

Data Grid for Fusion Simulations and Experiments

Svetlana Shasharina, Robert Eger, John Cary

Tech-X Corporation

5541 Central Avenue, Suite 135, Boulder, CO 80301

Simulations and experiments in the fusion community generate large data sets at remote sites. Visualization and analysis of these data sets is then difficult, as the data comes in different formats and downloading it to local computers can take hours over even the fastest network connections. Grid technologies are capable of providing solutions for fusion data transfer and access, but need to be customized to the fusion-specific needs and merged with distributed technologies currently used by the community.

Hence, the full scope of our projects is to develop the Fusion Grid Service specialized for fusion data transfer and access for visualization. This service will connect heterogeneous data sets coming from simulations and experiments, bring data into well-established analysis tools and provide means to switch between data-transfer protocols depending on size and type of transferred data. These protocols will include MDSip (the native MDSplus protocol), GridFTP (from Globus) and GridPST (a protocol recently developed at MIT for potential use in MDSplus). The proposed Fusion Grid Service will have HDF5 and MDSplus servers specialized to deal with the corresponding type of data. The client interface will be uniform and allow transparent and easy access to either kind of data. For ease of visualization and simplification of the interface, the underlying C++ client will be wrapped into Interactive Data Language and AVS/Express for accessing remote HDF5 files.

The first step of this work is to figure out the appropriate data transfer mechanisms for various data formats and sizes. For this purpose, we are performing an extensive set of performance tests. They include:

- full and partial file transfer of MDSplus data using Globus I/O over WAN and LAN,
- full and partial file transfer of MDSplus data using MDSplus with GridPST (an add-on service for MDSplus that provides for high-speed socket communication over multiple socket connections between hosts) instead of MDSip over WAN and LAN;
- full and partial HDF5 file transfer using GridFTP with multiple parallel streams over WAN and LAN;
- Extraction of MDSplus data into a binary file, transferring it using GridFTP and writing it into a local tree versus direct writing into a local MDSplus tree from a remote tree;
- Partial HDF5 and MDSplus file transfer using the `globus_twrite` application, developed at PPPL. This application allows researchers to have applications running on supercomputers stream their output data directly to HDF5 files to supercomputers at a different site and uses GridFTP.

We will present results of these tests, outline the recommendations for the data transfer mechanisms for the Fusion Data Grids under construction and report the progress on the project as a whole.