Contribution ID: **5** Type: **Oral Presentation**

Database-aided Workflows on HPC for Plasma Physics Simulation

Thursday, 20 March 2025 14:20 (20 minutes)

Models of physical systems simulated on HPC clusters often produce large amounts of valuable data that need to be efficiently managed both in the scope of research projects and continuously afterwards to derive the most benefit for scientific advancement. Database management systems (DBMS) and metadata management can facilitate transparency and reproducibility, but are rarely used in scientific supercomputing. Reasons include organizational overhead and low performance when migrating data to DBMS which is originally written to files on the parallel file system attached to cluster nodes.

Using first results of the Horizon Europe Project EXA4MIND, the work presented here explores different approaches and system set-ups for managing Plasma Physics data, considering interoperation of HPC systems/filesystems, databases and object stores. We evaluate post-processing workflows for physics simulations run on supercomputing systems at LRZ (Garching b.M./DE) in collaboration with LMU Munichs Chair of plasma and computational physics. The use cases we focus upon in this contribution are simulated many-body systems of elementary particles in plasma physics, produced from outputs from the Plasma Simulation Code (Ruhl et al., Ludwig Maximilian University of Munich/DE). When conducting parameter studies with HPC applications, much work goes into post-processing, visualizing and discussing the simulated data, often several times in an iterative process. TBs of resulting data then have to be processed (e.g. aggregation of domain patches, extraction of statistical information), and evaluated, on various levels – from ensembles of simulations down to single trajectories or timesteps.
Our focus includes testing the performance of typical data queries and processing steps with different execution methods. We strive to facilitate faster and more flexible access to both the raw and processed data by exploring the properties of different storage and database systems. These range from data access schemes facilitated by common python environments over row-based DBMS such as PostgreSQL to column DBMS like MariaDB columnstore, where live queries on large array-based datasets can be executed in memory, or functionalities like Data Vaults can provide access to external repositories. We conclude our contribution stating that modern data storage concepts involving DBMS are also an optimal basis for data sharing and systematic metadata management. Thus, we aim at facilitating a research data management according to the FAIR (findable, accessible, interoperable, reusable) principles from the start.

Primary authors: HACHINGER, Stephan (Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities); PAUW, Viktoria (Leibniz Rechenzentrum)

Presenter: PAUW, Viktoria (Leibniz Rechenzentrum)

Session Classification: Physics & Engineering

Track Classification: Track 1: Physics and Engineering Applications