



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

The Program for Climate Model Diagnosis and Intercomparison (PCMDI) Software Development: Applications, Infrastructure, and Middleware/Networks

D. N. Williams

June 30, 2011

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

The Program for Climate Model Diagnosis and Intercomparison (PCMDI) Software Development: Applications, Infrastructure, and Middleware/Networks

Software Project Lead:

Dean N. Williams

Submitted to:

U.S. Department of Energy Office of Science
Office of Biological and Environmental Research (BER)

Abstract

The status of and future plans for the Program for Climate Model Diagnosis and Intercomparison (PCMDI) hinge on software that PCMDI is either currently distributing or plans to distribute to the climate community in the near future. These software products include standard conventions, national and international federated infrastructures, and community analysis and visualization tools. This report also mentions other secondary software not necessarily led by or developed at PCMDI to provide a complete picture of the overarching applications, infrastructures, and middleware/networks. Much of the software described anticipates the use of future technologies envisioned over the span of next year to 10 years. These technologies, together with the software, will be the catalyst required to address extreme-scale data warehousing, scalability issues, and service-level requirements for a diverse set of well-known projects essential for predicting climate change. These tools, unlike the previous static analysis tools of the past, will support the co-existence of many users in a productive, shared virtual environment. This advanced technological world driven by extreme-scale computing and the data it generates will increase scientists' productivity, exploit national and international relationships, and push research to new levels of understanding.

Table of Contents

1 Introduction3

2 Vision3

2.1 Advisory boards.....5

2.2 Strategy.....5

3 Leading Development Activities and Collaborations 7

4 Hardware.....8

4.1 Procured Items.....9

4.2 Future Ongoing Costs (Responsibility of PCMDI).....10

4.3 System Description10

5 Future..... 10

6 Reference..... 12

1 Introduction

Inspired by new software and digital technologies, the software team of the Program for Climate Model Diagnosis and Intercomparison (PCMDI) is leading the charge to juxtapose national and international ultrascale climate simulations with observations and reanalysis data sets. These new developments are the neophyte stages of assembling disparate data collections under one worldwide-federated infrastructure. As we have seen since our last report in 2009, data volumes have exploded, and rapid demands for their use are on the rise, pushing the need for new and improved data intensive software and hardware [Overpeck 2011]. From PCMDI's perspective, we have seen a significant rise in requests to hold more data. The 2007 Third Phase Coupled Model Intercomparison Project (CMIP3) required 35 terabytes (TB, or 10^{12} bytes). In contrast, the upcoming 2013 Fifth Phase Coupled Model Intercomparison Project (CMIP5) will need 3.5 petabytes (PB, with 1 PB equaling 10^{15} bytes), with total federated CMIP5 data holdings expected to reach between 5 to 10 PB. To compliment the CMIP5 data archive, other high-profile data sets are scheduled to coexist within the federated infrastructure: the Community Earth System Model (CESM), the North American Regional Climate Change Assessment Program (NARCCAP), the Transpose-Atmospheric Model Intercomparison Project (TAMIP2), the Carbon-Land Model Intercomparison Project (C-LAMP), and the National Aeronautics and Space Administration's (NASA's) Atmospheric InfraRed Sounder (AIRS), to name only a few. To provide the international climate community with convenient access and to maximize scientific productivity, these data need to be replicated and cached at multiple locations around the world. Managing a distributed data system of this magnitude presents several significant challenges not only to system architectures and application development but also to the existing wide area networking infrastructures. Connecting people, organizations, bureaucracies, government agencies, and academia for the betterment of climate research is at the heart of our software development.

As the starting point for new and innovative research and discovery, software development plays a key role in managing (e.g., organizing, cataloging, publishing), analyzing (e.g., ensembles, uncertainty quantification, zonal means), and visualizing (e.g., Taylor diagram, skew-T, iso-surfaces) data. Because data are becoming too large to move, the need and use of remote compute servers to reduce data before bulk data transfers take place has increased. Input/Output (I/O) must be parallelized for more efficient compute processing and data delivery. Our research shows that 75% of processing time is due largely to I/O tasks. Although much of the computing is projected to be done server-side in the future, large quantities of data must be moved over the Internet for replication and long-term data preservation and for better accessibility. Today's network speeds average 5-10 Megabits per second (Mbps) for cable Internet users and 700–3,000 Mbps between national and international institutions. With a transfer rate of 1,000 Mbps, we can expect to move only 10 TB a day in a world driven by petabyte—and eventually exabyte—scales. This will hardly suffice. Therefore, to move petabytes of data from one institution to another, we need to utilize the network and ship 2-100 terabyte disks. In either case, human intervention is needed at every phase of the bulk-data movement.

2 Vision

Climate scientists collaborate worldwide to inter-compare knowledge gained by examining disparate data. For this to happen efficiently, PCMDI envisions universal climate data access in which data curation, preservation, dissemination, analysis, and visualization can be appropriately, reliably, and readily managed. At PCMDI, we have developed dedicated software for these tasks in anticipation of a diverse user community, but scientists are still our primary focus. The investments made have enhanced science at PCMDI and the national and international research and development communities at large. Though in some cases the desired stages of development are just now being realized, these developments ensure future progress for the continued successful dissemination of global climate information. The software to

ensure our success for future climate understanding under the leadership of PCMDI includes the following:

- **NetCDF Climate and Forecast (CF) Metadata Convention** for Earth sciences data. Adopted by a number of projects and groups as a primary standard, this convention is designed to promote the processing and sharing of files created with the NetCDF API. PCMDI requires the Coupled Model Intercomparison Project (CMIP) to use CF, which is the basis of the Intergovernmental Panel on Climate Change (IPCC) assessment reports. Others in the community are working closely with PCMDI on the CF governance board [CF 2011].
- **Climate Model Output Rewriter 2 (CMOR-2)**. This comprises a set of C-based functions with bindings to both Python and FORTRAN 90 that can be used to produce CF-compliant NetCDF files that fulfill the requirements of many of the climate community's standard model experiments, such as CMIP3, CMIP5, C-LAMP, AIRS (level 3), NARCCAP, and others [CMOR 2011].
- **Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT)**. Written in Python and conforming to the CF conventions, this climate data analysis tool brings together CDAT, VisIt, Paraview, VisTrails, and R. They address the particular requirements and ultra-scale and comparative capabilities needed for access, analysis, and visualization of climate data sets for the Office of Biological and Environmental Research (BER) [UV-CDAT 2011].
- **Earth System Grid Federation (ESGF)**, a federated, distributed production infrastructure for the interchange of climate data and information, model codes, analysis, and visualization capabilities [ESGF 2011].
- **Diagnostics**. PCMDI is developing diagnostics, new techniques, and software to understand the nature and cause of underlying climate change, such as the integration of the Microwave Sounding Unit (MSU) into UV-CDAT for model and observational data intercomparison studies.
- **Climate Energy Fusion Simulation**, a community outreach interactive simulation focused on finding solutions for the energy-climate challenges. Players face the challenge of meeting the world's 21st century energy demands on a fixed budget while keeping carbon emissions at a minimum [Simulation 2011].
- **LibCF**. A C/Fortran library that supports the creation of scientific data files conforming to the Climate and Forecast (CF) conventions, it uses the netCDF API (Application Programming Interface). LibCF is being developed at Unidata and will be released as part of the NetCDF library distribution. PCMDI is working to integrate this library into the UV-CDAT community analysis tool [LibCF 2011].
- **GRIDSPEC**, a set of CF conventions for the representation of gridded data for Earth system models. V. Balaji and Zhi Liang developed a set of GRIDSPEC command-line tools at the Geophysical Fluid Dynamics Laboratory (GFDL). Dave Kindig and Alex Pletzer at Tech-X and PCMDI, as part of the libCF library, have further implemented the GRIDSPEC standard to be included in UV-CDAT and to be used in CMIP5 data description [Gridspec 2011].
- **Mosaic**. There are two pioneering implementations of the Mosaic GRIDSPEC. One is a complete XML schema developed on the basis of the Gridspec; the other is a complete netCDF-3 classic implementation. Working with V. Balaji and others at GFDL, Dave Kindig and Alex Pletzer at Tech X and PCMDI are working to integrate the Mosaic GRIDSPEC into the UV-CDAT community analysis tool [Mosaic 2011].
- **Quality Control (QC)**, a process that works within the ESGF environment. There are three levels of QC checking: Level 1 is performed at the data publishing process. Level 2 runs quality control code developed at the German Climate Computing Centre (DKRZ) by PCMDI, BADC, or

DKRZ. Level 3 involves visual inspections performed by designated personnel at PCMDI, BADC, or DKRZ [QC 2011]

- ***PyDAP with ESGF Security***, a Python library implementation of the Data Access Protocol (DAP), also known as OPeNDAP. The PyDAP security server is collocated with the data and allows DAP clients (such as UV-CDAT) secure access to the ESGF distributed data sets [PyDAP 2011].
- ***Live Access Server (LAS)***, the product server of the ESGF. Collocated with the data, LAS allows data providers to connect analysis and visualization tools as node services. PCMDI is working with members of the community to connect UV-CDAT as a product to deliver analysis and visualization [LAS 2011].
- ***Websites*** important to the community: www-pcmdi.llnl.gov, climate.llnl.gov, cf-pcmdi.llnl.gov, www2-pcmdi.llnl.gov/cmor, cmip-pcmdi.llnl.gov, esgf.org, uv-cdat.org.

Thanks to recent extensions to the libCF library, which implements the CF conventions in a C API, climate data producers and consumers can create and read arbitrary mosaic data sets. The libCF supports a layered approach that users can build applications by using the software at many levels: curvilinear grids and data, regridding, mosaic (for inferring inter-tile connectivity), and host objects (for spatial and time aggregation). The UV-CDAT software is currently in the process of being extended to logically aggregate spatially and temporally scattered data. This involves adding the libCF library to the build system, modifying UV-CDAT's Climate Data Management System (CDMS) objects to the detect role played by a NetCDF file in a mosaic aggregation, and creating new file and variable objects that behave similarly to the native CDMS objects. This is just one example of how one component of software interoperates with another. All software fits into the ESGF federated infrastructure or can exist on its own.

2.1 Advisory boards

In order to coordinate and gain a broader acceptance and input on the external projects' progress and priorities, PCMDI is on several advisory boards, such as the National Oceanic and Atmospheric Administration's (NOAA's) Data Archiving and Access Requirements Working Group (DAARWG), the international G8 Exascale Research Initiative, the PCMDI and National Aeronautics and Space Administration (NASA) Observation Intercomparison group, and the Global Organization for Earth System Science Portals (GO-ESSP). Members of the PCMDI software team have agreed to participate and provide essential coordination between external project activities and the work being led by PCMDI to deploy the CMIP5 archive. The PCMDI software team will advise on scientific priorities regarding the processing of COordinated Regional climate Downscaling Experiment (CORDEX)¹ data. Being involved with advisory boards helps our strategic perspective and provides interactions and synergy with other climate-change initiatives. As members on the advisory boards, we review project delivery plans and deliverables on periodic timetables. More recently, PCMDI has membership on the DOE Office of Advanced Scientific Computing Research's (ASCR's) Federally Chartered Advisory Committee to help push the frontiers of computing for scientific applications.

2.2 Strategy

We crafted the following strategy to achieve our vision of creating a comprehensive framework of policies, management, and organizational structures that provide reliable, effective access to the full

¹ The CORDEX project is designed to meet this requirement, producing downscaled data that will cover the majority of populated land regions around the globe at resolutions from 50km (initially), increasing to 10km in many regions as the project progresses [CORDEX 2011].

spectrum of scientific climate data and are transparent, evolvable, and extensible. Such frameworks will serve as a driving force for future PCMDI leadership in the climate community. For this, we are both community leaders and partners, providing maximum data access and utilities, implementing a global plan to efficiently manage, disseminate, and process data; empower the community of potential users now and into the future; support global capability; and promote best practices and standards in the community.

Figure 1 shows our ESGF infrastructure strategy with the national and international collaborations. Users can access ESGF data using Web browsers, scripts, and in the near future, UV-CDAT clients. ESGF infrastructure is separated into gateways (green) and data nodes (blue). Gateways handle user registration and management and allow users to search, discover, and request data. Data nodes are located where the data resides, allowing data to be published (or exposed) on disk or through tertiary mass storage (e.g., tape archive) to any gateway. They also handle UV-CDAT data reduction, analysis, and visualization. ESGF currently comprises eight national and international gateways, four of which hold special status in housing CMIP5/AR5 replication data sets: LLNL/PCMDI, BADC, DKRZ, and the Australian National University (ANU) National Computational Infrastructure (NCI). Users have access to all data from the federation regardless of which gateway is used.

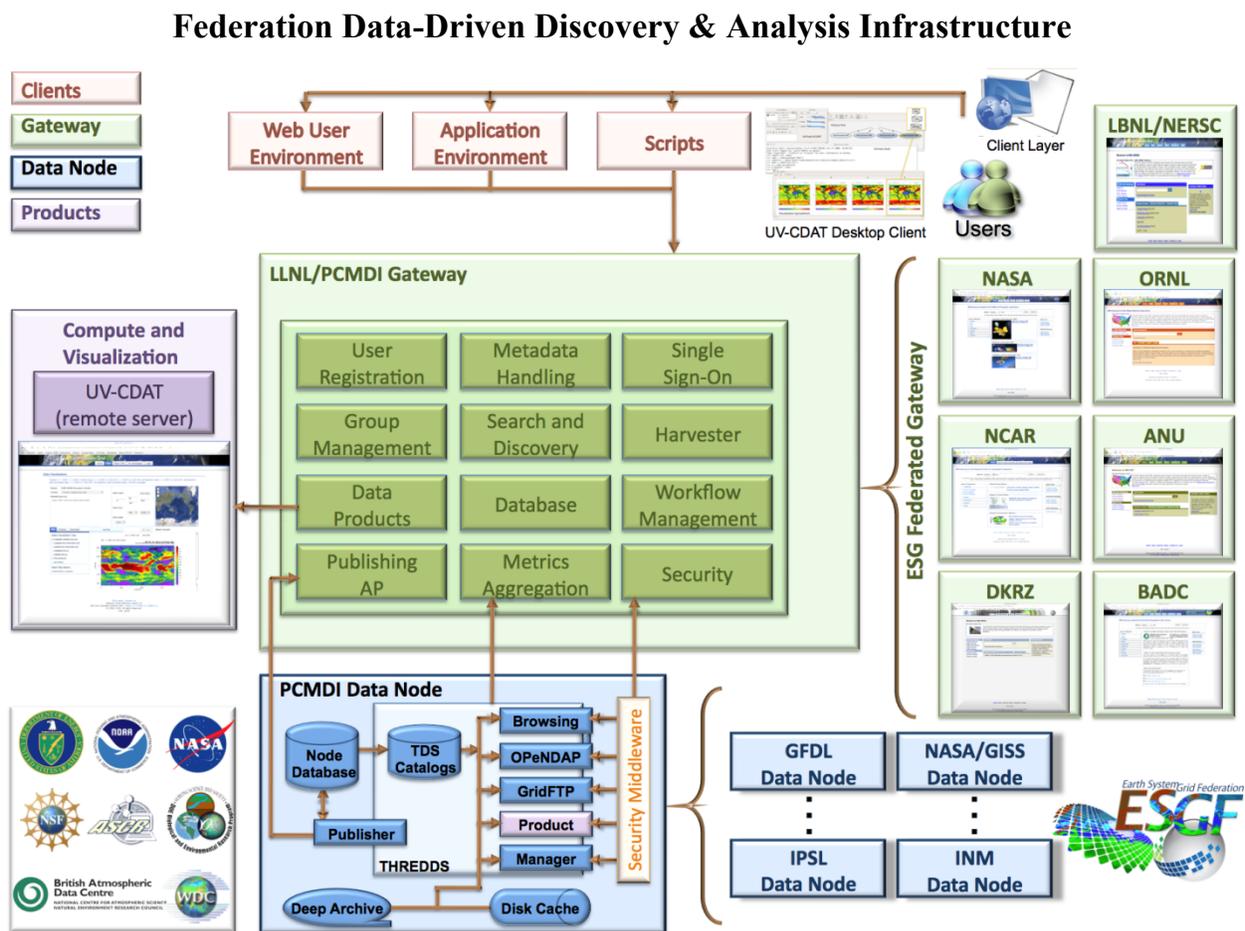


Figure 1: The culmination of hidden policies and standards, international collaborations, and the overall vision and strategy of a comprehensive infrastructure and framework for maximum data planning, management, and analysis

3 Leading Development Activities and Collaborations

PCMDI manages community activities and resources to successfully support PCMDI's mission and the international climate community as a whole. As mentioned, we lead a variety of interconnected projects aimed at data dissemination, analysis, and visualization of climate simulations, as well as observations and reanalysis. These activities and collaborations are of high risk and complexity, but are viewed as key resources for everyone. Worldwide data generation, cataloging, manipulation of climate data, pose a number of unique and difficult challenges, requiring many levels and types of interaction with DOE, NASA, NOAA, and NSF sponsors, as well as national and international climate program stakeholders, such as Atmospheric Radiation Measurement (ARM), Carbon Dioxide Information Analysis Center (CDIAC), BADC, ANU, and DKRZ, in addition to LLNL management and programming staff interactions at national and international climate sites.

From experience, we know that climate-science archives require a high degree of interoperability in order to integrate simulation, observation, and reanalysis data. The activities and special events are coordinated and organized under different domain projects that represent management and strategies, information technology, and climate science and data validation (include uncertainty quantifications and testbeds). For PCMDI software development, at the core of all underlying projects is the infrastructure and framework to support scientific analysis of large distributed data sets. Many of the informatics problems that need to be resolved are already well characterized, but it is essential that the solutions developed deal with managing and analyzing large-scale data flows.

The federated assembly of the CMIP5 archive is currently underway. Leading the way in large-scale data coordination, PCMDI has begun the process of assembling data for the next IPCC Fifth Assessment Report (AR5)—and, more importantly, science on this data is being collected. To accomplish our goal, PCMDI has led and coordinated the development of ESGF's international team of climate and computational scientists who are developing and delivering a production-quality environment for managing and accessing of ultrascale climate data. In this environment CMIP5 data will be augmented with access to other important data sets such as NASA's AIRS, Microwave Limb Sounder (MLS), Tropospheric Emission Spectrometer (TES), and CloudSat level 3 satellite observations; ARM data in the form of Climate Modeling Best Estimate (CBME)²; and NOAA's reanalysis data. Also included in the mix will be a number of diverse data sets representing downscaling and upscaling, such as CORDEX.

In data analysis and visualization, PCMDI leads and co-leads development of the slated UV-CDAT and the Visual Data Exploration and Analysis of Ultra-large Climate Data projects, respectively. These projects involve design, development, and deployment of an advanced and parallel-capable visual data exploration and analysis software infrastructure. This will meet the climate science mission of the Office of Biological and Environmental Research (BER) and the needs of the global climate community. Both projects will make use of the UV-CDAT framework and seamlessly integrate other software packages such as VisIt, ParaView, VisTrails, and R. Coordinating software design and engineering between the two projects is demanding but delivers to the community a unifying UV-CDAT software framework. A UV-CDAT software layer, as well as its Graphical User Interface, is being developed to directly access data from the ESGF distributed archive. This will allow users to directly browse, search, discover, and download distributed project data to their laptops/desktops for local manipulation. Conversely, UV-CDAT will be included as part of the ESGF back-end node configuration for remote execution and data manipulation. Both are shown in Figure 1.

Building on the work done for ESGF, we will be leading the effort to develop a data infrastructure and testbed component for the Climate Science for a Sustainable Energy Future (CSSEF) project. This effort

² CMBE is a new ARM data stream specifically tailored to climate modelers for use in evaluation of global climate models. It contains a best estimate of several selected cloud, radiation and atmospheric quantities from the ACRF observations and Numerical Weather Prediction (NWP) analysis (for upper air data only) [CMBE 2011].

involves seven laboratories (ANL, LANL, LBNL, LLNL, PNNL, ORNL, SNL) and a NSF facility (NCAR). Our primary assignment will be to research and develop software components that will fit into the ESGF infrastructure. More specifically, we will work with other organizations to develop a uncertainty quantification diagnostic testbed within the ESGF infrastructure for assessing new modeling components (atmosphere, land, ocean, and sea ice). The testbed and data infrastructure for the CSSEF project will also include modifications to the UV-CDAT analysis framework. Work has begun in UV-CDAT to look at new ways to visualize UQ in collaborations with the Polytechnic Institute of New York University and Kitware, Inc.

In the very collaborative, decentralized, problem-solving environment of ESGF, a faster network must be in place soon—one that is certain to achieve a minimum of 10 Gigabit per second (Gbps) for our petabyte-distributed archives. Climate and computational projects such as CMIP5, CSSEF, ESGF, and UV-CDAT have already incapacitated current networks. For better network performance, PCMDI is working with both national (e.g., Climate100) and international (e.g., Climate Data Network) groups to discover more efficient ways of transporting massive amounts of climate data quickly. We are partnering with LBNL and ORNL to test the emerging 100 Gbps network testbed's move of large-scale climate data, work that aligns with Energy Sciences Network's (ESnet's) Advanced Networking Initiative. The state-of-the-art network coupled with transport and management technologies will make realistic, at-scale experimentation of transport services and data management within the ESGF infrastructure possible. Along with Climate100, PCMDI has initiated the international Climate Data Network group, formed to track the status and availability and speed of international networks. If we are to move large portions of data from one location to another around the world, we will need this group to keep close track of networks, routers, and other related network resources.

As an integral part of the federated system, the success of the climate projects we have mentioned necessitate that applications, infrastructure, and middleware/networks function at their peak. Currently, our ESGF-connected testbed sites in the U.S. include LBNL, LLNL, NCAR, ORNL, and PNNL (with ANL, LANL, and SNL joining the list soon). These sites have all established ESGF data nodes where data and resources are also located. For international sites, we have established connections and tracking network progress with BADC, DKRZ, ANU, and the University of Tokyo. The Climate Data Network mailing list contains over 50 domestic and foreign network specialists and scientists concerned with network performance and data transport issues. In the U.S., ESnet is the primary provider of network connectivity for the Department of Energy (DOE) Office of Science and maintains the mailing list for the climate community at climatedata@es.net.

4 Hardware

In the first half of 2011, Livermore Computing (LC) procured a storage system on behalf of PCMDI. This entailed about one month of work effort in LC, encompassing vendor meetings, technology evaluations, developing selection criteria, writing the statement of work, evaluating the bids, and bringing in the new system. More work is needed to eventually bring the system online for scientific use. This storage system will be used exclusively for PCMDI's climate community projects, such as CMIP3, CMIP5, CSSEF and others. The cost of the system itself and its ongoing maintenance and support is the sole responsibility of PCMDI.

The main requirements described by PCMDI were 2.0 PB or more of usable space, high reliability, and low cost. To address reliability, LC focused on hardware RAID solutions with which PCMDI could layer ZFS software RAID. According to LC, this combination of hardware and software RAID has proven to be extremely reliable and stable in the current Green Data Oasis (GDO) environment, and we as PCMDI wanted the storage system to continue with this architecture type.

While procuring the storage and the necessary servers and peripherals at this cost presented a challenge, LC was able to meet the primary requirements and work within this reduced budget. The final system provides 2.5 PB of raw space, which we expect to provide slightly more than 2.0 PB of usable space. The

DataDirect Networks (DDN) systems can be expanded up 1.1 PB simply by adding 3 TB disk drives. Dividing the storage into two pools and using two separate Network File System (NFS) servers with large memory and multiple network cards will provide the PCMDI Linux application servers with excellent NFS performance (see Figure 2). These servers will be used to host ESGF nodes and be load balanced to smoothly distribute the workload of tens of thousands anticipated users.

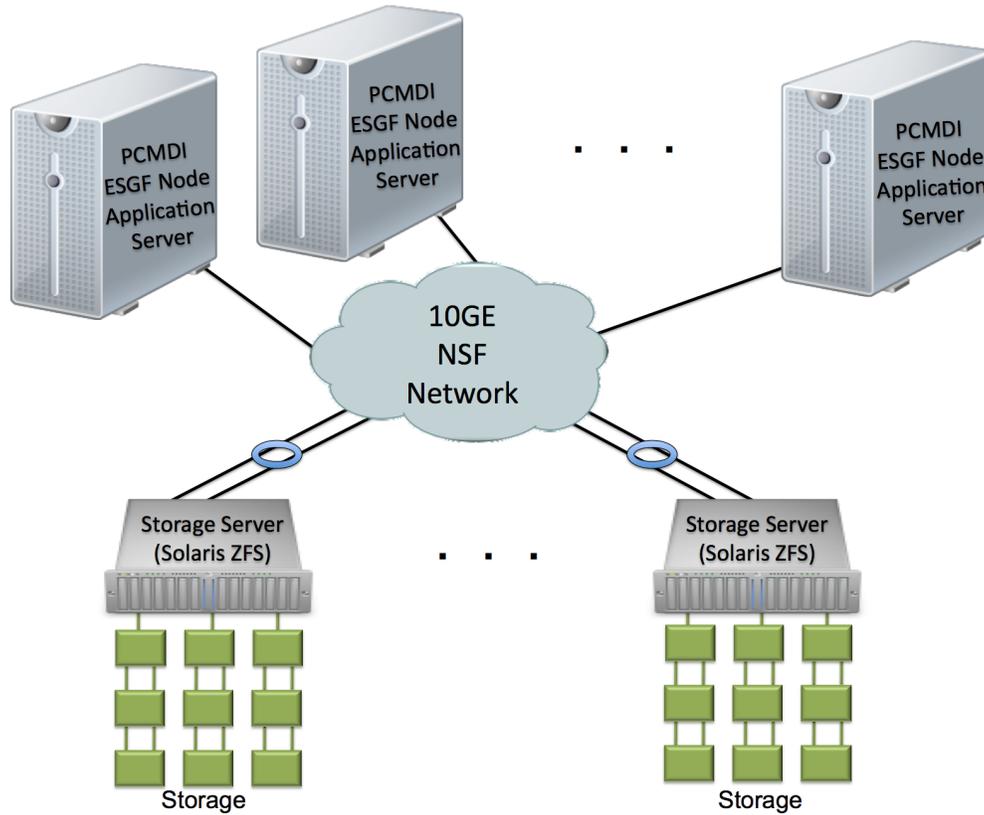


Figure 2: New storage facility for housing incoming climate data

4.1 Procured Items

Quantity	Item	Cost
2	DDN S2A9900 storage units, each with 420 3TB drives for a combined total of 2.5 PB of space, with easy expansion capability up to a total of 3.6 PB. Final contract was signed on May 12, 2011.	\$689,262
1	Year of support for DDN hardware	Included
2	NSF/Storage servers with 128GB of RAM each for running ZFS and providing NFS service. Includes solid-state drives for improving NFS caching performance. Requisition was issued on May 31, 2011.	\$75,287
4	10 Gbps network cards and cables from Myricom for the NFS servers	\$4,950
1	Year support for Oracle for OS support (Solaris, ZFS) on the storage servers	\$8,000
	Total	\$777,499

4.2 Future Ongoing Costs (Responsibility of PCMDI)

Item	Yearly Cost
Annual maintenance cost for DDN storage systems	\$34,088
Annual software support for Oracle (Solaris, ZFS)	\$8,000
LC system administration costs	0.1 FTE

Note: An option to contract with DDN allows purchases of multiple years of maintenance at the above rate. This option expires in May 2012.

4.3 System Description

The climate storage system will be located in B439 on the existing GCE (Green Collaborative Environment) network, where the current GDO storage unit and the GLCC/ALICE cluster currently reside. The GCE is on its own segment of the green network—a direct offshoot from the Lab's ESNet router—on a 10-Gbps network connection. Data in this environment must be reviewed and released for unlimited distribution if generated onsite. Externally produced data must bear no legal responsibility for the Lab to protect it. The long-term expectation is to move the GCE into the Livermore Valley Open Campus (LVOC)³.

The climate storage system will be accessed via a private 10-Gb NFS network between the two NFS servers and the various PCMDI Linux application servers. Two front-end NFS servers will each provide about 1 PB of usable storage via dual 10 Gbit networking and up to 4 8-Gbit FC connections to the storage. Storage servers can also provide direct access via applications such as the File Transport Protocol (FTP) if needed via Solaris containers, such as is done with GDO currently.

The network location of the PCMDI application servers is yet to be determined. We (PCMDI) have been given the option of placing the servers on the GCE network, but this requires changes on LC's part as outlined by LC's network manager. If those changes cannot be made, the PCMDI servers will remain as they are on the unrestricted network. In either case, it is likely that a 10 Gb private network would be established between the PCMDI application servers and the PCMDI storage servers for NFS support as shown in Figure 2.

5 Future

Through efforts such as CF, CMOR, ESGF, and UV-CDAT, PCMDI's software development has earned wide recognition for providing world-class resources and analysis tools for climate data. Building on that reputation, we created a collaboration of multi-institutional cyber-infrastructures serving resources and products worldwide. The result is a trusted, federated system of *science gateways* and *client analysis tools* for general public consumption. Six current and future programs and projects exist at the center of PCMDI's software develop activities:

- The Coupled Model Intercomparison Project, Phase 5 for scientists contributing to the IPCC Fifth Assessment Report (AR5)
- The Climate Science for a Sustainable Energy Future (CSSEF) for scientists developing a new climate modeling methodology to reduce uncertainty through the provision of a testbed

³ With the support of the U.S. Department of Energy's National Nuclear Security Administration and Office of Science, Sandia National Laboratories/California and Lawrence Livermore National Laboratory are working collaboratively to create an open, unclassified research and development space called the Livermore Valley Open Campus (LVOC).

- The PCMDI Diagnostics for helping PCMDI scientists develop and test hypotheses on climate data; including product verification and validation, data analysis, and report writing
- The PCMDI-NASA for supporting scientists wanting to identify and access NASA-relevant observations useful for the upcoming IPCC AR5
- The ARM CMBE for climate modelers interested in inter-comparing cloud, radiation, and atmospheric quantities from Numerical Weather Prediction analysis with global climate models
- The NARCCAP for scientists working to produce high-resolution climate change simulations in order to investigate uncertainties in regional scale projections of future climate.

Our objective is to continue responding effectively to significant and unfolding *grand challenges* in data management, analysis, and visualization. This work supports both the crucial collaborative research and the communities that are stewards of all distributed, scientific processes associated with extreme-scale data sets. At the same time, the infrastructures and tools we are using transcend geographical and organizational boundaries. Developing and integrating a federated data repository from multiple sources (including fixed servers and a virtual network of servers) is recommended for future large-scale interdisciplinary scientific advances and is recognized as a major enterprise that necessitates careful, thoughtful design. The comprehensive ESGF infrastructure will federate many data centers into a virtual data repository. Use of UV-CDAT will provide a full suite of analysis and data manipulation tools and integrate model, observational, and reanalysis data. It will also provide model inter-comparison metrics, user support, and life-cycle maintenance for the fundamental climate community projects that form the cornerstone of PCMDI's mission.

6 Reference

- [CF 2011] CF. NetCDF Climate and Forecast (CF) Metadata Convention, 2011 Available: <http://cf-pcmdi.llnl.gov>.
- [CMBE 2011] CMBE. Climate Modeling Best Estimate, 2011. Available: <http://www.arm.gov/data/pi/36>.
- [CMOR 2011] CMOR. Climate Model Output Rewriter 2, 2011 Available: <http://www2-pcmdi.llnl.gov/cmor>.
- [CORDEX 2011] CORDEX. COordinated Regional climate Downscaling Experiment, 2011. Available: http://wcrp.ipsl.jussieu.fr/SF_RCD_CORDEX.html.
- [ESGF 2011] ESGF. Earth System Grid Federation, 2011. Available: <http://www.esgf.org>.
- [Gridspec 2011] Gridspec. A standard for the description of grids used in Earth System models, 2011. Available: <http://www.gfdl.noaa.gov/~vb/gridstd/gridstd.html>.
- [LAS 2011] LAS. Live Access Server, 2011. Available: <http://ferret.pmel.noaa.gov/LAS>.
- [LibCF 2011] LibCF. NetCDF CF library, 2011. Available: <http://www.unidata.ucar.edu/software/libcf/>.
- [Mosaic 2011] Mosaic. Contains all the connectivity information between tile grids, 2011. Available: http://www.unidata.ucar.edu/software/libcf/html/group_gs_mosaic_grp.html.
- [Overpeck 2011] Overpeck J. T., G. A. Meehl, S. Bony and D. R. Easterling, Climate Data Challenges in the 21st Century, Science 11 February 2011: Vol. 331 no. 6018 pp. 700-702, DOI: 10.1126/science.1197869.
- [PyDAP 2011] PyDAP. Python library implementing the Data Access Protocol, 2011. Available: <http://www.pydap.org>.
- [QC 2011] QC. Quality Control Level 2 Tools, 2011. Available: https://verc.enes.org/collaboration/projects/cmip5-qc/wiki/Qc_level_2?version=19.
- [Simulation 2011] ClimateEnergyFusion Simulation. Climate Energy Fusion Simulation, 2011. Available: <http://climate.llnl.gov/simulation.html>.
- [UV-CDAT 2011] UV-CDAT. Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT), 2011. Available: <http://www.uv-cdat.org>.