

Unidata Report related to the proposal

Establishment of a Multiplexed Thredds Installation and a Ramadda Collaboration Environment for Community Access to Climate Change Data

Authors: G. Aloisio, S. Fiore, O. Marra

This proposal aimed at establishing a data management platform for climate change data related to the Mediterranean area, at the High Performance Computing Laboratory (University of Salento, Lecce - Italy).

The hardware purchased with the UNIDATA grant is an IBM x3630 M3 system equipped with a large amount of RAM, a fast disk storage and a good processing power. In particular the server is configured with two Intel Xeon six core E5645 CPU 2.40GHz, 24GB of RAM and 8TB of disk storage in RAID 5 configuration.

The server manages a Virtual Machine-based environment hosting climate change datasets provided by CMCC and running a (multiplexed) THREDDS and RAMADDA installations.

The UNIDATA equipment has been used for scientific and academic purposes. The students attending last year (2010-2011) the "Advanced Data Management" course at the University of Salento learned through a series of seminars titled "eScience data management", about the Common Data Model and the NetCDF Java and C libraries, the ISO19115 and the ISO19139 standards.

The seminars topics provided them the proper foundations:

- to design and implement software accessing and analyzing climate change NetCDF data,
- to better understand the multidimensional nature of these datasets,
- to manage THREDDS and RAMADDA services.

The students were also supported by the CMCC personnel, which has a strong knowledge on (i) Virtual Machine environments (e.g. based on the ESXI software), (ii) THREDDS software installation and configuration, (iii) NetCDF format, libraries, tools. This helped a lot to quickly set up the environment plus the THREDDS and RAMADDA service learning more about the UNIDATA software both from an end-user and administration point of views.

A Dashboard system developed at CMCC and providing monitoring capabilities was also setup on the VM-based environment to monitor the deployed service instances.

The UNIDATA equipment is being used this year too (2011-2012) with a stronger focus on the NetCDF C library and parallel data mining applications exploiting MPI and OpenMP. The multicore platform is being used to implement a parallel data access software for NetCDF files.

The UNIDATA equipment established with this grant has represented a strong facility for the HPC Laboratory at the University of Salento, where students have been able to test and learn more about Unidata software like THREDDS, RAMADDA, IDV, NetCDF libraries.

There are several educational and research benefits coming from this experience.

In particular, the students:

- worked on a real environment, starting from the setup of the hardware to the configuration of the services and the implementation & test of the code;
- learned a lot about tuning a VM-based environment with multiple machines, cores, services and data.
- deployed and managed a multiplexed configuration for the THREDDS installation, analyzing the performance benefits coming from such an architectural choice in terms of load balancing. They used an apache service (*configured with the mod_proxy_ajp module*) with several tomcat service instances running behind it, to balance the load related to multiple incoming requests.
- designed and implemented some simple software applications running on climate changed datasets and exploiting the C and Java NetCDF libraries.

- worked on real data understanding in a concrete way the nature of these multidimensional datasets, as well as the concepts of variable, dimensions, metadata, etc.

This year, the students of the new course on Advanced Data Management taught at the University of Salento, inherited the UNIDATA environment already set up during the last academic year. So they are focusing their attention much more on (i) designing and developing data-mining applications which analyze NetCDF files and (ii) understanding the NetCDF storage structure, format and libraries (C and Java).

This UNIDATA proposal was a great and successful experience. Overall the students were so excited and interested in working on such challenging climate data management topics with strong scientific tools, software and libraries.