GA–A22730

DEVELOPMENT OF IMPROVED METHODS FOR REMOTE ACCESS OF DIII–D DATA AND DATA ANALYSIS

by
K.L. GREENE and B.B. McHARG, JR.

NOVEMBER 1997

DISCLAIMER

GA–A22730

DEVELOPMENT OF IMPROVED METHODS FOR REMOTE ACCESS OF DIII–D DATA AND DATA ANALYSIS

by
K.L. GREENE and B.B. McHARG, JR.

This is a preprint of a paper to be presented at the 17th IEEE/NPSS Symposium on Fusion Engineering, October 6–11, 1997, San Diego, California and to be published in the *Proceedings*.

Work supported by
the U.S. Department of Energy
under Contract No. DE-AC03-89ER51114

GA PROJECT 3466
NOVEMBER 1997

Development of Improved Methods for Remote Access of DIII–D Data and Data Analysis*

K.L. Greene and B.B. McHarg, Jr.
General Atomics
P.O. Box 85608, San Diego, California 92186-9784

Abstract — The DIII–D tokamak is a national fusion research facility. There is an increasing need to access data from remote sites in order to facilitate data analysis by collaborative researchers at remote locations, both nationally and internationally. In the past, this has usually been done by remotely logging into computers at the DIII–D site. With the advent of faster networking and powerful computers at remote sites, it is becoming possible to access and analyze data from anywhere in the world as if the remote user were actually at the DIII–D site. The general mechanism for accessing DIII–D data has always been via the PTDATA subroutine. Substantial enhancements are being made to that routine to make it more useful in a non-local environment. In particular, a caching mechanism is being built into PTDATA to make network data access more efficient. Studies are also being made of using Distributed File System (DFS) disk storage in a Distributed Computing Environment (DCE). A data server has been created that will migrate, on request, shot data from the DIII–D environment into the DFS environment.

INTRODUCTION

The DIII–D National Fusion Facility, a tokamak experiment funded by the U.S. Department of Energy and operated by General Atomics (GA), is an international resource for plasma physics and fusion energy science research. There is a long history of collaborations with scientists from laboratories and universities from around the world. That collaboration has mostly been conducted by participation at the DIII–D site. New developments in computing and technology are now facilitating collaboration from remote sites, thus reducing some of the needs for traveling to the experiment. These developments include higher speed wide area networks, more powerful workstations, a distributed computing environment, network based audio/video capabilities, and the use of the World Wide Web.

An important part of the ability to collaborate is to have the ability to access data both for examination and for analysis. The DIII–D raw shot data is accessed via a single subroutine called PTDATA [1]. This routine has evolved over the years from initially accessing data on a single computer to becoming a client server process within the DIII–D environment whereby data on any computer could be accessed from any other computer. In recent years, changes were also made to facilitate wide area access as well. The problem with wide area access is that generally the greater the distance the less network bandwidth there is. On the other hand, accessing data directly from a remote site permits use of cpu cycles at the remote site, thus freeing up cpu cycles at GA. In order to alleviate the problem of slow access, a data caching capability has been built into PTDATA.

Another way of improving wide area access to data is by using the Distributed File System (DFS). This was explored recently as part of the Remote Experimental Environment (REE) project which was a joint study by GA, Princeton Plasma Physics Laboratory (PPPL), Lawrence Livermore National Laboratory (LLNL), and Oak Ridge National Laboratory (ORNL) [2]. The purpose of this study was to introduce tools into the DIII–D environment and to study their effectiveness. These tools included the use of audio/video for communication from the DIII–D control room, the broadcast of meetings, the use of interprocess communication software (IPCS from LLNL [3]) to post events to the network during a tokamak shot, and, in particular, the creation of a DCE cell for creating a common collaboratory environment. This DCE cell then used DFS disks to facilitate access to both raw and processed data.

This paper will discuss the improvements being made for data access. The changes to PTDATA involve the concept of caching. This is augmented by the use of DFS disks. This work is still in an experimental stage, but is expected to be in more common use within the next year.

THE PTDATA DATA ACCESS MECHANISM

When data is acquired from the DIII–D experiment, it is written into a shot file. Once data is in the shot file, it is available for data access from any computer system via a PTDATA call. PTDATA is the universal data interface for DIII–D data. It was originally written for the VAX but has been transported to UNIX systems for accessing data residing either on VAX computers or UNIX computers. The user calls PTDATA with a shot number, a data type (file extension), a pointname (which is a unique signal name), and certain parameters describing what data is to be returned.

Given the calling parameters, PTDATA must then find out

*Work supported by U.S. Department of Energy under Contract No. DE-AC03-89ER51114.

where the data is located. It should be noted that PTDATA can run on a variety of UNIX computers as well as VAX and Alpha VMS systems and supports both DECnet (locally) and TCP/IP (locally and wide area) for communications. The PTDATA search algorithm consists of the following "rules":

0 Search all file extensions in a given location

This is called rule 0 because it actually applies to the later rules. Since data resides in multiple files, if a connection already exists, whether local or remote, it is worthwhile to search all files for a given shot number in order to find a given pointname.

1 Search where data was last found

If pointname A was found on a particular computer, then it is likely that pointname B will be found there also. Thus, the previous computer will be searched first, even though that may be over the network (the network connection, which requires the most overhead, will already have been made on the previous call).

2 Search locally

If data resides on a disk local to the computer, then it will be found there without having to make a network connection. The overhead of always searching locally is minimal, particularly since disk directories are most often cached in memory anyway.

3 Query for location of the shot and pointname

Information is maintained on the User Service Center VAX computers for the location of a particular shot and shot file extension. If the shot.extension is known to exist on a particular computer, then PTDATA will search that computer. In the process of checking for the shot.extension location, other information is also obtained. Two tables are maintained on the User Service Center VAX computers which relate a pointname to a particular shot file extension and relate an extension to a particular computer. That computer is considered to be the origin point for that data and thus enables PTDATA to find out where it might be immediately after a tokamak shot.

4 Search remote locations

There are three locations where any shot file might be found and one additional location where the shot file originates. PTDATA searches these locations for the data. The first three are the shot server system where most restored data resides, the DIII–D VAX cluster, and the USC VAX cluster. The origin computer is determined from information obtained in step 3. It should be noted that network connections to the three known locations are always kept open in order to avoid the overhead of making those connections again later. In addition, up to two other connections are also kept open until needed for a different computer.

If data is still not found, then it is not currently on disk and the caller will need to request that it be restored.

PTDATA CACHING AND USE OF DFS DISKS

PTDATA works very well within the DIII–D environment and is actually in use at some other labs as well (note only TCP/IP is used over the wide area network). However, much greater efficiency in wide area access can be obtained if accessed data is cached locally after the first access. Thus subsequent accesses by the same or other users for a particular pointname would come from a local rather than remote source. A secondary effect is that data which is of more interest to users and is accessed frequently will tend to remain on disk locally. Also note that only pointnames that are accessed are cached locally, not the entire shot file. This caching is enhanced further by the use of DFS disks in the DCE environment and its automatic caching capabilities.

As part of the REE project mentioned previously, a DCE cell was installed with cell management services provided by the LLNL computing center. The fusion DCE cell forms a nationwide distributed computing "cluster" of workstations located at GA, PPPL, LLNL, and ORNL. Disks that are part of this cell are DFS disks. A data server has been written to migrate data from the DIII–D environment into the DCE environment. Once data is on the DFS disks, access by other nodes can take advantage of DFS's automatic caching capabilities thus further improving data access.

Recent changes to PTDATA allow it to determine when it is being executed under DCE. When this is the case and data is not available on DFS disks, PTDATA will no longer perform a search for data on remote disks but will instead send a request for data migration to the data server. After receiving a reply from the data server, PTDATA will make another attempt to access data from a DFS disk. Once the data for a requested pointname has been migrated to a DFS disk, all subsequent users interested in this pointname will see the data without further requests being sent to the data server.

The data server actually consists of two programs which though usually on different computers could be on the same computer. This was done so that all PTDATA calling processes did not have to be IPCS processes. The local data server is intended to run on DCE nodes where PTDATA is available. The master data server will run on a DCE node at GA known as TRITON. Communication between PTDATA and the local server uses UNIX message routines. Communication between the local server and the master server uses LLNL's IPCS.

The steps for PTDATA to find data are now modified after step 2 discussed previously. If PTDATA cannot find data locally, no search of remote disks is performed, as was described earlier. A private mail box is created and the ID of the mail box is included with the request for data that is sent to the local server via a mail box known to be associated specifically with the local server. When the local server receives a request for data, the request is put into IPCS format and is then sent, via IPCS routines, to the master server, located on TRITON. It is the responsibility of the master server to perform the search for data on remote disks. If the data is located remotely, a copy of the data is placed into the appropriate shot.extension file located on a DFS disk. After the search for data is complete, the appropriate reply is created and sent, via IPCS, to the specific local server which made the request. The local server, in turn, sends the reply to the original requestor. If PTDATA receives a positive response from the local server, then it will make one more attempt to find the data locally.

These concepts are illustrated by Fig. 1. Keep in mind that what is called "local disks" on the DFS node may in fact be anywhere in the DCE cell. If data is not found there, PTDATA contacts the local server which in turn contacts the master server at GA. The master server makes a normal PTDATA call which searches the local GA environment for the requested data. If found, then that data is written to the DFS disk which appears local to all computers in the DCE cell. The local PTDATA then again tries the DFS disk to access the data. Any subsequent accesses will of course find the data on the DFS disks.

In principle, a node does not have to be part of the DCE cell or use DFS disks. The local and master server can run on a local node anywhere and thereby cache data on disks that are truly local to the node. This is particularly important since not all computers will be or even can be part of a DCE cell and have DFS disks.

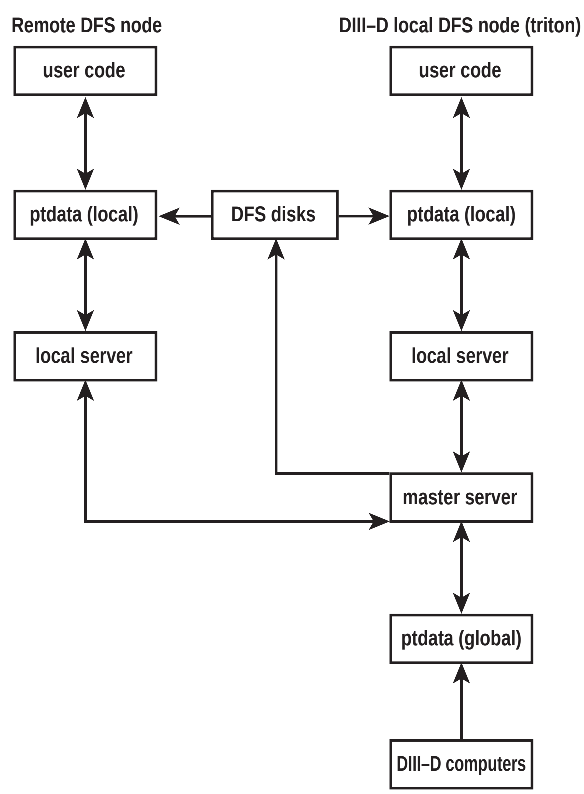

Fig. 1. Diagram of data flow for local and remote systems.

CONCLUSIONS

Modifications have been made to PTDATA to permit a data caching on a remote node. This provides a more efficient mechanism for accessing data since after the first access, all subsequent accesses will be from a local disk. Also, the more important data will be retained on the local system. This mechanism involves the addition of a local server and master server to the PTDATA mechanism. The mechanism is further enhanced if DFS disks are used in the DCE environment since a large DFS disk farm can exist for data which will appear local to other computers in the cell. Thus, the caching capabilities of DFS are an aid to data access.

The PTDATA caching work is still in experimental stages. Future work will focus on getting it into production use. There are also some efficiency issues that need to be addressed. Work thus far has focused on UNIX systems, but similar software needs to be developed for VMS systems. Some preliminary work was done for managing the cached data space (file deletion/retention for example) [4], but further work needs to be done.

REFERENCES

[1] B. McHarg, "Access to DIII–D data located in multiple files and multiple locations," in *Proc. 15th Symp. on Fusion Engineering*, p. 123, 1993.

[2] B. McHarg, T. Casper, S. Davis, D. Greenwood, "Tools for remote collaboration on the DIII–D National Fusion Facility," presented at the IAEA Technical Committee Meeting on Data Acquisition and Management for Fusion Research, Garching, Germany, July, 1997, unpublished.

[3] G. Preckshot, D. Butner, "A simple and efficient interprocess communication system for actually using a laboratory computer network," IEEE Trans. Nuclear Science, NS-34, 1987, p. 858.

[4] J. Freeman, private communication.