

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

Thursday, 19 March 2026 11:44 (22 minutes)

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. Consequently, optimizing the operation of these computing systems is a crucial issue. Adopting public cloud services is expected to reduce the Total Cost of Ownership (TCO) and allow the use of cutting-edge technologies (e.g., the latest GPUs). However, a systematic methodology for designing an optimal hybrid architecture to fully realize these advantages is currently lacking.

The National Astronomical Observatory of Japan and the National Institute of Informatics have been conducting case studies to demonstrate best practices for designing and implementing a hybrid cloud architecture dedicated to storing and analyzing ALMA radio telescope data. We collected and analyzed both system operation data (e.g., storage usage, analysis execution time) and application data (e.g., observed data from ALMA 12m and 7m antennas) to conduct two primary experiments.

First, we estimated the storage cost for petabyte-scale ALMA observation data, which requires high-speed access for hot data (frequently accessed data for analysis) and low-cost, long-term archiving for cold data (infrequently accessed data), based on the actual statistics such as data capacity increase of each year and the relationship between data age and access frequency. Although the cost of public cloud storage is often considered a barrier, our studies demonstrated that a three-tier storage architecture, comprising on-premises storage, cloud object storage, and cloud cold storage, with optimized data life-cycle management, reduces the overall cost.

Second, we developed an optimization method for selecting optimal server instances on public cloud services. Choosing the right instances requires considerable domain knowledge of both the application programs and the available service instances. Our method utilizes machine learning models to estimate the required CPU performance and memory capacity for ALMA data analysis based on the metadata of observations, such as the telescope resolution and data size. Crucially, the method adaptively chooses different server instances for different computing phases; for example, assigning a single-core instance for the calibration phase and switching to a multiple-core instance for the imaging phase. We also developed a procedure to execute the switching of those instances automatically by utilizing the cloud service API. Our experimental results indicate a 60% cost reduction accompanied by only a slight increase in execution time compared to the unoptimized execution, particularly for 7m antenna observation data.

Our case studies demonstrate that a hybrid architecture improves the operational efficiency of computing systems for astronomical observatories. While the cost of public cloud services remains a significant concern, our optimization methods reduce this barrier while preserving the necessary performance. Our presentation will detail the systematic design, methodologies, and quantitative experimental results that validate our findings.

Primary author: YOSHIDA, Hiroshi (National Institute of Informatics)

Co-authors: AIDA, Kento (National Institute of Informatics); KOSUGI, George (National Astronomical Observatory of Japan); MORITA, Eisuke (National Astronomical Observatory of Japan); NAKAZATO, Takeshi (National Astronomical Observatory of Japan); SHIZUGAMI, Makoto (National Astronomical Observatory of Japan); MIEL, Renaud (National Astronomical Observatory of Japan); TAGAWA, Hiroaki (National Astronomical Observatory of Japan); SUGIMOTO, Kanako (National Astronomical Observatory of Japan)

Presenter: YOSHIDA, Hiroshi (National Institute of Informatics)

Session Classification: Infrastructure Clouds and Visualisations - I (11:00 - 12:40)

Track Classification: Track 8: Infrastructure Clouds and Virtualizations