Status report on the prototype system of the cloud-based online data analysis environment for HSC-SSP

Takahiro Morishima*,1,   Hisanori Furusawa[1], Michitaro Koike[1], George Kosugi[1], Sogo Mineo[1], Yuki Okura[1], Masayuki Tanaka[1],   and   The HSC Software Team

* author (speaker)
[1] NAOJ

You can see the electric version in the following URL :
https://y2.nao.ac.jp/index.php/s/637weT6sWDzCfLQ

Abstract

This is the status report of R&D on the cloud-based online data analysis environment for Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP). HSC-SSP is a large imaging survey of the sky aimed to address a wide array of astrophysical questions, and we have indeed made a large number of significant astrophysical discoveries so far. To further promote and support the HSC-SSP user's scientific activity, we plan to provide a cloud-based on-line data analysis environment. For this, we are building a prototype system with an open-source container orchestration system, Kubernetes, and an open-source web-based interactive compute & development environment, Jupyter Notebook/JupyterLab. On this integrated system, the user can execute various operations of data analysis with a web-based user interface on Jupyter Notebook/JupyterLab, where most of the HSC-SSP data access tools are pre-installed. These tools allow the user to search/request/retrieve the data, display/map the data, etc. Those use can be available anytime and anywhere if you want.

Discussion

key feature:
- ☑ UI is web-based user interface on Jupyter Notebook/JupyterLab.
- ☑ It does not depend on device and OS.
- ☑ Most of the HSC-SSP data access tools are pre-installed. (and optimized)
- ☑ It is not necessary to download HSC data to local storage.
- ☑ Not imaging data (s21a, s20a, etc) and also spectroscopic data (PFS) are available online, simultaneously.

coordination/future plan:
- ☑ We have carried out various evaluation and benchmark tests under realistic severe conditions.
- ☑ Secure authentication/authorization, security measures.
- ☑ Tests under more realistic severe conditions.
  (i.e. trial term by developer/tester will begin before long.)
- ☑ Extension considering other observed data in multi-band frequency
  (now on being planning with ADC members, etc)

We are actually building a prototype systems, and we have carried out various R&D tests. For more realistic tests under severe conditions, trial operation period by developer/tester will begin before long.