Contribution ID: **7**　　　　　　　　　　　　　　　　　　Type: **Oral Presentation**

# The HEPS Data and Computing System

*Thursday, 20 March 2025 17:02 (20 minutes)*

The 14 beamlines for the phase I of High Energy Photon Source(HEPS) will produces more than 300PB/year raw data. Efficiently storing, analyzing, and sharing this huge amount of data presents a significant challenge for HEPS.

HEPS Computing and Communication System(HEPSCC), also called HEPS Computing Center, is an essential work group responsible for the IT R&D and services for the facility, including IT infrastructure, network, computing, analysis software, data preservation and management, public services etc. Aimed at addressing the significant challenge of large data volume, HEPSCC has designed and established a network and computing system, making great progress over the past two years.

For the IT infrastructure, A deliciated and high-standard machine room, with about 900⬚ floor space for more than 120 high-density racks in total has been ready for production since this August. The design of the network utilizes RoCE technology and a spine-leaf architecture. The data center network's bandwidth can support speeds of up to 100Gb/s, fully meeting the demands of high- speed data exchange. To meet the requirements of data analysis scenarios for HEPS, a computing architecture is designed and deployed in three types, including Openstack, Kubernetes, and Slurm. Openstack integrates the virtual cloud desktop protocol to provide users with remote desktop access services, and supports users to use browsers to access windows/Linux desktop, running commercial visualization data analysis software. Kubernetes manages container clusters, and starts multiple methodological container images according to user analysis requirements. Slurm is used to support HPC computing services and meet users' offline data analysis needs.

Additionally, HEPSCC designed and developed two softwares for the data management and analysis, DOMAS and Daisy. DOMAS (Data Organization, Management and Accessing Software stack), which is aimed for automating the organization, transfer, storage, distribution and sharing of the scientific data for HEPS experiments, provides the features and functions for metadata catalogue, metadata ingestor, data transfer, data web portal. Daisy (Data Analysis Integrated Software System) is a data analysis software framework with a highly modular C++/Python architecture. Some online data analysis algorithms developed by HEPS beamlines have been integrated into Daisy successfully most of which were validated at the beamlines of BSRF (Beijing Synchrotron Radiation Facility) for the real-time data processing. Other data analysis algorithms/software will be continuously integrated to the framework in the future.

This year, the data and computing system has been deployed at HEPS Campus (Huairou District, Beijing). The integration and the verification of the whole system at HEPS were finished and achieved great success.

**Primary author:**　HU, Hao (Institute of High Energy Physics)

**Presenter:**　HU, Hao (Institute of High Energy Physics)

**Session Classification:**　Data Management & Big Data

**Track Classification:**　Track 6: Data Management & Big Data