EMS7/ECAM8 Abstracts,
Vol. 4, EMS2007-A-00420, 2007
7th EMS Annual Meeting / 8th ECAM

GRID distributed computation of nested climate simulations and data-mining. On behalf of the EELA project

V. Fernandez-Quiruelas (1), J. Fernández (1), A. S. Cofiño (1), C. Baeza (2), M. Carrillo (3), F. Garcia-Torre (1), R. M. San-Martín (3), R. Abarca del Rio (2), J. M. Gutiérrez (1), Rafa Mayo (4)

(1) Dept of applied mathematics and computing sciences, University of Cantabria. Spain, (2) Universidad de Concepción, Chile. (3) Servicio Nacional de Meteorología e hidrología, Peru, (4) CIEMAT, Spain

The EELA (E-infrastructure shared between Europe and Latin America) project aims at establishing a bridge between the existing e-Infrastructures in Europe and those emerging in Latin America, through the creation of an interoperable Grid Infrastructure - based on the RedCLARA and GÉANT networks - for the development and deployment of advanced applications in Biomedicine, High Energy Physics, e-Education and Climate. EELA is expected to help reducing the digital division in Latin-America, making available to researchers a high performance e-Infrastructure for advanced investigations, later extendable to a larger community of users.

The different activities included in climate prediction can be roughly classified in numerical model simulation and forecast production, and data access/exchange and data analysis (so called data mining).

Modern climate science deals with different sources of data both from oceanic, surface and upper air observing networks, and also from satellites. This data is geographically distributed in various locations and it is stored in different kind of systems and formats. On the other hand, an increasing number of global climate predictions are available from numerical atmospheric and oceanic model integrations (reanalysis projects, ensemble model and multi-model experiments, etc.), which anticipate the ocean and atmospheric description of future climate. These sources of data can jointly help to solve

many important problems, such as regional climate change projections, i.e., the effects of climate change on different regions of interest. To this aim, efficient problem-driven statistical analysis tools are required for discovering knowledge, or useful information, within the huge amount of information. Data mining and machine learning techniques have been developed in the last decades to deal with this task, and different alternatives have been studied to make easier the process in a distributed environment such as the GRID.

Within the EELA project, we have successfully adapted the Community Atmosferic Model (CAM) to run on the associated Grid testbed between Europe and Latin America countries, based on EGEE middleware. Due to long run experiments (months for a secular simulation) a metadata information on the status of the experiments and broken simulations can be restarted anywhere on the testbed. One of the features enabling the distributed computation is the file Catalogue, which manages the location of files available to the Grid and their replicas. On top of the global runs with CAM, we have implemented the possibility of running regional simulations by means of the Weather Research & Forecasting (WRF) high-resolution numerical model for any region on the globe. The distributed computing enviroment favours the generation of ensembles of simulations, which can mostly run unattended by the user.

As El Niño phenomenon is a key factor for Latin-American climate understanding, and as well very important in terms of economic impacts. here we will show an example of this application on a region sensitive to the ENSO phenomenon.