Parallel I/O in Simulation Workflows

A. Galonska[1], P. Gibbon[1], F. Imbeaux[2], Y. Frauel[2], B. Guillerminet[2], G. Manduchi[3], A. H. Nielsen[4], B. Scott[5], F. Wolf[6]

[1] Forschungszentrum Jülich GmbH, Jülich Supercomputing Centre
[2] Association EURATOM-CEA, CEA/DSM/IRFM, CEA Cadarache, France
[3] Istituto Gas Ionizzati del CNR, Consorzio RFX, Associazione EURATOM/ENEA sulla Fusione
[4] Association Euratom-Risø DTU, DK-4000 Roskilde, Denmark
[5] Max-Planck-Institut für Plasmaphysik, EURATOM-Association, D-85748 Garching, Germany
[6] German Research School for Simulation Sciences, D-52062 Aachen, Germany
Contact: A.Galonska@fz-juelich.de

Abstract

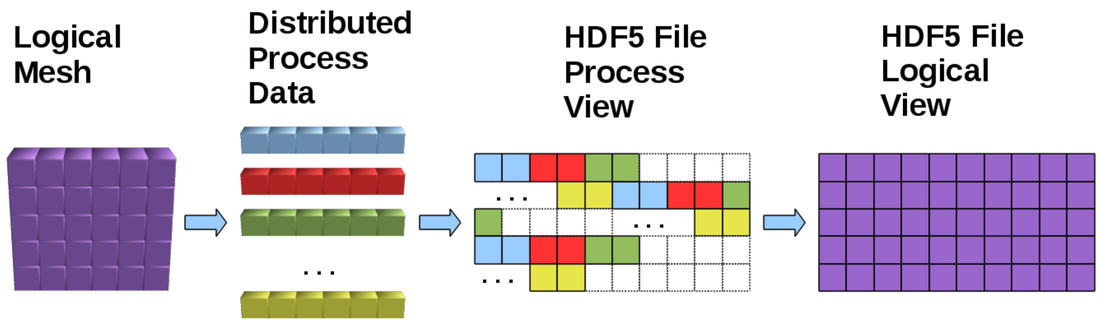
Turbulence code actors (and future HPC actors in general) exchange large size data sets inside KEPLER workflows and motivate the need for parallel I/O. A parallel I/O approach will be developed as an extension of the UAL to optimize the writing of large size CPOs (Consistent Physical Objects) as used for the output of turbulence codes. Parallel I/O for reading and writing code-specific restart files will be provided in addition.
The development of this parallel I/O extension of the UAL is just starting now, in collaboration between ISIP, IMP4 and the JARA-HPC SCIPIO project. The approach will be tested on HPC-FF with the codes ATTEMPT and GEMR.

Turbulence codes characteristics

- Densities, velocities, potentials, …
- Large meshes (ITER)
- Up to 250 GByte output data
- Parallelization based on domain decomposition
- Appear as HPC actors in simulation workflows

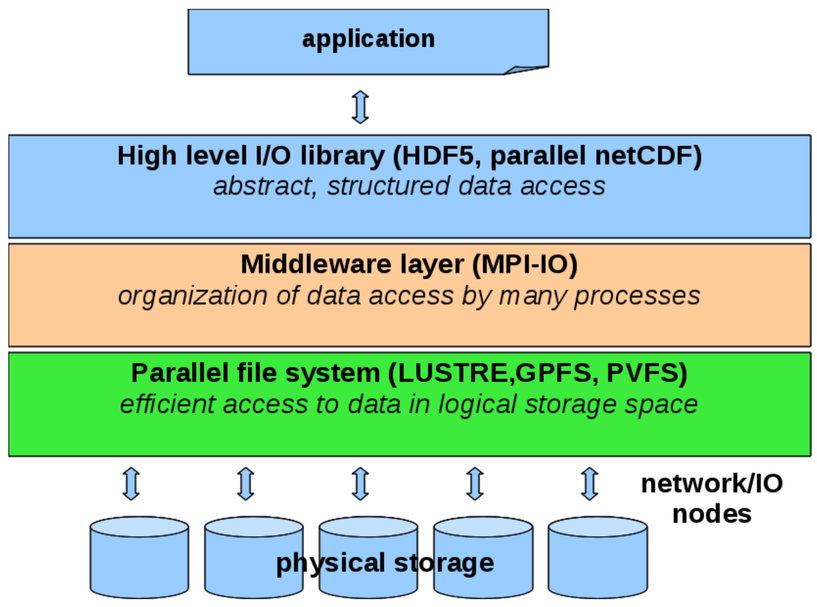
Data organization

- Mesh distributed across participating processes
- Local parts of calculation domain residing on processes
- Force processes writing their local data to assigned positions which results in global mesh inside HDF5 file



Parallel I/O with HDF5

- Utilization of all processes for I/O
- Built on top of MPI-IO
- Provides performance features for large data sets and parallel file systems



Parallel Database Access

- Usage on HPC systems
- Avoid gathering of mesh data
- Collective write of CPO signals to a local ITM HDF5 pulse file (similar to low-level UAL instructions)
- Transport of HDF5 file to gateway
- Link-in to the “run” database of the workflow

Restart Files

- Restart points for HPC actors
- SIONLib foreseen for parallel I/O
  - task local I/O to “multi-file”
- Collective writing of restart data
- Interface to KEPLER
- Easy restart point selection inside KEPLER workflow