Contribution ID: **145** Type: **分会报告**

面向 HEPS 的 IO 方法设计与优化

Sunday, 30 June 2024 14:45 (15 minutes)

北京在建的同步辐射光源装置预计每天产生数百 TB 的数据量，每年的数据量达到 PB 量级，对 IO、存储和科学计算带来极大压力。实验过程中需要在线处理用于快速判断数据采集质量，目前从磁盘读取海量实验数据读取存在严重的 IO 瓶颈，因此 HEPS 亟需稳定高效的 IO 方法克服以上困难，首先分析光源下计算任务的读取模式，结合 HDF5 分块存储特性，减少数据跳读，结合并行异步策略加速读写，减少 IO 在计算过程中的占比；其次通过压缩的方式减少数据体积，为保证数据完整性，压缩采用无损的方式，引入压缩会带来额外的时间和资源消耗，而不同的数据压缩效果也有所不同，所以，以加速整个科学计算为目标，综合评价引入压缩的提升，自动触发压缩过程及压缩方法。因此本文拟通过以上方法优化 HDF5 在 HEPS 科学计算过程中的 IO 速度，加速科学结果产出。未来以流处理的方式可以规避海量数据落盘再读取导致的 IO 瓶颈问题，因此最后本文首先介绍了未来 HEPS 场景下 IO 方法的设计思路。

Summary

Primary authors: Dr 符, 世园 (中国科学院高能物理所计算中心); Dr 胡, 誉 (中国科学院高能物理所计算中心); 刘, 建利; 齐, 法制 (高能所); Dr 孙, 浩凯 (中国科学院高能物理所计算中心); 王, 磊; 刘, 锐

Presenter: Dr 符, 世园 (中国科学院高能物理所计算中心)

Session Classification: 先进光源数据与软件