

Exploiting the Grid for collaborative ensemble studies using GENIE Earth System Models

A. R. Price¹, Z. Jiao¹, T. M. Lenton², G. Williams³, P. J. Valdes³, S. J. Cox¹ and the GENIE team⁴

GENIEfy: Grid ENabled Integrated Earth system model for the Community

This article can be found at www.e-science.soton.ac.uk/posters

Abstract

GENIE is a Grid-enabled framework for composing an extensive range of Earth System Models (ESMs) and simulating multi-millennial timescales, primarily for the study of ice age cycles and long-term human induced global change. Grid computing technology enables the flexible coupling of constituent models, execution of the resulting ESMs and the management of the data that they generate. We demonstrate how the Grid software deployed for the project is being used to execute large-scale simulations of an ESM that would not be feasible on a single computational resource.

Within the spectrum of GENIE models a simulation over multi-millennial timescales can require anything from 2 hours to ~1,500 hours of sequential compute time. To perform ensemble runs of the more computationally demanding models therefore requires a significant amount of resource. Whilst an individual project member may own sufficient resource on an exclusive basis, this may not be cost-effective if the resource is then only used in intermittent bursts with time in between idle as the experimental results are analysed, reported, and the next experiment designed. However it is very cost-effective if a multi-disciplinary, multi-site team such as GENIE, each of whom has similarly intermittent need for bursts of compute power, pool their resources.

The GENIE data management system provides programmatic access to a dedicated database and file repository from within the Matlab problem solving environment. Exploiting the database system we demonstrate how large scale simulations can be performed by staging intermediate results in the repository and using a number of distributed "worker" clients to autonomously progress the model states. Through the common system interface and integration into the scientists' familiar tools / environments we enable the project team as a whole to cost-effectively undertake much larger studies than could have been contemplated individually.

Introduction

GENIE is a Grid-enabled modelling framework GENIE that allows the flexible coupling of constituent models of varying resolution, dimensionality and comprehensiveness to form new integrated ESMs. Grid technology is a key enabler for the flexible coupling of constituent models, subsequent execution of the resulting ESMs and the management of the data that they generate.

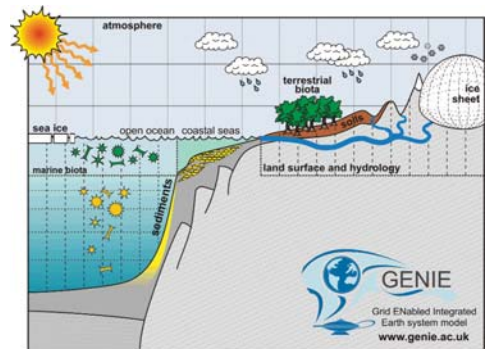


Figure 1: Schematic of the components modelled in GENIE.

Grid computing technology is required to ease the construction of new instances of Earth system model, automate the process of model tuning, speed up the execution of individual long integrations, enable large ensembles to be run, ease their execution, and feed and recycle data back into model development. It is particularly important to be able to study large ensembles of models, which are increasingly necessary for model tuning and data assimilation, statistical quantification of model error and investigative parameter sweeps.

¹ School of Engineering Sciences, University of Southampton, Southampton SO17 1BJ, UK
² School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

³ School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK
⁴ <http://www.genie.ac.uk/people>

Grid Software

GENIE has adopted software developed by the Geodise project (<http://www.geodise.org>) to build a distributed collaborative problem solving environment for the study of new Earth system models. The Geodise Toolboxes integrate compute and data Grid functionality into the Matlab and Jython environments familiar to scientists and engineers. In particular, we have built upon the Geodise Database Toolbox to provide a shared Grid-enabled data management system to provide a central repository for output from GENIE Earth system models.

The Geodise data model allows users to create descriptive metadata in Matlab / Jython data structures and associate that metadata with any file, variable or datagroup (logical aggregation of data) archived to the repository. The Geodise XML Toolbox is used to convert the metadata to a valid XML document which is then stored and made accessible in the database. The GENIE database system augments this model by defining XML schemas that constrain the permissible metadata and improves subsequent query performance. The interface to the database is exposed through web services and all access is secured and authenticated using X.509 certificates. Through a common client members of the project can access the database both programmatically and with a GUI interface.

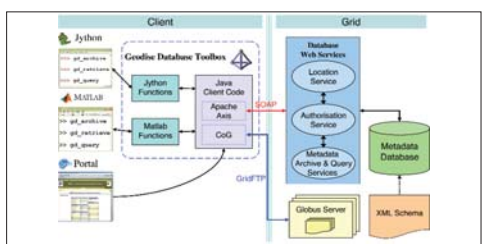


Figure 2: Schematic of the data management system

Ensemble Studies

The GENIE framework enables the composition of a large hierarchy of models capable of simulating the Earth system over millennial timescales. While many GENIE models can simulate such timescales in a few hours on a modern desktop the more complex models (e.g. GENIE-2) can require up to a week of sequential compute time to simulate 1000 years. To perform an ensemble of such models can therefore present a significant challenge to a single user.

We exploit the GENIE database and the restart capabilities of the models to break down ensembles of long simulations into manageable compute tasks. By staging the restart files in the shared repository all members of the project can access the ensemble study, query against the metadata in the database to find new work units, submit that work to compute resources available to them and upload new results as they become available. User's can target local institutional clusters, condor pools or remote resource available to them on the Grid (Globus Toolkit 2.4).

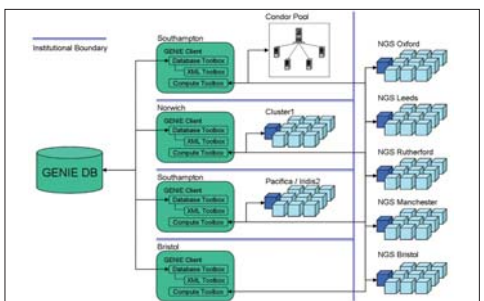


Figure 3: Collaborative study mediated by the GENIE database

Collaborative Study

The science team create an experiment "datagroup", a logical entity in the database describing the purpose of the study and acting as parent to a set of Simulation definitions. The simulation entities record the details for the specific instantiation of the model including the total number of timesteps to be performed and what output data to generate. This is all the information that is required and amounts to about 100 lines of configurable script for the scientist to edit. The system mediates the subsequent execution of the study.

To contribute to a study a project member just needs know the unique system identifier for the study as defined in the database. They then create a time-limited proxy certificate (to authenticate their actions on the Grid), retrieve a resource definition from the database and invoke an autonomous "worker" script. The "worker" allows a user to specify the number of timesteps to progress the members of the ensemble by and the number of compute jobs to submit to the specified resource.

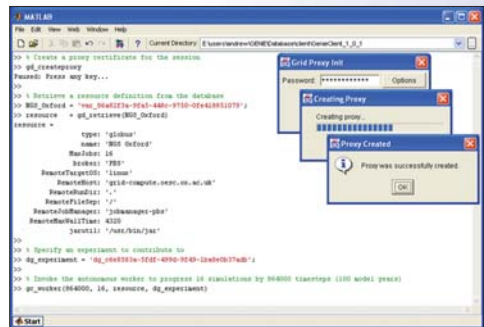


Figure 4: Typical user session contributing to an ensemble study

The unique system generated experiment identifier is the only piece of information that a user requires in order to contribute to a model study. Users inform the system of the amount of work they are willing to execute and the client does the rest. The worker script progresses through four stages of execution:

- 1) Updates the users local experiment setup
- 2) Post-processes any completed compute tasks
 - Queries the database for active jobs and polls their status
 - Post-processes and uploads job output to the database
- 3) Requests new work units from the database
 - Queries the database for a point-in-time assessment of the progress of the study
 - Prioritises available work units to progress the study evenly
- 4) Submits new work to the computational Grid
 - Retrieves the restart files from the repository and transfers the model to the remote resource
 - Submits a job to the specified resource and archives the details in the database

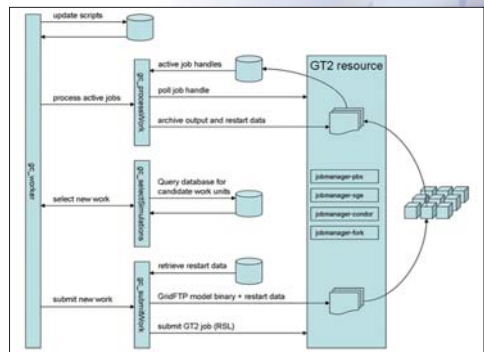


Figure 5: Diagram of the actions performed by the "worker" script

Performance

Twelve ensemble studies were executed by three members of the project using database client installations at their local institutions. In order to define the ensembles a single script was created for each study, specifying the appropriate metadata describing the simulations and uploading the model binary and input data and the model specific scripts for the experiment. The unique identifier for each study was circulated amongst members of the project willing to contribute resource and regular invocations of the client "worker" script were scheduled at each site to submit model runs to a number of computational resources around the UK. These resources included five nodes of the UK National Grid Service (<http://www.ngs.ac.uk>), three Beowulf clusters available at the Universities of Norwich and Southampton and a large Condor pool (>1,400 nodes) at the University of Southampton

The ensemble studies of the GENIE-2 model presented here were carried out during a 14 week period from December 2005 until March 2006. In total the system was used to perform:

- 352 simulations
- 3,736 compute tasks
- 46,992 CPU hours (some timings estimated)
- 428,000 GENIE-2 model years

The daily breakdown of resource usage is plotted in Figure 6. The rate of progression of the studies was largely determined by the amount of available work in the system at any given time. The overheads in performing queries on the database and the upload and retrieval of model data files did not adversely impact the performance of the system.

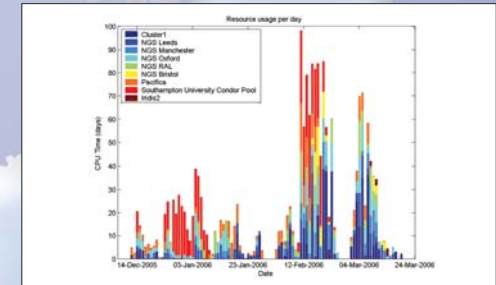


Figure 6: Resource usage for the 12 ensemble studies

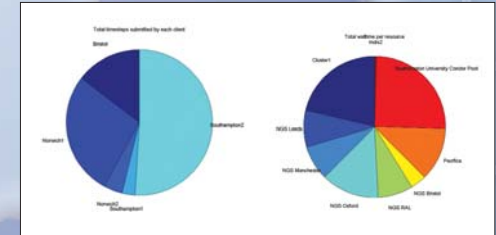


Figure 7: Distribution of client contributions and resource usage

The breakdown of client contributions reflects the amount of resource available to the user in each case. E.g. the client responsible for ~50% of the work was the only submission node on the large condor pool and was also used to submit jobs to the National Grid Service. The distribution of jobs across the computational resources illustrates that work was distributed evenly to the available resources. By mediating the study through a shared central repository the coordinating scientist has had the work performed over a collected pool of resource, much of which is behind institutional firewalls and probably not directly accessible to him/her. The system also enables us to introduce new resource as a study is performed and target the most appropriate platforms available.

Results

The effects of varying a fresh water flux correction parameter in the atmosphere component (IGCM) of a suite of GENIE-2 models with different ocean grid resolutions has been studied. Six of the ensembles are plotted in Figure 8 showing the strength of the maximum Atlantic overturning circulation for ensemble members across the 3 grids (32x32, 64x32, 72x72) initialised from states with default circulation (r1) and collapsed circulation (r0). The 64x32 surface grid avoids interpolation and needs the smallest flux correction. We are now tuning it to give a default climate state with the Atlantic MOC 'on'.

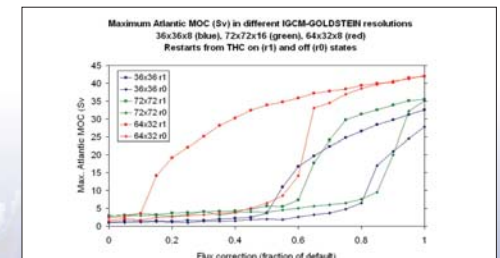


Figure 8: Maximum Atlantic overturning circulation (MOC) across six ensembles varying ocean grid resolution. Shows bi-stability of the Atlantic MOC under a dynamical atmosphere model in all resolutions tested.

Discussion

Fault tolerance The data management system provides the project with a resilient environment to perform distributed computing. Simulations are mediated through the database and are only progressed by the successful upload of completed work units into the repository. This means that the system is largely unaffected by any problems encountered by jobs on remote resources. If the configuration, submission, execution, post-processing or data upload fail for a work unit then the client system will detect a problem and attempt corrective action. If completed work units are not successfully uploaded to the database then the study is unchanged and the work can be attempted again.

Brokering In the absence of a brokering service on the UK National Grid our data management system provides a means to maximize our responsible use of the available resource. A user provides metadata about the resources they want to use including the maximum number of jobs that they would like to submit to the job manager (PBS, SGE, Condor) for any experiment. The database maintains a record of where a user's jobs are active and the client will respect the usage limits and refrain from submitting work to a resource if it has already reached its maximum number of allowed jobs. A user can then set up client invocations to automatically attempt to push work to a list of resources. Once each resource reaches its limit the client will move on to the next and the system therefore keeps all available compute power busy (assuming sufficient work exists) without exceeding the user's specified usage limits.

Conclusions

Members of the GENIE project collaborated to perform large-scale ensemble calculations on hierarchies of Earth System Model from the GENIE framework. Using a common Grid-enabled client interfacing to the shared GENIE database users pooled their resource to contribute to ensemble studies defined in the database. Armed with the unique system identifier for a particular study a user simply invoked an autonomous "worker" script specifying the amount of work they would like to contribute on a given resource. Through regular invocations of the client by multiple members of the project the ensembles were progressed through to completion. The system provides a robust and fault tolerant means to carry out large ensembles of lengthy model runs because progress is only achieved by the successful upload of data to the database.