

Design and Implementation of an IO Framework for HEPS (Remote Presentation)

Thursday, March 28, 2024 11:00 AM (20 minutes)

HEPS (High Energy Photon Source) is expected to generate a massive and diverse amount of data, and the data IO bottleneck severely affects computational efficiency. In order to address these issues, we have designed and implemented an IO framework specifically for HEPS, serving as the data IO module of Daisy (Data Analysis Integrated Software System), which is a software framework developed for HEPS. Firstly, to solve the problem of diverse data formats, based on the analysis of HEPS scientific tasks, we provide a unified IO interface to the computation, effectively shielding the underlying data format differences. Secondly, to improve the batch processing speed, we parallelize the IO read and write operations based on the characteristics of different data formats. Additionally, we design a prefetching strategy to asynchronously read the data required for subsequent computations into memory, further reducing the time-consuming data IO in the computation process through pipeline technology. Lastly, we introduce streaming data IO to avoid the IO bottleneck caused by writing data to disk and then reading it back. Moreover, we design an online data repository based on distributed memory, providing two forms of support for real-time online processing: one is data-triggered computation, where data is processed as it arrives, and the processing methods are compatible with those registered in Daisy; the other is remote reading the streaming data for computation, retrieving data from the online data repository.

Overall, our proposed IO framework addresses the challenges posed by the massive and diverse data generated by HEPS, significantly improving the computational efficiency and providing support for real-time online processing.

Primary authors: FU, Shiyuan (IHEP); HU, Yu (IHEP); WANG, Lei (IHEP); LIU, Rui (IHEP); SUN, Haokai (IHEP); LIU, Jianli (IHEP); WANG, Shuang (IHEP)

Presenter: FU, Shiyuan (IHEP)

Session Classification: Physics & Engineering Application

Track Classification: Track 1: Physics and Engineering Applications