DIII–D DATA MANAGEMENT*

B.B. McHarg, Jr., J.R. Burruss, J. Freeman, C.T. Parker, J. Schachter, and D.P. Schissel

General Atomics, P.O. Box 85608, San Diego, California 92186-5608
email: mcharg@fusion.gat.com, Phone: (858) 455-4249, Fax: (858) 455-2454

The DIII–D tokamak at the DIII–D National Fusion Facility now routinely acquires ~500 Megabytes (uncompressed) of raw data per pulse of the experiment. It is expected that in FY01, over one Terabyte of data will be acquired. In addition, there is a growing suite of codes running between pulses that produce analyzed data. Currently this amounts to ~10 Megabytes per pulse with total disk usage of about 80 Gigabytes. Most recently a relational database system has been introduced which is also adding to the overall data load. These three classes of data have evolved into three separate server systems, one for raw data, one for analyzed data, and one for database data. In addition there are several diagnostics, which are not part of the normal data system because they alone acquire as much data as the rest of the experiment's diagnostics combined. The management of all of the data and providing timely access to researchers is a significant and growing problem as the quantities of data are only expected to continue to grow in the future. In recent years there has been an order of magnitude increase in magnetic disk space devoted to raw data in order to accommodate this growth. In addition a Hierarchical Storage Management system (HSM) was implemented. Without these, management of the raw data and access by researchers would have been much more difficult and less efficient. As the analyzed data quantity grows, it too will require more management. This paper will examine the experiences of the approaches that have been taken in management of the data and plans for the continued growth of the data quantity. In particular the HSM system will be described as well as the hardware/software configurations of the three server systems and how they fit into the DIII–D environment. Also to be discussed is how these different classes of data are accessed: PTDATA for the raw data, MDSplus for the analyzed data, and MS SQL for the database data.

Topic: Data base techniques for information storage and/or retrieval