



東京大学
THE UNIVERSITY OF TOKYO

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システムの ご紹介

東京大学情報基盤センター

スーパーコンピューティング研究部門

<http://www.cc.u-tokyo.ac.jp/>

問合せ先: uketsuke@cc.u-tokyo.ac.jp

- 背景
- 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システム (Reedbush) 概要
- スケジュール
- 運用・サービス
 - トークン制
 - 教育利用, 若手利用
 - 企業利用
 - トライアルユース
- 試験運転期間のサービス
- その他
- 質疑

東京大学情報基盤センター

スーパーコンピューティング研究部門・本部情報システム部
情報基盤課スーパーコンピューティングチーム
(スーパーコンピューティング部門)(1/3)

<http://www.cc.u-tokyo.ac.jp/>

- 東京大学大型計算機センター(1965-1999)
 - 東京大学情報基盤センター(1999-)
 - スーパーコンピュータの運用, 利用支援
 - Hitachi SR16000(Yayoi)(2011.10-)
 - Hitachi HA8000(T2K東大)(2008.6-2014.3)
 - Fujitsu PRIMEHPC FX10(Oakleaf-FX)(2012.4-)
 - Fujitsu PRIMEHPC FX10(Oakbridge-FX)(2014.4-)
- 研究開発型, 市場調達型
- 合計約2,000人のユーザー(学外が半分)
- 大学(研究, 教育), 研究機関, 企業

東大センターのスパコン

2基の大型システム, 6年サイクル

FY

08 09 10 11 12 13 14 15 16 17 18 19 20 21 22

Hitachi SR11K/J2
IBM Power-5+
18.8TFLOPS, 16.4TB

Yayoi: Hitachi SR16000/M1
IBM Power-7
54.9 TFLOPS, 11.2 TB

Hitachi HA8000 (T2K)
AMD Opteron
140TFLOPS, 31.3TB

Oakleaf-FX: Fujitsu PRIMEHPC
FX10, SPARC64 IXfx
1.13 PFLOPS, 150 TB

Oakbridge-FX
136.2 TFLOPS, 18.4 TB

Peta

K

K computer

Post-K



東大情報基盤センターのスパコン 利用者総数 > 2,000

Oakleaf-FX (Fujitsu PRIMEHPC FX10)

Total Peak performance : 1.13 PFLOPS
Total number of nodes : 4800
Total memory : 150 TB
Peak performance / node : 236.5 GFLOPS
Main memory per node : 32 GB
Disk capacity : 1.1 PB + 2.1 PB
SPARC64 lxfx 1.84GHz

since April 2012

Oakbridge-FX (Fujitsu PRIMEHPC FX10)

Total Peak performance : 136.2 TFLOPS
Total number of nodes : 576
Total memory : 18.4 TB
Peak performance / node : 236.5 GFLOPS
Main memory per node : 32 GB
Disk capacity : 147TB + 295TB
SPARC64 lxfx 1.84GHz

since April 2014
Special System for Long-Term Jobs up
to 168 hours

一週間までの長時間利用可能

Yayoi (Hitachi SR16000/M1)

Total Peak performance : 54.9 TFLOPS
Total number of nodes : 56
Total memory : 11200 GB
Peak performance / node : 980.48 GFLOPS
Main memory per node : 200 GB
Disk capacity : 556 TB
IBM POWER 7 3.83GHz

since November 2011



東京大学情報基盤センター

スーパーコンピューティング部門(2/3)

● 研究

- 教員数11名(兼任・客員含む)
- システムソフトウェア, 数値アルゴリズム, 自動チューニング
- 利用者との共同研究, JHPCN(共同利用・共同研究拠点)
- 国内・国際プロジェクト(次ページ)

● 普及・人材育成

- 学際計算科学・工学人材育成プログラム, 計算科学アライアンス
 - 全学的なHPC(High-Performance Computing)教育
- お試しアカウント付き講習会(Oakleaf-FX)
- 若手・女性育成制度
- RIKEN AICS Summer/Spring School 共催

● 広報活動

- スーパーコンピューティングニュース(年6回+特集号)

東京大学情報基盤センター

スーパーコンピューティング部門(3/3)

● 研究プロジェクト

- Framework and Programming for Post Petascale Computing (FP3C) (JST-ANR, 2010-13)
 - 日仏共同プロジェクト, T3K(T2K+東工大)
- 将来のHPCIシステムのあり方の調査研究(文科省, 2012-13)
- ppOpen-HPC(JST-CREST, 2011-15)
- ESSEX-II (JST/DFG-SPPEXA, 2016-2018) 日独共同
- 理化学研究所共同研究(ポスト京)(2014-)

● 国内

- 学際大規模情報基盤共同利用・共同研究拠点(JHPCN)(中核拠点)(2010-)
- HPCIコンソーシアム(資源提供機関)(2012-)

● 国際協力(非居住者:利用手続き簡略化, ホワイト包括)

- Lawrence Berkeley National Laboratory (USA)
- 國立臺灣大學(2015・2016年冬季講習), 國立中央大學(台湾)
- Intel Parallel Computing Center (IPCC)

革新的ハイパフォーマンス・ コンピューティングインフラ・(HPCI) 文部科学省委託事業

<http://www.hpci-office.jp/>

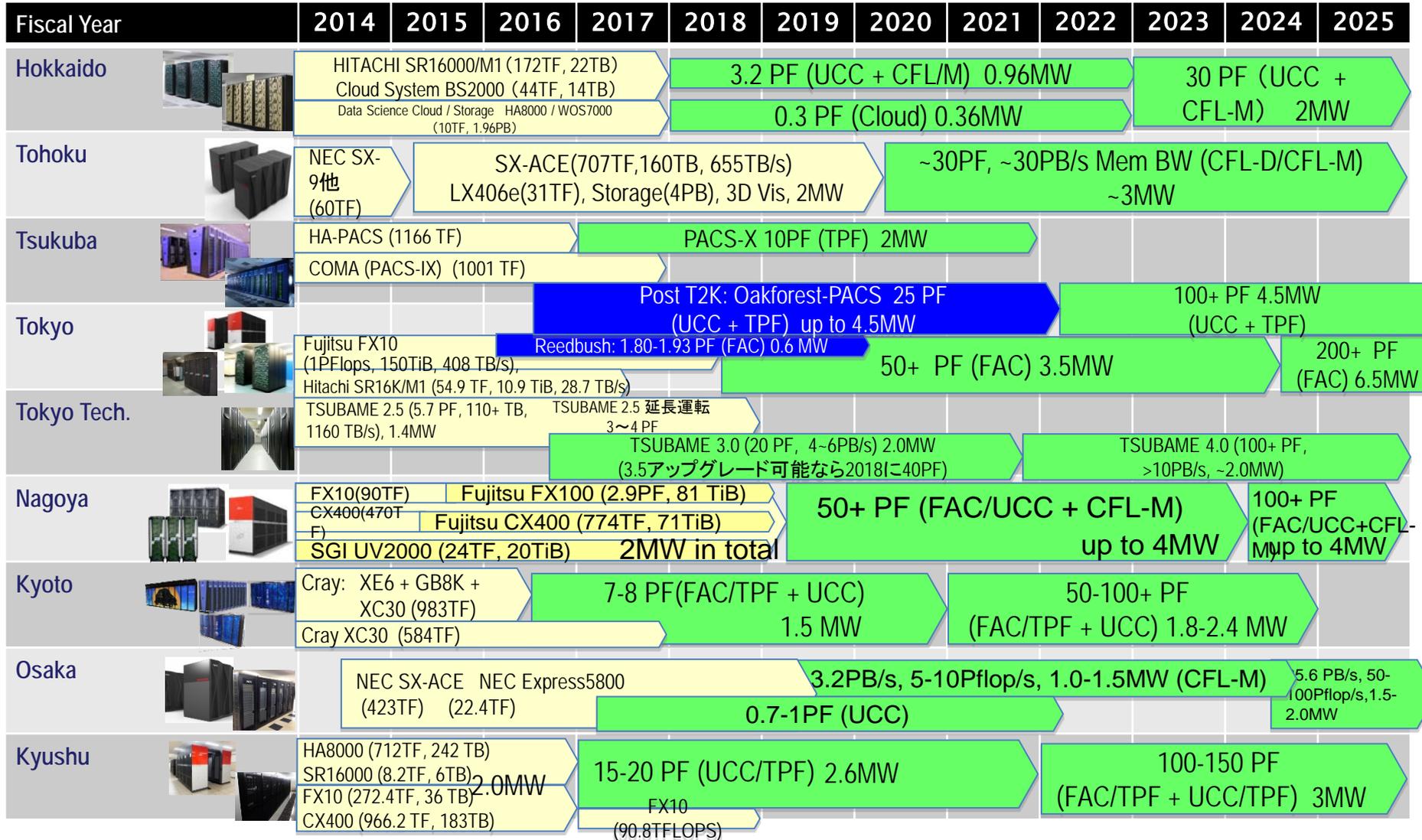
情報基盤センター群以外の会員リスト

- 使命：我が国における
 - － 計算資源(スパコン,
大規模ストレージ(東西拠点))
 - － 計算科学推進(HPCI戦略プログラム
⇒ポスト京重点課題)
- HPCIコンソーシアム(2012～)
 - － HPCI計算資源運用
 - － 産官学
 - － 資源提供者・利用者によるコミュニティ
 - － 2012年度発足

一般社団法人日本流体力学会
財団法人計算科学振興財団
特定非営利活動法人バイオグリッドセンター関西
自然科学研究機構核融合科学研究所
スーパーコンピューティング技術産業応用協議会
神戸大学
東京大学物性研究所計算物質科学研究センター計算物質科学イニシアティブ(分野2「新物質・エネルギー創成」)
東京大学生産技術研究所(分野4「次世代ものづくり」)
計算基礎科学連携拠点(分野5「物質と宇宙の起源と構造」)
名古屋大学 太陽地球環境研究所
独立行政法人宇宙航空研究開発機構宇宙科学研究所
独立行政法人海洋研究開発機構
一般社団法人日本計算工学会
計算生命科学ネットワーク
国立研究開発法人理化学研究所計算科学研究機構
高エネルギー加速器研究機構 共通基盤研究施設・計算科学センター
情報・システム研究機構 国立情報学研究所
一般財団法人高度情報科学技術研究機構
筑波大学 計算科学研究センター
大阪大学 核物理研究センター
国立研究開発法人産業技術総合研究所 情報技術研究部門
東京大学 物性研究所
東北大学 金属材料研究所
情報・システム研究機構 統計数理研究所
自然科学研究機構分子科学研究所 計算科学研究センター
独立行政法人宇宙航空研究開発機構 情報計算工学センター

9 情報基盤センター運用&整備計画 (2016年5月時点)

(参考)



電力は最大供給量(空調システム含む)

HPCI共用ストレージ

- 文科省委託費
- 東拠点(東京大学 柏キャンパス) 13PB
- 西拠点(理研) 10PB

ストレージ

- W1 storage
 - Gfarm
 - DDN SFA10000(Total10PB)
 - メタデータサーバ2台
 - データサーバ16台
 - 10GbE ネットワーク
- 60 PB tape archive
- データ解析システム
 - 理論ピーク性能 12.37TFlops
 - 総主記憶容量8.4TB
 - 計算ノード88台 ログインノード2台



理研

- E1 storage
 - Gfarm
 - DDN SFA10000 (Total 8PB)
 - データサーバ36台
 - 10GbE ネットワーク
- E2 storage
 - Gfarm
 - DDN SFA10000(Total 5.5PB)
 - メタデータサーバ2台
 - データサーバ8台
 - 10GbE ネットワーク
- 20 PB tape archive



東京大学情報基盤センター



HPCI: High Performance Computing Infrastructure

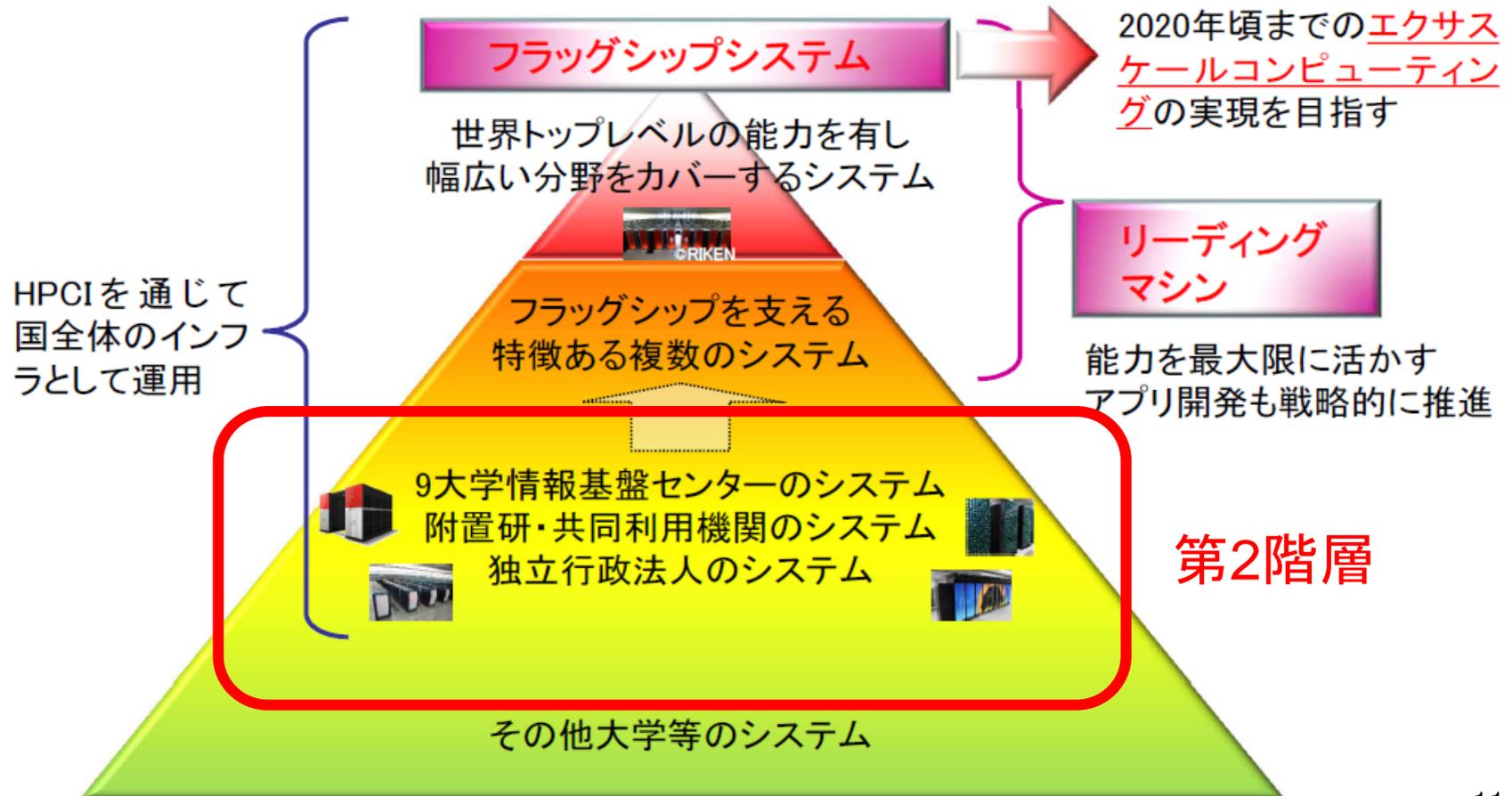
日本全体におけるスパコンインフラ

今後のHPCI 計画推進の在り方について(H26/3)より

我が国の次期スパコン開発の方向性

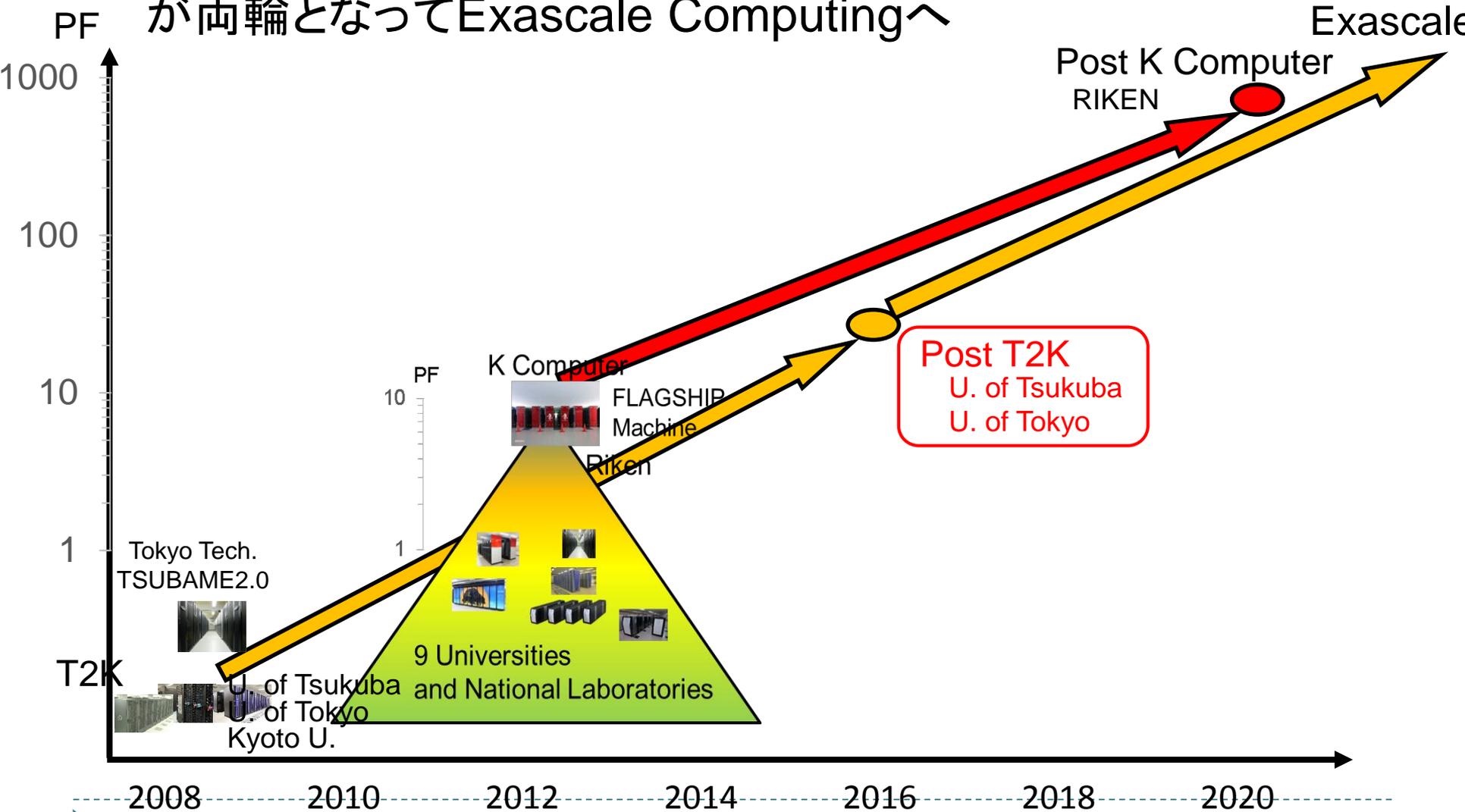
<我が国の計算科学技術インフラのイメージ>

http://www.mext.go.jp/b_menu/shingi/chousa/shinkou/028/gaiyou/1348991.htm



フラグシップを支えるリーディング(第2階層)マシン

第1階層(理研:京コンピュータ)と第2階層
が両輪となってExascale Computingへ



東大センターのスパコン

2基の大型システム, 6年サイクル

FY

08 09 10 11 12 13 14 15 16 17 18 19 20 21 22

Hitachi SR11K/J2
IBM Power-5+
18.8TFLOPS, 16.4TB

Yayoi: Hitachi SR16000/M1
IBM Power-7
54.9 TFLOPS, 11.2 TB

Hitachi HA8000 (T2K)
AMD Opteron
140TFLOPS, 31.3TB

Oakforest-PACS
Fujitsu, Intel KNL
25PFLOPS, 919.3TB

Post T2K
JCAHPC:
筑波大・
東大

Oakleaf-FX: Fujitsu PRIMEHPC
FX10, SPARC64 IXfx
1.13 PFLOPS, 150 TB

Oakbridge-FX
136.2 TFLOPS, 18.4 TB

Peta

K

K computer

Post-K



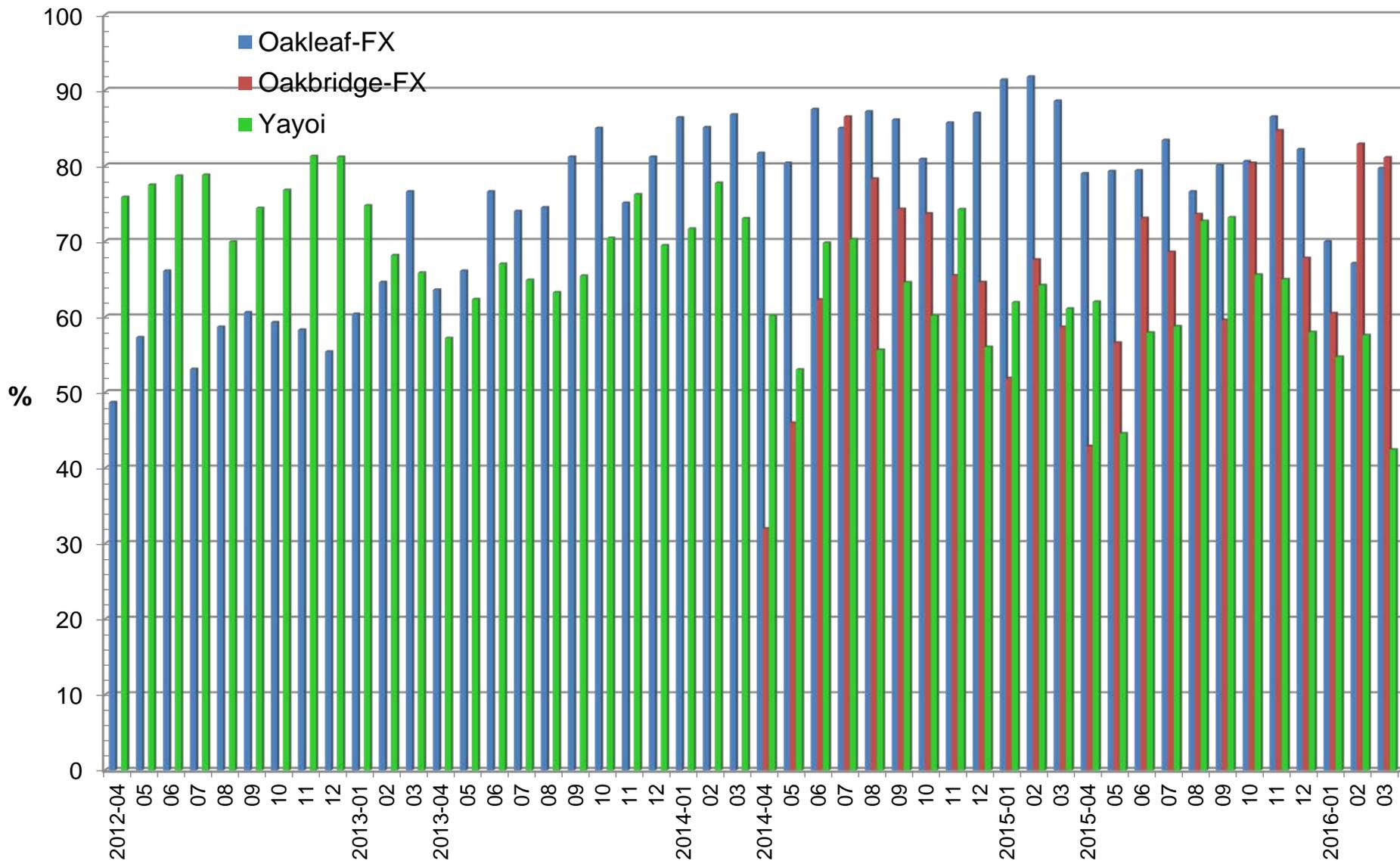
ポストT2K: Oakforest-PACS

<http://www.cc.u-tokyo.ac.jp/system/ofp/>

- 2016年12月1日稼働開始
- 8,208 Intel Xeon/Phi (KNL), ピーク性能25PFLOPS
 - 富士通が構築
- **最先端共同HPC 基盤施設(JCAHPC: Joint Center for Advanced High Performance Computing)**
 - 筑波大学計算科学研究センター
 - 東京大学情報基盤センター
 - 東京大学柏キャンパスの東京大学情報基盤センター内に、両機関の教職員が中心となって設計するスーパーコンピュータシステムを設置し、最先端の大規模高性能計算基盤を構築・運営するための組織
 - <http://jcahpc.jp>

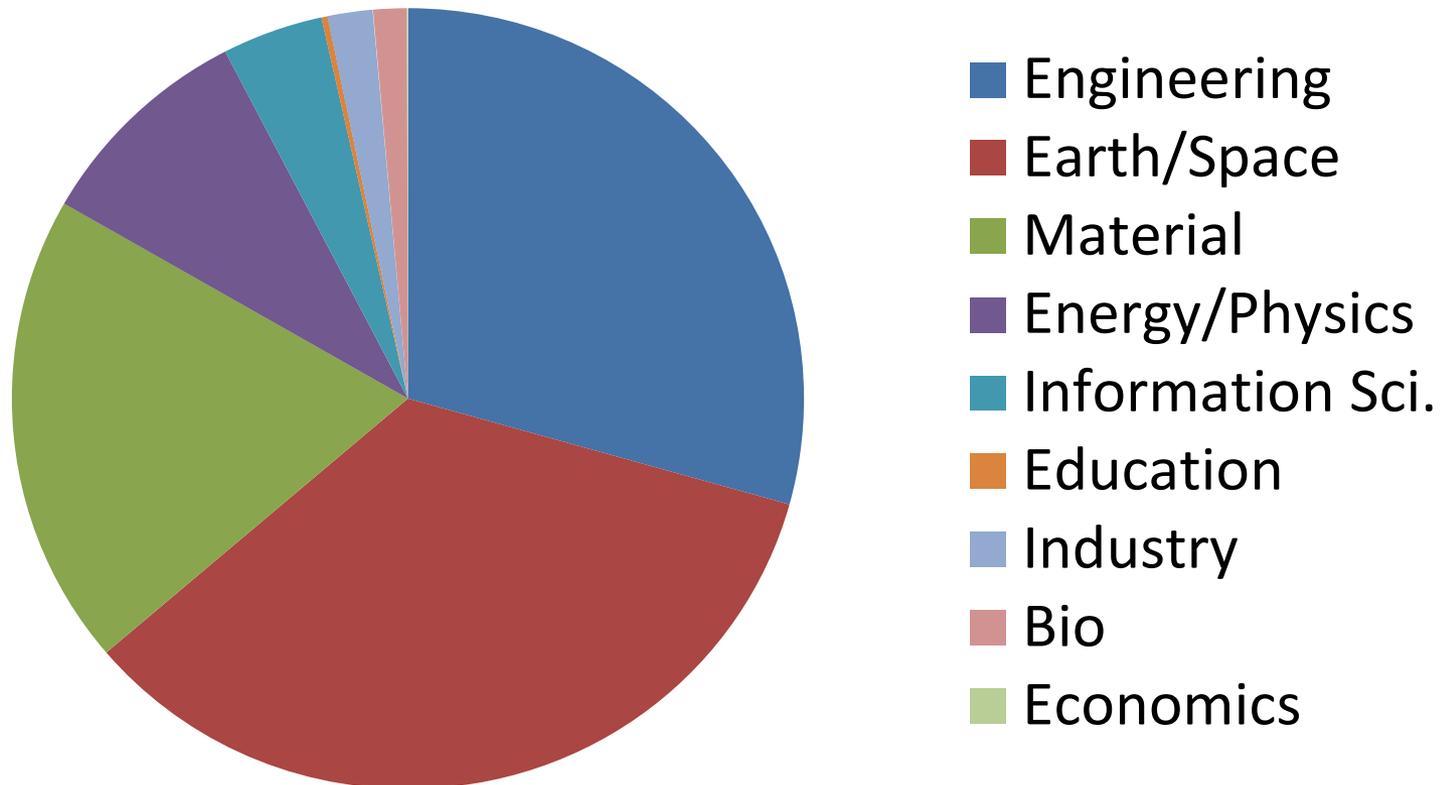
利用率

80+% Average
Oakleaf-FX + Oakbridge-FX



利用ノード時間積による利用分野 FX10 (Oakleaf/Oakbridge-FX) : FY.2015

グループ+教育+企業+HPCI
学内:47%, 学外:53%



- 背景
- **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システム(Reedbush)概要**
- スケジュール
- 運用・サービス
 - トークン制
 - 教育利用, 若手利用
 - 企業利用
 - トライアルユース
- 試験運転期間のサービス
- その他
- 質疑

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システム：導入の背景

- Post T2K (Oakforest-PACS)
 - 25 PF with KNL: 筑波大と共同で導入 (JCAHPC)
 - 2016年12月1日運用開始予定
- Mini Post T2K (MPT2K): 東大のみで導入
 - FX10混雑緩和, PT2K移行のための中間的なシステム
 - Post FX10 (2018年秋稼働開始予定) パイロットシステム
 - ✓ データ解析, Deep Learningなど、新規利用分野開拓
- スケジュール
 - RFI(資料提供招請公告): 2015年8月18日
 - RFP(入札公告): 2016年1月15日
 - 入札: 2016年2月29日, 開札: 2016年3月22日
 - **運用開始: 2016年7月1日: PT2Kより早い(全系は来年3月1日)**

東大センターのスパコン

2基の大型システム, 6年サイクル

FY

08 09 10 11 12 13 14 15 16 17 18 19 20 21 22

Hitachi SR11K/J2
IBM Power-5+
18.8TFLOPS, 16.4TB

Yayoi: Hitachi SR16000/M1
IBM Power-7
54.9 TFLOPS, 11.2 TB

Hitachi HA8000 (T2K)
AMD Opteron
140TFLOPS, 31.3TB

Oakforest-PACS
Fujitsu, Intel KNL
25PFLOPS, 919.3TB

Post T2K
JCAHPC:
筑波大・
東大

Oakleaf-FX: Fujitsu PRIMEHPC
FX10, SPARC64 IXfx
1.13 PFLOPS, 150 TB

Post FX10
50+ PFLOPS (?)

Oakbridge-FX
136.2 TFLOPS, 18.4 TB

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

Reedbush, SGI
Broadwell + Pascal
1.80-1.93 PFLOPS

Peta

K

K computer

Post-K

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システム：要求仕様(概要)

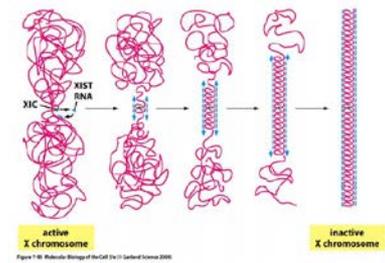
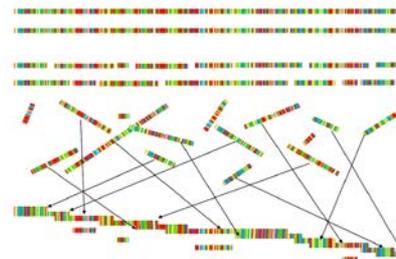
- 2タイプの計算ノード
- 汎用計算ノード：CPU Only
 - 各ノード：1.2TF, 256GB, 150GB/sec, 合計：400TF+
- 演算加速ノード：CPU+アクセラレータ：東大初のアクセラレータ搭載システム
 - ホストCPUのアーキテクチャは汎用計算ノードと異なってよい
 - 各アクセラレータ：GPUを想定, 4TF, 16GB, 1TB/sec, 合計：960TF+(240枚以上)
- ファイルシステム
 - 共有ファイルシステム：4PB, 75GB/sec
 - 高速ファイルキャッシュシステム：150TB, 200GB/sec
- 空冷, 500kVA

何故，東大情報基盤センターはGPUを導入してこなかったか？

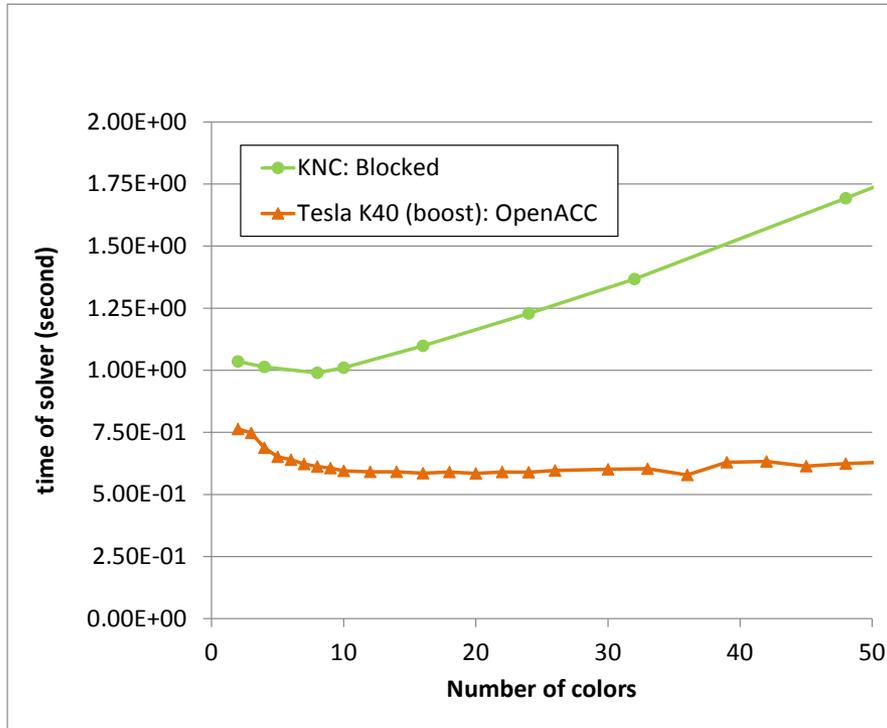
- CUDAによるプログラミング：複雑
- 2,000人を超えるユーザー
 - 非常に優れたユーザーとは言え，CUDAを勉強してもらうのはなかなか大変
- そのような背景もあり，ポストT2K(Oakforest-PACS)については2010年夏の段階で既にIntel Xeon Phiの最新機種を使用することを決めていた

今回GPUを入れる気になった理由

- OpenACC
 - OpenMPと類似したインタフェース: 使いやすいが性能悪かった
⇒ 昨今の性能向上, CUDAとそれほど大きな差がなくなった
 - NVIDIA研究者との共同研究, Atomic演算HWサポート等
 - OpenACC専門家の情報基盤センターへの加入(2016年1月)
- データ科学, 深層学習(Deep Learning)
 - 従来の計算科学, 計算工学分野とは異なった分野の新規ユーザー開拓が急務: 電気代=負担金
 - 東京大学ゲノム医科学研究機構
 - 東京大学病院
 - 医療画像処理への深層学習適用



ICCG Solver



Matrix Assembly in FEM: sec (GFLOPS)

	KNC	NVIDIA Tesla K40
Coloring + OpenMP/OpenACC	0.690 (94.0)	0.675 (96.2)
NO-Color/Atomic + OpenMP/OpenACC	0.576 (112.7)	0.507 (128.1)
Atomic + CUDA	-	0.362 (179.4)

	KNC	NVIDIA Tesla K40
GFLOPS (Peak)	1,011	1,430
GB	8	12
GB/sec. (Peak)	320	288
GB/sec. (STREAM Triad)	159	218

Reedbush (Mini PostT2K改め) (1/2)

- システム構成・運用: SGI
- Reedbush-U (CPU only)
 - Intel Xeon E5-2695v4 (Broadwell-EP, 2.1GHz 18core,) x 2
ソケット (1.210 TF), 256 GiB (153.6GB/sec)
 - InfiniBand **EDR**, Full bisection BW Fat-tree
 - システム全系: 420 ノード, 508.0 TF
- Reedbush-H (with GPU)
 - CPU・メモリ: Reedbush-U と同様
 - **NVIDIA Tesla P100** (Pascal世代 GPU)
 - (4.8-5.3TF, 720GB/sec, 16GiB) x 2 / ノード
 - InfiniBand **FDR x 2ch**, Full bisection BW Fat-tree
 - 120 ノード, 145.2 TF(CPU)+ 1.15~1.27 PF(GPU)=
1.30~1.42 PF

“Reedbush”って何？



Blaise Pascal
(1623-1662)

- L'homme est un roseau pensant.
- Man is a thinking reed.
- 人間は考える葦である

Pensées (Blaise Pascal)



Reedbush (Mini PostT2K改め) (2/2)

- ストレージ/ファイルシステム
 - 並列ファイルシステム (Lustre)
 - 5.04 PB, 145.2 GB/sec
 - 高速ファイルキャッシュシステム: Burst Buffer (DDN IME (Infinite Memory Engine))
 - SSD: 209.5 TB, 450 GB/sec
- 電力, 冷却, 設置面積
 - 空冷, 378 kVA(冷却除く)
 - < 90 m²
- データ解析、ディープラーニング向けソフトウェア・ツールキット
 - OpenCV, Theano, Anaconda, ROOT, TensorFlow, Torch, Caffe, Chainer, GEANT4

計算ノード: **1.795-1.926 PFlops**

Reedbush-U (CPU only) 508.03 TFlops

CPU: Intel Xeon E5-2695 v4 x 2 socket
(Broadwell-EP 2.1 GHz 18 core,
45 MB L3-cache)
Mem: 256GB (DDR4-2400, 153.6 GB/sec)

× 420

SGI Rackable
C2112-4GP3

InfiniBand EDR 4x
100 Gbps /node

Reedbush-H (w/Accelerators)

1287.4-1418.2 TFlops

CPU: Intel Xeon E5-2695 v4 x 2 socket
Mem: 256 GB (DDR4-2400, 153.6 GB/sec)
GPU: NVIDIA Tesla P100 x 2
(Pascal, SXM2, 4.8-5.3 TF,
Mem: 16 GB, 720 GB/sec, PCIe Gen3 x16,
NVLink (for GPU) 20 GB/sec x 2 brick)

× 120

SGI Rackable C1102-PL1

Dual-port InfiniBand FDR 4x
56 Gbps x2 /node

InfiniBand EDR 4x, Full-bisection Fat-tree

145.2 GB/s



Lustre Filesystem
DDN SFA14KE x3

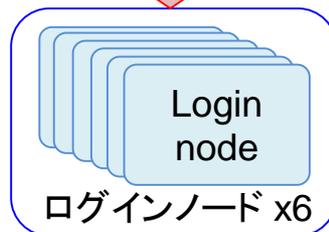
436.2 GB/s



DDN IME14K x6

Mellanox CS7500
634 port +
SB7800/7890 36
port x 14

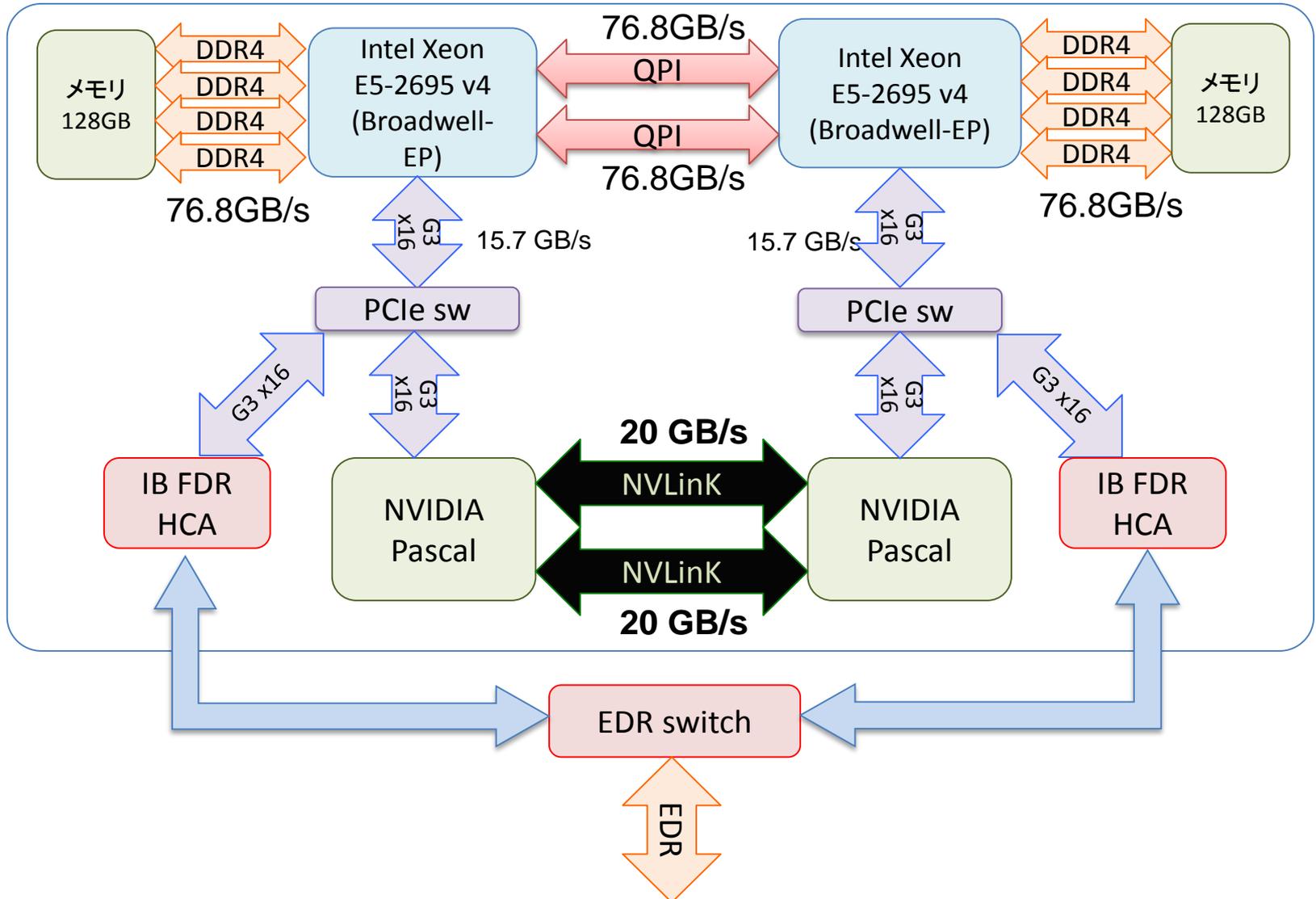
管理サーバ
群



UTnet

ユーザ

Reedbush-Hノードのブロック図



全体構成

項目		Reedbush-U	Reedbush-H	
システム全体 (計算ノード)	総理論演算性能	508.03 TFlops	1297.15～1417.15 TFlops	
	総ノード数	420	120	
	総主記憶容量	105 TByte	30 TByte	
	ネットワークポロジ	Full-bisection BW Fat-Tree		
	並列ファイルシステム	システム名	Lustreファイルシステム	
		サーバ(OSS)	DDN SFA14KE	
		サーバ(OSS)数	3	
		ストレージ容量	5.04 PB	
		帯域幅	145.2 GB/sec	
	高速ファイルキャッシュシステム	サーバ	DDN IME14K	
		サーバ数	6	
		容量	209 TByte	
		帯域幅	436.2 GB/sec	

項目		Reedbush-U	Reedbush-H
マシン名		SGI Rackable C2112-4GP3	SGI Rackable C1100シリーズ (開発中)
CPU	プロセッサ名	Intel Xeon E5-2695v4 (Broadwell-EP)	
	プロセッサ数(コア数)	2 (36)	
	周波数	2.1 GHz (Turbo boost 時最大 3.3 GHz)	
	理論演算性能	1209.6 GFlops	
Memory	容量	256 GB	
	メモリ帯域幅	153.6 GB/sec	
GPU	プロセッサ名	None	NVIDIA Tesla P100 (Pascal)
	コア数(単体)		56 (SM)
	メモリ容量(単体)		16 GB
	メモリ帯域幅(単体)		720 GB/sec
	理論演算性能(単体)		4.8~5.3 TFlops
	搭載数		2
	CPU-GPU間接続		PCI Express Gen3 x16レーン (16 GB/sec)
	GPU間接続		NVLink 2 brick (20 GB/sec x2)
インターコネクト		InfiniBand EDR 4x (100 Gbps)	InfiniBand FDR 4x 2リンク (56 Gbps x2)

ソフトウェア構成

項目	Reedbush-U	Reedbush-H
OS	Red Hat Enterprise Linux 7	
コンパイラ	GNU コンパイラ Intel コンパイラ (Fortran77/90/95/2003/2008、C、C++)	
		PGI コンパイラ (Fortran77/90/95/2003/2008、C、C++、OpenACC 2.0、 CUDA Fortran) NVCC コンパイラ (CUDA C)
メッセージ通信ライブラリ	Intel MPI, SGI MPT, Open MPI, MVAPICH2, Mellanox HPC-X	
		GPUDirect for RDMA: Open MPI, MVAPICH2-GDR
ライブラリ	Intel 社製ライブラリ(MKL): BLAS、LAPACK、ScaLAPACK その他ライブラリ: SuperLU、SuperLU MT、SuperLU DIST、METIS、MT-METIS、ParMETIS、 Scotch、PT-Scotch、PETSc、FFTW、GNU Scientific Library、NetCDF、PnetCDF など	
		cuBLAS、cuSPARSE、cuFFT、MAGMA、OpenCV、ITK、 Theano、Anaconda、ROOT、TensorFlowなど
アプリケーション	OpenFOAM、ABINT-MP PHASE、FrontFlow、FrontISTR、REVOCAP、ppOpen-HPC など	
デバッガ、プロファイラ	Total View, Intel VTune, Trace Analyzer & Collector	

ソフトウェア構成: データ解析向け

- OpenCV
 - コンピューター・ビジョン・ライブラリ
- Theano
 - Python数値計算ライブラリ
- ROOT
 - ビッグデータ向けのライブラリ
- TensorFlow
 - Google開発の機械学習向けライブラリ
- NVIDIA Deep Learning SDK

など...

- 背景
- 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システム (Reedbush) 概要
- **スケジュール**
- 運用・サービス
 - トークン制
 - 教育利用, 若手利用
 - 企業利用
 - トライアルユース
- 試験運転期間のサービス
- その他
- 質疑

スケジュール: Reedbush-U

- 試験運転

- 7月1日(金)10:00～8月31日(水)09:00
- 負担金無料
- 試験運転期間中は、システムの設定変更等のため、予告なく運用の停止、運用仕様の変更を行う場合がありますので、予めご了承ください。

- 正式運用

- 9月1日(木)09:00開始

- 新規利用申込み

- 6月1日開始予定
- 申込み状況によっては、新規利用申込みを打ち切ることがあります。
- 詳しくは、HP(<http://www.cc.u-tokyo.ac.jp/>)でご確認ください。



スケジュール: Reedbush-H(予定)

- 試験運転

- 3月1日(水)10:00～3月30日(木)09:00(予定)
- 負担金無料(無料はReedbush-Hのみ、-Uは有料)
- 試験運転期間中は、システムの設定変更等のため、予告なく運用の停止、運用仕様の変更を行う場合がありますので、予めご了承ください。

- 正式運用

- 4月3日(月)09:00開始

- 新規利用申込み

- 2017年1月上旬(予定)
- 申込み状況によっては、新規利用申込みを打ち切ることがあります。
- 詳しくは、HP(<http://www.cc.u-tokyo.ac.jp/>)でご確認ください。

情報・問い合わせ

- 全般
 - <http://www.cc.u-tokyo.ac.jp/system/reedbush/>
- 試験運転開始のお知らせ
 - http://www.cc.u-tokyo.ac.jp/system/reedbush/reedbush_test.html
- 利用コース
 - http://www.cc.u-tokyo.ac.jp/system/reedbush/reedbush_course.html
- ジョブクラス
 - http://www.cc.u-tokyo.ac.jp/system/reedbush/reedbush_job.html
- 利用申込・利用負担金
 - <http://www.cc.u-tokyo.ac.jp/support/application/>
- FAQ
 - http://www.cc.u-tokyo.ac.jp/support/faq/reedbush_faq.html
- 東京大学情報システム部情報戦略課研究支援チーム
 - 電話(平日09-12, 13-17) 03-5841-2717
 - uketsuke@cc.u-tokyo.ac.jp

- 背景
- 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システム (Reedbush) 概要
- スケジュール
- **運用・サービス**
 - **トークン制: Reedbush-Uの話が中心です**
 - 教育利用, 若手利用
 - 企業利用
 - トライアルユース
- 試験運転期間のサービス
- その他
- 質疑

利用コース

- Reedbushの利用コース
 - パーソナルコース: 研究者個人単位(大学・公共機関)
 - グループコース(一般): 研究グループ単位(大学・公共機関, 企業)
 - グループコース(ノード固定): 研究グループ単位
 - 事前審査あり, 全体の15%程度を上限
- 利用するコース (パーソナル・グループコース), 利用申込したノード数に応じて, 計算ノードの利用可能時間である「トークン」を割当てます。
 - 割り当てられたトークン内であれば (一部のコース、サービスを除き) 利用できるノード数制限はなく、最大利用可能ノード数まで、バッチジョブの実行を可能。
 - FX10(Oakleaf/Oakbridge-FX)の方法を踏襲

トークン(token) : FX10と同じ

参考: http://www.cc.u-tokyo.ac.jp/system/fx10/fx10_token.html

- トークンは、バッチジョブ(インタラクティブノードを利用したバッチジョブを除く)実行ごとに消費。
 - ノード時間積「経過時間 × ノード数 × 消費係数」により消費
 - 消費係数は、申込ノード数までは 1.00
 - 申込ノード数を超えた範囲について 2.00
- トークンを使い果たすとジョブ実行不可: 1日単位モニター
 - トークン不足の場合, ジョブはsubmitできない
 - 計算資源に余裕がある場合にのみ, トークンを追加可能
- トークンは, 利用を許可された有効期間内に全量が利用できることを保証するものではありません。
- 利用を許可された期間のみを有効期間としているため, 次年度への繰り越しや返金等は不可。
- できるだけ長期間(1年単位, 本年度は7ヶ月)の契約が
得, 短い期間でのトークン追加は割高です。

パーソナルコース

- パーソナルコースは研究者が個人単位でお使いいただくためのコースです。
- 次々頁に示すバッチジョブキューを使います。

コース	利用負担金 (年額, 税込) 大学・公共機関 等	利用可能 ノード数	トークン	ディスク量 /lustre
パーソナル コース	200,000円/年 (7ヶ月:145,800 円)	最大 16ノード	17,280 ノード時間 (2ノード×360日相当) 消費係数： 4 ノードまでは 1.00 4 ノード超過は 2.00	1TB

グループコース(一般)

- グループコース(一般)は、研究グループ等で利用するためのコースで、**4ノード・または8ノード単位**で申込みます。
 - 提供できる資源量に限りがあるため、利用申込単位(ノード数)によってはご希望に添えない場合があります。
- 標準で割り当てられる**ディスク容量は4ノードあたり4TB**
 - それ以外にもグループに所属する利用者ごとに**500 GB**のディスク容量が割り当てられます。
- **グループに登録できる利用者数は無制限**
 - 割り当てられたトークンをグループに登録された利用者で共有
- **グループ管理者(利用申込み時に設定)にはグループ内の利用者ごとに割り当てるトークン量を変更できる仕組みを導入。**
- **次頁に示すバッチジョブキューを使用します。**

バッチジョブキュー: Reedbush-U

- パーソナルユーザー, グループユーザー(一般)で共通
- u-interactive-1/4はトークンを消費しない
- 以下は本運用(2016年9月以降)の設定

代表キュー名	キュー名	最大ノード数	実行制限時間 (経過時間)
u-interactive	u-interactive_1	1	30 min
	u-interactive_4	2-4	10 min
u-debug	u-debug	1-24	30 min
u-short	u-short	1-8	4 h
u-regular	u-small	4-16	48 h
	u-medium	17-32	48 h
	u-large	33-64	48 h
	u-x-large	65-128	24 h

グループコース(ノード固定)(1/2)

- グループコース(ノード固定)は, 研究グループ等で利用するためのコースで, **8ノード単位**で申込みます。
 - 申込み分のノードを占有できる
 - **パーソナルコース, グループコース(一般)とは別のバッチジョブキューを使います。**
 - グループコース(一般)と同じ計算機資源を使用することも可能: その分トークンは減る
 - **トークンがなくなった時点で占有ノードも使えなくなります(資源に余裕があればトークン追加購入可能)**
 - 審査制, 全体の15%程度を上限とする
- **設定のカスタマイズが可能となる**
 - **個別のログインノード・ストレージ等柔軟に対応**
- **ディスク容量設定, グループ登録者数, グループ管理者権限はグループコース(一般)に準ずる**

グループコース(ノード固定)(2/2)

利用例: 16ノードで申し込んだ場合

- 16ノード x 360日分=138,240ノード時間のトークン
- 専用に割り当てられたバッチジョブキューを使用
 - グループ内合計で同時に利用できるノード数は最大16
 - 例えば, 8ノードのジョブを同時に2本実行することも可能
 - 使用ノード時間分のトークンが消費されます
- 専用のバッチジョブキューで既に16ノード使用中の場合
 - 同じキューを指定すると、必要なノードが空くまで待たされる
 - 同時に計算を実行したい場合, 前々頁で示したバッチジョブキュー(グループ利用(一般)用)で計算を実行することも可能
 - 使用ノード時間に応じてトークンは消費, ノード固定分に使えるトークンはそれだけ減少
- 16ノード以上の計算をしたい場合
 - 前々頁で示したバッチジョブキュー(グループ利用(一般)用)で計算を実行することも可能
 - 例: 128ノードの計算を4時間実行したい場合
 - u-x-largeジョブキューで実行可能(u-regularキューを指定)
 - 16ノードx4時間x係数1.0 + 112ノードx4時間x係数2.0 = 960ノード時間を消費

グループコース負担金(年)

	大学・公共機関等	企業	トークン	ディスク量 /lustre
一般 申込4ノード	400,000 (7ヶ月: 291,600)	480,000 (7ヶ月: 350,000)	34,560ノード時間 (4ノード×360日相当) 消費係数: 4ノードまでは 1.00 4ノード超過は 2.00	4TB(グループ あたり)/ 500GB(1人あ たり)
一般 申込8ノード以上 (8ノード当り)	700,000 (7ヶ月: 510,400)	840,000 (7ヶ月: 612,500)	69,120ノード時間 (8ノード×360日相当) 消費係数: 申込ノードまでは 1.00 申込ノード超過は 2.00	8TB(グループ あたり)/ 500GB(1人あ たり)
ノード固定 申込8ノード以上 (8ノード当り)	1,000,000 (7ヶ月: 729,100)	1,200,000 (7ヶ月: 875,000)		

追加ディスク, トークン

		大学・公共機関等	企業	備考
追加ディスク		6,480		1TB／年あたり
追加 トークン	パーソナル コース	16,600	-	1,440ノード時間 (2ノード×30日相当)
	グループ コース (一般)	33,300	40,000	2,880ノード時間 (4ノード×30日相当)
	グループ コース(ノード 固定)			

Reedbush-U, Reedbush-H 利用

- Reedbush-U, Reedbush-Hは分離して運用され、1つのジョブで両方のノードの同時使用はできません。
- パーソナルコース, グループコース(一般)の利用者は, Reedbush-U, Reedbush-Hの両方を利用可能です。
 - Reedbush-U, Reedbush-Hではトークン消費率が異なります
 - Reedbush-Hのトークン消費率は未定ですが, Reedbush-Uの約2.5倍程度になる予定です
 - Reedbush-HのホストCPU部のみ使用した場合にもGPUを使ったのと同じトークン消費率が適用される
- グループコース(ノード固定)はReedbush-U, Reedbush-H別々に申し込む必要があります。
 - グループコース(ノード固定)の利用者が, グループコース(一般)のリソースを使うことは可能です(トークン消費率は同じ)
 - ただしトークンは減り, ノード固定分に使えるノード時間は減ります

FX10⇔Reedbush-U トークン互換性(1/2)

- FX10(Oakleaf/Oakbridge-FX), Reedbush-Uはトークンに互換性があり, 原則として一方から他方へ移行可能。
- どちらか一方に負担金を払っていれば, もう一方はトークンを移行するだけで使うことができる。
 - 申請により, 毎月の月末処理時に移行処理を行います。
 - 申請方法については検討中
 - 2ノード時間@FX10=1ノード時間@Reedbush-U となります。
 - 「申込ノード数」もFX10とReedbush-Uでは1:2になります。
 - 鍵登録はそれぞれのシステムについて必要です
- 多様な計算機資源を容易に使いこなすことができるようになります。

FX10⇔Reedbush-U トークン互換性(2/2)

- トークン互換が可能なのは以下の場合のみです。
 - FX10: パーソナル1⇔Reedbush-U: パーソナル
 - FX10: パーソナル2⇔Reedbush-U: パーソナル
 - FX10: グループ⇔Reedbush-U: グループ(一般)
 - 企業利用の場合はトークンの移行はできません
 - HPCI, JHPCNについてもトークンの移行はできません
 - Reedbush-Uのグループユーザー(ノード固定)はトークンをFX10に移すことはできません
 - トライアルユースの場合, トークン移行はできません
- FX10⇔Reedbush-Hについてもトークン互換とする予定です。

バッチジョブ実行の条件

- 利用者は、「トークン量」がある限り、好きなキューにジョブを投入することができる
 - 最大利用ノード数
 - パーソナルコース: 16ノード
 - グループコース(一般): 128ノード
 - 申込ノード数を超過したジョブについては、超過部分について消費係数が高め(=2.00)に設定されます。
- 「u-regular」を指定してジョブ投入すると、ノード数によってu-small, u-medium, u-large, u-x-large に振り分けられる

バッチジョブキュー: Reedbush-U

- パーソナルユーザー, グループユーザー(一般)で共通
- u-interactive-1/4はトークンを消費しない
- 以下は本運用(2016年9月以降)の設定

代表キュー名	キュー名	最大ノード数	実行制限時間 (経過時間)
u-interactive	u-interactive_1	1	30 min
	u-interactive_4	2-4	10 min
u-debug	u-debug	1-24	30 min
u-short	u-short	1-8	4 h
u-regular	u-small	4-16	48 h
	u-medium	17-32	48 h
	u-large	33-64	48 h
	u-x-large	65-128	24 h

- 背景
- 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システム (Reedbush) 概要
- スケジュール
- **運用・サービス: FX10 (Oaleaf/Oakbridge-FX) と Reedbush (RB) の併用へ**
 - トークン制
 - **教育利用, 若手利用**
 - **企業利用**
 - **トライアルユース**
- 試験運転期間のサービス
- その他
- 質疑

教育・人材育成(1/2)

- お試しアカウント付き並列プログラミング講習会
 - 現在Oakleaf-FX⇒Reedbush併用へ
 - 既存利用者に限定せず，企業の技術者・研究者も受講可能
 - ✓ 受講者の3分の2以上は企業から受講：裾野拡大に大きな貢献
 - ✓ PCクラスタコンソーシアム(実用アプリケーション部会)と共催
 - 1～2日間の講習，1週間有効な「お試しアカウント」
 - ✓ MPI基礎，MPI応用(並列有限要素法)，マルチコアプログラミング
 - ✓ ライブラリ利用(センター教員開発のライブラリ普及)，OpenFOAM
 - 2016年度からの新企画
 - ✓ GPUプログラミング基礎，OpenMP+OpenACC応用⇒2016年度はHAPACS(筑波大)使用(筑波大・JCAHPC共催)，2017年度～Reedbush-H
 - ✓ 有限要素法で学ぶ並列プログラミング，MPI上級編
- 学部・大学院・高専の講義での利用(学外含む)
 - 提案書ベース，無料，専用のバッチジョブキュー(FX10:12ノード，15分)，(Reedbush-U:8ノード，10分)
 - 年10件程度(うち2～3は学外から)，センター教員の講義

講習会予定(終了したものも含む)

<http://www.cc.u-tokyo.ac.jp/support/kosyu/>

- 第55回 2016年5月18日
 - ✓ ALPS入門
- 第56回 2016年5月24日～25日
 - ✓ 有限要素法で学ぶ並列プログラミング基礎(新企画)
- 第57回 2016年5月31日
 - ✓ OpenFOAM入門
- 第58回 2016年6月8日
 - ✓ GPUプログラミング入門(新企画)(筑波大CCS, JCAHPC共催)
- 第59回 2016年6月14日～15日
 - ✓ OpenMP/OpenACCによるマルチコア・メニコア並列プログラミング入門(新企画)(筑波大CCS, JCAHPC共催)
- 第60回 2016年6月28日～29日
 - ✓ 並列有限要素法とハイブリッド並列プログラミング
- 第61回 2016年7月7日～8日
 - ✓ MPI基礎:並列プログラミング入門(Reedbush-Uで実施予定)

教育・人材育成(2/2)

- RIKEN AICS Summer/Spring School (2011~)
 - 理研AICS, 神戸大等と共催
 - プログラム策定, 教材作成の大半を当センターで実施
- 若手・女性支援
 - 40歳以下(女性は年齢制限無し): 無料でFX10, RBを使用可能
 - 公募型(年2回, 各半年間, 連続して2回応募可能⇒1年間無料)
 - ✓ 科研費, JHPCN公募型研究へ進展期待
 - 学生を対象とした「インターンシップ」制度, グループ制度
- 計算科学アライアンス(2014~)
 - 学際的研究の拠点, 全学的なHPC教育プログラムの策定
 - H28概算要求(正式に始動), 関連部局
 - Double Degreeを見据えた横断型・学際型プログラム
 - 情報理工学系・理学系・情基セの協力が元になっている

背景

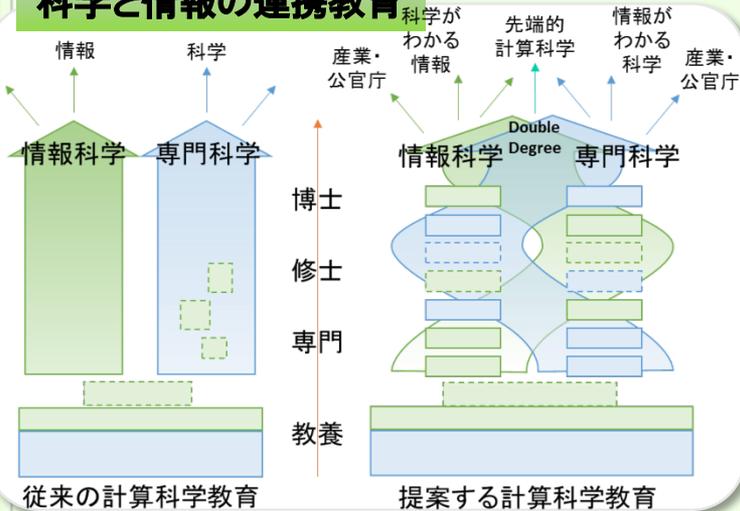
- 多数の社会活動・学術分野で計算科学・情報の知見の必要度が増大しかし、計算科学、自然科学、情報科学が専門分化して俯瞰力が低下
- 計算科学ソフトウェアが高度化・大規模化する中、並列処理など計算効率の原理が変化
- 海外ソフトの利用が多く、中国・ブラジルなど新興国も追い上げ、ポスト京アプリ開発が始動するも、国産ソフト開発・競争力強化が必要
- 企業もシミュレーションを重要視しているが、企業内の専門家はわずか

目的

- 自然・社会科学と情報科学を俯瞰した標準教育体系を整備
- 大規模ソフトウェア、並列計算などの手法を教える
- 国産シミュレーションソフトの開発力・競争力を強化
- 企業のシミュレーション能力・活用を向上する人材を育成

提案：計算科学アライアンス

科学と情報の連携教育



自然・社会科学と情報科学が連携して 計算科学の体系的教育・人材育成を行う

- 柔軟なコース設計：4学期・国際化対応
- 学部から大学院まで、習熟度別・英才教育
- キャリアパス整備：企業へのアピール

英才教育

- 自然・社会科学とソフト・ハード・計算を横断・俯瞰して理解・活用できる人材育成
- ダブルディグリー制度を目指す
 - 例：物理博士・情報修士
 - 入試や学位のありかた検討
- スーパーコンピュータ、ソフトウェア工学などの新しい情報技術を教育

基礎教育

- 自然・モデル・計算の基礎教育
- バイリンガル教材 → 国際的に通用
- 社会人教育対応コース、リテラシー教育

学術的必要性

- ポスト京の計算科学で突出した成果を挙げる
- 国際的な先導的立場を維持、発展させる
- あらゆる学問分野で高性能計算を活用

社会的必要性

- 日本の産業と科学を支える人物の輩出
- 我が国の国際社会における優位性
- 計算科学の力を活用する産学官リーダー

期待される成果

- 自然・社会科学と情報科学を俯瞰した標準体系カリキュラムの整備
- 世界を先導する先端的成果と国際連携拠点の形成
- 計算科学で世界的成果を挙げ日本の科学・産業を先導する人材の育成
- ハード・ソフト・自然科学を横断して先導できる人材の育成
- 本学のあらゆる専門学問領域において計算機活用による高度化を促進
- 計算科学を理解し、強みを活かす企業トップ、学術リーダー、官僚等の輩出
- 企業等においてシミュレーション等を駆使して競争力を発揮する人材の輩出

国際連携先端研究

- 海外トップ研究者を招聘
- 国際共同研究を探る WS
- 国際的に開かれた計算科学サマースクール、オンライン講座の提供
- 学生の留学・短期派遣支援
- 優秀な留学生・女性獲得活動
- 外国人・女性の教員・研究員採用

組織

大規模部局連携で
本学全体を底上げ

理学系研究科
工学系研究科
情報理工学系研究科
情報基盤センター
新領域創成科学研究科
数理科学研究科

外部連携

共同利用共同研究拠点
(7大学、筑波大、東工大)

生産技術研究所
物性研究所
地震研究所
大気海洋研究所
IPMU

理化学研究所 AICS
官庁・自治体・企業等
海外大学・研究所等

部局資源の活用・強みを集中
・再配分されたポストを活用
・基盤センターのスパコン利用

資源

本学・我が国の競争力強化

企業利用(1/3)

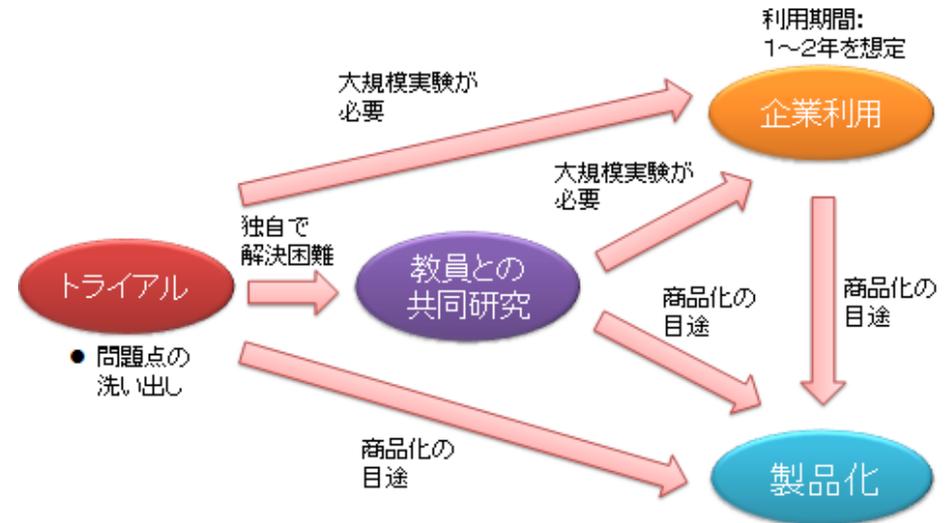
<http://www.cc.u-tokyo.ac.jp/service/company/>

- 2008年度より開始
 - 大規模並列計算普及, 社会貢献
 - ビジネスへの萌芽的段階での支援, データセンターと競合しない
 - 成果は原則公開, 全資源の10%以下
 - アカデミック利用者と異なる審査基準(年2回募集), 負担金体系
 - ✓ HPCI企業利用枠へは計算資源を拠出していない: 2017年度以降未定
- 様々な利用体系
 - 通常グループ利用(いわゆる「企業利用」)
 - ✓ 毎年3-4グループ, 基礎的な研究が多い
 - トライアルユース(グループ(有償, 無償), 個人)
 - ✓ お試しアカウント付き講習会を受講するとパーソナルトライアルユース可能
 - 大学等との共同研究(アカデミック料金で利用可能): 2-3件
 - オープンソース, 自作コードに限定(ISVアプリ無し): 基礎的研究

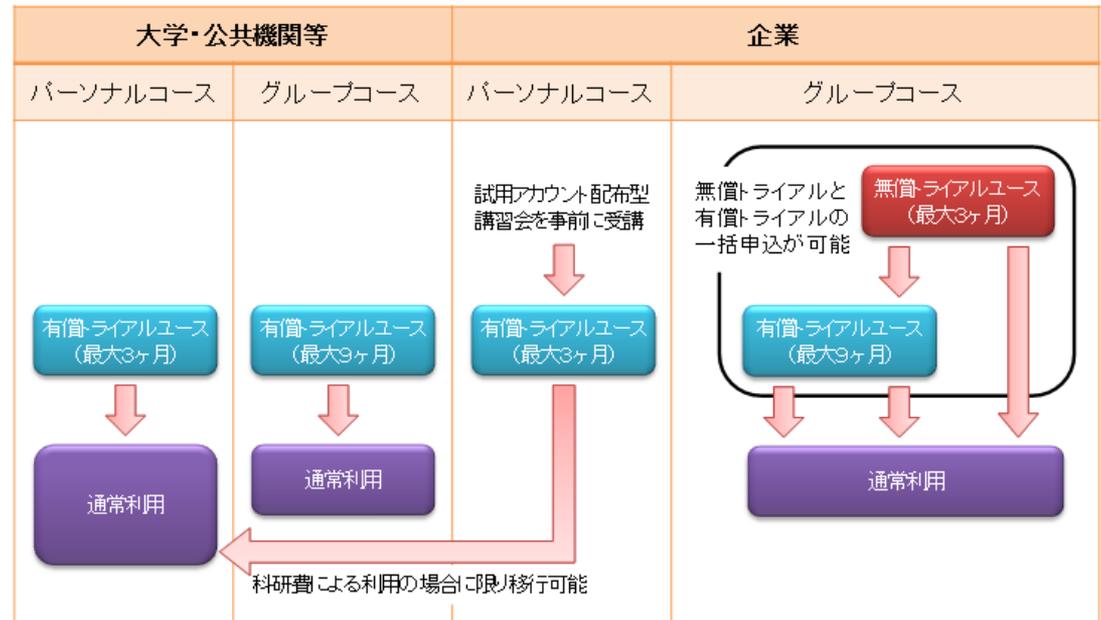
企業利用(2/3)

<http://www.cc.u-tokyo.ac.jp/service/company/>

様々な利用形態



トライアルユース



企業利用(3/3)

<http://www.cc.u-tokyo.ac.jp/service/company/>

- 通常有償利用, トライアルユース(グループ)は審査あり
 - 年2回募集(次回は第二期募集)
 - 次回予定: 8月19日締切、9月上～中旬審査、10月1日利用開始
- 現在はFX10(Oakleaf/Oakbridge-FX)を企業利用に拠出, 次回からReedbush-U申込可能(トークン互換無し)
- 問題点: 企業からの要求
 - セキュリティ: 企業だけでなく外部データを扱う研究では重要
 - ✓ 個別: ログインノード, ストレージ
 - ノード占有: Reedbushでは可能となる
 - ISV(NASTRAN等)
- 企業利用説明会(7月1日午後@浅野キャンパス)(予定)

大規模HPCチャレンジ

Oakleaf-FXで実施中, Reedbushでは未定:ノード固定

- <http://www.cc.u-tokyo.ac.jp/service/4800hpc/>
- 月1回1日(24時間), 4,800ノード(全計算ノード)を1グループで占有して実行できる, 公募制, 無料。
- FX10ユーザー以外も応募可能。
- 成果公開を義務づける
 - センター広報誌への寄稿
 - センター主催各種催しでの発表, 各種外部発表への情報提供
 - 速報結果の査読付国際会議への投稿等による迅速, 国際的な成果公開が望ましい。
- 企業からの申し込みも受け付ける(成果公開を義務づけ)
- 自作プログラム, オープンソースプログラム利用に限定⁶⁰

- 背景
- 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システム (Reedbush) 概要
- スケジュール
- 運用・サービス
 - トークン制
 - 教育利用, 若手利用
 - 企業利用
 - トライアルユース
- **試験運転期間のサービス**
- その他
- 質疑

試験運転期間中のサービス・お願い

- 基本的に9月以降利用するコース, ノード数を申し込んでいただくようお願いいたします。トークンの途中追加は可能。
- 試験運転期間中の残余トークンは本運用に持ち越せない
- 試験運転期間中は「トライアルユース」は設けない
- 試験運転期間中の「ノード固定」
 - システムの変更を伴うような利用は9月以降
- ジョブの実行時間は本運用時より短くする
 - u-debug, u-tutorial, u-lectureを除き、2分の1から4分の1程度

試験運転期間中の バッチジョブ実行制限時間の設定 Reedbush-U

- 利用状況により途中で変更する可能性もあります

代表キュー名	キュー名	最大ノード数	実行制限時間 (経過時間)
u-interactive	u-interactive_1	1	15 min
	u-interactive_4	2-4	5 min
u-debug	u-debug	1-24	30 min
u-tutorial, u-lecture (教育用)	u-tutorial, u-lecture	1-8	10 min
u-short	u-short	1-8	2 h
u-regular	u-small	4-16	12 h
	u-medium	17-32	12 h
	u-large	33-64	12 h
	u-x-large	65-128	6 h

- 背景
- 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システム (Reedbush) 概要
- スケジュール
- 運用・サービス
 - トークン制
 - 教育利用, 若手利用
 - 企業利用
 - トライアルユース
- 試験運転期間のサービス
- **その他**
- 質疑

まとめ

- 東大情報基盤センター概要
- Reedbushシステムの概要, サービスの概要
- 主としてReedbush-Uについて説明
 - Reedbush-H利用説明会: 2017年1月~2月(予定)
 - Oakforest-PACS利用説明会: 2016年10月~12月(予定)
- 6月~7月にかけて関東一円の各大学でも説明会開催の予定: ご希望があれば出張説明会いたしますので, お問い合わせください
 - 東京大学情報システム部情報戦略課研究支援チーム
 - 電話(平日9-12, 13-17時) 03-5841-2717
 - uketsuke(at)cc.u-tokyo.ac.jp
- 高速ファイルキャッシュシステムの割り当て方は, 試行を経て決定する
 - 実運用してみないとわからないこともある
 - 利用ユーザに対し、別途利用募集(審査あり)する予定

東大情報基盤センターのスパコン

- 2系列のスパコンシステム
 - 研究開発型: T2K⇒ポストT2K
 - 市場調達型: SR11K⇒SR16K・FX10⇒？(FY.2018)
- 何らかの形での「研究開発型」が今後のスタイル
 - CPUはコモディティである可能性はあるが
 - Application Driven, Co-Designを前提として利用者へのサービスを考えていかなければならない
 - センターとしての独自性, 研究の方向性

東大情報基盤センターのスパコンの将来

- 電力が拘束条件
- GPU
 - Reedbushは記念碑的な転回点
 - Post FX10ではPower 9+Voltaも選択肢の一つに
- 更なる新分野(新規顧客)の開拓
 - 負担金＝電気代
 - Entertainment: 中国ではスパコン利用重要分野に入っている
 - どのような能力が必要か？

東大センターのスパコン

2基の大型システム, 6年サイクル

FY

08 09 10 11 12 13 14 15 16 17 18 19 20 21 22

Hitachi SR11K/J2
IBM Power-5+
18.8TFLOPS, 16.4TB

Yayoi: Hitachi SR16000/M1
IBM Power-7
54.9 TFLOPS, 11.2 TB

Hitachi HA8000 (T2K)
AMD Opteron
140TFLOPS, 31.3TB

Oakforest-PACS
Fujitsu, Intel KNL
25PFLOPS, 919.3TB

Post T2K
JCAHPC:
筑波大・
東大

Oakleaf-FX: Fujitsu PRIMEHPC
FX10, SPARC64 IXfx
1.13 PFLOPS, 150 TB

Post FX10
50+ PFLOPS (?)

Oakbridge-FX
136.2 TFLOPS, 18.4 TB

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

Reedbush, SGI
Broadwell + Pascal
1.80-1.93 PFLOPS

Peta

K

K computer

Post-K



Reedbush-H利用アンケート

- GPUの利用希望について、アンケート調査にご協力をお願いいたします。
 - ジョブキューや利用コースの設定の際に参考にさせていただきます。
 - お知り合いでGPUを使っている方がいたら、Reedbushシステムをご紹介ください。是非アンケートにもご協力いただけると大変ありがたいです。
- アンケートURL
 - <https://jp.surveymonkey.com/r/GLFPF3J>
 - SurveyMonkey社のアンケートシステムを利用させていただきました。