

ABCI: World's First Large-Scale Open AI Infrastructure

Overview

- World top-level compute and data processing capability
 - **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

Peak Performance:

550 PFlops (FP16)

37.2 PFlops (FP64)

Effective Performance:

19.88 PFlops (#7 in Top500)

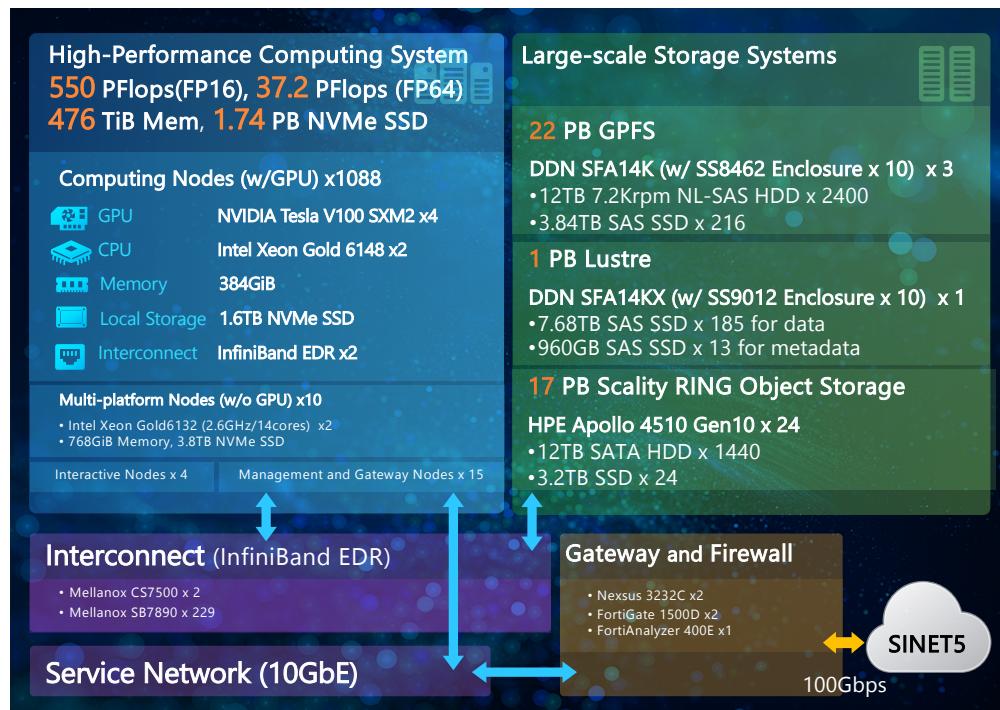
14.423 GFlops/W (#4 in Green500)

508.85 TFlops (#5 in HPCG)

ImageNet training 75 seconds

Average PUE: < 1.1 (Estimated)

ABC Hardware



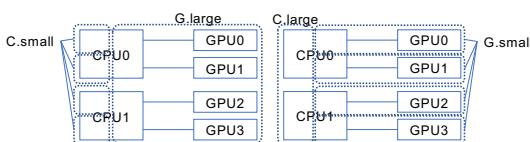
DC Facilities



- Single floor, cost effective building
 - Hard concrete floor 2t/m^2 weight tolerance for racks and cooling pods
 - Cooling capacity: 3.2MW
 - 70kW/rack: 60kW water + 10kW air
 - Warm water (32°C) free cooling
 - Power capacity: 3.25 MW
 - ABCI uses 2.3MW max

Services for Satisfying User Needs

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



- Not only providing traditional HPC libraries, but also users can deploy various DL frameworks via Pip
 - Run any NGC container images using Singularity

DL Performance on ABCI

- **World's Highest Speed in ImageNet-1k Training**
 - The current world record is by Fujitsu Lab and ABCI: 75.08% accuracy in 74.7 seconds.

