PROCEEDINGS OF SCIENCE

Hybrid Cloud Architecture for Astronomical Observation Data - Case Studies on ALMA Data Archive and Analysis

Hiroshi Yoshida[a]*, Kento Aida[a], George Kosugi[b], Eisuke Morita[b], Takeshi Nakazato[b], Makoto Shizugami[b], Renaud Miel[b], Hiroaki Tagawa[b], and Kanako Sugimoto[b]

a Center for Cloud Research and Development, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan

b National Astronomical Observatory of Japan, 2-21-1, Osawa, Mitaka, Tokyo, Japan

E-mail: h-yoshida@nii.ac.jp, aida@nii.ac.jp, george.kosugi@nao.ac.jp, eisuke.morita@nao.ac.jp, takeshi.nakazato@nao.ac.jp, m.shizugami@nao.ac.jp, renaud.miel@nao.ac.jp, hiroaki.tagawa@nao.ac.jp, kanako.sugimoto@nao.ac.jp

Adopting commercial cloud services for scientific research will reduce the total cost of ownership and enable access to state-of-the-art technologies. However, a methodology for designing an optimal hybrid architecture is currently lacking. We conducted case studies to demonstrate best practices for hybrid cloud architecture for archiving and analyzing ALMA telescope data. We focused on two key issues in developing a hybrid cloud design methodology: a tiered storage spanning on-premises and cloud environments, and the optimal selection of cloud server instances. Regarding the tiered storage, we estimated the storage cost for ALMA data using actual statistics, including data capacity growth and the relationship between data age and access frequency. The hybrid tiered storage reduces overall costs while mitigating the latency associated with cloud cold storage. For cloud server instance selection, we developed an optimization method that uses machine learning models to estimate required resources based on observational metadata such as telescope resolution, the number of antennas used in the observation, and data size. We also developed a workflow that dynamically switches to optimal instances via the cloud API. Our experimental results indicate a 60% cost reduction with a slight increase in execution time compared to the unoptimized execution, particularly for ALMA Compact Array observation data.

International Symposium on Grids and Clouds (ISGC2026)

15-20, March, 2026
Academia Sinica, Taipei, Taiwan

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0)
All rights for text and data mining, AI training, and similar technologies for commercial purposes, are reserved.
ISSN 1824-8039 . Published by SISSA Medialab.

PoS(ISGC2026)015

1. Introduction

An astronomical observatory requires not only state-of-the-art telescopes but also robust computing infrastructure to archive and analyze the vast amounts of astronomical observation data. For example, as of early 2026, the Atacama Large Millimeter/submillimeter Array (ALMA) radio telescope archive has surpassed 2 petabytes of total data and continues to expand. This growth creates a dual burden on research institutions: the need for vast storage and the computing resources required to process it. Consequently, optimizing the operation of these computing resources is crucial to curb rising operating costs.

While many institutions maintain robust on-premises infrastructure, there is growing anticipation for integrating commercial cloud services into research workflows. This hybrid approach offers several advantages. First, it allows researchers to utilize the latest generations of CPUs and GPUs immediately, avoiding the lengthy cycles of local hardware procurement and refreshes. Secondly, the cloud enables rapid scale-out during periods of peak demand ("cloud-bursting"). The final advantage is that organizations can reduce the Total Cost of Ownership (TCO) by shifting certain workloads to the cloud. This includes saving on the indirect costs of maintaining on-premises systems, such as electricity for cooling, physical facility space, and the specialized workforce required for hardware maintenance.

Despite these benefits, the transition to a hybrid model presents significant challenges. Currently, there is a lack of methodology for designing an architecture that optimally balances local and cloud resources. Furthermore, the scientific community often views unpredictable and uncontrollable cloud costs as a major obstacle, leading to hesitation to fully adopt these technologies.

To address these obstacles, the National Astronomical Observatory of Japan (NAOJ) and the National Institute of Informatics (NII) have been conducting collaborative research [1) 2) 3)]. NAOJ is currently exploring a hybrid cloud architecture and is looking to augment its existing on-premises systems with commercial cloud services to better handle massive astronomical data. On the other hand, NII has been helping Japanese universities and research institutes adopt cloud computing, so NII needs practical case experiences on adopting cloud services for scientific research.

By porting real ALMA data and its associated application software to a commercial cloud (Amazon Web Services, AWS), the team aims to establish practical and cost-effective guidelines. This research seeks to transform cloud computing from a source of financial uncertainty into a predictable, scalable resource for the astronomical community.

2. Hybrid Cloud Architecture and Design

Figure 1 shows the hybrid cloud architecture developed for this research. It comprises NAOJ's on-premises system and additional commercial cloud services, bridging the gap between fixed on-premises resources and the commercial cloud's elastic capabilities. The system is built to handle the ALMA-specific workflows, including long-term archiving and intensive computational analysis.

2.1 System components and integration

As depicted in Figure 1, the major components of the system are:

PoS(ISGC2026)015

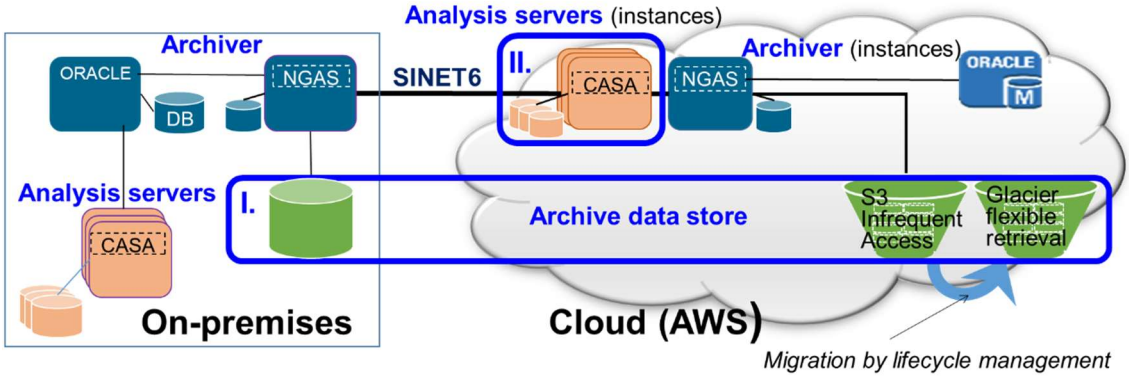

Figure 1: The Architecture of the hybrid cloud for archiving and analysis of ALMA data

- **Archiver:** This component manages the ALMA observation data and analysis result data and helps researchers search and retrieve that data. It consists of a server running an HTTP-based object storage system, NGAS (The Next Generation Archive System), and an Oracle RDBMS, which serves as a metadata repository.
- **Archive data store:** While the on-premises environment consists of a single-tier storage, the cloud environments utilize a tiered storage configuration using multiple types of object storage services, ranging from hot storage services to cold storage services, to balance performance and cost.
- **Analysis servers:** On these on-premises servers and cloud server instances, the CASA (Common Astronomy Software Applications) packages [4], the field-based open-source software for processing radio telescope data, are running.

Both the on-premises and cloud portions of the hybrid system include all those components. This symmetry allows for seamless data movement and workflow execution.

The NAOJ's local site and the cloud datacenters are connected via SINET6 [5]. SINET6 is a high-speed national research and education network backbone operated by NII, specifically designed for the Japanese research community. It provides rapid data transfer, which is essential for archiving and analyzing large astronomical datasets, as well as secure communication via L2VPN. It also provides direct connection services to cloud service providers' data centers, enabling those advantages even when using commercial cloud services.

2.2 Key considerations in developing a hybrid cloud design methodology

In the current ALMA on-premises environment, storage and servers consume the vast majority of resources. To manage those resources and consider leveraging cloud computing, NAOJ has identified several key issues. First, the petabyte-scale ALMA data archive demands extensive management effort. Second, there is currently no established methodology for determining the optimal scale of server instances during cloud migration. While on-premises systems are typically provisioned for peak loads, a direct "lift-to-cloud" approach often results in prohibitive costs due to underutilized capacity. Conversely, undersized instances risk prolonged execution times or, in extreme cases, process failures. Recognizing this, we developed a hybrid cloud design methodology specifically targeted at optimizing these two areas. To validate this approach, we conducted two case studies.

PoS(ISGC2026)015

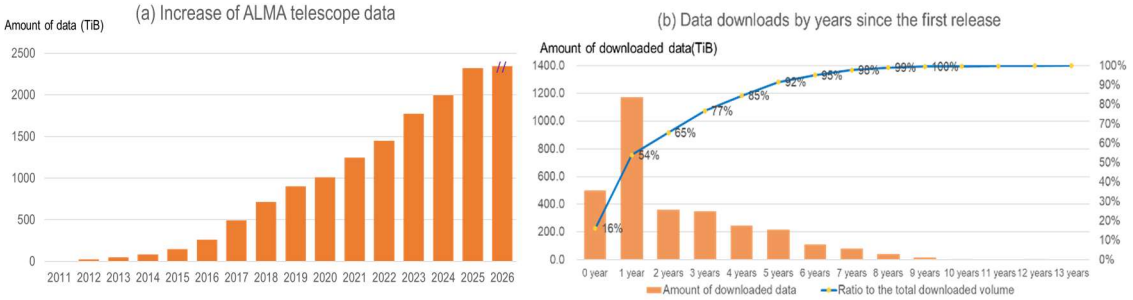

Figure 2: Statistics of the ALMA archive data

I. Data archiving in the tiered storage across on-premises and multiple cloud environments

We analyzed the costs associated with moving ALMA's petabyte-scale data archives to a tiered storage model. This configuration bridges the gap between on-premises hardware and multiple cloud environments, ensuring that data is stored in the most cost-effective location based on access frequency while balancing retrieval performance.

II. Selection of optimal cloud server instances for performing analyses in the cloud

We focused on identifying the optimal cloud server instances for data analysis. By moving computational workloads to the cloud, we aimed to select server instances that offer the best balance of processing power and cost efficiency relative to maintaining static on-site servers.

3. Case Study of the Tiered Storage and Cost Estimation for ALMA Data Archiving

The first case study focuses on storing the ALMA archive using a hybrid tiered approach. The goal of this study is to prevent storage costs from scaling linearly with total data volume while maintaining efficient access for scientists.

3.1 Analysis of data access patterns

Figure 2(a) shows the overall growth of the ALMA telescope archive. Currently, the archive contains more than 2 petabytes of data and has been growing by more than 300 terabytes per year. To design an effective tiered system, we first established a realistic model of how astronomical data is utilized over time.

Figure 2(b) shows the ALMA download statistics as of February 2026. By analyzing the statistics, we identified a clear decline in data access frequency over time. In the first 4 years after the data release, the data is accessed most frequently. Four years after the release, approximately 23% of the data is still being downloaded. Six years after the release, the download rate has dropped significantly to just 8%.

This clear transition from "hot" to "cold" data supports migrating older datasets to lower-cost storage tiers without significantly impacting most user requests.

3.2 Three-tiered storage configuration

Based on the access statistics, we translated the analysis results into a three-tiered simulation model, as shown in Table 1. The idea is to store frequently accessed data on local (on-premises) storage to maintain access performance and store infrequently accessed data in cloud storage. This configuration leverages the current AWS offerings [6].

PoS(ISGC2026)015

Table 1: Three-tiered storage configuration across the on-premises and cloud environment

Tier	Infrastructure		Characteristics
Tier 1	On-Premises Storage	Years 0–4	Immediate access on-premises with no transfer costs or egress delays
Tier 2	AWS S3 Infrequent Access	Years 5–6	Immediate access in the cloud (data transfer from the cloud is required), but retrieval and egress charges apply.
Tier 3	AWS Glacier Flexible Retrieval	Year 7+	Lowest storage cost, but requires a 3-hour restoration time before access.

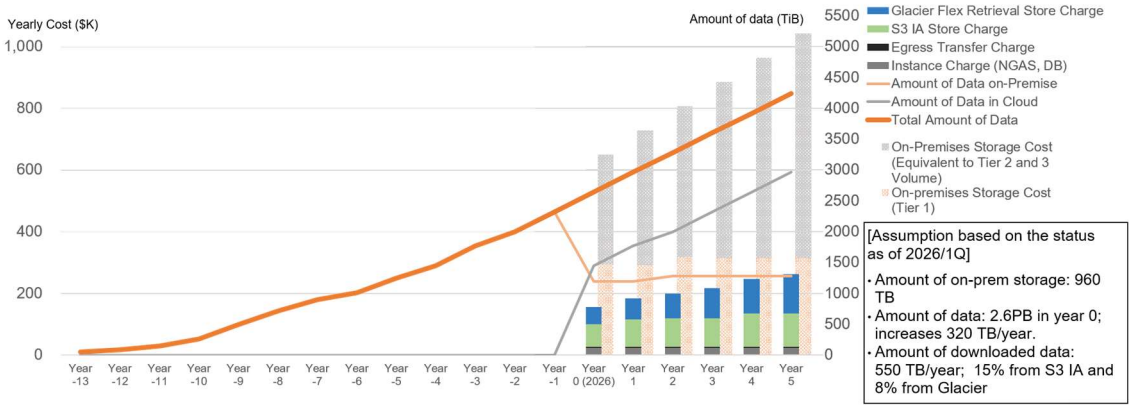

Figure 3: Estimation of the yearly cloud cost

To maintain efficiency, the cloud lifecycle management service automates migrations between Tier 2 and Tier 3 based on data age [7]. This service automatically migrates data objects from one cloud storage class to another based on predetermined policies that specify the source (Tier 2) and destination (Tier 3) storage classes, as well as migration criteria, such as the elapsed time since the data objects' creation date.

3.3 Cost estimation results

Using data growth statistics and estimates, we simulated the following scenario: before year 0 (assumed to be the current year, 2026), the ALMA archive operated only on the on-premises system, and then switched to a hybrid operation in year 0. Figure 3 shows the growth in data volume and the estimated annual commercial cloud cost from the simulation (represented by the left bar for each year). The simulation incorporated not only the base storage price but also associated costs based on our actual porting experience. These include running costs for NGAS archiver instances, fully managed Oracle database service fees, data retrieval fees, and egress transfer charges. The simulation shows the following results:

- By utilizing a hybrid model, investment in on-premises hardware can be capped based on the aggregate data volume, incorporating actual growth from the past three years alongside the current year's forecast, rather than constantly expanding local data centers to keep up with the data growth (in the simulation, 300TB/year).

PoS(ISGC2026)015

- Through the use of multiple tiers of cloud storage services (AWS S3 Infrequent Access and Glacier Flexible Retrieval), cloud costs do not scale linearly with total data growth, preventing "budget explosions".

From a performance standpoint, Tier 3 is low-cost cloud cold storage using AWS Glacier Flexible Retrieval, with a long restoration time before access (e.g., 3 hours in our actual measurements). Since our tiering strategy retains most of the active data (from the first six years) in on-premises or cloud storage without retrieval delays, the archive system ensures that the majority of users experience no retrieval delays, effectively mitigating the 3-hour restoration time required for cloud cold storage.

For reference, we also estimated the annual cost assuming the entire system is deployed on-premises (represented by the right bar for each year). Because actual hardware expenses are highly volatile due to quotation specifications, foreign exchange rate fluctuations, and recent memory cost spikes, we adopt a working assumption of $0.02/GB/month, taking commercial quotes as the primary reference. This cost per GB-month is lower than that of AWS S3 Standard but higher than those of the AWS cold storage services adopted in Tier 2 and Tier 3. This estimate includes only the required storage appliances, software, and their 5-year maintenance fee (capital expenditure, CAPEX) and excludes power and facility costs, as well as operational workload costs (operational expenditure, OPEX); therefore, actual on-premises deployment must incorporate OPEX. Additionally, actual costs for the hybrid cloud configuration are shown as the sum of the left bar and the Tier 1 portion of the right bar (excluding on-premises OPEX) for each year.

Ultimately, this case study shows that a structured hybrid architecture mitigates the financial issues associated with data growth in astronomy.

4. Case Study of Optimal Cloud Server Selection for ALMA Data Analysis

The second case study addresses the challenge of deploying cost-effective computing resources for data analysis. Unlike storage, which is relatively static, computing requirements for astronomical data vary widely depending on the datasets to be analyzed.

4.1 The complexity of server instance selection

Selecting the server instances requires considerable domain knowledge of both the application software and the available service instances. Selecting the right cloud instance is a complex task that requires bridging the gap between those two domains.

On the application software side, we must understand the specific resource needs of our CASA package. The CASA analysis workflow is split into two phases: Calibration, which primarily uses only a single core, and Imaging, which is much more resource-intensive and requires multiple cores and significantly more memory. Furthermore, these resource requirements are not fixed. They vary based on observation settings and the size of the dataset being processed.

On the cloud side, we have to navigate the service catalog. Instance prices are usually determined by the number of cores and memory size [8]. However, cloud providers offer these in set combinations, usually multiples of 2 to the n-th power [9]. For example, the memory size of a 1-physical-core instance is typically 4GB, 8GB, or 16GB, and when more memory is required, the number of cores has to be increased.

PoS(ISGC2026)015

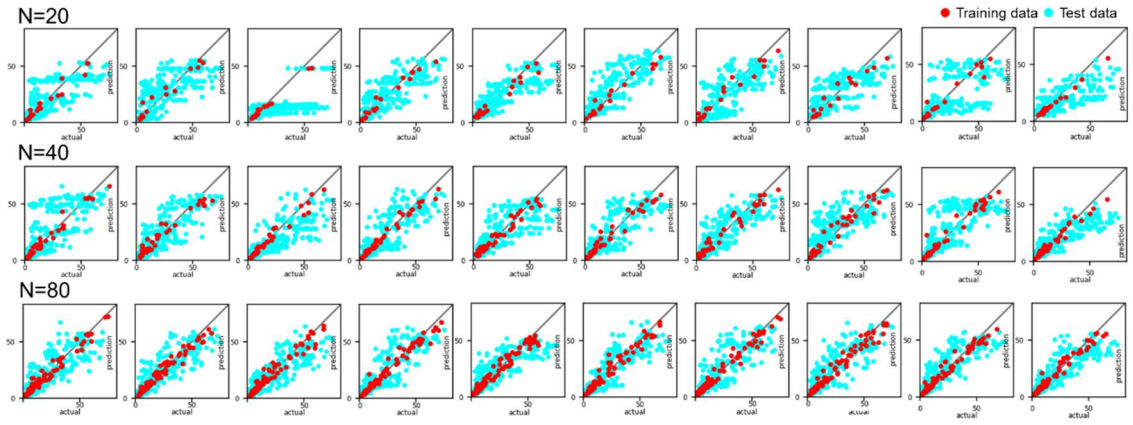

Figure 4: Actual (measured) vs. predicted CPU usage (N samples, 10 trials)

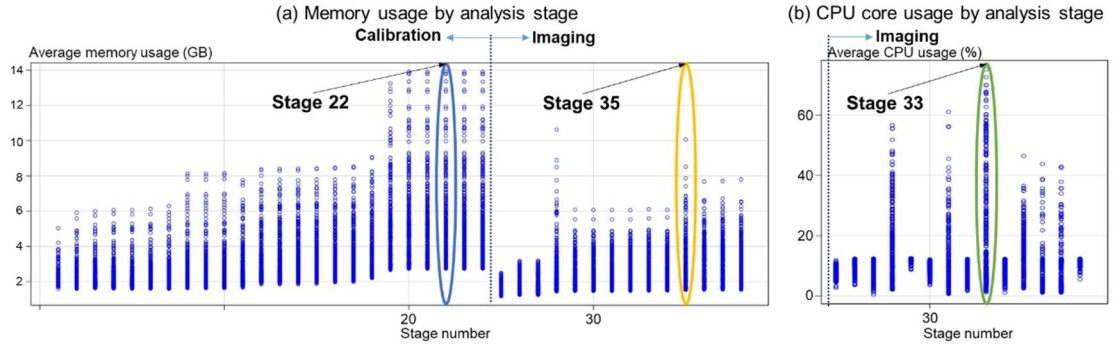

Figure 5: Resource usage by analysis stage

4.2 A three-step methodology for instance selection

To solve this problem, we developed a three-step methodology for selecting optimal server instances on commercial cloud services based on the resource usage prediction. Since the server instance type and price are determined by the number of cores and memory size, the following resource usage parameters must be estimated:

- Required average memory size for Calibration phase
- Required average memory size for Imaging phase
- Required average number of CPU cores for Imaging phase

(Calibration is excluded because it always requires a single core, as mentioned earlier)

The following three steps determine instances.

4.2.1 Step 1: Gathering training data

Resource usage metrics, including CPU and memory consumption, are collected from well-resourced server instances. Since these well-resourced instances are usually over-provisioned, determining how much training data to collect is a cost concern. Figure 4 presents our experimental results, showing the relationship between actual and predicted resource usage for training data sizes of N=20, 40, and 80. The red dots represent training data randomly selected in each trial, while the blue dots represent the remaining collected data (test data). These experimental results demonstrate that collecting 40 resource-usage samples is sufficient to train predictive models accurately.

PoS(ISGC2026)015

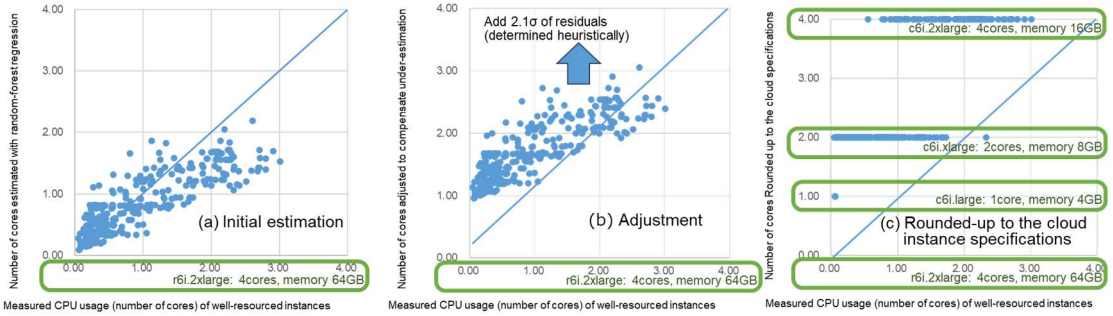

Figure 6: Estimation of the number of cores for Imaging using random forest regression and adjustment

4.2.2 Step 2: Machine learning model creation and initial estimation

We identify the most resource-intensive stages of Calibration and Imaging and heuristically apply a specific regression algorithm of machine learning to these stages. Figure 5 shows how these processes are carried out in practice. These charts show the distribution of average usage for memory and CPU cores by processing stages.

Regarding Calibration, stage 22 was identified as the step with the highest memory usage. We used scikit-learn's linear regression to estimate the required average memory size. Similarly, for Imaging, stage 35 exhibited the highest memory consumption, and we applied linear regression to estimate its average memory size. Additionally, stage 33 was found to have the highest CPU core usage in Imaging. Since random forest regression performed better for CPU core estimation, we utilized scikit-learn's Random Forest Regressor to estimate the average core requirement.

As candidate explanatory variables for estimation, we initially consider the following observational metadata and dataset attributes: Nms (number of measurement sets in the dataset), Nfield (number of observed fields), Nspw (number of spectral windows), TNchan (total number of spectral channels), TNrow (total number of spectra), Ninteg (number of integrations), aveR (average value of spectral resolutions), MOUSSize (size of Member Observation Unit Set), NumberofAntennas (number of antennas used for the interferometric observation), MOSAIC (whether the observation is mosaic or not), and storage_raw (dataset size on the working disk space). To avoid overcomplicating the model while preserving predictive accuracy, we select optimal subsets of explanatory variables by minimizing the Akaike Information Criterion (AIC) [10]. Consequently, the calibration memory model incorporates Nms, TNchan, TNrow, Ninteg, MOUSSize, and storage_raw; the imaging memory model uses Nfield, TNchan, MOUSSize, NumberofAntennas, MOSAIC, and storage_raw; and the imaging CPU usage model relies on Nfield, MOUSSize, and MOSAIC. These models provide the initial estimates.

4.2.3 Step 3: Instance selection

Initial estimates are adjusted to compensate for statistical bias (underestimation). This adjustment is done by adding the product of the standard deviation of prediction residuals (σ) and a heuristic coefficient (k) to the estimate. Finally, these values are rounded up to the nearest available cloud instance specifications (e.g., 4GB, 8GB, 16GB for memory).

PoS(ISGC2026)015

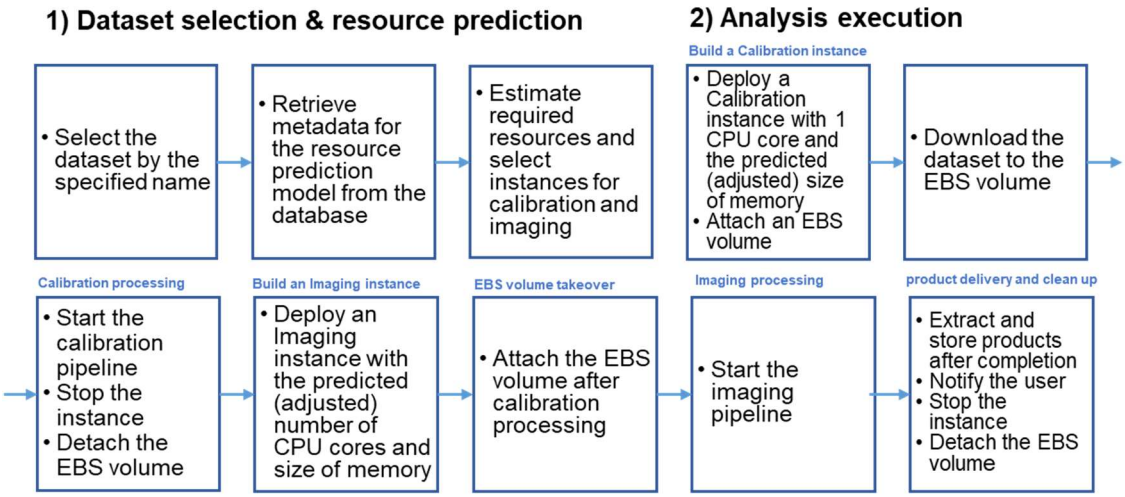

Figure 7: End-to-end workflow of ALMA data analysis

As an example, the process of selecting instances based on CPU core usage is shown in Figure 6. Figure 6(a) shows the relationship between actual (measured) CPU usages and the model's initial estimates, as explained in Step 2. About half of the initial estimates are underestimates. Figure 6(b) shows the result of the adjustment, and Figure 6(c) shows the results of the final round-up of the number of CPU cores to 1 core, 2 cores, or 4 cores. In Figure 6(b), $k\sigma$ is added to each initial estimate, with k = 2.1, chosen heuristically to ensure that the ratio of resource-shortage samples in Figure 6(c) falls below the predetermined 5% criterion in this trial.

This systematic rounding ensures we are not paying for a 'well-resourced' server that is 2 to 8 times the size we actually need, while still maintaining the performance required for astronomical data processing.

4.3Automated end-to-end workflow

This server instance selection logic is integrated into an automated analysis workflow. To make this human-readable and easy to maintain, the entire process is described in a Jupyter Notebook using the Literate Computing for Reproducible Infrastructure (LC4RI) technology developed by NII [11) 12)].

Figure 7 shows the overall workflow. The workflow begins with dataset selection and resource prediction. The user specifies the dataset name. The system then retrieves metadata from the archive database. Observation setting parameters are included in that metadata. Then the system selects the server instances for Calibration and Imaging based on the resource predictions we discussed in the previous section.

The next part of the workflow is executing the analysis. The system builds a Calibration server instance with 1 physical core and the predicted and adjusted memory size, attaching a block storage volume (Elastic Block Store, EBS). After the dataset is downloaded into the EBS volume and Calibration is completed, the system switches instances. It transitions from the 1-core Calibration setup to a multi-core Imaging setup, utilizing native cloud APIs. Once the analyzed products are saved to object storage, the instance and EBS volume are immediately released to prevent further charges.

PoS(ISGC2026)015

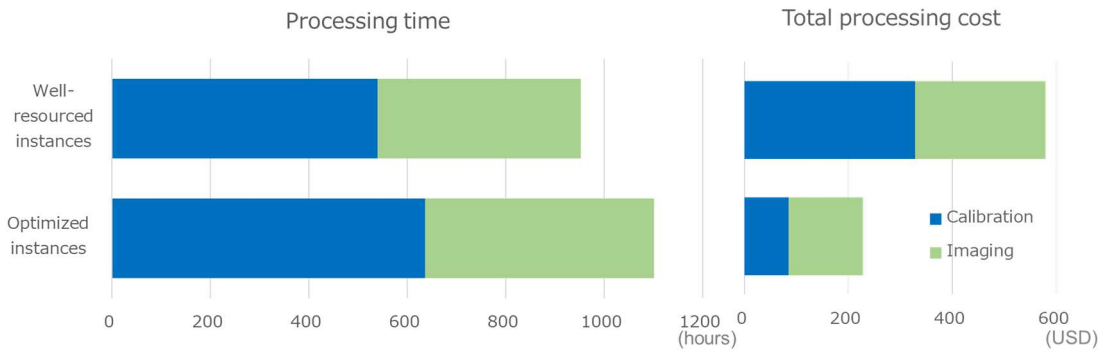

Figure 8: Comparison of the processing time and total processing cost of 372 analyses of the Compact Array data between optimized instances and well-resourced instances

In a commercial cloud, workflows can be executed on demand and in parallel, allowing the cloud to be used effectively in response to the load on on-premises computing resources.

4.4Analysis of experimental outcomes

The effectiveness of the proposed methodology was validated through a series of trials of analyses conducted on Amazon Web Services (AWS). The ALMA telescope consists of 4 12-meter antennas (Total Power Array), 12 7-meter antennas (Compact Array), and 50 12-meter antennas (Main Array). As a first step toward establishing a model, we focused our initial experiments on the Compact Array datasets because the analysis execution times are relatively short and their performance is primarily driven by CPU and memory usage, providing a controlled environment for testing our instance selection methodology.

The trial consisted of 372 actual analyses performed on the cloud. Figure 8 shows the total processing time and cost of those 372 analyses. By comparing our optimized methodology against a traditional "well-resourced" instance approach, where a single high-spec server instance is used for the entire workflow, the following results were achieved:

- The optimized server instance selection and dynamic switching of instances resulted in 60% reduction in costs.
- This cost saving was achieved with a 15% increase in total processing time, as a performance trade-off.

The well-resourced approach is essentially identical to a "lift-to-cloud" strategy for on-premises servers. Our results demonstrate that our managed hybrid architecture significantly improves efficiency compared with the simple lift-to-cloud strategy.

5. Conclusion and Future Work

5.1 Summary of findings

The collaborative case studies between NAOJ and NII have demonstrated that a hybrid architecture improves the operational efficiency of computing and archive systems for astronomical observatories. While the cost of commercial cloud services remains a significant concern, our optimization methods reduce this barrier without compromising necessary performance.

PoS(ISGC2026)015

5.1.1 ALMA data archiving

Our tiered storage model spanning on-premises and multiple cloud environments enables controlled investment in on-premises hardware. Also, by using the cloud for older archived data, we prevent storage costs from scaling unpredictably and uncontrollably as data volume grows. Crucially, this strategy also mitigates the long wait times associated with restoring data from cloud cold storage.

5.1.2 Selection of optimal cloud server instances for analyses of ALMA data

Specifically for the Compact Array cases, our server instance selection based on machine learning models, along with a dynamic instance-switching analysis workflow, resulted in a 60% reduction in total cost compared to using unoptimized server instances. This significant saving was achieved with only a minor, 15% increase in execution time.

5.2 Future works

As we look toward the next phase of this framework, the goal is to refine and scale these results to meet the demands of an even broader range of observational data.

- **Refinement of estimation methodology**

  To increase the accuracy of our predictive models, we plan to incorporate a broader range of observational metadata and system operational data. This includes adding metrics such as required storage throughput and IOPS (Input/Output Operations Per Second) to ensure performance stability across a wider range of datasets.

- **Scaling to Main Array data**

  The current success with the 7-meter Compact Array data serves as a stepping stone for the 12-meter Main Array data, which presents a larger challenge. The Main Array data is typically an order of magnitude larger than the Compact Array's. Processing these datasets requires substantially more computing power and higher I/O performance. According to our primary experiments, server instance selection based on predictions of maximum (not average) CPU and memory usage, and enhancements to I/O performance (throughput and IOPS), are effective for several Main Array datasets.

- **Broadening application scope**

  Finally, the research aims to establish a generalized methodology based on these practical insights. The objective is to help other research areas beyond radio astronomy adopt similar hybrid approaches to manage and analyze petabyte-scale data efficiently. By making these best practices more broadly applicable, we hope to foster a more flexible and cost-effective environment for data-intensive science.

Acknowledgement

This research was supported by ROIS NII Open Collaborative Research 2022-(22S0202), 2023-(23S0207), 2024-(24S0106), and 2025-(251S1-22758).

PoS(ISGC2026)015

References

[1] George Kosugi, Eisuke Morita, Takeshi Nakazato, Renaud Miel, Hiroaki Tagawa, Kanako Sugimoto, Kento Aida, Hiroshi Yoshida, *Efficient use of commercial cloud for astronomical data archive and data analysis,* in Proceedings SPIE Astronomical Telescopes + Instrumentation: Software and Cyberinfrastructure for Astronomy, VIII 1310112 (July 2024), https://doi.org/10.1117/12.3019514, accessed on 27/04/2026.

[2] Hiroshi Yoshida, Kento Aida, George Kosugi, Eisuke Morita, Takeshi Nakazato, Yohei Hayashi, Renaud Miel, *Performance and Cost Evaluation of Public Cloud Cold Storage Services for Astronomy Data Archive and Analysis,* In presentations of *International Symposium on Grids & Clouds 2021 (ISGC2021)*, March 2021, https://indico4.twgrid.org/event/14/contributions/393/attachments/161/201/Performance_and_Cost_of_Cloud_Cold_Storage_for_Astronomy_Data_Archive_and_Analysis.pdf, accessed on 20/04/2026.

[3] Hiroshi Yoshida, *Experiments in Storing Scientific Research Data in Cloud Cold Storage Services,* In presentations of *Storage Developers Conference 2018*, https://www.snia.org/sites/default/files/SDC/2018/presentations/Cloud_Storage/Yoshida_Hiroshi_Experiments_in_Storing_Scientific_Research_Data_in_Cloud_Cold_Storage_Services.pdf, accessed on 20/04/2026.

[4] National Radio Astronomy Observatory, *Common Astronomy Software Applications,* https://casa.nrao.edu/, accessed on 20/04/2026.

[5] National Institute of Informatics, *Science Information Network SINET6,* https://www.sinet.ad.jp/en, accessed on 26/04/2026.

[6] Amazon Web Services, *Amazon S3 Storage classes,* https://aws.amazon.com/s3/storage-classes/, accessed on 20/04/2026.

[7] Amazon Web Services, *Transitioning objects using Amazon S3 Lifecycle,* https://docs.aws.amazon.com/AmazonS3/latest/userguide/lifecycle-transition-general-considerations.html, accessed on 27/04/2026.

[8] Amazon Web Services, *Amazon EC2 Pricing,* https://aws.amazon.com/ec2/pricing/, accessed on 20/04/2026.

[9] Amazon Web Services, *Amazon EC2 instance types,* https://aws.amazon.com/ec2/instance-types/, accessed on 20/04/2026.

[10] Hirotsugu Akaike, *Information theory and an extension of the maximum likelihood principle,* In Proceedings of the 2nd International Symposium on Information Theory, 267-281 (1973).

[11] National Institutes of Informatics, *Literate Computing for Reproducible Infrastructure*, https://ccrd.nii.ac.jp/events/sc/sc24/docs/9_SC24_Literate_Computing-v1.2_A4.pdf, accessed on 20/04/2026.

[12] National Institutes of Informatics, *Literate Computing for Reproducible Infrastructure*, https://literate-computing.github.io/, accessed on 20/04/2026.

PoS(ISGC2026)015