

Case Studies of Hybrid Cloud Architectures for Astronomical Observatory Data

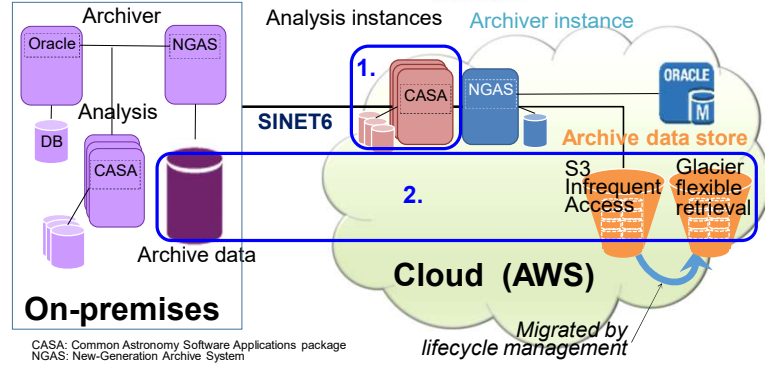
Adopting public cloud services for scientific research will reduce the total cost of ownership, allow the use of state-of-the-art technologies (e.g., the latest GPUs), and establish BCP. However, no methodology exists for designing an optimal architecture to realize these advantages. We have been conducting case studies of storing and analyzing ALMA radio telescope data in public cloud services in collaboration with the National Astronomical Observatory of Japan (NAOJ) to demonstrate the best practices and discuss the design of a suitable architecture using cloud services.

Case Study Overview

NAOJ is considering a hybrid cloud architecture comprising its on-premises system and additional public cloud services. To establish the criteria for optimal data allocation between the cloud and on-premises system, we ported the ALMA data and analysis/archiver software to public cloud services to investigate the cost-effective usage of computing resources and storage.



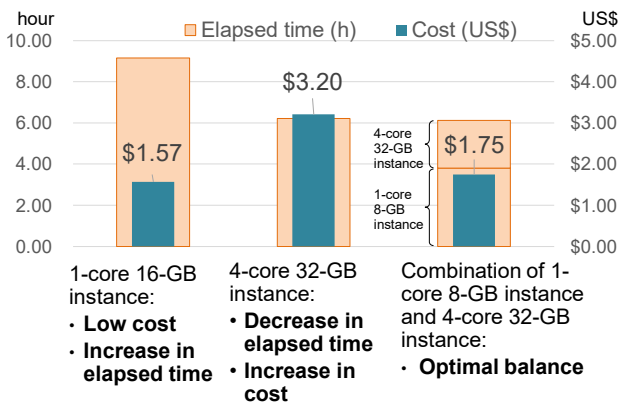
Source: NAOJ



Examples of Case Study Results

1. Selection of optimal server instances for analyses of ALMA data

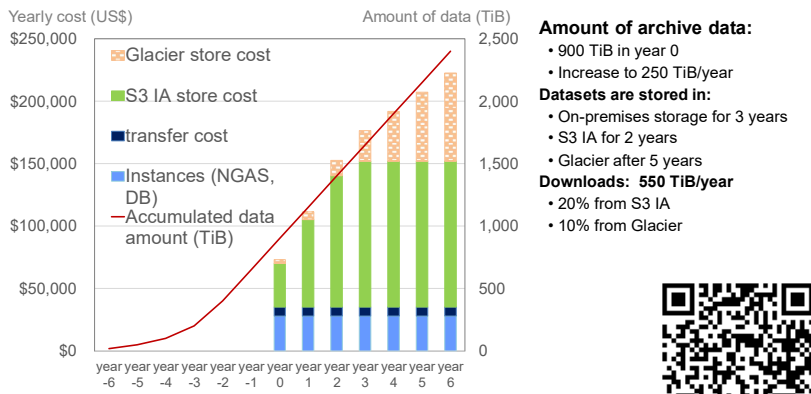
- We developed a technique for dynamically attaching instances with the optimal number of cores and RAM capacity according to the analysis stage.
- We established a resource estimation model based on the metadata (e.g., observation conditions) for the dataset to be analyzed.



2. Estimation of cost of storing ALMA archive data in tiered storage consisting of on-premises and cloud storage

The estimation results show the following advantages:

- Investment in on-premises storage can be kept constant.
- Using two cloud storage tiers (S3-IA and Glacier) inhibits the cloud cost from increasing in proportion to the total amount of data.
- The influence of the long restore time for Glacier (200 minutes) can be mitigated (limited to 10% of downloads).



Acknowledgment: We would like to thank the PoC members at NAOJ for providing data and support.