

Building and Evaluation of ABCI Cloud Storage Service

Yusuke Tanimura, Shinichiro Takizawa, Hiroataka Ogawa and Takahiro Hamanishi
National Institute of Advanced Industrial Science and Technology, Japan

1 INTRODUCTION

AI Bridging Cloud Infrastructure (ABCI) [1] is the world's first large-scale open AI computing infrastructure, constructed and operated by the National Institute of Advanced Industrial Science and Technology (AIST) in Japan. The ABCI service started in August, 2018 and the number of users reached more than 1,000 as of September, 2019. The ABCI integrates technologies for supercomputing, cloud and big data platforms, and supports container-based application deployment and fine-grained resource allocation for encouraging all kinds of AI uses. Recently, in order to accelerate the integration and enhance research activities of adding a high value to data and developing data oriented AI by using a wide variety of data sets in real world, the ABCI Cloud Storage service has been launched. The new service is expected to contribute creation of a supply chain of data and AI models trained from the data on ABCI, which make ABCI achieve a role of the open innovation platform for AI.

This poster presents the detail of design and implementation of the ABCI Cloud Storage service, including the performance evaluation for the basic operations.

2 ABCI CLOUD STORAGE SERVICE

2.1 Features

Figure 1 shows an overview of the storage components and services in ABCI. Unlike two POSIX shared file systems which are accessible only from inside of ABCI, the ABCI Cloud Storage service is accessible from both inside and outside of ABCI and provides the following features to satisfy efficient and secure data sharing requirements of AI and big data applications.

- **Data harbor functionality**

Users can use the service as a gateway depot for import/export data and as a means of data publishing and sharing with outside. ABCI connects to the outside through SINET5 which is a Japanese academic backbone network [4], and thereby takes advantages of the connectivity with many universities and research institutes.

- **Amazon S3 compatible interface**

Users can benefit from the ecosystem of Amazon S3 [2] when accessing data on the service. Various S3 clients that provide faster and/or efficient data transfer in each targeted use case are available.

- **Enterprise-grade encryption**

Users can use in-transit and at rest encryption in a transparent manner, when they are needed.

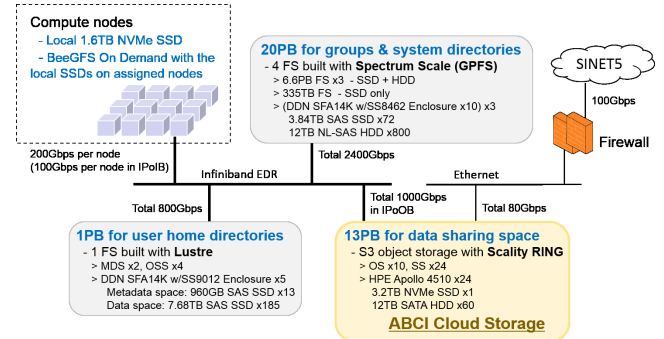


Figure 1: An overview of storage components and services in ABCI

2.2 Design and Implementation

One of the largest design issues is coordination of two account management systems. It is necessary to associate accounts for the cloud storage service to the main ABCI system, and to grant proper user privileges to each account for data publishing and sharing. Because S3 has different semantics for access control from the permission based of the POSIX file system, the authorization is carefully considered with a trade-off between control flexibility and security.

The service has been built with Scality RING [3]. Approximately at most 13 PB storage space is provided by 24 storage nodes, in our data durability setup, and 10 of the nodes also serve as an S3 connector. For the external access from the outside, 2 proxy nodes are setup for now, with a redundant configuration. The server-side encryption is provided by Scality RING though it is not compatible with S3.

These design and implementation details are explained in the poster. The performance evaluation is made by measuring aggregated throughput for both internal and external accesses, an encryption overhead and etc. The results are presented in the poster, too.

ACKNOWLEDGMENTS

This work is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). A part of this work is conducted as research activities of AIST - Tokyo Tech Real World Big-Data Computation Open Innovation Laboratory (RWBC-OIL).

REFERENCES

- [1] AI Bridging Cloud Infrastructure (ABCI). [n.d.]. <https://abci.ai>.
- [2] Amazon S3. [n.d.]. <http://aws.amazon.com/s3/>.
- [3] Scality RING. [n.d.]. <https://www.scality.com>.
- [4] Science Information Network 5. [n.d.]. <https://www.sinet.ad.jp>.