

Cloud Adoption and Utilization Support in Academic Research Community

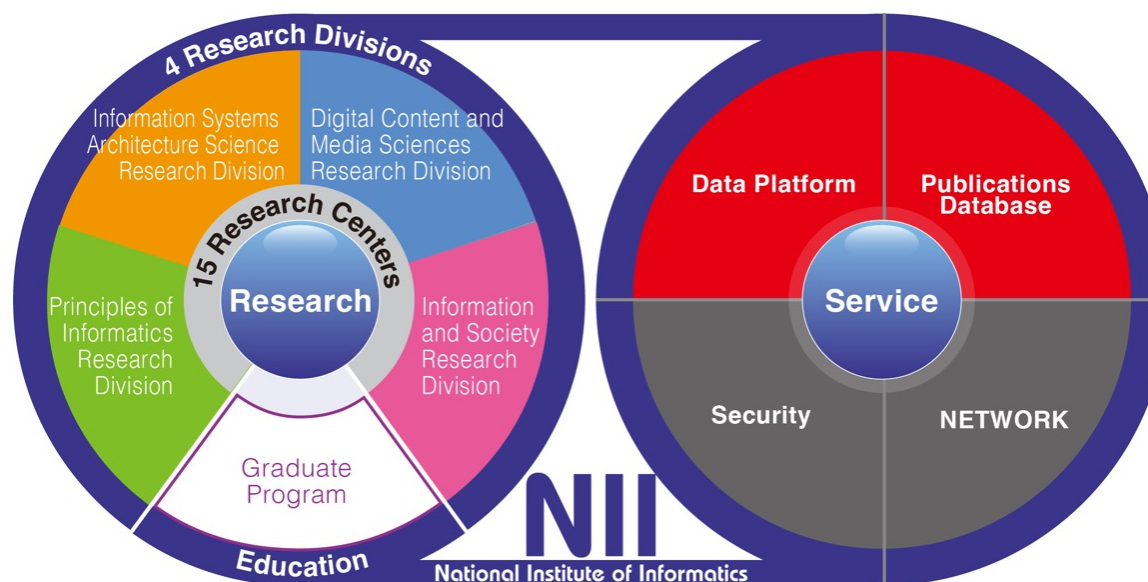
- PoC for Astronomical Observation Data -

Kento Aida

Center for Cloud Research and Development
National Institute of Informatics

National Institute of Informatics

- National Institute of Informatics (NII) seeks to **advance integrated research and development activities in information-related fields**, including networking, software, and content.
- As an inter-university research institute, NII promotes the creation of **a state-of-the-art academic-information infrastructure** (the Cyber Science Infrastructure, or CSI) that is essential to research and education within the broader academic community



Supporting Academic Research Infrastructure

- NII builds and operate the **Science Information NETWORK (SINET)**, high speed (400Gbps), high reliability and multifunctionality network.
- Leveraging the SINET, NII provides an **authentication federation platform, cloud adoption and utilization support**, and **academic content platforms** as well as develops **NII Research Data Cloud** to promote open science. Through those services, NII is working to maintain and provide the **Scientific Research Digital Platform**.
- Furthermore, NII **Security Operation Collaboration Services** contribute to building the framework enabling national universities and other academic institutions to respond quickly to cyber security



<https://www.nii.ac.jp/en/service/>

Center for Cloud Research and Development



Cloud

NII provides universities and research institutions with information and consulting services to support the adoption of clouds, as well as tools to support the use of clouds for education and research, and the development of wide-area data collection and analysis programs.

The center aims at establishment of the national advanced R&E infrastructure taking advantage of clouds. We research and develop advanced cloud infrastructure technologies such as inter-cloud, which utilizes multiple cloud platforms in a federated manner.

- Research and development of infrastructure technologies such as cloud architecture and middleware.
- Research and development of advanced IoT (Internet of Things) system infrastructure technologies using the Mobile SINET and clouds.
- Research and development of cloud and IoT application software in cooperation with researchers in universities.
- Support for cloud adoption and use in academic organizations.

Gakunin Cloud Adoption Support Service

NII collects, disseminates, and shares standard processes and information required when universities and institutes adopt and use cloud services.

- plan cloud adoption
- develop specification and run procurement

- review checklist responses
- conduct individual consultation, etc.

- provide checklist responses
- propose products and services



- Access to checklist responses
- Request for individual consultation
- Others (Sharing information through workshops/seminars)

cloud service checklist



A screenshot of a spreadsheet-style checklist with multiple columns and rows, likely used for tracking cloud service adoption across various institutions.

- Provide checklist responses to universities and research institutes
- Understand the needs of universities and research institutes
- Others (sharing information, attend workshops, etc.)

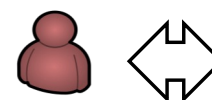
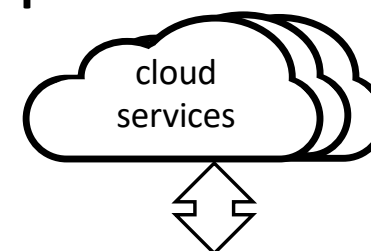
Cloud Gateway

portal system enabling to access cloud service in one-stop manner

- University users can access cloud services (contracted by the university/user) through the service menu.

- University administrators can configure the service menu for the users.
 - avoiding shadow IT
 - localization for the university (API)

- **single sign-on with GakuNin**



Admins

register cloud services contracted with the university



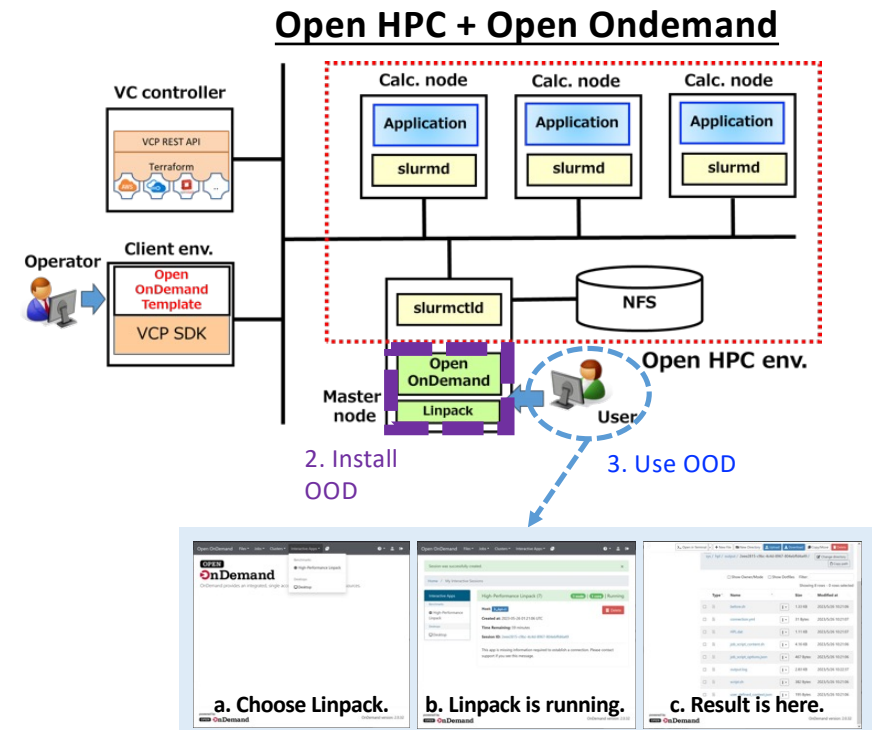
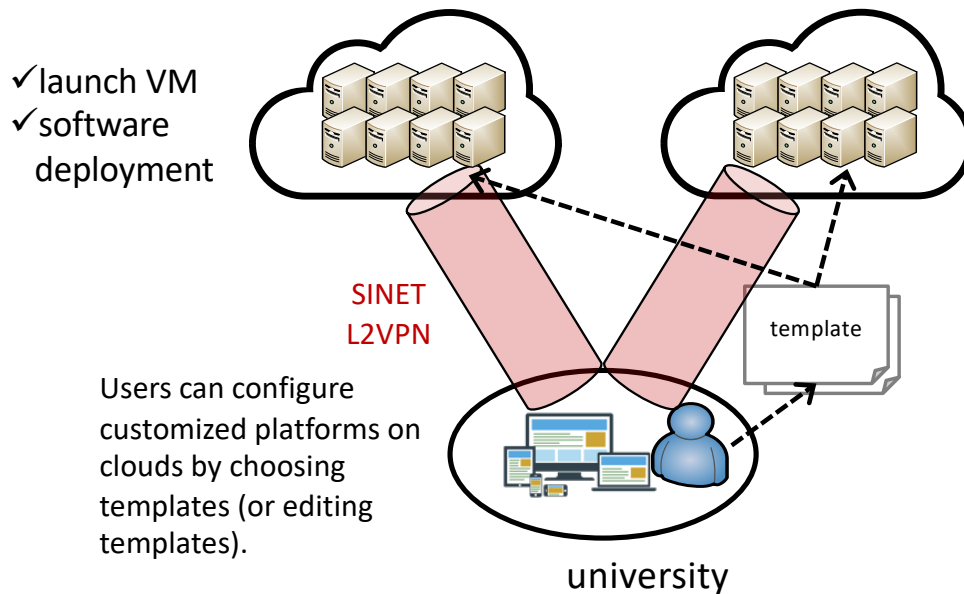
Users (staff/students)

- access the registered cloud services
- register cloud services contracted by the user

On-demand Cloud Configuration

Supports the construction of application environments for research and education on a cloud (IaaS).

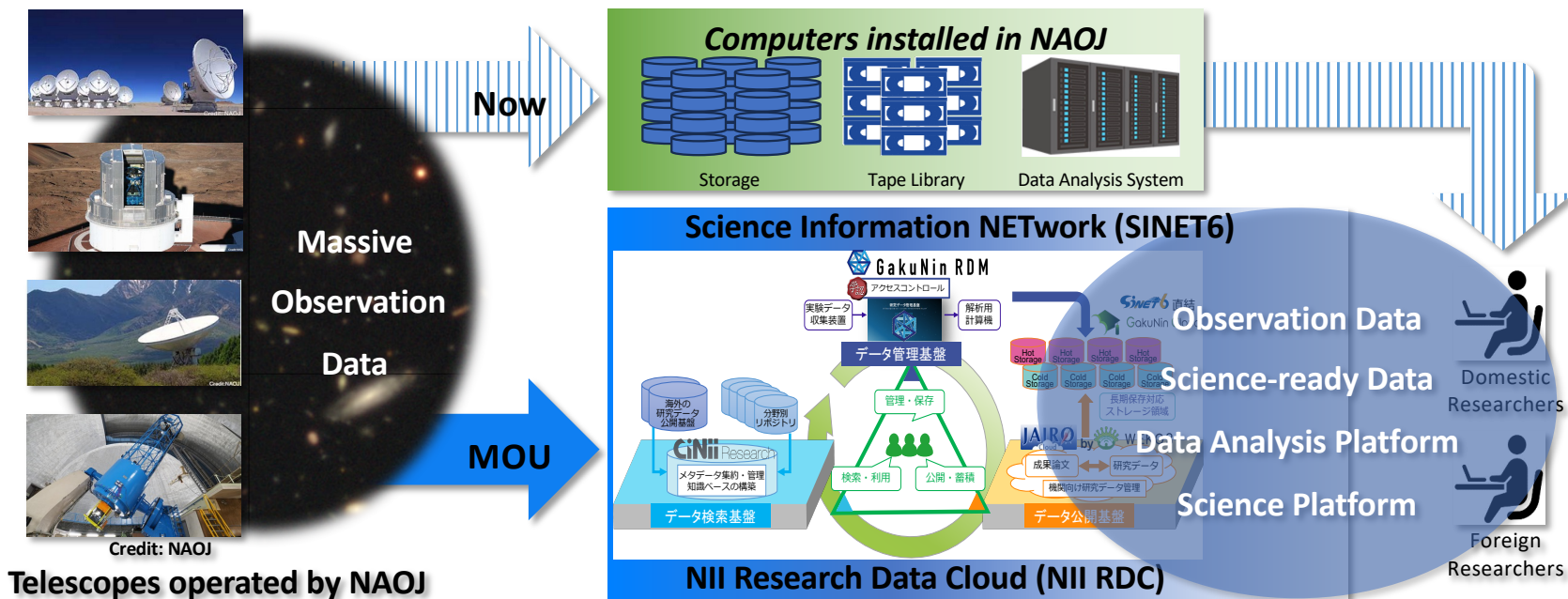
- We use our research product, VCP, as NII's infrastructure service.
- Provides Jupyter notebook-based application templates, e.g., HPC, OpenOnDemand, coursewareHub.



PoC for Archive and Analysis of Astronomical Observation Data

Collaboration with NAOJ regarding Astronomical Observation Data on Cloud

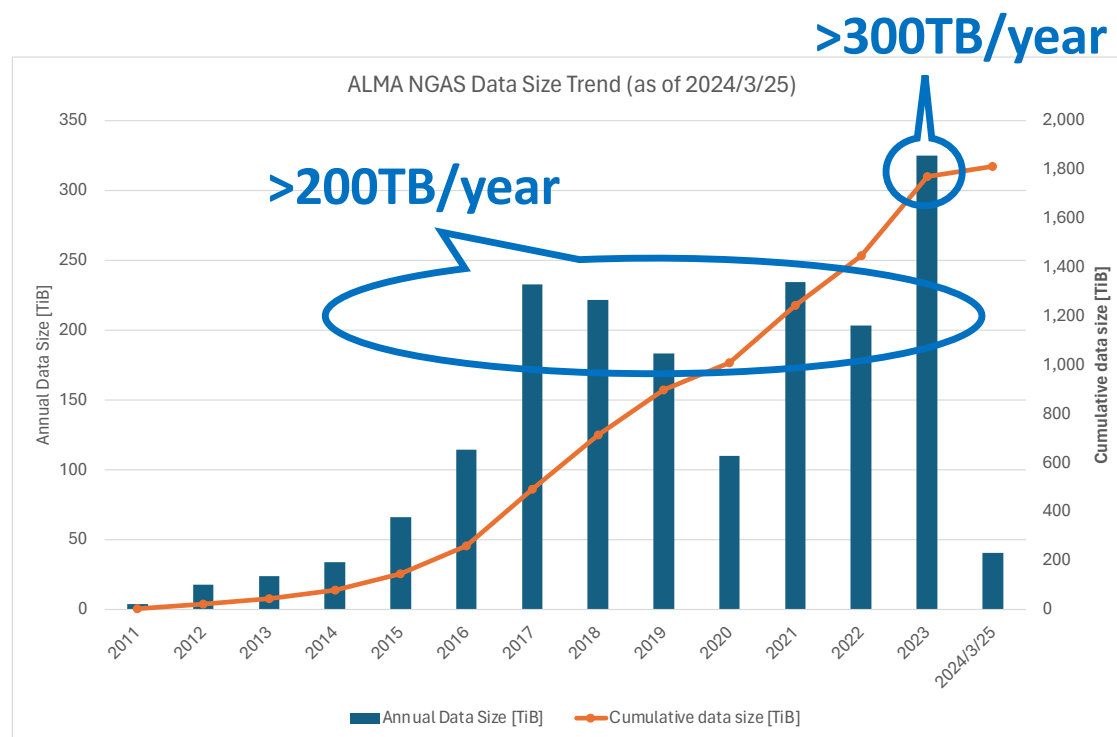
- Astronomy Data Center (National Astronomical Observatory of Japan : NAOJ) and Center for Cloud Research and Development (National Institute of Informatics : NII) concluded an engagement in long-term archiving and publishing of growing astronomical observation data, as well as data analysis... (October 18, 2023)



George Kosugi@NAOJ

ALMA

- Atacama Large Millimeter/submillimeter Array (ALMA)
 - astronomical interferometer installed in Atacama of Chile
 - 66 antennas (12m x 54, 7m x 12)



Issues

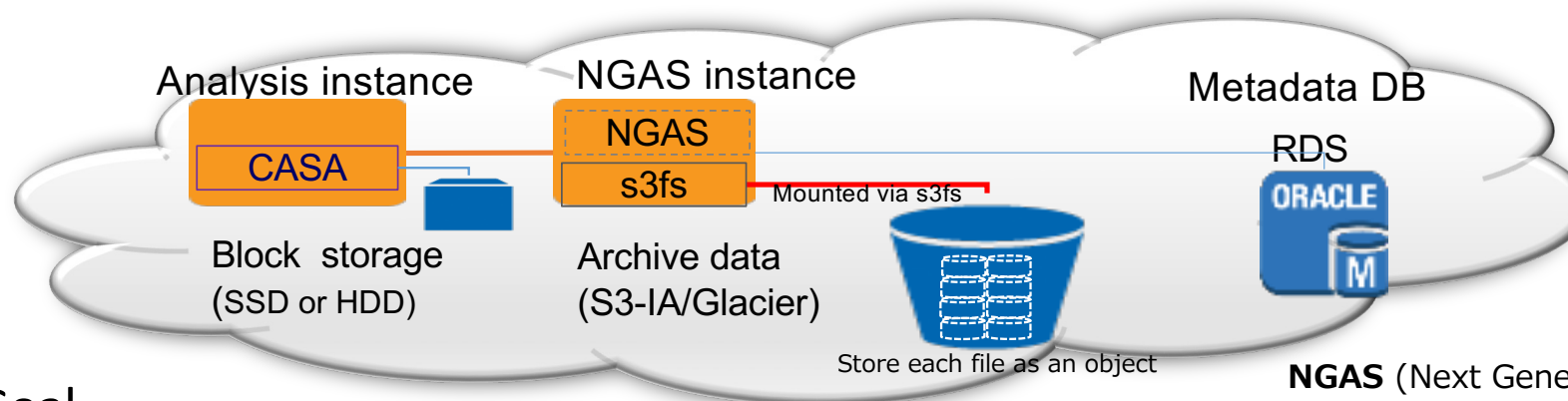
- ✓ storage system for long-term data archive
- ✓ computing system for data analysis



How can we utilize cloud service to solve the issues?

Archive and Analysis for Astronomical Data on Clouds

PoC for designing platforms for ALMA telescope data archive and analysis using **public clouds** (NAOJ and NII)



NGAS (Next Generation Archive System):
Archive system for ALMA telescope data

- **Goal**

Acquire practical information for:

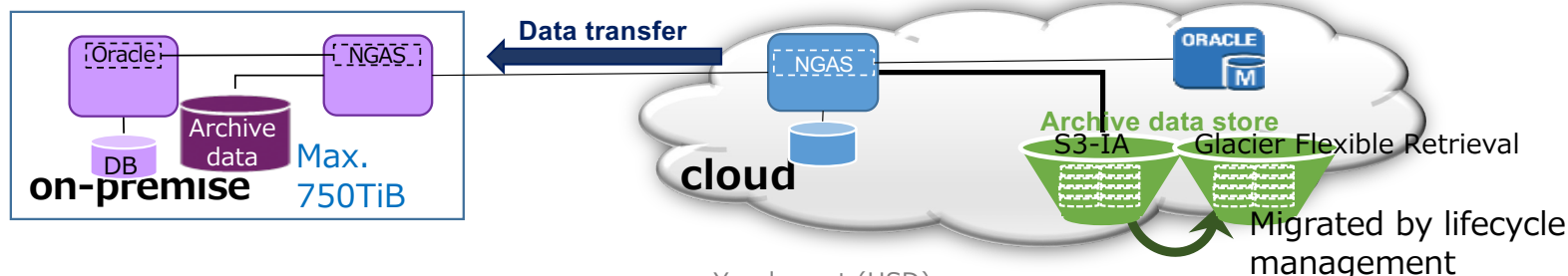
- Making decisions about storing and analyzing research data in public cloud services
- Designing an overall hybrid cloud architecture

- **Evaluation**

- Cost of storing ALMA telescope data in cloud storage services (object storage and cold storage) for a long period
- Performance/Cost of analyses of ALMA telescope data using computational resources of public cloud services (VM instances and block storage services)

Cost Estimation of Hybrid Cloud for ALMA Archive System

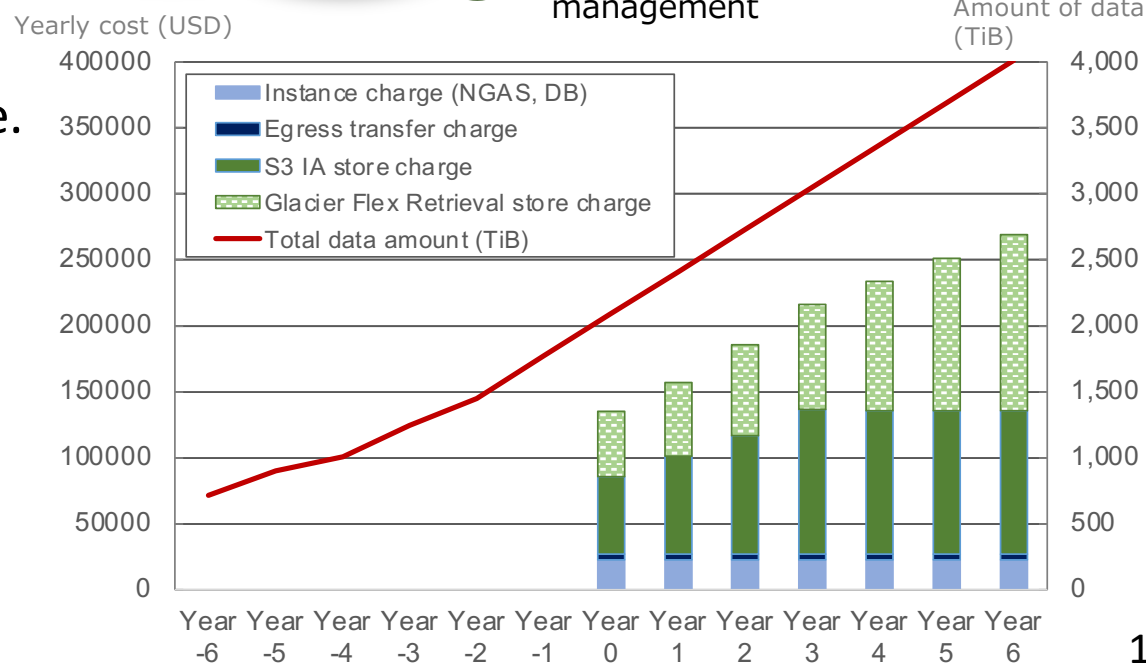
Tiered storage including Glacier, S3 IA, and on-premise storage



Tiered storage architecture restrains the extreme increase of the cloud storage charge.

[Assumptions derived from actual data usage statistics]

- Amount of on-premise storage: 960 TB.
- Amount of archive data: 2,092TiB in year 0; increases 320 TB/year.
- Datasets are stored in on-premise storage for 3 years, in S3 IA for 2 years, and then migrated to Glacier Flexible Retrieval
- Amount of downloads is 550 TB/year; 13% from S3 6% from Glacier Flex Retrieval

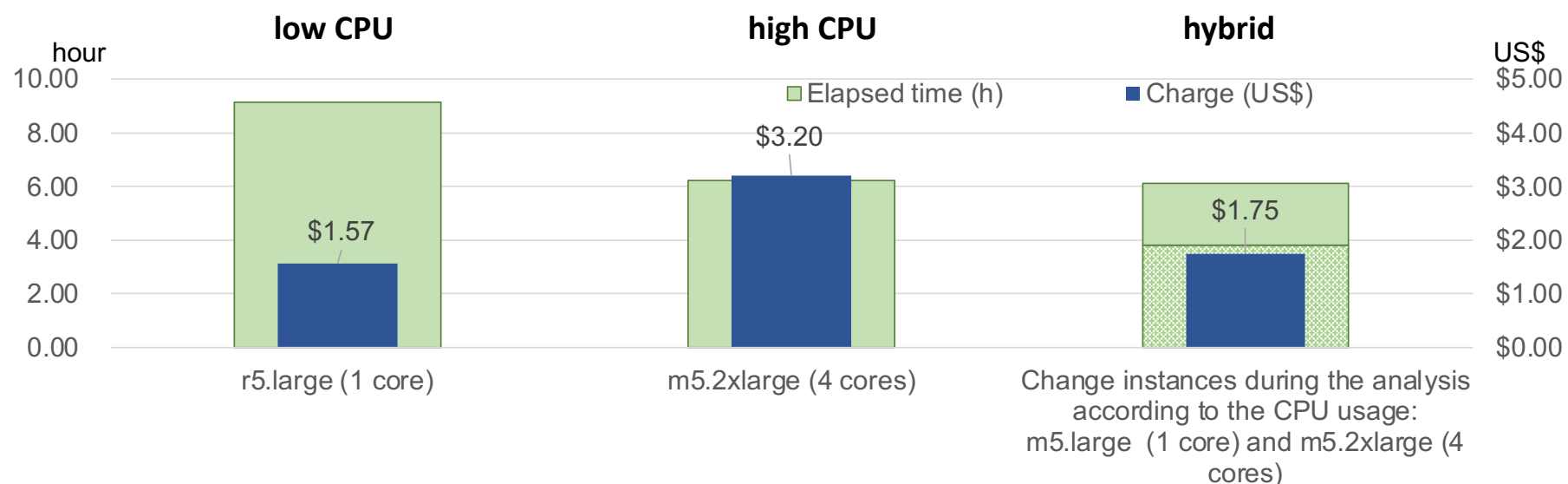


Performance and Cost Evaluation of ALMA

Observation Data Analyses

Observations on instance selection

- Instances with appropriate number of cores should be deployed according to the CPU usage characteristics of each analysis in order to assure analysis time and to save cloud charge.



Computing Resource Optimization for Data Analysis

Estimate optimal resources (instance types) through correlation analysis between observation parameters and required resources (CPU/memory)

- comparison of exec. time and cost for data analysis of 372 observation data-sets (7m ALMA antenna)
 - w/ resource optimization (choose instance type with minimum cpu/memory required by applications)
 - for calibration
 - c6i.large (1core/2vcpus/4GB) \$0.107/h or
 - m6i.large (1core/2vcpus/8GB) \$0.124/h or
 - r6i.large (1core/2vcpus/16GB) \$0.152/h
 - for imaging
 - c6i.xlarge (2cores/4vcpus/8GB) \$0.214/h or
 - c6i.2xlarge (4cores/8vcpus/16GB) \$0.428/h
 - w/ resource optimization (choose instance type with rich cpu/memory)
 - r6i.2xlarge (4cores/8vcpus/64GB) \$0.608/h



Conclusions

- NII promotes the creation of a state-of-the-art academic-information infrastructure that is essential to research and education within the broader academic community.
- GakuNin Cloud service provides universities and research institutions with information and consulting services to support the adoption of clouds, as well as tools to support the use of clouds for education and research, and the development of wide-area data collection and analysis programs.
- As the support for advanced utilization of clouds, NII collaborate with NAOJ for running PoC of Archive and Analysis of Astronomical Observation Data.