International Symposium on Grids & Clouds (ISGC) 2026

Contribution ID: **136** Type: **Poster Presentation**

LightPacker: An Adaptive Tool for Data Compression and Serialization in HEPS

The experimental data generated by the High Energy Photon Source (HEPS) is featured with massive scale and high diversity, which imposes severe challenges on the real-time efficiency and long-term storage of data processing pipelines. As efficient data serialization and compression are critical to data transmission and storage, traditional fixed strategies fail to adapt to the diverse characteristics of HEPS data and application scenarios. To address these critical issues, this paper designs and implements an adaptive serialization and compression decision-making system, namely LightPacker, which automatically selects the optimal combination of serialization and compression algorithms for specific data through five core modules: a Feature Analysis Module, an Offline Analysis Module, an Online Evaluation Module, a Decision Execution Module, as well as a supporting Algorithm Library and Historical Information Repository. Specifically, the Algorithm Library integrates various serialization and compression methods, providing a unified calling interface and standardizing the integration rules for new algorithms. The Feature Analysis Module extracts key features of real-time data (e.g., metadata, structure, distribution, and sparsity) and calculates the matching degree between real-time data and historical data. The Offline Analysis Module conducts comprehensive benchmark tests of algorithm combinations on accumulated historical data, storing the optimal solutions corresponding to different feature patterns in the Historical Information Repository to support rapid online decision-making. For real-time data, the Online Evaluation Module first retrieves the optimal algorithm combination from the Historical Information Repository based on the feature matching degree; in the absence of matching records, it performs rapid benchmark tests on candidate algorithms to generate preliminary screening results. Finally, the Decision Execution Module invokes the selected algorithm combination to complete data processing. By adopting strategies such as sampling, the system effectively controls additional performance overhead while ensuring the accuracy of decision-making. Experimental results demonstrate that in typical HEPS scenarios, LightPacker significantly outperforms traditional fixed strategies in both data transmission and storage tasks. This system can effectively reduce network bandwidth occupancy, improve preprocessing efficiency, and minimize storage resource requirements, thereby providing an efficient and adaptive solution for data processing pipelines in HEPS and other similar large-scale scientific facilities.

Keywords
data serialization; data compression; algorithm selection; performance testing; intelligent optimization; HDF5

Primary author: LIU, Dian (⊠⊠⊠⊠⊠⊠⊠)

Co-authors: HU , Yu (IHEP); FU , Shiyuan (IHEP); SUN , Haokai; LIU , Jianli (IHEP); WANG, Lei

Presenter: LIU, Dian (⊠⊠⊠⊠⊠⊠⊠)

Track Classification: Track 1: Physics and Engineering Applications