

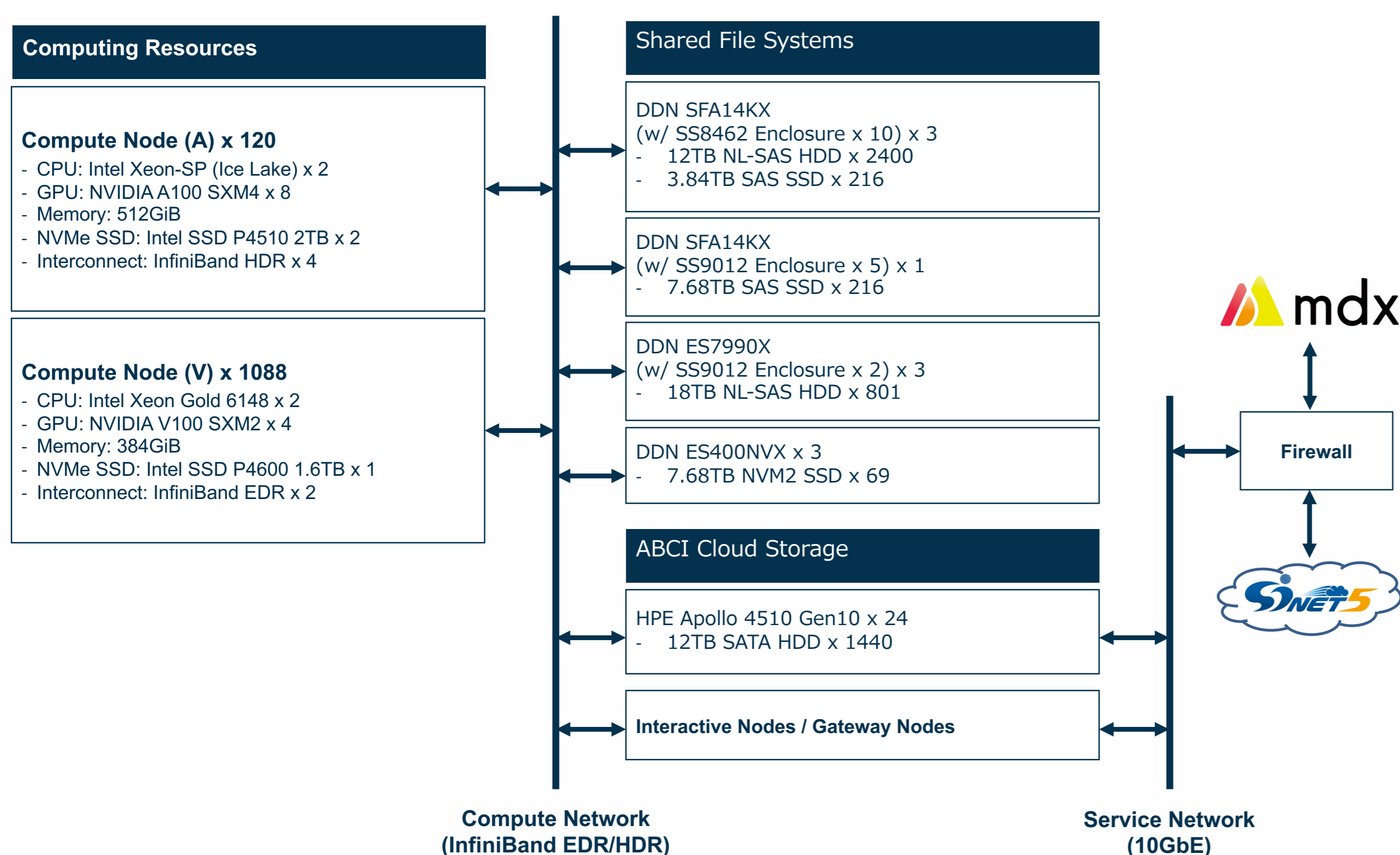
# ABCI: World Class Large-Scale Open AI Infrastructure

## Overview

- **Open, Public, and Dedicated** Infrastructure for AI and Big Data Algorithms, Software, and Applications
- Bridge the gap between AI Research and Industry: Open Innovation Platform to accelerate joint academic-industry R&D for AI
- Started operation in August 2018 and upgraded to ABCI 2.0 in May 2021

Total Performance / Capacity of ABCI 2.0	
HPC (FP64)	56.6 PFLOPS
DL Training (FP32/TF32)	226.0 PFLOPS
DL Training (FP16/BF16)	851.5 PFLOPS
Memory Capacity	573.5 TiB
Memory Bandwidth	5.73 PB/s
Local Storage	2.2 PB

## ABCI Hardware



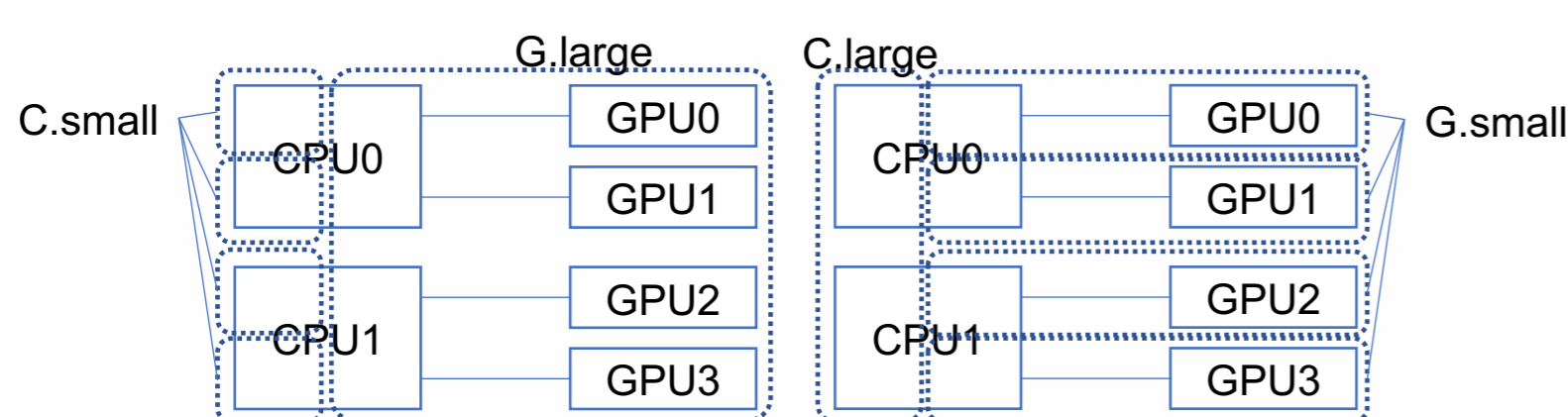
## DC Facilities



- **Single floor, cost effective building**
- **Hard concrete floor 2t/m<sup>2</sup> weight tolerance for racks and cooling pods**
- **Cooling capacity: 3.2MW**
  - 70kW/rack: 60kW water + 10kW air
  - Warm water (32 Celsius) free cooling
- **Power capacity: 3.25 MW**
  - ABCI uses 2.3MW max
  - Average PUE: 1.1 (Estimated)

## Services for Satisfying User Needs

- **Provide various types of resources by separating each Compute Node on demand**
  - Full node, 1 GPU, 4 or 8 GPUs, no GPU instances
  - Each instance has different #GPU, #CPU core, amount of Memory and amount of NVMe SSD



Resource types and allocation in Compute Node (V)

- Provides traditional HPC libraries as well as enables users to deploy various DL frameworks via Python package management software like Pip
- Run any NGC container images using Singularity
- SingularityPRO and Singularity endpoint service are available.

## Dataset and Model Sharing

- **Data Harbor concept for retrieving values by sharing**
  - More efficient, flexible exchange of ML models, input data, and derived data
  - Connect to the real world, through SINET5 in Japan
- **Provide a set of sharing service**
  - ABCI cloud storage service
  - ABCI datasets service: <https://datasets.abci.ai>
  - Workflow tools for using ML models on ABCI (under development)

