYC 7V12 48C 2.45GHz, NVIDIA A100 80GB, Mellanox HDR Infiniband	253,440	30,050	39,531	
16	<u>ABCI 2.0, 2021, Japan</u> AIST	Fujitsu PRIMERGY GX2570 M6, Xeon Platinum 8360Y 36C 2.4GHz, NVIDIA A100 SXM4 40 GB, InfiniBand HDR	504,000	22,208	54,341	1,600
17	<u>Wisteria/BDEC-01 (Odyssey), 2021, Japan</u> ITC, University of Tokyo	Fujitsu PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu interconnect D	368,640	22,121	25,952	1,468

Green 500 Ranking (Nov., 2021)

	TOP 500 Rank	System	Accelerator	Cores	HPL Rmax (Pflop/s)	Power (kW)	GFLOPS/W
1	301	MN-3, Preferred Networks, Japan	MN-Core	1,664	2.181	55.39	39.38
2	291	SSC-21 Scalable Module	NVIDIA A100 80GB	16,704	2.274	103.01	*33.98
3	295	Tethys	NVIDIA A100 80GB	19,840	2.255	71.50	*31.54
4	280	Wilkes-3, U. Cambridge, UK	NVIDIA A100 80GB	26,880	2.287	74.26	30.80
5	30	HiPerGator AI	NVIDIA A100	138,880	17.200	582.63	*29.52
6	403	Snellius Phase 1 GPU	NVIDIA A100	6,480	1.818	62.59	29.05
7	5	Perlmutter	NVIDIA A100	761,856	70.870	2,589	27.373
8	44	Karolina, GPU partition	NVIDIA A100	71,424	6.752	310.99	*27.213
9	45	MeluXina - Accelerator Module	NVIDIA A100	99,200	10.520	390.25	*26.96
10	262	NVIDIA DGX SuperPOD, USA	NVIDIA A100	19,840	2.356	90	26.195
15	106	Wisteria/BDEC-01 (Aquarius), Fujitsu, Japan	NVIDIA A100	42,120	4.425	183.93	24.06

HPCG Ranking (Nov., 2021)

	Computer	Cores	HPL Rmax (Pflop/s)	TOP500 Rank	HPCG (Pflop/s)
1	Fugaku	7,630,848	442.010	1	16.004
2	Summit	2,414,592	148.600	2	2.926
3	Perlmutter	761,856	70.870	5	1.905
4	Sierra	1,572,480	94.640	3	1.796
5	Selene	555,520	63.460	6	1.622
6	JUWELS Booster Module	449,280	44.120	8	1.275
7	Dammam-7	672,520	22.400	15	0.881
8	HPC5	669,760	35.450	9	0.860
9	Wisteria/BDEC-01 (Odyssey)	368,640	22.121	17	0.817
10	Earth Simulator -SX-Aurora TSUBASA	43,776	9.990	48	0.747
23	Oakforest-PACS	556,104	13.555	39	0.385

各種ベンチマーク(2/2)

- Graph 500 BFS
 - グラフ処理におけるデータ処理効率
 - BFS (Breadth-First-Search, 幅優先探索)
- HPL-AI
 - AI・機械学習向けHPL (for TOP500)
 - 科学技術計算 (HPL) ⇒ 倍精度演算 (FP64)
 - HPL-AI ⇒ 低精度演算を含む混合精度演算

HPL-AI (Nov. 2021)

R_{\max} : Performance of Linpack (TFLOPS)

低精度演算利用による性能向上

	Site	Computer/Year Vendor	Cores	HPL-AI (EFLOPS)	Top500	HPL R_{\max} (TFLOPS)	Speedup
1	<u>Fugaku, 2020, Japan</u> R-CCS, RIKEN	Fujitsu PRIMEHPC FX1000, Fujitsu A64FX 48C 2.2GHz, Tofu-D	7,630,848	2.0	1	442,010 (= 442.0 PF)	4.5
2	<u>Summit, 2018, USA</u> DOE/SC/Oak Ridge National Laboratory	IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR InfiniBand	2,414,592	1.411	2	148,600	9.5
3	<u>Selene, 2020, USA</u> NVIDIA	NVIDIA DGX A100 SuperPOD, AMD EPYC 7742 64C 2.25GHz, NVIDIA GA100, Mellanox Infiniband HDR	555,520	0.63	6	63,460	9.9
4	<u>Perlmutter, 2021, USA</u> DOE/NERSC/LBNL	HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10	761,856	0.59	5	70,870	8.3
5	<u>JUWELS Booster Module, 2020, Germany</u> Julich (FZJ)	Bull Sequana XH2000, AMD EPYC 7402 24c 2.8GHz, NVIDIA A100, Mellanox InfiniBand HDR	449,280	0.47	8	44,120	10
6	<u>HiPerGator, 2021, USA</u> U. of Florida	NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Infiniband HDR	138,880	0.17	30	17,200	9.9
7	<u>Christofari Neo, 2021, Russia</u> SberCloud	NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100 80GB, Infiniband HDR	98,208	0.123	43	11,950	10.3
8	<u>Polaris, 2021, USA</u> DOE/SC/Argonne National Laboratory	NVIDIA DGX A100, AMD EPYC 7532 32C 2.4GHz, NVIDIA A100, Slingshot-10	259,840	0.114	12	23,840	4.8
9	<u>Wisteria/BDEC-01 (Odyssey), 2021, Japan</u> ITC, University of Tokyo	Fujitsu PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu interconnect D	368,640	0.10	17	22,121	4.5
10	<u>Berzelius, 2021, Sweden</u>	NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, A100, Infiniband HDR, Atos	59,520	0.05	95	44,120	9.9

Graph500 BFS, Nov. 2021

	Site	Computer/ Vendor	Cores	Scale	GTEPS
1	<u>Fugaku, 2020, Japan</u> R-CCS, RIKEN	Fujitsu PRIMEHPC FX1000, Fujitsu A64FX 48C 2.2GHz, Tofu-D	7,630,848	41	102956
2	<u>Sunway TaihuLight, 2016, China</u> National Supercomputing Center in Wuxi	Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway	10,599,680	40	23755.7
3	<u>Wisteria/BDEC-01 (Odyssey), 2021, Japan</u> ITC, University of Tokyo	Fujitsu PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu interconnect D	368,640	37	16118
4	<u>Toki-Sora, 2021, Japan</u> JAXA	PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu interconnect D	276,480	36	10813
5	<u>LUMI-C, 2021, Finland</u> EuroHPC/CSC	Cray EX, SlingShot-10	190,976	38	8467.7
6	<u>Summit (CPU Only), 2018, USA</u> DOE/SC/Oak Ridge National Laboratory	IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR InfiniBand	2,414,592	40	7665.7
7	<u>SuperMUC, 2018, Germany</u> Leibniz	ThinkSystem SD650, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path	196,608	39	6279.47
8	<u>Lise, 2021, Germany</u> ZIB	Bull Intel Cluster Intel Xeon Platinum 9242 48C 2.3GHz Intel Omni-Path	121,920	38	5423.94
9	<u>Cori - 1024 haswell partition, 2017, USA</u> NERSC	Cray XC40	32,768	37	2562.16
10	<u>Tianhe-2 (MilkyWay-2), 2013, China</u>	MPP	196,608	36	2061.48

一般利用コース

<https://www.cc.u-tokyo.ac.jp/supercomputer/wisteria/service/course.php>

- グループコース(1人またはそれ以上から構成されるグループ)

- Wisteria/BDEC-01から「パーソナルコース」は廃止
- 代表者は大学・公共機関所属者



- トークン(ノード時間)を購入

- Odyssey, Aquariusを利用できる
- O/Aでそれぞれ消費係数が異なる

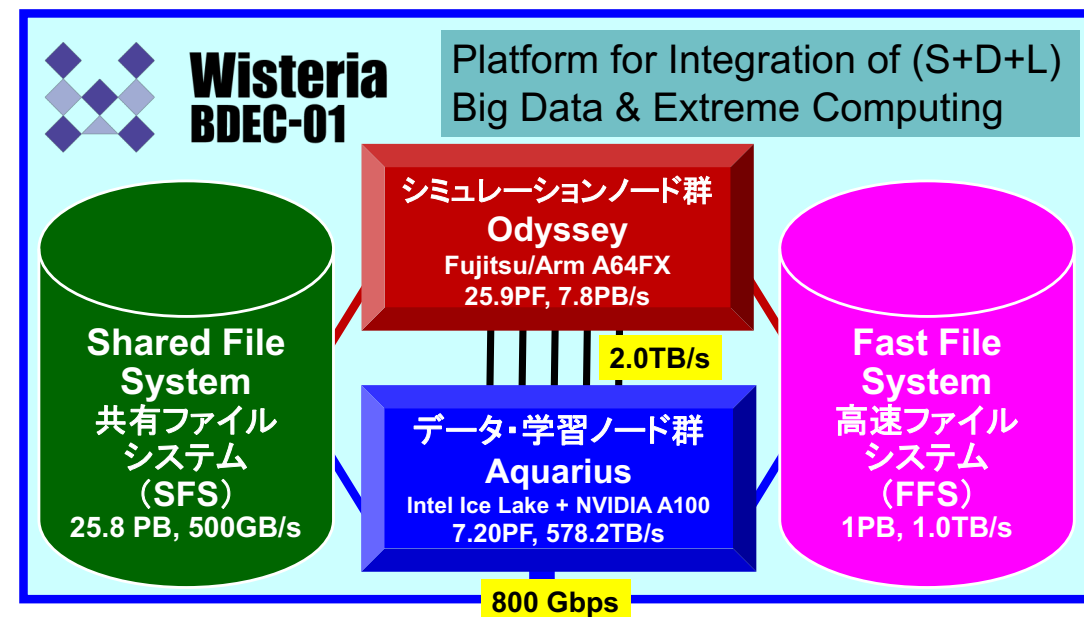
- 一般利用

- ノード固定

- Aquariusの1ノード(8GPU)を占有して利用
- 審査有り(ヒアリング)

- GPU占有

- Aquariusの1・2・4GPUを占有して利用
- 審査無し



External Network
外部ネットワーク



External Resources

外部リソース

ジョブクラス

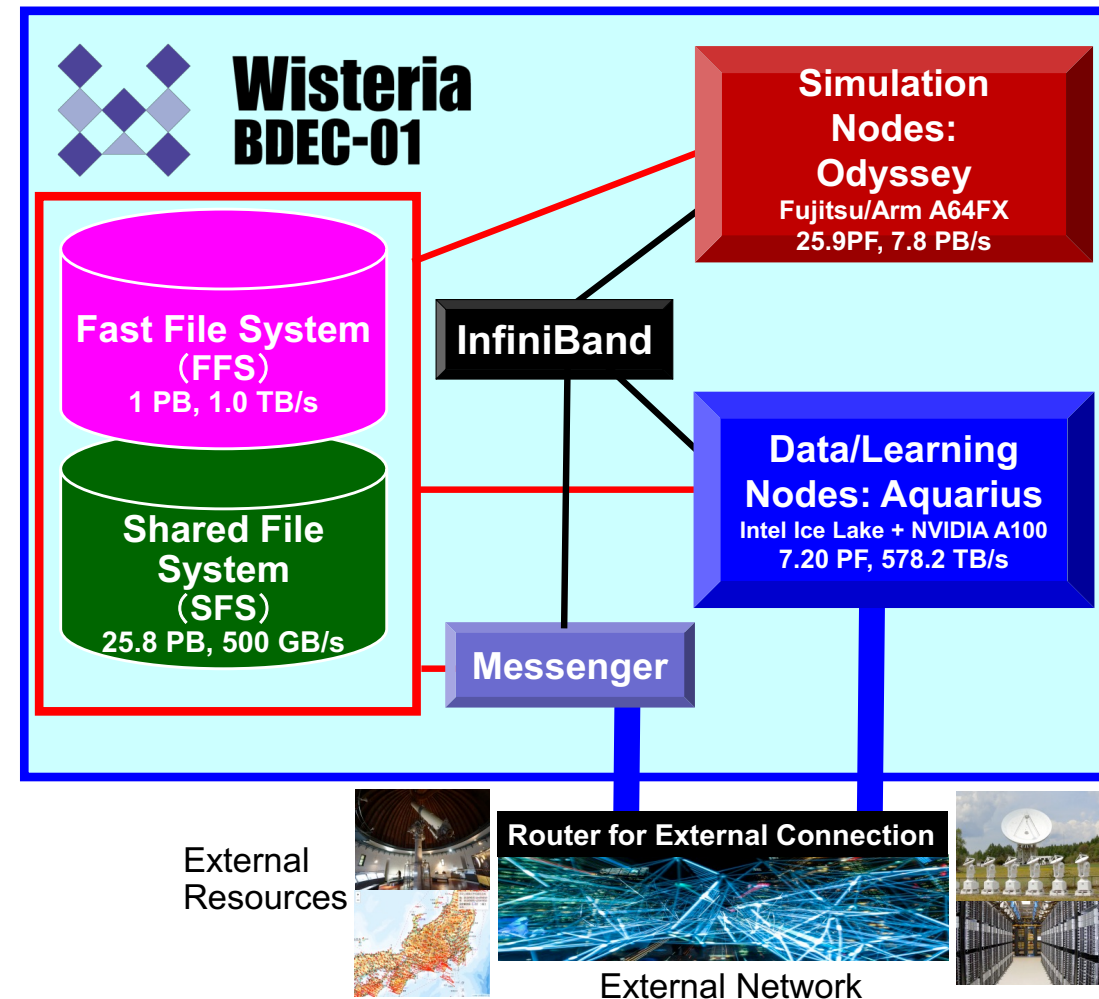
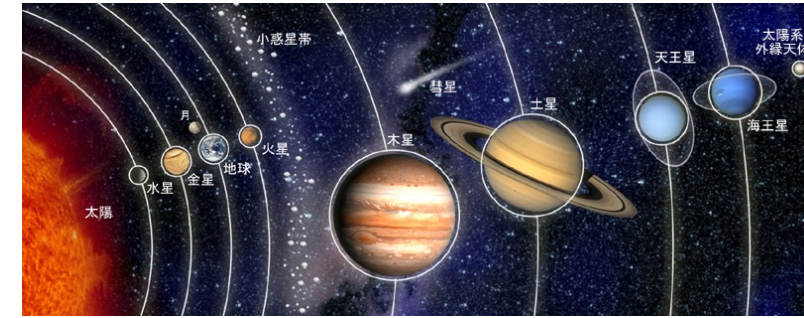
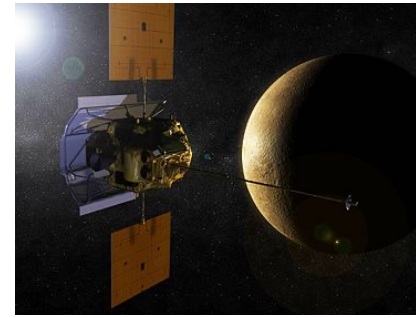
<https://www.cc.u-tokyo.ac.jp/supercomputer/wisteria/service/job.php>

- インタラクティブ
- バッチジョブ
- プリポスト
- 現状ではOdyssey, Aquariusは同時利用はできない

- Wisteria-O (Odyssey): シミュレーションノード群
 - XXX-o
 - priority-o: 優先キュー, トークン消費量1.5倍
- Wisteria-A (Aquarius): データ・学習ノード群
 - XXX-a ノード単位
 - share-XXX GPU単位
 - MIG (Multi-Instance GPU)により, GPU内を更に分割可能だが, 本システムでは採用せず

Wisteria-Messenger

- Intel Xeon Gold 6348 (IceLake)
 - 2.6 GHz, 28 cores/socket
 - 2 sockets × 6 nodes
 - IB-HDR
- Directly working with “Odyssey” for supporting “Aquarius”
 - can access external resources directly
 - e.g. “Filtering” in the Earthquake Simulation
 - h3-Open-SYS/WaitIO-Socket
- Operation has started
 - JHPCN users can access
 - Plan for public use is pending
- 本当は「Mercury」という計画があって、そのプロトタイプとしてMessengerがある。



- 東大情報基盤センターのスパコン概要
- Wisteria/BDEC-01
- **Ipomoea-01**
- h3-Open-BDEC

大規模共通ストレージシステム「Ipomoea」(1/2)

- スーパーコンピュータの処理能力の向上に伴い、扱うデータ量も増加の一途
 - 「計算・データ・学習」融合分野では、大量の観測データ、パラメータスタディの結果ファイルなどを処理する必要がある
- 当センターでは従来ストレージは各システムに附属して導入され、各システムのストレージは独立
 - 近年は当センターのシステム数も増加、利用者も目的や手法に応じて複数のシステムを同時に利用する事例が増加
 - システムがリプレースされる場合には大量のデータをバックアップする必要があった
 - 個別のワークロードのデータ量も増加
- このような状況(注:ストレージがシステム毎に独立)は利用者に多大な不便を強いることになり、当センターの全システムからアクセス可能な共通ストレージの導入が強く求められていた

大規模共通ストレージシステム「Ipomoea」(2/2)

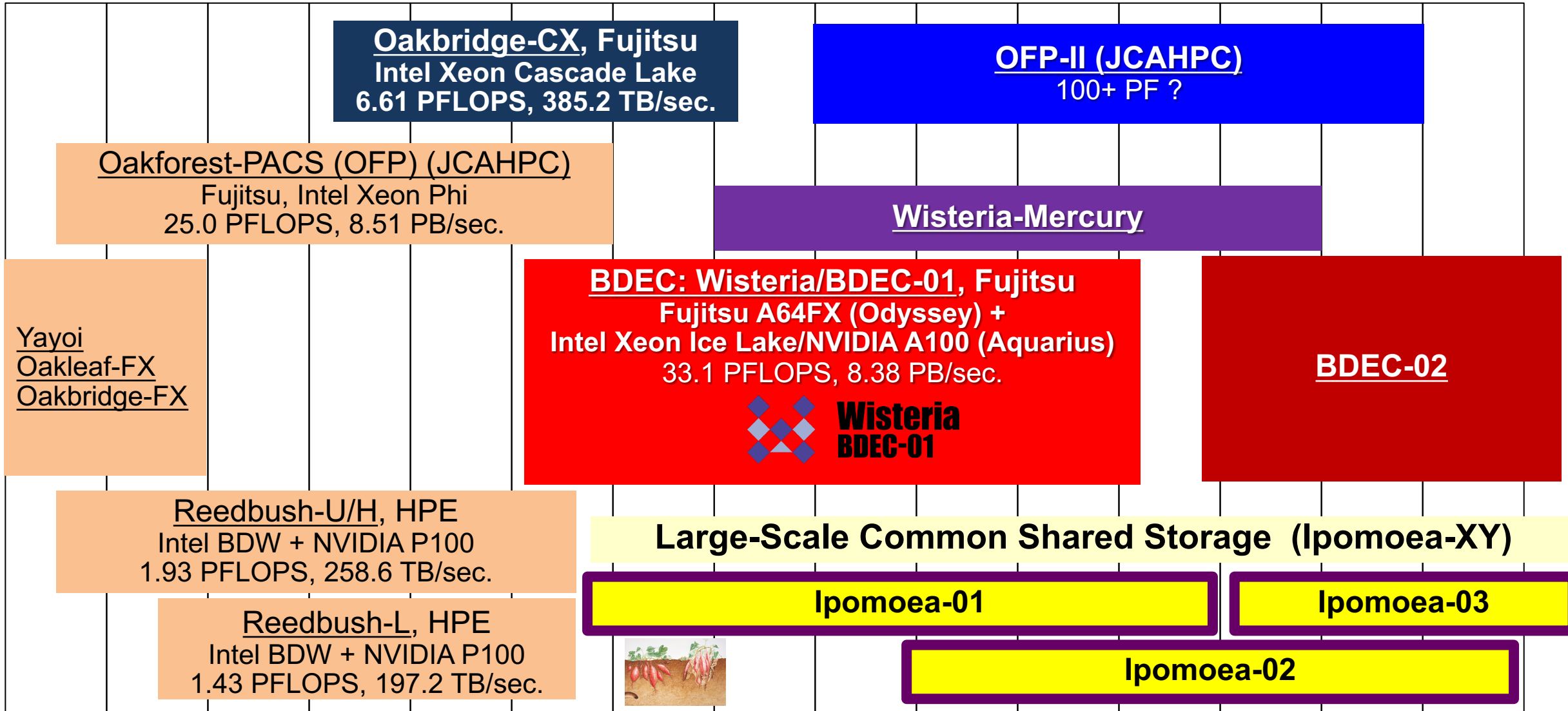
- 筑波大学と共同で最先端共同HPC 基盤施設(JCAHPC)によって運用しているOakforest-PACS(OFP)は, ピーク性能25PFLOPS を有するナショナルフラッグシップ級のシステム、2022年3月末に運用終了予定
- 2021年春, mdx , Wisteria/BDEC-01 が運用を開始
- 各システムからアクセスできる「大規模共通ストレージ(Ipomoea)」導入を決定
 - 「大規模共通ストレージシステム(Ipomoea)」は約5-6年使用し, 約3年ごとに新しいストレージシステム(25+PB)を導入し, 入れ替えることを想定している



Supercomputers in ITC/U.Tokyo

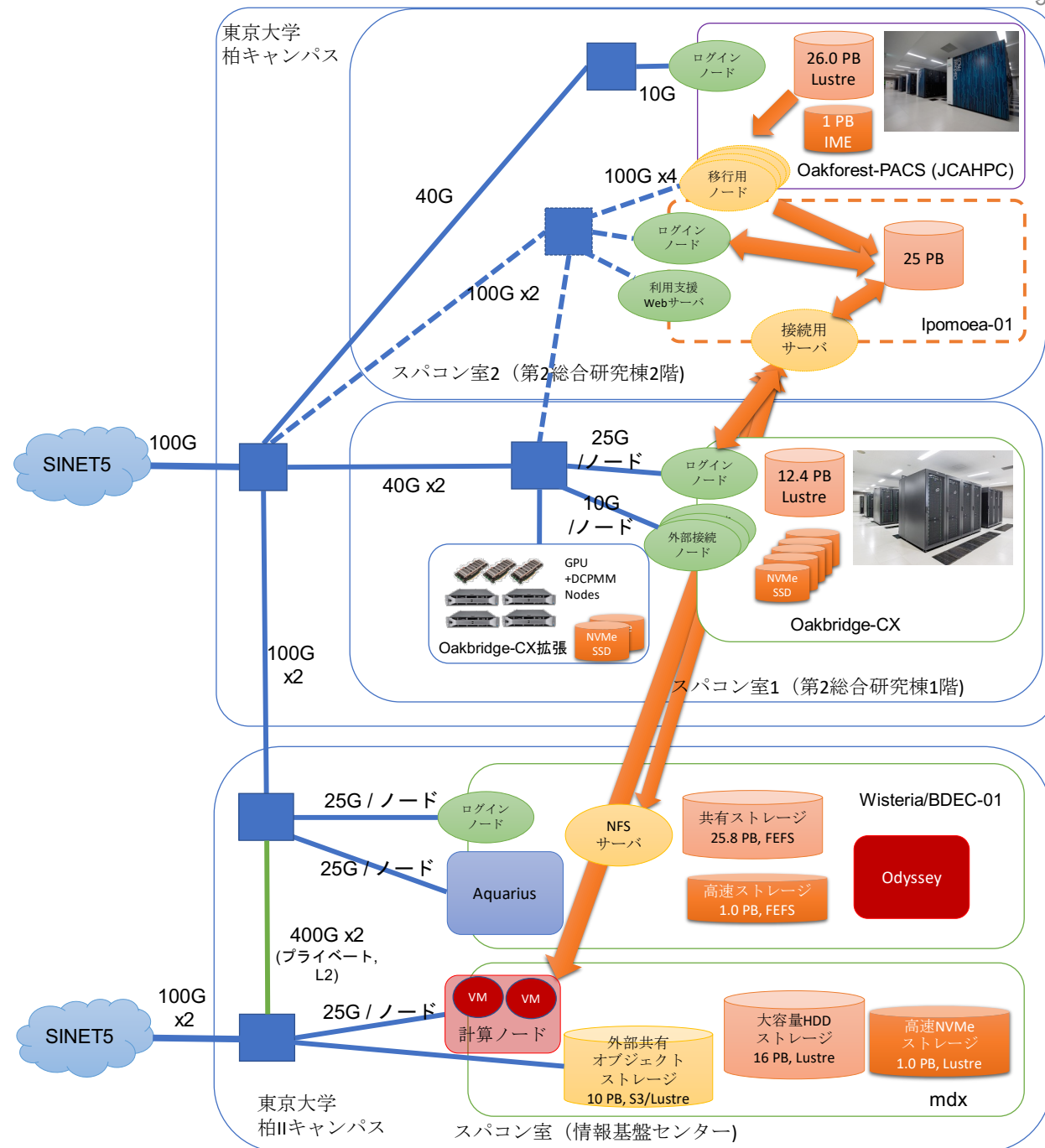
Information Technology Center, The University of Tokyo

FY16 17 18 19 20 21 22 23 24 25 26 27 28 29 30



Ipomoea-01

- 2022年1月運用開始予定, 25.7PF
- 富士通
- 2022年6月末までにOFPのLustre領域の必要ファイルについて移行完了
- 既存システム(2021年3月末)
 - OFP 50億ファイル 11 PB
 - OBCX 8.5億ファイル 2.6 PB
 - OPA



- 東大情報基盤センターのスパコン概要
- Wisteria/BDEC-01
- Ipomoea-01
- **h3-Open-BDEC**



(計算+データ+学習)融合によるエクサスケール時代の革新的シミュレーション手法

- エクサスケール(富岳+クラス)のスパコンによる科学的発見の持続的促進のため、計算科学にデータ科学、機械学習のアイデアを導入した(計算+データ+学習(S+D+L))融合による革新的シミュレーション手法を提案
 - (計算+データ+学習)融合によるエクサスケール時代の革新的シミュレーション手法(科研費基盤S, 代表: 中島研吾(東大情基セ), 2019年度~2023年度)
- 革新的ソフトウェア基盤「h3-Open-BDEC」の開発: 東大BDECシステム(Wisteria/BDEC-01), 「富岳」等を「S+D+L」融合プラットフォームと位置づけ、スパコンの能力を最大限引き出し、最小の計算量・消費電力での計算実行を実現するために、下記2項目を中心に研究
 - 変動精度演算・精度保証・自動チューニングによる新計算原理に基づく革新的数値解法
 - 階層型データ駆動アプローチ(hDDA: Hierarchical Data Driven Approach)等に基づく革新的機械学習手法
 - Hierarchical, Hybrid, Heterogeneous \Rightarrow h3



**Wisteria
BDEC-01**



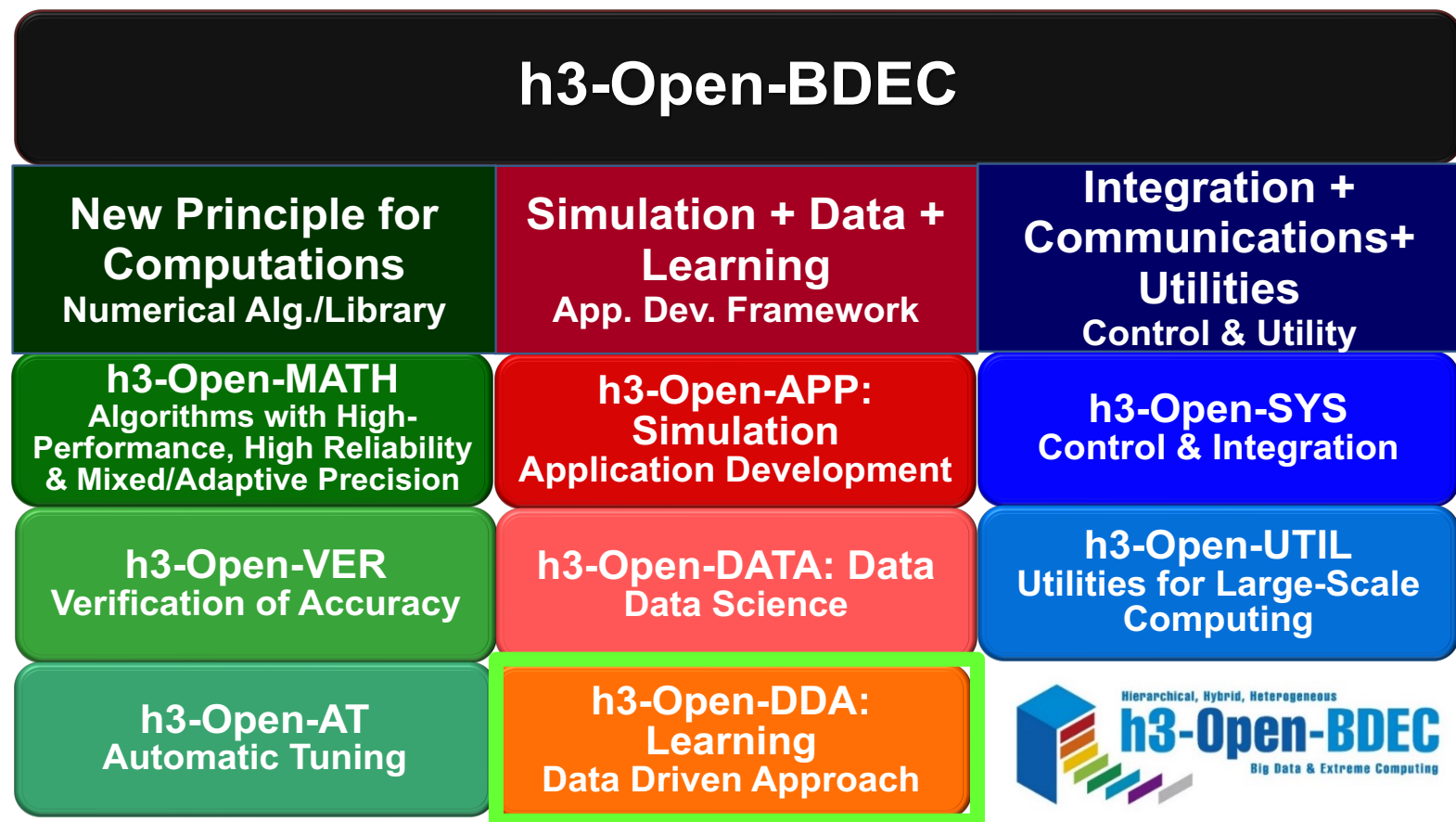
h3-Open-BDEC

「計算＋データ＋学習」融合を実現する革新的ソフトウェア基盤

<https://h3-open-bdec.cc.u-tokyo.ac.jp/>

- ① 変動精度演算・精度保証・自動チューニングによる新計算原理に基づく革新的数値解法
- ② 階層型データ駆動アプローチ (hDDA: Hierarchical Data Driven Approach) 等に基づく革新的機械学習手法

✓ Hierarchical, Hybrid, Heterogeneous ⇒ h3



期待される成果と意義

<http://nkl.cc.u-tokyo.ac.jp/h3-Open-BDEC/>

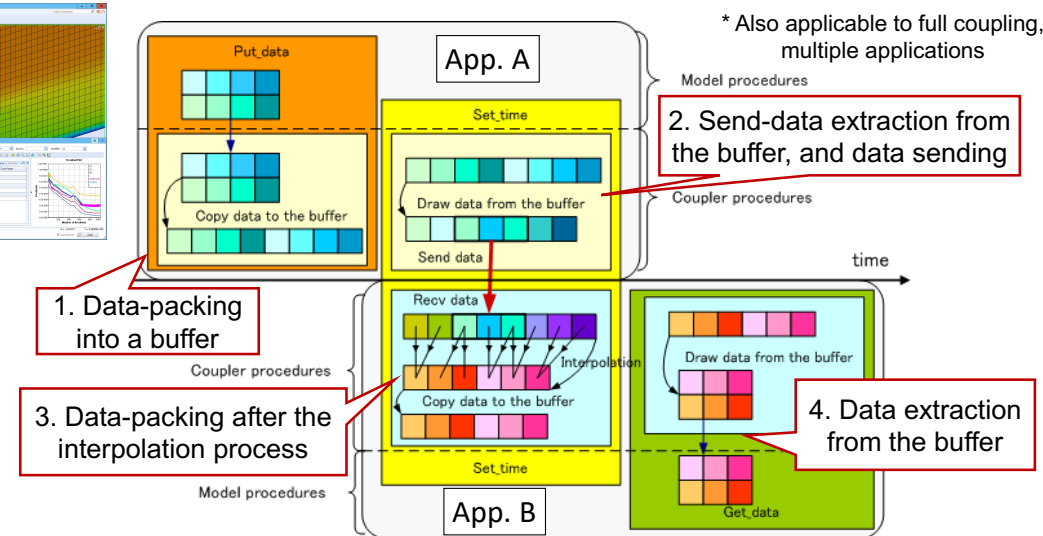
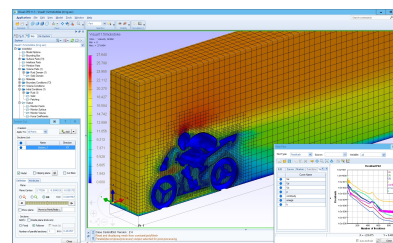
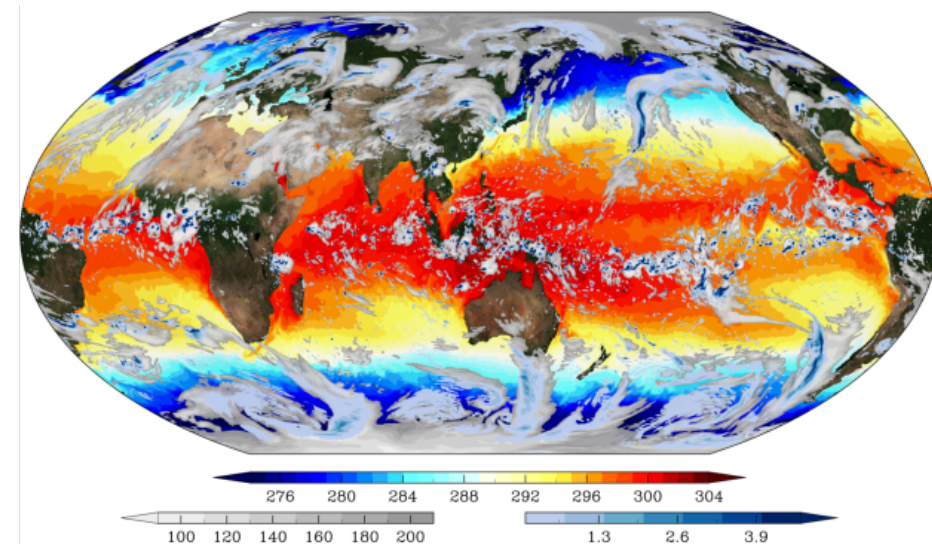


- 計算科学の専門家のみで(S+D+L)融合を容易に実現
 - 機械学習の専門家のサポートを必要としない
- ソースコード, マニュアル類も含めて一般に公開, 様々なエクサスケールシステムでの普及を目指す
 - ポスト富岳も含めたポストムーア時代への展開
- h3-Open-BDEC利用による(S+D+L)融合シミュレーションにより従来手法と同等の正確さを保ちつつ, 大幅な計算量・消費電力削減を目指す(10分の1が目標)。
- シミュレーション高度化: パラメータスタディのケース数を削減できる
- リアルタイム災害シミュレーション等への適用

Wisteria/BDEC-01上における

h3-Open-BDECを使用した(S+D+L)融合

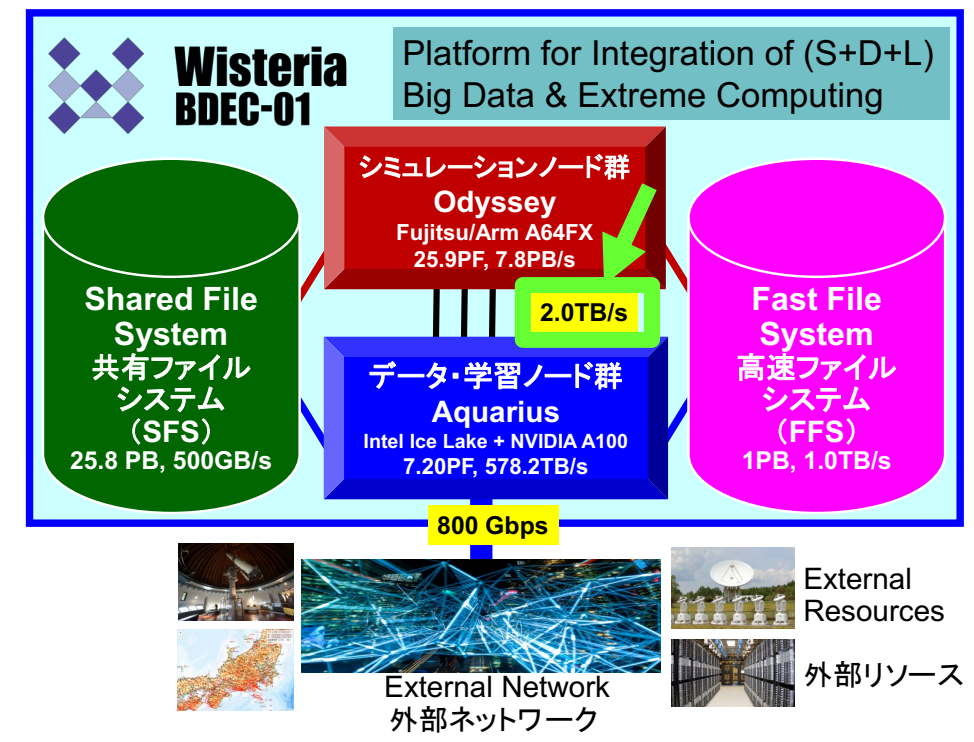
- シミュレーションとデータ同化の融合
 - 典型的・伝統的な(S+D+L)融合
- 気候・気象のための大気海洋連成シミュレーション
 - 東大大気海洋研, 理研, 国立環境研他
- **リアルタイム同化+三次元強震動シミュレーション**
 - **東大地震研**
- リアルタイム災害シミュレーション
 - 洪水, 津波
- 既存シミュレーションコードの(S+D+L)融合による高度化
 - OpenFOAM



AI for HPC の実現



- Odyssey-Aquarius連携
 - MPIによる通信は不可
 - O-Aを跨いでMPIプログラムは動かない
 - Odyssey-Aquarius間はInfiniband-EDR (2TB/sec)で結合されている
- ソフトウェア開発
 - O-A間通信: h3-Open-SYS/WaitIO
 - IB-EDR経由 (WaitIO-Socket)
 - 高速ファイルシステム (FFS) 経由連携 (WaitIO-File)
 - 高機能カプラー: h3-Open-UTIL/MP

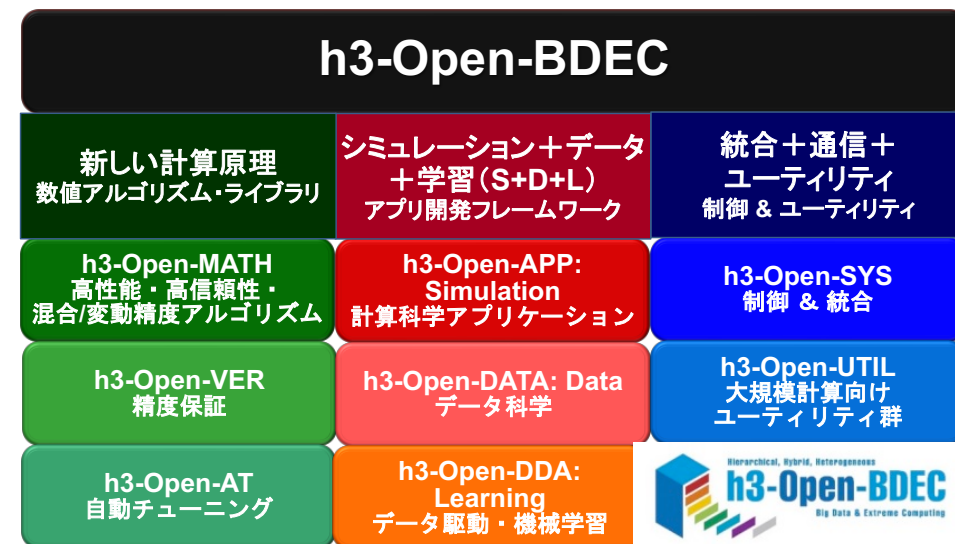
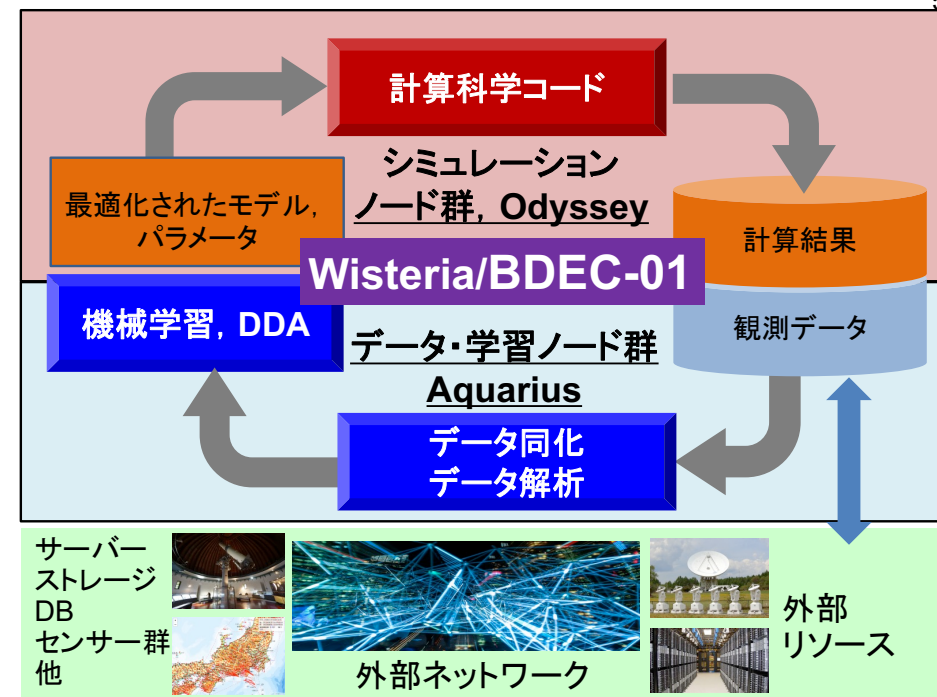


h3-Open-BDEC		
新しい計算原理 数値アルゴリズム・ライブラリ	シミュレーション+データ +学習 (S+D+L) アプリ開発フレームワーク	統合+通信+ ユーティリティ 制御 & ユーティリティ
h3-Open-MATH 高性能・高信頼性・ 混合/変動精度アルゴリズム	h3-Open-APP: Simulation 計算科学アプリケーション	h3-Open-SYS 制御 & 統合
h3-Open-VER 精度保証	h3-Open-DATA: Data データ科学	h3-Open-UTIL 大規模計算向け ユーティリティ群
h3-Open-AT 自動チューニング	h3-Open-DDA: Learning データ駆動・機械学習	

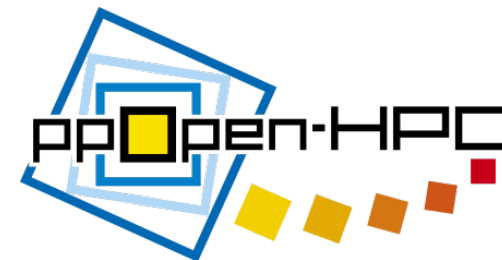
h3-Open-SYS/WaitIO

データ受け渡しライブラリ[松葉, 2020]

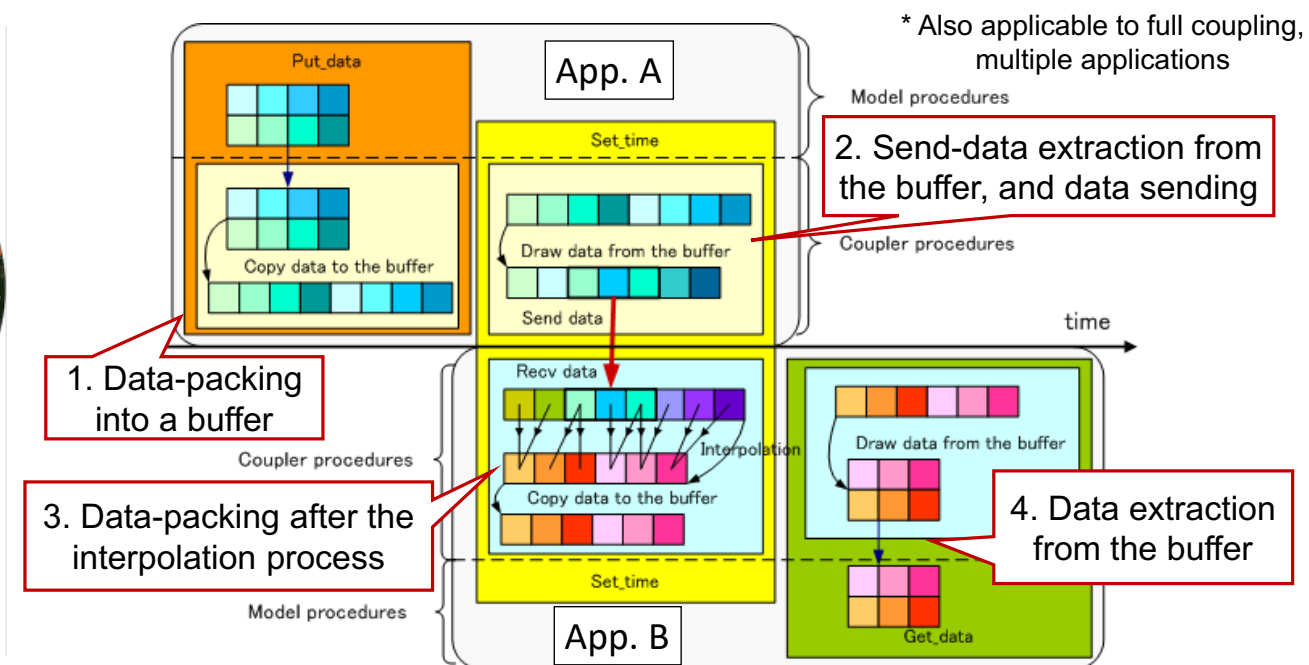
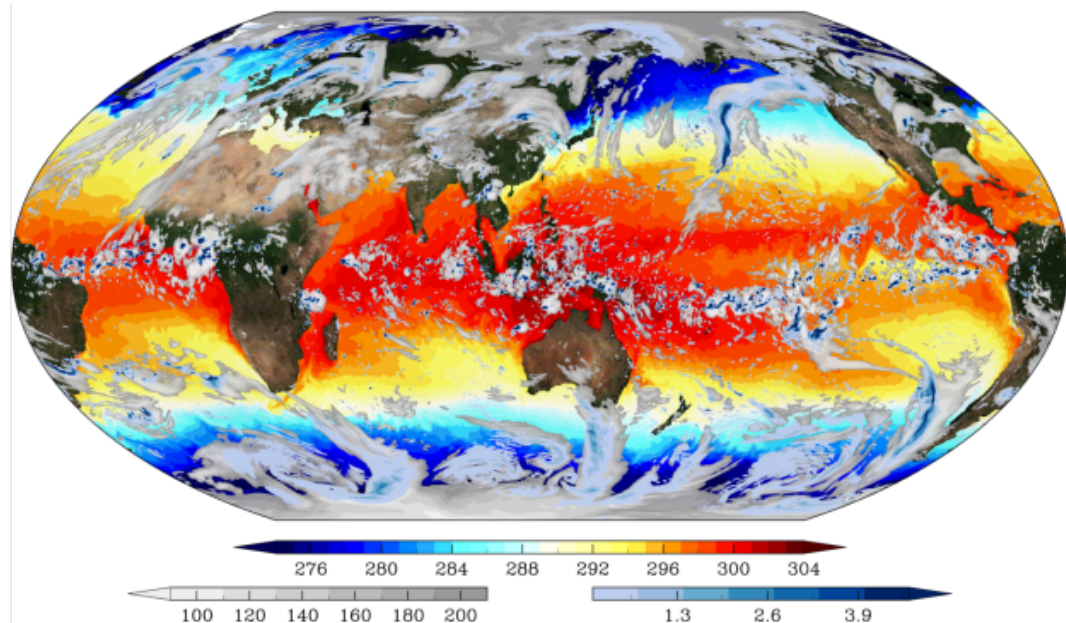
- ヘテロジニアス環境下での異なるコンポーネント間ファイル経由連携ライブラリとして考案
- 機能
 - ✓ Odyssey～Aquarius間連携
 - IB-EDR経由通信 (WaitIO-Socket)
 - ファイル経由 (WaitIO-File)
 - ✓ 外部からのデータ取得 (観測データ等)
 - ✓ 読み込み・書き出しの同期
- API: C/C++, Fortranから呼び出し可能
 - ✓ MPIライクなインタフェースを提供
- 多機能カプラー(h3-Open-UTIL/MP)との連携



連成シミュレーションのためのカプラー 〔荒川, 八代〕



- 従来のカプラー (Coupler) : ppOpen-MATH/MP
 - 複数 (通常2つ: 大気 (NICAM) + 海洋 (COCO)) のアプリケーションの弱連成 (Weak Coupling) をサポート
 - 各アプリケーションは1種類の計算をやる



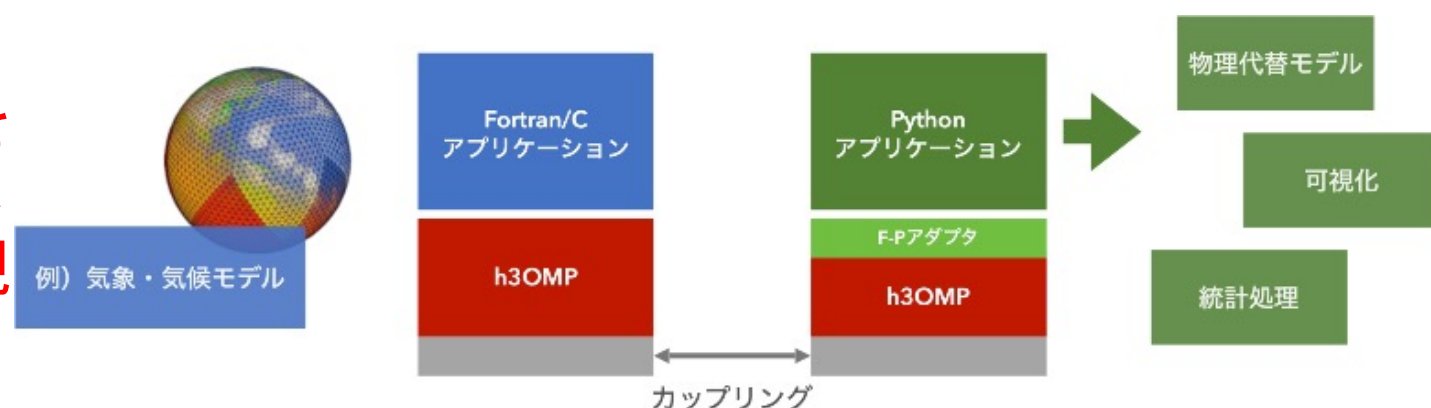
「計算＋データ＋学習」融合を支援する 多機能カプラーh3-Open-UTIL/MP



- 異なる物理モデル連成のアンサンブル実行を支援・統合するための機能
 - MPI通信、時刻同期、格子系間マッピング等の管理機能の他、従来のカプラーには無い、複数の弱連成結合シミュレーションのアンサンブル実行、片側のモデルのみをアンサンブル実行する多対1の弱連成結合が可能
 - スパコン上で、全地球大気海洋連成シミュレーションによって動作検証済み

Fortran/Cコード(物理モデル)とPythonコードの弱連成を実現する機能

- FortranやCで記述されたプログラム同士の間での連成計算に限って開発を行ってきたカプラーを、Pythonによって記述されたAI・機械学習、可視化処理系のワークロードから活用できるように機能拡充。



Fortran/CアプリとPythonアプリの連成計算の模式図
〔八代・荒川 2020〕

- O-A利用 : WaitIOとの連携