

A Review of UK HEC Grid Infrastructure State of the Art and Next Steps

W T Hewitt, R J Allan, J Brooke (Editors)

Authors

CLRC Daresbury Laboratory	Edinburgh Parallel Computing Centre	Manchester Research Centre for Computational Science	CLRC Rutherford Appleton Laboratory
R J Allan	M Westhead	W T Hewitt	D S Boyd
S Andrews	A Trew	J M Brooke	J Gordon
K Kleese	M Brown	J Chruszcz	R Fowler
M F Guest	A Kennedy	M Foster	C Greenough
		S Pickles	
		F Costen	
		R Hughes-Jones	

The UKHEC Collaboration prepared this report for EPSRC.

A Review of UK HEC Grid Infrastructure State of the Art and Next Steps

W T Hewitt, R J Allan, J Brooke (Editors)

Authors

Computational Science & Engineering Department CLRC Daresbury Laboratory Daresbury Warrington WA4 4AD	Edinburgh Parallel Computing Centre University of Edinburgh JCMB King's Buildings Mayfield Road Edinburgh EH9 3JZ	Manchester Research Centre for Computational Science University of Manchester Manchester M13 9PL	CLRC Rutherford Appleton Lab Chilton Didcot OX11 0QX
R J Allan	M Westhead	W T Hewitt	D S Boyd
S Andrews	A Trew	J M Brooke	J Gordon
K Kleese	M Brown	J Chruszcz	R Fowler
M F Guest	A Kennedy	M Foster	C Greenough
		S Pickles	
		F Costen	
		R Hughes-Jones	

The UKHEC Collaboration prepared this report for EPSRC.

Table of Contents

A Review of UK HEC Grid Infrastructure State of the Art and Next Steps	2
Table of Contents	3
Executive Summary.....	5
1 Introduction	7
2 The Importance of the Grid.....	9
2.1 Importance to industry.....	9
2.2 Importance to science and engineering.....	9
3 What is a Computational Grid?.....	10
4 Computational Grid Software.....	11
4.1 GLOBUS.....	12
4.2 Seamless Thinking Aid.....	12
4.3 LSF.....	12
4.4 UNICORE.....	12
4.5 Legion	13
4.6 Jini	13
5 Key Developments.....	13
5.1 Current UK Grid Activity.....	13
5.2 Related Initiatives	14
6 Why should the UK do anything?.....	14
6.1 What can we do with the Grid that we can't do otherwise?.....	14
7 What should the UK do?.....	15
7.1 What should we do?.....	15
7.2 What are the issues?.....	16
8 Conclusions.....	17
9 Acknowledgements	17
10 Bibliography.....	18
11 URLs.....	19
Appendix Summary of Activity	21
1 Summary of UK Activities - University of Manchester.....	21
1.1 Pan-European meta-computing projects	21
1.2 Grid-aware message passing libraries.....	21
1.3 Virtual Reality, Visualization and Computational Steering.....	21
1.4 Managed Bandwidth Links.....	22
1.5 Manchester Single Site Experiments.....	22
1.6 Manchester/ EPCC two site experiment	22
1.7 SC'99 Experiment	22
1.8 EuroGrid	23
2 Summary of UK Activities - EPCC.....	23
2.1 Write once run anywhere applications.....	23
2.2 Java Benchmarking	24
2.3 Network Quality of Service experiments	24
3 Summary of UK Activities - Daresbury Laboratory	24
3.1 Science Portals for the Collaborative Computational Projects.....	25
3.2 DAMP - CLRC Data Management Project	25
3.3 Daresbury and Manchester Experiments.....	26
3.4 Daresbury and RAL Experiments.....	26
3.5 Other Data Management Projects	26
3.5.1 Development of an Interdisciplinary Round Table for Emerging Computer Technologies.....	26
3.5.2 European Spatio-Temporal Data Infrastructure for High-Performance Computing.....	26
3.5.3 Access Point for NERC Data Centres.....	26
3.5.4 ESDANET	27

3.5.5	Web Operating System (WOS).....	27
3.6	Experimental Facilities on the Grid.....	27
4	Summary of UK Activities - RAL.....	28
4.1	HEP Applications.....	28
4.1.1	LHC Tier 1 Regional Centre.....	28
4.1.2	BaBar and CDF.....	28
4.1.3	Computational analysis.....	28
4.2	Other Applications.....	28
4.3	RAL, Daresbury and Manchester Experiments.....	28
4.3.1	RAL-DL experiments.....	29
5	Rest of Europe.....	29
5.1	European Grid Forum.....	29
5.2	EU R&D Projects.....	29
5.2.1	METHODIS.....	29
5.2.2	EuroGrid.....	29
5.2.3	UNICORE.....	29
6	USA Grid Development Activities.....	30
6.1	US Grid Forum.....	30
6.2	NSF PACI.....	30
6.2.1	NPACI.....	31
6.2.2	NCSA.....	31
6.3	GUSTO Consortium.....	31
6.4	NASA Information Power Grid.....	31
6.5	ASCI Problem Solving Environment.....	31
6.6	iGrid and StarTap.....	32
6.7	Illinois HPVM Project.....	33
6.8	DoE2000 Programmes.....	33
6.8.1	National Collaboratories.....	33
6.8.2	Advanced Computational Testing and Simulation.....	33
6.9	Data Management Grids in High-Energy Physics.....	33
6.9.1	US National Scalable Cluster Project.....	34
6.9.2	US Data Analysis Grid.....	34
6.9.3	GriPhyN.....	34
6.9.4	PPDG and HENP Grand Challenge.....	35
6.9.5	China Clipper Project.....	35
7	Japan.....	35
7.1	JAERI/STA.....	35
7.2	Real-World Computing Partnership.....	36
7.3	Waseda University Parallel and Distributed Computing Environment.....	36
	URLS.....	36

Executive Summary

Rapid developments in the next generation of Internet technology require urgent, concerted action from the UK academic and service community to ensure that we are not left behind. These developments are currently being driven by the needs of computational science but will have far reaching impact on the global information infrastructure. It is vital that the UK is a leader, not a follower, in this revolution.

The Internet is expanding rapidly in many different ways. For example mobile phone access to the Web represents a large qualitative shift in the way in which it will be used. The use of the Internet and Web for e-Commerce and particularly Business-to-Business transactions represents another similar shift. Ultimately the Internet will offer a proliferation of e-Services delivered via TV, phone, pager, car or virtually anything with a microchip inside. This Internet will have to be a reliable infrastructure that can provide transparent access to remote computational and data storage resources.

The same technologies offer a huge opportunity to the academic world. The ability to link and access in a common way, HPC computational facilities, large data storage facilities and experimental equipment (such as particle accelerators and telescopes) will significantly increase the quality and quantity of science that can be achieved.

The UKHEC Collaboration was funded as a Technology Watch Vehicle to carry out in-depth evaluations and demonstrations of computing technology that may enhance the speed of development and productivity of computational science research in the UK. This next generation of Internet services, the Grid, has been identified as the most important current area for evaluation.

This report outlines the state of the art in this area worldwide and identifies how UK academic research community must react to meet this challenge. Computational Science is currently leading this field because its needs represent a well defined problem, its users are used to working with "bleeding edge" technology and they understand the advantages such an infrastructure can offer them. It is clear that the UK has a very great deal to contribute, given its world-renowned achievements in pure computer science research, and a very great deal to gain from deployment of a National Computational Grid Infrastructure.

We summarise here the key findings and recommendations of this report.

- The development of a globally networked community will transform the way computational and experimental science is carried out. The demands for machine performance and network bandwidth required by the commercial sector are comparable with those of highly demanding scientific applications with the chief exception being that the commercial sector does not yet need the same tight coupling and synchronization of computational resources.
- However many sectors of the scientific community are not currently ready to exploit the GRID based model of computation. A twin track policy of continued provision of large specialist machines for scientific computation, alongside a massive collaborative effort of computational and computer scientists to produce E-Science that is firmly founded on the rigour of the scientific method, appears to be the most cost-effective approach.
- There are crucial areas in which GRID middleware is still lacking or is insufficiently tested. Electricity in a power grid is metered and ultimately paid for by the end user, it is still entirely unclear how this applies to computing resources or how the problems of funding and accountability will be solved (many large scale computational resources are currently financed by national or regional funding bodies). For industrial and defence industry use, security issues are vital. This may lead to the development of dedicated Grids accessible only to certified users.
- The UK must participate in and benefit from the world-wide development of Grid software. UKHEC can play an important role by disseminating information and by arranging seminars and workshops where contacts can be made and ideas exchanged.
- Identifying the right kind of scientific challenges will be a vital factor in the development of E-Science. The international effort of the particle physics community in preparing for the huge volumes of data from the LHC is an excellent example. The science cannot be carried out without the Grid, and in turn will drive forward the data transfer, storage and mining technology of the Grid. However this data-based model may not be suitable for areas

of science that need to steer and interact with computational simulations, such as computer-aided engineering (CAE) or virtual medicine. Other communities may thus need to form their own collaborations based around areas of common interest. Again this will require exchanges between those whose interest is the scientific vision and those whose interest is in providing the software and middleware to facilitate this.

- All Grid computing is predicated on first-class networking and this will continually need to improve as the Grid develops. The development of the UK academic network via SuperJanet4 provides an excellent opportunity in the near future and good quality international links will also be vital.

These issues are discussed more fully in sections 7, 8 and 9 of this document. However, a central lesson of the development of the Grid so far, is that the creation of a culture in which individuals and organisations are encouraged to participate in the building of Grid test-beds, is at least as important as the technical challenges of providing the middleware and software. Thus in the UK test-bed experiments described in the Appendix, the ready collaboration between the four leading computational sites in the UK is as encouraging as the technical success of the experiments themselves. In Europe, Germany has taken a lead because of a political will to unite the regional computing centres. In the UK the JREI-funded centres may be (but are not yet) a source of similar resources with many of their users already collaborating in scientific research through the CCPs. Add to this the existing high-end computing centres and a new national Tera-flop/s supercomputing system and the UK would have a pre-eminent e-Science infrastructure to enable internationally competitive research in a very wide range of disciplines.

1 Introduction

The Grid is whatever we want it to be! At the time of writing *“the grid is an emerging infrastructure that will fundamentally change the way we think about – and use – computing. The Grid will connect multiple regional and national computational grids to create a universal source of computing power”*[8].

The World Wide Web provides a ubiquitous global information space. The aim of the Grid is to extend this network functionality to provide ubiquitous computation and data storage functionality. Just as the web continues to change the way we communicate, the Grid aims to change the way we access and think of data creation via computational, experimental and observational facilities.

The term "Grid" is coined by analogy with the idea of the national power grid that provides a highly robust, standard source of electrical power to individuals and organisations. The Grid will provide the same kind of robust, standard access to a range of e-Services including knowledge, information, computational, experimental and data storage systems. The architecture of the grid is based on several layers:

- **The Computational Grid** provides raw computing power, high-speed bandwidth interconnection, and large-scale data storage. This basic layer is the subject of much research and development throughout the world and is the main topic of this paper. We note that the concepts on which the computational grid is based can be easily extended to include other non-computational facilities;
- **The Information Grid** allows easily accessible connections to major sources of information (e.g. databases storing the results of computation or experiment) and tools for its analysis and visualization;
- **The Knowledge Grid** involves using special techniques such as data mining and machine learning that give added value to the information and also provide intelligent guidance to decision makers. It would also enable services to be coupled to create new scientific opportunities, e.g. experimental or medical steering with real-time simulation.

Each community of users will use its own logical grid but all need to be based upon one common physical infrastructure (network, HPC, ...) that is a system of inter-operable components with publicly-specified interfaces. Any particular facility could be a member of more than one logical grid.

Most importantly is that whilst there is a great deal of both research and development work being done on the computational and data grid layer, and we can expect to deliver robust and usable facilities within a few years, a great deal of computer science research is needed to realise the information and knowledge layers which will create new scientific potential. Benefits may accrue from taking note of the work of related national and international initiatives such as the JISC's DNER (Distributed National Electronic Resource" programme and Inernet 2 in the USA.

E-Commerce is ruthlessly sweeping through industries and revolutionising the way in which business is carried out. It already represents a multi-billion pound sector of IT and it has only just started. This area will require a new generation of Internet technology if it is to continue to develop both in size and in the range of services it provides. It requires a highly robust, high performance grid network infrastructure with guaranteed Quality of Service.

The aspects described above represent different facets of the complex technological infrastructure that will be called the Grid. The Grid will provide common access to a range of "e-Services". An e-Service in commerce is a broad term, e.g. as defined by Hewlett-Packard, to include services offered to customers, business processes, software applications and hardware resources. Such service will allow transactions and provide solutions to problems for individuals, businesses and software agents and applications.

Commercial IT departments are increasingly building their infrastructures in a modular way, enabling them to plug into e-Services on the Net. E-Services will therefore be put to use to meet a wide range of needs and to solve a wide range of problems - in the consumer realm and in the business world. They will have the ability to discover, negotiate and transact with one another to complete a single task or set of tasks.

Whilst a lot of development work still has to be done to bring the Grid to a mature state with the (arguably) more demanding academic requirements, and the UK must contribute to this immediately, the major long-term challenge is one of culture.

The main driver of the academic Grid technology is the development of an e-Science Grid involving the connection of HPC facilities with large-scale data storage facilities and scientific equipment, such as the Jodrell Bank radio telescope or the CERN LHC particle accelerator. The scientists gain these main features:

- Common user interface to a large variety of computing and data resources to provide seamless access to these resources allowing users to store, search and process experimental data as well as run comparative simulation experiments;
- Better use of resources - jobs are run at the most suitable site, e.g., the one which will give the quickest turn around, the one which is the cheapest;
- Facilities for new ways of working, e.g., real-time processing of data from a radio telescope, including collaborative working of a group of distributed scientists;
- Aggregating computing resources to provide meta-computers.

There are three major groups of people who will contribute to the development and use of this Grid:

- **Scientists:** Those who will use these Grid technologies to solve their problems (e.g. the CCP and HPC communities);
- **Service providers:** typified by computer centre staff who provide day to day support for the infrastructure and provide both day-to-day and in-depth support to the scientist. This group of people will focus on the computational and data grid layers;
- **Computer Scientists:** who will develop new algorithms and methodologies that will be made available to the scientists by the service providers. In particular they must focus on the Information and Knowledge Grids. There is also a great deal of work behind the scenes to ensure that the Grid contains the components needed for deployment in the UK, e.g. security and accounting mechanisms and component interfaces to adapt our application layers to emerging infrastructures likely to be developed in the USA and Europe.

The next section of this report provides more information on the importance of this new technology. Section 3 describes in more detail the computational grid layer, and is followed by a short description of the major middleware packages that are under development.

Section 5 provides a short summary of the activities already being undertaken in the UK. It focuses mainly on Edinburgh, Daresbury, Rutherford and Manchester, the partners of the UKHEC collaboration. More details of this work and other work in the UK, Europe, Japan and the USA are given in an extensive appendix.

The final two sections of this report address why the UK scientific community should make a contribution and what it should be to these developments. We also aim to identify how Grid technology will be adapted and deployed in the UK and what particular issues must be addressed by the computer science community.

It is important to identify what areas this report is not addressing. Firstly we are not trying to re-present the first few chapters of [8] which describe in much more detail a vision of the Grid and its technologies. We are not addressing the scientific requirements and applications. The various research councils have that in hand, in particular to identify **all** the uses and applications of the grid technology. We are trying to focus on the infrastructure issues: what is available today, what is going on in the UK and elsewhere and what we need to do next.

This report has been prepared on a short time-scale. We hope we have mentioned all the pertinent work but do apologise if we have missed something out.

2 The Importance of the Grid

2.1 Importance to industry

The current attempts to build e-Science Grids are important not only because of the new scientific methodologies that they can facilitate but also because they will lead the way in solving critical problems necessary to build an infrastructure of much broader applicability.

The need for this infrastructure is clear. Web and e-Commerce servers require up times in excess of 99.999%. Downtime can be represented directly in lost sales. It is a difficult challenge for a company that builds (say) cars to provide this reliability in their IT infrastructure. There is a strong argument therefore for third party organisations that specialise in hosting these services over the Grid.

Furthermore to see the real benefits of this new economy, back-end business systems need to be tightly integrated into the e-Commerce system. Cisco provides an archetypal example of how to do this. Over 50% of Cisco orders are never touched by Cisco staff. They are taken automatically by the Website and transferred directly to sub-contractors. The orders are built from chips to a shipped product in three days. The system is so well integrated into Cisco's business structure that it can automatically produce, within a day, a set of auditable accounts showing the company's complete financial position.

Today Cisco is the exception. Tomorrow it will be the norm, and to achieve this many organisations will need to outsource their whole IT infrastructure.

The vision is that of client machines that provide an interface to the services and applications available on the Grid. A user can log in to the Grid anywhere in the world and be presented with the same personalised interface and access to their file space. A third-party organisation will maintain the system back up your data and install new applications for you.

This model can already be seen happening in the states with the success of Applications Service Providers (ASPs). Once the bandwidth is available it seems likely that Europe will follow. This is also the direction in which companies like Microsoft are starting to move. Their Next Generation Windows products emphasise what they call the "Programmable Web" targeted at supporting mainly e-Commerce applications using XML-aware code that can be executed remotely over the network. It is no coincidence that Microsoft is hosting the Grid Forum meeting this year at their head quarters in Redmond.

2.2 Importance to science and engineering

Recent discussions at a number of meetings have identified key points relevant to the future grid environment likely to be deployed in the UK. These include:

- Grid environments will play a pivotal role in all aspects of Engineering and Physical Science, with the potential to enhance the quality, effectiveness and timeliness of research in numerous disciplines. This impact will be felt across a broad spectrum of both experimental and simulation activities;
- The development of a viable e-Science infrastructure will require, not just an investment in hardware and network infrastructure, but also a significant investment in software development. The right software is required to glue the disparate components of the Grid together in a common framework. It is clear from the ongoing UK Grid experiments that the existing software is still very experimental.
- Many key application areas of industrial relevance in areas such as chemistry and engineering will require grid technology, e.g. in molecular design, the grid may provide a computational analogue to the technique of "high throughput screening" employed in many pharmaceutical companies. Similar combinatorial approaches can be applied to materials modelling;

- Data generation and processing associated with experimental facilities present demanding requirements that are not currently addressed, e.g. the data collection demands of the CCD detector at the ESRF are comparable with stated PPARC LHC requirements in this area. Whilst current data requirements on both the ISIS and SRS facilities are modest in comparison, the development of time-resolved experiments promises a substantial increase -- the GEM instrument on ISIS has the ability to generate 1TB per day of useful information;
- The availability of the grid will radically enhance the interaction of experiment and simulation, providing the experimentalist with a rich, easy to use portfolio of simulation and analysis tools;
- Access to a variety of GUIs and PSEs promises more effective utilisation and demand for both mid-range and high-end computing resources. Such an environment will also stimulate the industrial usage of simulation techniques by drastically reducing the effort required by the non expert to harness high-end computing technology;
- Whilst the grid will provide a drastic increase in throughput and effective utilisation of mid-range and local facilities, provide a viable alternative to the stated top-end requirements of the Science and Engineering research community. The closely coupled nature of many key applications can only be met through provision of an internationally competitive high-end machine. Any such resource should, however, be fully integrated into the future grid environment to provide flexible access to all available resources;
- The UK Research Councils must work closely with the US equivalents to exploit grid-based technology.

The potential of the grid will however only be realised through a level of effective and sustained interaction and collaboration between computer scientists and computational scientists that is not in place today. Incentives to promote this interaction represent a crucial short-term requirement. Effective exploitation of Grid-based technology will only be achieved through the provision and retention of well-trained staff. Sustaining such provision against opportunities in e-Commerce may indeed represent the greatest challenge to face e-Science.

3 What is a Computational Grid?

Grid computing describes the linking together of distributed computational resources to provide flexible access and a common interface for the user. Meta-computing extends this concept to enable distributed systems and/or supercomputers to aggregate their resources to out perform the limitations of a single computing system. To achieve these goals software systems must be provided which use Internet technology, now common in e-Commerce, for the benefit of the computational science community.

Distributed computing systems offer more than just a large CPU resource. A software environment of unprecedented quality and functionality is emerging along with the use of the Internet for E-commerce and leisure purposes. This is driven by a combination of the computer industry and the loose collection of worldwide "freeware" programmers. Geoffrey Fox has referred to this as the "Distributed Commodity Computing and Information System" [5]

In the USA and Japan there are several alliances of computing centres separated by large distances. In Europe, Germany has taken a lead because of the regional computing centres. In the UK the national facilities (CSAR, EPCC) and the JREI-funded centres may be (but are not yet) a source of similar resources.

The whole concept is often referred to as a "Computational Grid". Computers on a grid can solve very large problems requiring, for instance, more main memory than is available on a single machine. The use of these systems as single computational platforms is an active area of research, however given the high latency of wide area connections and the problems of heterogeneity such use is unlikely to be very widespread for most applications. The real value of the Grid comes instead from the ability to access remote heterogeneous resources in a common way.

A key concept is that of "ownership". A Grid is a "federation" of resources that may be accessed in a transparent way by grid users. This raises the fundamental question of "accounting" for resource usage, whether it is CPU time, disk, memory, licensed software or preserved data. Whilst this is perhaps the most important issue to be considered in

implementing a national Grid environment we do not consider it further in this report. Instead we focus on how the scientific user might benefit from such an ideal environment.

Grid-based computing is likely to become an important key technology for future UK High End Computing. We give a very brief description of some current international developments which were presented at SuperComputing'99 held in Portland, Oregon and the IEEE Workshop in Cluster Computing held in Melbourne, Australia. Further information is available from larger technical surveys [18],[3] and the book by I. Foster and C. Kesselman [8]. We do not attempt to provide an introduction to all the underlying distributed-computing techniques that are both complex and diverse. There are numerous discussions in the computer-science literature that should be consulted for background information (see e.g.,[18][11][16]).

4 Computational Grid Software

Some projects are beginning to provide the basic functionality of a computational Grid infrastructure on a distributed set of computers. We briefly describe the better-known examples to provide an idea of what type of application environments are being developed. More extensive surveys of computational grid and hierarchical clustering projects are available separately [18][3].

The table below shows how the range of facilities might be structured within the current UK funding model.

Centre	Number/Location	Purpose	Functionality	Capability/Size	Current Funding Source
Tier 0	1-2 National Centres	Ground-breaking simulations; Mission-led projects	High-end CPU engine Data store and research centre WAN/LAN	Tflop/s system 1,000+ CPUS	RCs (OST)
Tier 1	4 Regional Centres	Production work; visualization	Large CPU engine Data store and VR analysis systems WAN/LAN	250 Gflop/s 256+ CPUS	JIF/JREI
Tier 2	16 Large University Departments,	Development; visualization	Medium CPU Engine Data store and visualization systems WAN/LAN	64 Gflop/s 64+ CPUS	JREI/MPEF
Tier 3	Many Ubiquitous	Desktop access to the Grid	Competitive workstation; Collaborative working	1 Gflop/s 1+ CPUS	Research Grants

The Grid will develop as a combination of new and existing facilities with an enhanced potential to make new technology widely available in a short time.

At Tier 3 the research group and departmental level are workstations or PC "commodity" clusters that would be harnessed to both access the Grid and to provide some of the computational resource. Many universities and commercial companies have hundreds of PCs or workstations that could be coupled to the Grid to provide a range of capabilities. Some of the larger universities (Tier 1 and 2) have substantial resources that could also be integrated into the Grid for both throughput, mid-range and special purpose computing (e.g., VR facilities). The top-end platforms at one or two centres are essential to address capability computing. **The computational/data grid is essentially the integration of all of these resources.** As outlined in the Introduction this physical Grid may be partitioned into a number of logical grids.

There is no technical reason why we should stop at the national level, and UKHEC is working with a number of institutes to evaluate the opportunities offered by a world-wide Grid. These activities must, and do, involve government laboratories, large national facilities and industry and commerce.

Distributed software tools, and especially those which facilitate very complex "coupled" applications to be constructed and used are likely to be of growing interest over the coming few years. They are however difficult to implement, and it is more likely that data management or throughput services will be more common in the short term. There is however already a very wide range of packages both commercial and public domain that are relevant to Grid computing. We outline the five most significant here: GLOBUS, STA, LSF, UNICORE, Legion and Jini. A more comprehensive account can be found in [18].

4.1 GLOBUS

GLOBUS [L] is probably the largest current academic project. It involves joint work of Argonne National Laboratory and the University of Southern California's Information Science Institute with many additional contributors. Researchers are developing a basic software infrastructure for computations that integrate geographically distributed computational and information resources.

GLOBUS distributed and meta-computing concepts are being tested on a global scale by participants of the Globus Ubiquitous Supercomputing Test bed Organization (GUSTO). This is an agreement between US PACI sites to develop a Grid-computing test bed at a cost of around \$64M per year.

The Globus Grid programming toolkit is designed to help application developers and tool builders overcome the challenges in the construction of "Grid-aware" scientific and engineering applications. It does so by providing a set of standard services for user authentication, resource location, resource allocation, configuration, communication, file access, fault detection, and executable management. These services can be incorporated into applications and/ or programming tools in a mix-and-match fashion to provide access to needed capabilities.

4.2 Seamless Thinking Aid

Seamless Thinking Aid (STA) [F] is a Web-aware Java-based environment which includes a number of tools to assist parallel programming. The goal is to allow larger calculations and to couple applications with different memory or architectural requirements.

4.3 LSF

Load Sharing Facility (LSF) is a product of Platform Computing, widely used for corporate computational resource management, especially in the engineering industry [CC]. Platform aim to provide the best application resource management solutions for enterprise, allowing administrators to intelligently harness and leverage the maximum power from their existing computing systems by using idle cycles in a flexible and dynamic manner. The system has a broad range of academic and commercial users.

As well as monitoring load information such as, CPU queue length and utilisation, available user memory, paging and disk I/O rate, etc. LSF provides facilities to transfer work between locally managed or remote systems, e.g. to access machines with particular software licenses. This can work over autonomous and widely separated sites. Platform Computing is pioneering an open distributed resource management initiative with a number of other partners.

4.4 UNICORE

Uniform Access to Computing Resources [FF] was originally a project funded by the German Federal Republic to connect together several important regional super-computing centres. The strong federal political structure of the DBR makes this grid-based model particularly relevant and provides a grid environment that is also a very suitable model for a grid connecting the supercomputing centres of the whole EU.

UNICORE lets the user compose and edit structured jobs with a graphical job preparation client on a local workstation or PC. Jobs can be submitted to run on any platform in the UNICORE grid, and the user can monitor and control the submitted jobs through the job monitor client.

4.5 Legion

Legion [B] is an integrated grid-computing system that, like Globus, has been deployed at a number of sites in the USA. It arose from an object-based software project at the University of Virginia beginning in 1993.

Legion supports existing codes written in MPI and PVM, as well as "legacy" binaries. Key capabilities include:

- Eliminating the need to move and install binaries manually on multiple platforms;
- Providing a shared, secure virtual file system that spans all the machines in the system;
- Providing strong PKI-based authentication and flexible access control for user objects;
- Supporting remote execution of legacy codes, and their use in parameter space studies.

Legion is the second grid project that has been adopted by the US NSF at its National Partnership for Advanced Computational Infrastructure (NPACI) sites. NASA and the DoD are also running Legion test beds.

4.6 Jini

Jini [T] is a Java middleware technology designed to support the general requirements of federating network resources. Whilst Jini will not currently support an HPC grid it represents the natural direction for grid technology and it is likely that either Jini, or something that builds on its design concepts will be an integral part of the Grid of the future.

A Jini system is a distributed system based on the idea of federating groups of users and the resources required by those users. Jini leverages the Java environment to provide systems that are far more dynamic than is currently possible in networked groups where configuring a network is a centralised function done by hand. Although Jini uses Java, a Jini Grid could support the execution of code written in an arbitrary language.

5 Key Developments

5.1 Current UK Grid Activity

The main activities within the UK are summarised in the first four sections of the Appendix that details the work going on at Manchester, Edinburgh, Daresbury and Rutherford. These groups are unique in that they have a dual role in the Grid: that of service providers and that of research and development. Much of the work reported follows the investigations they are undertaking addressing "What type of service should we be offering in a few years time?"

All have installed Globus and are undertaking a number of single-site, and dual-site experiments with some success. Unicore is installed at Manchester. The use of these environments is growing but major problems come from immaturity of the software. None of the sites would envisage either Globus or Unicore being used in a production environment yet.

Daresbury is developing science portals for the collaborative computational projects (CCPs) and is involved in a number of large data management projects.

Edinburgh is also working on the use of Java for the development of Grid middleware and the development of tools to support scalable network quality of service

Rutherford is active in developing Grid-based computing and data infrastructures for a number of PPARC projects, particularly the LHC experiments at CERN, BaBar at SLAC and CDF at Fermilab. They are also working on distributed computational analysis and collaborative visualization.

Manchester is part of the EU funded EuroGrid project that is developing Unicore, is working on Globus, and is developing virtual reality, visualization, computational steering and meta-computing technologies. The latter is

through a trans-continental network between Stuttgart, Pittsburgh and Tsukuba. Work is also being undertaken to investigate the performance of managed bandwidth links between Manchester and RAL.

5.2 Related Initiatives

Over the last three years much work has been done in the UK and the USA regarding shared access to strategic datasets and information resources over the web. In the UK an initiative called the "Distributed National Electronic Resource - DNER" has been launched by the Joint Information Systems Committee (JISC). There is a close cooperation with the National Science Foundation in the States and it is now beginning to realise solutions to issues such as interoperability and authentication for distributed data resources. JISC is also working closely with the Coalition for Networked Information in the States and with Internet 2. Ken Kligenstein, Director of the Internet 2 Middleware initiative acknowledges the role of middleware as "managing complexity" and it is clear that we can gain from others experience in areas that are attracting such international interest and R&D.

MIMAS (formerly MIDAS) at Manchester Computing is the largest of the JISC supported national data centres hosting more than 40 strategic datasets and offering services to over 11,000 users from more than 180 institutions in the UK and beyond. As such it is a key player in the DNER and is involved in a number of projects addressing accessibility, interoperability, and authentication. This requires observation of relevant standards: protocols and profiles for the transmission and storing of data, metadata, and data formats. Also the adoption of a common national authentication service "Athens". The data collections policies of both the JISC and ESRC are currently important to the strategic development of national data centres such as MIMAS. As data and information become an intrinsic layer in the grid it would be appropriate for a cooperative collections policy to emerge that would drive forward the strategic direction for data and information provision in the UK. Further information about the organisations and initiatives mentioned above can be found at [II], [JJ], [KK] and [LL].

6 Why should the UK do anything?

It is critically important that the UK research community makes significant contributions to the emerging Grid technology. There are two reasons for this:

- In the medium term the Grid represents the next step in the development of the global information infrastructure. It is very important for industry and commerce that the UK is a leader in the development and exploitation of this new technology;
- In the short term the Grid provides important opportunities for researchers in the physical, environmental and biological sciences to collaborate and develop new inter-disciplinary projects. In order to retain our international standing in these areas we need to provide the infrastructure to be able to compete.

6.1 What can we do with the Grid that we can't do otherwise?

The introduction mentions briefly what the Grid might offer over and above the existing infrastructure. Here we present a number of scenarios on how an individual scientist might benefit from at least a computational and data grid.

Quicker/easier job turnaround: The aim of the Grid is to provide a seamless interface to many compute resources. Thus, **in the future**, a user may request from his or her workstation that the Grid undertake some simulation. A super-scheduler would identify the most suitable high-end computer, arrange for the relevant databases or files to be migrated to the supercomputer centre, the job to run, the results to be stored at the most appropriate data centre and the user to be informed when the job has completed.

Real-time access and processing of experimental data: A researcher at site X may, for instance, be using the Jodrell Bank telescope to browse galaxies. He or she may reserve a number of supercomputers to provide a meta-computer, the telescope, and network bandwidth, so that as the (Gigabytes/second) of data are collected from the telescope, they can be processed in real-time on the meta-computer and the results appear also in real-time on the

researcher's desk. He or she would use the Web interface to control the movements of the telescope. Similar scenarios are possible on other instruments, e.g. Synchrotron, and have already been implemented, for instance in the French BARNES Web software project.

Collaborative Working: Imagine this time that the user identified above needs to discuss with other colleagues throughout the UK where to direct the telescope. Using a collaborative working environment, with virtual reality and audio facilities and dedicated bandwidth to multiple sites, would enable the above experiment to benefit from the collected experience of many researchers in controlling the instrument.

All of these scenarios lead to better use of resources and quicker response time for the user, environments which foster better and quicker insight into the science and hence improve the “throughput” of the science, and better return on investment for the resource providers.

7 What should the UK do?

7.1 What should we do?

The UK scientific community, including the service providers and vendors, needs to respond to address the short, medium and long-term issues.

In the short-term we must:

- Coordinate the UK activities in this area. UKHEC provides a convenient forum to undertake this work and we would wish to continue for a longer period;
- Undertake a detailed requirements analysis of the scientific needs;
- Contribute to the development of a number of technologies, e.g., Globus, Unicore to ensure they have the facilities the UK requires;
- Foster the relationship between the application scientists, service providers and computer scientists to develop a proper understanding of the capabilities and needs of each other;
- Foster international relationships, particularly to the rest of Europe and the USA;
- Invest resources in a coordinated research, development and service medium-term programme that establishes a UK Grid test bed with a small number of nodes. These nodes should contribute to centres of excellence and expertise and should focus on demonstrating opportunities arising from the computational and data grid layers. These centres of excellence should bring together both service providers and computer scientists to work on the further development of the infrastructure and to address issues peculiar to the UK. Service providers should work with application scientists to develop applications using today's technologies that can capitalise upon the near-term capabilities of the Grid.
- Building a UK grid is not just a matter of providing networking and computational hardware. The problem of gluing these resources together in the right way is still an open research area. There is a great deal of software available but it is all experimental, incomplete and has been developed in advance of any standards in this area. There is a great deal of work to be done in developing the next generation of Grid middleware and in particular issues of inter-operability of components and interfaces must be addressed;
- Establish a long-term IT programme to undertake research into the Information and Knowledge Grids.

By the medium term we mean the application of today's computing technology to scientific problems, and by the long term we mean the development of new computing methodologies that can be applied to the scientific areas in 5 to 10 years

It must be emphasised that Grid technologies will not be mature enough for some years to come to enable traditional access to large national data and compute resources to be replaced by Grid resources. A number of scientists have

identified their immediate need as that of a multi Tflop/s sustained capability that cannot be provided through distributed aggregated resources. However it is also important that national high performance facilities, including experimental and observational instruments, be integrated into the Grid and form the core of it.

7.2 What are the issues?

There are a number of policy issues that need to be solved that can be exemplified by the following scenario. Suppose that the existing "JREF" machines can be connected to form a Grid. These machines are typified by a number of 32 or 64 processor systems including SGI Origin 2000, IBM SP, Sun and Compaq systems. (In reality these would be combined with the national HPC and data resources at CSAR, MIMAS, EPCC and RAL). But:

- In principle there is no spare resource on these "JREF" machines. To use for example 1 Tflop/s somebody has to pay for it;
- Who will install, maintain and support the Grid software?
- Who will be responsible for allocation of resources, both the peer review aspects and the system administration aspects;
- Who funds user X's use of another system?
- When a user's job has failed who does he or she contact to find out why and where?
- What is the unit of work that can migrate around the Grid? Is it worth migrating a 32 processor ten-minute job to another site if waiting would enable the job to finish in another two minutes?
- How does use of this aggregated facility interact with the national facility, particularly as the Research Councils have a contract for this facility until 2004?

Some of the technology gaps that have become focus areas for computational Grid and related research are:

- Execution environments that are portable and scalable;
 - Resource management
 - Data storage and movement
 - Security and authorisation
- Tools that enable the use of the execution environment;
 - Automated tools for porting legacy code
 - Collaborative problem solving environments for complex scientific and engineering tasks to extend the capacity of teams
 - Formal, portable programming paradigms, languages and tools that express parallelism and support software synthesis and re-use
- Development and execution environments to support applications of the future;
 - Application software that can make use of up to 10,000 processors
 - Methods for coupling multiple physics applications for analysis and optimisation and multi-disciplinary research
 - Access to inter-disciplinary data
- Design and architecture to integrate execution environment, user environment and applications.

Software implementations for meta-computing are often described in distinct software layers. Typical of this approach is the Integrated Grid Architecture proposed by the Ian Foster for the US Grid Forum [P]. Its four components are:

- Grid Fabric: the lowest level with primitive mechanisms provide support for high-speed network I/O, differential services, instrumentation, etc.
- Grid Services: the typical middleware level with a suite of grid-aware services implementing authentication, authorisation, resource location, event services, etc.

- Application Toolkit: provides more specialised services for diverse application classes, e.g., data-intensive, visualisation, distributed computing, collaborations, problem solving environments (PSE);
- Grid-aware Applications: implemented in terms of grid services and application toolkit components.

Some technical issues that need to be discussed include:

- End-to-end resource management and adaptation techniques able of provide application-level performance guarantees despite dynamic resource properties;
- Automated techniques for negotiation of resource usage, policy and accounting in large-scale grid environments;
- High-performance communication methods and protocols;
- Infrastructure and tools to support data-intensive applications, advanced tele-immersion concepts and new problem solving environments;
- Meta- or super-schedulers;
- Resource accounting and billing mechanisms;
- How to involve IT experts in enhancing the grid infrastructure e.g. new s/w, tools, algorithms;
- How to adapt applications to use it e.g. steering, visualisation;
- How to encourage e-service providers to make facilities available: CPU, storage, VR, instrumentation;
- What are the networking implications for data retrieval, computational steering, remote visualisation and "collaborative working";
- Role of "top end" system: flexible access;
- Role of 3rd party software providers e-services and licenses on demand;
- Network quality of service;
- Security issues, including distributed access and authentication.

8 Conclusions

There is significant body of work being undertaken in the UK by both the service providers and the computer scientists in developing the next generation of infrastructure required by the UK science base. This work must be coordinated and developed but it must be tempered with realism. The Grid will deliver many things and some of these elements are available now, albeit in an experimental way, and some will be delivered incrementally. Most importantly a fully featured Grid will not be available for some years, and only then if sufficient investment is made to research and develop the required infrastructure and new technologies.

Achieving such a goal requires a high level of immediate funding to:

- address the infrastructure deployment and related computer-science issues through basic research and evaluation and infrastructure funds;
- to create an increased awareness of the potential of the Grid for end users and adapt their existing applications;
- to provide ongoing support for new applications, methodologies, Web interfaces and inter-disciplinary projects; and
- promote awareness and integrate deliverables from other key national and international initiatives such as JISC's DNER programme and Internet 2 in the States

9 Acknowledgements

The preparation of this report, and some of the work described, was funded by EPSRC through grant GR/N09688 to the UKHEC Collaboration at Daresbury Laboratory and the Universities of Edinburgh and Manchester. We thank

colleagues at RAL, Manchester and Edinburgh who are involved in the PPARC HEP Data Initiative and with whom many discussions have taken place. We also thank staff of Hewlett-Packard and Platform Computing for information about e-Services in commerce.

10 Bibliography

- [1] A.S. Grimshaw, W.A. Wulf, J.C. French, A.C. Weaver & P.F. Reynolds Jr. *A Synopsis of the Legion Project* Technical Report CS-94-20 University of Virginia, 1994
 A.S. Grimshaw, W.A. Wulf, J.C. French, A.C. Weaver and P.F. Reynolds *Legion: the next Logical Step toward a Nationwide Virtual Computer* Technical Report CS-94-21 (University of Virginia, 1994)
 A.S. Grimshaw and W.A. Wulf *Legion -- a view from 20,000 feet* in ``Proc. 5th IEEE International Symposium on HPDC'' IEEE Computer Society Press (1996)
 A.S. Grimshaw, W.A. Wulf, J.C. French, A.C. Weaver and P.F. Reynolds Jr. *The Legion Vision of a Worldwide Virtual Computer* CACM 40 (1997)
- [2] E. Gabriel et al. *Implementing MPI with optimised algorithms for meta-computing* in ``Proc. 3rd MPI Developers' and Users' Conference'' (MPI Software Technology Press, Mississippi, 1999) [Q].
- [3] F Costen, J M Brooke, R J Allan and M Westhead *Grid Based High Performance Computing* Technical Report (UKHEC, 2000) , see www.ukhec.ac.uk
- [4] F. Costen, S. Pickles and M. Pettipher *Hierarchical Clustering* Technical Report, Manchester Computing, University of Manchester, 2000. For more information, see also: [J], [Z] and [AA].
- [5] G.C. Fox and W. Furmanski *Petaops and Exaops: Supercomputing on the Web* IEEE Internet Computing 1 (1997) 38-46. Also available at [Y].
 G.C. Fox, W. Furmanski, T. Haupt, E. Akarsu and H. Ozdemir *HPcc as High-Performance Commodity Computing on top of Integrated Java, CORBA, COM and Web Standards* in ``Euro-Par'98: Parallel Processing'' D. Pritchard and J. Reeve (eds.) LCNS 1470 pp55-74 (Springer 1998) ISBN 3-540-64952-2.
 G. Fox and W. Furmanski *Java for Parallel Computing and as a general Language for Scientific and Engineering Simulation and Modelling* Concurrency: Practice and Experience 9 (1997) 415-26.
 G. Fox et al. *Using WebHLA to integrate HPC FMS Modules with Web/ Commodity based Distributed Objects Technologies of CORBA, Java, COM and XML* in ``High Performance Computing 1999'' A. Tentner (ed.) (Society for Computer Simulation International, 1999) pp 273-8 ISBN 1-56555-166-4
- [6] H. Koide et al. *MPI-based communication library for a heterogeneous parallel computer cluster, STAMPI* JAERI [E].
- [7] H. Koide *STAMPI* see [N].
- [8] I. Foster and C. Kesselman (eds.) *The Grid: Blueprint for a new Computing Infrastructure* (Morgan Kaufmann Publishers, 1998) ISBN 1-55860-475-8. Abstracts of chapters and ordering information from [V]
- [9] I. Foster and K. Kesselman *GLOBUS: a Meta-computing Infrastructure Toolkit* Int. J. Supercomputing Applications (1997) 115-28
- [10] I. Foster and K. Kesselman *The Globus Project: a status report* IPPS/SPDP'98 Heterogeneous Computing Workshop S.4-18 (1998) [GG].
- [11] K. Hwang and Z. Xu *Scalable Parallel Computing: Technology, Architecture, Programming* (WBC/McGraw Hill, 1998) ISBN 0-07-031798-4
- [12] K. Kleese *Requirements for a Data Management Infrastructure to support UK High-End Computing* Technical Report DL-TR-99-04 (Daresbury Laboratory, 1999)
- [13] L.H. Turcotte *A Survey of Software Environments for Exploiting Networked Computing* Technical Report MSU-EIRS-ERC-93-2 (Mississippi State University, 1993)

- [14] Matthias A. Brune, Graham E. Fagg, Michael Resch, *Message-Passing Environments for Meta-computing* Future Generation Computer Systems 15 (1999) 5-6 pp. 699-712
- [15] PACX Edgar Gabriel, Michael Resch, Thomas Beisel, Rainer Keller, *Distributed Computing in a heterogenous computing environment* in "Recent Advances in Parallel Virtual Machine and Message Passing Interface" Vassil Alexandrov, Jack Dongarra (Eds.) Lecture Notes in Computer Science (Springer, 1998) pp 180-8
- [16] Presidential Information Technology Advisory Committee Report (1998)
- [17] R. Buyya (ed.) High Performance Cluster Computing: Architectures and Systems, Volume 1 (Prentice Hall, NJ, USA, 1999)
R. Buyya (ed.) High Performance Cluster Computing: Programming and Applications, Volume 2 (Prentice Hall, NJ, USA, 1999)
- [18] R. Orfali and D. Harkey *Client/Server Programming with Java and CORBA* (Wiley, 1997) ISBN 0-471-16351-1
- [19] R.J. Allan *Survey of Computational Grid, Meta-computing and Network Information Tools* Edition 2. Technical Report DL-TR-99-01 (Daresbury Laboratory, 2000), see www.dl.ac.uk/TCSC/HPCI/reports.html
- [20] CSAR Web pages [H].

11 URLS

- [A] <http://hla.dmsi.mil>
- [B] <http://legion.virginia.edu/overview.html>
- [C] <http://now.cs.berkeley.edu>
- [D] <http://nscp.upenn.edu>
- [E] <http://ssp.koma.jaeri.go.jp/en/stampi.html>
- [F] <http://stasrv1.koma.jaeri.go.jp/en>
- [G] <http://www.ca.metsci.com>
- [H] <http://www.csar.cfs.ac.uk>
- [I] <http://www.csar.cfs.ac.uk/staff/costen/sc99.html>
- [J] <http://www.dgs.monash.edu.au/~rajkumar/cluster>
- [K] <http://www.egrid.org>
- [L] <http://www.globus.org>
- [M] <http://www.globus.org>
- [N] <http://www.globus.org/mpi/related.html>
- [O] <http://www.gridforum.org>
- [P] <http://www.gridforum.org/iga.html>
- [Q] <http://www.hlr.de/people/resch/PAPER/pubs.html>
- [R] <http://www.ietf.org>
- [S] <http://www.jaeri.go.jp/english/index.cgicomp/comp.html>
- [T] <http://www.javasoft.com/jini>
- [U] <http://www.lanl.gov/asci>
- [V] <http://www.mkp.com/grids>
- [W] <http://www.nas.nasa.gov/Groups/Tools/IPG>
- [X] <http://www.nas.nasa.gov/NAS/Tools>
- [Y] <http://www.npac.syr.edu/users/gcfpetastuff/petaweb>
- [Z] http://www.phptr.com/ptrbooks/ptr_0130137847.html
- [AA] http://www.phptr.com/ptrbooks/ptr_0130137855.html
- [BB] <http://www.phys.ufl.edu/~avery/mre>
- [CC] <http://www.platform.com>

- [DD] <http://www.rwcp.or.jp/lab/mpperf>
- [EE] <http://www.startap.net>
- [FF] <http://www.unicore.org>
- [GG] <http://www-fp.globus.org/documentation/papers.html>
- [HH] <http://www-fp.globus.org/testbed>
- [II] <http://www.jisc.ac.uk/nsf/index.html>
- [JJ] <http://www.jisc.ac.uk/pub99/internet2.html>
- [KK] http://www.jisc.ac.uk/pub99/dner_desc.html
- [LL] <http://www.mimas.ac.uk>

Appendix Summary of Activity

1 Summary of UK Activities - University of Manchester

Manchester Research Centre for Computational Science (MRCCS) has been involved in a number of European projects aimed at developing pan-European Grid computing. MRCCS has also developed active links with major Grid projects in the USA, Europe and Japan which culminated in a trans-global meta-computing experiment exhibited at SuperComputing'99. This experiment has aroused worldwide interest and it was awarded first place in the HPC Games challenge at SC'99 and the Best Paper award at the recent HPCN'2000 conference. The active participation of a world-class radio astronomy facility at Jodrell Bank was an important factor in this success. The collaboration between MRCCS and Jodrell is continuing and will be a component of the EuroGrid project. MRCCS is involved in several major European projects utilising the Grid for medical applications and a Virtual Reality Centre is currently being installed which will add an extra dimension to the use of the Grid.

1.1 Pan-European meta-computing projects

MRCCS worked with a local computational engineering company to utilise the power of metropolitan and pan-European broadband ATM networks. These projects involved a number of technical work packages concerned with network QoS issues. They also included a cost benefit analysis of Grid enabled computational engineering and an analysis of the changes in business practice that access to the Grid would herald[A].

MRCCS has established ongoing working partnerships with major European (METHODIS), Japanese (STA) and US (Globus) projects. These are described elsewhere in this document, as is the recently established pan-European EuroGrid project that will make the UNICORE [B] technology available to the ongoing UK Grid experiments. MRCCS is also involved with users of the CSAR service running coupled ocean-atmosphere models running on different architectures and have discussed extending this work to European collaborations with CERFACS (France) who are a world-leading site in coupled environmental models.

1.2 Grid-aware message passing libraries

MRCCS is working with HLRS (Stuttgart) and JAERI (Japan) to Grid-enable codes using MPI. This work involves comparative testing of the PACX-MPI and STAMPI libraries that enable MPI codes to run between different machines and different architectures while retaining the performance and functionality of the vendor-tuned MPI implementations within each machine.

1.3 Virtual Reality, Visualization and Computational Steering

The Grid will make possible access to specialist facilities in virtual reality and visualization. However, there are problems to be solved in delivering the output of these specialist facilities in formats that do not place unacceptable demands on network bandwidth and on the power and sophistication of the end-user software. MRCCS is working with European partners to deliver remote access to specialist facilities via standard WWW browsers. Thus much of the rendering and graphical manipulation is carried out centrally and only moderate technology is necessary at the user end. This is an excellent illustration of the concept of the Grid. For more details see the NOVICE project WWW pages at [A]. MRCCS is also producing reports on such technology as part of the UKHEC, currently working with the Terra project which produces more data from 512-processor simulations than can be handled by local visualization tools alone.

An exciting extension of such work is to allow the steering of applications on high-end machine via specialist visualization and VR environments that can interact with the simulation, over Grid-enabled links. MRCCS is collaborating with workers at Oxford and Salford Universities to develop applications and demonstrations of steered computations.

1.4 Managed Bandwidth Links

Experiments are being undertaken to measure the performance of the managed bandwidth links between Manchester and RAL, and comparing performance with the shared bandwidth of SuperJANET.

1.5 Manchester Single Site Experiments

Globus 1.1.2 has been installed on a variety of large and small-scale machines at Manchester including Linux/x86, Solaris/sparc, Irix/mips (Origin 2000) and Unicos-mk/alpha platforms. The software was found; in its present state to be portable and installation was time consuming rather than problematic. Operation was found to be overly reliant on a centralized remote LDAP (grid-information) server, with frequent problems arising through timeouts as a result of high traffic levels and service outage. Installation of a mirror server using Netscape LDAP server 4.11 was found to alleviate these problems and allow normal use of the software.

Job submission tests between machines proved successful, together with the redirection of output and error files using both NQE/NQS batch systems and foreground execution with a variety of MPI based codes.

MRCCS is currently in the process of performance testing the Globus aware MPI implementation (MPICH/G) against the wide area MPI implementation e.g., PAC/X and Silicon Graphics' MPI for coupling various Origin 2000 servers across the local campus. Results are expected in June. Discussions are also in progress to determine whether Globus can be used in the pulsar search code employed in the SC'99 experiments [A]. Here a loosely coupled code is decomposed based on available bandwidth between nodes and is reasonable latency tolerant. Globus' ability to report network properties is clearly advantageous in this area and may ultimately lead to coupling of the Jodrell Bank radio telescope with real-time simulation and computational steering.

MRCCS is also working with the UNICORE development team to install UNICORE on various machines at Manchester and will become the Unicore certification centre for the UK. An important point here is that the UNICORE encryption technology is developed in Europe and is free from US encryption licensing restrictions.

1.6 Manchester/ EPCC two site experiment

We have built and deployed Globus 1.1.2 on two Cray T3E machines in distinct administrative domains (EPCC, Manchester). Job submission is enabled to each machine via the NQE job-manager supplied with Globus. This is found to allow job submission to either of the two machines, from any Globus-aware machine.

We have constructed a Globus-aware job submission tool in prototype form, which can record the specifications of a batch job, including file transfers that may be needed before the job can commence. This user interface is able to submit the specified job to either machine based on manual selection. Additionally the choice of machine can in principle be made by a broker, either as part of the client, or preferably a third party service and this is simulated within the client. The introduction of a third party broker raises important issues for Grid middleware, in that existing batch queue systems are not able to provide either a guaranteed time of execution or even a realistic estimate of same. Solutions that are based on usage history or a learning algorithm have been proposed.

In conclusion we find that the Globus software is able to run on the T3Es and batch job submission is possible between the two machines though the implementation of the job-manager may be restrictive to users. We also note that Globus provides sufficient infrastructure to allow the resources to be load balanced, though development of this environment is probably a few years away. Globus itself also shows signs of immaturity in some important areas, which suggests that it is not yet ready for production service use. Particular areas include resource discovery (Grid Information Services), which is liable to change in the next release.

1.7 SC'99 Experiment

At the SC'99 "HPC Games", an intercontinental team consisting of computational scientists, networking and systems specialists in Stuttgart, Manchester, Pittsburgh and Tsukuba (Japan) was awarded the top prize for the most challenging scientific applications, executed live across the planet from the Portland Convention Centre [A].

A molecular dynamics simulation with over two million particles ran concurrently on a Hitachi SR8000 at ETL (Tsukuba), and on CRAY T3E's at the Pittsburgh Supercomputing Center, Manchester and HLRS Stuttgart. This Ter(r)acomputer spanning more than 10,000 miles has a total peak performance of 2.2 Tflop/s.

A second application demonstrated was a flow solver called URANUS. A simulation of the crew-rescue vehicle (X-38) of the international space station with 3.6 Million cells on 1536 T3E processors was accompanied by a visualisation of the flow around the vehicle in a collaborative session with the European Networking Demonstration booth.

A third application was chosen that analysed radio-astronomy data in search of pulsars. For this application, sufficient bandwidth between the different computers is crucial. Between the three T3E systems, a system of networks consisting of JANET and Teleglobe (UK), DFN (Germany), and Abilene and vBNS (USA) delivered sustained bandwidths in excess of 1 Megabit per second.

The Manchester application adapts to actual bandwidth conditions by varying the amount of work it assigns to each machine. The molecular dynamics and fluid dynamics applications are optimised to mask latency by overlapping communication and computation.

Message passing between the heterogeneous machines comprising the Ter(r)acomputer is done by means of PACX-MPI, a library developed at HLRS (Stuttgart). This is implemented as a large subset of the MPI-1 standard, allowing immediate "grid-enabling" of most application codes that use MPI.

1.8 EuroGrid

EuroGrid has just been accepted for funding by the EU and will be the first pan-European Grid project. Participating sites are MRCCS (UK), IDRIS (Fr), FZ Juelich (Ger), Parallab (N), ICM (PI). Pallas, who is also involved with UNICORE, coordinates the project and technical input will be provided by FECIT who helped develop UNICORE. Industrial partners include AEROMATRA-CCR and government organisations are represented by DWD, the German Meteo.

EuroGrid aims to provide the middleware needed to make job submission across the European Grid a reality and to provide middleware to act as a resource broker to give the European Grid the possibility of becoming self-financing. This illustrates an important point in the development of the Grid, initially governmental funding will be used to put the Grid in place but the concept will only be successful if it can be utilised commercially and increases the quality of life (via innovations such as tele-medicine and tele-education). EuroGrid also recognises that, alongside technical developments, it is vital to foster an atmosphere of cooperation between those sites which participate in the Grid, and these organisational issues can sometimes be forgotten in enthusiasm over the Grid concepts.

2 Summary of UK Activities - EPCC

EPCC has been working on grid-based technology for over three years. Its activities have focused on two areas:

- The use of Java, both for the development of grid middleware and as a way of producing highly portable applications to run in a heterogeneous grid environment;
- The development of tools to support scalable network quality of service (Differentiated Services).

2.1 Write once run anywhere applications

The Hitachi Parallel Taskfarm (HPT) was a project involving collaboration between EPCC and Hitachi Europe Ltd. It demonstrated some of the enormous potential power of the "write once run anywhere" paradigm by extending it to parallel applications.

The HPT was a middleware solution that allowed the same task-farm code to run in serial, or in parallel on an SMP or MPP platform without even the need for recompilation. The system also provided a client server interaction so

that a user could export their code from the desktop to a compute server for execution. This code could be the same, unmodified code that ran on their desktop.

2.2 Java Benchmarking

The Java Grande Forum (JGF) provides a unified community voice to address Java language design and implementation issues relevant to "Grande" (large-scale or HPC) programming.

EPCC leads the Java Grande Forum Benchmarking effort. The aim is to produce a benchmark suite aimed at testing aspects of Java execution environments (JVMs, Java compilers, Java hardware etc.), pertinent to Grande applications. The work involves constructing a framework for the benchmarks, designing a Java instrumentation class to ensure standard presentation of results, and seeding the suite with existing and original benchmark codes.

The aim of this work is ultimately to arrive at a standard benchmark suite that can be used to:

- Demonstrate the use of Java for Grande applications. Show that real, large-scale codes can be written, and provide an opportunity for performance comparison against other languages;
- Provide metrics for comparing Java execution environments, thus allowing Grande users to make informed decisions about which environments are most suitable for their needs;
- Expose those features of the execution environments critical to Grande Applications, and in doing so encourage the development of the environments in appropriate directions.

This work adopts a standard approach, ensuring that metrics and nomenclature are consistent, which is important in order to facilitate meaningful comparisons in the Java Grande community. The work could be extended to evaluate a fuller Grid based computing environment for HPC applications.

2.3 Network Quality of Service experiments

The ability to provide guarantees of network performance is crucially important to the support of a reliable Grid infrastructure. The new IETF standard for Differentiated Services (DiffServ) is the obvious way to meet this need. However, the architecture and configuration of DiffServ networks is a complex and, as yet, poorly understood problem. Furthermore the specific requirements of Grid based network traffic have yet to be addressed by the standard.

EPCC is involved in a large project in collaboration with Cisco systems to develop modelling software to address this problem. The Intersim simulator will enable network engineers to investigate the effects of network changes or increased traffic demands on the quality of service provided by a network. This tool will be extremely useful in designing Grid networks. It will also be very important in contributing to the definition of DiffServ standards for Grid traffic and EPCC will be involved in leading these standards.

3 Summary of UK Activities - Daresbury Laboratory

The Computational Science and Engineering Department has been evaluating the Globus software at Daresbury Laboratory. Experiments have been carried out under the auspices of the UK High-End Computing Collaboration (grid-based HPC applications) and the EPSRC Distributed Computing Programme (resource management software).

Globus v1.1.1 has been installed on a cluster of four IBM power-PC desktop systems and a 32-processor Pentium Beowulf. The former uses LoadLeveler to share work locally and the latter uses PBS from MRJ Inc. Both these systems appear to work successfully with *mds.csun.ecs.edu* as the MDS server but an upgrade to v1.1.3 is required for more inter-site tests.

3.1 Science Portals for the Collaborative Computational Projects

We are currently assessing "active" portals to grid-enabled scientific tools for specific areas of research based on knowledge of the capabilities of computational Grid "middleware" technology and activities in e-Commerce where a very powerful and diverse environment involving contracted-out e-Services is being deployed. Very clear reasons for this sea change in the use of computational and knowledge acquisition resources include:

- Cost effectiveness through enhanced efficiency in hardware resource management and software payment on demand (sometimes referred to as "apps-on-tap");
- New ways of working;
- Business-to-business service negotiation leading to opportunities for new market or research areas as yet unexplored (inter-disciplinary science).

In extending this concept to the academic research arena by developing application portals based on the distributed computational Grid we may look to enhance the following key capabilities:

- Transfer of efficient IT best practice from commercial R&D to academic research;
- Transfer of latest academic research capabilities and results to the commercial sector via Internet portals for direct integration into e-Commerce models.

A portal consists of:

- Single-point Web access to information and "active" resources (accessible via payment transfer);
- Collection of all resources relevant to a particular "theme" including data, simulation facilities, experimental instruments, collaborative working and publication tools.

Active resources are defined to be those that generate new knowledge by simulation, measurement or data fusion and exploration.

The aim of a portal as we define it here is to escape from dependency on specific computational platforms, experimental instruments or conserved data but to make best use of resources available with guaranteed quality of service (through resource and network management middleware) and accreditation of services to give scientific dependability.

The existing CCPs and their Web sites may form a good basis for UK science-oriented portals.

In the immediate future we need to generate cost estimates for transforming these pages into interfaces to a range of simulation and other facilities with intelligent means of checking the input requests, validating and logging the output.

A Working Group and Web site has been set up for this project to act as a focus for development of "HPC Applications and Testbeds". See URL www.dl.ac.uk/TCSC/UKHEC/GridWG.

3.2 DAMP - CLRC Data Management Project

General information about the DAMP project can be found under [E]. Results of surveys and research projects have been published, for a complete list of publications see [F]. A list of all UK data holdings containing climate research data has been compiled and is available under [G] and a similar survey is being carried out for other scientific areas. A Working Group has been established to coordinate and promote data management activities within CLRC and UKHEC. More information on data services and research projects can be found at [H].

3.3 Daresbury and Manchester Experiments

Serial jobs have been submitted to the four-processor IBM cluster at DL from a Linux system at Manchester. A simple test of a portable Fortran FFT package (GPFA) was used. Jobs were run in several modes on a 120 MHz Power-PC tci23.

- Submitted interactively with a peak performance of 61 Mflop/s;
- Submitted to LoadLeveler from tci18 with a peak performance of 67 Mflop/s;
- Submitted to Globus and LoadLeveler from tci18 with a peak performance of 64 Mflop/s;
- Submitted to Globus and LoadLeveler from bert (Manchester) with a peak performance of 65 Mflop/s

All files were stored in NFS on the host machine of the cluster, which happens to be the same as the gatekeeper tci18. Clearly there is some variability in the time because the systems were also being used as desktop machines. There however seems to be no degradation in performance introduced by the additional software layers.

Similar tests were performed between Manchester and the Daresbury Pentium Beowulf system.

3.4 Daresbury and RAL Experiments

Experiments between Daresbury and RAL focus on the Linux clusters - a 32-processor Pentium system using PBS and 16 x 2-processor Alpha system using QSW's RMS and LSF at Daresbury and a 16x2-way Athlon system using PBS at RAL. At a later stage the 50-processor IBM SP at DL and 24-processor Compaq cluster at RAL would be introduced. The SP is using LoadLeveler that has already been tested on a four-processor IBM cluster at DL.

3.5 Other Data Management Projects

There are a number of other data management projects in which EPCC and Daresbury and RAL are partners over and above the PPARC HEC Challenge. Most of these projects use Grid technology in one form or another.

3.5.1 Development of an Interdisciplinary Round Table for Emerging Computer Technologies

DIRECT provides a forum for organisations representing all classes of users of large scale computing and data Centres to discuss future needs, see [H]. Together, they address the role of emerging computer technologies, such as HPCN, in defining the future of scientific computing. DIRECT has three major working groups: Data Storage and Management, Data Inter-Operability and Visualisation and Emerging Computing Technologies .

3.5.2 European Spatio-Temporal Data Infrastructure for High-Performance Computing

ESTEDI is a collaboration between numerous well-known research institutes in Europe, namely: Active Knowledge (D), FORWISS (D), University of Surrey (UK), CLRC (UK), CINECA (IT), DKRZ (D), DLR (D), IHPC&DB (RU) and NUMECA International (B). The project will establish a European standard for the storage and retrieval of multi-dimensional HPC data. It addresses a main technical obstacle, the delivery bottleneck of large HPC results to the users, by augmenting high-volume data generators with flexible data management and extraction tools for spatio-temporal data. The multi-dimensional database system RasDaMan developed in EU Framework IV will be enhanced with intelligent mass storage handling and optimised towards HPC. The project participants will operate the common platform and evaluate it in different HPC fields (Engineering, Biology, Astrophysics and Climate Research). The outcome will be a field-tested open prototype platform with flexible, standards-based, contents-driven retrieval of multi-terabyte data in heterogeneous networks. See [J] for more details.

3.5.3 Access Point for NERC Data Centres

The Earth Observation Data Group (EODG) provides services to several projects within the Space Science and Technology Department at Rutherford-Appleton Laboratory. These services range from advice on dealing with a particular dataset to a long-term commitment to support particular projects. The EODG is taking the lead in developing a meta-data system to link the on-line catalogues of data held at each of the NERC designated data centres. When complete, this will allow users to obtain answers to complex search queries. A first prototype can be accessed via the NERC web pages under [K].

3.5.4 ESDANET

The European Climate Data Network for Climate Model Output and Observations including the High Performance Mass Storage Environment (CLIDANET) is a collaboration of several high profile data and computing centres: DKRZ (D), UEA (UK), CCLRC (UK), FZI (D), PIK (D), CINECA (IT), CERFACS (F), CESCA (F), DMI (DK). The project plans to link existing mass storage data archives at European climate modelling centres and related observed data sources by a climate data network. The network will allow for fast and easy access to both categories of data, catalogue and climate data. The inter-comparison of climate model results and of observational data will be supported and encouraged.

3.5.5 Web Operating System (WOS)

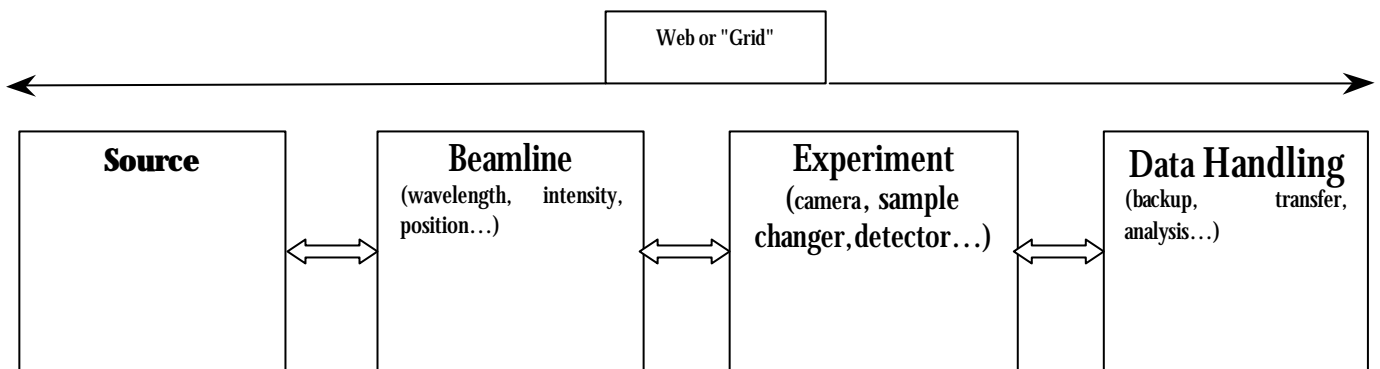
WOS is a component-based approach to demand-driven services provision on heterogeneous and dynamic resources. Communication replaces the notion of servers. Different versions of services and WOS may be running on a given network. Services are provided by "education engines" and "warehouses" connected by a discovery/ location protocol (WOSRP) [L]. There is also a generic services protocol (WOSP). This could be implemented as an interface to CML, CBL or CORBA. A proposal has been submitted to extend the WOS architecture developed in a previous EU supported project to include data management with Daresbury and other partners.

3.6 Experimental Facilities on the Grid

The development and evaluation of Grid-aware tools within the UK is timely since it coincides with the awakening within the scientific constituency to the kind of work they want to do and how they would like to do it. Thus there is potential for convergence of strands that at Daresbury are being exploited in two ways.

One approach is through collaboration between the Computational Science and Engineering Department and the Collaborative Computing Projects funded by EPSRC, PPARC and BBSRC. Two requirements emerging here concern access to high end computing facilities and the handling and analysis of time-resolved data collected from 2D detectors.

Our second thrust is to listen to the needs of the experimental scientists here on the Synchrotron Radiation Facility. For example, the protein crystallography community proposes to develop automation and remote operation for their data collection as illustrated below:



Such a system would need tight linkage between the components (which may not necessarily be geographically proximate) with maximum feedback between the stages through a Web, or Grid, infrastructure to enable go/ no-go decisions to be made interactively.

It is vital for the success of the Grid infrastructure that such requirements are given a high profile and we intend to work closely with the scientific community to achieve this.

4 Summary of UK Activities - RAL

4.1 HEP Applications

4.1.1 LHC Tier 1 Regional Centre

RAL provides large-scale computing resources for UK particle physicists. The next major development for this community will be the installation of a prototype Tier 1 Regional Centre for analysis of data from the CERN Large Hadron Collider (LHC). This will be part of an integrated computing infrastructure, based on a Grid model, involving many centres in a large number of countries, including Tier 2 centres in the UK, all linked by state-of-the-art networking. Subject to funding approval, the Tier 1 centre in the UK will be built up over the next three years to hundreds of CPUs (equivalent in power to thousands of today's CPUs) and hundreds of Terabytes of disk and tape. This resource will also act as a test bed for developing Grid technologies and user applications.

4.1.2 BaBar and CDF

In the meantime, Grid work has started with the UK users of current and near-future experiments like BaBar at SLAC in California and CDF at Fermilab. The BaBar experiment is now bringing a subset of data to the UK via RAL from where it is accessed by or copied to nine UK universities. They would like to use the Grid to locate data of interest, which may be located at any of the ten sites, and then either run a job at the relevant site to access it or replicate/cache it at the user's site or at a third site which has available computing resources. The distributed nature of their data, together with the possibility of multiple copies and the need for remote access and job submission, make this a good application for existing Grid technologies but due to their immaturity, the implementation will be challenging. The CDF experiment will use a similar model to BaBar starting in 2001, with additional equipment at Fermilab and RAL and a dedicated network link between the two sites to facilitate data access. This will enable QoS tests to be carried out over a dedicated wide area network. Ian Foster's group will be collaborating with these projects and Globus tools will be used.

4.1.3 Computational analysis

As part of the UKHEC programme, RAL has been investigating the use of Globus for computational analysis applications. Meta-computing based on Grid technologies enables distribution of a large simulation over a distributed set of parallel systems.

4.2 Other Applications

Since RAL provides many computing resources supporting a range of applications for external users, Grid techniques can offer a convenient way to provide a uniform user interface to the services at RAL and to others in the UK. Investigations have started with most services at RAL to define what form this grid interface should take. For some, like the Columbus super scalar service, remote job submission and data retrieval are obvious and these are being tested. For others, like the Atlas Datastore and RAL Visualisation services, there are several alternative solutions and projects have started to evaluate these.

4.3 RAL, Daresbury and Manchester Experiments

The RAL site has been registered with Globus and the Globus Security Infrastructure implemented on a range of services. Several batch services have been interfaced to Globus job submission including Columbus, a Linux Farm and a Beowulf PC cluster. Remote job submission between different RAL services and to/ from QMW have been carried out. An LDAP server has been implemented for Grid directory services but not yet integrated with the Globus MDS. A GASS server has been implemented. The bulk transfer of BaBar data has highlighted the limitations of standard transfer protocols. Experiments have started with a PVC across JANET to test QoS methods for optimising data transfer and work has also started on the robust file transfer methods required for reliable transfer of large datasets. QoS tests from Abilene and ESnet into JANET are underway. A managed bandwidth link between RAL and CERN is being set up.

4.3.1 RAL-DL experiments

Globus V1.1.1 has been installed on a four processor SPARC10 at RAL. After installation, some basic job submission and communication tests have been performed between this system, the Beowulf system at RAL and a Power PC cluster at DL. As many parallel and distributed applications are implemented using MPI, the Globus aware version of MPICH, MPICH-G, has been installed on all three systems and tested with several simple programs including the communications benchmark from the BECAUSE Benchmark Set (BBS 1.1.1). The bandwidth from RAL to DL, as measured through MPICH-G, is approximately 200 KBytes/s, with a latency of typically 10 ms to 50 ms. This compares with a bandwidth of 7.5 MBytes/s between processors on the SPARC10 system and 10 MBytes/s between nodes of the RAL Beowulf system using the same software. The three different job managers available across these systems (fork, PBS and LoadLeveler) have been tested and found to work together. One-site, two-site and three-site tests have been performed successfully using these MPI programs and the three different job managers. Distributed MDS servers, potentially possible in Globus V1.1.3, will be needed to overcome problems with non-availability of the US MDS server. Work is now continuing on two fronts, modifying some computational applications to run under the Globus system using these three systems and in reviewing other potential Grid and meta-computing systems such as UNICORE and Legion.

5 Rest of Europe

5.1 European Grid Forum

The European Grid Forum, EGrid, aims at fostering the co-operative use of distributed computing resources that are accessible via wide area networks. EGrid was formed in late 1999 and an organisational structure and a charter have been formulated. EGrid is an open forum - the community includes individuals from European research institutes, universities and companies working in the field of wide area computing and computational grids. EGrid members come from both worlds: the application-oriented end-users and the system software developers. In its early phase EGrid is meant as a discussion platform for interested parties in Europe. Multiple workshops are planned in the near future throughout Europe. Working Groups have recently been formed on topics including: Data Management; Resource Management; Testbeds; Programming Models; Performance Analysis. EGrid Web pages can be found at [M]. They also list a number of European middleware and application development projects.

5.2 EU R&D Projects

A number of EU projects are either already funded or in the proposal stages. These include (not inclusive): METHODIS, UNICORE, EuroGrid, WOS Systeme, ESTEDI.

5.2.1 METHODIS

The Metacomputing **T**ools for **D**istributed Systems project is an EU-funded project involving collaboration between HLRS (Stuttgart, Germany), CRIHAN (France), Pallas (Bonn, Germany), DASA (Germany) and Aerospatiale (France). The aim is to build an ATM-based meta-computing system for aerospace applications. Tools include COVISE, PACX-MPI and VAMPIR. This project runs alongside the UNICORE project to provide a seamless interface for submitting jobs to German regional supercomputers.

A large-scale implementation of these tools was demonstrated at SC'99 and involved Stuttgart, Manchester, Pittsburgh, San Diego and Tsukuba.

5.2.2 EuroGrid

The EuroGrid project has been funded by the EU as the IST strand of Framework 5 proposals. It is scheduled to begin in October 2000. The project is described in section 1.8 of this appendix.

5.2.3 UNICORE

UNICORE (**U**niform **I**nterface to **C**omputing **R**esources) was funded by the German Federal Government. It implements a complementary approach to that proposed by Globus. The idea is not to link machines as a single meta-computer but to provide a uniform and easy-to-use interface to machines on the Grid, making submission of jobs

transparent to the user. This approach has the advantage that it makes very few assumptions about the operating policies of the participating sites. Also, because the concept of a batch job is retained, it is almost certainly easier to implement accounting and billing users who submit work via the UNICORE interface. This is attractive as a funding and as a political model in a European context because it integrates while respecting national autonomy of the leading national centres. Also, the prospect of generating income via remotely submitted batch jobs gives an incentive for the large supercomputing centres to use this technology. The security arrangements of UNICORE are very stringent, making this an attractive option for industrial users of the Grid for whom issues of commercial security are paramount.

UNICORE is based on object-oriented Java technology and has an elegant abstraction of the concept of a job. The drawback of this approach is that it makes meta-computing applications more awkward to arrange because UNICORE's scheduling concepts are inherently asynchronous. Given the complementary nature of Globus and UNICORE and the strong financial and political backing of each, some rapprochement between them seems likely (see EuroGrid). As noted before, UNICORE's encryption technology is not US-derived and is free from US licensing conditions.

6 USA Grid Development Activities

The US is currently the main arena for the development of Grid technology and it was in the US that the term Grid was first coined. We give here a list of some of the major projects, this is by no means exhaustive but it gives a picture of the range of activities in the US.

6.1 US Grid Forum

The US Grid Forum is an informal consortium of institutions and individuals working on wide-area computing and computational grids: the technologies that underlie such activities as the NCSA Alliance's National Technology Grid, NPACI's Meta-systems efforts, NASA's Information Power Grid, DoE ASCI's DISCOM program, and other activities worldwide.

The Grid Forum is modelled, in many respects, on the Internet Engineering Task Force IETF [N] and focuses on the promotion of Grid computing via the documentation of "best practices" and "standards", with an emphasis on rough consensus and running code.

The Grid Forum is still in its early days and its terms of reference are still being defined. International participation is encouraged. So far the Grid Forum has selected a number of working groups and organised several workshops, including and "Birds of the Feather" meeting at SC'99. Working groups currently include: Scheduling; Grid Information Service; Security; Remote Data Access; Application and Tools Requirements; Grid Performance; Advanced Programming Models; Account Management; and User Services. These working groups emphasise the current active research areas where Information Technology input can provide additional grid software to the benefit of end application users. We will return to consider these areas of IT research at the end of this report.

The Grid Forum Web pages are at [O].

6.2 NSF PACI

The Partnerships for Advanced Computational Infrastructure (PACI) programme funded by the USA NSF Advanced Scientific Computing Division. PACI has established two national centres at the Universities of California at San Diego (NPACI) and Illinois (NCSA) which involve over a hundred academic institutions in collaborative HPC projects and costs around \$64M per year.

The core idea of the Alliance-PACI project is to establish a leading-edge computational grid, termed the "National Technology Grid", linking partner sites including NCSA and SDSC. Another project is termed the "National-Scale Machine Room".

6.2.1 NPACI

The National Partnership for Advanced Computational Infrastructure is based at the SDSC, University of California San Diego. It includes CalTech, U Texas, U Michigan, UC Berkeley, UC Davis, UCLA, UC Santa Barbara, U Houston/Keck, U Maryland and U Washington.

NPACI are developing the Legion software.

6.2.2 NCSA

The National Computational Science Alliance (the Alliance) is based at the University of Illinois at Urbana-Champaign. It includes OSC, UIUC, U Kentucky, UI Chicago, U Boston, Rice, Stanford, Princeton, MIT, U Wisconsin, U Minneapolis plus Allstate Insurance, American Airlines, AT\&T, Caterpillar, Dow Chemical, Eastman Kodak, Eli Lilly, FMC, F.P. Morgan, McDonnell Douglas, Motorola, Phillips Petroleum, Schlumberger Ltd., Sears, Shell Oil, Tribune Co., and United Technologies.

At SuperComputing'99 (Portland, Oregon, USA), Alliance research teams showed how the Alliance is developing a prototype virtual workspace, called the Access Grid that can be used for collaborative scientific research, distance education, and remote meetings and seminars. Some demonstrations utilised the Access Grid, connecting to remote locations either on the SC exhibit floor or in other cities. The Alliance also showed how its work in developing a national-scale technology Grid is enabling science and will exhibit new computational tools and infrastructure that are being integrated into the Grid. Projects of the Alliance Partners for Advanced Computational Services were also demonstrated.

NCSA are developing the GLOBUS software.

6.3 *GUSTO Consortium*

GLOBUS distributed and meta-computing concepts are being tested on a global scale by participants of the Globus Ubiquitous Supercomputing Test bed Organization (GUSTO). This is an agreement between PACI sites to develop a meta-computing test bed.

GUSTO is based at Argonne National Laboratory and the University of Southern California and started in 1997. It currently spans over twenty institutions and includes some of the largest computers in the world. Both dedicated and commodity Internet services are used.

GUSTO is further described in the GLOBUS pages [P] and [Q].

6.4 *NASA Information Power Grid*

The NASA IPG project is designed to implement seamless access to resources between NASA sites and a few NPACI sites. This followed from a number of workshops and reviews in the period autumn 1997. It grew from the Advanced Computing Networks and Storage (ACNS) and Computation Aerospaces (CSA) programmes at NASA. Goals of the project are to provide access to all resources for a single large simulation and to include virtual reality and access to large-scale data stores. A number of middleware implementations and demonstrator applications are being developed in phase II of the project starting in 3Q99 and continuing until 3Q04. The full project was planned to develop over a seven-year time scale and cost around \$63M per year.

Further information is available at [R]. There is also an "Information Power Grid Hotlist" from the NASA Web site which includes information on distributed computing, meta-computing and Java [S].

6.5 *ASCI Problem Solving Environment*

A major activity driven from Los Alamos National Laboratory is the PSE for the Accelerated Strategic Computing Initiative (ASCI). It also includes Lawrence Livermore National Laboratory and Sandia National Laboratory. Information on the ASCI projects is available at [T]. Components of PSE relevant to network-based computing are:

- **High Performance Computing Support:** High Performance Computing Support's (HPCS) role is to provide a supporting infrastructure between platforms and applications for effective high-end application execution and tera-scale data management.
 - Archival storage,
 - Scientific data management,
 - High speed interconnect,
 - Scalable I/O,
 - Distributed resource management
 - Platform and service integration;
- **Tri-Lab Networking:** Designing and implementing this wide/local area network architecture that enables uniform, transparent, and efficient distributed classified and unclassified computing among the three defence programs laboratories continues to be a formidable technical and administrative task that involves every aspect of networking.
 - Tri-lab connectivity,
 - New secure service and encryption upgrades,
 - Network modelling;
- **Distributed Computing Environment:** The purpose of the Distributed Computing Environment team is to provide a common set of key core services throughout the ASCI community, common both inter-organisationally (within a single lab) and between the ASCI computing environments at each of the three laboratories.
 - Production DCE core services,
 - Tri-lab distributed services/support,
 - DCE secure web pilot
 - Tri-lab DFS deployment,
 - DFS/HPSS integration testing,
 - Expanded desktop deployment,
 - Distributed objects,
 - Assessment study of PKI and DCE,
 - ASCI application support.

6.6 iGrid and StarTap

iGrid is a collaboration between University of Illinois at Chicago, Indiana University, Tokyo University and Keio University with the aim of "empowering global research community networking".

iGrid is part of the StarTap initiative. StarTap - Science, Technology, And Research Transit Access Point - is a persistent infrastructure, funded by the National Science Foundation Advanced Networking Infrastructure and Research division, which is part of the Computer and Information Sciences and Engineering (CISE) directorate, to facilitate the long-term interconnection and interoperability of advanced international networking in support of applications, performance measuring, and technology evaluations. The StarTap anchors the international vBNS connections program.

Physically, StarTap connects with the Ameritech Network Access Point (NAP) in Chicago, as does the vBNS and other high-speed research networks. It enables traffic to flow to international collaborators from over 100 U.S. leading-edge research universities and supercomputer centres that are now, or will be, attached to the vBNS or other high-performance US research networks.

StarTap is documenting the international collaborations it helps foster. These applications are among the most computation demanding and/or data-intensive today, and serve as test cases for the various network features StarTap deploys. Not only do these science applications help promote the exciting research being carried out worldwide, but they serve as a reference for others interested in computational science and engineering problems, or in the computer and communication technologies used to help solve them.

Major demonstrations have been held at SuperComputing'97, Alliance'98, iGrid'98 and one is planned at iGrid'2000. In the past these have included meta-computing using the Globus software and collaborative virtual reality using the CavernSoft software. In an example of the latter engineers at Caterpillar Inc. at NCSA were able to demonstrate a new tractor design to customers in Germany using an Immersadesk facility at GMD, Bonn.

Very informative Web pages are maintained at [U].

6.7 Illinois HPVM Project

The goal of this project, which started around 1997, is to develop shared controllable high-performance components for distributed systems. This includes predictable communication, management of heterogeneity, stable performance models and adaptive resource management. Virtual reality is supported using high-speed networking and an ability to manipulate large data sets. A variety of compute and networking components are being evaluated.

Software includes Illinois Fast Messages with APIs to MPI, SHMEM and Global Arrays, Dynamic co-scheduling resource management, FM-QoS heterogeneous communication layer and front-end administration tools using Java.

6.8 DoE2000 Programmes

6.8.1 National Collaboratories

The DoE2000 National Collaboratories are developing a set of tools and capabilities which will permit scientists and engineers working at different US Department of Energy and other facilities to collaborate on solving problems as easily as if they were in the same building. The programme supports research in tools that a virtual laboratory requires: collaborative tools; information surety (authentication plus security); and high-performance networking and one pilot implementation of these tools in partnership with other DoE programmes (e.g., ASCI).

6.8.2 Advanced Computational Testing and Simulation

The Advanced Computational Testing and Simulation programme is developing an integrated set of algorithms, software tools and infrastructure that will enable computer simulation to be used in place of experiments when real experiments are too dangerous, expensive, inaccessible or politically infeasible.

6.9 Data Management Grids in High-Energy Physics

Grid infrastructures may be, and are being, applied for data management enabling large data sets stored and indexed at remote sites to be analysed and re-used in inter-disciplinary projects.

Foremost in data grid developments are the particle physics, astrophysics and climate modelling communities. The first of these is stimulated by the imminent appearance of very large quantities of data from the CERN Large Hadron Collider (LHC). A large number of countries will participate in analysis of the data, with interacting grids organised as follows:

- Tier 0 -- CERN, Geneva where the ATLAS, CMS and other experiments will be run on the LHC;
- Tier 1 -- independent national centres in the USA and Europe;
- Tier 2 -- a number of regional centres in each country, probably deployed at universities or national laboratories;
- Tier 3 -- computing resources of an individual university group;
- Tier 4 -- an individual workstation.

This structure is typical of any grid organisation [BB].

6.9.1 US National Scalable Cluster Project

NCSP is developing a prototype meta-computing system including three university clusters in Illinois at Chicago, Maryland at College Park and Pennsylvania. Goals are to develop software and demonstrate scalable clustered computing enabling data transfer between geographically remote sites.

vBNS (very Broad Network System) is used to construct a fast network. It uses Asynchronous Transfer Mode protocols to achieve transfer speeds, sufficient to link nodes in local and wide area computing clusters, with the power to transfer Terabytes of data within minutes.

Other activities include data mining, data warehousing and medical supercomputing.

Project DataSpace is a five-year project that started in 1999 with the goal of establishing protocols and standards for high performance and distributed data mining. Protocols for mining distributed data were demonstrated at SC'99 and it was established that these protocols are effective for distributed workstation clusters connected with high performance networks (super-clusters) and with commodity networks (meta-clusters) see [CC].

6.9.2 US Data Analysis Grid

This project is planned to run during 2001-2005 to enable collaborative analysis of experimental data coming from the CERN LHC (see GriPhyN below). It is a joint proposal of the CMS/US ATLAS/LIGO experimental groups, involving individuals from Florida, FNAL, Northeastern, Caltech

The project aims to build an ensemble of Tier2 Centres, well coordinated with Tier1 Centres. It includes three projects on a shared network infrastructure.

6.9.3 GriPhyN

US physicists from the CMS and Atlas experiments at the Large Hadron Collider (LHC) at CERN, the LIGO experiment and the Sloan Digital Sky Survey are submitting a large-scale computing proposal to NSF the **Grid Physics Network**. The proposal is motivated by the fact that, whilst all four experiments have been approved for running and have received substantial construction funds, the computing needs of the US based physicists who will be analysing the data have barely been addressed.

The scale of the required computing resources is enormous: each of the four experiments requires enormous computing capacity and has a massive dataset (up to Petabyte size) that must be accessed by a widely distributed (international) user base served by networks having bandwidths that vary by orders of magnitude. Clearly, a computing solution for the four experiments requires dedicated, large-scale funding.

All four experiments receive considerable federal support, approximately \$1.4 Billion by the time the experiments come online (SDSS begins data taking in 2000, LIGO starts in 2002 while Atlas and CMS commence around 2005).

Whilst the LHC, LIGO and SDSS experiments plan to generate massive datasets, they are not unique. It turns out that many scientific and financial endeavours involve the rapid generation and analysis of large datasets. These include:

- The Earth Observing System Data Information System (EOSDIS) (3 PB by 2001);
- The Human Brain Project. Time series of 1 Terabyte scans of the human brain, generating of the order of a Petabyte of data in a short period of time;
- The Human Genome Project;
- Automated astronomical scans Geophysical data;
- Satellite weather image analysis, where chaotic processes are studied;
- Point of sale receipts, in which patterns of consumer spending are tracked;
- Banking records, which are analysed for spending cycles or unusual transactions which may relate to illegal activities.

The proposers are seeking funding based on previous work on large-scale distributed computing in the form of a computational grid. The problem of large computing and data resources being accessed by a large user base is the subject of several funded projects, namely GLOBUS, GIOD, Nile, and PPDG.

For LIGO, CMS and Atlas the one thing that stands out is the massive dataset that must be managed and accessed. While huge CPU resources must be used to analyse this data, the overwhelming problem is posed by the data itself - the proposed solution is to deploy a Data Grid.

It is assumed that each experiment will have one (or more) so-called Tier 1 computing centres within the US, e.g., Fermilab for CMS, Caltech for LIGO. For LHC these Tier 1 centres might have roughly 20% of the CPU and storage capacity available at CERN. The LHC Tier 1 centres are expected to have about 105 SpecInt95s in compute capacity and several PB in storage, along with perhaps several hundred TB in disk cache. The corresponding Tier 1 compute site for LIGO will be about an order magnitude smaller.

6.9.4 PPDG and HENP Grand Challenge

Several data grids are being constructed for analysis of experimental results that will come from the CERN LHC (see GriPhyN above).

PPDG is a DoE/NGI funded initiative involving US High Energy Nuclear Physics US laboratories FNAL, BNL, ANL, LBNL, SLAC, JLAB, and Universities Caltech and CACR, SDSC, Wisconsin (CS) are aiming to exploit expertise and existing tools for distributed data management; Globus, SRB, Condor matchmaking etc.

HENP Grand Challenge will be a DoE IT2/SSI project building on the PPDG work.

6.9.5 China Clipper Project

This is a US joint project between Argonne National Lab. (ANL), Lawrence Berkeley National Lab. (LBNL) and Stanford Linear Accelerator Center SLAC). It focuses on developing technologies for widely distributed data-intensive applications, mostly for particle physics experiment analysis. Software used includes a distributed parallel storage system, GLOBUS on the accessible networks. As well as demonstrating the feasibility of high data transfer rates to participating sites the project has developed network instrumentation, optimisation and debugging tools.

Work in the Clipper project is now being extended in the US Particle Physics Data Grid (PPDG).

7 Japan

7.1 JAERI/STA

The Center for Computational Science and Engineering (CCSE) was established within the Japanese Atomic Energy Research Institute (JAERI) in 1995. It is playing a leading role in the research and development of computational science and engineering in Japan. This is continuing the work started in the Science and Technology Agency (STA) and will continue to satisfy their requirements.

Principal strands of the research and development at CCSE are:

- Development of parallel basic software;
- Development of parallel algorithms;
- Development of parallel processing tools;
- Studies of numerical simulations on complex phenomena by particle and continuum methods;
- New computer architectures.

These feed into applications of special interest to the STA laboratories and Japanese Universities and software is available on JAERI and STA computers. Fortran 90 and MPI is used and the software is portable across many platforms including: Intel Paragon, Fujitsu VPP, Hitachi SR2201, Fujitsu AP3000, IBM SP, NEC SX4, Cray T90 etc.

A specific deliverable, relevant to this report, is the Meta-computing environment STAMPI and its associated tools. Further information is available from the Web site at [Z]

7.2 Real-World Computing Partnership

A project is under way to build distributed systems with shared resources. For further information contact K. Kubota et al., Real-World Computing Partnership, Tsukuba, Ibaraki 305-0032, Japan or see [AA].

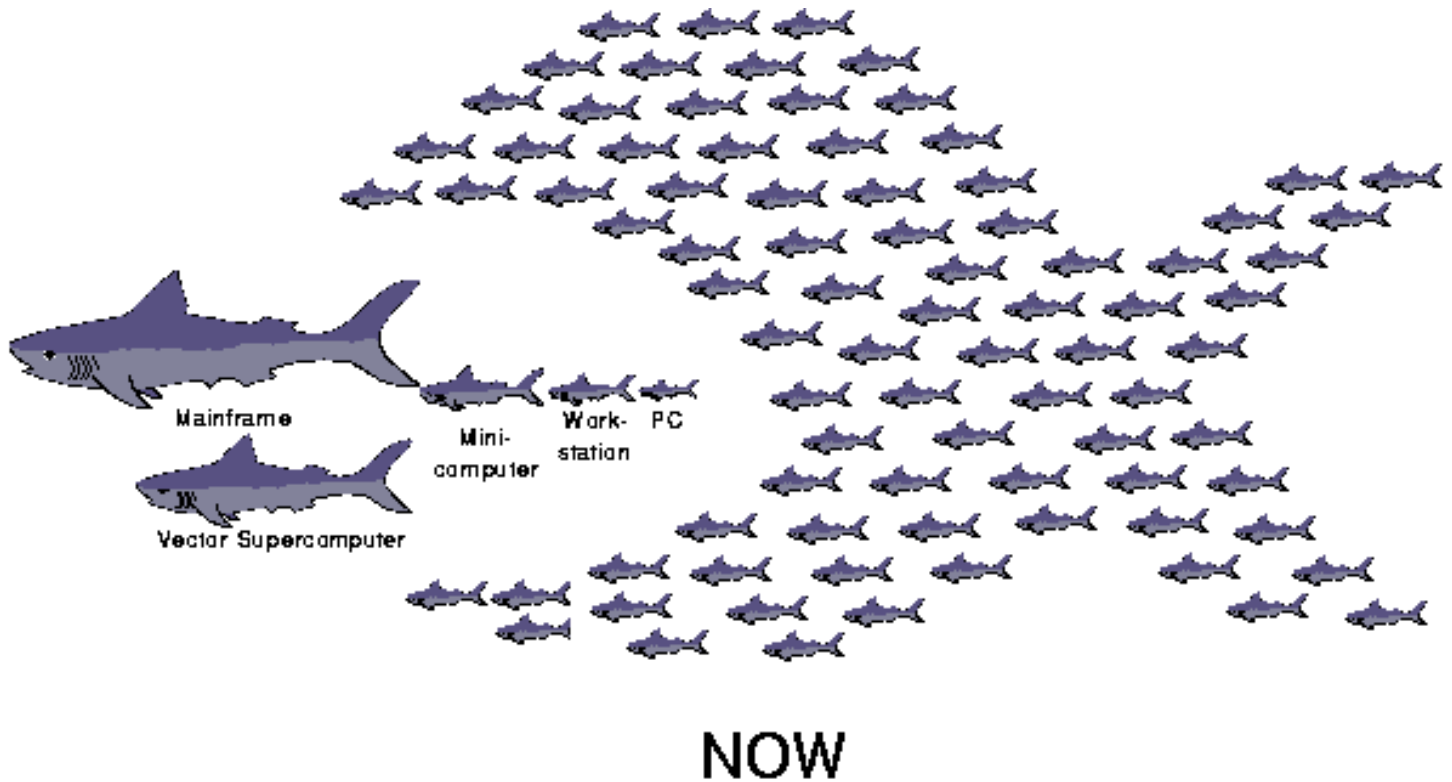
7.3 Waseda University Parallel and Distributed Computing Environment

The Parallel and Distributed Computing Environment Project is a project supported by the Japanese government through the Japan Society for the Promotion of Science. Its objective is to develop a parallelising restructuring compiler and related tools for parallel and heterogeneous distributed computing environment. The project puts equal emphasis on both practical and theoretical sides. To pursue the project, they built a network of high performance computers as a research infrastructure. The final goal of the project is to build a "super Grid", which can be interconnected to an international Grid.

URLS

- [A] <http://www.man.ac.uk/MVC/research>
- [B] <http://www.unicore.de>
- [C] <http://www.man.ac.uk/MVC/projects/NOVICE/>
- [D] <http://www.csar.cfs.ac.uk/news/grid>
- [E] <http://www.dci.clrc.ac.uk/Activity/DAMP>
- [F] <http://www.dci.clrc.ac.uk/ActivityPublications/226>
- [G] <http://www.dci.clrc.ac.uk/Activity/DAMP+958>
- [H] <http://www.dci.clrc.ac.uk/Activity/DMC>
- [I] <http://www.epcc.ed.ac.uk/direct>
- [J] <http://www.estedi.org>
- [K] <http://www.nmp.rl.ac.uk>
- [L] <http://www.wos-community.org>
- [M] <http://www.egrid.org>
- [N] <http://www.ieft.org>
- [O] <http://www.gridforum.org>
- [P] <http://www.globus.org>
- [Q] <http://www-fp.globus.org/testbeds>
- [R] <http://www.nas.nasa.gov/Groups/Tools/IPG>
- [S] <http://www.nas.nasa.gov/NAS/Tools>
- [T] <http://www.lanl.gov/asci>
- [U] <http://www.startap.net>
- [V] <http://www.dmsi.mil>
- [W] <http://www.ca.metsci.com>
- [X] <http://dsl.cs.uchicago.edu.beta>
- [Y] <http://now.cs.berkeley.edu>
- [Z] <http://jaeri.go.jp/english/index.cgicomp/comp.html>
- [AA] <http://www.rwcp.or.jp/lab/mpperf>
- [BB] <http://www.gridforum.org/iga.html>
- [CC] <http://nscp.upenn.edu>
- [DD] <http://www.phys.ufl.edu/~avery/mre>

The Berkeley NOW Project



"What can anyone give you greater than now"
- William Stafford

The Berkeley Network of Workstations (NOW) project seeks to harness the power of clustered machines connected via high-speed switched networks. By leveraging commodity workstations and operating systems, NOW can track industry performance increases. The key to NOW is the advent of the killer switch-based and high-bandwidth network. This technological evolution allows NOW to support a variety of disparate workloads, including parallel, sequential, and interactive jobs, as well as scalable web services, including the world's [fastest web search engine](#), and commercial workloads, such as NOW-Sort, the world's [fastest disk-to-disk sort](#). On April 30th, 1997, the NOW team achieved over 10 GFLOPS on the [LINPACK](#) benchmark, propelling the NOW into the top 200 fastest supercomputers in the world! Click [here](#) for more NOW news. The NOW Project is [sponsored](#) by a number of different contributors.

● [Project Overview](#)

Brief overview of the NOW project, including the [Case For NOW](#)

Click [here](#) for the latest **NOW news!**

● [Research Topics](#)

Research on High-Speed Communication, Operating Systems, File Systems, The Web, Programming Environments, and Applications

● [Papers and Slides](#)

Paper topics include [General NOW](#), [Fast Communication](#), [Distributed Operating Systems](#),

[Scalable File Service](#), [High-Performance Applications](#), [Architecture](#), and [NOW on the Web](#).

Click [here](#) for **recent NOW papers**, and [here](#) for **NOW dissertations!**

- [Software Release](#)

NOW Software is available: AM-2, Glunix, MPI-AM2, NameServer.

- [Project Information](#)

Pictures, People, Sponsors, Private Working Directory

- [Tutorial - How to use the NOW Cluster](#)

How to use [AM-II](#) and [MPI](#)

[AM-II Network Map](#)

- [Retreat Information](#)

All you need to know about NOW retreats

[\[Top Level\]](#) [\[Overview\]](#) [\[Research\]](#) [\[Papers & Slides\]](#) [\[Information\]](#) [\[Tutorial\]](#) [\[Software\]](#) [\[Log\]](#) [\[Retreat\]](#)

This page is casually maintained by remzi@cs.berkeley.edu