Project of Strategic Interest NEXTDATA

Scientific Report for the reference period **01/01/2012-31/12/2012**

Unit CASPUR

WP 2.5 - Archive of numerical simulations and projections

1. Scheduled activities, expected results and Milestones (as indicated in the Executive Plan)

The availability of results produced by numerical climatical, environmental and paleoclimatic simulations is an essential tool to complement and interpret the information provided by the measured data and develop methodologies for forecasting and impact estimating. The activities of CASPUR supercomputing center will contribute to the construction of a digital archive of the results of global and regional climate simulations, both existing and realized during the project time interval.

In terms of global climate conditions, in collaboration with the CNR-ISAC Unit, will be built an archives of the results of the simulation covering the industrial period (1850-2005) and future projections (RCP 4.5, RCP 8.5, RCP 3-PD) for the period 2006-2100, obtained using the global state-of-the-art numerical model EC-Earth installed and validated at CASPUR. The global simulations will be partly derived from simulations that have already been done and partly from new simulations of EC-Earth that will be implemented on CASPUR machines during the first year. CASPUR will also provide, whenever possible, support to the optimization of numerical codes.

In collaboration with the Unit ICTP and CNR-ISAC, CASPUR will support the production of simulations of ICTP regional areas of interest during the project, using the numerical model RegCM, possibly nested within EC-Earth; those simulations will be included in the digital archive. The archive will also include a series of simulation results of non-hydrostatic, high-resolution (1-10 km), mountain areas, where the complex terrain require the development of specific modeling solutions, made by CNR-ISAC and other structures , in part, on the machines of CASPUR, using the WRF numerical model. In these areas, the comparison between the numerical results and the data from measurement networks will be crucial for the calibration and validation of models at a local scale. These models will be engaged in global simulations to obtain an archive of future

scenarios with high spatial resolution in the mountainous regions of interest. CASPUR will also provide, whenever possible, support the optimization of numerical codes.

First year:

Participation in the census of the results of numerical simulations and re-analysis of global and regional simulations available at the partner of this project, the harmonization of protocols for the storage of huge numerical data and how to access the data and procedures for transferring large data sets and start the activities of storing existing numerical results datsets.

Participation in the formulation and preparation of specific numerical experiments to be carried out during the period of the project and obtain the first results of the new global simulations and regional areas of interest of the project.

By the end of the first year, in cooperation with the Unit CMCC, CNR-ISAC, ENEA UTMEA and ICTP, CASPUR will contribute to the development of the first version of the digital archive of numerical simulations.

2. Deliverables expected for the reference period

D2.5.1 (PM12): report on the census of the results of numerical simulations.

3. Activities which have been actually conducted during the reference period

3.1 Research activities

Nothing

3.2 Applications; technological and computational aspects

CASPUR has made available all the technology for high-performance computing, storage and data processing for the *state-of-the-art* applications of this project.
The infrastructure used for NextData includes a typical distributed memory supercomputer based on traditional X86 processors and a system server for the intensive I / O services (and for the external access service, virtualization and DBMS) which are interfaced to a Enterprise-class storage subsystem(Figure1).

Schematically, it follows the block diagram of the entire hardware infrastructure:

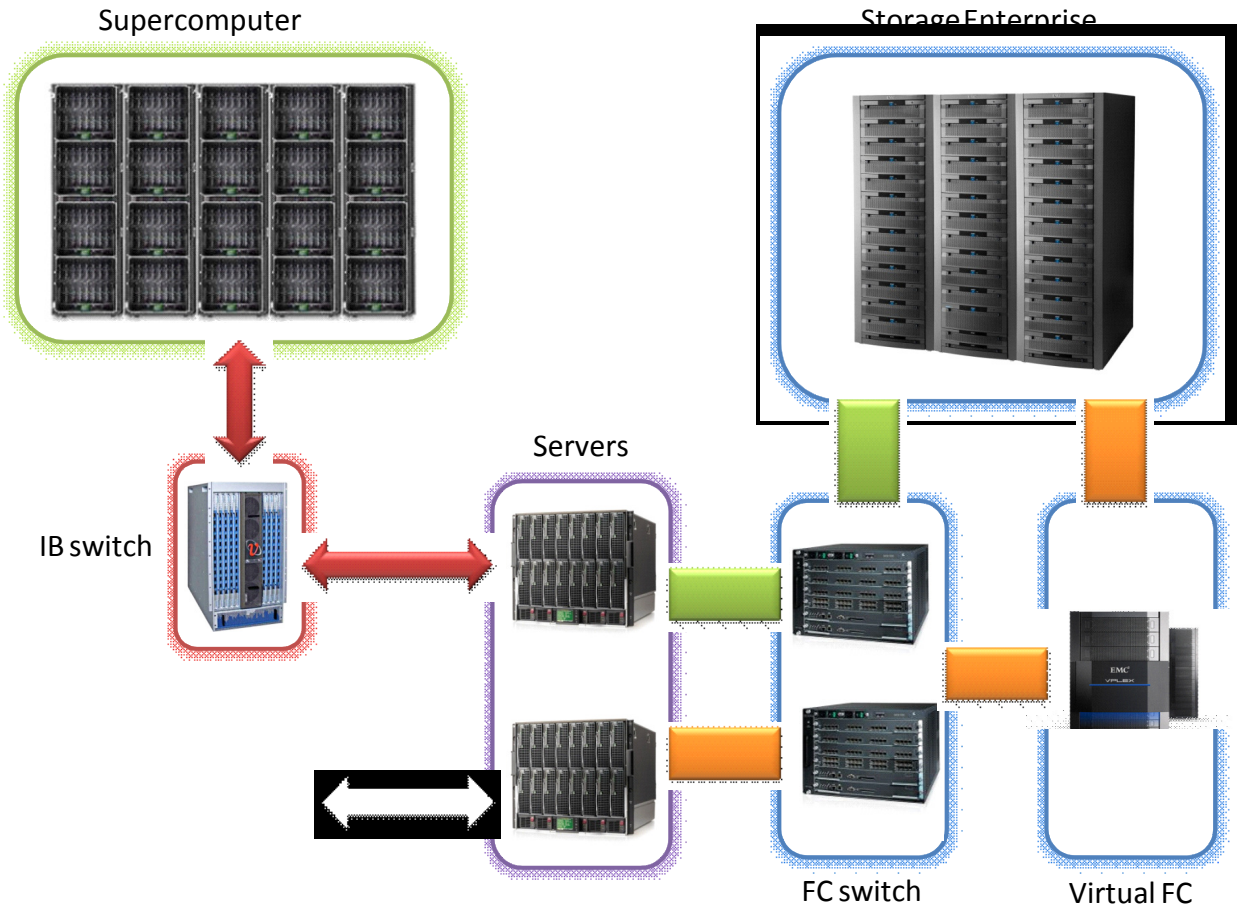


Figure 1. Block Diagram of the ICT / HPC for NextData.

This infrastructure of high-performance computing and storage has allowed us to support all of the simulations provided in CASPUR and start storing all of the datasets provided from simulations carried out elsewhere. The transfer of large amounts of data available elsewhere has greatly benefited from the research network GARR, in which CASPUR is one of the strategic access point. This synergy between CASPUR and GARR has provided and will guarantee users with a NextData available bandwidth, continuous and specialist support for the resolution of issues related to breaks, data transfer bottlenecks and massive datasets (of the order of terabytes).

To make the computational data, identified and / or products within the NextData project, available to the scientific community, it was decided to use the THREDDS Data Server (TDS) version 4.2, recently updated to 4.3, which is an acronym for THREDDS Thematic Real-time Environmental Distributed Data Services tool.

The TDS is a web server that provides access to data and metadata in science, using different protocols to access remote data, including OPeNDAP, OGC WMS and WCS, HTTP. Thredds data server relies on the open source Apache Tomcat web server, which is an implementation of the Java Servlet and JavaServer Pages. The way it works TDS is shown in figure 2:

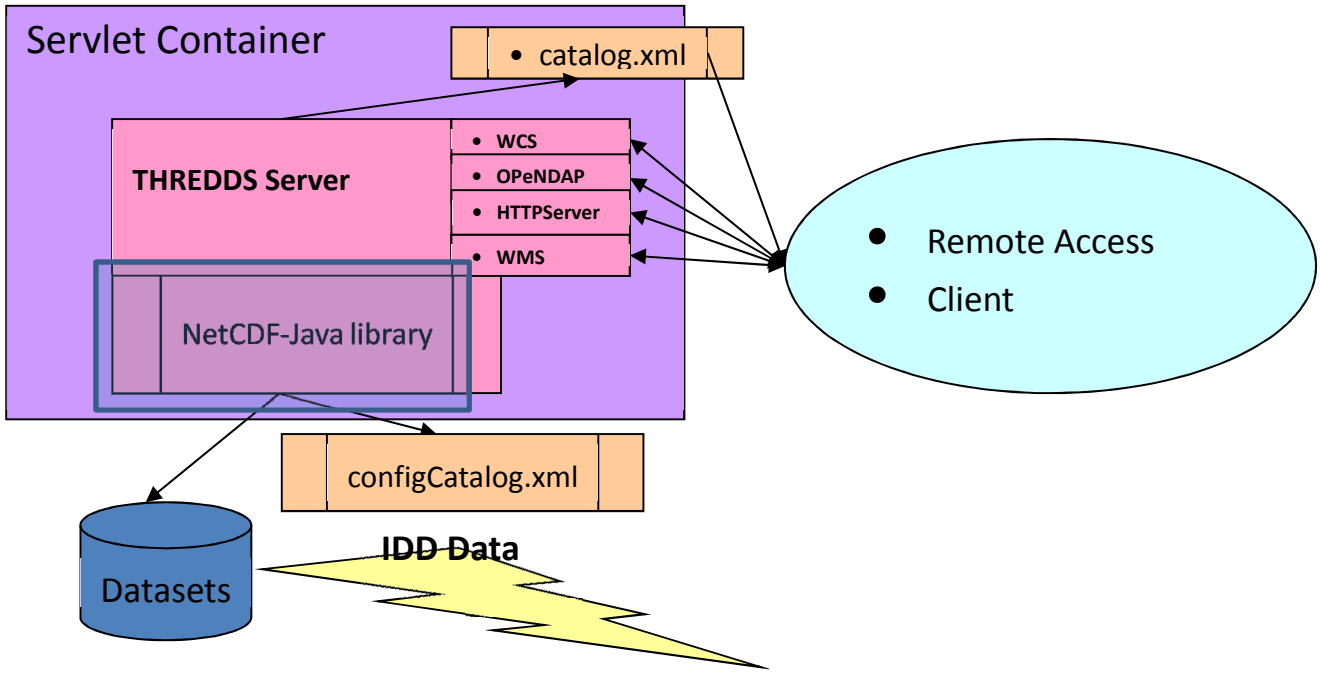
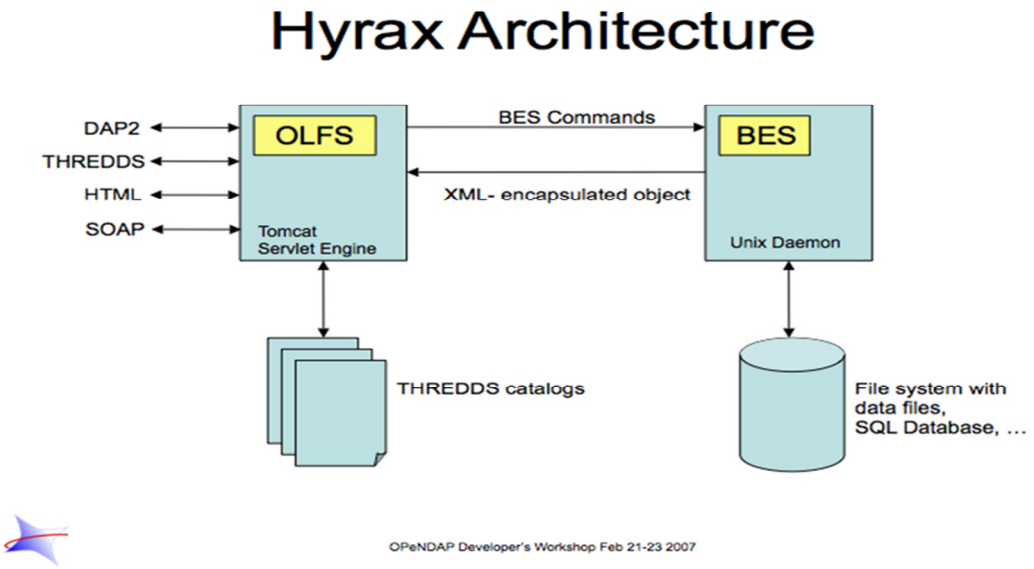


Figure 2. Scheme of the TDS server

At CASPUR was installed and tested 4.2.10 TDS OPeNDAP protocol; the version of Apache Tomcat is 7.0.30 (updated to 7.0.32) and Java version is the 1.6. The machine hosting the currently installed TDS is a server, namely bl102.caspur.it, which is a machine equipped with 4 Dual-Core AMD Opteron 2.8 GHz. The data are, however, are physically stored on Supercomputer storage and made available from a Lustre file system, which allows access from different computers to files on a remote computer over a network, potentially in a simultaneous way. More precisely, Lustre is a parallel distributed high performance file system capable of supporting cluster networks with thousands of nodes and petabytes of data to be stored. The file system resides on the Enterprise storage system with a of total capacity greater than 1PB described above. In particular, we implemented the structure described in the following figure, assuming that more than 60 TB of raw data are available offline (staging system).



Hyrax is an alternative name for the OPeNDAP Data Server

- OLFS (OPeNDAP Lightweight Front-End Servlet)
- BES (Back-End Server)

3.3 Formation

Nothing.

3.4 Dissemination

Nothing.

3.5 Participation in conferences, workshops, meetings

Nothing.

4. Results obtained during the reference period

4.1 Specific results (Data libraries, Measurements, Numerical simulations, etc)

Nothing.

4.2 Publications

Nothing.

4.3 Availability of data and model outputs (format, type of library, etc)

ICTP and CNR-ISAC data simulations are partially available onto CASPUR thredds Data Server.

4.3 Completed deliverables

Census Data arising from climate simulations is an ongoing activity.

5. Comment on differences between expected activities/results/deliverables and those which have been actually performed

Nothing.

6. Expected activities for the following reference period

When fully operational, the CASPUR Thredds Data Server should allow a simple, standardized (according to certain criteria) and efficient tool towards NextData datasets, both if those datasets are physically present in Rome (at CASPUR) or elsewhere, produced and /or stored by research groups of CNR-ISAC (Turin) and ICTP (Trieste) participating in the project.