



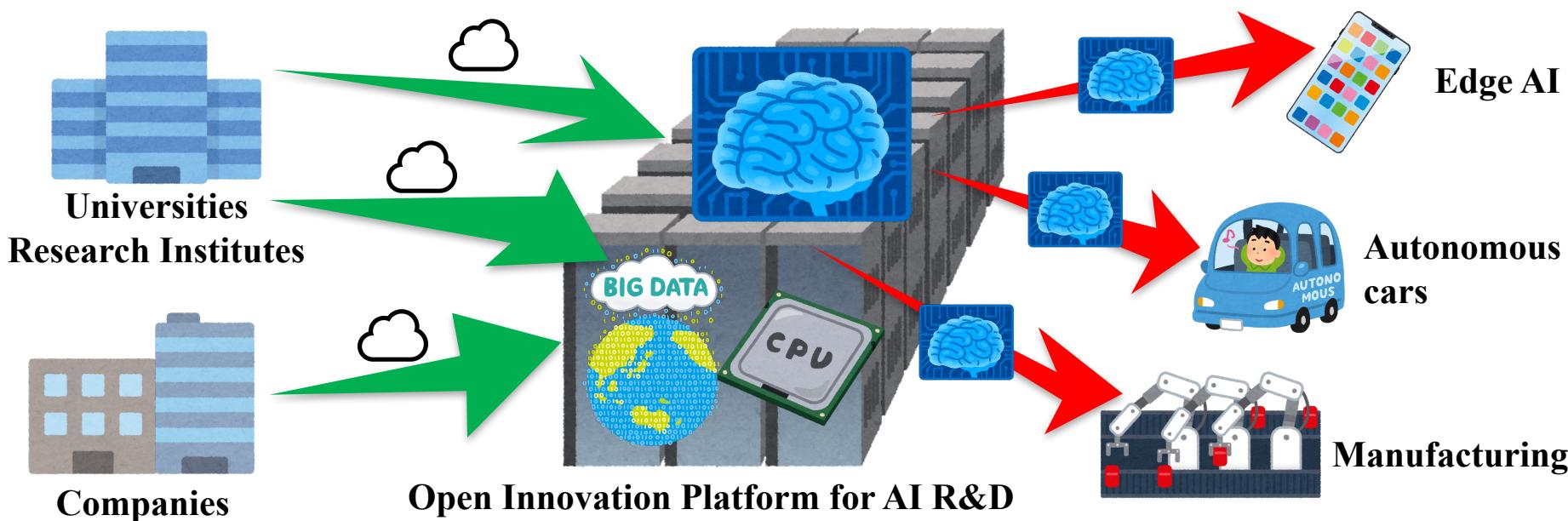
The world's first large-scale Open AI Infrastructure
<http://abci.ai/>

Hirotaka Ogawa

**Deputy Director, Real-World Big Data Computation OIL
Leader, AI Cloud Research Team, AI Research Center**

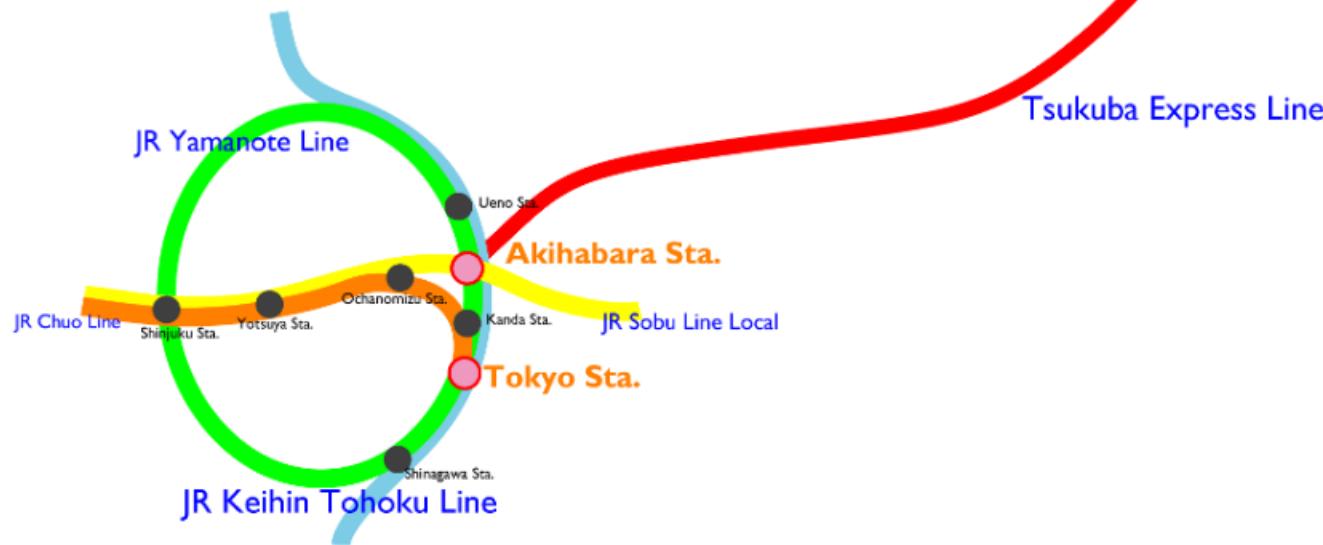
ABCI: AI Bridging Cloud Infrastructure

- **Open, Public, and Dedicated** infrastructure for AI & Big Data Algorithms, Software, and Applications
- **Open Innovation Platform** to accelerate joint academic-industry R&D for AI, international collaborations are also welcome
- Top-level compute capability: **0.55 EFLOPS (DL), 37 PFLOPS (DP)**
- Top-level energy efficiency: lower PUE
- All design and implementations will be open-sourced

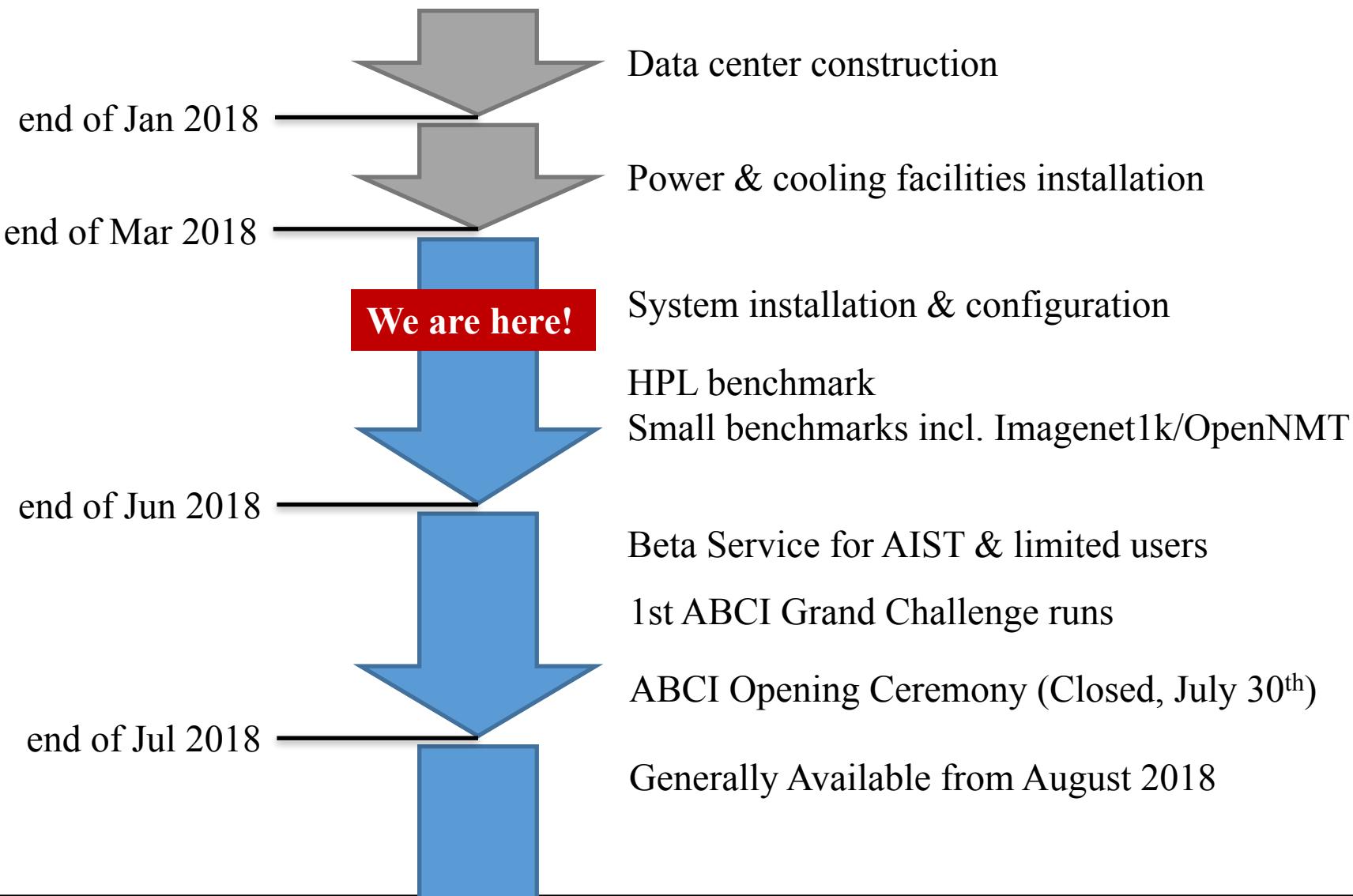


ABCI: Current State

- Vendors: FUJITSU, Fuji-Furukawa Engineering & Construction
- Beta Release: July 2018
- Generally Available: August 2018
- Now on installation, configuration, and benchmarking



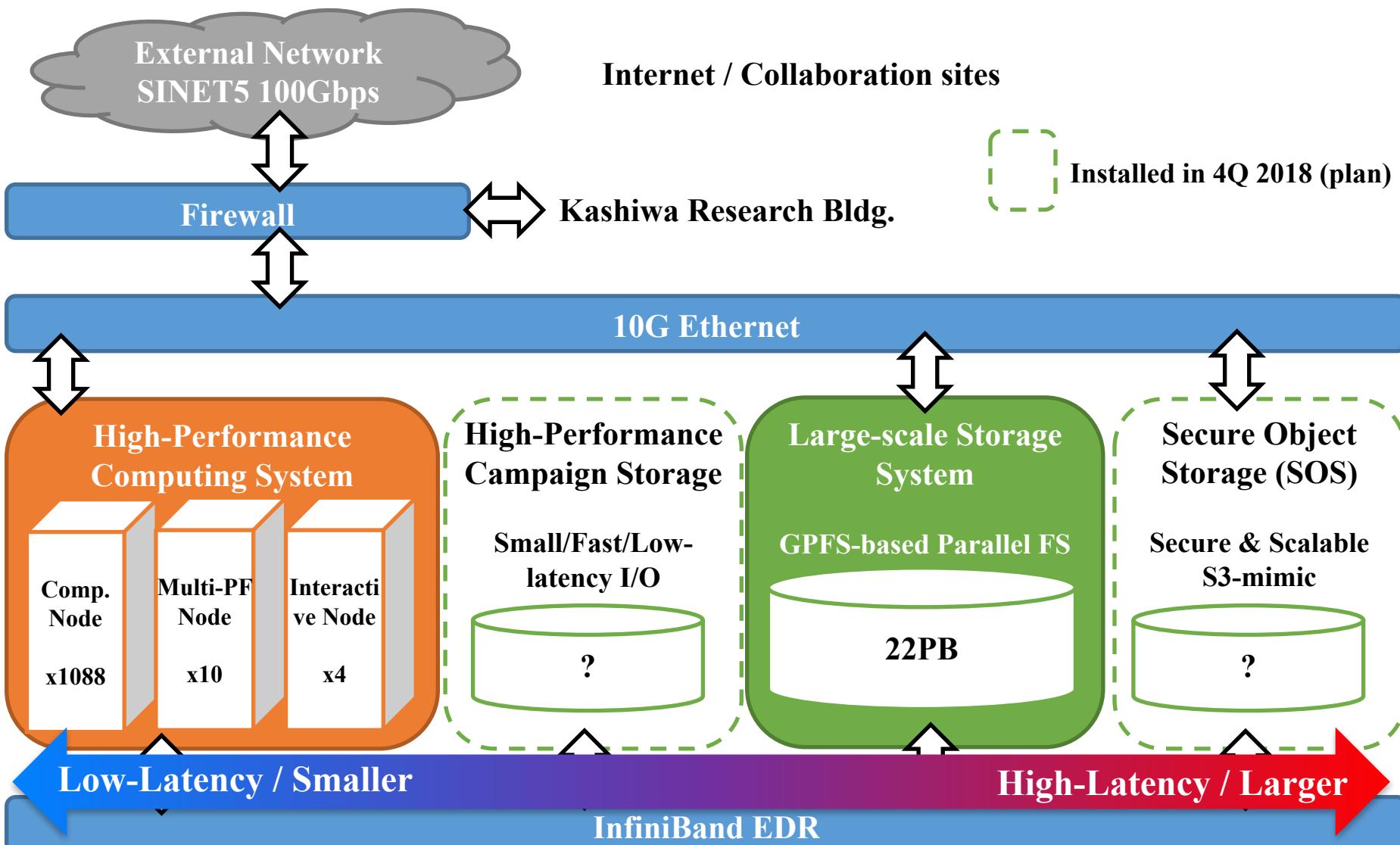
Timeline



4 Major Design Issues on ABCI

- **Massive Compute Capability**
 - Capable to serve for >100 projects, >1000 users
 - Propels forward cutting-edge & scalable AI R&D
- **Leveraging Software ecosystem**
 - State-of-the-art AI software properties built for de facto commodity architecture, e.g., Intel CPUs, NVIDIA GPUs
 - Container & repository support, to ease development & deployment, as well as to share/reuse codes among community
- **Data happy but secure**
 - Large-scale Shared FS that can be served for >100 projects
 - Secure Object Storage (SOS) for sharing/reusing & publishing data and learnt model data
 - To support relatively confidential data, SOS also supports encryption of both comm. channel and data
- **Green AI**
 - Hybrid water-air cooling system up to 70kW/rack
 - Earthquake-tolerant, inexpensive build

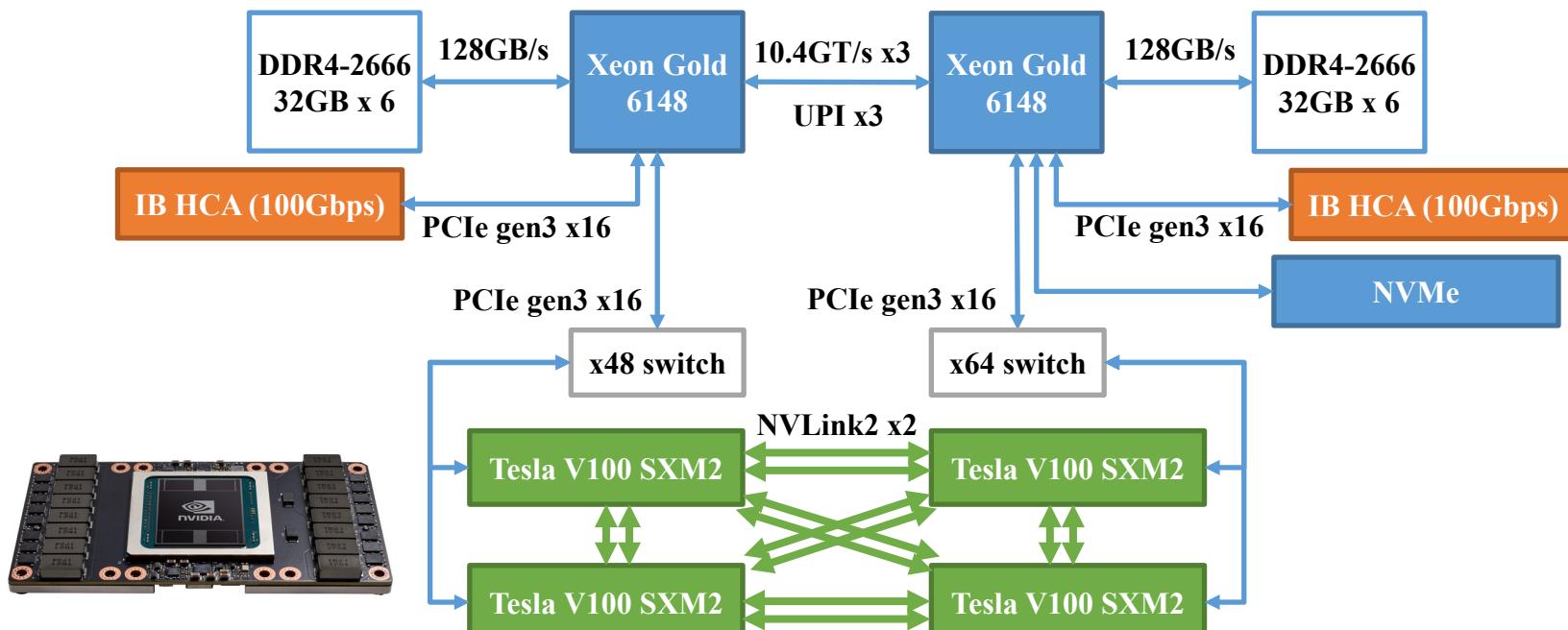
ABCi System Overview



Computing Node

FUJITSU PRIMERGY Server (2 servers in 2U)

CPU	Intel Xeon Gold 6148 (27.5M Cache, 2.40 GHz, 20 Core) x2
GPU	NVIDIA Tesla V100 (SXM2) x4
Memory	384GiB
Local Storage	1.6TB NVMe SSD (Intel SSD DC P4600 u.2) x1
Interconnect	InfiniBand EDR x2



Software and Services

■ Software (Natively installed)

Operating System	CentOS, RHEL
Job Scheduler	Univa Grid Engine
Container Engine	Docker, Singularity
MPI	OpenMPI, MVAPICH
Development tools	Intel Parallel Studio XE Cluster Edition, PGI Professional Edition, Python, Ruby, R, Java, Scala, Perl
Deep Learning	Caffe, Caffe2, TensorFlow, Theano, Torch, PyTorch, CNTK, MXnet, Chainer, Keras, etc.

■ Service Types and Container Support

Applications

Native / user-installed software

System-provided / User-defined containers

Job Bootstrap

Direct run

Singularity

Univa Docker

Service Type

On-Demand

Spot

Reserved

Campaign

Interactive use

Batch use

Advance reservation

Resource Alloc.

Univa Resource Allocator

ABCI Datacenter Overview

Commoditizing supercomputer cooling technologies to Cloud (70kW/rack)



- Single floor, inexpensive build
- Hard concrete floor with 2 tons/m² weight tolerance
- Racks
 - 144 racks max.
 - ABCI uses 43 racks
- Power capacity
 - 3.25 MW max.
 - ABCI uses 2.3MW max.
- Water-Air Hybrid Cooling
 - Water Block: 60kW/Rack
 - Fan Coil Unit: 10kW/Rack
 - Total: 3.2MW min. (summer)

ABCI Grand Challenge

<http://abci.ai/GrandChallenge/>

- “ABCI Grand Challenge” program encourages researchers and students to leverage ABCI’s enormous compute capability to challenge a broad range of critical AI problems
- Important Dates

	Submission deadlines	Acceptance notification	Challenge runs
#1	Apr 30 th , 2018 (ended)	May 30 th , 2018	Last week of Jul 2018
#2	Aug 31 st , 2018	Sep 30 th , 2018	Last week of Oct 2018
#3	Nov 30 th , 2018	Dec 22 nd , 2018	Last week of Jan 2019

- Please don't forget to apply next time, folks!

Coming Soon!

- ABCI will be launched in early summer 2018
- More information will be available
 - <http://abci.ai/>

