Introduction of Miyabi: AI for Science Based on Integration of “Simulation, Data, and Learning"

Toshihiro Hanawa
Lead, Operation Support Division, JCAHPC
Information Technology Center
The University of Tokyo

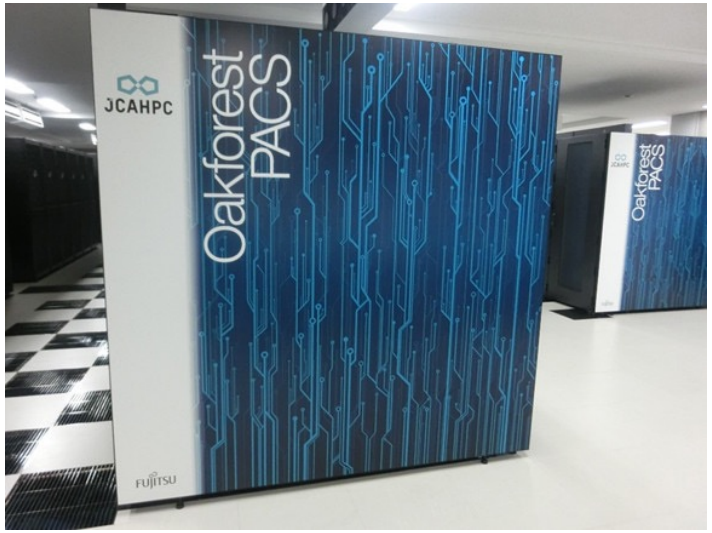
JCAHPC

http://jcahpc.jp/eng/index.html

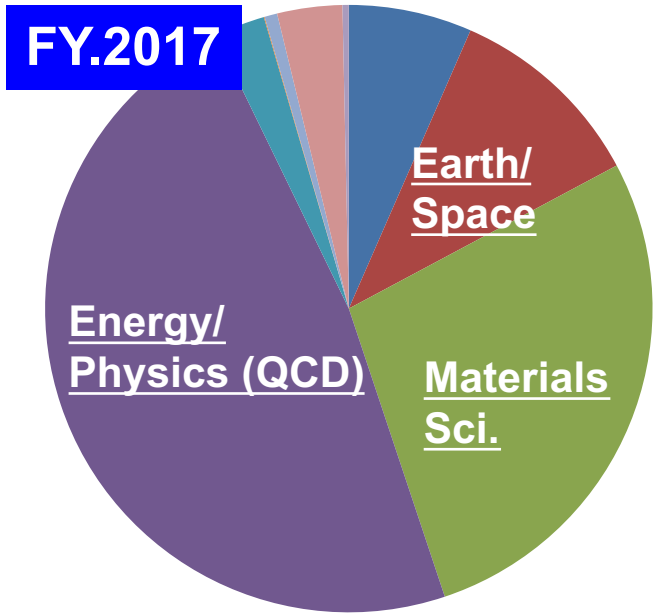
筑波大学 University of Tsukuba　東京大学 THE UNIVERSITY OF TOKYO

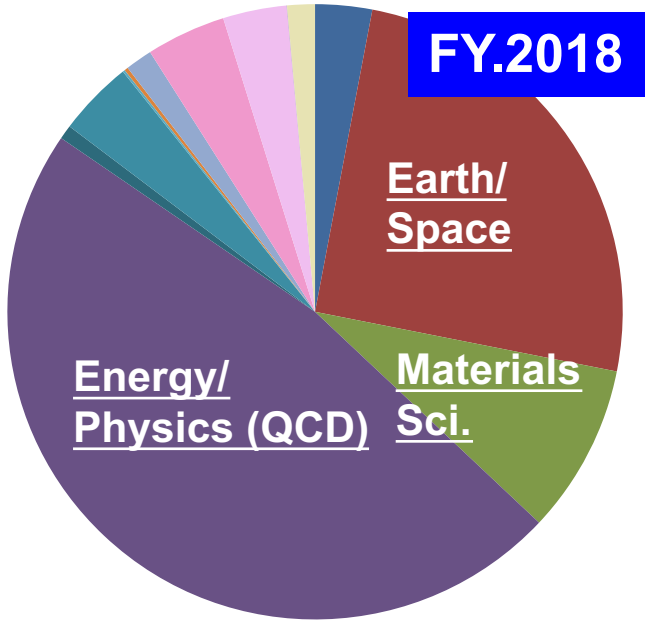
JCAHPC

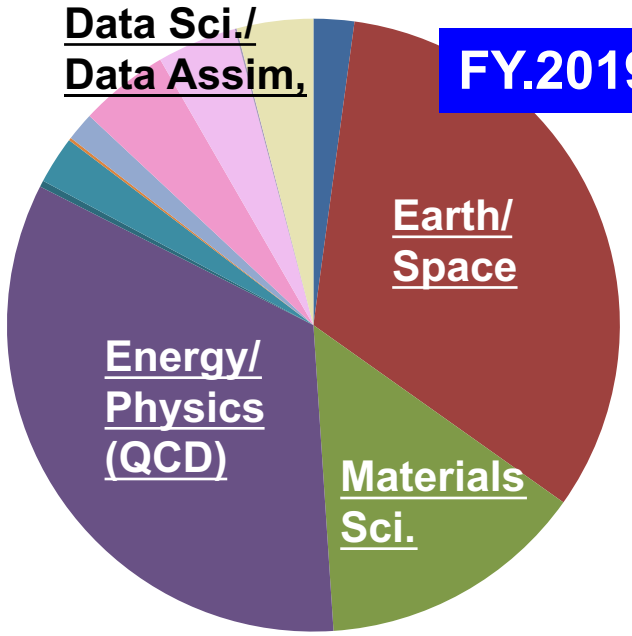
- Joint Center for Advanced High Performance Computing, since 2013
  - University of Tsukuba & University of Tokyo
    - Budgets of 2 Centers are combined
  - Promotion on Computational Science, Design/Procurement/Operation of Large-scale Systems
- Oakforest-PACS (OFP), 1st System of JCAHPC
  - 8,208 Intel Xeon Phi, 25 PF, Fujitsu
  - Top 500 (#6 (Nov.2016), #1 in Japan)
    - Retired in the end of March 2022 (#39 (Nov.2021))
- National Flagship System (in fact) in FY.2019/2020
  - Between K and Fugaku

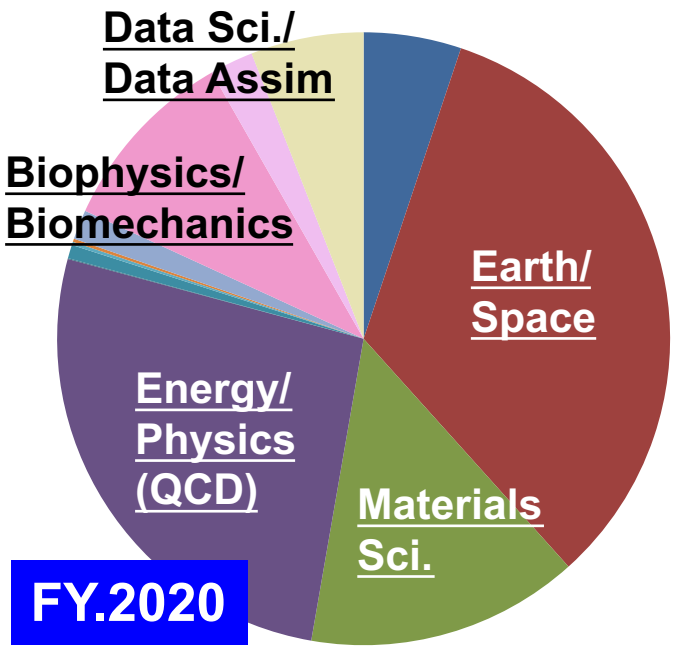


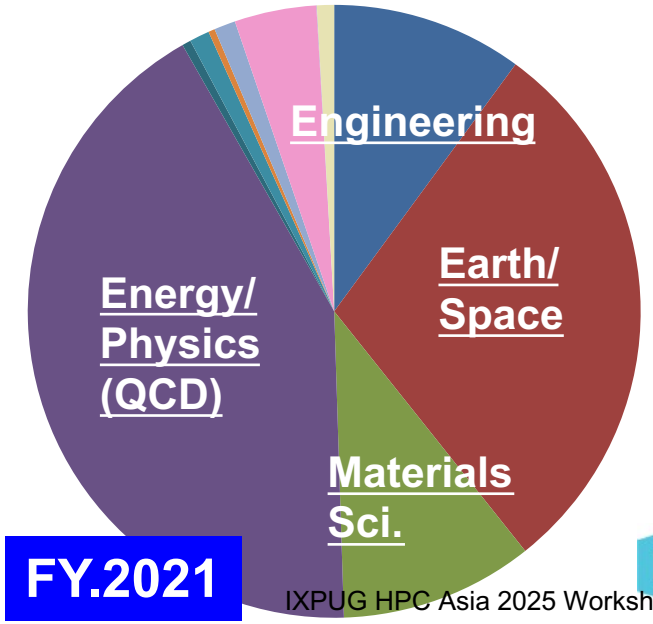
Research Area based on Machine Hours : OFP





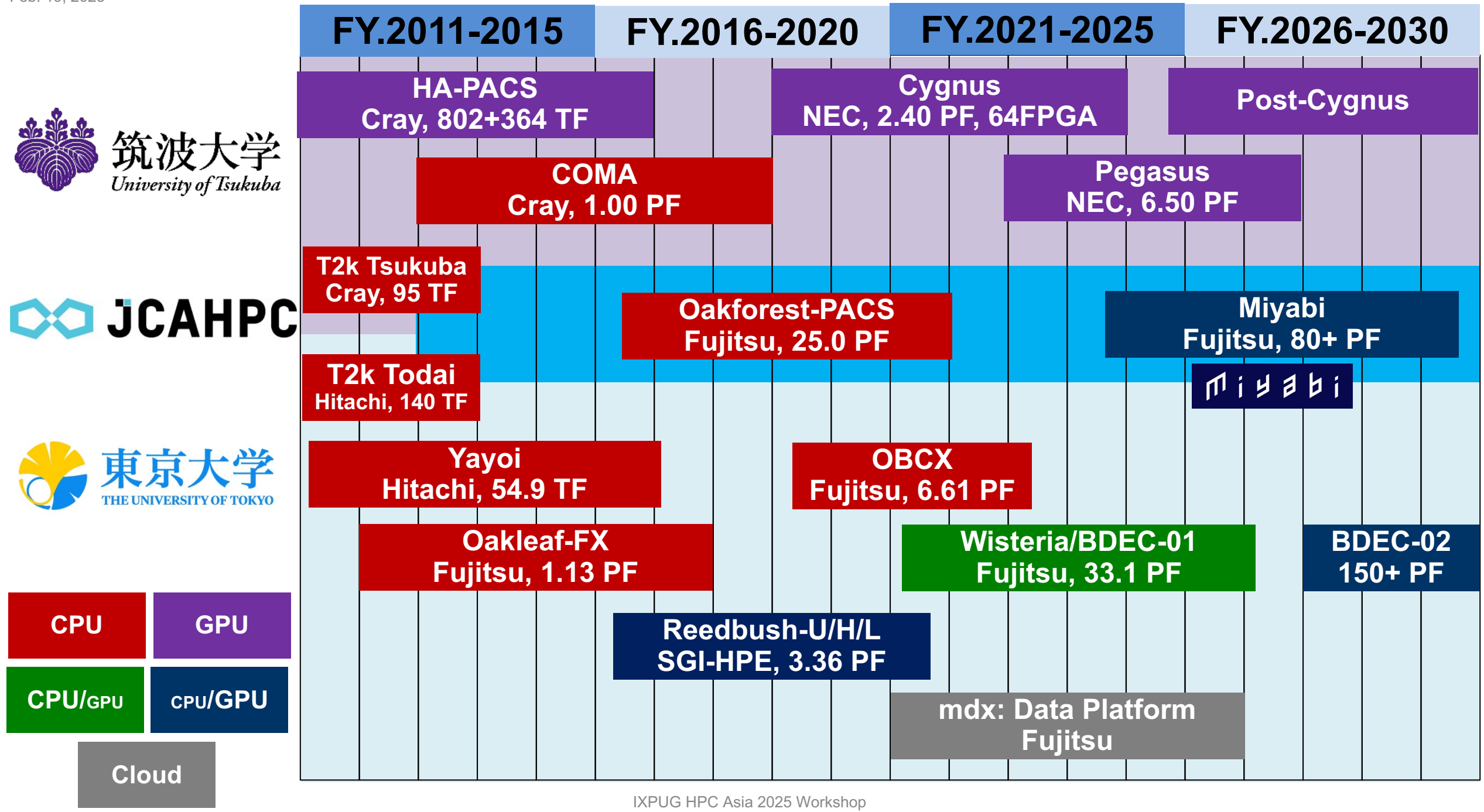






- Engineering
- Earth & Space
- Materials Science
- Energy & Physics (QCD)
- Info. Sci.: System
- Info. Sci. : Algo.
- Info. Sci. : AI
- Education
- Industry
- Biophysics & Biomechanics
- Bioinformatics
- Social Science & Economics
- Data Science & Data Assimilation

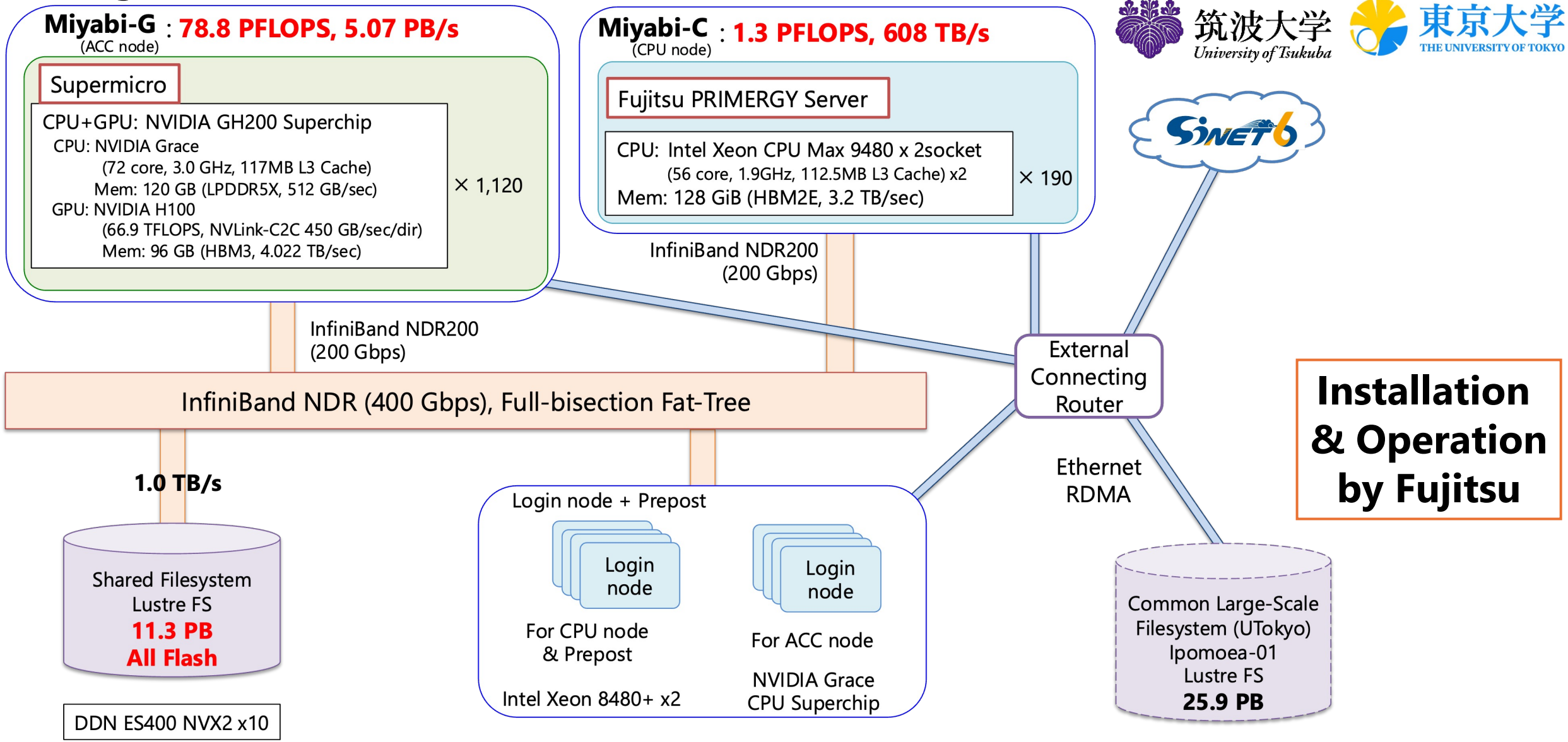
JCAHPC


FY.2011-2015
FY.2016-2020
FY.2021-2025
FY.2026-2030
筑波大学
University of Tsukuba
JCAHPC
東京大学
THE UNIVERSITY OF TOKYO
HA-PACS
Cray, 802+364 TF
Cygnus
NEC, 2.40 PF, 64FPGA
Post-Cygnus
COMA
Cray, 1.00 PF
Pegasus
NEC, 6.50 PF
T2k Tsukuba
Cray, 95 TF
Oakforest-PACS
Fujitsu, 25.0 PF
Miyabi
Fujitsu, 80+ PF
T2k Todai
Hitachi, 140 TF
Yayoi
Hitachi, 54.9 TF
OBCX
Fujitsu, 6.61 PF
Oakleaf-FX
Fujitsu, 1.13 PF
Wisteria/BDEC-01
Fujitsu, 33.1 PF
BDEC-02
150+ PF
Reedbush-U/H/L
SGI-HPE, 3.36 PF
mdx: Data Platform
Fujitsu
CPU
GPU
CPU/GPU
CPU/GPU
Cloud

About Miyabi

- System name was released in Apr. 2024.
- The name "**Miyabi**," which means "**elegance and refinement**" in Japanese, carries more than just excellent theoretical performance; it encapsulates the intention to manifest its inherent abilities with less effort.
- Both centers will undertake the introduction and operation of the new "**Miyabi**" system in tight collaboration. By advancing highly sophisticated computational science through integrating "simulation, data, and learning," we aim to contribute to the realization of a secure, safe, and reliable society.

Miyabi Overview



Installation & Operation by Fujitsu

Miyabi Overview

Miyabi

JCAHPC

筑波大学 University of Tsukuba

東京大学 THE UNIVERSITY OF TOKYO

Started from Jan. 14, 2025

Front racks are for management use


m
i
y
a
b
i
JCAHPC
FUJITSU

Miyabi-G Overview

Miyabi

JCAHPC

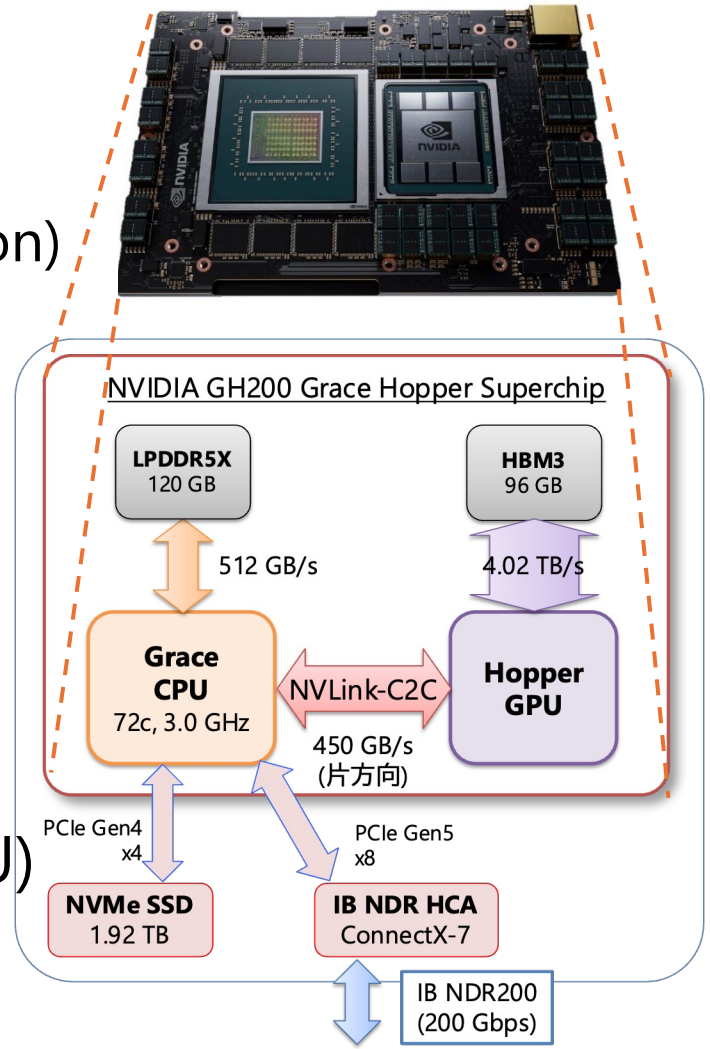
筑波大学 University of Tsukuba

東京大学 THE UNIVERSITY OF TOKYO

There are also same-sized racks installed on the back

Miyabi-G: Accelerator node

- **Compute node：CPU+GPU: NVIDIA GH200 Grace-Hopper Superchip**
  - CPU: **Grace** 72 core, 3.0 GHz, 3.45 TFLOPS
    - Memory：120 GByte, 512 GByte/sec (LPDDR5X)
  - GPU: **Hopper** 66.9 TFLOPS (by Tensor core with double precision)
    - Memory：96 GByte, 4.02 TByte/sec (HBM3)
  - CPU-GPU：NVLink-C2C, 450 GB/sec(single dir.), Cache-coherent, porting of existing application is easier
  - NVMe SSD: 1.9 TB, 8.0GB/sec, GPUDirect Storage
  - InfiniBand NDR200
- **Miyabi-G Overall**
  - **1,120 node** (18 rack、64 node/rack)
  - Theoretical peak performance: **78.8 PFLOPS** (CPU+GPU)
  - Aggregate memory BW: 5.07 PByte/sec



JCAHPC

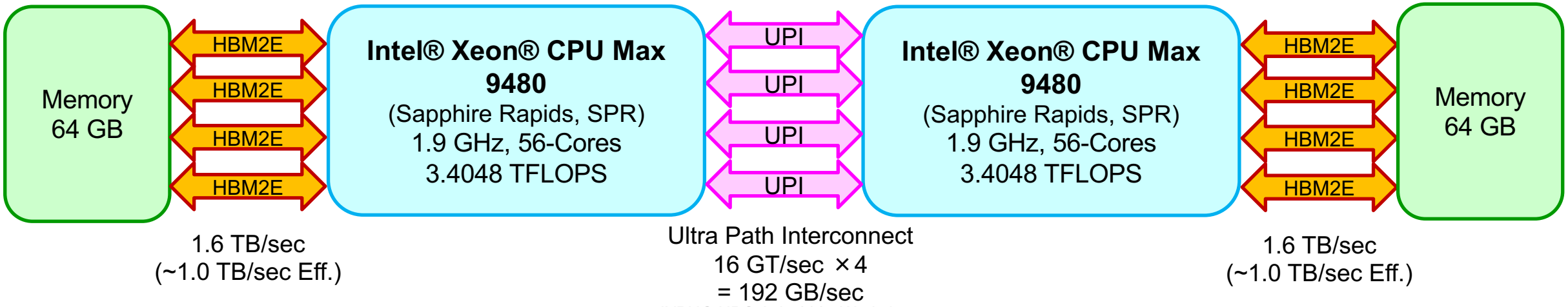
筑波大学 University of Tsukuba

東京大学 THE UNIVERSITY OF TOKYO

Miyabi-C overview

Miyabi-C: General Purpose CPU node

- **Compute node: CPU: Intel Xeon CPU Max 9480**
  - CPU: (Intel Xeon 9480: 56 core, 1.9 GHz) x 2 socket, 6.8 TFLOPS
  - Memory：128 GB, 3.2 TB/sec (**HBM2E only**)
  - InfiniBand NDR200
- **Miyabi-C**
  - 190 node (3 rack, 64 node/rack)
  - Theoretical peak performance: 1.3 PFLOPS
  - Aggregate memory BW: 372 TByte/sec. (Stream Triad, Theoretical peak: 608 TByte/sec)



Miyabi-C Compute Node

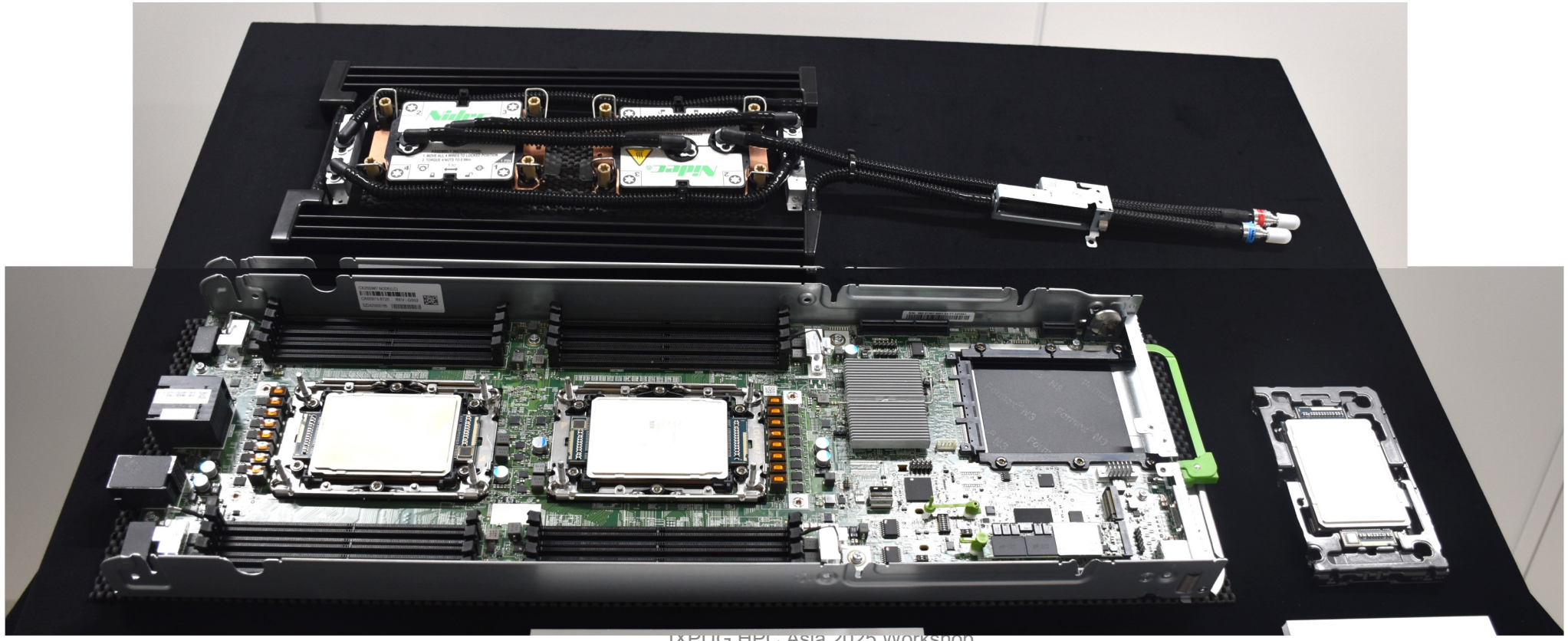
Miyabi

JCAHPC

筑波大学 University of Tsukuba

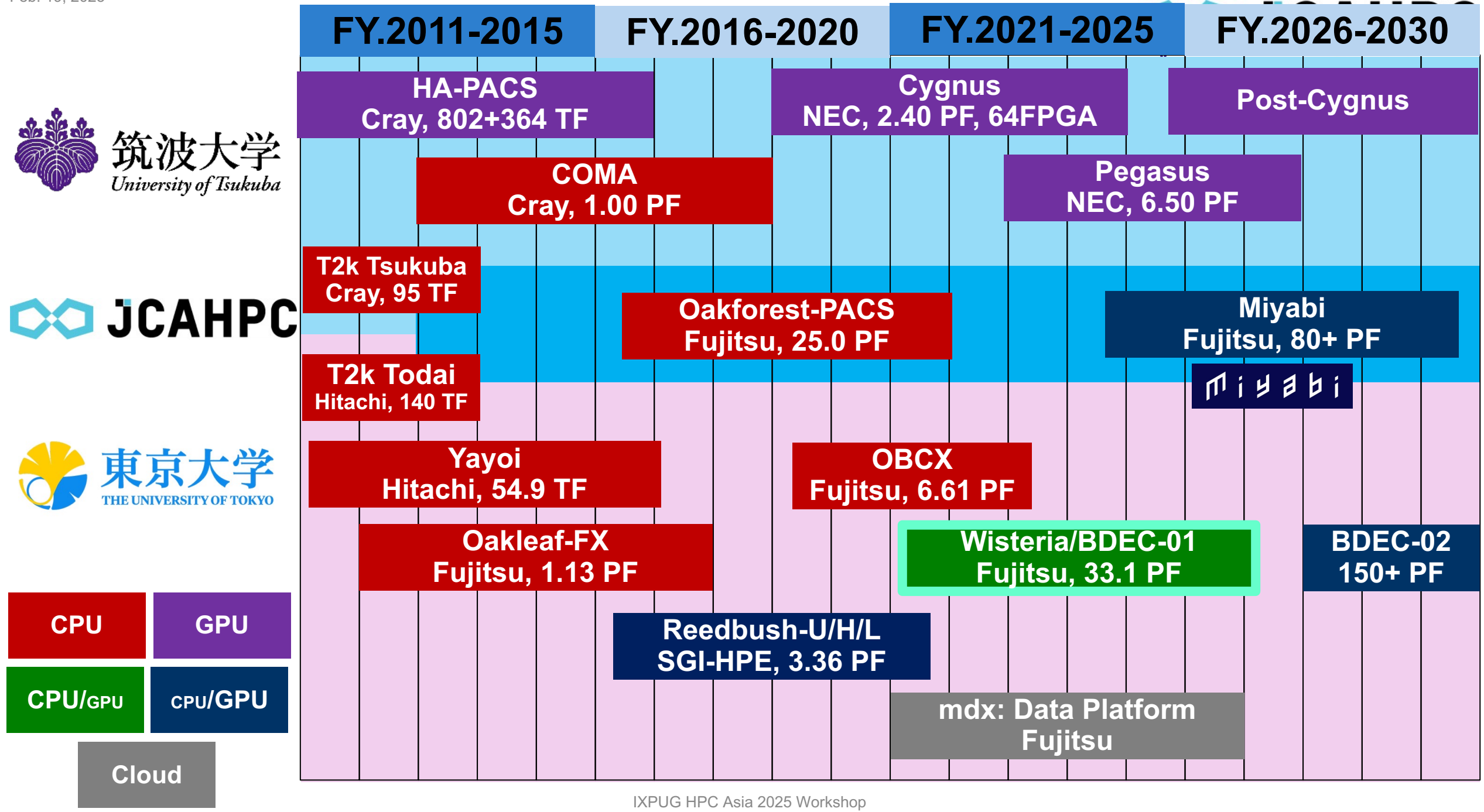
東京大学 THE UNIVERSITY OF TOKYO

- **Fujitsu PRIMERGY CX2550 M7**
  - The memory is only HBM2E in the package, DIMMs are not implemented
  - 4 nodes in 2U Chassis



What we will do with Miyabi

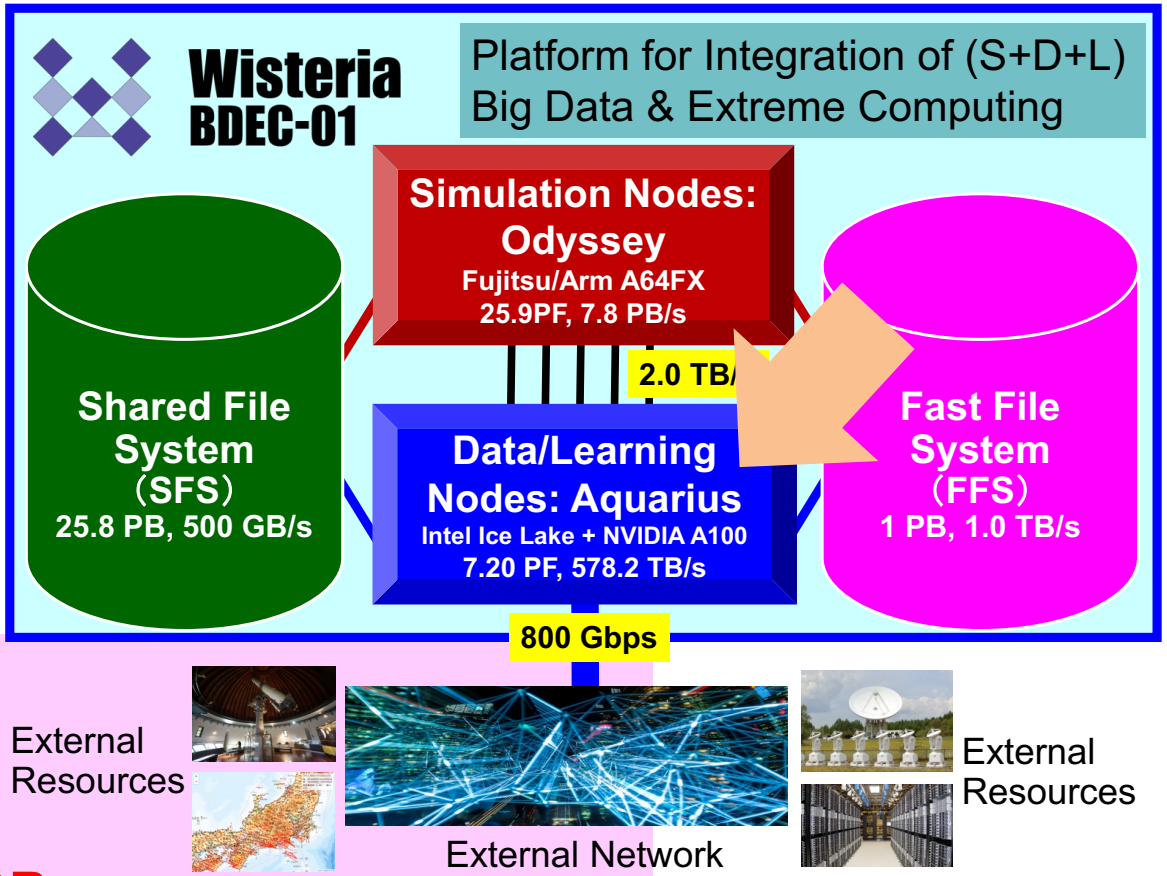
- 1. Computational Science !!!
- 2. AI for Science based on Integration of (Simulation/Data/Learning)
- 3. QC-HPC Hybrid Computing


FY.2011-2015
FY.2016-2020
FY.2021-2025
FY.2026-2030
筑波大学
University of Tsukuba
JCAHPC
東京大学
THE UNIVERSITY OF TOKYO
HA-PACS
Cray, 802+364 TF
Cygnus
NEC, 2.40 PF, 64FPGA
Post-Cygnus
COMA
Cray, 1.00 PF
Pegasus
NEC, 6.50 PF
T2k Tsukuba
Cray, 95 TF
Oakforest-PACS
Fujitsu, 25.0 PF
Miyabi
Fujitsu, 80+ PF
Miyabi
T2k Todai
Hitachi, 140 TF
Yayoi
Hitachi, 54.9 TF
OBCX
Fujitsu, 6.61 PF
Oakleaf-FX
Fujitsu, 1.13 PF
Wisteria/BDEC-01
Fujitsu, 33.1 PF
BDEC-02
150+ PF
Reedbush-U/H/L
SGI-HPE, 3.36 PF
mdx: Data Platform
Fujitsu
CPU
GPU
CPU/GPU
CPU/GPU
Cloud

Wisteria/BDEC-01

The 1st BDEC System
(Big Data & Extreme Computing)
HW Platform for Integration of (S+D+L)

- Operation started on May 14, 2021
- 33.1 PF, 8.38 PB/sec by **Fujitsu**
  - ~4.5 MVA with Cooling, ~360m²
- **2 Types of Node Groups**
  - Hierarchical, Hybrid, Heterogeneous (h3)
  - **Simulation Node Group: Odyssey**
    - **Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF**
      - **7,680 nodes (368,640 cores), Tofu-D**
      - **General Purpose CPU + HBM**
      - **Commercial Version of "Fugaku"**
  - **Data/Learning Node Group: Aquarius**
    - **Data Analytics & AI/Machine Learning**
    - **Intel Xeon Ice Lake + NVIDIA A100, 7.2PF**
      - **45 nodes (90x Ice Lake, 360x A100), IB-HDR**
    - **DL nodes are connected to external resources directly**
- File Systems: SFS (Shared/Large) + FFS (Fast/Small)



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

Simulation Codes

Optimized Models & Parameters

Simulation Nodes Odyssey

Wisteria/BDEC-01

Results

Observation Data

Machine Learning, DDA

Data/Learning Nodes, Aquarius

Data Assimilation Data Analysis

Server, Storage, DB, Sensors, etc.

External Network

External Resources

Wisteria BDEC-01

h3-Open-BDEC: Innovative Software Platform for Integration of (S+D+L)

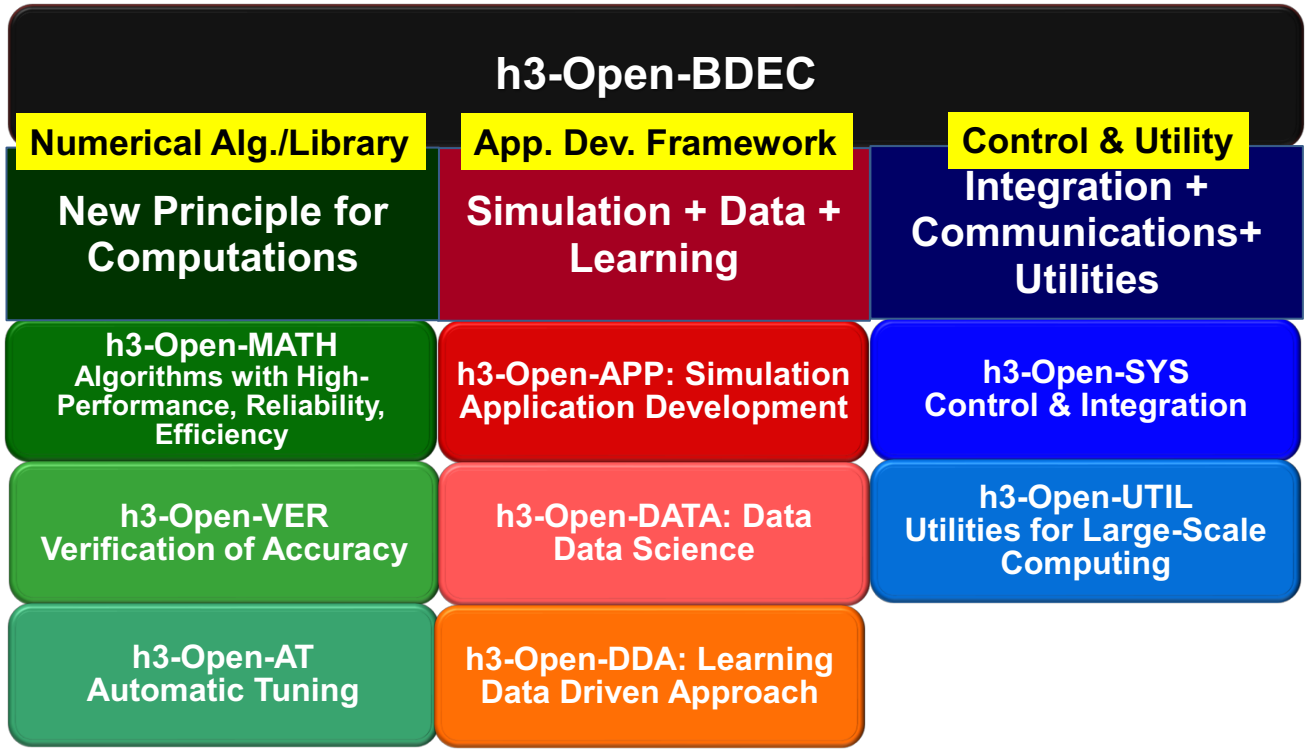
FY.2019-2023 supported by JSPS

Leading PI: Kengo Nakajima

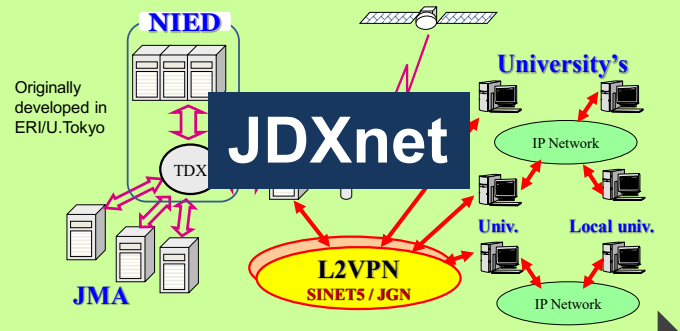
- "Three" Innovations
  - New Principles for Numerical Analysis by Adaptive Precision, Automatic Tuning & Accuracy Verification
  - Integration of (S+D+L) by Hierarchical Data Driven Approach (*h*DDA)
  - Software & Utilities for Heterogenous Environment, such as Wisteria/BDEC-01

Hierarchical, Hybrid, Heterogeneous
h3-Open-BDEC
Big Data & Extreme Computing

JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE · FOUNDED 1932 ·



3D Earthquake Simulation with Real-Time Data Observation/Assimilation on Wisteria/BDEC-01



Observed Data

Obs. Data

SINET6

External Server for Observed Data

Wisteria BDEC-01

Data/Learning Nodes: Aquarius
Intel Ice Lake + NVIDIA A100
7.20 PF, 578.2 TB/s

2.0 TB/s

Simulation Nodes: Odyssey
Fujitsu/Arm A64FX
25.9PF, 7.8 PB/s

Output & Movie

Filter

Visualizer

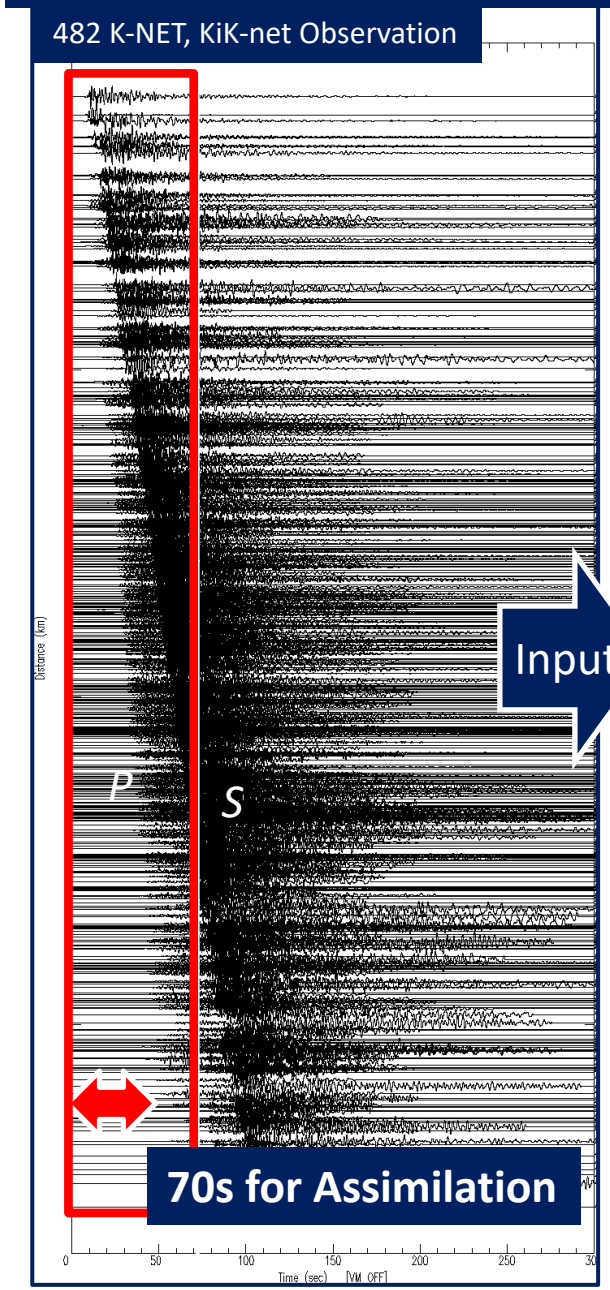
h3-Open-SYS/WaitIO-Socket

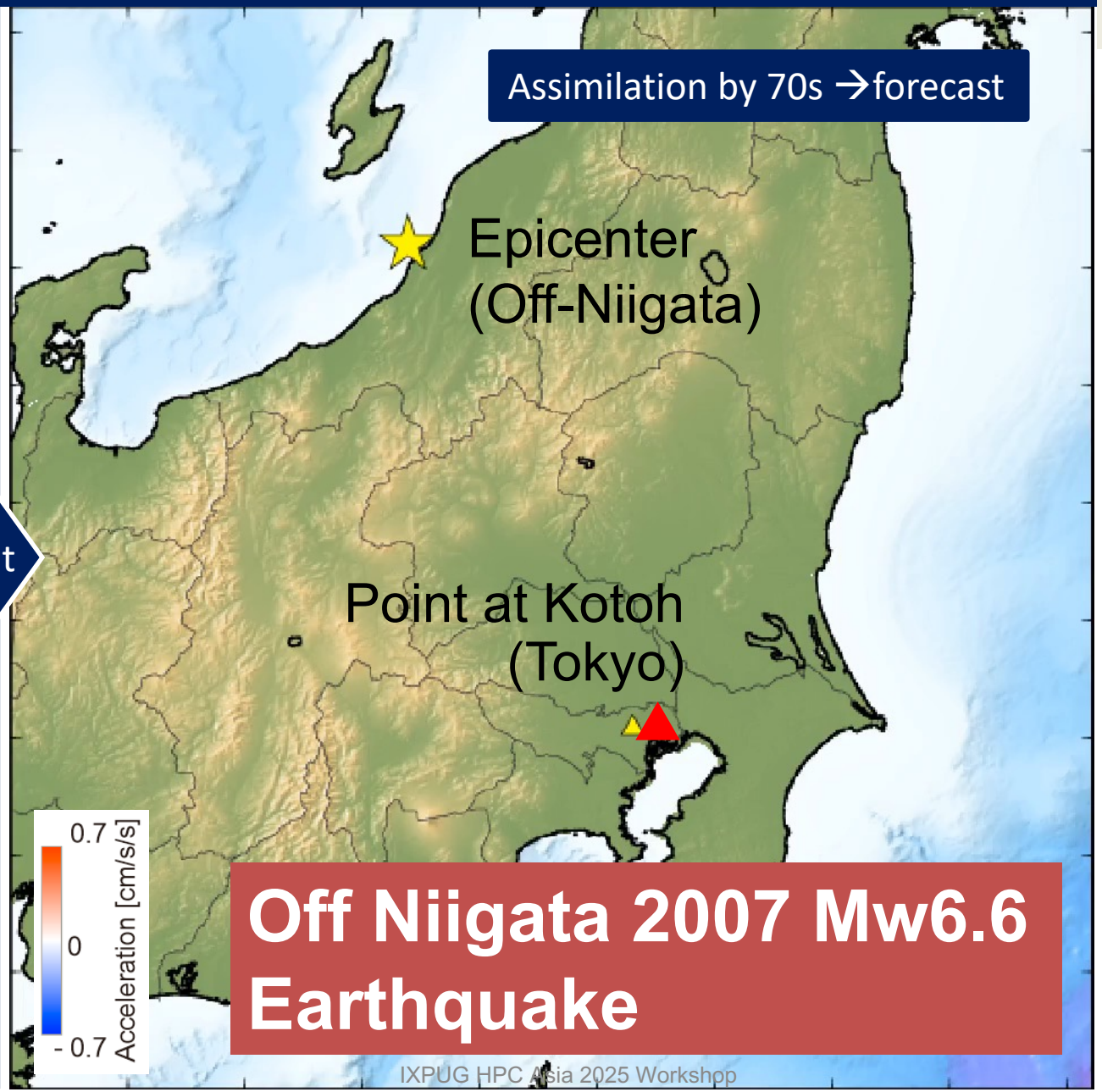
Seism3D/ OpenSWPC-DAF

Mesh Data

UG Model

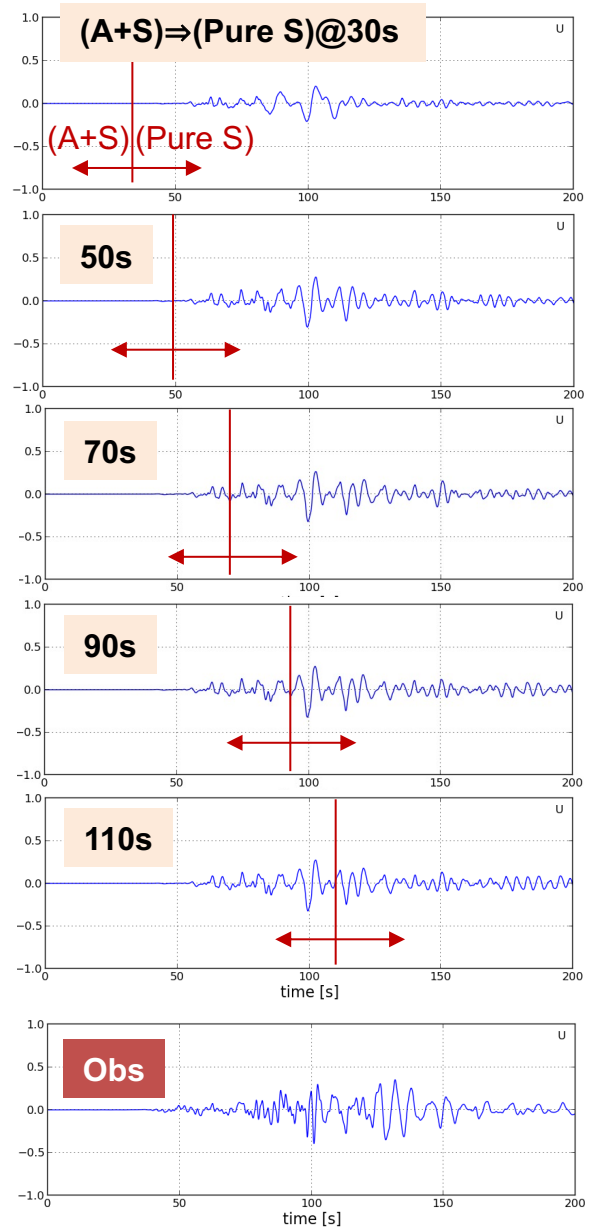
Data Assimilation + Pure Simulation/Forecast





Off Niigata 2007 Mw6.6 Earthquake

Results at Koto ▲ (N.KOTH)
N 35° 37.0'
E 139° 46.9'



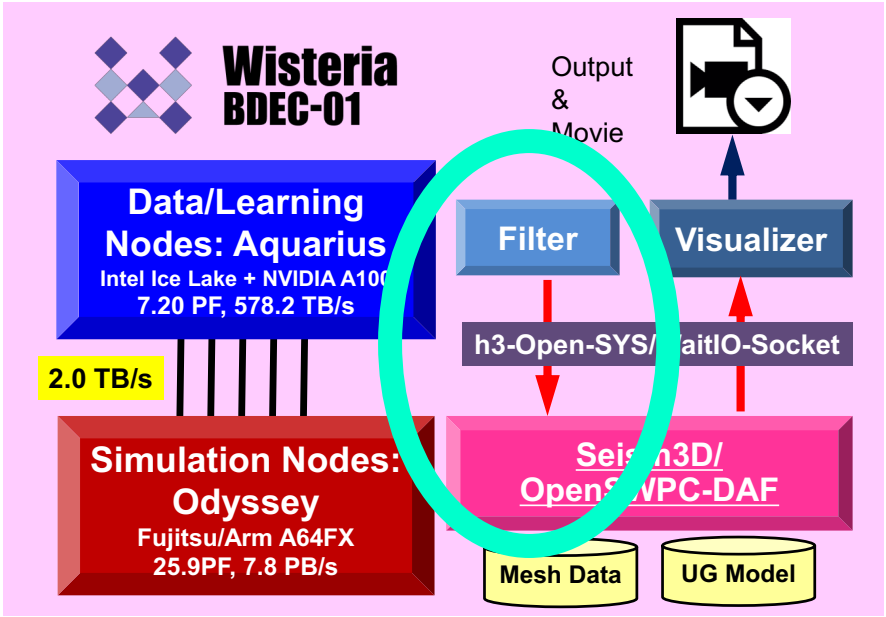
Communications by WaitIO-Socket [Kasai et al. 2021]

Aquarius: SEND

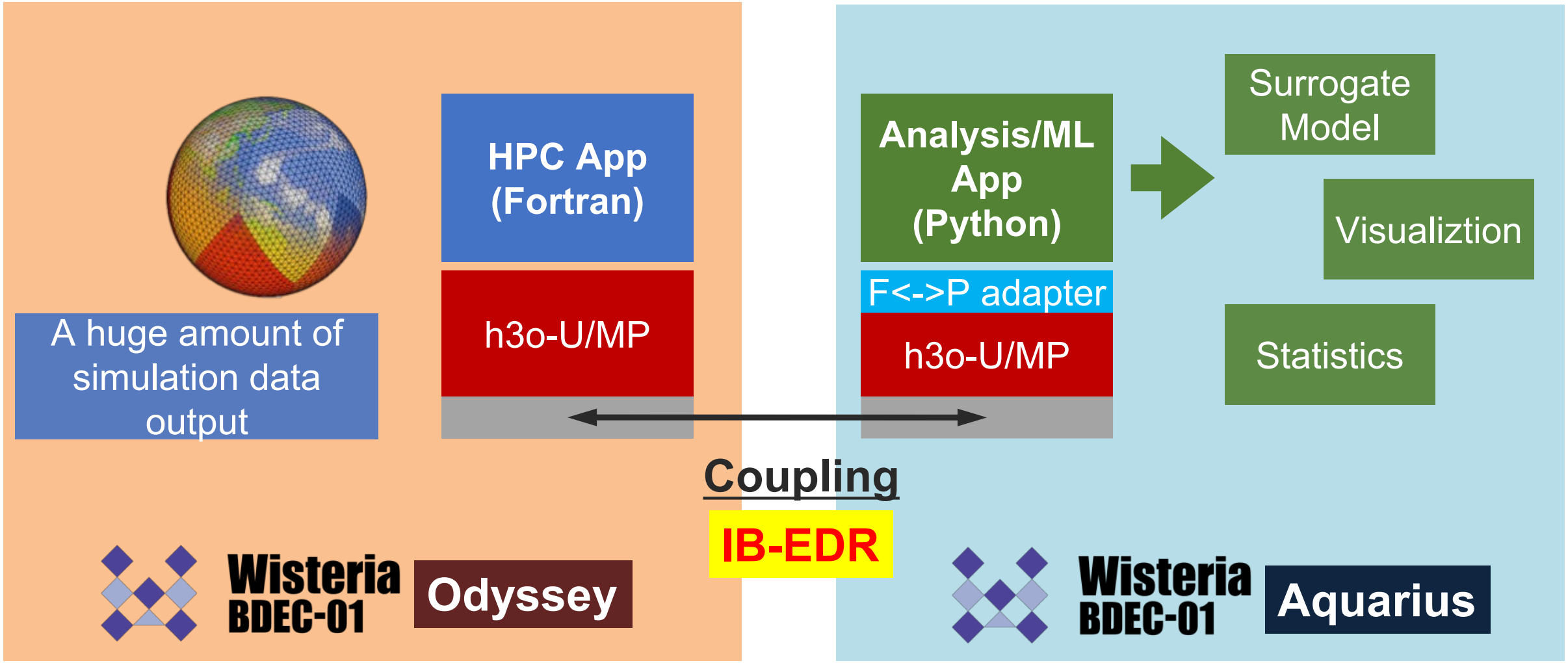
```
program dmy_filter
<省略: 型宣言等>
call mpi_init (ierr)
call mpi_comm_size (MPI_COMM_WORLD, nprocs, ierr)
call mpi_comm_rank (MPI_COMM_WORLD, myrank,  ierr)
call WAITIO_CREATE_UNIVERSE (WAITIO_COMM_UNIVERSE, ierr)

if (myrank==0) then
  open(100,file='./obsfile_list.txt', form='formatted', status='old', iostat=ierr)
  do i=1,300
    <省略: obsデータ読み込み処理>
    print *,"Send obs data ...... "
    call WAITIO_MPI_ISEND (NTMAX1_o, 1,           WAITIO_MPI_INTEGER,  2,1, WAITIO_COMM_UNIVERSE,req(1,1), ierr)
    call WAITIO_MPI_ISEND (DT_o,     1,           WAITIO_MPI_FLOAT,    2,2, WAITIO_COMM_UNIVERSE,req(1,2), ierr)
    call WAITIO_MPI_ISEND (NST_o,    1,           WAITIO_MPI_INTEGER,  2,3, WAITIO_COMM_UNIVERSE,req(1,3), ierr)
    call WAITIO_MPI_ISEND (AT_o,     1,           WAITIO_MPI_FLOAT,    2,4, WAITIO_COMM_UNIVERSE,req(1,4), ierr)
    call WAITIO_MPI_ISEND (T0_o,     1,           WAITIO_MPI_FLOAT,    2,5, WAITIO_COMM_UNIVERSE,req(1,5), ierr)
    call WAITIO_MPI_ISEND (ISO_X_o,  NSMAX,       WAITIO_MPI_INTEGER,  2,6, WAITIO_COMM_UNIVERSE,req(1,6), ierr)
    call WAITIO_MPI_ISEND (ISO_Y_o,  NSMAX,       WAITIO_MPI_INTEGER,  2,7, WAITIO_COMM_UNIVERSE,req(1,7), ierr)
    call WAITIO_MPI_ISEND (ISO_Z_o,  NSMAX,       WAITIO_MPI_INTEGER,  2,8, WAITIO_COMM_UNIVERSE,req(1,8), ierr)
    call WAITIO_MPI_ISEND (ISTX_o,   NST,         WAITIO_MPI_INTEGER,  2,9, WAITIO_COMM_UNIVERSE,req(1,9), ierr)
    call WAITIO_MPI_ISEND (ISTY_o,   NST,         WAITIO_MPI_INTEGER,  2,10,WAITIO_COMM_UNIVERSE,req(1,10),ierr)
    call WAITIO_MPI_ISEND (ISTZ_o,   NST,         WAITIO_MPI_INTEGER,  2,11,WAITIO_COMM_UNIVERSE,req(1,11),ierr)
    call WAITIO_MPI_ISEND (STC_o,    6*NST,       WAITIO_MPI_CHAR,     2,12,WAITIO_COMM_UNIVERSE,req(1,12),ierr)
    call WAITIO_MPI_ISEND (VxAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,    2,13,WAITIO_COMM_UNIVERSE,req(1,13),ierr)
    call WAITIO_MPI_ISEND (VyAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,    2,14,WAITIO_COMM_UNIVERSE,req(1,14),ierr)
    call WAITIO_MPI_ISEND (VzAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,    2,15,WAITIO_COMM_UNIVERSE,req(1,15),ierr)
    call WAITIO_MPI_WAITALL (15,req, status, ierr)
    call sleep(1)
  enddo
  close (100)
endif
call WAITIO_FINALIZE (ierr)
call mpi_finalize (ierr)
end
```

Odyssey: RECV

```
call WAITIO_MPI_IRECV (NTMAX1_o, 1,           WAITIO_MPI_INTEGER, 0,1, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (DT_o,     1,           WAITIO_MPI_FLOAT,   0,2, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (NST_o,    1,           WAITIO_MPI_INTEGER, 0,3, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (AT_o,     1,           WAITIO_MPI_FLOAT,   0,4, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (T0_o,     1,           WAITIO_MPI_FLOAT,   0,5, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISO_X_o,  NSMAX,       WAITIO_MPI_INTEGER, 0,6, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISO_Y_o,  NSMAX,       WAITIO_MPI_INTEGER, 0,7, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISO_Z_o,  NSMAX,       WAITIO_MPI_INTEGER, 0,8, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISTX_o,   NST,         WAITIO_MPI_INTEGER, 0,9, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISTY_o,   NST,         WAITIO_MPI_INTEGER, 0,10,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISTZ_o,   NST,         WAITIO_MPI_INTEGER, 0,11,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (STC_o,    6*NST,       WAITIO_MPI_CHAR,    0,12,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (VxAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,   0,13,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (VyAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,   0,14,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (VzAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,   0,15,WAITIO_COMM_UNIVERSE,...)
```



h3-Open-UTIL/MP (h3o-U/MP) Extended Multiphysics Coupler



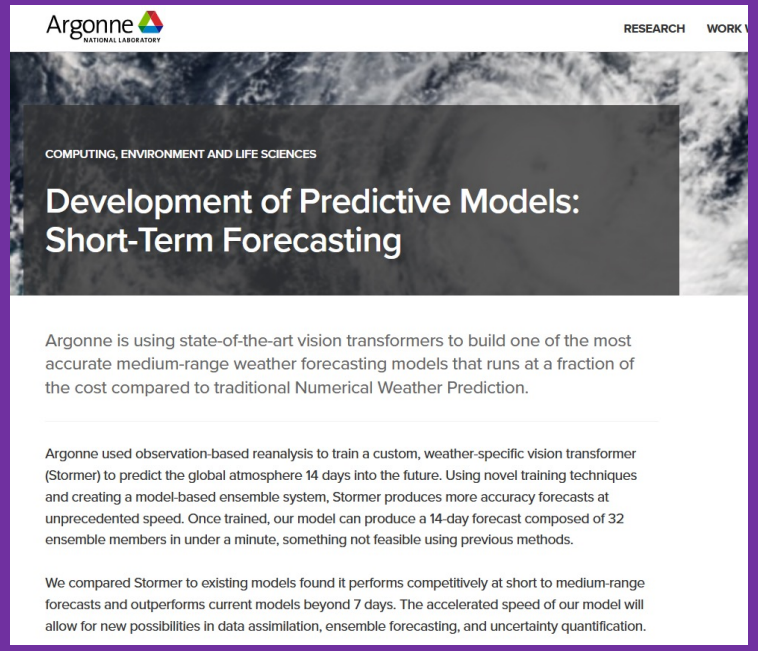
[Sumimoto, Arakawa, Yashiro et al. 2022]

2. AI for Science (1/2)

based on Integration of (Simulation/Data/Learning)

miyabi

- **Innovation in Scientific Research using Generative AI**
  - **Traditional "AI for Science": Data-driven Science**
- ANL's "Stormer (Short-Term Weather Forecasting)"
  - https://www.anl.gov/cels/development-of-predictive-models-shortterm-forecasting
- We support this type of activities of users, who want to utilize AI to CSE



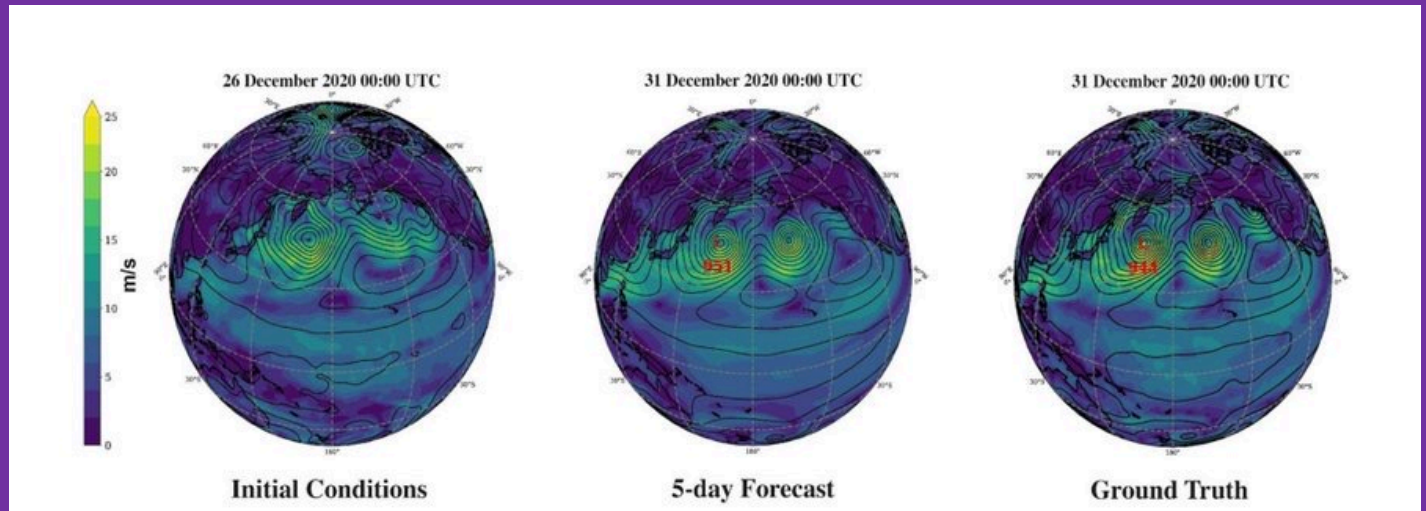
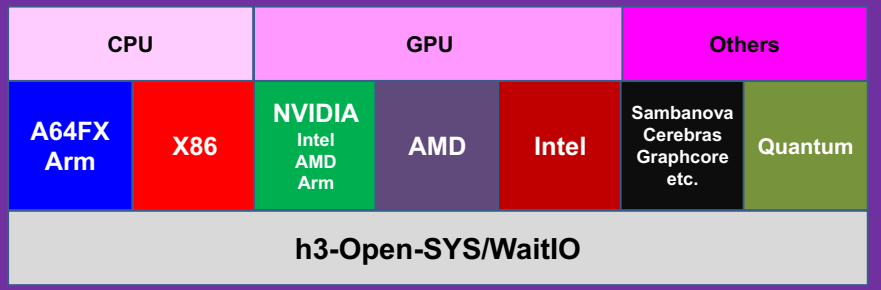
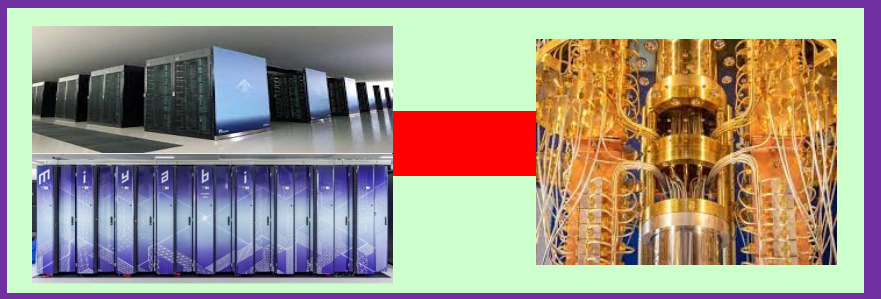


Figure 1: On December 31, 2020, an extratropical cyclone impacted Alaska setting a new North Pacific low-pressure record. Stormer forecasts both the location and strength of this extreme event with great accuracy 5 days in advance.

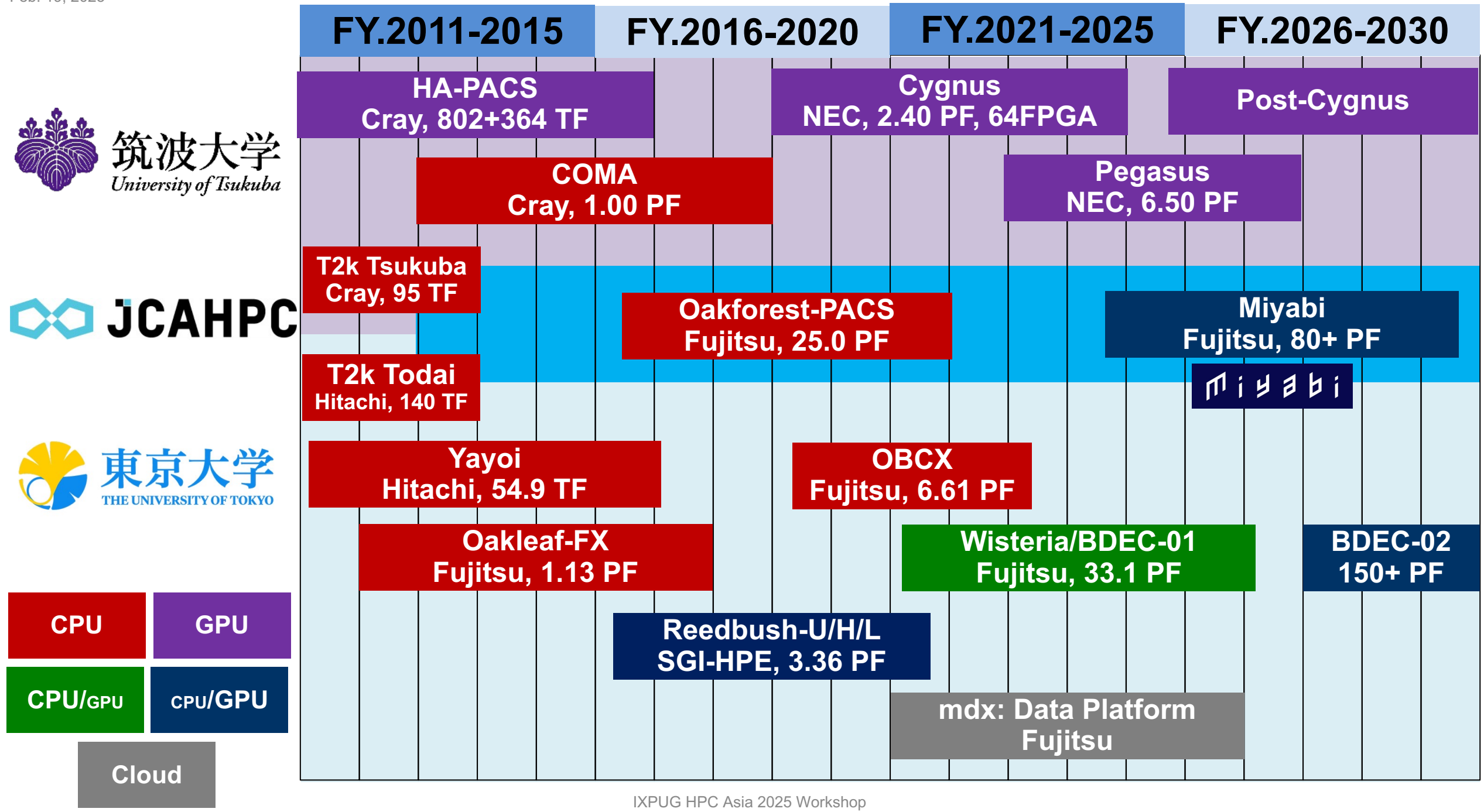
3. QC-HPC Hybrid Computing

- JHPC-quantum (FY.2023-FY.2028)
  - https://jhpc-quantum.org/
  - RIKEN, Softbank, U.Tokyo, Osaka U.
    - supported by Japanese Government (METI/NEDO)
  - Two Real Quantum Computers will be installed
    - IBM's Superconducting QC at RIKEN-Kobe (100+Qubit)
    - Quantinuum's Ion-Trap QC at RIKEN-Wako (20+Qubit)
  - Target Applications
    - Quantum Science, Quantum Machine Learning etc.

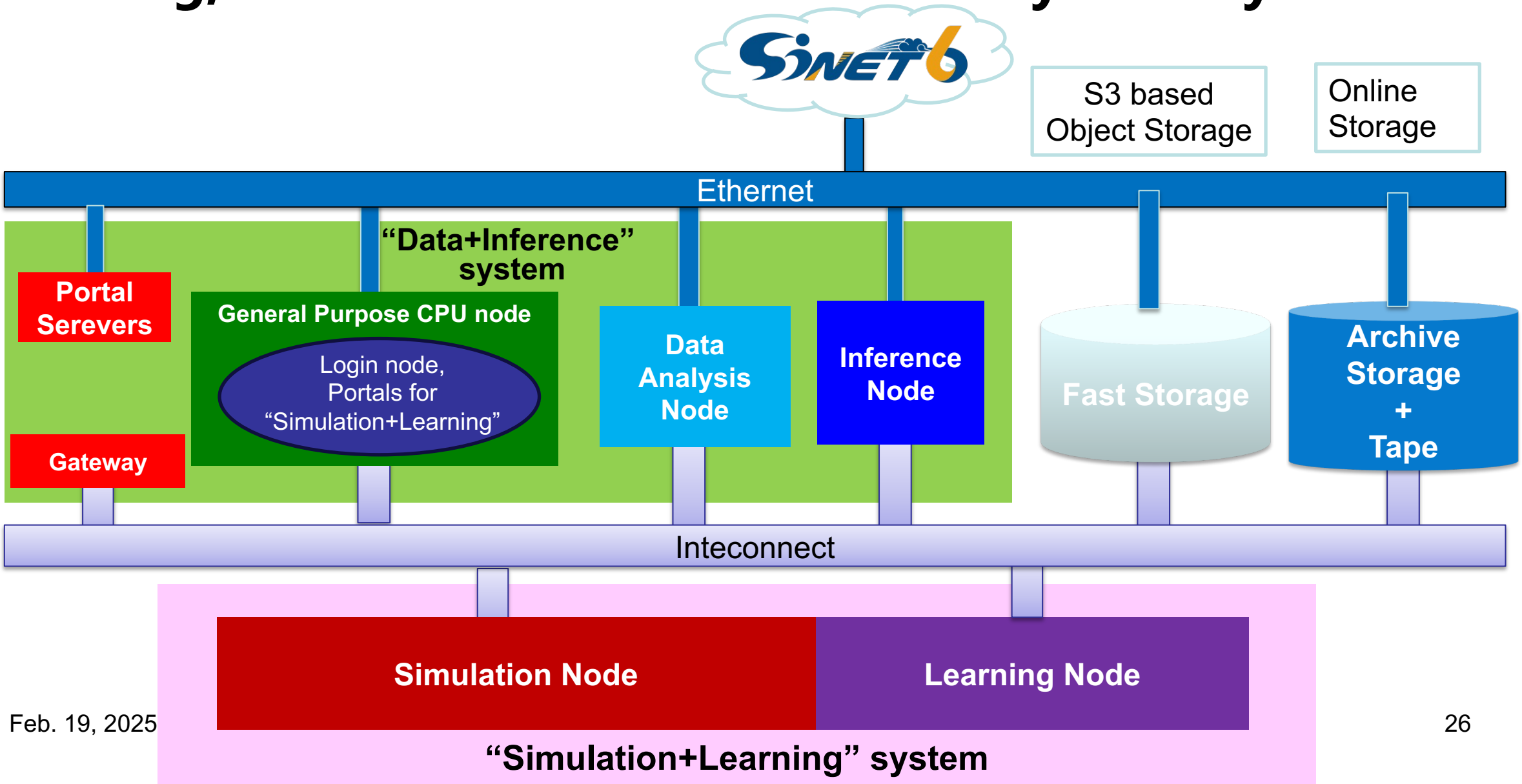




- Role of U.Tokyo
  - R&D on System SW for QC-HPC Hybrid Environment (QC as Accelerators)
    - Extension of h3-Open-BDEC
- **Fugaku & Miyabi will be connected to QCs in March 2026 (or before)**

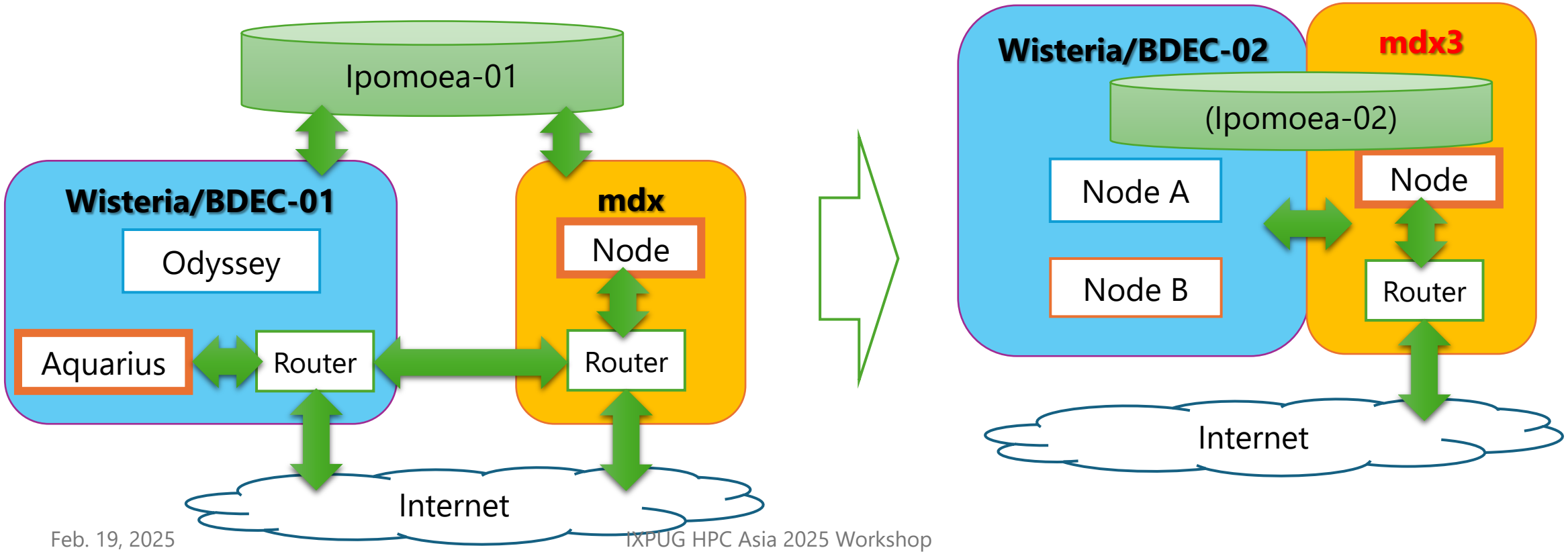

FY.2011-2015
FY.2016-2020
FY.2021-2025
FY.2026-2030
筑波大学
University of Tsukuba
JCAHPC
東京大学
THE UNIVERSITY OF TOKYO
HA-PACS
Cray, 802+364 TF
Cygnus
NEC, 2.40 PF, 64FPGA
Post-Cygnus
COMA
Cray, 1.00 PF
Pegasus
NEC, 6.50 PF
T2k Tsukuba
Cray, 95 TF
Oakforest-PACS
Fujitsu, 25.0 PF
Miyabi
Fujitsu, 80+ PF
T2k Todai
Hitachi, 140 TF
Yayoi
Hitachi, 54.9 TF
OBCX
Fujitsu, 6.61 PF
Oakleaf-FX
Fujitsu, 1.13 PF
Wisteria/BDEC-01
Fujitsu, 33.1 PF
BDEC-02
150+ PF
Reedbush-U/H/L
SGI-HPE, 3.36 PF
mdx: Data Platform
Fujitsu
CPU
GPU
CPU/GPU
CPU/GPU
Cloud

“Integrated Infrastructure System for Simulation, Data, Learning, and Inference” in the University of Tokyo



Integrated Infrastructure System for Computation, Data, Learning, and Inference

- Saving costs by sharing computer resources as much as possible and by avoiding redundant configurations.
  - Currently, Wisteria/BDEC-01 and mdx are separate systems in the same room



Summary

Miyabi

JCAHPC
筑波大学 University of Tsukuba
東京大学 THE UNIVERSITY OF TOKYO

- JCAHPC, Oakforest-PACS
- Integration of (Simulation/Data/Learning)
- **What we will do with Miyabi**
  - **Computational Science, Computational Science, Computational Science !!!**
  - **AI for Science**
  - **QC-HPC Hybrid Computing**