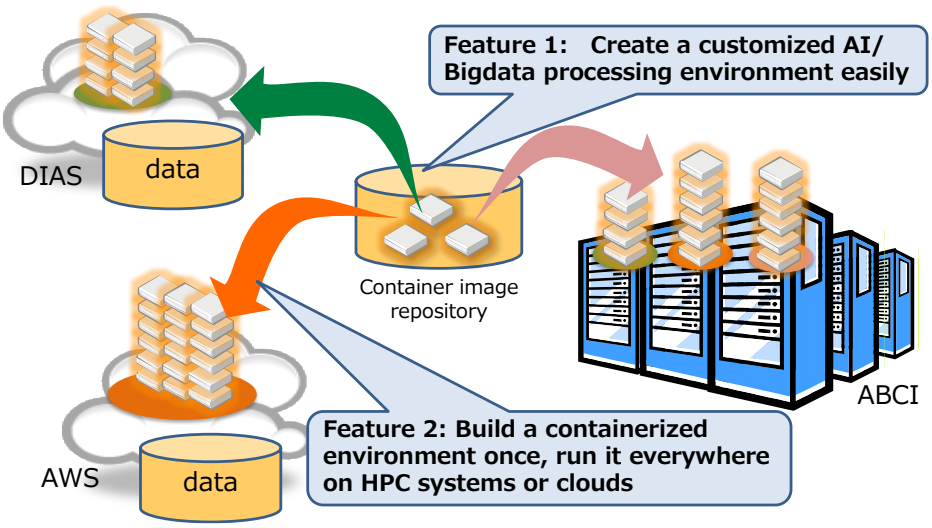
Proposal to a BDEC2 Platform Demonstrator

Contact Person: Ryousei Takano <takaon-ryousei@aist.go.jp>
National Institute of Advanced Industrial Science and Technology, Japan

[Overview]

We propose a demonstration of distributed and containerized AI/Bigdata processing environment in "build once, run everywhere" manner. We have constructed and operates ABCI, an open innovation platform for advancing AI research and deployment [1]. ABCI allows the user to submit a large-scale distributed deep learning job on compute nodes through Singularity container engine and Univa Grid Engine [2]. We expand it on a global scale for research reproducibility and productively. Our primary application is satellite image processing using deep learning techniques (e.g., [3]). Each institution has huge amount of image data and container images are shared with them for extracting knowledges and insights from such bigdata. We just started collaboration with international partners in this field. It is a very early stage of software development.



[Answer for questionary]

1. What innovative capabilities/functionalities will the proposed candidate platform demonstrate (e.g. transcontinuum workflow, edge computing, data logistics, distributed reduction engine, etc.)?
   - Singularity container on HPC systems
   - A data access abstraction layer to seamlessly access data from a container to underlying storage systems such as parallel file system and object store.
   - A persistent identifier (PID) management system for data, workflows, and container images. It is important to guarantee research reproducibility.
2. What applications/communities would/could be addressed?

- Geoscience and remote sensing. But this idea is not limited for this area.

3. What is the "platform vision," i.e. what kind of shared cyberinfrastructure (CI) for science would the further research/design/development of this platform lead to?
   - A distributed and containerized AI/Bigdata processing environment in "build once, run everywhere" manner
4. How available/ready/complete is the set of software components to be used to build the demonstrator?
   - The software development is undergoing.
5. As far as one can tell at this early date, to what extent can this be done with existing and/or otherwise available hardware/software/human resources?
   - Uncertain.
6. What is the potential international footprint of the demonstrator?
   - UCSD/Pacific Research Platform
   - NCHC, Taiwan

[References]

[1] ABCI, https://abci.ai/ (2019)

[2] H. Mikami, et al. "Massively Distributed SGD: ImageNet/ResNet-50 Training in a Flash," https://arxiv.org/abs/1811.05233 (2019)

[3] K. Enomoto, et al. "Image Translation Between Sar and Optical Imagery with Generative Adversarial Nets," IGARSS 2018 (2018)