Verification of the use of Google Cloud Platform by calculating a real-time forecast model

Takeshi Terui[1], Yasushi Fujiwara[2] Takehiko Nose[2]
[1]*National Institute of Polar Research*
[2]*Graduate School of Frontier Sciences, The University of Tokyo*

When the forecast model run in real-time, a high-performance computing environment with stability and high-speed operation is used. The calculation time of the real-time forecast model constrains by the schedule because the calculation must be completed earlier than the next forcing data update. Therefore, it is crucial to confirm the start and end times of the model calculation and finish the calculation within that time. This time is called the time quota for the calculation. The schedule defines when the update time of the input data and the calculation result are output to clarify this quota, and the time available for the model is described. This time is used not only for the forecast model's calculation time but also for pre- and post-processing of the model, the extension of the unexpected calculation time, and the free time for the next processing. The demand for computer resources to meet the quota is determined by creating a test case based on the actual model and running a test run while changing the CPU architecture and number of cores. Thus, the real-time forecast model's schedule and the demand for the computing environment are determined, and we can estimate the cost of the production operation of the forecast model.

We delivered sea ice forecast data to R/V Mirai for the Arctic Sea observation cruise from Sep. to Oct. in 2020 (MR20-05C). We forecasted every day and used VENUS as a data transfer platform. Sea ice forecast data was output from sea ice model IcePOM. We were using ECMWF 10-day weather forecast data as input data for IcePOM, high-resolution sea ice forecast data around the observation area in MR20-05C was calculated. We used the Google Cloud Platform (GCP) for the calculation environment of this forecast model.

In general, researchers use supercomputers owned by research institutes or high-performance servers built-in laboratories for real-time forecasting models. The cruise plan of MR20-05C was reconsidered due to the impact of Covid-19. The delivery period and area of the forecast data decided just one month before the cruise. There was not enough time to create an on-premises environment for calculating the forecast model. Therefore, we have selected GCP because of the flexibility of setting about computing resources and the ease of constructing a calculation environment. Next, we built the computer environment demanded the calculation of the forecast model on the GCP. We connected this environment to VENUS, and the data delivered to R/V Mirai. Figure 1 showed this relation diagram of the entire system.

In this presentation, we would like to introduce the technical results obtained in these construction processes. In particular, we would like to report on the results of comparing on-premises with GCP. This study was a case of using GCP to run the real-time forecast model in the scientific field, and it introduces the design, construction, and operation. For researchers considering the use of cloud services, this case can expect to be beneficial.

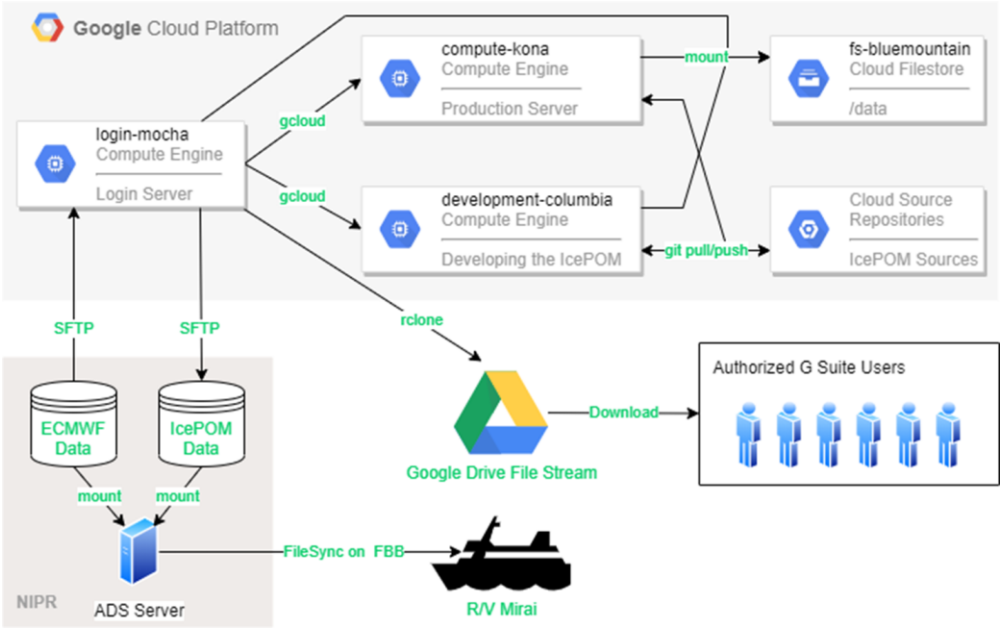

Figure 1. The diagram of each related component about the forecast model and delivery environment for MR20-05C.