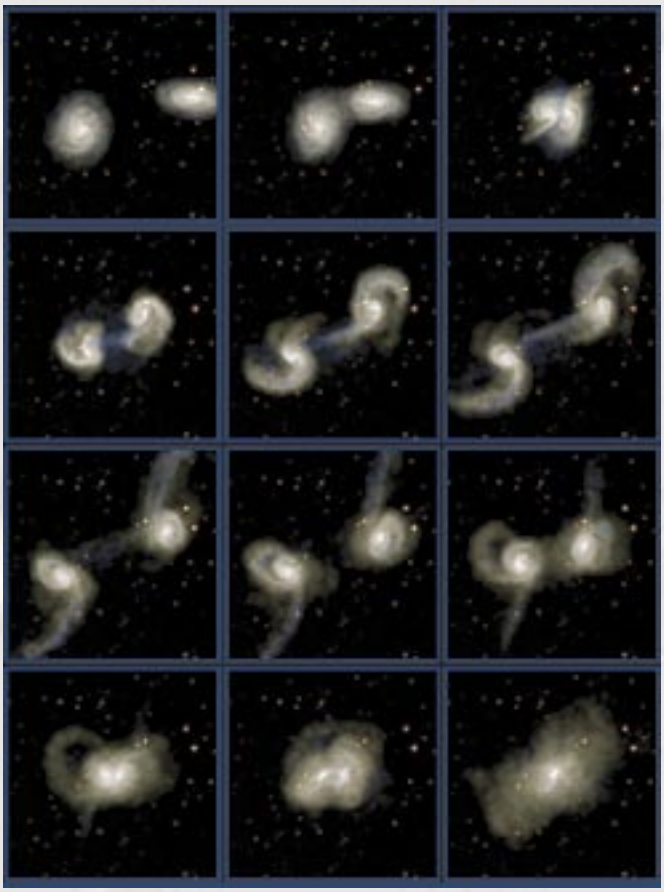
epcc news

The University of Edinburgh
www.epcc.ed.ac.uk

Issue 52, Autumn 2004

CONTENTS



Technological leadership: EPCC in 2004

MSc in High Performance Computing

EPCC at the University of Edinburgh offers a one-year taught MSc in High Performance Computing. This well-established programme provides an excellent grounding in HPC technologies and their practical application.

EPCC is an institute within the School of Physics at the University of Edinburgh. It has an international reputation as a leading centre of expertise in HPC and Grid computing. EPCC is the lead partner in the HPCx consortium, which provides the largest academic supercomputing resource in Europe for the UK research community.

If you have a keen interest in programming and would like to learn about high performance computing then this MSc will appeal. You will have access to leading-edge HPC platforms and technologies. The course has a strong practical focus, and lectures and tutorials are closely co-ordinated with associated practical sessions. You will also undertake a 16-week independent research project leading to a dissertation.

The MSc provides a doorway to a wide variety of future careers. Previous graduates have gone on to further research for a PhD in HPC and in areas of science which utilise HPC technologies. Other students have found employment directly.

Applications are encouraged from people who are competent programmers in Java, C++, C or Fortran and are graduates in science, engineering, computer science and mathematics or are currently working in a relevant field.

The entrance requirement is a good honours degree or equivalent work experience. No prior HPC knowledge is assumed.

For more information and application details see:
www.epcc.ed.ac.uk/msc/
Email: msc@epcc.ed.ac.uk

The MSc in HPC is supported by EPSRC, who offer a number of studentships.

Taught Courses

- Fundamental Concepts of HPC
- Practical Software Development
- Shared Memory Programming
- Message Passing Programming
- Parallel Decomposition
- Core Topics in HPC and e-Science
- Object Oriented Programming for HPC
- Hardware, Compilers and Performance Programming
- Applied Numerical Algorithms
- Exploiting the Computational Grid
- Scientific Visualisation
- Advanced Topics in HPC and e-Science

This MSc is supported by

EPSRC
Engineering and Physical Sciences Research Council

Editorial

Ratnadeep Abrol

Technical leadership is one of the founding principles of EPCC, and this issue's look back at 2004 demonstrates how this leadership permeates all aspects of EPCC's work.

In advanced computing, this issue brings you a taste of what it takes to create a home for high-end research computing by way of news of the Advanced Computing Facility (ACF) (p4). Housed within the ACF is the Storage Area Network (SAN) (p6), commissioned by EPCC, which contributes to making the University's e-Science infrastructure the most advanced in the UK. The ACF will also house a number of research machines, with QCDOC (see issue 51) being one of the most exciting.

Capability computing within EPCC has continued apace this year through collaborative projects such as RealityGrid and VirtU (p11). The Blue Gene/L project (see below) links capability and advanced computing by investigating the novel IBM Blue Gene/L platform. Also, EPCC has been supporting researchers within the UK and Europe through the continuing HPCx service (p7) and the new HPC-Europa visitor programme (p8).

This year has also seen a strengthening of the Grid effort within EPCC. The long running SunDCG project has culminated in JOSH (p17), global data and compute grid scheduling software contributed to the Grid Engine project at Sun; QCDGrid (p12), RealityGrid and VirtU have been looking at the applicability of the Grid to capability computing; and the FirstDIG (p15) and INWA (p16) projects have been proving there is a place for the Grid in industry.

With the Grid forming such a large part of our lives at EPCC, it is surprising to realise that it has only been four years since it was first discussed as a fledgling technology in the September 2000 issue of *EPCC News* (issue 40). In fact, the Grid is still a maturing technology. The RealityGrid article (p13) highlights this by considering some of the adversities faced by the pioneering Grid user. Even here EPCC is not resting on its laurels: the edikt project is successfully creating industrial strength Grid middleware to tackle such problems (p14); and by taking part in the NextGRID consortium EPCC is helping to shape the next generation of Grid middleware (p18).

2004 has also been a forward-looking year, with investigations into FPGAs (p7), investigations into Grid use in the financial services industry (p19) and NextGRID ensuring that EPCC will continue its technological leadership in coming years.

Blue Gene/L

Lorna Smith

EPCC is to collaborate with IBM in porting and optimising a range of scientific libraries onto the Blue Gene/L architecture [1].

This innovative new system is one of the next-generation of massively-parallel supercomputers. The first of these systems, to be installed in Lawrence Livermore National Laboratory [2], will consist of 65,536 nodes and provide nearly 400 teraflops performance.

Similar in design to the QCDOC architecture [3], each node comprises a single compute ASIC (an application-specific integrated circuit) and SDRAM-DDR memory chips. The ASIC is a complete system-on-a-chip, including all the required network interfaces and a small amount of fast on-chip memory. External memory chips are accessed through an on-chip memory controller.

ASIC provides a very compact, low-power building block. It allows Blue Gene to offer a high compute-density system with modest power and cooling consumption at a competitive price.

However, moving many key applications to this type of petascale system will be a challenge, one that will require new techniques to be developed which will scale to hundreds of thousands of processors. In this project we are focusing on porting and optimising key scientific libraries such as Scalapack and BLAS to this system, and look forward to solving many of these challenges.

This project is funded by IBM.

References

[1] www.research.ibm.com/bluegene/

[2] www.llnl.gov/asci/platforms/bluegenel/bluegene_home.html

[3] www.ph.ed.ac.uk/ukqcd/community/qcdoc/

The Advanced Computing Facility at the University of Edinburgh

Mike W. Brown

The Advanced Computing Facility (ACF) on the Edinburgh Technopole site at the Bush Estate is the new home for much of the University's high-end research computing equipment.

History

Completed in 1976 for the Edinburgh Regional Computing Centre (ERCC), which later evolved into the Edinburgh University Computing Service (EUCS), the ACF was built to house the large system procured for the regional service that was run by ERCC on behalf of the Universities of Edinburgh, Glasgow and Strathclyde. At that time this was the most powerful and significant computing installation in Scotland.

Contrary to expectation, computers began to get smaller, and the regional service was replaced in the early 1980's by a smaller but substantially more powerful one, located alongside the local services run by ERCC at King's Buildings.

The building was leased out to the Distiller's Company (later taken over by Guinness, and finally known as United Distillers), who used it to run their corporate IT until it was returned to the University. EPCC looked at this building for its HPC-97 bid, and the building went through a number of short-term leases until it became available again for University computing requirements towards the end of 2003.

Originally, the building consisted of an office area, a large (7,000 sq ft) computer room, and an extensive plant room. Distiller's completely refurbished the office area, and partitioned off a third of the computer room into an open-plan office. However, no upgrade was done to the support infrastructure, and by 1997 the multi-stage refrigeration plant was completely life-expired, and beyond repair.

A new lease of life

Early in 2003, with the existing accommodation at King's Buildings under constant pressure, a University working party selected the Bush to provide housing for the University's current and future large research computing equipment.

The building's remit was to provide an environment that was technically suited to house the envisaged types of computing equipment while also meeting the insurers' strict fire protection and physical security requirements.

Around the middle of 2003 an opportunity arose to apply for money to refurbish the building, and the go-ahead was given in November.

After the final tenant moved out in that month, the building was cleared of asbestos and the computer room areas were stripped. The main contractor took possession at the beginning of January, and works were substantially completed by the end of June. The quality of the building work, and the mechanical and electrical works, is excellent, and we

Continued opposite.

were greatly relieved that the works were completed to specification, substantially to time and also under budget.

The former single large computer room was taken back to its concrete shell, with all fixtures and services removed. Two independent secure, fireproof zones were established within, and one of them fitted out as a 3,000 sq ft (285 sq m) computer room, with appropriate floor, power distribution, cooling provision, fire-detection/suppression and security infrastructure. The other was left as a bare concrete box ready for future expansion.

The original plant room was stripped from end to end, and all services removed. In their place has come a modular chilled-water based system, with two 300kW capacity rotary-screw compressors supplying chilled-water to a high-capacity main. The in-room air-conditioning units are tapped directly off this chilled-water main, and valved connections were put in place within the plant room to enable two additional chillers if the need arises. All services have been designed with upgradeability and redundancy in mind, so that for the initial load we have duty and standby provision for all key elements of the cooling infrastructure. In addition there is a 2,500 litre chilled-water buffer vessel that will enable chilled-water to be provided for a reasonable period even after loss of both duty and standby chillers.

The QCDOC specialised system that we are constructing in collaboration with Columbia University and Brookhaven National Laboratory is an entirely water-cooled system. In order to support this system, cooling-water flow and return manifolds have been installed in the Computer Room, and attached to the chilled-water infrastructure. The QCDOC requires an independent secondary cooling-water circuit and, as with the chilled-water primary, there is duty and standby provision in the secondary circuit, and also its own 2,500 litre buffer tank.

In addition, it is expected that the cooling water return temperature will be above the outside ambient temperature for 70% of the time. In those circumstances, the cooling water will be re-directed up onto the roof and through a large dry-air cooler, so that the heat within the return water does not need to pass via the primary circuit to the chillers. We are expecting significant savings in plant running costs by using this 'free cooling' provision.

On the electrical side, a large main LV (low-voltage) switchboard was installed. This enables input to be switched between mains supplies and multiple standby generators, and the output to be distributed as raw power (for mechanical services) or to multiple UPSs (for computer services requiring conditioned power).

As the facility is not intended to house mainstream University services, the role of the UPS is not to sustain a 7x24 service in the event of total loss of site power, but to condition the input power such that 'brown-outs' do not cause disruption to services.

The provision of UPS-conditioned supplies is considered to be essential if services are to be maintained against the ever-present backdrop of fluctuating input voltages and phase-differences.

Upstream on the electrical provision, power is supplied to the building from its own high-voltage (11kV) substation, the 500kVA transformer being on the ScottishPower side, and supplying 415V to the main LV switchboard.

With only 500kVA available from the existing HV (high-

Continued overleaf.

Tera-scale Storage Area Network (SAN) for e-Science Research

Arthur Trew

Early in February 2004, EPCC completed the commissioning of a significant addition to the large-scale compute facilities available to major scientific research projects within and beyond the University: a 155 terabyte storage area network (SAN) funded by the Scottish Higher Education Funding Council (SHEFC), through its Science Research Investment Fund (SRIF) programme, and the edikt project.

The SAN provides high-performance storage in quantities currently unavailable to researchers in the most data-intensive scientific disciplines. It also enhances the University of Edinburgh's leadership in the operation of the most advanced e-Science infrastructure in the UK.

Located at the University's Advanced Computing Facility, the SAN will be fully operational by end of September 2004.

The system

The SAN has been designed to provide optimal multi-platform, multi-protocol and multi-site access to extremely large (~10s of terabytes) high-performance storage. Thus, research projects wishing to take advantage of the SAN are not restricted in the platforms they have, the location of their machines, or the protocols they use.

The SAN comprises both the raw storage and the multi-gigabit network fabric to enable multiple HPC systems to connect directly and simultaneously to portions of the available storage. The system also includes a tape-library, providing rapid online backup, archiving and recovery capacity.

In addition, the system's management software will allow EPCC's Systems Group staff to allocate storage dynamically, both to individual machines and to different research groups, as their storage requirements vary throughout their projects.

While the volume of storage available to researchers is in itself impressive, its performance will also significantly improve the rate at which scientific discoveries can be made and verified. Initial tests reveal minimum data transfer rates of 450MB/s between the SAN fabric and EPCC's existing SunFire 15K machine. Planned upgrades to the hardware and software of both the SAN and the 15K will provide almost double this throughput.

Supported research

Already, the availability of the SAN has prompted requests from a variety of the most data-intensive research groups within the University. These include Astronomy and Astrophysics, Particle Physics and edikt, who will be using the SAN to stress-test their novel data management and interpretation tools in a large-scale production environment.

Credits

The commissioning, operation and management of this major new facility for the University is the responsibility of Mike Brown, assisted by Linda Strefford and Craig Morris from EPCC's Special Systems Group.

---

voltage) supply, an alternative source was required for future provision, and we are shortly to commence work to install a new 11kV link to the new site HV ring which is being installed as part of the Technopole infrastructure upgrade.

ACF research services

The principal items of research computing equipment that have been, or soon will be, installed at the ACF are the University's 150TByte SAN (to provide large-scale storage to major research groups), the ScotGrid server and storage system, and the 10Tflop QCDOC system (for particle physics). The 10Gbit SRIF high-speed network backbone has been extended out to the ACF, and suitable equipment will be attached to this network for normal access, and for local (or remote) access to storage on the SAN.

As the QCDOC system will be, to some extent, almost a 'home build' type of machine, module test and repair facilities are required, and will be housed adjacent to the system itself.

The large, and somewhat esoteric, systems that will be housed at the ACF need a degree of close attention in excess of that required for more conventional mainstream services, and accordingly a 'Special Systems Group' is being set up to configure, manage and operate the equipment within the facility.

HPCx: the most powerful academic supercomputer in Europe

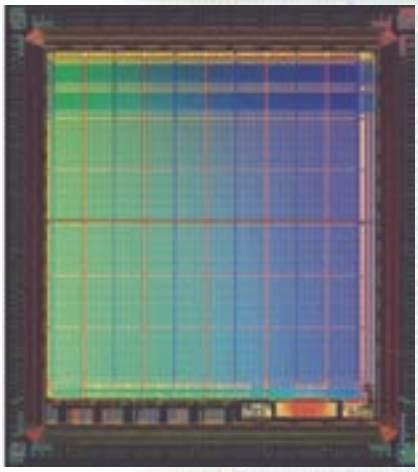
Alan Simpson

The University of Edinburgh, through EPCC, is the lead partner in HPCx, which is the UK's flagship supercomputing service. This £53m project runs from December 2002 for six years and its focus is firmly on capability science exploiting hundreds of processors. The support for the service is provided jointly by EPCC and CCLRC and the technology is provided by IBM.

2004 has been a particularly busy year for HPCx. The system underwent a significant upgrade which involved replacing both the interconnect and the processors as well as major upgrades of the system software. This was probably the most technically demanding part of the whole project and we were very happy that it was completed on schedule and with a minimum of disruption to users. The overall performance of the system has been doubled to more than 6 Tflops sustained. HPCx is ranked 18th in the recently released TOP500 list of the world's fastest supercomputers which makes it once again the most powerful academic supercomputer in Europe.

The HPCx system has been very popular throughout the last two years with utilisation averaging around 80%. The service now supports more than 30 national research consortia from many different applications areas and more than 350 users. These consortia have already produced many scientific results which are highlighted in the upcoming issue of the HPCx newsletter, *Capability Computing* (this will soon be available online at www.hpcx.ac.uk/about/newsletter). The upgraded system has significantly increased capability and the new interconnect has high bandwidth and much-improved latency. We are confident that this should allow an even broader range of applications to effectively use 512 processors or more and that this should ensure that the remaining four years of the project see the system deliver even more high quality science. This should be a very exciting and productive period for UK computational research.

More information about HPCx can be found on the website: www.hpcx.ac.uk



Reconfigurable computing

Andrew Murdoch

A new technology is emerging in high performance computing (HPC), involving chips that can change their configuration to match their task.

HPC applications have benefited from the exponential rise in the power of general purpose processors, which are cheap, readily available and use the latest semi-conductor processes. However, modern processors have significant cooling demands which limit the density of clusters and increase the running costs. Custom logic chips can be designed to provide very high performance for fixed tasks with low power requirements, but manufacturing is expensive, particularly for small batches.

One possible solution is to use Field Programmable Gate Array (FPGA) chips, whose digital circuits can be reconfigured to best match the demands of a given problem. FPGA devices are often used to verify custom chip designs and in applications that require flexible function. These include digital TV receivers that can decode and display new formats or network routing equipment with flexible rules to enable fast and secure communications.

FPGA devices have a regular structure composed of a large number of interconnected logic cells whose behaviour may be configured. An FPGA device configuration is produced by feeding a circuit design into a compiler which places and connects logic components. The circuits are designed using similar tools to those used to design custom chips, although higher level C-like programming languages are being developed. A programmable device may not perform as well for a given task as a fixed function custom chip, but the expectation is that performance and efficiency should be better than a general purpose processor running software.

EPCC is currently planning an experimental HPC platform using a number of reconfigurable devices in parallel to address problems in scientific and technical computing. This should allow a practical evaluation of the technology and provide valuable information and experience.

HPC-Europa

Promoting collaborative computational science in Europe

Catherine Inglis

HPC-Europa overview

HPC-Europa is a consortium of 11 leading HPC centres [1], funded by the European Commission (EC) for 4 years from January 2004, with a budget of 13 million.

The consortium will exploit the resources and expertise of each of the partners in order to create an innovative continental-scale facility for advanced computing, thus providing an integrated range of services to the European research community.

The cornerstone of HPC-Europa is its multi-disciplinary research visitor programme, which is underpinned by a series of inter-related research and networking projects. EPCC co-ordinates the Transnational Access (research visitor) programme, and is also active in all the other projects.

Visitor programme

Through the Transnational Access visitor programme, members of European research groups can visit any research institute associated with one of the participating centres [2]. While collaborating with researchers working in a similar field to their own, visitors also benefit from access to HPC-Europa's world-class computing facilities.

This visitor programme follows on from EPCC's TRACS programme, and similar programmes at CEPBA and CINECA. Over a 10-year period, TRACS brought over 500 researchers to some 30 departments in seven Scottish universities. Over four years, some 800 researchers will undertake HPC-Europa visits to research institutes associated with the six participating centres [2].

In 2004, EPCC has already welcomed HPC-Europa visitors from Austria, Bulgaria, the Czech Republic, France, Poland, Portugal, Spain and Switzerland. These visitors have come from the fields of informatics, engineering, chemistry, astrophysics and atomic physics, and have worked on projects such as 'Grid computing in 3D reconstruction', 'Global optimization of methane and water clusters' and 'A comparative study of the redox properties of thioredoxin-like enzymes'.

The first Transnational Access Meeting (TAM) took place in Edinburgh on 17th September. Researchers who have visited any of the centres will come to present their work and the results of their visit. Similar meetings in the past have been extremely successful, with new academic links often forged between visitors who had not previously met.

The bigger picture

Working in a consortium brings many benefits, the most obvious of which is an economy of scale. In HPC-Europa, resources and expertise are pooled to offer a higher quality of service to users than any individual centre could provide alone.

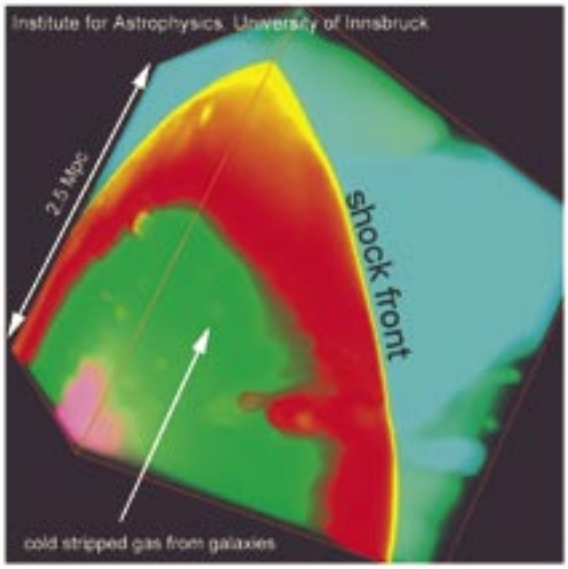
A Single Point of Access Grid portal is under development, to allow users to access HPC-Europa services in the same seamless manner, regardless of physical location. A centralised environment for technical support will be developed, including a shared repository of documentation and technical discussion forums. This will offer a greater pool of expert help to our users.

AccessGrid technology will allow tutorials, workshops and seminars at any given centre to be opened up to the visitors at all of the other centres. In this way, we will bring together visitors and local researchers at each of the centres, to create 'virtual communities' of researchers.

The ultimate goals of HPC-Europa are not only to develop a better, more integrated, service for all our users, but for the partner centres to learn to work effectively with each other

Continued opposite.

HPC-Europa



Temperature distribution in a simulation of the Coma cluster. Outgoing shock fronts caused by subcluster mergers are shown. The cold stripped gas from galaxies can be seen.

(This image was produced as part of the HPC-Europa research project described overleaf.)

and to co-ordinate their action for the long-term benefit of both the users and the participating centres themselves.

Networking and Joint Research Projects

The following networking and joint research projects underpin the visitor programme:

Collaboration Support Infrastructure and Tools in AccessGrid

AccessGrid is fundamental to HPC-Europa, as it allows remote collaboration between visitors and staff at different sites. This project will improve the scalability and functionality of essential tools used in AccessGrid collaborations, such as performance analysis tools and 3D visualisation software. The aim is to allow groups of researchers across Europe to be able to work together as efficiently as if they were in the same workplace.

Data management and portability

Building on existing expertise, this project will contribute to emerging tools and standards for efficient and portable data management and information retrieval. The project will address two main types of problems: those of interoperability between the data and the software that processes it, and those created by the working practices of the users themselves. The resulting tools and recommended working practices will increase the potential for reuse, sharing and interoperability of data and applications.

Performance analysis tools

To allow users to make more effective use of our HPC infrastructures, we must offer powerful and flexible tools for analysing the performance and behaviour of applications. Uniform tools for advanced performance analysis, based on CEPBA's Paraver technology, will be delivered across all sites. These tools will have very powerful analytical capabilities, will be scalable, and will have a uniform interface for analysing different programming models.

Joint analysis methodologies and tools to monitor systems performance will also be developed. These will be of value not only to users of the infrastructure, but also to the operators, who will be able to assess how efficiently their systems are being used, and make informed decisions on how to tune their configuration.

Single point of access

This project will develop a uniform, flexible and intuitive environment to access the consortium's Grid resources in a seamless manner, regardless of the user's location. The final infrastructure will include tools such as centralised accounting of distributed resources and user administration.

Further information

Further information about the consortium's activities can be found at www.hpc-europa.org

Closing dates for the visitor programme are held every three months, although researchers may apply at any time, using the application form available at www.hpc-europa.org/ta.html

If you have any questions about any aspect of the programme, please contact the HPC-Europa team at EPCC at europa@epcc.ed.ac.uk

Footnotes

1 The HPC-Europa consortium members are: CASPUR (Rome), CEPBA (Barcelona), CINECA (Bologna), EPCC (Edinburgh), HLRS (Stuttgart), IDRIS (Paris), NTUA (Athens), Parallab (Bergen), PSNC (Poznan), SARA (Amsterdam), TCD (Dublin)

2 The HPC-Europa partners participating in the Transnational Access programme are: CEPBA, CINECA, EPCC, HLRS, IDRIS and SARA.

Continued overleaf.

Here we feature the research project of one recent HPC-Europa visitor to EPCC.

Enrichment of the intra-cluster medium

Dr Eelco van Kampen is a lecturer at the Institute for Astrophysics at the University of Innsbruck (Austria). His HPC-Europa visit to EPCC was hosted by Dr Max Ruffert of the School of Mathematics at the University of Edinburgh.

Project description

My main focus is on galaxy formation and cosmology, and especially on galaxy clusters. The project aims to better model the interaction between the hot gas component of a cluster and its member galaxies. From X-ray spectra it is evident that metals are abundant in this hot gas. As heavy elements are only produced in stars the processed material must have been ejected by cluster galaxies. Several different mechanisms are possible but their efficiency and time evolution are unknown. Suggested processes include ram-pressure stripping, galactic winds, superwinds driven by starburst activity and jets from active galaxies.

In order to simulate the interaction between the ICM and the cluster galaxies, we devised a combined N-body, hydrodynamic, and phenomenological approach. Large-scale structure formation is simulated using an N-body tree code with an additional semi-numerical model for galaxy formation (van Kampen et al. 1999). The large-scale dark matter potential and the trajectories and evolving properties of the galaxies are stored, to be used as input for a hydrodynamical simulation of the ICM. These simulations are performed on cluster scale (20 Mpc), and employ a shock capturing Grid-based Piecewise Parabolic Method (PPM for short) on four nested grids. This allows us to cover the cluster centre, where most of the stripping is expected to happen, with high enough resolution.

Because the large-scale environment around the cluster needs to be modelled as well as the cluster itself, and we need to simulate a whole range of physical processes (from cosmology to star formation), the simulation set-up can only be performed on a high-performance computer. Many parts of the problem have existing solutions, and corresponding codes, but the challenge is to combine these parts to solve the whole problem. The various codes not only use different techniques, but are written in different languages, and have different data structures and units. My HPC-Europa visit was mainly aimed at finding the best solution for this using the knowledge available at the EPCC, and they managed to point me in all the right directions.

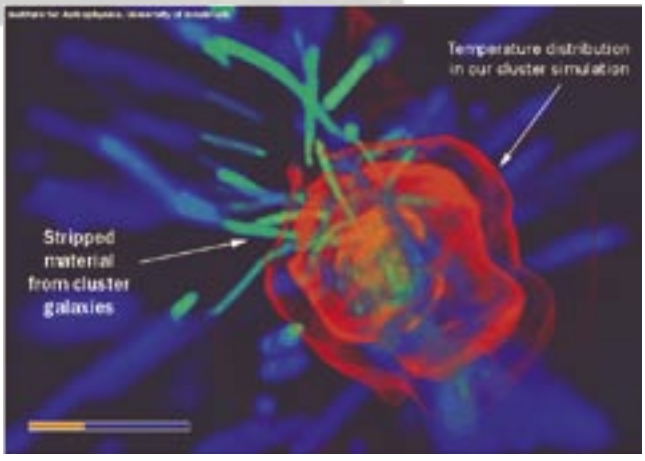
Specifically, the use of IDL was discouraged, even though a parallel library (MPIDL) exists which can in principle be used. The various codes which are solving a specific part of the problem (for example, the cosmological evolution of the dark matter component) are best left alone, and made to talk to each other through the MPI interface. This should also aid load balancing. However, some of this is still under discussion, and needs further investigation. Over the next few months various solutions will be tested on the machines available to EPCC, notably HPCx, to see which one works in practice.

References:

van Kampen E., Jimenez R., Peacock J.A., 1999, MNRAS, 310, 43

For more information on the project, please refer to: http://astro.uibk.ac.at/astroneu/hydroskiteam/index.htm

Metallicity distribution (blue-green) and temperature distrbution (red colourmap) in a (10 Mpc)^3. The stripped material from galaxies within the cluster and the Interaction with shockfronts going outwards due to a merger process can be seen. The simulation starts at half the age of the Universe and extends to roughly 10 Gyrs into the future.



VirtU: Developing the Virtual Observatory

Gavin J. Pringle

EPCC has been involved with computational cosmology for many years now, through our work with the Virgo Consortium, whose Principal Investigator (PI) is Prof. Carlos Frenk [1]. In issue 50 of *EPCC News* we reported on our involvement with the DEISA project (Distributed European Infrastructure for Supercomputing Applications, [2]). Part of this project involves Grid-enabling two Virgo simulation codes, namely GADGET [3] and Hydra_MPI [4], and a set of pre- and post-processing tools. Part of the funding of this work comes from the VirtU programme [5].

The VirtU programme, funded by PPARC, is an international project to construct a virtual universe (VirtU). The PI is Prof. Frenk and the author is one of the ten co-investigators.

As stated in [5], this virtual universe will consist of the Theoretical Virtual Observatory (TVO) and associated diagnostic tools making up the Theory/Observations Interface (TOI). VirtU will be a computing infrastructure to enable direct and rigorous comparisons of realistic simulations of cosmic structures, based on the best current theoretical understanding, with real data. VirtU will provide a set of tools to:

- remotely access and manipulate a large, dynamic archive of simulated objects (galaxies, clusters, etc) using AstroGrid [6] standards and procedures
- compare simulations carried out using different techniques
- compare simulated data held in the TVO with real data held in the Virtual Observatory
- enable parallel simulations to be executed remotely using Grid technology
- visualise multi-dimensional simulation output (including particle-based data) and observational data.

To achieve these goals, we propose to:

- develop a standard data format for storing simulation output
- develop and Grid-enable a set of powerful tools for the analysis of simulated and real datasets
- create a Virtual Telescope Simulator to aid in the planning of astronomical observations and the development of astronomical instruments
- adapt and expand existing visualisation tools to the specific requirements of VirtU.

The VirtU initiative is a collaboration that involves most of the groups active in cosmological and gas-dynamic simulation work in the UK, including EPCC, as well as collaborators in Germany, the USA and Canada. It will form a key component of the International Virtual Observatory (IVO) [7].

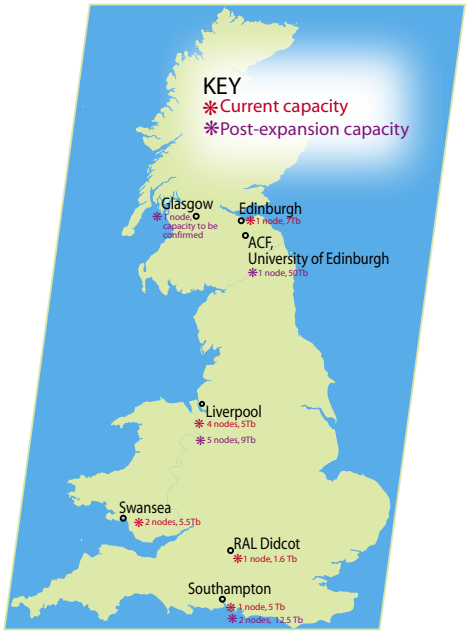
The team at EPCC is excited to be part of such a worthwhile, challenging project and look forward to continuing our close relationship with Virgo.

References

[1] www.virgo.dur.ac.uk
[2] www.deisa.org
[3] www.epcc.ed.ac.uk/t3e/virgo
[4] www.mpa-garching.mpg.de/~volker/gadget/index.html
[5] http://star-www.dur.ac.uk/~csf/virtU/virtU-final.pdf
[6] www.astrogrid.org
[7] www.ivoa.net

QCDgrid: Probing the building blocks of matter with the power of the Grid

George Beckett



Quantum chromodynamics (QCD) is the study of the basic constituents of matter, the building blocks for our universe. Both in the UK and around the world, scientists are currently developing techniques to quantify the complex behaviour of fundamental particles called 'quarks' and 'gluons'. Together, these particles completely define the properties of hadrons (for example, protons and neutrons) and hence the constitution of all nuclear matter.

Through complex and computationally intensive simulations, based on highly non-linear models, QCD physicists are striving to fill current gaps in our understanding of these particles. The sheer size and scale of these simulations dictates many months of computational time on the most powerful computers available today. In addition, these simulations generate terabytes of data, which then have to be analysed by a comparably intensive computation, in order to extract the physical quantities that scientists seek.

UKQCD is a collaboration of leading particle physicists from centres around the UK who have commissioned a HPC system, called QCDOC (QCD On a Chip), to accelerate the scale and rate at which these simulations can be performed.

EPCC, in collaboration with UKQCD, has developed and is continuing to enhance QCDgrid, a data management software suite that combines the distributed storage resources of the collaborators into a robust, unified facility called a 'data Grid'. QCDgrid provides a simple and intuitive environment that hides the many complexities of the underlying Grid technologies and presents what is essentially a standard file system to the user. QCDgrid incorporates a robustness metric that automatically disperses the simulation datasets across the Grid, providing a resilience that ensures data is not affected by the loss of one (or possibly more) storage nodes.

In addition to direct access to the file system, QCDgrid allows the user to query an associated metadata catalogue using a GUI browser tool that can identify and automatically retrieve 'files' from the data grid, based on a query definition.

The software also provides a job submission system that allows a user to schedule simulations and post-processing tasks on remote HPC systems from the comfort of their PC. As an example, QCDgrid allows a user to submit a post-processing job to a remote compute resource and delivers the associated simulation data directly to the resource from the data Grid, obviating any need for the user to transfer data manually via their local system.

In the future, we aim to join the UKQCD Grid with other similar activities in the International Lattice Data Grid (ILDG), a collaboration of research groups in the United States, Japan, Europe and Australia. This activity will allow like-minded scientists around the world to share their data and benefit from the scientific progress that other groups have made. The multi-national data Grid will be built on web service technologies and EPCC will be responsible for the definition of a web service interface specification that will allow national resources (such as the UKQCD Grid) to be integrated seamlessly into the ILDG system.

Some facts and figures

The UKQCD Grid has been operational since autumn 2002. At the time of writing, it consists of seven storage elements hosted at four sites: Edinburgh, Liverpool, Swansea and Rutherford Appleton Laboratories. An expansion phase is planned for the near future that will add new nodes at Southampton, Glasgow and the University of Edinburgh Advanced Computing Facility.

The storage capacity of the data grid is 12 terabytes (which will grow to 100 terabytes as a result of the expansion). This hosts approximately 30,000 simulation datasets occupying one third of the current capacity.

Further information about the project can be found by visiting our website: www.epcc.ed.ac.uk/qcdgrid

Funding

QCDgrid is part of the GridPP/GridPP2 project, a PPARC-funded collaboration between particle physicists and computational scientists from the UK and CERN, who are building a grid for particle physics.

Grid reality? The progress of RealityGrid

Kevin Stratford

At a recent conference devoted to mesoscopic methods in science and engineering in the sleepy Saxon setting of Braunschweig in Germany [1], there assembled a wide range of fluid dynamics and condensed matter physics experts from Asia, Europe and the Americas to discuss their latest work. A large number of presentations at the meeting were given over to the lattice Boltzmann equation, a crafty method which has been developed in the last ten years or so to solve the Navier-Stokes equations for fluid flow, and one which is particularly popular in the high performance computing community because of its amenability to parallelisation. The versatility of the lattice Boltzmann method was evident in the scope of the applications appearing in the talks: these ranged from the study of car and motorcycle aerodynamics by local automotive giant BMW, to the design of ion thrust engines in NASA's latest space probes. More mundanely, other presentations considered the search for better low-fat milk, food mixers or treatments for heart disease.

What better place to gauge the impact of Grid computing in the everyday work of computational scientists? Surely no computer-shy Luddites here! So, did anyone appear to be making use of the Grid? Well, no.

If recent experience in the RealityGrid [2] project here in the UK is anything to go by, this observation is hardly surprising. This project is one of a small number of e-Science pilot projects funded by EPSRC, has been running for over two years and involves collaboration between universities in London, Loughborough, Manchester and Edinburgh (along with EPCC), together with a number of industrial partners. Many members of the project have a strong scientific interest in running large-scale computations such as those based on lattice Boltzmann methods and molecular dynamics, which causes a hunger for the enhanced resources the Grid promises. However, going has been tough.

At a recent internal project meeting, a couple of talks detailed the sometimes excruciating pain involved in running Grid applications for real scientific problems with current resources. Jonathan Chin of University College London reported [3] on a number of serious obstacles that project members have encountered in getting Grid applications off the ground. The list was long: middleware of ferocious complexity which is difficult both to install and to run, the need for intervention by individual system administrators, remote job submission well-nigh impossible, and so on. Masochistic pride in the pain of the frontier explorer aside, there was a serious message: who can blame anyone for not being interested in using the Grid while it's so difficult? Perhaps the most fundamental allegation against the current machinery was that it was 'conceptually heavy' – it just takes far too long to comprehend everything that's involved. Surely the whole business could be less burdensome to the user?

Well, it's easy to criticise. RealityGrid has tried to concentrate its efforts on lightweight middleware for the specific purpose of computational steering – interaction with a running computation remote from the user to guide or merely monitor its progress.

This has been achieved using a combination of standard linkable libraries which the user adds to their application code to enable steering via a Grid service, client interfaces with which the user interacts, and an OGSI::Lite [4] container which hosts the services and handles SOAP communication. Clients have been developed as stand-alone applications for desktops, for web-browsers via the Gridsphere [5] portal framework, and even for PDAs using .NET. For users within the project, this lightweight approach has helped alleviate the pain of the Grid.

In the well-circumscribed aim of steering then, the project has been successful in providing a usable and useful tool for Grid computing, which culminated in the award of an HPC Challenge award at SuperComputing '03 in Arizona. More heavyweight pieces of software that are being developed within the project for the purpose of model coupling (to allow the interaction of different codes to compute different part of a unified problem) will clearly take more time. However, the aim is the same: to produce usable middleware with real problems in mind.

So, in years to come, it will no doubt be commonplace to reminisce at meetings like those in Braunschweig about the good times when it was painful to use Grid computing. Then we will know that the Grid has truly succeeded in helping scientists do their job.

References

[1] www.icmmes.org/

[2] www.realitygrid.org/

[3] www.nesc.ac.uk/technical_papers/UKeS-2004-01.pdf

[4] www.sve.man.ac.uk/Research/AtoZ/ILCT/

[5] www.gridsphere.org/

e-Science Data, Information and Knowledge Transformation: edikt in 2004

Rob Baxter

The edikt project at EPCC's sister organisation – the National e-Science Centre – entered its third year in 2004, continuing its quest to transfer leading-edge data management techniques to e-Science.

Edikt's mission is to apply software engineering 'hardening' to computer science prototypes and to build new tools and systems to support the emergent possibilities in scientific data analysis brought about by the Grid. Improving instrumentation and techniques in fields like astronomy, life science and geo-science is resulting in a tsunami of data. Edikt's challenge is to tame this tsunami and to distill new scientific knowledge from it.

2004 saw edikt release two base data manipulation technologies to the e-Science community: Eldas, an implementation of the Global Grid Forum's Data Access and Integration Service (DAIS) specifications based on Enterprise Java Beans technology; and BinX, an XML Schema language and accompanying software library for describing and transforming binary files. Edikt plans to extend and enhance these generic technologies through applications in a range of e-Science disciplines.

During the last year, for instance, Eldas has been put to work in the fields of astronomy, bioinformatics and geographical information systems. EdSkyQuery-G is a collaboration between edikt, the OGSA-DAI team at EPCC and AstroGrid. Its aim is to build a system for astronomical data federation based on the OpenSkyQuery standards emerging from Virtual Observatory initiatives in the US, UK and beyond. EdSkyQuery-G will fulfil this aim by building 'virtual query services' on top of Eldas and OGSA-DAI – both based on the DAIS specifications but with very different technology platforms underneath – with the eventual goal of integrating fully with the UK's AstroGrid Virtual Observatory infrastructure.

The same techniques of data integration using Eldas will also be applied to gene sequence databases in edikt's BioDAS project, and to virtual geographical data systems in GridOGC. This latter work, in collaboration with the UK's national data centre EDINA, plans to develop services to draw together data based on the Open GIS Consortium's interface standards for geographical data.

2004 will also see the successful conclusion of the OSAGE project, edikt's collaboration with the MRC Human Genetics Unit to re-architect and re-engineer the Edinburgh Mouse Atlas Project gene database.

More projects are in the pipeline and edikt is looking ahead to 2005 with keen anticipation!

More information on edikt, including current software downloads, is available from www.edikt.org

---

EPCC at ISC2004

Gavin J. Pringle

The 19th annual European supercomputing summit, the International Supercomputer Conference (ISC2004), was held in Heidelberg in June, and EPCC was there as attendees and exhibitors. Many delegates were particularly interested in our 'History of HPC at EPCC' display: several got down on the floor for a closer look at a photo of the first EPCC HPC platform: the ICL Distributed Array Processor.

In general, ISC2004 was very successful, with a record number of participants. The conference featured a wide range of academic, industrial and governmental speakers. However, it was disappointing that the majority of speakers were based in only two countries: USA and Germany.

The new Top500 list was announced and discussed at length. Jack Dongarra talked about a new list, where platforms are ranked by processors, but also memory system and interconnect. My personal highlight was the vendors' Hot Seat Sessions, where Microsoft announced it intends to become a plausible HPC vendor. AMD, the event's main sponsors, described its Opteron processor, and Cray described its involvement with DARPA and its impressive Red Storm platform at Sandia.

For an overview of the three-day conference by Ad Emmen, see: www.enterthegrid.com/primeur/live

ISC2005 looks like being bigger and better. For further information – and also photographs of ISC2004 – see: www.isc2005.org

FirstDIG: Using the Grid for data in business

Paul Graham

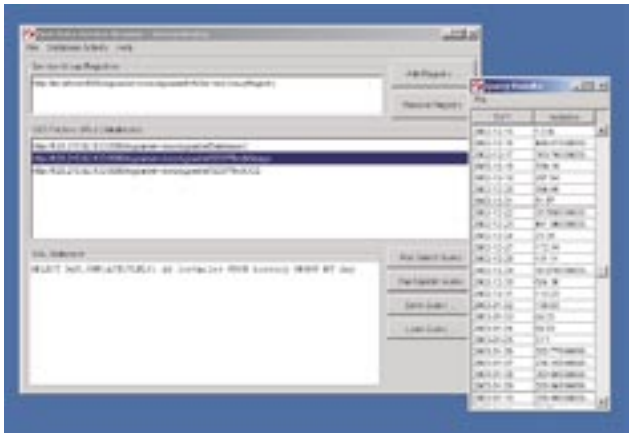
The FirstDIG project was a collaboration between the National e-Science Centre (represented by EPCC) and FirstGroup plc, represented by First South Yorkshire. First is a major company in the transport sector. It operates worldwide and is the UK's largest operator of bus services, with over 10000 vehicles in the UK alone. In running its services, First collects many different kinds of operational data, ranging from revenue and fuel consumption through to customer contact and scheduling. This data is collected in different ways, such as via manual entry, ticket machines or even GPS. The data is also stored in many different databases and systems: this is where First hoped to utilise Grid technologies.

The heterogeneous nature of First's data repositories can be traced to reasons such as earlier acquisitions of other companies and the incremental construction of systems. Now, this does not cause a problem for the day-to-day running and querying of the databases, but First was interested in using data mining to discover non-obvious and valuable information. One of the techniques of data mining involves correlating and comparing the data from various sources in order to find this non-obvious information. However, computing these correlations is clearly going to be difficult if the databases involved are on disparate operating systems, different database management systems, have different internal representations of the data, and even are geographically distributed between various sections of the company. This is an issue common to many organisations.

EPCC's solution was to provide a mechanism so that First could access their data via a single interface in a uniform manner, protecting the user from the differences between, and the location of the underlying databases. This was achieved by 'Grid-enabling' the databases via the OGSA-DAI (Open Grid Services Architecture – Data Access and Integration) software (www.ogsadai.org.uk). This Grid middleware exposes the databases as Grid Services, thus enabling access from other machines in a secure manner. Two of First's databases were identified as suitable candidates via a preliminary data mining exercise, and EPCC deployed the OGSA-DAI software for these databases. The software incorporated a GUI front-end to the Grid Services making access to the data easy and straightforward for the user. Functionality was included in the GUI to enable a 'JOIN' SQL query to be performed, allowing data from the different databases to be combined for data mining analysis.

In conclusion, EPCC successfully demonstrated the use of Grid middleware in a 'real-world' business environment, enabling First to access their data in a straightforward and repeatable manner. First has discovered valuable information from its data that would otherwise have been almost unobtainable due to the heterogeneous nature of the data resources and the impracticality of exporting individual databases. Now First's data is available 'in-situ' via a simple GUI, and it is keen to extend the technology to other databases. Also, the GUI has been taken up by the OGSA-DAI team and is being used in the INWA project (see next page).

The success of this project can be summarised by a quote from Darren Unwin, Divisional Manager, First South Yorkshire: 'The results of this exercise will revolutionise the way we do things in the bus industry.'



INWA: informing business and regional policy

Alastair Hume

Grid-enabled fusion of global data and local knowledge

Data mining projects often require distributed analysts to submit jobs to distributed compute resources that process data from distributed data resources. These requirements, along with others such as secure communications and access control, make data mining an ideal application of Grid technologies.

The INWA project [1] has investigated the suitability of existing Grid technologies for secure commercial data mining. This project has been funded under the Pilot Projects in E-Social Science programme [2] of the Economic and Social Research Council (ESRC). The project partners in the UK are EPCC, The University of Edinburgh Management School and Lancaster University Management School. In Australia the project partners are Curtin Business School in the Curtin University of Technology in Perth and Sun Microsystems.

This project has successfully built a Grid infrastructure linking compute and data resources in the UK and Australia, enabling distributed public and private data sources to be fused and analysed using data mining techniques. This infrastructure uses various Grid technologies from the UK e-Science Core Programme: The Transfer-queue Over Globus (TOG) [3] software from the EPCC Sun Data and Compute Grids project; The First Data Service Browser from FirstDIG [4] (see previous page); and OGSA-DAI [5].

The project has proved successful in identifying the strengths and weaknesses of the Grid technologies for commercial data mining. Where possible, valuable feedback has been given to the development teams responsible for the adopted Grid technologies, and this had led to bug fixes and enhancements in later versions of these products.

From a data mining point of view the project has successfully integrated public and private data using Grid technologies. This and other data has been successfully mined by remote analysts using the secure Grid infrastructure and has produced results that are of real interest to the companies that supplied the data.

References

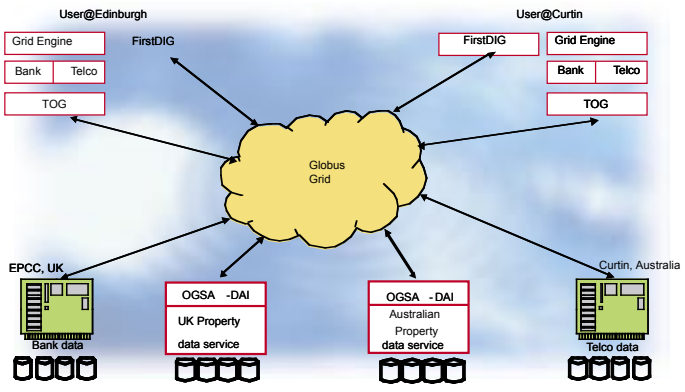
[1] The INWA project, www.epcc.ed.ac.uk/projects/inwa

[2] ESRC Pilot Projects in E-Social Science, www.esrc.ac.uk/esrccontent/researchfunding/e-Science.asp

[3] Transfer-queue Over Globus (TOG), http://gridengine.sunsource.net/project/gridengine/tog.html

[4] FirstDIG, www.epcc.ed.ac.uk/~firstdig/

[5] OGSA-DAI, www.ogsadai.org/



SunDCG: JOSH leaves home

Geoff Cawood

EPCC's Sun Data and Compute Grids project (SunDCG) was successfully completed in Spring 2004. The project team was tasked with developing an industry-strength compute and data scheduler based around Grid Engine, Globus and a variety of data technologies.

Grid Engine [1] is an open source distributed resource management system or 'local scheduler' that allows efficient use of compute resources within an organisation. However, Grid Engine alone does not provide the means to share resources among collaborating organisations. SunDCG aimed to achieve this wider goal by integrating Grid Engine with the Globus Toolkit's [2] Grid API to form a 'Grid scheduler'.

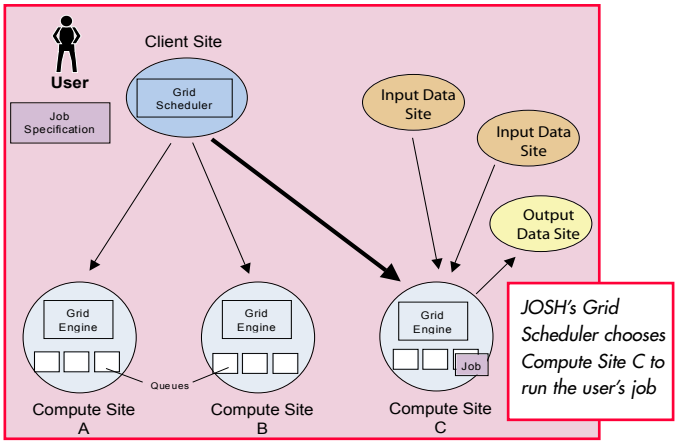
As mentioned in *EPCC News 49*, the project's first software deliverable was TOG (Transfer-queue Over Globus) [3]. This is an effective and usable integration, but it suffers from scalability issues. To improve on this, the team based its final Grid scheduler on the concept of hierarchy, and so JOSH (JOb Scheduling Hierarchically) was born.

Any Grid scheduler faces a difficult challenge: assign an incoming job to one of potentially many thousands of compute queues so as to satisfy all the requirements of the job (eg memory, disc space, operating system) while balancing machine utilisation, data transfer time, cost and a range of other metrics.

The idea with JOSH is divide-and-conquer: break one hard decision into a number of easier decisions. Specifically, a Grid scheduler in a JOSH network makes a coarse-grained decision about which compute site should execute the job, and then the local scheduler at that site decides on the actual queue.

This is shown in the diagram. A user at a client site defines their job (eg. a script for execution, parameters, input/output files, resource requirements) by writing a job specification file. The user submits this to a JOSH Grid scheduler, which queries each of the compute sites and ranks them according to:

- The physical ability of their queues to execute the job and satisfy its requirements
- The current load on those capable queues
- The estimated time to transfer the job's inputs and outputs from/to their respective sites.



In this case, the Grid scheduler selects compute site C as the best choice. The job specification is then forwarded to that site, where the Grid Engine decides which local queue at C should execute the job.

Just before the job starts, JOSH uses the data transfer facility of Globus to pull the specified input files to the compute site. Similarly, output files are pushed to their destinations at the end of the job. Supported protocols for this automatic staging are GridFTP, anonymous FTP and HTTP.

JOSH has been available from the open source Grid Engine community web site [4] since February 2004. News of the release has generated considerable interest and hundreds of downloads for the JOSH documentation. As a result, a number of organizations are considering using JOSH in their future Grid projects (eg. ULGrid [5], Poseidon [6], HPC-Europa JRA2 [7]).

Unfortunately, reported issues with the performance and robustness of Globus Toolkit V3.0 used by JOSH have deterred some potential users. However, modifying JOSH to use future releases of Globus should be relatively straightforward due to JOSH's modular design. Indeed, the open source nature of the code should allow interested parties to adapt JOSH to meet future challenges and ensure he has a long and useful career.

This work has been funded by the EPSRC/DTI Grid Core Programme. For further information on the project please see www.epcc.ed.ac.uk/sungrid.

References

[1] Grid Engine, http://gridengine.sunsource.net/

[2] The Globus Alliance, www.globus.org/

[3] TOG on Grid Engine site, http://gridengine.sunsource.net/project/gridengine/tog.html

[4] JOSH on Grid Engine site, http://gridengine.sunsource.net/project/gridengine/josh.html

[5] ULGrid, www.liv.ac.uk/e-science/ulgrid/

[6] Poseidon, http://czms.mit.edu/poseidon/new1/

[7] HPC-Europa, www.hpc-europa.org/ra2.html

FP6 Grid projects launch in Brussels

Diana Engesser

Following the success of the FP5 IST Grid Projects, the European Commission has recently announced the selection of a series of new Grid projects funded under the 2nd IST Call of FP6.

Further to the numerous achievements and continued success of the FP5 GRIDSTART project, EPCC has been invited by the Commission to organise a three-day event 'European Grid Technology Days 2004' taking place in Brussels from the 15th-17th of September.

This event has two goals. The first is to launch the FP6 Projects formally. The second is to pass on the achievements, knowledge and lessons learnt from the FP5 projects to the FP6 projects. This unique event is expected to attract around 200 participants.

The project launch meeting on 15th September is open to all and will feature keynote speakers from the European Commission and leading Grid experts from industry. There will also be an industry forum which will give participants an opportunity to hear views and contribute to the discussion on the direction that R&D in Grid computing should take in order for it to be a useful business tool. Representatives of each FP6 project will give an overview of their projects. One of these is NextGRID, led by EPCC.

NextGRID (see below) focuses on developing the architecture for the next generation of the Grid and aims to prepare the way for a mainstream use of Grid technologies. The project will run for three years and involves over 20 partners across Europe.

The second two days of the meeting (16th and 17th September) are open to FP5 and FP6 participants and will consist of plenary sessions and focused technical sessions to define common research directions among the projects, and ways in which they can collaborate.

Full details can be found at the event website: www.nextgrid.org/events

NextGRID: An architecture for the next generation Grid

Mark Sawyer

EPCC is the coordinator for NextGRID; a project to produce the architecture for the next generation Grid. The three-year project started on 1st September 2004, and involves twenty-two partners representing the elite of Europe's Grid research community, together with leading lights from industry.

The Grid has the potential to make a significant advance beyond the Internet, by turning it from a passive information medium into an active tool for creating and exploring new knowledge. This potential is currently unrealised, and without cost-effective and universally applicable technology, will remain so. Although Grid computing is already delivering benefits to scientists, this is only with a high cost in skilled human resources for the deployment and operation of Grids and the software installations they depend on. NextGRID aims to improve the technology in key areas such as security, development of applications, interoperability, quality of service and robustness.

NextGRID will result in programming environments, resource management tools, economic and business models for new services, and customisable, generic middleware and standards. These will be released in a six-monthly cycle and the reference implementations will be available as open-source.

For NextGRID to succeed, it must be involved in the global standardisation process for Grid computing. The partners in NextGRID will work with international standards bodies and community efforts such as the Global Grid Forum (GGF), to ensure that widespread take-up of the results will be possible.

Website: www.nextgrid.org

Financial services and the Grid

Mark Parsons

One of the main business domains where Grid solutions are currently being sold is the financial services sector. A number of banks in the US have already seen how they can realise much better utilisation of hardware resources, improve data and information flow within their business, respond rapidly to changing business conditions and improve productivity, competitiveness and profitability.

The financial services sector has a long history of deploying innovative IT solutions to try and steal a march on their competitors. The opportunities promised by the Grid are obviously attractive to them and according to one of the major players, over 50% of all US mortgages are now processed using Grid technology. Because a number of the applications used in the sector lend themselves easily to Grid computing some impressive claims are being made. For instance, one financial services customer, following a four month deployment project, saw a 14% reduction in their annual IT budget, which amounted to savings in excess of $30 million thanks to a 30% reduction in the cost of future hardware purchases.

At EPCC, we decided three years ago that the compute-on-demand space was very crowded. We have therefore focussed on Data Access and Integration Services for the Grid and have several world-leading technologies in this area. For instance our OGSA-DAI technology allows users to access SQL and XML databases from anywhere in the world in a standardised fashion that is independent of the underlying database management systems. This enables us to build virtual data warehouses to solve previously intractable business and scientific problems. We believe there are huge opportunities for the financial services sector in this area.

To bring a number of the issues described in this article to the attention of the major Scottish financial services organisations, EPCC hosted a seminar in Edinburgh in May this year. With guest speakers including Steven Neiman, Head of Innovation for JP Morgan Chase in New York – which has already deployed Grid solutions to great benefit – and Steve Ross-Talbot from London-based Enigmatec, the senior attendees from across the sector learnt what the Grid could mean to their businesses.

EPCC and a number of Scottish software companies who were represented at the meeting are now taking forward a series of proposals for pilot projects with some of Scotland's largest financial institutions.

We are just at the beginning of a major transformation in the way we build and deploy applications within businesses. Over the next decade, Grid and Web Services technologies are going to drive this transformation. Scottish financial services organisations, and businesses from many other sectors, can be in the vanguard of these developments if they start deploying them now. In the US, companies are already beginning to understand the true value of these developments. The core message of our seminar was that Scottish businesses should make sure they aren't left behind while their competitors reap the rewards.

The Grid: Software Efficiency and Re-usability

The uptake of computer simulation by emerging computational science disciplines and the constant quest for larger and more efficient computer resources by traditional disciplines lead to frequent evolution and revolution in software technologies that aim to support these developments. The Software Efficiency and Re-useability Study as carried out by the ENACTS project focuses on recent developments and emerging software technologies as these affect the efficiency and reusability of new and existing applications. In addition, the Study looks at portability and interoperability of software and data between facilities and how these impact program efficiency.
The Study is performed by ENACTS partner Parallab, Bergen Center for Computational Science at the University of Bergen (Norway) with input from the other partners in the ENACTS consortium.

Increasing software reuse leads to higher human productivity and increased software quality. However, it is not always pursued in practice. The report produced by the Study reviews some general considerations related to software reuse, including its place in the software development life cycle but also the influence of non-technical aspects (like economics and management). The importance of standards and standardisation in modern programming and scripting languages, protocols, mark-up languages, and meta languages is discussed as well as the impact of Open Source.

The efficiency of sequential, parallel, and distributed applications depends on the infrastructure on which it is executed. Application performance is influenced by the details of the underlying hardware topology, network topology, software topologies, and most of all data topology.
Measuring and tuning the performance of such applications is a complex task since it requires a good understanding of potential bottlenecks that can occur at the various levels in the computational infrastructure.

From a user's perspective, the development cycle and turnaround time of an application in a grid infrastructure are important. Successful reuse of existing applications in an efficient manner is a major acceptance criterion for emerging grid environments by the computational science community. Heterogeneous grid environments however pose daunting reuseability and efficiency problems. The Study addresses all of the above issues in detail.

The report is freely available to download:
http://www.enacts.org/softwareefficiency.pdf