

Grid computing for wildfire danger assessment: porting RISICO in glite

Monday, February 11, 2008 5:10 PM (20 minutes)

RISICO works on a set partitioned on identical and algorithmically independent squared cells. Computing a cell requires "cell status" and meteo data and produces a "next status" as also the wanted output of the simulation. A RISICO run on the Italian territory with cells of 1 km^2 requires approximatively the computing of 330.000 cells and 150Mb of input data, producing 1Gb of output data in a 20 min of time on a common workstation. Finer simulations using 0.01 km^2 sized cells, leads to quadratically increased input and output data size as also needed computation time. We can afford these higher needs through Grid. We divide a run into a few tens of jobs, each using Storage Elements for input and output data, and the requirement that every job successfully ends in a given maximum time. This gets achieved thanks to a "job status monitor" program running at UI level which polls about jobs termination, retrieve outputs and resubmits failed or late to finish jobs.

3. Impact

The RISICO application running at WN level is a quite straightforward rebuild of the original C++ source code, with no need for particular libraries. A GRisico wrapper script takes care of launching the executable after downloading three needed input compressed archives: the list of cells to compute, their status, meteo data for the given cell-set. The actual Logical File Names are provided through the InputSandbox. End of job execution gets recognized from our job status monitor program by polling LFC catalog for output file existence. This permits to retrieve it earlier than the OutputSandbox, whose readiness needs official LB answer. After execution ends it uploads the new computed status and the output as compressed archived. Time statistics for every step are taken by the wrapper script and returned through the OutputSandbox. They'll be eventually useful when verifying the submission strategy.

URL for further information:

<http://www.cyclops-project.eu/>

4. Conclusions / Future plans

Tests on a production Grid environment with data from real case scenario (VO cyclops) leads to satisfying results. Attention must be paid however to an effective ranking requirements choice in jdl. Up to now failed or unfinished jobs get resubmitted on provenly fast queues. Plans are to refine the resubmission strategy in an attempt to ensure a maximum completion time. A further goal is to integrate RISICO with geospatial services for input and output data sharing.

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

Forest fires, Risk Assessment, Civil Protection, Disaster Management

1. Short overview

The RISICO application, use-case selected by the CYCLOPS project estimates distribution of wildland fire risk over the Italian territory helping civil protection agencies to plan the firefighting system. A short description of RISICO is presented, along with motivation for the Grid porting of the application. A submission strategy for both application input data and output retrieval is considered. A mechanism to ensure automatic resubmission of eventually failed jobs is also described.

Primary authors: Dr D'ANDREA, Mirko (CIMA Foundation); Dr DAL PRA, Stefano (INFN)

Co-authors: Dr GAETANI, Francesco (CIMA Foundation); Dr VERLATO, Marco (INFN); Dr FIORUCCI, Paolo (CIMA Foundation); Dr ANGELINI, Valerio (CNR-IMAA)

Presenter: Dr DAL PRA, Stefano (INFN)

Session Classification: Earth Science

Track Classification: Application Porting and Deployment