JCAHPC

Basic Specification of Oakforest-PACS

Joint Center for Advanced HPC (JCAHPC)

by Information Technology Center, the University of Tokyo

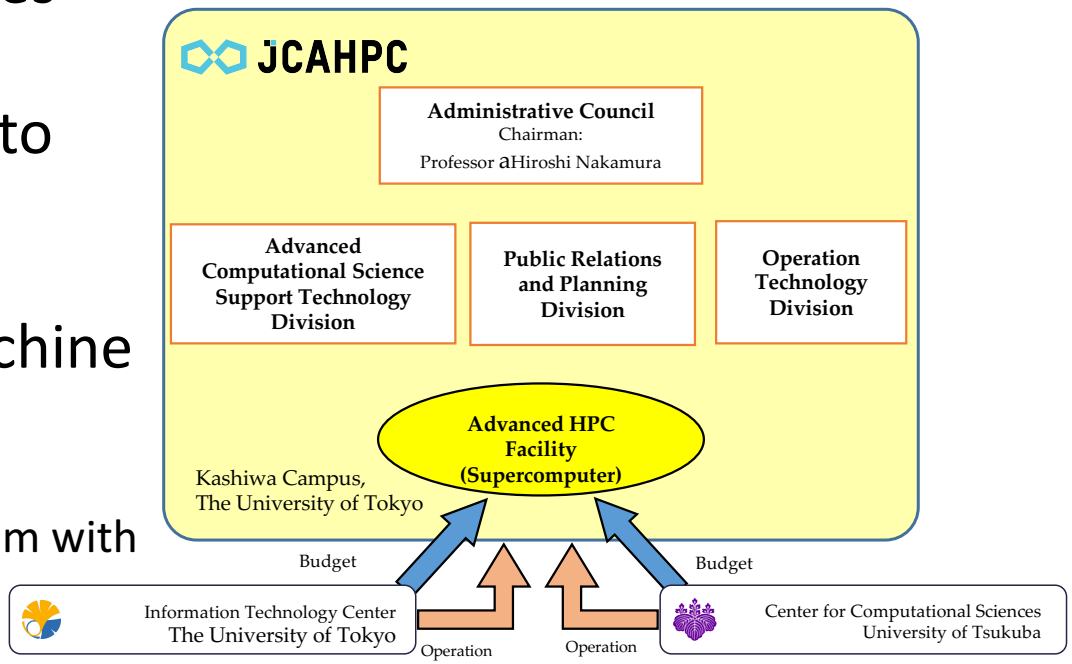
and Center for Computational Sciences, University of Tsukuba

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

筑波大学
計算科学研究センター
Center for Computational Sciences

Oakforest-PACS in JCAHPC

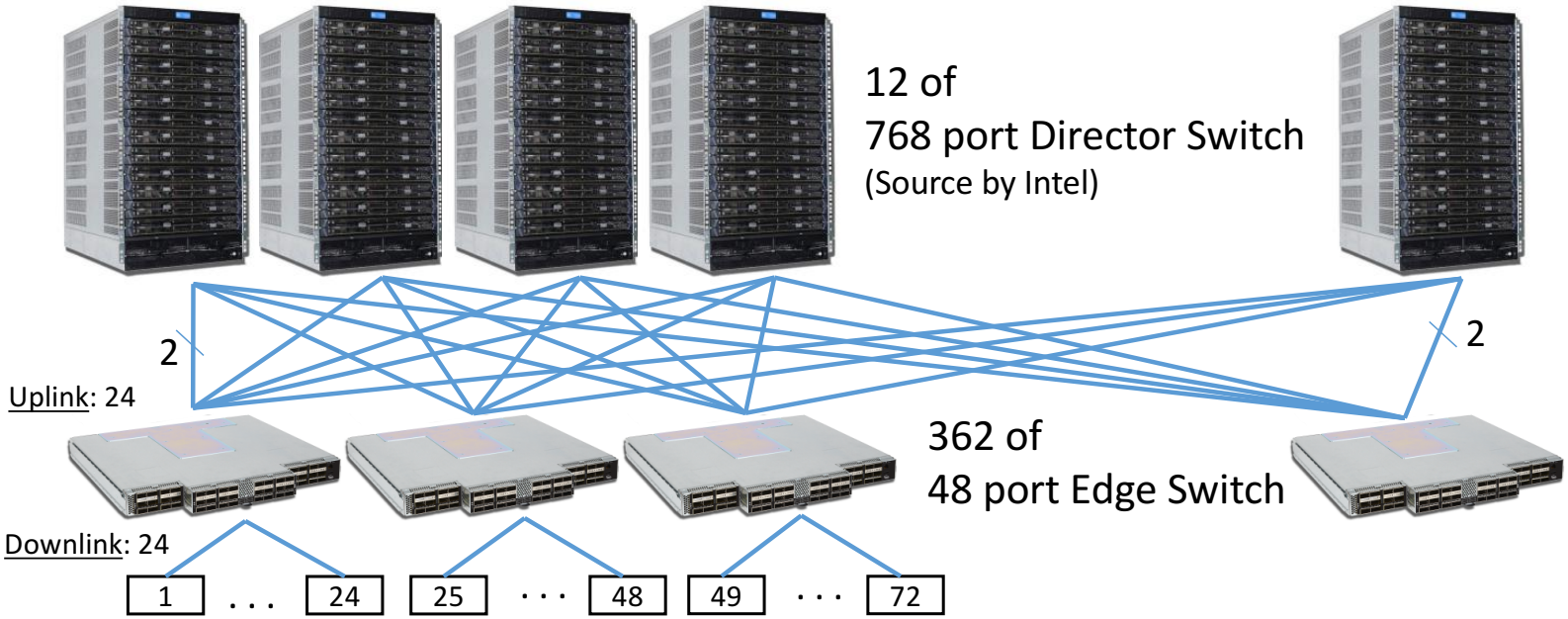
- Very tight relationship and collaboration with two universities
- For primary supercomputer resources, *uniform specification* to ***single shared system***
- Each university is financially responsible to introduce the machine and its operation
  - First attempt in Japan
  - Unified procurement toward single system with ***largest scale in Japan***



Specification of Oakforest-PACS system

Total peak performance			25 PFLOPS
Total number of compute nodes			8,208
Compute node	Product		Fujitsu PRIMERGY CX600 M1 (2U) + CX1640 M1 x 8node
	Processor		Intel® Xeon Phi™ 7250 (Code name: Knights Landing), 68 cores, 1.4 GHz
	Memory	High BW	16 GB, 490 GB/sec (MCDRAM, effective rate)
		Low BW	96 GB, 115.2 GB/sec (peak rate)
Interconnect	Product		Intel® Omni-Path Architecture
	Link speed		100 Gbps
	Topology		Fat-tree with (completely) full-bisection bandwidth

Full bisection bandwidth Fat-tree by Intel® Omni-Path Architecture



Firstly, to reduce switches&cables, we considered :
- All the nodes into subgroups are connected with FBB Fat-tree
- Subgroups are connected with each other with >20% of FBB

But, HW quantity is not so different from globally FBB, and globally FBB is preferred for flexible job management.

Specification of Oakforest-PACS system (Cont'd)

Parallel File System	Type	Lustre File System
	Total Capacity	26.2 PB
	Product	DataDirect Networks SFA14KE
	Aggregate BW	500 GB/sec
File Cache System	Type	Burst Buffer, Infinite Memory Engine (by DDN)
	Total capacity	940 TB (NVMe SSD, including parity data by erasure coding)
	Product	DataDirect Networks IME14K
	Aggregate BW	1,560 GB/sec
Power consumption		4.2 MW (including cooling)
# of racks		102

JCAHPC

Software of Oakforest-PACS system

	Compute node	Login node
OS	CentOS 7, McKernel	Red Hat Enterprise Linux 7
Compiler	gcc, Intel compiler (C, C++, Fortran)	
MPI	Intel MPI, MVAPICH2	
Library	Intel MKL LAPACK, FFTW, SuperLU, PETSc, METIS, Scotch, ScaLAPACK, GNU Scientific Library, NetCDF, Parallel netCDF, Xabclib, ppOpen-HPC, ppOpen-AT, MassiveThreads	
Application	mpijava, XcalableMP, OpenFOAM, ABINIT-MP, PHASE system, FrontFlow/blue, FrontISTR, REVOCAP, OpenMX, xTAPP, AkaiKKR, MODYLAS, ALPS, feram, GROMACS, BLAST, R packages, Bioconductor, BioPerl, BioRuby	
Distributed FS		Globus Toolkit, Gfarm
Job Scheduler	Fujitsu Technical Computing Suite	
Debugger	Allinea DDT	
Profiler	Intel VTune Amplifier, Trace Analyzer & Collector	

JCAHPC

Software of Oakforest-PACS

- OS: Red Hat Enterprise Linux (Login nodes),
  CentOS or McKernel (Compute nodes, dynamically switchable)
  - McKernel: OS for many-core CPU developed by RIKEN AICS
    - Ultra-lightweight OS compared with Linux, no background noise to user program
    - Will be deployed to post-K computer
- Compiler: GCC, Intel Compiler, XcalableMP
  - XcalableMP: Parallel programming language developed by RIKEN AICS and University of Tsukuba
    - Easy to develop high-performance parallel application by adding directives to original code written by C or Fortran
- Application: Open-source software
  - OpenFOAM, ABINIT-MP, PHASE system, FrontFlow/blue, and so on

JCAHPC

Photo of computation node

Computation node (Fujitsu PRIMERGY CX1640 M1)
with single chip Intel Xeon Phi (Knights Landing, 3+TFLOPS)
and Intel Omni-Path Architecture card (100Gbps)

Chassis with 8 nodes, 2U size
(Fujitsu PRIMERGY CX600 M1)

Machine location: Kashiwa Campus of U. Tokyo