

Grid-enabling the Midlands: The Midlands e-Science Centre of Excellence

Yi Zhang, Georgios Theodoropoulos
Midlands e-Science Center of Excellence
School of Computer Science
University of Birmingham
United Kingdom
{yxz | gkt}@cs.bham.ac.uk

Abstract – This paper provides a short overview of the UK e-Science Programme, introduces MeSC, the Midlands e-Science Center of Excellence, one of the newly established e-Science centers which, together with the National and Regional e-Science Centers, constitute the UK National Grid infrastructure.

Key-Words - e-Science, Grid, Midlands, Distributed Simulation

1 Introduction

Science is increasingly done through large-scale international collaborations and information sharing enabled by the Internet using enormous data collections, terascale computing resources and high performance visualization. The term e-Science refers to this type of large scale science and engineering global distributed collaborations ('collaboratories'). According to Dr John Taylor, Director General of the UK Research Councils, *'e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it.'* The infrastructure needed for e-Science is known as The Grid.

Grid computing provides a new platform for distributed computing allowing people to treat pools of computers, storage systems and networks as one big virtual computing system. The technological challenge of the Grid is to ensure transparent and secure access to heterogeneous computing resources. The aim is to design the underpinning hardware and software architecture - the grid middleware (e.g. Globus/OGSA/Condor) – that will ensure interoperability of components and resource management, enabling global e-Science collaborations [1].

2 The UK e-Science Programme

The UK e-Science Core Programme [2] is an initiative funded by the UK research councils (EPSRC, PPARC, BBSRC, NERC, ESRC, MRC) and the DTI. Central to the programme is a network of e-Science Centres, each operating under their own IT policies, whose goal is to enable uniform, secure access and load balancing on The Grid. Each Centre 'donates' equipment and network connectivity to the UK Grid to enable test bed experiments.

The UK e-Science Programme was announced in November 2000 and has, so far, gone through two phases.

The first phase started in 2001 and lasted for three years. In 2000 UK Government's Spending Review, the Office of Science and Technology (OST) was allocated £98M for a 3-year e-Science R&D Programme (2001-2003). This initiative was meant to be a cross-council activity, which involved all the Research Councils. Each Research Council had to make a specific allocation for the programme. The funding information for each year and funding source can be found in Table ??.

The main goals of the first phase of the e-Science Programme were to: establish a national e-science grid based on a network of e-science centers; promote the development of grid software and support the e-Science pilot projects of the different Research Councils; set up a National and a number of Regional e-Science centers.

In 2004 OST provided an additional £115 million over two years (2004-2005) to extend the work of the UK e-science Programme. This second phase focuses on "research challenges in processing, communication and storage of large volumes of data; provide generic solutions to needs of individual disciplines and applications; establish best practice across disciplines; provide infrastructure and facilities needed for next major stages of international collaborative research; and build on the leading international role" [3]. During the second phase of the programme, more e-Science Centres were established, this time focusing on specific themes of excellence.

3 The UK National e-Science Grid

The UK national e-Science Grid consists of a number of e-Science Centres, several working groups, a Grid support Center and production nodes (Figure 1). The next sections give a brief introduction of the structure of the UK national e-science grid and its software and network infrastructure.

3.1 The e-Science Centers

The National e-Science Centre [4] is situated in Edinburgh and is a joint venture between the Universities of

Research Councils	2001	2002	2003	2004	2005	Total
MRC	1	2	5	6.9	6.2	21.1
BBSRC	1	2	5	5	5	18
NERC	1	2	4	4	4	15
EPSRC	6	12	22	17.2	19	77.7
PPARC	3	8	15	16.4	15.2	57.6
ESRC	0	1	2	5.5	5.1	13.6
CLRC	1	1	2.5	2.5	2.5	10
Total	13	29	55.5	57.5	57.5	213

Table 1: UK e-Science Funding (£M)



Figure 1: The UK Grid

	Compute Clusters	Data Clusters
Processors	64 dual node Intel 3.06 GHz (1MB cache)	20 dual node Intel 3.06 GHz
Memory	2GB	4GB
Disks	2x 120GB IDE disks (1 boot, 1 data)	2x 120GB IDE disks
Network structure	Gigabit network Myrinet M3F-PCIXD-2 Front end (as node)	Gigabit network Myrinet M3F-PCIXD-2 Front end (as node)
External Storage	Disk server (as node) with 2x Infortrend 2.1TB U16U SCSI Arrays (UltraStar 146Z10 disks)	18TB Fibre SAN (Infortrend F16F 4.1TB Fibre Arrays (UltraStar 146Z10 disks)
Operating System	RedHat ES 3.0	RedHat ES 3.0
Software	PGI compilers and Intel Compilers	Oracle 9i RAC (16 CPU license) Oracle Application server (16 CPU license)

Table 2: Specification of the Production Grid

Belfast, Manchester, Cardiff, Cambridge, Oxford, Southampton and Imperial College London host the Regional Centres. Besides these basic eight regional centres, CLRC also established an e-Science Center to develop grid-enabled infrastructure for their scientific facilities and lead the UK Grid Support Centre.

As part of the second phase of the e-Science Core Programme, seven new e-Science Centres of Excellence were established. They are at the universities of Birmingham, Bristol, Lancaster, Lancaster, Reading and UCL and the White Rose Grid, which includes Leeds, York, Sheffield.

3.2 The e-Science Working Groups

To address various technologic issues related to the development of the grid, several working groups have been established:

- *Engineering Task Force* (ETF): focuses on the construction, testing and demonstration of different grid technologies for the UK national grid.
- *Security Task Force* (STF): addresses security issues of for the UK national e-science grid.
- *Architecture Task Force* (ATF): works on long-term architectural issues for the UK's contribution to e-Science infrastructure
- *Database Task Force* (DBTF): deals with developing software components for database access and integration with the Grid.
- *Grid Network Team* (GnT): helps with network development during grid construction.

3.3 The Grid Support Centre

The UK Grid Support Centre is funded by the Research Councils' Core e-Science programme. The centre supports all aspects of development UK e-science Grid, from deployment, operation to maintenance and distribution of Grid resources. Most importantly, the support centre functions as the UK e-Science Grid Certification Authority for the UK e-Science community.

3.4 The Production Grid

At present the UK e-Science Programme is entering a production phase. A production grid with four clusters, two compute and two data clusters, is established. The Clusters are hosted by the University of Manchester, CCLRC (Rutherford Appleton Laboratory), University of Oxford and the White Rose Grid and are available to users from April 2004. The configuration of the production clusters is given in Table 2.

3.5 Networking Infrastructure

The UK e-Science Grid relies on the UK University network SuperJanet4 as its network infrastructure. The backbone bandwidth of SuperJanet4 is currently 20Gbps.

3.6 Middleware

The current middleware standard for the UK e-Science Grid is Globus toolkit 2. The e-Science Grid will move to GT3/GT4 in the near future. Furthermore, testbeds are being set up to evaluate the OGSi infrastructure and identify migration paths for existing projects.

As part of this activity, an e-Science OGSA testbed is investigating the migration path from GT2 to OGSA. The project involves the Daresbury Laboratory of CLRC and the Universities of Manchester, Reading, and Westminster. The aim is to install and test the middleware based on the Open Grid Services Architecture (OGSA) [5,6] and deploy a number of applications on the testbed.

4 MeSC: The Midlands e-Science Center of Excellence

The Midlands e-Science Centre (MeSC) [7] is one of seven Centres of Excellence established during the second phase of the UK e-Science programme to augment the existing national and regional e-Science Centres. The purpose of these Centres of Excellence, as envisioned by

the e-Science Core Programme, is to:

- Add and coordinate expertise in technologies and applications important to e-Science
- Add experience and resources to the existing UK e-Science Grid
- Add regional coverage to the existing UK e-Science Grid
- Host Access Grid nodes for multi-site remote conferencing



Figure 2: The Midlands e-Science Center

The Midlands e-Science Centre of Excellence (Figure 2) is a virtual centre with a base in the School of Computer Science, University of Birmingham, and involvement of a large number of departments from across the University of Birmingham. The Centre links to e-Science activities at the University of Warwick (which also hosts an Access Grid Node) and has partnership with the Universities of Coventry and Wolverhampton. It receives support from the "first tier" e-Science Centres on the National UK Grid.

4.1 Area of Excellence

The Centre focuses on "Modelling and Analysis of Large Complex Systems" as its area of excellence. It pools together wide-ranging expertise in data- and CPU-intensive techniques for computational modelling, enabling multi-disciplinary collaboration on a scale not previously realised in the region. Its uniquely distinguishing aspect is the focus on long-term research into developing effective algorithms and software technology (simulation, numerical computation and evolutionary computation methods) for the Grid and their application to complex problems leading to scientific advancement or industrial benefit. MeSC has four primary aims:

- To provide accessibility and connectivity for the Grid for the Midlands region.
- To create a focal point for the activity of compu-

tational modelling of very large complex systems and a source of relevant expertise for industry.

- To enable long-term research into numerical algorithms and simulation techniques for the Grid.
- To foster collaboration between different academic disciplines in science, engineering and industry.

4.2 Specialist Themes

There are four specialist themes of research in the Centre:

Simulation of evolving systems of interacting elements.

This theme focuses on problems where the computational task is to simulate the dynamics of systems comprised of large numbers of interacting computational elements to observe their emergent behaviour. The theme encompasses a wide spectrum of simulation approaches, from discrete element method to cellular automata, neural networks and agent-based discrete event simulations.

Large-scale grid-enabled distributed simulation and interoperability.

Work in this area focuses on: building Grid-aware distributed simulation engines, addressing issues related to resource management, load balancing, synchronization protocols; developing discrete event simulations of large complex systems in different domains; Federated simulations and interoperability middleware (such as the High Level Architecture - HLA/RTI).

Distributed algorithms and compact storage for mathematical solutions of large complex systems.

This theme focuses on: the modelling and simulation of complex engineering and scientific problems using the full spectrum of numerical and simulation approaches: deterministic, Monte Carlo, linear or non linear, search and optimization, eigenvalue, etc; parallelisation techniques; techniques for efficient data storage and effective distribution.

Data mining and large-scale visualisation. This theme encompasses all work of the center related to Grid-aware data mining and visualisation. Issues addressed include: data cleaning, transportation integration and reduction, languages, mining association rules, classification and prediction techniques, cluster analysis, multidimensional data analysis, and the corresponding visualisation techniques.

4.3 Infrastructure-The Midlands Grid

The University of Birmingham is the Regional Network Entry Point (RNEP) to JANET for the Midlands Mid-MAN network, and has recently made a substantial investment in IT infrastructure (£3.14M SRIF grant supplemented by £1.67M of University funds), facilitating the installation of a campus-wide Gigabit Ethernet network and a large-scale visualization system established

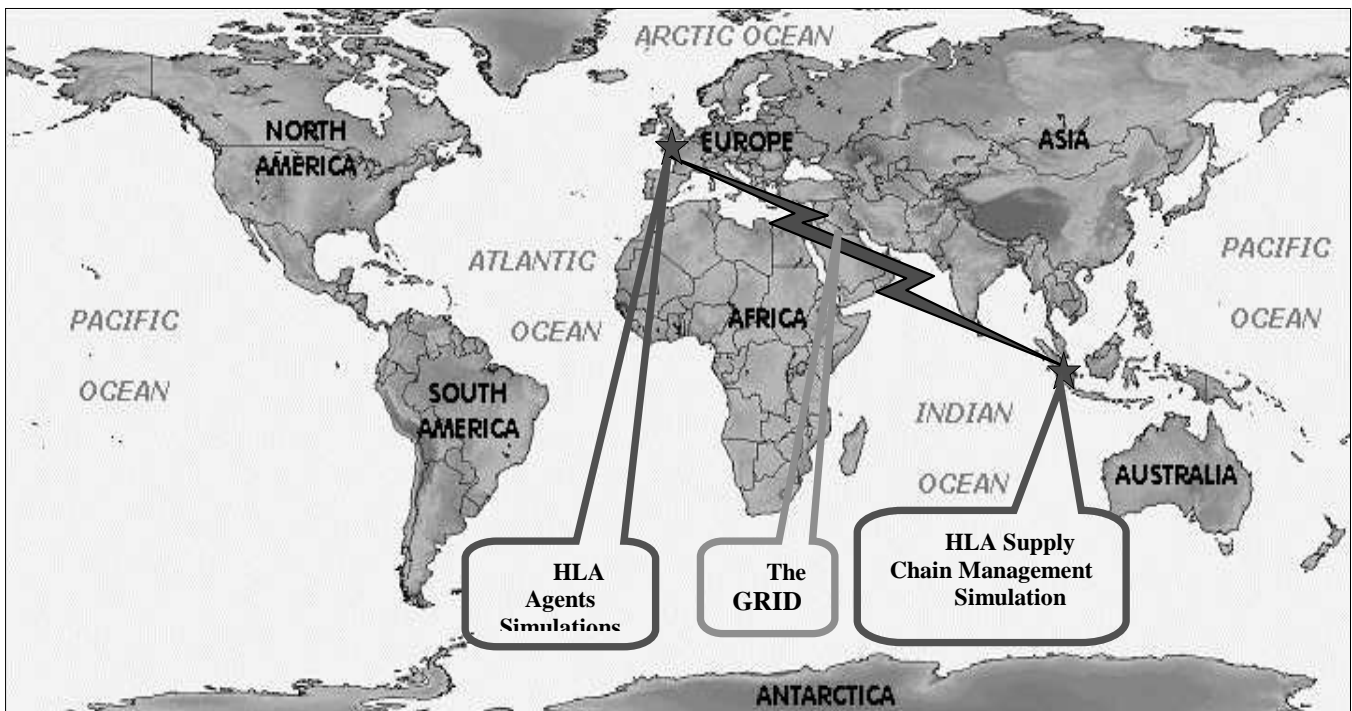


Figure 3: The DS-Grid Project

with the support of Hewlett Packard. The Network is based on Foundry hardware, and fiber or Category 6 copper cabling while the visualization system is an SV6 system (Mindstorm) with 5 HP J-6700 dual processor nodes. The Centre hosts three Access Grid Nodes (two in Birmingham and one in Warwick) running the InSORS software. With regard to compute resources, in addition to the University's Central Applications Service (a cluster of 6 dual-processor HP J6700 workstations running at 750 Mhz under HP's HP-UX) and the various clusters available in the different departments involved, a dedicated HP cluster (52 compute Zeon nodes, 1 TBy storage, Myrinet Switch) is currently being procured, while future University research strategic plans include a 1000 node cluster for MeSC).

4.4 Current Research and Projects

There are currently several e-Science R&D projects funded under the UK e-Science Core Programme or the Research Councils involved. The National e-Science Center maintains an up-to-date list of these projects in its www site [1]. MeSC and its partners participate in a number of e-Science projects including GridPP (a collaboration between particle physicists and computing scientists from the UK and CERN to build a Grid for Particle Physics), e-Tumor (an EU-funded project to develop a Web/Grid support system for brain tumours diagnosis and prognosis), an e-Science Pilot Project in Integrative Biology and projects on Grid resource management and resource prediction undertaken by Warwick [1].

The DS-Grid Project: Large Scale Distributed Simulation on the Grid. A distinctive and unique feature of

MeSC within the UK e-Science Grid is its focus on Distributed Simulation as one of its specialist themes. MeSC builds on strong expertise in this field in the School of Computer Science at the University of Birmingham (SysMoS Group).

Distributed Simulation (historically referring to the execution of discrete event simulation models onto parallel and distributed platforms) has in the last decade witnessed an explosion of interest and innovation, not only for speeding up simulations but also as a strategic technology for linking simulation components of various types (e.g. discrete or continuous, numerical or discrete event etc.) at multiple locations to create a common virtual environment (e.g. battlefields, virtual factories and supply chains, agent-based systems, games etc). The culmination of this activity has been the development of the High Level Architecture (HLA), a framework for simulator interoperability [18]. The High Level Architecture (HLA) for Modelling and Simulation was developed as an IEEE standard to facilitate interoperability among simulations and promote reuse of simulation models. Using HLA, a large-scale distributed simulation can be constructed by linking together a number of geographically distributed simulation components (or federates) into an overall simulation (or federation).

The development of such complex simulation applications usually requires collaborative effort from researchers with different domain knowledge and expertise, possibly at different locations. Furthermore, these simulation systems often require huge computing resources and the data sets required by the simulation may also be geographi-

cally distributed (e.g. in a supply chain simulation involving different companies, the most up-to-date data will be in the individual companies.). In order to support collaborative model development and to cater for the increasing complexity of such systems, it is necessary to harness distributed resources over the Internet.

The Grid provides an unrivalled technology for large scale distributed simulation as it enables collaboration and the use of distributed computing resources, while also facilitating access to geographically distributed data sets. The vision is a "Grid plug-and-play distributed simulation system", a distributed collaborative simulation environment where researchers with different domain knowledge and expertise, possibly at different locations, develop, modify, assemble and execute distributed simulation components over the Grid. However a number of important new research challenges have to be addressed before this vision is realized, including support for collaborative development of simulation applications, mechanisms for fault-tolerant, coordinated, secured simulation executions, advanced model and service discovery mechanisms, novel resource management and load balancing mechanisms to meet the different requirements of the different models etc.

The DS-Grid Project aspires to address the above challenges. It is a collaborative project between the Midlands e-Science Center of Excellence (MeSC) at the School of Computer Science, University of Birmingham (Dr Georgios Theodoropoulos), the School of Computer Science & Information Technology, University of Nottingham (Dr Brian Logan) and the School of Computer Engineering, Nanyang Technological University, Singapore (Dr Stephen Turner). The Project is one of the four "Sister Projects" funded by the e-Science Programme in 2003

As part of the project, a transnational, long distance Grid infrastructure between the UK and Singapore is set up (Figure 3) and a case study in Grid-aware large-scale distributed simulation is developed, namely an HLA oriented, agent based large scale distributed simulation (e.g. supply chain management).

5 Summary

This paper has provided a short overview of the UK e-Science Programme, and has introduced MeSC, the Midlands e-Science Center of Excellence, one of the newly established e-Science centers which, together with the National and Regional e-Science Centers, constitute the UK National Grid infrastructure. MeSC has Modelling and Analysis of Large Complex Systems as its area of Excellence, with Grid-aware Large Scale Distributed Simulation its distinctive expertise.

References

- [1] Ian Foster and Carl Kesselman *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, July 1998.
- [2] The UK e-Science Core Programme. <http://www.rcuk.ac.uk/escience/>
- [3] The Science Budget, UK Department of Trade and Industry, Office for Science and Technology, 2004. <http://www.ost.gov.uk/research/funding/budget03-06/dti-sciencebudgetbook.pdf>
- [4] The National e-Science Center. <http://www.nesc.ac.uk>
- [5] I. Foster, C. Kesselman, J. Nick, S. Tuecke. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. *Open Grid Service Infrastructure WG, Global Grid Forum*, June 22, 2002.
- [6] S. Tuecke, K. Czajkowski, I. Foster, J. Frey, S. Graham, C. Kesselman, T. Maguire, T. Sandholm, P. Vanderbilt, D. Snelling. *Open Grid Services Infrastructure (OGSI) Version 1.0*. Global Grid Forum Draft Recommendation, 6/27/2003.
- [7] The Midlands e-Science center of Excellence. <http://www.mesc.bham.ac.uk>