Preprint
UCRL-136058

The PCMDI software system and the next generation Internet project

G. L. Potter, D. N. Williams

This article was submitted to
Annual Bureau of Meteorology Research Centre's Workshop on
Parallel Computing in Meteorology and Oceanography
Melbourne, Australia, November 9-16, 1999

October 15, 1999

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

The PCMDI software system and the next generation Internet project

GERALD L. POTTER AND DEAN N. WILLIAMS

Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory
Livermore, California USA

PCMDI Software

One problem facing many scientists is not the absence of tools to analyze data, but rather a shortage of interrelated diagnostic software that is consistent, flexible, portable, adaptable, efficient, sharable, and easy to use. Consequently, many scientists are writing their own programs to ingest, manipulate and display data. Debugging and enhancing special purpose software diverts time that otherwise would be spent on research. The result is often not "friendly", reusable, or portable, nor does it promote standards within the research community.

In response to the needs of the scientific community, PCMDI has developed a suite of software tools for the storage, diagnosis, and visualization of data. PCMDI's principal tools are the Climate Data Analysis Tool (CDAT), the Climate Database Management System (CDMS), and the Visualization and Computation System (VCS). The design goal of this suite of software is to reduce the redundancy encountered so often in scientific analysis and to allow researchers to concentrate on their science.

One obstacle to sharing analysis software is the wide variety of data file formats that are in use. Programs must be written to convert data to a user's preferred file format and conventions. This data conversion requires additional expenditure of efforts on testing and quality assurance. Modular and interrelated software performs such tasks transparently.

Table 1 presents an abbreviated list of the PCMDI software system listing version and availability.

Table 1: Short-List of Software Products

Name	Acronym	Description	Distributed	Version
Climate Data Analysis Tool	CDAT	Utilizes an interpreted language to manipulate data and provide climate scientists with diagnostic routines.	Externally	2.0
Climate	CDMS	Designed to automatically locate and extract	Internally	Beta

Database Management System		metadata (i.e., variables, dimensions, grids, etc.) from PCMDI's collection of model runs and analysis files.		
Common Data UNIform File	cdunif	A library of input functions for accessing netCDF, HDF, DRS, GrADS/GRIB and VPOP data files.	Externally	1.7
Data and Dimension Interface	DDI	A Graphical User Interface (GUI) that reads and writes available file formats, and sends data to a variety of visualization systems.	Externally	1.2
Data Retrieval and Storage	DRS	A library that supports direct access I/O, and multi-dimensional array variables.	Externally	1.5
EzGet		A FORTRAN application programmer's interface (API) for reading various data file formats via cdunif, and for performing grid transformations, data masking, and some calculations.	Externally	1.1
Library of AMIP Data Transmission Standards	LATS	A collection of software routines to output gridded data in either netCDF or GRIB formats for the Atmospheric Model Intercomparison Project (AMIP).	Externally	1.1
Quality Control Software	QCS	Software that checks the correctness of AMIP model data.	Internally	Beta
Visualization & Computation System	VCS	A graphics software package that allows the display, animation, and manipulation of scientific data.	Externally	3.1

Data from simulations may be made up of several data formats. For instance, one model may produce a dump of all variables in a file for each time interval. Without rewriting the data, the PCMDI software can extract one or more variables for all time intervals as if they existed in one file. If necessary, the data can be rewritten in netCDF for later use. More organized data may be put into a database and located at other sites or left in original form. The general layout of the PCMDI software is shown in Fig. 1. CDAT is built around the PYTHON language, which enables it to use modules written in a variety of languages.

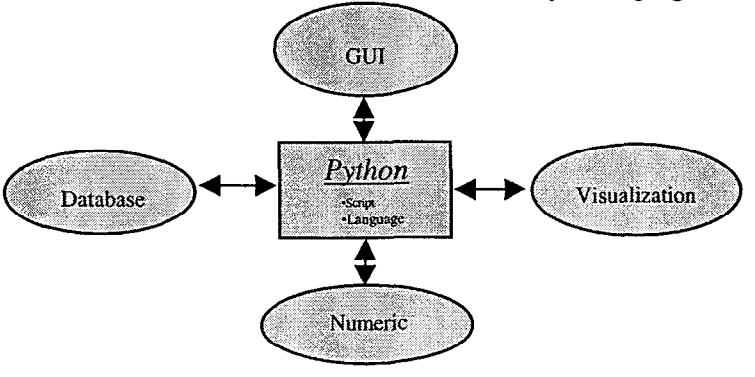

Fig. 1: The PCMDI Software system.

The Next Generation Internet (NGI)

PCMDI software was originally designed to be used in-house but because of the modular design and expandability, considerable interest has been generated to make it part of the US Department of Energy Next Generation Internet (NGI). The NGI is focused on network requirements that will enable data-intensive scientific research not now possible because of network limitations. It is anticipated that the results and "spin-offs" of this research, after testing and prototyping by the scientific community, will impact broad commercial use of networks.

One example of the need for this new network capability comes from high-resolution, long-duration simulations performed with future climate models. They will produce tens of petabytes of output. The output in turn will be made available to global change impacts researchers nationwide through a network of diagnostic and regional climate centers (Patrinos, 1998 and Gates 1999). These distributed centers, users, models, and data will be connected in a virtual collaborative environment called the Earth System Grid (ESG). User requests for data products will be translated into appropriate combinations of accesses to data caches, requests to central data archives, and new large-scale simulations.

The Earth System Grid will provide scientists with virtual proximity to the distributed data and resources comprising this collaborative environment. The simulation data to be maintained at central sites and/or at regional centers will likely total tens of Petabytes. The effective management of the required data movement operations will tax even the highest performance and most advanced networks. A prototype of an early version of the ESG will be demonstrated at Supercomputing 99 in the US. Fig. 2 is a schematic of the demonstration set-up showing the three sites involved.

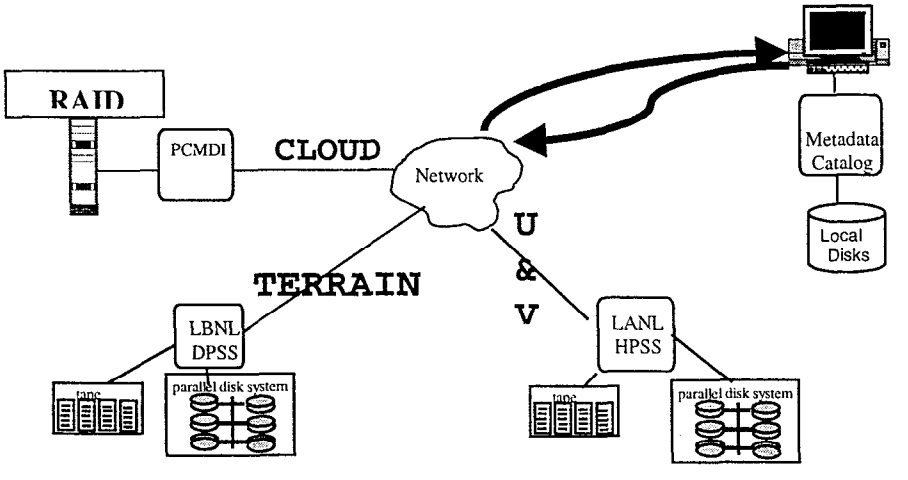

Fig. 2 shows a schematic of the ESG network as it will appear at Supercomputing 99. Note, that even if the hardware configurations are very dissimilar, the data appears the same to the requestor.

References

Patrinos, A., 1998. *The Accelerated Climate Prediction Initiative: Bringing the Promise of Simulation to the Challenge of Climate Change*, Committee Report. (http://www.epm.ornl.gov/ACPI/Documents/ACPIfinal.html.

Gates, W. L., 1999. *Recommended Implementation of the Accelerated Climate Prediction Initiative (ACPI)*, Report of the Ad Hoc Inter-agency Committee for ACPI Implementation.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.