BDEC Japan update for Open High Performance Computing and Big Data / Artificial Intelligence Infrastructure

Satoshi Matsuoka

Professor, GSIC, Tokyo Institute of Technology /

Director, AIST-Tokyo Tech. Big Data Open Innovation Lab /

Fellow, Artificial Intelligence Research Center, AIST, Japan /

Vis. Researcher, Advanced Institute for Computational Science, Riken

BDEC 2017

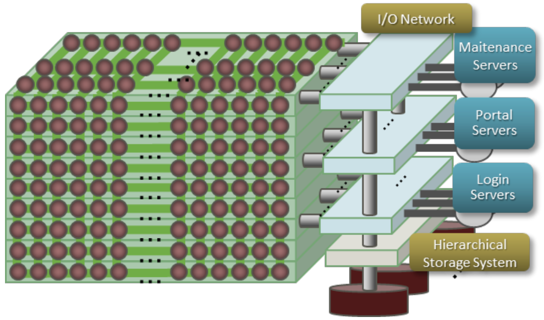
UPDATE:
Post K development

Yutaka Ishikawa
RIKEN AICS

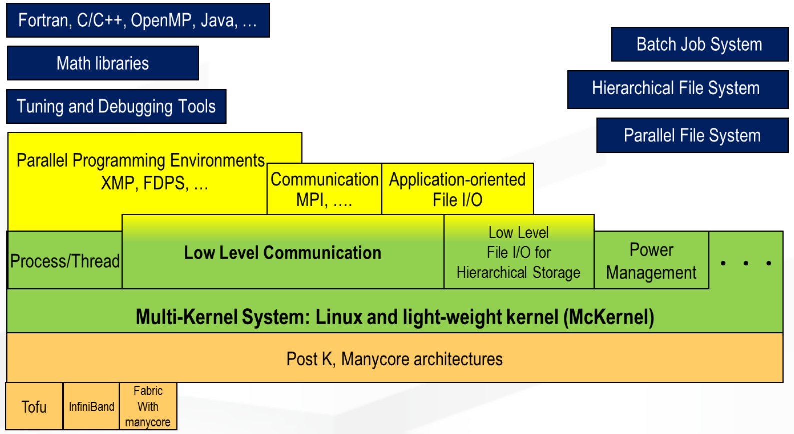
An Overview of Post K

- **CPU architecture**
  - ARMv8-A + SVE + Fujitsu's extension

    FP64/FP32/FP16



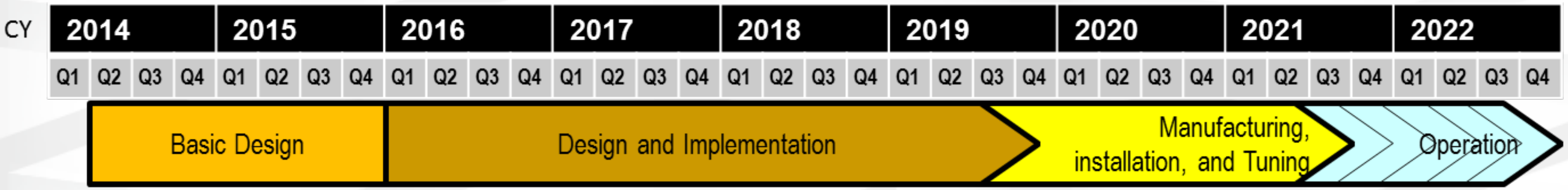
- **Completion of Functional design of system software and start of implementation**



- McKernel is a light-weight kernel with Linux API.
  - New features, such as for manycore and deep memory hierarchy, can be implemented without modification of Linux
  - It runs on Intel Xeon and Xeon phi, and Fujitsu FX100 (SPARC)

- McKernel is running on the Oakforest-PACS supercomputer, 25 PF in peak, at JCAHPC organized by U. of Tsukuba and U. of Tokyo

- **Schedule**



Collaborations

- **DOE-MEXT**
  - Optimized Memory Management, Efficient MPI for exascale, Dynamic Execution Runtime, Storage Architectures, Metadata and active storage, Storage as a Service, Parallel I/O Libraries, MiniApps for Exascale CoDesign, Performance Models for Proxy Apps, OpenMP/XMP Runtime, Programming Models for Heterogeneity, LLVM for vectorization, Power Monitoring and Control, Power Steering, Resilience API, Shared Fault Data, etc.
- **CEA, France**
  - Programming Language
  - Runtime Environment
  - Energy-aware batch job scheduler
  - Large DFT calculations and QM/MM
  - Application of High Performance Computing to Earthquake Related Issues of Nuclear Power Plant Facilities
  - KPIs (Key Performance Indicators)
- **RIKEN AIP (Center for Advanced Intelligence Project)**
  - Massively parallel and distributed search, Machine Learning, etc.

Japanese Open Supercomputing Sites Aug. 2017 (pink=HPCI Sites)

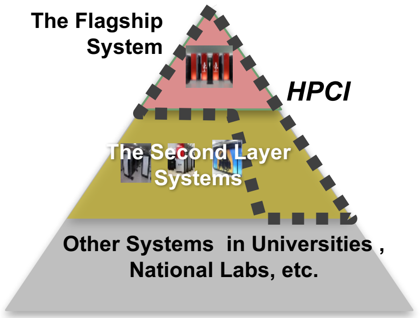
Peak Rank	Institution	System	Rpeak	Nov. 2016 Top500
1	U-Tokyo/Tsukuba U JCAHP	Oakforest-PACS - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path	24.9	6
2	Tokyo Institute of Technology GSIC	TSUBAME 3.0 HPE/SGI ICE-XA custom NVIDIA Pascal P100 + Intel Xeon, Intel OmniPath	12.1	NA
3	Riken AICS	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	11.3	7
4	Tokyo Institute of Technology GSIC	TSUBAME 2.5 - Cluster Platform SL390s G7, Xeon X5670 6C 2.93GHz, Infiniband QDR, NVIDIA K20x NEC/HPE	5.71	40
5	Kyoto University	Camphor 2 – Cray XC40 Intel Xeon Phi 68C 1.4Ghz	5.48	33
6	Japan Aerospace eXploration Agency	SORA-MA - Fujitsu PRIMEHPC FX100, SPARC64 XIfx 32C 1.98GHz, Tofu interconnect 2	3.48	30
7	Information Tech. Center, Nagoya U	Fujitsu PRIMEHPC FX100, SPARC64 XIfx 32C 2.2GHz, Tofu interconnect 2	3.24	35
8	National Inst. for Fusion Science(NIFS)	Plasma Simulator - Fujitsu PRIMEHPC FX100, SPARC64 XIfx 32C 1.98GHz, Tofu interconnect 2	2.62	48
9	Japan Atomic Energy Agency (JAEA)	SGI ICE X, Xeon E5-2680v3 12C 2.5GHz, Infiniband FDR	2.41	54
10	U-Tokyo- Inst. for Solid State Physics	**Sekirei** - SGI ICE XA, Xeon E5-2680v3 12C 2.5GHz, Infiniband FDR HPE/SGI	1.52	86

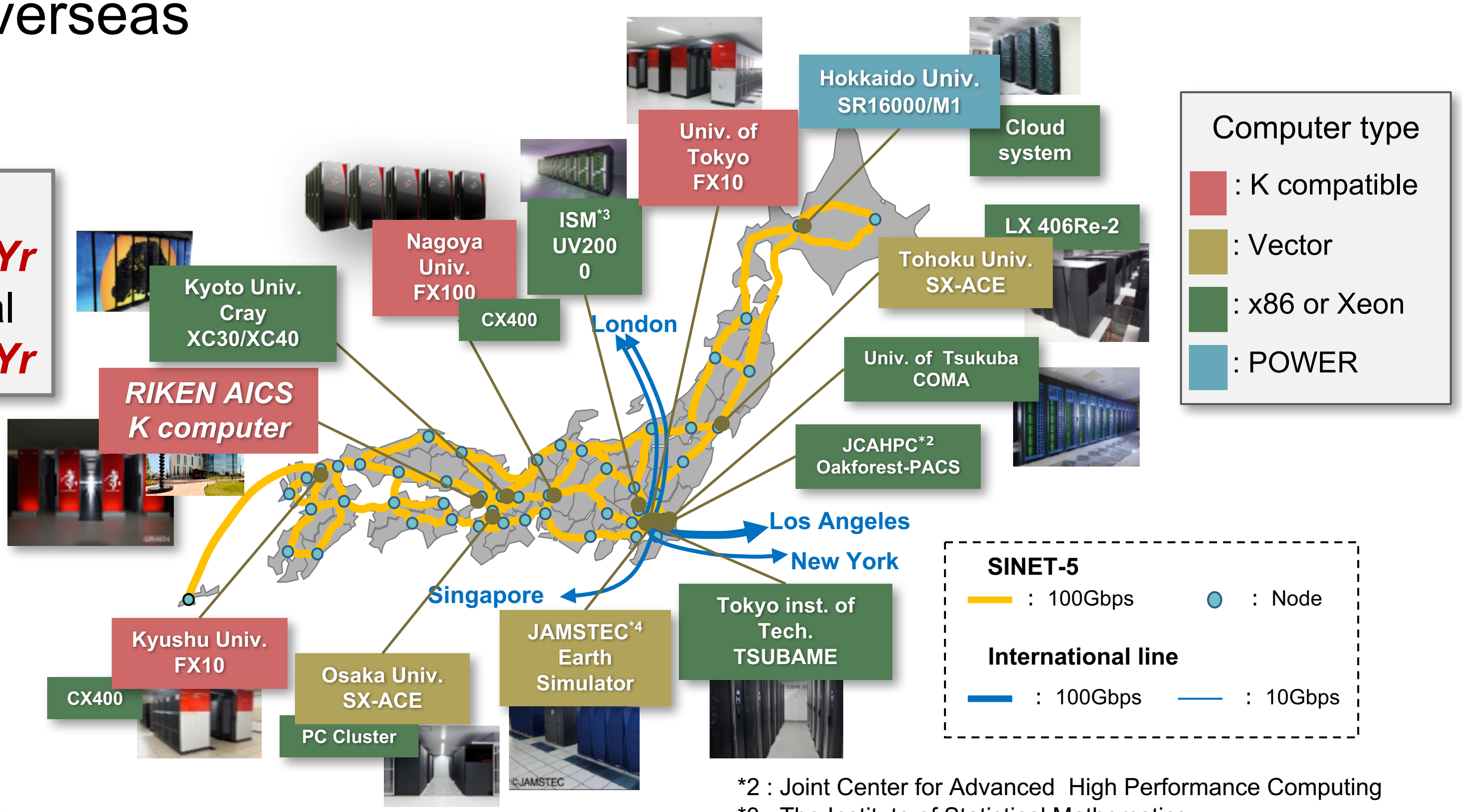
HPCI : High Performance Computing Infrastructure

- Established as Japanese integrated high performance computing infrastructure in 2011
- Variety of computer systems are connected via high speed academic backbone network and provided as ***HPCI*** resources to users in Japan and overseas

FY2017 Allocated computing resources

K computer
~4 Pflops* x *Yr
Other systems in total
~5.6 Pflops* x *Yr





*2 : Joint Center for Advanced High Performance Computing
*3 : The Institute of Statistical Mathematics
*4 : Japan Agency for Marine-Earth Science and Technology

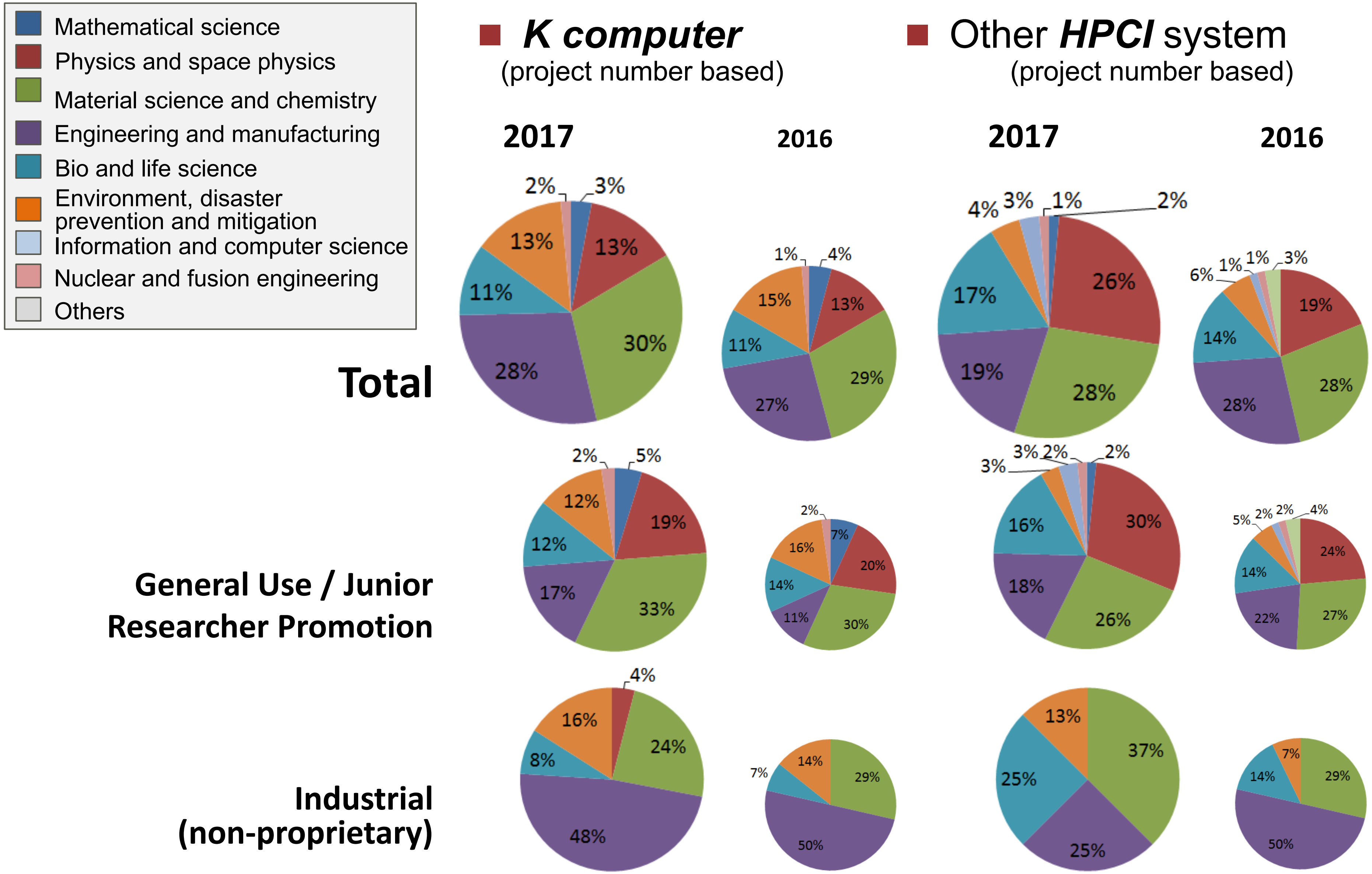
HPCI projects call results for FY 2017

- Number of submitted & awarded proposals for FY 2017 regular call projects[*1]

		Submitted[*3]	Awarded[*3]	Ratio[*3]
K computer[*2]	General Use	51(53)	31(31)	61(58)%
	Junior Researcher Promotion	16(21)	11(13)	69(62)%
	Industrial (non-proprietary)	29(30)	25(28)	86(93)%
	Total	96(104)	67(72)	70(69)%
Other ***HPCI*** system[*4]	General Use	141(128)	64(59)	45(46)%
	Industrial (non-proprietary)	14(11)	5(10)	36(91)%
	Total	155(139)	69(69)	45(50)%

*1 : Trial call projects are not included.
*2 : Results of "Term A" projects. "Term B" projects call will start from April.
*3 : Numbers in parentheses indicate those for FY 2016
*4 : Includes "concurrent use with ***K computer***"

Research application areas of awarded projects



U-Tokyo/Tsukuba-U JCAHPC "Oakforest-PACS"
24.9 Petaflops KNL/OmniPath

Chassis with 8 nodes,
2U size

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

FUJITSU

Specification of Oakforest-PACS system

Total peak performance			25 PFLOPS
Total number of compute nodes			8,208
Compute node	Product		Fujitsu Next-generation PRIMERGY server for HPC (under development)
	Processor		Next-generation of Intel® Xeon Phi™ （Code name: Knights Landing），>60 cores
	Memory	High BW	16 GB, > 400 GB/sec (MCDRAM, effective rate)
		Low BW	96 GB, 115.2 GB/sec (DDR4-2400 x 6ch, peak rate)
Inter-connect	Product		Intel® Omni-Path Architecture
	Link speed		100 Gbps
	Topology		Fat-tree with (completely) full-bisection bandwidth
Login node	Product		Fujitsu PRIMERGY RX2530 M2 server
	# of servers		20
	Processor		Intel Xeon E5-2690v4 (2.6 GHz 14 core x 2 socket)
	Memory		256 GB, 153 GB/sec (DDR4-2400 x 4ch x 2 socket)

Specification of Oakforest-PACS system (I/O)

Parallel File System	Type		Lustre File System
	Total Capacity		26.2 PB
	Meta data	Product	DataDirect Networks MDS server + SFA7700X
		# of MDS	4 servers x 3 set
		MDT	7.7 TB (SAS SSD) x 3 set
	Object storage	Product	DataDirect Networks SFA14KE
		# of OSS (Nodes)	10 (20)
		Aggregate BW	500 GB/sec
Fast 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
	# of servers (Nodes)		25 (50)
	Aggregate BW		1,560 GB/sec

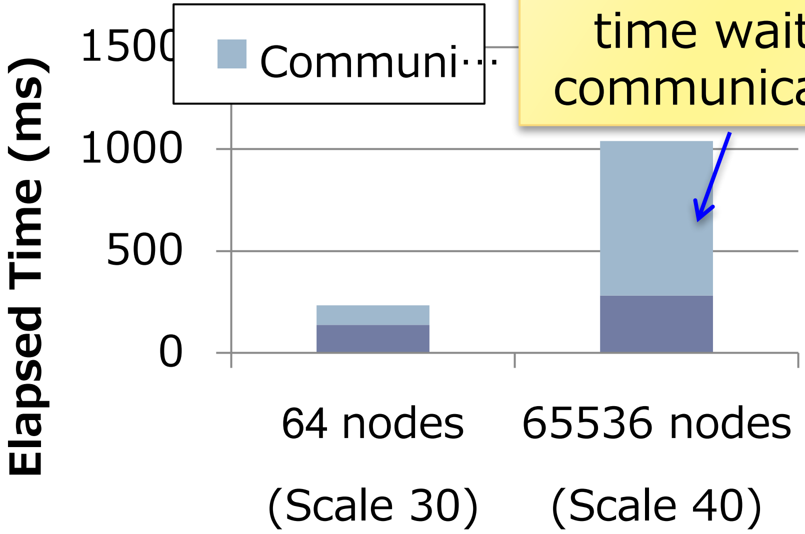
K computer “Still the best” for Bandwidth (Data-centric) workloads (It’s the Bandwidth!)

FUJITSU

	2011	2012	2013	2014	2015	2016
1.TOP500 List		2	4	4	4	7
2. Gordon Bell Prize						Finalist
3. HPC Challenge Awards (HPC、Random Access、STREAM、FFT)						
4. Graph500		4	2			

The Graph500 – 2015~2016 – 4 Consecutive world #1

K Computer #1 Tokyo Tech[EBD CREST] Univ. Kyushu [Fujisawa Graph CREST], Riken AICS, Fujitsu



73% total exec time wait in communication

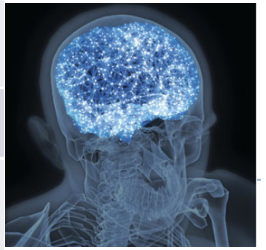
88,000 nodes,
660,000 CPU Cores
1.3 Petabyte mem
20GB/s Tofu NW

K computer

Effective x13 performance c.f. Linpack

List	Rank	GTEPS	Implementat…
November 2013	4	5524.12	Top-down o…
June 2014	1	17977.05	Efficient hybrid
November 2014	2		Efficient hybrid
June, Nov 2015 June Nov 2016	**1**	**38621.4**	**Hybrid + Node Compression**

*Problem size is weak scaling
"Brain-class" graph



LLNL-IBM Sequoia
1.6 million CPUs
1.6 Petabyte mem

IBM

≫

TaihuLight
10 million CPUs
1.3 Petabyte mem

Two Big Data CREST Programs (2013-2020) ~$60 mil

Advanced Core Technologies for Big Data Integration

Research Supervisor: Masaru Kitsuregawa
Director General, National Institute of Informatics

Advanced Application Technologies to Boost Big Data Utilization for Multiple-Field Scientific Discovery and Social Problem Solving

Research Supervisor: Yuzuru Tanaka
Professor, Graduate School of Information Science and Technology, Hokkaido University

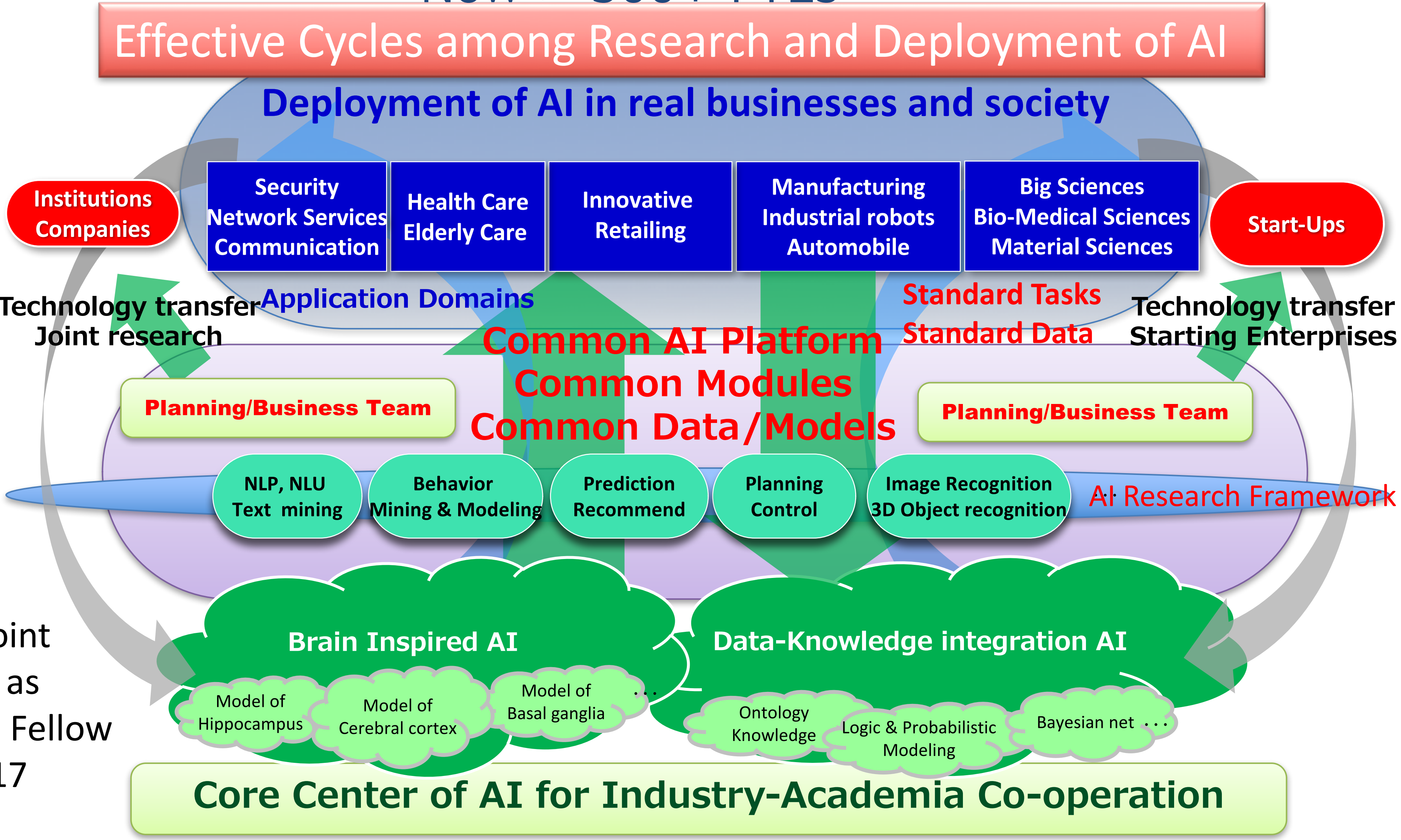
Tremendous Recent Rise in Interest by the Japanese Government on Big Data, DL, AI, and IoT

- Three national centers on Big Data and AI launched by three competing Ministries for FY 2016 (Apr 2015-)
  - METI – AIRC (Artificial Intelligence Research Center): AIST (AIST internal budget + > $200 million FY 2017), April 2015
    - Broad AI/BD/IoT, industry focus
  - MEXT – AIP (Artificial Intelligence Platform): Riken and other institutions ($~50 mil), April 2016
    - A separate Post-K related AI funding as well.
    - Narrowly focused on DNN
  - MOST – Universal Communication Lab: NICT ($50~55 mil)
    - Brain –related AI
  - $1 billion commitment on inter-ministry AI research over 10 years

Vice Minsiter Tsuchiya@MEXT Annoucing AIP estabishment

AI Research Center (AIRC), AIST (under METI)

Now > 300+ FTEs



Matsuoka : Joint appointment as “Designated” Fellow since July 2017

Two AI CREST Programs (under AIP, MEXT) (2016-2023) ~$40 mil x 2

Intelligent Information Processing Systems Creating Co-Experience Knowledge and Wisdom with Human-Machine Harmonious Collaboration

Research Supervisor: Norihiro Hagita (Board Director, Director, Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International)

Development and Integration of Artificial Intelligence Technologies for Innovation Acceleration

Research Supervisor: Minoru Etoh (Senior Vice President, General Manager of Innovation Management Department, NTT DOCOMO, INC.)

Estimated Compute Resource Requirements for Deep Learning [Source: Preferred Network Japan Inc.]

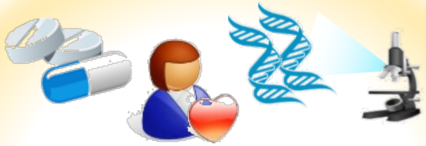
To complete the learning phase in one day

P:Peta
E:Exa
F:Flops

Image/Video Recognition

10P（Image）～ 10E（Video） Flops
学習データ：1億枚の画像 10000クラス分類
数千ノードで6ヶ月 [Google 2015]

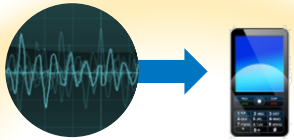
Bio / Healthcare



100P ～ 1E Flops
一人あたりゲノム解析で約10M個のSNPs
100万人で100PFlops、1億人で1EFlops

It's the FLOPS too! (in reduced precision)

Image Recognition



10P～ Flops
1万人の5000時間分の音声データ
人工的に生成された10万時間の
音声データを基に学習 [Baidu 2015]

Auto Driving

1E～100E Flops
自動運転車１台あたり1日 1TB
10台～1000台, 100日分の走行データの学習

Robots / Drones

1E～100E Flops
1台あたり年間1TB
100万台～1億台から得られた
データで学習する場合

機械学習、深層学習は学習データが大きいほど高精度になる
現在は人が生み出したデータが対象だが、今後は機械が生み出すデータが対象となる

各種推定値は1GBの学習データに対して1日で学習するためには
1TFlops必要だとして計算

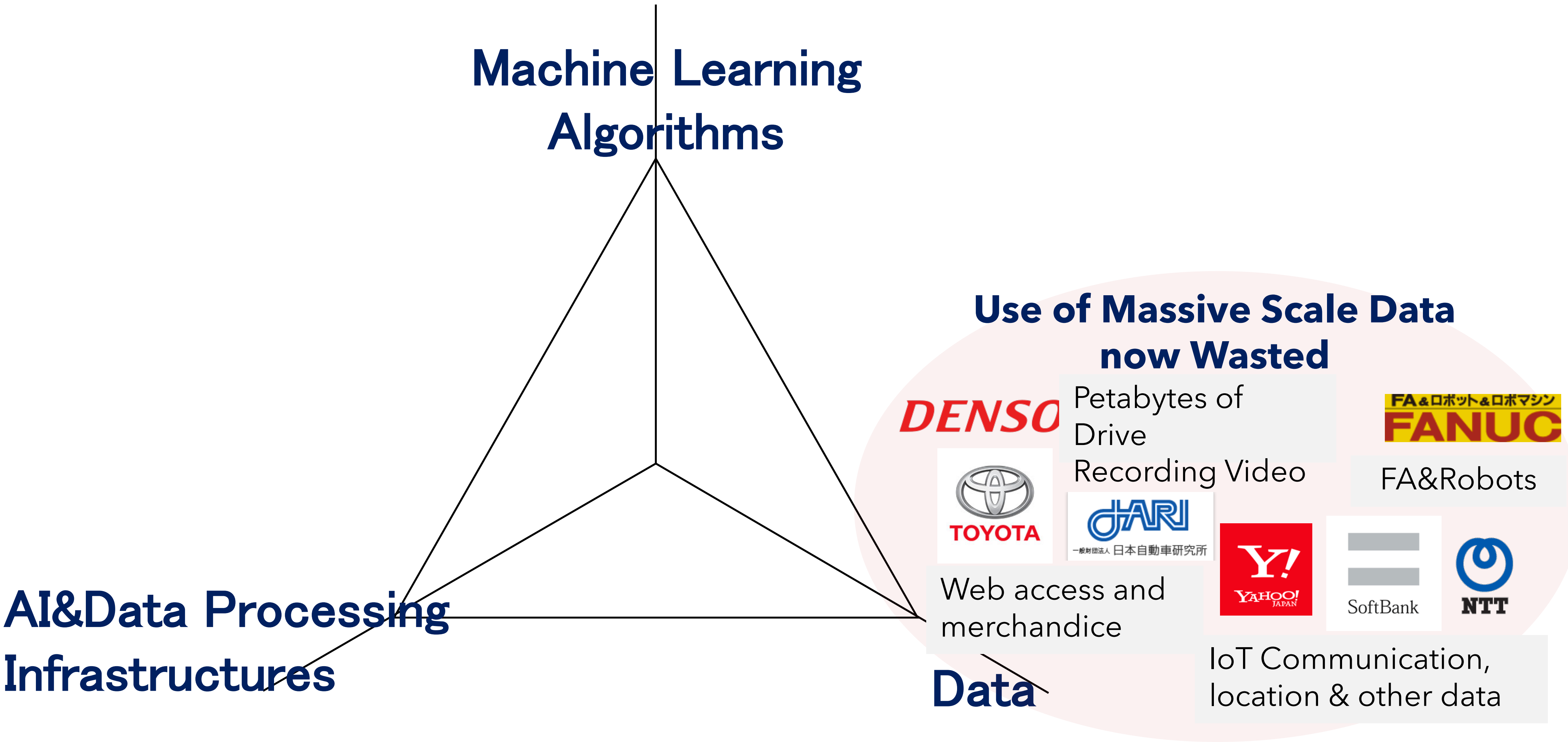
So both are important in the infrastructure

10PF 100PF 1EF 10EF 100EF

2015 *2020* *2025* *2030*

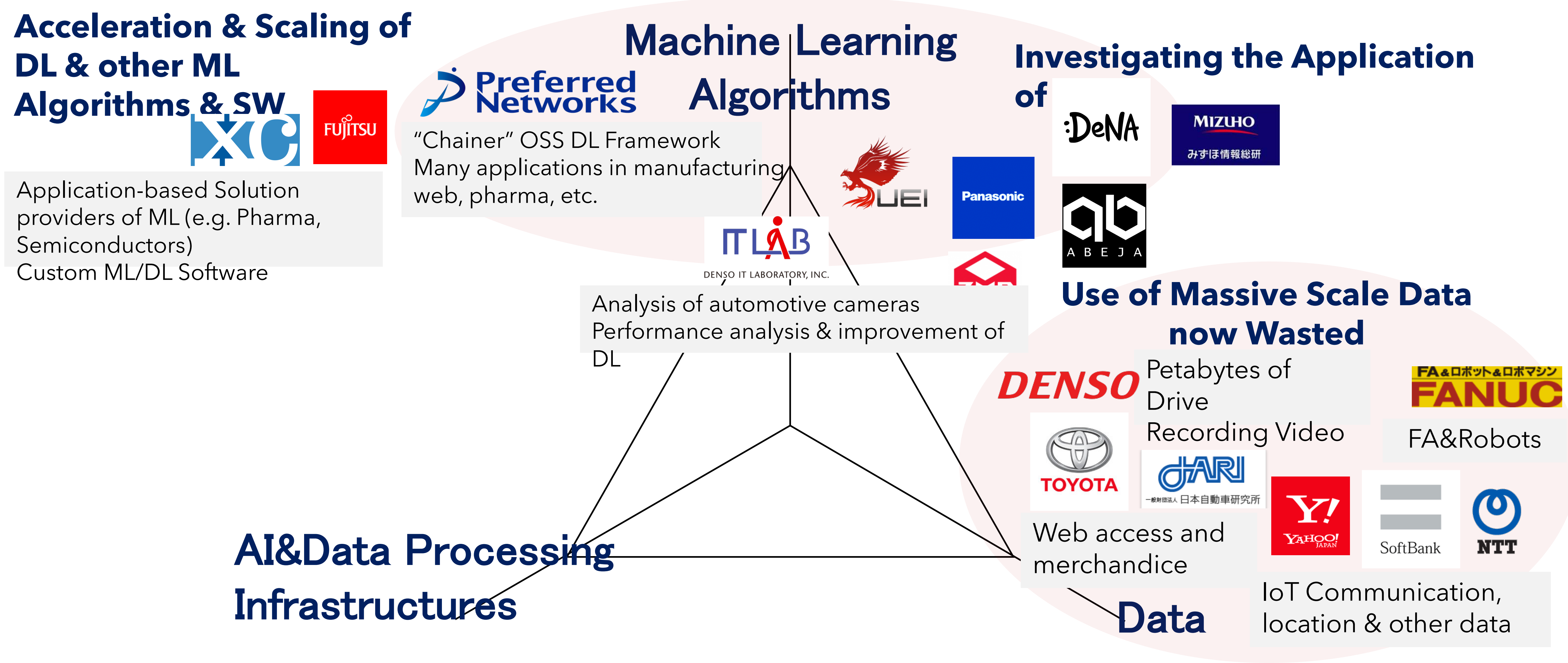
The current status of AI & Big Data in Japan

We need the triage of algorithms/infrastructure/data but we lack the infrastructure dedicated to AI & Big Data (c.f. Google)



The current status of AI & Big Data in Japan

We need the triage of algorithms/infrastructure/data but we lack the infrastructure dedicated to AI & Big Data (c.f. Google)



The current status of AI & Big Data in Japan

We need the triage of algorithms/infrastructure/data but we lack the infrastructure dedicated to AI & Big Data (c.f. Google)

深層学習処理の高度化・
高速化を模索

Application-based Solution providers of ML (e.g. Pharma, Semiconductors)
Custom ML/DL Software

Machine Learning Algorithms

Preferred Networks

"Chainer" OSS DL Framework
Many applications in manufacturing, web, pharma, etc.

Investigating the Application of

DeNA

MIZUHO みずほ情報総研

UEI

Panasonic

ABEJA

IT LAB DENSO IT LABORATORY, INC.

Analy 車載カメラ映像解析
Perfo 深層学習高性能化高速化 ent of
に関する基礎研究

Massive Rise in Computing Requirements

amazon web services

Windows Azure

IBM Cloud

SAKURA Internet

Insufficient to Counter the Giants (Google, Microsoft, Baidu etc.) in their own game

AI&Data Infrastructures

Massive "Big" Data in Training

Use of Massive Scale Data now Wasted

DENSO

Petabytes of Drive Recording Video

FA&ロボット&ロボマシン FANUC

FA&Robots

TOYOTA

JARI 一般財団法人 日本自動車研究所

Web access and merchandice

YAHOO! JAPAN

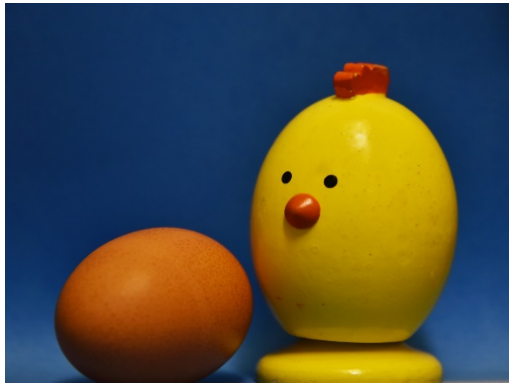
SoftBank

NTT

IoT Communication, location & other data

Data

The “Chicken or Egg Problem” of AI-HPC Infrastructures



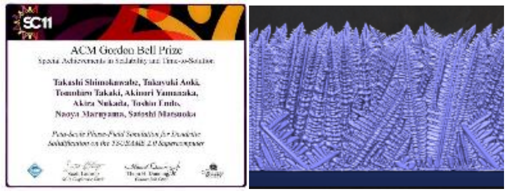
- “On Premise” machines in clients => “Can’t invest in big in AI machines unless we forecast good ROI. We don’t have the experience in running on big machines.”
- Public Clouds other than the giants => “Can’t invest big in AI machines unless we forecast good ROI. We are cutthroat.”
- Large scale supercomputer centers => “Can’t invest big in AI machines unless we forecast good ROI. Can’t sacrifice our existing clients and our machines are full”
- Thus the giants dominate, AI technologies, big data, and people stay behind the corporate firewalls…

2017 Q2 TSUBAME3.0 Leading Machine Towards Exa & Big Data

1. “Everybody’s Supercomputer” - High Performance (12~24 DP Petaflops, 125~325TB/s Mem, 55~185Tbit/s NW), innovative high cost/performance packaging & design, in mere 180m^2...
2. “Extreme Green” – ~10GFlops/W power-efficient architecture, system-wide power control, advanced cooling, future energy reservoir load leveling & energy recovery
3. “Big Data AI – HPC Convergence” – Extreme high BW & FLOPS, deep memory hierarchy, extreme I/O acceleration, for machine learning, graph processing, …
4. “Cloud SC” – dynamic deployment, container-based node co-location & dynamic configuration, resource elasticity, assimilation of public clouds…
5. “Transparency” - full monitoring & user visibility of machine & job state, accountability via reproducibility

2006 TSUBAME1.0
80 Teraflops, #1 Asia #7 World
“Everybody’s Supercomputer”

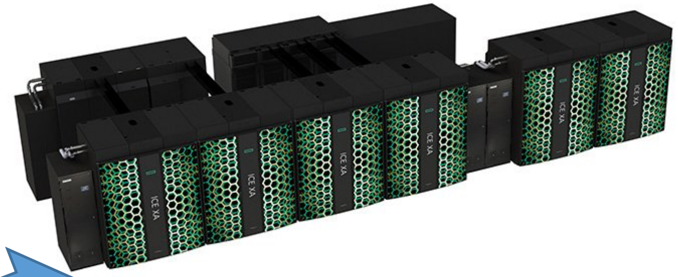
2010 TSUBAME2.0
2.4 Petaflops #4 World
“Greenest Production SC”



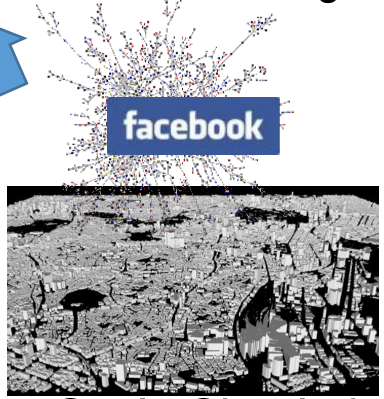
2011 ACM Gordon Bell Prize

2013
TSUBAME2.5
upgrade
5.7PF DFP
/17.1PF SFP
20% power
reduction

2013 TSUBAME-KFC
#1 Green 500



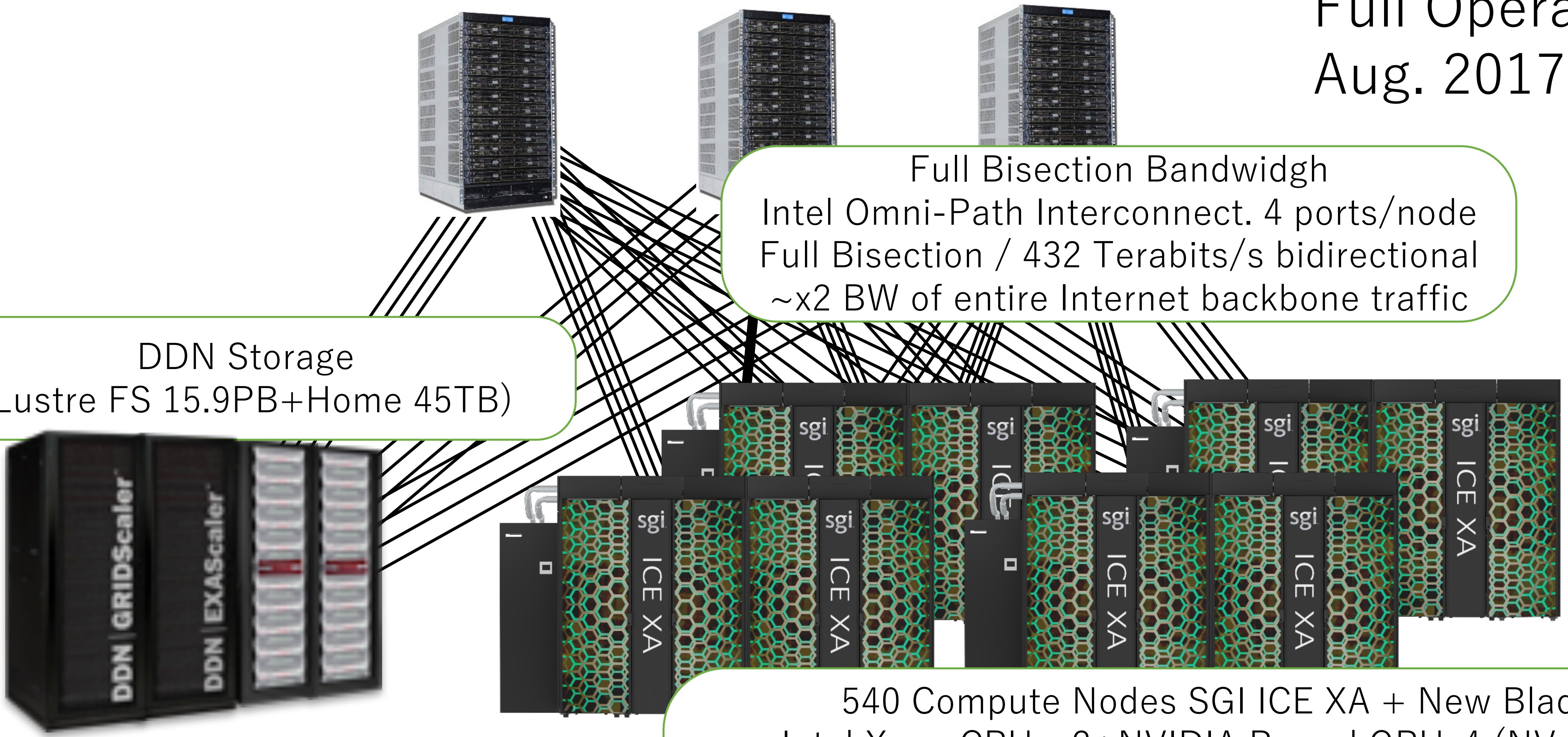
2016 TSUBAME3.0+2.5
~18PF(DFP) 3~4PB/s Mem BW
10GFlops/W power efficiency
Big Data & Cloud Convergence



Large Scale Simulation
Big Data Analytics
Industrial Apps

Overview of TSUBAME3.0

Full Operations
Aug. 2017

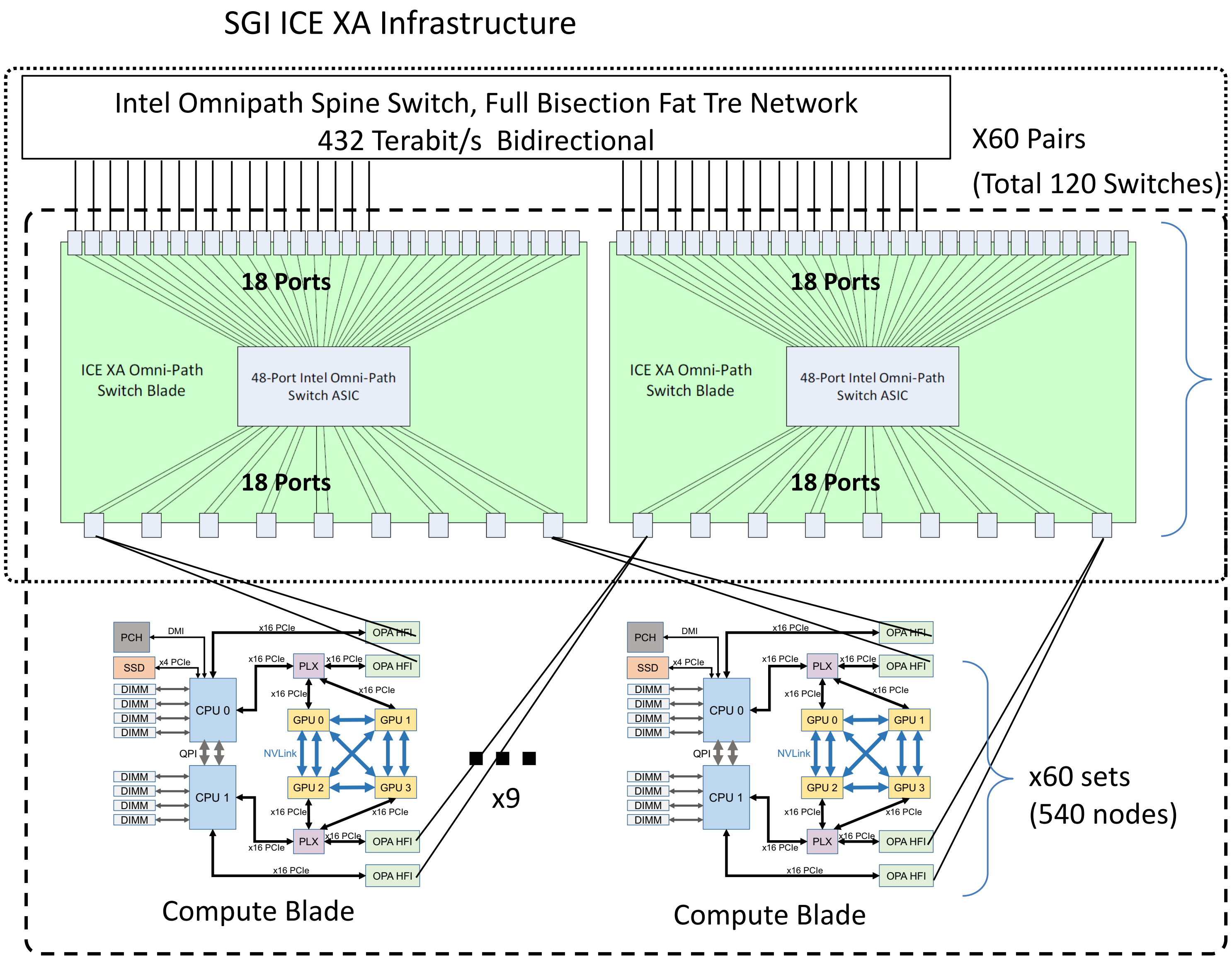


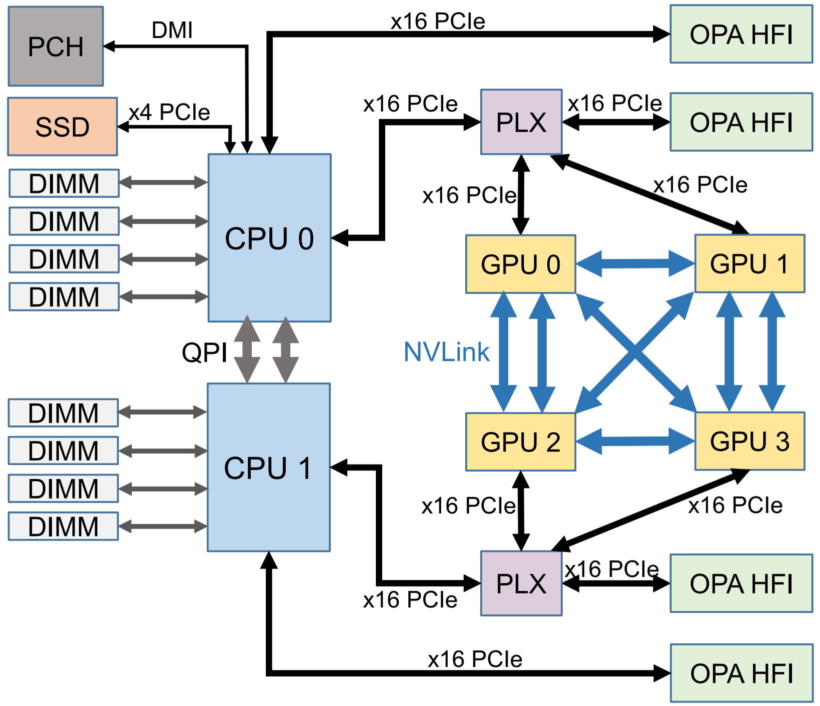
Full Bisection Bandwidgh
Intel Omni-Path Interconnect. 4 ports/node
Full Bisection / 432 Terabits/s bidirectional
~x2 BW of entire Internet backbone traffic

DDN Storage
(Lustre FS 15.9PB+Home 45TB)

540 Compute Nodes SGI ICE XA + New Blade
Intel Xeon CPU x 2+NVIDIA Pascal GPUx4 (NV-Link)
256GB memory 2TB Intel NVMe SSD
47.2 AI-Petaflops, 12.1 Petaflops

TSUBAME3.0 Compute Node SGI ICE-XA, a New GPU Compute Blade Co-Designed by SGI and Tokyo Tech GSIC

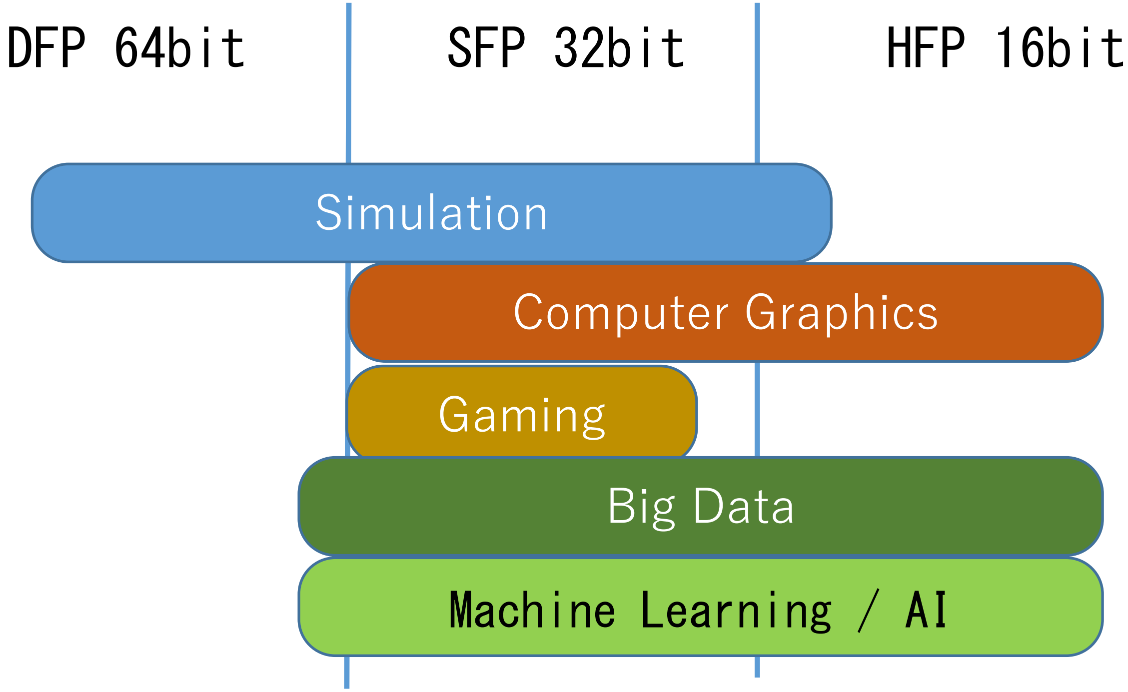


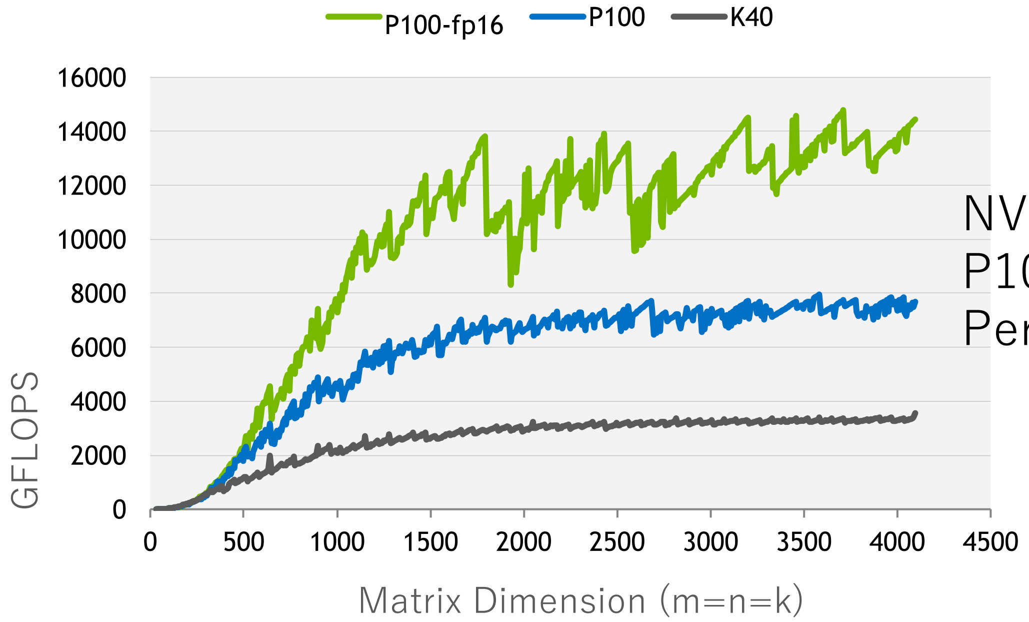


High performance "Fat Node"

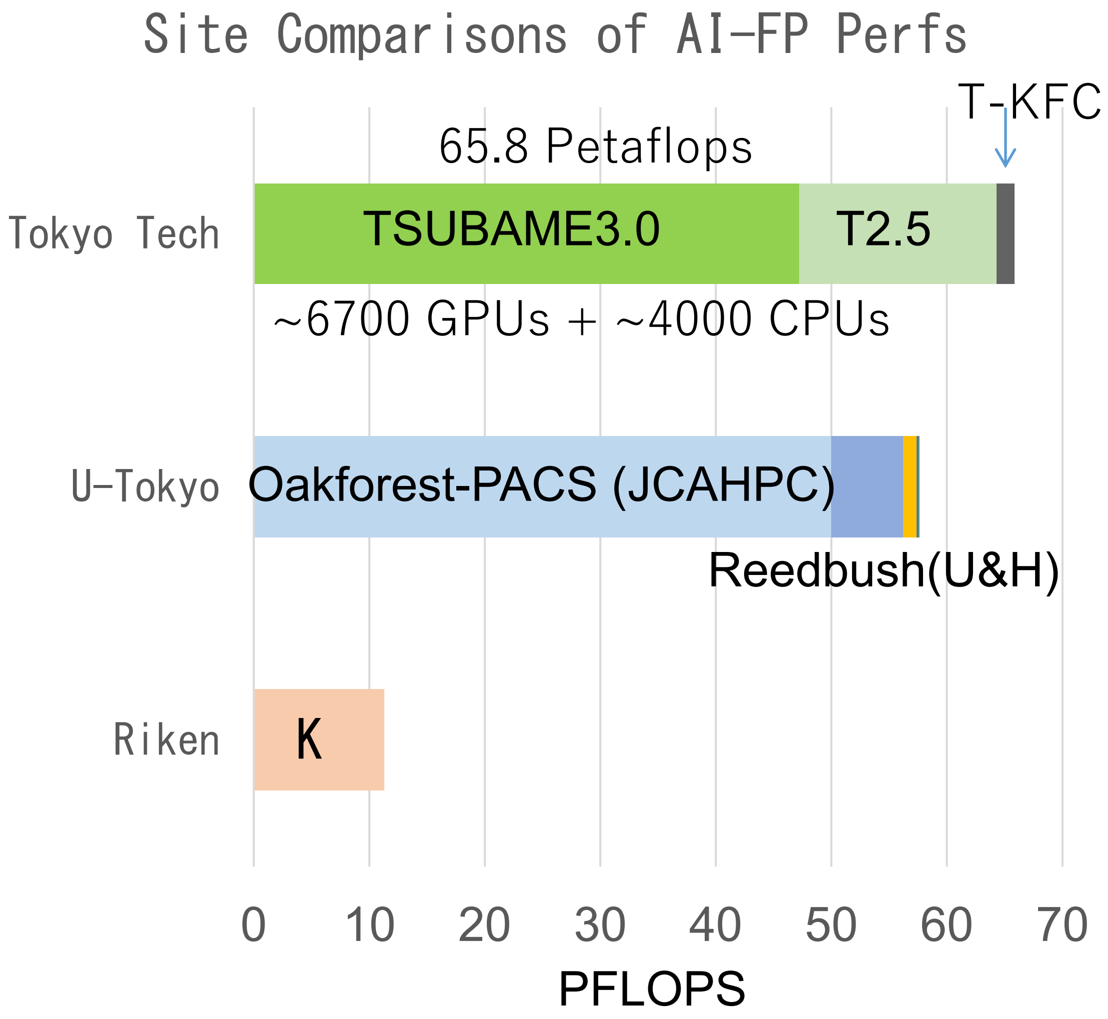
- High Performance 4 SXM2(NVLink) NVIDIA Pascal P100 GPU + Xeon
- High Network Bandwidth – Intel Omnipath 100GBps x 4 = 400Gbps
- High I/O Bandwidth - Intel 2 TeraByte NVMe
  - > 1PB & 1.5~2TB/s system total
- Ultra High Density, Hot Water Cooled Blades
  - 36 blades / rack = 144 GPU + 72 CPU, 50-60KW, x10 thermals c.f. IDC

Tokyo Tech GSIC leads Japan in aggregated AI-capable FLOPS TSUBAME3+2.5+KFC, in all Supercomuters and CloudsNV

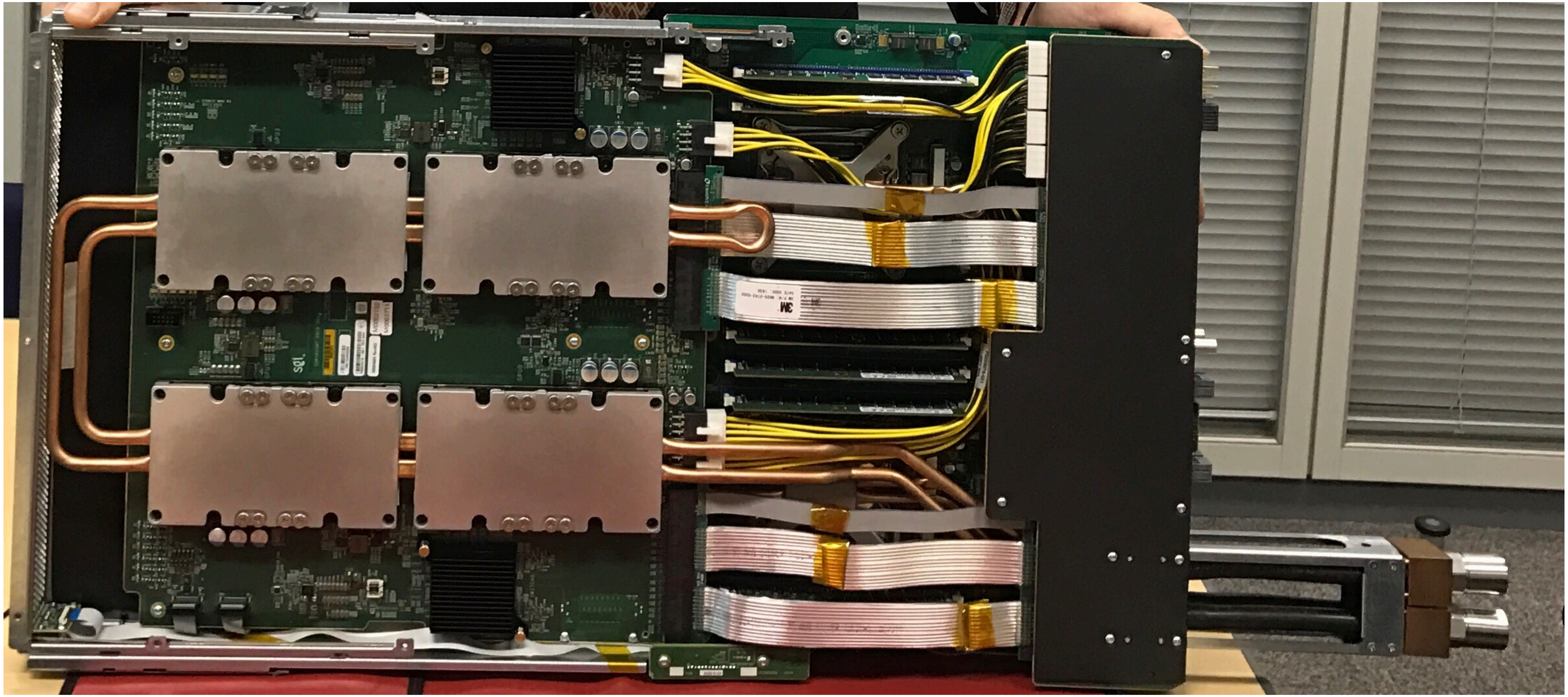




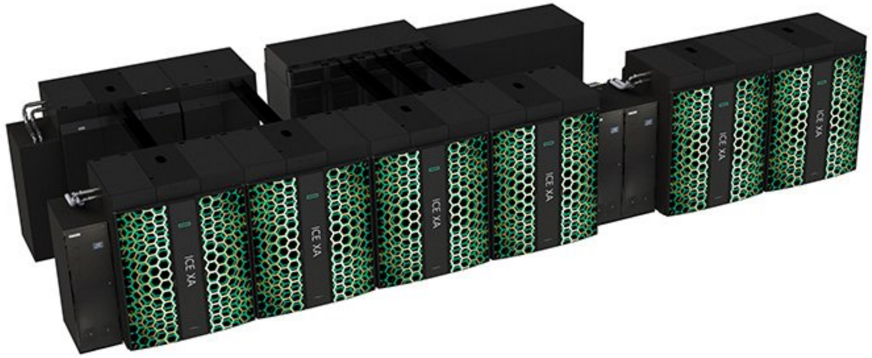
NVIDIA Pascal P100 DGEMM Performane



TSUBAME3.0 SGI ICE-XA Blade (new)
- Plan to become a future HPE product



TSUBAME3.0 Datacenter



15 SGI ICE-XA Racks
2 Network Racks
3 DDN Storage Racks
20 Total Racks

Compute racks cooled with
32 degrees warm water,
yearound ambient cooling
PUE = 1.033

AI R&D Investments in METI

FY2015	FY2016	FY2017	FY2018	FY2019

Next-Generation AI & Robotics Core Technology Development

(5 yr National Project)

10M

30M US$

30M US$

30M US$

30M US$

Foundation of AIRC @AIST

9M

Acceleration of AI R&D (FY 15 Supplementary Budget)

AAIC, AIST AI Cloud →400x Tesla P100, Spark-based

~175M US$ (ABCI+D
emo Lab)

Global Open Innovation Arena for AI R&D

(FY16 Supplementary Budget)

- **ABCI, AI-Bridging Cloud Infrastructure →130PFLOPS(AI), PUE < 1.1, < 3MW**
- **Demonstration env. for Robotics/Industry 4.0**
- **R&D "base" for AI-accelerated Nanofabrication and Medical technologies**

ABCI & Datability R&D (Plan)

(5 yr National Project)

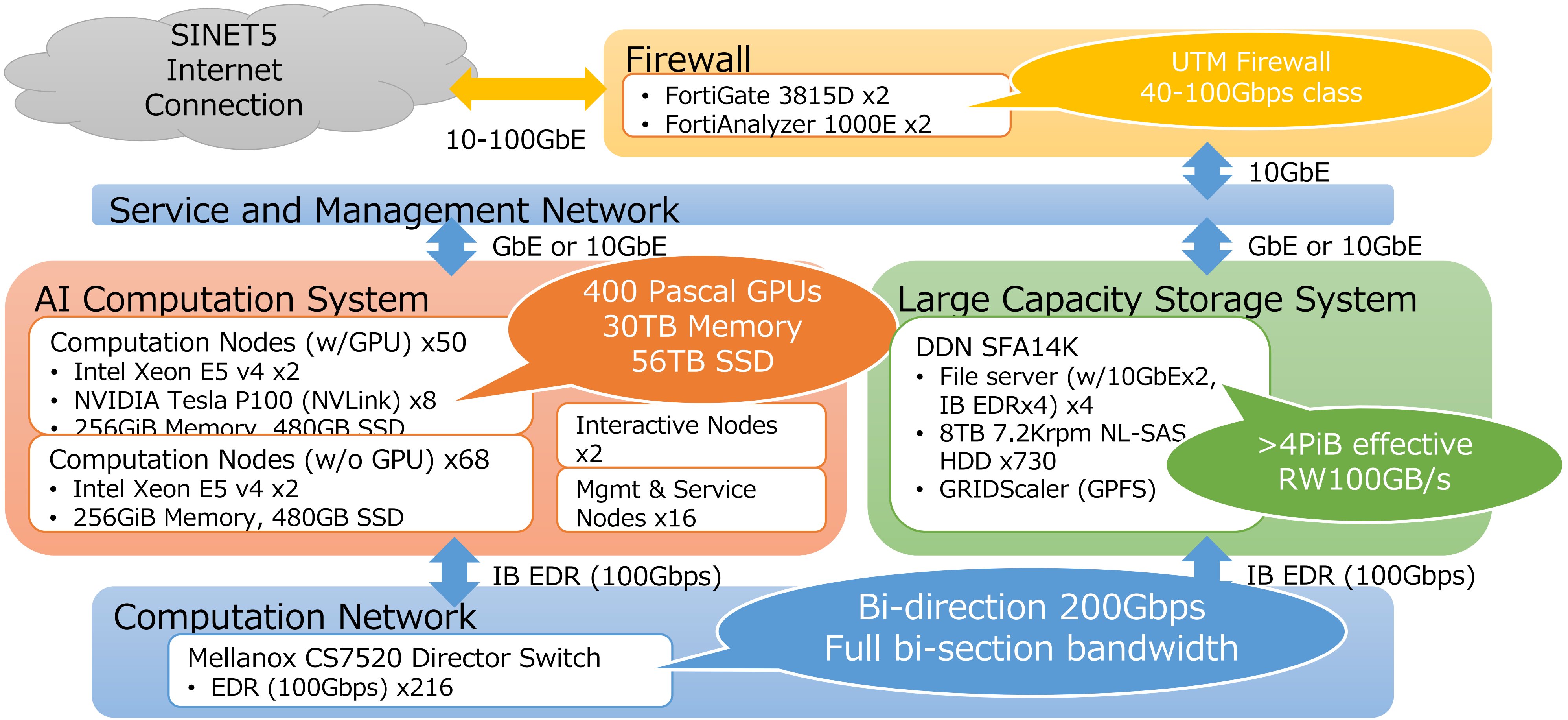
?0M US$

?0M US$

?0M US$

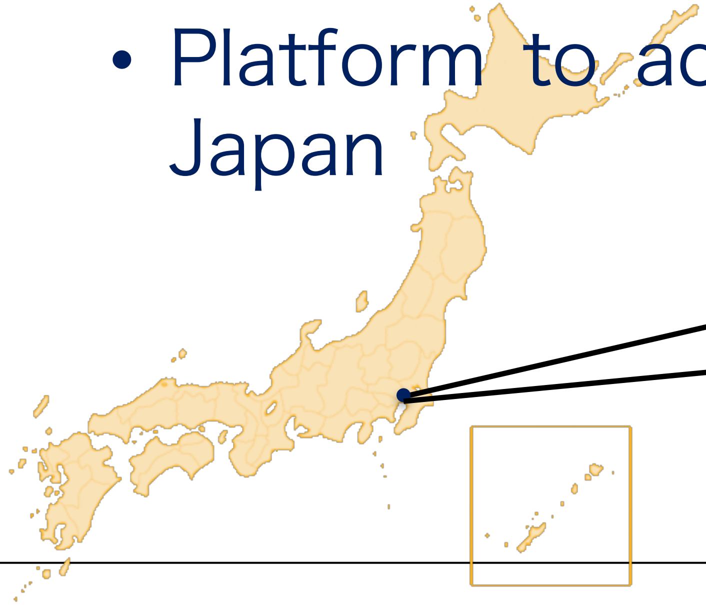
ABCI Prototype: AIST AI Cloud (AAIC) March 2017 (System Vendor: NEC)

- **400x NVIDIA Tesla P100s and Infiniband EDR** accelerate various AI workloads including ML (Machine Learning) and DL (Deep Learning).
- Advanced data analytics leveraged by **4PiB shared Big Data Storage and Apache Spark** w/ its ecosystem.



METI AIST-AIRC ABCI as the *worlds first large-scale OPEN AI Infrastructure*

- **ABCI:** AI Bridging Cloud Infrastructure
  - Top-Level SC compute & data capability (130~200 AI-Petaflops)
  - Open Public & Dedicated infrastructure for AI & Big Data Algorithms, Software and Applications
  - Platform to accelerate joint academic-industry R&D for AI in Japan



- 130~200 AI-Petaflops
- < 3MW Power
- < 1.1 Avg. PUE
- Operational 2017Q3~Q4

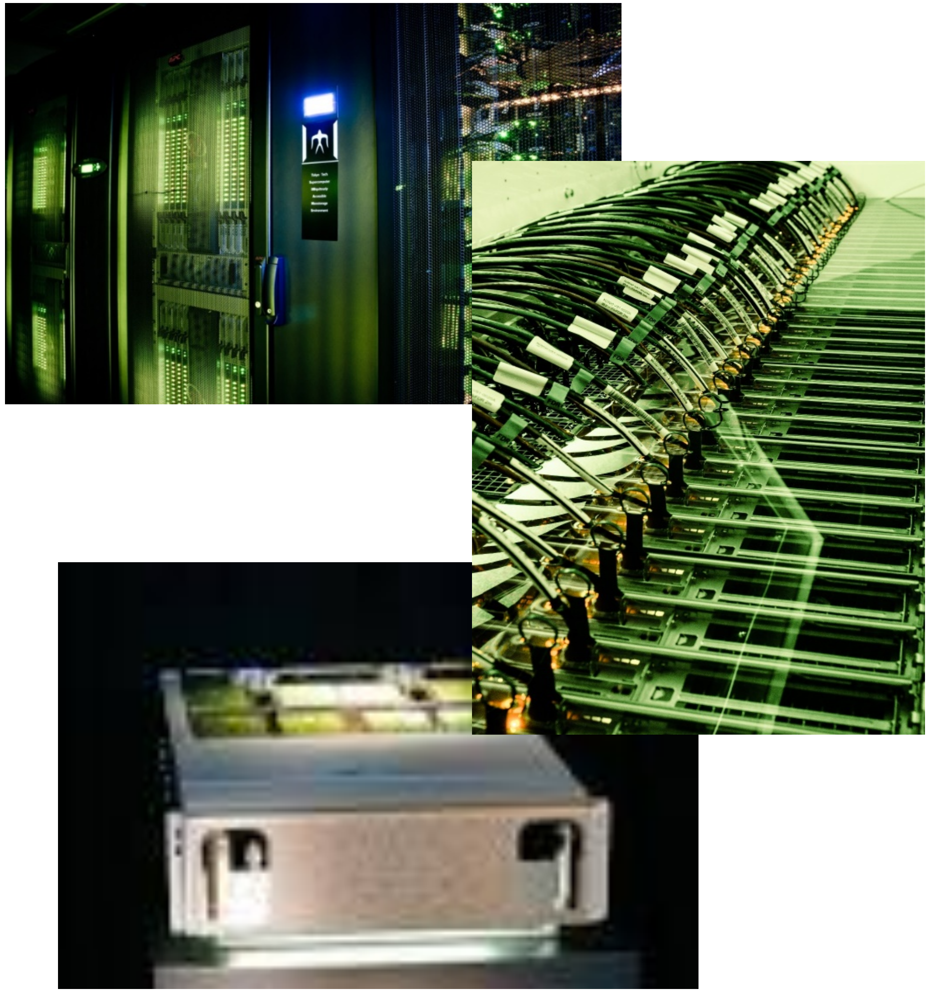
東京大学 THE UNIVERSITY OF TOKYO

AIST NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

Univ. Tokyo Kashiwa Campus

ABCI – 2017Q4~ 2018Q1

- **Extreme computing power**
  - w/ **130~200 AI-PFlops** for AI, ML, DL
  - **x1 million speedup** over high-end PC: 1 Day training for 3000-Year DNN training job
  - TSUBAME-KFC (1.4 AI-Pflops) x 90 users (T2 avg)
- **Big Data and HPC converged modern design**
  - For advanced data analytics (Big Data) and scientific simulation (HPC), etc.
  - Leverage Tokyo Tech's "TSUBAME3" design, but differences/enhancements being AI/BD centric
- **Ultra high bandwidth and low latency in memory, network, and storage**
  - For accelerating various AI/BD workloads
  - Data-centric architecture, optimizes data movement
- **Big Data/AI and HPC SW Stack Convergence**
  - **Incl. results from JST-CREST EBD**
  - **Wide contributions from the PC Cluster community desirable.**
- **RFC just out, includes 10 BD/ML benchmarks**
  - **No HPC benchmarks**



ABCI Cloud Infrastructure

- **Ultra-dense IDC design from ground-up**
  - Custom inexpensive lightweight “warehouse” building w/ substantial earthquake tolerance
  - **x20 thermal density of standard IDC**
- **Extreme green**
  - Ambient warm liquid cooling, large Li-ion battery storage, and high-efficiency power supplies, etc.
  - **Commoditizing supercomputer cooling technologies to Clouds (60KW/rack)**
- **Cloud ecosystem**
  - Wide-ranging Big Data and HPC standard software stacks
- **Advanced cloud-based operation**
  - Incl. dynamic deployment, container-based virtualized provisioning, multitenant partitioning, and automatic failure recovery, etc.
  - Joining HPC and Cloud Software stack for real

ABCI AI-IDC CG Image

Reference Image

引用元: NEC導入事例

TSUBAME3.0&ABCI Comparison Chart

	TSUBAME3 (2017/7)	ABCI (2018/3)	C.f.: K (2012)
AI-FLOPS Peak AI Performance	47.2 Pflops (DFP 12.1 PFlops) 3.1 PetaFlops/rack	130~200 Pflops, (DFP NA) 3~4 PetaFlops/rack	11.3 Petaflops 12.3 Tflops/rack
System Packaging	Custom SC (ICE-XA), Liquid Cool	19 inch rack (LC), ABCI-IDC	Custom SC (LC)
Operational Power incl. Cooling	Below 1MW	Approx. 2MW	Over 15MW
Max Rack Thermals & PUE	61KW, 1.033	50-60KW, below 1.1	~20KW, ~1.3
Node Hardware Architecture	Many-Core (NVIDIA Pascal P100) + Multi-Core (Intel Xeon)	Many-Core AI/DL oriented processor (incl. GPUs)	Heavyweight Multi-Core
Memory Technology	HBM2+DDR4	On Die Memory + DDR4	DDR3
Network Technology	Intel OmniPath, 4 x 100Gbps / node, full bisection, optical NW	Injection/bisection scaled down c.f. to save cost & IDC friendly	Copper Tofu 6-D torus custom NW
Per-node non volatile memory	2TeraByte NVMe/node	> 400GB NVMe/node	None
Power monitoring and control	Detailed node / whole system power monitoring & control	Detailed node / whole system power monitoring & control	Whole system monitoring only
Cloud and Virtualization, AI	**All nodes container virtualization, horizontal node splits, Cloud API dynamic provisioning, ML Stack**	**All nodes container virtualization, horizontal node splits, Cloud API dynamic provisioning, ML Stack**	None
Procurement Benchmarks	HPC-Oriented Benchmarks	BD & DNN Benchmarks	HPC Benchmarks

Fujitsu Deep Learning Processor (DLU™)

FUJITSU

K computer

FY201
8～
D
T
M
(Deep Learning
Unit)
L
U

Supercomputer K technologies

DLU™ features

- Architecture designed for Deep Learning
- High performance HBM2 memory
- Low power design

→ Goal: 10x Performance/Watt compared to others

- Massively parallel ： Apply supercomputer interconnect technology

→ Ability to handle large scale neural networks

→ TOFU Network derivative for massive scaling

"Exascale" AI possible in 1H2019

Software Ecosystem for HPC in AI

Different SW Ecosystem between HPC and AI/BD/Cloud

How to achieve convergence—for real, for rapid tech transfer

Existing Clouds

- BD/AI User Applications
- Machine Learnig MLlib/ Mahout/Chainer
- Graph Processing GraphX/ Giraph /ScaleGraph
- SQL/Non-SQL Hive/Pig
- Java · Scala · Python + IDL
- MapReduce Framework Spark/Hadoop
- RDB PostgresQL
- CloudDB/NoSQL Hbase/Cassandra/MondoDB
- Distributed Filesysem HDFS & Object Store
- Coordination Service ZooKeeper
- VM(KVM), Container(Docker), Cloud Services (OpenStack)
- Linux OS
- Ethernet TOR Swtiches High Latency/Low Capacity NW
- Local Node Storage
- x86 CPU

Application Layer

- Cloud Jobs often Interactive w/resource control REST APIs
- HPC Jobs are Batch-Oriented, resource control by MPI

System Software Layer

- Cloud employs High Productivity Languages but performance neglected, focus on data analytics and dynamic frequent changes
- HPC employs High Performance Languages but requires Ninja Programmers, low productivity. Kernels & compilers well tuned & result shared by many programs, less rewrite
- Cloud focused on databases and data manipulation workflow
- HPC focused on compute kernels, even for data processing. Jobs scales to thousands of jobs, thus debugging and performance tuning
- Cloud requires purpose-specific computing/data environment as well as their mutual isolation & security
- HPC requires environment for fast & lean use of resources, but on modern machines require considerable system software support

OS Layer

Hardware Layer

- Cloud HW based on Web Server "commodity" x86 servers, distributed storage on nodes assuming REST API access
- HPC HW aggressively adopts new technologies such a s GPUs, focused on ultimate performance at higher cost, shared storage to support legacy apps

Existing Supercomputers

- HPC User Code
- Numerical Libraries LAPACK, FFTW
- Various DSLs
- Workflow Systems
- Fortran · C · C++ + IDL
- MPI · OpenMP/ACC · CUDA/OpenCL
- Parallel Debuggers and Profilers
- Parallel Filesystem Lustre, GPFS,
- Batch Job Schedulers PBS Pro, Slurm, UGE
- Linux OS
- InfiniBand/OPA High Capacity Low Latency NW
- High Performance SAN+Burst Buffers
- X86 + Accelerators e.g. GPUs, FPGAs

Various convergence research efforts underway but no realistic converged SW Stack yet => achieving HPC – AI/BD/Cloud convergence key ABCI goal

National Institute of Advanced Industrial Science and Technology AIST

独立行政法人 産業技術総合研究所

National Institute for Advanced Industrial Science and Technology (AIST)

経済産業省 Ministry of Economy, Trade and Industry

Ministry of Economics Trade and Industry (METI)

ラボ長（産総研研究職 or 東工大教員/クロスアポ）

- 副ラボ長（産総研研究職）
- 副ラボ長（産総研事務職）
- ラボ研究主幹（産総研研究職）
- ラボ構成員

Director: Satoshi Matsuoka

TOKYO TECH

GSIC (HPC)

Tsubame 3.0/2.5 Big Data /AI resources

Resources and Acceleration of AI / Big Data, systems research

Joing Organization@Odaiba

AIST-TokyoTech AI/Big Data Open Innovation Laboratory (OIL) "Real World Big Data"

AIST Artificial Intelligence Research Center (AIRC)

Joint Research on AI / Big Data and applications

School of Informatics

Basic Research in Big Data / AI algorithms and methodologies

Other Big Data / AI research organizations and proposals

Application Area
Natural Langauge Processing
Robotics
Security

Matsuoka appointed 15% to AIST AI-OIL

Industrial Collaboration in data, applications

Industry

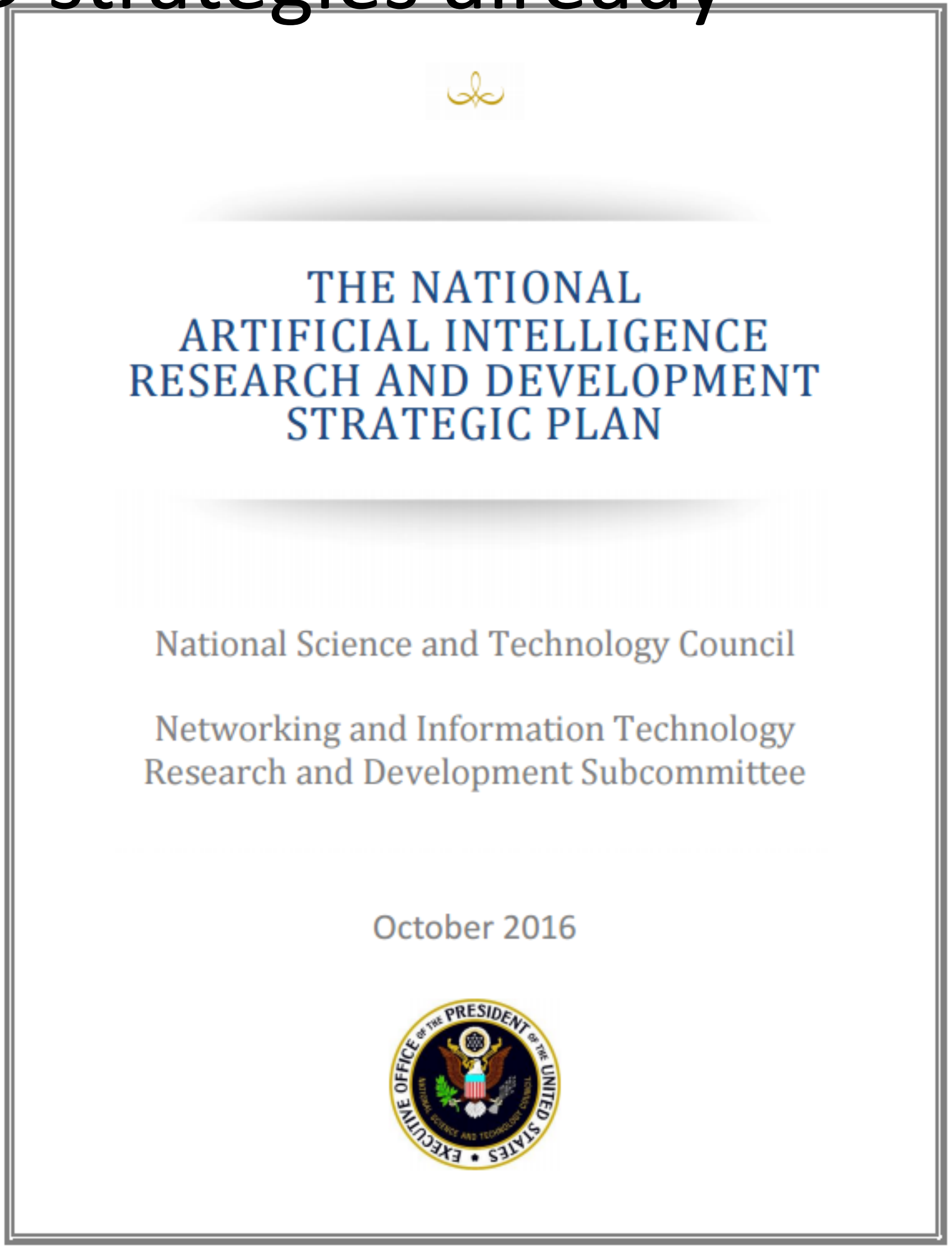
YAHOO! JAPAN

IT LAB DENSO IT LABORATORY, INC.

DENSO ••

We are implementing the US AI&BD strategies already …in Japan, at AIRC w/ABCI

- Strategy 5: Develop shared public datasets and environments for AI training and testing. The depth, quality, and accuracy of training datasets and resources significantly affect AI performance. Researchers need to develop high quality datasets and environments and enable responsible access to high-quality datasets as well as to testing and training resources.
- Strategy 6: Measure and evaluate AI technologies through standards and benchmarks. Essential to advancements in AI are standards, benchmarks, testbeds, and community engagement that guide and evaluate progress in AI. Additional research is needed to develop a broad spectrum of evaluative techniques.


THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

National Science and Technology Council

Networking and Information Technology Research and Development Subcommittee

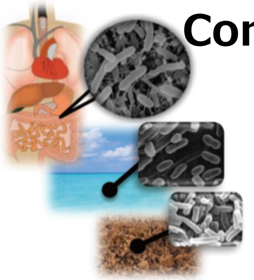
October 2016

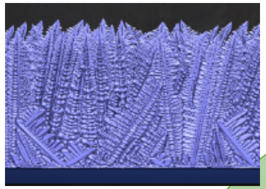
Co-Design of BD/ML/AI with HPC using BD/ML/AI

- for survival of HPC

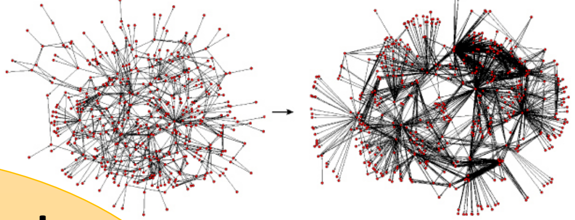
Acceleration and Scaling of BD/ML/AI via HPC and Technologies and Infrastructures

Accelerating Conventional HPC Apps





Large Scale Graphs



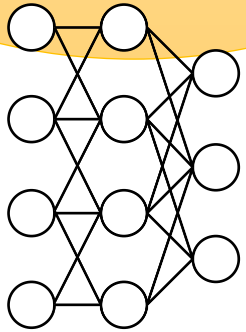
Big Data AI-Oriented Supercomputers

Mutual and Semi-Automated Co-Acceleration of HPC and BD/ML/AI

Big Data and ML/AI Apps and Methodologies

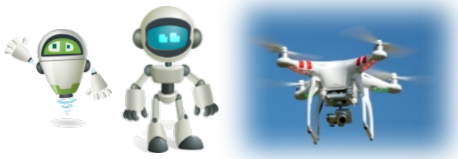
Optimizing System Software and Ops

Image and Video



Acceleration Scaling, and Control of HPC via BD/ML/AI and future SC designs

Robots / Drones

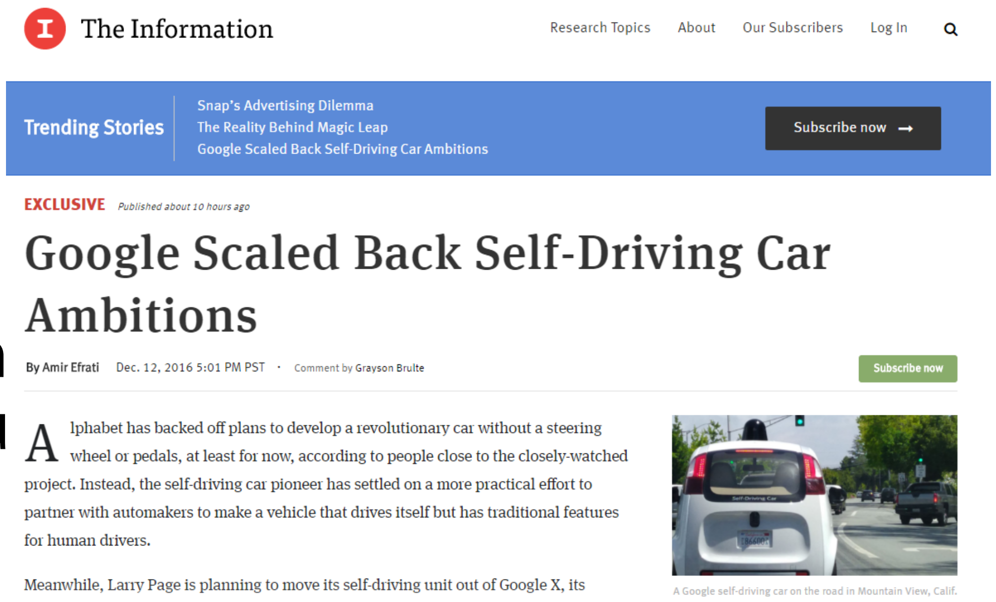


Future Big Data･AI Supercomputer Design

ABCI: World's first and largest open 100 Peta AI-Flops AI Supercomputer, Fall 2017, for co-design

But Commercial Companies esp. the "AI Giants"are Leading AI R&D, are they not?

- Yes, but that is because their shot-term goals could harvest the low hanging fruits in DNN rejuvenated AI
- But AI/BD research is just beginning--- if we leave it to the interests of commercial companies, we cannot tackle difficult problems with no proven ROI
  - Very unhealthy for research
- This is different from more mature fields, such as pharmaceuticals or aerospace, where there is balanced investments and innovations in both academia/government and the indu


The Information

Research Topics About Our Subscribers Log In

Trending Stories

Snap's Advertising Dilemma
The Reality Behind Magic Leap
Google Scaled Back Self-Driving Car Ambitions

Subscribe now →

EXCLUSIVE Published about 10 hours ago

Google Scaled Back Self-Driving Car Ambitions

By Amir Efrati Dec. 12, 2016 5:01 PM PST · Comment by Grayson Brulte

Subscribe now

Alphabet has backed off plans to develop a revolutionary car without a steering wheel or pedals, at least for now, according to people close to the closely-watched project. Instead, the self-driving car pioneer has settled on a more practical effort to partner with automakers to make a vehicle that drives itself but has traditional features for human drivers.

Meanwhile, Larry Page is planning to move its self-driving unit out of Google X, its

A Google self-driving car on the road in Mountain View, Calif.